Hardware prototyping and thermal simulations of strip sensor, and simulation of transmission lines for the LHCb VELO upgrade.

Iain Longstaff
Primary Supervisor: Lars Eklund
Secondary Supervisor: Paul Soler

Thesis submitted for the degree of Master of Science by Research

Particle Physics Experiment
School of Physics and Astronomy
College of Science and Engineering

January 17, 2022

Contents

Abstract

The LHCb experiment based at CERN's Large Hadron Collider (LHC) particle accelerator is undergoing upgrades to replace end of life detectors and take advantage of the luminosity upgrades that the LHC machine is gaining.

The VErtex LOcator (VELO) is a particle tracking sub detector of LHCb and its upgrade is designed to record up to an order of magnitude more data from proton-proton collision events. During the development of the upgrade, silicon strip sensors and pixel detectors were investigated as options.

A prototype of the strip sensor was assembled with readout hardware to characterise its performance. Cooling methods were considered and thermal simulations were performed to verify that cooling an irradiated silicon strip sensors after 50 fb^{-1} of integrated luminosity is viable. Using the technologically least aggressive and hence easier to implement cooling solution required liquid CO_2 to be below -18°C to avoid thermal runaway, the upgraded CO_2 cooling infrastructure is capable of -30°C hence thermal control is viable.

The silicon pixel sensor technology was chosen for its superior track reconstruction performance as the upgrade technology. It requires around 2.85 Tbits/s of data to be transferred for processing off-site. The sensor's readout electronics require multiple parallel transmission lines capable of each carrying 5.12 Gbits/s of data without significant signal loss and also have sufficient flexibility to accommodate the 3 cm lateral movement of the VELO sub detector halves.

A flexible copper multilayer ribbon cable design, so called "flex cables", is a compact solution, with a thickness just over 400 µm where each transmission line is 1200 µm wide. Simulations were performed to find the optimal characteristic differential impedance and transmission properties which were used to guide the prototype development. Comparisons of impedance measurements against real world prototype versions showed a systematic bias of around 10 Ω that could be explained by various factors explored within. Comparisons of transmission versus frequency showed good agreement between simulation and measurement. The flex-cable design has been finalised and the VELO upgrade is being assembled in preparation for on site installation into the LHCb detector.

Declaration of Authorship

The results presented in this thesis are the product of my own work. Appropriate references are provided when results of third parties are mentioned. The research presented here was not submitted for another degree in any other department or university.

To be more specific, the first two chapters are background to bring context to my work and credit goes to the thousands of people involved in making the LHCb experiment possible.

Chapter 3 includes my work on the LHCb VELO upgrade on strip detector development where I assembled a working strip detector prototype where Joe Ashby, a detector development technician, performed the wire bonds necessary for the assembly. Alex Morton was a masters student when we set up the measurements for the prototype. He did the data collection for the results shown in this section. I did the thermal-electric simulations of the strip sensor.

Chapter 3 also includes work on transmission lines for the VELO pixel upgrade. I ran electrical simulations to predict its performance. Dr Cameron Dean recorded the measurements of the flex-cable prototypes. Sneha Naik finalised the designs of the later versions of the flex cable using some of my results as input.

Acknowledgements

I first would like to thank Lars Eklund for his guidance, good humor and being generous with his time throughout the ups and downs of this journey, this wouldn't be possible without you. My second supervisor Paul Soler has also been helpful and given good advice and support in difficult times.

Daniel Hynds has been a fantastic mentor into the world of pixel detectors and late nights at test beams. He was always happy to answer Alex's and my well-formed and not-so-well-formed questions! I'm grateful to Alex Morton for the good company at test beams and for his work on the strip sensor prototype.

Thanks you to the people at the University of Liverpool for the help on thermal simulations. Tony Affolder, Tony Smith, Ashley Greenall, Peter Sutcliffe and Graham Beck, you have been great!

And thanks to Pablo Rodrguez and Eliseo Perez Trigo when I visited the department in Santiago de Compostela, it was nice to be face to face after collaborating at long distances. I'll cherish the memories of being on the roof of the Cathedral!

The academic and support staff at the Physics department have all provided a welcoming atmosphere and I'd like to say cheers in particular to Kenneth Wraight, Michael Alexander, Richard Bates, Fiona Grant, Liam Cunningham, Aidan Robson, Craig Buttar, Fred Doherty and Joe Ashby for their help.

The people who have at some point been PhD and masters students with me in Glasgow have been wonderful to be around. Will Breaden Madden, Stephen Ogilvie, Paul Sail, Sarah Karodia, Cameron Dean, Murdo Traill, Dana Bobulska, Patrik Hallsjo, Lauren Douglas, John Nugent, Sneha Naik and Leyre Flores amongst many more. It has been a pleasure to have lunch together, chat at the many Glaswegian pubs, go bouldering, play badminton, fly drones where they probably shouldn't have been flown, enjoy insightful fikas and play table-tennis in the common-room.

On LTA at CERN I met many amazing people and thank you, Jan Buytaert, Paula Collins, Kazu Akiba, Patrick Spradlin, Raphael Dumps, Jibo He and Vladimir Gligorov for your insights on the many aspects of how the VELO works and the principles behind it. There were many people who gave good company whether that was walking up mountains, sliding down mountains, or simply lunch, dinner and boardgames at R1. Thanks to Hugo Day, Patrik Thuiner, Tom Bird, Heinrich Schindler, Will Kalderon, Faye Cheung, Lewis Armitage, Luca Pescatore, Duncan Leggat and many more who were part of the LTA community or otherwise.

Outside of physics I've developed fond friendships in Glasgow, thanks for the help to improve the urban landscape, the madness at Improv, and the adventures around the table. Euan and Alastair who have in particular have been close flatmates and friends, I've enjoyed sharing the meals and the time together!

I'm grateful for a loving family who have cared about me, thank you Mum, Dad and Chris! During my time with the physics department in Glasgow, I've had the pleasure of meeting Johanna who has given so much love and support, its been a wonderful journey with you and I don't know what I'd do without you!

And last but not least, Ghillie my hearing dog has been bringing a smile to me and everyone else around everyday; life is made richer by his companionship!

List of Figures

List of Tables

1 Introduction

1.1 The Large Hadron Collider

The introduction gives a general description of CERN, the LHC accelerator, ATLAS, CMS and ALICE detectors. Details about the LHCb detector are given in Chapter 2.1.

The European Organization for Nuclear Research, known as CERN[1], is an international research organisation with 22 member states and runs the world's largest particle physics laboratory located north-west of Geneva.

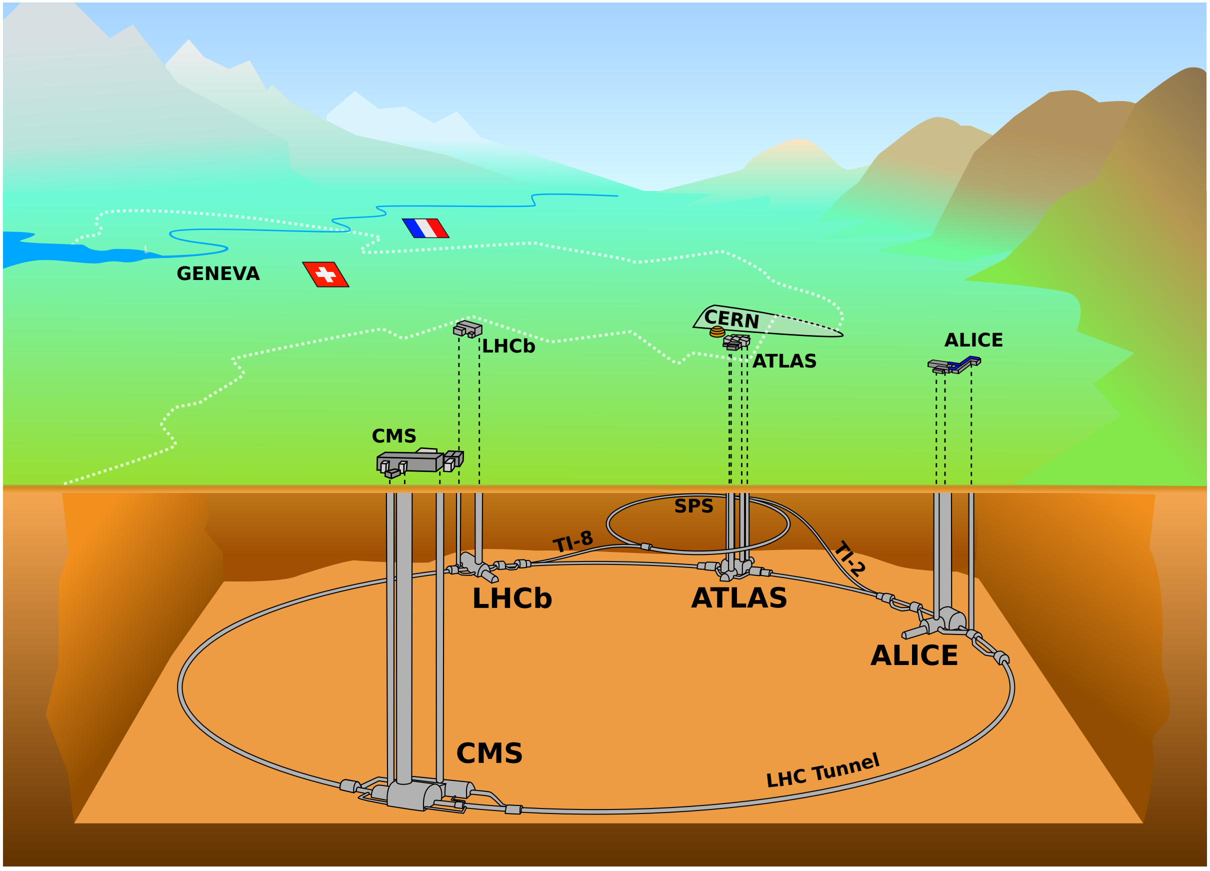


Figure 1: An overview of the LHC and the four main experiments, ATLAS, CMS, ALICE and LHCb

The facility hosts The Large Hadron Collider (LHC) [15], the most powerful synchrotron particle collider ever built, and aims to reveal physics beyond the Standard Model. It is built in the same 27.6 km underground tunnel that the Large Electron Position (LEP) collider had been constructed in [16]. It has achieved a centre of mass energy of 13 TeV and a luminosity of $2{\times}10^{34}$ cm$^{-2}s^{-1}$. It has four main underground experiments located at each of the collision points, two of them had new underground chambers excavated to accommodate their large size. Figure 1 shows the four main underground experiments located at the beam collision points along the 27 km long tunnel.

Hydrogen is ionised and accelerated to 50 MeV in a linear accelerator which is directed into the Proton Synchrotron and its booster ring before being injected to the Super Proton

[1] the acronyn comes from the original name, Conseil Européen pour la Recherche Nucléaire.

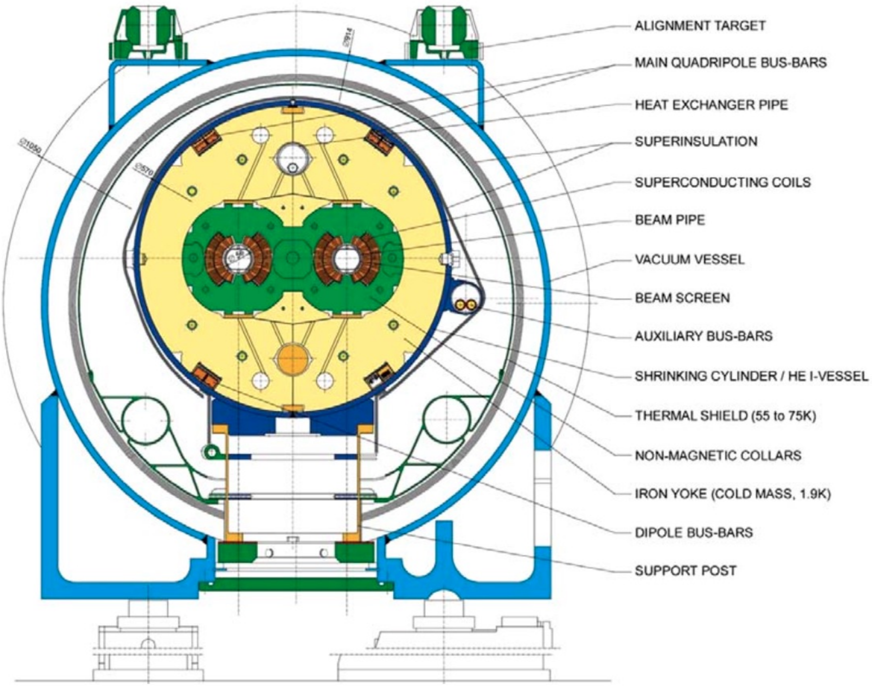


Figure 2: A cross-section of the main dipoles in the LHC.

Synchrotron at 25 GeV where the momentum in increased to 450 GeV. The final injection goes into the LHC ring where a 400 MHz Radio Frequency Acceleration System boosts the beam energy up to 7 TeV.

The LHC ring has 1232 main dipole magnets sections, about 16 m long, shown in cross section in Figure 2 which takes up the majority of the 27.6 km circumference. The NbTi superconductors are cooled with super fluid helium to around 1.9 K and provides a field of 8.3 T that allows each proton beam to be guided inside the LHC curvature and reach 7 TeV.

The maximum fill with the LHC proton beams is organised into 2808 bunches per ring separated by 25 ns with a length of 1 ns (within 4σ) and transverse emittance of 3.75 µm [15].

There are plans to increase the luminosity of the LHC machine in order to increase the data collected by all the LHC experiments. There is a new linear accelerator being commissioned and other infrastructure upgrades will happen over the next decade to 2030 to increment the luminosity up to an order of magnitude by 2030. The linear accelerator for example will be completed by the end of the second Long Shutdown (LS2) of the LHC system in 2021 [17]. The underground experiments will also have their respective upgrades to take advantage of the higher luminosity.

1.2 A summary of the four main LHC detectors

The ATLAS (A Toroidal LHC ApparatuS) is a general purpose detector fully enclosed around the collision [18]. It is able to take advantage of the LHC collision energies to measure a wide variety of phenomena related to Higgs, electroweak, top and flavour physics. The detector consists of a series of nested cylindrical sub detectors each with its own function. At the core is the inner tracking detector which is within a 2 T solenoid, the calorimeters are constructed outside the solenoid. Finally a large toroidal magnet is built within the outer muon detector layer. The whole detector is 44 m in length and 25 m high.

The CMS Compact Muon Solinoid [19] is also a general purpose cylindrical detector that has a similar physics program to ATLAS. Its silicon inner tracker and main calorimeters are within a 4 T solenoid, the muon detectors are built between steel yoke layers that return the solenoid's magnetic field. The CMS is 21.6 m long and 15 m high.

The ALICE (A Large Ion Collider Experiment) detector is a general-purpose heavy-ion detector and is used when the LHC is colliding Lead - Lead and Lead - proton collisions. The aim is to study the physics of quark gluon plasma and strongly interacting matter; various Quantum Chronodynamics questions will be explored and will also improve understanding of the time in the early universe before gluons and quarks condensed into hadronic matter [20].

The LHCb (Large Hadron Collider Beauty) experiment is a forward angle spectrometer designed for precision measurements of *CP* violation, Standard Model parameters and clues beyond the Standard Model [1]. More information about the detector is in Chapter 2.1.

2 The LHCb Experiment

2.1 LHCb overview

This section describes the main objectives of the LHCb detector, the hardware of the detector and its performance up to the end of the second LHC run in 2018.

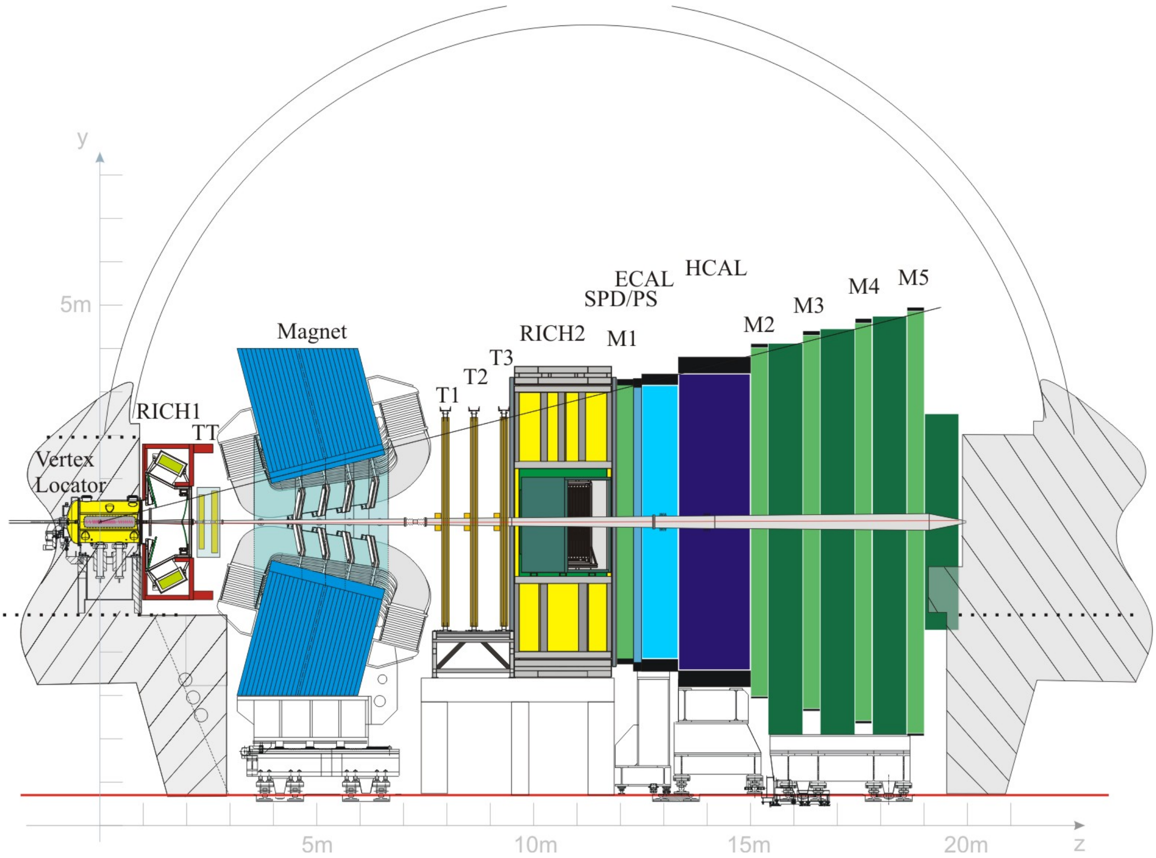


Figure 3: A side view of the LHCb detector [1].

The Large Hadron Collider beauty (LHCb) experiment [1] is designed to take advantage of the high number of b and c quarks produced in the LHC proton-proton collisions. The LHCb is a forward arm spectrometer where the proton-proton collision happens at the side of the underground chamber seen to the left in the Vertex Locator in Figure 3. Most b quark events are produced in a narrow angle to the LHC beam, Figure 4 shows a demonstration of $b\bar{b}$ pair production distributions generated by PYTHIA [21, 2]. About 27% of $b\bar{b}$ production can be reconstructed from the collisions while covering 1.8% of the solid angle of the collision point [22].

When discussing the angular coordinates of the detectors, pseudorapidity is a commonly used spatial coordinate and is defined as: $\eta = -\ln(\tan(\theta/2))$ where θ is the polar angle from the z-axis of the detector along the LHC beam direction

A right-handed coordinate system is used in LHCb where the beam axis is the z axis, the polar angle vertically from the beam line is in the y direction and the azimuthal angle is from the beam in the x-z plane. The acceptance of the detector is designed to cover ±250 mrad in the y-z vertical plane and ±300 mrad in the x-z horizontal plane where charged particle deflections take place. Thus the LHCb has a pseudorapidity range $\eta = -\ln(\tan[\theta/2])$ of about 2-5 where θ is the angle from the beam. The coverage of LHCb sensors over the solid angle range and its gaps is known as the detector acceptance, it can also imply different efficiencies in some detector regions.

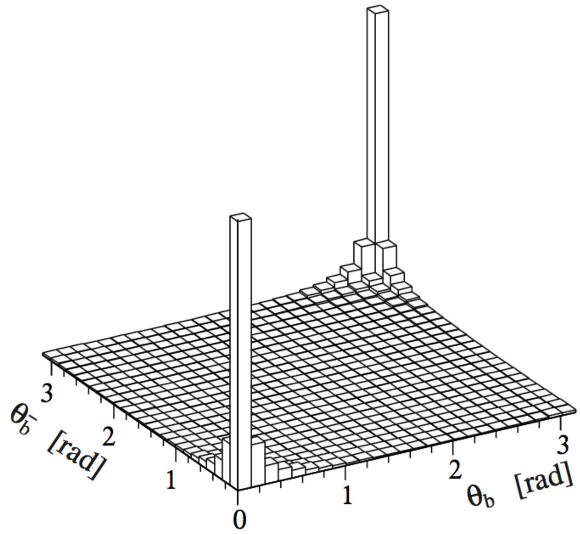


Figure 4: The production of $b\bar{b}$ pair production is confined to a narrow solid angle around the LHC beam, according to hadronic event generator PYTHIA [2].

Data taking started in 2010 using 3.5 – 4 TeV energy beams with a Long Shutdown from 2012-2015 where necessary upgrades were done to allow the LHC magnets to operate at full strength. Data collection resumed in 2015 with 6.5 TeV energy beams. The design luminosity of the LHCb was 2×10^{32} cm^{-2}s^{-1} but due to better than expected performance in event reconstruction, the LHCb can run at about twice design luminosity. [23] Figure 5 shows that over 9 fb^{-1} has been collected before the second Long Shutdown period.

Magnets are used to steer and focus the beams to control the interaction rate of each pair of counter-rotating proton bunches; this gives control on the collision rate and hence data rate of the LHCb. There is a compromise between higher available luminosity available and the LHCb detector occupancy limit. The LHCb upgrade will be able to improve the capability of using higher luminosity by increasing this limit. More details about the LHCb upgrade is found in Section 3.

The LHCb detector, seen in Figure 3, is made of several sub-detectors that specialise in an aspect of particle measurement. The LHC beams collide at the centre of the VErtex LOcator (VELO), Section 2.4, where precision tracking can locate decay positions of short lived hadrons like b-mesons. The Ring Imaging CHerenkov (RICH) detector, Section 2.9, allows measurement of particle velocity where, coupled with momentum information, allows identification of hadrons, primarily kaons and pions. The other subdetectors are more common to high energy detectors; the dipole magnet with more tracking sensors allow momentum measurements of charged particles, Section 2.3, the calorimeters measure particle energy, Section 2.7 and detectors specialised for muons are located at the far end, Section 2.8.

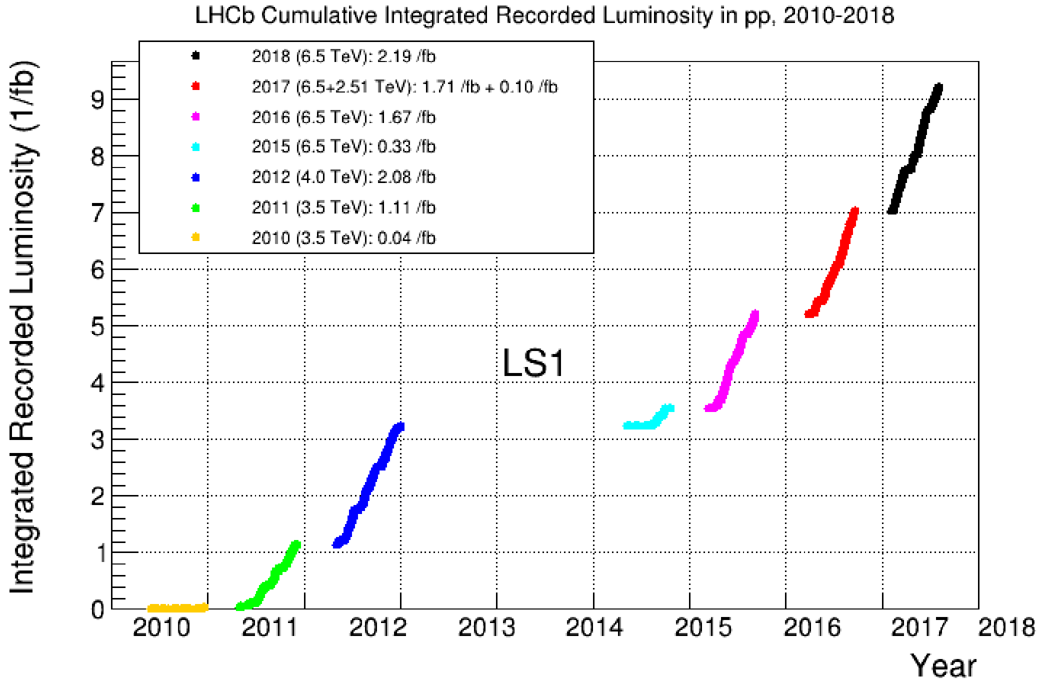


Figure 5: Cumulative integrated luminosity of the LHCb detector. LHCb collaboration.

2.2 Beampipe

The LHCb detector covers a high pseudorapidity region where there is a higher flux of high energy particles. As a consequence, there are secondary collisions in the beam pipe that increase the occupancy of detector modules nearby which can make event reconstruction harder.

When the LHC proton bunches travel down the beam pipe, the charge of the protons attract the electrons in the surrounding pipe which creates a mirror current that follows each bunch. To minimise proton bunch disturbance, the wake charge has to be allowed to travel freely around the LHC ring beam pipe. This also applies to Lead ion beams.

The Beryllium used to make the beam pipe is toxic, brittle and expensive, but it has a low material density to minimise multiple scattering of particles created in the collisions, is strong enough to hold the LHC vacuum and is electrically conductive for the wake charge. Figure 6 shows the layout of the beam pipe within the LHCb detector.

The VELO sub-detector is an unusual part of the LHC vacuum system. It is placed inside an extra cylindrical vacuum tank 1.4 m long and 1.1 m diameter. This allows a design where the sensors are separated from the beam vacuum, around 10^{-12} atmospheres and 2 orders of magnitude lower than the detector vacuum, by a aluminium foil 0.3 mm thick in the thinnest regions. The nearest foil and active sensor edges can then be moved as close to the beam axis as 4.5 mm and 8.2 mm respectively [22].

The foil also acts as an electromagnetic prophylactic, allowing mirror charges generated by the proton bunches to pass through the VELO detector with minimal impedance which reduces the disturbances to the proton bean and beam pipe heating. The mirror charges also induces wake fields from which the foil protects the VELO sensors electronics from disruption as a Faraday cage. Its is called the Radio Frequency foil or RF foil for this reason.

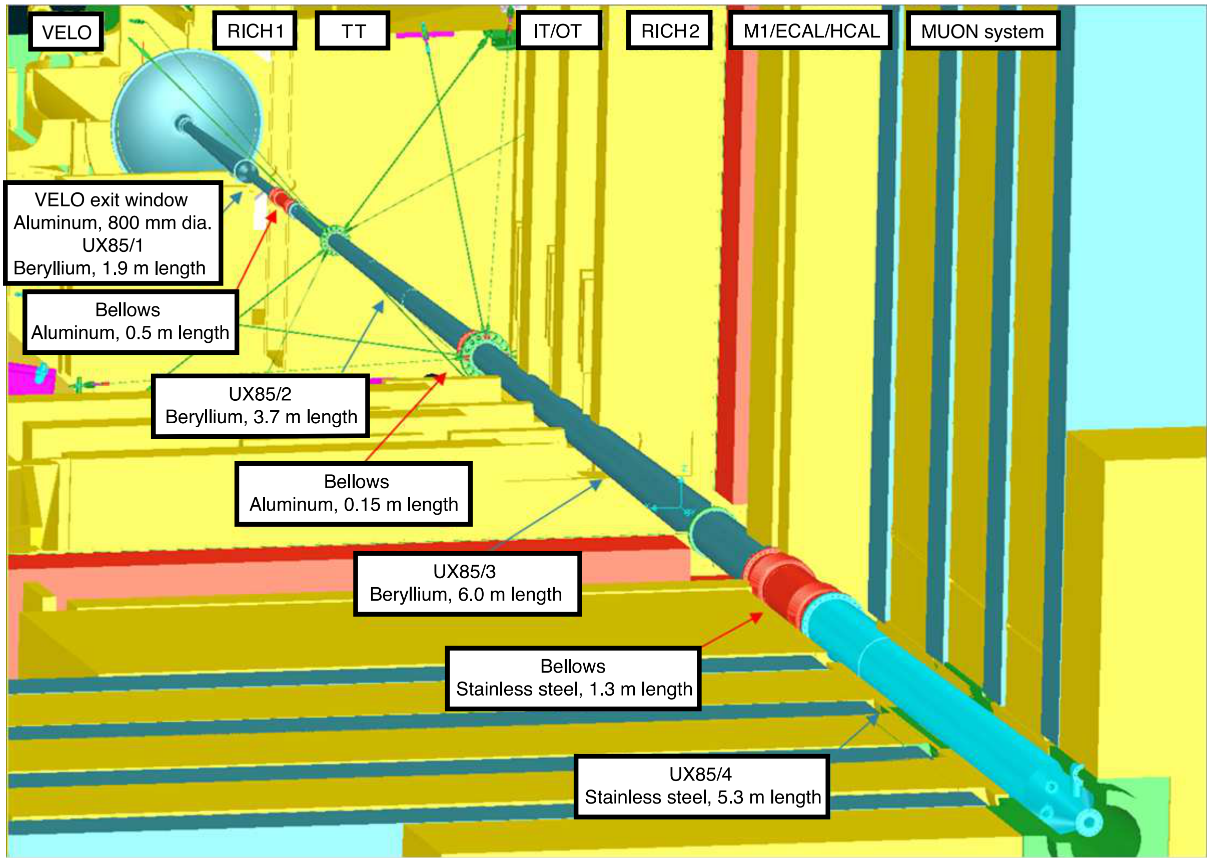


Figure 6: Cutaway view showing the beam pipe through the LHCb [1].

