Development, design and optimization of a novel Endcap DIRC for PANDA

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
der Justus-Liebig-Universität Gießen,
Fachbereich 07
(Mathematik und Informatik, Physik, Geographie)

vorgelegt von

Oliver Merle

Gießen, Dezember 2014

II. Physikalisches Institut der Justus-Liebig-Universität Giessen

Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der “Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis” niedergelegt sind, eingehalten.

Zusammenfassung

$\overline{\text{P}}$ANDA, ein Experiment an der zukünftigen FAIR Anlage in Darmstadt, strebt die Spektroskopie von Hadronen mit hoher Präzision und Statistik an. Dazu wird die gluonenreiche Proton-Antiproton Annihilation im Bereich von 1.5 GeV/c bis 15 GeV/c genutzt. Eine möglichst vollständige Abdeckung des Raumwinkels mit allen Detektoren ist eine Grundvoraussetzung um dieses Ziel zu erreichen.

Gegenstand dieser Dissertation ist das Systemdesign eines neuartigen DIRC Cherenkov Detektortyps zur Identifikation geladener Hadronen im Bereich der vorderen Endkappe des $\overline{\text{P}}$ANDA Target-Spektrometers. Ein besonderes Merkmal dieses Detektors ist die kompakte, planare Bauweise welche im Akzeptanzbereich ($\theta < 22°$) weniger als 5 cm Platz in Strahlrichtung und ca. 20 cm am äußeren Rand benötigt, wo einzelphotonenabbildende Kameras platziert sind.

Nach der Definition von Systemanforderungen wurde das System in logische Komponenten zerlegt. Für jede Komponente wurden mögliche Hardware- und Designoptionen identifiziert, analysiert und hinsichtlich ihrer Konformität mit den Systemanforderungen sowie ihrem Einfluss auf die Leistungsfähigkeit bewertet. Das optische System und Sensor-Layout wurden optimiert um einen Kompromiss zwischen Komplexität und Auflösung zu erlangen. Dieser Ansatz führte zu einem hochgradig modularen Detektordesign.

Die resultierende Systemdefinition umfasst die Spezifikation des optischen Systems, der Sensoren und der Ausleseelektronik. Weiter wurde ein Konzept zur Implementierung eines Musterrekonstruktionsalgorithmus zur Online Analyse vorgeschlagen.

Das neuartige Konzept erforderte die Entwicklung von speziellen Algorithmen zur Teilchenidentifikation, welche die effiziente Analyse der gemessenen zeitkorrelierten Photonenmuster ermöglichen. Diese Algorithmen wurden mit Signalen getestet, welche mit zeit-basierten Monte Carlo Simulationen generiert wurden um die Zeitcharakteristik der quasikontinuierlichen Interaktion bei der höchsten an $\overline{\text{P}}$ANDA erwarteten Rate nachzubilden. Die daraus folgende Schätzung der Detektoreffizienz ergibt eine π/K–Separation von bis zu 4σ bei einem Impuls von 4 GeV/c.

Abstract

$\overline{\text{P}}$ANDA, an experiment at the upcoming FAIR facility in Germany, aims at hadron spectroscopy with high precision and rate by exploiting gluon-rich proton-antiproton collisions at momenta from 1.5 GeV/c to 15 GeV/c. Almost 4π coverage by all detector components is a prerequisite to realize this goal.

The objective of this thesis is the system design of a novel type of DIRC Cherenkov detector for particle identification at the forward endcap of the $\overline{\text{P}}$ANDA target spectrometer. A key feature of this detector is its compact, planar design which occupies less than 5 cm in beam direction at the acceptance region ($\theta < 22°$) and ~20 cm at the outer rim where single-photon imaging cameras are located.

After the definition of system requirements, the system has been dissected into individual logical components. For each component, possible hardware and design options have been identified, analyzed and evaluated for compliance with the requirements and their impact on the system performance. The optical system and sensor layout have been optimized to compromise complexity and performance, leading to a highly modular detector setup.

The resulting definition of the design comprises the specification of the optical setup, the photosensor and the front-end electronics. Further, a concept for the implementation of the pattern reconstruction algorithm for online reconstruction has been proposed.

The novel concept also required the development of dedicated particle identification algorithms which permit the efficient analysis of the measured time-correlated photon patterns. These algorithms have been tested with signals generated by means of time-based Monte Carlo simulations which resemble the time characteristics of the quasi-continuous interaction at the highest rate expected at $\overline{\text{P}}$ANDA. The resulting performance estimations predict a π/K–separation up to 4σ at 4 GeV/c.

Contents

Part I.

Overview

1. Introduction

> *System design* – To develop a model on the basis of which a real system can be built, developed, or deployed that will satisfy all its requirements.
>
> *(Wymore)*

The topic of this thesis is the system design of a novel type of DIRC detector, commonly entitled *Disc DIRC* or *Endcap DIRC*. This detector is, broadly speaking, a large scale precision optics device exploiting the time correlated imaging of single photons to determine the emission angle θ_c of Cherenkov radiation with a precision better than 2 mrad.

The DIRC detector is a relatively young subclass in the family of Ring Imaging Cherenkov (RICH) detectors. RICH systems exploit the unique angular emission characteristics of Cherenkov radiation to precisely measure the relativistic velocity of fast charged particles up to 99.999 % c and higher*. The velocity $\beta = v/c$ can be directly obtained from the measured Cherenkov angle θ_c via the formula $\beta = 1/(n\cos\theta_c)$, where n denotes the refractive index of the radiator material. This velocity information is crucial for the discrimination of charged hadrons in particle physics experiments. An introduction to this topic is given in chapter 3.

The name DIRC is an acronym for "Detection of internally reflected Cherenkov light", what expresses the working principle of the detector which has been pioneered by the BaBar collaboration. A team around Blair Ratcliff developed and operated the first device, commonly known as the "BaBar DIRC" at the Stanford Linear Accelerator Center (SLAC) in the late 90s. At the time of writing, this is still the only DIRC device which has been used in a particle physics experiment.

The Endcap DIRC is based on the very same basic principles, but differs significantly in terms of geometry. While the BaBar DIRC has the geometry of an

*e.g. 50 GeV $\pi^{\pm}$ in LHCb RICH2

open barrel made from individual radiator bars, the Endcap DIRC is designed to cover the open ends of that barrel. Thus, it completes the DIRC technology in terms of geometrical acceptance. This is especially advantageous for fixed target experiments, where a larger fraction of the reaction products is emitted in forward direction due to the Lorentz boost.

While the BaBar DIRC camera was basically a huge pinhole camera with an aperture defined by the face of the coupled radiator bars, the Endcap DIRC makes use of radiator plates combined with high resolution focusing optics and dispersion correction. The whole system has to be designed to fit completely inside a magnetic spectrometer. These design choices also imply the need of different pattern reconstruction methods.

The aim of this work is to develop and evaluate the feasibility of an Endcap DIRC system which can be used for particle identification in the upcoming $\overline{\text{P}}$ANDA experiment at the new FAIR facility in Darmstadt. Further details on the motivation of this development are presented in chapter 4, following the introduction to the $\overline{\text{P}}$ANDA experiment (chapter 2) and Cherenkov detectors in general (chapter 3).

This development effort comprises the conceptional design and optimization of the optical system, sensors and readout-electronics, the development of models and computer simulations for requirement engineering and performance evaluation as well as the development of pattern analysis methods to reconstruct θ_c from the measured data in a high rate environment.

Major design challenges are posed by the tight geometrical constraints paired with the need of high rate single photon detection and imaging in an environment with a magnetic field of about $0.5 \ldots 1\,\text{T}$ as well as radiation levels which can easily harm electronic devices and a wide range of optical materials.

The core system which has been defined, optimized and evaluated in this thesis, meets the projected requirements and has been proposed as solution to the $\overline{\text{P}}$ANDA collaboration. A higher level SysML [44] diagram of the resulting Endcap DIRC design is shown in Fig. 1.1 (core system) and Fig. 1.2 (infrastructure).

As many details on the investigated DIRC related techniques are not covered by existing literature, strong focus has been put on the explanation of the working principles, advantages and disadvantages of different dispersion correction techniques and optical designs which have been investigated on the way to the final system, as well as on the motivation for the choice of each individual detector component. The key components of the system as well as the design synthesis are described in chapter 5. This document shall also serve as future reference for the

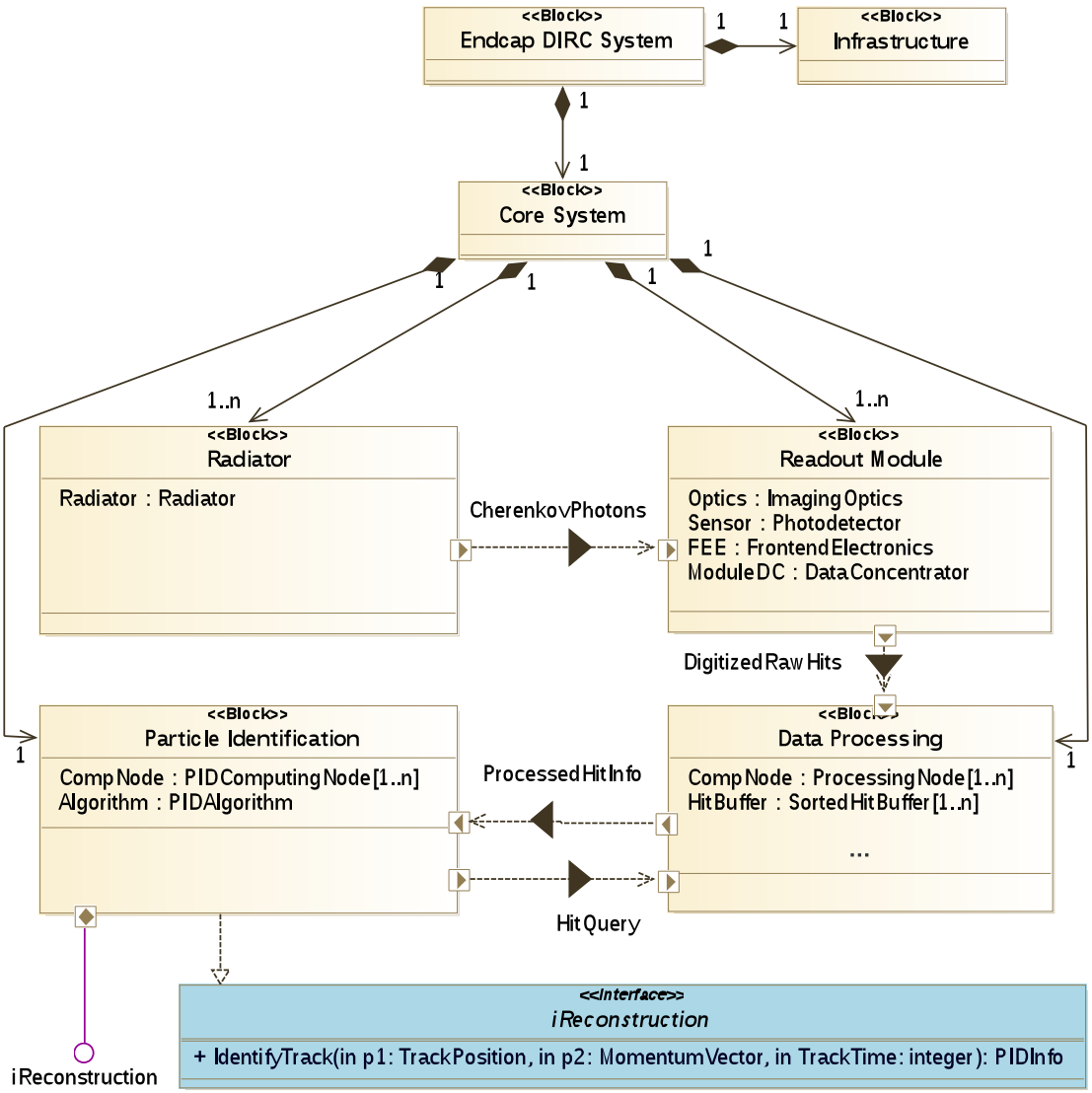


Figure 1.1.: A brief overview of the designed Endcap DIRC System (SysML Block Diagram). The core system design is the topic of the work presented in this document. A list of the infrastructure components needed is shown in Fig. 1.2.

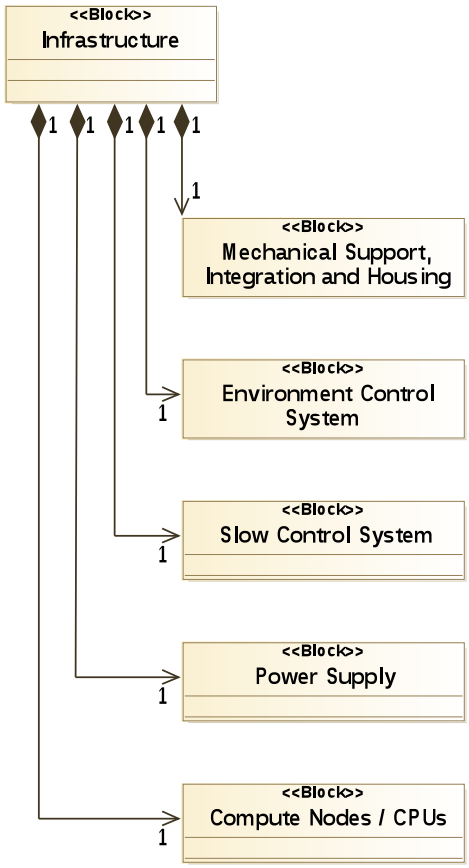


Figure 1.2.: List of the necessary infrastructure for the Endcap DIRC System (SysML Block Diagram). Besides the optomechanical support these are standard components which are used in many different detectors and do not involve a major development effort.

ongoing detector development.

Monte Carlo software, approaches for detector modeling and the developed pattern reconstruction algorithms are briefly introduced in chapter 6. Further optimization and performance studies as well as the final detector design are presented in chapter 8.

A first design iteration, based on a different sensor type, is also presented in chapter 7. This intermediate step is especially interesting because it demonstrates that a DIRC can also work with the intrinsically noisy silicon photo-multipliers by using precise time-correlated imaging to reject the noise.

Key topics of this thesis:

- Requirement analysis.
- Development of a logical detector concept and associated model.
- Review, development and optimization of focusing optics.
- Development of a hardware concept including optics, sensors and electronics.
- Development of a reconstruction algorithm.
- Performance evaluation using detailed simulations.

Software developed for this purpose:

- Standalone Geant4 simulations.
- A “time based” PandaROOT simulation.
- A software package for the optimization of focusing optics.

2. The $\overline{\text{P}}$ANDA experiment

2.1. $\overline{\text{P}}$ANDA in the context of FAIR

$\overline{\text{P}}$ANDA* is a modern hadron physics experiment dedicated to high precision spectroscopy of antiproton reactions. It will be part of the international Facility for Antiproton and Ion Research (FAIR) which is currently under construction in Darmstadt, Germany. FAIR extents the already existing GSI† facility by an additional proton linear accelerator (p-LINAC), two synchrotrons (SIS-100/300) and six cooler/storage rings to provide the connected experiments with proton, antiproton and heavy ion beams. A schematic overview of the planned facility is shown in figure 2.1. $\overline{\text{P}}$ANDA is one out of fourteen initial experiments dedicated to the so called "*four scientific pillars of FAIR*", which were defined as [4]:

APPA	Atomic and plasma physics. Applied sciences in the bio, medical and material sciences.
NuSTAR	Structure of nuclei, physics of nuclear reactions, nuclear astrophysics and radioactive ion beams.
CBM	Physics of hadrons and quarks in compressed nuclear matter and hypernuclear matter.
$\overline{\text{P}}$ANDA	Hadron structure and spectroscopy, strange and charm physics, hypernuclear physics with anti-proton beams.

To generate antiprotons at FAIR, protons from the p-LINAC will be further accelerated by the SIS-18 and SIS-100 synchrotrons to a kinetic energy of 29 GeV/c^2 and guided into a metal target where inelastic proton-nucleus collisions lead – beside a lot of other final state particles – to the production of antiprotons [29]. These

*The name is an acronym for antiproton ($\overline{\text{p}}$) annihilation at Darmstadt

†Gesellschaft für Schwerionenforschung mbH

will be collected by a magnetic horn, separated from other particles, fed into the collector ring (CR) to be pre-cooled and finally delivered to the RESR storage ring for accumulation [2, 122]. From there, antiprotons can be injected into the HESR (high energy storage ring) and NESR (new experimental storage ring).

The HESR will be instrumented with $\overline{\text{P}}$ANDA to cover physics with high energy antiprotons, while the NESR will provide decelerated antiprotons to in-ring experiments and the low energy facility FLAIR. By further deceleration in the low- and ultra-low-energy storage rings LSR/USR, energies as low as 20 keV will be accessible at the FAIR facility [1].

At the HESR, a combination of stochastic and electron cooling* will provide the unique opportunity to study antiproton reactions at high precision and rate [71]. Though, due to the modular construction plan of the FAIR facility, the RESR, NESR and FLAIR parts will not be available from the early beginning [4] and the HESR has to be filled directly from the collector ring. The HESR will not be able to provide highest luminosity until the RESR is available.

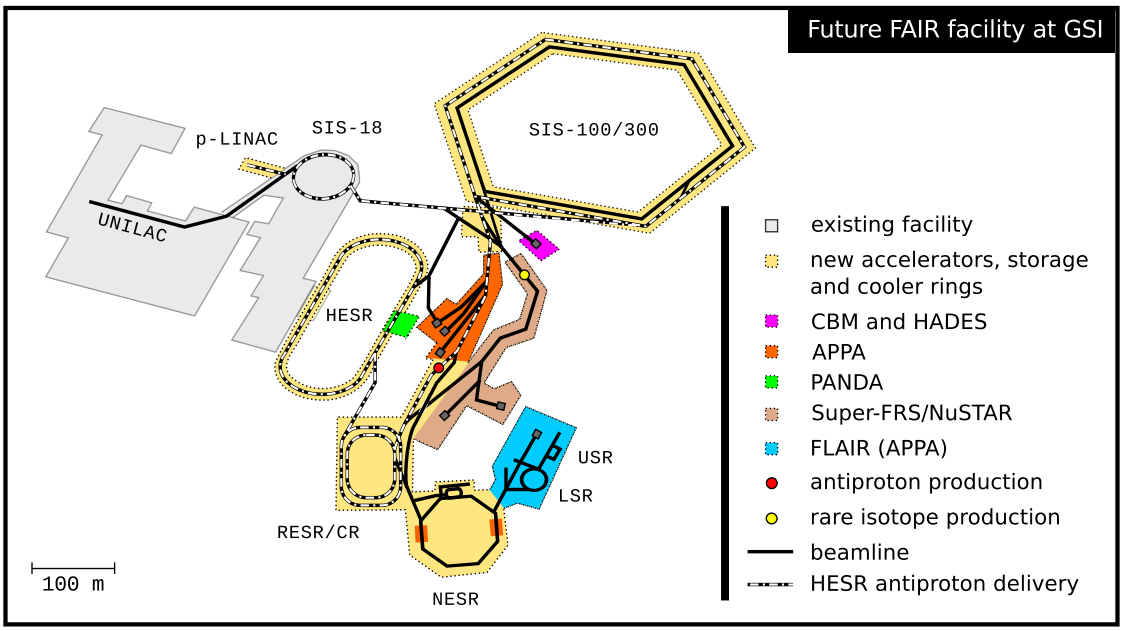


Figure 2.1.: Overview of the planned FAIR facility providing proton, antiproton and heavy ion beams for a multitude of different experiments with diversified physics programs. Antiprotons and rare ions are generated at production targets in the beamline.

*Electron cooling will only be available up to momenta of 8.9 GeV/c. Stochastic cooling can be provided also for momenta above 3.8 GeV/c

Two HESR operation modes are foreseen as summarized in table 2.1. A high resolution mode will be available in the momentum range covered by the electron cooler. The beam momentum spread σ_p/p increases with the number of antiprotons $N_{\bar{p}}$ stored in the ring, so $N_{\bar{p}}$ will be limited in this mode. A ten times higher peak luminosity can be reached in high luminosity mode where the HESR will be filled with the maximum number of antiprotons at the expense of an increased momentum spread. Latter mode relies on the RESR for particle stacking and will thus only be available in the full FAIR version.

2.2. In-ring target systems and interaction rate

The $\overline{\text{P}}$ANDA spectrometer can use several fixed internal targets. An internal target directly interacts with the beam inside the storage ring. The term fixed expresses that the target is at rest in the laboratory, which is a valid assumption as long as the target velocity can be neglected in the physics of the ongoing reactions. The simplest targets are conventional nuclear targets used to study antiproton-nucleus reactions ($\bar{p}$N). These are very thin wires or foils which can be placed in the beam halo. For the study of proton-antiproton reactions ($p\bar{p}$) two complex hydrogen systems are in development: the cluster-jet target and the pellet target [62].

Both hydrogen targets are specifically tailored to HESR conditions and offer the option to be operated with heavier gases, e.g. O_2, N_2, Ne, Ar, Kr and Xe for $\bar{p}$N studies. The operating principle of both targets is outlined in figure 2.2.

In the cluster jet target a de Laval nozzle is used to produce a uniform stream of nano-droplets, called clusters. Precooled gas is injected into vacuum through the nozzle which tightest part has a radius of about 10-30 µm. The expansion of the

	high resolution	high luminosity
Momentum range [GeV/c]	1.5 to 8.9	1.5 to 15
Max. number of $\bar{p}$ stored	10^{10}	10^{11}
Momentum spread σ_p/p	$< 4 \cdot 10^{-5}$	$\approx 10^{-4}$
Peak luminosity [$\text{cm}^{-2}\text{s}^{-1}$]	$2 \cdot 10^{31}$	$2 \cdot 10^{32}$
assuming ρ_{target} of	$4 \cdot 10^{15}\,\text{atoms/cm}^2$	

Table 2.1.: HESR operation modes

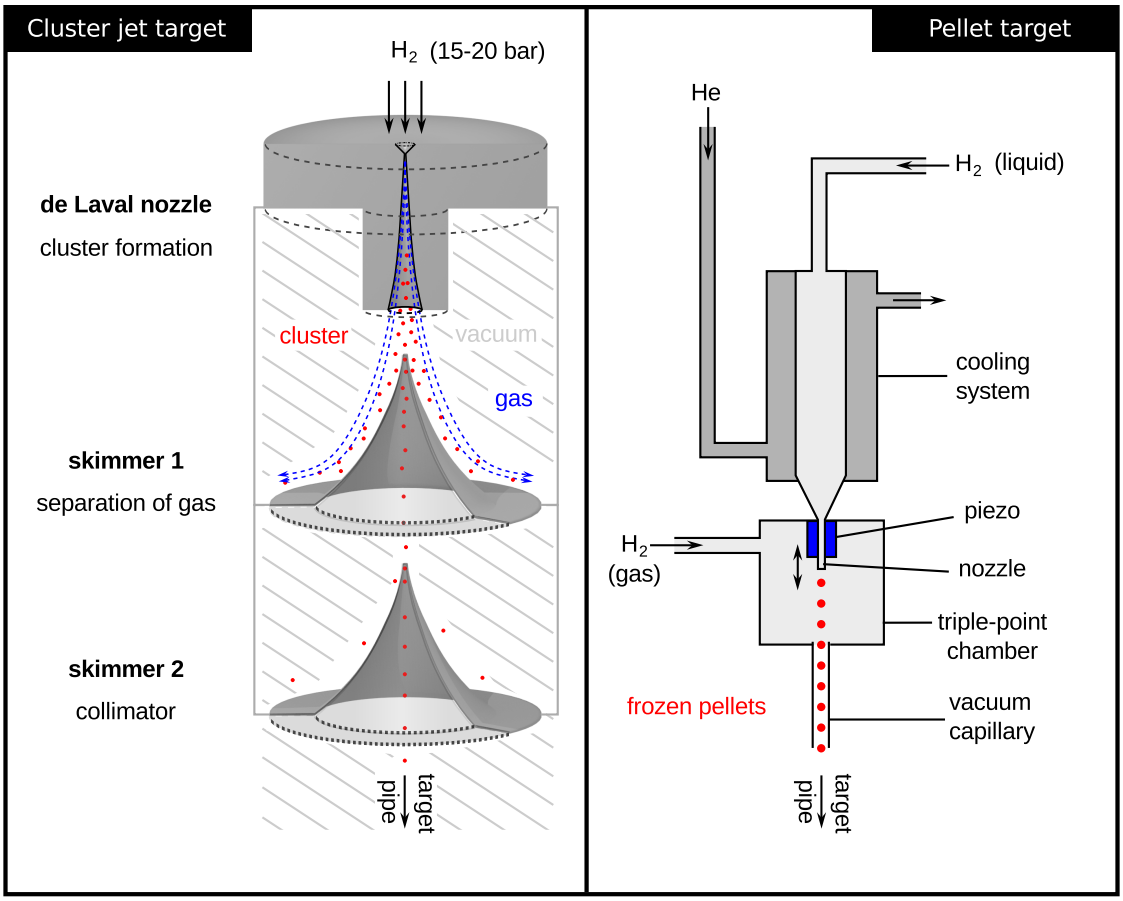


Figure 2.2.: Principle of operation for the foreseen cluster and pellet targets.

.

gas inside the nozzle results in a drop of its internal energy which, under proper preconditions, leads to condensation and thus cluster formation. Cluster size and yield depend strongly on such preconditions like gas temperature, pressure and nozzle shape ([130, 64]). In two consequent vacuum chambers, funnel-shaped skimmers are used to separate clusters from the gas stream and to collimate the resulting beam. Due to their high mass, cluster jets have an advantage over common supersonic gas jets: collisions with residual gas in the vacuum system do not significantly widen the target beam. Therefore cluster jets can be produced far away from the target as it is required by $\overline{\text{P}}$ANDA where the interaction region is about 2.1 m away from the target source to maximize the spectrometer acceptance.

In case of the pellet target, cryogenic liquid hydrogen is injected into a triple point chamber by means of a vibrating nozzle. An axial nozzle oscillation at

the correct frequency causes the liquid jet to break up into droplets of a size similar to the nozzle diameter. The triple point chamber, also filled with hydrogen, diminishes perturbing effects like evaporation and aerodynamic interactions and thus provides an optimal environment for a uniform formation of droplets [20]. Generated droplets leave the triple point chamber through a small capillary into vacuum where evaporation lowers the droplet temperature and causes them to freeze. The created pellets have a diameter of approximately 10-30 µm and an average distance in the order of several millimeters.

$\overline{\text{P}}$ANDA aims for average interaction rates up to 10 MHz in high luminosity mode*. According to HESR simulations [71], this rate corresponds to a target thickness of $4 \cdot 10^{15}$ atoms/cm^2. Currently, the cluster jet system is able to reach an effective target thickness of only $1 \cdot 10^{15}$ atoms/cm^2 and thus cannot be used when highest rates are mandatory. However, the primary advantage of the cluster jet target is the quasi-homogeneous target beam which can be fine controlled during operation to vary the effective target thickness over several orders of magnitude, e.g. by moving the nozzle and skimmers as well as varying process parameters like gas temperature and pressure. This fine-grained control is especially advantageous when a luminosity below the maximum is required. This feature could also be used to compensate for antiproton consumption during a HESR cycle by continuously increasing the target thickness and thus stabilizing the luminosity.

It has been claimed that the pellet target will be able to deliver an average target thickness of $\geq 4 \cdot 10^{15}$ atoms/cm^2 at a pellet rate ≥ 150 kHz [62] which would allow to reach the targeted 10 MHz interaction rate. A major disadvantage of the pellet target compared to the cluster jet is the occurrence of significant temporal fluctuations in luminosity which are caused by the coarse granularity of the target beam. It is envisaged to keep these fluctuations below a factor of two so the maximum instantaneous rate does not exceed 20 MHz. At pellet rates below 15 kHz the granularity can be turned into an advantage by adding a system for precise optical tracking of individual pellets. Such a system can provide a sub-mm primary vertex position which can be used to reduce background or, as shown in [93], to improve the identification and reconstruction of D-meson decays.

*As shown in [71], figure 8, at $p = 15$ GeV/c the maximum cycle-averaged luminosity $\bar{L}_{\text{cycle}} = 1.6 \cdot 10^{32}$ cm^{-2}s^{-1} is reached for a cycle time of $T_{\text{cycle}} = 2500$ s. Taking into account that the beam preparation time $T_{\text{prep}} = 290$ s is part of the cycle time and assuming a hadronic cross section of $\sigma = 51$ mb, the expected rate is $\bar{L}_{\text{cycle}} \frac{T_{\text{cycle}}}{T_{\text{cycle}} - T_{\text{prep}}} \sigma \approx 9.2$ MHz. Due to the barrier bucket system 10 % of the HESR will be empty, resulting in a 10 % rate increase.

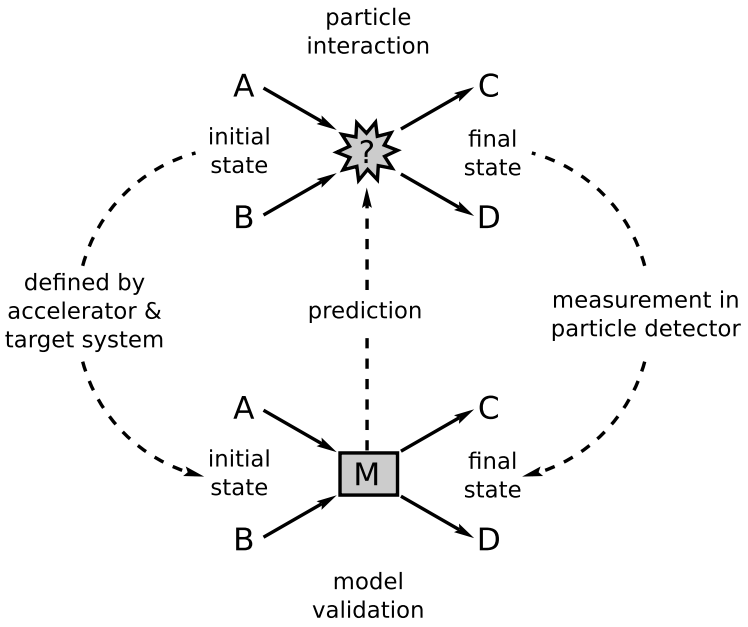
In conclusion both hydrogen target systems complement one another due to their very disjunct strengths. Both will be used in $\overline{\text{P}}$ANDA to provide optimum experimental conditions depending on the physics topic under investigation.

2.3. The $\overline{\text{P}}$ANDA spectrometer and sub-systems

2.3.1. The spectrometer

At large scale, general purpose hadron physics experiments like $\overline{\text{P}}$ANDA one examines what fundamentally happens when hadrons interact with each other in so called inelastic collisions. Unfortunately, these fundamental processes cannot be observed directly. To access this information anyway, it has to be deduced from the course of the reaction. This can be accomplished by measuring the initial and final state of single reactions and comparing this data to a theoretical model.

The initial state is defined by the parameters of the colliding particles such as quantum numbers, charge, energy and the momentum vector. These are usually well defined by the design of accelerator and target systems. The final state however, defined by the parameters of the particles produced in the reaction*,



*in this section the term particles is meant to include high energy photons

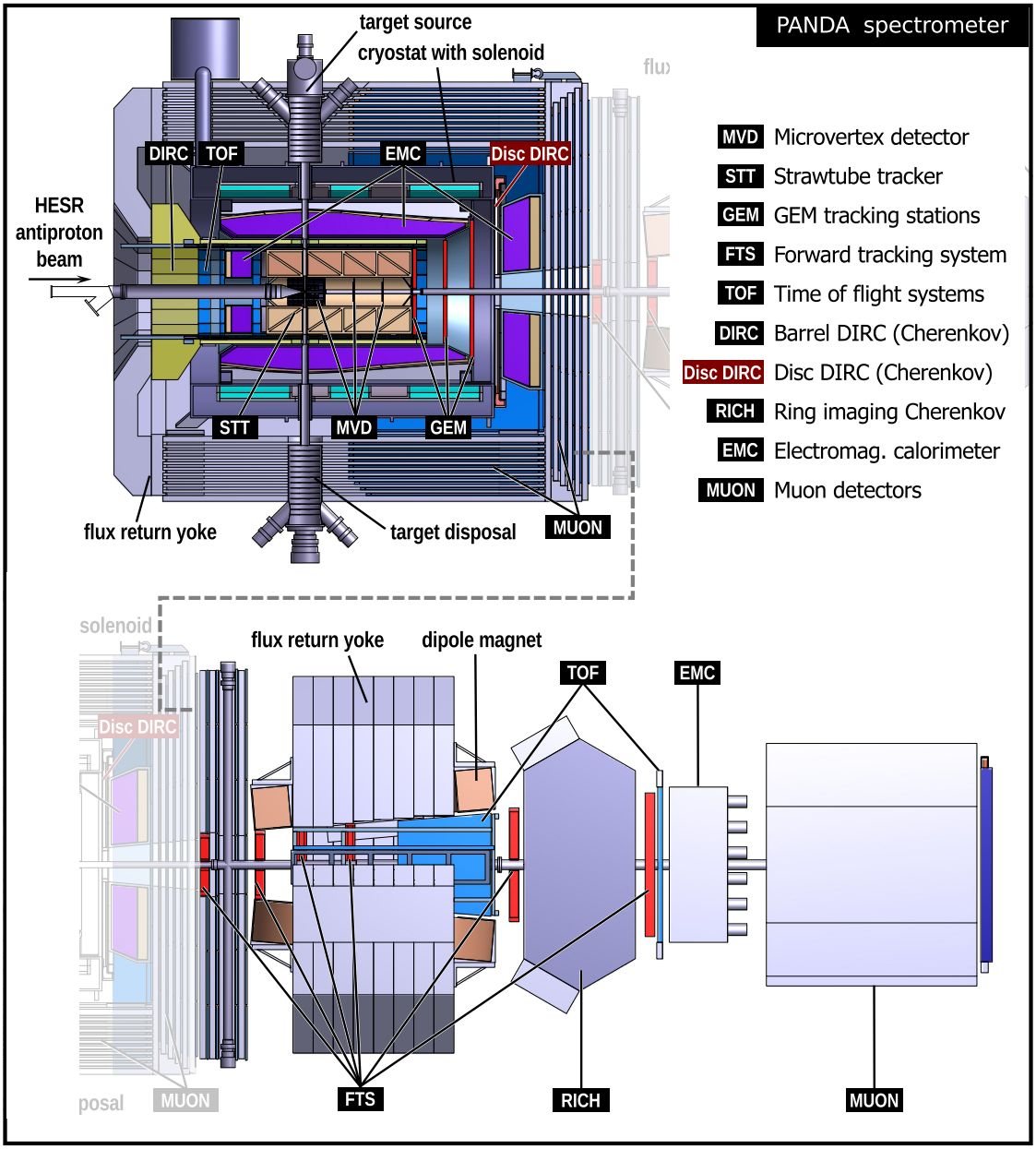


Figure 2.3.: The $\overline{\text{P}}$ANDA spectrometer and subsystems.

.

has to be measured by a dedicated detector system. Typically, a large apparatus composed of several specialized detectors is necessary to satisfy the experimental needs. These sub-systems work together to determine the particles relativistic four-momentum vector $P = (E, p_x, p_y, p_z)$. Its components are the three-momentum vector $\vec{p} = (p_x, p_y, p_z)$ and the total energy $E^2 = p^2 + m^2$ of the particle. Depending on the type of reaction and experiment, it is either aspired to measure P for all of the final state particles ("exclusive measurement") or only a subset ("inclusive measurement").

In case of charged particles, the three-momentum spectrum can be efficiently obtained by measuring the deflection of the particle trajectories inside a magnetic field. A setup to achieve this is called a "magnetic spectrometer" and consists at least of two components. A strong magnet to provide the field and a precise tracking system to measure the trajectory. Common tracking systems exploit a medium which can be easily ionized by the traversing charged particles. A position sensitive measurement of the resulting charges along the particle track yields the position information which is then used to reconstruct (compute) the trajectory. As the bending of the track inside the magnetic field must be sufficiently large to reach an acceptable momentum resolution it can be necessary to use more than one spectrometer. This is also the case in $\overline{\text{P}}$ANDA.

As the energy component E of the four-momentum cannot be determined by the magnetic spectrometer, it has to be provided by additional detector systems for calorimetry and/or particle identification (PID). The former measure E directly while the latter determine E indirectly by revealing the particle type and thus its mass m. The main components in large physics detectors like $\overline{\text{P}}$ANDA can be classified in three categories: tracking, calorimetry and particle identification. These key components are completed by detectors for luminosity measurement and the creation of trigger signals.

$\overline{\text{P}}$ANDA is equipped with two different magnetic spectrometers to cover high as well as low transverse momentum (p_T) regions with sufficient tracking precision. The target spectrometer employs a superconducting solenoid magnet providing a 2 T ($\pm$2%) field for momentum measurement of high p_T-tracks. The momentum of low p_T tracks in forward direction is determined by the forward spectrometer which is based on a dipole magnet and has an elliptical acceptance of $\theta_{\text{horiz.}} < 10°$, $\theta_{\text{vert.}} < 5°$. An overview of the setup is presented in figure 2.3

2.3.2. Tracking systems

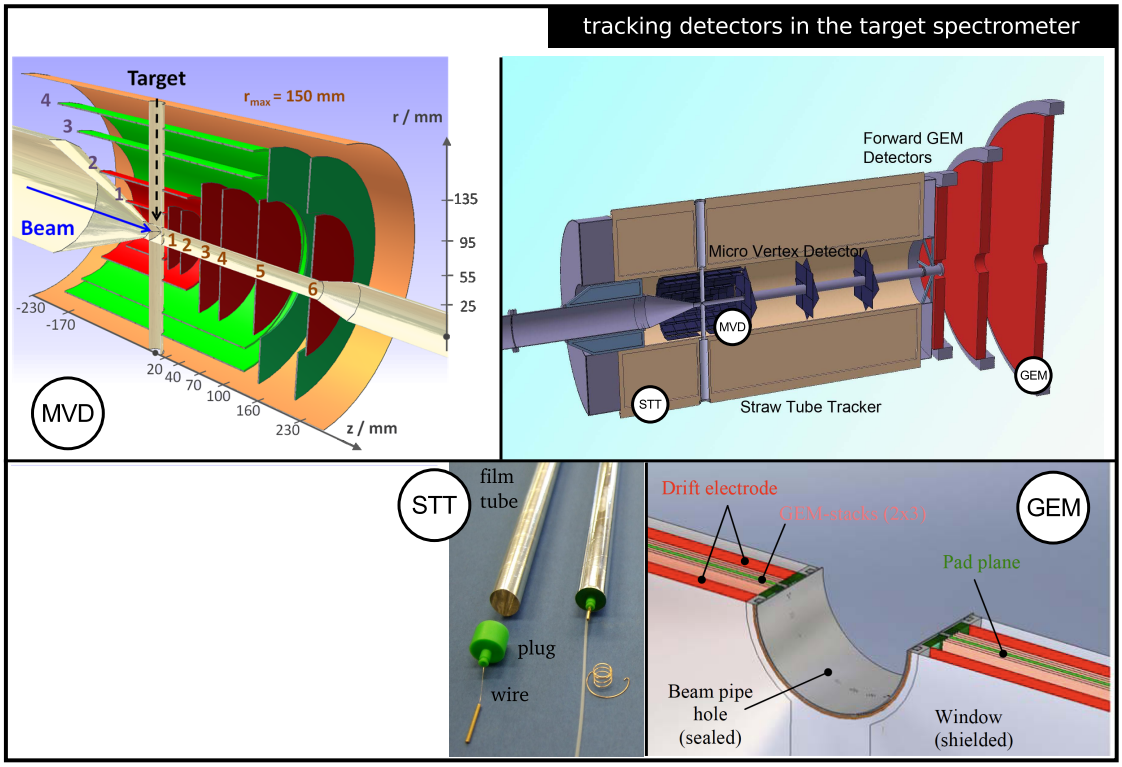


Figure 2.4.: The tracking detectors in the PANDA target spectrometer. Image sources: [98]: MVD picture, [142]: GEM tracker, [100]: STT and overview pictures.

The PANDA tracking system consists of four subsystems. Three of them are presented in figure 2.4 in more detail. A semiconductor based Microvertex detector (MVD) encloses the interaction region where very high position resolution is mandatory. The MVD is surrounded by a gaseous tracker based on straw tubes (STT) which provide good resolution in radial direction but limited resolution in beam direction. Three planar stations with Gas Electron Multiplier (GEM) foils cover the forward part of the target spectrometer, thus completing the STT in terms of acceptance. The forward spectrometer will employ several tracking stations (FTS, see figure 2.3) using the same type of straw tubes as the STT.

The Microvertex Detector (MVD) will provide 3D track positions accompanied with timestamps of about 20 ns precision. According to the technical design report [98], the envisaged spatial resolution is better than 100 µm longitudinal to the beam and tens of µm in the transverse plane. The targeted vertex resolution is

~ 100 µm. This choice is mainly motivated by the decay length of open charm mesons as the detection of displaced vertices is one of the MVD key tasks.

To maintain a good compromise between high rate capability, low radiation length and resolution, the MVD makes use of two different silicon sensor technologies: hybrid pixel detectors and double-sided micro-strip detectors (DSSD). The principle detector layout consists of four barrel layers and six forward "wheels" as presented in figure 2.4. The inner parts (red) of the barrel and wheels are equipped with pixel detectors, the outer ones (green) with DSSDs. The use of two additional forward discs between the MVD and the GEM detectors is also under discussion.

The Straw Tube Tracker is an assembly of 4636 gaseous proportional counters called "straws" which are aligned parallel to the beam-axis to form a barrel like geometry as illustrated in figure 2.4. The straws are tubes of 10 mm diameter made from 27 µm thick aluminized BoPET foil* serving as cathode (fig. 2.4). Each tube contains a 20 µm thick anode wire of gold plated tungsten and is filled with gas at a pressure of approx. 2 bar making the tube self-supporting and causing a close-to-perfect cylindrical shape [100]. Traversing charged particles loose energy by ionizing gas molecules inside the tube. The resulting electron/ion pairs are accelerated in the radial field between anode and cathode. In proximity to the anode wire the electric field is strong enough to allow drift electrons to ionize more gas which causes an avalanche which yields the amplification needed to generate a measurable signal at the anode. The distance of the track to the anode wire can be obtained by exploiting the drift time of the fastest electrons. This way, a position resolution better than 150 µm can be reached in the plane perpendicular to the tube. Along the beam-axis, a resolution of 3 mm is expected.

Three micropatterned gas detectors based on GEM foils [115] will provide the track information in the endcap region of the target spectrometer, e.g. at polar angles below 22°. The principle of operation is illustrated in figure 2.5. A GEM is basically a metallized, highly resistive polymer film which has been microstructured using a conventional etching process. The result is a grid of many tiny holes, typically of diameters below 100 µm. A high voltage applied to the GEM causes the formation of strong electric fields inside the holes. These fields are used for electron multiplication in a gas volume (fig. 2.5, bottom right). The impinging particle track ionizes the gas in the detector. An external electric field between cathode (drift electrode) and anode (pad plane) accelerates the electrons

*Biaxially-oriented polyethylene terephthalate. A specially treated thin PET film well known for it's high tensile strength and chemical stability. Typical Young's modulus: ≈ 4 GPa.

towards the three successive GEM stages where they are multiplied. The resulting electrons are collected on a position sensitive anode structure. Ions drift either to the cathode or one of the GEMs. A similar detector has been used in the COMPASS experiment [61]. Due to the Lorentz boost introduced by the fixed target setup in $\overline{\text{P}}$ANDA, rates are significantly higher at small polar angles. This makes GEM based detectors a perfect fit as they combine high resolution with high rate capabilities at a low material budget.

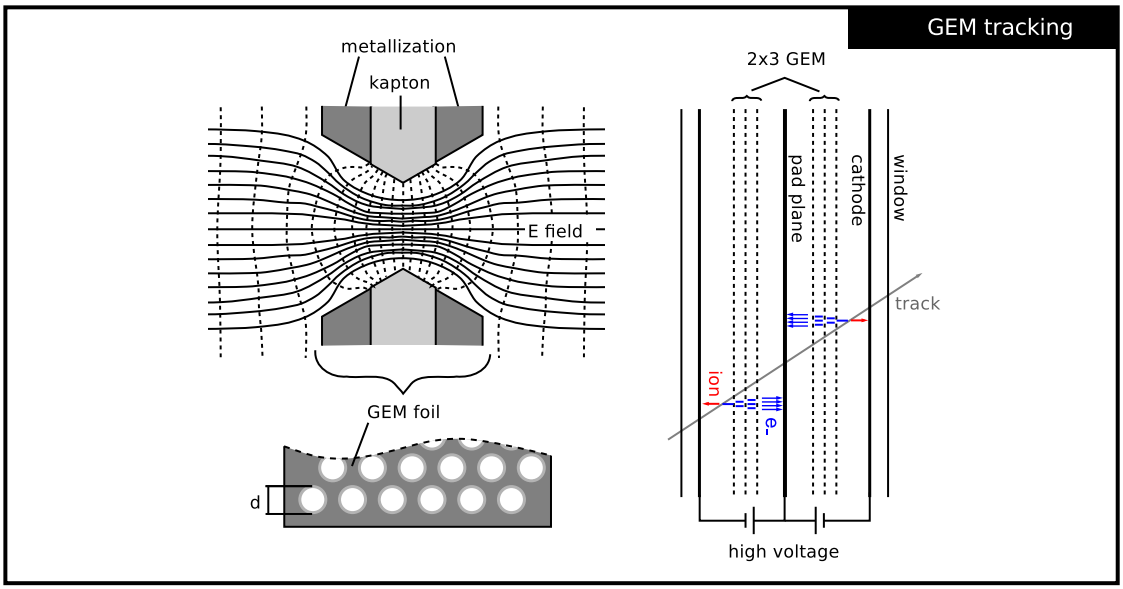


Figure 2.5.: Principle of a GEM tracking station.

The forward tracking stations (FTS) use the same straw-tubes as the STT. These will be grouped in modules of 32 parallel tubes arranged in two layers. Each module has its own front-end electronics and high voltage supply. The modules are used as basic elements to construct the tracking stations. Each station has four layers of modules. The first and last layer are oriented vertical while the inner layers are rotated by $+5^\circ$ and -5° around the beam axis. Again, the envisaged resolution is better than 100 µm.

2.3.3. Calorimetry

The target and forward spectrometers of $\overline{\text{P}}$ANDA are equipped with different types of electromagnetic calorimeters (EMC). The main task of the EMC is the reconstruction of electrons, positrons and photons with high efficiency and low

background [97]. In addition,the EMC has to discriminate $e^{\pm}$ and $\pi^{\pm}$ as there is no threshold Cherenkov detector in $\overline{\text{P}}$ANDA.

In case of the target spectrometer, the material of choice is lead tungstate ($PbWO_4$), a dense inorganic scintillator which has already been chosen by the LHC experiments CMS and ALICE. This material shows an interesting combination of properties, especially a high density and small Molière radius, which permit the construction of compact calorimeters with high granularity, as well as a fast response and radiation hardness, which are a requisite for high rate applications. A detailed overview of the parameters and a comparison with other scintillator materials can be found in [8]. Crystals of 20 cm length will be arranged to form the barrel, forward- and backward endcap geometries consisting of 11360, 3600 and 592 crystals respectively. The readout is realized by using large area avalanche photodiodes and phototriodes. The envisaged energy resolution is $\frac{\sigma_E}{E} \leq 1\% \oplus \frac{\leq 2\%}{\sqrt{E/\text{GeV}}}$. The constant term is driven by the electron pion separation by E/p measurement. In this case the energy resolution has to be better than the momentum error of the tracking system which is in the order of 1%. The latter term is introduced to assure a good mass determination of light mesons, e.g. π_0 and η [97].

A disadvantage of $PbWO_4$ is the rather low light yield with a strong temperature dependence of about -3% per $^{\circ}C$. The $\overline{\text{P}}$ANDA EMC will therefore be cooled to -25°C within $\pm 0.1^{\circ}$C to increases the yield by a factor of four compared to room temperature. The cooling introduces a negative side effect as it slows down the annealing rate of radiation induced defects inside the crystal, thus reducing the radiation hardness. Recently, the effect of stimulated recovery of irradiated $PbWO_4$ under illumination with visible light has been observed. This technique is under investigation and might be applied online during operation of the EMC [94]. Even without this method, the improved $PbWO_4$ crystals are considered to be sufficiently radiation tolerant to withstand the 6 month duty cycles of $\overline{\text{P}}$ANDA.

The forward spectrometer uses a shashlik type calorimeter made from alternating layers of lead and scintillator material. Wavelength shifting fibers lead through the scintillators and are coupled to photomultipliers for the readout of the scintillation signal. A resolution of $\frac{\sigma_E}{E} = 1.3\% \oplus \frac{2.8\%}{\sqrt{E/\text{GeV}}}$ has been measured with a prototype [88]. More information is given in [97] and references therein.

2.3.4. Particle identification

$\overline{\text{P}}$ANDA does not use a hadron calorimeter. Charged hadrons are discriminated by the particle identification (PID) system which is composed of time-of-flight, Cherenkov and muon detectors. Additional PID input is provided by the tracking system which measures the energy loss ($\mathrm{d}E/\mathrm{d}x$) of traversing particles.

Two DIRC Cherenkov detectors surround the tracking systems of the target spectrometer. The Barrel DIRC covers the polar angle range from 140° to 22° and the Endcap DIRC, also called Disc DIRC, the range from 22° to 5°/10° (θ_x,/θ_y). Both detectors exploit the Cherenkov effect to provide a discrimination between pions, kaons and protons at momenta above ~ 1 GeV/c. The targeted π/K-separation power at 4 GeV/c is 3.5σ for $\theta > 22°$ and 4σ for $\theta < 22°$. The Barrel DIRC design is very similar to the first DIRC detector which has been successfully applied in the BaBar experiment [144]. More details on the topic will be discussed in the section 3.4.

While the Endcap DIRC is based on similar principles, it aims for better resolution at higher momenta while using less space for optics and electronics. Due to the different geometry, it is significantly different to the BaBar like setup in terms of pattern reconstruction. The conceptual design, the developed analytical reconstruction approach and performance studies based on Monte Carlo simulations are the topic of this thesis.

Particle identification in the forward spectrometer may be provided by an aerogel RICH detector [11]. For the identification of slower particles, a time-of-flight detector is located between the Barrel DIRC and EMC. This detector consists of an array of small tiles of plastic scintillators (about $30 \times 30 \times 5\mathrm{mm}^3$) which are read out using two silicon photomultipliers per tile. A time resolution better than 100 ps is envisaged. There will be no start time available in PANDA. This originates from the quasi-continuous beam as well as the fact that a start detector would deteriorate the overall performance due to its radiation length close to the IP. Therefore the TOF system has to use relative timing.

To separate pions from muons, a range tracking muon system is integrated in the solenoid flux return yoke. It consists of several alternating layers of iron as absorber and mini drift tubes (MDT) to track the particles and measure energy loss. The iron itself is also part of the yoke. Absorber layers at the forward endcap are twice as thick due to the higher momentum caused by the Lorentz boost. A muon filter of additional four layers is placed between the target and forward spectrometer to

further enhance the system in this region. The forward spectrometer is equipped with technically equivalent muon system which is designed for higher momenta. More details can be found in the related technical design report [99].

2.3.5. Luminosity monitor

In case of low four-momentum transfer the cross section for elastic $\bar{\text{p}}$p scattering can be computed exactly. This process can therefore be used to measure the relative luminosity. A tracking system consisting of four layers of silicon microstrip detectors will be placed close to the beam inside the vacuum at a distance of 11 m upstream to the IP. This enables the measurement of the angles of antiprotons scattered at small polar angles which correspond to a low four-momentum transfer. The expected absolute precision is about 3% for the time integrated luminosity.

2.4. Physics program

The major part of the $\overline{\text{P}}$ANDA physics program is dedicated to the exploration of the strong interaction which is one of the four fundamental forces in nature. The other three are the electromagnetic and weak interaction as well as the gravitational force. Beside gravity, all fundamental interactions have been described by quantum field theories. Quantum electrodynamics (QED) is a very successful example of this approach as it has been probed many times and gives reliable and precise predictions. One can say that QED is a sophisticated and well established theory describing the electromagnetic interaction on the quantum level. This enormous success raises confidence that other interactions may also be described precisely by their field theories. Quantum chromo dynmamics (QCD), the field theory related to the strong interaction, works well for high energies where perturbation theory can be applied. However, at lower energies where perturbation theory cannot be applied, QCD is still far away from being such a sophisticated theory. $\overline{\text{P}}$ANDA aims at contributing to a deeper understanding of the strong interaction in the non-perturbative energy regime.

The $\overline{\text{P}}$ANDA experiment has been designed with an emphasis on the combination of high precision and high statistics. This goal will be realized by combining intense, cooled antiproton beams stored in a dedicated storage ring (section 2.1) with internal hydrogen targets (section 2.2). The maximum beam momentum of 15 GeV/c corresponds to a center of mass energy of $\sqrt{s}$ =5.47 GeV/c^2. The

antiproton-proton reaction ($\bar{p}p$) allows the direct formation of many states which can not be formed directly in e^+e^--collider experiments. This combination of experimental parameters opens up unique possibilities for precision measurements in the charmonium region.

The X state resulting from the reaction $\bar{p}p \to X_{\bar{p}p}$ or $e^+e^- \to X_{e^+e^-}$ can either be a single resonance ("formation mode") or a resonance accompanied by other particles ("production mode"). In formation mode $X_{\bar{p}p}$ can have any non-exotic quantum numbers. This is a significant difference to e^+e^-–colliders where the quantum numbers of $X_{e^+e^-}$ are restricted to those of the virtual photon ($J^{PC} = 1^{--}$) due to their conservation by the electromagnetic interaction. This difference in the formation process distinguishes $\overline{\text{P}}$ANDA from competing experiments like BELLE II and BESS III. In $\bar{p}p \to X$ all non-exotic states are accessible in formation mode what allows to apply beam energy scans to precisely measure the width and mass of resonances with $J^{PC} \neq 1^{--}$.

The resonance scanning technique has been pioneered by the R704 collaboration at the CERN-ISR [10] and further improved by the E760 and subsequent E835 collaborations at Fermilab. Fig. 2.6 illustrates the basic principle of this method. It involves the consequent measurement of the production rate at different beam energies. The beam energy is scanned in fine intervals over the resonance so that the resulting datapoints trace the shape of the resonance. The measured shape is a convolution of the beam profile and the width of the resonance. This leads to

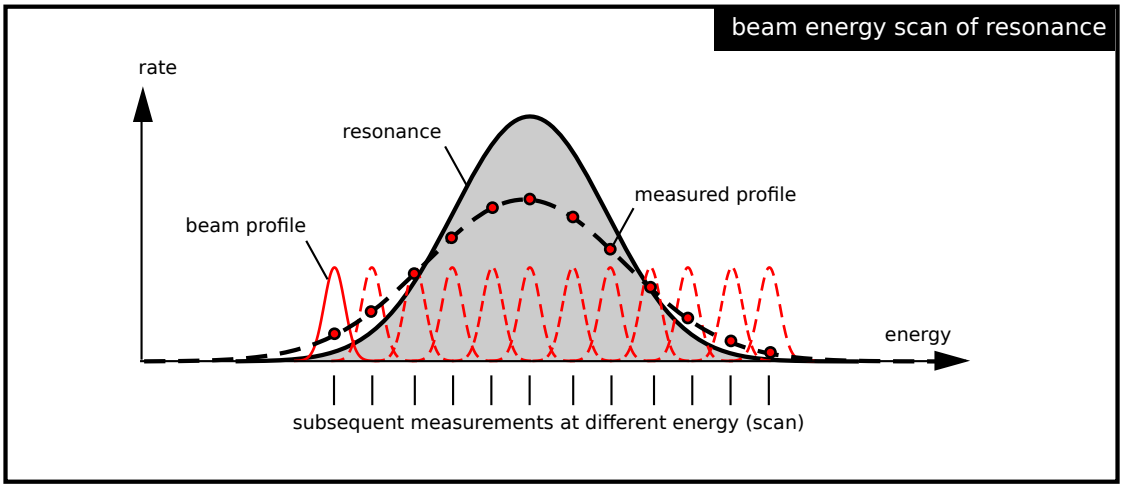


Figure 2.6.: Illustration of the resonance scanning technique.

superior resolution as the beam precision is much better than the energy resolution of current particle detectors. $\overline{\text{P}}$ANDA aims at a mass resolution in the order of 100 keV and width measurements to ~10 % [35].

These experimental conditions build the foundation to successfully address the major physics goals of $\overline{\text{P}}$ANDA which are summarized below.

One goal is the spectroscopy of charmonium states. Charmonia are mesons consisting of one charm and one anti-charm quark which are often described as QCD analogy to the positronium system of the QED. A listing of charmonium states is shown in Fig. 2.7. The E760/E835 experiments have explored the region below the open charm threshold* but had no access to states at higher energies. Meanwhile, experiments like CLEO, BaBar, Belle and BESS have found many of the predicted states above the open charm threshold. In addition, they discovered "XYZ states" which do not fit in the conventional charmonium spectrum. $\overline{\text{P}}$ANDA will contribute to the field by performing precision measurements of charmonia in this region up to an energy of 5.47 GeV/c^2. The beam momentum resolution will be at least equal to the resolution at E835 ($\Delta p/p \sim 10^{-4}$). In high resolution mode, the beam resolution should be one order of magnitude better. The precise measurement of the X(3872) width is one example where $\overline{\text{P}}$ANDA can exploit its unique features [69].

Another important topic is the spectroscopy of open charm mesons (D-mesons). As discussed in a previous section, $\overline{\text{P}}$ANDA will provide a good secondary vertex resolution of ~ 100 µm and good kaon identification over the full solid angle. This combination should enable a good reconstruction efficiency for open charm reactions. Again, resonance scans can be used to provide high precision mass and width measurements.

Further, $\overline{\text{P}}$ANDA searches for gluonic excitations. The QCD allows states with gluons as constituents. Even pure gluonic systems can exist due to the self-interaction of the gluon. States which consist only of gluons are referred to as glueballs, mixed states from quarks and gluons as hybrids. These states are allowed to have exotic quantum numbers. As these states do not mix with non-exotic states in the same mass region, the measured properties should be closer to theoretical predictions. Furthermore, as the constituents of glueballs couple solely to the strong interaction, the spectroscopy of glueballs might also shed some light on the question how hadron masses are generated by QCD.

*Defined by the mass of the lightest D meson pair. 2 M($D^{\pm}$) = 3.74 GeV/c^2

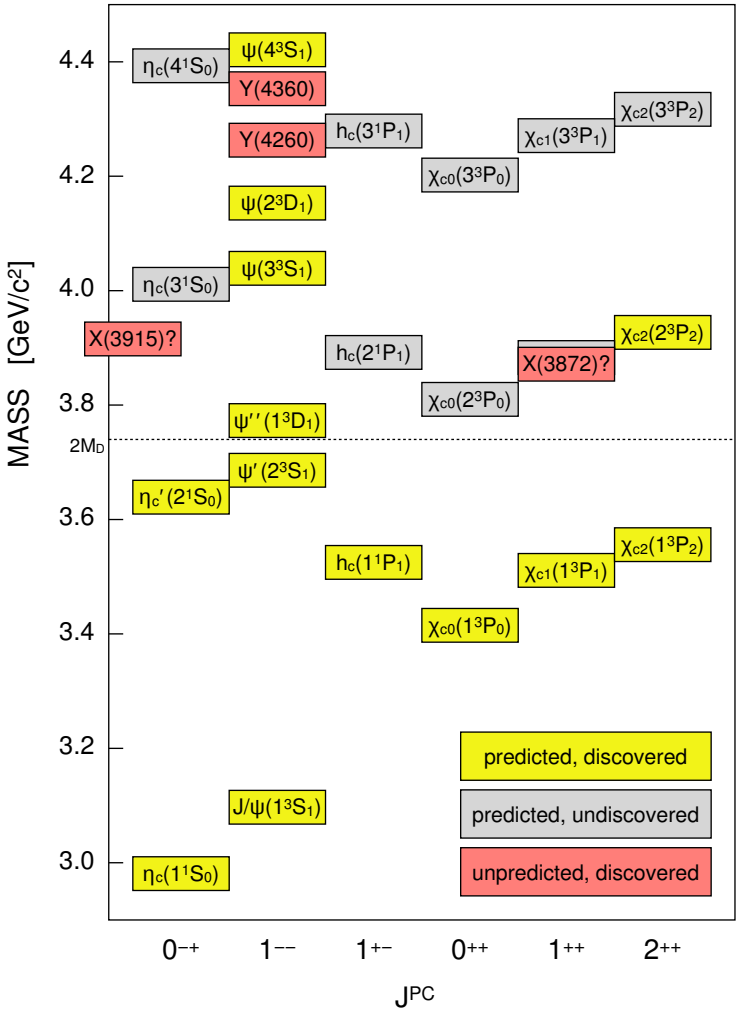


Figure 2.7.: Summary of charmonium states as published in [83]. The open charm threshold is indicated by the dashed line.

The use of heavier targets complements the field of research by gamma-ray spectroscopy of hypernuclei and the study of open and hidden charm in nuclear matter. The reader is referred to [74] for further details on the physics program.

2.5. Technical challenges

The choice of stored antiproton beams as probe implies several technical challenges. The duration of one HESR cycle is $\sim 2\,\mu$s. One bunch of antiprotons is stretched over the full ring, excluding a gap of 400 ns. Hence, interactions are

distributed over a 1.6 μs interval which is followed by a 400 ns gap. The time difference between consequent events follows an exponential distribution, so the time structure of the interaction gives no information on the event time T_0.

A start detector is not advisable as the additional material close to the interaction point would have a major impact on the detector performance. A fine start time with a precision better than 100 ps as needed by time of flight methods will not be available in $\overline{\text{P}}$ANDA. Time of flight detectors will have to work with relative measurements by using two tracks from the same event. However, a coarse event time with a precision in the order of ns can be provided by the MVD.

The significant advantages of the $\bar{p}p$ reaction over e^+e^- have been highlighted in the preceding section. However, there is also a significant disadvantage. The gluon rich $\bar{p}p$ process leads to strong hadronic background. At $\overline{\text{P}}$ANDA energies the total $\bar{p}p$ cross section $\sigma_{\bar{p}p}$ varies from 50 mb to 100 mb while the expected cross section of many reactions of interest σ_{phy} is lower than 100 nb. Hence the fraction of the interesting data in the overall data stream is only in the order of 10^{-6}.

A typical estimation of the total event size in $\overline{\text{P}}$ANDA is $\sim 10^4$ Bytes. This corresponds to a data rate of 200 GB/s at the nominal interaction rate of 20 MHz. As the storage of high amounts of data is impracticable and expensive, the data rate has to be reduced. The anticipated data reduction factor is $\leq 10^{-3}$.

Data reduction is a common problem which all high energy physics experiments have to solve. It can be realized by either selecting only interesting events from the data stream or by neglecting events which have been identified as uninteresting. The typical approach is a multi-level trigger system where the output of each level forms the input of the next one. In these setups, the first trigger level is usually a hardware trigger which takes only a subset of the detector data to decide whether the recorded data may contain interesting physics. In this case, the whole record is pushed to the succeeding trigger level.

The diverse $\overline{\text{P}}$ANDA physics program with its many channels of interest makes the utilization of such a first level hardware trigger impractical. Especially, because most channels have many hadrons in the final state and cannot be easily separated from background signals.

Therefore the $\overline{\text{P}}$ANDA collaboration decided to skip the first level trigger and work only with a special high level trigger which involves the online reconstruction of particle tracks, event building and online PID in a push-only architecture. The resulting information is fed into event selection algorithms which extract the

physically relevant events. The system can be understood as an event filter which operates on the full raw data stream of 200 GB/s.

A key component of this trigger infrastructure is the time distribution system SODA which delivers precise timestamps to all sub-detectors. The sub-detectors have to operate self triggered. Signal digitization and preprocessing has to be performed at the detector front-end and the following data-concentrators. SODA time stamps have to be assigned to all data. The resulting data stream is then fed into the high level trigger network. Due to the complicated time structure of the interactions, the precise time tagging is needed to assign individual hits to tracks and events.

While this trigger architecture has the advantage of being very flexible, the implementation of low latency online reconstruction algorithms for precise tracking and PID is extremely challenging. Hence, these topics are still under research.

3. Cherenkov detectors and their role in particle identification

3.1. Particle identification (PID)

In particle physics, most phenomena are short-lived and thus observed indirectly by their decay products. In many cases it is necessary to identify the particle species of these decay products to understand the underlying physics of a certain reaction. In practice, only particles with a sufficiently long lifetime are relevant. These are photons (γ), long-lived leptons ($e^{\pm}$, $\mu^{\pm}$) and hadrons (p, $\pi^{\pm}$, $K^{\pm}$, K^0_L, n). Modern detectors in particle physics use a combination of different approaches to identify these particles. A typical detector setup is shown in figure 3.1.

Discrimination between charged and neutral particles is provided by a tracking detector in a magnetic field. Charged particles interact with the tracking medium mainly by ionization processes and therefore leave a track in this detector while the minimum-ionizing neutrals are invisible to the device. Further, the bending of a charged particle track in the magnetic field reveals the charge-sign and momentum of the particle. Some decayed particles can be identified by high resolution trackers due to characteristic features in the track topology. Decays like $K^{\pm} \rightarrow \mu^{\pm}\nu_{\mu}$ and

$K^{\pm} \rightarrow \pi^{\pm}\pi_0$ show a "kink" in the track which is caused by the decay of a visible charged track to a visible charged and invisible uncharged particle. The decay of an invisible neutral particle in two visible charged tracks, e.g. $K^0_s \rightarrow \pi^+\pi^-$ or $\Lambda_0 \rightarrow p\pi^-$, can be detected by the "vee" structure formed by the charged tracks.

Another important detector is the electromagnetic calorimeter (EMC) which

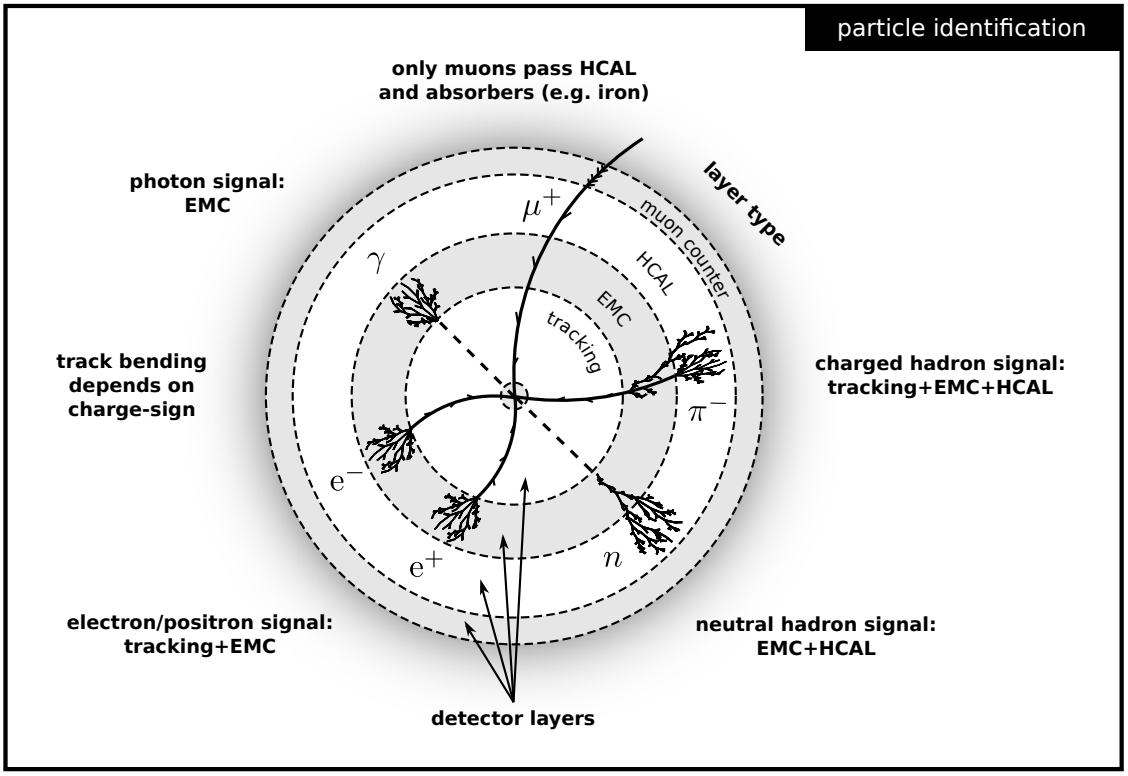


Figure 3.1.: Illustration of particle identification concepts in high energy physics experiments. The innermost detectors provide charged particle tracking, the following electromagnetic and hadronic calorimeters (EMC, HCAL) are used to measure the energy of particles. Muons can traverse all detectors and are identified by detecting a charged track behind the calorimeters. Other particle species can be identified by the different ways they interact with the detectors. To separate charged hadrons like pions and kaons at momenta > 1 GeV/c an additional velocity measurement by time-of-flight and/or Cherenkov detectors is necessary (not shown).

measures the energy deposit of traversing particles by means of the light emitted by a scintillating material. Photons, electrons and positrons are stopped, i.e. they deposit all their energy in the device, while muons and hadrons loose only a fraction of their energy and traverse the calorimeter. While EMC-signatures of $e^{\pm}$ and γ are indistinguishable they can be easily discriminated using the charge information provided by the tracking detector.

Often a hadronic calorimeter (HCAL) is placed behind the EMC to stop hadrons. The only particles which are able to pass the HCAL are muons. These can be detected and thus identified in a subsequent detector. The presence of a HCAL signal discriminates hadrons from $e^{\pm}$ and γ, which have already been stopped

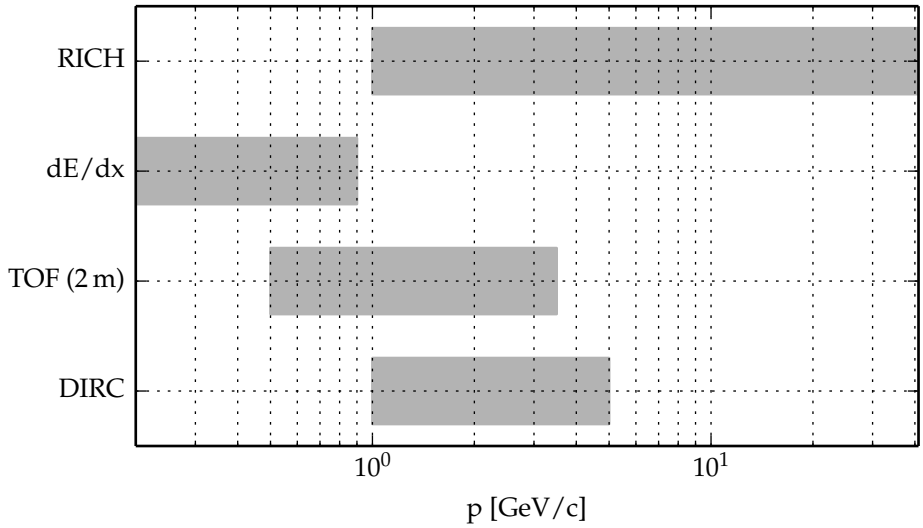


Figure 3.2.: Approximate momentum coverage of different PID techniques for π/K-separation.

in the EMC. However, individual hadron species can not be separated in current hadronic calorimeters due to their limited energy resolution*.

In case the detector is realized without a HCAL the separation of $e^{\pm}$ from charged hadrons is still feasible by using other observables. In contrast to hadrons, $e^{\pm}$ deposit all their energy in the EMC so that the measured track momentum must be consistent with the deposited energy. In this case, tracking information can be used to discriminate $e^{\pm}$ and charged hadrons. Additionally, the lateral shower shape can be analyzed to differ between hadronic and EM cascades. The latter technique can also be used to separate photons from neutral hadrons.

In any case charged hadron identification cannot be effectively provided by calorimeters. The identification of these particles is the task of dedicated detectors. Several detector concepts exist which differ in resolution, momentum range, radiation length, space requirements and financial budget.

At momenta below 0.7 GeV/c, the energy loss in the tracking detector can be

*During the hadronic cascade, neutral pions are created which initiate an EM-cascade via the decay $\pi_0 \rightarrow \gamma\gamma$. As the HCAL response is usually different for EM and hadronic cascades, the fluctuations in the neutral pion production degrade the energy resolution. The reason for the different response is given by the circumstance that a fraction of the energy deposited by hadronic showers is invisible to the calorimeter. This energy is consumed in nuclear processes which predominantly produce low energetic particles.

used for charged hadron identification. Above 0.7 GeV/c up to 1 GeV/c, PID by energy loss is not possible at all (crossover region) and at higher momenta the PID performance is rather limited. Good PID at higher momenta can be provided by time-of-flight (TOF), Cherenkov and transition radiation detectors. The latter are mainly used for e/π-separation. A comparison of the momentum coverage of different PID techniques for π/K-separation is given in Fig. 3.2. Not all π/K can be directly identified due to the limited lifetime, as illustrated in Fig. 3.4.

Considering the momentum coverage of a TOF detector, one could raise the question whether the Endcap DIRC can be replaced by such a system. The $\overline{\text{P}}$ANDA Endcap DIRC will be 2 m away from the interaction point (IP). The PID-performance of a 2 m TOF detector for different time resolution and momenta is presented in Fig. 3.3. A TOF detector would require a time resolution of 10–15 ps to reach a competitive performance. This would also involve the installation of a start-counter close to the IP. The additional material of such a start counter is likely to have a negative impact on the overall performance of the spectrometer.

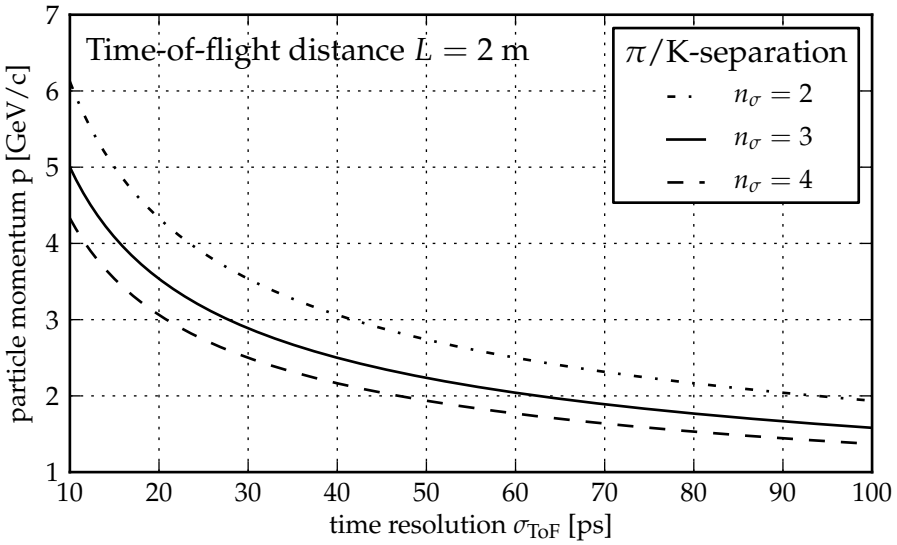


Figure 3.3.: Theoretical π/K-separation for a time-of-flight measurement over a distance of 2 m which corresponds to the IP-endcap distance in the $\overline{\text{P}}$ANDA target spectrometer.

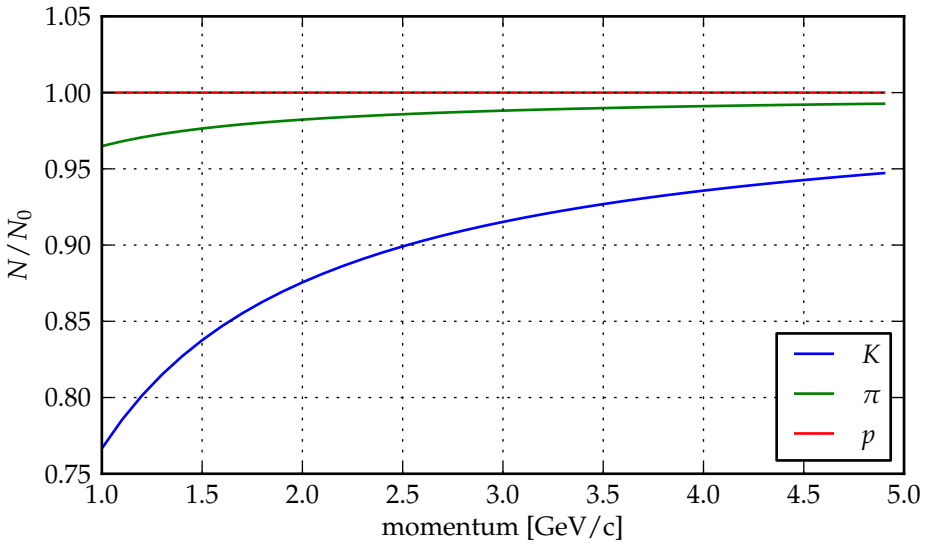


Figure 3.4.: Fraction of particles which are not decayed after traveling a distance of 2 m with a given momentum.

3.2. Separation Power

To compare the performance of different PID systems, it has to be quantified in a way which is independent of the underlying technical details. A common measure is the *separation power* n_σ, which is defined as the significance of the measured quantities:

$$n_\sigma = \frac{|\mu_1 - \mu_2|}{1/2(\sigma_1 + \sigma_2)} \tag{3.1}$$

with mean μ_i and standard deviation σ_i of two measured distributions ($i \in \{1,2\}$). Each distribution corresponds to a certain class of events. A graphical representation of this definition is shown in Fig. 3.5. A classifier $C = \mu_1 + \frac{n_\sigma}{2}\sigma_1 = \mu_2 - \frac{n_\sigma}{2}\sigma_2$ can be used to judge whether a measured quantity x is belonging to class 1 ($x < C$) or class 2 ($x > C$). In this case, the probability $P_{\text{misid.}}$ that a quantity is assigned to the wrong class (*misidentification*) is equal for both classes. The choice of a classifier closer to μ_1 (μ_2) will increase the misidentification for events of class 1 (2) in favor of class 2 (1) while increasing the purity $1 - P_{\text{misid.}}$ for events classified as class 1 (2).

In case of a Cherenkov detector, the two distributions correspond to the distribution of reconstructed Cherenkov angles θ_c for two particle species to separate,

e.g. pions and kaons. Assuming that these distributions are gaussian without systematic errors, the mean is equal to the expected Cherenkov angle at the given momentum p. If the error of the Cherenkov angle is independent of the particle species, (3.1) simplifies to:

$$n_\sigma = \frac{\theta_{c,\pi}(p) - \theta_{c,K}(p)}{\sigma_{\theta_c}} \tag{3.2}$$

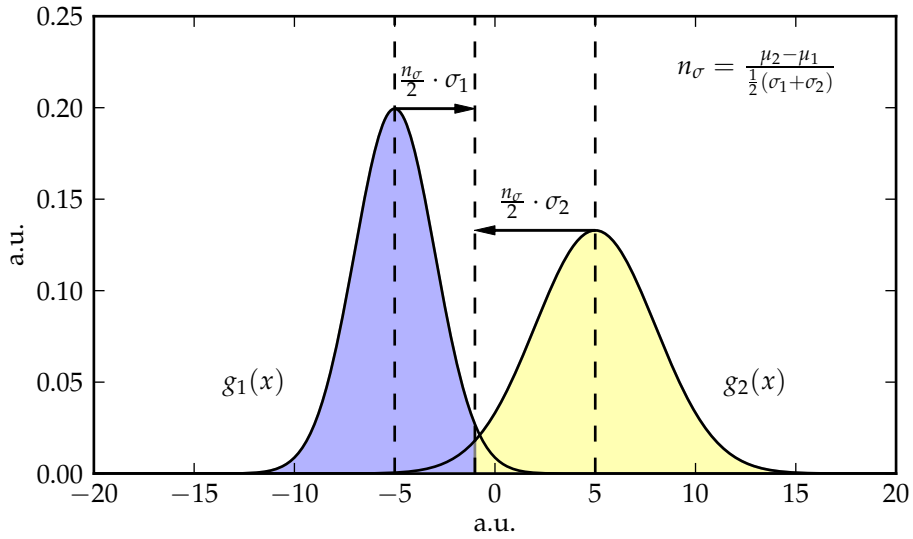


Figure 3.5.: Definition of the separation power n_σ.

.

The corresponding probability of misidentification $P_{\text{misid.}}$ for the classificator C can be computed by integrating the standard normal distribution from $n_\sigma/2$ to ∞.

$$P_{\text{misid.}}(n_\sigma) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{n_\sigma}{2\sqrt{2}}\right)\right] \tag{3.3}$$

3.3. Cherenkov detectors

About the year 1934 the doctoral student Pawel A. Cherenkov accidentally discovered a weak emission of visible light by purified liquids under γ- and β-irradiation. He found that the effect was independent of the liquid's purity and that the emitted light had a puzzling spatial asymmetry. His supervisor Sergey I. Vavilov interpreted *"the radiation observed as a result of the retardation of the Compton*

electrons liberated in liquids by γ-rays." [54]. This finding inspired I. M. Frank and I. E. Tamm to develop a theory which explains the emission characteristics in detail, including it's asymmetry, intensity and wavelength dependence. The phenomenon has been named *Cherenkov effect** and enriched the scientific community by enabling a new class of detectors to measure the velocity of fast charged particles.

