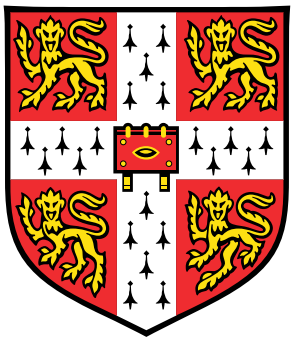
Convolutional neural networks and photonic crystals for particle identification at high energy collider experiments



Michele Piero Blago
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Churchill College September 2021

CERN-THESIS-2021-215
11/11/2021

Abstract

The LHCb RICH detection system at CERN has been established as an effective way to identify particles in high energy physics experiments. Future increases in collision rate and luminosity make limitations of both the particle identification (PID) algorithm and the Cherenkov radiators apparent. These two challenges are addressed in this thesis.

In the first part, a convolutional neural network (CNN) approach to particle identification in LHCb RICH is introduced. Changes in the LHCb computing infrastructure towards multi-core architectures provide the opportunity to benefit from the quickly evolving field of machine learning. The calculations of the presented neural network are well suited for quasi-real-time reconstruction through parallel processing. It takes binary input images from the two RICH detectors to classify particle species. The model is hyperparameter-optimised and trained on classification accuracy with simulated collision data for the upcoming LHC operation starting in 2022, reaching a PID performance comparable to the conventional algorithm. Under conditions of reduced combinatorial background, as expected from the introduction of timing resolution to the RICH detectors in future upgrades, the network achieves a particle identification performance close to 100 %, with simultaneous misclassification of the most prevalent particle species approaching 0 %.

The second part of the thesis addresses the limitations of conventional Cherenkov detectors, which require large gas chambers to distinguish between particles at high energies and to reach sufficient photon yield. Photonic crystals have been shown to emit photons under specific angles similar to the conventional Cherenkov effect. The properties of this novel Cherenkov radiator offer the possibility to distinguish particles in a much wider range of energies while requiring significantly less space as a Cherenkov radiator.
The work presented in this chapter introduces a framework to manufacture, simulate, and study photonic crystals under high-energy beam conditions at CERN. In particular, the manufacturing process of a one-dimensional photonic crystal, suitable for high-energy applications, is described, as well as its tests in a 200 MeV electron beam at the CERN CLEAR facility.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit for the Degree Committee for the Faculty of Physics.

Michele Piero Blago

Preface

I joined CERN and the LHCb experiment as a summer student in 2015, where I continued my journey as a technical student and later as a doctoral researcher. I am grateful to my CERN supervisor Carmelo D'Ambrosio for his encouraging guidance during all these years. The members of the CERN LHCb RICH group provided me with the best research environment I could imagine. I would like to thank Giovanni Cavallero, Christoph Frei, Thierry Gys, Silvia Gambetta and Antonino Sergi. Special thanks to Didier Piedigrossi for his tremendous help with any technical enquiries, and to Floris Keizer for our joint RICH performance research and for making the sleepless test beam weeks a memorable adventure.

I am also thankful to Sajan Easo for his unwavering motivation regarding the photonic crystal project. My gratitude extends to our collaborators, especially Xiao Lin, for providing the analytical framework and helpful explanations and Ido Kaminer for the many fascinating discussions and ideas. I thank Sune Jakobsen for contributing essential components to the laboratory set-up and Thomas Schneider for his friendship and lending his expertise to manufacture the crystal. From the CLEAR group, I thank Wilfrid Farabolini and Luke Dyks for their immense help and patience, as well as Michele Bergamaschi, Antonio Gilardi, and Thibaut Lefevre for their welcoming collaboration.

Throughout my PhD studies at the University of Cambridge, I am infinitely grateful to my supervisor Stephen Wotton. Thanks a lot, Steve, for always having my back and for your trust in my abilities.
I also thank the Cambridge High Energy Physics group members, in particular Val Gibson, for her leadership and Chris Jones for all the discussions and advice on the LHCb software. I gratefully acknowledge Daniel Hugo Campora Pérez for sharing his ideas regarding the neural network project, as well as IBM and CERN Openlab for providing the computing resources.

The past years would not have been the same without a few special people. I thank Simon Junker for the meticulous planning of our adventures in the Swiss Alps. Thomas Reitenspieß for sharing much more than just a fridge. Livio Verra for teaching me that endurance can be more than a metaphorical concept. Andre Salzinger and Marvin Holten for expanding my horizon beyond physics. Yuting Gu for the endless encouragement and support. And finally, my brother Antonio and my parents Sylvia and Nicola, without whom none of this would have been possible.

Contents

1 The Large Hadron Collider and the LHCb experiment **3**
1.1 The LHC and its four main experiments . . . 3
1.2 The LHCb experiment . . . 6
1.2.1 The LHCb physics case . . . 6
1.2.2 The LHCb detector . . . 8
1.3 The LHCb upgrades . . . 10
1.3.1 The particle tracking systems . . . 10
1.3.2 Particle identification . . . 13
1.3.3 Calorimeters . . . 18
1.3.4 Muon stations . . . 19
1.3.5 Changes to the data processing . . . 19
1.4 The evolution of the LHCb experiment . . . 21

2 Deep learning in LHCb RICH **23**
2.1 Simulation and reconstruction of LHCb events . . . 24
2.1.1 Simulation . . . 24
2.1.2 Reconstruction . . . 25
2.1.3 RICH PID performance . . . 27
2.2 The case for neural network RICH PID . . . 29
2.2.1 The advancement of deep learning in LHCb . . . 29
2.2.2 Experimental conditions and hardware change . . . 30
2.3 Machine learning in a nutshell . . . 32
2.3.1 Model training . . . 32
2.3.2 Performance metrics, loss functions, and Occam's razor . . . 33
2.3.3 No free lunch and regularisation . . . 35
2.3.4 Hyperparameters and cross-validation . . . 38
2.3.5 Deep learning . . . 39
2.3.6 Error backpropagation . . . 46

2.3.7 Convolutional neural networks . . . 47
2.4 A convolutional neural network for RICH PID . . . 49
2.4.1 The input data . . . 49
2.4.2 Feature engineering . . . 53
2.4.3 The PID neural network . . . 57
2.4.4 Computing resources . . . 61
2.5 CNN optimisation and PID performance . . . 61
2.5.1 Performance benchmark . . . 62
2.5.2 CNN hyperparameter optimisation . . . 62
2.5.3 richNet PID performance . . . 73
2.5.4 Reducing the background . . . 80
2.6 Outlook for CNNs in LHCb RICH . . . 85

3 Photonic crystals as alternative Cherenkov radiators for particle identification 87
3.1 Particle identification with the LHCb RICH detector . . . 88
3.2 Photonic crystals . . . 90
3.2.1 Field properties . . . 91
3.2.2 Light propagation . . . 93
3.3 Photonic crystals as effective Cherenkov radiators . . . 95
3.3.1 The potential of photonic crystals PID in particle physics . . . 95
3.3.2 Resonance transition radiation simulation . . . 98
3.4 Polymer photonic crystal laboratory measurements . . . 99
3.4.1 Experimental setup . . . 99
3.4.2 Laboratory measurements and results . . . 104
3.4.3 Polymer photonic crystal irradiation . . . 108
3.5 Manufacturing a high-energy photonic crystal . . . 110
3.5.1 Simulated SiO_2 | TiO_2 effective Cherenkov radiation . . . 110
3.5.2 Thin-layer coating . . . 116
3.6 SiO_2 | TiO_2 photonic crystal high energy beam tests . . . 122
3.6.1 Beam test setup . . . 122
3.6.2 Analysis and results . . . 128
3.6.3 Interpretation and future experiments . . . 141
3.7 Outlook for photonic crystals at CERN . . . 144

Appendix A RICH PID CNN figures 147

Appendix B A Brewster route to PID 159

Bibliography 165

Introduction

In this thesis, I present two potential routes to advance the particle identification of Ring-Imaging Cherenkov detectors to overcome their intrinsic limitations and prepare the experiment for the challenging high-luminosity environment of the Large Hadron Collider (LHC). Although both proposed innovations are suited for other experiments and laboratories, they are best understood in the context of the evolution of the LHCb (Large Hadron Collider beauty) experiment. Chapter 1 provides an overview of the ongoing detector upgrade and future changes to its subdetectors.

The RICH particle identification (PID) algorithm combines information from the two RICH photodetector planes and the tracking stations to formulate a particle hypothesis, iterating through a multitude of combinations. To meet the demands of the steep increase in occupancy of future LHC runs, changes to the LHCb computing structure towards multi-core architectures have been proposed. Although the RICH PID algorithm has performed reliably, it is not well suited for efficient parallelisation.
In Chapter 2, I introduce a convolutional neural network approach to particle identification in LHCb RICH. The parallelisable calculations of the network make it suitable for fast online event reconstruction and align well with technological developments. Its PID performance in the upcoming LHC run period is assessed and compared to the current algorithm in simulation. The impact of combinatorial background mitigation is explored, as potentially provided by timing information in future detector upgrades.

In Chapter 3, I investigate the possibility to use photonic crystals as a Cherenkov radiator alternative for particle identification, such as in the LHCb RICH detectors. Photonic crystals are periodic optical nanostructures that allow controlling the flow of photons. In recent years, the features of photonic crystals have been exploited to create an effect analogous to the conventional Cherenkov effect without many of its constraints. Following laboratory studies of a polymer photonic crystal, the manufacturing process of a one-dimensional dielectric photonic crystal is presented together with the first tests in a high energy particle beam.

Chapter 1

The Large Hadron Collider and the LHCb experiment

The research presented in this thesis was conducted at the European Organisation for Nuclear Research (CERN). Here, I review the current state of the LHC and LHCb and the plans for their future evolution. The LHC and its detectors undergo substantial upgrading periods during their lifetime. After a brief overview of the accelerator and the four major experiments, LHCb is introduced in more detail. The focus lies on its sub-detectors and their transformation over time, which is motivated by improving the physics performance and adapting to the increasingly demanding environment of the LHC.

1.1 The LHC and its four main experiments

CERN was inaugurated in 1954 by its 12 founding member states in Geneva, Switzerland, with the mission to study nuclear physics. Soon, the research focus shifted to high-energy physics, which motivated the creation of particle colliders for the understanding of subatomic structures.

Since its foundation, CERN has enabled numerous discoveries and developments. To name a few, the detection of neutral currents in 1973 [1], the discovery of W and Z bosons in 1983 [2, 3], the first antihydrogen atoms in 1995 [4], and the discovery of the Higgs boson in 2012 [5, 6].

In 1989, the Large-Electron-Positron Collider (LEP) started running, the world's largest particle accelerator with a circumference of 27 km, up to 175 m below the surface, straddling

the border between France and Switzerland. After the decommissioning of LEP in 2000, the tunnel was re-purposed to host the Large Hadron Collider, to this day the world's largest and most powerful particle accelerator. Since its start in 2010, the centre-of-mass collision energy has been increased from 7 TeV to 13 TeV, by the end of 2018. To reach these high energies, the particles undergo a series of acceleration stages, as illustrated in Figure 1.1. First, they are brought up to 50 MeV in the LINAC 2 linear accelerator. The Proton Synchrotron Booster (PSB) accelerates the particles to 1.4 GeV, the Proton Synchrotron (PS) to 25 GeV, and the Super Proton Synchrotron (SPS) to 450 GeV. From there, the protons are injected into the LHC, where they circulate in bunches, each containing about 115×10^9 particles. Up to 2,808 bunches circulate simultaneously per beam, in opposite directions and in separate beam pipes. The beams intersect in four interaction points where the bunches cross a rate of 40 MHz and a luminosity of 2×10^{33} cm^{-2}s^{-1}.

The four largest experiments are ALICE, ATLAS, CMS, and LHCb. ATLAS (A Toroidal LHC ApparatuS) [8] and CMS (Compact Muon Solenoid) [9] are general-purpose detectors. Their main goal is to test the Standard Model (SM) of Particle Physics, particularly investigating the Higgs Boson and searching for new physics. They are positioned on opposite sites of the LHC. They are independently developed, and their data are analysed separately to validate or falsify discoveries. ATLAS is the world's largest detector, with a length of 46 m and a diameter of 25 m. CMS is the heaviest detector, with a weight of 14.000 t, mainly due to its enormous superconducting solenoid magnet generating a 4 T field.

The focus of the ALICE (A Large Ion Collider Experiment) experiment [10] is the study of quantum chromodynamics (QCD) by investigating quark-gluon plasma, created at extreme energies and temperatures. That is mainly achieved through the collision and measurement of Pb-Pb collisions during dedicated heavy-ion runs.

The LHCb experiment [11] is dedicated to the understanding of the matter-antimatter asymmetry present in the universe by studying rare charge-parity (CP) symmetry violating hadronic decays, involving charm (c) and beauty (b) quarks. LHCb and its goals and upgrade plans are described in more detail in Section 1.2.

To increase the search range for physics beyond the Standard Model and the statistical significance for rare processes, the LHC and its experiments are shut down for extended periods to upgrade their components. During the first long shutdown (LS1) in 2013 and 2014, the LHC maintenance work was performed on its superconducting magnets. In the subsequent LHC Run 2, the centre-of-mass collision energy was gradually increased, reaching 13 TeV at the end of the run in 2018. Since then, the LHC has entered the long shutdown 2 (LS2), which is foreseen to last until the beginning of 2022. One of the main objectives of the LS2

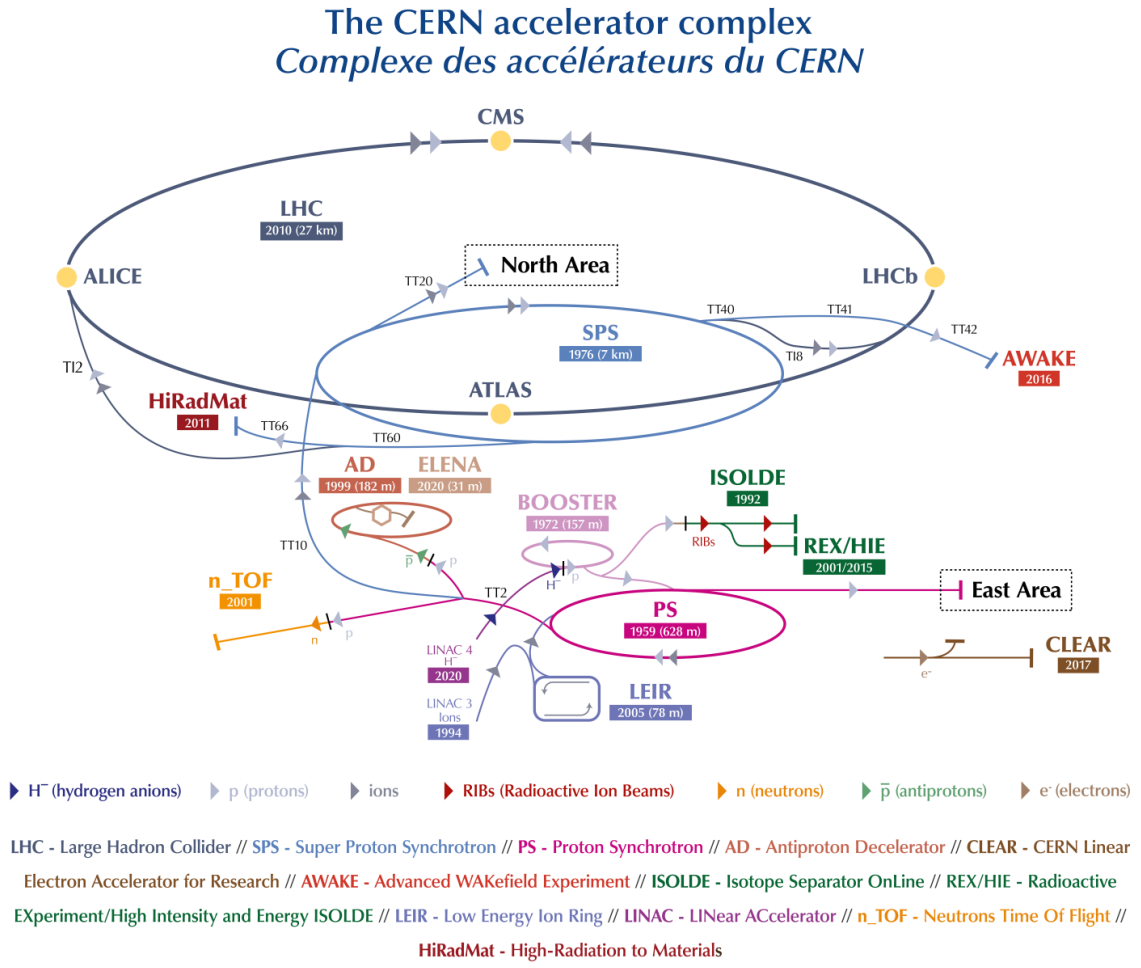


Fig. 1.1 The CERN accelerator complex. The particles are boosted to high energies by a set of acceleration stages before reaching its final energy in the LHC (dark blue line). The independent CLEAR accelerator structure, in which this thesis' beam tests are conducted, is illustrated on the bottom right. Source: [7].

is the preparation of the LHC for its high-luminosity operation (HL-LHC). For that purpose, the LINAC 2 accelerator stage will be replaced by LINAC 4, where low-energy negative Hydrogen ions (H^{-2}) will be accelerated instead of protons. With the new linear accelerator, the energy of the particles delivered to the PSB increases from 50 MeV to 160 MeV and more beam parameters can be adjusted, such as its energy spread. The commissioning of the HL-LHC will take place after LS3, where the machine is expected to deliver a luminosity of 5×10^{34} cm^{-2}s^{-1}. The upgrade plans of the HL-LHC are described in Reference [12].

The LHCb experiment is subject to significant modifications during LS2, called Upgrade Ia. The major change will see a triggerless readout to allow operating at the full potential of the LHC luminosity and 40 MHz collision rate. ATLAS and CMS will undergo an extensive revision during the next long shutdown 3 (LS3), which is currently foreseen to commence in 2025.

1.2 The LHCb experiment

The LHCb experiment aims to understand the underlying reasons for the matter-antimatter asymmetry in our universe. Section 1.2.1 provides a brief overview of the questions the LHCb experiment is designed to investigate. The LHCb detector and its upgrade plans are described in 1.2.2. The software tools used to simulate and reconstruct LHCb events are described in the introduction to Chapter 2.

1.2.1 The LHCb physics case

The stark contrast between matter and antimatter in the universe represents one of the great mysteries that the Standard Model of particle physics does not explain. This disparity has been the focus of attention of several generations of physicists ever since the first observation of antimatter in 1932 [13]. A source of *CP* violation beyond the SM is required to explain this asymmetry, among other reasons [14].

CP symmetry violation was first observed in neutral kaons in 1964 [15] and its discovery was awarded with a Nobel Prize in 1980. The Standard Model accounts for this source of CP violation in its description of weak interactions. The Cabibbo-Kobayashi-Maskawa (CKM) matrix describes quark flavour transitions in weak interactions [16]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.1}$$

The CKM matrix hence connects the mass eigenstates for the quarks (d,s,b) with their weak eigenstates (d',s',b'). The values of the CKM matrix V_{CKM} are not defined by the Standard Model but are constraint by the condition of unitarity ($V_{CKM}^{\dagger}V_{CKM}=1$), on the basis of which the matrix is constructed. The CKM matrix can be expressed in the so-called *Wolfenstein parameterisation* [17], defining $\lambda := |V_{us}| \approx 0.2$ and the parameters A,ρ,η:

$$V_{CKM} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho-i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4) \tag{1.2}$$

The parameterisation reflects the hierarchy, suppressing transitions between different generations by powers of λ.

The values of the matrix parameters have to be measured experimentally. From the unitarity condition of the CKM matrix, six constraints follow. These constraints impose that the products of the matrix elements from different rows and columns equal 0. Each of these can be represented by a closed triangle in a complex plane. The triangle following from the condition $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$ is known as the *Unitary Triangle*. All sides of the triangle are of the order of magnitude λ^3.

The Unitary Triangle angles can be expressed in terms of CKM matrix elements:

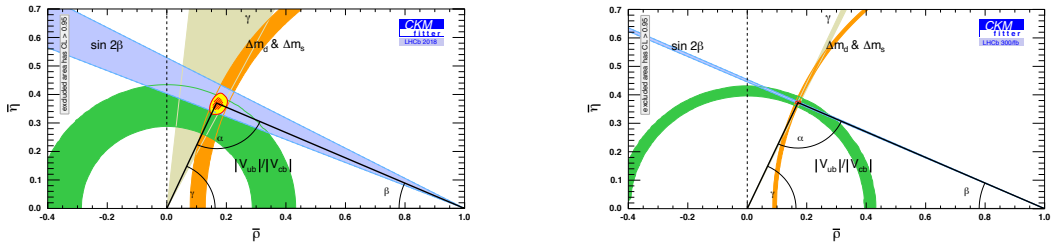
$$\alpha := arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta := arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \gamma := arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{1.3}$$

Correspondingly, the sides of the triangle are given by:

$$\left|\frac{V_{ud}V_{uc}}{V_{cd}V_{cb}}\right| = \sqrt{\rho^2+\eta^2}, \qquad \left|\frac{V_{td}V_{tb}}{V_{cd}V_{cb}}\right| = \sqrt{(1-\rho)^2+\eta^2}. \tag{1.4}$$

The Standard Model CP violation is a defining feature of the Unitary Triangle. Any deviations of the Unitary Triangle parameters from the aforementioned constraints would hint at physics beyond the Standard Model. An experimental gateway to explore potential deviations is beauty hadron decays, particularly decays of B mesons. The comparatively long lifetime of B mesons represents a distinct signature that allows for an excellent selection of relevant decays and rejection of background events in an experiment.

LHCb is designed to exploit the experimental signature of B-meson decays. It expands on previous and ongoing results from other experiments dedicated to the study of B mesons, BaBar, and Belle [19, 20]. Figure 1.2a illustrates the constraints of the Unitary Triangle after



(a) (b)

Fig. 1.2 LHCb experimental constraints of the unitary triangle in the $\bar{\rho} - \bar{\eta}$ plane after LHC Run 2 (a) and projection for after HL-LHC Run 6 (b). Source: [18].

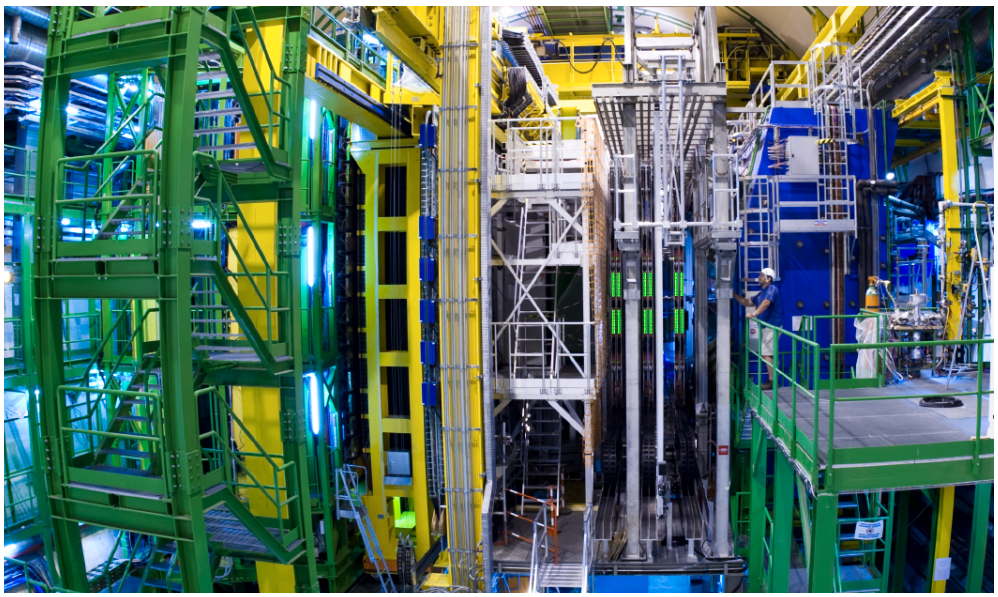
LHC Run 2. Figure 1.2b extrapolates the statistical precision obtained after Run 6 when LHCb is expected to have collected 300 fb^{-1} of data. To discover deviations from the SM in the Unitary Triangle, large amounts of data and high precision are essential, such as in identifying particles, which is the subject of this thesis.

1.2.2 The LHCb detector

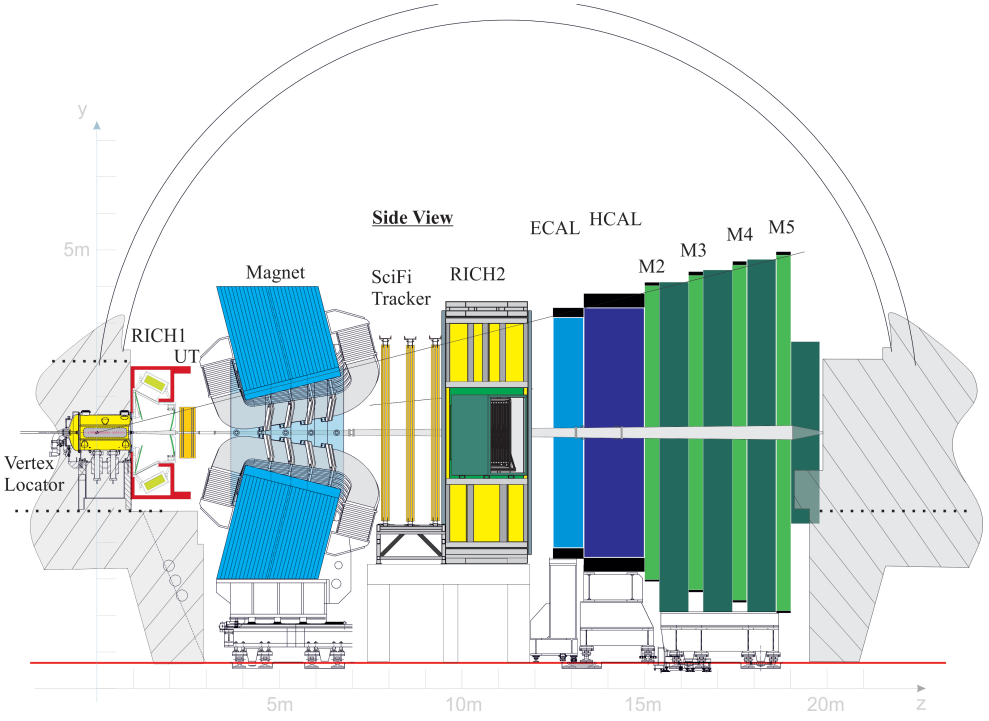
The LHCb research programme has the primary goal of investigating interactions involving beauty and charm hadrons primarily produced at small angles with respect to the beam. The single-arm forward spectrometer geometry of the LHCb detector is ideal for this type of measurement. It covers a polar angular range between 10 and 250 mrad, corresponding to a pseudorapidity range of $\eta \in [2,5]$ [21], defined as

$$\eta = -\log\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{1.5}$$

where θ denotes the polar angle. A picture of the LHCb detector is shown in 1.3a. LHCb consists of several sub-detectors, each serving a distinct purpose. Tracking layers are closely surrounding the interaction point (VELO), and large-area tracking stations, the Tracker Turicencis (TT) and three more tracking layers (T1-T3) are located upstream and downstream of the dipole magnet, respectively. The magnet creates a field with an average bending power of 4 Tm. The two Ring-Imaging Cherenkov detectors (RICH 1 and 2) provide particle identification information. RICH 1 is located close to the interaction point, next to the VELO, and RICH 2 downstream of the three tracking layers, T1-T3. Two calorimeters measure the energy of the particles (ECAL and HCAL). Muon stations are located downstream of the calorimeters. Apart from muons (or particles decaying to muons), all particles are generally absorbed by the previous detector stages, making these last stations ideal for identifying muons, representing a valuable decay signature.



(a)



(b)

Fig. 1.3 Picture of the LHCb detector (a) (reprinted and modified from [22]), with the *pp* collision point on the right. Schematic of the LHCb Upgrade Ia detector and its sub-detectors (b): VELO (Vertex Locator), RICH1 and RICH2 (Ring Imaging Cherenkov detectors), UT (Upstream Tracker), SciFi (Scintillating Fibre) Tracker, ECAL (Electromagnetic Calorimeter), HCAL (Hadronic Calorimeter), M1 to M5 (Muon stations). The *pp* collision point is on the left. Source: [23].

1.3 The LHCb upgrades

LHCb is currently undergoing a major revision called Upgrade Ia. It is neither possible nor desirable to record all data at the nominal collision rate of 40 MHz. The LHCb trigger system selects bunch crossings containing interactions of interest to the physics programme. The changes to the trigger system during Upgrade Ia are the major driver behind the modifications of LHCb's sub-detector stages. The RICH photodetectors and entire read-out electronics will be exchanged. Additionally, the TT and the three downstream tracking stations T1-T3 will be replaced by the Upstream Tracker (UT) and the Scintillating Fibre (SciFi) tracker, respectively.

While only minor changes are foreseen for the Upgrade Ib during LS3 (around 2025-2027), plans and R&D for the Upgrade II during LS4 (around 2031-2032) include significant changes to the SciFi tracker (then called Mighty Tracker) and adding a system to identify charged hadrons in the low momentum range between 1 to 10 $\mathrm{GeV/c}$, called TORCH (Time of Internally Reflected Cherenkov light).

The sub-detectors of LHCb and the upgrade plans are discussed in this section. The detector functions can broadly be categorised as particle tracking and particle identification. The track reconstruction system will be described in Section 1.3.1 and the particle identification detectors in Section 1.3.2. The RICH PID system (Section 1.3.2) will be discussed in more detail since it is the subject of this thesis. Replacing the hardware-trigger system by event-selection in software represents one of the major changes in Upgrade Ia (Section 1.3.5). That requires modifications of the front-end electronics of all sub-detector systems. A schematic overview of the Upgrade Ia detector is shown in Figure 1.3b.

1.3.1 The particle tracking systems

The LHCb tracking system measures the momentum of charged particles with an uncertainty ranging from 0.5 % for low momenta to 10 % at 200 $\mathrm{GeV/c}$. The impact parameter, which refers to the closest distance between a particle track and a primary vertex (PV), can be measured with a precision of $(15 + 29/\mathrm{pT})$ μm, where pT is the transverse momentum component (in GeV/c) with respect to the beam direction [24], as shown in Figure 1.4a.

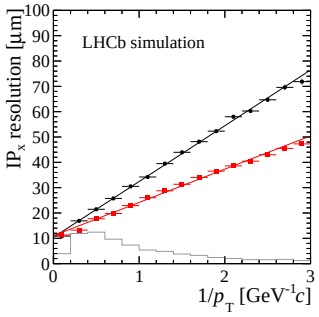
The Vertex Locator

The VELO [25, 26] sensors are located as close as possible to the interaction point. That reduces the uncertainty of the extrapolation of the first measurement to the interaction point. It consists of a series of 26 pixel modules, arranged along the beam direction.The beauty and

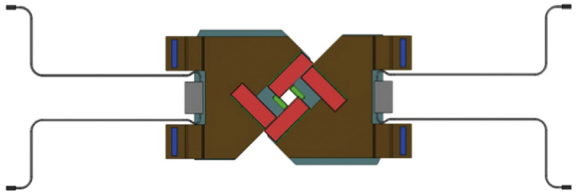
charm hadrons, which LHCb is designed to study, typically decay after a duration of $O(1\text{ ps})$, travelling several millimetres. A precise measurement of the perpendicular distance of the secondary vertex to the beam axis, the impact parameter, is essential. It allows determining whether a track originates from a primary or secondary vertex, the latter indicating decay of a particle with a long lifetime, such as b-hadrons.

A thin aluminium foil (RF foil) separates the beam region and the detector instead of a beam pipe. The RF foil is as thin as possible to minimise elastic scattering of the particles before reaching the silicon sensors, which would decrease the angular resolution.

The sensitive area of the Upgrade Ia VELO pixel sensors starts at 5.1 mm distance from the beam, a reduction of the distance compared to the previous VELO silicon strips, which were positioned at 8.2 mm. The detector consists of two parts that can be retracted during beam injection to prevent sensor damage. The sensors, shown in Figure 1.4b, are slightly overlapping to facilitate detector alignment.



(a)



(b)

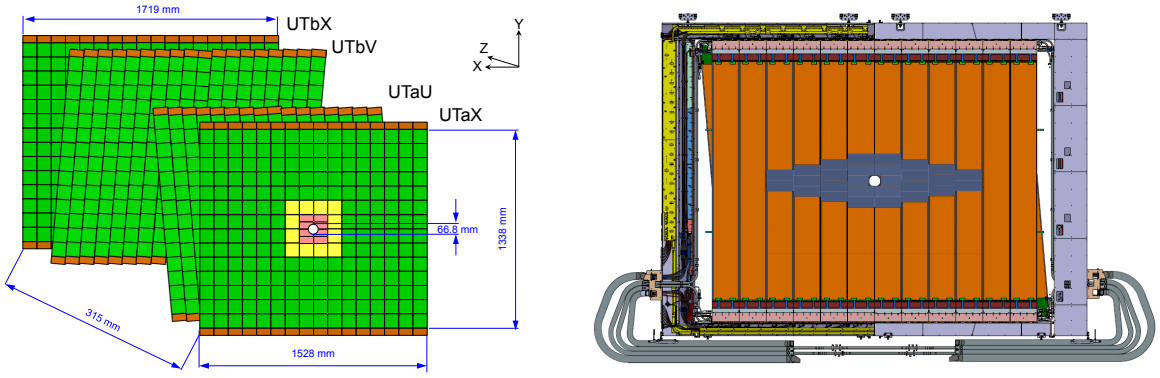
Fig. 1.4 (a) Impact parameter resolution before (black) and after (red) LHCb Upgrade Ia with the b-hadron child track population for each $1/p_T$ bin. Source: [25]. (b) Cross-section of a closed Upgrade Ia pixel sensor module in the xy plane. Source: [26].

The $55 \times 55\ \mu\text{m}^2$ pixel sensors are read out at 40 MHz by a custom ASIC, called VeloPix, at 20.4 Gb/s [27]. The close proximity of the sensors to the interaction points will lead to a very high particle flux. The sensors are hence designed to withstand a radiation dose up to $8 \times 10^{15}\text{MeV n}_{eq}/cm^2$ during their lifetime while maintaining a 99 % hit efficiency.

The high-occupancy environment of the HL-LHC during Run 5 will require additional time information to avoid mis-association of primary vertices [28]. Including the time dimension is a recurring theme for HL-LHC detectors. That way, the combinatorial background can be reduced much more efficiently than by improving the spatial resolution solely. In the case of the VELO, time information can be handed as a PV timestamp to the downstream detector systems to facilitate track association.

The Upstream Tracker

The Upstream Tracker will replace the Tracker Turicencis during Upgrade Ia [23]. The TT silicon sensors are not suited for the high radiation doses during Run 3, and its integrated front-end electronics can not be operated at the LHC collision frequency of 40 MHz. The UT will consist of four silicon-strip sensor layers with the *x*-*u*-*v*-*x* layout that has been used for the TT, where layers *u* and *v* are rotated by $-5°$ and $5°$, respectively, as shown in Figure 1.5a. Since the dipole magnet bends the particle trajectory in the horizontal plane, the silicon strips



(a) (b)

Fig. 1.5 (a) Schematic of the Upstream Tracker silicon strip layer modules. Source: [23]. (b) The Mighty Tracker, a modification of the SciFi tracker (orange) for Upgrade II, with added silicon modules (blue) close to the beamline. Source: [29].

in the UT modules are arranged vertically. To accommodate the higher track density near the beam axis, sensors closer to the beam are thinner and shorter, providing a better spatial resolution and reducing the occupancy: the sensors in the regions coloured in yellow in Figure 1.5a are of nominal length but with a pitch of 95 μm, half of the nominal pitch. The sensor regions coloured in pink, which are closest to the beamline, are about 5 cm long in the Y direction, which is half the nominal pitch and half the nominal length in comparison to the sensors coloured in yellow.

The Scintillating Fibre tracker

The Scintillating Fibre tracker will provide momentum information, replacing the downstream tracking stations for LHC Run 3 [23]. It consists of 6 modules, each composed of 8 sub-layers of fibres with a length of 2.4 m and a fibre diameter of 250 μm [30]. The SciFi tracker covers a surface of 6 m $\times$ 5 m in the *xy* plane. The light generated in the scintillating fibres will be detected using 128-channel Hamamatsu[1] Silicon Photomultiplier (SiPM) arrays, positioned

[1]Hamamatsu Photonics: `https://www.hamamatsu.com`

on the top and bottom end of the fibre mats. While SiPMs feature an excellent quantum efficiency of up to 45 %, their Dark Count Rate (DCR) increases linearly as a function of neutron flux and exponentially with temperature [23]. To mitigate these two effects, the SciFi SiPMs will be cooled to a temperature of $-40\,^{\circ}\mathrm{C}$.

The high-radiation environment of the HL-LHC will pose further challenges to the SciFi tracker. The SiPMs will be annealed during shutdown periods to reduce the DCR. Studies are ongoing to determine whether heating of the SiPMs during annealing is beneficial. Additionally, it is foreseen to replace the region of the fibre mats closest to the beampipe and most affected by the radiation by silicon modules. The new hybrid between fibre and silicon will be renamed Mighty Tracker (MT), shown in Figure 1.5b.

1.3.2 Particle identification

The particle identification systems are essential to differentiate between decay final states. Three types of detector are used for this purpose: the two ring-imaging Cherenkov detectors (Section 1.3.2), the calorimeters (Section 1.3.3), and the muon stations (Section 1.3.4). An additional detector, TORCH, has been proposed for installation during Upgrade Ib, to cover PID in the momentum region below $10\,\mathrm{GeV/c}$, and will be introduced in this section.

The RICH detectors

The two Ring-Imaging Cherenkov detectors are responsible for charged hadron identification. In particular, the RICH detectors distinguish between pions (π), kaons (K), and protons (p), covering the momentum range 10 to $100\,\mathrm{GeV/c}$, and even down to $2.5\,\mathrm{GeV/c}$ for kaon – pion separation. Most decay products are pions in the b and c hadron decays that are of interest to LHCb. Distinguishing these from other charged hadrons is essential to reduce the combinatorial background and different final states.

As mentioned in Section 1.2.1, a primary goal of LHCb is studying CP violation. A vital tool for these studies is two-body hadronic decays of the type $B \to h^+h^-$, where h represents a charged hadron. Examples of such decays would be $B_s^0 \to K^+K^-$ or $B^0 \to \pi^+\pi^-$. These decay channels illustrate the power of RICH PID increasing the sensitivity to CP violation, as demonstrated for the latter decay in Figure 1.6 [31]. The decay signal is illustrated by the light-blue dotted line, indistinguishable in the total distribution (solid blue line), the sum of all final states, as shown in the left plot. The other contributions are suppressed once the PID is performed, as shown in the right plot.

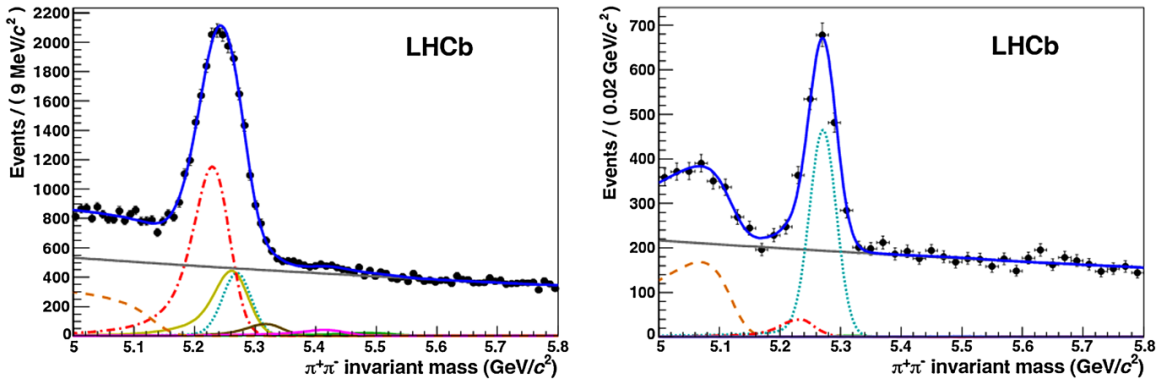
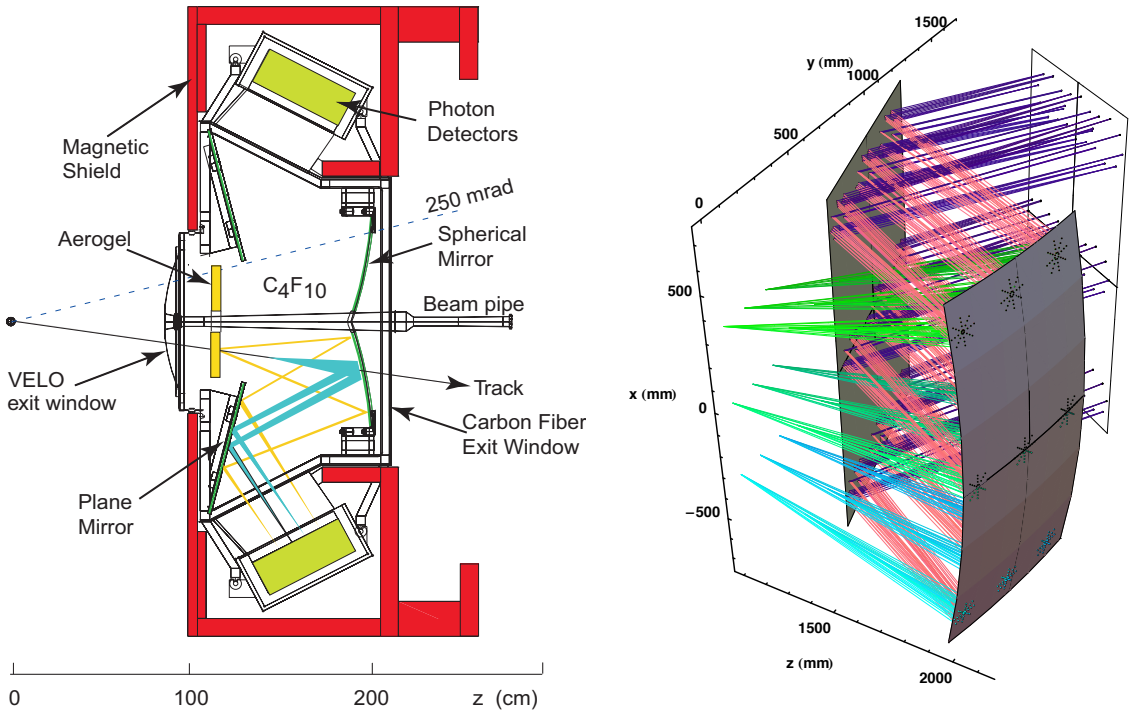


Fig. 1.6 Invariant mass distribution for $B^0 \to \pi^+\pi^-$ (turquoise dotted line) before (left) and after (right) adding RICH PID information. Various other b-hadron decay modes involving kaons, protons, or 3-body decays are eliminated through the PID. Source: [31].

The RICH system consists of two detectors. RICH 1 is placed close to the interaction point to minimise its size and therefore every effort has been taken to keep the material budget to a minimum before the track measurement while still covering the full angular acceptance of 25 to 300 mrad. RICH 1 covers the momentum range up to 40 GeV/c. RICH 2 is positioned downstream of the dipole magnet and tracking stations to reduce the material budget before the track measurement. It covers the momentum range from 15 to 100 GeV/c.

Both detectors contain fluorocarbon gas as Cherenkov radiators (C_4F_{10} in RICH 1 and CF_4 in RICH 2). The refractive indices are $n_{C_4F_{10}}(\lambda) = 1.0014$ and $n_{CF_4}(\lambda) = 1.0005$, for a wavelength of $\lambda = 400$ nm. To suppress scintillation, 5 % of CO_2 are added to the CF_4 radiator. During Run 1, 50 mm silica aerogel tiles were also present as Cherenkov radiators in RICH 1 to extend the PID performance below the C_4F_{10} kaon threshold ($n_{aerogel}(\lambda) = 1.03$, $\lambda = 400$ nm). It was removed during the LHC shutdown period between 2013 – 2014. The higher luminosity of Run 2 and the resulting increase in background photons from the aerogel rendered removing the aerogel and therefore reducing the material budget in the acceptance as overall beneficial to the PID and tracking performance of LHCb [32].

The two detectors feature a similar optical design. The Cherenkov photons generated in the gas are reflected by a set of spherical focussing and flat mirrors outside the detector acceptance and on to the photodetector planes, as shown in Figure 1.7 for RICH 1.

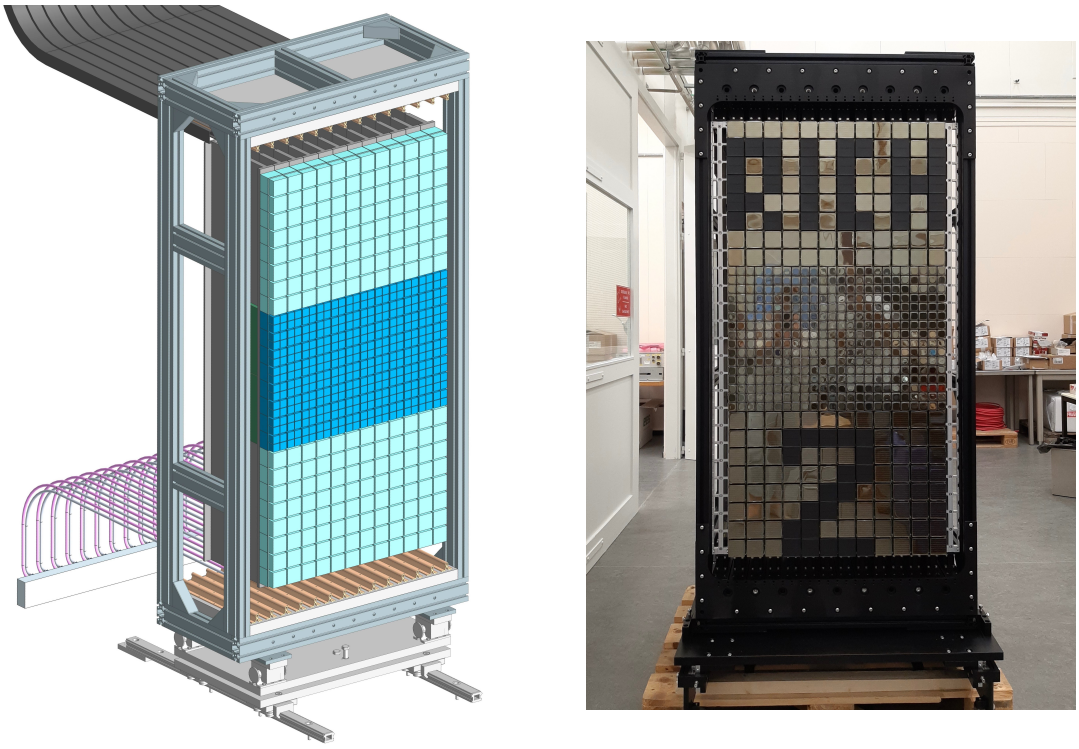


(a) (b)

Fig. 1.7 Schematic side-view of the RICH 1 detector (a). The aerogel was removed after Run 1. Source: [11]. Simulated Cherenkov photon path in the Upgrade Ia RICH 1 upper detector half (b). Source: [33].

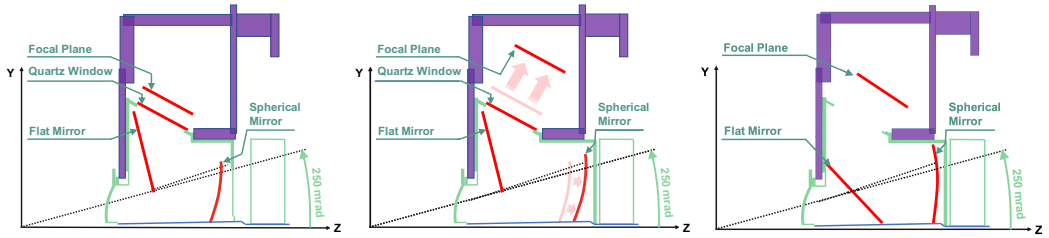
The photon detectors are sensitive to single photons. In LHC Run 1 and 2, hybrid photon detectors (HPDs) were used, with 1 MHz readout electronics inside their vacuum tube. To achieve a readout rate matching the LHC collision frequency of 40 MHz for Run 3 and beyond, the HPDs are replaced during Upgrade Ia by two types of 8×8 channel Multi-Anode Photomultiplier Tubes (MaPMT) by Hamamatsu and new readout electronics. A 1-inch R13743 MAPMT will populate RICH 1 and the inner region of RICH 2, while a 2-inch R13743 MaPMT will be located in the low-occupancy regions of RICH 2. One of the two RICH 2 photodetection planes featuring the 1- and 2-inch MaPMTs is shown in Figure 1.8.

In addition to changing the photon detectors, the optical system is modified in Upgrade Ia to adjust for the increase in luminosity and hence photon occupancy. By moving the spherical and flat mirrors as shown in Figure 1.9, the focal length is doubled, which increases the total photon detection area and reduces the pixel occupancy. Further changes are foreseen for Upgrade II, such as replacing the flat mirror by a mirror with a very low material budget. This allows placing the mirror inside the acceptance, hence reducing the overall emission point error of the system, as illustrated in Figure 1.9c.



(a) (b)

Fig. 1.8 Schematic view of one RICH 2 detector plane with mechanical assembly with 1-inch MaPMTs (dark blue) in the centre and 2-inch MaPMTs (light blue) in the low-occupancy regions (a). Picture of a fully assembled and tested RICH 2 detector plane half (b). Reprinted and modified from [34].



(a) (b) (c)

Fig. 1.9 (a) Schematic of RICH 1 upper half during Run 2. (b) Changes to the optical system after Upgrade Ia, doubling the focal length and reducing the photon occupancy per pixel. (c) Upgrade II modifications to RICH 1, moving the flat mirrors in the acceptance which reduces the optical aberration and hence the emission point error. Source: [35].

TORCH

To extend the *p*-*K* separation below the kaon threshold in C_4F_{10} gas of 9.3 GeV/c, the TORCH detector has been proposed, foreseen for installation during Upgrade Ia [36]. TORCH would be placed between the SciFi and RICH 2 detectors, identifying particles with a combination of time-of-flight measurement and Cherenkov radiation.

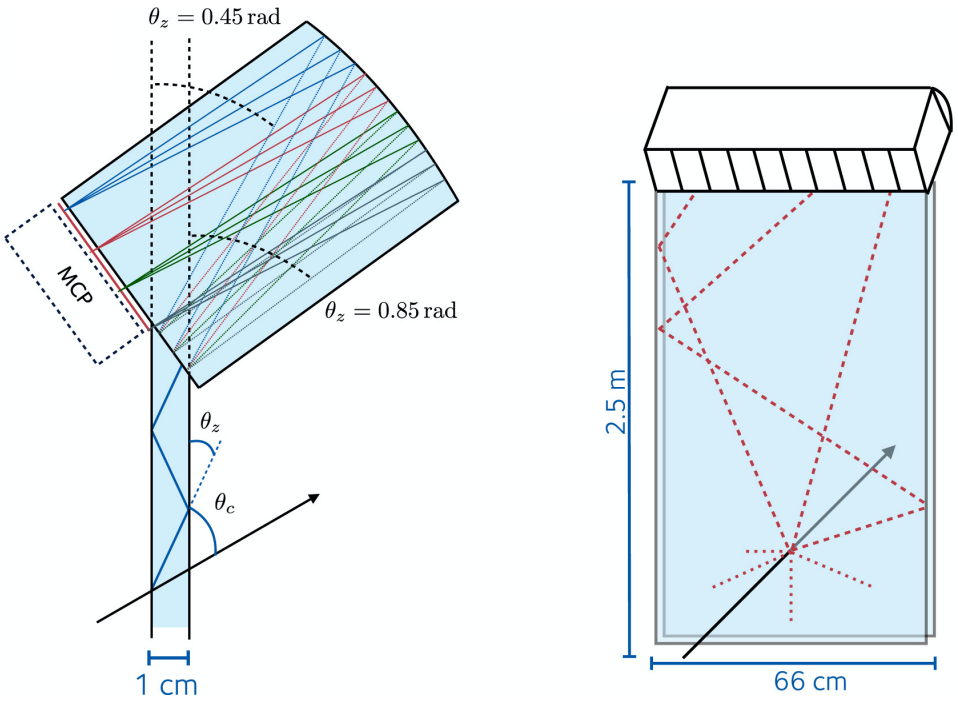


Fig. 1.10 (left) Schematics of a TORCH focussing block and MCP-PMT plane. (right) Potential Cherenkov reflection paths in a single module. Source: [37].

Cherenkov photons are generated by charged particles traversing the TORCH quartz modules. They are internally reflected and focussed onto a photon detector plane as shown in Figure 1.10. The TORCH PID is based on reconstructing both the Cherenkov angle and photon path length for a given particle track by combining the photon hit position and the information from the LHCb tracking system. Similar to the RICH reconstruction algorithm, explained in the introduction to Chapter 2, the algorithm iterates through several particle ID hypotheses and calculates a likelihood for each mass hypothesis.

R&D to develop Micro-Channel Plate Photomultiplier Tubes (MCP-PMTs) is continuing in collaboration with Photek[2]. Their excellent time resolution makes them good candidates

[2]Photek (UK) Ltd.: `https://www.photek.com`

for TORCH, but their ageing under conditions of high detector occupancy is still under evaluation [38].

1.3.3 Calorimeters

The calorimetry system of LHCb consists of the scintillating pad detector (SPD), pre-shower detector (PD), electromagnetic calorimeter (ECAL), and hadronic calorimeter (HCAL). The system serves two primary purposes: particle identification (mainly of electrons, photons, and neutral pions) and provides information for the level-0 (L0) trigger by selecting candidates with high transverse energy. Since a software trigger replaces the L0 trigger during Upgrade Ia, the SPD and PD are removed.It is hence foreseen to replace the HCAL with iron shielding for the muon during Upgrade II.

The granularity of the ECAL and HCAL is sufficient to operate at Run 3 luminosity. Hence, only their front-end electronics will be replaced during Upgrade Ia to transmit data at the total LHC bunch crossing frequency of 40 MHz [39]. The ECAL follows a shashlik design with alternating scintillator material and lead layers, with a total thickness of 25 radiation lengths (X_0). PMTs read out the wavelength-shifting fibres. To slow their ageing in the 5 times higher luminosity of Run 3, their gain is decreased by a factor ~ 5. Radiation damage affects the ECAL tiles around the beam pipe, which are planned to be replaced during Upgrade Ib. The HCAL consists of iron and scintillator planes, arranged parallel to the beam axis, similar to the design of the ATLAS TileCal technology [40]. The design of the ECAL and HCAL is shown in Figure 1.11.

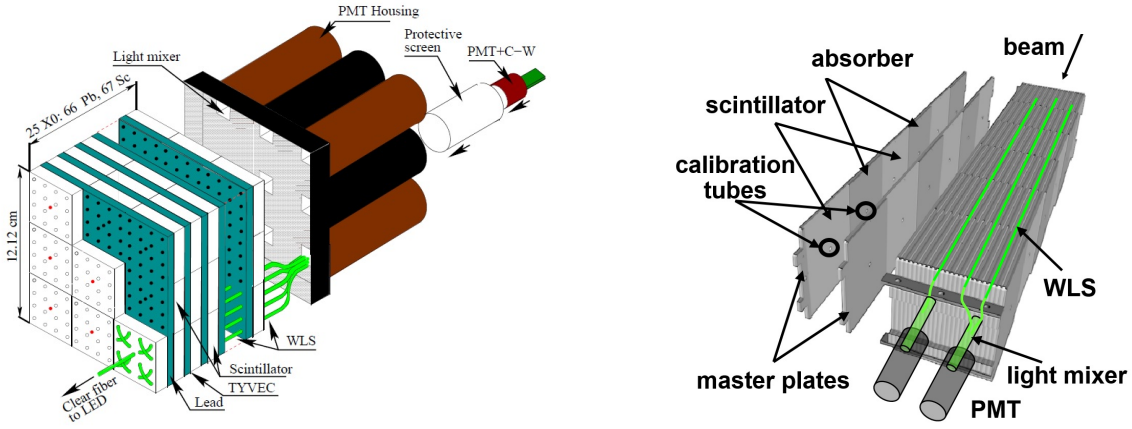


Fig. 1.11 Schematics of the ECAL (left) and HCAL (right). Source: [41].

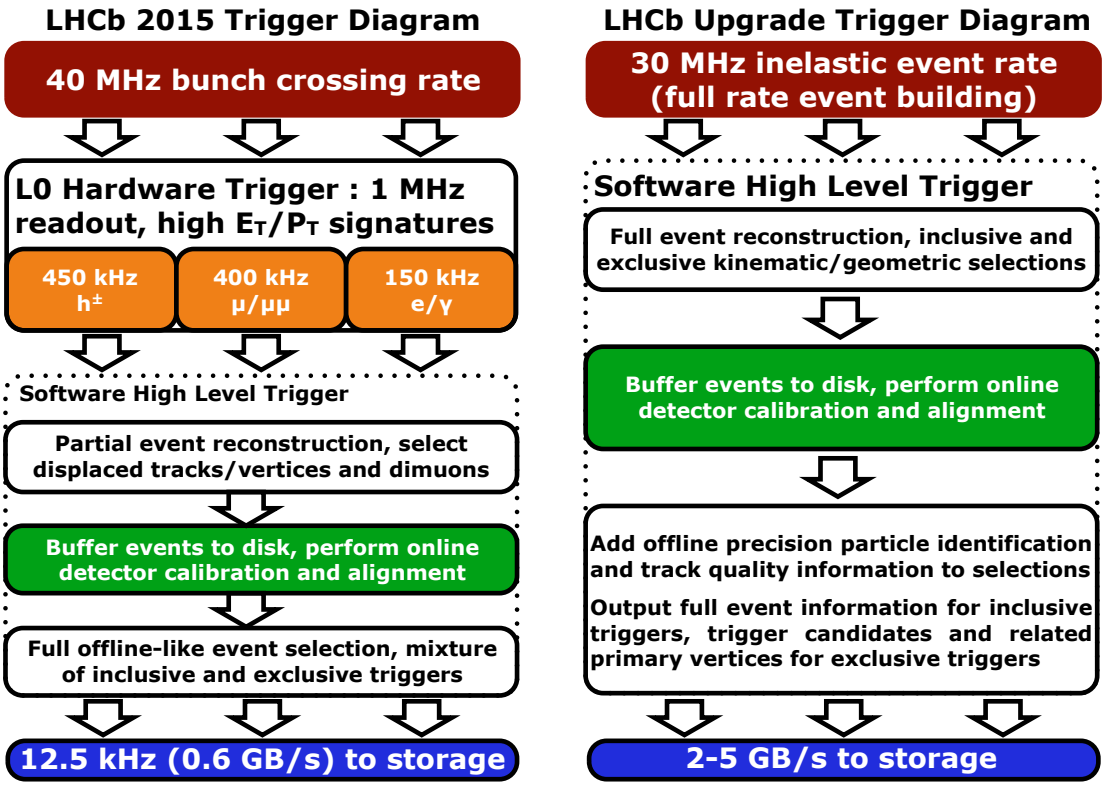
1.3.4 Muon stations

The 5 muon stations were used to provide information to the L0 trigger during Run 1 and 2 [42]. The first of the rectangular detectors was positioned downstream of RICH 2 and is replaced by neutron shielding during Upgrade Ia. During the upgrade, the off-detector readout electronics of the remaining 4 muon stations (M2 – M4) will be exchanged to increase the readout speed to 40 MHz. Their detection mechanism relies on Multi-Wire Proportional Chambers (MWPC). A second phase upgrade is planned to replace the inner regions of the M2 – M4 detectors with higher granularity detectors [43], either during LHCb Upgrade Ib or II.

1.3.5 Changes to the data processing

During Run 2, the trigger system consisted of two stages, the Level-Zero (L0) and the High-Level Trigger (HLT), shown in Figure 1.12a. The principle of the L0 trigger is to use prompt signals from the calorimeters and muon stages in order to filter events and reducing the rate from the collision rate of 40 MHz to 1 MHz. The 1 MHz rate is divided into the adjustable triggers for hadrons (450 kHz), the muons (400 kHz) and the photons/electrons (150 kHz). The L0-selected events are passed on to the HLT, operating on an Event Filter Farm (EFF). The HLT is divided into two stages; HLT1 is a quasi-real-time track-based trigger with an output rate of about 100 kHz, which are subsequently handed to the HLT2. HLT2 asynchronously reconstructs the events of the aligned and calibrated sub-detector systems, using the same reconstruction algorithms as used in offline processing.

