Proposal for LHCb RICH detector enhancements during LHC Long Shutdown 3

Edited by: Rafael Ballabriga, Carmelo D’Ambrosio, Sajan Easo, Floris Keizer*, Alessandro Petrolini and Steve Wotton

Abstract

The prompt Cherenkov radiation and focusing optics of the RICH detectors result in time characteristics which are unique among large-volume detector systems. The time-of-arrival at the photon detectors of the Cherenkov photons corresponding to a given primary vertex can be predicted to within ten picoseconds. This property can be used to significantly improve the signal to noise ratio and thereby the PID performance of the detector and will ultimately allow the present system to withstand luminosities in excess of $10^{34}$ cm$^{-2}$s$^{-1}$. To this end, we propose to integrate a new readout ASIC, the FastRICH, into the present system during the Long Shutdown 3 (LS3, 2026-2028). This will allow the system to timestamp each photon with a $\sim$ 150 ps time resolution within a short gate of $\sim$ 2 ns. This enhancement can be achieved at a limited cost, prepares for the Upgrade II RICH system overhaul and improves the hadronic PID performance for the physics programme of LHCb during Run 4.
1 Introduction

High quality particle identification (PID) is essential for almost all precision flavour measurements [1]. The ring-imaging Cherenkov (RICH) system consists of the upstream RICH1 detector with $C_4F_{10}$ radiator with a kaon low-momentum threshold of $\sim 10\text{ GeV}/c$ although identification of heavy particles was possible down to $\sim 2\text{ GeV}/c$ in veto mode [2].

The downstream RICH2 detector with CF$_4$ radiator covers the higher momentum range with good $\pi$-K separation up to $\sim 100\text{ GeV}$. As outlined in [1] and the references therein, the RICH system for the future LHCb detector will be a natural evolution of the current detectors. During Long Shutdown 4 (LS4), currently scheduled for 2033, the LHCb detector will undergo a major overhaul called Upgrade II to be able to maintain its excellent performance at the $\sim 1.5 \times 10^{34}\text{ cm}^2\text{ s}^{-1}$ luminosity during Run 5, which is a factor 7.5 increase in luminosity with respect to Run 3 and 4, with the aim of integrating $\sim 300\text{ fb}^{-1}$ throughout the HL-LHC era. This will require a substantial improvement in the precision of the measurements of the space and time coordinates of the detected photons. LS3 scheduled for the period 2026-2028 provides an opportunity to consolidate the Run 3 RICH detectors with the aim to (a) enhance the performance during Run 4 and to (b) introduce concepts and detector changes that are key in preparation for the major LHCb Upgrade II during LS4.

The addition of a Cherenkov photon timestamp with a resolution of better than 100 ps is one of the cornerstones to maintain the performance of the RICH detectors in the HL-LHC Run 5 environment. Additionally, the photon detector needs to have improved radiation hardness, channel density and bandwidth efficiency whilst remaining compact with a low power consumption. These challenging requirements will require a new front-end detector design. The proposed LS3 enhancements target specifically the front-end ASIC and the interface to the optical links. This way an important conceptual and technological foundation is laid for the future while most of the Run 3 detector and

![Figure 1: Schematic of the electronic readout chain evolving from Run 3 to Run 5 showing the future FastRICH ASIC with data-compressed (DC) outputs containing fast-timing information.]

<table>
<thead>
<tr>
<th>Sensor</th>
<th>ASIC timewalk</th>
<th>FE time gate</th>
<th>TDC time bin</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHC Run 3</td>
<td>150 ps</td>
<td>&lt; 4 ns</td>
<td>None</td>
</tr>
<tr>
<td>LHC Run 4</td>
<td>150 ps</td>
<td>CFD correction</td>
<td>2 ns</td>
</tr>
<tr>
<td>HL-LHC Run 5</td>
<td>$\sim 50\text{ ps}$</td>
<td>CFD correction</td>
<td>2 ns</td>
</tr>
</tbody>
</table>

Table 1: Overview of detector time resolutions from Run 3 to Run 5. The HL-LHC will begin operation at the start of Run 4 but the interaction rate at LHCb will not be increased to allow Upgrade I to complete its programme.
services remain unchanged.

The LHC Run 3 photon detector chain, represented in Figure 1, consists of multi-anode photomultiplier tubes (MAPMTs) read out by the CLARO ASICs on front-end boards (FEBs) [6]. The digital board provides the interface between the CLAROs and the Versatile Links (VLs) to the LHCb readout. The FPGAs capture the digital signals, format the data and transmit them using GBTx transceiver ASICs on the optical link plugins (labeled DTM for data transmission module and TCM for trigger and control module in Figure 2). The programmable FPGA logic samples the CLARO signals at 320 Mb/s using the deserialiser embedded in every input-output logic block. Towards the future, this Run 3 approach is limited by (a) the clocking resources and (b) radiation hardness of the design. Increasing the sampling rate beyond ~ 1 Gbit/s would require a redesign of the digital board and higher performance FPGAs. The poor radiation hardness of FPGAs requires minimal use of logic resources already during Run 3 in order to lower the probability of radiation upsets [3]. A limited number of logic failures is expected during Run 3 [7]. For this reason, the addition of multi-channel time-to-digital (TDC) converter logic in the FPGA is not feasible, and the increase in luminosity and radiation fluence in Run 5 will require an ASIC solution. The proposed enhancements during LS3 anticipate this shift towards a highly integrated TDC ASIC. The proposed multi-channel FastRICH

![Figure 2: Design of the Run 3 photon detector column showing in the yellow shaded region the electronics to be replaced during the LS3 enhancements [3–5].](image-url)
chip will perform multi-channel discrimination, apply data-compression techniques and timestamp each hit with $\sim 25$ ps time bins. As represented in Figure 1, the ASIC becomes the single active component between the photon sensor and the next-generation optical links resulting in a simplified readout scheme. Data compression (labelled 'DC' in the figure) at the front-end is one of the techniques that will be used to reduce the bandwidth, which is a key challenge arising from the higher luminosity and added timing information.

The highlighted region in Figure 2 includes the FEB and digital board, which would both be replaced during LS3 without changing the remainder of the photon detection column and infrastructure. As a result, the cost for this consolidation would be relatively low compared to Upgrade II during LS4. The anticipated duration of LS3 of around 3 years starting in 2026 would seem to be an ideal opportunity to improve the detector.

