



**Dissertation zur Erlangung des
akademischen Grades Dr. rer. nat.**

SciFi meets GPU

Tracking performance and GPU trigger studies for the SciFi Tracker

**vorgelegt von
Holger Stevens
geboren am 16.08.1988 in Nordhorn**

Lehrstuhl für Experimentelle Physik V
Fakultät Physik
Technische Universität Dortmund

Dortmund, April 2021

Der Fakultät Physik der Technischen Universität Dortmund zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften vorgelegte Dissertation.

Gutachter:

Prof. Dr. Bernhard Spaan
Prof. Dr. Dr. Wolfgang Rhode

Abstract

This thesis presents studies towards the LHCb detector upgrade for the LHC Run III. Mostly the entire detector is renewed to profit from the changed beam conditions. Two of the main changes are a fivefold instantaneous luminosity and a pure software trigger. Because of the higher luminosity, the track density is increased. Therefore, components with higher granularity are used, and one of them is the SciFi tracker. It is made of 250 μm thick scintillating fibres. The light output of these fibres decreases through radiation damages and natural material ageing. These effects are added to the detector simulation. Thereby the tracking efficiency of the SciFi is simulated for the entire planned lifetime. The first stage of the LHCb software trigger reduces the incoming data-rate of 40 Tbit s^{-1} by a factor of 40. The necessary computations are parallelisable, and for this reason, a GPU based implementation is suitable. A GPU version of the SciFi hit decoder was implemented during the baseline architecture change process from CPU to GPU. The LHCb detector has a diamond-sensor based safety system, which monitors the LHC beam. This system is renewed in parallel to the detector upgrade. Measurements towards the diamond sensor efficiency characterisation with a strontium source are presented.

Kurzfassung

Diese Arbeit präsentiert Studien bezüglich des LHCb Detektor upgrades für den LHC Run III. Fast der gesamte Detektor wird erneuert, um von den geänderten Strahlbedingungen zu profitieren. Zwei der Hauptänderungen sind die fünffache instantane Luminosität und der ausschließlich softwarebasierte trigger. Aufgrund der höheren Luminosität erhöht sich die Spurdichte. Deshalb werden Komponenten mit höheren Granularität verwendet wovon eine der SciFi tracker ist. Dieser besteht aus 250 μm dicken szintillierenden Fasern. Die Lichtausbeute der Fasern verringert sich durch Strahlenschäden und natürlicher Materialalterung. Diese Effekte wurden der Detektorsimulation hinzugefügt. Dadurch kann die Spurfindungseffizienz des SciFi trackers für die gesamte geplante Laufzeit simuliert werden. Die erste Stufe des LHCb Software triggers reduziert die Eingangsdatenrate von 40 Tbit s^{-1} um einen Faktor 40. Die notwendigen Berechnungen sind parallelisierbar und deshalb für eine GPU basierte Implementierung geeignet. Eine GPU Version des SciFi hit Dekodierers wurde, während des Prozesses zur Änderung der grundsätzlichen Architektur von CPU zu GPU, implementiert. Der LHCb Detektor verfügt über ein Diamantsensor basiertes Sicherheitssystem, welches den LHC Strahl überwacht. Dieses System wird parallel zum Detektor erneuert. Messungen bezüglich der Diamantsensor Effizienz Charakterisierung mittels einer Strontium Quelle werden vorgestellt.

Contents

1	Introduction	1
2	Physics at the Large Hadron Collider	3
2.1	The Standard Model of Particle Physics	3
2.2	Testing Physical Models	6
3	The LHCb Collaboration and Experiment	9
3.1	The Physics at LHCb	9
3.2	The LHCb Upgrade	12
3.3	The LHCb Upgrade Detector	13
3.4	The LHCb Upgrade Software Framework	18
3.5	LHCb Online Farm and Trigger	20
4	The Scintillating Fibre Tracker	25
4.1	The Scintillating Fibre	25
4.1.1	Scintillating Mechanism	25
4.1.2	Light Guidance	26
4.1.3	Light Loss	27
4.2	Detector Layout	28
4.2.1	Light Detection	30
4.2.2	SciFi Readout Electronics	32
4.3	The Clustering Algorithm	33
4.4	SciFi Tracker Raw Banks	35
4.5	Simulation of the SciFi Tracker	36
4.6	SciFi Decoding	37
4.6.1	Decoding Version 5	38
4.6.2	Decoding Version 6	39
4.6.3	Decoding Version 4	39
5	Simulation of the SciFi Over the Lifetime	41
5.1	Particle Matter Interactions	41
5.2	Light Yield Reduction Processes	43
5.2.1	Radiation Damage of Scintillating Fibres	43
5.2.2	Natural Fibre Ageing	45
5.3	Light Yield Attenuation Maps	46
5.4	Monte Carlo Sample Generation	48
5.5	LHCb Track Types	49

6 Long-term Studies	53
6.1 SciFi Cluster	53
6.2 Tracking Efficiency During the Detector Lifetime	58
6.3 Conclusion and Outlook	64
7 GPU Based Trigger	67
7.1 General Purpose Graphics Processing Unit	67
7.1.1 Hardware	68
7.1.2 Software	69
7.2 The Allen Project	71
7.2.1 Integration of Allen	72
7.2.2 Allen Sequence	75
7.3 SciFi Decoding on GPU	77
7.3.1 GPU Raw Bank Decoder Version 4	77
7.3.2 GPU Raw Bank Decoder Version 6	80
7.3.3 GPU Raw Bank Decoder Version 5	82
7.3.4 Conclusion of the GPU based Raw Bank Decoder	82
7.4 Allen Performance	82
7.5 Conclusion and Outlook	85
8 Diamond Measurements for the BCM Upgrade	87
8.1 BCM Upgrade	87
8.2 Diamond as Radiation Detector	88
8.2.1 Diamond Production	89
8.2.2 Diamond Surface and Contacts	91
8.2.3 Diamond Sensor Performance Characteristics	91
8.2.4 Strontium-90 Radiation	91
8.3 Diamond Measurements	92
8.3.1 Current-Time Curves	92
8.3.2 Proton Beam	99
8.4 Conclusion and Outlook	102
9 Conclusion	105
Appendices	
A SciFi performance studies	107
Bibliography	109
Acknowledgements	119

1 Introduction

The Standard Model of particle physics (SM) is a well-tested theory, which describes the interaction of particles and forces on a subatomic level [1, 2, 3]. With the discovery of the predicted Higgs particle in 2012 [4] the model was completed. Nevertheless, there are effects which are not described by the SM. For example, the gravity, which has no significant effect on the subatomic scale is not a part of the SM. Further inconsistencies are found in the CP -violation and baryon asymmetry.

Only $\approx 5\%$ of the universe consists of ordinary observable matter and energy. The remaining parts are the so-called Dark Matter and Dark Energy, and their contribution is 24 % and 72 % respectively. Through the mismatch of observed and expected galaxy rotation velocities, the Dark Matter was discovered, which interaction is only gravitational [5]. The Dark Energy is theoretically needed to explain the accelerated expansion of the early universe [6].

Another known tension of the SM is the mass of the neutrinos. They are predicted to be mass-less, but for the observed neutrino oscillation a mass is essential [7]. Also it is unclear whether neutrinos are Majorana or Dirac particles.

During the Big Bang matter and antimatter should be produced in equal shares. Though almost everything we can observe is matter, and only a small amount of antimatter can be observed. To explain this the emergence of this imbalance, Andrei Sakharov postulated three conditions in 1967 [8]. First baryon number violation, second interactions out of the thermal equilibrium and third violation of C and CP -Symmetry, with C and P being the conjugation and parity transformation of the charge like quantum numbers respectively. Even though CP -violation is allowed in the SM and proven by measurements, the amount is too small to cause the observed matter-antimatter asymmetry.

For all these reasons other or new theories are needed to describe the physics beyond the SM, so-called New Physics. The experiments at the Large Hadron Collider (LHC) can test and constrain some of the model predictions. The LHC and the standard model are introduced in chapter 2. One of the experiments at the LHC is the Large Hadron Collider beauty experiment (LHCb), which will be described in detail in chapter 3. It is dedicated to: perform high precision measurements in the heavy flavour sector, measure rare decays of charm and beauty hadrons and find new sources of CP -violation. During the period of 2011 to 2018 a dataset of 8.7 fb^{-1} was recorded [9]. This is the world largest sample of exclusive charm and beauty decays, and the LHCb Collaboration has performed many analyses with it. The very rare decay $B_s^0 \rightarrow \mu^+ \mu^-$ was found [10] and CP -violation in the charm sector was discovered [11]. However, the majority of analyses are limited by statistics. More significant results will reduce the possible parameter room of New

1 Introduction

Physics models. Therefore, the dataset of recorded decays have to be enlarged significantly, as the statistical error scales with $1/\sqrt{N}$, where N is the number of events.

For this reason the LHCb detector is undergoing major upgrades at the moment. The new detector will record data with a 5 times higher instantaneous luminosity [12]. With the new detector it is expected to record a dataset of 50 fb^{-1} during Run III, which lasts 10 years. This thesis will focus two parts. First the long-term performance study of the scintillating fibre tracker (SciFi). The SciFi is a sub-detector within the tracking system of the upgrade detector, cf. [chapter 4](#). The scintillating fibres, get cloudy by radiation damages. The implementation of this damage and an additional ageing effect into the detector simulation are presented in [chapter 5](#). With the detector simulation sample for different conditions are generated. The effect on the general tracking efficiency of the overall detector is presented in [chapter 6](#).

In the second part studies on the acceleration of the trigger system with Graphics Processing Units (GPU) are presented. These base on the new data taking concept of the LHCb experiment. These concept and the upgrade detector are introduced in [chapter 3](#). The new trigger and the basic concept of graphic cards programming are introduced in [chapter 7](#). Here, also the contribution to this project, which are the implementations of various version of the SciFi hit decoder are presented.

The Beam Condition Monitor is a safety system of the LHCb detector. Diamonds are used as the active material. Similar to the detector, this component is also upgraded. Studies for the characterisation of the future diamond sensors are presented in [chapter 8](#). Finally, the thesis concludes with a summary in [chapter 9](#).

2 Physics at the Large Hadron Collider

At the 29th September of 1954, the European Organisation for Nuclear Research was founded. It is located near Geneva, at the border between the Swiss and France. The first accelerator there was a 600 MeV Synchrocyclotron. Meanwhile, the latest named the Large Hadron Collider accelerates up to 7 TeV. Between these two accelerators, many others were built. Some of these are still used as pre-accelerator for the LHC and can be seen in the overview in [Figure 2.1](#). The LHC, with its 27 km circumference, is the world's largest and most powerful particle accelerator [13]. It is placed in a tunnel under the earth in a depth between 50 and 100 m. This positioning enables a natural radiation shield that ensures safe usage of the area above the accelerator.

The main object which are accelerated in the LHC are protons. They can be accelerated to an energy of 7 TeV which leads to a collision energy in the centre of mass system of 14 TeV, to reach and hold this beam energy is quite challenging, as the particle have to be hold on track. Therefore, super conducting magnets are used. Furthermore, two separated beam pipes are next to each other inside of the magnets, there the protons are accelerated in opposite directions.

At four places in the LHC it is possible to collide the two proton beams. These proton-proton collisions are the subject of interest of the experiments which are located there. Namely these are ATLAS [14], ALICE [15], CMS [16] and LHCb. The last one, LHCb, will be explained in detail in [chapter 3](#), while the other will only be introduced briefly here. All of them are general purpose detectors, this means that the detector is build around the interaction point of the collisions. All of them are looking for New Physics, but they have focuses in different fields. ATLAS and CMS are looking for new particles or effects. As mentioned before, the higgs particle completed the SM and these two experiments have discovered it. The third experiment, ALICE, is studying quark-gluon-plasma, which is produced in heavy ion collisions. Therefore, mostly lead is accelerated in the collider instead of protons.

2.1 The Standard Model of Particle Physics

One major goal of the various experiments at the LHC is to prove the predictions of the Standard Model of particle physics. Therefore, the theory and the corresponding concepts are introduced briefly. The actual formulation was finished in the 1970s and is one of the best proven theories in the world. The SM includes three of the four fundamental forces. These are the electromagnetic, weak and strong interaction, while gravity is not included. The fundamental particles on which

2 Physics at the Large Hadron Collider

The CERN accelerator complex Complexe des accélérateurs du CERN

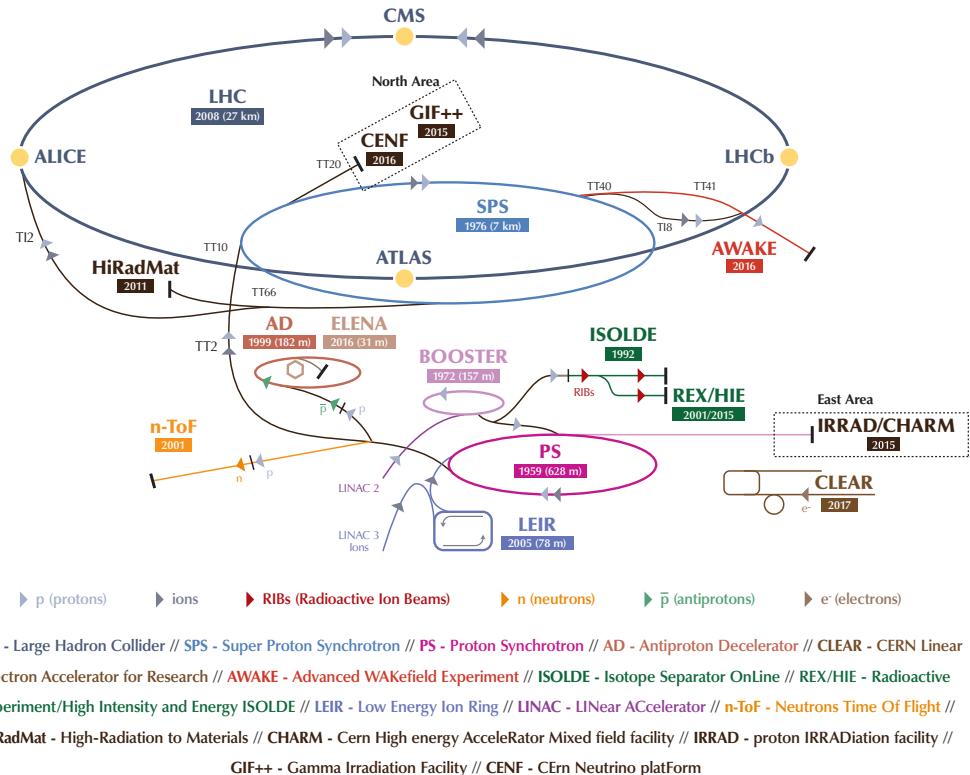


Figure 2.1 – ”The LHC is the last ring (dark blue line) in a complex chain of particle accelerators. The smaller machines are used in a chain to help boost the particles to their final energies and provide beams to a whole set of smaller experiments, which also aim to uncover the mysteries of the Universe.”[17]

2.1 The Standard Model of Particle Physics

the theory is based on are shown in [Figure 2.2](#). The particles can be split into fermions with a half integer spin and bosons with a integer spin. The fermions are split into six quarks and six leptons whereby these are split into three generation each. The six quarks are named up, down, charm, strange, top and bottom. Due to their colour charge they are the only particles which can interact with the gluon. Electron, muon and tau are the fundamental leptons, and the other three leptons are the corresponding neutrinos. The neutrinos can only interact via the weak force while the other leptons are charged and can couple therefore also to the photon. As already indicated, the bosons are the force carrier of the fundamental forces. An exception is the Higgs boson. For a long time, it was the last missing particle of the SM until the discovery in 2012. The Higgs boson provides the mass of the fermions due to the Higgs mechanism [\[18, 19\]](#). The Higgs mechanism causes a spontaneously symmetry breaking and therefore generates the fermion masses. For the complete theory also antiparticles for the fermions are needed. These are not listed here to simplify the picture. The antifermions have the same characteristics; only the quantum numbers are inverted.

Standard Model of Elementary Particles

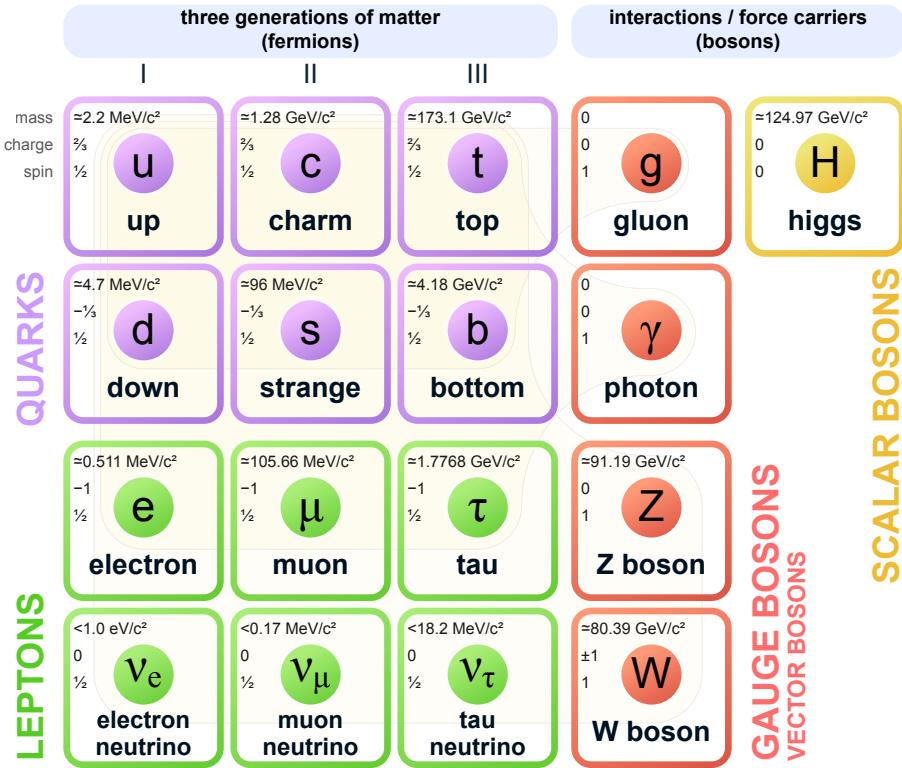


Figure 2.2 – Fundamental particles in the SM. The quarks and leptons are arranged in three generations. [20]

2.2 Testing Physical Models

In this section possibilities to test and constrain physics model are presented briefly. Thereby the focus is on measurements the LHCb detector can perform. The major model to test is the SM. One implication of the SM is the Cabibbo–Kobayashi–Maskawa (CKM) matrix [21, 22]. This matrix describes the magnitude of quark mixing due to flavour-changing weak interactions. Because of the three quark generation, the matrix dimension is 3×3 . A necessity of the theory is the unitarity of the CKM matrix.

This condition can be used to form the so-called unitarity triangles. An example triangle is shown in Figure 2.3, and the image is provided by the CKM-fitter group [23]. This group combines results from all available experiments in this area of science. The CKM matrix can be broken down to 3 mixing angles and a complex phase. A common way to test the SM is to measure the sides and angles of a CKM triangle. The results of the measurements then should add up to a closed triangle.

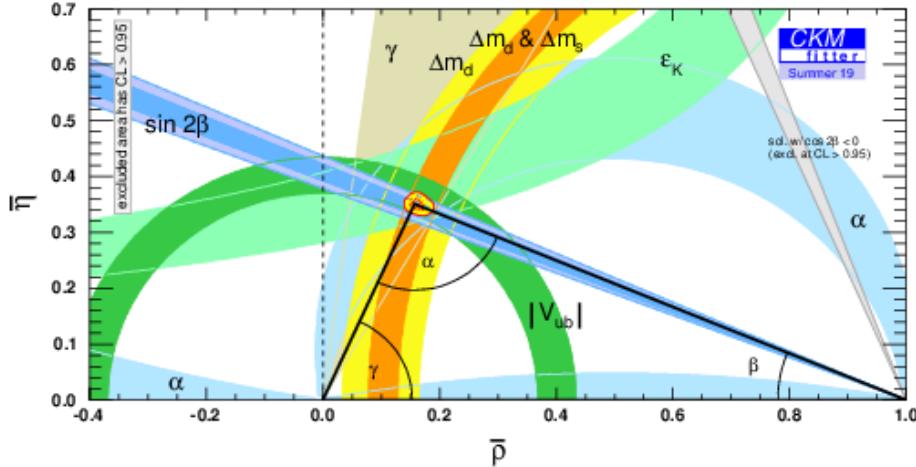


Figure 2.3 – CKM unitarity triangle fit [23].

In the plot of the CKM triangle a little tension at the apex can be seen. More significant results are needed to constrain the apex.

The LHCb experiment is one of the major contributors to the value for $\sin(2\beta)$. The value can be calculated from the measurement of the decay-time-dependent CP -violation in the decay $B^0 \rightarrow J/\psi K_S^0$ [24]. This measurement belongs to the direct proofs of the Standard model. Another possibility are indirect measurements or effects. No significant deviations from the Standard Model are found yet. Therefore, the manifestation of New Physics effects is expected to be small. That's why very rare Standard Model processes are measured. For them unpredicted effects are relatively significant. The probability for a decay scales with the squared values of CKM matrix and can be depicted in Feynman-diagrams. These diagrams are also a mathematical expression for the probability calculation. All matrix elements within the diagram are multiplied and as no element is greater than 1 the probability decreases for more complex decays. The simplest decays have one vertex or matrix element, these are named tree-level. With the next higher order, two vertices, it is possible to create so-called loop processes. In these loops particles can exist virtually. At this point also unknown particles can be created. If so they will effect the probability of the decay. One of these very rare processes in the Standard Model is the decay of a B-Meson into two muons. Another prediction of the Standard Model is the lepton-flavour-universality. This means that the coupling of the weak interaction is equal to all the three lepton generations. The LHCb experiment has measured the ratio of processes including the emission of 2 leptons, either an electron and a positron, or a muon and an antimuon. Thereby, the result of this ratio measurements is in tension with the SM prediction [25].

3 The LHCb Collaboration and Experiment

In this chapter, the LHCb collaboration is presented. The physics program and the motivation for an upgrade are shown. Also, the components of the upgraded detector itself are explained. The experiment is located at the eighth interaction point of the LHC (point 8). The LHCb collaboration was founded in 1996. Today 108 institutes and laboratories belong to the collaboration and are listed on the author list in December 2020. In total, around 1400 people are a member of the group, thereof about 900 are author. It is an international collaboration, and the distribution of the institutes is shown in [Figure 3.1](#).

3.1 The Physics at LHCb

As explained in the previous chapter, there are some open questions or puzzles in the field of high energy physics. The Large Hadron Collider beauty experiment tries to solve some of them. As the name indicates, the experiment is focused on the b-physics sector. Also charm quark containing decays are studied. Two main fields are the precision measurement of CP violation, and rare decays [26]. This particular physics program is the reason for the special shape of the detector. The detector is designed as a so-called single-arm forward spectrometer. Compared to the other detectors at the LHC, the acceptance is limited, but within these, the precision is high. This optimisation to a small acceptance results from the production process of the b-quarks in the LHC.

The most common process for the production of bb-quark pairs, at the LHC centre of mass energy, is the gluon-gluon fusion. It is likely that the two gluons have a significant different momenta. This difference leads to a boost of the produced particles into the direction of the beam-axis. In other words, the angle between the two quarks in a bb-pair is small and therefore it is likely that both are in detector acceptance. In [Figure 3.2](#) the angle distribution is shown, and the LHCb-detector acceptance is coloured in red. Nearly 25 % of the produced $b\bar{b}$ quark pairs are within the detector acceptance [27].

For later analyses, it is important to know the flavour of some particles. Therefore, the collaboration has so-called flavour tagging algorithms [28]. For these, both quarks from the fragmentation process must be in the acceptance of the detector. Also in the previous mentioned measurement of $\sin(2\beta)$ the presence of both b-quarks is essential. The neutral B meson oscillates during time in the corresponding anti meson and vice versa. If one of the initial b-quarks decays flavour specific,

3 The LHCb Collaboration and Experiment

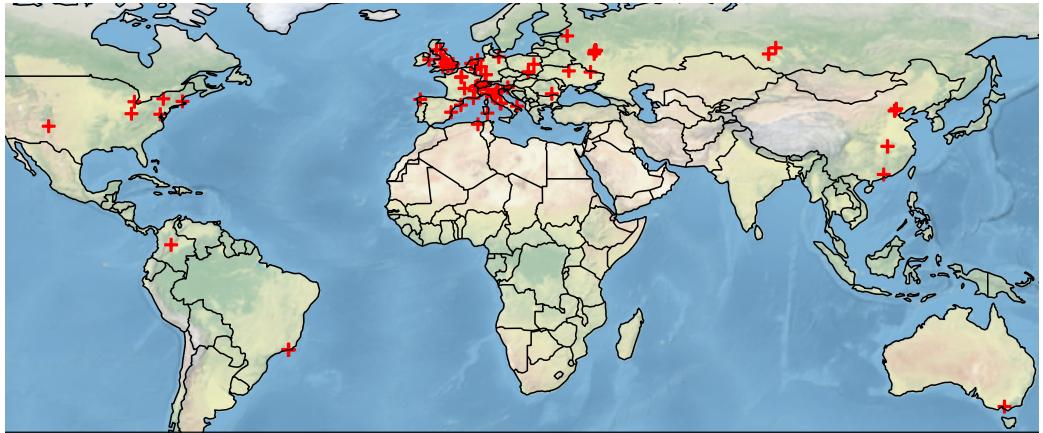


Figure 3.1 – Locations of institutes participating in the LHCb collaboration are marked on the map.

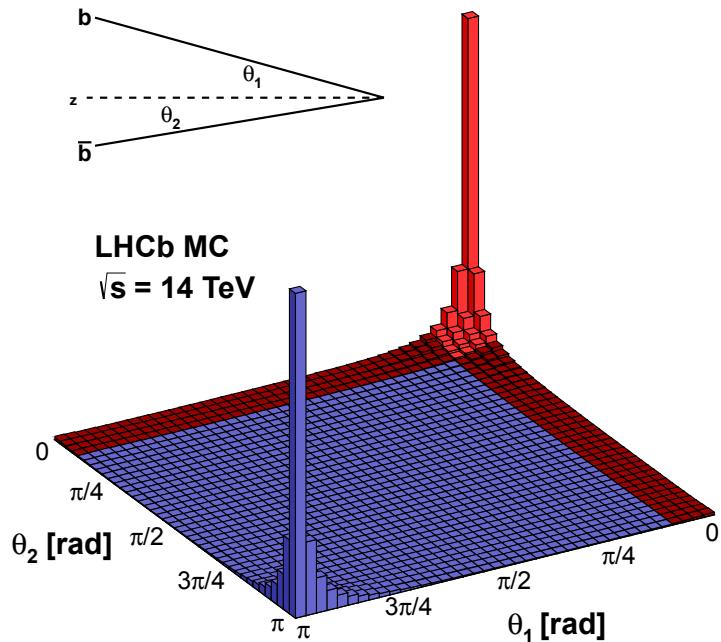


Figure 3.2 – Angular distribution of $b\bar{b}$ quark pairs produced in pp collisions at a centre of mass energy of 14 TeV. The acceptance of the LHCb detector is highlighted in red [27].

than the nature of the other b-quark is determined. Another critical factor for the detector performance is the track density. The resolution of the detector decreases with increasing track multiplicity in the detector. As the multiplicity is directly correlated with the luminosity the LHCb detector is not using the full instantaneous luminosity of the LHC beam. A system, the so-called Lumi-leveling, is adjusting the overlap of the two beams [29]. The luminosity is adjusted in that way, that the sweet-spot of best performance is reached. In the last years of data recording the average number of inelastic scattering protons per bunch crossing μ was set to one. The evaluation of the best setting for the value of μ takes the track multiplicity dependant performance of the sub-detectors into account and also the possible throughput of the Trigger system. Due to the usage of the Lumi-leveling system, the conditions in the detector stay constant, which improves the quality of the detector calibration and alignment. In Figure 3.3 the instantaneous luminosity at LHCb, ATLAS and CMS is shown for a LHC fill. For ATLAS and CMS the luminosity decreases exponentially while for LHCb the value is levelled to a constant value. At the end there is also a small decrease at LHCb, but this is due to a very long run. A typical run lasts about 12 hours.

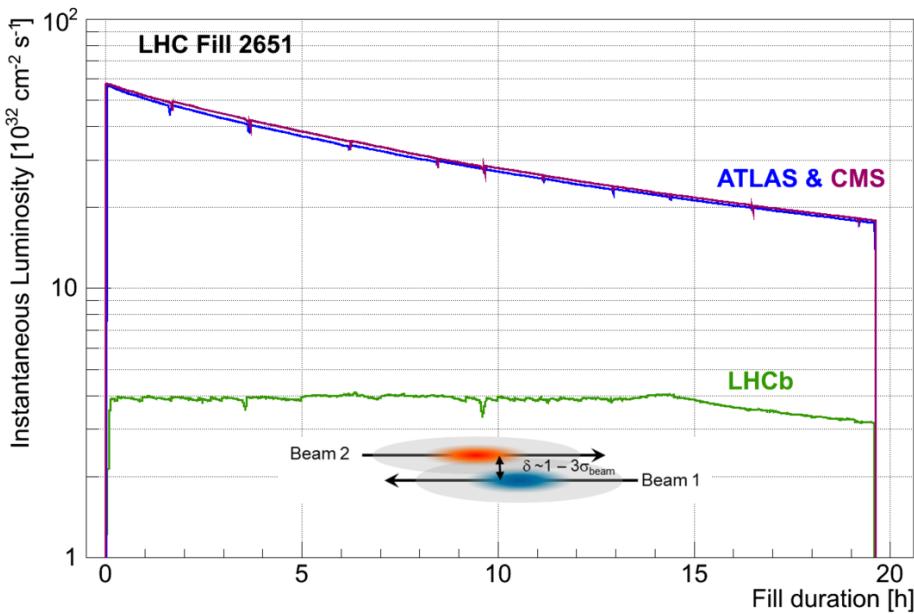


Figure 3.3 – Trend of the instantaneous luminosity during a LHC fill for ATLAS, CMS and LHCb [30].

The collaboration recorded data during Run I (2010-2013) and Run II (2015-2018) of the LHC. During this time a dataset of 8.8 fb^{-1} was recorded [9]. Based on these data, the collaboration has published over 500 paper so far. One of them describes the discovery of the decay $B_s^0 \rightarrow \mu^+ \mu^-$ [10], this decay is strongly suppressed in the

SM. Another result is the discovery of pentaquarks [31]. Also R_K , which is a factor for the lepton flavour universality, was measured. Nevertheless, the big questions which LHCb was targeted are still not answered. The most limiting factor is the statistical uncertainty. Therefore, an upgrade of the detector was planned, details toward this will be given in the next part.

3.2 The LHCb Upgrade

The main goal of the LHCb upgrade is to increase the dataset massively. It is planned to take 50 fb^{-1} in ten years in the LHC Run III with the original start in 2021 [32]. Due to the corona crisis the LHC schedule will be shifted at least to 2022. A complete 40 MHz readout of the detector is needed for the planned pure software trigger. The former trigger system was a combination of hardware and software, and the output rate of the hardware part (L0) was 1 MHz [33]. Another change will happen in the run conditions, the instantaneous luminosity will be increased by a factor of 5 to $2 \cdot 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. This leads to a higher pile-up ν , which is the average number of interactions per bunch crossing. In Figure 3.4 the luminosity dependant trigger yield for different decay channel is shown. Hereby, the L0 trigger stage is still present. It can be seen that the trigger yield of the non leptonic modes saturates. This saturation takes place at a much lower value than the planned upgrade luminosity. Therefore, the L0 trigger stage removal is a necessity to profit from the increased luminosity.

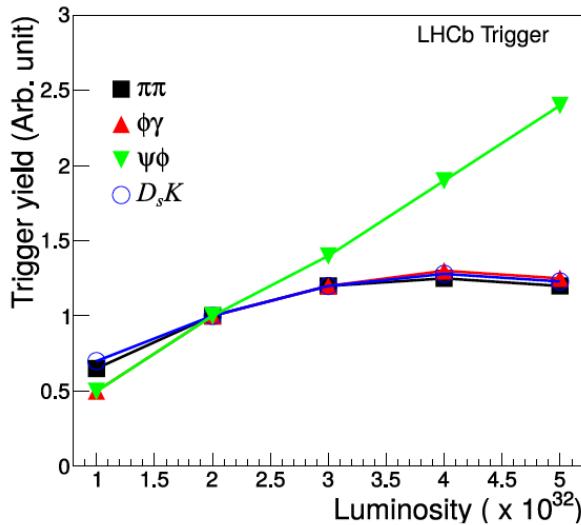


Figure 3.4 – Decay mode trigger yield for variable luminosity with L0 trigger [34].

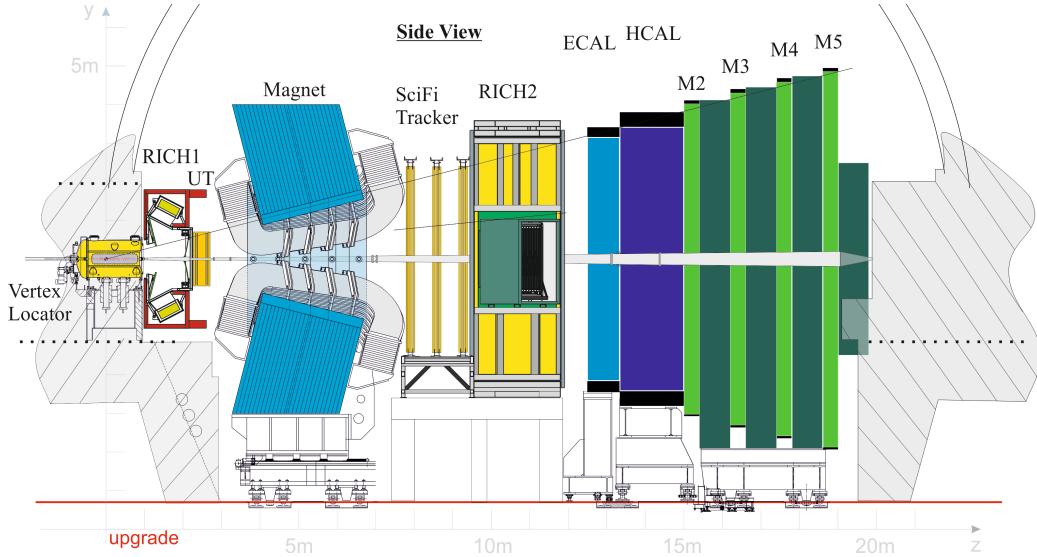


Figure 3.5 – Layout of the LHCb upgrade detector [35].

3.3 The LHCb Upgrade Detector

The following description of the LHCb detector is focused on the upgrade detector. It will be a total new detector, as only $\approx 5\%$ of the old are reused. But the general shape and concept stay the same and like the former detector it will be a single-arm forward spectrometer. The angular range which is covered is from approximately 10 mrad to 300 mrad (250 mrad) in the bending (non-bending) plane. This corresponds to a pseudo-rapidity range of $1.8 < \eta < 4.9$, with $\eta = -\log(\tan(\Theta/2))$ where Θ is the angle between the particle momentum and the beam axis.

In Figure 3.5 the LHCb detector with the corresponding coordinate system is shown. The z axis is in the direction of the beam-pipe, the y axis points to the top and the x axis is in the horizontal plane. Mainly four kinds of components are in the detector, tracking, calorimeter, muon detector and particle identification systems. Respect to the picture, the detector will be described from left to the right. The regions before and after the magnet are named upstream and downstream, respectively.