In Upgrade Ia, the L0 trigger is removed. The HLT strategy remains the same but with significant performance improvements, allowing to process the entire 30 MHz inelastic event rate. Signals from the sub-detector front-end electronics are transferred to PCIe40 readout boards at a rate of 100 GBit/s. Data is passed to HLT 1 at a 30 times higher rate and 5 times of the pileup compared to Run 2. HLT 1 performs the full charged particle track reconstruction and reduces the event rate by approximately a factor of 30. Events, including bunch crossing and sub-detector information, are buffered to CPUs. On HLT 2, the events are processed with offline-quality track reconstruction, particle identification, and complete track fitting, using the sub-detector calibration and alignment. The filtered and reconstructed data is stored at a rate of 2 – 5 GB/s.



(a) (b)

Fig. 1.12 LHCb trigger schemes for Run 2 (a) and Run 3 (b). The L0 hardware trigger is removed during Upgrade Ia and replaced by a software high level trigger, driving the change of the sub-detectors' readout electronics to allow a rate of 40 MHz. Source: [44].

A GPU-based implementation of the HLT1 has been selected as the baseline for Run 3 and beyond [45]. In addition to cost reductions compared to a complete CPU infrastructure, future gains in performance are expected due to the quickly advancing technology of GPUs as general-purpose processors. These changes motivate exploring new software approaches which benefit from this type of technology, as is discussed in Chapter 2.

1.4 The evolution of the LHCb experiment

The LHCb experiment has lead to many fundamental discoveries to date. To extend the success of its physics programme into the future, the experiment will have to operate at a much higher intensity. This is provided by the high-luminosity LHC, where the track multiplicity and hit occupancy will increase by an order of magnitude.

One crucial element in handling the drastic increase in multiplicity is exploiting parallel data processing during the reconstruction and selection of events. This is implemented in the ongoing upgrade by exchanging part of the computing infrastructure from Central Processing Units (CPUs), which are optimised for complex tasks and low latency, to multi-core architectures. That change provides an opportunity to benefit from the quickly evolving fields of machine learning and neural networks, which are well suited for high-throughput processor applications. The future LHCb experiment is likely to gradually replace or complement many of its existing software elements with machine learning counterparts. As a contribution to this evolution, the studies presented in Chapter 2 of this thesis demonstrate reliable particle identification with a convolutional neural network approach. The network performs particularly well under conditions of reduced combinatorial background, as expected by the integration of hit time information during LHCb Upgrade Ib and II, which may render it an essential addition to the online reconstruction framework of the experiment.

Maintaining the excellent particle identification performance of the LHCb RICH detectors in the environment of the HL-LHC or future hadron colliders may call for a major redesign of the Cherenkov detector systems. Increasing the sensitivity to a broader momentum range and minimising the spatial footprint required by the large gas chambers of conventional RICH detector could be achieved by photonic crystals as Cherenkov radiator alternatives. Their potential to generate an effect similar to Cherenkov radiation that can be flexibly tuned in the forward and backward direction while requiring significantly less space enables new detector designs. Chapter 3 lays the groundwork for manufacturing photonic crystals at CERN, as well as laboratory and beam environment studies. The ability to manufacture and test photonic crystals at CERN represents a powerful tool for R&D programmes to study their potential future applications in particle physics experiments.

Chapter 2

A deep learning approach to particle identification in Cherenkov detectors

In this chapter, I present a deep learning approach for particle identification in Cherenkov detectors. The method is developed for the LHCb RICH detector but may be extended to similar detector types and other experiments. The PID of the RICH detectors is an essential element of the LHCb experiment, and its performance has been excellent in LHC Run 1 and 2, a trend that is expected to continue in Run 3.

The simulation and reconstruction in LHCb, with a focus on RICH, are outlined in Section 2.1. Changes to the LHCb computing infrastructure have been proposed, shifting from Central Processing Units (CPUs) to Graphics Processing Units (GPUs). These are better suited for high-throughput applications and become vital in the high-occupancy environment of the HL-LHC. Section 2.2 motivates the use of neural networks for RICH PID in future LHC runs.

The underlying principles of machine learning and, in particular, convolutional neural networks for high-energy physics applications are introduced in Section 2.3. The proposed neural network for RICH PID is introduced in Section 2.4, together with the input data generation and pre-processing.

The standard metric to assess the RICH particle identification performance is the ability to separate pions and kaons. The neural network's PID efficiency and hyperparameter optimisation using simulated data for LHC Run 3 are presented in Section 2.5. Subsequently, the performance with reduced combinatorial background is explored, as potentially provided

by timing information in future detector upgrades. The chapter concludes with a summary and outlook in Section 2.6.

2.1 Simulation and reconstruction of LHCb events

The studies presented here rely on an accurate and reliable simulation of the apparatus in order to generate realistic datasets under controlled conditions. This section introduces the simulation and reconstruction frameworks at LHCb together with the current RICH PID algorithm. Its understanding will motivate that, although performing exceptionally well, it may be challenging to scale to higher occupancy due to the increasing number of photon-track combinations.

2.1.1 Simulation

The LHCb simulation software framework is called GAUSS [46, 47] and it simulates datasets in three main steps. Firstly, it generates simulated proton-proton collisions, using a customised version of the event generation framework PYTHIA [48, 49]. The decay of the generated hadrons is simulated using the framework EVTGEN [50] and PHOTOS [51] for final-state QED corrections.
Following this, the particle's path through the detector and its interaction with material is simulated, using the GEANT4 toolkit [52–54]. The entry and exit point of each particle through detector elements is recorded. These points, called Monte Carlo hits (MCHits), contain much more complete information of the processes inside a detector than the data collected during LHC run periods. This Monte-Carlo (MC) *truth* information is the key for physics analyses or detector optimisation.
Finally the detector response is simulated in a digitisation step using the BOOLE framework [55]. The reconstruction is the same for simulated data and LHC collision data and is discussed in the following section.

The study uses simulated data to assess the neural network's performance reliably and for its training. The accuracy of the simulation allows for training the neural network for PID with known particle labels and later using the weights of the trained network for PID of LHCb collision data.

2.1.2 Reconstruction

The event reconstruction in LHCb uses the custom BRUNEL software package [56], for both simulated data and data from the experiment. The goal is to determine the particle properties, such as the species, their energy and momentum, and from which primary vertex they originate. The track trajectories are reconstructed based on the combined information from the tracking system and the muon stations.

The RICH reconstruction relies on the input from the tracking detectors to identify charged particles. In summary, for each charged particle track, a particle ID is assumed. The algorithm calculates where Cherenkov photon hits would be expected on the RICH detector planes for this particle hypothesis. The algorithm then calculates the probability for the detected photons to originate from the track based on the particle hypothesis. The particle hypothesis that maximises this likelihood is then assumed to be the particle ID.

The reconstruction can be divided into four steps [57]. In the first step, the pixels on which hits were detected and their position in the RICH coordinate system are stored.

The second step operates on the tracks provided by the tracking system that intersect with at least one of the RICH radiator volumes. The particle trajectory is extrapolated through the RICH detectors, reflecting from the mirrors as if it were a photon, to obtain the extrapolated track position on the RICH detector panels. For each mass hypothesis (electron, muon, pion, kaon, proton, deuteron), the expected photon yield, Cherenkov angle and resolution are calculated. To determine in which range around the extrapolated track centre photons are expected, typically around $O(100)$ Cherenkov photons are ray traced from the track to the detector plane for each mass hypothesis which is the second most computing-intensive sequence of the algorithm. The ray tracing uses an analytical calculation of the reflections of the RICH mirrors.

In the third step, the track segment and pixel information of the previous steps are used to select Cherenkov photon candidates that are likely associated with a given track. The high number of combinations between pixels and segments and the complexity of the calculations make this the most computationally intensive part of the RICH reconstruction.

Finally, combining the results from the first three steps, the expected signal and background distribution for the entire event is calculated for a set of mass hypotheses. This distribution is compared to the observed distribution of photon hits in the detector for the respective event. From this comparison, a likelihood for the set of mass hypotheses is formulated.
Following [58], for a given event hypothesis $\boldsymbol{h} = (h_1, \ldots, h_N)$, where N is the total number of tracks and background sources and h_j either the mass hypothesis for track j or a background

parameter, the expected signal v_i for pixel i can be computed as

$$v_i(\boldsymbol{h}) = \sum_j^N a_{ij}(h_j), \tag{2.1}$$

where $a_{ij}(h_j)$ is the expected detected photoelectron number for the hypothesis h_j. The event likelihood can then be calculated as

$$\mathscr{L}(\boldsymbol{h}) = \prod_i^M \mathscr{P}_{v_i}(n_i), \tag{2.2}$$

where M is the total number of pixels and $\mathscr{P}$ the Poisson probability for the expected signal $v_i(\boldsymbol{h})$ and n_i the number of observed photoelectrons. The Poisson distribution is given by

$$\mathscr{P}_{v_i}(n_i) = \frac{e^{-v_i} v_i^{n_i}}{n_i!}. \tag{2.3}$$

The expected number of observed photoelectrons a_{ij} is calculated as

$$a_{ij} = \varepsilon_i \lambda(h_j) \int_M \int_i f_h(\theta, \phi) \mathrm{d}\theta \mathrm{d}\phi. \tag{2.4}$$

ε_i is the pixel detection efficiency, $\lambda(h_j)$ the expected Cherenkov photon number emitted by track j, and $f_h(\theta, \phi)$ the probability-density distribution for photons emitted by the track. This probability-density distribution presumes a Gaussian distribution of the Cherenkov angle θ_c and constant distribution of the azimuthal angle ϕ_c with respect to the track:

$$f_h(\theta, \phi) = \frac{1}{(2\pi)^{3/2}\sigma(\theta)} exp\left[-\frac{1}{2}\left(\frac{\theta - \theta_h}{\sigma(\theta)}\right)^2\right] \tag{2.5}$$

Here, $\sigma(\theta)$ is the single-photon resolution and θ_h the expected Cherenkov angle for hypothesis h.

The likelihood $\mathscr{L}$ is first calculated assuming all tracks are pions, representing the most abundant charged particle in the detector. The mass hypothesis is then changed for a track and $\mathscr{L}$ recalculated, choosing the mass hypothesis with the largest increase for the overall likelihood. This step is repeated iteratively for all tracks, maximising the likelihood for a final mass hypothesis of the entire event.

2.1.3 RICH PID performance

There are several intrinsic sources of error for the RICH detectors, which limit their PID abilities [31, 59]. The first one is the chromatic error. As explained in Section 1.3.2, Cherenkov photons are generated in the RICH gas radiators. The refractive index n of the radiator gases is a function of the photon energy, which is unknown. This results in an error of the reconstructed Cherenkov angle θ_c, which is a function of the refractive index: $cos\theta_c = \frac{1}{n(\nu)\beta}$.
Secondly, the pixel size of the photon detectors determines the resolution to measure the position of a photon hit, which is referred to as pixel error.
Thirdly, the focussing mirrors are tilted to move the photon detectors outside of the acceptance of the experiment. The resulting aberrations distort the focal plane, contributing to the performance error.
The fourth contribution, the emission point error, is due to the unknown position where Cherenkov photons are emitted from the particle traversing the radiators. The reconstruction assumes all are emitted from the mid-point of the particle trajectory through the gas. Also, the charged particle's trajectory is curved by the magnetic field, and the track's direction is not perfectly known at the point of emission.
All these effects limit the reconstruction of the Cherenkov angle, which is the critical parameter for the PID. In addition, there are fundamental limits given by the Cherenkov effect itself. These are explained in detail in Section 3.1. In summary, there is a lower and upper momentum limit for identifying particles using the Cherenkov effect. Cherenkov photons are generated only above a certain particle velocity $v_p > c/n$, setting a lower limit. Below that limit, particles can only be identified by the absence of Cherenkov photons, which is referred to as *veto mode*. For high momenta, the Cherenkov angle converges for all charged particles towards the same angle. Chapter 3 proposes a solution for these fundamental PID limitations dictated by the conventional Cherenkov effect. For now, they are acknowledged for a better interpretation of the RICH PID performance.

Performance metrics

The primary role of the RICH detectors is to distinguish between charged hadrons to allow for accurate analyses at LHCb. In particular, pions (π), kaons (K), and protons (p). The abundance of pions renders the K/π separation as the established stand metric in LHCb to assess the RICH PID performance. It is defined by the simultaneous measurement of the kaon efficiency (identifying a kaon as a kaon) and the pion misidentification (misID) rate (misidentifying a pion as a kaon). The efficiencies can be obtained by comparing the RICH

PID provided by the reconstruction algorithm to the MC information of a generated dataset from simulation.

As mentioned in Section 2.1.2, the working principle of the RICH reconstruction algorithm relies on maximising the overall likelihood of mass hypotheses. The K/π separation can be adjusted by the relative change to the logarithm of the likelihood (DLL), $\Delta\ln\mathscr{L}(\mathrm{K}-\pi)$, as a result of assigning a higher or lower preference of a kaon mass hypothesis with respect to the initial global pion hypothesis. This $\Delta\ln\mathscr{L}(\mathrm{K}-\pi)$ can be adjusted. A strict constraint results in a lower pion misID rate but also a lower kaon efficiency. A lower threshold has the opposite effect, a higher kaon efficiency but also a higher pion misID rate, as illustrated in Figure 2.1.

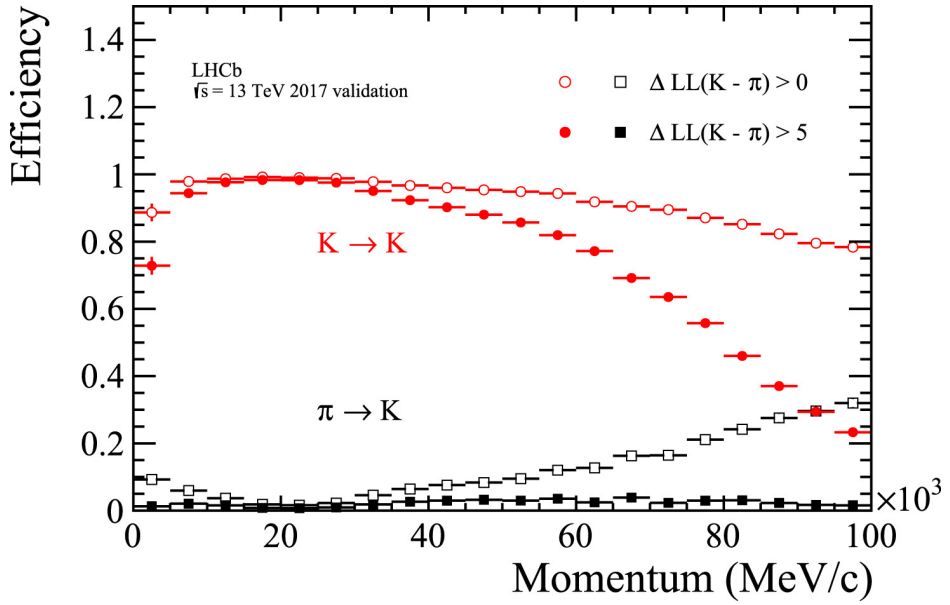


Fig. 2.1 Kaon ID efficiency and pion misID rate as a function of momentum for 2017 data. Results for a strict (DLL $>$ 5) and loose (DLL $>$ 0) K/π separation constraint are shown. Source: [60].

Figure 2.1 shows the performance for LHC Run 2 data for a momentum range from 0 to 100 GeV/c, combining the results from both RICH detectors. One can observe that the kaon efficiency is well above 90 % and pion misID below 10 % for momenta below 50 GeV/c. For momenta above 50 GeV/c, the performance quickly declines. The peak performance is reached for momenta around 20 GeV/c.

Another feature of the conventional RICH algorithm is the ability to identify kaons below the momentum threshold for generating Cherenkov photons. For kaons, this momentum threshold is at about 10 GeV/c. For these cases, the algorithm infers from the *absence* of

observed Cherenkov photons that the particle is likely a low energy kaon, allowing for a good K/π separation even at low energies.

2.2 The case for neural network RICH PID

The motivation for a neural network RICH PID is based on two reasons: Firstly, their excellent performance for a wide range of tasks, such as image processing and pattern recognition, and the continuous advancement of new techniques in this field. The underlying calculations of a neural network are most often more straightforward than analytic methods. Furthermore, they can be sensitive to hidden correlations that would have to be manually accounted for in an analytical model. Secondly, the change in hardware and technology is working in favour of computing parallelisation. Corresponding upgrades have already been proposed for LHCb.
In this thesis, I provide a proof of concept for a deep learning alternative to the conventional PID algorithm to motivate further research in this field.

This section motivates the need for new approaches to PID algorithms and the potential benefits of neural networks. It provides a brief overview of a selection of recent deep learning applications in the LHCb experiment.
In Section 2.1.2, the underlying principle of the conventional RICH reconstruction algorithm is explained, and the PID performance for Run 2 data is shown. While complex and computationally intensive, the algorithm performs exceptionally well, having been refined since its inception over 20 years ago. It was an essential component to all LHCb physics results and will continue its use during the upcoming LHC Run 3.

2.2.1 The advancement of deep learning in LHCb

Machine learning has introduced a multitude of new approaches to problem solving and pattern finding. One of the primary reasons for this development is the availability of much larger datasets and the steep hardware improvement to handle the data.

High-energy physics is the ideal place for deep learning applications. Data is produced and processed in enormous quantities. Precise physics simulations, as produced with GEANT4 and described in Section 2.1 are ideal candidates for this purpose. Their validity in replicating realistic physical processes has been rigorously tested [53, 54], allowing neural network training with both a large input dataset and matching labels provided by the MC information.

As a result, the application of deep learning techniques is quickly gaining traction in LHCb and other LHC experiments. One of the many examples is the implementation of neural networks for high-momentum track reconstruction in LHCb for the LHCb Run 2 trigger [61]. Or the use of neural networks for global particle identification in LHCb, which combines the likelihoods provided by the RICH detectors, the calorimeters, and the muon chambers [62]. Also, in offline physics analyses, the use of machine learning and deep learning is rapidly evolving. Recent examples include the two LHCb analyses [63, 64]. An overview of machine learning applications in LHCb is provided in [65].

2.2.2 Experimental conditions and hardware change

The environmental conditions of the LHCb experiment will become more challenging in the future. The expected HL-LHC luminosity in LHCb of 1.5×10^{34} cm^{-2}s^{-1}, will pose a significant challenge to LHCb in Run 5. In anticipation of these developments, changes to the data processing scheme in LHCb have been proposed [66, 67]. Figure 2.2 illustrates

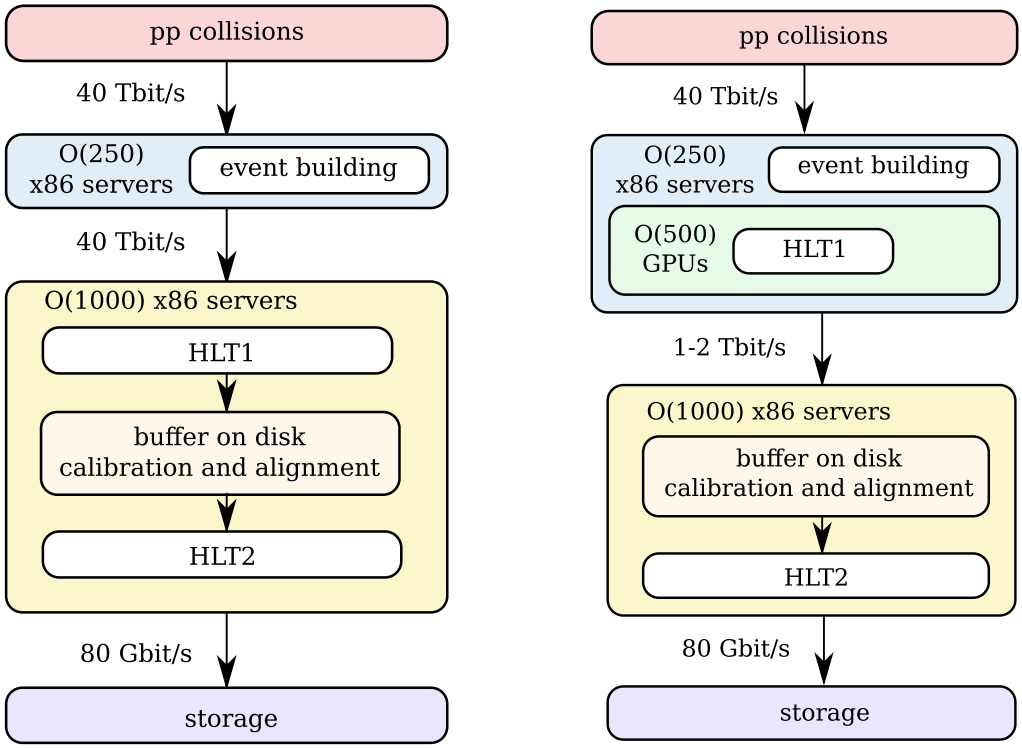


Fig. 2.2 Baseline proposal for the LHCb data acquisition system during Run 3 (a) and the GPU-enhanced proposal named project Allen (b). Here, the x86 event building servers host GPUs, processing HLT1, significantly reducing the data rate for HLT2. Source: [66].

these changes of the high-level trigger system, titled project Allen, after Frances E. Allen[1]. It proposes adding GPUs to the x86 servers, which process HLT1. Instead of buffering and processing all events, arriving at a rate of 40 Tbit/s, by both HLT1 and 2, only the events selected by the GPU supported event building servers are sent to the x86 servers, which are processing HLT2. This modification is expected to reduce the data rate between the two server farms by a factor 30 to 60.

The use of GPUs can be seen as a strong argument for the implementation of deep learning methods. CPUs (Central Processing Units) usually feature few cores, executing threads sequentially with a small number of threads. This favours complicated tasks optimised for low latency (memory access time). In contrast, GPUs (Graphics Processing Units) contain a larger number of cores, which enables many threads and computations in parallel, optimised for high throughput. That favours deep learning, which is based on a large number of simple calculations. Specialised libraries further facilitate the parallelisation of neural network computations on GPUs, and deep learning frameworks such as Caffe [68], TensorFlow [69], or PyTorch [70].

CPU- and GPU-based LHCb trigger systems have been directly compared in [45]. The deployment of GPUs has the potential to increase the throughput of the HLT filtering system while also promising further future gains in performance, anticipated by the swift technological advance of GPUs as general-purpose processors. That enables the possibility to implement a RICH PID neural network algorithm in the online reconstruction of HLT 1. A hybrid solution of both PID approaches could be implemented if a NN RICH PID exceeds the conventional algorithm in computing efficiency or performance only for specific momentum ranges or subsections of the RICH detector panels.
An example for such a hybrid solution would be using NN RICH PID for filtering in the online triggering of the HLT, where computing efficiency is the priority, together with the conventional algorithm for the offline reconstruction, where the PID performance is decisive.

The field of deep learning is rapidly advancing, providing new techniques and frameworks with increasing speed. Developments in this direction in terms of both software and hardware are necessary to prepare LHCb for the challenges of the high-luminosity environment. This study aims to demonstrate the feasibility of a neural network approach and develop a methodology to allow direct comparison of different algorithms. To this end, a novel convolutional neural network algorithm for RICH PID is developed based on well-established principles.

[1] American computer scientist pioneer and first woman to win the Turing Award.

2.3 Machine learning in a nutshell

Machine learning is sometimes described as "the process of computers changing the way they carry out tasks by learning from new data, without a human being needed to give instructions in the form of a program" [71].
The concept of an autonomous machine dates back a long time but became only practically realisable in the last few decades with the steep increase in computing power and data availability. The concept of a general artificial intelligence that can learn and perform any reasoning task a human can do (or even outperform the human) is not within reach at the time of writing this thesis. However, computer programs are capable of exceeding human performance in specific tasks. Prominent examples are the chess tournament defeat of the reigning champion Garry Kasparov against the IBM machine Deep Blue in 1997 [72]. Or the victory of DeepMind's AlphaGo against Lee Sedol in a five-game match of Go in 2016 [73], which after several iterations evolved into MuZero, with the capability to succeed in other games, such as chess and shogi, without previous knowledge of the rules [74]. These advancements raise the question which other tasks can be improved by applying machine learning, by various performance measures, such as accuracy or speed.

Machine learning can be understood as artificial intelligence, applied to a task with a limited scope, such as recognising an object from an image or targeted advertisement.
In the following, the concepts of machine learning most relevant to this thesis are outlined. This introduction, including sections 2.3.1 to 2.3.7, broadly follows [75] and [76] and is credited to these two references.

Let us define two vector spaces $\mathbb{X}$ and $\mathbb{Y}$, where both can be real vector spaces or finite sets. Let us call $\mathbb{X}$ input or feature space and $\mathbb{Y}$ label or target space. We call a pair of elements from the feature space and label space $(\boldsymbol{x}, \boldsymbol{y})$, where $\boldsymbol{x} \in \mathbb{X}$ and $\boldsymbol{y} \in \mathbb{Y}$, an example or a data point. We call the space of examples the training dataset.

2.3.1 Model training

The goal of machine learning is to train a model to represent a function f that takes an element $\boldsymbol{x}$ of the feature space and maps it onto the corresponding element $\boldsymbol{y}$ of the target space. In the case of parameterised supervised learning, first, a model has to be determined. The model can be expressed by a function f from a function space $\mathbb{F}$ which is a subset of $\mathbb{Y}$, $f \in \mathbb{F} \subseteq \mathbb{Y}$. To facilitate the task of finding a suitable function, the computer is lead to search for the parameters of the model instead of the function itself. That is possible by defining a parameter-function map $\mathbb{M}$. For a model with p parameters, one can find a set $\boldsymbol{\Theta} \subseteq \mathbb{R}^p$, such

that:

$$\begin{aligned} \mathbb{M} : \Theta &\to \mathbb{F} \\ \boldsymbol{\theta} &\mapsto f(\boldsymbol{\theta}), \end{aligned} \tag{2.6}$$

where $f(\boldsymbol{\theta})$ is the function representing the model with a choice of parameters $\boldsymbol{\theta}$.

One can broadly divide machine learning models into two categories: regression and classification. Other types of machine learning tasks are omitted for reasons of brevity, such as transcription or anomaly detection. Also, only supervised learning methods are discussed, meaning that for each element $\boldsymbol{x} \in \mathbb{X}$ a corresponding label $\boldsymbol{y} \in \mathbb{Y}$ exists. There are techniques to deal with unlabelled or incompletely labelled datasets, such as unsupervised or reinforcement learning. Since the datasets used in this thesis are all labelled, they will not be discussed in more detail.

Regression models are used for predicting a continuous quantity. The input variable can be real-valued or discrete. If there are multiple input variables, one refers to a multivariate regression problem.
Classification models are used to assign inputs to discrete output categories, such as for object recognition. Similar to regression problems, the input can have real-valued or discrete variables. If there are more than two output categories, the model type is called multi-class classification.

2.3.2 Performance metrics, loss functions, and Occam's razor

To determine how well $\boldsymbol{y}_{pred} \coloneqq f(\boldsymbol{\theta}, \boldsymbol{x})$ agrees with $\boldsymbol{y}_{true}$ for a set of examples, one has to define a performance metric $P(\boldsymbol{y}_{pred}, \boldsymbol{y}_{true})$. These metrics can either be used by a model optimiser, to improve the model, or to assess the model quality. A few commonly used metrics are introduced below.

Mean Squared Error (MSE)

Due to its mathematical simplicity and quadratic sensitivity to outliers, the MSE is often used to evaluate regression algorithms when data outliers are rare.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2, \tag{2.7}$$

where N is the number of examples (x_i, y_i) and $\hat{y}_i$ is the prediction made by the model.

Accuracy

Accuracy is frequently used in classification problems, where the labels are discrete values, such as numbers 0 to 9, or particle species.

$$\begin{aligned} Accuracy &= \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \\ &= \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i = y_i). \end{aligned} \tag{2.8}$$

The accuracy of a model is easy to interpret but subject to drawbacks. Firstly, for problems with multiple classes and an uneven distribution between the classes, the model accuracy is higher if the model predictions are biased towards the more common classes. This case is relevant for the thesis subject since some particle species are much more prevalent than others (i.e. pions). Secondly, classification is discrete, e.g. inferring 3 from 2.9 and 3.4, without being sensitive to their difference.

Logarithmic Loss (LogLoss)

Logarithmic loss is often used for multi-class classification problems, penalising wrong classifications. For N samples and M classes, the LogLoss is calculated as:

$$LogLoss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} y_{ij} * \log p_{ij}, \tag{2.9}$$

where y_{ij} specifies whether the sample i is part of class j and p_{ij} is the corresponding probability computed by the model. Since $L_{LogLoss} \in [0,\infty)$ and values closer to 0 signify higher accuracy, minimising $L_{LogLoss}$ increases the accuracy of the model. For machine learning problems, LogLoss is often conversely termed cross-entropy.

Over- and underfitting

To adequately approximate $\boldsymbol{y}_{true}$ one has to find the parameters $\boldsymbol{\theta}$ which minimise the error as defined by the loss function:

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \left[L(f(\boldsymbol{\theta}, \boldsymbol{x}), \boldsymbol{y})\right] \tag{2.10}$$

The goal is to find parameters $\boldsymbol{\theta}$, such that $f(\boldsymbol{\theta}, x_i) \approx y_i$ for example tuples (x_i, y_i) which were not included in the training dataset. This set of previously unseen data points is called *test dataset* and the ability to reliably perform on the test dataset *generalisation*.

Poor performance of the model is likely caused by either over- or underfitting. It is often caused by the complexity or capacity of the model not being adequate for a given set of data points to approximate. The capacity of a model can be understood as its degrees of freedom. For example, a polynomial of order 0 or 1 is unable to fit data points evaluating a polynomial of order 3, as illustrated in Figure 2.3 (top left and right). The result is underfitting, leading to poor performance of the model on both the training and the test dataset. An example of overfitting of a quadratic function is shown in Figure 2.3 (bottom right). A too high capacity of the model allows for "memorisation" input-output tuples from the training dataset rather than finding a general way of linking $\boldsymbol{x}$ to $\boldsymbol{y}$. Such a memorisation effect is often encountered when the model has more parameters than data points to train on. This raises the question, what is the optimal number of degrees of freedom of an algorithm. The answer is provided by statistical learning theory but also most famously by the principle of *Occam's razor* [77]. In broad terms, it states that the simplest model (i.e. with the fewest degrees of freedom) is favoured when multiple hypothetical models are available if each provides an equally suitable solution to a given problem. In the practical example, illustrated in Figure 2.3, the polynomial degree 3 (bottom left) is, therefore, best to generalise from the set of data points to unseen examples.

2.3.3 No free lunch and regularisation

That a machine learning algorithm is not able to perfectly generalise from a finite set of data points is intuitively logically sensible. The *no free lunch theorem* goes one step further, showing that all classification models have a similar error rate for unseen examples when averaging over all possible tasks [78]. That problem can be solved by defining a narrow space of applications for a classification algorithm. The set of functions it can choose from has to be limited to obtain such a specialised algorithm. This can be achieved by both limiting the number as well as the kind of available functions. As an example, one can limit the type of functions to linear functions and choose from those a suitable subset, e.g. setting a threshold for the degree of the polynomial. A practical approach for this limitation is achieved by the *regularisation* of a model.

To understand regularisation one can take the example of linear regression. The goal is to find a set of parameters $\boldsymbol{w}$, called weights, and the interception parameter b, such that the

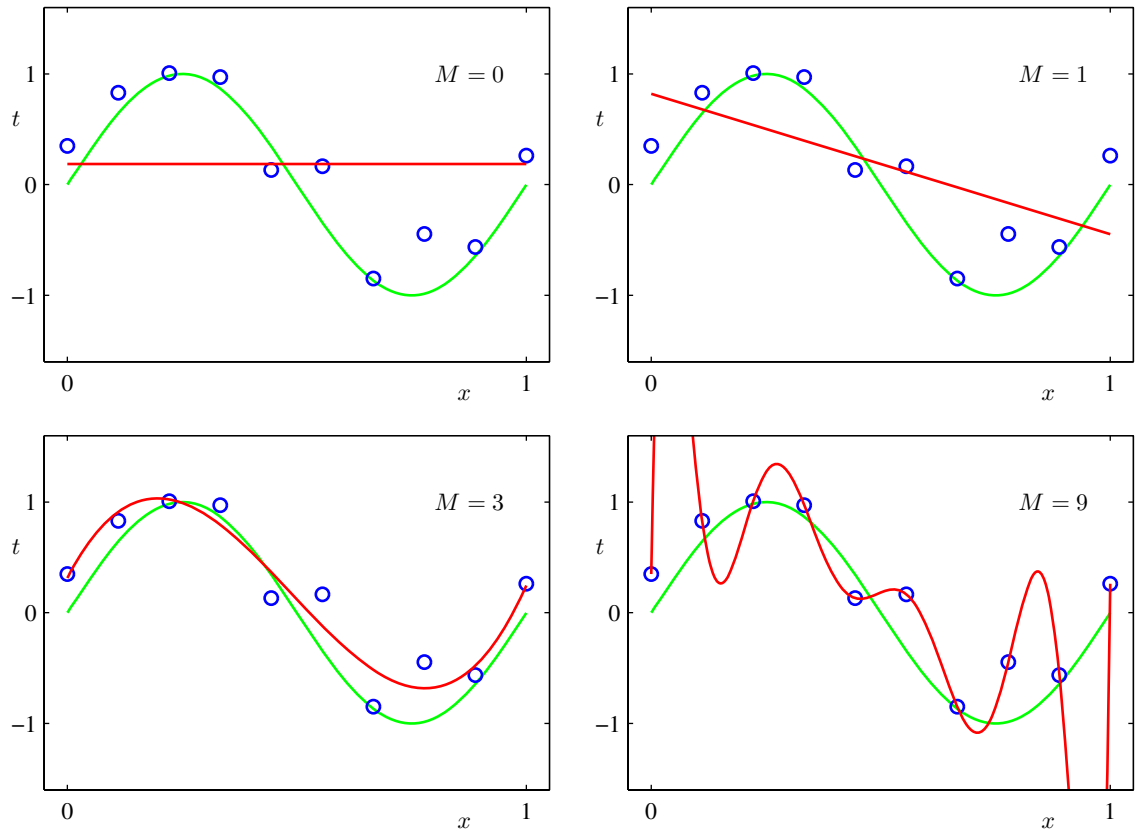


Fig. 2.3 Illustration of model fitting errors for a set of data points belonging to a polynomial of order 3. (Top left and right) models with too few parameters, in this case representing a linear function, unable to fit the data points and generalise. This is called underfitting. (Bottom right) a polynomial with too many degrees of freedom can "memorise" data points. Despite performing well on the training dataset, it does not replicate well the "true" function caused by overfitting. Both over- and underfitting are unable to generalise from training data to test data. (Bottom left) an example of a well-fitted model for the quadratic function. Source: [76].

function $\hat{\boldsymbol{y}} = b + \sum_{i=1}^{m} w_i x^i$ optimally approximates a set of data points $(x,y) \in (\boldsymbol{x},\boldsymbol{y})$. That can be achieved by minimising the MSE, as demonstrated in Equation 2.7.

As described in Section 2.3.2, the fewest required degrees of freedom to generalise from the training data are optimal. In most cases, it is difficult to identify such a suitable subset of hypothetically available functions. Regularisation allows modifying the training, biasing towards simpler functions by punishing large weights. That can be achieved by adding the product of the weights $\boldsymbol{w}$ to the data fidelity term, e.g. the MSE. The regularisation can be scaled by a factor λ:

$$J(\boldsymbol{w}) := MSE + \lambda \boldsymbol{w}^T \boldsymbol{w}. \tag{2.11}$$

There are multiple ways to regularise. The most suitable one has to be chosen individually for a given task. However, in general, regularisation denotes the attempt to minimise the generalisation error without reducing the training error. As illustrated in Figure 2.4, regularisation is an essential tool when choosing the capacity of a model. A too high capacity, if regularised, can still represent a satisfactory model. Commonly used regularisation

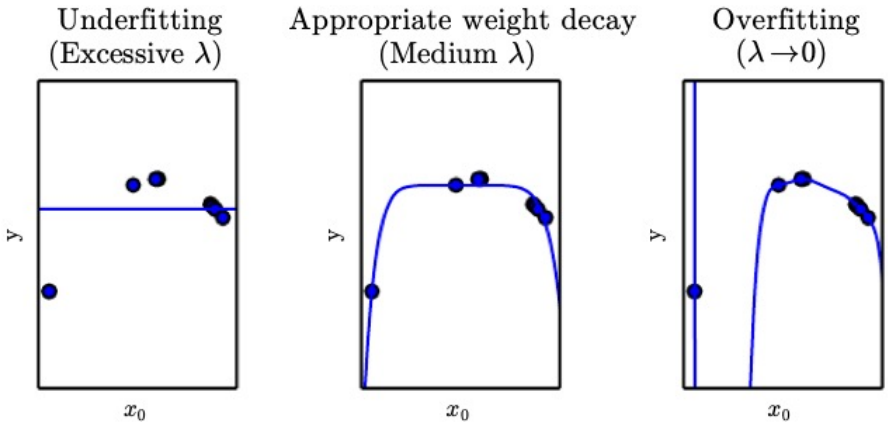


Fig. 2.4 For most problems, the ideal (i.e. smallest) capacity of a model to adequately map unseen data points is unknown. In practice, one chooses a model with a too high capacity and uses regularisation (e.g. Equation 2.11) to prevent over- or underfitting. In this example, a polynomial with degree 9 can model data points sampled from a quadratic function when an appropriate value for λ is chosen (middle). The regularisation penalises more complicated shapes. A too-large value for λ results in underfitting (left). Correspondingly, if λ is set too small, overfitting is the consequence. Source: [75].

estimators terms are the Lasso (least absolute shrinkage and selection operator) ($\lambda \|w\|_1$) and the ridge ($\lambda \|w\|_2^2$). They are l_1 and l_2 norms, respectively, scaled by a parameter λ. The Lasso estimator results in sparse models, zeroing out many of the model's parameters, due to the l_1 norm's sharp edges, as depicted in Figure 2.5.

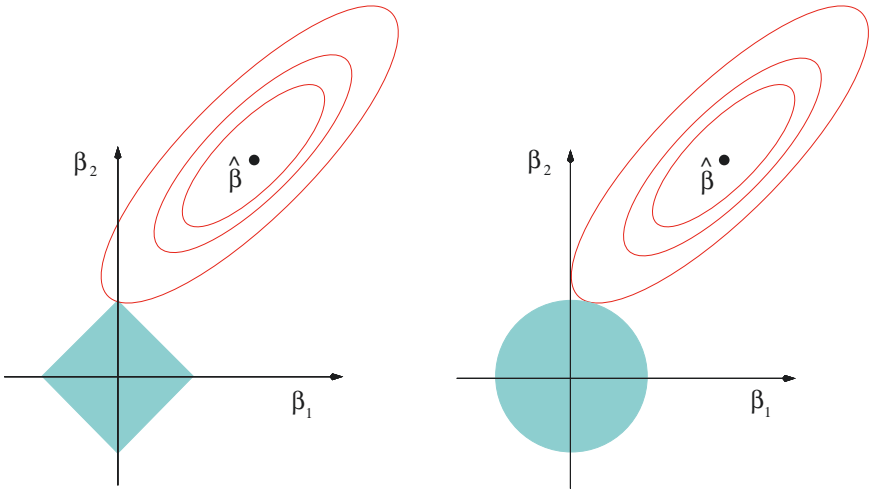


Fig. 2.5 Estimator contours for the Lasso (left) and ridge (right) regression. The blue areas represent the constraint regions, the red ellipses illustrate the contours of the MSE. Source: [79].

Other forms of regularisation include dropout layers, batch normalisation, and early stopping [79]. The latter is explained in Section 2.3.5. *Dropout* is a method that regularises models by ignoring (dropping) a certain percentage of randomly selected node outputs during training. *Batch normalisation* is a technique that scales the activations of a layer to have a mean of zero and standard deviation of one (e.g. a common Gaussian). This process is also called standardisation or whitening. The mean and variance standardisation stabilises the training while allowing the relationships between layers to change [75].

2.3.4 Hyperparameters and cross-validation

The term hyperparameter denotes a parameter not learned by the model but rather determined independently of the training. The regularisation parameter λ is an example of such a hyperparameter. To find suitable hyperparameters and estimate the power of a model to generalise from training to test data, one can split the training dataset into two disjunct sets of data points. The subset of data points for estimation of the generalisation ability is referred to as *validation* dataset. Typically, the validation dataset is small compared to the training dataset ($\approx 10-20$ %) since the amount of training data usually positively correlates with model performance.

Cross-validation is used to reduce the adverse effect on the model performance by reducing the training dataset and to avoid a large statistical uncertainty of performance estimation by a

small validation or test dataset. It describes the method to repeatedly train and test the model on randomly chosen disjunct subsets of the original dataset. That procedure is beneficial when relying on small sets of data points.

2.3.5 Deep learning

Deep learning (DL) is a field of machine learning using neural networks. In contrast to other areas of ML, which are motivated by rather by mathematical logic (i.e. decision trees), neural networks consist of several layers (deep neural networks or DNN), each containing several simple numerical calculations, so-called neurons. The name is inspired by a faint resemblance to biological neural networks. Mathematically, each layer can be represented by a function, e.g. $f^{(1)}$ for the first layer, $f^{(2)}$ for the second, and so on. A model with three layers could be written as $f(\boldsymbol{x}) = f^{(3)}(f^{(2)}(f^{(1)}(\boldsymbol{x})))$. The last layer, $f^{(3)}$ in this example, is called the output layer. The other layers are called hidden layers.

Typically, the large number of parameters in deep learning models requires both extensive training datasets and equally high computing power. While established already many decades ago, the fast technological advancements since the turn of the millennium have catalysed the development of DL research.

One application of a DNN is to approximate nonlinear functions of an input $\boldsymbol{x}$. The conventional approach is to manually find a nonlinear transformation $\phi(\boldsymbol{x})$. That allows to solve the problem with a linear model, applied on $\phi(\boldsymbol{x})$, rather than the original input $\boldsymbol{x}$. The manual process of finding a suitable transformation is called *feature engineering*. It often requires an expert understanding of the problem and time investment. The resulting problem is much easier to solve by the algorithm. An example for such a case is shown in Figure 2.6. However, it is not as flexible as before since the space of available functions was effectively limited by the transformation. An alternative approach is to let the DNN learn $\phi(\boldsymbol{x})$ by itself, allowing it to choose from the space of available nonlinear transformations. So instead of directly finding an appropriate solution f to the nonlinear problem, the algorithm finds ϕ, such that $\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{w}) = \phi(\boldsymbol{x}, \boldsymbol{\theta})^T \boldsymbol{w}$. Practically speaking, ϕ defines a hidden layer of the DNN and the parameters $\boldsymbol{w}$ map to the desired output $\boldsymbol{y}$.

To achieve the optimal compromise between the two concepts, feature engineering can also be included in defining a generic family of transformation functions ϕ to encode their understanding of the problem while still giving the algorithm flexibility in finding a suitable transformation. This method is applied to the particle finding algorithm, described in Section 2.4.2.

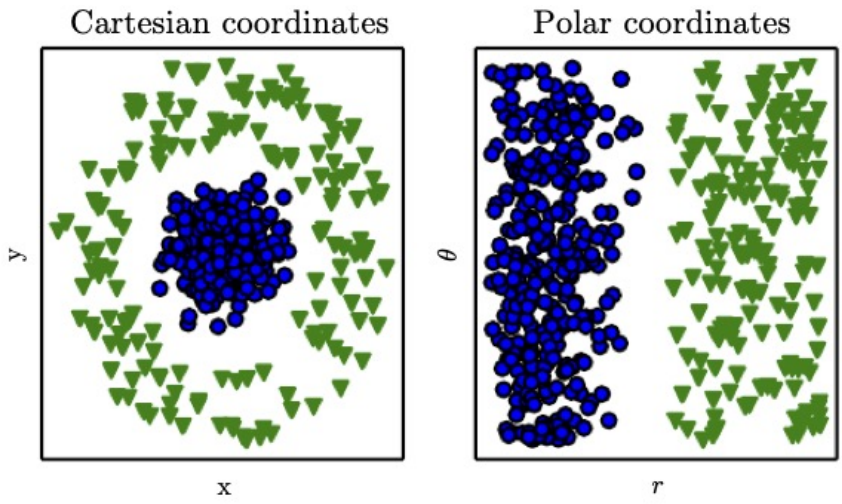


Fig. 2.6 With feature engineering, complicated problems can be translated into much simpler ones. The example shows how a polar transformation makes it possible to separate the two types of data, represented by circular and triangular shapes, with a linear function. Source: [75].

Neural networks and activation functions

A neural network is characterised as a series of functional transformations. Each layer corresponds to a linear combination of its input variables, called activation a. The activation of the first layer (denoted by the superscript (1)) of a network with input $x_1,\ldots,x_D$ would be

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + b_{j0}^{(1)}, \text{ where } j = 1,\ldots,M. \tag{2.12}$$

$w_{ji}^{(1)}$ are the weights of the first layer and $b_{j0}^{(1)}$ its biases, and M the total number of outputs. The activations are transformed by a nonlinear activation function $h(\cdot)$, to get the output $z_j = h(a_j)$. Each of these linear combinations of input variables to generate an output is called neuron. The first neuron of the first layer of a neural network is illustrated in Figure 2.7. Consequently, the second layer of the network takes the output of the first layer as input:

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)} z_i + b_{j0}^{(2)}, \text{ where } k = 1,\ldots,K. \tag{2.13}$$

To provide a compact mathematical representation of multiple layers of a network, the bias parameter is often absorbed into a weight vector $\boldsymbol{w}$, by adding an additional input variable x_0.

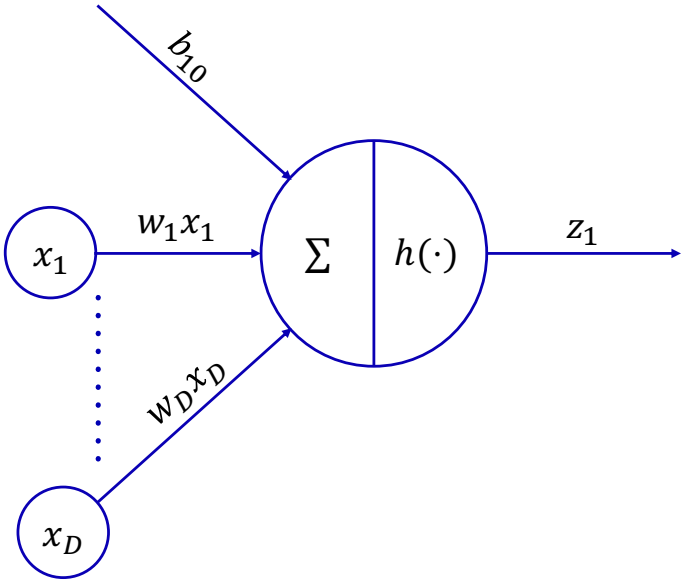


Fig. 2.7 The linear combinations of input variables and activation are called neurons. Here, the first neuron of the first layer of a network is schematically illustrated. The input variables $x_1, \ldots, x_D$ are multiplied by weights $w_{ji}^{(1)}$ and added up with a bias $b_{j0}^{(1)}$. An activation function $h(\cdot)$ is applied to the sum, to get the output z_1 of the neuron. The outputs $z_1, \ldots, z_M$ are the inputs to the next layer of the network.

Hence the two-layer network can be depicted as

$$y_k(\boldsymbol{x}, \boldsymbol{w}) = h_2 \left(\sum_{j=0}^{M} w_{kj}^{(2)} h_1 \left(\sum_{i=0}^{D} w_{ji}^{(1)} \right) \right) \tag{2.14}$$

The stacking of neural network layers can be continued, and the number of connections, output variables, activation functions, etc., adapted to achieve the optimal model for the problem at hand. This process is called the design of the network and usually requires human expert knowledge of the problem. A schematic of a neural network with two layers is illustrated in Figure 2.8.

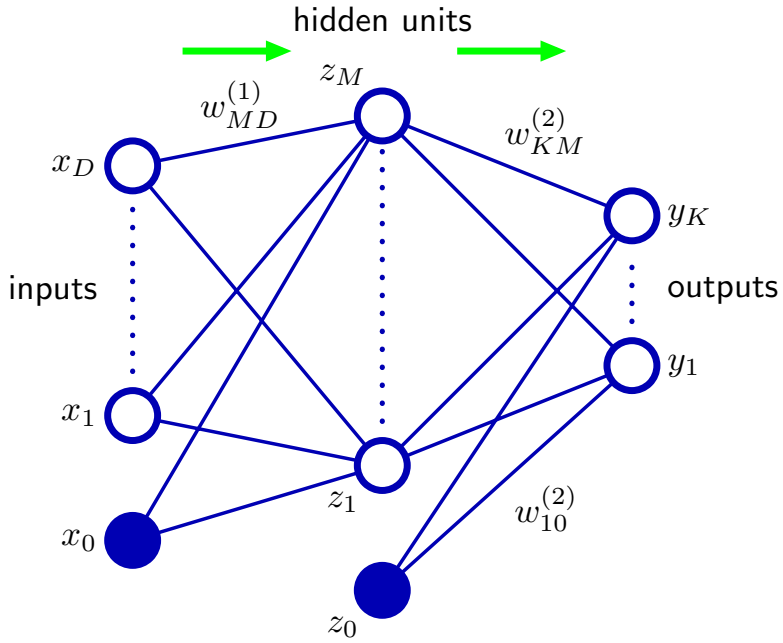


Fig. 2.8 Schematic of a neural network with two layers. Source: [76].

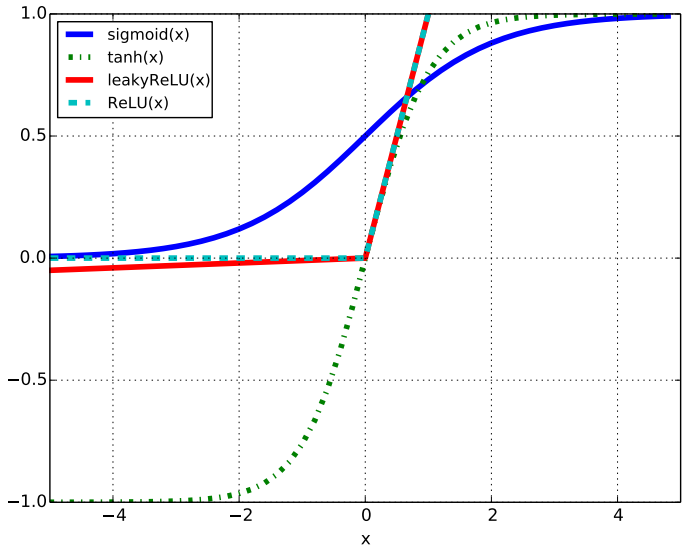


Fig. 2.9 Selection of four commonly used activation functions: $\sigma(x) = \frac{1}{1+exp(-x)}$, $\tanh(x) = 2\sigma(x) - 1$, $\mathrm{ReLU}(x) = \max(0,x)$, $\mathrm{leakyReLU}(x) = \max(0,x) + a\min(0,x)$. The choice of activation function is dependent on the type of problem and also to where in the model the activation function is used.

The choice of activation function depends on the problem type and the layer's position in the model. Common choices of activation functions are often sigmoidal, variations of the logistic sigmoid function, defined as

$$\sigma(x) = \frac{1}{1+exp(-x)}, \tag{2.15}$$

or the hyperbolic tangent function, which relates to the sigmoid function by $\tanh(x) = 2\sigma(x) - 1$. The logistic sigmoid function is most often used in the last layer of a binary classification network due to its properties to map all real inputs to a finite interval $(0,1)$, which can also be interpreted as a probability.

Activation functions are still subject to research, and no best solution can be universally defined. The most popular activation function for feed-forward networks is rectified linear units (ReLU) due to their superior model training capabilities [80]. ReLU combines two linear functions, which facilitates optimisation. Yet, it can account for nonlinear effects due to its nonlinear nature:

$$\text{ReLU}(x) = max(0,x). \tag{2.16}$$

A modification of the classical ReLU function is the so-called *Leaky* ReLU:

$$\text{leakyReLU}(x) = \max(0,x) + a\min(0,x). \tag{2.17}$$

It is the same for positive input values but has a constant negative slope for values below 0 which can be adjusted by a parameter a. Its benefit is that it is influenced by all input values x, not only the positive ones as for the standard ReLU function. Figure 2.9 shows a selection of activation functions.

Another type of activation is the *softmax* function [76]. It is often used in the last layer of a multi-label classification model. It can be interpreted as a generalisation of the logistic sigmoid function σ. It normalises the output of the network to a probability distribution for several output classes. For K classes and k being a member of these classes, the posterior probability of a class C_k given an input $\boldsymbol{x}$, is defined as:

$$p(C_k|\boldsymbol{x}) = \frac{\exp(a_k)}{\sum_j \exp(a_j)}, \tag{2.18}$$

where $a = \ln\left(\frac{\sigma}{1-\sigma}\right)$ represent the inverse of the logistic sigmoid. Or in other words, the logistic ratio of the posterior probability for a class $p(\boldsymbol{x}, C_k)$ and its prior $p(C_k)$:

$$a_k = \ln p(\boldsymbol{x}, C_k) p(C_k). \tag{2.19}$$

Gradient learning and early stopping

The nonlinearity of neural network (NN) activation functions means that one cannot use linear equation solvers or optimisation algorithms to train a NN. Rather, one has to use gradient-based iterative optimisers. The goal is to find the parameters $\boldsymbol{\theta}$ that minimise the cost or loss function $J(\boldsymbol{\theta})$.

The general idea of gradient-based optimisation is to follow the direction of the negative gradient ∇J, i.e. the steepest descend of a cost or loss function J. The size of iterative steps k towards a (local) minimum is determined by the learning rate parameter η, such that

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \eta \cdot \nabla J(\boldsymbol{\theta}^{(k)}). \tag{2.20}$$

The starting value $\boldsymbol{\theta}^{(0)}$ has to be manually initialised.

The learning rate η has to be chosen such that the algorithm converges. Choosing a too-large value for η will lead to divergence, while a too-small value slows convergence towards a minimum.

While mathematically appealing, the gradient-based optimisation of Equation 2.20 is computationally expensive due to the necessity to iterate over each example in a given dataset. The practical solution is to approximate the gradient using a randomly selected data subset. The result is a much faster optimisation in trade for a higher probability of divergence. This approach is called stochastic gradient descent (SGD) [81].

Variations of SGD are the most prevalent type of optimisation algorithms used for deep learning. Popular variations of SGD are *Adagrad* and *Adadelta*. During the optimisation, the learning rate is updated in each iteration, usually decreasing with each step. For Adagrad, the learning rate approaches zero comparatively quickly. Adadelta aims to improve on that with exponentially weighted averages of the gradient. One of the most popular optimisers, expanding on this concept, is called *Adam* (adaptive moment estimation) [82]. This algorithm combines several successful extensions of SGD. One of these extensions consists of adding a "momentum" [83] term to Equation 2.20: $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \boldsymbol{v}$. The momentum term $\boldsymbol{v}$ is

defined as

$$\boldsymbol{v}^{(k+1)} = \alpha \boldsymbol{v}^{(k)} - \eta \nabla_{\boldsymbol{\theta}} \left(\frac{1}{m} \sum_{i=1}^{m} J\left(\boldsymbol{f}\left(\boldsymbol{x}^{(i)}, \boldsymbol{\theta} \right), \boldsymbol{y}^{(i)} \right) \right). \tag{2.21}$$

The hyperparameter $\alpha \in [0,1)$ determines the contribution of the previous momenta. The more significant the relative difference between α and the learning rate η, the more previous momenta contribute. Momentum can be understood in this context as an analogy to the physics case. Here, unit "mass" is assumed, which sets momentum in this analogy equal to velocity. The Adam optimiser utilises momentum and an adaptive learning rate to converge faster. The Adam update rule can be summarised as follows:

$$\begin{aligned}
\boldsymbol{\theta}^{(k+1)} &= \boldsymbol{\theta}^{(k)} - \frac{\eta \cdot \hat{m}^{(k)}}{\sqrt{\hat{v}^{(k)}} + \varepsilon}, \text{ where} \\
\hat{m}^{(k)} &= \frac{m^{(k)}}{1 - \beta_1^{(k)}} \\
\hat{v}^{(k)} &= \frac{v^{(k)}}{1 - \beta_2^{(k)}} \\
m^{(k)} &= (1 - \beta_1) \cdot \nabla J(\boldsymbol{\theta}^{(k)}) + \beta_1 m^{(k-1)} \\
v^{(k)} &= (1 - \beta_2) \cdot \left(\nabla J(\boldsymbol{\theta}^{(k)}) \right)^2 + \beta_2 v^{(k-1)}.
\end{aligned} \tag{2.22}$$

ε is a small term to prevent division by zero. The original publication of the Adam optimiser suggests default values for $\varepsilon = 10^{-8}$, $\eta = 10^{-3}$, the first momentum term $\beta_1 = 0.9$, and the second momentum term $\beta_2 = 0.999$.

A sample of examples is randomly selected to estimate the gradient to optimise a neural network. Passing the entire dataset through the network once is called an epoch. Usually, multiple epochs are used for model optimisation. Between each epoch, the subset of examples is again randomly selected. The number of selected training samples is called the batch size. The total number of iterations is defined by how many batches are required to complete an epoch. Choosing a feasible number of epochs is part of the hyperparameter tuning of a neural network since there is no analytical way to determine its optimum. Too many epochs lead to overfitting, too few to underfitting.

To avoid both over- and underfitting, one can choose a high number of epochs and implement a rule for the network to stop the optimisation process. This procedure is called *early stopping*. The performance of the network is measured using a performance metric, such as accuracy (Equation 2.8), for the validation dataset. At the minimum validation error, the

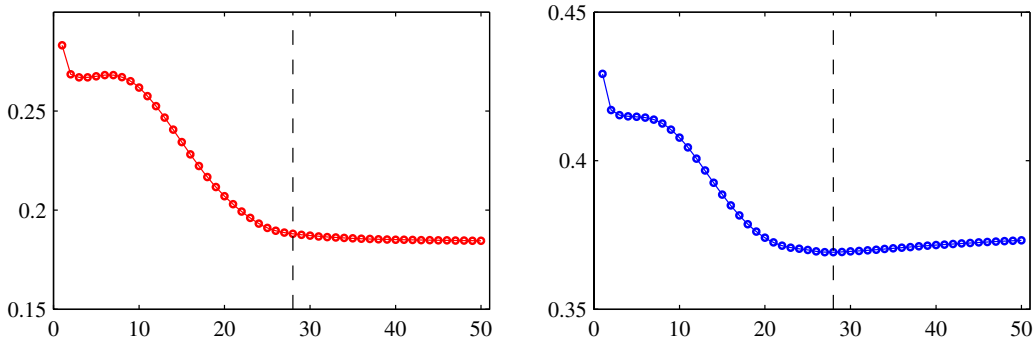


Fig. 2.10 Model error measured on the training set (left) and validation set (right). The dashed vertical line marks the point where the model has reached its optimal generalisation potential. Here, the validation error is the smallest. To the right of the line, the validation error increases again, while the training error continues to decrease due to overfitting. Source: [76].

neural network model can best generalise and perform well on unseen examples. If the network training continues, the training error still decreases, but only due to overfitting. This can be seen by an increase of the validation error, as shown in Figure 2.10.

