GPU-based Online Track Reconstruction for $\overline{\mathrm{P}}$ANDA and Application to the Analysis of $\mathrm{D} \rightarrow \mathrm{K}\pi\pi$

DISSERTATION

zur
Erlangung des Grades eines
Doktors der Naturwissenschaften
in der Fakultät für Physik und Astronomie
der Ruhr-Universität Bochum

von

Andreas Herten

aus

Eschweiler

Bochum (2015)

Versicherung gemäß §7 Abs. 2 Nr. 6 PromO 2011
Hiermit versichere ich, dass ich meine Dissertation selbstständig und ohne unerlaubte fremde Hilfen angefertigt und verfasst habe und keine anderen als die angegebenen Hilfsmittel und Hilfen benutzt habe. Meine Dissertation habe ich in dieser oder ähnlicher Form noch bei keiner anderen Fakultät der Ruhr-Universität Bochum oder bei einer anderen Hochschule eingereicht.

1. Gutachter: Prof. James Ritman, Ph.D.
2. Gutachter: Prof. Dr. Ulrich Wiedner

Datum der Disputation: 2. Juli 2015

Abstract

The $\overline{P}$ANDA experiment is a new hadron physics experiment which is being built for the FAIR facility in Darmstadt, Germany. $\overline{P}$ANDA will employ a novel scheme of data acquisition: the experiment will reconstruct the full stream of events in realtime to make trigger decisions based on the event topology. An important part of this online event reconstruction is online track reconstruction. Online track reconstruction algorithms need to reconstruct particle trajectories in nearly realtime. This work uses the high-throughput devices of Graphics Processing Units to benchmark different online track reconstruction algorithms. The reconstruction of $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ is studied extensively and one online track reconstruction algorithm applied.

Das $\overline{P}$ANDA-Experiment ist ein neues Hadronenphysikexperiment, das für das Beschleunigerzentrum FAIR in Darmstadt gebaut wird. $\overline{P}$ANDA wird eine neue Art der Datenaufnahme verwenden: Der gesamte Strom physikalischer Ereignisse wird in Echtzeit rekonstruiert und für Entscheidungen der Aufnahme von Daten verwendet. Ein Element dieser Online-Ereignisrekonstruktion ist Online-Spurrekonstruktion, bei der die Spuren von Teilchen in nahezu Echtzeit berechnet werden müssen. Die vorliegende Arbeit verwendet Grafikkarten, GPUs, als Geräte mit hoher Rechenleistung um verschiedene Online-Spurrekonstruktionsalgorithmen zu testen. Die Rekonstruktion von $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ wird analysiert und einer der Online-Algorithmen auf den Kanal angewendet.

Contents

Chapter 1

Preface

Since 100 years, the structure of the atom is largely known. Rutherford and other researchers of that era methodically studied the atom to understand the phenomenon of radiation. The assessed structure included a massive nucleus surrounded by charged electrons. Soon, the constituents of the nuclei were found: protons and neutrons. Initially thought to be the smallest possible particles at that time, a more complicated picture emerged when strangely behaving particle were observed in particle showers. Finally, in the 1960s, quarks were introduced as the new smallest, indivisible massive particles making up matter. The ground-work for the Standard Model of particle physics was laid.

Still today, 50 years after the formulation of the quark model, research in the realm of quarks is as exciting as ever, with discoveries which continue to surprise. Large-scale experiments are built to study the effects of the Standard Model with never-before seen energies down to the scales of individual quarks and the search for physics beyond the current Standard Model. Dedicated facilities investigate the larger objects of mesons and baryons, bound states of quarks which form at lower energies. Independent of the methods employed, the thousands of researchers in particle physics all over the world create more and more sophisticated machines to more finely resolve the processes underlying all nature.

The $\overline{\text{P}}$ANDA experiment, which is being built as part of the FAIR project, will perform research in the field of mesons and baryons near the charm threshold; at center-of-mass energies of up to about $\sqrt{s} = 5.5$ GeV. The hadron physics experiment is built with fixed-target kinematics to study antiproton-proton events with unprecedented detail and explore the region of physics which is described by the theory of Quantum Chromo Dynamics.

The setup of $\overline{\text{P}}$ANDA is unique among physics experiments. In the energy region of interest, signal events and background processes are very similar. To discriminate the large hadronic background from the interesting physics events, a novel method of data acquisition is chosen. $\overline{\text{P}}$ANDA will utilize a realtime reconstruction of the event topology to initiate data taking. This *online* reconstruction continuously evaluates information from $\overline{\text{P}}$ANDA's sub-detectors, compares it with a list of interesting physics channel, and eventually triggers the recording. In $\overline{\text{P}}$ANDA's high-luminosity mode, the reconstruction has, on average, 50 ns to decide to keep an event or to disregard it. No existing dedicated hardware equipment is suitable to match

$\overline{\text{P}}$ANDA's online reconstruction demands, making it a pioneering experiment to employ an online software trigger.

A computationally intensive task of online event reconstruction is online track reconstruction. The tracks, which connect the hit points from particles throughout sub-detectors, are computed to recover the parameters of the original particle. Tracks are the basis of all subsequent triggering decisions. Per second, on average about 1.6×10^9 hits are recorded in $\overline{\text{P}}$ANDA's STT sub-detector. The full amount of hits needs to be considered to reconstruct tracks, in the shortest timespan possible. This computing-intensive task requires new tools in the physics toolkit. Both the employed hardware needs to cope with the high-performance requirements and also the track reconstructing software needs to operate close to realtime.

This thesis presents first work towards online track reconstruction on Graphics Processing Units for $\overline{\text{P}}$ANDA. These high-performance devices, which originally were used to accelerate graphics of computer games, offer a platform of massively parallel computing. Exploiting the parallelism by suiting algorithms, a system of high-performance and low-budget can be set up to match the challenges of modern physics experiments.

The thesis is structured as follows.

In Chapter 2, the Standard Model is outlined and the most important theories are highlighted.

Chapter 3 begins by presenting the FAIR facility, where $\overline{\text{P}}$ANDA will be located. After that, the physics program of $\overline{\text{P}}$ANDA is outlined. Following, the different sub-detectors of the experiment are discussed. Special focus is laid on $\overline{\text{P}}$ANDA's novel data acquisition mechanism. Finally, the software tools employed during this thesis are presented.

In Chapter 4, the GPU platform is explained. The challenges of online track reconstruction are outlined and important terminology specified. After this, three algorithms are presented, capable of reconstructing tracks on the GPU: The Line Hough Transform, the Circle Hough Transform, and the Triplet Finder.

Chapter 5 analyzes the physics process of $\overline{\text{p}}\text{p} \rightarrow \text{D}^+\text{D}^- \rightarrow \text{K}^-\pi^+\pi^+\text{K}^+\pi^-\pi^-$ in $\overline{\text{P}}$ANDA. First, the simulation for the channel is analyzed. Subsequently, the detector response is studied. In the next parts, whole simulated events are reconstructed, first inclusively, then exclusively. Finally, background-like events are studied and different track reconstructions benchmarked.

Chapter 6 concludes this thesis and summarizes the results.

Chapter 2

Physics Introduction

This chapter gives a concise introduction into the basic principles of the physical effects of importance for $\overline{P}$ANDA.

2.1 The Strong Force in the Standard Model of Particle Physics

The mathematical description of all processes in particle physics is called the *Standard Model of particle physics*, or short the Standard Model (SM). Combining the theories of electromagnetic, weak, and strong interactions, it bundles the knowledge of all known forces and particles comprising matter[1]. It is being developed and refined since the 1960s and proven with predicted measurements during the years since. A recent highlight, the discovery of a Higgs-like boson by the two LHC experiments CMS and ATLAS, was awarded the Nobel Prize in Physics in 2013.

2.1.1 The Standard Model

The particles described by the SM can be categorized, among others, by their spin. Half-integer spin particles are called **fermions**, integer spin particles **bosons**.

2.1.1.1 Fermions

Elementary fermions with spin $s = {}^1\!/_2$ are the 24 particles making up all matter, disregarding dynamic effects. Fermions are categorized into leptons and quarks. Leptons have integer electrical charge and are not effected by the strong interaction; quarks, on the other side, interact strongly and having multiples of $q = {}^1\!/_3$ electrical charge. To each of the 6 leptons and 6 quarks, there is an antilepton and antiquark, respectively. Fermions are further structured into three generations of increasing mass. All elementary fermions are summarized in Table 2.1.

Per definition, fermions can also denote composite particles – non-elementary particles composed of an odd number of spin $^1\!/_2$ particles. Baryons, particles made-up of three quarks, belong to them. The proton p and the antiproton $\overline{p}$ are among the most prominent baryons.

---

[1]Gravity is notably absent from the SM. A Theory of Everything, combining all known forces, is still hypothetical.

Table 2.1: Overview of the fermions of the Standard Model, all with spin $s = 1/2$, sorted into leptons and quarks. The first column describes the elementary charge, q, of the particles of this row. Following, the three generations of leptons and quarks, ordered by their masses, m. To each fermion an antiparticle exists, summing up to 24 particles in total. Antiparticles have opposite charge-like quantum numbers. The color charge of the quarks is not shown. Masses of neutrinos are current upper limits. Data from [1].

| | | | Generation | | | | |
| | | **1** | | **2** | | **3** | |
Type	q / e	Name	m / $\mathbf{keV}/c^2$	Name	m / $\mathbf{MeV}/c^2$	Name	m / $\mathbf{GeV}/c^2$
Leptons	-1	e	511	μ	105.7	τ	1.7
	0	ν_e	<0.002	ν_μ	<0.19	ν_τ	<0.018
Quarks	$2/3$	u	2300	c	1275	t	173.21
	$-1/3$	d	4800	s	95	b	4.18

Table 2.2: Overview of the bosons of the Standard Model, all with integer spin s, sorted by their associated field. Given is also the electrical charge, q, and the mass m. Mass measurements from [1].

Associated Field	Boson	s / $\hbar$	q / e	m / GeV/c^2
Electromagnetic	γ	1	0	0
Weak	$W^\pm$	1	± 1	80.4
	Z	1	0	91.2
Strong	g (8)	1	0	0
Higgs	H	0	0	125.7

2.1.1.2 Bosons

Bosons denote particles with integer spin (e.g. $s = 0, 1, \ldots$).

Gauge bosons[2] are the mediators of forces in the SM. Five gauge bosons are known, each associated to an interaction: γ, W^+, W^-, Z, g. The photon, γ, is the mediator of the electromagnetic interaction. It has no electrical charge and is massless, the latter being responsible for the infinite range of the electromagnetic force. W^+, W^-, Z are the three bosons of the weak interaction, the W bosons being electrically charged and both of mass $80.4\,\mathrm{GeV}/c^2$, the Z having no charge and a mass of $91.2\,\mathrm{GeV}/c^2$. Due to their comparably high masses, the weak interaction is only relevant for short-range interactions. Compared to electromagnetic and strong forces, it appears *weak*, giving the interaction its name. Gluons, g, are the mediators of the strong interaction. There are eight types of massless gluons, distinguished by their Quantum Chromo Dynamics (QCD) color charge content. Since they carry this charge, they can themselves participate in the strong interaction and self-interact. Self interaction leads to the phenomenon of confinement: Quarks and gluons can not be measured as isolated particles,

---

[2]With spin equal to exactly 1.

they always form composite structures when observed from a distance. The strong force is explained in more detail in Section 2.1.2.

Another boson is the Higgs boson H, a scalar, spin-less boson with mass of approximately $125\,\text{GeV}/c^2$. It is the quantum responsible for mediation of the Higgs field and has a special role in the SM: The interaction with the field creates the masses of fermions and gauge bosons in the electroweak regime of the SM. Proposed a long time ago[3], the boson has recently been discovered at the Large Hadron Collider (LHC) [5, 6].[4] The *elementary* bosons are summarized in Table 2.2.

Similar to fermions, Bosons can denote composite particles of integer spin. Mesons, comprising a quark and an antiquark, are one kind of them.

2.1.1.3 Mathematical Formulation

On a mathematical level, the SM is described as a Quantum Field Theory (QFT), with the fundamental quantum fields of the different SM interactions. It is a gauge theory with free parameters that need to be fixed by experimental measurements. It combines knowledge from the electroweak and strong interaction, includes the basic fermion fields, and accounts for masses by means of the Higgs mechanism.

The dynamics of the SM can be summarized in the Lagrangian $\mathcal{L}$, related to the energy configuration of a system[5]. The Lagrangian of the SM is invariant under local transformations of the product group $\text{SU}(3) \times \text{SU}(2) \times \text{U}(1)$[6], where the latter two components act on the Higgs field, the fermion field, and the fields of the electroweak bosons, and the former, $\text{SU}(3)$ part acts on the QCD gluon and the fermion field.

The Lagrangian can be split up into parts:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Fermion}}^{\text{Kinetic}} + \mathcal{L}_{\text{Fermion}}^{\text{Masses}} + \mathcal{L}_{\text{Higgs}}.$$

The first part, $\mathcal{L}_{\text{Gauge}}$, refers to the different gauge bosons of the SM. It includes the dynamical properties (kinematics) as well as self interactions of the bosons. The second term, $\mathcal{L}_{\text{Fermion}}^{\text{Kinetic}}$, includes the dynamics of fermions coupling with gauge fields. $\mathcal{L}_{\text{Higgs}}$ holds information for the Higgs field. The masses of the fermions are created in the Yukawa interactions between

---

[3]The *Higgs* mechanism is more correctly called the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism and emerged over year-long research by the involved physicists, partly independently from each other. To specify a single paper is difficult in this context. The first paper of HIGGS was published in 1964 [2], succeeding papers shortly after. A good overview about the mechanism is given in [3] with a more history-focused version available at [4].

[4]It is yet to be found out if the Higgs-like boson discovered at LHC is one of many Higgs bosons or the only one.

[5]The Lagrangian $\mathcal{L}$ is, more precisely, the **Lagrangian density**. It can be transformed to the equation usually called the Lagrangian, L, by integrating over all space-like dimensions, leaving only the time-like. Integrating furthermore the time-like dimension gives the action S. From S the equations of motion can be derived. In this thesis, the Lagrangian $\mathcal{L}$ specifies the Lagrangian density, as it is usually the case in particle physics.

[6]$\text{U}(n)$, the unitary group, and $\text{SU}(n)$, the special unitary group, denote groups of $N \times N$ unitary matrices with defined mathematical structure. The *group operation* is matrix multiplication. For SU, a subgroup of U, the matrices have determinant 1 as an additional requirement. $\text{U}(1)$, the unitary group of dimension 1, describes rotations along the unit circle in the complex plane.

the Higgs field and the fermions through spontaneous symmetry breaking – $\mathcal{L}_{\text{Fermion}}^{\text{Masses}}$ holds this information.

Another way of splitting up the Lagrangian is by summarizing the individual parts by the interaction they are associated to:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}'_{\text{Higgs}}. \tag{2.1}$$

$\mathcal{L}_{\text{EW}}$ holds all electroweak-relevant information associated to its $\text{SU}(2)_\text{L} \times \text{U}(1)_\text{Y}$ symmetry group. This includes, among others, the electric charge[7], the weak isospin charge, the electroweak bosons, and the coupling constants. $\mathcal{L}_{\text{QCD}}$ holds the information about gluon and quark interactions, as defined through $\text{SU}(3)_\text{C}$. $\mathcal{L}'_{\text{Higgs}}$, as before, holds the Higgs field information needed for mass generation. $\mathcal{L}_{\text{QCD}}$ is explained in more detail in Section 2.1.2.4.

2.1.2 The Strong Force

Since the research of the P̄ANDA experiment (Antiproton Annihilation at Darmstadt) is mainly focused on aspects of the strong interaction, this part is highlighted in more detail below.

The strong interaction creates the strong force, the force operating effectively at sub-fm scales and responsible for binding the quarks of a hadron.[8] It is called *strong* as it exceeds the strength of all other fundamental forces at the scale of hadrons.

The strong force is described mathematically by the quantum field theory of Quantum Chromo Dynamics (QCD), a gauge theory similar to Quantum Electrodynamics (QED), the theory of the electromagnetic force. Starting in the 1960s, QCD emerged as the theory for the description of the strong interaction. In total, two Nobel Prizes have been awarded in direct relation to QCD theory: 1969 to Murray Gell-Mann, one of the *fathers* of QCD [7]; 2004 to David J. Gross, H. David Politzer, and Frank Wilczek [8]. Many other Nobel-Prize-awarded research touches topics of QCD as well.

2.1.2.1 Color Charge

In the realm of QCD, quarks carry color charge in addition to their electromagnetic charge. Although quarks carry the charge, they do not exhibit it measurably at a long distance – quarks always form bound states with other quarks to create colorless *hadrons*. No states with non-zero color charge were measured up to now. The phenomenon of the inability to isolate individual quarks is called *confinement*. When applying energy and trying to separate two quarks, the flux tube of force-carrying gluons between the quarks stretches until it finally breaks apart, creating a quark-antiquark pair from the vacuum. The quark and antiquark directly reattach to the original quarks, creating two hadrons from the original one. The reason for the behavior of the gluons is the strength of the force: After reaching past a certain distance (approximately the diameter of a hadron), the strength of the strong interaction is constant over the distance, approximately 1 GeV/fm. Unlike e.g. the electromagnetic force, it does not decrease with increasing separation distance. The strength needed to separate quarks then is sufficient to create new pairs of quarks from the vacuum.

---

[7]More precise: The weak hypercharge Y.

[8]The strong force is also responsible for holding nuclei of atoms together on a fm scale. In this case, this residual strong force is usually referred to as the nuclear force.
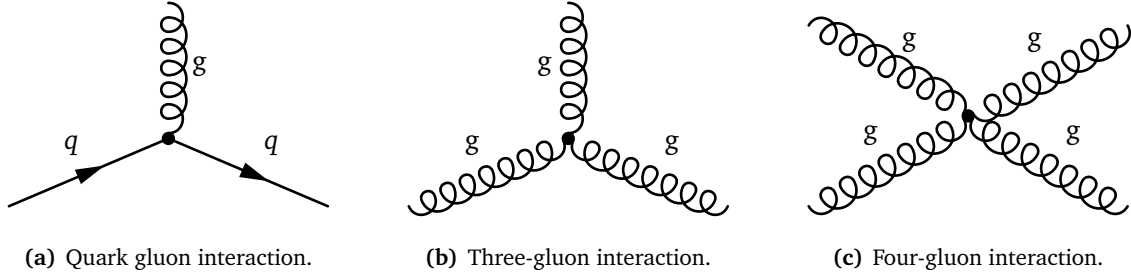
(a) Quark gluon interaction.   **(b)** Three-gluon interaction.   **(c)** Four-gluon interaction.

Figure 2.1: Feynman graphs of the possible quark (q) gluon (g) interactions of QCD. From left to right: **(a)** – a quark radiates a gluon; **(b)** – a gluon radiates another gluon / three gluons interact; **(c)** – four gluons interact. The vertices of the left two interactions are proportional to the strong coupling g_s, the right vertex is proportional to g_s^2.
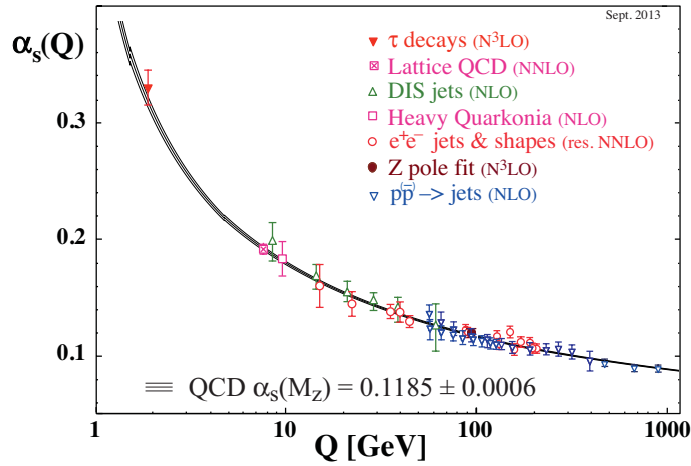

Figure 2.2: The running coupling constant of QCD, α_s, as a function of the energy scale Q. The different data points employ perturbative theories up to different levels of perturbation in their extraction of the value: *NLO* involves *next-to-leading order* corrections, *NNLO next-to-next-to-leading order* corrections, and so on [1].

2.1.2.2   Gluons

Another difference in comparison to QED is that the mediator particle of QCD, the gluon, itself carries the charge of the interaction, in contrast to the uncharged photon. A gluon is internally color-charged with a color and an anticolor. The charge enables gluons to *self-interact*. In total, eight kinds of gluons exist – mathematically this is included as QCD being a SU(3) group theory. Self-interaction introduces three-gluon and four-gluon vertices into the theory. A gluon can radiate another gluon of different color charge (three-gluon vertex) or interact directly with another gluon (four-gluon vertex). Gluon self-interaction creates the flux tube between quarks and is hence the reason behind the quark confinement at long distances. The Feynman diagrams of possible vertices of QCD are presented in Figure 2.1.

2.1.2.3   Strong Coupling Constant

The coupling constant of QCD, α_s (see Equation 2.4), varies depending on the energy region it is evaluated at. α_s gets smaller when going from $\mathcal{O}(\text{fm})$ to smaller distances between interacting quarks, e.g. by using a probing particle with high momentum transfer.

With high momentum transfer q^2, quarks can be considered not as bound states of mesons and baryons, but as quasi-free particles. The quarks experience *asymptotic freedom*. Thus, the coupling constant is not really a constant but depends on the momentum transfer q^2. A definition for α_s is[9]

$$\alpha_s\left(\mu^2\right) = \frac{1}{\beta_0 \ln(\mu^2/\Lambda^2)}, \tag{2.2}$$

with μ being the mass-dependent scale, α_s is evaluated at; β_0 a perturbatively calculated factor, depending on the number of quark flavors and colors[10]; $\Lambda = \Lambda_{\text{QCD}}$ the scale of QCD, the starting point of energy at which QCD can be considered as a perturbative theory[11] [9, 10]. α_s is often referred to as the *running coupling constant* of QCD. QED's coupling α_e is also running and depends on the energy scale, but α_e experiences the opposite effect compared to α_s: It becomes stronger with increasing energy scale.

α_s is usually evaluated at the scale of the mass of the Z boson, $91.2\,\text{GeV}/c^2$: $\alpha_s(M_Z^2) = 0.1185 \pm 0.0006$ [1]. Figure 2.2 displays the dependence of α_s on the scale it is evaluated at, with data points from different measurements and theoretical calculations.

The variation of α_s enables the possibility for particle physics experiments with extremely high energies, e.g. the LHC experiments, to perform QCD probes described by perturbative methods. Experiments with beam momenta of $\mathcal{O}\left(\text{GeV}/c\right)$ or lower need to employ other, less explicit theories to model the physical processes. The iterative approach of Lattice QCD is a prominent one. Prominent examples are Lattice QCD, an iterative approach modeling quarks and gluons on three-dimensional lattices, and chiral perturbation theory, which uses hadrons instead of gluons and quarks as degrees of freedom.

2.1.2.4   QCD Lagrangian Density

As mentioned in Section 2.1.1.3, the mathematical formulation of the SM is usually done with the Lagrangian density. It contains the static and dynamic information as quantum fields.

The QCD Lagrangian, as sketched in Equation 2.1, can be written as

---

[9]To derive Equation 2.2, α_s is considered in the realm of a perturbative approach: $\mu^2\,\mathrm{d}\alpha_s/\mathrm{d}\mu^2 = -\alpha_s^2(\beta_0 + \beta_1\alpha_s + \beta_2\alpha_s^2 + \ldots)$. β_0 corresponds to the highest contribution from a direct QCD vertex. Because QCD allows for self-interaction, another gluon vertex can contribute to the same process – a loop. Perturbatively, more loops can be added to expand and modify the *effective* coupling constant. Hence, it is important, at which scale, μ, α_s is evaluated. This process is called *renormalization*. Equation 2.2 is the leading order part of the infinite series that is the running a_s.

[10]β_0 is a normalizing factor of the first element of the running definition of α_s: $\beta_0 = (11N_c - 2N_f)/12\pi$, with N_c the number of colors, and N_f, the number of quark flavors.

[11]Λ_{QCD} as well depends on the scale it is evaluated at: $\Lambda_{\text{QCD}}(5\,\text{GeV}) \approx 217\,\text{MeV}$.

$$\mathcal{L}_{\text{QCD}} = \sum_q \bar{\psi}_{q,a} \left(i\gamma^\mu \partial_\mu \delta_{ab} - m_q \delta_{ab} - g_s \gamma_\mu t^C_{ab} \mathcal{A}^C_\mu \right) \psi_{q,b} - \frac{1}{4} \mathcal{F}^A_{\mu\nu} \mathcal{F}^{A\,\mu\nu}, \tag{2.3}$$

using Einstein notation summing over repeated indices. [1, 11]

Going through the parts of the Lagrangian from left to right, the individual parts are explained in the following.

$\sum_q$ The explicit sum runs over all quark flavors q, $q = 1 \ldots 6$, i.e. u, d, c, s, t, and b quarks – see Table 2.1.

$\psi_{q,a}$ The quark field, representing the quarks. q, again, denotes a quark flavor, a represents the strong color charge and runs from $a = 1$ to $a = 3$, the three possible colors of a quark.[12] b is another color charge.

$i\gamma^\mu \partial_\mu \delta_{ab} - m_q \delta_{ab}$ The propagator for quarks of mass m_q. The γ^μ are the Dirac matrices. This part of Equation 2.3 corresponds to the Dirac equation of a free field.

$g_s \gamma_\mu t^C_{ab} \mathcal{A}^C_\mu$ The interaction between gluons and quarks. $\mathcal{A}^C_\mu$ represents the gluon fields, with C between 1 and 8 referring to the number of gluons. There are eight kind of gluons, $N_C^2 - 1 = 8$,[13] generated by the eight Gell-Mann matrices in t^C_{ab} – a possible set of matrices as generators of the $\text{SU}(3)_C$ group. g_s is the coupling, connected to the QCD coupling constant α_s by

$$\alpha_s = \frac{g_s^2}{4\pi}. \tag{2.4}$$

$\frac{1}{4} \mathcal{F}^A_{\mu\nu} \mathcal{F}^{A\,\mu\nu}$ The dynamics of the gluon field, describing field propagation and gluon-gluon interactions. The gluon-gluon self-interaction distinguishes QCD from QED strongly, here the force carrier exerts the force it transmits. The field tensor $\mathcal{F}^A_{\mu\nu}$ is given by $\mathcal{F}^A_{\mu\nu} = \partial_\mu \mathcal{A}^A_\nu - \partial_\nu \mathcal{A}^A_\mu - g_s f_{ABC} \mathcal{A}^B_\mu \mathcal{A}^C_\nu$, with the structure constants f_{ABC} ($A, B, C = 1 \ldots 8$), connected to the Gell-Mann matrices[14]. The $\mathcal{A}^{A,B,C}_{\mu,\nu}$ are the gluonic fields as before.

In $\mathcal{L}_{\text{QCD}}$, the individual properties of the strong interaction can be seen – initially free quarks as QCD's massive particles, quark-gluon interaction and gluon-gluon self-interaction with different colors.

---

[12]A separation of the quark fields in up-type and down-type quarks, as well as left-handedness, is omitted here for brevity.

[13]Naively, one would expect nine gluons to exist. But one of the possible color states, the colorless singlet $(r\bar{r} + b\bar{b} + g\bar{g})/\sqrt{3}$, would be subject for unconfined measurement. In nature, it is not seen.

[14]The structure constants are proportional factors in the commutator relations of the generators T_α of $\text{SU}(3)_C$: $[T_a, T_b] = if^{abc}T_c$. They are either 0, 1, $\pm 1/2$, or $\sqrt{3}/2$. The eight Gell-Mann matrices λ^α are the generators, with $T^\alpha = \lambda^\alpha/2$.

(a) Meson $J^P = 0^-$ nonet.

(b) Meson $J^P = 1^-$ nonet.

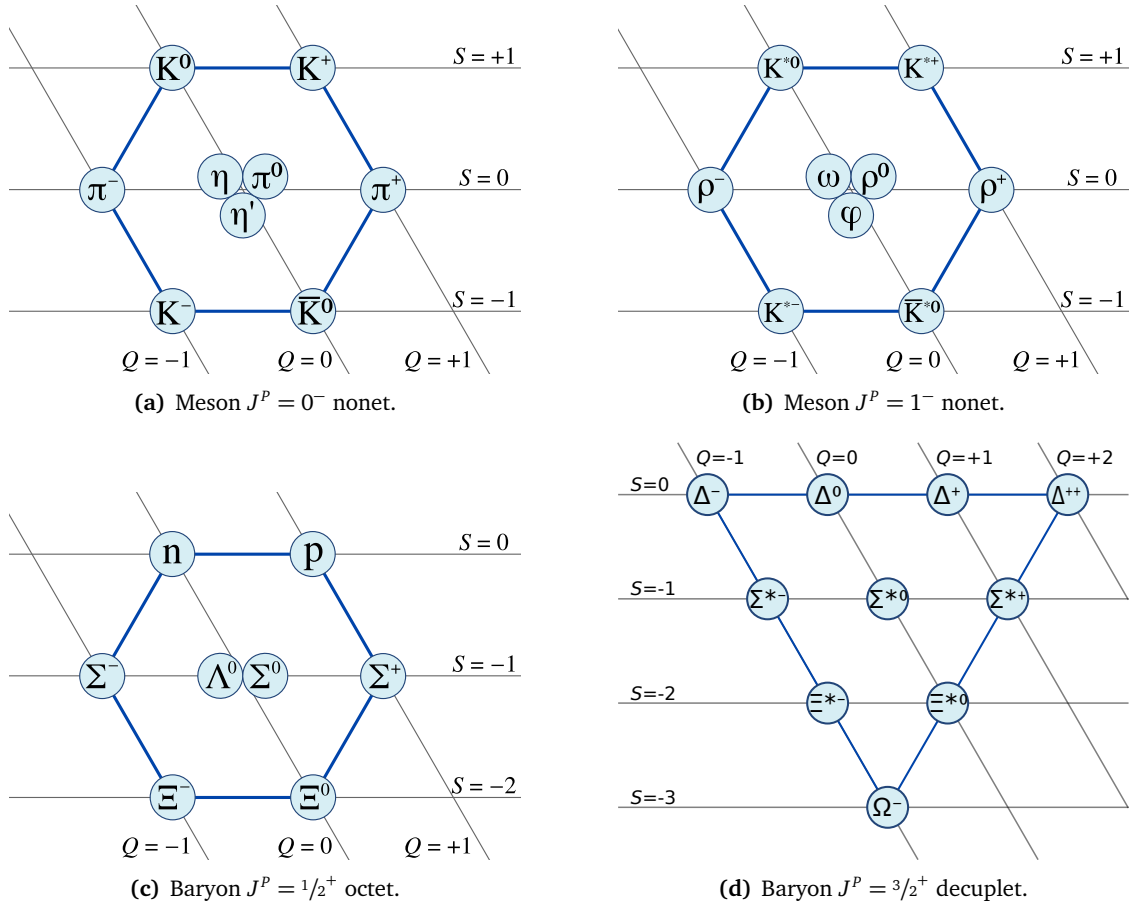(c) Baryon $J^P = 1/2^+$ octet.

(d) Baryon $J^P = 3/2^+$ decuplet.

Figure 2.3: The multiplets of ground state mesons (top) and baryons (bottom) with u, d, s quark content and no orbital angular momentum. The elementary charge Q is shown horizontally, the strangeness S vertically. Adapted from [12–15].

2.2   Hadronic States

Hadrons, particles consisting of quarks, are in the first level categorized by their number of constituting quarks: Mesons are comprised of two quarks (a quark and an antiquark) and baryons of three quarks.

Further classification can be achieved when sorting the hadrons by their quark flavor content, electromagnetic charge q, and total angular momentum, J – the combination of spin S and orbital angular momentum L, $J \leq |L \pm S|$. For mesons, J of the ground states, the states with zero orbital momentum ($L = 0$), can either be $J = 1$ or $J = 0$. Considering only the lightest mesons with u, d, and s quark content, this creates two groups: The nonet of $J^P = 1^-$ and the nonet of $J^P = 0^-$. The 1^- particles are called vector mesons, the 0^- pseudoscalar mesons, the P in the exponent of the notation refers to the behavior of a particle under parity transformation. Sorting each nonet by elementary charge and strangeness[15] provides the pictures in Figure

---

[15]Strangeness is the quantum number denoting the strange quark content of a hadron. One strange quark leads to strangeness $S = -1$, one strange antiquark to $S = 1$.

2.3(a) and Figure 2.3(b).

The same principle can be employed to sort baryons. The baryons with no orbital angular momentum have even parity ($P = +$), the possible total angular momentum values are $J = 1/2$ and $J = 3/2$. The eight and ten particles of the different angular momenta are shown in Figure 2.3(d) and Figure 2.3(c), respectively.

2.3 Heavy-Light Charmed Mesons

Mesons with a charm or a bottom quark (or antiquark) are labeled by the term *heavy-light mesons*. Compared to u, d, and s quarks, c and b quarks are at least $13 \times$ more massive. The top quark does not form bound states of mesons or baryons, as it decays before engaging in any kind of binding due to its high mass.

In the nomenclature of particle physics, mesons with one charm (or anticharm) quark are called D mesons. If the second quark is a strange quark, an indicating index is added: D_s. The same principle follows for the even heavier mesons comprising bottom quarks: B denotes a meson with one b quark (or antiquark) and either a u or d. B_s refers to the second quark being with strangeness, B_c labels additional *charm*.

D and B are also called mesons with **open charm** and *open bottom*, as they have a non-zero charm or bottomness, respectively[16]. Mesons with **hidden charm** or bottom quantum number are called *quarkonia*: Charmonium for $c\bar{c}$ states, bottomonium for $b\bar{b}$ states. Prominent examples of quarkonium mesons of the two quark flavors are the J/ψ, the first excited state of a $c\bar{c}$ meson ($J^P = 1^-$), and the Υ, the first excited $b\bar{b}$ state (also $J^P = 1^-$). Both particles are of importance since through their detection the c and b quarks were first seen, respectively: The J/ψ was first detected independently by Richter [16] and Ting [17] in 1974, the Υ by Lederman three years later [18].

2.3.1 Open Charm: D Mesons

Two years after the discovery of the J/ψ, mesons containing one heavy charm quark were found at the Stanford Linear Accelerator Center (SLAC): The neutral D^0 was found by GOLDHABER, PIERRE, ABRAMS, *et al.* [19], the charged $D^{\pm}$ by PERUZZI, PICCOLO, FELDMAN, *et al.* [20]. The papers were published a month apart – in June and July 1976, respectively. With masses of $(1864.84 \pm 0.07)\,\text{MeV}/c^2$ (D^0) and $(1869.61 \pm 0.10)\,\text{MeV}/c^2$ ($D^{\pm}$), the two particles are the lightest D mesons. Since their discovery, many new D meson states have been found. The Particle Data Group (PDG) summary currently lists 15 states with masses ranging up to $2750\,\text{GeV}/c^2$ (the D(2750) was found in 2010 [21]).

The spectrum for strange D mesons, D_s, continues to be measured as well, as more and more states are discovered beyond the lightest $D_s^{\pm}$ ($m = (1968.30 \pm 0.11)\,\text{MeV}/c^2$) – see [25, Table 1] for an concise overview over the current experimental status. The spectrum is based on calculations in the relativistic quark model, done by GODFREY and ISGUR in 1985 [22] and

---

[16]The charm quantum number (*charm* or *charmness*) is defined analogously but opposite to the number accompanying strange quarks (*strangeness*): +C for every included c quark, -C for every included $\bar{c}$ antiquark. Bottomness follows strangness with -B for every b quark and B for every $\bar{b}$.
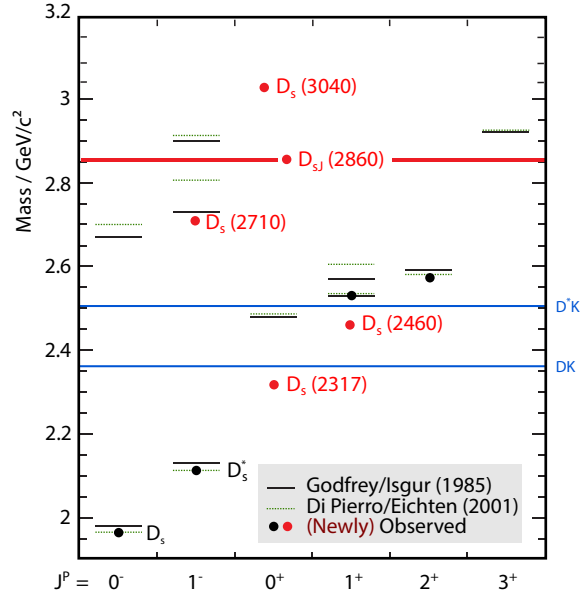
Figure 2.4: D_s spectrum predicted by GODFREY and ISGUR (solid lines) and DI PIERRO and EICHTEN (green dashed lines). Included are the DK and D*D thresholds (blue). The quantum numbers of D_s(2860) are yet to be determined. Adapted from [24].

DI PIERRO and EICHTEN in 2001 [23]. Many of the newly found D_s mesons do not fit the original model predictions, as can be seen in Figure 2.4. Above as well as below the DK mass threshold, states were found, for which no clear theoretical prediction exist. Hence, the open charm sector is still subject to intense research both experimentally as well as theoretically – even 40 years after the discovery of the first charmed meson. The work by SONG, CHEN, LIU, and MATSUKI gives an extension to the model by GODFREY and ISGUR to incorporate the newly found states by means of *screening*, see [25].

2.3.2 Hidden Charm: Charmonium

Since the detection of the first charmonium state in 1974, also the charmonium spectrum has grown largely. States of different angular momentum configurations and masses have been found. Figure 2.5 gives a graphical overview.

Starting ten years ago, a number of new states have been found in the charmonium spectrum, dubbed the XYZ states. Especially the two e^-e^+ experiments Belle and BaBar contributed with new measurements. The states, located over the DD threshold, usually lack description by a theoretical model. There are hints that these states could incorporate signs of new physics. Some theoretical suggestions for explanations of the new charmonium states are listed in the following, according to [26].

Molecular States A charmed meson loosely binds to another charmed meson, e.g. D^0 binding with $\overline{D}^0$ or $\overline{D}^{0*}$. The binding happens by means of interactions with quark/color exchange (short distances) or π (long distances, possibly dominating). The state is sometimes referred to as a *deuson*. Figure 2.6(a).
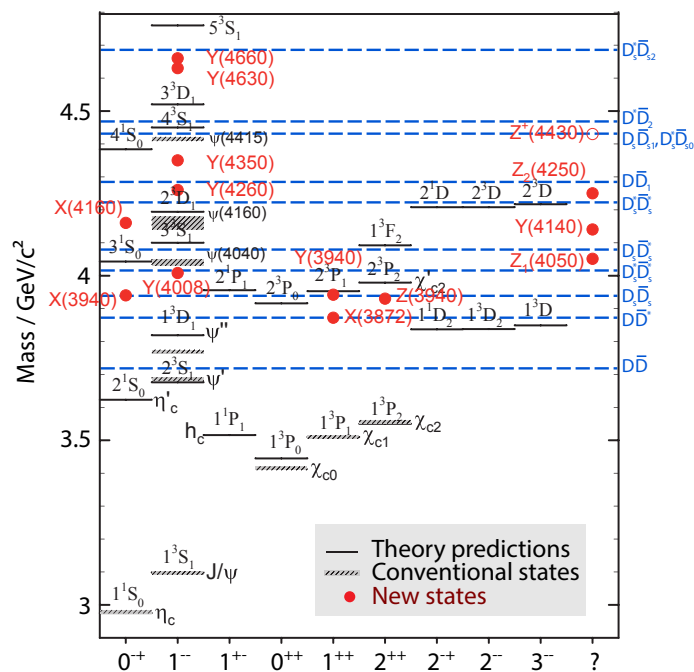
Figure 2.5: Charmonium spectrum. Solid, horizontal lines represent theoretical predictions, shaded lines conventional charmonium states. Red circles denote recently found XYZ states. D thresholds are shown in dashed blue lines. Adapted from [24].
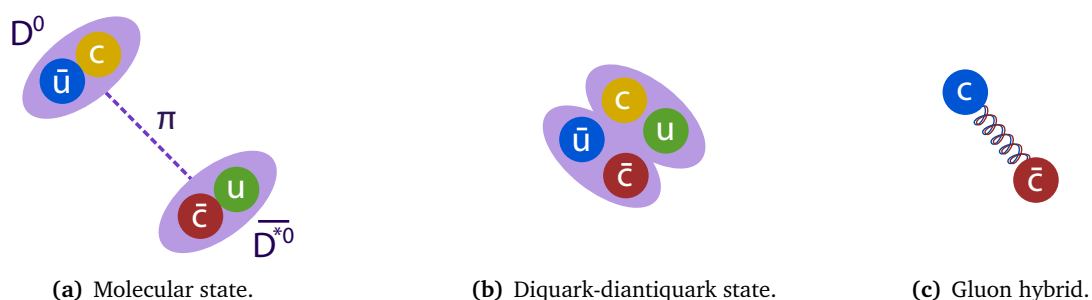

(a) Molecular state.

(b) Diquark-diantiquark state.

(c) Gluon hybrid.

Figure 2.6: Possible explanations of the internal configuration of the newly discovered XYZ charmonium states.

Tetraquarks Four quarks group together but are more strongly bound compared to a molecular
state. A quark pair binds with its anti-quark pair sister – a diquark-diantiquark system.
Binding might happen due to spin-spin interactions. Figure 2.6(b).

Hybrids A charmonium state has an additional possible gluonic degree of freedom, of which
the conventional states are the lowest excitations. Figure 2.6(c). See also Section 3.2.1.4.

Most of the new XYZ charmonium-like states decay either in J/ψ- or D$\overline{\text{D}}$-involving processes
and have narrow decay width of $\Gamma \approx 50\,\text{MeV}$ to $\Gamma \approx 100\,\text{MeV}$ [27]. For one prominent state,
the X(3872), there only exists an upper limit of $\Gamma < 1.2\,\text{MeV}$ for the width [1].

Chapter 3

The P̄ANDA Experiment

The P̄ANDA experiment is presented in this chapter. First, the accelerator facility supplying infrastructure to P̄ANDA and other experiments, FAIR, is introduced in Section 3.1. The targeted physics program of P̄ANDA is highlighted following in Section 3.2. In Section 3.3, the individual parts of the detector are shown. The novel data acquisition scheme of P̄ANDA is the topic of Section 3.4. Finally, Section 3.5 presents the software tools used and concludes this chapter.

3.1 Facility for Antiproton and Ion Research

The Facility for Antiproton and Ion Research (FAIR), the laboratory where P̄ANDA is located at, will establish a research facility for the investigation of *cosmic matter* in the laboratory. Offering a sophisticated accelerator complex with diverse beams, it will create an infrastructure for experiments with antiprotons, stable ions, and exotic nuclei. Both the experiments and the accelerators located at FAIR are explained in the following, an overview is given in Table 3.1. FAIR is being built in modules (Section 3.1.4).

3.1.1 FAIR Experiments

On the premises of FAIR, a number of diverse physics experiments will be situated. The experiments are structured into four groups, covering different scientific areas – FAIR's *four pillars*: At APPA, atomic, plasma, and applied physics is researched; NUSTAR produces and investigates radioactive nuclei; CBM measures dense baryonic matter; P̄ANDA studies hadrons produced in antiproton-proton reactions.

3.1.1.1 APPA

Atomic, Plasma Physics and Applications (APPA) is an umbrella collaboration of FAIR experiments dealing with atomic, bio, and plasma physics as well as material science.

BIOMAT   Sharing an experimental hall with SPARC, BIOMAT (Biology and Material Science) will research in matter irradiated with heavy ions. Both biological matter as well as solids are targeted [29].

15

Table 3.1: An overview of the different FAIR facilities. The facilities marked with »*« are not part of FAIR's initial start-up version (see Section 3.1.4).

| Device	Key Aspects	See Sec.
Experiments		
APPA	Physics with matter and plasmas	3.1.1.1
BIOMAT	Irradiation of matter (biological, solids) with heavy ions	3.1.1.1
FLAIR*	Slow antiprotons and heavy ions; trapping	3.1.1.1
HEDgeHOB	High-energy-density matter	3.1.1.1
SPARC	Heavy ion analysis	3.1.1.1
WDM	Laser-ion interactions	3.1.1.1
NUSTAR	Experiments with rare ions	3.1.1.2
Super-FRS	Production and separation of exotic nuclei	3.1.1.2
DESPEC/HISPEC	Spectrometry of exotic nuclei	3.1.1.2
ELISe	Collision of electrons with exotic nuclei	3.1.1.2
EXL*	Study of light ions with gas-jet target	3.1.1.2
ILIMA	In-beam analysis of exotic nuclei	3.1.1.2
LaSpec	Laser spectroscopy of exotic nuclei	3.1.1.2
MATS	Analysis of trapped exotic nuclei	3.1.1.2
R3B	Reactive properties of high energy exotic nuclei	3.1.1.2
CBM	External experiment for QCD probes with heavy ions	3.1.1.3
P̄ANDA	Internal experiment with antiproton collisions	3.1.1.4
Accelerators & Infrastructure		
UNILAC	Ion acceleration first stage	3.1.2.1
p-LINAC	Proton acceleration first stage	3.1.2.1
SIS18	Synchrotron for ion and proton acceleration	3.1.2.2
SIS100	Main FAIR accelerator for ions and protons	3.1.2.2
SIS300*	Further acceleration for ions	3.1.2.2
CR	Accumulation and cooling of antiprotons and ions	3.1.2.3
RESR*	Further accumulation of antiprotons and ions	3.1.2.3
NESR*	Experimental ring for ions and mid-energy antiprotons	3.1.2.3
eA Collider*	Electron ring associated to NESR	3.1.2.3
Antiproton Target	Metal-based target, length: $\mathcal{O}(10\,\text{cm})$	3.1.3.1
HESR	Acceleration and cooling of high-energy antirprotons	3.1.3.2
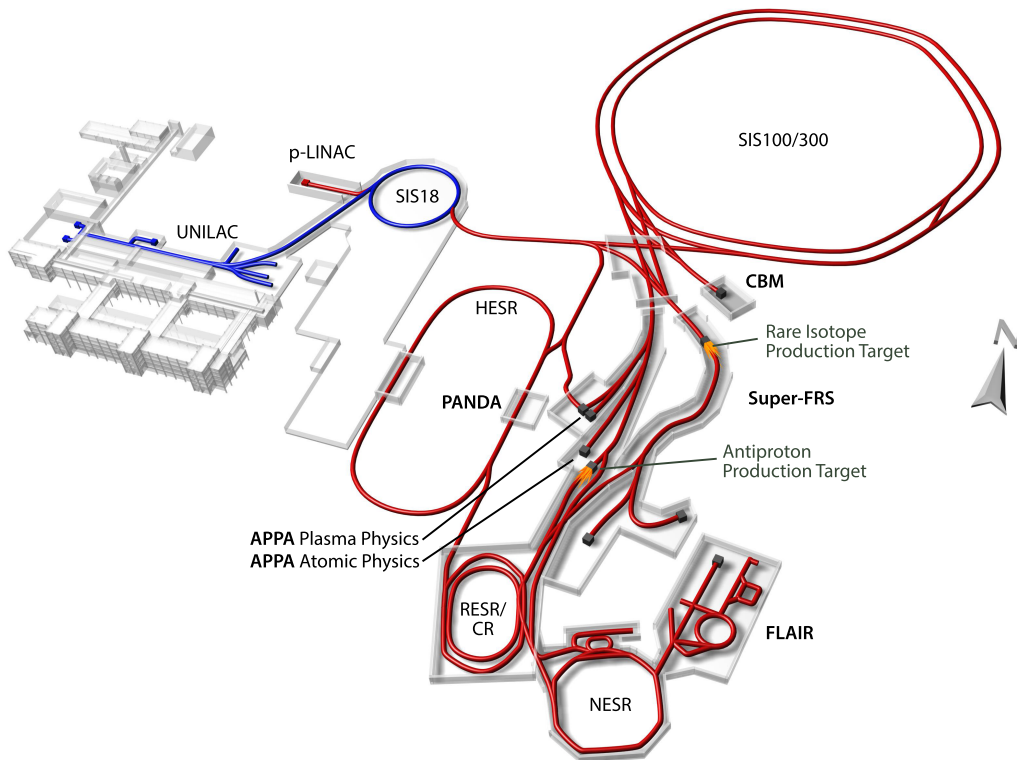
Figure 3.1: The future FAIR facility with different accelerators and experiments annotated. On the left side, the existing buildings of GSI can be seen. The accelerator chain outlined in blue is already existing, the red chain is planned for FAIR. Adapted from [28].

Effects of continuous radiological exposure are studied, e.g. during travels to Mars. Both for the involved materials and persons. The approach of using an heavy ion beam with a wide range of beam energies opens up the possibility of inexpensive, reproducible, in-detail research in comparison to the usual flight tests with airplanes. Additionally, particle therapy for cancer and non-cancer diseases is studied. BIOMAT will be the first radiobiology laboratory of its size in Europe [30].

FLAIR   The Facility for Low-Energy Antiproton and Heavy Ion Research (FLAIR) focuses on physics with antiprotons of lower energy, as a next generation experiment of the Antiproton Decelerator (AD), located at CERN (Conseil Européenne pour la Recherche Nucléaire). Investigated, among others, is: Symmetry breaking processes of antiparticles; the inner structure of hadrons as probed by antiprotons; gravitation of trapped antihydrogen [32].

FLAIR comprises a magnetic and an electrostatic storage ring , as well as a facility for trapping particles, HITRAP (Figure 3.2. The magnetic ring is called LSR (Low Energy Storage Ring), the electrostatic ring USR (Ultra Low Energy Storage Ring). From FAIR's accelerator chain extracted antiprotons can be cooled to kinetic energies of 300 keV (LSR), 20 keV (USR), and $\mathcal{O}$ (eV) (HITRAP). LSR is a modification of the Swedish CRYRING storage ring used at the
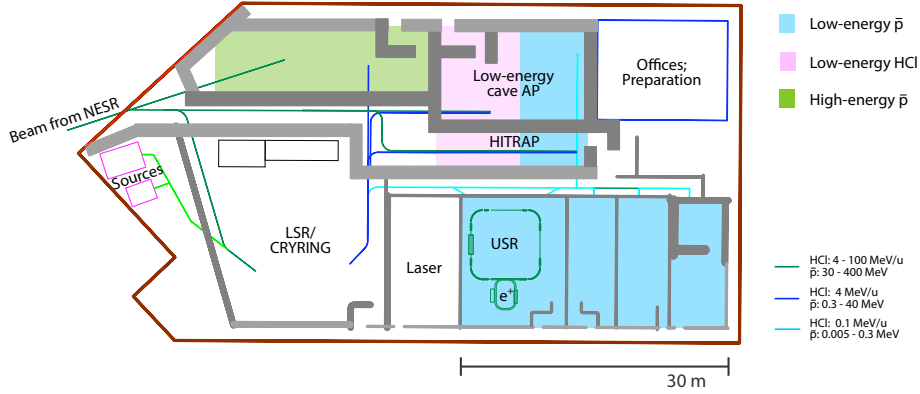
Figure 3.2: FLAIR's experimental hall. Adapted from [31].