Vertex Locator

The Vertex Locator (VELO), as the name indicates, detect the position of the pp -collision, the so-called primary vertex (PV). It is also possible to identify secondary vertices in case of short-living particles. The VELO is split in two halves, both have 26 modules of 200 μm thick silicon pixel sensors. The chips are operated at a temperature below -20°C . A special micro-channel cooling concept guides liquid CO₂ trough the chips. For the reconstruction of the PV the detector needs

3 The LHCb Collaboration and Experiment

to be placed close to the beam. Therefore, the VELO components are mobile and can be moved to a distance of 5.1 mm to the beam. This also means that the detector is inside the vacuum of the LHC beam pipe. In [Figure 3.6](#) the VELO is illustrated in two different planes. In the top part the $x - z$ -plane is shown. The interaction point and a PV are visible in the centre of the detector. The z -distance between the modules is becoming greater to the outer regions. In the bottom part the VELO is shown in the $y - x$ -plane. Both situation, VELO closed and open are shown. The opening of the detector is necessary as the beam of the LHC is less stable at the beginning of a fill. The VELO could be destroyed when hit by the beam. Only after the beam is declared to be stable, the detector moves to its operating position. One system which is controlling the LHC beam is the Beam Conditions Monitor (BCM). [\[36\]](#)

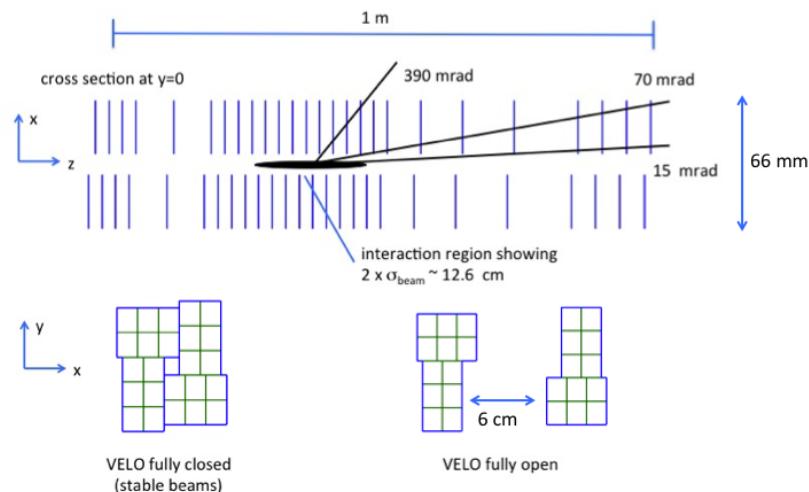


Figure 3.6 – VELO modules in the $x - y$ -plane are shown in the top. In the bottom part the fully open and closed VELO are shown, where the $y - x$ -plane is used. [\[36\]](#)

BCM

The entire energy of the LHC beam is about several hundred megajoule. Even a hit by a small beam fraction would destroy detector components. One important safety system, but not part of the detector itself, is the Beam Conditions Monitor. In the picture of the detector the BCM is not marked. The two BCM stations are mounted at different places. The first one, the upstream station, is right before the VELO. The second, the downstream station, is located between the later introduced UT and the magnet. In [Figure 3.7](#) this station is shown. The ring with the sensor diamonds is covered by a silver foil. In front of the BCM the beam-pipe

is made from Beryllium and the rear part is stainless steel. Both stations are build in general as identical systems. A station consists of eight diamond chips. The diamonds are the active material and are operated at a voltage of 200 V. When foothills of the beam are passing the diamond a current is produced and can be detected. If the beam is too unstable and the signal current is higher than defined threshold, the BCM triggers a LHC beam dump. The future thresholds are not defined yet. Simulations of the upgrade BCM within the new detector geometry concerning the changed beam profile are in progress. New diamonds will be used for the upgrade BCM, cf. [chapter 8](#). The characterisation of these diamonds is also missing. [37]

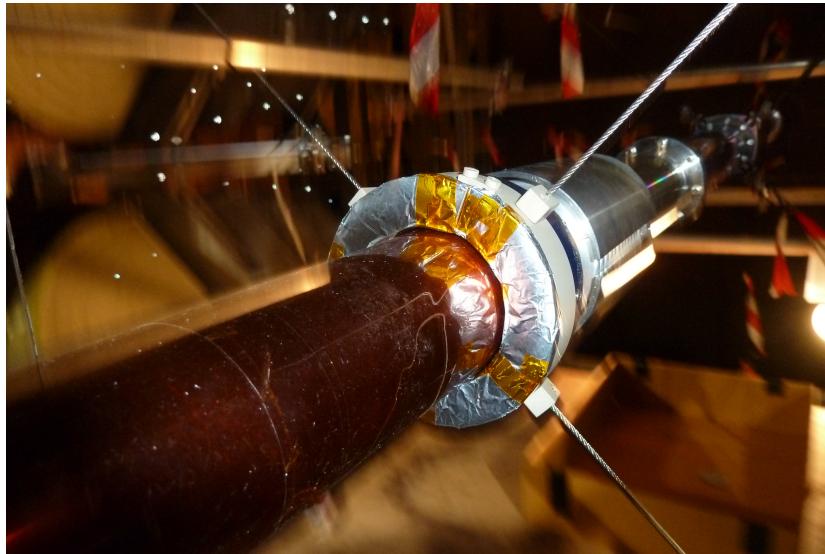


Figure 3.7 – BCM downstream station mounted around the beam-pipe.

Ring Imaging Cherenkov detector

Two Ring Imaging Cherenkov detectors (RICH) are placed at different positions in the LHCb detector. The first one RICH1, is located behind the VELO. The second RICH2 is between the SciFi-detector and the ECAL. The RICH system belongs to the Particle Identification (PID) system.

Due to the special characteristic of Cherenkov light, different particles can be distinguished. The main focus for LHCb is the kaon and pion separation. The emission angle Θ is defined by

$$\cos \Theta = \frac{m}{n \cdot \beta}$$

where n is the refractive index of the medium, m the mass of the particle and β is the ratio of the particle speed and the speed of light in vacuum. The emitted cherenkov light is guided by a mirror system to multi-anode photo-multipliers

3 The LHCb Collaboration and Experiment

where the light is detected. In the two RICH stations different radiators are used to cover a wide range of energy sensitivity. In RICH1 C_4F_{10} is used as radiator, it has a refractive index of $n = 1.0014$. This enables a $\pi - K$ separation up to about 50 GeV. The first RICH station covers the total LHCb acceptance, but RICH2 encloses only a reduced acceptance of 120 mrad (horizontal) and 100 mrad (vertical). The reduced acceptance area is due to the changed energy spectrum of the particles behind the magnet. The slow particles are bent out of the detector. That is also the reason for the different radiator in RICH2, which is CF_4 . With its smaller refractive index of $n = 1.0005$ it is possible to extend the operative region to 100 GeV. Behind RICH2 the calorimeters of the LHCb detector are positioned. [38]

Magnet

The magnet is an crucial part of the tracking system. A normal conducting dipole magnet is used which provides an integrated magnetic field of 4 Tm. This field enables the measurement of the particle charge and the momentum. Therefore, the particle track information are combined with the energy information from the calorimeter. As the magnetic field is mainly in the y -direction the bending occurs within the horizontal plane, which is referred to as the x -plane. The magnet field reaches into the SciFi which is positioned right behind the magnet. In Figure 3.8 the magnetic field along the z -axis is shown. The position of the sub-detectors along this axis is shown in the bottom part. This image refers to the old detector but neither the magnet field or the position of the sub-detectors is changed significant. [39]

Scintillating Fibre Tracker

The SciFi tracker is right in front of RICH2. It is part of the tracking system and composed of scintillating fibres. A detailed description is given in chapter 4.

Calorimeter

Like the RICH detectors the calorimeter provide PID information. Beside this they also measure the particle energy and provide some position information. Two different systems are installed, first the electromagnetic calorimeter (ECAL) and second the hadronic calorimeter (HCAL). Both systems use the same principle, an alternating system of absorber respectively scintillator plates. In the ECAL the absorber is lead, here the photons and electrons are detected. For the identification of hadrons the absorber plates are made from iron. All relevant particles except muons are stopped until this stage. [38]

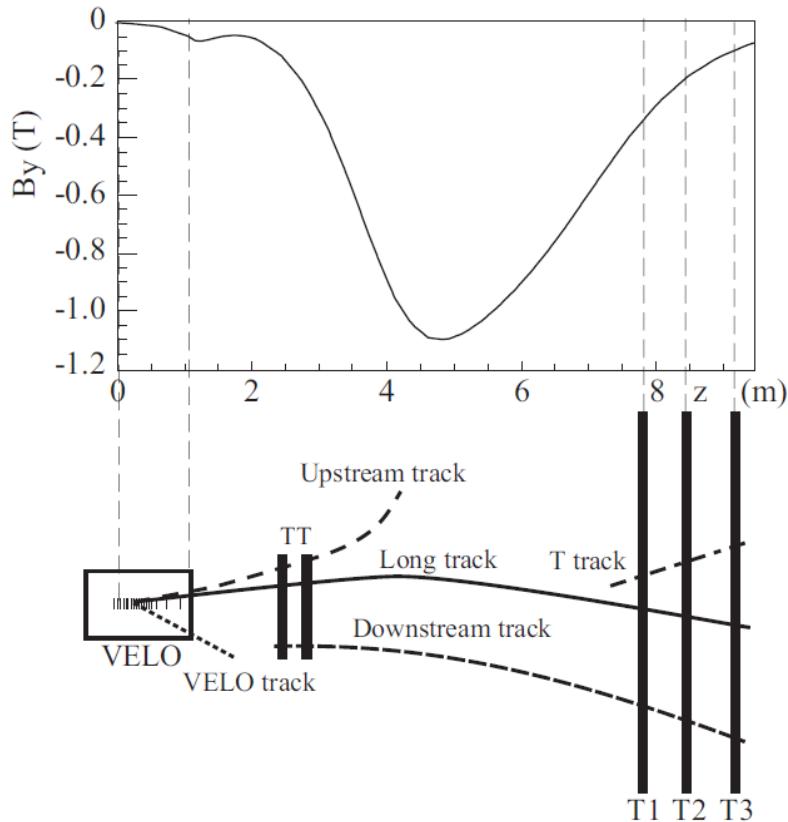


Figure 3.8 – Magnetic field along the z -axis [26].

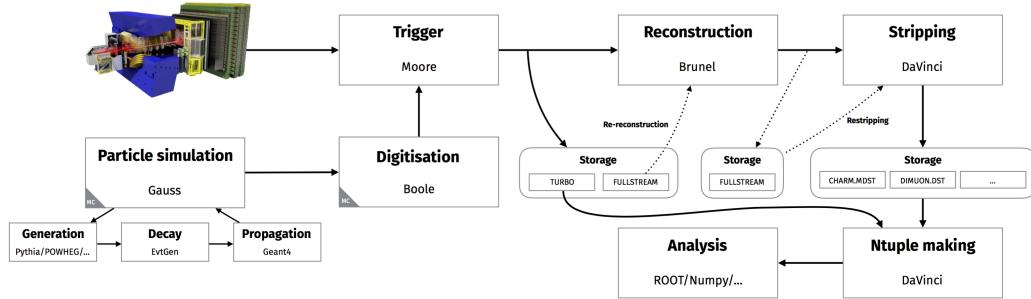
Muon system

As the name indicates the Muon system identifies muons. Beside the particle identification it provides some track information. In many parts of the LHCb physics programme muons are involved for example the decay $B_s^0 \rightarrow \mu^+ \mu^-$. Because of the important role of the muons the first station of the system was connected to the L0 trigger. This was possible as the muon chamber were fast enough to operate at 40 MHz. After the removal of the L0 trigger the station M1 is not needed anymore. For a reduced material budget in front of the calorimeters the station is removed. The remaining system is split into four stations (M2-M5). The base component of the stations are Multi-Wire Proportional Chambers (MWPC). As the occupancy in the system is higher in the inner regions, the MWPC are arranged to smaller units there. The chambers which are not modified during the upgrade only the readout is renewed. [38]

3.4 The LHCb Upgrade Software Framework

The LHCb collaboration has its own software packages and a tradition to name these software after renowned scientists. In this chapter the main parts of the LHCb software framework are explained.

For any physics program it is absolutely necessary to have a complete simulation of the detector and all following data processing steps. Cross checks for measurements need simulated events, so called Monte Carlo (MC) events. MC data sets are also used to calculate efficiencies. For example the performance of different trigger settings are tested with these data. Also for any planned changes on the detector or the data-flow or processing, it is imperative to simulate it first.



for the different components is generated. This is the digitisation step by BOOLE [50].

Boole

At this point a signal stream which is equivalent to the real detector output is generated. Therefore, all components needs to be simulated. In case of the SciFi these are: the light production based on the deposited energy; the light travelling in the fibres and the detection at the SiPM; For the read out also the thermal noise is simulated and the time behaviour of the electronic. For the other sub-detectors the corresponding parts are simulated. With this simulated detector signal it is possible to study the performance of the total detector before it is build. One example for this are the trigger studies within the Allen project, cf [section 7.2](#).

Brunel

The next step is the creation of physical objects, so called pseudo particles, from the digitised hits or information. BRUNEL performs the track reconstruction [51]. At this stage particle identification information are evaluated and combined with the reconstructed tracks. Therefore, BRUNEL needs the information of all subsystems. For example the SciFi-hit-decoder, which is ported to GPU in [section 7.2](#), is part of this sequence.

DaVinci

The output from BRUNEL is the input for the next part of the software stack, DAVINCI [52]. With DAVINCI a reconstruction of a user specified decay chain is possible. Before single analyses can be done, the total dataset is split into different categories of particle decays. This step is called stripping and it is needed to reduce the dataset size to an usable dimension. At the moment the collaboration use around 70 PB of tape storage [53]. With the so-called stripping the data is categorised and provided to the analysis groups. The size of this datasets is in the range of a few GB to some TB.

Moore

In the previous step stored data was handled. But before events get stored the trigger selection takes place to reduce the data-rate. The decision if a event is stored or not is based on trigger-lines within the trigger, which is emulated in MOORE [54]. A detailed description on the trigger can be found in [section 3.5](#). Because the decision is time critical only a partial reconstruction is done. Within MOORE, there are algorithms from the BRUNEL project used as well, but they are slightly adjusted for the online usage. A typical change is the width of search windows.

Tesla

Different data formats are used within the storage process. Up to now only the full-stream was mentioned. Here the entire event is stored and passed to the recon-

struction. Tesla provides another data format within the so-called turbo-stream [55]. Within the turbo the events are reconstructed and selection lines are applied. Due to the real time calibration of the detector the events can be stored direct without a second offline reconstruction step. An important point is that only the information of the triggered decay are stored. This means that only a few tracks are stored. As most of the information of the event are deleted the data size of a turbo event is much smaller. The average event size is about 7 kB. Therefore, it is possible to store more events at the limited bandwidth of 10 GB s^{-1} . But this also means, that later corrections or re-calibration are not possible.

3.5 LHCb Online Farm and Trigger

In this part the new LHCb online farm and it's tasks will be described. Also a brief introduction to the upgrade trigger which will use the farm is given. Because of the limitations like cooling power and space in the LHCb cavern, not all parts of the new infrastructure are placed within the cavern. It was decided to use a modular container based approach to host the Event Filter Farm (EFF). These containers are positioned at the surface of point 8 and are equipped with the needed cooling and have a network connection to the parts in the cavern.

This modular approach enables an easy possibility to scale the system up when needed. At the end of 2020 six container are bought and set up. Beside the application of the trigger the farm is also used to process Monte-Carlo (MC) events. This MC production is only possible in non data taking periods. For proper physics results the ratio of data and MC should be stay at least the same as in the previous Runs. This means that the amount of the MC production needs to grow in the same way as the stored data. For now the focus will be on the data taking application and not the MC.

[Figure 3.10](#) shows the basic scheme of the architecture for the data acquisition in Run III. First the data of the different sub-detectors needs to be merged, this happens in the Event builder (EB). The event builder server host the PCIe40 cards, which are also known as TELL40. These cards are connected to optical data cables, which are coming from the detector, for example the data links from the SciFi. In the Trigger and Online Technical Design Report (TDR) [33] it was foreseen to use 500 server in the EB. This number was chosen because of the 500 PCIe40 cards which are needed to handle the links. So one server can host one card. Due to general advances in server technology only 170 server are in the EB. As the picture is taken from the TDR the numbers are outdated. In general all descriptions are based on the TDR and in the case of major changes they are pointed out.

The reduction of EB nodes also mean that one server has to carry about three PCIe40 cards instead of one. As the EB only arrange data and not process them the input and output data-rate is unchanged at 32 Tbit s^{-1} . The events are sent from the EB to EFF. In the EFF an actual reduction of the data-rate takes place

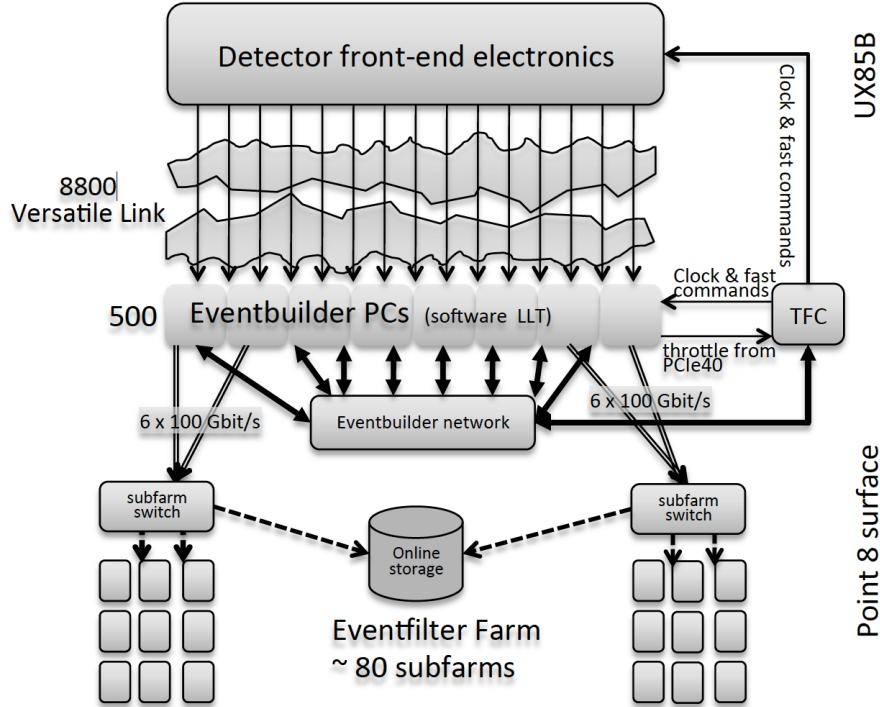


Figure 3.10 – Architecture of the upgraded LHCb readout-system [33].

due to the high level trigger (HLT) application. The incoming date-rate fluctuates but in average there will be a proton proton collision rate of $\approx 30\text{ MHz}$ at the interaction point (IP8) of LHCb. This is significant lower than the normal 40 MHz LHC frequency. Reasons for this are the placing of IP8 and the filling scheme of the LHC ring. The bunches are arranged in batches and these again in trains, between each are gaps. One reason for this is the controlling of the beam and another is the limitation of the injection kicker magnet rise times [56]. So the missing rate of $\approx 10\text{ MHz}$ are gap-gap or gap-bunch events where no collision can happen.

It is not possible to store all events which are coming from the detector. The size of a full reconstructed event is about 100 kB which leads to a total rate of 3 TB/s if every event is stored. Beside the fact that not all events are relevant for the physics program it is practically not possible to store everything. In an average year the LHCb raw data is in the order of 20 EB while the storage capacity is only tens of PB per year. As the LHCb physics is targeting rare processes a reduction by random choice is not wanted. Therefore, the task of the LHCb trigger is the reduction of the output data-rate while selecting events of interest.

One essential element of the upgrade is the switch to a software only trigger system with a trigger-less detector readout at 40 MHz. In the previous version the trigger had a hardware stage and a follow up software stage. [Figure 3.11](#) shows a sketch of the upgrade trigger scheme. The actual HLT implementation is split

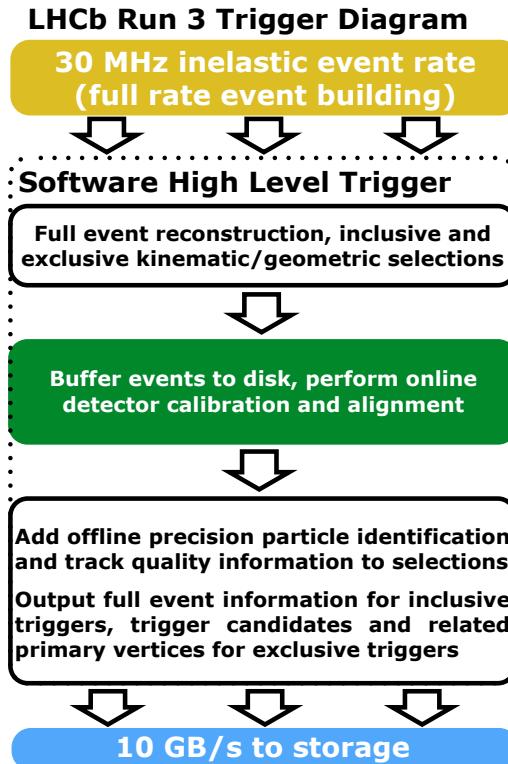


Figure 3.11 – Diagram of the LHCb upgrade trigger. The pure software trigger is split in two steps. First a quick event reconstruction and a reduction to 1 MHz. The second step is the high precision reconstruction after the detector calibration is available. [57]

in two parts HLT1 and HLT2. In the first stage the data is reduced to 1 Tbit s^{-1} respectively 1 MHz. This data is stored in a buffer until the full reconstruction can be done. The size of this buffer storage will be in the order of 20 PB. For the full reconstruction all conditions of the detector in the corresponding run needs to be known. Therefore, the real-time calibration is done, based on the taken data the needed parameters are generated. With these information the HLT2 will process the events from the buffer. The data-rate which is saved finally is in the range of 2 to 10 GB/s, depending on the actual settings of the trigger the so-called trigger configuration key (tck). In this configuration the actual values for the high level feature, which are used in the selection progress, are defined. These can be for example particle IDs, tracks with high momentum or reconstructed PV. The upper limit of 10 GB/s results from the limited offline storage resources [58].

In Figure 3.12 the upgrade reconstruction sequence, as described in the trigger TDR, is shown. The diagram is split in two areas, left the offline part HLT2 and on the right online which is HLT1. First of all the VELO tracks are reconstructed. Then the VELO tracks are extended to the UT under some kinematic conditions. Within the forward tracking the tracks are extended to the SciFi. Based on the

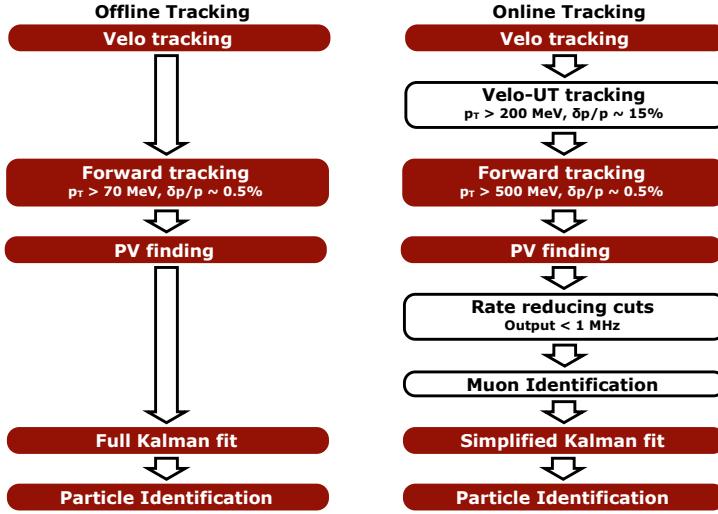


Figure 3.12 – Sequence within the HLT1 (right) and HLT2 (left) [33].

found tracks the PVs are reconstructed. In the end a simplified kalman fit is used and PID information are generated. In the offline step basically the same things happen, but due to the higher available computing time the precision is increased. One necessity for this higher precision is the detector calibration which is done between the two steps. The calibration and the alignment of the detector are done in so-called real time [59]. This real time refers to the time between event recording and final trigger decision in HLT2. The time and the needed statistics are reached due to a sufficient disk buffer size. This concept was already used during the former data recording periods of the detector. Because of the important role of the data processing LHCb has established the real time analysis project (RTA).

This section is completed with an important remark. The concept of the new pure software trigger was introduced. It might appear that this is an easy feasible project, but the opposite is the case. The idea of the software trigger was formulated in the trigger and online TDR published in the beginning of 2014 [33]. It was predicted, that the planned budget for the farm will allow to buy sufficient computing power. For the prediction of the future CPU performance is based on Moore's law. This law predicts the transistor density grow and the density is doubled about every two years. In terms of general server performance at equal cost, the performance growth-rate is estimated to 25 % per year [60]. For the LHCb application of the trigger a factor of about 16 was predicted. Hereby the performance is compared to the old HLT-server-nodes from 2010 and the acquisition of the new nodes is foreseen in 2019. The initial plan of the trigger sequence did also include a low level trigger stage within the software trigger. This stage should reduce the incoming data like the former hardware trigger. Due to the additional coarse cuts the efficiency of the trigger was reduced. When it become feasible to run the trigger without this stage it was removed.

4 The Scintillating Fibre Tracker

As mentioned in the description of the upgrade detector, cf. [chapter 3](#), the SciFi Tracker is one part of the LHCb tracking system. In total, the SciFi will cover an area of 360 m^2 and can be read out at a rate of 40 MHz. In this chapter, the detector components are described in detail. For the ageing studies in [chapter 5](#) the basic technology of the detector and the effects inside the material are presented. Also, the readout electronics and the data handling are shown, as a prerequisite for the trigger studies in [section 7.2](#).

4.1 The Scintillating Fibre

The scintillating fibre used for the SciFi Tracker is provided by the company Kuraray and has the product name SCSF-78MJ [61]. It has a diameter of $250\text{ }\mu\text{m}$ and is made of a polystyrene core with added dye and wavelength shifter. Two claddings enclose the core with lower refraction index to enable total reflection. A sketch of the fibre is shown in [Figure 4.1](#). Illustrated are the claddings and a traversing particle which produces light inside the fibre.

4.1.1 Scintillating Mechanism

The following description of the scintillating mechanism is a simplified representation; more details can be found in [62]. When a particle crosses the core material energy is deposited there. Actually, the energy is absorbed by the conjugated π orbitals of the benzene rings of the polystyrene. Through such an excited state polystyrene can emit scintillating light. However, the typical decay time of this process is around 20 ns. This is too slow when it is compared to the bunch crossing frequency of 40 MHz. Most of the produced light would be detected during later bunch crossings. The light travel time in the fibre is also a non negligible factor, assuming the speed of light the photons travel approximately 30 cm/ns.

To speed up the scintillation process the polystyrene core is doped with a dye. In case of our fibre the energy is transferred to a dye called p-Terphenyl via the non-radiative so-called Förster transfer [63]. The light yield of a scintillating fibre depends on the difference between the emission and absorption spectra of the used materials. If the overlap is too big, all the produced photons get absorbed again. This reabsorbed light will be re-emitted isotopically again where only a certain fraction will match the angle conditions, cf. [Equation 4.2](#). Also the quantum efficiency of polystyrene is low, this means that ratio of created photons and deposited energy is small. Because of this two facts all light or energy would be

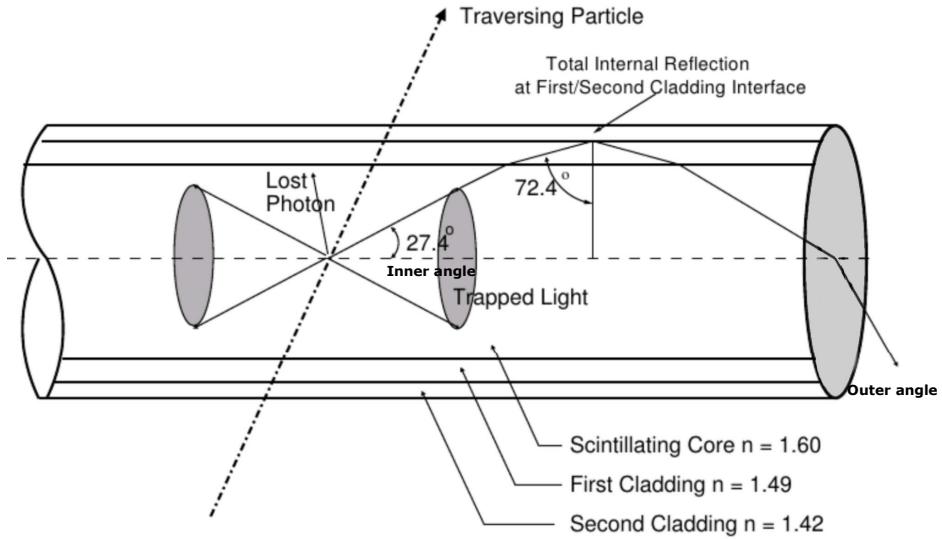


Figure 4.1 – Particle traversing a fibre. Whether produced photons are trapped or not depends on their angle. Total reflection is possible due to the decreasing refraction indices of the material layers. [35]

lost soon. To ensure the needed gap between the spectra, the wavelength shifter TPB (tetraphenylbutadiene) is added. The absorption and emission spectra of the scintillator, polystyrene, the dye, p-Terphenyl, and the wavelength shifter, TPB, are shown in [Figure 4.2](#). An additional benefit is an increased photon detection efficiency. This results from the fact that the spectrum was shifted to a region where the photon detector has a higher photon detection efficiency, cf. [subsection 4.2.1](#).

4.1.2 Light Guidance

In the previous part the process of the light production was explained. Hereinafter the light travel inside the fibre is explained. Not all of the isotropically emitted photons are trapped inside the fibre because the conditions for total reflection are not always fulfilled. As the diameter of the fibre is large compared to the wavelength of the produced photons geometrical optics is used to approximate the photon propagation. When n_1 is the refractive index from the inner material and n_2 the corresponding of the outer most cladding, than the relation between the incidence angle α_1 , and the refraction angle α_2 is defined by Snell's law as

$$\sin(\alpha_1) \cdot n_1 = \sin(\alpha_2) \cdot n_2. \quad (4.1)$$

Than the maximal angle α_c where total internal reflection is possible can be calculated as

$$\alpha_c = \arcsin\left(\frac{n_2}{n_1}\right) \approx 21^\circ. \quad (4.2)$$

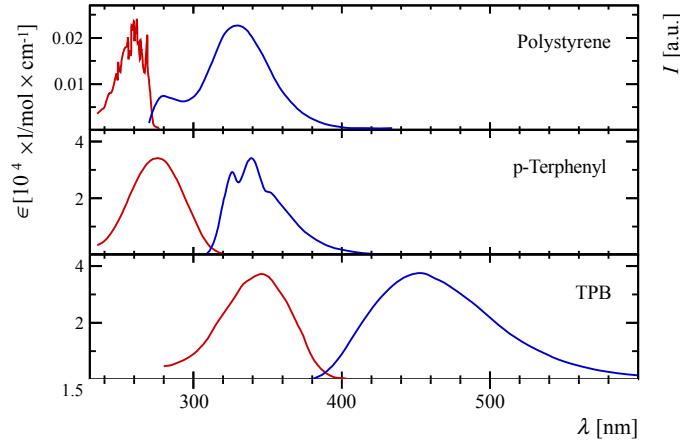


Figure 4.2 – Absorption and emission spectra of the components of the used fibre materials. The red and blue line corresponds to the decadic molar extinction coefficient ϵ and the emission intensity, respectively. Taken and modified from [64].

At this point it is important to recap that the outer cladding has a lower refractive index than the inner one. Because of this and the assumption that the surfaces normal's are parallel, the inner cladding can be ignored for the calculation of the α_c . In case of the used fibre the angle is 27° . [65]

With this value the caption efficiency of the isotropically produced photons can be determined. At the point of the photon production, the fibre is split vertically into two hemispheres. Every hemisphere represents one direction in the fibre. In the later detector, these will be the direct and reflected light. In every hemisphere about 5 % of the photons are trapped inside the fibre due to the total reflection. But not all of these reach an end of the fibre, different processes lead to photon losses inside the fibre, these are explained in the following.

4.1.3 Light Loss

Until this point an ideal fibre in terms of material and geometric properties was assumed, but for several reasons this is insufficient. The shape of the fibre, respectively the surfaces of the different layers, is not perfectly cylindrical. So the angle between photon and surface can vary which might allow a photon to leave the fibre. Also impurities in the fibre can absorb photons. These are the extrinsic effects but there are also the intrinsic ones.

There is a certain amount of absorption by the basic materials. In addition to this there is Rayleigh scattering, which cannot be avoided. So far the wavelength of the photons was ignored, but of course the amount of scattering is correlated to the

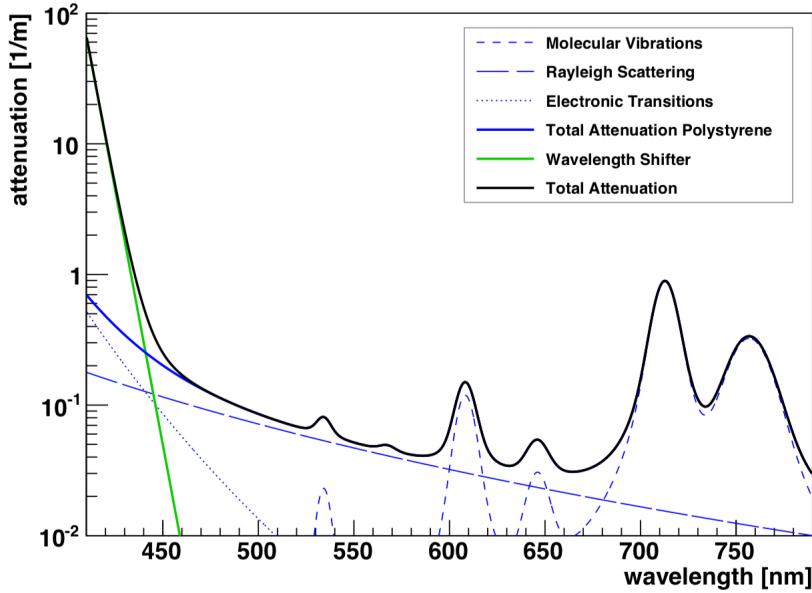


Figure 4.3 – Wavelength dependent attenuation coefficient of the scintillating fibre. [65]

wavelength. As well the refraction index of a material is wavelength dependent. The light loss for light travelling through the fibre can be described by an exponential function with an attenuation coefficient α . This coefficient is plotted in Figure 4.3 as a function of the wavelength.

The light intensity in the fibre after a certain fibre length x is described by

$$I(x) = I_0 \cdot \exp(-\alpha \cdot x) \quad (4.3)$$

where I_0 is the initial light intensity and α the effective attenuation coefficient. To increase the speed of the needed simulation an effective model is used, more details on this in the corresponding section 4.5. This effective model is a necessity for the later performed simulations, more details concerning this are presented in section 4.5.

4.2 Detector Layout

The primary material of this detector is a scintillating fibre with a diameter of 250 μm . Further details on the fibre are given in section 4.1. The fibre is wrapped into a six-layer fibre-mat with a length of 2500 mm and a width of 520 mm. At one end of the mat a mirror is applied to increase the light yield. At the other end, the light can leave the fibre and is detected by a SiPM array, see subsection 4.2.1. The next bigger unit in the detector is a so-called module. It is composed of 8 mats, four mats side by side and two mats on top of each other to cover the acceptance of 5 meters. Thereby the side with the readout is on top and bottom of the module,

respectively. Consequently, the mirror sides are in the middle of the module. An essential feature of the module is the structural stiffness. This stiffness is needed for a proper knowledge about the fibre position in the detector coordinate system. The shape of the modules is a major factor for the position resolution of the tracker. The modules are mounted in a so-called C-Frame. The overall detector is split by the beam-pipe in the two regions *A* and *C*. Two layers per side and station are mounted in one C-Frame. This mounting includes all connections like power supply, network and dry gas. The frames slide into the old support structure of the outer tracker.

For the total Scifi tracker, 12 C-Frames are needed. Compared with the previous tracking station, the general shape is not changed. The tracker consists of three stations, each with four layers. Thereby the layers are arranged in the so-called $x - u - v - x$ configuration. This means that outer layers are parallel to the y axis and the inner ones are tilted by $\pm 5^\circ$ and named stereo layer. This leads to an asymmetric hit resolution, in x it is better than in y . The particular arrangement is motivated by the main task of the SciFi tracker, the momentum measurement. Because in LHCb a dipole magnet is used, charged particles are only bent in one plane. In this case it is the $x - z$ -plane. Therefore, the resolution in x is more important than in the y -direction. The chosen angle is a compromise between 3D resolution and hit combination ambiguity. A greater angle would increase the overlap between the normal and the stereo layers. The 5° will lead to an overlap about 21 cm.

The third station, T3, is wider than the first two stations. T1 and T2 consist of 10 modules, while T3 has 12. This enables better tracking efficiencies for low momentum tracks, as they become more bend. [Figure 4.4](#) shows the three stations where the first one is illustrated with more details for the explanation of the SciFi numbering scheme. The naming of the elements and their range is listed below where the elements are sorted from the large to small.

- **Station:** **T1, T2, T3** increasing in positive z direction
- **Layer:** **X1, U, V, X2**
- **Quarter:** **Q0, Q1, Q2, Q3** start at negative x and y increased in mathematical negative around the beam pipe.
- **Module:** **M0-M5** M0 is always closed to the beam pipe. T1 and T2 consist of only M0-M4
- **Mat:** **Mat0-Mat3** For the readout boxes on top of a module, looking towards the C frame, the mats are numbered from left to right.
- **SiPM:** **S0-S3** SiPMs number follows the mat direction
- **Channel:** **C0-C127** The channel number follows the SiPM

The information about the ranges of the different detector level elements is essential for the compression of the data during the readout, cf. [section 4.4](#).

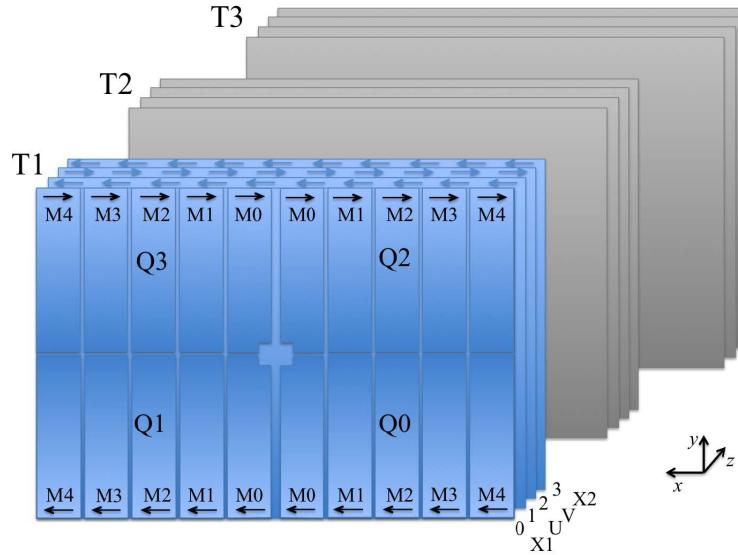


Figure 4.4 – Layout of the SciFi tracker. In the first of the three stations the underlying elements are labelled. These are the Quarter Q and the Modules M. The direction of the SiPM channel numbering is indicated by arrows. [66]

4.2.1 Light Detection

Within a typical event only a few photons, approx 10, reach the fibre end. Consequently, it is essential to maximise the photon detection efficiency to make a signal detection possible. In the SciFi Tracker silicon photo-multiplier (SiPM) are used for the photon detection. This technology fulfils all the requirements like 40 MHz readout and radiation hardness.