2.3.6 Error backpropagation

To train a neural network, information is repeatedly transmitted forwards and backwards. The forward-feeding of information has already been explained and illustrated in Figure 2.7. The backward propagation (or backpropagation) of information can be best understood by separating the neural network function into its layers:

$$f(\boldsymbol{x},\boldsymbol{\theta}) = f^{(k)}(f^{(k-1)}(\ldots f^{(1)}(\boldsymbol{x},\boldsymbol{\theta}^{(1)}),\ldots,\boldsymbol{\theta}^{(k-1)}),\boldsymbol{\theta}^{(k)}), \tag{2.23}$$

where $\boldsymbol{\theta}^{(i)}$ denote the parameters of the i-th layer. One can compute the gradient of the loss function $J(f,y)$ using the chain rule to obtain the derivative of the loss function with respect to the weights of a given layer:

$$\frac{\partial J_n}{\partial w_{ji}} = \frac{\partial J_n}{\partial a_j}\frac{\partial a_j}{\partial w_{ji}} \tag{2.24}$$

To simplify the notation, we introduce

$$\delta_j := \frac{\partial J_n}{\partial a_j}, \tag{2.25}$$

where J_n represents an element of the overall loss for a given batch of the training set, computed as $J = \sum_{n=1}^{N} J_n$, and a_j the weighted sum of its inputs, a generalisation of Equation 2.13:

$$a_j = \sum_i w_{ij} z_i, \tag{2.26}$$

where $z_j = h(a_j)$ is the activation of unit j given by the activation function $h(\cdot)$. δ_k denotes the partial derivative of the loss with respect to the output units. The δ's for all units of the network can be computed using the chain rule:

$$\begin{aligned} \delta_j &= \sum_k \frac{\partial J_n}{\partial a_k} \frac{\partial a_k}{\partial a_j} \\ &= h'(a_j) \sum_k w_{kj} \delta_k \end{aligned} \tag{2.27}$$

where $h'(\cdot)$ is the derivative of the activation function. From Equation 2.26 one obtains

$$\frac{\partial a_j}{\partial w_{ji}} = z_i. \tag{2.28}$$

Combining Equations 2.27, 2.28, the required derivatives can be evaluated:

$$\frac{\partial J_n}{\partial w_{ji}} = \delta_j z_i. \tag{2.29}$$

The total loss J can be obtained by summing over the individual elements, e.g. combining the loss from several batches:

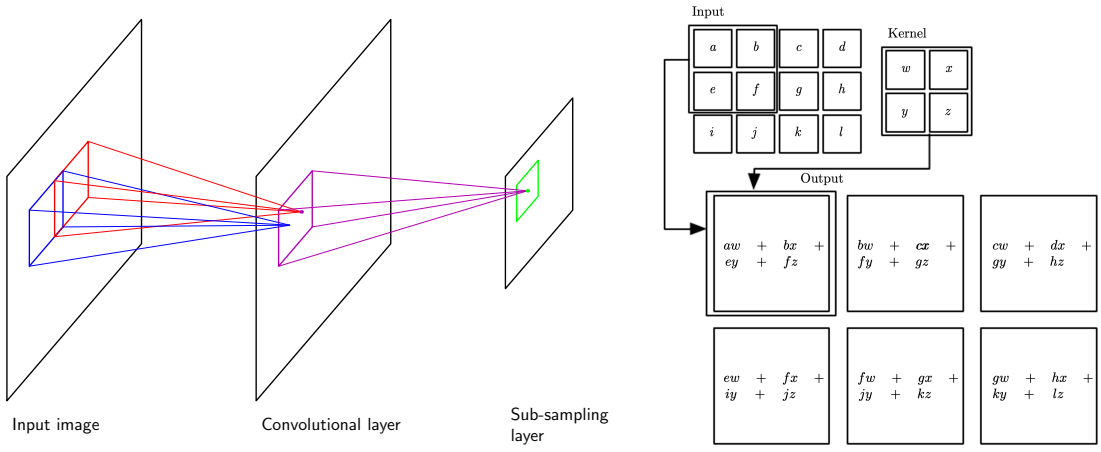
$$\frac{\partial J}{\partial w_{ji}} = \sum_n \frac{\partial J_n}{\partial w_{ji}}. \tag{2.30}$$

2.3.7 Convolutional neural networks

An image can be represented as a set of pixel intensity values. For black and white images, pixels either have the value 0 or 1, for greyscale images a value between 0 and 1, and for colour-images a value between 0 and 1 for the three components red, green, and blue. To analyse images using a neural network, one can take the approach of a simple, fully connected network of multiple layers, as described above. That would, however, neglect the correlation of neighbouring pixels. Pixels close to each other usually correlate more strongly than more distant ones.

Convolutional neural networks (CNN) take the correlation of neighbouring pixels into account, using a grid-like topology for a particular type of linear operation, called convolution,

in at least one of the network layers. The grids or outputs of a convolutional layer are called feature maps. They receive input only from a subregion of the image or previous plane. Feature maps use the same weight values to achieve an approximate invariance against translations or distortions of image features. A schematic of a CNN is illustrated in Figure 2.11a and an example of a CNN computation is shown in Figure 2.11b.



(a) (b)

Fig. 2.11 (a) Diagram of CNN with a convolutional layer and subsampling units. (b) Example of a convolutional layer operation. Source: [75, 76].

Pooling layers reduce the size of an input plane while amplifying its most essential features, such as edges. For example, a 2×2 sub-sampling pooling layer can take the average of all four values or their maximum, thereby reducing the output to a single number and halving the image size. That reduces the computational burden of the following layer and is essential for the final layers of a network to summarise the last layer to a small number of classes or even a single number.

Convolutional neural networks have proven to be successful in image recognition and other two-dimensional grid-structured tasks. From their first application in automating check reading in 1998 [84], to their first success in the ImageNet object recognition competition [85] (followed by many more successful competitions in the subsequent years).

The grid structure of the RICH detector planes used for particle identification is a strong motivation for using convolutional neural networks to approach this problem. The feasibility of other neural network types may be determined by future studies and are not covered in this thesis.

2.4 A convolutional neural network for RICH PID

In this section, the neural network for particle identification for LHCb RICH is introduced. The framework is explained, starting from data generation in Section 2.4.1, to feature engineering in Section 2.4.2, to the structure of a proposed neural network in Section 2.4.3. This work extends the existing LHCb reconstruction algorithms to generate input data for the network, following a suggestion from Campora Pérez and Jones [86].

2.4.1 The input data

High-energy physics simulations are ideal for building and training neural networks due to their empirically tested agreement with real-world data and the additional Monte Carlo truth information. The studies presented here use simulated data.

To generate data for the neural network training and performance assessment, the LHCb simulation and digitisation infrastructure, described in Section 2.1 is used. In particular, GAUSS v54r3 with B_s^0 and $\bar{B}_s^0$ signal events for simulation and BOOLE v41r0 for digitisation. The database and simulation are configured to model the LHCb Run 3 experiment with RICH MaPMT photon detectors[2].

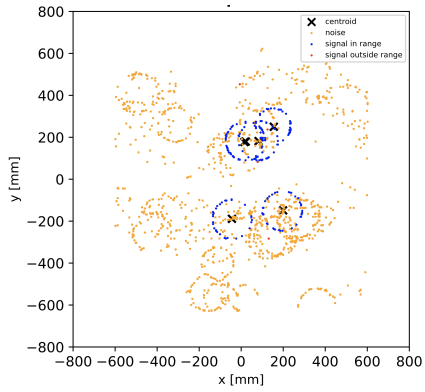
The reconstruction framework BRUNEL v61r0 is used to obtain information such as the particle trajectory from tracking and the particle momentum. The reconstruction of the dataset also includes the conventional RICH PID algorithm, which allows for a direct comparison between the two approaches on the same dataset.

Data visualisation

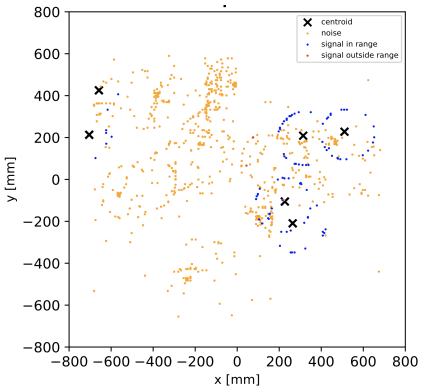
Examples of observed hits on the RICH planes for single events can be seen in Figure 2.12, for both RICH 1 and RICH 2. The coordinate systems of the two physically separated detector planes for RICH 1 and RICH 2 are joined in this representation. For RICH 1 plots, the upper (lower) detector plane spans the space for which $y > 0$ ($y < 0$). In RICH 2 plots, the right (left) detector plane refers to the coordinates $x > 0$ ($x < 0$).

A cross marks the extrapolated charged particle track locations. Photon signals associated with one of the tracks are shown in blue or red, depending on whether they are within a specified radius search range. The radius search ranges are described in more detail in Section 2.4.2.

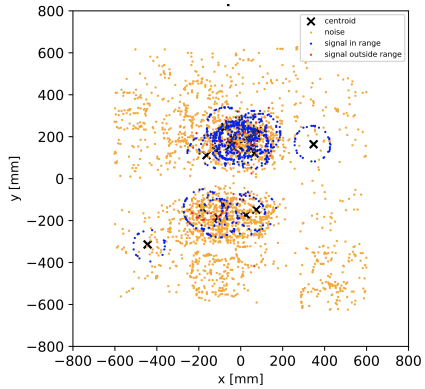
[2]Database configuration “upgrade/dddb-20200508”,
simulation configuration “upgrade/sim-20200508-vc-md100”.



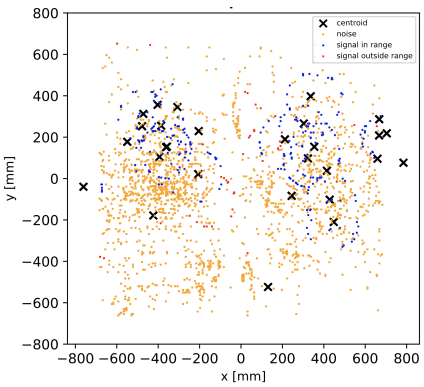
(a) RICH 1, one primary vertex.



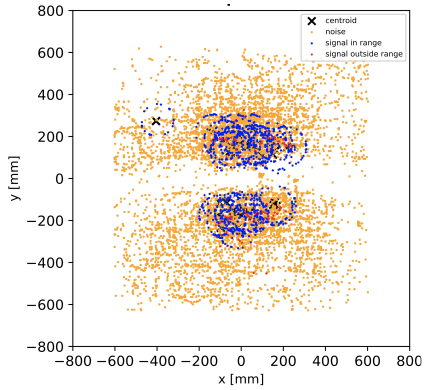
(b) RICH 2, one primary vertex.



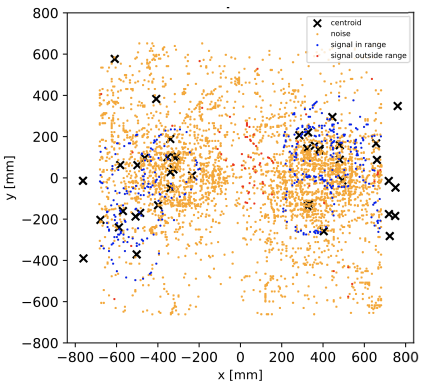
(c) RICH 1, five primary vertices.



(d) RICH 2, five primary vertices.



(e) RICH 1, ten primary vertices.



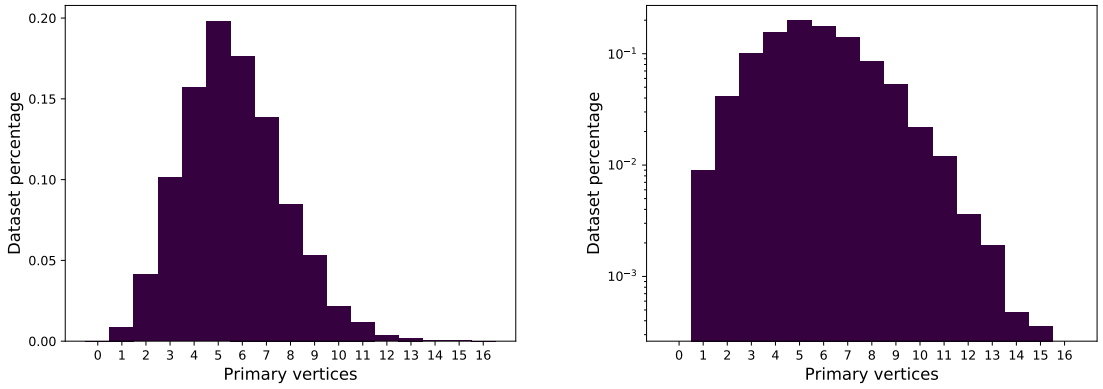
(f) RICH 2, ten primary vertices.

Fig. 2.12 A single event example for the RICH 1 (left-hand side) and RICH 2 (right-hand side) detector planes, for a single primary vertex (a,b), five primary vertices (c,d), and ten primary vertices (e, f). The extrapolated centres of tracks with signals in all tracking detectors (so-called long tracks) are marked by black crosses, the MC detector hits within the radius search range in blue, MC hits outside the search radius in red, and noise in orange.

The detector signal includes background hits, such as electronic noise or noise from particles other than photons reaching the photon detectors. Orange dots illustrate the background hits that are not associated with the reconstructible tracks.

Three cases of single events are shown. The first, in Figures 2.12a and 2.12b, features only one primary vertex. For this type of event, it is comparatively simple to distinguish between individual rings from tracks, six in this example, even by eye. The second case, in Figures 2.12c and 2.12d, contains 5 primary vertices. The concentration of detector signal hits in the centre of RICH 1 is visible, while the tracks are more separated in the RICH 2 detector due to the deviation of the LHCb magnet. The third case, in Figures 2.12e and 2.12f, shows an event with 10 primary vertices. Individual Cherenkov circles are difficult to distinguish. Events with such a high number of primary vertices will become much more frequent with the increase in luminosity of the HL-LHC.

Two points are important to stress in the discussion of Figure 2.12. First, the colour separation of the detector signal in blue (MC hit inside range), red (MC hit outside range), and orange (noise), are only for the benefit of the reader. The PID algorithm is not aware of this information. Second, the MC signal hits for a given track represent a background in the reconstruction of another track. This combinatorial background is especially important for events with a large number of primary vertices.



(a) (b)

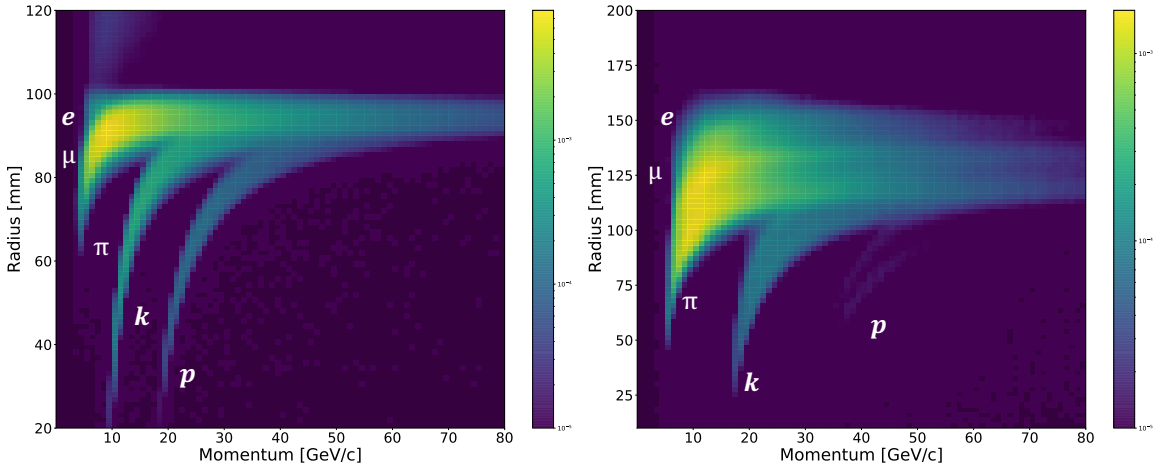
Fig. 2.13 Normalised distribution of number of primary vertices per event on a linear (a) and logarithmic scale (b).

In Figure 2.13 a normalised distribution of the number of primary vertices per event for the dataset used in this study is shown. The number of proton-proton interactions n per

bunch-crossing follows a Poisson probability distribution:

$$P_n(\mu) = \frac{e^{-\mu}\mu^n}{n!}, \tag{2.31}$$

where μ represents the expected value of n and the variance of the distribution. For the dataset in this study, the binned mean is $\mu = 5.63$. The mean value of the distribution is proportional to the luminosity of the LHC [87]. One can see from the distribution that events with 3 to 8 primary vertices comprise over 85 % of the events in the dataset.



(a) RICH 1 (b) RICH2

Fig. 2.14 Cherenkov radii versus momentum in RICH 1 (a) and RICH 2 (b) on a logarithmic colour scale. These distributions are not used for PID but to demonstrate the particle separation and saturation towards high momenta.

Figure 2.14 shows the Cherenkov radius versus momentum for the particle species in the dataset. These figures are not used for PID but to visualise trends in the radius and momentum distribution of the particles in the dataset and differences between the two detectors. Deuterons are rare in the dataset and hence not visible in the figure. The plots illustrate several features that have been discussed above. Firstly, the Cherenkov radii reach similar values at high momenta, setting a physical limit to the PID abilities. Secondly, at lower momenta, the threshold for Cherenkov generation is visible. The particle momenta above which Cherenkov photons are generated are depicted in Table 2.1 for the RICH 1 and RICH 2 radiator gases.

	RICH 1 (C_4F_{10}) [GeV/c]	**RICH 2 (CF_4) [GeV/c]**
e	0.01	0.02
μ	2.00	3.34
π	2.55	4.27
K	9.41	15.7
p	17.7	29.7
D	35.5	59.3

Table 2.1 Cherenkov momentum threshold for the six particle species in the RICH 1 and 2 radiator gases.

2.4.2 Feature engineering

In the previous section, the dataset was illustrated and explained. This section describes how features from the dataset are selected, which serve as the input for the convolutional neural network.

The data input, CNN training, and performance assessment are divided into momentum bins to assess the neural network's performance separately for individual momentum regimes. The width of the momentum bins is motivated by two factors. The first factor is the changes of the particle radii of the most prevalent particle types in a given range. As a result, momentum bins up to 20 GeV/c are chosen to be of a smaller width of 1 GeV/c, while at higher momenta particles approach similar radii where they can be grouped into larger momentum bins of 5 or 10 GeV/c bin width. The second factor is the number of available charged particle tracks for a given range. Towards higher momenta, the number of particles declines and hence also the number of available tracks to train the network. The number of charged particle tracks per momentum bin for the dataset of 10^6 events is shown in Figure 2.16a.

As described above, the data structure after the BRUNEL reconstruction is event-based, defined by the LHC bunch-crossings, usually including multiple proton-proton interactions. The format for neural network training consists of pairs of input and label. In this case, the desirable label is the PID. Hence, the event-based structure of the data is further broken down into pairs of particle information and label. Particle information is selected as input based on *availability* and *feasibility*. Availability denotes in this context any information that exists in detector data from LHC collisions. Therefore, no MC information can be used. Feasibility is determined by whether a piece of information results in a significant PID performance improvement. Any redundant information should be left out to minimise the algorithm's computing and data processing requirements. While availability is clearly

defined, the feasibility of information as input to the neural network is subject to investigation in this chapter.

A non-exhaustive list of available variables is

1. the position of the extrapolated charged particle track on the detector plane,
2. the charged particle momentum,
3. the position of the detector signals,
4. the detector (RICH 1 or RICH 2),
5. the number of primary vertices in the event.

Additional variables are not considered because either their benefit is expected to be negligible, as the LHCb magnet polarity, or they would rely on a computationally expensive reconstruction component, such as the hypothetical Cherenkov angle on which the conventional RICH PID algorithm relies.

The approach presented in this thesis requires the extrapolated position of the track on the detector planes and the momentum of the particles. This expands on the information provided by the tracking stations. While the particle momenta are in any case determined in the LHCb reconstruction, the track extrapolation is solely for the RICH reconstruction. The track extrapolation comprises only a small part of the computational budget of the RICH reconstruction while greatly benefiting most particle identification approaches. Trackless ring finding algorithms have been proposed, however, their feasibility is limited by long computation times and required low combinatorial backgrounds [88, 89].

Cherenkov photon search radii

In Section 2.3.7, the motivation behind using convolutional neural networks for the RICH PID was explained. It relies on the fact that the Cherenkov rings, detected by the RICH photon detectors, represent spatially correlated information: the photon detector pixels which are within the Cherenkov radial range around an extrapolated track centre, contain the most relevant information. To exploit this fact, an annular region around each charged particle is defined, such that it includes the Cherenkov photons for each particle hypothesis, discarding the irrelevant information of all the other pixels on the detector plane. These radial ranges are statically defined for each momentum bin.

To determine the annular acceptance, the distances of MC Cherenkov photons are plotted with respect to the extrapolated track centres. This is illustrated in Figure 2.15 for the momentum

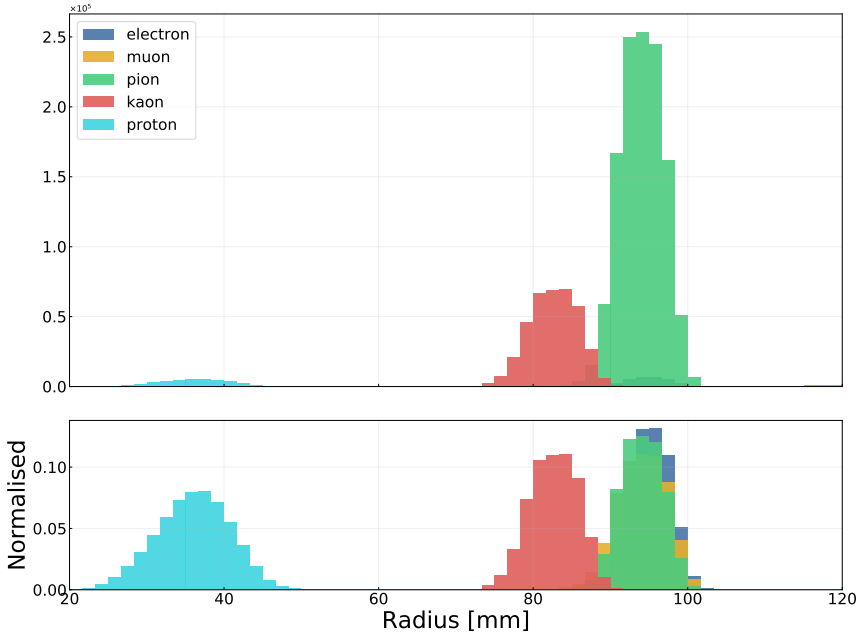
range 19 to 20 GeV/c. In the plot, the absolute (top) and normalised (bottom) frequency of occurrence of each particle with respect to the total number of particles of this species can be seen.

One can make several observations from this figure. The first one is that pions are by far the most prevalent particle species in this momentum range. The second one is that the *light* particle types, namely electron, muon, and pion, appear at similar radii. Thirdly, one can note that kaons make up the second-largest part of the distribution and are well separated from the light particles in this momentum range. The fourth observation is that protons start appearing in RICH 1 since the momentum is just above the Cherenkov threshold of 17.7 GeV/c for the C_4F_{10} radiator (see Table 2.1). In RICH 2, the momentum of the protons is too low to generate a significant number of Cherenkov photons.
The annular region of interest for the 19 to 20 GeV/c momentum bin is selected to include the majority of particles. For RICH 1, two momentum ranges are set: 20 to 50 mm and 70 to 104 mm. For RICH 2, a single range is used: 50 to 160 mm.

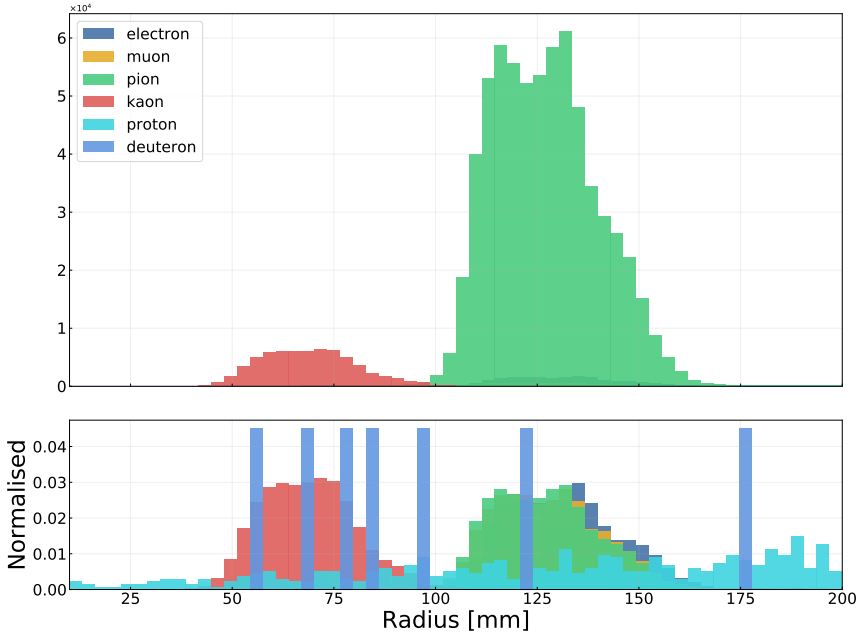
The radius search ranges are noted in Table A.1 and a complete list of Cherenkov radii figures for each momentum range can be found in Figure A.4, Section A.2. Figure 2.16b shows the fraction of each particle species as a function of the momentum. Especially at low momenta, pions are by far the most frequent particle in the dataset. Towards higher momenta, the kaon fraction increases up to 40 %. Protons make up about 15 % of the particles above the Cherenkov threshold. Electrons and muons make up only a small portion of the dataset. The number of deuterons in the dataset is negligibly small.

Image transformation

The accepted pixels within the radius search range around each track centre provide the two-dimensional input for the convolutional neural network. To translate the radial information of the search ranges to a more compact format with simpler features, a polar transformation is applied, as illustrated in Figure 2.18. Each resulting image has a size of 64×64 pixels. An example of the polar transformation for a single event in RICH 1 and 2 is shown in Figure 2.17. This event features only one primary vertex and three charged particle tracks: one kaon with 23 GeV/c momentum and two pions with 7 GeV/c and 8.5 GeV/c. The resulting polar transformed images for the three particles are shown on the right-hand side of the figure. To demonstrate the amount of signal and background, even for non-overlapping tracks, the signal is coloured in blue and the background in orange. The convolutional neural network receives these as input, as monochrome images, without the colour-coded information.

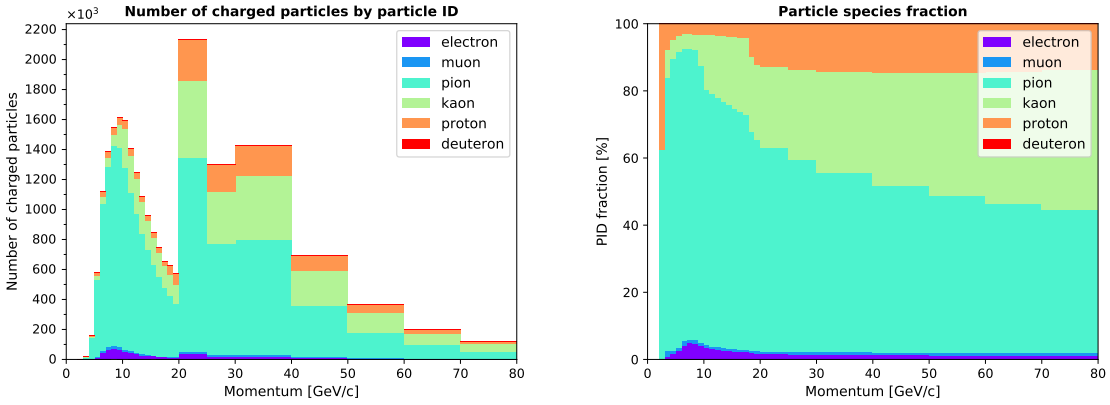


(a) RICH 1, Cherenkov radii for 19 to 20 GeV/c.



(b) RICH 2, Cherenkov radii for 19 to 20 GeV/c.

Fig. 2.15 Histograms of distance in mm between the extrapolated track centres and MC Cherenkov photons, for the momentum range 19 to 20 GeV/c, in RICH 1 (a) and RICH 2 (b). The top part of the plot shows the absolute incidence of the particle for this radial range and the bottom part the normalised histogram for each individual particle type.



(a) (b)

Fig. 2.16 (a) The number of charged particle tracks for a given momentum range sub-divided into particles species. The dataset contains 10^6 events. The bin widths are chosen to ensure sufficient events for effective training. (b) The relative number of each particle species in a momentum bin.

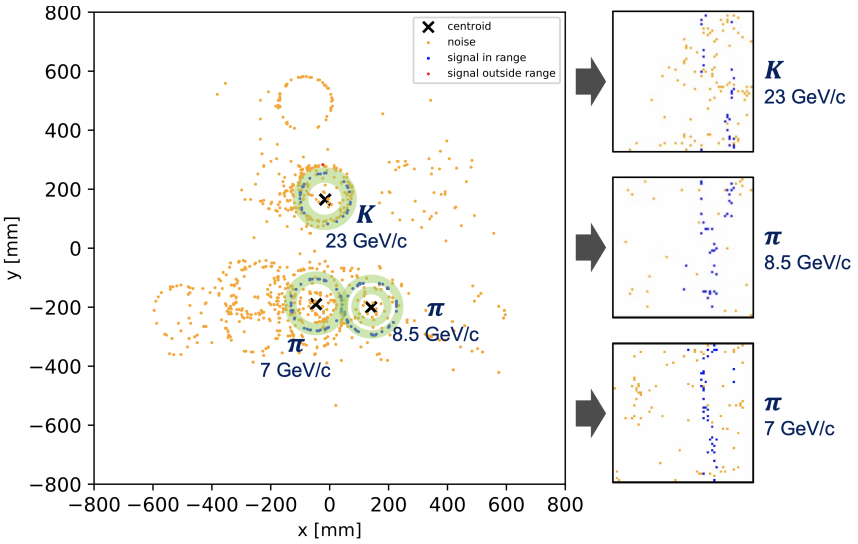
In Figure 2.13 one can see that single primary vertex events are comparatively rare. Figure 2.19 shows examples of 64×64 images from events with 3, 5, and 10 primary vertices, respectively, which are much more common in the dataset.

2.4.3 The PID neural network

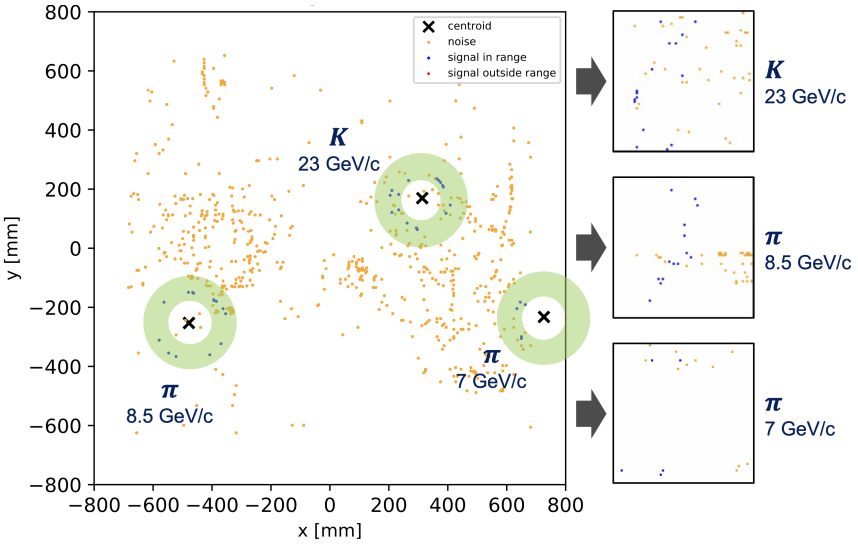
For each momentum range, the neural network is individually trained and assessed. 80 % of the dataset is used to train the network, 10 % to validate it, and 10 % to test and assess its performance. The primary input is the 64×64 monochrome image for a given charged particle track. In the beginning, the network receives only RICH 1 images. With this configuration, the hyperparameters of the network are tuned. After the tuning, the effect of additional inputs, such as the corresponding RICH 2 image, or scalar variables, such as the particle's momentum, the $x-y$ coordinates of the extrapolated track centre and the number of primary vertices in the event are investigated.
The momentum is normalised with respect to the momentum range it falls in:

$$p_{norm} = \frac{p - p_{min}}{p_{max} - p_{min}}, \tag{2.32}$$

where p is the momentum of the particle and p_{min} and p_{max} the lower and upper boundaries of the corresponding momentum range, respectively.



(a) Single RICH 1 event with one primary vertex.



(b) Single RICH 2 event with one primary vertex.

Fig. 2.17 Visualisation of a single event in RICH 1 (a) and RICH 2 (b), with radius search ranges shown in light green. This event contains only 1 primary vertex and three charged particle tracks creating Cherenkov signal in both detectors. Their energy and PID is illustrated next to the respective extrapolated track centre. The images resulting from the polar transformation of the green search ranges are depicted on the right. Background hits are shown in orange and signal hits in blue to guide the eye. The neural network images do not contain the colour information, but monochrome images.

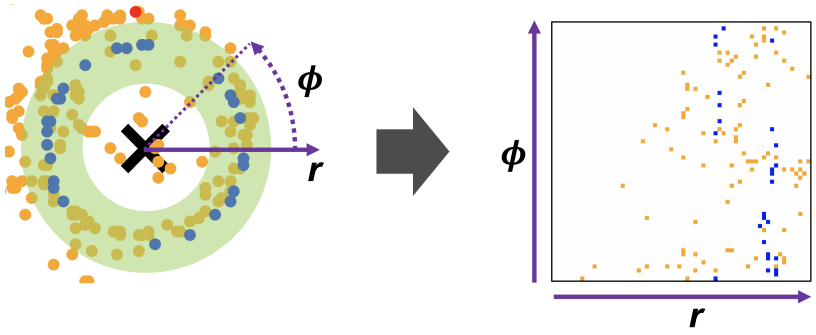


Fig. 2.18 Demonstration of polar transformation to a 64×64 pixel image for a single charged particle.

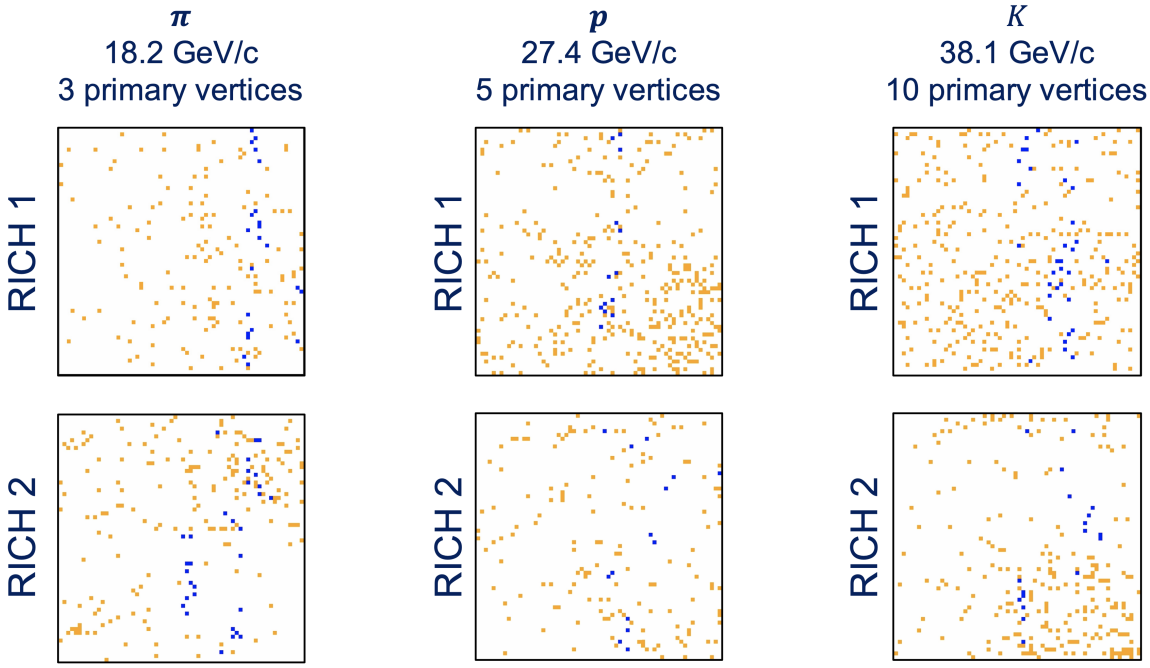


Fig. 2.19 Examples for neural network input images with increasing combinatorial background from left to right, for RICH 1 and RICH 2. (left) pion and 3 primary vertices, (centre) proton and 5 primary vertices, (right) kaon and 10 primary vertices. Background is coloured in orange, and signal in blue.

In Figure 2.20, the structure of the neural network PID classifier is laid out. It was chosen as a starting point with a simple structure that is common to image pattern recognition, based on discussions with collaborators [86]. The optimisation of the network is discussed in Section 2.5.

A 64×64 monochrome image is provided as the primary input. The input image is operated on by three convolutional layer pairs of increasing filter size (32, 64, and 128, respectively), separated by max-pooling layers (pool size 2 by 2). Each convolutional layer features a Leaky ReLU activation function with a negative slope coefficient of 0.1 (Equation 2.17).

The output of the convolutional network is flattened and forwarded to two fully connected layers with 256 and 16 nodes, respectively. Both fully connected layers are regularised by a 25 % node dropout and followed by a Leaky ReLU activation.

The output layer holds 6 numbers, equal to the number of particle species, with a *softmax* activation function.

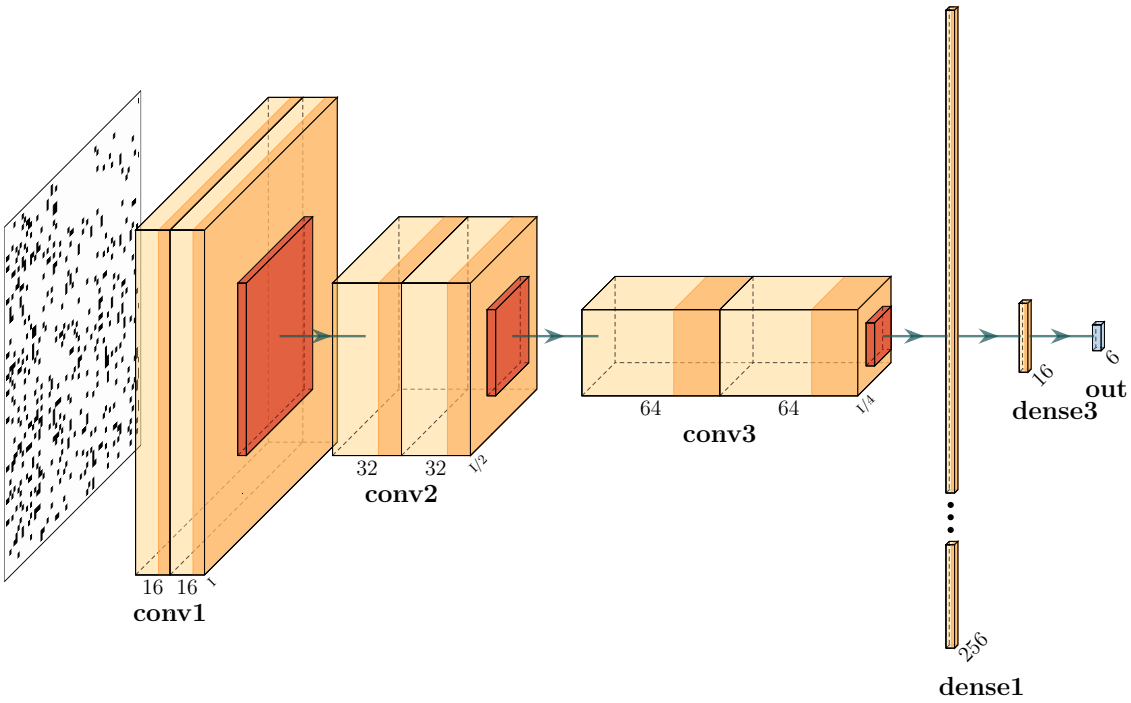


Fig. 2.20 Visualisation of the convolutional neural network architecture before optimisation. The network receives 64×64 monochrome images as primary input. The image is processed by three convolutional layer pairs with increasing filter size, followed by three fully connected layers with 256, 16, and 6 nodes, respectively.

During network training, the output is compared to the label of the charged particle track, which is represented as a six-tuple of zeros and a one for the corresponding particle type. The loss function used is *categorical cross-entropy*, the optimiser *Adadelta*, and the performance metric *accuracy* (Equation 2.8) .

Early stopping is applied to avoid overtraining. The loss during the cross-validation step is used as a measure. If it does not improve for a set number of epochs, the training is stopped. The weights with the optimal validation loss are selected and stored for network testing.

2.4.4 Computing resources

The studies presented in this thesis use the IBM Minsky Power 8+ cluster [3] in collaboration with CERN Openlab [90]. The cluster features four NVIDIA[4] Tesla P100 GPUs with over 3.500 cores. Details of its infrastructure are presented in Table 2.2.

ML IBM cluster infrastructure		
login node	compute nodes	software
IBM S821LC (8001-12C)	IBM S822LC for HPC (8335-GTB)	CentOS 7.4 ppc64le
8 P8 cores at 2.32 GHz	16 P8 cores at 3.26/3.86 GHz	CernVM-FS
64 GB of DDR4 memory	4x NVIDIA P100	NVIDIA CUDA 9.1
	3584 CUDA cores, 16 GB HBM2	
8 TB HDD	4 TB HDD + 1.6 TB NVMe	IBM PowerAI 1.5
	IB EDR between nodes	slurm
1 GbE (/ 10 GbE)	1 GbE to the CERN network	IBM Spectrum Conductor
	1 Xilinx ADM-PCIE-8K5 (CAPI)	Deep Learning Impact

Table 2.2 The IBM Minsky cluster infrastructure. Source: [91].

2.5 CNN optimisation and PID performance

In this section, I demonstrate the potential power of the CNN approach by using a standardised benchmark of PID performance to allow direct comparison with the existing LHCb RICH algorithm.
The performance of the convolutional neural network, described in Section 2.4.3, is improved by hyperparameter optimisation. The optimisation does not aim to be exhaustive but rather to indicate the potential that may be reached with a CNN.

[3] IBM Switzerland Ltd: `https://www.ibm.com/ch-de`
[4] NVIDIA Corporation: `https://www.nvidia.com/`

I explore the impact of background on the CNN performance and demonstrate that it is possible to retain the excellent PID performance in a high occupancy environment by exploiting the capabilities of future RICH detectors in reducing the effective occupancy using high-resolution time-resolved readout.

2.5.1 Performance benchmark

Performance plots from the conventional RICH PID algorithm are used as a benchmark. These are generated with the same dataset as used for the CNN studies. They indicate the expected RICH PID performance for LHC Run 3.

The kaon ID efficiency versus pion misID rate is shown in Figure 2.21. Reducing the delta-log-likelihood threshold results in an improved kaon ID efficiency but at the same time a poorer pion misID rate. Figure 2.21a and 2.21b show two examples, for $DLL > 4$ and $DLL > 0$, respectively. DLL values ranging from -10 to 10 are typically compared to find a suitable threshold. Below $10\,\mathrm{GeV/c}$, kaon and pion separation is difficult due to the absence of Cherenkov photons for kaons (see Table 2.1).

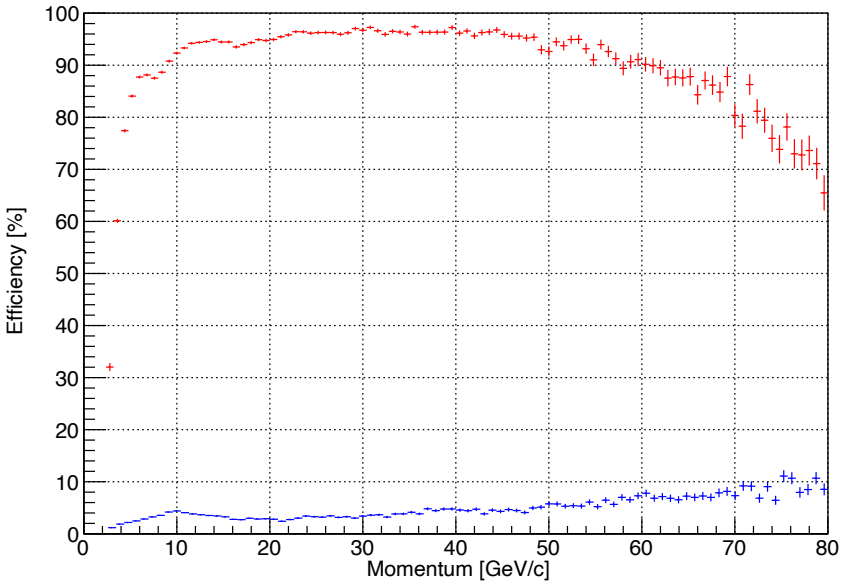
2.5.2 CNN hyperparameter optimisation

A selection of hyperparameters of the convolutional neural network presented in Section 2.4.3 is optimised to explore the potential of a CNN RICH particle identification.

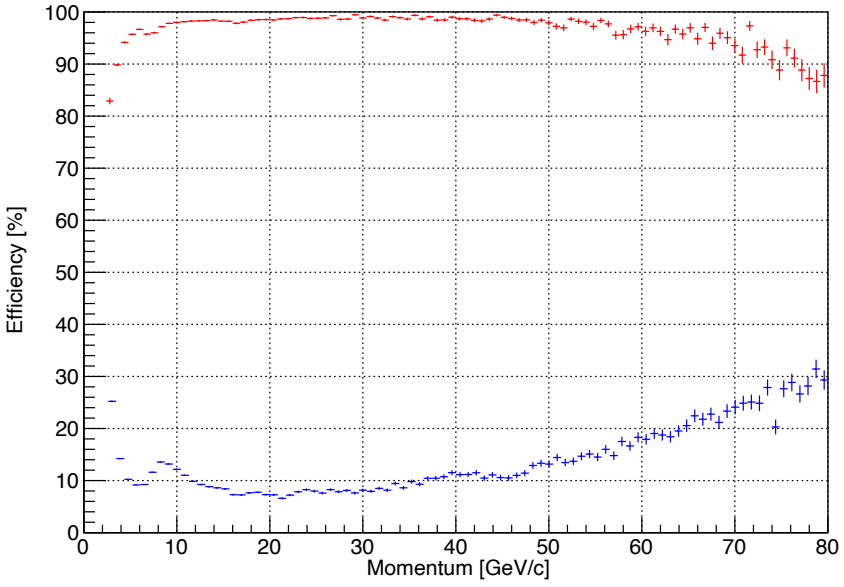
Number of training images

The simulated dataset contains 10^6 proton-proton collision events. As shown in Figure 2.16a, the number of charged particle tracks varies for different momenta, following a Poisson distribution. Only a small fraction of particles and hence images to train the network are available at low and high momenta. A larger momentum bin size is chosen above $20\,\mathrm{GeV/c}$ to counter this effect.

The impact of the number of input images is assessed. While a larger number of training examples usually yields better network performance, the improvement is expected to diminish towards large values. Network training is computationally expensive, and its duration increases with the dataset size. The training duration of the CNN for a single momentum bin with 10^6 images is in the order of several hours. The goal is to find a training dataset size for which a reasonable network performance is reached while minimising the required training time.



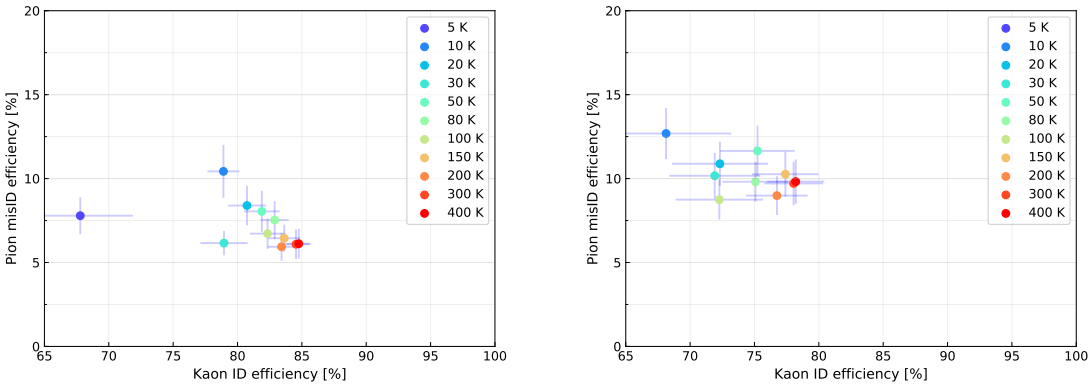
(a) Efficiencies vs. momentum, for $DLL > 4$.



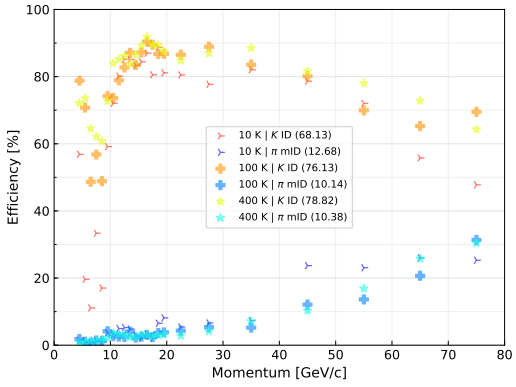
(b) Efficiencies vs. momentum, for $DLL > 0$.

Fig. 2.21 Benchmark RICH PID performance of the conventional algorithm, using the same dataset as for the CNN studies. The kaon ID efficiency is shown in red, the pion misID rate in blue, for $DLL > 4$ (a) and $DLL > 0$ (b). A low threshold increases kaon ID efficiency, adversely affecting the pion misID rate and vice versa.

A second reason to study the effect of the number of training examples is the limited number of well-defined decay topologies in LHC collision data. As suggested in Section 2.4.1, such events could be used to train the network with collision data rather than from simulation. Since only a few decays allow PID without the RICH, the network training would have to rely on a potentially much smaller dataset.



(a) (b)



(c)

Fig. 2.22 Effect of dataset size on the average pion misID vs. kaon ID efficiencies with their respective standard errors, for the momenta between 10 and 60 GeV/c (a) and all momenta (b). Figure (c) shows the efficiencies as a function of momentum for three dataset sizes: 10 K, 100 K, and 400 K. The corresponding average efficiency values are noted in the legend.

In Figure 2.22a and b the average pion misID rate is plotted against the average kaon ID efficiency for different dataset sizes. The data is split into 80 % training, 10 % validation, and 10 % testing tuples. The number of images in the dataset is illustrated in thousands (K). The network training time for a single momentum bin ranges from one hour for 5 K to 40 hours for 400 K images.

The closer a point is to the bottom right of the plot, the better the network's overall performance. In Figure 2.22a, the average is taken for the momentum ranges 10 to 60 GeV/c and in Figure 2.22b for all momenta. The error bars denote the standard error of the corresponding average efficiencies in both dimensions.

For the small dataset size of only 5 K images per momentum range, the low value for both pion misID and kaon ID efficiency can be interpreted as the network predicting almost exclusively pions for all inputs. The 30 K efficiency point in Figure 2.22a is separated from the group. Its relatively good pion misID rate is compensated by a lower kaon ID efficiency. Such statistical outliers may be a feature of neural networks. An option to tune the network towards high kaon ID efficiency or a lower pion misID rate are presented in Section 2.5.2 by using class weights.
Apart from these two outliers, a clear trend is noticeable. A larger dataset yields an improvement in kaon ID efficiency and pion misID fraction, albeit with diminishing gains, as seen in Figure 2.22a. In Figure 2.22b, other factors affect the performance: as illustrated in Figure 2.16a, the number of training tuples is limited by the dataset above 60 GeV/c and below 5 GeV/c. Below 10 GeV/c the performance is influenced by the absence of kaon Cherenkov photons.

In Figure 2.22c, ID efficiencies are shown for the entire momentum range, for 10 K, 100 K, and 400 K images per momentum bin. The smaller the amount of training tuples, the more the performance fluctuates, as also noticeable in the larger error bars in Figures 2.22a and b. Between 10 and 50 GeV/c the kaon ID (pion misID) is close to or above 80 % (below 10 %) for 100 K, and 400 K images.

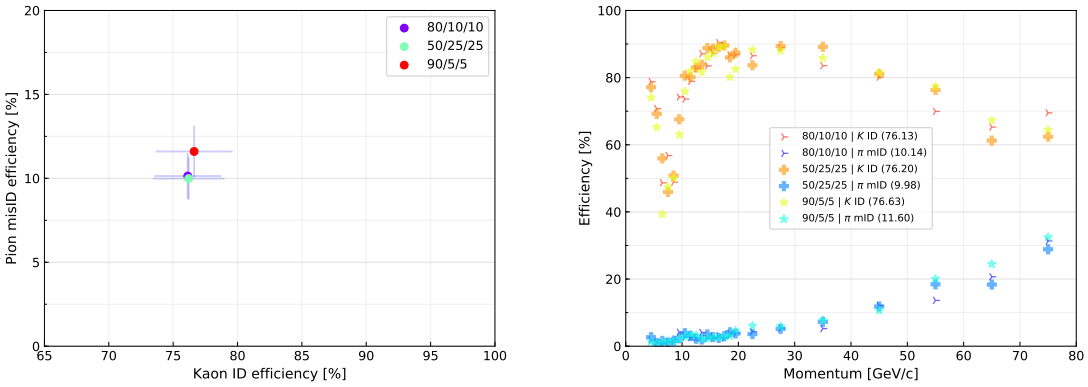
The purpose of the hyperparameter optimisation in the following sections is to improve this performance. Since only relative changes in the performance are important for this optimisation, it will be performed with a dataset size of 100 K images, unless stated otherwise. This is sufficient to achieve a reasonably high PID performance while keeping network training time at a minimum. The full dataset will be used for the final assessment of the PID performance once an optimal configuration of the individual parameters has been established.

Validation and test dataset size

The training dataset is used to adjust the weights and bias parameters of the neural network. The purpose of the validation dataset is to assess the generalisation ability of the model with unseen data during training. The function of the test dataset presents is to evaluate the network performance.
A larger validation dataset represents more input to tune the hyperparameters of a neural

network. In general, this could improve the predictive power of a model, the ability to generalise for unseen data. The test dataset is not used to augment the model parameters but only for its assessment.

Three dataset divisions are investigated, expressed as percentages of the total sample: 80|10|10, 50|25|25, and 90|5|5. The training dataset size is kept at 80 K. Consequently, the 50|25|25 split results in 80 K, 40 K, 40 K images, and the 90|5|5 split in 80 K, 4 K, 4 K images.



(a) (b)

Fig. 2.23 Effect of training and validation set size on the average pion misID and kaon ID efficiency. (a) shows the average pion misID vs. kaon ID efficiencies with the corresponding standard errors. (b) shows the efficiencies as a function of the momentum for three set sizes.

Figure 2.23a shows the average pion misID versus kaon ID efficiencies with their respective standard deviations. In Figure 2.23b, the efficiencies are displayed for the entire momentum range.

No clear trend is visible with increasing validation and training dataset sizes, and the differences are minor compared to the corresponding standard errors.The pion misID rate is marginally worse for the 90|5|5 division. The performance for a 50|25|25 split is similar to the 80|10|10 default split while requiring considerably longer training times due to the larger validation and test sets.

Batch size

This section explores the effect of the batch size on the number of required epochs (the number of times the dataset is passed through the network), stability (the steadiness of the validation loss as a function of epochs), and accuracy (the fraction of correct classifications)

of the network training. During network training, the data is divided into subsets, called batches, used to calculate the model's error. The error is used to update the weights using a gradient descent optimiser variant.

A relatively large batch size usually improves the estimation of the error gradient. However, large batch sizes often cause the training function to converge to sharp local minima, limiting the network's generalisation ability [92].
A relatively small batch size results in a noisy adjustment of the model weights. The noisy updating of the weight can have a regularising effect and hence lead to better generalisation. Additionally, small batch sizes reduce memory use and can improve training stability.
To identify which batch size is most suited for the PID network, the model is trained using eight different batch sizes, ranging from 8 to 1024. An example of the training and validation loss for the momentum range 19 to 20 $\mathrm{GeV/c}$ is shown in Figure 2.24.
Figure 2.25 shows the average validation loss minima as a function of the average number of required epochs, with their respective standard deviations. On average, the network converges to lower minima for smaller batch sizes, requiring a larger number of epochs.

The model with batch size eight does not follow this trend. A reason may be the training patience value, which is set to 10 epochs. If the optimiser does not find a lower minimum within that limit, the training is stopped, and the best loss weights are selected. A batch size of 8 may require more iterations to escape a local minimum due to its small gradient descent step size.

The effect of the batch size on the kaon ID and pion misID performance is illustrated in Figure 2.26. No significant difference is discernible for batch sizes between 16 and 256. A very low or high batch size, as 8 or 1024, leads to unstable performance, as shown in Figure 2.26b. An increase in the number of epochs substantially raises the total training time of the network.
The batch size of 64 represents an adequate compromise between performance, stability, and training time.

Loss function class weights

As illustrated in Figure 2.16b, the relative frequency of particles is unbalanced. At lower momenta, pions make up $\sim$80 % of all particles, decreasing to $\sim$60 % towards 80 $\mathrm{GeV/c}$. The kaon fraction increases from $\sim$ 15 to $\sim$40 %. Protons comprise $\sim$10 % of the dataset, above the proton Cherenkov threshold. As mentioned at the beginning of this chapter, the focus of this study lies on distinguishing kaons and pions, following the conventional RICH PID.

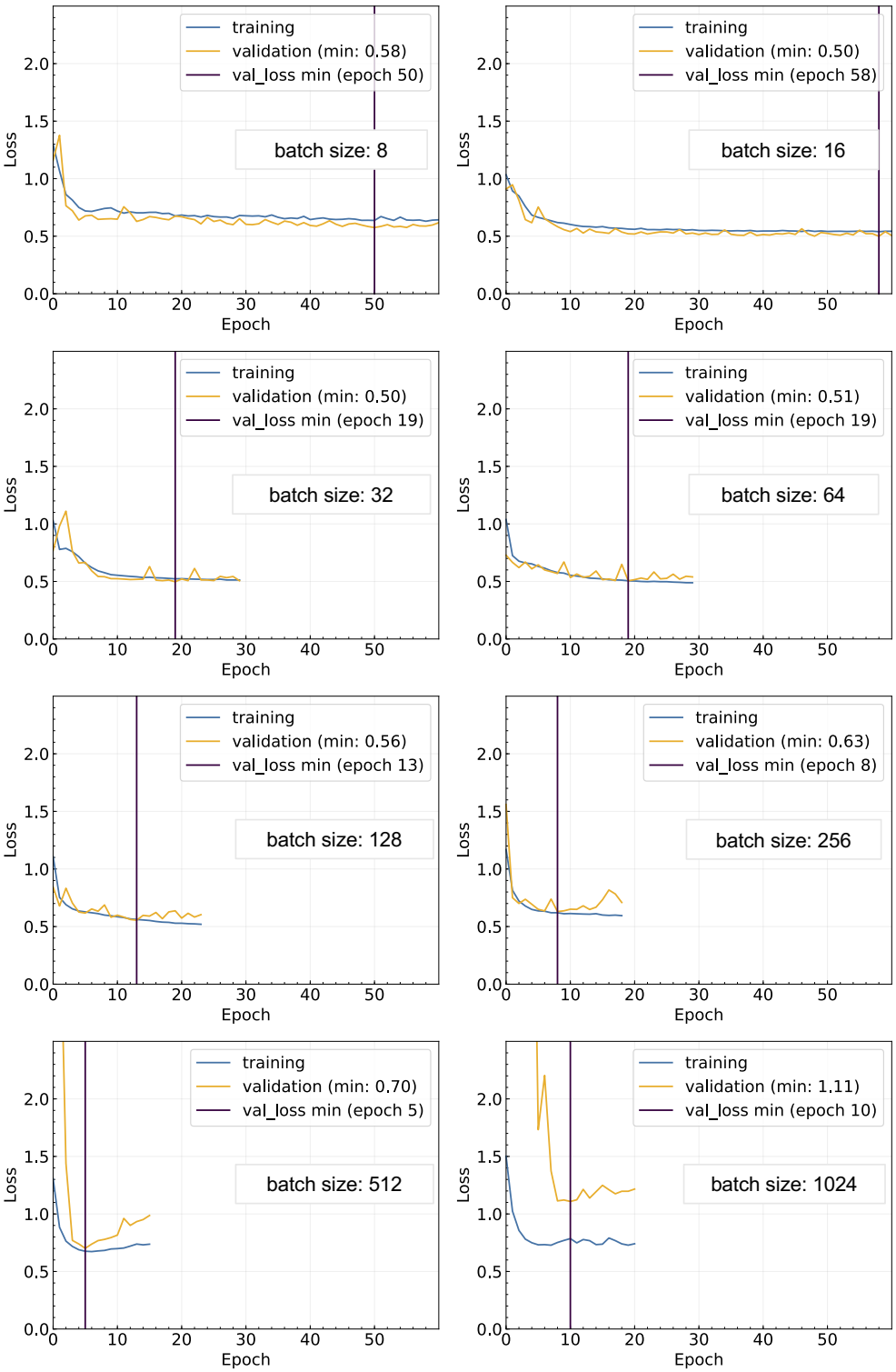


Fig. 2.24 Training (orange) and validation (blue) loss for the momentum range 19 to 20 GeV/c, as a function of epochs, for batch sizes ranging from 8 (top left) to 1024 (bottom right). The smallest validation loss is marked by a vertical line and noted in the corresponding legend.