Manne Siegbahn Laboratory in Stockholm. While LSR decelerates both antiprotons and ions, USR only decelerates antiprotons [33].
FLAIR is not part of FAIR's startup version but will be built later on as part of module 4 of the Modularized Start Version (MSV).

HEDgeHOB   High Energy Density Matter generated by Heavy Ion Beams (HEDgeHOB) will make use of FAIR's intense heavy ion beam to produce high-energy-density matter. The basic properties of this matter is analyzed, including both stationary as well as dynamical properties [34].

   In the past, high power lasers, chemical, and even nuclear explosions were used to generate this warm dense matter. The approach of a heavy ion beam allows for a very fast and homogeneously heating of target volumes, resulting in low gradients inside the material [35].

SPARC   The research program of SPARC (Stored Particles Atomic Research Collaboration) is two-fold, centering around highly-charged heavy ions analyzed at the border between nuclear and atomic physics [36].

   In its first use-case it will use relativistic heavy ions to study collision dynamics with strong, ultra-short electromagnetic fields. Different beam energies will be used to investigate a broad spectrum of collision processes. Secondly, SPARC will research in highly charged ions, both stable and unstable. The ion trap HITRAP is used to perform, among others, high sensitivity tests to the SM.

   Recent proposals foresee the inclusion of SPARC into the High-Energy Storage Ring (HESR) ring to enable a very early realization of the experiment [37, 38].

WDM   The Warm Dense Matter (WDM) project will build a facility to combine FAIR's heavy ion beams with a kilojoule laser, PHELIX (Petawatt High-Energy Laser for Heavy Ion EXperiments), to study warm dense matter and dense strongly-coupled plasmas. WDM will produce macroscopic samples of warm dense matter to be analyzed by means of X-ray scattering [39].
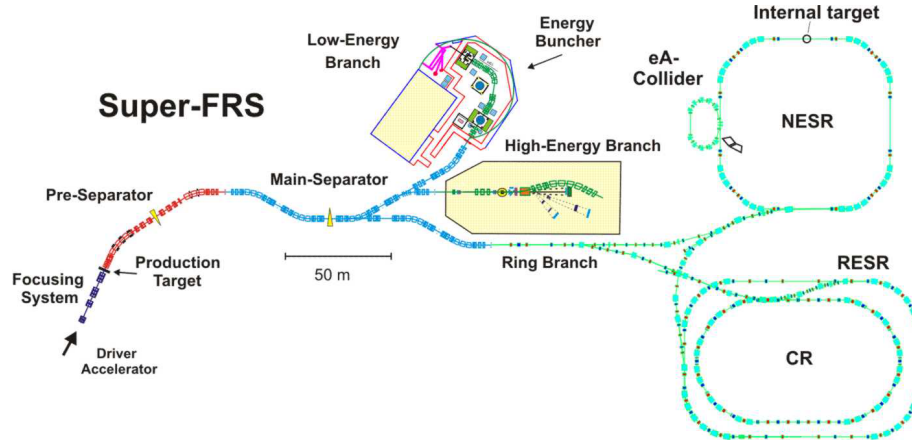
Figure 3.3: Structure of the NUSTAR experiments [39].

3.1.1.2 NUSTAR

NUSTAR, short for Nuclear Structure, Astrophysics and Reactions, is a collection of experiments investigating the properties of heavy elements (heavier than iron) with intensive beams made of these rare isotopes. See Figure 3.3 for an overview.

Central to NUSTAR is the Super-FRS, providing beams of radioactive ions to the experiments. Targeted scientific research is, among others, the understanding of the inner structure of nuclei, as well as connected astrophysical topics.

Super-FRS   The Super Fragment Separator (Super-FRS) is the successor of the Fragment Separator (FRS) at GSI, providing the input particles to all following NUSTAR experiments. A set of superconducting magnets enables the in-flight separation of exotic ions produced in a target. Compared to FRS, Super-FRS provides a larger accepted phase space of transmitted ions, important for transmitting radionuclei produced through projectile fission in the target [40]. The experimental setup allows for choosing a thick production target, offering a high luminosity of particles. Special magnets with ceramic insulation need to be used due to the high radiation surrounding the target. Isotopes from all elements, up to uranium, can be produced and separated. This is done in $\mathcal{O}(100\,\text{ns})$ – also short-lived nuclei can be supplied.

Super-FRS provides the beam of radionuclei to three different experimental areas via the low energy branch (incorporating the experiments DESPEC/HISPEC, MATS, and LaSpec), the high energy branch (with R3B), or the ring branch (with the rings CR, RESR, NESR, and eA Collider, and the experiments EXL and ILIMA) [41].

DESPEC/HISPEC   In DESPEC/HISPEC (Decay Spectroscopy/High-Resolution Spectroscopy), low energy beams ($100\,A\,\text{MeV}$ to $300\,A\,\text{MeV}$) of radioactive ions from Super-FRS are studied by high precision gamma ray spectroscopy.

In the HISPEC part, gamma rays emitted by excited nuclear states are measured in-flight using a germanium detector and a calorimeter. DESPEC slows down these radioactive ions in a

silicon detector, surrounded by additional instrumentation specialized for measuring gamma rays and neutrons [42].

ELISe   The Electron-Ion Scattering experiment (ELISe) probes for nuclear properties (charge, transition charge, matrix elements, and more) of radioactive ions by means of electron scattering. As a first of its kind experiment, ELISe will collide electrons from the Electron Ion Collider (eA Collider) with ions from the New Experimental Storage Ring (NESR) in a collider setup. Instrumentation at the interaction point records the scattered electrons and other collision remnants [43].

EXL   As EXL's unabbreviated name, Exotic nuclei studied in light-ion induced reactions at the NESR storage ring, suggests, the experiment will be an internal experiment in the NESR storage ring. It will be a fixed-target experiment with a gas-jet target to study nuclear structures with reactions induced by light ions. Capitalizing on the storage ring beam structure, EXL is able to measure its unstable nuclei at low momentum transfer with high resolution. Measurements are performed in the target region by a silicon recoil, a gamma ray, and a slow neutrons detector. A forward detector is also foreseen (for neutral and charged particles), a heavy ion spectrometer will be installed into the beam [44, 45].

ILIMA   Isometric Beams, Lifetimes and Masses (ILIMA) will be an experiment, researching the different unstable isotopes produced in Super-FRS and stored in the Collector Ring (CR), the Recuperated Experimental Storage Ring (RESR), and NESR. The masses of the stored exotic nuclei are measured by means of two methods: The *Schottky Mass Spectroscopy* targets long-living isotopes involving different in-beam pickup signals, the *Isochronous Mass Spectroscopy* focuses on short-lived nuclei by incorporating time of flight detectors within the beam of the CR [46].

LaSpec   LaSpec (Laser Spectroscopy) will determine nuclear parameters of radioactive isotopes and isomers produced in Super-FRS using laser spectroscopy. Spectroscopy with lasers allows for high precision measurement of properties like hyperfine structures and electronic transitions, giving access to nuclear spins, as well as dipole and quadrupole moments. Exotic nuclei with high neutron numbers are targeted by this experiment [47].

MATS   MATS (Precision Measurements of very short-lived nuclei with Advanced Trapping System) is the partner experiment of LaSpec, located in the same experimental hall as LaSpec of the low energy branch of Super-FRS. For mass measurements and different kinds of spectroscopy of short living radionuclides, MATS consists of three different ion traps: an electron beam ion trap for charge creation, a pre-trap for ion preparation, and an eventual Penning trap for the actual measurements [47].

R3B   Reactions with Relativistic Radioactive Beams (R3B) uses a similar system as GSI's current LAND experiment to study instable nuclei created in a fixed target and boosted into the forward direction. It will be located at the high energy branch of Super-FRS and consist of different sub-detector system. Physics objectives include research into properties of heavy ions, scattering with light ions, and breakup reactions [48].
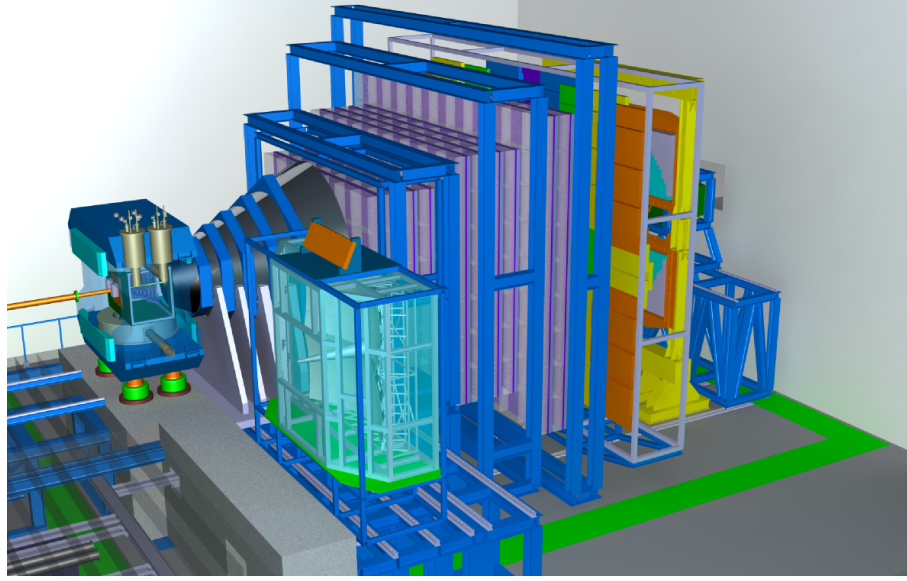
Figure 3.4: Rendering of the CBM experiment in its experimental hall, with the beam entering from the left. In the front, the gas RICH module is moved out of the experiment and instead the muon detection system is included. The experiment is approximately 10 m long. Taken from [49].

3.1.1.3 CBM

The CBM experiment (Compressed Baryonic Matter) is an external fixed-target experiment researching the QCD phase diagram at high baryon densities and moderate temperatures. QCD-related questions, like confinement, chiral symmetry breaking and restoration, and baryonic matter in neutron star-like properties are part of the experimental program.

Fixed-target reactions of heavy ions with beam energies of $10\,A$ GeV to $40\,A$ GeV from the SIS100/SIS300 accelerators will be investigated by CBM. A detector system with fine granularity is needed in order to resolve the high multiplicity events of the experiment. To detect the various particles of interest (e.g. ρ, Λ, Ξ, Ω, D, J/ψ) and discover rare signals in the event topology, CBM will gather a large amount of data. The experiment will run with a reaction rate of 10 MHz and involves several high speed detectors and read outs. Techniques for real-time reconstruction of events are planned [49].

As SIS300 will not be part of the initial start version of FAIR (see Section 3.1.4), a reduced version of CBM will run with lower beam energies of $2\,A$ GeV to $11\,A$ GeV. The High-Acceptance Dielectron Spectrometer (HADES) detector will also be installed in front of CBM to operate at these lower beam energies [50].

3.1.1.4 $\overline{\text{P}}$ANDA

Antiproton Annihilation at Darmstadt ($\overline{\text{P}}$ANDA) is an internal experiment at HESR. Antiprotons of 1.5 GeV/c to 15 GeV/c beam momentum collide with a fixed proton target. The interaction of antiprotons with protons enables research for a wide variety of physics as particles of all quantum numbers can be produced. Particular research targets are meson spectroscopy (notably
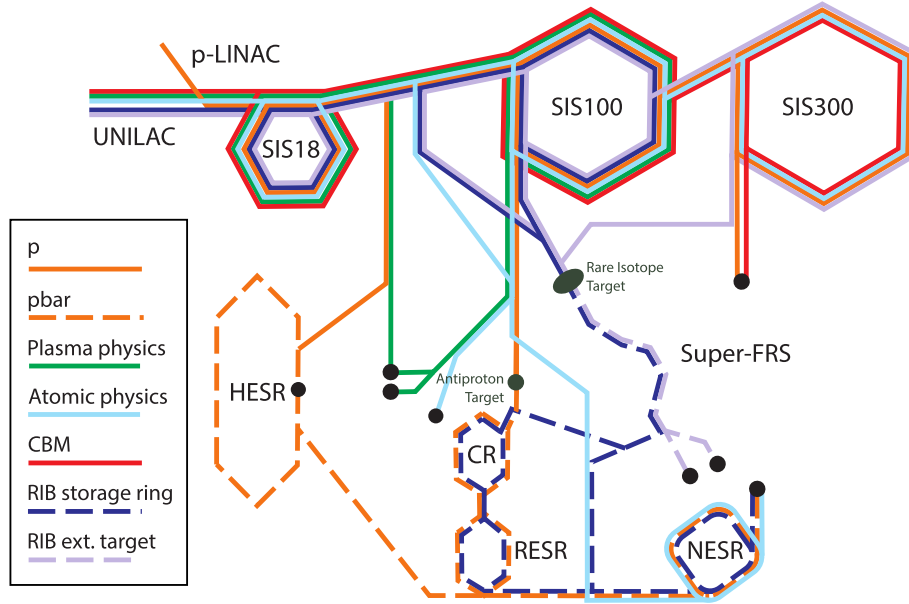
Figure 3.5: Schematic view of the different accelerator lanes of FAIR. The machines can be operated in parallel, serving up to four scientific programs at once with different beam properties. Experiments are outlined as black spots. Adaption from [39].

charmonium and open charm), baryon production, nucleon structure and hypernuclei, and exotic states (e.g. glueballs or quark molecules).

This wide range of physics goals imposes requirements to the detector design and physics runs. The detector is build up of a number of sub-detectors organized into the Target Spectrometer, surrounding the interaction region in a nearly 4π fashion, and into the Forward Spectrometer, detecting forward-boosted particles at shallow angles. P̄ANDA will run at an average event rate of 20 MHz and incorporates a real-time event reconstruction mechanism to distinguish background events from signal events.

The experiment is described in detail in Section 3.3.

3.1.2   Facility

The FAIR facility provides infrastructure to a number of different experiments (see Section 3.1.1). Different accelerators create the combination of particles and momenta needed for the diverse experimental objectives. The accelerator chain builds up on existing GSI facilities and uses the proven machines for injection devices. The descriptions in this section are taken from [39, 51].

Figure 3.5 outlines the different types of beams intended for different experiments and their routes through the accelerator chain. The setup of the chain allows for serving up to four different beams in parallel for different scientific programs. Not all of the displayed rings are part of FAIR's start version.
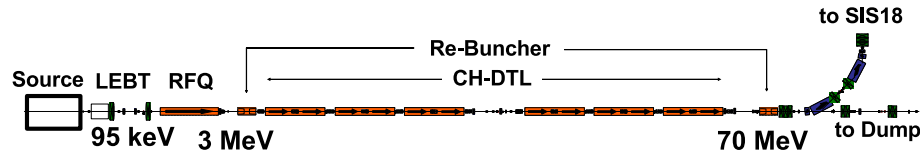
The different accelerators are outlined in the following.

Figure 3.6: The proton linear accelerator p-LINAC. A low energy beam transport system (LEBT) transports protons from a duoplasmatron source to a radio-frequency quadrupole (RFQ). From there, a system of drift tubes (DTL) and cavities (CH for *Crossedbar H-Cavities*) accelerates protons for injection into SIS18 [39].

3.1.2.1 Linear Accelerators

For ions and protons, FAIR comprises two distinct linear accelerators, starting off the acceleration to the different experimental facilities. Both UNILAC and p-LINAC inject their accelerated ions to SIS18.

UNILAC UNILAC is the first stage of GSI's chain for ion acceleration. In its basic version it is already in operation since 40 years. UNILAC is able to deliver beams with energies of $3\,A\,\text{MeV}$ to $13\,A\,\text{MeV}$ for heavy ions and up to $15\,A\,\text{MeV}$ for light ions. The accelerator uses different ion sources (MEVVA, ECR, PiG). For FAIR, UNILAC will be upgraded in many parts to match the required beam properties (e.g. beam intensities, power, and emittances).

p-LINAC In addition to UNILAC, FAIR features a new injection stage for protons – the p-LINAC. Protons, ionized in a duoplasmatron are accelerated to $70\,\text{MeV}$ before being transferred to SIS18, see Figure 3.6. 7×10^{12} protons per cycle are injected into SIS18.

3.1.2.2 Synchrotrons

Three synchrotrons are responsible for increasing the energy of protons and ions produced in the linear accelerators.

SIS18 Part of the legacy GSI accelerator chain and subject to technical upgrades for FAIR, SIS18 is the first ring accelerator in the chain. Light ions are accelerated to $2\,A\,\text{GeV}$, heavy ions to $1\,A\,\text{GeV}$, and protons to $4.7\,\text{GeV}$. Acceleration is done in two possible modes: With a fast ramp rate of $10\,\text{T/s}$ up to $12\,\text{T\,m}$ maximum magnetic bending power (magnetic rigidity), or with $4\,\text{T/s}$ to a higher maximum bending power of $18\,\text{T\,m}$. The latter mode injects ions directly to Super-FRS, the former to SIS100 for additional acceleration.

SIS100 The main accelerator for FAIR is SIS100. Ions and protons, pre-accelerated by SIS18, are brought to beam energies of $2.7\,A\,\text{GeV}$ (U^{28+}) and $29\,\text{GeV}$ (protons). 4×10^{11} ions per pulse or 2×10^{13} protons per cycle can be accelerated and compressed to bunch lengths of $60\,\text{ns}$ and $25\,\text{ns}$. SIS100 features a circumference of $1084\,\text{m}$ and gets its name from its magnetic rigidity of $100\,\text{T\,m}$. Superconducting magnets ($2\,\text{T}$) ensure fast ramping at a $4\,\text{T/s}$ rate. 14 accelerator cavities are installed [52].

SIS300   Located next to SIS100, SIS300 will increase the magnetic rigidity by a factor of three compared to SIS100. It is used for additional acceleration of ion beams – for U^{92+}, an energy of $34\,A$ GeV is planned. The accelerator will be equipped with $6\,$T magnets, ramping at $1\,$T/s. Ions are delivered to CBM.

SIS300 is not part of the planned initial start version of FAIR, see Section 3.1.4.

3.1.2.3   Storage Rings

Storing, accumulation, cooling, and further preparation of the different beams is performed by a number of rings.

CR   The Collector Ring (CR) collects antiprotons and ions from SIS100 and Super-FRS, respectively, and pre-cools the beams stochastically. The targeted beam energies are $0.74\,A$ GeV for U^{92+} and $3\,$GeV for antiprotons. Normal conducting dipoles are used for this task in the 211 m long ring. The emittance of the beam of antiprotons after production in the target and insertion into CR is reduced from 240 mm mrad to 5 mm mrad within 10 s, for rare isotopes from 200 mm mrad to 0.5 mm mrad within 1.5 s. In addition to its use as a cooling facility, CR will be used as a mass spectrometer for short-lived isotopes in a *isochronous mode*, employing an internal Time Of Flight (TOF) detector [53].

Antiprotons are extracted and accumulated in RESR, ions are transfered to NESR.

RESR   The Recuperated Experimental Storage Ring (RESR) is dedicated to accumulation and deceleration of antiprotons and short-lived ions. The ring's location surrounds the CR, it has a circumference of 245 m, and re-uses technical equipment[1] from GSI's ESR.

Up to 1×10^{11} antiprotons of fixed energies of 3 GeV can be accumulated and stochastically cooled in the RESR. The particles are delivered in sets of 2×10^8, coming pre-cooled from CR every 10 s. The antiprotons are either ejected to NESR (for further transport to FLAIR) or into HESR.
Rare isotopes are decelerated from $740\,A$ MeV to ranges from $100\,A$ MeV to $500\,A$ MeV within 1 s with fast ramping dipoles of 1 T/s ramp rates. Ions are exclusively delivered to NESR.

RESR is not part of the initial start version of FAIR, it will be built as module 5 of the MSV (see Section 3.1.4).

NESR   The New Experimental Storage Ring (NESR) is a storage ring for stable and instable ions, as well as antiprotons. It is the successor of GSI's ESR. Different experimental facilities are connected to NESR (FLAIR, EXL, ELISe, ILIMA) and make use of its high intensity beams. Instable ions can either be used with their injection energy or decelerated to lower energies. Antiprotons are always decelerated to momenta below the initial $3\,$GeV/c. Deceleration is performed with a fast ramp rate of 1 T/s. NESR incorporates both stochastic and electron cooling to optimize beam properties, especially during deceleration. The ring's circumference

---

[1]Quadrupoles and beam diagnostics are re-used from the Experimental Storage Ring (ESR), the dipole design is used from the NESR.

is 222 m, including four straight sections of 18 m each. In one straight section, NESR joins with an electron ring (eA Collider) to offer electron ion beam interaction for ELISe.

NESR gets beam from different sources. Stable ions can be delivered by SIS100/SIS300, instable ions by RESR, CR, or directly by Super-FRS, antiprotons by RESR/CR (see Section 3.1.4). NESR is not part of FAIR's initial start version.

eA Collider The eA Collider, part of the ELISe collaboration, stores electrons, initially accelerated in an electron linear collider. In one of NESR's straight sections, the eA Collider overlaps with the ion storage ring and allows for elastic, inelastic and quasielastic scattering of electrons off the instable ions. The electron beam energy is 125 MeV to 500 MeV, the circumference of the ring is about 50 m [43, 54].

HESR The High-Energy Storage Ring (HESR) is dedicated to storage, acceleration, and deceleration of antiprotons. P̄ANDA is located in one of HESR's straight sections and the ring is explained in more detail in the following section.

3.1.3 Antiprotons in HESR

The main ring for storing and accelerating antiprotons in FAIR is HESR. A proton beam is produced in the p-LINAC and further accelerated with SIS18 and SIS100. At the antiproton production target, antiprotons are created. They are collected and cooled in CR and RESR and then transferred to HESR. In the FAIR start version (see Section 3.1.4), RESR is absent and a combination of CR and HESR is used to prepare the beam.

3.1.3.1 Antiproton Production Target

Antiprotons are produced from protons accelerated by SIS100, which inelastically collide with a metal target. Behind the target, the antiprotons are collected in a magnetic horn and separated from collision residue particles in a subsequent 58 m long beam line. Thirteen quadrupoles and two sextupoles modify the beam to match the properties needed for injection into CR – for both magnet types the CR magnet designs are used.

The basic design of the antiproton production target is based on the design of the Antiproton Accumulation Complex (AAC) at CERN. It is modified to match the requirements of FAIR's accelerator chain: Instead of 25 GeV source protons (AAC), 29 GeV protons are delivered by SIS100; roughly 2×10^{13} protons are delivered at both AAC and FAIR, but a proton bunch at FAIR is much shorter (25 ns) than at AAC (4 bunches within 400 ns).

Antiprotons at 3 GeV with $\Delta p/p = \pm 3\%$ momentum spread are collected and transferred to CR. Approximately 5×10^{-6} antiprotons are produced per proton. The target itself is of roughly 10 cm length and made of either copper, nickel, or iridium. The material is chosen to be not too light, in order to produce a sufficient number of antiprotons in a short distance, but also to be not so dense that it melts under the deposited proton energy [55].
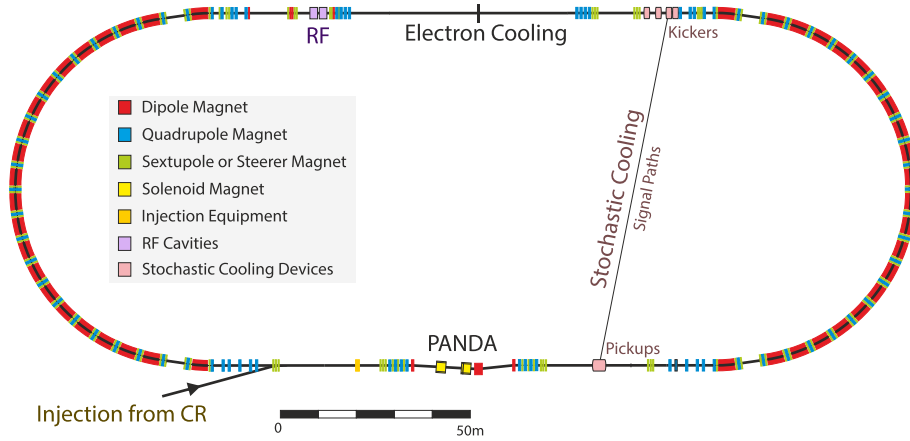
Figure 3.7: Schematic view of the High-Energy Storage Ring. Adapted from [56].

3.1.3.2 HESR

HESR (Figure 3.7) is designed for experiments with antiprotons of high energies[2] – beam momenta from $1.5\,\text{GeV}/c$ to $15\,\text{GeV}/c$ will be reached with a phase-space cooled beam [57, 58].

The circumference of the ring is $574\,\text{m}$, with two straight sections of $132\,\text{m}$ each. The beam is cooled by both electron and stochastic cooling. The ring will work in two modes: Either in high luminosity mode, $\mathcal{L} = 2 \times 10^{32}/(\text{cm}^2\,\text{s})$, with a comparably large beam momentum spread, or in the high precision mode, $\Delta p/p = 4 \times 10^{-5}$, with lower luminosity. 44 dipole magnets with a total bending power of $50\,\text{T\,m}$ are used.

P̄ANDA is housed in one of the straight sections. For detection of particles ejected from its internal target, P̄ANDA houses both a solenoid and a large dipole magnet. Especially the last device imposes challenges onto the accelerator design, as the magnet has to be rampable by the accelerator control system, but each different field strengths results in different particle detection properties inside of the experiment. Also, the magnet has to have a large aperture and its bending of the beam needs to be corrected by additional magnets. The bending of the beam can be seen schematically in the sketch of Figure 3.7.

Two RF cavities for beam acceleration/deceleration are installed in the straight sections of HESR. Furthermore, kickers for injection from RESR/CR are installed. Antiprotons are always injected at $3\,\text{GeV}$ and then brought to the desired experimental energy.

The equipments and instruments used in HESR are currently being tested and evaluated at the Cooler Synchrotron (COSY) facility in Forschungszentrum Jülich (FZJ).

Electron Cooling HESR will be equipped with an electron cooler with high energies of $4.5\,\text{MeV}$ located in the straight section opposite of P̄ANDA. Upstream and downstream from the electron cooler, solenoid magnets for beam correction are installed. In the start version of

---

[2]FLAIR, in contrast, investigates low energy antiprotons, see Section 3.1.1.1.

HESR, the 4.5 MeV electron cooler is preceded by a version with lower energy of 2 MeV. For cooling, the device uses a longitudinal magnetic field to guide the electrons. In addition to antiprotons, heavy ions can also be cooled (see Section 3.1.3.2) [59].

Stochastic Cooling Systems for stochastic cooling are installed in both straight sections. The signal pickups are located downstream of $\overline{\text{P}}$ANDA, the beam kickers at the beginning of the next straight section. The high-sensitivity devices operate with a bandwidth of 2 GHz to 3 GHz, with extensions to 4 GHz to 6 GHz possible. The longitudinal and transversal spreads of the beam are cooled [60].

Clearing Electrodes To neutralize charged ions from interaction of the beam with surrounding residual gas, clearing electrodes are installed. Especially downstream of $\overline{\text{P}}$ANDA, continuously working electrodes are needed to clear away ions from the experiment's internal target [58].

Impact of Modularized Start Version on HESR Within the Modularized Start Version (MSV) (see Section 3.1.4), RESR will not be built initially and will not be available for antiproton accumulation at the startup of FAIR. Only CR is available for pre-cooling and collection of antiprotons.

HESR has methods in place for antiproton accumulation (stochastic cooling, cavities), but the particle numbers needed for the high luminosity mode cannot be reached without additional accumulation in RESR: Instead of 10^{11}, only 10^{10} antiprotons will circulate in the ring. Consequently, HESR will only run in high resolution mode in the MSV. Also, the duty cycle of $\overline{\text{P}}$ANDA is lowered, as HESR needs time for antiproton accumulation, in which no physics experiments can occur.

The scheme for antiprotons reaching $\overline{\text{P}}$ANDA in the MSV is as follows: Protons from SIS100 create antiprotons at the antiproton production target. 10^8 antiprotons are collected in CR. They are cooled for 10 s, then transferred to HESR. There, the antiprotons are cooled, while the next 10^8 antiprotons are collected in CR. After 10 s cooling, the CR antiprotons are, again, transferred to HESR. This procedure is repeated for 100 times until 10^{10} antiprotons are accumulated in HESR. Since the particles are cooled 100 times for 10 s, the accumulation procedure in HESR takes 1000 s. Afterwards, the beam needs to be cooled further and brought to the desired energy for the specific measurement [61].

Induced by the MSV, the HESR's performance has recently been studied with respect to heavy ions [62]. Both stable and instable ions could be injected through the accelerator chain, accelerated, stored, and studied with new internal targets and associated experimental devices. SPARC experiments could be installed [37, 63].
Ions, e.g. U^{92+}, would be injected through CR at 740 A MeV and accelerated to 2 A GeV, although higher energies are possible (up to 5 A GeV). These nuclei would also be subject to stochastic and electron cooling [64].
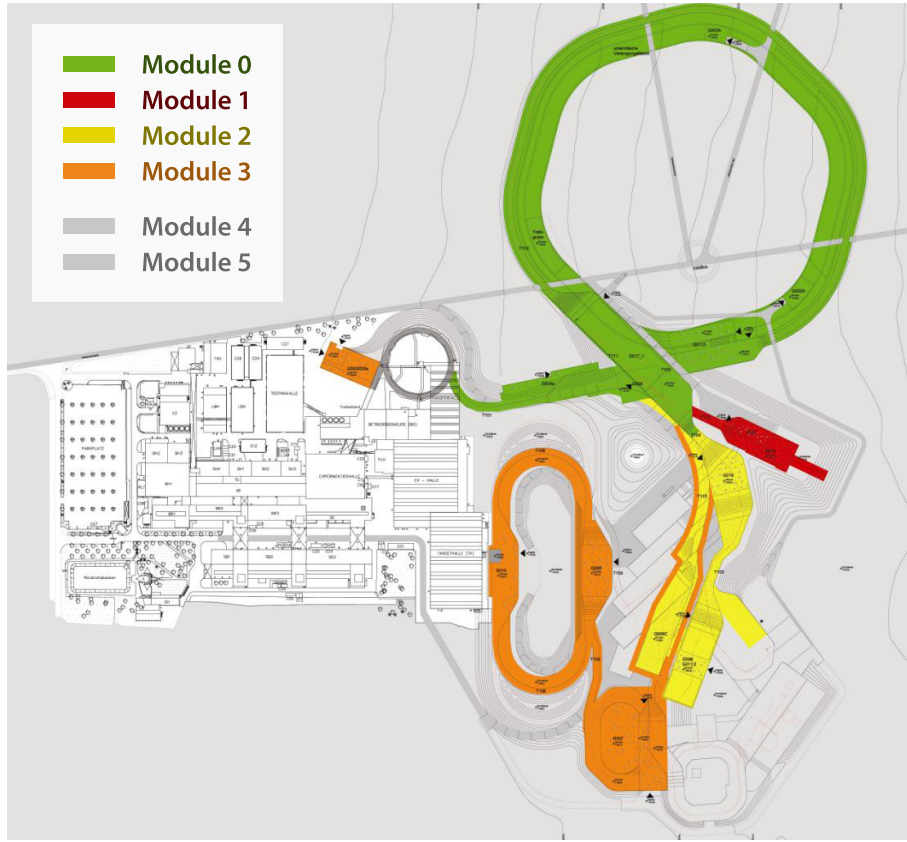
Figure 3.8: The different modules of FAIR's Modularized Start Version. Adapted from [65].

Table 3.2: The different modules of the Modularized Start Version of FAIR.

Module Number	Module Content
0	SIS100, connection to GSI accelerators (SIS18, UNILAC)
1	Halls for CBM, HADES, APPA
2	Super-FRS
3	Antiproton infrastructure: p-LINAC, antiproton target, CR, HESR, hall for P̄ANDA
4	Halls for NUSTAR, FLAIR; NESR
5	RESR
6	SIS300, eA Collider

3.1.4   Modularized Start Version

To realize the project as soon as possible, amid funding constraints from all involved countries, FAIR was divided into modules in 2009. The different modules are summarized in Table 3.2.

Modules 0 to 3 are part of FAIR's initial version, the so-called Modularized Start Version (MSV). Although important parts are missing initially, this structure enables most of the experiments
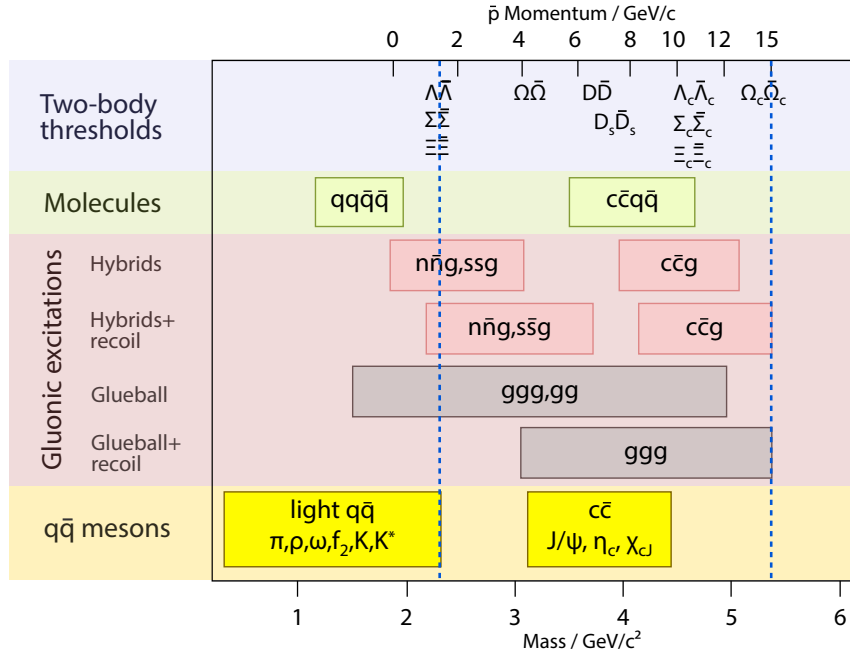
Figure 3.9: The physics spectrum accessible by HESR and, partly, by $\overline{\text{P}}$ANDA. Blue vertical lines denote the range of the beam momentum delivered to the experiment, ranging from $1.5\,\text{GeV}/c$ to $15\,\text{GeV}/c$ (upper x axis). The mass of the physics systems is shown on the lower x axis. Adapted from [67].

to start in a timely manner. Modules 4 to 6 are desired to be built soon, in order to intensify the experimental operations with updated machinery. They are planned, »as additional funds become available« [66].

Modules 0 and 1 are built first, as module 0 includes the connection of the FAIR facility to GSI and module 1 consists of comparably easy to construct and low-priced buildings. Modules 2 and 3 are built thereafter in parallel [65].

3.2 Physics Program

The physics program of $\overline{\text{P}}$ANDA centers around research of effects of the strong force (see Section 2.1.2), as depicted in Figure 3.9. The center of mass energy of $\sqrt{s} = 2.26\,\text{GeV}$ to $\sqrt{s} = 5.48\,\text{GeV}$ enables measurements in the charm mass region.

The phase-space-cooled high-precision antiproton beam from HESR enables a diverse physics program, covering effects of hadron physics up to the charm energy region. Antiproton-proton annihilation ($\overline{\text{p}}$p collisions) enables many special opportunities, as hadrons of all kind of quantum numbers can be produced directly. The limiting factors of e^-e^+ and pp colliders are avoided: e^-e^+ colliders can only produce vector mesons with $J^{PC} = 1^{--}$ (e.g. J/ψ) directly, as the quantum numbers are limited by the intermediate photon. Other hadrons can only be reached by other, more indirect processes (e.g. photon-photon fusion), lowering the measurement precision strongly. In pp colliders (LHC), collisions occur among the hadronic constituents of the protons, particles are mainly produced out of the quark sea with a large

momentum spread. In case of P̄ANDA, the created particles are directly formed from the energy of the antiparticle-particle (p̄p) interaction, in which every constituting antiquark of the antiproton has a charge-conjugated quark to annihilate with in the proton. A broad spectrum of accessible quantum states is opened up. The chosen setup of a circulating beam hitting a fixed target enables a precise study of decaying particles boosted into the forward direction.

The different physics disciplines researched by P̄ANDA are highlighted in the following. For a full, in-depth overview of P̄ANDA's planned physics program, see e.g. the »Physics Performance Report for PANDA: Strong Interaction Studies with Antiprotons« [68].

3.2.1   Hadron Spectroscopy

A large part of the physics investigations of P̄ANDA focuses on measurements and classification of hadrons. Of interest are mesons with open charm (D, D_s), charmonium states (e.g. J/ψ), heavy baryons (e.g. Λ^*, Ξ^*), and the search for signs of new physics.

3.2.1.1   Open Charm Spectroscopy

Mesons with non-vanishing charm number have already been introduced in Section 2.3.1. P̄ANDA will measure the well-known D and D_s mesons, but also refine the more recent discovered states by BaBar, Belle, and CLEO, which do not fit in any well-established quark model theories (Figure 2.4). The mesonic system of a heavy quark, c, and a light quark, u or d[3], offers the possibility to gain insight into the strong force, when the heavy quark is assumed stationary and the dynamic effects summarized into the light quark. The method was used in the hydrogen atom to better understand the electromagnetic force. Operating at the D production threshold, $6.4\,\text{GeV}/c$, a large quantity of D mesons can be produced with favorable background conditions, as the phase space for additional hadrons is small.

One particular interesting meson is the $D_s(2317)$, first seen by BaBar [69], decaying into $D_s^+\pi^0$. The decay width of the state is small, as the particle's mass is lower than the DK threshold and it hence can not decay by K emission. Currently, only an upper limit of $\Gamma < 3.8\,\text{MeV}/c^2$ for the width is known. Depending on the decay width, it is possible to further constrain theoretical models for the physics process creating the *new* mesons. Simulations show that P̄ANDA will be able to measure the width with resolutions of $\mathcal{O}\left(100\,\text{keV}/c^2\right)$ [70]. A reason for this high resolution is the narrow beam momentum spread ($\Delta p/p \simeq 4 \times 10^{-5}$) of HESR, allowing for precise threshold scans. Similar results are hoped to be obtained for the new, higher mass D_s mesons, e.g. $D_s(2460)$ and $D_s(2860)$.

3.2.1.2   Charmonium Spectroscopy

As a key physics goal, P̄ANDA will also explore the charmonium mass region, introduced in Figure 2.3.2 and depicted in Figure 2.5 and Figure 3.9. For both the well-established states as well as the newly-found states, the experiment will refine the current measurements and contribute with original ones. The singlet states, e.g. η_c, are measured by different experiments but have comparably large differences concerning the mass and width. Here, P̄ANDA can

---

[3]Possibly also s.

contribute with a high-precision resonance scan, measuring masses and widths with precision up to $\mathcal{O}\left(100\,\text{keV}/c^2\right)$. In addition, the newly-found XYZ states are expected to be measured and classified concerning their quantum number configuration and nature.

Compared to previous charmonium-targeted experiments at Fermilab (E760 and E835), $\overline{\text{P}}$ANDA will run with a factor 10 higher luminosity, a factor 10 better beam momentum resolution, and a detector with greater spatial coverage, magnetic field, and a variety of specialized sub-detectors.

3.2.1.3 Baryon Spectroscopy

The »Review of Particle Physics« lists in its Baryon Summary Table a large number of baryons, for which the evidence is *only fair* or even *poor* [1]. The excitation spectrum is not fully understood; the experimental measurements for the ud-containing baryons (nucleons) do not agree very well with the theoretical predictions of the quark model – the predicted masses differ from the measured masses, some states are not seen at all. Additionally, the internal dynamics of the baryonic states are not well understood. In the realm of baryons with strangeness (Λ, Σ, Ξ, Ω), experimental data is scarce. Established states are not seen with high precision and new states are measured but need confirmation.

$\overline{\text{P}}$ANDA will run with a comprehensive baryon spectroscopy program. In the momentum range of HESR, the ratio of produced baryons to produced mesons is at least $1:1$ (for $p_{\overline{\text{p}}} = 3\,\text{GeV}/c$) and exceeds $2:1$ ($p_{\overline{\text{p}}} > 12\,\text{GeV}/c$). Compared to pp colliders, the energy threshold for creation of baryons with strangeness or even charm is lowered, as no additional D or K mesons need to be produced to conserve strangeness or charm. $\overline{\text{P}}$ANDA expects measuring Ξ, including resonances and diverse decay channels. In addition, the number of measured Ω baryons will be raised by $\overline{\text{P}}$ANDA, with an hourly rate of about 700 in the high-luminosity mode[4]. Charmed baryons, like Λ_c^+ and Σ_c, are also expected to be measured at $\overline{\text{P}}$ANDA, although being more challenging due to their small decay lengths and lower production rate.

3.2.1.4 Gluonic Excitations

The nature of the strong force introduces the possibility of the force-carrying mediator, the gluon, to contribute to the valence quantum numbers. The Standard Model (SM) allows for bound states with excitations of gluons – either as gluonic excitations of valence quark systems (*hybrids*) or as sole gluon states (glueballs). None of the predicted gluonic states has yet been seen clearly and unambiguously in experiments. $\overline{\text{P}}$ANDA is in a good position to measure the states and deepen the understanding of them.

Hybrids In hybrid states, the additional gluonic excitation appears as an excitation of the gluon flux tube between two quarks (Figure 2.6(c)). It is predicted that this excitation adds the quantum number of the gluon ($J^P = 1^{\pm}$) to the quark-antiquark system. For pseudoscalar and vector mesons without any orbital angular momentum (S wave), eight hybrid states are generated. One of them, $J^{PC} = 1^{-+}$, was possibly seen in $\pi_1(1400)$ and $\pi_1(1600)$ mesons at the Crystal Barrel Experiment at LEAR (CERN) [71, 72]. Theoretical predictions foresee it lying at $1.6\,\text{GeV}/c^2$ [73].

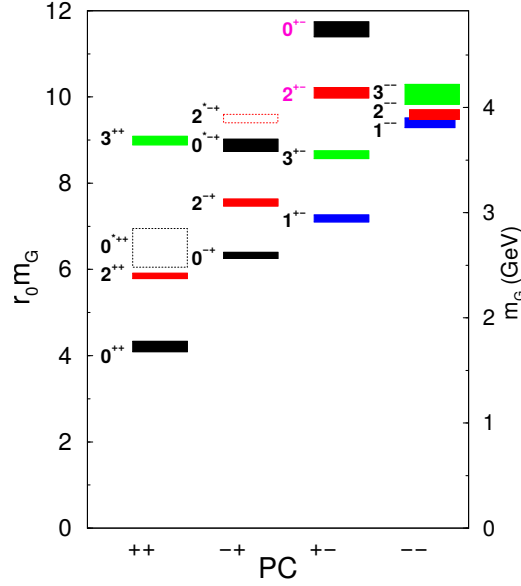[4]A estimated cross section of 2 nb is used to obtain this number.

Figure 3.10: Mass spectrum of glueballs from $SU_C(3)$ gauge theory. Left vertical axis gives masses in multiples of the hadron scale $r_0^{-1} = 410\,\text{MeV}$, right axis gives masses in GeV/c^2. The vertical spread of boxes refers to uncertainty on predicted mass and does not include uncertainty on r_0. The boxes are color-sorted concerning the total angular momentum [73, 74].

Also for heavier mesons with charm, gluonic excited states are expected. For charmonium, the lightest exotic state is expected to have $J^{PC} = 1^{-+}$ and to be located at the $4.3\,\text{GeV}/c^2$ mass region. It is estimated that at P̄ANDA, exotic charmonium hybrids are produced with similar cross sections as the conventional states, around $120\,\text{pb}$ [68]. The challenge in the reconstruction of the hybrid mesons is the discrimination against established states of similar signature.

Glueballs Gluonic excitations without any valence quarks are called *glueballs*. In these cases, bound states of gluons are formed without any mass-bearing quarks. If existing, glueballs create the possibility to study effects of the strong interaction directly. Lattice QCD can predict values and quantum number configurations of glueballs without any additional input to the SM, see Figure 3.10. The ground state is $J^{PC} = 0^{++}$, the first excitation at $J^{PC} = 2^{++}$. The mass is expected to be around $1700\,\text{MeV}/c^2$ and $2400\,\text{MeV}/c^2$, although the calculations still have large uncertainties.

Of particular interest are glueballs with exotic quantum numbers – numbers, which can not be held and mimicked by mesons. The lightest predicted *oddball* has $J^{PC} = 2^{+-}$ and is located at $4.3\,\text{GeV}/c^2$. Decays of glueballs include channels going to ϕ, η, and J/ψ states. P̄ANDA's investigated momentum region is fitting for finding some of the proposed states, $\bar{p}p$ annihilations a distinguished setup for finding also heavier glueballs. A challenge lies in identifying these glueballs, as they mix with ordinary meson states [1, 68, 73].

3.2.2 Nucleon Structure

Apart from physics in the energy regime, at which QCD processes can not be treated pertur-
batively anymore, $\overline{\text{P}}$ANDA will study processes in the transition region towards perturbative
energies. In lepton scattering experiments, Generalized Parton Distributions (GPDs) can be
used to describe hard processes in $\overline{\text{p}}$p annihilations. They deliver insight into the internal
configurations of the nucleon. In $\overline{\text{P}}$ANDA's investigated intermediate region, wide angle Comp-
ton scattering processes ($\overline{\text{p}}$p $\rightarrow$ γγ) can be divided in parts described by GPD (*soft* parts),
and parts for which a perturbative QCD approach is valid, as the involved photon has very
high momentum (*hard* parts) – the so-called handbag approach. The high luminosity and
sub-detector configuration (especially the Electromagnetic Calorimeter (EMC)) enable $\overline{\text{P}}$ANDA
to probe the nucleon structure by studying $\overline{\text{p}}$p $\rightarrow$ γγ with a rate of $\mathcal{O}(1000)$ events per month[5]
despite the small reaction cross section. The division of the total scattering process into hard
and soft parts is valid as well, when one photon is replaced by a pion, $\overline{\text{p}}$p $\rightarrow$ π^0γ, or another
neutral meson, e.g. ρ^0. The π^0 process is another targeted process by $\overline{\text{P}}$ANDA.

Additional nucleon structure analysis at $\overline{\text{P}}$ANDA is achieved by Drell-Yan processes ($\overline{\text{p}}$p $\rightarrow$
$\mu^-\mu^+ X$) and electron-positron creation processes. Drell-Yan processes allow for analysis of
momenta of quarks bound in a nucleon. Electron-positron final states, $\overline{\text{p}}$p $\rightarrow$ e^+e^-, give access
to the time-like region of the proton form factor, and allow for the independent extraction of
the electric (G_E) and magnetic (G_M) form factors.

3.2.3 Hadrons in Matter

$\overline{\text{P}}$ANDA will investigate the effects of hadrons implemented into nuclear matter. The in-medium
spectral function of a meson and its anti-meson partner appear to be split and unequal compared
to the in-vacuum masses. The reason for this appears to be the chiral symmetry breaking pattern
of QCD, and its in-medium restoration. The largest mass shift effects happen when hadrons
are produced at rest, having only little momentum compared to the hadronic matter they are
implemented in. A mass split of $140\,\text{MeV}/c^2$ is predicted for implemented D^+D^- states [75].

$\overline{\text{P}}$ANDA's setup of annihilating antiprotons and protons, releases an energy of about $2\,\text{GeV}$
and enables low-momentum implementation of hadrons into nuclei. Beam momenta up to
$15\,\text{GeV}/c$ allow for production of in-medium D mesons and charmonium mesons.

In addition, contributions to the measurement of the J/ψ absorption cross section in J/ψ-
nucleon reactions are expected. Here, especially the energy dependence of the absorption is
still an open topic [68].

3.2.4 Hypernuclei

Hyperons, i.e. baryons containing strange quarks (Λ, Σ, Ξ, Ω), can be implemented into nuclei,
replacing a proton or neutron. Usually, Λ baryons are considered. Since these baryons only
decay weakly, they form quasi-stable *hypernuclei*, which can be studied spectroscopically. Due to
its strangeness, the particle is not restricted by the Pauli principle for the states it can populate
in the nucleus. Thus, hyperons give insight into the structure of the nucleon, and, vice versa,

---

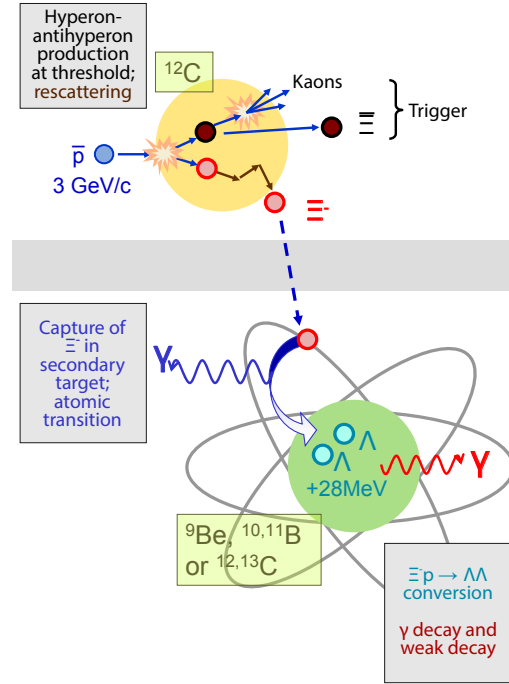[5]Running in high-luminosity mode and $\sqrt{s} = 3.2\,\text{GeV}/c^2$.

Figure 3.11: The principle scheme for hypernuclei creation at P̄ANDA. Ξ^- are produced and then stopped in a secondary target, caught by a nucleus, which is transformed to a double Λ hypernuclei. Adapted from [68].

hyperons can be easily studied when bound in hypernuclei. Hypernuclei extend the chart of nuclides by a third dimension, forming (unstable) related elements with strangeness – matter possibly seen in neutron stars. Since 40 years already, not only hypernuclei with one Λ, but two Λ are known. But only six types of them are measured up to now.