Section 2 of this document describes the use of fast-timing information in the event reconstruction to improve the RICH PID performance. The photon sensor and housing, which are not modified during LS3, are described with the emphasis on time information in Section 3. The novel FastRICH ASIC is introduced in Section 4 followed by the redesigned digital board and optical links in Section 5.

2 Event reconstruction including time

Simulation studies show that the proposed photon time information can significantly improve the PID performance, particularly by reducing combinatoric background and dark counts. As demonstrated in [8], for a given track the time-of-arrival (ToA) of Cherenkov photons at the detector plane can be predicted to within 10 ps. This prediction includes the LHCb tracking information, the reconstructed photon paths in the RICH detector and the primary vertex (PV) time often referred to as the ‘t-zero’. While the tracking resolution and RICH spatial resolution remain the same from Run 3 to Run 4, the addition of photon timestamps adds a new dimension to the RICH detector and its estimate of the PV t-zero would be a first for the LHCb experiment. The development of algorithms including time information is an important step towards Run 5, where time information will be a necessity rather than an enhancement. The gathered experience on the recording and processing of time information during Run 4 will make this a seamless transition.

2.1 Use of time in software and hardware

The RICH PID is performed using a log-likelihood algorithm in the LHCb high-level trigger (HLT) [2]. A ‘photon object’ is created for each combination of a track and RICH hit that is allowed within spatial constraints posed by the various mass hypothesis. Each of these objects is a candidate for a true combination and it contains all the required information regarding the track and the hit, including what is needed to reconstruct the corresponding Cherenkov angle. This information is also accessible to calculate the time-of-flight of the tracks and the photon candidates. This makes it possible to predict when the hit is expected for each photon object. In software, the measured detector hit timestamp associated to the photon object can be directly compared to the prediction by means of a time gate of width $\Delta t_{\text{gate}}$. If a detected hit is more than $\frac{1}{2} \Delta t_{\text{gate}}$ from the predicted hit time, then the photon object is considered to correspond to a fake combination and it is removed from the log-likelihood calculation. This improves the PID
as demonstrated in Section 2.2 and reduces the number of calculations and CPU time required by the algorithm. The width $\Delta t_{gate}$ is determined by the photon sensor and readout electronics resolutions.

Although alternative methods of using time information in the reconstruction algorithms are being investigated, they would be mostly focused on PID by means of (combinatoric) background reduction. Considering these RICH detectors aim for PID at relatively high momenta, the variations in the time-of-flight between different particle types at such momenta are generally too small for the hit time information to contribute directly to the discrimination between the different mass hypothesis.

In addition to the aforementioned software time gate, a hardware shutter can be implemented in the front-end electronics with the additional benefit of reducing the bandwidth and power consumption. However, while the software gate can benefit from the reconstructed time information per photon to be adjusted to a narrow time window, the hardware shutter has to register Cherenkov photons from the entire duration of the bunch crossing and therefore must have a wider window. Taking into account the PV time spread and the range of possible track and photon paths through the experiment, a typical shutter time of $\Delta t_{shutter}$ of 2 ns is envisaged. Within this period, each photon is timestamped with high resolution in order to apply a refined software time gate.

### 2.2 Expected performance

The PID curves in Figures 3 to 4 provide a standardised and sensitive probe of the RICH performance, where the bottom right corner represents the ideal performance of high kaon identification and low pion misidentification [2]. Generally, there is a clear trend when it comes to the use of timing information: technologies with improved time resolution allow narrower software time gates to be applied, which improves the PID performance [8]. This was demonstrated using the LHCb simulation framework 1, simulating the Run 3 detector with added timing information at a luminosity of $2.0 \times 10^{33}$ cm$^{-2}$ s$^{-1}$. As a proof of principle, Figure 3 assumes the ideal photon detector with zero time resolution and a range of different software time gates $\Delta t_{gate}$ from 25 ns to 20 ps [9]. In practice however, the PID curve that can be achieved depends on the sensor and readout electronics resolution. For Run 3, the readout electronics allow $\Delta t_{shutter} = \Delta t_{gate} = 6.25$ ns.

The additional hit timestamps of the proposed LS3 enhancements would allow a smaller time gate to be chosen. Adding the MAPMT transit time spread (TTS) $\sigma_{mapmt} \sim 150$ ps (for details refer to Section 3) to the simulated curves in Figure 4 shows that the optimal time gate for Run 4 corresponds to $\pm 2\sigma \sim 600$ ps. At narrower time gates the loss of signal photons from the gate degrades the performance and at wider gates the background reduction is less. The studies were performed at two luminosity scenarios of $2.0 \times 10^{33}$ cm$^{-2}$ s$^{-1}$ (Figures 3 and 4) and $3.0 \times 10^{33}$ cm$^{-2}$ s$^{-1}$ (Figure 5). In both scenarios, the PID curves clearly show the same trend that a 600 ps time gate applied for Run 4 results in a performance improvement through the reduction of combinatoric background. While the absolute PID performance depends on a range of parameters and can therefore change, the trend of improved PID performance using time gating is robust and a strong motivation for the proposed LS3 enhancements.

Using the samples produced for Figure 5, an initial verification of the effect of the

---

1LHCb detector simulation package: Gauss v54r5
Figure 3: PID curves at $\mathcal{L} = 2.0 \times 10^{33}$ cm$^{-2}$ s$^{-1}$ and ‘ideal’ photon detector time resolution, for software time gates ranging from 25 ns to 20 ps.

Figure 4: PID curves at the same conditions as Figure 3 but introducing $\sigma_{mapmt} \sim 150$ ps.
Figure 5: PID curves at $L = 3.0 \times 10^{33} \text{cm}^{-2}\text{s}^{-1}$ and $\sigma_{\text{mapmt}} \sim 150 \text{ps}$, comparing 6.25 ns (Run 3) and 600 ps (Run 4) time gates.

proposed LS3 enhancements on the signal to background ratio [10] was performed. The results from this study already indicate that the use of a time resolution of 150 ps allows an improvement up to 50% on the signal-to-background ratio, depending on the final state, the various criteria used for data selection and assuming the same HLT retention rates between Run 2 and Run 3.

