Teilchenidentifikation mit RICH Detektoren: Algorithmen und ihre Optimierung

Am Beispiel der Experimente CERES und HERMES

Axel Torsten Maas

Diplomarbeit

am Fachbereich Physik der Technischen
Universität Darmstadt

Gutachter: Professor Dr. P. Braun-Munzinger

25.10.2000

Particle Identification with RICH detectors: Algorithms and their optimization

A case study on the experiments CERES and HERMES

Axel Torsten Maas

GSI Darmstadt, DESY Zeuthen

25.10.2000

Table of Contents

Chapter 1
Introduction

Ring Imaging Čerenkov, abbreviated RICH, detectors are powerful tools for the identification of relativistic and ultrarelativistic particles. The idea to use the Čerenkov effect in combination with imaging was invented in 1960 by A. Roberts and a first working detector was developed by J. Seguinot and T. Ypsilantis in 1977. Today they are used in a wide range of physics applications from middle energy physics to high energy nuclear and high energy physics, as well as astrophysics.

Starting in the 1980s, the first experiment to work with a RICH was E605 at Fermilab, followed by the Ω-RICH at CERN, considered the first large RICH detector. This first generation was followed up by the second generation ones including DELPHI, SLD and CERES. Today the third generation is beginning to operate with BaBar, HERA-B, HADES and HERMES and the fourth generation with ALICE and LHCb will soon be constructed.

This work is concerned with the problems of developing and optimizing strategies for RICH detectors to identify particles from raw data. It deals with general methods and optimization in chapters three and six as well as with two different implementations for specific experiments. Chapter four describes the design and development of algorithms for the CERES experiment, a high energy nuclear physics experiment at the SpS. Chapter five contains a method to optimize the particle identification algorithms of the HERMES experiment, a medium energy particle physics experiment at HERA.

According to the partial review character of this work, chapter two is an introduction into RICH detector basics, and an index and an exhaustive bibliography are included as well.

Although RICH detectors have more purposes than just particle identification, as will be described in chapter two, this work concentrates on this application. It also deals with the problem from a point of view of a perfect detector in many sections, ignoring problems originating for example from misalignment. Common methods to deal with such problems will be referenced in chapter three, but not treated in detail. Data read out and hardware issues are not treated within this work.

Chapter 2
RICH Detectors

2.1 Čerenkov Effect

When a charged particle traverses matter, it loses energy through the excitation of bound electrons, ionizations and other processes, depending on the material. The particle's energy loss for electronic excitations and ionizations is described by the Bethe-Bloch relation. If the particle's speed exceeds the speed of light in this medium, that is

$$v > \frac{c}{n} \tag{2.1}$$

where c is the speed of light in vacuum, n is the refractive index of the traversed material and v is the speed of the particle, it is possible for the particle to radiate electromagnetic waves, the Čerenkov radiation. The medium in which a particle radiates is called radiator. This Čerenkov effect can already be derived on a classical basis, as shown in [89]. However, the amount of energy radiated diverges if treated purely with classical physics. A quantum mechanical treatment ensures convergence; the effect is hence only semiclassical. It is not necessarily limited to relativistic particles, since media are available which have an index of refraction to reduce the speed of light to the order of several centimeters per second. In the semi-acoustics case, a particle moving faster than the speed of sound in a medium develops a Mach cone. An analogous effect occurs for Čerenkov radiation: It is emitted like a shock wave along a cone behind the particle, called the Čerenkov cone. The wave vector $\vec{k}$ and the particle's direction enclose an angle $\theta_{\check{C}}$ with

$$\cos\theta_{\check{C}} = \frac{1}{n\beta} \tag{2.2}$$

where

$$\beta = \frac{v}{c} \tag{2.3}$$

is the Einstein beta [89]. Since photons are directed along the wave vector $\vec{k}$, they are emitted from the particle's trajectory under the Čerenkov angle $\theta_{\check{C}}$. This behavior is another way to see why the Čerenkov effect only sets in above a threshold defined by (2.1). Below it, there is no coherent radiation, $\theta_{\check{C}}$ is not well defined, no cone is created. Since β is bound to be smaller than one[1], there is an asymptotic angle $\theta_{A\check{C}}$, given by

$$\cos\theta_{A\check{C}} = \frac{1}{n} \tag{2.4}$$

From (2.2) the threshold for a particle to radiate can be calculated:

$$\gamma_{th} = \frac{1}{\sqrt{1-\frac{1}{n^2}}} \tag{2.5}$$

where γ_{th} is the Einstein gamma

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \tag{2.6}$$

[1] All charged particles known up to now have mass.

at the threshold. Using (2.4) and (2.5), the asymptotic and threshold behavior can be set in relation and yield

$$\sin^2\theta_{A\check{C}} = \frac{1}{\gamma_{th}^2} \tag{2.7}$$

The spectrum of the Čerenkov radiation during emittance is given by

$$\frac{d^2N_{Ph}}{dE'dx} = \frac{\alpha Z^2}{\hbar c}\sin^2\theta_{\check{C}} \tag{2.8}$$

per unit of energy and distance, where α is the fine structure constant, $\hbar = h/2\pi$ is Planck's constant and Z is the charge of the particle [30].

The wavelength observed is strongly dependent on the opaqueness of the medium for certain wavelengths. The integral average yield of photons N_{Ph} when traversing a medium of length L is given by

$$N_{Ph} = N_0 L Z^2 \sin^2\theta_{\check{C}} \tag{2.9}$$

N_0 is a constant describing the number of radiated photons observed after unit length from a radiating unit charge. This includes medium effects like absorption. It can be calculated as

$$N_0 = \frac{\alpha}{\hbar c}\int_0^\infty T_{TM}dE' \tag{2.10}$$

where T_{TM} is the transmission of the medium for a photon of energy E'. The number of photons from the Čerenkov effect obey Poissonian statistics [49]. The probability to observe N photons is

$$P(N) = N_{Ph}^N \frac{e^{-N_{Ph}}}{N!} \tag{2.11}$$

Important relations are

$$m = \frac{p}{c}\sqrt{1 - n^2\cos^2\theta_{\check{C}}} \tag{2.12}$$

$$p = \frac{mc}{\sqrt{1 - n^2\cos^2\theta_{\check{C}}}} \tag{2.13}$$

$$\theta_{\check{C}} = \cos^{-1}\left(\frac{1}{n}\sqrt{1 - \frac{m^2c^2}{p^2}}\right) \tag{2.14}$$

where p is the momentum of a radiating particle and m is its mass. This results in that knowing the momentum of the particle and the Čerenkov angle identifies the particle up to its charge. The threshold momentum

$$p_{th} = \gamma_{th}\beta_{th}mc \tag{2.15}$$

$$= \frac{mc}{\sqrt{n^2 - 1}} \tag{2.16}$$

is important as well. The resolution with which two particle types can be separated by a detection system at a given common momentum p is measured in standard deviations of the single photon resolution. The single photon resolution describes how well the emission angle from a given track can be measured, and is denoted by σ_γ. The resolution R is then defined by

$$R = \frac{\theta_{1\check{C}} - \theta_{2\check{C}}}{\sigma_\gamma} \tag{2.17}$$

$$= \frac{1}{\sigma_\gamma}\left(\cos^{-1}\frac{\sqrt{1 + \left(\frac{m_1c}{p}\right)^2}}{n} - \cos^{-1}\frac{\sqrt{1 + \left(\frac{m_2c}{p}\right)^2}}{n}\right) \tag{2.18}$$

where m_1 and m_2 are the masses of the different particle types. Equation (2.17) gives theoretical limits attainable for the identification probability, if it can be described as a Gaussian in the Čerenkov angle with half width σ_γ. Another theoretical limit is the maximum separation momentum given by

$$p_{\max} = \sqrt{\frac{m_2^2 - m_1^2}{2k_f n_\sigma}} \tag{2.19}$$

where n_σ gives the number of standard deviations σ_γ by which the species 1 and 2 should be separated. k_f is the RICH detector constant defined as

$$k_f = \frac{\sigma_\gamma \tan\theta_{\check{C}}}{\sqrt{N_{Ph}}} \tag{2.20}$$

By choosing the requested number of standard deviations, the theoretical limit for the maximum separation momentum is given. Note that this is an implicit equation since $\theta_{\check{C}}$ itself depends on the momentum. A good approximation is however achieved for $\theta_{\check{C}} = \theta_{A\check{C}}$.

2.2 Threshold Čerenkov Detectors

The simplest way to utilize the Čerenkov effect is to make use of its threshold behavior (2.1). Particles can be identified by whether they are supposed to radiate or not if their speed or momentum is known. Each species starts to radiate above a threshold given by (2.16). Since (2.16) depends on the mass, and species can be ordered by their mass, a separation band can be defined as

$$\begin{aligned} \Delta p &= p_{th}^{(2)}(m_2) - p_{th}^{(1)}(m_1) && (2.21)\\ &= c\sqrt{\gamma_{th}^2 - 1}\,(m_2 - m_1) && (2.22)\\ &= c\sqrt{\frac{1}{n^2 - 1}}\,(m_2 - m_1) && (2.23) \end{aligned}$$

where m_2 and m_1 are the masses of the heavier and lighter particle respectively. Within this separation band all species with mass m smaller or equal than m_1 can be separated from all species with mass m greater or equal than m_2. If electrons are not separated by other detectors (for example transition radiation detectors (TRD) or electromagnetic calorimeters [30]) the lower mass is the electron mass and the heavier commonly the charged pions' mass or in some cases the muon mass. If leptons can be separated, pions and a heavier hadron can be distinguished in the momentum range Δp. Although limited by equation (2.23), threshold Čerenkov detectors have been used widely, since they provide a relatively easy way of particle identification. Especially if using a stack of detectors with different radiators, particle identification can be provided over a larger band with more than just two species. In some cases only one or two species are of interest. In this case, a threshold Čerenkov detector can be quite useful. If only electrons shall be detected, it is easy to build a hadron-blind detector using a threshold counter. A detector is hadron blind if it only responds to non-hadronic particles[2]. By choosing a high enough γ_{th}, that is, a radiator with low n, a threshold can be imposed that is high enough that due to their low mass only electrons can radiate within the observable momenta.

The Čerenkov radiation is detected by a photo sensitive array, like photomultipliers or photocathodes. This results in a modification of the number of photons, which are observed by modification of equation (2.10) to be

$$N_0 = \frac{\alpha}{\hbar c}\int_0^\infty Q_{QE} T_{TM}\, dE' \tag{2.24}$$

[2] This is a little exaggerated. Since the mass is the important feature of the detector, it is also blind for heavier leptonic flavors.

where Q_{QE} is the quantum efficiency of the detector for a photon of energy E'.

Threshold Čerenkov detectors are still widely used. Some newer experiments for neutrino and high energy astrophysics and astronomy use or plan to use this concept [183]. A reason for this is that water can be used as a radiator, making the detector cheap and safe and allowing the necessary large active detector volume or use under dangerous conditions like in balloons [110, 146].

Čerenkov light emission is also an effect used to measure high energetic, several GeV to more than PeV, cosmic particles. The methods used are neither of the threshold type nor of the RICH type described below, but depend on the Čerenkov light emitted by the shower of charged particles generated by cosmic rays. The largest project of this kind is the Auger project using 1500 identical Čerenkov light detectors [26]. It is important to notice that these other options exist [183].

2.3 Ring Imaging Čerenkov Detectors

To extend the possibilities of the RICH detector as a Particle IDentification, abbreviated PID, detector, it is natural to use the second observable the Čerenkov effect provides: The Čerenkov angle given by equation (2.2). If an arbitrarily precise identification system would be available, PID would be possible for all momenta above the threshold momentum (2.16), limited only by Poissonian statistics and by the Rayleigh scattering of the photons in the radiator resulting in a Gaussian spread in $\theta_{\check{C}}$.

Two main detector concepts are currently widely used to extract the angular information: The proximity-focused ring imaging technique and the mirror-focused ring imaging technique.

Both principles aim at transforming the Čerenkov cone into a geometrical two-dimensional shape which allows the extraction of $\theta_{\check{C}}$. Originally the shape was a ring, hence the name of the detectors, and the radius was uniquely mapped to the Čerenkov angle. Current detectors do not necessarily use rings; ellipses or conic sections are also in use. Rings are most common.

Both concepts use a photon detector to record the Čerenkov photons.

2.3.1 Proximity-Focused RICH

The main problem of extracting $\theta_{\check{C}}$ is that the radiation is a continuous process: While the particle traverses the radiator, it radiates. Hence the Čerenkov cone illuminates a disc with a size only limited by the radiator length L. The idea of the proximity focusing RICH is to reduce this disc to a ring by shortening the radiator. Equivalently, the radiator can be so large that the particle to be observed decays or stops at a distance small compared to L and far away from the photon detector. In these cases, only a ring is detected, as illustrated in figure 2.1. A particle radiates along its trajectory under the Čerenkov angle $\theta_{\check{C}}$ in the radiator. The resulting Čerenkov cone becomes a ring with a finite width w outside the radiator and is detected by the photon detector.

If the radiator's length is L and the distance between the end of the radiator and the photon detector is d, then the radius r at the center of the ring will be

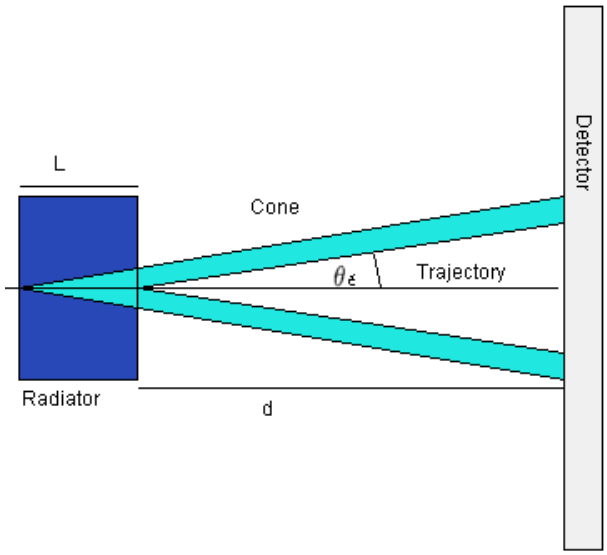
$$r = \left(d + \frac{L}{2}\right) \sin\theta_{\check{C}} \tag{2.25}$$

and its width will be

$$w = L \sin\theta_{\check{C}} \tag{2.26}$$

independent of the distance d. Since Čerenkov radiation is a continuous process, the photons will be spread uniformly over the ring.

The Čerenkov angle can be extracted by measuring the ring width and/or its radius. Usually it is easier to measure the ring radius with satisfactory precision, since the ring width is subject to scattering of photons, refraction and other medium effects. Additionally, the number of photons may be small, between one and ten, making a measurement of the width complicated, especially if the resolution is bad.



2.1. Simple model of a proximity-focused RICH

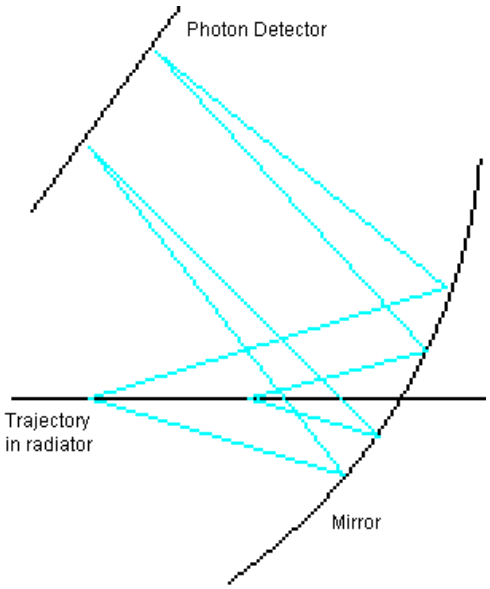
An important property of proximity-focused RICH detectors is that if particles do not move perpendicular to the photon detector, the rings become conical sections [29]. The incident angle can then only be reconstructed by analysis of the properties of the section.

It is important to notice that for a track impacting the radiator under a finite dip angle, the ring is deformed towards an ellipse and the number of photons produced cannot be described by (2.24) anymore. An additional factor of $1/\cos\theta_D$ has to be included, where θ_D is the dip angle [46]. The reason for this is that the traversed radiator length becomes longer by this factor.

2.3.2 Mirror-Focused RICH

It is often not possible to place the photon detector in the particle's trajectory. This might be because of heavy radiation or to reduce the radiation length. There might not be enough space to have a large enough distance to produce rings that are resolvable according to equation (2.25). If there is a high probability of particles incident under an angle, a proximity-focused design may also be not desirable. This is for example the case in forward spectrometers. In this case the radiator and photon detector would have to be build curved, which is not always feasible or possible. Another problem is that in order to have rings of acceptable widths the radiator must be short, but must provide a high photon statistics. If a material with these properties and the required index of refraction can not be found, a proximity-focused detector is ruled out.

In theses cases an alternative design can be used, the mirror-focused RICH. In this case a spherical mirror is placed inside or in front of the radiator. Because of the optical properties of a spherical mirror, the Čerenkov cone is projected upon a ring on the focal surface. This focal surface is a concentric sphere with half the radius of the mirror itself. A photon detector mounted on it will find these rings. Note that the definition of the focal sphere is only valid, if the optical system of the RICH can be approximated by Gaussian optics. This is true for nearly all of the current RICH detectors, so it will be assumed further on. The principle of operation is displayed in figure 2.2. A particle on its trajectory emits a Čerenkov cone, which is reflected by the mirror and detected by the photon detector. Because of the spherical



2.2. The operation principle of a mirror-focused RICH

surface of the mirror, the cone is focused on a ring.

Usually the photon detector surface cannot be curved because of technical reasons. Even if it would be curved, chromatic effects would not allow an optimal position for each possible photon energy. If the bending radius of the mirror is large enough, the focal sphere can be well approximated by a plane. A large mirror with large radius aids in remaining within the limits of the Gaussian optics, which requires that the differences in incident angle for two rays is not too different, and guarantee focusing and prevent astigmatisms [197].

For a given focal length f of the mirror, the radius of a ring r on the focal surface can be calculated as [121]

$$r = f\sqrt{\frac{1}{\gamma_{th}^2} - \frac{1}{\gamma^2}} \tag{2.27}$$

assuming that the origin of the track is in the center of the mirror and the mirror is aligned with the beam axis.

The mirrors used are often only a segment of a sphere, large enough to reflect the Čerenkov cones. However, in most cases the mirror does not provide an acceptance large enough to reflect the complete cone of any particle, and some cones might be missed completely or partially. This introduces a mirror cutoff resulting in incomplete rings in the detector, a problem which has to be dealt with eventually, if the acceptance of the experiment plus the size of the largest cone is not smaller than that of the mirror. An example of this problem will be reviewed in chapter four.

A further modification of equation (2.24) is necessary to include the mirror:

$$N_0 = \frac{\alpha}{\hbar c}\int Q_{QE} T_{TM} M_{MR} dE' \tag{2.28}$$

in which M_{MR} describes the mirror's reflectivity for a photon with energy E'.

Remaining in the focal plane is not necessary in principal, but moving out of it results in rings with a larger width. Since the resolution is finite this does not affect the purpose of the RICH, if the effect is below the single photon resolution limit.

2.3.3 Figure Of Merit

As a figure of merit for the performance of a RICH[3] detector the number of photons per unit length and unit charge particle is widely accepted, which is obtained for an asymptotic particle by

$$\left.\frac{dN_{Ph}}{dL}\right|_{\theta_{\check{C}}=\theta_{A\check{C}},Z=1} = N_0 \sin^2\theta_{A\check{C}} \tag{2.29}$$

from equation (2.9) [71]. An asymptotic particle is defined as a particle having infinite momentum. Within the context of PID, an asymptotic particle is also a particle which has a momentum large enough that the generated ring cannot be distinguished within the resolution from the ring of a particle with infinite momentum.

2.4 Use of Čerenkov detectors

This section describes the purposes for which Čerenkov detectors are used.

Particle detection. The first thing a Čerenkov detector does is to detect the presence of a charged particle satisfying (2.1). This basically simple problem can be quite challenging at high multiplicities, since charged particles crossing the photon detector give rise to a signal. This problem is especially important to proximity-focusing RICH detectors, where all tracks cross the photon detector.

Particle identification. The primary task Čerenkov detectors where invented for was to identify relativistic and ultrarelativistic charged particles. This task is performed either by detection and identification of particles or by identifying particles found by other components of an experiment. Particle identification is the main subject of this work and will be discussed in great detail in the following.

Particle species filter. Čerenkov detectors can be used to filter particle species. This concept was already introduced as hadron blindness: It can be implemented either by the use of an appropriately chosen threshold or by restricting to rings with a specified size, mostly asymptotic rings. These are rings which cannot be distinguished, within the resolution, from a ring generated by a cone emitted under the asymptotic angle. In combination with other detectors, the particle species to be filtered can be a priori any, but common are two options: Hadron blindness by choosing a very large threshold or pion identification by using other detectors like TRDs, electronic calorimeters or muon hodoscopes to previously filter out light leptons [30].

Threshold dependent tracking. RICH detectors are commonly not used for tracking due to the fact that the resolution cannot be made good enough because of scattering effects. In some cases it is interesting or necessary to use them for this purpose nevertheless. This especially applies if there is a large amount of background particles. These can be removed from tracking by requiring a signal in the RICH. One option is to use the RICH as a trigger for tracking. A particle is then tracked if it has been found by the ring it produced. Using more than one RICH, the tracking can principally be done using the RICHes. In the latter case, it is commonly assumed that the track of interest originates from some interaction point, and then two RICH detectors deliver with two rings two points in space which together with the vertex are enough to find the full track. The additional use of a magnetic field between both RICH detectors or between the interaction point and the target makes a momentum determination possible. This has, for example, been realized by the CERES experiment.

Measurement of particle speed. The relevant variable on which the Čerenkov angle depends is β, the speed of the particle. To measure it by measuring the angle is therefore quite interesting. This is especially true in comparison with conventional time of flight (TOF) detectors, because RICH detectors can be built in a much more compact way than those [70]. This concept was demonstrated with the Heavy Ion RICH (HIRICH) at the Gesellschaft für Schwerionenforschung, GSI. It has been shown that a RICH can determine the speed of a particle as well as a TOF detector while being much smaller [70].

[3] If it is not explicitly noted, RICH stands from here on for both proximity- and mirror-focused RICH detectors.

Measurement of particle momentum. As already mentioned, it is feasible to measure particle momentum with a RICH detector. In this case a magnetic field is used, just as in conventional tracking detectors. A mathematical derivation of such an approach for the HERA-B experiment described in the next section is demonstrated in [55].

Another approach requiring high photon statistics makes use of the momentum dependence of the multiple scattering of the particle in the radiator. Like the Čerenkov effect, multiple scattering is a medium effect that influences charged particles traversing matter [89]. It is an elastic scattering process that changes the direction of the particle. This directional change is Gaussian distributed in three dimensions and the root mean square is given by

$$\theta_{MS} = \frac{13.6\ MeV}{\beta cp} Z \sqrt{\frac{x}{X_0}} \left(1 + 0.038 \ln \frac{x}{X_0}\right) \tag{2.30}$$

where x is the distance covered in the radiator and X_0 is the radiation length of the radiator [30]. p must be measured in MeV/c. Such a scattering smears out the ring, having a larger width and a non-homogenous illumination as a consequence. If multiple scattering is the dominant part of all scattering this is called multiple scattering dominance (MSD) and the width of the ring can be used to determine the momentum. This is for example considered by the AQUA experiment [18]. The mathematical derivations can be found in detail in [195].

2.5 Overview of RICH Detectors in Experiments

This section gives an overview of existing, planned, or proposed experiments using RICH detectors. Since RICH detectors are quite common today even for small experiments, it would be beyond the scope of this work to mention all of them. Most of the larger ones will be covered nevertheless[4], giving an interesting spectrum of physics studied with the RICH technology.

The detectors have been separated into two groups: analog and digital detectors. Defined here as analog are those RICH detectors which are able to resolve the number of photons incident in one cell of spatial resolution for one track, while digital ones are not able to do so. A cell of spatial resolution is the area of the photon detector, where two different photon impact points cannot anymore be distinguished. A pixel is defined by the minimum spatial cell area of resolution. Usually a pixel corresponds to a single photon detection unit, like a single photo-multiplier tube, and is the same for all events, but not necessary of the same size and shape at all positions. In some cases the signal of a photon covers multiple pixels.

There is no principal hardware difference between both groups, although some hardware parts are more common for a detector in one of the groups. Besides the different photon detection systems, the way the signals are processed by the electronics is important. A detection system with more than two output states, on and off, is considered analog. Without going into detail on the hardware, typical realizations of the detection systems are given below. A more detailed overview of current technology and developments is given in [166].

2.5.1 Analog Detectors

Typical readout techniques for analog detectors are photocathodes. They use a photo conversion gas to generate electrons from photons, which can then be measured, like, for example, Tetrakis- diMethylAminoEthylen (TMAE) or solids like CsI. This is somewhat similar to a multi-wire proportional chamber (MWPC) and is also referred to as a MWPC readout. The readout itself can be done either by an array of photo cathodes, called a pad readout, or by a time-projection technique, called TPC readout. Pad readouts are faster than the TPC version due to the smaller drift space, but need more readout modules. There are several other possibilities and improvements [166].

[4] Some of the not mentioned RICH experiments are E605 at Fermilab being the first experiment with an operating RICH, UA2 at CERN [101, 187] and SNO [180].

An advantage of analog readout for proximity-focused detectors is the possibility to distinguish between a photon hit and a hit by a charged particle, which crosses the detector [50].

ALICE, A Large Ion Collision Experiment, is an approved experiment at the Large Hadron Collider (LHC) at the european nuclear research center (CERN) in Geneva, and is planned to start operation in 2005/6. It is a heavy ion collision experiment and will have the highest center of mass energy, resulting with $2.4A$ TeV lead on lead to about 1 PeV. The physics under investigation is QCD under extreme conditions and Quark Gluon Plasma (QGP). ALICE will have a proximity-focused RICH detector [47] to identify hadrons between 1 and 5 GeV [51, 60]. A very high occupancy of the photon detector of estimated 12% hit pixels in full central collisions is expected, being a very high value [119], but typical for heavy ion collisions [7]. The readout will be a pad readout with a solid photocathode and the radiator will be liquid C_6F_{14} [50]. The PID will be performed trackwise with an indirect raytracing algorithm, an approach described in the next chapter. The necessary tracking information will be delivered by a TPC.

BRAHMS is an experiment at the Relativistic Heavy Ion Collider (RHIC) at the Brookhaven National Laboratory (BNL) in Brookhaven. It is a two arm spectrometer to measure QCD under extreme conditions in gold on gold collisions at $100A$ GeV. One arm will measure particles at large angles while the other is positioned in forward direction. The particle identification in the forward arm consists of a TOF, a threshold Čerenkov, and a RICH detector to identify particles in the range from 2 to 25 GeV/c momentum. The RICH works with four photomultipliers with 100 pixels each. The radiator is gaseous perfluorbutane, C_4F_{10}. Using a TPC and drift chambers, the PID can rely on the tracking information. The optics is designed to have as few distortions as possible. The RICH will be used for PID up to 9 GeV/c [48].

CAPRICE, Cosmic AntiParticle Ring Imaging Čerenkov Experiment, is a balloon mounted experiment to investigate the cosmic particle flux. It uses a magnetic spectrometer for momentum determination and a proximity focusing RICH detector for PID. The radiator used is solid NaF, the readout is a MWPC readout [29, 101, 183, 187].

CDF II. A considered part of the upgrade of the CDF experiment at the Tevatron collider at the Fermi National Laboratory (Fermilab) at Chicago for the second run of the Tevatron after 1999 was a RICH detector. It would have been used for PID. Technical problems made it difficult to design an appropriate detector, since its thickness would have been limited to 5 to 13 cm. A possible design would have used liquid LiF_2 as radiator and a sprayed CsI pad readout. This would have been a proximity-focused RICH detector [64]. The CDF II experiment researches QCD in $p\overline{p}$-collisions at 2 TeV center of mass energy.

CERES/NA45 [7, 8, 101], ČErenkov Ring Electron Spectrometer, is an experiment at the Super proton Synchrotron (SpS) at CERN. It is a heavy-ion fixed-target experiment, using a gold target in a proton, sulfur or lead beam with up to $160A$ GeV/c. It investigates QCD under extreme conditions with electronic probes. It uses two independent RICH detectors mounted behind each other with a pad readout with 50000 pads each. It will be described in more detail in chapter four.

CLEO III is an upgrade of CLEO II at the Cornell Electron Storage Ring (CESR) at the Cornell University. It studies the CP-violation in the b-system, being a so-called beauty farm. The center of mass (cms) energy is chosen such as to be at the $\Upsilon(4S)$ resonance, to use the b-decay modes of this particle. As part of the upgrade a RICH detector for PID was added. The radiator is solid LiF, the readout is based on a MWPC with gaseous Triethylamine (TEA) as photo conversion medium. The RICH is built proximity focusing [22, 23, 144].

COMPASS/NA58 is a multi-purpose experiment under construction and will use both the μ-beam and the p-beam at the SpS at CERN. It uses a fixed polarized hydrogen or deuterium target. To use them as a solid state target, they are used in form of NH_3 and ^6LiD. The physics will be nucleon spin physics and hadron spectroscopy. When all stages will be completed it will consist of two spectrometers mounted in a line. The first part of the spectrometer will have a large angular acceptance while the second part is limited to smaller angles. Both will be equipped with RICH detectors, the wide angle version to be finished 2000. The radiator will be C_4F_{10} and it will have MWPC readout. It will be used

for π-K separation up to 60 GeV/c [32].

DELPHI is located at the Large Electron Positron collider, LEP, at CERN. The physics investigated in the e^+e^--collisions at up to 209 GeV cms energy are electroweak physics and QCD, ranging from Z^0-physics to the determination of the Cabbibo-Kobayashi-Maskawa (CKM) matrix and higgs search.

The experiment is equipped with three RICH detectors. One is a barrel around the interaction point using a MWPC readout and gaseous C_5F_{12} and liquid C_6F_{14} as radiators. This RICH is called BRICH. The two end cap RICH detectors, called FRICH, are also using a MWPC readout and have similar technical specifications. Their radiators are gaseous C_4F_{10} and liquid C_6F_{14}. All three operate with proximity focusing for the liquid radiators and with mirror focusing for the gaseous radiators.

The reconstruction uses tracking data from the DELPHI TPC and the drift chambers [10, 31, 63, 101, 187].

HADES [196], High Acceptance DiElectron Spectrometer, is an experiment studying QCD under extreme conditions. It is located at the SchwerIonenSynchrotron (SIS) at the GSI. Its main topics in this field are dileptons from ρ and ω decays in nuclear and hadronic matter [68]. It is a fixed-target experiment using the heavy ion beam with $(1-2)\,A$ GeV/c. To measure the expected changes in the mass and widths of these mesons, it is designed for a high mass resolution $\Delta m/m$ below 1%. Since the desired processes have a rate of 10^{-6} per event it is also optimized for high counting rates and good background rejection. Part of this is the hadron-blind RICH detector, working with a C_4F_{10} radiator and a MWPC readout with CsI as photo conversion medium. The RICH is also part of the second level trigger, using a ring search implemented in hardware. The algorithm is a pattern recognition method, as described in chapter three [108].

JETSET [130] is an experiment at the low energy anti proton ring, LEAR, at CERN. It uses a fixed internal gas target for $p\overline{p}$ collisions to study the properties of the $p\overline{p} \rightarrow \phi\,(1020)\,\phi\,(1020) \rightarrow 2K^+2K^-$ chain. It uses a solid fused silica, i.e. quartz, radiator in a proximity-focusing RICH to identify the Kaons and reduce the background by a factor of 10^5. The readout is MWPC based with small pixel size to allow fast operation to achieve high statistics.

LHCb is an approved collider experiment at the LHC at CERN. It is a forward spectrometer, designed for beauty physics. It will have two RICH detectors measuring in different momentum ranges. RICH 1 will have an aerogel and a C_4F_{10} radiator and will operate between 1 and 60 GeV/c. The second RICH detector will cover a smaller acceptance with a CF_4 radiator and measures up to 150 GeV/c. They have different designs because at smaller forward angles the expected event tracks will have a higher momenta [66, 67].

Omega RICH. The Ω-RICH [122] at CERN was used by several experiments [101], WA69, WA82, WA89 [121, 122] and WA94 [1], in different fields of physics research, like photoproduction, hadroproduction and heavy ion and hyperon physics. It was used to identify pions, kaons and protons between 12 and 100 GeV/c. The radiator used was nitrogen or a nitrogen/freon mixture and the readout utilizes a TPC.

SLD [101, 187] at SLAC has studied Z^0 physics. The RICH[5] detector of the experiment is designed similar to the DELPHI RICH detectors, using two radiators but a TPC for readout. It is used to identify hadrons from 1 to 30 GeV/c.

Super-Kamiokande [167,187] in Japan is the successor of the Kamiokande experiment. The physics under examination are atmospheric neutrinos, neutrino astrophysics and proton decay. The detector is one large proximity focusing RICH detector with a liquid water radiator. With 22500 tons active detection medium it is the largest existing RICH detector. The water is also the medium for search for proton decays. In this case the proximity focusing is achieved by the fact that the mean path of the particles is small compared to the detector. The readout consists of 11146 photomultiplier tubes having a diameter of 50 cm.

The detector is large enough to allow for an angular resolution. This is necessary since it is expected

[5] The name used for the RICH is CRID, Čerenkov Ring Imaging Detector

that the neutrino flux exhibits an angular variation. At the same time it is possibly the upper limit for the size of RICH detectors with separated radiator and readout, since for even larger radiators the absorption of the Čerenkov photons would be prohibitive. To use larger volumes requires an integrated readout as e.g. for the Čerenkov threshold counter of the Antarctic Muon And Neutrino Detection Array, AMANDA [183]. However, the reconstruction gets more complicated in such a case.

To allow shielding and separation of incoming and outgoing particles, the RICH itself is surrounded by water with a second outward directed photo-multiplier tube matrix with 20 cm photomultiplier tubes. The surrounding water is used as the radiator for this second matrix.

The used photo-multiplier tubes have a good time resolution and together with the size of the detector allow the separation of single photons through analog detection. A second benefit of the size is the possibility of vertex reconstruction by a time-of-flight measurement for the photons. These vertices are then connected to rings, which are found by a spatial Hough transformation. This transformation is described in chapter three. After PID and measurement of the Čerenkov angle by counting the number of photons, the momentum can be determined by equation (2.13). The PID is performed utilizing the fact that of the two observed species, muons and electrons, the electrons have badly distorted rings while the muons have well defined rings. The muons are the signal particles for which the momentum measurements are performed [167].

2.5.2 Digital Detectors

Digital detectors commonly use Photo-Multiplier Tube (PMT) readout. In this case each pixel of the detector measures only the presence or absence of at least one photon hit per event.

AQUA is a proposed experiment for neutrino astrophysics and astronomy and proton decay research. First studies regarding the technical principles have been completed [18]. As for Super-Kamiokande a single RICH detector would be used for PID, vertex reconstruction and momentum measurement. The latter would be achieved by measuring the changes in the single photon resolution by multiple scattering. Water would be the radiator and active medium. The readout would consist of PMTs.

BaBar [5, 40, 144] is a collider experiment to examine the CP violation in the b sector of the standard model. This experiment, also investigating b-, c-, τ- and 2γ-physics, is located at the Stanford Linear Accelerator Center (SLAC) at the PEP II B-factory. This collider operates asymmetrically with an $e^- e^+$ beam on the $\Upsilon\,(4S)$ resonance.

The RICH detector DIRC [46, 111], Detection of Internally Reflected Čerenkov Light, has an unusual geometry. The Čerenkov light produced during an event is guided by light-guides outside of the detector and registered by a PMT matrix at the endcaps. This matrix consists of about 11000 PMTs. Because of the geometry, no rings are produced, but two conical sections[192]. The radiator used is solid quartz. The DIRC is used for separation of pions and kaons between 0.7 GeV/c and 4 GeV/c and operates since October 1999.

Gran Sasso. The principle of measuring momenta by multiple scattering is used for a proposed long baseline RICH experiment in Gran Sasso using neutrinos produced by CERN in a distance of 731 km [195]. In this case a mirror-focused system would be used. The unified radiator and active medium should consist of 10^6 tons of water. The readout is designed to use photomultipliers.

HERA-B [102, 105, 144] studies b-physics [182]. It operates on the 920 GeV/c proton beam of the Hadronen Elektronen Ring Anlage (HERA) accelerator at the Deutsches Elektronen SYnchrotron (DESY) in Hamburg. The fixed target consisting of copper wires is positioned in the beam halo. The radiator is C_4F_{10}, the readout consists of a PMT matrix built from PMTs of two different sizes. The area of the first type is four times as large as that of the second type. The first type is used at the edge of the matrix, while the second one makes up the center of the detection area. The PMTs are using multi-cathodes.

Remarkable is the fact that the Čerenkov light is reflected by two mirrors, first by a spherical one and then by a planar one. The detection plane hence faces the direction of the beam. The reason for this is two-fold: First to reduce radiation damage for the readout because of the high frequency of collisions

and high multiplicity of fast charged particles, and second to reduce background of the readout to an acceptable level. The RICH system was originally only intended to separate pions and kaons from 5 GeV/c to 50 GeV/c. Currently it is also used for momentum measurement, since the tracking system has turned out to be too sensitive against radiation damage and undergoes replacement [57].

HERMES, HERa MEasurement of Spin, is an experiment investigating nucleon spin physics and nuclear physics. It uses the 27.56 GeV/c electron/positron beam of the HERA accelerator at DESY with an internal gas target. The experiment is separated in a top and a bottom half, both having a RICH detector of identical design. The detectors are dual-radiator RICHes, i.e. they use the same mirror and readout for two radiators. These radiators are gaseous C_4F_{10} and an amorphous silica-aerogel. It is the first RICH that uses aerogel as a radiator material. The readout consists of 1934 PMTs each. More details about this experiment will be given in chapter five.

IMB [183,187] in Ohio, U.S.A., is an experiment to study the existence of proton decay and neutrino physics, the latter especially from atmospheric pion decays. Its RICH detector uses water as a radiator, which is at the same time the active medium of the experiment. The readout consists of PMTs.

Kamiokande [183,187] in Japan was an experiment to study atmospheric and cosmic neutrinos and general neutrino astrophysics. The radiator was 4500 t of water and the readout used about 1000 PMTs. This experiment has been replaced by Super-Kamiokande.

NA35. The NA35 experiment at the CERN SpS investigates QCD under extreme conditions in nucleus-nucleus collisions up to $200A$ GeV/c [28]. The RICH detector is built proximity focusing using a liquid C_6F_{14} radiator. The readout is a two step design. In a first step the photo conversion gas TMAE produces an electron avalanche which then produces light. This light is then measured in a second step by a charged-coupled device (CCD) camera. The RICH is used for particle identification from 1.2 to 3.8 GeV/c. The streamer chamber of the experiment provides the tracks used for PID [27].

PHENIX [19] is an experiment at RHIC. The main task is the study of QCD in extreme conditions, mainly of the quark gluon plasma. Dileptons are one of the main signals in this context, and nearly hadron-blind RICH detectors will be used to identify them. The radiators could be either ethane or CO_2, having as a consequence different degrees of hadron blindness. The readout is a PMT matrix of 5120 PMTs.

SELEX/E781 is an experiment on charm physics at Fermilab. For particle identification up to 320 GeV/c it uses a RICH detector with a readout consisting of 2848 PMTs. The experiment achieves an efficiency of 90% over the whole momentum range of interest [62].

SPHINX [103] studies diffractive production processes and searches for pentaquarks, color neural configurations of five quarks, $qqqq\overline{q}$. It is located at the Serpukhov accelerator. The RICH detector uses as radiator SF_6 and the readout consists of photo tubes.

SQUASH, Strangelet, QUark and Antinuclei Search at High Sensitivity [127], is a proposed experiment at the CERN SpS. The main topic is the search for exotic particles like strangelets or free quarks produced in heavy ion collisions with a $160A$ GeV/c lead beam. A magnetic separator would exclude particles with charges of e or higher so only quarks or stranglets would enter a RICH detector which should be used to identify these particles. A gaseous C_4F_{10} radiator would be used with a PMT readout. To deal with acceptance effects, a flat mirror on the beam side wall would be added to reflect Čerenkov light which would be absorbed otherwise. The reflected part of the cone generates a part of the ring which has inverse curvature as the rest of the ring, but is connected continuously to the rest of the ring.

Chapter 3
Particle Identification Algorithms And Optimizations

This chapter presents the main principles for algorithms for particle identification with RICH detectors. Each RICH detector and sometimes different physics topics to be studied with the same RICH detector require different approaches, therefore it is not possible here to describe each algorithm in a general way in detail. However, the main thoughts and foundations are presented. In the second part of this chapter possible optimizations to cope with specific problems will be discussed. Examples for CERES and HERMES of how such algorithms and their optimizations are structured in detail will be covered in chapter four and five. The judgement of the performance of an algorithm is crucial and the last section of this chapter will be devoted to the question of how to qualify and quantify the performance of an algorithm.

3.1 Observables

It is important to consider which are the observables accessible by a RICH detector. RICH detectors, in principle, produce a geometric two-dimensional output, the response of the photon detector plane. This output is a lattice, not necessarily a regular one, where each lattice node has an amplitude. The other information is the output range of these nodes. This lattice constitutes the detector information. To gain information about the type of the particles measured, it is necessary to transform this lattice.

But already at this point a general subdivision of the approaches must be made. This splitting is created because a varying amount of additional information from other detector parts is available. There may be no other information, if the RICH detector is the only detector, or up to complete tracks with momentum information and perhaps partial PID information.

This divides the RICH reconstruction methods already into two categories, the track-dependent and track-independent algorithms.

Track-dependent algorithms use external tracking information to generate predictors e.g. for the estimated centers of rings or focal points of conic sections. Then the local information of the lattice is processed. Because these predictors are known, a natural cutoff of the lattice is given by the asymptotic radius.

Track-independent algorithms have to survey the complete lattice to find geometric structures of the expected shape. Recognized shapes are then identified with particles and the properties of these shapes are used for particle identification.