With the HESR's antiproton beam, P̄ANDA will be able to study hypernuclei and double Λ hypernuclei in large quantities. For this, the initial setup of the detector is modified. A carbon target serves as the primary target. The MVD, the innermost tracking detector, is removed and substituted by an instrumented secondary target. Upstream, a Germanium-based photon detector is installed to detect the γ-rays of the excited hypernuclei. The process of producing 2Λ nuclei, as depicted in Figure 3.11, starts with a production of a low momentum Ξ^- in the reaction $\bar{p}p \rightarrow \Xi^-\bar{\Xi}^{+}$[6]. The Ξ^- re-scatters in the target and is eventually stopped in the secondary target. There, the Ξ^--hypernucleus is converted into a 2Λ-hypernucleus. Simulations show, that for a $\bar{p}$ momentum of $3\,\mathrm{GeV}/c$, a cross section per nucleon of $2\,\mu\mathrm{b}$ is expected for $\Xi^- \bar{\Xi}^+$ [68, 76].

---

[6]The quark content of Ξ^- is dss, providing the two strange quarks for the process.

Table 3.3: Overview of the different P̄ANDA sub-detectors and associated components. Green-colored boxes (■) identify parts of the Target Spectrometer and orange-colored boxes (■) identify parts of the Forward Spectrometer.

Sub-detector		Key Aspects	See Sec.
Targets			
Cluster-jet target	■	Initial option, well-known	3.3.1
Pellet target	■	Option for higher luminosities	3.3.1
Tracking Detectors			
MVD	■	Fast, precise tracking close to interaction point	3.3.2.1
STT	■	Large tracking volume, low material budget	3.3.2.2
GEM	■	Fast detectors, in forward direction	3.3.2.3
FTS	■	Straw tube-based system	3.3.2.4
Magnets			
Solenoid	■	2 T field, superconducting	3.3.3
Dipole	■	Large aperture	3.3.3
Particle Identification			
SciTil	■	Low-footprint, fast TOF	3.3.4.1
F-TOF	■	Wall of plastic scintillators	3.3.4.1
Barrel DIRC	■	Cherenkov counter, reflecting light upstream	3.3.4.2
Disc DIRC	■	Wall of fused silica	3.3.4.2
RICH	■	Aerogel- and gas-based ring detection	3.3.4.2
Muon System	■	Four parts, barrel: instruments magnet return yoke	3.3.4.3
Calorimeters			
EMC	■	Two parts, lead-tungstate crystals	3.3.5.1
FSC	■	Sampling calorimeter	3.3.5.2
Luminosity			
LMD	■	Four discs of HV-MAPSs, in vacuum	3.3.6

3.3 P̄ANDA Detector System

P̄ANDA's system of sub-detectors is arranged into two parts. The Target Spectrometer covers the central region around the interaction point and instruments nearly the full 4π solid angle. Polar angles greater than 22° (with respect to the beam axis) are covered by the barrel part of the Target Spectrometer, angles lower than 22° are covered by detectors built as endcaps. Vertically, angles as low as 5° are instrumented, horizontally, down to 10°. The Forward Spectrometer is built downstream of the interaction point to measure forward-boosted particles at even shallower angles. The two parts of the detector are shown in Figure 3.12 and Figure 3.13, an overview is also given in Table 3.3.

P̄ANDA is about 13 m long and has a diameter of about 5 m in the Target Spectrometer. The experiment employs two magnets to discriminate trajectories of charges particles: A 2 T superconducting solenoid in the target region and a large-aperture dipole magnet in the Forward
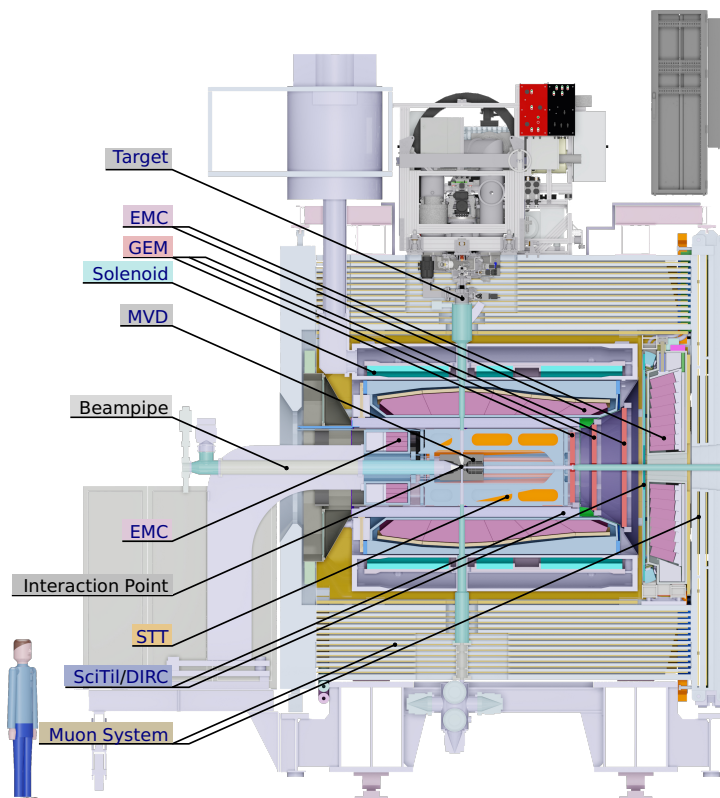
Figure 3.12: P̄ANDA's Target Spectrometer.

Spectrometer. Four tracking detectors, of which three are located in Target Spectrometer, reconstruct tracks of particles with high precision. For particle identification, three basic techniques are used: Time Of Flight, Cherenkov radiation detectors, and a large sampling muon system. Two calorimeters measure the particle energies, one calorimeter system per detector part. Located downstream from the experiment, the Luminosity Detector sub-detector measures the luminosity at the interaction point.

3.3.1   Targets

P̄ANDA employs an internal target to interact with the circulating antiproton beam of HESR. Particles are injected vertically into the beam, inducing the p̄p-processes P̄ANDA studies. In the course of the experimental run, also other target material, e.g. Xe-targets, will be used.

Two targets are in development for P̄ANDA. The *cluster-jet target* comprises a well-known design and will be installed in the first period of P̄ANDA's experimental run. A second target, the *pellet target*, is being designed for runs with higher luminosities.

Cluster-Jet Target   The target material for the cluster-jet target is either hydrogen (H_2) or deuterium (D_2). The target type gets its name from the nano-scaled particle clusters formed from condensed gas. Typically, 10^3 to 10^5 atoms comprise a cluster.
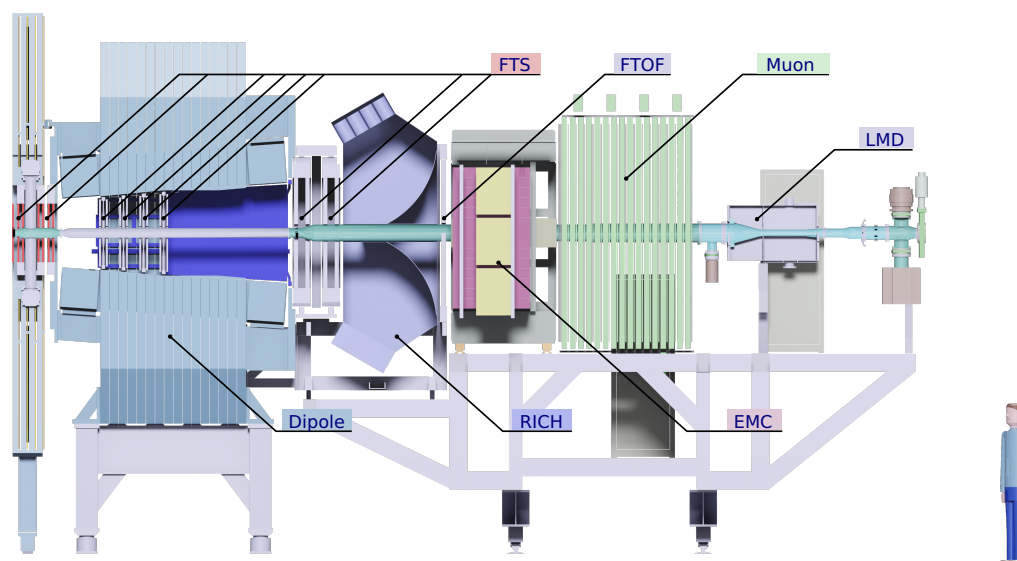
Figure 3.13: P̄ANDA's Forward Spectrometer.
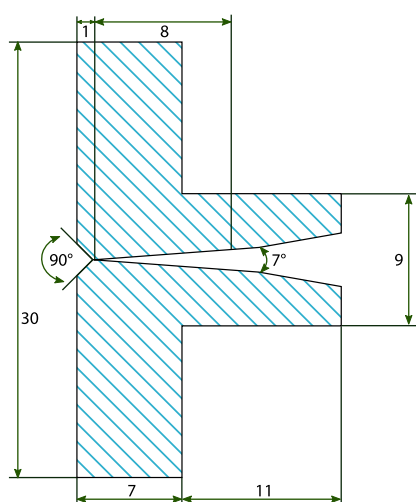
(a) Cut-through drawing of a cluster-jet nozzle. The tapered opening can be seen. Numbers are in mm [77].
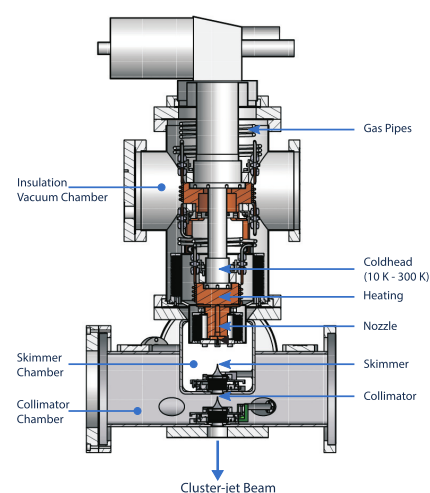
(b) Mechanical drawing of the cluster-jet target [77].

Figure 3.14: Cluster-jet target illustrations.

Clusters are formed when forcing a pre-cooled gas (25 K to 35 K) through a narrow nozzle, adiabatically cooling it in the process. The throat of the nozzle is from 10 μm to 100 μm diameter, see Figure 3.14(a). Under appropriate conditions, a supersonic stream of clusters forms after the nozzle. A skimmer and a collimator shape the jet stream, before it is inserted into the cluster-jet vacuum pipe (Figure 3.14(b)). This pipe interfaces by means of a vacuum valve with the HESR vacuum. A number of turbomolecular pumps, partly in differential setups, create the needed high vacuum with low residual gas for the storage ring at different stages of
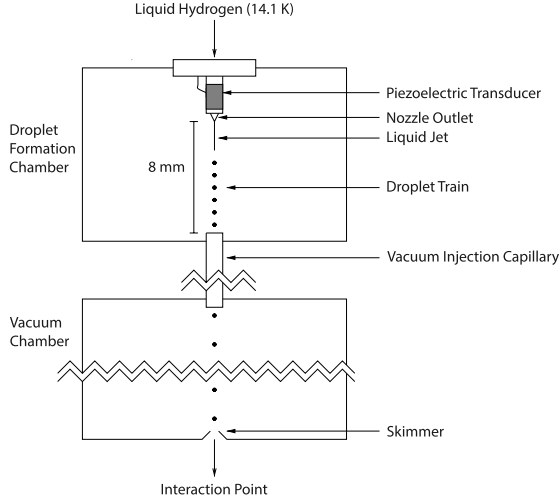
(a) Sketch of a pellet target [78].
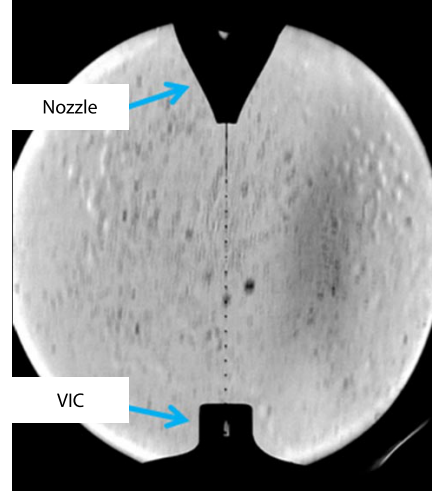
(b) Photograph of pellets emitted by a nozzle into the vacuum injection capillary, VIC [77].

Figure 3.15: Pellet target illustrations.

the particle beam. Below the interaction region, another vacuum valve with associated pumps interfaces to the beam dump.

To align the stream of clusters with the antiproton beam in the interaction region, the individual parts of the target system are adjustable spatially. The nozzle head is tiltable, skimmer and collimator movable.

The areal density of the clusters is 2×10^{15} atom/cm^2, the spread of the cluster in transverse direction is 2 mm to 3 mm, its spread in the longitudinal directions is 15 mm. The size of a cluster is of $\mathcal{O}(\text{nm})$.

Comparing P̄ANDA's target with targets in use in previous rings, the P̄ANDA target will operate with higher pressures: While, for example, targets at CERN used hydrogen at pressures of 10 bar, P̄ANDA's target operates at approximately 25 bar. At the operational temperatures, hydrogen is in its liquid form, creating higher cluster densities, as needed for P̄ANDA's physics program.

Pellet Target The second target in development for P̄ANDA is a pellet target. Instead of a broad jet of target particles, the pellet target provides individual frozen droplets (*pellets*) of greater effective thicknesses for beam interaction. Compared to cluster-jet targets, a higher areal density of 5×10^{15} atom/cm^2 is reached. Only with this can the higher luminosities demanded by P̄ANDA's high luminosity mode be achieved. The large pellet size makes realtime tracking of the pellets feasible – the primary interaction vertex can be measured more precisely, giving high-precision constraints for subsequent analyses of event topologies. The pellets have diameters of 20 μm.

The pellets are produced in the following scheme. The different parts can be seen in Figure 3.15.

1. Gas (hydrogen, deuterium, or elements with higher Z) is cooled and liquified under pressure.

2. The cryogenic liquid is injected into a nozzle. The tip of the nozzle is vibrating vertically using a piezoelectric transducer, cutting the continuous stream of liquid into distinct droplets. The droplet rate, equivalent to the nozzle vibration frequency, is around 40 kHz to 110 kHz. The nozzle ends in a chamber filled with the same initial gas, with conditions, close to its triplet-point.

3. The droplets travel through the vacuum into an injection capillary. During this transit, the surface of the droplets evaporates and cools the droplets down below the freezing point.

4. A skimmer shapes the stream of pellets into a well-defined flux.

5. A pellet catcher is located below the interaction point, which pumps away the evaporating pellets by dedicated vacuum pumps [78].

For manual monitoring, the *triplet-point chamber* is equipped with windows. The stream of droplets can be photographed using a stroboscopic light. For determination of the exact position of the pellet beam interaction, a tracking system of detection lasers will be used.

The pellet target is the choice for $\overline{\mathrm{P}}$ANDA's experimental runs beyond the experiment's starting phase. However, research is still needed, to fix the exact parameters of the device. The first version of a pellet target was in use in the Wide Angle Shower Apparatus (WASA) since 1995 [79].

Other Targets In addition to the previously discussed targets, $\overline{\mathrm{P}}$ANDA may also run with solid targets. Wires or foils of different materials (e.g. copper, silver) are brought into the beam statically. For studies of hypernuclei, a secondary target is needed in the center of the detector, see Section 3.2.4. These alternative targets are options for $\overline{\mathrm{P}}$ANDA's later experimental runs.

3.3.2 Tracking Detectors

The system to reconstruct the tracks of charged particles in $\overline{\mathrm{P}}$ANDA is divided into four parts. The MVD, STT, GEM are located in the Target Spectrometer and the FTS tracks particles in the Forward Spectrometer.

3.3.2.1 Micro Vertex Detector

Located close to the interaction point, the MVD is optimized to detect tracks and vertices of hadrons with strange and charm quarks. In particular, a good efficiency in D meson reconstruction is important for $\overline{\mathrm{P}}$ANDA's physics program. The MVD provides precise timing information and the sub-detector's high granularity read-out is ideal for measuring all charged particles with a high acceptance, providing full three-dimensional hit point information. The information provided in this chapter is based on the »Technical Design Report for the: PANDA Micro Vertex Detector« [80]. Figure 3.16 shows illustrations of the detector.
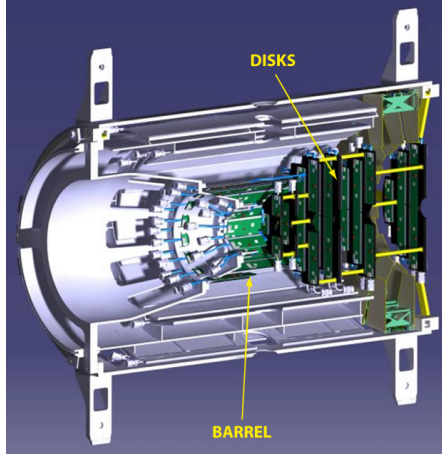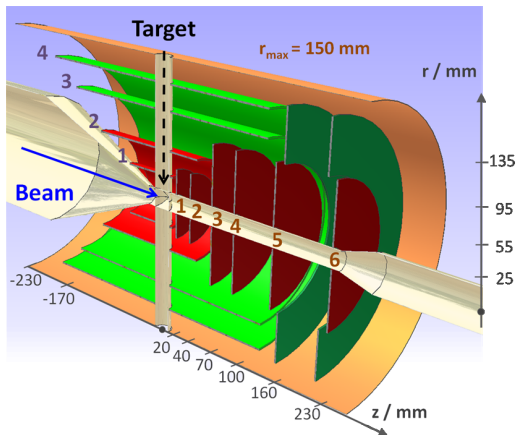
(a) CAD drawing of a cut-through of the MVD. The beam enters from the left. The diameter of the device is about 25 cm.
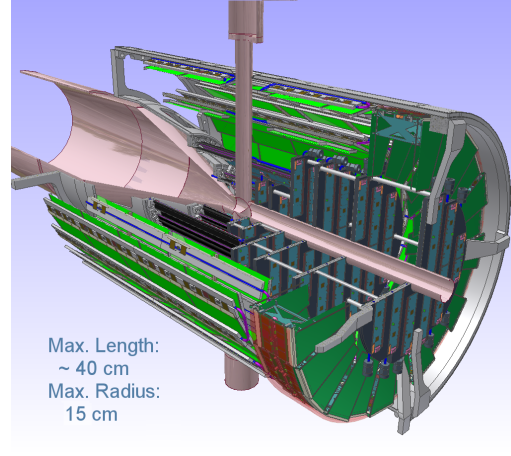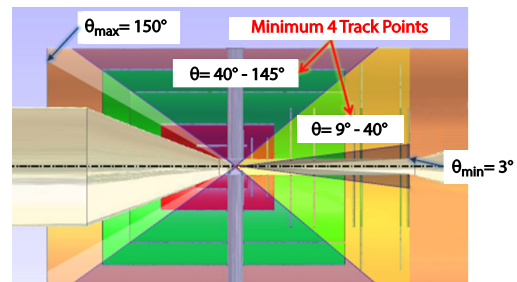

(b) ³/₄ CAD drawing of the MVD, looking upstream. Pixel sensors are colored in turquoise, strip sensors in green.

Figure 3.16: Schematic drawings of the MVD. The interaction point is located where the tapered beam pipe intersects with the vertical target pipe [80].

(a) Positions of individual detectors. Red: pixel sensors; green: strip sensors.

(b) Covered angles by the MVD.

Figure 3.17: Geometrical properties of the sensors of the MVD [80].

The sensor material of the MVD is made of silicon, providing radiation hardness, fast response time, and low material budget for P̄ANDA's innermost detector close to the interaction point. The sub-detector is structured into a barrel and forward disk part. The barrel part consists of four layers, radially located around the center of the detector. Their locations are at radii of 2.5 cm, 5.5 cm, 9.5 cm, and 13.5 cm, see Figure 3.17(a). The two inner layers have a pixel segmentation, allowing for spatially precise measurements in an environment of high rates and track densities, the two outer layers are built as strip detectors, enabling a coverage of a large detector surface. A double-layered design is chosen for the strip detectors, keeping the material budget of the detector read-out to a minimum. Downstream from the interaction point, six forward disks are positioned at 2 cm, 4 cm, 7 cm, 10 cm, 16 cm, and 23 cm. The four closest to the interaction point are built entirely as pixel detectors, the two furthest away feature a design combined of pixel detectors in the inner and strip detectors in the outer regions.

In the polar angle range from 9° to 145° (symmetrical around the beam pipe), four hit points are found at minimum (Figure 3.17(b)). Instrumented are the ranges of 3° to 40° (disk part) and 40° to 150° (barrel part). The active detector silicon area is 0.106 m^2, 0.397 m^2, and 0.073 m^2 for the pixel sensors, the strip sensors in the barrel part, and the strip sensors in the disk part, respectively. The pixel sensors are read out by 10^7 individual channels, the strip sensors by 1.6×10^5 (barrel part) and 5×10^4 (disk part) channels.

The MVD is operated at about 30 °C and normal pressure. All sensors are cooled actively by a low-pressure system (depression system) of water at 16 °C. With this, leaks are avoided and the mechanical influence on the cooled components is kept to a minimum.

As P̄ANDA's basic read-out concept does not rely on triggered data taking, the MVD runs freely. Special read-out chips are developed to work under these special circumstances – ToPix for the pixel part, PASTA for the strip part.

The MVD enables vertex reconstruction with spatial resolutions $< 100\,\mu$m, sufficient for resolving decays of D mesons.

Figure 3.18: Structure of a MVD pixel sensor [80].

Pixel Sensors One silicon pixel sensor is built as a matrix of 116×110 cells, each square cell has a side length of $100\,\mu$m. One sensor, hence, incorporates an area of 1.3 cm^2. The pixel detectors are built as hybrid detectors: The sensor with the active silicon material, running as a reverse-bias diode, is soldered head-to-head on to the accompanying read-out chip (ASIC). Figure 3.18 illustrates the technique. Silicon is epitaxially grown to a maximum thicknesses of $150\,\mu$m on top of a Czochralski (Cz) substrate. After the growth process, most of the Cz substrate is removed. Also, the silicon layer is reduced to $100\,\mu$m. The layer, with its $100\,\mu$m $\times 100\,\mu$m grid structure, is flipped and bump-bonded onto custom developed ASICs. The ASIC is located on top of a special carbon foam, allowing for efficient heat exchange with the internal cooling pipe. The read-out chip, called ToPix, is able to measure the spatial position of a particle passage, the time, and the energy loss. It is made in radiation tolerant 130 nm CMOS technology.
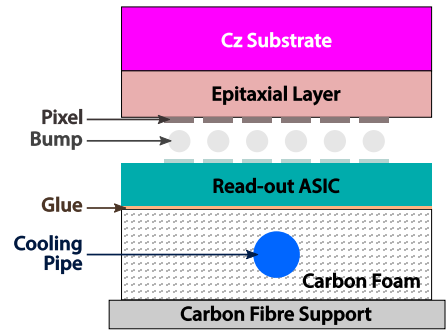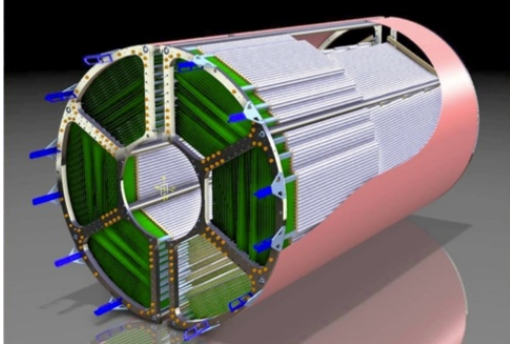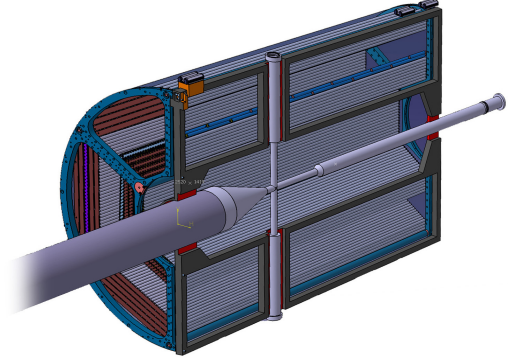
(a) Dimensions of an individual trapezoidal sensor.

(b) Schematic assembly of a MVD disk.

Figure 3.19: The trapezoidal strip sensors of the MVD [80].

Strip Sensors   Three different forms for the strip sensors are used. In the barrel part, square and rectangular shapes are employed, in the disk part, trapezoidal sensors are chosen. The latter sensors are shown in Figure 3.19. To properly measure two-dimensional hit points, all strip sensors are double-sided. The barrel sensors use a stereo angle of 90° between the two sides, the disk sensors an angle of 15°. The *strip pitch*, the distance between two adjacent strips, is chosen to be 130 μm in the barrel strip sensors and 90 μm in the disk strip sensors with an intermediate passive strip.

The active area of the rectangular barrel sensors is 1941 mm^2, the area of a trapezoidal sensor is 1517 mm^2. As the edges of the disk sensors do not have active sensor material, the individual trapezoids are arranged with a slight overlap.

Currently, a dedicated read-out ASIC for the strip sensors of the MVD is in development [81]. PANDA Strip ASIC (PASTA) is connected to one side of the sensors by means of wire bonding and pitch adapters and features a small footprint (21 mm^2). The chip can cope with the high hit rate per channel of 40 kHz – 20× higher than in the pixel part – while delivering fast timing information (< 10 ns). The read-out is cooled by the same cooling system as used in the pixel part.

3.3.2.2   Straw Tube Tracker

The Straw Tube Tracker (STT) is P̄ANDA's central tracking detector. Filling a volume of 1.26 m^3, 4636 small straws enable hit point measurements with resolutions better than 150 μm in the *x-y* plane and a combination of tubes provides better than 3 mm in the *z* direction. The STT is illustrated in Figure 3.20 [82].

Location and Structure   The sub-detector surrounds the MVD and adjoins on the outside to the barrel DIRC. The STT, hence, covers radial distances from 15 cm to 41.8 cm at a length of

(a) CAD drawing of the STT, including the low-weight holding structure.

(b) Half view of STT. The intersection between the target pipe and beam pipe can be seen, the MVD, which surrounds the interaction point, is not shown.

Figure 3.20: Drawings of the STT. The sub-detector is 165 cm long and has a radius of 42 cm [82].

165 cm parallel to the beam axis (z-direction). The detector starts upstream of the interaction region at -55 cm and covers the z-axis to 110 cm. The volume covers polar angles from $10°$ to $140°$.

The straws are organized in 19 layers parallel to z (*axial straws*) and 8 layers skewed by $\pm 2.9°$ with respect to the z-axis (*skewed straws*). The skewed straws are located in a hexagonal shape in the central part of the STT (see Figure 3.21(a)) and enable measurements in the z-direction with a resolution of 3 mm. The straws are closely packed with a minimal space of 20 μm between each straw.

Drift Tubes   Each straw is a small drift tube with an inner diameter of 1 cm. A gold-plated tungsten wire of 20 μm thickness in the center of the straw functions as the anode, the inner wall of the straw is the cathode. The wall is made of a double-layered, aluminized polyester foil (*Mylar foil*) of 12 μm thickness each. The first layer of aluminization is used for electrical conduction, the second for shielding against external light.

The wire inside each straw is kept straight by a tension of 50 g, leading to a maximal sag of < 35 μm. For all the straws of the STT, this adds up to 232 kg. Instead of holding this tension with an external frame that would add to the detector's material budget, the STT choses a light-weight, self-supporting structure: The straw tubes are held at 1 bar overpressure, giving tension to the anode. The pressuring gas, at the same time the drift medium, is a mixture of argon and carbon dioxide (Ar+10 %CO_2). This mixture has beneficial properties in terms of spatial resolution (< 150 μm for all drift distances), measurement rate (80 ns for 4 mm drift path, smaller than P̄ANDA's average time between two events (100 ns)), and radiation hardness (no aging effects in irradiated test straws). The maximum drift time inside a straw is 200 ns, of which about 50 ns are a result of P̄ANDA's magnetic field, since the drift paths inside the straw are curved. With knowledge of the event start time t_0, the drift time inside a straw can be measured. Using a calibration curve, $r(t)$, the isochrone radius is computed. Figure 3.21(b) shows a calibration curve for a test straw with different drift properties.
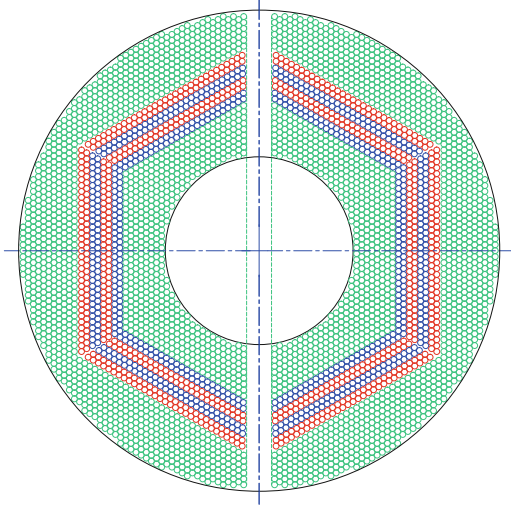
(a) Structure of the STT. Colored in green are the axial straws, colored in red and blue are the skewed straws, with each color denoting a different inclination angle.
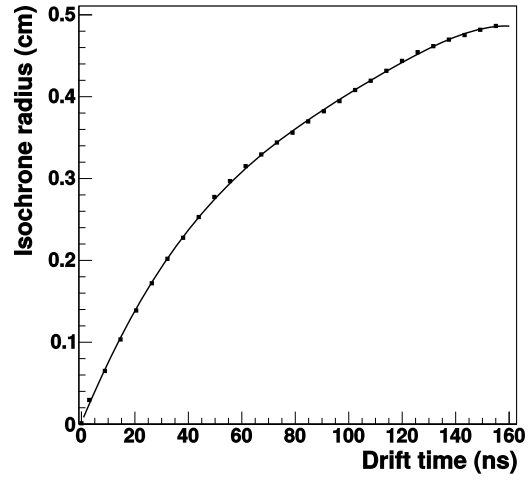
(b) Exemplary calibration curve to connect a given drift time to a isochrone radius. No external magnetic field is present.

Figure 3.21: Drawings of the STT **(a)** and calibration curve **(b)** [82].
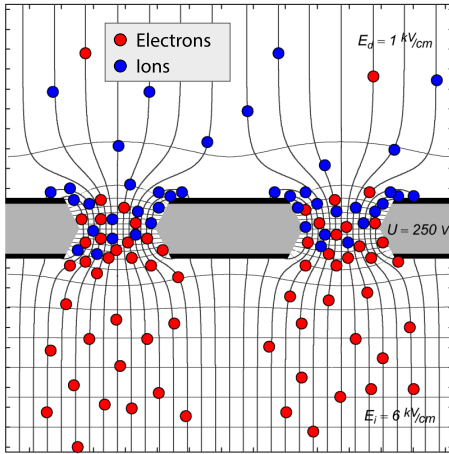
(a) Principle of electron amplification through small holes with applied high voltages [83].
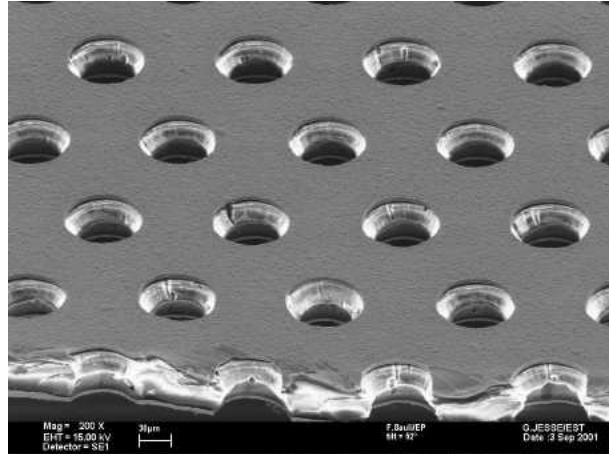
(b) Electron microscope picture of the GEM hole pattern [84].

Figure 3.22: GEM working principle **(a)** and a picture of the GEM foils **(b)**.

The overall radiation length X/X_0 for a particle traversing all 27 layers of the STT radially is 1.23 %, reducing the effects of multiple scattering and energy loss, and enabling precise track reconstruction.
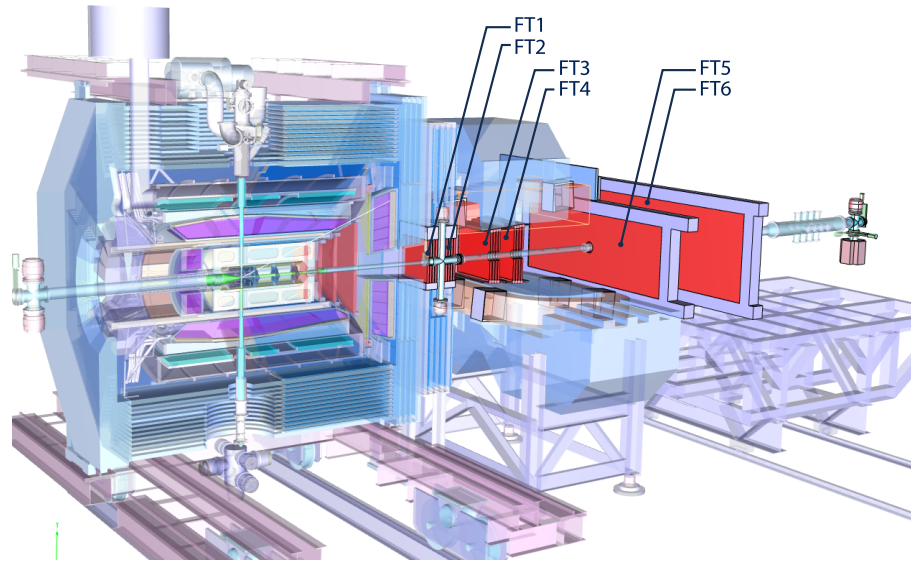
Figure 3.23: The positions of the six stations of the FTS.

3.3.2.3 Gas Electron Multiplier

The GEM (Gas Electron Multiplier) detector tracks particles in the forward direction with three disks. They are located 1.1 m, 1.4 m, and 1.9 m downstream from the target in the Target Spectrometer. Shallow polar angles from 3° to 22° are instrumented. The diameters of the disks are 0.9 m, 1.1 m, and 1.5 m.

For hit detection, the technique of electron multiplication in gaseous micro-pattern foils is employed, see Figure 3.22(a). A system of three stacked GEM foils amplifies electrons to measurable quantities. A foil is typically made from a plastic material and coated with a conducting surface. The CERN-developed GEM detectors use 5 μm thick copper coating on a 50 μm Kapton base. Holes of 70 μm diameter, aligned with a pitch of 140 μm in the foils (Figure 3.22(b)), provide the needed electron amplification due to the different voltages applied to the two conducting sides.

In each GEM disk, a read-out pad is located in the longitudinal center. Both downstream and upstream from the pad, foil systems amplify signals. A cathode layer eventually closes up the disk. Per station, four projections are available: A radial and a concentric fragmentation, and one each in the x and y direction. About 35,000 channels are read out.

The GEM detector measures hit points with resolutions < 100 μm.

3.3.2.4 Forward Tracking System

The Forward Tracking System (FTS) is dedicated to measuring tracks of particles boosted in the forward direction. In particular, the amount of deflection of a particle's trajectory due to the Forward Spectrometer's dipole is analyzed.

As shown in Figure 3.23, six FTS tracking stations are planned. The first two, dubbed FT1 and FT2, are located in front of the dipole and directly after the Target Spectrometer, the last two, FT5 and FT6, are located after the dipole and in front of the RICH detector. FT3 and FT4 are located inside the dipole. Built perpendicular to the beam direction, the planes cover angles of $\pm5°$ in the vertical direction and $\pm10°$ in the horizontal direction. The area is defined by the aperture of the dipole.

Each tracking station is equipped with four double layers of straw tubes. The outer layers have vertically-aligned straws, while the straws of the two center layers are inclined by $+5°$ (second layer) and $-5°$ (third layer). With this, three projections per station are provided to enable two-dimensional position measurement. The straws themself are identical to the ones used for the STT: A Mylar tube of 10 mm inner diameter, with a gold-plated tungsten anode wire in its center, operated at 1 bar overpressure with Ar+10 %CO_2 gas. The wire is operated at 1800 V.

FT1 and FT2 both have 1024 straws, FT3 and FT4 1536 straws; FT5 has 3200 straws, FT6 4736 straws. In total, over 13,000 straws instrument the FTS, more than twice as many as in the STT. The stations are located at 2.954 m, 3.274 m, 3.945 m, 4.385 m, 6.075 m, and 7.475 m from the interaction point. They span areas of 0.83 m^2 to 8.9 m^2 for FT1 and FT6, respectively.

The maximum drift times in the straws operated outside the dipole are 130 ns. The magnetic field adds more time to this value for the stations located inside the dipole, as drifting electrons inside of the straws are forced to curved trajectories – for the maximum beam momentum[7] of 15 GeV/c another 20 ns are added. Positions can be resolved with $\mathcal{O}\left(100\,\mu m\right)$, leading to momentum resolutions $< 1\,\%$. On average, one charged particle per event will reach the FTS – the maximum multiplicity is eight [85].

3.3.3 Magnets

P̄ANDA will employ two magnets – in the Target Spectrometer a superconducting solenoid, in the Forward Spectrometer a dipole magnet. Both magnets are described in the »Technical Design Report for the PANDA Solenoid and Dipole Spectrometer Magnets« [86].

Solenoid   To provide a homogeneous magnetic field in the beam direction to the tracking detectors, a superconducting solenoid is chosen. NbTi superconducting cables create a field of 2 T longitudinal strength. The iron return yoke is external and houses the muon chambers (see Section 3.3.4.3) in a sampling fashion. Illustrations of the magnet is shown in Figure 3.24(a). The total weight of the solenoid is > 300 t, most of it from the yoke.

The magnetic field is supplied over a length of 4 m in the central opening hole of the magnet of 1.9 m diameter. Excluding the muon chambers, all sub-detectors of the Target Spectrometer are housed inside the solenoid. In the volume of the STT, a field inhomogeneity $\leq 2\,\%$ is achieved (see Figure 3.24(b)). Three sub-coils create the magnetic field, operating with 5000 A current and incorporating a gap for the target pipe.

---

[7]As the magnetic field of the dipole also influences the antiprotons circling in HESR, the dipole field is ramped with the accelerator machines. See Section 3.3.3.
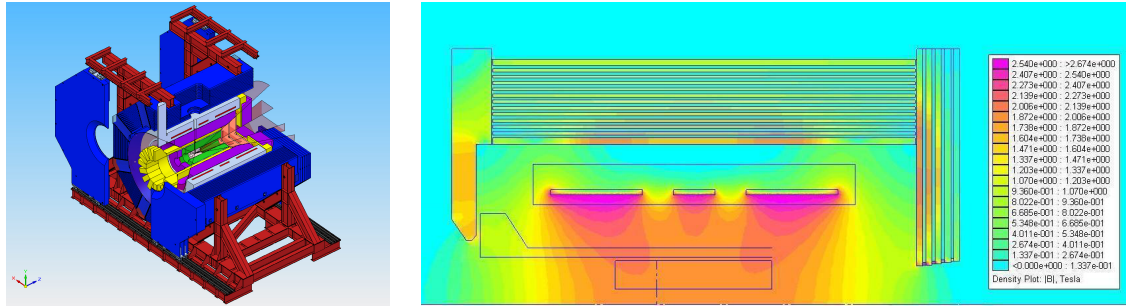
(a) CAD drawing. The return yoke is colored in blue, the sub-coils in red, surrounded by the accompanying coil system (cryostat, coil former) in light blue.

(b) Magnetic flux density in the Target Spectrometer. On top: The return yoke; on bottom: STT.

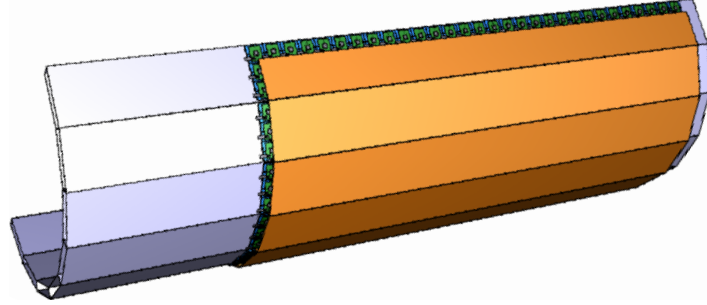**Figure 3.24:** Illustrations for the solenoid magnet [86].

(a) CAD drawing of the closed magnet, looking from the downstream direction.

(b) CAD drawing of an exploded view.

(c) Strength of the y component of the magnetic field in the z - y plane.
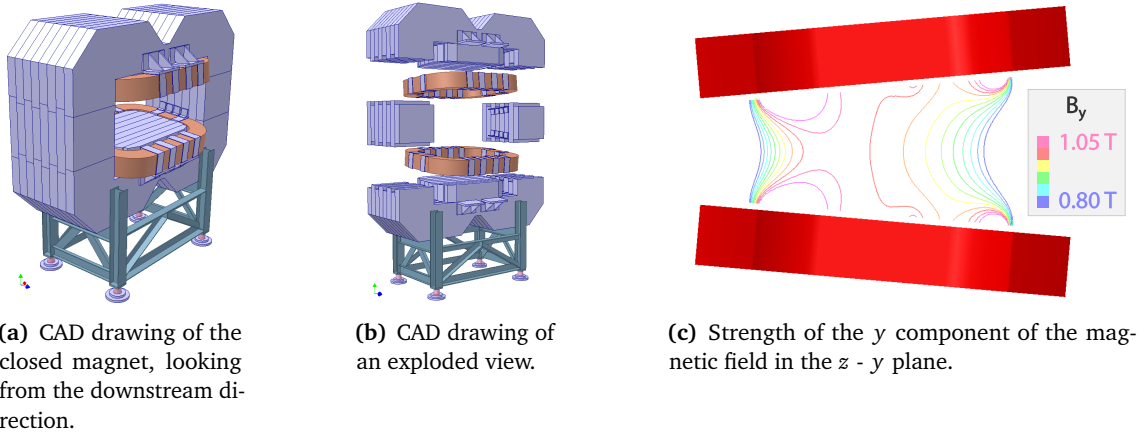
Figure 3.25: Visualizations of the dipole magnet. The magnet has a length of 2.5 m in z-direction [86].

The flux return yoke is structured into 13 layers, with an outer radius of 2.3 m. Each steel plate is interleaved with a 30 mm gap for the individual muon chambers. The plates are 30 mm thick, with the inner and outer-most plates twice as thick, see also Figure 3.29.

Dipole Located in the Forward Spectrometer, the dipole magnet bends trajectories of particles emitted into forward angles – angles of ±5° vertically and ±10° horizontally are covered. In contrast to the solenoid, a resistively conducting magnet is chosen. Reasons are a lower cost, easier operation, a well-known design, straight-forward handling of the magnet's weight, and no need for cooling (with additional structures, lowering the magnet's acceptance). The magnetic field of the dipole is shielded against the field of the solenoid by a field clamp. The magnet is ramped together with HESR, as its field influences also circulating particles in the beam. The magnet, hence, is part of the accelerator's lattice. The dipole deflects the beam by 2.2° at a beam momentum of 15 GeV/c and bending power of 2 T m. To compensate this, one correction dipole is placed before P̄ANDA, one after.
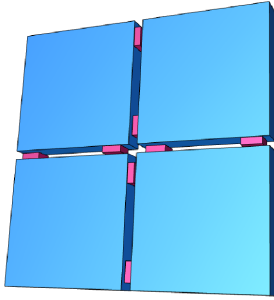
(a) A module: Four tiles (blue), read out with two SiPMs (purple) each.

(b) Half of the SciTil barrel. In orange: super-modules.

Figure 3.26: Illustrations for the SciTil. The side length of a module **(a)** is 6 cm, the lengths of the super-modules **(b)** are 1.8 m [87].

Two race track-type coils with an opening angle of ±5° form the magnetic field of 2 T m, see Figure 3.24. The gap between the coils is 0.80 m × 3.10 m for the opening and 1.01 m × 3.10 m at the end. They are made of copper and weigh 18 t.

The dipole is 2.5 m long, 3.9 m high, and 5.3 m wide. The return yoke is made of steel and built in segments. This gives beneficial properties for ramping the magnet together with HESR and also simplifies construction in the experimental hall. In total, the dipole weighs about 220 t.

3.3.4 Particle Identification

3.3.4.1 Time Of Flight Detectors

P̄ANDA is equipped with TOF detectors in the barrel part and in the forward part.

Scintillating Tiles Located in the Target Spectrometer between the DIRC on the inside and the calorimeter on the outside, scintillating tiles give precise timing information of traversing particles, e.g. for software triggering. The sub-detector is usually referred to by the name SciTil. It has a low material budget (<2 % radiation length), ensuring good performance of the calorimeter, and a low total radial thickness (<2 cm), fitting into the tight space between DIRC and calorimeter. The position is particularly beneficial for detecting photon conversions in the sub-detectors enclosed in the calorimeter (e.g. DIRC), improving the resolution of neutral particles, e.g. π^0 [87].

The tiles are made of a plastic scintillator, *BC-408*, and have a 28.5 mm × 28.5 mm surface area and 5 mm thickness. The material is chosen to have a good light yield and timing resolution: A minimum ionizing particle losing 1 MeV in the 5 mm creates 10,000 photons, of which 120 photons can be measured. The timing achieved with this is $\mathcal{O}(100\,\mathrm{ps})$. Two Silicon Photo Multipliers (SiPMs) read-out a tile on adjacent sides.

Four tiles are combined into a module (Figure 3.26(a)) and read out by an 8 channel ASIC – one for each SiPM in a module. 3 × 30 modules form a *super-module* (3 in the azimuthal
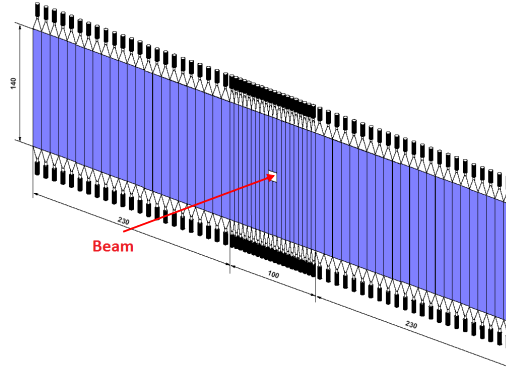
Figure 3.27: The structure of the Forward TOF. The total width is 5.6 m [88].

direction, 30 in z), from which 16 are used to build up the whole SciTil barrel (Figure 3.26(b)). In total, 5760 tiles and 11,520 SiPMs comprise the sub-detector. The acceptance is −50 cm to +150 cm in the z direction. A super-module is cooled by flowing cooled dry air through it.

Forward TOF The relative timing information of particles to enable separation of π/K and K/p is provided by a wall of plastic scintillators that is located 7.5 m downstream from the interaction point, right after the RICH detector. The targeted time resolution of the Forward TOF (F-TOF) is 50 ps to 100 ps [89].

As in the barrel part, BC 408 is used as the scintillating material. The wall is made of 66 plates of 140 cm length and 2.5 cm thickness, see Figure 3.27. The center-most 20 plates have a width of 5 cm, the surrounding plates a width of 10 cm. Each plate is read out by two photomultiplier tubes (PMTs), one on each side.

Also, the side walls of the dipole magnet are equipped with scintillators to measure low momentum tracks not able to leave the magnet.

3.3.4.2 Cherenkov Detectors

For particle identification, PID, detectors of two different kinds are installed to detect Cherenkov light of particles emitted when passing a material faster than the speed of light in this medium.

DIRC Detectors Detectors employing Detection of Internally Reflected Cherenkov Light (DIRC), detect Cherenkov light created in confined spaces. The light is transferred to photon detectors by means of total internal reflexion, preserving the opening angle of the light cone, and thus information on the velocity of the creating particle. The opening angle θ_c of the light cone is connected to velocity of the particle by $\cos\theta_c = 1/(\beta n)$, with β being the particle velocity normalized to the vacuum speed of light, and n the refraction index of the material. The combination of DIRC information of a particle's velocity with momentum information measured in the tracking detectors yields the particle's mass, fully identifying the kind of particle [92].

P̄ANDA's DIRC-employing detectors are segmented into a barrel DIRC and a disc DIRC in the forward direction. See Figure 3.28(a) and Figure 3.28(b), respectively.
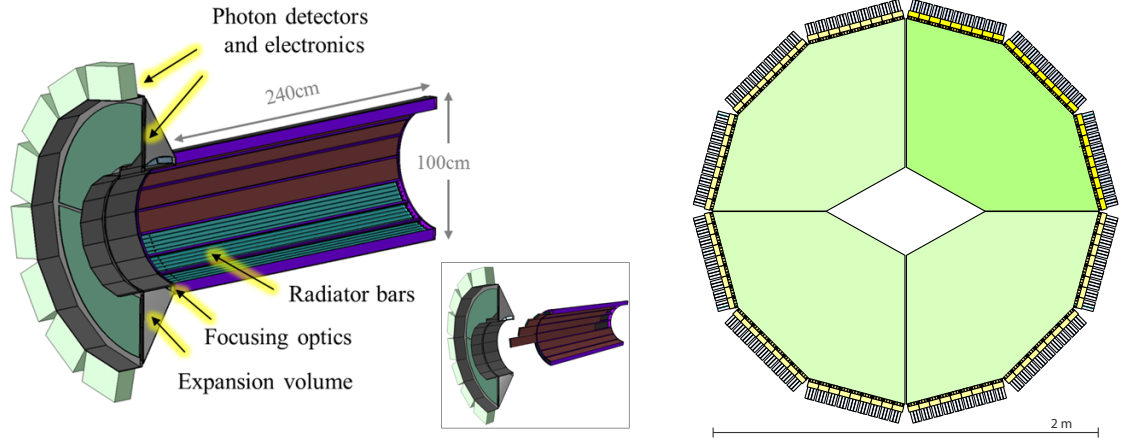
(a) The barrel DIRC system. The radiator bars (turquoise) are arranged in sets of five to form bar boxes (red). The expansion volume in the backward direction is colored in a dark shade of green, with the photon detection in lighter green [90].

(b) The disc DIRC. The radiator is colored in green and segmented into four pieces. They are read out through the surrounding instrumented focal plane. One quarter section is highlighted [91].