2.3 Primary Vertex time

The PID curves in this document were produced with the PV t-zero from truth-matched simulated events. During Run 4, the RICH reconstruction algorithm can instead be used to estimate this t-zero. Preliminary simulation studies have shown that it is feasible to estimate the t-zero using the RICH reconstruction, especially since the Run 4 luminosity will be at a level similar to Run 3. Here, the 3D spatial reconstruction can perform sufficiently well to associate Cherenkov photons to each track. The t-zero for the tracks and PVs can then be deduced and used to apply the software time gate. While the best results are expected from running the likelihood maximisation twice (once to obtain the most likely photon-to-track associations to calculate t-zero and again after the subsequent software time gate), this would also significantly increase the computation time in the high-level trigger. Therefore studies are seeking for quick indicators of correct photon-to-track association, where for example the amplitude of the probability density function (PDF) calculated at the stage prior to the likelihood maximisation is a promising indicator of a true photon object. By applying a cut on the PDF amplitude, a subset of photon objects can be selected for t-zero determination. Since a multitude of photons can be matched to each track, and multiple tracks to each PV, the error on the PV t-zero is reduced significantly. The principle here is that when photons can be correctly associated
to tracks, the track time resolution scales with $\sqrt{N}$ with $N$ being the number of Cherenkov photons. Preliminary results have shown a resolution better than 100 ps at least for a large subset of PVs. During HL-LHC Run 5, it is critical that the PV time will be provided as part of the LHCb tracking information, because the order-of-magnitude increase in PV and track multiplicity will mean that the RICH detector can no longer estimate the PV times. Instead, the reconstruction must receive these as an input for the application of software time gates.

The additional uncertainty has the effect of applying an offset to the centre of the software time gates. The 600 ps curve in Figure 4 is therefore a best-case scenario. The optimisation of the width of the time gate in relation to the t-zero resolution that can be achieved is part of the foreseen software R&D.

3 Photon sensor and elementary cell

An exploded view of the Run 3 elementary cell (EC) was shown in Figure 2 [11]. The EC consists of MAPMTs with magnetic shielding that plug into a baseboard with bias voltage divider. The FEBs contain CLARO ASICs which are connected through the backboard to the digital board. The metal casing gives the EC mechanical support and provides safety ground and thermal paths. For the LS3 enhancements, the MAPMT sensor with its baseboard and casing are unchanged and its time resolution has therefore been studied in detail in the lab and test beam setups. The remainder of the column infrastructure, such as the cooling, mechanics and power distribution will also remain the same.

The MAPMT transit time spread (TTS) is specified in the Hamamatsu datasheets as $\sigma_{mapmt} \sim 120$ ps for the 2-inch R12699-406-M64 and 150 ps for the 1-inch R11265-103-M64 MAPMT models [12]. This has been verified by directly coupling a 12 GHz oscilloscope with active probe to the MAPMT pins with single-photon light input from a picosecond pulsed laser source. A fixed threshold was applied below the single photoelectron peak and the ToA of the signal was registered as shown in Figure 6 for the 2-inch device, where a Gaussian fit to the peak gives $\sim 140$ ps. Similarly, for the 1-inch MAPMT a
TTS of \(\sim 160-190\) ps was measured where the upper limit corresponds to the MAPMT with signal-induced noise mitigation \([13]\). An \(\sim 15\%\) improvement was observed using constant-fraction discrimination (CFD), resulting in \(\sim 120\) ps for the 2-inch MAPMTs and \(\sim 135-160\) ps for the 1-inch MAPMTs as expected from the datasheet. As part of the simulation R&D studies, the optimal use of the better time resolution of RICH2 than RICH1 will be studied.

The CLARO ASICs in the Run 3 EC are operated at \(0.7\) mW/ch and \(\sim 80\) ps RMS jitter \([6]\). Figure 7 shows the measured time walk in the CLARO using a test charge input, where the error bars represent channel variations \([9]\). The time walk reaches up to \(\sim 4\) ns, which could be corrected using a time-over-threshold (ToT) measurement. The CLARO is optimised specifically for MAPMT signals (with negative polarity) in 350 nm technology.

In the Run 4 scenario, the FEB is replaced by a new FEB containing the FastRICH ASIC. Time walk can be corrected using CFD. A successful implementation of CFD in the FastRICH would therefore mean that no ToT information needs to be transmitted and the hardware shutter time can be narrower than that for the CLARO, which reduces bandwidth requirements. The FastRICH would allow positive or negative input signal polarities and has a wide dynamic range, which meets the long-term goal to read out faster SiPM or MCP-type sensors. A FastRICH optimised for only negative input signals is also being explored although this would only allow SiPMs to be read at the cathode. The FEB replacement would require minor modifications to the resistor configuration at the backboard. A simplified layout without backboard can also be considered as part of the R&D for the new boards arrangement.

### 4 Front-end ASIC

In this section, the criteria for the future RICH readout ASIC are established. These cover a range of requirements such as bandwidth reduction, power consumption and radiation hardness in addition to the timing performance. A major requirement of this novel readout chip will be its integrated time-to-digital converter (TDC) with bin size of \(\sim 25\) ps in order to match the (future) sensor resolution. A candidate ASIC known as the FastRICH is currently under development in a collaboration between Barcelona and CERN microelectronics and with input on the ASIC specifications from the RICH group.

#### 4.1 Requirements

While the timing requirements for Run 4 are related to the MAPMT TTS discussed in Section 3, the foreseen improvements in sensor time resolution during Upgrade II are also anticipated in the ASIC design. Additionally, the use of time information in the RICH reconstruction as demonstrated in Section 2 is on a per-photon basis. The timing specification is therefore defined in terms of electronic readout bin width. Typically, the best performance was observed with a \(\pm 2\sigma_{\text{sensor}}\) symmetric time gate around the expected ToA. Since the ToA is not synchronised to the electronic readout bin edges, this means that in most cases one or more neighbouring readout bins may also need to be read out and larger readout bins may result in an unnecessarily wide time gate. For example for a future 50 ps sensor, the ideal time gate would be centred on the expected time with a width of 200 ps. To achieve this, an order 100 ps or better electronic readout bin size is...
ensitigated. Furthermore, oversampling of the TTS distribution using the proposed TDC resolution would be beneficial to find the distribution of photons and to apply a fit to obtain the PV t-zeros.