In both cases the properties of the geometric objects are used for particle identification. Three properties are available from these objects. The first is the number of photon hits registered, the second one is a property, e.g. the radius of a ring, which can be mapped uniquely to the Čerenkov angle $\theta_{\check{C}}$. In principal an angle is found for each photon hit. These are the first two observables which can be found. However, because of (2.9) both are basically equivalent. Nevertheless each one is important on its own. While the Čerenkov angle is used to identify the particle, the number of registered photons is used to separate a true signal from background. The latter part is commonly performed by requiring a

minimum number of registered photon hits, the background threshold N_{BT}. This value might depend on various quantities like Čerenkov angle, momentum, spatial position etc. or it might be a constant. This effectively makes use of the RICH detector as a threshold Čerenkov detector. The third observable is the degree to which the geometric object agrees with the expected shape. In the case of a ring this would be e.g. how uniformly the photons are spread about the circumference. This does not directly provide information about the particle type, but it can be important to eliminate incorrect identifications. If for example noise is generated as a single large noise cluster, this might lead to a misidentification. The reason for this is that if the cluster lies on the expected circumference of a ring, it provides enough hits at the appropriate Čerenkov angle. Problems of this kind can be treated by using geometrical information. This is equivalent to requiring a uniform distribution of Čerenkov photons in the azimuthal emission angle. For convenience this information will be called "ringness", although there might be various degrees of extraction, the geometric shape may not be a ring, or the usefulness may be limited by the Poissonian statistics of the emission.

The extraction of these three types of information is the first of two parts of the PID process. The second part is to map these observables on the particle type of a track. A strict separation of these two parts is generally not the case, but they are in principle separated. In the simplest case the second part can be completed by measuring the momentum and applying equation (2.12). This yields directly the mass of the particle. Equally simple in hadron-blind detectors is the restriction to a specific range of observables, commonly to asymptotic values of $\theta_{\check{C}}$. It is then sufficient for the PID to find an asymptotic ring. In such cases the measured number of photons N and the ringness r_{rn} are used to validate the PID.

Other methods rely on testing all possible particle hypotheses against the lattice to select the hypothesis which yields the best result. A particle or mass hypothesis in this sense uses a given track with momentum p and assumes that it would be of type t with mass m. From this the expected Čerenkov angle $\theta_{\check{C}}^{\text{exp}}$ can be calculated by equation (2.14) and the expected number of photons N_{Ph} by (2.9). By comparing with the measured values, a likelihood for this hypothesis can be defined. By repeating this for all possible particle types, and comparing the likelihoods, the most probable particle type can be found. The details of the definition of an appropriate likelihood can be quite different for different applications and possibly depend on information like momentum or spatial position. A simple version can be constructed by

$$\ell\left(N, \theta_{\check{C}}\right) = N_{Ph}^{N} \frac{e^{-N_{Ph}}}{N!} \exp\left(-\frac{\left(\theta_{\check{C}}^{\text{exp}} - \langle\theta_{\check{C}}\rangle\right)^2}{2\sigma_\gamma^2}\right) \tag{3.1}$$

where σ_γ is the single photon resolution of the photon detector and $\langle\theta_{\check{C}}\rangle$ is the average measured Čerenkov angle. The latter one enters this function as a Gaussian. The reason for this is that for the true particle type, the photons will be distributed about the true Čerenkov angle as a Gaussian, since effects like Rayleigh scattering impose a Gaussian variation in the polar angle. Rayleigh scattering is the coherent scattering of a photon by a whole atom. Because this process is elastic, the probability of detection is not changed as long as the photon remains within the acceptance. The probability of scattering scales with Z_R^2, the atomic number of the radiator. Hence radiators with low Z_R are preferable [197].

Note that this likelihood is not normalized, since it is no probability. Note also that this is a single track likelihood, i.e. it ignores any other tracks in the same event. An event likelihood can be constructed to simplest order by a product of all track likelihoods. Maximization of the likelihood corresponds to finding the true particle type. Hypotheses which predict no radiation will be assigned a sub-threshold value for the likelihood. If this is the highest likelihood, the particle will be assumed as not radiating. If two or more possible hypotheses are below threshold, and these are the highest likelihoods, no identification is possible. If any two highest likelihoods have the same value, no identification is possible as well.

From these likelihoods it is possible to derive a value to describe the confidence in the result by taking the ratio of the highest and the second highest likelihood. This value can be shown to be the best possible cut parameter to achieve optimum efficiency in finding particles of species t in a sample at the same level

of not misidentifying other particles as being of type t. The possibilities of a likelihood-based analysis and the last assertion are treated in more detail in [37, 87, 165]. Some examples of application can be found in [5, 49, 96, 118, 121].

In the following two sections some methods to either extract the observables and/or identify the particles in more detail will be discussed. The next section will deal with track-dependent methods while the one after this will deal with track-independent methods.

3.2 Track-Dependent Algorithms

If track information is available, there are two possibilities to treat an event. The first one is to use only the information local to the track, that is to investigate only an area in which the Čerenkov photons could have been emitted by a single track and to ignore the rest of the data. The second approach does also make use of nearby or all tracks and tries to find a configuration of particle types which results in optimal agreement with the measured data.

In the following it is assumed that if tracks are available, the momentum is available as well.

3.2.1 Direct Raytracing

A relatively straight forward method is Direct RayTracing (DRT). This method uses Monte Carlo methods to generate the expected measurement for a given set of tracks and particle hypotheses and then compares it to a given measurement. To speed up the process, the Monte Carlo can be coarse, as long as the uncertainties remain of the order of the pixel resolution.

Each track is treated separately. For a given track a large number of Čerenkov photons is generated with emission vertices distributed regularly in azimuthal angle and along the trajectory in the radiator. The number of photons has to be larger than the expected photon yield by more than an order of magnitude. Each single photon is then tracked to the photon detector and its impact position is mapped on a pixel. The tracking is dependent on the geometry of the detector. An example for the mathematical methods can be found in [49]. As a result of the large number it is possible to find a true hit probability P_i^k for each single pixel i of the detector to register k photons. A likelihood can then be defined by

$$\ell = \sum_{\text{all pixels } i} \ln\left(\sum_{k=0}^{\infty} \delta_{kN_i} P_i^k\right) \tag{3.2}$$

where the sum is over all pixels and N_i is the number of photons registered in the pixel. δ_{kN_i} is the Kronecker δ-function. It is 1 for $k = N_i$ and 0 else. Instead of ∞ an appropriate cutoff is selected in practice. This likelihood can be calculated for each particle hypothesis by reiterating the procedure for each hypothesis. If a hypothesis is below threshold, no photons are generated, but the likelihood is calculated anyway to find the sub-threshold likelihood. As described above, the comparison of the likelihoods delivers the particle type.

3.2.2 Indirect Raytracing

An inversion of the DRT is the Indirect RayTracing (IRT). For each photon hit which could have been emitted by the track, i.e. which is inside the maximum radius of the ring[6] for the lightest particle under consideration, plus some additional width due to resolution and scattering, the Čerenkov angle is calculated. This determination is done for each possible mass hypothesis for each track. The determination is performed by back-tracking from the position of the pixel to an estimated vertex on the particle's trajectory. The estimate is typically the middle of the trajectory inside the radiator, because there is no

[6] Although the presented methods would also work with conic sections etc., from now on the geometric object will be denoted as a ring.

possibility to determine the true emission point, and using the middle any effect of this approximation on the Čerenkov angle should vanish on average. An artifact is that the distribution of the angle becomes wider. If absorption is significant, the emission point may be nearer to the end of the radiator; this can be calculated and taken into account to find the true average emission point [47]. The back-tracking itself depends on the geometry of the detector. Examples can be found in [49, 66, 157]. A likelihood can then be normally constructed by

$$\ell = \prod_{\text{all hit pixels } i} \exp\left(-N_i \frac{\left(\theta^i_{\check{C}} - \theta^{\text{exp}}_{\check{C}}\right)^2}{2\sigma^2_\gamma}\right) \tag{3.3}$$

for each mass hypothesis. Additionally windows can be imposed around $\theta^{\text{exp}}_{\check{C}}$ to measure the number of photons and to use this information for rejection of background. Alternatively the sum of the contributions can be used, or by taking the logarithm of (3.3) the quadratic average angle

$$-\ln \ell = \sum_i N_i \frac{\left(\theta^2_{\check{C}} - \theta^{\text{exp}}_{\check{C}}\right)^2}{2\sigma^2_\gamma} \tag{3.4}$$

can be determined. In the latter case the smallest value of $-\ln \ell$ delivers the particle type.

3.2.3 Reconfiguration

A method to extend IRT and DRT to include event information is the reconfiguration or maximization of likelihoods. This method assigns in a first stage each particle an estimated type, either by assigning each track simply the most common type or by IRT, DRT or other methods. Then the particle type for each track is varied and the variations are kept, if they increase the event likelihood [2, 66, 67] which can be determined in the simplest case by multiplying all individual likelihoods. This process is iterated until a maximum has been found. Random variations can be included to avoid local extrema.

A possible modification to improve the behavior for multiple overlapping rings is to assign the hits to specific rings or split them after each step between rings in the neighborhood to recalculate the likelihoods. The splitting can be an equal splitting or a more sophisticated method. Such an approach is also necessary if the initial particle estimation and the one used in each iteration is the same, otherwise the likelihood will not change. Commonly this approach is used in high multiplicity environments, and the initial method is to assign each track the most common particle type. The true configuration is then searched, for example by variation and calculation of (3.3).

3.2.4 Yield Determination

A straightforward method to identify the particle type is the yield determination [44, 49]. The principle is to calculate for each mass hypothesis the appropriate Čerenkov angle and then count the number of photons found inside a ring of a width determined by the geometry and resolution of the photon detector to find a large enough fraction of the photons. The mass hypothesis yielding the number of photons nearest to the number expected by (2.9) will be chosen. Alternatively a likelihood can again be determined by (2.11). The resolution is limited by the fact that the expected number of photons must be different by at least one. This gives a resolution in the Čerenkov angle of

$$\begin{aligned} 1 &= \Delta N_{Ph} = N_0 L Z^2 \left(\sin\theta_{\check{C}} - \sin\theta'_{\check{C}}\right)\left(\sin\theta_{\check{C}} + \sin\theta'_{\check{C}}\right) && (3.5)\\ &\approx 2N_0 L Z^2 \theta_{\check{C}} \Delta\theta_{\check{C}} && (3.6)\\ \Delta\theta_{\check{C}} &\approx \frac{1}{2N_0 L Z^2 \theta_{\check{C}}} && (3.7) \end{aligned}$$

where $\theta_{\check{C}}$ is the expected angle for one particle type and $\theta'_{\check{C}}$ is the hypothesis for the particle type next to it in mass. It is assumed that $\theta_{\check{C}}$ is small, as it is normally the case for a RICH used in particle physics[7]. This gives for values of $\theta_{\check{C}}$ near the asymptotic value a resolution of

$$\Delta\theta_{\check{C}} \approx \frac{1}{2N_0LZ^2\sqrt{1-\frac{1}{n}}} \tag{3.8}$$

$$\frac{\Delta\theta_{\check{C}}}{\theta_{A\check{C}}} \approx \frac{1}{2N_0LZ^2\left(1-\frac{1}{n}\right)} \tag{3.9}$$

where Z^2 of the particle to be identified is given by physics under investigation, L is fixed by the available space and n by the chosen radiator, only N_0 is left to increase the angular resolution. Additionally, N_{Ph} is not a fixed number. The possibility for statistical fluctuations of the order of 1 is given by

$$P^{\Delta N=\pm 1}(N) = P(N+1) + P(N-1) = \frac{e^{-N_{Ph}}}{N!}N_{Ph}^{N}\left(\frac{N_{Ph}}{N+1}+\frac{N}{N_{Ph}}\right) \tag{3.10}$$

For the average number N_{Ph} this results in

$$P^{\Delta N=\pm 1}(N_{Ph}) = \frac{e^{-N_{Ph}}}{N_{Ph}!}N_{Ph}^{N_{Ph}}\left(1+\frac{N_{Ph}}{N_{Ph}+1}\right) \tag{3.11}$$

which is a function that slowly decreases with N_{Ph}. For a given value of $\Delta\theta_{\check{C}}$ this fluctuation limits the maximum available efficiency. Therefore, a high yield of photons is essentially for this method. Furthermore the use of this method can be complicated by acceptance effects, e.g. due to the finite size of a mirror, and other effects that make the yield strongly position dependent, e.g. a different track length in the radiator.

3.2.5 Pattern Comparison

Another simple method is pattern comparison [49,108]. The method relies on a database of expected hit patterns, which have been generated by Monte Carlo and describes for a given set of particle parameters, like momentum and spatial position, the expected probabilities for a pixel to give a signal. Since no continuous database can be generated, the pattern is binned in the parameters. For each particle and each mass hypothesis the actual measured pixel pattern is compared with the one expected in this bin. The particle type that provides the highest agreement with the expected pattern is chosen. This method is in principle a static version of DRT, resulting in a more coarse examination. However, if the expected particle types are asymptotic, like in a hadron-blind RICH detector, and the detector provides translational invariance, the database can be small and the comparison is fast, providing the possibility to use the RICH as a trigger detector [108].

3.2.6 Two-dimensional PID

The methods presented thus far do not incorporate the ringness as defined in the beginning of this chapter, but only the number of photons and the Čerenkov angle $\theta_{\check{C}}$. The azimuthal angle ϕ is not used. By using optical methods, deploying the method of Lagrange multipliers, it is possible to include this information and model the complete trajectory of each photon, which imposes additional constraints on a hit to belong to a given track. This is effectively an inclusion of the requirement that the found ring is indeed a ring. A detailed treatment of this method with respect to the DELPHI experiment is given by [131] and shows that this would work in principle.

[7] This is not true for some of the larger astrophysical RICH detectors.

3.2.7 Angular Hough Transformation

The principle of a Hough transformation is to map an image to a feature space, in which characteristic properties of objects show a strong peaking. This method from visual pattern recognition can be used for RICH PID [47, 60]. In this case the most prominent feature is the Čerenkov angle. If the track is known, an angular Hough transformation is best suited, while a spatial one can be used if the track is not known, as it will be described in the next section.

The angular Hough transformation proceeds as follows: For each track a histogram for the Čerenkov angle is generated, the size of the bins are of the order of the single photon resolution. For each hit the Čerenkov angle is calculated, for example using the same technique as for IRT, and each hit is then entered in the appropriate bin in the histogram. At the true Čerenkov angle a peak should appear. By finding this peak, either by finding the maximally populated bin or by a center-of-gravity or a Gaussian fit[8] the Čerenkov angle can be measured and therefore the particle type be determined. Particles below threshold should exhibit no clear peak and can be identified by this. Because of noise it is likely that they will have a finite number of entries. The highest values for sub-threshold particles are expected to be found at the highest bin in Čerenkov angle. Thus this bin in the histogram should not conincide with the asymptotic Čerenkov angle, but be located at a higher value. A center-of-gravity or maximum search will then find the peak at an entry corresponding to a Čerenkov angle larger than the asymptotic one, a signature for a sub-threshold particle. The expectation of an increase towards larger angles can be justified, if the assumption of a spatial isotropic background is made. The area of a ring scales as

$$A = \pi\left(\left(r_{\check{C}} + \frac{w}{2}\right)^2 - \left(r_{\check{C}} - \frac{w}{2}\right)^2\right) \tag{3.12}$$

$$= 2\pi\frac{w}{2}r_{\check{C}} \tag{3.13}$$

i.e. linear with the Čerenkov radius of the ring[9]. Since the noise density is uniform, the number of noise hits also scales linear with the Čerenkov radius. In a well-behaved geometry, the radius and the Čerenkov angle are connected by

$$\tan\theta_{\check{C}} = \frac{2r_{\check{C}}}{L} \tag{3.14}$$

and because of this the number of noise hits N_{BG} scales as

$$N_{BG} = \rho\pi\sigma_\gamma L\tan\theta_{\check{C}} \tag{3.15}$$

where ρ is the number of noise hits per unit area. Normally $\theta_{\check{C}}$ is small

$$N_{BG} \approx \rho\pi\sigma_\gamma L\theta_{\check{C}} \tag{3.16}$$

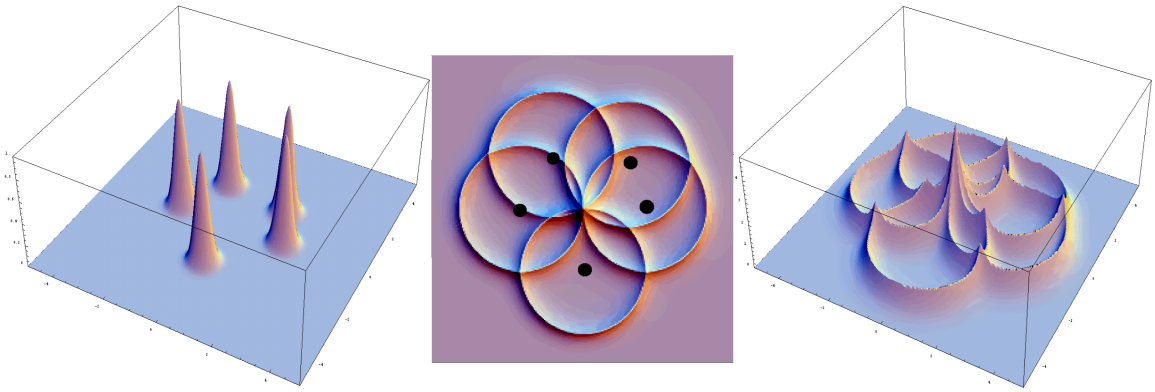
and hence the increase at large angles. Depending on the size of ρ, the number of bins may not be too large, to avoid that noise peaks at the edge of the histogram are larger than the ones expected for a particle. This is important, if the noise scales more than linear with $\theta_{\check{C}}$, which happens for example in the case of high ring and track multiplicity environments [47].

3.3 Track-Independent Algorithms

The principal task of a track-independent approach is the same as for visual image recognition: The identification of geometric objects which are obscured by other parts of the picture. The situation for RICH detectors is in some sense more complicated, since the shapes are normally fragmented and large

[8] A Gaussian shape is expected, since the Čerenkov photons should be smeared out by scattering in polar angle in a Gaussian way.

[9] In cases of shapes which are not rings, similar calculations can be made.



3.3. The spatial Hough transformation. The left panel shows five registered hits, the right panel shows the transformed signal. The center panel contains an overlay of both.

background is present in combination with low resolution. On the other hand, the shape is known in advance, and it is geometrically simple.

However, track-independent methods typically have to deal with an unknown radius in the case that non-asymptotic rings should be treated as well, making it more difficult to identify circles. Additionally, the spatial resolution of a RICH is in general inferior to that of a dedicated tracking device, making determination of ring centers difficult. Nevertheless some sophisticated algorithms have been developed and applied with considerable success.

3.3.1 Spatial Hough Transformation

The spatial Hough transformation, also called dual space technique [42, 163], is a simple algorithm to find ring centers directly in raw data. It is based on the geometrical properties of the ring to have a circular shape of a given radius. In a perfect detector all photon hits of an asymptotic particle would be positioned on a perfect (circular) Čerenkov ring. If a circle is drawn around each hit of asymptotic radius, all circles will intersect at the center of the Čerenkov ring. If compared to all other intersection points, this center will have the highest number of intersecting rings. An example is shown in figure 3.3. The prominent feature used in this case is the fact that the signal consists of rings with a defined ring center, in comparison to the Čerenkov angle of the angular Hough transformation.

In general there is not the signal of only a single particle, but of several particles plus background. The procedure then is the same, around all hits a ring will be drawn, and the number of rings intersecting in a pixel will be counted in a two-dimensional histogram, also called the dual space. All peaks above a certain threshold in such a histogram will then be associated with an asymptotic particle.

To find particles which are not asymptotic, it is necessary to vary the radius of the circles placed around the hits according to the expected Čerenkov radius. A pronounced peak, called Hough amplitude, will only then evolve if the assumed Čerenkov radius and the position is the one of a true particle of the corresponding radius.

The disadvantage of this method, and in general of all track-independent methods, is the susceptibility to background. A high level of background hits generates a certain probability to create a fake ring, that is to generate a peak, although no Čerenkov ring of a particle is present. This problem gets even more severe when non-asymptotic particles are searched, since in this case the number of expected hits decreases and the peak gets smaller and therefore the background fake probability increases.

If the background is isotropic with an area density of ρ, the limiting density of this method can be

found with (3.12) by

$$\rho_{\text{lim}} = \frac{N_{cut}}{2\pi\sigma_\gamma r_{\check{C}}} \tag{3.17}$$

where N_{cut} is the number of rings around hits required to intersect to let a peak be identified as a particle. The limiting number of background hits is then given by

$$N_{\text{lim}} = \frac{N_{cut} A_D}{2\pi\sigma_\gamma r_{\check{C}}} \tag{3.18}$$

where A_D is the total area of the detector. This is the limiting number to find at least one fake particle per event. This number is decreased in situation were correlations are present. These correlations for example can be overlapping rings. The number is also dependent on the resolution.

This method is only the first step to find ring candidates. In a second step the hits which seem to make up a ring are used to perform a fit of the possible ring center. The simplest method to do so is a center-of-gravity fit, but more sophisticated methods are available [7].

This method does not make use of ringness; if a large amount of hits is concentrated at a small segment of the ring, also the correct ring center will be found - and hence a concentration of background hits will be able to generate a fake ring. The only signal of this would be a high level of signals in the dual space, but a separation will be complicated.

3.3.2 Elastic Net

Another approach to ring search is the elastic net, based on a principle described in [58]. Originally this method was developed to provide an approximative solution to the class of NP complete problems. NP complete problems are combinatorially solvable, but scale for a characteristic size M of the problem stronger than any power of M. The most popular version of such a problem is the traveling salesman. In this case the shortest way to visit M cities in a plane without visiting any city twice should be found. The combinatorial solution would be to test any possible path. Since this scales as described, it is not practical, and there are some approaches to find an approximate solution, one of them being the elastic net. Originally the development of the net was aimed at neurobiology, but it can be used for the problem of track-independent ring recognition for RICH detectors as well.

The algorithm starts with a closed geometric shape positioned in the plane. This shape would be a ring in case of a RICH. It is subject to two forces. The first is an attractive force originating from each city or hit. The second is a tension in the ring to keep its size small. In the case of an application to the RICH a third force would be necessary to ensure that the shape of the ring is conserved or only changed up to an extent necessary to respect optical deformations of rings in the RICH. The evolution of the ring under theses forces is then simulated in an iterative procedure. During the iteration, the attractive forces decrease in range, so that the cities or hits next to the ring start to dominate with time. The result is a ring matched to the hits creating a ring. This procedure corresponds to a constant minimization of the energy functional given by the potential of the different forces.

An important requirement is that no crossovers may occur, since this changes the topology of the net and would prevent a minimization. The forces have to be chosen in a way to respect this requirement. This is also important for the RICH application where the topology of the ring must also not change.

An implementation with respect to the traveling salesman problem can be found in [58]. For RICH PID, a proof of principle has been conducted with good results [72,73]. In this case the forces to maintain the ring shape where implemented as to stabilize the Fourier transform of the ring up to a few orders.

3.3.3 Statistical Search

If the number of expected background hits inside an area $\pi r_{\check{C}}^2$ is much smaller than the number of hits

per ring

$$\frac{N_{BG}}{\pi r_{\check{C}}^2} \ll \frac{N_{Ph}}{\pi r_{\check{C}}^2} \tag{3.19}$$

a simple statistical search can be used. In this case a window of the geometric shape of the expected ring with a radius slightly larger than the asymptotic radius is moved pixel by pixel over the detector area. If inside the window a threshold number of hits is found, and this number only decreases for any variation of the position, a ring is found, which then can be analyzed further by fits like a center of gravity fit to find the ring center. This method is fast but only applicable if (3.19) is fulfilled and the particles to be found are well above threshold. Since it is fast, this method would be for example usable as a trigger, in suitably clean environments, like a hadron-blind detector with low background level.

3.3.4 Cluster Algorithm

The cluster algorithm has not yet been described in literature and was an experimental method investigated with data as part of this work at the HERMES experiment.

The algorithm assumes as preconditions that the rings are only weakly fragmented. Ring fragmentation is defined as

$$F_R = \frac{P_{Hit}}{P_C} \tag{3.20}$$

where P_{Hit} is the number of pixels with a signal on the circumference of a ring and P_C is the number of all pixels on a ring. A weakly fragmented ring has F_R of the order of 0.5 or higher. In this case it is likely that at least one neighbor of each hit pixel is hit as well and that spaces in between hit pixels are of the order of one or two pixels without a hit. At HERMES the rings of the gas radiator fulfill these requirements, if the particles are about 1 GeV/c above threshold, as will be described in more detail in chapter five.

The algorithm consist of several stages and the first one is the statistical evaluation to find hits which possibly belong to a cluster. For this purpose, the number of hits in an area around the hit is counted and each hit is weighted by a function of the distance, for example an exponential decay

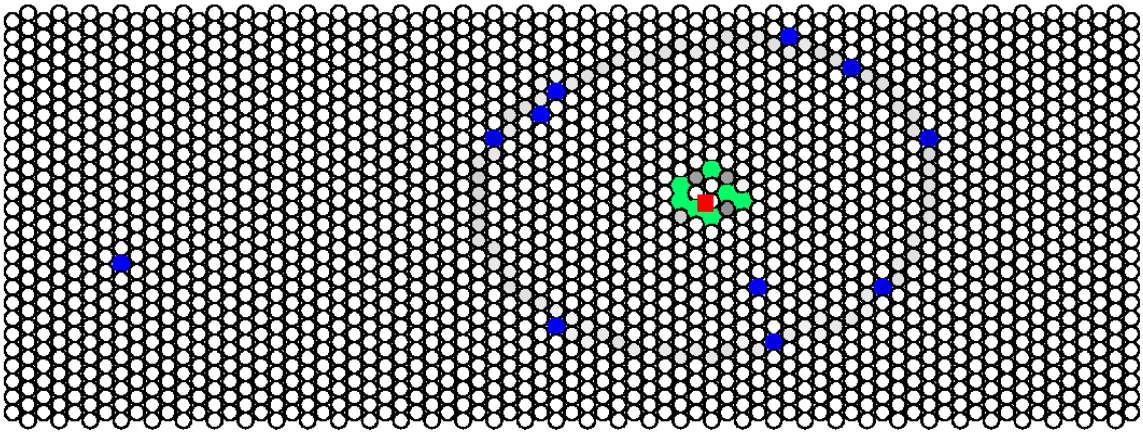
$$C_i = \sum_{j \text{ with } 0<r_{ij}<r_{\text{lim}}} e^{-r_{ij}} \tag{3.21}$$

where r_{ij} is the distance between both hits and r_{lim} limits the area and should be at least of the order of the asymptotic radius. This area is called the disc of influence of the hit. Only hits with C_i above a certain threshold are admitted, which is an implicit cut on noise. Since smaller momentum particles have smaller rings, the reduction in number of hits should be compensated by the reduction in distance. This does however only work up to a limit where rings and noise become indistinguishable. The hits found by this algorithm are called cluster hits.

The second stage is the line stage. There each hit is connected by a line segment to a neighboring cluster hit or to a cluster hit that is maximally two pixels away. This line segment is not necessarily bidirectional, but all hits have up to two connections. The lines may therefore already be close to each other, but not necessarily have to, and may jump over neighboring pixels.

The lines are the primitives for the next stage, the polygon stage. In this stage, cluster hits with only one connection are tested if they can be connected to another cluster hit with only one connection, in this way producing closed polygons. These polygons may have interconnections between hits within and are not necessarily convex. However, each polygon is now referred to as a cluster.

The last stage now operates on these new primitives and therefore is called the cluster stage. Here the clusters are examined if they have the necessary number of cluster hits. This is again a cut on noise. Also all collinear clusters are rejected, since rings are not collinear. Finally small clusters, which are separated but not too far apart, are joined to create a single cluster, since in this case it is probable that a ring has been splitted in two parts. Clusters being too large to be a valid ring are split symmetrically,



3.4. Result of the cluster algorithm. See text for details

assuming that they stem from two closely overlapping rings. This last process shows one limitation of this algorithm: It is restricted to a low ring-multiplicity environment.

The result from a typical run of the cluster algorithm are presented in figure 3.4. Each circle corresponds to a pixel. The square is the position where the reflected track would have impacted on the detector, the virtual track hit. The gray shading is to guide the eye where to expect a ring. The dark hits are noise or part of the aerogel ring, see chapter five. The light hits belong to the gas ring, which has to be found. The color assignment has been done by the cluster algorithm to mark these hits as belonging to one cluster. The figure is taken from the Hermes Rich Event display HeRE, developed in the course of this thesis, and shows a real data event of the 1998 data. In this case all hits have been found. A visual check with 1998 data demonstrated that at least for gas rings the method operates correctly, resulting in an efficiency of $(89 \pm 8)\,\%$ with a fake probability of $(15 \pm 2)\,\%$, which corresponds to a contamination of 15% or a purity of the sample, as defined below, of 85%. The cluster algorithm finds 15% more rings than tracks have been successfully detected by the tracking detectors.

3.4 Optimization

All the approaches described in the previous two sections are designed for an ideal RICH detector. For a real detector, however, additional problems have to be handled, namely:

- Finite resolution. The size of the pixels is limited and hence the spatial resolution of the hits is limited as well. The resolution in time is also limited with regard to resolving multiple hits. Usually the more limiting problem is the spatial resolution.
- Noise. Any hit that does not belong to a track or originally belonged to a track but has been scattered away is considered as noise. It may originate from electronics noise or charged particles crossing the radiator or photon detector without producing rings. This problem might be the most severe in high multiplicity environments. If the background is uncorrelated, it scales linearly with the size of a ring, as shown in equation (3.16) [7]. However, in an environment with high multiplicities of charged particles crossing the detector it might scale with a higher power of the radius [47].
- Overlapping rings. Even more severe than noise can be the situation that two rings overlap at least partially. In this case it is often complicated or even impossible to decide which hit belongs to which

track.

Additionally, the following problems may influence the PID [71]:

- The detector alignment or mirror alignment is not known to the required precision. This problem has to be addressed and can be cured by software using real data [74, 90].
- Optical errors:
  - Non-uniform reflectivity of the mirrors. A solution of this problem can be incorporated for example in a DRT method.
  - Spherical aberrations resulting in deformed rings.
  - Chromatic aberrations. They increase the width of the rings, since the nearly white Čerenkov spectrum is subject to different focal lengths. This increase is enhanced by the fact that the Čerenkov effect itself depends on the index of refraction which in turn depends on the wavelength. Both kinds of aberrations are hard to correct but may be small in reality. They become the dominant problem, if the optics of the RICH detector is at the technical limit [71].
  - Focal plane. It is technically problematic to fit the photon detector to the focal plane, resulting usually in distorted rings. Most photon detectors are planar and not bent. Depending on the resolution this may give rise to severely distorted rings or may have no measurable impacts. This problem can also be addressed by software corrections as it will be described below.
- Multiple scattering. As described by equation (2.30), particles traversing matter are subject to scattering. If the particles are slow or the radiator has a short scattering length, deviations of the particles from their original path will also give rise to rings with a larger width or may even distort the rings in severe cases. However this effect can also be utilized as described in chapter two.
- Stray fields originating from other detector components may have a similar effect. This and the previous problem are hard to address, but the effect may be small, depending on the individual experiment and the available shielding [71].
- Variation of the refraction index. This may be induced by pressure or temperature changes. The variations may be over time or spatial. The result is a non-uniform behavior of the Čerenkov angle and variations of the asymptotic ring size. These effects are normally small and below spatial resolution.
- Ageing. Variation of the refraction index might occur also because of ageing. This is especially important for solid or amorphous radiators. A similar effect may result from absorption of impurities or as a result of mechanical stress.
- Diffraction. If there are windows or structures of the radiator, like pipes or segments for liquid or solid radiators, diffraction at the boundaries might influence the result. This problem can also be solved by software for IRT- and DRT-like algorithms [124].

For track-dependent methods the efficiency of the tracking is of primary importance. Small inefficiencies for the track position of the order of a few pixels, however, can in principle be cured in software by a variational method, varying the hit position to maximize the likelihood [27].

Note that not all of the above listed problems appear for both mirror- and proximity-focused RICH detectors.

The following subsections describe some approaches to cope with these problems, if not already noted above. Since most of the problems mentioned above are of very different importance to each single experiment, in general each experiment will require its own approach, but some principles may be useful for several applications. All of these optimizations require an already existing method for the PID which

they try to improve.

3.4.1 Process Type

Each optimization tries to improve the methods in a different manner, modifying different aspects of the PID. The possible access points are described in this subsection. An optimization may not necessarily be part of only one of the categories, but may belong to several at once.

Preprocessing. Assuming that the method is in principle able to do PID, if not special properties of the data would reduce its effectiveness, a preprocessing of the data is a valid approach. In this case the data is preprocessed and only the cleaned up data is fed to the main method. A possible example is to clear noise from the data by identifying it by its properties [69], like creating large clusters. An extreme example would be to already find interesting areas around a track, identify which hits belong to a possible ring by some pattern recognition method and then feed only the hits belonging to the ring to the PID method which then extracts the particle's type. Since the identification of the ring and the PID are tasks that can be subdivided, this is possible.

Postprocessing. If a method has a high efficiency to identify rings but has a high misidentification or fake probability, the postprocessing of the found rings is an appropriate method. The easiest way is to only accept rings which satisfy some given cuts, like a minimum number of photons, if this is not already included in the method. Other possibilities are a geometrical check if the ring is really a ring or if the density of hit per ring segment is constant. The CERES experiment used for this purpose for example a neural net trained to separate fake and true rings [190].

Parallelprocessing. This type, also known as N-versioning, is based on the availability of several algorithms. In this case all are executed in parallel, and the results are selected upon a democratic voting of the algorithms. This requires at least three methods in order to have a sensible result. If only two are available, the loss of statistics may be prohibitive, at maximum of the sum of fails of both methods. This approach is widely used for controlling purposes in situations with small failure tolerance like airplane, spacecraft or nuclear power plant operation.

Veto is a combination of parallel- and postprocessing. It uses a standard algorithm and checks whether the output makes sense and, if not, activates another method to redo the PID. This is repeated until a result makes sense or the system runs out of methods. In the latter case the event/track is flagged as a failure.

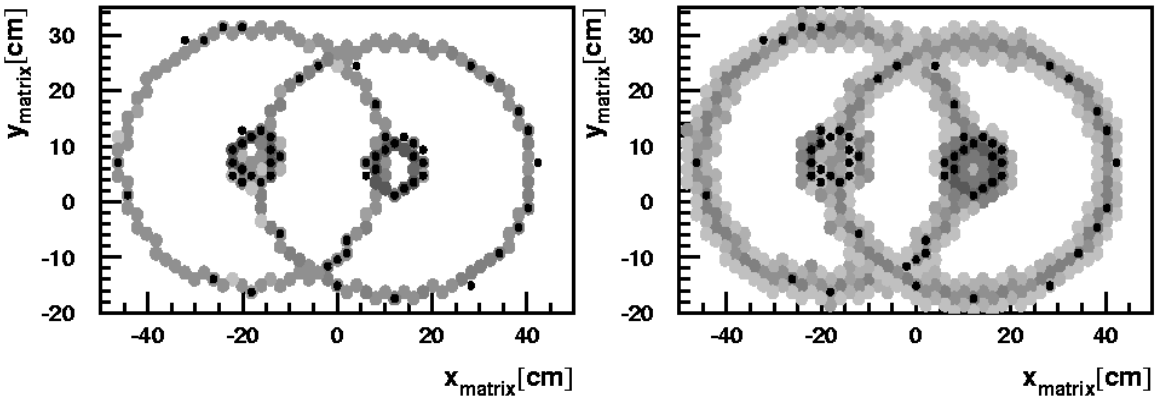
Internal Changes. An optimization may improve some part of the algorithm during its operation compared with the methods described above. This internal change is also a way to optimize a method.

3.4.2 Photon Detector Area

A main problem of the optics of a mirror-focused RICH detector is that a perfect circle can only be expected on the focal surface, which is a concentric sphere of half the radius of the mirror for a wave length independent optical system and a large mirror, as described in chapter two. Ignoring chromatic and scattering effects, only a photon detector on this plane would be able to detect perfect circles. If the displacement of the photon detector from this sphere exceeds a limit which is given by the available resolution, it may disturb the PID. To place it on the focal sphere is normally not feasible because of construction problems connected to a spherical photon detector. Raytracing methods solve this problem implicitly, since the complete optics is traversed. On the other hand, this problem is especially important to be taken into account for track-independent methods. If such a method finds a possible ring, the fit of the ring center has to take the deformation into account. This procedure is strongly dependent on the given optics and detectors. An example is described in [56]. This approach to optimization belongs to the post processing type.

3.4.3 Blobbing

In order to keep the speed acceptable for DRT it is required to use a rough Monte Carlo description. Effects like diffraction, stray fields and scattering may be ignored in such a simplified approach. However, all these effects more or less only result in an enlargement of the ring width. To take this into account, blobbing is an easy and fast possibility. This approach was originally developed during this work in cooperation with P. Jung and R. Kaiser. The aim was to reduce the effect of diffraction at the boundaries of the tiles of the aerogel radiator and of the mirror alignment for the HERMES detector on the PID. These effects were already addressed by an analytic correction, but blobbing allowed to further improve the result.



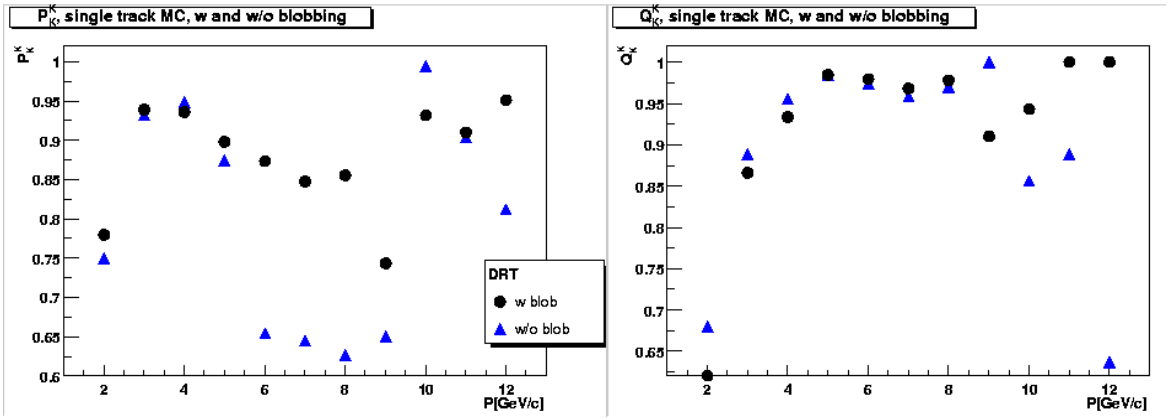
3.5. Blobbing. Left panel shows the event unblobbed, the right panel blobbed. From [90]

For each pixel hit by a photon generated by DRT, each adjacent pixel is assigned the hit of a fractional photon. This effectively increases the area one photon covers by enlarging the ring with an appropriate fall-off towards the outer edges of the ring, as shown in figure 3.5. It results in an effective increase of the number of photons impacting on the detector surface. These photons reach areas of the detector which are inaccessible to photons generated by the standard method. This increase must be corrected for by renormalization so that the total probability will remain to be one. The effect of this internal change to DRT at HERMES was to increase the efficiency to identify a Kaon while at the same time maintaining its purity and did not affect any of the other particles, as shown for Kaons in figure 3.6.

An alternative approach for implementation is to provide a random kick in polar angle for each generated photon before raytracing it. This kick would have to be spread Gaussian-like around the true Čerenkov angle to model the expected spread in angle.

3.4.4 Track Dependence

A fundamental modification is to change from a track-independent method to a track-dependent method. It may become possible, if the layout of an experiment changes severely. It constitutes an obvious improvement because additional information always increases the knowledge about an event and therefore the possibility to reconstruct it. The inclusion of track dependence in this case can be either as severe as a total exchange of a previous method or by additionally pre- or postprocessing. In the case of preprocessing, only regions where a track is pointing to are fed to the method. In the case of postprocessing, found rings are rejected, if no tracks are pointing to it. The combination of parallelprocessing or veto with a second track-dependent method should also allow to improve the results.



3.6. Improvement of a DRT algorithm by blobbing. The left panel shows the kaon efficiency and the right panel shows the kaon purity (HERMES). Statistical error bars are omitted, but are of the order of 10%.

3.4.5 Iterative Procedures

If tracks are available, it is possible to extract for each hit the Čerenkov angle under which the photon has been emitted by each track, as described above. Using this information, PID is possible. However, situations with a high ring multiplicity or with overlapping rings may cause considerable problems. In this case wrong angular information may be incorporated by using hits from another track nearby. In such a situation it would be advantageous to know which hit belongs to which track.

An iterative procedure, belonging to the class of postprocessing optimizations, is a solution. Two possible implementations are presented here.

A first approach uses an iteration with respect to the connection between hit and track [106,185]. As a first step, for each hit and each track the Čerenkov angle will be calculated. For each track all hits will be entered into a histogram with bins in $\theta_{\check{C}}$, similar to the method of angular Hough transformation described above. The iteration proceeds then as follows:

Each photon i in bin k of histogram j gets a weight which is given by

$$w_i^{jk} = \frac{n_j}{\sum_l n_l} \tag{3.22}$$

where n_l is the number of photons in the bin in which photon i is situated in histogram l. For example, if there are two histograms with two bins, and the occupancy per bin is $2,3$ in histogram 1 and $1,4$ in histogram 2, then photon 1, which is in histogram 1 in bin 1 and in histogram 2 in bin 2 would get a weight of $\frac{2}{2+4} = \frac{1}{3}$ in histogram 1 and in histogram 2 a weight of $\frac{4}{2+4} = \frac{2}{3}$. By this the peaks are enhanced while the background is reduced. This is iterated until a suitable peak-to-background ratio is achieved.

If the approach is based on a likelihood, another ansatz is given by [164]. It proceeds through the modification of the influence of the hits on the likelihood itself in each iteration step. A generalized version operates as follows: In a first step all Čerenkov angles for each track will be reconstructed, and a likelihood will be defined by

$$-\ln L \quad = \quad \langle N_{xPh} \rangle - \sum_{\text{All hits}} \ln \phi_i^x \tag{3.23}$$

$$\phi_i^x = \frac{r_{x\check{C}}^{\exp}}{\sigma_i} \exp\left(-\frac{1}{2}\left(\frac{r_{x\check{C}}^{j} - r_{x\check{C}}^{\exp}}{\sigma_i}\right)^2\right) + \phi_H \tag{3.24}$$

Here $r_{x\check{C}}^{\exp}$ is the Čerenkov radius expected for track x. Each track has at this time an initial particle type assigned either by the method to be optimized or, if this optimization is rather used as a stand-alone method, as the most numerous particle type. $r_{x\check{C}}^{i}$ is the actual radius for hit i, σ_i the resolution at the pixel of hit i and ϕ_H a background contribution determined empirically or by Monte Carlo. N_{xPh} ist the number of hits expected for track x.