Figure 3.28: The DIRC detectors of P̄ANDA.

Barrel DIRC The radiator bars of the barrel DIRC are located between the SciTil on the outside and the STT on the inside. They cover polar angles of 22° to 140° and an azimuthal range of nearly 2π, leaving only a gap for the target pipe. Each bar is made of synthetic fused silica, with beneficial properties with respect to radiation hardness, light transmission, and dispersion. A bar has the dimensions of 2400 mm × 17 mm × 32 mm and is arranged in a set of five bars to form a *bar box*. 16 of the boxes are placed around the beam line at a radial distance to the beam axis of 476 mm.

To keep the material budget inside of the calorimeter as low as possible, the barrel DIRC is read out in the upstream direction. Mirrors attached to the forward ends of the bars reflect the Cherenkov light backwards. There, it enters an instrumented expansion volume through a focusing lens. The 30 cm deep volume is filled with mineral oil with a refraction index close to the index of the silica. Photodetectors, based on PMTs, detect the Cherenkov light with 15,000 channels.

With the envisaged setup, pions and kaons can be separated with ≥ 3σ for a momentum range up to 4 GeV/c [91, 93].

Disc DIRC The disc part of the DIRC detector covers polar angles between 22° and 5° (vertically) / 10° (horizontally) and is installed in an endcap-fashion in the Target Spectrometer, between the GEM detectors and the calorimeter. Like the barrel part, the disc as well is made from fused silica. The disc has a diameter of 2200 mm and a thickness of 15 mm [94].

The disc is approximated by a polygon with 128 sides. On each straight surface on the outside, a focusing light guide is glued to the silica, with a dichroic filter in the transmission region, see

(a) The four parts of the muon system, colored in blue. The number of Mini Drift Tubes (MDT) is given.

(b) The four modules of the forward muon range system.
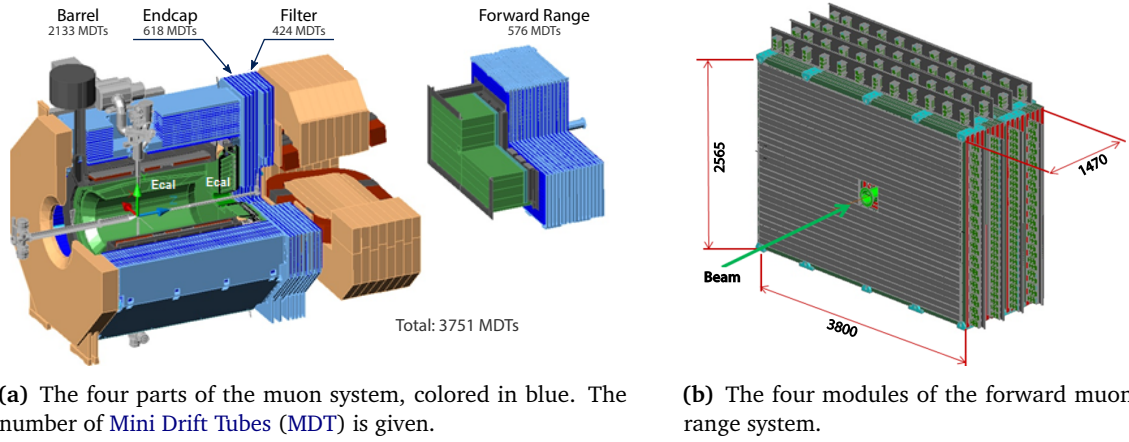
Figure 3.29: Overview of the muon system of P̄ANDA [96].

Figure 3.28(b). The light guide is read out by a number of PMTs in the respective focal planes. With this, pions and kaons can be separated up to 4 GeV/c momenta.

RICH Detector   In the Forward Spectrometer, a detector employing detection of Cherenkov rings, RICH (Ring-Imaging Cherenkov), is installed. In the polar angle range of 0° to 10°, it supports the separation of π and K mesons.

Located approximately 6.5 m from the interaction point between the last two forward TOF stations, the detector uses Aerogel and gas for Cherenkov light creation. The light is reflected with a mirror to PMTs, located outside of the magnetic field of the dipole magnet [95].

3.3.4.3   Muon System

P̄ANDA's system for muon detection is divided into four parts – three in the Target Spectrometer, one in the Forward Spectrometer, see Figure 3.29(a). The **barrel muon system** surrounds the solenoid magnet cylindrically and functions as its return yoke. The same function is held by the **muon endcap** in the downstream direction. Located after the endcap, a removable **muon filter** additionally discriminates forward-boosted particles and forms the magnetic field lines in the intersection region of the solenoid and the dipole. Finally the **forward muon range system**, which is followed by only the Luminosity Detector, detects high-momentum muons at low angles.

The muon system is built as a range system, with detecting layers alternating with passive iron layers. Muons in the momentum range of 1 GeV/c to 10 GeV/c can be measured.

The active sensors are Mini Drift Tubes (MDTs), operated in the proportional mode. They provide the needed resolution (in both space and time), a flexible and simple design, are robust and proven already in other experiments. The principle structure is shown in Figure 3.30(a). Gold-plated tungsten anode wires reside in slots of a aluminum profile with wall thicknesses of >0.45 mm. The pitch between the wires is 10 mm, optimized for the needed resolution. For gas tightness, the profile is covered in a plastic envelope. This classic MDT design does not
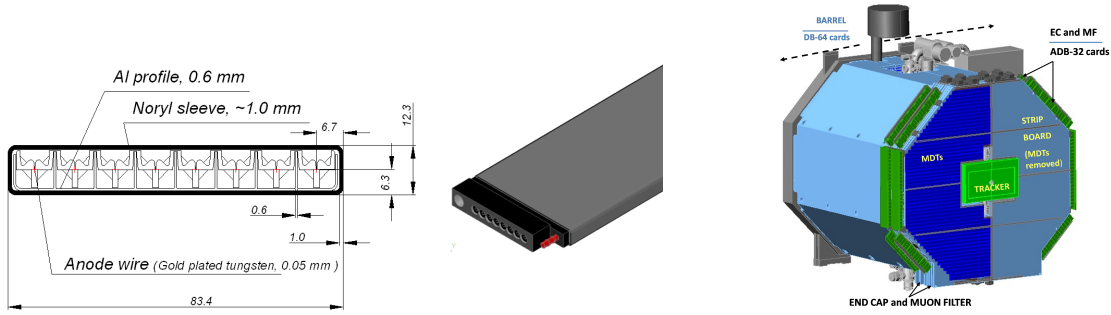
(a) Structure of the Mini Drift Tubes. In a comb-like structure of aluminum, anode wires are embedded. For insulation reasons, the module is covered in plastic, e.g. Noryl. On the right, a ready-built module with its connectors is shown.

(b) The Target Spectrometer part of the muon system, with the endcap and filter as disk-like structures in front, and the barrel part in back.

Figure 3.30: Overview of the muon system [96].

provide information about the position of the hit along the wire and can, hence, deliver only one-dimensional data. For a two-coordinate measurement, external strip electrodes are added to the MDTs to measure the induced signal by a particle transition. The system of MDT and strip sensors is 15 mm thick.

Barrel Muon System   13 layers of sensors, each 3 cm thick, alternate with iron absorbers of the same thickness. For mechanical reasons, the first and last iron layers are 6 cm thick. The first active layer on the inside, the *0-layer*, is made of two MDTs with a strip layer in between and the outer MDT shifted by half a channel pitch. With this, accurate three-dimensional information can be provided for particles entering the muon systems. In total, 2133 MDTs will be installed.

The barrel part is built in eight sections, each with a mass of 23.7 t – nearly solely the mass of the iron of the return yoke (97 %). In total, the barrel part weighs 189.5 t.

Muon Endcap   The barrel muon endcap consists of 6 detecting layers, intersecting with 5 absorbing iron layers. To accommodate the higher momentum of forward-boosted particles, the iron layers are twice as thick as their barrel counterparts, 6 cm. 618 MDTs are installed. Weighing in total 46.5 t, the muon endcap is part of the downstream return yoke door, see Figure 3.30(b).

Muon Filter   The muon filter is an additional disk-like absorber structure, located directly behind the muon endcap. It is very similar to the endcap, the only difference being one absorber layer and one sensor layer less. 424 MDTs are installed.

Forward Muon System   Located approximately 9 m downstream from the target, directly after the forward calorimeter, the forward muon range system discriminates pions from muons for high momenta, detects pion decays, and, with a coarse resolution, energies of neutrons and antineutrons, functioning as a hadronic calorimeter.
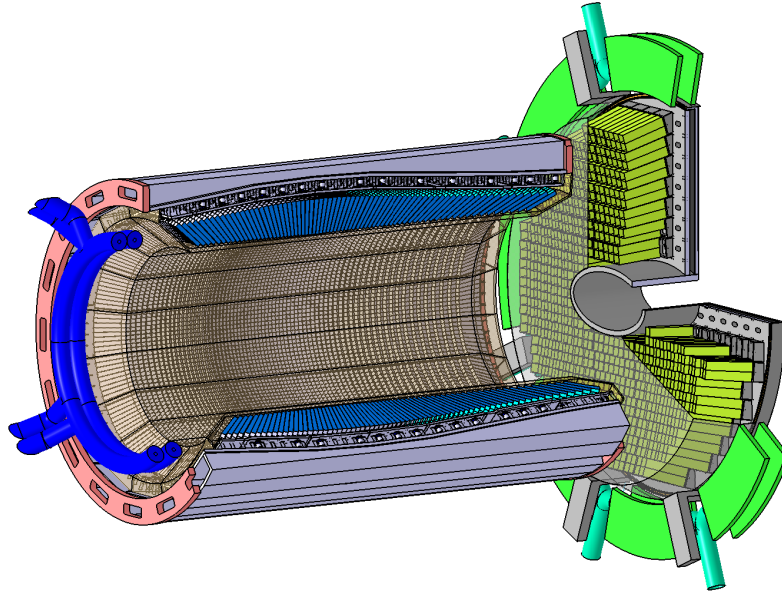
Figure 3.31: Cut through a CAD drawing of the target Electromagnetic Calorimeter. The barrel part of the calorimeter is colored in shades of blue, the forward endcap is colored in shades of green. The backward endcap is not shown, it is located as an inset in the very left of the barrel part [97].

The general structure is just like the muon detectors in the Target Spectrometer: 16 layers of iron are intersected with the same amount of MDT-instrumented layers (Figure 3.29(b)). The first layer, again, is a double measuring *0-layer*. Per layer, 43 MDTs are installed. The forward muon system has a mass of 77 t.

To accompany the slightly skewed beam line, the forward muon system is built in four identical modules, assembled around the beam pipe each with a small offset.

3.3.5 Calorimeters

The electromagnetic calorimeter is structured two-fold in P̄ANDA. The Target Spectrometer has a big, nearly 4π-covering calorimeter to measure particles close to the interaction point. In the Forward Spectrometer, located right before the muon system, a sampling calorimeter, built in a *Shashlik* fashion, measures forward-boosted particles.

3.3.5.1 Electromagnetic Calorimeter

The EMC, as the calorimeter in the Target Spectrometer is called, is divided into three parts: The **barrel calorimeter** is located in the central region, inside of the solenoid and outside of the SciTil. It covers a length of 2.5 m, starting at an inner radial distance of 0.57 m and ends at 0.94 m. The polar range of 22° to 140° is instrumented. The **forward endcap** closes the barrel up in the downstream direction, the **backward endcap** upstream. The forward endcap is located 2.1 m downstream from the interaction point and has a diameter of 2 m, covering angles of 5° to 23.6° (vertically) and 10° to 23.6° (horizontally). The backward endcap is
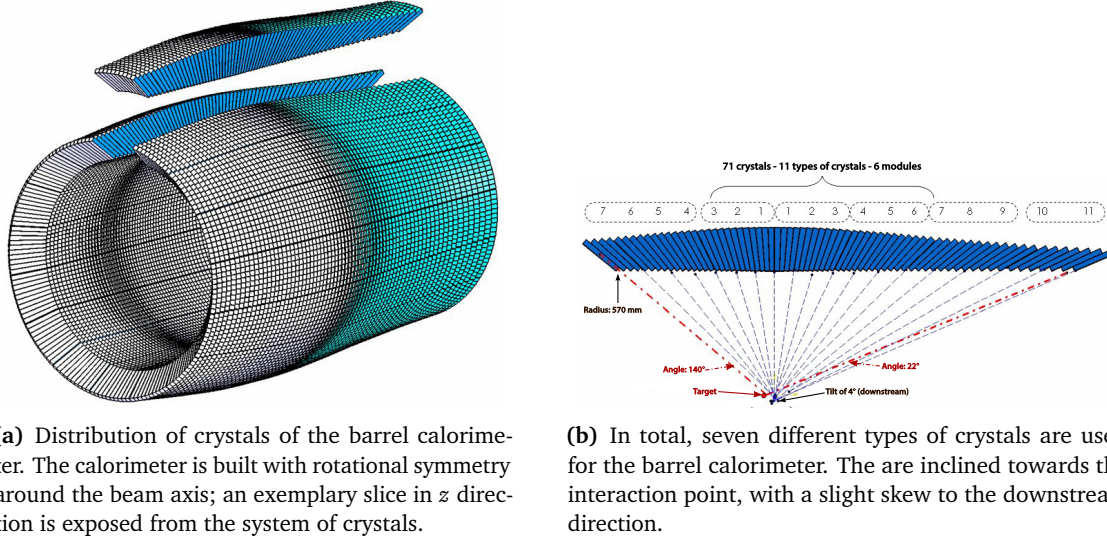
(a) Distribution of crystals of the barrel calorimeter. The calorimeter is built with rotational symmetry around the beam axis; an exemplary slice in z direction is exposed from the system of crystals.

(b) In total, seven different types of crystals are used for the barrel calorimeter. The are inclined towards the interaction point, with a slight skew to the downstream direction.

Figure 3.32: The crystals of the barrel electromagnetic calorimeter [97].

located at $z = -1\,\text{m}$ and has a diameter of $0.8\,\text{m}$. It covers polar angles from $151.4°$ to $169.7°$. The barrel and the forward endcap is shown in Figure 3.31.

The calorimeter has a low photon energy threshold of $10\,\text{MeV}$, with single crystals at $3\,\text{MeV}$. An energy resolution of $\sigma_E/E = \leq 1\,\%$ is possible.

Readout   As the sub-detector is located inside the $2\,\text{T}$ magnetic field, a readout of the crystals with a PMT is not possible. Silicon-based avalanche photodiodes with large detection areas of $14\,\text{mm} \times 7\,\text{mm}$ have been developed. Two photodiodes are installed on the outside surface area of each crystal. In the forward endcap, the high particle rate is the main challenge. Here, vacuum phototriodes are used, able to measure the projected $500\,\text{kHz}$ rate while also being radiation hard.

Crystals   The calorimeters use $PbWO_4$ as a high-density, inorganic scintillating material. The compound is already in use in the electromagnetic calorimeter of CERN's Compact Muon Solenoid (CMS) experiment for several years. Its light signal is short ($10\,\text{ns}$) and it is radiation hard. Compared to the initially developed crystals for CMS, PWO, the crystals used for $\bar{\text{P}}$ANDA, dubbed PWO-II, have a greater light output. The crystal quality is improved, its crystalline structure more perfect, and the operation at $-25\,°\text{C}$ additionally increases the light yield compared to operation at room temperature.

Structure   The system is cooled by silicone oil and insulated from the surrounding sub-detectors by a thermal shield. Each crystal measures $200\,\text{mm}$ in length, equivalent to 22 radiation length X_0, built in a truncated pyramid shape. They have a surface area of approximately $20\,\text{mm} \times 20\,\text{mm}$, the exact dimension depending on the location of each crystal in the calorimeter.
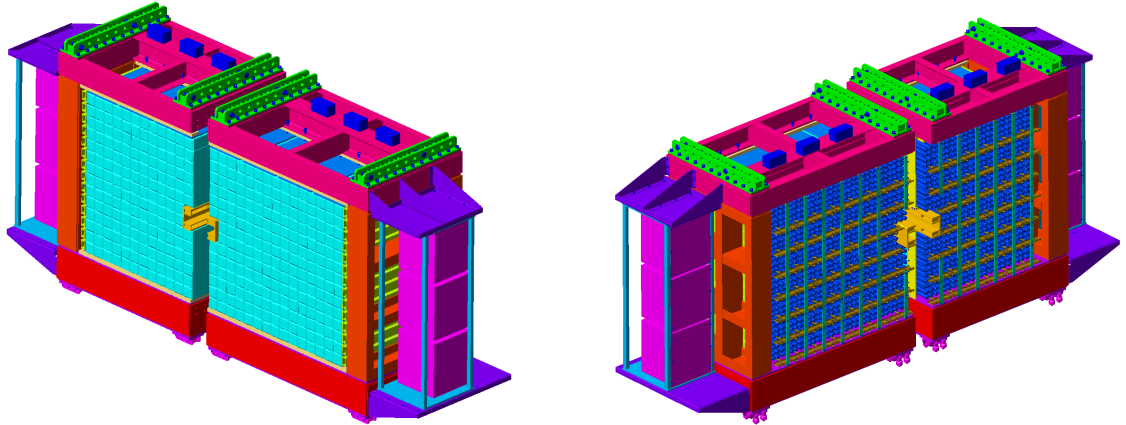
(a) Front view of the Forward Spectrometer Calorime-
ter. The cells with the scintillating plates and ab-
sorbers are colored in turquoise, surrounding the
yellow-colored hole for the beam pipe. The holding
structure cases the modules.

(b) Back view of the Forward Spectrometer Calorime-
ter. In blue, PMTs for readout of the crystals on the
backside.

Figure 3.33: CAD renderings of the Forward Spectrometer Calorimeter, located 7.8 m downstream from
the interaction point. The sub-detector is 3.6 m wide and 1.5 m high [98].

The crystals are installed pointing into the direction of the target, with a slight off-set of a
few degrees to reduce the dead zone between the crystals for particles originating form the
interaction point. They are held by a carbon fiber holding structure and wrapped into highly
reflective foil. The distance between two crystals is 600 μm.

Each crystal weighs on average 0.98 kg, depending on the exact shape. In the barrel, the
crystals are arranged in 16 slices of 710 crystals each, totaling to 11,360 crystals. The forward
endcap employs 3600 crystals, the backward cap 592.

3.3.5.2 Forward Spectrometer Calorimeter

the Forward Spectrometer Calorimeter (FSC) is located in the Forward Spectrometer, after
the last TOF wall and before the forward muon system – starting 7.8 m downstream from the
interaction point and spanning to 8.95 m. With dimensions of 3.6 m × 2.2 m it instruments the
area of angles <10° horizontally and <5° vertically left unoccupied by the forward endcap of
the target calorimeter. The active area is 2.97 m × 1.54 m. In total, the FSC has a mass of 3.7 t.
Views from the upstream and from the downstream direction are shown in Figure 3.33.
Due to its design with Wavelength-Shifting Fibres (WSF), the FSC is also called the *Shashlik
Electromagnetic Calorimeter* [98].

To cope with the potentially high-momentum particles in the forward direction and measure
energies in an optimized space, the FSC is constructed as a sampling calorimeter. Absorber
plates alternate with scintillating plates to form a detection cell of the calorimeter. A radiation
length of 19.6 X_0 is achieved.

(a) Fully assembled module. Four cells (turquoise), inhabit the scintillating and absorbing layers. Plugs for attaching the PMTs are colored in orange.

(b) Exploded view of the backside of a module. The Wavelength-Shifting Fibres can be seen coming out (green). Blue screws hold the structure together. Colored in a lighter blue is a light monitoring system.

(c) Exploded view of the front side of a module. The Wavelength-Shifting Fibres (green) form loops and re-enter the cell. Colored in purple are protective and shielding cases.

Figure 3.34: CAD visualizations of the modules of the Forward Spectrometer Calorimeter. The length of a module is 1.2 m [98].

Module Structure   1512 cells, divided into 378 modules, make up the detector. The cells are aligned into 54 pieces horizontally and 28 pieces vertically. Each cell's face has an area of 5.5 cm × 5.5 cm.

In a module, four cells share a common holding and read-out structure, see Figure 3.34(a). The size along the beam axis is 1.2 m. 380 layers of absorbers and scintillators alternate over a length of 684 mm: each 1.5 mm thick tile of scintillator is accompanied by a 0.3 mm thick absorber. The scintillator layers of a cell are placed in between reflective foils, with the sides painted with reflective paint. While the scintillating layers are part of a cell, the four cells of a module share the same absorber plate. The sandwiches of each module are held together by pressure plates on the front and on the back.

An organic plastic, polystyrene, is used as the scintillating material in the FSC. It is doped with 1.5 % para-terphenyl and 0.04 % POPOP. Lead, doped with 3 % antimony to increase rigidity, is used for the absorber plates.

Wavelength-Shifting Fibres   WSF are used to collect light in the scintillating layers and transport it out to the photon detectors at the backplate. 18 fibers of 1 mm diameter pierce through each cell in the longitudinal direction (Figure 3.34(b)), giving the Shashlik Electromagnetic Calorimeter its name. Each WSF is bent on the front side of the cell to form a loop, see Figure 3.34(c). With this, 36 fiber ends are located on the back side of each cell, combined to a bunch. The fibers are connected to fast PMTs for photon detected, one PMT per cell, located behind the crystals in a module. The PMTs offer the dynamic range and high rate capabilities needed in the very forward region of the FSC. The PMTs are read out by adapted sampling Analog-to-Digital Converter (ADC) modules, developed for the calorimeter
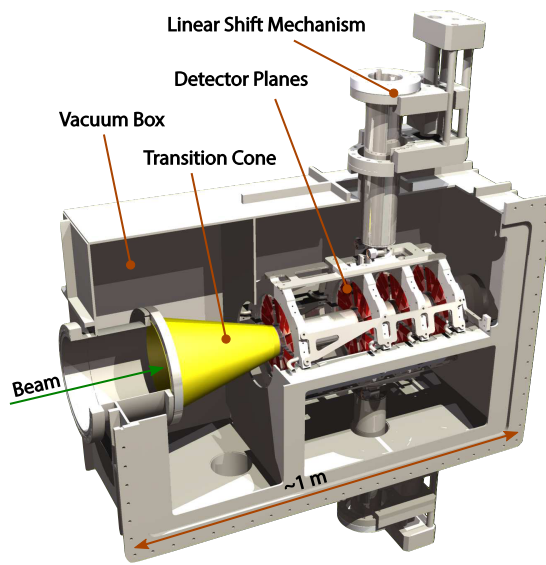
(a) The Luminosity Detector, as it is installed 10.5 m downstream of the interaction point.
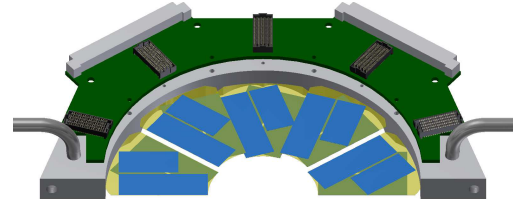
(b) One of the retractable half discs. Sensor modules are shown in blue.
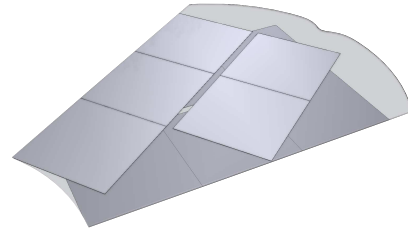
(c) A sensor module, consisting of ten High Voltage Monolythic Active Pixel Sensors. Five are shown in reflective silver, five are installed partially overlapping on its back in a mirrored alignment.

Figure 3.35: CAD drawings of different aspects of the Luminosity Detector [99].

in the Target Spectrometer. Instead of 40 MHz, the FSC uses a different ADC chip to measure 120 MHz.

3.3.6 Luminosity Detector

To measure the luminosity provided at P̄ANDA, a dedicated sub-detector is installed downstream from the interaction point. Located between 10.5 m and 12.5 m downstream of the interaction point, the Luminosity Detector (LMD) measures tracks of elastically scattered antiprotons at very shallow polar angles of 3 mrad to 8 mrad. The relative luminosity is monitored, from which the absolute luminosity can be deduced. The sub-detector is shown in Figure 3.35(a).

Sensors The LMD consists of four instrumented discs of approximately 15 cm diameter. The first disc is located 11.24 m from the interaction point, the succeeding discs 20 cm, 30 cm, and 40 cm away from the first one. For detection of traversing particles, thin, silicon-based pixel sensors with integrated front-end electronic, High Voltage Monolythic Active Pixel Sensors (HV-MAPSs), are used – 100 HV-MAPSs per disc. The HV-MAPSs are produced as 2 cm × 2 cm squares with 50 μm thickness and grouped into 10, partially overlapping pieces to form a *module*, see Figure 3.35(c). Each HV-MAPS is segmented by means of *smart diodes* to form a pixel array. About 247 columns and 242 rows, aligned with a pitch of 80 μm in both directions, create an array with, in total, approximately 60,000 channels. 96 % of the surface of a HV-MAPS is covered with active detector material.

The HV-MAPSs are installed rotationally symmetric onto half discs, discs cut horizontally. The sensors and attached electronics are cooled actively. Since the sensors are located very

close to the beam of HESR, a measure is taken to compensate for bad beam conditions: The half discs are retractable by means of a *linear shift mechanism*. While this induces safety for the detector components, it also imposes challenges onto the flexibility of all connections of the sensors to outside of the LMD. A retractable half disc with five modules and support structure is shown in Figure 3.35(b)

Vacuum Chamber   The LMD is operated in a vacuum to minimize interactions of antiprotons with traversed matter and to separate the device against the circulating antiproton beam. Since the components of the sub-detector are subject to degassing, the vacuum is two order of magnitude worse than HESR's vacuum of $\mathcal{O}\left(10^{-9}\,\text{mbar}\right)$. The LMD's vacuum chamber is sectioned off from the beam pipe vacuum by a transition cone, allowing scattered particles close to the beam to hit the sensors. The cone is located before the first disc and reduces the beam pipe diameter from 200 mm to 80 mm (see Figure 3.35(a)). To introduce as little material as possible, the cone is made of a laminated foil of plastic (boPET), for robustness, and aluminum, for shielding against the antiproton beam. In order not to burst while vacuum pumping, the LMD chamber is operated together with the HESR vacuum system.

3.4   Data Acquisition

P̄ANDA will employ a novel scheme for acquisition and processing of detector data: The experiment will run without a classical hardware trigger mechanism. In this section, the requirements of the experiment are outlined and the chosen system presented, with focus on the online event reconstruction. The description of the Data Acquisition (DAQ) is based on [67].

3.4.1   Requirements

As outlined in Section 3.2, the physics program of P̄ANDA is quite diverse. With its unique position as a high-precision experiment, detecting collisions of cooled antiprotons both in a nearly 4π-fashion and in the forward direction, it will run with a diverse physics program. The different physics channels investigated have different signatures, which are not easy to distinguish from the huge background. Additionally, a high amount of statistics is needed for high precision measurements. For this, P̄ANDA will operate with a mean reaction rate of 20 MHz, in high-luminosity mode. A comparison of reaction rates and event data sizes of P̄ANDA and other particle physics experiments is given in Figure 3.36. While having comparably small event sizes, the reaction rate is very high for P̄ANDA, leading to a large data rate to be processed.

The signature of the signal events is very similar to the background events. Such, that no simple combination of fast sub-detectors exists which could trigger the readout of the remaining detectors. Therefore, P̄ANDA will run without a conventional first level trigger – it will sample the incoming data stream continuously and search for interesting physics events constantly in realtime; instead of a hardware-level trigger, P̄ANDA will run with *online event filtering*.[8]

---

[8]Sometimes, P̄ANDA is referred to as running *triggerless*. This, though, is not really true, as P̄ANDA still employs a triggering mechanism, it is just located much further down the read-out chain compared to traditional particle physics experiments.
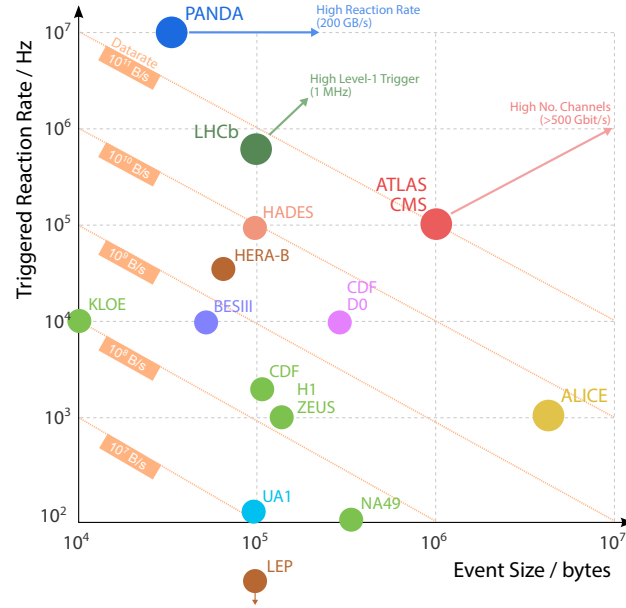
Figure 3.36: Comparison of the read out reaction rates and event sizes of different particle physics experiments. Adapted from [100].

For the »Technical Progress Report for: PANDA (AnitProton Annihilation at Darmstadt) Strong Interaction Studies with Antiprotons« [67], raw data rates of different sub-detectors were estimated. An upper limit of 200 GB/s for raw data from the front-ends will be produced when running at 20 MHz with 15 GeV/c beam momentum, including electronic noise and background particles. This rate needs to be reduced to a rate of $(100-200)$ MB/s (a factor of 1000) by the online event filter to match the available data storage capacity.

3.4.2 Read-Out Scheme

The scheme that $\overline{\text{P}}$ANDA has chosen to match the requirements for online event reconstruction and triggering is shown in Figure 3.37 and outlined in the following. This section focuses on the hardware-structure allocated to DAQ. The algorithmic side of the data flow is explained in Section 3.4.3.

The DAQ relies on a multi-layered, parallel approach, starting with front-end electronics of the sub-detectors. A push infrastructure is employed to transfer data from the sub-detectors to the different stages of the DAQ.

$\overline{\text{P}}$ANDA's chosen scheme offers possibilities to change triggering algorithms on-the-fly, e.g. when new physics channels of interest emerge, or when the type of the physics run is changed. Triggering on defined and comparably complex event signatures is enabled, e.g. triggering on displaced vertices.
The scheme is modular: Every stage of the read-out chain can be complemented by additional devices. A test-bed system has the same principle structure as the eventual fully-built read-out for $\overline{\text{P}}$ANDA. Computing power can be re-allocated for special physics runs searching for
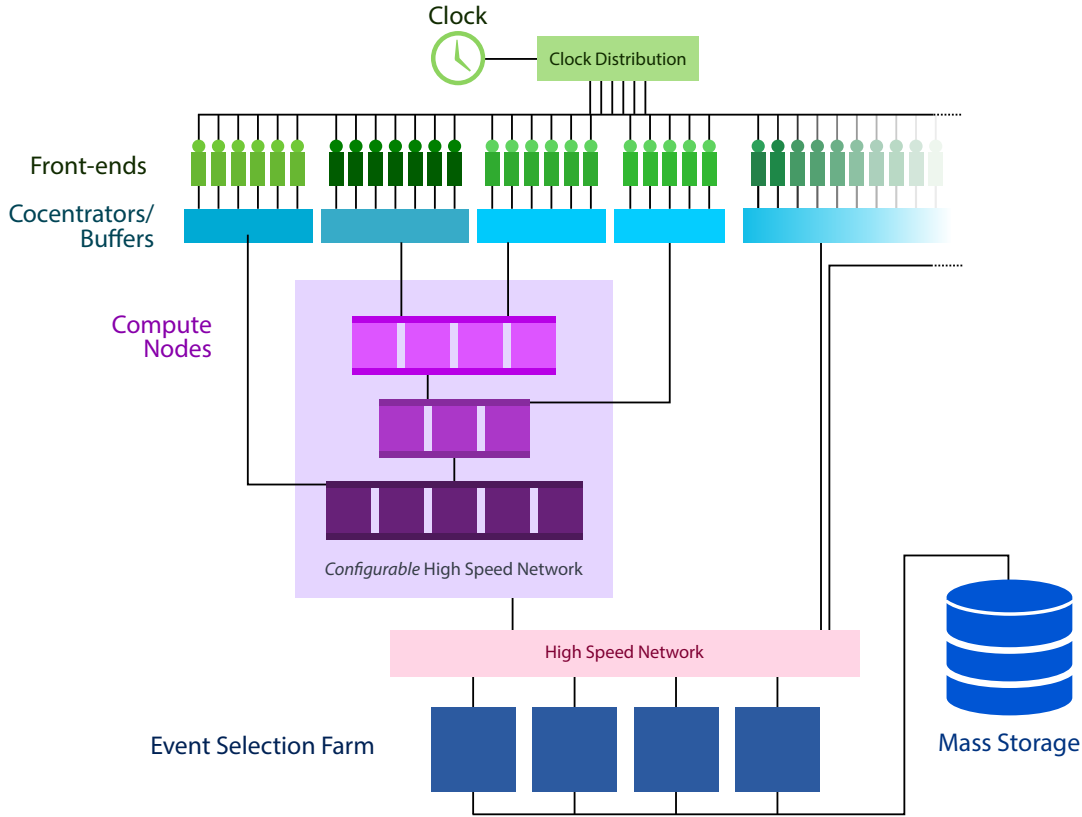
Figure 3.37: P̄ANDA's DAQ scheme, divided into the different devices responsible for data processing. The scheme is based on [67].

computational intensive event topologies. Scaling the system for runs with increased luminosity is possible by assigning new computing devices.

If feasible, the system relies on hardware produced for the mass-market – keeping the cost low.

Clock Signal   Since all data taking equipment (i.e. front-end chips) runs continuously and is its own collecting entity, it needs to be running in-sync with all other equipment. A global clock is introduced, giving a timestamp to every device participating in data collection and processing. The clock system uses the Synchronization Of Data Acquisition (SODA) architecture. A central clock generates a reference time which is then distributed by means of passive, optical network infrastructure to the individual devices. The time precision is <20 ps (RMS) [101]. The clock signal is distributed among many different devices: Front-end chips reading out individual sub-detectors in a radiative surrounding as well as data-reconstructing devices further away from the experiment. Since a conventional event structure is, a-priori, missing in the P̄ANDA data stream, the clock signal is used to associated time slices of different sub-detectors with each other.

Detector Front-Ends   The front-ends in each sub-detector are responsible to amplify the signal and convert it to digital data. In general, a first level of data pre-processing happens already here: In order to find valid hits, noise reduction and zero suppression is applied. Depending on
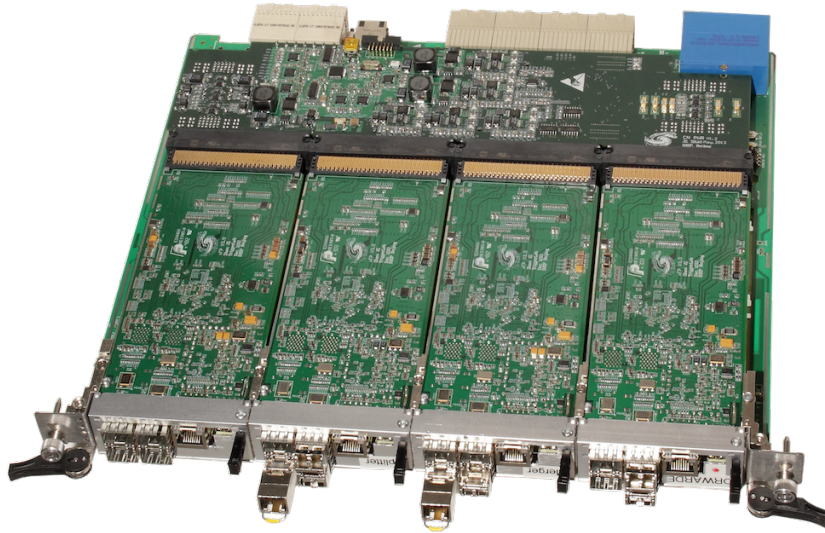
Figure 3.38: A prototype of the compute node to be used in P̄ANDA's DAQ. The enclosing hosts the motherboard part of the node, supplying infrastructure for the four µTCA daughterboards [102].

the sub-detector, even further simple processing can be done – patterns, e.g. clusters, can be found and extracted into the data stream. Front-ends reduce the amount of data to a maximum of <40 GB/s.

Data Concentrators and Buffers   Data from the front-ends of the same sub-detector is collected in data concentrators. The concentrators include memory for buffering data, since not all front-ends deliver data concurrently or are equally fast. As the concentrators have computing power as well, they are used for grouping data according to their time information. In addition, some tasks of more global, but still comparably easy pattern recognition can be performed here. Using a switchable high-speed network, data concentrators transmit data to the compute nodes – one individual time slice is sent to one node.

Compute Nodes   Compute nodes offer a large amount of dynamically allocated computing power to detect patterns and simple physical signatures. They operate on sub-sets of the global data set, determine and sort the incoming time slices. Among others, they can combine EMC clusters, compute particle tracks, or match both against each other; they can compute time-of-flight information; they identify particle types (see also Section 3.4.3.1).
Data processing in compute nodes is structured in levels. Fast algorithms, working potentially only on a small sub-set of data, are succeeded by more computing intensive ones. The computational complexity of employed algorithms can be increased, while the data is reduced at the same time. All in all, a reduction of $\mathcal{O}(10)$ is expected on compute nodes.

The current design of a compute node for P̄ANDA, shown in Figure 3.38, is structured into two parts. The *motherboard* part creates the infrastructure for four pluggable *daughterboards*. It is equipped with a Field-Programmable Gate Array (FPGA), controlling the actual computations on the plug-in boards. High-speed data transfer between the boards is provided by means

of *Rocket IO*. The daughterboards are compliant to the xTCA[9] / μTCA[10] standard. They use Virtex 5 FPGAs as central processing units and hold 4 GB RAM. They can be equipped with up to four Ethernet interfaces.

Computing Farm   The actual selection of an event for possible storage is the last step of the DAQ chain. Pre-assembled events with some of their properties already reconstructed enter the selection. There, (nearly) the complete event topology is computed, including quantities such as momenta, invariant masses, vertices.

An event is selected, if it is among the events of interest for the current physics run. This can vary, as some runs require more restrictive selection criteria, and others choose a less restrictive trigger. The software trigger eventually makes the decision when comparing the reconstructed event against a list of events of current interest – see Section 3.4.3.2. A final data reduction of $\mathcal{O}(10)$ is expected to match the available resources of a few PB/yr.

The event selection is located on a large farm of processors, connected to the compute nodes with a high-speed network. The architecture of the farm must still be designed beyond its principle drafting, but should be built largely from off-the-shelf hardware.

3.4.3   Online Trigger System

The online trigger in P̄ANDA is multi-layered, as outlined in the last section. A very first pre-selection happens on the front-ends directly. Data is further sorted when aggregated and combined time-wise. Continuing, tracks of particles are coarsely reconstructed and particles preliminary identified. Finally, the software trigger decides whether or not to keep an event.

The individual parts of the online processing scheme are shown in Figure 3.39 and described in this section. The design of the individual parts is not yet fully determined, thus the preliminary status is presented here. First, properties of the particles and the event are reconstructed in the different steps of the *online reconstruction*. Then, the data is employed by the *software trigger* to discriminate signal and background.

3.4.3.1   Online Reconstruction

In the *online reconstruction*, the process of aggregating, sorting, and processing data is described. Data is reconstructed to specify the type of the event for the succeeding software trigger (Section 3.4.3.2).

Data Collection   Hit points are generated when analog information is converted to digital data on the sub-detector front-ends. Noise is suppressed and a first pre-processing could be done: hit clusters are formed from adjacent strips in the MVD, energy clusters calculated in the EMC, et cetera. Data is time-stamped according to the global SODA clock.

---

[9]Extended Telecommunications Computing Architecture
[10]Micro Telecommunications Computing Architecture
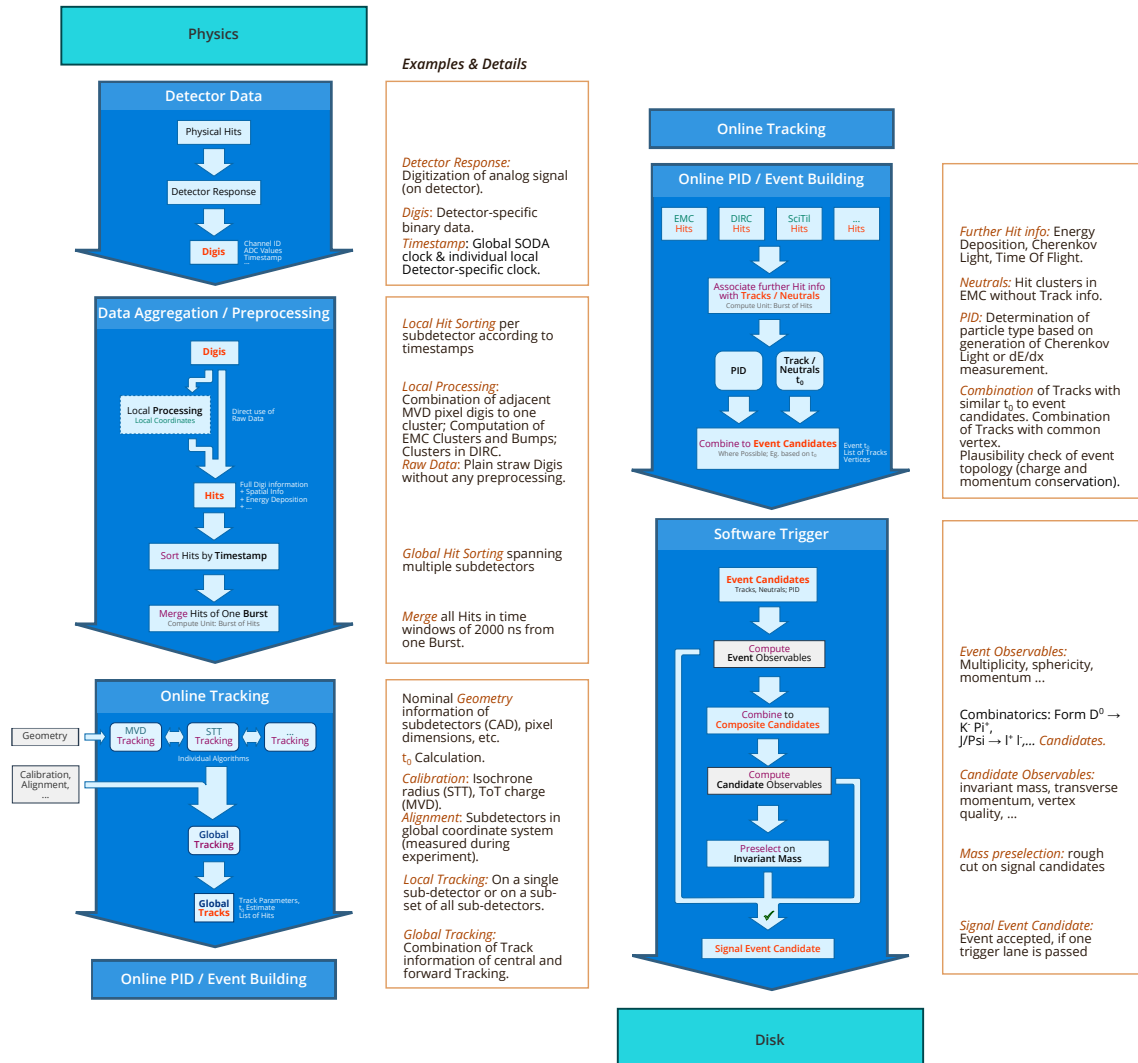
Figure 3.39: P̄ANDA's DAQ scheme with the individual event reconstruction steps outlined. The sorting into the different stages does not represent physical availability of the data – the hardware borders are not shown. Some of the processes happen in parallel, some data cross-transfers are omitted.

Data Aggregation   In the next step, data of all sub-detectors is sorted time-wise. The HESR beam structure can be helpful here: In the current running scheme, HESR will follow a 1600 ns beam spill (a *burst*) with a 400 ns gap. The period without interactions can be used to give an upper limit for the intertwined hit data. If possible, a first event building can be applied on the now time-sorted data to give first estimates of event candidates. A burst can be a natural choice of an undivided work data package.

Online Tracking   A particle's track needs to be reconstructed in order to determine a particle's path through the detector, its momentum and its charge. *Tracking* is first done individually on a local basis, before global tracking combines the tracklets to tracks.

   Local Tracking   The first stage of track reconstruction considers only a sub-set the all sub-detectors. Depending on the algorithm, this can be a single sub-detector, or a combination of sub-detectors. Primarily, tracking detectors (MVD, STT, GEM, FTS) are considered in this step, although information from additional sub-detectors might be useful (e.g. precise timing information from the SciTil).

   At this stage geometry information of the sub-detectors and other input is needed. The exact location and measurement uncertainties of each MVD pixel/strip need to be provided, et cetera. The data is provided by means of calibration tables located on the machines running the tracking algorithms. For the STT, the calibration curve of drift time and isochrone radius (Section 3.3.2.2) has to be present.

   The position of each participating sub-detector entity needs to be known from alignment. Both for the individual elements of a sub-detector (drift tube positions, MVD strip sensor locations) and for the position of whole sub-detectors in relation to one another. In general, it is possible that specialized algorithms can track particles through a sub-detector without being embedded into a coordinate system. The location information is then needed in the next step, when local tracks are related to each other and global tracks are formed.

   Track reconstruction, usually a comparably slow and iterative processing step, needs to be done swiftly when done in an online reconstruction environment. Specialized, heterogeneous algorithms need to be developed. A layered algorithm setup can divide the complicated tracking problem into fast-solvable parts, lowering the data complexity along the way. The continuous, only time-wise sorted flow of data, is a challenge special to $\overline{\text{P}}$ANDA's (at this stage) untriggered data taking approach. To minimize the combinatorial effects, a first, coarse algorithm could reconstruct tracks originating from the primary vertex. Subsequent algorithms then have a lower total amount of hits to consider to find more complex event topologies.

   As online track reconstruction is a main topic of this thesis, it is explained in more detail in Section 4.

   Global Tracking   Global tracking is the second part of online tracking. Tracks generated by different algorithms, considering different (sub-sets of) sub-detectors, are combined to common track candidates. *Track merging* tries to combine multiply found tracks of the same particle.

Beyond the individual track information, the tracking algorithms can also supply first estimates of further, secondary parameters. For example, a STT algorithm that does not need an event start time for track reconstruction could in turn produce a limit on that value, possibly lowering the run time of dedicated algorithms.

To the next stage of online reconstruction, online Particle Identification (PID), track parameters, associations to hits (including their timestamps), and possible secondary information is passed on.

Online PID The last step before the actual software trigger is online Particle Identification (PID) and event building. The tracks computed in online tracking are combined with further information from other sub-detectors: e.g. EMC clusters can be associated, DIRC signatures matched, SciTil timing information included. For uncharged particles, leaving no track points in the tracking detectors, hypotheses are formed from EMC information. For more on $\overline{\text{P}}$ANDA's general approach on PID, see Section 3.5.3.4.

The type of particle is determined by means of measurements of energy loss ($\mathrm{d}E/\mathrm{d}x$), the structure of the Cherenkov light in the DIRC / RICH detectors, by different energy cluster properties in the EMC, or by the flight time in the TOF detectors.

Particles belonging to a common event are associated, *building* the event. Information about the time of the occurrence of the event is extracted and associated – from the stream of time-wise sorted data, an event-based data stream is formulated. Also, a first, rough plausibility check can be applied on the resulting data. Only coarse discriminations are applicable here to not disregard any signal data.

Data is sent to the software trigger, responsible for the final decision to transfer data to storage (*disk*).

3.4.3.2 Software Trigger

The software trigger is the last element of the DAQ and located on high-performance computing farms. It tags events to disregard or to store on the mass storage system – available for storage is a few PB/yr. Important input parameters to the software trigger are a charged particle's track (momentum, vertices) from online tracking, information about clusters of neutral particles from the event building, and the type of the particle as of PID. Using this information, the whole event topology can be assembled.

With the help of $\overline{\text{P}}$ANDA's offline analysis framework PandaRoot, the performance of the software trigger has recently been evaluated for a first time in »Present Status of the PANDA Software Trigger« [100].

As online-running algorithms are still missing in the experiment's computing framework, the report analyzes a set of possible trigger channels with the available offline reconstruction algorithms. The results, hence, are an upper limit to the realistically achievable triggering efficiencies under the constraints of the currently available PID algorithms. Also, the report uses event-based data, as not all sub-detectors have yet included a realistic timely-sorted data

flow model into PandaRoot. Artifacts and inefficiencies introduced when building events from the time-sliced detector data will eventually also reduce the triggering capabilities.

Currently, a preliminary list of 25 interesting physics channel has been defined. Of this, 10 are evaluated in detail – electromagnetic and exotic channels, charmonium and open charm decays, Λ baryons. Background data is generated using the DPM model (see Section 3.5.3.1). 10 trigger lanes are specified depending on the (inclusive) final state signature of one of the reaction particles, e.g. $J/\psi \to \mu^- \mu^+$ triggers the $\overline{\text{p}}\text{p} \to J/\psi\,\pi^+\pi^-$ decay channel.

Variables used for discrimination in the trigger lanes include momentum-based and energy-based quantities, angles, missing mass, PID probabilities, particle multiplicities, shape of the event. Also, the center-of-mass energy, $\sqrt{s}$, influences the cut parameters, as the available energy in the system strongly changes its topology and the shape of the discriminated quantities. Hence, trigger lanes have to be defined for every channel of interest at every energy studied. If an input event passes any of the trigger lanes, it is considered an *event candidate* and tagged for storage.

Trigger efficiencies for the 10 benchmark channels range from approximately 50 % to 6 %, depending on the evaluated channel and energy[11]. The efficiencies are higher when the targeted suppression factor of 1/1000 can be loosened. Using an optimization for high efficiency with suppression of 1/100, these channels can both be tagged with efficiencies of 60 %, a vast increase in particular for the η_c.

The study of the software trigger emphasizes that well-reconstructing algorithms are of importance also in the online reconstruction part of the experiment. Effective trigger decisions can be made only with precisely reconstructed events. Adding the need for fast and low-latency reconstruction makes $\overline{\text{P}}$ANDA's online event reconstruction a very challenging task.

3.5   Software Tools

To evaluate and tune the detector performance, $\overline{\text{P}}$ANDA employs a large, integrated software frame work called PandaRoot. Currently used for simulations and detector assembly, PandaRoot is going to be utilized also for the experiment's run.