Frank and Tamm explained the emission of Cherenkov light in a medium with refractive index n as a shockwave-like effect caused by the transition of a fast charged particle with a velocity v_{particle} exceeding the phase velocity of light in this medium ($v_{\text{phase}} = c/n < v_{\text{particle}}$). The phenomenon can be understood qualitatively as an analogon to the supersonic boom. The resulting electromagnetic radiation is emitted isotropic on the envelope of a cone coaxial to the particle direction and with opening angle $2\theta_c$. The angle between each photon and the particle direction θ_c, named *Cherenkov angle*, is a function of the particle velocity:

$$\theta_c = \arccos\left(\frac{1}{\beta n}\right) \quad \text{with} \quad \beta = v/c \tag{3.4}$$

The velocity threshold for Cherenkov emission can be directly derived from this expression by solving for $\cos\theta_c = 1$:

$$\beta_{\text{thresh.}} = \frac{1}{n} \tag{3.5}$$

The number of emitted Cherenkov photons N_{ph} per length l of the particle track is given by the *Frank-Tamm equation*:

$$\frac{\mathrm{d}^2 N_{ph}}{\mathrm{d}E\,\mathrm{d}l} = \frac{\alpha z^2}{\hbar c}\left(1 - \frac{1}{(\beta n)^2}\right) = \frac{\alpha z^2}{\hbar c}\sin^2\theta_c \tag{3.6}$$

with energy E, the particles integer charge z and the fine structure constant α. Using $E = hc/\lambda$ one can transform this to:

$$\frac{\mathrm{d}N_{ph}}{\mathrm{d}l} = 2\pi\alpha z^2 \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} \frac{\sin^2\theta_c}{\lambda^2}\,\mathrm{d}\lambda \tag{3.7}$$

Note that due to dispersion, the refractive index is a function of the wavelength ($n = n(\lambda)$ and so the Cherenkov angle ($\theta_c = \theta_c(\lambda)$). By ignoring the dispersion

*sometimes also Vavilov-Cherenkov effect

and assuming a wavelength interval of 400 nm $< \lambda <$ 700 nm one can estimate the number of emitted photons per cm track-length as $N_{ph}/\text{cm} \approx 490 \sin^2\theta_c$. Typical values for n are in the range of 1.000295 (air at STP) to 1.67 (lead glass) corresponding to only $N_{ph}/\text{cm} \approx 0.3 - 314.3$.

The detection of such low levels of light can be a technical challenge and it was only by virtue of the outstanding sensitivity of the human eye that Cherenkov was able to discover this phenomenon. Because of the very low number of photons produced, it was not until the advent of photomultipliers that Cherenkov detectors became technically feasible. From then they have been used in many experiments including the setup by Emilio Segr'e, Owen Chamberlain et al. to discover the antiproton*. Cherenkov, Frank and Tamm were honored with the Nobel Prize in physics in 1958 "*for the discovery and the interpretation of the Cherenkov effect*". Segr'e and Chamberlain were honored one year later with the Nobel Prize in physics "*for their discovery of the antiproton"*.

In 1960, Arthur Roberts proposed a "*new type of Cherenkov detector for the accurate measurement of particle velocity and direction*" [111]

> *... in which the light emitted by a single particle traversing a radiator is imaged, by means of a lens or mirror focused at infinity, on the cathode of an image-intensifier tube. The image is a ring, whose diameter measures accurately the Čerenkov cone angle, and thus the particle velocity. In addition the coordinates of the center of the circular image accurately indicate the orientation of the particle trajectory (though not its position). ...*

This is the first proposal of a RICH detector, however with a small active area. Larger RICH detectors became available 10 years later due to the development of large area UV photon detectors based on multiwire proportional chambers (MWPC) by J. Seguinot and T. Ypsilantis [118]. These detectors showed a quantum efficiency of 50 % for UV-photons with an energy exceeding 9.5 eV. From there, the development of RICH detectors took still a long way with successes and disappointments until this technology reached a mature state. Since about the mid–90's RICH detectors are a reliable and well established component in many

* The experiment consisted of a magnetic spectrometer to select the particle momentum combined with two scintillation-counters for a time of flight measurement and two Cherenkov counters for an additional velocity selection which permitted to reject background caused by the mesons in the beam.

physics experiments. An overview about this development is given in [31]. Two recent RICH designs are presented Fig 3.6 to provide an example of the setup and form factor of modern ring imaging Cherenkov detectors.

The left figure, published in [60], illustrates the setup of one out of two RICH detectors at the LHCb experiment. This RICH comprises two different radiator materials, silica aerogel ($n = 1.03$ at $\lambda = 543.5$) and C_4F_{10} gas ($n = 1.0014$ at $\lambda = 400\,\text{nm}$), to cover a total momentum range of 1–40 GeV/c. At momenta below 3 GeV/c, the aerogel radiator yields the best particle separation in terms of Cherenkov angle due to the higher refractive index. π/K-separation by means of aerogel is feasible up to about 10 GeV/c. Cherenkov thresholds for π/K in the C_4F_{10} radiator are at momenta of about 3 GeV/c and 10 GeV/c respectively. Hence, the gas radiator will yield the best PID performance at momenta from 10–40 GeV/c. At momenta from 3–10 GeV/c the additional pion signal in the gas radiator can be used to support the aerogel signal.

The measurement of the Cherenkov angle is realized by an optical imaging system based on metal coated mirrors. The imaging plane is instrumented with hybrid photodetector tubes which have been specifically developed for this detector. Each tube provides 1024 logical pixels, corresponding to a pixel size of $2.5 \times 2.5\,\text{mm}^2$ at the entrance window [60]. The dimension of the whole RICH in beam direction is roughly 1 m.

Beside this classical RICH design based on spherical focusing mirrors, there are other interesting design alternatives which lead to a thinner detector, e.g. 30 cm in beam direction. Such an option is the Belle II focusing aerogel RICH, which is currently being developed. The working principle is illustrated at the right side of Fig. 3.6. The system uses multiple aerogel radiators with a slightly different refractive index $n_2 > n_1$. The Cherenkov light is simply imaged on a large plane covered with multi-pixel hybrid photodiodes*. The focusing of the Cherenkov ring is realized by tailoring n_1 and n_2 in such a way that the Cherenkov photons from both radiators are imaged onto the same ring at the image plane. This detector aims at a $4\,\sigma$ π/K-separation up to 4 GeV/c.

Because of its compactness and similar performance requirements, this detector could be considered as an alternative to the Endcap DIRC design presented in this thesis. However, in contrast to the Endcap DIRC, the photo-sensors of the focusing aerogel RICH have to cover the full acceptance and are thus exposed to much

*These devices will be introduced in section 5.5

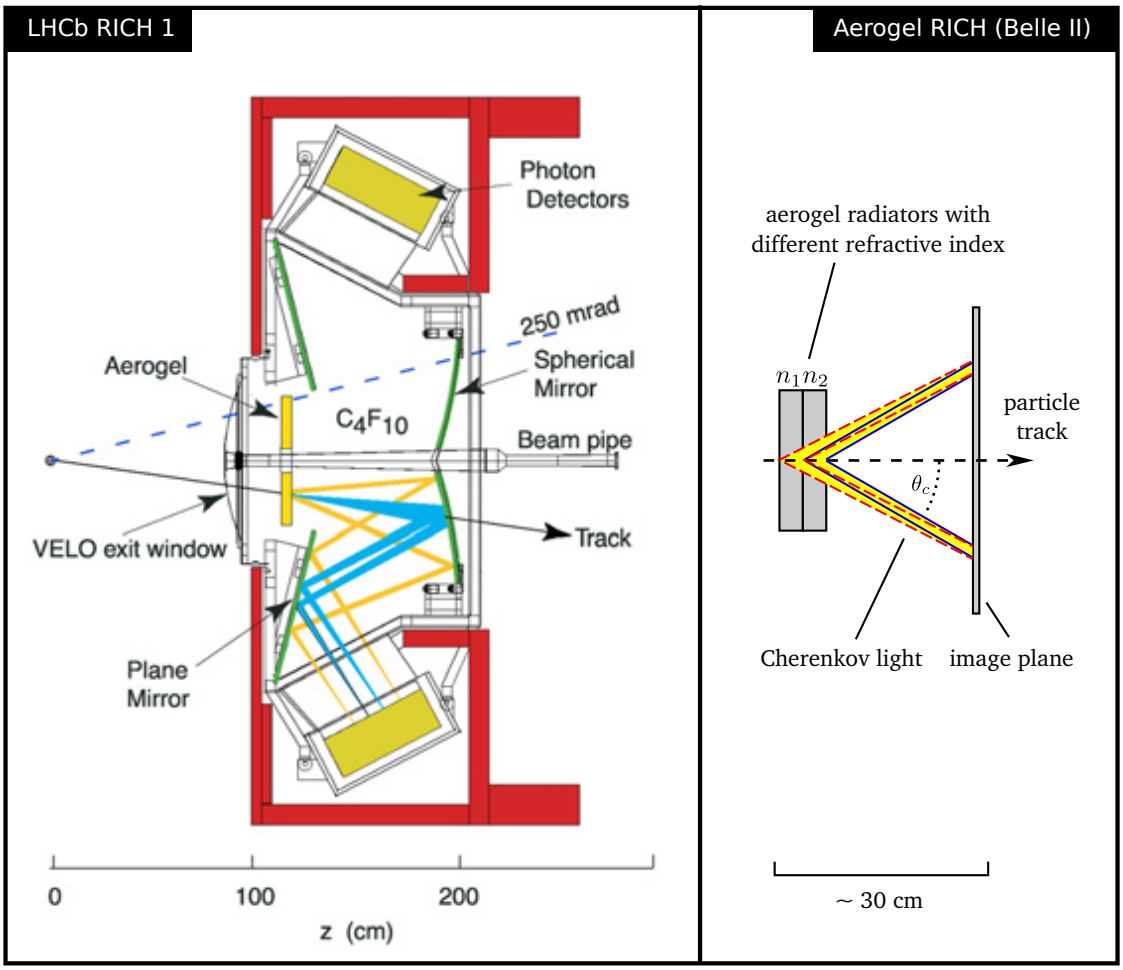


Figure 3.6.: Setup of one of the two LHCb RICH detectors (left) and the working principle of the proximity focusing aerogel RICH developed for Belle II. Left figure from [60].

higher radiation levels. This can pose major difficulties because the radiation load in $\bar{p}p$ experiments is much higher than in e^+e^- experiments. The Endcap DIRC geometry is a much better match to fixed target experiments with high neutron and charged hadron fluence.

3.4. DIRC

DIRC is an acronym for "Detection of internally reflected Cherenkov light". It's concept makes use of a few cm thick fused silica radiator which serves at the same time as an angle-preserving lightguide. A fraction of the emitted Cherenkov photons becomes trapped inside this radiator by total internal reflection and is transported by consecutive internal reflections, similarly to photons in an optical fiber. The key difference is that a DIRC radiator is realized as bar or plate with parallel and orthogonal surfaces which preserve the angular information during transport. At the end of the bar or plate, photons can exit into an optical system where they are detected. This technique allows to cover large areas inside a particle detector while, in opposite to most conventional RICH devices, the space-consuming optical imaging system is located outside the machine.

This is nicely illustrated in Fig. 3.7, taken from [73] which shows the standoff volume of the BaBar DIRC outside of the BaBar detector (upper illustration). The part inside the spectrometer is just a barrel out of 1.725 cm thick fused silica bars plus support and housing. The working principle of the BaBar DIRC is outlined in the illustration below. The Cherenkov photons are emitted, partially trapped and transported as explained above. At the end of the bars, the photons enter a fused silica wedge which is coupled to the expansion volume filled with purified water. The wall of the expansion volume is covered with about 11000 single channel photomultiplier tubes. The position of the tube and the exit window of the wedge define the angle of the detected photon.

One of the major advantages of this detector type is its low space requirement. By using a DIRC, the outer detectors in a spectrometer can be constructed closer to the interaction region what results in smaller radii R. In case of electromagnetic calorimeters, the reduced radius leads to a significant cost reduction as the area covered by expensive crystals is proportional to R^2. Another advantage is the low radiation length of the system combined with the fact that it can be located very close to the calorimeter what helps to minimize the EMC performance degradation

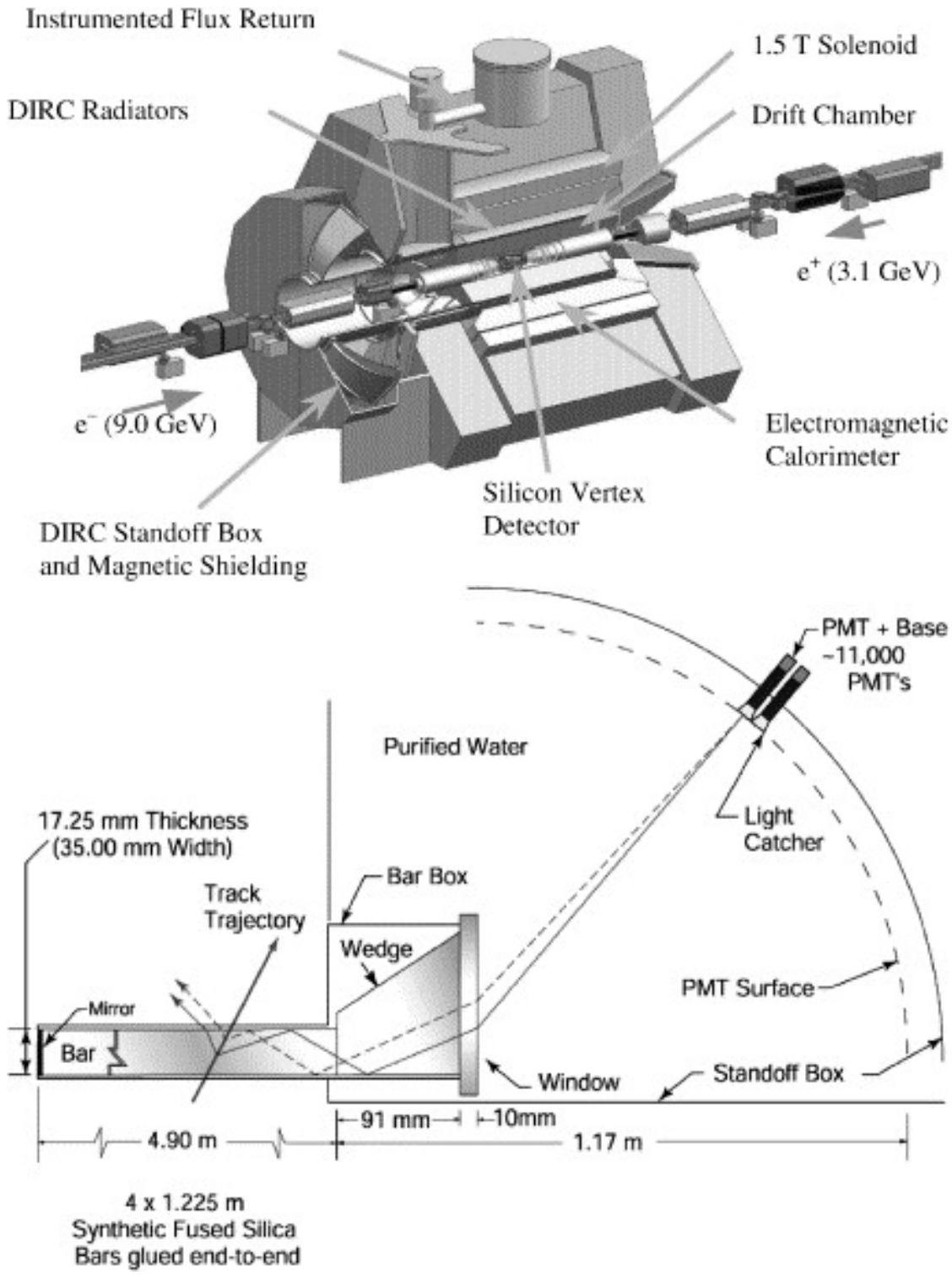


Figure 3.7.: BaBar DIRC setup and working principle. Figures from [73].

due to showering. These properties make DIRC detectors an interesting option for charged particle identification up to momenta of about 4 – 6 GeV/c, depending on the required separation power (compare also figure Fig. 5.3 on page 51).

Up to now, the BaBar DIRC is the only DIRC which has been used in a running experiment. Today, there are three ongoing detector developments which have been adapted from the original BaBar design. The fast focusing DIRC (FDIRC) for Super-B, the Belle II imaging time-of-propagation counter (iTop) and the $\overline{\text{P}}$ANDA Barrel DIRC. The concepts of these detectors are summarized in Fig. 3.8. These illustrations have been extracted from the publications [50, 112, 45].

The FDIRC has been researched as PID solution for the now canceled Super-B project in Italy. In the FDIRC design, the expansion volume has been replaced by a fused-silica focusing block with a aluminum mirror coating. Further, the design aims at the mitigation of chromatic errors by means of a precise time measurement. This technique will be discussed in more detail in section 5.4.4.

The iTop design [45] aims at the measurement of the photon reflection angle φ by means of the time of propagation (eq. 5.6). Additionally, this detector uses wider radiator plates instead of bars and a proximity focusing optics readout with a low spatial resolution. Such a detector requires a single photon time resolution better than 50 ps.

The design of the $\overline{\text{P}}$ANDA Barrel DIRC [50] comprises basically three design options. Radiator bars with either proximity focusing or a focusing lens and finally plates with proximity focusing. The latter option is similar to the iToP design, but with better spatial resolution. The detector aims also at a chromatic correction by means of fast timing.

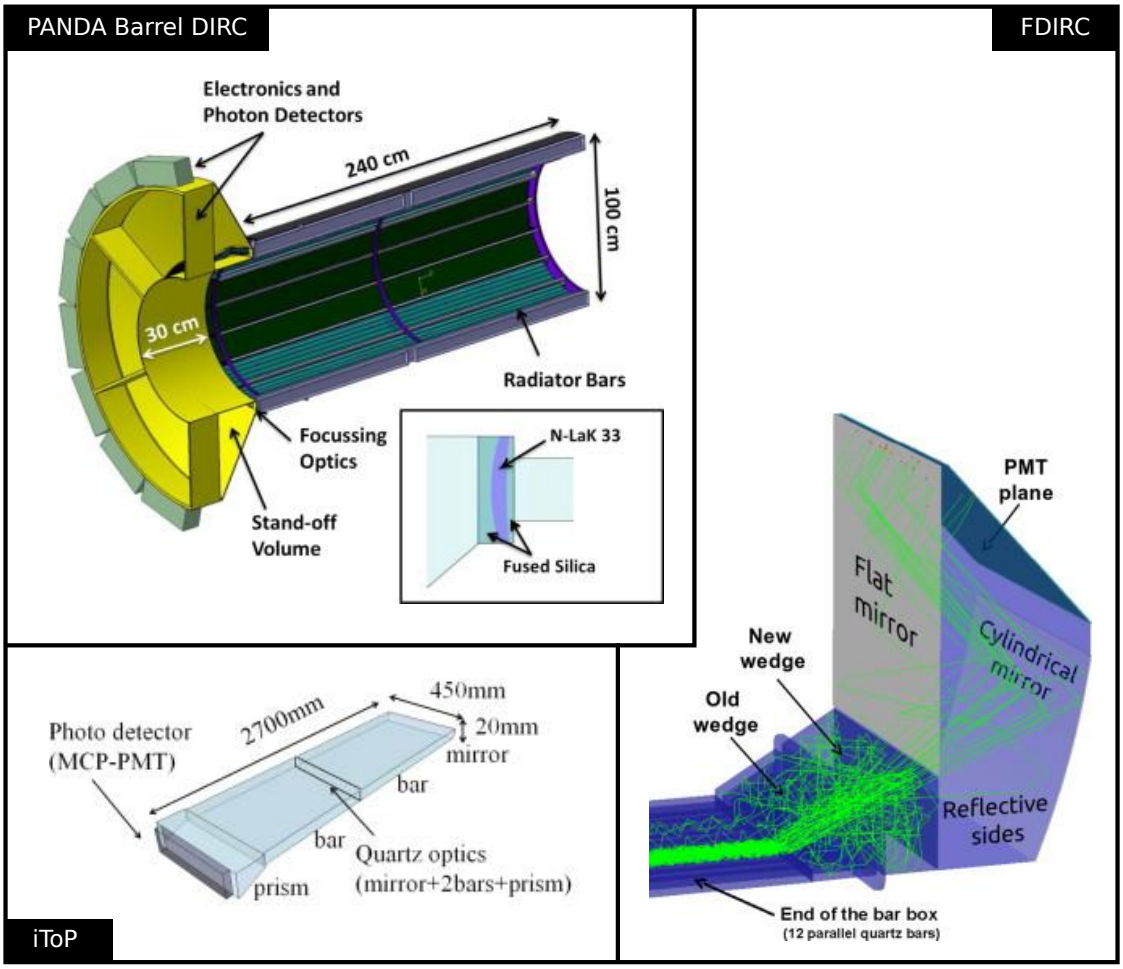


Figure 3.8.: DIRC designs which are currently developed for different experiments. Illustrations from [50] (Barrel DIRC), [112] (FDIRC), [45] (iTop)

Part II.

Detector design

4. The initial idea of a 3D Disc DIRC

Prior to the presented work, there were two competing proposals for an Endcap DIRC in $\overline{\text{P}}$ANDA [116]. A focusing Disc DIRC (FDD), proposed by Klaus Föhl and investigated by researches at Edinburgh and Glasgow universities, as well as the Time-of-Propagation Disc DIRC (TOPDD) proposed by the Gießen group around Michael Düren. The readout of both designs is shown in Fig. 4.1 together with two older Disc DIRC proposals.

The FDD works similar to the FDIRC which has been proposed for Super-B (section 3.4). However, the FDD-optics are significantly smaller and do not require a reflective coating. The curvature allows the incoming photons to be internally reflected. This comes at the price of a complicated, aspheric surface which is represented by a 5th-degree polynomial. Another feature of the FDD is the mitigation of chromatic errors by a LiF crystal, an approach which has been proposed earlier by Boris Morosov. Technical details of this technique will be discussed later in section 5.4.4.

While the FDD is an imaging DIRC detector which measures the photon angle directly, the TOPDD concept exploited the time information to reconstruct the photon angle (eq. 5.6). Hence, the TOPDD is similar to the iTOP counter which is being developed for the Belle II experiment.

Simulation studies performed by the author [79] showed that the TOPDD concept had severe problems in separating background from signal. Background photons can originate from dark counts, the Cherenkov emission of knock-on electrons (compare Fig. 4.3, 4.4 and related captions) and from coincident particle tracks. A typical 2d-signal of the TOPDD is shown in Fig. 4.2a.

To overcome this limitation, the author proposed to merge the TOPDD concept with imaging concepts into a *3D Disc DIRC* [79]. The aim of the 3D concept is to form a synthesis of the pure ToP approach, where the photon angle is reconstructed only from the time-of-arrival, and the focusing approach, where the time resolution is usually not competitive to the imaging resolution.

The advantage is best expressed by a comparison of the expected 2d TOPDD

hit-pattern with an expected 3d-pattern (Fig. 4.2). Structures which overlap in the 2d parameter space can be easily separated by including the 3rd dimension. The presented 3d-pattern is a simulated pattern of a dSiPM based detector design which has been researched in the beginning of this work, until it turned out that the sensors are not radiation tolerant enough to be operated at $\overline{\text{P}}$ANDA. One argument why the detector development started with this type of sensor anyway, is that MCP-PMTs have a limited lifetime as well. The efficiency of MCP-PMTs starts to degrade rapidly after a certain photon fluence.

In the following chapter, the DIRC technology will be dissected. Individual topics are then explained and discussed in detail with respect to available hardware and the applicability for the given Endcap DIRC application. This process will lead to different detector design, specifically tuned to meet the $\overline{\text{P}}$ANDA requirements. This design has been researched as alternative to the initial dSiPM option and finally proposed as Endcap PID solution to the $\overline{\text{P}}$ANDA collaboration. The advantage over the FDD design is the use of simpler, cylindrical optics, a higher spatial resolution and a reduced photon rate which permits to operate the device with current MCP-PMTs.

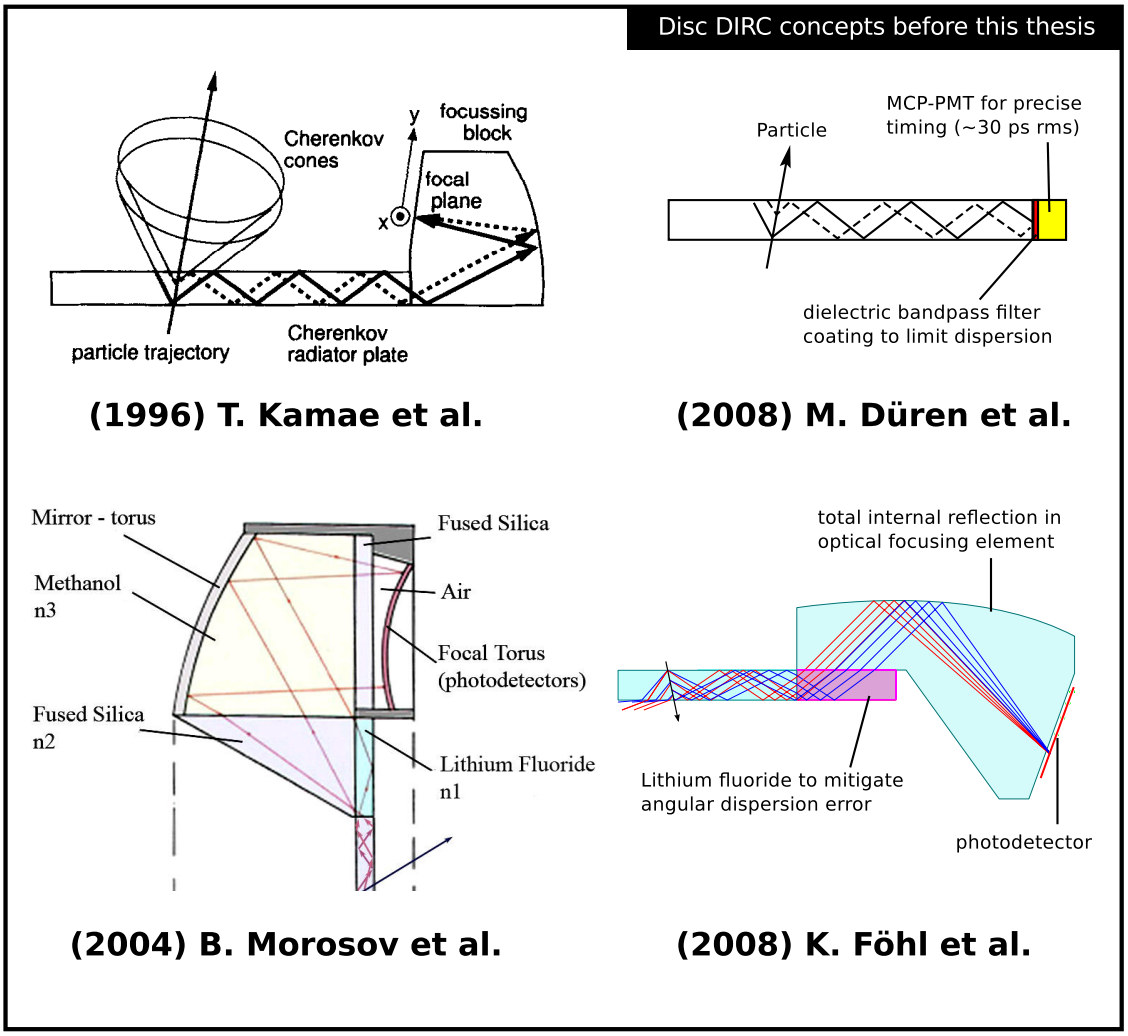
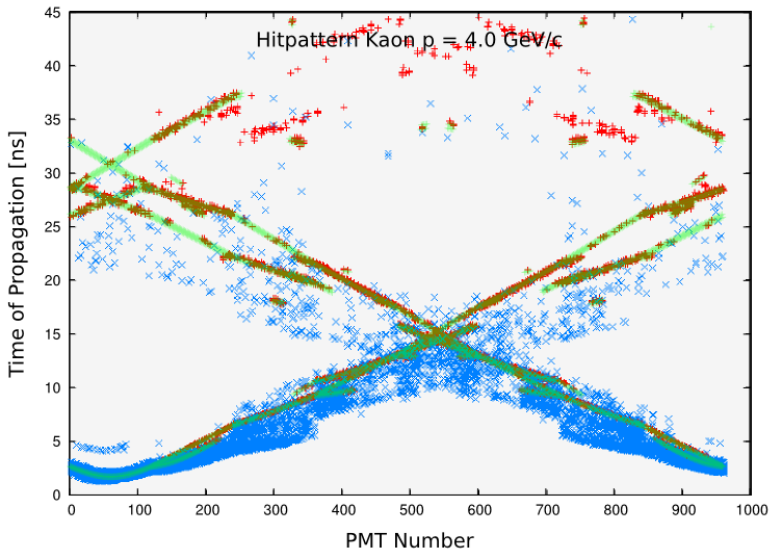
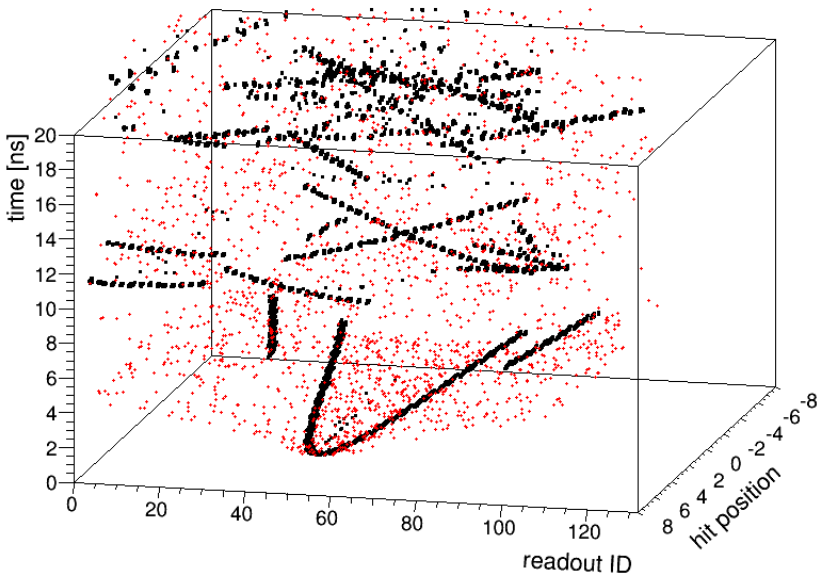


Figure 4.1.: An overview of Disc DIRC designs in literature. Up to now, none of them has ever been realized. All these ideas have been considered in the development of the Endcap DIRC design. Sketches were taken from following publications: T. Kamae et al.: [58], B. Morosov et al.: [87], K. Föhl et al.:[39]. Details on the 2008 PANDA design ideas at the right are presented in [116].

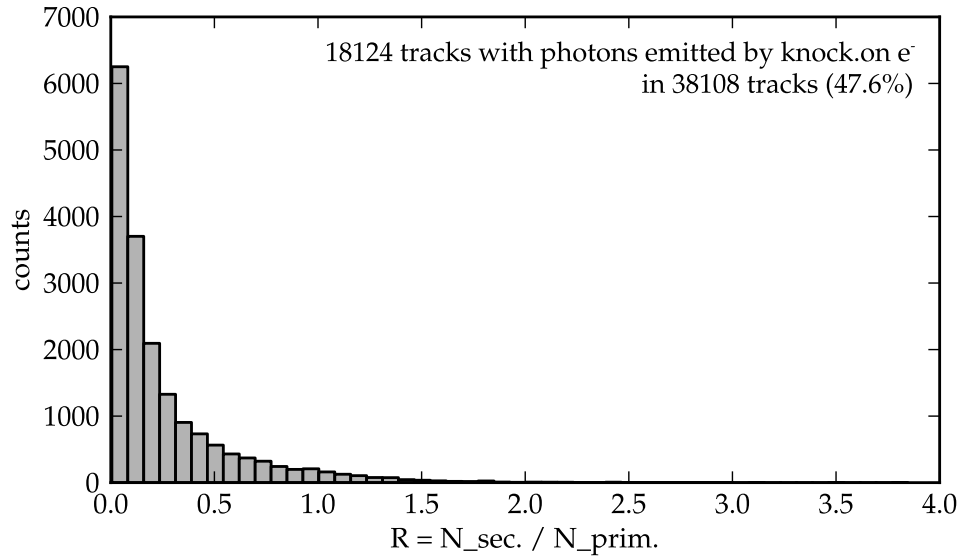


(a)

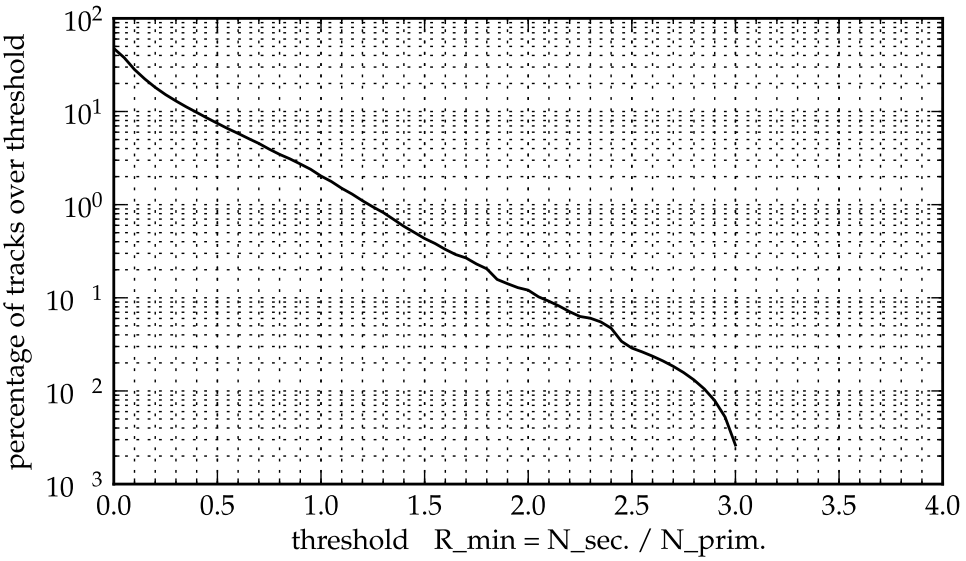


(b)

Figure 4.2.: (a): Simulated ToP DIRC pattern. Red: pattern of photons from 1000 kaons. Blue: hits from background photons emitted by knock-on electrons. Green: overlayed theoretical hitpattern [80]. (b): Simulated pattern of a 3D Disc DIRC, exploiting a precise time and position measurement. Black: photons from primary particles. Red: photons from knock-on electrons. — In two dimensions, photons from knock-on electrons show a strong overlap with the signal. In three dimension however, signal and background can be easily separated. This separation is not only mandatory for an efficient background suppression, but also to separate coincident particle tracks.

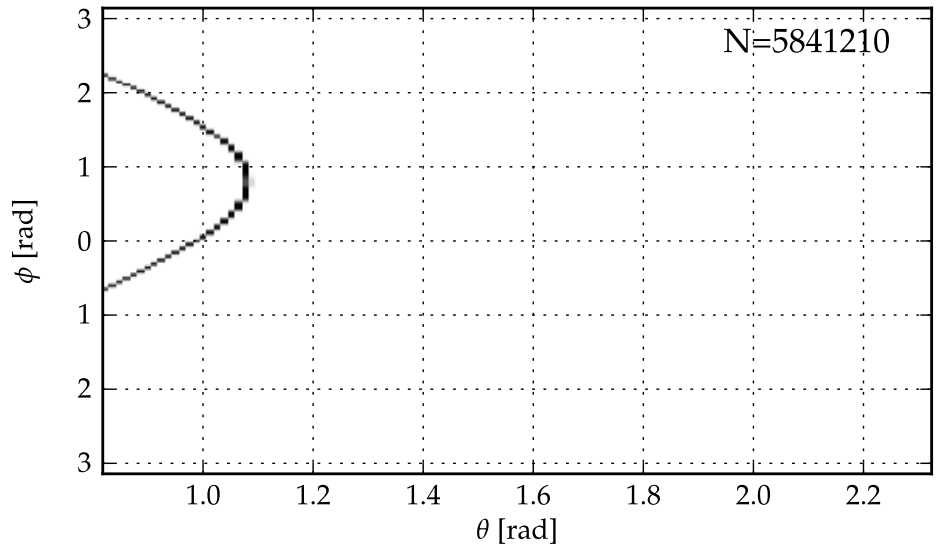


(a) Background level R (ratio of sec. and primary photons)

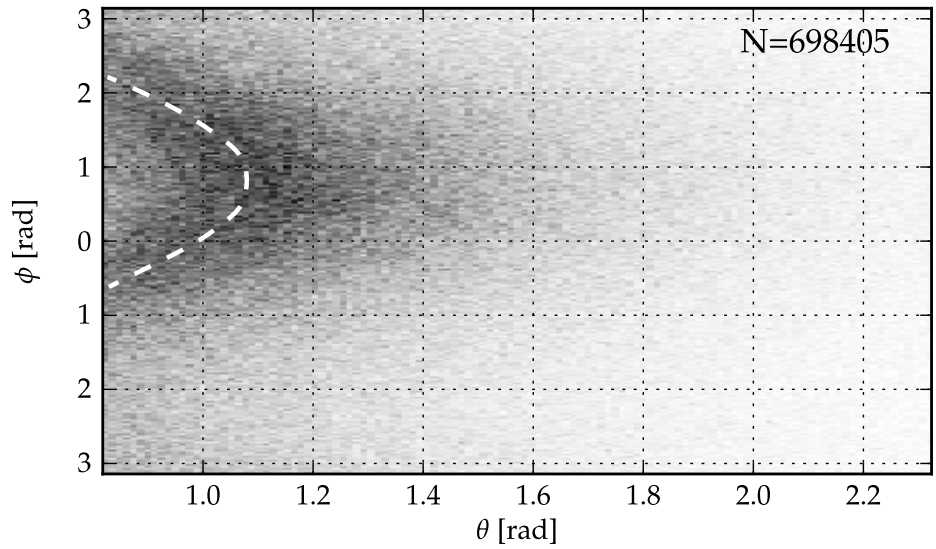


(b) Percentage of tracks with background level $R \geq R_{\text{min.}}$

Figure 4.3.: (a): Histogram of the ratio $R = N_{\text{sec}}/N_{\text{prim}}$ with the number of trapped photons emitted by knock-on electrons N_{sec} and the number of trapped photons emitted by the primary particle N_{prim} for individual tracks (40000 tracks minus 1892 tracks (4.7%) with inelastic collisions). (b): The percentage of tracks with $R > R_{\text{min.}}$ computed from the data histogrammed in (a) versus the threshold $R_{\text{min.}}$. About 7.5 percent of the tracks correspond to $R \geq 0.5$.



(a) photons emitted by primary π^+ (signal)



(b) photons emitted by knock-on e^- (background)

Figure 4.4.: Distribution of the direction vectors of trapped Cherenkov photons at emission by primary particles (top) and knock-on electrons (bottom). The color scale reaching from white (0) to black (max.) differs between both histograms. The z-axis of the coordinate system is defined perpendicular to the radiator plane. The histograms show only the region were the condition of total internal reflection is met. The white dashed line in the lower figure reproduces the emission direction of the primary photons shown in the upper figure. The data has been produced using a Geant 4 simulation (QGSP_BERT physics list with EM option 3). The incident particle is a π^+ with $p = 4$ GeV/c, $\theta = 15°$, $\phi = 45°$.

5. Development of a conceptual detector design

This chapter describes the development process, beginning with the analysis and definition of the system requirements, followed by a first conceptional definition of the system and a dissection into individual system components.

Each of the components is then described, explained and analyzed in a dedicated section. Physics processes and details required for the modeling of the detector are also explained in these sections, together with hardware options, constraints and compatibility regarding the system requirements.

This analysis of system components is then followed by the design synthesis. A concrete detector model including hardware will be defined in this step. This model is then fleshed out further in the succeeding chapter by a successive optimization of the components as well as the whole system.

5.1. Requirement definition and analysis

Considerations regarding the system design have to start with the identification of boundary conditions, design constraints and the required system performance. This information has to be bundled into a first specification of system requirements. These topics are addressed in the following sections, followed by a brief discussion and a summary table 5.4.

5.1.1. Functional and performance requirements

The functional requirements are defined by the setup and physics objectives of the $\overline{\text{P}}$ANDA experiment. The purpose of the Endcap DIRC system is to provide charged hadron identification for tracks at polar angles ranging from 22° to 10°/5° in x/y-direction. Here, the lower limits differ due to the asymmetric acceptance of the target spectrometer. An almost complete coverage of the momentum range

accessible by a DIRC detector (~0.8 GeV/c up to 4 GeV/c or higher) is needed to cope with the Lorentz boost at the highest beam momentum of 15 GeV/c. Further, the trigger-less $\overline{\text{P}}$ANDA DAQ scheme demands PID information to be provided online as input to the event selection algorithm. The minimum requirement for detector life is 10 years of operation at a 50 % duty cycle.

Among the charged, long-lived hadrons which reach the endcap region without decaying into lighter particles, pions and kaons are closest in mass. Since the separation power for all other species is higher, it is appropriate to define the performance requirement in terms of π/K-separation.

The identification of charged kaons in open charm decays is a typical physics case for the Endcap DIRC. For the set of $\overline{\text{P}}$ANDA benchmark channels [74] listed in table 5.1, events have been generated at beam momenta of 15 GeV/c and 10.2 GeV/c using the EvtGen package [68]. Further, events of the reactions $\bar{p}p \to D^+D^-$, $\bar{p}p \to D_s^+D_s^-$ and $\bar{p}p \to h_c \to \eta_c\gamma$ have been generated at their individual thresholds. Decay branches have not been restricted in any of the reactions. The resulting kaon distributions are displayed in figures 5.2 and 5.1, respectively. Individual contributions have not been weighted by their total cross sections to attain an equal contribution to the phase space figure.

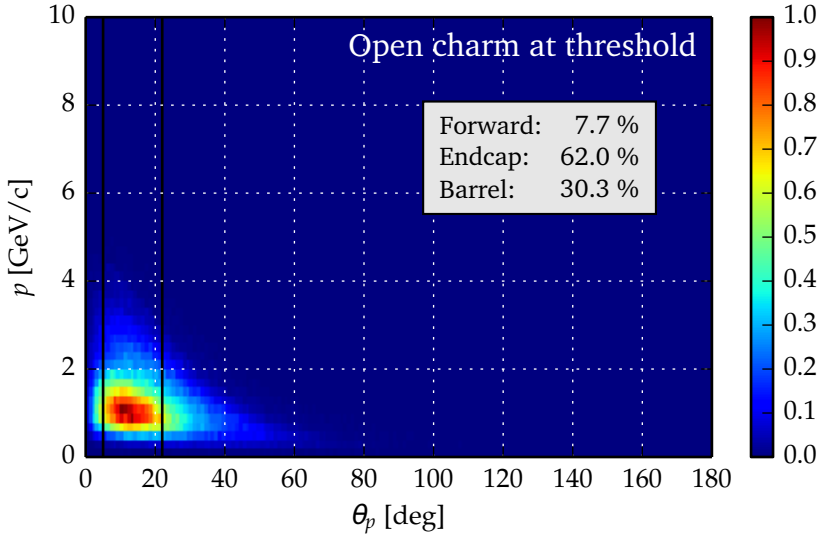
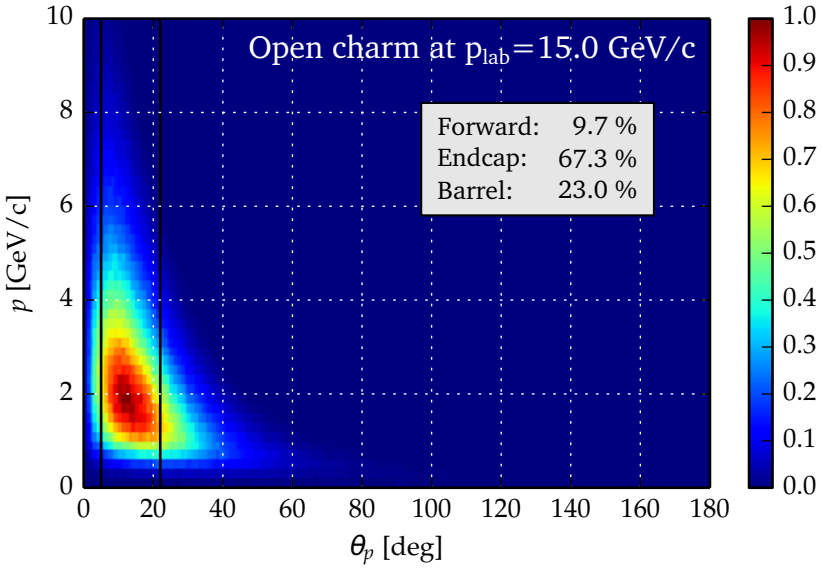
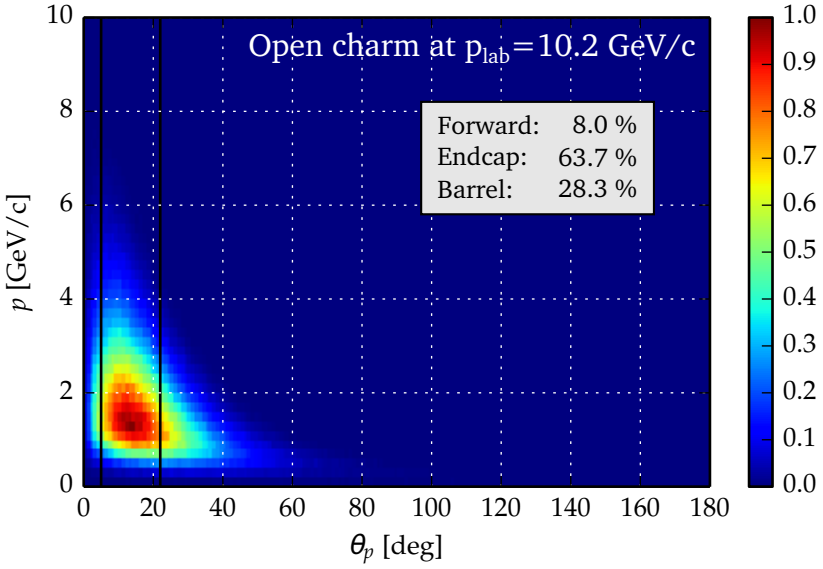


Figure 5.1.: Combined kaon phase space of the reactions $\bar{p}p \to D^+D^-$, $\bar{p}p \to D_s^+D_s^-$ and $\bar{p}p \to h_c \to \eta_c\gamma$ at the individual production threshold. The channels have not been weighted by their individual cross sections and contain all decay branches.



(a)



(b)

Figure 5.2.: Kaon phase space of the open charm channels listed in table 5.1 at a beam momentum of 15 GeV/c and 10.2 GeV/c. The channels have not been weighted by their individual cross sections and contain all decay branches.

- $\bar{p}p \to D^0\bar{D}^0$
- $\bar{p}p \to D^0\bar{D}^0\gamma$
- $\bar{p}p \to D^{*0}\bar{D}^{*0}$
- $\bar{p}p \to D^{*0}\bar{D}^{*0}\gamma$
- $\bar{p}p \to \Lambda_c\bar{\Lambda}_c$
- $\bar{p}p \to D^+D^-$
- $\bar{p}p \to D^+D^-\gamma$
- $\bar{p}p \to D_s^+D_s^-$
- $\bar{p}p \to D_s^+D_s^-\gamma$
- $\bar{p}p \to D^{*+}D^{*-}\gamma$

Table 5.1.: Channels used to generate Fig. 5.2.

In this specific set of reactions, most of the tracks are emitted in forward direction and the fraction of kaons in the Endcap DIRC acceptance increases with the beam momentum. In any of the shown cases, the major kaon fraction is emitted into the Endcap DIRC acceptance. This behavior is a simple consequence of the Lorentz boost of the particle system and can be expected in other reactions as well. Further, Fig. 5.2a shows clearly that the highest momenta are expected closer to the center of the Endcap DIRC. Hence, the performance requirements become θ dependent at momenta > 3.5 GeV/c.

As the DIRC acceptance is usually limited by dispersion, the initial request for π/K-separation is $> 4\sigma$ up to 4 GeV/c, following the definition of separation power presented in section 3.2. This value defines the performance target for system design and development. The realistically achievable performance has to be determined by an iterative development process.

While being a powerful tool, the term of separation power is rather abstract. To get a better imagination of the performance requirement, it is helpful to translate it into more descriptive quantities like misidentification or Cherenkov angle resolution σ_{θ_c}. The misidentification can be directly computed. According to equation 3.3, a separation power of $4\,\sigma$ corresponds to a misidentification of 2.3 %. To translate the $4\,\sigma$ requirement to a Cherenkov angle resolution σ_{θ_c}, one has to *substitute* the Cherenkov angle for pions and kaons at the maximum required momentum in the definition of separation power (eq. 3.2):

$$\sigma_{\theta_c} \leq \frac{1}{4} \cdot \left[\arccos\left(\frac{1}{n} \sqrt{1 + \left(\frac{m_\pi}{4\,\mathrm{GeV}} \right)^2} \right) - \arccos\left(\frac{1}{n} \sqrt{1 + \left(\frac{m_\mathrm{K}}{4\,\mathrm{GeV}} \right)^2} \right) \right] \tag{5.1}$$

For a fused silica radiator with n=1.47 the Cherenkov angle resolution has to be 1.6 mrad to reach a $4\,\sigma$ separation at 4 GeV/c. The correspondence of separation power and σ_{θ_c} at n=1.47 is shown in figure 5.3.

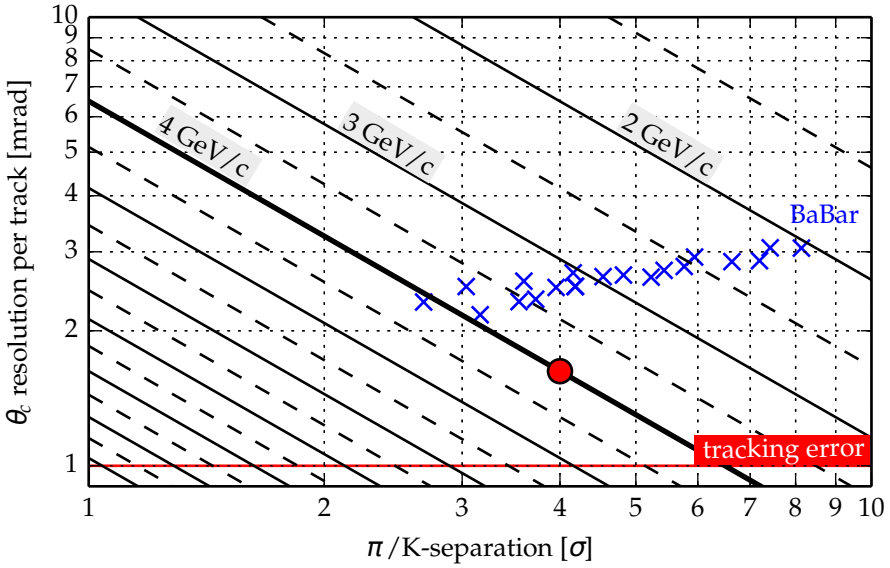


Figure 5.3.: Correspondence of required Cherenkov angle resolution and π/K separation in σ at momenta from 1.5 to 10 GeV/c. Solid lines show integer momenta, dashed lines 0.5 GeV/c steps in between. The targeted performance is marked by the red circle. For comparison, the BaBar separation power obtained from the channel $D^0 \to K^-\pi^+$ as shown in [9] has been added (blue crosses). A decrease of resolution towards lower momenta is expected behavior as the effect of multiple scattering of the particle tracks and energy loss increases with lower momenta.

.

In a DIRC detector, the Cherenkov angle θ_c is obtained by measuring the angular information of several, say $N_{\text{ph.}}$ individual photons. To compute θ_c from these measurements, the direction and position of the particle track is required. This information has to be either provided by the tracking system, as it was the case at BaBar, or reconstructed from the hit information of several photons. Assuming that the track information is provided by the tracking system with an error σ_{track} and no correction takes place to correct for tracking errors, a naive performance figure can be defined as

$$\sigma^2_{\theta_c} = \frac{\sigma^2_{\text{ph.}}}{N_{\text{ph.}}} + \sigma^2_{\text{track}} \tag{5.2}$$

where $\sigma_{\text{ph.}}$ denotes the measurement error of θ_c for a single photon. This error is reduced by taking $N_{\text{ph.}}$ independent measurements into account. As the tracking error is common to all the measurements, it does not scale with $N_{\text{ph.}}$. From

this formula, it is evident that σ_{track} defines the lower limit of the achievable resolution. A reasonable value is 1 mrad which is also shown as limit in figure 5.3. If σ_{track} becomes significantly larger, it will start to dominate σ_{θ_c} and the targeted performance cannot be reached. In consequence, the performance requirement on the Endcap DIRC directly implies a performance requirement for the tracking at the forward endcap, which will be primarily provided by the GEM tracking system with some support by the MVD. It has to be added that the tracking error is usually momentum dependent and the red line in Fig. 5.3 serves just as an indication.

In case of very large $N_{\text{ph.}}$ the tracking error might be reduced by exploiting the correlation of the single photon angles, e.g. by a correction fit which forces the angles to match a common Cherenkov angle. However, as this technique is purely hypothetical it cannot be considered while defining the basic requirements.

It has to be emphasized that equation 5.2 can only be used as a semi-quantitative approximation since the real resolution for single photons $\sigma_{\text{ph.}}$ is generally a complicated function depending on the emission direction of the photon. Also, important effects like the multiple coulomb scattering of the particles are neglected in this expression. Nevertheless, this simple model can be very helpful to understand the relevance of different error sources.

5.1.2. Environmental conditions and design constraints

Geometric constraints

The position of the Endcap DIRC inside the target spectrometer has already been defined prior to this thesis. A volume 194 cm downstream the interaction point, directly in front of the endcap electromagnetic calorimeter is reserved for the installation of the device (Fig. 5.4). This volume, which had to be allocated prior to the finalization of the detector design, defines the geometrical boundary conditions.

This choice of detector position is reasonable. A calorimeter destroys the velocity information of the incident particle, hence the DIRC has to be placed inside the tracking layer (Fig. 3.1). Here, the device can be either positioned at the interface of the tracking and the calorimeter layer (option I) or further inside the tracking layer, sandwiched between individual tracking detectors (option II).

The latter option can be advantageous for fast online reconstruction if the adjacent tracking stations are considerably close to the DIRC, so that the track

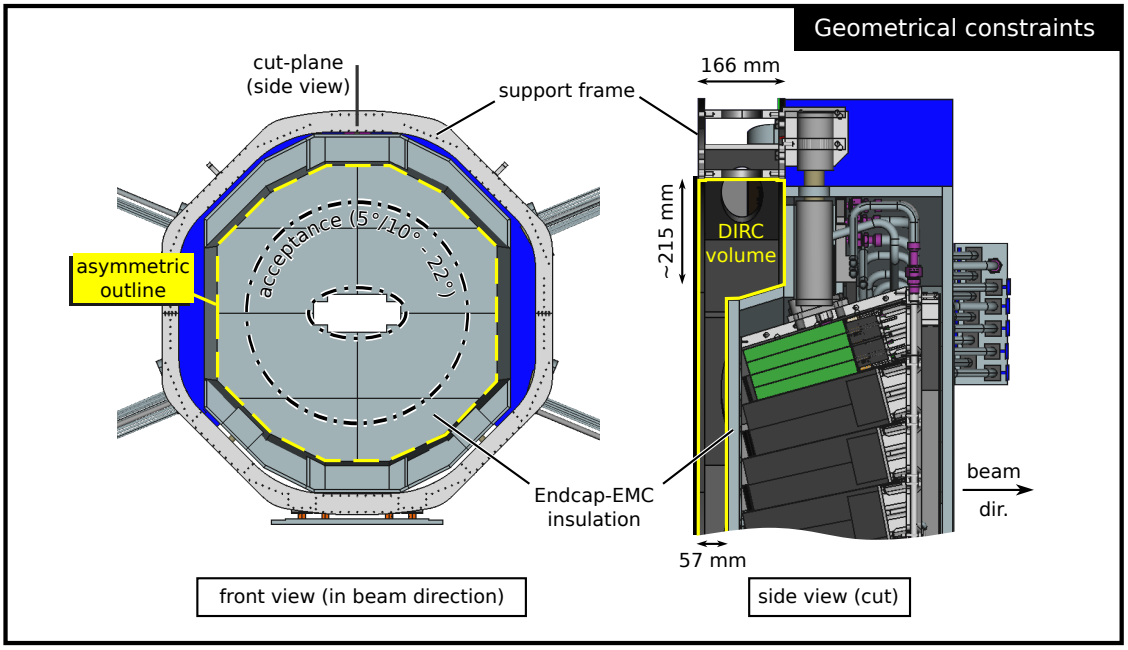


Figure 5.4.: Illustration of the geometric constraints defined by the Endcap EMC geometry. The complete Endcap DIRC system has to fit inside the yellow region marked in the side view. Optics, sensors, front-end electronics and support structures have to share a volume with a cross-section of $\sim 166 \times 215\,\mathrm{mm}^2$ close to the support frame. This shape runs along the asymmetric outline of the calorimeter insulation as shown in the front view (yellow line). The maximum dimension of the radiator in beam direction is 57 mm including housing.

segment in the DIRC can be approximated by using only information of the neighboring trackers. The disadvantage is, that the electromagnetic showers occurring in the DIRC radiator are created further away from the calorimeter what can distort the spatial distribution of the shower and hence the hit-cluster reconstruction of the EMC. The choice of option (I) minimizes these problems by bringing the DIRC close to the calorimeter.

The resulting geometric constraints are summarized in Fig. 5.4. The Endcap DIRC has to fit in the free space between the upstream edge of the EMC support frame and the EMC insulation. A cross section of this volume is marked in the side view. The transition between the thinner part of the volume with a depth of 57 mm in beam direction and the outer part with a cross section of about $166 \times 215\,\mathrm{mm}^2$ is indicated by a yellow dashed line in the front view. The insulation has a slightly asymmetric shape, what has to be taken into account. The maximum of the polar

angle acceptance is also shown in the figure by a black, dashed line defining the minimum area which has to be covered by the radiator. The hole in the middle of the endcap EMC is the acceptance of the forward spectrometer and has to be left open by the DIRC radiator. Any material in this region would distort the resolution of the forward detectors.

While not shown in the drawing, the upstream region in proximity to the support frame is needed to route cables from the inside of the target spectrometer to the outside. The Endcap DIRC has also the option to route a limited amount of cables through the EMC support frame, which also serves as DIRC support structure.

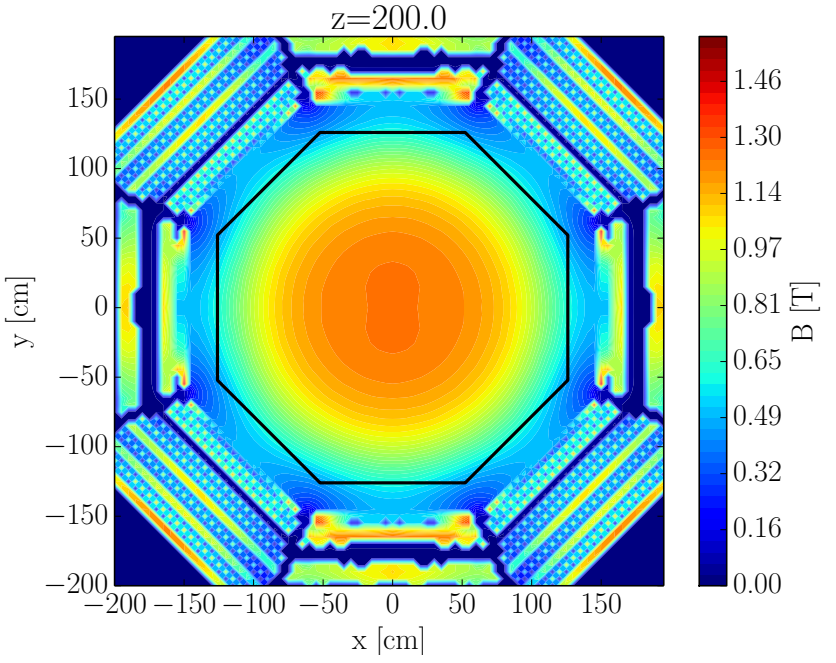
Magnetic field

The complete Endcap DIRC volume is inside the strong magnetic field of the solenoid magnet. This is different in respect to BaBar, where the expansion volume was outside of the spectrometer and photosensors were not exposed to strong fields. A map of the expected magnetic field is shown in Fig. 5.5. This map has been rendered using the field map provided with the PandaRoot package, the official simulation package of the PANDA Collaboration. The configuration of the magnetic field has to be considered during the design phase as it is limiting the operation of photosensors and voltage transformers, e.g. in high voltage supplies. In addition, the field has to be considered in the mechanical design.

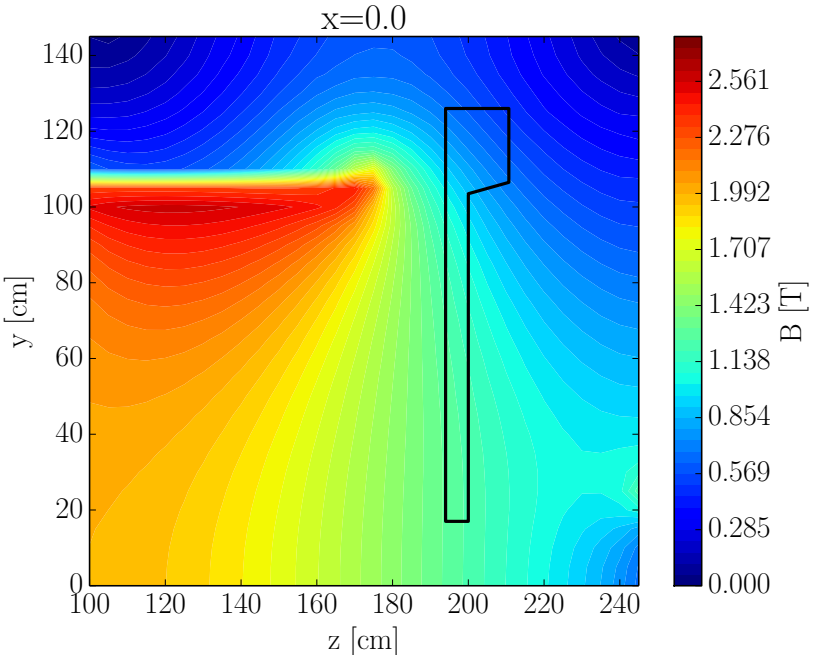
Rate requirements

Another important requirement is the rate capability of the detector. The nominal interaction rate which a detector in $\overline{\mathrm{P}}$ANDA should be able to handle is $R_{\mathrm{nom.}} = 20\,\mathrm{MHz}$. This value already accounts for fluctuations in target density up to a factor of two [109]. These fluctuations are an intrinsic problem of the pellet target, as explained in section 2.2.

However, the situation in $\overline{\mathrm{P}}$ANDA is too complex to be correctly expressed in a single number. Rate fluctuations on the μs scale originating from the statistics of the quasi-continuous interaction are not included in $R_{\mathrm{nom.}}$. These are important to understand the peak rate which has to be covered to avoid data loss. On the other



(a)



(b)

Figure 5.5.: The magnetic field in the xy-plane at z=200 cm (a) with the approximate shape of the EMC support frame (see Fig. 5.4). The corresponding field map in the yz-plane at x=0 (b) contains the outline of the DIRC volume to provide orientation. The z-axis is defined by the beam direction, x is horizontal and y vertical.

hand, the average of the expected rate over a full accelerator cycle

$$R_{\text{cycle}} = \frac{1}{t_{\text{cycle}}} \int_0^{t_{\text{cycle}}} R(t)\mathrm{d}t \lesssim 10\,\text{MHz}$$

is at least a factor of two lower. Hence the usage of $R_{\text{nom.}}$ for lifetime estimations would lead to an overestimate of a factor of two. This has to be kept in mind when comparing the lifetime estimations in this document to other $\overline{\text{P}}$ANDA-DIRC estimations, which are typically based on the nominal rate of 20 MHz instead of R_{cycle}.

The operational parameters for the planned HESR in high luminosity mode can be found in [71], [72]. According to these documents, the number of antiprotons stored in the HESR is limited to $N_{\bar{p},\text{max}} = 10^{11}$ and antiprotons of the previous cycle remain in the storage ring during a refill. The initial luminosity L_0 at the time when the target is turned on is given by

$$L_0 = f_0(p)\, N_0\, \rho_{\text{target}}$$

with the momentum dependent revolution frequency f_0, the target density ρ_{target} and the number of antiprotons N_0 stored in the ring. During the measurement, the luminosity will decrease exponentially due to the loss of antiprotons by the hadronic interaction and beam heating. The luminosity profile can be described using the beam loss rate τ^{-1}:

$$L(t) = L_0 e^{-t/\tau}.$$

Both, the hadronic cross section $\sigma_{\bar{\text{p}}\text{p}}$ as well as beam heating effects decrease significantly with increasing momentum, what has an impact on the initial and average rates. The beam lifetimes, total cross sections $\sigma_{\bar{\text{p}}\text{p}}$, as well as the corresponding initial luminosities and rates $R_0 = L_0 \cdot \sigma_{\bar{\text{p}}\text{p}}(p)$ are summarized in table 5.2. Due to the momentum dependence of the hadronic cross section, the highest R_0 is reached at 1.5 GeV/c. This rate of 19.5 MHz, which does not account for fluctuations in the target density, is almost identical to the nominal rate of 20 MHz. However, the instantaneous rate can be controlled to some extend by steering the beam target overlap or, in case of the cluster jet target, the target density. As the dominant loss contribution also scales with the beam target overlap, such an approach will not significantly reduce the average luminosity, but rather flatten the exponential luminosity profile.

p [GeV/c]	τ^{-1} [10^{-4} s]	f_0 [kHz]	$\sigma_{\bar{p}p}$ mbarn	L_0 [10^{32} cm^{-2} s^{-1}]	R_0 [MHz]
1.5	6.5	443	100	1.95	19.5
9.0	1.7	519	57	2.28	13.0
15.0	1.4	521	51	2.29	11.7

Table 5.2.: HESR parameters according to [71], $\bar{p}p$ cross sections from [20], and resulting luminosities/rates at the start of the measurement cycle.

In consequence to the stronger beam heating at lower momenta, the cycle averaged luminosity

$$L_{\text{cycle}} = \frac{1}{t_{\text{cycle}}} \int_0^{t_{\text{exp}}} L(t)\,\mathrm{d}t = L_0(t_{\text{exp}}) \frac{\tau}{t_{\text{cycle}}} \left(1 - e^{-\frac{t_{\text{exp}}}{\tau}}\right)$$

shows an opposite trend. The cycle time $t_{\text{cycle}} = t_{\text{prep}} + t_{\text{exp}}$ is the sum of the time needed for beam preparation and the actual measurement. The optimum cycle time which results in the maximum L_{cycle} can be found by solving

$$0 = \frac{\partial L_{\text{cycle}}}{\partial t_{\text{exp}}} = -L_0 \frac{e^{-\frac{t_{\text{exp}}}{\tau}} (\tau e^{\frac{t_{\text{exp}}}{\tau}} - \tau - t_{\text{cycle}})}{t_{\text{cycle}}^2}$$

if the maximum number of antiprotons $N_0 = N_{\text{max}}$ can be provided for every t_{prep}. Otherwise, N_0 and L_0 become a function of the experimental time t_{exp} which is available for accumulation. An approximation of the maximum N_0 is given by

$$N_0(t_{\text{exp}}) = \frac{R_{\text{prod}} \cdot t_{\text{exp}}}{1 - e^{-\frac{t_{\text{exp}}}{\tau}}}$$

with the nominal antiproton production rate $R_{\text{prod}} = 20\,\text{MHz}$. Using the preparation times given in [72] and this equations, one obtains the luminosity values shown in table 5.3. It is obvious, that lifetime estimations have to be based on the maximum cycle averaged luminosity or rate and not on $R_{\text{nom.}}$. If a safety factor is used, it should be stated explicitly.

The total runtime of the PANDA experiment is assumed to be 10 years at 50% duty cycle, what translates $T_{\text{life}} = 158 \cdot 10^6\,\text{s}$. Motivated by the presented estimations, based on the work of A. Lehrach et al. presented in [71], [72], a rate

p GeV/c	t_{cycle} s	t_{prep} s	L_{cycle} $\text{cm}^{-2}\text{s}^{-1}$	R_{cycle} MHz	R_{exp} MHz	R_0 MHz
1.5	4890	120	$5.32{\cdot}10^{31}$	5.3	6	19.5
15	2227	290	$1.59{\cdot}10^{32}$	8.1	10.3	11.7

Table 5.3.: Optimum cycle times and the corresponding cycle averaged luminosity L_{cycle} and rate R_{cycle}. In addition, the average rate over t_{exp} is given, including a factor of 1.1 to account for the barrier bucket system. For direct comparison, the initial rate R_0 has been added to the table.

of $R_{\text{life}} = 10\,\text{MHz}$ has been chosen as long term average rate. This value already includes a safety factor of >25% as it assumes an operation at highest luminosity over the full runtime. The specific properties of nuclear targets have not been considered.

To estimate the peak interaction rate, one has to account for statistical fluctuations caused by the continuous beam. The interaction can be considered as a Poisson process with rate R. The distribution of the time between consecutive events of such a process is given by the exponential distribution with the probability density function

$$P_{\text{exp}}(t;R) = Re^{-t\cdot R} \text{ for times } t \geq 0.$$

To find the peak rate, samples have been drawn from this distribution and converted to a train of consecutive timestamps. Then, the development of the rate in time has been computed for a gliding window of constant width δt_w, resulting in a sequence of rates and the corresponding time during which the rate is constant. Afterwards, the rates have been histogrammed while normalizing the contribution to the histogram by their duration.

The result of such a calculation for R = 23 MHz and $\delta t_w = 1\,\mu\text{s}$ is presented in Fig. 5.6. Based on this estimation, a maximum instantaneous rate* of 40 MHz can be specified for a window of 1 μs. This interval has been chosen to match

*In a strict mathematical sense, the term "instantaneous rate" (of change) corresponds to the slope of a variate function $\frac{\mathrm{d}f}{\mathrm{d}t}(t)$ at a particular point in time. However, in case of a sequence of discrete events which is not represented by a continuous function, this definition makes no practical sense without smoothing $f(t)$. As the resulting peak rate depends on the smoothing interval, it has to be mentioned together with the rate to form a usable specification. In this document, the term instantaneous rate is synonymous with the average rate over a sufficiently small interval.

the timescale relevant for front-end electronics which have to digitize the analog detector signals and transport the data.

When the actual rate exceeds the rate capability of the Endcap DIRC, signals will be lost due to the dead time of the electronics. This signal loss will not only affect single tracks, but all tracks in the particular time interval. In consequence, the DIRC will deliver either no PID information or PID information with significantly lower performance. In any way, the functional requirements cannot be met. Therefore it is advantageous to introduce an additional safety factor to account for the possibility that the target density fluctuations of the pellet target exceed the specified factor of two. Assuming a factor of three for these fluctuations and following the same arguments that lead to the 20 MHz definition of R_{nom}, one ends up with a maximum average rate of 30 MHz. The corresponding peak rate for a 1 μs window is ~55 MHz .

Radiation conditions

All components used in nuclear and particle physics experiments have to be tolerant to the irradiation with high fluxes of γ-rays, charged particles and neutrons. In case of the DIRC, special care has to be taken in respect to the optical materials and front-end electronics.

Glasses tend to darken when they are irradiated. Electronic devices, especially semiconductors, can either be destroyed due to the modification of the bulk material or disrupted in operation, e.g. due to the state change of single bits in digital circuits. These effects usually scale with one or more of the following quantities:

- amount of energy deposited by ionizing radiation
- 1 MeV neutron equivalent fluence
- charged particle flux

As all of these parameters are important to judge whether a single component can be used at a specific detector position or not, radiation maps have been generated. The most recent estimations are shown in Fig. 8.20, 8.21, 8.22 on page 221ff. In these figures, the geometry of the final detector design is used as radiation sensor.

The energy deposit in fused silica ranges from about 500 Gy close to the acceptance of the target spectrometer to 10 Gy at the outer rim of the detector. The outer part is shielded by the barrel electromagnetic calorimeter.

The expected 1 MeV equivalent fluence at the rim is in the order of $10^{11}/\mathrm{cm}^2$.

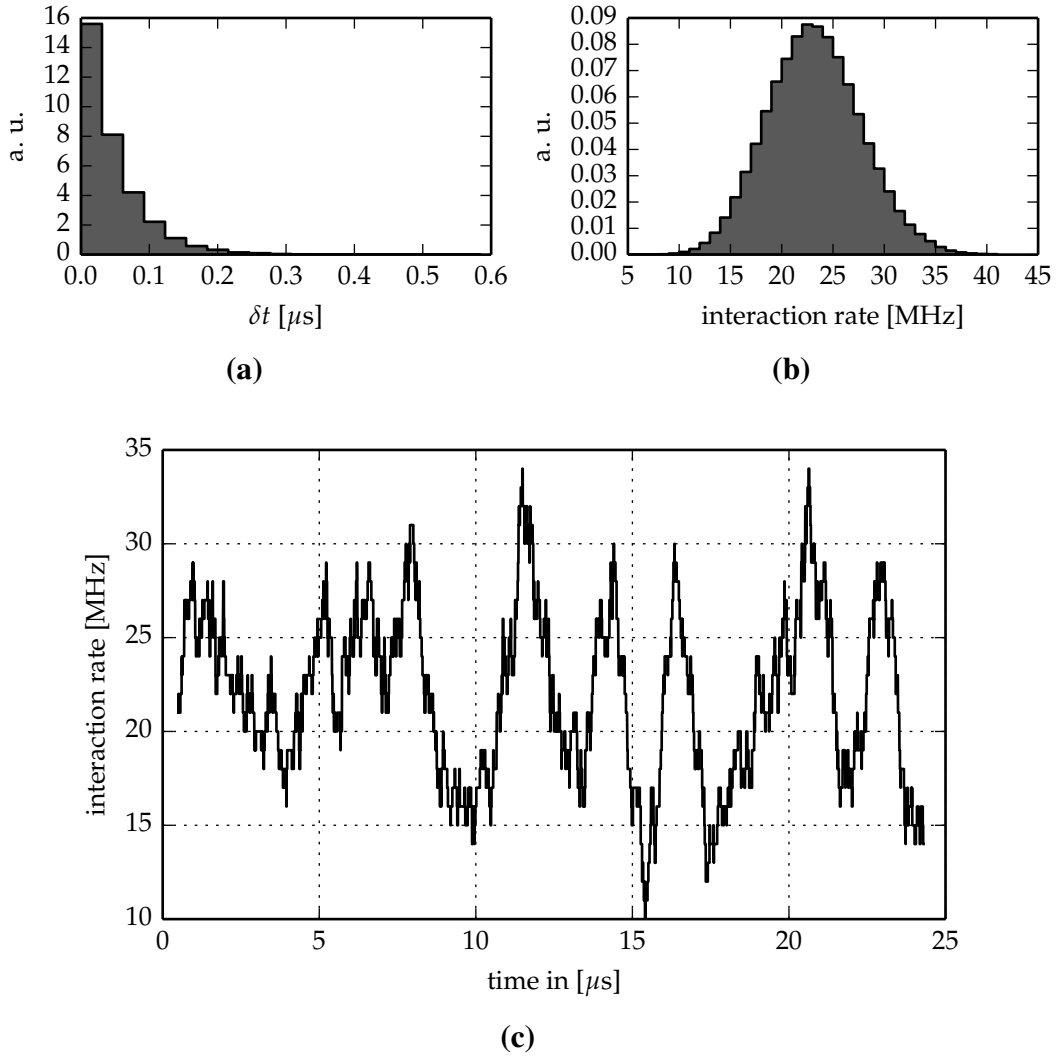


Figure 5.6.: Modeled development of the instantaneous event rate in time (c) resulting from the continuous wave like interaction of the antiproton beam at 15 GeV/c with a hydrogen target of the constant density $2*\langle\rho_{\text{targ}}\rangle = 8 \cdot 10^{15}$ atoms/cm^2. As the interaction is a Poisson process, the time between individual events can be described by an exponential distribution with rate parameter $\lambda = L_{\text{max}} * \sigma_{\text{had.}} \approx 23$ MHz. The time dependence of the luminosity L is neglected. 10^5 samples have been drawn from such a distribution as shown in (a). The instantaneous rate has been determined by averaging over a gliding time window of 1 μs width. The distribution of the different rates is shown in (b). A part of the data is plotted versus time (c). The model predicts a maximum expected rate of < 40 MHz in case of target density fluctuations ≤ 2.

This value can only serve as a lower limit for two reasons. The estimation is based on $\bar{p}p$ neglecting the operation with nuclear targets which could contribute significantly to the neutron fluence. The second reason is that the transport engine used for this estimation is Geant4, which is not as precise in neutron predictions as for example FLUKA*. However, recent comparisons of the Monte Carlo codes show only a difference of a factor of two, which is not utterly important when it comes to the order of magnitude.

The charged hadron flux is another important quantity as it helps to estimate the single event upset probability in electronic device. While the related problems are well under control in airborne applications, the failure rate in particle physics experiments can become so high that typical multiple redundancy approaches become useless. A rate of $\sim$100 Hz/cm^2 has to be assumed for this kind of estimations. This value includes an energy cut at 10 MeV, As the SEU cross section is zero below 10 MeV, an energy cut has been included.

5.1.3. Discussion

A closer view on Figures 5.3, 5.4 and 3.7 reveals that the targeted Endcap DIRC performance exceeds the performance of the BaBar DIRC significantly, while an extreme size reduction of the imaging system is required at the same time. Pin-hole focusing will not be sufficient due to the tight geometric constraints and more sophisticated imaging solutions have to be developed. The number of photosensors and thus the active area is also limited by these constraints. This leads to a higher photon rate per unit area which has to be handled by the sensors as well as the front-end electronics. Likewise, the spatial granularity of the sensors has to scale with the size of the optical system to maintain a compatible resolution. These tasks have to be solved for a system which operates inside a high magnetic field.

Due to these significant differences, the Endcap DIRC can be considered to be a novel technology. While being based on the sound foundation laid by the BaBar groups around Blair Ratcliff, it has to face its very own challenges. Thus, it cannot be assured that such a high performance target as 4 σ separation power up to 4 GeV/c can be met. The maximum possible design performance may turn out to be lower.

*FLUKA has not been used because the $\overline{\text{P}}$ANDA geometry is not ported to the FLUKA format and the media of all detectors have to be taken into account to obtain reliable results.

Class	Requirement	Definition
Function and performance	Charged hadron PID	has to be provided online
	Separation power (π,K)	$> 4\,\sigma$ over full acceptance
	Cherenkov angle resolution	$\leq$1.6 mrad in fused silica
	Momentum coverage	1.5 … 4 GeV/c
	Polar acceptance, min. max.	$\theta_x = 10°$, $\theta_y = 5°$ (elliptical) $\theta_{x,y} = 22°$ (circular)
	Azimuthal acceptance	2π
	Detector lifetime	$\geq$ 10 years in duty cycles of 6 months/year $\hat{=}\ 158 \cdot 10^6$ s
Environment	Allocated space	Volume in front of the endcap EMC as indicated in Fig. 5.4
	Distance to interaction point	~ 194 cm
	Magnetic field	0.5 … 1.3 T with field directions as shown in Fig. 5.5
	Interaction rate at target density $\leq 2 \cdot \rho_{nom}$ at target density $\leq 3 \cdot \rho_{nom}$	$\leq$ 10 MHz long term average $\leq$ 26 MHz averaged over 10 μs window $\leq$ 40 MHz averaged over 1 μs window $\leq$ 37 MHz averaged over 10 μs window $\leq$ 55 MHz averaged over 1 μs window
	Energy deposit in radiator in optics	~500 Gy for SiO_2 ~10 Gy for SiO_2
	1 MeV neutron equ. fluence	$> 10^{11}$ / cm^2 at sensor ($<10^{12}$ / cm$^{2\,(*)}$)
	Charged hadron flux	~100 Hz/cm^2 ($E_{kin} > 10$ MeV) at sensor

Table 5.4.: Summary of the specified basic requirements for the Endcap DIRC. $^{(*)}$: Upper limit defined by assuming a maximum estimation error of one order of magnitude.

A closer look on the rate requirements revealed that most lifetime estimations done by the PANDA DIRC groups are based on the nominal rate of 20 MHz, rather than on the cycle averaged rate of 10 MHz. Hence, these values are a factor of two larger than necessary. As the assumption of a operation at highest rates over the full lifetime, which is commonly used in $\overline{\text{P}}$ANDA, already includes a safety factor, this additional factor of two seems utterly redundant. As laid out in the preceding sections, this work will be based on a long term average rate of 10 MHz, which is consistent with the numbers published in several $\overline{\text{P}}$ANDA documents, e.g.

[98, 109].

While not explicitly mentioned as boundary condition, the total cost of the device is an important limit which has also included in the design process. As a matter of course, prices are not part of this document.

5.2. System decomposition

After defining a basic set of system requirements, one has to determine the components and design options which are compatible with this specification. To minimize the risk of overlooking or skipping possible options, it is helpful to start with a functional decomposition of the system. This is done in two steps:

- divide the high-level functional description of the system into the most general set of essential processes,
- identify the interdependence of these processes.

After the functional decomposition, the physical components necessary for the individual processes are identified. These components may also show an interdependence. The resulting information can then be used to define an abstract system model which can be used to evaluate different design options.

This section aims at providing a universal decomposition of a generic DIRC detector into individual processes and related components. The result is a simple, serial sequence of processes as shown in Fig. 5.7. Every item has an assigned item code consisting of a class prefix (letters) and a number.