This likelihood is then modified by

$$\phi_i^x \rightarrow \phi_i^x + \Phi_i^{xy} \tag{3.25}$$

$$\Phi_i^{xy} = \sum_y \frac{r_{y\check{C}}^{\exp}}{\sigma_i} \exp\left(-\frac{1}{2}\left(\frac{r_{x\check{C}}^{i} - r_{y\check{C}}^{\exp}}{\sigma_i}\right)^2\right) \tag{3.26}$$

This summation is performed for all other tracks y having a radius $r_{y\check{C}}^{\exp}$ for the most likely hypothesis. This is done for all tracks and then iterated, until the likelihoods are minimized. In each iteration step the most likely hypothesis for the other tracks y may have changed by the previous iteration; hence the change in each step.

3.4.6 Noise Reduction For Angular Hough Transformation

A simple way to reduce background in an internal fashion for the angular Hough method is to parametrize the background as a function of the Čerenkov angle. This function (it might be linear as in (3.16) or much more involved [47]) will then be subtracted for each track from the histogram and therefore removes a bias of the angle by the background. This bias is introduced by any non-constant background, for example a background increasing linearly with $\theta_{\check{C}}$ will bias the angle towards larger values.

The usefulness is increasing with the photon statistics N_{Ph} for a radiating track, since for a low number of entries, statistical fluctuations may dominate any background effects. This limitation has to be taken into account. If the statistical fluctuation per bin is much larger than the background contribution, the background reduction will be shadowed by a single photon. This is the case, if an additional photon enhances the contents of the histogram by a larger amount than the integrated noise, and the probability to find such an additional photon is as high as nearly unity per histogram. However, in these cases the average background is normally so weak that it can be neglected.

3.4.7 Complementary Particle Identification

The use of additional information is always useful to improve the abilities of particle identification. While the inclusion of the tracking allows to improve the RICH PID itself, the addition of another PID detector is going beyond a single detector PID. Since this is not part of this work, the possibilities to optimize the PID by other detectors will only be sketched.

A possible improvement can be the extension of the momentum range. This is possible in both directions, either to low momenta, where thresholds restrict a RICH detector, or to high momenta, where the rings become unresolvable. An extension to low momenta is possible using for example TOF detectors or the dE/dx characteristics of tracking detectors like a TPC [30]. At high momenta, there are two main options. The first one is to use another detector at all like a TRD, which is dependent on Einstein γ rather than on Einstein β [52]. The other is to delegate the identification of the one or more of the lightest species to another detector, for example electron identification to a TRD. For the muons, which are troublesome because of the small mass difference to the pions, special muon detectors are available, for example muon hodoscopes [30]. A general option for both is always the inclusion of a Threshold Čerenkov counter or another RICH, using other radiators. The improvement can be seen for example by

figure 5.35, where at low momenta the range is extended by a TOF detector and the lepton identification is separated to extend the maximum momentum where a pion can be identified. The details are described in chapter five.

The combination of several PID systems allows for a better performance, since more independent information is available. To build up an optimized PID system, the sensitive ranges and observables of each detector subsystem has to be taken into account properly. An option to do this are decision networks, which are described below.

3.4.8 Decision Networks

Decision networks, also referred to as committee networks, belong to the field of artificial intelligence (AI) and neural networks, well known tools to optimally combine results from different sources into a single result [37]. The approach is to have separate sources, either different detectors as described above or different algorithms for a single detector, in this case a RICH. The output is then fed to a decision network, taking into account all outputs from the different algorithms. It additionally may also take into account the data fed into the original algorithms or other external sources. It then decides either in a "winner takes it all" scenario or in a combinatorial way which is the best result. A "winner takes it all" scenario is that one single algorithm or source is identified as having the best result and the output of this is then made the result of the whole system. A combinatorial approach may combine the results of the different algorithms, providing for example the arithmetic mean of the different particle types identified, if the types are mapped in a sensible way to numbers. In case of RICH detectors, where the particles are explicitly arranged by their mass, an ordinal numbering scheme is appropriate. The true particle type is then determined by rounding off the result.

A simple example of such a decision network based system would be two algorithms, one being strong at low momenta, the other one at high momenta of the particle. The decision network then selects the appropriate algorithm by using the momentum information.

The internal architecture of a decision network may be as simple as "use algorithm A for $p < 5$ GeV/c and B else" or highly sophisticated as a neural network or a knowledge based AI system. A look into more detail of such a network, which effectively combines veto and postprocessing as well as parallelprocessing, is given in chapter five, where an implementation of such a method for the HERMES experiment is described.

3.5 Performance Parameters

If an algorithm has been developed, the most important question is the one of the performance of the algorithm. To allow a comparison of different algorithms, appropriate figures of merits have to be available. This section will develop such figures and the rest of the work will use them throughout. The following is generalized from [92].

The situation is that an algorithm identifies a track of a true type t as a particle of identified type i. If a sample of particles exists, the number N_t^i describes the number of particles of type t identified as particles of type i. All such values can be grouped together in a matrix $\underline{N}$. This matrix has $t_{\max}$ columns, the maximum number of particle species. There are $i_{\max}$ rows, the number of possible identifications. If all species can in principle be identified, $i_{\max} = t_{\max} + 1$. The addition of 1 results from the fact that in general an algorithm cannot always identify a particle, especially below its radiation threshold. In this case the identified type is "unidentified", denoted as X. All species which cannot be identified at all form for this purpose a single species, U. If for example protons and deuterons will never radiate in the observed momentum range, all protons and deuterons are automatically of type U. Additionally, some algorithms may create fake particles. That is, they identify particles which do not exist. Such fakes are always contaminations. They will be denoted by having a true particle type Y. This may also occur because of fake tracks delivered by the tracking. The N_Y^X component represents the amount of

unidentified fakes, and will normally be zero, but very specific algorithms may have a nonvanishing value.

The vector $\vec{N}$ with components

$$N_t = \sum_i N_t^i \tag{3.27}$$

can be connected to the differential particle flux $\vec{\Phi}$ with components Φ_t for each species by

$$\vec{\Phi} = \frac{d}{d\tau}\vec{N} \tag{3.28}$$

$\vec{\Phi}$ gives then in each component t the number of particles of type t per unit time τ. For comparison the relative flux factors are defined as

$$\phi_{t_1 t_2} = \frac{\Phi_{t_1}}{\Phi_{t_2}} \tag{3.29}$$

This quantity describes the relative abundance of species t_1 with respect to species t_2. The normalized flux factor is constructed by

$$\phi_t = \frac{\Phi_t}{\Phi} = \frac{N_t}{\sum\limits_{t'} N_{t'}} = \frac{\sum\limits_i N_i^t}{\sum\limits_{i',t'} N_{t'}^{i'}} \tag{3.30}$$

describing the relative abundance of a particle species in a given sample. Φ is the total particle flux defined as

$$\Phi = \sum_t \phi_t \tag{3.31}$$

$\vec{\phi}$ is hence the normalized flux. From $\underline{N}$ it is possible to define the probability of a particle of type t to be identified as of type i as

$$P_t^i = \frac{N_t^i}{\sum\limits_{i'} N_t^{i'}} = \frac{N_t^i}{N_t} \tag{3.32}$$

The result is the probability matrix $\underline{P}$. Note that the columns are normalized by definition to 1:

$$\sum_i P_t^i = 1 \tag{3.33}$$

If no particles of species t are in a sample, the efficiency is by definition 0.

The diagonal elements of $\underline{P}$ represent the probability of a particle of type t to be identified as its true type. This value is therefore the efficiency ε_t of a detector and a given algorithm to identify particles of species t:

$$\varepsilon_t = P_t^t \tag{3.34}$$

Note that this efficiency contains effects of detection as well as identification. Above the radiation threshold the detection efficiency δ_t bounds ε_t as an upper limit. It describes the probability to detect a particle. It is zero below the radiation threshold and beyond it is

$$\delta_t\left(N_{\min}\right) = 1 - \sum_{n=0}^{N_{\min}-1} N_{Ph}^n \frac{e^{-N_{Ph}}}{n!} \tag{3.35}$$

by virtue of equation (2.11). $N_{\min}$ is the minimum number of detected photons to allow identification. This number limits the detection probability especially in environments with large noise contaminations. If for example 10 is the average number of expected photons, and the detection limit is 3, then the detection probability is 99% within the fiducial acceptance. This limitation is imposed by Poissonian statistics. It includes all detector specific properties in the value N_{Ph}, which depends on the momentum and particle type and might depend on spatial position, charge etc. If the RICH is used as a threshold counter, the detection efficiency does not influence the result. It is replaced by the probability to find noise hits to fake another species, see for example equation (3.18).

Note that the X-type influences the efficiency, while the Y-type does not.

A rejection factor can be defined as the ratio of true number of particles of type t divided by the number of particles incorrectly identified as of type t but of true type t'

$$R_t^{t'} = \frac{\sum\limits_i N_t^i}{N_{t'}^t} \tag{3.36}$$

which are just the inverse of the elements of $\underline{P}$

$$R_t^{t'} = \frac{1}{P_{tt'}} \tag{3.37}$$

The total rejection factor for a species t is given by

$$\frac{1}{R_t} = \sum_{j \neq t} \frac{1}{R_t^j} \tag{3.38}$$

and denotes the true signal to identified background ratio. More interesting is the true signal-to-background ratio given by

$$M_{SBt} = \frac{N_t^t}{\sum\limits_{j \neq t} N_j^t} \tag{3.39}$$

which gives the true rejection power.

Another important quantity derivable from $\underline{N}$ is the true type probability Q_t^i

$$Q_t^i = \frac{N_t^i}{\sum\limits_{t'} N_{t'}^i} \tag{3.40}$$

The value Q_t^i describes the probability that a particle identified as i is of true type t. The off-diagonal elements of the matrix $\underline{Q}$ give the contaminations of a particle sample of type t by another species t'

$$c_{t'}^t = Q_{t'}^t \tag{3.41}$$

The diagonal elements Q_t^t are the purities of the sample, being

$$Q_t^t = 1 - \sum_{t' \neq t} c_{t'}^t \tag{3.42}$$

Note here that the purity is not influenced by the X-type but by the Y-type. It is also noteworthy, that the rows are normalized by definition to one

$$\sum_t Q_t^i = 1 \tag{3.43}$$

It is possible to construct the identified particle number vector $\vec{I}$ defined by

$$I_i = \sum_t N_t^i \tag{3.44}$$

In this case Q can be written as

$$Q_t^i = \frac{N_t^i}{I_i} \tag{3.45}$$

By this a relation between P and Q can be established

$$I_i Q_t^i = P_t^i N_t = N_t^i \tag{3.46}$$

$$\mathrm{diag}\left(\vec{I}\right) \underline{Q} = \underline{P}\,\mathrm{diag}\left(\vec{N}\right) = \underline{N} \tag{3.47}$$

This can be divided by the total particle flux Φ to obtain an equation using the true and identified fluxes $\vec{\phi}$ and $\vec{\phi}'$ respectively

$$\mathrm{diag}\left(\vec{\phi}'\right)\underline{Q} = \underline{P}\,\mathrm{diag}\left(\vec{\phi}\right) \tag{3.48}$$

In the special case where the Y-column is omitted, (3.48) can be reduced to [92]

$$Q_t^i = \frac{P_t^i \phi_t}{\sum\limits_{t'} P_{t'}^i \phi_{t'}} \tag{3.49}$$

It is now obvious that the connection between the efficiency and purity is flux dependent. It is also noteworthy, that while the purity depends directly on the relative fluxes of the particles, the efficiency does not. However, if the efficiency is depending on the event topology, and the event topology is flux-dependent, also the efficiencies become implicitly flux-dependent. For more than one track per event inside the acceptance or position dependent efficiencies this is usually the case.

No algorithm can be characterized by either the efficiency or by the purity alone, since it is possible to have 100% efficiency for species t by just identifying all particles as of species t. At the same time, if only one particle is identified and this one correctly, the purity is at 100%. Both outcomes, although optimal in terms of efficiency and purity alone, are not satisfactory. Only a combined quantity can give a valid indicator. A possible indicator is the effectivity

$$E_t = \frac{P_t^t + Q_t^t}{2} \tag{3.50}$$

This quantity has a nonvanishing value only if there is at least one particle identified correctly. The reason to define it as a sum rather than a product is, that an effectivity of 50% should correspond to 50% efficiency and 50% purity, as it is expected for such a term. Another possible way to estimate the performance of an algorithm is a PQ-diagram, in which the dependence of the purity on the achieved efficiency is plotted.

The effectivity is still not a satisfactory measure, since it gives only an estimate on how well an algorithm performs for a single species. A more convenient quantity is the quality of the algorithm defined as

$$F = \frac{\sum\limits_i C_i^P P_i^i + \sum\limits_i C_i^Q Q_i^i}{\sum\limits_{i,j} C_i^j} \tag{3.51}$$

where the coefficients C_i^j have been chosen to weigh the importance of the different species for the physics analysis. Hence it is possible that different analyses will judge the quality of an algorithm differently. However, within the framework of one analyses, the quality of the algorithm is given by (3.51).

For the last two quantities, effectivity and quality, it should be noted that they are not uncorrelated. $\underline{Q}$ can be expressed via equation (3.48) as a function of $\underline{P}$ and $\vec{\phi}$ and $\vec{\phi}'$, the last one being directly accessible as it is the result of the PID process. However, effectivity and quality are a convenient measure of the performance of an algorithm with respect to a single species in a given physics case.

3.5.1 Statistical Errors

Errors of a statistical nature can be calculated as usual also for efficiencies and purities. In these cases, the errors are given by

$$\Delta P_t^i = \sqrt{\left(\frac{\partial P_t^i}{\partial N_{it}}\right)^2 N_t^i + \sum_{i' \neq i}\left(\frac{\partial P_t^i}{\partial N_t^{i'}}\right) N_t^{i'}} \tag{3.52}$$

$$\frac{\partial P_t^i}{\partial N_t^j} = \frac{1}{N_t} - \frac{N_t^i}{N_t^2} = \frac{1}{N_t}\left(1 - P_t^i\right) \text{ for } i = j \tag{3.53}$$

$$\frac{\partial P_t^i}{\partial N_t^j} = -\frac{P_t^i}{N_t} \text{ for } i \neq j \tag{3.54}$$

$$\Longrightarrow \Delta P_t^i = \frac{1}{N_t}\sqrt{\left(1-P_t^i\right)^2 N_t^i + P_t^{i2}\left(N_t - N_t^i\right)} \tag{3.55}$$

the same can be done similarly for the purities resulting in

$$\Delta Q_t^i = \sqrt{\left(\frac{\partial Q_t^i}{\partial N_t^i}\right)^2 N_{it} + \sum_{t'\neq t}\left(\frac{\partial Q_t^i}{\partial N_{t'}^i}\right) N_{t'}^i} \tag{3.56}$$

$$= \frac{1}{I_i}\sqrt{\left(1-Q_t^i\right)^2 N_t^i + Q_t^{i2}\sum_{t'\neq t}\left(I_i - N_t^i\right)} \tag{3.57}$$

For the vectors of total number and fluxes, the error is just the normal statistical error. For the relative fluxes the errors are given by

$$\Delta\phi_i = \frac{1}{\Phi}\sqrt{\left(1-\frac{N_i^2}{\Phi^2}\right)^2 N_i + \frac{1}{\Phi^2}\sum_{j\neq i} N_i N_j^2} \tag{3.58}$$

$$\Delta\phi_i' = \frac{1}{\Phi}\sqrt{\left(1-\frac{I_i^2}{\Phi^2}\right)^2 I_i + \frac{1}{\Phi^2}\sum_{j\neq i} I_i I_j^2} \tag{3.59}$$

For effectivities the error is given by

$$\Delta E_t = \frac{1}{2}\sqrt{\Delta P_t^{t2} + \Delta Q_t^{t2}} \tag{3.60}$$

and for the quality it is given by

$$\Delta F = \frac{1}{\sum\limits_{i,j} C_i^j}\sqrt{\sum_i C_i^{P2}\Delta P_i^{i2} + \sum_i C_i^{Q2}\Delta Q_i^{i2}} \tag{3.61}$$

3.5.2 Extraction of Performance Parameters

The extraction of efficiencies and purities from an experimental and software setup is difficult. Since these values are influenced by the data which are delivered by the detector and by the physical particle flux, in principle these figures can only be extracted from real data. These values then characterize the performance of the algorithms as well as of the detector. However, they are useful for comparison of the algorithms although the detector influences them, since the detector is the same for all methods and provides only upper bounds.

The most desirable approach is to understand the detector so well and have sufficiently good estimates of the particle fluxes and their spatial and momentum dependences, that an appropriate Monte Carlo can be used, making the extraction of the values easy. However, this is in general not possible, or only up to an approximation which has to be checked with real data. For this purpose a clear sample of the different particle species has to be extracted, together with the flux values. To extract these, there are different possibilities, depending on the available resources, but very specific to the experiment. One possibility is to create a clean sample using other parts of the detector. The other uses the characteristics of decaying particles. The latter technique is well described for the special case of the HERMES detector in [125].

3.5.3 An Example

To illustrate these concepts, the following situation is considered. A RICH is used to discriminate electrons and pions and uses a non-perfect tracking device. In this case let the $\underline{N}$-matrix be for a given sample

$$\underline{N} = \begin{pmatrix} N_e^e & N_\pi^e & N_Y^e \\ N_e^\pi & N_\pi^\pi & N_Y^\pi \\ N_e^X & N_\pi^X & N_Y^X \end{pmatrix} = \begin{pmatrix} 100 \pm 10 & 4 \pm 2 & 2 \pm 1 \\ 20 \pm 4 & 400 \pm 20 & 1 \pm 1 \\ 10 \pm 3 & 30 \pm 5 & 0 \pm 0 \end{pmatrix} \tag{3.62}$$

The efficiency calculated by (3.32) would then be (rounded off)

$$\underline{P} = \begin{pmatrix} P_e^e & P_\pi^e & P_Y^e \\ P_e^\pi & P_\pi^\pi & P_Y^\pi \\ P_e^X & P_\pi^X & P_Y^X \end{pmatrix} = \begin{pmatrix} 0.77 \pm 0.04 & 0.01 \pm 0.00 & 0.67 \pm 0.27 \\ 0.15 \pm 0.03 & 0.92 \pm 0.01 & 0.33 \pm 0.27 \\ 0.08 \pm 0.02 & 0.07 \pm 0.01 & 0.00 \pm 0.00 \end{pmatrix} \tag{3.63}$$

the purity is then given by virtue of equation (3.40)

$$\underline{Q} = \begin{pmatrix} Q_e^e & Q_\pi^e & Q_Y^e \\ Q_e^\pi & Q_\pi^\pi & Q_Y^\pi \\ Q_e^X & Q_\pi^X & Q_Y^X \end{pmatrix} = \begin{pmatrix} 0.94 \pm 0.02 & 0.04 \pm 0.02 & 0.02 \pm 0.01 \\ 0.05 \pm 0.01 & 0.95 \pm 0.01 & 0.00 \pm 0.00 \\ 0.25 \pm 0.07 & 0.75 \pm 0.07 & 0.00 \pm 0.00 \end{pmatrix} \tag{3.64}$$

Notice now, that $\vec{N}$ and $\vec{\Phi}$ have the following values, if assumed that $\vec{N}$ has been measured in unit time:

$$\vec{N} = \begin{pmatrix} N_e \\ N_\pi \\ N_Y \end{pmatrix} = \begin{pmatrix} 130 \pm 11 \\ 434 \pm 21 \\ 3 \pm 2 \end{pmatrix} \tag{3.65}$$

$$\vec{\Phi} = \vec{N},\ \Phi = 567 \tag{3.66}$$

This results in the relative flux of

$$\vec{\phi} = \begin{pmatrix} \phi_e \\ \phi_\pi \\ \phi_Y \end{pmatrix} = \begin{pmatrix} 0.23 \pm 0.02 \\ 0.77 \pm 0.02 \\ 0.01 \pm 0.00 \end{pmatrix} \tag{3.67}$$

the missing normalization is due to rounding errors. The identified flux and the number of identified particles on the other hand is given by

$$\vec{I} = \begin{pmatrix} I_e \\ I_\pi \\ I_X \end{pmatrix} = \begin{pmatrix} 106 \pm 10 \\ 421 \pm 21 \\ 40 \pm 6 \end{pmatrix} \tag{3.68}$$

$$\vec{\Phi}' = \vec{I},\ \Phi' = \Phi = 567 \tag{3.69}$$

$$\vec{\phi}' = \begin{pmatrix} 0.19 \pm 0.02 \\ 0.74 \pm 0.02 \\ 0.07 \pm 0.01 \end{pmatrix} \tag{3.70}$$

Now (3.48) is fulfilled up to rounding errors

$$\underline{P}\,\mathrm{diag}\left(\vec{\phi}\right) = \begin{pmatrix} 0.77 & 0.01 & 0.67 \\ 0.15 & 0.92 & 0.33 \\ 0.08 & 0.07 & 0.00 \end{pmatrix} \begin{pmatrix} 0.23 & 0 & 0 \\ 0 & 0.77 & 0 \\ 0 & 0 & 0.01 \end{pmatrix} = \begin{pmatrix} 0.18 & 0.01 & 0.01 \\ 0.03 & 0.71 & 0.00 \\ 0.02 & 0.05 & 0.00 \end{pmatrix} \tag{3.71}$$

$$\mathrm{diag}\left(\vec{\phi}'\right)\underline{Q} = \begin{pmatrix} 0.19 & 0 & 0 \\ 0 & 0.74 & 0 \\ 0 & 0 & 0.07 \end{pmatrix} \begin{pmatrix} 0.94 & 0.04 & 0.02 \\ 0.05 & 0.95 & 0.00 \\ 0.25 & 0.75 & 0.00 \end{pmatrix} = \begin{pmatrix} 0.18 & 0.01 & 0.00 \\ 0.04 & 0.70 & 0.00 \\ 0.02 & 0.05 & 0.00 \end{pmatrix} \tag{3.72}$$

The effectivity vector for this case is given by

$$\vec{E} = \begin{pmatrix} E_e \\ E_\pi \end{pmatrix} = \begin{pmatrix} 0.86 \pm 0.04 \\ 0.94 \pm 0.01 \end{pmatrix} \tag{3.73}$$

which shows that the algorithm performs somewhat better for pions than for electrons. Assuming that the particles are equally important, the quality is then given by

$$F = 0.90 \pm 0.05 \tag{3.74}$$

which with 90% would be a good value for most physics analyses.

Chapter 4
CERES

CERES, ČErenkov Ring Electron Spectrometer, is an experiment at the CERN SpS and part of the CERN heavy ion program. CERES started to take data in 1992. It researches QCD under extreme conditions in nucleon-nucleus and nucleus-nucleus collision in a fixed target geometry. The beam is either a proton beam of up to 450 GeV/c or an ion beam of sulfur or lead up to 200A GeV/c and 160A GeV/c respectively. The segmented target is either beryllium or gold. The experiment mainly observes dileptons from the fireball created in the collision and uses for this purpose two nearly hadron-blind mirror-focused RICH detectors. It was upgraded twice, once from 1994 to 1995 to allow operation in the environment of the higher multiplicity of lead on gold collisions. A second time it was upgraded in 1998 to add a TPC as a tracking device. This was originally done to improve the invariant mass resolution. This addition also allows to use track-dependent methods to do PID instead of or additionally to the already existing track-independent method. This chapter describes the methods that have been developed, as well as their results. It is therefore also a demonstration on how to design and develop PID algorithms.

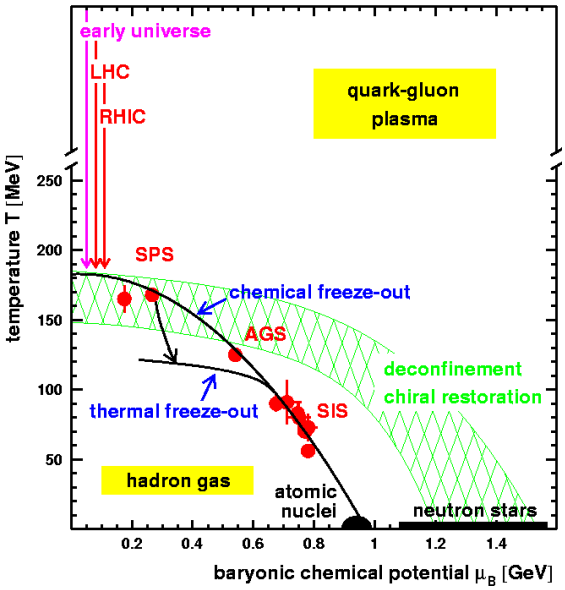
This chapter will first introduce the physics of CERES and afterwards its setup. It will then provide a short description of the current reconstruction algorithm. This will be followed by an introduction to a new reconstruction scheme with an emphasis on the PID aspect. The latter will then be expanded in the following section, demonstrating the basic concepts and obstacles as well as first Monte Carlo results. Finally, another short section will complete this chapter by describing how the additional tracking information can be used to upgrade the present strategy using the tracking capabilities of the TPC in a postprocessor fashion.

4.1 Physics

In heavy-ion collisions, many body systems obeying the strong interaction described by the QCD are under investigation. Of special interest to map the phase diagram of these systems, as shown in figure 4.7. From multi-fragmentation nuclear physics experiments it is known that nuclear matter has a phase transition from a liquid to a gas. Theoretical models predict a phase transition from a baryonic system to a system of free quarks and gluons, the QGP. Confinement forces colored objects like quarks and gluons normally to be confined within hadrons, the normal environment of everyday life. At temperatures of the order[10] of 150 MeV this normal matter should undergo a transition to QGP and will then be deconfined, so that the fundamental objects are colored. Additionally the broken chiral symmetry observed at nuclear conditions would be simultaneously restored. CERES and several other experiments at the SpS examine the possibility of such a phase transition in heavy ion collisions until 2000, when this heavy ion program at CERN ends. Their successors will be experiments at RHIC, which can profit from higher luminosity, more running time and higher center-of-mass energy. At an even higher center-of-mass energy, ALICE will follow up starting not earlier than 2005. An overview of the status of experimental mapping of the QCD phase diagram can be found in [41, 184], for the theoretical part [177, 179, 194] give an overview.

The transition to QGP was claimed to be found at CERN at the beginning of 2000. One of the experiments contributing to this discovery was CERES.

[10] In the following physics discussion in this section only $\hbar = c = k_B = 1$



4.7. The phase diagram of QCD, from [41]. The phases at high densities are not shown

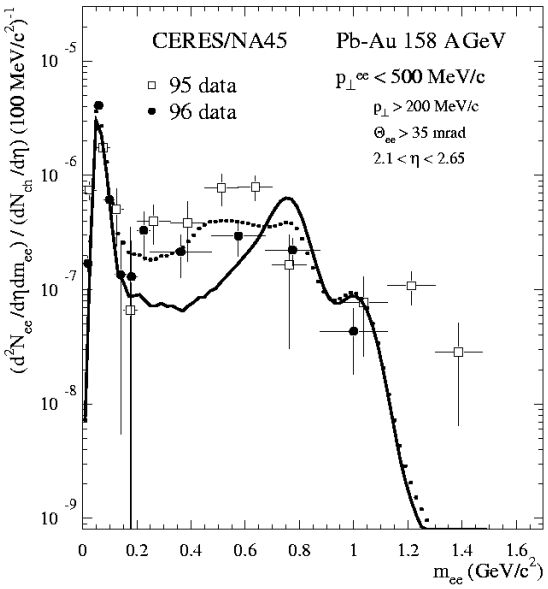
While for experiments heavy ion collisions are the method of choice to investigate the QCD phase diagram, there are several different methods which try to access this area from a theoretical point of view. The reason for this split-up is mostly the non-linear structure of the QCD Lagrangian, which makes analytical calculations extremely difficult, perhaps impossible. A numerical approach is lattice gauge theory [142]. However, lattice gauge theory is for technical reasons not yet able to perform calculations for non-vanishing net-baryon densities. For pure gauge or a $SU(2)_C$, that is without fermions or fewer colors respectively, lattice theory is possible even for finite net-densities, resulting in phase transitions as well and the production of glueballs, which are also calculated by phenomenological methods [160]. For finite densities other possibilities have been found and are mostly model systems: Instanton models [141, 161, 169, 191], random matrix theory [79], random phase approximation [142], percolation [65, 155], comparison with supersymmetric theories [139], or others [80], even as exotic as described in [61]. An overview is given for example by [154]. Some of these methods are quite successful for describing part of the current experimental observations. A thermal model, for example, to describe the dilepton yield in the intermediate mass region 1.5 GeV $\leq M_{l\bar{l}} \leq$ 3 GeV, is quite satisfactory [143]. However, no model is able to explain all observations simultaneously up to now and there are no first principal calculations of any of the observations. One approach is to find a universal equation of state, an example of this effort is contained in [189].

The understanding of the phases of QCD at high temperatures is important to understand the dynamics of the universe at early stages. The influence of a strong phase transition on the baryogenesis can be found for example in [85]. The influence of phase transitions in general on the baryogenesis is explored in [34, 112].

At the same time the physics of QCD at large baryonic densities and low temperatures is important for the understanding of compact stellar objects like neutron stars [14, 84].

In this sense, high energy nuclear physics is connected to cosmology and astrophysics.

The CERES experiment tries to probe this state of matter with dileptons. Because of the thermalization of the system, any information about the previous state is lost. Hence only weakly interacting particles like leptons can be used to directly analyze the QGP and its formation if it forms at all. An ex-



4.8. The measured dilepton excess from CERES [41]. The solid line is from the extrapolation of nucleon-nucleus collisions, the dashed line takes rho modifications into account [172]. The difference between the data points and the solid line is the measured excess

tensive discussion of the basic theoretical models of dilepton production can be found in [142]. CERES analyzes the electron/positron pairs from short-lived resonances like the ρ-meson, which are believed to decay during the collision and due to this their properties are influenced by the surrounding medium. If this influence leads to for example a reduction of their mass, this should result in a change of the dilepton yield. Exactly this is what is observed by CERES: An increase of the dilepton yield at the intermediate invariant mass range from 200 to 600 MeV, as shown in figure 4.8. In other mass regions the HELIOS-3 collaboration also measured an excess.

A possible explanation of the dilepton excess observed by CERES and HELIOS-3 is a change of the ρ mass and/or width, described by the Brown-Rho scaling [172]. Beside this explanation there are several others, for example a mass change in the η', η-system [98]. This excess is an indication for the formation of a QGP [162].

Taking all observations and all results from theoretical calculations together, the following picture is emerging, although yet no final confidence can be given to these results.

The phase diagram of QCD probably has a smooth crossover for small baryon densities and high temperatures of the order of 150 MeV and a first order phase transition for higher densities and low temperatures, ending in a tricritical point [35, 132, 136, 178, 186] at intermediate densities and temperatures, if a strange mass of the order of the observed one is deployed. For high baryon densities and low temperatures the appearance of three new phases beyond the hadronic phase is currently expected, being a color superconducting phase [11] involving up and down quarks called $CSC2$ [134], one phase additionally including the strange quark $CSC3$ [12] and at even higher densities a color flavor locking mode CFL. CFL is the breaking of color $SU(3)_C$ and flavor symmetry $SU(3)_F$, leaving only a combined color and flavor symmetry $SU(3)_C \times SU(3)_F$ unbroken [133,135]. In the superconducting phase, where quarks form Cooper pairs, the chiral symmetry is restored, but color symmetry is broken to $SU(2)_C$ [36]. Whether $CSC3$ and CFL are physically distinct is not clear and depends on the mass of the strange quark [176]. A possible explanation for this behavior can be given by an Instanton-Anti-Instanton liq-

uid model which generates Instanton-Anti-Instanton molecules at high densities and finite temperatures [173]. The existence of another diquark phase predicted by instanton calculation is vague and parameter dependent [140,178]. The transition at high temperatures and low densities leads to the QGP, where it is expected that confinement vanishes [99] and chiral symmetry is restored [152, 168, 170]. Another change could be the (partial) restoration of the $U(1)_A$ anomaly, as described in [179], a problem connected with the topological charge that is not yet solved.

For the transition to a QGP at high temperatures and low densities, three important indications have been found experimentally that can be justified theoretically. Additionally, measurements of the particle flow indicate that the system thermalizes and can be indeed treated as if in thermodynamical equilibrium. This also indicates that the involved degrees of freedom are not purely hadronic and is hence a fourth indication [174]. The first indication is the already mentioned dilepton excess [86] measured by CERES. The second is an enhancement of strange particle production, which was also measured during the SpS heavy ion program [41,181,184]. The last is a suppression of the hidden charm states J/ψ and its excitations χ and χ'. This suppression is starting from a certain centrality, or equivalently energy density, measured for J/ψ by NA50 [39, 171]. Theoretical models predict this behavior as observed at the SpS [41,184]. A possible explanation is that the gluon distribution rapidly changes from a distribution around 200 MeV to a distribution around 600 MeV. While the former are not able to dissolve the strongly bound J/ψ state, the tail of the latter distribution is [100, 153, 156, 175]. These indications are split into the groups of electronic or leptonic probes, being the dileptons, and in hadronic probes, being strangeness enhancement and J/ψ suppression [151].

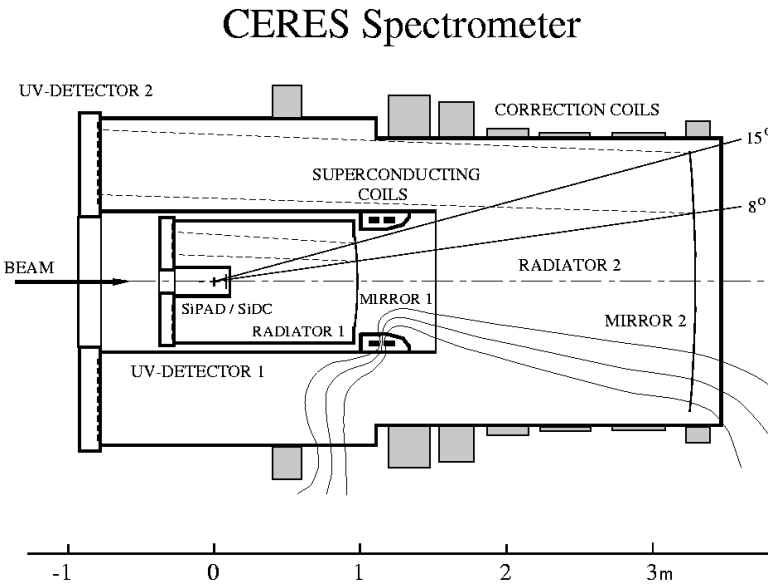
All these results led to the announcement that the QGP exists and that is has been measured at CERN. It will nevertheless be difficult to explore the QGP with the energies available at the SpS. It is hoped that the physics will be much more accessible with the experiments BRAHMS, PHOBOS, PHENIX and STAR at RHIC and afterwards with ALICE at LHC. Jets on the experimental and pQCD on the theoretical side are expected to be accessible then [123].

If looking further, an interesting connection with Deep Inelastic Scattering (DIS) can be seen for example in [114–116]. Perhaps there might even be a connection to polarized DIS , as performed at low energies by HERMES. Magnetic phase transitions could be measured in such an environment.

4.2 Setup

The main signals detected at CERES are leptonic probes, the electrons and positrons from meson decays. The main problem is to reject the huge hadronic background, which is of the order of 10^6 at CERES energies. To accomplish this goal, a detector design with optimal hadron rejection was chosen. The main part of the forward spectrometer CERES are two RICH detectors with azimuthal symmetry, an acceptance from 8° to 15° in the laboratory frame and a mirror-focused geometry. The radiator was chosen to have a radiation threshold of $\gamma_{th} \approx 32$. This allows electrons to radiate with a momentum of 16.3 MeV/c. Within the resolution, electrons will have an asymptotic radius for a momentum above 50 MeV/c. The lightest hadronic particle, the pion, would radiate starting from 4.48 GeV/c and reaches the asymptotic radius at 13 GeV/c. Kaons will start to radiate only from 15.8 GeV/c onwards. The phase space is only sparsely populated by hadrons at these momenta, so an electron or positron is clearly identified by an asymptotic ring in both RICH detectors, if a momentum cut of 50 MeV/c, which corresponds to transverse momentum cut of minimal $p_t > 7$ MeV/c, is applied. This effectively makes the RICH detectors hadron blind.

To measure the momentum of radiating particles, a magnetic field produced by superconducting coils was placed between the first and the second RICH detector. The field geometry was chosen such that it did not affect the polar angle θ but only the azimuthal angle ϕ of any track originating from the target. The momentum was then determined by measuring the deflection in ϕ together with the known mass of the electron.



4.9. The original layout of CERES

In the original setup of 1992 the only additional detector was a single silicon drift chamber in front of the RICH detectors to reconstruct the interaction vertex, see figure 4.9.

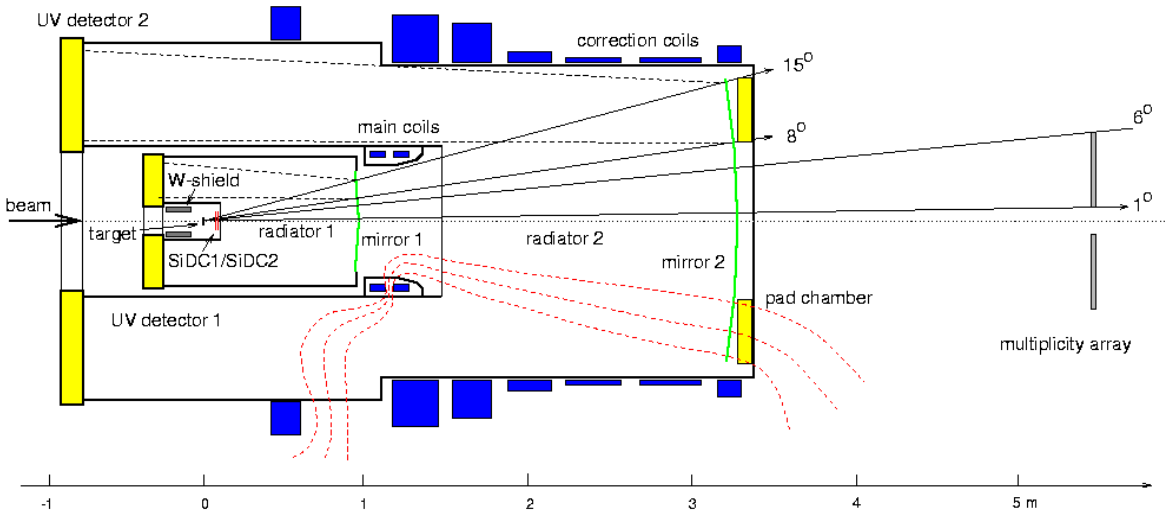
The largest problem in addition to the hadronic background are numerous electrons from pair creations, also called conversion, by photons and from Dalitz decays from mesons[11]. The only characteristic separating these and the leptons from decays of e.g. ρ mesons is the opening angle distribution, which is narrower for Dalitz decays and highly peaked at zero for conversions. To reject this background as well as the hadrons is the most complicate challenge at CERES.

The detector was upgraded once in 1994 to cope with the higher expected charged particle multiplicities as compared to proton or light ion induced reactions, since the available tracking devices would not allow an acceptable level of efficiency and signal-to-noise ratio. In this case the original single silicon drift chamber was replaced by two silicon drift chambers and an additional pad chamber was added. These two additions raised the PID abilities to an acceptable level [8]. The SiDC detectors are mostly used to reject conversions by their characteristics, like doubled dE/dx compared to normal tracks, because of the small opening angle on the order of or below the resolution of the SiDCs. By requiring a hit for each track in both SiDCs, the effective radiation length before the RICH detectors is reduced from originally 2.35% to 0.55%. This new setup is shown in figure 4.10.

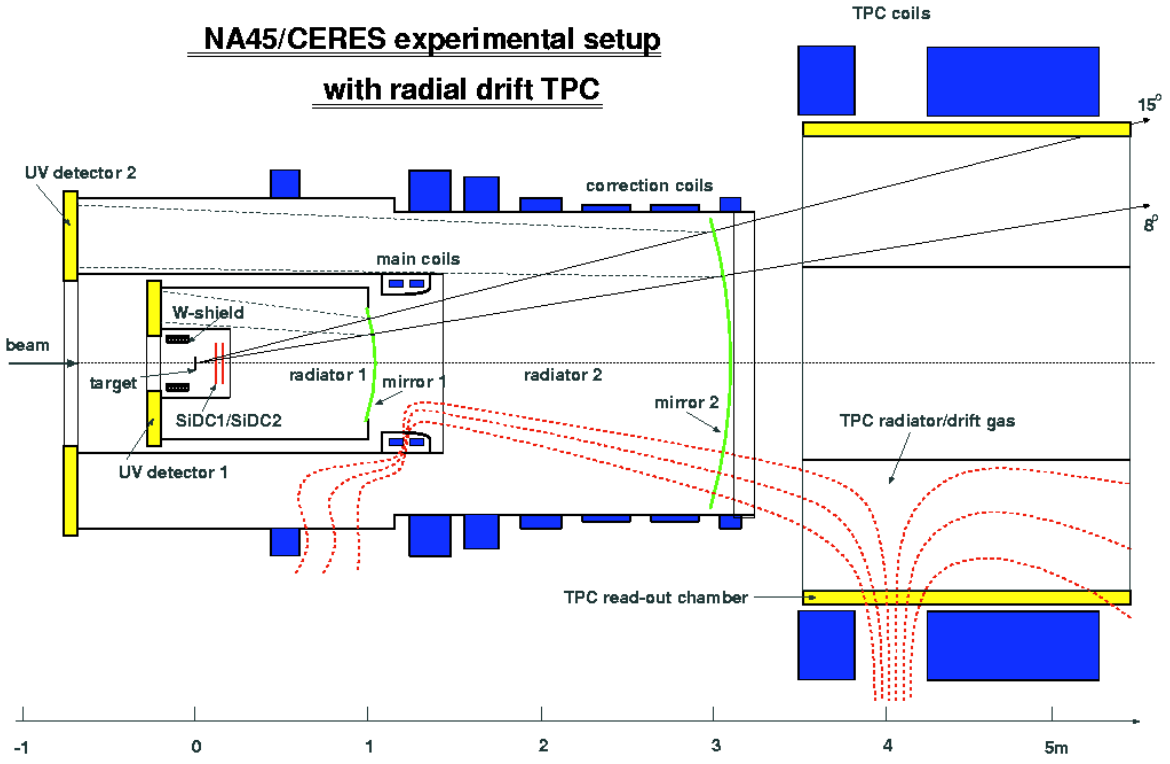
A second upgrade [9] was performed in 1998 and the pad chamber was replaced by a time projection chamber, to allow an increase in mass resolution, $\Delta m/m$. Since the TPC can also provide measurement of momenta, it was now possible to switch off the magnetic field between both RICH detectors. This allowed to combine both detectors effectively to one RICH detector with a doubled photon statistics because of the identical radiator and readout in both RICHes. The TPC also allows tracking with a higher resolution than with the RICH detectors and provides a second PID detector, using the dE/dx information of a track. This current setup is shown in figure 4.11.