In this section, the most important features and components of PandaRoot are highlighted. First, the tools on which PandaRoot is built upon are outlined. Then, important physics programs are presented, which are used during the course of a typical simulation data run.

3.5.1   ROOT

On its lowest level PandaRoot builds upon ROOT, CERN's data analysis framework. ROOT was first presented in 1997 by BRUN and RADEMAKERS [103] and has been significantly extended since. ROOT offers infrastructure for processing, displaying, and storing physics related data. It is written in C/C++ and offers the ability to run programs (*macros*) in an interpreted, uncompiled mode, enabling swift implementations of data analyses.

---

[11]$\overline{\text{p}}\text{p} \to J/\psi\,\pi^+\pi^-$ is evaluated at $\sqrt{s} = 4.5\,\text{GeV}$ and $\overline{\text{p}}\text{p} \to \eta_c\,\pi^+\pi^-$, which is triggered by $\eta_c \to K^0_S K^- \pi^+$, at $\sqrt{s} = 5.5\,\text{GeV}$.
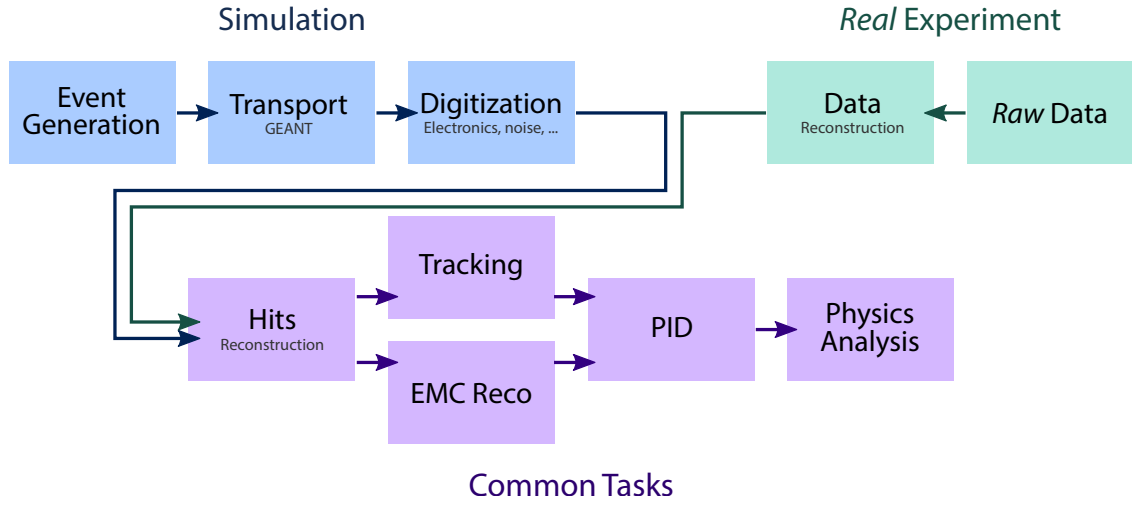
Figure 3.40: Different, important stages of dataflow in PandaRoot. The framework can be used both for working with simulation data as well as with actual experimental data.

3.5.2 FairRoot

To extend ROOT for the needs of the FAIR experiments, a common framework was developed: FairRoot [104]. The framework introduces a task-based structure: Tasks can be registered to a steering `FairRun` class, which are then called in the correct order at the correct time. Examples include complex tasks for event generation, track propagation, track reconstruction, but also more specific tasks for extraction of single event parameters, like track momenta. FairRoot offers interfaces to different event generators and geometrical description tools. It stores run parameters in a `RunTimeDataBase` and offers storage support for databases. It enables matching of reconstructed and original generated data (`FairLinks`) and has built-in functionality for time-wise event mixing, as needed for P̄ANDA's read-out scheme.

A recent addition to FairRoot introduces message-based data transport, called FairMQ. The additional abstraction layer of FairMQ enables a more versatile data flow through the tasks of FairRoot, both locally on one machine as well as through a network. Offloading a single computation to another network-connected computer is made possible seamlessly with FairMQ [105].

3.5.3 PandaRoot

PandaRoot is the specialization of FairRoot for the P̄ANDA experiment. The complete detector geometry, as presented in Section 3.3, is rebuilt digitally using the appropriate materials and dimensions. The beam pipes of HESR and maps of the magnetic fields are included as well. The P̄ANDA-specific digitization and reconstruction algorithms are also incorporated into the framework.
ROOT, FairRoot, and PandaRoot are programmed in C/C⁺⁺, with certain parts included in Fortran.

The software framework comprises the whole software in use for detector evaluation and

physics performance analysis in P̄ANDA. Specific reconstruction code is developed, as well as externally used packages embedded. PandaRoot is intended to be used for simulated data as well as for real physics data when P̄ANDA starts its experimental running. Important steps in the chain of reconstructing physical events are outlined in Figure 3.40. *Event Generation* creates physical events in accordance with physical models. *Transport* propagates the produced events through the virtual detector. In *Digitization*, the detector's response is simulated, as it would be for the *Real Experiment*. Subsequently, tracks are produced from hit points, energy clusters of neutral particles found, and the corresponding particles identified. Finally, the physics event is analyzed [106].

The following sections go through the steps of a typical PandaRoot event simulation and reconstruction chain, highlighting important points.

3.5.3.1   Event Generators

Event generators produce reactions in beam-target and beam-beam interactions in accordance to measured or theoretically predicted models. They consider the masses of particles, momenta and energies, quantum numbers, decay lengths and decay widths. The simulated particles are produced statistically distributed, filling the phase-space of the simulation due to their respective model. The events generated are *primary particles* and input for the next stage of particle propagation.

As different event generators specialize on different physical models, different generators are included in PandaRoot. They are included by means of interfaces and packaged with PandaRoot usually due to their Open Source license.

EvtGen   EvtGen was developed in the late 1990s to simulate events of B decays, targeting physics of the just-started B factories (BaBar, Belle, ...) [107]. Since then, EvtGen gained popularity also among other experiments. The initial development is continued as part of the *EvtGen Project* [108].

One of the main features of EvtGen is that for a generation of an event only the amplitudes of the simulated decays need to be provided. The resonant sub-structures and angular correlations between simulated particles are computed automatically. Probabilities and spin density matrices are calculated internally in EvtGen. The generator has a modular structure, is expandable, and can cover arbitrary decay chains. It has a large range of built-in decay models, for example Dalitz decays involving three daughter particles, but can also be expanded with customized decay models. In a decay file, the user can specify the type of the decay, the constituting particles, the decay widths and model. Usually EvtGen is used to simulate single targeted signal channels. EvtGen is the event generator employed in the event reconstruction of this thesis in Section 5.4.

DPM   The Dual Parton Model (DPM), generally, is used to describe high-energy collisions between two hadrons, two nuclei, or an hadron and a nucleus [109]. The model is phenomenologically motivated and can cover non-perturbative (*soft*) processes as well as semi-perturbative (semi-*hard*) processes at higher energies [110, 111].
In P̄ANDA, DPM is used to describe inelastic and elastic hadronic processes as well as Coulomb

elastic scattering processes. For elastically scattered $\bar{\text{p}}$p events, available data of different momenta has been analyzed specially for $\overline{\text{P}}$ANDA. A parametric description was found for the momentum range of $2\,\text{GeV}/c$ to $16\,\text{GeV}/c$ [112]. The DPM generator is usually used to describe background-like events in $\overline{\text{P}}$ANDA.

Other Also available in PandaRoot are **Pythia**, a multi-purpose event generator for high-energy collisions (e.g. LHC) [113] and **UrQMD**, a model for (ultra)relativistic collisions of heavy ions [114]. In addition, a `BoxGenerator` enables precise generation of single particles or a set of particles. The particles are shot into a specific solid angle of the detector with a defined momentum range. The `BoxGenerator` is used for simulated benchmarking of detector components and evaluation of algorithms.

3.5.3.2 Propagation

Once particles and their decays are simulated, they are transported through the detector. Two simulation frameworks that describe the transport are available, the Fortran-written GEANT3 and GEANT4, written in C++. Both have the same functionality, but they differ in performance and results as a result of their respective underlying models. A common interface to both GEANT versions is given by the Virtual Monte Carlo (VMC) [115]. The specific commands are replaced by more abstract ones of VMC which propagates them towards the user-chosen version. The abstraction allows for easy performance comparison of the two versions.

The track propagation works on each simulated particle. The propagating tool transports each particle through the programmed geometry step-wise[12], evaluating different actions along the step and after the step. At every step, the governing physics processes, according to the external material and magnetic field, are statistically computed; energy loss is calculated, tracks deviated due to scattering, particles radiated (bremsstrahlung, Cherenkov), or daughter particles created from particle decays. The tracks of the particles are adapted to the new circumstances and the step-wise propagation is continued until the border of the defined volume is reached. If active detector material is hit, position, time, and deposited energy are propagated to the next step of the simulation chain (*digitization*) [116].

3.5.3.3 Digitization

The last step of the simulation-only chain of PandaRoot is digitization. In this step the detector response to the hits created through the event propagation is simulated. The goal is to be as close to data from real hits coming from an actual experiment. Digitization is unique for the different detecting entities of the sub-detectors. It covers, e.g., artificial generation and suppression of electronic noise, signal thresholds, or inefficiencies.

Through the different steps of generation, propagation, and digitization of simulated data, the connection to the original created, *primary* particle is kept intact by means of `FairLinks`. With this, matching can be done, connecting reconstructed data with event generator input data.

---

[12]In order for precise particle propagation, exact detector geometry description is inevitable. Both active and passive detector material (e.g. measuring sensors and holding structures, respectively) have to be included with their individual materials.

Links to generator data are used for benchmarking the developed algorithms of the subsequent event reconstruction and analysis.

3.5.3.4 Reconstruction

The reconstruction step is the first common stage for simulated and actual data from P̄ANDA. This stage reconstructs physical particles from the digitized detector responses to eventually form a complete event. After converting the individual hits of each sub-detector to a common basis (x, y, z coordinates, associated time and deposited energy), the trajectories of the particles are computed. After this, the trajectories are associated to particle types in the PID, recreating the full information of the particle.

Track Reconstruction   Tracks are reconstructed using hits from the individual sub-detectors. Usually, the four tracking sub-detectors take part in track reconstruction (*tracking*): MVD, STT, GEM, FTS.

Tracking is a multi-layered process, as different algorithms have different input parameters and feature sets. Generally, tracking can be separated into two tasks: *Track finding* and *track fitting*. Track finding searches for hits created from one common particle. Track fitting creates the best-fitting track through a given set of hit points. In most cases, a track finding algorithm is followed by a track fitting algorithm. Some algorithms combine the two stages into one, and can deliver pre-fitted tracks from a large number of input hits.

P̄ANDA's default track reconstruction starts with a track finder using STT hits. A road-finding procedure is employed to link individual straw hits to a connected set of hits. These hits are used in a Hough transform (HT) (see also Section 4) to find a first track hypothesis as a circle in the xy plane. Using this hypothesis, hits in the STT and the MVD are associated and a new, χ^2-based track candidate is fitted in the two-dimensional space. To expand the track from a circle representation in the xy-plane to a helix in the xyz-space, the skewed STT hits are associated and iteratively tested for compatibility. Finally, piled-up tracks are cleaned and the final track candidate is formulated. A second stage associates GEM hits to the tracks. The assumption in this stage is that the magnetic field is constant, leading to helical trajectories of charged particles. Tracks in the FTS are currently being found ideally, using information from the event generator to associate hits to tracks. A realistic *forward* track finder, using HTs, is currently under development and soon to be included in PandaRoot [85].

The last step in track reconstruction is track fitting. Using the package Genfit, the best-fitting track is formed going through the hits from the previous track finding stage [117]. Genfit was originally developed inside the P̄ANDA collaboration. Internally, Genfit uses a Kalman filter[13] for precise track fitting, considering detector material, geometry effects, and also including the magnetic field. A Kalman filter is a widely used tool to estimate systematics underlying error-prone measurement points [118]. In particle physics, a Kalman filter is used for track reconstruction, as its iterative approach can consider scattering through matter and energy loss [119]. The Kalman filter incorporates the magnetic field map and considers the uncertainties of the different sub-detector measurements, usually yielding very precise track results.

---

[13]Genfit also provides an alternative to the Kalman filter: A deterministic annealing filter.

PID   In Particle Identification (PID) the particle species is determined.

$\overline{\text{P}}$ANDA incorporates different concepts to specify the type of the particle. For example the MVD and STT use measurements of specific energy deposition, $\mathrm{d}E/\mathrm{d}x$, and the DIRC detectors use the Cherenkov angle, θ_c[14]. The different sub-detectors can perform effective PID for a limited range of momentum. Each sub-detector targets different momenta. In the first stage, PID is performed on a per-sub-detector basis.

A charged particle is tested to determine how likely it is one of the following particles hypotheses: e, μ, π, K, p. These are the particles leaving most of the hit points in $\overline{\text{P}}$ANDA[15]. For every sub-detector, a particle is appointed a *PID number*, a set of probabilities for it to be one of the five particle hypotheses. Using a Bayesian approach and assuming that the probabilities do not influence each other, the individual PID probabilities can be combined to form an overall PID probability. The approach is fast but needs a clear separation between different particle types. For more complicated events, multivariate analyses are under consideration, covering in particular cases of the EMC.

Energy clusters in the EMC, which are not be associated to a charged track from the other sub-detectors, are tested for neutral particle hypotheses.

3.5.3.5   Analysis

By means of track reconstruction and PID, tracks are found and particles identified. PandaRoot also offers functionality to analyze the particles and events reconstructed in the previous chain on a high level. The analysis is enabled by means of the Rho package [120]. Rho builds up on standard, fast ROOT interfaces. It is included into PandaRoot and extended by specific and necessary functionalities.

Using the PID probability as computed before, Rho offers methods for combining the particles to composite particles, e.g. K^+K^- to φ, if the corresponding four-vectors are compatible. The package takes care of double counting due to combinatorics. Particles are structured in candidate lists, RhoCandList, to which selection criteria can be applied (e.g. momentum spread). Different fitting methods are also included into Rho. Notably is a vertex fitter and a mass fitter. The vertex fitter varies the measures track parameters of daughter particles within their measurement uncertainties so that thy all originate from one common vertex; the mass fitter fixes the mass of a composite candidate to the PDG value by varying the other kinematic variables accordingly [121].

Rho can apply selection criteria and run fitters on combined candidates sequentially and store the output in ROOT's TTree structure, enabling swift and interactive evaluation of the analyzed events. This is especially helpful if a computing grid or cluster is used to simulate, reconstruct, and analyze large data sets.

---

[14]The momentum of a given particle, p, can be determined from the radius of the curvature of its track in a magnetic field (Lorentz force). To fully identify the particle and its four momentum, a second observable is needed. The second observable can either be directly E from a calorimeter fully stopping a particle in it, or the relative velocity βγ, since $p = \gamma m_0 \beta c$. Access to βγ gives the energy loss through ionization, $\mathrm{d}E/\mathrm{d}x \propto \ln(a\beta\gamma)/\beta^2$, or the Cherenkov angle, $\cos\theta_c = 1/(\beta n)$. Also TOF detectors can give insight into β, as $\tau \propto 1/\beta$.

[15]Concerning $\overline{\text{P}}$ANDA, these particles are *quasi*-stable.

Chapter 4

GPU-based Online Track Reconstruction

$\overline{\text{P}}$ANDA's novel read-out concept is based on reconstructing the stream of events in real-time in order to suppress background in an online software trigger. The concept is outlined in detail in Section 3.4.3. Track reconstruction in the online event filter is an essential task and computationally intensive. Hits measured with a rate of $\mathcal{O}\big(\text{Ghit/s}\big)$ need to be assembled swiftly into tracks.

Graphics Processing Units (GPUs) are throughput-optimized computing devices which offer a large number of processors, able to compute with a high rate. Algorithms exploiting the parallel architecture of the device can benefit from the high performance and energy efficiency. Using GPUs for reconstruction of tracks in real-time is a promising approach to handle the immense stream of data of $\overline{\text{P}}$ANDA.

This chapter introduces the parallel computing architecture of GPUs in Section 4.1. In Section 4.2 the basis of online track reconstruction is presented. Three track reconstruction algorithms running on GPUs are shown subsequently; the Line Hough Transform in Section 4.3; the Circle Hough Transform in Section 4.4; the Triplet Finder in Section 4.5. Finally, in Section 4.6, the results are summarized.

4.1 General Purpose Graphics Processing Unit

Graphics Processing Units (GPUs), electrical circuits built into computers, were originally used to render imagery onto a screen. Data was transfered from the Central Processing Unit (CPU) to the GPU, operations as instructed by the CPU were performed and the result was output to a monitor. To match the needs of applications demanding more GPU-side computing power for forming increasingly realistic looking images, GPUs evolved to highly specialized computing devices. The number of shaders[1] grew to operate on ever larger image data with higher clock rates, enabled partly by Single Instruction, Multiple Data (SIMD)[2] execution patterns. With

---

[1]Shaders are used on GPUs to compute realistic lighting effects on computer imagery.

[2]SIMD refers to a pattern exploited when operating on large sets of similar data. A single instruction is used to manipulate not one single data entry, but a whole set of data at once. The instruction takes one or a few clock cycles.

the introduction of programmable shaders and floating point operations to GPUs, the devices quickly gained popularity among a more general community. Arbitrary computational tasks were translated into *GPU language*, making use of the limited instruction set offered for data manipulation on graphics cards. Channels to transfer data back to the CPU were introduced. The widely used APIs OpenGL and DirectX, which offered the original video card access patterns, were complemented by more general APIs focusing on exploiting the GPU's raw computing power to unspecialized computing tasks in the 2000s (e.g. BrookGPU [122]). Soon, higher evolved APIs were introduced, giving access to the whole GPU device in a programmable way without making use of the original GPU primitives. Two prominent programming platforms are available, the proprietary NVIDIA CUDA, introduced in 2007 for NVIDIA cards, and the vendor-agnostic OpenCL by the Khronos Group, introduced in 2009 [123, 124]. Both platforms offer programming of GPU-running applications by extensions to the C/C++ programming languages and APIs to directly communicate to the GPU device [125–129].

The use of GPUs for a general computing is called GPGPU[3] computing. During recent years, GPGPU computing gained large popularity, as it offers processing power with lower cost compared to large supercomputing farms. The GPU-provided performance gain is achieved if the computed problem can be parallelized and calculation operations can be performed on a large amount of data points at the same time. Hence, the grade of achievable parallelism depends on the specific problem. Ideally, it should be a computing-intensive task, operating on independent data points of similar structure.

For this thesis, the programming platform offered by NVIDIA, CUDA, is used. No definitively favorable solution emerges for GPU programming when comparing CUDA to OpenCL [130–132]. Both platforms have their respective benefits and drawbacks. For parallel programming, it is a matter of choosing the language better suited for the specific problem, the infrastructure, and the software support. While the employed programming scheme and language constraints the parallel code to devices of certain vendors, a large part of the effort of GPU programming is in stating the problem in an effective parallel way and transposing it from a serial, CPU-like description. Only applications optimally occupying the GPU will gain a significant speed-up from a change to a parallel environment. Once the port of the code is done, a change of programming platforms is largely a conversion effort, respecting device and platform specific optimization possibilities.

In this section, the GPU platform of NVIDIA GPUs is presented and the CUDA software infrastructure shown. If not specified otherwise, the information given in this section is based on »CUDA C Programming Guide« [133] and »CUDA by Example: An Introduction to General-Purpose GPU Programming« [129].

4.1.1 Hardware Platform

The adoption of GPUs for multi-purpose high-performance computing builds on a simple principle: Offering a large amount of processing cores, capable of executing computations in a massively parallel way. This is in stark contrast to the serial, or limited-parallel, way that CPU processes are executed traditionally. Figure 4.1 schematically sketches typical chip layouts of a

---

[3]GPGPU stands for General Purpose Graphics Processing Unit.

(a) CPU Structure.   **(b)** GPU structure.
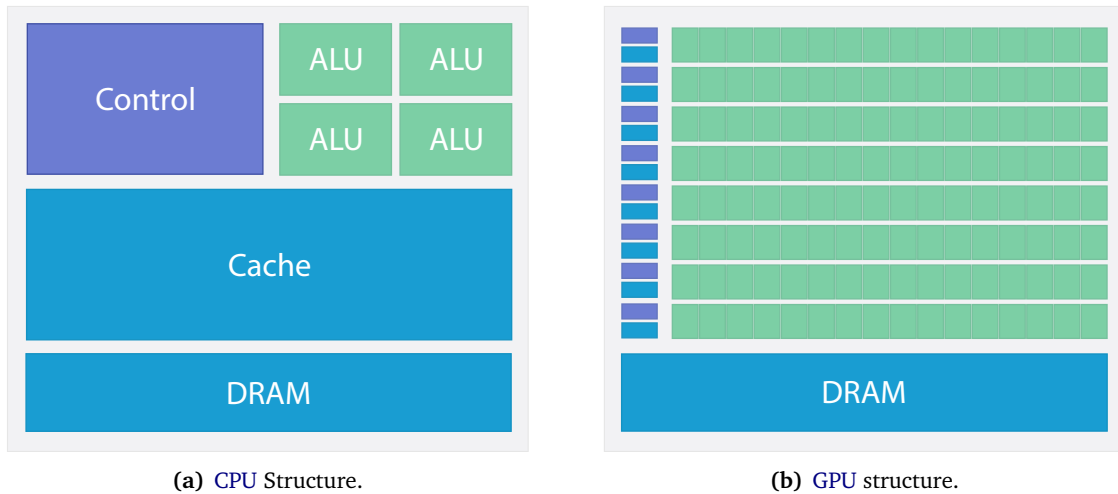
Figure 4.1: Structure of CPU and GPU integrated circuits. Outlined in blue are the structures available for memory, purple shows control. Green denotes Arithmetic Logic Units, on which the actual computation happens.

CPU and a GPU. A CPU is designed as a versatile processor, being equally performant for a vast spectrum of different types of computing requirements. To reduce latency, CPUs use complex, sophisticated instruction sets and a large amount of cache close the computation cores. The Arithmetic Logic Units (ALUs), on which the actual computations are performed, have a large amount of transistors and are clocked with high frequencies (typical CPU frequencies range from 2 GHz to 3 GHz). GPUs employ a different strategy for the chip. Instead of focusing on effective data handling with big control structures, they provide a large amount of bulk ALUs. The number of transistors used for caching and data flow control is greatly reduced, the freed space taken by ALUs. Instead of processing data with a low latency, GPUs are designed for processing data with high throughput. The processors operate with the same instructions on different data elements. Hence, GPUs are especially suited for computing-intensive tasks, where latency coming from memory access can be hidden by executing a large number of processes in parallel.

GPUs are connected to the CPU typically via the PCI-Express bus of the motherboard. Applications are launched to the GPU by software drivers, copying the needed data-set, launching computations, and retrieving computed data. Data typically rests in the GPU memory, GDDR RAM, with sizes from 1 GB to 10 GB, or even higher, depending on the device. Data can reside on different types of memory on the GPU, all different in speed and access pattern.[4] Different types of memory are cached differently. GPUs offer L1 cache, accessible by threads from a

---

[4]The definition, partitioning, and level of GPU memory changes with chip architectures. Generally, NVIDIA divides the different available levels of memory into: *global memory* (cached, main memory, accessible from anywhere, including the host), *constant* and *texture memory* (cached, read-only memory, accessible from anywhere), *local memory* (cached, abstracted global memory, dedicated to one thread), *shared memory* (fast on-chip memory, visible by all threads of one block), and *register memory* (on-chip memory dedicated to one thread). In *Maxwell* cards, the L1 and the texture cache are fused into one common cache, but shared memory resides in a dedicated part of the on-chip memory.

Figure 4.2: Sketch of the multiprocessor and memory distribution on a GPU. Cache is colored in blue, a Streaming Multiprocessor is colored in purple, with the individual cores sketched in green.

single multiprocessor, and L2 cache, shared among different multiprocessors. In modern GPUs, L1 cache is of the size $\mathcal{O}(\text{kB})$, L2 cache of $\mathcal{O}(\text{MB})$ size.

NVIDIA GPUs are categorized two-fold: The microarchitecture of the chip and a further defining version. The architecture is referred to by code names, from oldest to newest: *Tesla*, *Fermi*, *Kepler*, and *Maxwell*. The additional version is called *compute capability* and is 1 (Tesla), 2 (Fermi), 3 (Kepler), and 5 (Maxwell), with sub-versions (e.g. 5.2) referring to different feature sets of devices, independent from an architecture change.

The building block of GPU devices is the Streaming Multiprocessor (SMM)[5], a combination of processing cores (CUDA cores). Larger devices include more SMMs to provide more computing power. An exemplary scheme of the distribution of SMMs is shown in Figure 4.2. CUDA cores are belonging to one SMM, which offers the local memory access patterns. Cores of one SMM have access to L1/texture cache for caches from global and constant memory and to shared memory, to swiftly distribute data among the threads.

The SMM is responsible for creation, management, scheduling, and execution of *warps* (groups of *threads*). Threads are executed in sets of 32. Data is operated in a Single Instruction, Multiple Threads (SIMT) scheme, executing a computation on a set of threads, parallelized during runtime[6]. The collection of 32 threads executing the same computing instruction is called a warp. A warp is the only set of threads, which is guaranteed to be executed at the same time on a GPU. Warps are setup for execution by warp schedulers, which distribute the threads into cores. Sets of threads (*blocks*) are partitioned into warps consecutively by thread

---

[5]The Streaming Multiprocessor originally were abbreviated »SM«. For Kepler, the abbreviation changed to SMX. It changed again for Maxwell and is now called SMM. To differentiate the multiprocessor against the Standard Model (SM), the current abbreviation, SMM is used.

[6]Classical SIMD pattern vectorize the operation on data during compile-time.

Table 4.1: Overview of GPU cards tested here. The Devices Under Test (DUTs) comprise two consumer-grade (GeForce) and two HPC-grade (Tesla) devices. The peak performance refers to single-precision performance.

Note: The GTX 580 is not sold any more. The given price is referring to the last retailed version.

Property	Unit	Device			
		GeForce		Tesla	
		GTX 580	GTX 750 Ti	K20X	K40
Manufacturer			NVIDIA		
Release Year		2010	2014	2012	2013
Chipset		Fermi	Maxwell	Kepler	
Multiprocessors	#	16	5	14	15
CUDA **Cores**	#	512	640	2688	2880
GPU Clock	MHz	1544	1110	732	745
Global Memory	GB	1.5	2	6	12
Memory Clock	MHz	2004	2700	2600	3004
Peak Performance	GFLOPS	1580	1306	3950	4290
Price	€	190	150	3600	3800

index. If all data points can not be processed in an equal time by all the cores of one warp, *branch divergence* can occur. One thread takes more time than the other concurrently executed threads. When this happens, the warp has to wait for the last thread to finish its operation, blocking cores in the mean time. To keep cores busy, a SMM has multiple active warps. If one warp waits for memory access, the SMM control activates another warp which is already ready for execution and defers the memory-accessing warp later into the queue [134]. A property of current GPUs is that SMMs have inferior computing throughput for double precision numbers than for single precision numbers. For example, devices of compute capability 5 can instruct 128 single 32 bit operations per clock cycle in one SMM, whereas only one 64 bit operation [133]. Double-precision performance is a main factor of difference between consumer-grade and server-grade GPU devices.

The key to high performance computing on GPUs is to keep all CUDA cores as occupied as possible and to maximize the amount of running threads. Data sets should be large, needing arithmetic challenging computations with scarce memory access – the *arithmetic intensity*[7] should be high. Data should be computed with a high amount of parallelism, keeping the inter-thread dependencies as little as possible to reduce *race conditions*.

Examples of GPU devices with specifications, which are also used during the course of this thesis, are listed in Table 4.1. Both GeForce cards are consumer-grade GPUs cards intended for displaying video imagery, both Tesla cards are supercomputing-grade cards intended for high-performance computing in servers.

---

[7]The ratio of computing operations per memory transfer.

4.1.2   Software Platform

From the software side, NVIDIA GPUs are programmed using the CUDA platform. The CUDA toolkit creates the whole infrastructure needed for parallel GPU programming: It defines the programming language, gives access to the GPU by an API, and ships with an compiler and additional tools (debugger, profiler). The descriptions in this section are referring to the CUDA toolkit version of 6.5, but should be sufficiently general to be valid also for further version.

4.1.2.1   Programming Model

The programming language of NVIDIA's GPU toolkit is also called »CUDA C/C++«, as it builds up on the C++ language, extending it with GPU-specific statements. These *declaration specifiers* usually start with two underscores and specify different types and scopes of functions and methods.

The most prominent specifier is **__global__**, declaring a *kernel* function in CUDA, for example:

```cpp
__global__ void myKernel(float * a) {
        a[0] = 10;
}
```

A kernel is a C function which gets called in parallel a specified number of times. It is the prime keyword enabling parallel programming. When calling a kernel function from a usual C++ host-side program, a triple-chevron extension specifies how many times the kernel should be invoked, the *execution configuration syntax*.

```cpp
int main() {
        // ...
        myKernel<<<20, 10>>>(a_d);
        // ...
}
```

The first number (20) specifies the number of *blocks*, the last number (10) the number of *threads* per block. Blocks and threads are hierarchical ways of invoking large sets of parallel processes. As the number of threads is created per block, the total number of threads is a multiplication of the two parameters ($20 \times 10 = 200$). The index of the currently executed thread in a given block is known to the kernel by the global variables threadIdx and blockIdx[8]. Multiple blocks are structured on a *grid*. Blocks and threads can have up to three dimensions, to provide three unique indices at a time, e.g. threadIdx.x, threadIdx.y, and threadIdx.z.

Among further specifiers extending the C language are **__device__**, specifying functions and data only accessible on the GPU, **__host__**, for functions only available host-side, and **__constant__** and **__shared__**, specifying the respective types of memory, data of an object resides in.

---

[8]With the additional global variable blockDim, the unique id of the current thread can be computed by the following: **int** currentIndex = blockIdx.x * blockDim.x + threadIdx.x.

The workflow of a simple, minimalistic CUDA program using the CUDA runtime API[9] consists of three components: device-side memory allocation and transfer towards it; kernel invocation; transfer back to the host. The parts are exemplary shown in the following code snippet.

```cpp
1   int main() {
2           float * a_h = new float[1];
3           a_h[0] = 3.14;
4
5           float * a_d = new float[1];
6           size_t arrayLength = 1 * sizeof(float);
7
8           cudaMalloc((void **) & a_d, arrayLength);
9           cudaMemcpy(a_d, a_h, arrayLength, cudaMemcpyHostToDevice);
10
11          myKernel<<<1, 1>>>(a_d);
12
13          cudaMemcpy(a_h, a_d, arrayLength, cudaMemcpyDeviceToHost);
14
15          cudaFree(a_d);
16  }
```

Line 2 allocates an array of floats, residing on the host side. The array is filled in line 3. An array intended only for device-side usage is allocated in line 5. In line 8, a call to the CUDA global function `cudaMalloc()` allocates memory for the device-side array `a_d`. The content of the host-side `a_h` is copied to the device-sided `a_d` in line 9 using the built-in `cudaMemcpy()`. Finally, in line 11, the kernel is invoked with 1 block and 1 thread. The kernel function `myKernel()` expects a pointer to a device side float array. After the kernel computations finish, data can be copied back to the host with `cudaMemcpy()` in line 13. This time, the direction is reversed when `cudaMemcpyDeviceToHost` is given as an argument. The content of `a_h` is overwritten in this case. In the end, the device-side memory is freed with `cudaFree()`.

The runtime API exposes more global methods for GPU computing. Important examples are atomic functions and streams. Atomic function are special methods for manipulating data in shared or global memory. They are *read-modify-write* functions, which lock the targeted address during modification for other threads. After successful operation, e.g. by an `atomicAdd()`, the address is released again for other processes. Streams allow for easy ordering of data and processes, e.g. when two or more GPU devices are available.

4.1.2.2 Compilation

CUDA C/C++ code is compiled with NVIDIA's `nvcc` compiler. Since typical CUDA is a mix of code running on the CPU and code intended for the GPU, `nvcc` splits the parts accordingly. The part for the device is compiled into assembly or binary form. The call to the kernel in the host code (e.g. the <<<.,.>>> part) is modified to include runtime calls to the compiled device-side code.

---

[9]Two methods for GPU programming exist in the CUDA toolkit: The high-level runtime API and the low-level driver API.

Finally, the code is given to the host-side compiler, e.g. gcc, to compile the formed code and
link it.

In addition to compiling executable programs, nvcc can also compile shared libraries. Shared
libraries are used, e.g., to link the GPU-accelerated Circle Hough Transform code to PandaRoot.

4.1.3 Terminology

During the course of this thesis, GPU-specific terminology is used. The following list gives an
overview of the most important terms.

Host The computer hosting the GPU card. *Host-side code* is run on the CPU. Host-side code is
mostly run serially.

Device The GPU.

Streaming Multiprocessor A combination of processing cores on the GPU. The building block
of a GPU. A GPU is comprised of several Streaming Multiprocessor (SMM).

Kernel The function executed in a massively parallel fashion. In CUDA, a kernel is identified
by the `__global__` specifier.

Thread, Block, Grid Threads are the smallest units of execution on a GPU device. They process
kernels or launch other threads. Threads are grouped into blocks, blocks grouped onto
a grid, with respective unique indices. Threads of a block are processed on a single
SMM and can share memory. The number of active threads and total threads is limited,
depending on the compute capability of the device.

Warp A collection of 32 threads running concurrently in a SMM.

Block Size The number of threads invoked per block.

Grid Size The number of blocks.

Dynamic Parallelism A feature enabling kernels to spawn other kernels, allowing for dynami-
cally nested kernels, occupying GPUs efficiently.

Compute Capability The version number of an architecture of a GPU card. Usually also defines
the set of features of this specific architecture.

4.2 Online Track Reconstruction

In the DAQ of P̄ANDA, online track reconstruction (*tracking*) has an important role. It is an
essential stage for forming particle hypotheses, needed for online event building, and subsequent
online event triggering. P̄ANDA's online reconstruction has already been introduced in Online
Trigger System, Section 3.4.3. This section discusses specific aspects of potential algorithms for
online track reconstruction.

4.2.1 Objectives of Online Tracking

In the following, the properties of an idealized online track reconstruction algorithm are listed. The list intentionally gives idealistic statements in order to discuss the different points subsequently.

- The prime task of track reconstruction algorithms is to reconstruct particle tracks as *well* as possible. This means:

    - Reconstruction with highest **efficiency** possible, detecting and computing most of the physically present tracks.

    - Reconstruction with highest **purity** possible, creating particle tracks containing hits from only one particle and having minimal contamination by other particles.

    - Reconstruction with highest **resolution** possible, reconstructing the particle's original track as well as possible and, thus, defining the momentum of the reconstructed particle as close to the original momentum as possible.

- Track reconstruction should happen as **early** as possible in the online reconstruction chain. The earlier a track can be formed, the better.

- The reconstruction should run as **fast** as possible, to reduce the latency of computing and number of computing device needed.

- Track reconstruction should **rely on as little information** as possible.

- But in return, the reconstruction should **provide as much information** as possible.

A realistic tracking algorithm optimizes between the above statements, constrained through the algorithm's inherent reconstruction capabilities and the setup of the experiment and the DAQ. The different aspects of optimization are highlighted in the following.

4.2.2 Efficiency

Efficiency of track reconstruction can be defined as the ratio of the number of reconstructed tracks to all tracks present,

$$\varepsilon_{\text{total}} = \frac{N_{\text{reconstructed}}}{N_{\text{all}}}, \tag{4.1}$$

with N denoting the number of tracks.

Since not all tracks present in an event have hit points in any of the sub-detectors, the efficiency can be reformulated to cover only tracks that can be reconstructed:

$$\varepsilon_{\text{reco}} = \frac{N_{\text{reconstructed}}}{N_{\text{reconstructible}}}. \tag{4.2}$$

The kind of tracks that are considered *reconstructible* depends on the sub-detector, the physical event structure, and the algorithm. For $\bar{\text{P}}$ANDA, a loose constraint for reconstructible primary[10] tracks is that the track has at least 2 hits in the MVD or 3 hits in any other detector. For a reconstructible secondary[11] track, 3 MVD or 4 other hits can be a discrimination value.

For an online track reconstruction algorithm, achieving a high efficiency is of utmost importance. Only if an algorithm finds nearly all of the reconstructible tracks can a subsequent (online) event selection operate with high efficiency. In the case of the online software trigger, every not-reconstructed event is a lost event, and for full event reconstruction, every track of the event needs to be found. Partial event reconstruction drastically limits the ability to detect events of the targeted physics channels.

4.2.3   Track Contamination

Efficiency gives a measure of merit of the *quantity* of reconstructed tracks. The *quality* of the reconstruction is not specified a-priori. Different quantities can be defined to reflect the quality of the reconstruction and the contamination of the tracks.

4.2.3.1   Purity

Purity specifies which fraction of hits in one track come from the *correct* particle. The correct particle is the particle, which produces the large majority of hits in the track. The exact value of this large majority which defines the association to the correct particle is subject to definition. A common assumption is that a track with >80 % of its hits from one particle can be considered matching the track of the original particle. The original parameters of the simulated track are accessible via *truth matching*, where the reconstructed track is compared to the original track from the Monte-Carlo (MC) simulation.[12] A high-purity track has a large fraction of hits originating from the correctly associated true track; a track with low purity combines hits from different tracks to one reconstructed track, although the majority can still be from one single track. The hit purity can be defined as:

$$\pi = \frac{n_{\text{correct}}}{n_{\text{all}}}, \tag{4.3}$$

with n_{correct} being the number of hits of the correct MC track and n_{all} the total number of hits of the reconstructed track.

A track can be considered *pure* if all reconstructed hits come from one single MC track. Individual track reconstruction can accept tracks even if they have some external admixture, i.e. $\pi < 1$. A cut value can be 80 %, discarding all tracks with a lower fraction of correct hits. To consider also tracks with even higher fractions of extraneous hits, a category of *contaminated tracks* can be defined, e.g. with a *true* content between 50 % and 80 %.

---

[10]A track originating from the interaction point.
[11]A track of a secondary particle, not originating from the interaction point.
[12]The track of the original particle is also called the *MC track*.

A related quantity to purity is the **coverage** of a MC track. This quantity relates the total number of hits of a MC track to the number of hits in a reconstructed track.

$$c = \frac{n_{\text{reconstructed}}^{\text{correct}}}{n_{\text{MC}}},$$

here n_{MC} is the number of MC hits of the true MC track and $n_{\text{reconstructed}}^{\text{correct}}$ is the number of found hits belonging to the associated MC track in the reconstructed track. Ideally, all MC hits should be in exactly one reconstructed track. The lower limit of c, i.e. the percentage of reconstructed hits of a MC track, is a quantity to be chosen.

4.2.3.2 Ghost Ratio

A quantity relating to purity is the ghost ratio. Instead of referring to individual *hits*, the ratio quantifies *tracks*. It is defined as

$$g = \frac{N_{\text{reconstructed}}^{\text{wrong}}}{N_{\text{reconstructible}}}. \tag{4.4}$$

Here $N_{\text{reconstructed}}^{\text{wrong}}$ is the number of tracks which have been wrongly reconstructed, i.e. have a large fraction of hits not from the true track.

4.2.3.3 Clone Rate

Track reconstruction algorithms can find tracks multiple times when considering different sub-sets of hits of the same true track. The clone rate is the mean number of times one single true track is found among the reconstructed tracks. An effective and efficient algorithm should have a clone rate close to one.

4.2.3.4 Remarks

In contrast to efficiency, a high purity is not of highest importance for the online reconstruction. In later steps, with cost of additional processing time, the purity and clone rate can always be refined and improved. Usually, the effect of low-purity track reconstruction is an increase of track multiplicity and consequently multiply found particles.

A track reconstruction algorithm is usually tuned to maximize efficiency and purity. Increasing efficiency generally lowers purity, whereas increasing purity most often lowers the efficiency. In contrast to offline track reconstruction algorithms, in which the goal usually is to reconstruct events are *cleanly*, online tracking algorithms should always aim for high efficiency.

4.2.4 Momentum Resolution

The momentum resolution is the third parameter specifying the quality of a track reconstruction algorithm. The resolution is usually correlated with both the efficiency and purity. Resolution depends on the measurement capabilities of the sub-detectors, but can also be affected by the techniques employed by the track finding algorithm. While the former is an inherent quantity of each sub-detector, the latter can be improved by executing additional computation.

The momentum resolution (Δp) of a particle is defined as

$$\Delta p = p_{\text{reconstructed}} - p_{\text{correct}} \quad \text{and} \tag{4.5}$$

$$\Delta p_{\text{rel}} = \frac{p_{\text{reconstructed}} - p_{\text{correct}}}{p_{\text{correct}}}, \tag{4.6}$$

where the index *reconstructed* relates to the reconstructed momentum of a particle and *correct* to the original MC momentum of the particle. Equation 4.5 gives the absolute momentum difference, Equation 4.6 the relative difference. A closely related quantity to Equation 4.5 is the pull, for which Δp is divided by the uncertainty of the reconstructed momentum, i.e. $\Delta p / \sigma_{p_{\text{reconstructed}}}$.

In offline track reconstruction, different resolution enhancement algorithms can be applied The most common is track fitting with a Kalman filter, which greatly improves track resolution [119]. In online track reconstruction, every additional algorithm adds to the overall processing time needed for track reconstruction, which should be kept small. Especially a Kalman filter is a comparably slow algorithm, which should be omitted in an online environment, if possible. Ideally, online track reconstruction algorithms should find a collection of hit points from a track and fit the track at the same time; i.e. they should employ *track finding* and (implicit) *track fitting* and provide track candidates with reasonable resolution. The resolution often depends on the time needed for computation, with the acceptable resolution correlated to the acceptable computing time.

4.2.5 Position of Tracking Algorithms

An online tracking algorithm should be as early as possible in the DAQ chain. The sooner tracks can be formed, the sooner event building for background discrimination can happen, reducing the stream of data. The position is constrained by the available data and the structure of the data. Detector-specific hit points first need to be created and the sub-detector-ordered stream of hits merged and sorted by time. Since $\bar{\text{P}}$ANDA runs continuously, a track reconstruction algorithm should be able to form tracks from a time-wise sorted stream of hits. An event-based operating algorithm can only be employed after event building, an additional step, needing further input information and adding to the overall execution time. Adding this step is only feasible if the benefits (e.g. a very precisely reconstructed track) outweigh the drawbacks (additional processing time).

In general, the computing infrastructure also constrains tracking algorithms and their position in the DAQ chain. Different architectures retrieve data by employing different transfer mechanism and hence are more effective at different stages in the read-out chain. ASICs process detector data in its rawest form; GPUs and CPUs a-priori[13] expect data in higher-level formats; FPGAs lie in-between and can be embedded more deeply into networked systems. The choice of position in the chain affects the hardware, which affects the choice of the tracking algorithm, since not all algorithm perform equally well on all platforms.

---

[13]Hardware for directing raw data directly onto the GPU is currently developed, making an operation of GPUs closer to the sub-detector possible [135, 136].

4.2.6 Performance

Apart from finding tracks with high efficiency, algorithmic speed is essential for online algorithms. The need for running in real-time imposes limits on the algorithmic design. Several ways can be chosen to match performance limitations in online reconstruction algorithms.

As stated above, one choice could be to prefer coarse and fast reconstruction over pure and precise reconstruction, if it still allows for proper event building. If possible, slow track enhancement algorithms, as Kalman filters and other track fitters, should be omitted.

Parallelism should be exploited on all levels possible: A computing device should process more than one event or time-window at a time to keep the processors as busy as possible; multiple devices can be used to expand parallelism beyond a single device. It might be necessary to interconnect parallel-running devices for data exchange during computation. As this potentially has a negative effect on computing throughput, a more favorable solution is to run parallel devices as individual entities. The HESR's bunch structure (see Section 3.1.3.2) can provide a natural choice of data segmentation.

The computing architecture employed can also limit the performance of algorithms, as they might run with different efficiencies. While GPUs tend to offer a more flexible programmability and higher computing throughput potential, FPGAs operate closer to the raw data and hold potential for fast computation of specific tasks [137, 138]. Multi-CPU systems, like a Xeon Phi, are a recent addition to small-scale HPC devices and might find their specific usages as well.

A heterogeneous assembly of different computing devices, operating with specifically tuned, distinct track reconstruction algorithms at various places in the DAQ chain, seems to be a probable solution matching the demands of $\overline{\text{P}}$ANDA's online track reconstruction.

4.2.7 Available and Provided Information

Depending on the position of an online tracking algorithm in the DAQ chain, not all event information is available. Especially the event starting time t_0, needed for precise track reconstruction in the STT, requires significant computation before it is available. The employed track reconstruction algorithms need to be able to operate with limited event information, or be complemented with pre-processing algorithms. In general, the algorithm should have the ability to compute tracks with as little information as possible. As long as the resulting tracks can still be used to form track candidates, an *economical* algorithm is to be preferred.

The different tracking and PID sub-detectors of $\overline{\text{P}}$ANDA all have distinct specializations, as outlined in Section 3.3. Tracking algorithms should be tailored to match the available information. Hit points of the MVD have full 3D information, provide precise time of occurrence, and are taken close to the interaction point. But the limited space entails also a limited number of hit points. The STT covers a much larger volume and instruments it nearly fully, providing a stronger lever arm for tracks with many hit points. The z-information for the hit points is only determined when including the skewed straws; for the best measurement resolution, the event time t_0 needs to be present. Tracking algorithms can first use the MVD hits to define regions of interest in the STT, both spatially as well as time-wise. This reduces computing complexity and processing time. In addition, the time information can be propagated.

Track candidates originating from one algorithm can be input of another algorithm. While a first tracking algorithm can compute rough track candidates without t_0, the track information can in turn be used to constraint t_0, and function as a basis for a more precise subsequent tracking algorithm, operating with the t_0 assumption. Algorithms computing tracks originating not from the interaction point at $(0, 0, 0)$, but coming from particles with displaced vertices, are computationally more complex. While the interaction point provides a fixed point for primary-track finding algorithms, this information is missing when the complete inner detector volume might be the origin of the track. To reduce computational complexity, different algorithms can operate on different sub-sets of data. A fast tracking algorithm capable of finding primary tracks with high efficiency can be used to reduce the number of hit points as input for a second algorithm. The hit points already associated to a track candidate can be removed from the data stream, drastically lowering the overall number of possible tracks from displaced vertices. A second-stage algorithm for finding secondary tracks then needs to consider only a reduced number of hits. In general, this approach can be repeated with more and more specialized algorithms which reduce the number of hits along the way.

4.2.8   Summary

A **realistic online track reconstruction** contains a set of different specialized algorithms, which must find an optimal tradeoff between purity and performance, but keep maximum efficiency, as every not-reconstructed track is a potentially lost event. The algorithms are built in a multi-stage scheme, reconstructing first simple, then more and more complex track structures, reducing the number of not-associated hits along the way to decrease computational complexity. Track reconstruction algorithms are tailored for specific sub-detectors, for positions in the DAQ chain, and for computing architectures.

4.2.9   Terminology

Different track reconstruction terminology is used in this chapter. The following list gives an overview.

Efficiency ε The ratio of reconstructed tracks to all tracks or to all reconstructible tracks. See Section 4.2.2 and Equation 4.2.

Purity π The fraction of true hits among all hits of a track. See Section 4.2.3 and Equation 4.3.

Ghost ratio g The fraction of wrongly reconstructed tracks among all (reconstructible) tracks. See Equation 4.4.

Resolution For a track reconstruction algorithm, usually the momentum resolution. The difference of the momentum of a reconstructed track to the momentum of the original, simulated MC track. See Equation 4.6.

MC Track The original simulated track from the event generator, associated to the reconstructed track by mean of *MC matching*.

Tracklet A part of a track, still in need of assembly to a full track or track candidate.

Primary Track, Secondary Track, Displaced Track A primary track is a track originating from the interaction point. A secondary or displaced track is a track from a particle not originating from the interaction point, but from a displaced vertex.

4.3   Line Hough Transform

4.3.1   Introduction into Hough Transforms

The Hough transform (HT) is a numerical method for finding parameters which connect a set of measurements best, according to a specific model. In its simplest form, the HT is used for edge detection in images and finds straight lines connecting pixels.

The method was first patented by HOUGH in 1962 [139] for automated scanning of pictures of bubble chambers. It was adapted and generalized by DUDA and HART ten years later [140].

The method generates a set of parameters for each analyzed point. An *accumulator array* keeps track of the generated parameters: For every value of the parameter, the according bin of the accumulator array is incremented. The procedure is repeated for every analyzed point. Finally, the parameters with the highest frequency are extracted from the accumulator array – parameters fitting best to all analyzed points. Taking the original example for straight-line edge detection in images: a straight line is defined by two points, e.g. the slope and the y-intercept. One value is sampled while the other is computed, as the position of the currently analyzed pixel constrains the problem to having two degrees of freedom. Every parameter pair is incremented in a two-dimensional accumulator array (i.e. a histogram). Eventually, the set of line parameters equal to the line best connecting all pixels can be found as the array bin with the highest number of entries.

In particle physics experiments, a HT can be used to connect hit points by lines or curves. Compared to the HT's usual application for detection of edges in pictures with plenty of pixels, the usage in particle physics has a comparably low number of pixels (hit points) to be connected by a line. This complicates the extraction of the bin in the accumulator array with the highest number of entries.

In this section, the Line Hough Transform considered for $\overline{\text{P}}$ANDA is presented. A different Hough transform, the Circle Hough Transform, is the subject of the next section. In the following, the algorithm is presented and its formulation for $\overline{\text{P}}$ANDA derived. Two implementations on GPUs are presented and their performances benchmarked.

4.3.2   Algorithmic Description

The Line Hough Transform (LHT) reconstructs particle tracks as straight lines, close to the original formulation of the HT by DUDA and HART. A pre- and post-processing step is needed to consider the bending of the trajectory due to $\overline{\text{P}}$ANDA's solenoid magnetic field.