The structure in Fig. 5.7 evolves naturally when describing the function of the detector. The fundamental idea which distinguishes the DIRC from traditional RICH detectors is the creation of Cherenkov radiation in a thin radiator (P1), which is then partially trapped and transported by total internal reflection (P2). Angle and position of individual photons have to be determined to obtain the angular information at emission. Therefore the photon has to be imaged* (P3) onto a photodetector and detected (P4). The resulting signals have to be digitized (P5), transported (P6) and processed (P7) to be suitable as input for the particle identification process (P8).

*for the special case of a time-of-propagation approach, the imaging process is just a one-to-one relation with no physical component.

This is a simple chain, where every process depends just on the input of its predecessor. The assignment of the necessary components is also straight forward. A radiator (C1) is needed for photon emission and transport, followed by imaging optics (C2), photodetector (C3) and front-end electronics (C4). An optional data concentrator (C5) may be needed before the data is transmitted through a network (C6) to compute nodes (C7) for data processing. Finally an algorithm for particle identification (C8) is needed to extract the physics information from the signals.

Beside these main functional components (prefix C), auxiliary components (prefix CA) are needed for the functional components. Infrastructure components (prefix CI), as the mechanical support, gas systems for moisture and temperature control and calibration systems are not included in this model.

Table 5.5 shows the processes and related components together with important component parameters, the design constraints which affect the components, and the way external systems could be influenced by the components (interfaces).

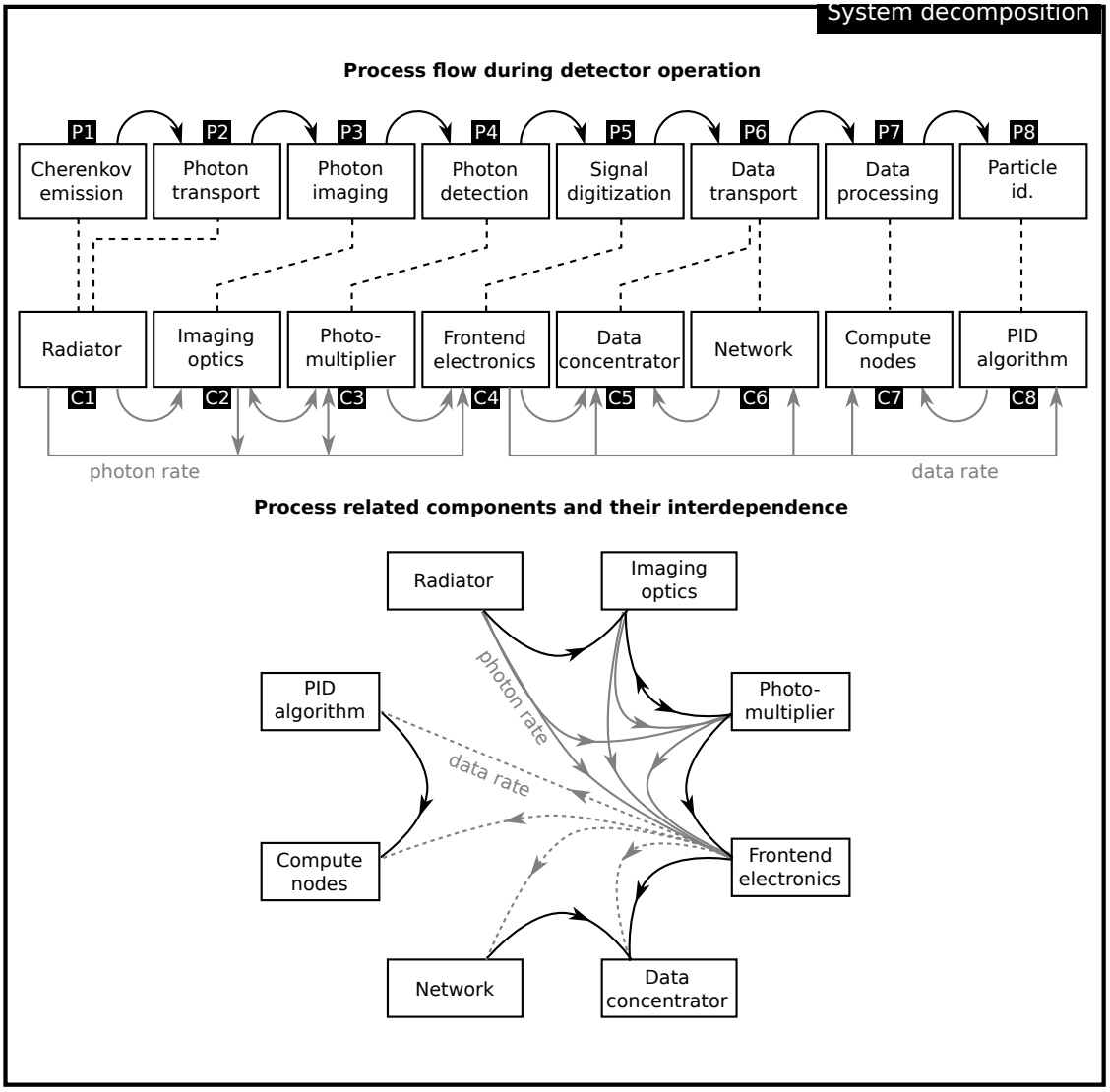


Figure 5.7.: Decomposition of the DIRC detector system in processes and related components. Besides the shown components, which are directly related to the function of the detector, infrastructure components like the mechanical support structure, environmental control system and online calibration system are required.

ID	Process	Components	Important parameters	Constraints	Interfaces
P1	Cherenkov emission	Radiator	Refractive index, thickness	Thickness (showering)	EMC (preshower)
P2	Photon transport	Radiator	Surface figure & roughness, transmission, dispersion, rad. hardness, fluorescence, size and shape	Geometric constraints Radiation conditions	
P3	Photon imaging	Imaging optics	as in P2	Geometric constraints Radiation conditions	
P4	Photon detection	Photomultiplier	Radiation hardness, rate capability, detection efficiency, aging	Magnetic field Geometric constraints Radiation conditions Rate requirements	
P5	Signal digitization	Front-end electronics	Radiation hardness, SEU cross-sections, power consumption, rate capability, efficiency, time and charge resolution, geometric size, channel density	Geometric constraints Radiation conditions Rate requirements	EMC (heat dissipation)
P6	Data transport	Data concentrator and network	Radiation hardness, SEU cross-sections, power consumption, bandwidth	Geometric constraints Radiation conditions Rate requirements	EMC (heat dissipation) DAQ/T infrastructure
P7	Data processing	Compute nodes (CPU/FPGA)	Bandwidth and throughput	Rate requirements	Tracking system DAQ/T infrastructure
P8	Particle identification	Algorithm / software	Throughput, efficiency	Performance requirements Rate requirements	Tracking system Global PID algorithm DAQ/T infrastructure

Table 5.5.: Decomposition of the DIRC function into processes and components.

5.3. An abstract model for DIRC detectors

Before the discussion of individual components it is helpful to derive an abstract model which describes a DIRC detector in a more general way. Terms which are defined in this section will be consistently used throughout the document. This model will also serve as basis for the reconstruction of Cherenkov angles.

A key feature of DIRC detectors is a thin radiator in terms of geometry and radiation length. In the following, it is assumed that the side faces of this radiator are aligned orthogonal to the top and bottom faces, which are perfectly parallel to each other. Top and bottom faces denote those sides of the radiator which are traversed by the incident charged particles. The *radiator plane* is defined as a parallel plane in the middle between the top and bottom surface (Fig. 5.8). Further it is assumed that the radiator has the shape of an arbitrary convex polygon. The edges of this polygon are named *radiator segments*. Optical elements with a width w can be attached at any position on such a radiator segment. These optical elements can be of any kind, ranging from simple expansion volumes to complex focusing optics (Fig. 5.9).

Under given assumptions, the photon transport inside the radiator can be elegantly described by means of 2D projections on the radiator plane. At emission, the photon and the radiator plane enclose an angle φ. The author has shown in [79] that φ can be computed using the equation

$$\cos\varphi = \frac{s_{2_d}}{s} = \frac{A\cos\theta_c}{B} \pm \sqrt{\frac{\cos^2\theta_p - \cos^2\theta_c}{B} + \left(\frac{A\cos\theta_c}{B}\right)^2} \tag{5.3}$$

with the terms $A = \sin\theta_p \cos(\phi_{\text{rel}})$ and $B = A^2 + \cos^2\theta_p$ and

- the particle angle relative to the normal of the radiator plane θ_p
- the Cherenkov angle θ_c
- the angle ϕ_{rel} enclosed by the photon and particle trajectory on the radiator plane

The reflection angle φ is hence fully determined by the particle trajectory, its velocity ($\cos\theta_c = 1/n\beta$) and the photon emission angle ϕ_{rel} relative to the particle trajectory in the radiator plane projection (RPP). In conclusion, if the particle trajectory and ϕ_{rel} are known, the Cherenkov angle becomes a function of φ:

$$\theta_c(\varphi, \theta_p, \phi_{\text{rel}}) = \arccos(\sin\theta_p \cos(\phi_{\text{rel}}) \cos\varphi + \cos\theta_p \sin\varphi) \tag{5.4}$$

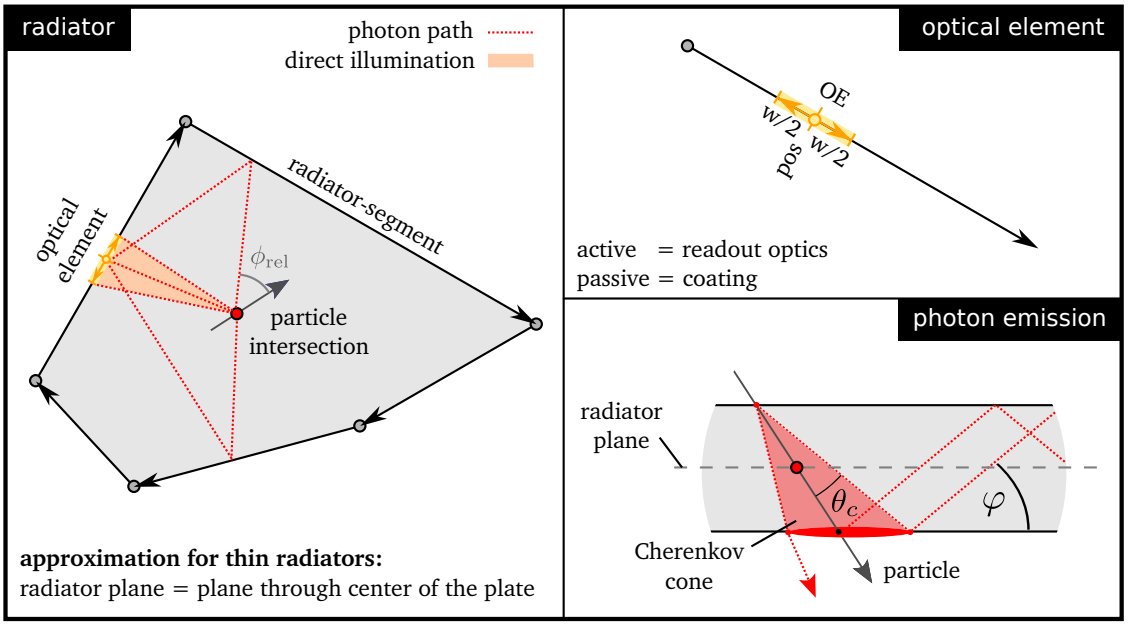


Figure 5.8.: Illustration of terms defined in the abstract DIRC model.

Due to the initial assumptions, φ is preserved during transport even if the zig-zag path of the photon is reflected at a radiator segment. It is hence sufficient to describe the photon path in the RPP.

In the following discussion it is assumed that the particle trajectory is known from the tracking detector. It is helpful to begin with the simplest path, a photon which travels directly to an optical element without being reflected at a radiator segment (*direct photon*). The corresponding path in RPP is a straight line from the emission point, given by the intersection of the particle trajectory with the radiator plane, to the optical element. The angle ϕ_{rel} can be directly obtained from the known location of the optical element. If the optical element is of an imaging type measuring φ, the Cherenkov angle follows directly from eq. 5.4.

However, the hit detected in the optical element is not necessarily a direct photon. It could also originate from a photon which has been reflected at a radiator segment (compare Fig. 5.8). This raises two questions. How to determine the possible reflected paths and how to resolve the ambiguities. The first question will be addressed in more detail later in section 5.9.

A common way to remove ambiguities is to exploit the time information. The length $s_{2\mathrm{d}}$ of the photon path in RPP is related to the real 3d photon track length

s_{3d} by:

$$s_{3d} = \frac{s_{2d}}{\cos\varphi} \tag{5.5}$$

The time of propagation $t_{prop} = t_{det} - t_0$ of a photon emitted at time t_0 and detected at time t_{det} can be computed by means of the group velocity v_g in the radiator material*:

$$t_{prop} = \frac{s_{3d}}{v_g} = \frac{s_{2d}}{v_g \cdot \cos\varphi} \tag{5.6}$$

However, this approach requires the knowledge of t_0 which corresponds roughly to the time when the particle arrives at the radiator plane. Another option to remove ambiguities is by a cut on the expected Cherenkov angle spectrum. The typical single photon resolution of a DIRC counter is better than 15 mrad. Ambiguities often result in computed Cherenkov angles which are far-off the expectation for any particle hypothesis.

So far, this model – including the path reconstruction algorithm discussed in section 5.9 – can be used to describe the photon parameters from emission up to the optical element in any DIRC detector using a thin radiator with polygonal outline. It can be used directly to compute Cherenkov angles in a proximity focusing setup as shown in (Fig. 5.9). In this case, the RPP is valid inside the expansion volume and the imaging plane can be described as radiator-segment. In such a setup, the Cherenkov angles can be directly computed from path information.

In case of focusing optics however, the RPP approach is invalid either due to the nonlinear refraction at a lens or reflections at a focusing mirror which is not orthogonal to the radiator plane. The different optics cannot be described in an abstract manner. In the following, the model will be extended by a description of cylindrical focusing mirrors. These mirrors map the angle φ' to a coordinate $z(\varphi')$ on the focal plane. The function

$$z(\varphi') = z(\varphi, \alpha_{FEL}) = z\left(\arctan\left(\frac{\tan\varphi}{\cos\alpha_{FEL}}\right)\right) \tag{5.7}$$

will be referred to as *imaging function* of the optical element. The connection between φ' and the angle α_{FEL} of the photon path to the optical element in RPP is

*Note that eq. 5.6 can also be used to compute the angle φ and hence the Cherenkov angle from the photons time of propagation.

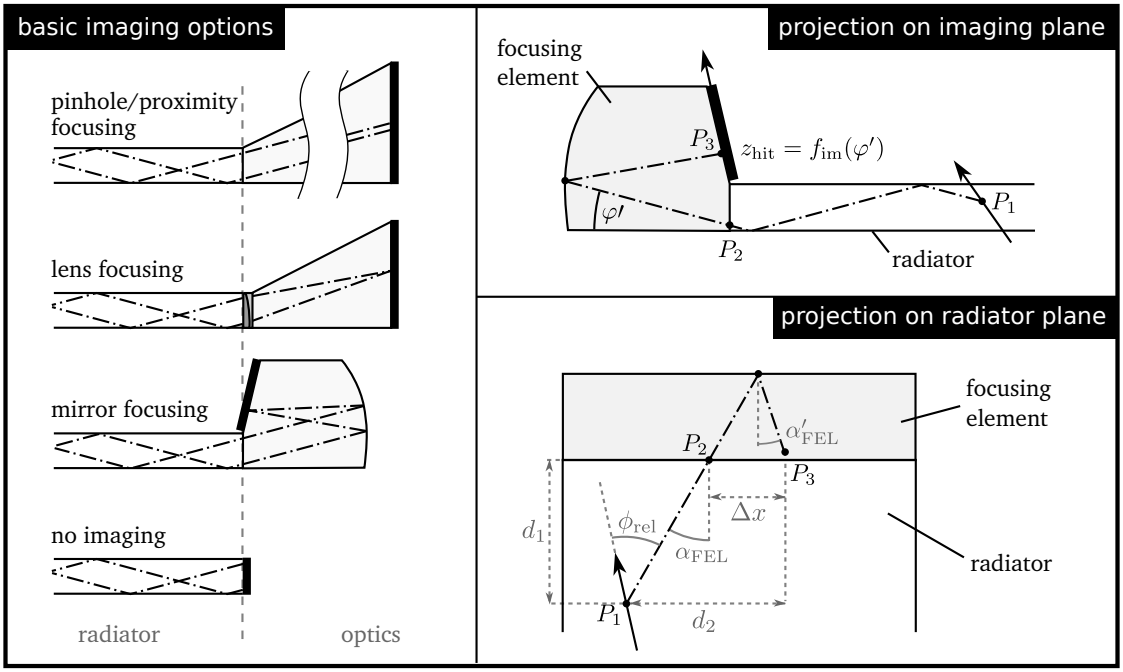


Figure 5.9.: Illustration of different optic options for the DIRC readout (left) and terms defined by the DIRC model (right).

given by

$$\tan\varphi' = \frac{\tan\varphi}{\cos\alpha_{\mathrm{FEL}}} \tag{5.8}$$

It is obvious that α_{FEL} has to be known to compute φ and thus the Cherenkov angle θ_c from eq. 5.4. One option is to assume that the photon entered the optical element at its center. α_{FEL} is then defined by the geometry and tracking information. The width of the optical element defines the error of α_{FEL}. Fig. 5.10 shows the dependence of the focus position on α_{FEL} for different φ angles. Each curve results from the incident angles $-45° \leq \alpha_{\mathrm{FEL}} \leq 45°$. In case of small focusing elements with reflecting side surfaces, the pattern is folded onto the smaller focal plane (dashed lines).

If α_{FEL} is determined by the position of the focusing element, the uncertainty depends on the distance of the emission point to the optical element. The opening angle of the direct light cone shown in Fig. 5.8 (top left frame) will become larger when the emission point gets closer to the optical element. Hence, it can be advantageous to reconstruct α_{FEL} from the pattern in a wide focusing element

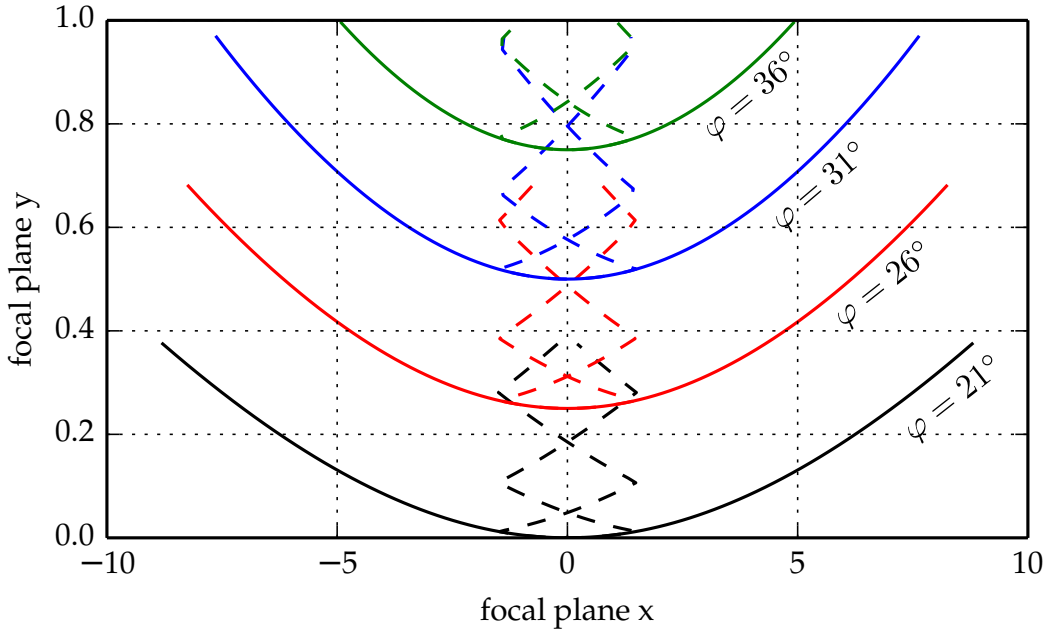


Figure 5.10.: Variation of the focus position in a focusing element with cylindrical mirror for different φ angles in dependence of α_{FEL} ($-45° \leq \alpha_{\mathrm{FEL}} \leq 45°$). Solid lines show the dependence for a focusing element of a width $w > 10$ and the dashed lines correspond to a focusing element of width $w = 3$ where the light is reflected at the side surfaces. The imaging function $z(\varphi')$ is linear, mapping $\varphi' = 21° \ldots 41°$ to $y = 0 \ldots 1$. The parameter $< L_{\mathrm{FEL}} >$ equals 10.

where the folding of the pattern can be neglected. This is achieved by the formula

$$\tan \alpha_{\mathrm{FEL}} = \frac{d_2}{< L_{\mathrm{FEL}} > \cos \varphi' + d_1} \tag{5.9}$$

with d_1 and d_2 as given in Fig. 5.9 and the average photon path length $< L_{\mathrm{FEL}} >$ in the imaging plane projection (IPP). This projection is equal to the view plane in Fig. 5.9, top right. $< L_{\mathrm{FEL}} >$ is the average distance which light travels from the entrance of the optics to the focal plane, e.g. from P_2 to P_3 in the given figure*. In

*Equation 5.9 can be derived by treating the light propagation in the optical element in the same manner as the photon propagation inside the radiator. The two angles φ' and χ in IPP determine the position $z(\varphi')$ on the focal plane and the displacement Δx along the axis perpendicular to the IPP-plane ($\Delta x =< L_{\mathrm{FEL}} > \tan \chi$). This displacement defines P_2 in respect to P_3. The condition that the resulting angle α_{FEL} has to be compatible with the angles φ' and χ results directly in eq. 5.9.

general, $< L_{\mathrm{FEL}} >$ will also be a function of φ'.

The presented model has been implemented in C++, resulting in a flexible, geometry-independent DIRC model which serves as basis for the implementation of reconstruction algorithms. Hence, the algorithms can work with any radiator shape and instrumentation, as long as a specific optics model is available.

5.4. Optical components (C1,C2,CA1)

General design parameters and properties which influence the design choice for the optical components (C1, C2) are discussed in this section. Important physical effects will be explained in place when necessary.

5.4.1. Radiator (C1)

has impact on	*key properties*
• Cherenkov production	• Material
• Photon transport	• Shape
• Downstream detectors	• Thickness d_R
• Imaging optics.	• Surface figure

The radiator of a DIRC detector has two important roles. It serves as the medium where the Cherenkov emission takes place and at the same time as a light guide to transport the resulting photons while preserving their angular information and hence the physics signal.

A material with high optical transmission is mandatory to avoid signal loss as the light path inside the radiator can easily exceed 1 m while the number of emitted Cherenkov photons is only in the order of a few hundred. To maintain this high transmission during the projected detector life-time the radiation induced absorption has to be negligible in the wavelength range relevant to the application.

Not only signal loss, but also the distortion of the signal has to be minimized. The relevant angular information can be distorted at Cherenkov emission and during transport. At emission, the major sources of uncertainty are given by the chromatic dispersion of the material and by small perturbations of the particle track direction introduced by multiple coulomb scattering. During transport, errors are caused by the non-perfect surface figure of the radiator and inhomogeneities in the bulk material.

These effects depend primarily on the material selection and the specification of the surface figure. In addition, the thickness of the radiator d_R has some influence on the amount of multiple coulomb scattering and the number of reflections during transport. d_R also affects the design and performance of the imaging optics and the influence of the DIRC device on the downstream detectors.

The following three subsections will discuss photon transport effects, the choice of the radiator material and the optimum radiator thickness.

Photon transport

This section summarizes the most important effects relevant to photon transport in optical materials and introduces the related equations which were used for detector modeling.

Trapping fraction: The fraction ϵ_{trap} of Cherenkov photons which are trapped inside a DIRC radiator by total internal reflection is denoted "trapping fraction". This number depends on the radiator material, the track velocity βc and the angle of the particle track in respect to the surface normal of the radiator (*dip angle*). In case of the Endcap DIRC, the radiator normal is parallel to the beam axis and the dip angle equals the polar angle θ_p of the track. According to eq. 3.7, the absolute number of photons N_{ph} emitted by a particle is proportional to its track-length L inside the radiator. For a given radiator thickness d_R, one obtains:

$$N_{\text{ph}} = \frac{L}{d_R} \cdot N_0 = \frac{N_0}{\cos\theta_p}.$$

The number of photons trapped inside the radiator is hence given by the product of the trapping fraction and the number of created photons

$$N_{\text{trap}} = \epsilon_{\text{trap}} \cdot N_{\text{ph}}.$$

Fig. 5.11 shows the absolute number of trapped Cherenkov photons in a fused silica radiator normalized to N_0:

$$\frac{N_{\text{trap}}}{N_0} = \epsilon_{\text{trap}} \cdot \frac{L}{d_R}$$

for several particle species, momenta and dip angles relevant to the Endcap DIRC. The absolute number of trapped photons is fairly constants at polar angles exceeding 10°. The only exception is given by slow protons at momenta below 2.5 GeV/c

due to the small Cherenkov angle. One can therefore expect a rather constant number of photons per track over the full acceptance, owing to the restricted range of dip angles.

Chromatic dispersion: The refractive index n of an optical material varies with the wavelength of the light. A popular parametrization is the *Sellmeier equation*

$$n(\lambda) = \sqrt{1 + \frac{B_1\lambda^2}{\lambda^2 - C_1} + \frac{B_2\lambda^2}{\lambda^2 - C_2} + \frac{B_3\lambda^2}{\lambda^2 - C_3}} \tag{5.10}$$

with the material specific constants B_i, C_i and the vacuum wavelength λ. These constants are commonly provided by manufacturers who obtain them by a fit to measured data. The Sellmeier equation can be derived directly from a Lorentz oscillator model for the dielectric constant. Hence it has a strong physical basis and can even be used to extrapolate the refractive index from measurements [13]. This is not necessarily true for other empirical formulas.

A useful measure for the strength of the chromatic dispersion is given by the *Abbe number* v_d which is defined as:

$$v_d = \frac{n_d - 1}{n_f - n_c} \tag{5.11}$$

with the refractive indices $n_{d,f,c}$ at the wavelength of the Fraunhofer D-line (589.2 nm), F-line (486.1 nm) and C-line (656.3 nm). A higher Abbe number corresponds to lower dispersion and vice versa.

The dispersion affects not only refraction, but also two key observables of the Endcap DIRC. These are the Cherenkov angle $\theta_c = \arccos\left((n(\lambda)\beta)^{-1}\right)$ and the *group velocity* v_g of the light inside the medium:

$$v_g = \frac{c}{n(\lambda) - \lambda\frac{\partial n(\lambda)}{\partial\lambda}} \tag{5.12}$$

which is experimentally accessed by measuring the time of propagation t_p needed by the photon to travel the path $s_{3\mathrm{d}}$ (eq. 5.6):

$$t_{\mathrm{prop}} = \frac{s_{3\mathrm{d}}}{v_g} = \frac{s_{3\mathrm{d}}}{c}\left(n - \lambda\frac{\partial n}{\partial\lambda}\right)$$

While the error of the Cherenkov angle can be partly corrected by using an apochromatic imaging system, the error of the propagation time cannot be corrected

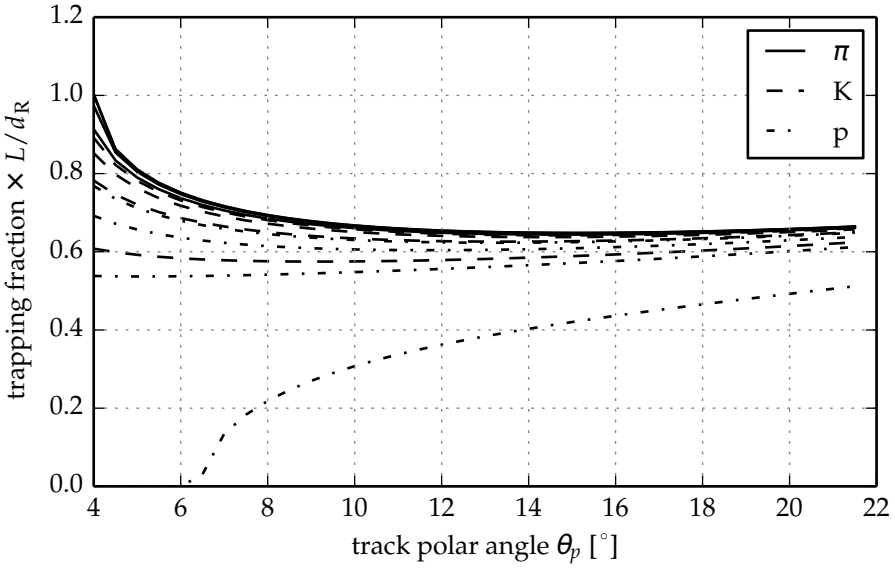


Figure 5.11.: Trapping fractions for photons emitted by particle tracks traversing a radiator of thickness d_R at an angle of θ_p relative to the surface normal. The trapping fractions are multiplied with the angle dependent increase of pathlength L/d_R of the track inside the radiator. The resulting value can be multiplied with the number of photons N_0 for a track at normal incidence to obtain the absolute number of Cherenkov photons which are trapped and transported inside the radiator. The four curves for each particle correspond to the trapping fraction at 1.5, 2.5, 3.5 and 4.5 GeV/c (trapping fraction increases with higher momentum).

without a measurement of the wavelength. However, due to the correlation of both errors, a precise measurement of one quantity can be used to correct for the chromatic error of the other. Dispersion correction techniques are discussed in section 5.4.4.

Photon losses: The term "photon loss" will be used throughout this document to describe the loss of information carried by signal (Cherenkov) photons. This information is lost if the photon is either absorbed or its angular information has been considerably distorted. Photon loss in the radiator can be categorized in two groups. Losses due to imperfection of the interface between different media (*surface losses*) and losses due to absorption and scattering inside the medium (*bulk losses*).

Bulk losses: The term "absorption of light" summarizes all processes where the energy of the photon is transferred to the material, e.g. by exciting lattice vibrations or electronic transitions in the band structure and in additional energy states introduced by lattice defects. In an homogeneous material, the probability for the absorption of a photon at a given wavelength is constant. Hence, the intensity follows the well known Beer–Lambert law:

$$I = I_0 e^{-x/\Lambda(\lambda)} \tag{5.13}$$

with the attenuation length Λ which is a function of the photon wavelength λ. This attenuation length is a material property and constant as long as the material remains unchanged. Ionizing radiation however can introduce additional lattice-defects, so called color centers or F-centers, which also contribute to the attenuation length (Λ_{rad}). In this case, one talks about *radiation induced absorbance*. Λ_{rad} is a non-trivial function depending on the absorbed ionizing dose, and – if the decay time of the color centers is considerably short – even on the dose rate and time.

$$I = I_0 e^{-x/(\Lambda+\Lambda_{\mathrm{rad}})}$$

More details on the radiation damage in glasses is given in section 5.4.5.

Photons are not only lost by absorption but also by scattering out of the original geometric path. Scattering processes in the medium can be either elastic or inelastic. In practice the inelastic processes, such as Raman and Brillouin scattering, are very weak phenomena which do not contribute significantly to the total scattering cross section. Elastic scattering of optical photons at spherical particles with diameter d can be described by solving the Maxwell equations. Two popular models are the Mie solution, an exact solution valid in case of $\lambda \geq d$, and *Rayleigh scattering*, an approximate solution valid in case $\lambda \ll d$. The Rayleigh scattering cross section is proportional to λ^{-4}:

$$\sigma_{\mathrm{Rayleigh}} = \frac{2\pi^5}{3}\frac{d^6}{\lambda^4}\left(\frac{n^2-1}{n^2+2}\right)^2 \tag{5.14}$$

while the wavelength dependence of Mie scattering is significantly lower. The Mie scattering cross section is expressed as infinite series and will not be discussed due to its complicated structure. The interested reader can find detailed information in [19].

In optical glasses, elastic scattering can originate from either particle-like impurities like bubbles or from density (and thus refractive index) fluctuations on

a sub-micrometer scale* [19]. The typical bubble size is larger than the photon wavelength so the Mie solution is valid in this case. The density fluctuations are on a scale smaller than the photon wavelength and the resulting scattering is therefore best described using the Rayleigh approximation. The related attenuation coefficient per unit length for a single component glass is given in [103]:

$$\alpha_{\text{rayleigh}} = \frac{8\pi^3}{3}\frac{n^8}{\lambda^4}p^2\beta_T(T_F)k_B T_F \tag{5.15}$$

with refractive index n, photon wavelength λ, photoelastic constant p, fictive temperature T_F, isothermal compressibility $\beta_T(T_F)$ and the Boltzmann constant k_B. The fictive temperature T_F, at which the density fluctuations were frozen in, can be lowered by annealing the glass at higher temperatures. This way, the scattering cross section can be slightly reduced.

The Rayleigh contribution to photon loss is generally low, but becomes important for applications with long light paths. In high purity, bubble-free synthetic fused silica Rayleigh scattering becomes the dominant source for bulk losses. The above equation has therefore been used to predict the Rayleigh cross sections for Monte Carlo simulations. The material data has been taken from [117].

Surface losses: Losses at realistic optical surfaces are primarily caused by reflections at the interface between different media and the scattering of light due to the roughness of the surface. The reflection loss is a simple consequence of Electrodynamics and well described by the Fresnel equations. To simplify the mathematical handling of complex refractive indices n_i one can define "effective" refractive indices $\eta_{s,p}$ for s/p polarization

$$\eta_{i,s} = n_i\cos\theta_i \qquad\qquad \eta_{i,p} = n_i/\cos\theta_i \tag{5.16}$$

where θ_i denotes the angle to the surface normal in the adjacent medium $i \in \{1,2\}$. The amplitude reflection coefficients can then be expressed in terms of $\eta_{i,s/p}$:

$$r_s = \frac{\eta_{0,s}-\eta_{1,s}}{\eta_{0,s}+\eta_{1,s}} \qquad\qquad r_p = \frac{\eta_{1,p}-\eta_{0,p}}{\eta_{1,p}+\eta_{0,p}}. \tag{5.17}$$

which are related to the intensity reflection coefficients by

$$R_s = r_s r_s^* \qquad\qquad R_p = r_p r_p^*. \tag{5.18}$$

*These inhomogeneities are frozen in during the rapid cooling in the glass manufacturing process.

The transmission for non-absorbing materials can be computed by $T_{s,p} = 1 - R_{s,p}$. For unpolarized light, the s and p contributions are equal and the reflection coefficient can be averaged:

$$R = \frac{1}{2}\left(R_s + R_p\right). \tag{5.19}$$

The Fresnel losses discussed above only account for reflections due to different refractive indices and assume a perfectly smooth surface. Real surfaces however are not perfectly smooth but have a certain degree of surface roughness which leads to additional scattering of the incident light in directions which differ from the path predicted by geometrical optics. A common measure for the microscopic surface profile is the *RMS surface roughness*

$$R_\mathrm{q} = \sqrt{\frac{1}{x_2 - x_1}\int_{x_1}^{x_2} z(x)\mathrm{d}x} \tag{5.20}$$

where the $z(x)$ denotes the surface height at a position x. A measurement of R_q by profilometric methods cannot reproduce this integral as the band width is limited by the systems spatial resolution, e.g. the sampling length Δx of the profilometer. A typical measurement might involve the measurement of surface heights z_i at points x_i with the sampling length $\Delta x = x_{i+1} - x_i$, resulting in an RMS roughness of

$$R_\mathrm{q} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} z_i^2} \tag{5.21}$$

Obviously, the result of the measurement is strongly biased by the sampling length.

The scattering characteristics of a surface can be expressed using the bidirectional reflectance distribution

$$\mathrm{BRDF}(\theta_i, \phi_i, \theta_s, \phi_s) = \frac{\mathrm{d}L_s(\theta_s, \phi_s)}{\mathrm{d}E_i(\theta_i, \phi_i)}, \tag{5.22}$$

which is defined as the ratio of the incident and scattered flux for different directions. The "practical" computation of the BRDF for a given surface profile is complicated and hence calls for the use of approximations. Common mathematical approximations to the scattering at imperfect surfaces are introduced in [14], chapter 8. An approximation for the scattering loss can be obtained by integrating

the obtained BRDF outside the specular lobe. For smooth surfaces, one obtains the *total integrated scatter* *

$$\mathrm{TIS} = \left(4\pi \cos\theta_i \, R_q n/\lambda\right)^2 \tag{5.23}$$

with the angle of incidence θ_i, the refractive index n, the wavelength of the incident light λ and the RMS surface roughness R_q. As the TIS is the integral over all the light scattered outside the specular lobe, it also defines the photon loss at total internal reflection. The reflection coefficient is then simply given by

$$R_{\mathrm{TIR}} = 1 - \left(4\pi \cos\theta_i \, R_q n/\lambda\right)^2 \tag{5.24}$$

and the intensity after N_{ref} reflections can be computed by

$$I = I_0 \cdot (R_{\mathrm{TIR}})^{N_{\mathrm{ref}}} \tag{5.25}$$

Faraday effect: Cherenkov radiation is linear polarized with the polarization vector perpendicular to the Cherenkov cone. Many fundamental optical processes depend on the light polarization. This applies also to the transmission of dielectric multilayer filters, which will be introduced in section 5.4.4. Hence it might be of interest in respect to filter design and reconstruction methods whether this polarization can be predicted reliably. Due to a magneto-optical phenomenon called the Faraday effect [26], the polarization of the light traversing a dielectric medium in strong magnetic fields is rotated by an angle

$$\beta = V \cdot d \cdot B \qquad \text{with} \qquad V(\lambda) = \frac{e}{2m_e c} \gamma \lambda \frac{\mathrm{d}n}{\mathrm{d}\lambda} \tag{5.26}$$

with the magnetic field component in direction of the light B, the length of the light path d and the Verdet constant V. In fused silica, the Faraday rotation is 70.5 mrad per cm path in a 1 T field (table 5.6). If the polarization of the light has to be reliably predicted somewhere, this has to be taken into account. However, this is not a trivial task because the effect depends also on the alignment of the light path in respect to the magnetic field. Currently, the Monte Carlo tools Geant3/4 do not account for this effect.

* The expression in [14] does not contain the refractive index as it assumes $n = 1$. The refractive index simply has to be added to the definition of the wave number k.

Choice of material

The DIRC principle requires a solid state radiator with high transmission in the VIS-UV region. Hence, the eligible material options are optical glasses and transparent plastics*. Most of the glasses and plastic materials do not meet the radiation requirements described in section 5.1.2 as the bulk absorption of the light increases strongly with the absorbed ionizing dose.

Only a few optical materials are known for their intrinsically high tolerance to ionizing radiation. These are synthetic fused silica, sapphire, and very pure alkali halide crystals like CaF_2 and LiF. From this list, fused silica is the only viable radiator material because crystals are too expensive and cannot be produced in the required geometry.

Fused silica radiators have been successfully employed in the BaBar DIRC, hence the technology is well understood. The material has a very high transmission in the VIS-UV region as well as a high Abbe number. Further, it is suitable for precision polishing and the raw material is available at large sizes exceeding 1 m diameter. Therefore it is the ideal candidate for the Endcap DIRC radiator.

Since BaBar times, the high demands of modern lithography lead to further improvements in the quality of synthetic fused silica. Hence, there is no doubt about fused silica being the primary material choice.

However, the choice of precision polished fused silica plates on the 1 m scale lead to high production costs. At this point, one can rise the question about alternatives. A valid alternative would have to provide a significant cost advantage or offer improved system performance at a compatible cost level.

Since bulk transmission and homogeneity of higher grade fused silica are outstanding, the detector performance could only be further improved by using a material with lower dispersion and hence a higher Abbe number. Figure 5.25 on page 108 shows the Abbe diagram containing all standard glasses from the glass catalog of Schott AG. One can see that there are only a few glasses with lower dispersion than fused silica. Measurements by Wirtenson and White (1997) [146, 145] show that these materials do not qualify in terms of radiation hardness. Schott also offers Ce-doped glasses which are more resistant to ionizing radiation, but the Ce content significantly reduces the transmission at shorter wavelengths ([151], section 2.2.11). In conclusion, there is no fused silica competition among the optical glasses.

*Obviously, the use of crystal material is beyond question due to the large size of the radiator.

When it comes to lowering production costs, the selection of a different glass will not have a significant impact since the total expense is driven in a large part by processing costs and not by the price of the raw material. Precision polishing of surfaces works mostly analog for different glasses and leads to comparable costs. Low cost optics are usually manufactured using optical polymers like poly-(methyl-methacrylate) (PMMA), cyclic olefin copolymer (COC), polystyrene (PS) and polycarbonate (PC). The transmission of PS and PC near the UV cutoff is lower than the PMMA transmission. More importantly, PS and PC show stronger chromatic dispersion and birefringence than PMMA [16]. The optical properties of COC are very similar to those of PMMA. COC has a slightly higher dispersion and refractive index, but lower birefringence than PMMA. Therefore PMMA and COC would be the preferred plastic materials for a DIRC radiator.

First of all, the material has to meet the requirements in terms of radiation hardness. Studies on the irradiation of PMMA based plastic optical fibers (POF) presented in [132] show that the material is only moderately radiation tolerant. A gamma ray induced attenuation of 0.5 dB/m has been measured after irradiation with a γ–dose of 600 Gy. An attenuation less than 1 dB/m can be estimated from the data by extrapolating the curve to a wavelength of 350 nm. This corresponds to a loss of about 11 to 21 percent per meter light path. Due to the lower density and radiation length, the expected accumulated dose in PMMA is lower than the maximum of 500 Gy which has been specified for fused silica. Also, the radiation load is not uniformly distributed over the radiator as shown in figure 8.20. Hence, the effective radiation induced loss will be lower as it has to be averaged along the photon path. In conclusion, the argument of radiation hardness does not exclude the use of PMMA as radiator. However, it might be necessary to exchange the radiator once during operation to limit the photon loss to $\lesssim 10\,\%$. Irradiation data of COC has not been found in literature, but the coloring of COC induced by gamma irradiation in medical sterilization devices is well known [133].

Depending on the total expense for material and processing, the use of PMMA radiators might be economical even if they have to be replaced once during the operation of the system. Another interesting feature of PMMA is the long radiation length of X_0 =34.07 cm which is a factor of 2.77 longer than the X_0 of fused silica [18]. Hence, the use of PMMA would also minimize the impact on the EMC performance.

A major disadvantage of polymeric optics in comparison to glass optical systems is the lower optical quality. Here, the most crucial point is the control of the surface

figure. Polymeric optics are usually produced by casting or injection/compression molding. Especially for large parts, the non-linear shrinkage of the material causes variations in the surface figure [13]. Unfortunately, precision polishing of optical polymers is complicated due to their softness. This makes it impractical to correct the unsatisfactory surface figure of the blanks after casting. This processing related limitation would have to be solved to produce a utilizable DIRC radiator. To the authors knowledge, there is no method available to precision polish PMMA blanks of 1 m diameter. Due to the low number of parts, the cost related to the development of such a technology would thwart the later cost saving during production. The same arguments apply to COC.

These arguments lead to the conclusion that fused silica is the only material of choice. The use of PMMA is not excluded by requirements on radiation hardness but by the lack of technology to produce PMMA sheets with the precise surface figure which is needed to meet the performance requirements.

Radiator thickness and implications

In $\overline{\mathrm{P}}$ANDA, the perturbation of $\mathrm{e}^{\pm}$ and γ tracks through the radiator has to be minimized to prevent a significant degradation of the EMC performance. The perturbation in the high energy limit is mainly caused by energy loss of the $\mathrm{e}^{\pm}$ via Bremsstrahlung

$$\frac{\Delta E}{E_0} = 1. - e^{-\frac{d}{X_0}} \tag{5.27}$$

and the conversion of high energy photons to $\mathrm{e}^{+}\mathrm{e}^{-}$ pairs, which probability can be parametrized by

$$P_{\gamma \to e^+ e^-} = 1. - e^{-\frac{7}{9}\frac{d}{X_0}} \tag{5.28}$$

what is a good approximation for photon energies down to 1 GeV. The amount of distortion scales with the traversed material thickness in radiation lengths d/X_0. The dependence of the relative $\mathrm{e}^{\pm}$ energy loss and the γ conversion probability on the radiator thickness is shown in Fig. 5.12. The probability for γ conversion varies by a factor of ~3 from ~6 % at 1 cm to ~18 % at 3 cm radiator thickness.

Another important effect which scales with d_R is the multiple coulomb scattering (MCS) of traversing particles from the nuclei in the material. The repeated

Quantity	Value	Unit	Source
Mechanical properties			
Density ρ_{SiO_2}	2.2	g/cm^3	Heraeus
Young's modulus E_{SiO_2}	$7.25 \cdot 10^4$	N/mm^2	Heraeus
Modulus of rigidity G_{SiO_2}	$3.0 \cdot 10^4$	N/mm^2	Heraeus
Poisson's ratio ν_{SiO_2}	0.17		Heraeus
Mean expansion coeff. α_L (-50 ... 0 °C)	$2.7 \cdot 10^{-7}$	K^{-1}	Heraeus
Mean expansion coeff. α_L (0 ... 100 °C)	$5.1 \cdot 10^{-7}$	K^{-1}	Heraeus
Particle interaction			
Z/A	0.4993		PDG
Mean excitation energy I_{SiO_2}	139.2	eV	PDG
Radiation length X_0	12.29	cm	PDG
Minimum ionization	3.737	MeV/cm	PDG
Nuclear collision length	29.64	cm	PDG
Nuclear interaction length	44.47	cm	PDG
Pion collision length	41.77	cm	PDG
Pion interaction length	58.56	cm	PDG
Optical properties			
Refractive index $n_e^{(546.23\,\mathrm{nm})}$	1.46004		Schott
Abbe number $\nu_e^{(546.23\,\mathrm{nm})}$	67.67		Schott
Stress optical constant	$3.4 \cdot 10^{-12}$	1/Pa	Schott
Verdet constant $V_{SiO_2}^{(457.9\,\mathrm{nm})}$	7.05	rad/Tm	[89]
Sellmeier coefficients for λ in μm:			Schott
B_1: $6.694226 \cdot 10^{-1}$, B_2: $4.345839 \cdot 10^{-1}$, B_3: $8.716947 \cdot 10^{-1}$			
C_1: $4.480112 \cdot 10^{-3}$, C_2: $1.328470 \cdot 10^{-2}$, C_3: $9.534148 \cdot 10^{1}$			

Table 5.6.: Material properties of fused silica

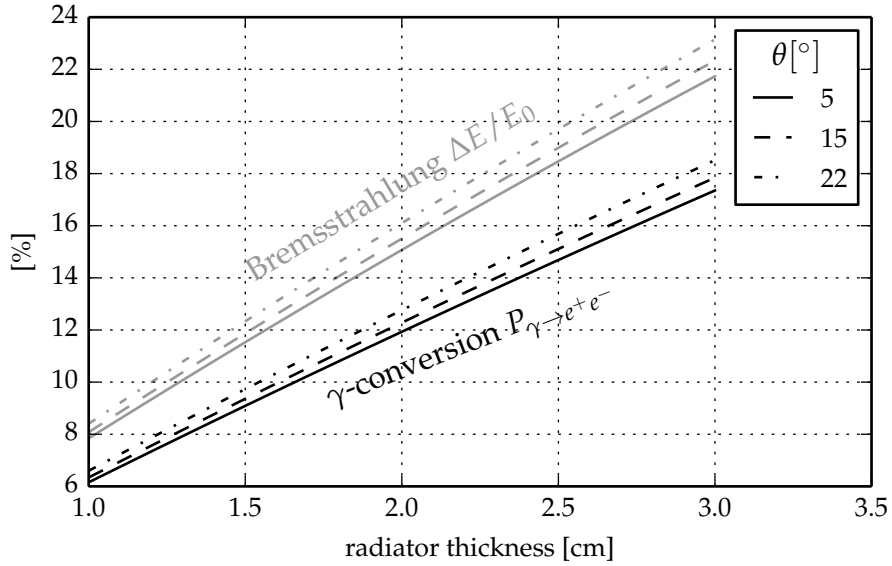


Figure 5.12.: Probability for gamma conversion ($E_\gamma > 1\,\mathrm{GeV}$, equation 5.28) and relative energy loss due to Bremsstrahlung (equation 5.27) in a fused silica radiator for different thickness and track polar angle θ.

scattering of the particle inside the radiator is degrading the DIRC performance by smearing the track angle at Cherenkov emission.

A Gaussian approximation of the scattering angle distribution has been provided by Highland, Lynch and Dahl [46, 75]. While underestimating the real distribution by neglecting its non-Gaussian tails, this approximation still agrees with Molière theory within 98 percent. The standard deviation in the plane-projection can be estimated by

$$\widetilde{\sigma}^{\mathrm{MCS}}_{\theta,\mathrm{track}} = \frac{13.6\,\mathrm{MeV}}{\beta c p}|z|\sqrt{d/X_0}\,(1 + 0.038\ln(d/X_0)) \tag{5.29}$$

Figure 5.13b shows $\widetilde{\sigma}^{\mathrm{MCS}}_{\theta,\mathrm{track}}$ for different particle species traversing a 2 cm thick fused silica radiator.

Besides MCS, the energy loss of hadrons is of importance. The associated deceleration of the Cherenkov emitting particle leads to a continuous decrease of the Cherenkov angle.

Fast charged particles – except the electron and its antiparticle – loose energy in matter primarily while interacting with the electrons in the material, causing excitation and ionization. The Bethe equation can be used to describe the mean

rate of energy loss of heavy charged particles, i.e. heavier than the electron, in the region of $0.1 \lesssim \beta\gamma \lesssim 1000$ [18]:

$$\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = K\frac{z^2}{\beta^2}\frac{\rho Z}{A}\left[\frac{1}{2}\ln\frac{2m_e c^2\beta^2\gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right] \tag{5.30}$$

with $K = N_A e^4/4\pi\epsilon_0^2 m_e c^2$, the integer charge of the incident particle z, the maximum energy transfer possible in a single collision $T_{\max}$, the mean excitation energy I and finally atomic number Z, atomic mass A and density ρ of the material. The term $\delta(\beta\gamma)$ is an empirical correction factor to account for the so called "density effect" which is related to polarization effects in media.

The shape of the energy loss distribution is asymmetric and the mean value given by the Bethe formula can be strongly influenced by the tail of this distribution. Thus, for many practical aspects it makes more sense to use the "most probable energy loss" [18] which, for media of moderate thickness x, is given by :

$$\Delta_p = \xi\left[\ln\frac{2m_e c^2\beta^2\gamma^2}{I} + \ln\frac{\xi}{I} + j - \beta^2 - \delta(\beta\gamma)\right] \tag{5.31}$$

with $\xi = (K/2)(\rho Z/A)(x/\beta^2)$ and $j = 0.2$.

Figure 5.13a shows the most probable energy loss in a 2 cm thick fused silica radiator for different particle species at different momenta, normalized to the kinetic energy of the incident particle. The energy loss is not only important for calorimetry, it is also connected to the Cherenkov angle which is a function of the velocity βc. When a particle slows down while traversing the radiator, the Cherenkov angle decreases with βc as shown in figures 5.13c, 5.13d.

To minimize the aforementioned perturbations, the radiator has to be as thin as possible. On the other hand, the choice of a thicker radiator results in a higher number of measured Cherenkov photons per particle track $N_{\text{ph.}}$ The reasons for this are two-fold. First, the number of emitted Cherenkov photons is proportional to the length of the path traversed by the emitting particle (equation 3.7 on page 32). Second, the number of reflections at the radiator surfaces during photon transport decreases with increasing radiator thickness what causes the reflection losses to be lower.

An increase in statistics leads to a reduction of the overall error σ_{θ_c} given in equation 5.2, which can be expressed in more detail using the aforementioned

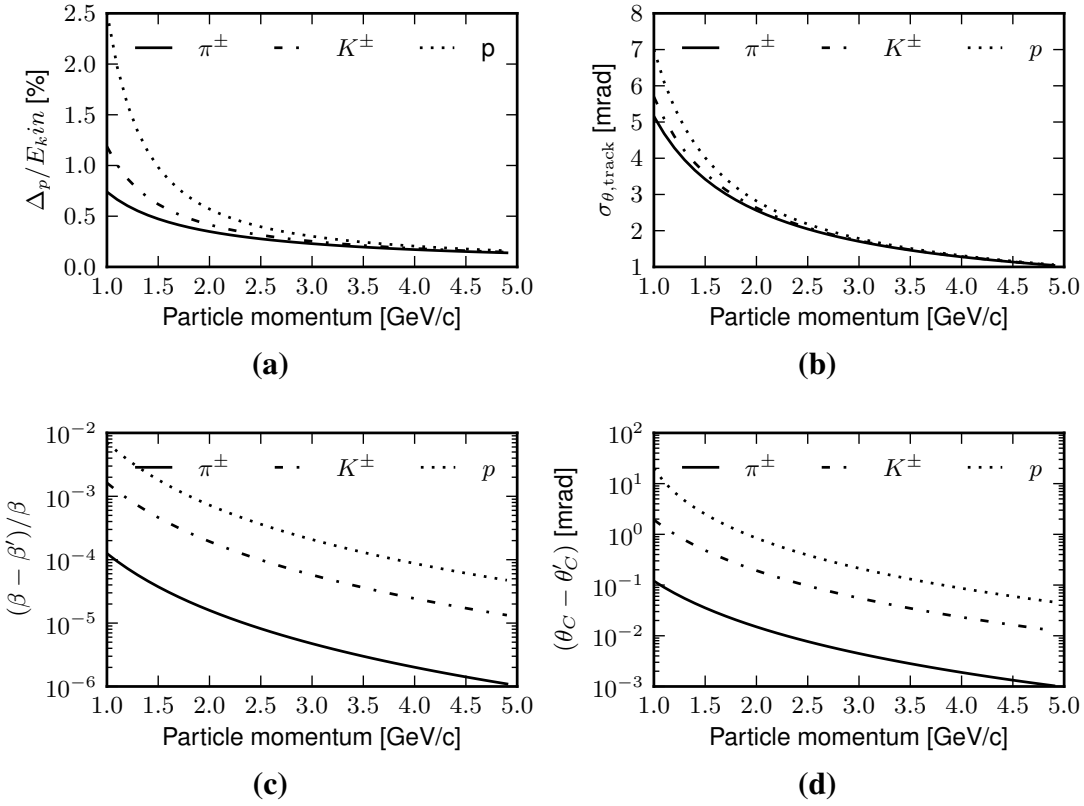


Figure 5.13.: Energy loss and multiple coulomb scattering in a 20 mm thick fused-silica radiator. (a): most probable energy loss Δ_p, computed using equation (5.31) and parameters for SiO_2 as published in [126]. (b): the standard deviation of the distribution of projected scattering angles, computed using equation (5.29). (c): relative difference of the velocity β before and β' after the particle lost the energy Δ_p. (d): the resulting difference of the corresponding Cherenkov angles θ_c, θ'_c.

errors:

$$\sigma^2_{\theta_c} = \underbrace{\frac{1}{N_{\text{ph.}}}\left[\sigma^2_{\text{im.}} + \sigma^2_{\text{chrom.}}\right]}_{\text{statistical term}} + \underbrace{\sigma^2_{\text{MCS}} + \sigma^2_{\text{dE/dx}} + \sigma^2_{\text{track}}}_{\text{track related errors}} \tag{5.32}$$

This equation can be used to analyze the impact of the radiator thickness on the detector performance. An increased number of photons $N_{\text{ph.}}$ will reduce the overall error as long as the statistical term is larger than the sum of the track related errors. The contributions of MCS and energy loss also grow with the radiator thickness.

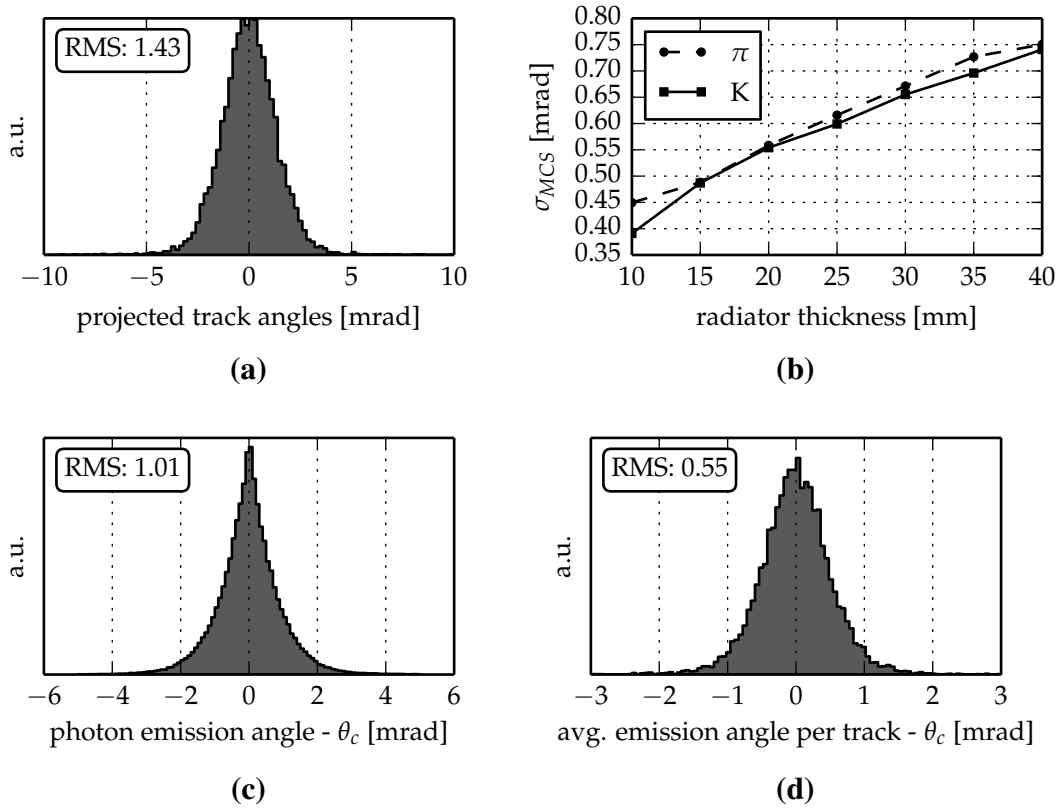


Figure 5.14.: Geant4 based multiple coulomb scattering simulation of 10^5 Kaons at a momentum of 4 GeV/c in a fused silica radiator. Figure (a) shows the distribution of the projected polar angles of the kaons after traversing a 2 cm thick radiator. The resulting RMS value is compatible with the expected value according to eq. 5.29. To determine the related emission error of the Cherenkov angle, a trapping fraction of 60 % has been emulated by a cut in the phi angle. The corresponding angular deviation of the Cherenkov emission angles from the undisturbed Cherenkov angle is shown in (c), the deviation of the mean Cherenkov angle per track from the undisturbed Cherenkov angle in (d). The dependence of the standard deviation of the mean Cherenkov angle per track on the radiator thickness is shown in (b).

A quantitative analysis is necessary to obtain a better understanding of the interplay of the individual terms. Therefore eq. 5.32 is evaluated for kaons at the acceptance limit of p =4 GeV/c. Kaons have been chosen as probe as they show stronger MCS due to the higher mass.

Some of the error contributions have already been discussed. A tracking error σ_{track} of 1 mrad has been defined as external requirement for the tracking system

in section 5.1. The contribution of the energy loss $\sigma_{\mathrm{dE/dX}}$ is negligible at higher momenta (Fig. 5.13d).

However, the error of the photon angle introduced by MCS of the incident particle is harder to estimate. The scattering model implemented in the Geant4 framework [5, 7] has therefore been exploited to get a better understanding on the impact of MCS on the Cherenkov angle distribution. For this purpose, the passage of pions and kaons at a momentum of 4 GeV/c through a fused silica radiator has been simulated using a modified physics model. To exclude the chromatic error from the analysis, the refractive index has been fixed to 1.47 over the whole wavelength range. Further, all other distorting effects have been deactivated in the physics list to focus on the influence of MCS only.

The photon error caused by MCS will also depend on the trapping fraction, which is fairly constant over the Endcap DIRC acceptance (Fig. 5.11). A trapping fraction of 60 % has been emulated in the analysis by applying an angular cut. Finally, the number of photons has been artificially reduced to 20. photons at a radiator thickness of 20 mm. The resulting error distribution of the photon angles (Fig. 5.14c) is non-Gaussian and does thus not scale with $\sqrt{N}$. The corresponding error of the mean Cherenkov angle per particle track is shown in (Fig. 5.14d). This error varies < 5 % for photon multiplicities of 20 to 100 per track and < 10 % for multiplicities down to 10 photons per track. MCS errors obtained for pions and kaons at different radiator dimensions are shown in Fig. 5.14b. These have been used as input for eq. 5.32.

To evaluate the statistical term in eq. 5.32, one has to know the number of detected photons. This number depends on the type of sensor which is used as well as on the type of dispersion correction. In the following, a filter approach is assumed which limits the chromatic error to approx. 1 mrad and results in a number of 20 detected photons per track at $d_R = 20$ mm. With the knowledge of σ_{MCS}, one can compute eq. 5.32 for different imaging errors $\sigma_{\mathrm{im.}}$. The result is shown in Fig. 5.15.

A realistic value for the imaging error is in the order of 3 to 6 mrad. In this scenario, a thicker radiator will always lead to an improved performance. The increase of σ_{MCS} is only relevant for unrealistically low imaging errors below 3 mrad.

As explained in section 5.5, it can be advantageous to limit the number of detected photons to deal with aging effects in the photodetector. In that scenario, one can define a bandpass filter for each radiator thickness which restricts $N_{\mathrm{ph.}}$ to

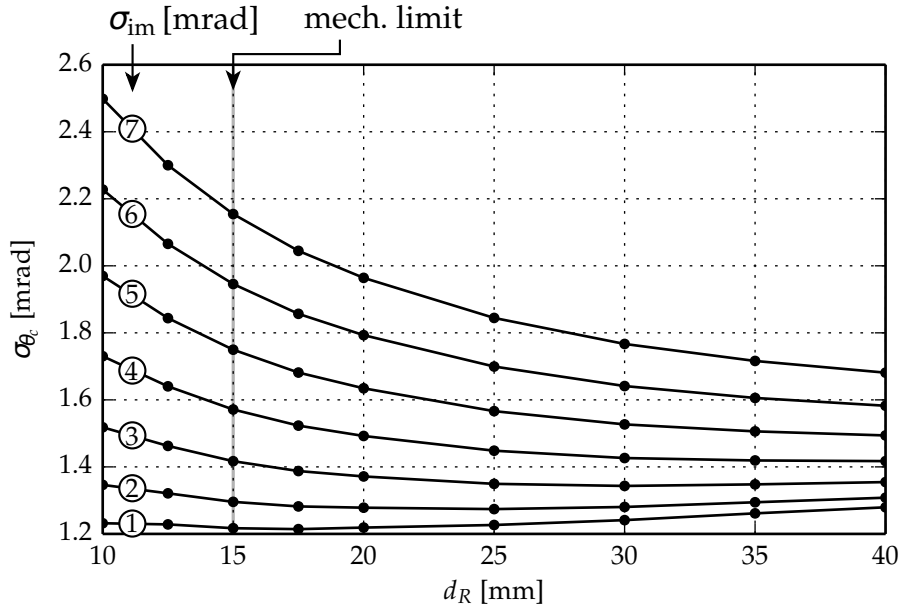


Figure 5.15.: Error of measured Cherenkov angle according to eq. 5.32 as function of the radiator thickness d_R, assuming a fixed wavelength band with a chromatic error of 1.6 mrad.

a fixed number of photons, e.g. 20. This means the number of photons is fixed, but the chromatic error varies with the radiator thickness. All filter pass-bands have been optimized with the approach discussed later in section X to yield the minimum chromatic error.

The resulting error σ_{θ_c} shown in Fig. 5.16 shows a wide plateau with a minimum in the vicinity of $d_R = 20$ mm. The increase of the chromatic error causes the rise towards smaller d_R while MCS starts to dominate the error at large d_R. In this scenario, the radiator can be as thin as 17.5 mm. Any increase in thickness will not lead to significant increase in performance. However, if the real system shows a slightly higher photon loss than the model used, the filter-bands would have to be wider and the chromatic term would be stronger, reducing the size of the plateau. Therefore, the thickness of 20 mm is advisable to include a certain safety margin.

Surface figure

Photons emitted in a 2 cm thick radiator by a 1 GeV/c kaon at 10° undergo more than 25 consecutive reflections at the radiator surfaces until they enter the optics.

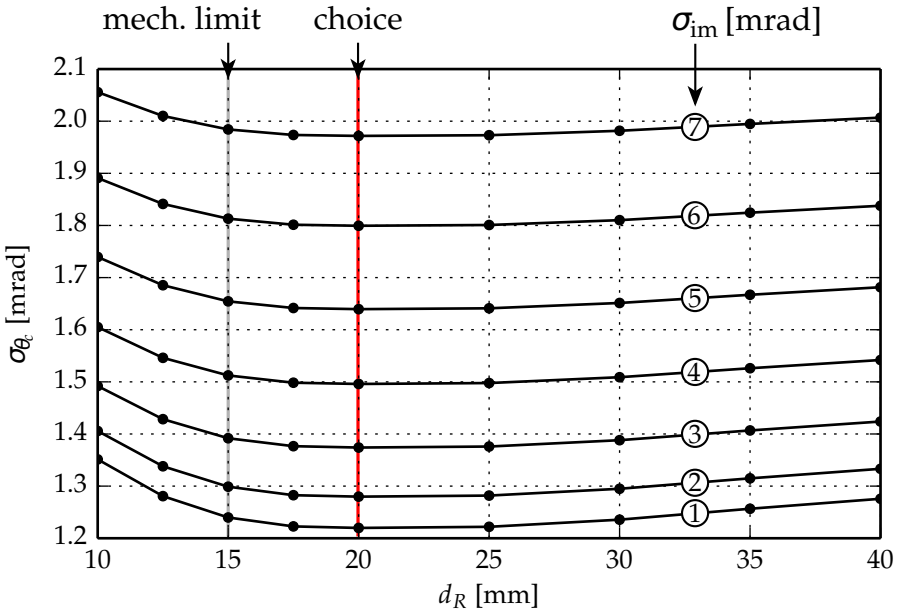


Figure 5.16.: Error of measured Cherenkov angle according to eq. 5.32 as function of the radiator thickness d_R, forcing a fixed number of photons $N_{\text{ph.}}$ =20 by altering the band width of the filter.

Hence, the surface figure of the radiator is overly important to reach the required resolution. An ideal radiator has perfectly flat and parallel surfaces. The quality of a real radiator however will be limited by the precision of the manufacturing process.

An optical surface can be characterized by its height profile $z(x)$. The surface error is the deviation Δz of the real profile from the design shape. Surface errors cause additional aberrations. The type of distortion depends strongly on the spatial frequency of the surface errors. As explained earlier, errors at the highest spatial frequency result in scattering losses (eq. 5.23). Hence, the surface roughness of the radiator has to be specified with respect to the expected photon loss. The reflection loss for $N = 50$ reflections $(1 - R_{\text{TIR}}^N)$ is shown in Fig. 5.17 as function of the photon wavelength for several values of rms surface roughness. The angle of incidence is 40°. These values match a photon which travels from a region close to the beam pipe to the outer rim of the Endcap DIRC. To cover this region with sufficient efficiency, the losses should not exceed 10 %. This corresponds to a required surface roughness between 1 nm and 1.5 nm.

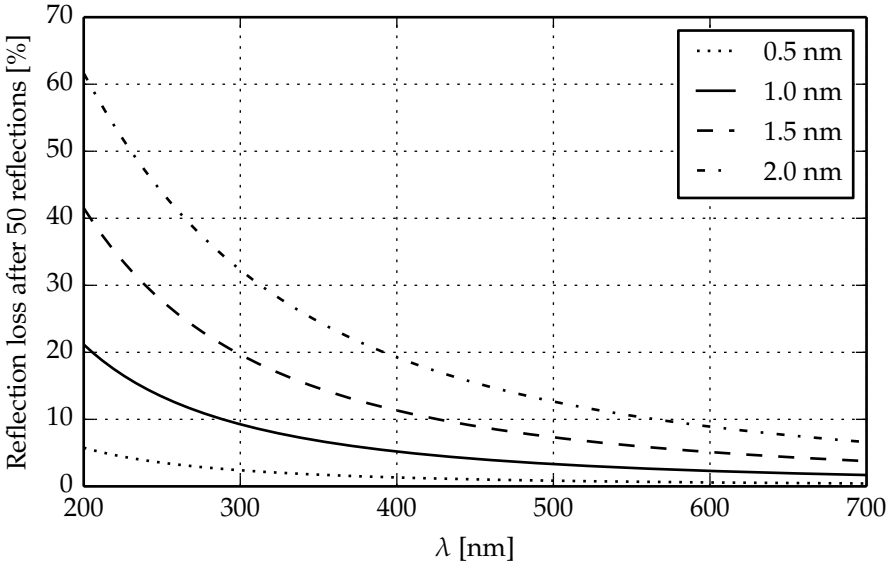


Figure 5.17.: Reflection losses in the radiator at different rms surface roughness (0.5 nm to 2 nm). The values correspond to the total integrated scatter after 50 reflections at an angle of incidence of 40°.

The definition of tolerances for errors with lower spatial frequencies is hardly possible without deeper knowledge of the surface topology which is expected for a given manufacturing process. The profile depends strongly on the type of machining applied as well as on the individual machine and processing parameters.

A suitable handle on the specification of the radiator surface is the RMS slope error σ_{slope} of the surface. In a naive model, one could assume that the slope error at each consecutive reflection is normally distributed. The error δ_{ph} of an individual photon after N reflections is then given by the sum of the errors from N samples out of s_i:

$$\delta_{\text{ph}} = 2 \sum_{i=1}^{N} s_i,$$

resulting in a statistical photon error of

$$\sigma_{\text{ph}} = \sqrt{N} \cdot 2\sigma_{\text{slope}}.$$

However, the surface profile is typically not random but has systematic errors caused by the tools used during processing.

Beside the statistical component, the radiator will show a certain deviation from parallelism. This can be included as a constant slope term δ_{wedge}, resulting in an

overall transport error of

$$\sigma_{\text{trans}} = \sqrt{N} \cdot 2\sigma_{\text{slope}} + N \cdot \delta_{\text{wedge}}$$

.

According to this toy model, a required maximum transport error of $\sigma_{\text{trans}} \leq 0.5\,\text{mrad}$ at $N = 50$ reflections leads to a specification of $\sigma_{\text{slope}} \leq 18\,\mu\text{rad}$ and $\delta_{\text{wedge}} \leq 5\,\mu\text{rad}$.

In practice, one can obtain the slope distribution from a 2D surface profile measured with an interferometer by computing the slopes between all combinations of points at a fixed distance l, which should be close to the resolution of the device. The rms slope can then be obtained from the distribution of these computed slopes.

A target wedge angle of $\delta_{\text{wedge}} = 5\,\mu\text{rad}$ (1 arcsec) is extremely challenging when polishing optical flats with a diameter of 1 m. Hence, it might become necessary to correct for this systematic error instead of avoiding it.

5.4.2. Imaging optics (C2)

has impact on	*key properties*
• Radiator thickness	• Material
• Photodetector size/granularity	• Geometry
• Photon transport	• Imaging performance
• Photon rate	• Efficiency

The imaging optics of the DIRC, often called camera, has the purpose to permit a precise measurement of the angle of Cherenkov photons entering from the radiator. In the Endcap DIRC, the reflection angle φ of the Cherenkov photon relative to the radiator plane is limited by the maximum track and Cherenkov angles

$$\varphi_{\text{min}} = 90^\circ - (22^\circ + \arccos(1/n)) = 20.86^\circ$$

as well as the limit for total internal reflection

$$\varphi_{\text{max}} = \arccos(1/n) = 47.14^\circ,$$

corresponding to a range of $\varphi_{\text{max}} - \varphi_{\text{min}} = 459\,\text{mrad}$.

Due to their steep angles, most Cherenkov photons stay trapped in the radiator and do not exit into air. While being advantageous during photon transport, this

behavior complicates the optical design. To avoid higher reflection losses, the light has to stay in a medium with a refractive index similar to or higher than the refractive index of fused silica.

As the DIRC operates at low light levels, optical systems should employ a low number of optical interfaces to keep scattering and reflection losses at a minimum. A set of basic options has already been introduced in Fig. 5.9 on page 70. The simplest option is a proximity focusing setup where the light is detected on a large sensitive area at a certain distance to the exit face of the radiator. In this case, the resolution is limited by the affordable active area as well as by the width of the light bundle which equals the radiator thickness. This approach has been successfully employed by BaBar. The restricted space for the Endcap DIRC however excludes the use of such a simple camera.

To reduce the size of the expansion volume and active area, the width of the light bundle can be reduced by means of focusing optics. Viable options are the use of single and compound lenses as well as focusing mirrors or any combination of those.

A case study of a lens based imaging system is presented in Fig. 5.18. Focusing is realized by a single plano-convex spherical lens made from undoped* SF6 glass. This glass has been chosen because it has a high refractive index (1.87 at 400 nm) while being more radiation tolerant than conventional glasses. The lens is realized as a doublet with a fused silica plano-concave counterpart, so that it can be easily bonded to the planar surfaces of both prisms. The lower edge of the prism at the radiator interface serves as mirror to cover photons with negative φ angles, which are not drawn for clarity.

The optical setup in Fig. 5.18 serves merely as an example how a lens based approach could be realized and to visualize the downside of lens based designs in general. The red and green dashed lines represent focused bundles of parallel rays at 500 nm and 300 nm respectively. The design shows a strong chromatic aberration, what can be expected as high index glasses tend to have higher dispersion than low index glasses (see also Fig. 5.25). In standard environments, the chromatic errors can be limited by constructing achromatic doublet lenses from materials with different Abbe numbers. In a high radiation environment however, the number of available optical materials is very limited and many of the radiation hardened glasses have a low transmission below 500 nm what is not a good match

*Some glasses can be doped with Ce to reduce radiation induced absorbance (see also section 5.4.5

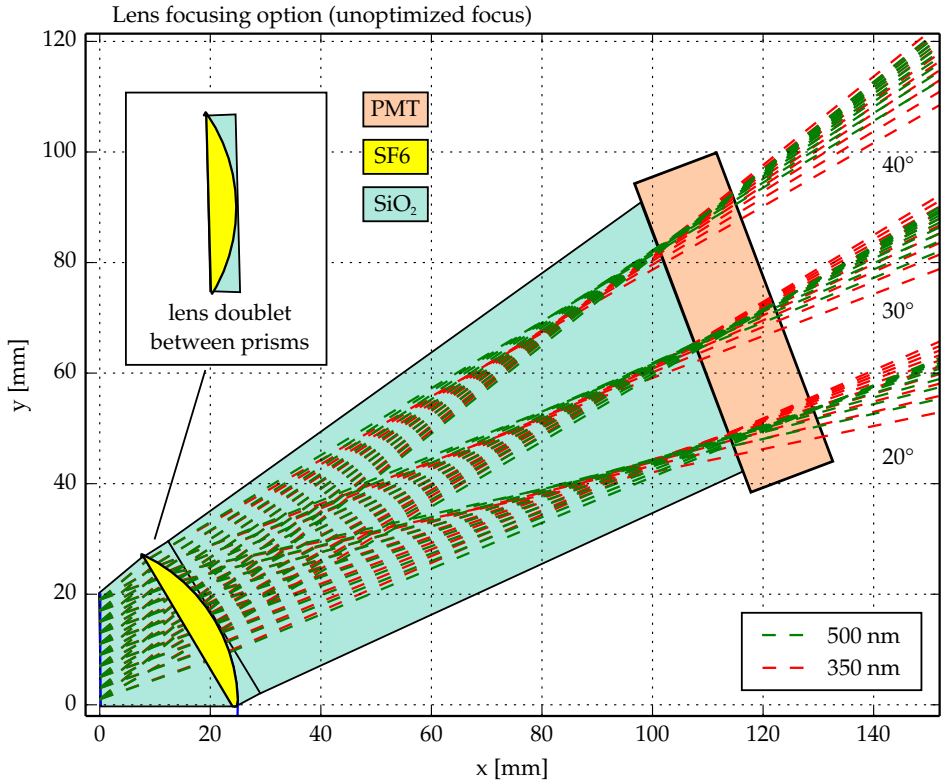


Figure 5.18.: Example of a lens based focusing setup for the Endcap DIRC, traced with PyOptics (section 6.4). Cherenkov light is entering from a 20 mm thick radiator at x=0 mm. Only rays with positive slope have been drawn. The corresponding rays with negative slope will be reflected at y=0 mm in the first fused silica prism and focused in a similar way. Note that this system has not been optimized using multivariate optimization routines, as it has been done for the final optics. The focus can be further reduced. However, the use of refractive elements introduces additional chromatic aberration (the focal length varies with the wavelength) which can be larger than the wavelength dependence of the Cherenkov angle itself. Achromatic doublets can hardly be designed due to the limited number of radiation hard optical materials. These complications render lens based optics rather unattractive.

to the sensitivity of available photosensors (see also Fig. 5.30).

The second option is the use of focusing mirrors. As the light cannot be easily extracted from the radiator into air, the mirrors have to be coated on a fused silica part so that the reflection happens at the glass-metal interface. The higher refractive index of the glass leads to a reduced reflectivity compared to the air-metal interface (Fig. 5.19). For applications which need good reflectivity below ~400 nm, the metal type is restricted to aluminum. At higher wavelengths, aluminum is outperformed by silver. Alternative or in addition to metal coatings, dielectric multilayer coatings can be applied. This type of coating is briefly discussed in section 5.4.4.

Focusing mirrors have the advantage that they do not introduce any additional chromatic aberration. However, the use of metal mirrors implies absorption losses in the order of 10 %. As the reflection happens from within a fused silica part, total internal reflection can be exploited as well. Such *focusing lightguides* with a 5-th order polynomial surface have been proposed by K. Föhl [39]. The 10 % gain in photon statistics come at the price of acylindric surfaces which can hardly be manufactured using standard processing. Custom tooling and metrology have to

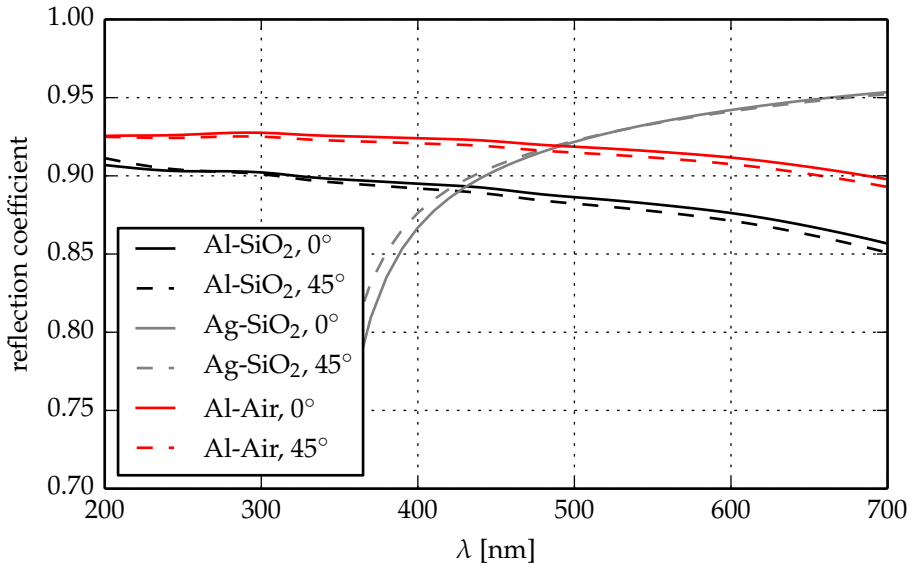


Figure 5.19.: Computed reflectivity of aluminum/air, aluminum/fused-silica and silver/fused-silica interfaces for the incidence angles 0° and 45°. The complex refractive index of Al and Ag has been taken from [104] who used the source [106, 107].

be developed at the production site.

An imaging approach based on cylindrical focusing shapes has been proposed by T. Kamae in 1996 [58]. Cylindrical surfaces can be produced and tested by means of standard tooling. Disadvantages are the photon loss at the coating as well as a lower number of degrees of freedom – three for the cylindric surface versus five in case of the polynomial.

Experience showed that most manufacturers are willing to produce cylindrical optics, but only few are interested in the production of acylindirc optics in small lot sizes. Hence, the choice of the surface type has a significant impact on the availability of suppliers and cost.

In conclusion, optics based on focusing mirrors are the most attractive solution as they comprise a low number of components, e.g. one precision polished block made from fused silica with an applied coating, and do not cause additional chromatic aberration. Cylindrical optics are preferable because they can be produced with standard tooling. However, polynomial optics allow a more flexible design due to the increased number of degrees of freedom. The design of lens based optics is basically limited by the availability of radiation tolerant glasses with a high transmission below 500 nm.

5.4.3. Optical coupling (CA1)

has impact on	*key properties*
• Photon transport • Photon rate	• Refractive index • Transmission • Outgassing

The term optical coupling describes the process of connecting two optical parts at a shared boundary in a way that the distortion of the traversing light is minimized. To realize this goal, the optical properties of the coupling material have to match the properties of the parts as close as possible. Independent of the specific type of coupling, optical losses are primarily caused by three effects:

- low transmittance of the coupling medium (absorption losses)
- difference in refractive index (reflection losses)
- gas bubbles (μm scale) at the interfaces or inside the coupling medium (scattering losses)

The first two points can be controlled by a careful selection of the coupling medium. It must have a good transmission in the application specific wavelength range and should closely resemble the refractive index and dispersion of the connected parts. The last point is related to the bonding process itself and thus not material specific.