A drawback for the TPC is multiple scattering. Since the momentum distribution of the particles covers the spectra down to some ten MeV/c, multiple scattering as described by (2.30) is important. The

[11] Particles from Dalitz decays will be called Dalitz for convenience. Particles from conversion process will be similarily called conversions.



4.10. CERES after the first upgrade



4.11. The current CERES setup

largest scatterer in the setup is the RICH 2 mirror with $x/X_0 = 0.0488$ [33], giving rise to an angular deviation width of 163 mrad for a threshold electron and 53 mrad at 50 MeV/c . A scattering up to an asymptotic ring radius is possible up to about 85 MeV/c. Scattering widths of the order of the single photon resolution of 1 mrad [104] are present up to 265 MeV/c. Since the TPC is placed behind the RICH detectors, tracking using only the TPC is limited, at least at small momenta. Using the SiDC detectors for vertex reconstruction to find one additional point for particles originating from the interaction allows a sensible tracking. Since all other scatterers are smaller by an order of magnitude, back-tracking is possible as described below.

4.3 Present Reconstruction

The present reconstruction strategy uses an integrated approach which combines PID and rejection of physical background.

The starting point are the RICH detectors. By a simple geometric operation both RICH detectors can be projected into one plane. An approach to do so is described below.

The result is a three-dimensional array describing the spatial position and the amplitude for each pixel. It should be remarked that the avalanche generated by each photo electron created in the photo conversion gas is detected by several neighboring pads. Since the pads are the basic objects of the readout and therefore of the data, each photon signal is spread Gaussian over several pixels. It is possible to reconstruct the position of the photon impact from this information, but this is neither used for the reconstruction method as of 1999/2000 nor is it used by the methods presented in this work. Hence throughout this chapter each pixel with an amplitude which is different from zero will be considered as a hit. This differs from the standard conventions used for CERES, where a hit is the fitted photon impact position. It is convenient for the following, since each pixel with a non-zero amplitude is treated as a single hit and only the integrated value is used for counting of photons, but no position fit is performed.

In this plane, all hits are contained and a spatial Hough transformation is performed to find ring candidates with asymptotic radius. The result of this first transformation is then used in an iterative manner to generate weights for each pixel for a second spatial Hough transformation. This second Hough transformation, termed the non-linear Hough transformation, is again spatial. In the resulting dual space, peaks above the Hough threshold amplitude are searched for. These are then considered as asymptotic rings.

A problem of this approach is that it becomes complicated to resolve rings which overlap, because the Hough peaks are not clearly separated anymore. The problem is somewhat similar to the Rayleigh criteria of wave optics. If they overlap completely, only a doubled Hough amplitude is an indication of the overlap. These single tracks are then a source for a combinatorial background, since the dileptons are the observed signal. The problem of unrecognized pairs is the central problem of this algorithm [7].

The SiDC detectors can be used to aid in the rejection. First, all rings which do not have a track in both silicon detectors originate most probably from conversions in later stages of the detector. Second, spatially close tracks produce a characteristic peak in the measured dE/dx in the silicon detectors. This allows to reject conversions and close Dalitz pairs.

This limits the contamination from conversion electrons to those produced in the target and the first silicon detector and reduces the background from close Dalitz pairs. A residual background is nevertheless always present, due to pairs with large opening angles and if statistical fluctuations prevent the rejection of conversions in the SiDCs. An additional unremovable background are δ-electrons, which are kicked out of their shell by charged particles from the interaction. If this happens within the target or the first SiDC, they produce a track with two rings and silicon hits. The problem of the background is worsened by the momentum distribution of Dalitz electrons. Because mostly one of the particles from the Dalitz decay is hard and the other one is soft, one can be affected by multiple scattering, which may lead to the loss of a ring in both detectors or it may be scattered out of acceptance. The remaining hard particle

will then contribute to the combinatorial background [190].

If a pair has survived the rejection, a track is searched for in the TPC, which can be connected to the track to determine its momentum. The connection is made by extrapolating the rings to the RICH 2 mirror, which is assumed planar, and to match it to TPC tracks that are back extrapolated to the same plane. The track from the RICH detectors is constructed by fitting the ring candidates found previously to a ring to find its center with a higher resolution than by just using the peak in the dual space [7].

After this, electron and positron tracks with their momentum are available and the remaining particles should be from the desired source.

For hadrons the approach is similar, although a variable radius spatial Hough transformation is used [109]. The rejection steps are not necessary as far as they are concerned with the pair characteristic. The requirement for a silicon track remains. The momentum determination is analogue to that as for electrons.

4.4 A New Strategy

4.4.1 Track Dependence

The new TPC of CERES is capable of stand-alone tracking with a high resolution of the order of 1 cm at a distance of 350 cm from the target before the magnetic field. This is much better than the single photon resolution of the RICH detectors. It is therefore natural to develop strategies to use the TPC as the primary tracking device. The main problem of this approach remains multiple scattering in the RICH 2 mirror for low momenta tracks, and the main issue is the reliability of the tracks found in the TPC before the mirror. This problem can be solved as already indicated by the fact that all particles of interest originate from the target so that one point is known with high precision, since the sub-target in which the event occurred can be found with the silicon detectors. Together with the impact point on the RICH 2 mirror of the track as extrapolated from the TPC an average trajectory can be found, which smoothes out any local changes due to scattering within the smaller scatterers like the RICH 1 mirror.

Due to this it is feasible to take a track-dependent reconstruction as an additional reconstruction method, which can be combined with the present method in a veto or parallel processing fashion.

4.4.2 New Strategy

The new strategy is fundamentally different from the present one in two ways besides the fact that it starts from the TPC and not from the RICHes. First, it strictly separates between PID and rejection. It first does a best effort PID, trying to identify as many tracks as possible with as low contamination as possible. Rejection on physical grounds to separate for a special signal source is an independent process. The second point regards the difference between electrons and hadrons. Although not described in much detail before, the hadronic tracking is in many details different from the electron tracking, especially because of the variable radius and the changes in rejection strategy [109]. The new approach does not make any difference between both, and uses a unified PID. This would in principle allow to identify also non-asymptotic electrons, but as already calculated before, the multiple scattering makes it very unlikely to get any results from this.

This subsection will describe the different stages of the new reconstruction strategy, and in this leads beyond RICH PID. However, the tracks are necessary for a track-dependent PID, which makes it necessary to deal with the problems of tracking.

The first part is the tracking using the TPC to find tracks and the SiDC to allow then a full tracking inside the detector. The second part is the particle identification using both RICH detectors and the TPC for a track-dependent PID. The last part is concerned with the rejection for electrons and positrons.

4.4.2.1 TPC Seed Searching

The first step of the new strategy is the TPC tracking. Any tracks not found by the TPC will not be used in the next stages of the reconstruction. The current tracking strategy at CERES is to first generate hits from the raw data, since a charged track will deliver a Gaussian distributed hit over several pixels of the TPC. The pixels of the TPC in this case are azimuthal bins and radial time bins according to the time projection principle. The pixels defined by each azimuthal and time bin can be mapped uniquely into cartesian space. The next step of the tracking is to connect these hits to a track. The problem in this approach is the ambiguity which may arise if tracks are crossing each other during the influence of the magnetic field. However, this should occur not too often, since the deflection by the magnetic field is purely in ϕ-direction. Another problem are δ-electrons, which either cross the TPC radially or if they are influenced by the magnetic field create circular paths at a given z-position, where z is measured from the target. Both topologies create a multitude of hits and thus affect the tracking.

The field is contained within the center half of the TPC while the first and last quarter are nearly field-free. Since the TPC is divided into 20 planes, there are four to five planes where the field effects can be neglected at the entrance. To avoid the problems of crossing tracks, a new approach is currently under development and part of the first segment of the implementation was part of this work.

The idea is to find tracks from the raw data directly within the first five planes where crossings of trajectories should be rare. In these five planes track seeds are searched for, giving a first indication where tracks may be at all. One plane of a typical event is shown in figure 4.12[12]. The information from theses seeds are then used to extrapolate the track backwards for combination with the rest of the experiment and are also used to complete the tracking in the TPC to yield momentum and dE/dx information.

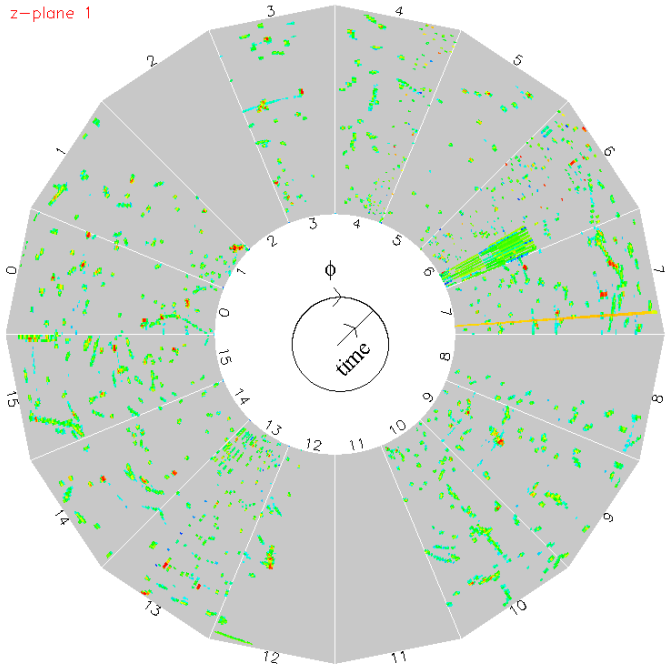
The search for track seeds is done by a Hough transformation, which takes the presence of a signal in the TPC over five planes as the characteristic feature. The algorithm operates as follows:

- A grid of macro pixels is created in the azimuthal- and time-bin space. Each macro pixel is large enough to contain all pixels of an average single hit of a track. For the CERES TPC the size of a macro pixel is two azimuthal bins and three time bins
- In each macro pixel the amplitudes from all covered pixels are summed up
- The macro grid is transferred from the azimuthal- and time-bin space to spherical coordinates, which can be done for each of the five planes separately and uniquely. A track coming from the origin, the target, should now have the same spherical coordinates in each plane
- All five macro grids are summed up

The result of the last summation gives a map of the measurements of the TPC at arbitrary radius in spherical coordinates. Tracks which originate from the target and did not experience too large scattering do not change their azimuthal and polar position. Therefore all hits generated from these tracks will be summed up in one macro pixel with a very high amplitude. δ-electrons or electronic noise will only appear in one plane, and the height of the amplitude will thus be very limited with a maximum height of 1/5 of the maximum amplitude. An appropriate minimal requirement for the amplitude like four times the average amplitude per plane should be good enough to well separate noise and true tracks at a low level of fake tracks. The result of an early implementation where the tracks were signified by a trace in θ direction is shown in 4.13.

It is possible that a track lies at the border of two or four macro pixels. This indeed happens because of the size of the macro pixels in azimuthal direction in 50% of the cases. To prevent this, it is necessary to use a second macro grid which is displaced against the first one by one pixel in both azimuthal and time direction. Each track will then be visible by a full amplitude in exactly one of both grids, and will

[12] A colored version is available from the author or from the CERES and HERMES web pages.



4.12. A typical CERES event in the second plane of the TPC. Each plane is divided azimuthally in 16 chamber, denoted by their number in azimuthal direction. Each one is subdivided in 48 bins. In radial direction the time bins are shown, ranging from 0 to 255. A color scale indicates the amplitude in each azimuthal and time bin.

be distributed in the other grid between at least two macro pixels. Hence all tracks are admitted which have a large amplitude in either of the grids.

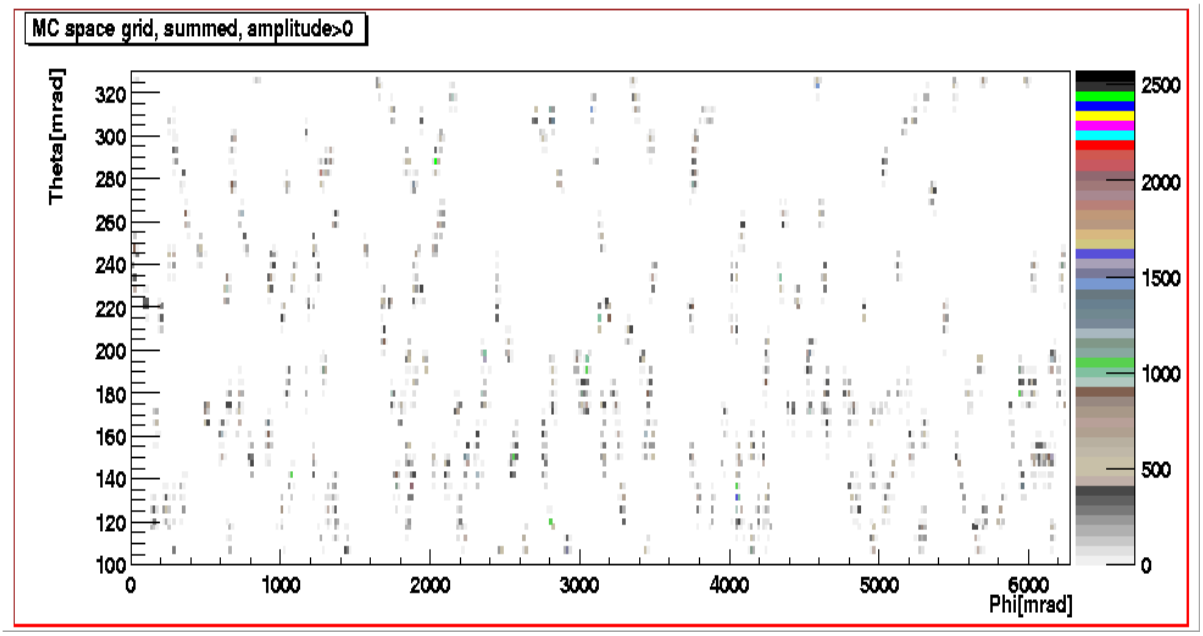
In a next step the contents of the pixels of the macro pixel are used for a fit of the track to increase the precision compared with the single macro pixel. A possible tracking for momentum determination could be added by a template tracking strategy which assumes a momentum hypothesis and checks with templates if the track behaves as expected. These points will not be discussed here any further.

4.4.2.2 Back-Tracking

The next step is back-tracking. Under the continuing assumption that the track originates from the target, but may have been scattered in the RICH 2 mirror by some amount, the track is first extrapolated back to the mirror. The impact point on the mirror together with the event vertex defines, an average track in the front part of the detector. The vertex can be found by the present methods with the SiDCs. To do the back-tracking, it is assumed that the TPC delivers the track as a function of z. The track is then described as

$$\vec{r}(\lambda) = \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} \frac{dx}{dz} \\ \frac{dy}{dz} \\ \frac{dz}{dz} \end{pmatrix} \lambda \tag{4.1}$$

dz/dz sets the scale and can be chosen arbitrarily. Derivatives will be denoted as x', y' and z' from here



4.13. Result from the CERES TPC seed searcher. In this early implementation the tracks are signified not by a single amplitude but by a trace in theta direction because of an arbitray origin. This plot is from Monte Carlo data.

on. In the same coordinate system the mirror surface can be expressed by

$$\vec{s}(\alpha,\beta) = \begin{pmatrix} 0 \\ 0 \\ m \end{pmatrix} + \begin{pmatrix} r\sin\alpha\cos\beta \\ r\sin\alpha\sin\beta \\ r\cos\alpha \end{pmatrix} \tag{4.2}$$

m is the distance between the vertex and the center of the mirror and r is the radius of the mirror. It is assumed spherical instead of planar as in the present method. Any constant offsets which are necessary due to detector alignment can be included in the offset and the scale of the track. If the mirror is not aligned correctly, additional non-zero offsets would have be to be introduced, but effects comparable to the resolution would distort the rings which is not visible in experimental data and is therefore ignored. The track impact point is described by the intersection points of (4.1) and (4.2)

$$\vec{r}(\lambda) = \vec{s}(\alpha,\beta) \tag{4.3}$$

This is a non-linear system of three equations for the three parameters λ, α and β. The third equation yields

$$\lambda = \frac{m - z + r\cos\alpha}{z'} \tag{4.4}$$

(4.4) changes the first two equations to

$$x + \frac{m - z + r\cos\alpha}{z'}x' = r\sin\alpha\cos\beta \tag{4.5}$$

$$y + \frac{m - z + r\cos\alpha}{z'}y' = r\sin\alpha\sin\beta \tag{4.6}$$

Dividing (4.6) by (4.5) gives an equation for β:

$$\beta = \tan^{-1} \frac{y + \frac{m-z+r\cos\alpha}{z'} y'}{x + \frac{m-z+r\cos\alpha}{z'} x'} \tag{4.7}$$

The parameter α can now be determined by adding (4.5) and (4.6) squared

$$r^2 \sin^2\alpha = \left(x + \frac{m-z}{z'} x' + \frac{rx'}{z'} \cos\alpha \right)^2 + \left(y + \frac{m-y}{z'} y' + \frac{ry'}{z'} \cos\alpha \right)^2 \tag{4.8}$$

For convenience a and b are defined as

$$a = x + \frac{m-z}{z'} x' \tag{4.9}$$

$$b = y + \frac{m-y}{z'} y' \tag{4.10}$$

which are constant. Using $\sin^2\alpha = 1 - \cos^2\alpha$ yields a quadratic equation in $\cos\alpha$

$$r^2 \left(1 - \cos^2\alpha\right) = a^2 + \frac{2rx'a}{z'} \cos\alpha + \frac{r^2 x'^2}{z'^2} \cos^2\alpha + b^2 + \frac{2ry'b}{z'} \cos\alpha + \frac{r^2 y'^2}{z'^2} \cos^2\alpha \tag{4.11}$$

which can be solved by first collecting terms

$$\cos^2\alpha \left(r^2 + \frac{r^2}{z'^2} \left(x'^2 + y'^2 \right) \right) + \cos\alpha \left(\frac{2r}{z'} \left(x'a + y'b \right) \right) + a^2 + b^2 - r^2 = 0 \tag{4.12}$$

and then solving

$$\alpha = \cos^{-1} \left(-\frac{1}{z'r} \frac{x'a + y'b}{1 + \frac{x'^2+y'^2}{z'^2}} \pm \sqrt{\frac{1}{z'^2 r^2} \left(\frac{x'a + y'b}{1 + \frac{x'^2+y'^2}{z'^2}} \right)^2 + \frac{r^2 - a^2 - b^2}{r^2 \left(1 + \frac{x'^2+y'^2}{z'^2} \right)}} \right) \tag{4.13}$$

One of the values of (4.13) corresponds to the searched track impact point. The second corresponds to the point where the track would impact, if the mirror were a true sphere and the track would cross the target to the other side. It is therefore no problem to find the correct results for λ, α and β.

4.4.2.3 Vertex Connection

The next part of the strategy (which is not yet implemented and is not part of this work) will be to create tracks from the back-tracked TPC tracks and from the tracks of the SiDC, which describe a particle's propagation through the complete detector. These tracks are called "full tracks" further on. In first order the connection is quite simple. For a given TPC track the track from the back tracking is compared to the hits in the SiDC. A hit in each one with a small distance to the average track will then indicate a full track. A "small distance" is in this case momentum dependent and should allow for variations by multiple scattering. The performance of the matching will therefore be dependent on the momentum. If no multiple scattering would be present the deviation should be of the order of the resolution of the SiDCs. This implies a implicit rejection, because all tracks which originate after the SiDCs like conversions in the RICH are not able to achieve a full track status. However, these tracks do not carry physical meaning but are only artifacts from the measurement. Therefore no physically important information is lost. At the same time a full track requirement aids in the rejection of background which originates from the detector.

It is more problematic that this introduces an implicit momentum cut due to multiple scattering. However, this problem is always present for the CERES experiment, for the new and the present strategy, and has to be taken into account.

4.4.2.4 Track-Dependent PID

For each track PID can be performed. It is technically not necessary to restrict the PID to full tracks, also averaged tracks from back tracking can be used. Since average tracks, which are not full tracks, are mostly not of physically relevant origin, from here on only full tracks are considered and will be referred to as tracks.

Each track can be treated by two PID detectors.

- The TPC can use the measurement of dE/dx to identify particles [30]. A good separation between electrons/positrons and hadrons can be achieved for low momenta, but at higher momenta this is not possible anymore. The TPC is the only detector in CERES to allow at least in some momentum ranges a separation between the hadron species: Pions, kaons and protons. However, this will be not be examined in any detail here.
- The second PID is provided by the RICHes. To weaken the limitations resulting from (3.35), it is natural to combine both detectors to one effective detector without the magnetic field to double the photon statistics. Possible PID methods with their advantages and disadvantages will be described in more detail in the next section.

Both detectors can assign a particle type to each track. There are many ways to combine them, as described in the previous chapter, but this is not yet under investigation.

4.4.2.5 Rejection

The last part of the new reconstruction strategy is the rejection, which is already more or less part of the physics analysis. All methods already developed for the current strategy can be deployed again, and hence well known approaches can be used.

New rejection strategies may be possible, especially if the present and the new strategy are combined.

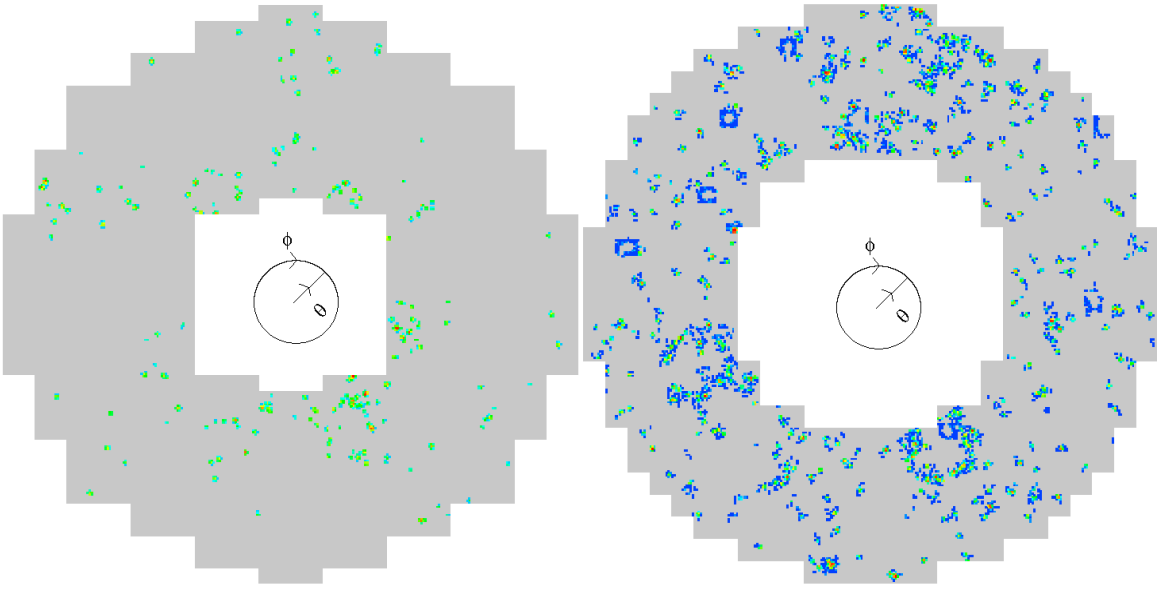
4.5 PID Methods

This section describes four possible PID methods for the CERES double RICH detector system. It shows the results obtained with two of these methods using samples of Monte Carlo data, because the described strategy has not yet been implemented in the data production. Two kinds of Monte Carlo data are used: The full Monte Carlo simulation including background and a Dalitz decay Monte Carlo with only two tracks to study overlap effects. A typical event display is shown in figure 4.14. For the methods theoretical limits in terms of efficiency and purity are found by numerically optimizing the available cuts in an automated learning fashion. It should be noted that the initial direction and momentum is used for each track and therefore the tracking performance is less than expected from a tracking implemented as described above. These results are very preliminary.

The presented methods are feasibility tests, in order to demonstrate the possible performance for a track-dependent PID at CERES. It shall be noted that no rejection is included; all results quoted describe the plain PID performance of the algorithms.

4.5.1 Data Preparation

All algorithms add the response from both RICH detectors to increase the photon statistics. Additionally it is necessary to calculate the position, where each track would impact the detector plane, if it would have been reflected by the mirror: The virtual track hit point. Both parts of the data preparation are common to all methods and are described here.



4.14. A typical (Monte Carlo) RICH event. The left panel shows RICH 1 and the right panel RICH 2. Each non-gray dot corresponds to an non-zero amplitude. The radial direction corresponds to the theta angle and ranges from 8 to 15 degrees. The detector is symmetric about the azimuthal direction. This is from an older version of the CERES GEANT simulation than the acutally used one below. The newer version has less hits in RICH 2.

4.5.1.1 RICH Summation

The RICH detectors are not summed up on a per pixel basis. Indeed all hits from each detector maintain their individuality and are all placed in a projective plane as are the tracks. The transformation is simple, but already treats all differences in the scales of both RICH detectors correctly. For each pixel it is known in which direction in spherical coordinates a photon must travel to impact on it. Given this direction as θ_{Pix} and ϕ_{Pix}, the cartesian coordinates of a pixel are given by

$$\vec{r}_{Pix} = \begin{pmatrix} x_{Pix} \\ y_{Pix} \end{pmatrix} = \begin{pmatrix} \tan\theta_{Pix} \sin\phi_{Pix} \\ \tan\theta_{Pix} \cos\phi_{Pix} \end{pmatrix} \tag{4.14}$$

The units are arbitrary and any scaling factor could be added. For convenience this factor was set to 1. The size of a pixel in this plane and in this units is about 0.002 which corresponds to an average angular size of 2 mrad. The amplitude of each hit is unchanged. In the projective plane it is possible for two hits to occupy the same spatial position.

4.5.1.2 Track Transfer

The projection of the tracks is equally simple. Given the momentum of a track $\vec{p}$ with components p_x, p_y and p_z, its position in the plane is given by

$$\vec{r}_t = \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \begin{cases} x < 0 : \pi + \tan^{-1}\frac{p_x}{p_y} \\ x \geq 0 : \tan^{-1}\frac{p_x}{p_y} \end{cases} \\ \tan^{-1}\frac{\sqrt{p_x^2+p_y^2}}{p_z} \end{pmatrix} \tag{4.15}$$

The stepwise function is necessary because of the definition of momentum in the present CERES coordinate system used in the reconstruction code. Again an arbitrary scaling factor was set to 1.

4.5.2 IRS

The Indirect Ring Search (IRS) is based on an angular Hough transformation. The Čerenkov angles are extracted using an indirect ray tracing method. Given the virtual track hits and the hit distribution in the projective plane, the following algorithm was developed.

For the angular Hough transformation, a histogram in radius is generated with a number of bins n_h and bins k_h. The size of each bin is one pixel. The number of bins n_h has been chosen to be 25 as the best value after experimentation with several sizes. An asymptotic ring in the projective plane has a size of 15.63 pixels for a given threshold $\gamma_{th} = 32$ and an asymptotic ring radius of 31.25 mrad. To fill the histogram the following simplified IRT was used. The simplifications were possible due to the good optical properties of the CERES RICH.

For each hit i and each track j their mutual distance d is calculated by

$$d = \sqrt{\left(x_t^j - x_{Pix}^i\right)^2 + \left(y_t^j - y_{Pix}^i\right)^2} \tag{4.16}$$

The histogram entry is then given by the floor function

$$k_h = \left\lfloor \frac{d}{\sigma_{Pix}} \right\rfloor \tag{4.17}$$

where σ_{Pix} is the pixel size. In k_h the amplitude a^i of each hit i will be entered. After all hits for a given track have been entered or discarded, if they are too far away, a peak is searched for in the histogram. This is done by a center of gravity fit with a sliding window. For each three consecutive bins the center-of-gravity and the sum of the bins are calculated, yielding a radius r_{kh} and an amplitude a_{kh}. The radius with the largest a_{kh} is selected as being the Čerenkov radius $r_{\check{C}}$ of this track. The largest average amplitude of these three bins a_{kh} is called ringsum a_h and should be of the order of

$$\langle a \rangle = N_{Ph} \langle a_k \rangle \tag{4.18}$$

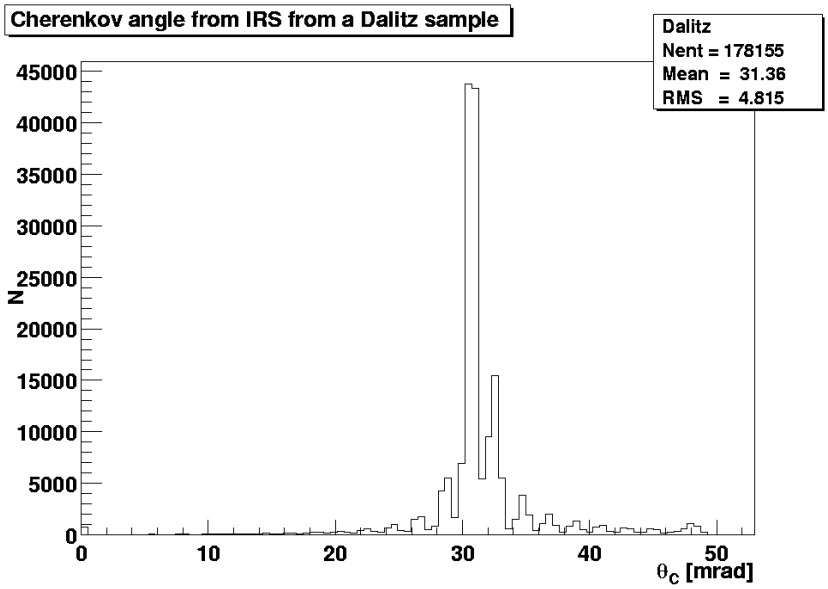
for an asymptotic ring where $\langle a_k \rangle$ is the average amplitude of a photon signal deposited on all pixels. This quantity can be used for cuts, which are effective cuts on N_{Ph}. The radius $r_{\check{C}}$ is the second observable. Using the momentum as the third observable, both can be combined to yield the Čerenkov angle and the mass squared by

$$\theta_{\check{C}}^j = \tan^{-1} \frac{r_{\check{C}}^j}{r_{Fix}} \tag{4.19}$$

$$\beta^j = \frac{1}{n \cos \theta_{\check{C}}^j} \tag{4.20}$$

$$m_j^2 = p_j \left(\beta^{j2} - \beta^{j4}\right) \tag{4.21}$$

r_{Fix} sets the scale of the asymptotic Čerenkov angles. All changes in the refraction index or other reasons for changes are contained in this factor. The rest of the calculation is then approximately independent of the true asymptotic values. The distribution of Čerenkov angle for a Dalitz sample is shown in figure 4.15. Note that the mass squared may be negative, if the Čerenkov radius is larger than the asymptotic value. This can easily happen because of scattering, especially if the mass of the particle is small. Altogether IRS yields three observables, a_h, $r_{\check{C}}$ and m_j^2. By cuts on these quantities particle species may be selected.



4.15. Angle distribution for a Monte Carlo Dalitz sample from IRS. The expected angle is 31.25 mrad. The peak structre originates from the bins.

4.5.3 eIRS

IRS is a purely trackwise method, which is problematic if overlapping rings are common. This is the case for Dalitz and conversions in CERES, and the effect of an eventwise treatment would be of some importance. A simple eventwise extension of IRS is eventwise IRS (eIRS), given by the first iterative method described in the previous chapter. If for each hit i it is kept in which histogram it occupied which bin, the total contents of theses bin can be calculated as

$$a_w^i = \sum_{\text{all tracks } j} a_{kh}^{ij} \tag{4.22}$$

where a_{kh}^{ij} is the content of the bin in which hit i is located in the histogram of track j. Then the bin content of each histogram is recomputed where the total bin content is then given by

$$a_{kh}^{j} = \sum_{\text{all hits } i \text{ in bin } k} \frac{a^i}{a_w^i} \tag{4.23}$$

These modified histograms are then used to extract a_h, $r_{\check{C}}$ and m. This weighting can be iterated an arbitrary number of times.

4.5.4 DRS

The direct ring search (DRS) is based on a modified direct raytracing approach. It extracts observables for each possible mass hypothesis k; in the case of CERES these are electron and pion. Basically DRS looks where a ring is to be expected and creates a ringsum from what it finds.

For each track j it is possible to calculate the expected radius for each hypothesis k

$$\beta^{jk} = \frac{p_j}{\sqrt{p_j^2 + m_k^2}} \tag{4.24}$$

$$\theta_{\check{C}}^{jk} = \cos^{-1}\frac{1}{\beta^{jk}n} \tag{4.25}$$

$$r_{\check{C}}^{jk} = r_{Fix}\tan\theta_{\check{C}}^{jk} \tag{4.26}$$

Again r_{Fix} is used to compensate for small changes in the asymptotic values. The ringsum b_k^j is now calculated by summing up all hit amplitudes for all hits $\aleph$ satisfying

$$r_{\check{C}}^{jk} - \sigma_{DRS} \leq \sqrt{\left(x_t^j - x_{Pix}^i\right)^2 + \left(y_t^j - y_{Pix}^i\right)^2} \leq r_{\check{C}}^{jk} + \sigma_{DRS} \tag{4.27}$$

σ_{DRS} is the size of the DRS window. A value of one pixel for σ_{DRS} has turned out to be quite acceptable, since photon signals are spread Gaussian over two to three pixels. The contribution b^{ij} of each hit to the ringsum is weighted by a Gaussian as

$$b_k^{ij} = a^j \exp\left(-\frac{\left(\sqrt{x_{Pix}^{j2} + y_{Pix}^{j2}} - r_{\check{C}}^{jk}\right)}{\Sigma_{DRS}}\right) \tag{4.28}$$

where Σ_{DRS} describes the half width of the photon spread. Its value is one pixel. The reason for this weight is to enhance the importance of the hits on the true radius. These will contribute strongest. The ringsum is then found by

$$b_k^j = \sum_{i\in\aleph} b_k^{ij} \tag{4.29}$$

This result is then normalized by $r_{\check{C}}^{jk}$ to yield a likelihood for each hypothesis

$$\ell_k^j = \frac{b_k^j}{r_{\check{C}}^{jk}} \tag{4.30}$$

This seems to be phenomenologically better than what is expected for the true behavior of $\left(r_{\check{C}}^{jk}\right)^3$. The latter would be expected using the fact that the ring diameter is constant and the Čerenkov angles are small for CERES: The hit density on a ring scales by virtue of (2.9) and (3.12) as

$$N_R = AN_{Ph} \sim \tan\theta_{\check{C}}\sin^2\theta_{\check{C}} \approx \theta_{\check{C}}^3 \tag{4.31}$$

The scaling allows the comparison of rings of any radius. The particle is then identified by a simple mechanism. If none of both likelihoods of each hypothesis is above a background value, the particle cannot be identified. If only one is higher than the background value, its type is the result. If both are higher, the larger one gives the particle type. If both are above the unidentified threshold and equal, it cannot be identified. Particles which cannot be identified are assumed to be non-radiating particles.

4.5.5 A Likelihood Method

Combining both IRS and DRS, another simple likelihood method can be constructed. Using DRS to calculate the ringsum, the number of measured photons for each mass hypothesis k for each track j can be calculated by

$$N_{jk} = \frac{b_k^j}{\langle a_k\rangle} \tag{4.32}$$

assuming $\langle a_k\rangle$ is constant, which is likely to be the case, since the detector has a well defined response to photons. The value of $\langle a_k\rangle$ is around 60, corresponding to 1200 for b_k^j if an average photon statistic of 20 for an asymptotic electron is assumed. For each hit and each track the Čerenkov angle θ_{ji} can be

calculated under which the corresponding photon would have been emitted. This is combined for all hits within the asymptotic radius plus σ_{DRS} together with (4.32) to a likelihood as

$$\ell_k^j = (N_{Phk})^{N_{jk}} \frac{e^{-N_{Phk}}}{N_{jk}!} \sum_{\aleph} \exp\left(-\frac{\left(\theta_{\check{C}}^{jk} - \theta_{ji}\right)^2}{\Sigma_{DRS}} \right) \tag{4.33}$$

where σ_{Cut} is a upper cutoff to limit the range where hits can influence the likelihood. N_{Phk} is the number of photons expected for this hypothesis, which is given by

$$N_{Phk} = N_{APh} \frac{\sin^2 \theta_{\check{C}}^{jk}}{\sin^2 \theta_{A\check{C}}} \tag{4.34}$$

and N_{APh} is the asymptotic photon yield, with a value of 20. Note that the faculty in (4.33) is the continuum faculty, since N_{jk} does not need to be an integer.

4.5.6 Acceptance Correction

A significant problem for CERES is the acceptance. If rings only would be accepted within the fiducial acceptance, i.e. where the whole ring is contained inside the RICH detectors, this would narrow the acceptance from [0.14 mrad, 0.26 mrad] to [0.17 mrad, 0.23 mrad] which would be a loss of 50% of the detector area. It is therefore necessary to correct for the edge of the acceptance. A possibility to include tracks up to the edge of the detector are "shadow hits", which are an extrapolation from a found ring segments at the limit of the acceptance. The correction is a factor which can be used to correct IRS, eIRS and DRS ringsums and N_{jk}. Since the radius is different, different degrees of approximations are sufficient for the inner and outer edge. The correction factor is calculated and applied for IRS and eIRS for each single bin separately, for DRS and for the likelihood method for each hypothesis. The factor is only applied if the bin or radius of the hypothesis is indeed describing a ring which is cut by the acceptance. This can be determined by checking the radial distance to the outer and inner edge e of the ring, which is given by

$$e_{o/i} = \pm r + \sqrt{x_t^2 + y_t^2} \tag{4.35}$$

where r is the radius of the bin or hypothesis.

Outer Edge. The geometry is shown in figure 4.16. For the outer edge the limit of $r_0 \to \infty$ is used. The angle θ gives a measure for the amount of cutoff. The correction factor f can then be calculated as

$$\theta = \cos^{-1} \frac{r}{r_C} \tag{4.36}$$

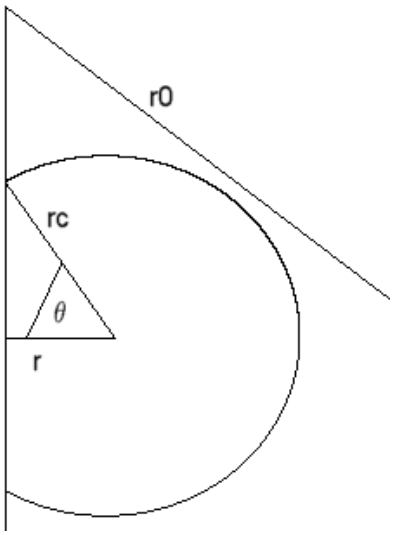
$$f = \frac{1}{1 - \frac{\theta}{\pi}} \tag{4.37}$$

Inner Edge. At the inner edge this limit is no longer valid and the curvature has to be taken into account. Vector algebra delivers the shadow angle θ as

$$\theta = \cos^{-1} \frac{(r + r_0)^2 + r_C^2 - r_0^2}{2 r_C (r + r_0)} \tag{4.38}$$

and the correction factor can be calculated analog

$$f = \frac{1}{1 - \frac{\theta}{\pi}} \tag{4.39}$$



4.16. Geometry for the acceptance correction

4.5.7 Ringness

An important property for a RICH ring is that it is a ring. As simple as this sounds, none of the methods presented above respects this feature. All project the two-dimensional geometric information upon a one-dimensional information, the radius or angle. To obtain information on how well the rings conform to a circular shape, the ringness can be defined. Since RICH rings are likely fragmented, it is not possible to require a complete ringness. This would be to require that the hit density is constant for arbitrary small areas of the ring.

A first approximation to this ideal ringness test is the monopole approximation. In this case the ringness is measured by the distance of the virtual track hit and the center of gravity of the ring. For an ideal ring the distance will be zero. The ring's center of gravity is calculated as

$$\vec{r}_{CoG} = \frac{\sum\limits_{j \text{ satisfying (4.27)}} a^j \vec{r}_j}{\sum\limits_{j \text{ satisfying (4.27)}} a^j} \tag{4.40}$$

where $\vec{r}_j$ is measured with respect to the virtual track hit point. The ringness in monopole approximation r_{ma} is then given by

$$r_{ma} = |\vec{r}_{CoG}| \tag{4.41}$$

This is an additional observable for all of the methods mentioned above.

4.5.7.1 Acceptance correction

The ringness calculation is again influenced by the acceptance cutoff. To calculate at least an approximate ringness, a virtual hit is introduced. The amplitude of this hit is given with f for the outer edge by (4.37) and for the inner edge by (4.39) by

$$a^{VH} = A^{VH} \cdot f - A^{VH} \tag{4.42}$$

$$A^{VH} = \sum_{j \text{ satisfying (4.27)}} a^j \tag{4.43}$$

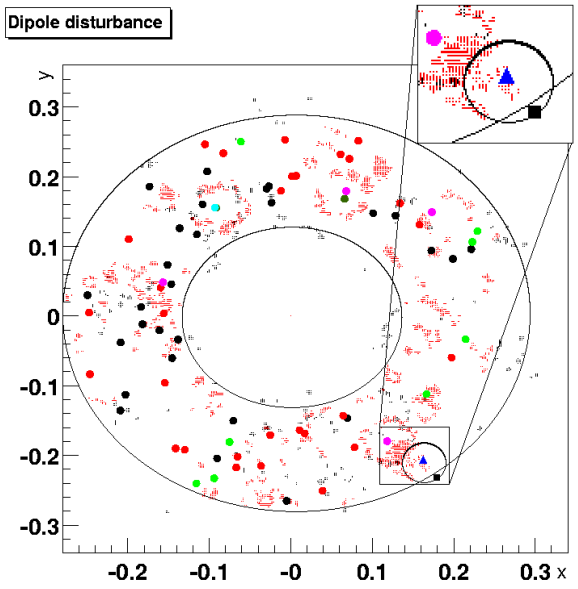
The virtual hit is positioned at the center of gravity of the missing part of the ring. The appropriate position for the virtual hit with respect to the track position is

$$R_{VH} = \sqrt{x_t^{i2} + y_t^{i2}} \tag{4.44}$$

$$\vec{r}_{VH} = \mathrm{sgn}\,(r_0 - R_{VH})\, R_{VH} \frac{\sin\theta}{\theta} \begin{pmatrix} \sin\left(\tan^{-1}\frac{x_t^i}{y_t^i}\right) \\ \cos\left(\tan^{-1}\frac{x_t^i}{y_t^i}\right) \end{pmatrix} \tag{4.45}$$

where r_0 is defined as above, that is infinite for the outer edge and as the inner radius of the acceptance for the inner edge. θ is given for the outer edge by (4.36) and for the inner edge by (4.38).