SiPMs are semiconductors which can detect single photons, and they have an internal gain of $\approx 10^6$. An additional benefit is the low bias voltage in the range of 25 V to 75 V. Before the usage of SiPM photo-multiplier tubes were used to detect single photons, which need high voltage of ≈ 1000 V. The SiPMs which are used for the photon detection in the SciFi are provided by the company Kuraray. A SiPM is an array of several avalanche photo diodes which are used in the Geiger mode. Such an array is shown in [Figure 4.5](#).

Single pixel are connected to a unique readout unit, a so-called channel. Here 4 times 16 pixel build a channel, with a pixel size of $62 \mu\text{m} \cdot 57 \mu\text{m}$. This leads to a channel widths of 250 μm including the support structure.

An illustration of the photon detection is shown in [Figure 4.6](#). A particle is crossing some scintillating fibres. While traversing the fibre material photons are created, coloured green. Photons reaching the fibre end are coloured black. Some of these hit SiPM pixels which are coloured yellow then. The signal amplitude of the three channel is shown at top. The further steps of the cluster position calculation which are done by a clustering algorithm are explained in [section 4.3](#).

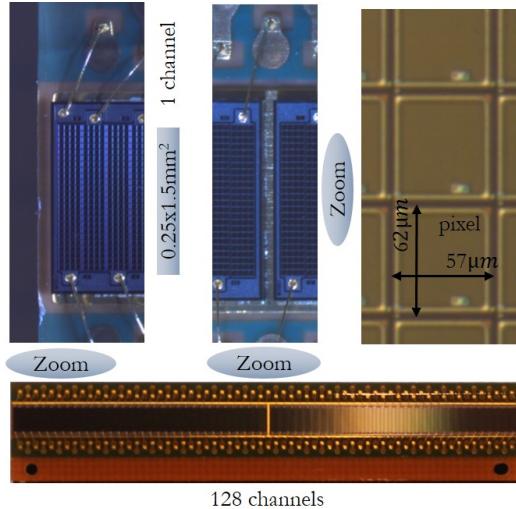


Figure 4.5 – In the bottom a 128 (2 times 96) channel SiPM is shown. In the top row different parts are magnified. left: SiPM channel. middle: Gap after the first 96 channel. right: single pixel within a channel. [67]

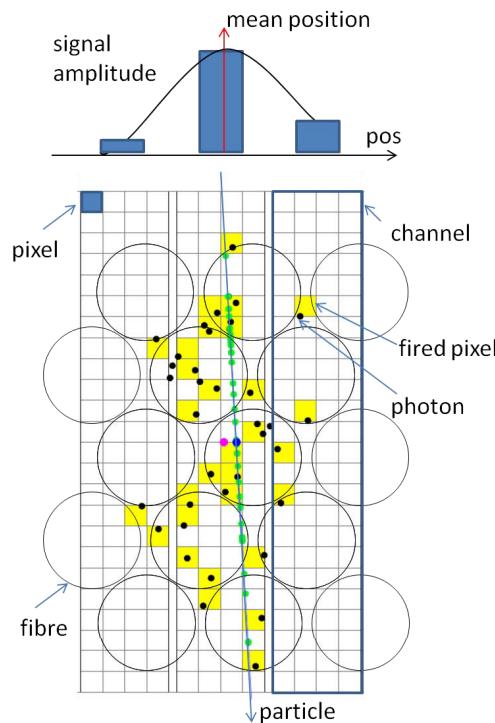


Figure 4.6 – Particle traversing a Scifi-mat with five layers (old design). Three SiPM channels with their pixels are overlaid by the cross-section of the fibre ends. Produced photons are green and detected black. Fired pixels are coloured yellow. A histogram of the signal amplitude per channel is shown on top. [35]

As mentioned in [subsection 4.1.1](#) the photon detection efficiency of the SiPM is wavelength dependent. But the chosen SiPM version has its maximal detection efficiency near the region of scintillating light output spectrum. Furthermore, the wavelength dependence of the SiPM has a significant impact on the simulation of the SciFi. For an accurate description of the light detection, all previous simulation steps need to be wavelength dependent. The electronics which receive the signal of the SiPM are described in the next section.

Over the lifetime of the detector the SiPM will be effected by radiation damages. This effects will lead to a significantly increased dark count rate. Nevertheless, the effect is reduced by an operation at low temperature. Therefore, the operational temperature for the SiPMs is -40°C . In total the tracker consists of 4096 SiPM arrays which corresponds to 524k readout channels. The readout and the data handling of these is explained in the following part.

4.2.2 SciFi Readout Electronics

The data of the SciFi tracker is managed by custom made electronic boards. There are two different boards, first the Front-End Electronics (FE) which is located directly at the detector and second the Back-End Electronic (BE) placed in the data centre. The connection between the two sides is done via optical links.

Front-End Electronics

Under average data-taking conditions only a small fraction of channels will produce a signal. The average number of total hits is around 8000, with an average cluster width of three only $\approx 5\%$ of the 524k channels receive a signal. To reduce the data-rate which comes from the detector, only the information from triggered channel is sent out. This is a common way of data handling and is called zero suppression.

The just mentioned FE is located in a readout-box, a sketch of the board is shown in [Figure 4.7](#). At the bottom part the SiPMs are located, coloured in orange, one unit represents 128 channels. From the SiPM the signal is passed to the so-called PACIFIC¹ board. Here the signal is integrated over 25 ns and the resulting charge is digitised by a 2 bit ADC. In the next step the clustering takes place on a FPGA (blue), details to this process can be found in [section 4.3](#). After the signal clusters are built, they are packed and sent out to the BE via optical links (red). The total control of all connected parts is done by the Master Board. It handles control commands, which are received by the timing and fast control system (TFC).

Beside the FE, cooling and vacuum systems are part of the readout-box. The cooling is needed for the SiPMs, which have to be operated at -40°C to reduce the dark count rate to a minimum cf [subsection 4.2.1](#). The vacuum is needed as an isolation for a proper cooling.

¹Low Power ASIC for the SCIntillating FIbre TrACker readout

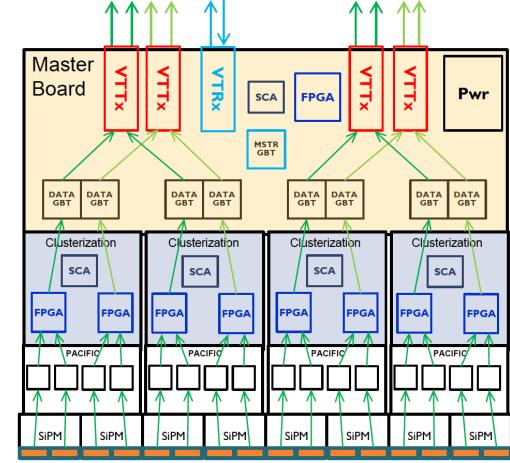


Figure 4.7 – Scheme of the readout board. The light detecting SiPMs are located at the bottom and their signal is digitised by the PACIFIC. The clustering algorithm is implemented on a FPGA. The found clusters are packed and then sent to the BE electronic via optical links. [35]

Back-End Electronics

Data from the detector is received by the BE in the data centre. At this point the focus is on the receiver board and not the centre itself as it already was introduced in [section 3.5](#). Because the board is connected via PCIe to the server of the online farm and the detector uses a 40 MHz trigger-less readout, the board is named PCIe40 and the integrated FPGA TELL40. All sub-detector send their data to a PCIe40 board. Therefore, it has a universal skeleton and custom parts for every sub-detector.

4.3 The Clustering Algorithm

The data of the SiPM arrays is processed for two major reasons. First the reduction of background hits and second for a higher resolution. The data rate of the FE/BE link is limited, therefore zero suppression is utilised. This means that only the data from channels, where a signal respectively cluster was, is sent to the BE. Also for the subsequent tracking it is important to reduce the amount of background hits, referred to as ghost hits. These enlarge the amount of combinatorial possible tracks which slows down the computational operations. It is also possible to create ghost tracks. These are reconstructed tracks which are not originating from any real track, cf. [Equation 5.4](#). Thermal noise is the most common source of the dark count rate (DCR). Even though the sensors are cooled to -40°C they have a non negligible DCR. Another source are photons which come from neighbouring parts of the SiPM, so-called cross-talk. The DCR per channel is expected to be about 13 MHz after 5 years of data-taking, which corresponds to a neutron fluence of

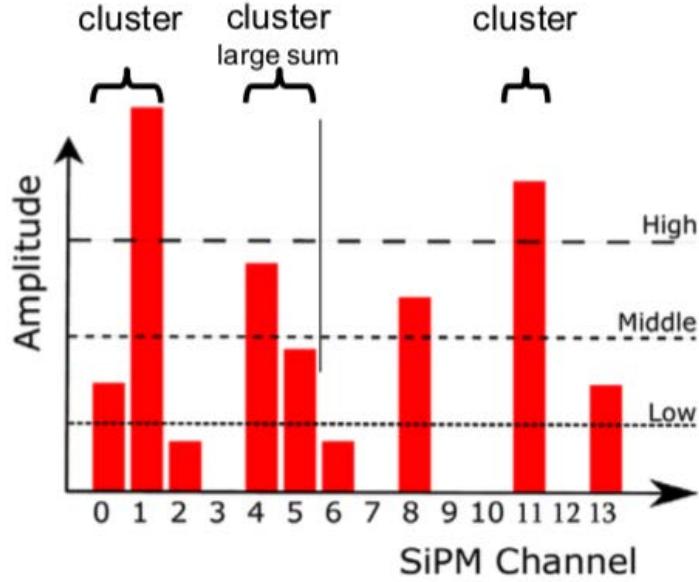


Figure 4.8 – Illustration of the clustering. The signal amplitude for different SiPM channel is given. [68]

$6 \cdot 10^{11}$ neq/cm². Hereby the damage is proportional

The clustering algorithm on the FE board filters the signal so that in the best case only hits from real tracks are sent out to the BE. The procedure of the algorithm is shown in Figure 4.8 and the information are taken from [68].

In the Pacific chip the data is digitised into a 2-bit value. Three different thresholds for the detected light amount can be set in the firmware of the FE, at the moment the values are 1.5, 2.5 and 4.5 photoelectrons. In reversed order these are the high-, seed- and low-threshold. The low-threshold is just above the thermal noise so that the edges of a cluster can be still separated from background. The amount of channel contributing to a cluster is the cluster size. Found clusters can have different sizes, the mean is 2.5. This value is determined from simulation. For the particles of interest the cluster size can go up to 4, even bigger values are possible. The special handling of so-called big clusters will be explained later. The clustering is split in three steps. First the cluster candidates are defined, therefore a channel needs a signal higher than the seed threshold. In the second step all neighbouring channels with a signal higher than the low threshold are considered to belong to the same cluster. In the third and last step the sum of signals is compared to the high threshold. The cluster candidate is rejected when the sum is lower than the high threshold.

The mean cluster position is calculated as a weighted mean by

$$\Sigma(\text{position}_i \times \text{weight}_i) / \Sigma(\text{weight}_i)$$

where the weight is the value of the reached threshold. This mean cluster position

is decoded into a 8-bit number. The first 7 bit represent the channel number and the last one defines the inter-channel position. When this bit is 0, the mean is in the range of [-0.25, 0.25]. In the other case when the bit is 1 it is in the range of [0.25, 0.75]. Clusters with a size bigger than 4 are called big, these are handled in a special way, details to this will be explained in the decoding part [section 4.6](#).

4.4 SciFi Tracker Raw Banks

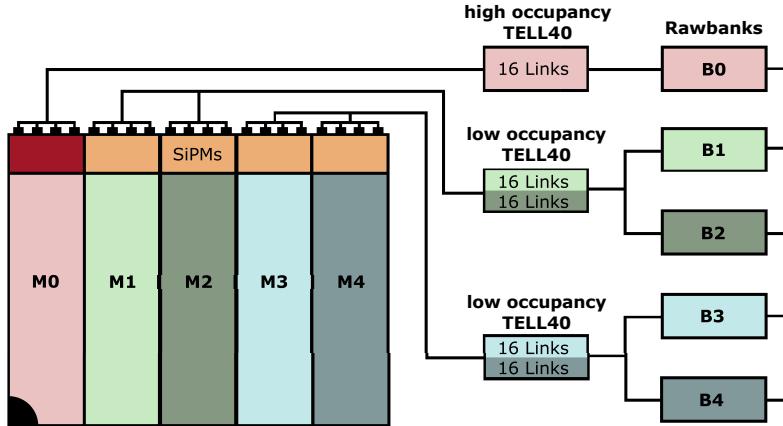


Figure 4.9 – Illustration of a SciFi tracker quadrant from T1 and the connection of the FE with the raw banks [69].

The data which is sent from the FE to the BE via optical links is grouped in so-called raw banks. Due to the significant gradient of occupation in the SciFi tracker different data-formats are used. In the inner modules of the stations the hit density is the highest and the data rate is more than twice times as the rate of the outer most module. A sketch of one detector quarter is shown in [Figure 4.9](#) the receiving TELL40 boards are split in high and low occupancy.

This differentiation is needed for a proper cost-performance ratio. Therefore, only the minimum amount of needed links and raw banks is used. For the 16 links in the high region a fixed header variable data format is used. The fluctuation in the number of hits is very high here, through this the length of the data string is variable. The header of the data is fixed, here global information like the corresponding bunch-crossing ID are stored. In the low occupancy region the length of the data-string is fixed as well as the header. With this fixed header fixed data format it is possible to use the bandwidth of the link more effective. The number of links to a single TELL40 is limited by the computation capability of the FPGA. In the low occupancy region 32 links per TELL40 are used. For the high occupancy region the number is reduced to 16. This reduction is necessary because of the more complex handling of the variable data format and the higher data rate.

The TELL40 boards are located in the EB server, the data of the hits needs to be provided to the HLT, this is done via the raw banks. One high occupancy TELL40 board translate the data into one raw bank. The same ratio, 16 links per raw bank, is used in the low occupancy region. This means that there will be 5 raw banks per quadrant. As the SciFi tracker has 12 layer or 48 quadrants the total amount of raw banks is 240.

The clusters are stored in 16-bit words, as those can be aligned fast in the TELL40 [70], an illustration of this format is shown in [Figure 4.10](#). For the numbering of the optical link 6 bit are used and for the channel additional 7 bit. There is also one fraction bit, a size flag and one bit which is unused at the moment.

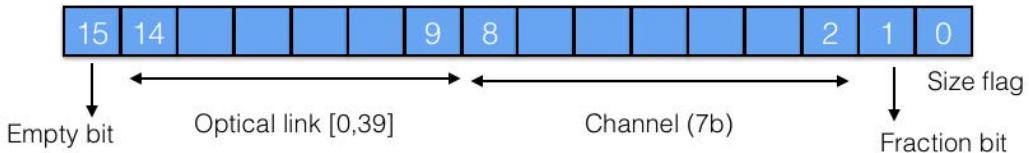


Figure 4.10 – Format of the cluster data in the 16-bit word [68].

4.5 Simulation of the SciFi Tracker

In this section, the simulations of the SciFi tracker at different levels of detail are presented. The most detailed or lowest level is the simulation of a single fibre. At the roughest scaling, the whole response of the complete detector is simulated. Hereby the results from the single fibre simulation are used to simplify the fibre mat simulation. Also, some major implications of the simulation results are presented.

A single fibre simulation was implemented with the Geant4 toolkit [71]. In this simulation, photons can be created at every point of the fibre. All the relevant material parameters can be set flexible. Due to this, the simulation can be tuned and checked by single fibre measurements in the lab [65]. This simulation also calculates the time after which the photons arrive at the fibre end, so the signal shaping of the readout can be optimised [72].

The next higher level is the simulation of a fibre mat. Fibres are put next to each other based on the single fibre simulation. Now it is possible to study effects like cross-talk between fibres. Photons which leave a fibre can enter another one and fulfil the conditions of total reflection there. It was decided to add titanium dioxide to the glue that holds the fibres, to reduce this effect [73, 74]. Photons can use different paths to reach the fibre end. Some stay in the core all the time other are trapped in the cladding still others use the total fibre diameter. The used path depends on the initial angle and the wavelength. A wavelength-dependent reflection loss coefficient was found. This value represents the probability to lose a photon at a single reflection. With this value the number of photons which reach the fibre end is calculated in later analysis steps. Based on all this an effective

simulation was developed, the light yield attenuation map [75, 76]. This approach of simulation is explained in detail in [section 5.3](#). The effective model is needed for the full simulation of pp collisions in the LHCb-detector. A Geant4 simulation for every fibre where energy gets deposited would be too slow. The tool for the attenuation maps is implemented in BOOLE, like the rest of framework for the simulation of the detector response. The maps itself, with the attenuation values, are stored in the standard LHCb conditions database.

From the Gauss Simulation, the previous step in the LHCb simulation chain, the energy deposit of each MC-Hits is known. With a conversion factor, which is known from test beam campaigns, the amount of produced photons, the light yield, in the fibre is calculated. The attenuation map provides the factor between photons produced along the track of the primary particle and photons which are detected by the SiPM. The maps are used as a lookup table, for every position (x, y) in the detector an attenuation factor is calculated by the single fibre simulation. This factor includes the attenuation from the point of excitation to the SiPM. For the indirect light, which first goes to the mirror, a dedicated map is produced. This is needed as the spectrum is shifted due to the longer path lengths of the photons and they have a significant later arrival time. With this extra map it will also be possible to change the mirror reflectivity to improve the quality of the performance simulations. For example it would be possible to use the values for the reflectivity which are measured during the fibre mat production.

Because of the exponential function, cf. [Equation 4.3](#), which is used for the light loss model, attenuation coefficients from different sources can be summed up. One example of this is the independent treatment of direct light and light, which first travels in the direction of the mirror and then gets reflected there. Another example is the later following study about radiation damage and ageing effect, cf. [chapter 6](#). These are also handled like additional sources of attenuation.

4.6 SciFi Decoding

In this section the basic concept of the SciFi hit decoding is presented. Three different versions are introduced. The decoding for all version will be converted in [section 4.6](#) into the new GPU trigger framework.

The cluster data from the FE which is packed in raw banks needs to be decoded. This means that the cluster information is transformed to hits within the LHCb coordinate system. All tracking algorithms in the trigger need this hit position information. The general concept of the decoding is explained on basis of the format version five (v5) in [subsection 4.6.1](#). Furthermore, the versions v4 and v6 are also relevant for this thesis. The differences will be explained in the corresponding parts [subsection 4.6.3](#) and [subsection 4.6.2](#).

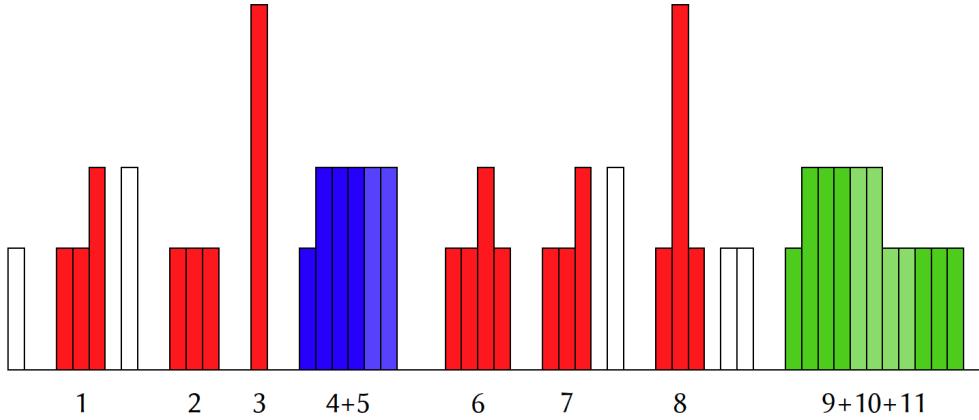


Figure 4.11 – Illustration of SiPM channel signals for clusters of different sizes. Normal clusters with a width up to 4 are coloured red. Big clusters are blue, up to 8 channel, and green more than 8. The different treatment of the big clusters is explained in the corresponding decoding sections. [68]

4.6.1 Decoding Version 5

The information to decode is the 16-bit cluster word, cf. [Figure 4.10](#). Here it does not matter if the cluster is coming from a high or low occupancy region, as the cluster format is the similar. From the 16-bit word 6 bits are used for the link-ID to the TELL40 board. A clear assignment of the 144 TELL40 boards respectively 240 raw banks to actual detector regions is necessary. This is done via the so-called ReadoutMap which is used in the FTReadoutTool. For every raw bank five values are stored. These are station, layer, quarter, first module and first mat. With these information the assignment is explicit.

In [Figure 4.11](#) eleven clusters with different sizes are illustrated. Clusters with a channel width up to 4 channel are red, for those the size bit is 0. The category of big clusters has two member, first clusters up to a width of 8 in blue and in green clusters bigger than that. For small clusters the decoding is a easy conversion of information due to the ReadoutMap. For the small, red, clusters the normal weighted mean position calculation is used. For the big clusters, green and red, the clustering algorithm uses the geometric mean as the cluster position. The fraction bit is 0 when the size is odd and 1 for even cluster sizes. For big clusters two different cases can take place. Fragmented clusters like the blue 4 and 5 are likely to belong to the same hit, this is assumed up to a cluster size of 8. This procedure is called merging. In the second case, cluster size > 8, only the first and the last cluster are stored, the middle ones are deleted and this information will be recovered in the decoding step. The decoder can recognise the case on the basis of the size flags. If the following cluster is not marked as a big cluster as well, it actually forms the rest frame or end of the previous cluster and will be merged cf. blue. In the other case, the next cluster is also flagged big, the distance

between these two is calculated by the difference of their channel IDs. Based on this distance modulo 4, the normal small cluster size, new cluster are created. Due to this the green cluster number 10 can be recovered.

4.6.2 Decoding Version 6

The decoding version six is very similar to v5 and will provide the same output. There is a difference in the format for the cluster which are at the end of large cluster. Another difference is the channel number in the end fragments, the last channel is stored and not the mean like in v5. As the fraction bit is not needed anymore it can be used for the marking of the end. All parts of a big cluster have the size flag 1 the begin is marked by fraction bit = 1 and the end with 0.

4.6.3 Decoding Version 4

For the version 4 the data is encoded in the same way as for v5. But the two features merging and recovering are not used. This means that some information are lost. Due to this fact the hit efficiency of the detector is decreased.

5 Simulation of the SciFi Over the Lifetime

In this chapter, the simulation of the LHCb SciFi tracker for the planned lifetime is presented. Therefore, the interaction of particles with matter is described in [section 5.1](#). Afterwards, the focus is on the two main processes, radiation damage and natural fibre ageing, which both lead to a light yield reduction of the scintillating fibres. The conditions cover different combined scenarios of detector lifetime and absorbed radiation.

First the basic effects of these processes are explained, cf. [section 5.2](#). Then the concept of the light yield attenuation maps is presented in [section 5.3](#). These maps are used to implement the effect of radiation in the SciFi simulation. This is followed by a description of the different simulation steps for the event sample generation within certain condition, cf. [section 5.4](#). Afterwards the different track types and the criteria for these are shown in [section 5.5](#). Here also efficiency parameters are defined. The evaluation of the data is presented in [chapter 6](#).

5.1 Particle Matter Interactions

Particles, charged or neutral, can only be sensed by interacting with matter. The presented information are based on the book [77]. The usual processes in a detector material are:

- ionisation and excitation by charged particles
- bremsstrahlung
- scattering and absorption of photons
- Cherenkov and transition radiation
- hadrons with nuclear matter (nuclear reactions)
- weak interactions

All of these processes are relevant for the overall LHCb detector. Nevertheless, in this section only the ionisation and excitation process is introduced, because this process is most relevant for the radiation damage and light production of the scintillating fibres of the SciFi tracker. Besides this it is relevant for the signal production in the BCM diamonds.

In general all mechanisms can occur when a particle traverse matter. Thereby the dominant process is defined by the particles energy. The momentum dependent stopping power for muons in copper is shown in Figure 5.1. In the high energy

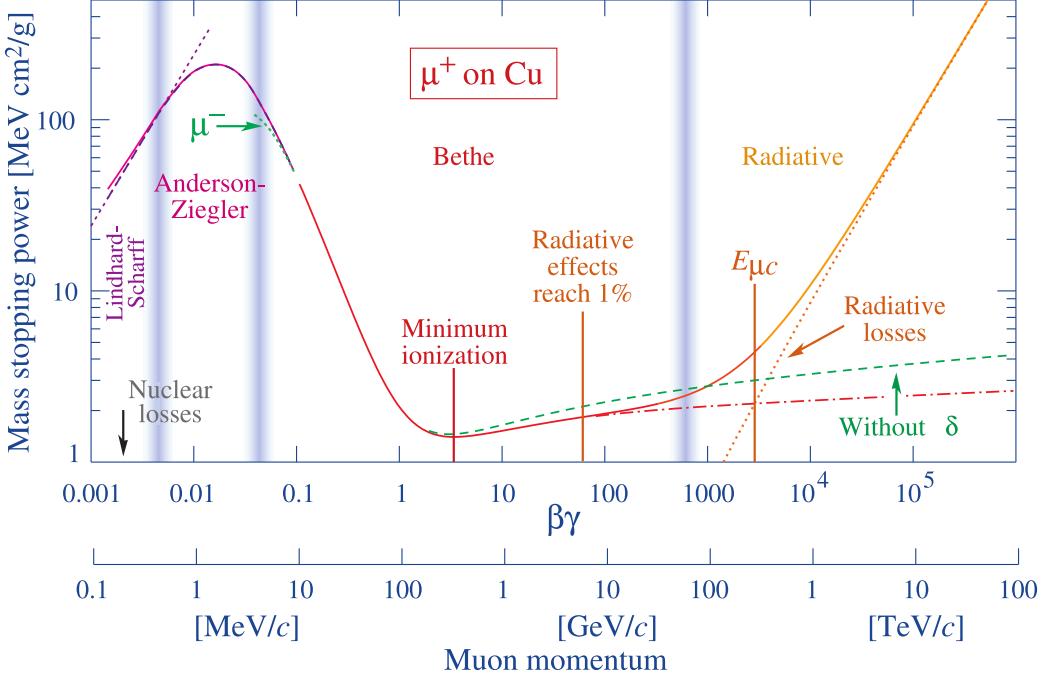


Figure 5.1 – Stopping power for positive muons in copper as a function of the momentum. The vertical bands separate different approximations. [78]

region above 100 GeV, radiative losses are the dominant process. Below this energy, the main process is ionisation. A minimum in the stopping power curve is visible at around 0.2 GeV. Particles with an energy in this region are called minimum ionising particle (MIP). This particular energy region is described by the Bethe approximation. For heavy particles the energy loss is described by the Bethe-Bloch formula

$$\left\langle -\frac{dE}{dx} \right\rangle = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] [79],$$

and the maximum single-collision energy transfer is

$$W_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e/M + (m_e/M)^2}.$$

Hereby the used variables are: Avogadro number N_A , classical electron radius r_e , electron mass m_e , incident particle charge number z , material atomic number Z , material atomic mass A , mean excitation energy I and the particle mass M . The term $\delta(\beta\gamma)$ is a correction for density effects, details towards this can be found in [78]. It can be seen that the ionisation strongly depends on the particle mass. As

mentioned before this approximation is only valid for heavy particles. In case of the SciFi, this is assumed to be correct, but not for the ${}^{90}\text{Sr}$ source. The electrons from the β decay are light particles in the sense of this formula. Besides the mass aspect, the fermion character of the electrons needs to be taken into account. Because the incoming electrons are indistinguishable from the material electrons additional modifications are necessary. The modified Bethe-Bloch formula for electrons is

$$\left\langle -\frac{dE}{dx} \right\rangle = 2\pi N_A r_e^2 m_e c^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{m_e c^2 \beta^2 \gamma^2 \{m_e c^2 (\gamma - 1)/2\}}{I^2} + (1 - \beta^2) - \frac{2\gamma - 1}{\gamma^2} \ln 2 + \frac{1}{8} \left(\frac{\gamma - 1}{\gamma} \right)^2 \delta(\beta\gamma) \right] \quad [79].$$

5.2 Light Yield Reduction Processes

In the following section, the theory behind the two light yield reduction processes are explained: The first is the reduction caused by radiation damage. The second is the natural ageing of the fibre.

5.2.1 Radiation Damage of Scintillating Fibres

Before the radiation damage effects are explained, a short overview about the radiation and their simulation in the LHCb SciFi tracker is given. For the simulation of the expected dose in the detector the FLUKA package is used [80, 81]. Within FLUKA, different types of radiation are taken into account. The types of fluences are neutron, photon, electron/positron and charged hadrons. With a scoring approach all information can be combined. The LHCb simulation was cross checked and tuned with dosimetry measurements in the cavern [82]. An example distribution of the expected dose in the SciFi tracker is shown in Figure 5.2. The calculations are done for a $x-y$ plane at a z position of 920 cm to 940 cm. The black rectangles which overlay the map indicate the module positions. The simulation was done for an integrated luminosity of 50 fb^{-1} . As expected the radiation is highest in the centre part around the beam-pipe. From here the profile is almost symmetric but in x direction the gradient is smaller than in y . This is due to the bending direction of the magnet. In the acceptance of the tracker the dose ranges from 0.03 kGy to 20 kGy. This map shows the total ionising dose after 50 fb^{-1} .

Because of the long experience with the FLUKA package the simulation of the expected dose is quite precise, nevertheless an additional safety factor of two is common. In contrast to the FLUKA simulation the radiation damage simulation of scintillating fibres is new. These fibres were never used before in such a radiation environment. Within the SciFi collaboration many studies on the radiation damage model were done [64, 83]

The actual damage from ionising radiation can occur in different ways. Optical changes in the base material reduce the light output [84]. Also new centres of

5 Simulation of the SciFi Over the Lifetime

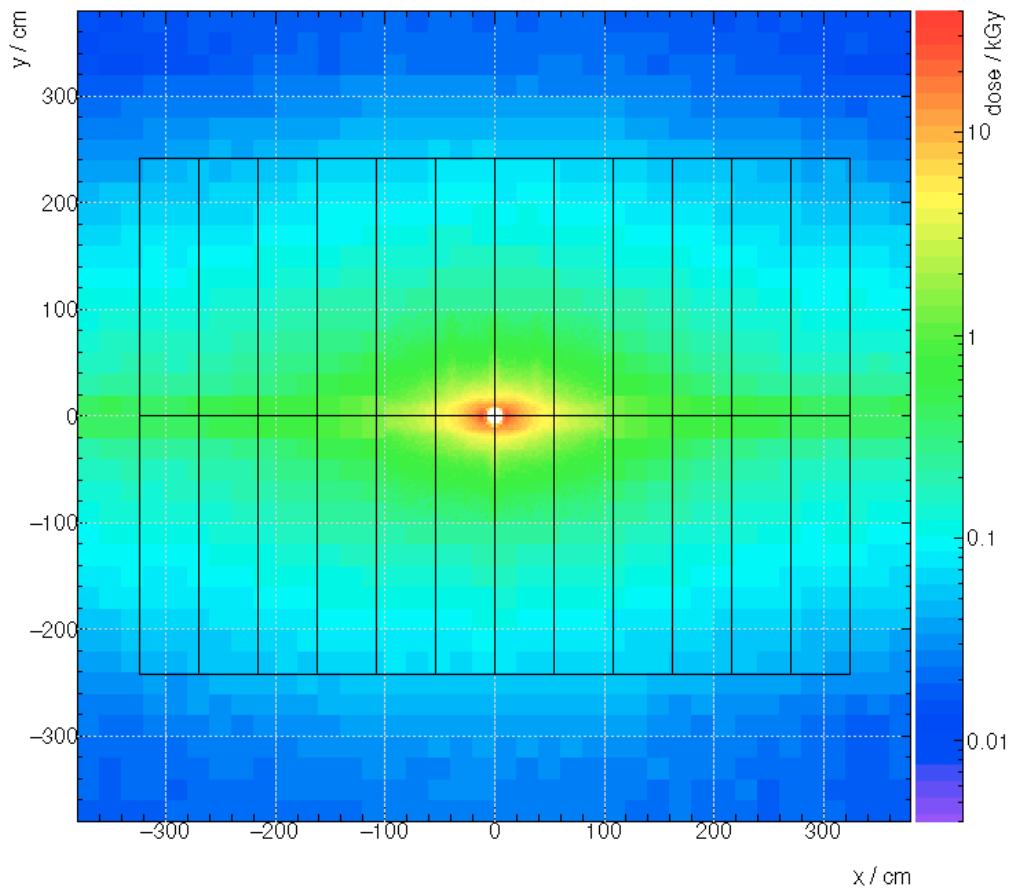


Figure 5.2 – FLUKA simulated dose map of a SciFi $x - y$ plane positioned at $z = 920 \text{ cm}$ to 940 cm . [76]

absorption or scattering due to the creation of free radicals [85]. These radicals have free valence electrons which can absorb light in the blue and UV region. This region of the spectrum is crucial for the light yield as the transfer from the dye to the wavelength shifter is in the UV region and a major fraction of the actual light is blue.

As the dose profile in the SciFi tracker is unique and accumulated over a long time period of 10 years, studies under similar conditions are not possible. Therefore, radiation studies with a broad range of dose and dose rate were performed during the development of the SciFi project. One result of this is a dose rate dependent scaling of dose and the light yield reduction. Two models were found by the collaboration to describe the radiation damage. First a power law model and second a linear model [76]. The attenuation α scales like

$$\begin{aligned}\alpha_{\text{pow}} &\propto 0.4 \cdot D^{0.8}, \\ \alpha_{\text{lin}} &\propto 0.38 \cdot D,\end{aligned}$$

where D is the dose in kGy. Although the dose rate dependency on the damage is not fully understood a partial explanation is possible: Some healing processes are possible inside the fibre. Oxygen can bond to the free radicals, mitigating their negative effect on the attenuation length. This healing needs some time and a certain oxygen concentration within the fibre. It is possible that additional oxygen diffuses into the fibre material, but at high dose rates the diffusion process is too slow. For the calculation of the later used attenuation maps the linear model is used, as this provides a worst case estimation. Another result is that the radiation is not affecting the intrinsic photon production. Therefore, this factor will not be changed within the following studies.

5.2.2 Natural Fibre Ageing

Even in an environment without ionising radiation the light yield of the scintillating fibres decreases over time. This is a natural effect and the impact is small, but over the foreseen lifetime of 10 years becomes a relevant factor.

The estimated lifetime of 10 years represents only the time of operation. But the fibre mats are produced a certain time before the data taking period, Run III, starts. Most fibre mats were produced in 2016, which is four years earlier than the planned starting time of Run III. Due to the corona crisis the time between production and start of usage increases additionally.

Chemical and physical structure damaging processes occur in the polymeric base material of the fibre over time [86]. The manifestations are minute cracks and dotty defects. In case of the fibres of the SciFi tracker the base material is PS. One major process here is the production of tertiary methylbenzene radicals. In the chemical reaction the radical is created by tearing off the tertiary atom of hydrogen.

Factors which influence the natural ageing are temperature, UV light, oxygen and the humidity. All these factors are expected to be sufficiently constant during

the time of operation. The mats are packed light tight in the modules. Only via the light injection system it is possible to couple light into the fibre. This system is used for the calibration of the system, but the usage is limited and the amount of light is not high. The temperature in the cavern is regulated to a constant value and the humidity underlays the usual seasonal fluctuations but both should not effect the mats inside the modules.

Due to measurements by the collaboration a linear model was found. The attenuation by ageing is 1.4% per year [87].

5.3 Light Yield Attenuation Maps

In this section, the concept of the so-called light yield attenuation maps (LYAM) is introduced. The LYAM is a tool used within the simulation of the SciFi tracker to predict the light output after a certain energy deposit within the fibre. Originally the tool was developed to speed up the general simulation, cf. [section 4.5](#). The information about the LYAMs are based on [76, 88, 73].

When a particle in the simulation deposits energy in a fibre, then the energy amount and the position (x, y) are known. The deposited energy is transformed to a number of photons and the conversion factor is about 8000 photons per MeV [67]. The reduction or attenuation of this photon number is looked up in the LYAM. The attenuation takes various effects into account. For example absorption, scattering and reflection losses in the fibre. Also the effect of the different wavelengths and the corresponding photon detection efficiency of the SiPM are included. At the end the tool provides a number of the detected photons, which is the light yield. This procedure is fast, as the detailed run-time intensive calculations are done beforehand. Therefore, the single fibre simulation of the previous chapter was used. For the (x, y) positions in the lookup table a binning is needed. The bin size is set to 2.5 cm \times 2.5 cm. These value was chosen because the same is used for the inner part of the FLUKA map, cf. [Figure 5.2](#). This leads to 120 bins in the x -direction and 101 bins in the y -direction. Because of the SiPM arrival-time difference of direct and indirect light, two different maps are used, one for the direct photons and another one for the photons which are reflected at the mirror. Both maps are plotted in [Figure 5.3](#), they show the attenuation coefficients for the start of the data taking period. Displayed is only one quarter (Q2) of a station. This reduction, to simulate only one quarter to safe computing time, was used because the detector and the radiation profile are symmetric around the beam pipe. Because of the beam pipe shorter mats are used within module zero (M0). This cutout is visible in the lower left edges in the LYAMs.