During HL-LHC Run 5, the radiation fluence will increase by an order of magnitude requiring more robust readout electronics. As the FPGAs on the Run 3 digital boards are expected to fail at these radiation levels, a shift of technology to an ASIC solution is foreseen for Run 4 and 5 [3]. The ASIC is required to withstand a fluence in RICH1 of $\sim 10^{13} \text{n}_{\text{eq}}/\text{cm}^2$ and around 5 kGy for a total integrated luminosity of $350 \text{fb}^{-1}$ [1]. Furthermore, it should be compatible with the lpGBT and VL+ radiation-hard chipset used for the LHCb optical readout.

While the novel ASIC will have additional timing and digital functionality compared to the Run 3 CLARO chip, its power budget has to remain similar and at the level of a few mW per channel. More details on the power consumption and the comparison between the novel ASIC and the ‘CLARO + FPGA’ combination will be discussed in Section [5.1].

The photon occupancy across the photon detector plane is highly non-uniform, especially in the RICH1 detector where the average occupancy is $\sim 5\%$ while the peak occupancy reaches about 40%. In the low-occupancy regions, data compression in the ASIC will be important to maintain the bandwidth during Run 4 at a similar level to Run 3 whilst adding hit time information. In order to be cost effective, this non-uniformity requires the ASIC to have a configurable number of output links. In the high-occupancy region, all links will be connected to the lpGBT while in the low-occupancy regions, some of the output links can be disabled in the ASIC and left unconnected on the digital board, resulting in a smaller number of optical link components on the board. The number of input channels per chip is a balance between a complex PCB design to route many analogue signals to the same ASIC and cost and chip density for fewer channels. The optimal is expected at 16 channels per chip.

In addition to zero-suppression, the ASIC should implement a configurable shutter time of $\Delta t_{\text{shutter}} \sim 2\text{ ns}$ as well as CFD in order to keep the bandwidth to a minimum. The shutter time with configurable width and offset with respect to the LHC clock is

Figure 8: Elementary cell with FastIC ASICs mounted on the Front-End Board first used during test beam studies at the SPS facility.
used to exclude hits from data that arrive outside the few nanoseconds window associated with the signal from the bunch crossing. The shutter reduces the required TDC range, which decreases the power consumption in the chip (since the oscillator is only active for the duration of the conversion) and the number of bits to encode the hit timestamp. In order to compensate for time-walk errors in the ASIC, the ToT information or CFD needs to be used. Since the ToT information would require additional bits to be transmitted to the back-end, where the time-walk correction can be made, CFD is preferred as it corrects for time walk directly at the front-end. While CFD reduces bandwidth, it adds complexity and power consumption in the ASIC. Studies are being carried out to assess both techniques before a decision will be taken.

The possibility of an additional threshold in the ASIC to discriminate multiple photons by amplitude could be an advantage in particular for Upgrade II with an order of magnitude higher photon detector occupancies. It may therefore already be a feature of the ASIC during Run 4.

4.2 FastIC, FastIC+ and FastRICH ASICs

The Fast Integrated Circuit (FastIC) is an ASIC designed in 65 nm CMOS technology for the readout of precise timing detectors with intrinsic signal amplification [14]. The FastIC is developed by ICCUB and CERN-EP-ESE based on the architecture of the HRflexToT chip [15]. The ASIC forms the basis for the FastIC+ development and the FastRICH customisation for the RICH upgrades. The first available FastICs have been successfully mounted on custom FEBs and tested during the LHCb RICH beam tests at the SPS facility at CERN in October 2021. An EC with a 2-inch type MAPMT and FastIC FEB is shown in Figure 8. The FastIC was also coupled to a 1-inch MAPMT and an SiPM array and the recorded timing data are being analysed (with a paper in preparation).

The 8-channel (single-ended) FastIC reads the signals delivered by a sensor and processes these in a current mode method, the low frequency input impedance being $\sim 16\,\Omega$. As shown in Figure 9, the precise timing input stage sends each incoming signal to a fast current discriminator which compares the signal with a programmable threshold, which can be set at a level below the single photoelectron signal. The leading edge of the signal at the comparator output retains the signal ToA information. In this non-linear time-over-threshold mode, the maximum rate is above 50 MHz depending also on the details of the input pulse shape. The power consumption for the time output branch is approximately 6 mW per channel at 1.2 V.

The input stage can be programmed to work in both positive or negative polarity and covers a wide dynamic range of 5 uA to 20 mA. This provides important flexibility to couple the ASIC to different sensor types such as MAPMTs with a negative input signal or SiPMs with a positive signal and larger input capacitance during future upgrades. Using a 330 pF 3x3 mm$^2$ SiPM at 4.5 V overvoltage the FastIC has been measured to contribute no more than 31 ps RMS degradation to the single photon time resolution of the sensor. The FastIC has also been tested not to degrade the pulse shape or time response for PMT-based detectors [14]. The ASIC is a 65 nm technology and is therefore expected to be intrinsically more radiation hard for TID than for example the 350 nm Run 3 ASIC technology.

The FastIC can be programmed to output the signal for each channel or to combine 4 channels and sum their signals before processing. This active summation functionality
is integrated to explore the impact of segmenting large SiPMs (with large capacitance) into smaller ones to achieve lower jitter while covering large detector areas. The output driver can be configured either in CMOS single-ended or differential SLVS mode. The configuration interface is based on I2C and therefore compatible with the lpGBT chipset for example.

The FastIC+ is a next-generation ASIC based on the FastIC developed by the micro-electronics section at CERN and the University of Barcelona with medical applications being one of the main targets. Developments are ongoing to add to the FastIC chip a TDC with $\sim 25\text{ ps}$ time bins based on a ring-oscillator as shown in Figure 10. The FastIC+, similar to the original FastIC, will not be designed to be radiation hard although the 65 nm technology is expected to be radiation hard in terms of TID. The digital logic does not use triple redundancy techniques. While the chip is therefore suitable for prototype applications in HEP, it is not designed to be used in HEP experiments with high irradiation.