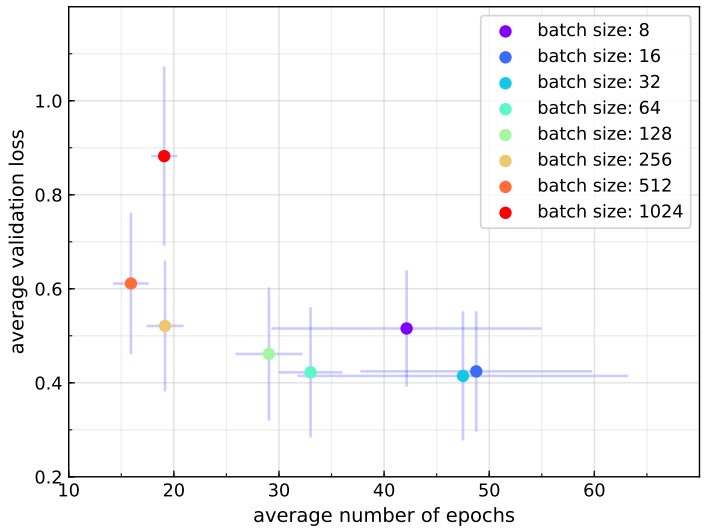
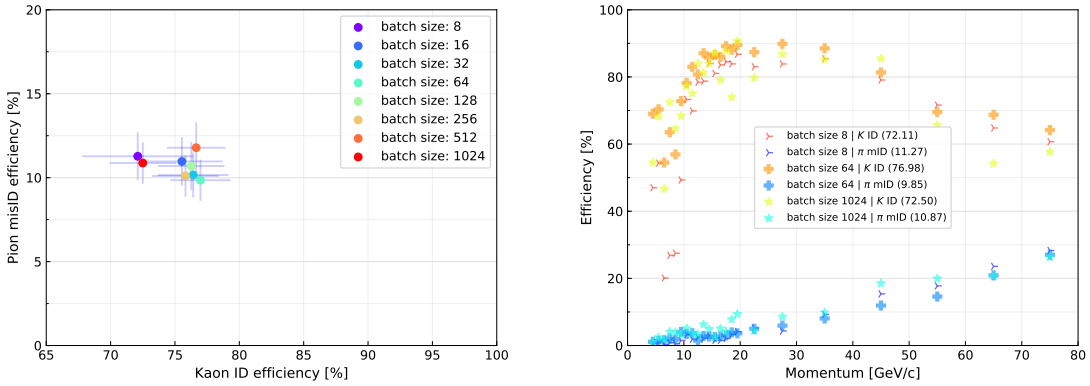


Fig. 2.25 Minimal validation loss versus the number of epochs to reach the minimum. Averaged over all momentum ranges with standard deviations as error bars for batch sizes ranging from 8 to 1024. A smaller batch size results in a lower validation loss but a larger number of required training epochs.



(a) (b)

Fig. 2.26 Effect of batch size on kaon ID and pion misID. Figure (a) shows the average efficiencies with standard deviations and (b) the kaon ID efficiency and pion misID fraction versus momentum for the batch sizes 8, 64, and 1024.

An unbalanced dataset may result in a biased model: a high accuracy and a correspondingly low loss can be achieved by classifying all particles as the most frequently occurring class. There are three ways of addressing that challenge: class rebalancing, over-/undersampling, and loss function class weights.

Class rebalancing is achieved by removing input examples from the over-represented classes to match the number of examples from the least populated one. When limiting the particle identification to pions, kaons, and protons, class rebalancing would lead to removing up to 90 % of training examples to match the proton input.
Over- or undersampling may lead to over- or underfitting, and may be preferred if one or more of the classes of interest feature only a few examples.

In this study, loss function class weights represent a suitable approach to address the particle class imbalance. This can be achieved by modifying Equation 2.33 to

$$LogLoss = -\frac{\sum_{i=1}^{N} w_i \log p_{ij}}{\sum_{i=1}^{N} w_i}, \tag{2.33}$$

where w_i are the weights for class i.

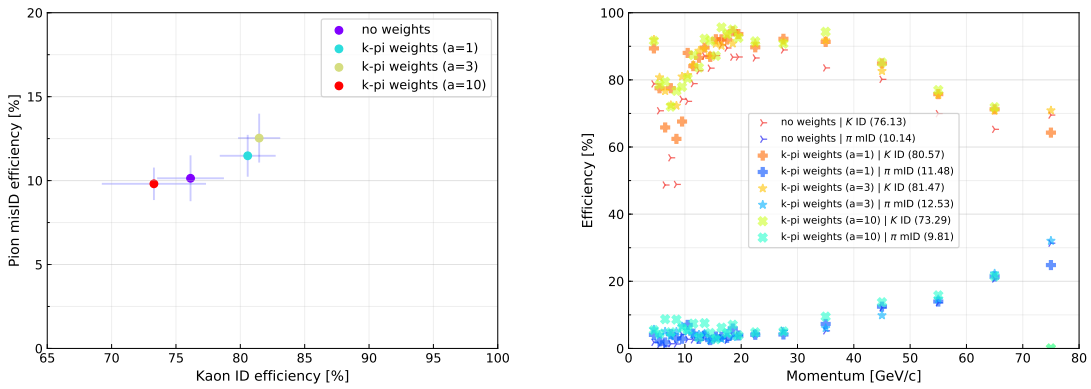
For the RICH PID, two factors have to be considered when choosing suitable weights.
The first one concerns the relative frequency of the particles. Consequently, the weights would be set to

$$w_i^{equal} = \frac{1}{N_i}\frac{N}{K}, \tag{2.34}$$

where N_i denotes the total number of occurrences for the particle of class i, N the total number of particles, and K the number of classes.

A second factor is the importance of the particle for the RICH PID. When choosing the weights as in Equation 2.34, the kaon and pion PID performance sharply decreases since high weights would be assigned to electrons, muons, and deuterons, which are rare in comparison to the other species.
To adjust the weights to the particles of interest, kaons and pions in this case, they are multiplied by an additional factor: $w_i = \alpha_i w_i^{equal}$. This factor is chosen to be small for electrons, muons, and deuterons: $\alpha_{e,\mu,d} = 10^{-3}$.

Different values for $\alpha_{\pi,k,p}$ are compared in Figure 2.27, with an unweighted class model as a reference. The average efficiencies in Figure 2.27a show that a moderate weight factor of 1 or 3 leads to a similar improvement of the kaon ID efficiency while the pion misID slightly



(a) (b)

Fig. 2.27 Effect of adding class weights to the loss function on the average pion misID and kaon ID efficiency. (a) shows the average pion misID vs. kaon ID efficiencies with their respective standard deviations. (b) shows the efficiencies as a function of momentum for three set sizes.

decreases with respect to the unweighted reference. A high class weight factor of 10 leads to a loss in kaon ID efficiency.

Figure 2.27b provides a complete picture of the performance for the entire momentum range. The high class weight factor of 10 leads to misclassification at the high end of the spectrum, classifying all particles as pions.
Both weight factors 1 and 3 result in improved classification. The kaon ID efficiency is consistently higher than the unweighted case, with only minor differences in pion misID. The gain in efficiency is especially pronounced below the kaon Cherenkov threshold of 9.4 GeV/c, and higher for $\alpha_{\pi,k,p} = 3$ than 1.

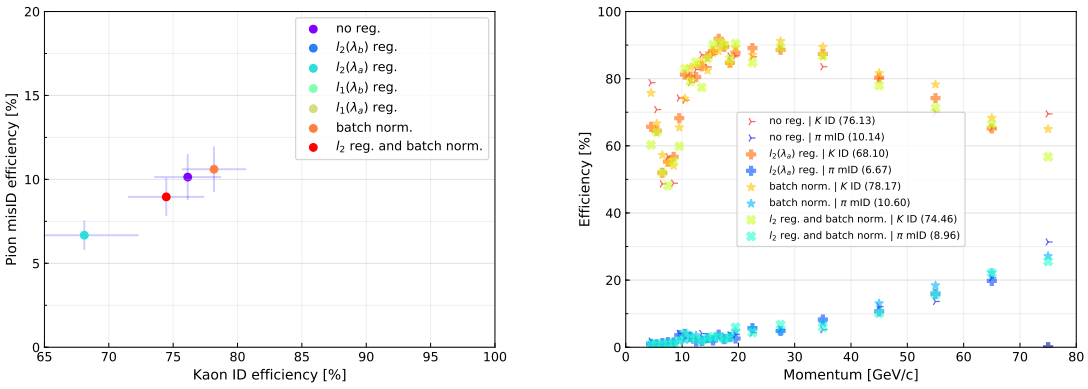
The relative size of the weights for the particle species may be used to bias the PID towards higher kaon ID efficiency at the cost of increasing the pion misID rate and vice versa. Such a mechanism would correspond to the DLL threshold of the conventional RICH PID algorithm and may be similarly used for tuning.

Regularisation

Regularisation is used to minimise overtraining and improve the generalisation ability of a model. As mentioned above, two regularisation techniques are already implemented in the default model, namely early stopping and dropout layers. In this section, two additional regularisation methods are assessed: kernel regularisation and batch normalisation, as de-

scribed in Section 2.3.3. Kernel regularisation is achieved by applying a penalty term to the layer weights. l_1 (Lasso) and l_2 (ridge) terms are compared, each with two regularisation parameters: $\lambda_a = 1 \cdot 10^{-4}$ and $\lambda_b = 5 \cdot 10^{-4}$.

The average pion misID versus kaon ID efficiencies are shown in Figure 2.28a. $l_1(\lambda_b)$ regularisation and the combination of $l_2(\lambda_a)$ with batch normalisation yield a performance slightly below the default model. The other options move the kaon ID efficiency close to zero.
Batch normalisation without kernel regularisation leads to the highest kaon ID efficiency, with a slight increase in pion misID, compared to the default model. Figure 2.28a shows the performance for all momenta, for the non-zero models. Batch normalisation results in a consistently higher kaon ID efficiency, especially below 10 GeV/c.



(a) (b)

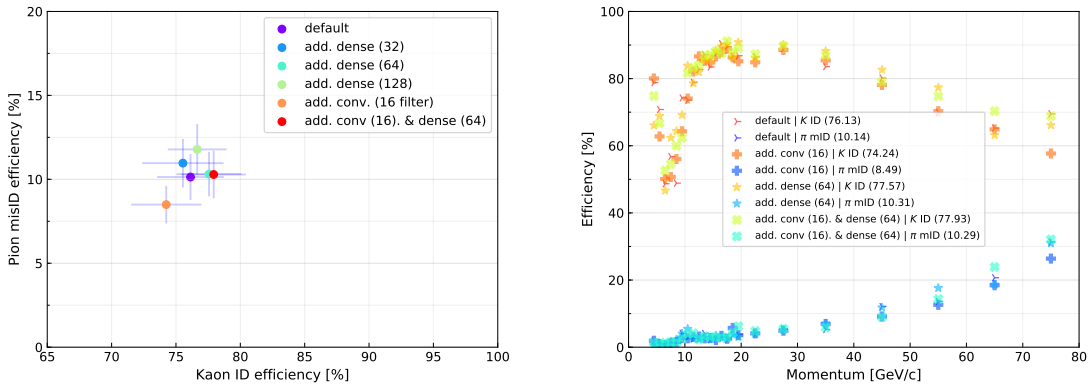
Fig. 2.28 Comparison between different regularisation techniques. The average pion misID versus kaon ID is shown in (a) and for the entire momentum range in (b) for the regularisation methods that yield efficiencies comparable or better than the default model.

Additional network layers

Additional network layers introduce further variables to recognise patterns but may also lead to a higher chance of overfitting. The default network uses three pairs of convolutional layers with 32, 64, and 128 filter sizes, respectively. The convolutional layers are followed by two fully connected (dense) layers, with 256 and 16 nodes.

The choice of network depth determined by the number of hidden layers or different architectures is limitless. In this section, a small subset of all possible options is investigated: an additional dense layer with 32, 64, or 128 nodes, an additional convolutional layer pair with

filter size 16, and the combination of the 16 filter convolutional layer pair and a 64 node dense layer.
The corresponding efficiencies are presented in Figure 2.29a as average values and a selection for all momenta in Figure 2.29b.



(a) (b)

Fig. 2.29 Comparison between different regularisation techniques. The average pion misID versus kaon ID is shown in (a) and for the entire momentum range in (b) for the regularisation methods that yield efficiencies comparable or better than the default model.

The addition of a convolutional layer pair alone achieves a lower pion misID rate. However, combined with a 64 node dense layer, a kaon ID efficiency higher than the default network is achieved.
Since the additional layers increase the total number of network parameters, a larger number of input tuples may improve performance.

2.5.3 richNet PID performance

In the previous sections, a selection of hyperparameters and their effect on the PID performance were assessed. Now, these optimised parameters are used in combination, and the modified CNN is evaluated.

Afterwards, the impact of other input variables is investigated. Namely, momentum, detector plane hit coordinates, and number of primary vertices.
The performance of the resulting optimised CNN, referred to as richNet, will be compared to the conventional PID algorithm for RICH 1 and RICH 2 input images.

Optimised CNN

The previous section examined a non-exhaustive selection of hyperparameters. A schematic of the resulting optimised convolutional neural network with a single input image, referred to as richNetS, is shown in Figure 2.30.

The network receives a single 64×64 monochrome image at a time as input. It is trained for each momentum bin separately, using 300 K images, which are divided into training, validation, and test data, in a ratio of 80 %, 10 %, 10 %. The available maximum of 400 K images did not offer a significant performance improvement, as shown in Figure 2.23, while substantially increasing the computational load.

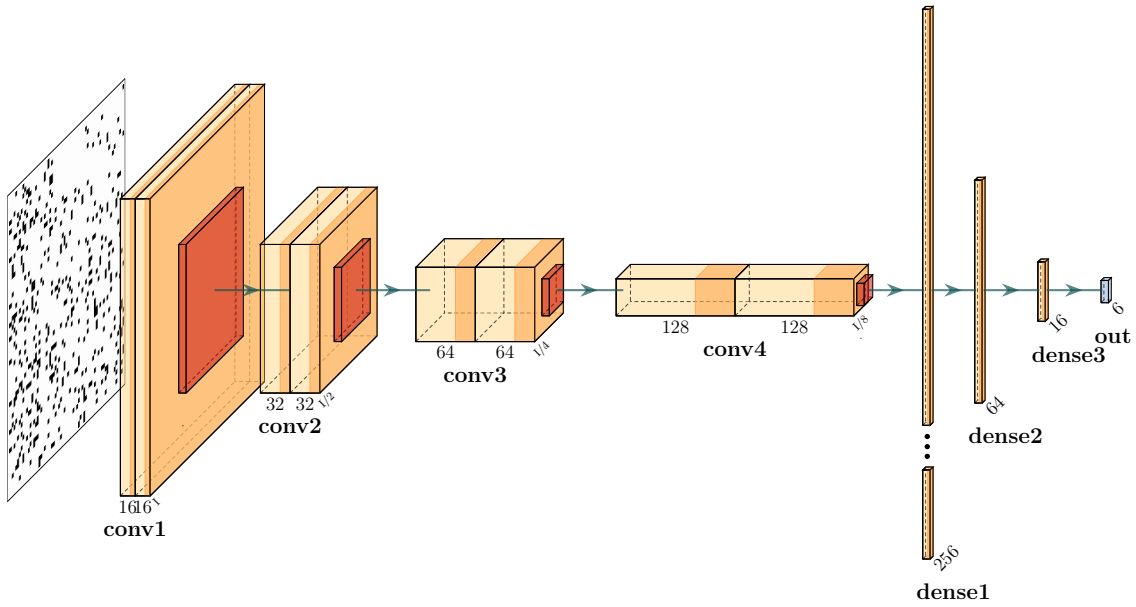
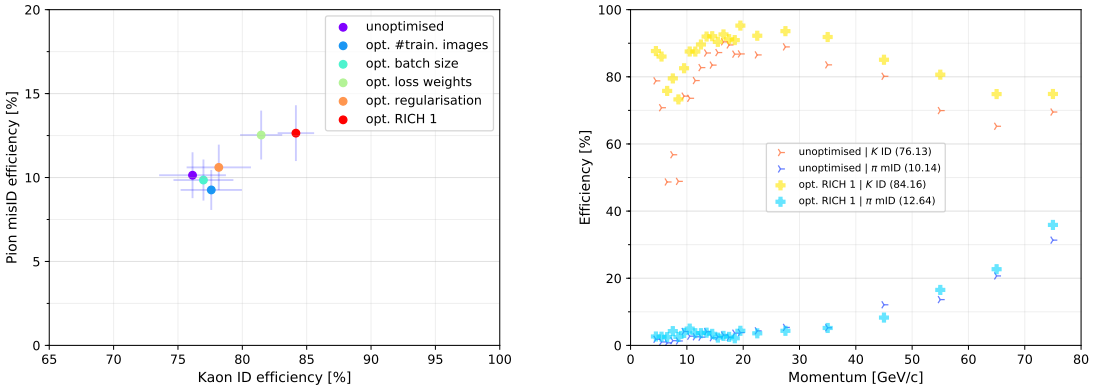


Fig. 2.30 Visualisation of richNetS with optimised parameters. The network receives 64×64 monochrome input image. Four convolutional layer pairs with increasing filter size (16,32,64,128) process the input image. A 2×2 max-pooling follows each convolutional layer pair. The convolutional network output is flattened and followed by four fully connected layers with 256,64,16, and 6 nodes, respectively. The output layer is activated by a softmax function, all other layers by Leaky ReLu activation. Every layer output is batch-normalised.

For the momentum bins 3 to 4 GeV/c and 4 to 5 GeV/c only about 13 K and 130 K images are available, respectively. Similarly, the ranges 60 to 70 GeV/c and 70 to 80 GeV/c only contain about 160 K and 95 K images, respectively. The corresponding distribution of the number of charged tracks as a function of momentum is illustrated in Figure 2.16a. The smaller training set for these momenta is expected to have an adverse effect on the performance, as demonstrated in Figure 2.22a.

Four convolutional layer pairs with increasing filter size (16, 32, 64, 128), stride 3, and zero padding operate on the input. Each convolutional layer couple is followed by two-dimensional max-pooling (2×2 pooling size). The output of the convolutional network is flattened and followed by three fully connected layers with 256, 64, and 16 nodes, respectively. These are regularised by 25 % dropout layers. Leaky ReLU activation functions with a negative slope coefficient of 0.1 are used throughout the network, apart from a sigmoid activation at the output layer with six nodes, corresponding to the number of particle species. Each activation layer output is batch normalised. Categorical cross-entropy is used as loss function and Adadelta as the optimisation algorithm.



(a) (b)

Fig. 2.31 Comparison of models with individual optimised hyperparameters. Average pion misID vs. kaon ID (a) and the efficiencies for all momenta for the default and fully optimised network (b).

Figure 2.31a shows the performance of the models with a single optimised parameter in comparison to the default and the completely optimised network. The optimisation leads to a significant improvement in performance, evident in the direct comparison between the default and optimised CNN in Figure 2.31b

Benchmark vs. richNet PID

To process the images for both RICH 1 and RICH 2, the network's convolutional layer structure is duplicated, and their output is combined in three fully connected layers. The resulting network structure is shown in Figure 2.32 and will be referred to as richNet.

The kaon ID and pion misID efficiencies for richNetS are shown for RICH 1 input images in Figure 2.33a and for RICH 2 images in 2.33b. The performance for richNet with both RICH 1 and RICH 2 input images is shown in Figure 2.34.

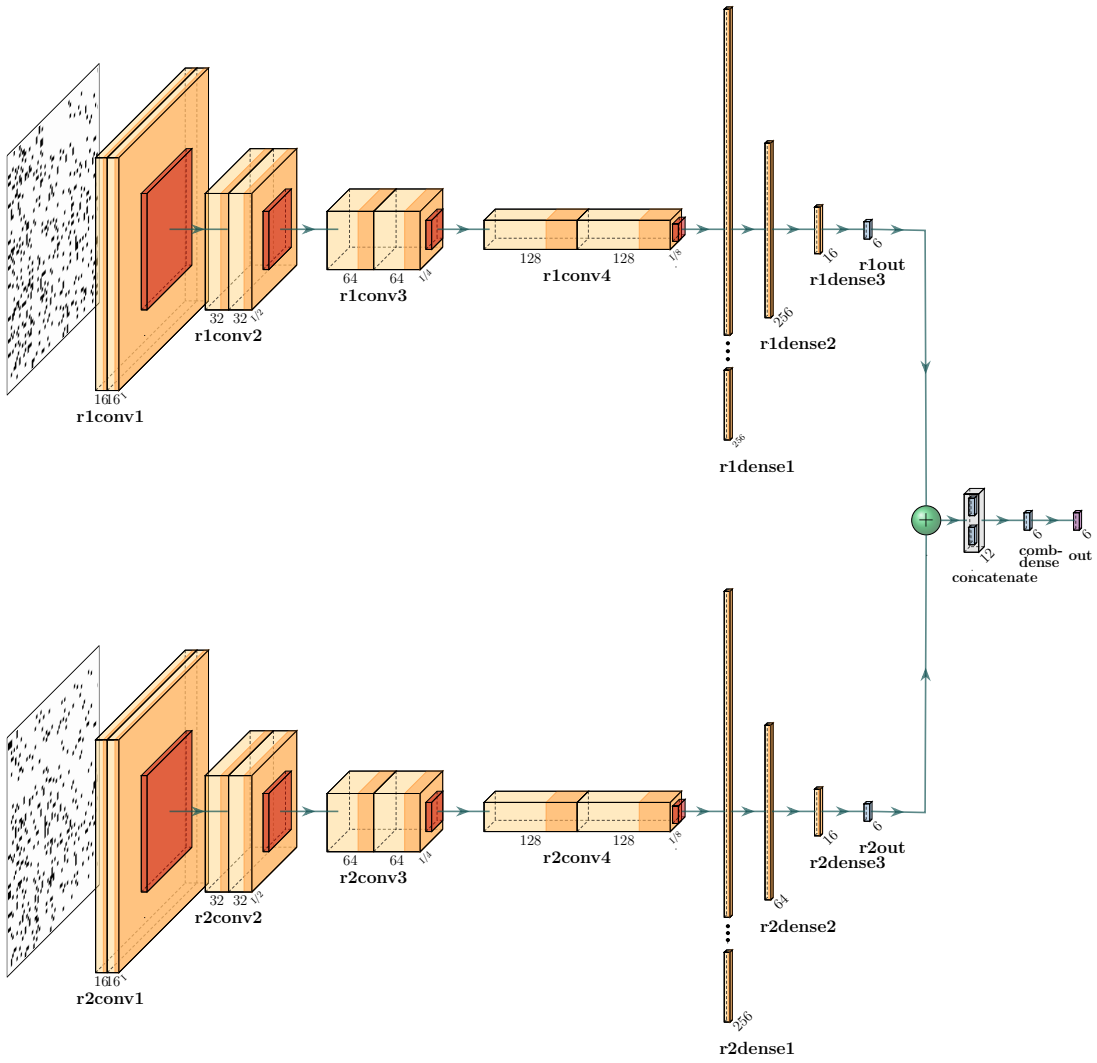
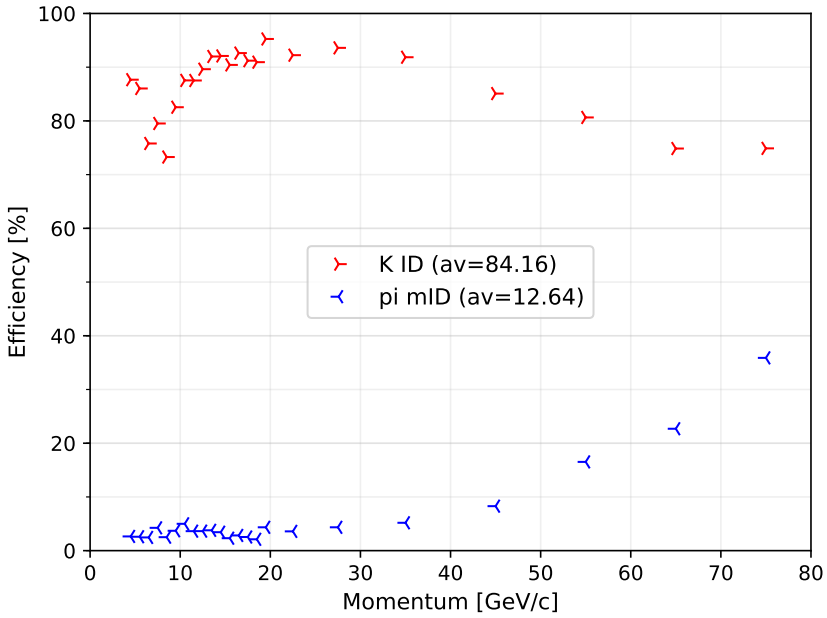
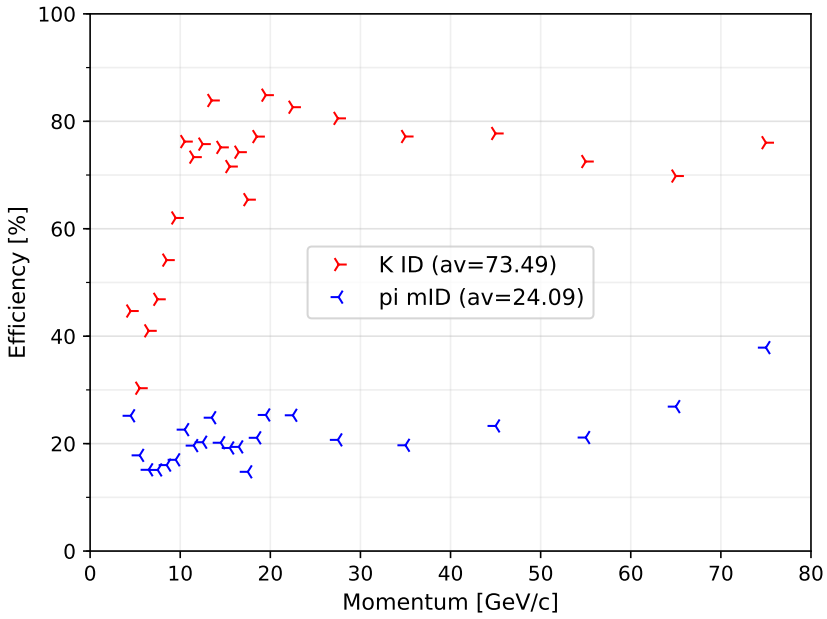


Fig. 2.32 Visualisation of richNet with optimised parameters. The network receives 64×64 monochrome input images from RICH 1 and RICH 2. The structure consists of two richNetS networks, one for each input image, followed by three fully connected layers, with 12, 6, and 6 nodes, respectively. Similarly to richNetS, every layer output is batch-normalised, the output layer is activated by a softmax function and all other layers using Leaky ReLu activation.



(a)



(b)

Fig. 2.33 Optimised convolutional network kaon ID and pion misID efficiencies for RICH 1 (a) and RICH 2 (b) individual image inputs.

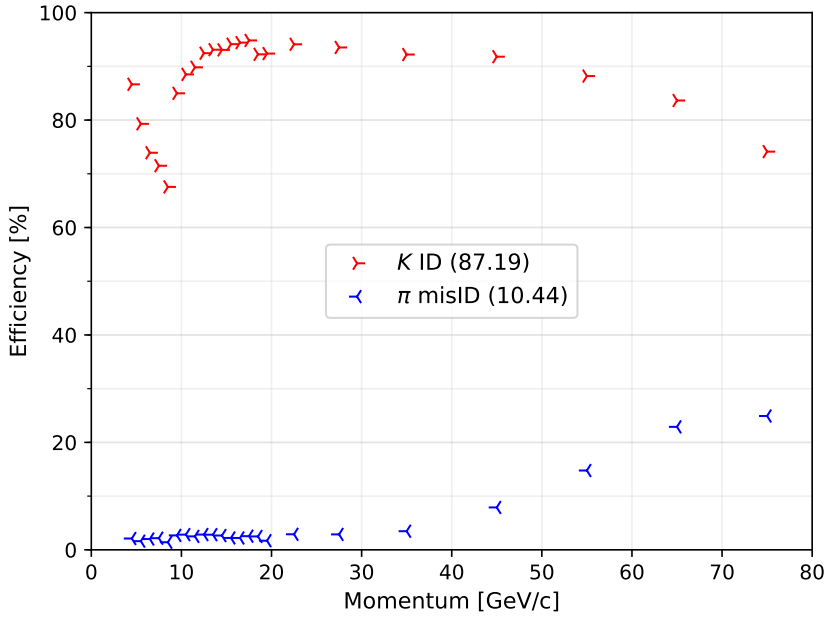


Fig. 2.34 richNet kaon ID and pion misID efficiencies for simultaneous RICH 1 and RICH 2 image inputs.

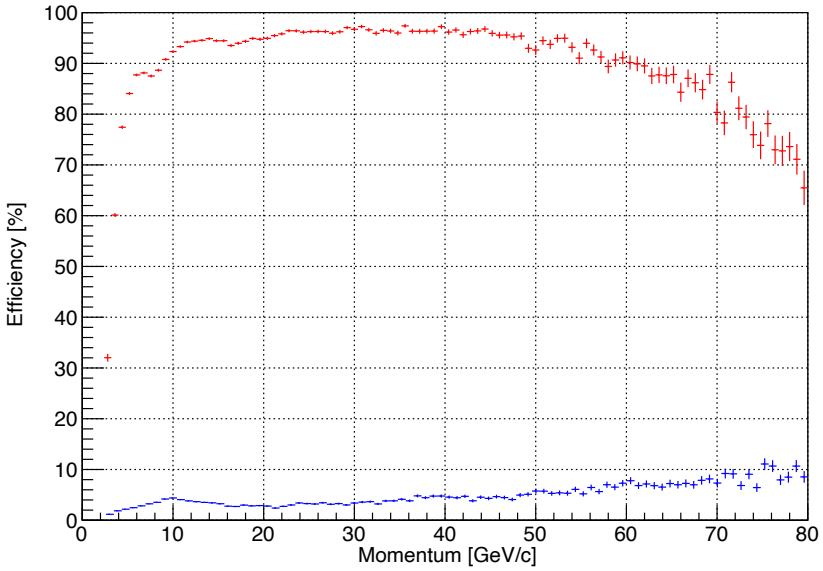


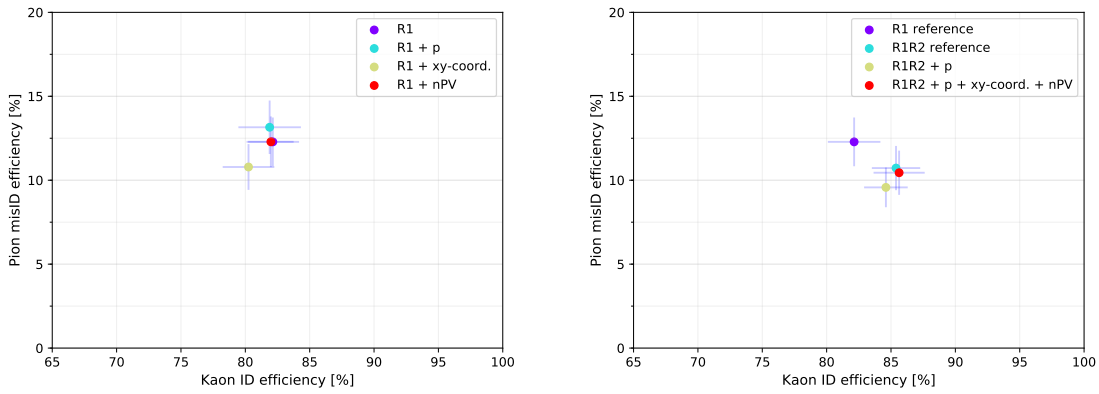
Fig. 2.35 Benchmark RICH PID kaon ID (red) and pion misID efficiencies (blue) for $DLL > 4$, to compare to the richNet performance.

The richNetS performance with RICH 1 input images, shown in Figure 2.33a, is close to the performance of the richNet network up to 40 GeV/c. As expected, the performance decreases when using only RICH 2 input images (Figure 2.33b). The lower performance with RICH 2 inputs may be a result of the optical distortions of its mirror system, as observable in Figures 2.17b and 2.19.
The richNet kaon identification and pion misidentification efficiencies with inputs from both RICH detectors is comparable to the benchmark, shown in Figure 2.35. The average efficiencies are well above 90 % and below 10 % for charged particles with momenta between 11 and 50 GeV/c. For the rest of the range, the kaon ID is close to 80 % and the pion misID increase only up to ∼20 % above 50 GeV/c.

Below 5 and above 60 GeV/c only a fraction of the number of training images is available, which also harms the performance. Further improvement to the performance is expected if all momentum bins receive an equally high number of training images (300 K per bin in this case).

Additional input variables

Three supplementary input variables are provided to the network: the particle momentum (normalised to its corresponding momentum bin), the x-y coordinate of the extrapolated track centre on the detector plane, and the number of primary vertices in the event.



(a) (b)

Fig. 2.36 Average kaon ID and pion misID efficiencies. (a) shows the richNetS performance with RICH 1 image input in comparison to individual additional scalar inputs. (b) shows the richNet efficiencies with RICH 1 and 2 image inputs only, with additional momentum input, and with the combination of all inputs: the two images, the momentum, the xy-coordinates of the track centres, and the number of primary vertices.

Additional inputs increase the network complexity, resulting in longer training times and increased required computing resources. The scalar inputs are processed by three fully connected layers, separated by batch normalisation and dropout layers.

The results of the performance comparison are shown in Figure 2.36. The average kaon ID and pion misID efficiencies are shown in Figure 2.36a for richNetS with

1. RICH 1 image,
2. RICH 1 image & momentum,
3. RICH 1 image & track xy-coordinates,
4. RICH 1 image & number of primary vertices.

Figure 2.36b shows the average efficiencies for richNet (with richNetS as benchmark) for the inputs

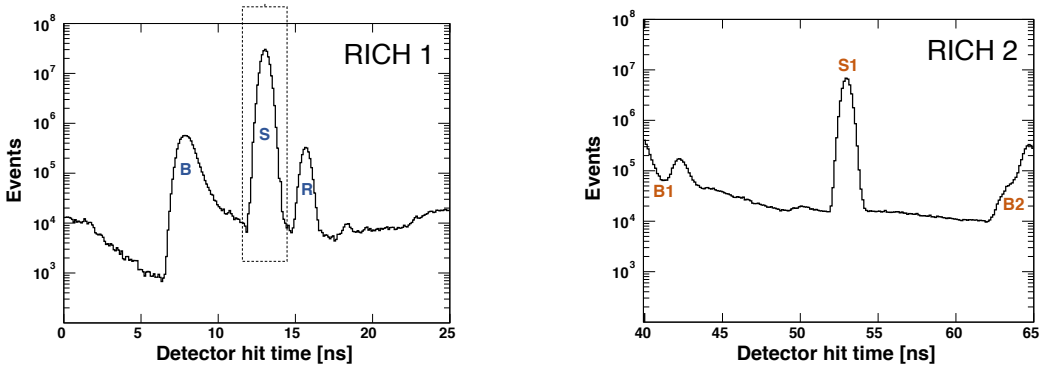
1. RICH 1 & 2 images,
2. RICH 1 & 2 images & momentum,
3. RICH 1 & 2 images & momentum & track xy-coordinates & number of primary vertices.

The investigated additional scalar inputs do not yield a significant ID efficiency improvement. The additional increase in required computational resources discourages their use.

2.5.4 Reducing the background

In this section, the richNet performance is investigated in the light of high precision time-resolved readout in the RICH detector, as foreseen for future upgrades. It has the potential to reduce the background occupancy drastically, and detailed studies are ongoing [93, 94]. In LHCb Upgrade Ib, where MaPMTs will continue to be used as photon detectors, the intrinsic MaPMT transit time spread of about 300 ps determines the time resolution limit. It has been shown that in Run 3 and 4, an estimate of the primary vertex time can be achieved by matching track segments from the RICH system with trajectories associated with a primary vertex by the VELO detector. The RICH detector timestamp associated with the track segment enables calculating the primary vertex time by fitting the time distribution of the photons [94]. An upgrade of the electronic readout chain may introduce a time gate to remove tracks and photons that travel directly to the detector plane or photons reflected multiple times in the mirror systems. A gate of approximately 2 ns would eliminate these

hits, which would arrive well before or after the signal photons, as shown in Figure 2.37. An additional sub-detector or timing layer may be used to reconstruct the primary vertex time. The photon detectors are expected to be replaced during LHCb Upgrade II. R&D on suitable candidates, such as silicon photomultipliers or micro-channel plates, is ongoing. One of the main criteria in the selection of new photon sensor technologies is their timing resolution. Even without significant steps in technology, a resolution of 150 ps seems achievable [94]. The increase in occupancy in the HL-LHC environment during LHC Run 5 necessitates the use of timing information.



(a) (b)

Fig. 2.37 Hit time distribution for RICH 1 (a) and RICH 2 (b). The signal hit peak is marked by an S, background with B, and photons reflected twice in the mirror system by an R. A time gate of 2 ns would acutely reduce the background occupancy in the detectors. Source: [93].

The reduction in the combinatorial background will be estimated in two ways: by pre-selecting input images from events with a specified number of primary vertices and by eliminating non-signal hits from the input images. Both methods do not accurately represent timing background reduction. However, they may indicate the trend and abilities of the richNet concept to identify particles with reduced background. The two approaches and their drawbacks are addressed in their respective sections.

Primary vertex threshold

An increase in timing resolution can be mimicked by selecting events with fewer primary vertices. This has the effect of reducing the combinatorial background and background from scattered and secondary particles but also eliminates some hits that would be retained with a timing cut due, for example, to primary vertices occurring at almost the same time. This approach may be closer to the environment expected from a RICH detector with timing, potentially identifying the primary vertex time or at least separate events in time such that each sub-event contains particles from a small number of primary vertices.

The pre-selection drastically reduces the number of available images to train the network. Figure 2.38a shows the number of images per momentum bin on a logarithmic scale, divided into tracks from events with a given number of primary vertices. A horizontal line is drawn at 50 K (red) and at 100 K (orange). Below a threshold of 6 primary vertices per event, no comparison under similar conditions is possible. The number of tracks from events with fewer primary vertices sharply declines due to its Poisson distribution (Figure 2.13). More than 70 % of the charged particles originate from events with at least seven primary vertices, as visible in Figure 2.38b. Only approximately 10 K input images are available per momentum bin for single primary vertex events. The network is not able to generalise from such a small dataset.

The richNet PID performance is compared in Figure 2.39 with input image limits of 50 K and 100 K per momentum bin.
Restricting the number of input images to 50 K reduces the network's ability to generalise, as shown in Figure 2.22. As a result, the network's loss function is more likely to converge to a higher local minimum and to feature a larger variance when comparing the accuracy between runs. That results in a broader spread of efficiency values.

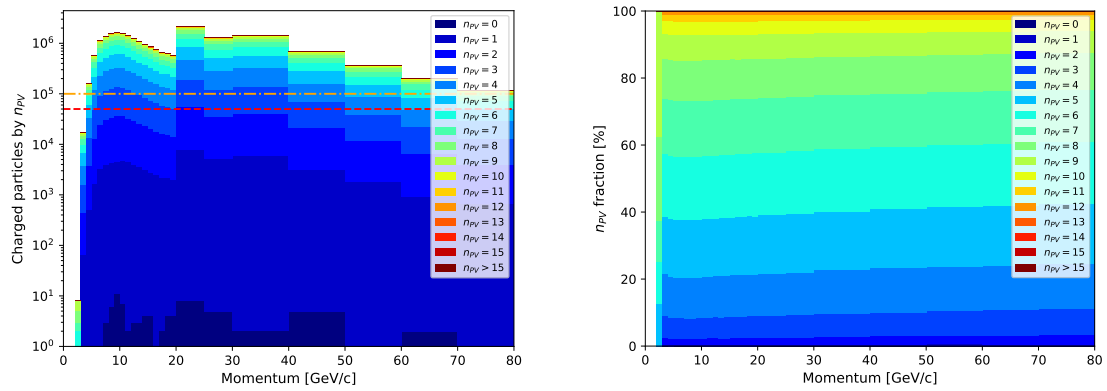


Fig. 2.38 Absolute (a) and relative (b) number of charged particle tracks per momentum bin on a logarithmic scale. The dashed horizontal lines in (a) at 50 K (red) and 100 K (orange) illustrate that for a primary vertex threshold of 5, not enough images are available to train the network equally for all momenta. The coloured bars indicate the number of training images for events with a given number of primary vertices.

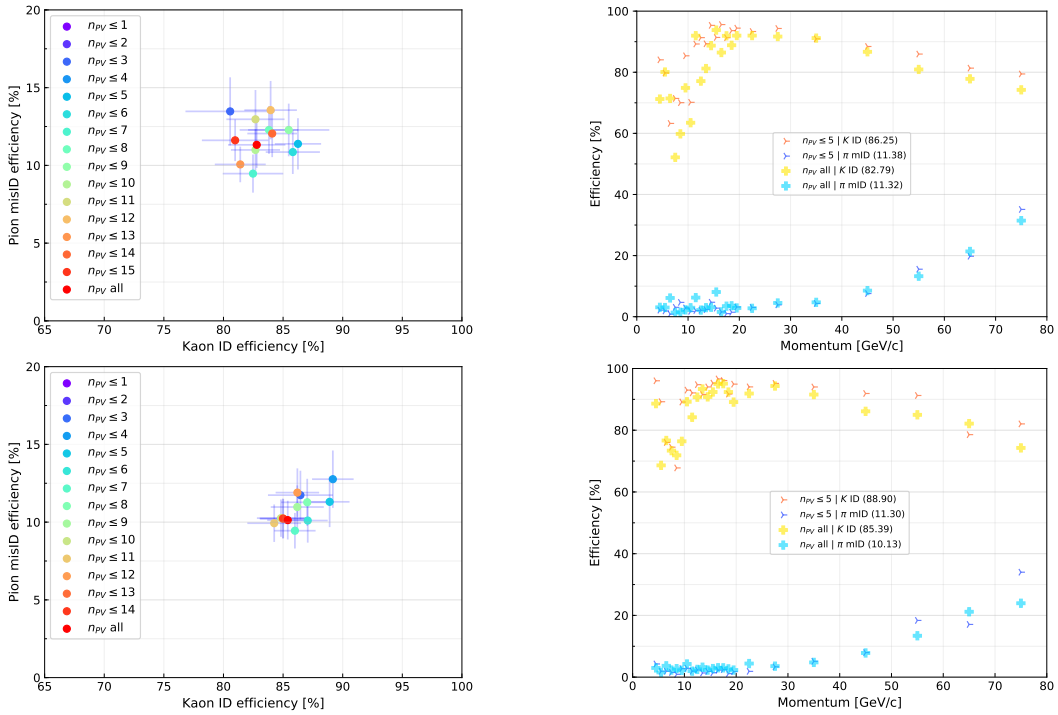


Fig. 2.39 richNetS average PID efficiencies with restrictions of the number of primary vertices per event (left). Comparison between a five primary vertices threshold and default for the entire momentum range (right). The upper limit of images per momentum bin is set to 50 K (top) and 100 K (bottom). Primary vertex thresholds below 5 restrict the dataset to fewer images, impairing the network training.

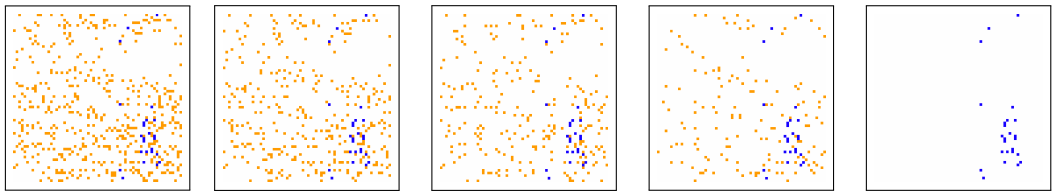
For a 100 K image limit per momentum bin, better overall performance is observable. Events with fewer primary vertices yield an improvement in kaon ID efficiency before the primary vertex threshold limits the number of images to below 100 K.
For both limits, 50 K and 100 K, the performance decreases below a primary vertex threshold of 5 due to the low number of training images. Primary vertex thresholds between 5 and 10 show a minor performance gain, albeit within the standard errors. For thresholds larger than ten primary vertices the performance remains approximately consistent since only a small number of events feature that many primary vertices. As visible in Figure 2.38b, such events comprise less than 5 % of the dataset.

Background elimination

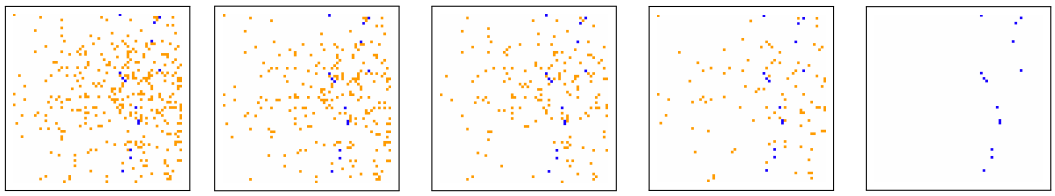
The second approach to reducing noise occupancy is by eliminating background hits for a given input image. This includes hits from separate particle tracks, as well as background such as photons from secondary particles or electronic noise. As illustrated in Figure 2.12,

the majority of hits do not belong to charged particles tracks with a signal in both detectors. However, they may still originate from charged particle tracks or their secondary particles, arriving within a 2 ns time window. The method of eliminating a fraction of non-signal hits may not be directly comparable to the effect of a specific time gate. Instead, it indicates the potential CNN performance for scenarios ranging from a realistic LHC Run 3 environment to ideal conditions without any background.

An example of the background elimination is displayed in Figure 2.40. The input image for a 16.4 GeV/c pion is displayed, taken from an event with 10 primary vertices. As before, Cherenkov photons associated to this pion are coloured blue, all other hits in orange. The percentage indicates how many non-signal hits remain, for RICH 1 (2.40a to 2.40e) and RICH 2 (2.40f to 2.40j).



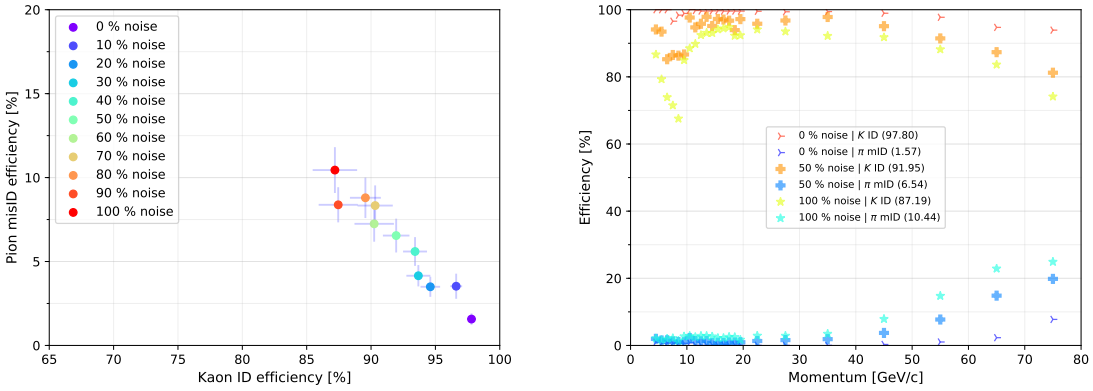
(a) 100 % (b) 70 % (c) 50 % (d) 30 % (e) 0 %



(f) 100 % (g) 70 % (h) 50 % (i) 30 % (j) 0 %

Fig. 2.40 Example of noise reduction for the polar-transformed input image of a 16.4 GeV/c pion from an event with 10 primary vertices, for RICH 1 (a - e) and RICH 2 (f - j). The percentage below an image represents the relative amount of remaining non-signal hits.

The richNet performance with reduced background as described above is shown in Figure 2.41. The average pion misidentification efficiency versus the kaon identification efficiency is displayed in Figure 2.41a. The performance values are plotted for all momenta in Figure 2.41b for 100 %, 50 %, and 0 % background. A clear trend is visible towards 100 % kaon identification and 0 % pion misidentification with reduction of the background. While the 0 % case represents an idealistic scenario, it testifies to the network's ability to reliably distinguish the signature of kaons and pions.



(a) (b)

Fig. 2.41 richNetS performance with reduction of background by elimination of non-signal hits. Average efficiency values (a) and comparison between the default and a signal-only scenario (b).

As mentioned in the introduction of the section, it is not feasible to directly link a background reduction percentage to a timing scenario. That said, any improvement of the RICH detector system that reduces the background occupancy, as anticipated with the introduction of timing, can be expected to affect the richNet PID performance positively.

2.6 Outlook for CNNs in LHCb RICH

Machine learning has proven to be a powerful tool for many pattern and image recognition tasks. The rapid development of computing hardware and technologies supports this trend. Changes in the LHCb computing infrastructure towards GPU computing are under way which motivate a shift from computationally complex CPU algorithms to parallelisable approaches, as provided by artificial neural networks.

This chapter explored the potential of using a convolutional neural network for particle identification in the LHCb RICH detectors. Following the conventional RICH PID, the ability to distinguish between kaons and pions was used to measure performance.
The studies demonstrated a robust method of objectively tuning the behaviour of the CNN through hyperparameter optimisation. The optimised network with image inputs from both RICH 1 and 2 achieves a PID performance comparable to the conventional RICH PID algorithm under LHC Run 3 conditions, while utilising highly parallelisable calculations applicable to GPU processing.

A distinct trend towards optimal kaon ID and minimal pion misID was established for scenarios with reduced combinatorial background as may be expected by the introduction of time information. The effect of time stamp noise reduction was approximated for this purpose.

The work in this chapter presented a path that has the potential to maintain the excellent RICH PID performance under conditions that could be realistically obtained in future upgrades, while reducing the demands on the computing infrastructure. Further work will be needed to confirm the results with a new generation of the simulation framework, including timing, which will increase the predictive power of the background reduction methods.

Chapter 3

Photonic crystals as alternative Cherenkov radiators for particle identification

Photonic crystals and their potential impact on society can be best understood in an analogy to the semiconductor revolution. The term photonic crystal refers to optical materials with a periodic structure. This can lead to novel behaviours, some of which are analogous to the effects of the crystalline lattice in semiconductors. Transistors have shaped the last century's advancements in electronics and are today present in the great majority of devices that shape our daily life. Very similar to controlling the electrical properties of materials due to band-gaps in semiconductors, photonic crystals allow controlling *optical* properties. One can engineer materials that modify light waves through perfect reflection or steer them in well-defined directions using their characteristics. In recent years, the features of photonic crystals have been analytically simulated to create an effect analogous to the conventional Cherenkov effect without many of its constraints.
The goal of the studies outlined in this chapter is to explore the properties of photonic crystals in the realm of high energy physics experiments. In particular, to investigate the possibility to use photonic crystals as alternative Cherenkov radiators in particle identification detectors such as the LHCb RICH detectors.

The two ring-imaging Cherenkov detectors are essential for particle identification at the LHCb experiment. C_4F_{10} and CF_4 gases are used to generate Cherenkov photons in RICH 1 and RICH 2, respectively. While the RICH detectors have demonstrated excellent particle identification performance, the gas radiators are intrinsically limiting components in terms of

both the PID momentum range and the large space requirements.
The working principle of the LHCb RICH detectors and ring-imaging Cherenkov particle identification is introduced in Section 3.1. This is followed by a brief outline of photonic crystals properties in Section 3.2. Their potential as novel radiators for particle identification in particle physics experiments is motivated in Section 3.3.

Studies of photonic crystals as promising radiator materials for future RICH-like high-energy experiments are presented, including the first tests of one-dimensional dielectric photonic crystals as alternative Cherenkov radiators for particle identification in a high-energy particle beam.
The research is divided into three parts: polymer photonic crystal laboratory measurements, the manufacturing of a photonic crystal suitable for high-energy conditions, and the beam campaign of the manufactured photonic crystal. These studies represent the first steps to realising a practical photonic crystal particle identification detector. An alternative approach to PID with photonic crystals where I contributed as co-author is outlined in Appendix B.
In Section 3.4, I introduce a laboratory setup and measurements of a one-dimensional polymer photonic crystals using a radioactive source. The simulated response is compared to the experimental setup results. The section concludes with irradiation tests of the polymer photonic crystal in a high-energy particle beam.
In Section 3.5, I describe the manufacturing and simulation studies of a $SiO_2 \,|\, TiO_2$ photonic crystal, suitable for high-energy beam tests. The effective Cherenkov radiation of the photonic crystal and its sensitivity to parameter variations is simulated using an analytical model.
In the last section of the chapter, Section 3.6, the high-energy beam test of the manufactured photonic crystal as effective Cherenkov radiators are presented. The chapter closes with an interpretation of the results and proposals for future experiments.

3.1 Particle identification with the LHCb RICH detector

The primary purpose of the two LHCb Ring Imaging Cherenkov detectors is to distinguish between charged hadrons (primarily π, p, K), making use of the Cherenkov effect [95, 31]. This effect occurs when charged particles traverse a transparent medium with a velocity, v_p, greater than the phase velocity of photons in that medium, c_m:

$$v_p > c_m = \frac{c}{n(\omega)}, \tag{3.1}$$

where c represents the speed of light in vacuum, and $n(\omega)$ the refractive index of the traversed medium, which is a function of the photon frequency. In this case, photons are emitted along

the particle track with a characteristic angle (θ_{Ckv}):

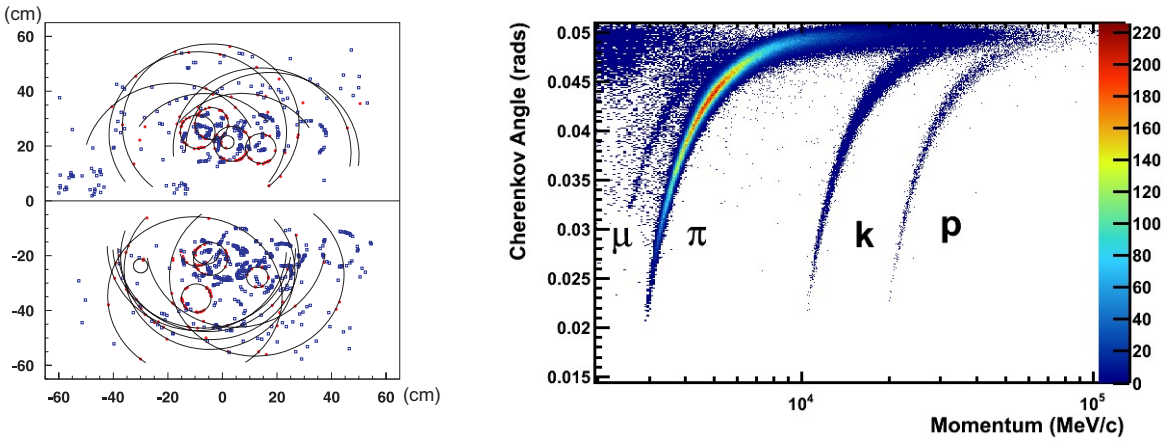
$$\cos\theta_{Ckv} = \frac{1}{n(\omega)\beta}. \tag{3.2}$$

The relative velocity $\beta = v_p/c$ denotes the fraction of particle speed divided by the speed of light in vacuum. This conventional Cherenkov radiation results from constructive interference of electromagnetic particle waves emitted by the polarised atoms in the medium. Their polarisation is caused by the charged particle crossing the medium. The effect was first observed in 1934 by P. A. Cherenkov [96, 97]. The corresponding theoretical background was established by I. Frank and I. Tamm in 1937 [98].
While the polar photon emission angle θ_{Ckv} is constant, the azimuthal angle is random. As a result, a light cone is formed by the Cherenkov photons. From Equation 3.2 and the condition $|\cos\theta| \leq 1$, it follows that there is a minimum velocity $\beta_{thr} = 1/n$, called Cherenkov threshold, below which Cherenkov radiation is not emitted. For large particle velocities $\beta \to 1$ the photon emission angle does not exceed θ_{Ckv}^{max}, which can be calculated from $\cos\theta_{Ckv}^{max} = 1/n$. θ_{Ckv}^{max} is called the saturation angle and depends on the particle mass and the refractive index of the medium. It sets the upper limit for Cherenkov detectors to distinguish particles since the difference in emission angle is no longer resolvable. These two particle velocity conditions, set by β_{thr} and $\beta \to 1$, define the momentum limits of particle identification in RICH detectors.
Another limiting factor is the photon emission efficiency for low-index refractive Cherenkov radiators, as required for sensitivity in the gigaelectronvolts regime. According to Frank and Tamm, the photon yield is proportional to $1 - \frac{1}{n^2\beta^2}$. As a result, RICH detectors with gas radiators of refractive indices close to 1 require large chambers to achieve a sufficient photon yield.

RICH-like particle identification systems measure the Cherenkov angle θ_{Ckv}. With this information, the particle velocity β can be determined, as indicated by equation 3.2. The refractive index of the radiation medium is known, and the particle momentum p is determined by a tracking system measuring the curvature of the track of a charged particle in a magnetic field. The track curvature additionally indicates the particle charge. Using

$$p = \gamma m \beta c, \tag{3.3}$$

where γ denotes the Lorentz factor, the particle mass m can be deduced and the particle identified.



(a) (b)

Fig. 3.1 Example of a simulated event in RICH 1 including rings to guide the eye (a). Particle identification performance of the reconstructed Cherenkov angle as a function of particle momentum (b). Source: [95] and [31], respectively.

Charged particles which fulfil the condition stated in Equation 3.1 emit Cherenkov photons along their tracks. These photons are reflected by spherical mirrors onto flat mirrors and then onto photon detectors. Besides focussing, this set of mirrors assures that the components of the optoelectronic read-out chain are located outside the acceptance of the highly energetic particle tracks. That has the benefit of reducing radiation damage to the electronics while at the same time avoiding influencing the particle track direction and energy measured in subsequent detector stages and also avoiding background photons generated by particles traversing the photon sensor windows.
Using the information of the tracking system, a likelihood is determined to formulate a particle hypothesis for each photon. A typical LHC Run 2 event in RICH 1 is shown in Figure 3.1a with circles fitted according to the particle hypothesis. With this information, the Cherenkov angle is reconstructed. In Figure 3.1b, one can see that Pions, Kaons, and Protons can be well distinguished before saturation.

3.2 Photonic crystals

Light propagation in photonic crystals is, to a large extent, analogous to electron behaviour in semiconductors. The electron flow is prevented in certain directions and domains due to the energy bandgap structure of the lattice featuring a periodic potential. In the case of photonic crystals, the periodic structure is not given by a potential but by a dielectric function or a periodic index of refraction. As a result, light propagation can be controlled by photonic

bandgaps, and if the light loss in the material is small enough, the reflection, refraction, and interference of photons allow for novel light phenomena.

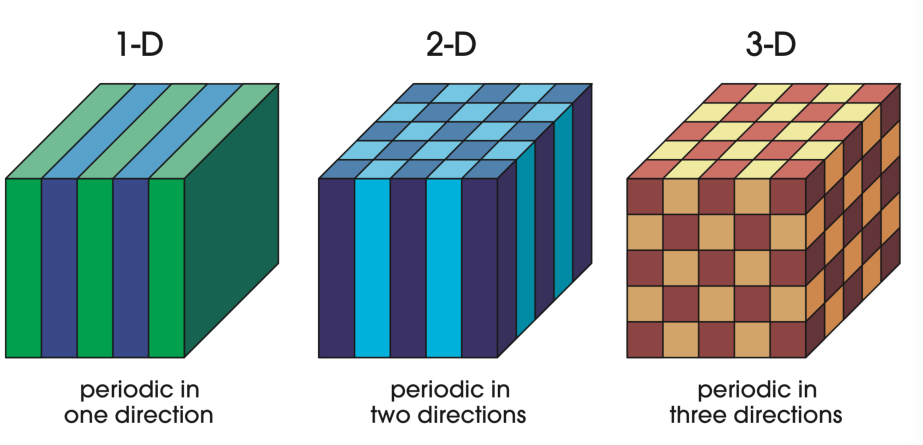


Fig. 3.2 Representation of 1-, 2-, and 3-dimensional photonic crystals. The colours illustrate different indices of refraction. Source: [99].

Figure 3.2 shows examples of 1-, 2-, and 3-dimensional periodicity of photonic crystals. A higher dimensional periodicity allows for more well-separated bandgaps and hence more control of the light propagation. However, the more complex the structure, the more complex the simulation of its physical properties and manufacturing becomes increasingly difficult.

3.2.1 Field properties

For the sake of conciseness, the mathematical background for the theory of light propagation in photonic crystals will only be outlined and follows the treatment given in [99].