The algorithm first reconstructs tracks in the projection onto the xy plane, the plane perpendicular to the beam axis. To expand the tracks from their two-dimensional (2D) representation as a circle into a three-dimensional (3D) helix, an additional step is needed. The first step,
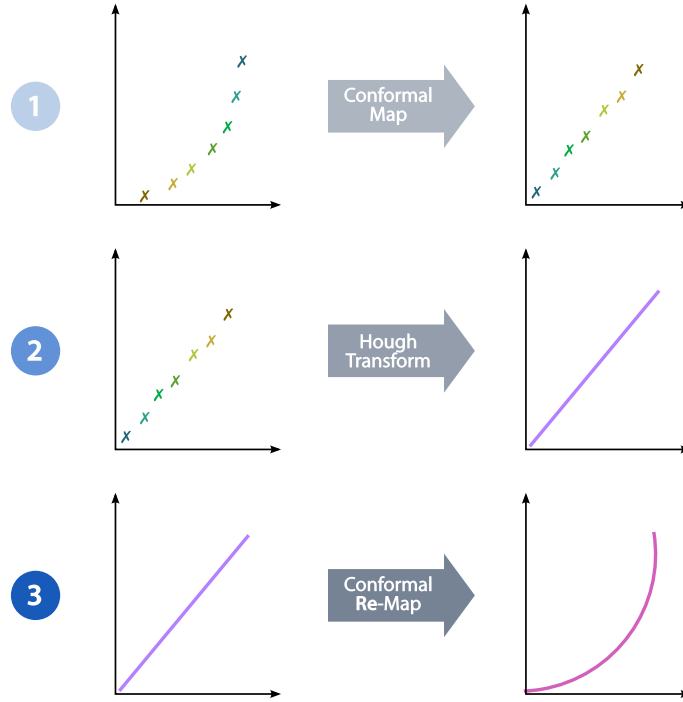
Figure 4.3: The individual steps of the Line Hough Transform.

though, is the more complicated and computing-intensive one, as a multitude of hits have to be considered to reconstruct meaningful track candidates. Once the 2D candidates exist, the 3D expansion has only to consider a problem of very limited, if at all, combinatorial complexity. A way to expand 2D track candidates into the third dimension is to apply a $s-z$ fit, a fit along a track's circular segment s and the z coordinate. In a solenoidal magnetic field, this functional dependency is linear.

Figure 4.3 outlines the different stages of the LHT. Step ① pre-processes the considered hit points and removes the circle shape the hit points are distributed on. A conformal map is chosen for this. The actual HT follows in step ②, which determines the continuous line connecting all hit points. Step ③ recovers the information about the bending radius from step ①.

In the following, the different equations needed for the LHT of $\overline{\text{P}}$ANDA are derived.

4.3.2.1   Conformal Map ①

A conformal map is a projection of a point into an image domain, preserving (oriented) angles of curves going through the point. The angles are preserved locally, while the lengths of the lines or the curvatures can change. This feature is exploited when applying a conformal map as a LHT pre-step: The hit points keep their orientation to each other, while the curvature of the connecting line (*track*) changes.

Using an appropriate conformal map, hit points lying on a common circle are transformed to hit points lying on a common line. Usually, hit points lying close to the coordinate origin on the
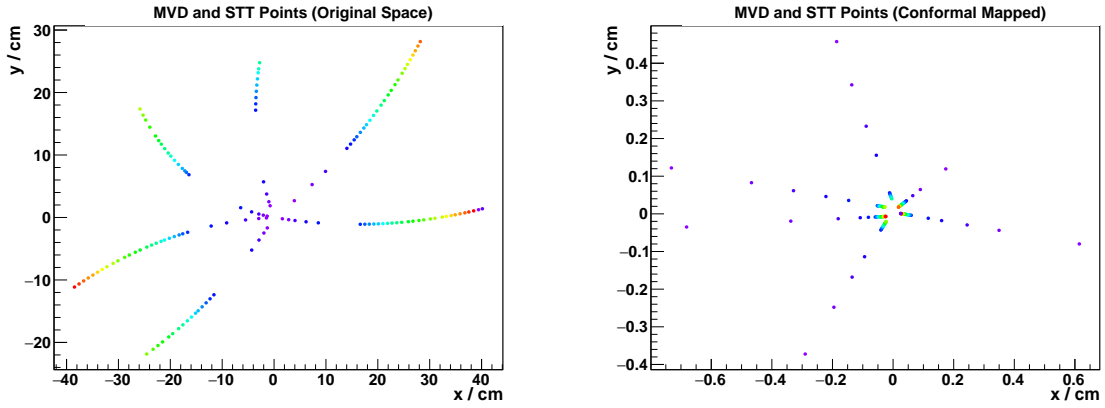
(a) Hit points in the MVD and the STT **before** the conformal mapping.

(b) Hit points in the MVD and the STT **after** the conformal mapping.

Figure 4.4: Conformal map of MVD and STT hit points in the xy plane of $\overline{\text{P}}$ANDA (neglecting STT isochrone information). The hit points originate from six charged tracks. The left image shows the spatial distribution of the hit points before conformal map, the right image after. Left, hits from the more densely instrumented STT can be seen further outside, hits from the MVD are closer to the interaction point. The points are colored from inside to outside in the left picture. The color is kept for the right image – the inversion of the order can be seen.

curved trajectory are transformed to hit points lying at large distances from the origin on the straight line – their radial order is essentially inversed. An illustration is given in Figure 4.4.

In the following, the simplest case introduces a conformal map for (x, y) point-like hit points. After that, a more general conformal map for extended hit points is derived.

Point-like Hit Points   A hit point with the coordinates (x, y) is mapped into a conformal space noted with primes by

$$x' = \frac{x}{x^2 + y^2} \quad \text{and} \tag{4.7}$$

$$y' = \frac{y}{x^2 + y^2} \quad \text{, or combined:} \tag{4.8}$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{x^2 + y^2} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{4.9}$$

This mapping can be used for hit points of sub-detectors providing 2D hit points with x and y value – e.g. hit points of the MVD or GEM.

Extended Hit Points (Circles)   Hit points from the STT are defined ambiguously: A straw's anode wire position (x, y), is complemented by the length the electrons drift in the straw. This improves the spatial resolution of a hit point from the STT. Due to the cylindrical geometry, the electron drift distance is actually a drift radius around the anode wire. A circle around the wire's center is created, spanning all possible passage points. The drift radius, r, is also referred to as the *isochrone radius*, as all passage points are isochronous around the wire center.

The conformal map from Section 4.3.2.1 needs to be extended to consider the isochrone radii, as only with this can the STT's achieve nominal resolution. An isochronous hit point is parameterized by (x_0, y_0, r_0).

Equation 4.7 rearranges to the following:

$$x' = \frac{x}{x^2 + y^2} \tag{4.10}$$

$$\Rightarrow y = \sqrt{\frac{x}{x'} - x^2}. \tag{4.11}$$

This can be inserted into Equation 4.8:

$$y' = \frac{\sqrt{\frac{x}{x'} - x^2}}{x^2 + \left(\sqrt{\frac{x}{x'} - x^2}\right)}$$

$$\Rightarrow x = \frac{x'}{x'^2 + y'^2}. \tag{4.12}$$

Inserting this in Equation 4.11 yields the second part of the inversion:

$$y = \frac{y'}{y'^2 + x'^2}. \tag{4.13}$$

In addition to these $x(x', y')$ and $y(x', y')$ conformal map inversions, a combined relation is easily found:

$$x^2 + y^2 = \frac{1}{x'^2 + y'^2}. \tag{4.14}$$

The equation of a circle around (x_0, y_0) with radius r_0 is

$$(x - x_0)^2 + (y - y_0)^2 = r_0^2. \tag{4.15}$$

It rearranges to:

$$x^2 - 2xx_0 + x_0^2 + y^2 - 2yy_0 + y_0^2 = r_0^2$$

$$\Rightarrow x^2 + y^2 - 2xx_0 - 2yy_0 + x_0^2 + y_0^2 - r_0^2 = 0.$$

Equations 4.12, 4.13, and 4.14 can now be used to transform the circle into conformal space:

$$0 = \frac{1}{x'^2 + y'^2} - 2x_0 \frac{x'}{x'^2 + y'^2} - 2y_0 \frac{y'}{x'^2 + y'^2} + x_0^2 + y_0^2 - r_0^2$$

$$= 1 - 2x_0 x' - 2y_0 y' + \underbrace{(x_0^2 + y_0^2 - r_0^2)}_{\equiv n} \cdot (x'^2 + y'^2)$$

$$= \frac{1}{n} - \frac{2x_0}{n} x' - \frac{2y_0}{n} y' + (x'^2 + y'^2)$$

$$= \frac{1}{n} + \left(x' - \frac{x_0}{n}\right)^2 - \frac{x_0^2}{n^2} + \left(y' - \frac{y_0}{n}\right)^2 - \frac{y_0^2}{n^2}$$

$$\Rightarrow \left(x' - \frac{x_0}{n}\right)^2 + \left(y' - \frac{y_0}{n}\right)^2 = \frac{x_0^2}{n^2} + \frac{y_0^2}{n^2} - \frac{1}{n}$$

$$= \frac{r_0^2}{n^2}$$

Resubstituting the intermediate definition $n \equiv x_0^2 + y_0^2 - r_0^2$ yields:

$$\left(x' - \underbrace{\frac{x_0}{x_0^2 + y_0^2 - r_0^2}}_{\equiv x_0'} \right)^2 + \left(y' - \underbrace{\frac{y_0}{x_0^2 + y_0^2 - r_0^2}}_{\equiv y_0'} \right)^2 = \left(\underbrace{\frac{r_0}{x_0^2 + y_0^2 - r_0^2}}_{\equiv r_0'} \right)^2 \tag{4.16}$$

Comparing this with Equation 4.15, it becomes clear that this is a circle in conformal space around new center coordinates x_0' and y_0' with radius r_0' as defined in Equation 4.16.

Summarizing: To conformal map STT hit points (x, y) with isochrones of radii r analogously to Equation 4.9, the following relation is to be used:

$$\begin{pmatrix} x_\odot' \\ y_\odot' \\ r_\odot' \end{pmatrix} = \frac{1}{x^2 + y^2 - r^2} \begin{pmatrix} x \\ y \\ r \end{pmatrix}. \tag{4.17}$$

Conformal Map of Circles Not From the Origin   The above equations (Equations 4.17 and 4.9), are valid for mapping points lying on circles going through the origin at $(0,0)$. This applies to tracks originating from the interaction point, *primary tracks*. To recover tracks from secondary vertices – particles decaying $\mathcal{O}(\text{cm})$ from the interaction point – the conformal map needs to be adapted.

For a mapping with the hypothesis of a circle not going through $(0,0)$, but through (x_d, y_d), and choosing the point-like conformal map for simplicity, one can substitute the variables:

$$\begin{aligned} x_\rightarrow &= x - x_d \\ y_\rightarrow &= y - y_d \, . \end{aligned} \tag{4.18}$$

Inserting this in Equation 4.9 yields the following map for circles not going through the origin,

$$\begin{pmatrix} x_\rightarrow' \\ y_\rightarrow' \end{pmatrix} = \frac{1}{\left(x - x_d\right)^2 + \left(y - y_d\right)^2} \begin{pmatrix} x - x_d \\ y - y_d \end{pmatrix}. \tag{4.19}$$

It can easily be seen that the case for circles from the origin (Equation 4.9) is a special case of Equation 4.19 for $x_d = y_d = 0$.

Currently, this conformal map is not used in tracking algorithms at $\overline{\text{P}}\text{ANDA}$, as it significantly increases the amount of computations necessary. In the simple case of circles from the origin, Equation 4.9, one (x', y') pair for every (x, y) pair is computed; a linear function. The conformal map for displaced tracks (Equation 4.19) computes one $(x_\rightarrow', y_\rightarrow)$ pair for every hit point and every passing point (x_d, y_d). If the latter is not known, a large number of point hypotheses need to be tested, resulting in a large number of computations, also in any subsequent algorithms.

4.3.2.2 Hough Transform ②

As outlined in the introduction of this section, the LHT consists of the following steps:

- Consider a hit point $i = 1$.
- Generate a line, $j = 1$, going through that hit point.
- Fill the line parameters in an accumulator array / histogram.
- Generate another line, $j = 2$, going through the same hit point, but inclined slightly differently than the previous line.
- Fill these line parameters $j = 2$ into the array / histogram.
- Continue generating (*sampling*) lines through the hit, until a terminating condition is reached (e.g. the generated line is equal to the initial line). Fill the line parameters into the array.
- Repeat the previous steps for all hit points $i = 2, 3, \ldots$.
- After sampling all hit points, extract the bin of the accumulator array / histogram, which has the most entries.

The generation of a line going through a point transfers the point to the parameter space, the *Hough space*. The choice of the Hough sampling function determines the structure of the Hough space. The space needed for a LHT is two-dimensional, as a line can be defined by two parameters (e.g. its slope and the y-intercept). More parameters could be sampled and filled into Hough spaces of higher dimensions. In practice, this approach greatly increases computing time and quickly becomes impractical.

One parameter of a line equation is sampled to calculate the other. The choice of the sampling granularity determines the resolution with which the track can be reconstructed. Sampling with higher granularity (i.e. more sample parameters) increases the resolution, but also the number of computations. An upper limit to the sampling granularity is given by the sub-detector resolution. In this way, the HT considers measurement uncertainties implicitly by the choice of the sampling step size.

Similar to the conformal map, also the LHT is slightly modified for point-like hit points and spread-out hit points. In the following, first the line equation for point-like hit points is derived, followed by the equation for extended hit points.

Point-like Hit Points   It is useful to replace the usual parameterization of a line, $y = mx + b$, with a *normal* parameterization.

DUDA and HART [140] introduced the normal parameterization that is shown in Figure 4.5. It uses the minimum distance between the line and the origin of the coordinate system ρ, and the angle θ between the line and the x axis. The equation is:

$$\rho = x \cos \theta + y \sin \theta, \tag{4.20}$$
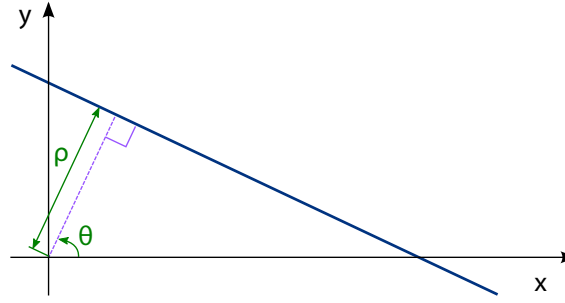
Figure 4.5: The parameters used for the line equation of the Line Hough Transform.
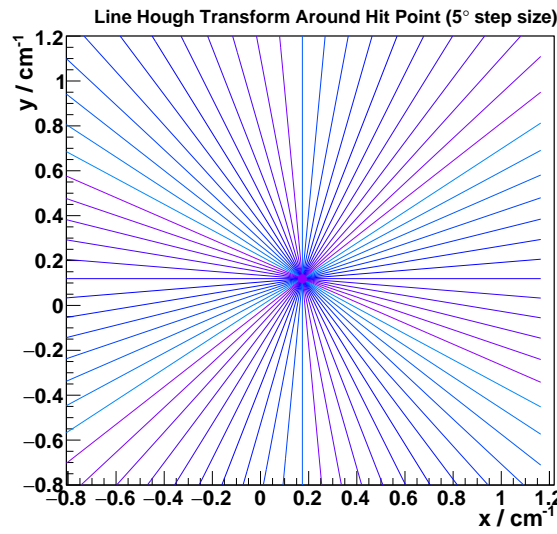
Figure 4.6: A visualization of lines generated in the Line Hough Transform. A conformal-mapped MVD hit point without any isochronous information is chosen. The chosen step size is $\Delta\theta = 5°$, with $\theta \in [0°, 180°)$.

where (x, y) are the coordinates of a point.

θ is the sampled value within the HT. Using i to denote the hit points considered and using j to mark the current sampled value, Equation 4.21 becomes:

$$\rho_{ij} = x_i \cos\theta_j + y_i \sin\theta_j. \tag{4.21}$$

The *Hough angle* θ is sampled between $0°$ and $180°$. Greater values of θ are redundant. The choice of the sampling step size of θ defines the resolution of the reconstruction and the number of computations. In Figure 4.6, a conformal-mapped MVD hit point from Figure 4.4(b) is sampled with lines in steps of $5°$.

Extended Hit Points (Circles) As in the case of the conformal map in Section 4.3.2.1, the extended hit points of the STT need special treatment for the Hough transform.
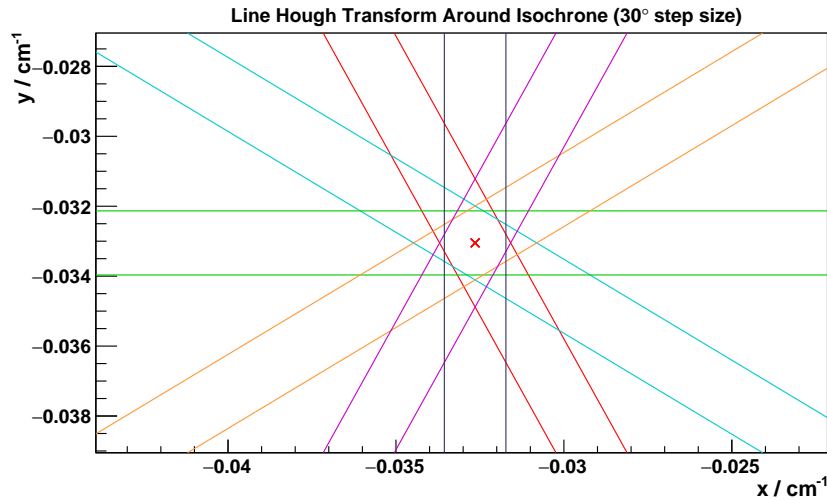
(a) Hough transform with a step size of $\Delta\theta = 30°$.
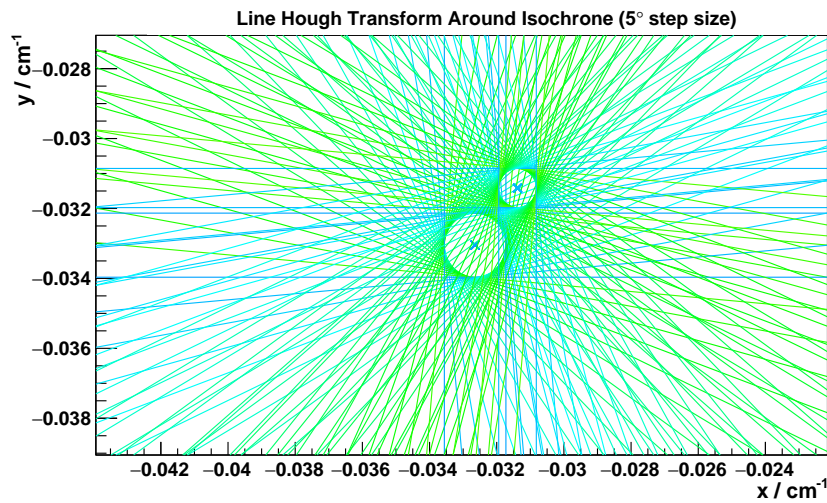
(b) Hough transform for two hit points with a step size $\Delta\theta = 5°$ and an additional STT hit point.

Figure 4.7: Visualization of Line Hough Transform around STT hit points with isochrone information, after conformal map. For clarity, (a) shows only a small number of lines, equivalent to a large step size. In (b), the granularity is higher and an additional second hit point is shown. The coincidental generated tangential lines on two sides of the isochrone have the same color as can be seen especially in (a).

The HT does not generate a line going through the center of the drift tube – the wire at (x, y) – but one tangential to the isochrone. To derive a formula, the point to be Hough transformed is moved *virtually* from the straw's center to lie on the isochrone circle. With $x_\odot$ and $y_\odot$ denoting this shifted point and r the isochrone radius, one finds:

$$x_\odot = x + r \cos \theta \text{ and}$$
$$y_\odot = y + r \sin \theta \ .$$

This can be inserted into the usual HT of Equation 4.20:

$$\begin{aligned}
\rho_\odot^+ &= x_\odot \cos \theta + y_\odot \sin \theta \\
&= (x + r \cos \theta) \cos \theta + \left(y + r \sin \theta\right) + r \sin \theta \\
&= x \cos \theta + y \sin \theta + r \underbrace{(\cos^2 \theta + \sin^2 \theta)}_{=1} \\
&= x \cos \theta + y \sin \theta + r.
\end{aligned}$$

Since the range of θ is limited, $\theta \in [0°, 180°)$, one needs to account for the two sides of the isochrone. This can either be done by extending the range of θ to 360° to go completely around the wire center, or by calculating both sides at once. To keep the same structure for both the point-like HT and the isochronous HT, the latter is chosen. Hence, the final equation with indices as before for Hough transforming isochronous hit points is:

$$\rho_\odot^{ij} = x_i \cos \theta_j + y_i \sin \theta_j \pm r_i. \tag{4.22}$$

A visualization is given with Figure 4.7. The coarse step size in θ used in Figure 4.7(a) illustrates the two simultaneous generated lines. Figure 4.7(b) shows two STT points with isochrones next to each other, sampled more finely.

Hough Space   Each sampled pair of line parameters (ρ_{ij}, θ_j) is incremented in an accumulator array. Since one parameter is sampled and the other calculated by means of the line equation, the array is of dimension 2. The resolution of the track is determined by the size of each cell (bin) of the array and by the LHT sampling step width. The cell size needs to be chosen carefully, as a too large size leads to many line parameter incrementing the same bin. A possible representation of an accumulator array is a histogram, like the TH2 classes provided by ROOT.

Figure 4.8 shows histograms of Hough spaces for MVD **(a)** and STT **(b)** hits after conformal mapping. The Hough sinusoids of the MVD points are spread out further, as there are larger gaps between the layers of the MVD than for the the tightly structured STT straws. The STT sinusoids lie much closer to the x-axis vertically because the hit points are further away from the interaction point. The Hough space of a track in the STT has more entries, since the STT has more layers than the MVD and two parameter pairs are generated per STT hit. To extract track parameters from the bin with the highest content, the MVD and STT hit points should be considered separately, as a common histogram would be dominated by STT values.
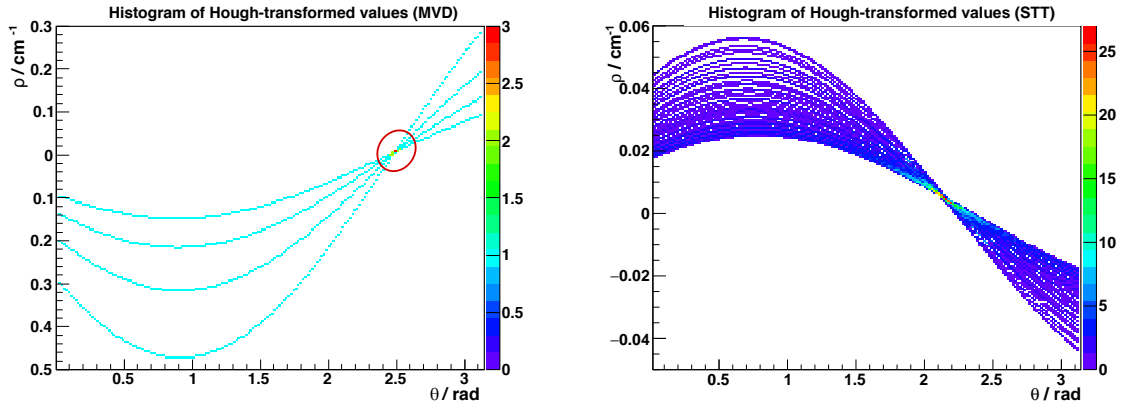
(a) Hough transform for four MVD hit points. The hit points are lying comparably far away from each other, leading to spread out Hough sinusoids. The effect is amplified further by the conformal map. The point of overlap can be seen clearly at approximately $(2.5\,\text{rad}, 0\,\text{cm}^{-1})$.

(b) Hough transform for 24 STT hit points with isochrones. Compared to the MVD points in **(a)**, the sinusoids are much closer together, even with a five times smaller y-scale.

Figure 4.8: Histograms of Hough spaces for hit points, Hough-transformed per Equation 4.22 (after applied conformal map). The sampling angle is chosen to be $\Delta\theta = 1°$. The x-axis displays radians and ranges from $\theta = 0$ to $\theta = 180°$. The histogram has 180 bins on both the x and y-axis, corresponding to one bin for each θ value.

Properties and Relations   Comparing the real (or conformal) space and the Hough space, different properties can be noted [140]:

- A hit point in the real space is transformed to a sinusoidal curve in the Hough space – a *sinusoid*.

- A point in the Hough space is equal to a line in the real space.

- Hit points lying on the same line in the real space correspond to sinusoids intersecting at the same point in the Hough space.

- Sinusoids of hit points of the same line form bands in the Hough space. The width of the band is smallest at the point equal to the line's parameters (the intersection) and largest, $\pi/2$ away from it.

Extraction of Line Parameters   The task of extracting line parameters of the Hough space is non-trivial. As seen in Figure 4.9, the space for tracks from one single event is filled with many values: Here, 140 hit points create 140 sinusoids, sampled every 1° in $\theta \in [0°, 180°)$ creates $140 \times 180 = 25{,}200$ parameter pairs filled into a histogram. The Hough bands are overlapping, sinusoids of hits from different lines intersecting.

Figure 4.9 shows tracks from only one event. For a first- or second-stage algorithm the event building is not necessary complete, thus this image is a best-case scenario.

(a) Large ρ range. **(b)** Zoom to ρ values close to ρ = 0.
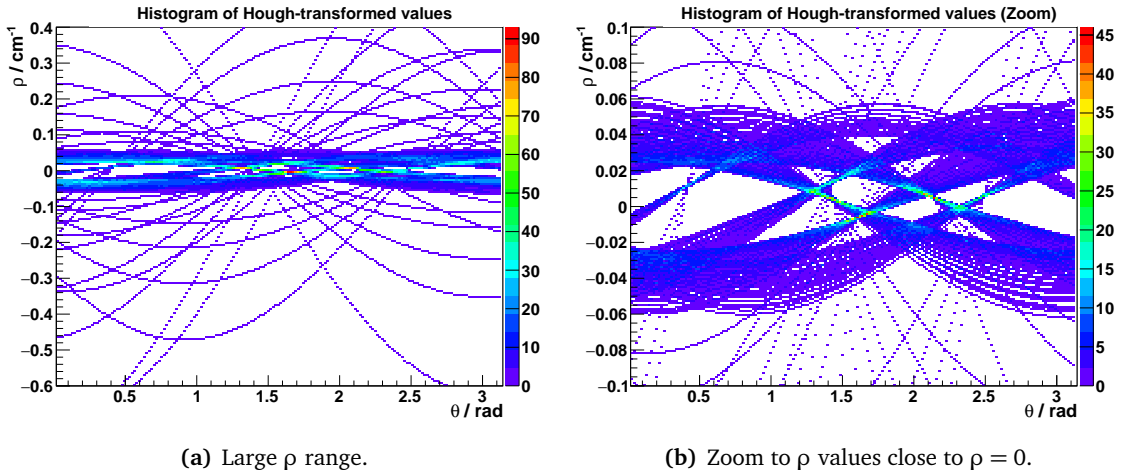
Figure 4.9: Hough histogram of the MVD and STT hit points shown in Figure 4.4(b). A step size of $\Delta\theta = 1°$ is chosen. In total, the sinusoids of 140 hit points from 6 tracks are shown.

Different simplification techniques may be applied: loose threshold cuts remove bins with low number of entries (the majority) and limit the scope of the parameter search; contrast filters to heighten the peaks of band overlaps; iterative global peak finding; *hill climbing* along highest gradients; event sub-division; a combination of all. All approaches differ in efficiency and performance and are suited only for specific problems. For this thesis, the simplification of per-track peak finding and an iterative maximum finding is used, since the focus lies on the track reconstruction algorithms.

4.3.2.3 Conformal Re-Map ③

After finding the parameters of a line using the HT, the initial conformal mapping has to be inverted, as illustrated in part ③ of Figure 4.3.

General Derivation The most general form of a straight line equation is

$$y' = mx' + b. \tag{4.23}$$

Since the track is found in the conformal-mapped Hough space, primed variables are chosen. Using Equation 4.9, a solution using the *real* space variables x and y can be found:

$$\frac{y}{x^2 + y^2} = m\frac{x}{x^2 + y^2} + b.$$

This equation can be rearranged to the following:

$$
\begin{aligned}
0 &= mx - y + bx^2 + by^2 \\
&= \frac{m}{b}x - \frac{1}{b}y + x^2 + y^2 \\
&= x^2 + \frac{m}{b}x + y^2 - \frac{1}{b}y.
\end{aligned}
$$

Completing the square, further rearrangements can be done:

$$
\begin{aligned}
0 &= x^2 + \frac{m}{b}x + y^2 - \frac{1}{b}y \\
&= x^2 + \frac{m}{b}x + \frac{m^2}{4b^2} - \frac{m^2}{4b^2} + y^2 - \frac{1}{b}y + \frac{1}{4b^2} - \frac{1}{4b^2} \\
&= \left(x + \frac{m}{2b} \right)^2 - \frac{m^2}{4b^2} + \left(y - \frac{1}{2b} \right)^2 - \frac{1}{4b^2} \\
&= \left(x + \frac{m}{2b} \right)^2 + \left(y - \frac{1}{2b} \right)^2 - \frac{1+m^2}{4b^2}.
\end{aligned}
$$

This is the equation of a circle:

$$
\left(x - \underbrace{\frac{-m}{2b}}_{\equiv x_0} \right)^2 + \left(y - \underbrace{\frac{1}{2b}}_{\equiv y_0} \right)^2 = \underbrace{\frac{1+m^2}{4b^2}}_{\equiv r_0^2}, \tag{4.24}
$$

with the coordinates of the center (x_0, y_0) and a radius r_0 equal to

$$
\begin{aligned}
x_0 &= \frac{-m}{2b} \\
y_0 &= \frac{1}{2b} \\
r_0 &= \frac{\sqrt{m+1}}{2b}, \text{ or combined}
\end{aligned}
$$

$$
\begin{pmatrix} x_0 \\ y_0 \\ r_0 \end{pmatrix} = \frac{1}{2b} \begin{pmatrix} -m \\ 1 \\ \sqrt{m+1} \end{pmatrix}. \tag{4.25}
$$

This fully defined circle sets the parameters of the track connecting the initial hit points, neglecting any deviations of the track, e.g. due to multiple scattering.

Re-Map for Normal Line Parameterization   In Section 4.3.2.2 a normal parameterization for a line was used (see Equation 4.20). It can be brought to the general relation of Equation 4.23:

$$
\begin{aligned}
\rho &= x' \cos\theta + y' \sin\theta \\
\Rightarrow y' &= -\underbrace{\frac{\cos\theta}{\sin\theta}}_{\equiv m} x' + \underbrace{\frac{\rho}{\sin\theta}}_{\equiv b}.
\end{aligned}
$$

Using the substitutes for m and b, the center and radius relation of Equation 4.25 rearranges to

$$
\begin{pmatrix} x_0 \\ y_0 \\ r_0 \end{pmatrix} = \frac{\sin\theta}{2\rho} \begin{pmatrix} \frac{-\cos\theta}{\sin\theta} \\ 1 \\ \sqrt{\frac{-\cos\theta}{\sin\theta} + 1} \end{pmatrix}. \tag{4.26}
$$

4.3.3 Performance Measurements on GPUs

The LHT is widely used in many track reconstruction applications. The transform is part of P̄ANDA's default offline track reconstruction in the Target Spectrometer and in the Forward Spectrometer. There are versions running on classical CPU architectures and specialized versions on FPGAs [141].

The reconstruction capabilities of physical events of LHTs have been shown previously, e.g. [142]. This section hence focuses on benchmarking the GPU implementations of LHTs.

Two different GPU implementations are presented. The first uses Thrust as a basis, the second is built directly in CUDA, using advanced techniques like dynamic parallelism.

4.3.3.1 **Thrust** Implementation

Thrust is a package for CUDA which adds a templated programming pattern and a multitude of pre-programmed algorithms to the programming environment. The package is built in reference to the C⁺⁺ Standard Template Library (STL) and extends it for GPU tasks. Sets of data points are combined in *vector*s, but instead of the STL `std::vector<int>`[14], Thrust offers a vector for host-side data (`thrust::host_vector<int>`), and a vector type reserved to reside on the device side, `thrust::device_vector<int>`. Setting a `device_vector` to a `host_vector` transfers the data from the host to the device. Subsequently, one of the pre-built functions can be used on a vector, e.g. `thrust::sort()`.

Using `thrust::transform()`, Thrust's iterators, and custom operators, a version of the LHT is programmed. The scheme of the algorithm is as follows:

Conformal Map Using `thrust::transform()`, the x, y and r values of the studied hit points are mapped.

Angle Generation A `thrust::device_vector<>` containing one angle value θ per entry is generated in parallel using `thrust::sequence()`. This vector is the basis for the following parallelism.

Hough Transformation With Thrust's `thrust::transform()`, Hough values are generated. The process is parallel in the angles: For every θ value from the previous step, a LHT value is computed in parallel. Due to limitation of Thrust, the algorithm is serial in the hits.

Accumulator Array Filling As an optional step, a matrix is filled in parallel as an accumulator array. Another template-oriented package for linear algebra with sparse matrices, cusp [143], is employed for this step. Array filling is omitted for the following time measurements.

Different Numbers of Hits In Figure 4.10, normalized run times of the different parts of the algorithm are shown in dependence of the number of processed hits. The scope of the parallelism can be seen, as the graph for the actual HT is constant for varying number of hits.

---

[14]`int` is only chosen as an example.
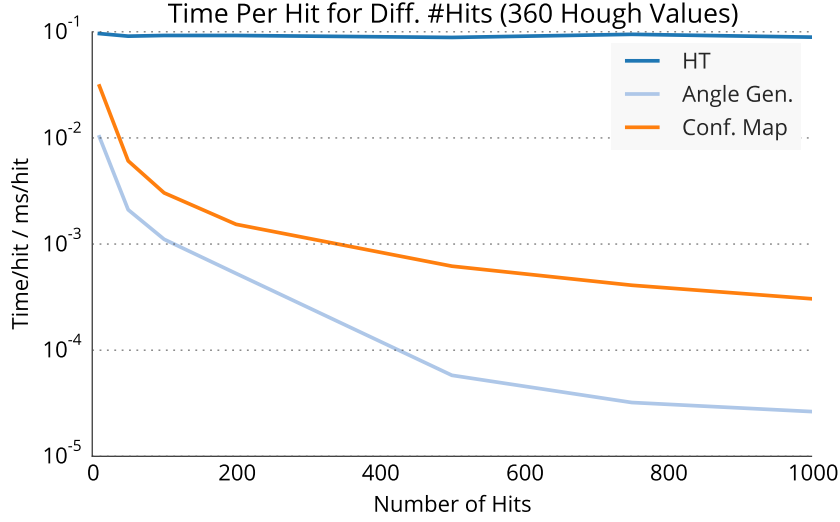
Figure 4.10: Time measurement of the different steps of the Thrust-based Line Hough Transform. Normalized to the number of hits. A GeForce GTX 750 Ti GPU with floating-point precision data types is used, evaluated for 360 Hough values. A performance plot is given in Figure A.1.
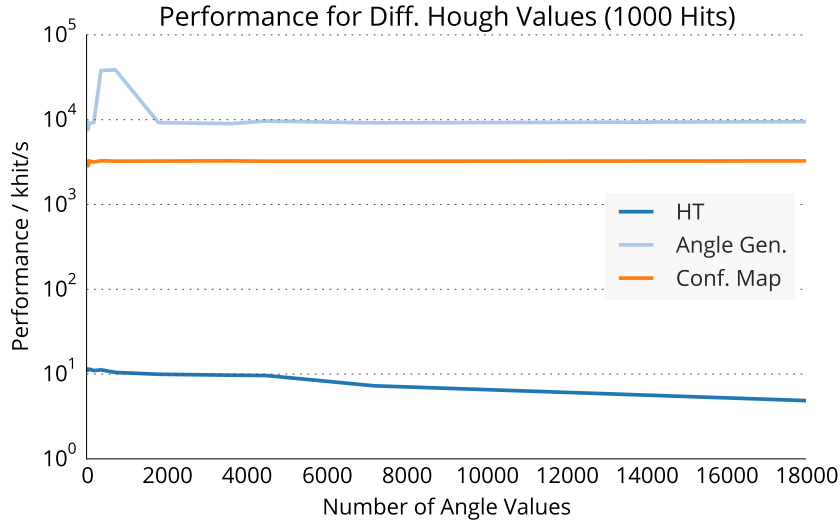
Figure 4.11: Performance in khit/s of the different steps of the Thrust-based Line Hough Transform. A GeForce GTX 750 Ti GPU is used with a fixed number of 1000 hits.

The normalized time for a conformal map of hits is smaller for larger sets of hits. Since the algorithm is parallel in the hits, the parallel GPU infrastructure is used more effectively. Notable is that the measured time is close to the reliably measurable threshold. The angle generation part of the algorithm is independent of the number of hits and executed much faster. Since Figure 4.10 displays the time *per hit,* the corresponding angle generation graph falls with increasing numbers of hits.
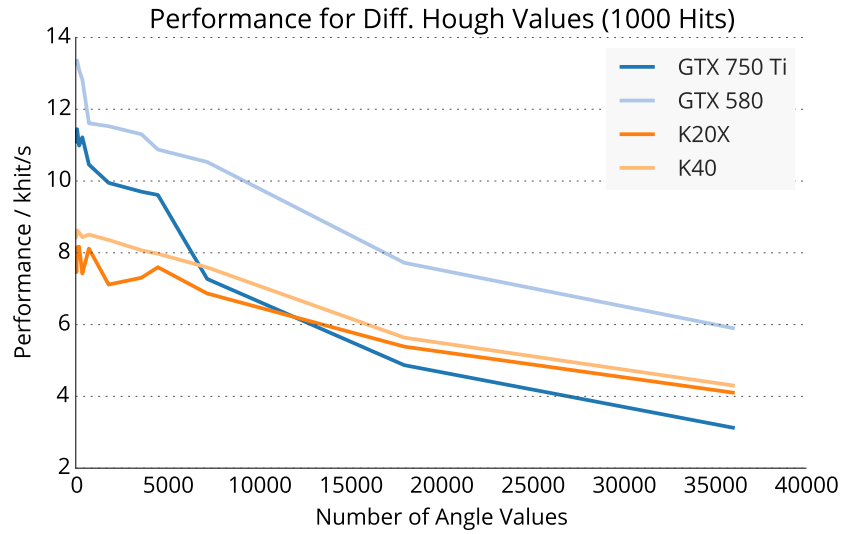
Figure 4.12: Comparison of performance of the Thrust LHT for different GPUs as a function of the number of angle values. The number of processed hits is fixed to 1000.

Different Number of Angles The Thrust LHT is parallel in the number of angles. A larger number of generated θ values leads to a larger number of computable ρ quantities. The performance for different numbers of θ values is benchmarked in Figure 4.11. While the HT part of the algorithm can process approximately 10,000 hits per second when evaluated at 360 θ values, it has only a little lower performance (8000 hit/s) for a 50 fold increase in θ numbers, as the algorithm is parallel in this quantity. The conformal map is parallel in the number of hits, leading to a constant distribution of this part in the graph. The performance of the angle generation is also constant, suggesting that this process does not fully utilized the offered parallelism of the GPU or is too small to measure reliably. The bump on the left is around 360 θ values results from a spot with beneficial data size and computation properties. In any case, the study of the angle granularity is a mere benchmarking of the achievable computational performance. 18,000 angle points are surely beyond the detector resolution. A realistic number of θ values is between 360 and 1080.

Different GPU Devices The total performance of the Thrust LHT is shown in Figure 4.12 for four different hardware cards in dependency of the number of θ values. Common to all cards is the decline in performance in sections. A large difference, though, is notable when comparing the HPC Tesla cards with the consumer-grade GTX cards; the latter ones are comparably fast or faster, over a large range of the studied angles. Especially the GeForce GTX 580 is approximately twice as fast as the Tesla K20X/Tesla K40. The high GPU clock rate seems to have beneficial properties for invoking the large number of automatically generated kernels by Thrust. The package uses heuristics to programmatically determine kernel configurations and launch parameters, leading to different settings for the different GPUs.

4.3.3.2   **CUDA Implementation**

The second implementation of the LHT is done directly in CUDA. It uses advanced GPU programming techniques like dynamic parallelism, has various kernels for the different steps of the algorithm, and is tuned for fast memory access. Dynamic parallelism is used to make the HT *adaptive*: In a first step, a coarse Hough space is filled. In subsequent steps, only bins with non-zero content are considered for further finer transformation; the LHT *zooms* into regions of interest and only calculates parameters more finely in these regions.

   The steps of the algorithm are:

Conformal Map All hits of an event are mapped. A kernel is called for all hit points with 256
      threads and number of blocks accordingly.[15]

Hough Transform A kernel is launched with 128 threads and a block size equivalent to the
      number of events. The Hough space is partitioned into 256 values, a Hough angle $\theta^{(1)}$ is
      calculated for each bin of the partition. The kernel computes the Hough value, $\rho^{(1)}$, for
      each hit of the current event for each angle. The parameter pair, $(\theta^{(1)}, \rho^{(1)})$, is filled into
      the accumulator array by means of CUDA's `atomicAdd()`. The bin of the accumulator array,
      in which $(\rho^{(1)}, \theta^{(1)})$ fall, is the basis for a next kernel call. The cell is partitioned into 256
      bins and the kernel itself is called again recursively on it, leading to effective bin sizes of
      $360°/(256 \times 256) \approx 0.005°$ in this second layer. The procedure is repeated, calculating
      new sampling angles, $\theta^{(2)}$, for new Hough values, $\rho^{(2)}$. Depending on the desired depth
      of the kernel invocation, the recursion stops and an array with track parameters is filled.

      In the current version, the recursion breaks after a depth of tow, as tracks of good
      precision are obtained. Adding another level would lead to effective angle steps sizes
      of $360°/(256^3) = 0.00002°$, which is beyond the measurement resolution of the sub-
      detectors.

   The *dynamic Hough transform* is programmed for $\overline{\text{P}}$ANDA by the NVIDIA Application Lab of
the Supercomputing Centre Jülich.

   Figure 4.13(a) shows the time for the dynamic LHT for different GPUs. Measured is the
whole »Hough Transform« part, as depicted in the description. The GeForce GTX 580 is not
included in the graph, as it does not support the dynamic parallelism feature, needed for the
algorithm. A comparable trend as for the Thrust-based LHT can be seen: The consumer-grade
GPU is faster than the HPC-grade GPUs. The performance, the number of hits processed per
second, is shown in Figure 4.13(b). For the full amount of studied hits, the GeForce GTX 750 Ti
is approximately 2 × faster than the Tesla K20X/Tesla K40. For small numbers of processed hits,
the distribution peaks to about twice the performance (see Figure A.2), due to fast memory
access for the low amount of data. Contrary to the Thrust approach, the performance is not yet
saturated, suggesting that the GPU still has unoccupied cores. Apart from further increasing the
number of parallel process hits, different techniques for increasing parallelism can be employed,
e.g. the bunching wrapper presented in Section 4.5.2.2.

---

[15]The number of blocks is $\lceil N_{\text{hit}}/256 \rceil$.

(a) Time. **(b)** Performance.
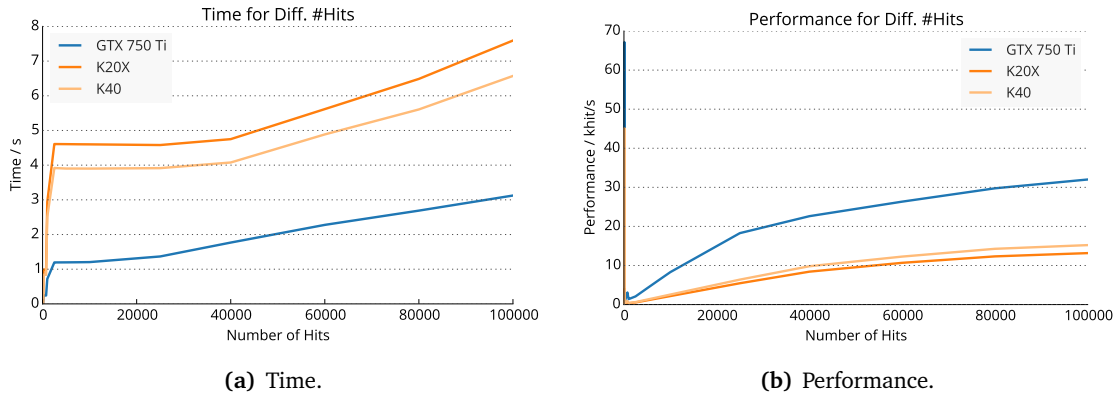
Figure 4.13: Run time and performance for the adaptive Line Hough Transform for different GPU devices. See Figure A.2 for a version of **(b)** with a logarithmic axis along x.

4.3.3.3 Comparison and Discussion

For the Line Hough Transform, two different GPU implementations with basically different approaches are presented. The Thrust approach shields the user from explicit usage of kernels. No kernel invocation parameters need to be chosen, no interior kernel structures programmed. The HT is relying solely on the usage of built-in Thrust functions, which operate on device-side data by means of iterators. In the second approach, a more *classical* CUDA application is used, with advanced features of recursively calling kernel structures.

One benefit of using Thrust for a parallel algorithm is its easy accessibility. The functions mimic STL functions and have a low threshold for productive programming. When complicated data types and data operations are used, the benefit quickly vanishes, and a lot of time is consumed to fit the problem into the architecture of Thrust. More severely, Thrust does not allow for arbitrary parallelism on data; extending the per-angle parallelism of the Thrust LHT to a per-hit parallelism can not be done easily, if at all. The initial offset for writing fully customizable CUDA code is rewarded quickly when the problem is of complex nature. Also, only this approach allows for fine-tuning and optimization of many details of the program. The anticipated high performance of the Thrust approach is limited by its scope of parallelism; the CUDA approach is approximately a factor ten times faster and is still not performance-saturated.

From a programming platform view, the CUDA approach is the favorable solution. Thrust can be used for very specific, simple tasks, where a pre-defined function can be exploited. With the dynamic HT, the Line Hough Transform as presented in Section 4.3.2 can be performed with a rate of up to 30,000 hit/s, including accumulator array filling and track parameter extraction. Further performance increases are expected by adding additional means of parallelism.

4.4 Circle Hough Transform

A complementary HT to the Line Hough Transform from Section 4.3 is the Circle Hough Transform (CH). It builds on the principles presented in Section 4.3.1, sampling track values and finding the best-fitting one as the highest-voted bin of an accumulator array. Instead of
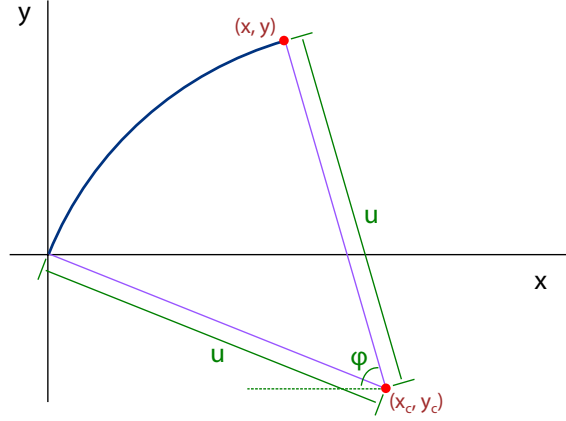
Figure 4.14: The parameters used for the circle determination of the Circle Hough Transform.

sampling a line equation, different circles as track representations are sampled directly. The pre- and post-processing step involving conformal map is avoided. The presented Circle Hough Transform (CH) is capable of reconstructing tracks originating from the interaction point at $(0, 0)$. The basic assumption of $(0, 0)$ as a fix point laying on the circle greatly reduces the complexity of the algorithm. A more general CH can be imagined, not constrained by this approach (see Section 4.4.2.3).

In the course of the this section, a general formulation of the CH is given first, then differentiations for $\overline{\text{P}}$ANDA are introduced. The CH is benchmarked for its capabilities of reconstructing tracks and physical events. In addition, the performance of the current GPU implementation of the algorithm is shown.

4.4.1 Algorithmic Description

Two different methods of sampling the parameter space of the CH can be noted. The first, initial approach samples an angle in close reference to the LHT from Section 4.3. Building up on this basis, a second approach samples directly certain p_t values, giving immediate access to a track's momentum. In the following, the general CH is derived first, before the p_t version is introduced subsequently.

4.4.1.1 General Circle Hough Transform

Hit points with finite 2D spatial information, e.g. MVD or GEM hit points, are considered differently in the CH than isochronous hit points from the STT. Similar to Section 4.3.2.2, first equations for point-like hit points are derived, then the extended hit points of the STT are considered.

Point-like Hit Points The CH assumes that tracks originate from the interaction point $(0, 0)$. A circle going through both $(0, 0)$ and a hit point at (x, y) can be parameterized by the center of the circle, (x_c, y_c), and the radius. The radius of the circle is directly defined by the distance

of either (x, y) to (x_c, y_c) or $(0, 0)$ to (x_c, y_c). The CH then computes x_c and y_c as a function of an angle φ[16]. The setup with different relations outlined is shown in Figure 4.14.

The position of the circle center, (x_c, y_c), can be found:

$$\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + u \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} \quad , \tag{4.27}$$

with the radius of the circle u and the angle φ as defined in Figure 4.14. For a given circle center point, the radius is $u = \sqrt{x_c^2 + y_c^2}$. Taking the distance to $(0, 0)$ from Equation 4.27 yields

$$\left| \begin{matrix} x_c \\ y_c \end{matrix} \right| = \sqrt{(x + u \cos \varphi)^2 + (y + u \sin \varphi)^2} = \sqrt{x_c^2 + y_c^2}.$$

Inserting $\sqrt{x_c^2 + y_c^2} = u$, the equation can be simplified to get u:

$$u = \sqrt{(x + u \cos \varphi)^2 + (y + u \sin \varphi)^2} \tag{4.28}$$
$$\Rightarrow u^2 = (x + u \cos \varphi)^2 + (y + u \sin \varphi)^2$$
$$= x^2 + 2xu \cos \varphi + u^2 \cos^2 \varphi + y^2 + 2yu \sin \varphi + u^2 \sin \varphi$$
$$\Rightarrow -x - y = 2u (x \cos \varphi + y \sin \varphi)$$
$$\Rightarrow u = -\frac{x + y}{2 (x \cos \varphi + y \sin \varphi)}$$

Adding indices, a circle going through $(0, 0)$ and a hit point (x_i, y_i), can be parameterized by

$$\begin{pmatrix} x_c^{ij} \\ y_c^{ij} \end{pmatrix} = u_{ij} \begin{pmatrix} \cos \varphi_j \\ \sin \varphi_j \end{pmatrix} + \begin{pmatrix} x_i \\ y_i \end{pmatrix} \quad \text{with} \tag{4.29}$$
$$u_{ij} = -\frac{x_i^2 + y_i^2}{2 (x_i \cos \varphi_j + y_i \sin \varphi_i)},$$

with the circle center (x_c^{ij}, y_c^{ij}) for angle φ_j and hit point coordinates (x_i, x_i). The two center values are the values filled into the Hough space, while φ is sampled for $\varphi \in [0°, 360°)$. The sampling extends to 360° since circles opening in both directions, equivalent to tracks of positively and negatively charged particles, need to be considered.