The ongoing design of the bespoke FastRICH chip takes into account the specific requirements of the RICH detector as specified in Section 4.1. The specifications of the FastRICH are detailed in Appendix A. The FastRICH will be radiation hard by design. Similar to the FastIC+, the design includes a TDC with $\sim 25\text{ ps}$ time bins. The ring oscillator will be started and stopped for the duration of the required RICH shutter time to reduce bandwidth and power consumption. The existing FastIC front-end will be modified in order to include the CFD and data compression. The possibility of a branch with a second threshold level is being studied in order to distinguish single and multiple photons per channel. A TDC bin width of $25\text{ ps}$ with CFD results in a 7-bit timestamp over a 2 ns shutter time and the 16-channel ASIC would require 4 additional bits to identify the channel number. The ASIC, with up to four configurable outputs, will be designed for

![Figure 9: Block diagram of the FastIC ASIC channel when programmed in a single ended architecture. The FastIC can be configured to select time, energy and trigger channels. Only the time channel with the fastest response and non-linear ToT information is used for the RICH application.](image-url)
maximal compatibility with the lpGBT (and if possible GBTx) readout. Additionally, the ASIC will need to be compatible with the architecture and data transmission protocols of the Upgrade II data acquisition system. Work is ongoing to define the additional requirements for data synchronisation across the LHCb apparatus as well as the detailed specification of the data and controls interfaces.

The more generic FastIC+ and the custom FastRICH projects will share resources in terms of the digital design in order to optimise resources. Although the RICH requirements drive its design, the FastRICH ASIC meets a more general demand for radiation-hard TDC readout ASICs for a range of sensor types in HEP. Hence there is significant scope for synergy with other sub-detectors and wider applications. In the future, a pixel detector consisting of a 2D channel array tightly integrated with SiPMs can be envisaged.

5 Digital board

A new digital board will be installed during the proposed LS3 enhancements. Similar to the Run 3 design, it is proposed that the board would connect to the backboard shown in Figure 2 and contain plugins with lpGBT or GBTx ASICs (labelled DTM for ‘data transmission module’ and TCM for ‘trigger control module’ in Figure 2). The FPGA will no longer be present for Run 4. The controls and signals from the FastRICH outputs on the FEBs would be directly routed to the optical plugin, which means that the motherboard is simplified and lower cost than the Run 3 version. In order to save overall bandwidth and component costs, two or more versions of the motherboard can be produced with more plugins in the high-occupancy photon detector regions and less plugins in the low-occupancy regions.

![Schematic of the additional TDC readout block for the FastIC+ and FastRICH chip developments.](image)

Figure 10: Schematic of the additional TDC readout block for the FastIC+ and FastRICH chip developments.
5.1 Power consumption

The proposed readout would simplify the number of required voltage levels. During Run 3, four different regulator voltages are required, namely 2.5 V for the CLARO and VTRx, 1.5 V for the GBTx and 1.0 V, 1.8 V and 2.5 V for the FPGA [5]. This would be reduced to two voltages, 1.2 V for the FastRICH and lpGBT and 2.5 V for the VTRx+. This relatively simple and low-voltage scheme is a benefit also for Upgrade II, where low power and compact electronics designs will be required. There would be an interest in more compact, surface-mount and rad-hard DC-DC regulators for the high-occupancy detector region should these become available. For LS3, the 2.5 V regulator plugins may be reused from the Run 3 detectors, while the other regulators would need to be modified to or replaced by 1.2 V regulators.

The removal of the Run 3 FPGAs (since their functionality is included in the FastRICH chip) saves about 4.4 W power consumption per digital board reading out 1-inch type MAPMTs, which equates to ~ 8.6 mW per channel. Although the FPGAs are further from the EC, which is the more critical part for power consumption, this reduction in power consumption is still beneficial and may compensate some of the increased power consumption of the FastRICH (< 6 mW/ch target for MAPMTs) compared to the CLARO ASIC (0.7 mW/ch). Furthermore, a typically 50% margin was built into the Run 3 electronics power consumption due to irradiation damage during Run 3 and Run 4 combined. With the replacement of components during LS3 enhancements, the TID is set back to zero at the start of Run 4 and smaller margins can be kept. Furthermore, R&D is ongoing to evaluate minor adjustments to the EC structure and boards layout to improve the thermal coupling between the cold bar and the FastRICH ASICs.

While this document proposes LS3 enhancements, these are intended as a first step towards the HL-LHC Run 5 in particular for the FastRICH chip, which is foreseen to be used also for the major Upgrade II. The Upgrade II detector will have a higher channel density due to the proposed reduction in pixel size. The addition of more channels within similar spatial constraints necessitates lower power electronics during Run 5. Additionally, the analogue signal integrity generally requires the ASIC to be nearby the sensor. In case of SiPMs operated with cryogenic cooling during Run 5, the heat generation near the cold region should therefore be as small as possible. For Upgrade II a complete redesign of the detector is foreseen including an improved cooling and mechanical layout, which will help to operate the FastRICH within the more demanding constraints of the HL-LHC RICH detector design.

5.2 Optical plugins

Part of the design specifications and appeal of the FastRICH is its close-compatibility with the lpGBT avoiding the need for an FPGA (or ASIC-equivalent) for data formatting. The digital outputs of the FastRICH will be directly connected to the lpGBT eLinks and the I2C can be used for configuration of the FastRICH. The 160 MHz clock lines from the lpGBT plugin can be distributed across the front-end.

The choice for an optical link plugin is driven by (a) flexibility to easily design different motherboards for bandwidth optimisation, (b) modular replacement in case of component failure and (c) PCB requirements for the lpGBT. In particular for (c), the 0.5 mm BGA of the lpGBT requires a high-end PCB technology. It therefore reduces the cost to limit
the area of this PCB technology to the plugin rather than the entire motherboard. It is being investigated whether the readout can be given the flexibility to reuse GBTx plugins ($\sim 5$ Gb/s) from Run 3 in conjunction with novel lpGBT plugins ($\sim 10$ Gb/s). For example, the DTMs and TCMs from Run 3 could be used in the low-occupancy regions while faster plugins are installed in the high-occupancy region in RICH1.

A prototype version of an lpGBT / VTRx+ optical plugin has been designed as shown in Figure 11 and is ready for production in small quantities. The layout is close to the final envisaged design and fits the mechanical constraints of the Run 3 column. On this version, two lpGBTs are used while the VTRx+ could accommodate the output of up to four lpGBTs. The number of lpGBTs per plugin and the number of plugins per variant of the motherboard will be optimised for a more detailed LS3 technical design report in the near future. In the region of the detector requiring the highest bandwidth and to make most cost-effective use of the 4-channel VTRx+ modules it may be desirable to develop a plug-in variant having 4 lpGBT ASICs. It may be challenging to achieve this within a realistic plug-in footprint due to the large number of connections required to the motherboard. However, the option remains to design a motherboard for this limited region having the components directly integrated onto the motherboard in order to achieve the required density.