Light propagation in a photonic crystal is a phenomenon of macroscopic electromagnetism and hence described by the corresponding Maxwell equations (in SI units):

$$\begin{aligned} \nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \\ \nabla \cdot \mathbf{D} = \rho, \quad \nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}. \end{aligned} \tag{3.4}$$

$\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic fields, $\mathbf{D}$ and $\mathbf{B}$ are the displacement and magnetic induction fields, respectively. ρ is the free charge and $\mathbf{J}$ the current density. If the medium is assumed to have no sources of light and does not change with time, and therefore also does not have any charges or currents, one can set $\mathbf{J} = 0$ and $\rho = 0$. Four more restrictions can be made. Firstly, the field strengths are assumed small. The higher-order terms of the

displacement field $\mathbf{D}$ vanish, when being expressed as a power series in the components of $\mathbf{E}$. Together with the second assumption, that the material is macroscopic and isotropic, the relation is simplified to $\mathbf{D}(\mathbf{r},\omega) = \varepsilon_0\varepsilon(\mathbf{r},\omega)\mathbf{E}(\mathbf{r},\omega)$, where ε_0 is the vacuum permittivity and $\varepsilon(\mathbf{r},\omega)$ the relative permittivity describing the properties of the dielectric. The dependence on the frequency, denoted by ω, can be omitted if material dispersion is not taken into account. In the case of lossless, transparent materials, $\varepsilon(\mathbf{r})$ is real and positive[1]. Analogously we can relate $\mathbf{B}$ to $\mathbf{H}$ by $\mathbf{B}(\mathbf{r}) = \mu_0\mu(\mathbf{r})\mathbf{H}(\mathbf{r})$, where μ_0 is the vacuum permeability. The relative magnetic permeability is further approximated to unity as is the case for most materials of interest. This allows relating ε to the refractive index n by $n = \sqrt{\varepsilon\mu}$.
Following these assumptions, the Maxwell equations 3.4 can be formulated as

$$\begin{aligned} \nabla\cdot\mathbf{H}(\mathbf{r},t) = 0, \qquad & \nabla\times\mathbf{E}(\mathbf{r},t) + \mu_0\frac{\partial\mathbf{H}(\mathbf{r},t)}{\partial t} = 0, \\ \nabla\cdot[\varepsilon(\mathbf{r})\,\mathbf{E}(\mathbf{r},t)] = 0, \quad & \nabla\times\mathbf{H}(\mathbf{r},t) - \varepsilon_0\varepsilon(\mathbf{r})\frac{\partial\mathbf{E}(\mathbf{r},t)}{\partial t} = 0. \end{aligned} \tag{3.5}$$

The time dependence can be separated from the spatial dependence by expanding $\mathbf{H}$ and $\mathbf{E}$ into harmonic modes:

$$\begin{aligned} \mathbf{H}(\mathbf{r},t) &= \mathbf{H}(\mathbf{r})\,e^{-i\omega t}, \\ \mathbf{E}(\mathbf{r},t) &= \mathbf{E}(\mathbf{r})\,e^{-i\omega t}. \end{aligned} \tag{3.6}$$

Using this decoupling of spatial and time dependence, the divergence equations of 3.5 require that the electromagnetic waves in the medium are *transverse* [2].
Additionally, the harmonic representation simplifies the relation of $\mathbf{E}$ and $\mathbf{H}$:

$$\begin{aligned} \nabla\times\mathbf{E}(\mathbf{r}) - i\omega\mu_0\mathbf{H}(\mathbf{r}) = 0, \\ \nabla\times\mathbf{H}(\mathbf{r}) + i\omega\varepsilon_0\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r}) = 0. \end{aligned} \tag{3.7}$$

This allows to decouple $\mathbf{E}$ from $\mathbf{H}$, making use of the relation $c = 1/\sqrt{\varepsilon_0\mu_0}$, and to arrive at the final result:

$$\nabla\times\left(\frac{1}{\varepsilon(\mathbf{r})}\nabla\times\mathbf{H}(\mathbf{r})\right) = \left(\frac{\omega}{c}\right)^2\mathbf{H}(\mathbf{r}). \tag{3.8}$$

This equation allows to derive, for a material with structure $\varepsilon(\mathbf{r})$, the modes $\mathbf{H}(\mathbf{r})$, that fulfil the transversality condition. From the first curl equation of 3.7 $\mathbf{E}(\mathbf{r})$ can be recovered.

[1] The complex components of the dielectric constant describe losses, while $\varepsilon(\mathbf{r}) < 0$ account for materials difficult to penetrate by light, as for example metals.

[2] Meaning that in the case of a plane wave $\mathbf{H}(\mathbf{r}) = \mathbf{a}\,\exp(i\mathbf{k}\cdot\mathbf{r})$, the wave vector $\mathbf{k}$ has to be orthogonal to $\mathbf{a}$, $\mathbf{a}\cdot\mathbf{k} = 0$.

Having now established the transversality constraint $\nabla \mathbf{H} = 0$ of the harmonic modes and how to derive the $\mathbf{H}$ and $\mathbf{E}$ field from a given structure $\varepsilon(\mathbf{r})$, another qualitative property can be derived from the Variational Principle. Equation 3.8 can be formulated as an eigenvalue problem by defining the operator $\hat{\Theta}$[3]:

$$\hat{\Theta}\mathbf{H}(\mathbf{r}) \coloneqq \nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r}) \right). \tag{3.9}$$

The smallest eigenvalue of Equation 3.8 is ω_0^2/c^2. This value corresponds to the lowest-frequency mode, which formulated as a variational problem, represents the field pattern minimising the functional:

$$\mathrm{U}_f(\mathbf{H}) \coloneqq \frac{\left(\mathbf{H}, \hat{\Theta}\mathbf{H}\right)}{(\mathbf{H}, \mathbf{H})}, \tag{3.10}$$

where $(\cdot, \cdot)$ denotes the inner product[4]. This functional can be formulated in terms of the electric field $\mathbf{E}$ using Equation 3.7:

$$\begin{aligned} \mathrm{U}_f(\mathbf{H}) &= \frac{(\nabla \times \mathbf{E}, \nabla \times \mathbf{E})}{(\mathbf{E}, \varepsilon(\mathbf{r})\mathbf{E})} \\ &= \frac{\int d^3\mathbf{r} \, |\nabla \times \mathbf{E}(\mathbf{r})|^2}{\int d^3\mathbf{r} \, \varepsilon(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2}. \end{aligned} \tag{3.11}$$

From this equation, we can derive the following helpful characterisation of the modes: in order to minimise U_f the electric field $\mathbf{E}$ has to be largest in the region of high dielectric constant ε.

3.2.2 Light propagation

Light propagates in photonic crystals as **Bloch modes**. That is in analogy to solid-state physics and due to the periodicity of the crystal lattice. Formulated as Bloch states, each value of the wave vector $\mathbf{k}$ defines an eigenstate of the operator $\hat{\Theta}$ and frequency $\omega(\mathbf{k})$. The corresponding eigenvector is then given by

$$\mathbf{H}_\mathbf{k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{u}_\mathbf{k}(\mathbf{r}). \tag{3.12}$$

The amplitude term $\mathbf{u}_\mathbf{k}(\mathbf{r})$ takes on the same periodicity as the lattice the wave is propagating through: $\mathbf{u}_\mathbf{k}(\mathbf{r}) = \mathbf{u}_\mathbf{k}(\mathbf{r} + \mathbf{R})$, where $\mathbf{R}$ is the lattice vector.

[3] Without proof, it is noted that $\hat{\Theta}$ is hermitian and its eigenvalues are real numbers.

[4] For two wave functions $\mathbf{F}(\mathbf{r})$ and $\mathbf{G}(\mathbf{r})$, $(\mathbf{F}, \mathbf{G}) \coloneqq \int d^3\mathbf{r} \mathbf{F}^*(\mathbf{r}) \cdot \mathbf{G}(\mathbf{r})$, "*" denoting complex conjugation.

To understand how band structures form inside a photonic crystal, it is simplest to consider a one-dimensional photonic crystal of alternating media with refractive indices n_1 and n_2 and spatial period a, as shown in Figure 3.3. No mode in the crystal is allowed to have a

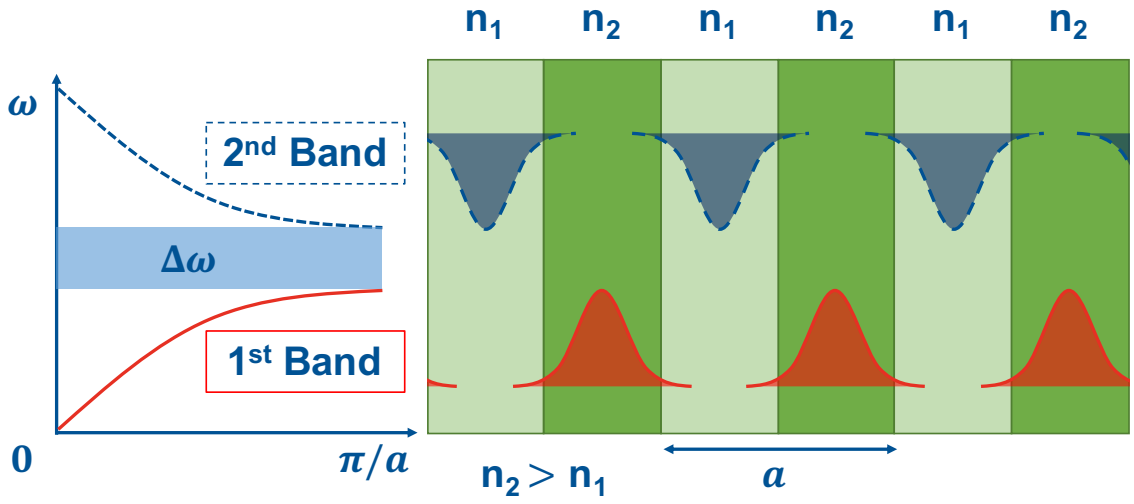


Fig. 3.3 Illustration of the band structure of the lowest two modes in a one-dimensional photonic crystal (left). Electric field energy density of the first (red) and second (blue) band.

frequency within the illustrated gap. That can be understood by considering the first two bands and taking into account the result that was obtained by using the variational principle in Equation 3.11: the low-frequency mode, here the first band, has the maximum of its energy in the region of high ε, and hence in this example the higher refractive index n_2, whereas the higher frequency modes concentrate their energy in the low ε region. The transversality constraint of the modes requires them to occupy different regions of ε.
The resulting bandgap is analogous to the conduction and valence band of semiconductors. The bandgap size can be increased by a larger difference between the refractive indices of the alternating media. A mathematical description of the two- and three-dimensional photonic crystal structures is omitted since it would go beyond the scope of this introduction, but the above-explained principles generally still hold.

3.3 Photonic crystals as effective Cherenkov radiators

This section outlines the potential of photonic crystals as effective Cherenkov radiators for particle identification in RICH-like experiments. The analytical simulation used to calculate the effective Cherenkov radiation of a one-dimensional photonic crystal is introduced.

3.3.1 The potential of photonic crystals PID in particle physics

The motivation to use photonic crystals as Cherenkov radiators arises from the limitations of conventional Cherenkov materials in the case of high particle energies. As explained in Section 3.1, the Cherenkov angle saturates for high velocities and hence does not allow distinguishing between particles in this regime. That is the case for typical transparent dielectrics. Additionally, the interface between air and dielectric results in total internal reflection for a relative permittivity $\varepsilon_r > 2$. This follows from the resulting inequality $\lim_{v\to c}\left(\frac{\varepsilon_r\omega^2}{c^2}-\frac{\omega^2}{v^2}\right)^{1/2} > \frac{\omega}{c}$ concerning Cherenkov radiation fields inside the dielectric medium for which Cherenkov radiation outside the medium disappears [100]. Consequently, to identify particles in the multi-gigaelectron regime, permittivities close to 1 are required, which limits conventional Cherenkov detectors to the use of large gas chambers.

The properties of photonic crystals rely heavily on the principle of bandgap structures, as explained in more detail in Section 3.2. These bandgaps allow restricting the direction of propagation of light for specific frequencies. This property can be used to create photonic crystals in which the transition radiation that is emitted when a charged particle crosses the boundary between different media (i.e. differing in their index of refraction) interferes constructively for a specific angle [101, 102]. The periodicity of the photonic crystal structure leads to resonance effects of the transition radiation waves emitted at each interface. This results in the emission of light at a certain angle similar to the conventional Cherenkov effect [103]. In contrast to conventional Cherenkov radiation, the effective Cherenkov radiation in photonic crystals relies on the intensity of the transition radiation and is hence not limited to the usual radiation angles (in the following, Cherenkov radiation refers to the above-described effect and not the conventional one, if not noted otherwise).
The flexibility of this effect becomes particularly evident considering that by using multilayer structures with suitable properties, also backward Cherenkov radiation can be created [104]. Tailoring Cherenkov angles to adapt to the requirements of high energy physics experiments is hence a logical consequence, as was proposed in 2018 in [100].

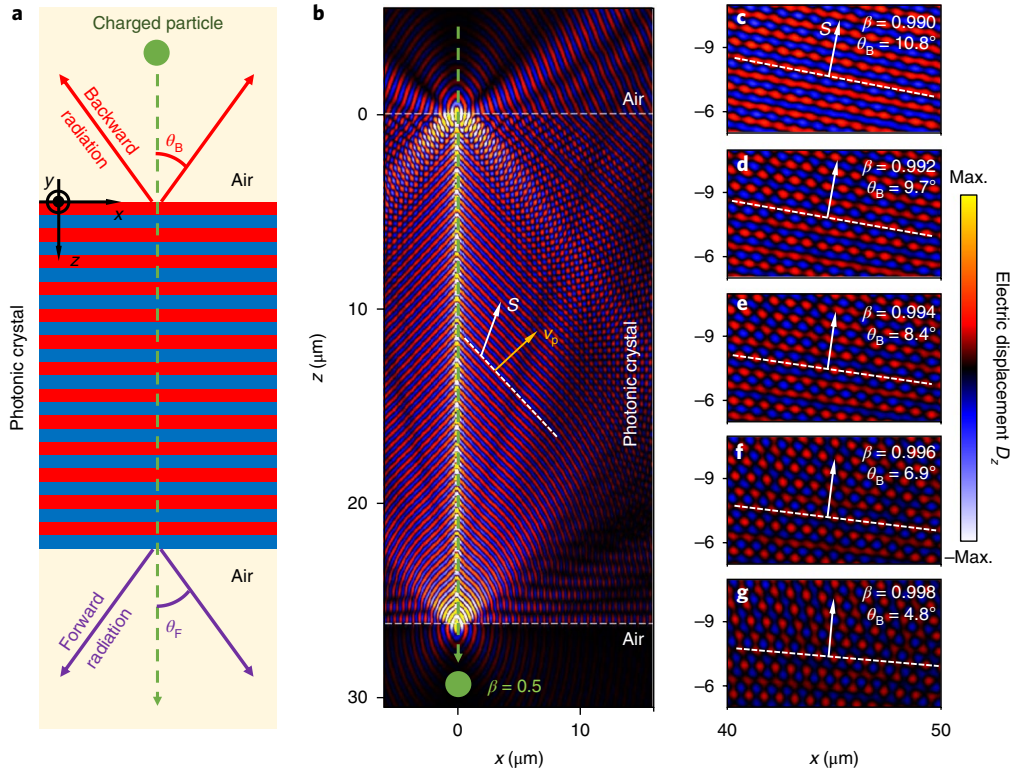
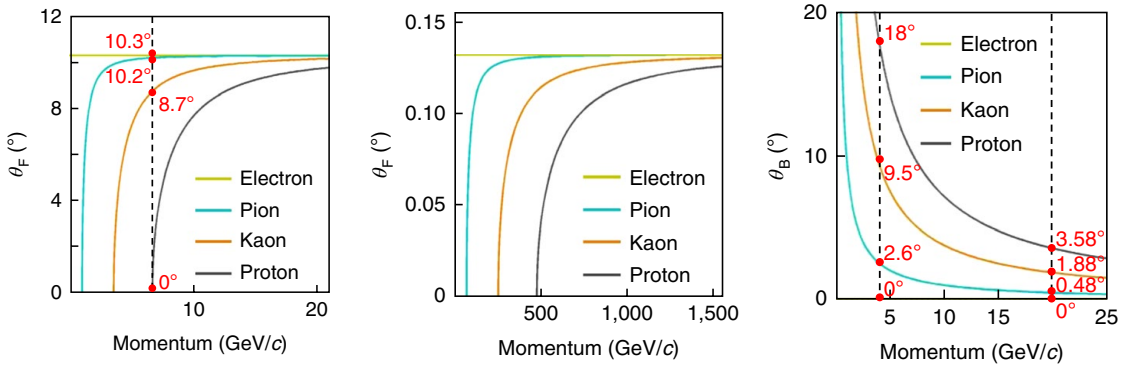


Fig. 3.4 Forward and backward Cherenkov radiation from a charged particle inside a one-dimensional photonic crystal (a). Cherenkov radiation field distribution in backward direction as a result of resonance transition radiation (b-g). Source: [100].



(a) (b) (c)

Fig. 3.5 Cherenkov angle versus particle momentum for a photonic crystal of 2 mm thickness and alternating layers of GaP ($\varepsilon_{GaP} = 10.6$) and SiO_2 ($\varepsilon_{SiO_2} = 2.1$). By varying the thickness of the dielectric layers[5], Cherenkov angle separation can be achieved in the moderate (a) and high momentum (b) regime, using a forward Cherenkov configuration, or in the low momentum regime using a backward Cherenkov configuration (c). Source: [100].

The resonance transition radiation can be best understood by a graphical illustration, as shown in Figure 3.4b. Figure 3.4a shows the corresponding schematic of a charged particle, here an electron, traversing a transparent periodic structure and creating Cherenkov radiation in both the backward and forward direction. In the case illustrated in Figure 3.4b the particle velocity of $\beta = 0.5$ is not large enough to create conventional Cherenkov radiation. Instead, the plane waves of transition radiation at each dielectric interface interfere constructively. For this particular case, more radiation energy is emitted in the backward direction, as indicated by the power flow Poynting vector **S**. Figures 3.4c – g indicate the angular sensitivity to the particle velocity. It is only when the Bloch modes couple to air constructively that effective Cherenkov radiation can be observed. If they interfere destructively, the effect is similar to conventional transition radiation.

The potential to use photonic crystals in high energy physics particle identification becomes apparent when looking at the angular separation of Cherenkov radiation, as shown in Figure 3.5. For a one-dimensional photonic crystal of 2 mm thickness with alternating layers of GaP ($\varepsilon_{GaP} = 10.6$) and SiO_2 ($\varepsilon_{SiO_2} = 2.1$), particles can be identified in various energy regimes by modifying the thickness of the individual dielectric layers[5]. Figures 3.5a–b show a forward Cherenkov configuration for the moderate and high momentum regime, respectively. The dashed line in Figure 3.5a demonstrates the different Cherenkov angles for $6.6\,\mathrm{GeV/c}$ particles, allowing to formulate a particle hypothesis. By adapting the length d_{unit} of the GaP/SiO_2 compound, one can perform particle identification in the regime $500\,\mathrm{GeV/c}$. A crystal of the same thickness and dielectric materials but with yet another configuration of their individual layer thickness can produce backward Cherenkov radiation allowing to distinguish between particles with a momentum of $4\,\mathrm{GeV/c}$. This is particularly relevant for particle identification in high energy physics experiments, such as LHCb. The momentum regime between 1 –$10\,\mathrm{GeV/c}$ can usually only be covered by materials such as silica aerogel, as was present in the LHCb RICH 1 detector until 2013 (Figure 1.7a). Optimisations of the overall LHCb performance resulted in an increase of the particle multiplicity causing light losses in the aerogel due to scattering.

A particle identification detector using photonic crystals as a radiator would solve many of the constraints of the current generation of conventional Cherenkov detectors. One can even think of a detector using a combination of backward and forward Cherenkov radiators [100]. The range of energy for particle identification could be increased, and the size of the detector

[5] Layer composition in 3.5a: $d_{unit} = 1.0205\lambda$, $d_{GaP} = 0.3\,d_{unit}$, $d_{SiO_2} = 0.7\,d_{unit}$, for $\lambda = 700$ nm.
Layer composition in 3.5b: $d_{unit} = 1.0144\lambda$, $d_{GaP} = 0.3\,d_{unit}$, $d_{SiO_2} = 0.7\,d_{unit}$, for $\lambda = 700$ nm.
Layer composition in 3.5c: $d_{unit} = 0.2792\lambda$, $d_{GaP} = 0.6\,d_{unit}$, $d_{SiO_2} = 0.4\,d_{unit}$, for $\lambda = 700$ nm.

significantly reduced. As a result, the subsequent detector stages would move closer to the interaction point and decrease in size, reducing their costs significantly.

Several nanophotonics-based Cherenkov experiments have found evidence for controlling effective Cherenkov radiation through metamaterials and the feasibility for particle detection applications has been suggested [105–111]. However, so far no experiments have been performed studying backward Cherenkov radiation in the optical spectrum of an all-dielectric photonic crystal and in conditions of a RICH-like experiment, as provided by a multi-megaelectronvolt particle beam.

3.3.2 Resonance transition radiation simulation

The simulations of the effective Cherenkov radiation presented in this thesis rely on a framework of analytical calculations developed by Lin et al., with whom I had frequent discussions about my work, providing insight into their model [112]. The calculations expand on Ginzburg and Frank's transition radiation theory within macroscopic electrodynamics, allowing to analytically derive the radiated fields and spectrum of the emitted photons. They are explained in detail in [113, 100], and full credit regarding the analytical model remains with its authors.

In this thesis, two properties of the effective Cherenkov radiation are simulated: the angular spectral energy density (ASED) and the energy spectrum of emitted photons as a function of wavelength. The ASED represents the distribution of the backward radiation in angle and frequency and is defined in the supplementary material section S7 of Reference [113], given by:

$$U_1(\omega,\theta) := \frac{\varepsilon_{1r}^{3/2} q^2 cos^2\theta}{4\pi^3 \varepsilon_0 c sin^2\theta} |a_1|^2, \tag{3.13}$$

where ε_{1r} is the relative permittivity of medium 1 into which the energy of the particle is radiated (backwards with respect to its motion), ε_0 the permittivity of free space, q the particle charge, c the speed of light in vacuum, and a_1 the amplitude of the radiation field in medium 1 (see equation E7 in the supplementary material of Reference [113]).
The total energy radiated by the particle into medium 1 can be obtained by integrating the emitted field energy density over space:

$$W_1 = \int_0^{+\infty} W_1(\omega) d\omega \tag{3.14}$$

$$= \int_0^{+\infty} \int_0^{\pi/2} U_1(\omega,\theta) \cdot (2\pi \sin\theta) d\theta d\omega \tag{3.15}$$

$W_1(\omega)$ represents the resonance transition radiation energy spectrum in the backwards direction:

$$W_1(\omega) = \int_0^{\pi/2} U_1(\omega,\theta)\cdot(2\pi\sin\theta)d\theta \tag{3.16}$$

A significant purpose of the experiments presented in this thesis is to test the predictive power of the analytical framework to predict effective Cherenkov radiation from resonance transition radiation.

3.4 Polymer photonic crystal laboratory measurements

A one-dimensional polymer photonic crystal is studied in a laboratory setup, using a Strontium-90 radiation source. The sample is described together with its expected properties as predicted by the simulation. The laboratory setup and the results of the measurements are presented. To test the radiation robustness of the polymer compound, it is irradiated in a high-energy beam.

3.4.1 Experimental setup

The experimental setup design is inspired by the RICH detector layout, featuring a mirror that focusses and reflects the light onto the photodetection plane. Figure 3.6 shows the GEANT4 [6] [52, 53] simulation and an enlarged view of the region where the photonic crystal is positioned. Here, photon tracks are shown in green, the electron track in red.

Electrons are emitted from a Strontium-90 source via beta-decay:

$$^{90}\mathrm{Sr} \xrightarrow{\beta_1\text{: 0.546 MeV/c}} {}^{90}\mathrm{Yr} \xrightarrow{\beta_2\text{: 2.28 MeV/c}} {}^{90}\mathrm{Zr}$$

The decay energy spectrum is continuous with the maximum at $\beta_1 = 0.546$ MeV/c from the decay to Yttrium-90. The spectrum extends up to $2.28\,\mathrm{MeV}/c$ due to the following decay to Zirconium-90. This energy spectrum is important to bear in mind since the Cherenkov angle is a function of the energy for the tested samples and is hence responsible for the radial width of the Cherenkov ring.

The photonic crystal is mounted close to the source on a plastic holder that also serves as a collimator for the electrons. The collimator can be exchanged, and hence its aperture varied. For the presented results, a diameter of 3 mm is used.

[6] Simulation toolkit for the interaction of particles with matter.

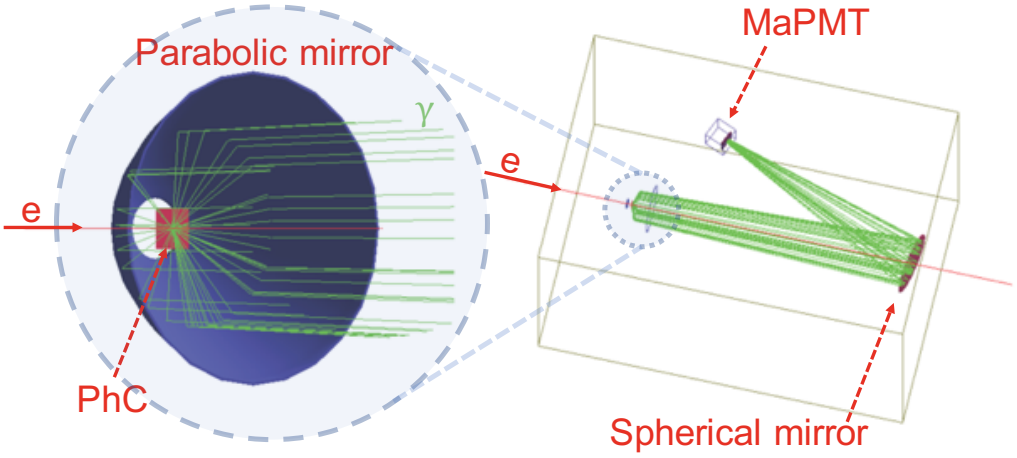


Fig. 3.6 GEANT4 simulation of the experimental setup with enlarged view of the parabolic mirror in which the photonic crystal is located. Photons are shown in green, the track of the beta-particle in red.

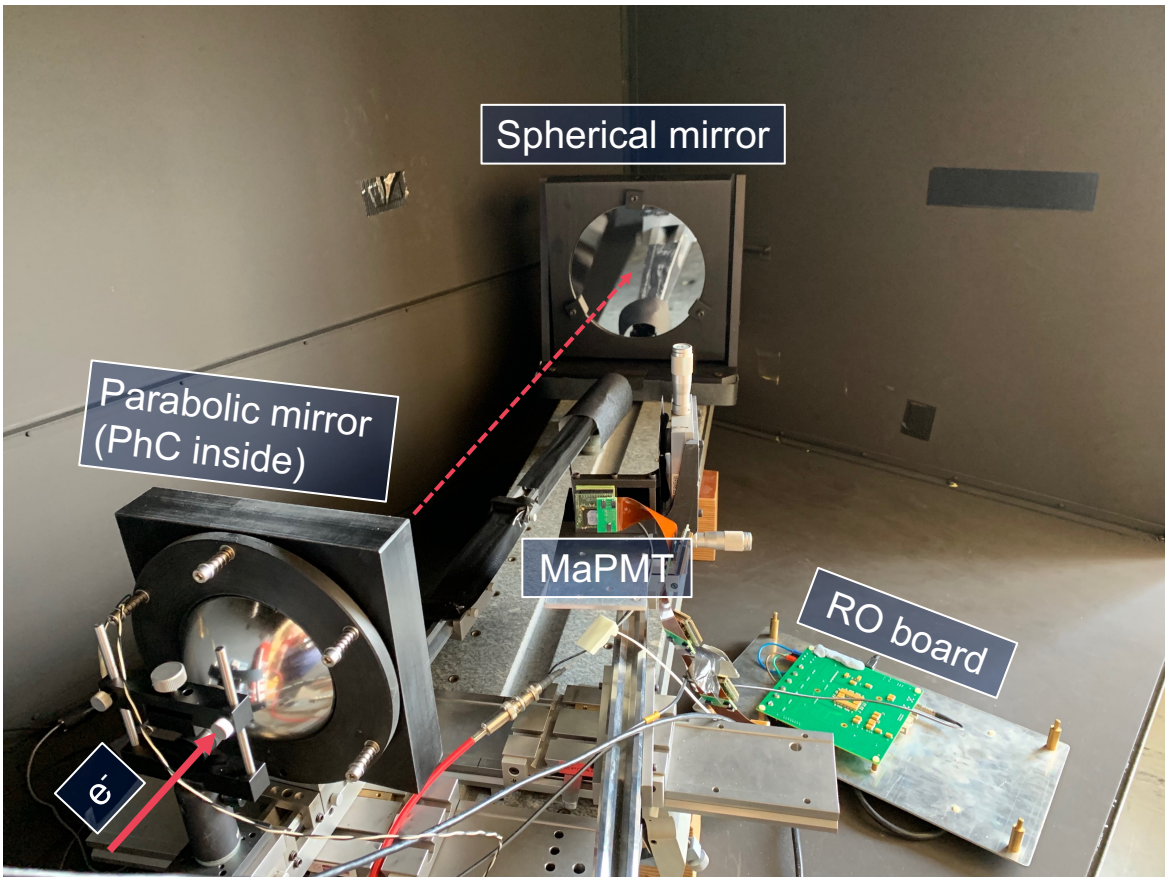


Fig. 3.7 setup to test photonic crystal samples inside a light-tight box showing the principal components. The photonic crystal is located in the centre of the parabolic mirror. The electron path is illustrated by a red dashed line.

Figure 3.7 shows a picture of the setup as installed inside a light-tight box. A Strontium-90 electron source is positioned on an optical bench inside a light-tight box. The source is fixed inside a self-shielding container, respecting standard precautions regarding radiation safety. Electrons traverse the photonic crystal, which is attached to a collimator downstream of the source. A trigger scintillator is installed in the electron beam to send a signal to the read-out board. Photons are reflected from the parabolic mirror onto the spherical mirror, focussing the light onto the MaPMT photon detector.

The setup uses an adapted data acquisition system that was originally designed for the ALFA experiment [114]. The 64 channel 1 inch $\times$ 1 inch multi-anode photomultiplier tube[7] (MaPMT) is connected to a custom-built voltage divider, a passive, and an active board. The passive board routes the signal from the MaPMT pins to the active board, which uses a MAROC 2 read-out chip [115]. In parallel for all channels, the MAROC chip passes the MaPMT signals through slow shaper circuits, and the peaks of the shaped signals are captured by sample-and-hold circuits in response to an external trigger. The sampled peaks of all channels are then output through an analogue multiplexer. The signal is then transferred to the read-out board, where the analogue-to-digital (ADC) converter is located. When the read-out board receives a signal from the trigger scintillator, all channels of the PMT are read-out.

1D polymer photonic crystal

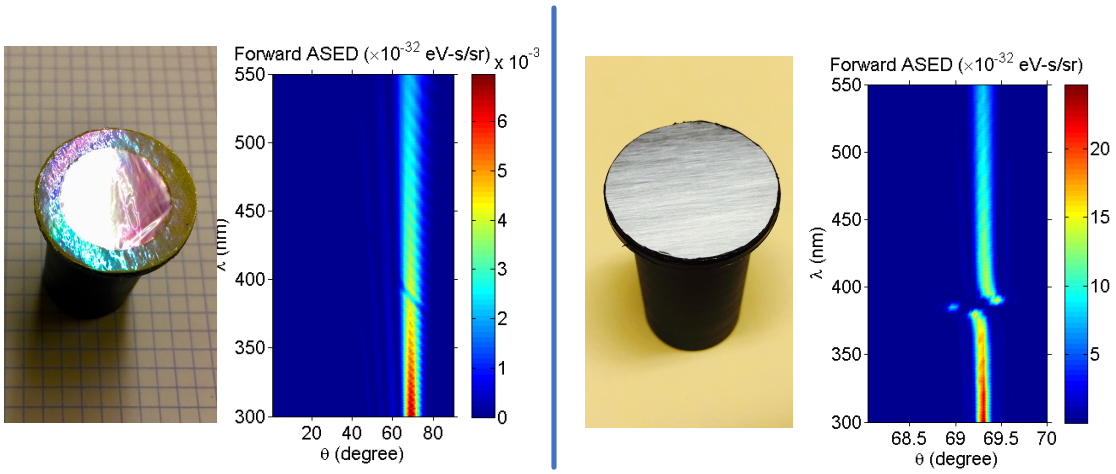
Two types of photonic crystal samples were tested, which are shown attached to the collimator in Figure 3.8. The first sample (sample A) consists of in total 16 layers and the second sample (sample B) of 1024 layers of alternating PVDF and PET[8]. The use of polymers for photonic crystal applications has been suggested due to the large dielectric difference that can be achieved with inorganic compounds [116]. Another benefit of polymer samples is their cost-effectiveness. The polymer samples were acquired from Peak Nano [9].

Next to the samples shown in the figure, one can see their respective simulated forward angular spectral energy density (simulated by Lin et al. [117]), as described in Section 3.3.2. For both samples, one can note that the angle is consistent independent of the wavelength, which enhances the probability to observe a Cherenkov ring without the necessity of a wavelength bandpass filter. That is not a general feature of photonic crystals but of these particular types of samples designed to avoid a chromatic error. One can also note in

[7] Manufactured by Hamamatsu (http://www.hamamatsu.com)

[8] Polyvinylidene fluoride ($n_{PVDF} = 1.415$) & Polyethylene terephthalate ($n_{PET} = 1.567$).

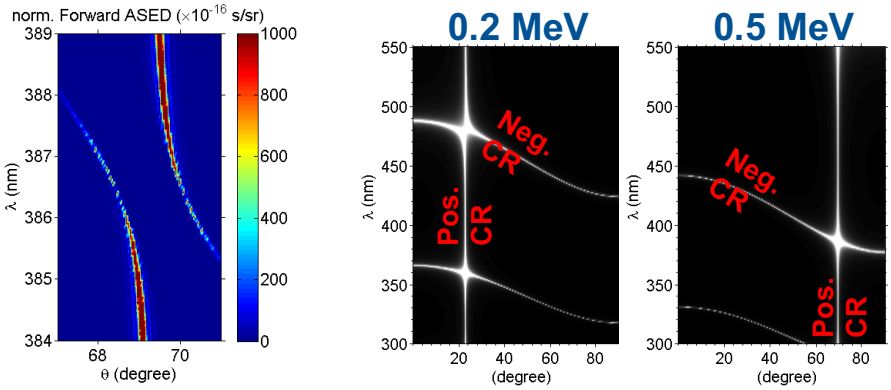
[9] Peak Nano (formerly Polymer Plus LLC): `https://www.peaknano.com/`



(a) (b)

Fig. 3.8 Photonic crystal samples (16 layer sample (a) and 1024 layer sample (b)) and their respective forward Cherenkov emission angle as a function of the wavelength for 5 MeV/c electrons. Plots produced by Lin et al. [117].

Figure 3.8 that there is a gap in the band. This anti-crossing effect is shown in more detail for sample B in Figure 3.9a. It can be attributed to the interaction between two modes with negative and positive group velocity (i.e. negative and positive Cherenkov radiation) [118]. In Figure 3.9b the contribution of backward Cherenkov radiation for sample B is visible for different electron energies. These two characteristics could potentially be investigated later to gain further insights into the resonance transition radiation Cherenkov effect and examine the accuracy of the simulations.



(a) (b)

Fig. 3.9 Features of sample B of 1024 layers and 0.5 MeV/c electron energy. Detailed view of anti-crossing feature in photonic crystal sample (a). Forward and backward contribution of Cherenkov radiation for varying electron energies (b). Plots produced by Lin et al. [117].

GEANT4 simulation

The experimental setup is described using the GEANT4 toolkit, which allows simulating the interaction of particles with matter. A picture of the simulation is shown in Figure 3.6. A 0.5 MeV/c electron beam is crossing a light-tight box, filled with air, in which the components of the experimental setup are located, including the parabolic mirror, the photonic crystal, the spherical mirror, and the 1×1 inch^2 MaPMT. Using the GEANT4 toolkit allows simulating the mirror reflections and the MaPMT photon conversion process. The latter is adapted from the LHCb RICH Upgrade beam test simulation [119]. The resonance transition radiation Cherenkov effect is not implemented in GEANT4. Hence, a Cherenkov radiation material is used to feature a refractive index corresponding to the expected effective refractive index of the photonic crystal. As a result, Cherenkov light is emitted at the angle expected from the resonance transition radiation simulations, while at the same time, effects such as scintillation and conventional transition radiation are simulated. Figure 3.10 shows the

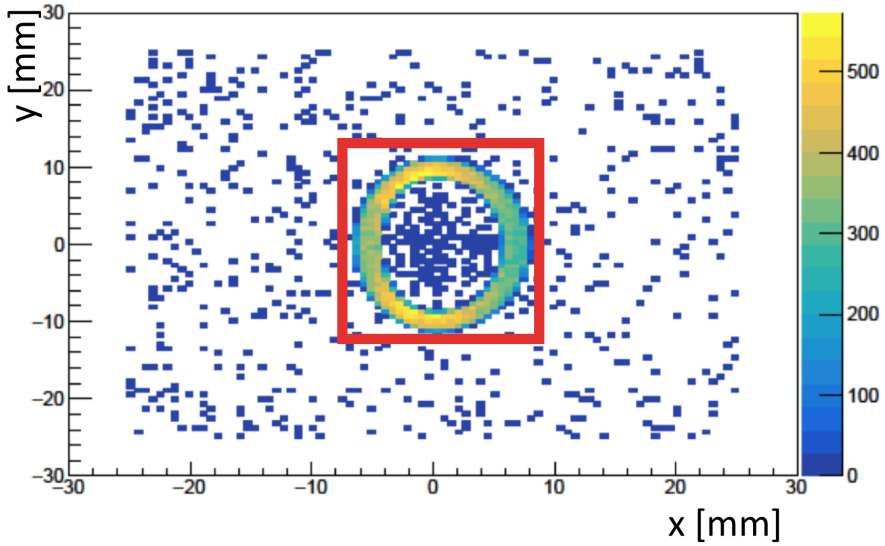


Fig. 3.10 Cherenkov light signal simulated using GEANT4. A 0.5 MeV/c electron traverses a Cherenkov material with a refractive index such that the Cherenkov angle matches the angle expected from the simulations. This allows checking the positioning of the optical components and to estimate the background from scintillation or transition radiation. The red square indicates the size of the MaPMT.

signal as obtained by the MaPMT in the GEANT4 simulation. It represents an idealised scenario in terms of electron energy and pixel size. The signal shown here is produced by an electron beam with 0.5 MeV/c and no energy spread, and much finer binning than the actual 8×8 channels of the MaPMT. That allows estimating the optimal positioning of the optical components and the location of the photonic crystal with respect to the parabolic

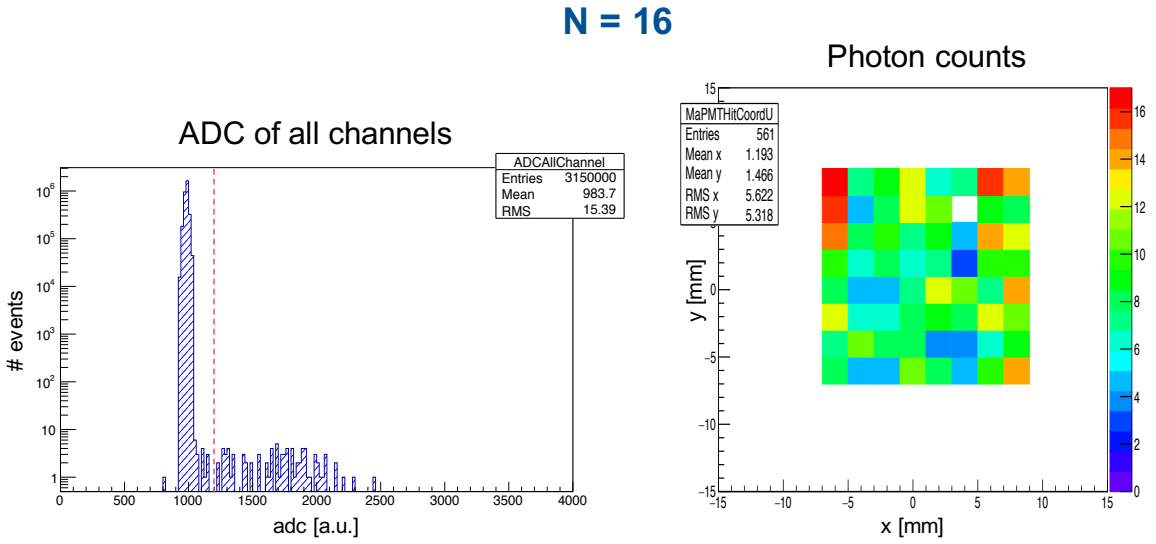
mirror, while at the same time, background effects of scintillation and transition radiation can be evaluated. The low background is due to the fact that the majority of background photons are assumed to be internally reflected in the photonic crystal and exit it from the sides. To avoid background from this internally reflected light, the outer edges of the photonic crystal sample are covered by a black tape in the experimental setup.

3.4.2 Laboratory measurements and results

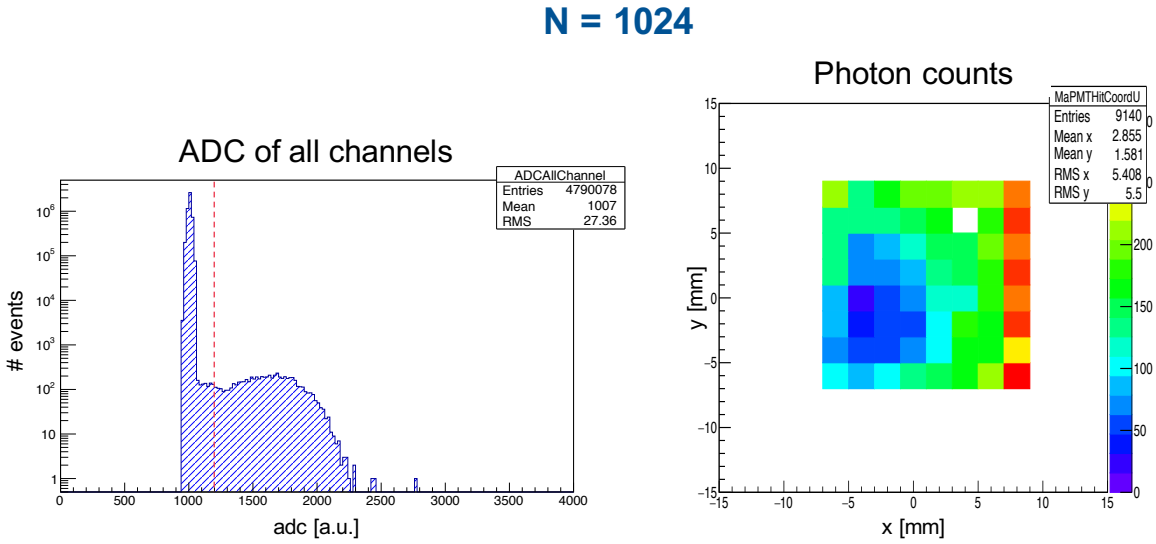
Figures 3.11a and 3.11b show examples of data collected with the 16 and the 1024 layer sample, respectively. On the left, a histogram of the signal strength in arbitrary units is shown. The pedestal of events corresponding to zero-incident photons exceeds the photon peak on its right by several orders of magnitude. The high probability of detecting zero photons in an event indicates that the peak on the right side of the pedestal can almost exclusively be attributed to single-photon events. The shape of the single-photon peak is much more pronounced and features far more events in the case of the 1024 layer sample in Figure 3.11b. While the events are homogeneously distributed in Figure 3.11a, in Figure 3.11b, evidence of a circular pattern is visible that might be interpreted as part of a ring with a low hit-density region in the centre. One can draw several conclusions from the presented measurements. The low number of hits produced with sample A with only a few layers of dielectric material agrees with the conventional transition radiation effect regarding the fact that a higher number of optical interfaces yields a higher number of photons [101]. The periodicity of the photonic crystal suggests that layer number and photon yield might scale linearly. However, this would only hold true in the case of a perfectly transparent material. One can see from the picture of sample B in Figure 3.8b that this is not the case in the visible optical wavelength range. The MaPMT is most sensitive in the *red* part of the visible spectrum. Hence one can not expect the photon yield to scale linearly for sample B with respect to its number of layers in comparison to sample A.

Even though the properties of sample A do not allow to see a Cherenkov ring pattern, the optimal positioning of the optical components and characteristics of the data acquisition system, such as the trigger delay, can be deduced from the number of events in a given time: a higher number of events corresponds to a better focussing of the photons created by the sample onto the MaPMT. The locations of the optical components agree with the ones set in the GEANT4 simulation to obtain a signal as shown in Figure 3.10 for an effective Cherenkov angle of 69°, within the precision of mechanical positioning.

With this characterisation of the setup, only minimal changes in the positioning of the photonic crystal and the photomultiplier are required to adapt to the second sample. A ring



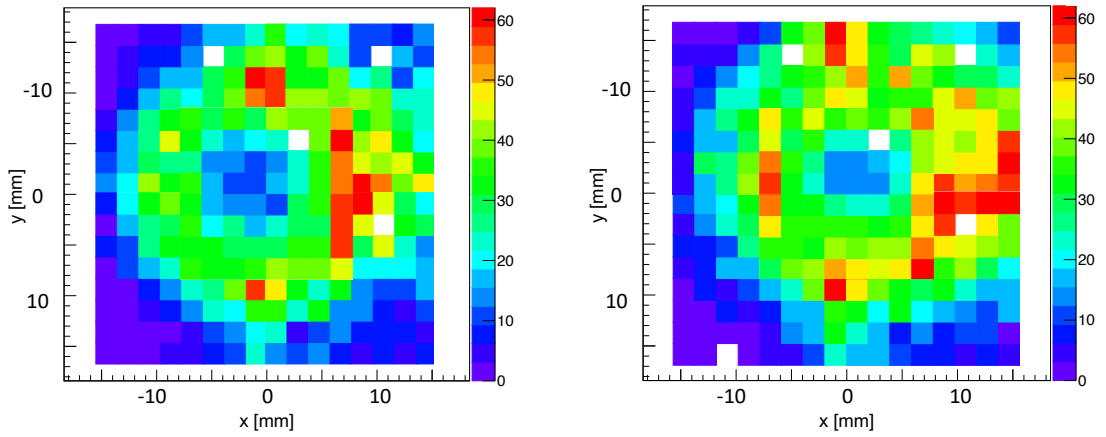
(a)



(b)

Fig. 3.11 Example data collected using the 16 (top) and 1024 (bottom) layer photonic crystal. The histograms on the left show the electron charge from the MaPMT in arbitrary units. The pedestal peak caused by the charge of events with mostly zero photons dominates the histogram by several orders of magnitude. On the right side of the pedestal, the single-photon peak is visible. Only events above a threshold of 1200 are used for the two-dimensional hitmaps, shown on the right. While for the 16 layer sample the light is distributed on the MaPMT without a noticeable pattern, a circular shape is visible for the 1024 layer sample.

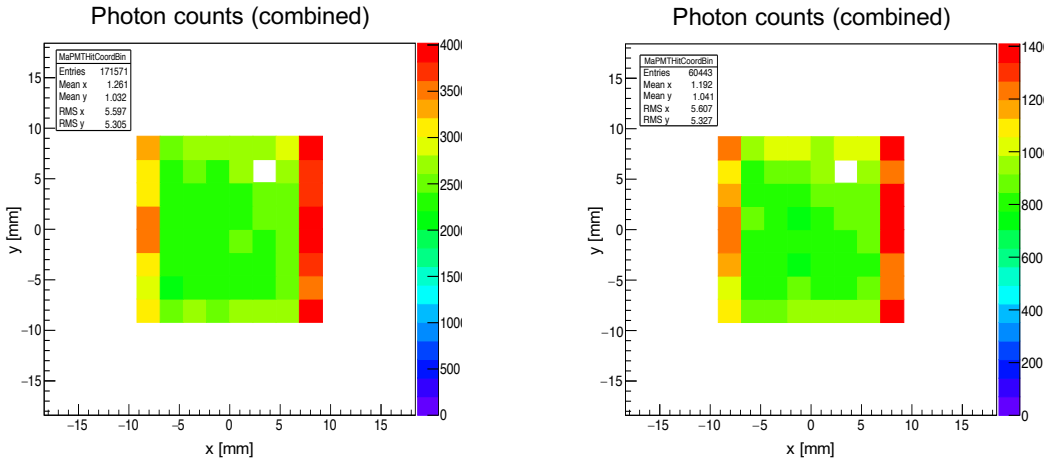
shape is visible for sample B. By moving the MaPMT while leaving all other parameters unchanged, the entire ring can be captured, as shown in Figure 3.12. Figure 3.12 shows two separate measurements of a ring shape, each consisting of five MaPMT hitmaps. The hitmaps are superimposed at the correct relative locations corresponding to the physical MaPMT positions. For the second measurement, shown in Figure 3.12b, the position of the photonic crystal inside the parabolic mirror was moved by a millimetre, slightly changing the focal position and broadening the ring pattern.



(a) (b)

Fig. 3.12 Two-dimensional hit-maps, superimposed corresponding to the respective position of the MaPMT for two sets of measurements.

To test the hypothesis that the observed ring is caused by background scintillation light, one can use a material producing only scintillation light. In case that the observed light is due to scintillation, this test should yield a similar result. Two example MaPMT plots of such a study are shown in Figure 3.13, where the MaPMT is in the same position as for Figure 3.11b. Various positions of the plastic sample inside the parabolic mirror were examined. In Figure 3.13a the position is similar to Figure 3.11b, in Figure 3.13b it is arbitrarily shifted. One can see that light is smoothly distributed on the PMT apart from the sides, which can be attributed to the non-uniform response of the MaPMT pixels. One can also note a non-responsive pixel in the top-right corner of the MaPMT.
An empty region in the centre was not observed for any position of the plastic sample, supporting the hypothesis that scintillation is unlikely to create a ring pattern.



(a) (b)

Fig. 3.13 Two example histograms using scintillating plastic instead of the photonic crystal to determine whether scintillation light can produce a similar pattern as the photonic crystal samples due to the focussing of the mirrors. No ring pattern was found for all tested positions of the plastic sample.

Laboratory results interpretation

The laboratory measurements suggest a good geometrical agreement between the behaviour expected from the resonance transition radiation simulations and the experimental results for sample B. Sample A does not yield enough photons to generate a Cherenkov ring distinguishable from the background noise of the MaPMT with the available statistics. Other properties of sample A, such as variable flatness due to its thinness, might also be responsible for not observing a pattern.

The observed ring may be interpreted as a potential Cherenkov ring. That is taking into account possible background effects that might create a similar pattern. Conventional transition radiation is emitted mainly at low angles with respect to the particle trajectory and hence not integrated by the parabolic mirror. This effect is considered negligible. Scintillation is delayed due to the time span the excited molecule requires to fall back onto its original energy by emitting a photon. This time delay is, however, within the length of the data acquisition gate and hence can emit photons reaching the MaPMT in the same read-out window as potential Cherenkov photons. However, as seen in Figure 3.13, scintillation light causes a smooth distribution with no low-hit density region in the centre. In the case of the photonic crystal samples, scintillation photons are expected to be internally reflected and absorbed by the black tape attached to the outer part of the sample. GEANT4 simulation studies using scintillation materials instead of the photonic crystal can provide further insight

into possible background effects such as propagation of particles through materials that have not been anticipated in the setup.

The fact that the ring pattern is susceptible to the position of the sample inside the mirror and can be seen only for a sample position within a range of 0.2 mm with respect to the parabolic mirror also suggests that light is emitted from the sample within a narrow range of angles, consistent with the Cherenkov radiation hypothesis rather than a scintillation effect. The width of the ring, broader than expected from the GEANT4 simulation, may be attributed to the electron energy range emitted by the ^{90}Sr source.

3.4.3 Polymer photonic crystal irradiation

The polymer photonic crystal was configured for 0.5 MeV electrons. To test such a polymer photonic crystal in a high-energy particle beam, not only a new configuration of the layer thicknesses would be required, but it would also have to withstand the radiation damage of the beam. In this section, the polymer photonic crystal is subjected to an 200 MeV electron beam at varying charges to investigate its radiation robustness.

The CLEAR facility at CERN

The CERN Linear Electron Accelerator for Research (CLEAR) is an electron linear accelerator facility, independent of the PS, SPS, and LHC beam-line complex [120, 121]. The CLEAR acceleration structure and in-air test stand are shown in Figure 3.14.

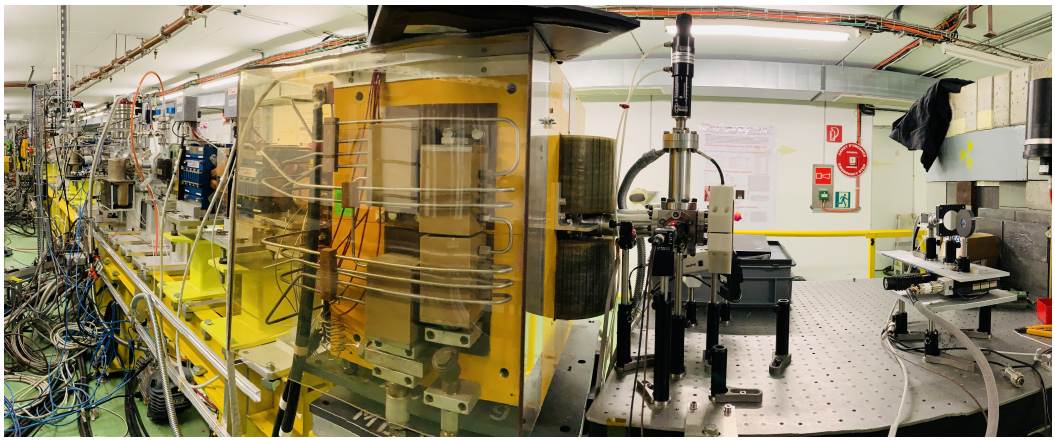


Fig. 3.14 Panoramic picture of the CLEAR beam line (left) with in-air test stand (right).

The electrons are extracted by photoelectric emission using a laser and accelerated to between 60 and 220 MeV energy. The laser mechanism allows selecting single electron bunches of

1 to 4 ps length and trains of up to 200 bunches. A bunch can carry a charge in the range of between 10 pC to 2 nC, the beam charge may be varied with the number of accelerated bunches. The beam can be operated at a repetition rate between 0.83 and 10 Hz, focussed to a width with a Gaussian spread of about $\sigma_b = 1\,\text{mm}$.

Irradiation setup

To irradiate separate sections of the polymer with different exposures, without entering the beam area, the photonic crystal is mounted on a translation stage orthogonal to the beam on the in-air test stand. The beam size is set to $\sigma_b = 5\,\text{mm}$ and the sample is irradiated at five positions with exponentially increasing charge, at $0.1\,\mu\text{C}$, $1\,\mu\text{C}$, $10\,\mu\text{C}$, $100\,\mu\text{C}$, and 1 mC.

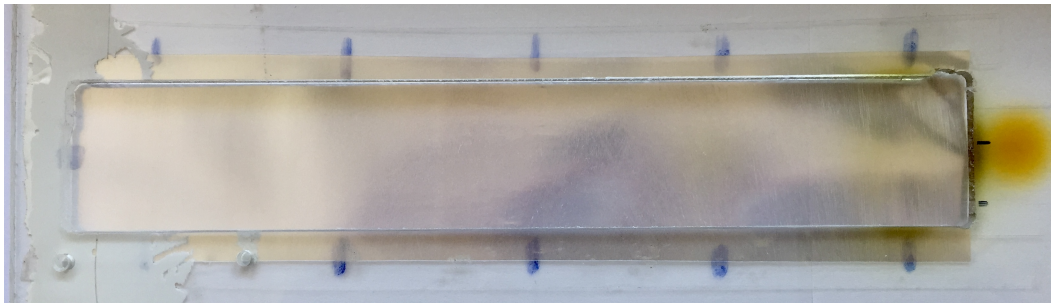


Fig. 3.15 Picture of the irradiated polymer photonic crystal. The sample was irradiated at five spots, marked by blue lines, with increasing deposited charge from left to right.

A picture of the irradiated sample is visible in Figure 3.15. The irradiation points are marked by blue lines, with the deposited charge increasing from left to right. A slight colour change is visible for $100\,\mu\text{C}$ and significant yellowing for 1 mC, indicating polymerisation of the material.

To irradiate a larger surface area, the beam was set to a width of $\sigma_b = 5\,\text{mm}$. To measure a sharp resonance transition radiation Cherenkov signal, the beam would have to be focussed. A size of $\sigma_b = 1\text{mm}$ would yield an area-factor 25 in radiation damage within the same time-span of the irradiation test.
Since a beam test experiment may by far exceed accumulated charge depositions of 1 mC, a polymer photonic crystal, as tested in the laboratory setup, is not suited for testing in a high-energy particle beam. The polymer material can neither withstand the radiation dose of a high-energy experiment, such as LHCb, nor that of a beam test campaign.

3.5 Manufacturing a high-energy photonic crystal

As analytically demonstrated by Lin et al. [100], the angle of emission of effective Cherenkov radiation can be adapted in a wide range, even in the backward direction. The study aims to take the first steps in demonstrating backward radiation by designing an apparatus at the CERN beam facilities, simulating the expected behaviour and comparing it to the measurement results.
It is important to note that an important objective is to gather experience about the photonic crystal manufacturing process for future project expansions. While a sample could be ordered from a manufacturer, probably featuring a higher production precision, in-house expertise has the benefit of considerable cost savings and a much shorter lead-time between a simulation concept and beam testing. The manufacturing process should therefore be considered as an R&D undertaking.

The project can be divided into three parts. The first part concerns identifying materials and a configuration by simulation that is expected to generate backward effective Cherenkov radiation for a 200 MeV electron beam. In the second step, a sample is manufactured according to these specifications. Finally, the manufactured photonic crystal is studied in a high-energy beam, conducted at the CLEAR facility at CERN. The beam test will be described in detail in Section 3.6.

3.5.1 Simulated $SiO_2 \,|\, TiO_2$ effective Cherenkov radiation

Ideal materials for a photonic crystal suited for a high-energy beam environment are resistant to the effects of ionising radiation, optically transparent, and feature a large difference in their reflective indices. The two materials chosen for this study are silicon-dioxide (SiO_2) and titanium-dioxide (TiO_2), with $n_{SiO_2} = 1.47$ and $n_{TiO_2} = 2.81$ at 450 nm [122], due to their compatibility with the deposition process. Their refractive index contrast makes them suitable for optical thin-film applications.

Due to the thinness of the photonic layers, they must be deposited on a substrate for mechanical stability. A reflective chromium layer is applied to shield any potential signal generated in the photonic crystal from the background produced in the comparatively large substrate. The thickness of the reflective layer is chosen as thin as possible to avoid adverse adhesion effects while reducing photon transmission to below 1 % (see Figure 3.24).

The properties of a multi-layer $SiO_2 \,|\, TiO_2$ structure are simulated.The approach extends Ginzburg and Frank's transition radiation theory [101, 102] to a one-dimensional photonic crystal by analytical calculation [100].

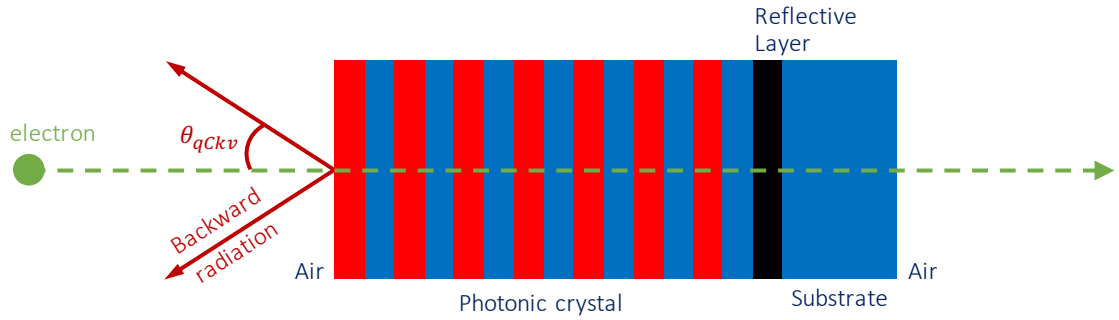
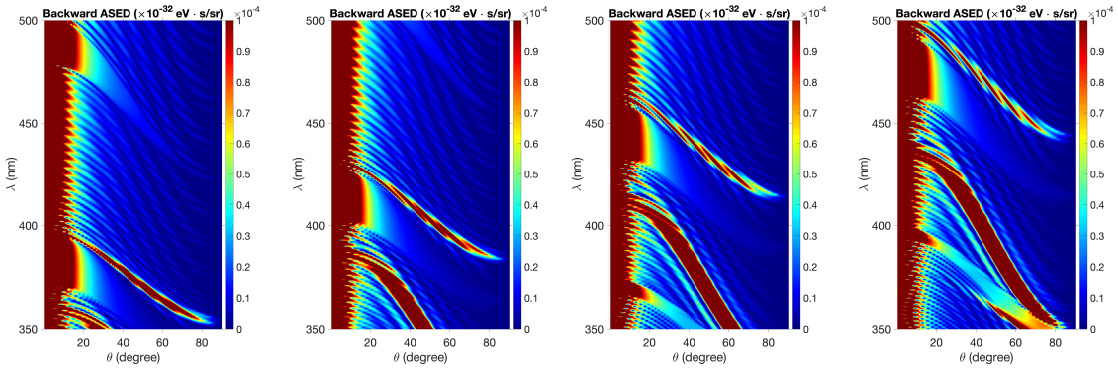
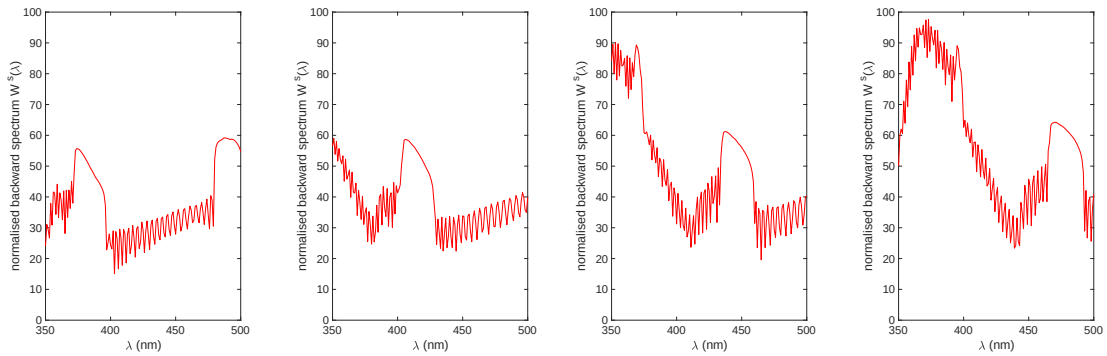


Fig. 3.16 Schematic of the 1D photonic crystal structure (not to scale). Electrons traversing the photonic crystal generate resonance transition radiation at the layer interfaces, of which only photons at a distinct angle couple to air in the backward direction with respect to the particle beam. A reflective layer isolates the photonic crystal from the background radiation generated in the substrate.