The choice of a coupling medium further depends on the mechanical requirements, whether the coupling should be reversible or not and finally on the environmental conditions. In the case of the Endcap DIRC, the radiation hardness is the most important requirement. Also the effect of the coupling medium on the environment has to be considered. Some coupling substances can pollute the system by outgassing. A condensation of these substances on the radiator surface would deteriorate the reflection probability.

The coupling of optical elements can be either realized reversible or permanent. Transparent liquids, gels or silicone rubber polymer films are frequently used for reversible connections. Common optical cements for permanent bonding are epoxy resigns, acrylics, urethane adhesives and silicone polymers. The latter two options result in more flexible bonds what can be advantageous when bonding materials with different coefficients of thermal expansion (CTE).

The refractive index, absorption length and radiation tolerance of several optical cements have been measured in the visible part of the electromagnetic spectrum by Montecchi and Ingram [84] (published in 2001). Nearly all of the mentioned adhesives are still available on the market.

Name	***type***	***n (430 nm)***	***Λ [cm]***	***λ_{min} [nm]***	***curing***
GE/Bayer RTV 615	Silicone rubber	1.47	> 15	< 300	curing agent
NOA-61	UV curing polymer	1.59	0.55	~350	UV light
EPOTEK 301-2	two-comp. epoxy	1.60	> 10	300	curing agent
Bicron BC-630	silicone grease	1.47	> 10	< 300	–

Table 5.7.: Properties of radiation tolerant adhesives published in [84] and BC-630 optical grease (n = refractive index, Λ = absorption length, λ_{min} = cutoff wavelength). Transmission below the cutoff wavelength is negligible. Note that a very thick layer (~ 1 mm) of NOA-61 is required to efficiently block light below 350 nm. EPOTEK 301-2 has a very steep edge but the NOA-61 transmission raises slowly from 350 nm to 380 nm. Hence, NOA-61 cannot be used as edge filter.

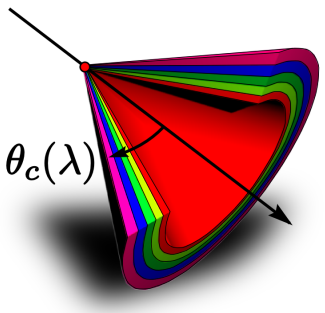
The closest match to the refractive index of fused silica (1.467 at 430 nm) is reached by silicones, followed by the epoxies and polymers. The most interesting options are summarized in table 5.7.

The NOA-61 adhesive is also considered as option for the Belle 2 iTop counter. BaBar has used the EPOTEK-301-2 epoxy to glue the radiator bars of the DIRC. Both are also viable solutions for the Endcap DIRC. While most of the properties of these two adhesives are similar, they have a slightly different cutoff wavelength. This difference may be exploited to tailor the wavelength acceptance of the DIRC.

The silicone adhesive or grease is an interesting option to form a reversible bond between the photodetector and the optical element. Reversible bonds are favorable to allow the replacement of the sensor in case of failure. This contact could also be realized by using a silicone rubber pad.

5.4.4. Dispersion correction

One of the major factors which limit the resolution of DIRC detectors at higher momenta is the broadening of the Cherenkov angle distribution caused by the chromatic dispersion of the radiator material: $\theta_c(\lambda) = \arccos\left[1/\left(\beta n(\lambda)\right)\right]$ The Cherenkov angle is smeared out to a band of Cherenkov angles. In section 5.4.1 it has been explained that the material choice is restricted to synthetic fused silica, primarily because of the high radiation levels. Fig. 5.20 shows the Cherenkov angle bands in the wavelength interval from 300 nm to 700 nm for electrons, pions, kaons and protons in this material.



Dispersion affects the Cherenkov cone

The distance between Cherenkov angles of different particle types becomes smaller with increasing momentum. This fact defines the ultimate performance limit of a DIRC detector. However, as a consequence of the unknown photon wavelength, the performance saturates much earlier due to the chromatic broadening of the Cherenkov angle. The θ_c-bands for pions and kaons show a significant overlap at higher momenta, what starts to become an increasing difficulty at momenta above ~3.5 GeV/c.

The Endcap DIRC has to provide π/K–separation up to momenta of 4 GeV/c. Hence, it is advisable to mitigate the chromatic error $\sigma_{\text{chrom.}}$ caused by dispersion.

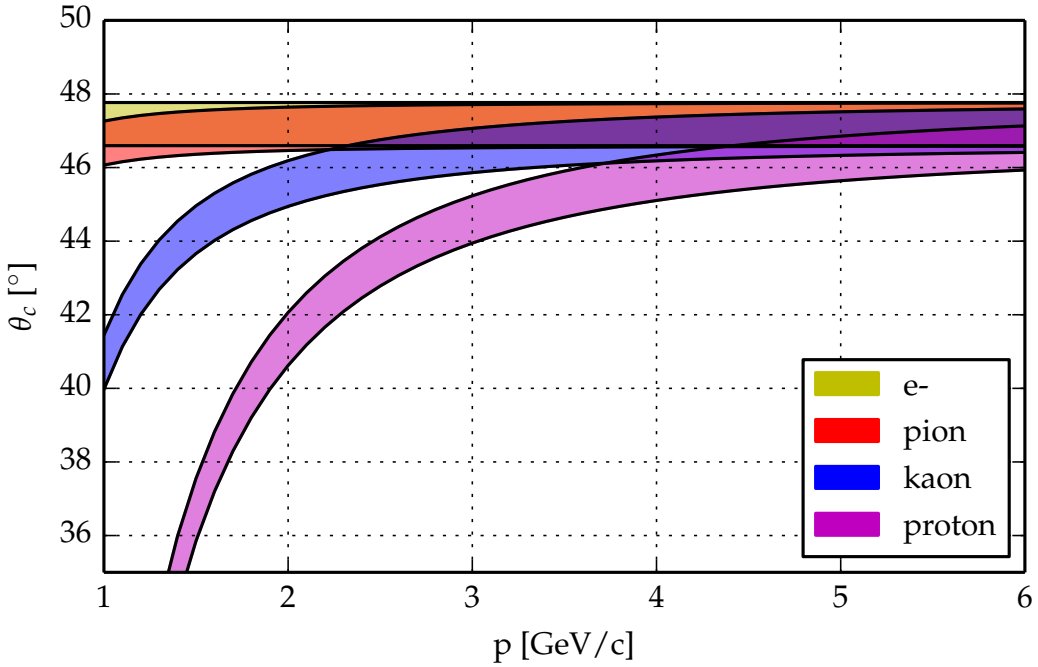


Figure 5.20.: Cherenkov angles in a fused silica radiator for a wavelength band from 300 nm to 700 nm. At higher momenta, the bands start to overlap. In this region the detector resolution begins to be limited by chromatic dispersion. The error can be mitigated by counting statistics, by a reduction of the accepted wavelength range or by apochromatic optics which correct the angular deviation by a combination of materials with different refractive index and dispersion. The Endcap DIRC has to separate pions from kaons up to momenta of >4 GeV/c.

Basically, there are four handles on the reduction of $\sigma_{\text{chrom.}}$:

1. higher photon statistics ($\sigma_{\text{chrom.}}/\sqrt{N_{\text{ph.}}}$).
2. reduction of the wavelength acceptance (spectral selection).
3. correction of the chromatic broadening by achromatic optics.
4. correction by means of the photons time of flight.

Higher photon statistics (1.) is always advantageous as long as the count rate stays compatible with other constraints. A higher count rate tightens the requirements on the sensor and front-end electronics and increases the data rate.

A limited wavelength acceptance (2.) leads to a reduced count rate. This can be an advantage if the expected count rates are not compatible with other system

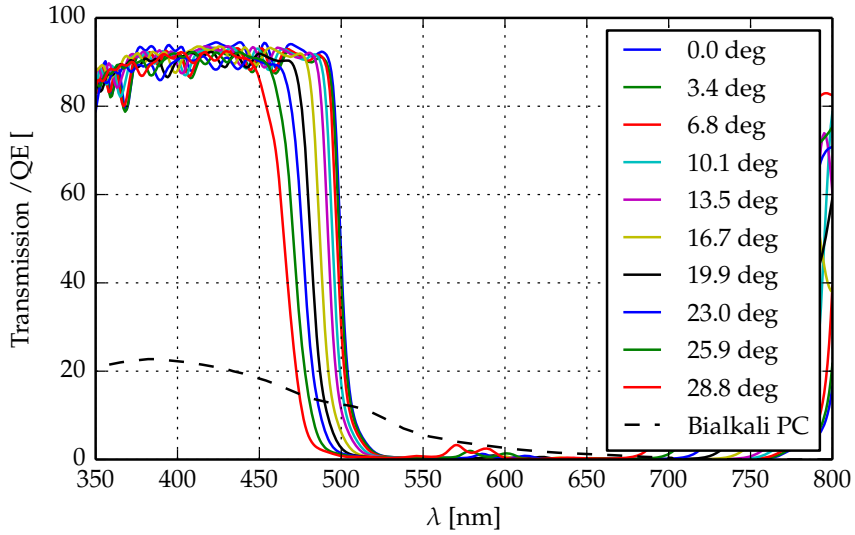
requirements. Furthermore, the term $\sigma_{\text{chrom.}}/\sqrt{N_{\text{ph.}}}$ becomes smaller even with the reduction of $N_{\text{ph.}}$. Hence, if the imaging error $\sigma_{\text{im.}}$ in eq. 5.32 is very small, a reduced acceptance can still lead to an increase in the overall performance of the detector. In practice however, the imaging errors are too large to benefit from this trade-off between chromaticity and statistics.

The reduction of the wavelength acceptance can be realized by using materials with different bulk transmission or reflection characteristics. The first class comprises the glass types and coupling media used in the optical design as well as absorptive optical filters, e.g. colloidally colored glasses. The second class covers different types of mirror coatings, e.g metals and interference mirrors/filters based on dielectric multilayer systems.

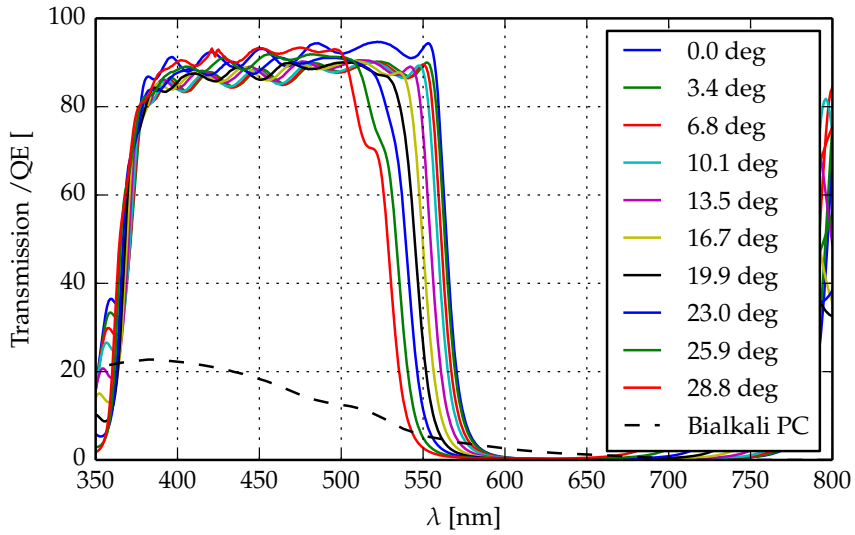
The highest flexibility is provided by optical filters as these are already available for many different wavelength ranges and can be tailored to user requirements as well. Absorptive filters are made from base glasses which are usually not radiation hard. The radiation hardness of commercial dielectric multilayer mirrors and filters has been investigated in the context of applications in fusion reactors [96, 76] and space missions [102]. In fact, no or only negligible degradation has been observed after irradiation with γ-rays up to several MGy, a fluence of $> 10^{10}$ protons/cm^2 at 60 keV and 30 MeV and a fluence of 10^{19} fast neutrons/cm^2. These values exceed the expected radiation in $\overline{\text{P}}$ANDA by orders of magnitude.

The drawback of interference filters in comparison to absorptive filters is the strong dependence of the filter characteristics on the angle of incidence (AOI). As shown in Fig. 5.21, the edges of a bandpass or edge filter are shifted towards shorter wavelengths with increasing AOI. This intrinsic behavior is a direct consequence of the working principle of interference filters. Basically, these consist of a stack of thin-film Fabry–Pérot interferometers which are tuned to reinforce the reflection or transmission of certain wavelengths. Broad band filters consists of hundreds of these layers with different thickness and refractive index. The effective thickness of each layer depends on the AOI, and so the wavelength at which destructive or constructive interference is observed. These filters can be mathematically optimized and hence precisely tailored to the customers requirements.

A possible use in a DIRC detector implies that the light is entering the filter from glass or a coupling medium and not from air. As filter characteristics are typically measured in air, the angles given in product specifications have to be



(a)



(b)

Figure 5.21.: Transmission of two commercial dielectric multilayer filters sold by Thorlabs Inc, FD1B (a) and FD1C (b), at different angles of incident. The plotted data has been provided by this supplier. The angle of incidence shown in the legends has been converted from angles in air (0°–45° in 0.5° steps) to the corresponding angles in fused silica. The rising transmission at wavelengths above 650 nm is insignificant for sensors with a bialkali photocathode (dashed line).

converted to the corresponding angle in glass:

$$\alpha_{\text{glass}} = \arcsin[\sin(\alpha_{\text{air}}) / n_{\text{glass}})].$$

Two examples of the transmission of commercially available filters are shown in Fig. 5.21. These filters are suitable to be used in the Endcap DIRC. The lower wavelength cutoff of filter (b) at 350 nm can be reinforced by choosing NOA-61 as glue (table 5.7) for the whole optical system. To reduce the number of optical interfaces which have to be coupled using glue or gels, one can think about coating the filter system directly on the frontglass of the MCP-PMT prior to the assembly of the tube. A first discussion of this proposal with a PMT manufacturer indicated that the materials commonly used in interference filters are compatible with their vacuum system [27]. This choice of filter placement does not increase the number of optical interfaces which have to be coupled. An alternative approach would be to coat the optical element attached to the PMT, given that the size of these elements is compatible with the manufacturers process.

Option (3.) on the list of approaches to dispersion correction is the use of achromatic optics. A common option to correct for the chromatic aberration in lens system is to construct an achromatic doublet, consisting of a positive and negative lens, which are made from materials with different Abbe numbers and cemented at a shared surface. Even better correction can be achieved by using fluorite crystals or fluoro-crown glasses, as these show slightly different dispersion characteristics than most other glasses. Similar effects can be achieved by using prisms instead of lens systems [43].

An early Endcap DIRC proposal by B. Morozov [87] included LiF single crystals for the mitigation of the chromatic error. This technique has later been adapted by a proposal of K. Föhl. LiF shows excellent transmission in the visible and UV part of the spectrum as well as anomalous dispersion, but is also relatively expensive due to the complex production process. Even though LiF is used for neutron imaging, exploiting the capture reaction $^6\text{Li} + \text{n} \rightarrow \alpha(2.05\text{MeV}) + 3\text{H}(2.73\text{MeV})$, its radiation hardness is compliant with the system requirements.

However, standard low-cost achromatic lens systems are designed completely without expensive low-dispersion glasses or crystals. This should also be possible in case of achromatic prism setups. A simple 2D model, sketched in Fig. 5.22, has been set up to review this problem and to identify alternative material combinations. Cherenkov light from the radiator is traced through two surfaces of a prism which enclose an angle α. The second surface forms the interface to the focusing optics.

The radiator material is fixed, but the other two materials can be selected freely to form the analogon of an achromatic doublet. The difference between the exit angles φ for two rays at 400 nm an 500 nm has been chosen as figure of merit (FOM). The plot in Fig. 5.22 shows the resulting FOM for various material combinations versus the prism angle at different particle angles θ_p. A variation of θ_p shifts the resulting v-shaped curves along the α-axis. In conclusion, to minimize the FOM for all theta angles between 10° and 22°, one has to choose the α angle were the solid and dashed curves intersect (points marked by the colored dots).

All material combinations presented in Fig. 5.22 show a comparable FOM of ≤ 1 mrad. The combination of fluorite single crystals LiF and CaF_2 with fused silica results in an opening angle between 90° and 100°. The performance of an fluorite crown glass (FK51A) with the radiation hardened BK7G18 glass results in a similar prism angle. Unfortunately, these fluorite glasses are not radiation tolerant. To form an achromatic doublet from radiation hard glasses, one has to select a high index glass to reach sufficient difference in the Abbe number. A viable solution is the combination of BK7G18 with LF5G19 at a prism angle close to 70°. As radiation hardened glasses trade transmission for stability, this solution can only be used at relatively long wavelengths > 500 nm. As discussed in section 5.5, this wavelength range can only be accessed efficiently by means of photodetectors with GaAsP-photocathodes. In case of an already dispersion limited design, this restriction on longer wavelength is counterproductive because the number of Cherenkov photons per wavelength interval increases strongly with shorter wavelengths. Further, as radiation hard glasses are produced on request, the price will also be high. Hence, the LiF or CaF_2 based setups are still the best choice.

The last option (4.) for dispersion correction is the precise measurement of the photon propagation time t_p. As explained earlier, the dispersion of the glass affects the phase and group velocity. Hence, the velocity of the light inside the radiator becomes wavelength dependent (eq. 5.12) and thus t_p (eq. 5.6). Conversely, the measurement of t_p can be used to determine the photon wavelength. In conclusion, this technique is analog to a coarse wavelength measurement.

To obtain the time of flight one has to know the correct length of the photons path. In a DIRC however, the path length depends on the photon angle – the quantity one aims to correct. For sufficiently small wavelength ranges $\Delta\lambda$, the

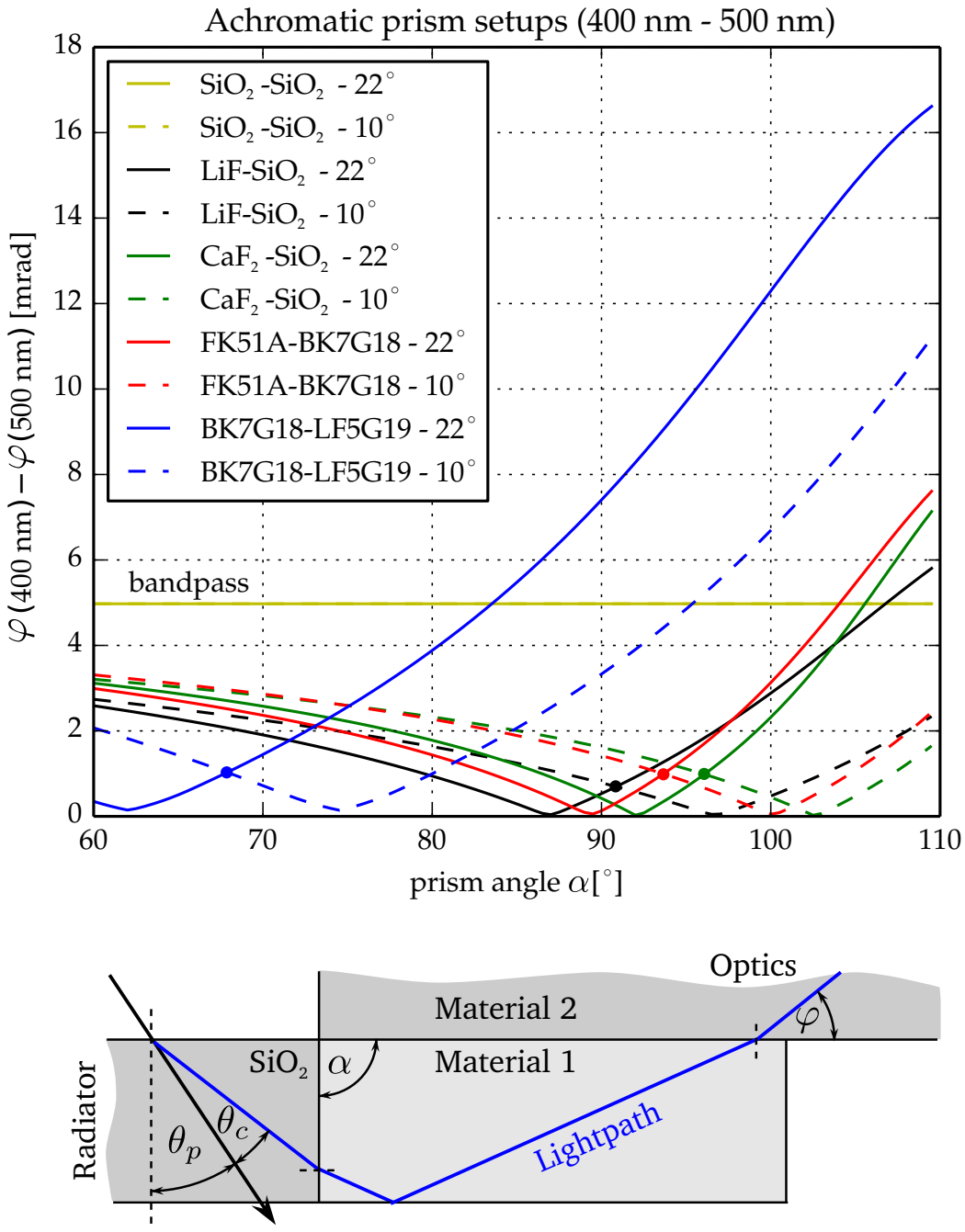


Figure 5.22.: Achromatic prism setups by means of different optical materials. The model is restricted to the 2d case and neglects the influence of the optical coupling between individual materials. The material of the prism and optics as well as the particle angle θ_p are given in the legend.

chromatic error $\Delta\theta_c$ of the Cherenkov angle can be estimated by

$$\Delta\theta_c \approx \frac{\partial\theta_c}{\partial\lambda}\Delta\lambda = \frac{\partial\theta_c}{\partial n}\frac{\partial n}{\partial\lambda}\Delta\lambda$$

with

$$\frac{\partial\theta_c}{\partial n} = \frac{\partial}{\partial n}\left[\arccos\left(\frac{1}{n\beta}\right)\right] = \left(n^2\beta\sqrt{1-\frac{1}{(n\beta)^2}}\right)^{-1} = \frac{\cos\theta_c}{n\sin\theta_c} = \frac{1}{n\tan\theta_c}$$

resulting in

$$\Delta\theta_c \approx \frac{1}{n\tan\theta_c}\frac{\partial n}{\partial\lambda}\Delta\lambda. \tag{5.33}$$

An expression for the chromatic error of the propagation time for a known path s can be obtained in the same manner:

$$\Delta t_p \approx -\frac{s}{c}\lambda\frac{\partial^2 n}{\partial\lambda^2}\Delta\lambda \tag{5.34}$$

By measuring light from a radiator plate however, one does not know s but only its projection s_{2d} in the radiator plane. The full path is obtained by taking the angle of the photon into account. The angle φ between the photon vector and the radiator plane can be computed from the particles polar and azimuth angles θ_p, ϕ_p (z-axis perp. to the radiator plane) and the Cherenkov angle using eq. 5.3. The real path of a photon can then be computed by the known projected path s_{2d} from the point of emission to the point at the radiator rim. The resulting expression for the propagation time is then

$$t_{p,c} = \frac{s_{2d}}{c\cdot\cos\varphi}\left(n-\lambda\frac{\partial n}{\partial\lambda}\right) \tag{5.35}$$

and the corresponding chromatic error

$$\begin{aligned}\Delta t_{p,c} &\approx \frac{s_{2d}}{c}\left[\frac{\partial(\cos\varphi)^{-1}}{\partial\lambda}\left(n-\lambda\frac{\partial n}{\partial\lambda}\right)-(\cos\varphi)^{-1}\lambda\frac{\partial^2 n}{\partial\lambda^2}\right]\Delta\lambda \\ &= \Delta t_p + \frac{s_{2d}}{c}\frac{\partial(\cos\varphi)^{-1}}{\partial\theta_c}\frac{\partial\theta_c}{\partial\lambda}\left(n-\lambda\frac{\partial n}{\partial\lambda}\right)\Delta\lambda \\ &= \Delta t_p + \Delta_c\end{aligned} \tag{5.36}$$

has an additional term Δ_c due to the wavelength dependence of s. The error σ_λ of a wavelength measurement by means of a t_p measurement with precision σ_{top} can be computed by

$$\sigma_\lambda = \sigma_{\text{top}} \cdot \left(\frac{\partial t_{p,c}}{\partial \lambda} \right)^{-1} \tag{5.37}$$

The resulting resolution is shown in Fig 5.23, computed for an incident pion at 3 GeV/c, $\theta_p = 15°$, $\phi_p - \phi = 0°$. Note that the particle parameters have a negligible influence on the overall performance figure of this method. In a realistic scenario, the measured photon angle φ and the projected path s_{2d} will have errors as well. The performance estimated using eq. 5.37 gives the resolution limit.

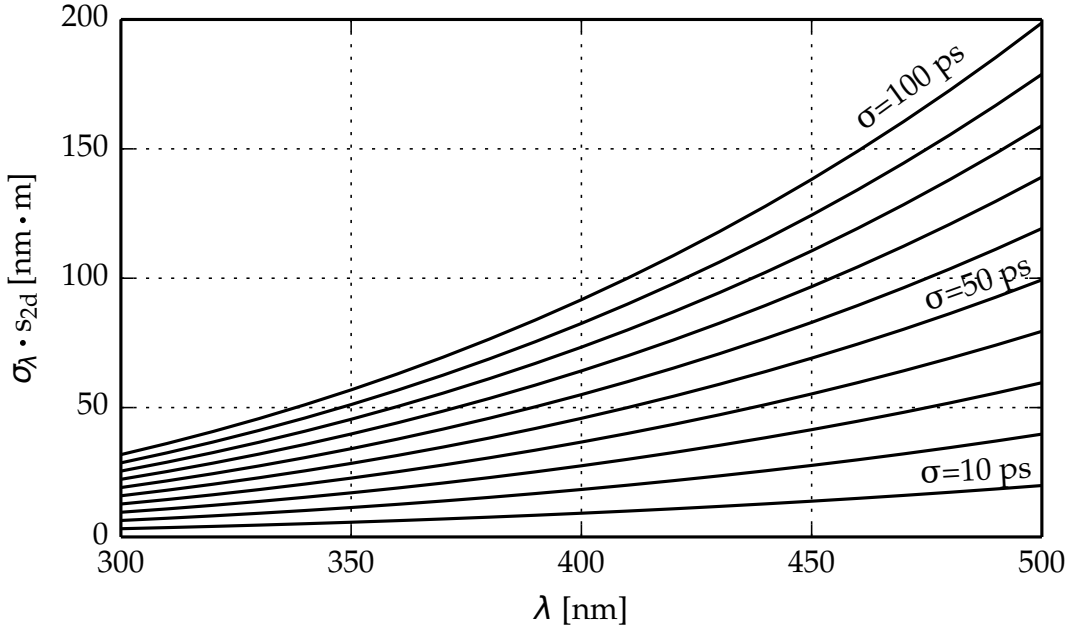


Figure 5.23.: Wavelength error resulting from a time of propagation measurement over a one meter projected path with time resolutions ranging from 10 ps to 100 ps.

This approach to dispersion correction was part of the fast focusing DIRC concept planned for the now canceled SuperB-facility in Italy. The reverse way, correcting the time by means of a precise angle measurement, is also feasible. However, this is only helpful if the detector is used as a time-of-flight stop. A time of propagation approach is more complicated to use in $\overline{\text{P}}$ANDA than in most collider experiments because of the missing reference time.

5.4.5. Radiation hardness of glasses

Radiation damage is a complex phenomenon which can affect a multitude of different material properties. For glasses, the most evident and widely known effect is the formation of color-centers. These color centers form absorption bands which occur predominantly in the VIS-UV region of the spectrum. Hence, radiation effects have to be considered during the design of optical components for particle physics experiments as well as spaceborne applications. Especially publications originating from the latter area of research are a valuable source of information.

The radiation hardness of synthetic fused silica has been measured by M. Hoek et al. [48], [49]. Figure 5.24 displays some of the results presented in [48]. The curves show the degradation of the bulk transmission for samples with different hydrogen content. In all cases, the transmission loss is only significant for wavelengths below 350 nm. If this part of the spectrum is needed, a synthetic fused silica with proper hydrogen content has to be selected. On the whole, synthetic fused silica is extremely radiation hard and fulfills all Endcap DIRC requirements.

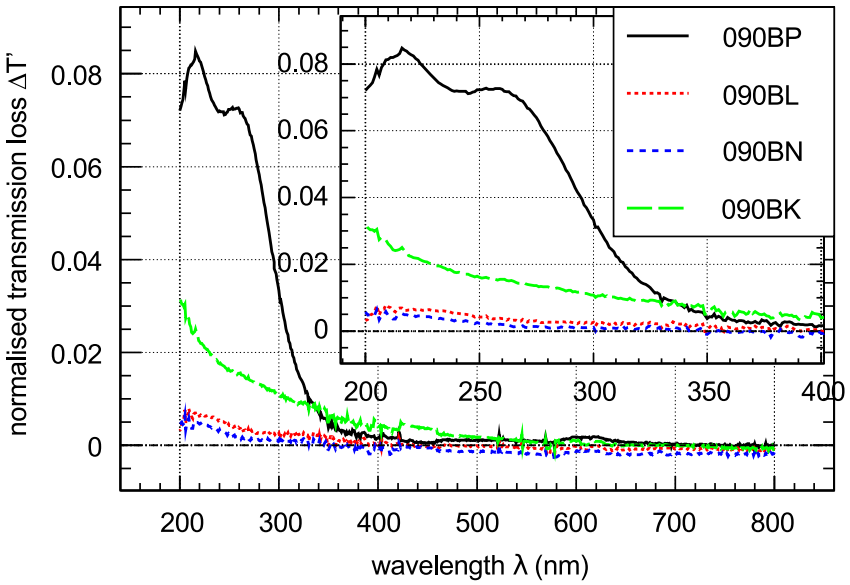


Figure 5.24.: Measurement of the transmission degradation in fused silica samples with different hydrogen content, irradiated with a γ-dose of approximately 100 krad. $\Delta T' = (T_0 - T_1)/T_0$, with transmissions T_0 before and T_1 after irradiation. [48]. Hydrogen content of the samples [mol/cm^3]: 090BP $< 1.0 \cdot 10^{15}$, 090BL $1.3 \cdot 10^{16}$, 090BN $1.4 \cdot 10^{17}$, 090BK $1.7 \cdot 10^{18}$.

The authors of [30] present extensive studies concerning the radiation hardness of glasses with a focus on space optical instruments. Their measurements include the irradiation of glasses with γ-rays and protons. The results indicate that nearly all of the standard optical glasses are not radiation tolerant enough to fulfill the Endcap DIRC requirements. Low dispersion glasses, which can be used to design dispersion correcting optics, are also not qualified to be used in the Endcap DIRC optics [146, 145].

Schott offers special Ce doped glasses which are more resistant to ionizing radiation. The trade-off is a low transmission at shorter wavelengths. This makes these materials incompatible with the wavelength acceptance of bi- and multialkali photocathodes (Fig. 5.30). The available Ce-doped glasses are shown as red squares in the Abbe Chart in Fig. 5.25. Optical designs based on these glasses would require to select a wavelength range starting at ~400 nm.

Single crystals like CaF or LiF are radiation tolerant if they show a low level of impurities. The formation of color centers in LiF induced by irradiation with γ-radiation, 3 MeV protons and thermal neutrons has been investigated in [78, 141],

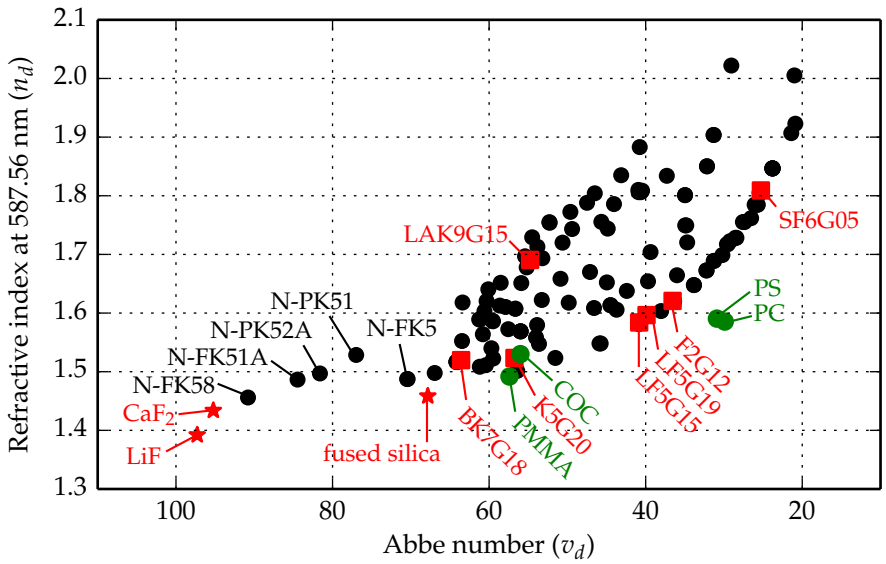


Figure 5.25.: Abbe chart showing optical glasses from the Schott catalog (black circles) together with intrinsically radiation hard materials (red stars) and radiation hardened glasses from Schott (red squares). The latter are stabilized by adding Ce to the glass matrix. The green circles show common optical plastics.

[101] and [56], respectively. Based on the results presented in these publications, one can conclude that the expected total ionizing dose and neutron fluence at the Endcap DIRC perimeter will only slightly deteriorate the optical properties of a LiF crystal. Several absorption bands are formed by the creation and aggregation of color centers. The most prominent absorption bands are the F ($\lambda_F = 247\,\text{nm}$) and M ($\lambda_M = 443\,\text{nm}$) band. Two less dominant bands R_1 ($\lambda_{R_1} = 315\,\text{nm}$) and R_2 ($\lambda_{R_2} = 374\,\text{nm}$) are located in between. These bands cover the wavelength range from 200 nm to 500 nm which is of interest for the Endcap DIRC optics. These color centers will also cause fluorescence background at several wavelengths in the VIS range.

In conclusion, the optical system of the Endcap DIRC can be designed by using synthetic fused silica, Ce-doped glasses and single alkali-halide crystals (LiF, CaF_2). Other materials do not qualify due to the strong radiation induced absorbance paired with the long required light path inside the material.

5.5. Photosensors (C3)

has impact on	*key properties*
• Imaging optics • Front-end electronics • Hit rate	• Rate capability • Spatial and time resolution • Spectral response • Efficiency • Lifetime

Cherenkov applications require the detection of light at lowest intensities, equivalent to the detection of single photons, on a large active area, typically in the order of thousands of cm^2. The energy of these optical photons ($E_{ph.} = hc/\lambda$) ranges from 2 eV at 600 nm to 6.2 eV at 200 nm. Photons in this regime deposit energy either via the photoelectric effect or by inelastic scattering involving phonons. Scattering processes are not relevant in photon sensing due to the lower cross section and the fact that the involved phonons are by no means easier to detect than the incident photon itself. Hence, all sensors for low light imaging have to exploit the photoelectric effect.

The term *photoelectric effect* describes the absorption process of a photon by a bound electron which is then excited to a higher energy level in an atom, molecule,

or energy band. Due to energy conservation, the difference between the electron energy levels before and after absorption of the photon has to match the photon energy $E_{\text{ph.}}$. To detect a wide, continuous range of wavelengths it is therefore necessary to exploit transitions in or between energy bands of solid state materials, e.g. metals and semiconductors, as these provide a quasi-continuous set of energy levels. Actually, most materials which are commonly used to convert optical photons to electrons are semiconductors.

An electron excited by an incident photon is called *photoelectron*. If its energy exceeds the vacuum level, it has the ability to exit the material at the boundary surface (Fig. 5.29). This process is named *external photoelectric effect* and related detectors are labeled to be *photoemissive*. The prefix external has been chosen to discriminate this process from the case where the electron remains inside the material, entitled *inner photoelectric effect* (Fig. 5.26). In such detectors, the resulting electron-hole pair is available as free charge carrier and the device is referred to as being *photoconductive*.

In both cases, the resulting electron charge has to be multiplied to generate a measurable output pulse. The typical amplification needed for low noise single photon detection is in the order of $10^5 - 10^6$. The amplification factor is referred to as *gain* of the sensor.

Photodetectors can be divided in three categories depending on the type of charge multiplication they use (Tab. 5.8). Photoconducting solid state sensors exploit impact ionization inside a specifically designed semiconductor device as multiplication mechanism. Vacuum based photodetectors are of the photoemissive type and multiply the photoelectrons by accelerating them onto secondary emission layers. A third type is represented by the gaseous photodetectors which are also photoemissive, but exploit impact ionization in the gas volume for charge multiplication.

Sensor Category	Conversion Type	Multiplication Process
Solid state	photoconducting	impact ionization in the bulk
Vacuum based	photoemissive	secondary emission layers
Gaseous	photoemissive	impact ionization in the gas

Table 5.8.: Categories of eligible photodetectors.

Technical details, advantages and drawbacks of the first two detector classes

are presented in the following sections, followed by a section dedicated to the selection of possible sensor options for the Endcap DIRC.

A discussion of gaseous sensors has been omitted because the established versions use photosensitive gas mixtures or photocathodes which are only sensitive to UV light. As already discussed in section 5.4, the UV range has the drawback of increased optical dispersion and scattering loss. Novel, gaseous PMTs with solid alkali photocathodes sensitive to visible light are under development [24, 21] but can not yet compete with existing vacuum based sensors.

5.5.1. Solid state sensors

This section outlines the function, basic properties and available sensor options in the field of solid state photosensors. One specific detector type, the Visible Light Photon Counter (VLPC) [143], has been excluded from the discussion as it needs extreme cooling down to temperatures below 10 K and is therefore impractical for most applications.

Photodiode

One of the simplest solid state sensors is the p-i-n photodiode (Fig. 5.26). This semiconductor device consists of a slightly doped near-intrinsic region (i) between an n- and p-doped region. An applied reverse bias voltage V_B leads to the formation of a large depletion region of a thickness which is to first order defined by the thickness of the intrinsic layer. The diode does not conduct in this state. However, the resulting internal electric field leads to a drift current if electron-hole pairs are generated inside the depletion region. This way, the inner photoelectric effect causes a drift current which is proportional to the photon flux. Because a p-i-n photodiode does not provide an internal mechanism for charge multiplication, its application is limited to photon fluxes which lead to a sufficiently high current.

Avalanche Photodiode

The sensitivity of the diode can be increased by adding a multiplication region to the p-i-n structure. A common choice is the use of an additional, highly doped p-layer to create a very high field over a small p-n junction (Fig. 5.26). This field has to be strong enough to accelerate the electrons entering from the intrinsic layer so that they reach sufficient energy to excite other electrons by impact ionization.

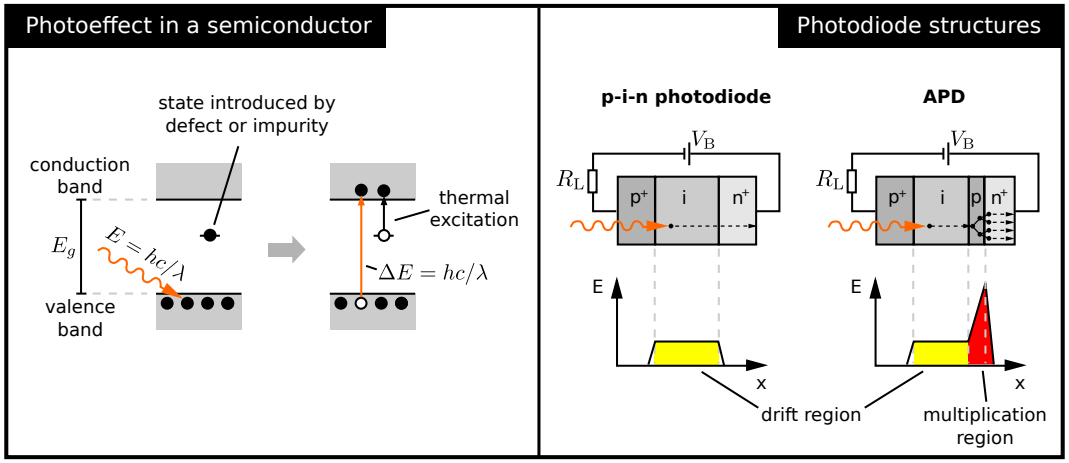


Figure 5.26.: The left pane illustrates the inner photoeffect where an incident photon gets absorbed by an electron in the valence band which is in turn excited to a level in the conduction band. This process is only possible if the photon energy exceeds the energy of the bandgap E_g. A concurrent process is the thermal excitation of electrons to the conduction band. Defects or impurities in the semiconductor can introduce new states closer to the conduction band. These states have a much higher probability of being thermally excited. The right pane illustrates the structure of common sensors, the p-i-n diode and the avalanche photodiode (APD), which has an additional junction with high field to cause charge multiplication by impact ionization. Photons absorbed in the thick intrinsic layer create electron-hole pairs which drift in the electric field and contribute to the current. The shown structure is "n-on-p", which works well in the red part of the spectrum. By exchanging p and n, one obtains a "p-on-n" structure which is better suited to detect blue photons.

These secondary electrons are again accelerated and generate more charge carriers what effectively results in an avalanche effect. With this technique a gain up to about 10^4 is commonly achieved. Due to the nature of the multiplication process, this class of devices is called avalanche photodiode (APD). The gain of these devices is still not sufficient for photon counting applications.

Geiger Mode Avalanche Photodiode

To achieve the necessary single photon sensitivity, APDs can be further optimized to be operated at a reverse bias voltage which exceeds the diodes breakdown voltage. These devices are then called single photon (SPAD) or Geiger-mode

avalanche photodiodes (G-APD). Due to the high bias voltage, a single photo-generated electron-hole pair initiates a self-sustaining avalanche breakdown in the multiplication region which leads to a large current. The high sensitivity comes at the price of an additional quenching mechanism which has to be implemented to interrupt this avalanche process.

The simplest solution, named *passive quenching*, is realized by placing a resistor R_q in series with the diode. The high diode current at the beginning breakdown leads to an increased voltage at the resistor while the voltage over the diode is reduced. This results in a lower internal field and the avalanche is quenched. More sophisticated approaches can be designed by using active circuit elements. These *active quenching* techniques have usually the advantage of lower dead time.

After quenching, a G-APD needs a certain time τ to recover the internal field. This time depends on the device capacitance C_cell and the design of the quenching circuit. In case of passive quenching, the recovery time is simply given by the RC-time ($\tau \approx C_\mathrm{cell} \cdot R_\mathrm{q}$). As the capacitance of an G-APD is roughly proportional to its area $C_\mathrm{cell} \propto A$, an increase in active area A leads to a longer recovery time τ. This size dependence of the recovery time limits the practical size of a single G-APD. Typical G-APDs are much smaller than $0.5\,\mathrm{mm}^2$ [90]. For larger devices, τ becomes so large that the intrinsically high thermal noise would effectively occupy the device at room temperature.

Silicon Photomultiplier (SiPM)

Due to the size limitation, larger G-APD based sensors have to be constructed as arrays of hundreds or thousands of small G-APD cells including corresponding cell electronics. While the G-APD itself is a binary device, high density G-APD arrays can again deliver a proportional* response to multiphoton events [90]. In the high energy physics community such devices are entitled *silicon photomultiplier* (SiPM). The gain of these devices is in the order of 10^5–10^7.

Today, a variety of interesting products is offered by several manufacturers, e.g. Hamamatsu Photonics K.K. (HPK), Philips Digital Photon Counting (PDPC), Photonique SA, SensL and Zecotek Photonics. Some of these, e.g. the HPK Multi-Pixel Photon Counter (MPPC) [150], are fully analog G-APD arrays as described above. Others, like the PDPC digitial silicon photomultiplier (dSiPM), provide sophisticated active quenching circuits and readout electronics on the

*this is only true for low count rates where pile-up effects are not significant (eq. 5.40)

same microchip [42]. Many of the available products are designed to be arranged in larger tiled patterns.

This technology is a very interesting candidate for Cherenkov applications like the Endcap DIRC as it offers single photon sensitivity in combination with a large, scalable active area, small size and insensitivity to magnetic fields. Furthermore, the small feature size of the single G-APD cells and the resulting low cell capacitance permits the design of detectors with excellent spatial and time resolution. However, there are also significant disadvantages which will be addressed in the following section.

The *photon detection efficiency* (PDE) of a sensor is the probability that an incident photon of a given wavelength λ is detected. The PDE of an SiPM depends on the geometric efficiency ϵ_{geo}, which is the ratio of the active and inactive area of the device, the quantum efficiency QE, which is the probability for a photoelectron or electron-hole pair being created, and finally the probability $P_{\text{avalanche}}$ that the charge carrier triggers an avalanche:

$$\text{PDE}(\lambda, V_{\text{B}}, T) = \text{QE}(\lambda) \cdot \epsilon_{\text{geo}} \cdot P_{\text{avalanche}}(\lambda, V_{\text{B}}, T) \tag{5.38}$$

The PDE is not only a function of the wavelength but also of the bias voltage V_{B} and the temperature T [90]. PDE curves for an Hamamatsu MPPC and a Philips dSiPM are shown in Fig. 5.30. Assuming that pile-up is negligible, the expected number of detected photons N_{exp} for a given number of incident photons N_{photons} is then given by

$$N_{\text{exp}} = \int N_{\text{photons}}(\lambda) \cdot \text{PDE}(\lambda)\, \mathrm{d}\lambda. \tag{5.39}$$

As explained above, a G-APD cell on a SiPM can only detect a single photon at a time. After the hit the cell is insensitive until it has been quenched and at least partially recharged. At higher rates, this leads to pile-up effects and hence to a nonlinear detector response. The number of detected photons N_{det} becomes smaller than the expected number of photons N_{exp}. This effect can be described using Poisson statistics. For pulses shorter than the recovery time, the number of registered hits is scaled by the probability of having $N \geq 1$ hits in a single cell with an average number of $N_{\text{exp}}/N_{\text{cells}}$ hits ($P(N \geq 1) = 1 - P(0)$):

$$N_{\text{det}} = N_{\text{cells}} \left(1 - e^{-N_{\text{exp}}/N_{\text{cells}}}\right) \tag{5.40}$$

This equation demonstrates that the dynamic range and linearity of an SiPM depend strongly on the granularity of the device. While a higher number of cells

yields better linearity and wider dynamic range, it also worsens the geometric efficiency ϵ_{geo} for most* devices. This is simply a consequence of the necessary gaps between individual cells. Hence, the granularity has to be a trade-off between efficiency and dynamic range.

To be correct, the value of N_{exp} in eq. 5.40 should also include the contribution of dark count and optical crosstalk, quantities which will be discussed in the following section. For continuous operation or longer pulses, the recovery time τ of individual cells has to be taken into account.

Disadvantages of SiPMs

The two primary drawbacks of SiPMs are the high thermal noise and the sensitivity to radiation damage which leads to a further noise increase or even a fatal failure of single cells [105, 91, 33, 148]. The noise of photodetectors which occurs even when the device is not illuminated is usually named *dark count*. For SiPMs, the typical dark count rate is in the order of MHz/mm^2 [40].

The major reason for dark count is the thermal excitation of electrons in the intrinsic region. Such an electron leads to an avalanche and effectively produces a signal which is indistinguishable from the pulse generated by a converted photon. The dark count rate (DCR) depends strongly on the temperature as well as on the number of available charge carriers which can be thermally excited. Due to the silicon bandgap of 1.11 eV at 300 K [127], these carriers originate mainly from donor levels. Ionizing radiation can induce additional, donor-like states inside the semiconductor, e.g. by lattice displacement, which can be thermally excited and contribute to the dark count rate.

Lattice defects can also form states where electrons become trapped for a certain amount of time τ' and are then reexcited (trapping center, [127]). If these traps become filled during an avalanche and $\tau' \gtrsim \tau$, the decay of the traps can trigger consecutive avalanches. This phenomenon is commonly known under the term *afterpulsing*. Again, the concentration of this kind of lattice defects can be increased by radiation induced defects. The high sensitivity of these devices makes them especially sensitive to this kind of defects.

Another drawback is a phenomenon which is typically referred to as *optical crosstalk*. The recombination of charge carriers in a single G-APD cell can lead

*There are some architectures which are less sensitive to this problem, e.g. the Zecotek Microchannel APD [113].

to the emission of photons. These photons have the ability to either trigger a neighboring cell or to be emitted from the bulk. In the latter case, they can even hit a different SiPM in the detector assembly. In modern SiPMs, the optical crosstalk between individual cells is usually minimized by separating the cells using structures like trenches in the Si-wafer.

As outlined above, the high dark count rate and especially the sensitivity to ionizing radiation of the SiPMs are intrinsic problems of the underlying technology. The dark count rate can be minimized by cooling the devices down to temperatures as low as $-20°C$. As a rule of thumb, the DCR changes by a factor of 2 every 8°C [108]. However, due to the operation principle of the G-APD it is unlikely that a single SiPM-cell can be realized in a radiation-hard fashion. Radiation induced lattice defects cannot be avoided while the device has to be sensitive to single electrons. In conclusion, any application of analog SiPMs in an environment with higher radiation levels has to deal with high dark count rates.

Digital SiPM

Nearly all SiPMs deliver an analog output signal which corresponds to the sum of the individual cell signals (Fig. 5.27, top right). These signals have to be processed by additional front-end electronics. In contrast to this approach, PDPC has developed a fully digital SiPM (dSiPM) by integrating the G-APDs in a conventional CMOS process and connecting the output of each individual cell to a dedicated digital electronic block (Fig. 5.27, left). The detection of a photon directly generates a digital trigger signal. Subsequently, the cell is actively quenched and the hit information remains temporally stored in the cell electronics. Cells can be recharged by toggling a recharge line. This makes the photon detection fully digital down to the level of a single cell.

This approach opens up interesting possibilities. Each cell circuit is equipped with a 1-bit SRAM which is used to activate or deactivate a single G-APD cell. PDPC makes use of this technique to reduce the dark count rate of their devices. The initial dark count rate is usually dominated by a few noisy pixels. By masking these "hot" pixels, e.g. switching them off, the overall DCR of the device can be reduced significantly.

The current PDPC solutions utilize a counter and a time-to-digital converter (TDC) to determine the arrival time of the first photon together with the total number of fired cells in a configurable time window. Therefore, the trigger signal

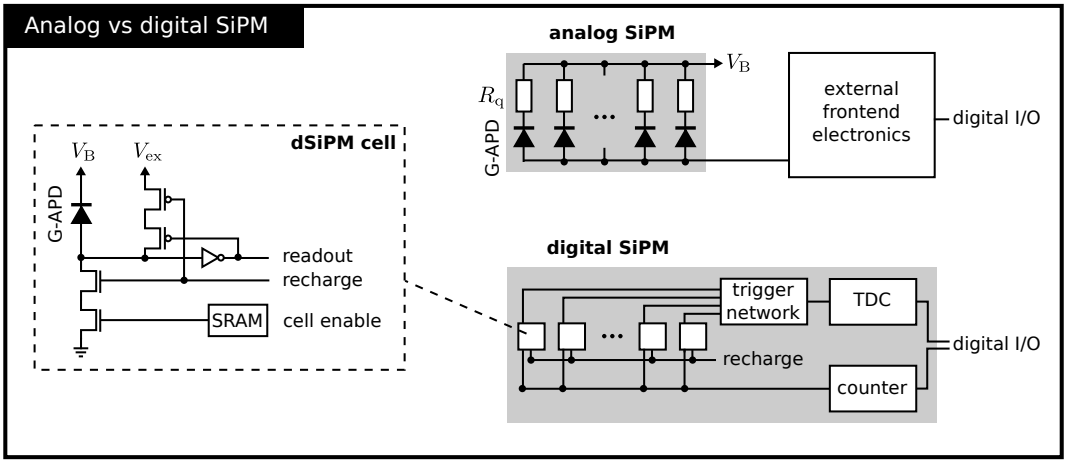


Figure 5.27.: Illustration of the digital SiPM (dSiPM) concept by PDPC in comparison to a conventional analog SiPM. The figures are adopted from various illustrations in [42].

of each individual cell is connected to a global counter via a synchronous bus and to a global TDC via a configurable trigger network [42] (Fig. 5.27).

At the time of writing, two dSiPM variants are on the market. These are functionally equivalent but differ in the size of the single G-APD cells and related parameters. Both ASICs have a die size of about $7.2 \times 7.9\,\mathrm{mm}^2$ and provide 4 logical pixels with an area of $3.2 \times 3.88\,\mathrm{mm}^2$ each. The DPC-6400-22 ASIC comprises G-APD cells with an area of $59.4 \times 32\,\mathrm{mm}^2$, resulting in 6396 cells per pixel with a fill factor of 54 %*. The DPC-3200-22 ASIC uses twice the cell width, resulting in 3200 cells per pixel with a fill factor of 78 %.

The ASICs have been optimized to provide a good geometric efficiency. Bonding pads are only at two opposite sides of the die, so these ASICs are two side buttable. A good example for the tiling capability are the tiles offered by PDPC. These consist of a grid with 4×4 ASICs (8×8 pixels) where 75 % of the area is covered by logical dSiPM pixels.

As shown in Fig. 5.30, the PDE of a DPC-6400-22 pixel is similar to the PDE of an analog MPPC if corrections for optical cross talk and afterpulses are applied. The dSiPM data has been obtained for individual cells by deactivating all other cells on the chip. Hence, optical crosstalk is excluded. According to PDPC, the geometric efficiency of the pixel is already included in the PDE data.

*The fill factor values in [42] have been accidentally swapped.

Despite the non-trivial signal routing through the trigger network, dSiPM sensors are reasonably fast. A single photon time resolution of 61 ps RMS has been measured using a coincidence setup of two DPC-6400-22 [42]. This value includes all uncertainties implied by the ASIC electronics.

As current dSiPMs are optimized to readout scintillators, they do not live up to their full potential regarding imaging applications. In principle, the ASICs can be improved to also deliver spatial coordinates on the single cell level [42]. In addition, the acquisition cycle can be modified to deliver single photon data at a relatively short dead time in the order of hundreds of ns. However, such ASICs are not yet on the market.

5.5.2. Vacuum based sensors

Long before the advent of sensitive semiconductor detectors, light sensors were realized as vacuum tubes. These consist of a vacuum envelope usually made from glass or ceramics with a transparent entrance window (Fig. 5.28, left). A photocathode applied inside the vacuum converts incident photons to electrons. These photoelectrons are accelerated towards the anode of the tube by means of an external applied voltage, resulting in a small anode current. Such a device is also known as *vacuum photodiode*. Electron multipliers can be incorporated in the tube to reach single photon sensitivity. This type of detector is widely known as *photomultiplier tube* (*PMT*).

The spatial and time resolution of a PMT depends on the structure of the electron multiplier as well as the anode design. The electron multiplier can imply changes to the electron trajectories and increase the *transit time spread* (TTS) of the signal. Typical anodes consist of a metal coating connected to the outside via a vacuum feedthrough. Such anodes can be structured to provide a position sensitive readout of the charge cloud which exits the electron multiplier. Anodes can also be instrumented with semiconductor devices like APDs or Si-pixel detectors. Detectors with active anode elements are commonly entitled *hybrid photodetector*.

Photocathodes

The emission of a photoelectron from the photocathode can be well described by the Spicer three step model [124]. The separate steps:

(1) *absorption:* photon absorption inside the cathode layer

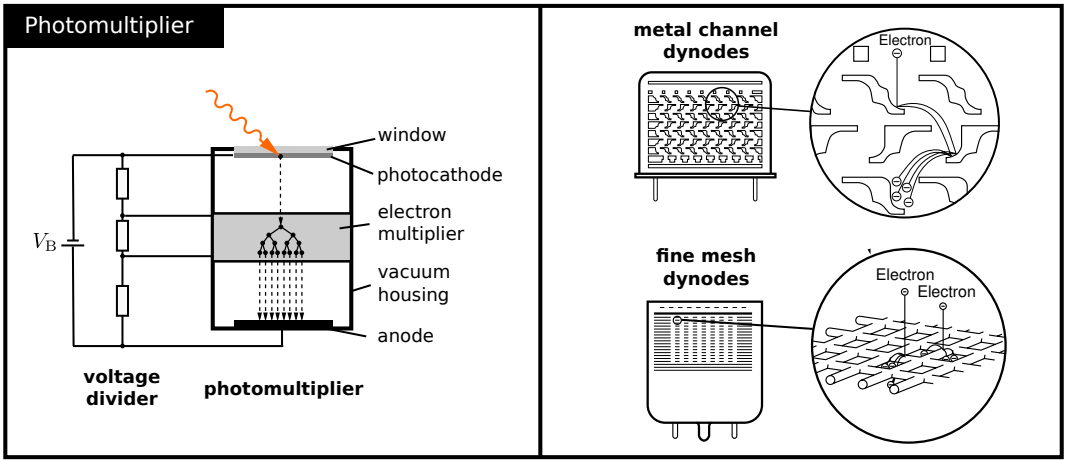


Figure 5.28.: The left figure shows the basic setup of a photomultiplier. It consists of a vacuum tube with an entrance window, a photocathode, an electron multiplier which is based on materials with high secondary emission yield and an anode to collect the charge. A high voltage, usually in the order of several kV, is needed to accelerate the electrons from the cathode successively over the individual dynodes down to the anode. The right figure shows two selected dynode structures. These figures have been adapted from [3]

(2) *diffusion:* electron transport to the vacuum boundary

(3) *emission:* electron escape into the vacuum

are labeled in Fig. 5.29, left, which illustrates the external photoelectric effect in a standard semiconductor photocathode like the widely used bialkali types.

As the absorption of electromagnetic radiation is a bulk effect, the resulting photoelectron has to diffuse to the geometric boundary of the photocathode layer to be able to leave the material. During this diffusion process the electron loses energy, e.g. by electron-phonon or electron-electron scattering with an effective scattering length l_{scat}. After reaching the boundary, the electron can only escape if its remaining energy exceeds the vacuum barrier. Hence, the electron diffusion process limits the depth d_{max} at which a created photoelectron can still exit the photocathode. This depth limit has a direct impact on the quantum efficiency as the fraction of absorbed light in a layer of thickness x is $\propto e^{-x/l_{\text{abs}}}$ and $x \leq d_{\text{max}}$.

Based on the three steps listed above, Spicer derived a comprehensive parametrization of the internal quantum efficiency, e.g. the number of emitted photoelectrons

per absorbed photon:

$$\mathrm{QE_{int}} = (1-R)\mathrm{QE} = \frac{\alpha_{\mathrm{PE}}}{\alpha}\left(1+\frac{l_{\mathrm{abs}}}{l_{\mathrm{scat}}}\right)^{-1} P_{\mathrm{E}} \tag{5.41}$$

where R is the reflectivity of the photocathode and $\frac{\alpha_{\mathrm{PE}}}{\alpha}$ denotes the ratio of excited electrons which have an initial energy above the vacuum level. $\frac{l_{\mathrm{abs}}}{l_{\mathrm{scat}}}$ is the ratio of the absorption length of the photons and the scattering length of the excited electron during diffusion. P_E is the probability for an electron at the geometric boundary and with an energy above the vacuum level to be emitted. The value of P_E is typically < 0.5. All of these parameters are a function of the incident photon energy.

Obviously, the quantum efficiency can be increased by either increasing $\frac{\alpha_{\mathrm{PE}}}{\alpha}$, e.g. by lowering the vacuum level, or by decreasing $\frac{l_{\mathrm{abs}}}{l_{\mathrm{scat}}}$, e.g. by increasing the scattering length. The first option is actually used to produce negative electron affinity (NEA) photocathodes (Fig.5.29, right). Typical materials are GaAs or GaAsP. These photocathodes are then activated by an additional Cs or CsO monolayer, which effectively lowers the vacuum level below the lower edge of the conduction

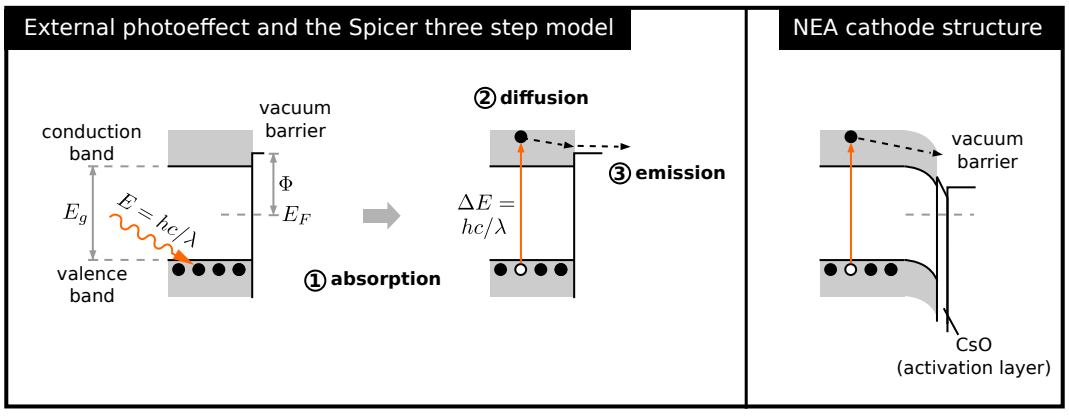


Figure 5.29.: Left: the three steps of the Spicer model, describing the external photoelectric effect in a standard semiconductor photocathode. The absorption of a photon (1) in the bulk can excite an electron to the conduction band where it may diffuse to the geometric surface of the cathode (2). At the surface, the electron can be emitted with a certain probability (3) if its remaining energy exceeds the vacuum level. Right: a special high-sensitivity photocathode with an additional CsO layer which lowers the vacuum energy. In this case, the probability of step 3 is vastly improved and thus the quantum efficiency.

band inside the bulk. The resulting $\frac{\alpha_{\mathrm{PE}}}{\alpha}$ becomes almost unity what leads to highly efficient photocathodes. As the low vacuum level permits more thermally excited electrons to be emitted to the vacuum, NEA photocathodes cause a higher dark count rate than conventional photocathodes.

Spicer showed by means of his model that the typical photoelectron yield for metal photocathodes is several orders of magnitude lower than the yield for semiconductors and insulators. The best yield can be reached by using NEA cathodes [124].

Commonly used photocathode materials are alkali antimonides like K-Sb-Cs (a.k.a. Bialkali) or Sb-Na-K-Cs (a.k.a. Multialkali), NEA cathodes based on GaAs, GaAsP and InGaAs as well as solar blind cathodes like Cs-Te or Cs-I. The wavelength dependent quantum efficiency of some selected photocathodes is shown in Fig. 5.30 (scaled by 70 % to account for collection efficiency) in comparison to the PDE of two SiPMs.

Fig. 5.30 also shows enhanced Bialkali photocathodes developed by HPK which they named Super- and Ultrabialkali. These photocathodes show an enhanced crystallinity [92] and an improved quantum efficiency. While typical Bialkali photocathodes show a peak-QE of about 25 %, the Super- and Ultrabialkali types peak at 35 % and 43 % respectively. This increase in sensitivity is assumed to be a direct consequence of the improved crystallinity, as this property is related to the scattering length. Beside HPK, Photonis has also shown improved Bialkali cathodes [59]. Unfortunately, the availability of these improved Bialkali types depends on the tube in question as the adaption of the process to different tubes is non-trivial [27].

The performance of photocathodes is extremely sensitive to changes in the chemical composition and material structure [28]. Hence, diffusion of foreign atoms into the cathode layer can lead to a degradation of quantum efficiency. Such chemical poisoning can be caused by residual gases in the vacuum as well as materials in the cathode substrate or ions emitted by electron stimulated desorption. This chemical poisoning causes the decay of quantum efficiency in current PMTs based on microchannel plate electron multipliers which will be introduced in the following section.

Photocathode configurations as shown in Fig. 5.28 are named *transmission-type photocathode* because the front side is illuminated while the photoelectrons are extracted at the backside. This configuration where the cathode is applied directly on the front glass of the tube is the most common in position sensitive detectors.

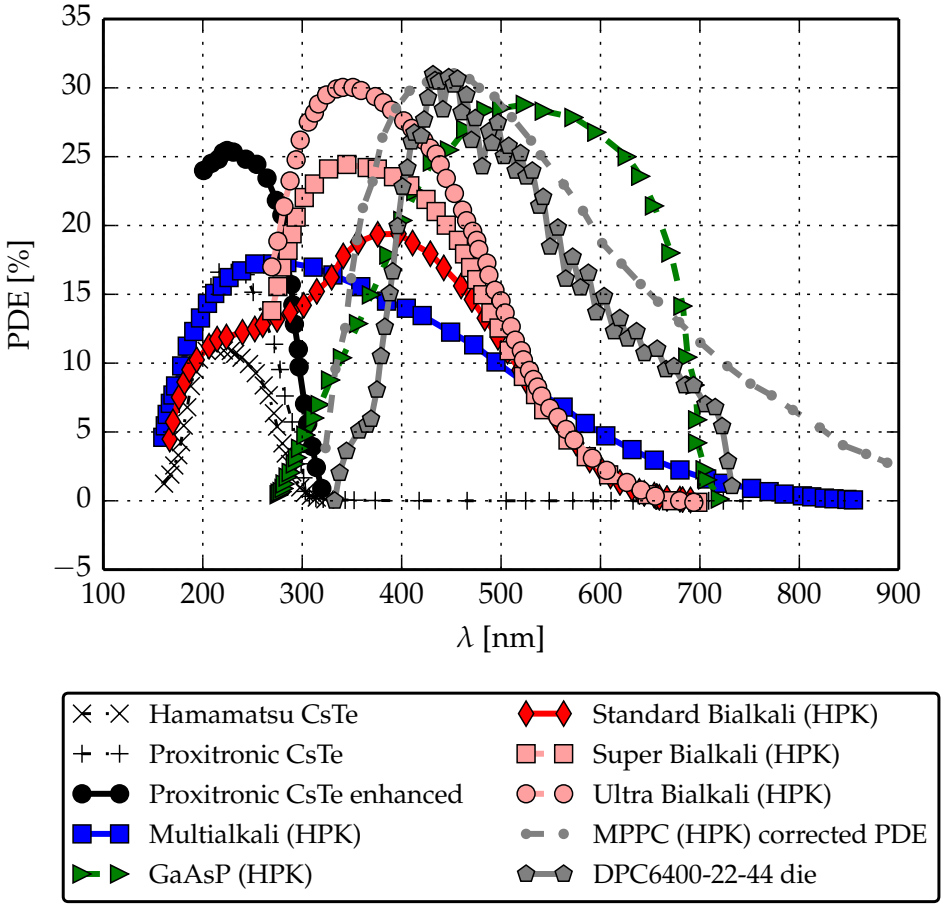


Figure 5.30.: Comparison of the expected photon detection efficiencies of an MCP-PMT with 70 % collection efficiency and different photocathode materials. The PDE of two SiPM sensors, a Hamamatsu MPPC and a Philips dSIPM, are shown for comparison (gray). The PDE of the MPPC has been taken from [137] where the authors applied corrections to account for the contribution of afterpulses and darkcounts. The other data has been taken from datasheets of the individual manufacturer.

In a second configuration, the *reflection-type photocathode*, photoelectrons are extracted at the illuminated side. The advantage is a slightly improved quantum efficiency, but the setup is not well suited for the precise determination of the photon impact point what is a crucial requirement in imaging applications. Reflection-type photocathodes which are coated on metals also show higher saturation currents than transmission type photocathodes due to the lower resistance of the metal

substrate. However, in most applications the saturation of the photocathode is not a limiting factor.

Electron multipliers

Electron multipliers in PMTs are based on a phenomenon called *secondary electron emission*. This phenomenon is closely related to the photoemission process which has been described in the preceding section. Secondary electron emission can be modeled by a similar three step model where the first step is replaced by the "generation of internal secondary electrons due to energy loss of the primary particle" [119]. It is therefore no surprise, that the alkali antimonide materials used for photocathodes are also excellent secondary electron emitters. Materials like PbO, BeO, MgO are also commonly used [3].

To reach a good secondary electron yield, the incident electron has to be accelerated towards the secondary emissive material, which is usually called dynode. Typical yield values are in the order of 10 electrons, hence the process has to be repeated several times to reach an effective gain of 10^5 – 10^7 which is needed for single photon imaging. Due to the statistic nature of the secondary emission process, the yield varies at each multiplication step what leads to gain fluctuations.

Such multiplication structures can be constructed in a variety of ways which differ in collection efficiency, gain, transit time spread, spatial resolution, uniformity and magnetic field resistance. A review of different structures is given on [3]. In the following, the focus will be put on structures usable for large area imaging applications and the influence of an external magnetic field.

Two dynode examples from HPK PMTs, the metal channel dynode and the fine-mesh dynode, are shown in Fig. 5.28, right. The metal channel dynode is a highly optimized, micromachined electron multiplier featuring a short electron path length between the individual dynodes. Further, it uses a focusing mesh between the photocathode and the first dynode. This combination leads to a low spatial spread during multiplication and hence to a small transit time spread as well as a low spatial spread of the output charge on the anode. This dynode type is available in a 2” square tube with Bialkali photocathode (HPK H9500).

Nearly all of the available dynode structures are very sensitive to distortions of the electron trajectories. Such distortions are introduced by external fields such as the high magnetic field inside or in the vicinity of a magnetic spectrometer. The additional gyration of the electrons in presence of higher magnetic fields has

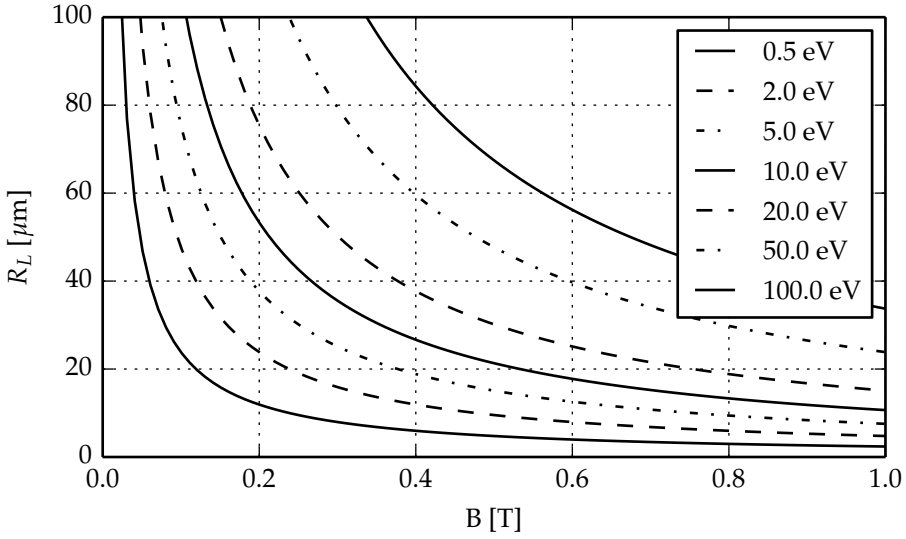


Figure 5.31.: The Larmor radius R_L of electrons at different kinetic energies. The typical transverse energies of the electrons in an electron multiplier are in the range of several to tens of eV.

a strong impact on the gain of the device. Metal channel dynodes can work in magnetic fields up to about 10 mT, depending on the alignment of the tube to the field.

To reach sufficient gain at higher magnetic fields, the feature size of the dynode structure has to be reduced so that it remains compatible with the gyration of the electrons (Fig. 5.31). The fine mesh dynode (Fig. 5.28, right) can be operated in magnetic fields up to ~1.5 T. Specific versions for single photon detection have been developed by HPK in cooperation with the Belle II collaboration [47, 51]. These devices showed a low transit time spread ≤ 100 ps and a spatial resolution in the order of 0.5 mm measured with a field parallel to the tube axis.

To work at even higher fields or to achieve faster timing, a *microchannel plate* (*MCP*) can be used as electron multiplier. Ever since their development in the 1960s, MCPs have been used for electron multiplication in plasma and heavy ion physics experiments. The major commercial application is their use in image intensifier tubes for night vision systems. The working principle is presented in figure 5.32. MCPs are operated in vacuum and consist of a thin, highly resistive substrate with many tiny holes in it. These holes, which are called channels, have a typical diameter of 5 to 40 μm at densities of 10^4 to 10^6 channels per cm^2. For a

given channel diameter, the substrate thickness is defined by the channel length to diameter ratio $\alpha = l/d$ which is an important performance parameter because most of the key properties are a function of α. Commonly used values for α are 40 to 100, resulting in a MCP thickness in the order of 0.4 up to several mm.

The channels have semiconducting walls with a high secondary electron emission yield (SEY) *. Top and bottom surfaces of the substrate are coated with a conducting material resulting in two electrodes. When a high voltage (typically 1 kV) is applied at these electrodes, every channel functions as a single channel electron multiplier. An MCP can thus be considered as a high density array of millions of tiny electron multipliers. A charged particle hitting the channel wall at sufficiently high energy releases one or more secondary electrons which are accelerated along the channel by the electric field between the top and bottom electrodes. These electrons strike the channel wall again, each of them releasing more secondary electrons. This multiplication process repeats until the charge cloud has left the channel.

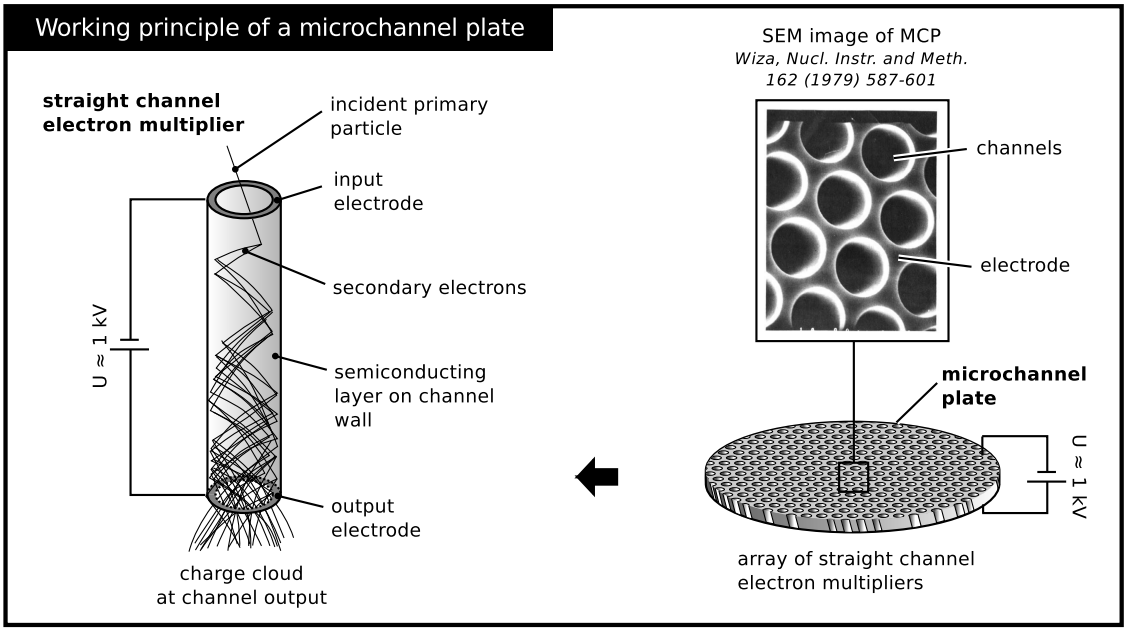


Figure 5.32.: Working principle of a microchannel plate.

According to Wiza [147], the gain of a straight channel electron multiplier is practically limited to 10^3 – 10^5 due to the onset of ion feedback at higher gains. The term "ion feedback" describes the situation when molecules in the residual

*They have to be semiconducting or highly resistive so that the channel can be recharged after electron emission but the electrodes are not short-circuited.

gas get ionized, travel back through the channel, hit the channel wall and release additional electrons due to secondary emission. The influence of ion feedback can be reduced by avoiding a straight path from the ion source (the channel exit) to the channel entrance. This can be either realized by using a curved or twisted channel geometry, which is harder to manufacture, or by stacking two MCPs with a certain bias angle so that the channel direction in the first MCP is in the opposite direction to the channels in the second MCP (Fig. 5.33). In the case of a twisted channel, space charge effects cause a saturation of the gain at about 10^6 for 25 μm pores. In Chevron and Z-stacks, higher gains are possible because the electrons from the first MCP can excite multiple channels in the second MCP. Typically, about 6.5 channels of the second MCP are excited by one channel in the first MCP.

Photoelectrons which do not instantly enter one of the channels have a low probability of being detected. The electron collection efficiency of the MCP can be approximated by the ratio of the area occupied by the holes A_{holes} to the total MCP area A_{MCP}, entitled *open area ratio* (OAR). Pores in a conventional MCP are aligned in a hexagonal closest package. Assuming that the amount of blocked pores is negligible, the open area ratio can therefore be expressed as

$$\text{OAR} = \frac{A_{\text{holes}}}{A_{\text{MCP}}} = \frac{\pi}{2\sqrt{3}} \left(\frac{d_{\text{hole}}}{d_{\text{pitch}}} \right)^2 \tag{5.42}$$

with the hole diameter d_{hole} and the channel pitch d_{pitch}. A typical MCP with 10 μm pores has a channel pitch of 12 μm resulting in an open area ratio of 63 %.

A reduction of the channel pitch leads to thinner channel walls. Hence, the pitch is limited by the required mechanical stability of the MCP. The OAR can be further increased by techniques like funneling, where the channels are etched in a way that one end becomes shaped like a funnel which effectively increases the channels acceptance while maintaining the wall thickness in the rest of the plate. This approach usually leads to an open area fraction in the order of 70 % to 80 %. Recently, even 90 % have been realized using so called tapered channels [77].

MCP-based photomultipliers

The scheme of a typical MCP-based photomultiplier (MCP-PMT) is illustrated in Fig. 5.33. The most common dynode structure is a Chevron setup consisting of two MCPs which provide a combined gain of 10^6 – 10^7. The gain can be increased to $\sim 10^9$ by using a Z-stack configuration with three MCPs, e.g. for tubes with a

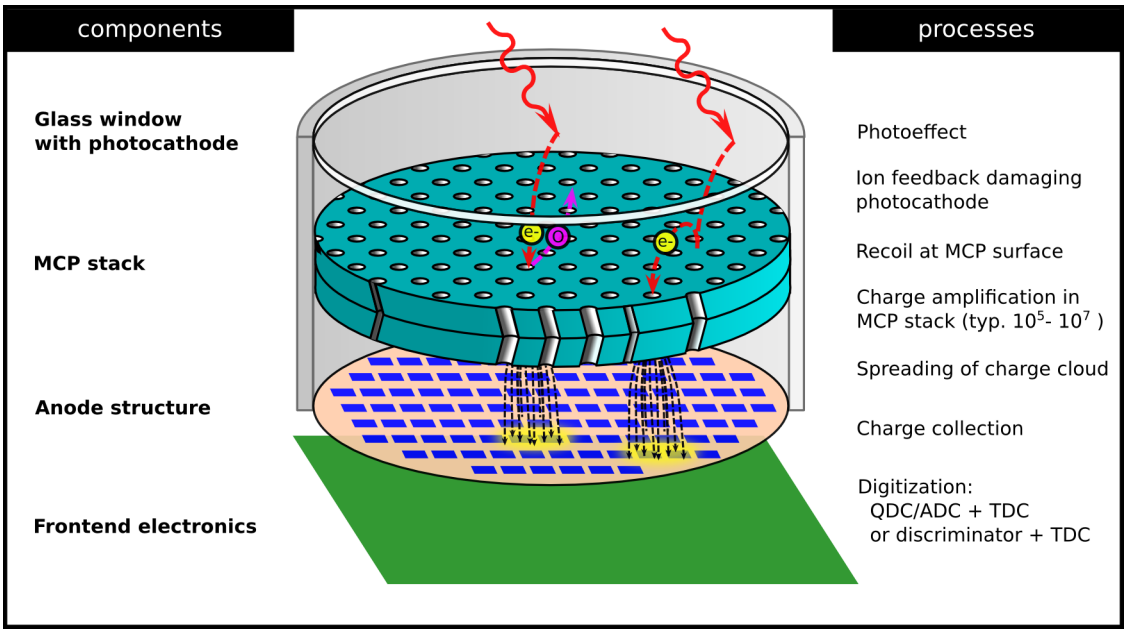


Figure 5.33.: Illustration of a photomultiplier tube with microchannel plate dynode. The components are named at the left while the processes which happen are listed at the right.

capacitive readout [25]. The charge cloud exiting the last MCP is collected by an anode which can be realized in many different ways reaching from a simple metal layer over delay lines to active elements like Si pixel detector ASICs inside the vacuum. These anode options will be discussed in the subsequent section.

The major shortcoming of MCPs as dynode structure is a faster decay of QE during operation. This behavior has been studied by different groups, especially at Belle 2 and PANDA, and reported in [55, 70]. As explained in the preceding section, photocathodes can easily be poisoned by impurities. It is widely accepted that the MCPs show enhanced outgassing during operation what finally leads to gradual poisoning which is then observed by an decay of QE.

Such problems have already been observed a decade ago in MCP-PMTs used for night vision devices. One countermeasure was the introduction of a passivation layer on top of the first MCP which works as an ion barrier [32]. Such a passivation layer has later been introduced by HPK to improve the lifetime* of the SL10 MCP-PMTs developed for the Belle 2 experiment from initially < 0.1 C/cm^2 to ~ 1 C/cm^2

*In this context, the term lifetime had been defined as the extracted anode charge up to the point where the QE has been decreased to 80 % of its initial value.