At this stage the map contains all the basic effects of attenuation in the fibre mat and the detection efficiency at the SiPMs. The expected number of detected photons \bar{n} is calculated by

$$\bar{n}_{\text{dir/refl}} = k \cdot \alpha_{\text{dir/refl}} \cdot E_{\text{hit}} \quad (5.1)$$

5.3 Light Yield Attenuation Maps

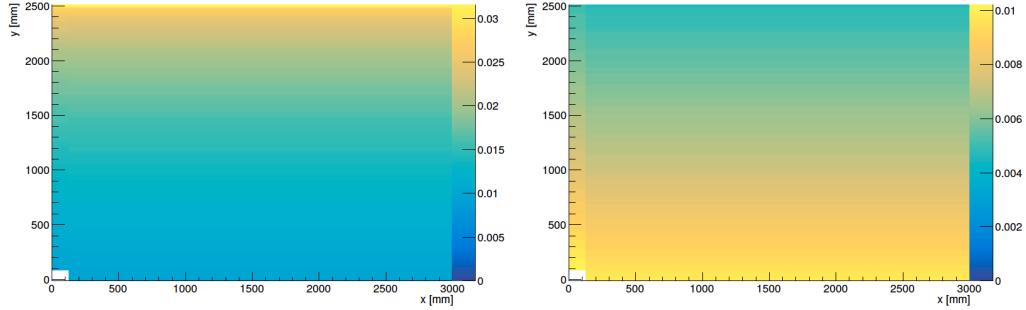


Figure 5.3 – Attenuation coefficient within the LYAM for direct light (left) and reflected (right) [76].

where k is the calibration factor, 6400 photons per MeV, and E_{hit} the energy deposited by the hit. The number of direct and reflected photons is calculated independently with the corresponding coefficient α_{dir} and α_{ref} respectively.

Now the integration of the radiation damage into the LYAM is described briefly. Single fibre simulations were performed whereby FLUKA maps are added. Due to the radiation damage in the fibre the light output is reduced. Here the linear radiation damage model is used, cf. subsection 5.2.1. For the calculation of a LYAM after a certain dose 12120 ($120 \cdot 101$) single fibre simulations are needed. To avoid these computing intensive simulations, an interpolation function for the LYAM was found. The light yield attenuation A for a given integrated luminosity L is calculated with

$$A(L) = a \cdot L + b \cdot \exp(-c \cdot L + d).$$

For the parameter a, b, c and d there is no global solution. Due to the sharp dose profile in the detector the fit is performed for each bin and the set of the parameters is stored. With this linear interpolation it is possible to provide a LYAM for any value of L without additional fibre simulations. The only missing part now is the natural ageing of the fibre. This effect is also treated like an additional attenuation process. So the total attenuation a_{total} is the sum of all effects.

$$a_{\text{total}} = a_0 + a_{\text{rad}} + a_{\text{age}}$$

Thereby a_0 is base attenuation of fibre and SiPM, a_{rad} the attenuation due to radiation damage and a_{age} the attenuation because of natural fibre ageing. Now the implementation of the remaining part a_{age} is explained. The effect of ageing courses a light loss of 1.4 %/yr cf. chapter 5. As the effect is homogeneous over the fibre the path length, d , of the photons, respectively the excitation point is relevant. Similar to the base attenuation of light in the fibre, cf. Equation 4.3 an exponential function is used to describe the effect of light reduction.

The time and Luminosity dependent attenuation coefficient is

$$\alpha_{\text{total}}(t, L) = \alpha_{\text{interpol}}(L) \cdot \exp(k_{\text{age}} \cdot d \cdot t) \quad (5.2)$$

with $k_{\text{age}} = -4.2 \cdot 10^{-4}$, the time t is in months and the distance d in metre.

This time and luminosity dependent LYAMs are used in the Monte Carlo sample generation in the following section.

5.4 Monte Carlo Sample Generation

In this section the generation of the used sample is presented and the needed settings are listed. Also the basic strategy of the long term studies is explained.

For the generation of the Monte Carlo sample the LHCb software stack is used, cf. section 3.4. To recap, a pp collision is simulated by GAUSS. The detector response, which includes the light guidance in the fibres of the SciFi tracker, is done via BOOLE. And for the final reconstruction BRUNEL is used. Here the former Run II framework is used as the MOORE application for Run III is not available yet.

The impact of changed detector conditions on the physics performance is evaluated easily when the simulated collisions stay the same. This means that a set of proton-proton collisions is generated, in total 3000 events. These sample of events are digitised with different settings in BOOLE, which represent the different phases in the lifetime of the tracker. After this step, all events are reconstructed with the same algorithms. The output of the reconstruction contains information about the number of tracks. The different track types will be explained in section 5.5. Besides the number of tracks, there are other output variables. Due to the known MC information, it is possible to match the output, tracks and hits, to the generated ones. The different categories are listed in chapter 6.

Coming back to the settings of the initial collision. The sampled decay is set to $B_s^0 \rightarrow \phi\phi$. This is a standard decay within the LHCb collaboration for tracking and efficiency studies. It contains a B meson and tracks in all tracking stations. Due to a small angle between the ϕ 's the tracking system has to separate close tracks. This decay is forbidden at tree level in the SM and the predominantly process is via a gluonic $\bar{b} \rightarrow \bar{s}s\bar{s}$ loop level process [89]. Therefore, this decay channel can probe for new heavy particles entering the quantum loop. The amount of interactions per bunch crossing is set to the planned value for the upgrade detector, $\nu = 7.6$. Delayed signals from interactions in the detector can be read out in the next bunch crossing; this effect is called spillover and is also enabled in this simulation. The radiation damages and the light reduction are assumed to be independent of the magnetic field direction. Therefore, only one magnet polarity is used, magnet down.

For the description of the overall SciFi tracker the geometry version 6.4 is used. The data is processed with the GAUSS version v51r0. In the following step BOOLE in the version v40r2 creates the detector response. Here the LYAMs, which were introduced in section 5.3 are used. The map provides the factor of reduction which is applied to the produced photons dependent on the position in the tracker. This procedure of the sample production is a common way even without any radiation

or ageing studies.

To the normal attenuation coefficients in the map a time dependent factor is added. This underlays the assumption that the dose is applied to the tracker constant over time. Of course this is not totally true. There are shorter breaks between the fills of the LHC and longer breaks like the end of year shutdown. But as the used radiation damage model is only dose dependent and not dose rate, this simplification is acceptable. It is possible to activate an ageing function. If so, an additional attenuation factor is added. Also radiation damage effects in the readout are taken into account. For the SiPM a noise rate of 14 MHz after 50 fb^{-1} is expected [90]. This value is also scaled linear in time like the fibre damage. In the range of 0 to 70 fb^{-1} sample are produced, thereby a step size of 5 fb^{-1} was used. Within these 15 sample no ageing was taken into account. In addition to this the same was done but with activated ageing effect. For the case that the ageing effect was underestimated a third set of sample was produced with doubled ageing effect. The produced sample go beyond the planned data taking of 50 fb^{-1} . One reason for this is to cover unexpected changes in the LHC schedule. Another one is, that it is foreseen to use parts of the SciFi in the next upgrade phase.

After this step the reconstruction of the event is done with BRUNEL version v60r3. For all samples the exact same algorithms are used. The different track types which are build are explained in the next section.

5.5 LHCb Track Types

Depending on the information from the different parts of the tracking system tracks are grouped into different categories. These categories and the underlying conditions are explained now. Figure 5.4 shows a simplified sketch of the LHCb tracking system. The different track types are illustrated in blue. Tracks within the VELO are called VELO tracks. Within the forward tracking, these are extended to the other tracking stations. The most constrained track is a Long track, the extended VELO track matches hits in the UT and the SciFi. Tracks from particles which leave the detector acceptance after passing the UT are called Upstream tracks. There are two other track types without hits in the VELO. Therefore, different algorithms besides the forward tracking are needed. Tracks in the UT and SciFi are so-called Downstream tracks, reconstructed by the downstream tracking. This downstream tracks are important for long-lived particles like K_S^0 and Λ . These decay after leaving the VELO, therefore the decay products create downstream tracks.

The last category is TTrack, and they are located only in the SciFi, here the seeding algorithm is used. Of course, overlaps within the different types of tracks have to be taken into account as well. TTracks which are found by the seeding can also be found as a long track by the forward tracking. A special algorithm will delete the duplicates.

In general the SciFi tracker has a single hit efficiency of 99 %, taking into account

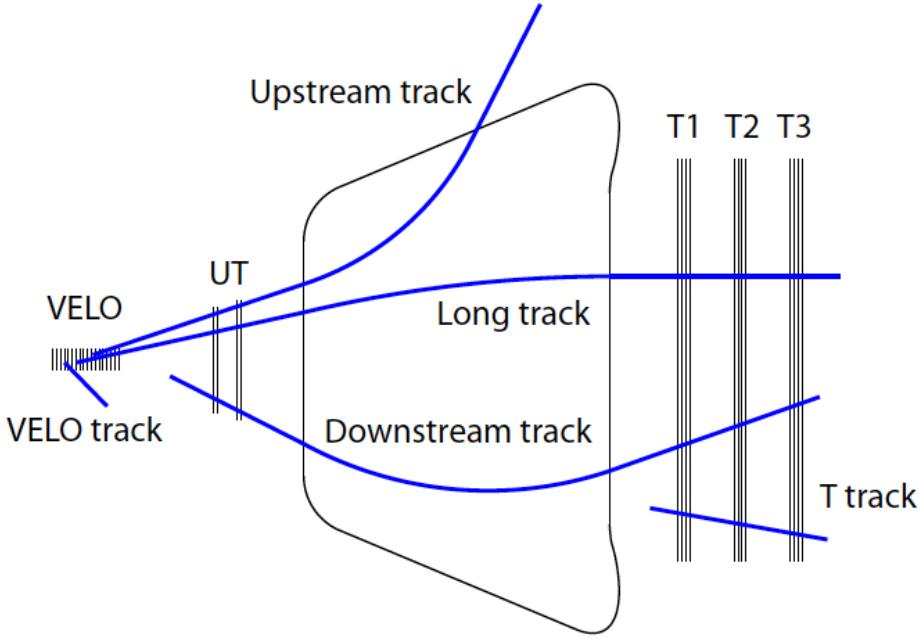


Figure 5.4 – Tracking components of the LHCb detector. The different track types are illustrated in blue [35].

the later steps in the detection chain a effective hit conversion rate of $\approx 96\%$ will be achieved. So not all, but only 61 % of the tracks are expected to have hits in all the twelve layers. Different conditions are required by the algorithms, the forward one requires at least four stereo hits and four in the x layer. In case of the seeding the minimum hit number in the final fit for the track is nine.

Within the simulation a Monte Carlo particle is defined as *reconstructible* in the SciFi when there is at least one x- and one stereo-hit in each station. The *reconstructed* tracks found by the algorithm are compared to the ones which are declared as *reconstructible*. The two tracks are matching when they share at least 70 % of the hits of the *reconstructed* track [91].

Based on the definitions the following parameters for the efficiency are calculated. The tracking efficiency ($\epsilon_{\text{Tracking}}$) is the ratio of the *reconstructed* and matched tracks to the total amount of *reconstructible* tracks.

$$\epsilon_{\text{Tracking}} = \frac{\text{reconstructed matched}}{\text{reconstructible}} \quad (5.3)$$

Another important factor is the *ghost rate*. It is the ratio of *reconstructed* tracks, which could not be matched to the Monte Carlo information, and the number of

reconstructed tracks

$$\text{ghost rate} = \frac{\text{reconstructed not matched}}{\text{reconstructed}} \quad (5.4)$$

Ghost tracks occur most often due to the combination of hits from different particles or noise. It is possible that the random combination of hits fulfil the quality checks in the fits.

The detector tracking performance over time is compared to the initial one at the time $t=0$. Due to the reduced light cluster which were detected at $t=0$ may disappear in later simulation with $t>0$. This can cause that tracks are not *reconstructible* anymore because of the lack of hits. As less tracks are *reconstructible* the efficiency can increase. To avoid this artefact a corrected efficiency ϵ_{corr} is calculated by

$$\epsilon_{\text{corr}}(L, t) = \frac{\text{reconstructed}(L, t) \text{ matched}}{\text{reconstructible}(L = 0, t = 0)} \quad (5.5)$$

where t is the time after start of the data taking period and L the integrated luminosity. The loss of the efficiency, ϵ_{loss} , is the difference between ϵ_{corr} and the start efficiency. Because of the identical simulated collisions and the relative efficiency comparison, systematic uncertainties from the detector cancel out. Therefore, no error bars are added to the plots in [section 6.2](#). A statistical error based on the number of tracks would not be visible.

The quality parameter introduced above are calculated in the following track categories:

- long
- long $> 5 \text{ GeV}$
- long from $B > 5 \text{ GeV}$
- TTrack

The category with the long tracks has two sub categories. The first has an energy cut on the momentum. The second has the same cut and requires in addition that the track is coming from a B meson. Within BRUNEL, three different algorithms are used to reconstruct tracks. These are *forward* tracking, *seeding* and so-called *best*. The first two are HLT1 like and the last has offline precision and is HLT2 like. Therefore, the *best* algorithm can use the entire detector data. The output of this algorithm is the match of all available tracking algorithms. For all off three introduced algorithms, the categories above are applied. Hereby, the TTrack can only be evaluated in the seeding algorithm, because the other do not create standalone TTracks. More detailed information about the different algorithms can be found in [\[33\]](#).

6 Long-term Studies

This chapter will present the results of the long-term studies in the different scenarios. Therefore, at first the basic features and distribution of the clusters itself are shown in [section 6.1](#). These numbers are compared to the corresponding ones at the end of Run III. This is followed by the studies on the tracking efficiency change during the planned lifetime of the SciFi in [section 6.2](#). In the end a conclusion is given and a outlook is presented, cf. [section 6.3](#). All results are based on the MC samples described in the previous chapter. These samples are upgrade simulations and not LHCb official.

6.1 SciFi Cluster

The number of clusters in the SciFi and their distribution over the 3000 events is shown in [Figure 6.1](#), within BOOLE the SciFi-hits are called lite clusters. The shape of the average number of clusters is Gaussian-like and the mean value is 4775 with a standard deviation of 1868. This broad range of the cluster number will be relevant in the GPU decoding, cf. [section 7.3](#).

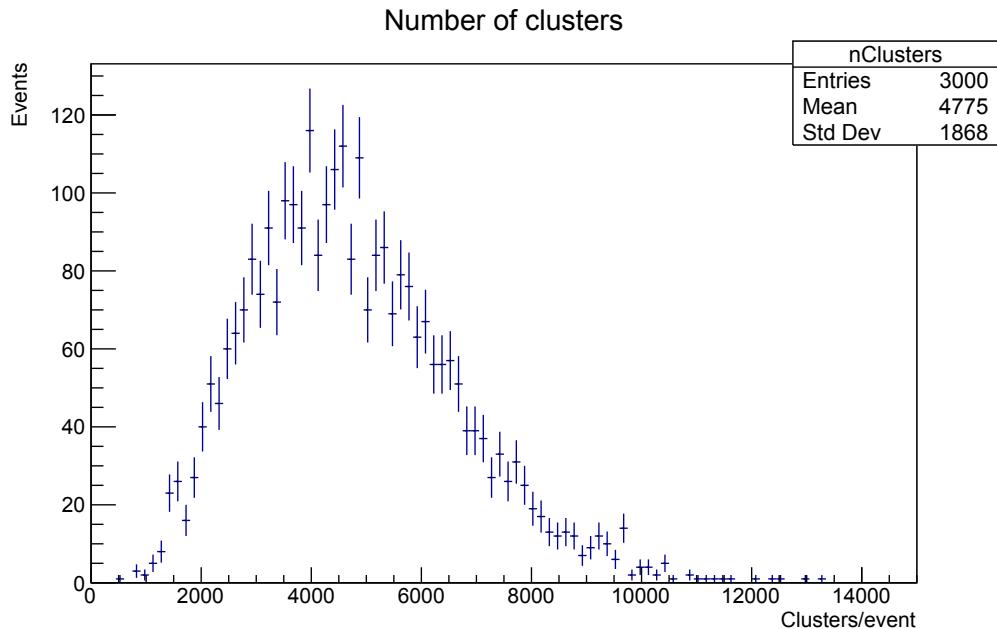


Figure 6.1 – Distribution of the number of cluster per Event.

[Figure 6.2](#) illustrates the distribution of the clusters in the detector. Therefore, a pseudo channel ID is used. Here the channel from the four quarters are matched depending on their distance to the beam-pipe. The distribution of the cluster in the detector matches the expectation from the FLUKA maps. Two discontinuities are visible: the first one results from the shorter beam-pipe module and the second from the extra module in station T3. A similar plot is shown in [Figure 6.3](#) but here only cluster from spillover are considered.

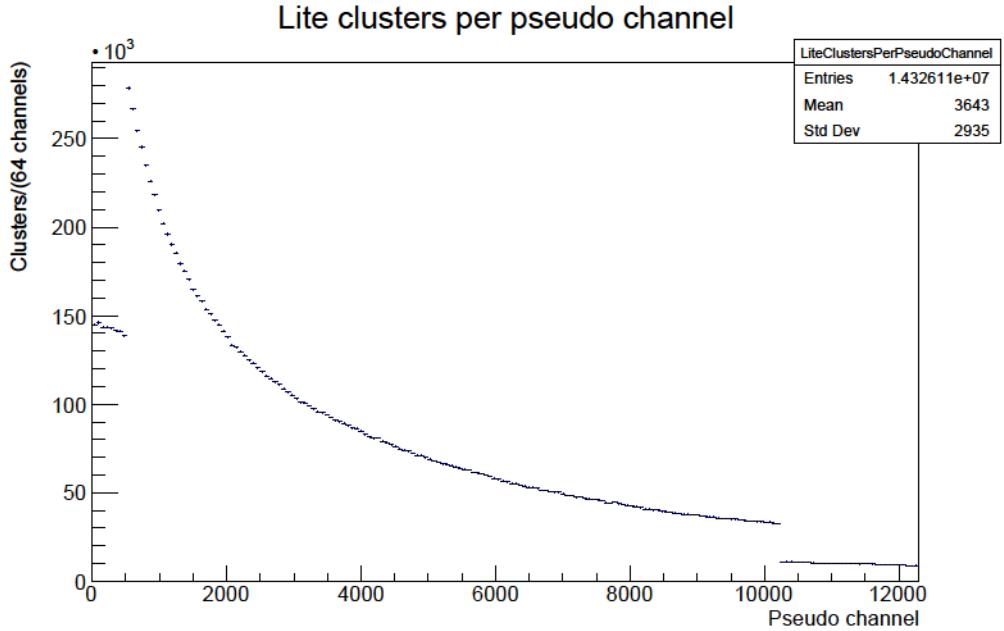


Figure 6.2 – Histogram of the cluster per pseudo channel condition (0,0). The short beam-pipe module causes the first discontinuity and the extra module in T3 the second.

For the initial condition of the SciFi the number of cluster and their distribution throughout the detector is known. Also the amount and distribution of spillover clusters is known. These numbers are compared to the corresponding ones after an integrated luminosity of 50 fb^{-1} . All three scenarios of ageing factors are considered and the conditions are named (0,0), (50,0), (50,1) and (50,2) where the first number represents the integrated luminosity and the second is the multiplication factor which is applied to the ageing effect.

In [Figure 6.4](#) the average number of clusters per event is shown. Due to the damages in the fibre a smaller number could be expected, but due to the increased noise in the SiPM the number is higher. But as expected the cluster number per event is decreasing over time when the ageing effect is added and even more in the sample with double ageing. The values of the mean cluster per event are listed in [Table 6.1](#). After 50 fb^{-1} with normal ageing the average number of clusters is increased by a factor of 1.09.

The same factors can be obtained from the entry numbers of the histograms

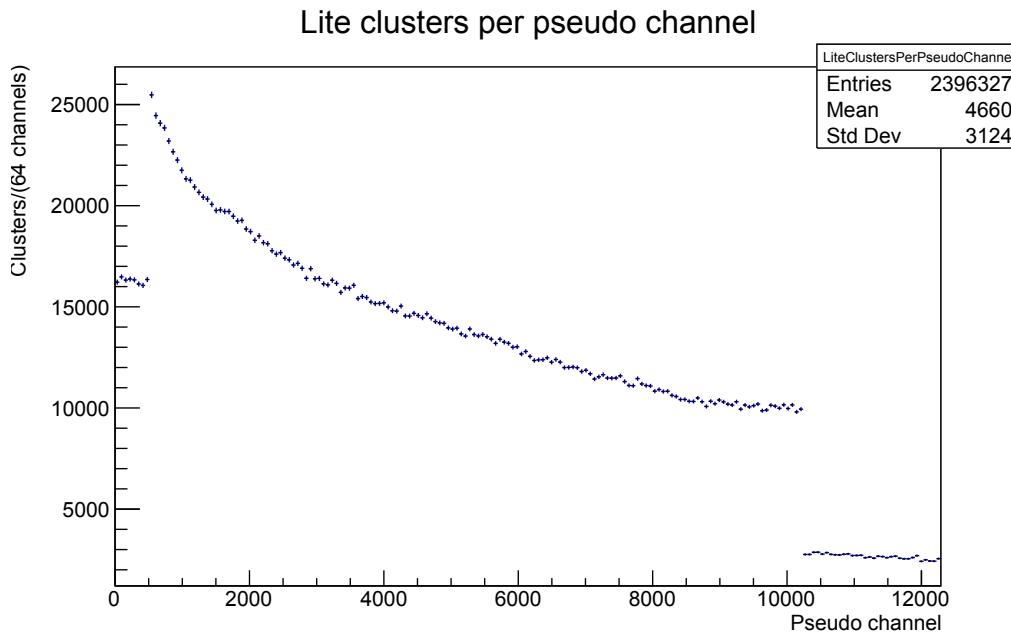


Figure 6.3 – Histogram of the cluster per pseudo channel whereby only spillover signals are taken into account, condition (0,0).

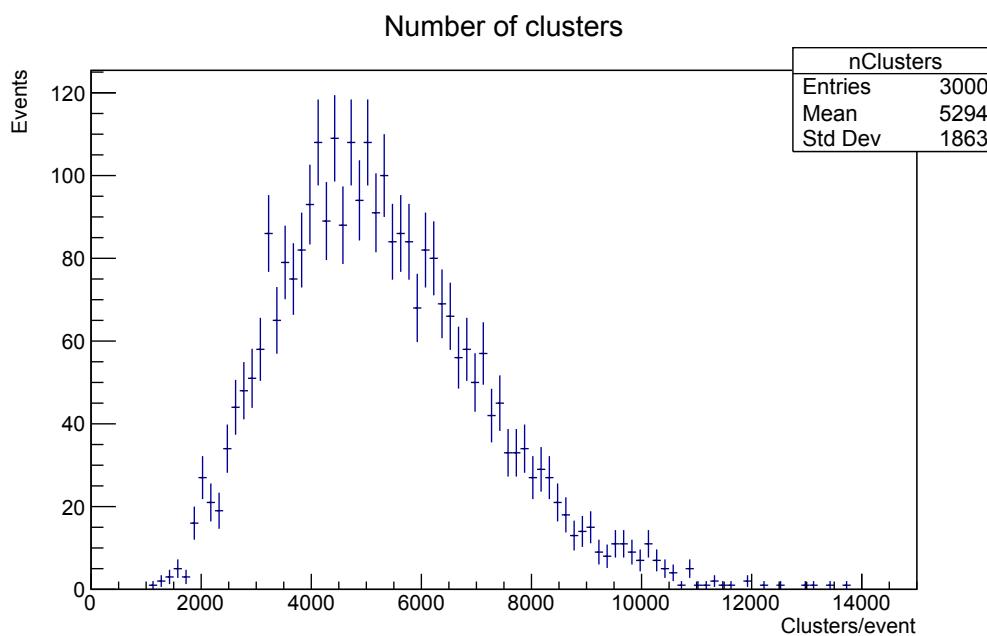


Figure 6.4 – Number of cluster per event with condition (50,0).

6 Long-term Studies

Table 6.1 – Mean number of cluster per event and the corresponding standard deviation for different conditions.

condition	mean #cluster	err	modification factor
0-0	4775	1868	-
50-0	5294	1863	1.11
50-1	5203	1837	1.09
50-2	5101	1808	1.07

from the pseudo channel cluster distribution. [Figure 6.5](#) shows this distribution for 50 fb^{-1} and normal ageing effect. The plot for condition (50,2) is similar and can be found in the appendix [Figure A.1](#). In [Table 6.2](#) the results from different scenarios are summarised. The mean value is shifted to higher pseudo channel after the 50 fb^{-1} but there is no significant difference between the different ageing sample. The shift is expected as the damage caused by radiation is higher in the inner part of the tracker.

Table 6.2 – Results from the cluster by pseudo channel distribution for the start condition and 50 fb^{-1} with the three ageing sample.

condition	# cluster	mean pseudo channel	mean change factor
0-0	1.43E+07	3643	-
50-0	1.59E+07	3850	1.06
50-1	1.56E+07	3846	1.06
50-2	1.53E+07	3846	1.06

The LYAM was split in direct and reflected light. Through that it is possible to evaluate the effect of the different ageing factors on the amount of spillover. The distribution of the clusters per pseudo channel from spillover is shown in [Figure 6.6](#), here the normal ageing effect was used. In [Table 6.3](#) the results from the four conditions are listed.

Table 6.3 – Comparison of the amount of cluster and cluster from spillover respectively.

condition	# cluster	# spillover cluster	mean spillover	spillover ratio in %
0-0	1.43E+07	2396327	4660	16.73
50-0	1.59E+07	4014029	5016	25.28
50-1	1.56E+07	3837755	5031	24.59
50-2	1.53E+07	3666316	5050	23.96

The amount of spillover has increased from 16.7% to 24.6% in case of the normal ageing.

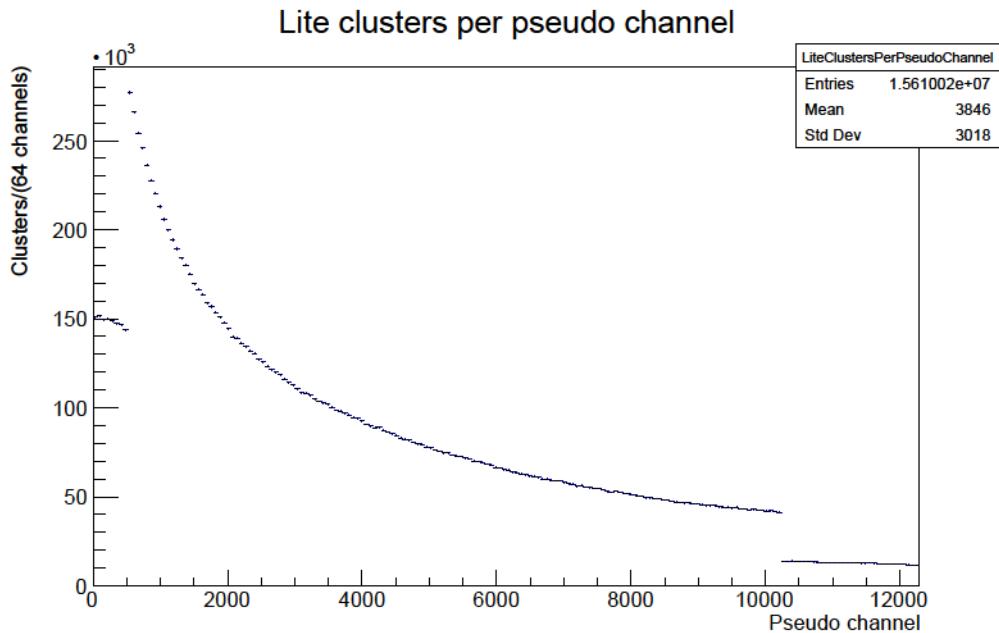


Figure 6.5 – Distribution of the cluster in the SciFi per pseudo channel after 50 fb^{-1} with normal ageing. condition (50,1)

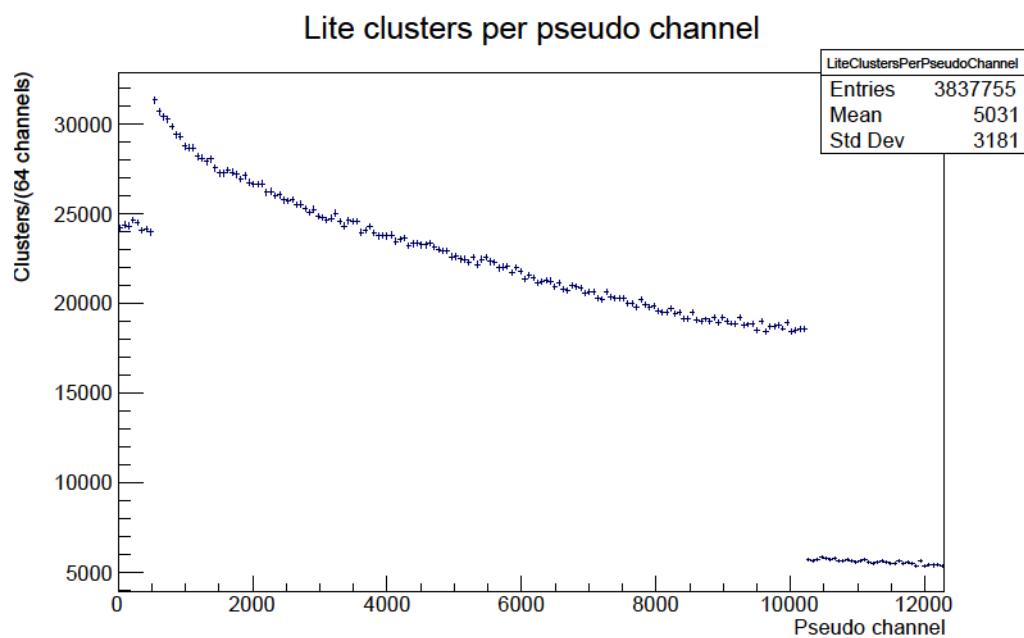


Figure 6.6 – Cluster from spillover in the (50,1) sample per pseudo channel.

6 Long-term Studies

To conclude here, after 50 fb^{-1} there will be on average 9% more clusters in an event. Also the ratio of spillover will be increased. The effect on the tracking performance is evaluated in the next section.

6.2 Tracking Efficiency During the Detector Lifetime

In this section the condition dependent tracking efficiency for various track types and algorithms is presented. The key value which is presented is the ϵ_{loss} , the difference of start efficiency and the value for a given condition. As all the following results are compared to the performance at start of the detector data taking period the efficiency for this condition is shown in detail.

In [Table 6.4](#) characteristic numbers from the reconstruction step of the initial sample are listed. For the *seeding* algorithm the overall tracking efficiency is the highest. Within the *best* performing category the efficiency is $\approx 95\%$.

Table 6.4 – Number of tracks *reconstructed* and *reconstructible* for different categories in the starting sample with 3000 events. The corresponding efficiencies are given in percent.

	<i>seeding</i>	<i>forward</i>	<i>best</i>
hasT <i>reconstructed</i>	311694	-	-
hasT <i>reconstructible</i>	383183	-	-
$\epsilon_{\text{Tracking}}$ in %	81.3	-	-
long <i>reconstructed</i>	228428	204211	216172
long <i>reconstructible</i>	250204	250204	250204
$\epsilon_{\text{Tracking}}$ in %	91.30	81.62	86.40
long $>5\text{GeV}$ <i>reconstructed</i>	154186	144020	148313
long $>5\text{GeV}$ <i>reconstructible</i>	162603	162603	162603
$\epsilon_{\text{Tracking}}$ in %	94.8	88.6	91.2
long from B $>5\text{GeV}$ <i>reconstructed</i>	11018	10742	10947
long from B $>5\text{GeV}$ <i>reconstructible</i>	11618	11618	11618
$\epsilon_{\text{Tracking}}$ in %	94.8	92.5	94.2
Ghost tracks	36807	102052	37539
Total tracks	312900	326775	273163
<i>ghost rate</i> in %	11.7	31.2	13.7

As mentioned before, samples under different conditions were produced. In contrast to the previous section all sample are used and not only 0 fb^{-1} and 50 fb^{-1} .

First the results for the *seeding* algorithm are presented. The efficiency loss for the four track categories is illustrated in [Figure 6.7](#). The vertical red dashed line indicates the planned 50 fb^{-1} after 10 years. In each plot the three different ageing models are included. The following convention is used for all these plots.

6.2 Tracking Efficiency During the Detector Lifetime

Blue indicates no ageing, orange normal ageing and green doubled ageing effect. The efficiency of the detector is decreasing as expected and the double ageing

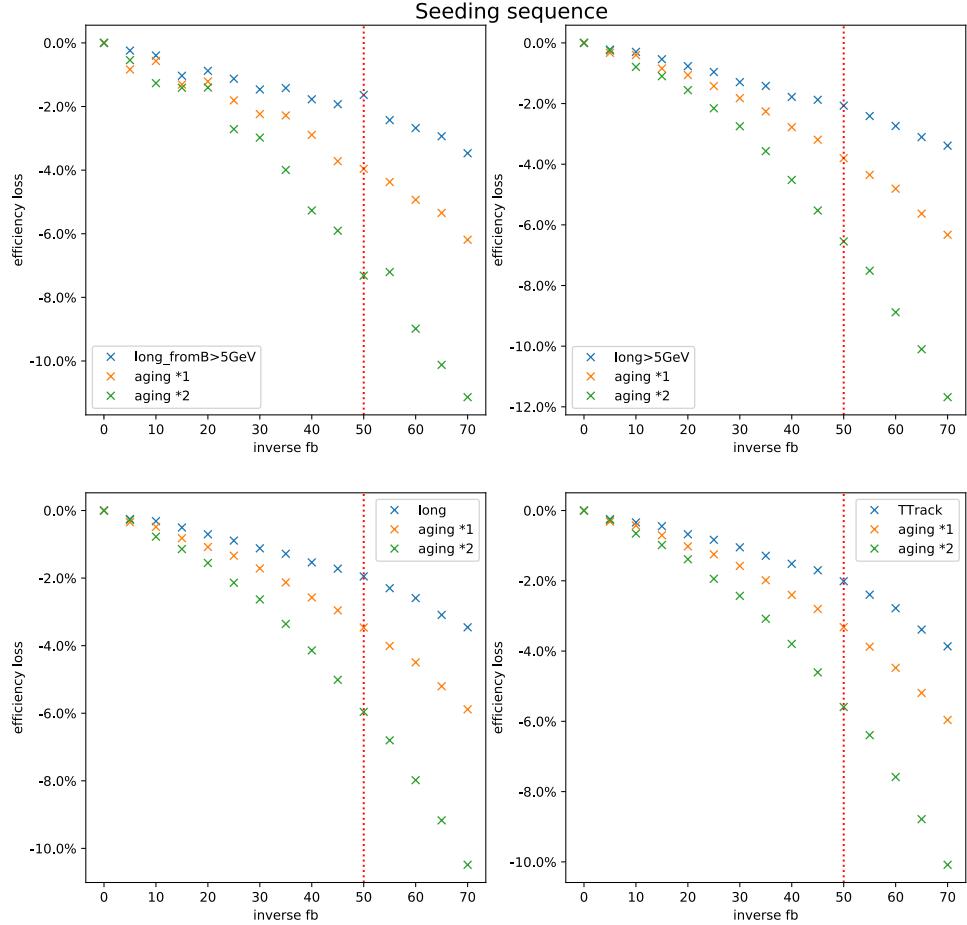


Figure 6.7 – Efficiency loss compared to the start of data-taking. The planned detector lifetime is indicated by the red dashed vertical line. Blue indicates no ageing, orange normal ageing and green doubled ageing effect. Different track categories are used for the four plots, these are label in the corresponding legends.

is worse than the normal. The shape of the curves is smooth beside the high energetic B meson category. But in this category the statistics are limited the most. The ϵ_{loss} for TTracks is 3.3 % after 50 fb^{-1} with normal ageing. An overview of the performance loss after this time is listed in [Table 6.5](#). In addition to the

reconstruction efficiency the rate of ghost tracks is calculated. The trend of this rate is shown in [Figure 6.8](#). In the beginning the rate was 11.8 %, this is increased up to 14.6 % for normal ageing. It seems that the amount of ageing has a small effect on the ghost rate compared to the other effects.

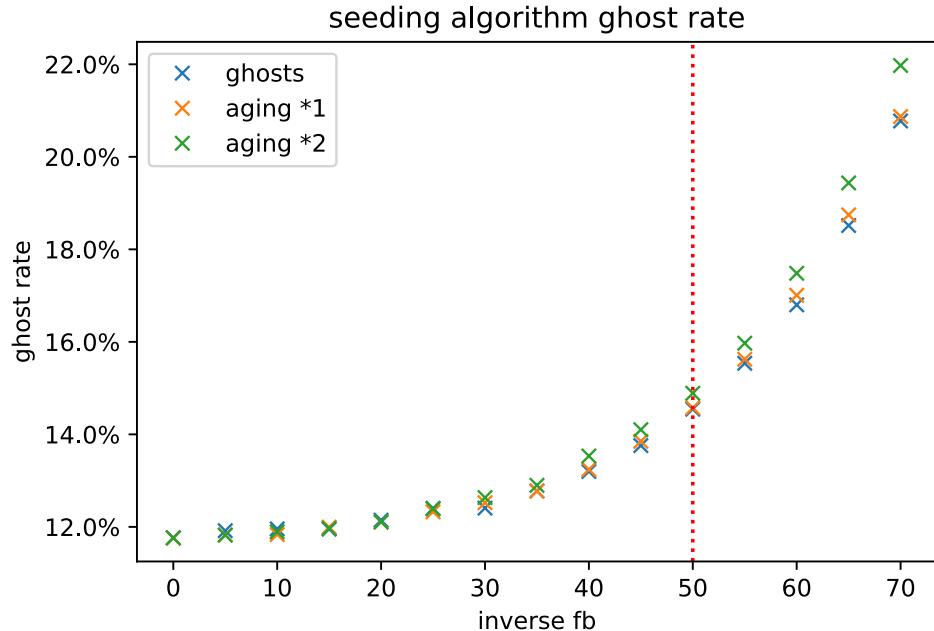


Figure 6.8 – Ghost rate of the *seeding* algorithm over time for the three ageing scenarios.