(a) 200 nm (b) 220 nm (c) 240 nm (d) 260 nm

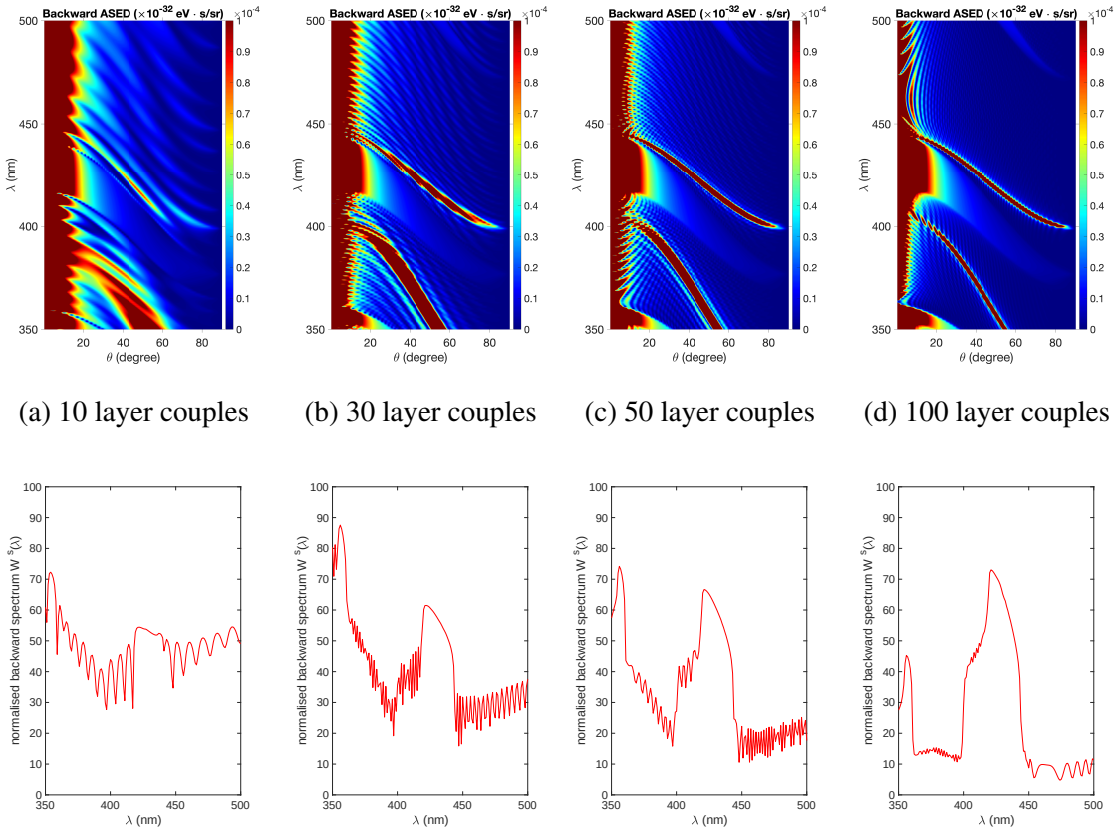


(e) 200 nm (f) 220 nm (g) 240 nm (h) 260 nm

Fig. 3.17 Angular spectral energy density (top) and normalised backward spectrum (bottom) for 25 layer couple SiO_2 | TiO_2 photonic crystals. The layer thickness is indicated below the figure.

Figure 3.17 shows the signal calculated by the model for a 200 MeV electron traversing a photonic crystal with 25 layer couples of $SiO_2 \mid TiO_2$, in addition to the chromium layer, as illustrated in the schematic in Figure 3.16. The effect of different global layer thicknesses is shown. Meaning, that each individual SiO_2 and TiO_2 layer has a thickness of either 200 nm, 220 nm, 240 nm, or 260 nm.
Figure 3.17a to 3.17d show the backward angular spectral energy density (ASED), as described in Section 3.3.2. The radiation distribution is displayed for angles between 0 and 90 degrees on the x-axis and wavelengths from 300 to 500 nm. Figure 3.17e to 3.17h show the corresponding normalised spectra in the backward direction, for the same wavelength range.

The radiation close to 0° is expected to be optical transition radiation (OTR) in the backward direction. Additionally, two bands are visible. The upper one corresponds to effective backward Cherenkov radiation, caused by the constructive interference of backward resonance transition radiation. The lower band is the result of effective forward Cherenkov radiation, reflected from the chromium layer [112]. With increasing layer thickness, the effective Cherenkov radiation bands shift towards higher wavelengths. For thicknesses larger than 240 nm, a second, faint effective backward Cherenkov radiation band appears in the displayed wavelength range.
To detect the effective backward Cherenkov band visible in Figure 3.17, a layer thickness above 200 nm is desirable for a signal in the sensitive range of the CMOS camera sensor used for the experiment.
The larger the layer thickness, the more material has to be evaporated during the coating process. As explained in Section 3.5.2, the amount of material in the coating machine limits the number of layers, motivating an upper limit to the layer thickness of 260 nm.



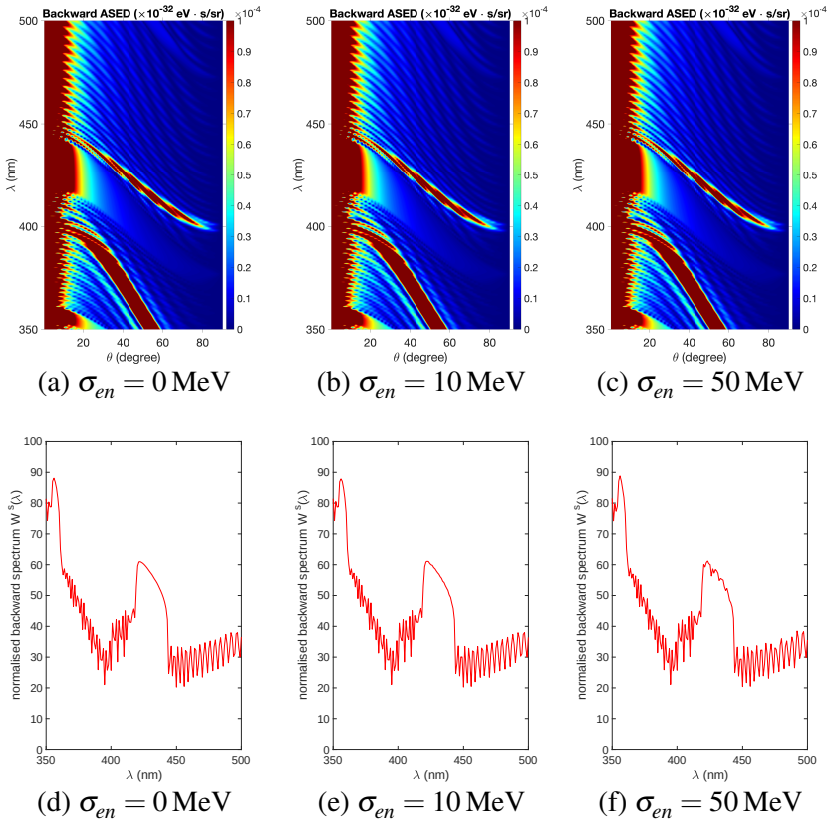
(a) 10 layer couples (b) 30 layer couples (c) 50 layer couples (d) 100 layer couples

(e) 10 layer couples (f) 30 layer couples (g) 50 layer couples (h) 100 layer couples

Fig. 3.18 Angular spectral energy density (top) and normalised backward spectrum (bottom) for a $SiO_2 \,|\, TiO_2$ photonic crystal with 230 nm layer thickness. The number of layer couples is indicated below the figure.

Figure 3.18 shows the ASED and normalised backward spectrum for a fixed 230 nm layer thickness and varying number of layer couples, ranging from 10 to 100. A higher number of layer couples results in a sharper signal, as visible in 3.18a to 3.18d and better signal to noise ratio, illustrated by the normalised backward spectrum in 3.18e to 3.18h.

The simulated effective Cherenkov radiation as a function of the electron energy is shown in Figure 3.19. The signal is clearly evident even for large variations in beam energy.



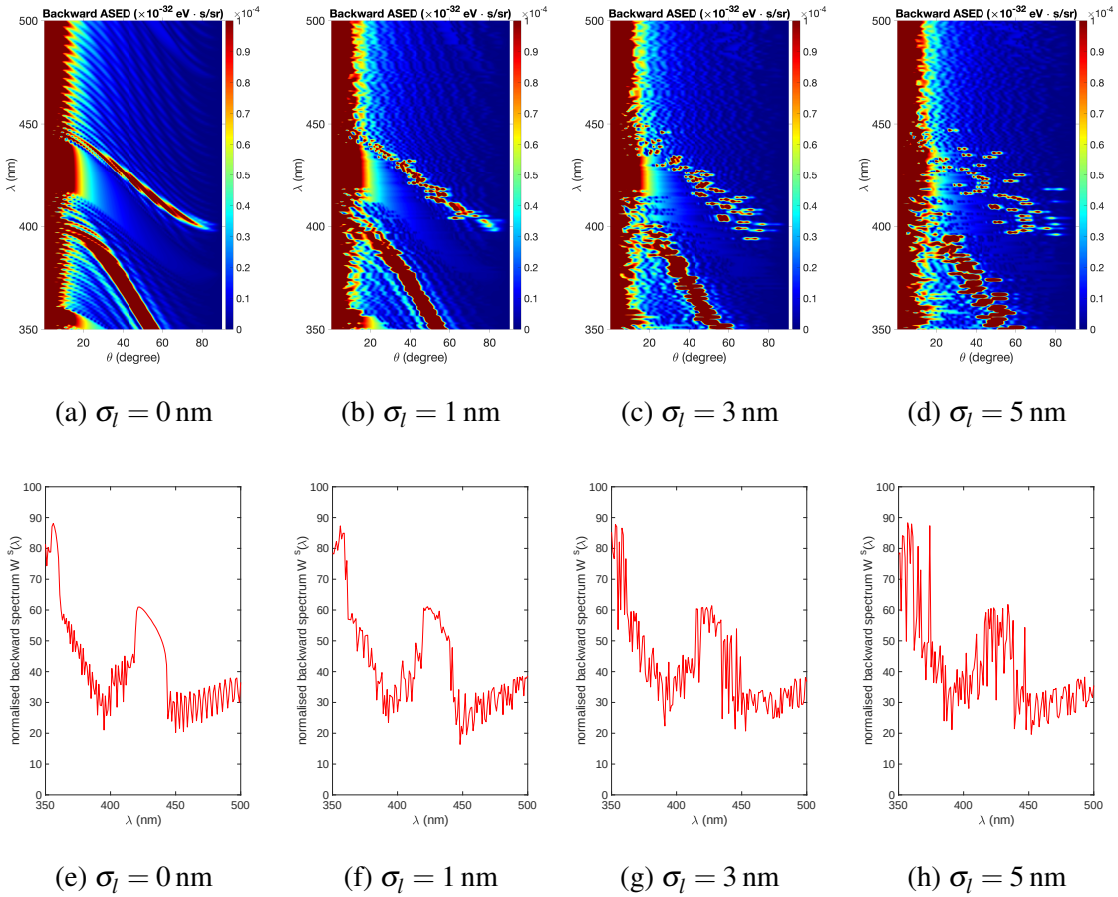
(a) $\sigma_{en} = 0\,\mathrm{MeV}$ (b) $\sigma_{en} = 10\,\mathrm{MeV}$ (c) $\sigma_{en} = 50\,\mathrm{MeV}$

(d) $\sigma_{en} = 0\,\mathrm{MeV}$ (e) $\sigma_{en} = 10\,\mathrm{MeV}$ (f) $\sigma_{en} = 50\,\mathrm{MeV}$

Fig. 3.19 Angular spectral energy density (top) and normalised backward spectrum (bottom) for a $SiO_2\,|\,TiO_2$ photonic crystal with 28 layer couples of 230 nm layer thickness. The electron energy is varied with a Gaussian spread, indicated below the corresponding figure.

Variations in layer thickness might be expected to cause the signal to deteriorate. To explore this, the layer thickness is varied. This is demonstrated in Figure 3.20 for 28 layer couples of 230 nm mean layer thickness and 200 MeV electrons. The thickness of each layer is sampled from a Gaussian distribution. Even for fluctuations as small as $\sigma_l = 3\,\mathrm{nm}$, a significant deterioration is observed and underlines the importance of exact process control during manufacturing. Larger variations make the signal likely undetectable.

The instability of the signal with thickness variations represents a notable concern for manufacturing.



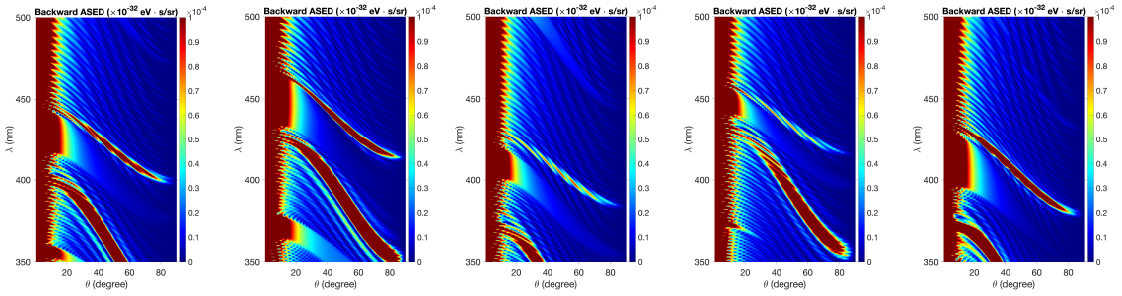
(a) $\sigma_l = 0\,\text{nm}$ (b) $\sigma_l = 1\,\text{nm}$ (c) $\sigma_l = 3\,\text{nm}$ (d) $\sigma_l = 5\,\text{nm}$

(e) $\sigma_l = 0\,\text{nm}$ (f) $\sigma_l = 1\,\text{nm}$ (g) $\sigma_l = 3\,\text{nm}$ (h) $\sigma_l = 5\,\text{nm}$

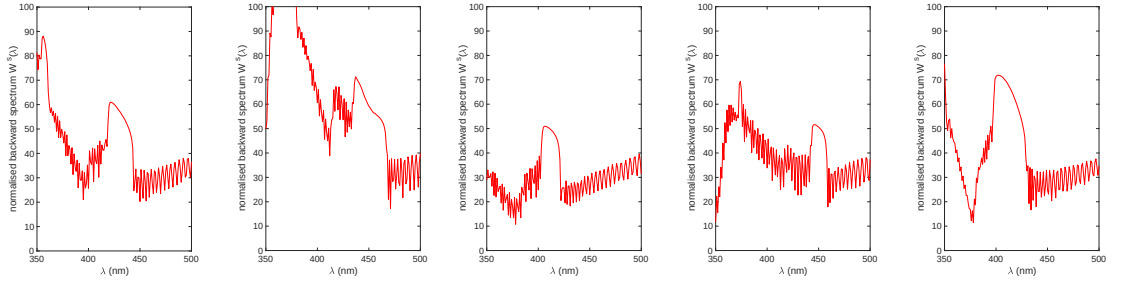
Fig. 3.20 Angular spectral energy density (top) and normalised backward spectrum (bottom) for a $SiO_2 \,|\, TiO_2$ photonic crystal with 28 layer couples of 230 nm layer thickness, varied with a Gaussian spread, indicated below the corresponding figure.

The photonic crystal configuration is also sensitive to variations of the refractive index, as shown in Figure 3.21. The refractive indices of SiO_2 and TiO_2 are functions of the wavelength. As described in Section 3.5.2, the refractive index depends on the relative amount of the different stoichiometric forms of titanium-oxide arrangements.

Figure 3.21b and 3.21c show the effect of a positive and negative refractive index offset of 0.2 for both SiO_2 and TiO_2. The positive offset results in a shift of the band structures towards higher wavelength, and the opposite for the negative offset. A positive n_{TiO_2} or negative n_{SiO_2} offset creates a larger refractive index contrast, resulting in a stronger signal. In figure 3.21d and 3.21e, the offset of 0.2 and -0.2 is applied to the refractive index function of SiO_2. The shift of the signal bands is similar, as well as its reduction in strength when decreasing the difference to the refractive index of TiO_2 with a positive offset, and vice versa.



(a) Default (b) $n_{TiO_2} + 0.2$ (c) $n_{TiO_2} - 0.2$ (d) $n_{SiO_2} + 0.2$ (e) $n_{SiO_2} - 0.2$



(f) Default (g) $n_{TiO_2} + 0.2$ (h) $n_{TiO_2} - 0.2$ (i) $n_{SiO_2} + 0.2$ (j) $n_{SiO_2} - 0.2$

Fig. 3.21 Angular spectral energy density and backward spectrum for a $SiO_2 \mid TiO_2$ photonic crystal with 28 layer couples of 230 nm layer thickness. The refractive index is varied by a constant offset, indicated below the figures.

3.5.2 Thin-layer coating

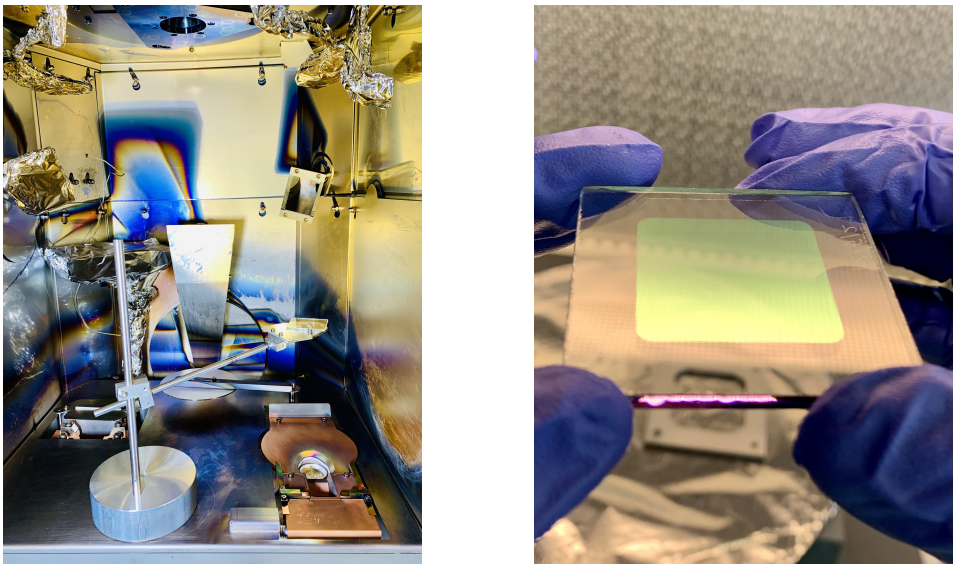
In this section, the manufacturing process of a one-dimensional $SiO_2 \mid TiO_2$ photonic crystal at CERN is described. As illustrated in 3.16, the photonic crystal layers are deposited on a quartz substrate. A chromium layer coating is added between the substrate and the photonic crystal to reflect background radiation generated in the substrate. As described in Section 3.5.1, an ideal thickness of the photonic crystal layers is between 200 and 260 nm, to generate a backward effective Cherenkov signal above 400 nm wavelength and to minimise the use of evaporated material in the coating process.

Electron-beam evaporation setup

A Polyteknik[10] Flextura M5808 E evaporator is used for coating, installed at the CERN Thin Film & Glass laboratory. The machine uses electron-beam physical vapour deposition and can hold up to 6 different materials in separate pockets. Inside the vacuum chamber, the

[10] Polyteknik AS: `https://www.polyteknik.com/`

materials are transformed into the gaseous phase by a 10 keV beam. The beam current can be adjusted to regulate the material deposition rate.



(a) Evaporation chamber. (b) Single-layer sample.

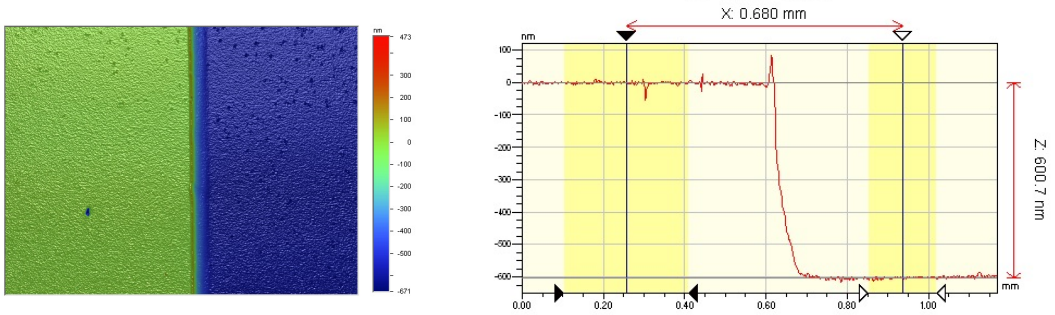
Fig. 3.22 Polyteknik evaporator with sample holder (left) and single SiO_2 | TiO_2 layer couple deposition for coating calibration (right).

Figure 3.22a shows the evaporator chamber. The coating substrate in the sample holder is positioned close to the material pocket on the bottom right to maximise the angular acceptance of the evaporated material. Attaching the sample to the rotating top of the coater would increase the layer uniformity, however, drastically increase the required material per layer.
At the same time, the substrate is kept far enough away to avoid damage by sparks from the heated material and at an angle orthogonal to the line of sight to the material pocket for an even layer thickness.

Nanometre layer calibration

The evaporator features a quartz crystal monitor (QCM) at the top of the chamber, directly above the material pocket, to control the relative layer thickness. A geometric factor allows estimating the deposition depth on the sample relative to the quartz monitor as a calibration starting point. The layer thickness is calibrated through a series of single SiO_2 | TiO_2 layer couple depositions to determine a correction factor for thickness monitoring.

Figure 3.23 illustrates a calibration sample measurement with 100 nm chromium, 250 nm SiO_2, and 250 nm TiO_2, resulting in a total thickness of 600 nm. The thickness is measured using a Veeco 3D optical profilometer[11] [123].



(a) 3D optical profile.

(b) Thickness measurement.

Fig. 3.23 Optical profile and thickness measurement of a chromium | SiO_2 | TiO_2 sample with layer thicknesses of 100 nm, 250 nm, and 250 nm, respectively, for a total of 600 nm.

When evaporating TiO_2, the resulting vapour and hence deposited layer may consist of various titanium-oxide stoichiometric forms. To optimise the TiO_2 film quality with stable and reproducible optical properties, Ti_3O_5 is evaporated as starting composition, following Reference [124]. As a result, the vapour should exclusively consist of TiO molecules. To achieve a consistent TiO_2 deposition, oxygen is added to the vacuum at a constant flow rate of 5 sccm[12], to achieve a refractive index of $n(\lambda = 550\,\text{nm}) = 2.41$, as stated in the reference. The granular raw materials are pre-melted by the electron beam without a substrate installed to minimise sparks during the evaporation process.

Photonic crystal manufacturing

The manufacturing goal is a SiO_2 and TiO_2 photonic crystal with a layer thickness between 200 and 260 nm. More important than the exact thickness is their consistency throughout the sample.
A 100 nm chromium layer thickness is chosen to minimise transmission while not negatively affecting the adhesion and thermal expansion properties of the material compound. This thickness leads to only negligible transmission throughout the entire spectrum of relevant wavelengths, as demonstrated by the transmission measurements of Figure 3.24.

The coating procedure is programmed to stop once the thickness measured on the QCM reaches the value pre-determined by the calibration. The number of layers is limited by

[11] Veeco Instruments Inc.: `https://www.veeco.com`
[12] Standard cubic centimetres per minute, indicating cm^3/min

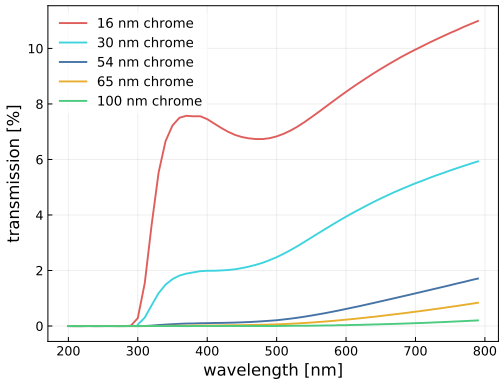
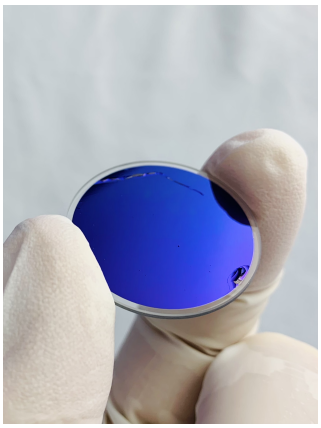


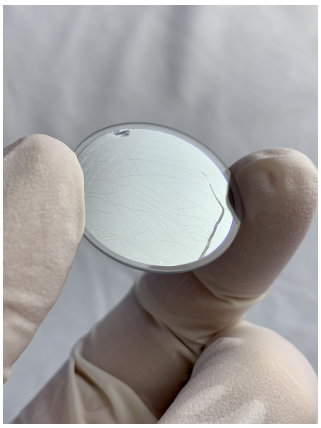
Fig. 3.24 Light transmission percentage through glass substrate samples covered with a single layer of chromium of varying thickness. At 100 nm thickness only a negligible fraction of light passes through.

the material in the coater and the adhesion to the substrate, for which a SiO_2 disk of 1 mm thickness and 30 mm diameter is used.
Two photonic crystal samples were produced: the first one in a two-stage process (30L-PhC) and the second one in a single coating session (28L-PhC).

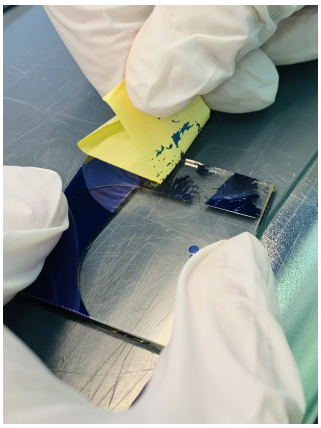
For the 30L-PhC, the coating was interrupted after 10 SiO_2 | TiO_2 layer couples to investigate and refill the evaporation material pockets. In the second coating session, 20 more layer couples were added.
As visible in Figure 3.25a, the sample features two defects on the front. Since they are located at the edge of the 30 mm substrate disk, they are not expected to affect the beam measurements. The rear view of the crystal in Figure 3.25b shows fractures of the chromium layer.
An additional substrate was installed next to the photonic crystal to test the binding strength of the thin films. The second set of layers do not adhere well, as displayed in Figure 3.25c. That discourages going beyond 30 layer couples and a multi-stage coating process in general. The temperature varies during coating process temperature fluctuations and the cooling phase to room temperature after the two coatings. The chromium layer cracks are likely caused by thermal stress due to the different thermal expansion properties of the materials.
A compound with many layers, as the 61 in this case, is prone to adhesion difficulties. Compromising the vacuum after the first coating stage could have introduced impurities which may have an additional adverse effect on the adhesion.



(a) 30L-PhC front view.



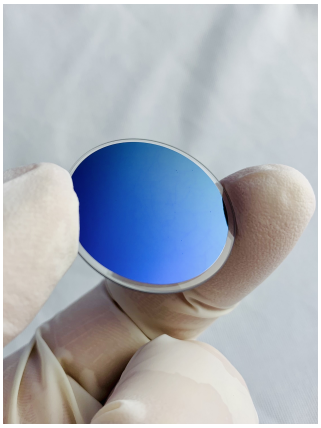
(b) 30L-PhC rear view.



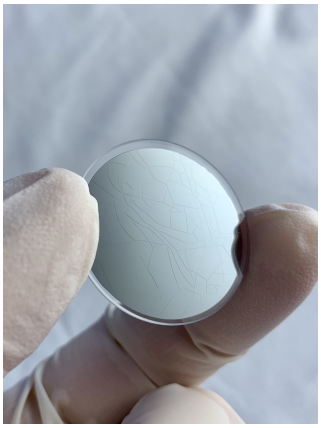
(c) Adhesion test.

Fig. 3.25 Photonic crystal A with 30 layer couples, manufactured in a two-stage process, front (a) and back (b) view. The layer adhesion is tested on a separate substrate (c).

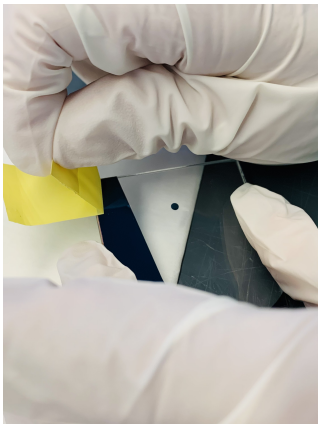
A second photonic crystal was produced, this time in a single coating process, shown in Figure 3.26. This sample features fewer chromium layer cracks, indicating reduced thermal stress on the materials. The test on the parasitic substrate in Figure 3.26c demonstrates an improved layer adhesion.



(a) 28L-PhC front view.



(b) 28L-PhC rear view.



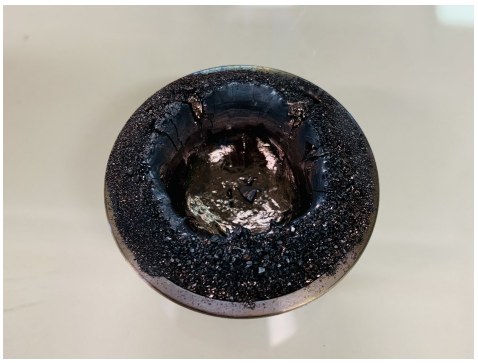
(c) Adhesion test.

Fig. 3.26 Photonic crystal B with 28 layer couples, manufactured in a single coating process, front (a) and back (b) view. The layer adhesion is tested on a separate substrate (c).

Out of the six available pockets one was filled with chromium, two with SiO_2, and three with Ti_3O_5. For both the SiO_2 and Ti_3O_5 only approximately 2 % were evaporated.
The coating had to be concluded after 28 layer couples to avoid the electron beam from heating the pocket material. Despite varying the electron beam location in a pre-defined star pattern, the cavities approached the depth of the pocket, as illustrated in Figure 3.27.



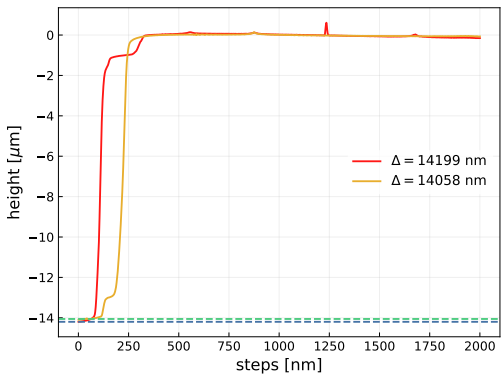
(a) Photonic crystal B front view.



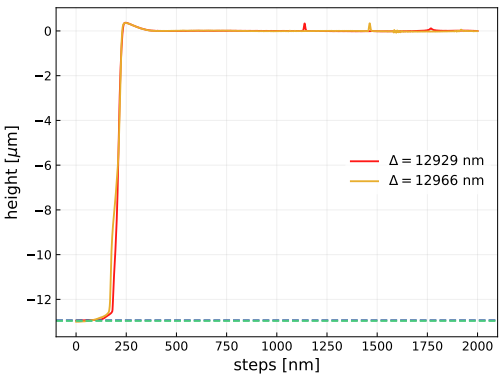
(b) Photonic crystal B rear view.

Fig. 3.27 SiO_2 (a) and Ti_3O_5 (b) material pockets after coating. The electron beam created a cavity down to the pocket's depth, limiting the number of layer depositions.

A measurement of the photonic crystals' total thickness is shown in Figure 3.28a and 3.28b. The height is measured in two locations on opposite sides of the sample. The distance relative to the baseline is depicted in the legend. The similarity of the measurements on the two sides of the sample indicates a uniform thickness of the layers. Subtracting the 100 nm chromium layer, sample A has an average layer thickness of 233.8 nm and photonic crystal B of 229.4 nm.



(a) 30L-PhC total thickness.



(b) 28L-PhC total thickness.

Fig. 3.28 Photonic crystal thickness measurement for 30L-PhC (a) and 28L-PhC (b).

3.6 SiO_2 | TiO_2 photonic crystal high energy beam tests

The behaviour of photonic crystals, when exposed to high energy particles, is expected to be extremely sensitive to their detailed structure. Therefore, comparing their response with the expected behaviour deduced from simulations is a very powerful way of testing the theoretical model. In this section, I present an experimental apparatus that represents the first attempt to measure backward effective Cherenkov radiation from a dielectric 1D photonic crystal in a high energy particle beam and provides a laboratory to explore strategies for mitigating sources of experimental backgrounds.

3.6.1 Beam test setup

The experimental setup for the CLEAR beam tests of the two photonic crystal samples is illustrated in Figure 3.29. The 200 MeV electrons will penetrate the photonic crystal with its multi-layer structure facing the beam. A Basler camera[13] with a SONY[14] CMOS sensor is mounted on a precision rotation stage. The angle of acceptance reaches from 10 to 90°, limited by the proximity of the camera to the particle beam.

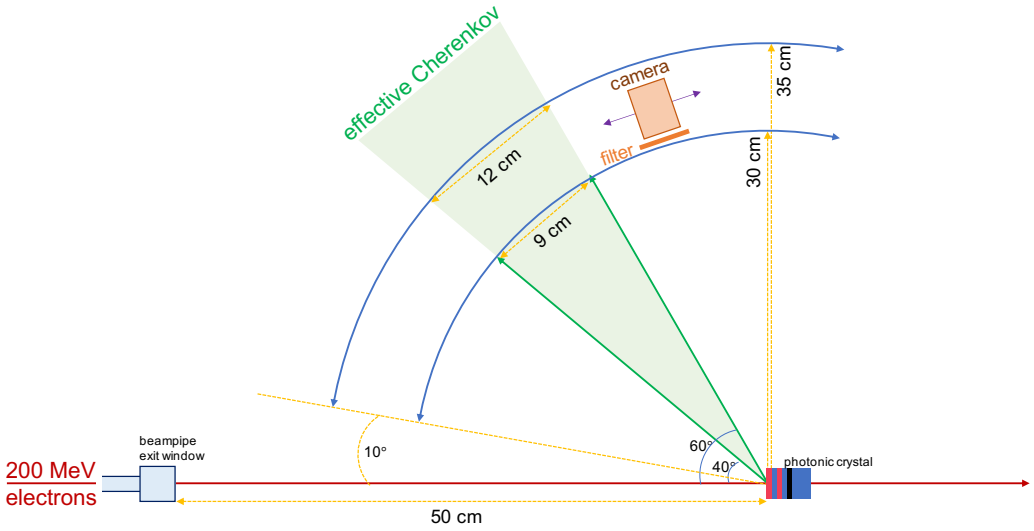
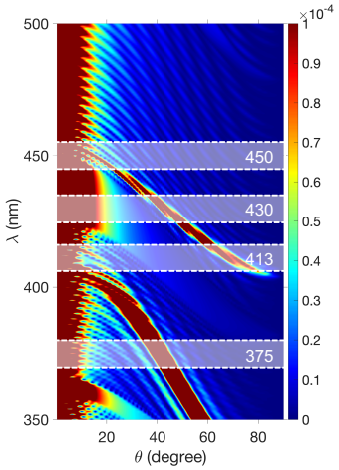


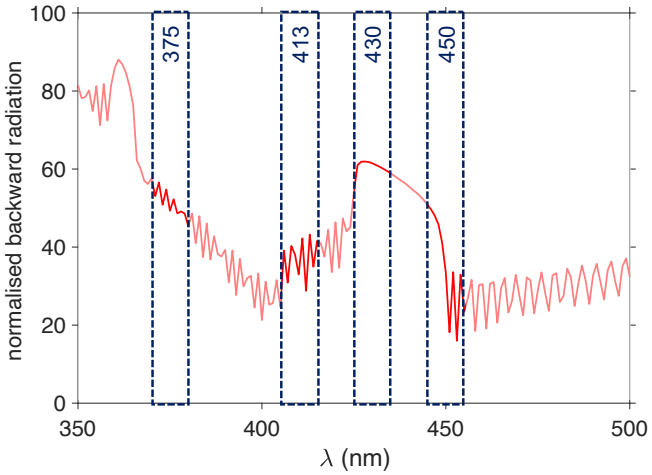
Fig. 3.29 Schematic of the beam test setup at CLEAR. The electron beam penetrates the photonic crystal structure. A CMOS camera and bandpass filter (BPF) are mounted on a rotation stage. A potential emission range of an effective Cherenkov signal between 40 and 60° is illustrated in green.

[13] Basler AG: `https://www.baslerweb.com/`
[14] Sony Semiconductor Solutions Corporation: `https://www.sony-semicon.co.jp/e/`

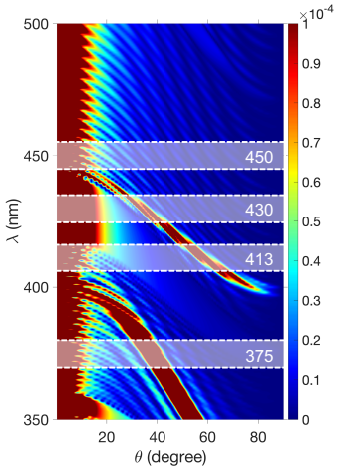
A bandpass filter is mounted in front of the camera, to select a narrow wavelength region. Four Edmund Optics[15] bandpass filters with 10 nm full width at half maximum are used, with centre wavelengths of 450 nm, 430 nm, 413 nm, and 375 nm.



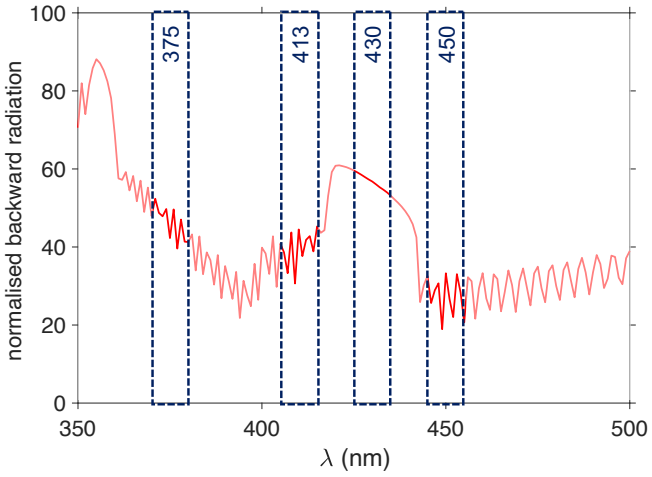
(a) 30L-PhC ASED.



(b) 30L-PhC spectrum.



(c) 28L-PhC ASED.



(d) 28L-PhC spectrum.

Fig. 3.30 Photonic crystal ASED and backward spectrum for 30L-PhC (top) and 28L-PhC (bottom). The bandpass filter wavelength selection ranges are indicated by the dashed boxes.

[15] Edmund Optics Ltd: `https://www.edmundoptics.co.uk`

CLEAR in-air test stand setup

Figure 3.31 shows a top view of the experimental setup installed on the CLEAR in-air test stand. The beam direction is from left to right. The photonic crystal is installed in a PVC plastic holder on a manual translation stage for position adjustments along the beam axis, visible on the right side. The CMOS camera and bandpass filter holder are mounted on the rotation stage, which is positioned at a 10° angle. The camera is oriented towards and focussed on the photonic crystal with the bandpass filter in its line of sight. The setup is laser-aligned with respect to the beam axis and the camera rotation around the photonic crystal. The data acquisition of the CMOS camera is gated with a 30 ns time window and synchronised with the beam. The ambient light in the beam area was extinguished during data taking.

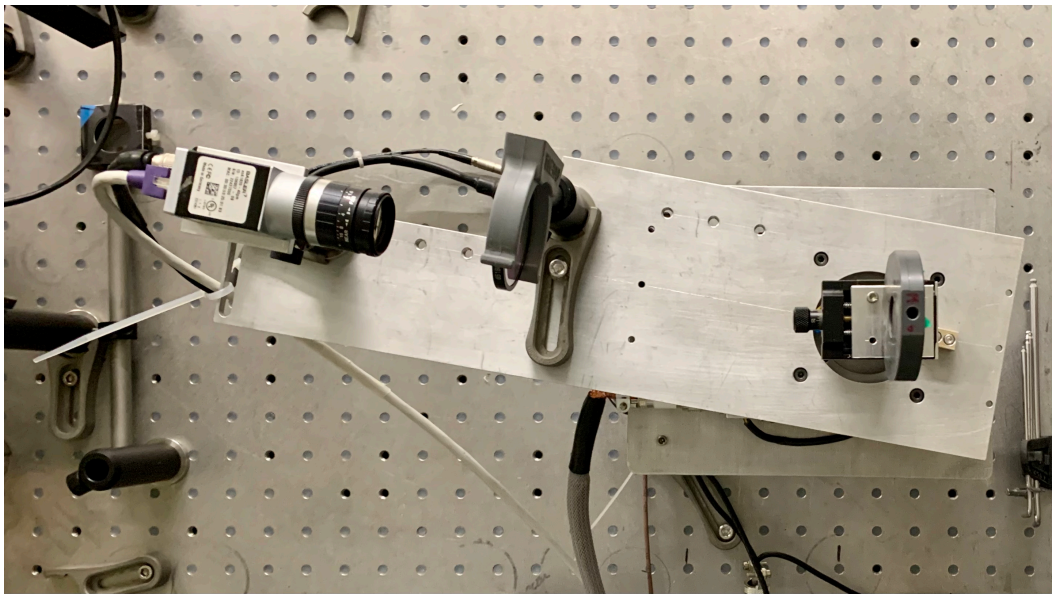
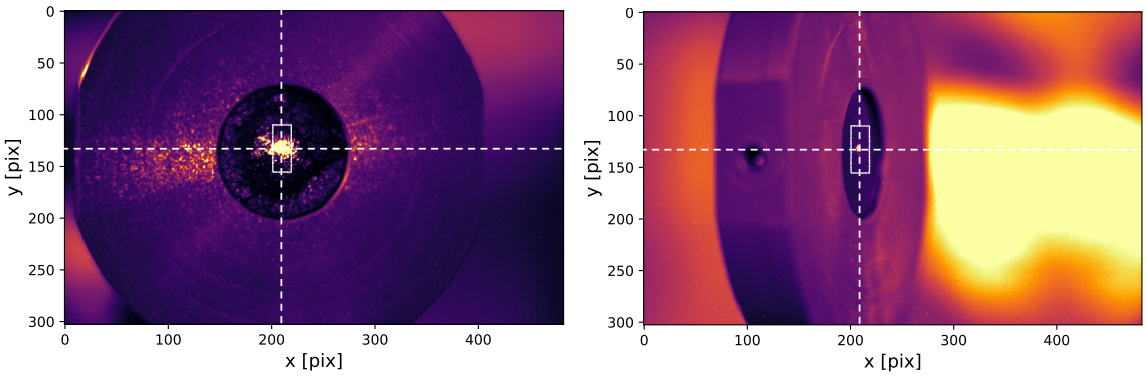


Fig. 3.31 Top view on the CLEAR beam test setup. The photonic crystal is in the centre of the rotation stage, mounted on a precision stepping motor. The photonic crystal is orthogonally aligned to the beam inside a PVC plastic holder, mounted on a manual translation stage to adjust its position to the centre of rotation. The CMOS camera is mounted at 30 cm distance on the far end of the rotation stage. A bandpass filter is located in the line of sight between the camera and the photonic crystal.

Figure 3.32 shows image examples taken with the CMOS camera at 10° and 70°. The beam alignment laser is visible in the centre of the sample. The laser spot and an alignment cross, visible as a white dashed line, are used to position the crystal in the centre of rotation with respect to the camera. At angles above 75°, the plastic holder obstructs the camera's view of the sample.



(a) 10° (b) 70°

Fig. 3.32 Beam test camera view on the photonic crystal at 10° (a) and 70° (b). The photonic crystal is installed in a plastic holder. A beam-direction laser is visible in its centre, which is used for the rotational alignment of the camera, as marked by the dashed lines. The fiducial region is marked by a white rectangle around the centre.

In a second variation of the experimental setup, an optical spectrometer is installed. It is coupled to a fibre and positioned close to the photonic crystal, as displayed in Figure 3.33.

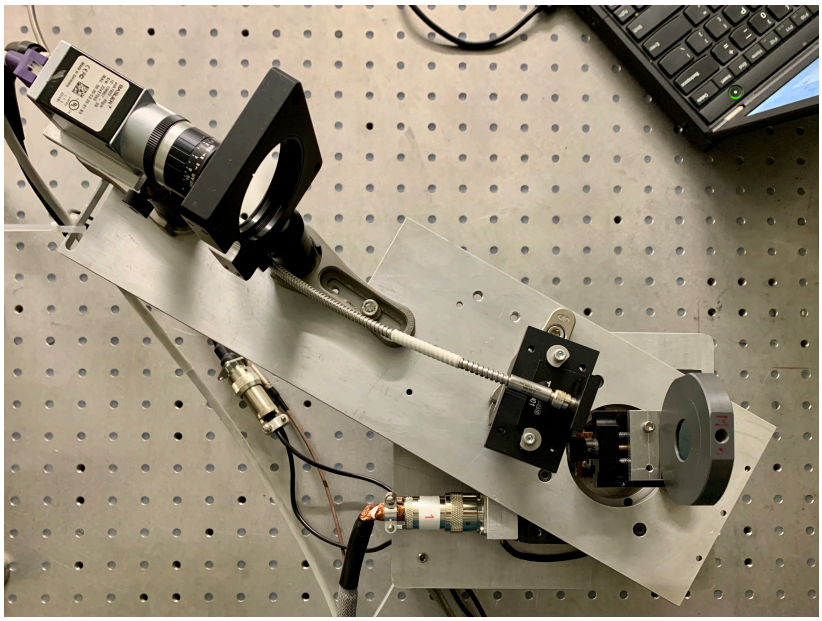
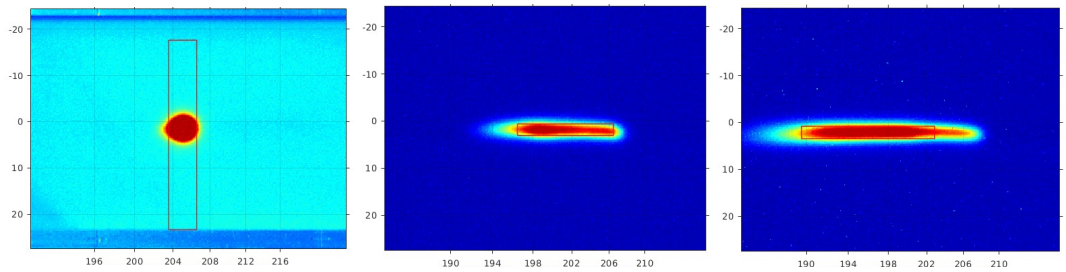


Fig. 3.33 Top view on the spectrometer setup. The spectrometer is coupled to a light-guiding fibre, which is mounted on the rotation stage, close to the photonic crystal.

Beam setup

The CLEAR electron beam was operated with a beam energy of 200 MeV and a repetition rate of 0.83 Hz. The beam energy spread varies with the number of bunches. Beam spectrometer monitor examples are shown in Figure 3.34 for a single bunch, 20, and 40 electron bunches. The bunch numbers approximately correspond to train charges of 0.5, 10, and 20 nC, respectively, depending on the beam transport quality. The saturation of the beam monitor only allows for an estimation of the energy spread of $\Delta_1 \approx 2$ MeV, $\Delta_{20} \approx 5$ MeV, and $\Delta_{40} \approx 10$ MeV. As shown in Figure 3.19, the effective Cherenkov signal is expected to be robust to electron energy variations.



(a) Single bunch. Beam energy 204 MeV.

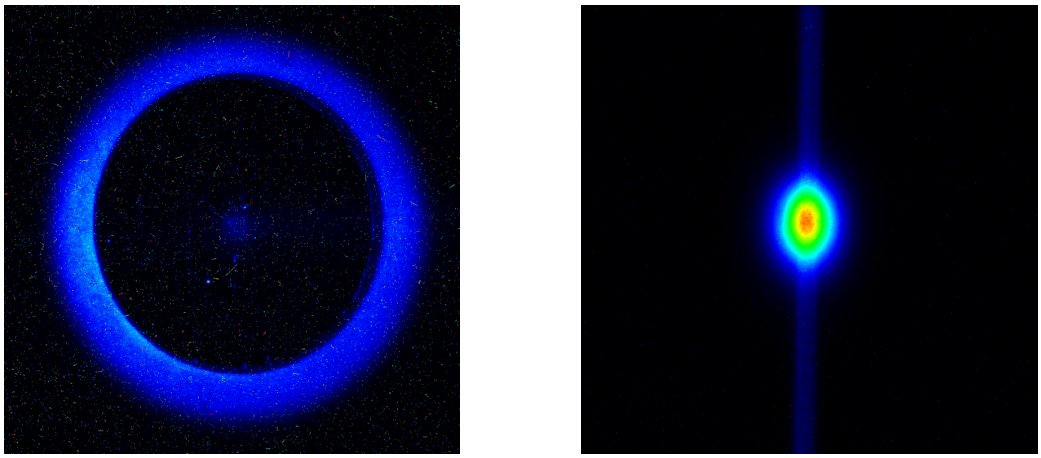
(b) 20 bunches. Beam energy 201 MeV.

(c) 40 bunches. Beam energy 198 MeV.

Fig. 3.34 CLEAR beam energy monitoring for a single bunch (a), 20 (b), and 40 (c) electron bunches. The beam energy is tuned to 200 MeV, and illustrated on the x-axis.

In addition to the alignment of the position of the photonic crystal surface relative to the camera rotation, the beam position and focus was calibrated. A fluorescent screen was attached to the rear of the photonic crystal holder, emitting light when traversed by charged particles. Figure 3.35b shows a view on the rear side of the photonic crystal holder during beam operation. The vertical alignment laser is enabled and used to position and focus the beam.

Figure 3.35a shows the CMOS camera view of the sample during beam operation and without a bandpass filter. Considerable light emission is visible from the inner part of the plastic holder, around the photonic crystal. That background may originate from scintillation of the plastic or from light that escapes the edges of the photonic crystal sample and traverses the thin holder front. To reduce the background, a fiducial region is defined, which is a rectangular section in the centre of the CMOS camera pictures, as indicated by the white boxes in Figure 3.32.



(a) Frontal view. (b) Rear view.

Fig. 3.35 Frontal view of the sample holder during beam operation without a bandpass filter (a). Background light is emitted from the inner part of the photonic crystal holder. Rear-side view of photonic crystal holder with focussed beam on the attached fluorescent screen and vertical alignment laser (b).

Background mitigation

During the beam test campaign, the setup was modified to mitigate various background sources. These are described in this section and will be referred to for the corresponding results, presented Section 3.6.2.

Two bandpass filter holders were used, a metal and a plastic holder. The first one, visible in Figure 3.33, was replaced by a plastic holder, which is present in Figure 3.31 to minimise its obstruction of the beam at low angles.

The background observed by the CMOS camera significantly increased in proximity to the beamline. A major noise reduction was achieved by installing lead shielding to protect the camera from electrons scattered along the beam direction and back-scattering from the beam dump.

Additionally, black velvet tape was installed on the photonic crystal holder front to absorb its light emission, and a tube was fitted in front of the camera to reduce its angular acceptance.

The fluorescent alignment screen on the rear side of the photonic crystal holder was removed. Instead, a separate screen at the end of the beamline was used for beam focussing and alignment. Figure 3.36 shows the modified experimental setup with the described background reduction measures.

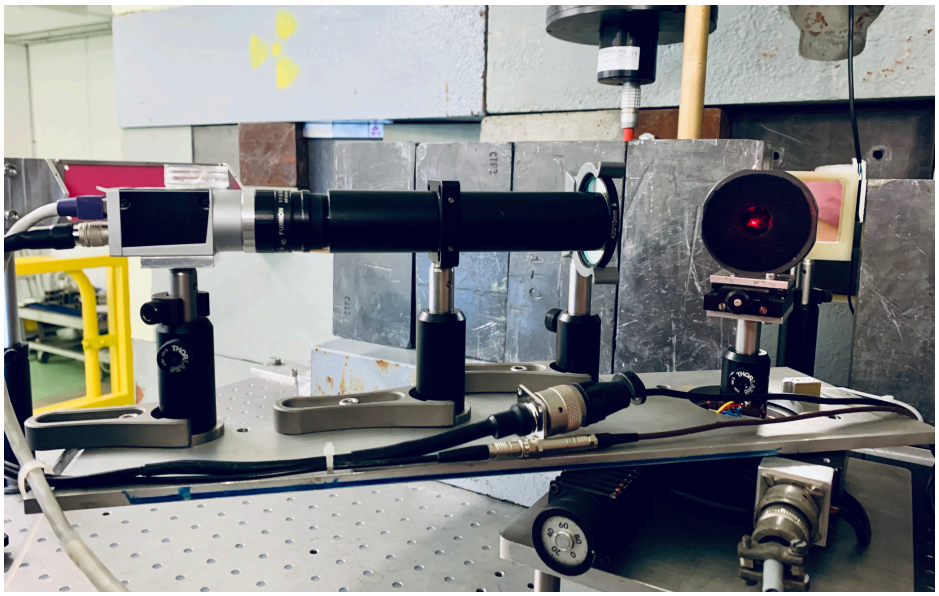


Fig. 3.36 Front view on CLEAR in-air setup with background reduction measures.

3.6.2 Analysis and results

The duration of the beam campaign was two weeks, denoted as phases 1 and 2. The results will be presented together with the corresponding modifications made to the experimental setup. Measurements of the 28 layer photonic crystal were prioritised due to its higher manufacturing quality and beam-time limitations.

For each bandpass filter and photonic crystal configuration, measurements in the angular range from 10 to 90° as illustrated in Figure 3.29 were taken. The number of measurements was limited by the beam time provided by the CLEAR facility and the beam stability. The beam transportation quality and consequently the bunch charge, while generally consistent during individual runs, varied between measurements.

Beam phase 1 - Camera setup

During the first data-taking week both the 30L-PhC and 28L-PhC were studied in the CMOS camera configuration shown in Figure 3.31, using the four bandpass filters.
For the CMOS measurements, the camera was moved through the angular range, taking a series of measurements in 2° intervals. A fiducial region is defined, restricting the area of interest (AOI) to the centre of the photonic crystal. The pixel intensity values of the AOIs are integrated, and the mean is taken over all measurements for a given angle.

The bandpass filters were mounted on a metal holder for the 28L-PhC measurements, which was changed to a plastic holder during the 30L-PhC installation to minimise obstruction of the beam line at low angles.

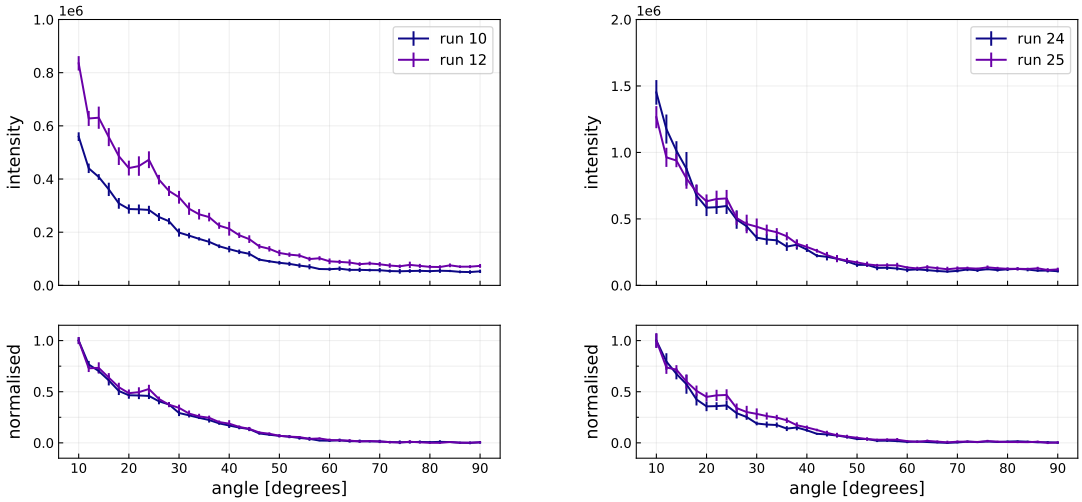
The integrated pixel intensity measurements for each bandpass filter are shown for the 28L-PhC on the left and the 30L-PhC on the right in Figures 3.37 to 3.40. The min-max normalised intensities are shown below the corresponding figures. Since beam conditions change between runs, the shape of the curves is more informative than their absolute height. Additional factors affecting the absolute intensity are the CMOS sensor response curve and the background light spectrum with respect to a given bandpass filter. During runs, the beam conditions were monitored, and datasets were rejected when the deviations were too large.
A general trend of decreasing intensity with increasing angle is visible. Additionally, all curves show an increase in intensity between 20 and 30°. As shown in Figure 3.30, the expected signal should move with changing bandpass filters, which is not the case in the data. It is possible that this is due to obstruction of the beam path and its related background for small angles may change the otherwise continuous background distribution.
An increase in intensity above 75° is visible for one of the 30L-PhC measurements with a 375 nm bandpass filter, shown in Figure 3.40b. As mentioned in Section 3.6.1 and demonstrated in Figure 3.32, the photonic crystal holder obstructs the view of the camera on the sample. The increase, visible in the 30L-PhC 375 nm BPF measurement, is therefore unlikely to originate from an effective Cherenkov signal.

Beam phase 1 - spectrometer setup

Spectrometer measurements were taken during the first beam test phase, as shown in Figure 3.33. At each angle setting, the optical spectrum was integrated over three beam pulses and for the same duration without beam. A spectrometer scan of the angular range required a considerably longer time than the camera measurements. To account for possible hysteresis effects caused by changing beam conditions, the spectrometer fibre was first moved from 20 to 90° and then from 85 to 25°, both in steps of 10°. Below 20°, the optical fibre holder obstructed the beam path due to its proximity to the sample.

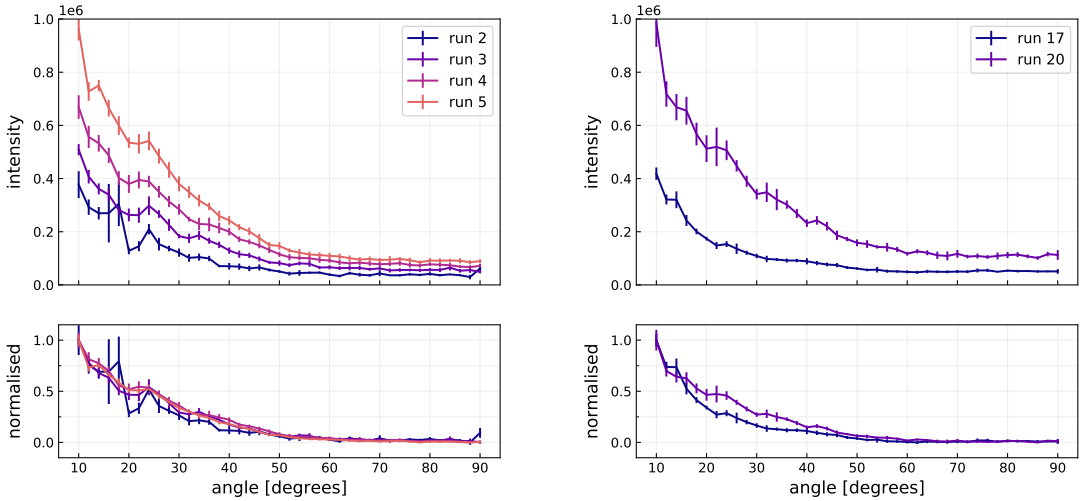
Figure 3.41 shows the spectrum from 180 to 880 nm for the 28L-PhC (left) and 30L-PhC (right), with and without beam, and the subtraction of the two respective measurements, for 20, 45 and 70°.