2.3 Particle tracking and dipole magnet overview

The tracking system of the LHCb plays an important role in providing precise positions of where particles travel in collision events. These include locations of the primary vertex (PV), where the original LHC beam collision occurs, secondary or displaced vertex, where heavy hadrons originating from the PV decay into daughter particles, and deflection through the magnetic dipole revealing momentum information.

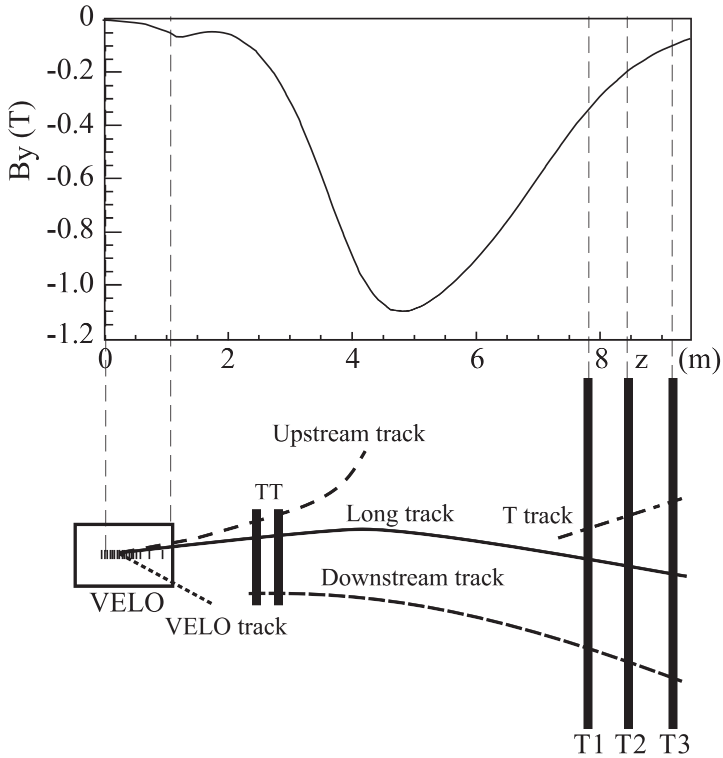


Figure 7: All the possible track types that the LHCb detector can measure and the magnetic field magnitude along the z axis [1].

There are three different technologies used in the tracking system. The original VELO detector uses strip silicon sensors with a circular geometry and has the highest spatial resolution, more details are provided in Section 2.4. There are linear silicon strip detectors installed upstream of the dipole magnet called the Tracker Turcensis (TT), and downstream of the dipole magnet called the Inner Tracker (IT). The detector planes at T1, T2 and T3 have the IT installed close to the beampipe where higher spatial resolution is needed, while the Outer Tracker (OT) covers the rest of the larger area with gas straw tube modules using drift time. More details of the TT, IT and OT are found in Section 2.6.

The LHCb uses a large water-cooled dipole magnet to deflect charged particles so that their momentum can be measured. It has a forward acceptance of $\pm$ 250 mrad vertically and $\pm$ 300 mrad horizontally. While it is not a superconducting magnet as originally proposed, it provides an integrated magnetic field of 4 Tm for 10 m tracks with a peak field of 1.1 T. It had to balance the need for less than 2 mT at both RICH detectors but maximise the field strength between the VELO and Trigger Tracker to get the best momentum resolution, the field profile can be seen in Figure 7.

The field of the magnet is reversed periodically to control for systematic biases in the LHCb detector; since charged anti-matter particles bend in the opposite direction, its important to avoid biases due to any subtle difference in one half of the detector. Hall probes mapped out the field during commissioning to account for any variability due to effects like hysteresis.

2.4 VErtex LOcator (VELO)

The VErtex LOcator (VELO) is the LHCb sub-detector that has precise enough tracking to locate displaced secondary vertices from heavy b and c-hadron decays. These displaced vertices will be up to a few cm away from the primary vertex of the proton-proton collision. It is made of 42 modules of silicon strip detectors. Each module has two back-to-back sensors, one measures radial distance from the beam called the r sensor and the other the azimuthal angle called the ϕ sensor to give a coordinate on the x-y plane, Figure 9 shows a detailed layout of the strips on each sensor. Figure 8 shows the arrangement of the modules in relation to the LHC beam collision volume with a cluster of 32 modules close to the interaction region and 10 further downstream in the positive z direction. The pile-up detectors, as mentioned in Section 2.10 is located upstream in the negative z direction for fast vertex detection.

The VELO gains precision of particle tracking due to its proximity to the interaction region, but it means it has to endure high levels of ionising radiation. It also had to be retractable when the LHC beam is being injected into the beam pipes before the proton bunches can be declared stabilised once full energy is reached. When the two halves close, seen in Figure 8, both sides give full sensor coverage around the beam where the innermost strips are 8.2 mm away from the beam.

2.5 VELO performance

The VELO was optimised to find displaced vertices with enough precision. It is designed to cover the forward region of 15-300 mrad, or pseudorapidity from $1.6 < \eta < 4.9$, where all tracks will cross a minimum of three VELO stations. The tracking precision needs to be sufficient to measure the impact parameter[2] and the flight distance of heavy mesons like B^0 or B^0_s. The track reconstruction also needed to be efficient so that the High Level Trigger can reduce the event rate from 1 MHz to a few kHz in a manageable time frame.

The performance numbers in this section is based on the 2014 paper on VELO performance [22], which included 3 fb^{-1} of data from before the first Long Shutdown. The exceptions are decay time resolution and thermal performance figures found later in this section.

The average signal to noise ratio is around 20:1 and varies due each channel having a difference strip length and routing lines, as seen in Figure 9, which affects its capacitance and hence the noise for the electronic front-end.

The resolution of a particle track hit through the sensor is dependant on the angle of the track with respect to the normal of sensor plane and the strip pitch. The best precision is when the angle is around 8° with the minimum pitch of 40 μm. Having at

[2] impact parameter is the closest distance of a track to a primary vertex

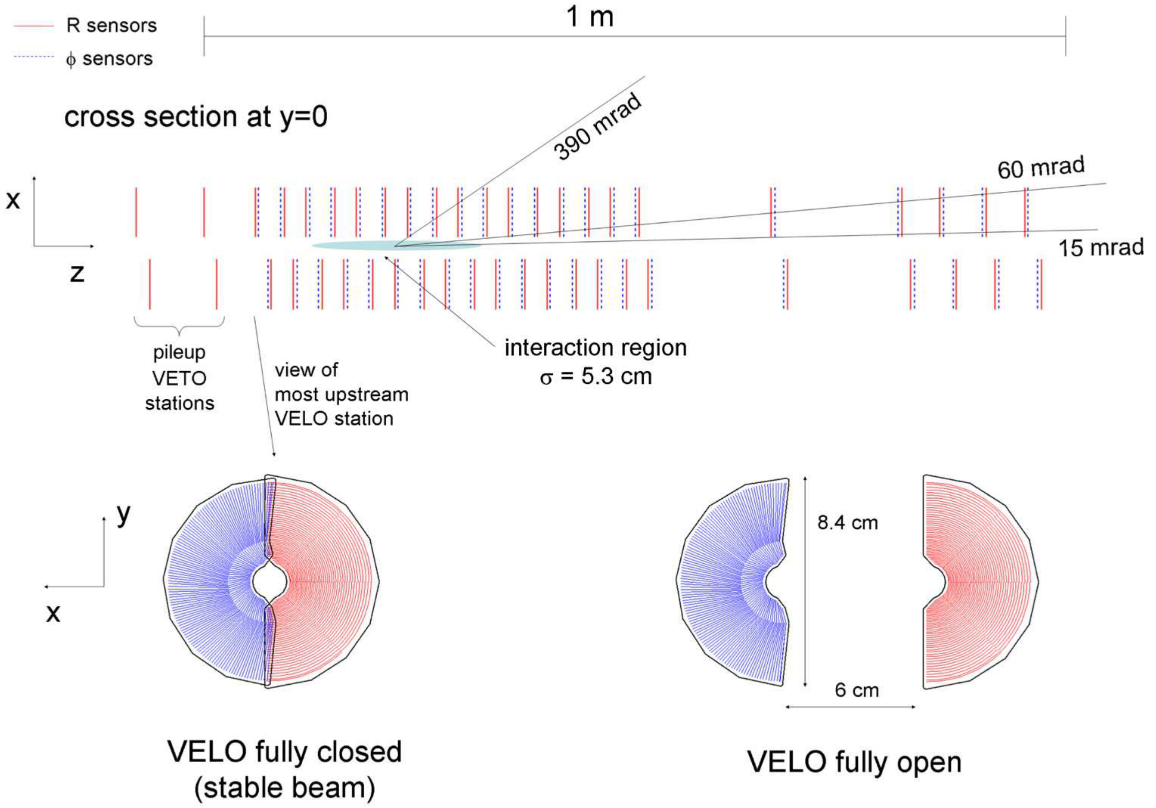


Figure 8: 42 VELO modules and 4 pileup detectors, each giving a position on the x-y plane. Each half of the detector can retract and close in to give full coverage around the LHC beam [1].

least 3 hits on the VELO station further constrains the precision of the tracks down to 4 µm.

The detector occupancy describes the fraction of channels hit per event in a sensor. If it is too high it will increase the rate of incorrectly reconstructed tracks and the time taken when reconstructing an event. Having more strips per unit area reduces occupancy but increases cost and complexity, so higher density sensors are only placed where the particle flux is expected to be higher. The occupancy in the VELO sensors were found around 0.7% to 1.2% depending on the distance from the beam and position on the beam axis from the collision point. If channels are less than 10% or three times the normal occupancy, they are deemed dead or noisy and can be disabled in the front-end electronics. Only about 1.1% of channels have been lost but the cluster finding efficiency remains at 99.97%.

The tracking reconstruction efficiency is found by using a tag-and-probe method where $J/\psi \rightarrow \mu^+\mu^-$ decays are fully constructed on one muon, an example of a full track is seen in Figure 7, while the other muon is only reconstructed in the VELO. If most J/ψ masses are correctly reconstructed, it gives an indicator on the tracking efficiency of the VELO. It was found that there is 98% efficiency across the range of pseudorapidity, momentum of 10 to 150 GeV/c, azimuthal angle and number of tracks in the same event.

A good accuracy of the primary vertex (PV) resolution is critical to determining the decay time for *CP* violating and rare processes. It is highly dependent on the number of tracks produced from the PV. From the 2011 data the resolution in x-y plane is 35-10 µm with the number of tracks from 5-40 respectively. It ranges from 270-60 µm in the z direction.

The impact parameter (IP) allows for identifying tracks of heavy mesons and their decay products; background tracks can then be eliminated to give a clean data set. The IP resolution of a track is inversely proportional to the transverse momentum of the particle. The low momentum around 5 GeV/c yields a maximum of 90 µm resolution but levels out 30 to 20 µm in the 30 to 70 GeV/c momentum range.

The decay time resolution is critical for precision measurements of decay times, a key example was the measurement of B_s^0- $\overline{B}_s^0 \rightarrow D_s^- \pi^+$ which revealed the B_s^0-$\overline{B}_s^0$ oscillation frequency to be 17.63 ± 0.11 (stat) ± 0.02 (syst) ps $^{-1}$ [24]. For short decay times, where typically heavy mesons decay within the VELO detector, secondary vertex resolution dominates. The average decay resolution from 2012 data is about 50 fs using the $B_s^0 \rightarrow J/\psi$ decay channel as a control [22].

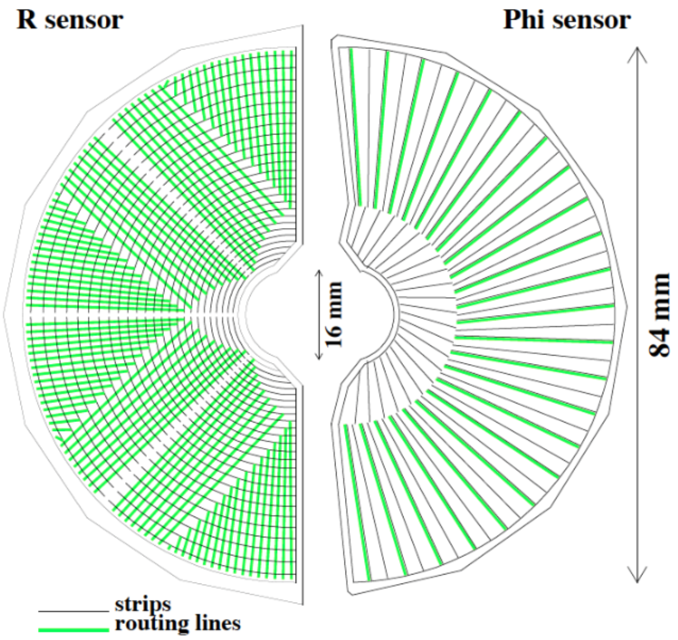


Figure 9: Layout of strips and routing lines in R and Φ sensor. The routing lines carry the charge pulses from the strip to wire bonds onto the Beetle chips.

2.5.1 Electronic front-end

Each VELO module use on each side, 20 Beetle Application Specific Integrated Chips (ASIC). They are radiation hardened and use 0.25 µm CMOS technology. Each chip has 128 inputs analogue coming from the silicon strip sensor with signals arriving at a rate of 40 MHz from the LHC bunch crossings, or every 25 ns. The output is an analogue signal which needs to match the 1 MHz accept rate of the L0 trigger. The Beetle uses I2C protocol for configuration [25].

Each channel has a charge sensitive pre-amplifier and shaper and has a buffer That holds 160 analogue signals that are stored for up to 4 µs which allows for the latency of the 1 MHz L0 trigger. The trigger will select an event and the Beetle will read out the appropriate slot before it is shifted out of memory. The 128 channels are multiplexed into four output lines, hence 80 lines go out of each side of a module.

Figure 10 show how kapton cables are used to connect the output lines and some of the input lines related to chip control. The long kapton cables have to be flexible to accommodate the movement from the two VELO halves. The Repeater board is located just outside the VELO vacuum tank and while repeating output and control signals, houses voltage regulators for the front-end and trigger electronics.

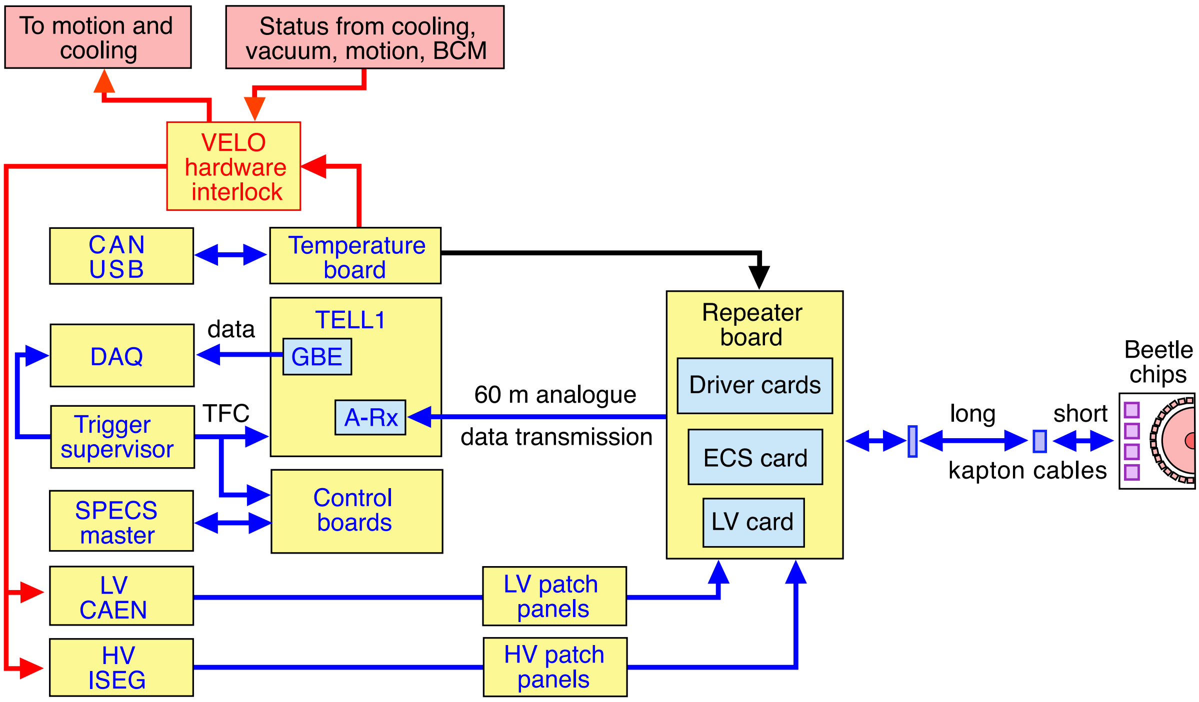


Figure 10: The electronic chain from the VELO module [1].

2.5.2 VELO cooling

When the VELO sensors are irradiated, the conductivity of the silicon changes, and the sensors are prone to rapidly overheating. More on the effects of silicon irradiation on the VELO and thermal runaway is discussed in Section 3.2.2.

The VELO uses a CO_2 cooling system to prevent the silicon sensors going at least above -5 °C. At the base of each VELO module, five aluminium cooling blocks are attached to the substrate which has stainless-steel cooling capillaries plumbed into the blocks. The core of the substrate is 400 µm thick thermal pyrolytic graphite (TPG) which has excellent thermal conductivity parallel to the planes of the TPG molecular layers. The silicon sensors and the hybrid electrical circuit layer with its components are glued onto this TPG substrate [26].

The CO_2 cooling is bi-phase; when performing nominally, the capillaries at the VELO modules will have a mix of liquid and gaseous CO_2. The liquid CO_2 is designed to boil while absorbing heat from the aluminium blocks to take advantage of the latent heat of evaporation. The CO_2 is around -28 °C and the sensors are kept at -7 °C. The heat produced by each module is about 20 W, most of which is coming from the beetle chips.

When the silicon material is irradiated, there is a process of self-annealing naturally occurs over a period of time and is highly temperature dependant. Occasionally during maintenance opportunities the VELO sensor is brought up to room temperature for a few days where the thermal energy is enough to change the characteristics of lattice defects accumulated during irradiation and reduces the required depletion voltage. However, there is also a reverse-annealing effect that increases the depletion voltage and so it is important to otherwise keep the silicon sensors below -5 °C to stave off the unwanted self-annealing [10]. The cooling system has redundant systems to ensure 24/7 cooling.

2.6 Silicon and straw trackers

Downstream of the VELO there are other tracking detectors. The Tracker Turicensis (TT) [27] and the Inner Tracker (IT) [28] using silicon strip technology suited for the higher occupancy from higher particle flux. The Outer Tracker (OT) [29] surrounds the IT stations and uses straw tube drift chambers to cover a larger area. Figure 11 shows relative positions of the trackers; the TT is upstream of the dipole magnet and the rest downstream.

Both the TT and IT uses silicon micro-strips with a pitch of 200 µm, the strip lengths vary with the expected occupancy depending on its position the beam pipe. Each station of the tracking detectors have four layers in a $(x-u-v-x)$ pattern where x are vertical strips, u is tilted $-5°$ and v is $+5°$ from the vertical. By having slanted planes of strip detectors, it is optimised for higher horizontal precision than vertical; a desirable feature for measuring momentum from charged particle deflection from the dipole magnet.

The TT and IT have similarities to the VELO detector. They use Beetle chips for readout, Kapton cables and have an active cooling system although C_6F_{14} is used. The tracker detectors are isolated in boxes from heat, electrical interference and external light and is flushed with nitrogen to avoid condensation on the detectors which is kept below 5°C.

The outer tracker uses 70% Argon and 30% CO_2 which gives a drift time below 50 ns and a drift-coordinate of about 200 µm. The OT is also enclosed in a box and water cooled. Like the silicon detectors both halves are able to be moved horizontally to give maintenance access to other detectors like the RICH detector.

2.7 Calorimeters

The Calorimeters provide measurements for particle identification, trigger and some energy and position information. A pre-shower detector (PS) is placed in front of the Electromagnetic Calorimeter (ECAL) to improve discrimination from a large charged pion background. A scintillator pad detector (SPD) located in front of the PS to distinguish between electrons and high traverse momentum π^0 s. Photons and charged particles can be distinguished by a lead sheet placed between the SPD and PS; the lead causes showering from the photon and lights up the SP, but charged particles will light up both.

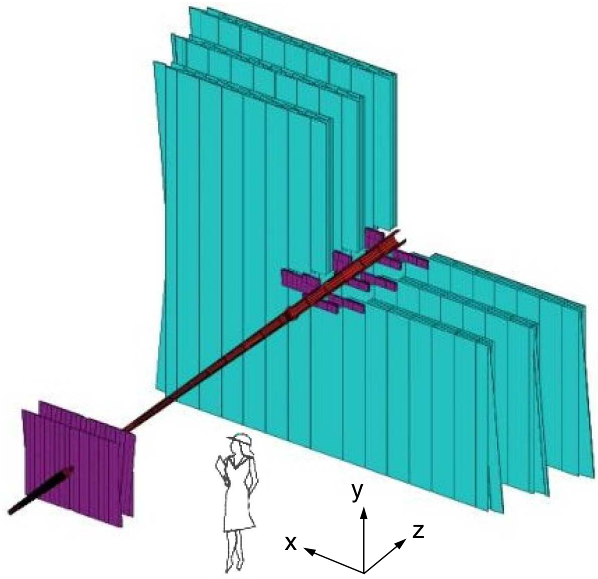


Figure 11: Layout of tracking detectors beyond the VELO, purple is the silicon detectors including Tracker Turicensis and Innner Tracker, light blue is the straw tube Outer Tracker, red is the beam pipe [1].

The ECAL allows identification of potential electrons, photons and hadrons candidates. It is a Shashlik layout where there are 66 layers of lead and scintillator plates normal to the LHC beam, 2 mm and 4 mm thick respectively. It is 25 radiation lengths for photons to ensure that they are fully contained in the ECAL for better energy resolution. The light from the scintillators are transferred via fibre optic cables to Photo Multiplier Tubes (PMT). Fibre optic cables and PMTs are also used for the PS and HCAL [30].

The hadronic calorimeter (HCAL) is 1.65 m thick and has iron and scintillator plates parallel to the LHC beam and is segmented in a manner to avoid a straight line of sight through the scintillator plates.

Its main purpose is to provide useful data for the L0 hadron trigger. There is also some data for particle identifications; electrons can be ruled out in HCAL events for example. The material in both the ECAL and HCAL attenuates almost all of LHCb's interacting particles except for muons which continue through to the downstream muon detectors.

2.8 Muon detectors

The muon detectors are an important part of the first level trigger together with the calorimeters. The time resolution must be below 25 ns to distinguish from separate LHC bunch crossings. The first muon detector station situated in front of the calorimeters is dedicated to the first level trigger only. During event reconstruction the four remaining muon detectors provide key information due to the clarity of the signal, there is virtually no background making the trigger especially effective. It also is useful for opposite side tagging for semileptonic B decays [31].

Multi-wire proportional chambers (MWPC) are used, using an $Ar/CO_2/CF_4$ gas, with

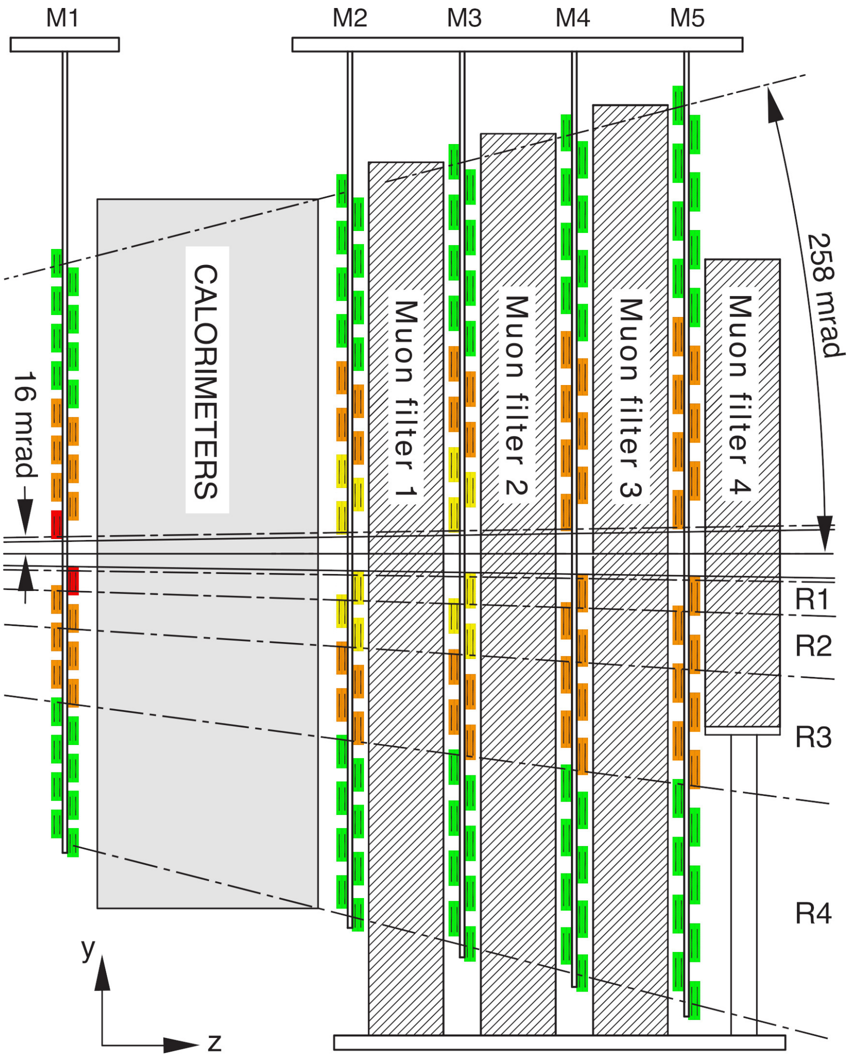
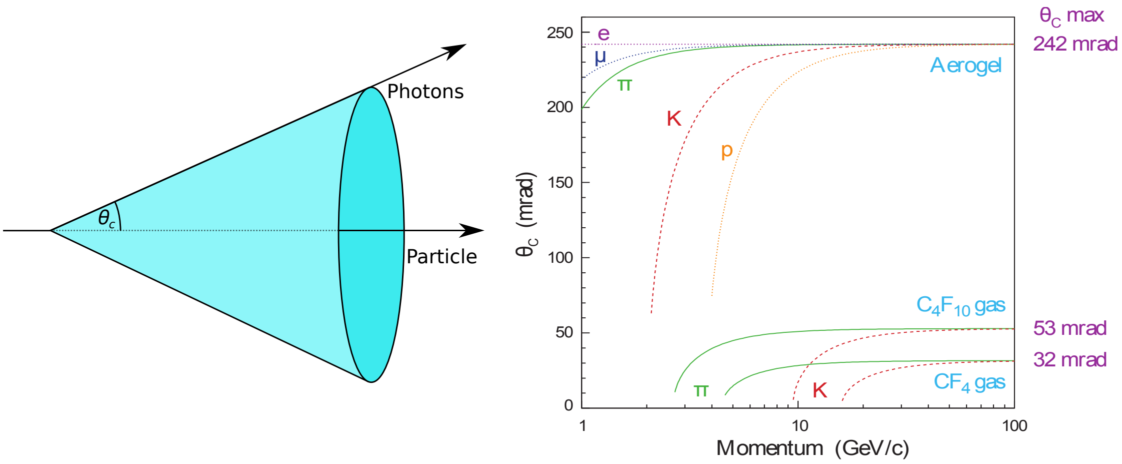


Figure 12: A side view of the muon system. The different regions from R1 to R4 are shown, based on the expected occupancy of the detector space. The R1 region in the M1 station are Gas Electron Multiplier (GEM) detectors in red. The rest are Multi-wire proportional chambers (MWPC) sensors with different readout technology; mixed wire-cathode pads in yellow, cathode pads in orange and wire pads in green [1].

276 chambers per station. The exception is the first trigger station that uses 12 triple layer GEM detectors in the inner most region which is subjected to the highest radiation and occupancy. The other areas have different readout systems for the MWPC technology to adapt to the different rates of occupancy and required spacial resolution which vary over the acceptance of the LHCb as shown in Figure 12.

Each station is split in two halves vertically which allows them to slide outwards to allow access to the beam pipe and other inner detectors for installation and maintenance, similar to the calorimeters [32].



(a) Cherenkov Angle θ_c. **(b)** Cherenkov angles versus momentum.

Figure 13: The Aerogel and C_4F_{10} is used in RICH1, CF_4 is used in RICH2, the Aerogel was removed in 2013. Using the momentum and the Cherenkov angle, likelihood values can be calculated for particles like pions or kaons.

2.9 Ring Image Cherenkov (RICH) detectors

The RICH detectors are important for distinguishing primarily protons, kaons and pions in final state charged particles and improves the yield for many decay channels of interest to the point that analysis is possible.