In addition, the same sample were reconstructed with the *forward* algorithm. The efficiency for the three possible track categories and the ghost rate are shown in [Figure 6.9](#). The efficiency for long tracks is reduced about 3.2 %. Thereby the *ghost rate* is increased by 1.3 %.

The last used algorithm is *best*, which is HLT2 like. The corresponding trend plots are in [Figure 6.10](#). Similar to the other algorithms the overall efficiency is reduced and the ghost rate increased. For an overview all performance numbers after 50 fb^{-1} for all ageing scenarios are listed in [Table 6.5](#).

Normally the manifestation of the added ageing effect is similar for all algorithms. Only the ghost rate deviates from this trend. In general the ghost rate is increased for all algorithms after 50 fb^{-1} , but for the *seeding* algorithm the rate is increasing slightly with increased ageing effect, while for the two other algorithms the rate is decreasing. This means that the *seeding* algorithm is creating more ghost tracks than the other algorithms. The actual number of tracks which are reconstructed in the different algorithms are listed in [Table 6.6](#). Based on these numbers the ghost rates were calculated. The three ageing scenarios are listed for every algorithm

6.2 Tracking Efficiency During the Detector Lifetime

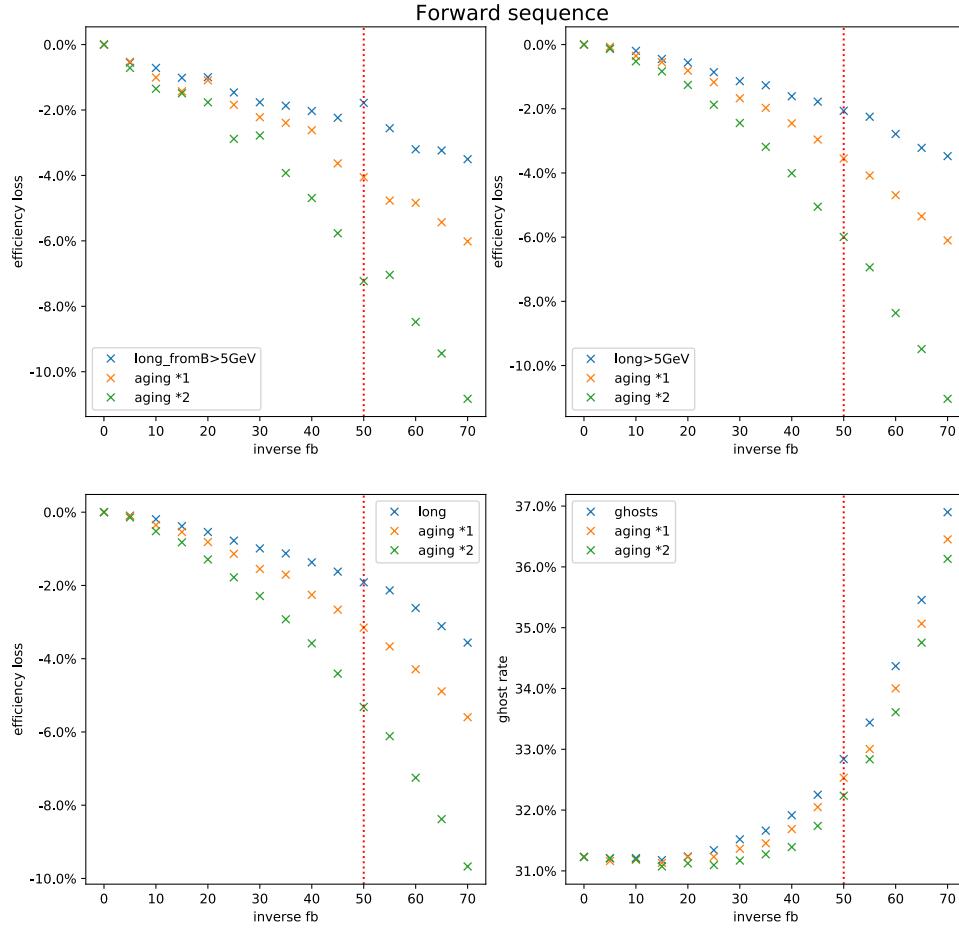


Figure 6.9 – Evolution of the ϵ_{loss} and the ghost rate for the *forward* algorithm.

starting from 0 ageing to double ageing. Also the difference between end and start is calculated. For the *best* sequence the difference of the matched tracks is negative. So less tracks are reconstructed. But for the two other algorithms the number of found tracks is increasing. It seems that the hits from noise are helpful to reconstruct some tracks. But this trend is cancelled by the added ageing effect. The absolute number of ghost tracks is increased after 50 fb^{-1} especially for the *seeding*. It can be concluded that the ghost rate for *best* sequence is not effected that much by the changed conditions. This is due to the tighter constraints in the track fit, therefore it is more unlikely that a noise hit can be added to a track

6 Long-term Studies

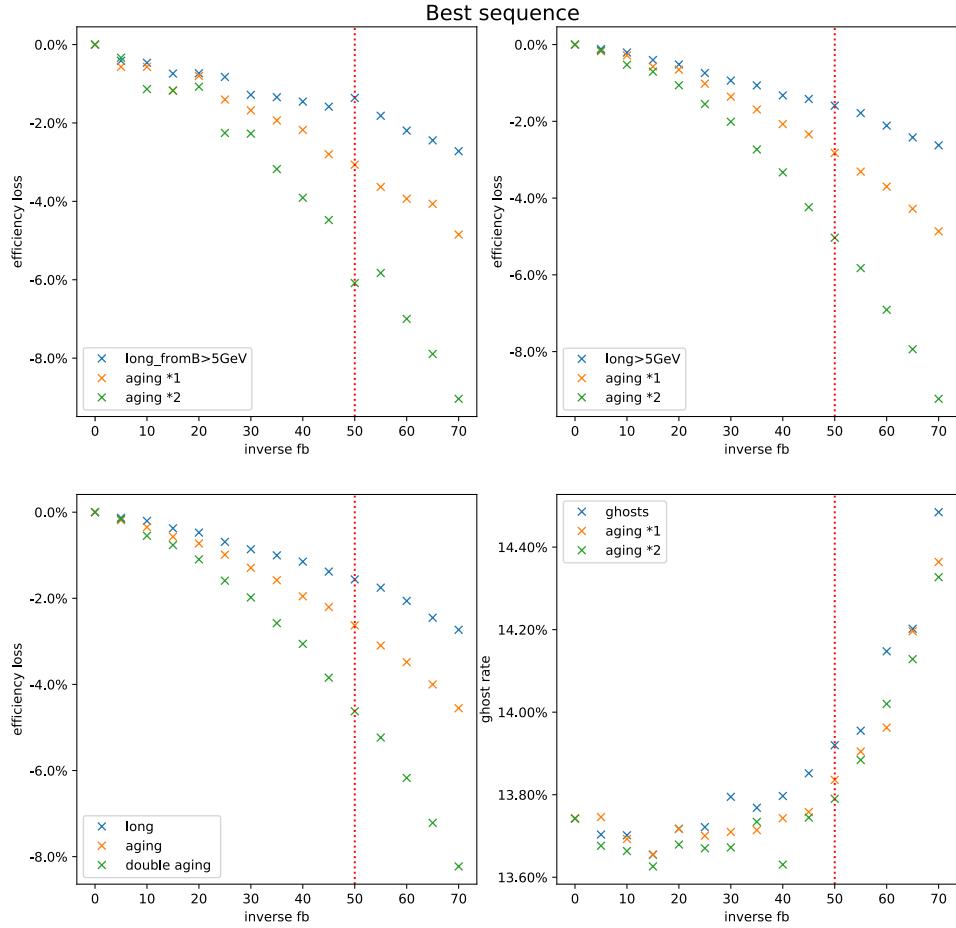


Figure 6.10 – Evolution of the ϵ_{loss} and the ghost rate for the *best* algorithm.

compared to the other algorithms.

6.2 Tracking Efficiency During the Detector Lifetime

Table 6.5 – Overview of the ϵ_{loss} and ghost rate for the different algorithms after 50 fb^{-1} . All numbers are given in percent

track cat	ageing factor	ϵ_{loss}	seeding	ϵ_{loss}	forward	ϵ_{loss}	best
start $\epsilon_{\text{Tracking}}$		94.8		88.6		91.2	
	0	-1.9		-1.9		-1.6	
	1	-3.5		-3.2		-2.6	
long	2	-9.0		-5.3		-4.6	
		94.8		88.6		91.2	
	0	-2.1		-2.1		-1.6	
long from B	1	-3.8		-3.6		-2.8	
	2	-6.6		-6.0		-5.0	
		94.8		92.5		94.2	
start $\epsilon_{\text{Tracking}}$	0	-1.6		-1.8		-1.4	
	1	-4.0		-4.1		-3.1	
	2	-7.3		-7.2		-6.1	
TTracks		81.3		-		-	
	0	-2.0		-		-	
	1	-3.3		-		-	
	2	-5.6		-		-	
start ghost rate		11.8		31.2		13.7	
	0	14.5		32.8		13.9	
	1	14.6		32.5		13.8	
	2	14.9		32.2		13.8	

Table 6.6 – Absolute number of the reconstructed tracks, matched and not matched, within the 3000 events. The numbers are given for the three algorithms and all ageing effects.

sequence	ageing factor	matched			not matched		
		start	end	difference	start	end	difference
<i>best</i>	0	273163	270073	-3090	37539	37594	55
	1	273163	266776	-6387	37539	36911	-628
	2	273163	261159	-12004	37539	36014	-1525
<i>forward</i>	0	326775	328459	1684	102052	107864	5812
	1	326775	322260	-4515	102052	104837	2785
	2	326775	312877	-13898	102052	100862	-1190
<i>seeding</i>	0	312900	315396	2496	36807	45873	9066
	1	312900	309847	-3053	36807	45158	8351
	2	312900	300882	-12018	36807	44797	7990

6.3 Conclusion and Outlook

The interpolated LYAMs were used to simulate the detector over the planned lifetime and even beyond. Different magnitudes of the natural ageing effect were included. Based on these samples, performance numbers were calculated. This performance evaluation was done first for the raw cluster data and second for the track reconstruction sequences.

The studies on the cluster data showed that the amount of clusters is increasing during the expected data-taking period. This is mainly due to the noise of the SiPM. Thereby the radiation damage and fibre ageing are counteracting this increase. At the end of Run III, the number of clusters per event are 9 % higher than at the beginning. Also, a shift of the mean cluster position away from the beam-pipe is visible. Because of the higher radiation in the inner part this is expected.

Furthermore, the arrival time of the photons was tracked to study effects on the spillover rate. For that reason, the LYAM treat direct and reflected light independently. The rate of clusters which arrive the SiPM late, so that they are considered to belong to the next event is increased from 16.7 % to 24.6 %. Because of the light loss due to radiation and ageing, less clusters are detected. Therefore, fewer tracks can be reconstructed. This trend is visible for all algorithms and track categories. Within the collaboration studies on the tracking efficiency were done based on the single fibre simulation and the FLUKA map for 50 fb^{-1} . As expected, the results, from the interpolated map with the same radiation level, are compatible. But with included ageing effect the ϵ_{loss} is 1.7 times higher than as predicted by the model which included only radiation damage, which was assumed in earlier studies.

At this point, it should be mentioned that the initial light yield of the fibres is outdated. The assumed value based on a test beam is from 2018. The initial time gap between fibre module production and the start of Run III is enlarged significantly due to the corona crisis. Therefore, it is necessary to measure the light yield of the mats again. Due to the lack of tools which describe the performance change of the other LHCb sub-detectors, the performance loss is probably underestimated. Because of the greater damage in the inner region, the performance loss is higher for tracks with high momentum as these are more likely to be in the inner part. As a result the ϵ_{loss} is higher in the *long from B* $> 5\text{ GeV}$ category than in *long from B*.

A possible solution for this could be to change modules from the outer part with inner ones. Thereby the replacement of the beam-pipe modules is not possible due to their different geometry. Another possibility to counteract the performance loss can be the change of the limits in the clustering. Because of the higher amount of spillover it is might be worth to use a multi event format for the cluster like in the former outer tracker. However, this would cause a significant change in the overall event model.

In the first years of data-taking, the ghost rate increases slightly for the *seeding* while the rate gets smaller for the *forward* and *best* algorithm. Nevertheless, at the end of the planned lifetime, the rate is increased significantly for all cases. The

6.3 Conclusion and Outlook

seeding is found to be very sensitive to the changed conditions. Adjustments within the reconstruction algorithms could help to suppress the ghost tracks.

Beside the actual performance loss, there is another reason for adjustments in the algorithms. The changed conditions effect the per event run-time in the trigger. Because of the increased amount of cluster the combinatorics of hits is increased. This slow down in the reconstruction will cause a throughput reduction, as the trigger computing power is a limiting factor. In the used reconstruction sequence, the run-time was increased by about 23 %. This value is just an estimation as the numbers are taken from the internal timing tool an no fixed conditions are used.

Beyond the the planned data-taking of 50 fb^{-1} the performance of all algorithms is getting worse. The slope of the ϵ_{loss} curves is increasing over time. A follow-up upgrade of the SciFi tracker is already under study. It is foreseen to replace old modules and use silicon strips in the inner part of the tracking station, and this is the so-called Mighty Tracker [92]. The mighty tracker will consist of fibre mats like the SciFi tracker. The geometry of the new fibre mats will probably be slightly different. Nevertheless, the tools for the simulation of the tracking performance evolution can be used to simulate the mighty tracker.

7 GPU Based Trigger

The task of the LHCb Trigger system is to reduce the data rate and choose which events are stored, cf. [section 3.5](#). An essential part of the LHCb upgrade is the pure software trigger. Even before the upgrade, the computing task of the trigger was quite challenging and computing-intensive, with the new conditions coming with the upgrade it will be even more challenging. The input rate is 30 times higher than before and in addition, the instantaneous luminosity is 5 times higher. This leads to a peak trigger data input rate of 40 Tbit s^{-1} at the average event rate of 30 MHz [93]. The Allen project, named after Francis E. Allen, was developed in a subgroup of the LHCb real-time analysis project. It is a Graphics Processing Unit based implementation of the first trigger stage, the HLT1.

A general introduction on GPUs and their programming methods is given in [section 7.1](#). The overall Allen project is described in [section 7.2](#). The implementation of the GPU based SciFi decoder is presented in [section 7.3](#). This is followed by an overview of the performance in [section 7.4](#). The chapter is concluded in [section 7.5](#) and an outlook is given.

7.1 General Purpose Graphics Processing Unit

A Graphics Processing Unit is a special electronic circuit for computing devices. These are equipped with a large number of processing units and called throughput-optimised. Originally, the task of a GPU was to render images to display them on a screen. Correspondingly, the first usage of the term GPU was by Sony in 1994 for the image processing chip used in the PlayStation 1 [94]. Over the years, GPUs have evolved significantly, but the general working principle is still the same. Data and instructions needed for the calculations are transferred from the Central Processing Unit (CPU) to the GPU. The result of the calculations is the output of the GPU, which is sent to a display device. Based on this image processing example, the high number of processors can be motivated. Assuming the task, that the colour of an image needs to be changed. As every pixel in the picture is independent, the calculation of the new colour can be done in parallel. As long as there are more processors than pixels, this simple example works out. At this point, it is important to mention that the clock frequency of a GPU processor is significantly lower than in a CPU. But the complexity of colour calculation is relatively low. Just a simple linear function, for example, addition of the three primary colours, is evaluated. Therefore, the reduced clock frequency is still sufficient. Summarising, a GPU is best suited for computations when the operation can be split into many calculations with low complexity.

This concept of computation acceleration is called Single Instruction Multiple Data (SIMD). There, the computing task stays the same and just the input data changes. This concept and the corresponding Single Instruction Multiple Threads (SIMT) are used within the Allen project.

A shader is a computer program which was originally used for the shading in 3D images. At some point, these shaders became programmable, after this, the usage of GPUs for non-image calculations became more common. With the integration of floating point units, the needed precision for scientific calculations was enabled. It was shown that some computations are more efficient on a GPU than on a CPU. One of the first examples for this is the LU factorisation [95]. Non-graphic calculations on a GPU are called General-purpose computing on graphics processing units (GPGPU). Nowadays, many computing centres are equipped with GPU accelerators. Also, custom made solutions for the training of neural networks are on the market. This ranges from TensorFlow processing USB sticks to full equipped servers with GPUs. Six of the top 10 supercomputer worldwide have GPUs [96].

For the development of GPU software, two major platforms are available, OpenCL and CUDA. The first one is an open standard defined by the Khronos Group, and the second is the proprietary by NVIDIA Corporation. Within the Allen project, the primary programming language is CUDA. Therefore, only this platform is described in the following.

7.1.1 Hardware



Figure 7.1 – General chip architecture of CPU (left) and GPU (right) [97].

As mentioned in the previous part, there are differences in the chip design of CPUs and GPUs. Sketches of the respective designs are compared in [Figure 7.1](#). The GPU possesses many more Arithmetic Logic Units (ALU). Within these, the actual computations take place. CPUs have to handle a wide range of different tasks quickly and mostly in serial order, therefore the base frequency is high, typically in the range of 2 to 4 GHz. Compared to this, the GPU has a low clock rate. Typical clock frequencies are in the order of 1 GHz. Because the GPU is throughput

optimised, no large cache is needed, the CPU cache is larger and located physically closer to the ALU than the one of the GPU. Also, a large and respectively complex Control unit is not required by the GPU. The standard interface between a GPU and the motherboard is PCI Express. Via this connection, data is copied to the memory of the GPU. Similar to a CPU, a GPU has RAM (random-access memory). A GPU uses a special variant, the so-called GDDR SDRAM, which is short for graphics double data rate type synchronous dynamic random-access memory. The memory size of consumer GPUs is in the range of 1 GB to 24 GB. With the newest generation, GDDR6x, a per pin bandwidth between 19 GB/s to 21 GB/s is possible [98].

Memory caching is a common way to speed up computing operation on CPU and GPU as well. For GPUs there are different ways to cache memory, these are illustrated in [Figure 7.2](#). There are differences in the overall accessibility of the different memory types. They are described starting with the most restricted. Local memory is only available for single threads, while shared memory is accessible for all threads within a block. The global memory is accessible for all threads. There are also global read only memory types. These are the constant and texture memory. The base component of the GPU is the Streaming Multiprocessor (SM), an array of CUDA cores. Over the years, different architectures were brought to the market by NVIDIA, beside the architecture they differ in the amount of SM and memory. Also, new enhancements and capabilities came with the latest versions, these features are summarised under the term computing capability. Ordered from old to new the architecture versions are Tesla, Fermi, Kepler, Maxwell, Pascal, Volta, Turing and Ampere. Besides the increased computing power, a benefit is improved power efficiency. A major reason for this is the smaller chip process size. For the latest Ampere GPUs, a 7 nm manufacturing process is used, while for Volta the size was 12 nm. Usually, the execution of operations on a GPU is asynchronous, only within a warp the execution is guaranteed simultaneous. A Warp is a group of threads, here the maximal number of threads is 32. A common way to process data is single instruction multiple threads. Hereby, it is essential to avoid a prominent branching of the code. If one thread of a warp runs significant longer, than the other threads of the warp have to idle for a long time. Also, the memory access of single threads can interfere and slow down the execution. A proper management of the memory accesses is needed. This problem applies mainly to the global memory and not to the block specific shared memory. For a fast memory access, CUDA has some special features like the coalesced memory access. Here, the access of a set of consecutive threads to a consecutive memory block is coalesced into a single transaction.

7.1.2 Software

The software framework CUDA is used for programming NVIDIA GPUs. The CUDA Toolkit provides all necessary tools for the development of high-performance GPU code. It has debugging and optimisation tools, a compiler (nvcc) and an

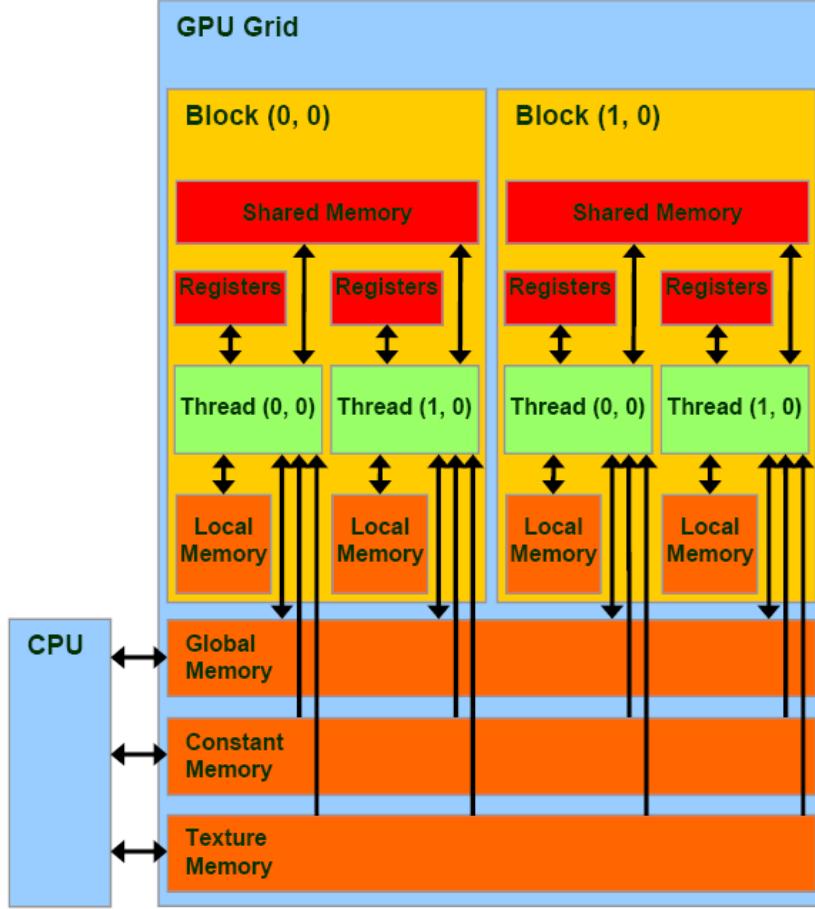


Figure 7.2 – CUDA memory hierarchy model [99].

Application Programming Interface (API). The programming language CUDA C++ is an extension with GPU specifications of the basic C++. At this point, a brief introduction of the programming with CUDA is given. The code examples are based on the book CUDA-by-example [97]. Kernel are functions executed on a device or GPU. The qualifier for GPU specific statements usually starts with two underscores, like `__global__`, and is followed by a scope of functions and methods, for example:

```
__global__ void myKernel(float *a){
    a[0] = 5;
}
```

This function is then called by the host in the following way:

```
int main(){
    // ...
```

```

myKernel<<<23,42>>>(a_d);
//...
}

```

On the host side of a program, the execution configuration takes place. The parameters between the upright chevrons define how often this kernel is invoked. In the example, a two-dimensional structure is used. First specified is the number of blocks, 23, and second the number of threads, 42. Every block will create the requested number of threads and therefore 966 ($23 \cdot 42$) instances are invoked in total. This is the basic way to start parallel running threads. For the processing of data, it is important to have a unique index for every thread. This is provided via the global variables `threadIdx` and `blockIdx`. With the previous set dimension of the blocks, `blockDim.x` the unique processID (pId) is calculated by:

```
int pId = blockIdx.x * blockDim.x + threadIdx.y;
```

As blocks and threads can either have up to 3 dimensions denominated by `threadIdx.x`, `threadIdx.y` and `threadIdx.z`. Hereby it is not needed to use all dimensions. The set of multiple blocks is called a grid and this is always located on the GPU. It is possible to specify functions or data which are only accessible by the GPU via the qualifier `__device__`. The opposite option would be only accessible by the host. This is indicated by `__host__`. The usage of location attributes was already used before in `a_d`. Here the index shows that this array is stored on the `device`. CUDA has an own memory management system, and memory is allocated by using `cudaMalloc`. Data to process is copied from the host to the device by `cudaMemcpyHostToDevice`, and results are transferred back to the host by `cudaMemcpyDeviceToHost`.

7.2 The Allen Project

The Allen project is a GPU-based implementation of the first stage of the LHCb Upgrade trigger, HLT1. At the beginning of the LHCb upgrade plans, a CPU based trigger with massively increased computing capacity was planned [33]. This plan leads to a similar online farm structure as in Run I and Run II. Since the CPU computing performance did not grow like expected, alternative approaches for a trigger acceleration became necessary. In the year 2015, the first ideas for the usage of GPU within the LHCb trigger came up. Subsequently, a standalone project within the LHCb RTA group raised from this idea. In the beginning, it was foreseen to use the GPU as an additional accelerator within the CPU farm. The VELO decoding and primary vertex finding were the first components of the trigger ported to GPU. Because of the good performance, it was decided to port the entire HLT1 sequence to GPU. A homogeneous architecture will provide an additional performance benefit because fewer copy processes are needed. Also, a homogeneous is easier to maintain. Like the switch from a CPU to a novel GPU

trigger, significant experiment changes need approval from the collaboration. The corresponding committee is the LHCb Technical Board.

For a positive decision concerning the GPU based trigger, different things need to be demonstrated. First of all, the feasibility needs to be shown, as Allen is the first complete software GPU trigger with high throughput in the field of high energy physics experiments. Also, a significant performance benefit due to the switch is necessary. Within the LHCb group, technical design reports (TDR) are used to present the status of a project to the collaboration. To that effect, the Allen group published a TDR, cf. [93]. During the completion of this thesis, it was decided to choose the GPU trigger as the new baseline strategy. Based on the TDR and a corresponding publication in *Computing and Software for Big Science* [100] the requirements on the project and their realisation, which lead to the positive decision, are presented.

7.2.1 Integration of Allen

Before the technical details about the project are presented, the planned physical integration is shown. Events in the LHCb computing farm are grouped into so-called multi-event packets. This packaging is done within the Event Builders, cf. section 3.5, where the GPUs are planned to be installed. On the one hand, it is almost for free to implement GPUs here, as the servers have empty PCIe slots. On the other hand, a cost reduction for the EB and EFF network is possible. The first data reduction takes place in the EB, therefore a slower connection is sufficient. This slower network connection is significantly cheaper, because the on-board network cards are sufficient. A sketch of the online farm with GPUs is illustrated in Figure 7.3b. The GPUs are located in the EB and reduce the incoming rate of 40 Tbit s^{-1} down to 1 Tbit s^{-1} to 2 Tbit s^{-1} . The events are stored in a disk buffer. When the online detector calibration and alignment is done the EFF with the HLT2 sequence reduces the rate further to 80 Gbit s^{-1} . The corresponding scheme without GPUs is shown in Figure 7.3a, here the HLT1 is located in the EFF.

During longer periods without data recording, the old HLT farm was used for the Monte Carlo sample production. Of course, this task is also planned for the new HLT. Due to the significant replacement of CPUs by GPUs, the scheduled computing capability for the MC sample production is reduced. Besides the LHCb Allen project, GPUs are becoming more relevant in supercomputers, and they deliver a significant amount of the computing capacity. Therefore, work is ongoing within the community to enable GPU based MC sample production.

Another critical point to demonstrate was the long term stability of the GPUs within the farm. This point is even more crucial when consumer graphic cards are purchased. These are not specified for a continuous usage. Within the TDR, a long-term stability study with NVIDIA GTX 2080 Ti cards is presented. Cards with and without fan were used, and no errors or problems occurred.

Besides the physical implementation the software integration into the LHCb framework is essential. The Allen project is managed via a Git repository like

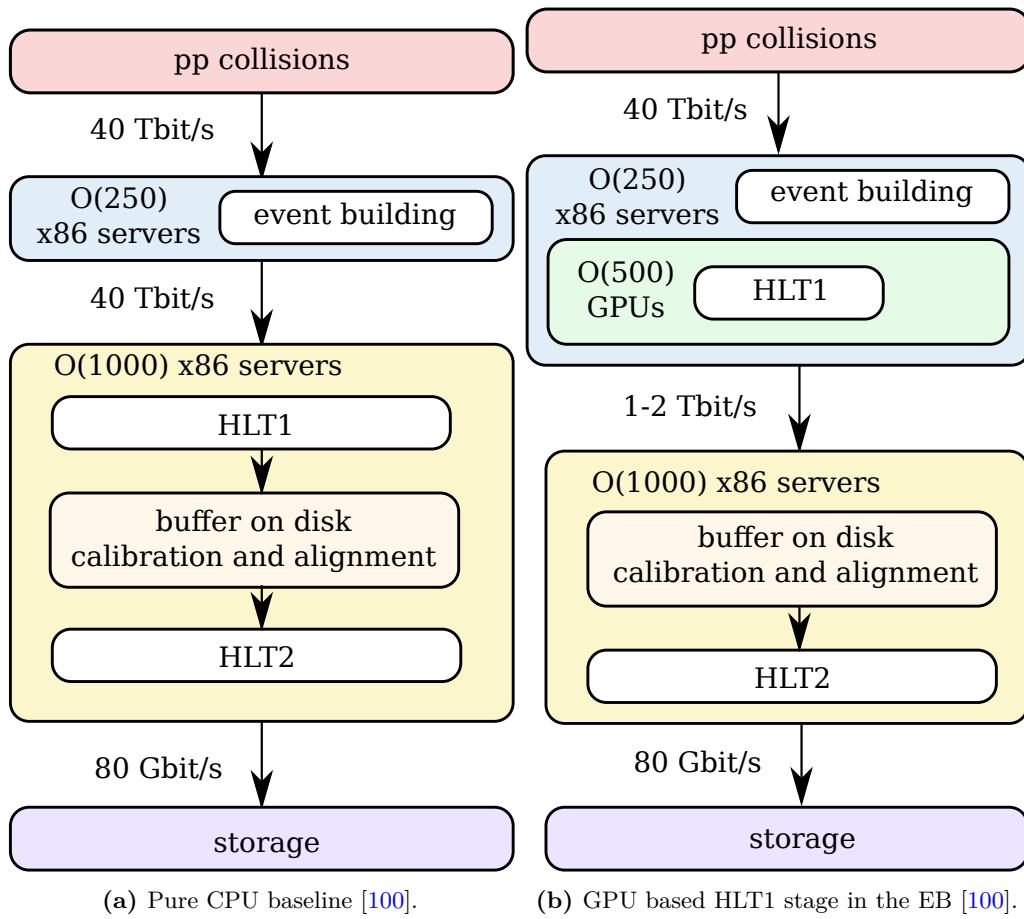


Figure 7.3 – LHCb upgrade farm data processing scheme.

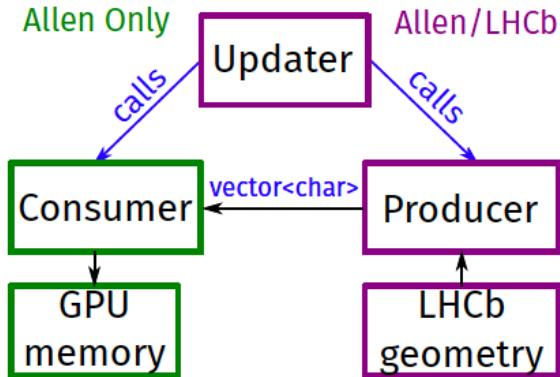


Figure 7.4 – Diagram of the update service of non-event data on the GPU [93].

other LHCb projects. This includes, for example, code reviews and continuous integration tests. It is possible to run Allen standalone and as a GAUDI project. The overall project code can be compiled for a typical x86 platform. Therefore, the physics performance can be controlled by non GPU users. Also, the entire framework is compiled and tested daily on a CPU architecture as part of the nightly LHCb software testing.

The integration of information from the framework during the usage of the trigger is essential. One example is the TCK settings, which have to be available in Allen. There is an update service for the conditions which is illustrated in Figure 7.4. When the conditions are outdated, the updater calls the Producer and the Consumer functions on the GPU. The Producer is using the current information about the geometry and the alignment and puts them into the format which Allen expects. The information also contains the mapping of the SciFi Raw banks. This mapping is needed for the decoding. Another example is the magnetic field map which is essential for the track finding.

The overall architecture of the Allen framework is shown in Figure 7.5. The framework is a multi-threaded system in which every thread steers a CUDA stream. Such a stream is a pipeline for memory operations and algorithm executions on the GPU. A key feature of those streams is the guaranteed asynchronous execution of events. As the reconstruction per event is independent, no cross-stream communication is needed. The input reader provides the MEP data to the Streams. The scheduler runs the defined sequence of algorithms. The sequence can be modified easily; an example is shown in Figure 7.6. An essential part of the Allen project is the MemoryManager. It is connected directly with the Scheduler. Dynamic memory allocations are possible in CUDA, but they require a synchronisation of all streams. Such synchronisations cancel the performance benefits of GPU usage. Before a thread is launched, the required memory is reserved by the MemoryManager. Therefore, all data dependencies in the algorithms need to be known a priori. The memory amount is set to a standard value, but due

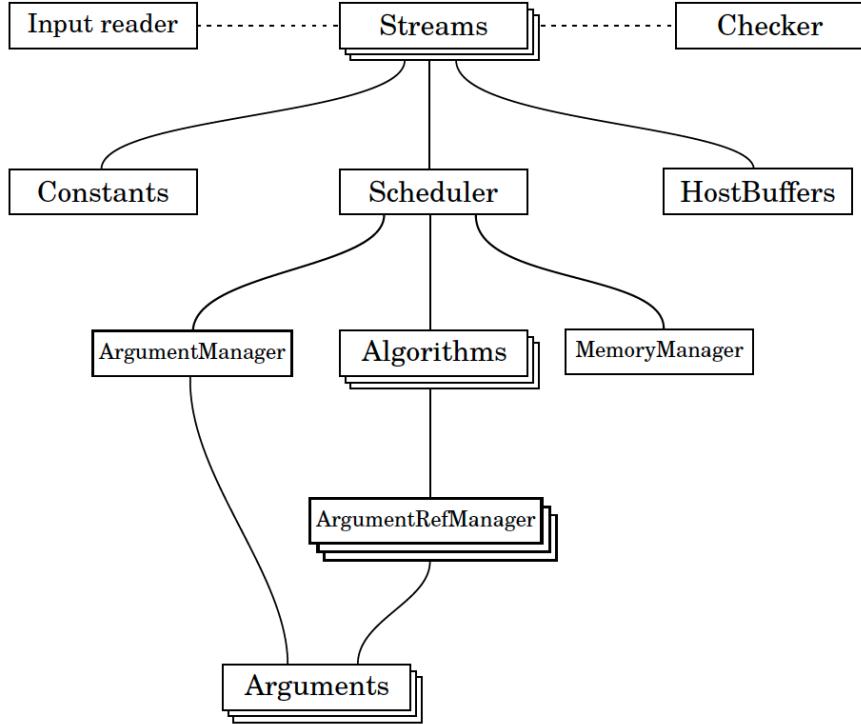


Figure 7.5 – Allen framework schematic. Multiple Streams are executed in parallel. Every stream runs a sequence of algorithms independently. The needed memory is managed by the MemoryManager through the Scheduler. [93]

to the wide range of occupancy in the detector the estimated memory can be insufficient. Therefore, a safety mechanism is implemented. This will detect the lack of memory and rerun the event with more memory space.

The reconstruction of events with very high occupancy takes much longer, because of the increased combinatorics. Due to the significant amount of tracks in a high occupancy event a successful reconstruction of a decay is unlikely. It is expected that most of the high occupancy events are rejected in the later offline analysis steps anyway. Therefore, a global event cut (GEC) is part of the HLT1. The GEC deletes events based on the number of hits in the SciFi and UT. Through this the limited computing capability of the trigger is not wasted for crowded events. The settings of the GEC is adjustable, and standard condition is the cut of the 10 % busiest events.

7.2.2 Allen Sequence

Besides the feasibility of the project and the integration, the performance of the HLT1 GPU trigger implementation is important. This includes the number of processed events and the reconstruction and selection performance. Before numbers

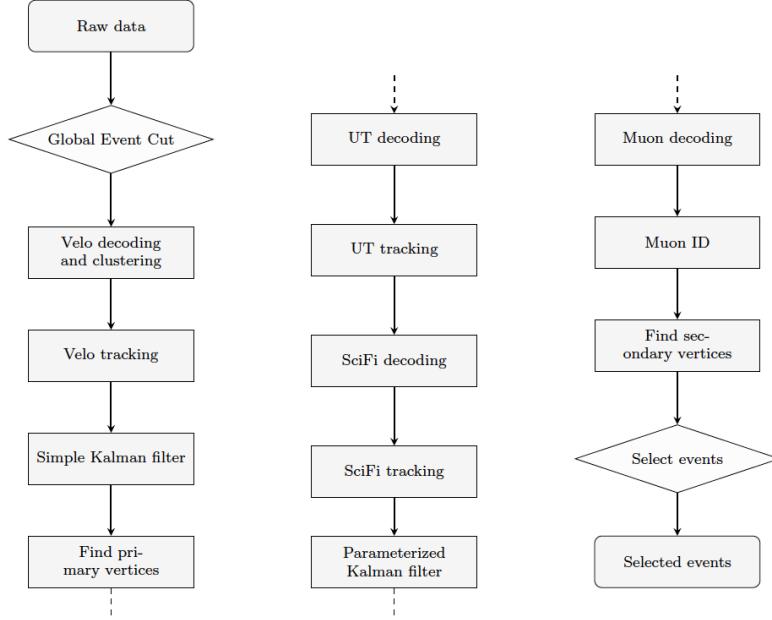


Figure 7.6 – HLT1 Allen standard sequence. Rhombi represent actual trigger decisions, rectangles are data processing algorithms. [100]

for these are presented in [section 7.4](#), the algorithms in the standard sequence of Allen are explained. [Figure 7.6](#) depicts the just mentioned standard sequence.