The broad wavelength range of the optical spectrometer would allow measuring a signal similar in shape to the simulations in Figure 3.30, also if it occurred at a different wavelength



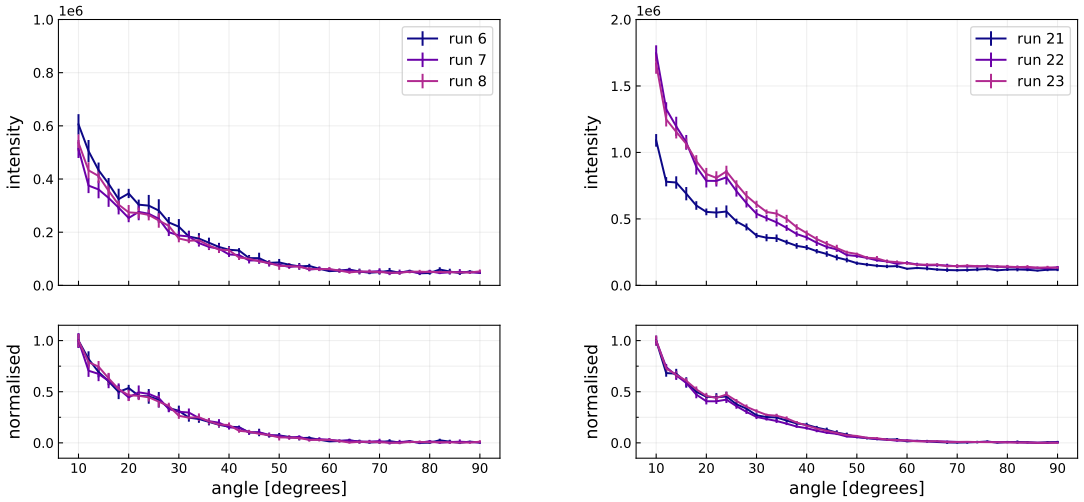
(a) 28L-PhC & 450 nm BPF (b) 30L-PhC & 450 nm BPF

Fig. 3.37 Beam test phase 1 CMOS camera measurements with a 450 nm bandpass filter of the 28L-PhC (a) and 30L-PhC (b).



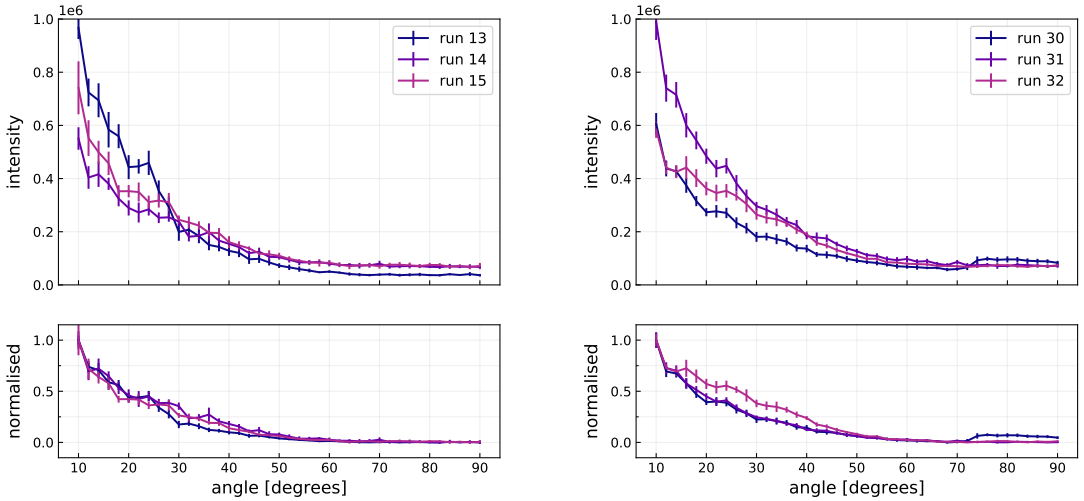
(a) 28L-PhC & 430 nm BPF (b) 30L-PhC & 430 nm BPF

Fig. 3.38 Beam test phase 1 CMOS camera measurements with a 430 nm bandpass filter of the 28L-PhC (a) and 30L-PhC (b).



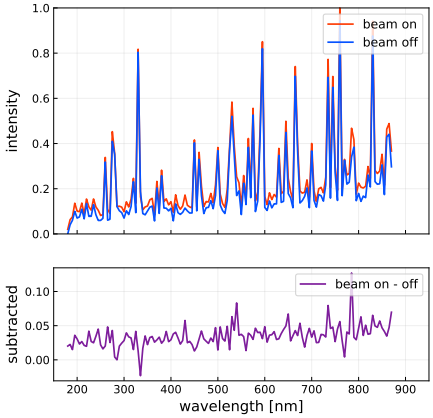
(a) 28L-PhC & 413 nm BPF (b) 30L-PhC & 413 nm BPF

Fig. 3.39 Beam test phase 1 CMOS camera measurements with a 413 nm bandpass filter of the 28L-PhC (a) and 30L-PhC (b).

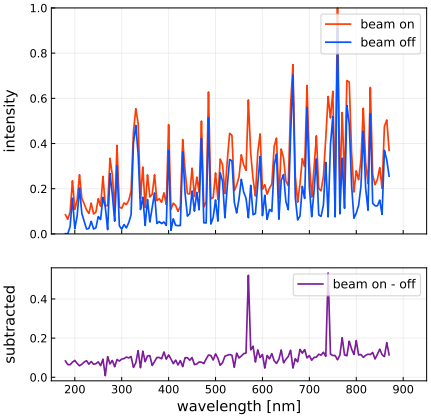


(a) 28L-PhC & 375 nm BPF (b) 30L-PhC & 375 nm BPF

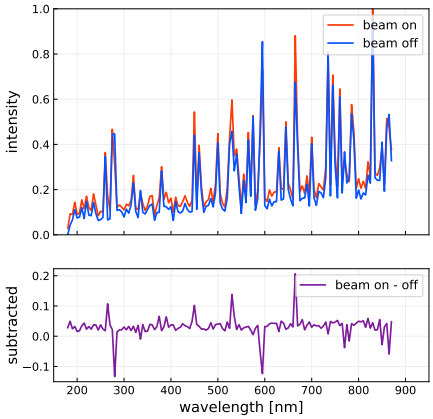
Fig. 3.40 Beam test phase 1 CMOS camera measurements with a 375 nm bandpass filter of the 28L-PhC (a) and 30L-PhC (b).



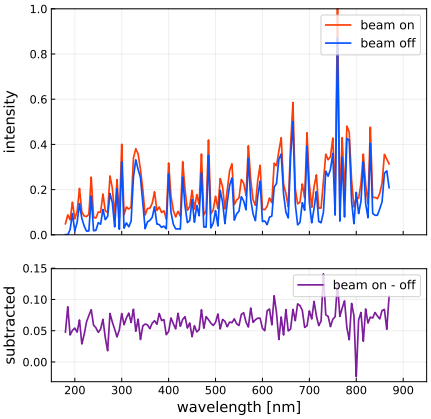
(a) 28L-PhC spectrum at 20°.



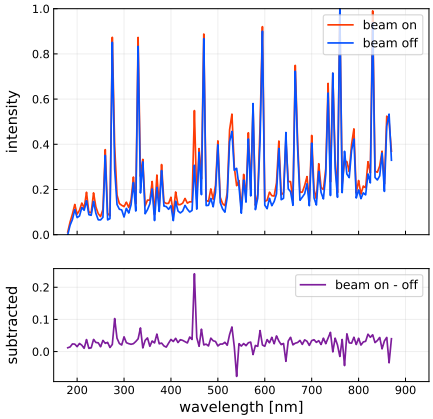
(b) 30L-PhC spectrum at 20°.



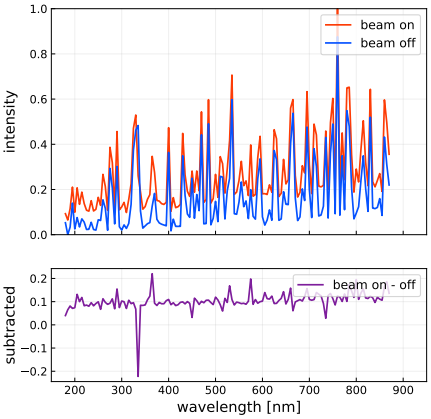
(c) 28L-PhC spectrum at 45°.



(d) 30L-PhC spectrum at 45°.



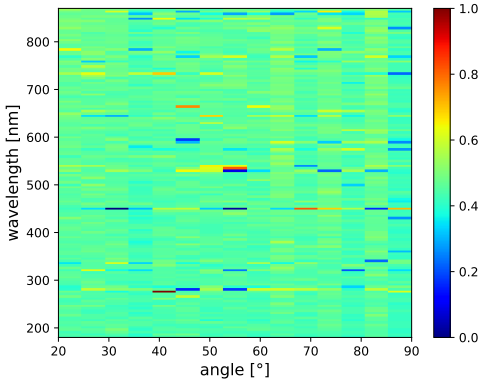
(e) 28L-PhC spectrum at 70°.



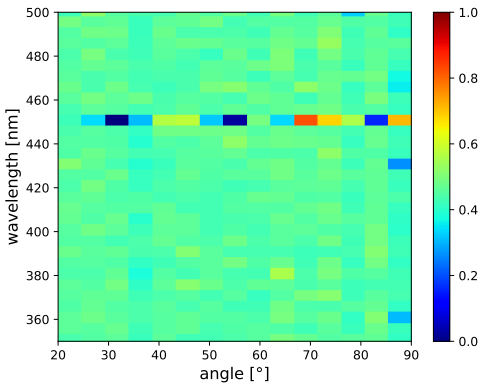
(f) 30L-PhC spectrum at 70°.

Fig. 3.41 Measurement of the 28L-PhC (left) and 30L-PhC (right) optical spectra with and without beam and their respective difference at 20°, 45° and 70°.

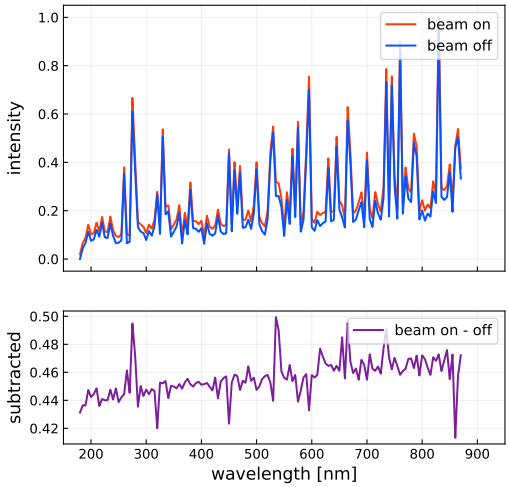
region. Only narrow intensity peaks are observable, which do not remain consistent for different angular positions and are likely caused by sensor noise.



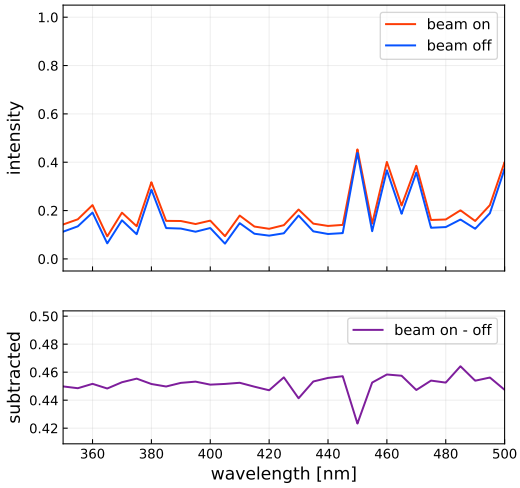
(a) 28L-PhC spectrum vs. angle.



(b) 28L-PhC spectrum vs. angle, 350 to 500 nm.



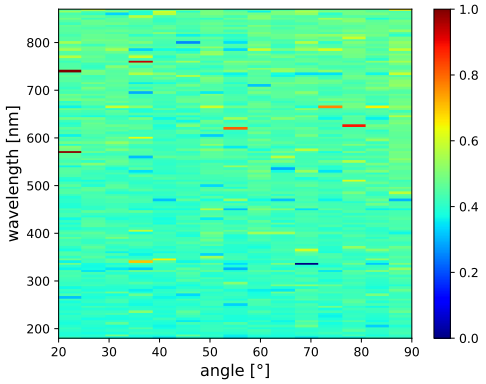
(c) Integrated 28L-PhC spectrum.



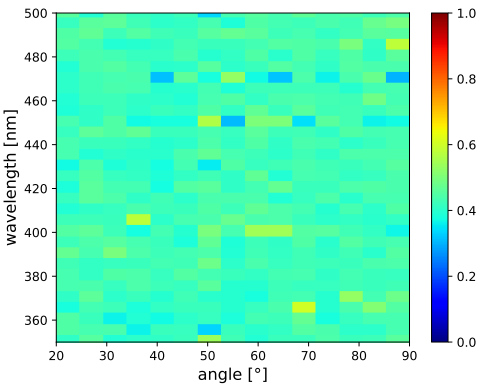
(d) Integrated 28L-PhC spectrum, 350 to 500 nm.

Fig. 3.42 Measurement of the 28L-PhC normalised optical spectrum (beam on - beam off) as a function of the angle (top) and integrated over all angles (bottom), for the full wavelength range (left) and wavelengths 350 to 500 nm (right).

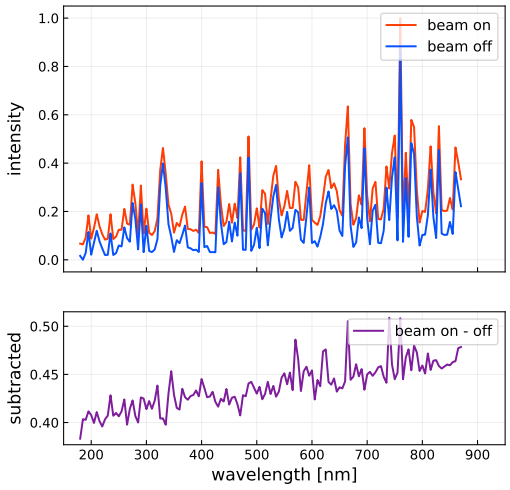
In Figures 3.42 and 3.43, the optical spectrum is shown as a function of the angle for the 28L-PhC and 30L-PhC, respectively. For each bin, the no-beam measurements are subtracted from the ones with beam. The colour indicates the normalised intensity. The spectrum is uniform, apart from a deviation around 450 nm of the 28L-PhC. While the observed structure may motivate further spectrometer studies, it may be an instrumental effect and is not of a significant magnitude.



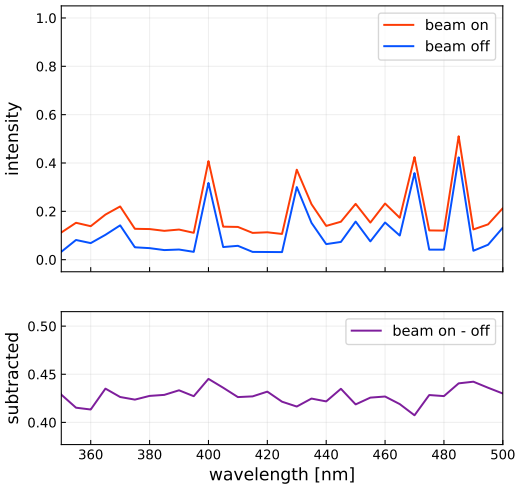
(a) 30L-PhC spectrum vs. angle.



(b) 30L-PhC spectrum vs. angle, 350 to 500 nm.



(c) Integrated 30L-PhC spectrum.

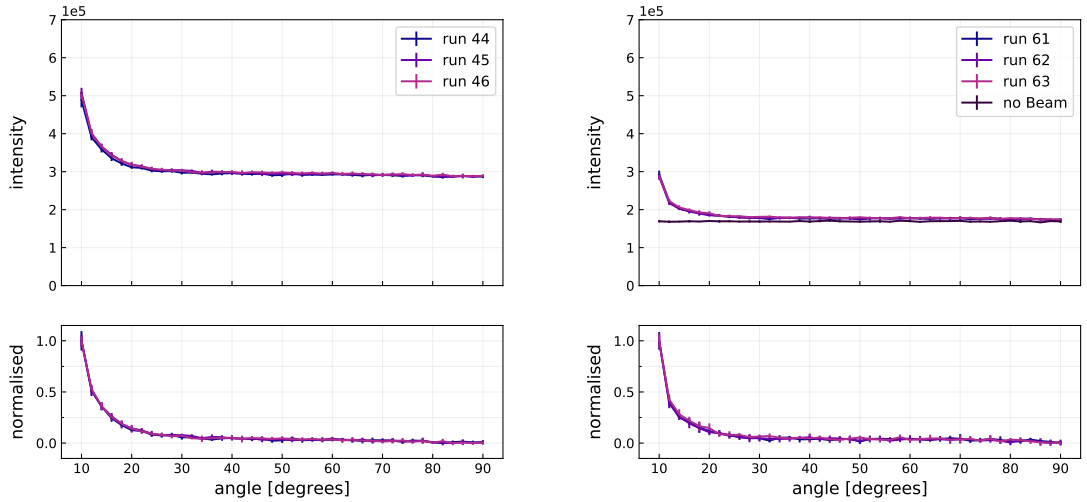


(d) Integrated 30L-PhC spectrum, 350 to 500 nm.

Fig. 3.43 Measurement of the 30L-PhC normalised optical spectrum (beam on - beam off) as a function of the angle (top) and integrated over all angles (bottom), for the full wavelength range (left) and wavelengths 350 to 500 nm (right).

Beam phase 2 - lead-shielded CMOS setup

In the second data-taking week, the background mitigation measures, outlined in Section 3.6.1, were installed. In the beginning, the lead shielding was mounted, and after the second day, additional changes were made. The 28L-PhC was exclusively studied, and priority was given to measurements with bandpass filters between 400 and 450 nm, due to beam-time constraints.



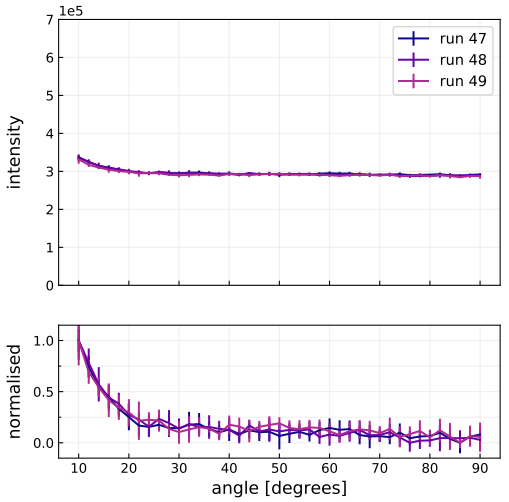
(a) 28L-PhC & 450 nm BPF. Lead-shield background reduction only.

(b) 28L-PhC & 450 nm BPF. Lead-shield and additional background reduction.

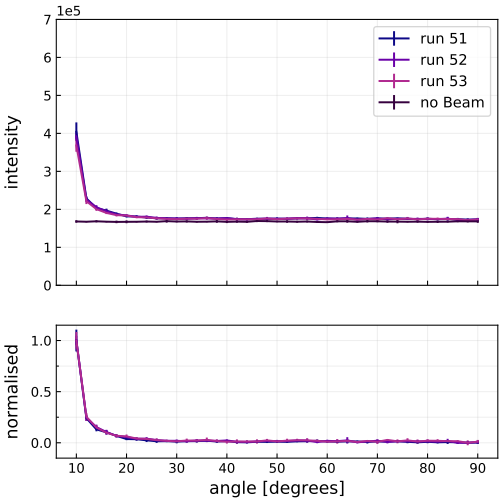
Fig. 3.44 Beam test phase 2, 28L-PhC CMOS camera measurements with a 450 nm bandpass filter: lead-shielding only (a) and additional background reduction measures (b).

Figures 3.44, 3.45, and 3.46 show, on the left, the camera measurement with lead-shielding only, and on the right, full background reduction measures. For the latter, an additional measurement of the background without beam was taken.
The intensity decreases more rapidly with the angle in comparison to the corresponding measurements of beam phase 1, reaching a steady plateau above 25°. The additional background mitigation leads to a reduction of the plateau, only slightly above the intensity of measurements without beam, representing the general background noise of the CMOS camera.

One of the two 375 nm runs (run 54), shown in Figure 3.47, features a peak at 26°. While the ten measurements taken by the camera at this angle vary significantly in their respective intensities, as indicated by the error bar, they are consistently higher than the baseline.
The missing drop from high intensities at 10° and the irreproducibility of the feature, shown

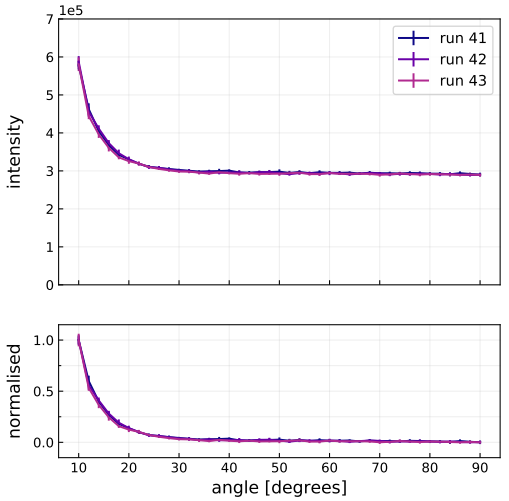


(a) 28L-PhC & 430 nm BPF. Lead-shield background reduction only.

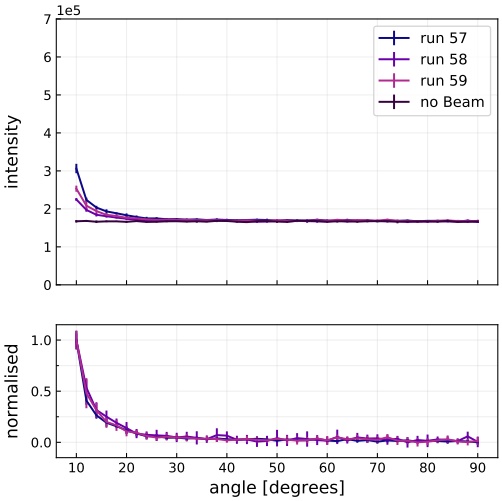


(b) 28L-PhC & 430 nm BPF. Lead-shield and additional background reduction.

Fig. 3.45 Beam test phase 2, 28L-PhC CMOS camera measurements with a 430 nm bandpass filter: lead-shielding only (a) and additional background reduction measures (b).



(a) 28L-PhC & 413 nm BPF. Lead-shield background reduction only.



(b) 28L-PhC & 413 nm BPF. Lead-shield and additional background reduction.

Fig. 3.46 Beam test phase 2, 28L-PhC CMOS camera measurements with a 413 nm bandpass filter: lead-shielding only (a) and additional background reduction measures (b).

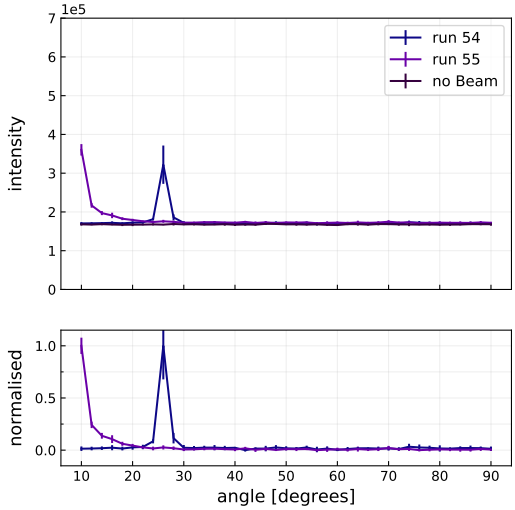


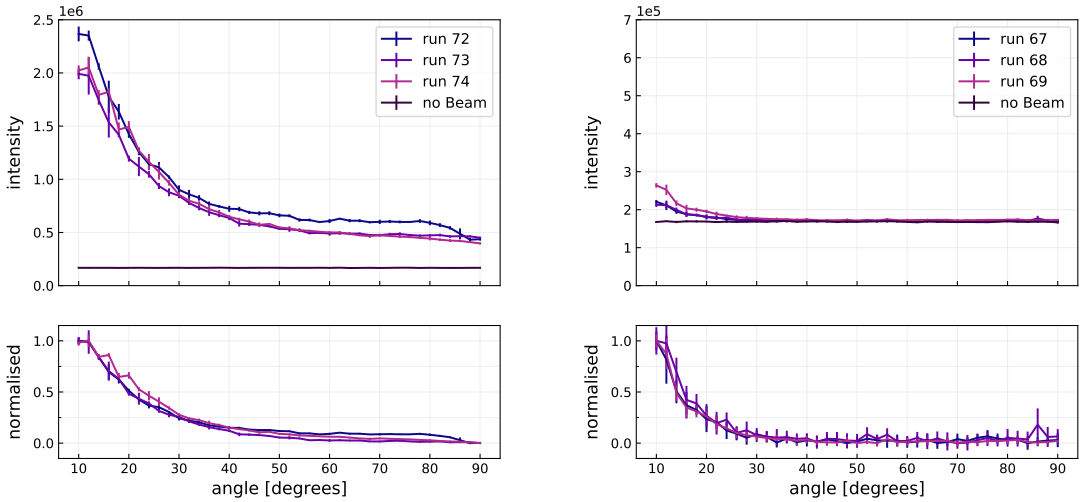
Fig. 3.47 Beam test phase 2, 28L-PhC CMOS camera measurements with a 375 nm bandpass filter and lead-shielding with additional background reduction measures. Run 54 shows an non-reproducible peak.

by the subsequent run 55, suggest a problem with the dataset for run 54 rather than evidence for a signal.

Beam phase 2 - background measurements

In the previous Section 3.6.2, background measurements with a bandpass filter and disabled beam were included in the corresponding figures. Two additional sets of measurements were taken to gain a better understanding of the background present in the experiment: runs without a bandpass filter, as shown in Figure 3.48, and runs with a tilted photonic crystal, presented in Figure 3.49.

The measurements without a bandpass filter were performed without and with a camera lens cover, shown in Figure 3.48a and 3.48b, respectively. Figure 3.48a illustrates the light measured from the photonic crystal, shaped only by the spectral response curve of the camera. With the camera lens covered, as for the data in Figure 3.48b, the observed intensity reduces by an order of magnitude. The shape and absolute intensity are still comparable to the bandpass filter measurements. While a plastic lens cover does not provide an absolute light-tightness, the measurements indicate that at least a fraction of the intensity recorded with the bandpass filters is not necessarily caused by light. Other potential sources are beam electrons traversing the sensor or the camera lens, producing background in the form of photons or electronic noise.



(a) 28L-PhC & no BPF.

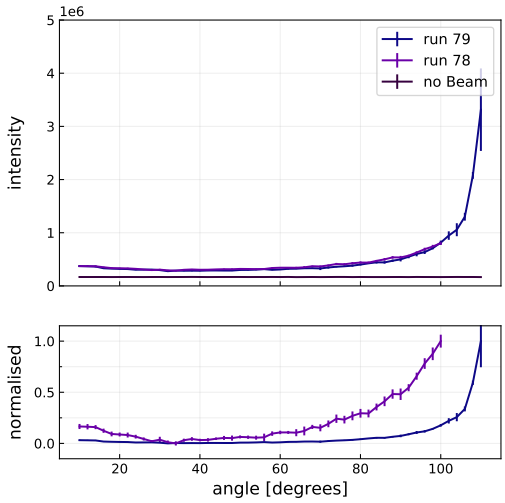
(b) 28L-PhC & no BPF. Camera covered.

Fig. 3.48 Beam test phase 2, 28L-PhC CMOS camera measurements without a bandpass filter (a) and with a cover on the camera (b).

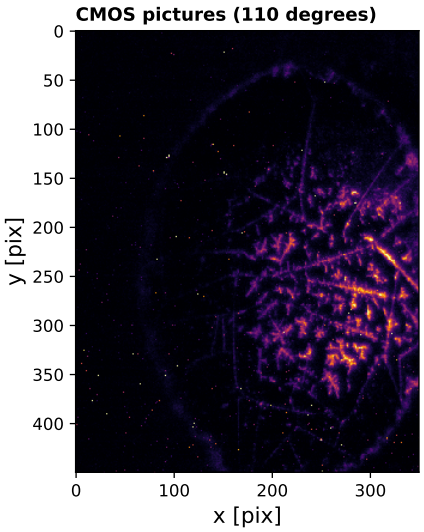
To understand better the background at angles close to the beam axis, the photonic crystal was tilted by 55° and measured without a bandpass filter. The angular range of the camera was increased up to 110°. The corresponding intensities are presented in Figure 3.49a.
The shape and intensity of the measurements are similar to the ones without a bandpass filter of Figure 3.48. For low angles, the intensities are similar to the intensities with a lens cover, shown in Figure 3.48b, which may correspond to the background generated by beam electrons.

The steep intensity increase towards large angles and the similarity to the measurements without a bandpass filter and lens cover suggest a large contribution of beam-generated background light reflected by the photonic crystal. This light is most likely produced by air-Cherenkov radiation in the 50 cm path between beam exit window and photonic crystal. Another contribution may be the scintillation light generated in the beam exit window. The background could in the future be explored by extending the GEANT4 simulation to the beam test setup.

Figure 3.49b shows the light captured by the camera at 110°. This arrangement is intended to be approximately equivalent to a hypothetical camera at 0° for the conventional setup with the photonic crystal orthogonal to the beam.



(a) Tilted 28L-PhC & no BPF.

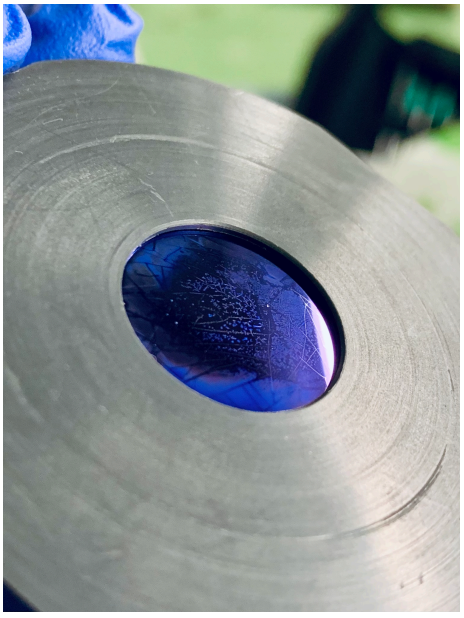


(b) Run 79 CMOS image at 110°

Fig. 3.49 The 28L-PhC is tilted by 55° and no BPF installed. At high camera angles, the photonic crystal acts like a mirror for the light generated by the beam. The corresponding CMOS pixel integration area at 110° is shown for run 79 (b).



(a) 28L-PhC.

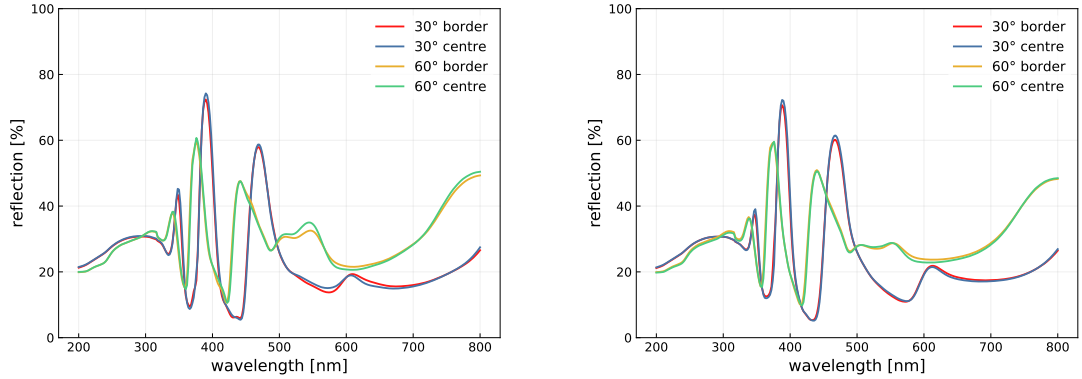


(b) 30L-PhC.

Fig. 3.50 Visible radiation damage of both photonic crystals, 28L-PhC (a) and 30L-PhC (b).

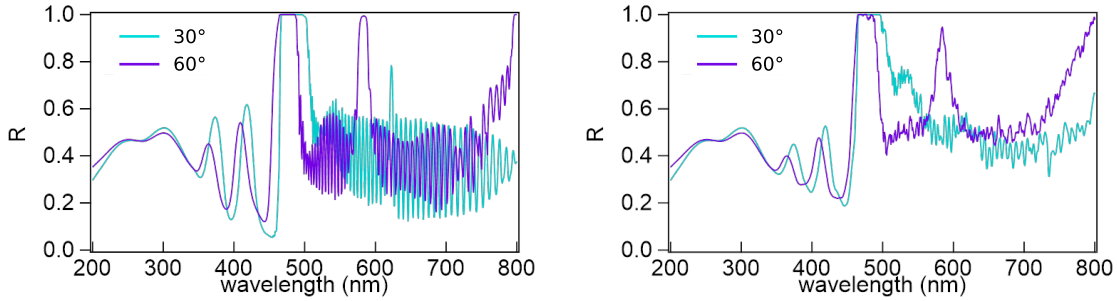
Reflection measurements

The two photonic crystals display visible deterioration due to the radiation, shown in Figure 3.50. Measurements of the reflection of both crystals are presented in Figure 3.51 at 30° and 60° angles of incidence. Although the centre regions were subject to considerably more radiation than the border region of the 30 mm photonic crystal disks due to the narrow beamwidth of approximately 3 to 5 mm, no significant difference between the two is observable.



(a) 28L-PhC. (b) 30L-PhC.

Fig. 3.51 Reflection measurements of the 28L-PhC (a) and 30L-PhC (b) at 30° and 60°, both at the centre and the border of the samples.



(a) 230 nm layer thickness,2 nm Gaussian error. (b) 230 nm layer thickness, 10 nm Gaussian error.

Fig. 3.52 Simulated reflection spectrum of a 30 layer $SiO_2|TiO_2$ photonic crystal with 230 nm layer thickness, at 30° (blue) and 60° (purple). The layer thickness is varied by a Gaussian thickness error of 2 nm (a) and 10 nm (b). Plots generated by [125].

In Figure 3.52, the corresponding simulated reflection spectra for a 30 layer $SiO_2|TiO_2$ photonic crystal with 230 nm layer thickness, at 30° (blue) and 60° and with a Gaussian

variation of the layer thickness of 2 nm (a) and 10 nm (b) are shown [125].
The reflection measurements and simulations show a consistent shift towards shorter wavelengths when going from 30° to 60° angle of incidence. When comparing measurements and simulations, similarities of the modes around 400 nm are visible. These may indicate increasing absorption in TiO_2 below 450 nm, with light penetrating only through the first few layer couples. The differences for longer wavelength suggest inconsistencies in the crystal structure below the top layers, either from variations in thickness or voids from delamination.

3.6.3 Interpretation and future experiments

Although no backward Cherenkov signal was detected, the studies provided a multitude of valuable insights which may lay the groundwork for future experiments. Potential improvements to the setup and its components are formulated below.

Photonic crystal improvements

Two main difficulties regarding the photonic crystal samples may harm a potential signal: the manufacturing and the radiation damage.

As shown in Figure 3.20, already a variation of above 1 % in layer thickness deteriorates the signal quality to a point where it is unlikely to be detectable. The reflection measurements, presented in Section 3.6.2, indicate inconsistencies of the crystal structure below the top layers. The manufacturing process utilised a quartz crystal monitor for a consistent thickness deposition and an automated shutter to interrupt the coating. While the coating uncertainty is expected to be minor, it may be enough to deteriorate the signal.
The precision of the layer deposition may be improved with an in-situ reflection measurement of the optical spectrum after each layer coating. This additional monitoring would allow for higher precision by comparing the measured spectrum to a simulated response.

The cracks caused by thermal stress may be mitigated by a controlled heat source and a temperature sensor on the coating substrate. Temperature fluctuations during the coating process may be reduced, and the duration of cooling to room temperature increased.

The third obstacle in manufacturing is that a specified refractive index must match the simulation requirements. As stated in Reference [124], especially the refractive index of TiO_2 depends on the stoichiometric ratio of the deposited titanium-oxide variants, as shown in Figure 3.21. The coating procedure used for the manufacturing of the two photonic crystal samples used Ti_3O_5 evaporation with a controlled oxygen influx to minimise deviation from the specified refractive index. An iterative process of measuring the refractive index of

a single coating layer under repeatable and controlled coating conditions and simulation adjustment could minimise deviations between the simulation and the real sample. In this study, no access to a facility with suitable instruments to measure the refractive index with sufficient precision was available. However, a future project would likely benefit from such a calibration.
Measurements with additional bandpass filters may allow for deviations of the sample's refractive index with respect to the simulation.

Ordering a sample from an industrial manufacturer may provide more minor uncertainties with respect to the specifications. This option would require careful consideration of the resulting costs and lead times.
While the configuration of the photonic crystals presented in this thesis resulted from an extensive search, a different composition may provide more robustness against uncertainties in the manufacturing process and less dispersive signals.

The second challenge concerning the photonic crystals is radiation robustness. During the beam test period, the two samples were subjected to a considerable amount of radiation. Figure 3.50 displays the two samples after the beam tests. The thermal tension cracks of the chromium layer, visible in Figure 3.25b and 3.26b, seem to have propagated to the surface. The outer surface area colour changed, and Lichtenberg patterns are visible in the centre where the beam was focussed.
Beam-induced electric discharges may cause the visible deterioration of the photonic crystals in the sample or by thermal stress due to the energy deposited in the sample. The damage was already visible after the first beam test phase and may negatively affect a potential effective Cherenkov radiation signal.

Experimental setup sensitivity

The second potential area of improvement is the sensitivity of the experimental beam test setup. Firstly, the measurements suggest a high amount of beam-generated background, as illustrated in Section 3.6.2. The reflection of in-air Cherenkov light from the surface of the photonic crystal and its holder seems to have generated a significant signal in the CMOS camera at angles below 25°. Although a backward effective Cherenkov signal may still be detected at higher angles, the background diminishes the probability to detect a signal in a considerable angular range.
Beam generated photons from Cherenkov or transition radiation may be avoided with an experimental setup in vacuum.A photon detector may either also be positioned in the

vacuum chamber or outside a beampipe window. The latter option would require considering reflection or absorption of the window material.

Despite the high number of electrons per bunch, the signal may be too low to being detected by the CMOS camera. Alternative detectors with single-photon sensitivity may be used, such as multi-anode photomultiplier tubes.

A comparatively simple change to the setup should be considered: the distance between the CMOS camera and the beam may be increased by a flat mirror positioned at the corner of an L-shaped rotational stage. That would reduce noise originating from scattered beam electrons.
Mounting the bandpass filters on a remotely controllable filter wheel would be an additional improvement to the setup. While not increasing its sensitivity, far fewer accesses to the beam area would be required, which would increase the data-taking time.

Analytical model

Lastly, the analytical model of effective Cherenkov radiation from resonant transition radiation may have to be revised. One possibility may be field interferences caused by the electron density in a beam bunch. The analytical calculations underlying the simulation assume a low electron current resulting in well-separated fields [113]. However, since a significant purpose of the experiment is to find evidence for the predictions of the simulation, any corresponding conclusion is not sensible until the uncertainties of the photonic crystal manufacturing and experimental setup have been considerably improved, as suggested above.

Future experiments

A future beam test experiment may implement the suggestions outlined above to increase the likelihood to observe backward Cherenkov radiation from a dielectric photonic crystal.
In summary, a suitable photonic crystal configuration has to be determined by simulation. An ideal configuration features radiation robust materials and minimal dispersion.
Secondly, the samples have to be manufactured with minimal uncertainties regarding the material specifications. The most decisive parameters are precise refractive indices, determined by the material stoichiometry during the crystal growing process and the layer thickness.
Thirdly, the signal to noise ratio of the experimental setup has to be increased. The signal sensitivity may be improved by single-photon sensitive photon detectors and a larger area of acceptance. However, noise reduction is the main priority. Beam induced background could likely be reduced to a minimum with a setup in vacuum.

Given the complexity of such a setup, an alternative may be to use the background-mitigated setup as presented in this thesis in combination with a suitable photonic crystal sample.
As the measurements from the second beam phase show, the beam-induced background may be reduced to the level of the CMOS sensor noise for angles above 25°. A photonic crystal configuration that produces an effective backward Cherenkov signal at higher angles could be studied with the presented experimental setup, given that the photon yield within the angular acceptance of the sensor exceeds its electronic noise.

3.7 Outlook for photonic crystals at CERN

This chapter introduced a framework to study effective Cherenkov radiation from photonic crystals at CERN. The commissioning of a laboratory setup was described using a polymer photonic crystal and a radioactive source. A ring-shaped signal was observed, resembling Cherenkov radiation, in the region expected by GEANT4 simulations.
An irradiation campaign was conducted to measure the deterioration of the polymer photonic crystal, concluding that the material is not suitable for use in the LHC environment.

A process to manufacture photonic crystals was described, relying exclusively on CERN facilities, together with analytical simulations of their expected behaviour under various scenarios. The first tests of one-dimensional dielectric photonic crystals in a high-energy particle beam were performed using the manufactured samples at the CLEAR beam facility. Although no evidence for backward Cherenkov light from resonance transition radiation was observed in the beam tests, the presented studies provided a wealth of important information.

Several challenges have to be addressed on the path from the first tests presented in this thesis to an application in a collider experiment. To name only a few, the photonic crystal radiator would have to consist of radiation-robust compounds with a low overall material budget, generate a sufficient number of photons, and feature low material loss and dispersion. Strategies to overcome these obstacles have been discussed [100] and entail using optically transparent materials to reduce loss, limiting the frequency range in signal detection to minimise chromatic aberration effects, and improving the photon yield through increasing the photonic crystal thickness or by using two- or three-dimensional crystal structures [99]. The radiation hardness and material budget are well-established aspects in high energy experiments with abundant existing expertise. A mechanism to obtain photonic crystal particle identification without obstructing the particle trajectory has been proposed in [126], to which I contributed as co-author and is summarised in Appendix B.

Constructing a particle identification detector with photonic crystal radiators would enable novel designs. The ability to control effective Cherenkov radiation in the forward and backward direction enables addressing a wider momentum range by separating momenta into different regimes through the direction of the emitted photons. Curved structures can be used to achieve an angle close to orthogonal between the charged particle and photonic crystal layers, either through flexible photonic crystal compounds [127–129] or by applying thin-layer deposition methods to curved substrates.

The work in this thesis demonstrates the ability to manufacture, simulate, and study photonic crystals under high energy beam conditions at CERN. The cost-efficiency and rapid turnaround of this infrastructure may represent a valuable asset for CERN's R&D programmes and could prove indispensable for future studies to explore the potential of photonic crystals in hadron collider experiments.

Appendix A

RICH PID CNN figures

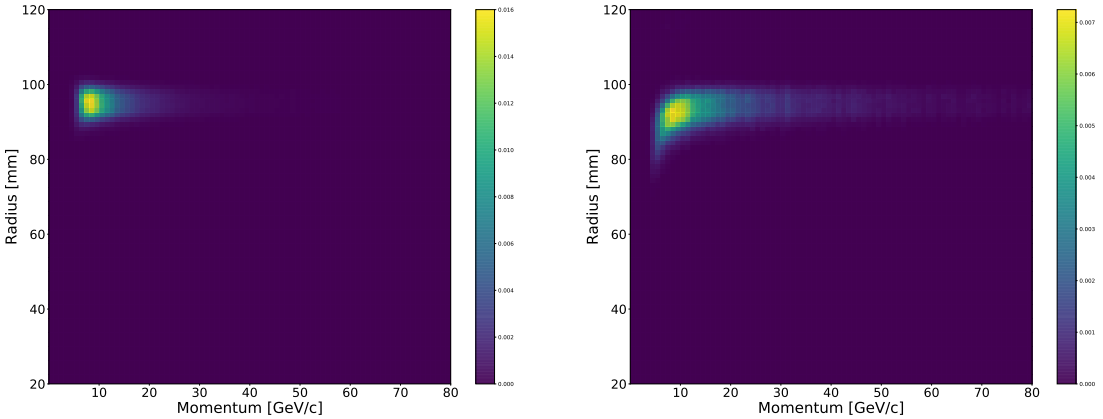
This appendix contains figures related to the RichNet chapter of this thesis (Chapter 2).

A.1 Data visualisation figures

Figure A.1 and A.2 show the Cherenkov radii versus momentum for the individual particles in RICH 1 and RICH 2 respectively.

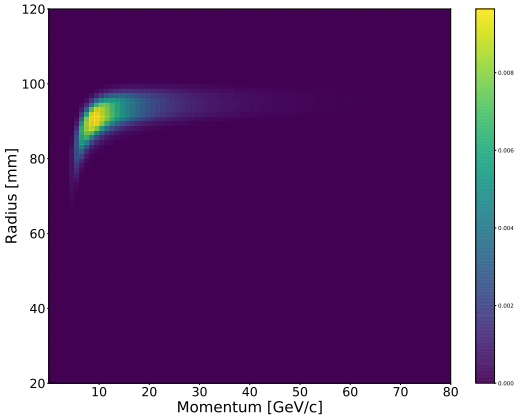
A.2 Feature engineering figures

Figure A.4 shows the distance between the extrapolated track centre and associated Cherenkov photons for each momentum bin are shown for both RICH 1 and RICH 2. The corresponding radius search ranges are noted in Table A.1. The top part of the plot shows the absolute incidence of the particle for this radius range, and the bottom one the normalised histogram for each particle type. Electrons are coloured in blue, muons in orange, pions in green, kaons in red, protons in purple, and deuterons in brown.

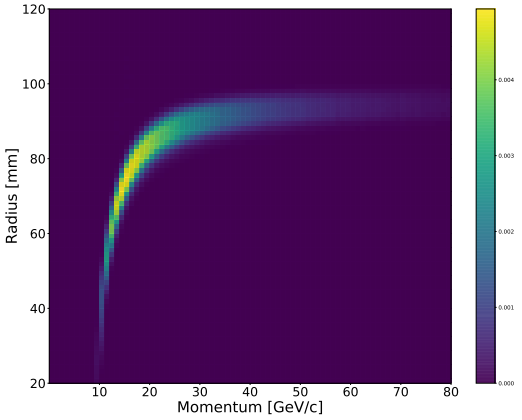


(a) Electron radii.

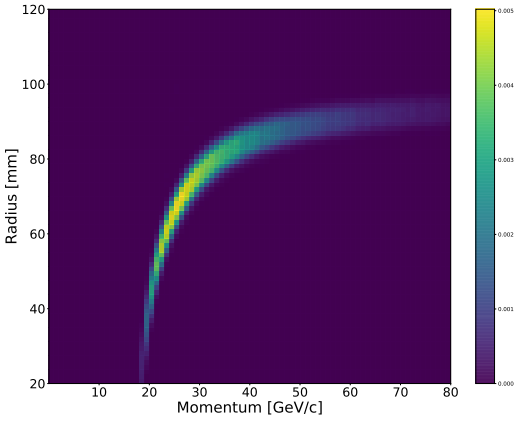
(b) Muon radii.



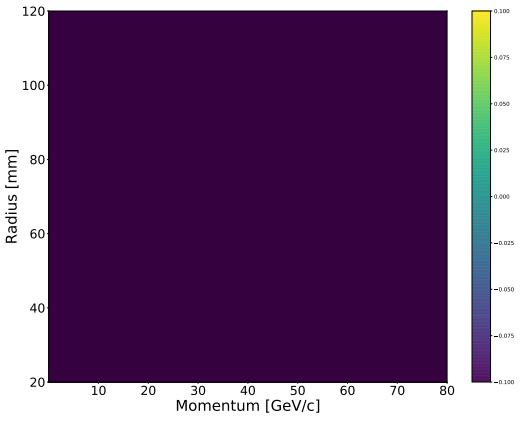
(c) Pion radii.



(d) Kaon radii.

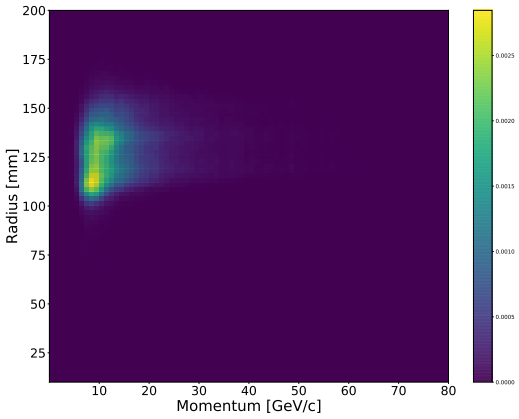


(e) Proton radii.

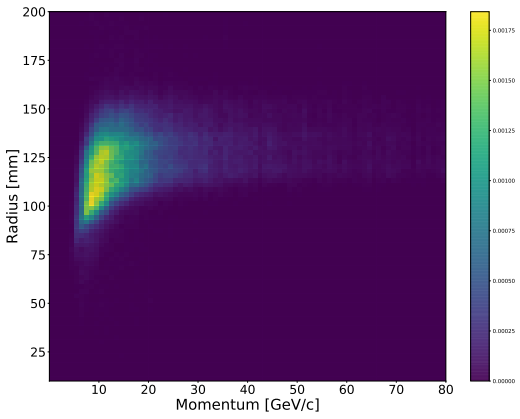


(f) Deuteron radii.

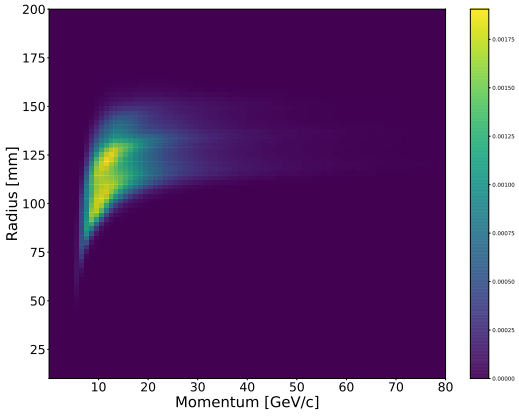
Fig. A.1 Rich 1 Cherenkov radii vs. momentum for the individual particle species in the dataset.



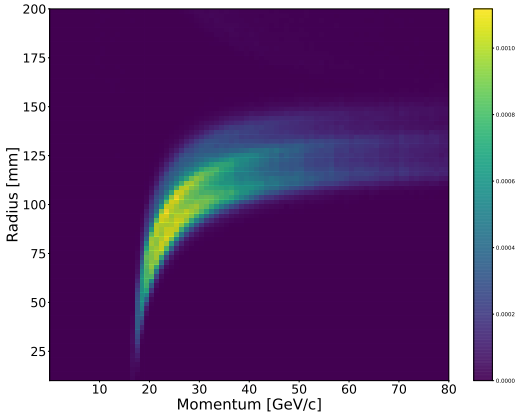
(a) Electron radii.



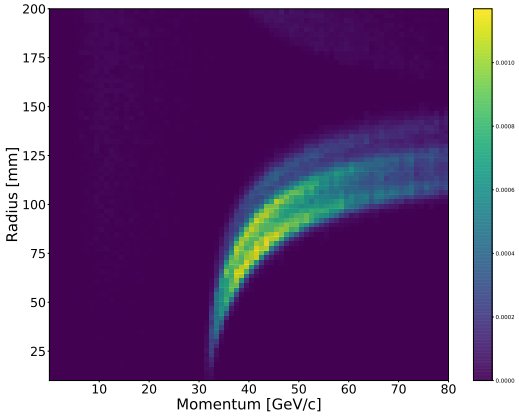
(b) Muon radii.



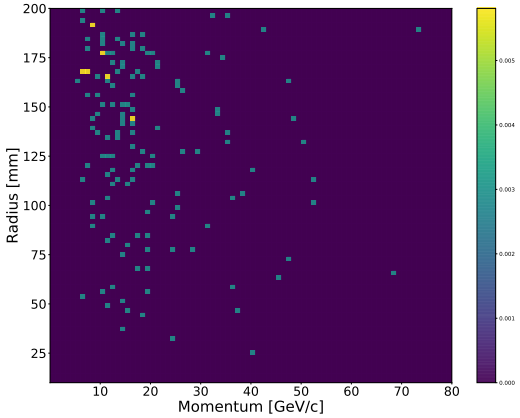
(c) Pion radii.



(d) Kaon radii.



(e) Proton radii.

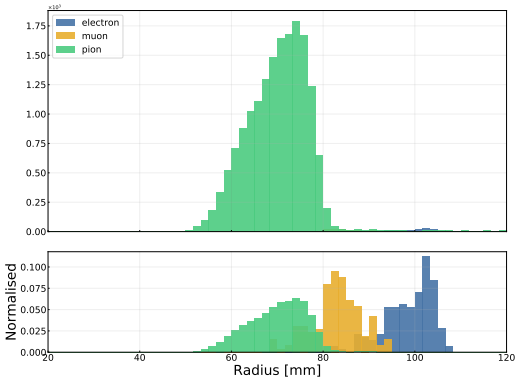


(f) Deuteron radii.

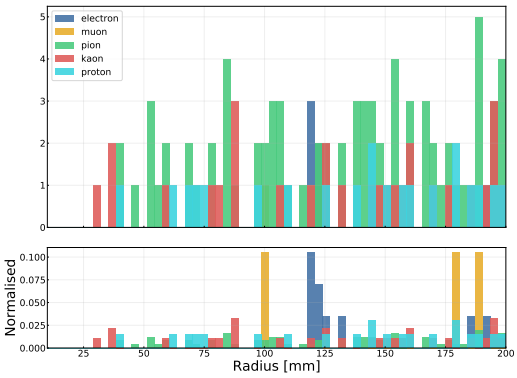
Fig. A.2 Rich 2 Cherenkov radii vs. momentum for the individual particle species in the dataset.

Momentum range [GeV/c]	RICH 1 [mm]	RICH 2 [mm]
2-3	20-50 & 80-104	20-160
3-4	50-104	20-160
4-5	30-104	20-160
5-6	30-104	30-160
6-7	30-104	70-160
7-8	30-104	70-160
8-9	45-65 & 75-104	80-160
9-10	20-40 & 75-104	80-160
10-11	40-55 & 75-104	80-160
11-12	45-65 & 75-104	80-160
12-13	50-70 & 80-104	80-160
13-14	50-70 & 80-104	80-160
14-15	55-104	90-160
15-16	60-104	90-160
16-17	60-104	10-160
17-18	10-40 & 65-104	20-160
18-19	10-40 & 70-104	40-160
19-20	20-50 & 70-104	50-160
20-25	30-104	60-160
25-30	30-104	60-160
30-40	65-104	20-160
40-50	70-104	70-160
50-60	80-104	80-170
60-70	80-104	80-170
70-80	80-104	100-179

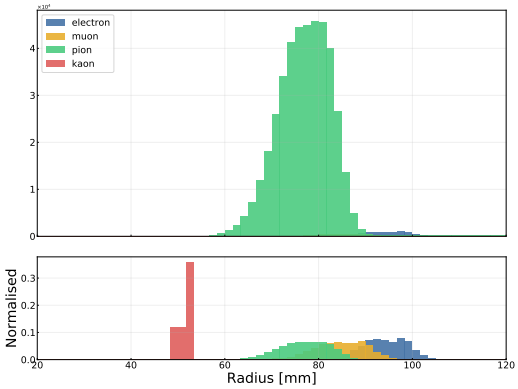
Table A.1 Table of search range radii for RICH 1 and RICH 2.



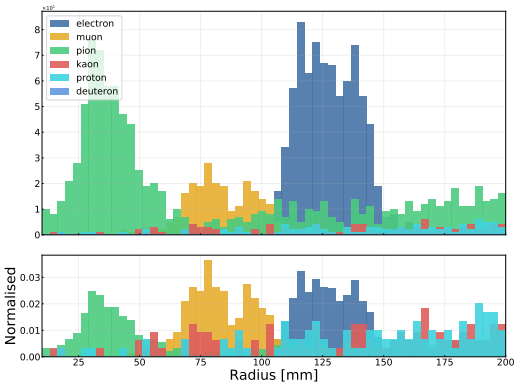
(a) RICH 1, radii for 3 to 4 GeV/c.



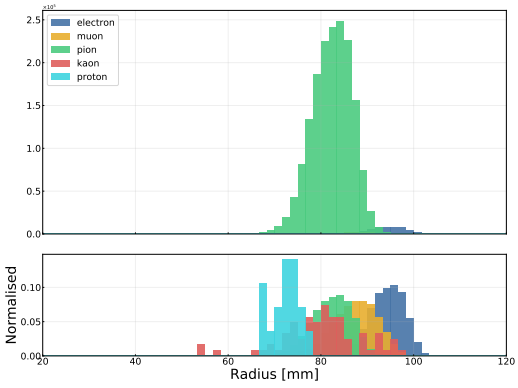
(b) RICH 2, radii for 3 to 4 GeV/c.



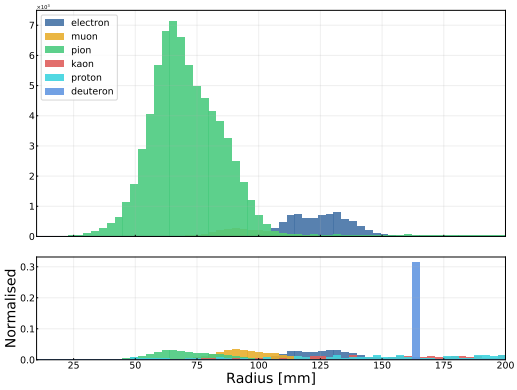
(c) RICH 1, radii for 4 to 5 GeV/c.



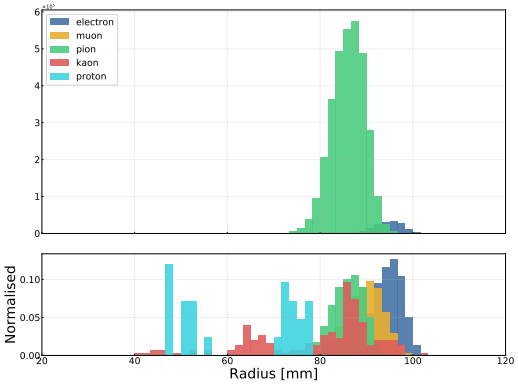
(d) RICH 2, radii for 4 to 5 GeV/c.



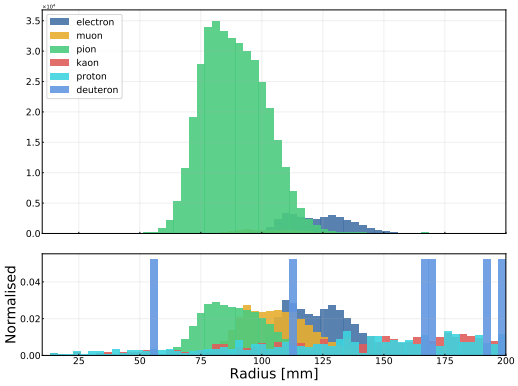
(e) RICH 1, radii for 5 to 6 GeV/c.



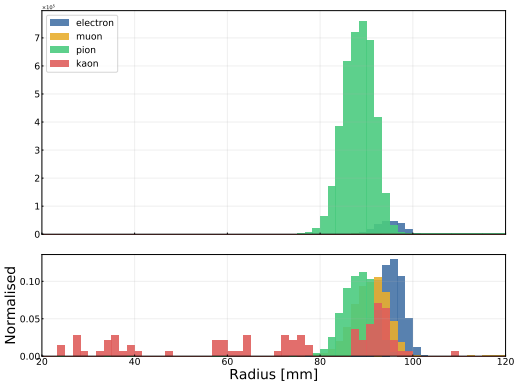
(f) RICH 2, radii for 5 to 6 GeV/c.



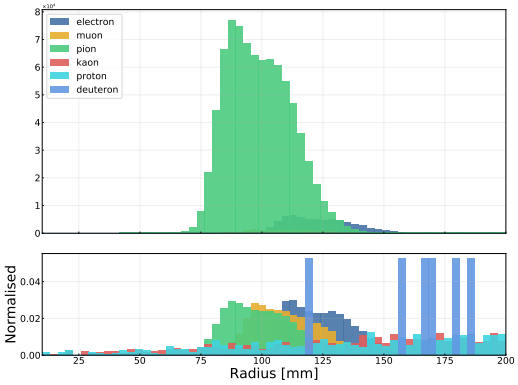
(a) RICH 1, radii for 6 to 7 GeV/c.



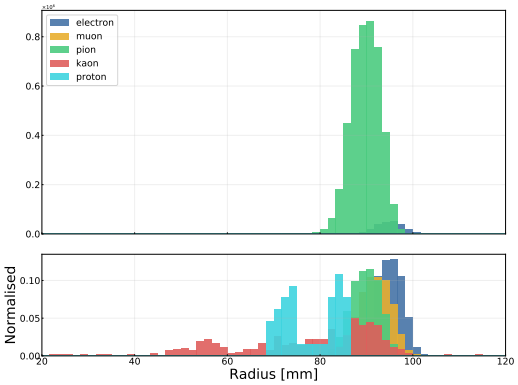
(b) RICH 2, radii for 6 to 7 GeV/c.



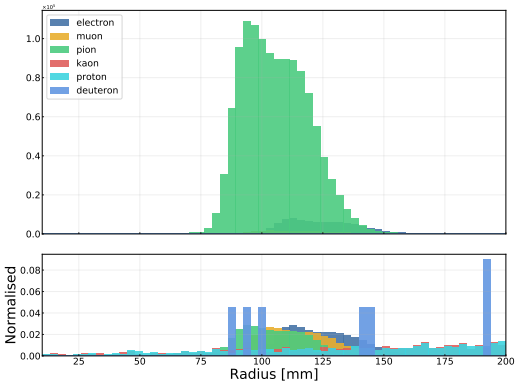
(c) RICH 1, radii for 7 to 8 GeV/c.