Extended Hit Points (Circles) When considering hits from the STT, the isochrone radius r is an additional coordinate to x and y. The constraints for the circle is slightly modified: The circle goes through $(0, 0)$ but is now <u>tangential</u> to the isochronous hit point in (x, y, r). As shown in Figure 4.15, the distance from the circle center at (x_c, y_c) to the hit point's center at (x, y) is extended by the amount of the drift radius. The distance from the circle center to the origin and the circle center to the hit point center (wire position) is not equal any more.

---

[16]φ is chosen as the CH sampling parameter to not confuse it with θ from the LHT.
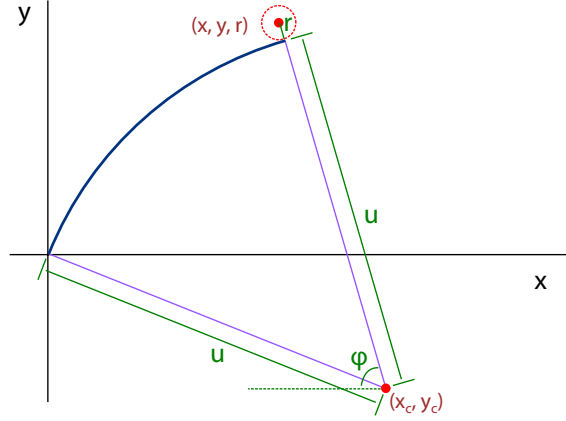
Figure 4.15: The parameters used for the circle determination of the Circle Hough Transform around an isochronous hit point, e.g. from the STT.

u needs to be extended on the hit point's side by r. Using the definition of u from Equation 4.28:

$$u = \sqrt{\left(x + u\cos\varphi\right)^2 + \left(y + u\sin\varphi\right)^2} + r$$

$$\Rightarrow (u-r)^2 = x^2 + y^2 + u^2 + 2u(x\cos\varphi + y\sin\varphi)$$

$$\Rightarrow -2u(x\cos\varphi + y\sin\varphi + r) = x^2 + y^2 - r^2$$

$$\Rightarrow u \equiv u_\odot = \frac{r^2 - x^2 - y^2}{2\left(x\cos\varphi + y\sin\varphi + r\right)}.$$

Hence, the CH's sampling equation for circle center coordinates of circles going through $(0,0)$ and being tangential to (x_i, y_i, r_i) is

$$\begin{pmatrix} x_c^{ij} \\ y_c^{ij} \end{pmatrix} = u_\odot^{ij} \begin{pmatrix} \cos\varphi_j \\ \sin\varphi_j \end{pmatrix} + \begin{pmatrix} x_i \\ y_i \end{pmatrix} \quad \text{with} \tag{4.30}$$

$$u_\odot^{ij} = \frac{r_i^2 - x_i^2 - y_i^2}{\mp 2\left(x_i\cos\varphi_j + y_i\sin\varphi_j \pm r_i\right)}, \tag{4.31}$$

with the circle center coordinates (x_c^{ij}, y_c^{ij}) for a hit point of index i and angle of index j. The $\pm$ sign in front of r_i accounts for both cases of circles tangential to the hit point: one case for the isochrone side close to the circle center, one case for the isochrone side further away from the circle center.

4.4.1.2 Momentum-Dependent Circle Hough Transform

Equation 4.30 defines circles as a function of a sampled angle, $(x_c, y_c)(\varphi)$. This approach is straight-forward and keeps the CH similar to the LHT. Considering circles instead of lines with circle centers as parameters of the Hough space leads to one essential drawback. An equidistant angle sampling step accounts for unequal spatial distributions of circle centers. For sampled

angles which are shallow with respect to the coordinate origin, circle centers are generated which lie very far away from the origin. A large range of Hough parameters is generated, filling a Hough space unequally. Physically, this leads to tracks with high transverse momentum to be covered only with a coarse sampling resolution, while low-p_t tracks have more finely grained sampling.

A way to better cover the large dynamic range is to sample the circle centers not as a function of φ, but directly as a function of p_t. Further benefits include the gain in accessibility of the method: instead of sampling an abstract and arbitrary angle value, the to-be-reconstructed momentum of the track is directly input. The procedure is as follows: For a given value of p_t, the radius of the corresponding circle is computed. Based on the radius, φ values for the hit point are computed. The φ values are input into the Hough transform Equation 4.30.

Momentum to Radius   A particle with transverse momentum p_t forms a track described by a circular arc in the transverse xy-plane. Neglecting all deflection effects (multiple scattering, energy loss, ...), the different circle parameters for different parts of the arc of the track can be approximated by a single circle. The relation between p_t and the radius of this approximated circle R yields from the equality of the Lorentz force F_L and the centripetal force F_c:

$$\overrightarrow{F_L} = q\vec{v}\vec{B}, \qquad F_c = \frac{mv^2}{R} = \frac{pv}{R}$$

$$|F_L| = F_c$$

$$\Leftrightarrow qvB = \frac{pv}{R}$$

$$\Leftrightarrow R = \frac{p}{qB},$$

with the momentum p, the magnetic field strength B, and the charge q, giving the bending direction of the circle. For particles of unit charge and homogeneous magnetic fields, the equation can be written as:

$$R/\text{cm} = \frac{100}{0.3 B/\text{T}} p_t/\text{GeV}/c. \tag{4.32}$$

The units of the quantities are specified.

Radius to Angles   A circle of radius R, equivalent to a specific transverse momentum p_t as of Equation 4.32, can have four sets of parameters for a given isochronous STT hit point, as illustrated in Figure 4.16. Depending upon the sign of the charge ($\pm q$) a track can be on either the outside or the inside of the isochrone. The four φ values, that are compatible with the isochrone, are derived in the following when solving Equation 4.31 for φ[17]:

---

[17]For simplicity, only the $+r$ part is taken for the derivation. The second side of the isochrone is added at the end.
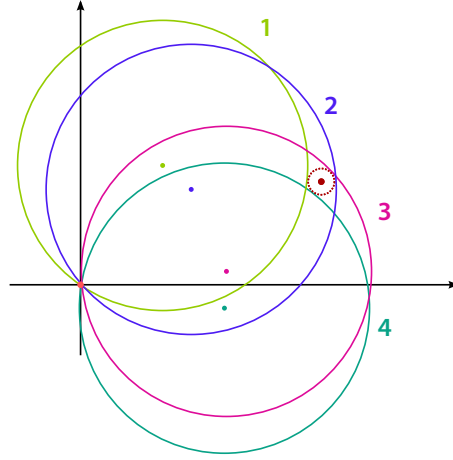
Figure 4.16: Visualization of the four possible circle center locations of a circle of fixed radius R around an isochronous hit.

$$u = \frac{r^2 - x^2 - y^2}{2\left(x \cos\varphi + y \sin\varphi \pm r\right)}$$

$$\Rightarrow x \cos\varphi + y \sin\varphi = \frac{r^2 - x^2 - y^2 - 2ur}{2u} \tag{4.33}$$

For the left side, a trigonometric relation can be used:

$$x \cos\varphi + y \sin\varphi = \sqrt{x^2 + y^2} \begin{cases} \sin\left(\varphi + \arctan\left(\frac{x}{y}\right)\right) & \text{, for } y > 0 \\ \cos\left(\varphi - \arctan\left(\frac{y}{x}\right)\right) & \text{, for } x > 0 \end{cases}$$

With this, Equation 4.33 is split for two cases

$$y > 0: \ \sin\left(\varphi + \arctan\left(\frac{x}{y}\right)\right) = \frac{r^2 - x^2 - y^2 - 2ur}{2u\sqrt{x^2 + y^2}}$$

$$x > 0: \ \cos\left(\varphi - \arctan\left(\frac{y}{x}\right)\right) = \frac{r^2 - x^2 - y^2 - 2ur}{2u\sqrt{x^2 + y^2}},$$

which can finally be solved for φ

$$y > 0: \ \varphi_1 = \arcsin\left(\frac{r^2 - x^2 - y^2 - 2ur}{2u\sqrt{x^2 + y^2}}\right) - \arctan\left(\frac{x}{y}\right)$$

$$x > 0: \ \varphi_2 = \arccos\left(\frac{r^2 - x^2 - y^2 - 2ur}{2u\sqrt{x^2 + y^2}}\right) + \arctan\left(\frac{y}{x}\right).$$

Summarizing the equations, adding indices for bookkeeping, and accounting for both isochrone sides, one finds:

$$\varphi_{1,3} = \arcsin\left(\frac{r_i^2 - x_i^2 - y_i^2 \mp 2u_{ik}r_i}{\pm 2u\sqrt{x_i^2 + y_i^2}}\right) - \arctan\left(\frac{x_i}{y_i}\right), \text{ for } y_i > 0$$

$$\varphi_{2,4} = \arccos\left(\frac{r_i^2 - x_i^2 - y_i^2 \mp 2u_{ik}r_i}{\pm 2u\sqrt{x_i^2 + y_i^2}}\right) + \arctan\left(\frac{y_i}{x_i}\right), \text{ for } x_i > 0,$$

(4.34)

with index i for a certain hit and k for a radius of a certain momentum, with $u_{ik} = R_k + r_i$, where $R_k = R_k(p_t^k)$ is given by Equation 4.32. To account for hit points in the negative quadrants, $x_i < 0$ and $y_i < 0$, the φ_j values have to be corrected by π: $y_i < 0 \rightarrow \varphi_{1,3}^* = \varphi_{1,3} + \pi$ and $x_i < 0 \rightarrow \varphi_{2,4}^* = \varphi_{2,4} + \pi$.

The four φ values per hit and momentum can then be directly inserted into Equation 4.30 to obtain circle centers for filling into the Hough space. For MVD and GEM hit points, the according equations are obtained when the isochrone radius is set to zero, $r_i = 0$. $\varphi_{1,3}$ is reduced to φ_1 and $\varphi_{2,4}$ to φ_2, since not two isochrone sides need to be considered.

Momentum Limit Choosing Equation 4.34 to sample the momentum range instead of the φ range has two distinct advantages: First, the momentum range is sampled in equidistant steps, filling the Hough space uniformly. Second, it allows the momentum to be restricted to *relevant* ranges. For $\overline{\text{PANDA}}$, the maximum transverse momentum of a particle depends on HESR's beam momentum. For the full beam energy, the theoretical upper limit of the transverse momentum is $p_t^{\max} = \sqrt{s}/2 = 2.735\,\text{GeV}/c$. The momentum yields a maximum track circle radius of:

$$R(p_t^{\max} = 2.735\,\text{GeV}/c) = \frac{100}{0.3 \times 2}5.47$$
$$\approx 456\,\text{cm} = R_{\max}.$$

$R_{\max}$ constraints the Hough space. In both x and y direction the space extends to $\pm 456\,\text{cm}$.

4.4.1.3 Procedure

In Figure 4.17 the method of the CH is depicted. The track with the lowest possible momentum creates a circle which has the center exactly at the halfway position between the hit point and the coordinate origin. Tracks with higher momenta create larger circles with center points further outside. For a given track radius, there are always four possible circle centers. Two on the *inside*-facing surface of the isochrone (green in Figure 4.17(a)), and two on the *outside* surface (purple in Figure 4.17(b)), with respect to the coordinate origin. Per side, there is a circle for positive-charged tracks, and a circle for tracks with negative charge. The four different possibilities are shown in Figure 4.16.

Sampling through the φ values and noting the circle center values, two loci of points emerge for the STT, Figure 4.17(d); the outer set of points for the circles tangent to the outside of the isochrone, the inner set of points for inside-tangential circles. The loci form two hyperbolas
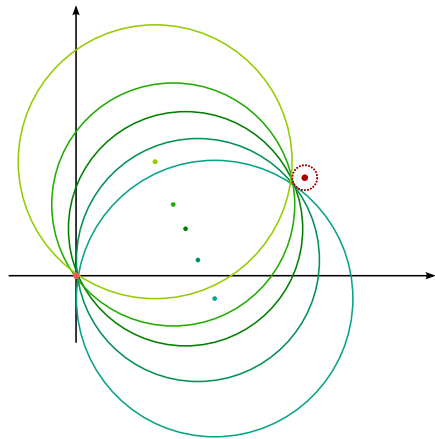
(a) Circle centers on the isochrone inside for low-momentum tracks.
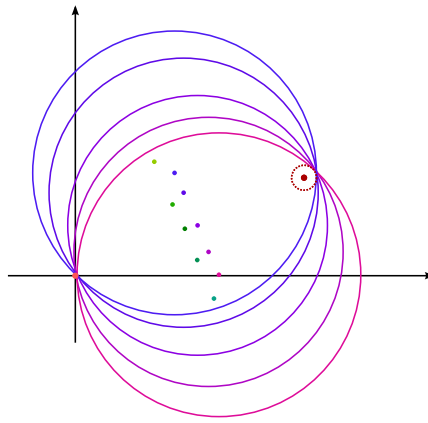
(b) Circle centers on the isochrone outside for low-momentum tracks.
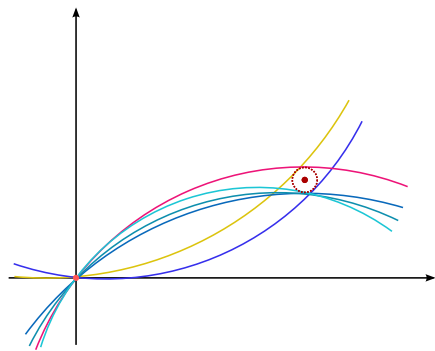
(c) Segments of circles from higher momentum tracks. The center points are out of the frame.
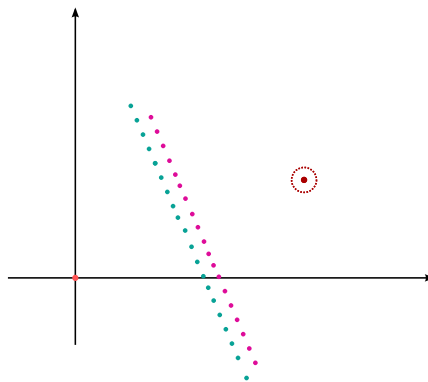
(d) Different circle centers. For circles going through the inner side of the isochrone (blue) and for the outer side (red).

Figure 4.17: Schematic sketch of the procedure of the CH for constructing circle centers for a isochronous hit point (dark red, with circle around).

(a) Larger range. The two bands of the center points are visible.

(b) Zoom around the center of the hit point.
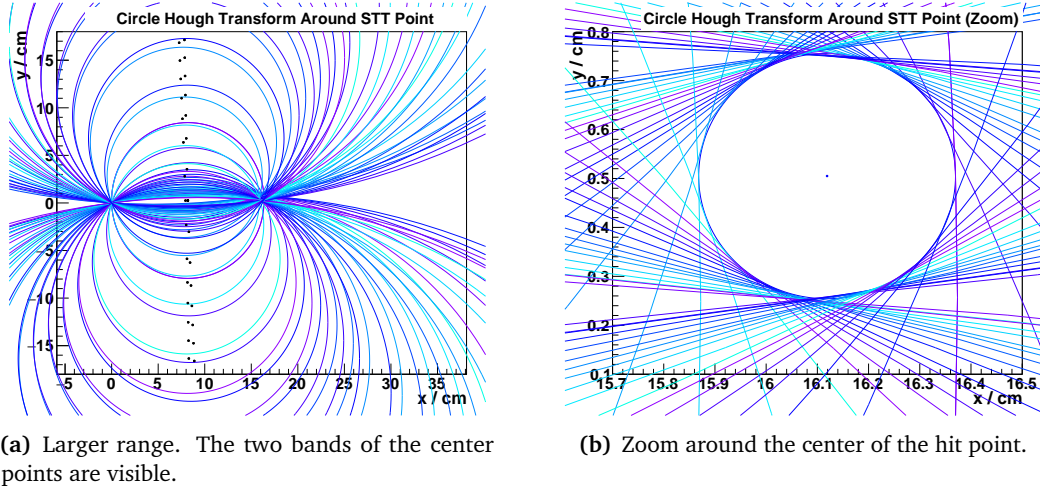
Figure 4.18: Visualizations for the Circle Hough Transform. Circle centers (black) for circles of different radii. **(b)** is a zoom around the isochronous hit point from **(a)**. Colors of circles are chosen to simplify differentiating them.

opening to two different sides. For MVD or GEM hit points, the isochrone radius is $r = 0$ and the two-line pattern reduces to a single straight line. The line is perpendicular and bisects the line connecting the hit point with the origin.

An exemplary transform for a STT hit point is shown in Figure 4.18. In Figure 4.18(a) it can be seen that the scope of the sampled p_t range is limited; to the right of the plot, a space without circle arcs emerges. The space would be filled by tracks with unphysically large transverse momenta. Figure 4.18(b) highlights the transform around an isochrone radius.

4.4.1.4 Hough Space

The pictures of Figure 4.19 are generated by sampling φ and filling every obtained circle center into the Hough space. The data shown is for the hits of the event already introduced in Figure 4.4, an event with comparable small p_t. Restraining the assumed p_t for smaller values, $p_t \in [0\,\mathrm{GeV}/c, 1\,\mathrm{GeV}/c]$, yields Figure 4.19(b). Here, different overlapping regions emerge, equivalent to the circle centers of the tracks. In combination with the position of each hit, which gives the bending radius of the track, the circle centers are used to fully define a track. To account for empty bins between two consecutive φ steps, the Hough space is interpolated linearly; the bins of the Histogram, lying on a straight line between two filled Hough values are also filled.

4.4.2 Performance Measurements

The CH is a relatively new algorithm which has been developed during the course of this thesis. In this section, the algorithm is first benchmarked in terms of its performance to reconstruct tracks and physical events. Then, the performance of the GPU implementation is shown.

(a) Large range, $p_t = 0\,\text{GeV}/c$ to $p_t = 2.74\,\text{GeV}/c$.   **(b)** Zoom to $p_t \leq 0.6\,\text{GeV}/c$.
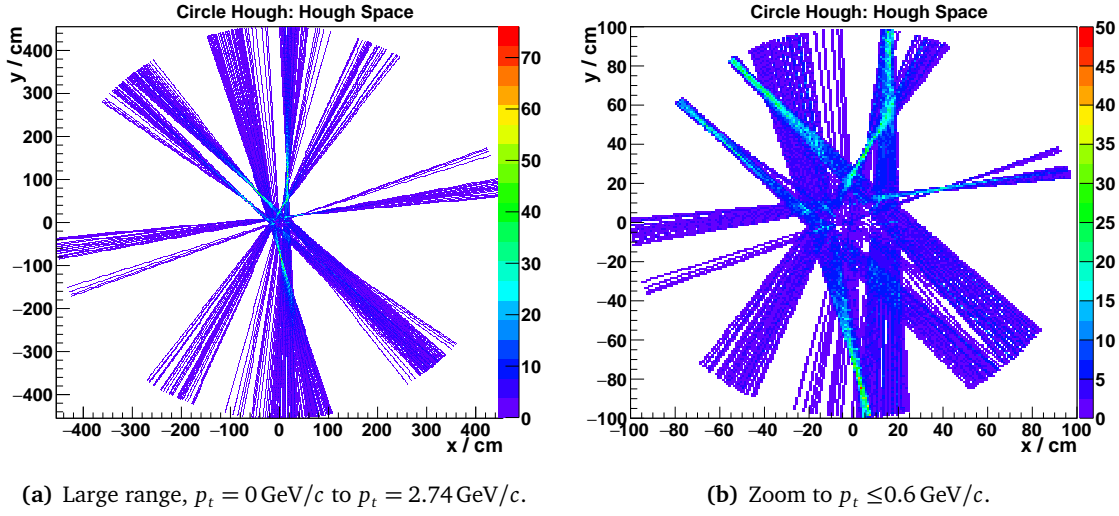
Figure 4.19: Hough spaces of the Circle Hough Transform for the tracks shown in Figure 4.4(a). In **(a)**, the full transverse momentum range of $0\,\text{GeV}/c$ to $2.74\,\text{GeV}/c$ is shown, **(b)** shows a zoom from $0\,\text{GeV}/c$ to $0.6\,\text{GeV}/c$.

Table 4.2: Overview of reconstruction efficiency for the Circle Hough Transform. *All* refers to the full number of reconstructed tracks, including particles produced in scattering through matter.

	ε / %	
	Circle Hough	$\overline{\text{P}}$ANDA Offline Tracking
All	90	92
K	77	79
π	90	92

4.4.2.1 Reconstruction Capabilities

The CH, as presented in Section 4.4.1.2, is tested for its reconstruction capabilities. The physics channel of this thesis, $\overline{\text{p}}\text{p} \rightarrow \text{D}^+\text{D}^- \rightarrow \text{K}^-\pi^+\pi^+\text{K}^+\pi^-\pi^-$ (see Chapter 5), is used as the example reconstruction channel. Although the D decay has a non-negligible length and creates the daughter particles from a displaced vertex, the vertex is sufficiently close to the interaction region for the CH to reconstruct the tracks. If needed, the displacement can subsequently be reconstructed using vertex fitting.

In the following section, first, the reconstruction quality of individual tracks is analyzed. After this, the algorithm's abilities for reconstructing whole events in an online environment is investigated.

Track Reconstruction   The CH is investigated for its performance in reconstructing tracks. To quantify relations to the simulated original track, matching to the true track, *MC matching* is performed.
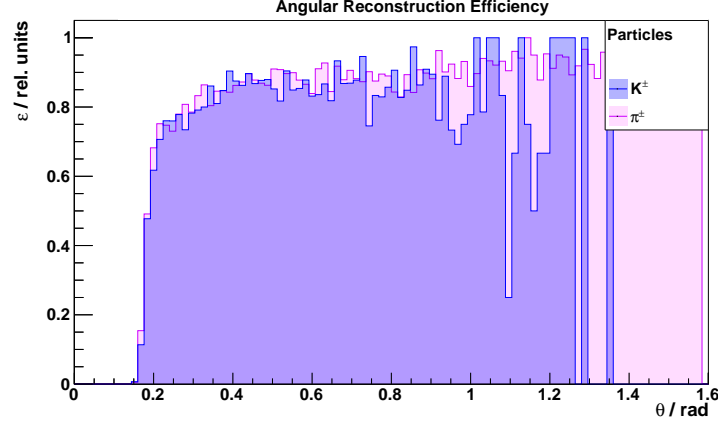
Figure 4.20: Track reconstruction efficiency as a function of the polar angle for pions (purple) and kaons (blue). The efficiency is given with respect to reconstructible tracks.

Reconstruction Efficiency In total, the CH reconstructs particle tracks with 90 % efficiency. Pions can be reconstructed better (90 %) than kaons (77 %), as pions leave more hit points in the tracking sub-detectors. The full reconstruction efficiency of the CH is higher than the efficiency for kaon and pion reconstruction, as the algorithm also reconstructs particles from other sources, e.g. particles which are produced by scattering through matter. The efficiencies are summarized and compared to $\overline{P}$ANDA's offline track reconstruction in Table 4.2.

The reconstruction efficiency of the CH as a function of the polar angle is shown in Figure 4.20. The total efficiency is divided into efficiencies for reconstructed pion tracks and kaon tracks. Both efficiencies start rising rapidly for $\theta > 0.15$ rad. For values greater than 0.2 rad, the efficiencies are between 80 % and 90 %. While the kaon efficiency decreases slowly for increasing angles, with a lack of statistics for very large polar angles, the pion efficiency increases over the full θ spectrum to $\theta = \pi/2$. The reconstruction efficiency for pions is better than for kaon and covers a larger polar angle. The main reason for this effect is the higher number of hits from pions in the tracking sub-detectors.

The efficiency is measured not in relation to all tracks, but in relation to all *reconstructible tracks*. Normalizing to this number gives a realistic impression of the reconstruction capabilities of the algorithm, as tracks, which can not be reconstructed, are not considered. Defining a track to be reconstructible must be done with careful consideration: Choosing to tight discrimination parameters over estimates the reconstruction capabilities of the algorithm. On the other hand, choosing too loose criteria accepts too many tracks that can not be reconstructed by the algorithm. For the CH, the following criteria have been chosen to define a track as reconstructible. It is a primary track, it has at least three hits in the MVD, and it has at least six hits total in the MVD, STT, and GEM sub-detectors.

Figure 4.21 visualizes the importance of defining a sub-set of *reconstructible* tracks, by plotting the ratio of $N_{\text{reconstructed}}/N_{\text{reconstructible}}$ as a function of the MC true transversal momentum. Reconstructible is defined as noted in the previous section. The orange distribution is over 80 % for the whole p_t range, suggesting well-reconstructed tracks. The values close to 1 for
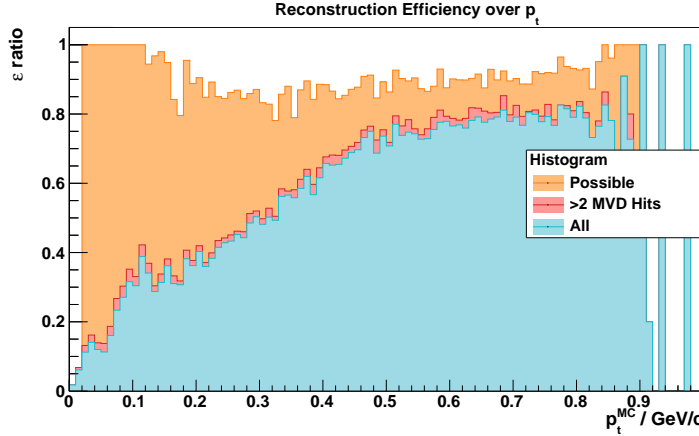
Figure 4.21: Reconstruction efficiency for all particles as a function of MC transverse momentum p_t. Different definitions of *reconstructible* tracks are shown.
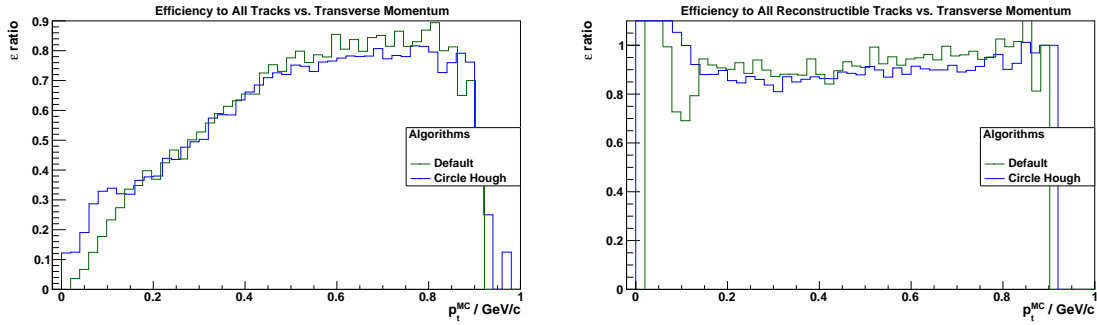
(a) Fraction of reconstructed tracks to all tracks as a function of p_t.

(b) Fraction of reconstructed tracks to reconstructible tracks as a function of p_t.

Figure 4.22: Comparison of reconstruction efficiencies of Circle Hough Transform (blue) and the offline track reconstruction of $\overline{P}ANDA$ (»default«, green) as a function of the track momentum. The excess over 1 in **(b)** is due to reconstruction of tracks which were considered not reconstructible.

small p_t originate from the small number of tracks falling into this range. The red distribution defines *reconstructible* differently: at least 3 hits in the MVD sub-detector are required. The distribution of the ratio of the number of reconstructed tracks to the number of reconstructible tracks is shaped differently. As the number of possible reconstructible tracks is increased, the fraction of tracks that have been reconstructed is reduced over the full p_t spectrum. Finally, the blue distribution shows the fraction of reconstructed tracks relative to all available tracks in this simulation channel.

The reconstruction efficiency of the CH is compared to that of $\overline{P}ANDA$'s offline track reconstruction in Figure 4.22. In Figure 4.22(a) the number of reconstructed tracks is normalized to all MC tracks in the event. In Figure 4.22(b) the number of tracks is normalized to the number of reconstructible tracks. Comparing the two tracking algorithms, a similar shape of the efficiency is seen. For most of the p_t range, the reconstruction efficiency of the offline track reconstruction is slightly superior. An inverse effect can be noted for tracks with very small

(a) Transverse momentum resolution.

(b) Relative transverse momentum resolution.
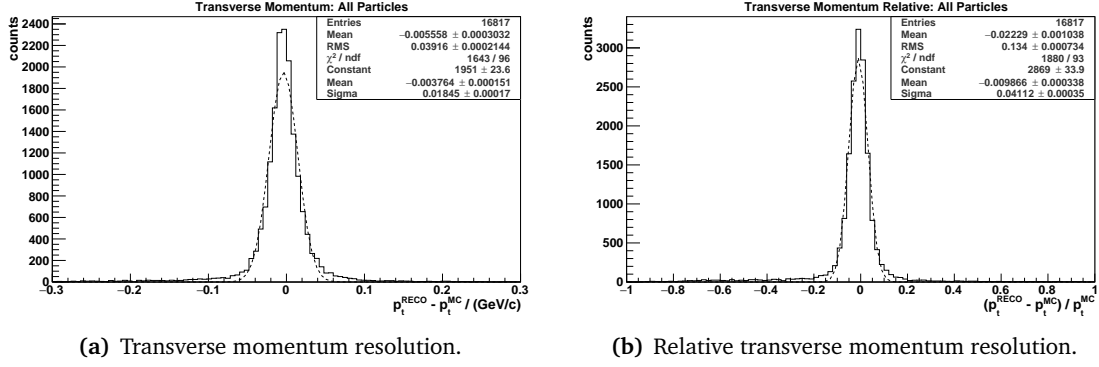
Figure 4.23: Transverse momentum resolution of reconstructed tracks, $\Delta p_t = p_t^{RECO} - p_t^{MC}$. The distributions include momenta from all particle species. Gaussian fits to the distributions are indicated by dashed lines.

(a) Δp_t for kaons.

(b) Δp_t for pions.
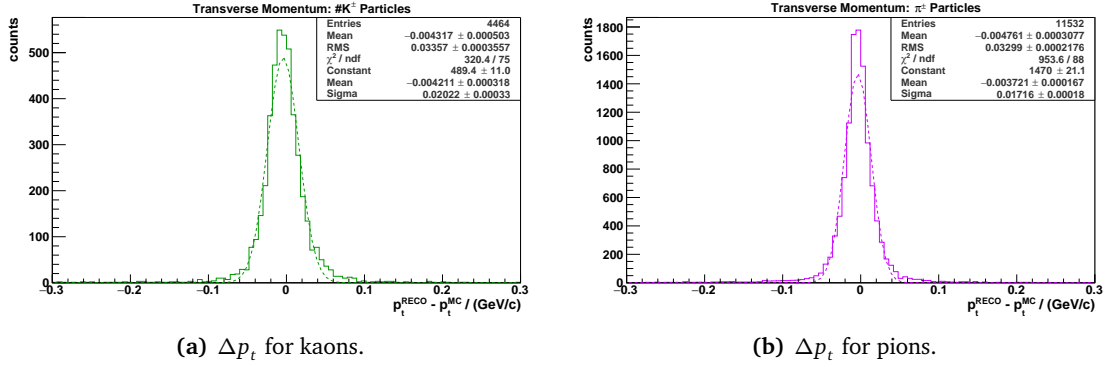
Figure 4.24: Track reconstruction quality split by particle species into K and π. The distributions of $\Delta p_t / p_t^{MC}$ are shown in Figure A.3.

$p_t < 0.2\,\text{GeV}/c$. Here, the CH reconstructs more tracks.

Transverse Momentum Resolution   In important quantity of track reconstruction algorithms is the resolution to reconstruct the momentum of the original particle, see also Section 4.2.4. Figure 4.23 shows the difference between the reconstructed transverse momentum, p_t^{RECO}, minus the transverse momentum of a particle's original track, p_t^{MC}. The mean transverse momentum resolution for K and π is determined to be $\sigma_{\text{tot}} = 19.6\,\text{MeV}/c$ (Figure 4.23(a)) or $\sigma_{\text{rel}} = 4.1\,\%$ (Figure 4.23(b)). The measured quantities of the reconstruction resolution are summarized for this graph in Table 4.3. In the table, data is also quoted for the same quantities from the $\overline{\text{P}}$ANDA offline track reconstruction[18]. The offline reconstruction runs with a Kalman track fitting step. Since an online track reconstruction should reconstruct tracks without a Kalman filter, numbers are given as well for $\overline{\text{P}}$ANDA's offline track reconstruction without a Kalman filter.

---

[18]The offline tracking is that included in version scrut14 of PandaRoot.

Figure 4.24 shows the reconstruction capabilities divided into the two particle species. The resolution for pions ($\sigma_{tot} = 18.2\,\text{MeV}/c$) is slightly better than that for kaon ($\sigma_{tot} = 21.6\,\text{MeV}/c$). Pions have more hits points in the tracking detectors and fill a greater θ range (see also Figure 5.14 and Figure 5.15). Analysis of tracks of positively and negatively charged particles shows that the CH reconstructs both charge types with equal resolution and efficiency (see Figure A.4).

A reconstructed track can be put into one of three categories: it can be **fully reconstructed**, **partially reconstructed**, and **spuriously reconstructed**.

Fully Reconstructed All hit points in a reconstructed track come from one MC true particle, and all hits of the MC true particle are in one reconstructed track.

Partially Reconstructed All hit points in a reconstructed track come from one MC true particle, but not all MC hits are in one reconstructed track.

Spuriously Reconstructed Most hit points of one reconstructed track come from one MC particle. The value of *most* must be defined for the analysis. The discrimination value is studied later in this section. Here, a spurious track must have $\geq 70\,\%$ true hit content, otherwise it is considered not to be reconstructed.

Using these definitions, the three plots of Figure 4.25 are obtained. As expected, the fully reconstructed tracks best reconstruct the original momentum ($\sigma_{rel} = 3\,\%$). For these tracks, no hits from other particles contaminate the reconstruction and pull the reconstructed parameters away from the original values. Partially and spurious reconstructed tracks have about one or a half percentage point worse resolution, respectively. Only about 20 % of the tracks are reconstructed fully, spurious and partially found tracks are found equally frequent. While reconstructing clean tracks with a good resolution should be a prime objective of an tracking algorithm, the CH's reconstruction capability for contaminated tracks is still good, as only a small decrease in resolution is observed.

Degree of Contamination The degree of contamination is the parameter that defines how much admixture of hits from other tracks is allowed to be in a reconstructed track. Figure 4.26 illustrates the distribution of tracks with different *spuriousness* fractions, the fraction of the number of hits in a reconstructed spurious track which come from the associated original MC track. The gaps in the distribution come from the discreteness and small value of the number of hits in a reconstructed track and number of hits in a MC track. A rise in this distribution to larger fractions is notable.

The quality of tracks with certain associated spuriousness fractions is shown in Figure 4.27(a) as a function of the relative transverse momentum resolution. The mean value and the spread along the y-axis of each fraction bin can be seen in Figure 4.27(b). The error bars represent the standard deviations of the mean along the x-axis. The deviation is largely constant over the spectrum, with a slight trend to smaller values for tracks with higher purity. Although the multiplicity of entries is larger for higher fractions, also a non-negligible number of tracks have spuriousness fractions between 0.5 and 0.8. Choosing a low discrimination value (e.g. 0.5)

(a) *Fully* reconstructed tracks.

(b) *Partially* reconstructed tracks.

(c) *Spurious* reconstructed tracks.
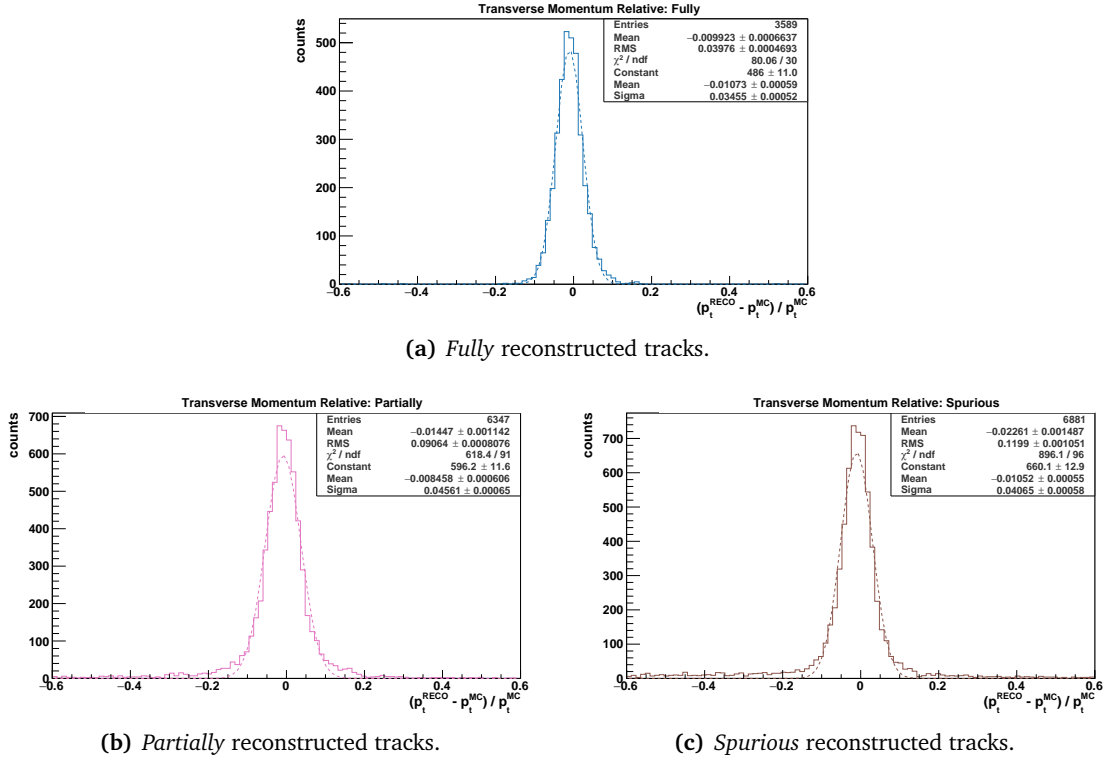
Figure 4.25: Track reconstruction resolution for the three purity categories, fully reconstructed, partially reconstructed, and spurious reconstructed.
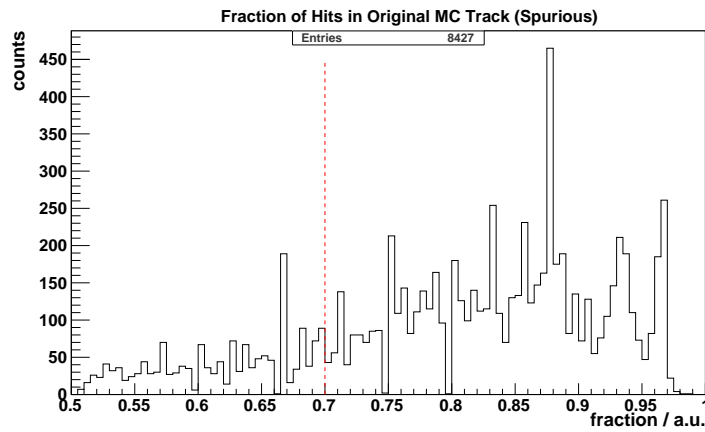
Figure 4.26: Distribution of the fraction of hits in a reconstructed track that come from one original MC particle. The red line defines the minimum value for a track to be accepted as a *spurious* track.

(a) $\Delta p_{t,\text{rel}}$ as a function of the spurious fraction

(b) Profile histogram of **(a)**.
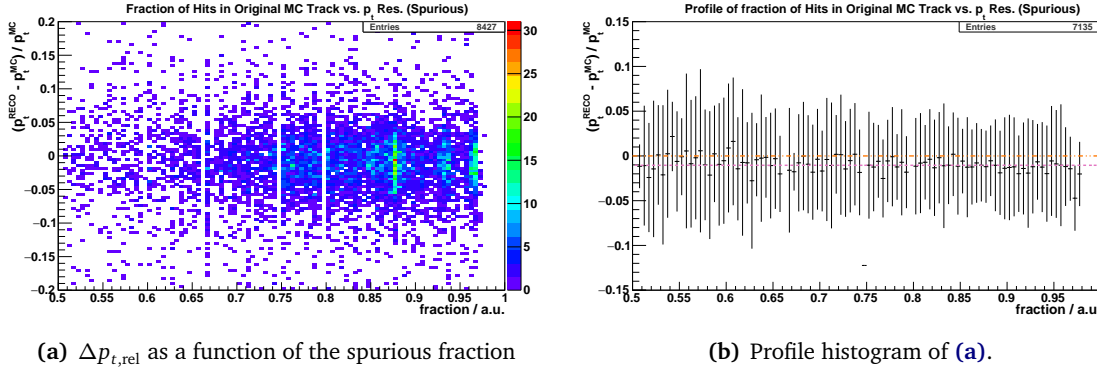
Figure 4.27: The spurious discrimination fraction in comparison to the relative transverse momentum resolution of a track. The purple line in **(b)** denotes the (weighted) mean along the x-axis. The orange line is located at $y = 0$ to guide the eye.

increases the efficiency of the track reconstruction, since in total more tracks are reconstructed, but decreases the purity and resolution of the reconstructed tracks slightly. For an online track reconstruction algorithm, the number should not be taken chosen too strictly. A parameter of 0.7 is used as the working point to match the requirements of the CH.

Figure 4.27(b) displays another feature of the spurious reconstructed tracks by the CH: their relative momentum is not distributed around zero but systematically displaced. The mean displacement (purple line) is located at $y = -1.0\,\%$. The reconstructed p_t of spurious tracks has a tendency to be too low in comparison to the original track. A possible explanation is energy loss of particles in the detector. This feature is not considered in the CH, as it would complicate the algorithm both algorithmically and computationally. While this effect is not a serious reconstruction limitation, it should be monitored closely and considered in a full online event reconstruction.

Event Reconstruction   The CH appears to be able to reconstruct particle tracks with reasonable quality. The capability of the algorithm to produce tracks, which can be combined to composite particles in an online environment, is benchmarked in this section. The offline reconstruction capabilities of the CH are shown in Section 5.5. It is necessary to use fast algorithms for event building in an online software trigger, like the one $\overline{\text{P}}$ANDA will employ. If possible, the comprehensive offline event reconstruction chain in PandaRoot should be avoided. Currently, no software for online event building exists in the software framework.

In the following, an impression of the performance of the CH for creating events online is given. The simple online-like event reconstruction, shown here, solely relies on information provided by the tracking algorithm and uses no further PID information. It is a coarse formulation of a lower limit of the event building performance.

Combinatorial Assembly   The basic approach of this online-like, tracking-only event building lies in combining all tracks of certain charges to D meson candidates. To reconstruct a D^+ (which decays to $K^-\pi^+\pi^+$), the algorithm initializes with the negative charged track as the

Table 4.3: Momentum resolution of tracks reconstructed by the CH and of $\overline{\text{P}}$ANDA's offline reconstruction (with and without Kalman filter). *N* is given in relation to the *all* tracks, 16,817 for the CH and 8594 for the $\overline{\text{P}}$ANDA offline tracking. The *missing* percentages to 100 % come from secondary reconstructed particles, which are part of *all*, but are neither K nor π.

	Circle Hough		$\overline{\text{P}}$ANDA Offline Tracking		
			w Kal.	w/o Kal.	
	N	$\Delta p_t/p_t$	N	$\Delta p_t/p_t$	
	%	%	%	%	
All	100	4.12 ± 0.04	100	2.03 ± 0.03	5.30 ± 0.08
K	27	4.68 ± 0.07	27	2.30 ± 0.06	5.52 ± 0.01
π	68	3.87 ± 0.04	69	1.91 ± 0.03	5.14 ± 0.09
$K^+ \pi^+$	47	4.06 ± 0.05	48	2.05 ± 0.04	5.21 ± 0.11
$K^- \pi^-$	48	4.09 ± 0.05	48	1.98 ± 0.04	5.26 ± 0.11
Fully	21	3.50 ± 0.05	51	1.89 ± 0.03	4.13 ± 0.06
Partially	38	4.61 ± 0.07	12	2.11 ± 0.07	13.69 ± 0.60
Spurious	41	4.11 ± 0.06	37	2.76 ± 0.12	7.53 ± 0.22
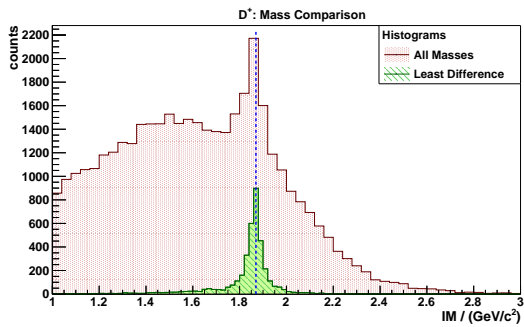
(a) Invariant mass distribution of D^+ candidates.
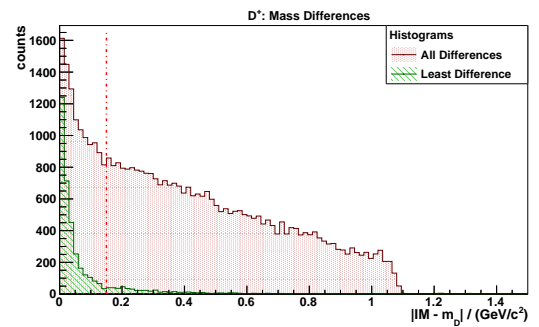
(b) Absolute mass difference distribution to m_D.

Figure 4.28: Candidate selection for online event building. The green distributions are the masses of D candidates, which have the least difference to the nominal D mass per event.

(a) Mass distribution of D$^+$ candidates.

(b) Mass distribution of D$^-$ candidates.
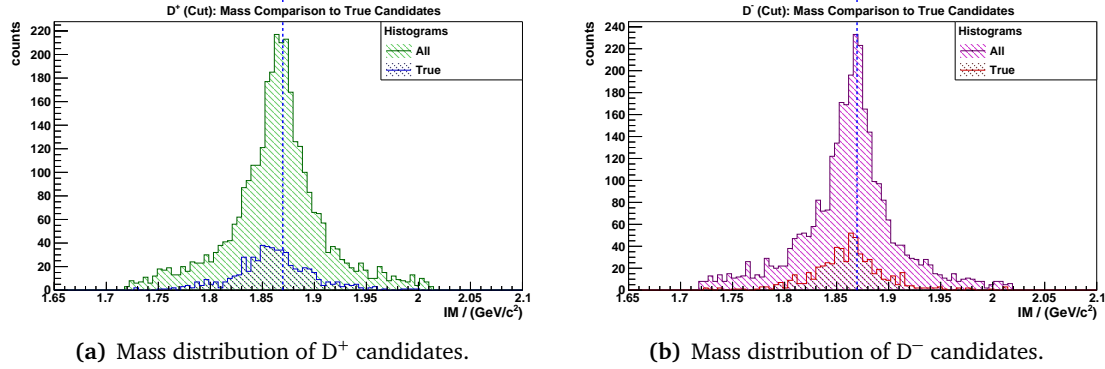
Figure 4.29: Comparison of masses of D$^\pm$ candidates with the least different mass to the nominal D mass, compared with MC-matched candidates. For 4000 D$^+$ and D$^-$ events.

kaon hypothesis. All combinations of two positive tracks in the current event are associated to this K$^-$, as the pions always have the opposite charge compared to the kaon. Using the momenta provided by the CH and the K and π mass hypotheses, the four-vectors of the particles can be constructed. The D$^+$ is the sum of the three daughter particles. The procedure is repeated for the D$^-$, starting with the assembly of a K$^+$ candidate and iteratively combining all tracks with negative charge in the event. Since no further information is used to identify the daughter particles, this simple online D building relies solely on the fact that the true D combination is one of all the possible D combinations.

The invariant mass distribution of the D$^+$ candidates is shown in Figure 4.28(a) in red. A large combinatorially constructed set of composite particle candidates is created – on average about 10 per event. Lacking other information, the best D candidate per event is taken as the candidate which is closest to the nominal D mass, $m_D = 1.869\,\text{GeV}/c^2$. The distributions in Figure 4.28(b) show the absolute mass difference to the nominal mass. The green histogram is filled with the one candidate per event with the *least difference* to the D nominal mass. The red line marks another applied cut: A mass window of $\pm0.15\,\text{GeV}/c^2$ around the nominal D mass is chosen to suppress outliers. In Figure 4.28(a) the distribution for the one candidate with the least mass difference per event is also shown in green. The blue line shows that they are reasonably well distributed around the nominal D mass. A difference can be seen when comparing the heights of the distributions. The larger number of entries in the distribution of *all* D candidates results from the combinatorial assembly of the meson candidate. They match the nominal D mass, but are not comprised of the best set of daughter tracks per event. This is an artifact of the simple, combinatorial assembly wihout PID the algorithm employs.

The mass distributions of the D$^+$ and D$^-$ candidates, which have the least mass difference to the nominal D mass per event, are shown in Figure 4.29. The mass cut around $m_D = \pm0.15\,\text{GeV}/c^2$ is also applied. A coarse MC-truth matching is shown as well. The D candidate is considered *true* if it has all three K π π daughter tracks coming from the original true particles with a majority of hits. The resolutions of the individual distributions are shown in Figure A.5. They are summarized in Table 4.4 and compared to results obtained from applying the same technique on $\overline{\text{P}}$ANDA's default tracking (with and without an additional Kalman filter). The

(a) Missing mass squared before cuts.