The data flow within this sequence is described in the following starting in the top left. The raw data packed in MEPs is copied to the GPU. After passing the GEC, VELO hits are decoded and clustered. Triplets of hits are created and used as seeds for straight-line tracks, and this process is parallel. The tracks are fitted by a Simple Kalman filter. Afterwards, a PV search is performed. Hits from the UT also need to be decoded. Based on the x coordinate, the strips of the detector are arranged in regions. For a fast extrapolation of the found VELO tracks, the hits get an additional sorting in y . Next algorithm in the chain is the SciFi decoding, details regarding the implementation follow in [section 7.3](#). The decoded hits are then used to extend the previously found tracks through the SciFi detector. The extended tracks have passed the magnet. The magnetic field map was copied within the conditions to the GPU. Due to this, the track extrapolation can be corrected with the magnetic field. A non-negligible amount of the field is still present inside the SciFi, and this is taken into account by a linear model. With a parametrised Kalman filter, the distance of closest approach from a track to a PV is calculated. Here, only VELO tracks with a momentum estimation from the other steps are used. Also information from the muon stations are decoded. Then, an algorithm calculates the probability that a particle is a muon. This calculated PID information are matched to the tracks extrapolated from the SciFi. For the finding of secondary vertices, two track combinations are build. This is done for

the different combinations in parallel. The secondary vertex provides a momentum and a mass estimate which is used in the selection. Based on the TCK settings, different criteria are applied. As an example, the set of five trigger lines that were used for some performance measurements within the TDR is presented:

- 1-Track: A single displaced track with $p_T > 1$ GeV.
- 2-Track: A two-track vertex with significant displacement and $p_T > 700$ MeV for both tracks.
- High- p_T muon: A single muon with $p_T > 10$ GeV for electroweak physics.
- Displaced dimuon: A displaced dimuon vertex with $p_T > 500$ MeV for both tracks.
- High-mass dimuon: A dimuon vertex with mass near or larger than the J/Ψ mass with $p_T > 750$ MeV for both tracks. [100]

In Figure 7.7, a breakdown of the used resources is shown; the SciFi decoding related algorithms are highlighted. At this stage, it can be concluded that the SciFi decoding uses a significant part of the overall resources.

7.3 SciFi Decoding on GPU

In section 4.6 the three different SciFi decoder versions (v4, v5 and v6) were introduced. In this section, the implementation of these in the Allen framework is presented. The key challenges while porting a CPU algorithm to the GPU are pointed out. Besides the overall requirement to have a fast and correct decoding, the later tracking steps require that the hits are sorted by their x position. Also additional information besides the actual cluster position are needed. Because of the fibre mat tilt in the stereo layer, the y axis start coordinate is provided. For an optimal computing time it is essential only to execute needed computations and avoid idling threads. This is one reason for the split in different kernels. Due to the split the performance of the single kernel can be optimised independently. The entire chain of all kernels is explained on basis of the decoder for version four in subsection 7.3.1. As the skeleton is the same for the versions six and five, only the differences are presented in subsection 7.3.2 and subsection 7.3.3, respectively.

7.3.1 GPU Raw Bank Decoder Version 4

The first step is the calculation of the actual number of clusters within an event, as the memory allocation is needed for the later decoding step. Hereby the calculation is parallelised with the number of events in the MEP, which was copied to the GPU by a stream. Therefore, the `blockIdx.x` is set to the corresponding event number. Within a block, the code iterates over all raw banks and the iterator is

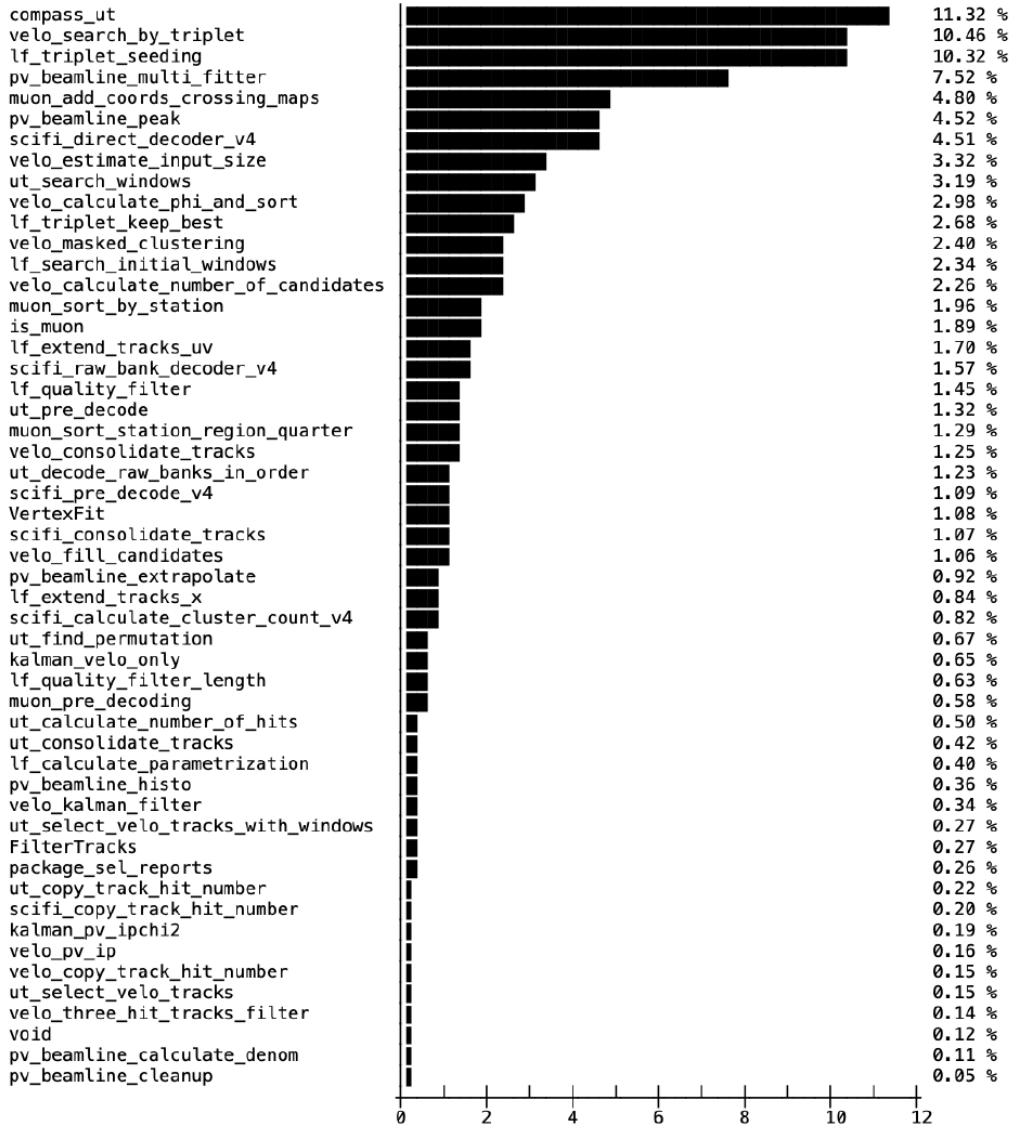


Figure 7.7 – Breakdown of the used resources in the TDR Allen sequence [93].

set to `threadIdx.x`. Besides the actual cluster counting, a first ordering is applied. The raw banks are ordered by their x position within the detector. This is needed as the modules are counted from the centre within each quarter and this is not consistent with the normal x axis. After this step, the total amount of clusters is known and the raw banks are ordered. In the next step, the pre-decoding a data arranging takes place. Thereby no actual data is shifted, only the memory references are arranged. This step is needed because the third SciFi station is wider than the other. Figure 7.8 shows the readout scheme of the third station. There

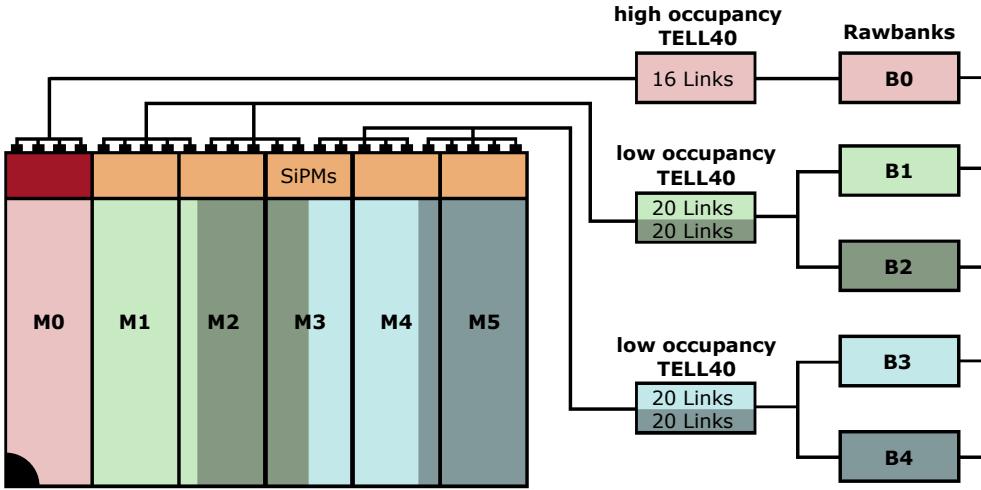


Figure 7.8 – Raw bank cabling scheme of the third SciFi station quarter 4. Taken and modified from [69].

are more modules than raw banks. Therefore, the convention that the first mat of a raw bank is also the first mat in a module is not valid any more. For this reason, the parameter first mat per raw bank was introduced within the ReadoutMap in subsection 4.6.1. With this value, a unique matID is generated. Based on this, the actual mat numbers are corrected. In the end, the arranged clusters reference information are stored to be passed to the next step. Of course, the execution of the pre-decoder is also parallelised the corresponding code snippet is:

```
//...
const uint number_of_events = gridDim.x;
const uint event_number = blockIdx.x;
const uint selected_event_number = event_list[event_number];
//...
```

The next and last step is the actual raw bank decoding. Here, the transformation of the clusters information into geometric hit coordinates takes place. After the previous steps, it is now possible to take the cluster data and store the calculated hit in the output memory. Only the `threadIdx.x`. is needed to store the decoded hit at the right position in the memory.

The actual decoded hit is defined as:

```
struct Hit {
    float x0;
    float z0;
    float endPointY;
    uint32_t channel;
    uint32_t assembled_datatype;
}
```

These are the variables required by the tracking. At this stage, additional information from the detector geometry is needed, these can be accessed through *geom*. Here the physical start position of the mirror per mat and the slope of the mat are used. The SiPMs are made of two halves with a gap in between, cf. [Figure 4.5](#). If the SiPM channel is from in the second half, the size of the die gap is added to the channel position value. The corresponding code for these calculations is:

```
//...
const float dxdy = geom.dxdy[mat];
const float dzdy = geom.dzdy[mat];
float uFromChannel = geom.uBegin[mat] + (2 * id.channel() + 1 + fraction)
    * geom.halfChannelPitch[mat];
if (id.die()) uFromChannel += geom.dieGap[mat];
uFromChannel += id.sipm() * geom.sipmPitch[mat];
const float endPointX = geom.mirrorPointX[mat] + geom.ddxX[mat] * uFromChannel;
const float endPointY = geom.mirrorPointY[mat] + geom.ddxY[mat] * uFromChannel;
const float endPointZ = geom.mirrorPointZ[mat] + geom.ddxZ[mat] * uFromChannel;
const float x0 = endPointX - dxdy * endPointY;
const float z0 = endPointZ - dzdy * endPointY;
//...
```

These calculations are not GPU specific and similar to the CPU decoding. The TDR standard sequence contains a kernel, which was not mentioned yet. This is the so-called direct decoder. It is an approach by the collaboration. The pre-decoding step is only needed for the arrangement within the third station. Therefore, the decoding sequence is split. The clusters from station T1 and T2 do not need a pre-decoding and are decoded directly in a separate kernel. But the clusters from T3 are passed to the pre-decoder and the standard decoder. As a sanity check, the GPU algorithm's decoded hits have been compared to the output of the CPU decoder, and no deviations occurred.

7.3.2 GPU Raw Bank Decoder Version 6

The overall concept of the decoding for the version v6 is similar. The sequence is: calculate cluster count, pre-decoder and raw bank decoder. For this version, a direct decoding is not possible, due to the fact that clusters can be recovered,

cf. subsection 4.6.2. For the detection of such clusters, the size and fraction bits within the clusters data are analysed. Therefore, it is not possible to map directly the input data to an output. It is possible to decode the hits like in the version 4 and ignore the recovering first. Afterwards, in an additional step, the missing cluster can be recovered. But the performance of this approach is expected to be insufficient, because the clusters have to be arranged in the overall array of hits at the right position. If one cluster is added, all following ones have to be shifted. In general this is possible, but such a memory access process is expensive in the sense of run-time. Another problem of this approach is the thread safety. When the recovering step is parallelised, the different threads needs to synchronise and wait for each other. The expected slow run-time of this approach was validated.

As the pre-decoding is needed for the third station anyway an integration of the recovering step in this kernel is tested. Beforehand a change in the cluster count kernel is needed. The naive count of cluster data frames from v4, would underestimate the cluster amount. After the pre-decoding, the cluster amount is given per module. Here the affiliation of cluster to different modules from raw banks within the third station is taken into account. The actual hunt for cluster to recover is added to the cluster count function. This kernel iterated over all hits from a raw bank anyway. After the description of the sequence architecture, details on the kernel processes are explained. If the case occurs that the size flag of a cluster is 1, the size flag of the next cluster is checked. When this is also flagged as large, the cluster channel distance is calculated, and based on this value the amount of clusters which will be recovered is defined. When the distance is not greater than eight, no cluster is recovered. The distance is calculated by the channel numbers. This procedure of checking the amount of clusters to recover does not affect the parallelisation of the calculate cluster count kernel. With the known amount of clusters, the pre-decoder and raw bank decoder kernels are invoked similar to v4. Enough space for the recovered cluster is reserved, but the usual direct mapping from cluster data to the converted hit will not work. Only the amount of expected clusters was increased but not the number of clusters in the data. Within the pre-decoder, the data references for the recovered cluster are generated. After this step the amount of cluster references is equal to amount which was calculated within the cluster count. This means that every recovered cluster has a pair of memory reference for the decoded hit and reference to the clusters data.

Because the raw bank decoder directly maps the information which are provided by the *store sorted cluster reference* step, the different cases of cluster are identified there. For every of the five possible cases a unique value is added to the hit data. In addition to the case a value, named delta, is stored for the cluster channel calculation in the decoder.

After this, the actual raw bank decoder can be invoked. Similar to v4, the kernel contains the decoder and a *make cluster* function. Within the last mentioned function there are no changes, because the geometry of the SciFi tracker is not changed and the computing task is the same. Nevertheless, the decoder is different. Depending on the case value stored in the cluster data, the case corresponding

functions are called.

7.3.3 GPU Raw Bank Decoder Version 5

The procedure of the GPU based SciFi hit decoding for version 5 is similar to the just described implementation of v6. This is due to the fact that the versions only differ in some definitions. The two flags (large, fraction) are used in a different way. Therefore, a second condition value is added to the hit data in the pre-decoding step. The calculation of the cluster channel number for fragmented and recovered clusters is numerically not identical. But these differences have no major effect on the overall performance of the algorithm.

7.3.4 Conclusion of the GPU based Raw Bank Decoder

All three versions of the GPU decoder are functional. Within the TDR, only version 4 is used. This is because this data format version was standard for a long time and v5 and v6 are new. Therefore, the major part of the MC simulation sample for the upgrade are based on v4. It was important to show that the decoding is also working for the other versions. It is likely that v4 will not be the standard version for the entire Run III. Because when version 4 is used, then a small fraction of information is lost. This in turn leads to a physics performance reduction. The information loss comes from the fact that the clusters are not merged and no recovery is applied. The fraction of these clusters will become higher during the lifetime of the detector, because of the increasing SiPM noise rate. These processes, merging and recovering of clusters, influence the performance of the decoding significantly. Because of the extra calculation within the pre-decoder the sequence is slower. Additional studies are needed to evaluate the trade off between physics performance and throughput when v4 or v6 is applied. Another factor which is slowing down the performance of the decoding is the arrangement of the raw banks in the third station. With a SciFi backend electronic extension, as illustrated in [Figure 7.9](#), it would be possible to use the direct decoder for the total SciFi and not only for the first two stations, under the condition that v4 is used. Through additional raw banks for the third station, the assignment of the first mat in a raw bank would be unambiguous again. This idea requires additional TELL40 boards. Therefore, additional cost performance benefit studies are necessary. Besides changes of the detector readout, modifications in the data format of the SciFi are possible. At the moment a new version, v7, is under development. A new data ordering is planned within v7. With an unique ordering, the arrangement and ordering steps becomes unnecessary.

7.4 Allen Performance

In this section the performance of the Allen project at the time of the TDR is presented. The tracking efficiency for *forward* tracks is shown in [Figure 7.10](#) and

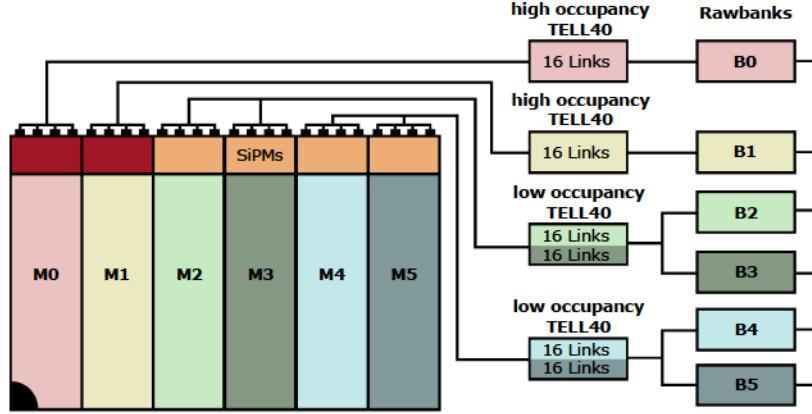


Figure 7.9 – Proposal for additional raw banks in station 3 of the SciFi. The overlap from raw bank over multiple modules is avoided. Taken and modified from [69].

the overall efficiency of the GPU sequence is comparable with the corresponding CPU sequence.

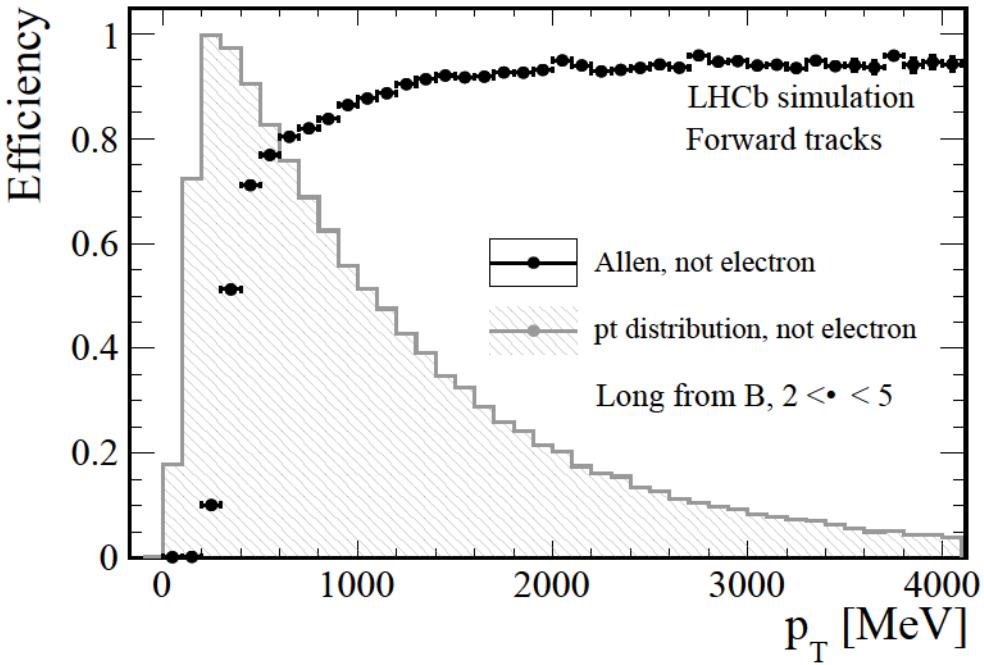


Figure 7.10 – Allen tracking efficiency for *forward* tracks [101].

Another important factor besides the tracking efficiency is the event throughput of the HLT1. Therefore, Figure 7.11 shows the throughput rate in kHz for different devices. Besides the four GPU, the sequence was also executed on a CPU. The conditions for these measurements are 12 input streams with 1000 events each and

7 GPU Based Trigger

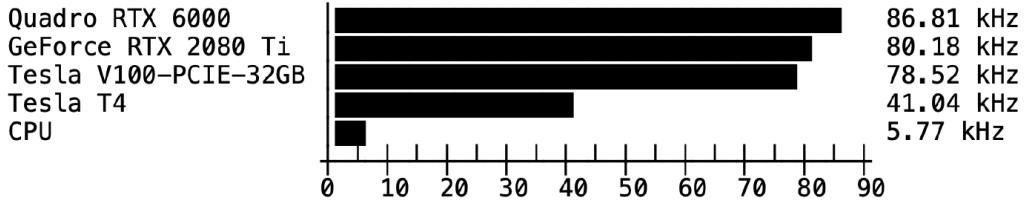


Figure 7.11 – Throughput of the HLT1 Allen sequence for different architectures [93].

this calculations are processed 100 times. For the total input rate of 30 MHz a single GPU needs a throughput rate of at least 60 kHz, assuming 500 GPUs in total. Three of the listed GPUs are above this lower throughput limit. From these only the V100 tensor core GPU is designed for scientific usage as the total tesla series. The GeForce RTX series contains gaming cards an the Quadro RTX are designed for professional image processing applications, and the used CPU was an AMD EPYC 7502. At the first glance it is surprising that the two consumer GPUs have a comparable or even better throughput than the professional GPU. But the computations which are part of the sequence are not complex, therefore the full card potential can not be used. The scientific cards are made for complex calculations mainly the training of neuronal nets. The same trend can be seen in [Figure 7.12](#), here the trigger rate of different GPUs is plotted against their theoretical computing capability given in floating point operations per second (FLOPS). It can be seen

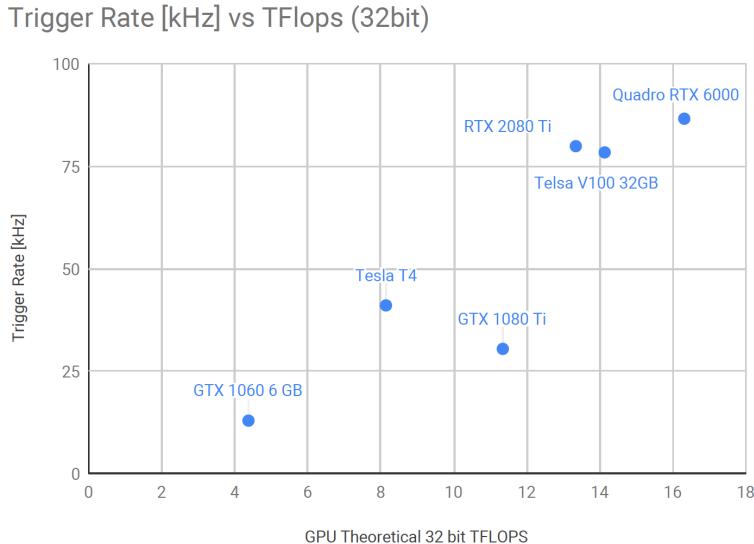


Figure 7.12 – Throughput of the Allen TDR sequence for various GPUs over their theoretical performance in FLOPS [93].

that the trigger rate scales almost linearly with the theoretical performance of the GPU. It is expected that this trend will stay the same for the next generation of

GPUs. Therefore, the plan of the collaboration is to postpone the GPU purchase as much as possible. In the previous part of this thesis, it was claimed that an increased amount of hits in the SciFi will cause a trigger throughput reduction. Figure 7.13 shows the throughput of the Allen sequence on a Tesla V100 GPU as a function of the SciFi raw data volume. For these measurements the GEC on the 10 % busiest events was enabled and this lead to a throughput drop of about 20 % [100]. All events used for this performance measurement are simulated with standard detector conditions. But the fundamental dependency between hit amount and throughput is expected to be the same after the planned data taking of 50 fb^{-1} . On average the number of hits will be 9 % higher, this is a result from the ageing studies cf. chapter 6. This increased number will shift the total throughput spectrum more to the right.

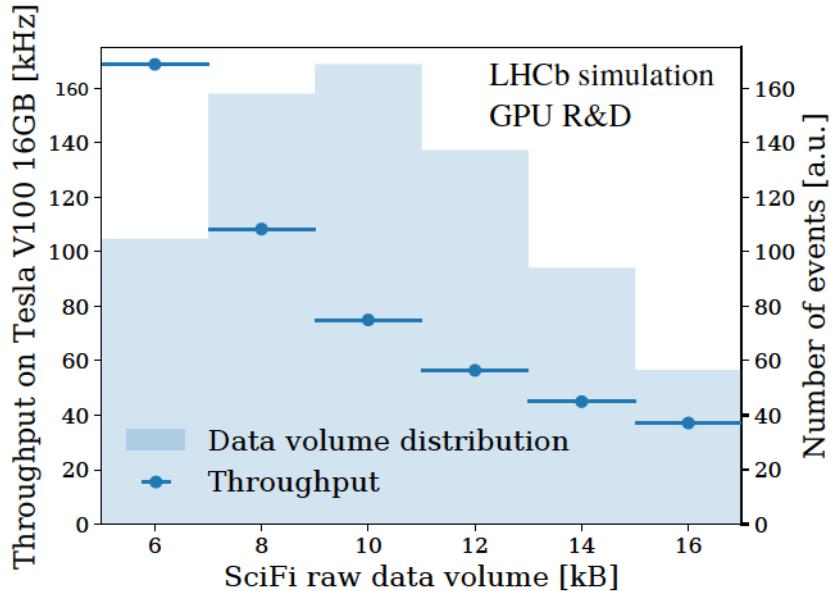


Figure 7.13 – Allen sequence throughput in bins of the SciFi raw data volume. The GEC was deactivated, so the busiest events are not removed. [100]

7.5 Conclusion and Outlook

In this chapter, the general concept for the usage of GPUs for non-graphics calculations was introduced. Computing tasks which can be parallelised are ideal for execution on a GPU. Because of the GPU architecture and processing power, a speed-up in the execution time can be expected. The basic programming concepts transforming CPU to GPU code were shown based on NVIDIA's CUDA framework. Afterwards, the Allen project of the LHCb collaboration was introduced. The idea of the project is a GPU based implementation of the HLT1 stage for the upgraded

7 GPU Based Trigger

LHCb detector. At this trigger stage an average data rate of 32 Tbit s^{-1} needs to be processed.

The baseline concept for the upgrade trigger was foreseen as a CPU farm. For the trigger baseline change, the Allen group has to show a significant benefit of the GPU trigger compared to the CPU one. Therefore, a TDR was written wherein all the needed information about the project are collected. This includes the integration of the Allen project with the rest of the LHCb framework and studies about the long term usage of GPUs. Also financial aspects were taken into account. The Allen project is at least cost neutral or even cost saving. Because the GPU are integrated in the event builder nodes, the first data reduction takes places there and not only in the event filter farm. Therefore, the required bandwidth between EB and EFF is lower and a cheaper technology is sufficient. The most important point is the throughput of the GPU sequence, though. For these throughput measurements the complete HLT1 sequence has to be ported to GPU code. One part of this sequence is the decoding of the SciFi clusters. This is done by the SciFi raw bank decoder. A working algorithm was implemented for this decoding. The decoding sequence consists of three steps. These are: calculate cluster count, pre-decode and raw bank decoder. Thereby the first kernel counts the clusters and reserves memory for the decoded hits. The second kernel arranges the data, which is unsorted overall, because of a disadvantageous raw bank structure in station 3. The decoder sequence is available for three different SciFi raw data format versions. The versions differ in the data format and support different feature. The newer versions 5 and 6 can recover hits. A major problem thereby is the fact that the amount of hits to recover is unknown *a priori*. This is problematic for the parallel memory access. An insertion of hits is in this parallel process not suitable, but all hits need to be ordered by their position. A good performing solution was found. Here the count of cluster to recover is added to the pre-decoder and the order process is a part of the pre-decoder.

Within the TDR it is planned to equip the GPU trigger with 500 cards, this leads to a minimum throughput rate of 60 kHz per GPU. Multiple GPUs were found to reach this value while executing the Allen TDR HLT1 sequence. Besides scientific cards also consumer GPUs are sufficient, for example the RTX 2080 Ti. Because trigger throughput and theoretical card performance scale in a proper way, no cards are purchased. It is planned to wait as long as possible to save cost and wait for newer GPU generations.

Since the TDR and the decision to use the GPU trigger as baseline project for the HLT1 significant changes and developments happen. Due to server technology performance growth the number of server in the EB is reduced from 500 to 170. Here every server can host two GPUs. This reduces the total number of GPU to 340. Nevertheless, this number of cards is sufficient to achieve the needed throughput in the HLT1. The overall performance of the Allen project was increased significantly. At the end of 2020, the performance is more than doubled compared to the TDR [102].

8 Diamond Measurements for the BCM Upgrade

The Beam Condition Monitor was already introduced briefly as a part of the LHCb detector safety system in the upgrade detector description, cf. [section 3.3](#). In this chapter, the plan for an upgrade of this component is presented, cf. [section 8.1](#). Therefore, a brief introduction to the usage of diamonds for radiation detection is given in [section 8.2](#). Afterwards, measurements for the diamond characterisation are shown in [section 8.3](#). The possible usage of these measurements to characterise diamonds for the BCM upgrade are concluded in [section 8.4](#).

8.1 BCM Upgrade

The BCM at the LHCb detector measures the particle flux around the beam-pipe continuously. This is needed as components, especially the VELO, are placed close to the beam. The distance of the VELO sensors is only 3.5 mm and a small undetected deviation of the beam position will destroy the detector. Synthetic diamonds are used as active material in the BCM. They detect the fluence of charged particles, more details regarding this are given in the next section. When the BCM detects an increased particle fluence, it initiates a beam dump. Therefore, the BCM is connected to the Controls-Interlocks-Beam-User system of the LHC [103]. The beam dump threshold values are not defined yet for two reasons. First, the BCM simulation within upgrade conditions is not completed. Second the efficiency of the diamond sensors is not known. Nevertheless, an estimation based on the values from Run II leads to thresholds of about 10 nA.

A beam dump affects all LHC users significantly because a lot of data recording time is lost. The filling of the LHC with proton bunches and the following acceleration of them takes a few hours. Therefore, false-positive beam dumps are undesirable. Nevertheless, the false-negative case is worse, as detector components get destroyed. That's why the sensor characterisation is important.

During the operation time of the BCM, no significant problems occurred, but at the end of Run II the signals from two diamonds showed excessive currents. Therefore, these were excluded from the BCM beam-dump decision logic. In general, diamond sensors are known to be radiation hard, which is the reason for this technology was chosen [104]. Nevertheless, one possible reason for the strange behaviour of the problematic diamonds can be radiation damages. To ensure a properly working BCM for Run III it was decided to build a new BCM with new diamonds. The basic layout of the BCM, which consists of two stations with eight

diamonds each, stays the same. Since the area near the VELO is also interesting for other small detector systems, the space for the BCM upstream station is more limited than before. For example, the PLUME (Probe for LUminosity MEasurement) detector is planned to be placed there [105].

Besides the pure layout aspects, there are also some parts within the readout system which motivate an upgrade. The signal from the old BCM was processed in a TELL1 board. This board is the predecessor version of the TELL40. As the readout systems of all other sub-detectors are upgraded to the newer version, the BCM would be the only component using the old standard. Due to the readout change, the BCM system will stay easily compatible with the LHCb framework.

The installation of the new BCM is scheduled early in the overall assembly of the LHCb upgrade detector. When the C-side of the detector is ready, this means for example that all C-frames of the SciFi are in place, the beam pipe can be installed again. After this step, it is foreseen to install the BCM. Afterwards, it can be continued with the A-side of the LHCb detector. The development for a setup to characterise diamonds at the university is motivated by two facts: First, the tight detector commissioning schedule and second the interest to study the defective diamonds of the old BCM. Of course, there are possibilities to characterise diamonds in external facilities. For the old BCM, this was done *inter alia* at the ELBE facility [106].

The Dortmund BCM group is developing the new BCM and the new diamond characterisation setup. Therefore, some of the later presented results were already used in corresponding theses [107, 108, 109].

8.2 Diamond as Radiation Detector

Diamond is a solid-state of carbon atoms arranged in a crystal structure. This structure is named diamond cubic and consists of two intersecting face-centred-cubic lattices. The diamond structure is illustrated in Figure 8.1, and every atom has four neighbours, each with a sp^3 bond.

Because of the strong bond in the diamond structure, the energy to displace an atom from the lattice is high compared to other materials. The actual value, as the average of all three directions, is about 43 eV [110]. This is almost twice as high as for silicon. Therefore, diamond is known to be radiation hard. Pure diamond is classified as insulator because of the high band gap of 5.47 eV [111]. Due to the high band gap, the amount of intrinsic charge carriers is low. As a result of this, the dark current is low, which is a benefit for the usage as a detector. This low dark current leads to a higher signal to noise ratio when measuring the sensor's signal. Therefore, it is easier to measure low currents accurately.

Diamonds can also be considered as semiconductors, and this is the way the BCM operates. Because of the described diamond characteristics, usage as a semiconductor is possible without additional doping. Most common semiconductors need a p-n junction, and this is created via doping. The fundamental principle

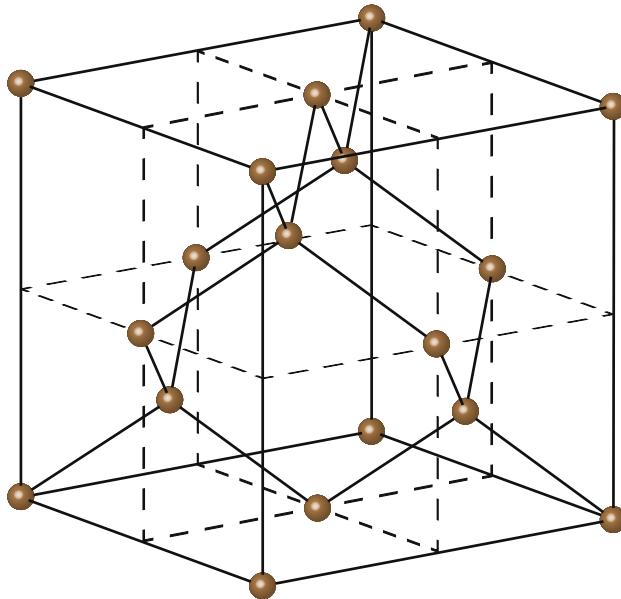


Figure 8.1 – Structure of the diamond cubic crystal system [77].

of this semi-conductor detector type is that a traversing particle interacts with the atoms and deposits some energy in the material. Due to this, electrons are excited and raised from the valence into the conduction band. The underlying processes are described in [section 5.1](#). The electron leaves a hole in the valence band. Both electron and hole will drift when an electric field is present. The mobility of them is slightly different. Electrons have a mobility of $1800 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and holes $1600 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ [111]. Within diamond, an average energy of 13.19 eV is necessary to create such an electron-hole pair [112]. This value is much higher than the bandgap, due to additional energy losses in the material for example by phonons [113]. When the diamond is connected with electrodes, the movement of the charge carriers will effect a signal in the electrodes. This signal current can be measured. The *Shockley-Ramo theorem* describes the signal mechanism quantitatively [114, 115].

8.2.1 Diamond Production

Diamonds that are used as detectors have often special geometric properties. For the BCM small cuboids or plates, with a basic side of $10 \text{ mm} \times 10 \text{ mm}$ and a thickness of 0.5 mm are used. These can be produced from natural diamonds, but this would be very expensive. Moreover, the amount of impurities in natural diamonds is hard to control. Therefore, diamond sensors are made from synthetic diamonds which are known as laboratory-grown diamonds.

A common way to grow diamonds is the chemical vapour deposition (CVD) method. The carbon providing gas is methane. A plasma state is created via

heating by microwaves. A substrate is needed as a starting point where the diamond can grow. The substrate defines the later type of crystal, the diamond can become a polycrystal or a monocrystal. When the substrate has a monocrystalline structure, the diamond will extend this single crystal, and the final product is named scCVD. When the substrate contains diamond powder the resulting structure is a polycrystalline (pCVD). The pCVD diamond is a composition of multiple monocrystals with boundaries between them. Because of these boundaries, pCVD show slightly inferior properties as a detector than scCVD. It is possible that electrons and holes recombine at a boundary. But the production of pCVD is simpler than scCVD, and therefore they cost less compared to scCVD. Another advantage of the polycrystalline ones is their maximal size. For them, a wafer size up to 140 mm is possible while scCVD are limited to 8 mm [116]. For this reasons, the BCM consist of pCVD diamonds.