The minipods on the existing PCIe40 can go up to 10 Gb/s and should therefore be compatible with the higher lpGBT and VL+ rates. The suitability at this higher rate of the data fibres from the LHCb cavern to the computing farm are being studied within LHCb and may require the PCIe40(++)s that operate at 10 Gb/s to be installed closer to the cavern. Alternatively, the lpGBT plugins may be able to operate at half-speed using the existing fibres or the GBTx plugins may be reused. Detailed studies are starting up to assess the exact number of optical fibres required, which depends on the precise data compression format, hit occupancy and layout of the (different versions of) front-end digital boards.
5.3 Bandwidth considerations

Bandwidth is a key consideration for the future upgrades, especially with the addition of precise-timing information and the increase in luminosity and readout channel density in Run 5. The aim of the LS3 enhancements is to add time information whilst staying at a similar or only slightly increased bandwidth with respect to the Run 3 detector. Several requirements on the front-end ASIC were defined in Section 4.1 in order to minimise the output bandwidth. These include a zero-suppressed data format, CFD and configurable hardware shutter time of 2 ns.

The Run 3 detector contains $\sim 124 \times 10^3$ channels in RICH1 and $74 \times 10^3$ in RICH2. For every bunch-crossing period, the front-end data are sent triggerless with no data compression (NDC) to the back-end. This results in a raw data bandwidth of $\sim 8$ Tb/s ignoring the header data. This data is transmitted across 2496 GBTx ASICs operating at 4.8 Gb/s. Therefore $\sim 65\%$ of the optical link bandwidth is used by the two versions of the digital board, where one version reads out 512 channels from 1-inch MAPMTs across six links and the other version 256 channels from 2-inch MAPMTs across four links. The optical links are operated at two-thirds of their maximum capacity due to the overhead from header data ($\sim 12.5\%$) as well as the digital board design optimisation to route the MAPMT channels to optical links. For the Run 4 enhancements, the boards will need to be re-optimised to achieve the best data link usage for the zero-suppressed data, which will be a critical experience also in preparation for Upgrade II.

At the back-end Run 3 configuration, the RICH detector has 72 PCIe40 cards with 3456 input links. The cards process the front-end data at every bunch crossing. The PCIe40 output bandwidth is 100 Gb/s, resulting in a total maximum bandwidth capability of 7.2 Tb/s. The back-end therefore processes the incoming data through an optimised combination of DC in the low-occupancy and NDC in the high-occupancy detector regions.

![Figure 12: Photon detector hit occupancy (%) in the RICH1 detector.](image-url)
This reduces the incoming NDC data from \(~8\) Tb/s to \(~1.1\) Tb/s. The DC is particularly effective due to the low average photon detector occupancy, which is simulated in the LHCb framework to be 5.3\% for the RICH1 detector and 5.5\% for RICH2 over the 25 ns bunch crossing period. A small region of the detector in RICH1 has a higher photon occupancy reaching up to \(~40\%\) as shown in Figure 12.

As a novel feature during Run 4, the radiation-hard ASIC solution will output compressed data from the front-end. This was not possible for the Run 3 FP\(\text{GAs}\) as the data compression logic would add to the resources used in the FPGA and as a result decrease its radiation tolerance. As outlined in Section 4.2 and Appendix A, the FastRICH would require 7 bits for timing plus 4 bits for channel identification per hit. This channel identification relies on accurate bookkeeping for the mapping from the ASIC to the optical link and back-end, so that only the channel (out of 16) on the chip needs to be encoded. Based on the average occupancies and a 30 MHz bunch crossing rate, a total DC front-end output data bandwidth of 3.5 Tb/s is expected. Although this is less than half of the NDC Run 3 bandwidth, sufficient margin needs to be built in to account for the following contributions in the DC data:

- Photon detector occupancy fluctuations typically caused by ‘busy’ events.
- Bunch count synchronisation and header data for the compressed data format.
- Additional challenges in the ASIC-to-optical link mapping imposed by the highly non-uniform detector occupancy.
- The possibility of increased luminosity operation during Run 4 (for example from \(2.0 \times 10^{33} \text{cm}^{-2} \text{s}^{-1}\) to maximum \(3.0 \times 10^{33} \text{cm}^{-2} \text{s}^{-1}\)).

Taking all into account, a factor 3 margin in the bandwidth can be chosen. This margin, which will be subject to further scrutiny in simulation studies for several digital board layout options, would give a data bandwidth of \(~10.5\) Tb/s. Hence the front-end to back-end bandwidth can be expected to remain within the 12.2 Tb/s bandwidth limit of the Run 3 optical fibres, albeit with a challenging re-optimised data fibre routing to compensate for the non-uniform photon occupancy on the detector.

At the back-end, the \(~10.5\) Tb/s Run 4 bandwidth (including margin) would exceed the output bandwidth capacity of the Run 3 PCIe40 cards of 7.2 Tb/s. In contrast to Run 3, DC has already been applied at the front-end and cannot be used to further reduce the back-end output. However, studies will be performed to evaluate potential gains from using a NDC format at the back-end for the high-occupancy data links and the possibility to use smaller margins for fluctuations per PCIe40(++) than per optical link, because each PCIe40 card hosts up to 48 links. Nevertheless, a modest expansion of the back-end resources is anticipated in the form of addition PCIe40++ cards. These cards have a four-fold output bandwidth compared to the PCIe40 and can therefore be optimally used to process the high-occupancy regions of the RICH1 detector.