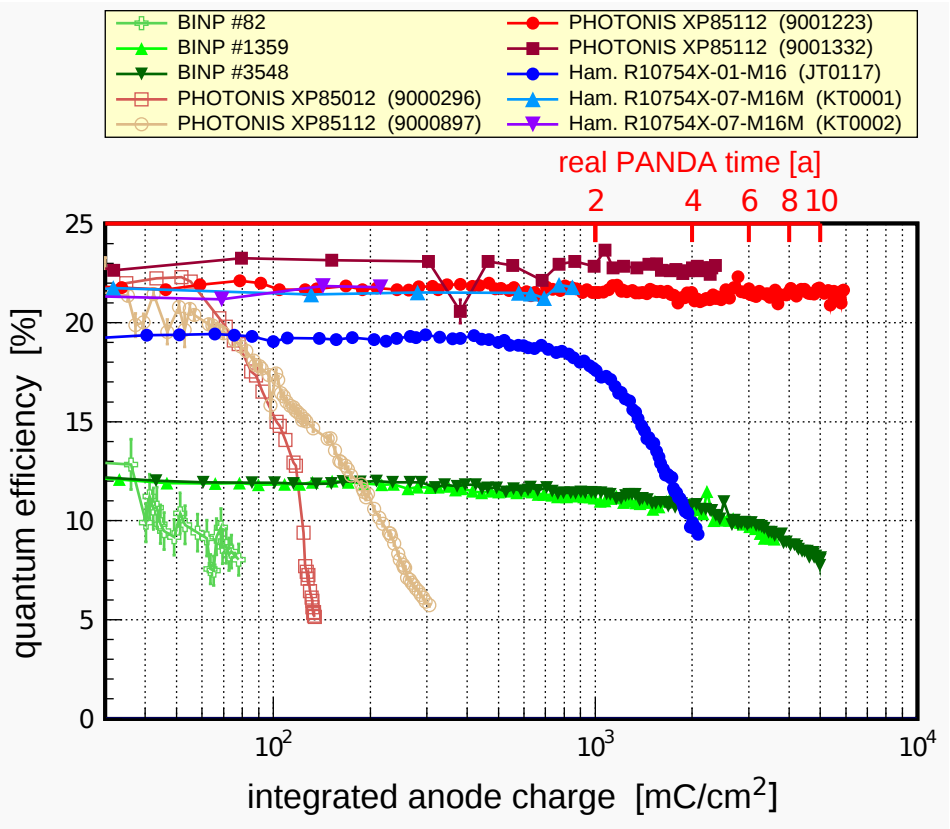


Figure 5.34.: QE decay in miscellaneous lifetime enhanced MCP-PMTs as function of the extracted anode charge Q_{out} for illumination with lower photon rates (typically 1–5 mC/cm^2 per day [22], corresponding to 72–362 kHz/cm^2). Figure from [70].

[55] and later to 2–3 C/cm^2 [86]. The passivation layer at the input of the first MCP comes at the price of a reduced collection efficiency of 37 % instead of 65 %. However, it has been shown in [86] that the layer is also effective if it is placed at the input of the second MCP while the collection efficiency is only slightly reduced to 60 % instead of 65 %.

A second, even more efficient approach is given by the possibility to apply functional coatings on the MCP via atomic layer deposition (ALD), a process developed by Arradiance Inc. [134, 128]. This technique allows to coat not only the MCP surface but also the inside of the pores. Hence, the ALD process can further be used to activate MCPs made from any substrate material, e.g. silicon [17] or borosilicate glass [121]. New substrates open up a wide area of future possibilities, such as the integration of electronic devices directly within the MCPs.

Lifetime measurements of PMTs featuring ALD-enhanced MCPs have been

reported in [70]. An ALD–enhanced Photonis Planacon XP85112 survived up to an extracted charge of $\gtrsim 6\,\mathrm{C/cm^2}$ without a noticeable QE degradation (Fig. 5.34). HPK tubes with ALD coating are also under test, but the integrated charge has not yet reached values $> 1\,\mathrm{C/cm^2}$. While not definitely proven, it is assumed that the homogeneous ALD coating leads to a significant reduction of outgassing of the MCP during operation what in turn diminishes the chemical poisoning of the cathode.

Both coating techniques address the outgassing of the MCP and the related ion backflow. Another handle to lifetime enhancement is the option to lower the photocathodes sensitivity to chemical poisoning. The Budker Institute of Nuclear Physics (BINP) has taken this route and successfully decreased the QE decay of their MCP-PMTs [12] at the price of an increased dark count rate. Fig. 5.34 shows the difference between two tubes with enhanced photocathode (BINP #3548 and BINP #1359) and a tube without enhancement (BINP #82). It has to be mentioned that the peak QE of the BINP photocathode is about 20 % at $\lambda \approx 480\,\mathrm{nm}$ while the QE in Fig. 5.34 has been measured at a wavelength of 372 nm.

Beside these aging issues, MCP-PMTs have many advantages over other PMTs. Intrinsically, MCP-PMTs show excellent imaging and timing properties. Due to the high granularity of the MCP, the image at the MCP output closely resembles the image at the input side. Hence, the contribution of the electron multiplier to the overall image distortion is negligible. This is also the reason why night vision devices use MCPs. The major contributions to image distortion are effects which increase the uncertainty of electron trajectories from the point of photon absorption to the MCP pore and from the MCP output to the anode (compare Fig. 5.33):

- Due to diffusion, the angle between the photoelectron momentum and the cathode normal has a random component.
- Photoelectrons which do not enter a channel at the first strike can recoil and enter a different pore.
- More than one channel is excited in a subsequent MCP.
- The momentum spread of the exiting charge cloud and the coulomb force causes the electrons to drift apart on their way to the anode.

All these effects contribute to a blurring of the image at the anode. The maximum displacement of electrons which recoil at the first strike can be minimized by

reducing the input gap, e.g. the gap between the photocathode and the first MCP. A reduction of the output gap between the last MCP and the anode structure can be used to limit the spreading of the charge cloud. These "proximity focusing" techniques are commonly used in image intensifiers for night vision devices. Fig. 5.35 summarizes some common modifications used to enhance the imaging properties as well as some anode options.

The footprint of the charge cloud on the anode is also significantly reduced by strong magnetic fields in the order of 1 T [65]. In this case, the electron trajectories curl around the magnetic field lines what effectively limits the expansion of the charge cloud.

As the electron trajectories do not differ significantly, the transit time spread of an MCP-PMT is very small, typically in the order of a few 10 ps RMS for a single photoelectron. This makes the MCP-PMT also an excellent choice for timing applications. However, these superior properties can only be exploited if the anode structure has a sufficient granularity and does not introduce additional signal distortion. An overview of possible solutions is given in the subsequent section.

Position sensitive readout techniques

The anode is one of the most important components of the PMT as it has to pick up the signals while minimizing any further signal distortion. Most manufactures offer multiple anodes from stock as well as the option to tailor the anode to the parameters of the customer application. This section will provide a short review of the possibilities based on a selection of representative anode technologies. Some common anode options are shown in Fig. 5.35.

Many image intensifiers as well as intensified CCD cameras (iCCD) use a phosphor screen coated with a thin metal electrode to convert the charge cloud to photons. The incident electrons can traverse the coating and excite the phosphor. Afterwards, the charge is collected by the metal film. Generally, the phosphor is applied to a fiber optic plate (FOP) to preserve the position information. Intensified CCD cameras use a fiber optic taper which is directly coupled to a CCD which records the phosphor image. Such systems can provide a very high spatial resolution in the order of 50 μm. However, the frame based operation of the CCD limits the single photon time resolution to milliseconds. A replacement of the CCD by faster imaging detectors would allow to improve the time resolution up

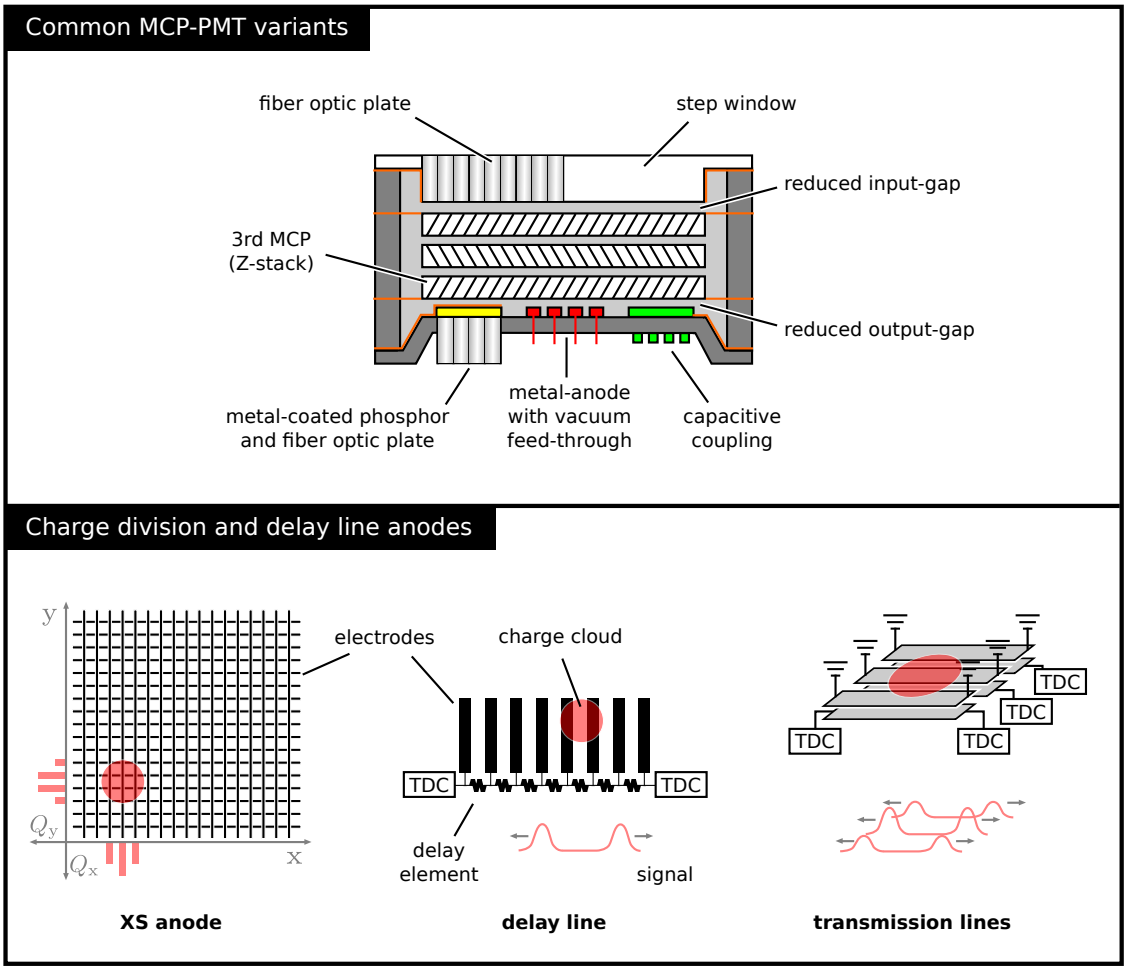


Figure 5.35.: Common variants of MCP-PMT components. Step face windows can be used to reduce the input-gap between the photocathode and the first MCP. Sometimes the window is realized as fiber optic plate (FOP) to minimize image distortion at the photocathode. However, FOPs have a limited transmission of typically 60 to 70 %. To increase the gain, a third MCP can be added at the cost of additional image blur. The charge distribution at the anode can be converted to light by a phosphor layer and transported to the outside via a FOP and detected by an attached CCD. Another common anode type is a metal pattern for charge collection which is coupled to the outside via vacuum feed-throughs, e.g. Kovar pins. Instead of individual feed-throughs, pickup electrodes can be capacitively coupled to an internal metal layer. The latter configuration is also known as resistive sea tube. The lower pane shows some advanced anode designs for sub-mm resolution which are addressed later in the text.

to nanoseconds, depending on the decay characteristics of the phosphor which ultimately limits the time resolution (rise time) as well as the rate capability (decay time).

The most common anode structures used in PMTs comprise either a single or an array of metal electrodes (a.k.a. multi-anode) which are directly connected to the outside of the tube via vacuum-feedthroughs (Fig. 5.35). As these electrodes are read out individually, the spatial resolution is practically limited by the number of necessary electronic channels and vacuum feedthroughs.

A different class of anodes based on charge division structures has been introduced to overcome this limitation. These structures are designed in a way that the charge cloud is spread over several anode elements. The position of the cloud is then computed by the amount of charge deposited on the individual electrodes. Due to this principle of operation, such setups are generally incapable to detect coincident events what limits the rate capability. However, these anodes provide very good spatial resolution at a low number of electronic channels. Usually, charge division anodes do not provide a time-stamp. In this case, a precise event time can be obtained from the output electrode of the last MCP. This results in a time resolution of ~ 0.1 ns FWHM.

An example of a charge division structure is the cross-strip (XS) anode [120, 135] (Fig. 5.35). It consists of two stacked, orthogonal layers of resistive strips collecting the anode charge which is spread across several of these electrodes. After measuring the charge on all strips, the centroid of the charge cloud is computed separately for both layers resulting in a precise 2d-coordinate. Coincident hits are difficult to deconvolve on the fly what limits the rate capability. Early versions of this readout used the signal at the output electrode of the last MCP as trigger and to obtain a time-stamp. This event based approach can be replaced by a more complicated parallel processing which reduces the dead time and hence increases the rate capability. This version does not use the MCP electrode for timing what leads to a lower time resolution in the order of 1 ns.

Beside a concrete charge measurement, one can use the propagation time of an induced pulse in a delay line to compute the position of the charge cloud in one dimension (Fig. 5.35). The second dimension can be measured by an additional, rotated delay line layer similar to the XS-anode. Typical two-layer delay lines are limited to count rates of a few hundred kHz due to the high pile-up probability. A variant of the delay line is the transmission line which can be considered as a continuous delay line with a higher signal propagation speed, e.g. ~ 0.2 mm/ps.

Due to the tough timing requirements, the transmission line anodes are typically read out using waveform sampling ASICs. A prototype readout of a Planacon MCP-PMT with 32 delay lines read out by DRS4 waveform sampling ASICs has been presented in [63]. The reported spatial resolution along the transmission line is 2.8 mm FWHM measured with a scintillator. A transmission line anode for a large scale MCP detector delivered a 2 mm spatial resolution and a time resolution better than 100 ps [41] for single photons. The charge of the pulses can be extracted from the measured waveforms what permits the determination of the charge cloud position orthogonal to the transmission lines by centroiding.

An interesting alternative to anodes with vacuum feed-throughs are capacitively coupled anode structures, an approach developed by Roentdek-Handels GmbH [25]. The vacuum side of this anode consists of a highly resistive layer coated on the ceramic backplane of the tube. Electrodes placed at the air side of this backplane couple capacitively to the resistive layer and can thus be used to measure the charge footprint. As the structured part of the anode is outside the vacuum, it can be modified at any time. An example for such an anode is a three layer delay line which has been demonstrated to deliver hit information with a spatial precision of about 150 μm FWHM and a time resolution of $\sim$ 100 ps FWHM up to a global count rate of 1 MHz.

Anodes can also incorporate active elements like APDs or pixel detector ASICs to measure the charge cloud. Hamamatsu and Belle 2 have developed a single photon sensitive hybrid photodiode with APD readout. The photoelectrons are accelerated by a voltage of several kV before they hit an APD at the anode where they generate multiple charge carriers by impact ionization. The internal amplification in the APD results in a measurable output pulse [66]. As this device does not use any dynode structure, it can be operated in high axial magnetic fields and has been tested up to 1.5 T. The magnetic field even improves the imaging properties as it confines backscattering electrons spatially due to the Larmor precession around the axial field lines.

A second noteworthy example is the readout of a Planacon MCP-PMT using a 2×2 pattern of Timepix 2 ASICs [138]. These ASICs, developed for silicon pixel detectors, consist of 256×256 pixel cells with a pitch of 55 μm. Each pixel has a discriminator with an adjustable threshold of $> 1000\,\mathrm{e}^-$ and can provide a time-of-arrival (ToA) with the precision of the clock, e.g. up to 10 ns LSB. Hence, these ASICs can easily measure the charge cloud of typically $> 10^5\,\mathrm{e}^-$. Alternatively, the ASIC can provide an energy measurement via time over threshold (ToT) but

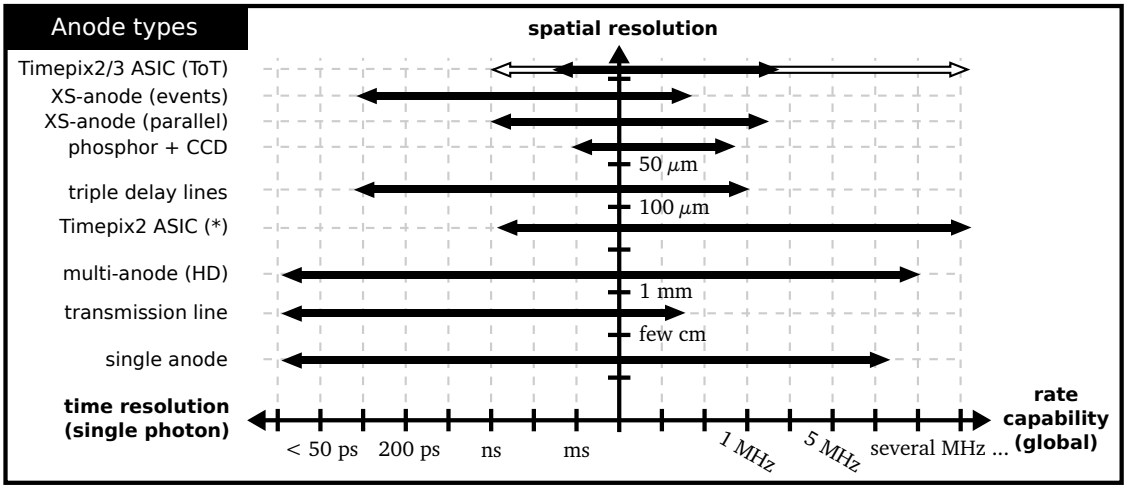


Figure 5.36.: Comparison of selected anode designs discussed in this section. The arrows point to the corresponding coordinates in the left and right axes while the length of the arrow is just to guide the eye. Note that this figure shows a qualitative comparison and the axes have varying scales. The count rate has to be understood as the global count rate for continuous readout. The figure shows that current anode designs do not permit to obtain high spatial resolution at high global count rates. (*): The Timepix 2 readout is frame based with a readout/dead time $\gtrsim 300\,\mu$s what leads to event loss at higher rates. This problem will be addressed in the upcoming Timepix 3 which will provide a data driven readout, improved time resolution and simultaneous time of arrival and time over threshold measurements (white arrow).

without ToA. This mode can be used to determine the centroid of the charge cloud. Measurements presented in [138] show that the Timepix 2 readout has a higher spatial resolution than the intrinsic resolution of the MCP-PMT which has been measured to be about 165 μm FWHM. A drawback of this solution is the frame based readout of the ASIC with a readout time of $> 300\,\mu$s. No signals are detected during this time what leads to a dead time of 23 % at a 1 kHz framerate. However, the next generation of the ASIC will feature a data-driven readout mechanism which will eliminate this restriction. In addition, the time resolution of the Timepix 3 will be improved to a design value of 1.562 ns LSB.

Fig. 5.36 has been compiled to ease the selection of an anode technology. It shows a qualitative comparison of the different anode options in terms of three fundamental requirements of photon counting applications:

- time resolution
- rate capability

- spatial resolution

It is clearly visible that all sub-mm readouts except the high density multi-anode and the Timepix solution are limited to a global count rate in the order of 1 MHz or lower. Applications which require a high spatial resolution but want to avoid a high channel-density have to sacrifice rate capability and time resolution.

5.5.3. Sensor options for the Endcap DIRC

Already the requirement of single photon detection in a magnetic field of about 0.5 T to 0.9 T (Fig. 5.5, page 55) limits the choice of a vacuum based photosensor to three options:

- The HAPD developed by HPK in cooperation with Belle 2
- An MCP-PMT
- A PMT with fine mesh dynode

A second important restriction is the radiation at the sensor area, especially the neutron equivalent fluence of $> 10^{11}$ n/cm^2 which will cause damage to any of the SiPMs. Damaged cells in analog SiPMs will increase the dark count rate of the whole sensor. On the contrary, the Philips dSiPM allows to switch off these "hot" cells, so that the noise increase can be turned into a decrease of efficiency. As displacement damage is localized to a region smaller than the size of a cell, one can assume that this damage will lead to a continuous PDE decay which is proportional to the particle flux. If the rate of this PDE decay is sufficiently low one can also consider the dSiPM as a valid sensor option, assuming that the damage to the logic part of the ASIC is negligible.

For a further comparison of the four viable sensor technologies, key performance parameters of exemplary devices have been collected in table 5.9. All listed package dimensions are compatible with the Endcap DIRC requirements and, except the fine mesh type, provide a good geometric efficiency.

The HAPD has a very low gain and time resolution. This device is read out by a dedicated, triggered ASIC which measures only the charge but not the time of arrival. In the Endcap DIRC, at least a coarse time stamp in the order of 1 ns is required to assign photon hits to incident particle tracks. A higher time resolution in the order of 100 ps or better is envisaged to suppress noise and resolve ambiguities during pattern analysis. Hence, the HAPD in its current state

	HAPD	***MCP-PMT***	***Fine-Mesh***	***dSiPM die***
Peak PDE	24 % (340 nm)	15 % (380 nm)	~ 20 % (400 nm)	30 % (450 nm)
Single photon RMS time res.	–	~ 35 ps	~ 100 ps	~ 60 ps
Active area [mm^2]	62.4 × 62.4	53 × 53	26.5 × 26.5	~ 49.6
Package area [mm^2]	72 × 72	59 × 59	39 × 39	7.8 × 7.2
Package height [mm]	30	24	> 30	~ 1
Geometric efficiency	74 %	81 %	46 %	88 %
Standard pixelization	12 × 12	8 × 8, 32 × 32	24 × 1	2 × 2
Vacuum gap dim. [mm]	~ 20	~ 8	~ 30	0
Radiation hardness	$>10^{11}$ n/cm^2 , >1 kGy	very high	very high	not sufficient
Aging effects	no	yes	no	no
DCR/cm^2 at room temp.	some kHz	some kHz	some kHz	< 40 MHz
Gain at B=1 T	$\geq 5 \cdot 10^4$	10^6	$10^6 - 10^7$	–
Digital output	no, but dedicated ASIC available	with Timepix anode	no	yes

Table 5.9.: A comparison of B-Field insensitive sensor options for the Endcap DIRC. Problematic parameters are highlighted. The PDE of the HAPD is the product of the peak QE (27 %) and the geometric efficiency of the APD array (89 %). A value of the time resolution has not been published as the detector is not designed for timing. In case of the dSiPM (DLS-6400-22), the pixel fill factor has been included in the PDE and the geometric efficiency is the ratio of the total pixel area to the chip size plus two 0.25 mm wide bonding lanes. The data of the fine mesh tube corresponds to measurements of an old prototype [47], published in 2001, but has been found to be compatible with current versions. However, current fine mesh PMTs are only available in a cylindrical envelope. The MCP-PMT data corresponds to a Photonis Planacon.

is unsuitable to be used in an Endcap DIRC design. It is unlikely that these shortcomings can be easily improved as the current device is already the result of many years of development.

The MCP-PMT has the lowest PDE of these detectors, what is a consequence of the limited collection efficiency (eq. 5.42). A second, more severe disadvantage is the aging of the photocathode which has been addressed in a preceding section. Beside these issues, all parameters are excellent.

The fine mesh PMT has a better collection efficiency, a slightly higher transit time spread and does not suffer from aging effects. Its main drawback is the long anode-cathode distance which results from the high number of dynodes needed in these devices. As explained below, the large gap will increase the imaging error.

A closer look on Fig. 5.5 in the requirements section on page 55 reveals that the magnetic field at the readout region is not homogeneous. Further, the sensor cannot be placed at an arbitrary position. The position and alignment will be defined by the imaging optics. It is therefore non-trivial or even impossible to perfectly align the axes of all sensors to the local magnetic field. In short, the sensors have to work in a non-axial magnetic field.

Unfortunately, the magnetic forces are much stronger than the electric forces in the tube and will thus guide the electron trajectories along the magnetic field direction. To underline this argument, an electron trajectory (Fig. 5.37) has been computed for an electron traveling from an MCP output electrode to the anode in presence of a homogeneous 0.5 T field. This trajectory has been obtained by solving the ordinary differential equation:

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = e \cdot (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{5.43}$$

for the relativistic case (together with $\mathbf{v} = \mathbf{p}/E$) using a standard Runge-Kutta solver. The choice of input parameters, as listed in the figure caption, is based on the charge cloud model published in [114].

To mitigate such effects, it is helpful to choose a sensor with a short cathode to anode distance. In case of the 30 mm anode-cathode gap of the fine-mesh PMT, severe displacement can be expected if the sensor is inclined to the field by several degrees. In an MCP-PMT however, the relevant distance is the sum of the input and output gaps which is usually less than 8 mm in standard tubes and less than 3 mm in case of proximity focusing.

Beside the gap size, there are further strong arguments for an MCP-PMT detector. The Photonis Planacon MCP-PMT is a well established device which can be easily

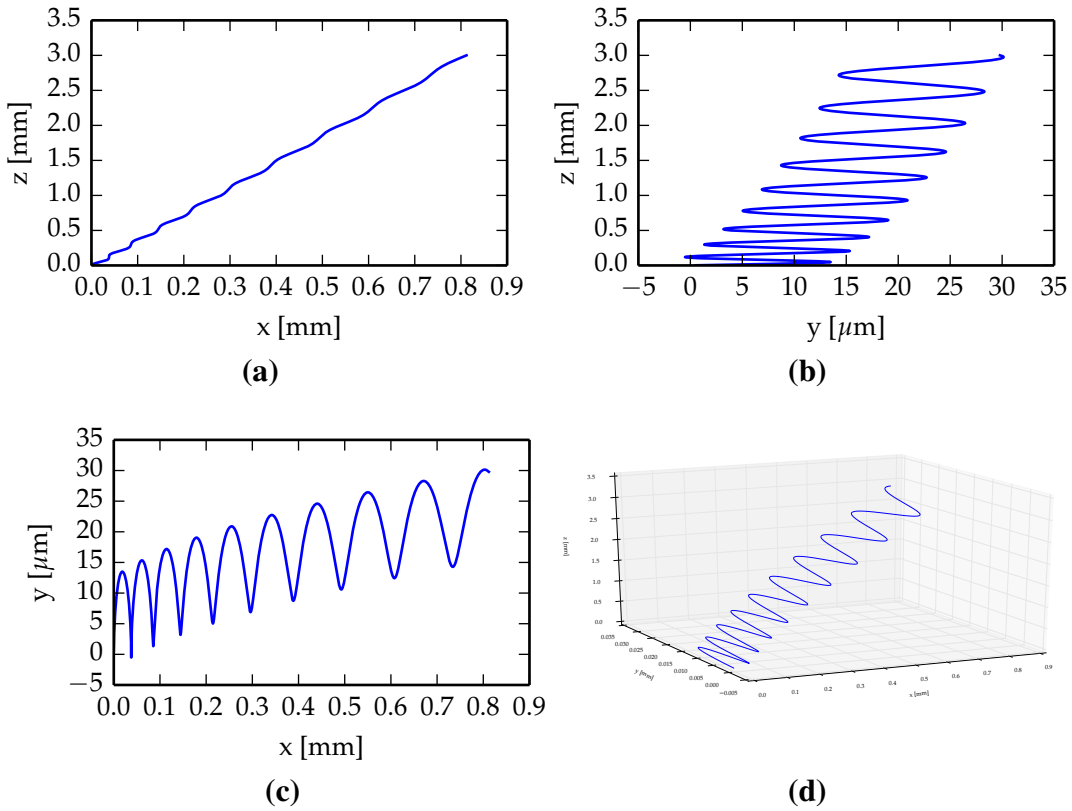


Figure 5.37.: Computed trajectory of an electron with an initial kinetic energy of 10 eV, emitted in the yz-plane at an angle of 15 ° relative to the tube axis (z). The acceleration voltage is 150 V over an output gap of 3 mm and the B-Field of 0.5 T is inclined by 15 ° rel. to z axis (xz-plane). The projection to the xz-plane (a) clearly shows how the electron trajectory is guided along the B-field by the additional gyration, resulting in the expected offset of 0.8 mm. This effect will lead to a linear displacement of the image (homogeneous field) and to nonlinear distortions in case of inhomogeneous fields.

customized, e.g. by changing the anode structure or the size of the input and output gaps, while the current fine mesh PMTs are not even commercially available in a suitable package. Thus, the development effort for fine-mesh tubes would be much higher. Further, the $\overline{\text{P}}$ANDA Barrel DIRC as well as the Belle iTOP detectors foresee to use MCP-PMTs. Due to the large interest of the scientific community, there is a strong ongoing R&D–effort to mitigate the aging effects. Hence, the MCP-PMT is the most promising choice among the vacuum based sensors. The Planacon is the largest tube on the market and other manufacturers have reported

to develop MCP-PMTs with the same form-factor. The Planacon can therefore be used as reference sensor for the detector design.

While the Philips dSiPM has otherwise excellent properties, it suffers from the intrinsically high DCR and sensitivity to radiation damage. A reduction of the DCR can be achieved by cooling which is non-trivial to integrate, especially if the sensors have to be coupled to glass. At a temperature of $\sim 0°$, the DCR would still be about $4\,\mathrm{MHz/cm^2}$. On the other hand, the excellent single photon time resolution of the ASIC permits high background suppression if precise time correlated imaging techniques are being used.

The current implementation of the dSiPM does not live up to its full potential as the hit position is not encoded on the cell level. The granularity of one die could be as high as 100 rows. This corresponds to a DCR of about 10 kHz per row. Such a modification would lead to a reduction of the overall number of dSiPMs and hence the total DCR. A reduction of active area leads also to smaller optical systems, but at the cost of extreme requirements on imaging performance. Smaller pixels are only useful as long as the point spread function of the optics matches the spatial resolution of the sensor.

In conclusion, modified dSiPMs with higher granularity could be an alternative to vacuum based sensors, but they will pose different problems such as the production of high precision optics, a cooling system, high DCR and susceptibility to radiation damage.

At the beginning of this research, the radiation hardness of the dSiPM was unknown and the lifetime of MCP-PMTs limited to less than $1\,\mathrm{C/cm^2}$. Hence, both routes were investigated in the hope that either the dSiPM turns out to be radiation tolerant enough to survive at least half of the $\overline{\mathrm{P}}$ANDA run-time or the lifetime of the MCP-PMTs improves significantly. A later irradiation campaign finally showed that the tested dSiPMs would be severely damaged on the time scale of a few $\overline{\mathrm{P}}$ANDA days [67] and it remained unclear if these sensors can be modified to achieve the required radiation hardness. Fortunately, the MCP-PMT lifetime has been improved up to $6\,\mathrm{C/cm^2}$ as discussed in a previous section. These experimental facts settled the final sensor decision.

5.5.4. Anode options for the Endcap DIRC

While the MCP-PMT has been identified as sensor choice, the required anode layout is still open. As discussed in a preceding section, anode technologies differ

strongly in their rate capability. Hence, the maximum expected hit rate has to be known. An upper rate limit is directly imposed by the lifetime limit of the sensor. This allows to estimate the peak rate at the anode.

At first one has to estimate the size of the total active area, what can be realized without introducing the final design. The tubes will be aligned to the edge of the radiator which has the shape of a polygon. An octagonal radiator with an inner radius of 1055 mm would reach close to the space foreseen for the optics and sensors. This radiator can be equipped with the total amount of 112 tubes, resulting in an active area of $A = 2800\,\mathrm{cm}^2$.

Assuming a homogeneous illumination of the sensors, the maximum integrated number of detected photons at a gain of $G = 10^6$ is

$$N_{\mathrm{ph,life}} = 6\,\frac{\mathrm{C}}{\mathrm{cm}^2} \cdot \frac{A}{e \cdot G} = 1.05 \cdot 10^{17}.$$

According to the system requirements, the long term average event rate is 10 MHz with an average of three tracks in the Endcap region. This results in $N_{\mathrm{track,life}} = 4.74 \cdot 10^{15}$ tracks over the total projected $\overline{\mathrm{P}}$ANDA runtime of $158 \cdot 10^6$ s. In conclusion, the maximum average number of photons per track compatible with the sensor lifetime is

$$N_{\mathrm{ph,max}} = \frac{N_{\mathrm{ph,life}}}{N_{\mathrm{track,life}}} = 22.15$$

This means that the detector design has to be restricted to about 25 detected photons per track. In the following we will assume that the target density fluctuations are not higher than a factor of two. The 10 μs averaged interaction rate has been determined to be 26 MHz, resulting in a peak hit rate of

$$R_{\mathrm{peak},10\mu\mathrm{s}} = \frac{26\,\mathrm{MHz} \cdot 3 \cdot N_{\mathrm{ph,max}}}{112\,\mathrm{tubes}} = 15.4\,\frac{\mathrm{MHz}}{\mathrm{tube}}.$$

For a 1 μs window, the peak interaction rate is 40 MHz, resulting in

$$R_{\mathrm{peak},1\mu\mathrm{s}} = 23.7\,\frac{\mathrm{MHz}}{\mathrm{tube}}.$$

A comparison with the rate capability of the anodes (Fig. 5.36 page 134) shows that these rates are clearly incompatible with all anode designs except the Timepix and the multi-anode. However, one could think about a tiled arrangement where the anode consists of a pattern of individual delay lines or XS anodes (Fig. 5.38, right). The readout of XS anodes requires a precise charge measurement as well

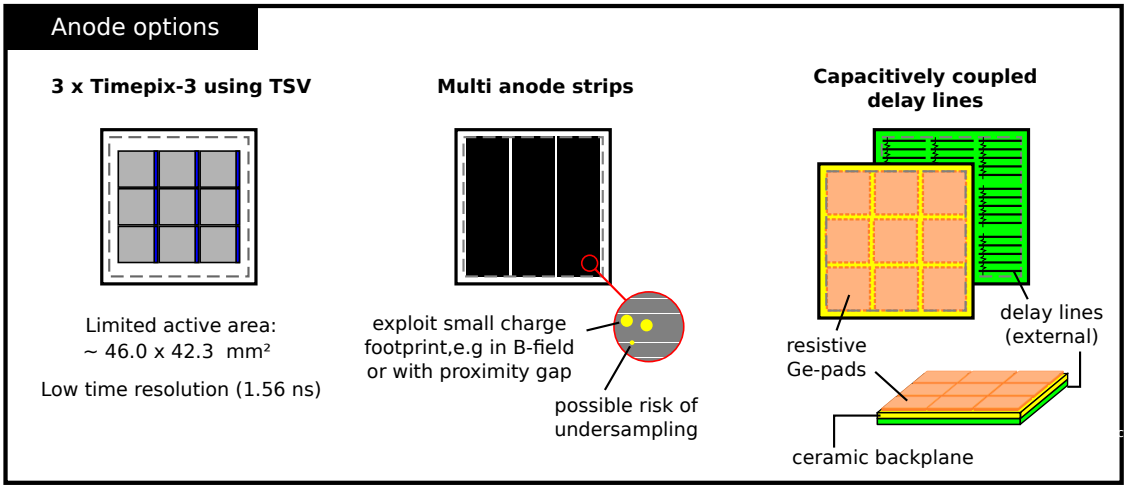


Figure 5.38.: Different conceptional anode layouts for high rate and spatial resolution. A Timepix 3 based readout could deliver the highest rate and spatial resolution, but only a coarse time-stamp. A classical multi-anode pattern can provide a spatial resolution in the sub-mm range as long as the charge footprint on the anode is sufficiently small. The spacing between individual pads has to be large enough to avoid electronic cross-talk. Hence, there is a certain risk that the anode becomes undersampled in presence of a magnetic field where the charge cloud is potentially smaller than the gap. This problem can be avoided by measuring the induced pulse via an anode which is capacitively coupled to a resistive layer. The pile up in a capacitively coupled 3×3 delay line readout would be 5 % at the expected peak rate.

as complicated centroiding, while the hit-position x and time of arrival t_a on a single delay line can be simply computed from the two timestamps t_1, t_2 at either end of the line:

$$x = \frac{t_1 - t_2}{t_1 + t_2} \quad ; \quad t_a = \frac{t_1 + t_2}{2}$$

Further, the pile up probability in the delay line decreases with the length. Assuming a delay line with the typical propagation speed of 100 μm per 100 ps [139] and a length of 50/3 mm (3×3 pattern in Fig., 5.38), one obtains a pile up probability of only

$$P = \left(1 - e^{-\frac{26.7}{9}\,\mathrm{MHz}\cdot 16.7\mathrm{ns}}\right) \cdot 100 \approx 5\,\%$$

at the 1 μs averaged peak rate.

A problem with the delay line as well as with multi-anode structures is the risk of undersampling in a strong magnetic field. If the charge cloud becomes much smaller than the feature size of the anode, part of the signal could be lost at the

gaps. A possible countermeasure is the use of capacitively coupled anodes. As previously discussed, this approach uses a resistive layer inside the tube for charge collection. The charge footprint is then detected by an anode structure which is capacitively coupled to the resistive layer through the ceramic backplane of the vacuum housing. The size of the induced charge footprint on the anode is defined by the thickness of the backplane. This readout requires a sufficient gain $> 10^6$.

Another option is the detection of the charge cloud with the Timepix 3 pixel-detector. The shape of the largest array of ASICs which fits into the tube body is 3×3, corresponding to an active area of $46 \times 42.3\,\mathrm{mm}^2$. Hence, the active area is reduced by 30 %. However, this readout option is insensitive to undersampling and works at a low gain. The > 80 MHz rate capability of the ASIC exceeds that of the MCP and there is no need for additional analog front-end electronics.

A downside is the relatively high power consumption of ~ 1.5 W per ASIC what poses the engineering challenge to design a suitable heat sink. The existing Timepix 2 Planacon readout used a 2×2 array with a total power consumption of ~ 4 W. Simulations during the design phase had shown that the indium seal of the tube would be damaged in absence of cooling [131].

In conclusion, the most straightforward anode option is the multi-anode, even if there is some risk of undersampling. Custom layouts can be produced on a short timescale and at relatively low cost. The Timepix 3 solution would be elegant, but requires a vast amount of development. A capacitively coupled anode like a delay line can be considered as an alternative in case of undersampling of the multi-anode structure or to reduce the required number of readout channels.

5.5.5. Summary

This section provided a preferably complete overview of state of the art single photon sensitive detectors as well as the underlying physical mechanisms and fundamental differences of individual technologies. Based on this information, possible sensor candidates have been identified and further evaluated in respect to their compatibility with the application.

From the available sensor options, which all show certain limitations, the MCP-PMT has been identified as the most promising candidate. While this class of devices suffers from aging effects related to chemical poisoning of the photocathode, they have otherwise outstanding properties such as a superior resistance to magnetic fields, an excellent time resolution, intrinsically high rate capability and

the possibility to reach sub-mm spatial resolution.

A comparison of available anode options further revealed that not all of these properties can be exploited at the same time. Applications with global rates in the order of several MHz have to choose a compromise between time resolution and spatial resolution. The highest spatial resolutions of $< 100\,\mu$m can only be exploited at rates up to 1 MHz. A future exception to this rule could be introduced by a Hybrid MCP-PMT using the Timepix 3 ASIC.

The Photonis Planacon is currently the largest commercially available MCP-PMT with a measured* lifetime of $\sim 6\,\mathrm{C/cm^2}$. It is therefore used as reference sensor for the Endcap DIRC design. As other manufacturers work on tubes with a compatible form factor, the market will be more competitive in the near future. Due to the simplicity and low cost, the preferred anode type is the multi-anode. The possible undersampling inside the magnetic field has to be studied in a prototype.

5.6. Front-end electronics (C4)

has impact on	*key properties*
• Data rate	• Rate capability
• Data loss	• Time resolution
• Cooling infrastructure	• Sensitivity/efficiency
	• Noise and crosstalk
	• Channel density
	• Power consumption
	• Radiation hardness

The term front-end electronics (FEE) typically addresses all electronic parts needed to read-out and process the detector signals before they are transmitted to a data-concentrator, buffer or network. In a wider sense, this also includes voltage supplies and systems for slow control. This section will only focus on the electronics needed to process and digitize the sensor signals, while data-concentration and control will be handled in subsequent sections.

*This measurement includes only one sample. Currently, there is no information about the production spread of this parameter.

5.6.1. Available ASICs for the PMT readout

The requirements for the FEE can be easily derived from the system requirements. Due to environmental constraints, the FEE has to be very compact (compare Fig. 5.4, page 53). As this space limitation complicates cooling, the total FEE power consumption should be low to prevent heat accumulation. The peak photon rate has been estimated to be 17.3 MHz/tube if smoothed with a 10 μs sliding window and 26.7 MHz/tube in case of a 1 μs window. Radiation hardness as defined in the requirement section is prerequisite.

The information required by the Endcap DIRC is the time-of-arrival (ToA) of individual photons with a precision in the order of 100 ps as well as the spatial coordinate on the imaging plane. Both preferred analog anode options (section 5.5.4), the multi-anode and the delay line readout, require electronics with fast timing capabilities but no charge measurement, though it can be of interest to correct for time-walk. The requirements on space and power consumption call for a very high level of integration. Thus, a solution based on ASIC* and FPGA† is preferable.

A precise time measurement in the ps-regime is commonly realized by using time to digital converter (TDC) ASICs. These can easily measure the time of arrival down to a bin size of 10 ps. The input for a TDC is provided using a discriminator which generates a well defined output pulse when the input signal crosses a certain threshold. As PMT signals are rather low, with an average input charge of 160 fC, they have to be amplified before they can be fed into the discriminator. As previously explained, the gain of the PMT results from a multiplication process and is hence subject to stochastic variations. To reach a high efficiency, the circuit should be able to detect pulses down to 1/3 of a single photoelectron pulse (50 fC at a gain of 10^6).

These gain variations translate also directly into pulse-height fluctuations and hence to time-walk. This error can be corrected by using charge information. To avoid the need for additional analog channels, the charge can be approximated by the total time in which the pulse is above a given threshold (*time-over-threshold*, *ToT*). As the ToT can be determined with the same TDC which measures the ToA, this approach does not increase the channel count.

Table 5.10 summarizes key properties of available readout ASICs. The first group comprises ASICs which include at least a preamplifier and discriminator.

*Application specific integrated circuit
†Field Programmable Gate Array

ASIC	*No. of channels*	*time res. [ps]*	*Rate [MHz/ch]*	*Power cons. [W]*	*Channels per cm^2*	$Q_{\text{min.}}$ *[fC]*
MAROC-3	64	$\sim 10^3$ RMS	–	0.22	410	50
NINO	8 32	< 25 RMS	–	0.24 0.68	100 127	50 ($\diamond$) 30
CLARO-CMOS	4	10 RMS	–	< 0.008	100	50
HPTDC	32 8	100 LSB 25 LSB	2.5	0.8–1.5	4.4 1.1	–
ACAM	8 2	81 LSB 27 LSB	10 40	0.15 – 0.5	9 2.5	–
THS788	4	13 LSB	1.6	0.7	2	–
TDC@FPGA	64	<20 LSB	$\gg 5$	1.5*	8.8	–
TOFPET	64	50 LSB	0.15	0.4	257.5	60 (†)
STiC-v3	64	50 LSB	0.15	1.6	256	60 (†)

Table 5.10.: Comparison of front-end ASICs. The first group of three ASICs features at least a preamplifier and discriminator. The second group includes fast TDCs which can be used in combination with the former mentioned ASICs. The last group lists two interesting ASICs which are optimized for the readout of SiPMs. However, their characteristics are also compatible with the signal of PMTs. Latter ASICs offer the huge advantage of combining all parts on a single chip with digital readout. Except for the ACAM TDC, the rate capability accounts for the latency imposed by the digital readout – The computation of channel density is not exactly fair as half of the ASICs are not packaged and will occupy more space when bonded. ($\diamond$): The NINO can accept charges of $Q_{\text{min.}} \approx 10\,\text{fC}$ but at a much higher jitter of $\sim 50\,\text{ps}$. (*): This power consumption is just an estimate based on the FPGA used. (†): Estimated from SiPM requirements. The minimum input charge is possibly lower.

Some of these have also built-in ADCs. The next group lists pure TDC ASICs which can be used in combination with the ASICs in the first group. The last two ASICs are not designed for PMTs but for the readout of analog SiPMs. They combine the preamplifier, discriminator and TDC in one chip, what leads to a very high channel density. Further, they provide ToT information.

The TOFPET ASIC has also the lowest power consumption of all listed ASICs which comprise a TDC. Hence, the TOFPET ASIC is a preferable choice for a multi-anode readout with 64 channels or more. While a compatibility with PMT signals is expected [110], it has still to be verified that the ASIC works flawlessly in this case. A fallback solution can be the TDC on FPGA developed at GSI, in

combination with two 32 channel NINO-ASICs.

Another prominent method to measure the time and charge precisely is the use of switched-capacitor-arrays. These sample the signal at very high rates up to several 10^{10} samples/s what enables applications based on online pulse processing. The trade-off one has to accept is a long dead time in the order of several μs, which is simply a consequence of the large amount of samples which have to be read out at a lower data rate. Such a device is perfectly well suited for triggered application but unusable where continuous sampling is mandatory. In future, there will be exceptions to this rule. The DRS-5 ASIC, which will not be available before 2015, will feature a dead-time-free acquisition mode. In this mode, the self-triggered channels record only short regions of interest which will be stored in a readout buffer. The reduced sampling depth in combination with buffering allows a dead-time-free operation up to hit rates in the order of some MHz. This future ASIC is especially interesting for the read out of a delay line anode.

At the time of writing, the radiation hardness of the TOF-PET ASIC has not been measured. As the ASIC is not optimized for a high radiation environment one can expect problems during operation. However, the $\overline{\text{P}}$ANDA-MVD group is working on the radiation hardening of the digital part of this ASIC design, which will be reused in the PASTA ASIC. This work could be also used to improve the TOFPET ASIC in a future iteration. The only, definitive radiation hard readout option would be a combination of HPTDC or TDC@FPGA* and NINO or CLARO-CMOS. The CLARO-CMOS is sufficiently radiation hard (400 kRad, 10^{14} n/cm^2 [36]) but to the authors best knowledge, the HPTDC and NINO ASICS have not been tested up to the fluence expected in $\overline{\text{P}}$ANDA. Radiation damage is hence an important aspect which has to be understood to select possible hardware.

5.6.2. Radiation induced damage and soft errors

High fluxes of charged particles, neutrons and γ-rays lead to functional and performance degradation of semiconductor circuits and ultimately cause the failure of the device. This kind of permanent errors, also called “hard errors”, are primarily introduced by lattice displacement in the bulk material or the build-up of charge at the oxide and interfaces due to ionization. While the device degradation caused by these mechanisms is usually a long term effect depending on the accumulated

*ported to a rad-hard FPGA like the Altera Arria GX (details in section 5.6.2).

dose, it can also happen that the device is destroyed instantaneously by a single event effect.

Besides the permanent damage of the device, the operation of integrated circuits can be temporarily disturbed by ionization processes which alter the charge in the semiconductor. A typical example is a single event upset (SEU) where a bit in a storage cell is flipped from state 0 to state 1*. This type of non-permanent single event error (SEE) is commonly named "soft error". Soft errors can, for example, randomly change the result of a computation. This can lead to undetected data corruption or race conditions which completely interrupt the operation of the device. The latter situation is called "single event failure interrupt" (SEFI). Another type of soft error is the single event transient (SET) which describes the situation when ionization creates a voltage pulse in the device.

In case of SRAM-based FPGAs, soft errors become especially important because the circuit logic itself is defined by millions of bits on the device (configuration memory, LUTs). A change of this configuration memory can alter the circuit and thus corrupt the operation.

Both, hard and soft errors have to be considered carefully when selecting electronic devices which have to be operated in a high-radiation environment. In general, one has to discriminate between the following mechanisms:

(a) Continuous degradation due to ionization effects,
(b) Continuous degradation due to lattice displacement,
(c) Destructive single event errors,
(d) Non-destructive single event errors.

The contribution of a specific radiation field to the individual damage mechanisms strongly depends on the type and energy of the radiation. The total ionizing dose (TID) is by definition an adequate measure for (a). However, it cannot be used to predict the lattice damage effects (b).

It has been established experimentally that in many cases the lattice damage is approximately proportional to the non ionizing energy loss (NIEL) [129, 140]. NIEL is the amount of energy transferred from the incident particles to the lattice atoms by elastic or inelastic collisions with the target nuclei. As these collisions lead to the generation of lattice defects, NIEL is roughly proportional to the number of defects and thus a good measure for (b). As there are many exceptions

*In case of multiple affected bits, the effect is called multi bit upset (MBU).

to this linear behavior, the NIEL scaling hypothesis should be considered with care. Nevertheless, NIEL is the correct quantity to specify maximum fluence limits as it is directly related to the bulk damage.

Due to this manifold spectrum of effects, the applicability of every electronic device which is used in a particle detector has to be verified be means of two steps:

- The radiation field at the device location has to be known either by measurements or by means of simulations.
- The single event cross sections, total ionizing dose effects and maximum non ionizing energy loss have to specified and verified for the electronic device by means of measurements.

The radiation levels at the sensor region have therefore been determined by using the PandaRoot framework including all available geometries related to the target spectrometer. The resulting specifications are listed in the system requirements table. The energy loss in silicon has been obtained directly from the Monte Carlo code, but the NIEL of the particles had to be computed from the total fluence of the different species.

To compare the damage caused by different particle species X at different kinetic energies E_{kin}, their fluence has to be normalized by weighting the fluence by the individual damage cross sections $D_X(E_{\mathrm{kin}})$. A common standard (ASTM) is to normalize to the damage caused by 1 MeV neutrons ($D_n^{1\mathrm{MeV}} = 95\,\mathrm{MeVmb}$). The result is called the "1 MeV neutron equivalent flux":

$$\phi_{\mathrm{eq.}}^{1\mathrm{MeV}} = \sum_X \int \frac{D_X(E_{\mathrm{kin}})}{95\,\mathrm{MeVmb}} \cdot \phi(E_{\mathrm{kin}}) \cdot \mathrm{d}E_{\mathrm{kin}} \tag{5.44}$$

Unfortunately, data on damage cross sections exists only for a few materials, mainly Si and $GaAs$. The ROSE collaboration (RD48) recommends several datasets for the damage cross section in Si [140] which are shown in Fig. 5.39. This data has been used for fluence normalization. Lighter hadrons have been treated as pions. The simulation results for different radiation related quantities have also been visualized and shown in the figures 8.21, 8.22 and 8.20.

The determined particle flux allows also an estimation of the expected SEU upset rate in FPGAs located at the readout region of the Endcap DIRC. While the SEU cross section increases with the particle energy, it follows a Weilbull

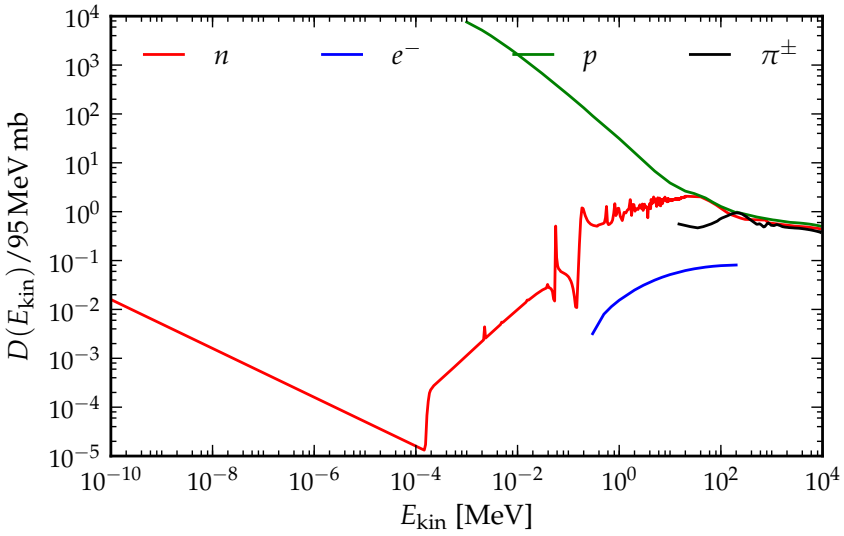


Figure 5.39.: Silicon damage cross section normalized to 1 MeV neutron damage. Data from several sources recommended by Vasilescu and Lindstroem. See [140] and references therein.

.

function and quickly saturates at a certain, device dependent value [38]. The SEU cross section for a single bit in the radiation hardened Virtex XQVR300 FPGA (release year ~1999, 250 nm CMOS process) under proton irradiation has been measured to be $\sigma_{\mathrm{SEU,bit}} = 2 \cdot 10^{-14}\,\mathrm{cm}^2$ [38]. Under the simplifying assumption that all charged hadrons have the same SEU cross section, one can estimate the failure rate of a device with $2 \cdot 10^6$ bits:

$$R_{\mathrm{device}} = 100\,\frac{\mathrm{Hz}}{\mathrm{cm}^2} \cdot 2 \cdot 10^{-14}\,\frac{\mathrm{cm}^2}{\mathrm{bit}} \cdot 2 \cdot 10^6\,\mathrm{bit} = 4 \cdot 10^{-6}\,\mathrm{Hz}$$

In case of one FPGA per sensor and about 100 sensors, one obtains a total rate of $R_{\mathrm{tot}} = 1.4$ SEUs per hour corresponding to a *mean time between failure* (*MBF*) of $1/R_{\mathrm{tot}} = 42$ min. This failure rate is quite high but could still be handled by error detection techniques which correct corrupted data in the SRAM or by reprogramming the whole FPGA in case of a detected SEU.

However, modern FPGAs can have much lower SEU cross sections. More recently, the SEU cross section of an Arria GX FPGA by Altera (release year ~2007, 90 nm CMOS process) has been measured to be $\sigma_{\mathrm{SEU,bit}} = 1.6 \cdot 10^{-16}\,\mathrm{cm}^2$ [34]. The device under test provided close to 10^7 configuration bits, corresponding

to $R_{\text{device}} = 1.6 \cdot 10^{-9}$ Hz and a MBF of 72 days for 100 devices. Further, this device is radiation hard up to a TID exceeding 350 krad and a 1 MeV equivalent neutron fluence* of about $6 \cdot 10^{11}/\text{cm}^2$. These values are compatible with the Endcap DIRC requirements and the estimated failure rate can be considered to be acceptable.

This estimation does not account for a contribution of ions which have a several order of magnitude higher cross section, e.g. saturating at $\sim 10^{-7}\,\text{cm}^2$ for the XQVR300. The rate of high energetic ions is expected to be negligible, but the generated amount of Monte Carlo data is not sufficient to fully support this assumption.

The failure rate can be further reduced by using flash based FPGAs in which the configuration SRAM is replaced by SEU resistant flash memory. A third alternative is the antifuse based FPGA where each configuration bit is represented by a single antifuse†.

In conclusion, modern SRAM based FPGAs can be used as readout option for the Endcap DIRC. A viable solution is the Altera Arria GX FPGA, which is fully compatible with the environmental requirements. The use of less SEU prone FPGAs with flash or antifuse based configuration memory is not mandatory.

5.7. Data concentration and network (C5,C6)

has impact on	*key properties*
• Data rate	• Maximum bandwidth
• Data loss	• Power consumption
• Cooling infrastructure	• Radiation hardness

To stay compatible with the environmental constraints, the FEE will be based on ASICs or FPGAs which ultimately deliver the digitized signal. The output of individual FEE entities has to be sent to the DAQ system by means of a network. The task of the *data concentrator* is to provide a network link and to manage the communication between individual FEE parts and the outside systems, e.g. DAQ and *slow control*, over this network. In this context, slow control signals are commands to start, stop, reset or monitor the operation of individual FEE entities. Additionally, the $\overline{\text{P}}$ANDA DAQ infrastructure demands the time tagging

*Estimated for 350 krad by linear interpolation from the numbers given in [34]

†An antifuse is the opposite of a fuse. Initially, it has a high ohmic resistance. This resistance can be lowered non-reversible by applying a certain voltage.

of individual hits based on a common $\overline{\text{P}}$ANDA reference time which is distributed by a system named SODA (Synchronization of Data Acquisition).

The functionality of the data concentrator can either be implemented on an FPGA, e.g. the Altera Arria GX which has been found to be compatible with the environment (compare preceding section), or by using dedicated ASICs.

In the recent years, the CERN community has developed a feature-rich architecture for data transmission named Gigabit Transceiver (GBT) [85], comprising a radiation hard optical link ("*Versatile Link*"), a SERDES (serializer-deserializer) ASIC (GBTX) and an optional slow control interface ASIC (GBT-SCA).

The GBT chipset provides a user data rate up to 3.28 Gb/s in GBT mode. This mode includes a Reed-Solomon based forward error correction. FEE entities can be connected to the GBTX ASIC using up to 40 SLVS* links ("e-Link"). All e-Links use geographic addressing. The clock of these e-Links is generated by dedicated phase-locked-loops and can be configured individually. Higher clock rates result in a lower number of available links (40 e-Links at 80 Mb/s, 20 e-Links at 160 Mb/s, 10 e-Links at 320 Mb/s). While the e-Link receiver accepts also LVDS signals, the output is SLVS only. Another interesting feature of the GBTX is the possibility to provide an external reference for the jitter-cleaning phase-locked-loop which generates the internal clock. This input can be used to synchronize the GBTX clock, and hence the e-Link clocks, with SODA.

The GBT-SCA can be connected to an GBTX using one of the e-Links. This slow control ASIC provides interfaces to I2C bus, JTAG and SPI (both master) as well as 32 digital I/O lines, 32 ADCs and 4 DACs.

The TOFPET ASIC is completely compatible with the GBTX. It provides two LVDS output ports at a 80 MHz or 160 MHz clock. These can be directly connected to an e-Link. The configuration of the ASIC is realized over an SPI compatible link which can be interfaced using the SPI master on the GBT-SCA together with the digital I/O for chip selection/multiplexing. Further, by using the e-Link clock at 160 MHz as clock input for the TOFPET ASIC, the TDC can be synchronized to SODA by means of the earlier mentioned GBTX clock input.

This scenario allows to read out up to nine TOFPET ASICs at 320 MBit/s (DDR). Due to the clock synchronization, the TDC timestamps can be directly converted to $\overline{\text{P}}$ANDA time. Due to the simplicity of the solution and the minimal development effort in comparison with the implementation of such features on FPGA, the

*Scalable Low Voltage Signaling – like LVDS but with a common voltage of 400 mV instead of 1.2 V.

presented TOFPET/GBT scenario is the favorite solution for an MCP-PMT based Endcap DIRC.

In case of FPGA based TDCs, the preferred solution would be to use the communication features which are already implemented in these FPGAs. As the SODA receiver is also implemented on an FPGA, this device could serve as data concentrator, similar to the GBTX. This approach also serves as fallback solution in case that the GBTX chip is unavailable to $\overline{\text{P}}$ANDA or the Endcap DIRC.

5.8. Compute nodes (C7)

has impact on	*key properties*
• Data rate • Data loss	• Maximum bandwidth • Power consumption • Radiation hardness

In the context of the Endcap DIRC system, a *compute node* is a device which accepts data from the DAQ network and performs actions on this data. These actions can be of any kind of data processing ranging from cluster finding and centroiding methods to particle identification. Such a compute node can be a simple computer, a multi-processor system or an FPGA board.

In $\overline{\text{P}}$ANDA, the term compute node is also related to a multipurpose, ATCA* compliant FPGA board which has been developed for the $\overline{\text{P}}$ANDA DAQ system. This board will be referred to as *PCN* to avoid confusion between the terms. A block diagram of the PCN is shown in Fig. 5.40.

The heart of the PCN is the processing part which offers four user-programmable FPGAs, interconnected by fast serial links. Each of these FPGAs provides two optical links and one gigabit Ethernet port at the frontpanel of the ATCA board. FPGAs on one PCN can also communicate with different PCNs over the backplane of the ATCA shelf by means of a data switch. However, the interconnection between the switching FPGA and the processing part are not realized as fast RocketIO connections. Each FPGA has access to dedicated DDR2 RAM up to 4 GB.

Due to its fast communications and potentially high computing power, this board is an ideal building brick for computing intensive online processing tasks.

*Advanced Telecommunications Computing Architecture

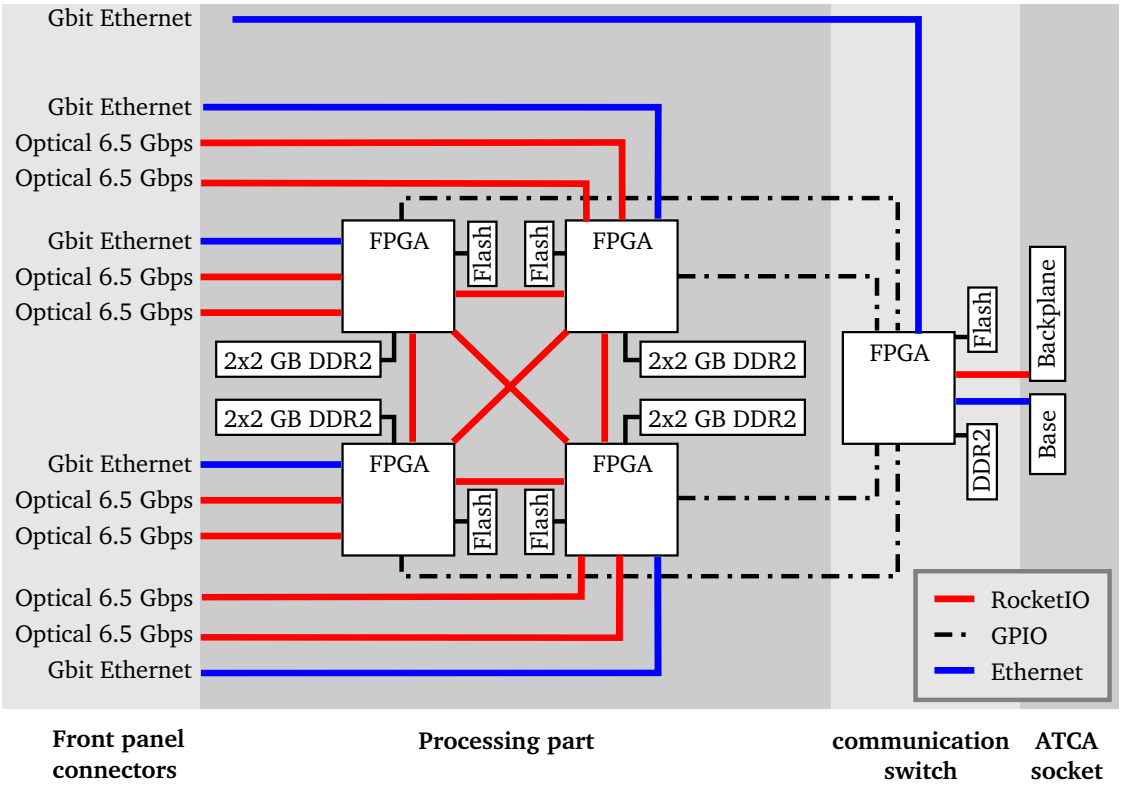


Figure 5.40.: Simplified block diagram of a PCN, adapted from [149] and updated to newer hardware specifications. The board offers 4 Virtex-5 FX70T FPGAs which are interconnected by RocketIO (3.75 Gb/s line rate). Each FPGA is equipped with 4 GB DDR2 SRAM and provides two fiber links as well as one Gbit Ethernet port. A fifth FPGA manages the communication between the four others and different PCNs located in the same ATCA shelf through the ATCA backplane. A newer version of the board is separated in an xTCA carrier board equipped with the switching part and four AMC (Advanced Mezzanine Card) slots. The FPGAs of the processing part and related periphery are located on individual AMC mezzanine cards.

Possible applications in the Endcap DIRC FEE/DAQ chain are online signal processing of digitized MCP-PMT signals, e.g. the computation of the charge cloud centroid on the anode, and online reconstruction of measured patterns. The FPGAs are especially interesting to implement many parallel processing chains to derandomize time consuming computations during pattern analysis.

The downside of the PCN is the high effort involved in the implementation of such algorithms on an FPGA which are programmed using a hardware description language. Implementations in software are almost always easier to develop and

maintain than hardware solutions. It is therefore advisable to consider FPGA solutions only in cases where software solutions on conventional compute nodes are limited in performance or cost.

5.9. Particle identification (C8)

Particle identification is realized by discriminating different particle species due to their dissimilar response in a detector. Basically, the difference in response of a Cherenkov detector can be quantified in two ways:

- comparison of the photon hit-pattern to a computed prediction or calibrated samples, e.g. by means of likelihood methods.
- direct reconstruction of the Cherenkov angle.

At BaBar, the reconstruction of the single photon Cherenkov angle has been achieved by a rather simple method, rooted in the uncomplicated nature of the pinhole camera. Tracking information identifies the bar in which the emission took place. The photon direction vector is computed simply from the hit position, identified by a PMT, and the exit face of this bar (Fig. 5.41). The Cherenkov angle is then computed as the angle between the photon direction and the track direction. Complications arise only due to the kaleidoscopic nature of the image at the end of the bar, leading to a 16-fold ambiguity including reflections at the wedge (Fig. 3.7). These ambiguities can be partly resolved by time cuts and the exclusion of unphysical photon paths. The number of remaining ambiguities is typically ~ 3 [144]. The Cherenkov angle per track is then determined by a maximum likelihood fit to the single photon Cherenkov angles as well as the number of detected photons.

The use of wide radiator plates instead of narrow bars reduces the kaleidoscopic effect because different paths with or without reflections at the radiator rim become more separated in space and time. However, as argued in section 5.3, the possible paths have to be determined to reconstruct the angle of the photon in the radiator and hence the Cherenkov angle. This becomes especially complicated in case of multiple rim reflections in radiators with an arbitrary convex polygonal outline as defined in the abstract DIRC model.

An algorithm for path reconstruction in such radiators had already been developed by the author [79]. In the following, this algorithm is referred to as VIL

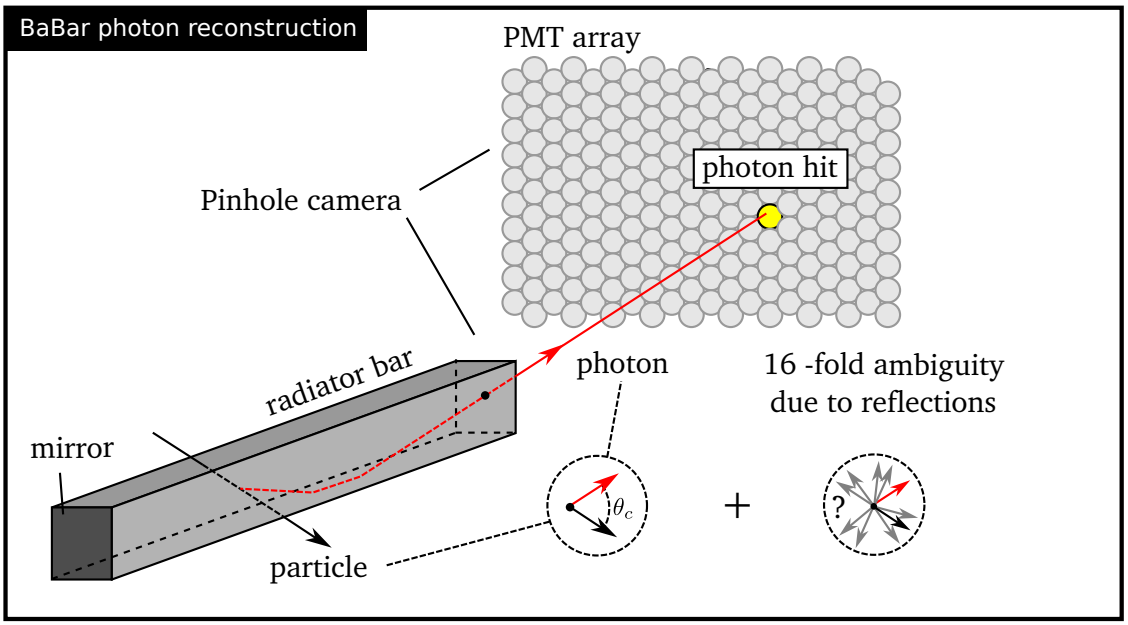


Figure 5.41.: Reconstruction of the single photon Cherenkov angle at the BaBar DIRC. From tracking information, the particle track and correct bar are known. The photon is detected by a single anode PMT. Its path is extrapolated to the exit surface of the bar, resulting in a photon direction vector. The angle between this vector and the direction vector of the particle track is considered to be the Cherenkov angle. However, there are several ambiguities due to the kaleidoscopic nature of the image. The last reflection could have happened at a top or bottom, left or right surface, the photon could have been reflected at the mirror or not, or at the wedge (not drawn) or not [144].

(virtual image lookup). Photon paths are predicted by VIL using a lookup table, tree or dictionary of geometric structures (recon-nodes). Each of these structures belongs to a distinct reflection sequence. A reflection sequence is defined as the sequence of radiator-segments at which the photon has been reflected prior to detection. Recon-node structures are obtained from the radiator shape by a consequent construction of virtual images.

An example of such a structure is presented in Fig. 5.42. The real photon path is shown in radiator plane projection at the left side of the figure. This photon is first reflected at the blue radiator segment, then at the red one and finally detected in an optical element located on the green segment.

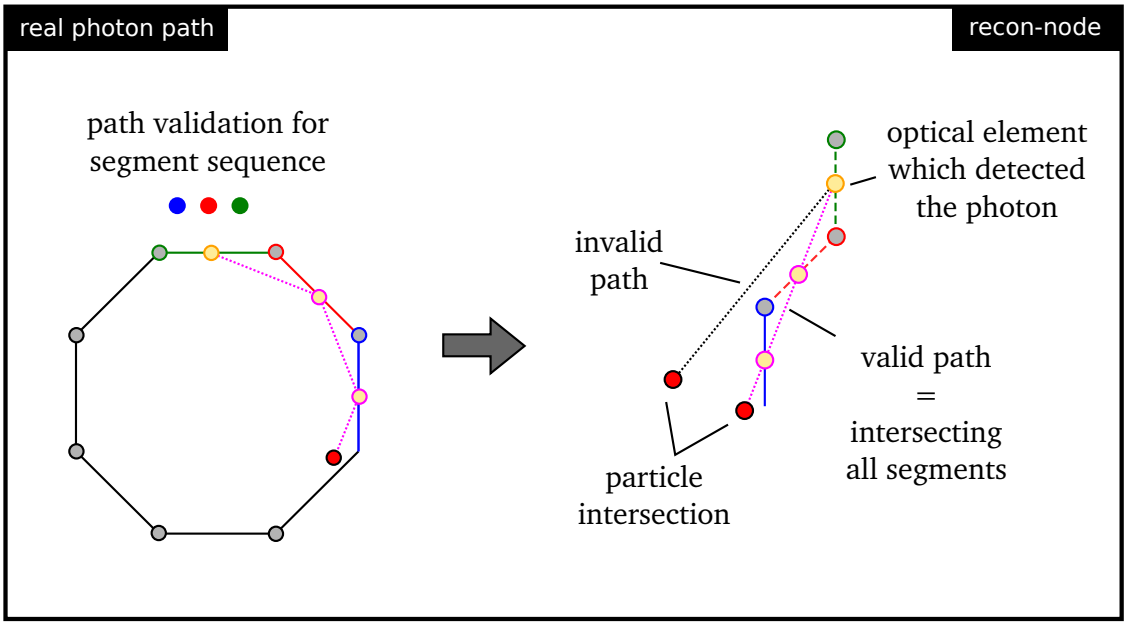


Figure 5.42.: Construction of a recon-node in radiator plane projection and validation of possible particle tracks.

The corresponding recon-node structure for N segments $S_1 \dots S_N$ is created by mirroring the segments S_i with $i > j$ at S_j for $j = (N-1) \dots 1$. In the given example: the green radiator-segment is first mirrored at the red segment. In the intermediate structure, the green segment and the red one are mirrored at the blue segment, resulting in the final structure shown at the right side of Fig. 5.42. This structure corresponds to the virtual image which an observer would see by looking into the blue mirror. The photon path corresponds to the straight line of view.

A recon-node can be used to probe whether a path to a optical element on the green segment is physically possible or not. The path exists if the line from the emission point in RPP to the optical element on the green segment intersects all radiator segments in the node.

A major advantage of this approach is that the recon-node structure preserves all relevant coordinates needed for the computation of the Cherenkov angle by means of the DIRC model presented in section 5.3:

- the coordinates of the origin of the photon path (particle intersection)
- the emission angle relative to the particle track ϕ_{rel}
- the pathlength $s_{2\mathrm{d}}$ of the photon

- the angle α_{FEL} to the normal of the optical element

All these values can be determined directly when probing a path node and do not have to be computed separately.

In summary, VIL is based on a dictionary of virtual images which can be used to probe whether a track is physically possible for a specific reflection sequence. All reconstruction relevant geometric quantities can be computed during this test.

The number of reflection sequences depends on combinatorics. Given that the instrumented radiator segments can be considered to be non-reflecting, the number of nodes needed to predict up to M rim reflections in a radiator with N reflecting segments is:

$$1 + \sum_{i=1}^{M} N \cdot (N-1)^{(i-1)} \tag{5.45}$$

per instrumented segment. The number of intersections which have to be computed to probe all paths to a given optical element is:

$$\sum_{i=1}^{M} N \cdot (N-1)^{(i-1)} \cdot i \tag{5.46}$$

In a radiator with three reflecting segments, one obtains the numbers given in the following table. It is obvious, that this generic algorithm becomes very compu-

rim reflections	1	2	3	4	5	6
nodes	4	10	22	46	94	190
intersections	3	15	51	147	387	963

tation intensive in case of radiators with many reflecting edges or a high number of expected rim reflections. However, since the computation of the nodes is completely independent, these can be executed in parallel. The structure of the recon-node is only geometry dependent and hence static. Further, the computation of line intersections can be efficiently implemented as matrix operation on GPUs as well as on FPGAs. By evaluating all nodes in parallel, the computation time can be drastically reduced. The numbers listed in the above table are still small in comparison to the capacity of current hardware. The $\overline{\mathrm{P}}$ANDA compute node is an excellent candidate to implement the parallel evaluation of recon-nodes in FPGAs for online reconstruction.

VILs computational complexity originates from the fact that it applies to radiators with any convex polygonal outline. It is not able to use eventual symmetries in the radiator shape to simplify its calculations. In case of radiators with very high symmetry, e.g. a rectangular plate, the algorithm should be replaced by simpler descriptions which exploit these symmetries.

An example for such a simplification is presented in Fig. 5.43. The figure shows a proximity focusing DIRC with a rectangular radiator plate, four photon paths with 0 to 4 rim reflections and the related virtual images (gray). Due to the regular pattern in the virtual image, the angle ϕ_{ph} and thus the track length and ϕ_{rel} can be described by the trivial expression in the bottom right corner of the figure. Similar expressions can be found for photons reflected at the rear end of the plate.

With this expression, the problem is already solved. The internal reflection angle φ can be obtained by extrapolating the pixel position to the radiator center:

$$\tan\varphi = \frac{z}{L} \cdot \cos\phi_{\mathrm{ph}}$$

and the Cherenkov angle can be computed using eq. 5.4 with $\phi_{\mathrm{rel}} = \phi_{\mathrm{ph}} - \phi_{\mathrm{track}}$.

VIL would solve the problem in Fig. 5.43 by computing all intersections in all virtual images (gray dots) what is highly impractical in this case. In conclusion, VIL is only interesting in case of complex radiator geometries with low symmetry.

So far, the presented tools in this section permit only the reconstruction of the Cherenkov angles for a single detected photon. In the next reconstruction step, the angles of the detected photons have to be combined to a common track Cherenkov angle.

One option is to compute the truncated mean of the Cherenkov angle. Particle identification can then be realized by choosing an appropriate classifier, as explained in section 3.2.

A powerful alternative is the application of a maximum likelihood fit to the measured single photon angles. In case of many, e.g. several ten detected photons, the maximum likelihood fit allows to account for the expected background and number of detected photons for every individual particle hypothesis. This approach has been applied at BaBar [144]. The maximum likelihood fit can not only be applied in Cherenkov angle space, but also in pattern space. In this case, the fit is applied to the time stamps and positions of detected photons. This approach is used at the Belle II iTop counter [125].

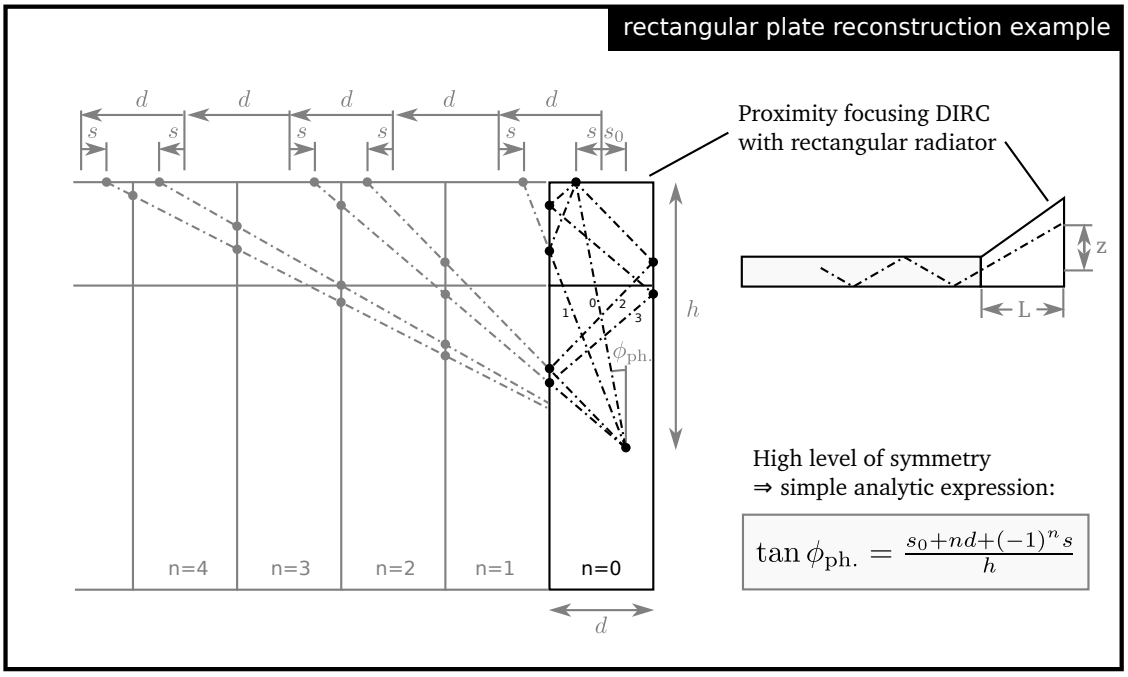


Figure 5.43.: The high symmetry of rectangular plates allows to formulate very simple analytic expressions which are faster to evaluate than the recon-node based approach. The VIL algorithm would evaluate every single virtual image in detail, as it is incapable of detecting and simplifying for symmetries in a given shape.

5.10. Conceptual design options and design synthesis

In the preceding sections, alternative options for the practical realization of each logical detector component have been discussed in respect to their applicability and compliance with the system requirements as defined in section 5.1. Table 5.11 summarizes the most viable options identified in this process, accompanied by additional design constraints which are implied by this specific choice.

As the Endcap DIRC has to operate continuously at high track rates, it is important to assure that all relevant hit information can be read out without data loss. It is apparent from Fig. 5.7, that the information flow in the Endcap DIRC can be described as a single pipeline from C1 to C8. The performance figure of all components up to the FEE (C1–C4) influence the total rate of detected photons. The resulting data rate is then determined by the FEE choice which defines the size of the digitized hits. This information has to be transmitted to C6–C8 via C5.

The components C6–C8 are considered to be uncritical because the rate capa-

ID	*Name*	*Options*	*Implied constraints*
C1	Radiator	Fused silica plate	Machining limit for the precision polishing of the large surfaces is a circular area of R < 60" (1524 mm)
C2	Imaging optics	Focusing mirror on fused silica + LiF/CaF_2 + Interference filter	– Prism length $\lesssim$20 cm Angle of incidence dependence
C3	Photosensor	MCP–PMT ~~dSiPM~~	Detector life $Q_{\text{anode}} < 6\,\text{C/cm}^2$ and orientation to B-field radiation damage
C4	Front-end electronics	TOFPET ASIC TDC@FPGA + NINO32 (STiC–v3) (TDC@FPGA + Claro–CMOS)	Rate capability and form factor according to section 5.6
C5	Data concentration	GBT chipset Altera Arria GX or similar	User data rate 3.28 Gbps Raw data rate 3.125 Gbps
C6	Network	Optical link up to 6.5 Gbps	–
C7	Compute Nodes	Multi-processor system $\overline{\text{P}}$ANDA compute node (PCN)	–
C8	Particle identification	VIL + mean + classifier VIL + max. likelihood fit	–
CA1	Optical coupling	NOA-61 EPOTEK 301-2 BC 630	λ-cutoff at 350 nm λ-cutoff at 300 nm λ-cutoff < 300 nm

Table 5.11.: List of viable options for individual detector components which have been identified in the preceding sections. Note that the term fused silica refers always to synthetic fused silica. The given size limits of optical components result from discussions with different manufacturers. The STiC–v3 ASIC and a Claro–CMOS ASIC with more than 4 channels are not yet available.

bility can be scaled out any time by adding a higher number of nodes for digital signal processing and reconstruction. These parts are outside the spectrometer and thus not subject to tight spatial constraints. At this stage, only cost will be a limiting factor.

Components C2–C5 however have to share the limited space located close to the EMC frame which is highlighted in Fig 5.44. Hence, the option for horizontal scaling, i.e. increasing the number of entities to increase the performance, is limited. Vertical scaling, i.e. increasing the performance of each individual entity, would only be possible by using different hardware. However, to the authors best

knowledge there is no superior alternative to the listed options, except for the type of FPGA.