4.5.7.2 Resulting Problems



4.17. Dipole disturbance in ringness calculation. The triangle is the hadron track and circles are electrons. The black ring sketches the ring caused by dipole contributions. The black square is the position of the virtual hit.The common acceptance of RICH 1 and RICH 2 is sketched by the large black concentric circles. The color code for the electron momentum is the same as in figure 4.25. The units for x and y are arbitrary as described in subsection 4.5.1.

The ringness introduced is only an approximation. It is called monopole approximation because two large and equal hits, a dipole configuration, on the opposite site of the virtual track hit will also yield a zero ringness in monopole approximation. This situation is unlikely and has not been observed. A larger problem are dipoles induced by virtual hits.

Figure 4.17 displays the situation. A hadron track lies between a ring and the acceptance at half distance. The virtual hit of the acceptance correction of equation is placed by (4.45) such that a dipole configuration emerges. The dipole is made of the electron ring towards the inner edge and this virtual hit, which is outside the acceptance. The result is that the track is considered to have perfect ringness. This indicates on the one hand, that the principle works, but is on the other hand a contamination unremovable with the available observables. Monte Carlo studies indicate that the fraction of the contamination at a 20% efficiency level of IRS for a momentum above 1 GeV/c of this type of error is below 5%, so that the

importance is small.

4.5.8 Performance

To study the performance of the algorithm, a full Monte Carlo sample of 3336 events has been used. It contained about 750000 tracks. The track information for Monte Carlo data consists of the original momenta and the origin of the particle. Due to multiple scattering in the target and the silicon, the direction of the particle is severely changed for low momenta. Therefore it turned out not to be useful to study low momenta, and only particles with more than 100 MeV/c linear momentum are included. The particles lost are mainly electrons from conversions. The same problem will be weakened with the final tracking, since it will measure the final momentum. Then the only large scattering which changes the direction is the RICH 2 mirror, which can be solved as described before, while the scattering at the target has no importance, since it is essentially point-like as seen from the TPC.

To simulate a full track, only tracks which originate at z smaller than 10 cm, that is in the vicinity of the target, have been accepted. Only tracks within the acceptance of 8° to 15° degrees have been taken into account. No further cuts were applied. The Monte Carlo is the newest GEANT Monte Carlo available for CERES. It includes all background effects and should estimate the noise and background approximately correct. However, the simulated asymptotic radius is 10% to 20% of a pixel too small[13]. This deviation has been included in the r_{Fix} parameter.

4.5.8.1 IRS and DRS

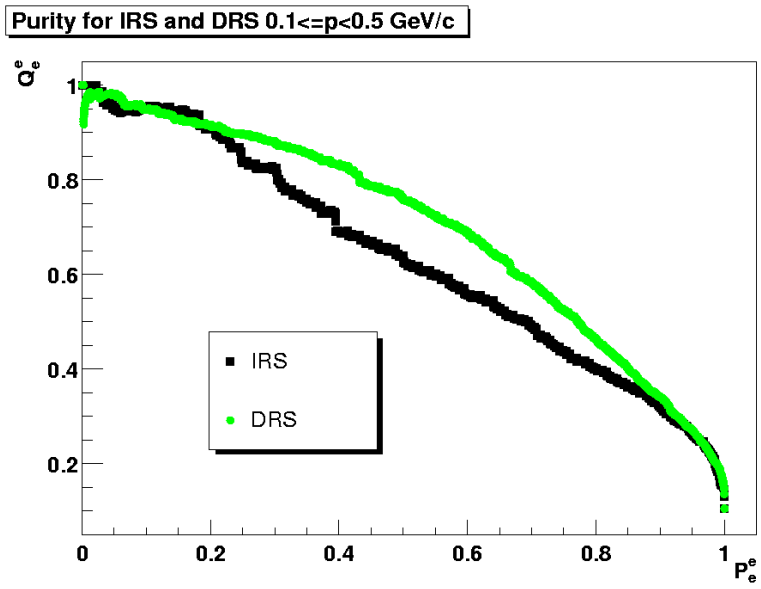
The methods IRS and DRS are the main reconstruction methods, and the likelihood and eIRS methods are derivations of them. Hence IRS and DRS will be presented first. Since no finalization with respect to a quality can be made, PQ-diagrams as defined in chapter three are used to visualize the performance of the algorithms. $1 - Q$ is then again the contamination of electrons by any other particle species.

The interesting behavior is the dependency on the efficiency of the purity for electrons, if the cuts on the available variables are changed. For IRS these are a_h, $r_{\check{C}}$ and m_j^2 and for DRS ℓ_k^j. Both methods have additional cuts on the ringness r_{ma}. The cuts have been varied using a feedback system to choose an optimal set of cuts. I.e. cuts have been tested using the Monte Carlo knowledge of the true particle type and those cuts have been chosen which optimized the result. A comparison with a selection by hand showed that the qualitative features are the same and optimal cuts are the following: Cuts on the ringness and for IRS on $r_{\check{C}}$ and m_j^2 allow to remove a large amount of hadronic background, reducing it by one order of magnitude. This is the steep ascent at the highest efficiency or signal value. Continuous cuts on all parameters allow then a slow further increase. To allow for the removal of nearly all background tracks, a_h and ℓ_k^j [14] have turned out to be the most important. At very high purity values it becomes increasingly difficult to remove remaining contaminations. The reasons for these problems will be detailed later. At a level of 20% efficiency at more than 1 GeV/c it is possible to reduce the initial hadronic background by a factor of 10^4. The ratio of surviving electrons to background hadrons can be increased at the same time by only a factor of about 10^3. These values decrease somewhat as the momentum is decreased, but stay in the same order of magnitude. For the physics case the most interesting particles are electrons above 1 GeV/c. In this case the initial ratio of signal electrons and background hadrons is of the order of 10^{-3}, so that at a level of 20% efficiency the ratio is about one, actually a little better.

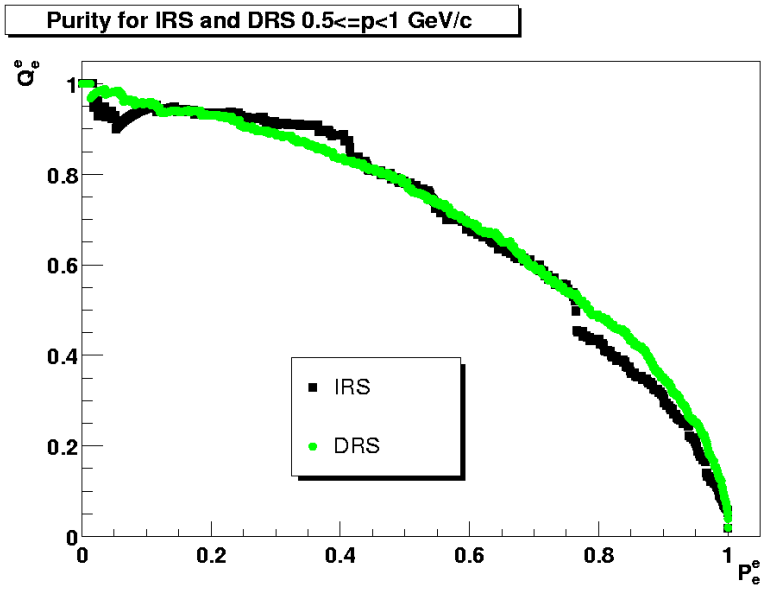
Figure 4.18 is the PQ-plot for electrons between 100 MeV/c and 500 MeV/c. For IRS after the steep start the behavior is nearly linear and then flattens out. DRS behaves better in this momentum range. Since scattering is still an issue at this momentum, it seems that the approach of DRS is better suited. The reason for this is that if the virtual track hit is displaced against the center of the circle, the angular Hough transformation will yield two peaks, both of small height at too low and too high radius. It then

[13] This has been measured with IRS and agrees with what is known from the GEANT properties

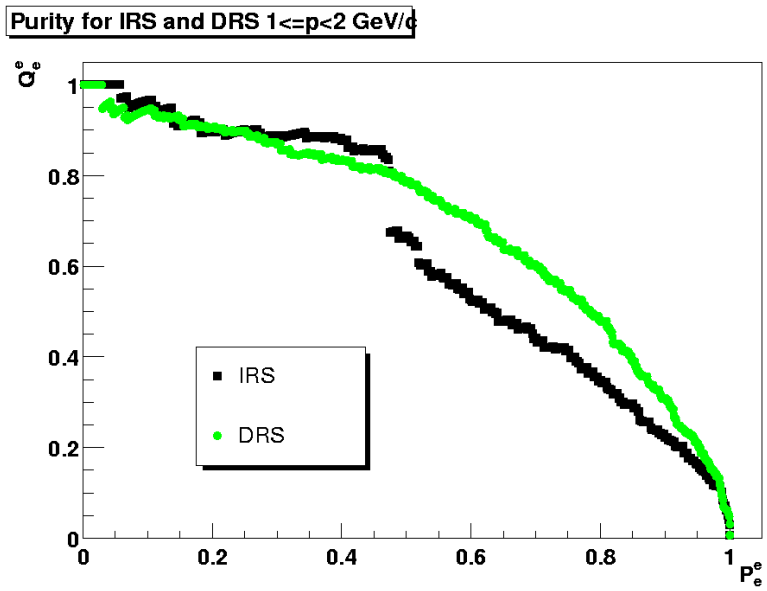
[14] Or the background likelihood, which is equivalent.



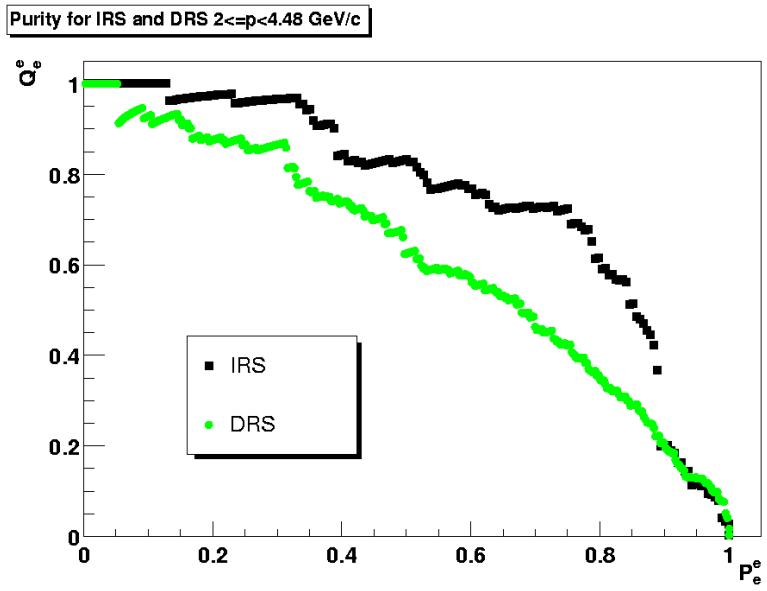
4.18. Comparison of IRS and DRS between 0.1 GeV/c and 0.5 GeV/c for electrons



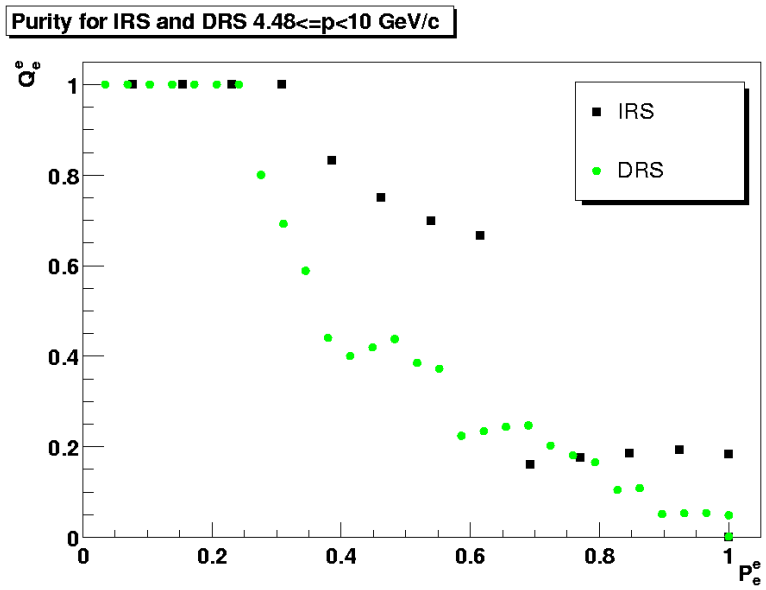
4.19. Comparison of IRS and DRS between 0.5 GeV/c and 1 GeV/c for electrons



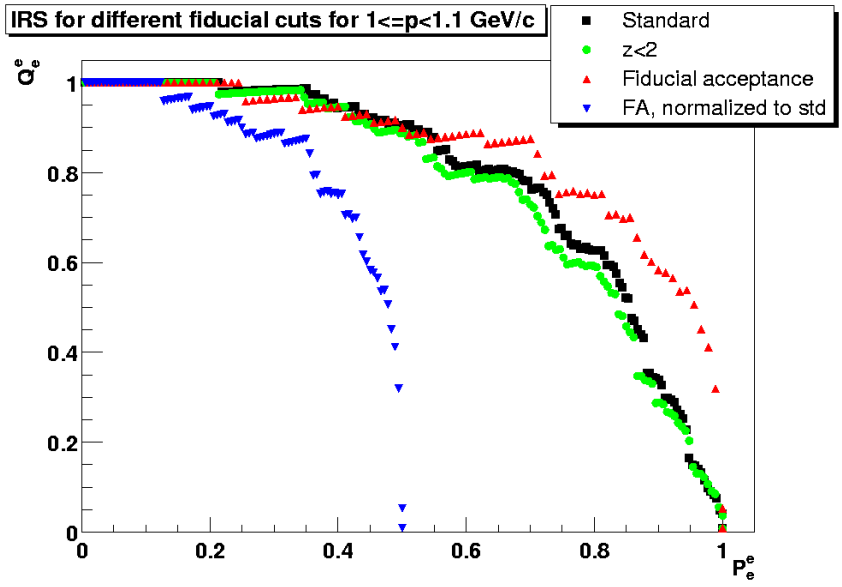
4.20. Comparison of IRS and DRS between 1 GeV/c and 2 GeV/c for electrons



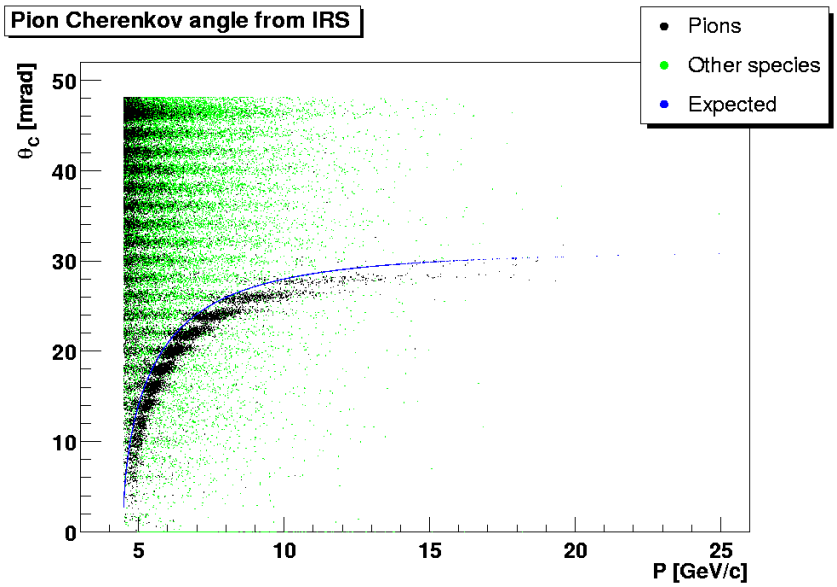
4.21. Comparison of IRS and DRS between 2 GeV/c and 4.48 GeV/c for electrons



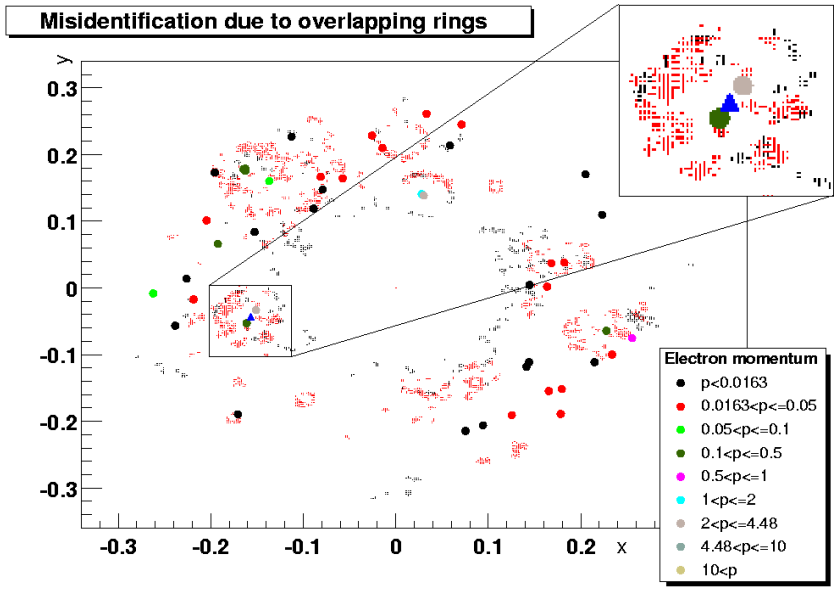
4.22. Comparison of IRS and DRS between 4.48 GeV/c and 10 GeV/c for electrons



4.23. Comparison of IRS, DRS and different fiducial cuts



4.24. Pion Čerenkov angle from IRS. The light dots are non-pionic background.



4.25. Remaining contaminations due to overlapping rings. Triangles are hadrons, circles are electrons. Red dots are hits from RICH 2 and black dots are hits from RICH 1. Note that red dots may shadow black dots. The units for x and y are arbitrary as described in subsection 4.5.1.

becomes increasingly difficult to determine the true radius by a simple center of gravity fit. For DRS only the total ringsum decreases, but since the scattering is not too strong, this effect is smaller than that for IRS.

Figure 4.19 is the PQ-plot between 500 MeV/c and 1 GeV/c. Both IRS and DRS show roughly the same behavior, since the importance of the scattering is now strongly reduced. The DRS behavior is much smoother because the observables are more strongly correlated. A track with high likelihood has most probably also a good ringness, because the hits are more likely evenly spaced. For IRS the observables are not as strongly correlated, and some cuts may spontaneously eliminate large parts of the background.

Figure 4.20 is the PQ-plot for electrons between 1 and 2 GeV/c. DRS does not change too much, but IRS exhibits a large jump. The reason for this step, which has already been seen in 4.19 in some rudimentary form at about 0.4, is due to the m^2 distribution of the tracks. True electron tracks show a double peak at the electron mass and a smaller one at the negative electron mass squared, while hadron tracks fill the mass squared spectrum continuously. If the cut is jumping over this gap a large amount of background tracks are removed. By this the smaller peak is lost, the reason why it happens only at about 0.4. Because of scattering both peaks are smeared to one at lower momenta even for electrons, and no step is observed.

Figure 4.21 shows the PQ-plot between 2 GeV/c and the pion radiation threshold of 4.48 GeV/c. In this regime IRS is better than DRS, and the jump of IRS can be found at higher efficiencies. At these momenta the rings are better formed and the ability to use four against two variables to cut is an advantage. The loss of purity at high values is a statistical artifact. In this regime, the number of contaminating hadrons is of the order of one. While the cuts are always stronger, removing electrons, single hadrons can survive longer, and by virtue of equation (3.40) the purity then starts to worsen. For higher statistics the curves would be monotonous.

Figure 4.22 shows the behavior between 4.48 GeV/c and 10 GeV/c for electrons. Although DRS can now remove about 50% of the pions beforehand because their likelihood is larger, it is now even more inferior to IRS, which can, at these momenta, fully profit from its double number of control parameters. However, statistics is small, as can bee seen by the small number of data points, and the results are not necessarily very reliable.

Figure 4.23 shows a PQ-plot to compare different fiducial cuts for IRS in the momentum range of 1 GeV/c to 1.1 GeV/c. In this range DRS behaves very similar to IRS. A change of the requirement of the track to emerge from the target, i.e. below 2 cm, does not significantly affect the behavior. To reduce the acceptance to the fiducial area of the detector gives rise to an improvement at high efficiency. This is as expected. As good as the acceptance correction is, it is still an extrapolation, which works in general, but is likely to fail in some cases, as has been demonstrated above for the ringness. However, for large purities, the effect of the fiducial cut is even weaker than the difference between DRS and IRS, and therefore the acceptance correction does a good job, especially if it is considered that the fiducial cut reduces the acceptance by 50%. The normalization of the efficiency to the full acceptance is also shown in the figure 4.23 and denoted as normalized to standard. This signifies clearly the advantage of the analytical acceptance correction.

For pions, the situation is easier, since the pions' flux dominates all species above 4.48 GeV/c, where they start to radiate. Without any cuts at 100% efficiency already a purity of 37% exists. DRS is able to improve this to 48% at a level of 91% efficiency or to 79% at a level of 52%. IRS produces the familiar plot of Čerenkov angle against momentum, but with large background at low momenta as can be seen in figure 4.24. It provides 40% purity at 98% efficiency, but decreases for 75% purity to 37% efficiency.

Looking at a 20% level of electron efficiency, the remaining contaminations can be separated in two classes. The first is a topology class for the hits. By creating large noise clusters or by multiple rings, as shown for example in figure 4.25, a ring is generated for the track that does not exist but has the correct ringsum, radius and ringness. About 25% to 30% of the remaining contamination belong to this class, split equally between noise clusters and multiple rings. About one third of the noise is due to the already mentioned failure of the acceptance correction. The second class belongs to the topology of tracks and

their correlations described in the next subsection.

4.5.8.2 Track Correlations

The main problem for the remaining contaminations is the high multiplicity of non-radiating tracks in CERES. Figure 4.26 shows an event where the hits from both RICH detectors are shown and all full tracks within the acceptance are displayed. As can be seen, there are numerous hadronic and leptonic tracks.

There are now two situations, both visible in figure 4.27.

Two or more tracks, at least one of them being an electron with a ring, are within a distance of one to two pixels. In this situation all windows of IRS and DRS are large enough to grant all tracks the status of an electron with ring. The possibility of finding two electrons from a conversion together is high, but this does no harm, since both are correctly identified as electrons. A problem emerges if one is a hadron, which is then misidentified as an electron. A rough estimation to find the probability for a configuration of an electron and a hadron with the required topology is given by

$$P = \frac{\pi r^2 N_h N_{e^+e^-}}{A} \tag{4.46}$$

where r is the radius within another track should be located, $N_{e^+e^-}$ is the number of electron tracks per event, N_h is the number of hadron tracks per event and A is the total area of the detector. This does not include cases when a hadron track takes the place of an electron which has been scattered away. This latter topology is a priori unresolvable but very rare.

The average number of full track electrons with asymptotic rings that is above 50 MeV/c, is 9.2 per event in the full acceptance. The actual number of rings is slightly higher, since average tracks originating form the SiDCs are not taken into account, and neither are overlapping double rings from conversions in RICH 1. The number of hadron track per event is 197.6. Even if all radiating pions were identified correctly, this number would only slightly change downwards. This results in a probability of 17.8% to find a hadron track within one pixel radius from an electron track and 59.4 % within two pixels. This corresponds to each 6th and each 2nd event respectively. Experimentally numbers can be found by simple counting of these situations. For the case of the Monte Carlo sample the values are 16.5% and 61.8% in excellent agreement with the calculated values.

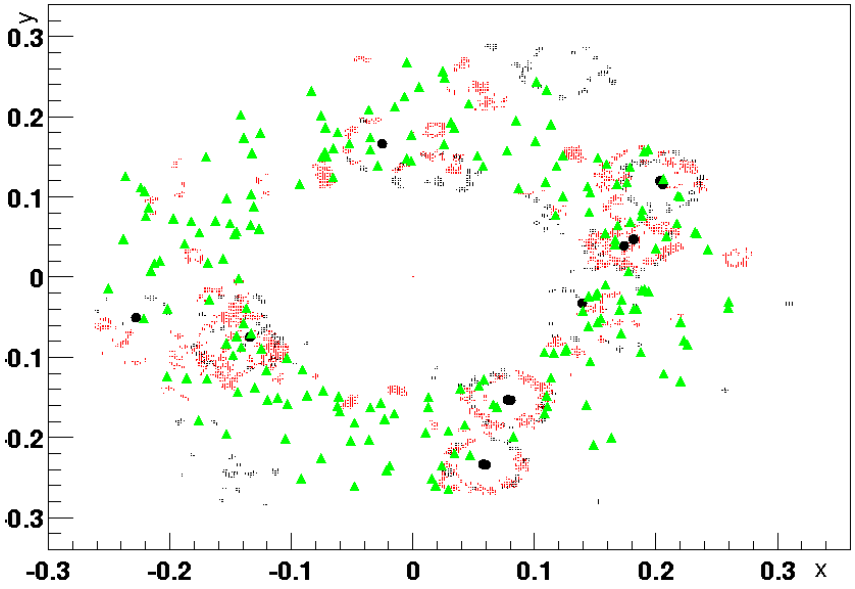
Above 1 GeV/c, the numbers are each 70th and each 20th event. At a level of 20% efficiency this situation occurs for IRS only for each 150th and 100th event respectively, indicating that at this level already weak rings with this configuration have been removed.

A simple strategy to remove these situations would be to divide the measured ringsum by the total number of tracks which share the same virtual track hit point for this ring. Conversion electrons would then be further recognized correctly, since they should have a doubled amount of photon hits, if both radiate. In the situation of an electron and a hadron, both would have reduced ringsum, so that they will be likely to be cut away and the electron will be lost. There is a good chance to reduce the remaining contamination with this to a level which is at least a factor two or larger smaller, corresponding to a rejection power of 10^5. However, due to the numbers, the initial efficiency will be limited for 2 pixels to 94% and for one pixel to 98%.

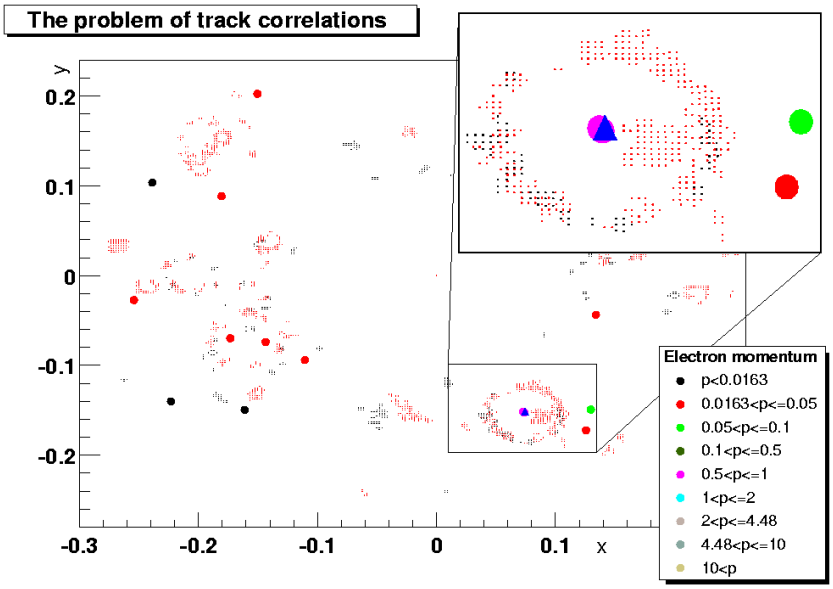
First tentative studies with Monte Carlo data indicate, that DRS is not significantly affected by this treatment at all and IRS looses too much performance, so that a more sophisticated treatment seems to be necessary.

4.5.8.3 Overlapping Rings

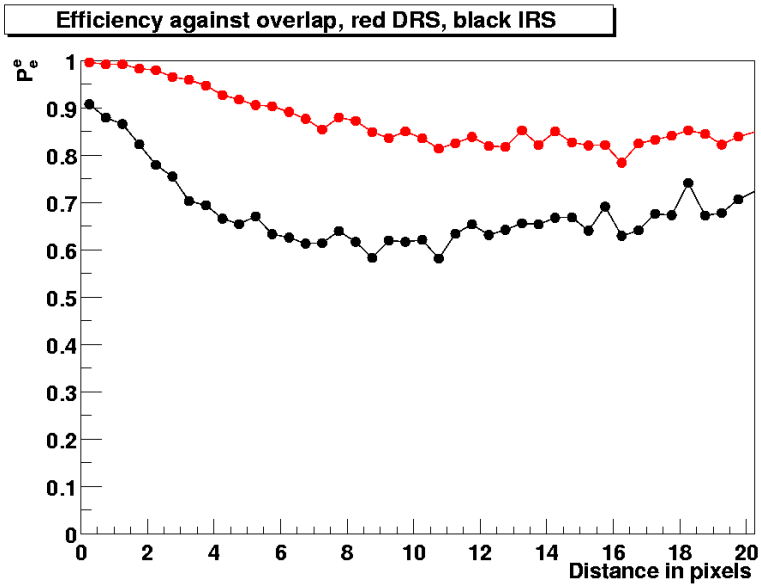
For physics purposes it is important to be able to identify overlapping rings correctly. For the present algorithm the reconstruction efficiency for rings decreases down to half the value when compared to the one of non overlapping rings, if the overlap is complete. Figure 4.28 shows the dependence for IRS and DRS against the distance in pixels between both tracks.



4.26. The problem of CERES. Dots are RICH hits, circles are electron tracks and triangles are hadron tracks. The units for x and y are arbitrary as described in subsection 4.5.1.



4.27. The problem of close tracks. The Triangle is the hadron, circles are electrons. Light dots are hits from RICH 2 and black dots are from RICH 1. The inset shows that the hadron accompanies two electrons. The units for x and y are arbitrary as described in subsection 4.5.1.



4.28. Dependence of the efficiency on the ring overlap for IRS and DRS

The results are from a Monte Carlo with only Dalitz decays and 20000 events. Additional to the cuts mentioned above, both tracks were required to be within the fiducial area, and the ringness criterium was not used. The level of efficiency was adjusted by cuts on the other parameters. By changing the cuts the curves are shifted up and downwards, but the qualitative features were not changed. No momentum cut was imposed. The asymptotic radius is 15.63 pixels, and the last bin includes all distances up to infinity. The features are that for IRS a slight decrease can be noticed at half overlap, because a second peak in the Čerenkov spectrum is then possible by the hits of the other ring. For DRS, a pion hypothesis cannot made be strong enough by these hits, and hence the efficiency remains nearly constant. As the hits of one track enter the window for the other track to be an electron, in both cases the efficiency starts to increase up to very high values, since now both tracks mutually enhance each other. This behavior is qualitatively different from the present algorithm.

4.5.8.4 Other methods

For eIRS and the likelihood method no clear results are available. While reproducible results could not yet be achieved for the likelihood method, eIRS seems to suffer from the fact that most of the tracks do not have a ring and the photon is split more or less equally, so no improvement is visible for an acceptable number of iterations. However, in principle it should work for even a large number of not radiating tracks, but the changes decrease like $1/n$ with n the number of tracks and the convergence becomes very slowly.

4.6 A Possible Optimization Strategy For The Present Method

The original starting point of this work was to find an optimized rejection strategy using the TPC. Although this has been changed to a completely new approach, the original ideas should be sketched here. The algorithm is multistage:

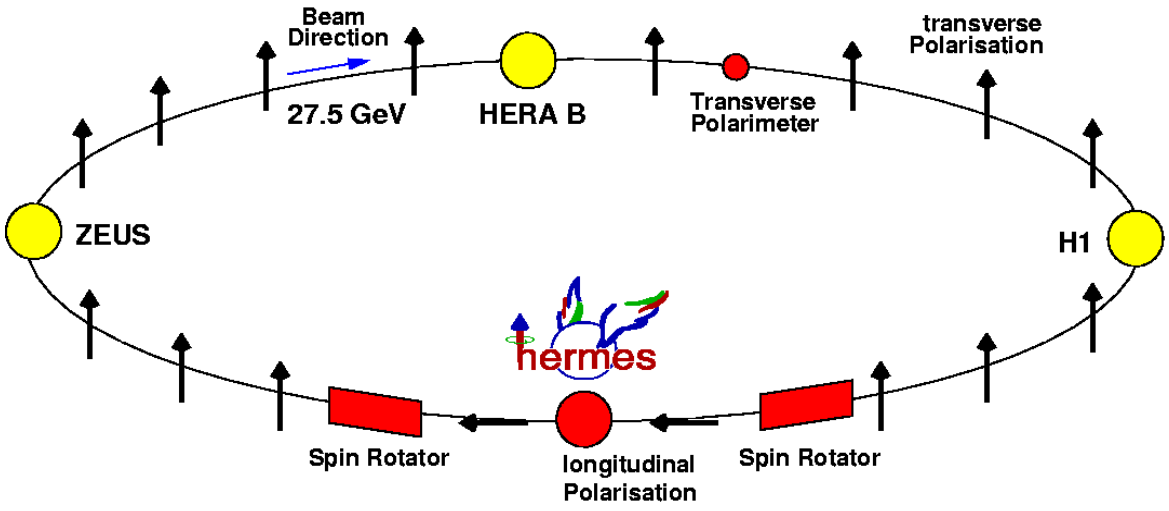
- The first step was to use the conventional reconstruction method and to operate as a post processor. Both RICHes were treated individually.

- The next step was to use the dE/dx information to select electron candidates from the TPC. These were called back tracks.
- Front tracks were constructed by the requirement that for each track a ring in RICH 1 and RICH 2 must have been found.
- Using the back-tracking described above, back and front tracks were checked if they can be connected to a primary track. Each primary track represented the postprocessor hypothesis of a single track, which has to be tested. If another electron track could be found close enough, the primary track would be classified as from a conversion or Dalitz and therefore be rejected.
- The next step is to find electron tracks in the TPC which are close enough to belong to the primary track. These were the associated tracks. A primary track could be associated to another primary track. This was done by selecting parallel tracks in the TPC and tracks with near impact points on the backside of the RICH 2 mirror.
- Additionally single rings in RICH 1 or RICH 2 are searched, which overlap with the ring of the primary track, even if they themselves do not belong to a primary track and do not have a correspondence in the other RICH detector.
- A similar approach is used to identify conversions, their typical signal being a doubled number of counts for a primary track in the sum for the rings in RICH 1 and RICH 2. If the factor was two or larger than the expected number, a conversion was marked, for a factor of 1.5 a possible conversion was marked. The latter may occur if the second conversion track was scattered away in the RICH 1 mirror.
- Additionally ring search methods should be deployed around each non-primary secondary track to identify track-dependent rings, which may have been missed earlier. If two rings for an associated non-primary track could be found or if the associated track was a primary track, it became a secondary track to the primary track it was formerly associated to. Possible implementations are IRS and DRS or the likelihood method.
- As a support a more sophisticated track-independent ring search was considered to search in the neighborhood of a primary track without associated tracks. The idea was to use the information which hits already belonged to the primary track. This was not implemented.
- From all observations a rejection factor was built, which described the level of confidence that the primary track was from a conversion or Dalitz. A secondary track or a conversion already sets the rejection factor to a high value. In other cases only multiple uncorrelated observations could achieve a high rejection factor.

The project was closed in favor of the approach described above before the inclusion of SiDC could be considered. There are only rough data estimations of the performance, but they indicated already that the method was at least worth to be pursued.

Chapter 5
HERMES

HERMES, HERa MEasurement of Spin, is an experiment studying nucleon spin physics and nuclear physics at the electron/positron beam of the HERA accelerator at DESY. The 27.56 GeV/c beam of the accelerator collides with a fixed gas target. One advantage of using a circulating electron/positron beam is the Sokolov-Ternov effect: Charged particles spin-polarize themselves by emission of synchrotron radiation due to a slightly asymmetric spin-flip probability. For the given energy and ring diameter the polarization reaches a level of up to 60% in transverse direction shortly after injection of the beam. By spin rotators this is changed to the longitudinal direction needed for the physics [3]. The situation is depicted in figure 5.29.



5.29. Spin polarization and rotation at HERA

The HERMES detector [3] is a classical forward spectrometer, with a tracking system consisting of drift chambers and a magnet for momentum measurement. It possesses components for electron/positron-hadron separation. Up to 1997 a threshold Čerenkov counter was used for pion/hadron separation. It was replaced in 1998 by a mirror-focused RICH detector, which uses two radiators in conjunction with one photon detector. The first radiator is gaseous C_4F_{10}, the second one an amorphous silica aerogel.

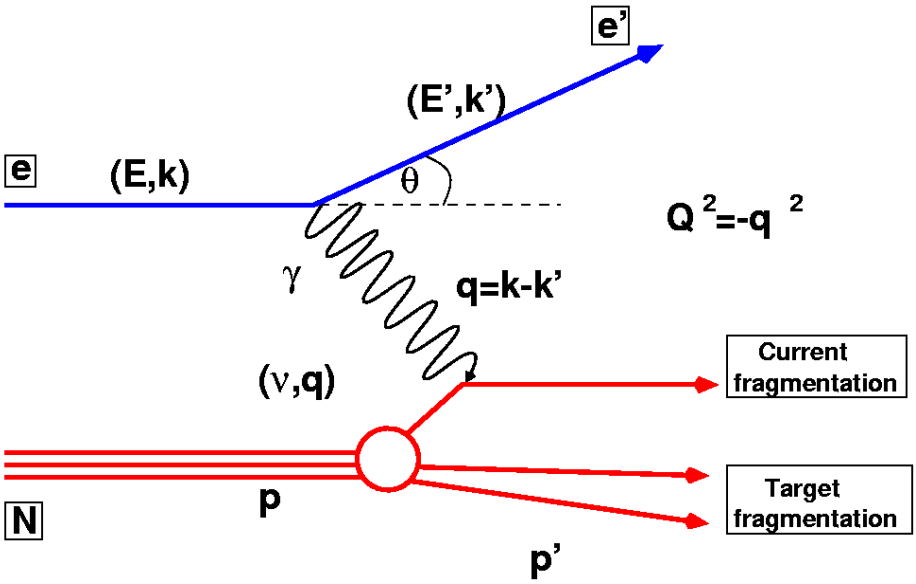
For PID with the RICH detector two methods, an indirect and a direct ray tracing method, have been implemented. A main topic of this work was the optimization of this PID by combining both methods with a decision network, as will be described in the last section of this chapter.

5.1 Physics

About two decades ago, the European Muon Collaboration (EMC) at CERN measured for the first time the contribution of the valence quarks to the nucleon spin [24]. From relativistic quark models a contribution of about 60% was expected. The measurements of EMC were even compatible, however, with no contribution of the valence quarks to the nucleon spin. This unexpected and at this time unexplainable result was termed the "spin crisis". Later and more precise measurements by the spin muon collaboration (SMC) at CERN and several experiments at SLAC have shown that the spin of the valence quarks gave a contribution of about 30% to the total nucleon spin, but this is only a fraction of the total value. The more recent understanding takes into account that further contributions are coming from the gluon and sea quarks, as well as from their orbital angular momentum. These measurements gave also indirect hints that the contribution by the strange sea quark does not vanish, although it was expected to do so, since the nucleon is a non-strange object.

HERMES started in 1995 to investigate the spin puzzle, as it was now called. HERMES measures the final states from Deep Inelastic Scattering (DIS), a process shown in figure 5.30. HERMES is the first experiment able to perform semi-inclusive measurements of polarized DIS with high statistics.

HERMES covers a wide field of particle and nuclear physics. Below covers only part of it, and briefly describes three and a few more will be listed. Further information can be found for example in [4, 16, 45, 94, 120] and references therein.

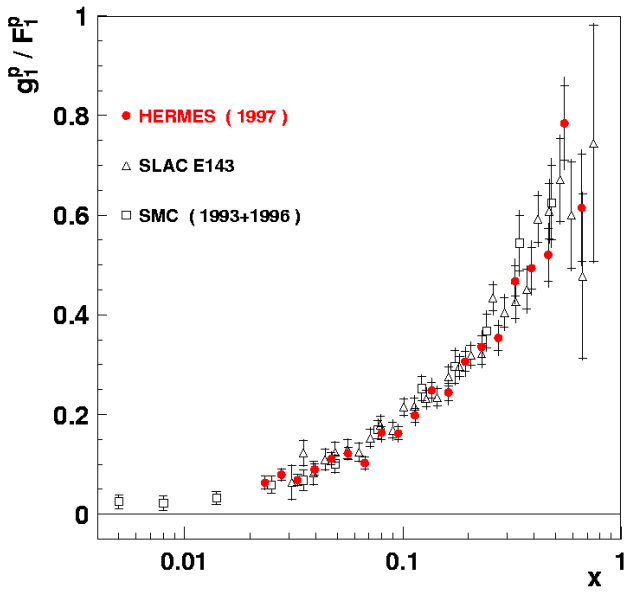


5.30. Deep inelastic scattering

Two of the main physics subject currently under investigation by HERMES are the direct measurement (from inclusive data) of the spin structure functions g_1 for the proton and neutron, as a function of Bjorken x and its decomposition (from semi-inclusive data) into the different parton contributions in form of the polarized Parton Distribution Functions (PDF). HERMES was able to increase the precision of these polarized PDFs [120] and using the RICH will further improve them and make it possible to access the strange quark polarization for the first time.

g_1 describes the spin contents of the nucleon. It is a particular combination of the polarized quark distribution functions Δu, Δd and Δs [17]. The direct sum $\Delta\Sigma$ of the Δq_f is the contribution of the

quarks to the nucleon spin. The Δq_f enter g_1 summed up weighted by their charges. An example for the measurement of g_1^p for the proton normalized to the structure function F_1^p of the proton is shown in figure 5.31. A direct measurement of g_1 is possible by measuring the asymmetry of the scattering cross-section of the lepton.



5.31. The ratio of the structure functions g1 and F1 for the proton. From [94]

To measure the individual functions Δq_f, it is necessary to perform semi-inclusive measurements. The asymmetry of the measured hadrons is directly related to the polarized quark distribution functions. For Δu and Δd lepton/hadron separation is sufficient, but for Δs kaon identification is necessary [4]. The measurement of the other structure functions also benefit from the hadron identification [94]. Semi-inclusive measurements also allow the separation of the contribution of valence and sea quarks. Therefore the RICH detector of HERMES gives access for the first time to the measurement of the individual flavor contributions [94, 120].

A first measurement was performed by HERMES of the polarized gluon distribution function ΔG although with large theoretical uncertainties.

Another topic are aspects of nuclear physics. Due to the momentum and therefore energy range covered by HERMES, it is possible to observe the process of hadron formation. If this formation happens within a nucleus, the new hadron is influenced by strong multiple scattering. This reduces the observed yield compared to production of the hadron from a free nucleon. The observed suppression yields information about the formation time of hadrons. An unexpected dependence on the nucleus mass number A was observed. This is already known as the HERMES effect [15].