When a charged particle passes through the medium of a RICH detector, it polarises the gas and causes the wake to emit light. Since the particle is travelling faster than the speed of light in the medium, the photons constructively interfere to create a series of light cones emerging at a characteristic angle as seen in Figure 13a. The Cherenkov angle θ_c is related by

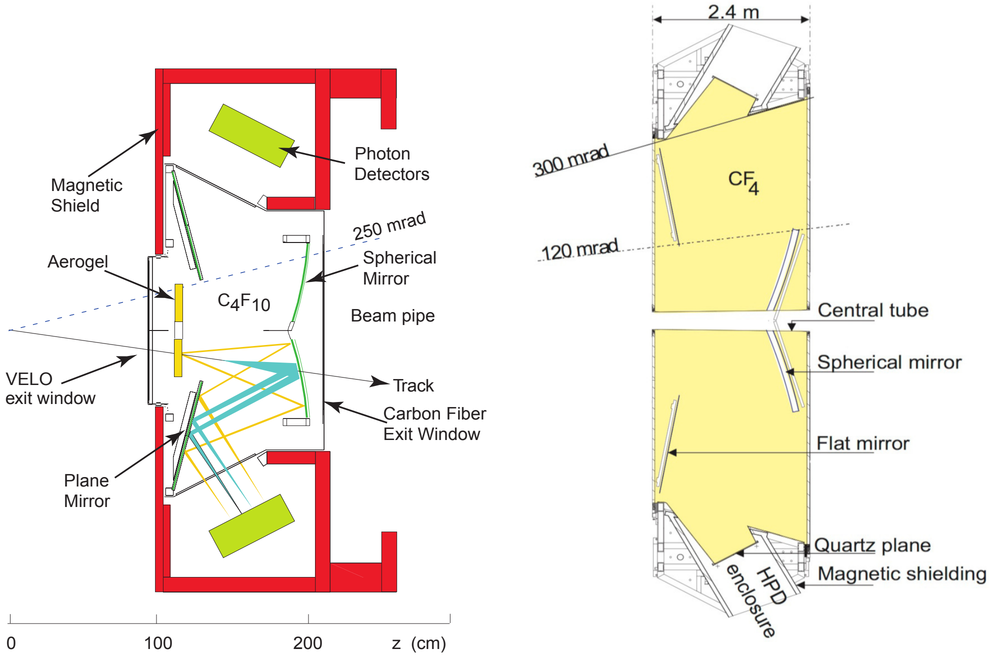
$$\cos\theta_c = \frac{c}{nv} \tag{1}$$

where v is the speed of the particle, c is the speed of light and n is the refractive index of the medium.

The Cherenkov light is collected on a spherical mirror, reflected on flat mirrors and onto a plane of photon detectors positioned at the focal plane of the spherical mirror. The optics of the mirrors ensure that the focal image on the photo-detector plane is independent of where the source of the light cone is along the particle's flight path. All the light from a charged particle will therefore form a ring onto the photo-detector plane where the size of the ring indicates the Cherenkov angle θ_c.

The detector mediums of RICH1 is C_4F_{10} and RICH2 is CF_4. Figure 13b shows how θ_c varies with particle momentum, by comparing the particle's deflection through the dipole magnet, to the Cherenkov angle from the RICH detectors, a likelihood of whether it is a pion, kaon or proton can be calculated and used for the trigger systems.

The aerogel has a promising refractive index that has potential to distinguish protons, kaons and pions in a lower momentum range that the fluorocarbons gasses could not give. However, its performance was not as good as expected and its mass reduced the momentum precision. The aerogel was removed in the long term shut down of 2013 [33].



(a) Layout of RICH1 (side view) **(b)** Layout of RICH2 (top view)

Figure 14: The RICH1 subdetector is located just after the VELO module, while the RICH2 is just after the tracking stations as seen in Figure 3.

The photon detectors used are Hybrid Photon Detectors (HPD) which are able to convert photons around 200-600nm with a quartz input window into photo-electrons which are accelerated in the HPD's vacuum vessel to strike onto a 32×32 silicon pixel array [1].

2.10 LHCb trigger

The LHCb uses a trigger system to discriminate the millions of events per second during its operation in real time [34, 35]. For example at LHCb design luminosity, 2×10^{32} cm^2 s^{-1}, it was expected that the rate of $b\bar{b}$ pair events occur at the rate of 100 kHz with a 15% subset with its decay products measurable with the LHCb's acceptance; *CP* violating and rare decays of interest have cross-sections that reduce the event rate three orders of magnitude. The hardware trigger layer, Level-0 (L0), operates at a rate of 40 MHz input and 1 MHz output. The software based High Level Trigger (HLT) was designed to output event data around 2 kHz [1] but it has been improved to output around 12.5 kHz and the luminosity has be increased, as detailed later in this section.

The L0 trigger takes 2048 channels from the pile-up detector in the VELO detector, 19420 channels from the calorimeter detectors, and 25920 channels from the muon detector and combines the data the Decision Unit (DU) to select potential event candidates.

The pile up detector is four planes of silicon strip detectors as seen in Figure 8, similar to the VELO modules detailed in section 2.4, and located in the negative z direction to the beam collision centre. The pile-up data can quickly determine if an beam crossing has multiple primary vertices and reject events that are too busy. It turns out the LHCb

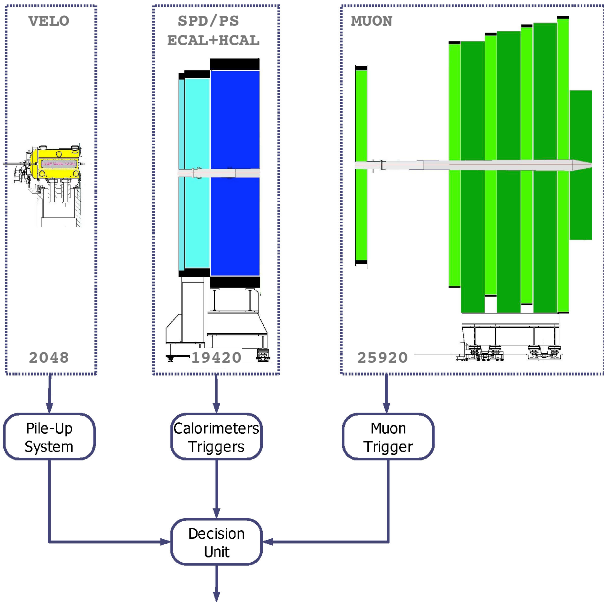


Figure 15: Level-0 hardware trigger which operates at 40 MHz and selects for high transverse momentum, energy and displaced vertices[1].

detector can cope with a bigger occupancy than expected which meant Run Two ran at twice the design luminosity. The pile up detector is not used as intended but it has been useful for specialised triggers focusing on low multiplicity events.

The calorimeter channels provide data on the transverse energy (E_T) of electrons, photons, π^0s and hadron candidates. The threshold for E_T is 3 GeV for e and γ, and 3.7 GeV for hadrons. These give an indication that the candidate event has some heavy flavour of interest.

The muon trigger gives information on high transverse momentum (p_T) which also indicates heavy flavour events. To pass the selection, muon tracks need to pass through all five planes of the muon detectors and have a p_T threshold of 1.5 GeV/c for single muon and average 1.3 GeV/c for di-muon candidates [36].

The L0-DU collects and buffers the channels coming from all the L0 sensors, with the large size of the LHCb detector, signals from the same event take different times to arrive at the DU depending on the location of its local sensor. The DU does simple logic within 2 µs to pass or fail each event and pulls out its data from the buffer to be passed on to the HLT

The HLT is a C++ software based trigger which is run on the Event Filter Farm (EFF) with thousands of CPU cores each executing an algorithm related to the HLT. There are two subsections to the HLT, the HLT1 and HLT2.

The HLT1 does partial reconstruction of events using similar data that L0 uses, such as transverse momentum estimation from track reconstruction, but in a more complete way. In this respect it confirms the L0 trigger and applies more stringent cuts and about 15% of events pass a subset of the sub-triggers of the HLT1.

The HLT2 uses all the detector information to reconstruct composite particles from each event's final state particles such as $K^* \to K^+\pi^+$ or $J/\psi \to \mu^+\mu^-$. More sophisticated selection like reconstructed mass cuts or the kinematics of heavy hadrons can be used.

These reconstructions and particle identifications can be organised into trigger lines, meta data that future software can use for sorting events.

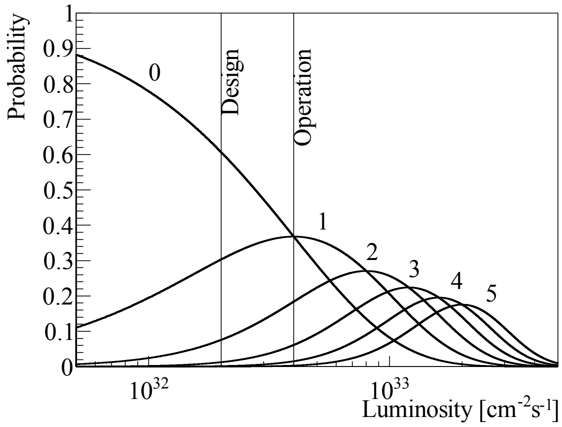


Figure 16: Rate of interaction probabilities of LHC bunch crossings at the LHCb [3].

It was found that the trigger and event reconstruction could handle higher pile-up, defined as the average number of visible interactions per bunch crossing. This allowed for higher luminosities around 3.5-4$\times10^{32}$ cm^2 s^{-1} [37], about twice the design luminosity as indicated in Figure 16.

As a software process, HLT has undergone many improvements to improve its efficiency of selecting a high volume of events at high speed. The amount of time in stable beam and data collection mode is only 30% of the time.

To take advantage of the rest of the 70% of down-time, the EFF was equipped with inexpensive consumer hard drives as a data buffer. The HLT software also gained flexibility defer incoming events while using its available CPU time to do more sophisticated event selection and have resiliency in case of data flow jams further down the line. The automatic alignment and calibration process of the LHCb also allowed the HLT to be more efficient. Overall, the design rate of the HLT improved from 2 kHz to 13.5 kHz by 2015 [37, 38], schematics of the trigger in 2011 and 2015 can be compared in Figure 17.

2.11 Data processing

From LHCb run 2, at a event rate of 12.5kHz, raw data is stored to tape drives on the Tier 0 and Tier 1 on the LHC computing grid. From here, the data is processed, or "stripped" in a similar way to the on-line trigger software where the output is stored as a Data Storage Tape (DST) file. In each event the values in each active channel is reconstructed into physical phenomena objects such as particle tracks, Cherenkov rings and calorimeter clusters with associated measurement uncertainties. The file sizes range from 200-250 kB per event with zero suppressed raw data down to a few kB for compressed data [39].

The DST file is processed to add PID and other information to give researchers the tools to query a set of trigger lines and events that meet conditional situations.

A compromise can be made to store a stream of compressed reconstructed events, about 5 kB per event, that cannot be reprocessed, but is good enough for most analysis. This is particularly useful for Charm physics where large numbers of events are produced

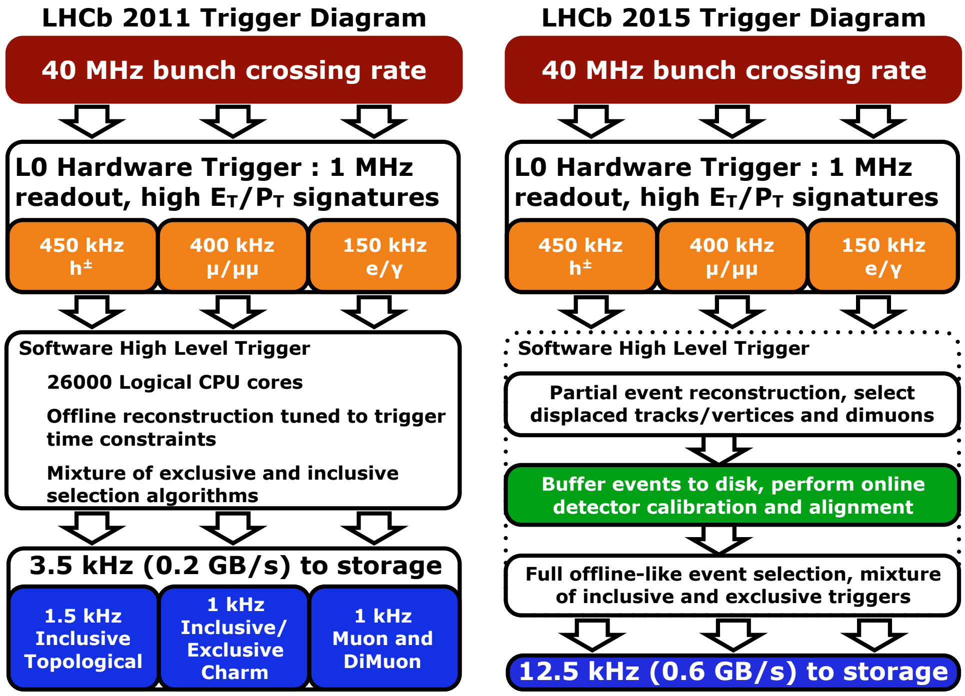


Figure 17: Schematic of the LHCb trigger system in 2011 and 2015. The HLT has had large performance improvements from having a buffer system.

and can now be stored for analysis with big data sets. Of the 12.5 kHz data rate from the EFF, about 2.5 kHz is allocated to a Turbo stream using these compressed events and the rest is Full Stream. The number of Turbo events outnumber the Full events [40].

The raw data from the Full stream is re-processed at intervals to provide new data sets that take advantage of improved or more sophisticated trigger lines.

In the future with luminosity upgrades of the LHC and LHCb, the amount of data output will increase by at least a factor of 30, so the Turbo stream will become the choice for managing large amounts of data with limited storage resources [39].

2.12 LHCb upgrade

The LHCb has completed two data collection runs, 3 and 6 fb^{-1} of integrated luminosity has been collected respectively as show in Figure 5, at a luminosity of up to 4×10^{32} cm^{-2} s^{-1}. For the LHCb upgrade [41] the aim is to collect about 50 fb^{-1} and the data rate to be 2×10^{33} cm^{-2} s^{-1}. The existing hardware trigger is a bottle neck; due to the way that the L0 trigger works, the yield of hadronic decays as luminosity increases would saturate as seen in Figure 18. The L0 trigger uses a threshold of transverse energy (E_T) in the calorimeter to reduce the rate of accepted events down to 1 MHz. The LHCb upgrade will have increased hadronic events but in order to avoid exceeding the 1 MHz limit, the E_T threshold would have to increase at a cost of a worse efficiency and overall there are diminishing returns in the event yield above a certain luminosity. The upgrade therefore will have the L0 trigger removed and data from all sub detectors will go direct to the software trigger at a rate of 40 MHz [42].

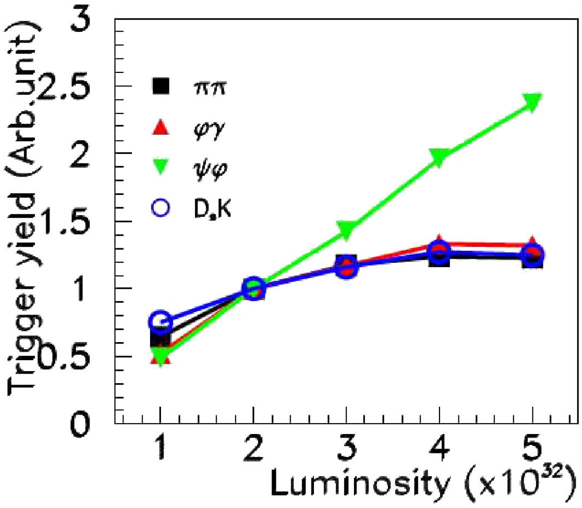


Figure 18: Trigger efficiency vs luminosity. In purely hadronic decays, the yield saturates at higher luminosity due to the existing LHCb hardware trigger.

2.12.1 VELO upgrade

The VELO upgrade [41] is a silicon pixel detector where some components of the old VELO are reused; such as the vacuum tank and motion stages for the VELO halves. Figure 19 shows a cross section of a 3D CAD model of the upgrade. There are 26 VELO modules in each VELO half, each have twelve ASIC sensors arranged in a L-shape as seen in Figure 20. When the VELO halves close, the L-shapes from both sides overlap slightly leaving a square gap in the middle where the beam passes.

The new detector module shape requires a completely new RF foil to follow the new geometry of the sensor upgrade. The inner diameter of the foil is also decreased to 3.5 mm from the beam centre due to improved confidence of the LHC beam stability. Within the acceptance of the VELO, the RF foil contributes more than half of the material budget before the second measured point; with more material comes a worse impact parameter resolution. Much effort has been done to engineer the thinnest possible foil that still has sufficient structural integrity and negligible leakage between beam vacuum and VELO vacuum. It is precision machined from a solid block of aluminum to a thickness of 250 μm and chemically etched in critical regions. The aim is to reduce the thickness to 150 μm in these areas which will be half that of the existing RF foil.

The silicon pixel sensors, 200 μm thick, are bump bonded onto a set of new radiation hardened ASICs called VeloPix designed to output event data at a rate of 40 MHz [5]. With 624 VeloPix ASICs across all modules, there is a total of 41 million pixels with a size of 55 x 55 μm^2 allowing an occupancy of around 0.125% for the pixels closest to the beamline, as close as 5.1 mm. There are six VeloPix ASICs on each module side which go to two output data cables per side.

Sets of 8 pixels are grouped 2×4 into super-pixels where the outgoing data packet includes 13 bits for the super-pixel address, 9 bits for event ID and a 8 bit hit map for each of the member pixels. This increases the likelihood that particle hits on the sensor that create a hit cluster will be included just in one outgoing data packet.

The VELO ASIC has 4 serialisers, Gigabit Wireline Transmitters (GWT) [43], that

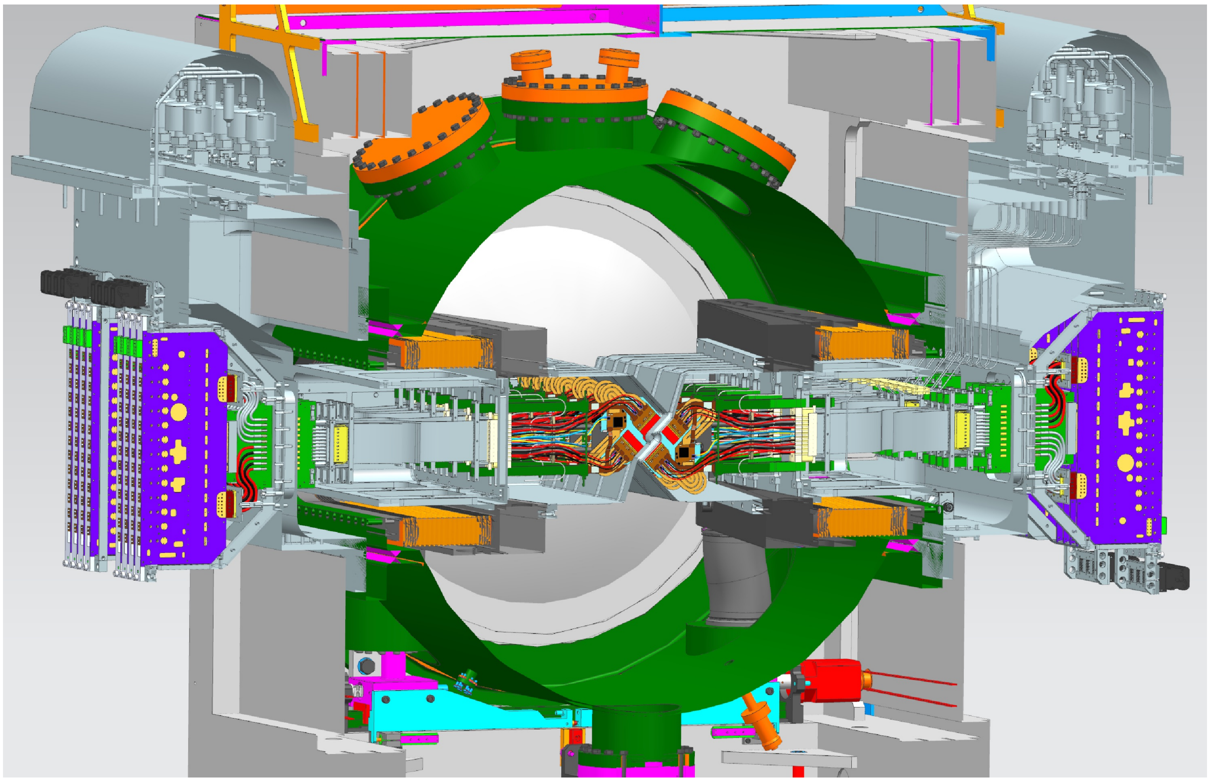


Figure 19: A cross-section of the VELO upgrade. The existing vacuum chamber (green) is reused. Starting at the centre where the LHC beam travels. The new RF foil (grey) fits the geometry of the pixel modules. The silicon pixel detectors (red) sit on the CO_2 cooled silicon substrate (light blue). Data is routed with cables (brown) from the sensor to the flex cable (green) which extends continuously to the vacuum-feedthrough (light green) and interfaces through the vacuum chamber side wall (grey)). This image does cut out the mid-section of the flex cables. The vacuum-feedthrough connects to the Opto Power Board (OPB) (purple) where external fiber optic cable take the data to the CPU farm where the High Level Trigger is executed [4]. Figure 47 shows a simplified CAD model and Figure 21 shows a photo of an assembled module.

group four of the 30 bit super-pixel packets into one larger data frame 128 bits long including a header. Each GWT can output up to 5.12 Gbits/s, this will meet the demand of the busiest ASIC outputting around 16 Gbits/s, Figure 20 shows the expected track density. Each output will be connected to a differential pair line on a flexible dataline tape as described in section 3.3.

The sensor/ASIC assembly is attached to a 400 μm thick silicon substrate with microchannel CO_2 cooling integrated within to provide effective cooling for the sensors [44, 45, 46, 47]. The cooling infrastructure for delivering liquid CO_2 at -30°C will be rebuilt with more powerful hardware with extra redundancy and will share with the new Upper Tracker sensors that also uses CO_2 cooling [42].

Figure 21 shows one of the first assembled modules which shows how the sensor fits with the cabling for the power supply, control and monitoring cables, and data lines. The separate circuit boards and curved kapton lines are designed to accommodate differences in thermal expansion across different material types to minimise stress on the silicon

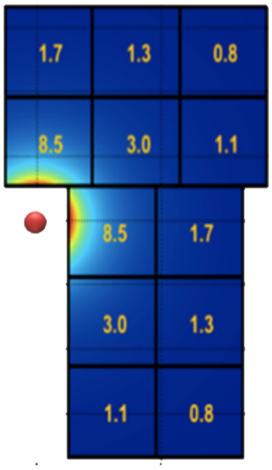


Figure 20: Expected track density, million tracks per second, in each of the 12 VeloPix ASICs of one VELO module. Peak data rate will approximately 2.8 Tbit/s. The inner most pixel will be 5.1 mm from the beamline centre [5].

substrate which contains liquid CO_2 of up to 50 atmospheres.

Cables with sufficient flexibility serve a conduit for control signals, power lines and outgoing data allowing the VELO halves to move between open and closed positions. Opto Power Boards (OPBs) are mounted on the side of the vacuum chamber and manage power supply and conversion of data from electric to light so that 1040 fibre optic cables can be used to transport 2.85 Tbit/s to the CPU farm, a solution that is physically manageable in the limited space available in the VELO corner of the LHCb underground chamber [41].

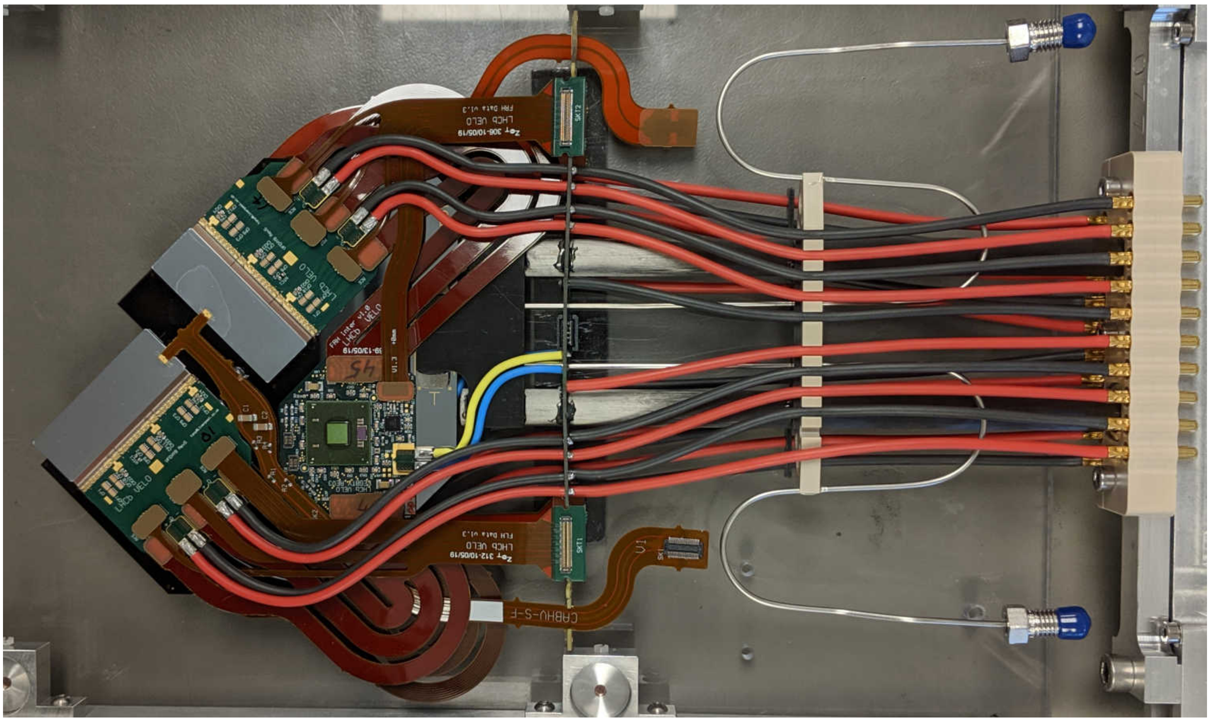


Figure 21: Photo of one of the first assembled VELO upgrade modules. The red and black cables are the VeloPix power supply while the brown kapton cables route high voltage to the sensors, control and monitoring signals, and sensor data. The steel S-bend pipes house the CO_2 for microchannel cooling.

2.12.2 Upgrade of the rest of the LHCb detector

The existing Tracker Turicensis (TT) [1] and Tracking stations downstream of the dipole magnet will be replaced with new detector technology.

The TT will be replaced with a new silicon strip sensor detector called the Upstream Tracker (UP) [48] which will, in the new luminosity, have better radiation tolerance and occupancy, 40 MHz readout and improved coverage close to the beam pipe.

The three downstream Tracking stations, which includes the Inner and Outer Tracker, will have a similar three station configuration, will use a novel detector technology; the Scintillation Fibre Tracker (SciFi) [48]. The panels are full of close packed embedded scintillating optic fibres 2.5 m in length, light induced by particle interaction are directed to the fibre ends where Silicon PhotoMultipliers (SiPM) will read out the signals at the 40 MHz rate.

The benefits of having a single detector technology for the whole tracker, with sufficient performance and cheaper than silicon sensors per unit area was considered worth the time and effort in developing this new system. There is also less dead time compared to the old straw tube trackers so SciFi will have a better spill over performance.

The polystyrene fibres are doped with a fluorescent dye chosen to have a fast response time within several nano seconds, good quantum efficiency over 95 % and useful emission spectra. It is also to be radiation tolerant enough to perform sufficiently over an integrated luminosity of 50 fb^{-1} with around 11×10^{14} $\mathrm{MeV/cm}^2$ Neutron Equivalence fluence per area at the hottest part [48].

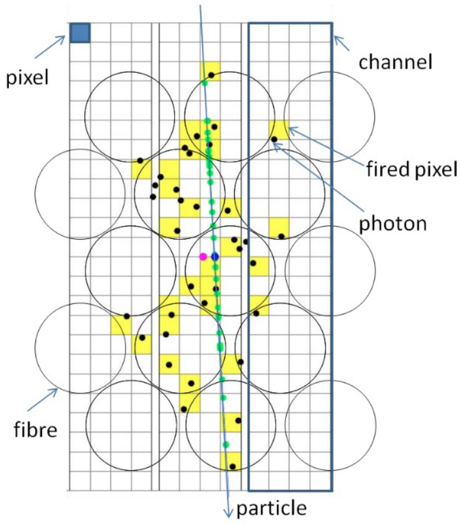


Figure 22: A schematic cross-section of the SciFi detector. The fibres are closed packed and the Silicon PhotoMultipliers (SiPM) are placed on the fibre ends. The SiPM channels indicate particle position on one axis while time of flight of photons down the fibres yield measurement on the other axis normal to the beam direction.

It was found from irradiation tests that the fibre becomes less transparent, resulting in up to 40% signal loss but is still sufficient for the SiPM to perform. Having mirrors at the far end of the fibres from the SiPMs is necessary to enhance the light signal so it can travel the full length of the irradiated fibre, but it does complicate measurements for signal timing and spillover. The SiPM has a new cooling system to keep it at -40°C in order for an irradiated SiPM system to function correctly

The fibre core has two cladding layers of decreasing refractive index, with a total diameter of 250 µm, to increase the range of light angles up to 27.4° that will undergo total internal refraction in the fibre and thus will be captured and detected by the SiPM.

The SciFi modules are made from 6 layers of close packed fibres, mirrors mounted at the fibre ends closer to the beam and SiPM mounted at the other end. To convert the analog signals from the SiPMs, the PACIFIC chip [49] is used to output data packets compatible for the software trigger.

The RICH detectors will have the HPDs replaced with commercially available Multi-anode Photo Multiplier Tubes (MaPMTs), specifically the Hamamatsu R11265-103-M64 MaPMT. They will have a custom ASIC readout chip, the "CLARO" which will digitise the MaPMT output at a rate of 40 MHz [50].