Another more common method for the diamond synthesis is high pressure high temperature [117]. These production variant mimics the conditions in which diamonds are formed naturally. A disadvantage of this method is the higher incorporation amount of foreign atoms. The higher purity of pCVD is another reason to use them instead of scCVD. A typical impurity is nitrogen and diamonds are classified by their nitrogen content. When the number of nitrogen N is very low, $N < 5 \text{ ppm}$, the diamond is type IIa. The pCVD diamonds belong to this type, while the high pressure diamonds belong to type Ib. An overview of the different categories is shown in Figure 8.2.

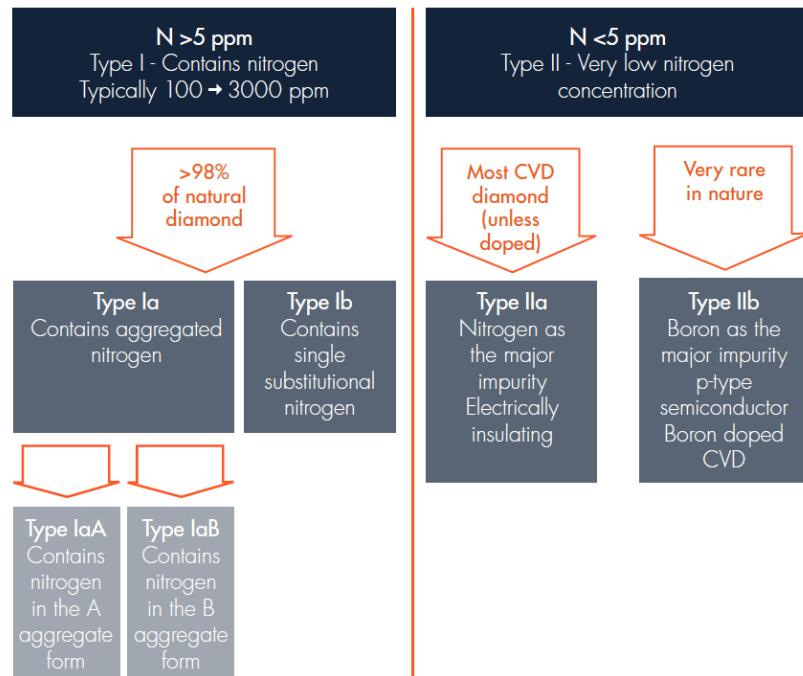


Figure 8.2 – Diamond purity categories based on the nitrogen concentration N [116].

8.2.2 Diamond Surface and Contacts

As mentioned before diamond sensors need contact electrodes to provide the signal for a measurement device. In case of the BCM diamonds, metal is applied to the two large sides. The metal is applied via vapour deposition. Typical contact metals are gold or platinum, but due to flat diamond surface the mechanical adhesion is low. For a proper bound of the contact material an intermediate layer is applied. For the BCM diamonds, this layer is made from titanium. But for the diamonds which are studied in this thesis, another approach is used. The interface between the platinum-gold contact is diamond-like carbon [118]. This is an amorphous carbon variant which also provides sufficient contact. Thereby only the centre 8 mm × 8 mm of each diamond side is provided with the contact material and not the total area.

8.2.3 Diamond Sensor Performance Characteristics

Different features are used within the characterisation of diamond sensors. One of them is the charge collection distance (CCD) d_{sep} . It represents the mean free path length of the charge carriers. Here the mobility μ is multiplied with the mean lifetime τ for both electron and hole. The CCD is calculated by

$$d_{\text{sep}} = (\mu_e \tau_e + \mu_h \tau_h) \cdot E , [111] \quad (8.1)$$

where E is the electric field strength of an applied field.

Another characterisation parameter is the charge collection efficiency (CCE). The CCE is the ratio of detected charge Q_{det} and released charge Q_{rel} . In the case of parallel electrode plates and a detector thickness d , which is much bigger than d_{sep} , the CCD can be approximated as

$$\text{CCD} \approx \frac{Q_{\text{det}}}{Q_{\text{rel}}} \cdot d = \text{CCE} \cdot d , [111]. \quad (8.2)$$

The conditions for the approximation are valid for the BCM diamonds. Also the dark current is a factor which quantifies the quality of the diamond structure. More defects within the crystal structure lead to a higher dark current.

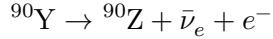
8.2.4 Strontium-90 Radiation

Radioactive nuclides are a common source for a particle beam in laboratories. For the later shown measurements a ^{90}Sr was used. Due to the two beta decays in the decay chain it is an electron source. Because of their energy range, the electrons are considered to be minimal ionising particles.

With a half-life of 28.8 a the Strontium decays into Zirconium via Yttrium



and an energy of 549.9 keV is shared among the electron and the neutrino [119]. The following decay



has a half-life time of 64 h and the released energy is 2278 keV [120].

The energy spectrum was simulated with Geant4 as part of a thesis of the Dortmund BCM group [108]. This spectrum is plotted in Figure 8.3. In 1980

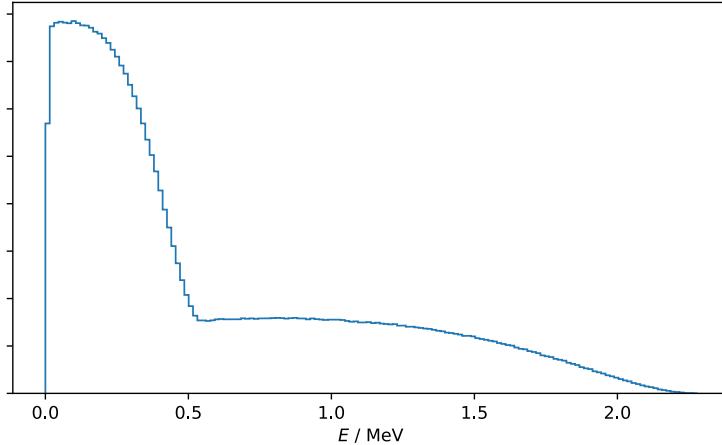


Figure 8.3 – Simulated ^{90}Sr energy spectrum [108].

the ^{90}Sr source had an activity of 33 MBq which leads to an activity of about 13 MBq in 2020, where the measurements were done. The source is provided by the dosimetry company PTW and has the specification number T48010. A sketch of the source is shown in Figure 8.4.

8.3 Diamond Measurements

In this section measurements of 19 diamonds named D01-D19 are presented. They have the same geometric properties as the BCM diamonds. They come from a spare production batch in 2010 and were never used in the BCM and not radiated anywhere else.

8.3.1 Current-Time Curves

Absolute measurements of CCE and CCD are not possible yet with the available setups in the Dortmund laboratory. Therefore, measurement results are compared relatively between the different diamonds. With the setup described later, current-time curves were recorded. These measurements can be done with and without a particle beam. A measurement without beam provides the dark current as a function of time, this should be low, while in the measurements with beam a high current is appreciated for good diamonds. At a given particle flux, a high current

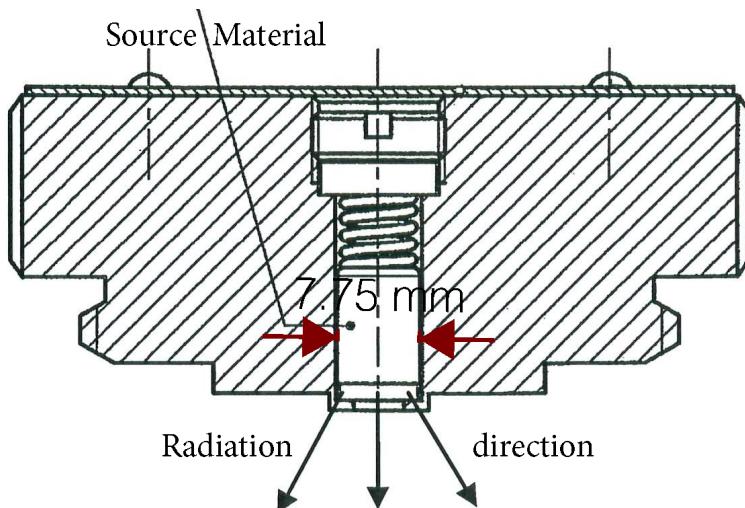


Figure 8.4 – Geometric properties of a T48010 ^{90}Sr source. Taken and modified from [121].

means that the CCE of the diamond is high. The first setup was developed in an earlier thesis [107]. As the setup showed some disadvantages, the following lines describe the further improvements.

As both, dark current and signal current, are low about a few nA, the overall setup is placed in a Faraday cage. This grounded aluminium box also provides a radiation shielding. Within the cage there is a support structure to place the diamond under test. A clamp is applied to the contact sides. The signal is provided to a picoamperemeter of type Keithley 6487 [122]. It is also used to provide the operating bias voltage to the diamond. Next to the diamond support structure a mount for the ^{90}Sr source is placed. The mount has a cutout for the source. Thereby the source is centred above the diamond. In addition the mount is connected to a sledge which is controlled by a stepping-motor. Through this the distance between diamond and source can be varied. The entire setup is managed via a LabView¹ program.

The normal BCM operating bias voltage is 200 V. Therefore, all measurements are done at this voltage level if nothing contrary is said. Measurements are done with both positive and negative polarity. Due to the pCVD production process the diamond has a growth direction and this direction is marked by the manufacturer. With the polarity change it is possible to measure in both directions without moving the diamond or removing the clamp.

Former studies showed a strong time dependent signal after start of a measurement. When the bias voltage is applied to the diamond and the source is in place the measured current increases significantly. After a certain time the current saturates. Therefore, a waiting time of 30 minutes is applied in the data

¹Laboratory Virtual Instrument Engineering Workbench

8 Diamond Measurements for the BCM Upgrade

acquisition sequence. This concept is known as pumping. The traversing source particles produce free electrons in the diamond. These electrons travel through the diamond and some of them are captured by particle traps. These traps are mainly lattice defects. After the waiting time almost all traps are filled with electrons. Thenceforth the particle beam produce an adequate constant signal.

In the standard measurement the diamond is measured at three different distances to the source. First at 5 mm, than 10 mm and in the end at 15 mm distance. The distances are varied to enable measurements with different particle fluences. After the 30 minute waiting time the current is measured every second for 60 minutes. Such a standard measurement is shown in [Figure 8.5](#) for diamond D03.

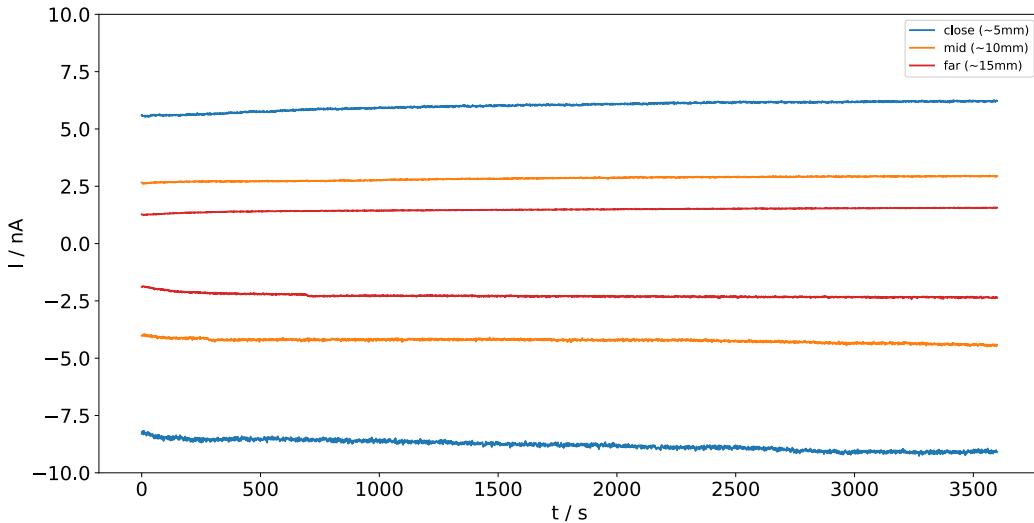


Figure 8.5 – Current of diamond D03 at three positions. The bias voltage is 200 V for the three upper curves and -200 V for the lower ones.

When the current is compared at equal source distances, it is striking, that the absolute value is higher for the negative polarity. This trend is supposed to be caused by the diamond growth direction. In addition, unequal *Schottky* contacts can cause this. Both reasons are unwanted, therefore the polarity trend can be used as an additional feature to characterise the diamonds.

Selected diamonds were measured again after some days, to ensure a reproducible measurement setup. In [Figure 8.6](#) the overlay of two measurements of diamond 02 is shown.

Especially at the closest position with negative polarity, coloured blue, the curves have a non negligible gradient. This shows that the pumping is not completed finished after 30 minutes. But this effect is negligible within the last 100 data points which are averaged in the later analysis steps. In general the curves of the second measurement match the results from the first measuring run. At the end of the measurement still a small offset between the corresponding curves is visible. Here the result from the second run is always higher than the first. This

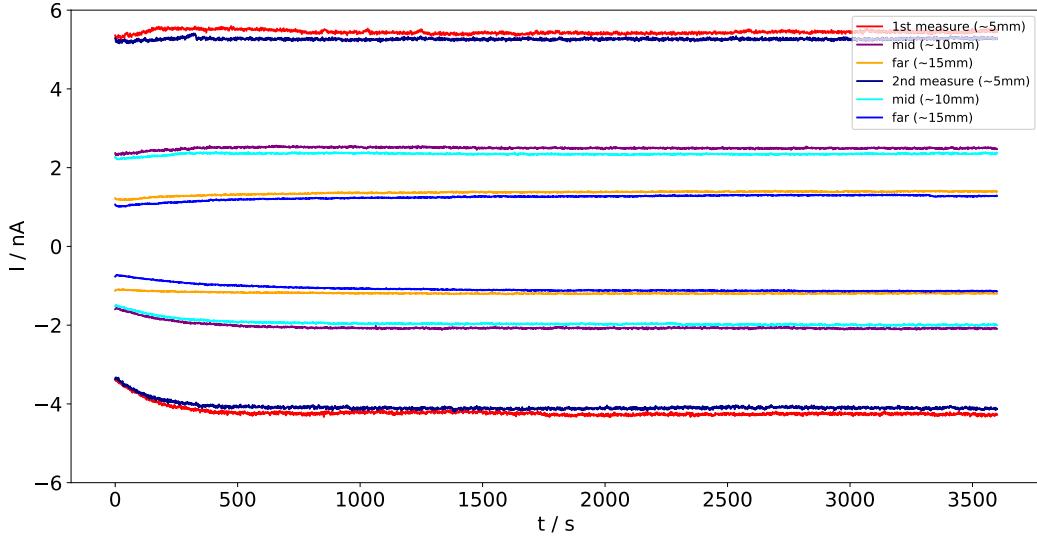


Figure 8.6 – Comparison of two measurements with diamond 02 for both polarities.

is probably within the systematic uncertainties of the setup. The time between the measurements are some days and the diamond was removed from the setup in the meanwhile. Therefore, one reason for the deviation is the manual placing of the clamp. A small clamp is used to minimise effects by blocking the beam with the clamp material. As a result the contact area between clamp and diamond is small. Also the force of the clamp is important for the contact quality. This contact problem is avoided in later studies at different setups by applying an Indium foil between diamond and a clamp [108, 109]. Another uncertainty is the distance positioning of the source. The slide performs a reference run before every measurement run. Thereby the stepping motor goes to a mechanical limit switch. This mechanical switch and slip of the sledge itself lead to a small position uncertainty. Due to the experience with the stepping motor, this error is expected to be within the micrometre range. Besides the systematic effect from the setup itself, the laboratory environment can cause current deviations. A slightly changed room temperature will affect the measured current. The dark current of the diamonds within the shielded setup is below 1 pA. The values were measured with applied bias voltage in both polarities but without source. As the dark current is three magnitudes smaller than the signal current at the greatest source distance, the dark current is neglected in the following.

For a better understanding of the effect of the source position uncertainty a depth profile is measured. Therefor a measurement every 0.5 mm was performed in the range of 5 mm to 15 mm. This was done only for the negative polarity with diamond D17. The results of these measurements are shown in [Figure 8.7](#).

For the comparison of different diamonds and measurements the last 100 current values are averaged (I_{inf}). Thereby the effect from statistical fluctuations is minimised. In [Figure 8.8](#) the I_{inf} values of the depth profile measurement are plotted

8 Diamond Measurements for the BCM Upgrade

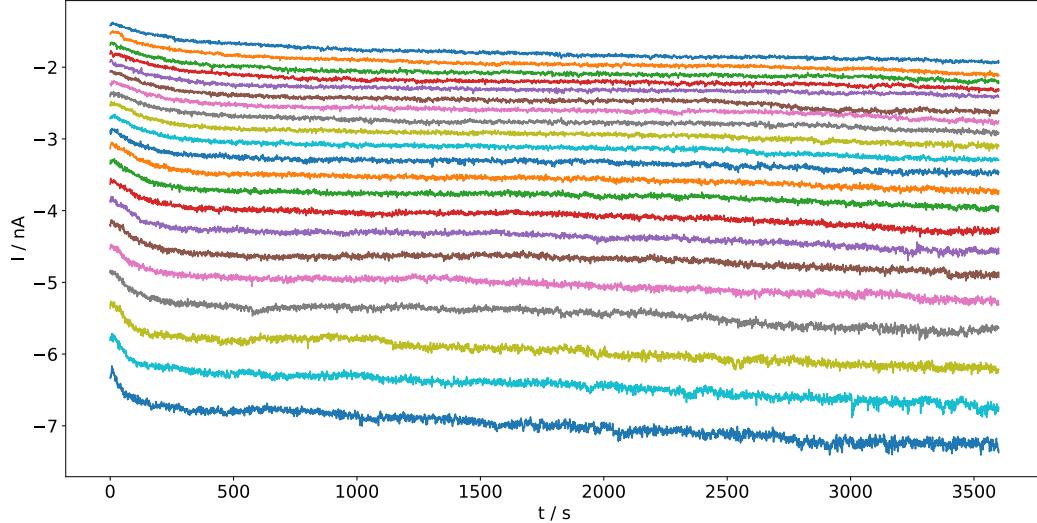


Figure 8.7 – Depth profile of D17 with negative polarity. The source distance ranges from 5 to 15 mm with a step size of 0.5 mm. Each colour represent another distance, and the bottom curve is from the closest position.

against the source distance. An exponential function $f(x) = a \cdot \exp -b \cdot x + c$ is fitted to the data. Based on the fit results the effect of a position uncertainty on the signal current is calculated. The evaluation is done as a worst case scenario at the 5 mm position, because this is the most sensitive position. The result is, that a deviation of about 60 μm leads to a signal change of 1 %.

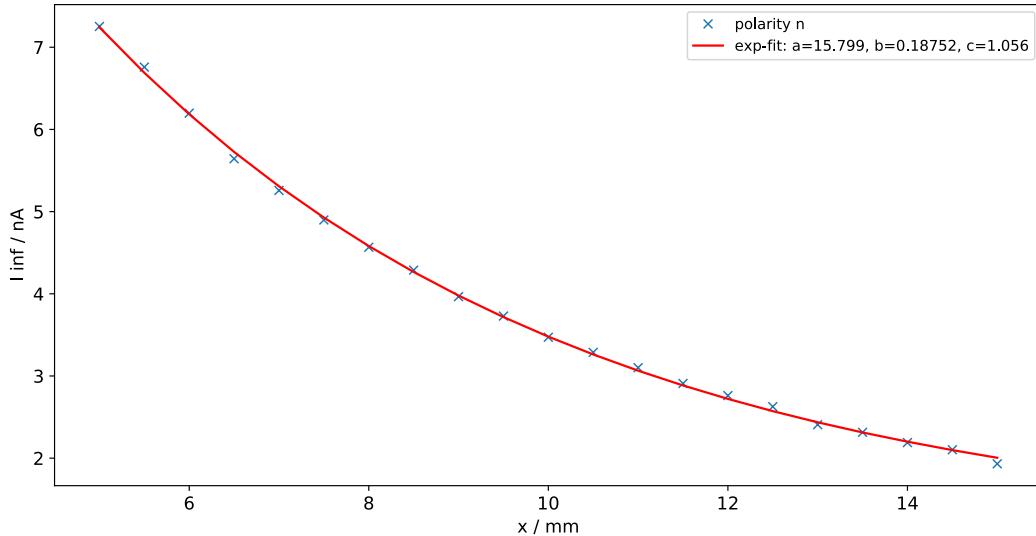


Figure 8.8 – Absolute I_{inf} values from the D17 depth profile.

As the setup is not calibrated, absolute statements are limited. Nevertheless,

for a rough categorisation within a batch of diamonds, the setup is considered to be accurate enough. Therefore, standard measurements with all 19 diamonds are performed and the results are presented in the following. For all diamonds I_{inf} is calculated at both polarities and the three distances. As an example this is shown for the D03 in [Figure 8.9](#). The results from all diamonds at the 5 mm

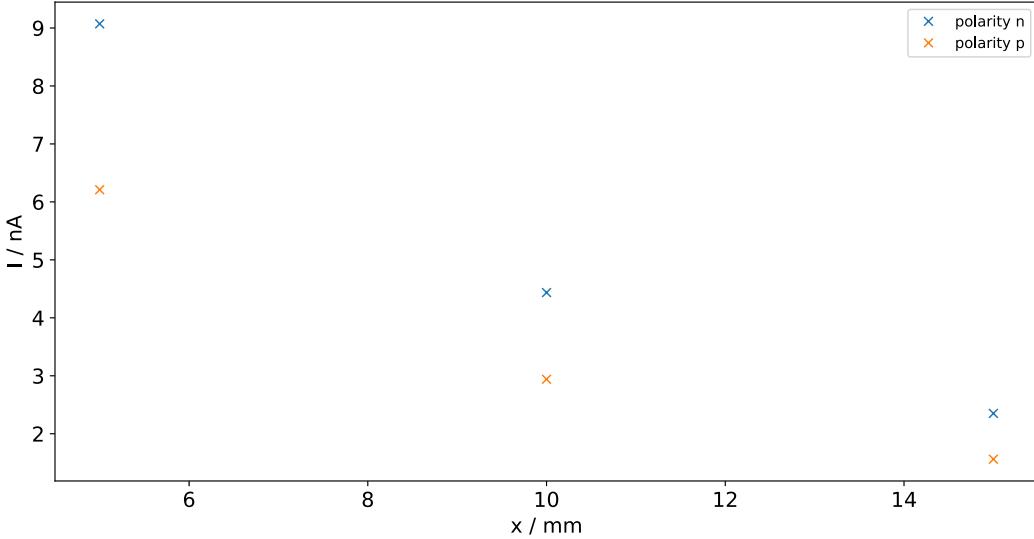


Figure 8.9 – Steady-state current I_{inf} as for different source positions of diamond D03.

position are summarised in a histogram, cf. [Figure 8.10](#). One result is, that the diamonds performance respectively efficiency has a wide range. There is a factor of two between the diamond with the highest and lowest signal current.

Another important fact is that not all 19 diamonds contribute data to this plot. Three diamonds are found to be defect or problematic. This is detected by significant different signal shapes. The corresponding data from D06 is shown in [Figure 8.11](#). The measured current at negative polarity increases abruptly by more than a factor of 10 and fluctuates significantly. This measurement was repeated to exclude external errors. But this strange behaviour occurred also in the repeated measurements for all three broken diamonds. Broken only refers to their abnormal signal, there are no visible defects at the diamond or the contacts. Until now it is still unclear what causes these current breakthroughs. Also the fact that it mainly appears at negative polarity is not understood yet. For positive polarity only little steps show up in the signal shapes.

Beside the fact that the reason is unknown it is likely that this effect is correlated to lattice defects. Such a signal increase without an higher flux can cause an unnecessary beam dump in the BCM. Therefore, it is recommended not to use diamonds with such problematic behaviour. Even if this effect does not occur during data taking, it is likely that these diamonds get earlier effected by radiation damages. This conclusion is passed on the following assumptions: The radiation damage is linear with the dose. When a diamond reaches a certain lattice defect

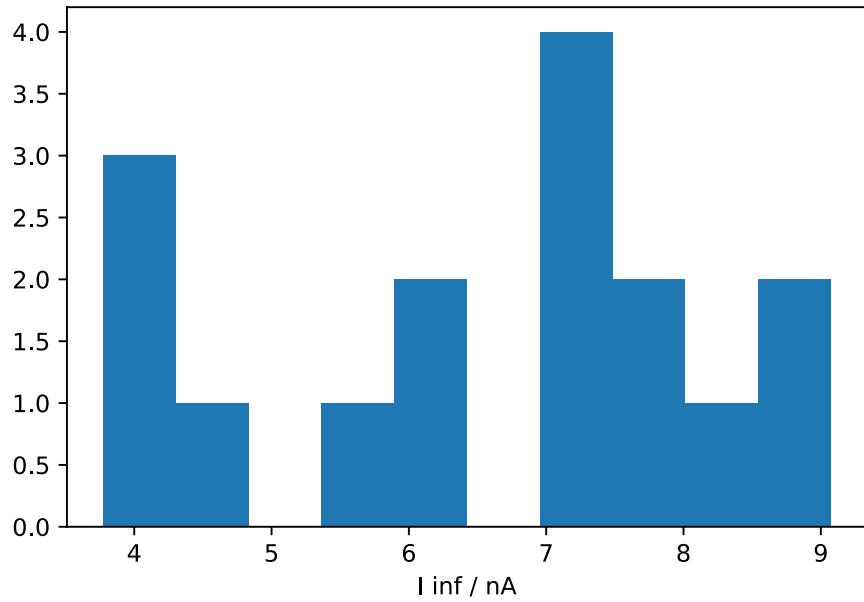


Figure 8.10 – Comparison of I_{inf} at 5 mm distance and negative polarity for 16 diamonds. The data from 3 broken diamonds is excluded.

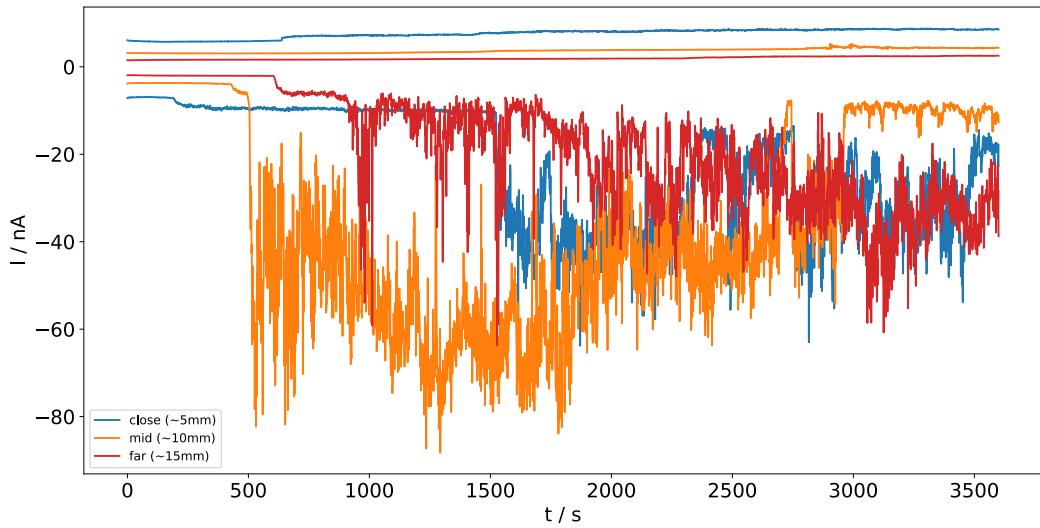


Figure 8.11 – Current time curves of the problematic diamond 06. The top three curves for positive polarity look normal, while the bottom curves for negative polarity have significantly increased currents.

level, excessive signal currents occur. Consequently a bad diamond reaches this level earlier than a good one under equal conditions.

Another result from the measurements is the fact that the preferred direction in the sense of signal level is not always the diamond growth direction. This means that I_{inf} for positive polarity is not bigger than for the negative polarity all the time or vice versa. It can be that the contact on the diamond is the leading factor for the difference between the two polarities. The absolute difference between the two polarities can also be an estimator for the diamond and contact quality. The spread is significantly higher for the diamonds with the highest signal currents.

8.3.2 Proton Beam

Thankfully, the West German Proton Therapy Centre Essen (WPE) provided us the possibility to use their proton beam for some measurements with our diamonds. The results of the WPE and the Dortmund BCM group are not published yet. The studies are ongoing but some preliminary results are shown here. The data at the WPE was recorded together with Martin Bieker.

A proton beam mimics the LHC beam conditions the BCM is monitoring better, than electrons from a ^{90}Sr source. The used beam energy at the WPE was 100 GeV. Other significant differences are the particle mass and the known flux. With a known flux absolute efficiency measurements are possible.

The used setup is shown in [Figure 8.12](#). The diamond holder is mounted on a sledge. Thereby it is possible to measure at different positions within the provided beam profile. The position can be changed remotely during data recording without the need to enter the room. Of course it is not possible to enter the room during beam operations. In the image the beam is coming from the right. The position of the diamond is marked within the picture.

The measurement conditions are equal to the ^{90}Sr source measurement. So the bias voltage is 200 V and the current is recorded every second with the same device. Only the mounting of the diamond is different.

The beam of the WPE is not continuous, it is provided in shots. In [Figure 8.13](#) the data of 10 shots on D13 is presented. The diamond is positioned in the middle of the beam and did not move in between the shots. With a peak detection, the start and endpoint of a shot are detected. The currents in this time window are numerically integrated. This calculated collected charge can be used as a feature to compare the diamonds among themselves.

The possibility to move the sensor lateral in the beam profile is used for two reasons. At first this enables an easy solution to change the proton flux through the diamond. The second reason is to test the resolution of these diamond sensors and crosscheck the width of the provided beam. So the diamond is moved between the single shots to different positions. The data of such an lateral profile scan is shown in [Figure 8.14](#). The blue curve represent the measured current while the diamond position is coloured orange. For the calculation of the beam width the collected charge is calculated for every position.

These data pairs are shown in [Figure 8.15](#). In addition a Gaussian function is fitted to the data and the resulting curve is plotted in blue. This fit determines

8 Diamond Measurements for the BCM Upgrade



Figure 8.12 – Setup at the WPE beam line. The beam is provided from the right side.

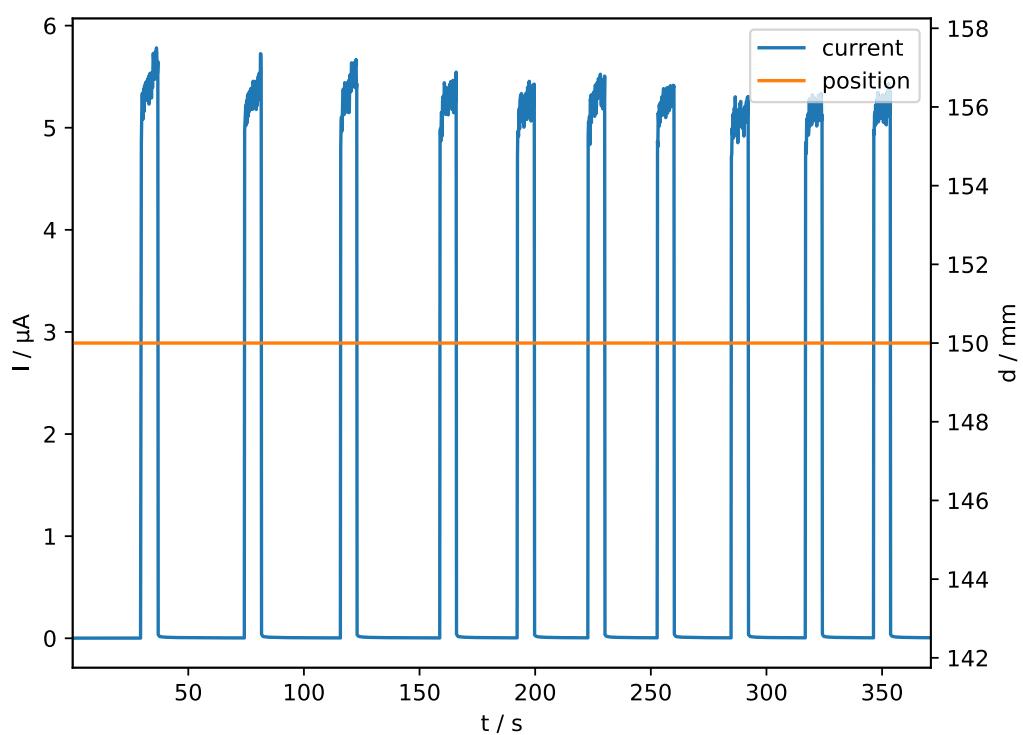


Figure 8.13 – 10 shots with a 100 GeV proton beam on D13 with constant diamond position.

the width of the beam to $\sigma = 13.6$ mm. This is a little bit higher than the value of 12.3 mm which is provided by the WPE MC simulations. Due to the dimension of the diamond a deviation is expected. The spacial resolution of the diamonds is not

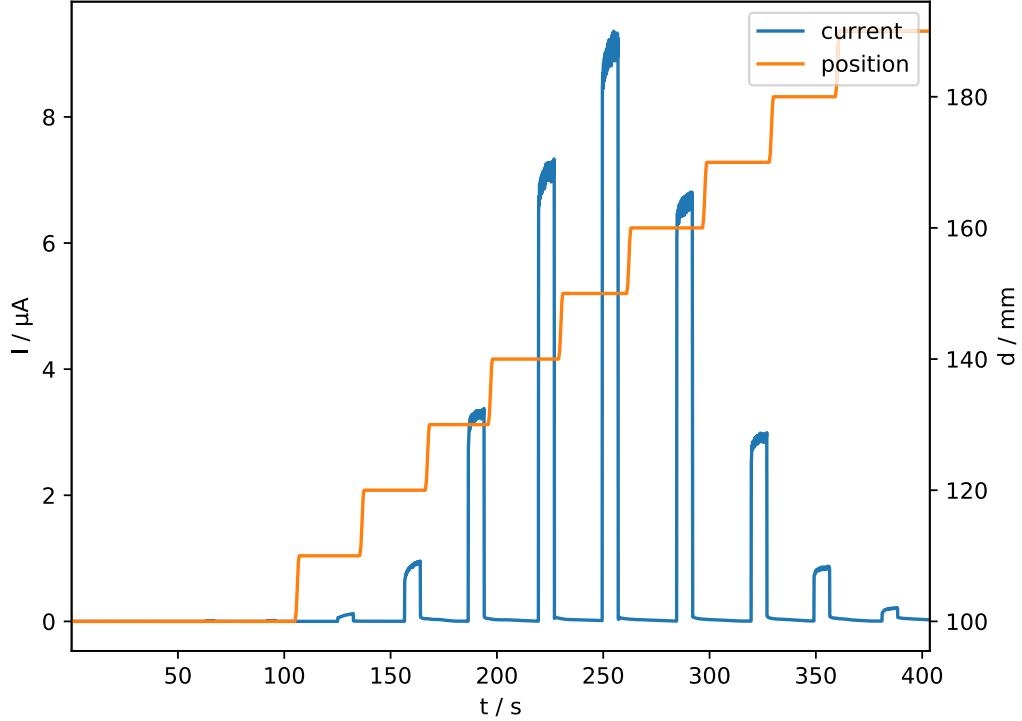


Figure 8.14 – Lateral beam profile scan with D01.

relevant for this thesis or the BCM, therefore no further studies towards this are presented. But the results of the WPE and ^{90}Sr measurements are compared for a crosscheck. Therefore, the collected charge of a diamond at the centre position is compared to the I_{inf} value. The results are listed in Table 8.1. Also the ratio of both measurements is calculated and is almost constant. This result is not very significant due to the small sample size. In the limited beam time it was not possible to measure all diamonds. Therefore, it would be useful to complete the measurement.

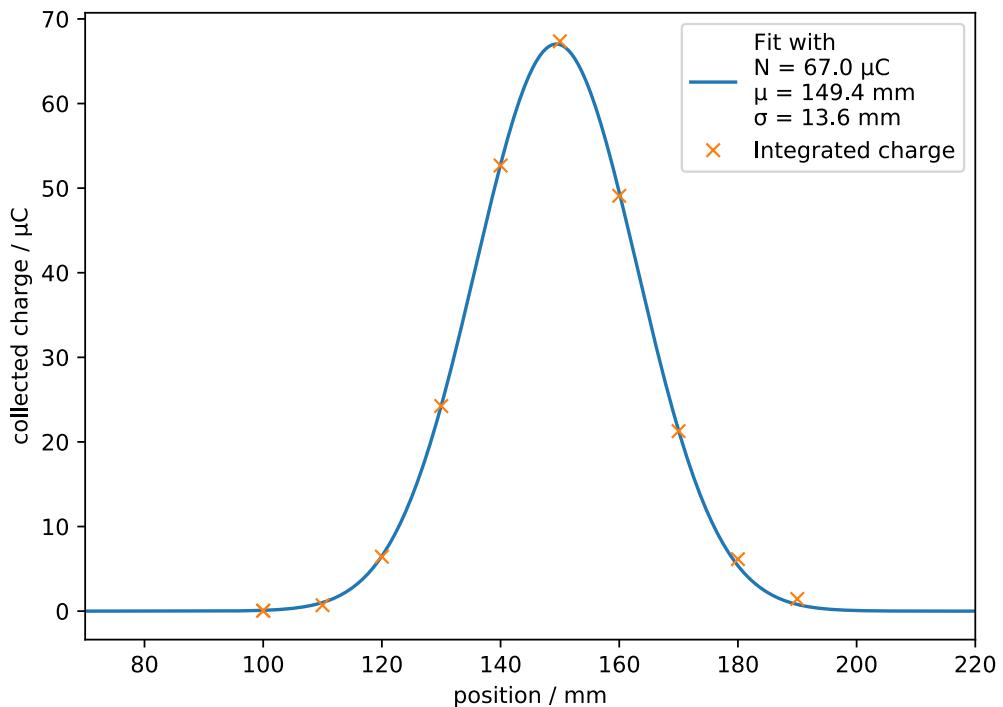


Figure 8.15 – Collected charge of the lateral scan. A Gaussian function is fit to the data.

diamond	${}^{90}\text{Sr} / \text{nA}$	WPE / μC	ratio
1	7.20	67.0	9.31
2	5.30	48.5	9.15
5	4.31	39.0	9.05
6	8.55	80.8	9.45

Table 8.1 – Comparison of ${}^{90}\text{Sr}$ and WPE results. The ratio of the results is almost constant.