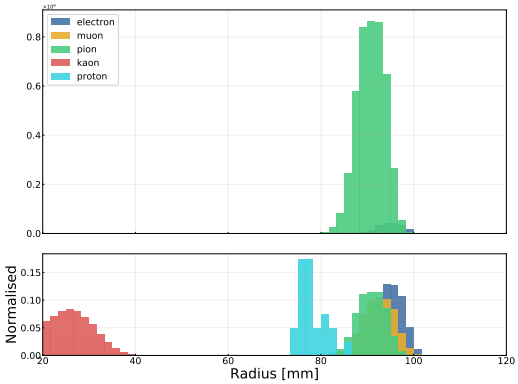
(d) RICH 2, radii for 7 to 8 GeV/c.



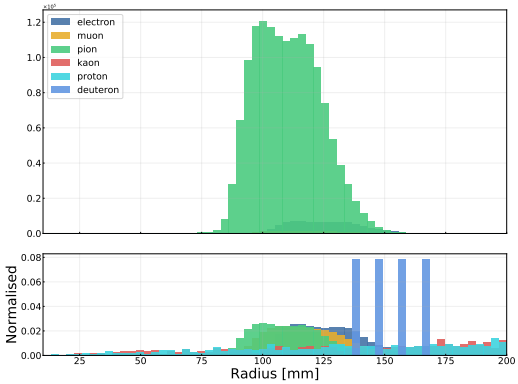
(e) RICH 1, radii for 8 to 9 GeV/c.



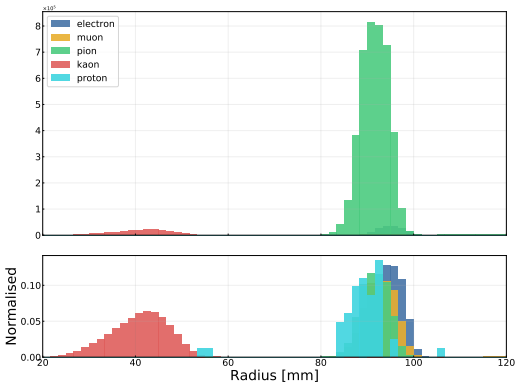
(f) RICH 2, radii for 8 to 9 GeV/c.



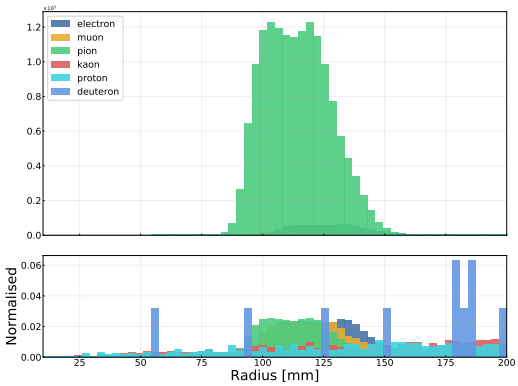
(g) RICH 1, radii for 9 to 10 GeV/c.



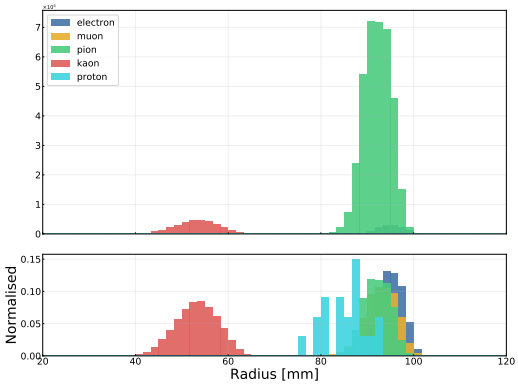
(h) RICH 2, radii for 9 to 10 GeV/c.



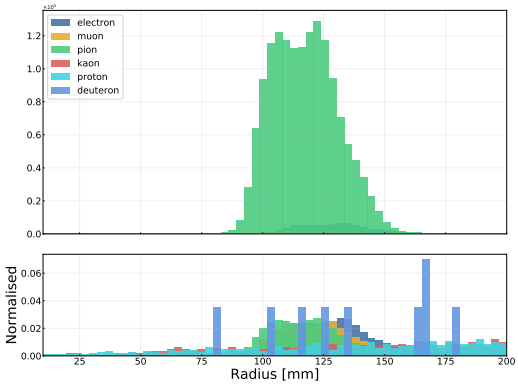
(i) RICH 1, radii for 10 to 11 GeV/c.



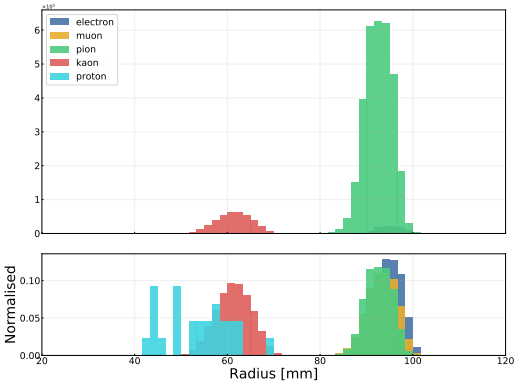
(j) RICH 2, radii for 10 to 11 GeV/c.



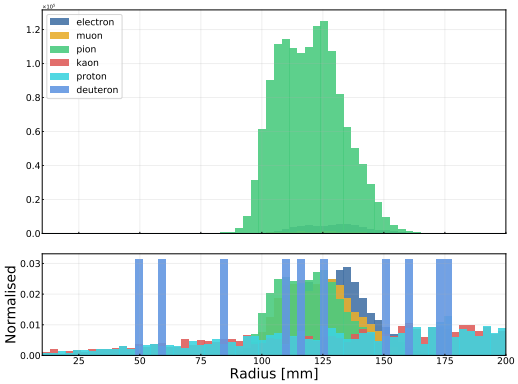
(k) RICH 1, radii for 11 to 12 GeV/c.



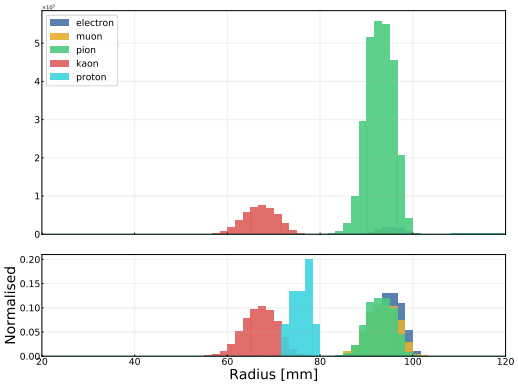
(l) RICH 2, radii for 11 to 12 GeV/c.



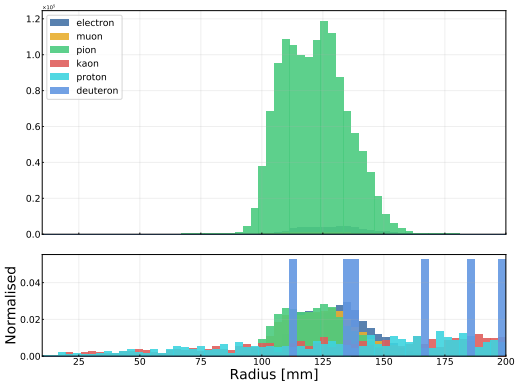
(m) RICH 1, radii for 12 to 13 GeV/c.



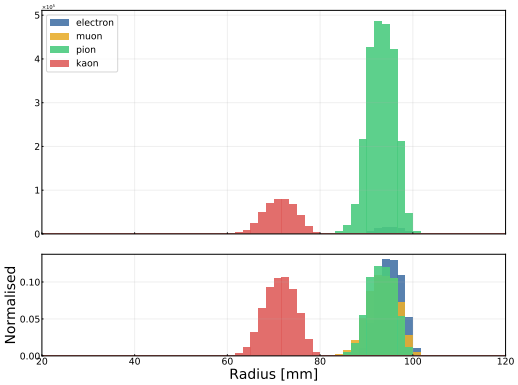
(n) RICH 2, radii for 12 to 13 GeV/c.



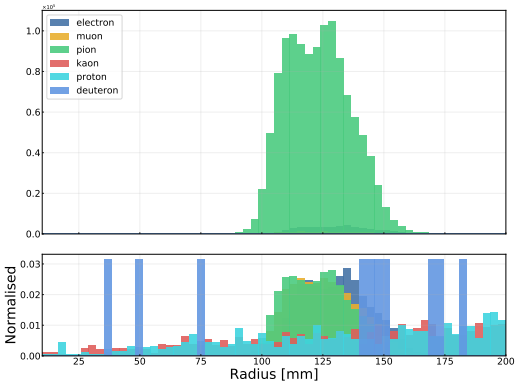
(o) RICH 1, radii for 13 to 14 GeV/c.



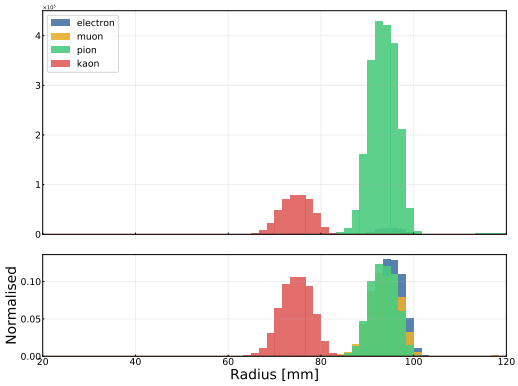
(p) RICH 2, radii for 13 to 14 GeV/c.



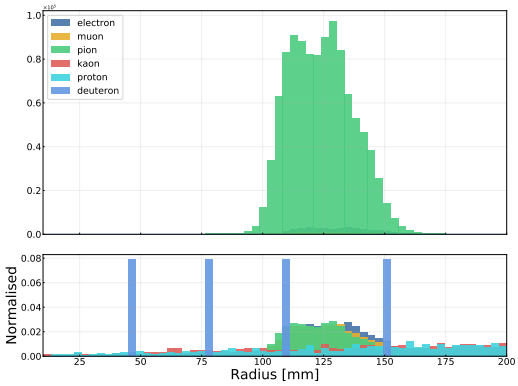
(q) RICH 1, radii for 14 to 15 GeV/c.



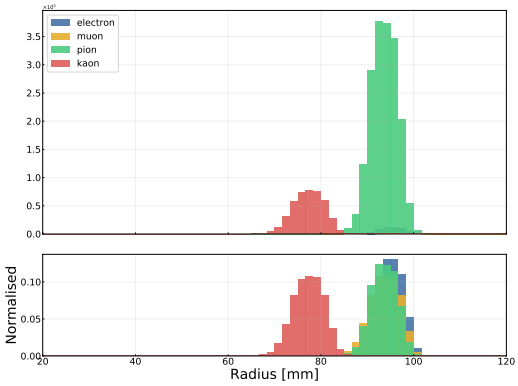
(r) RICH 2, radii for 14 to 15 GeV/c.



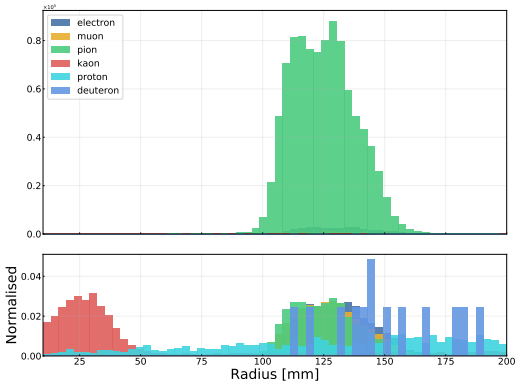
(s) RICH 1, radii for 15 to 16 GeV/c.



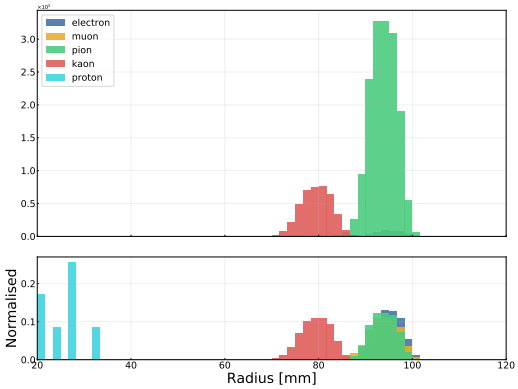
(t) RICH 2, radii for 15 to 16 GeV/c.



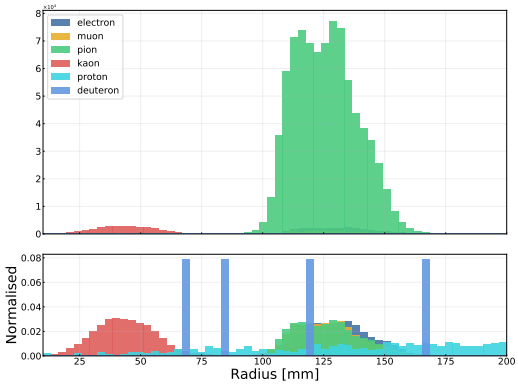
(u) RICH 1, radii for 16 to 17 GeV/c.



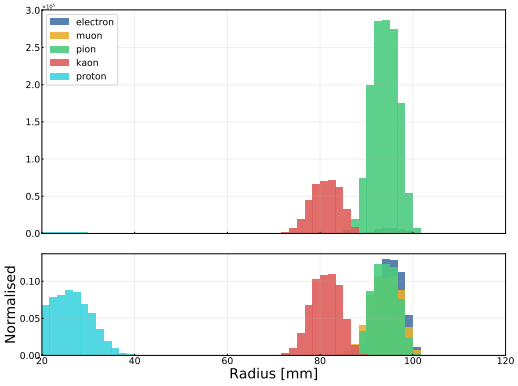
(v) RICH 2, radii for 16 to 17 GeV/c.



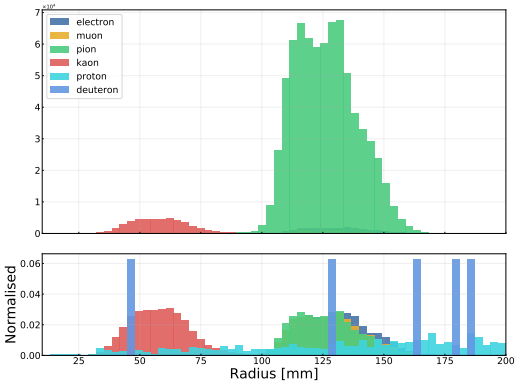
(w) RICH 1, radii for 17 to 18 GeV/c.



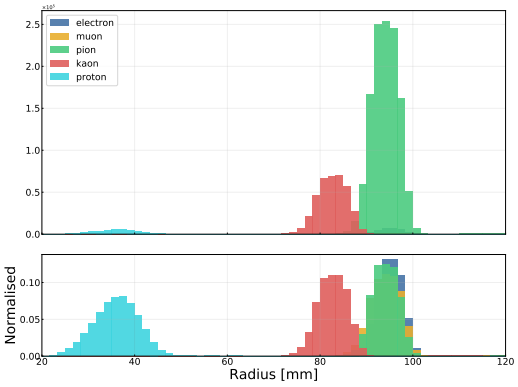
(x) RICH 2, radii for 17 to 18 GeV/c.



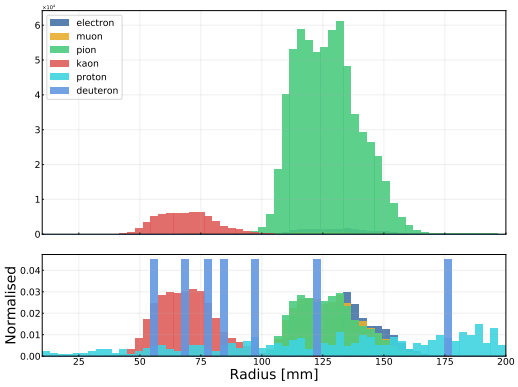
(y) RICH 1, radii for 18 to 19 GeV/c.



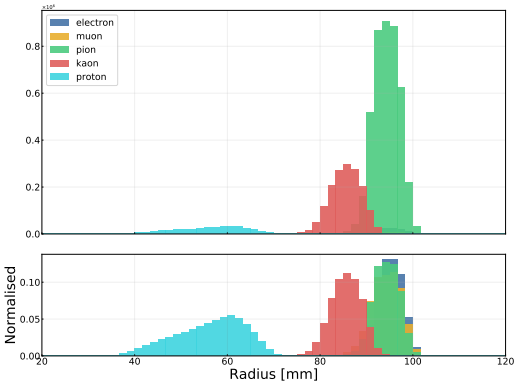
(z) RICH 2, radii for 18 to 19 GeV/c.



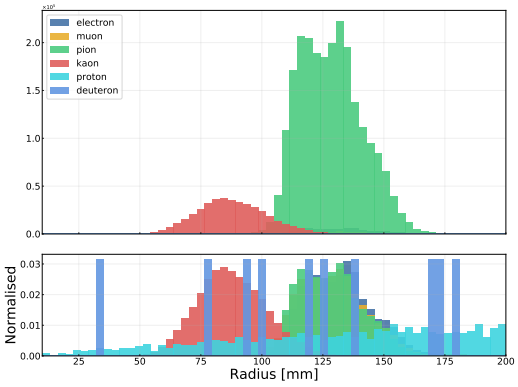
(aa) RICH 1, radii for 19 to 20 GeV/c.



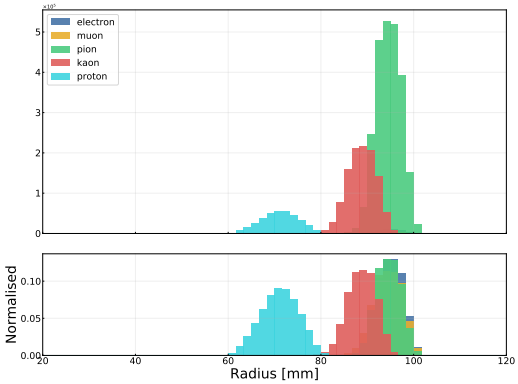
(ab) RICH 2, radii for 19 to 20 GeV/c.



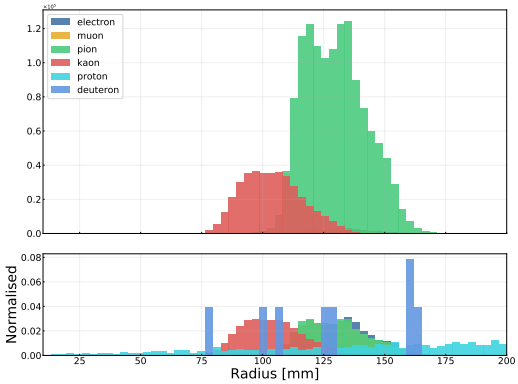
(ac) RICH 1, radii for 20 to 25 GeV/c.



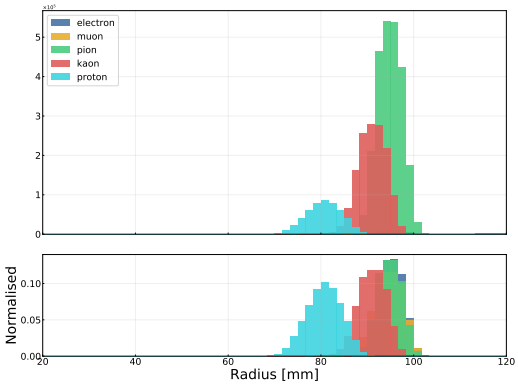
(ad) RICH 2, radii for 20 to 25 GeV/c.



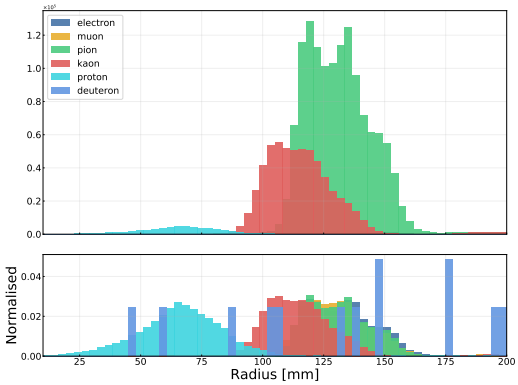
(ae) RICH 1, radii for 25 to 30 GeV/c.



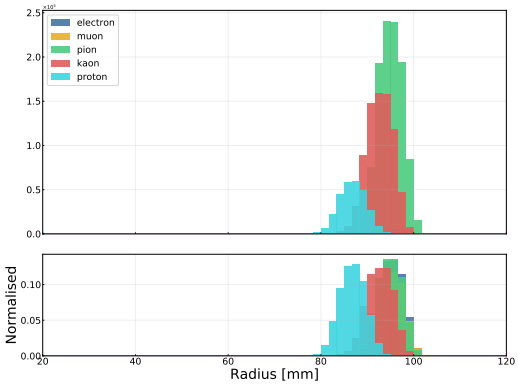
(af) RICH 2, radii for 25 to 30 GeV/c.



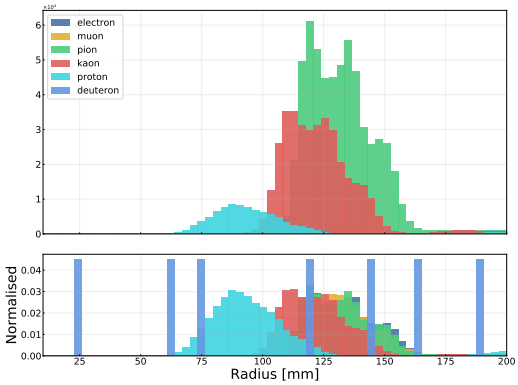
(ag) RICH 1, radii for 30 to 40 GeV/c.



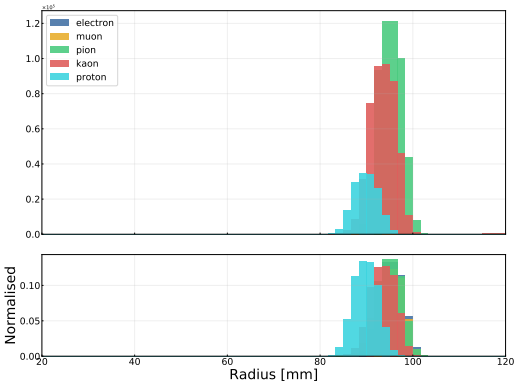
(ah) RICH 2, radii for 30 to 40 GeV/c.



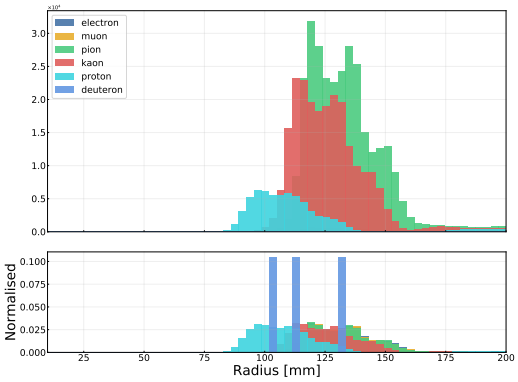
(ai) RICH 1, radii for 40 to 50 GeV/c.



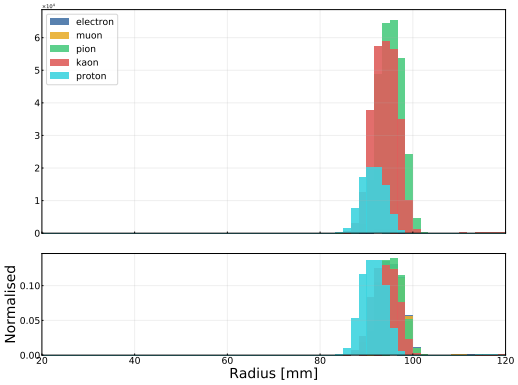
(aj) RICH 2, radii for 40 to 50 GeV/c.



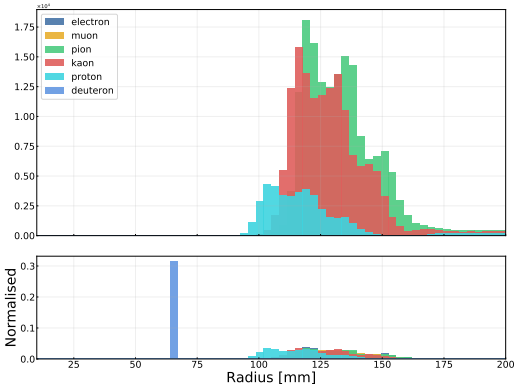
(ak) RICH 1, radii for 50 to 60 GeV/c.



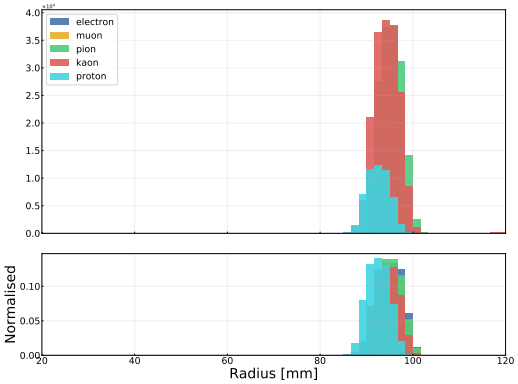
(al) RICH 2, radii for 50 to 60 GeV/c.



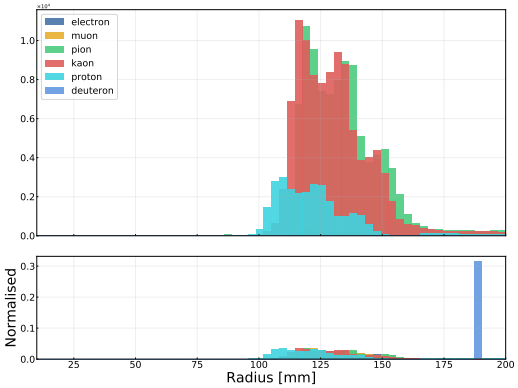
(am) RICH 1, radii for 60 to 70 GeV/c.



(an) RICH 2, radii for 60 to 70 GeV/c.



(ao) RICH 1, radii for 70 to 80 GeV/c.



(ap) RICH 2, radii for 70 to 80 GeV/c.

Fig. A.4 Histograms of distance in mm between the extrapolated track centres and MC Cherenkov photons, for all momentum ranges, in RICH 1 (left) and RICH 2 (right).

Appendix B

A Brewster route to PID

This appendix introduces an alternative approach to particle identification using the optical filtering properties of photonic crystals and the Brewster effect. It provides a summary of the published method to which I contributed as a coauthor [126].

A schematic of the working principle is presented in Figure B.1. The mechanism makes use of a broadband angular filter consisting of multiple stacks of one-dimensional photonic crystals. Conventional Cherenkov photons, generated by a charged particle travelling parallel to the filter surface and photon detection plane, pass the filter only with a precise incident angle, following the Brewster effect. The transmitted light is therefore p-polarised (transverse-magnetic), and all other photons are totally reflected. It should be stressed that there are no special requirements regarding the Cherenkov radiator material. It can hence be chosen to maximise the number of emitted photons.

By tailoring the bandgap structure of the photonic crystals, a non-dispersive pseudo refractive index can be achieved with angular selectivity over the entire spectrum of visible light. The light passing through the broadband filter hits the photon detection plane with a particular Brewster-Cherenkov angle θ_{BCR}, given by the angle between the particle trajectory and the component of the wavevector parallel to the detection plane, as shown in Figure B.1. Particles can be identified by measuring this angle, with sensitivity over an extensive momentum range and optical bandwidth.

Figure B.2 shows the intensity distribution on the photon detection plane for $2\,\mathrm{GeV/c}$ particles with and without the broadband photonic crystal filter. After traversing the filter, two tails of light are visible, at an incident angle θ_{BCR}, as indicated also in Figure B.1, which can be identified by the position of the peak intensities, as demonstrated in the top right plot.

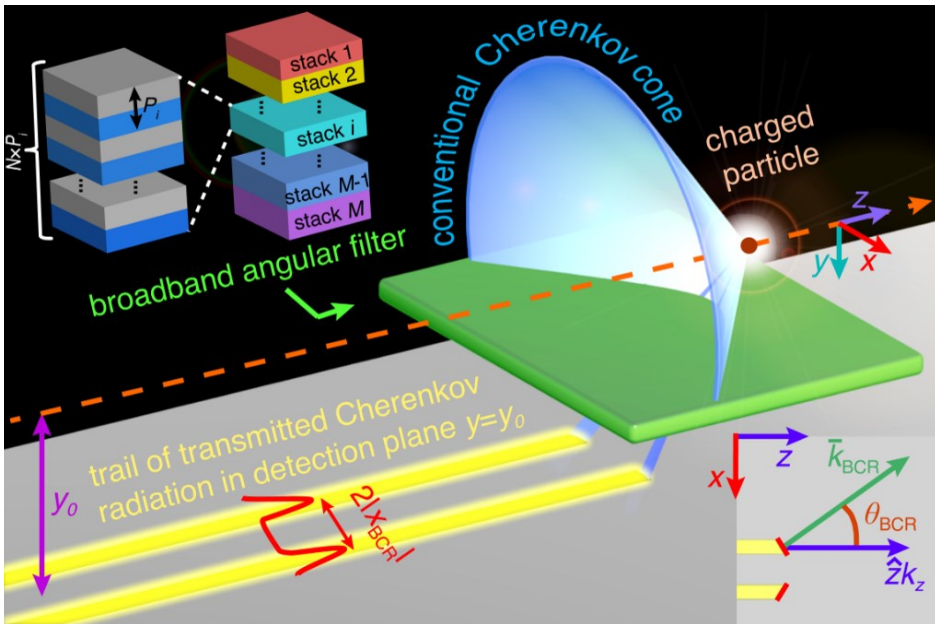


Fig. B.1 Schematic of the PID approach using the conventional Cherenkov effect and photonic crystals as broadband angular filters through the Brewster effect in multiple 1D photonic crystal stacks. The stacks differ in their periodic structure but feature identical materials. Source: [126].

The angle is given by

$$\cos\theta_{BCR} = \frac{k_z}{|\bar{k}_{BCR}|} \tag{B.1}$$

$$= \frac{1}{n_{BCR}\beta} \tag{B.2}$$

$\bar{k}_{BCR} = \hat{x}k_x + \hat{z}k_z$ is the tangential wavevector of the Cherenkov radiation transmitted through the filter, parallel to the detector plane. The component $k_z = \omega/v$ is given by the charged particle [100]. Following from the Brewster effect, $|\bar{k}_{BCR}| = \sqrt{\frac{\varepsilon_{r1}\varepsilon_{r2}}{\varepsilon_{r1}+\varepsilon_{r2}}}\frac{\omega}{c}$, with the relative permittivities of the photonic filter crystal as ε_{r1} and ε_{r2}. The pseudo refractive index of the filter is hence given by $n_{BCR} = \sqrt{\frac{\varepsilon_{r1}\varepsilon_{r2}}{\varepsilon_{r1}+\varepsilon_{r2}}}$.

The bandgap properties of photonic crystals allow tuning the pseudo refractive index n_{BCR} to achieve sensitivity in different momentum ranges. In Figure B.2, the intensity distribution for a pseudo refractive index of $n_{BCR} = 1.13$ is shown, which would be suitable for PID below 10 GeV/c. Pseudo refractive indices close to unity can be achieved using polymers to identify particles at a higher momentum range. Since the filter may be positioned far away from the particle beam, the materials do not have to be radiation robust. Figure B.3 shows

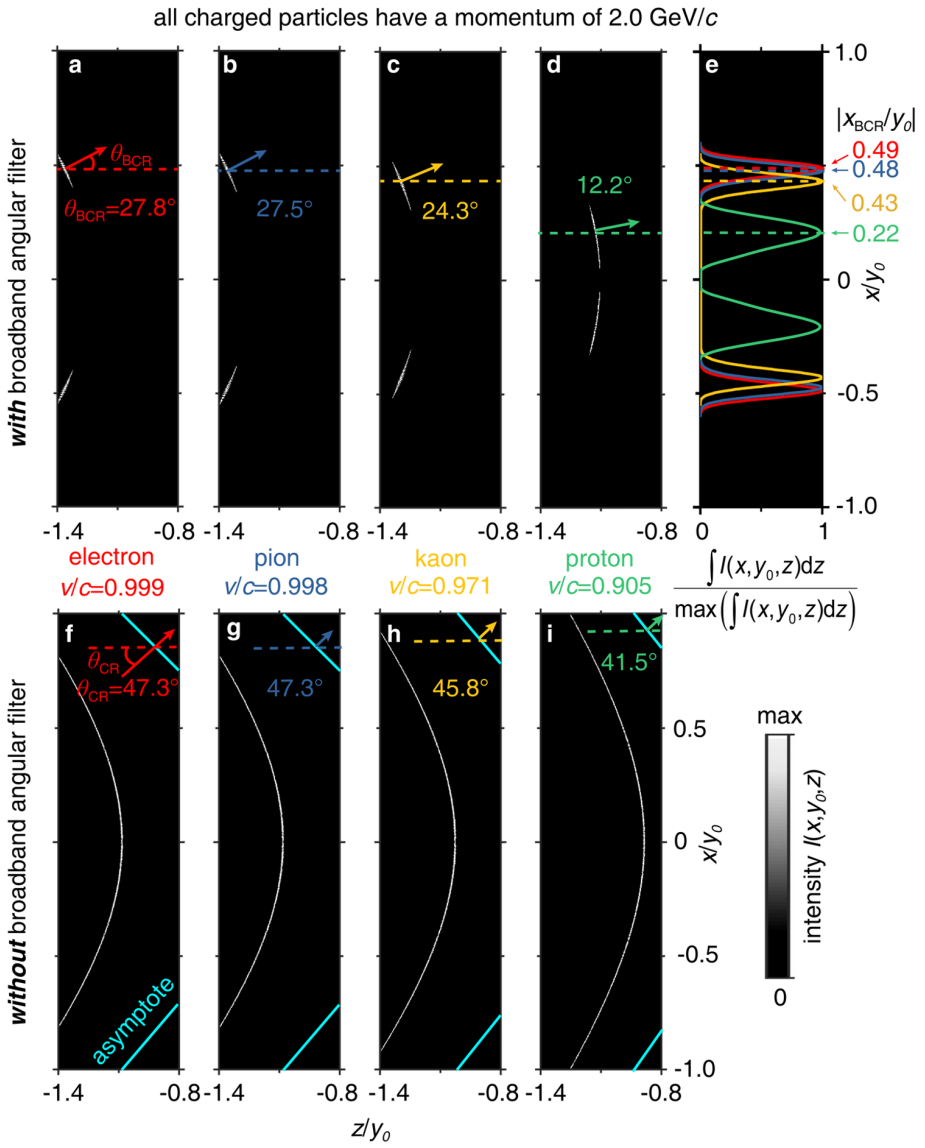


Fig. B.2 Intensity distribution on the photon detection plane for electrons, pion, kaons, and protons with a momentum of 2 GeV/c. The top distributions show the intensity with the broadband filter, the bottoms ones without. Source: [126].

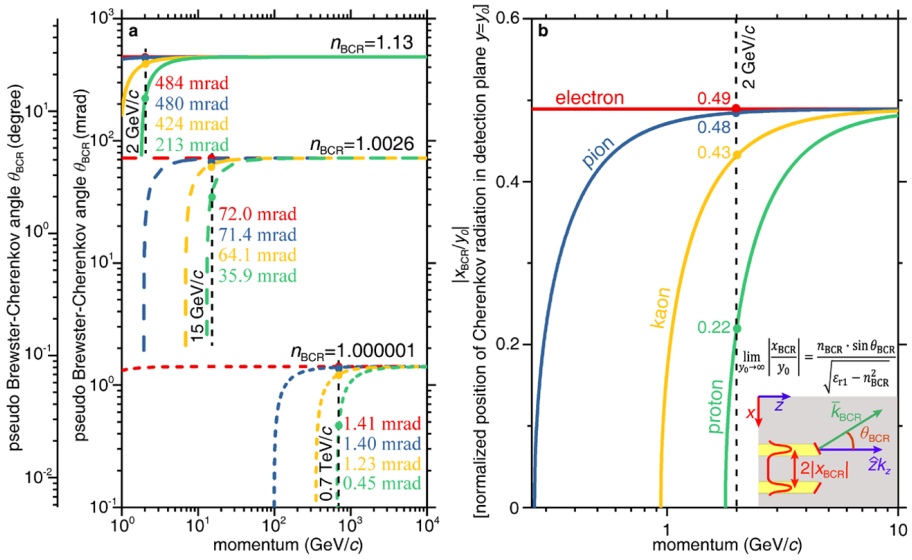


Fig. B.3 θ_{BCR} for different pseudo refractive indices of the photonic crystal filter, to identify particles at different momentum ranges (left) and the corresponding normalised distribution (right). Source: [126].

the pseudo-Brewster-Cherenkov angle θ_{BCR} for different pseudo refractive indices of the photonic crystal broadband filter.

While the constraint on the direction of travel of the particle parallel to the surface of the broadband filter is not suited for a RICH-like detection system, it may allow identification of charged particles from a narrow particle beam, even with different momenta and high flux. The studies presented in Chapter 3 do not focus on the Brewster-Cherenkov PID method but rather on the approach exploiting the constructive interference of resonance transition radiation in one-dimensional photonic crystals.

Bibliography

[1] Hasert, F. et al. "Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment." *Phys. Lett. B*, 46(1):138 – 140, 1973. doi:10.1016/0370-2693(73)90499-1.

[2] Arnison, G. et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV." *Phys. Lett. B*, 122(1):103 – 116, 1983. doi:10.1016/0370-2693(83)91177-2.

[3] Banner, M. et al. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider." *Phys. Lett. B*, 122(5):476 – 485, 1983. doi:10.1016/0370-2693(83)91605-2.

[4] Baur, G.a. et al. "Production of antihydrogen." *Phys. Lett. B*, 368(3):251–258, 1996. doi:10.1016/0370-2693(96)00005-6.

[5] CMS Collaboration. "A new boson with a mass of 125 GeV observed with the CMS experiment at the Large Hadron Collider." *Science*, 338(6114):1569–1575, 2012. doi:10.1016/j.physletb.2012.08.021.

[6] ATLAS Collaboration. "A particle consistent with the Higgs boson observed with the ATLAS detector at the Large Hadron Collider." *Science*, 338(6114):1576–1582, 2012. doi:10.1016/j.physletb.2012.08.020.

[7] Mobs, E. *The CERN accelerator complex*. CERN-GRAPHICS-2019-002-1, 2019. `https://cds.cern.ch/record/2684277`, accessed: 2020-12-24.

[8] ATLAS Collaboration. "The ATLAS experiment at the CERN large hadron collider." *JINST*, 3: S08003, 2008. doi:10.1088/1748-0221/3/08/S08003.

[9] CMS Collaboration. "The CMS experiment at the CERN LHC." *JINST*, 803:S08004, 2008. doi:10.1088/1748-0221/3/08/S08004.

[10] ALICE Collaboration. "The ALICE experiment at the CERN LHC." *JINST*, 3(08):S08002, 2008. doi:10.1088/1748-0221/3/08/S08002.

[11] LHCb Collaboration. "The LHCb detector at the LHC." *JINST*, 3(08):S08005, 2008. doi:10.1088/1748-0221/3/08/S08005.

[12] Apollinari, G. et al. "High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1." 2017. doi:10.23731/CYRM-2017-004.

[13] Anderson, C.D. "The Positive Electron." *Phys. Rev.*, 43(6):491, 1933. doi:10.1103/PhysRev.43.491.

[14] Sakharov, A.D. "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe." *Sov. Phys. Usp.*, 34(5):392–393, 1991. doi:10.1070/pu1991v034n05abeh002497.

[15] Christenson, J.H. et al. "Evidence for the 2 π Decay of the K 2 0 Meson." *Phys. Rev. Lett.*, 13 (4):138, 1964. doi:10.1103/PhysRevLett.13.138.

[16] Tanabashi, M. et al. "Review of Particle Physics." *Phys. Rev. D*, 98:030001, Aug 2018. doi:10.1103/PhysRevD.98.030001.

[17] Wolfenstein, L. "Parametrization of the Kobayashi-Maskawa Matrix." *Phys. Rev. Lett.*, 51: 1945–1947, 1983. doi:10.1103/PhysRevLett.51.1945.

[18] LHCb Collaboration. "Physics case for an LHCb Upgrade II: Opportunities in flavour physics, and beyond, in the HL-LHC era." *CERN-LHCb-PUB-2018-009*, 2018. `https://cds.cern.ch/record/2320509`.

[19] Belle II Collaboration. "The Belle II Experiment." *Nucl. and Part. Phys. Proc.*, 260:233 – 237, 2015. doi:10.1016/j.nuclphysbps.2015.02.049.

[20] Bevan, A. et al. "The Physics of the B Factories." *Springer Nature*, 2017. doi:10.1140/epjc/s10052-014-3026-9.

[21] LHCb Collaboration. "LHCb detector performance." *Int. Journ. of Mod. Phys. A*, 30(07): 1530022, 2015. doi:10.1142/S0217751X15300227.

[22] Brice, M. *View of the detector LHCb.* Aug 2008. `http://cds.cern.ch/record/1124325`, accessed: 2021-01-07.

[23] LHCb Collaboration. "LHCb Tracker Upgrade Technical Design Report." (CERN-LHCC-2014-001), Feb 2014. `https://cds.cern.ch/record/1647400`.

[24] LHCb Collaboration. *Standard set of performance numbers*. 2020. `https://lhcb.web.cern.ch/speakersbureau/html/PerformanceNumbers.html`, accessed: 2021-01-07.

[25] LHCb Collaboration. "LHCb VELO Upgrade Technical Design Report." *CERN-LHCC-2013-021*, Nov 2013. `https://cds.cern.ch/record/1624070`.

[26] LHCb Collaboration. "LHCb VELO upgrade." *Nucl. Instrum. Meth. A*, 845:97 – 100, 2017. doi:10.1016/j.nima.2016.04.077. Proceedings of the Vienna Conference on Instrumentation 2016.

[27] Poikela, T. et al. "VeloPix: the pixel ASIC for the LHCb upgrade." *JINST*, 10(01): C01057–C01057, Jan 2015. doi:10.1088/1748-0221/10/01/c01057.

[28] Williams, M. *VELO Upgrade II: Overview*. Mar 2018. https://indico.in2p3.fr/event/16795/contributions/60724/, accessed: 2021-01-08.

[29] Leverington, B. "SciFi R&D for Upgrade II." *4th Workshop on LHCb Upgrade II*, April 2019. https://indico.cern.ch/event/790856/contributions/3368476/. accessed: 2021-01-09.

[30] Hopchev, P. "SciFi: A large Scintillating Fibre Tracker for LHCb." *Proceedings of the Fifth Annual LHCP*, 2017. https://arxiv.org/abs/1710.08325.

[31] LHCb Collaboration. "Performance of the LHCb RICH detector at the LHC." *Eur. Phys. J. C*, 73(5):1–17, 2013. doi:10.1140/epjc/s10052-013-2431-9.

[32] Papanestis, A. and D'Ambrosio, C. "Performance of the LHCb RICH detectors during the LHC Run II." *Nucl. Instrum. Meth. A*, 876:221–224, 2017. doi:10.1016/j.nima.2017.03.009.

[33] LHCb Collaboration. "LHCb PID Upgrade Technical Design Report." *CERN-LHCC-2013-022*, (CERN-LHCC-2013-022. LHCB-TDR-014), Nov 2013. https://cds.cern.ch/record/1624074.

[34] Gambetta, S. *RICH 2 A-side*. Jul 2020, Private communication.

[35] Blago, M.P. "The LHCb RICH Upgrade." *PoS*, ICHEP2018:485. 4 p, 2019. doi:10.22323/1.340.0485.

[36] Forty, R. "The TORCH project." *JINST*, 9(04):C04024–C04024, apr 2014. doi:10.1088/1748-0221/9/04/c04024.

[37] Harnew, N. et al. "Status of the TORCH time-of-flight project." *Nucl. Instrum. Meth. A,*, 952: 161692, 2020. doi:10.1016/j.nima.2018.12.007. RICH 2018 Conf.

[38] Gys, T. et al. "Latest results from the TORCH R&D Project." *Nucl. Instrum. Meth. A*, 912, 10 2017. doi:10.1016/j.nima.2017.10.034.

[39] Guz, Y. "LHCb Calorimeter Upgrade." (LHCb-PROC-2013-051. CERN-LHCb-PROC-2013-051):8 p, Sep 2013. https://cds.cern.ch/record/1602198.

[40] ATLAS Collaboration. "ATLAS tile calorimeter: Technical Design Report." 1996. https://cds.cern.ch/record/331062.

[41] Pereima, D. "Calibration of the LHCb calorimetric system." *JINST*, 12(06):C06016–C06016, jun 2017. doi:10.1088/1748-0221/12/06/c06016.

[42] Alves Jr, A.A. et al. "Performance of the LHCb muon system." *JINST*, 8(02):P02022, 2013. doi:10.1088/1748-0221/8/02/P02022.

[43] Cardini, A. "The LHCb Muon Upgrade." *JINST*, 9(02):C02014–C02014, feb 2014. doi:10.1088/1748-0221/9/02/c02014.

[44] LHCb Collaboration. *Trigger Schemes*. `http://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html`, accessed: 2021-01-30.

[45] LHCb Collaboration. "A Comparison of CPU and GPU implementations for the LHCb Experiment Run 3 Trigger." *tbs. to Computing Software for Big Science*, 2021. doi:arXiv:2105.04031.

[46] LHCb Collaboration. "The LHCb simulation application, Gauss: design, evolution and experience." 331(3):032023, 2011. doi:10.1088/1742-6596/331/3/032023.

[47] Easo, S. et al. "Simulation of LHCb RICH detectors using GEANT4." *IEEE*, 52(5): 1665–1668, 2005. doi:10.1109/TNS.2005.856766.

[48] Sjöstrand, T., Mrenna, S. and Skands, P. "PYTHIA 6.4 physics and manual." *Journal of High Energy Physics*, 2006(05):026, 2006. doi:10.1088/1126-6708/2006/05/026.

[49] Belyaev, I. et al. "Handling of the generation of primary events in Gauss, the LHCb simulation framework." *J. of Phys. Conf.*, 331(3):032047, 2011. doi:10.1088/1742-6596/331/3/032047.

[50] Lange, D.J. "The EvtGen particle decay simulation package." *Nucl. Instrum. Meth. A*, 462 (1-2):152–155, 2001. doi:10.1016/S0168-9002(01)00089-4.

[51] Golonka, P. and Was, Z. "PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays." *Europ. Phys. J. C-Part. and Fields*, 45(1):97–107, 2006. doi:10.1140/epjc/s2005-02396-4.

[52] Agostinelli, S. et al. "GEANT4—a simulation toolkit." *Nucl. Instrum. Meth. A*, 506(3): 250–303, 2003. doi:10.1016/S0168-9002(03)01368-8.

[53] Allison, J. et al. "Geant4 developments and applications." *IEEE Trans. Nucl. Sci.*, 53(1): 270–278, 2006. doi:10.1109/TNS.2006.869826.

[54] Allison, J. et al. "Recent developments in Geant4." *Nucl. Instrum. Meth. A*, 835:186–225, 2016. doi:10.1016/j.nima.2016.06.125.

[55] The Boole Project. `http://lhcbdoc.web.cern.ch/lhcbdoc/boole/`, accessed: 2021-03-24.

[56] Corti, G. et al. "Software for the LHCb experiment." *IEEE*, 53(3):1323–1328, 2006. doi:10.1109/TNS.2006.872627.

[57] Albrecht, J. et al. "Upgrade trigger and reconstruction strategy: 2017 milestone." *LHCb-PUB-2018-005*, 2018. `https://cds.cern.ch/record/2310579`.

[58] Forty, R. "RICH pattern recognition for LHCb." *Nucl. Instrum. Meth. A*, 433(1):257–261, 1999. doi:10.1016/S0168-9002(99)00310-1.

[59] Papanestis, A. and D'Ambrosio, C. "Performance of the LHCb RICH detectors during the LHC Run II." *Nucl. Instrum. Meth. A*, 876:221–224, 2017. doi:10.1016/j.nima.2017.03.009. The 9th international workshop on Ring Imaging Cherenkov Detectors (RICH2016).

[60] Gambetta, S. et al. "The LHCb RICH detectors: Operations and performance." *Nucl. Instrum. Meth. A*, 2020. doi:https://doi.org/10.1016/j.nima.2019.02.009.

[61] LHCb Collaboration. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC." *JINST*, 14(04):P04013–P04013, apr 2019. doi:10.1088/1748-0221/14/04/p04013.

[62] Aaij, R. et al. "Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2." *EPJ Tech. Instrum.*, 6 (LHCb-DP-2018-001. 1):1. 18 p, Mar 2018. doi:10.1140/epjti/s40485-019-0050-z.

[63] LHCb Collaboration. "Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays." *Phys. Rev. Lett.*, 118(CERN-EP-2017-041): 191801. 11 p, Mar 2017. doi:10.1103/PhysRevLett.118.191801.

[64] Benson, S. et al. "Real-time discrimination of photon pairs using machine learning at the LHC." *SciPost Phys.*, 7(arXiv:1906.09058):062. 16 p, Jun 2019. doi:10.21468/SciPostPhys.7.5.062. Submitted to SciPost Physics.

[65] Delaney, B.R. "Machine Learning @ LHCb." *LHCb-TALK-2020-153*, Sep 2020. `https://cds.cern.ch/record/2729794`.

[66] Aaij, R. et al. "Allen: A high-level trigger on GPUs for LHCb." *Comp. and Softw. for big Science*, 4(1):1–11, 2020. doi:10.1007/s41781-020-00039-7.

[67] LHCb Collaboration. "LHCb Upgrade GPU High Level Trigger Technical Design Report." (CERN-LHCC-2020-006. LHCB-TDR-021), May 2020. `https://cds.cern.ch/record/2717938`. CERN-LHCC-2020-006.

[68] Jia, Y. et al. "Caffe: Convolutional architecture for fast feature embedding." *Proc. 22nd ACM int. conf.*, pages 675–678, 2014. doi:10.1145/2647868.2654889.

[69] Abadi, M. et al. "Tensorflow: A system for large-scale machine learning." *12th USENIX symp.*, pages 265–283, 2016. `https://tensorflow.org`.

[70] Paszke, A. et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." *CoRR*, abs/1912.01703, 2019. doi:arXiv:1912.01703.

[71] Cambridge Dictionary. *Definition of machine learning*. `https://dictionary.cambridge.org/dictionary/english/machine-learning`, accessed: 2021-02-01.

[72] Campbell, M., Hoane Jr, A.J. and Hsu, F.h. "Deep blue." *Artificial Intelligence*, 134(1-2): 57–83, 2002. doi:10.1016/S0004-3702(01)00129-1.

[73] Silver, D. et al. "Mastering the game of Go with deep neural networks and tree search." *Nature*, 529(7587):484–489, 2016. doi:10.1038/nature16961.

[74] Schrittwieser, J. et al. "Mastering atari, go, chess and shogi by planning with a learned model." *Nature*, 588(7839):604–609, 2020. doi:10.1038/s41586-020-03051-4.

[75] Goodfellow, I. et al. *Deep Learning*. MIT Press, 2016. https://www.deeplearningbook.org.

[76] Bishop, C.M. *Pattern recognition and machine learning*. springer, 2006.

[77] Encyclopædia Britannica. *Occam's razor*. https://www.britannica.com/topic/Occams-razor, accessed: 2021-02-27.

[78] Wolpert, D.H. and Macready, W.G. "No free lunch theorems for optimization." *IEEE*, 1(1): 67–82, 1997. doi:10.1109/4235.585893.

[79] Friedman, J. et al. *The elements of statistical learning*. Springer series in statistics, 2009.

[80] kaggle. *Rectified Linear Units (ReLU) in Deep Learning*. https://www.kaggle.com/dansbecker/rectified-linear-units-relu-in-deep-learning, accessed: 2021-03-06.

[81] Kiefer, J., Wolfowitz, J. et al. "Stochastic estimation of the maximum of a regression function." *The Ann. of Math. Stat.*, 23(3):462–466, 1952. doi:10.1214/aoms/1177729392.

[82] Kingma, D.P. and Ba, J. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*, 2014. doi:arXiv:1412.6980.

[83] Polyak, B.T. "Some methods of speeding up the convergence of iteration methods." *USSR comp. math. and math. phys.*, 4(5):1–17, 1964. doi:10.1016/0041-5553(64)90137-5.

[84] LeCun, Y. et al. "Gradient-based learning applied to document recognition." *Proc. IEEE*, 86 (11):2278–2324, 1998. doi:10.1109/5.726791.

[85] Krizhevsky, A., Sutskever, I. and Hinton, G.E. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*, 25: 1097–1105, 2012. doi:10.1145/3065386.

[86] Campora Perez, D. and Jones, C. *Private communication*. 2018.

[87] LHCb Collaboration. "Precision luminosity measurements at LHCb." *JINST*, 9(12): P12005–P12005, dec 2014. doi:10.1088/1748-0221/9/12/p12005.

[88] Lester, C. "Trackless ring identification and pattern recognition in Ring Imaging Cherenkov (RICH) detectors." *Nucl. Instrum. Meth. A*, 560(2):621–632, 2006. doi:10.1016/j.nima.2006.01.073.

[89] Frühwirth, R. and Strandlie, A. "Robust circle reconstruction with the Riemann fit." 1085(4): 042004, 2018. doi:10.1088/1742-6596/1085/4/042004.

[90] Di Meglio, A., Gaillard, M. and Purcell, A. "CERN openlab Whitepaper on Future IT Challenges in Scientific Research." 2014. doi:10.5281/zenodo.8765. https://cds.cern.ch/record/1704394.

[91] Azzolin, V. et al. "Improving data quality monitoring via a partnership of technologies and resources between the CMS experiment at CERN and industry." *EPJ Web of Conf.*, 214:01007, 01 2019. doi:10.1051/epjconf/201921401007.

[92] Keskar, N.S. et al. "On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima." *CoRR*, 2016. doi:arXiv:1609.04836.

[93] Keizer, F. *Prospects for using sub-nanosecond time information to improve the LHCb RICH performance*. 2020. `https://indico.cern.ch/event/871479/contributions/3744203/attachments/1990355/3318089/Keizer_LHCC_2020.pdf`.

[94] Keizer, F. "Sub-nanosecond Cherenkov photon detection for LHCb particle identification in high-occupancy conditions and semiconductor tracking for muon scattering tomography." 2020. doi:10.17863/CAM.45822.

[95] LHCb Collaboration. "LHCb reoptimized detector design and performance." *LHCb-TDR-009*, 2003. `https://cds.cern.ch/record/630827`.

[96] Cherenkov, P.A. "Visible emission of clean liquids by action of γ radiation." *Dokl. Akad. Nauk*, 2:451, 1934.

[97] Cherenkov, P.A. "Influence of Magnetic Field on the Observed Luminescence of Fluids Induced by Gamma Rays." *Dokl. Akad. Nauk*, 3(9):413–416, 1936. doi:10.1103/PhysRev.52.378.

[98] Frank, I. and Tamm, I. "Coherent Visible Radiation of Fast Electrons Passing Through Matter." pages 29–35, 1991. doi:10.1007/978-3-642-74626-0_2.

[99] Joannopoulos, J.D. et al. *Photonic crystals: molding the flow of light*. Princeton university press, 2011.

[100] Lin, X. et al. "Controlling Cherenkov angles with resonance transition radiation." *Nature Physics*, page 1, 2018. doi:10.1038/s41567-018-0138-4.

[101] Ginzburg, V.L. "Transition Radiation and Transition Scattering." *Physica Scripta*, T2A: 182–191, jan 1982. doi:10.1088/0031-8949/1982/t2a/024.

[102] Ginzburg, V. and Tsytovich, V. "Several problems of the theory of transition radiation and transition scattering." *Physics Reports*, 49(1):1–89, 1979. doi:10.1016/0370-1573(79)90052-8.

[103] Kaminer, I. et al. "Quantum Čerenkov Radiation: Spectral Cutoffs and the Role of Spin and Orbital Angular Momentum." *Phys. Rev. X*, Jan 2016. doi:10.1103/PhysRevX.6.011006.

[104] Luo, C. et al. "Cerenkov radiation in photonic crystals." *Science*, Jan 2003. doi:10.1126/science.1079549.

[105] Xi, S. et al. "Experimental Verification of Reversed Cherenkov Radiation in Left-Handed Metamaterial." *Phys. Rev. Lett.*, Nov 2009. doi:10.1103/PhysRevLett.103.194801.

[106] Vorobev, V.V. and Tyukhtin, A.V. "Nondivergent Cherenkov Radiation in a Wire Metamaterial." *Phys. Rev. Lett.*, May 2012. doi:10.1103/PhysRevLett.108.184801.

[107] Ren, H. et al. "Nonlinear Cherenkov Radiation in an Anomalous Dispersive Medium." *Phys. Rev. Lett.*, May 2012. doi:10.1103/PhysRevLett.108.223901.

[108] Genevet, P. et al. "Controlled steering of Cherenkov surface plasmon wakes with a one-dimensional metamaterial." *Nature Nanotech.*, Jul 2015. doi:10.1038/nnano.2015.137.

[109] Hummelt, J.S. et al. "Coherent Cherenkov-Cyclotron Radiation Excited by an Electron Beam in a Metamaterial Waveguide." *Phys. Rev. Lett.*, Dec 2016. doi:10.1103/PhysRevLett.117.237701.

[110] Antipov, S. et al. "Observation of wakefield generation in left-handed band of metamaterial-loaded waveguide." *J. of Appl. Phys.*, 2008. doi:10.1063/1.2948929.

[111] Duan, Z. et al. "Observation of the reversed Cherenkov radiation." *Nature comm.*, 2017. doi:10.1038/ncomms14901.

[112] Lin, X. *Interdisciplinary Center for Quantum Information, State Key Laboratory of Modern Optical Instrumentation, ZJU-Hangzhou Global Science and Technology Innovation Center, College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China.* Private communication, 2020.

[113] Lin, X. et al. "Splashing transients of 2D plasmons launched by swift electrons." *Science Advances*, 3(1), 2017. doi:10.1126/sciadv.1601192.

[114] Jakobsen, S. et al. "Performance evaluation and optimization of the luminosity detector ALFA." *CERN-THESIS-2010-091*, 2010. `https://cds.cern.ch/record/1277003`.

[115] Barrillon, P. et al. ""MAROC: Multi-Anode ReadOut Chip for MaPMTs"." *IEEE Nucl. Sc. Symp. Conf.*, Oct 2006. doi:10.1109/NSSMIC.2006.355974.

[116] Edrington, A.C. et al. "Polymer-based photonic crystals." *Advanced Materials*, 2001. doi:10.1002/1521-4095(200103)13:6<421::AID-ADMA421>3.0.CO;2-%23.

[117] Lin, X. *Interdisciplinary Center for Quantum Information, State Key Laboratory of Modern Optical Instrumentation, ZJU-Hangzhou Global Science and Technology Innovation Center, College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China.* Private communication, 2017.

[118] Törmä, P. and Barnes, W.L. "Strong coupling between surface plasmon polaritons and emitters: a review." *Rep. on Prog. in Phys.*, page 013901, Dec 2014. doi:10.1088/0034-4885/78/1/013901.

[119] Blago, M. et al. "Determining the photon yield for the LHCb RICH Upgrade photodetection system." *Springer Proc. in Phys.* doi:10.1007/978-981-13-1316-5_46.

[120] Gamba, D. et al. "The CLEAR user facility at CERN." *Nucl. Instrum. Meth. A*, 909:480–483, 2018. doi:10.1016/j.nima.2017.11.080.

[121] Sjobak, K. et al. "Status of the CLEAR electron beam user facility at CERN." 2019. doi:10.18429/JACoW-IPAC2019-MOPTS054.

[122] Polyanskiy, M.N. *Refractive index database*. accessed: 2021-07-22.

[123] Glaude, D. *Metrology Section, Mechanical Materials Engineering Department, CERN, Geneva, Switzerland.* Private communication, Jun.2021.

[124] Duyar, O., Placido, F. and Durusoy, H.Z. "Optimization of TiO_2 films prepared by reactive electron beam evaporation of Ti_3O_5." *Jour. of Phys. D*, 41(9):095307, apr 2008. doi:10.1088/0022-3727/41/9/095307.

[125] Baumberg, J. *NanoPhotonics Centre, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.* Private communication, Jul 2021.

[126] Lin, X. et al. "A Brewster route to Cherenkov detectors." *Nature comm.*, Sep 2021. doi:10.1038/s41467-021-25822-x.

[127] Liu, R.Y.F. et al. "Forced Assembly of Polymer Nanolayers Thinner Than the Interphase." *Macromolecules*, 38(26):10721–10727, 2005. doi:10.1021/ma051649x.

[128] Pursiainen, O.L.J. et al. "Compact strain-sensitive flexible photonic crystals for sensors." *Applied Physics Letters*, 87(10):101902, 2005. doi:10.1063/1.2032590.

[129] Arsenault, A.C. et al. "Photonic-crystal full-colour displays." *Nature Photonics*, 1(8):468–472, 2007. doi:10.1038/nphoton.2007.140.

Copyright © 2021 Michele Piero Blago
All rights reserved

The work in this dissertation is funded by the CERN Doctoral Student programme.