The RICH1 will have a larger light sensitive area in the new build and without the aerogel, the one third area dedicated to aerogel ring images can be devoted to C_4F_{10} gas images. Thus a new spherical mirror has been constructed for a RICH1 upgrade with lower occupancy. The changes to RICH2 will only be to change the MaPMTs, readout electronics and new associated mountings.

The calorimeter subdetectors will largely be the same except for the replacement of the readout electronics. The PMTs used in the calorimeters will be retained, but to extend their lifetime until the next upgrade cycle, their gain will be reduced by a factor of five while the signal gain will be increased in the front-end electronics. The electronics will also deal with the higher chance of spill-over of PMT signal between two consecutive beam collisions 25 ns apart.

It is expected that the calorimeter detector modules will tolerate the expected radiation exposure between Long Shutdown 2 and 3, and then only the inner cells will need replacing after that. The SPD and PS will be removed, while it was useful in the L0 trigger it is not required in the context of a software trigger and will simplify the calorimeter system.

The Muon detectors front-end electronics actually already read out at 40 MHz but will need new off-detector readout electronics to be compatible with the software trigger. The M1 station will be removed; it's usefulness was in L0 trigger and won't be relevant for 40 MHz readout. While the remaining muon detectors are well shielded, there is a region near the beam pipe just downstream of the calorimeters which will need extra shielding due to increased hadronic showers. Another region that is sensitive to background is the outermost regions of M5 which has channels representing large areas up to 0.5 m^2; this makes it vulnerable to back-scattered particles from LHC infrastructure, even more so for higher luminosity. Therefore the granularity of this region will be increased by modifying the front-end [51].

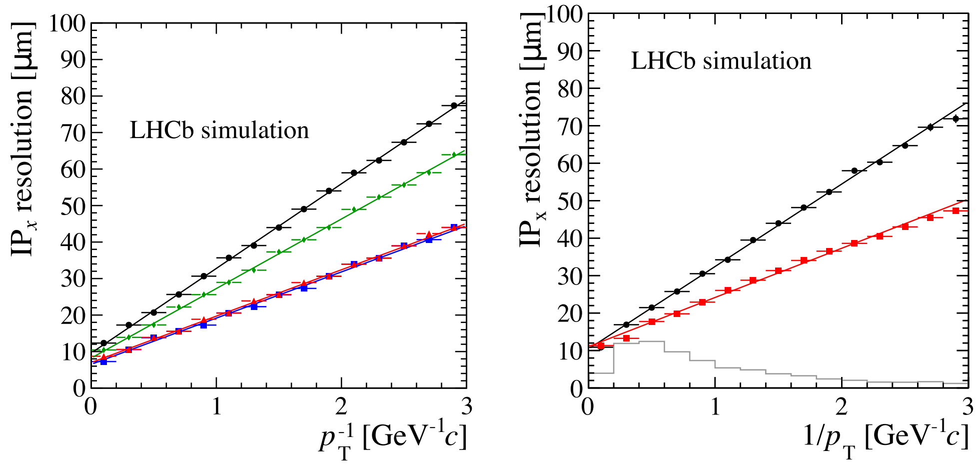
The new software trigger is designed to handle 4 TByte/s. To minimise cost of the hardware needed to process the new data rate, the Event Filter Farm (EFF) will be located on the planet surface in a new building where nearly 20 000 fibre optic cables 300 m in length will connect from the LHCb sub detectors. Cooling, power and accessibility will be easier than if the EFF was based in the underground LHCb chamber.

As described in Section 2.10, the software trigger structure is similar with HLT1 and HLT2 but will be implemented on a larger scale. The systems of data buffering and Turbo streams, proven in Run 2, will be used to enhance data processing and storage. While the LHC beam bunches are separated at 25 ns intervals and the LHCb front-end reads out at 40 MHz, there are gaps in the bunch "trains", so the average frequency of events is around 30 MHz. On a larger time scale of orders of hours and days there is downtime between each LHC beam fill and other technical reason that prevent active beam collisions. Therefore during data collection, the computing resources will be prioritised to the HLT1 software trigger and excess intermediate data can be buffered so it can be processed through the HLT2 software when computing resources are available. The EFF is expected to have about a week's worth of HLT1 processed data giving plenty of time to work through any backlog. If there is downtime longer than this, computing resources can be devoted to Monte Carlo production of simulated events. This general flexibility allows for a smaller CPU farm that what would be needed without a buffer and allows for slower but more sophisticated HLT2 algorithms to be executed in good time [52, 53].

3 VELO upgrade work

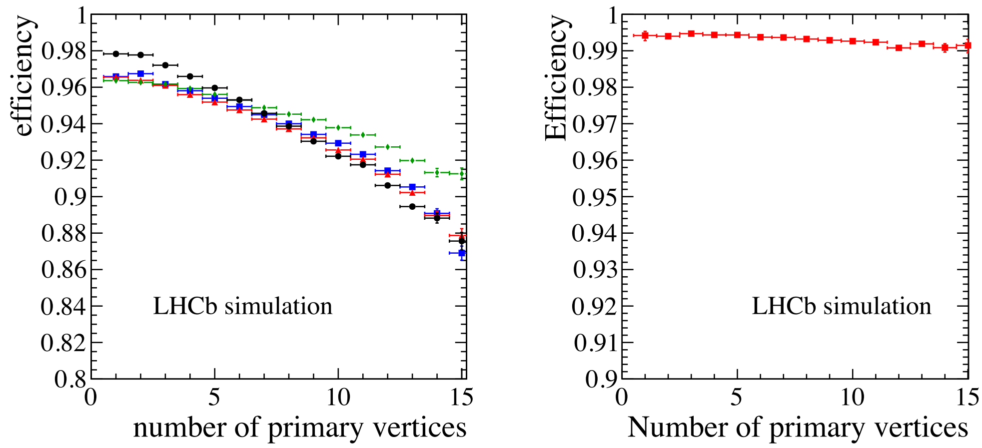
3.1 Strip prototype assembly

At the time of the hardware prototyping, the pixel detector described in Section 2.12.1 and a strip silicon detector [54] were being prototyped. The pixel sensor has larger material budget due to the ASIC being directly under the sensor layer and a reduction in impact parameter resolution is shown in Figure 23 compared to a strip sensor.



(a) Strip detector simulation [54]. **(b)** Pixel detector simulation [41].

Figure 23: Comparison of Impact Parameter resolution on x-axis vs inverse transverse momentum for tracks within $2 < \eta < 5$. Figure 23a: black - original VELO, green - Letter of Intent layout, red compact - layout with TPG cooling, blue - compact layout with microchannel cooling. Figure 23b: black - original VELO, red - pixel upgrade.



(a) Strip detector simulation [54]. **(b)** Pixel detector simulation [41].

Figure 24: Reconstruction efficiency vs number of primary vertices using simulated events of $B^0 \to K^{*0}\mu^+\mu^-$ decays with particle momentum greater than 5 GeVc^{-1}. The coloured lines in Figure 24a are: black - existing VELO, green Letter of Intent layout, red - compact layout with TPG cooling, blue - compact layout with microchannel cooling. Figure 24b: Reconstruction efficiency for the pixel upgrade, estimated from simulation.

The strip sensor had a 25 µm minimum strip width closest to the beam compared to the pixel's uniform 55 µm size. When considering the reconstruction efficiency of tracks for high number of primary vertices, important in the higher luminosity situation of Run 3, the pixel sensor is significantly better as shown in Figure 24. Hence this was the general basis for selecting the pixel detector for the VELO upgrade.

This section will focus on the strip sensor prototype, an evolved design compared to the existing VELO detector manufactured by Hamamatsu designed for the high luminosity environment. It looks into the performance of the strip prototype and thermal simulations to see if it was a feasible option at the time.

New features compared to the existing VELO strip sensors are, a smaller inner radius to improve impact parameter resolution, smaller pitch and strip length to improve occupancy and tracking resolution, and for the Φ sensor a different routing design that allows for a much simpler mapping of strips to channel number onto new readout chips.

3.1.1 Module construction

The aim when building the prototype sensor module was to use the hybrid readout chip from the current LHCb silicon tracker to read out a small proportion of the sensor shown in Figure 25; this speeds up assembly using a proven solution. The hybrid accommodates Beetle chips [25] which are readout chips used in the current VELO detector and can then be read out to the Alibava [55] or TELL1 [56] electronic readout assemblies that use Field Programmable Gate Arrays (FPGA) to process the Beetle output to data files.

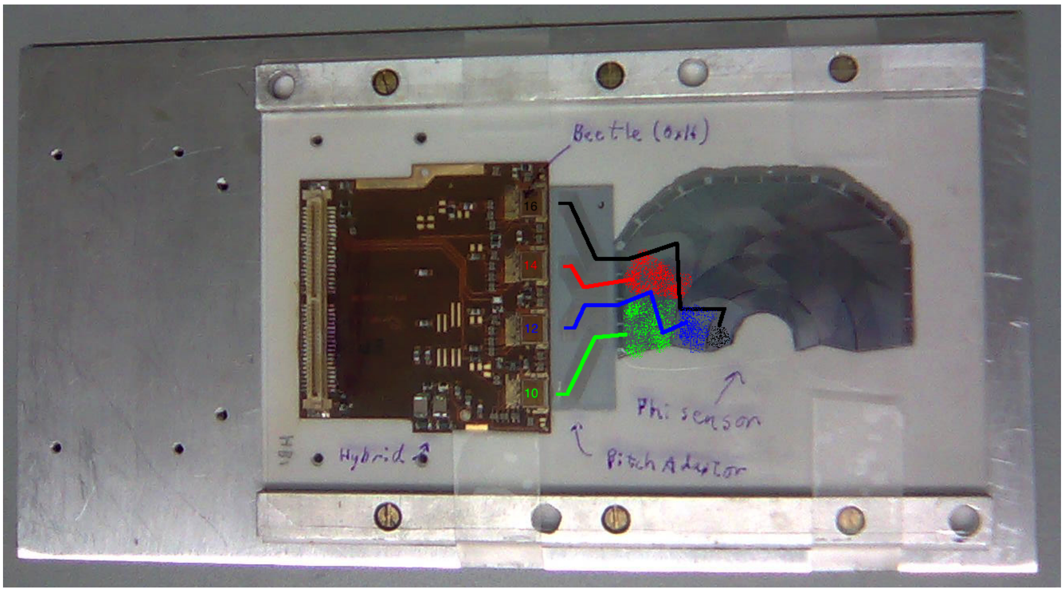
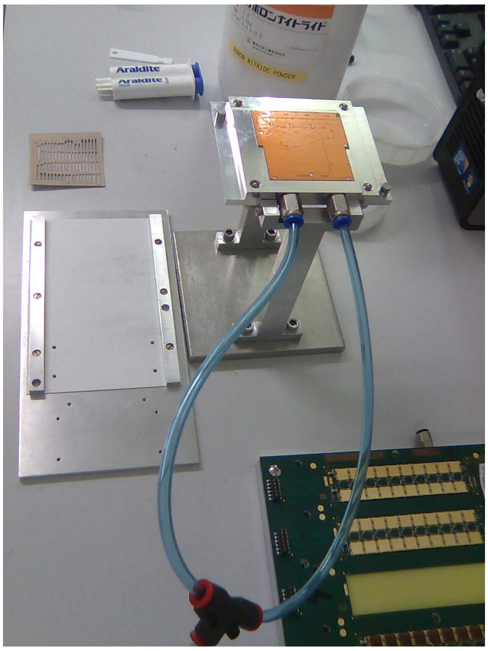
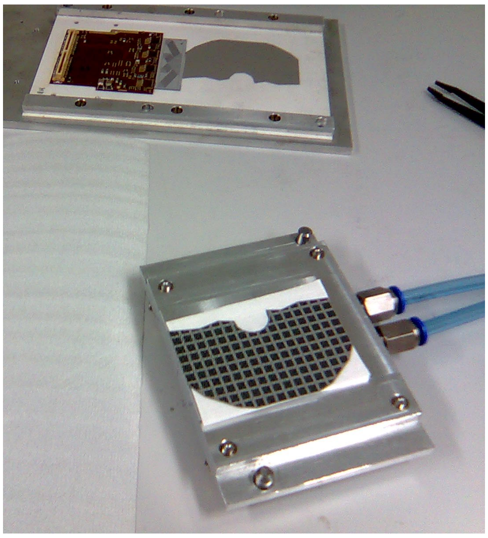


Figure 25: Ceramic baseplate attached onto aluminium baseplate. Coloured lines show routing from sensor sector to Beetle chips.

The module was assembled onto an aluminium nitride (AlN) ceramic base 0.65 mm thick that provides a heat conductive flat surface. It consists of the front-end hybrid to house the Beetle chips, a pitch adaptor to route the channels from the sensor to the Beetle chips and the prototype sensor itself.



(a) Just before gluing hybrid.



(b) Staystik® glue is easily cut into complex shapes, clean room paper protects the sensor from the aluminium jig.

Figure 26: The two vacuum gluing jigs in action

Two gluing jigs were designed and made to aid placing the hybrid and sensor in the right place consistently as seen in Figure 26. Two different gluing options were tried; Araldite® an epoxy resin mixed with boron nitride powder in a 2:1 ratio to give good thermal conductivity and Staystik®, a thermoplastic glue available in 127 µm thin sheets. Both have their own advantages: Staystik® is much easier to handle since a complex shape can be cut out as seen in Figure 26b, the thickness of glue is inherently uniform within the manufacturing tolerances and it is also removable. The glue has to be heated over 125°C to bond which rules out its use with thermally sensitive items such as irradiated sensors. Araldite® has to be carefully dispensed to avoid excess or trapped air bubbles and effort has to be taken to ensure the right thickness and uniformity, the gluing jigs aid this process with 4 micro-jacks.

Once the hybrid was glued to the ceramic, surface mount components are soldered onto the hybrid and cleaned before Beetle chips are glued in position with electrically conductive Staystik® 571 to ground the chip. A pick and place machine was used to give high accuracy; a tolerance of approximately 10 µm is desirable to help make the wire bonding process reliable.

The pitch adaptor used is a 300 µm glass substrate with a metal layer that has been laser etched to give the correct routing as seen in Figure 25. This was made at the University of Santiago de Compostela and is a cost effective solution that gives sufficient precision. The close proximity of the bond pads seen in Figure 27 is a challenge to wire bond reliably; the small 40 µm pitch is not an issue from left to right, the wires can loop above each other but require more space in the parallel direction, about 150 µm. A new pitch adaptor design factored in the extra space, with reference marks and dedicated power routing which is ready to be used in the next prototype module. Overall about 95% of channels are functional between the Beetle chips and sensor. Some issues encountered

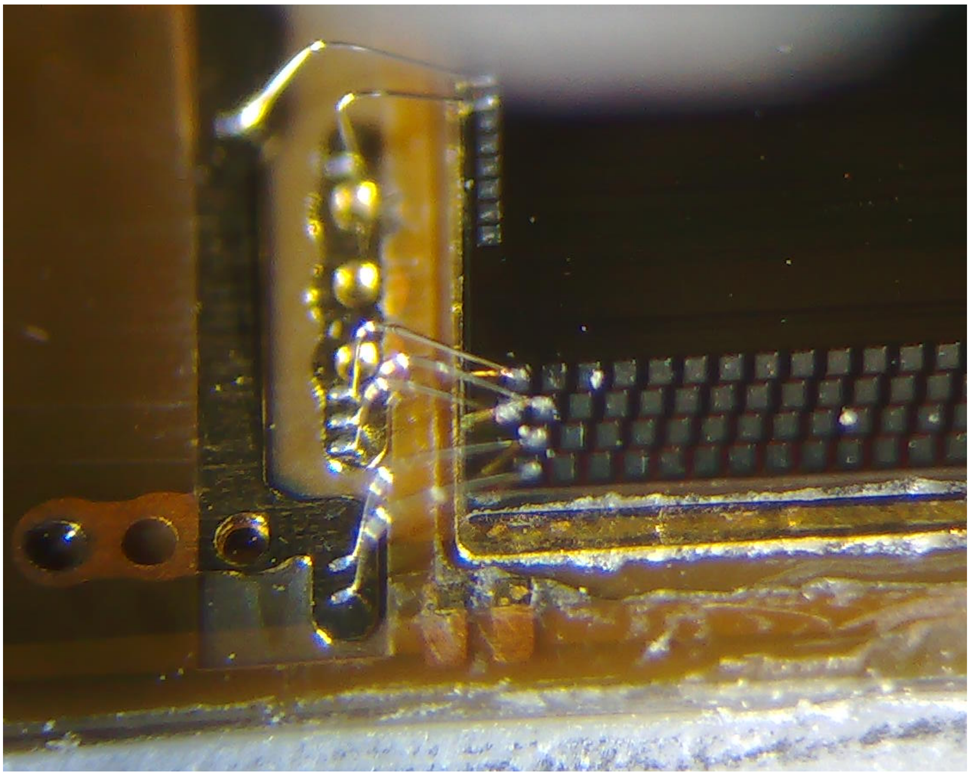


Figure 27: Close up of beetle wire bonds - before wire bonding to pitch adaptor

are that the metal layer is sometimes broken and the substrate can be chipped.

The Φ sensor was glued using Staystik® 672, seen in Figure 26b, which is electrically non-conductive to insulate the ground plane of the sensor.

The steps required in the assembly of the strip prototype module are summarised as follows:

- Hybrid glued to ceramic;
- Hybrid populated and cleaned;
- Beetles chips glued with Staystik;
- Beetles chips wire bonded;
- Beetle pedestals verified with Alibava;
- Pitch adaptor and sensor glued in place;
- Pitch adaptor and sensor wire bonded.

3.1.2 Characterising the Φ strip sensor

The characterisation of the Φ sensor was undertaken by Alex Morton, a MSci student at the University of Glasgow [6]. The sensor is equivalent to a diode and when a reverse bias voltage is applied to the sensor, very little current flows. The depletion volume in the silicon does increase as voltage increases; with a larger depletion volume, more electron hole pairs created from thermal energy can conduct current. Therefore a leakage current increases up to the point where the entire volume of the silicon material is depleted and the leakage current saturates. More information about depletion volume in silicon is discussed in section 3.2.2.

The plot of current vs bias voltage of the sensor in Figure 28 should show a plateau in current as the reverse bias increases. However, the leakage current continues to increase, albeit more slowly, after 60 V. This could be due to second order effects arising from the increased field in the device and a slight increase in depletion volume as that volume extends laterally close to the edges of the sensor. The depletion voltage can also be estimated by measuring the sensor performance as a function of voltage.

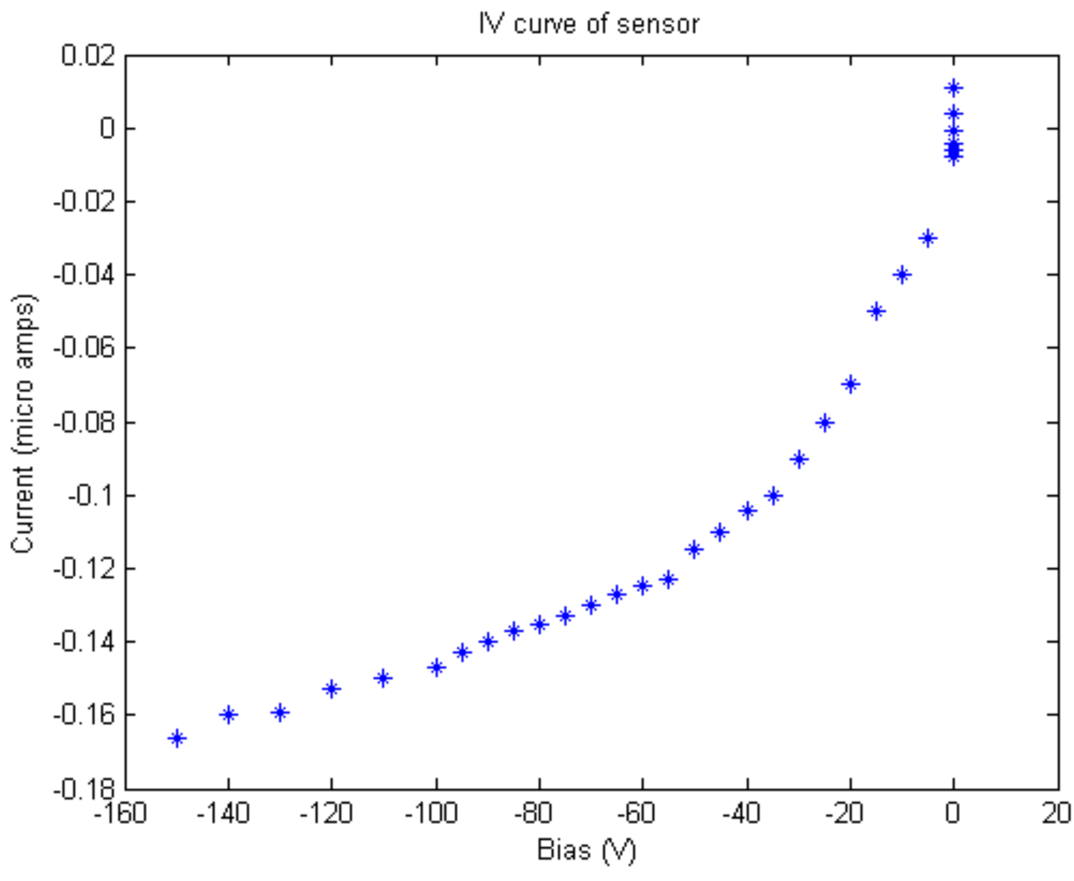


Figure 28: A plot of current vs bias voltage of Φ sensor shown in Figure 25 [6].

The prototype module was placed in a dark box where a ^{90}Sr source was placed above the sensor, while a scintillator was placed underneath to provide an asynchronous trigger to the Alibava readout system. There are two populations of beta decay; from ^{90}Sr with end point energy 0.54MeV and ^{90}Y with endpoint energy 2.28 MeV [57]. The 200 µm thick silicon and 650 μum thick AlN ceramic will filter out low energy electrons to the scintillator, the lower energy threshold was not quantified.

For Figure 29 and 30 the measurements came from the middle region of the prototype sensor, shown in blue highlighting in Figure 25. A pulse shape can be accumulated with the timing information between the asynchronous trigger and 40 MHz readout as seen in Figure 29. A spectrum of particle energy is shown in Figure 30.

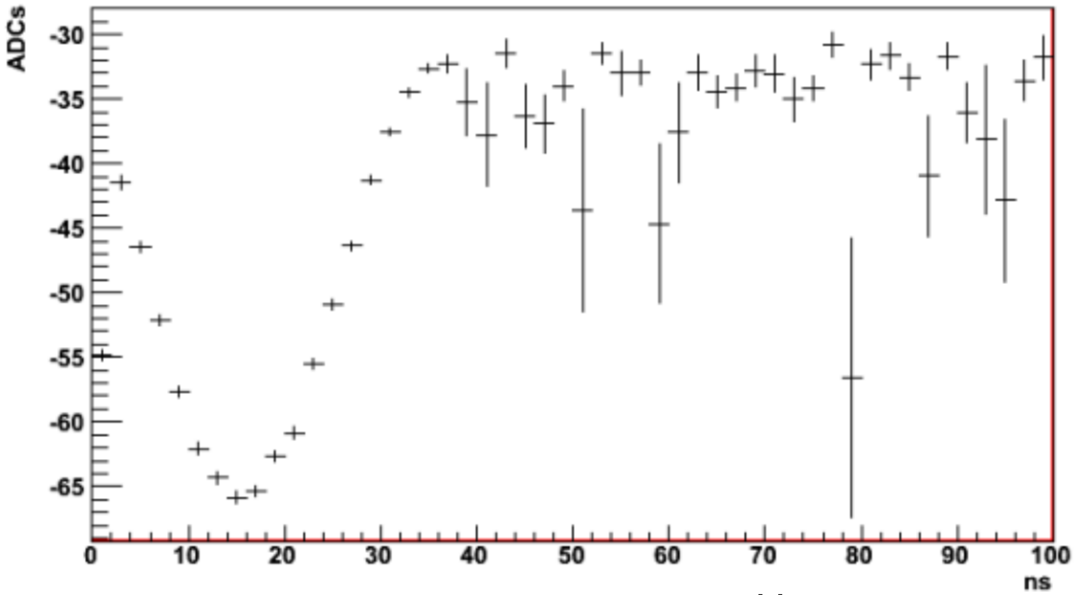


Figure 29: Pulse shape histogram [6].

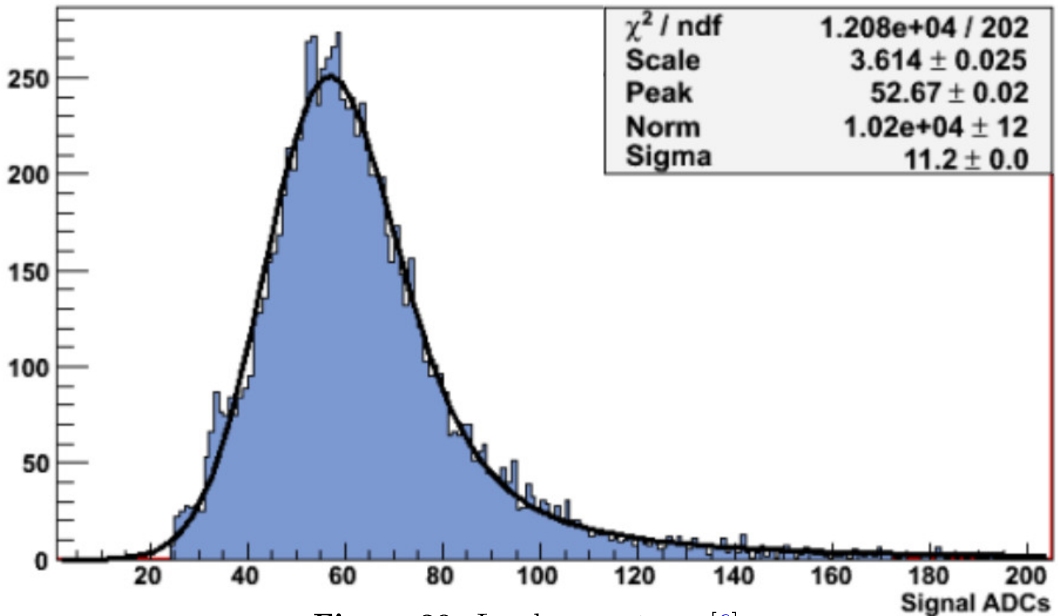
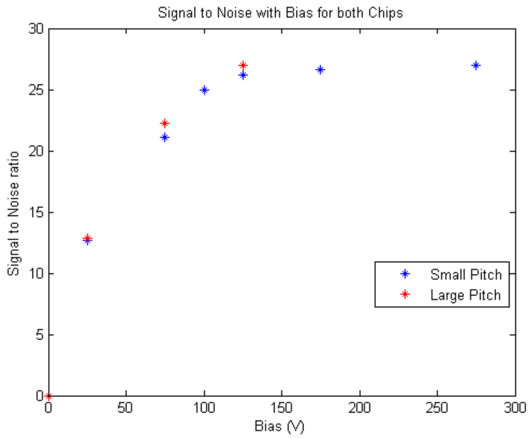
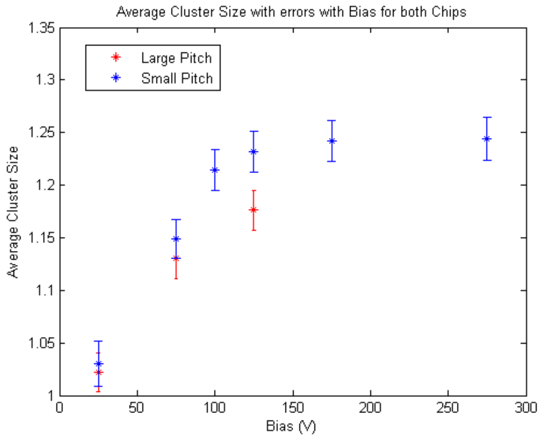


Figure 30: Landau spectrum [6].

The signal to noise ratio as a function of voltage is shown in Figure 31a, the ratio saturates around 125 V as the depletion volume is maximised and thus more charge is collected from electron hole pairs caused by the ^{90}Sr emission.



(a) Signal to noise. At 125V; signal is 16000 electrons, noise is 590-610 electrons [6].



(b) The depletion voltage is consistent with Figure 31a [6].

Figure 31: Voltage scans for S/N ratio and average cluster size. Small and Large pitch refers to the blue and green highlighted region in Figure 25.

The final measurement of average cluster size vs voltage is shown in Figure 31b. As the voltage increases, so does the depth of the depletion region which allows charge to drift laterally towards more than one diode. A narrower pitch will also facilitate charge sharing. The measurement confirms a consistent depletion voltage and a narrower pitch effect.

3.1.3 Concluding remarks on module construction

A prototype strip sensor for the LHCb upgrade was assembled with a pre-existing readout board, custom pitch adaptor. It was shown that the prototype reasonably performed as expected during reverse bias tests. Basic tests were done with a ^{90}Sr source and a Landau spectrum was re-created. Tools for the assembly were designed and manufactured and the process of assembly was mapped out, identifying challenging aspects such as needing a reliable pitch adaptor better suited for wire bonding.

3.2 VELO strip prototype cooling

An active silicon sensor has a cooling challenge; the high voltage supply required for the sensor to function will generate heat. As the sensor gets irradiated over its lifetime, the voltage requirement gets higher and more heat is generated. A cooling solution will keep the sensor functional and prevent the sensor from overheating

This section explores the work on exploring the different cooling solutions, the nature of irradiated silicon and implementing a thermo-electric simulation of a likely module design to see if the strip sensor prototype could be kept at a stable temperature towards the end of its irradiated life. This work concludes that after a 50 fb^{-1} lifetime it does look viable.