8.4 Conclusion and Outlook

The general upgrade of the LHCb detector will also include an overhaul of the BCM system. Therefore, a new set of diamonds was ordered. These new diamonds need to be characterised before they are used in the BCM. The characterisation in a beam facility can not be guaranteed due to the tight time schedule of the upgrade and the corona crisis. Therefore, the possibility to perform the needed characterisation locally is evaluated. With a setup the diamond signal currents, induced by a ${}^{90}\text{Sr}$ source, are measured. These current signals are used to classify

8.4 Conclusion and Outlook

the diamonds. This procedure was tested with 19 spare diamonds. One result is that the setup can provide a relative classification of the diamonds. It is also possible to detect diamonds with a probably defect in the lattice structure or bad contacts. To conclude, a diamond characterisation in the sense of BCM usability is possible in the university.

In addition some diamonds are measured in a proton beam at the WPE. The results from these measurements scale sufficient linear to the results from the ^{90}Sr setup. It is planned to use this proton beam to get an absolute calibration for the in house measurements. This is possible due to the known flux at the WPE.

In parallel another setup was developed. With the so-called DONNA board it is possible to measure single signal pulses [109]. This board amplifies the signals induced by the ^{90}Sr source. The same source as before is used. With the DONNA board it is possible to measure the CCE without knowing the source flux. The presented measurements and results were used during the development of the DONNA setup.

9 Conclusion

The LHCb experiment performs an upgrade at the moment. Almost the entire detector is renewed. This new detector is needed as the proton-proton collision conditions will be changed. A fivefold instantaneous luminosity will be used. One of the new sub-detectors is the SciFi tracker, which is an essential component of the tracking system. Besides the detector itself, there is a significant change in the data acquisition system. The former hardware stage in the trigger was removed, and the upgrade trigger will be software only. At the average event rate of 30 MHz, the trigger's input rate is 32 Tbit s^{-1} , and the output bandwidth is limited to 10 Gbit s^{-1} .

The first contribution described in this thesis is about the SciFi tracker. This detector's base material is scintillating fibre, and the SciFi will cover an area of 360 m^2 . It is planned to record a dataset of 50 fb^{-1} in 10 years. During this time, the fibres are damaged by radiation. Because of the huge dimension of the tracker, the dose profile varies significantly. Besides the radiation damage, the light yield of the fibres is reduced by natural material ageing. The two effects, radiation damage and natural ageing, were added to the BOOLE SciFi simulation. In the simulation, it is possible to use any combination of radiation damage and fibre age. The radiation damage effect is applied by light yield attenuation maps, which can be interpolated in the dose. The light yield reduction caused by ageing is added as an additional attenuation.

This new SciFi simulation approach enables for the first time the possibility to simulate the tracking performance evolution of the SciFi over the planned lifetime. Furthermore, these are the first studies, which take the fibre ageing into account. Compared to the initial SciFi estimations, the ϵ_{loss} is 1.7 times higher. Also, the *ghost rate* will increase over time, and thereby the fibre ageing will reduce the rate slightly. Due to the increased SiPM noise, the total number of hits is increased by 9 %. An implication of the increased hit number is a longer per event computing time in the trigger. With constant trigger computing performance either the accuracy of the reconstruction or the throughput of the trigger needs to be reduced. The second contribution is towards LHCb's new trigger system, and this is software only. Now the first trigger stage, HLT1, is GPU based. Within the baseline architecture change process, a complete HLT1 sequence on GPU was needed. One part of the sequence is the SciFi decoder. A GPU based version of this decoder is presented in this thesis. Thereby three different data format versions are covered. The performance of the entire GPU HLT1 trigger was reported to the LHCb collaboration via a TDR. This document convinced the collaboration to realise the

9 Conclusion

GPU trigger.

The third contribution is towards the BCM, a safety system that monitors the LHC beam position within the LHCb detector. Within the scope of the detector upgrade the BCM is renewed. The sensors of the BCM are made from diamonds. A batch of diamonds is ordered to build new BCM stations. Before the sensor are assembled to the support structure a characterisation is necessary. For this task, a source based setup is developed, which is located in the local laboratory. By that a characterisation independent of external facilities is possible.

In conclusion, this thesis covers various topics relevant to the future LHCb experiment. The SciFi simulation studies deal with the detector and performance evolution. Contributions to the data-recording process or the trigger process were made via the implementations of SciFi hit decoder in the new GPU trigger, Allen. Furthermore, safe data acquisition is supported by the studies towards the BCM diamond characterisation.

A SciFi performance studies

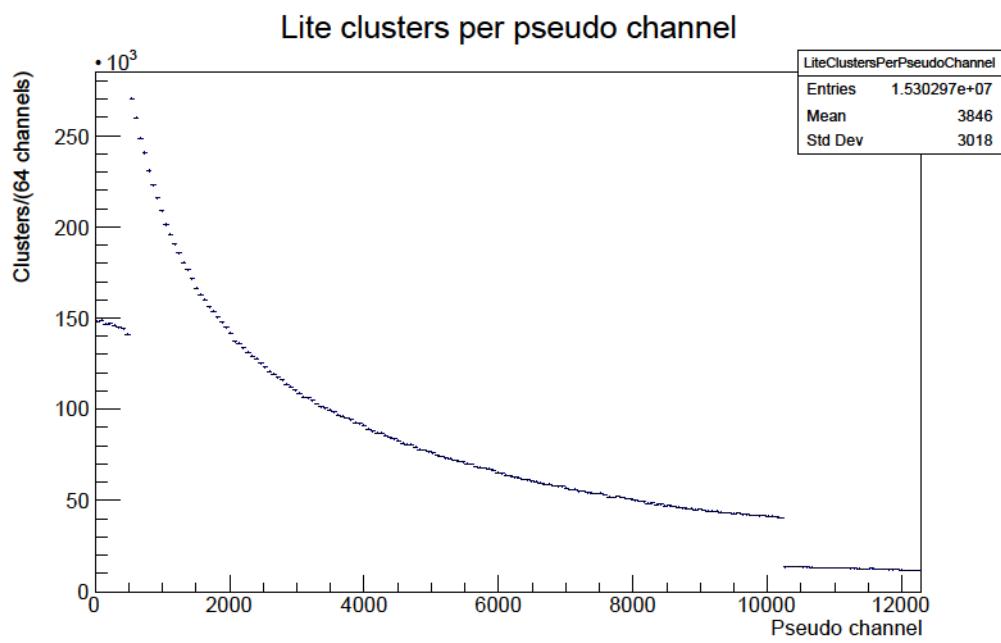


Figure A.1 – Distribution of the cluster in the SciFi per pseudo channel after 50 fb^{-1} with double ageing. condition (50,2)

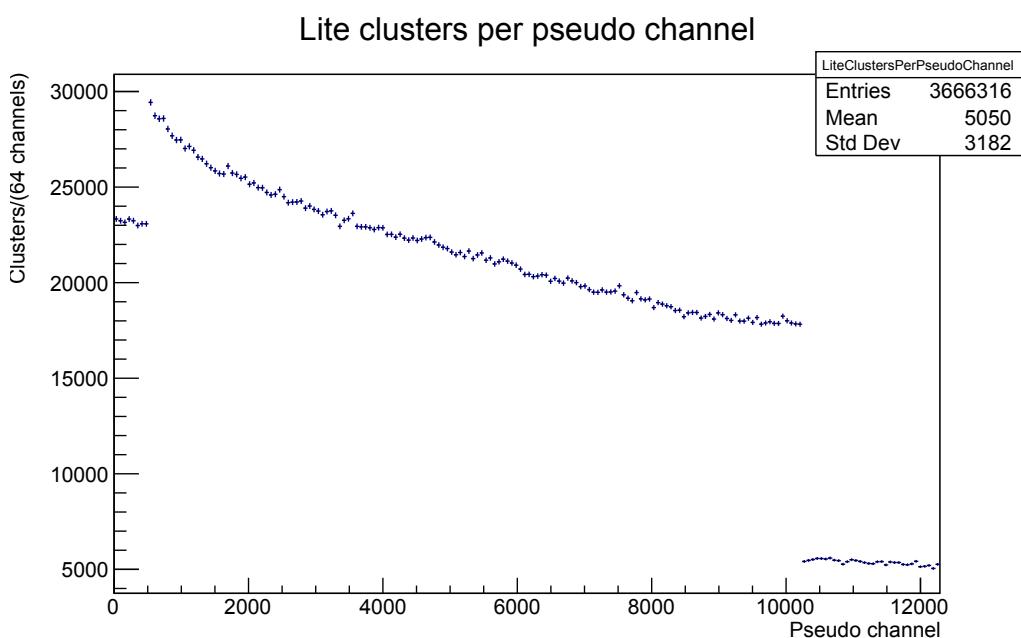


Figure A.2 – Cluster from spillover in the 50-2 sample per pseudo channel.

Bibliography

- [1] S.L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588, DOI: [10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [2] Steven Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266, DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [3] Abdus Salam, *Weak and Electromagnetic Interactions*, 8th Nobel Symposium, *Elementary Particle Theory*, ed. by Nils Svaratholm, **C68-05-19**, Lerum, Sweden, 1968.
- [4] Georges Aad *et al.*, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, DOI: [10.1103/PhysRevLett.114.191803](https://doi.org/10.1103/PhysRevLett.114.191803), [arXiv:1503.07589 \[hep-ex\]](https://arxiv.org/abs/1503.07589).
- [5] E. Corbelli and P. Salucci, *The extended rotation curve and the dark matter halo of M33*, Monthly Notices of the Royal Astronomical Society **311**.2 (2000) 441–447, ISSN: 1365-2966, DOI: [10.1046/j.1365-8711.2000.03075.x](https://doi.org/10.1046/j.1365-8711.2000.03075.x).
- [6] Joshua A. Frieman, Michael S. Turner, and Dragan Huterer, *Dark Energy and the Accelerating Universe*, Annual Review of Astronomy and Astrophysics **46**.1 (2008) 385–432, DOI: [10.1146/annurev.astro.46.060407.145243](https://doi.org/10.1146/annurev.astro.46.060407.145243).
- [7] K. Eguchi *et al.*, *First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance*, Physical Review Letters **90**.2 (2003), ISSN: 1079-7114, DOI: [10.1103/physrevlett.90.021802](https://doi.org/10.1103/physrevlett.90.021802).
- [8] A.D. Sakharov, *Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32–35, DOI: [10.1070/PU1991v034n05ABEH002497](https://doi.org/10.1070/PU1991v034n05ABEH002497).
- [9] LHCb Collaboration, *LHCb performance numbers*, <http://lhcb.web.cern.ch/lhcb/speakersbureau/html/PerformanceNumbers.html> (visited on 12/01/2020).
- [10] LHCb Collaboration, *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ Branching Fraction and Effective Lifetime and Search for $B^0 \rightarrow \mu^+ \mu^-$ Decays*, Phys. Rev. Lett. **118** (19 2017) 191801, DOI: [10.1103/PhysRevLett.118.191801](https://doi.org/10.1103/PhysRevLett.118.191801).
- [11] Roel et al. Aaij, *Observation of CP violation in charm decays*, Phys. Rev. Lett. **122**.arXiv:1903.08726. LHCb-PAPER-2019-006 (2019) 211803. 12 p, DOI: [10.1103/PhysRevLett.122.211803](https://doi.org/10.1103/PhysRevLett.122.211803), <https://cds.cern.ch/record/2668357>.

Bibliography

- [12] *Letter of Intent for the LHCb Upgrade*, tech. rep. CERN-LHCC-2011-001. LHCC-I-018, Geneva: CERN, 2011, <https://cds.cern.ch/record/1333091>.
- [13] Christiane Lefèvre, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, 2008, <https://cds.cern.ch/record/1260465/files/0812015.pdf> (visited on 12/01/2020).
- [14] G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003, doi: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [15] K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002, doi: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [16] S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004, doi: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [17] Esma Mobs, *The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018* (2018), General Photo, <https://cds.cern.ch/record/2636343>.
- [18] Peter W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509, doi: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [19] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323, doi: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [20] <https://en.wikipedia.org/wiki/Quark> (visited on 12/01/2020).
- [21] Nicola Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531–533, doi: [10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531).
- [22] Makoto Kobayashi and Toshihide Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657, doi: [10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [23] J. Charles *et al.*, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. **C41** (2005) 1–131, doi: [10.1140/epjc/s2005-02169-1](https://doi.org/10.1140/epjc/s2005-02169-1), eprint: [hep-ph/0406184](https://arxiv.org/abs/hep-ph/0406184), updated results and plots available at: <http://ckmfitter.in2p3.fr>.
- [24] R. Aaij *et al.*, *Measurement of CP violation in $B^0 \rightarrow J/\psi S\ 0$ and $B^0 \rightarrow (2S)K\ S\ 0$ decays*, Journal of High Energy Physics **2017**.11 (2017), ISSN: 1029-8479, doi: [10.1007/jhep11\(2017\)170](https://doi.org/10.1007/jhep11(2017)170).
- [25] LHCb collaboration, *Test of lepton universality in beauty-quark decays*, 2021, [arXiv:2103.11769](https://arxiv.org/abs/2103.11769).
- [26] Jr. Alves A. Augusto *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005, doi: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [27] Christian Elsässer, *$b\bar{b}$ production angle plots*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html (visited on 11/22/2020).

- [28] Miriam Calvo Gomez, *Flavour Tagging at LHCb*, LHCb-PROC-2012-041. CERN-LHCb-PROC-2012-041 (2012) 6 p, <http://cds.cern.ch/record/1478153>.
- [29] F Follin and D Jacquet, *Implementation and experience with luminosity levelling with offset beam* (2014), Comments: 5 pages, contribution to the ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders, CERN, Geneva, Switzerland, 18-22 Mar 2013 5 p, DOI: [10.5170/CERN-2014-004.183](https://doi.org/10.5170/CERN-2014-004.183).
- [30] LHCb Collaboration, *LHCb Detector Performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, DOI: [10.1142/S0217751X15300227](https://doi.org/10.1142/S0217751X15300227).
- [31] R. Aaij *et al.*, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115** (7 2015) 072001, DOI: [10.1103/PhysRevLett.115.072001](https://doi.org/10.1103/PhysRevLett.115.072001).
- [32] CERN, *LHC long term schedule*, <https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm> (visited on 12/01/2020).
- [33] LHCb Collaboration, *LHCb Trigger and Online Upgrade Technical Design Report* (2014), <http://cds.cern.ch/record/1701361>.
- [34] J Albrecht *et al.*, *The upgrade of the LHCb trigger system*, Journal of Instrumentation **9**.10 (2014) C10026–C10026, DOI: [10.1088/1748-0221/9/10/c10026](https://doi.org/10.1088/1748-0221/9/10/c10026).
- [35] LHCb Collaboration, *LHCb Tracker Upgrade Technical Design Report* (2014), <http://cds.cern.ch/record/1647400>.
- [36] LHCb Collaboration, *LHCb VELO Upgrade Technical Design Report* (2013), <http://cds.cern.ch/record/1624070>.
- [37] Ch. Ilgner *et al.*, *The Beam Conditions Monitor of the LHCb Experiment* (2010), [arXiv:1001.2487](https://arxiv.org/abs/1001.2487).
- [38] LHCb Collaboration, *LHCb PID Upgrade Technical Design Report* (2013), <http://cds.cern.ch/record/1624074>.
- [39] *LHCb magnet: Technical Design Report*, Geneva, 2000, <https://cds.cern.ch/record/424338>.
- [40] Marco Clemencic *et al.*, *Recent developments in the LHCb software framework Gaudi*, 17th International Conference on Computing in High Energy and Nuclear Physics, ed. by Jan Grunorad, **219**, 2010 042006, DOI: [10.1088/1742-6596/219/4/042006](https://doi.org/10.1088/1742-6596/219/4/042006).
- [41] G. Barrand *et al.*, *GAUDI - The software architecture and framework for building LHCb data processing applications*, Comput. Phys. Commun. **140**.1-2 (2001), ed. by Mirco Mazzucato 92–95.
- [42] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011), ed. by Simon C. Lin 032023, DOI: [10.1088/1742-6596/331/3/032023](https://doi.org/10.1088/1742-6596/331/3/032023).

Bibliography

- [43] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011), ed. by Simon C. Lin 032047, DOI: [10.1088/1742-6596/331/3/032047](https://doi.org/10.1088/1742-6596/331/3/032047).
- [44] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026, DOI: [10.1088/1126-6708/2006/05/026](https://doi.org/10.1088/1126-6708/2006/05/026).
- [45] Torbjörn Sjöstrand *et al.*, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159–177, DOI: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024).
- [46] D.J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152–155, DOI: [10.1016/S0168-9002\(01\)00089-4](https://doi.org/10.1016/S0168-9002(01)00089-4).
- [47] S. Agostinelli *et al.*, *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303, DOI: [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [48] John Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270, DOI: [10.1109/TNS.2006.869826](https://doi.org/10.1109/TNS.2006.869826).
- [49] Markus Frank *et al.*, *DD4hep*, 2018, DOI: [10.5281/zenodo.592244](https://doi.org/10.5281/zenodo.592244).
- [50] LHCb Collaboration, *The Boole Project*, <http://lhcbdoc.web.cern.ch/lhcbdoc/bole/>.
- [51] LHCb Collaboration, *The Brunel Project*, ed. by Marco Cattaneo and Rob Lambert, <http://lhcbdoc.web.cern.ch/lhcbdoc/brunel/>.
- [52] LHCb Collaboration, *The DaVinci Project*, <http://lhcbdoc.web.cern.ch/lhcbdoc/davinci/>.
- [53] Concezio Bozzi, *LHCb Computing Resource usage in 2019*, tech. rep. LHCb-PUB-2020-002. CERN-LHCb-PUB-2020-002, Geneva: CERN, 2020, <https://cds.cern.ch/record/2710435>.
- [54] LHCb Collaboration, *The Moore Project*, <http://lhcbdoc.web.cern.ch/lhcbdoc/moore/>.
- [55] R. Aaij *et al.*, *Tesla: An application for real-time data analysis in High Energy Physics*, Computer Physics Communications **208** (2016) 35–42, ISSN: 0010-4655, DOI: [10.1016/j.cpc.2016.07.022](https://doi.org/10.1016/j.cpc.2016.07.022).
- [56] R. Bailey and Paul Collier, *Standard Filling Schemes for Various LHC Operation Modes*, tech. rep. LHC-PROJECT-NOTE-323, Geneva: CERN, 2003, <https://cds.cern.ch/record/691782>.
- [57] LHCb Collaboration, *RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector* (2020), <https://cds.cern.ch/record/2730181>.
- [58] LHCb Collaboration, *Computing Model of the Upgrade LHCb experiment*, tech. rep. CERN-LHCC-2018-014. LHCB-TDR-018, Geneva: CERN, 2018, [http://cds.cern.ch/record/2319756](https://cds.cern.ch/record/2319756).

- [59] R. Aaij *et al.*, *A comprehensive real-time analysis model at the LHCb experiment*, Journal of Instrumentation **14**.04 (2019) P04006–P04006, ISSN: 1748-0221, DOI: [10.1088/1748-0221/14/04/p04006](https://doi.org/10.1088/1748-0221/14/04/p04006).
- [60] ECFA High Luminosity LHC Experiments Workshop: Physics and Technology Challenges. 94th Plenary ECFA meeting (2013), <https://cds.cern.ch/record/1631032>.
- [61] Kuraray Co. Ltd, *Scintillator Material Product Catalogue*, <http://kurara.ypsf.jp/psf/sf.html> (visited on 01/12/2020).
- [62] Mirco Deckenhoff, *Scintillating Fibre and Silicon Photomultiplier Studies for the LHCb Upgrade*, CERN-THESIS-2015-318, PhD thesis, Technische Universität Dortmund, 2015.
- [63] Th. Förster, *Zwischenmolekulare Energiewanderung und Fluoreszenz*, Annalen der Physik **437**.1-2 (1948) 55–75, DOI: [10.1002/andp.19484370105](https://doi.org/10.1002/andp.19484370105).
- [64] Robert Ekelhof, *Studies for the LHCb SciFi Tracker – Development of modules from scintillating fibres and tests of their radiation hardness*, CERN-THESIS-2017-098, PhD thesis, Technische Universität Dortmund, 2016.
- [65] Martin Stefan Bieker, Robert Jan Ekelhof, and Robin Manderfeld, *Description of Light Guidance in Dual Clad Scintillating Fibres for the LHCb SciFi Tracker*, tech. rep. LHCb-PUB-2019-006. CERN-LHCb-PUB-2019-006, Geneva: CERN, 2019, <https://cds.cern.ch/record/2671964>.
- [66] Jeroen Van Tilburg, *SciFi readout numbering scheme*, tech. rep. LHCb-INT-2016-044. CERN-LHCb-INT-2016-044, Geneva: CERN, 2016, <https://cds.cern.ch/record/2229004>.
- [67] Christian Joram *et al.*, *LHCb Scintillating Fibre Tracker Engineering Design Review Report: Fibres, Mats and Modules*, tech. rep. LHCb-PUB-2015-008. CERN-LHCb-PUB-2015-008, Geneva: CERN, 2015, <https://cds.cern.ch/record/2004811>.
- [68] Sevda Esen *et al.*, *Clustering and rawbank decoding for the SciFi detector*, tech. rep. LHCb-INT-2018-024. CERN-LHCb-INT-2018-024, Geneva: CERN, 2018, <https://cds.cern.ch/record/2630154>.
- [69] David Rolf, *Optimierung der Ausleseelektronik des LHCb-SciFi-Trackers - Studien zur Laufzeit einer Grafikkarten basierten Decoding-Sequenz im HLT*, B.Sc. thesis, Technische Universität Dortmund, 2019.
- [70] Olivier Le Dortz *et al.*, *LHCb Upgrade SciFi Tracker TELL40 Data Processing* (2018), <https://edms.cern.ch/ui/#!master/navigator/document?D:100228958:100228958:subDocs>.
- [71] Mirco Deckenhoff, *Simulation of Scintillating Fibres in Geant4*, tech. rep. LHCb-PUB-2014-023. CERN-LHCb-PUB-2014-023. LHCb-INT-2014-009. LHCb-INT-2014-009, Geneva: CERN, 2014, <https://cds.cern.ch/record/1662547>.

Bibliography

- [72] Mirco Deckenhoff, *Signal Shape and Time of Light Propagation in Scintillating Fibre SCSF-78MJ from Kuraray*, tech. rep. LHCb-INT-2013-008. CERN-LHCb-INT-2013-008. LHCb-PUB-2014-016, Geneva: CERN, 2013, <https://cds.cern.ch/record/1517050>.
- [73] Martin Bieker, *Simulation Studies for the Scintillating Fibre Tracker of the LHCb Experiment*, M.Sc. thesis, Technische Universität Dortmund, 2018.
- [74] Stefan Escher, *Simulation studies of scintillating fibre mats for the LHCb upgrade*, M.Sc. thesis, Technische Universität Dortmund, 2017.
- [75] Moritz Demmer *et al.*, *Simulation of Light Yield Attenuation Maps for the LHCb SciFi Tracker Upgrade*, tech. rep. LHCb-INT-2016-015. CERN-LHCb-INT-2016-015, Geneva: CERN, 2016, <https://cds.cern.ch/record/2141533>.
- [76] Sarah Beranek *et al.*, *Simulation of the light yield attenuation maps for the LHCb SciFi Tracker Upgrade*, tech. rep. LHCb-PUB-2019-007. CERN-LHCb-PUB-2019-007, Geneva: CERN, 2019, <https://cds.cern.ch/record/2673602>.
- [77] Hermann Kolanoski and Norbert Wermes, *Teilchendetektoren*, Springer Spektrum, 2016, DOI: [10.1007/978-3-662-45350-6](https://doi.org/10.1007/978-3-662-45350-6).
- [78] Particle Data Group *et al.*, *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020**.8 (2020), 083C01, ISSN: 2050-3911, DOI: [10.1093/ptep/ptaa104](https://doi.org/10.1093/ptep/ptaa104).
- [79] E. A. Uehling, *Penetration of Heavy Charged Particles in Matter*, Annual Review of Nuclear Science **4**.1 (1954) 315–350, DOI: [10.1146/annurev.ns.04.120154.001531](https://doi.org/10.1146/annurev.ns.04.120154.001531).
- [80] T. T. Böhlen *et al.*, *The FLUKA Code: Developments and Challenges for High Energy and Medical Applications*, Nucl. Data Sheets **120** (2014) 211–214, DOI: [10.1016/j.nds.2014.07.049](https://doi.org/10.1016/j.nds.2014.07.049).
- [81] Alfredo Ferrari *et al.*, *FLUKA: A multi-particle transport code (Program version 2005)*, CERN-2005-010, CERN Yellow Reports: Monographs (2005), DOI: [10.5170/CERN-2005-010](https://doi.org/10.5170/CERN-2005-010).
- [82] Matthias Karacson, *Evaluation of the Radiation Environment of the LHCb Experiment*, Presented 12 Dec 2016, 2016, <https://cds.cern.ch/record/2243499>.
- [83] David Müller, *Performance studies of irradiated scintillating fibre modules and quality checks during serial production for the LHCb SciFi Tracker*, Universität Heidelberg M. Sc. thesis, 2017.
- [84] C. Zorn, *A pedestrian's guide to radiation damage in plastic scintillators*, Radiation Physics and Chemistry **41**.1 (1993) 37–43, ISSN: 0969-806X, DOI: [https://doi.org/10.1016/0969-806X\(93\)90040-2](https://doi.org/10.1016/0969-806X(93)90040-2).

- [85] W. Busjan, K. Wick, and T. Zoufal, *Shortlived absorption centers in plastic scintillators and their influence on the fluorescence light yield*, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms **152**.1 (1999) 89–104, ISSN: 0168-583X, DOI: [https://doi.org/10.1016/S0168-583X\(98\)00974-4](https://doi.org/10.1016/S0168-583X(98)00974-4).
- [86] A. Artikov *et al.*, *Properties of the Ukraine polystyrene-based plastic scintillator UPS 923A*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **555**.1 (2005) 125–131, ISSN: 0168-9002, DOI: <https://doi.org/10.1016/j.nima.2005.09.021>.
- [87] A.B.R. Cavalcante *et al.*, *Refining and testing 12,000 km of scintillating plastic fibre for the LHCb SciFi tracker*, Journal of Instrumentation **13**.10 (2018) P10025–P10025, DOI: [10.1088/1748-0221/13/10/p10025](https://doi.org/10.1088/1748-0221/13/10/p10025).
- [88] Ole Gerber, *Simulation studies to investigate the effects of radiation dose on the LHCb SciFi detector*, M.Sc. thesis, Technische Universität Dortmund, 2018.
- [89] LHCb Collaboration, *Measurement of CP violation in the $B_s^0 \rightarrow \phi\phi$ decay and search for the $B^0 \rightarrow \phi\phi$ decay*, JHEP **12**.arXiv:1907.10003. LHCb-PAPER-2019-019 (2019) 155. 34 p, <https://cds.cern.ch/record/2684085>.
- [90] Olivier Goran Girard *et al.*, *Read-out of irradiated SiPMs attached to short scintillating fibre module*, tech. rep. LHCb-INT-2017-009. CERN-LHCb-INT-2017-009, Geneva: CERN, 2017, <https://cds.cern.ch/record/2262339>.
- [91] Renato Quagliani *et al.*, *The Hybrid Seeding algorithm for a scintillating fibre tracker at LHCb upgrade: description and performance*, tech. rep. LHCb-PUB-2017-018. CERN-LHCb-PUB-2017-018, Geneva: CERN, 2017, <https://cds.cern.ch/record/2265114>.
- [92] R. Aaij *et al.*, *Expression of Interest for a Phase-II LHCb Upgrade: Opportunities in flavour physics, and beyond, in the HL-LHC era*, tech. rep. CERN-LHCC-2017-003, Geneva: CERN, 2017, <https://cds.cern.ch/record/2244311>.
- [93] The LHCb Collaboration, *LHCb Upgrade GPU High Level Trigger Technical Design Report*, tech. rep. CERN-LHCC-2020-006. LHCb-TDR-021, 2020, <https://cds.cern.ch/record/2717938>.
- [94] Jon Peddie, *Is it Time to Rename the GPU? Used for Far More than Graphics Now*, <https://www.computer.org/publications/tech-news/chasing-pixels/is-it-time-to-rename-the-gpu> (visited on 11/12/2020).
- [95] Peng Du *et al.*, *From CUDA to OpenCL: Towards a performance-portable solution for multi-platform GPU programming*, Parallel Computing **38**.8 (2012), APPLICATION ACCELERATORS IN HPC 391–407, ISSN: 0167-8191, DOI: <https://doi.org/10.1016/j.parco.2011.10.002>.

Bibliography

- [96] Prometheus GmbH, *Top500 Supercomputer*, <https://www.top500.org/lists/top500/2020/11/> (visited on 01/08/2021).
- [97] Jason Sanders and Edward Kandrot, *CUDA by Example: An Introduction to General-Purpose GPU Programming*, 1st, Addison-Wesley Professional, 2010, ISBN: 0131387685.
- [98] Inc. Micron Technology, *GDDR6X feature*, <https://www.micron.com/products/ultra-bandwidth-solutions/gddr6x> (visited on 12/01/2020).
- [99] NVIDIA Corporation, *CUDA programming guide*, <https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html> (visited on 12/01/2020).
- [100] R. Aaij *et al.*, *Allen: A High-Level Trigger on GPUs for LHCb*, Computing and Software for Big Science **4** (2020), doi: [10.1007/s41781-020-00039-7](https://doi.org/10.1007/s41781-020-00039-7).
- [101] *Performance of the GPU HLT1 (Allen)* (2020), <http://cds.cern.ch/record/2722327>.
- [102] V. Gligorov, *RTA project report at the 98th LHCb week*, <https://indico.cern.ch/event/973067/#73-rta-report>.
- [103] Benjamin Todd, *A Beam Interlock System for CERN High Energy Accelerators*, PhD thesis, 2006, <http://cds.cern.ch/record/1019495>.
- [104] Harris Kagan, Shaun Roe, and Peter Weilhamer, *RD42 Status Report: Development of Diamond Tracking Detectors for High Luminosity Experiments at the LHC*, tech. rep. CERN-LHCC-2008-005. LHCC-RD-016, Geneva: CERN, 2008, <http://cds.cern.ch/record/1098155>.
- [105] Sergey Barsuk *et al.*, *Probe for LUMinosity MEasurement in LHCb*, tech. rep. LHCb-PUB-2020-008. CERN-LHCb-PUB-2020-008, Geneva: CERN, 2020, <https://cds.cern.ch/record/2743098>.
- [106] A Büchner *et al.*, *THE ELBE-PROJECT AT DRESDEN-ROSSENDORF*, 2010, <https://accelconf.web.cern.ch/e00/PAPERS/WEP3B04.pdf>.
- [107] R. Kallo, *Untersuchung von bestrahlten und unbestrahlten Diamantsensoren für das Beam Condition Monitor System des LHCb-Experiments*, Title in English: “Investigation of unirradiated and irradiated diamond sensors for the Beam Condition Monitor System of the LHCb experiment”, B.Sc. thesis, Technische Universität Dortmund, 2019.
- [108] M. Kaluza, *Charakterisierung von Diamantsensoren mittels Vermessung von Einzelpulsen*, Title in English: “Characterization of diamond sensors through single-pulse measurements”, B.Sc. thesis, Technische Universität Dortmund, 2020.
- [109] L. Funke, *Data acquisition and diamond detector pulse shape measurements for the upgrade of the LHCb Beam Condition Monitor*, M.Sc. thesis, Technische Universität Dortmund, 2020.

- [110] J. Koike, D. M. Parkin, and T. E. Mitchell, *Displacement threshold energy for type IIa diamond*, Applied Physics Letters **60**.12 (1992) 1450–1452, DOI: [10.1063/1.107267](https://doi.org/10.1063/1.107267).
- [111] Hermann Kolanoski and Norbert Wermes, *Particle Detectors: Fundamentals and Applications*, Oxford University Press, 2020, DOI: [10.1093/oso/9780198858362.001.0001](https://doi.org/10.1093/oso/9780198858362.001.0001).
- [112] C. Canali *et al.*, *Electrical properties and performances of natural diamond nuclear radiation detectors*, Nuclear Instruments and Methods **160**.1 (1979) 73–77, ISSN: 0029-554X, DOI: [10.1016/0029-554X\(79\)90167-8](https://doi.org/10.1016/0029-554X(79)90167-8).
- [113] Claude A. Klein, *Bandgap Dependence and Related Features of Radiation Ionization Energies in Semiconductors*, Journal of Applied Physics **39**.4 (1968) 2029–2038, DOI: [10.1063/1.1656484](https://doi.org/10.1063/1.1656484).
- [114] W. Shockley, *Currents to Conductors Induced by a Moving Point Charge*, J. Appl. Phys. **9** (1938) 635, DOI: [10.1063/1.1710367](https://doi.org/10.1063/1.1710367).
- [115] S. Ramo, *Currents Induced by Electron Motion*, Proceedings of IRE **27** (1939) 584, DOI: [10.1109/JRPROC.1939.228757](https://doi.org/10.1109/JRPROC.1939.228757).
- [116] Element Six Technologies, *CVD diamond handbook*, 2020, https://e6cvd.com/media/wysiwyg/pdf/CVD_Diamond_Handbook_digital_01.07.20.pdf.
- [117] O Voronov and Gary Tompa, *High Pressure-High Temperature Device for Making Diamond Materials*, Diamond Materials VII **2001** (2001).
- [118] A. Galbiati *et al.*, *Performance of monocrystalline diamond radiation detectors fabricated using TiW, Cr/Au and a novel ohmic DLC/Pt/Au electrical contact*, 2008 IEEE Nuclear Science Symposium Conference Record, 2008 190–196, DOI: [10.1109/NSSMIC.2008.4775151](https://doi.org/10.1109/NSSMIC.2008.4775151).
- [119] Laboratoire National Henri Becquerel, *Strontium 90 decay*, http://www.nucleide.org/DDEP_WG/Nuclides/Sr-90_tables.pdf (visited on 12/01/2020).
- [120] Laboratoire National Henri Becquerel, *Yttrium 90 decay*, http://www.nucleide.org/DDEP_WG/Nuclides/Y-90_tables.pdf (visited on 12/01/2020).
- [121] PTW The dosimetry company, *Gebrauchsanweisung Radioaktive Kontrollvorrichtungen T48010 und T48012*, D467.151.00/06de, 2017.
- [122] *Picoammeter / Voltage Source 6487 datasheet*, 6487, Keithley, <https://download.tek.com/datasheet/6487.pdf> (visited on 09/12/2020).

Acknowledgements

An dieser Stelle möchte ich die Gelegenheit nutzen und mich bei einigen der vielen Menschen, die mich auf dem Weg zur Promotion begleitet haben, bedanken. Zunächst danke ich meinem Doktorvater Prof. Dr. Bernhard Spaan, der mir die Promotion ermöglicht hat, nachdem ich bereits meine Bachelor- und Masterarbeit an seinem Lehrstuhl schreiben konnte. Des Weiteren danke ich Prof. Dr. Dr. Wolfgang Rhode für seine Bereitschaft als Zweitgutachter zu fungieren.

Des Weiteren danke ich dem SFB 876, dieser hat meine Promotion und einen anregenden interdisziplinären Meinungsaustausch ermöglicht.

Großer Dank gilt der Gemeinschaft des Lehrstuhls E5 für die überaus angenehme Arbeitsatmosphäre und Hilfsbereitschaft. Danken möchte ich den ehemaligen Kollegen Robert, Mirco und Florian, die mich damals an den Lehrstuhl gelockt haben. Ebenfalls möchte ich mich bei den Studierenden bedanken, die ich ausbilden und betreuen durfte. Diese Arbeiten haben meine Forschung unterstützt.

Die Arbeitsgruppe, der ich angehöre hat öfters ihren Namen geändert, aber egal ob ihr der Hardware-, Detektorentwicklungs- bzw. BCM-gruppe von E5 zugehörig wart oder seid, ich danke euch.

Aufgrund der ungewöhnlichen Zeit hat sich eine besondere Gruppe herauskristallisiert, ich nenne sie den Corona Kern. Matthias, Martin, Henning und David ich danke euch, dass ihr ein angenehmes Arbeitsumfeld ermöglicht habt und trotz Pandemie immer mit Rat und Tat zur Seite standet.

Großer Dank gilt auch all denen, die sich bereit erklärt haben, meine Arbeit Korrektur zu lesen.

Meiner Familie danke ich für ihre Unterstützung und die Möglichkeit immer in meine Heimat und zu meinen Wurzeln zurückkehren zu können. Abschließend danke ich meiner geliebten Partnerin Hannah.

Zu guter Letzt folgt noch eine willkürlich geordnete Liste mit Gruppen denen ich danke:

- der neuen Gruppe
- den HSP Spielen spielen und Frisbee Freunden
- meinen alten Freunden aus dem Studium, die Dortmund verlassen haben
- der SV Veldhausen 07 Schach Abteilung
- dem EF3 Spieleabend
- Sophia und Alex, Marius und Esther