Hadronic final states are also a subject studied at HERMES. The determination of total cross-sections in polarized and unpolarized DIS is pursued, for example for the exclusive production of ρ^0 with the result of a spin asymmetric cross-section [16, 94].

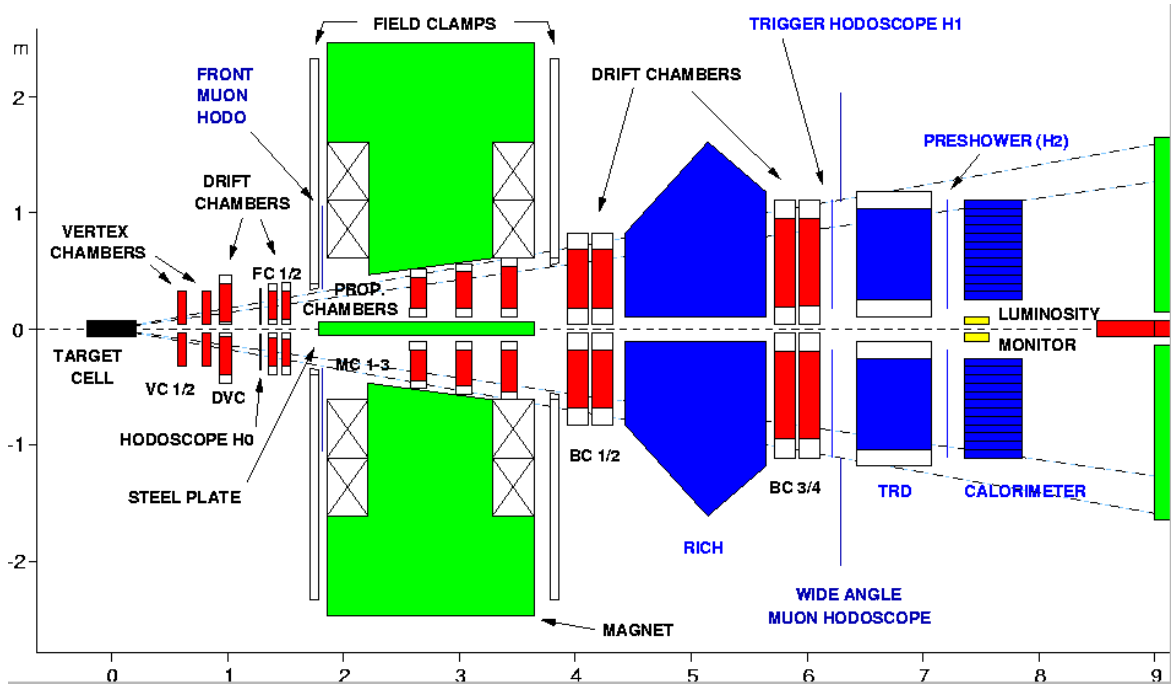
Other subjects currently under investigation are measurements of the DIS contribution to the generalized GDH integral and deeply virtual Compton scattering. The first is a well developed subject [94] while the second is relatively new.

Future measurements of azimuthal asymmetries of hadrons will allow to investigate the (polarized) Collins fragmentation functions. An important future topic will be the hitherto unexplored transverse

spin physics, and skewed parton distribution functions. These studies will be performed by HERMES during its running up to about 2006 [45]. With COMPASS at CERN starting in 2001, another experiment will investigate the physics of the nucleon spin.

Even further studies of the details of polarized nucleon/nucleus-electron scattering may be possible with the presently discussed projects eRHIC and TESLA-$\vec{N}$ [17] using a higher center of mass energy to cover a wider kinematical range and higher statistics with a higher luminosity. These experiments would provide a precision test of QCD in the spin sector.

5.2 Setup



5.32. The current setup of HERMES

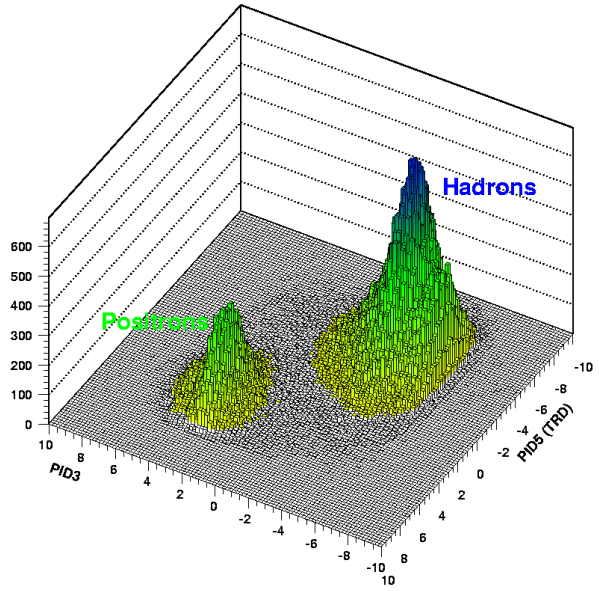
The HERMES detector is a classical forward spectrometer [3]. Several upgrades have been installed since 1995 and the current setup is shown in figure 5.32. The spectrometer has an identical top and bottom half, above and below both beam pipes of the HERA accelerator.

The target is a thin internal gas target held in a tubular open-ended storage cell that is mounted coaxially to the beam tube. For nucleon spin physics, the target can be polarized or unpolarized hydrogen or deuterium as proton targets and 3-helium as an effective neutron target. For nuclear physics also nitrogen or the noble gases up to krypton are used. Since the storage cell is cooled below 100 K down to 25 K, only gases which do not liquefy at these temperatures can be used. They can be polarized up to more than 85% by a weak magnetic field. The target is not separated from the beam pipe but a differential pumping system prevents the target gases from reducing the beam pipe vacuum

HERMES uses a set of several detectors for both tracking and PID. The first is a micro strip gas detector, called vertex chamber to reconstruct the vertex. The two front and four back tracking drift chambers are used for the main tracking. Additionally drift chambers inside the magnet are used to

improve tracking and momentum determination. The latter is performed using the large 1.3 Tm magnet. An additional drift chamber, DVC in figure 5.32, is used for large angle tracking. The tracking efficiency is at the level of 99%, the momentum determination is done on the level of $\Delta p/p$ of 1% [3].

A luminosity monitor measures Bhabba and Møller scattering for positrons and electrons, since the luminosity directly enters the asymmetry calculations. Not shown are polarimeters measuring the longitudinal and transverse polarization of the electron/positron beam, i.e. within and outside of the region of the spin rotators.



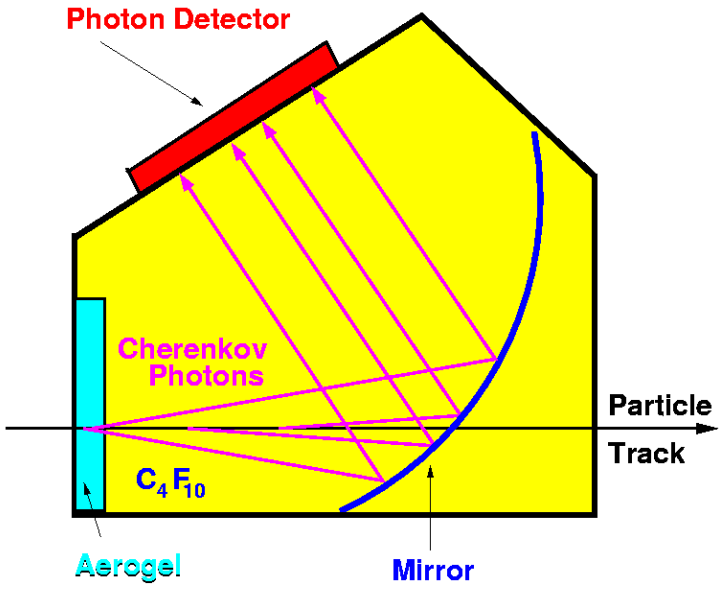
5.33. The calorimeter, the preshower detector and the TRD of HERMES allow an effective electron/positron-hadron separation. Positrons indicates the type of beam particle used for this plot. PID 3 is a quantity derived from the measurements of the calorimeter and the preshower, the other axis depends on the TRD alone. For details see [3]

The particle identification system is divided into a lepton and a hadron identification system. An electromagnetic lead calorimeter, a preshower detector and a transition radiation detector are used to separate positrons/electrons from hadrons. Using a probability analysis and a two-dimensional cut, an electron/positron-hadron separation up to 99.7% in certain kinematical ranges is possible [3]. Figure 5.33 shows the plane for the two-dimensional cut for electron/positron-hadron separation.

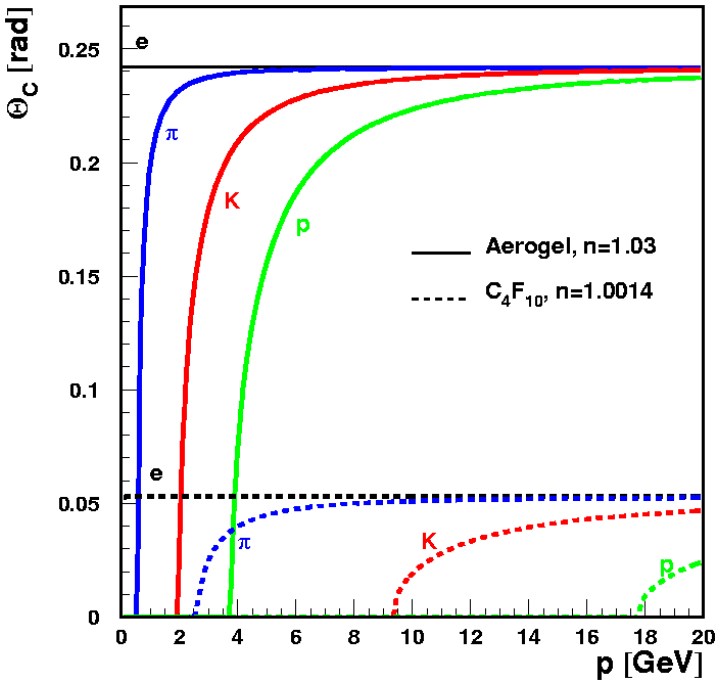
Muons are identified by additional muon hodoscopes. This is necessary because their mass is too close to that of the pion to separate them by a RICH detector besides within a small momentum range [95]. Only their small energy loss in matter is a good observable to separate them from the pions.

For hadron identification two detectors in different momentum ranges are available. At low momenta of up to $2-3$ GeV/c the three hodoscopes from the trigger can also be used as a TOF detector with reasonable performance [3]. For pion identification and additional electron/hadron separation, a threshold Čerenkov detector was used up to 1997. It was replaced by a RICH detector, which has been installed in the same place and body as the original threshold counter.

A detailed description of the RICH detector shown in figure 5.34 can be found in [20,149]. The RICH uses one PMT photo detector matrix for two radiators to increase the momentum range. One is gaseous C_4F_{10} and the second is an amorphous silica aerogel. This is the first RICH to operate with an additional aerogel radiator, and only the availability of this medium made a RICH in HERMES useful [88]. The



5.34. Sketch of the top HERMES RICH detector



5.35. The dependence of the Čerenkov angle on the momentum for both radiators of the HERMES RICH

readout consists for each detector, the top and bottom one, of 1934 single PMTs arranged in a hexagonal close packed grid. By use of funnels the effective detector area is increased considerably [20].

The RICH is used for identification of pions, kaons and protons in a momentum range from 2 to 18 GeV/c. The momentum dependence of the Čerenkov angle for both radiators for all particles is shown in figure 5.35. The HERMES RICH is not used for lepton-hadron separation. It is assumed that the TRD, calorimeter and preshower detector are able to solve this task. This improves the range for particle identification significantly as can be seen in figure 5.35. Muon-pion separation is only feasible over a very limited momentum range with the RICH, and is therefore solved using the muon hodoscopes.

The use of two radiators and of an external lepton-hadron separation both extend the momentum range for particle identification with the RICH detector, an optimization strategy already presented in chapter three.

5.3 Present RICH PID Methods

Currently HERMES has two methods available for hadron identification. One is an IRT method [49, 90, 125], the other a DRT method [49]. For both a much more complicated mathematical algorithm is necessary to compute the Čerenkov angle $\theta_{\check{C}}$ than in the CERES case since due to construction requirements the optics of HERMES produces not rings, but distorted ellipses. An example is shown in figure 5.38, which will be explained in more detail later. Both methods exist in parallel, with the IRT method being the primary reconstruction method, while DRT is not yet used routinely for all tracks and presently not as far optimized as IRT. The major drawback for DRT originally was the computing time necessary which is about one order of magnitude larger than that for IRT. This problem, however, does no longer exist due to an upgrade of the reconstruction farm. Mathematical details of the ray tracing algorithm will not be detailed further, but can be found in [49,90] and only the main features for PID will be discussed. Besides the geometric distortions it is important to note that the HERMES RICH operates digitally. Since two radiators are involved, particles radiating in both are characterized by two concentric ellipses. This results from the fact that the same optics and detectors are used for both radiators. The larger ring with a lower threshold is from the aerogel radiator, the smaller one is from the gas radiator. It is important to notice that the gas ring is so small that it is likely that PMTs are hit by several photons, while this is not the case for the aerogel ring, which thus appears fragmented.

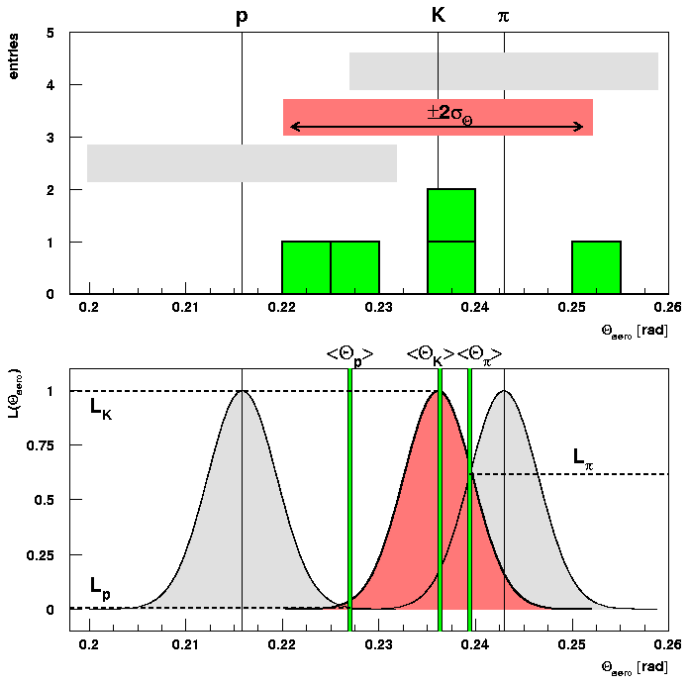
5.3.1 IRT

The IRT method operates quite classically. For each event, the multiplicity of HERMES is about 1.3. For all hits in the appropriate detector half inside or up to the asymptotic radius the Čerenkov angle is constructed for each particle hypothesis for each track [20]. Within a window of $\sigma_\theta = 2\sigma_\gamma$ the single hits angles are averaged to $\langle\theta_{\check{C}}\rangle$, and a likelihood is created as

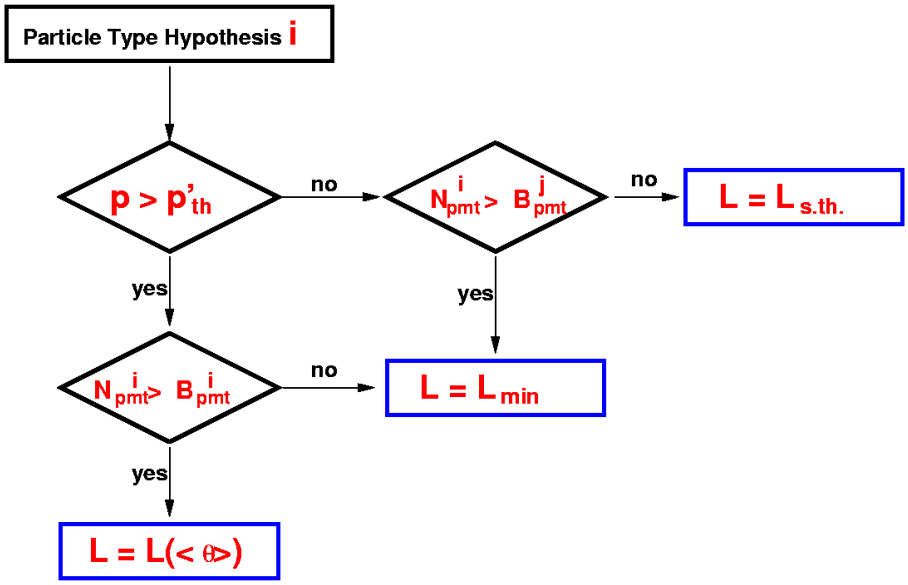
$$\ell_R^k = \exp\left(-\frac{\left(\theta_{\check{C}}^k - \langle\theta_{\check{C}}\rangle\right)^2}{2\sigma_\gamma^2}\right) \tag{5.1}$$

for each hypothesis k for each radiator R. This is shown in 5.36 for the aerogel radiator. Because the Čerenkov angles for both radiators are very different for a given hypothesis, it is easy to determine from which radiator a hit originates for a given track.

The likelihoods for the different radiators are combined by multiplication, if the particle for a given hypothesis is above the corresponding radiation threshold. For the gas ring, due to the small size, the gas likelihood is calculated starting 0.6 GeV/c above the threshold. Additionally, a minimum number of hits within the window is required for a hypothesis to use the angular likelihood (5.1). If this number is not met, or the particle is sub-threshold, a constant likelihood for this radiator and this hypothesis



5.36. The windows used for IRT for a kaon. From [90].



5.37. The likelihood construction for IRT, from [96].

is assigned. The complete likelihood determination for each radiator and each hypothesis is shown in figure 5.37. In this figure N^i_{pmt} is the number of registered hits, B^j_{pmt} is the background threshold, $L_{\min}$ is the minimum likelihood, i.e. the lowest observed likelihood value, $L_{s.th.}$ is the constant likelihood for a sub-threshold particle, $L(\langle\theta\rangle)$ is given by (5.1) and L is the resulting likelihood. The hypothesis with the largest likelihood yields the particle's type. If the highest likelihood is achieved by more than one hypothesis, the particle is flagged as unidentified. This happens in about 1% of the cases, depending on momentum [125]. A detailed description of the IRT likelihood method is given in [96].

A quality parameter is introduced as

$$IRT_{rqp} = \log_{10} \frac{\ell^{k_1}}{\ell^{k_2}} \tag{5.2}$$

where ℓ^{k_1} is the highest likelihood, and hence the one for the identified particle, and ℓ^{k_2} is the second-highest likelihood. IRT_{rqp} assumes the value 0 for unidentified particles and is positive for identified particles. By cuts on IRT_{rqp} a minimum requirement on the quality of the result can be imposed. From a point of efficiency, a $IRT_{rqp} > 0$ cut is optimum [125], but to increase purity a cut of $IRT_{rqp} > 0.5$ may be better, although resulting in a loss of 10% of all particles to remain unidentified. However, a high purity is more important for the physics analysis than efficiency [4].

The IRT method was improved by a correction for photon refraction at the aerogel tile boundaries [124] and for misaligned mirrors [90]. These two corrections resulted in an improvement of the single photon resolution up to the design value.

5.3.2 DRT

The second method for hadron PID is based on DRT [49]. Figure 5.38 shows a typical result of DRT in the HERMES event display. The shading of the PMT pixels corresponds to the probability of this PMT to be hit by a photon emitted from the appropriate radiator if the track is of the type indicated at the lower end of the figure. The dark pixels are the actual hits and the symbols are the virtual track hit points. Part of the aerogel ring is missing because of the mirror acceptance.

For each track, each hypothesis k and each radiator, 360 photons are created, regularly distributed along the trajectory and in azimuthal direction. From the generated probabilities a likelihood is defined as in (3.2), given by

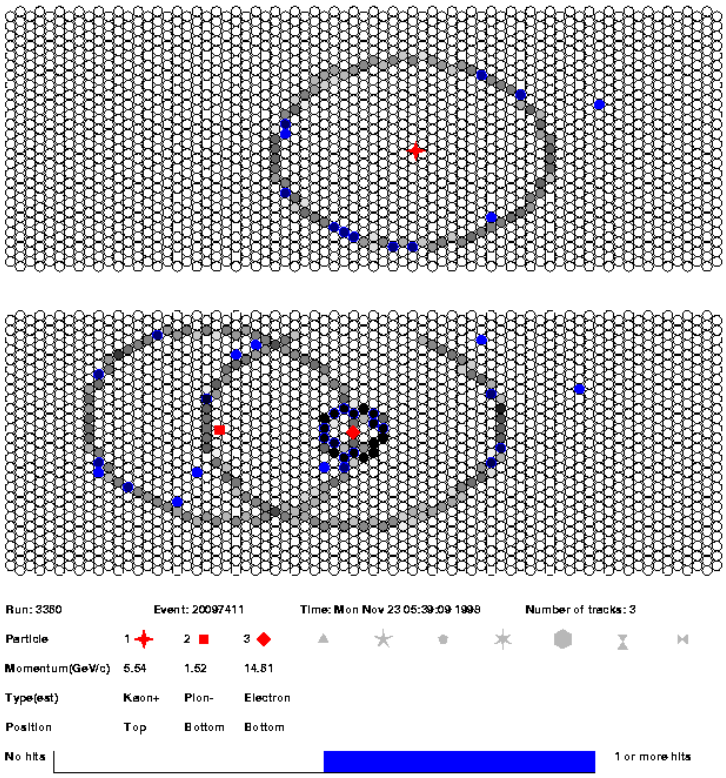
$$\ell^k = \sum_{\text{all pixels } i} \ln\left((P_i C_i) + (1 - P_i)(1 - C_i)\right) \tag{5.3}$$

where C_i is 1 if pixel i is hit and 0 otherwise. There are no distinct likelihoods for the gas and the aerogel radiator as in the case of IRT, and there is no detailed sub-threshold treatment or minimum number of hits. The latter issues are treated by the second term in the above likelihood, which corresponds to a probability for a pixel not to be hit. As in (5.2), a quality parameter is defined as

$$DRT_{rqp} = \log_{10} \frac{\ell^{k_1}}{\ell^{k_2}} \tag{5.4}$$

where again ℓ^{k_1} is the highest and ℓ^{k_2} is the second-highest likelihood. By construction it is extremely unlikely, but not impossible, for DRT to have two equal likelihoods which also results in a zero quality parameter. For identified particles DRT_{rqp} is also positive. A sensible cut for DRT to optimize purity and efficiency seems to be of the order of 2, which also results in a 10% loss of statistics to unidentified.

The DRT algorithm is also equipped with an aerogel boundary refraction correction, and with a mirror correction. In the course of this work it has been improved by a blobbing optimization. This is not turned on in 5.38, but figures 3.5 and 3.6 in chapter three show the effect of this optimization.



5.38. A typical event from an open charm decay candidate as seen from DRT without blobbing. The degree of shading indicates the probability of firing for this PMT for the given particle hypothesis. The darker a PMT is, the more likely it has been hit. This figure is taken from the HeRE event display.

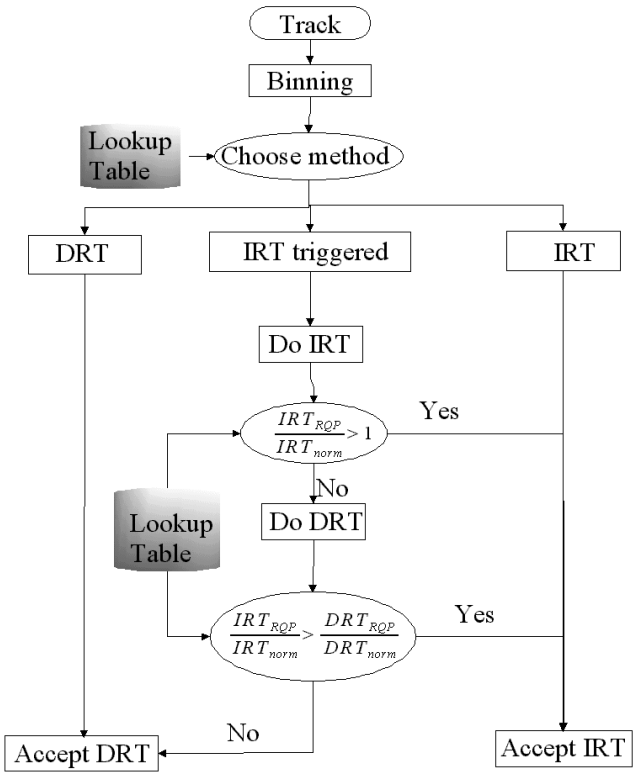
5.4 Optimization

IRT and DRT exhibit different performance in different kinematical areas. IRT is stronger at larger momenta, while DRT is better for low momenta, and there is a different behavior for different event topologies. However, this behavior is again dependent on particle species and event topology, and there is no simple rule to select one or the other algorithm as superior. This is shown in figures 5.43 to 5.48, 5.49, 5.50, 5.51 and 5.52. Since the differences are large it seems nevertheless worthwhile to find a combination of both to profit from the advantages of both. This section describes an approach based upon a variation of a committee network [37], a decision network. The basic principle has already been sketched in chapter three, and the technical details for using this optimization at HERMES are given in [113]. The first part of this section will describe the basic operation, the second part will treat the results and the third part will look into the underlying design philosophies and the motivation for several decisions on the way the optimization was implemented.

5.4.1 Operational Flow

The decision network is a combined veto and parallelprocessing optimization and its main task is to schedule IRT and DRT to do PID. It is therefore termed the RICH PID Scheduler (RPS). The operational flow of the RPS is shown in figure 5.39.

The algorithm operates trackwise, but respects part of the event information during the first step,



5.39. The network topology of the RPS

where the momentum and the event topology are discretized. The momentum is binned in equal bins of currently 1 GeV/c each. The event topology is respected by two features. The first is the track multiplicity for each detector half. The second applies only to situations with two tracks per detector half. In this case the amount of overlap between both rings is calculated using the overlap parameter x_{olp}. This quantity is defined as

$$x_{olp} = 1 - \frac{|\Delta\vec{x}|}{(r_1 + r_2 + r_{PMT})} \tag{5.5}$$

where $|\Delta\vec{x}|$ is the euclidean distance of both virtual track hit points, r_{PMT} is the radius of a PMT and r_1 and r_2 are the radii of the rings in the direction of the other virtual track hit. These radii are approximated by

$$r = 107.7 \cos^{-1}\left(\frac{m^2 + p^2}{pn}\right)(1 + 0.1275\cos(2\phi)) \tag{5.6}$$

where m is the mass or mass hypothesis, p is the momentum and ϕ is the angle against the horizontal on the PMT plane for the direct connection of both ring centers. n is the index of refraction of the aerogel, since the larger aerogel rings are the first to overlap and gas rings should not be without aerogel rings in a perfect world. This definition is from [93]. The overlap parameter was chosen such that $x_{olp} \in [0, 1]$ for overlapping rings is a measure of the amount of the overlap. For rings that are just touching $x_{olp} = 0$, for rings with the same virtual track hitpoint $x_{olp} = 1$. Rings that do not overlap at all have negative x_{olp}. The overlap parameter is discretized as well during the binning. In the case where the rings do not overlap, all possible overlap parameters are projected onto one bin. If they do, the corresponding range of the overlap parameter from 0 to 1 is binned into 15 bins. This number was chosen to have approximately

one bin for one PMT overlap. Rings have a width of roughly one PMT, so this is the relevant scale.

For one track the overlap parameter cannot be defined. For more than two tracks, the amount of data for storage and the amount of necessary statistics needed to extract the rules explained below behave uneconomical. For three tracks the relation is already $1:289$ compared to two tracks. Because of an average track multiplicity of 1.3 for the complete HERMES detector, top and bottom, the PID optimization using the overlap parameter has been restricted to 2 rings per detector half.

Finally, the following binning results:

Multiplicity bin	# of momentum bins	# of overlap bins
1	31	1
2	31	17, see[15]
3	31	1
n>3	31	1

The spatial position of the rings is not taken into account, since effects of acceptance cutoff and position do not change the performance of IRT [126] and is implicitly covered by DRT as seen in figure 5.38. This is visible for the electron ring in form of the missing part of the aerogel ring. Photons from this part of the ring would not have been reflected by the mirror and therefore the hit probability was set to zero.

The next step of the RPS is to choose the appropriate method to do PID for a track in the given bins. This method is selected by a bin-dependent rule which uses a lookup table generated from Monte Carlo data as described below. As a result either IRT, DRT or a combination of both, called IRT triggered and treated below, is applied to the track. The result of the algorithm then is the result of the PID. Only a particle type is delivered, since a quality parameter cannot be coherently defined anymore.

5.4.1.1 IRT triggered

Besides IRT and DRT as stand-alone methods, RPS also provides a combination of both for those cases where one of them is not clearly superior. The important properties for this combination are the quality parameters IRT_{rqp} and DRT_{rqp}. The flow of IRT triggered is also included in figure 5.39. For each track first IRT is performed, yielding besides the particle type the quality IRT_{rqp}. RPS checks then, if

$$\frac{IRT_{rqp}}{IRT_{norm}} > 1 \tag{5.7}$$

where IRT_{norm} is a constant from the rules lookup table, which is also bin-dependent. It is a normalization constant. Its definition is that the result from IRT is trustworthy, if IRT_{rqp} is greater than this value, or equivalently if their ratio (5.7) is greater than 1. If this is the case, the result from IRT is accepted, and IRT triggered is finished. If not, and that is the reason why this method is called IRT triggered, DRT is executed, triggered by IRT. Again besides the type a quality DRT_{rqp} is obtained. If

$$\frac{IRT_{rqp}}{IRT_{norm}} > \frac{DRT_{rqp}}{DRT_{norm}} \tag{5.8}$$

the result from IRT is still accepted as the valid outcome of the PID. If (5.8) is however not fulfilled, the result of DRT is accepted as the valid outcome of the RPS and the IRT result is discarded. DRT_{norm} is as IRT_{norm} a normalization constant with the same properties.

Since IRT_{norm} and DRT_{norm} are chosen with respect to their task in IRT triggered, the fact that IRT is performed first and not DRT is arbitrary. A DRT triggered method could replace this equally well.

Besides trying to obtain the best possible result by this comparison, this approach also reduces the number of unidentified particles, because now both methods have to be unable to identify a track for it to be unidentified.

[15] There are 15 bins in $[0,1]$, one if no overlap is present and one if it is not possible to calculate the overlap parameter.

It is notable that IRT_{norm} and DRT_{norm} cannot be used as optimized cuts for IRT and DRT alone, because they have been extracted to optimize IRT triggered. Nevertheless they still are a measure of how trustworthy the results of the algorithms are, and it could be expected that optimal cuts would not change qualitatively with respect to the bins.

5.4.1.2 Speculative Overlap Parameter

The main drawback of the overlap parameter x_{olp}, as calculated by (5.5), is that by (5.6) the mass of the particle is explicitly used. However, during PID the mass is obviously not known. The problem is resolved by using a speculative overlap parameter. The idea behind this approach is that not really the true overlap parameter is necessary, but a value similar to the true overlap parameter is necessary. Due to this not the true particle type is needed, but only a particle type producing a ring of approximately the same size as the true one. To find such an approximation, an iterative approach is used. Since IRT and DRT do not inherently use the overlap, but are pure trackwise algorithms, they are for the case of two tracks calculated first. For tracks below 5 GeV/c, the resulting DRT type is used to calculate the overlap parameter; above, the type as resulting from the aerogel alone from IRT is used. The reason to only use the aerogel type is that the aerogel ring determines the size of the complete ring, and the overlap parameter is only dependent on the aerogel ring. Additionally, gas rings relatively soon reach a size where they cannot be separated from asymptotic rings due to the resolution of the PMT matrix. To switch from IRT to DRT at low momenta has two reasons. First, DRT generally provides better results for low momenta, as can bee seen from figures 5.43 to 5.48. Secondly, empirically the deviation of the speculative overlap parameter from the true overlap parameter is reduced below 5 GeV/c by about 25% if DRT is used instead of the IRT aerogel type. The remaining deviation decreases with momenta because of the increased photon statistics and aerogel size, and is at the lowest momenta of the order of 1.5 overlap bins.

The resulting overlap parameter is then used as an input to the rule to select the method and for the IRT triggered normalization constants. This is possible because, although parallelprocessing is induced by IRT triggered, the main operation of RPS is to evaluate the results and is hence postprocessing.

It should be noted, that these considerations are somewhat academic, as it has turned out that at least for Monte Carlo the effect of the difference between the true and the speculative overlap parameter does not change the results with statistical significance. The changes are well below 0.5% for all quantities.

5.4.1.3 Rules Extraction

The rules used in the RPS during reconstruction are taken from a lookup table. This lookup table is generated beforehand in a training procedure. The RPS is trained by reconstructing a Monte Carlo sample using different sets of rules to find which are the optimal ones for a given bin. To provide a feedback, the training procedure is supervised. The feedback is attained by the requirement to maximize the training functional. The obvious functional to use for RICH PID is given by the method quality (3.51). To maximize this value means to maximize the reconstruction success.

The training procedure for RPS is to analyze a set of Monte Carlo events, containing in the present case 1.8 million tracks, with a reasonable set of rules for each bin. The set maximizing the method quality in a bin is chosen. This results in a set for each individual bin. These sets then compose the lookup table.

The next problem is to select an appropriate sample of rules to test for each bin. The method rule is simple, since only three discrete values are available, corresponding to IRT, DRT and IRT triggered. This rule can hence be explored completely. This is not the case for the continuous IRT_{norm} and DRT_{norm} variables, where a complete coverage of the available phase space is not possible. A solution is a regularly discretized phase space. The possible ranges[16] for an identified particle of IRT_{rqp}, $(0, 15]$ and DRT_{rqp} $(0, 150]$ have been divided into cells of 0.1 and 0.2 respectively and the lowest values are 0.1 and 0.2. These divisions are fine with respect to the structures which IRT_{rqp} and DRT_{rqp} exhibit [90] if they are

[16] These are empirical values

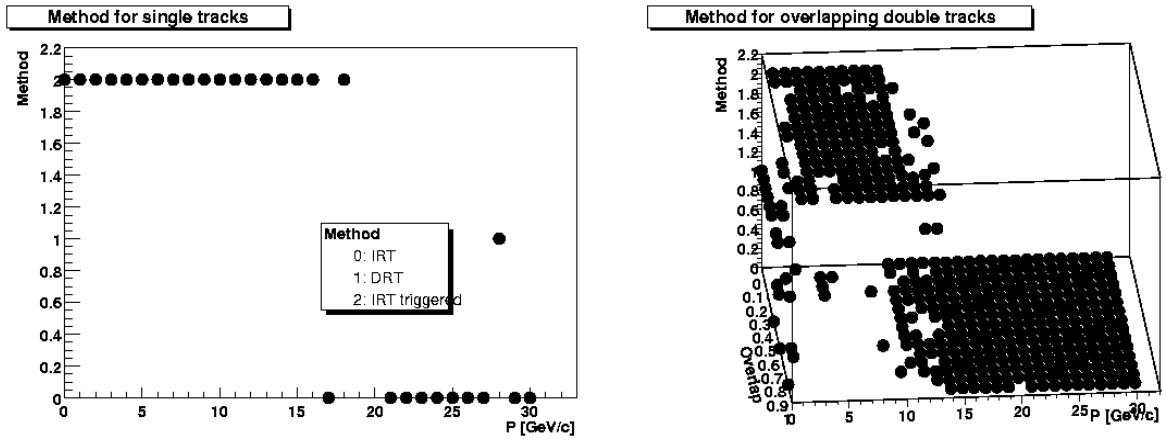
plotted against momentum or momentum and overlap parameter. This discretization allows to use the full possible phase space to obtain the optimal rules for the three degrees-of-freedom method, IRT_{norm} and DRT_{norm}.

For the coefficients of the method quality, initially the values

C_π^P	C_K^P	C_p^P	C_π^Q	C_K^Q	C_p^Q
1	2	1	2	4	2

(5.9)

had been chosen. These values reflect the fact that for all ongoing analyses kaons are considered to be more important than the other species and purity is more important than efficiency [4]. However, it turned out during systematic studies described below, that the RPS is very insensitive to the chosen coefficients, and that no statistically significant change can be observed between the set of coefficients in (5.9) and a flat method quality where all coefficients are equal. All following results have been obtained with a rules set maximizing (5.9), but no visible changes in purity or efficiency will be seen, if a rules set from a flat functional would be used instead.



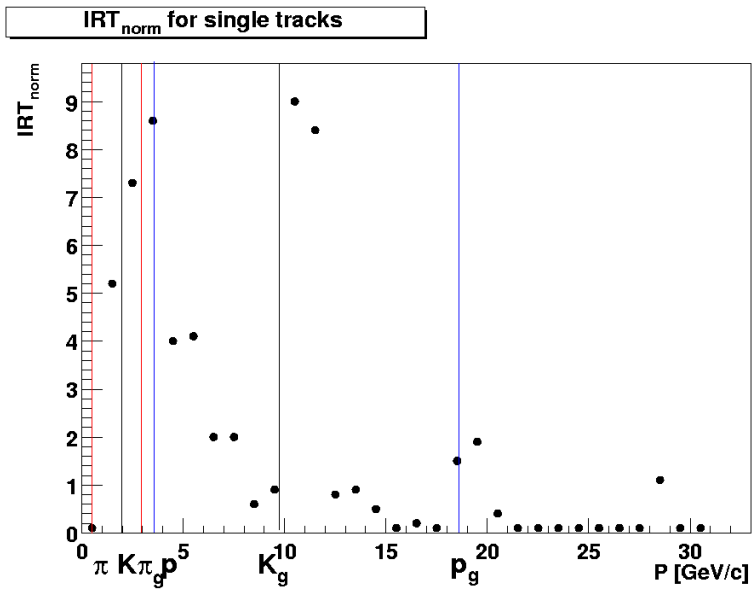
5.40. The methods used by RPS

To see if the RPS rules show a reasonable behavior, the rules resulting from the training have to be studied. The reason for this is, that effectively the RPS models a system of three non-linear functions, one for each degree of freedom. There is no analytical way to determine the resulting set of rules.

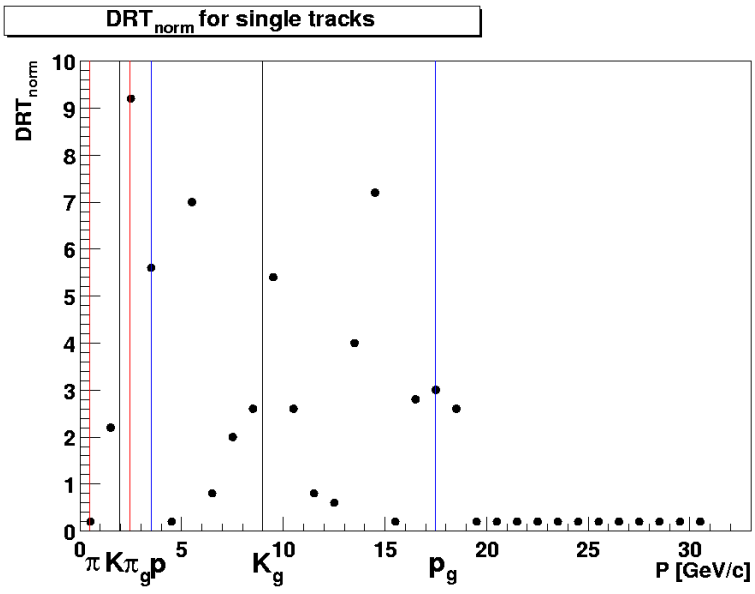
Figure 5.40 shows the methods chosen for single and double tracks. Obviously IRT triggered dominates for the bulk of the data, which is below 10 GeV/c. This clearly shows that no single algorithm is presently suited to solve each possible situation, but the most successful strategy is to use their combination.

Figure 5.41 shows the values for IRT_{norm}, figure 5.42 the values for DRT_{norm} for single tracks. If compared with figure 5.35, it is visible that the algorithms are trusted most, that is the normalization is lowest, if the momentum is far away from any threshold. The normalization constants are highly peaked at the threshold momenta for the different particle species for both radiators. Even the fact is visible that for IRT the gas threshold is moved up by 0.6 GeV/c.

These results clearly demonstrate that the extraction of the rules was able to find the most troublesome regions of phase space: By design the onset of radiation from the gas radiator is placed at momenta where the asymptotic aerogel rings of two species become no longer separable. This causes difficult areas in the vicinity of the threshold, where the rings from one radiator are nearly identical and the other rings are not yet significant due to the onset. These results clearly demonstrate the ability of the extraction process to classify the obstacles for the RICH PID correctly.



5.41. The IRT normalization for single tracks. The vertical lines are the thresholds of the indicated particle types, the index g stand for the gas radiator. The IRT momentum shift for the gas radiator is respected.



5.42. The DRT normalization for single tracks. The vertical lines are the thresholds of the indicated particle types, the index g stand for the gas radiator. The peak at 14 GeV/c is located where the kaon and pion aerogel ring become indistinguishable

It has to be added that for two-track events the extraction has been done using the true overlap parameter instead of the speculative one. The reason is that errors in its determination are not inherently part of the reconstruction, but of the approximation described above. It was decided not to fold any dependences of this process into the rules extraction. However, since it has turned out that the difference between true and speculative overlap is insignificant, this consideration is of no further importance in this case. Nevertheless, for other experiments, especially for those with higher spatial resolution, this might be of importance.