The FastRICH will have 16 channels and a programmable number of output links of 1, 2 or 4. At the highest output link frequency of 1.28 GHz, the output bandwidth is therefore 5.12 Gb/s per chip. In total, the RICH detectors have about \(~200 \times 10^3\) channels requiring 12500 FastRICHs. The configurable number of output links has the advantage that in the low-occupancy regions fewer eLinks per FastRICH can be used, which reduces the number of required lpGBT and VTRx+ components significantly. In the high-occupancy region, the approach will be taken to enable four output links at 1.28 Gb/s in the high-occupancy region.
6 R&D

The R&D for the LHCb RICH hardware and software are outlined in Sections 6.1 and 6.2 respectively. These activities provide a clear path towards the LS3 enhancements, which is outlined in more detail with a timeline in Appendix C. Additionally, the R&D lays an important foundation for Upgrade II in terms of technologies, algorithms and experience with detector operation using fast timing information. The extensive RICH test beam campaign, which was started in 2021 and will continue on a regular basis in the years to come, is central to these R&D developments as it provides the unique opportunity to study prototypes of the optoelectronic chain and develop and test small-scale detector systems in conjunction with new DAQ, software and reconstruction algorithms.

6.1 Hardware

An active programme of hardware R&D is underway. The two main activities to date relate to the specifications of the FastRICH ASIC and the development of prototype readout modules and their evaluation in the laboratory and CERN beam facilities.

The development of the FastRICH ASIC is a collaboration between ICCUB and CERN. Successful exploitation in the LHCb RICH in LS3 requires careful attention to the detailed specifications of the ASIC, in particular in relation to the characteristics of the analogue front-end and to the capabilities of the readout interface which must be compatible with the LHCb readout architecture and capable of maintaining the required data transmission bandwidth. The RICH group is therefore closely collaborating with the FastRICH developers at this early stage of its development in order to define the critical specifications. Detailed studies of the timing precision and resolution necessary to achieve the required PID performance have been performed and this information has been used to define a compact encoding scheme for the data. These studies have also led to proposals for new features such as CFD of the analogue input with the aim of further reducing the required output bandwidth. Detailed studies have also been made to estimate the output data bandwidth in different regions of the detector in order to define an optimal strategy for the design of the readout interface with the aim of achieving the most cost-effective use of the data links. This modelisation for the number of optical link components for the varying photon occupancy across the detector will be useful also for Upgrade II.

The development of readout prototypes and their evaluation has also been an area of significant activity. We have taken the approach to develop prototypes that incrementally improve the time resolution in parallel with the development of the FastRICH ASIC. In the absence of a suitable prototype multi-channel ASIC with embedded TDC at this early stage of the project, we have therefore developed a TDC in an FPGA for our first campaign of measurements. Our initial measurements in the laboratory and at the CERN beam facilities have given encouraging results and suggest that the prototype is working as expected with the expected resolution. Our main focus is to now make additional measurements in the laboratory in order to better characterise the behaviour of the front-end ASIC with the aim of making further measurements with beam in the coming year. We are also developing an alternative readout board based on the CERN picoTDC ASIC that will serve as a very precise reference system for future measurements [16].

A further strand to the hardware R&D will be the integration of our readout electronics with the LHCb readout architecture using the CERN data transmission chipsets and the
thermomechanical integration of the new electronics with the existing RICH mechanics. Additionally, an irradiation campaign of the ASIC is foreseen to study its operation at radiation conditions similar to the LHCb RICH detector.

6.2 Software and reconstruction

The software packages for LHCb simulation and reconstruction are undergoing a major upgrade for Run 3 which includes using the DD4HEP [17] detector description package and enabling the multithreading features of the GEANT4 software toolkit. The software developments aimed at Run 4 also need to be upgraded to make use of this new framework.

Studies will continue to develop and optimise the algorithms that make use of the RICH hit timestamps in order to find the primary vertex time. These studies can help to achieve a reduction in the size of the time gate that can be applied in reconstruction software as indicated in Section 2.3. A detailed simulation of the electronics resolutions and digitisation will be implemented, which is an important development also for Upgrade II and the ability to simulate in more detail the expected performances during Run 5. Additionally, the effect on the PID performance of a two-level threshold in the ASIC can be studied using the RICH reconstruction algorithms. Fast timing information is a cornerstone of the LHCb and RICH upgrade programme, and the LS3 enhancements would provide a unique opportunity to develop and use the novel timing algorithms for Run 4 and Run 5 with LHC data.
7 Conclusions

The LHCb Upgrade II will fully exploit the flavour-physics opportunities of the HL-LHC during Run 5 and Run 6 in line with the recommendations of the 2020 update of the European Strategy for Particle Physics, and will study additional physics topics that exploit the forward acceptance of the LHCb spectrometer. The main Upgrade II detector will be installed in LS4 of the LHC (2033) and will build on the strengths of the former LHCb experiment and the current Upgrade I. It will operate at a luminosity of up to $1.5 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$, with the aim of integrating $\sim 300 \text{fb}^{-1}$ throughout the HL-LHC era.

By bringing forward to the period 2023-2028 a modest part of the investment foreseen for the Upgrade II, it is possible to already install in LS3 some of the detector improvements foreseen for Upgrade II. This is the case for the RICH system, which we propose to enhance during LS3 by installing a new time-sensitive front-end readout, thus anticipating at a minimal cost what is expected for LS4 and the LHCb Upgrade II.

The proposed enhancement consists of replacing the present readout architecture with a novel ASIC (FastRICH) and a direct connection to the newly developed optical links components (lpGBT/VL+). It has the capability to gate the acquisition to $\sim 1\text{ns}$ (and not $25\text{ ns}$ of the LHC crossing rate) and to provide each photon detected inside this gate with a precise timestamp, by using a TDC with readout bins of $25\text{ ps}$. This results in a time resolution of the Run 4 detector of $\sim 150\text{ ps}$, limited only by the MAPMT transient time spread. Moreover, it simplifies the data flow paths and takes advantage of its intrinsic data compression capabilities, thus requiring only a small number of extra links to the back-end DAQ network. The same ASIC would be integrated in the final system foreseen for the LHCb Upgrade II.