As stated in section 5.7, a total number of nine ToFPET ASICs can be read out with one GBTX ASIC at a 320 Mbps line rate. This results in a total data rate of 2.88 Gbps corresponding to a sustained hit rate of 57.6 MHz at a 50 bit data size*. The form-factor of the ASICS would permit to use 9 TOFPET ASICS and one GBTX per MCP-PMT to read out a 24×24 multi-anode structure. This would allow to digitize hit rates up to 2.3 MHz per cm^2 active area.

The maximum hit rate per MCP-PMT can be derived from the lifetime constraint, defined as the maximum extracted anode charge per active area at a gain of $G = 10^6$. For a Planacon tube with $A = 25\,cm^2$ one obtains a sustained rate of

$$\frac{Q_{\text{anode}} \cdot A}{G \cdot e \cdot 158 \cdot 10^6\,\text{s}} = 5.9\,\text{MHz}$$

over the nominal $\overline{\text{P}}$ANDA runtime. This is the maximum rate, averaged over the total runtime, which the device can handle before the quantum efficiency starts to drop. Hence, the components C1–C2 have to be designed in such a way that the average hit rate does not exceed this limit.

Following the preceding estimations, it is obvious that the 5.9 MHz photon hit rate poses no challenge to the readout electronics and data concentrators as it is one order of magnitude lower than the maximum rate these can handle. However, the 5.9 MHz are an average rate corresponding to an interaction rate of 10 MHz. To estimate the peak rates due to target density fluctuations and beam statistics, one has to scale this value according to the interaction rate.

A TOFPET channel has a dead time of several 10 ns in case of small signals. This dead time is valid for 4 consecutive pulses due to de-randomization. The data of all channels is collected in frames which are transmitted every 1024 clock cycles (6.4 μs), corresponding to 40 hits at 320 Mbps. In case of uniform illumination, the ASIC is limited by the readout rate and not the channel dead time. The relevant timescale for the computation of the average rate is thus 6.4μs. The peak interaction rate averaged over a 6.4 μs wide gliding window has been determined† to be 27 MHz for a target density factor of 2. Scaling by the ratio of the interaction rates results in a peak rate of

$$5.9\,\text{MHz} \cdot \frac{27\,\text{MHz}}{10\,\text{MHz}} = 15.9\,\text{MHz}.$$

*40 bits in 8B/10B encoding

†using the procedure explained in section 5.1

This rate can already be handled by three TOFPET ASICs.

These estimations show clearly that the Endcap DIRC is not limited by the rate capability of components C4–C5. The rate capability can always be adjusted to operate without data loss up to the nominal interaction rate of 20 MHz, including statistical fluctuations due to the continuous beam.

As stated above, the remaining components C1–C3 have to be tailored to be compatible with the rate limit of C3. A reduction of the number of detected photons can be approached in three different ways:

1. Reduction of the radiator thickness d_R.
2. Reduction of the wavelength acceptance.
3. The use of components with higher photon losses.

Approach 1 will increase the average number of reflections during photon transport and hence the related reflection losses (eq. 5.25). On the other hand, a thinner radiator is equivalent to a smaller aperture of the focusing optics, what can enhance the imaging resolution.

Approach 2 reduces the overall dispersion error. It has been underlined in previous sections that the Endcap DIRC will need some mechanism for dispersion correction to reach the required performance at high momenta. Hence, this approach is very appealing as it addresses two problems at the same time.

Approach 3 can be useful if the specific component brings an advantage to the system. Such a component is a fiber optic plate (FOP). As discussed in section 5.5, the magnetic field will contribute to the image distortion in the tube by its influence on the charge trajectories along the input and output gaps. This effect can be mitigated by using smaller gap sizes. Due to mechanical constraints, the input gap is usually reduced by the use of step windows (compare Fig.5.35) which reach inside the tube. In case of off-axis illumination, the edges of the step window can cause shadowing and unwanted reflections. These distortions can be eliminated completely by realizing this window as fiber optic plate. The trade-off is a reduced transmission of 60 % to 70 %.

Several combinations of these approaches are listed in tab. 5.12 for comparison in terms of photon yield as well as chromatic error. The combinations with a 20 mm thick radiator show the lowest chromatic errors. The dielectric bandpass filter (DBP) is clearly the best option out of these, followed by a dielectric longpass (DLP) / SFP combination. The use of a FOP requires also a 20 mm radiator if the

d_R [mm]	PMT window	Filter/λ range [nm]	$\sigma_{\text{chrom.}}$ [mrad]	N_{ph}	$\frac{\sigma_{\text{chrom.}}}{\sqrt{N_{\text{ph}}}}$
20	SFP	DBP / 360–465	1.9	22	0.41
20	SFP	DLP / >350	2.9	31	0.52
20	FOP	DLP / >350	2.9	21	0.63
15	SFP	DLP / >350	2.9	23	0.60
15	FOP	EPO / >300	4.6	25	0.92
10	SFP	EPO / >300	4.6	23	0.96
20	SFP	EPO / >300	4.6	48	0.66

Table 5.12.: Expected number of detected photons N_{ph} per charged track with $\beta = 1$ for different radiator thickness d_R, PMT faceplate (SFP = standard fused silica plate, FOP = fiber optic plate) and wavelength acceptance (DBP = dielectric bandpass filter, DLP = dielectric longpass filter, EPO = EPOTEK 301-2). The rms error for the given wavelength range $\sigma_{\text{chrom.}}$ has been folded with the PDE of the Photonis Planacon and the Cherenkov emission spectrum. The underlying model is based on numerical integration of the product of following quantities: Frank-Tamm formula (eq. 3.7), Planacon PDE, wavelength dependent reflection efficiency (eq. 5.25), bulk transmission (1. - eq. 5.13), Al-mirror reflectivity from fused silica (Fig.5.19) and finally a trapping fraction of 64 %. The last row of the table shows the maximum number of detected photons which can be achieved by choosing a 20 mm thick radiator and the widest wavelength band.

dispersion has to be limited by the wavelength band. These results, together with the arguments concerning the radiator thickness discussed in section 5.4.1, lead to the decision that a 20 mm radiator is the best choice. This decision leaves also all options open for the future. For example, an already ordered radiator prototype does not become obsolete in case that it turns out during development that an FOP is required.

The next point to address is the radiator shape as well as the design of the focusing optics. Available options are summarized in Fig. 5.44. The shape of the radiator is primarily defined by the outline of the EMC insulation. However, the size limitation of the radiator does not permit to cover the whole acceptance with a single plate. One can consider to bond several radiator plates to cover the whole area. However, to minimize photon losses at the interfaces, these bonds have to be realized bubble-free and without any defects. Even then, unavoidable reflection losses remain due to the mismatch of refractive indices. Further, the individual radiator pieces have to be aligned much better than 1 mrad to avoid

significant distortion of the transported photons. Considering the length of the edges which have to be bonded, the fact that the identified glue options (CA1) are non-reversible, and finally the high cost of precision polished fused silica plates, it can be claimed that this process involves a high risk. To avoid this risk, it has been proposed to subdivide the whole detector into four sub-detectors, each with a monolithic radiator.

Such a radiator is shown in Fig. 5.44. The outline can be adapted to optimize the alignment of the attached optics and hence the PMTs in respect to the magnetic field. The number of required edges should be minimized to reduce the cost for polishing.

As discussed in section 5.4.2, lens systems are an ineligible option, mainly due to the restrictions implied by the small set of radiation hard optical materials. The choice is hence limited to focusing mirrors. These can be designed as classical, aluminum coated surfaces as proposed by T. Kamae in [58] or even as acylindric shapes which allow to exploit total internal reflection as proposed by K. Föhl in [39]. Dispersion correction can be realized by means of filters and achromatic prisms, as discussed in section 5.4.4. The correction of the chromatic error by means of a precise time measurement is not considered to be a preferable option for the Endcap DIRC design because of two reasons. The pathlength of the light is short, e.g. in the order of half a meter in radiator plane projection, and there is no precise reference time available to compute the time of propagation from the measured time of arrival.

The focusing optics will be attached directly to the edge of the radiator. As explained in section 5.3, the position on the focal plane depends on the angle α_{FEL} of the photon path to the optical element in radiator plane projection. Hence, it is either required to reconstruct α_{FEL} (5.9), requiring a very wide focusing element to mitigate ambiguities resulting from folding (compare Fig. 5.10), or to obtain α_{FEL} from the position of the optical element, requiring very narrow elements to reduce the uncertainty of α_{FEL}. A reduction of the focusing element width results in the need of more focusing elements to cover the same area and hence to more surfaces which have to be polished. As precision polishing has a major stake in the overall production cost of these elements, a reduction in width is not economic.

So far, the development of the discussed initial detector design as outlined in Fig. 5.44 has been realized mostly by means of numerical modeling. For the next system design steps however, more complex simulation and optimization routines are necessary to develop the optical system and to obtain a more realistic

picture of the detector response. Additionally, a figure of merit is required for the design process. The system performance has to be evaluated by means of a PID method. This method will be based on VIL and the DIRC model presented in section 5.3. The subsequent chapter will introduce the simulation, optimization and reconstruction software which has been developed for this purpose. The following chapter will then present the development and performance analysis of the final system design.

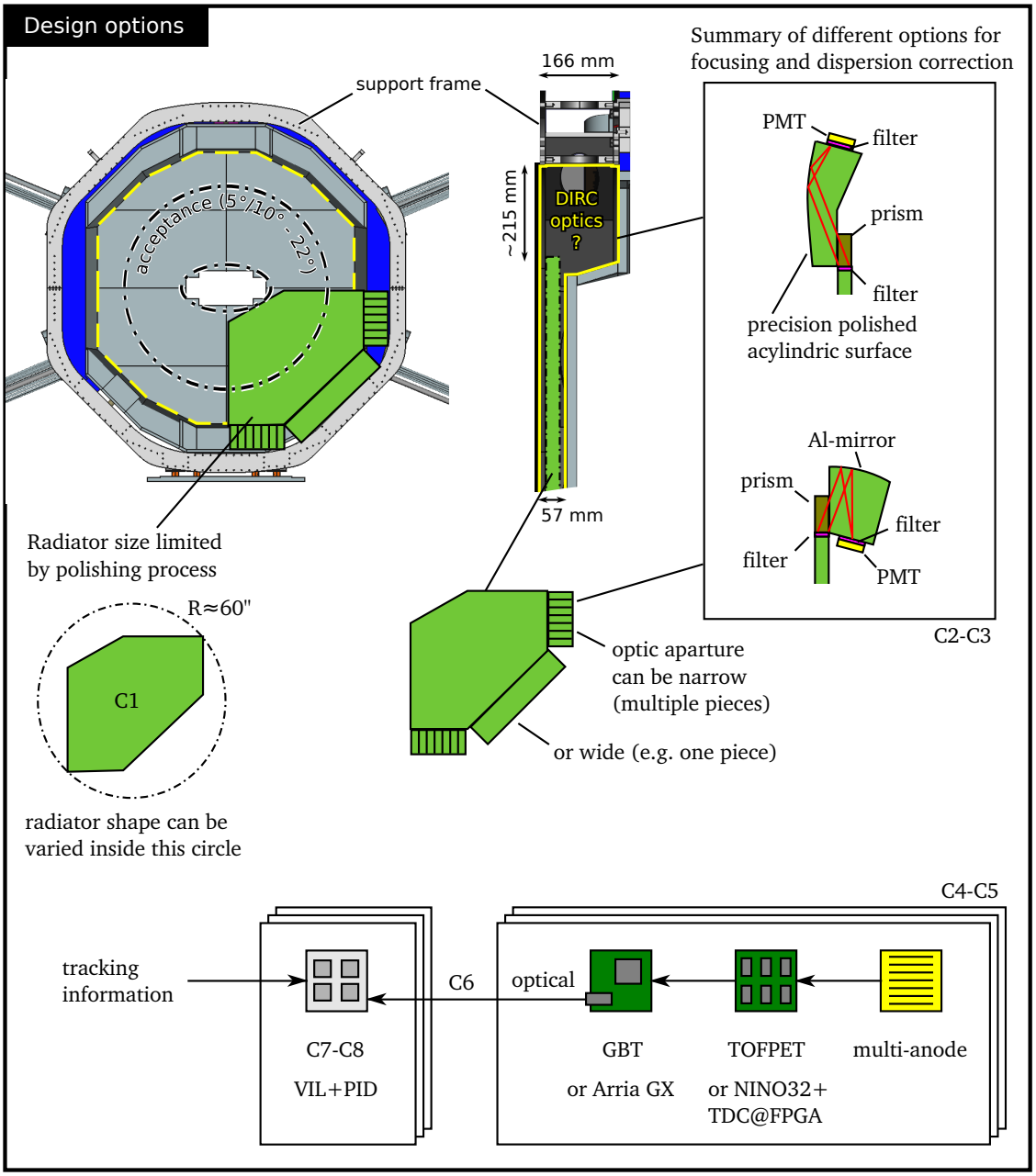


Figure 5.44.: Summary of detector design options. The constraints on the radiator size do not allow to cover the whole acceptance by one part. To avoid risks and complications implied by bonding several plates, the system has to be subdivided into four individual detector modules. Options for the focusing system are focusing mirrors realized either as acylindric total internal reflecting curvature or as cylindric aluminum mirror. Each option can utilize filters or achromatic prisms to mitigate chromatic errors. Filters can be located at the aperture or at the interface to the PMT. Prisms and filters are optional. The focusing elements can also be coupled directly to the radiator.

6. Simulation and detector modeling

This chapter starts with a brief introduction of the third party frameworks Geant 4 and PandaRoot which have been used to implement the microscopic detector simulations, followed by the description of a generic DIRC reconstruction package and a flexible optical design software which have been developed along this project.

6.1. The Geant4 transport code

Most of the detector simulations developed during this project are based on the Geant4 transport code [5, 7] which introduces itself as "*toolkit for the simulation of the passage of particles through matter*". Geant4 is a successor of the still frequently used GEANT (geometry and tracking) transport code developed at CERN. While GEANT has been implemented in Fortran, Geant4 is written in C++ based on an object-oriented architecture. It features tools for a complete description of the system geometry including the material parameters relevant to particle transport. The particle transport is completely handled by the internal tracking code, while the user can access the particle data at predefined states by implementing user specific functions similar to event handlers (inversion of control design-pattern). Such a function is called everytime when the simulation is in the specific state, e.g. before or after new secondary particles are added, or when a particle passes through a specific volume. These functions can not only be used to access the data but also to interact with the simulation, e.g. by removing particles or by associating metadata to them.

In addition, Geant 4 provides tools for visualization, serialization and fine grained control over the physics models. While Geant 4 can model nearly all relevant physical processes related to particle transport, the user can deactivate any part of the physics model. Recommended, tuned physics lists are readily available to be used for specific applications.

Typically, Geant 4 is used by writing a C++ application which is linked to the

Geant 4 libraries. The user implements the necessary initialization functions and finally transfers the control to Geant 4 by calling its run-loop. Hence, Geant 4 can be called from any C++ application. Such Geant 4 simulations have been used frequently during the design of the detector. The principle business logic is similar to the later implementation in the PandaRoot framework. However, a native Geant 4 application can access the full set of Geant 4 features. In PandaRoot, many of these features are hidden behind an abstraction layer. This lead to the decision to use Geant 4 during the design process and PandaRoot for verification of the performance under $\overline{\text{P}}$ANDA conditions. The basic principles of detector simulation are similar in both frameworks and are introduced in the next section.

6.2. The PandaRoot framework

Another way to access Geant4 is through the Virtual Monte Carlo (VMC) interface [53] which has its roots in the ALICE software project. VMC forms an abstraction layer on top of two out of the three most used particle transport codes in high energy physics, GEANT and Geant4. Support for the third package, Fluka, has been discontinued. The abstract VMC interface is provided by the ROOT software package. Geometries can be described using Root geometry. These geometries can be either converted to a native Geant4 geometry or used directly by the Geant4 transport code by means of the Geant4 VMC package. The user has to implement initialization routines and callback functions, similar to the procedure described for Geant4. The advantage of this abstraction layer is the direct comparison of results from different transport codes. However, the really interesting combination would have been a comparison with Fluka [15], which is rumored to be much more precise in case of neutron shielding applications. The trade-off in using VMC is the abdication of the flexibility provided by the Geant4 framework as the VMC interface provides only the common denominator of GEANT and Geant4. This hides many powerful features and makes the development experience less appealing.

The official $\overline{\text{P}}$ANDA simulation package PandaRoot [123] is based on VMC via FairRoot [6]. FairRoot forms basically another layer on top of the VMC interface, providing classes which permit the geometry definition and callback implementation on a sub-detector level, a common data structure and additional access to event generators (Fig. 6.1).

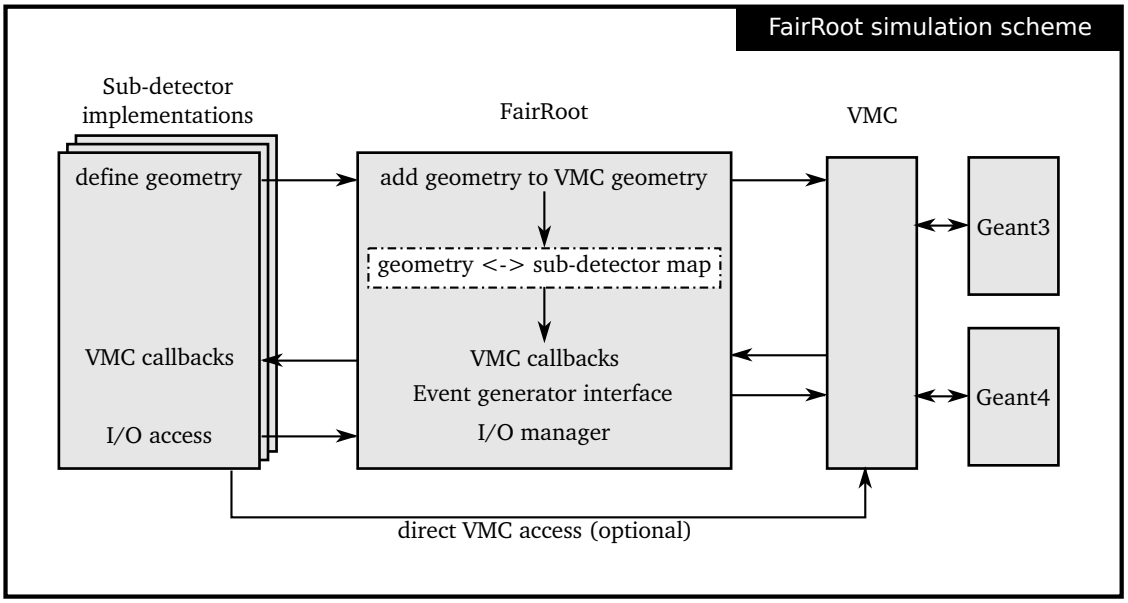
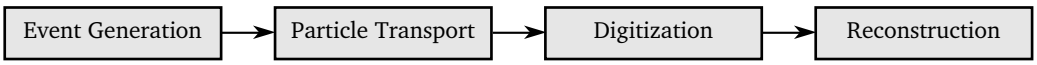


Figure 6.1.: Illustration of the FairRoot simulation scheme. Subdetectors can add their geometry via the FairRoot interface. FairRoot merges the geometries to a single VMC geometry while maintaining a map of volume identifiers to sub-detector identifiers. After initialization, the control is transferred to the Monte Carlo engine, which returns the control to FairRoot via predefined callback functions. In the FairRoot implementation of these callbacks, the volume is matched with the geometry map to find the associated sub-detector. If a sub-detector is found, the call is forwarded to its implementation. The code in the concrete sub-detector implementation is then similar to a standard VMC simulation. VMC can be accessed to obtain additional state information. The hit data is processed in the sub-detector code and written to a common root-file by means of the FairRoot I/O-manager.

Additionally, FairRoot provides a set of convenient processing features . A time-ordered buffer permits the implementation of dead-time filtering and pile-up handling. Data processing is implemented in user provided ROOT Tasks which obtain the relevant data containers for input and output from the I/O Manager. This flexible scheme allows also to mix several streams of Monte-Carlo data prior to processing, e.g. background and signal simulations.

Typical data processing steps required by virtually every detector are digitization and reconstruction. The term digitization describes the conversion from Monte Carlo hit information (e.g. time of arrival, position and energy deposit) to digital values, e.g. TDC and ADC values. At this step, the efficiency, noise and intrinsic

resolution of the detector have to be modeled. Finally, the reconstruction step involves the algorithms which compute the physical quantities of interest from the digitized detector signals.



The full simulation chain begins with the generation of events, which means the definition of the particles which have to be tracked through the geometry. Usually, one is interested in the analysis of final state particles resulting from a set of specific physics channels. These final states can be computed by several simulation codes, which are commonly referred to as *event generator*.

Two generators have been employed in the presented studies. The *EvtGen* package [68] has been used to compute the distribution of kaons from open charm channels (Fig. 5.2 and 5.1). The software needs a global decay file which specifies the decay chains for the particles involved, a *pdl* file to define the basic properties* of each particle and one optional user decay file which can be used to override settings from the first decay file. Typically, predefined decay and pdl file are used as input. These contain all important decay channels for unstable particles as well as the necessary particle definitions. The user can then specify individual decays in the user decay file. The following lines

```
Decay pbarpSystem
  1.0   D+ D- PHSP;
Enddecay
End
```

will simulate the reaction $\bar{p}p \rightarrow D^+D^-$ using the phasespace decay model for a center of mass energy provided at the command-line. EvtGen will compute the decays of the D-mesons according to the decay branches specified in the global decay file. Resulting unstable particles are further decayed iteratively until the complete final state consists of long-lived particles. EvtGen will output the complete decay tree for each simulated event.

The second event generator which has been used is the PandaRoot specific *DPM generator*. This software is developed by members of the $\overline{\text{P}}$ANDA collaboration

*e.g. mass, width, charge, spin.

and simulates the $\bar{p}p$ reaction by means of the DPM model [23] based on string fragmentation. This generator is used in $\overline{\text{P}}$ANDA to estimate the background as well as radiation load and detected rates at individual sub-detectors. The DPM generator needs only the beam momentum or center of mass energy as input.

Instead of a complex event topology, single probe tracks have been frequently used to estimate the performance of the Endcap DIRC during detector design. The simulation of 1000 tracks with equal track parameters (θ_p, ϕ_p, p, m) gives a clear picture of the detector response at a certain position. The whole response can then be mapped by simulating a set of 1000 probe tracks for several track parameters. Such a mapping results in a comprehensive figure of the detector performance over the full phase-space. The reconstruction of events with higher track multiplicity can then be studied by means of smaller sub-samples and by mixing the probe tracks with background from a DPM simulation.

During geometry definition, materials have been specified according to the data and models presented in chapters 5, 8 and 7. The high bulk absorption of the standard fused-silica material in PandaRoot has been removed. This absorption length closely resembled the attenuation length due to Rayleigh scattering computed using eq. 5.15 with parameters for fused silica from [117]. As Geant4 provides a model for Rayleigh scattering, the according parameters have been set there, so that the photon propagation is treated realistically and scattered photons remain in the radiator. The wavelength of dielectric filters has been approximated by a constant window because the Geant4 models for optical surfaces do not allow to define the complex wavelength dependence of the filter. The reflection losses at the top and bottom radiator surface have been taken into account in the digitization stage by computing the reflection loss according to eq. 5.25. This allows to change this parameter without the need to rerun the time consuming transport code.Detector specific details on the digitization and reconstruction will be covered in chapters 8 and 7.

6.3. DDRecon – a generic DIRC reconstruction package

A generic reconstruction software for plate based DIRC counters, named DDRecon*, has been developed on the basis of the DIRC model presented in section 5.3.

The software offers a set of predefined classes for the convenient description

*Disc DIRC Reconstruction

of the DIRC geometry in terms of this model, including instrumentation. The provided reconstruction algorithms comprise a pseudo-likelihood method for hypothesis based pattern matching as well as a routine for the direct computation of single photon Cherenkov angles. Both routines use the VIL algorithm which has been introduced in section 5.9 for path reconstruction. The VIL implementation has been adapted from the existing source code which had been developed by the author prior to this work [79, 80].

Both implemented reconstruction algorithms have been developed during detector design to obtain a performance figure for the detector system. The second algorithm is considered to be the choice for online PID in $\overline{\text{P}}$ANDA as it is simpler and provides a better performance, both in terms of particle identification and runtime. The algorithms will be outlined in the following two sections.

Pattern matching

The input parameters to the algorithm are:

- a particle mass hypothesis m_h
- the particle momentum vector $\vec{p}$
- the intersection of the particle track with the radiator plane (x_p, y_p)
- the measured hit-pattern

In this context, the hit-pattern is defined as a set of tuples (t_a, z, n_{ROE}) with the photon time-of-arrival t_a, the pixel coordinate on the focal-plane z as well as the id-number of the readout element n_{ROE} for a given hit. The output is a pseudo-likelihood value for the given hypothesis.

As shown in Fig. 6.2 this algorithm can be divided in four phases:

1. pattern prediction
2. reference time computation
3. hit-prediction-pairing
4. pattern matching

Pattern prediction The hypothesis m_h and track momentum p are used to compute the Cherenkov angle hypothesis $\theta_{c,h}$ at the average wavelength ($n = n(\langle\lambda\rangle)$) detected by a given readout element from eq. 3.4.

For each readout-element n_{ROE} which has been hit, all* possible projected photon paths are computed by VIL using (x_p, y_p) and the predefined detector geometry. In addition to the path-length $s_{2\mathrm{d}}$, VIL delivers α_{FEL} and ϕ_{rel}.

The momentum vector and $\theta_{c,h}$ are then used to predict the hit coordinates and arrival times for each of the hits according to the DIRC model (eq. 5.3, 5.7, 5.6). This step results in a theoretical hit-pattern of tuples $(t_{p,\mathrm{pred.}}, z_{\mathrm{pred.}}, n_{\mathrm{ROE}})$. Note that t_p is a time-of-propagation while the measured times are the time-of-arrival $t_a = t_p + t_0$.

Reference time computation To obtain the missing t_0, the measured and predicted hit-patterns are compared in z by the following procedure. Each hit in the measured pattern is compared to its multiple predictions in the theoretical pattern by computing the difference and applying a cut $|z - z_{\mathrm{pred.}}| < z_{\mathrm{thresh.}}$. The hits which have only one valid match in the theoretical pattern are then used for the computation of t_0 by computing the mean value of all differences $t_a - t_{p,\mathrm{pred.}}$.

To get rid of outliers, the hit with the largest deviation of $t_a - t_{p,\mathrm{pred.}}$ from t_0 is removed if $|(t_a - t_{p,\mathrm{pred.}}) - t_0| < t_{\mathrm{out}}$. In this case, the new mean is corrected and more outliers are removed successively.

Hit-prediction-pairing The derived t_0, which corresponds to a smeared track arrival time, is then used to compare the measured times t_a to the predicted time-of-propagation $t_{p,\mathrm{pred.}}$. Each hit in the measured pattern will be assigned to the prediction which is the closest in time-of-propagation, while applying a cut $|t_a - t_0 - t_{p,\mathrm{pred.}}| < t_{\mathrm{thresh.}}$.

Pattern matching The matching between the hits is then quantified by means of a pseudo-likelihood which assumes a gaussian probability density function $\mathcal{G}(x|\mu;\sigma)$ for each measured z or t_a with the mean $z_{\mathrm{pred.}}$ or $t_{p,\mathrm{pred.}} + t_0$ and a variance which matches the expected resolution. The combined likelihood is then given by:

$$\ln \mathcal{L} = \sum_{i=0}^{N} \left[\ln \mathcal{G}\left(z_i | z_{\mathrm{pred.},i}; \sigma_z\right) + \ln \mathcal{G}\left(t_{a,i} | t_{p,\mathrm{pred.},i} + t_0; \sigma_t\right) \right]$$

Note that this is not strictly correct, as z and t_a are not independent. The measurement error of the time at digitization is indeed independent of the spatial error, but z and t_p are correlated due to the chromatic dispersion (compare section 5.4.4). However, this correlation is not dominant in dispersion limited designs based

*The highest number of predicted rim-reflections is defined by a parameter in VIL.

on optical bandpass filters which reduce the chromatic error of both measured quantities.

Reconstruction of single photon Cherenkov angles

The input parameters to the algorithm are:

- the particle momentum vector $\vec{p}$
- the intersection of the particle track with the radiator plane (x_p, y_p)
- the measured hit-pattern

with the same hit-pattern definition which has been used in the pattern matching algorithm (PMA). The output is a set of single photon Cherenkov angles, named the Cherenkov angle candidates.

The flow chart of the algorithm is shown in Fig. 6.3. For each readout-element n_{ROE} that was hit, the photon paths in radiator-plane projection are predicted as in PMA. The single photon Cherenkov angles $\bar{\theta}_c$ are computed for each hit and every predicted path from the measured z coordinate and particle parameters by means of the DIRC model (inverse function of eq. 5.7 and eq. 5.4). The predicted path length $s_{2\mathrm{d}}$ and $\bar{\theta}_c$, t_p are returned for every hit as a set of candidates including all ambiguities.

The output of this algorithm does not involve t_0 reconstruction nor a PID approach. This has been left open for the user code. The following section describes the reference implementation for PID at the user side.

PID with single photon Cherenkov angles

A combined cut on the expected Cherenkov angles for different particle hypotheses is applied to remove outliers resulting from noise, scatter and Cherenkov emission originating from coincident tracks as well as from fast knock-on electrons. Basically, this cut rejects all unphysical Cherenkov angles. Further, this cut is used to remove ambiguities.

The expected Cherenkov angle $\theta_{c,h}$ for a mass hypothesis m_h is computed by eq. 3.4 using $n = n(\langle\lambda\rangle)$ and the momentum vector from tracking $\vec{p}$. But here $\langle\lambda\rangle$ denotes the average detected wavelength on *all* the readout elements. The Cherenkov candidates are then filtered by claiming

$$|\theta_{c,h} - \bar{\theta}_c| < \theta_{c,\mathrm{thresh.}} \text{ for at least one } \theta_{c,h}$$

with a threshold $\theta_{c,\text{thresh.}}$ which is chosen according to the detector resolution. Typically, three hypotheses are required: proton, pion, kaon. For each hit, the candidate with the minimum $\theta_{c,h} - \bar{\theta}_c$ is accepted as solution. However, this criteria is rarely used as the cut on the Cherenkov angle removes nearly all ambiguities.

The reference time t_0 is computed in the same way as in the PMA and finally used to apply a cut on the time-of-propagation of the selected hits.

The remaining sample, cleaned by Cherenkov angle and time cuts, is used to determine the track Cherenkov angle by computing the mean value $\theta_c = {}^1/_N \sum_{i=1}^{N} \bar{\theta}_{c,i}$.

PID can be realized by introducing a classifier as discussed in section 3.2. The choice of this classifier can be based on the reconstructed track Cherenkov angle distributions obtained from Monte Carlo data.

Further comments

The reference implementation for PID using the reconstructed Cherenkov angle candidates is specifically simplified to be applicable for online PID implementations. As discussed later, the final Endcap DIRC design does not require an explicit evaluation of the VIL algorithm. It will be sufficient to include only a few reflection levels, what can be directly implemented without the need of a full recon-node evaluation. Hence, the single photon Cherenkov angles can be computed using the DIRC model and then processed by three steps: a combined cut, the iterative t_0 computation and one mean value to obtain the track Cherenkov angle.

Attempts to replace the iterative outlier stripping for t_0 by a truncated mean did not yield the same PID performance. If a precise time reconstruction is required, this iteration can hardly be avoided due to its robustness. For coarse cuts in time in the ns-level, a truncated mean can be sufficient.

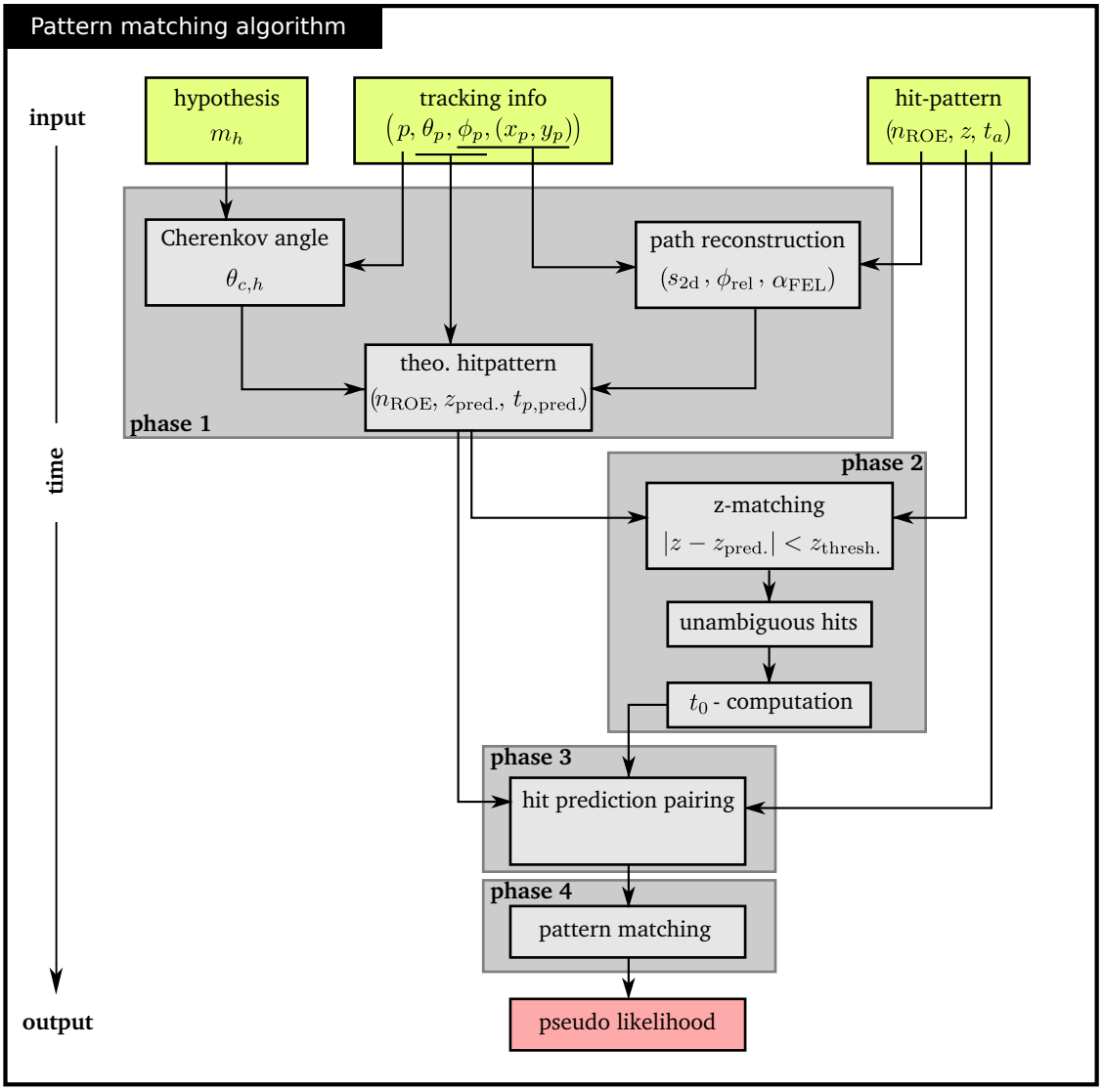


Figure 6.2.: Flow chart of the DDRecon pattern matching algorithm.

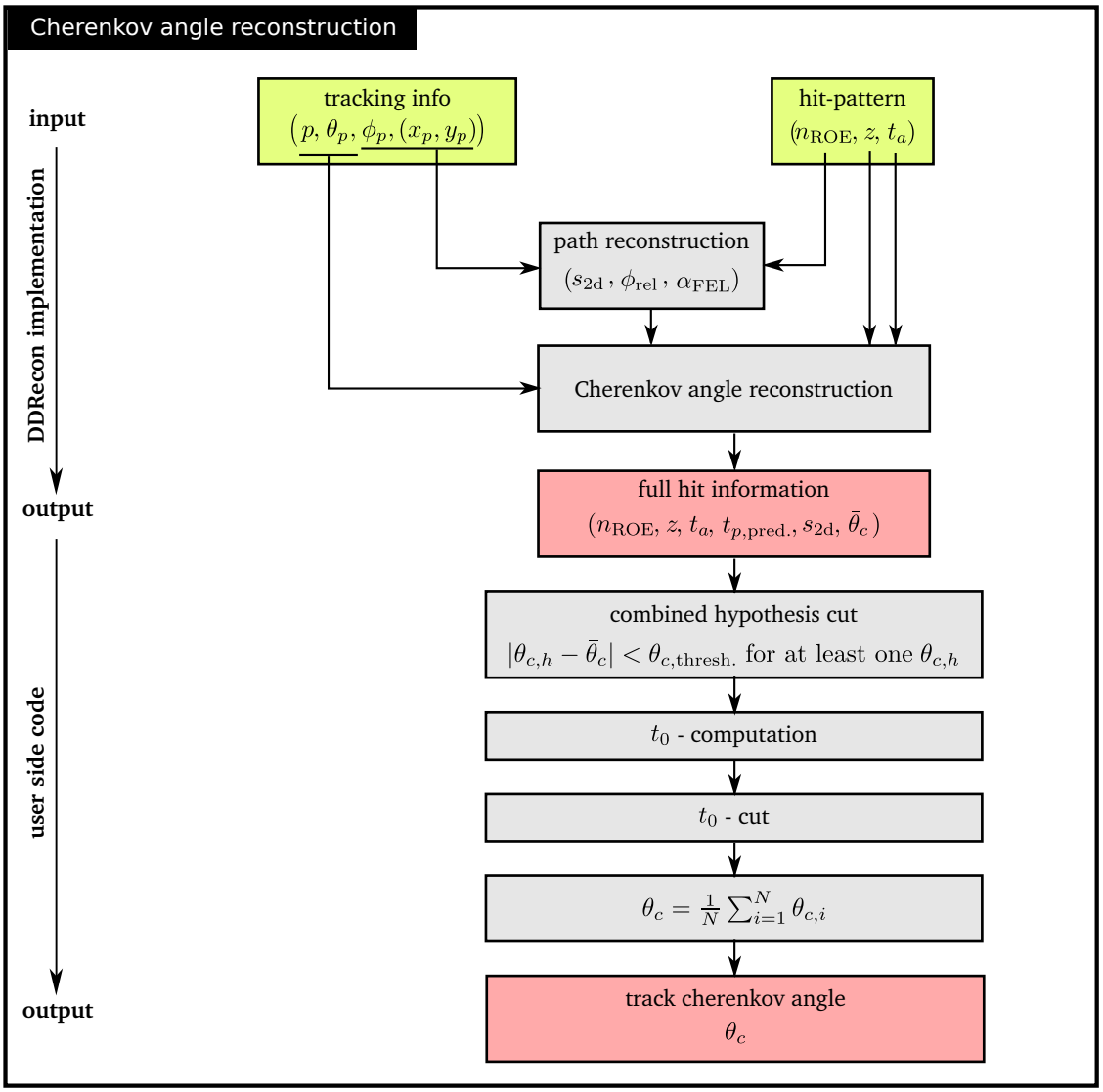


Figure 6.3.: Flow chart of the DDRecon Cherenkov angle reconstruction algorithm including user side reference implementation as explained in the text. The t_0-cut can be applied while computing the mean Cherenkov angle.

6.4. PyOptics – a tool for optical design and optimization.

As the institute did not possess a license for any commercial optics design software and declined to afford such a license, some time had to be spent on the development of a suitable optic design tool, in the following named "PyOptics". The basic requirements for the software were:

- 2D raytracing capability
- ability to process complex off-axis assemblies
- support for polynomial and "free-form" surfaces
- optimization of imaging optics
- not bound to a specific optical setup or application

Further, three principles shall guide the design of the software.

- Flexibility and usability over performance,
  as computation time is cheaper than engineering time.
- Easily adaptable and extendable software,
  as new problems frequently require to add new or change existing features.
- Functionality over GUI,
  as users are considered to have basic software engineering skills.

The programming language Python has been chosen for this project as it perfectly matches the requirements listed above. Python is by design a very clean and flexible language what comes at the expense of type safety and runtime performance. This is far outweighed by the fact that Python can be easily extended with modules written in C/C++ which permit the efficient implementation of time critical or computation intensive tasks. Another strong argument for Python is the availability of highly sophisticated scientific packages like numpy and scipy. They provide access to a variety of scientific functionality like fast and flexible handling of numerical arrays, multidimensional minimization, random numbers, a bridge to FORTRAN and many more [57, 95]. These are completed by matplotlib, a powerful package to generate scientific plots [52]. In addition to these python specific libraries, many popular packages, e.g. ROOT, provide python bindings which enable the integration of these tools right into the application.

At the time of writing, the implementation of PyOptics consists of a surface based 2d-raytracing code which is capable of handling generic on- and off-axis

optical assemblies. These may consist of an arbitrary number of elements with different shapes, bulk material and surface properties. As a fundamental design rule in PyOptics, each property of every element is allowed to be modified at runtime. This way, numerical optimization or Monte-Carlo methods can be easily connected to any feature of the optical design. This has been proven to be very useful during the development of different optimization approaches while designing the Endcap DIRC optics. The implemented surfaces are circular and elliptical arcs, arbitrary polynomials, lines and polygons.

To avoid redundancy and save development time, the PyOptics surface model is still based on ideal surfaces and does not yet account for perturbating effects like Fresnel losses and absorption or scattering at the surface and in the bulk. These effects are already part of the Geant 4 framework which provides a sound model for realistic optical surfaces. As the optical elements will be anyhow included in the full Geant 4 based detector simulation it was not a high priority to implement a more detailed surface model in PyOptics. The current model is sufficient to optimize the principle surface figure of the optical element and to determine the geometrical tolerances for manufacturing. However, a more detailed model for straylight analysis is foreseen for the future.

The software has been used for the optimization of cylindrical and polynomial "free-form" surfaces as well as for the tolerancing of optical elements. The general business logic of the optimization and tolerancing scripts is presented in figure 6.4.

Functionality which is independent of the individual optical design has been included in PyOptics. Only design specific parts have to be implemented at the client side. First of all, the individual optical setup has to be defined in terms of the PyOptics geometry model. Furthermore, a scalar merit function has to be provided for the optimization of the design.

The purpose of a merit function is to provide a numerical value which corresponds to the performance of the design under investigation. This quantity does not simply equal the resolution of the system. It is rather a combination of the resolution with all other design constraints considered to be important. Typically, these are geometrical constraints like the size of the illuminated area, but any parameter can be used. The merit function therefore quantifies the amount of agreement between an optical setup and the designers expectation. The variation of any parameter describing the optical design, e.g. surface position or radius, will affect the result of the merit function f. Thus, any sub-set of design parameters $\vec{p}$ can be considered to be a parameter of f.

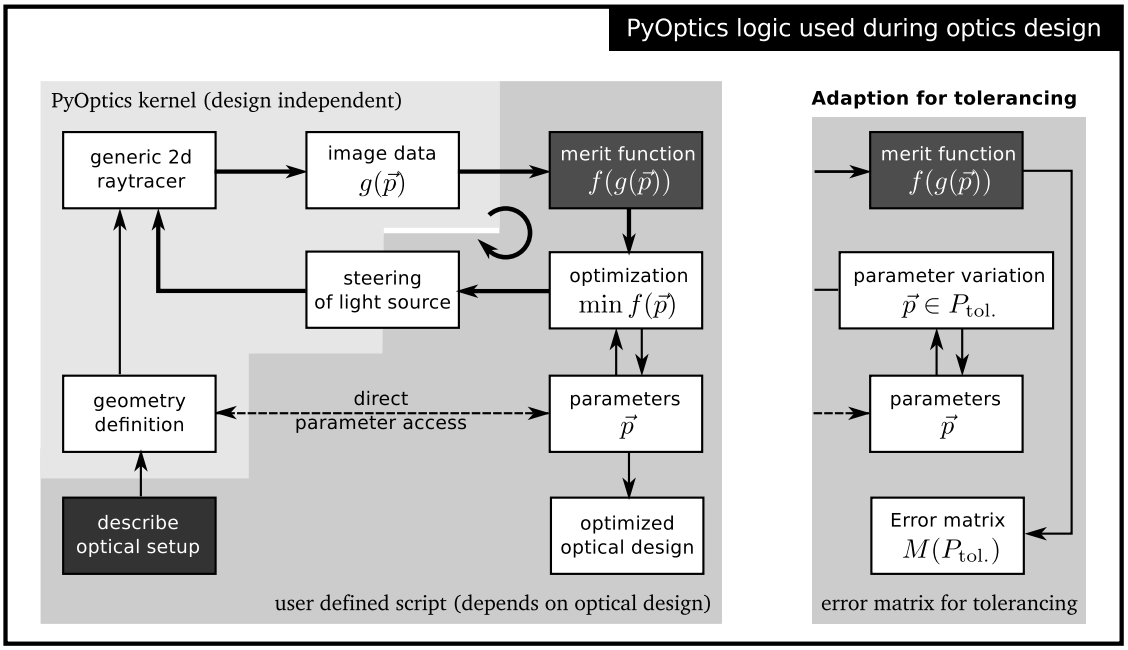


Figure 6.4.: Business logic of the PyOptics based scripts used for the optimization (left) and tolerancing (right) of optical elements. The light grey background marks functionality provided by PyOptics, while the dark grey background indicates code which has to be provided at the client side. In fact, only the geometry description and the scalar merit function have to be implemented, as any external multivariate minimizer can be used directly for optimization. The steering of the optical source is currently mixed with client code, but can be migrated to PyOptics as an independent configurable source. The code can be easily adapted to determine tolerances for the design parameters, as shown in the right panel. Errors are computed for many combinations of parameters. This error matrix can be used to query the maximum error in a given range of parameters.

The optimization of an optical design consists in the minimization of the related merit function $f(\vec{p})$. Because PyOptics permits access to all design parameters at runtime, any standard multivariate minimization algorithm can be used, e.g. those provided by scipy.

In most nontrivial cases, the prior condition of a successful multivariate optimization is a good guess of the initial parameters. If the start-parameters are far away from the global minimum, a higher risk exists that the fit will converge to a local minimum, thus delivering a sub-optimal solution. In case of the optimization of an optical design, the start-parameters correspond to the initial geometry.

For the Endcap DIRC optics, polynomial shapes have been studied as well as cylindrical shapes. Especially for the polynomial shapes, the determination of the start-parameters is non-trivial. The method developed to determine adequate start-parameters is explained in the appendix (A.1), followed by an algorithm to approximate the surface by elliptical arcs (A.2). The latter algorithm had to be developed to obtain an efficient representation of the polynomial surface in Monte-Carlo.

Part III.

Performance analysis

7. A dSiPM based 3D Disc DIRC design study

In the earliest stages of the presented research, the sensor options in components table 5.11 were the exact opposite. Lifetime enhanced MCP-PMTs with Al-layer had an insufficient extracted charge limit of $Q_{\text{max}} \lesssim 1\,\text{C/cm}^2$ and ALD-enhanced tubes were not available. On the other hand, the radiation hardness of the Philips dSiPM had not been measured.

Even though the SiPM technology is known to be susceptible to radiation damage (section 5.5.1), there were some strong arguments why dSiPMs could be a viable sensor choice for the use in the Endcap DIRC environment. The following arguments are based on the technical and physical details which have been discussed in section 5.5.

One argument for dSiPMs is their high internal granularity which can be exploited for imaging by modifying the ASIC as proposed by Philips [42]. The high spatial resolution allows to use a smaller active area what leads to a more compact optical design, but also to a lower global* dark count rate.

Another argument is the unique feature to activate or deactivate every single GAPD-cell on a dSiPM-die. Data on the irradiation of SiPMs show clearly that these devices are much more sensitive to irradiation with neutrons [105] than to γ irradiation [148]. This indicates that displacement damage is the major source for an increase in DCR and that surface effects at the Si-SiO_2 interface play only a marginal role, especially for radiation conditions expected at the Endcap DIRC readout (table 5.4). Displacement damage is usually localized to a volume well below the cell-size. A damaged cell, showing a significantly higher dark count rate (DCR), would increase the occupancy of the sensor and thus decrease the detection efficiency. The dSiPM however allows to switch off individual GAPD-cells. By deactivating all damaged cells, the displacement damage does not lead to a higher DCR but to a lower collection efficiency which increases proportionally with the

*global quantity = value for the whole system

number of damaged cells. This is appreciable because the signal-to-noise ratio remains unaffected.

In the end, the applicability of this sensor depends only on the timescale on which its efficiency decreases below a critical percentage which is not anymore compatible with the system requirements.

These arguments and the lack of rigid alternatives to the dSiPM option led the initial development in this direction while the MCP-PMT option has been considered as an alternative in case of a significant improvement of the sensor lifetime. The resulting dSiPM design study is still interesting as it demonstrates that a 3D-DIRC can yield stable performance, even in presence of the very high dark count rate of silicon photomultipliers.

7.1. Detector design

As explained in chapter 4, the dSiPM based detector design is an ancestor of the Time-of-Propagation Disc DIRC and forms a synthesis of the ToP and imaging approach. To correct for the chromatic error in both measurements simultaneously, the wavelength range has been restricted to two wavelength bands of 100 nm width, similar to the prior ToP design. In contrast to previous design attempts, the whole detector has been divided into four independent sub-detectors, following the arguments which have been laid out in section 5.10. An overview of the system is presented in Fig. 7.1 and 7.3.

The four radiators form an octagonal shape which follows the symmetry of the $\overline{\text{P}}$ANDA solenoid and the shape of the EMC frame. A radiator thickness of 20 mm has been chosen to maximize the photon yield. Each radiator is instrumented with 108 readout elements consisting of a symmetric focusing element with a filter and mirror coating and two 2×2 arrays of dSiPM sensors as shown in Fig 7.3. The setup uses two bandpass-filters: 400 nm–500 nm (A) and 500 nm–700 nm (B). The filter coatings are applied to the surface of the readout elements which will be coupled to the radiator. Each readout element has a different filter than its direct neighbors, leading to an alternating filter sequence along the radiator rim.

This alternating sequence of filters has the purpose to extend the path-length for a fraction of the detected photons. About half of the photons emitted in the wavelength range 400 nm to 700 nm is reflected at a filter (e.g. the red track in Fig. 7.3), while the other half enters a readout element (green and blue tracks).

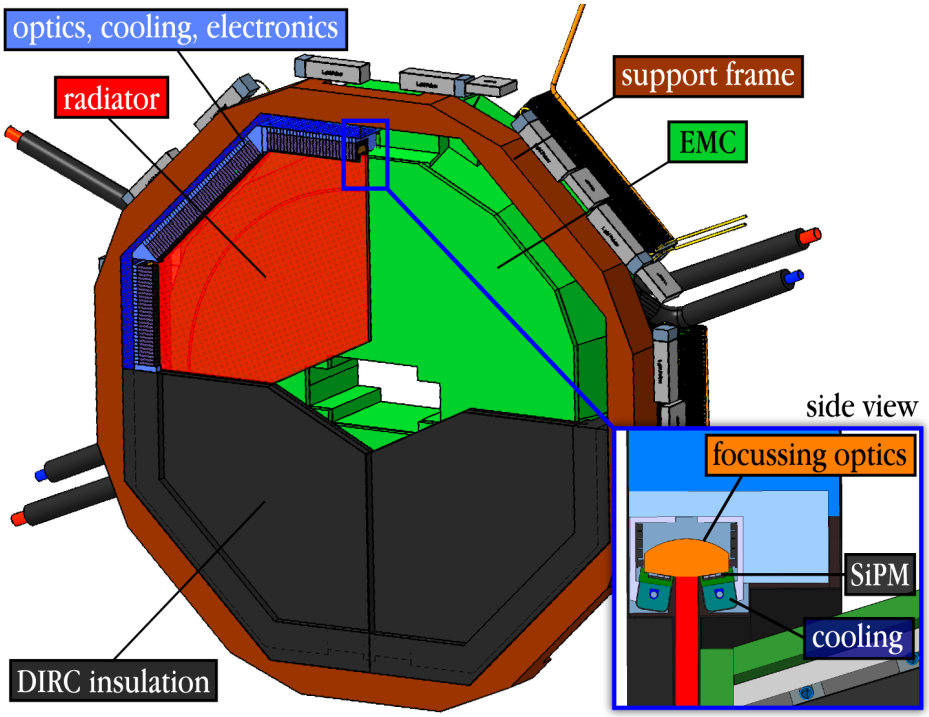


Figure 7.1.: Simplified overview of the dSiPM based Endcap DIRC design.

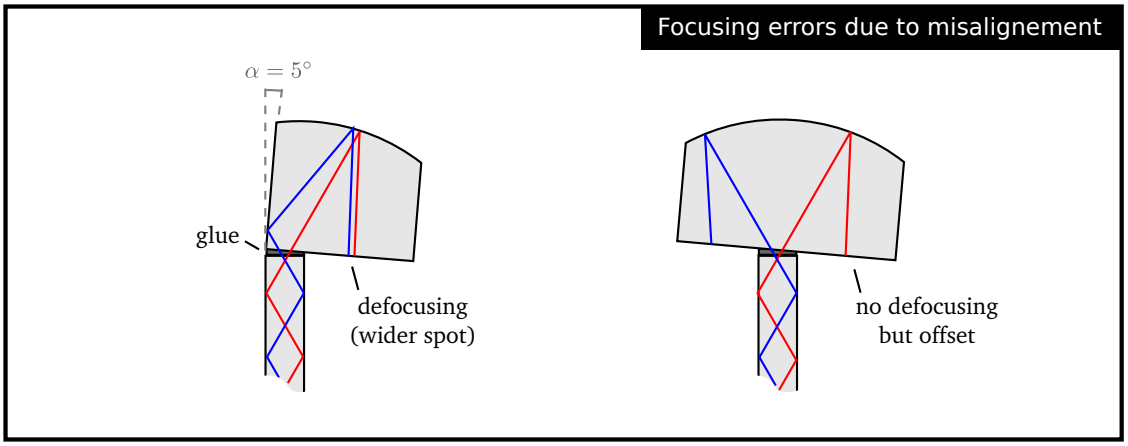


Figure 7.2.: The misalignment of asymmetric focusing elements (left) causes an additional blurring of the image because the reflected and not reflected rays are focused on a slightly different position. This can be avoided by choosing a symmetric setup (right) where the misalignment causes an offset at both sides which can be easily corrected during analysis.

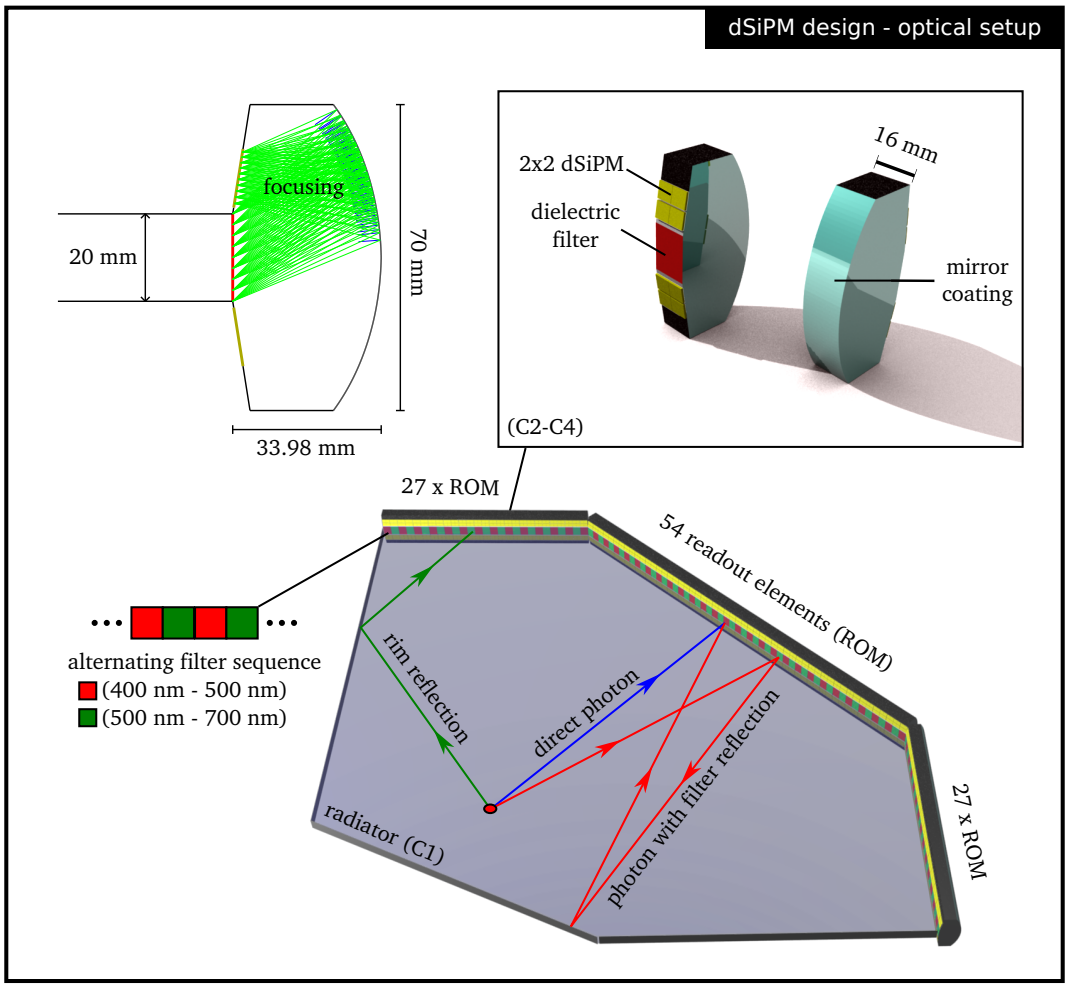


Figure 7.3.: Optical setup of the dSiPM based Endcap DIRC design. The radiator is instrumented with 108 readout modules (ROM) which consist of a 16 mm wide symmetric focusing element, coupled to two 2×2-arrays of dSiPMs. A dielectric multilayer coating is applied to the entrance side which will be coupled to the radiator. Two kinds of filter coatings are used to increase the average photon path length and arranged in an alternating sequence along the radiator. The focusing surfaces are coated with Aluminum. In total, a sub-detector is equipped with 864 dSiPMs.

Those of the reflected photons which return to the rim have again a probability of about 50 % to hit the correct filter to be detected (red track). Longer photon paths have the advantage of an increased propagation time what leads to a smaller relative error $\sigma_t / t_{\text{prop.}}$ as the time resolution σ_t of the sensor is constant.

The choice of a symmetric imaging system, which becomes practicable due to the small size of the dSiPM sensors, prevents a potential defocusing due to misalignment of the optical element. Such a defocusing can occur in non-symmetric DIRC optics which reflect one part of the light at a plane surface prior to focusing (Fig 7.2). If this plane surface and the radiator are tilted by an angle α, the reflected and not reflected photons will differ by an angle of 2α due to the reflection. Hence, the reflected part will be focused to a different position on the focal plane.

At the time of the dSiPM studies, PyRecon had not been developed. The shape of this specific focusing surface has therefore been optimized by Klaus Föhl, who had an existing Physica code for the optimization of polynomial surfaces*. Hence, the shape of this focusing element is acylindric. From a practical point of view, this is a disadvantage as most manufacturers showed no interest in a prototype development and rejected this shape right away. However, there are some smaller companies specialized on custom products who are willing to prototype such shapes. In the given case, the acylindrical optics yield a much better performance than cylindrical optics. The acylindrical focus is about 20 μm RMS if form errors are ignored. A similar cylindrical solution would result in a focus of 125 μm RMS†.

7.2. Performance studies

The detector has been modeled in Geant4 and PandaRoot to simulate the systems response to probe tracks in a clean environment as well as in the $\overline{\text{P}}$ANDA environment. Geometry and event displays are shown in Fig. 7.4. Key parameters of the detector model are summarized in table 7.1.

In the implemented model, the cells of the dSiPM matrix are grouped into 64 μm

*The same code which has been used to compute the polynomial surfaces for the Endcap DIRC Design proposed by Glasgow/Edinburgh [39].

†These values for cylindrical surfaces result from an optimization with PyOptics. However, a design where the light bundle which leaves the radiator in left (right) direction is focused on the right (left) dSiPM might lead to an improved performance for cylindrical surfaces. A study of this setup has been omitted as the dSiPM design is already deprecated.

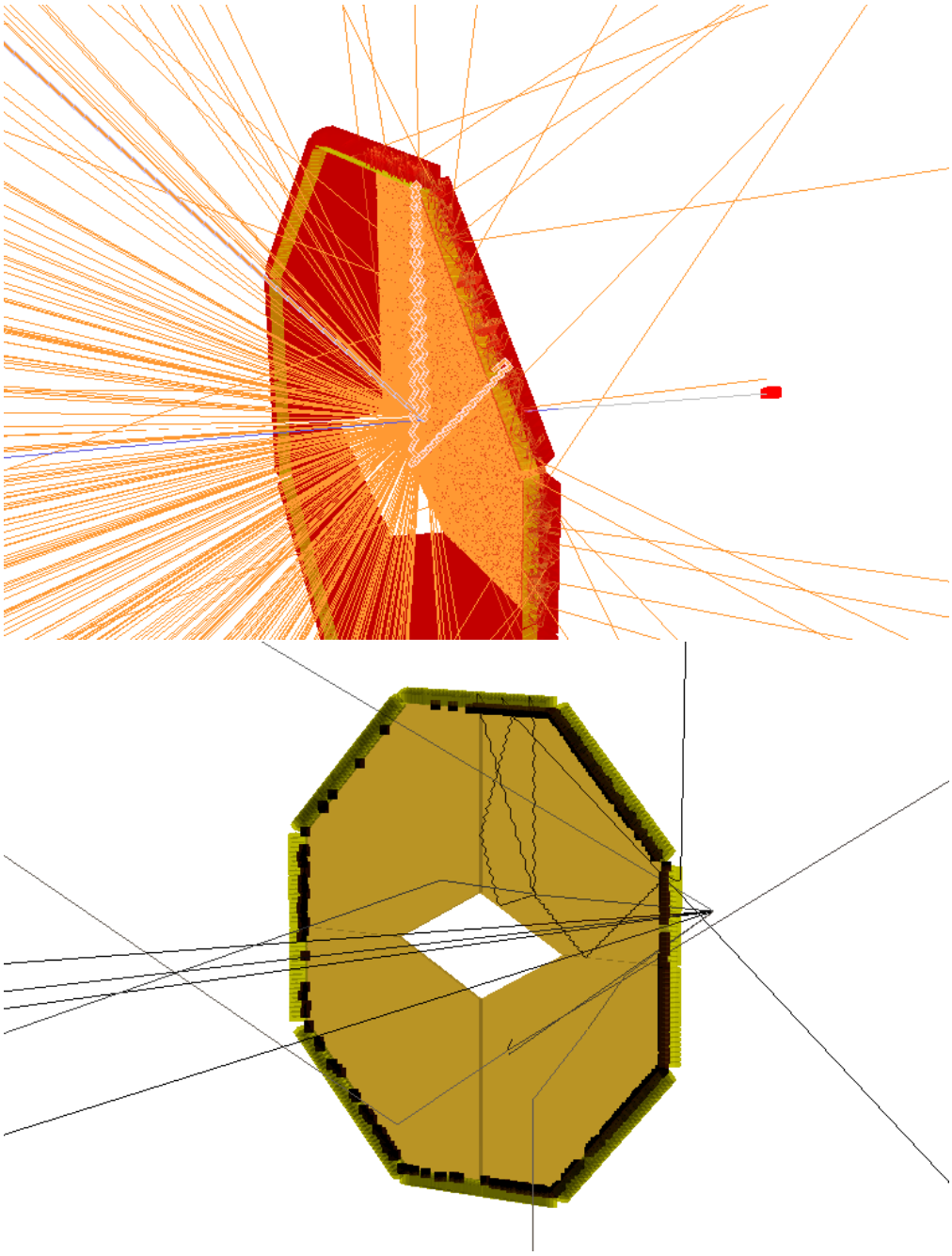


Figure 7.4.: Event display of charged particles passing the dSiPM based Endcap DIRC. The upper figure includes all photon paths. The lower figure shows only some selected photon tracks inside the radiator.

Time resolution (RMS):	60 ps
TDC binning (LSB):	10 ps
Pixel line height:	64 μm
Active cell ratio:	90 %
Per die dead time:	100 ns
Dark count rate per GAPD:	25 Hz
dSiPM temperature:	0°
Surface roughness:	1 nm

Table 7.1.: Simulation parameters of the dSiPM design.

high logical pixel-strips to emulate a modified Cherenkov readout of the sensor. Timing characteristics, dark count rate and PDE equal those measured for the existing dSiPM. A plot of the PDE curve is included in Fig. 5.30 on page 122. Due to the limited capability of Geant 4 to model the angular dependency of the filters, these have been implemented as angle-independent filter. The relevance of the angular dependency will be further discussed in the subsequent section.

A typical 3d hit-pattern resulting from a Geant 4 simulation of the detector is shown in Fig. 7.5a (red) together with the pattern prediction computed by the pattern-matching algorithm (section 6.3). The pattern-matching serves also as cross-check for the simulation as the reconstruction code is completely independent from Monte-Carlo. Markers further away from the prediction originate from photons emitted by knock-on electrons or scattered light. Dark counts are not shown in the figure.

Fig. 7.5b shows the π/K–misidentification obtained by reconstructing 1000 patterns for each particle species. These results correspond to clean patterns without dark count and coincident particles as shown in Fig. 7.5a. Further, the theoretical patterns have been computed from the true track information without the consideration of tracking errors. The misidentification is extraordinary low in this case, what is a consequence of the missing tracking error paired with the extremely high spatial resolution of the optical system and a photon yield of ~ 45 photons per track.

Such simulations of clean patterns have been used to tune the reconstruction algorithm and the design, e.g. the width of the focusing elements. Subsequently, the missing background sources have been added to the simulation to obtain a more realistic figure of the detector performance under final conditions. Fig. 7.6a

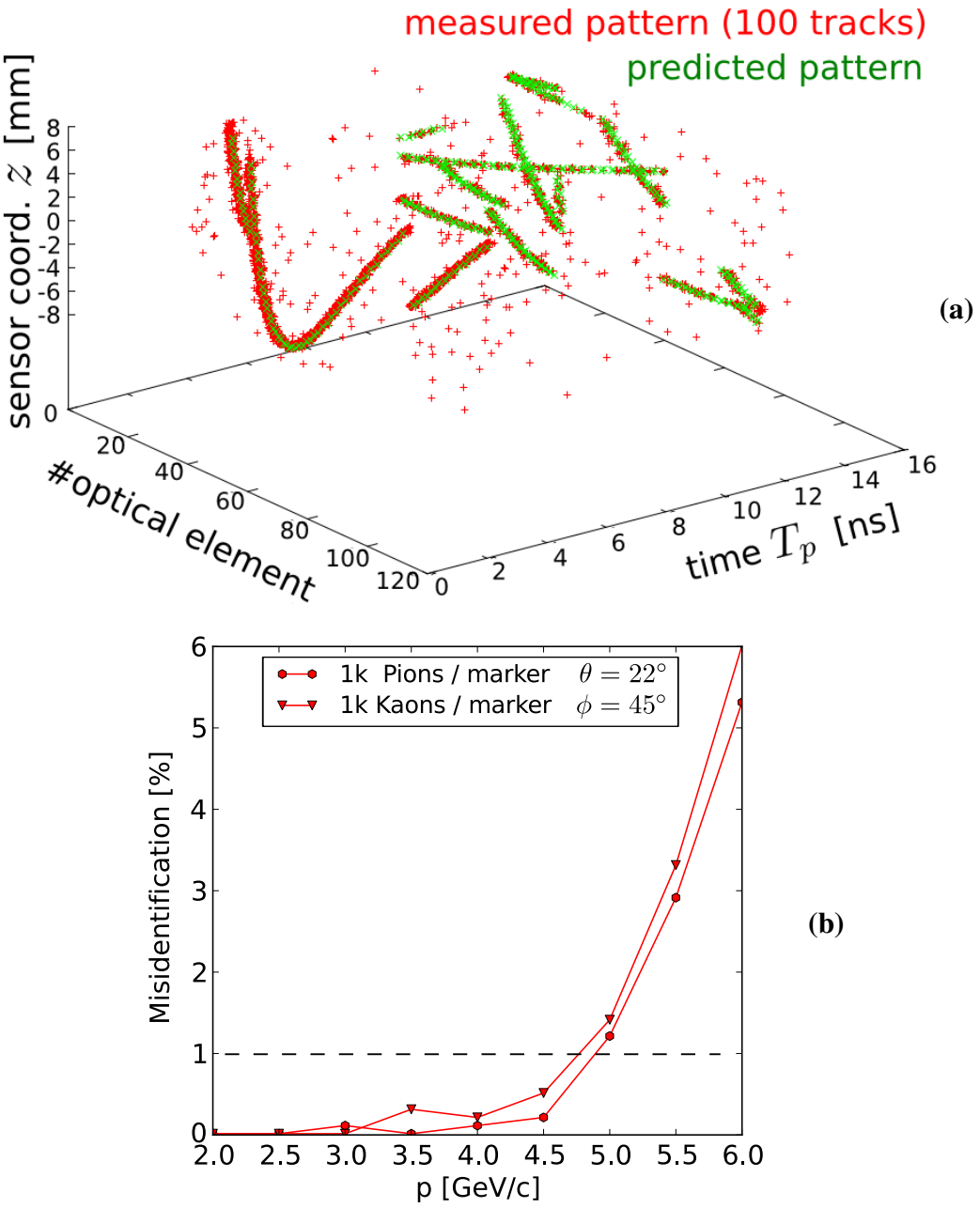
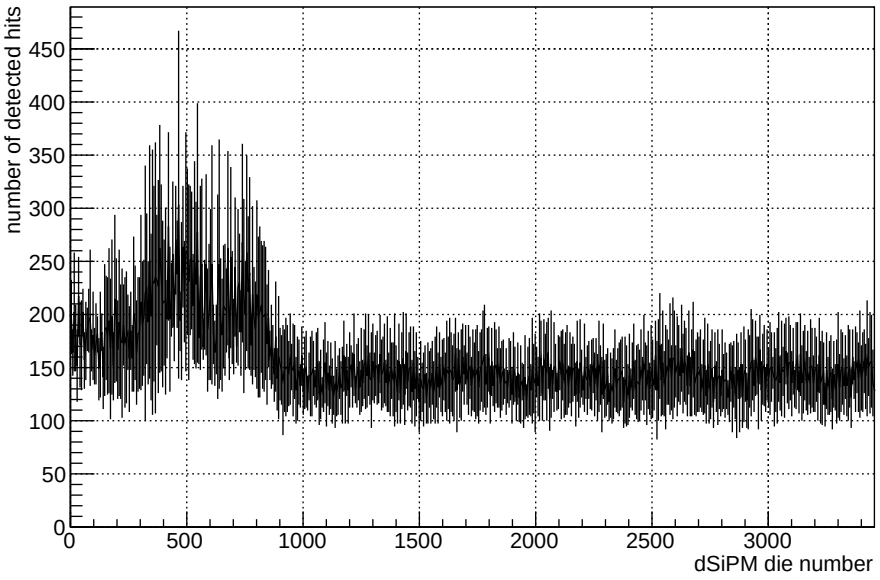
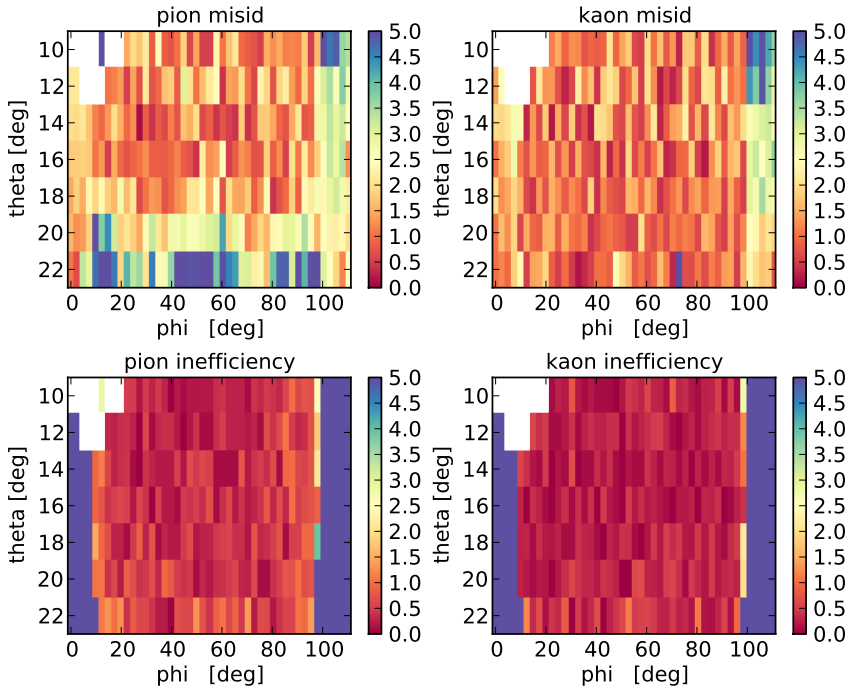


Figure 7.5.: (a): Simulated photon-hit pattern vs. hit-pattern predicted by the pattern matching algorithm discussed in section 6.3. (b): Misidentification obtained by reconstructing 1000 pion and 1000 kaon probe tracks with $\theta_p = 22°, \phi_p = 45°$ at different momenta by means of the pattern matching algorithm [81]. These results do not include dark counts, showers and background from coincident particles.



(a)



(b)

Figure 7.6.: Results from time-based simulations in presence of the full $\overline{\text{P}}$ANDA geometry. The data include dark-counts and coincident tracks from $\bar{p}p$ interactions. (a): Number of detected hits per dSiPM in a time-frame of 109 μs. The higher rate in the first sub-detector originates from probe tracks. (b): Results obtained by reconstructing 1000 probe tracks for each θ,ϕ combination at $p = 4$ GeV/c. The detector covers the ϕ range of 90°. The blue areas in the inefficiency plot indicates that $\geq 5\%$ of the tracks cannot be reconstructed. Tracks emitted under these angles cannot reach the sub-detector directly. The reconstructed tracks have been deflected.

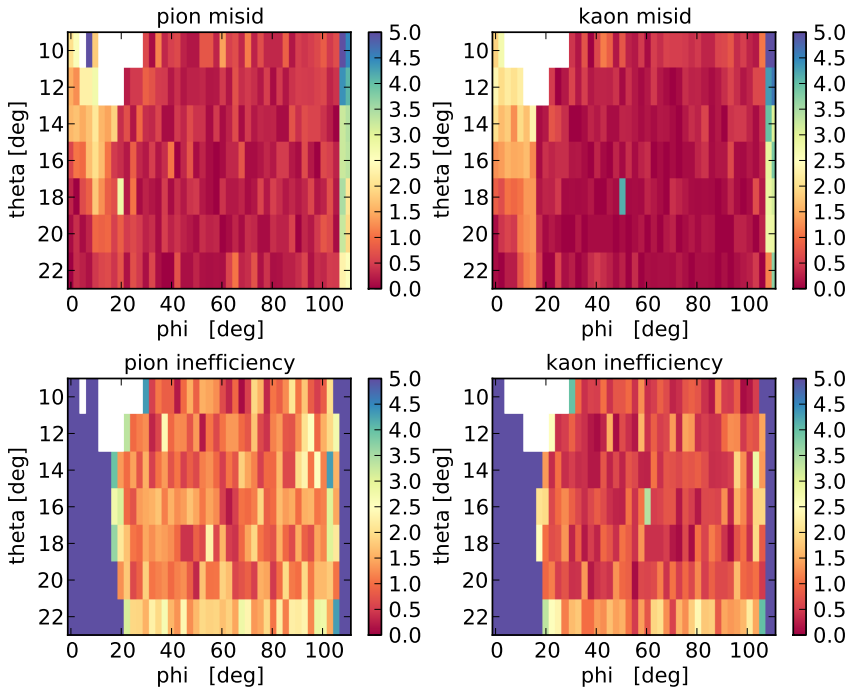


Figure 7.7.: Reconstruction results for 2 GeV/c. See caption of Fig. 7.6 for details.

shows the number of hits on individual dSiPM dies from dark counts, probe tracks and DPM-background for a data frame of 109 μs duration. These results have been obtained by performing a time-based simulation with the PandaRoot framework while including the material of all detectors within the target spectrometer, the beam pipe and the magnet yoke. Background from the $\bar{p}p$–interaction has been modeled using the DPM generator. A beam momentum of 15 GeV/c has been chosen to select the highest Lorentz-boost. At the digitization and event-mixing stage, the time between consecutive background events is randomly generated using an exponential distribution with a mean time of 50 ns. Hence, the mean rate is equal to the 20 MHz rate requirement as specified by the $\overline{\text{P}}$ANDA collaboration [109]. However, the simulated background events include only inelastic collisions, what leads to an overestimation of the background of 20 % according to the ratio of the elastic and total hadronic cross sections.

The dSiPMs are numbered per readout element (1–8, 9–16, …) along the rim of the whole detector. Hence, the dSiPMs 1–864 belong to the first, 865–1728 to the second, 1728–2592 to the third and 2593–3456 to the fourth sub-detector.

This dataset contains also hits from 1000 probe tracks with equal four-momentum vector, spaced 100 ns in time (10 MHz). Pile-up is emulated collectively for hits from probe tracks, background and dark counts. Probe tracks are later used to determine the reconstruction efficiency and PID performance for the given track parameters in presence of the background.

The number of hits from the simulation in Fig. 7.6a can also be estimated from the parameters in table 7.1. The expected raw dark count rate per dSiPM with 4×6400 GAPD cells* is $R_{\mathrm{DCR}} = 640\,\mathrm{kHz}$ at 25 Hz per SPAD cell. This DCR corresponds to a die temperature close to 0°. The expected raw signal hit rate per dSiPM at a 20 MHz interaction rate with 3 incident tracks per event is then:

$$R_{\mathrm{DPM}} = 20\,\mathrm{MHz} \cdot 3\,\mathrm{tracks} \cdot 44\,\mathrm{photons/track} \,/\, 3456\,\mathrm{dSiPMs} \approx 764\,\mathrm{kHz}$$

The expected background rate from the experiment is hence at a similar level as the dark-count rate. At a dead time of 100 ns, one expects a signal loss of

$$1 - e^{-(R_{\mathrm{DPM}} + R_{\mathrm{DCR}}) \cdot 100\,\mathrm{ns}} \approx 13\,\%$$

Based on these estimations, one expects 133 detected hits in a 109 μs frame what is in agreement with the simulation results in Fig. 7.6a. Note that the higher count rate in the first sub-detector originates from the probe-tracks.

The reconstruction of probe tracks on background has been performed for the first sub-detector. For every θ,ϕ combination shown in the colormaps in Fig. 7.6b, a sample of 1000 probe tracks has been generated for pions and kaons at 4 GeV/c in presence of background†. The term pion/kaon-inefficiency in Fig. 7.6b describes the percentage of tracks which were not reconstructed. This can happen if the algorithm rejects too many photons as outliers or if the number of detected photons is low. The blue areas are outside of the quarter detector, which covers a ϕ-range of 90°. Some tracks can still be reconstructed in this region because they are deflected by scattering. The track information for the reconstruction is generated from Monte-Carlo truth by smearing the angle by a gaussian with $\sigma_{\theta,\phi} = 1$ mrad. The position of the track is also smeared by $\sigma_{x,y} = 2.5$ mm.

Up to $\theta_p = 18°$ the misidentification at 4 GeV/c is better than 2 % on average. Judging from the shape of the kaon distribution in 5.2a, this performance is compatible with the system requirements. At 2 GeV/c, the misidentification is better than 2 % over the whole acceptance (Fig. 7.7).

*actually there are 6396 cells per logical sub-pixel, but the missing 4 cells have not been included in the detector model.

†Fig. 7.6a has been generated from one of these samples.

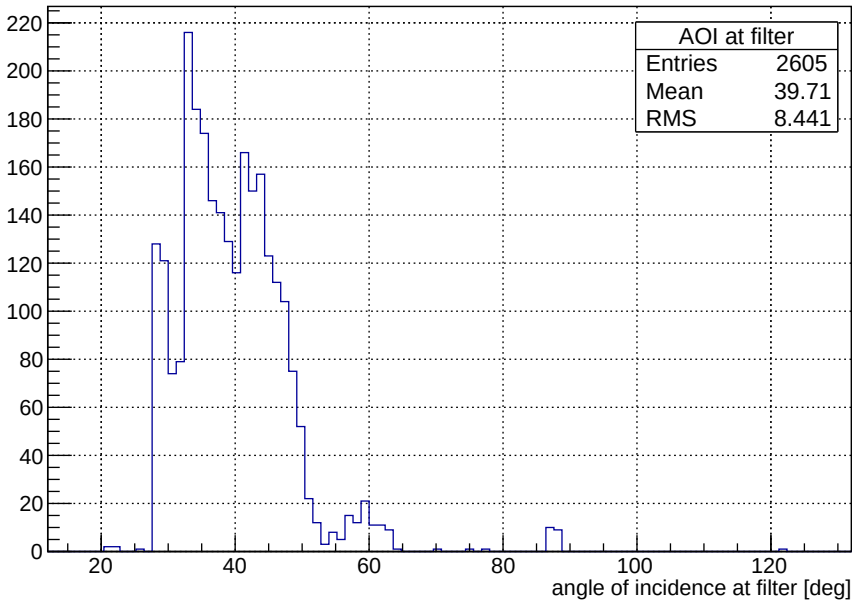
7.3. Design flaws

The presented dSiPM based design has still some flaws which would have to be corrected. The angle of incidence (AOI) at the bandpass-filters ranges from about 20° to 60° with an average value close to 40 ° (Fig. 7.8). Due to the strong AOI dependence of these filters, this detector cannot be built using standard filters. It is also questionable whether such a filter can be designed at all. Even though some filters are optimized for an AOI of 45°, the angular dependence of the filters increases with higher incident-angles. Hence, these filters are more sensitive to AOI changes than the standard filters which are optimized for normal incidence.