5.4.1.4 Monte Carlo Data

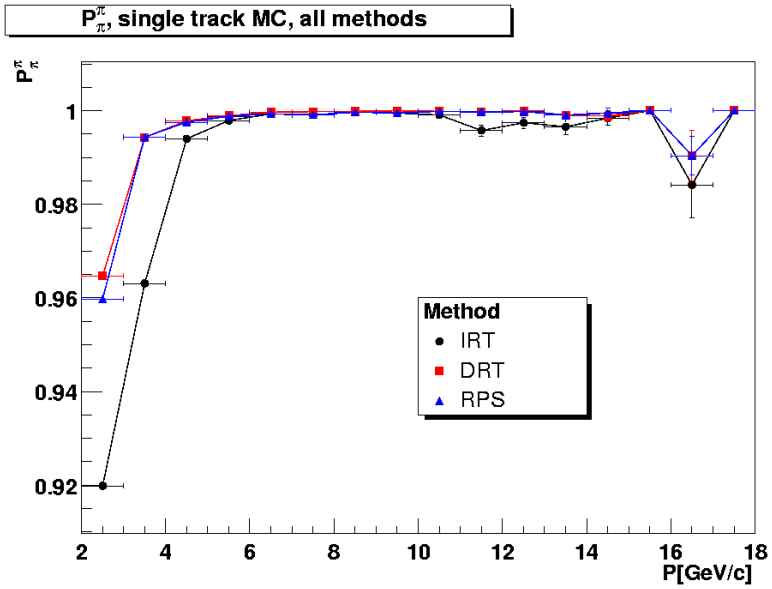
The RPS rules are extracted from Monte Carlo data, because the setup of HERMES, does not permit to produce a clean particle sample with the properties of average data using other spectrometer components. It is possible for decaying particles by invariant mass analysis, but these events only represent a certain topology class, and are not suited to produce a rules set for all data. Nevertheless it is possible to judge the quality of the Monte Carlo data from comparisons between Monte Carlo results and decaying particle data, as extracted in [125]. From a comparison of the results shown later and in [113] and from [125], a confirmation is at least partially achieved within the statistical errors. Where large differences occur, reasonable explanations are available. For example for the kaon two-track efficiency, the values from Monte Carlo are considerably lower than for data from decaying ϕ. This can be attributed to the fact that for ϕ decays the other track is mainly a kaon, while for an average DIS sample it is a pion or electron or positron. In the latter cases the rings of the particles at the same amount of overlap provide much more hits and therefore disturb the PID stronger. Since the results are overlap-integrated up to now because of statistics, the effect that overlap occurs much more often for average DIS than for decaying particle data is shadowed. This results from the fact that pions and electrons/positrons normally have an asymptotic ring while kaons do not most of the time. Hence the area occupied is larger and overlap is more probable. It can be seen from Monte Carlo data [113] that the performance for overlap with respect to single tracks drops.

As a result, it is necessary to use Monte Carlo data for rules extraction, if an average DIS sample shall be treated. The aim is to improve the Monte Carlo simulation to be as realistic as possible. To the standard Monte Carlo, which already includes refraction on the aerogel tiles, imperfect mirrors and a realistic photon statistics, an addition was made in the course of this work. What has not been included previously was a realistic background from noise or trackless rings. The latter refers to rings which are present but there is no track associated with them. They might originate from not reconstructed tracks because of a vertex cut in the track reconstruction or from particles which have been created elsewhere, like conversions. For HERMES a visual inspection of a few thousand events turned out, that the average ring multiplicity is 1.5 in comparison with a track multiplicity of 1.3. The corresponding background cannot be simulated by Monte Carlo. To include it, background from real data was mixed into the Monte Carlo data. This background was extracted from real data and consists of "empty" detector halves. A half is empty in this case, if no track points to it. This approach automatically includes hot channels as well.

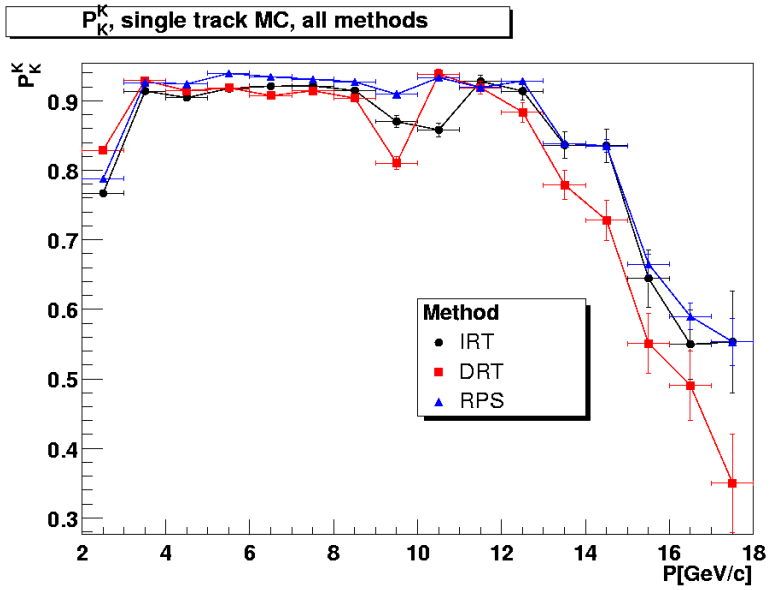
As a result, the measured performance of the PID algorithms for Monte Carlo data decreased up to 10% and the resulting values were in much better agreement with the results from decaying particles data.

5.4.2 PID Performance

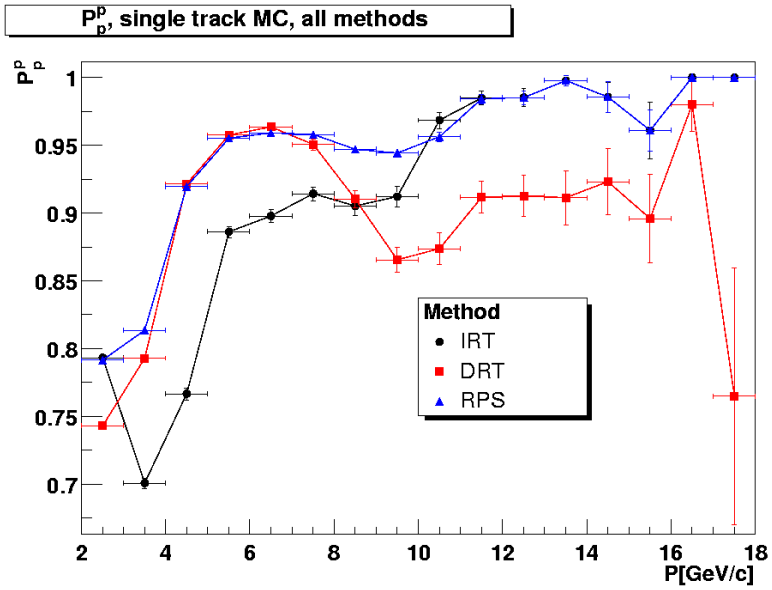
This section will present the results from Monte Carlo data and from decaying particle data for the RPS algorithm presented above. The interesting quantities are the efficiencies and purities for each species as well as the method quality.



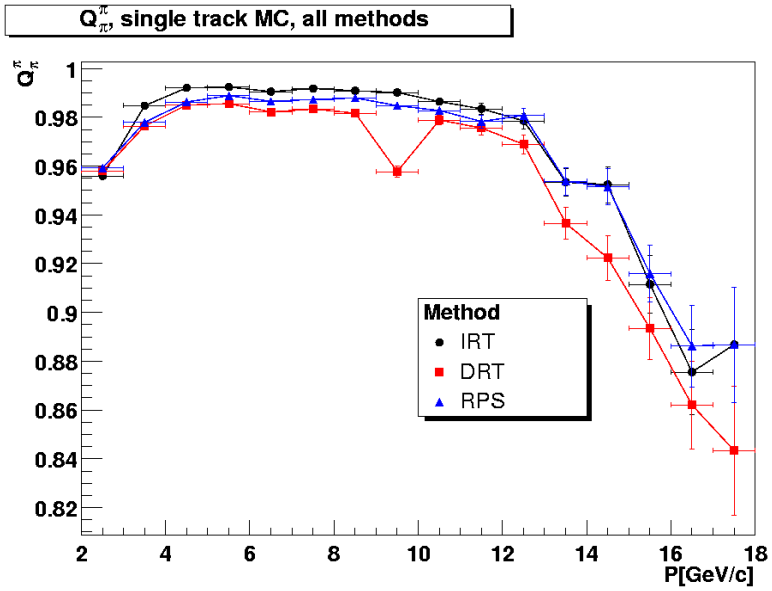
5.43. Pion efficiency for single track Monte Carlo



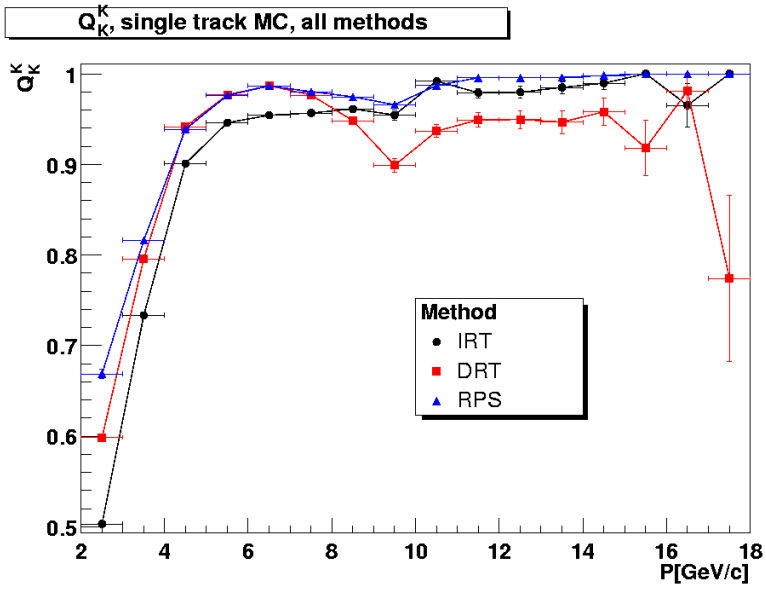
5.44. Kaon efficiency for Monte Carlo single tracks



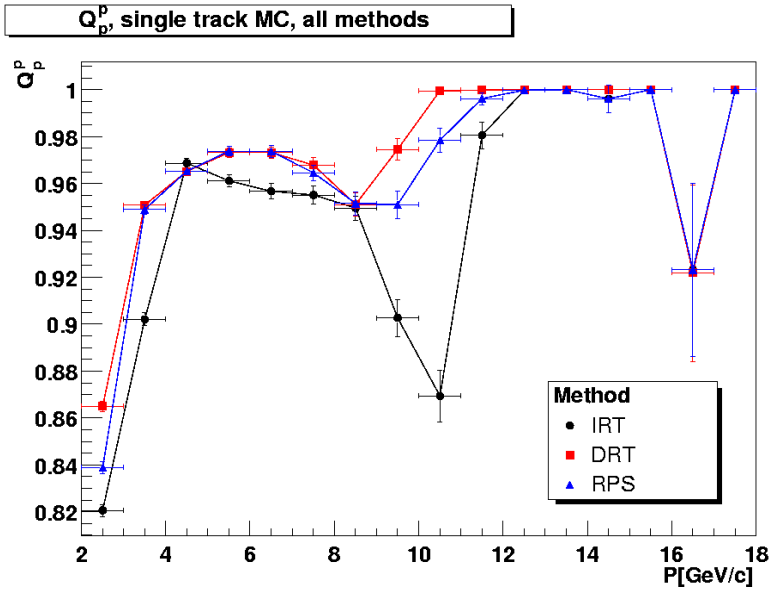
5.45. Proton efficiency for single track Monte Carlo



5.46. Pion purity for single track Monte Carlo



5.47. Kaon purity for single track Monte Carlo



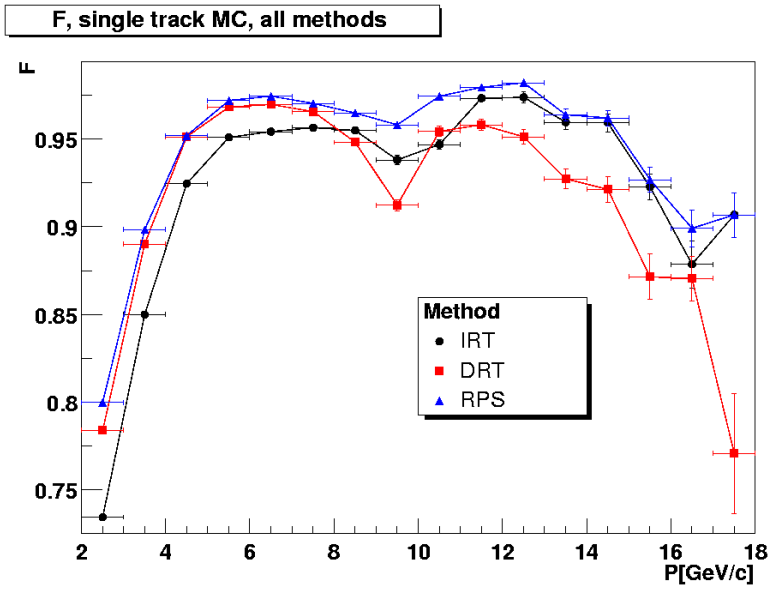
5.48. Proton purity for single track Monte Carlo

5.4.2.1 Results from Monte Carlo Data

For Monte Carlo the interesting quantities can be obtained directly using the formalism of chapter three.

Figures 5.43 to 5.48 shows the results for single tracks. Several interesting features are apparent. First IRT is strong at large momenta and weak at low momenta compared to DRT. Both algorithms show a structured behavior with respect to momenta. The difference between both methods is, depending on the bin, significant up to about 10%.

If compared to RPS, it is clearly visible that both methods are not competitive. This is true if the arguments for the physics case from above are taken into account and special attention is paid to kaon purity. RPS shows a behavior which is better at least than that of one of the algorithms, and in numerous bins is better than both of them. RPS smoothes the structures exhibited by IRT and DRT as well, if one of both methods does behave smoothly. If both methods show a similar structure, RPS is also forced to follow it. The reason is the dependence of RPS on IRT and DRT. It can only profit from the fact that one of both methods is better than the other. If both methods exhibit a similar behavior, RPS is by definition not able to produce a completely different one. Altogether, it is visible that RPS exhibits a higher integrated performance than either of the methods, which qualifies it for the use as the primary reconstruction method.



5.49. The quality for RPS, IRT and DRT for single track Monte Carlo data

Also visible from these plots are the artifacts from the maximization of the method quality. The most prominent example is the lowest bin for the kaon efficiency and purity. For the purity the improvement of RPS against IRT and DRT is large, while for the efficiency DRT is superior, but the difference is small. The reason is that at the cost of losing a few percent of efficiency the purity can be increased by an amount larger than a factor of two. Since this is supported by the maximization requirement, this is preferred compared to a solution where RPS is equal to DRT in both bins. It is noteworthy that pions in comparison to kaons and protons suffer from a nearly always reduced purity compared to IRT. The reason here are the fluxes. Because pions are about ten times more abundant than kaons or protons, it is advantageous for the purity of both particles if pions are identified correctly: One correctly identified pion is worth ten correctly identified kaons or protons for Q_K^K and Q_p^p. At the same time one correct

kaon or proton is only worth 1/10 pion for Q_π^π. It is therefore advantageous to identify pions as well as possible, even at the cost of reduced kaon and proton efficiency. The lost kaons and protons do not spoil the pion purity too much if they are identified as pions. But the other way around the effect is notable. As a result, the $\underline{P}$ matrix of the RPS shows a strong upper triangle character.

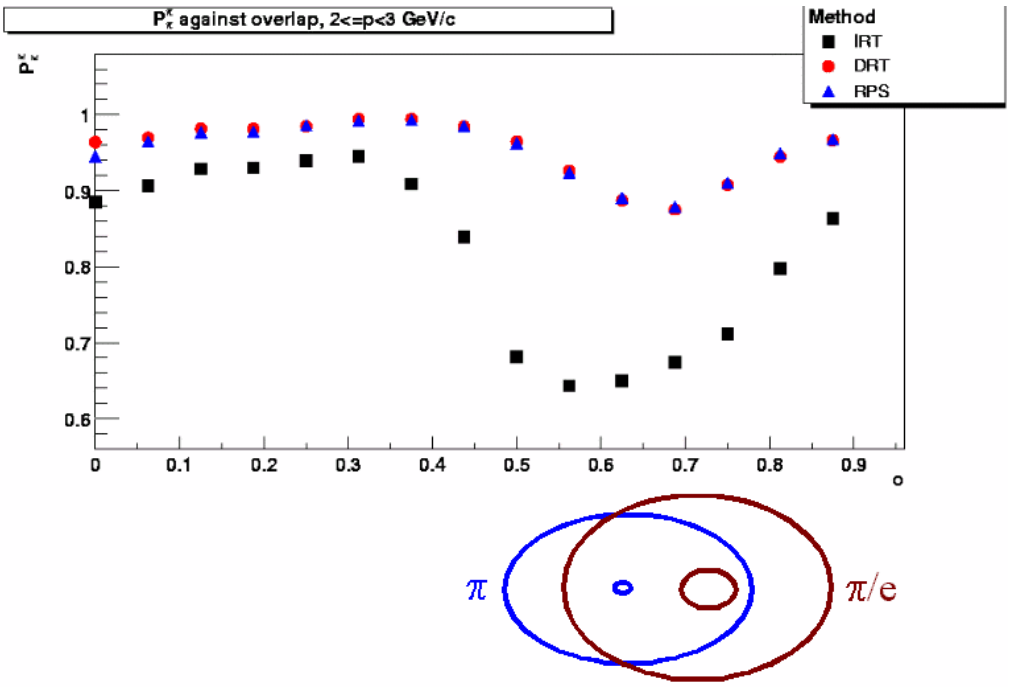
Note that all arguments given for the interpretation of the plots hold irrespective whether a flat method quality or the one given by (5.9) was used.

An interesting question is whether RPS is meeting the design goal and is able to be the best algorithm with respect to the method quality used to define it. As figure 5.49 shows, it is. The overlap-integrated results for two and more tracks look similar, besides the fact that for all methods the performance degrades with increasing track multiplicity. The reason is the increasing probability for overlap, which spoils the implicit assumption for IRT and DRT that only hits from the track to be identified are available. These results are shown in [113] and will not be reproduced here.

The dependences become more interesting, if the case of two tracks with resolved overlap dependence is treated.

5.4.2.2 Overlap

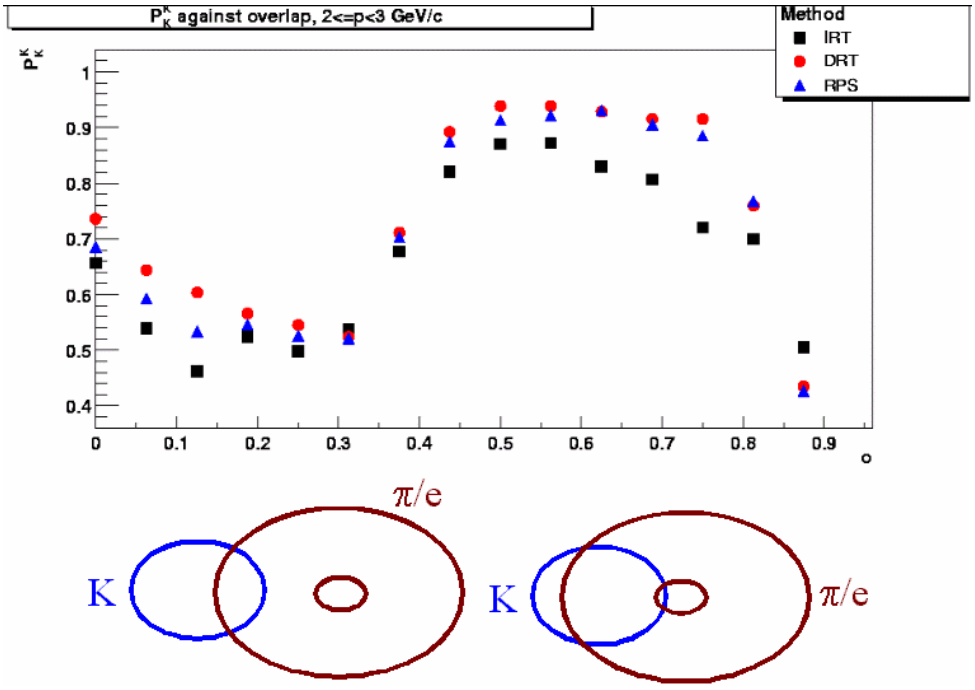
In this case a much richer structure in efficiency and purity is visible and the effects of the overlapping rings become clearly evident. This section will discuss the occurring phenomenons for three examples. The resulting explanation scheme can be extended to other cases for each bin. The examples all concern efficiencies. It is simpler to develop the scheme in this case because cross-correlations from other species are only involved due to the second track. The situation is more complicated for purity, but the basic principles remain the same.



5.50. Pion efficiency versus overlap parameter for RPS, IRT and DRT between 2 and 3 GeV/c

The first example shown in figure 5.50 is the behavior of P_π^π between 2 and 3 GeV/c. In this momentum

range pions do not have a fully evolved gas ring, and IRT ignores it because of the 0.6 GeV/c shift. For an overlap which corresponds to the virtual track hit of the pion being within the aerogel ring of the other particle[17], IRT exhibits a serious drop. The most probable other particle is another pion or an electron or positron, mostly with a larger momentum than 3 GeV/c and hence with a fully evolved gas ring. As a result, parts of the aerogel ring and parts of the gas ring of the second particle are at a distance of the virtual track hit of the pion where an aerogel ring for a kaon is expected. This leads to misidentification of the pion as kaon or as unidentified, if the hypotheses for kaon and pion have the same value. DRT includes the gas ring, and the additional hits provide enough support for the pion hypothesis on average to reduce the effect. It is nevertheless present.

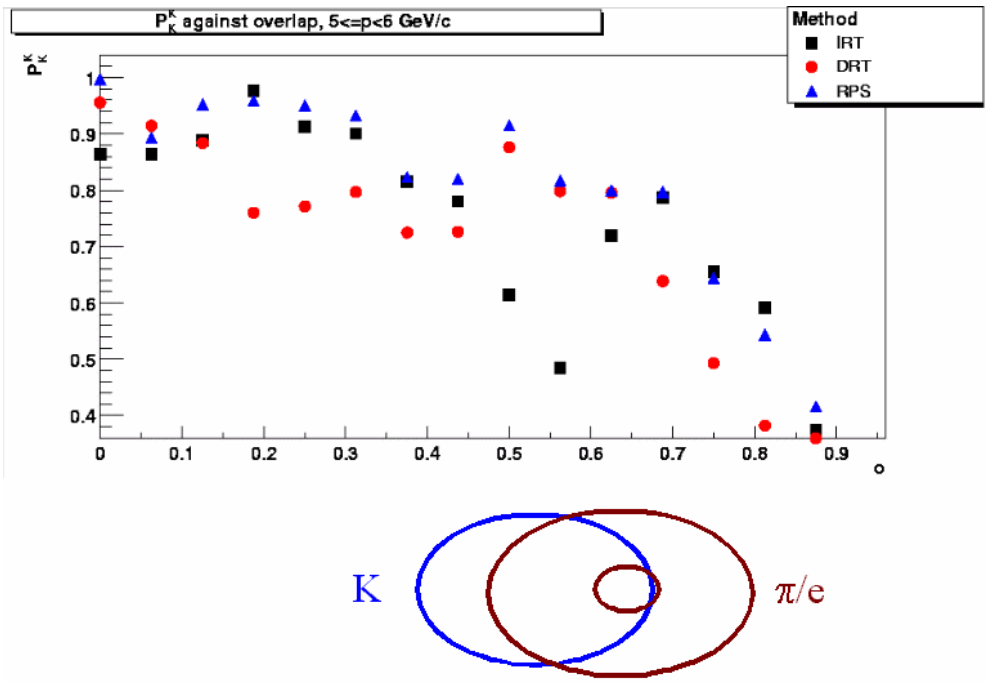


5.51. Kaon efficiency versus overlap parameter for RPS, IRT and DRT for momenta between 2 and 3 GeV/c

Figure 5.51 is the behavior P_K^K between 2 and 3 GeV/c. In this region the kaon has a small aerogel ring, but no gas ring. For a small overlap, corresponding to an overlap only of the aerogel rings, both IRT and DRT show a massive drop. Again the second particle is most likely a fully evolved pion or electron/positron. The aerogel ring of the second particle does not include the virtual track hit of the kaon, but is in a position where a gas ring for the pion would be expected. This gas ring is the reason for DRT to lose efficiency. IRT, which ignores the gas ring, is irritated by the gas ring segment of the second particle. The segment nearer to the virtual track hit provides numerous hits at a distance where a pion gas ring would be expected, and forces this hypothesis. The deficiency for IRT at large overlap corresponds to the other segment of the gas ring being at the position of the expected pion aerogel ring. The problem in this case is the low statistics of the kaon aerogel ring because of the small Čerenkov angle, see (2.9).

Figure 5.52 shows the behavior of P_K^K between 5 and 6 GeV/c. In this momentum range the kaon

[17] A value of 0.5 corresponds to the virtual track hit touching the aerogel ring of the other particle



5.52. Kaon efficiency versus overlap for RPS, IRT and DRT for momenta between 5 and 6 GeV/c

has a full aerogel ring but no gas ring. IRT shows a large loss at a configuration where the aerogel ring of the second particle, supposingly again a pion or electron, includes the virtual track hit. In this configuration IRT mistakes this aerogel ring segment for a pion gas ring, which leads to misidentification. This happens not symmetrically on the other side of the virtual track hit at lower values of x_{olp}, because of the curvature of the ellipses. In this case only a small part of the aerogel is in the gas window, resulting in a too weak support for the pion hypothesis. DRT does not have this problem, because it seems to be not too sensitive to the number of fake gas hits the aerogel ring of the other particle can provide. The reason is that DRT explicitly includes the probability of the absence of hits in the second term of the likelihood (5.3). IRT is then increasing again as the aerogel ring leaves the gas window, but as the gas ring of the other particle enters the gas window, it again loses performance. This is also true for DRT since the number of fake gas hits is now much too high because of the denser populated gas ring for the second particle.

The general scheme of all argumentations is always to observe the effects when the aerogel and gas rings enter the different windows of the particle. Depending on the sensitivity of IRT and DRT on a given number of hits, this may result in misidentification.

For all plots again the observations made for RPS for single tracks are true. As long as one method clearly dominates, RPS follows the behavior of this method, but if both methods fail, RPS fails as well. This is visible in figure 5.52. Again, depending on the bin, IRT or DRT show the better performance. Profiting from this behavior, RPS is able to be superior for the integrated multiplicity bin, even if not for each overlap or momentum bin. The reason is, that RPS is somewhere between both methods, but always nearer to the better one as can be seen for example in figure 5.50. As an integrated effect, this results in a better performance.

Note that the algorithm's performance massively depends on the event topology. If physics results are extracted that depend on this topology this has to be taken into account.

5.4.2.3 Statistics

A vital point to be reviewed if comparing IRT, DRT and RPS is the question of IRT_{rqp} and DRT_{rqp}. Both provide a tuneability for the purity. RPS has nothing similar. All results presented thus far are for IRT and DRT cuts greater 0, which effectively are no cuts. If the standard cuts mentioned before for IRT and DRT are used, both algorithms can improve their purity but at the cost of statistics and efficiency. Especially the loss of statistics is significant:

All %	π	K	p	Total
IRT	6.0	14.1	14.2	8.3
DRT	1.7	13.5	34.3	8.5
RPS	0.0	0.0	0.0	0.0

This table shows the percentage of unidentified particles for IRT, DRT and RPS. The total number takes the relative fluxes into account. For all methods a cut of $p > 2$ GeV/c has been imposed. For IRT and DRT additionally $IRT_{rqp} > 0.5$ and $DRT_{rqp} > 2$ has been imposed. No further cuts were made (or are possible) for RPS. The result is that to be at the same level of performance, both IRT and DRT lose about 10% compared to RPS. Kaons are the most important identified hadrons for the present HERMES physics analyses, and it is for them that the statistical losses are most severe. So IRT and DRT can only provide the same statistics or the same performance as RPS but not both at the same time. In other words, compared to IRT and DRT with standard cuts, using RPS corresponds to running HERMES one year longer, if seven years of data taking with RICH are assumed.

5.4.2.4 Results from Data

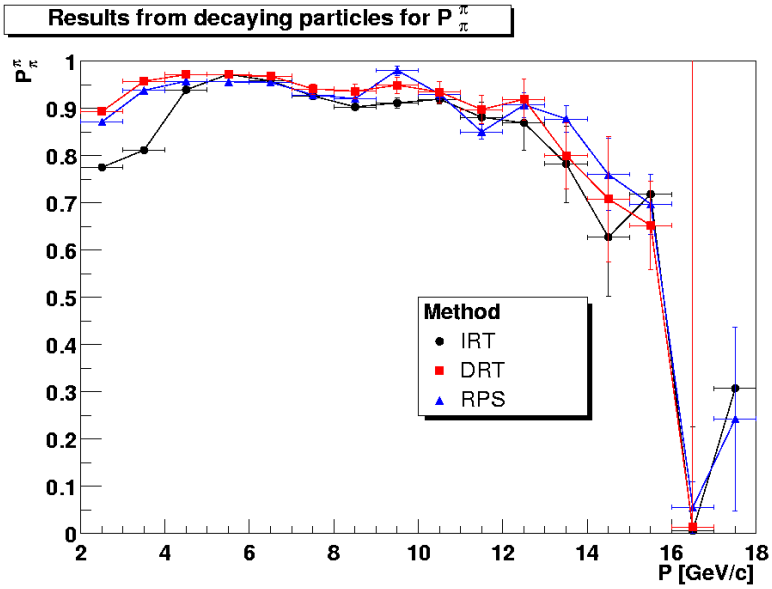
Even if the Monte Carlo simulation provides results which are comparable with real data, a cross-check with real data is still necessary. This cross check has been done using a decaying particle sample with the methods and the code described in [125]. The results shown here at not comparable to those shown in [125] because a higher multiplicity was required resulting in reduced efficiencies.

Preliminary tests with rules extracted from an earlier version of the Monte Carlo simulation already indicated an increase in momentum integrated kaon and proton efficiency by 10% compared to IRT. The pion efficiency stayed the same within error bars. The results were obtained from data from decaying ϕ for K, from K_s for π and from Λ for p.

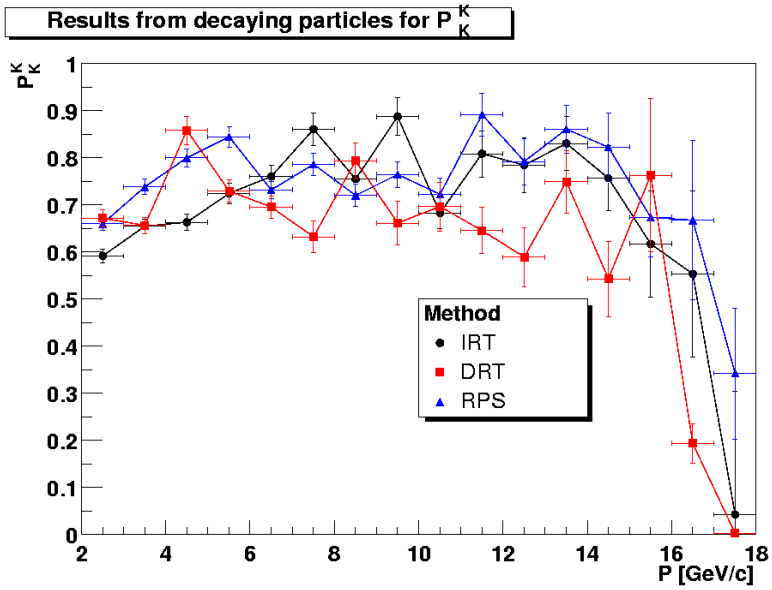
With a rule set obtained from the present Monte Carlo as described above, the results presented in figures 5.53 and 5.54 were achieved for P_π^π and P_K^K respectively. A new proton results was not yet obtained. For kaons it is obvious that the statistics are not sufficient above 7 GeV/c. The results shown are track multiplicity integrated.

RPS exhibits an advantage of up to 10% compared to the single methods up to 6 GeV/c for P_K^K. Since the bulk of the particles is located at these momenta [125], again the integrated efficiency for RPS is superior. The effect is smaller for P_π^π but RPS exhibits over the whole momentum range a better or at least within the errors equivalent result as the single methods. For P_π^π DRT is as efficient as RPS, as has already been seen for Monte Carlo data in figure 5.43. Again the largest advantage compared to IRT is present at low momenta, where the bulk of the pions is located. This is especially important with a view to purity. Since pions are the most abundant particle species, the advantage of RPS is able to improve the kaon and proton purity considerably. This observations become even more important when noted that RPS was optimized for an average DIS sample, while IRT was tuned[18] by the use of the decaying particle data. Because the average topology is different between decaying particles and average DIS, RPS is exhibiting results which are below the achievable optimum for decaying particles. A rules set from a dedicated Monte Carlo simulation of decaying particles may again improve the RPS results. This difference to average DIS makes it also difficult to compare the results from decaying particles with the

[18] DRT was not tuned yet.



5.53. Pion efficiency from real data against momentum. The results are from K short decays



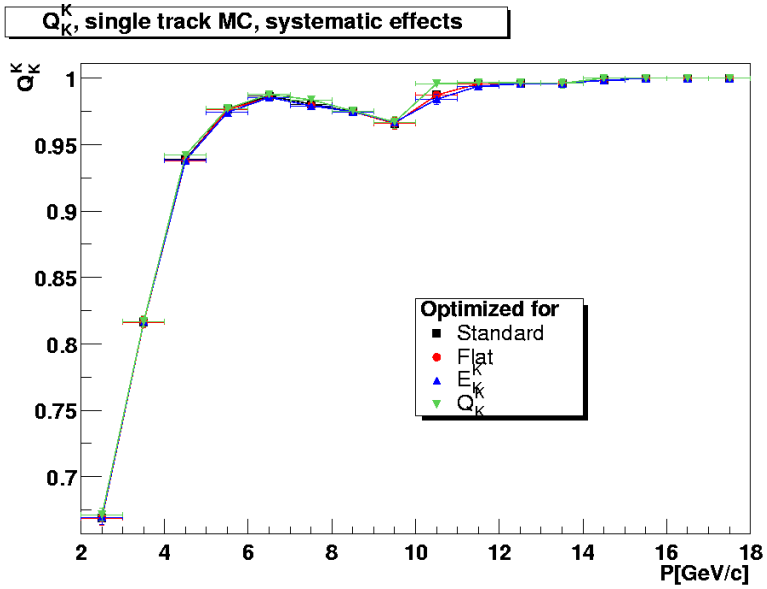
5.54. Kaon efficiency from real data against momentum. The results are from phi decays

Monte Carlo results shown previously. However the qualitative feature that RPS is superior or at least comparable to the better method is maintained. These result therefore prove again that RPS achieves the design goal and can be used as a standard PID.

However, a comparison of a dedicated decaying particle Monte Carlo simulation with the data will still be necessary for the use of this performance results in physics analyses. This Monte Carlo sample is not yet available. Therefore the only result from real data up to now is, that the results are at least qualitatively in agreement with the Monte Carlo data.

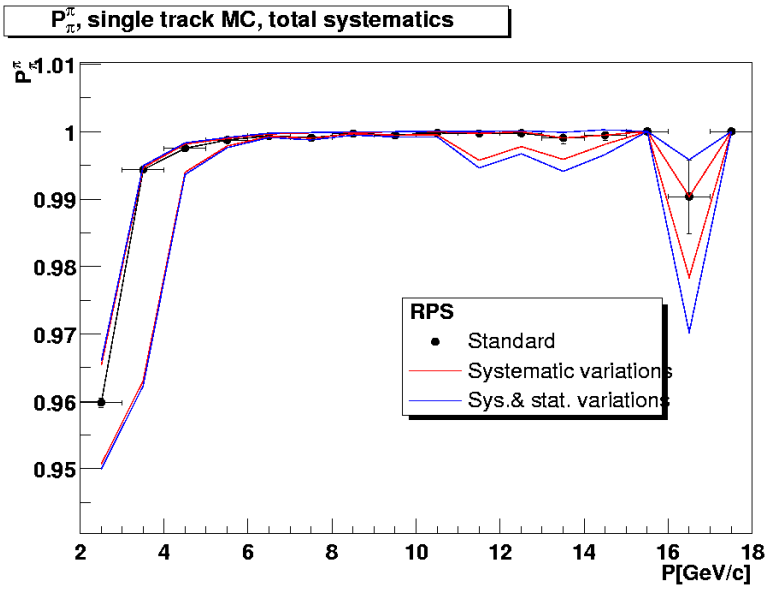
5.4.3 Systematic Effects

By use of the method quality as the governing quantity for the training procedure, a high degree of arbitrariness is introduced. Since six free parameters can be manipulated, it is important to study how robust the results are under changes. These systematic studies have been performed by analyzing the most extreme sets of parameters. Besides the initial set (5.9), a flat distribution is used. Additionally quenched sets of coefficients haven been used. In a quenched set one quantity is made infinitely more important than any other. This is simply achieved by setting the coefficient for this quantity to 1 and all others to 0. This has been performed for all possible effectivities, efficiencies and purities. All parameterizations result in similar observations. The result visible from figure 5.55 is that no significant difference is found between the flat and the initial parametrization (5.9). This indicates that the RPS is stable against small changes of the functional. Only for the quenched values a significant effect emerges in figures 5.55. The result is as expected. If Q_K^K is optimized, then Q_K^K increases. The effect nevertheless is small.

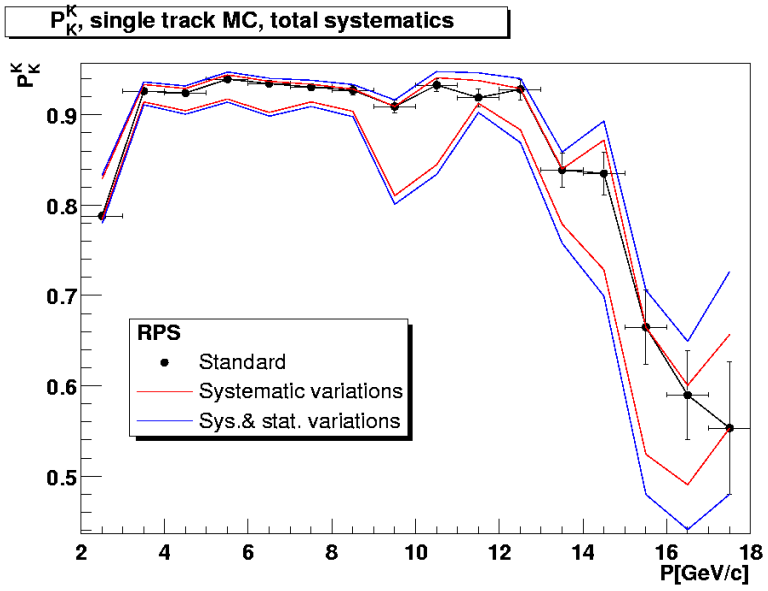


5.55. Systematic effects on kaon purity when optimizing for kaon purity and effectivity

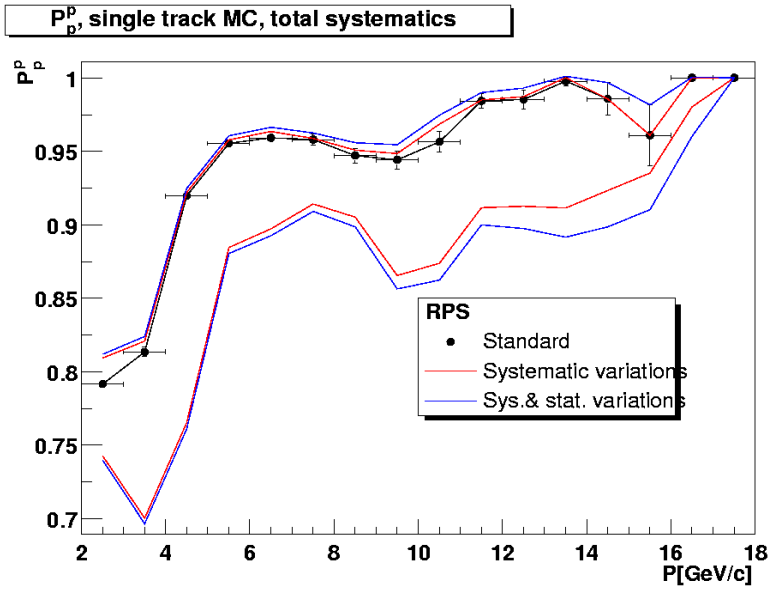
The figures 5.56 to 5.61 show the variations possible for each quantity. The points represent the standard and flat parametrization. The inner band shows the variations of the quantity for any of the parameterizations mentioned above. The outer band shows the statistical uncertainty on this variational band. Obviously all variations are quite small, although larger than the statistical errors. This is especially interesting when it is noted that these bands for example includes the effect of optimizing for P_π^π on Q_K^K. The result is surprising at first, especially in this case where an efficiency is optimized, which should not



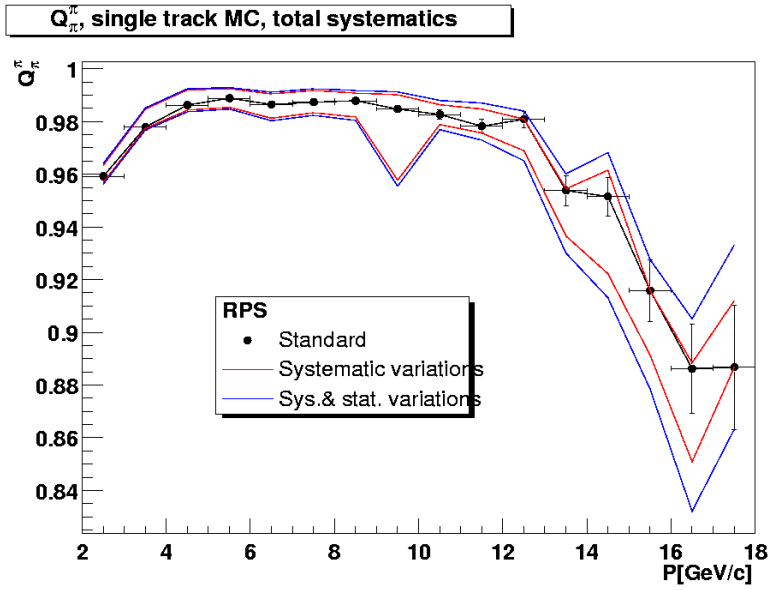
5.56. Systematic and statistic variations of the single track pion efficiency



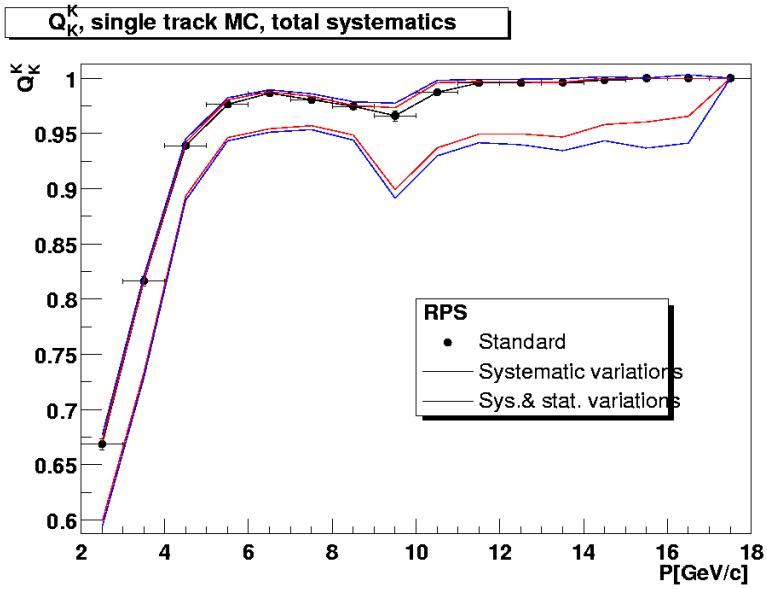
5.57. Systematic and statistic variations of the single track kaon efficiency



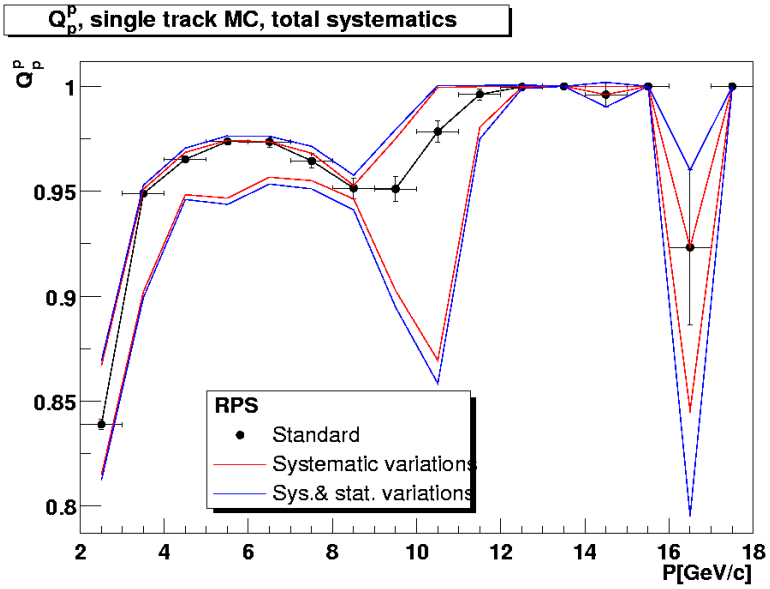
5.58. Systematic and statistic variations of the single track proton efficiency



5.59. Systematic and statistic variations of the single track pion purity



5.60. Systematic and statistic variations of the single track kaon purity



5.61. Systematic and statistic variations of the single track proton purity

depend on other species, at least for the single track shown. The reason is that the IRT and DRT do not distinguish anywhere between species, while RPS does so during the training. Even if the normalization values IRT_{norm} and DRT_{norm} are chosen appropriately to only pay attention to one quantity and the method which performs best for it, the chosen method is still good enough for other particle species too. Since the normalization values only increase the trust in the general results, but do not intentionally try to reduce a species, no type is neglected. Effectively, if the IRT-DRT combination RPS would treat each species equally, as they are designed to, any systematic study would show no difference between any parametrization. What is visible are only results from the fact that this equality is not perfectly achieved. In other words, there is no normalization combination for IRT and DRT which is able to significantly reduce the result for a given quantity. RPS profits from this robustness of both.