All services and mechanical, optical and electrical components remain the same as in the present system and are expected to be operational until LS4 and the full Upgrade II deployment. This enhancement not only prepares for the full upgrade II exercise, but is expected to have strong benefits for the particle identification and LHCb performance during Run 4.
References


## Appendix A: FastRICH specifications

Table 2: Preliminary specifications for the FastRICH ASIC for the LHCb RICH LS3 enhancements and Upgrade II.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>65 nm CMOS</td>
</tr>
<tr>
<td>Die dimensions / # of pads</td>
<td>$3 \times 4 \text{mm}^2 / \mathcal{O}(100)\text{^2}$</td>
</tr>
<tr>
<td>Package / sensor coupling</td>
<td>BGA with pitch $\geq 0.8 \text{mm}$</td>
</tr>
<tr>
<td>Radiation hardness</td>
<td>Yes (TID $&gt; 100 \text{Mrad}$ and triplication)</td>
</tr>
<tr>
<td># of channels</td>
<td>16</td>
</tr>
<tr>
<td>Channel type</td>
<td>Linear (i.e. not pixelated)</td>
</tr>
<tr>
<td>Channel connection</td>
<td>Single-ended</td>
</tr>
<tr>
<td>Polarity</td>
<td>Configurable positive or negative</td>
</tr>
<tr>
<td>Input signal attenuation</td>
<td>Configurable per channel: 1, 1/2, 1/4, 1/8</td>
</tr>
<tr>
<td>TDC time bin</td>
<td>25 ps</td>
</tr>
<tr>
<td>Electronics time jitter</td>
<td>$\sim 30 \text{ps RMS SPTR}$</td>
</tr>
<tr>
<td>Residual time walk</td>
<td>$&lt; 200 \text{ps pk-to-pk}$ (after CFD, over 50 $\mu\text{A}$ to 5 mA range)</td>
</tr>
<tr>
<td>Time gate</td>
<td>2 ns nominal, configurable width and offset to the 40 MHz clock</td>
</tr>
<tr>
<td>Power consumption analog</td>
<td>Target $&lt; 4 \text{mW}$</td>
</tr>
<tr>
<td>Power consumption digital</td>
<td>$\sim 2 \text{mW per channel}$</td>
</tr>
<tr>
<td>Energy resolution</td>
<td>Non linear (not required when CFD is implemented). Possibility of an additional threshold level, increasing the output bandwidth by 1 bit/hit</td>
</tr>
<tr>
<td>Dynamic range</td>
<td>5 $\mu\text{A}$ to 5 mA **</td>
</tr>
<tr>
<td>Maximum front-end rate</td>
<td>$&gt; 50 \text{MHz}$ (non-linear ToT mode, Sensor dependent)</td>
</tr>
<tr>
<td>Testing and calibration</td>
<td>Internal test charge generation controlled by digital signal</td>
</tr>
<tr>
<td>Slow control interface</td>
<td>I2C with multiple chips on the same I2C bus</td>
</tr>
<tr>
<td>VCO oscillation freq.</td>
<td>1.66 GHz</td>
</tr>
<tr>
<td># of VCO stages</td>
<td>12</td>
</tr>
<tr>
<td>Bits/event (ToA)</td>
<td>$\text{fToA @}500\text{ps}$: 2 (Assumes a 2 ns gate) $\text{uToA @}20\text{ps}$: 5</td>
</tr>
<tr>
<td>Total bits/event</td>
<td>7 ToA ($2\text{fToA}, 5\text{uToA}$)</td>
</tr>
<tr>
<td>Channel identification</td>
<td>4 Channel identification</td>
</tr>
<tr>
<td>1 Threshold high hit (only Upgrade II)</td>
<td></td>
</tr>
<tr>
<td>Output</td>
<td>Digital differential, lpGBT compatible</td>
</tr>
<tr>
<td>Output links freq.</td>
<td>160, 320, 640, 1280 MHz</td>
</tr>
<tr>
<td># of output links</td>
<td>Programmable at chip level to 1, 2 or 4</td>
</tr>
</tbody>
</table>

* Including CFD and a branch for second (‘2 bit’) threshold level.
** While 5 $\mu\text{A}$ can be reached in terms of electronic noise, the timing performance can be achieved over the range 50 $\mu\text{A}$ to 5 mA.
Appendix B: Costings

A preliminary estimate of the costs for improvements in LS3 is given in Table 3. These partially reduce the expenses for Upgrade II and its R&D. The main cost contribution is from the silicon engineering run to produce the FastRICH ASIC. The digital board includes DC/DC regulators, connectors, PCB production and assembly. The optical links plugins include component costs for $\sim 3k$ lpGBTs and $\sim 1k$ VTRx+ ASICs and requires some contingency due to the relatively advanced PCB technology required for the 0.5 mm pitch BGA of the lpGBT. An expansion of the back-end infrastructure with $\sim 25$ PCIe40++ boards is first estimated and subject to further investigation and simulation of the bandwidth. These boards are also foreseen to be used for Upgrade II.

Table 3: Preliminary cost estimates in kCHF for RICH1 and RICH2 combined during LS3 enhancements. These include the costs for the associated R&D.

<table>
<thead>
<tr>
<th>Description</th>
<th>LS3 cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronics</td>
<td></td>
</tr>
<tr>
<td>FastRICH production and packaging</td>
<td>900</td>
</tr>
<tr>
<td>Front-End Board (FEB)</td>
<td>150</td>
</tr>
<tr>
<td>Digital board (front-end)</td>
<td>325</td>
</tr>
<tr>
<td>Optical link plugins</td>
<td>400</td>
</tr>
<tr>
<td>DAQ boards (back-end)</td>
<td>325</td>
</tr>
<tr>
<td>Ancillary services and detector infrastructure</td>
<td></td>
</tr>
<tr>
<td>DCS/DSS</td>
<td>40</td>
</tr>
<tr>
<td>Calibration, monitoring and alignment</td>
<td>160</td>
</tr>
<tr>
<td>Total</td>
<td>2300</td>
</tr>
<tr>
<td>Contingency</td>
<td>+15%</td>
</tr>
<tr>
<td>Grand total</td>
<td>2645</td>
</tr>
</tbody>
</table>
Appendix C: Timeline

A proposed timeline for the LHCb RICH enhancements for LS3 is shown in Figure 13. Here, R&D refers to front-end electromechanic tests including cooling, FastRICH design and simulation studies, the FastRICH interface to the optical link chipset, PCB design, DAQ implementation and first DAQ firmware studies and test beams of prototype systems with comparisons to the present front-end. Additionally, R&D in software, simulation studies, full implementation of timing algorithms, new reconstruction algorithms, detector evaluation and PID and physics performance studies are included in the foreseen R&D. The timeline for the FastRICH ASIC project is shown in Figure 14.

Figure 13: LHCb RICH timeline for LS3 enhancements.

Figure 14: Timeline for the development and production of the FastRICH ASIC.