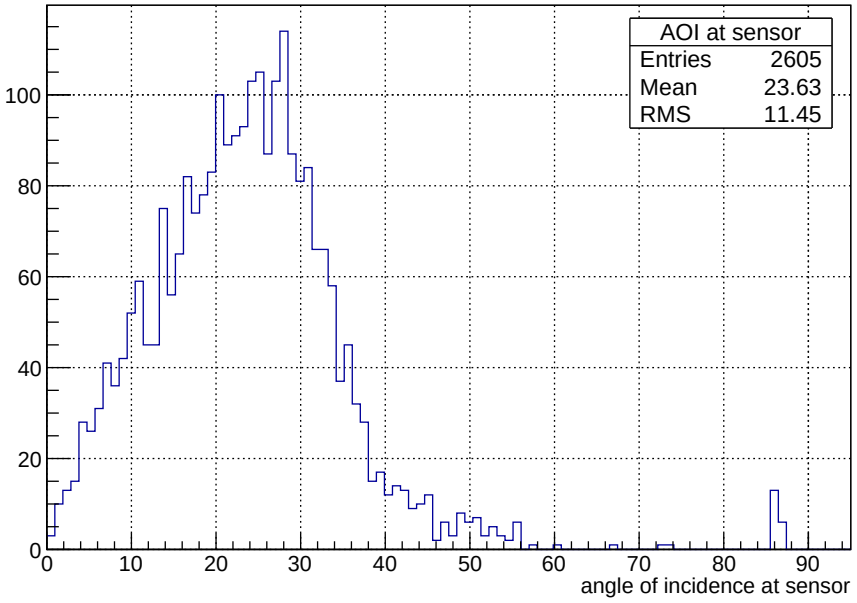
The AOI distribution at the sensors however is compatible with standard filters like those shown in Fig. 5.21 on page 101. Hence, it would make sense to move the filters to the sensor location and to abandon the mechanism to prolongate the photon paths. In this case, it is sufficient to use only a single wavelength-band.

Further, the light emission of dSiPMs could cause optical cross-talk between individual sensors through the optics. This effect has not been studied in detail. Filters in front of the dSiPM could help to diminish such effects, because the APD emission spectrum is stronger at longer wavelengths.

For the reconstruction of clean patterns, such a single band version of the detector yielded similar results as the two-band version. However, these studies have not been continued at this point as it turned out that the dSiPMs would survive only a few days in the $\overline{\text{P}}$ANDA environment. Nevertheless, the design principle could be used in an environment where the sensors are not exposed to a high 1 MeV-neutron equivalent fluence.



(a)



(b)

Figure 7.8.: (a): Photon angle of incidence at the bandpass filters located at the ROM-radiator interface. A filter design which covers such a wide range of steep angles is extremely challenging, if not impossible to realize. (b): The AOI distribution at the sensor area is closer to normal incidence. Such distributions can be handled by standard filers like the ones shown in Fig.5.21 on page 101.

8. An MCP-PMT optimized design option

Roughly at the same time as the dSiPM option turned out to be insufficiently radiation tolerant, the MCP-PMT lifetime has been improved due to the introduction of ALD-enhanced MCPs. The increase from $Q_{\text{max}} = 1\,\text{C/cm}^2$ to $6\,\text{C/cm}^2$ opened up the possibility to design an MCP-PMT based solution by matching the count rate to the lifetime.

8.1. Detector design

This section covers the details about the components C1–C5 which have been left open in section 5.10. An overview of the identified design options has been sketched in Fig. 5.44 (p. 166). One of the more challenging tasks during the development of the MCP-PMT based detector option was to design the optical system in a way that all of the constraints are met. The most important constraints are given by the alignment of the sensor to the magnetic field and the geometrical constraints as defined by the EMC insulation and support frame. Other design constraints arise directly from the sensor choice which has been motivated in the preceding sections. The 2" Planacon MCP-PMT has an active area of $53 \times 53\,\text{mm}^2$. Hence, in the imaging plane projection of the optics, the light bundles which enter with an angle $\varphi' \in [\varphi'_{\text{min.}}, \varphi'_{\text{max.}}]$ have to be focused to a line segment of 53 mm. However, this length has been reduced to 50 mm during optimization to allow a certain displacement of the image due to the interaction of the magnetic field with the charges inside the PMT (compare Fig. 5.37 on page 138). Hence, a linear image displacement up to 3 mm can be tolerated. The offset of the displacement has to be corrected by positioning the tube accordingly.

Due to the complications which arose during prototyping of the acylindric dSiPM optics, strong focus has been put on the research and comparison of various optical setups. The goal was to find a solution which provides precise imaging

while staying compatible with standard manufacturing processes. However, polynomial optics have been optimized as well to provide a complete and objective comparison.

The availability of radiation tolerant optical materials and the impact on the optical design of the Endcap DIRC have already been discussed in chapter 5. One of the consequences is that achromatic lens-systems can hardly be designed due to a lack of rad-hard materials with higher refractive index and good transmission in the wavelength range covered by bialkali photocathodes. Hence, the remaining options are focusing mirrors, either coated with aluminum or optimized for total internal reflection. Both can be combined with filters and prisms to provide additional dispersion correction, as discussed in section 5.4.4.

The optical curvatures have been optimized by defining an initial geometry and a merit function in PyOptics. The merit function is simply defined as the maximum spot width obtained by focusing several bundles of parallel light onto the imaging plane. Such a set of bundles is shown in Fig 8.1. The bundle with the widest spot defines the figure of merit. More complex merit functions have been tested, e.g. using the average of all spot widths or the rms of the point spread function. However, these functions did not lead to any improvement.

PyOptics validates the ray propagation at every step. In case of a material mismatch or rays which leave the setup without being focused, the merit function returns a very high penalty value. This confines the optimization algorithm to the physical meaningful geometry range. Further merit penalties are introduced to account for design constraints, e.g. the geometrical limits for the focusing surface, the possible sensor positions and the size of the active area on the focal plane.

Prior to an optimization run, the aperture as well as the focal plane position and tilt have been fixed. Initial surface parameters for the optimization have been determined by means of the method explained in section A.1 or, in case of cylindrical optics, by computing the focal length. The multivariate optimization of this initial surface is handled by a downhill-simplex algorithm which minimizes the merit function.

Fig. 8.1 shows two optimized focusing elements (FEL), TIR1 and TIR2, similar to those proposed by Klaus Föhl [39]. However, TIR2 has been optimized for an restricted angular range of 21°–41° instead of the full range of 21°–47°. The corresponding imaging performance is presented in Fig. 8.2. A comparison with the performance of the final cylindrical optics (Fig. 8.4) as shown in Fig. 8.3 indicates that the cylindrical design clearly outperforms TIR1 and TIR2.

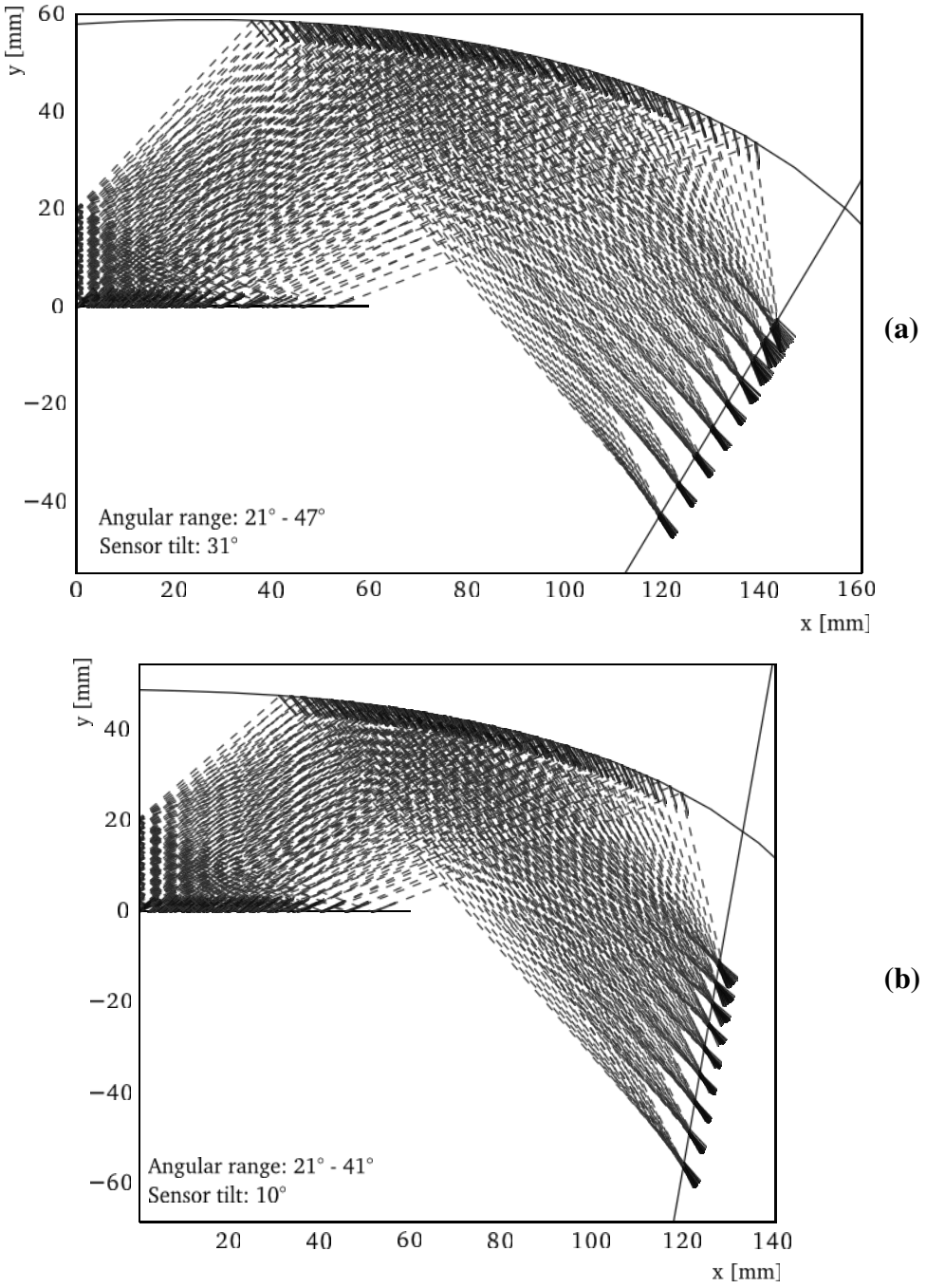


Figure 8.1.: Geometry output from PyOptics for the two representative polynomial optics TIR1 (a) and TIR2 (b) discussed in the text. TIR1 provides a very good alignment of the sensor to the field, at the cost of lower resolution and larger size. TIR2 is an example of a high resolution polynomial optics for a reduced angular interval. It is compatible with the geometrical constraints, but the alignment of the sensor to the field is critical.

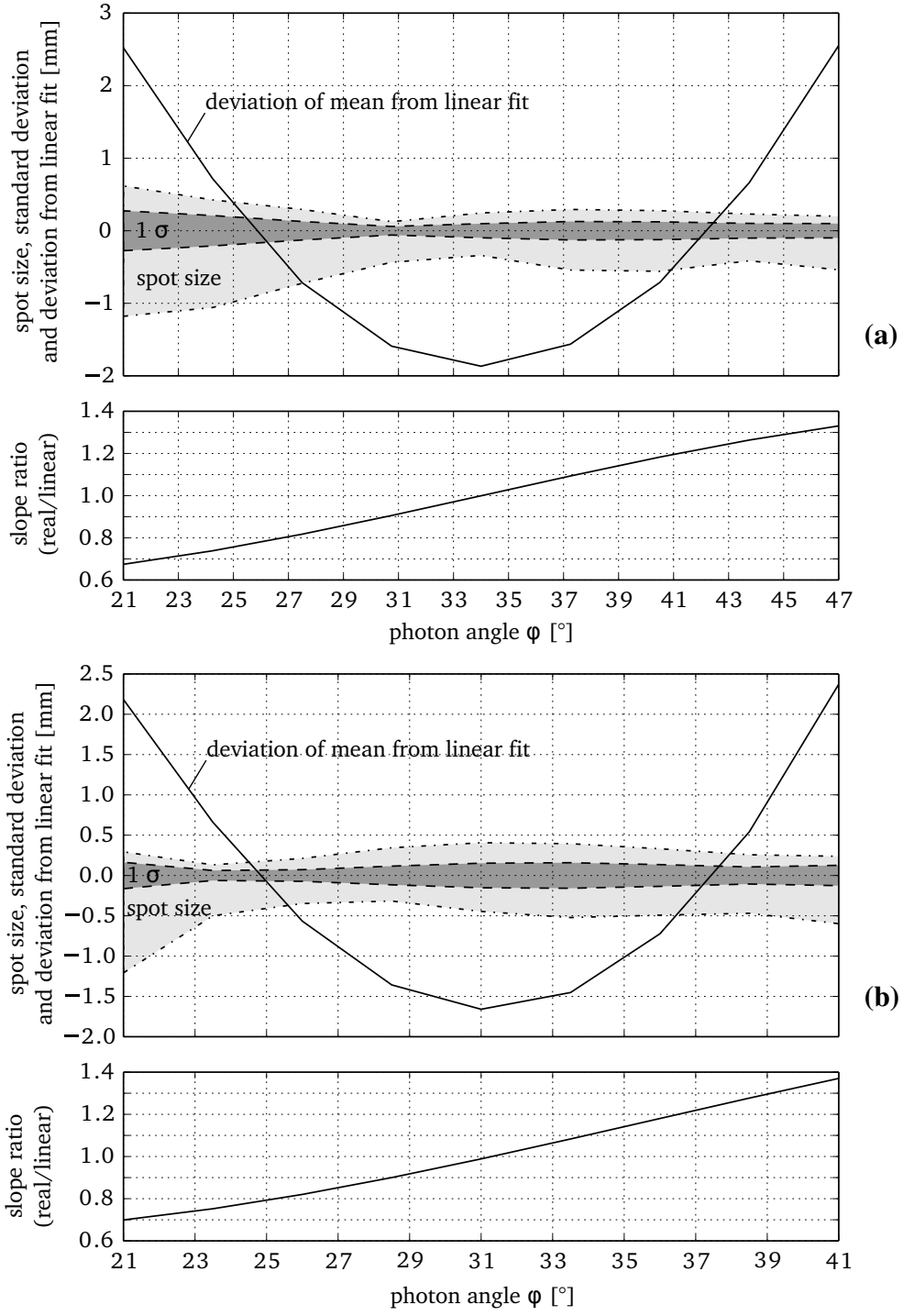


Figure 8.2.: Performance parameters of polynomial focusing elements optimized for the full φ' range from 21° to 47° (a) and a restricted range from 21° to 41° (b).

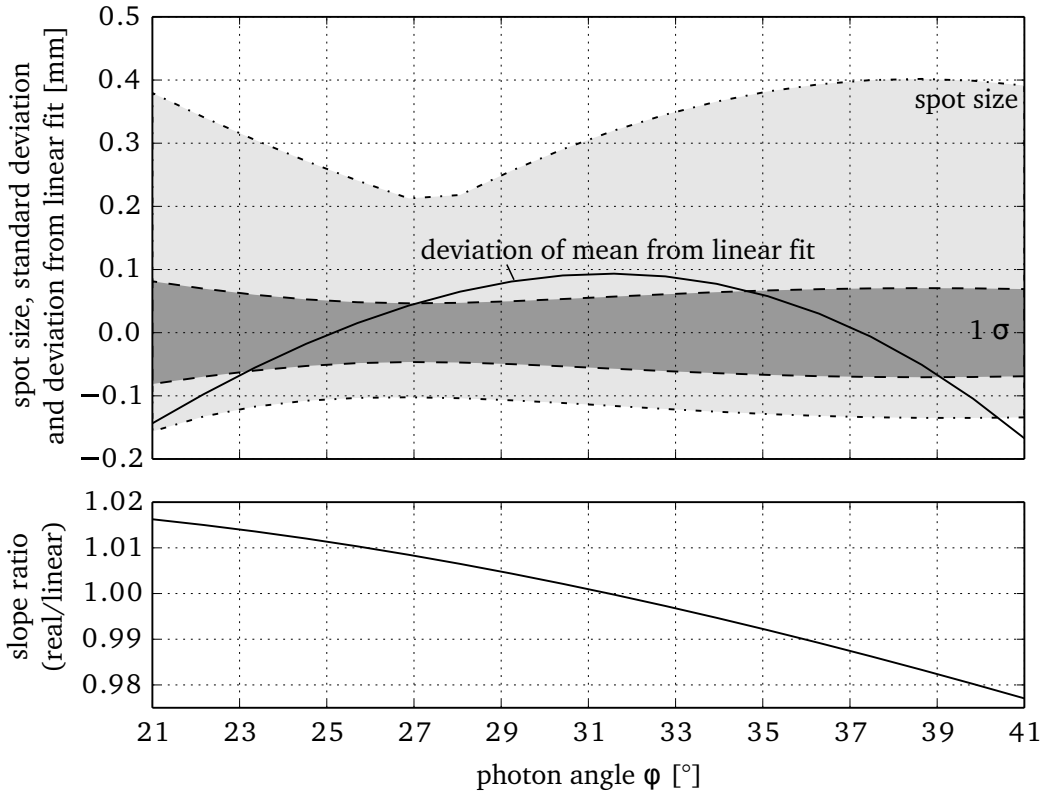


Figure 8.3.: Performance parameters of the cylindrical focusing element shown in Fig. 8.4. The maximum spot size (light gray band) is in the order of half a mm while the rms spot size (dark gray band) is about 150 μm. The deviation of the real imaging function from a linear fit is presented by the solid line in the upper plot. The ratio of the real slope and the slope of the linear fit is shown in the lower plot. As the slope difference is only in the order of 2 %, the non-linearity has a negligible effect on the resolution.

Design TIR1 is optimized to provide a good alignment of the PMT to the magnetic field. This alignment comes at the price of a larger size and inferior resolution. TIR2 is an example of a total internal reflecting FEL which is optimized to produce the smallest spot size. The alignment of the sensor is critical, but still in the acceptable range. The angular acceptance has been restricted to allow a fair comparison to the cylindrical design, which also profits from a restricted angular acceptance. TIR2 fulfills all requirements, but its performance is significantly worse than that of the cylindrical optics. The cylindrical shape yields better linearity, a smaller spot–size and a better alignment to the field. Further, the cylindrical shape leads to a narrower AOI distribution and can be produced using

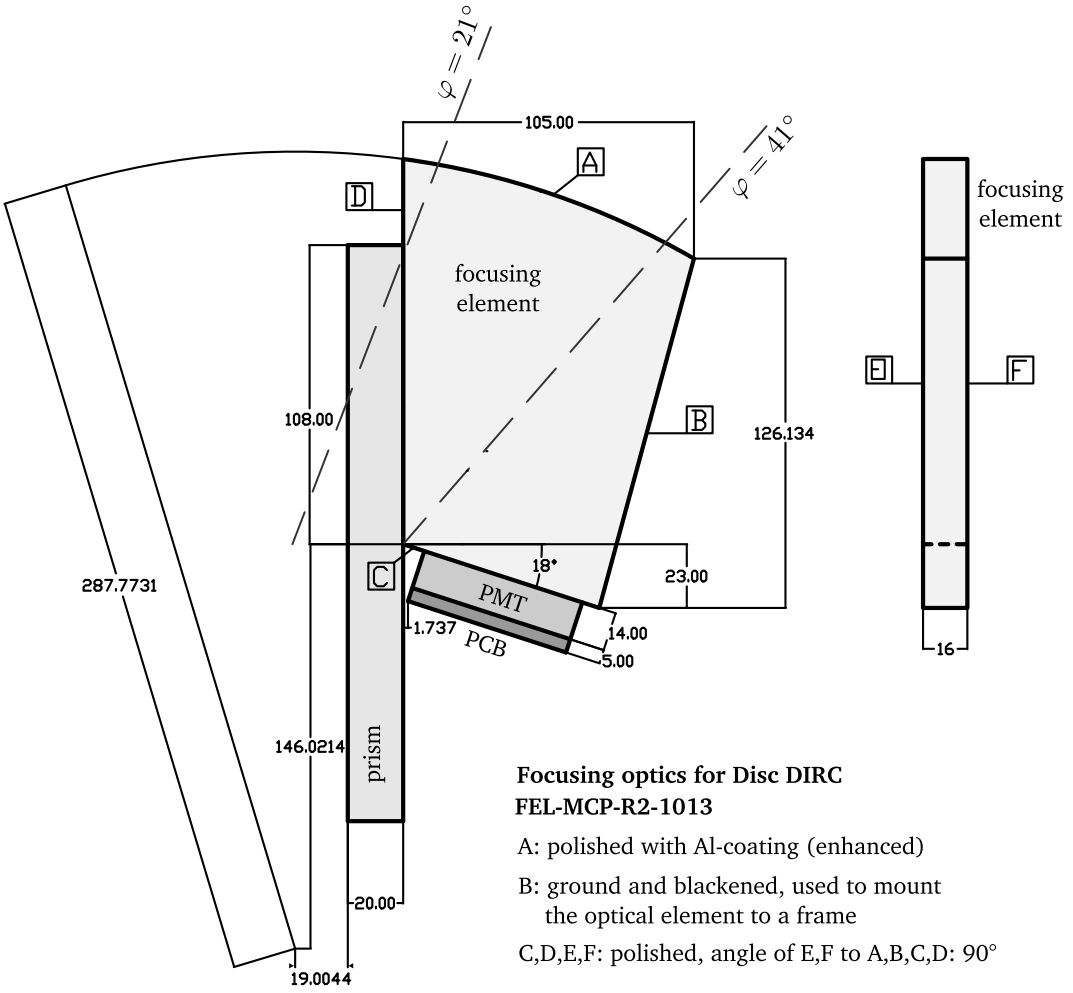


Figure 8.4.: Cylindrical imaging optics optimized for a 2″ MCP-PMT with a geometry similar to the Photonis Planacon. Due to the geometrical constraints and the limited size of the radiator, an additional prism element is needed to connect the focusing element and the radiator plate.

standard tools for polishing and surface metrology which are readily available at most manufacturers. In contrast, precision polishing of acylindric shapes requires custom tooling. For the cylindrical design, the expected orientation and strength of the magnetic field at the sensor positions is shown in Fig. 8.11. The gain dependence on these parameters is shown in Fig. 8.5. Tolerances for the optical element can be found in the appendix (Fig. A.4).

However, the acylindric FELs have also advantages. Due to the limited size of the radiator, the cylindrical optics cannot be attached directly to the rim (compare Fig. 5.44 on page 166). An additional connection has to be realized by means of a

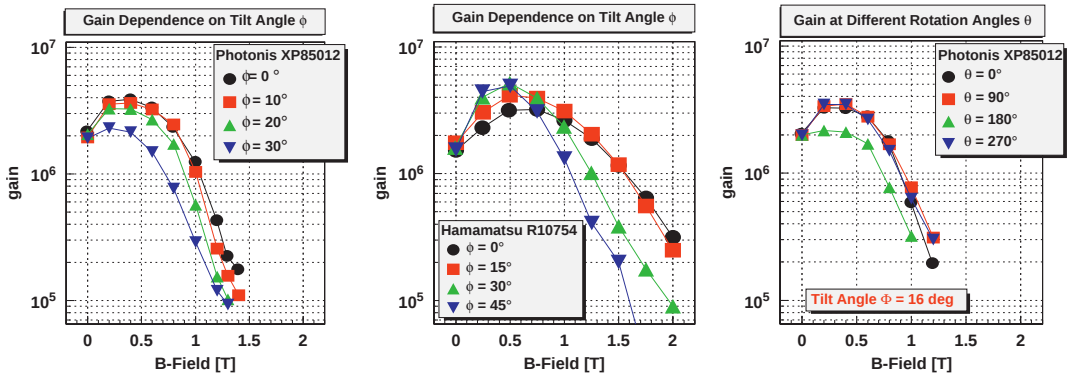


Figure 8.5.: Dependence of the MCP-PMT gain on the angle ϕ of the tube axis relative to the magnetic field. The tubes are a Planacon XP85012 with 25 μm pores and a MCP-PMT by HPK (R10754) with 10 μm pore size. The angle θ denoted the rotation of the tube around its own axis. In general, tubes with smaller pores perform better in a high magnetic field because the Larmor precession has less effect on the multiplication process. Figure from [136].

prism or bar. TIR2 is narrow enough to be attached directly to the radiator.

The prism connecting the cylindrical FEL to the radiator could be an achromatic prism or a fused-silica prism. To understand the advantages and disadvantages of such a prism, one has to look at the 3d imaging function $z(\varphi,\alpha_{\mathrm{FEL}})$. Geant 4 has been used to simulate the 3d imaging for a design with fused silica prism and a design optimized for a LiF prism. Fig. 8.6 shows the resulting imaging performance as a function of φ and α_{FEL} for a LiF-prism which is glued to the FEL with an epoxy, a fused-silica prism attached to the FEL by optical contact bonding and a fused silica prism glued with epoxy.

LiF based optics show a stronger α_{FEL} dependence of the focus mean position due to refraction. Because of Fresnel reflections at the prism-FEL interface, the rms of the point spread function increases with larger α_{FEL}. The reflected photons contribute to the stray light and deteriorate the resolution.

The influence of the epoxy can be seen by comparing the focus rms and width of the two fused-silica options. Even the epoxy leads to a slight degradation of the image quality. Hence, it is advisable to use optical contact bonding for the prism-FEL interface*.

The focus width in Fig. 8.6 corresponds to the spot dimension along the fo-

*Optical contact bonding of the prism and FEL in the final design (Fig. 8.4) has been tendered by manufacturers.

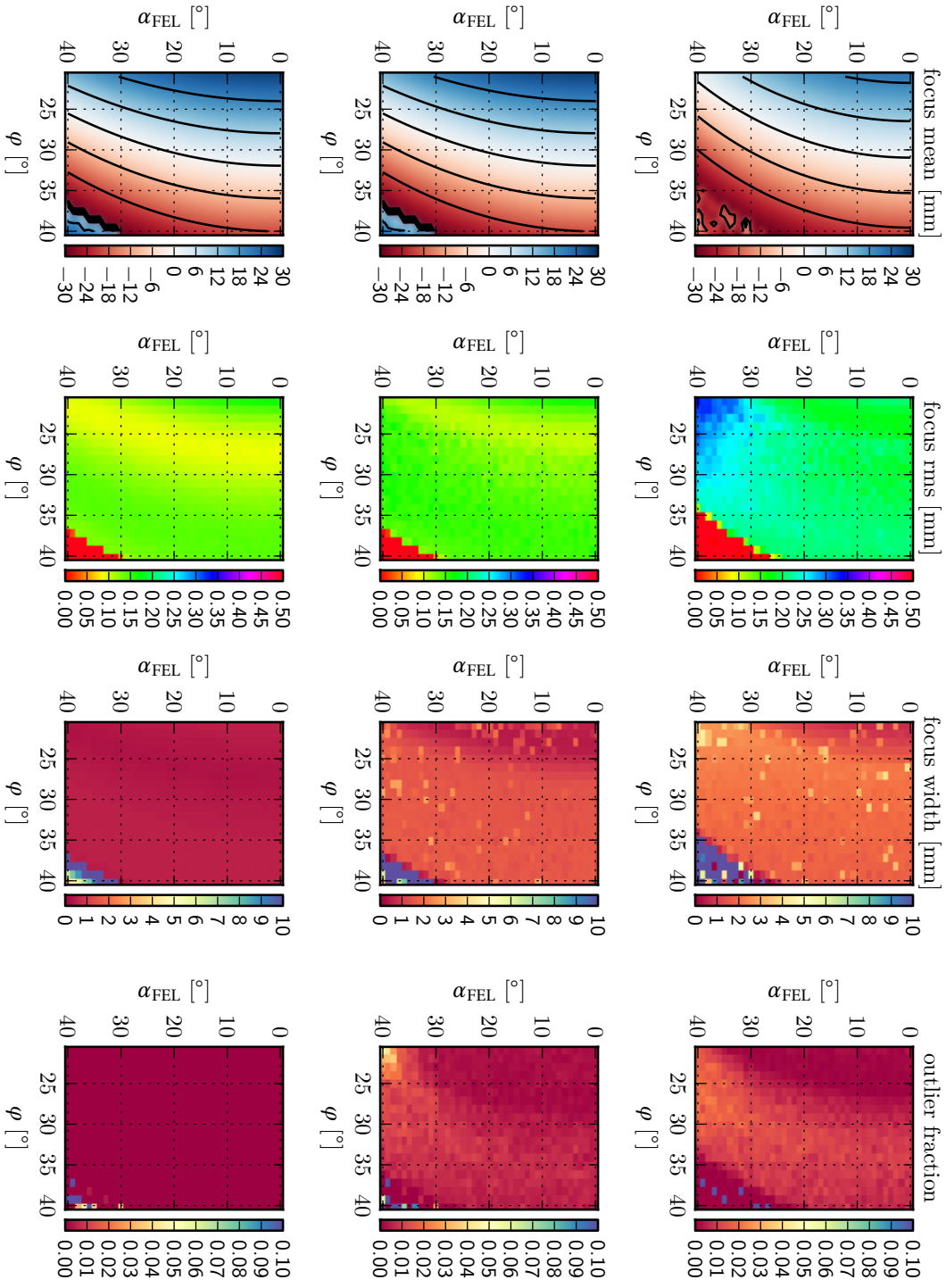


Figure 8.6.: 3d-raytracing results for the optics shown in Fig. 8.4 with (center column) and without (left column) EPOTEK 301-2 between prism and FEL. The right column displays results for a similar design with an epoxy–coupled LiF prism. Rows: $z(\phi,\alpha_{\mathrm{FEL}})$, spot rms, spot width corrected for outliers, percentage of removed outliers.

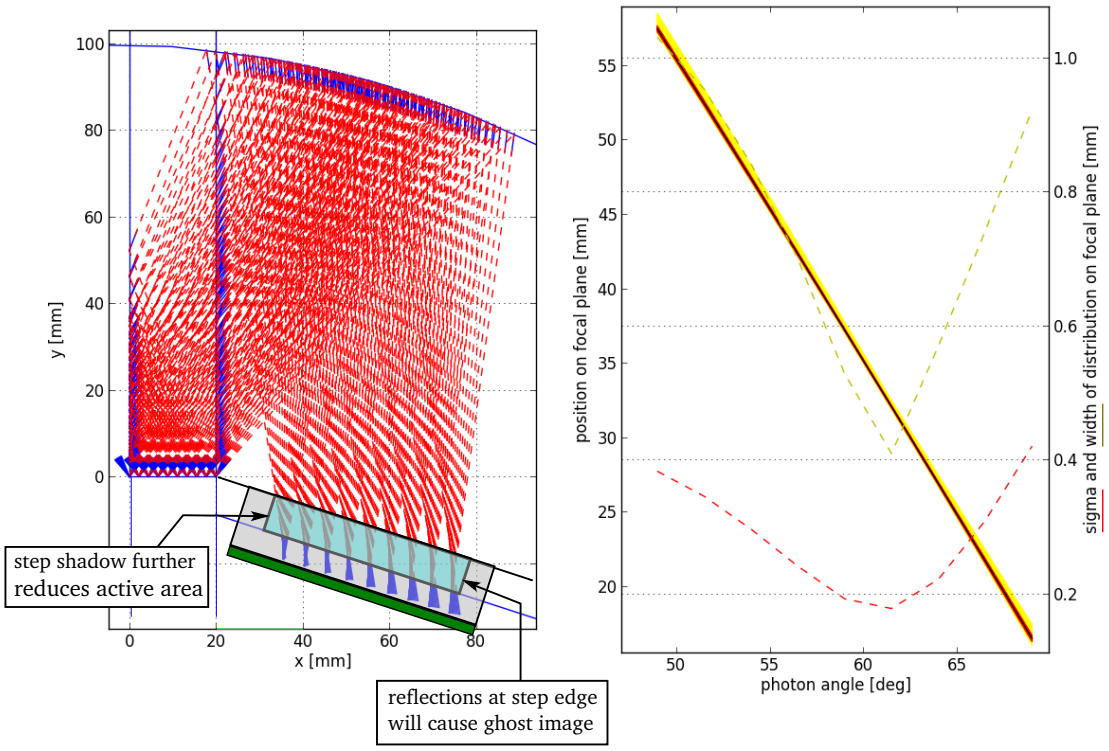


Figure 8.7.: A variant of the setup shown in Fig. 8.4, re-optimized for the use with a step window. The edges of the step window create shadowing and reflections. Such effects could only be avoided by using a fiber-optic plate as input window.

cusing direction, similar to the bundle width during optimization. In Fig. 8.6 however, the width has been cleaned from values which are far away from the spot. The corresponding percentage of removed hits is shown in the column "outlier fraction".

These results show that the LiF-prism leads to a degradation of the imaging quality. Further, LiF prisms are hardly available in the required size. Hence, the choice of the prism material is fused-silica. In this case, the prisms serve only as connection to the radiator. However, the FEL can be attached at different positions on the prism. This introduces an additional degree of freedom in the detector geometry which allows to shift individual FELs to better match the Endcap geometry.

Due to the necessary tilt of the sensor and the related angle-of-incidence, the design is not well suited for proximity focusing tubes with a step window *. The

*introduced in Fig. 5.35 on page 131.

problems are illustrated in Fig. 8.7 which shows a variant of the optical design which has been re-optimized for a step window of 8.4 mm thickness and an aperture of $47 \times 47\,\mathrm{mm}^2$. The edges of the step face cause shadowing at one side, resulting in a decrease of the effective active area to $41 \times 47\,\mathrm{mm}^2$, and reflections at the opposite side which lead to image distortions. An optimization of the angle-of-incidence is not possible due to the constraints in the alignment of the sensor with respect to the magnetic field. The problem could be solved by using a fiber-optic plate and accepting a reduction of 30 % to 40 % in detection efficiency. However, the baseline design does not include a proximity-focused tube.

The radiator outline has been altered during the optics design process to allow a smaller tilt angle of the sensor. The magnetic field plots in Fig. 8.11 do already include this modification. The final shape is shown in Fig 8.10.

To match the spatial resolution of the optics, the sensor should have a pixel size of $150\,\mu\mathrm{m} \times \sqrt{12} \approx 0.5\,\mathrm{mm}$. The final size of the pixels depends on the influence on the detector performance. The expected angle-of-incidence distribution at the sensor is shown in Fig. 8.8. This distribution has been obtained by means of a Geant 4 simulation of the final detector design. A comparison with the angular transmission-dependence in Fig. 5.21 on page 101 shows that this range is compatible with these filters.

The next step is to compute an optimum range for the wavelength filter. This has been realized numerically by computing the filter width needed to detect N photons for a given set of start-wavelengths (Fig. 8.9a). The resulting values have been interpolated by a spline. The minimum of the spline shows the optimum start wavelength. Fig. 8.9b shows the start wavelength vs the chromatic error and the width needed to detect $N = 20$ photons. It is clearly visible, that the narrower filter bands in the UV lead to stronger chromatic errors. Instead of the optimum of 355 nm to 437 nm, the wavelength range of 360 nm to 465 nm has been selected which is close to the optimum and to the existing filters in Fig 5.21.

The resulting detector design is presented in Fig. 8.10. All optical elements are made from synthetic fused silica. Every sub-detector module consists of a 2 cm thick radiator which is read out by 27 attached readout-modules (ROM). Each ROM consists of a single 2" MCP-PMT with a $50 \times 50\,\mathrm{mm}^2$ active area and bialkali photocathode, three cylindrical FELs and bars with a cross section as shown in Fig. 8.4, a front-end digitizer board with TOF-PET ASICs attached directly to the PMT and a front-end controller board with data concentrator and optical link.

The number of focusing elements is a trade-off between cost and impact on the

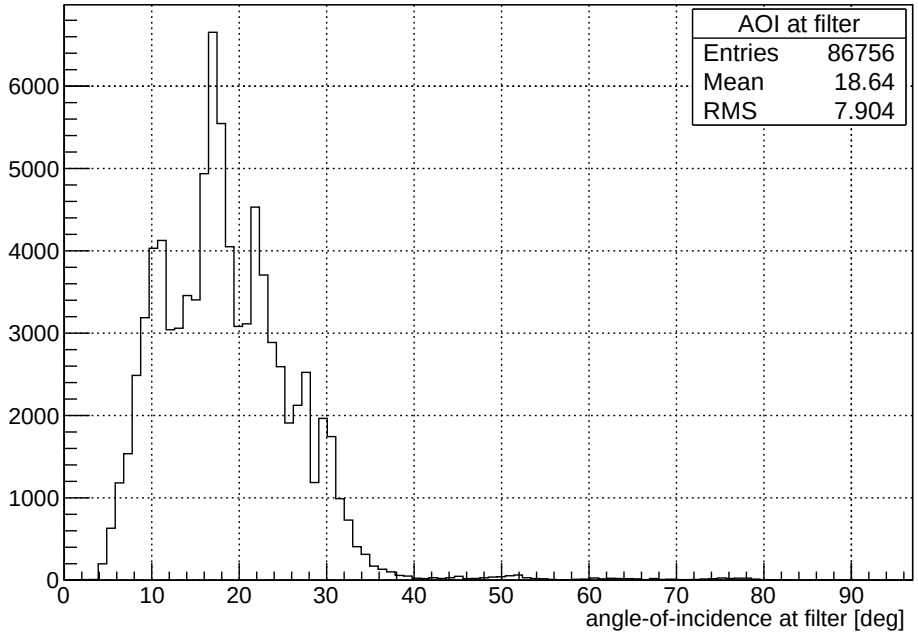


Figure 8.8.: AOI distribution at the sensor surface obtained by a Geant 4 simulation of the final detector design. This range is compatible with filters as shown in Fig. 5.21 (page 101).

final detector performance, as shown in Fig. 8.12. The use of very wide focusing elements for α_{FEL}-reconstruction conflicts with the combination of geometric constraints and constraints on the sensor alignment. In the final design, the distance of the FELs to the radiator has to be different for several ROMs to resolve geometry conflicts. It is foreseen to coat the interference filter directly on the front-glass of the PMT prior to the assembly of the tube*. The anode structure of the tube consists of 3×100 pixel strips at a 0.5 mm pitch which are read out individually. First sensor-prototypes have been delivered and are being characterized at the time of writing. More details on the front-end electronics and processing architecture will be discussed after the performance estimation.

*The filter coatings are usually compatible with the vacuum process and stable enough to withstand the high temperatures ($\sim 400^{\circ}$C) during processing

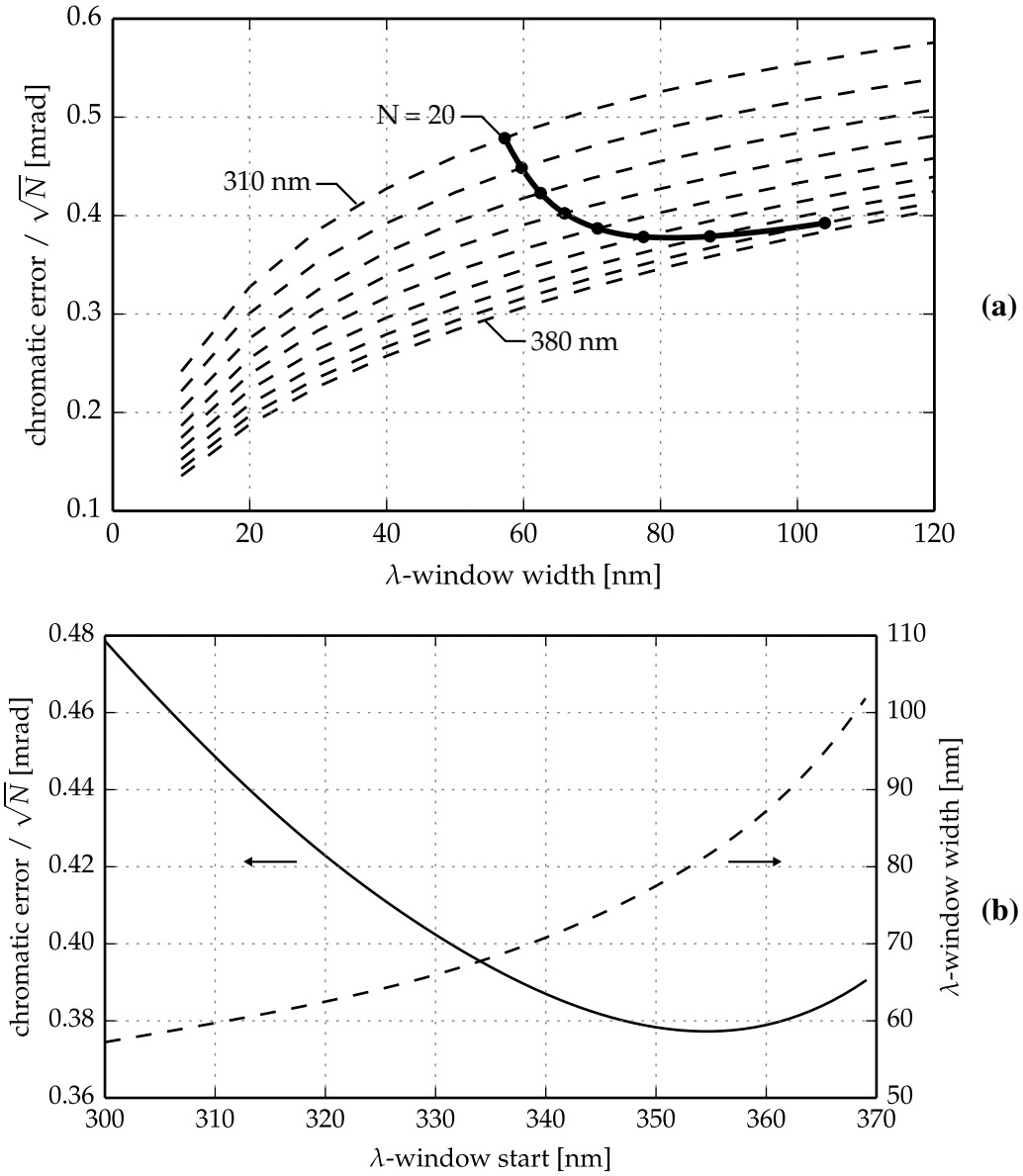


Figure 8.9.: (a): The dashed lines show the chromatic error per particle track for several wavelength bands starting at 310 nm to 380 nm (10 nm steps) versus the variable bandwidth. The underlying model includes the quantum efficiency of the Planacon MCP-PMT as well as the bulk and reflection losses in the radiator and at the mirror. The dots on each line mark the bandwidth at which the filter results in 20 detected photons. These dots are interpolated by a quadratic spline (solid line) whose minimum indicates that a optimum filter configuration exists (b). The reason for the existence of a minimum is the decline of quantum efficiency at higher wavelengths. The optimum is a 82 nm wide filter starting at 355 nm.

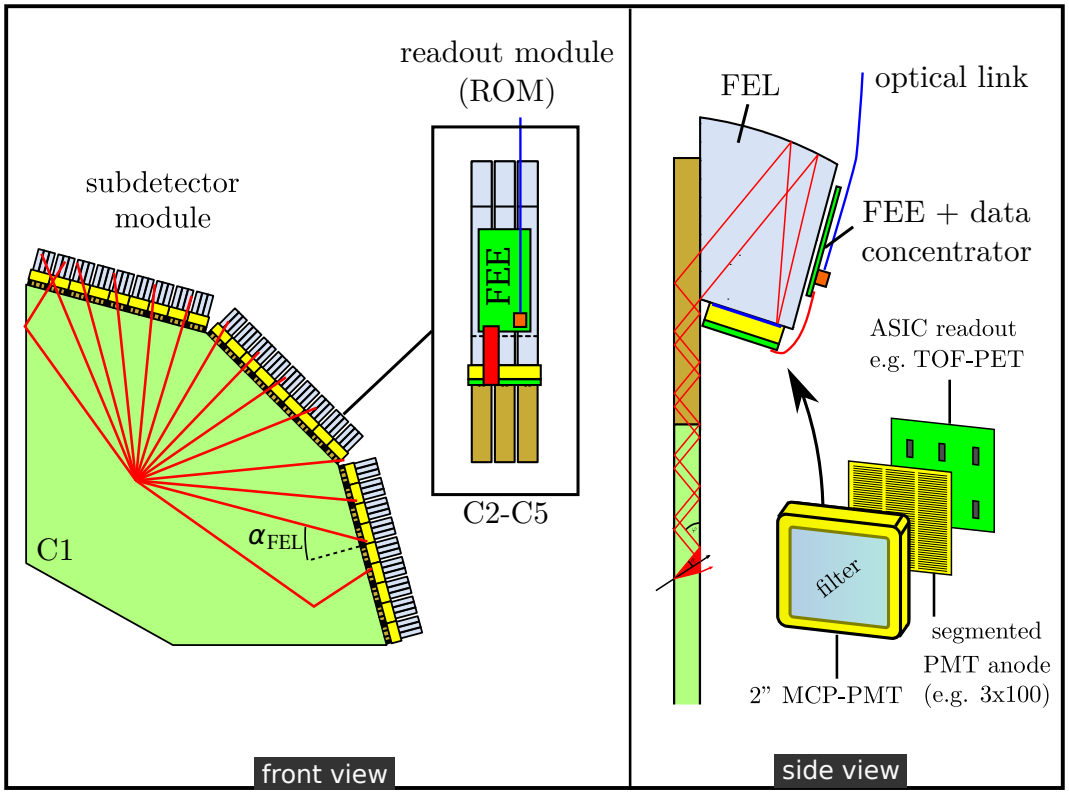


Figure 8.10.: The final MCP-based Endcap DIRC design consists of four subdetector-modules (SDM) which form a dodecagonal outline to improve the alignment of the sensor with the magnetic field. A SDM consists of one monolithic fused-silica radiator (C1) which is coupled to and read out by a set of 27 readout modules (ROM) attached to the rim. The ROMs combine all remaining components which have to be placed inside the spectrometer (C2–C5). It has been proposed to develop a common support structure for the ROM, so that every individual component can be mounted to this skeleton. With such a support structure, ROMs can be assembled and aligned individually prior to the assembly of the whole system. This permits a simpler quality-assurance during construction because the ROMs can be completely tested on a test-bench prior to the assembly of the full system.

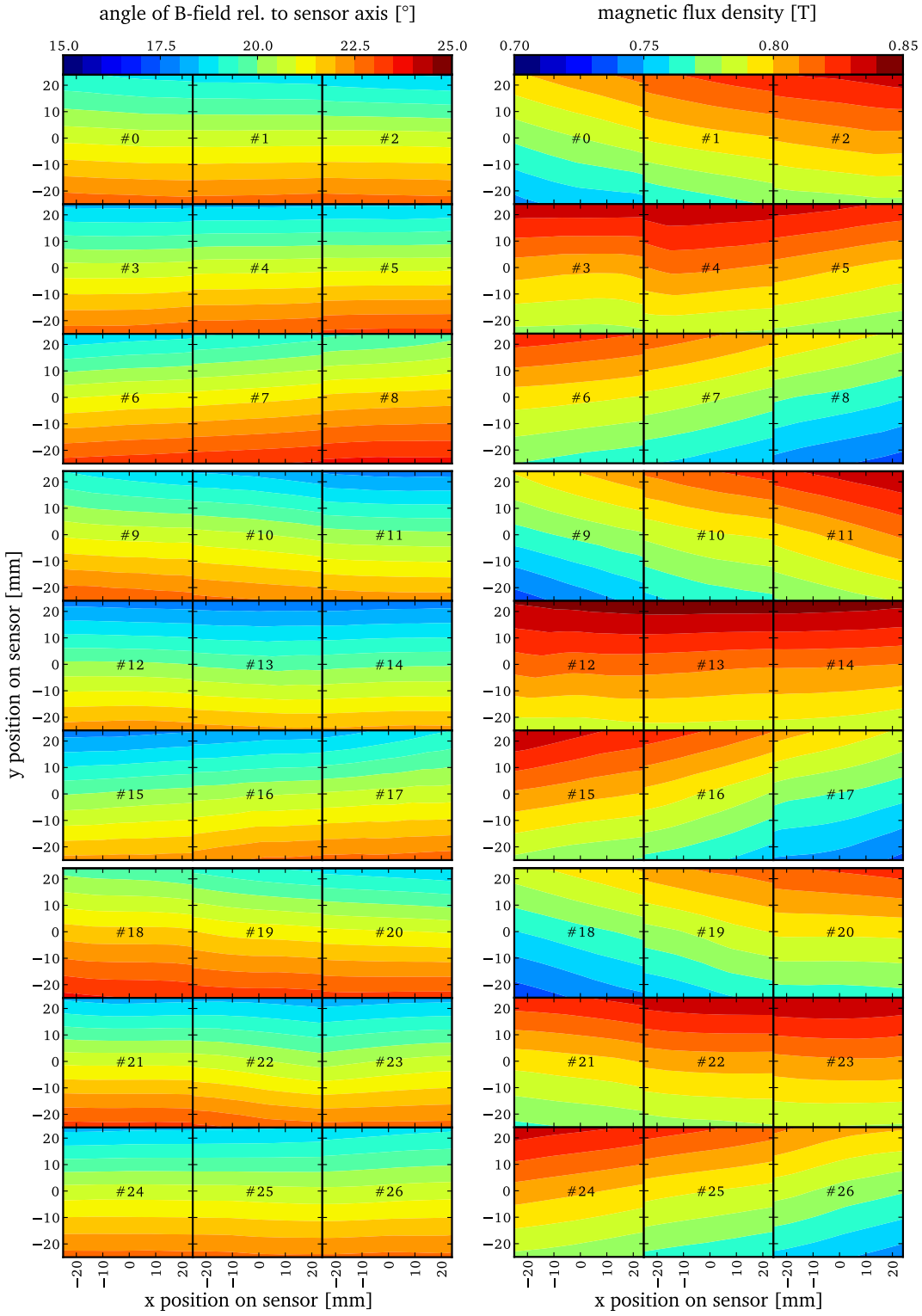


Figure 8.11.: Expected orientation (left column) and strength of the magnetic field (right column) on the surface of each of the 27 PMTs of one subdetector-module.

8.2. Performance studies

Similarly to the dSiPM design studies, two simulations have been developed for performance estimation. One is implemented by interfacing Geant 4 directly and the other is based on the PandaRoot framework. In both cases, the analysis of the resulting hit-patterns is conducted by the DDRecon library. The reconstruction algorithm developed for the MCP-based design, comprising the reconstruction of single photon Cherenkov angles, has been explained in chapter 6. To translate the Cherenkov angle resolution to a misidentification, a PID-classifier* has been computed from the reconstructed θ_c-distributions and applied to the reconstructed Cherenkov angles. Common simulation parameters are summarized in table 8.1.

Fig. 8.12 a shows the estimated detector performance for a super-bialkali photocathode with a quantum-efficiency of 30 % and different numbers of focusing elements per ROM at a particle momentum of 4 GeV/c, obtained from a direct Geant 4 simulation without background from the experiment. As discussed in section 5.3, a higher number of FELs leads to a reduced uncertainty of α_{FEL} and hence to a stronger reduction in separation power for tracks which are closer to the FELs. The results of this study motivated the choice of three FELs per ROM. A higher number of FELs would lead to a significant cost increase without a significant improvement of the overall performance.

An example of an accumulated hit-pattern for 10^4 tracks is given in Fig. 8.12 b. To visualize the results from the reconstruction algorithm, the reconstructed single photon Cherenkov angles for all these tracks are accumulated in the histogram shown in Fig. 8.12 c. The corresponding mean Cherenkov angles for each track are shown in Fig. 8.12 d. The separation power in Fig. 8.12 a has been computed from

Time resolution (RMS)	21 ps
TDC binning (LSB)	50 ps
Pixel line height	0.5 mm
Per pixel dead time	20 ns
Surface roughness	1.0 nm
Track position error $\sigma_{\mathrm{x,y}}$	1.5 mm
Track angular error σ_{θ_p,ϕ_p}	1 mrad

Table 8.1.: Simulation parameters of the MCP-PMT/TOFPET-ASIC design.

*see section 3.2 on page 30

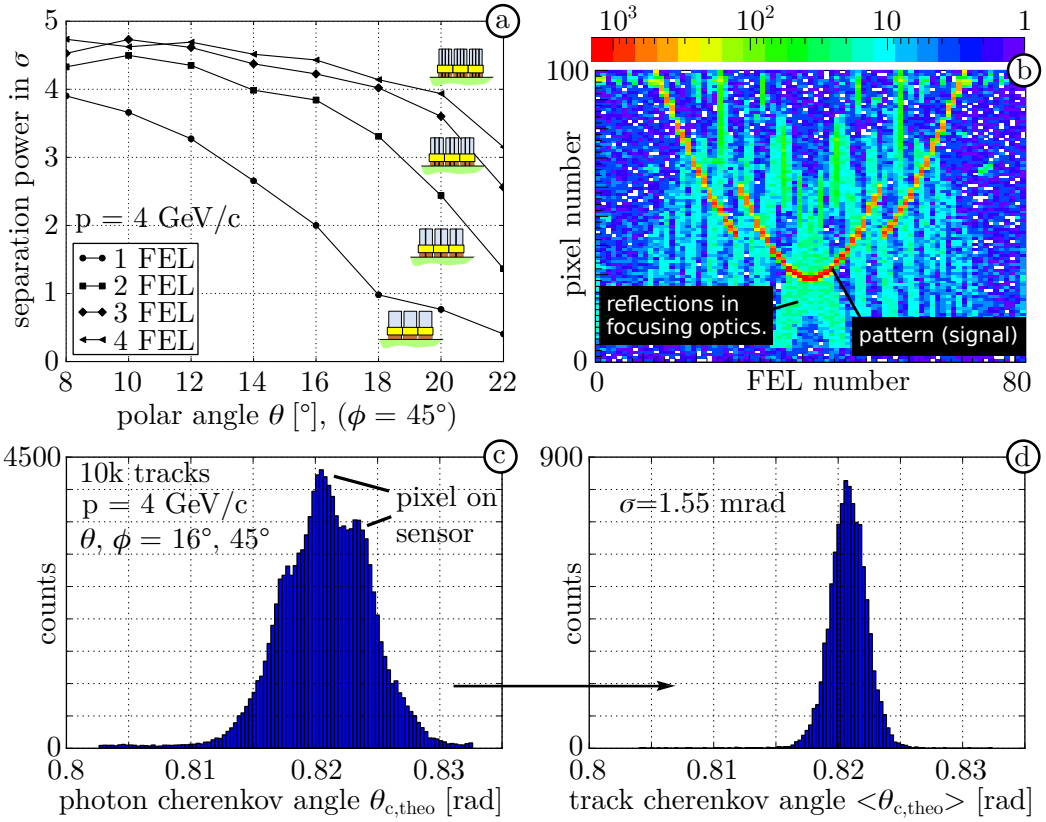


Figure 8.12.: (a): π/K separation at 4 GeV/c vs. polar angle for different numbers of FELs per ROM. Each dot corresponds to the reconstruction of 10^4 π/K-tracks. The separation power has been computed from the resulting track Cherenkov angle distributions using eq. 3.2. (b): a typical hitpattern ($\theta_p = 16°, \phi_p = 45°$). (c): Distribution of single photon Cherenkov angles, reconstructed from the pattern shown in (b). (d): The track Cherenkov angle distribution obtained from (c). Published in [82].

the track Cherenkov angle distributions for 10^4 pions and 10^4 kaons as explained in section 3.2. The multi-peak structure in Fig. 8.12 c is a consequence of the pixelized readout and vanishes for a continuous readout. The discrete information from the pixels is smeared out by the error of the track-angles (1 mrad), the unknown α_{FEL} and multiple coulomb scattering. Note that the histogram contains reconstructed Cherenkov angles for hits on all ROMs, where each ROM has a different view on the Cherenkov cone.

As the Planacon is not available with a super-bialkali photocathode, the measured quantum efficiency of a Planacon MCP-PMT (Fig. 8.13b) has been used for the final performance estimation. Results from a direct Geant 4 simulation

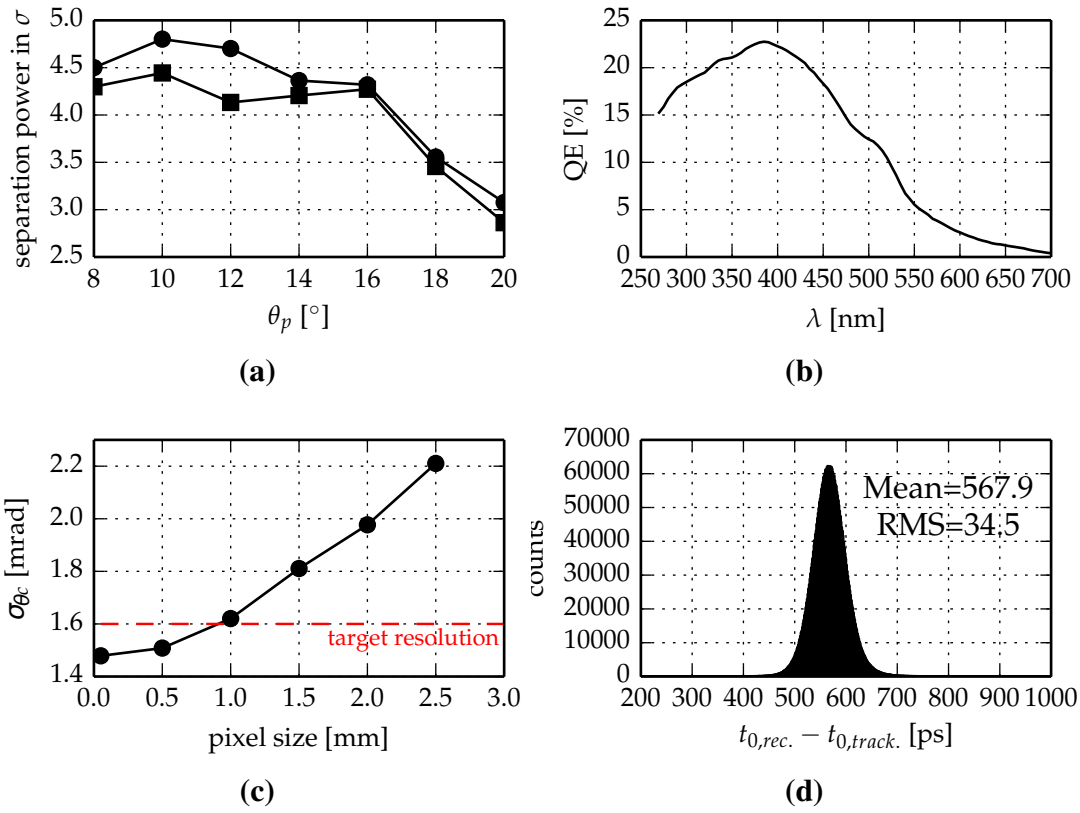


Figure 8.13.: (a): Separation power for a setup using a Planacon with the quantum efficiency as presented in (b). The curves show the performance of a design with bandpass filter (squares) and with an edge filter at 350 nm (circles). These results have been obtained by reconstructing 1000 π/K–tracks at $p = 4$ GeV/c and $\phi_p = 45°$. (b): A Planacon-QE curve which has been measured by $\overline{\text{P}}$ANDA collaborators in Erlangen. (c): Dependence of the Cherenkov angle resolution on the pixel size for the design with bandpass filter. The points correspond to the reconstruction of 1000 π at (p, θ_p, ϕ_p)=(4 GeV/c, 15°, 45°). (d): Error of reconstructed t_0 obtained from the study presented in Fig. 8.15. The momentum range is 1 to 4.5 GeV/c in 0.5 GeV/c steps.

of the presented design with bandpass filter as well as a design with $\lambda > 350$ nm are shown in Fig. 8.13a. Both designs are compatible with the performance requirements. Further, Fig. 8.13c shows the dependence of the Cherenkov angle resolution per particle track on the choice of the pixel-size. The design target of $\sigma_{\theta_c} < 1.6$ mrad cannot be met at a significantly larger pixel-size.

A time-based PandaRoot simulation of the Endcap DIRC has been implemented to include the influence of the experimental background originating from coinci-

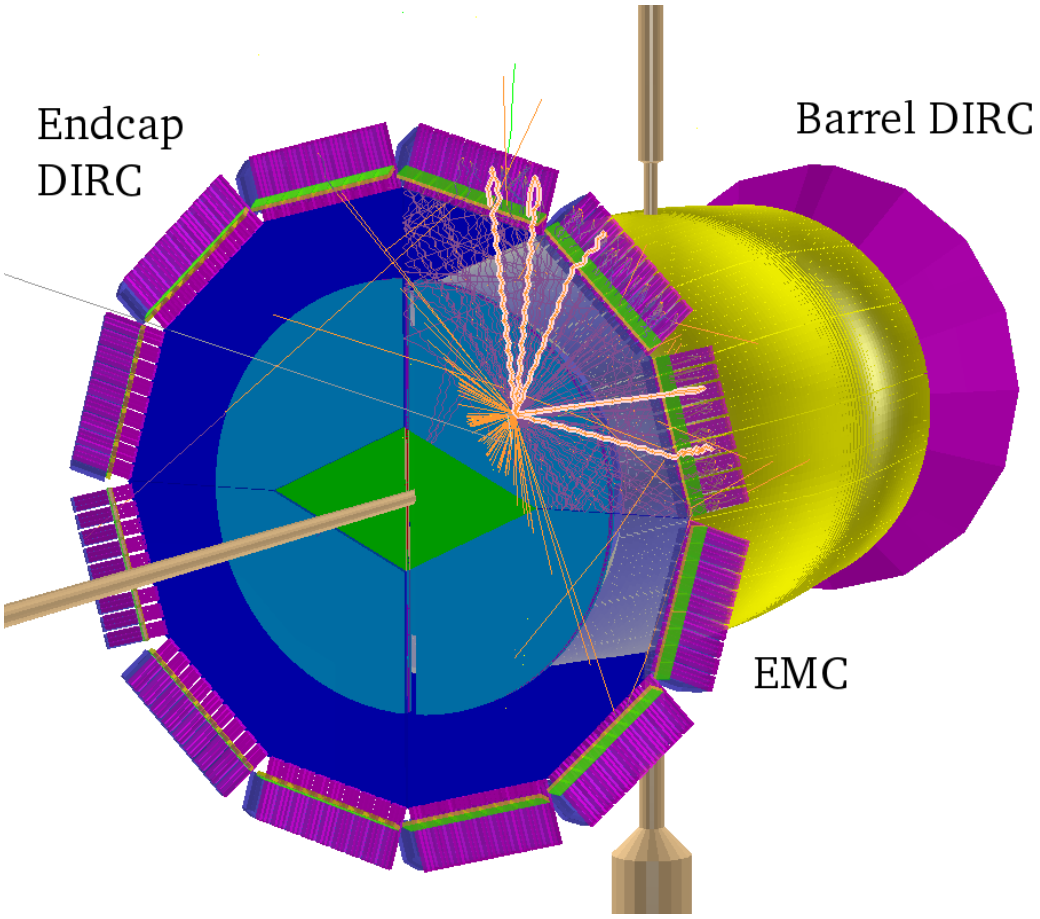


Figure 8.14.: PandaRoot event display showing photons originating from a charged particle track hitting the Endcap DIRC. For clarity, the geometry of the solenoid and flux return yokes has been hidden and only a few photon tracks are highlighted.

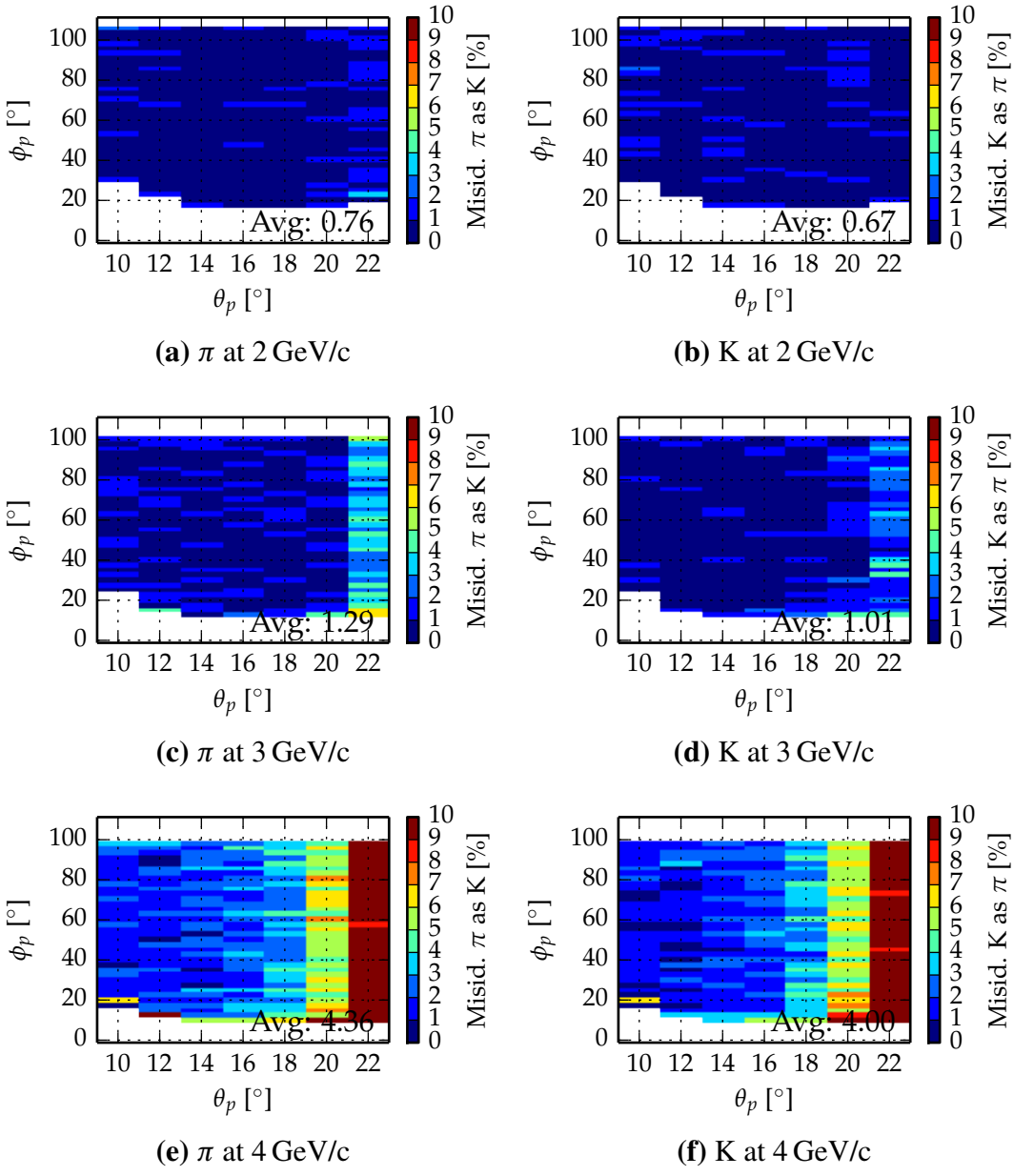
dent tracks and showers from the surrounding material. Fig. 8.14 shows an event display of the simulation. The mapping of the detector performance has been realized in the same way as described for the dSiPM based design. Tracks with different directions at the vertex are used to scan over one SDM. The simulation includes DPM-background at a 20 MHz interaction rate. In contrast to the dSiPM study, the background includes elastic and inelastic events and is thus not overestimated. The corresponding misidentification maps are shown in Fig. 8.15. Apart from minor statistical variations due to the low track count of 10^3 reconstructed tracks per bin, the results show a rather uniform misidentification in ϕ_p and the expected dependence in θ_p, which corresponds to the α_{FEL} uncertainty. One can consider these results to be compatible with the performance requirements because

the regions with higher misidentification are not critical. The kaon rate is low for this part of the phase-space (compare Fig. 5.2a). The deviation of the reconstructed start times $t_{0,\mathrm{rec}}$ from the time $t_{0,\mathrm{track}}$ where the track arrives at the radiator is presented in Fig. 8.13d. The data includes angular scans as in Fig. 8.15 for the momenta 1 GeV/c to 4.5 GeV/c in 0.5 GeV/c steps.

The reconstruction efficiency, defined as the number of reconstructed tracks divided by the number of incident tracks at the radiator, is typically in the order of 99 % over the full acceptance of the SDM, as shown in Fig. 8.16. Variations in the number of photons per reconstruction frame with a duration of 45 ns are about 10 % as shown in Fig. 8.17. Finally, the RMS values of the single photon Cherenkov angle distributions are shown in Fig. 8.18. The single photon RMS is reduced at higher momenta because the distortions of the particle track decreases with increasing momentum, as shown in Fig. 5.13 (p. 86).

In addition to the performance studies which include the experimental background, a specific test case for multiple coincident tracks has been constructed to proof the efficient background rejection of the Endcap DIRC system and the reconstruction algorithm. The 3d-hitpatterns for the test-case are presented in Fig. 8.19, where the probe-track pattern is drawn in black. Both scenarios include five coincident particle tracks in total. In Fig. 8.19a, the green pattern is so close to the probe track that both patterns start to overlap in 3d. In Fig. 8.19b however, the green pattern has been displaced, resulting in two separated patterns. Note that the scenario of overlapping patterns is extremely rare because tracks have to match closely in momentum, start-time, θ_p, ϕ_p and sign of the charge to show such an overlap. In all other cases, patterns are separated and show only crossings in 3d-space as it is the case for the pattern in the "separated" scenario. The reconstruction results for the probe tracks in both scenarios are summarized in table 8.2. Generally, patterns can be reconstructed with reasonable performance in presence of four tracks in the same time window as long as the patterns are separated. Only the overlapping pattern has a significant impact on the performance. A design with higher photon statistics is more robust in this case. However, this artificially created case of overlapping patterns can be considered to be insignificant in practice.

Finally, the DPM simulations have been used to estimate the radiation load and charged particle rates at the Endcap DIRC. The corresponding maps are shown in Fig. 8.20, 8.21 and 8.22.

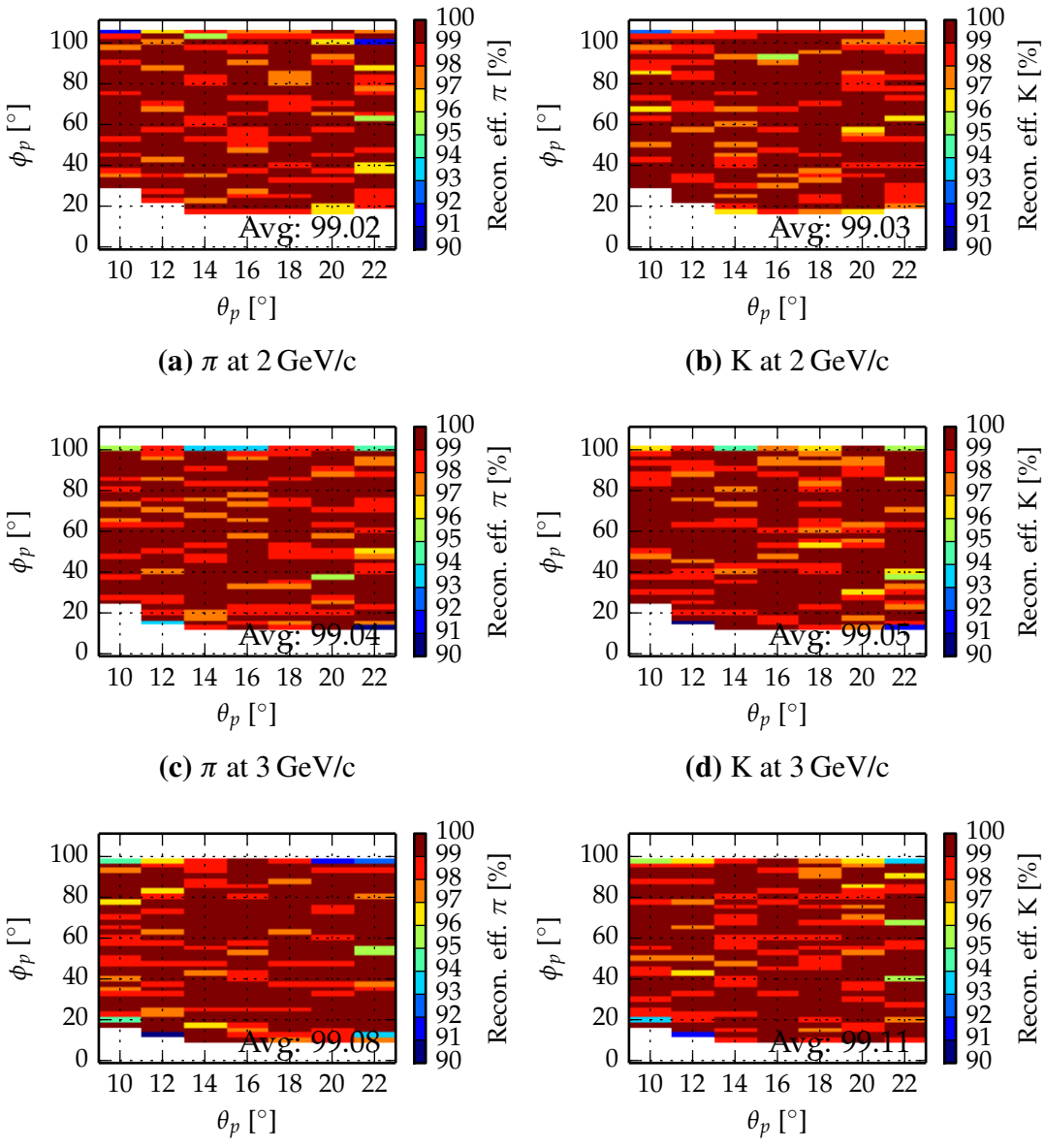


(a) π at 2 GeV/c **(b)** K at 2 GeV/c

(c) π at 3 GeV/c **(d)** K at 3 GeV/c

(e) π at 4 GeV/c **(f)** K at 4 GeV/c

Figure 8.15.: Misidentification obtained by reconstructing 10^3 probe-tracks per colormap-bin in presence of DPM-background at the nominal interaction rate of 20 MHz. Track angles at the interaction point are given by the bin coordinates θ_p, ϕ_p. Probe-tracks are scanned over one SDM. In the presented color-maps, bins with a track reconstruction efficiency below 90 % have been omitted. Hence, the acceptance of the SDM is clearly visible in the plots. The hole for the beam pipe is visible as empty bins in the lower left corner. Tracks at 22° pass through the GEM electronics. At 4 GeV/c, the misidentification for these tracks exceeds 10 %. The colormap has been restricted to 10 % to allow a comparison of the results for $\theta_p < 22°$.

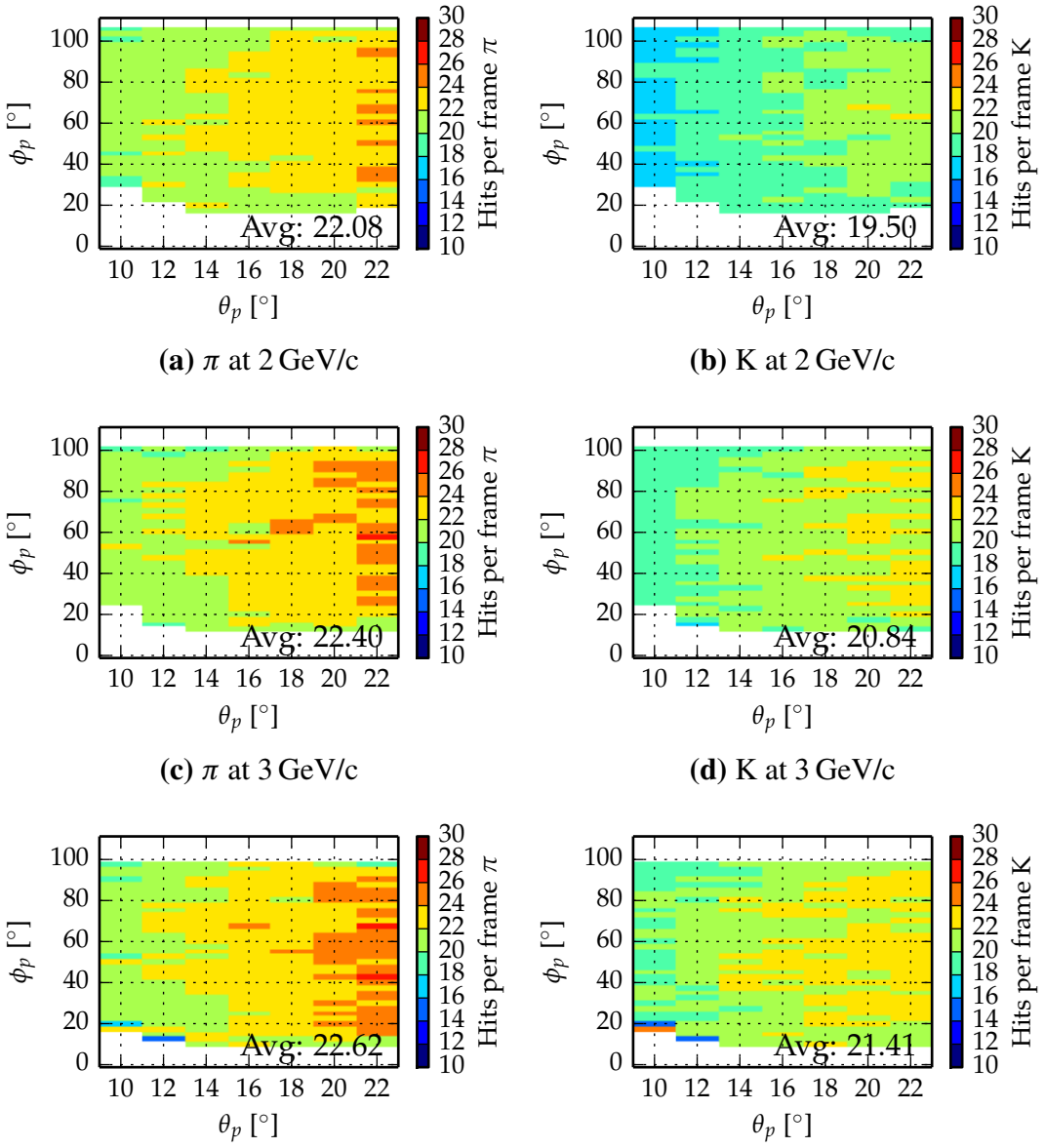


(a) π at 2 GeV/c **(b)** K at 2 GeV/c

(c) π at 3 GeV/c **(d)** K at 3 GeV/c

(e) π at 4 GeV/c **(f)** K at 4 GeV/c

Figure 8.16.: Track reconstruction efficiency for the results shown in Fig. 8.15. This value is defined as the percentage of reconstructed probe-tracks from all probe-tracks which reach the radiator. Note that the decay of probe tracks has been suppressed during simulation. Tracks can be lost due to inelastic collisions or simply because the reconstruction algorithm rejects all hits. However, the average track reconstruction efficiency is ~ 99 % for all momenta. This value is considered to be sufficient.

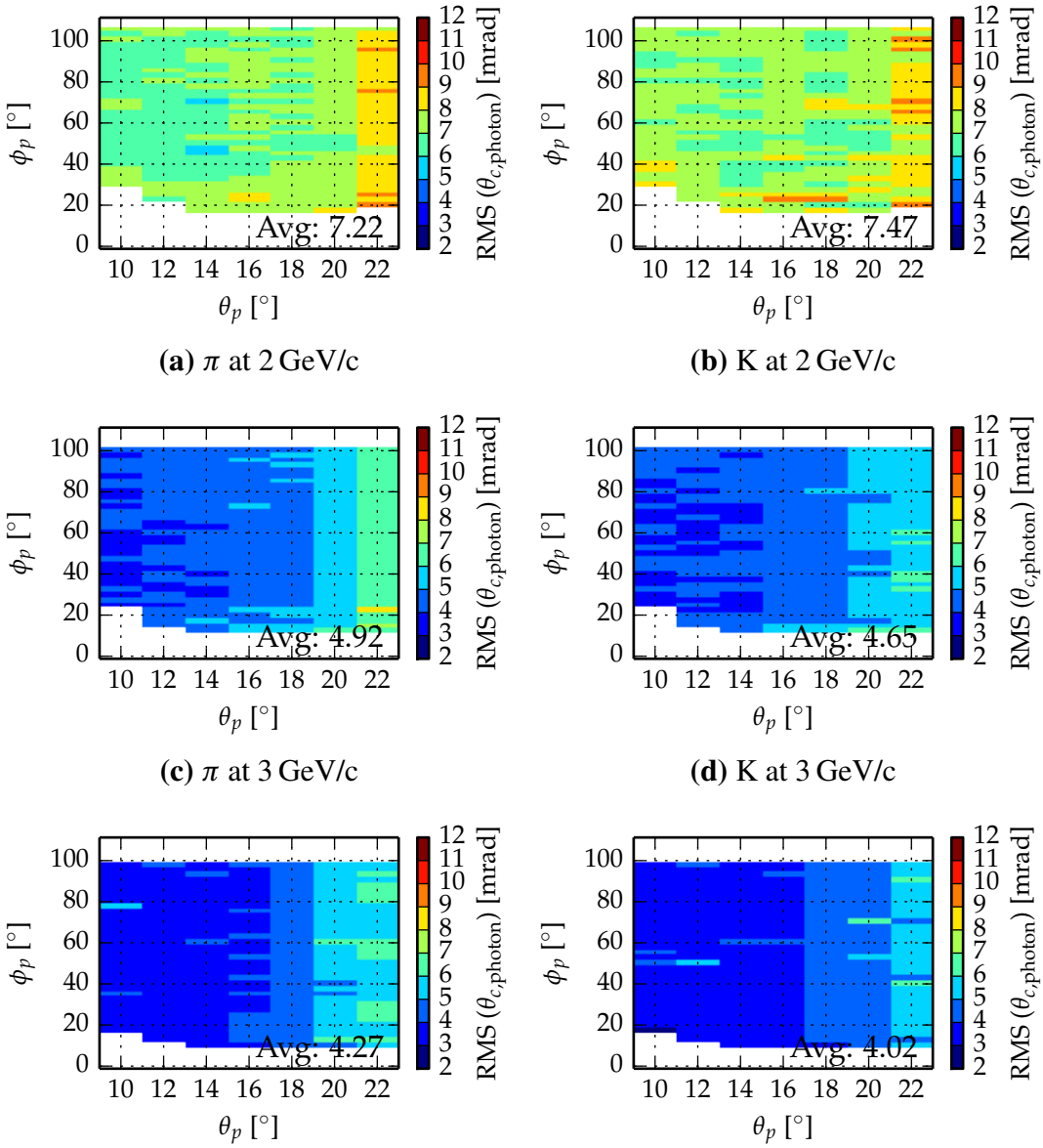


(a) π at 2 GeV/c **(b)** K at 2 GeV/c

(c) π at 3 GeV/c **(d)** K at 3 GeV/c

(e) π at 4 GeV/c **(f)** K at 4 GeV/c

Figure 8.17.: Average number of detected hits (probe tracks + background) per reconstruction frame with a duration of 45 ns. The given values correspond to the results shown in Fig. 8.15. The decreasing number of hits towards lower polar angles is a consequence of the limited angular acceptance of the focusing elements and the higher number of reflections at the radiator surface.