3.2.1 Conceptual cooling solutions

The conceptual designs of an upgrade VELO strip module and its cooling solutions were explored using the CAD software Solidworks. It is a useful tool to see how all the elements of the module may fit together and provides a 3D model for finite element thermal simulations.

The major difference from the existing VELO module is to have a more integrated cooling design. The current VELO has a Thermal Pyrolytic Graphite (TPG) substrate reinforced by carbon fibre and a separate cooling CO_2 plate bolted onto the end of the module. Thermal information is limited since there are only two thermistors.

Effective cooling is more imperative due to higher radiation levels in the expected lifetime of the upgrade sensors; this will increase leakage currents by an order of magnitude which will require removing more heat to avoid a thermal runaway as discussed in Section 3.2.2.

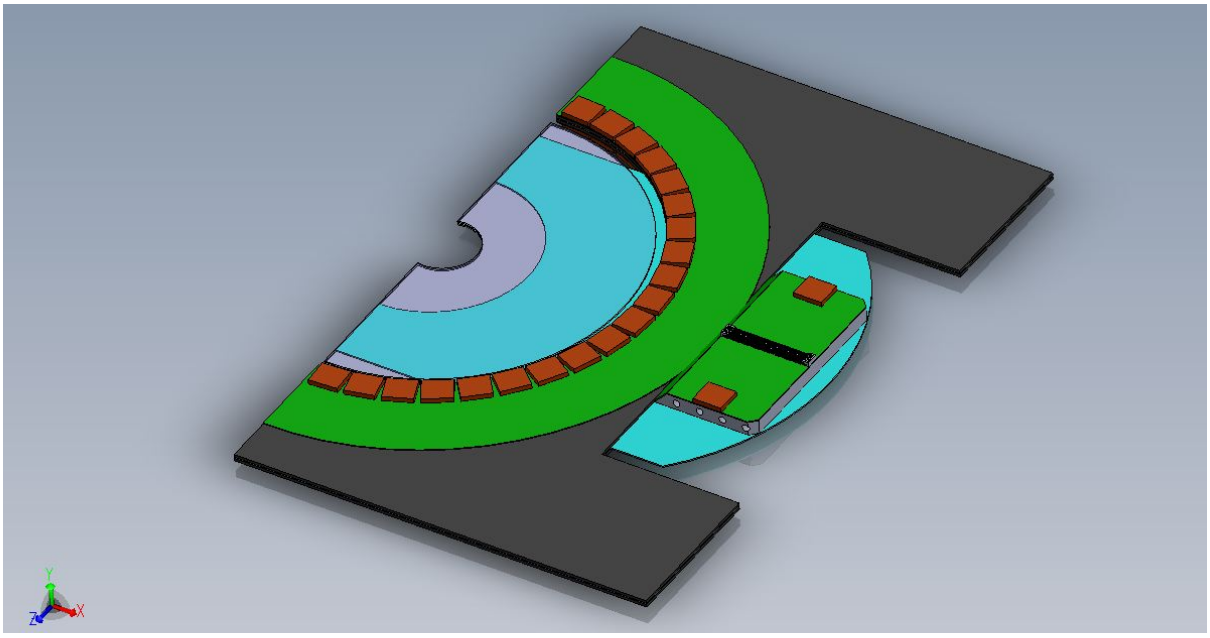


Figure 32: Blue - diamond, black - carbon fibre, green - kapton PCB, orange - readout chip, grey - sensor. This design works around the size limitation of a 50 mm radius diamond wafer.

Synthetic Diamond was considered as a substrate material as it offers an unprecedented thermal conductivity of about 2000W/mK. The type of diamond considered is polycrystalline and formed by Chemical Vapour Deposition (CVD), however it is expensive and is limited in size to 10 cm diameter wafers. In Figure 32, the diamond could potentially be attached to one large cooling block with invar pipes carrying liquid CO_2. Invar is an alloy designed to have a negligible coefficient of thermal expansion and is useful for avoiding mechanical stress due to temperature gradients in the materials.

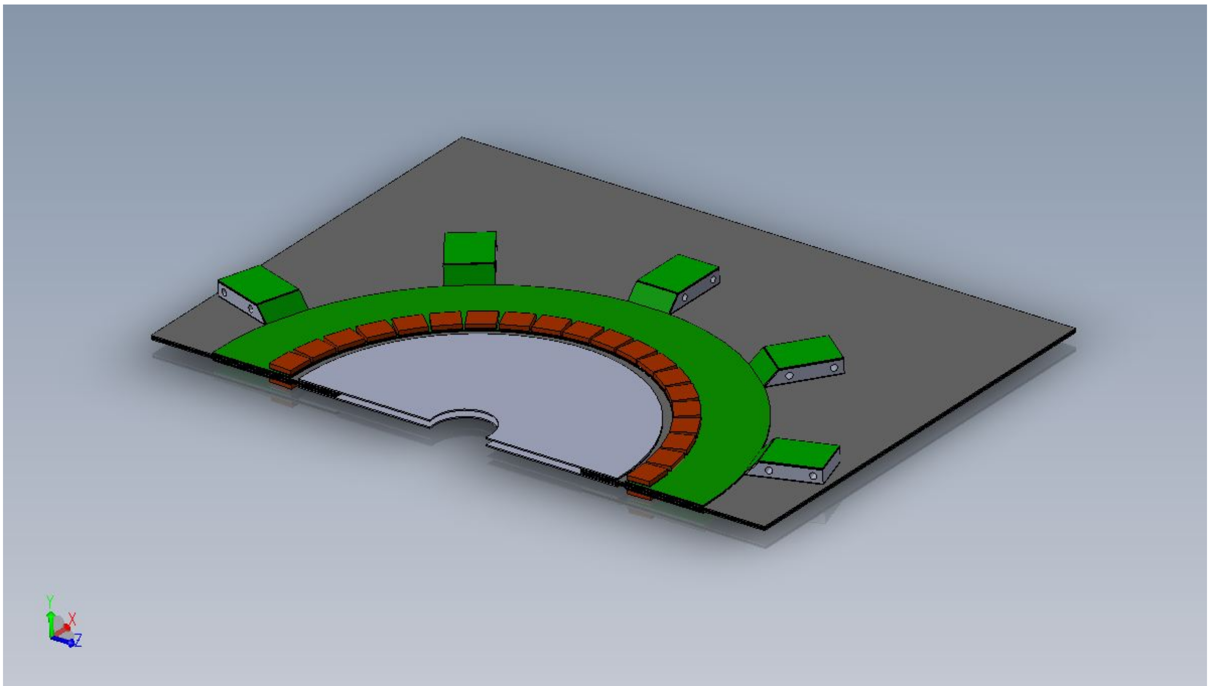


Figure 33: TPG substrate design. Connectors would be placed on top of the cooling block. A microchannel design would be very similar but without the cooling block

TPG as the thermal substrate does not have the size restriction and prohibitive cost of diamond and has a thermal conductivity around 1700 W/mK along the plane of the wafer. The conductivity in the perpendicular direction is poor, about 10 W/mK but that could be ameliorated by cutting diagonally into the TPG where the cooling block touches all layers of TPG. The thermal performance therefore is potentially similar to diamond. The cooling block itself would be made of TPG layers and it would be an efficient use of space to put connectors on its top surface as shown in Figure 33.

A third solution is to use microchannel bi-phase CO_2 cooling where a silicon substrate with hollow conduits transports the CO_2. Microchannels are constructed by using a deep reactive-ion etching process on a silicon wafer to cut out sub-millimeter grooves, a second silicon wafer is bonded on top and the ensemble is thinned to the desired thickness. CO_2 enters into narrow channels which then widen to larger channels causing the CO_2 to boil and utilises the latent heat of evaporation to aid the transfer of heat. It is the most effective cooling solution as the microchannels can be routed directly underneath the heat sources and has the advantage that silicon substrate on silicon sensors avoids mechanical stresses from mismatches in thermal expansion. It is a new concept however and has to be engineered well to avoid catastrophic rupture from high pressure CO_2.

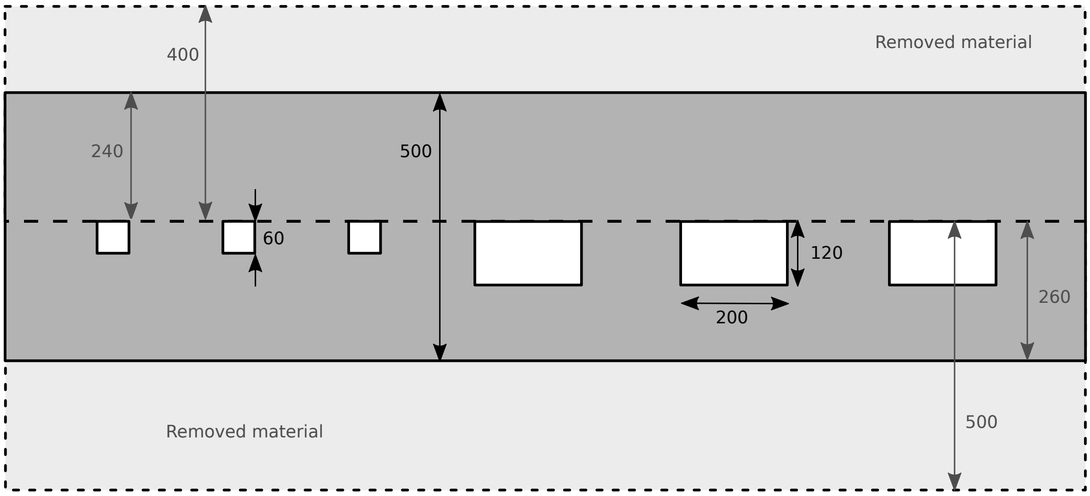


Figure 34: A microchannel cross section according to the chosen LHCb upgrade design. Dimensions in micrometers, the small channels are square.

In the implemented upgrade, microchannel cooling was the chosen technology where the total thickness of the silicon substrate is 500 μm. The operational coolant pressure is about 20 bar CO_2 but extensive testing was done to ensure plenty of safety margin even if the substrate is at room temperature where CO_2 pressure can go up to about 60 bar, thus the cooling system has been validated up to 186 bar [44, 45, 46, 47]. Figure 34 shows a visual representation of a cross section of the microchannel substrate.

The physical model shown in Figure 33 with TPG cooling blocks is used in the thermal simulation as while it is the simplest design to manufacture, it has the least capability to cool an irradiated silicon sensor. If it is found to be viable for this cooling method, the margin of safety with the other cooling designs should be larger.

3.2.2 Modelling irradiated silicon sensors

A thermal simulation of the strip prototype sensor needs to incorporate the behaviour of silicon detectors at high radiation exposure. The VELO's close proximity to the LHC beam means the radiation dose distribution is very non-uniform, varying both in radial position and location along the beam axis. The change in fluence in neutron equivalence n_{eq} in the radial direction is shown in Figure 35 [7].

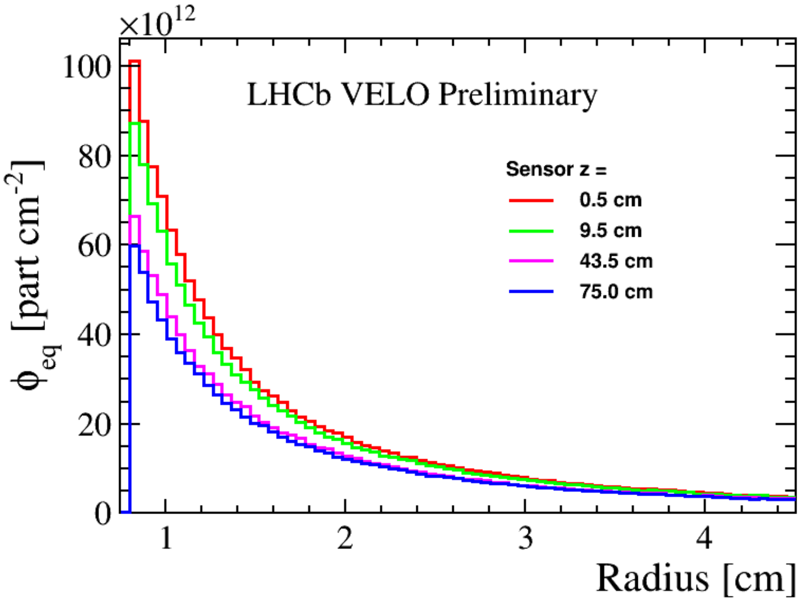
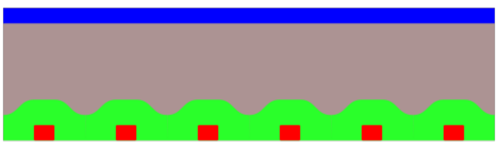
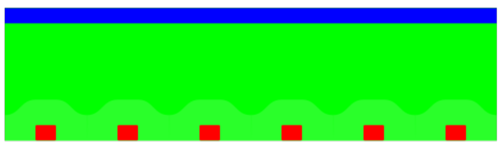


Figure 35: The expected fluence vs radial position of sensors from the beam axis. The Maximum fluence is around $1{\times}10^{14}$ 1 MeV n_{eq}/cm^2. Along the z axis the fluence is less further away from the LHC beam interaction point [7].

Almost all of the VELO silicon sensors have n-type segmented implants on a lower concentration n-type bulk and a p-type uniform implant on the back side, n^+-on-n; two sensors have n^+-on-p, just having p doped silicon bulk with n-type implants. The boundary between the different doping concentrations have natural depletion zone; free elections diffuse from the zone of higher concentration, leaving electron holes behind, and an electric field grows in the diffusion zone eventually reaching an equilibrium balancing the diffusion rate with the the electrostatic forces as shown conceptually in Figure 36a.



(a) No bias with natural depletion zone



(b) Fully depleted volume.

Figure 36: Schematic of silicon sensor when a bias voltage is increased between the top surface and the n-type junction. The depletion zone increases to the full volume as shown. Red is n^+-type with higher doping concentration, grey is n-type at lower concentration, green is the depletion zone and blue is p-type.

Applying a reverse bias voltage between the implants and the back plane increases the depletion zone to the full volume of the silicon sensor with an electric field, see Figure 36b. For the non-irradiated VELO sensors the depletion-voltage required to deplete the full volume is around 150 V.

When charged particles travel through this depletion volume, electron hole-pairs are created where the electrons drift in the electric field, collecting to the n-type implants and a charge impulse travels to the ASICs for readout.

The silicon bulk is not a perfect insulator, so new sensors will have a initial leakage current. For the VELO sensors when new, most leaked about 40 μA at a temperature of -7 °C. As the silicon gets more irradiated, permanent defects are accumulated in the silicon lattice which changes the electrical conductivity of the material; the leakage current consequently changes. Eventually the reverse bias voltage has to be increased to compensate for the extra leakage current. There is a limit of about 1000 V beyond which is not safe for the sensor and high voltage supply.

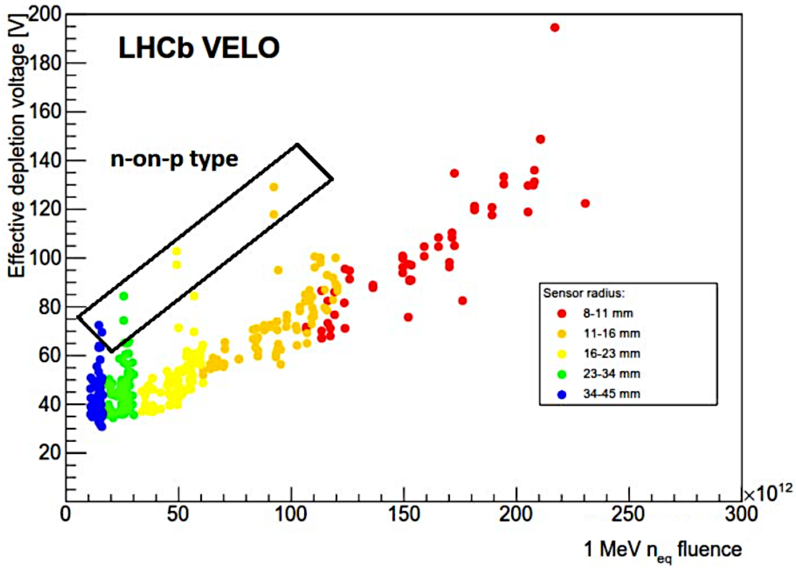


Figure 37: Radiation levels and effects on depletion voltage [8].

A way to buy time on the efficiency of irradiated VELO sensors was to use n^{+}-on-n. The overall effect is that as the silicon bulk gets more irradiated, the n-type bulk becomes p-type, a phenomena known as type-inversion shown in Figure 37. The depletion voltage changes as a result, reducing from its initial value until the silicon changes from n-type to p-type and thereafter the voltage increases with radiation exposure. The figure also shows data points within the diagonal rectangle the n^{+}-on-p sensors which gives a comparison on how successful this strategy was to reduce leakage current at high fluence.

A proxy for the depletion voltage of the VELO sensors that can be consistently measure in-situ is the Effective Depletion Voltage (EDV). By having four sensors at nominal operation and one as the device under test (DUT), track reconstruction can verify the efficiency of the DUT as a function of the DUT's reverse bias voltage. When the most probable value (MPV) of the analogue to digital conversion saturates to the maximum MPV of the voltage scan, it indicates the point where the sensor reaches maximum charge collection and full depletion. The EDV is chosen to be the bias voltage at the point the DUT reaches 80% of its maximum MPV [8].

The leakage current $I(T)$ is a mostly temperature driven process where

$$I(T) \propto T^2 e^{(\frac{-E_g^{\text{eff}}}{2k_B T})} \tag{2}$$

where E_g^{eff} is the effective band gap which accounts for the conductivity of the material, k_b is Boltzmann's constant and T is the temperature of the material.

It is critical to control the temperature of the silicon sensor to avoid a feedback loop of increased leakage current causing more heating which causes more leakage current and so on. When the silicon is irradiated, the higher power consumption from a higher required depletion voltage makes it more sensitive to thermal runaway.

The thermal simulation thus needs to incorporate the expected irradiation on the sensor closest to the LHC interaction point as a function of radius and the expected leakage as a function of irradiation. This sensor would also be at the highest irradiation exposure position along the beam axis; centred around the beam interaction region. The expected fluence has been characterised in units of 1 MeV neutron equivalent (n_{eq}) dose per cm^2 and characterised per fb^{-1} of integrated luminosity [58, 59]. Figure 38 scales to 50 fb^{-1} at the end of the expected lifetime of the upgrade.

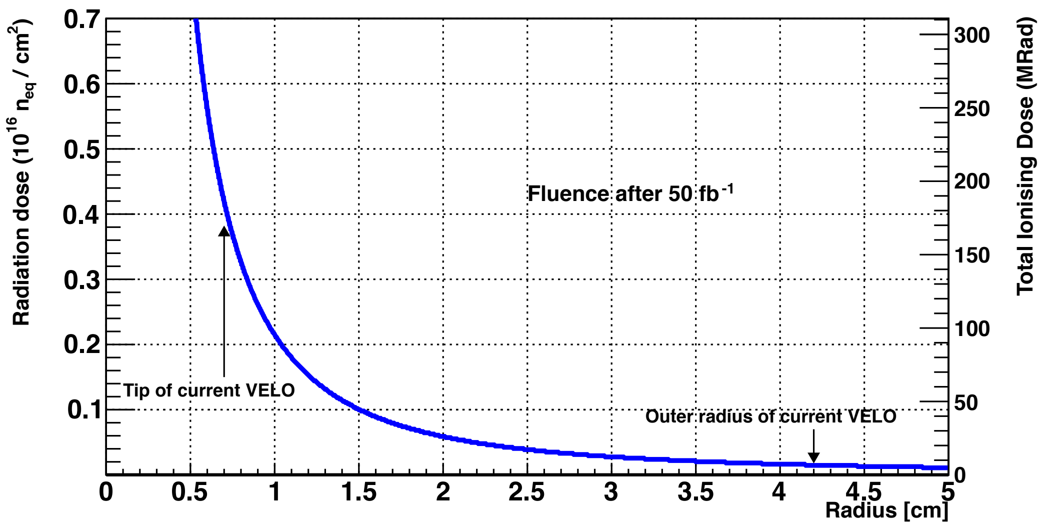


Figure 38: Fluence as a function of radial distance from beam after 50 fb^{-1}, not including saftey factors [9].

Extensive studies have been performed on the effect of leakage current as a function of fluence [10]. Assuming full depletion of the sensor the leakage current is linear with fluence and can be described as:

$$\frac{\Delta I}{V_{ol}} = \alpha\Phi, \tag{3}$$

where α is the current related damage rate, V_{ol} is the volume of the sensor and Φ is the fluence in neutron equivalence. This is known as the Hamburg Model in the detector development field.

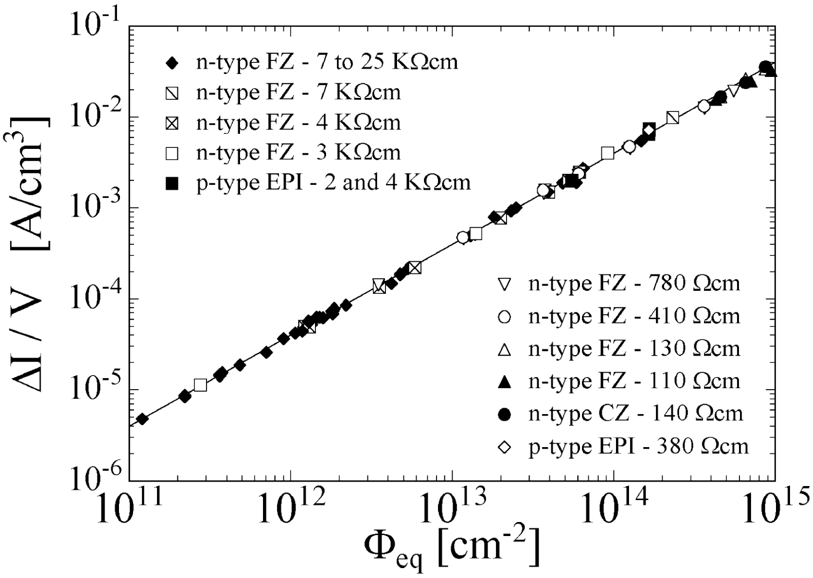


Figure 39: Current per volume to neutron equivalent fluence Φ. A linear fit of $\frac{\Delta I}{V_{ol}} = \alpha\Phi$ to data is superimposed. [10].

In Moll's thesis, The proportionality constant α was fitted to data shown in Figure 39 resulting in a value of 3.99 $\pm 0.03\times10^{-17}$ A/cm at 20°C [10]. Using the relationship between leakage current and temperature described in Equation 2 and the empirical relationship between leakage current and fluence, Equation 3, a thermal model of the VELO strip prototype can be implemented.

3.2.3 Implementation

The levels of fluence as a function of the radial distance from the LHC beam is shown in Figure 38 after 50 fb^{-1}, which means that the leakage current is greatest at the inner radius of the sensor. The sensor in the conceptual model in Figure 33 was split up into average fluence regions, as shown in Figure 40, with step sizes of about 10^{15} n$_{eq}$, the values are shown in Table 1.

The thermal model is a 200 µm sensor irradiated up to 7 $\times10^{15}$ n$_{eq}$ at the inner edge. The extrapolation of Equation 3 beyond 10^{15} n$_{eq}$ relies on the assumption that the sensor is fully depleted and voltage independent. At the time of the simulations there was not direct data to confirm leakage current of 200 µm thick silicon sensors at the expected fluences used in the thermal simulation.

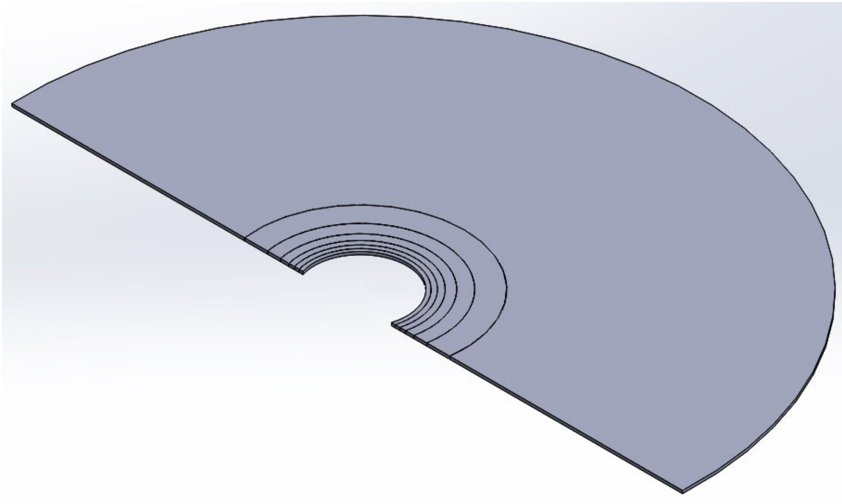


Figure 40: The sensor model has seven volumes of average fluence corresponding to Table 1.

Radii range (mm)	average fluence $10^{15} n_{eq}$
5.1 - 5.5	7.15
5.5 - 6.0	6.14
6.0 - 6.6	5.16
6.6 - 7.5	4.18
7.5 - 8.9	3.15
8.9 - 11.4	2.12
11.4 - 37	0.55

Table 1: Average fluence steps of 10^{16} n_{eq} were chosen as much as possible

There has recently been data from ATLAS's innermost silicon sensors showing a linear relationship between the sensor's leakage current and fluence up to 0.25 x10^{15} n_{eq} [60]. While it does not go to the fluence that the thermal model uses, it does show a good linear behaviour on sensors 200 µm and 230 µm thick [61] with the same species of high energy particles that VELO is exposed to.

If the extrapolation of the Hamburg model does not apply beyond 10^{15} n_{eq}, it is likely to be due to the sensor no longer being fully depleted and reaching the limits of high voltage supply. With less active volume in the silicon, the leakage current is likely to increase at a slower rate against increasing fluence. This will work in favour from the point of view of keeping the sensor temperature under control. The only known phenomenon that could increase leakage current above the Hamburg model is charge multiplication; this is a state of silicon that has high noise and is being avoided in silicon sensor designs irrespective of thermal considerations.

Extrapolating the Hamburg model is likely to be a conservative estimate of leakage current beyond 10^{15} n_{eq}. The model was run at twice and four times the expected leakage current to see how resilient the cooling solution of the sensor can be. The sensor modelled here is the compact sensor discussed in Reference [54] where the inner diameter is at 5.1 mm compared to the strip sensor prototype of 7.5 mm and is irradiated to about twice the fluence. There is therefore some thermal margin that can be gained by having a larger inner radius.

A finite element software package, Ansys [62], was used to calculate the heat generation with $Q = I^2R$, which requires putting a material resistivity in the model as function of temperature.

The data from Equation 3 were from sensors 300 µm thick and measured at 20°C. The prototype sensor is 200 µm thick so the proportionality constant needs to be scaled to keep the equivalent leakage current per volume consistent.

The leakage current at -25°C according to Equation 3, can be rescaled to a given fluence using Equation 2, hence seven different effective resistivity-temperature curves can be calculated for each section shown in Figure 40, via Ohm's law $J = V/d\rho$ where J is the current density per unit area, d is the thickness of the sensor and ρ is the resistivity in ohm metres. The effective band gap from Equation 2 for irradiated silicon is empirically shown to be 1.21 eV [63].

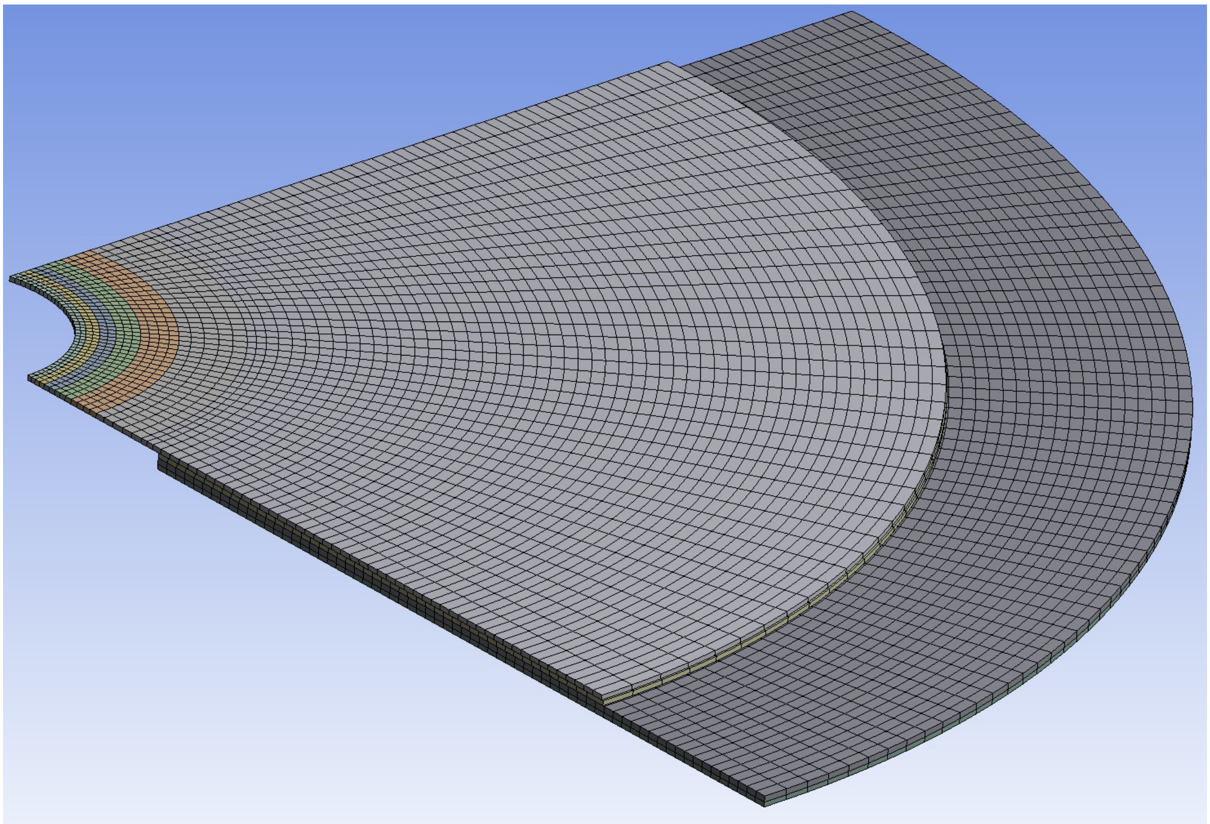


Figure 41: Simplified mesh model of sensor on top placed on substrate. An overhang is increased by cutting away from the substrate below, which minimises material but makes it harder to cool the silicon sensor.