The difference of RPS compared to each single algorithm results from the different performance of both compared to each other, but the changes seen here represent the differences for each species with respect to only one algorithm.

The other important observation is that in nearly all bins the standard parametrization is already at the upper edge of the variation band. This indicates that nearly all quantities have reached the optimum, only a very slight possible improvement can be seen. The interpretation is that for a flat or nearly flat parametrization all quantities are respected as good as possible. Due to the cross correlations induced by the purities between the species, it is important to increase the efficiencies for all species, which in turn optimizes the purities. As a result the drop of any quantity is connected to a large penalty in the method quality and is therefore disadvantageous. If possible, it is avoided. This is not always possible, resulting in bins where the upper edge of the variational band is above the standard point, but a good approximation is nearly always possible.

Altogether, the main observation is that RPS is very stable against a change of parametrization and the flat method quality already provides a nearly optimal performance for all quantities. This guarantees a high stability and optimum performance for any physics case, and hence a custom fit as explained in the next section will be hardly needed, if the properties of the data sample is comparable to the sample used to generate the rules. The result of this systematic studies is hence that the RPS is already in flat parametrization an omnipurpose algorithm, leaving only a small space for improvement.

5.4.4 Custom Fit

The RPS can be optimized rapidly for another physics purpose than kaons by a feature called custom fit. This is described in much detail in [113]. While conventional algorithms have to be optimized for different particle species and purposes tediously by hand, the RPS can be easily fitted to the need. The key to its flexibility is the automated training procedure which is governed by the simple functional (3.51). This functional can be understood from a physics point of view and the effects of changes are, as presented in the previous section, predictable. This allows an optimization guided by a given physics case.

Although the systematic studies of the previous section showed that this property will be hardly necessary for HERMES, this is not a general statement. The reason for this is that the species independence of the IRT-DRT combination is good. In general this is not true, and the available flexibility of the RPS or similar algorithms may be valuable.

This is also the case, if the PID should be optimized for a special case, e.g. vector meson production. In this case the generation of a new rules set will allow to optimize the PID.

The key feature is the automated training procedure, which has been mentioned several times. This feature is a heritage of the origin of the RPS as being a neural network, as discussed in the next section.

5.4.5 A Neural Net

The original design for the RPS was motivated by the principle of a committee network based on a neural net [37]. Because of the necessity to optimize performance, speed and use, only some artifacts remain at

the surface.

However, the underlying principle still is a neural network. The structure can be unfolded as follows: The discretization is not part of the net, but is a preprocessor stage. Besides discretization for technical reasons discussed below, it maps the event topology to two parameters, the track multiplicity and for two tracks additionally to the overlap parameter. This process is referred to as dimensional reduction and is helpful to reduce the necessary complexity of the network and the phase space of possible configurations. The input neurons are only taking a few parameters, the momentum and the overlap and multiplicity bin, instead of 1938 values, the 1934 PMTs, the position of the track, the multiplicity and the momentum. The problems to provide a reasonable training set for 1938 parameters would be tremendous, an effect known as the curse of dimensionality.

The few input values are fed by a classical feed forward mechanism through the hidden neurons for the choice of the methods and the evaluations of IRT triggered. The neuron function is in this case not continuous but a discretized numerical function in form of the rules. These numerical functions would normally again be modeled by a neural network rather than by a lookup table, but for the sake of simplicity this has not been implemented at the cost of the necessary discretization. The results then determine the output of the hidden neurons to the single output neuron which gives the type. An additional auxiliary output neuron has been introduced to record the used method, but this is not a necessary part of the net.

IRT and DRT with their two values, type and quality parameter, are the committee, whose output the network is translating into an optimized result.

The implemented training procedure is a little bit different from the classically one for a neural net. Although it has the classical properties of a supervised training, the extremalization of a governing energy functional, it is different by the fact that the training is complete. Normally a network is trained by usage of a large sample of examples, and the weights, or rules in this case, are adjusted by the changes of the energy functional. The aim is that the net develops a sufficient generalization to treat unknown data. In the case of the RPS, this method was replaced by a full testing of a set of discrete rules. Normally such a test is not possible, because the rules space is continuous. But in the case of the RPS, all relevant scales are coarse enough to allow a discretization. The resulting number of phase space points is small enough to test all possible network topologies, and therefore to perform the original idea of training in a statistical sense [37], which is normally not possible and the conventional network training is the appropriate approximation.

Chapter 6
Other Methods: A View To The Future

There are several methods which would in principle be of interest to be used for RICH PID, but are not used today. The reason for this is either that the computing power available today is not sufficient or the foundations of these methods are not yet as complete as necessary to adapt it to the problem of RICH PID or that just nobody has tried them yet.

With the present third and the future fourth generation of RICH detectors, the challenges to RICH PID will increase. To limit the size, RICH detectors are built as compact as possible, and the properties of the optics let the rings more and more deviate from genuine rings. The extreme example are the DIRC detectors, where each ring is transformed into two conical sections, depending on the number of reflections each photon experiences, whether even or odd. The new ultra high energy experiments at LHC like LHCb will experience particle and hence ring multiplicities that are an order of magnitude larger than for many of the present experiments. The studies on QCD under extreme conditions will give rise to an even higher background-to-signal ratio at the highest occupancies of RICH detectors.

With constant progress in resolution and optical quality, and the exponential growth in computing and computing knowledge, it is possible to meet this challenge.

In this last chapter some of the methods will be presented which make use of the current and short-term future development in computing sciences. These sections will not dwell on specific points of how to implement these methods for RICH detectors, but will describe the basic properties of the technologies and which properties of these system make them useful for RICH PID.

6.1 Massive Numerical Methods

In this section a number of possible extensions or new approaches are grouped, which have one property in common: The need for a large amount of computing time and space. Moore's law predicts an increase of computing performance by 2 every 18 months. Over several decades, the development has followed this law, or was even faster in some areas, like disk space. Using current technologies, this will be possible for another decade before the limits of physics have been reached. Several new concepts are currently developed to allow a further increase. Of special interest to the pattern recognition problem of RICH PID are devices which allow the simultaneous computation of multiple patterns at the same time. The most prominent are DNA and quantum computers. These are, however, far away in the future.

These developments do not only impact RICH PID, but will also allow to improve the hardware. This section will, however, only cover the PID software part.

The increase of computing performance will allow to overcome several limitations of today. For example a DRT-based algorithm would then be able to use a much more precise but time-consuming Monte Carlo, thus making approximations like blobbing unnecessary. It will also allow to increase statistics for this method beyond random fluctuations.

Iterative methods using an a-priori guess of the most probable particle species were originally developed to reduce time consumption by not exploring all possible hypotheses. This approximation to some extent limits the performance, even in the most sophisticated versions. Using a much higher computing performance, the complete search space is accessible, which removes the necessity for this approximation.

An increase of speed does as well allow to open up areas of multiplicities for methods, which scale more than linearly with the number of hits. The Hough algorithm for example scales as $O(N_{Hit})$ where N_{Hit} is the number of photon hits. A method like the cluster algorithm scales as $O\left(N_{Hit}^4\right)$ and is therefore not usable in cases of high hit density.

Another example is a track-independent method based on B-splines, which has not yet been explored to a published level. In this case any three hits define a triangle, if their mutual distances are below the asymptotic radius and they are not collinear. From these three hits a B-spline is formed. The technique of B-splines originates from image editing. It describes how to form curves from a set of discrete points. By adding more and more points, the changes of the B-splines can be observed. If a significant change in curvature is detected, which happens if a hit is included which is not part of the ring, the hit would be rejected. Particle identification would be done, if an appropriate number of hits would be found on the B-spline. The parameters of the B-spline would then determine the particle's type.

6.2 Neural Networks

Neural networks have already established their importance in modern image recognition applications. Their ability to generalize from a given sample even in a noisy environment [37] make them an excellent candidate for future RICH PID algorithms. The methods presented in this work thus far had the common property to be analytical solutions to the problem of RICH PID. Neural nets present a numerical-statistical approach. With technologies available today, it would be possible to use a complete RICH detector as the input neuron field for a neural net. Because the simplicity of the geometric shape to be recognized and the often given translational and sometimes rotational invariance, already a small number of hidden neurons would be able to provide a good PID. Together with fast computers and large disk space, it is easily possible to create and use the large amount of training samples, which would be necessary to use a direct neural network, using the pixels itself. If a preprocessing is used for example to find ring arcs, already simpler networks and smaller training samples would suffice.

A situation is yet missing where their fuzziness[19] provides a killer application compared to analytical methods, although neural nets already became competitive due to their development during the last decade. Their commercial spread may however be the key for their success. Ready-to-use neural networks which only have to be trained become a common trade good, and their use would perhaps reduce the long development time necessary for conventional algorithms. By training a neural net out-of-the-box with a sample from Monte Carlo data, it might be possible to achieve a reasonable PID within days or weeks rather than months or years with much less manpower.

6.3 Genetic Algorithms

Genetic algorithms [129] are another fuzzy method. They are deployed to optimize algorithms and aid in construction problems like processor design. Their basic principle is to start with a large population of possible solutions to a problem, from these then some are selected which best solve the given problem. All others are discarded. To those kept, a set of descendants with small variations is added. Again those best solving the problem are kept and so on.

If the result is clearly defined, like maximum speed while ensuring some results, it is easy to define a function which will guarantee the survival of the fittest. In the case of RICH PID, the problem is different, and it will be hard to define a selection function analytically to select the optimal result on an event-by-event basis for true data. In this case only a numerical solution would be possible to define a selection criterion. An interplay between a neural network and a genetic algorithm may be able to solve

[19] Fuzzines or fuzzy methods are mathematical approaches which are not analytical but contain random effects, e.g. random numbers or operations.

this problem. In this case the neural network would be trained to be the selection criterion for the genetic algorithm.

A much more likely application for genetic algorithms is the design of RICH PID methods. An algorithm like the RPS presented in the previous chapter could be optimized using genetic algorithms instead of a complete training. The same is true for analytical algorithms with several free parameters. Genetic algorithms may provide a solution where the phase space for the parameters of an algorithm becomes too large to be searched even numerically.

6.4 Artificial Intelligence

For a human, it is possible after some training to recognize particle types from a RICH event display right away. In these cases beside the optical reception a knowledge-based deduction is the important ability. From knowing the general properties for which to look, a human can find the solution and can use, if in doubt, appropriate tools like image editing. The only reason for not performing the RICH PID by hand is the vast large amount of events. A solution well into the future is perhaps true artificial intelligence [75]. Beyond neural nets as an extension of statistical methods [37], artificial intelligence research is directed to construct systems which do have the ability of a knowledge-based deduction. A system understanding the problem of RICH PID and using tools rather than executing a set of instruction would be able to perform as well as a human, but with a more acceptable speed. The necessary knowledge for RICH PID is very limited as are the number of useful tools. Hence already simple systems may be able to perform the task. Artificial intelligence is nevertheless a field which is not yet as far developed as the previously presented approaches, and there is still a long way ahead before its applications are beyond the level of a research project.

6.5 Custom Fit PID

Custom fit PID is in principle not a new method nor a new approach. It is more a quality of service which will become increasingly important in future RICH PID systems. With the increase of omnipurpose detectors, the physical analysis diversifies more and more, and each has different requests to the PID. Especially the particle type and/or kinematical range for which the PID should be optimized is likely to be different. Since any PID will have large regions of phase space where its performance varies with the degree of optimization for a given problem, fast adaptation is important.

If a PID has to be optimized carefully by hand for each problem set, the time scales necessary to perform preanalysis studies become unpractically large. PID methods will have to be versatile and easily optimizable. An example is the RPS presented in the previous chapter, where a simple, easy understandable parameter set controls the properties of the PID.

As a next step, the abstraction of PID would step up one level. As a result no longer single PID methods would be designed, but methods to design a PID for a given set of requirements. This would be for example an algorithm which translates the requirements of a given physics analysis to an energy functional for a neural network and a training sample and then train it. This development would shorten the implementation cycle drastically and would increase the "physics turn-around".

Chapter 7
Summary

This work reviews the present status of RICH PID in middle to high energy physics. Chapter two gives an introduction to the foundations of RICH detectors, from the Čerenkov effect to the proximity- and mirror-focused RICH detectors. It also provides a review of the past, present and future RICH detectors at experiments and their purposes.

Chapter three presents the foundations for the spectrum of standard RICH PID methods in use today. It looks at methods for detectors with and without tracking. It reviews the problems RICH PID has to face and presents optimizations of the previous methods to cope with these problems. It finally develops the necessary tools to judge and compare RICH PID methods independent of the physics case and the detailed implementation.

Chapter four presents a development of RICH PID methods from the design on, visualized in the example of the CERES experiment. The presented solutions showed how the common IRT and DRT principles are adapted to the given setup. Additionally a solution is demonstrated for the problem of the acceptance cut off. Finally the feasibility of a track-dependent method for CERES is shown using data from a full Monte Carlo.

Chapter five demonstrates how to optimize a set of given methods using the decision network technique. This is detailed in the example of the HERMES experiment, where the present IRT and DRT methods are combined, using the available resources in form of quality parameters the algorithms provided. It demonstrates with Monte Carlo data that the combined method, the RPS, is superior to both in terms of statistics or performance. These results are at least in qualitative agreement with results from real data. Extended systematic studies show the stability of the algorithm. From a physics point of view the flat parametrization of the algorithm allows an omnipurpose use.

Chapter six finally sketches a glimpse of the future and presents possible future algorithms and trends in RICH PID.

This work in total gives a full review of the technique of RICH PID and develops new concepts for the examples of the CERES and HERMES experiments.

Chapter 8
Acknowledgments

I would like to thank Professor Dr. P. Braun-Munzinger for the opportunity to do a diploma thesis with both the CERES and the HERMES experiment. In this context I would like to thank Dr. W.-D. Nowak for the possibility to join the HERMES group from a non-collaborating institute to do this thesis on general algorithms.

I would like to thank the CERES collaboration, especially the GSI group, D. Miskowiec, H. Sako and G. Hering. I would further like to thank P. Glässel, K. Filiminov, T. Wienold, H. Appelshäuser, S. Esumi, A. Marin and J. Stachel from the University of Heidelberg and I. Ravinovich from the Weizman Institute.

I would like to thank the HERMES collaboration, especially the DESY Zeuthen group. Of those I would like to mention P. Jung, K. Negodaeva, Y. Gärber and F. Ellinghaus. From the rest of the collaboration I would like to mention S. Yoneyama from the Tokyo Institute of Technology and B. Hommez from the University of Gent.

Especially I would like to thank R. Kaiser, who has been my supervisor at DESY Zeuthen. I would also like to thank his wife M. Kaiser.

For valuable comments and suggestions I would like to thank P. Braun-Munzinger, W.-D. Nowak, R. Kaiser and R. Knobloch.

I would like to thank the HEP theory group of the University of Darmstadt for providing me with an enjoyable theory seminar during my diploma thesis as an experimentalist.

I thank my parents for giving me the opportunity and the possibility to pursue my path up to this point.

I thank some friends, who will never read this, but should be mentioned anyway: Sarah, Michaela, Mark, Govinda, Simon, Melanie, Carmen for giving me the opportunity for enjoyable breaks during the last year.

The most I thank my girlfriend Renate Knobloch, who has suffered much from the last four years and especially during the last month. Thank you for everything.

Bibliography

The abbreviation NIM was used for Nuclear Instruments and Methods in Physics Research.

[1] S. Abatzis et al., *Study of charged particle production using Omega RICH in WA94 experiment,* NIM A371 p22, 1996

[2] K. Abe et al., *Obtaining physics results from the SLD CRID,* NIM A371 p195, 1996

[3] K. Ackerstaff et al., *The HERMES spectrometer,* NIM A417 p230,1998

[4] K. Ackerstaff et al., *Flavor decomposition of the polarized quark distributions in the nucleon from inclusive and semi-inclusive deep-inelastic scattering,* Physics Letter B464 p123, 1999

[5] I. Adam et al., *The DIRC detector at BaBar,* NIM A433 p121, 1999

[6] W. Adam et al., *Particle identification algorithms for the DELPHI RICH detector,* NIM A371 p240, 1996

[7] G. Agakichiev et al., *Čerenkov ring fitting techniques for the CERES RICH detector,* NIM A371 p243, 1996

[8] G. Agakichiev et al., *Performance of the CERES electron spectrometer in the CERN SPS lead beam*, NIM A371 p16, 1996

[9] Agakichiev et al., *Technical note on the NA45/CERES upgrade*, CERN/SPLSC 96-50, SPSLC/R110, 1996

[10] E. Albrecht, *Operation, optimization, and performance of the DELPHI RICH detectors,* NIM A433 p47, 1999

[11] M. Alford, K. Rajagopal, F. Wilczek, *Color superconductivity and signs of its formation,* hep-ph/9802284, 1998

[12] M. Alford, J. Berges, K. Rajagopal, *Unlocking color and flavor in superconducting stranger quark matter,* hep-ph/9903502, 1999

[13] M. Alford, J. Berges, K. Rajagopal, *Gapless color superconductivity,* hep-ph/9908235, 1999

[14] M. Alford, J. Berges, K. Rajagopal, *Magnetic fields within color superconducting neutron star cores,* hep-ph/9910254, 2000

[15] A. Airapetian et al., *Hadron formation in deep-inelastic positron scattering from ^{14}N and ^{2}H*, to be published, HERMES, 2000

[16] A. Airapetian et al., *Double-spin asymmetry in the cross section for exclusive ρ^0 production in lepton-proton scattering,* to be published in European Journal of Physics, HERMES, 2000

[17] M. Anselmino et al., *Electrons scattering with polarized targets at TESLA,* internal note, TESLA-$\vec{N}$, 2000

[18] P. Antonioli et al., *The AQUA-RICH atmospheric neutrino experiment,* NIM A433 p104, 1999

[19] Y. Akiba et al., *Ring imaging Čerenkov detector of PHENIX experiment at RHIC,* NIM A433 p143, 1999

[20] N. Akopov et al., *The HERMES Dual-Radiator Ring Imaging Čerenkov detector,* to be published in NIM, 2000

[21] B. Artmann, *Lineare Algebra,* Birkhäuser Skripten, Basel, 1991

[22] M. Artuso, Y. Mukhin, *A fast ring imaging detector for the CLEO upgrade*, NIM A343 p319, 1994

[23] M. Artuso, *The ring imaging detector for CLEO III,* NIM A371 p324, 1996

[24] J. Ashman et al., *A measurement of the spin asymmetry and determination of the structure function g_1 in deep inelastic muon-proton scattering*, Physics Letter B206 p364, 1998

[25] P. Atkins, *Physical Chemistry*, Oxford University Press, Oxford, 1978

[26] Auger project, *http://www.auger.org,* 2000

[27] J. Bächler et al., *Reconstruction of Čerenkov rings in imaging detectors,* NIM A343 p273, 1994

[28] J. Bächler et al., *Study of particle spectra with an optically readout RICH detector in the NA35 experiment,* NIM A343 p288, 1994

[29] G. Barbiellini et al., *Performance of the CAPRICE RICH detector during the 1994 balloon flight*, NIM A371 p169, 1996

[30] J. Bartels, D. Haidt, A. Zichichi, ed., *The european journal of physics C, particles and fields, number 1-4,* Volume 15, 2000

[31] M. Battaglia, P. Kluit, *Particle identification using the DELPHI RICH detectors,* NIM A433 p252, 1999

[32] G. Baum et al., *The COMPASS RICH project*, NIM A433 p207, 1999

[33] Baur et al., *Study of electron pair and photon production in lead-lead collisions at the CERN SPS,* CERN/SPSLC 94-1, SPSLC/P280, 1994

[34] G. Baym, D. Bödeker, L. McLerran, *Magnetic fields produced by phase transition bubbles in the electroweak phase transition,* hep-ph/9507429, 1995

[35] B. Berdnikov, K. Rajagopal, *Slowing out of equilibrium near the QCD critical point,* hep-ph/9912274, 1999

[36] J. Berges, K. Rajagopal, *Color superconductivity and chiral symmetry restoration at nonzero baryon density and temperature,* hep-ph/9804233, 1998

[37] C. Bishop, *Neural networks for pattern recognition,* Clarendon press, Oxford, 1995

[38] J.-P. Blaizot, *Signals of the Quark-Gluon-Plasma in nucleus-nucleus collision,* hep-ph/9909434, 1999

[39] J.-P. Blaziot, P. Dinh, J.-Y. Ollitrault, *Transverse energy fluctuations and the pattern of J/ψ suppression in Pb-Pb collisions,* nucl-th/0007020, 2000

[40] D. Boutigny for BABAR, *Status of the BABAR detector*, to be published in World Scientific, 2000

[41] P. Braun-Munzinger, *Towards the Quark-Gluon-Plasma,* nucl-ex/9909014, 1999

[42] J. Carvalho, *RICH rings pattern recognition algorithms - a few preliminary tests,* HERA-B internal note, 1999

[43] C. Caso et al., *Particle Physics booklet,* Springer, New York, 1998

[44] N. Chernov, E. Kolganova, G. Ososkov, *Robust methods for RICH ring recognition and particle identification*, NIM A433 p274, 1999

[45] H. Collad, *Long range plan,* HERMES internal note 00-003, 2000

[46] P. Coyle et al., *The DIRC counter: A new type of particle identification device for B factories,* NIM A343 p292, 1994

[47] D. Cozza et al., *Development of a pattern recognition method for the HMPID in ALICE (I),* ALICE internal note 98-XX, 1998

[48] R. Debbe et al., *The ring imaging Čerenkov counter for the BRAHMS experiment at RHIC,* NIM A371 p327, 1996

[49] D. De Schepper, R. Kaiser, E. Cisbani, *Particle identification with the HERMES RICH detector: Description of the different approaches,* HERMES Internal Note 98-008, 1998

[50] A. DiMauro et al., *A fast RICH detector for particle identification in ALICE at LHC,* NIM A343 p323, 1994

[51] A. Di Mauro et al., *Particle identification with a solid photocathode RICH in ALICE at LHC,* NIM A343 p284, 1994

[52] B. Dolgoshein, *Complementary particle ID: Transition radiation and dE/dx relativistic rise*, NIM A433 p533, 1999

[53] K. Dransfeld, P. Kienle, H. Vonach, *Physik I,* Oldenbourg Verlag, München, 1992

[54] K. Dransfeld, P. Kienle, *Physik II,* Oldenbourg Verlag, München, 1994

[55] D. Dujmic et al., *Measuring momentum with the HERA-B RICH and ECAL,* HERA-B internal note, 1999

[56] D. Dujmic, R. Eckmann, R. Schwitters, *Reconstruction of photon directions in the HERA-B RICH*, HERA-B internal note 00-019 RICH 00-005, 1999

[57] D. Dujmic, R. Eckmann, R. Schwitters, *Performance of the HERA-B RICH,* HERA-B internal note 00-020 RICH 00-006, 2000

[58] R. Durbin, D. Willshaw, *An analogue approach to the travelling salesman problem using an elastic net method,* Nature Vol. 236 p689, 1987

[59] D. Dwyer, K. Rajagopal, *Collision-induced decay of metastable baby skyrmions,* hep-ph/0004126, 2000

[60] D. Elia et al. für ALICE, *A pattern recognition method for the RICH-base HMPID detector in ALICE,* NIM A433 p262, 1999

[61] Ellis, Ghosh, Mavromatos, *On the thermodynamics of a gas of AdS black holes and the quark-hadron phase transition,* hep-th/9902190, 1999

[62] J. Engelfried et al., *The RICH detector of the SELEX experiment,* NIM A433 p149, 1999

[63] B. Erzen et al., *Analysis of the DELPHI RICH data at LEP II*, NIM A433 p247, 1999

[64] R. Fernholz, K. Ragan, K. Strahl, *A fast RICH detector for the CDF run II upgrade,* NIM A371 p305, 1996

[65] S. Fortunato, H. Satz, *Polyakov loop percolation and deconfinement in SU(2) gauge theory*, hep-lat/9911020, 2000

[66] R. Forty, O. Schneider, *RICH pattern recognition,* LHCb internal note 98-

040, 1998

[67] R. Forty für LHCb, *RICH pattern recognition for LHCb,* NIM A433 p257, 1999

[68] J. Friese für HADES, *HADES, the new electron-pair spectrometer at GSI,* internal note, 1998

[69] J. Pérez, A. Agüera, *Results from background reduction in the DELPHI forward RICH,* NIM A371 p232, 1996

[70] R. Gernhäuser et al., *HIRICH, a novel device for velocity measurement of relativistic ions*, NIM A433 p217, 1999

[71] P. Glässel, *The limits of the ring image Čerenkov technique,* NIM A433 p17, 1999

[72] Gorbunov et al., *Ring recognition method based on the elastic neural net,* COMPASS internal note, 1999-3, 1999

[73] Gorbunov et al., *Progress report on ring reconstruction with the elastic net,* COMPASS internal note, 1999-18, 1999

[74] A. Gorišek, P. Križan, *HERA-B RICH optical system alignment with data*, IJS report IJS-DP-7688, 1997

[75] G. Görz, *Einführung in die künstliche Intelligenz,* Addison-Wesley Publishing Company, Bonn, 1995

[76] N. Grewe, *Vorlesungskurs Quantenphysik,* Technische Universität Darmstadt, Sommersemester 1997 bis Wintersemester 1998/1999

[77] D. Gries, *The science of programming,* Springer Verlag, New York, 1981

[78] S. Großmann, *Mathematischer Einführungskurs für die Physik,* Teubner Verlag, Stuttgart, 1993

[79] T. Guhr, *Lecture on group theory in quantum physics,* TU Darmstadt, 1999

[80] M. Gyulassy, L. McLerran, *Yang-Mills radiation in ultra-relativistic nuclear collisions,* nucl-th/9704034, 1997

[81] D. Haidt, P. Zerwas, ed., *The European Physical Journal C, particles and fields, Volume 3, Number 1-4, pp1-749,* Springer Verlag, New York, 1998

[82] Halasz et al., *On the phase diagram of QCD,* hep-ph/9804290, 1998

[83] D. Halliday, R. Resnick, J. Walker, *Fundamentals of Physics extended, fourth edition,* John Wiley & Sons, 1993

[84] H. Heiselberg et al., *Hot neutron stars as a source for gamma ray bursts at cosmological distance scales,* astro-ph/9711169, 1997

[85] P. Huet et al., *Hydrodynamic stability analysis of burning bubbles in electroweak theory and in QCD,* hep-ph 9212224, 1992

[86] C. Hung, E. Shuryak, *Dilepton/photon production in heavy ion collisions and the QCD phase transition*, hep-ph/9608299, 1996

[87] J. Izen et al., *Some statistics for particle identification,* BaBar internal note, 1998

[88] H. Jackson, *Study of a RICH conversion in HERMES,* HERMES internal note 95-031, 1995

[89] J. Jackson, *Classical Electrodynamics, Second Edition*, John Wiley & Sons, New York, 1975

[90] P. Jung, *Pion-Kaon Separation mit dem HERMES RICH,* diploma thesis, Berlin University of Technology, 2000

[91] R. Kaiser, *HERMES PID,* HERMES internal note 97-025, 1997

[92] R. Kaiser, *Conventions for the HERMES RICH particle identification,* HERMES internal note 99-010, 1999

[93] R. Kaiser ed., *Workshop on HERMES RICH algorithms, Zeuthen July 1999,* 1999

[94] R. Kaiser, *Physics results from HERMES,* Czechoslovak Journal of Physics, Vol. 49, No. 0, 1999

[95] R. Kaiser, *Muon identification with the HERMES RICH*, HERMES internal note 99-018, 1999

[96] R. Kaiser, *Likelihood analysis for the RICH IRT method,* HERMES internal note 00-026

[97] G. Kalvius, K. Luchner, H. Vonach, *Physik IV*, Oldenbourg, München, 1995

[98] J. Kapusta, D. Kharzeev, L. McLerran, *The return of the prodigal goldstone boson,* hep-ph/9507343, 1995

[99] D. Kharzeev, H. Satz, *Color deconfinement and quarkonium dissociation,* hep-ph/9505345, 1995

[100] D. Kharzeev, M. Nardi, H. Satz, *Anomalous J/ψ suppression and the nature of deconfinement,* hep-ph/9707308, 1997

[101] P. Kluit, *Physics with ring imaging Čerenkov detectors,* Nuclear Instruments and ;Methods in Physics Research A371 p223, 1996

[102] S. Korpar et al., *The HERA-B RICH,* NIM A433 p128, 1999

[103] A. Kozhevnikov et al., *SPHINX phototube RICH detector for diffractive production experiments at Sepukhov accelerator,* NIM A433 p164, 1999

[104] H. Kraner et al., *CERES/NA45 status report,* CERN/SPSLC 94-2 SPSLC/M529, 1994

[105] P. Križan et al., *The RICH detector for HERA-B,* NIM A371 p295, 1996

[106] P. Križan, M. Starič, *An iterative method for the analysis of Čerenkov rings,* IJS report IJS-DP-7819, 1997

[107] Laermann, *Recent results from lattice QCD simulations,* p1c, Nuclear Physics A610, 1996

[108] J. Lehnert et al. for HADES, *Ring recognition in the HADES second-level trigger,* NIM A433 p268, 1999

[109] B. Lenkeit, *Produktion von geladenen Pionen mit hohem Transversalimpuls in 200A GeV S-Au-Kollisionen,* University of Heidelberg, diploma thesis, 1995

[110] E. Lorenz, *Air shower Čerenkov detectors,* NIM A433 p24, 1999

[111] C. Lu et al., *Detection of internally reflected Čerenkov light, results from*

the BELLE DIRC prototype, NIM A371 p82, 1996

[112] A. Lue, K. Rajagopal, M. Trodden, *Semi-analytical approaches to local electroweak baryogenesis,* hep-ph/9612282, 1997

[113] A. Maas, *RPS - RICH PID Scheduler,* HERMES internal note 00-029, 2000

[114] L. McLerran, *Small x physics and why it's interesting,* hep-ph/9705426, 1997

[115] L. McLerran, *Three lectures on the physics of small x and high gluon density,* hep-ph/9903536, 1999

[116] L. McLerran, *Two lectures on small x and high gluon density.* nucl-th/9911013, 1999

[117] H. Messinger, *Langenscheidts Handwörterbuch Englisch,* Langenscheidt, Berlin, 1996

[118] G. Močnik, *Particle identification in the HERA-B RICH detector,* HERA-B internal note 1994

[119] A. Morsch for ALICE, *Physics sources of noise in ring imaging Čerenkov detectors*, NIM A433 p235, 1999

[120] A. Most, *Proceedings of SPIN 97, Amsterdam 1997*, ed. C. de Jager et al., World Scientific 800, 1997

[121] U. Müller et al., *Particle identification with the RICH detector in experiment WA89 at CERN*, NIM A343 p279, 1994

[122] U. Müller et al., *The recent performance of the Omega RICH detector experiment WA89 at CERN,* NIM A371 p27, 1996

[123] G. Nayak et al., *Equilibration of the gluon-minijet plasma at RHIC and LHC*, hep-ph/0001202, 2000

[124] K. Negodaeva, *Inclusion of the Aerogel Boundary Refraction into the HERMES RICH Reconstruction*, HERMES internal note 99-022, 1999

[125] K. Negodaeva, *The PID efficiency of the HERMES RICH detector for the IRT method,* HERMES internal note 00-028, 2000

[126] K . Negodaeva, *Experimental PID efficiencies,* minutes of HERMES analysis meeting, 9th of August 2000

[127] R. Nóbrega, P. Sonderegger, J. Varela, *A RICH counter for trigger and detection of exotic particles,* NIM A371 p285, 1996

[128] W. Nolting, *Grundkurs Theoretische Physik Band 6: Statistische Physik,* Vieweg Verlag, Braunschweig, 1998

[129] S. Pal, P. Wang, ed., *Genetic algorithms for pattern recognition,* CRC Press, 1996

[130] S. Passaggio, R. Jones, *Charged K identification with the JETSET RICH,* NIM A371 p188, 1996

[131] G. Polok, *Results of the two-dimensional particle identification analysis applied for the RICH in the DELPHI experiment,* NIM A433 p289, 1999

[132] K. Rajagopal, *The QCD tricritical point: Beyond monotony in heavy ion physics,* hep-ph/9808348, 1998

[133] K. Rajagopal, *Three-flavor QCD at high density: Color-flavor locking and chiral symmetry breaking,* hep-ph/9807318, 1998

[134] K. Rajagopal, *QCD at finite baryon density: Chiral symmetry restoration and color superconductivity,* hep-ph/9803341, 1998

[135] K. Rajagopal, *Mapping the QCD phase diagram*, hep-ph/9908360, 1999

[136] K. Rajagopal, *How to find the QCD critical point,* hep-ph/9903547, 1999

[137] K. Rajagopal, S. Abe, *Statistical mechanical foundations for systems with nonexponential distributions,* cond-mat/0003493, 2000

[138] K. Rajagopal, E. Shuster, *On the applicability of weak-coupling results in high density QCD,* hep-ph/0004074, 2000

[139] L. Randall, R. Rattazzi, E. Shuryak, *Implication of exact SUSY gauge couplings for QCD,* hep-ph/9803258, 1998

[140] R. Rapp et al., *Diquark Bose condensates in high density matter and instantons,* hep-ph/9711396, 1997

[141] R. Rapp et al., *High density QCD and instantons*, hep-ph/9904353, 1999

[142] R. Rapp, J. Wambach, *Chiral symmetry restoration and dileptons in relativistic heavy-ion collisions,* hep-ph/9909229, 1999

[143] R. Rapp, E. Shuryak, *Thermal dilepton radiation at intermediate masses at the CERN-SpS,* hep-ph/9909348, 2000

[144] B. Ratcliff, *Future high energy physics experiments using RICH detectors: The next generation,* NIM A371 p309, 1996

[145] Rational, *http://www.rational.com,* 2000

[146] L. Resvanis, *High-energy neutrino telescopes,* NIM A433 p34, 1999

[147] ROOT, *http://root.cern.ch,* 2000

[148] Rational Rose, *http://www.rational.com/rose*

[149] D. Ryckbosch for HERMES, *The HERMES RICH detector*, NIM A433 p98, 1999

[150] Sarle, *Neural network implementation in SAS software*, Proceedings of the 19th annual SAS user group international conference, 1994

[151] H. Satz, *Hard probes of dense matter,* hep-ph/9502322, 1995

[152] H. Satz, *Color deconfinement in hot and dense matter,* hep-ph/9611366, 1996

[153] H. Satz, *Color deconfinement and J/ψ suppression in high energy nuclear collisions,* hep-ph/9711289, 1997

[154] H. Satz, *QCD & QGP: A summary*, hep-ph/9706342, 1997

[155] H. Satz, *Deconfinement and percolation,* hep-ph/9805418, 1998

[156] H. Satz, *A brief history of J/ψ suppression,* hep-ph/9806319, 1998

[157] P. Schiavon, *Studies of the Čerenkov emission angle reconstruction part 1,* COMPASS/RICH/PS/Report 2.b, 1997

[158] P. Schiavon, *Studies of the Čerenkov emission angle reconstruction part 2,* COMPASS/RICH/PS/Report 3.a 1997

[159] H. Schildt, *The complete reference C++,* Osborne McGraw Hill, Berkley, 1998

[160] T. Schäfer, E. Shuryak, *Glueballs and instantons,* hep-ph/9410372, 1995

[161] T. Schäfer, E. Shuryak, *Instantons in QCD,* hep-ph/9610451, 1998

[162] R. Schneider, W. Weise, *Thermal dileptons from quark and hadron phases of an expanding fireball,* hep-ph/0008083, 2000

[163] R. Schwitters, *Rings upon rings: A new approach to pattern recognition in Ring Imaging Čerenkov detectors,* HERA-B internal note, 1998

[164] R. Schwitters, *"Removal" of background rings from single-ring likelihood fits,* HERA-B internal note 00-084 RICH 00-010, 2000

[165] R. Schwitters, *Single-ring likelihood fit including background,* HERA-B internal note 00-021 RICH 00-007, 2000

[166] J. Seguinot, T. Ypsilantis, *Evolution of the RICH technique,* NIM A433 p1, 1999

[167] M. Shiozawa for Super-Kamiokande, *Reconstruction algorithms in the Super-Kamiokande large water Čerenkov detector,* NIM A433 p240, 1999

[168] E. Shuryak, *Which chiral symmetry is restores in hot QCD?,* hep-ph/9310253, 1993

[169] E. Shuryak, J. Verbaarschot, *Screening of the topological charge in a correlated instanton vacuum,* hep-lat/9409020

[170] E. Shuryak, *The QCD vacuum, chiral phase transition and quark-gluon plasma*, hep-ph/9503427, 1995

[171] E. Shuryak, *Physics of heavy ions collisions: The summary of Moriond-97*, hep-ph/9704449, 1997

[172] E. Shuryak, *The problem of chiral restoration and dilepton production in heavy ion collisions*, hep-ph/9812482, 1998

[173] E. Shuryak, *QCD phases at high density and instantons,* hep-ph/9807243, 1998

[174] E. Shuryak, *Equation of state, flow, fluctuations and J/ψ suppression,* hep-ph/9801393, 1998

[175] E. Shuryak, D. Teaney, *J/ψ suppression in heavy ion collisions and the QCD phase transition,* nucl-th/9801016, 1998

[176] E. Shuryak, *QCD at finite density and color superconductivity,* hep-ph/9908290, 1999

[177] E. Shuryak, *What we have learned and want to learn from heavy ion collisions at CERN SPS?*, hep-ph/9906443, 1999

[178] E. Shuryak, *New phases of QCD, the tricritical point; and RHIC as a "nutcracker",* hep-ph/9903297, 1999

[179] E. Shuryak, *The phases of QCD*, hep-ph/9609249, 2000

[180] SNO experiment, *http://www.snow.phy.queensu.ca,* 2000

[181] J. Sollfrank et al., *Canonical strangeness enhancement,* nucl-th/9802046, 1998

[182] J. Spengler, *HERA-B: Status and first results*, 3. Latin American Symposium on high energy physics, 2000

[183] C. Spiering, *Čerenkov imaging and timing techniques in astroparticle physics,* NIM A371 p162, 1996

[184] J. Stachel, *Towards the Quark-Gluon-Plasma,* nucl-ex/9903007, 1999

[185] M. Starič, P. Križan, *An iterative method for the analysis of Čerenkov rings in the HERA-B RICH,* NIM A433 p279, 1999

[186] M. Stephanov, K. Rajagopal, E. Shuryak, *Signatures of the tricritical point in QCD,* hep-ph/9806219, 1999

[187] S. Stone, *Physics results from RICH detectors*, NIM A433 p293, 1999

[188] Taligent naming scheme, *http://hpsalo.cern.ch/TaligentDocs/TaligentOnline /DocumentRoot/1.0/Docs/books/WM/WM_ 63.html#HEADING77,* 2000

[189] D. Teaney, E. Shuryak, *An unusual space-time evolution for heavy ion collisions at high energies due to the QCD phase transition,* nucl-th/9904006, 1999

[190] T. Ullrich, *Produktion von e^+e^--Paaren in ultrarelativistischen S-Au-Kollisionen bei 200A GeV/c,* University of Heidelberg, doctoral thesis, 1994

[191] M. Velkovsky, E. Shuryak, *QCD with large number of quarks: Effects of the instanton-anti-instanton pairs,* hep-ph/9703345, 1997

[192] R. Waldi, *Status of the BABAR experiment,* Heavy Flavours 8 Conference Proceedings, 1999

[193] T. Wienold, *Estimate of the dilepton signal from the 99 data sample,* CERES internal note, 2000

[194] F. Wilczek, *QCD in extreme conditions*, hep-ph/0003183, 2000

[195] T. Ypsilantis et al., *A long base line RICH with a 27 kton water target and radiator for detection of neutrino oscillations,* NIM A371 p330, 1996

[196] K. Zeitelhack et al., *The HADES RICH detector,* NIM A433 p201, 1999

[197] W. Zinth, H-J. Körner, *Physik III,* Oldenbourg Verlag, München, 1996

List of Figures

Index