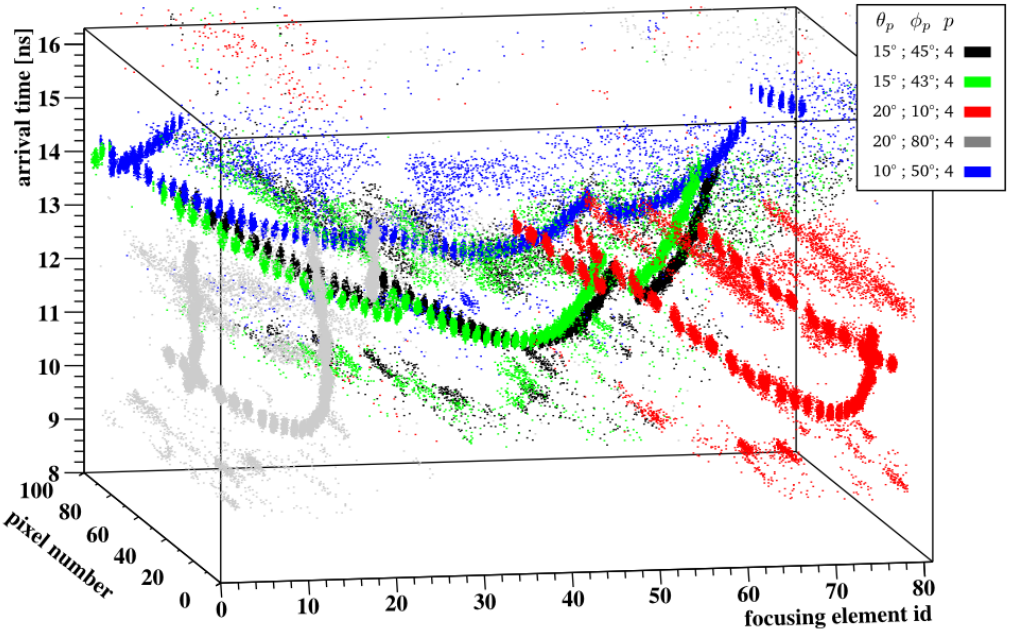


(a) π at 2 GeV/c **(b)** K at 2 GeV/c

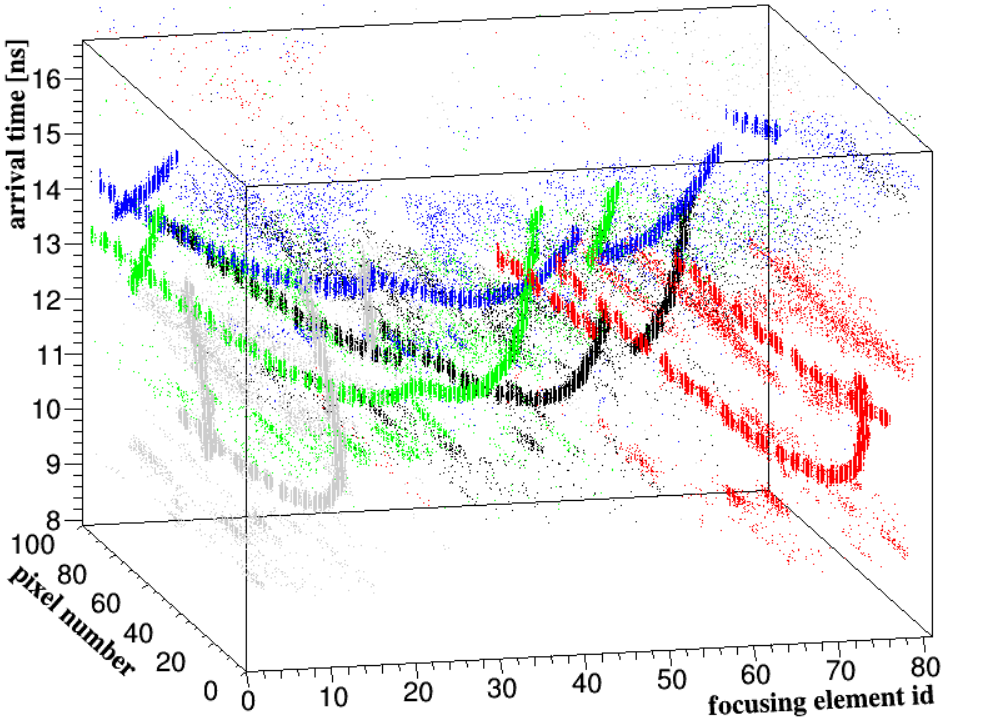
(c) π at 3 GeV/c **(d)** K at 3 GeV/c

(e) π at 4 GeV/c **(f)** K at 4 GeV/c

Figure 8.18.: RMS of the reconstructed single photon Cherenkov angle distributions corresponding to the results shown in Fig. 8.15. Note that these distributions are obtained after the application of time cuts and outlier removal.



(a)



(b)

Figure 8.19.: 3d-hitpatterns of a probe track (black) in presence of 4 coincident tracks. The difference between scenarios (a) and (b) is the green track which is very close to the probe track in (a) and farther away in (b).

λ-range	*360 nm – 465 nm*				*> 350 nm*			
scenario	separated		overlap		separated		overlap	
particle specie	π	K	π	K	π	K	π	K
Avg. hits in pattern	97.5	93.2	96.1	91.0	146.8	138.3	143.7	135.5
Avg. hits rejected	82.7	78.5	76.2	71.0	124.8	116.3	118.4	110.1
Avg. hits selected	14.8	14.7	19.9	20.0	22.0	22.0	25.3	25.4
Misidentification [%]	2.6	2.9	13.7	13.0	2.2	2.1	4.5	4.0
Missing tracks [%]	0.1	0.1	0	0	0	0	0	0
t_0 RMS [ps]	59.7	60.9	41.3	38.01	71.2	60.9	46.3	40.7

Table 8.2.: Reconstruction results for multi-track events with 5 coincident tracks. Patterns for the two compared scenarios are shown in Fig. 8.19a (entitled *overlap*) and Fig. 8.19b (entitled *separated*). It is clearly visible that the separated scenario can be handled easily by the reconstruction algorithm, what is a consequence of the good separation of the patterns in the 3d-space. The design with higher counting statistics is more robust in handling the rare case of overlapping tracks. The two overlapping tracks in the given example (hit-patterns drawn black and green in Fig. 8.19a) have equal momenta (4 GeV/c) and polar angles (15°) but differ in their azimuth angle by 2°.

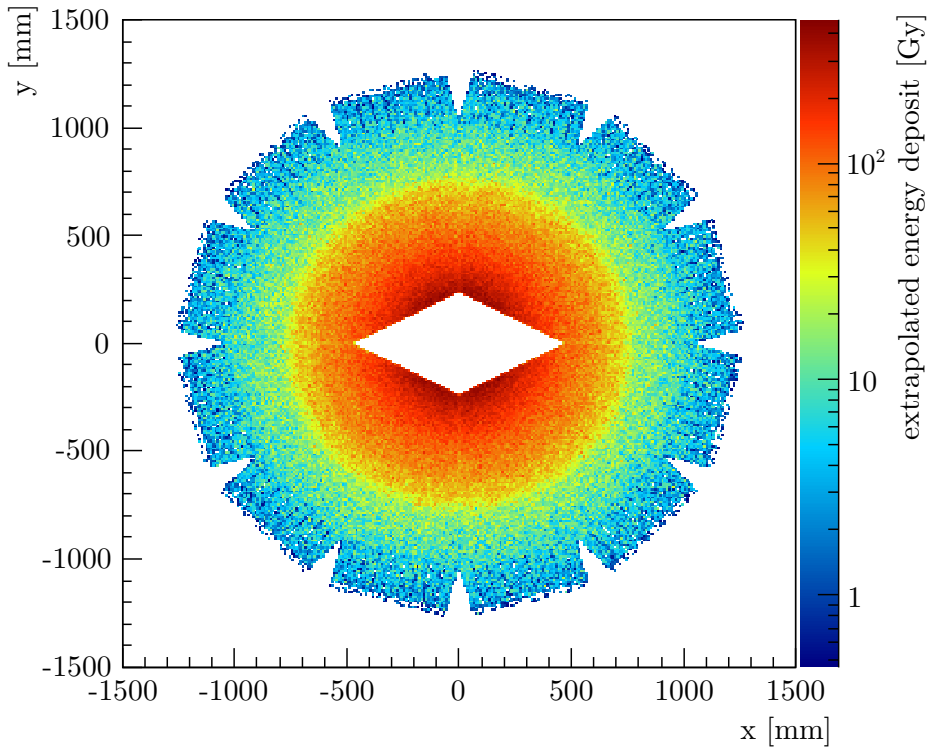


Figure 8.20.: Energy deposited in fused silica, extrapolated to a total lifetime of 10 years at a 50 % duty cycle and 10 MHz interaction rate. The extrapolation is based on a sample of $6.3{\cdot}10^5$ DPM Events at a beam momentum of 15 GeV/c, simulated using the Geant4 transport code via the VMC interface and the PandaRoot framework. The shadow of the electromagnetic calorimeter is clearly visible. The energy deposit at the outer perimeter, where sensors, electronics and focusing optics are located, is at a level of a few krad.

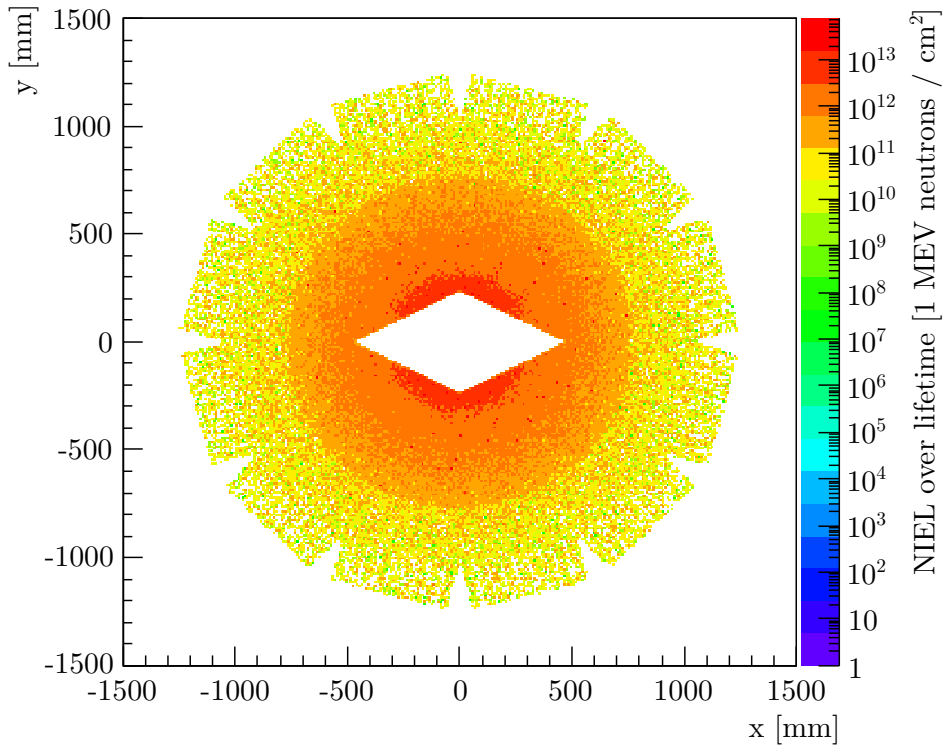


Figure 8.21.: 1 MeV neutron equivalent fluence per cm^2 at the Endcap DIRC area, extrapolated to a total lifetime of 10 years at a 50 % duty cycle and 10 MHz interaction rate. The extrapolation is based on the same sample as in figure 8.20. The fluence is computed by integrating the NIEL in silicon for several particle species using the Lindstroem data, as explained in section 5.6.2. The 1 MeV neutron fluence is a measure for the radiation damage in electronic devices. The interesting region is the outer perimeter, where the fluence is about 10^{11}. The fluence through the radiator is not important for the Endcap DIRC. It is important to denote that the operation with nuclear targets is not included in this estimation.

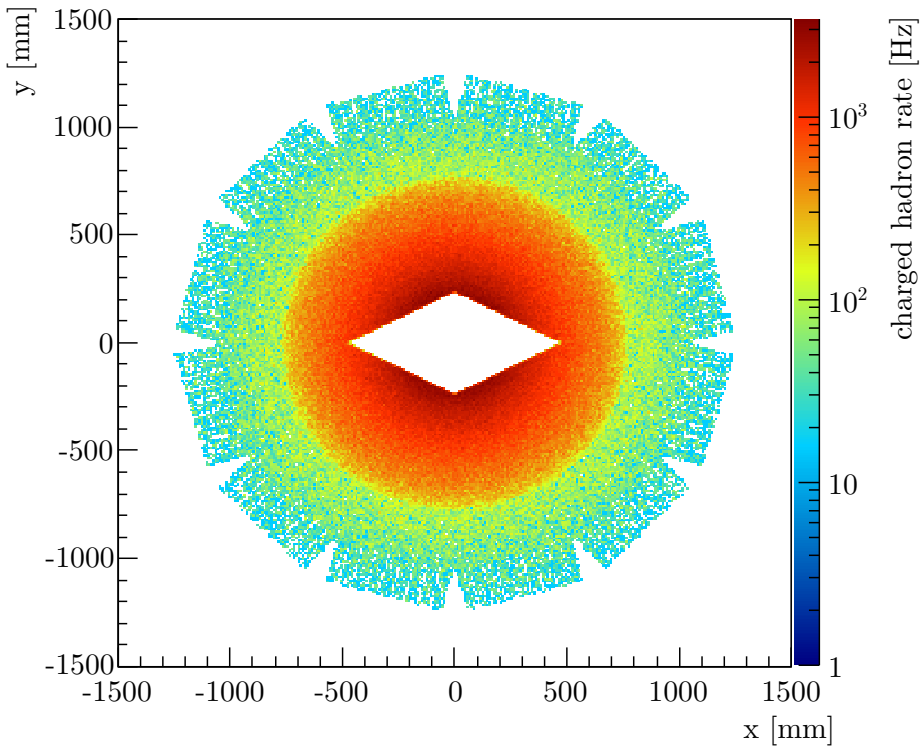


Figure 8.22.: Charged hadron rate ($E_{\text{kin.}} > 10\,\text{MeV}$) per cm^2 at the Endcap DIRC area, extracted from the same data sample as in figure 8.20. This rate can be used to estimate the rate of single event upsets in the frontend electronics.

8.3. FEE/DAQ concepts

The next step after the specification of components C1–C5 is the determination of the required number of ASICs and the anticipated data rate for components C4–C8. According to the PandaRoot simulation at the interaction rate of 20 MHz, the expected average hit rate per MCP-PMT is 5.6 MHz, corresponding to 18.7 kHz per channel. Hence, the device can operate at 10 MHz interaction rate over the full $\overline{\text{P}}$ANDA lifetime. In that case, the expected average rate per tube is only 2.8 MHz, well below the limit of 5.9 MHz which has been estimated in section 5.10.

The proposed front-end scheme is shown in Fig. 8.23. A minimum of five TOFPET ASICs would be required to read out the 300 channels of a single ROM. To keep the PCB design symmetric with equal signal paths for each of the three pixel-columns, it is proposed to use a set of six TOFPET ASICs, two for each column. The upper and the lower half of the 100 pixels are each connected to one ASIC. For small input signals and hence short dead times of the analog front-end, the TOFPET ASIC is limited by the data rate. Data is transmitted in 8b/10b encoded frames every 6.4 μs over one or two 160 MHz LVDS links at single or double data rate. Hence, the rate limitation depends on the specific link configuration.

As described in section 5.7, the GBT chipset is the best candidate for the readout as it is designed to work in presence of higher radiation levels, provides all necessary features and does not require the development of custom firmware, what would be the case if FPGAs were used. The GBT can provide 20 e-Links at 160 Mb/s or 10 e-Links at 320 Mb/s to interface the TOFPET ASICs. Due to the total number of six TOFPET ASICs, the data rate for the readout will be limited to 320 Mb/s in either case*. However, one link per ASIC at double data rate reduces the number of required LVDS lines between the front-end controller board and the ASIC board attached to the PMT. The 320 Mb/s link speed limits the number of hits per data frame to 40.

The hit multiplicity per TOFPET ASIC in 6.4 μs is shown in Fig. 8.24a. Because the typical multiplicity is lower than 20, no data loss is expected. There exist spurious events with multiplicities as high as 55, but as such high multiplicities are not observed in multiple ROMs at the same time, it can be assumed that these

*The scenario of five ASICs and 10 links does not work because one link is required to interface the slow-control ASIC of the GBT chipset. This ASIC provides the SPI link to control the individual TOFPET ASICs.

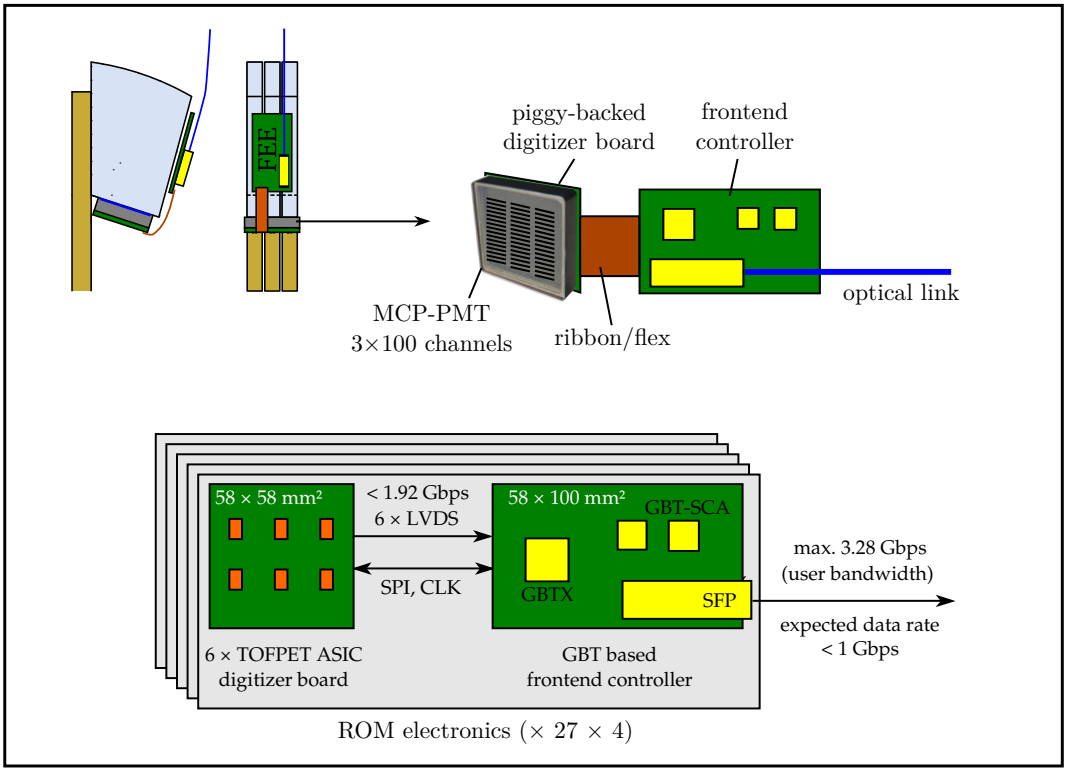


Figure 8.23.: Proposed configuration of front-end electronics and data concentrator (C4–C5).

are caused by showers in or close to the optics.

At an interaction rate of 20 MHz, the maximum (average) data rate is thus 938 Mb/s (280 Mb/s) for a single ROM and 25.3 Gb/s (7.6 Gb/s) for a SDM. Already one PCN ($\overline{\text{P}}$ANDA compute node, section 5.8) could be sufficient to accept the data from a single SDM. However, at the computing stage C7–C8 the throughput will be limited by the performance of the implemented algorithms and not by the link speed.

Fig. 8.25 illustrates the basic layout of the proposed approach to online processing. Each ROM has an individual fiber link which is connected to an FPGA on a PCN. Data transmitted by the ROM is continuously written to circular buffers. Each pixel has a dedicated buffer. This way, the data in the circular buffer is always ordered in time and older data will be overwritten by newer data. Hit sorting is realized implicitly by inserting the hits into the buffers. The buffer length has to be chosen according to the latency of the tracking algorithm. To buffer the 40 bit hits for a time of 10 ms at an average hit rate of 22 kHz one requires 110 kB per column

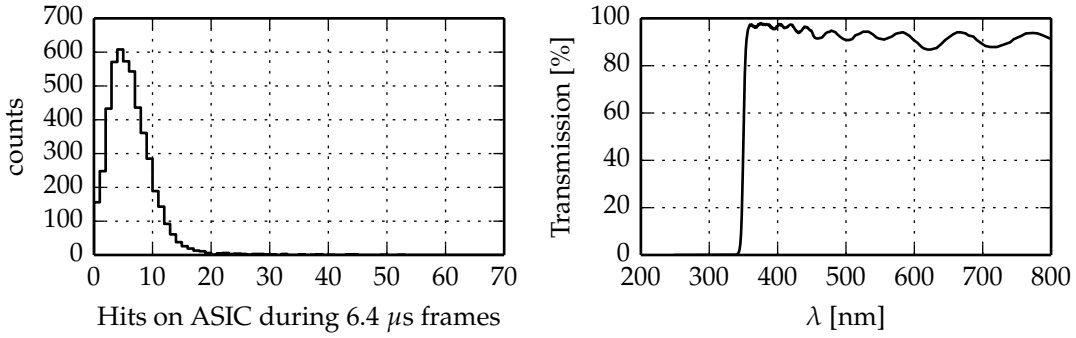


Figure 8.24.: (a): Hits per 6.4 μs data frame detected in a TOFPET-ASIC (50 connected channels) extracted from PandaRoot simulations at 20 MHz interaction rate. A maximum of 40 hits can be processed and transferred. (b): Transmission of a longpass edge filter (ZUL0350) from Asahi Spectra at normal incidence. This filter option will be referred to later in this document.

of 100 pixels and hence 330 kB per ROM. This value is insignificant compared to the 4 GB of RAM per FPGA.

The reconstruction logic is divided in parallel units comprising a controller block (C-block) and multiple reconstruction-blocks (R-blocks) which are executed in parallel. Each R-block comprises an intermediate, unsorted hit buffer and the reconstruction logic for a specific pixel-column. Geometric details of the individual column are directly encoded in the logic.

The reconstruction procedure is triggered when the tracking information is sent to a C-block. The time-stamp of the track determines the time window for hit selection from the ring-buffers.

All reconstruction results which have been presented in the previous section do account for only one rim reflection. As there exist three segments where the photon can be reflected prior to the detection, one has to evaluate three recon-nodes (e.a. Fig. 5.42 on page 156) in addition to the direct path. These nodes can be evaluated in parallel in each R-block. The resulting paths in radiator plane projection are then used to translate the hits in the R-block buffer to single photon Cherenkov angles. These are transmitted to the C-block where the computation of the track related t_0 and θ_c takes place. Particle identification is then realized by a predefined classifier. To mask the computation time required for the whole reconstruction process, multiple reconstruction units have to be implemented in parallel. Note

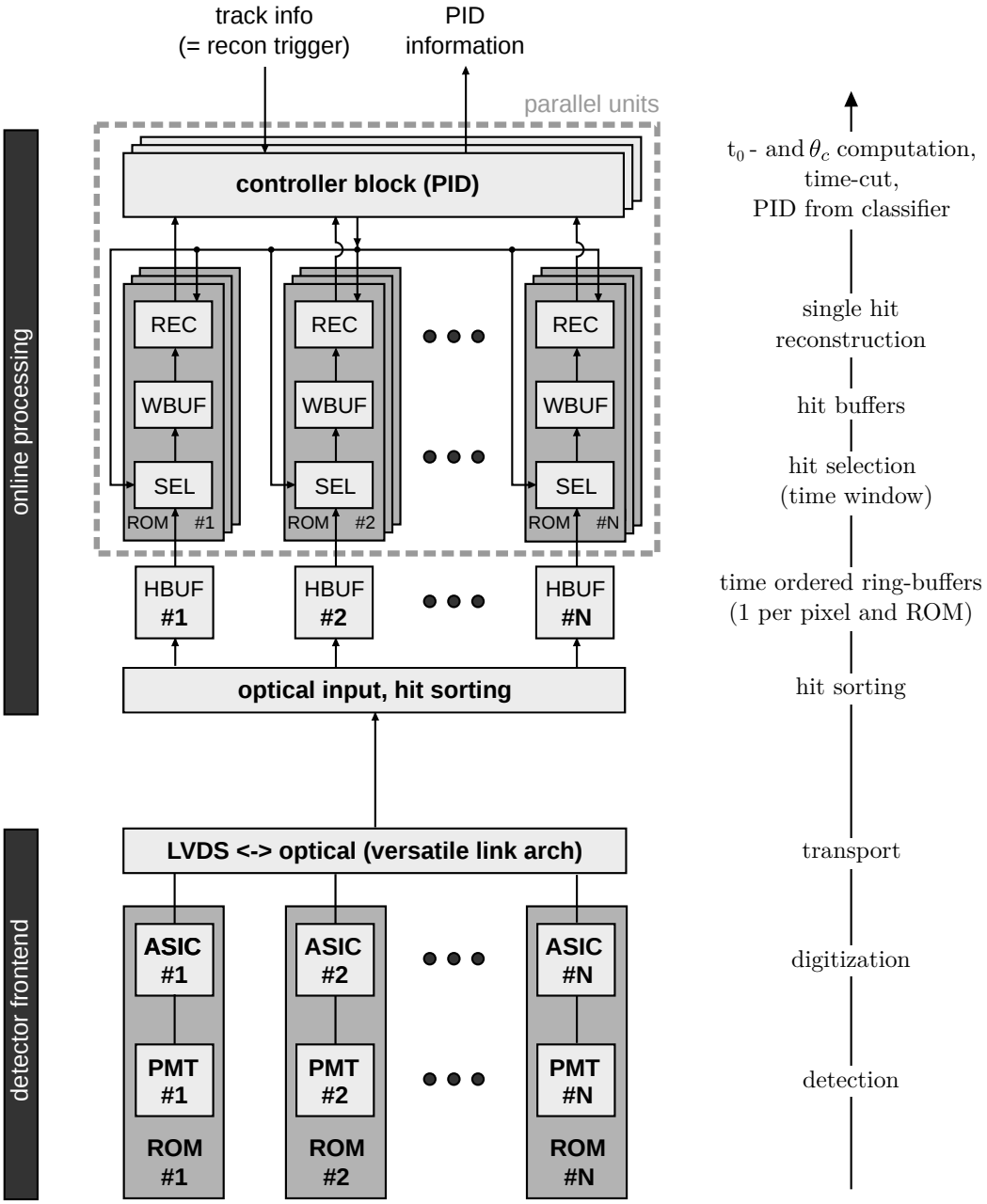


Figure 8.25.: Illustration of the proposed online processing scheme. The tasks of individual components is summarized in the simplified timeline at the right side of the figure.

that the latency of the reconstruction does not affect the size of the ring-buffers, as every R-block maintains a dedicated hit-buffer.

The number of FPGAs required can hardly be estimated in advance as it depends on the space which these blocks occupy on the FPGA as well as on the number of parallel units required to mask the processing time. The development of the FPGA firmware is a very complex task, spanning approximately one to two man-years. Alternatively, one can consider to implement the online processing on a fast multi-processor system.

Further, there is still room in hit-multiplicity and data-rate to allow for a design with a higher photon yield. A design without bandpass filter but a cutoff at 350 nm is a viable alternative (compare also table 5.12 on page 163). The performance of this design modification has been compared to the bandpass based design in Fig. 8.13a, table 8.2 and Fig. 8.19. The only hardware difference between both detector options is the different filter. A commercial filter option is the ZUL0350 longpass filter manufactured by Asahi Spectra Co., Ltd as $50 \times 50\,\mathrm{mm}^2$ filter on fused silica substrate. The transmission of this filter is shown in Fig. 8.24b.

8.4. Conclusion and outlook

A system design has been proposed which seamlessly describes and specifies the detector hardware, comprising the optical system, sensors and front-end electronics, as well as an approach to event reconstruction. Performance estimations obtained by reconstructing the detector response from time-based PandaRoot simulations show that the presented detector design is compatible with the requirements listed in section 5.1.

During design, strong emphasis has been put on modularization, while trying to minimize the number of required detector elements and thus complexity and cost. A wide range of hardware options has been considered and the choice for each option is motivated by the compliance with system requirements as well as by its impact on performance and cost.

First prototypes have been constructed and tested as a combined effort of the Gießen DIRC group [37]. These have been used successfully for particle identification and provide a first proof of principle for the Endcap DIRC technology. However, for budgetary reasons, these prototypes are far off the quality required for the final detector. At the time of writing, a sensor prototype with a 3×100

multi-anode structure is being characterized. Preliminary results demonstrate a low amount of cross-talk for single photon signals. Simulations show that the influence of $< 5\,\%$ of cross-talk between neighboring pixels on the expected detector performance is negligible. Prototypes for the optics have also been ordered according to the required specifications. Hence, the hardware required to realize this project will soon be available. A prototype of the system with near to final hardware is required to test and validate the system under realistic conditions.

A comparison of designs with bandpass filter and with longpass filter showed that a cutoff at 350 nm leads, despite the higher chromatic error, to a design with similar or better performance (e.g. Fig. 8.13a and table 8.2). As the sensor lifetime as well as the rate capability of the readout allow for more detected photons, one can consider to replace the bandpass filter by an edge filter, e.g. the ZUL0350 longpass filter manufactured by Asahi Spectra Co., Ltd. A further simplification of the design could be realized by a smart combination of coupling media which provide the wavelength cutoff.

Problems during prototyping may arise from the magnetic field at the sensor, which can cause image distortions, as well as from the ASIC development in case that the TOFPET ASIC is not sufficiently radiation tolerant in its current version. Further, the performance of the tracking system is a critical prerequisite. The Endcap DIRC will not be able to comply with the performance requirements if the tracking precision is significantly worse than specified in this document. The relatively low number of detected photons as well as the unknown event start-time do not allow to reconstruct the track parameters together with the hit-pattern.

Acknowledgments

Looking back at the recent years, I realize that scientific research can indeed be an adventurous journey. I have learned that research is not only about facts but also about people who form a unique kind of community.

There were ups and downs, unforeseen twists, lots of people to convince as well as inconvincible people, power struggles usually referred to as “politics” and finally a lot of dedication which is holding everything together. In the end, all of this is an integral part of the process resulting in the big picture which widens our horizon.

First of all, I want to thank all the great minds which have contributed indirectly to this work via their very own contribution to this big picture. Between countless publications, there are some enlightening papers which helped me a lot to understand the things which we cannot measure in the lab due to limited resources. Thanks to everyone who takes the time to write papers which target at the transport of knowledge, rather than the bare number of personal publications.

Next, I have to thank Prof. Dr. Michael Düren for the great opportunity to work on this project and the freedom he gave me during my research. I always appreciated his openness and his collaborative way of decision making.

Further, I wish to thank all the colleagues in Gießen for the nice time we had. Special thanks go to Roland Schmidt, Julian Rieke, Roberto Pérez Benito and Avetik Hayrapetian for all the interesting discussions, on- and especially off-topic, which enriched the time we’ve shared.

The most overwhelming experience during my PhD time was the birth of my two sons Johann and Jakob. I want to thank my family for the support and patience during all this time. Being a young father and researcher can be challenging, but I had the fortune of a very supportive environment at home and at work. Thanks to everyone who made this possible.

Bibliography

[1] LSR Low-energy Storage Ring: Technical Design Report. http://www.flairatfair.eu/typo3/fileadmin/files/documents/reports/DesignReportLSR_1.3.pdf. 8

[2] Technical Design Report – FAIR Antiproton Target and Separator, Draft March 2008. 8

[3] *PHOTOMULTIPLIER TUBES - Basics and Applications*. Hamamatsu Photonics, 3a edition, 2007. 119, 123

[4] Green Paper - The Modularized Start Version. http://www.fair-center.de/fileadmin/fair/publications_FAIR/FAIR_GreenPaper_2009.pdf, 2009. 7, 8

[5] S. Agostinelli et al. Geant4—a simulation toolkit. *Nucl. Instr. and Meth. A*, 506(3):250–303, July 2003. 88, 167

[6] M. Al-Turany et al. The FairRoot framework. *JPCS*, 396(2):22001, 2012. 168

[7] J. Allison et al. Geant4 developments and applications. *Nuclear Science, IEEE Transactions on*, 53(1):270–278, Feb. 2006. 88, 167

[8] A. Annenkov, M. Korzhik, and P. Lecoq. Lead tungstate scintillation material. *Nucl. Instr. and Meth. A*, 490(1–2):30 – 50, 2002. 18

[9] BABAR-DIRC Collaboration and J. Schwiening. Construction and performance of the BABAR-DIRC. *J. Instrum.*, 4(10):P10004, 2009. 51

[10] C. Baglin et al. Formation of the $\chi 1$ and $\chi 2$ charmonium resonances in antiproton-proton annihilation and measurements of their masses and total widths. *Phys. Lett. B*, 172(3–4):455–460, 1986. 21

[11] A. Barnyakov, M. Barnyakov, et al. Aerogel for FARICH detector. *Nucl. Instr. and Meth. A*, (0):–, 2014. 19

[12] M. Y. Barnyakov and A. V. Mironov. Photocathode aging in MCP PMT. *J. Instrum.*, 6(12):C12026, 2011. 129

[13] M. Bass et al. *Handbook of Optics, Third Edition Volume IV: Optical Properties of Materials, Nonlinear Optics, Quantum Optics (set)*. Handbook of Optics. McGraw-Hill Education, 2009. 74, 82

[14] M. Bass et al. *Handbook of Optics – Volume I*. McGraw-Hill, 2010. 78, 79

[15] G. Battistoni et al. The FLUKA code: description and benchmarking. *AIP Conference Proceedings*, 896(1), 2007. 168

[16] S. Bäumer. *Handbook of Plastic Optics*. Wiley, 2006. 81

[17] D. Beaulieu and N. Sullivan. Silicon microchannel plate devices with smooth pores and precise dimensions, July 20 2010. US Patent 7,759,138. 128

[18] J. Beringer et al. Review of particle physics. *Phys. Rev. D*, 86:010001, Jul 2012. 81, 85

[19] C. F. Bohren and D. R. Huffman. *Absorption and Scattering of Light by Small Particles*. Wiley-VCH Verlag GmbH, 2007. 76, 77

[20] A. V. Boukharov et al. Dynamics of Cryogenic Jets: Non-Rayleigh Breakup and Onset of Nonaxisymmetric Motions. *Phys. Rev. Lett.*, 100:174505, May 2008. 11, 57

[21] A. Breskin et al. Progress in gaseous photomultipliers for the visible spectral range. *Nucl. Instr. and Meth. A*, 623(1):318 – 320, 2010. 111

[22] A. Britting et al. Lifetime-issues of MCP-PMTs. *J. Instrum.*, 6(10):C10001, 2011. 128

[23] A. Capella et al. Dual parton model. *Physics Reports*, 236(4–5):225–329, 1994. 171

[24] R. Chechik and A. Breskin. Advances in gaseous photomultipliers. *Nucl. Instr. and Meth. A*, 595(1):116 – 127, 2008. 111

[25] A. Czasch et al. Position- and time-sensitive single photon detector with delay-line readout. *Nucl. Instr. and Meth. A*, 580(2):1066–1070, 2007. 127, 133

[26] C. G. Darwin and W. H. Watson. The constants of the magnetic dispersion of light. *Proceedings of the Royal Society of London. Series A*, 114(768):474–490, 1927. 79

[27] J. DeFazio. personal communication. 102, 121

[28] A. di Bona et al. Development, operation and analysis of bialkali antimonide photocathodes for high-brightness photoinjectors. *Nucl. Instr. and Meth. A*, 385(3):385–390, 1997. 121

[29] A. Dolinskii et al. Antiproton complex at the FAIR project. *Nucl. Instr. and Meth. A*, 629(1):16–24, 2011. 7

[30] D. B. Doyle et al. Space radiation sensitivity of glasses: first results toward a comprehensive dose coefficients database. In *Proc. SPIE*, volume 4134, pages 261–267, 2000. 108

[31] J. Engelfried. Cherenkov light imaging—Fundamentals and recent developments. *Nucl. Instr. and Meth. A*, 639(1):1–6, 2011. 34

[32] J. Estrera, M. Iosue, A. Giordana, and J. Glesener. Microchannel plate having an enhanced coating, May 28 2002. US Patent 6,396,049. 127

[33] R. Faccini et al. Silicon photo-multiplier radiation hardness tests with a beam controlled neutron source. *Nucl. Instr. and Meth. A*, 623(3):921–926, 2010. 115

[34] C. Färber et al. Radiation tolerance tests of SRAM-based FPGAs for the potential usage in the readout electronics for the LHCb experiment. *J. Instrum.*, 9(02):C02028, 2014. 149, 150

[35] E. Fioravanti and the PANDA Collaboration. Experimental overview of the PANDA experiment. *JPCS*, 503(1):12030, 2014. 22

[36] M. Fiorini et al. Radiation hardness tests and characterization of the CLARO-CMOS, a low power and fast single-photon counting ASIC in 0.35 micron CMOS technology. *Nucl. Instr. and Meth. A*, (0):–, 2014. 146

[37] K. Föhl et al. First particle identification with a Disc-DIRC detector. *Nucl. Instr. and Meth. A*, 732(0):346–351, 2013. 228

[38] E. Fuller et al. Radiation testing update, SEU mitigation, and availability analysis of the Virtex FPGA for space reconfigurable computing, presented at the IEEE Nuclear and Space Radiation Effects Conference. In *Proc. International Conference on Military and Aerospace Programmable Logic Devices*, 2000. 149

[39] K. Föhl. A Focussing Disc DIRC design for particle identification in PANDA. *Hyperfine Interactions*, 194:171–175, 2009. 43, 95, 164, 187, 197

[40] E. Garutti. Silicon photomultipliers for high energy physics detectors. *J. Instrum.*, 6(10):C10003, 2011. 115

[41] H. Grabas et al. RF strip-line anodes for psec large-area MCP-based photodetectors. *Nucl. Instr. and Meth. A*, 711(0):124–131, 2013. 133

[42] Y. Haemisch et al. Fully Digital Arrays of Silicon Photomultipliers (dSiPM) – a Scalable Alternative to Vacuum Photomultiplier Tubes (PMT). *Physics Procedia*, 37(0):1546–1560, 2012. 114, 117, 118, 183

[43] N. Hagen and T. S. Tkaczyk. Compound prism design principles, I. *Appl. Opt.*, 50(25):4998–5011, Sept. 2011. 102

[44] M. Hause. The sysml modelling language. In *Fifteenth European Systems Engineering Conference*, 2006. 3

[45] T. Hayakawa. Performance tests of a full scale prototype of the Belle II TOP counter with cosmic muons and 2.1 GeV/c positron beam. *Nucl. Instr. and Meth. A*, (0):–, 2014. 38, 39

[46] V. L. Highland. Some practical remarks on multiple scattering. *Nucl.Instr. and Meth.*, 129(2):497 – 499, 1975. 84

[47] M. Hirose et al. Development of a 24-anode linear-array fine-mesh PMT with 85% photoelectron detection efficiency and 100 ps TTS at B≤1 T. *Nucl. Instr. and Meth. A*, 460(2–3):326–335, 2001. 124, 136

[48] M. Hoek. Tailoring the radiation hardness of fused silica. *Nucl. Instr. and Meth. A*, 639(1):227 – 230, 2011. 107

[49] M. Hoek et al. Radiation hardness study on fused silica. *Nucl. Instr. and Meth. A*, 595(1):190 – 193, 2008. 107

[50] M. Hoek et al. The PANDA Barrel DIRC detector. *Nucl. Instr. and Meth. A*, (0):–, 2014. 38, 39

[51] T. Hokuue et al. R&D report on fine-mesh multi-anode PMT with TTS=100 ps under B≤1 T. *Nucl. Instr. and Meth. A*, 494(1–3):436–440, 2002. 124

[52] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science and Engineering*, 9(3):90–95, 2007. 178

[53] I. Hřivnáčová. The Geant4 Virtual Monte Carlo. *JPCS*, 396(2):22024, 2012. 168

[54] P. A. Čerenkov. Visible radiation produced by electrons moving in a medium with velocities exceeding that of light. *Phys. Rev.*, 52:378–379, Aug 1937. 32

[55] K. Inami et al. Lifetime of MCP–PMT. *Nucl. Instr. and Meth. A*, 564(1):204–211, 2006. 127, 128

[56] M. Izerrouken, A. Meftah, and M. Nekkab. Color centers in neutron-irradiated Y3Al5O12, {CAF2} and LiF single crystals. *J. Lumin.*, 127(2):696–702, 2007. 109

[57] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. 178

[58] T. Kamae, K. Arai, T. Ohtsuka, N. Tsuchida, and H. Aihara. Focussing DIRC, a new compact Cherenkov ring imaging device. *Nucl. Instr. and Meth. A*, 382(3):430 – 440, 1996. 43, 96, 164

[59] M. Kapusta et al. Breakthrough in quantum efficiency of bi-alkali photocathode PMTs. In *Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE*, volume 1, pages 73–77, Oct 2007. 121

[60] U. Kerzel. The LHCb RICH detectors. *Nucl. Instr. and Meth. A*, 617(1–3):283–286, 2010. 34, 35

[61] B. Ketzer et al. Performance of triple GEM tracking detectors in the COMPASS experiment. *Nucl. Instr. and Meth. A*, 535(1–2):314–318, 2004. 17

[62] A. Khoukaz. Internal targets for the PANDA experiment. *PoS*, STORI11:036, 2011. 9, 11

[63] H. Kim et al. A prototype TOF PET detector module using a micro-channel plate photomultiplier tube with waveform sampling. *Nucl. Instr. and Meth. A*, 662(1):26–32, 2012. 133

[64] E. Kohler. Design and performance of the future cluster-jet target for PANDA at FAIR. *PoS*, STORI11:063, 2011. 10

[65] S. Korpar et al. Photonis MCP PMT as a light sensor for the Belle II RICH. *Nucl. Instr. and Meth. A*, 639(1):162 – 164, 2011. 130

[66] S. Korpar et al. A 144-channel HAPD for the Aerogel RICH at Belle II. *Nucl. Instr. and Meth. A*, (0):–, 2014. 133

[67] B. Kroek. Entwicklung einer schnellen Photonenauslese für DIRC-Detektoren. PhD thesis, Justus-Liebig-Universität Gießen, (in preparation). 139

[68] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instr. and Meth. A*, 462(1–2):152–155, 2001. 48, 170

[69] J. S. Lange, M. Galuska, T. Gessler, W. Kuhn, S. Kunze, et al. Prospects for X(3872) Detection at Panda. *AIP Conf.Proc.*, 1374:549–552, 2011. 22

[70] A. Lehmann et al. Improved lifetime of microchannel-plate PMTs. *Nucl. Instr. and Meth. A*, (0):–, 2014. 127, 128, 129

[71] A. Lehrach et al. Beam performance and luminosity limitations in the high-energy storage ring (HESR). *Nucl. Instr. and Meth. A*, 561(2):289 – 296, 2006. 8, 11, 56, 57

[72] A. Lehrach et al. Beam dynamics of the high-energy storage ring (HESR) for FAIR. *International Journal of Modern Physics E*, 18(02):420–429, 2009. 56, 57

[73] D. W. G. S. Leith. DIRC—the particle identification system for BaBar. *Nucl. Instr. and Meth. A*, 494(1–3):389–401, 2002. 36, 37

[74] M. Lutz et al. Physics Performance Report for PANDA: Strong Interaction Studies with Antiprotons. 2009. 23, 48

[75] G. R. Lynch and O. I. Dahl. Approximations to multiple coulomb scattering. *Nucl. Instr. and Meth. B*, 58(1):6 – 10, 1991. 84

[76] P. Martin, T. Hernandez, and E. Hodgson. Commercial dielectric coated mirrors for ITER diagnostic applications. *Fusion Engineering and Design*, 84(7-11):1259–1262, June 2009. 100

[77] S. Matoba et al. Measurement of the absolute sensitivity of a high-sensitivity microchannel plate. *JPCS*, 388(14):142018, 2012. 126

[78] W. L. McLaughlin et al. Radiation-induced color centers in LiF for dosimetry at high absorbed dose rates. *Nucl. Instr. and Meth.*, 175(1):17–18, 1980. 108

[79] O. Merle. Development of reconstruction methods for the PANDA Disc DIRC. Diplomarbeit, Justus-Liebig-Universität Gießen, 2009. 41, 67, 154, 172

[80] O. Merle et al. Reconstruction methods for the PANDA Time of Propagation Disc DIRC. *J. Instrum.*, 4(09):P09008, 2009. 44, 172

[81] O. Merle et al. Simulation and reconstruction of photon patterns in the PANDA 3D Disc DIRC. *J. Instrum.*, 7(01):C01068, 2012. 190

[82] O. Merle et al. Development of an Endcap DIRC for PANDA. *Nucl. Instr. and Meth. A*, (0):–, 2014. 211

[83] J. Messchendorp. Physics with Charmonium - A few recent highlights of BESIII. *PoS*, Bormio2013:043, 2013. 23

[84] M. Montecchi and Q. Ingram. Study of some optical glues for the compact muon solenoid at the large hadron collider of CERN. *Nucl. Instr. and Meth. A*, 465(2–3):329 – 345, 2001. 97

[85] P. Moreira et al. The GBT Project. *Proceedings, Topical Workshop on Electronics for Particle Physics (TWEPP09)*, 2009. 151

[86] T. Mori et al. Lifetime-extended MCP-PMT. *Nucl. Instr. and Meth. A*, 629(1):111–117, 2011. 128

[87] B. J. Morosov. Proposal of a disc DIRC. `https://www-alt.gsi.de/documents/DOC-2004-Dec-68.html`, 12 2004. 43, 102

[88] D. A. Morozov et al. Test beam study of the PANDA shashlyk calorimeter prototype. *JPCS*, 160(1):012021, 2009. 18

[89] E. Munin, J. A. Roversi, and A. B. Villaverde. Faraday effect and energy gap in optical materials. *J. Phys. D*, 25(11):1635, 1992. 83

[90] Y. Musienko. Advances in multipixel Geiger-mode avalanche photodiodes (silicon photomultiplies). *Nucl. Instr. and Meth. A*, 598(1):213–216, 2009. 113, 114

[91] I. Nakamura. Radiation damage of pixelated photon detector by neutron irradiation. *Nucl. Instr. and Meth. A*, 610(1):110 – 113, 2009. 115

[92] K. Nakamura et al. Latest bialkali photocathode with ultra high sensitivity. *Nucl. Instr. and Meth. A*, 623(1):276–278, 2010. 121

[93] A. Nordhage et al. Tracked pellets; a way to improve the efficiency of charmonium studies. *Nucl. Instr. and Meth. A*, 568(2):561 – 565, 2006. 11

[94] R. W. Novotny et al. High-quality PWO crystals for the PANDA-EMC. *JPCS*, 293(1):012003, 2011. 18

[95] T. E. Oliphant. Python for scientific computing. *Computing in Science Engineering*, 9(3):10–20, 2007. 178

[96] I. I. Orlovskiy and K. Vukolov. Thermal and neutron tests of multilayered dielectric mirrors. *Fusion Engineering and Design*, 74(1–4):865–869, 2005. 100

[97] PANDA Collaboration. Technical Design Report for PANDA Electromagnetic Calorimeter (EMC). *ArXiv e-prints*, Oct. 2008. 18

[98] PANDA Collaboration. Technical Design Report for the: PANDA Micro Vertex Detector. *ArXiv e-prints*, July 2012. 15, 63

[99] PANDA Collaboration. Technical design report for the panda muon system. *Internal draft*, 2012. 20

[100] PANDA Collaboration. Technical design report for the PANDA (AntiProton Annihilations at Darmstadt) Straw Tube Tracker. *The European Physical Journal A*, 49(2):1–104, 2013. 15, 16

[101] M. Piccinini et al. Optical spectroscopy and imaging of colour centres in lithium fluoride crystals and thin films irradiated by 3 MeV proton beams. *Nucl. Inst. and Meth. B*, 326(0):72–75, 2014. 109

[102] A. Piegari et al. Optical performance of narrow-band transmittance filters under low-and high-energy proton irradiation. In *Proc. SPIE*, volume 8168, pages 816814–816818, 2011. 100

[103] D. Pinnow et al. Fundamental optical attenuation limits in the liquid and glassy state with application to fiber optical waveguide materials. *Appl. Phys. Lett.*, 22(10):527–529, 1973. 77

[104] M. N. Polyanskiy. Refractive index database. `http://refractiveindex.info`. 95

[105] Y. Qiang et al. Radiation hardness tests of SiPMs for the {JLab} Hall D Barrel calorimeter. *Nucl. Instr. and Meth. A*, 698(0):234–241, 2013. 115, 183

[106] A. D. Rakić. Algorithm for the determination of intrinsic optical constants of metal films: application to aluminum. *Appl. Opt.*, 34(22):4755–4767, Aug. 1995. 95

[107] A. D. Rakić et al. Optical properties of metallic films for vertical-cavity optoelectronic devices. *Appl. Opt.*, 37(22):5271–5283, Aug. 1998. 95

[108] D. Renker and E. Lorenz. Advances in solid state photon detectors. *J. Instrum.*, 4(04):P04004, 2009. 116

[109] J. Ritman. Event rates in PANDA. `https://panda-wiki.gsi.de/foswiki/pub/Tracking/MeetingMinutes/EventratesinPANDA.pdf`, 10 2011. 54, 63, 192

[110] A. Rivetti. personal communication. 145

[111] A. Roberts. A new type of Čerenkov detector for the accurate measurement of particle velocity and direction. *Nuclear Instruments and Methods*, 9(1):55–66, 1960. 33

[112] D. A. Roberts et al. Results from the FDIRC prototype. *Nucl. Instr. and Meth. A*, (0):–, 2014. 38, 39

[113] Z. Sadygov et al. A microchannel avalanche photodiode with a fast recovery time of parameters. *Technical Physics Letters*, 39(6):498–500, 2013. 115

[114] M. Saito, Y. Saito, K. Asamura, and T. Mukai. Spatial charge cloud size of microchannel plates. *Rev. Sci. Instrum.*, 78:023302, 2007. 137

[115] F. Sauli. Gem: A new concept for electron amplification in gas detectors. *Nucl. Instr. and Meth. A*, 386(2–3):531 – 534, 1997. 16

[116] P. Schönmeier et al. Disc DIRC endcap detector for PANDA at FAIR. *Nucl. Instr.*

and Meth. A, 595:108–111, Sept. 2008. 41, 43

[117] S. Schröder et al. Bulk scattering properties of synthetic fused silica at 193 nm. *Opt. Express*, 14(22):10537–10549, Oct 2006. 77, 171

[118] J. Seguinot and T. Ypsilantis. Photoionization and Cherenkov ring imaging. *Nuclear Instruments and Methods*, 142(3):377–391, 1977. 33

[119] A. Shih, J. Yater, C. Hor, and R. Abrams. Secondary electron emission studies. *Applied Surface Science*, 111(0):251–258, 1997. 123

[120] O. H. W. Siegmund et al. Large area event counting detectors with high spatial and temporal resolution. *J. Instrum.*, 9(04):C04002, 2014. 132

[121] O. H. W. Siegmund and LAPPD Collaboration. Atomic layer deposited borosilicate glass microchannel plates for large area event counting detectors. *Nucl. Instr. and Meth. A*, 695(0):168–171, 2012. 128

[122] P. Sievers et al. Concept for the Antiproton Production Target at FAIR. *Conf.Proc.*, C1205201:2570–2572, 2012. 8

[123] S. Spataro. Event Reconstruction in the PandaRoot framework. *JPCS*, 396(2):22048, 2012. 168

[124] W. E. Spicer and A. Herrera-Gomez. Modern theory and applications of photocathodes. 2022:18–35, 1993. 118, 121

[125] M. Starič. Pattern recognition for the time-of-propagation counter. *Nucl. Instr. and Meth. A*, 639(1):252–255, 2011. 158

[126] R. M. Sternheimer, S. M. Seltzer, and M. J. Berger. Density effect for the ionization loss of charged particles in various substances. *Phys. Rev. B*, 26:6067–6076, Dec 1982. 86

[127] B. Streetman and S. Banerjee. *Solid State Electronic Devices*. Prentice Hall Series in Solid State Physical Electronics. Pearson Prentice Hall, 2006. 115

[128] N. Sullivan, S. Bachman, P. De Rouffignac, A. Tremsin, and D. Beaulieu. Microchannel plate devices with tunable resistive films, Dec. 24 2009. US Patent App. 12/143,732. 128

[129] G. Summers et al. Damage correlations in semiconductors exposed to gamma, electron and proton radiations. *Nuclear Science, IEEE Transactions on*, 40(6):1372–1379, Dec 1993. 147

[130] A. Taschner, A. Khoukaz, E. Kohler, and H.-W. Ortjohann. Velocity determination of hydrogen clusters at a cluster-jet target. *PoS*, STORI11:065, 2011. 10

[131] T. Tick et al. Status of the Timepix MCP-HPD development. *J. Instrum.*, 5(12):C12020, 2010. 142

[132] K. Toh, K. Sakasai, T. Nakamura, K. Soyama, and T. Shikama. Effects of neutrons and gamma-rays on polymethylmethacrylate plastic optical fiber. *J. Nucl. Mater.*, 417(1–3):814–817, 2011. 81

[133] TOPAS Advanced Polymers. Cyclic olefin copolymer (COC) product information. `http://www.topas.com/sites/default/files/files/topas_product-brochure_english.pdf`, 2006. 81

[134] A. Tremsin, P. de Rouffignac, N. Sullivan, D. Beaulieu, and M. Potter. Microchannel plate devices with multiple emissive

layers, Aug. 27 2009. US Patent App. 12/038,254. 128

[135] A. Tremsin et al. High speed multi-channel charge sensitive data acquisition system with self-triggered event timing. *Nuclear Science, IEEE Transactions on*, 56(3):1148–1152, June 2009. 132

[136] F. Uhlig et al. Performance studies of microchannel plate PMTs. *Nucl. Instr. and Meth. A*, 695(0):68–70, 2012. 202

[137] A. Vacheret et al. Characterization and simulation of the response of Multi-Pixel Photon Counters to low light levels. *Nucl. Instr. and Meth. A*, 656(1):69–83, 2011. 122

[138] J. Vallerga et al. Optical MCP image tube with a quad Timepix readout: initial performance characterization. *J. Instrum.*, 9(05):C05055, 2014. 133, 134

[139] J. V. Vallerga and J. B. McPhate. Optimization of the readout electronics for microchannel plate delay line anodes. In *Proc. SPIE*, volume 4139, pages 34–42, 2000. 141

[140] A. Vasilescu and G. Lindstroem. Notes on the fluence normalisation based on the niel scaling hypothesis, rose/tn/2000-02. Technical report, CERN, 2000. 147, 148, 149

[141] M. A. Vincenti et al. Spectroscopic investigation of f, f 2 and f 3 + color centers in gamma irradiated lithium fluoride crystals. *Mater. Sci. Eng.*, 15(1):012053, 2010. 108

[142] B. Voss and J. Kunkel. GEM–trackers for the PANDA target spectrometer - design studies. *GSI scientific report 2008*, 2008. 15

[143] M. R. Wayne. Visible light photon counters and the DØscintillating fiber tracker. *Nucl. Instr. and Meth. A*, 387(1–2):278–281, 1997. 111

[144] R. Wilson et al. The DIRC particle identification system for the BaBar experiment. *Nucl. Instr. and Meth. A*, 538(1–3):281 – 357, 2005. 19, 154, 155, 158

[145] G. R. Wirtenson and R. H. White. Long-term bleaching of optical glasses darkened by Co60 ionizing radiation. In *Proc. SPIE*, volume 3124, pages 53–59, 1997. 80, 108

[146] G. R. Wirtenson and R. H. White. Rate of long-term bleaching in FK 51 optical glass darkened by Co^{60} ionizing radiation at dose rates of 10 krad/hr and 7 rad/hr [3124-08]. In E. W. Taylor, editor, *Photonics for Space Environments V*, volume 3124 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 60, 1997. 80, 108

[147] J. L. Wiza. Microchannel plate detectors. *Nucl. Instr. and Meth.*, 162:587–601, 1979. 125

[148] C. Xu et al. Influence of X-ray irradiation on the properties of the Hamamatsu silicon photomultiplier S10362-11-050C. *Nucl. Instr. and Meth. A*, 762(0):149–161, 2014. 115, 183

[149] H. Xu et al. An ATCA-based High Performance Compute Node for Trigger and Data Acquisition in Large Experiments. *Physics Procedia*, 37(0):1849–1854, 2012. 153

[150] K. Yamamoto et al. Development of Multi-Pixel Photon Counter (MPPC). In *Nuclear Science Symposium Conference Record, 2006. IEEE*, volume 2, pages 1094–1097, Oct 2006. 113

[151] P. Yoder. *Opto-Mechanical Systems Design, Third Edition*. Optical Science and Engineering. Taylor & Francis, 2005. 80

[152] J. H. Yoo and I. K. Sethi. An ellipse detection method from the polar and pole definition of conics. *Pattern Recognition*, 26(2):307–315, 1993. 246

A. Appendix

A.1. Algorithm to construct an initial focusing surface for optimization

The purpose of the focusing element (FEL) is to convert the angular information of the incoming light to a position on one or more sensors. Light of a given angle of incidence can enter at any point on the radiator-FEL interface. Therefore one has to consider a bundle of parallel rays which has to be focused on a certain spot (Fig. A.1, right). As the interval of possible angles $\varphi \in [\varphi_{\min}, \varphi_{\max}] \subset \mathbb{R}$ is continuous, adjacent bundles overlap each other and share a section of the focusing surface.

Perfect focusing of parallel light can be realized with a parabolic surface if the incident light is parallel to the axis of symmetry. At $x = 0$, a parabola $y = ax^2 + c$ is reflecting all rays to a single point $F = (0; (4a)^{-1} + c)$, called focus (fig. A.1, left). In fact, every continuous curve which focuses parallel light to a single point has to be a parabola*, so it is the only solution existing. There is no perfect solution for the focusing of non-parallel light. In conclusion, the FEL surface has to be a compromise of different parabola. This is the underlying idea of the developed algorithm which is outlined in the flow chart shown in figure A.2 and explained in the following paragraphs.

The parabola is completely defined by the focus F and the parameter a. For a given bundle and fixed focus F, the slope and position of the reflecting parabola segment can be adjusted by varying the parameter a as shown in figure A.1. This is also valid in case of a symmetry-axis at $x \neq 0$,as it is just a translation of the coordinate-system.

The prior constraints or "design parameters" of a focusing element can be

*Let F be an arbitrary focus and P a point on the focusing curve $f(x)$. It is always possible to construct a parabola $p(x)$ with focus F so that it goes through P. To reflect a ray at point P in the direction of F, $f(x)$ must have the same slope as $p(x)$: $f'(P_x) = p'(P_x)$. This is also true for $f(P_x + \epsilon) = p(P_x + \epsilon)$ and by induction for any other point, so $f(x) = p(x)$.

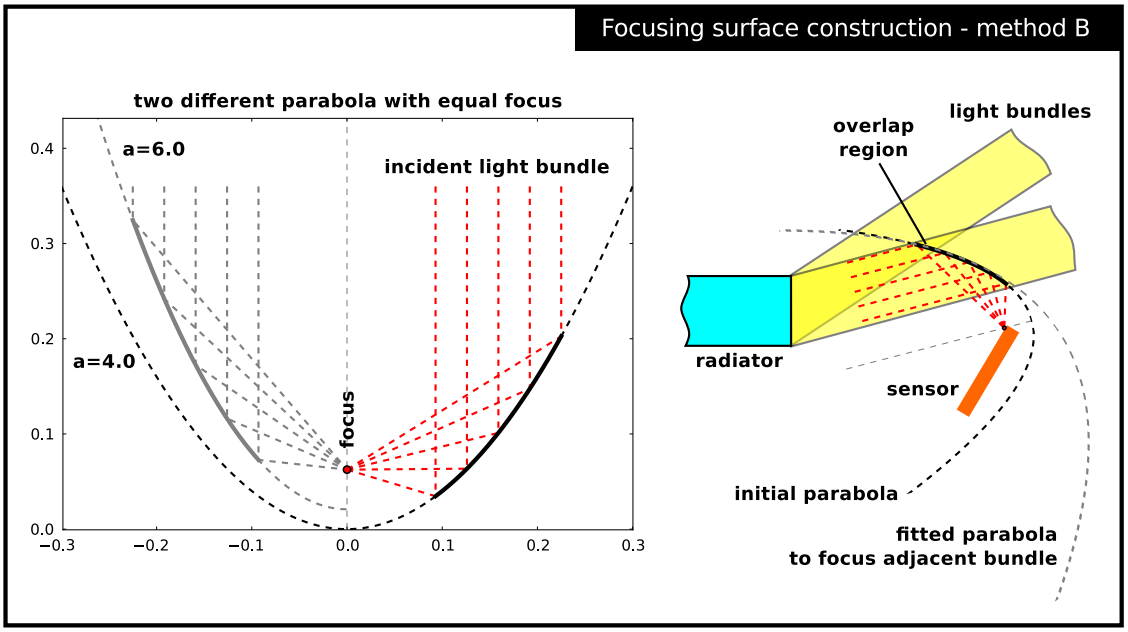


Figure A.1.: A parabola can be used to focus a ray bundle to a single point on the sensor. Because bundles of different angles overlap significantly in proximity to the radiator, the parabola segment is also illuminated with light which is not parallel to its axis of symmetry what leads to aberration. In the overlap region, two neighboring parabola have to be as similar as possible to minimize this error.

represented by a tuple $\mathbb{P} = (d_r, \varphi_{\min}, \varphi_{\max}, S, d_s, \alpha_s)$ where d_r denotes the radiator thickness, $\varphi_{\min}$ and $\varphi_{\max}$ the minimum and maximum angle between the x axis and the entering rays and $S = (s_x; s_y)$, d_s, α_s the sensors position, length and its orientation in respect to the y axis. The full sensor area shall be used for imaging. The sensor position S describes one endpoint of the sensor and therefore also the focus of the parabola for the bundle with angle $\varphi_{\min}$ (or $\varphi_{\max}$, depending on the side of the sensor where S is located). Also the symmetry-axis is defined by S and $\varphi_{\min}$ as illustrated in Fig. A.1. Only the parameter a has to be defined. This parameter, named "start parameter", can also be expressed as the distance from the lower radiator point, or as the tangent at the beginning of the focusing surface.

The algorithm uses a fixed number of different ray-bundles (e.g. three). The first and last bundles at $\varphi_{\min}, \varphi_{\max}$ are always taken into respect. After the definition of the start parameter the algorithm computes N_p points on the part of the reflecting parabola segment which is also intersected by rays from the adjacent bundle (the part on the surface which is shared by the two bundles). These points are used to

find the focus and parameter a of the parabola which focuses the adjacent bundle. This parabola is determined by a constrained fit of a two dimensional error function $f(u,a)$ where a is the parabola parameter and u the position on the sensor in a parametric form. This function computes the mean-squared distance to the points in respect to a and the focal point given by u. After the parabola is found, the points in the overlap zone to the next bundle are used to compute the following parabola. This iteration is repeated until a parabola is computed for each of the bundles. In a final step, N_q points on each parabola segment are computed and a 5-th degree polynomial is fitted through this point set. This polynomial is the compromise surface.

This surface is used as start for an iterative refinement step in which a merit function which computes the imaging error for $N_m b$ ray bundles is minimized by means of a downhill simplex algorithm. In some cases this leads to a significant better focusing quality, in others the effect is rather minimal. Hence, the construction of the initial surface is already close to the local optimum.

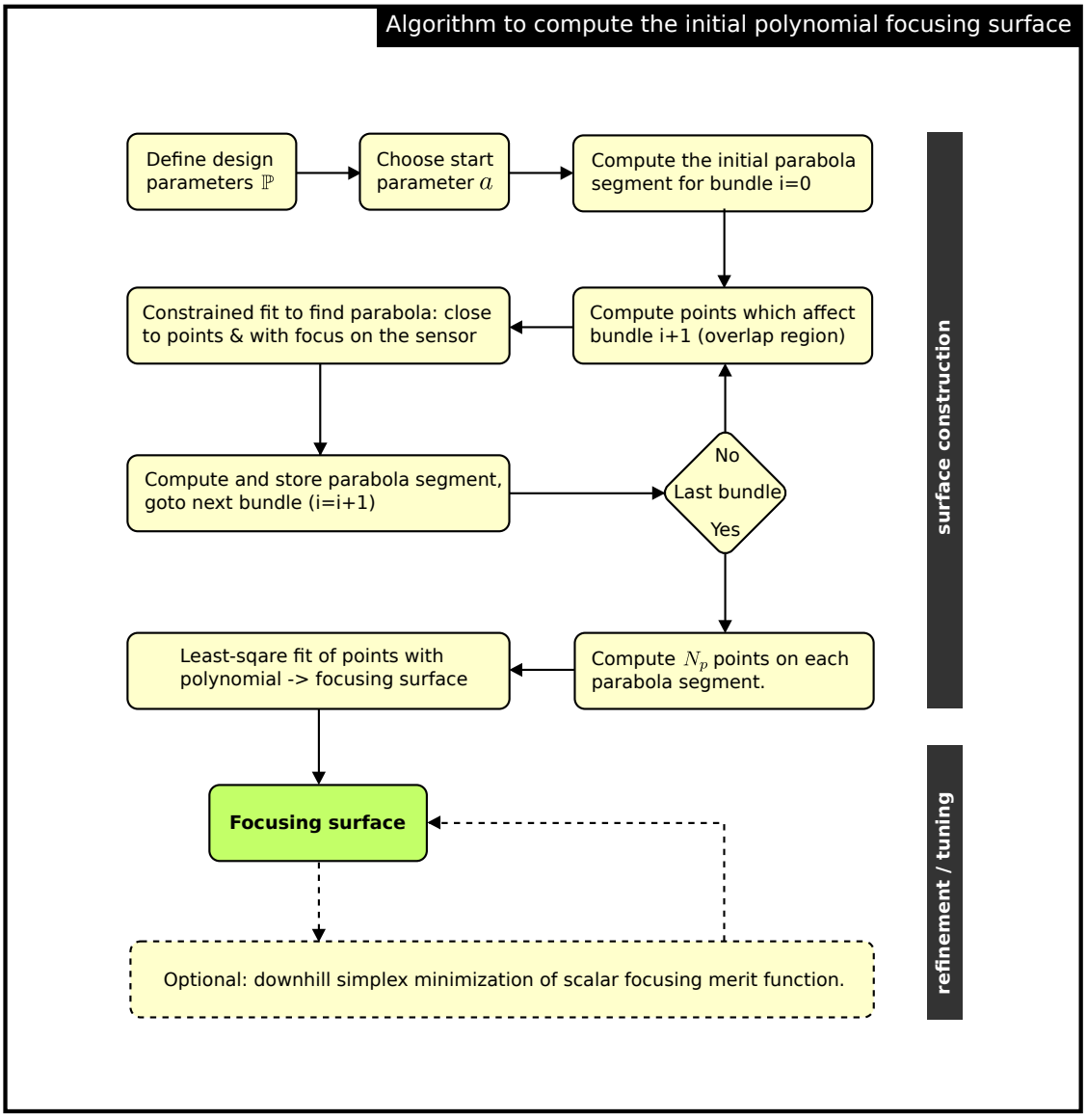


Figure A.2.: Algorithm to construct an initial focusing surface as start for the optimization.

A.2. Algorithm to approximate the optical surface polynomial using elliptical arcs

A major issue which comes along with the use of polynomial surfaces is that the Geant 4 geometry modeler does not support such a surface type. To implement these shapes, one has basically two options. The first one is to implement a new geometry object (*solid*) in Geant 4 which supports the polynomial surface. The second one is to use already available solids to approximate the polynomial $p(x)$. In doing so, the impact on the computational performance has to be minimized as the reflections at these surfaces have a strong influence on the total time needed to complete the full simulation of the detector.

The different options will be discussed briefly in this section and an algorithm for the approximation of the polynomial with elliptical arcs will be presented. This algorithm permits an efficient implementation of the polynomial surfaces in Geant 4 by using elliptical tubes and CSG objects.

The Geant 4 documentation provides detailed information about the interface a new solid implementation would have to provide. Thus, for an experienced developer, it is technically straight forward to add a polynomial shape to the framework. However, a closer look at the interface reveals that in case of a polynomial surface most navigation related functions will have to involve root-finding. Root-finding algorithms are relatively computation intensive iterative algorithms. Such a solid would therefore introduce a severe performance penalty. This is especially true in case of optical physics simulations which involve more particles (the optical photons) as conventional particle transport simulations.

To approximate the polynomial by means of already existing solids, the most trivial approach would be the use of a polygon with fine step-width. However, the step-width has to be in the order of µm to produce a sufficiently precise approximation of the shape. Obviously, the resulting high number of vertices will have a drastic impact on the runtime performance. Especially if the mesh implementation does not utilize space partitioning, as it is the case in Geant 4.

An attempt to implement such a shape using Geant 4's *G4TesselatedSolid* geometry further showed that its implementation behaves numerically unstable in case of such extreme geometries*. Functions which are essential for correct geometry navigation, like the computation of the closest distance of a point to the solids

*Meanwhile the implementation has been updated. This test has not been repeated with the newest version

surface, yield contradictory results due to round-off errors. Thus, photons are wrongly considered to be outside of the volume when being inside and vice versa, leading to physically impossible photon propagation.

Among the remaining solids, only the elliptical tube is of interest. Cylinders cannot be used for approximation as the polynomial has a variable curvature. Elliptical arcs can be constructed by an intersection of a box with the tube using the constructive solid geometry (CSG) feature implemented in Geant 4. The resulting solids can then be used to reassemble the polynomial surface.

In some few cases it is sufficient to fit the polynomial points with an ellipse. An example is the focusing element used in the dSiPM based Disc DIRC design, which is basically a single elliptical arc. In more complicated cases, the attempt to fit segments of the polynomial with an ellipse did not yield good results. The fits did not converge reliably for arc segments which were short in comparison to the circumference of the corresponding ellipse. A more stable algorithm has been developed to overcome these limitations.

This algorithm is based on the "pole and polar" construction of conic sections. The general equation for a conic section is

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0 \tag{A.1}$$

with the six real coefficients A,B,C,D,E,F. This equation has five degrees of freedom, as the equation can be divided by F in case that $F \neq 0$. Any conic section is completely defined by five independent constraints. To assure that a piece-wise elliptical interpolation of the polynomial $p(x)$ between the points $(x_n, x_{n+1}), n = 0 \ldots N-1$ results in a continuous curve, the adjacent ellipses need to have the same slope at the shared endpoints $x_1, \ldots, x_{N-1}$. Therefore the two endpoints and the slope of the polynomial in these points can be used as constraints. In this case, only one degree of freedom remains.

The "pole and polar" construction of a conic, which is frequently used to solve Hough based ellipse finding problems, allows to elegantly pick a quantity for this last degree of freedom. The boundary points and tangents form a triangle as shown in Fig. A.3. A conic can be constructed by choosing any control point P_C inside the triangle. In the following, this point will be on the line segment $\overline{P_A P_{M12}}$, defined by the intersection of the tangent lines P_A (pole) and the midpoint P_{M12} on the segment $\overline{P_1 P_2}$ (polar).

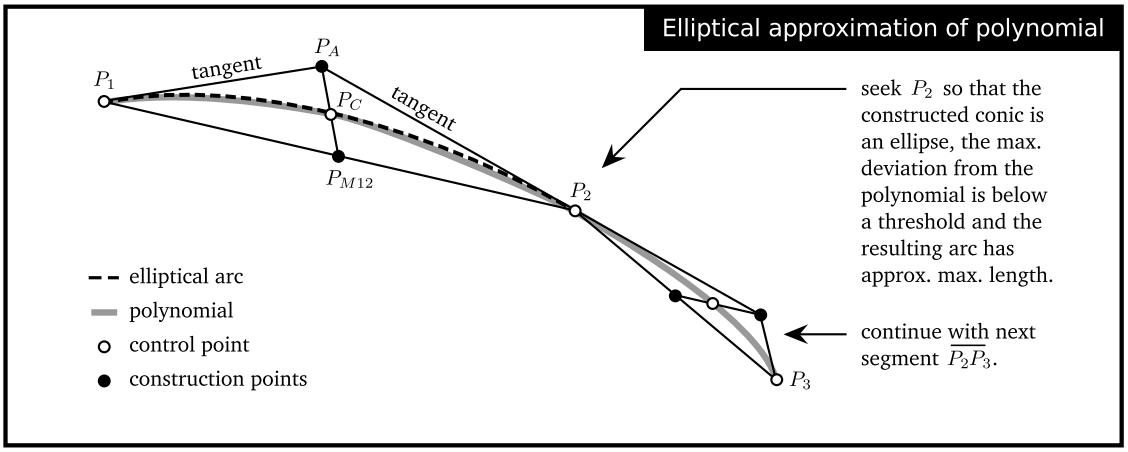


Figure A.3.: Approximation of a polynomial by elliptical arcs. Two points and two tangents on the polynomial define four constraints for the interpolating conic. The missing fifth constraint is provided by a third point on, or close to, the polynomial. P_2 can be varied to find the (approximate) maximum arc length where the maximum deviation from the polynomial is below threshold.

The family of conic sections through P_1 and P_2 is given by [152]:

$$\chi = -\frac{l_1(P_{C,x},P_{C,y})\,l_2(P_{C,x},P_{C,y})}{l_3(P_{C,x},P_{C,y})^2} \tag{A.2}$$

with the three lines:

$$\begin{aligned}
l_1(x,y) &= \overline{P_1P_A} = u_1x + u_2y + u_3 = 0\\
l_2(x,y) &= \overline{P_2P_A} = v_1x + v_2y + v_3 = 0\\
l_3(x,y) &= \overline{P_1P_2} = w_1x + w_2y + w_3 = 0
\end{aligned}$$

and the resulting coefficients of the conic are:

$$\begin{aligned}
A &= \chi u_3^2 + u_1u_2\\
B &= \chi u_3v_3 + {}^1\!/_2\,(u_1v_2 + u_2v_1)\\
C &= \chi v_3^2 + v_1v_2\\
D &= \chi u_3w_3 + {}^1\!/_2\,(u_1w_2 + u_2w_1)\\
E &= \chi v_3w_3 + {}^1\!/_2\,(v_1w_2 + v_2w_1)\\
F &= \chi w_3^2 + w_1w_2
\end{aligned}$$

The remaining degree of freedom is the position of P_C on $\overline{P_A P_{M12}}$. This choice ultimately defines the type of conic. The conic is an ellipse if $4B^2 - 4AC < 0$.

The elliptical arc approximating a polynomial between two points $P_1 = \big(x_n; p(x_n)\big)$ and $P_2 = \big(x_{n+1}; p(x_{n+1})\big)$ can be determined by computing the intersection P_C of the polynomial with the line $\overline{P_A P_{M12}}$. In case that the resulting conic is not an ellipse, P_C is shifted along $\overline{P_A P_{M12}}$ towards the pole P_A until the conic becomes an ellipse. The closest ellipse is then determined by a standard bisection algorithm.

To minimize the number of elliptical arcs per polynomial, the algorithm computes the elliptical arc approximation between the first point $P_1 = \big(x_0; p(x_0)\big)$ and a point $P_2 = \big(x_0 + \Delta x; p(x_0 + \Delta x)\big)$. The average deviation of both curves is computed and compared to a predefined limit. Δx is successively increased until the deviation is higher than this limit. The last valid approximation is selected as solution and its endpoint defines the start for the next elliptical arc. The process continues until the whole polynomial has been approximated.

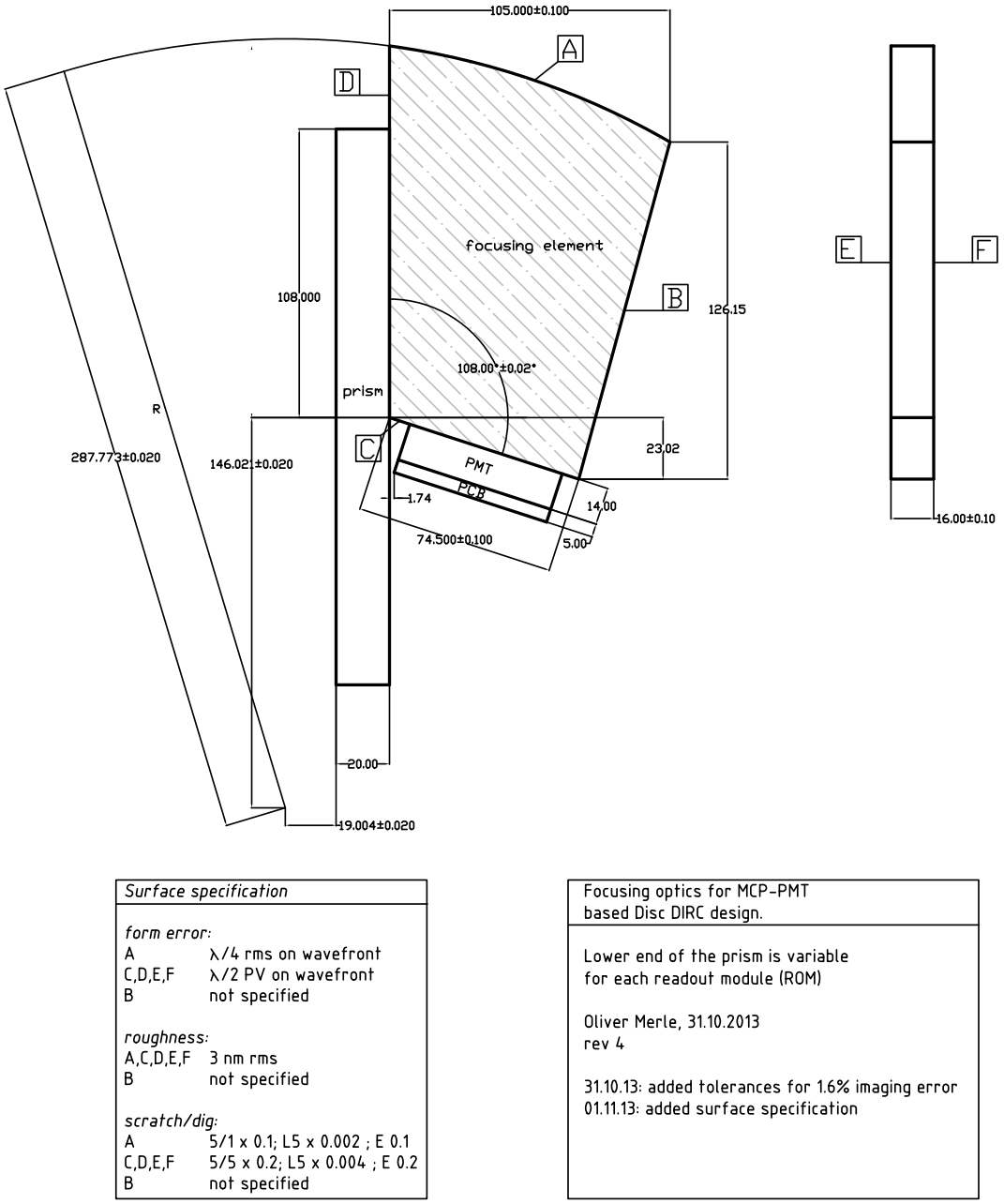


Figure A.4.: Tolerances for imaging optics. Form errors are merely a wish and can be relaxed slightly to adapt to the manufacturers capabilities. Displacement errors on the arc center and radius are strict. Errors correspond to 1.6 % of the ideal imaging error.