A simplified model was initially made to see how much effect changing the overhang of the sensor from the supporting substrate as seen in Figure 41.

Figure 42 shows how much cooling ability is lost by going from a conservative substrate cut to an extreme overhang. The critical cooling temperature T_c, is defined as being the cooling temperature at the point it is no longer able to prevent thermal runaway. In the minimal model the outermost surface of the substrate was set to the cooling temperature and was changed in one degree Celsius increments until thermal runaway occurred.

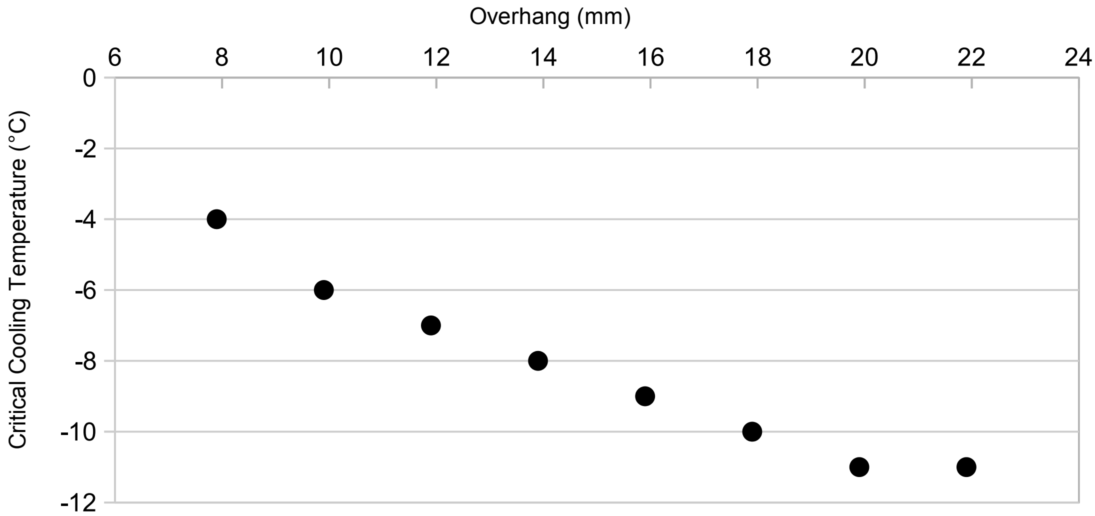


Figure 42: Increasing sensor overhang relative from the supporting substrate shows that lower temperatures are required to cool the inner edge of the silicon sensor to prevent thermal runaway.

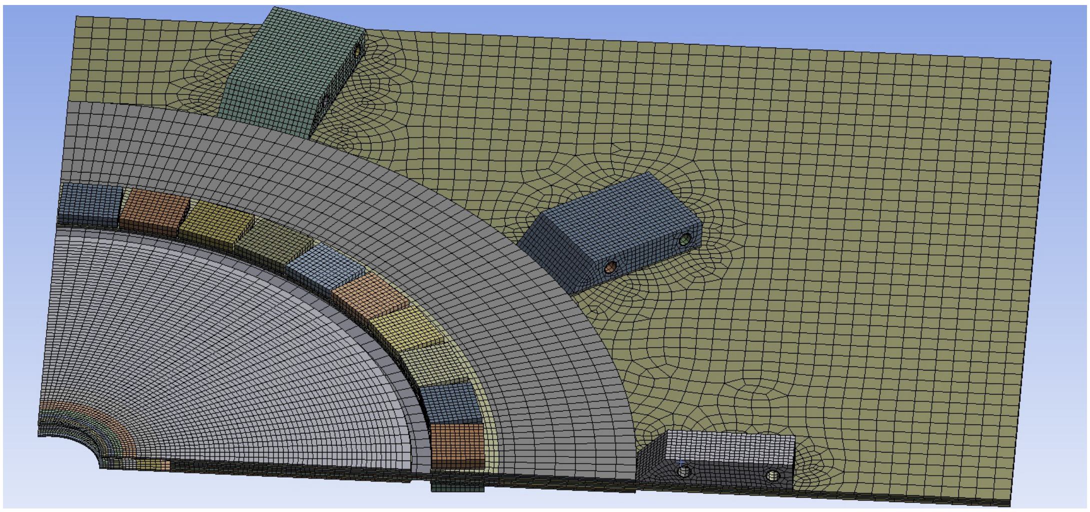


Figure 43: The full mesh model of the thermal simulation. This includes the sensor, glue layers, readout ASIC and cooling blocks where the CO_2 cooling liquid would pass through.

In the full model of the strip sensor thermal model, seen in Figure 43, the cooling temperature is set to the inside of the cooling pipes to simulate the CO_2 cooling liquid.

When the sensor is at a stable temperature state, as seen in Figure 44, most of the heat generated is coming from the ASICs. At the time of development, the likely readout ASIC for the strip sensor to be used was the SALT chip [64]. This is being used in the Upstream Tracker, mentioned in Section 2.12.2, and at the time the expected power consumption per chip was 1.1 W. With a total of about 22 W of heat from the ASICs, the power from cooled silicon is small in comparison, around 1 to 2 Watts.

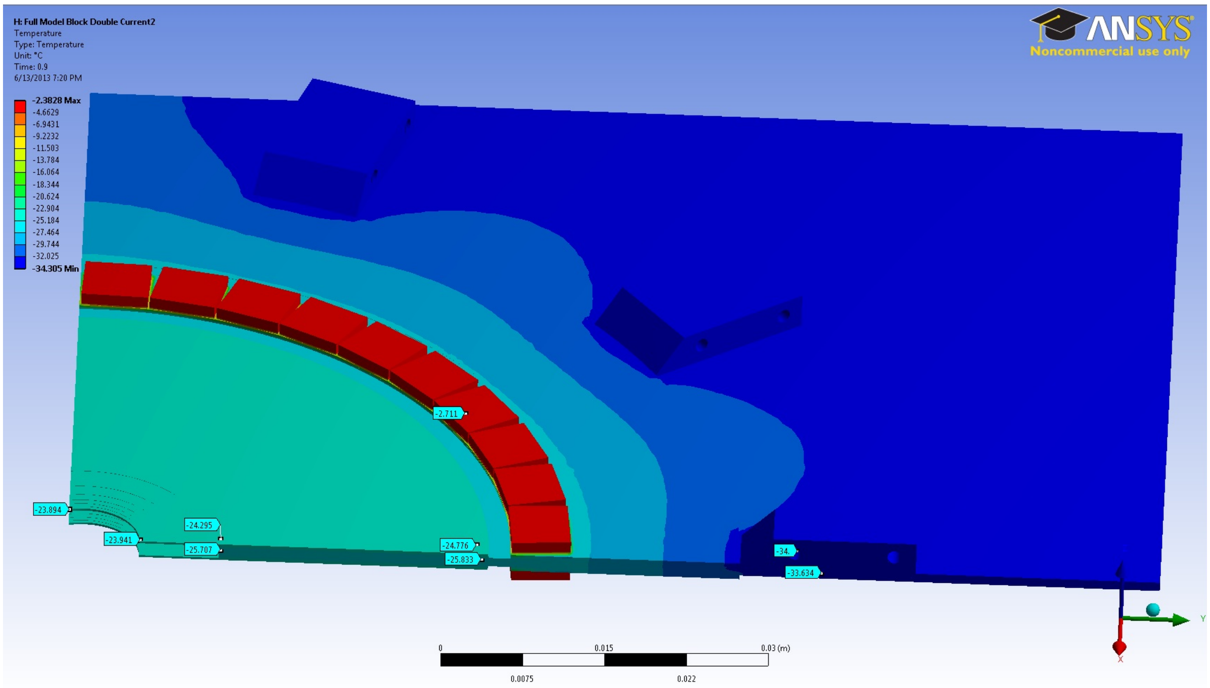


Figure 44: A module running at a steady nominal state. The TPG substrate is able to take away the heat from the ASICs and keep the sensor temperature under control.

When the silicon sensor undergoes thermal runaway, the thermal model attempts to reach a thermal equilibrium with the material and electrical properties given into the model as seen in Figure 45. The resistivity data of the silicon does not change above room temperature and it does not factor in other possible phenomena such as plastic deformation of materials as the temperature approaches a thousand degrees Celsius. Therefore like the simplified model, the model should only be used to find T_c. The cooling temperature was also tested in one degree increments to find T_c.

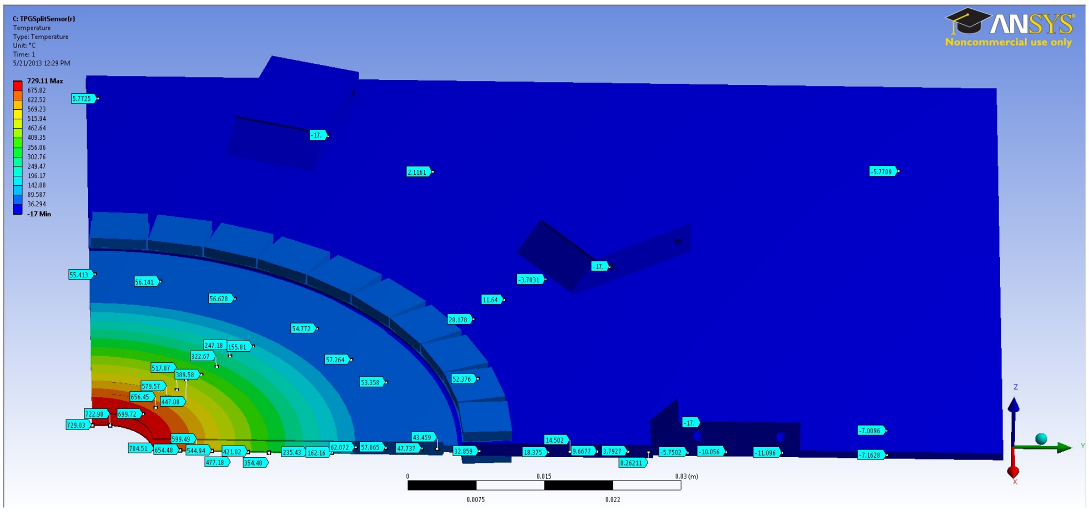


Figure 45: The thermal simulation shows thermal runaway when cooling is not sufficient.

To explore the margin of safety of cooling, the expected leakage current from the Hamburg Model, Equation 3, was scaled for all sections of the sensor. Figure 46 shows how the critical cooling temperature changes and how it is still stable when the cooling temperature is at -24°C.

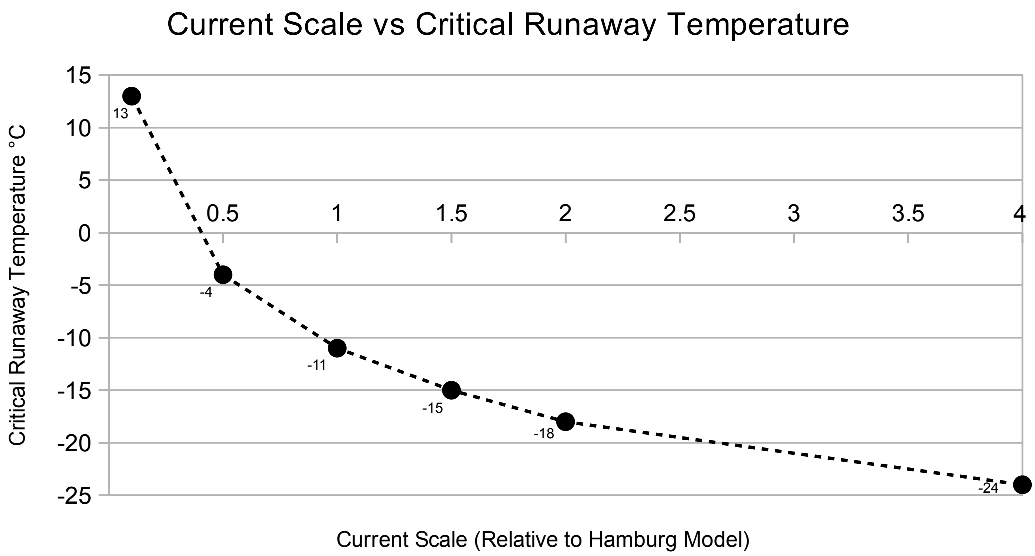


Figure 46: By scaling the expected current from the Hamburg Model in Equation 3 above and below, the effect on the critical runaway temperature is shown here.

3.2.4 Concluding remarks on thermal simulation of the VELO strip sensor

There were a number of engineering solutions considered to apply cooling to an irradiated silicon sensor from micro-channel cooling to cooling blocks. The thermal simulation took the worse case scenario of the least effective cooling solution, TPG cooling blocks on TPG substrate, higher radiation exposure by using the compact sensor with a smaller inner radius. The Hamburg model was also scaled to see if a doubling or quadrupling in expected leakage current still gave a viable cooling state. The ASIC chips were modelled to give 1.1 W of heat each and later developments of the SALT chip are actually expected to run at least half this power rating.

Taking the worst T_c in which thermal runaway occurs, -24°C, this still has a safety margin when compared to the new cooling infrastructure which will provide liquid CO_2 at -30°C.

In the end the VELO pixel detector was chosen which does make cooling more challenging due to the Velopix ASIC layer separating the sensor layer from cooling substrate. However, using -30°C CO_2 and micro-channel cooling should give a comfortable safety margin from thermal runaway.

3.3 VeloPix data lines

The VeloPix ASIC is expected to produce 31 Gbits/s per module [41]. Custom made flexible copper data cables, flex cables, will provide a good solution to carry parallel 5.12 Gbits/s digital signals from the module with minimal signal distortion, material and still accommodate the movement of the VELO halves.

Figure 47 shows that the flex cable will connect from the hybrid on the detector substrate to the vacuum feed-through (VF). The data line will continue to the Opto Power Board (OPB) where data transitions to optic fibre and towards the high level trigger infrastructure.

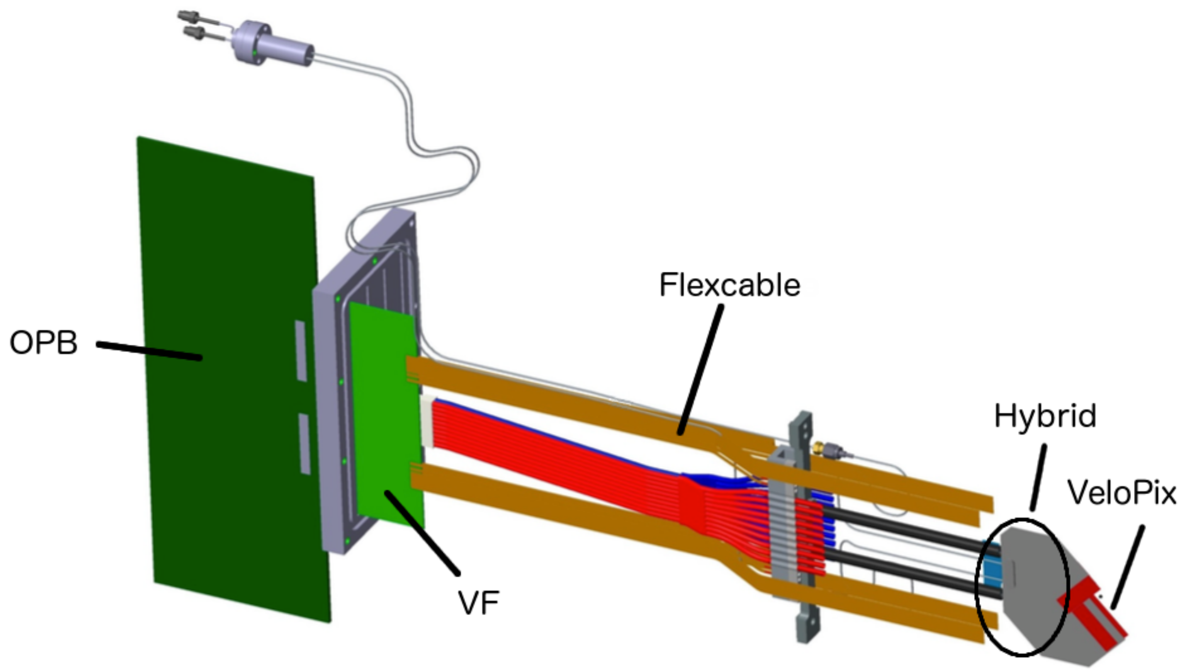


Figure 47: CAD model of the VELO upgrade, the brown indicates the flex cables that will be connected from the Hybrid (not visible in diagram) to the vacuum tank feed-through (VF). The Opto Power Boards (OPBs) will connect to the feed-through with edge connector.

The work in this section involves simulating the electrical properties of the flex cable using Ansoft Q3D Extractor [11] in order to aid the design and prototyping of the physical hardware. There was an iterative process where the computer modelling of impedance and attenuation informed the next steps in prototyping and the prototype measurements then informed the modelling and so on. This section will describe what aspects of the flex cable were modelled and then how the simulation results compare to the prototype measurements.

In each flex cable there is either five or seven pairs of edge coupled striplines, Figure 48 shows a pair in cross-section. Each conductor within a pair of striplines have a copy of the digital data signal except one is the inverse of the other, hence the term differential pair. The difference of the signals can be taken at the OPB and any common mode noise accumulated throughout the differential pair will be ideally subtracted out. The differential pair is also shielded by grounded conductors, shown as black in Figure 48, from external noise and neighbouring differential pairs. The material of the conductor is copper and the dielectric is DuPont™ Pyralux® AP-*PLUS*[3].

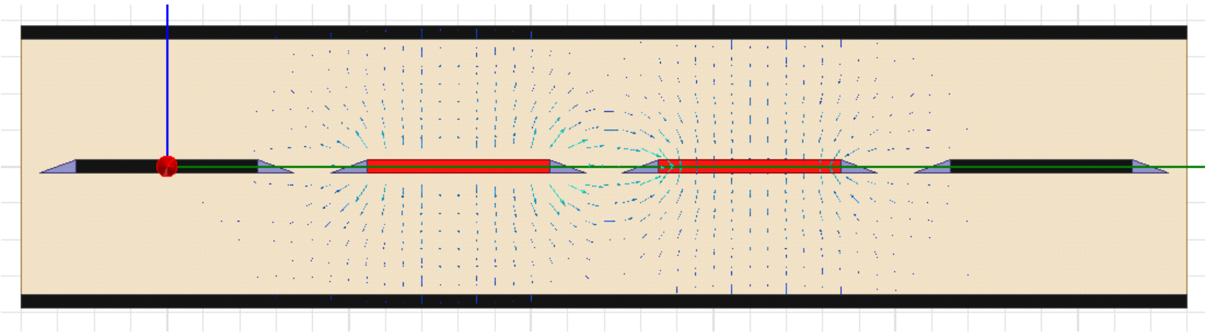


Figure 48: Cross-section of one differential pair, width and spacing of each stripline is 200 µm. Simulated with Ansoft Q3D Extractor [11]. Electric field lines are shown for the differential mode. The signal lines are in red, ground in black, dielectric in cream and glue in purple. The geometry of the glue was updated as shown in Figure 52.

The flex cable will be connected to the kapton hybrid with a SlimStackA™ Molex connector[4]. A similar connector will be installed on the vacuum side of the VF where the data lines continue to a PCI Express connector. This is where the OPB can be slotted into.

[3] https://www.dupont.com/products/pyralux-ap.html

[4] https://www.molex.com/molex/products/family/slimstack_fine_pitch_smt_board_to_board_connectors

3.3.1 Transmission line theory

The differential pair carries an approximation of square waves representing digital signals; these dynamic pulses of electrical energy can be characterised as a Fourier sum of sine waves and therefore sensitive to the inductance and capacitance of the stripline pairs. Using the transmission line model, where there is a unit length of 2 parallel conductors, a characteristic impedance Z_0 per unit length can be derived as

$$Z_0 = \sqrt{\frac{R + jwL}{G + jwC}}, \tag{4}$$

where R is the resistance along the conductor, L is the inductance along the conductor, G is the conductance across the dielectric between the conductors, C is the capacitance across the dielectric, w is frequency of the sine wave through the transmission line and j is the imaginary unit.

High quality transmission lines have low conductor resistance R and low conductance G from good insulation between each stripline. At higher frequencies the imaginary terms tend to dominate, the real terms R and G become negligible and the characteristic impedance becomes frequency independent, $Z_0 \simeq \sqrt{L/C}$; this trend as shown in Figure 49 which is good for maintaining signal integrity.

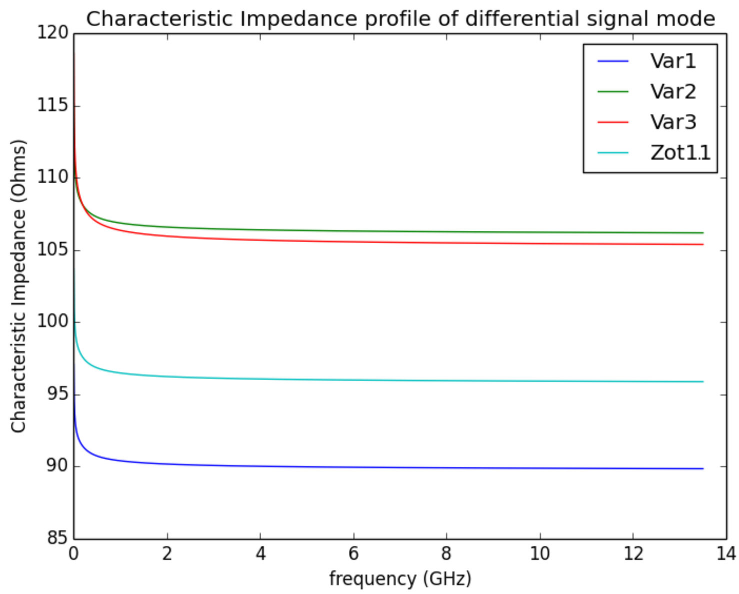


Figure 49: A plot of simulated impedances vs frequency for different geometries as described in Table 3. At lower frequencies the conductor's resistance and dielectric's conductance dominates, but as shown in Equation 4 the inductance and capacitance of the transmission line become more important at higher frequencies.

Figure 47 shows the different components that will carry the VeloPix signals to the OPB and it is important that the assembly has a characteristic impedance that is as consistent as possible to minimise electrical energy being reflected back. 100 Ω is the standard and therefore the geometry of the flex cable needs to be optimised to this.

As the signal goes through the transmission line, not all of the electrical energy is transported from input to output. The transmission is defined as $\mathrm{T} = V_{out}/V_{in} = P_{out}/P_{in}$ where V is the measured input and output voltage and P is the corresponding input and output power of the transmission cable. The transmission loss can be expressed as

$$\mathrm{TL}_{dB} = -20\log_{10}\frac{V_{out}}{V_{in}} = -10\log_{10}\frac{P_{out}}{P_{in}} \tag{5}$$

where TL_{dB} is transmission loss in decibels.

While some of the signal energy is reflected in a transmission line, some of the electrical energy will be lost as heat from the skin effect resistance of the conductor and some from the dielectric losses as transient electric fields rise and fall in the dielectric medium from the signal pulse. It can be shown that skin effect losses have a $\sqrt{f}$ dependency in frequency while dielectric losses have a linear dependency f [65]. The transmission loss can therefore be expressed as

$$\mathrm{TL}_{dB} = a\sqrt{f} + bf \tag{6}$$

where a is the coefficient from conductive losses and b is the coefficient from dielectric losses.

Non-trivial devices that can intake and output electrical signals can be characterised by what is known as scattering parameters [66]. Figure 50 shows a diagram of a 4 port system connected to what could be a simple or complex component. When the system is stimulated at a given frequency the total input current must be equal to total output current to satisfy the port condition so that using a scattering parameter is valid.

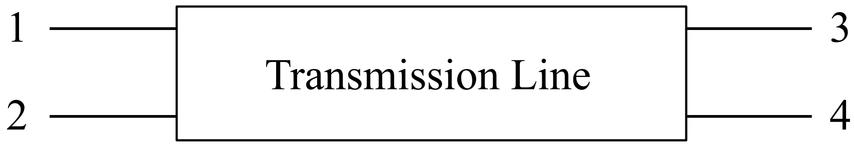


Figure 50: A schematic of a four port system.

The outputs of a device can be related to the inputs via a matrix such that

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \tag{7}$$

where a_{1-4} is a vector of inputs, b_{1-4} is the vector of outputs and S_{ij} are the scattering parameters that make up an S-matrix at a given frequency. Each input and output has an amplitude and phase to describe the signal

To find a particular scattering parameter, one input can be stimulated such that $a_1 \neq 0$ and $a_{2-4} = 0$, meaning that the four input terms summing to an output $b_i = S_{ij}a_j$ can be reduced to one term and a particular scattering parameter

$$S_{ij} = \frac{b_i}{a_j} \tag{8}$$

can be determined.

For example if a differential pair is stimulated at port one as in Figure 50, the signal reflection would be $b_1 = S_{11}a_1$, its transmission would be $b_3 = S_{31}a_1$ and the scattering parameters S_{41} and S_{21} would characterise cross-talk transmission and cross-talk with reflection.

While the scattering parameters can be found from the special case of a single input, they are valid for the general case of all combinations of inputs that satisfy the port condition. Scattering parameters also change with respect to frequency, so a series of S-matrices sampling a range of frequencies are needed to fully characterise a device.

Scattering matrices or S-matrices are powerful tools as multiple electrical components can each be represented by a S-matrix and multiplied together to characterise the behaviour of several components joined together as an electrical signal is passed through the whole system. It is a similar principle to representing optical components as matrices to predict how light travels through a complex optical system.

For the flex cable of the VELO upgrade, differential signals are being used as the desired transmitted signal. Instead of representing the transmission of signal from one end of the physical stripline to the other, Equation 7 can be abstracted so that there are two ports representing differential mode and two representing common mode where the two physical signals are added instead of subtracted. The desired behaviour in this case is to have good transmission from the first differential mode port to the second, minimise cross-talk to common mode ports and have reasonable transmission up to 5 GHz frequency.

A Vector Network Analyser (VNA) can be used to measure the individual components as seen in Figure 47 separately or all together to get a series of S-matrices representation of physical hardware. The Q3D Extractor software can also export a simulated component's S-matrices and therefore provides a way to directly compare simulated to measured signal properties of the flex cable.

3.3.2 Exploring the flex-cable cross-section

There are constraints to the dimensions of the flex-cable cross-section. The thickness of the middle copper layer was initially 18 µm but was later doubled due to a change in the manufacturing process. As the flex cable is connected to a Molex connector that has a 400 µm pitch between each contact, the position between each ground stripline needs to be 1200 µm apart and the differential stripline pairs fit within. The width of the striplines and the spacing between can optimised to achieve an impedance of 100 Ω.

An early flex-cable prototype was cut to see how the real world cross-section manifests from idealised designs. Figure 51 shows that the glue layer is more significant and gives an asymmetry with 12 µm more dielectric material above than below the striplines. The profile of each stripline is trapezoid with about 16 µm narrower on the top surface of copper. This sloped edge is defined as a chamfer with the top corner 8 µm horizontally separated from the bottom corner as show in Figure 52.

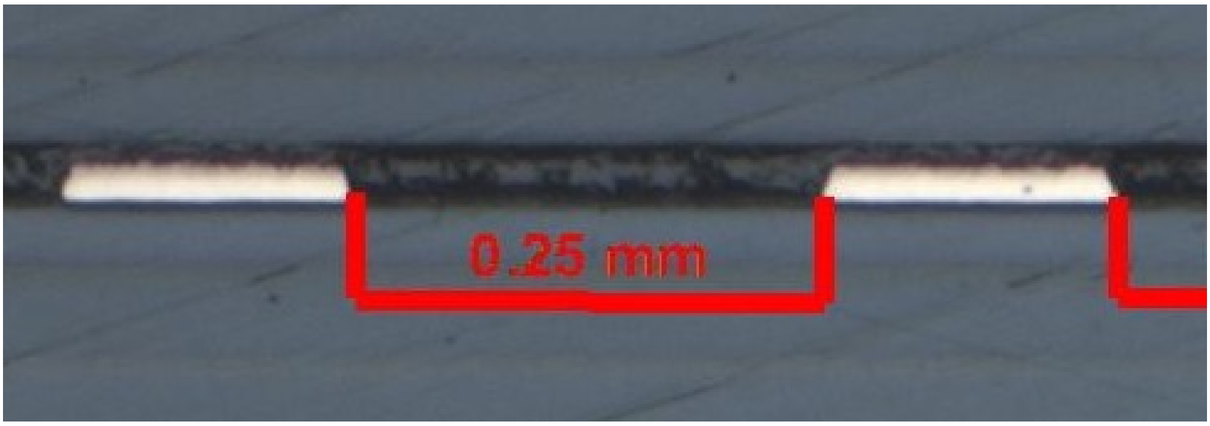


Figure 51: A flex-cable prototype cut in cross-section. Copper striplines are white, dielectric is light blue and glue is dark blue. The glue layer gives approximately 12 µm extra thickness to the dielectric layer. The striplines have a trapezoid profile 16 µm narrower on the top edge.

The model was incorporated to include these real world geometries and Figure 52 labels the physical parameters that are discussed in this section.

A parameter scan of the width of the striplines shows the variation in impedance and attenuation in Figure 53. The midpoints of the striplines were kept at 400 µm apart, Figure 48 shows such an example, therefore wider striplines had have a higher skin area but reduced pair spacing which likely would lead to stronger electric fields and hence greater dielectric loss.

To find what geometric or electrical properties were most sensitive to the performance of the differential pair, a simple model was compared to different plausible properties and geometries. The simplest model of the flex cable had 200 µm stripline width for both ground and signal which meant all the separations are 200 µm. There was no extra glue thickness and no chamfer.

Table 2 shows the most significant changes in impedance. Changing the dielectric constant by 10% changed the impedance the most and shows what scale of change of impedance can occur with a plausible tolerance of the material property. The build tolerance in thickness of the dielectric Pyralux was ± 5% and retrospectively using the

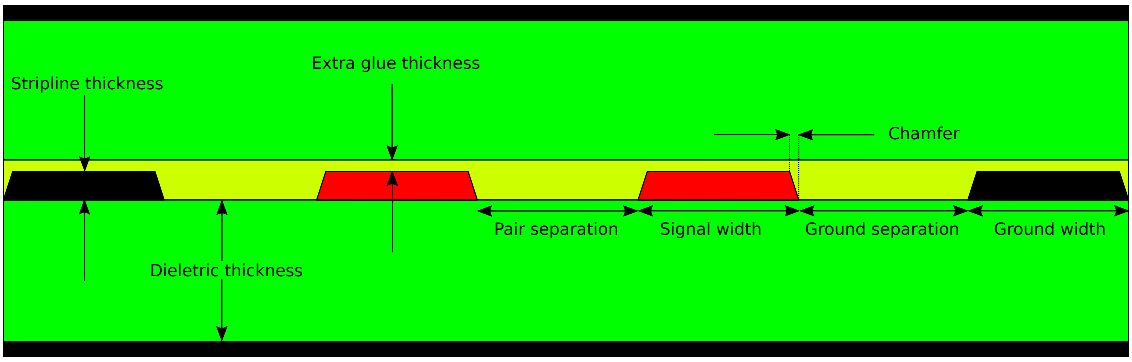


Figure 52: The physical parameters of the flex-cable model. The mid point of the ground striplines are 1200 μm apart.

EEWeb on-line calculator for edge coupled striplines [12] showed a similar amount of change in impedance.

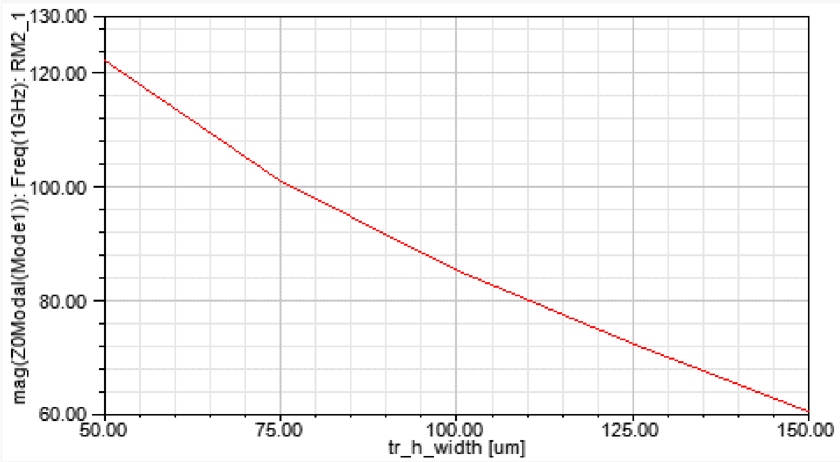
Permutation	Impedance (Ω)	Change (Ω)
Version 1 (Q3D Extractor)	87.1	0
Extra glue thickness 12 μm	88.4	1.3
8 μm chamfer	89.1	2.0
Pyralux dielectric constant -10%	91.8	4.7
Pyralux dielectric constant +10%	83.1	-4.0
Version 1 (EEWeb [12])	99.1	0
Pyralux dielectric thickness -5%	94.6	-4.5
Pyralux dielectric thickness +5%	103	3.9

Table 2: Permutations of Variant 1, listed are the properties that affect the characteristic impedance the most significantly. Using the EEWeb calculator [12], the change from thickness of the dielectric was calculated.

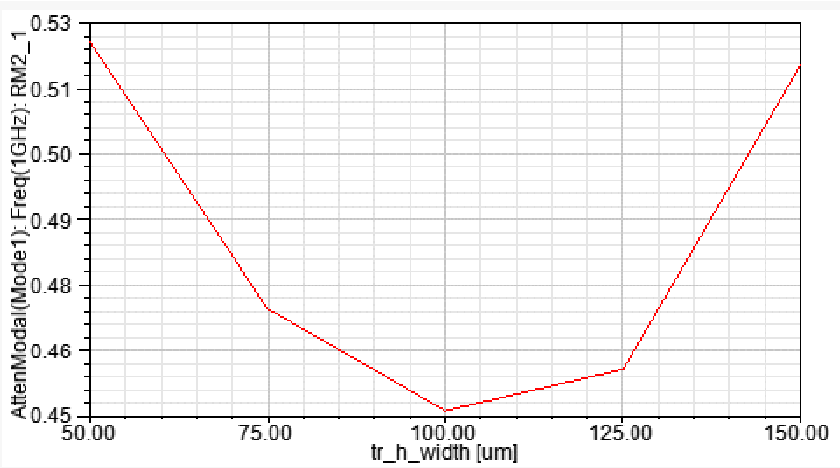
Other parameter changes that gave less than 0.1 Ω were the surface roughness (both the Hammerstad-Jensen and Huray model), the loss tangent and smaller mesh sizes. The roughness models try to emulate the real world surfaces of conductors, The Hammerstad-Jensen model [67] assumes a corrugated surface to factor in a longer path for electrons to travel and the Huray model [68] imitates surfaces found in scanning electron microscope studies of copper by using stacked pyramids of spherical shapes like a stack of snowballs. The surface roughness however did not however significantly affect the performance of the differential pair.

Q3D Extractor has the advantage over the on-line calculator in that the EEWeb calculator only models the differential pair and not the ground traces. By increasing the ground separation in the Q3D Extractor simulation, the capacitance of the differential pair should decrease and the characteristic impedance should increase following Equation 4. Figure 54 shows that this is a small effect that happens in the correct direction.

In the real flex cable, this would be making the ground trace narrower which does have a limit, but it does show how much the impedance is expected to change per micron



(a) Differential Mode impedance (Ω)



(b) Differential Mode attenuation (Np/m)

Figure 53: Scan over the width of the striplines against impedance and attenuation at 1 GHz where each stripline is placed at 400 µm pitch. The "`tr_h_width`" parameter shown on the x axis is half the actual stripline width, hence the optimal attenuation occurs at a width of 200 µm in Figure 53b.

(about 0.017 Ω) and therefore if there is a tolerance of 5% or 10% in the build of the separation or width of traces of 200 µm there may be a change of 0.17 Ω or 0.33 Ω. This is a negligible effect.

Within the finite element simulations for Figure 54, everything else is kept the same except for the ground separation, and so in theory there should be quite a smooth change in impedance as capacitance and distance is linear. What is different is each separation point has to be re-meshed. While Q3D Extractor does automatically optimise for mesh detail so that it can converge on a solution, the standard error of the impedance (0.15 Ω) from a linear fit give an indication of how good this mesh optimisation even if it is not perfect. Figure 55 shows how the mesh is automatically adjusted to higher density electric fields. In a real world build of the flex cable, an impedance within 5 Ω of the target is good enough.

Ultimately testing real life prototypes will give an indication of the consistency of

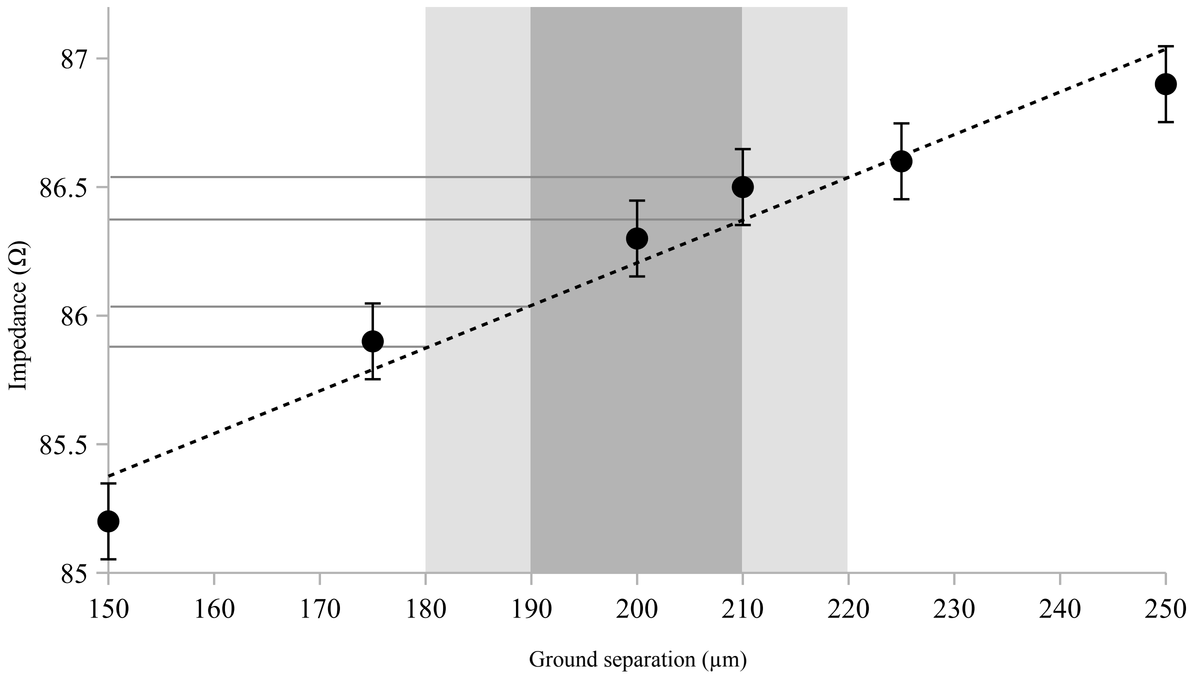


Figure 54: Using Q3D Extractor on Version 1.4, impedance vs ground separation defined in Figure 52. The standard error on impedance using a linear fit is 0.15 Ω which is likely due to the re-meshing of the finite element model on each data point. Grey areas are ± 5% and 10% of 200 µm ground separation which shows a negligible effect on impedance.

the build. Factoring in all the possible differences from Table 2 can give a plausible explanation for the difference in measured impedance and simulated as shown in the next section.

The digital signal that is transmitted in flex cables has a base frequency of 2.56 GHz, so it is important to ensure the signal losses are acceptable the region of the fundamental and the first harmonic at 5.12 GHz.

Figure 56 shows frequency scans of the transmission properties of the differential pair models, the chamfered asymmetric model shows reduced transmission performance. It is plausible that the more acute corners would create denser current flow and electric fields and hence greater losses. The models with poorer and better dielectric properties are also consistent with less losses with lower dielectric constant and vice versa.

By analysing the relationship between frequency and transmission, an insight can be found into whether the signal losses are dominated by skin effect on the copper conductor

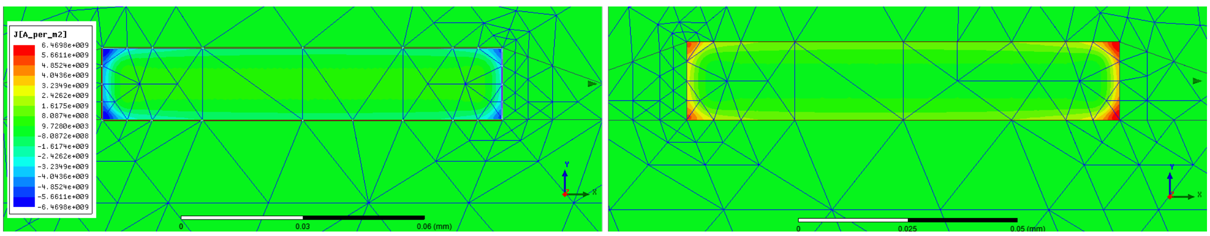


Figure 55: Q3D Extractor simulation of current density of rectangular differential lines. The software automatically adjusts the mesh density as needed.

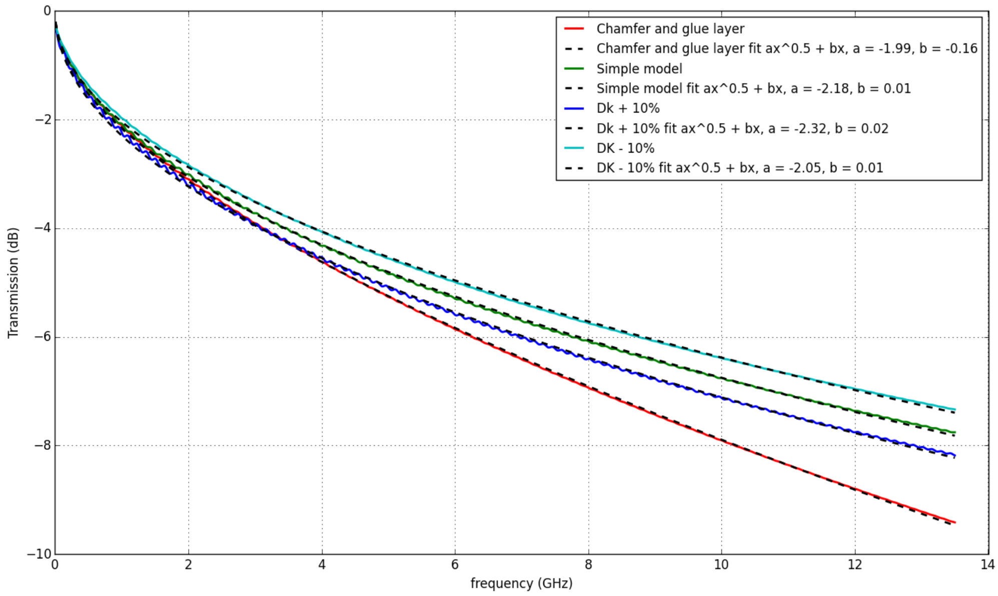


Figure 56: Permutations of Variant 1, this plot shows how the different perturbations affect transmission vs frequency. The more realistic 8 μm chamfer of the striplines and extra dielectric thickness is more lossy at higher frequencies. From Equation 6, the conductive losses appear to dominate.

or dielectric losses in the Pyralux medium. Using Equation 6 as an empirical model to fit to the transmission simulations in Figure 56 the conductive loss coefficient $a \approx -2$ is two orders of magnitude higher than dielectric loss coefficient $b \approx -0.01$; the exception is the simulation of the chamfered model where $b \approx 0.16$ is a magnitude higher. A higher b coefficient would make sense for the chamfered model because of a relatively higher electric field flux in the dielectric medium. There is no error assigned to each frequency point so the unweighted fit has an approximate error of 0.002 for a and 0.0005 for b.

The conclusion from the finite element model is that the conductive losses is the dominating factor for the overall transmission loss of the transmission cable. Therefore the prediction would be that wider traces, and thus larger surface area, will give a less lossy transmission line albeit with a lesser effect due to the concentration of current at the stripline edges as seen in Figure 55.

Since the cross-section pictures shown in Figure 51 reveal the real world geometry of the differential lines, the finite element simulation went forward using the 8 μm chamfer and 8 μm extra glue thickness to compare to measured prototype versions in the next section.

3.3.3 Comparison to prototype measurements

From preliminary simulation results in Q3D Extractor, three variants of flex-cable prototypes were designed, fabricated and tested. A further set of prototypes and final production versions were fabricated by a contractor[5]. The dimensions of the different design variants are described in Table 3.

Version	stripline width	pair separation	ground separation	stripline thickness	length
1	200	200	200	18	56
2	150	250	225	18	56
3	100	100	350	18	56
Zot 1.1	180	220	210	18	56
Zot 1.2	200	200	200	36	56
Zot 1.3	180	220	210	36	57.5
Zot 1.4	200	200	200	36	57.5
Final	200	200	200	36	55, 56.1, 57.5

Table 3: Dimensions of the flex-cable designs in µm but flex-cable length in cm; the final production version has different lengths to fit the geometry along the full length of the VELO sub detector. Figure 52 diagrammatically shows each dimension description.

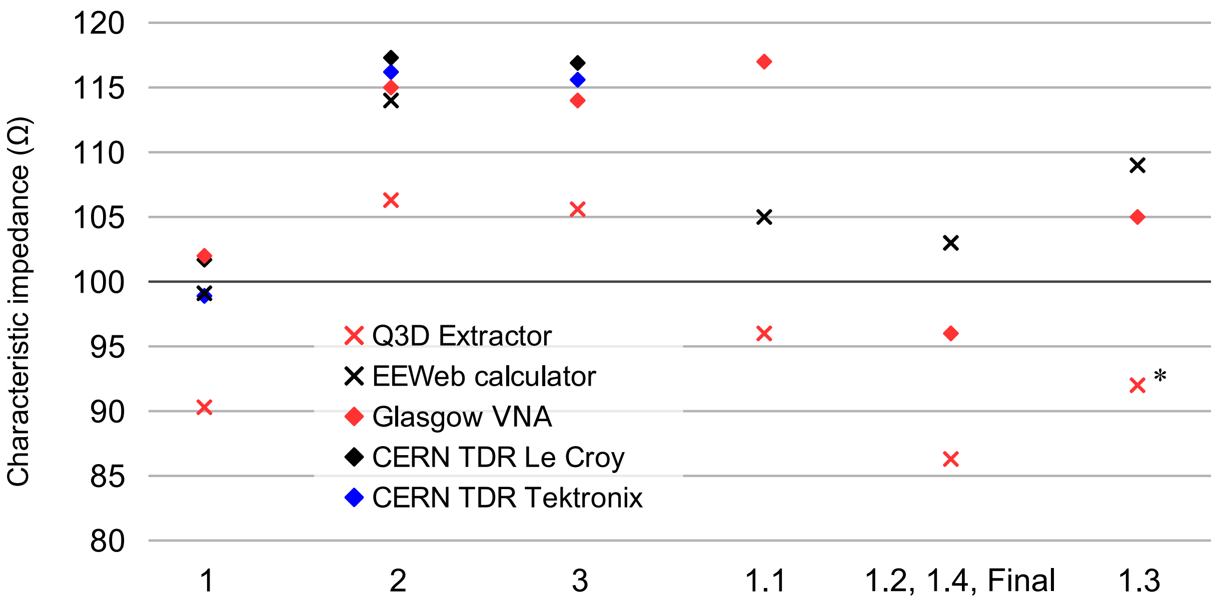


Figure 57: Comparison of measured and simulated values of characteristic impedance for different flex-cable versions. The EEWeb calculator [12] had a value of 170 Ω for Version 3. * The Q3D Extractor data point for version 1.3 was simulated by Kenneth Wraight [13].

The hardware measurements of the flex cables at the University of Glasgow physics department were performed by Cameron Dean [69] and Sneaha Naik [70]. This includes the Glasgow data in Figure 57 of characteristic impedance and the transmission vs frequency plots in the rest of the section, Figure 58 shows an experimental set-up at Glasgow of

[5] https://www.zot.co.uk/

the Vector Network Analyser (VNA) which is used for measuring scattering parameters. The CERN data in Figure 57 was measured by Lars Eklund and Jan Buytaert using a Time Domain Reflectometer (TDR).

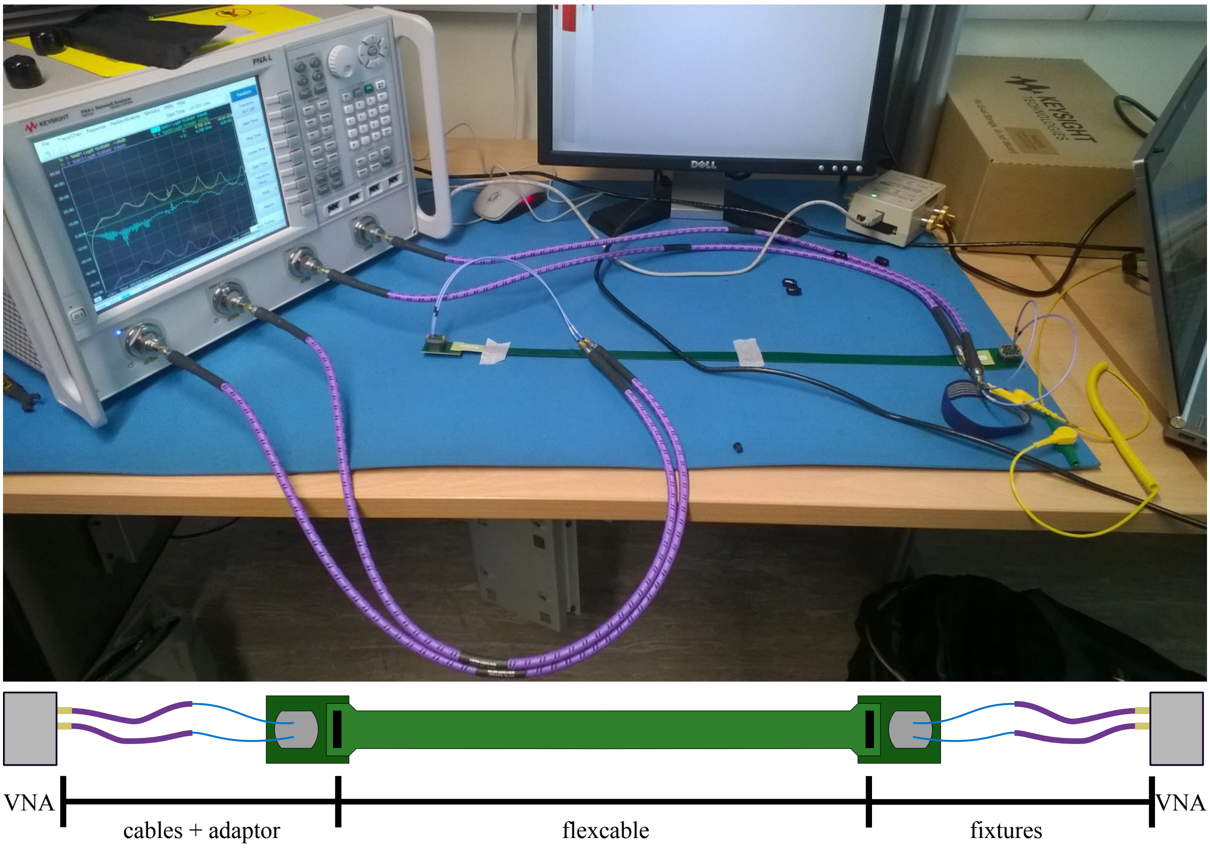


Figure 58: A picture and diagram of the flex-cable measurement. Time domain reflectometry can be performed to get a characteristic impedance measurement using the Vector Network Analyser (VNA). Calibration has to be done to remove the effects of the hardware connecting the flex cable to the VNA in order to only measure the properties of the flex cable itself.

A TDR measurement sends a step function with a very fast rise time and measures the timing of any reflections that come back. This allows a reading of the characteristic impedance along each part of the transmission line and also over any chain of devices connected together. The principle is analogous to radar.

The VNA measures scattering parameters by sending a sinusoidal signal into one port and recording the output of all four ports, the ratio of each output and input signal gives the corresponding scattering parameter as shown in Equation 8. This input signal stimulation is repeated on each port to get all sixteen scattering parameters as described in Equation 7. The process is repeated for all frequencies and the impedance can be mathematically derived.

The characteristic impedance measurements of the prototypes from different equipment is consistent with each other, shown in Figure 57, but the Ansys model always underestimates the impedance of around 10 Ω with the exception of version 1.1. Section 3.3.2 has discussed what could account for this discrepancy.

The EEWeb edge coupled stripline impedance calculator [12] is a simplified model which uses only the differential pair with rectangular cross section, two ground planes above and below and symmetrically separated by a dielectric medium. The approximate model is fairly close to measured values, but generates a value of 170 Ω for version 3, since its parameter space for calculating realistic values is limited and shows that it is valuable to corroborate impedance using alternative methods.

As discussed earlier in Section 3.3, scattering parameters can be used to compare transmission properties from simulation and measurement. Measuring the flex cable with the Vector Network Analyser as shown in Figure 58 requires calibrating and characterising the cables and adaptor, otherwise known as fixtures, that connects to the flex cable. Each part of the fixtures can be measured separately or together to get scattering matrix representations. The VNA can measure the adaptors separately in an open ended configuration or characterise the whole system in a closed circuit configuration.

Without compensating for the effects of the fixtures, the measurement as shown in Figure 59 is more lossy than the Ansys simulation as expected. Only versions 1,2,3 and Zot 1.1 are shown in the results as these had the most extensive measurements made.

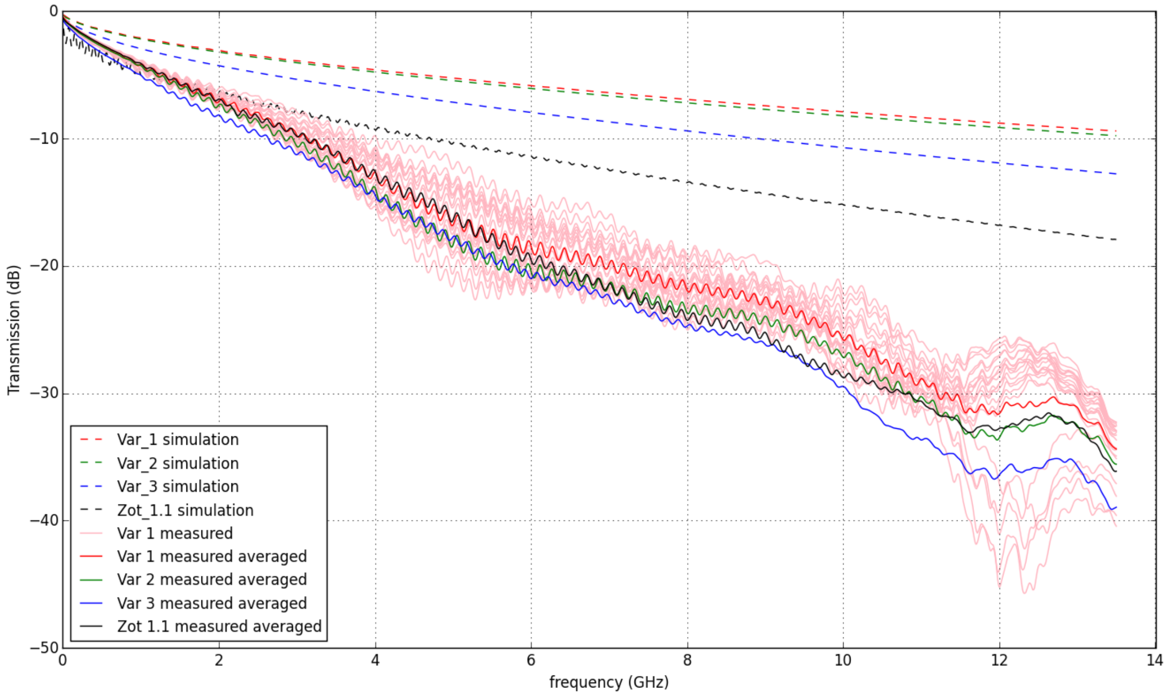


Figure 59: Transmission vs frequency of prototype variants. All measurements of version 1 are plotted to show the variance, and the noise at high frequency, with the averages of version 1,2,3 and Zot 1.1.

There were four copies of each flex-cable prototype version with seven differential pairs each. One copy of the version 1 flex cable was not available during measurement, therefore there were 21 differential pairs measured for version 1 and 28 pairs for versions 2 and 3. The variance in losses per differential channel increases to higher frequencies until signal to noise ratios reduce to the point that measurements are very sensitive to how well the measurement equipment is connected.

An observation can be made that the design variants with narrower trace widths tend to be more lossy which would not contradict a transmission line that has losses mostly from skin-depth conductive loss. However each flex-cable version has other different variables, so this particular comparison is not ideal.

There are oscillations seen in the measured data that correspond to a frequency difference of about 150 MHz. The signal will be partially reflected at the ends of the flex cable and different frequencies will constructively and destructively interfere, creating resonances at regular points in the frequency spectrum. This phenomena is dependent on the length of the flex cable and the velocity of the signal through the differential pair; Figure 60 shows data from flex cables of shorter lengths and have correspondingly larger spaces between resonances along the frequency axis.

The base signal frequency is at 2.5 GHz, any frequency above 10 GHz has so much loss that it will have negligible contribution to the output signal. Therefore subsequent plots will show up to 10 GHz in the frequency range.

The finite element simulation does not include the Molex connector which is still left in the measurement and only simulates the different cross-sections with a length of 56 cm. To make a proper comparison between simulation and measurement, the effects of the connector have to be removed.

To characterise the connector, a “stub” is used. This is a flex cable that was as short as possible with a transmission line around 2 cm long but with room for two connectors on each end. Stubs were created only with cross-section variants that match the first three prototype designs.

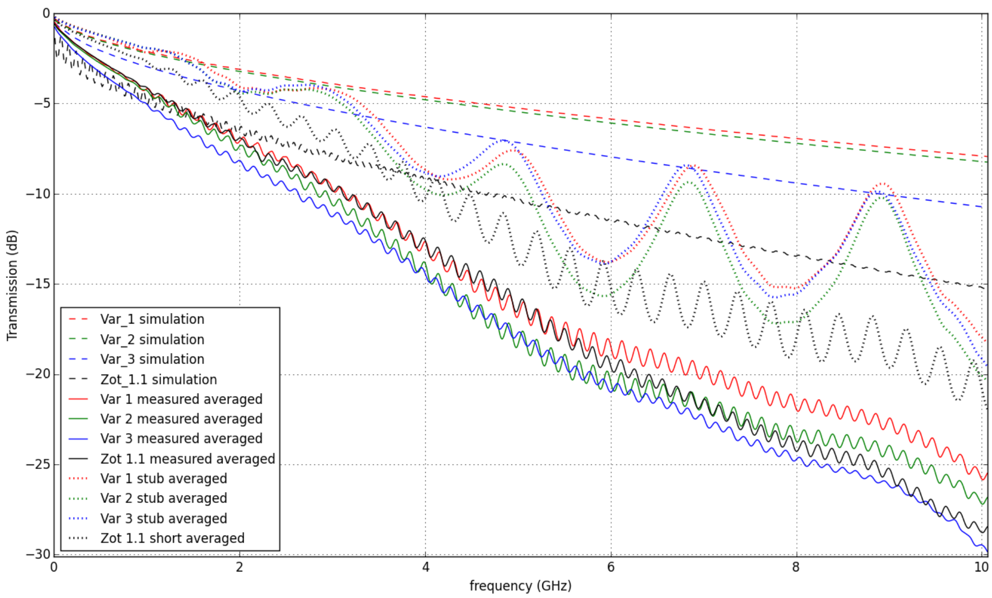


Figure 60: Transmission vs frequency of prototype variants of both long and stub versions compared to simulated transmission. The Zot 1.1 short is 20 cm long and has an intermediate loss between the 56 cm long length and the 2 cm stub length.

Figure 60 shows the transmission curves of the long and stub flex-cables which shows how the connectors do attenuate the signals. Because the stub is mostly the connector and very little length of the cross-sections of the flex-cable variants, the averaged plots of the stub transmission losses over versions 1,2 and 3 overlap closely. There is a significant 2 GHz frequency spacing in resonance pattern due to the short length of the transmission line of the stub. The Zot 1.1 short tape with a length around 20 cm has a frequency spacing around 350 MHz. Even though there was not a stub version of the Zot 1.1, the 20 cm long flex cable does demonstrate an intermediate example of both losses and frequency resonances.

The difference in losses between the long and stub versions should account for the main length of the flex cable as the effects of the connectors are removed. Figure 61 shows that while the 2 GHz resonance makes it difficult to compare accurately, the simulation is approximately consistent and it is easy to see that the difference between version 1 and 3 is consistent for both measurement and simulation.

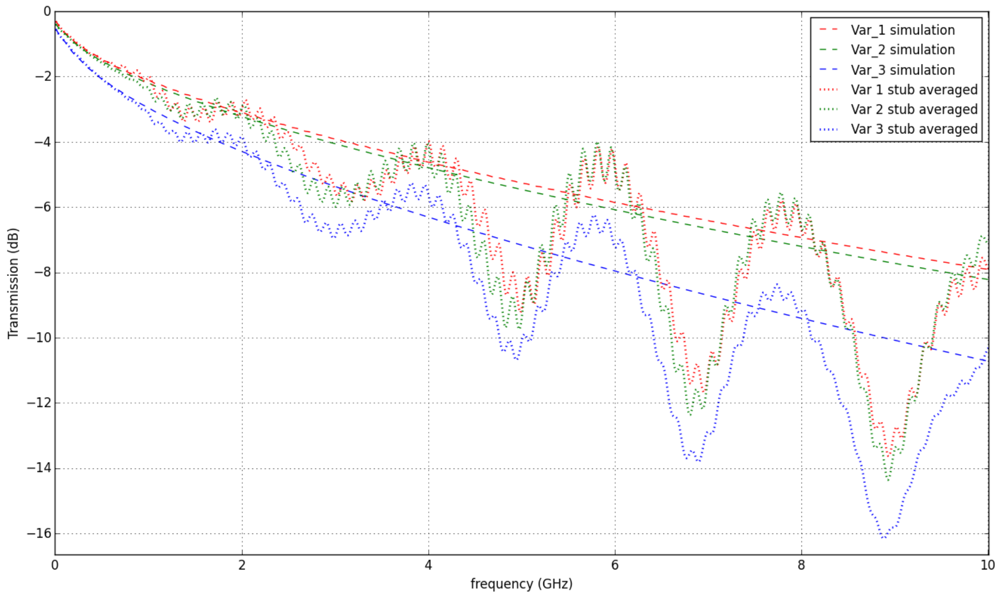


Figure 61: Comparison of simulated transmission to difference in measurement between full length flex cable and stub in order to observe losses without the effects of the Molex connectors.

The second way to subtract the effects of the Molex connector uses the short calibration using the stub to complete the circuit. The VNA is able to characterise the fixtures, the Molex connectors and the very short 2 cm of transmission line as a baseline. When the full length flex cable is measured off this baseline, what is left over should be the effects of the main length of the flex cable. There are various propriety algorithms and techniques used in this measurement that can remove the resonance effects seen in previous plots. Figure 62 shows smoother plots at lower frequencies, but gets noisier at a faster rate towards high frequencies.

The measurement technique using short calibration seems to be more sensitive to bad connections and noise as about half the fixture subtracted data was more lossy than the unmodified measurement of transmission. A simple algorithm was used to remove bad data; each unmodified measurement of each differential pair is matched to the short calibration that uses the stub's corresponding differential pair, if at any frequency below

10 GHz that the transmission loss of the fixture removed measurement go below the unmodified measurement, that is considered by definition “impossible” and is removed.

The data that is remaining is still highly variable, more variable than seen in Figure 59, and when the losses are greater at higher frequencies, the signal to noise ratio rapidly diminishes. Flex cable version 3 had the cleanest data and it shows a better alignment to its simulation at low frequencies.

In summary, two methods were used to compare the simulated transmission losses to measured losses and within the variability across 21 to 28 channels, or half that for the short calibration, of version 1,2 and 3, there is agreement. Version 1 appears to have the best transmission performance.

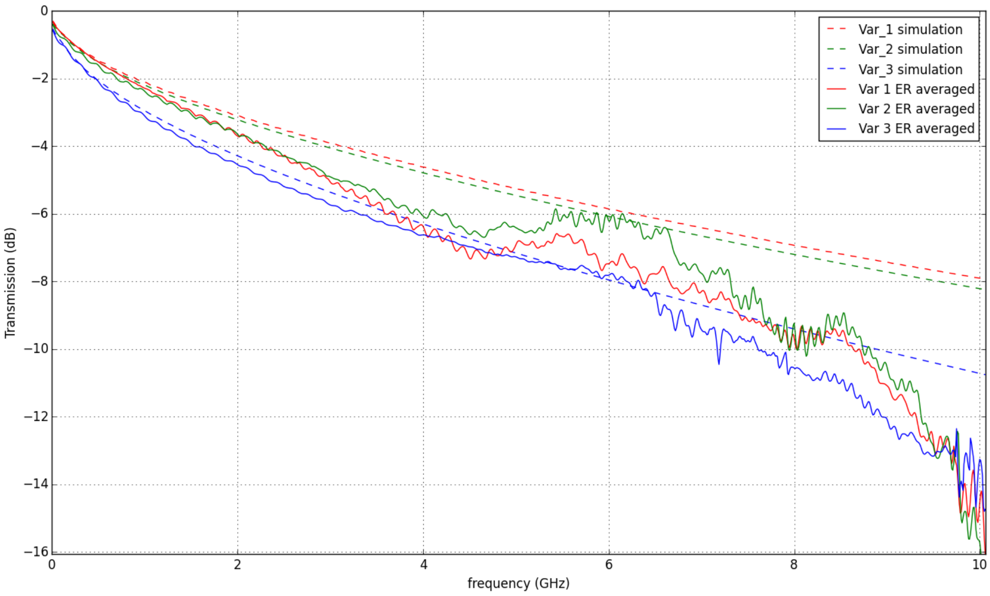


Figure 62: Comparison of simulated transmissions to “Everything Removed” ER averaged which is the measurement using fixture removal techniques that the VNA system offers.

3.3.4 Concluding remarks on flex cables

Using Q3D Extractor, a finite element software, simulations of the cross-section of the flex cable were performed to explore the parameter space and find an optimal design to match 100 Ω characteristic impedance. By using pictures of a cut flex cable, a more realistic cross-section could be constructed.

The transmission curve against frequency indicated that conductive losses were more dominant than dielectric losses. Design variants with narrower traces, and hence less surface area, both in simulations and measurements tended to have more lossy performance.

The impedance from simulations tended to underestimate the impedance by around 10 Ω with respect to physical prototype measurements but plausible effects were investigated which could account for the difference. These effects include possible variability in stripline thickness, slight changes in dielectric constant, changes in dielectric thickness due to glue layers and the shape of the stripline deviating from a rectangular cross-section due to construction methods. Despite the bias, they did help guide iterations of the prototypes to the final design.

When measuring the transmission loss vs frequency, there were 21 to 28 available differential pairs of version 1,2 and 3 which showed an increasing variability as a function of frequency. Using stubs and full length flex cables, it was possible to remove the effects of the Molex connectors and fixtures with two different measurement techniques based on scattering parameters to get a fair comparison to the simulation model. The transmission vs frequency comparison of measurement and simulation did show agreement with both techniques considering the variability of the measurement.

Version 1 was measured to have the most promising characteristics of characteristic impedance and transmission loss and the final design was based on this cross-section.

The VELO module moves about 3 cm from standby position during LHC beam injection to the data acquisition position which will leave a 3.5 mm aperture [41]. A flex-cable prototype has been subjected to 3000 bends without significant loss to performance [69].

To verify the radiation hardness of the flex cable, test flex cables of the same material were manufactured in parallel to the prototypes and subjected to over the expected irradiation that the flex cable would experience after 50 fb^{-1}. It was found that there was a negligible change to the performance [70].

The design of the flex cable has been finalised as described in 3, its connector designs optimised [70] and it has been assembled into the full electronics system and its performance verified at CERN's test beam facility shown in Figure 63. As of writing, the VELO upgrade is being assembled in full so that is ready to be installed on-site at the LHCb machine.

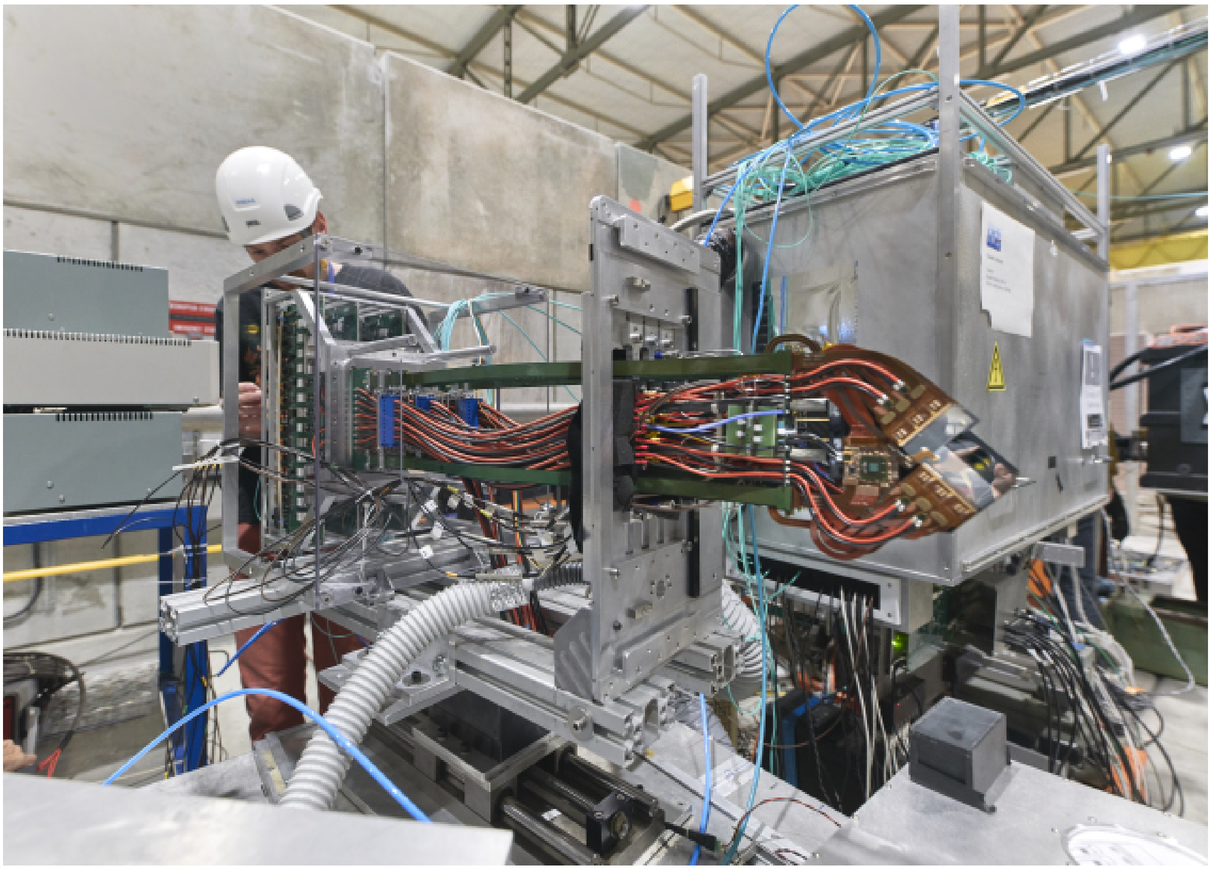


Figure 63: One of the VELO upgrade modules being assembled at the test beam. The green flex cable can be seen connecting the sensor output to the Opto Power Board at the rear. [14]

4 Conclusion

The research covered in this document is related to the upgrade of the VELO sub-detector of the LHCb on the LHC ring at CERN. This includes constructing a prototype module of a strip sensor and simulating the thermal behavior of a complete strip sensor module to verify the feasability of the cooling design. Simulations of transmission line cross-sections were used to aid the development of flex cables as part of the hardware that will transport data signals out of the new VELO readout system.

A prototype strip sensor for the LHCb upgrade was assembled with a readout board and basic tests were done with a ^{90}Sr source. It was shown that the prototype functions as expected and identified some of the challenging parts of the assembly, particularly in having a reliable pitch adaptor.

Tools were designed and made to pick up the fragile 200 μm sensor and place it accurately on a substrate where it was wire bonded to a preexisting readout board using Beetle ASICs; the readout chips used in the original VELO detector. The sensor behaved as expected during reverse bias tests and a pulse shape distribution and Landau spectrum was reproduced using the ^{90}Sr source.

Due to the higher radiation exposure expected in the LHC luminosity upgrade, simulations were done to understand the thermo-electric behavior of irradiated silicon upgrade conditions for the strip sensor. This involved making 3D CAD designs and using a model linking heat generation to leakage current and irradiation levels of silicon. The cooling solutions included microchannel cooling where liquid CO_2 evaporates within a silicon substrate and using cooling blocks on a Thermal Pyrolytic Graphite substrate - not dissimilar to the original VELO design.

Using the less powerful cooling solution, cooling blocks, it was found that CO_2 temperatures above -18°C will lead to thermal runaway. This indicates that -40°C CO_2 is a safe and feasible option for strip sensors. The final pixel based sensor will use a microchannel cooling system.

The VELO upgrade will output around 2.85 Tbits/s of data and specialised data cables are required. These flex cables can accommodate 3 cm of movement from the VELO halves and will have sufficiently low attenuation and bulk.

The data signals are transported on multiple parallel transmission lines, each capable of transmitting 5.12 Gbits/s at 2.56 GHz base frequency. Finite element electromagnetic simulations using Q3D Extractor were performed to assess what was the optimal cross sectional design to get the best characteristic impedance and attenuation performance.

The simulations seemed to have a systematic bias in characteristic impedance compared to measured values from physical prototypes but it did help guide the later iterations of flex cables. The frequency dependency of the attenuation also indicated that conductive skin depth losses is the dominant effect compared to dielectric losses.

The simulation only modelled a cross-section with a length of 56 cm, therefore the effects of the Molex connectors of the real-world flex cables had to be subtracted from measurement to compare to simulation in a fair way. Two different methods were used and showed that transmission in simulation and measurement were in agreement.

The flex-cable design is now finalised, fully qualified, produced and the new VELO upgrade is being installed in the LHCb.

Bibliography

[1] The LHCb Collaboration *et al.* The LHCb Detector at the LHC. *Journal of Instrumentation* **3**, S08005 (2008). URL http://stacks.iop.org/1748-0221/3/i=08/a=S08005.

[2] Pythia. http://home.thep.lu.se/ torbjorn/Pythia.html. Online: accessed 4-May-2018.

[3] Spradlin, P. The LHCb prompt charm triggers (2013). URL https://cds.cern.ch/record/1596742.

[4] Murray, D. The LHCb VELO Upgrade (2018). URL https://cds.cern.ch/record/2655463.

[5] Poikela, T. *et al.* VeloPix: the pixel ASIC for the LHCb upgrade. *JINST* **10**, C01057 (2015). URL https://cds.cern.ch/record/2189759.

[6] Morton, A. *Characterisation of Prototype Strip Detectors for the Upgrade of the LHCb VELO*. Master's thesis, University of Glasgow (2013).

[7] Akiba, K. *et al.* Radiation Damage Effects and Operation of the LHCb Vertex Locator. *IEEE Transactions on Nuclear Science* **65**, 1127–1132 (2018).

[8] Akiba, K. *et al.* Radiation Damage Effects and Operation of the LHCb Vertex Locator. *IEEE Trans. Nucl. Sci.* **65**, 1127–1132 (2018).

[9] Collins, P. Private communication.

[10] Moll, M. *Radiation Damage in Silicon Particle Detectors*. Ph.D. thesis, Fachbereich Physik der Universitt Hamburg (1999).

[11] Ansys Q3D Extractor (2014). URL http://www.ansys.com/Products/Simulation+Technology/Electronics/Signal+Integrity/ANSYS+Q3D+Extractor.

[12] EEWeb Calculator for Edge Coupled Stripline Impedance. https://www.eeweb.com/tools/edge-coupled-stripline-impedance. Accessed: 2020-07-03.

[13] Wraight, K. Private Communication.

[14] Ordan, J. M. LHCb Upgrade 2018. Beam test North Area in Prvessin (bldg 887) (2018). URL https://cds.cern.ch/record/2644707. General Photo.

[15] Evans, L. & Bryant, P. LHC Machine. *JINST* **3**, S08001 (2008).

[16] *LEP design report* (CERN, Geneva, 1984). URL https://cds.cern.ch/record/102083. Copies shelved as reports in LEP, PS and SPS libraries.

[17] G., A. *et al.* *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*. CERN Yellow Reports: Monographs (CERN, Geneva, 2017). URL https://cds.cern.ch/record/2284929.

[18] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST* **3**, S08003. 437 p (2008). URL https://cds.cern.ch/record/1129811. Also published by CERN Geneva in 2010.

[19] The CMS collaboration. The CMS Experiment at the CERN LHC. *JINST* **3**, S08004 (2008).

[20] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST* **3**, S08002 (2008). URL https://cds.cern.ch/record/1129812. Also published by CERN Geneva in 2010.

[21] Kulyavtsev, A. *et al.* BTeV: An Experiment to measure mixing, CP violation, and rare decays of beauty and charm at the Fermilab collider. In *High energy physics. Proceedings, 30th International Conference, ICHEP 2000, Osaka, Japan, July 27-August 2, 2000. Vol. 1, 2* (2000). URL http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-00-151. hep-ex/0006037.

[22] Aaij, R. *et al.* Performance of the LHCb vertex locator. *Journal of Instrumentation* **9**, P09007–P09007 (2014). URL https://doi.org/10.1088%2F1748-0221%2F9%2F09%2Fp09007.

[23] Jacobsson, R., Alemany-Fernandez, R. & Follin, F. The LHCb Online Luminosity Monitoring and Control 4 p (2013). URL https://cds.cern.ch/record/1610623.

[24] Aaij, R. *et al.* Precision measurement of the B_s^0-$\bar{B}_s^0$ oscillation frequency with the decay $B_s^0 \to D_s^- \pi^+$. *New J. Phys.* **15**, 053021 (2013). 1304.4741.

[25] Löchner, S. & Schmelling, M. The Beetle Reference Manual - chip version 1.3, 1.4 and 1.5. Tech. Rep. LHCb-2005-105. CERN-LHCb-2005-105, CERN, Geneva (2006).

[26] Jans, E. Operational aspects of the VELO cooling system of LHCb. *PoS* **Vertex2013**, 038. 10 p (2014). URL https://cds.cern.ch/record/1664281. PoS(Vertex 2013)038.

[27] Gassner, J., Needham, M. & Steinkamp, O. Layout and Expected Performance of the LHCb TT Station. Tech. Rep. LHCb-2003-140, CERN, Geneva (2004). URL https://cds.cern.ch/record/728548.

[28] Barbosa-Marinho, P. R. *et al. LHCb inner tracker: Technical Design eport*. Technical Design Report LHCb (CERN, Geneva, 2002). URL https://cds.cern.ch/record/582793. Revised version number 1 submitted on 2002-11-13 14:14:34.

[29] Barbosa-Marinho, P. R. *et al. LHCb outer tracker: Technical Design Report*. Technical Design Report LHCb (CERN, Geneva, 2001). URL https://cds.cern.ch/record/519146.

[30] Amato, S. *et al. LHCb calorimeters: Technical Design Report*. Technical Design Report LHCb (CERN, Geneva, 2000). URL https://cds.cern.ch/record/494264.

[31] The LHCb Collaboration. Opposite-side flavour tagging of b mesons at the lhcb experiment. *The European Physical Journal C* **72**, 2022 (2012). URL https://doi.org/10.1140/epjc/s10052-012-2022-1.

[32] Barbosa-Marinho, P. R. *et al. LHCb muon system: Technical Design Report*. Technical Design Report LHCb (CERN, Geneva, 2001). URL https://cds.cern.ch/record/504326.

[33] Ogilvy, S. *Branching ratios of charmed baryons in the LHCb experiment*. Ph.D. thesis, Glasgow U. (2015). URL http://theses.gla.ac.uk/6463/.

[34] Aaij, R. *et al.* Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. *JINST* **14**, P04013 (2019). 1812.10790.

[35] Aaij, R. *et al.* The LHCb Trigger and its Performance in 2011. *JINST* **8**, P04022. 31 p (2012). URL https://cds.cern.ch/record/1493820.

[36] Alves, A. A., Jr. *et al.* Performance of the LHCb muon system. *JINST* **8**, P02022 (2013). 1211.1346.

[37] Aaij, R. *et al.* LHCb Detector Performance. *Int. J. Mod. Phys.* **A30**, 1530022 (2015). 1412.6352.

[38] Sciascia, B. LHCb Run 2 Trigger Performance. *PoS* **BEAUTY2016**, 029. 7 (2016). URL https://cds.cern.ch/record/2208038.

[39] The LHCb Collaboration. Computing Model of the Upgrade LHCb experiment. Tech. Rep. CERN-LHCC-2018-014. LHCB-TDR-018, CERN, Geneva (2018). URL https://cds.cern.ch/record/2319756.

[40] Benson, S., Gligorov, V., Vesterinen, M. A. & Williams, J. M. The LHCb Turbo Stream. *Journal of Physics: Conference Series* **664**, 082004 (2015). URL http://stacks.iop.org/1742-6596/664/i=8/a=082004.

[41] The LHCb Collaboration. LHCb VELO Upgrade Technical Design Report. Tech. Rep. CERN-LHCC-2013-021. LHCB-TDR-013 (2013). URL https://cds.cern.ch/record/1624070.

[42] Bediaga, I. *et al.* Framework TDR for the LHCb Upgrade: Technical Design Report. Tech. Rep. CERN-LHCC-2012-007. LHCb-TDR-12, CERN, Geneva (2012).

[43] Gromov, V. *et al.* Development of a low power 5.12 Gbps data serializer and wireline transmitter circuit for the VeloPix chip. *Journal of Instrumentation* **10**, C01054–C01054 (2015). URL https://doi.org/10.1088%2F1748-0221%2F10%2F01%2Fc01054.

[44] Buytaert, J. *et al.* Micro channel evaporative CO2 cooling for the upgrade of the lhcb vertex detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **731**, 189–193 (2013). URL https://www.sciencedirect.com/science/article/pii/S0168900213003811. PIXEL 2012.

[45] Nomerotski, A. *et al.* Evaporative CO2 cooling using microchannels etched in silicon for the future LHCb vertex detector. *Journal of Instrumentation* **8**, P04004–P04004 (2013). URL https://doi.org/10.1088%2F1748-0221%2F8%2F04%2Fp04004.

[46] de Aguiar Francisco, O. *et al.* Evaporative CO2 microchannel cooling for the LHCb VELO pixel upgrade. *Journal of Instrumentation* **10**, C05014–C05014 (2015). URL https://doi.org/10.1088/1748-0221/10/05/c05014.

[47] Bitadze, A. & Akiba, K. A silicon microchannel bi-phase CO2 cooling for lhcb. In *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 1–5 (2019).

[48] The LHCb Collaboration. LHCb Tracker Upgrade Technical Design Report (2014). URL https://cds.cern.ch/record/1647400.

[49] Mazorra, J. *et al.* PACIFIC: the readout ASIC for the SciFi tracker of the upgraded LHCb detector. *Journal of Instrumentation* **11**, C02021–C02021 (2016). URL https://doi.org/10.1088%2F1748-0221%2F11%2F02%2Fc02021.

[50] Carniti, P. *et al.* CLARO-CMOS, a very low power ASIC for fast photon counting with pixellated photodetectors. *Journal of Instrumentation* **7**, P11026–P11026 (2012). URL https://doi.org/10.1088%2F1748-0221%2F7%2F11%2Fp11026.

[51] The LHCb Collaboration. LHCb PID Upgrade Technical Design Report. Tech. Rep. CERN-LHCC-2013-022. LHCB-TDR-014 (2013). URL https://cds.cern.ch/record/1624074.

[52] LHCb Trigger and Online Upgrade Technical Design Report. Tech. Rep. CERN-LHCC-2014-016. LHCB-TDR-016 (2014). URL https://cds.cern.ch/record/1701361.

[53] The LHCb Collaboration. Upgrade Software and Computing. Tech. Rep. CERN-LHCC-2018-007. LHCB-TDR-017, CERN, Geneva (2018). URL https://cds.cern.ch/record/2310827.

[54] Bird, T. *et al.* Simulated Performance of a Strip-Based Upgraded VELO. Tech. Rep. LHCb-PUB-2014-033. CERN-LHCb-PUB-2014-033, CERN, Geneva (2014). URL https://cds.cern.ch/record/1697944.

[55] Marco-Hernández, R. A portable readout system for silicon microstrip sensors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **623**, 207 – 209 (2010). URL http://www.sciencedirect.com/science/article/pii/S0168900210004821. 1st International Conference on Technology and Instrumentation in Particle Physics.

[56] Haefeli, G. *et al.* The {LHCb} {DAQ} interface board {TELL1}. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **560**, 494 – 502 (2006). URL http://www.sciencedirect.com/science/article/pii/S0168900205026100.

[57] URL http://www.nndc.bnl.gov/.

[58] Hickling, A. *et al.* Use of IT (current vs temperature) scans to study radiation damage in the LHCb VELO. Tech. Rep. LHCb-PUB-2011-021. CERN-LHCb-PUB-2011-021, CERN, Geneva (2011).

[59] The LHCb Collaboration. LHCb VELO TDR: Vertex locator. Technical design report. Tech. Rep. CERN-LHCC-2001-011 (2001).

[60] The ATLAS Collaboration. Modelling radiation damage to pixel sensors in the ATLAS detector. *JINST* **14**, P06012. 57 p (2019). URL https://cds.cern.ch/record/2673914. Submitted to JINST.

[61] The ATLAS Collaboration. Production and integration of the ATLAS insertable b-layer. *Journal of Instrumentation* **13**, T05008–T05008 (2018). URL https://doi.org/10.1088%2F1748-0221%2F13%2F05%2Ft05008.

[62] Ansys 14 (2013). URL http://www.ansys.com/.

[63] Chilingarov, A. Temperature dependence of the current generated in Si bulk. *Journal of Instrumentation* **8**, P10003–P10003 (2013). URL https://doi.org/10.1088%2F1748-0221%2F8%2F10%2Fp10003.

[64] Beteta, C. A. *et al.* 8-channel prototype of SALT readout ASIC for upstream tracker in the upgraded LHCb experiment. *Journal of Instrumentation* **12**, C02007–C02007 (2017). URL https://doi.org/10.1088%2F1748-0221%2F12%2F02%2Fc02007.

[65] Che, W., Tang, Y. F., Zhang, J. & Chow, Y. L. Formulas of dielectric and total attenuations of a microstrip line. *Radio Science* **45** (2010). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009RS004246.

[66] Pozar, D. *Microwave Engineering* (Wiley, 2004).

[67] E.Hammerstad, O. Accurate models for microstrip computer-aided design (IEEE, 1980). URL https://doi.org/10.1109/MWSYM.1980.1124303.

[68] Huray, P. *et al.* Fundamentals of a 3-d “snowball” model for surface roughness power losses. 121 – 124 (2007).

[69] Dean, C. *Time dependent studies of* $B \rightarrow h^{+}h^{'-}$ *decays and research and operation for the VELO project at LHCb*. Ph.D. thesis, University of Glasgow (2019). URL http://theses.gla.ac.uk/40969.

[70] Sneha, N. *On-detector electronics for high speed data transport, control and power distribution for the LHCb VELO and ATLAS Pixel Upgrades*. Master’s thesis, Glasgow U. (2020).