(b) Missing mass squared after cuts. The x range is limited.
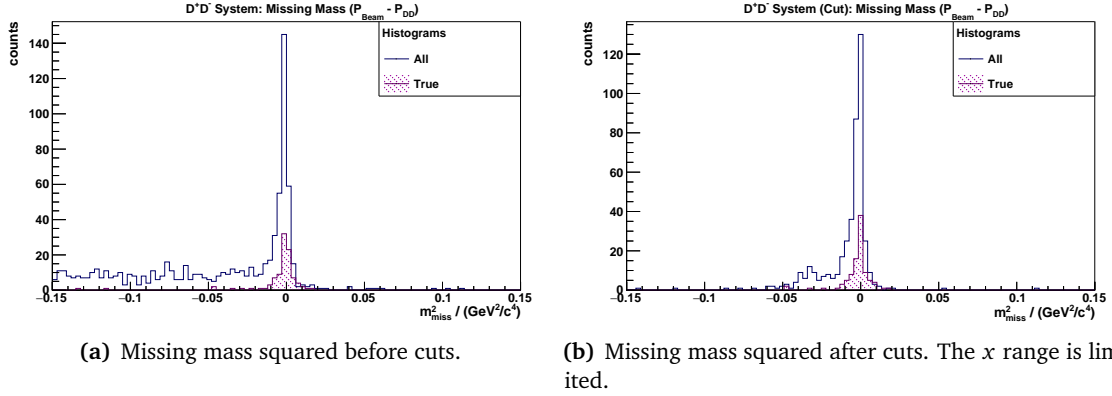
Figure 4.30: Distributions of the missing mass $m_{\mathrm{miss}}^2 = P_{\mathrm{b+t}} - P_{\mathrm{D^+D^-}}$ before and after cuts. The MC-checked candidates are superimposed. Selection criteria are given in Figure 4.31.

(a) p_t.

(b) Δp_z.
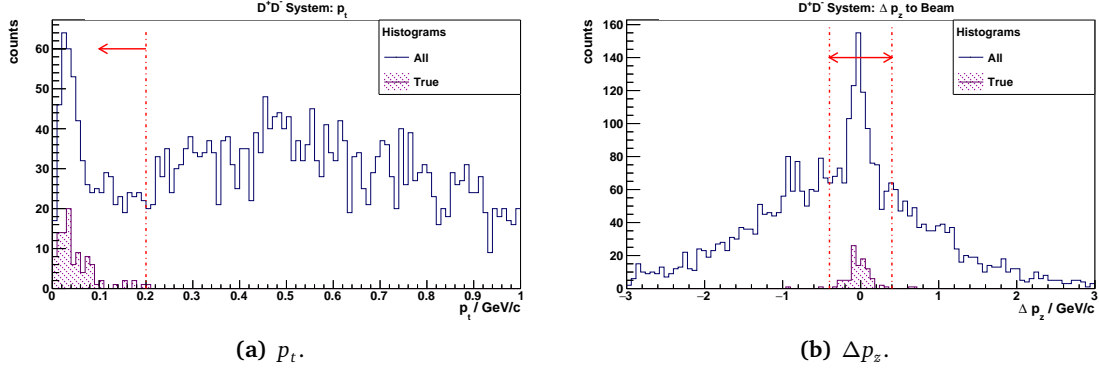
Figure 4.31: Transverse **(a)** and longitudinal **(b)** momentum distributions of the D^+D^- system. Cut values are outlined in red; the p_t cut accepts all candidates left of the red line, the p_z cut all candidates between the red lines. The MC-truth matched D mesons are shown as well.

plots, comparing the reconstructed D invariant mass distributions of this simple online event reconstruction using the default tracking, are given in Table A.6.

D^+D^- System If both D candidates of one event pass the mass pre-selection cut, their four-vectors can be combined to form a D^+D^- system. The distribution of the missing mass squared is shown in Figure 4.30(a). Figure 4.30(b) shows the distributions after applying cuts, which are introduced in the following. Before cuts, the distribution has the most entries around zero, suggesting a correct combination of candidates. A large tail to negative values can be seen, created by D meson candidates assembled from wrong combinations of daughter particles. These wrong candidates have four-vectors with too large magnitudes, compared to the beam. The cuts from Figure 4.31 remove most of the false combination to form a close peak around 0.

Figure 4.31 shows the distributions of transverse **(a)** and longitudinal **(b)** momentum for the D^+D^- system. The MC-matched candidates are outlined. To discriminate the D^+D^- candidates further against wrong combinations, two cuts are outlined in the graph. The first cut constrains

p_t to $p_t < 0.2\,\text{GeV}/c$ – Figure 4.31(a). The second cut restrains the longitudinal momentum around the beam momentum, $\Delta p_z = p_{z,\text{D}} - p_{z,\text{beam}} < \pm 0.4\,\text{GeV}/c$ – Figure 4.31(b). The cuts are motivated by the phase-space distribution of the produced D mesons, see Section 5.2.1. The corresponding graphs for the offline PandaRoot tracking are given in Figure A.7.

In 80 % of the cases, a candidate $D^\pm$ is found using the CH. The number is slightly lower for the offline tracking (73 % to 77 %). The mass resolution is approximately $35\,\text{MeV}/c^2$ for the unmatched, and $30\,\text{GeV}/c^2$ for the D mesons which have been MC-matched by the simple majority-based approach. Combining the two D mesons to an exclusive system, the pairs are found in 10 % of the cases for the CH and about 5 % for the offline tracking. The number of true candidates is only a fraction of the number of all candidates for all cases. The reason for this is two-fold. First, the basic approach of the simple online event builder combines every combination of pions and kaons with compatible charges in one event to a D meson candidate. A large number of combinatorially assembled D meson candidates is created which are discriminated in the following reconstruction steps. Still, after the employed cuts, the number of false-positive D candidates is comparably large, also in the signal region. The second reason lies in the determination of true candidates. Only a coarse, majority-based matching is done, which underestimates the number of true candidates.

The resulting distributions are reasonably located around the expected masses, showing a principle functioning of the CH algorithm, and also of the simple online event building. Not including any PID information of further sub-detectors of $\overline{\text{P}}$ANDA makes a pure and efficient reconstruction very challenging, as many wrong candidates need to be removed from the data stream. For the eventual online event building, the currently achieved efficiency should be increased further to reconstruct as many D meson candidates as possible. Potential efficiency enhancements are outlined in Section 4.4.2.3. Already in the current state, the resolution of the reconstruction is slightly better than the offline algorithm without the additional Kalman filter and is thus considered acceptable.

4.4.2.2 GPU Implementation

In its current state, the GPU port of the CH is generates angles φ and not p_t values. The principle of the circle center generation is equal for both approaches, the different sampling values are conceptually very similar.

The CUDA-based CH implementation comprises the following steps:

Angle Generation A number of φ values is generated with $\varphi \in (0°, 360°]$ and chosen granularity, e.g. $\Delta\varphi = 1°$.[19]

Circle Center Calculation Hit coordinates are copied to the device. A kernel for the circle center generation is invoked, with one thread computing one pair of circle center values per angle and hit. The kernel is launched with 256 threads and according block size. After completion, the circle center pairs are copied back to the host.

---

[19]In this CH approach, the angle values are equal for all hits. Going over to the p_t sampling method, every φ value is unique for a hit point. While this adds to the overall execution time, track curvatures can be targeted much more effectively.

Table 4.4: Overview of the mass resolution of D mesons from the simple online event reconstruction. $D^\pm_*$ refers to reconstructed particles, matched to the MC truth. For the D^+D^- system, the resolutions of the missing mass is given.

	Circle Hough		$\overline{P}$ANDA Offline Tracking			
			w Kal.		w/o Kal.	
	N	σ_m	N	σ_m	N	σ_m
	%	MeV/c^2	%	MeV/c^2	%	MeV/c^2
D$^+$	80	35.7 ± 1.0	73	42.3 ± 2.0	77	49.0 ± 1.4
D$^+_*$	14	29.5 ± 1.6	15	18.0 ± 1.0	17	49.5 ± 3.4
D$^-$	79	36.8 ± 1.1	74	47.3 ± 2.0	77	47.3 ± 1.2
D$^-_*$	15	28.2 ± 1.3	16	22.4 ± 1.4	18	56.0 ± 2.5
		σ_{miss}		σ_{miss}		σ_{miss}
		$(MeV/c^2)^2$		$(MeV/c^2)^2$		$(MeV/c^2)^2$
D$^+$D$^-$ (cut)	10	3.0 ± 0.3	6	0.4 ± 0.1	5	14.4 ± 1.8
D$^+$D$^-_*$ (cut)	2	4.3 ± 0.9	2	0.9 ± 0.2	3	9.7 ± 1.5

The computational complexity for each kernel is kept low, as only one circle center pair is computed. A high GPU occupancy is reached with this technique. Currently, no accumulator array filling is implemented on the GPU. The host-side infrastructure is used for this. As can be seen in Figure 4.32, the transfer back to the host is currently the bottleneck of the GPU for more than 50 hits. The point of optimal operation is at approximately $\mathcal{O}(1000)$ hits, see below.

By means of an interface, the GPU CH is included into PandaRoot. It is linked to the ROOT application as a shared object using CMake and the `extern` keyword of C++. If a GPU device is present, the user can decide to use the accelerated GPU version of the CH, or run with the conventional CPU code by changing a flag in the ROOT macro during runtime. The following time measurements have been done with a stand-alone version of the GPU version, to exclude unknown contributions from the PandaRoot framework. If not specified differently, the sampling granularity of θ is $1°$ and the number of threads per block is 256.

Performance of the Parts of the Algorithm Benchmarking of the algorithm was done with three time-wise intensive parts of interest: The time needed for copying all hit coordinates to the device (*Copy HD*); the time for execution of the kernel on the whole set of hits (*HT*); the time for copying Hough-Transformed circle centers back to the host (*Copy DH*). The time needed for the three steps as a function of the number of hits is shown in Figure 4.32(a) for the GeForce GTX 750 Ti card. It can be seen clearly that the process of copying data back to the host is the most time consuming for $N_{hits} \gtrsim 50$. The double-logarithmic plot of Figure 4.32(b) highlights the difference for small numbers of processed hits and also illustrates the three principle behaviors of the parts. For $\mathcal{O}(>100\,\text{hits})$ the HT kernel computation time per hit begins to saturate. The kernel has reached its least run time, a higher number of hits neither increases nor decreases the time needed per hit. While the same is true for the transfer of data back to the host, the time needed for the initial host to device transfer of hit points decreases over a large range of hit sizes. The maximum in parallel execution is reached only for approximately 50,000 hits.

(a) Time per hit.

(b) Time per hit, double-logarithmic plot.
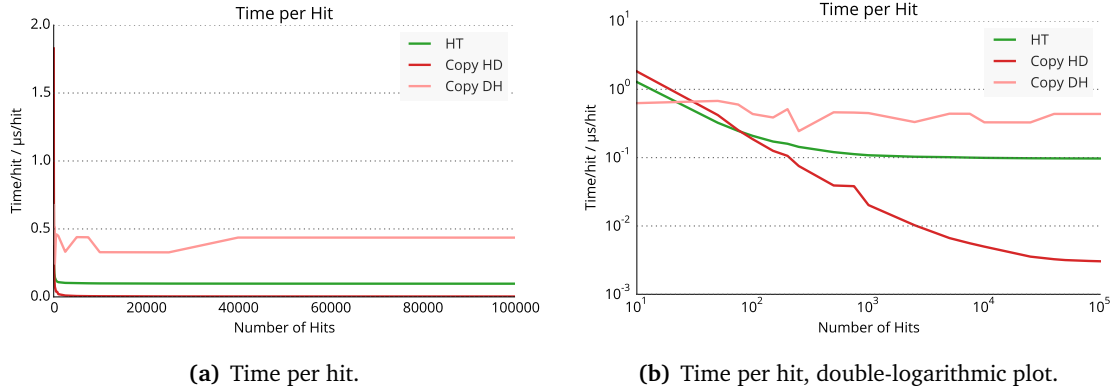
Figure 4.32: Run time, normalized to the number of processed hits, for the three parts of the GPU Circle Hough Transform on a GeForce GTX 750 Ti device: The data copy to the GPU device (*Copy HD*); the kernel call (*HT*); re-transfer of generated data to the host (*Copy DH*). **(b)** show a double-logarithmic plot to highlight the behavior for a small number of hits.

(a) Time per hit.

(b) Time per hit for limited range of hits.
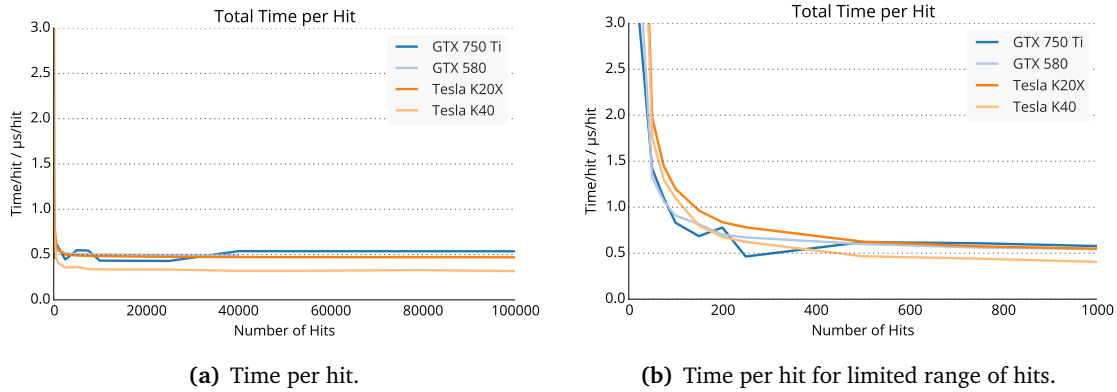
Figure 4.33: Total run time, normalized to the number of processed hits, for different GPU cards **(a)**. **(b)** limits the scope of the *x*-axis to highlight the behavior for lower numbers of hits. The memory of the GeForce GTX 580 limits its processing of hits to <40,000.

The copy of data back to the host is the slowest process in the algorithm, as it is largely constraint by the host CPU. Filling the accumulator array on the device side and extracting track parameters directly on the GPU circumvents the need for copying back.

Comparison of Different GPUs In Figure 4.33 the total runtime of four different GPU devices is compared. As the total run-time is largely dependent on the host-side-limited memory copy, the consumer-grade GTX GPUs and the Tesla cards are comparably fast. Minimal processing time is reached around 20,000 hits, after which the distribution slightly rises again.

The performance of the actual HT kernel, generating the circle center values, is shown in Figure 4.34. The full range of analyzed numbers of hits is given in **(a)**, **(b)** focuses on a smaller range of hits. A performance plateau is reached at approximately 20,000 hits for the consumer-grade and at 50,000 hits for the HPC devices. The performance is higher for the HPC

(a) Performance.

(b) Performance of limited range to 10,000 hits.
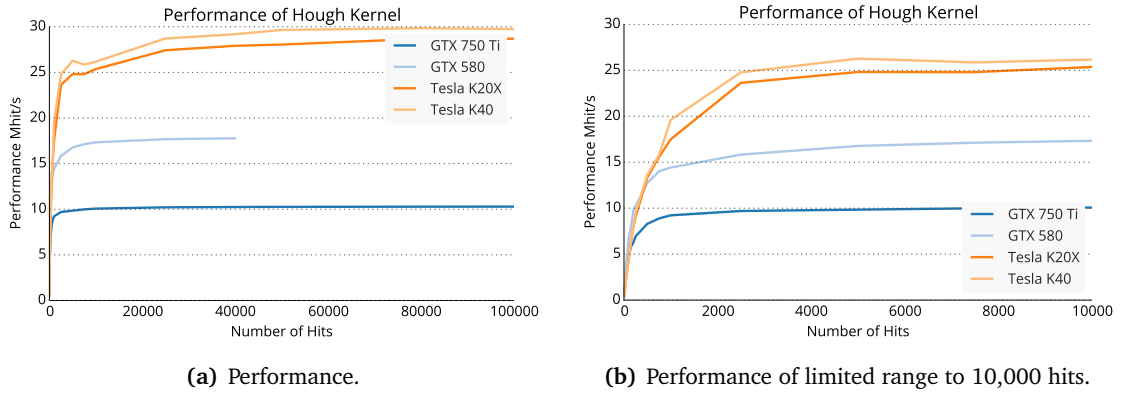
Figure 4.34: Performance in number of processed hits per second for the Hough kernel. The memory of the GeForce GTX 580 limits its processing of hits to <40,000

(a) Time per hit for copying data from device to host.

(b) Ratio of time for copying and kernel invocation.
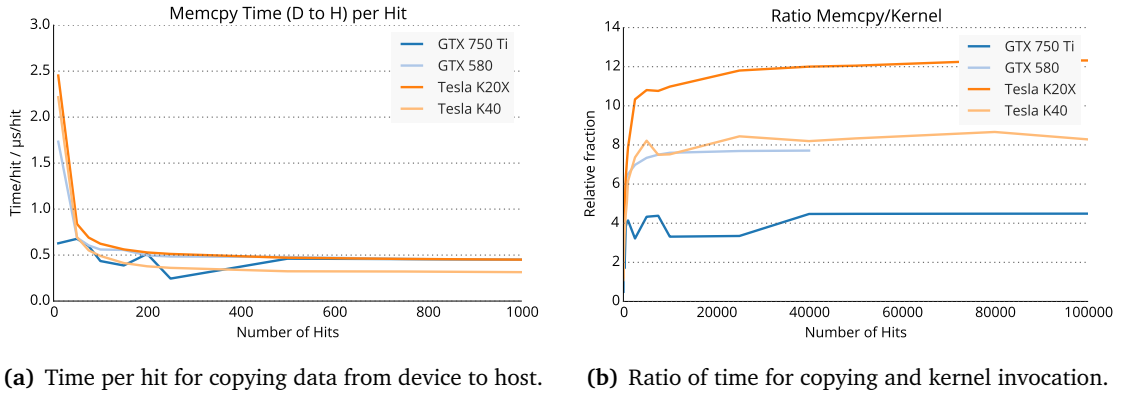
Figure 4.35: Illustrations for data copy from device to host.

cards, by up to a factor of 3, as the cards offer a larger amount of CUDA cores for more kernel throughput.

Figure 4.35 **(a)** and **(b)** illustrate the current bottleneck of the algorithm: The transfer back of the circle centers from the device to host. Pinned host memory has been used to accelerate device-side memory access. Still, the time needed for data copying is a factor of 4 higher for the GeForce GTX 750 Ti, and around a factor of 10 higher for the Tesla cards. The reason for the smaller impact on the GeForce GTX 750 Ti is the lower kernel performance of this device.

Dependence on Sampling Parameter The granularity with which θ is sampled directly determines the number of Hough transformation calculations done per circle center. Per θ step, one circle center (x, y) pair is computed. Four distributions for the four different GPUs are shown in Figure 4.36, with a double-logarithmic plot in Figure 4.36(b), highlighting the behavior for small numbers of angles. Since more computations are needed for higher number of angles, the overall performance decreases. A notable effect can be seen for the Tesla cards at 180

(a) Performance.

(b) Double-logarithmic performance.
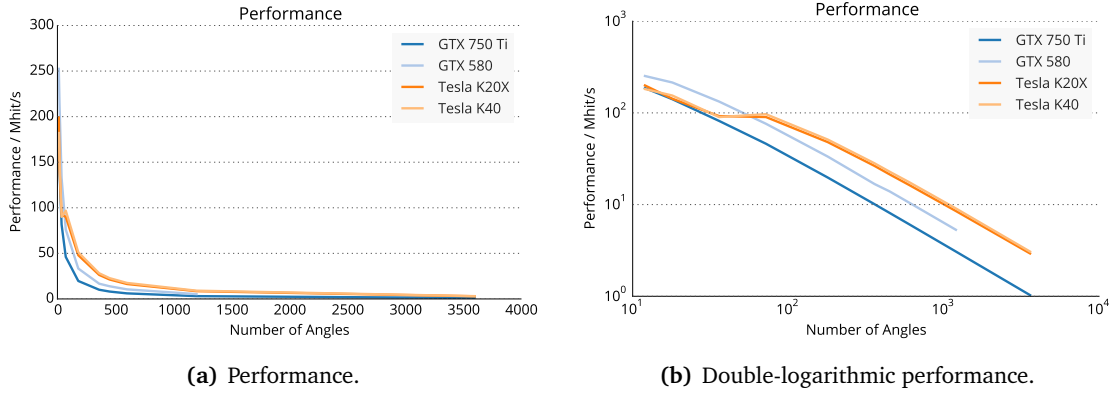
Figure 4.36: Performance for the HT kernel part of the CH for different numbers of angles and GPUs. For this plot, the number of processed hits is fixed to 5000 and the number of threads is fixed to 128.
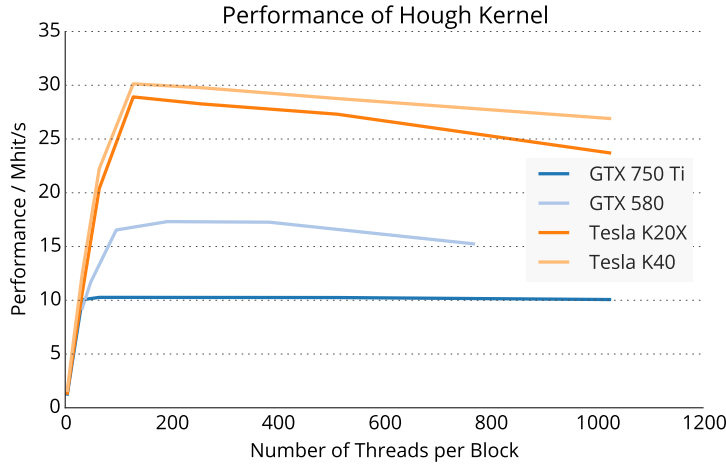
Figure 4.37: Number of threads per block (block size) for different GPUs. The maximum block size is defined by the compute capability.

angle values. Here, the performance is equal to that of 72 values. Apparently, a computationally efficient occupancy is reached.

The overall trend of all distributions is of decreasing performance with an increasing number of φ samples. With a finer sampling of the Hough space, more time for computations is needed. While this is an expected effect, it underlines the importance of carefully choosing the needed sampling precision. In the current implementation of the CH, a coarser sampling step width is always the better option in terms of computing time. Instead of increasing the sampling frequency, the sampled parameter can be chosen more carefully when sampling in p_t and not in θ. See the outlook in Section 4.4.2.3.

Dependence on Block Size   The number of threads per block, the block size, is a parameter which is subject to optimization. With an optimal block size, the GPU can be occupied so as to

leave the least number of computing cores idle. In the CH implementation, the number of hits, N_{hits}, and the number of threads per block, N_{threads}, determines the number of blocks (the grid size), N_{blocks}:

$$N_{\text{blocks}} = N_{\text{hits}} \times N_\theta / N_{\text{threads}},$$

with the number of θ samples N_θ. Grid size and block size are both limited for specific compute capabilities. Usually, the maximum block size[20] is $N_{\text{threads}}^{\max} = 2^{10}$. The grid size varies between $N_{\text{blocks}}^{\max} = 2^{31}$ (GeForce GTX 750 Ti, Tesla K20X, Tesla K40) and $N_{\text{blocks}}^{\max} = 2^{16}$ (GeForce GTX 580).

The performance in dependence of the block size is shown in Figure 4.37. The graph for the GeForce GTX 580 is limited by the maximum grid size. It can be seen that the ideal number of threads is located between 128 and 256. The latter has been chosen in all previous performance measurements.

4.4.2.3 Discussion and Outlook

A novel approach for parallel track reconstruction in an online environment has been presented in this section: the Circle Hough Transform. The algorithm samples track hypotheses as circle representations through sets of hits and extracts the circle parameters, which best connect the hits. In a benchmark channel, tracks are reconstructed with about 90 % efficiency and a 4 % momentum resolution. Masses of initial state particles can be reconstructed with a 35 MeV/c^2 resolution and 80 % efficiency, using a simple online event builder. The efficiency and resolution is slightly better when applying the simple event builder on events reconstructed with the offline track reconstruction algorithm of PandaRoot. In its first implementation on the GPU, the algorithm reaches a peak of 30 Mhit/s performance on a Tesla K40 card.

The reconstruction efficiency is not yet satisfying. Although the algorithm allows for reconstructing events from the tracks, the number of reconstructed initial state particles is lower than for $\overline{\text{P}}$ANDA's offline track reconstruction (see Section 5.4). Work still has to be done to improve the reconstruction efficiency. Ideas for improvement are presented in the following. The GPU implementation already achieves very good results, but up to now only parts of the algorithm are implemented on the device. Especially track parameters extraction is a needed algorithmic addition to the GPU version of the CH.

Algorithmic Aspects Tracks reconstructed with the CH have approximately half the resolution of Kalman-fitted tracks from the offline reconstruction in PandaRoot, see Table 4.3. The algorithm seems to be a feasible candidate for an online track reconstruction algorithm.

First analyses of tracks which were not reconstructed by the CH show that a large part of the efficiency loss is due to sub-detector issues. The MVD has more material than the STT which leads to deviations of tracks from perfect circles. Since hits from the MVD are close the coordinate origin, a small deviation has large effects on the track properties. A way to handle the different sub-detector properties is to form tracklets in the MVD and tracklets in the STT separately and combine them to a common track in a second step.

---

[20]Block size and grid size are of three dimensions. As the CH algorithm only makes use of the x-value, only the quantity of the first dimension is given here. See also: Section 4.1.2.1.

A modification of the CH, currently under development, incorporates this method and shows promising results. The modification uses geometrical properties of tracks in the MVD to define regions of interest in the STT. The modification both reduces the amount of computations needed and also increases the reconstruction efficiency. Especially, the number of spurious found tracks is reduced greatly.

Since MVD hit points define regions of interest in the STT, the approach can also be used for defining regions of interest in *time*. Up to now, the algorithm assumes that the drift isochrones of the STT are already pre-calculated. For a real online track reconstruction, there is a strong possibility that this isochrone information is not available yet. The needed event start time, t_0, is not present a-priori in the STT. But hits in the MVD have a very precise time-stamp and can constrain the STT window.

Currently, the algorithm operates in an event-based mode. Generally, there is no limitation for the algorithm to run in a time-wise environment. First tests show no principle hurdle in the track reconstruction. A difficulty is track parameter extraction from the accumulator array. In a time-based stream of hits, the Hough space is much more crowded with hits from different events overlapping one another. The definition of a region of interest in time can be very beneficial here.

A not yet explored possibility of the CH is its capability to reconstruct secondary tracks. At the moment, circles are always assumed to go through the origin. This reduces the number of computations, but in principle the algorithm does not need this constraint. Especially in a two-stage approach, in which the first stage removes *primary hits* from the hit stream. In this case, the computational complexity should be manageable.

GPU Implementation   The first GPU implementation ports only the computing-intensive core of the CH to the device, the Hough kernel. A kernel performance of 30 Mhit/s is reached for a Tesla K40 card. Currently, the algorithm is bandwidth bound, as most of the time is spend copying data between the device and host. The next development makes use of CUDA's `atomicAdd()` to fill the accumulator array directly on the GPU. With this, expensive memory transfers are avoided, which is the limiting factor especially for the HPC-grade cards.

To increase the performance, further methods can be applied. Currently, the kernel structure is flat. No nested kernels are invoked. Using one kernel per hit pair, which calls a daughter kernel for the actual Hough transform, can increase the GPU occupancy of the algorithm. The bunching wrapper, introduced in Section 4.5.2.2, can extend parallelism of the CH as well. A necessity is that track parameter extraction is also run directly on the GPU.

4.5   Triplet Finder

The Triplet Finder (TF) is designed as a part of $\overline{\text{P}}$ANDA's online track reconstruction. The algorithm is built to process hits from the STT in an continuous, time-wise way and not to rely on the event starting time t_0. The TF is initiated by combining a small number of neighboring STT hits without including drift time information.

The TF was first presented in [144] and [145] as a proof-of-concept, CPU-only implementation.
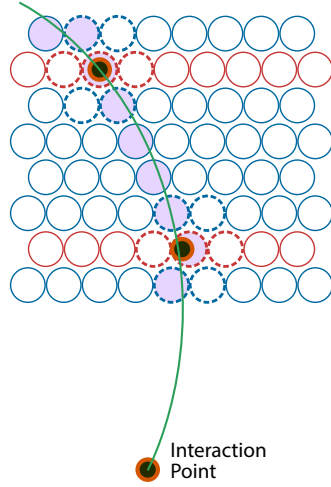
Figure 4.38: Sketch of the principle procedure of the TF (not to scale). As soon as a hit (filled circle) is detected in a pivot layer (red outlines), the algorithm searches for adjacent hits (dashed outline). If at least three hits are found, a center-of-gravity hit (*Triplet*) is formed (black dot with red outline). Two Triplets are combined with the interaction point to form a track candidate (green). Subsequently, hits lying close to the track are associated.
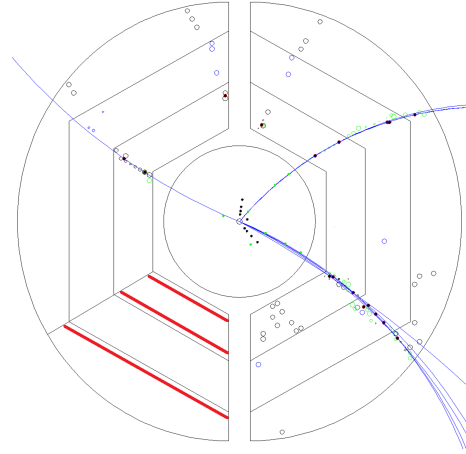
Figure 4.39: View of the STT along the beam axis. Outlined in red are the three pivot layers, exemplary for one of the six sectors of the STT. In blue tracks formed in a narrow time window can be seen. The small circles in the STT represent the drift isochrones of hit straw tubes. Additionally, hit points from the MVD can be seen in the center [145].

In the following, the algorithmic design of the TF is outlined and specific points of the optimization shown for the GPU. Finally, a outlook of the further possibilities of the algorithm is given.

4.5.1 Algorithmic Description

From the initial design of the TF, the algorithm is conceived to run time-wise. Instead of considering one event at a time, the algorithm searches for tracks in time windows. Tracks are reconstructed in the xy-plane normal to the beam axis. The depth information along the z-axis is not yet utilized.

Triplets are first created, then used to form circles as track candidates. Finally, the remaining hits are associated to the track. The scheme is explained in Figure 4.38. The following description is based on the CPU reference implementation, a first proof-of-concept design. It still has explicit and implicit parameters which need to be optimized for their algorithmic performance.

Triplet Generation   The straws of the STT are aligned rotationally symmetric as a barrel around the z-axis in the center. The straws are organized in layers, starting from the inside: 8 to 10 layers of *axial* straws are followed by 8 layers of *skewed* straws, which are followed by 5 to 11 axial straws on the outside, see Section 3.3.2.2. The algorithm is initiated by a hit point

in a straw belonging to a dedicated layer of straws, a *pivot layer*. Currently, three pivot layers are defined per STT sector, as illustrated in Figure 4.39: The second layer of straws from the inside; the second to last layer of straws in the inner axial part; and the layer of straws second on the outer section of axial straws. The pivot layers are chosen to provide a large lever arm and span a large distance of the STT. Further pivot layers are a possibility.

As soon as a hit is detected in one of the pivot layers, the TF searches for hits in the surrounding six straws. If at least three and less than six hits are found, the cluster of hits is used to formulate a virtual, center-of-gravity hit:

$$\vec{x}_{\text{cog}} = \frac{1}{N} \sum_{i=1}^{N} \vec{x}_i,$$

with $\vec{x}_i$ being the coordinates of the hit and i the hit's index in the array of all hits of the cluster of size $N \in [3, 4, 5, 6]$. In most cases, the number of hit points considered to form a virtual hit is $N = 3$, giving the center-of-gravity hits the name ***Triplet*** and the algorithm its title.

The center-of-gravity approach replaces the need for the event starting time t_0 for track reconstruction. Usually, t_0 is needed to calculate drift distances in straw tubes, as sketched in Figure 4.40. The drift time of an electron, t_{drift}, is calculated as the difference between the arrival time of the electron at the anode wire (t_{signal}) and the event starting time: $t_{\text{drift}} = t_{\text{signal}} - t_0$. By means of a calibration curve, t_{drift} can be converted into a drift distance, $r_{\text{iso}}(t)$, giving a circle describing the location of all possible drift start positions (see Section 3.3.2.2). The event starting time t_0 is essential to increase the STT's resolution from $\mathcal{O}(0.1\,\text{cm})$ (without usage of drift time) to $\mathcal{O}(0.015\,\text{cm})$ (with drift time). The TF makes an attempt to track particles through the detector without t_0, but improves the initial resolution by creating the combined Triplets.
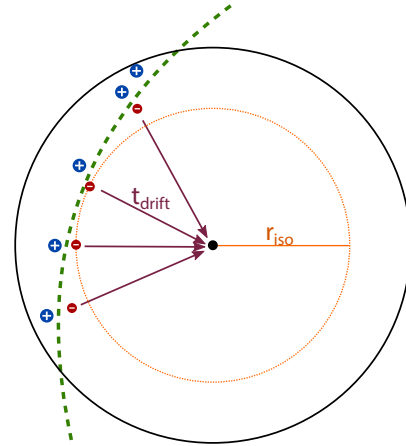
Figure 4.40: Sketch of a cross section of a STT drift tube (*straw*). Shown in green is a track traversing the straw, ionizing the gas (blue ions and red electrons). Electrons drift to the central anode wire (purple arrows), where an electrical current signal is generated upon arrival.

The TF is intended to work on a continuous flow of hit points, structured into sets of hits of certain size. For the initial version, a time window of HESR's burst of 2000 ns is used. During Triplet generation, it is checked, whether or not the hits are compatible time-wise. If they originate from the same track, their time stamps should be roughly equal. To account for the electron drift time inside the straws, a discrimination window of 200 ns is chosen (see Section 3.3.2.2), with 50 ns added for additional safety. Only if the neighboring hits in the Triplet generation are within this 250 ns time window are they combined to a Triplet.

Forming Track Candidates   Track candidates from the TF are described as circles, since in this stage the algorithm only considers the x and y position of hit points, and the STT has

a low material budget. A circle can be formulated, analytically, by specifying three points. The TF creates circles from two Triplets and the interaction point. Adding the coordinate origin constraints one degree of freedom and limits track candidates to tracks of primary particles. While this lowers the combinatorial complexity considerably, generally a third Triplet could be taken for circle creation as well. As the amount of processed hits is increased, it might be included in a version of the algorithm running in a later stage of the online track reconstruction chain, when many of the primary tracks are already removed from the data stream.

Hit Association In the last step of the TF, all STT hits are tested for combination with the track candidate from the previous stage. A hit is considered part of the track, if its distance to the track is $\leq 6\,\text{mm}$ – the 5 mm straw tube radius and a 1 mm safety margin. Additionally, the hit is checked to lie inside the considered time window.

A track candidate is finally accepted as a track if the number of hits is above a certain threshold.[21]

4.5.2 Performance

This section covers the performance of the TF. First, in Section 4.5.2.1, the physics performance is shortly presented. In Section 4.5.2.2, points of the GPU optimization of the algorithm are highlighted. Further GPU optimizations are shown in B.

4.5.2.1 Physics Performance

In [144] and [145], the performance of the track reconstruction of the TF is presented for the CPU reference implementation. In the following, the most important results are summarized.

The track reconstruction efficiency is measured relative to those tracks, that are *reconstructible*. The considered data set is a DPM-generated set at $\overline{\text{P}}$ANDA's maximum beam momentum of $15\,\text{GeV}/c$. The DPM generator is designed to simulate background-like events. Many tracks are included, which have very low transverse momentum or are forward-boosted. Particles of these events do not reach the STT at all and, hence, are excluded when calculating the efficiency. The chosen discrimination parameters for a track to be reconstructible are: It is a charged track from a primary or secondary particle; it has a transverse momentum of $p_t > 60\,\text{MeV}/c$; its relativistic velocity fulfills $\beta\gamma > 0.4$; it is not the track of a particle with displaced vertex – the origin of the track is close to the interaction point ($\vec{x}_{\text{start}} < 2\,\text{cm}$); at least one STT straw is hit. The parameters are intentionally chosen coarse to not bias the data set. Possibly, the parameters could be tightened further, especially the last requirement, as the TF needs a minimum of 2×3 hits to work properly.

Applying the selection criteria on the DPM dataset with approximately 2 million simulated tracks, 9228 tracks are considered reconstructible. Of those, the TF finds 6821 tracks, 74 %. In Figure 4.41, the momentum distribution of the reconstructible tracks **(a)** and the reconstructed tracks **(b)** is compared qualitatively. Two observations can be made. First, there are fewer

---

[21]Currently, the threshold is between 7 and 11 layers tubes, depending on the STT rows and sectors the track is lying in. It is a first assumption and subject to tuning, as it determines the algorithmic efficiency and purity.
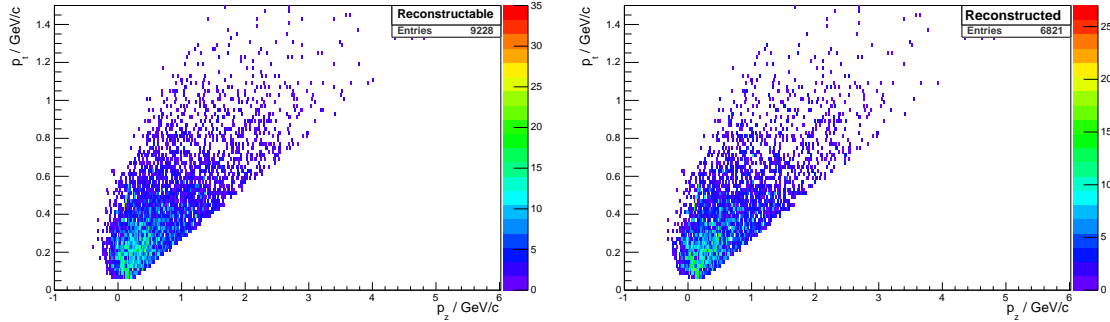
(a) Transverse momentum vs. longitudinal momentum of tracks considered for reconstruction.

(b) Transverse momentum vs. longitudinal momentum of reconstructed tracks.

Figure 4.41: Momentum distribution of tracks considered for the Triplet Finder and reconstructed by the algorithm [144].

events present, as limited efficiency. Second, the distribution of the events in the momentum space has the similar shape before and after reconstruction, suggesting well-reconstructed momenta.

The TF produces 38,445 tracks, of which 21,137 have more than 80 % correctly assigned hits. Comparing the number to the 74 % from the previous paragraph shows that the algorithm reconstructs each track multiple times – considering only the tracks with > 80 % *true* hit content, each original track is found three times on average. The reason for the multiple reconstruction is that the TF considers different sub-sets of hits of the same track and creates track candidates based on each sub-set. A track merger is a possible post-processing step combining multiply found tracks.

4.5.2.2   GPU Speed Performance

In the following, the speed performance of the Triplet Finder is shown. The performance is the result of many optimization steps. A few important optimizations are presented in the course of this section, a detailed description of all optimization steps is given in appended Section B. The GPU optimizations have been done in collaboration with the NVIDIA Application Lab of the Supercomputing Centre of the Forschungszentrum Jülich and have been partly presented in [146] and [147].

If not stated otherwise in the following, a Tesla K20X GPU is used for the performance study. The NVIDIA GPU with a Kepler chip has 2688 streaming processors and 6 GB RAM (see also Table 4.1). CUDA is used as the programming language in version 6.5. The data sample for testing is a DPM-generated set with approximately 150,000 hits. For individual performance optimizations, the results of the other optimizations are already included.

Initial Implementation   For the initial GPU implementation of the TF, the original CPU algorithm as described in Section 4.5.1 is ported to the GPU. The algorithm starts operating directly on hits, forms Triplets from hits, computes tracks from Triplets, and finally associates hits to tracks. The different stages are processed independently by individual kernels.

(a) Processing time.

(b) Rate of processed hits (*performance*).
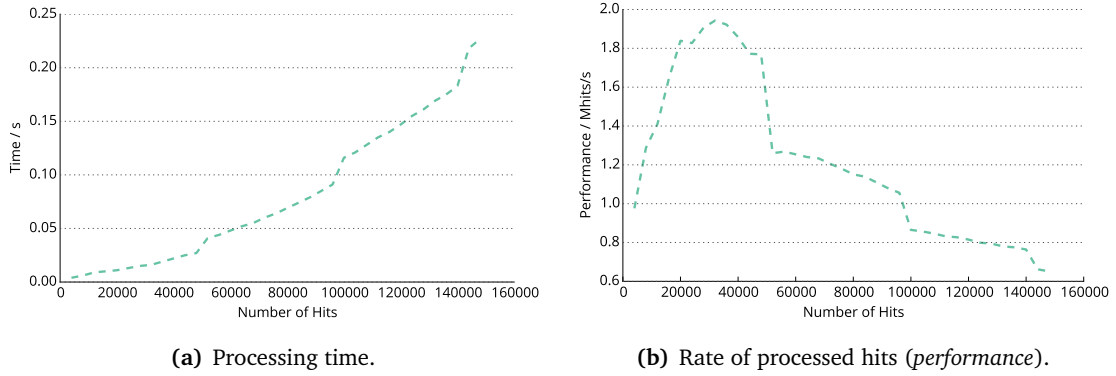
Figure 4.42: Processing speed of the initial GPU implementation of the Triplet Finder. Both plots are functions of the number of processed hits.
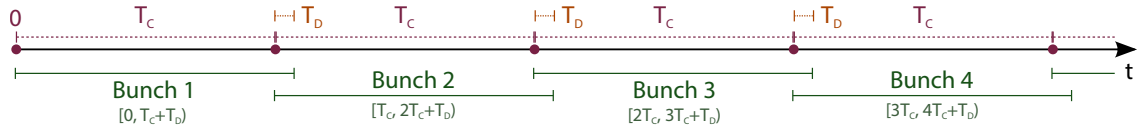
Figure 4.43: The sub-dividing method for the TF. A *bunch* contains the hits of a core time window, T_C, and the hits of an additional increment due to the drift time inside the STT's straws, T_D.

The computing time and performance of the initial implementation are shown in Figure 4.42. The graphs are functions of the number of hits processed at the same time: The number of hits the TF considers during one call while track finding. In Figure 4.42(a) the algorithmic complexity of the algorithm, $\mathcal{O}(N^2)$ can be seen, as the processing time rises quadratically with increasing number of processed hits. Two steps in the functions are visible, at 50,000 and at 100,000, due to GPU occupation inefficiencies: The GPU is fully loaded with a set of hits and only a few additional hits lead to an unproportionally large increase in processing time, as the invocation overhead becomes more pronounced. Figure 4.42(b) shows the rate with which the considered data-set is processed, its *performance*. It can be seen that the rate peaks at approximately 25,000 hits processed at once. The next optimization makes use of this fact.

***Bunching* Scheme** As seen in Figure 4.42(b), the TF has an optimal working point of 25,000 processed hits. This $\mathcal{O}(10^4)$ number of hits is needed to fill the GPU optimally in comparison to the algorithmic complexity. A smaller number of hits leaves processors unoccupied, resulting in a non-optimal performance. Increasing the number of processed hits further, beyond 25,000, results in a loss of performance. The algorithmic complexity still increases quadratically with the number of considered hits and some processors are idling during the computations. Again, a non-optimally occupied GPU is the result.

A sub-division of the considered data stream circumvents the loss in performance, as the computational complexity is reduced. In the context of the TF, the sub-division is called a *bunch* and the part of the algorithm, dividing the continuous data stream into bunches, is called the ***bunching wrapper***. The quadratic algorithmic complexity of the algorithm applies only on the sub-set of the hits in the bunch and does not extend beyond the working point to greater

(a) Processing time.

(b) Comparison of performance between the initial, not-bunched and the bunched version.
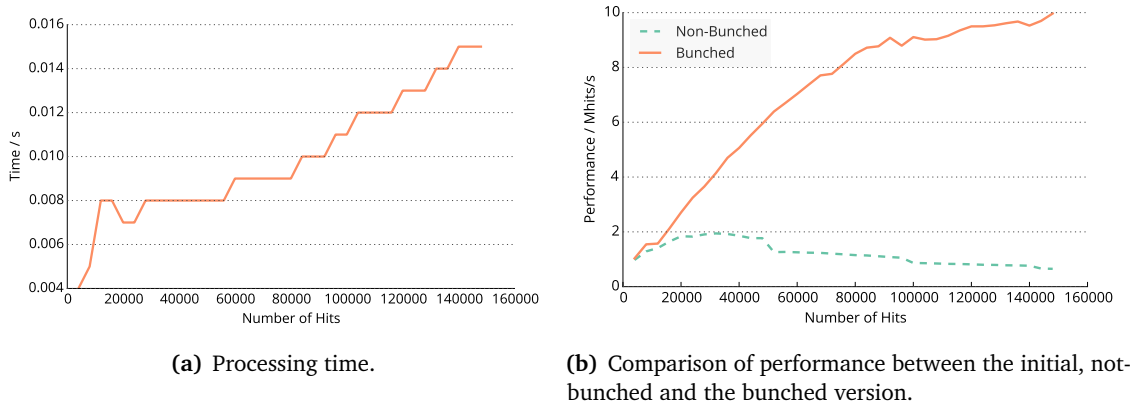
Figure 4.44: Processing speed of the GPU Triplet Finder with the *bunching wrapper* as a function of the number of processed hits. The dynamic parallelism bunching approach is used with sector-row binning.

number of hits. A small overhead is generated, as for every bunch division an overlap has to be considered. The overlap is the maximum drift time in the straws of the STT, 200 ns, which is added to the timestamp of the last hit in the bunch. An illustration is given in Figure 4.43. A higher number of bunch divisions adds more overlaps and, consequently more effective time is needed to process the algorithm. However, the performance benefits introduced by bunching vastly outweigh the time overhead.

With N_C being the average number of hits in a bunch and N_D the average number of hits in a time window equivalent to the drift time, the algorithmic complexity in dependence of the number of processed hits, N, is

$$\mathcal{O}\left(N \frac{(N_C + N_D)^2}{N_C} \right), \tag{4.35}$$

which is linear in N. Due to the overlap, it is possible for the TF to consider hits twice and reconstruct a track more than once. A track merging as a post-processing step is a possibility to combine doubly reconstructed tracks. For a bunch size equal to HESR's beam burst size of 1600 ns, adding the overlap time is not needed, as a 400 ns gap without events follows.

Calling the TF again with a number of certain hits, but now bunched into sub-sets of $2\,\mu s$[22], yields the time measurement shown in Figure 4.44(a). The expected linear behavior of Equation 4.35 can be seen after an initial plateau. Comparing the performance between the un-bunched version from the previous section with the bunched TF version visualizes the large gain introduced through the wrapper (Figure 4.44(b)). After a rise for small numbers of processed hits, still not filling the GPU, the bunched version reaches a plateau at about 90,000 hits and increases in performance only very slowly after this point. For the number amount of processed hits, the bunched version is a factor of 15 faster than the unbunched version – it reconstructs tracks with a rate of nearly 10 Mhit/s.

---

[22]The value of $2\,\mu s$ is subject to tuning. Refer to Section B and Figure B.2 for a determination.
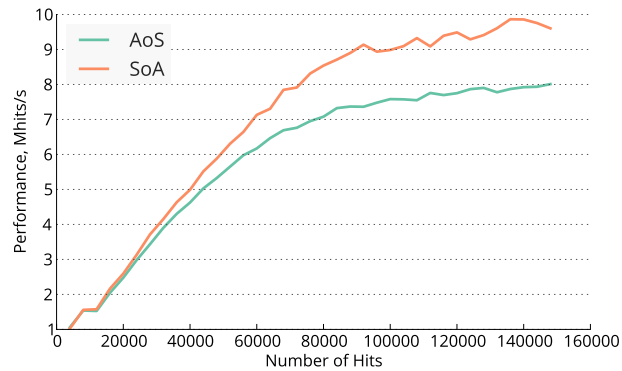
Figure 4.45: Performance comparison for data types using an array of structures (AoS) or a structure of arrays (SoA).

The bunching wrapper was developed in the scope for the TF, but is applicable on top of a variable algorithm. It can be used for other algorithms which have an ideal working point of input data sizes.

Three different ways of invoking bunches have been investigated. An approach, which starts an initializing GPU kernel which then subsequently calls all the individual TF stages (*dynamic parallelism* approach); an approach, which uses the method of host streams to launch the individual TF GPU processes (host streams approach); an approach, which directly launches one large GPU kernel per bunch, containing all the TF stages. The approaches are explained in more detail and benchmarked in appendix Section B. The dynamic parallelism approach is the fastest, as it requires the least amount of CPU-GPU communication and can occupy the GPU very well with the small kernels.

Structuring of Datatypes – SoA and AoS Arrays of complex data can be structured in different ways. Most of the time, a data class holds all the information for a specific entity, e.g. a hit, and the individual parameters are components of this class. A number of entities, e.g. a set of hits, is then a collection (array, vector) of this class. It is an array of this structure (AoS). An alternative approach is to embed the length of the array into the class and create collections from the smallest possible parameters inside of the class, a structure of arrays (SoA).

Array of Structures:

```
struct hit {
        float x;
        float y;
        float z;
        float r;
};
hit AoS[N];
```

Structure of Arrays:

```
struct hit {
        float x[N],
        float y[N],
        float z[N],
        float r[N];
};
hit SoA;
```

While an AoS is usually the more immediate programming approach and structures data accessibly, a SoA often has benefits for parallel architectures. SoA avoids loading unnecessary parameters and is better in terms of memory coalescing, as data of one type is not interleaved by others but allocated at subsequent addresses.

Going over from AoS to SoA increases the maximum performance by 20 %, as shown in
Figure 4.45. The gain in performance is especially notable in the track candidate creation and
the skewlet generation step – improving the invocation time by approximately 25 % and 40 %,
respectively.

Consumer-Grade Devices   In Section B, the TF is benchmarked on two different server-grade
GPUs, a Tesla K20X and a Tesla K40. For an impression of possible future performance gain,
both cards are also operated with an increased GPU clock rate. The maximum performance is
12.5 Mhit/s on the Tesla K40 with increased clock rate.

Both the Tesla K20X and Tesla K40 are supercomputer-grade GPUs. They have a large number
of computing cores and memory and employ ECC for error correction. The benefits in computing
power and features result in high retail prices. Also consumer-grade GPUs, which generally sell
at vastly lower prices, are enabled for GPGPU computing and can be employed for computing
with high performance.

The consumer card GeForce GTX 750 Ti is compared to the Tesla K20X in Figure A.8 for the
different presented optimizations. The same trends already seen for the Tesla K20X can generally
be noted. The performance differences between the host streams and the dynamic parallelism
approach for bunch invocation in the bunching wrapper are considerably smaller (see Section
B). Also, skewlet binning appears to not have the negative performance impact it does for the
Tesla K20X, but leads to performance results equal to the binning-less approach (see Section
B). Two reasons can be noted for the effects observed: The GPU of the GeForce GTX 750 Ti is
much smaller (bunching), and it is clocked with a higher frequency (skewlet binning).

The comparison of the TF performance after all optimization steps for the two devices is
shown in Figure 4.46(a). For the full number of processed hits, the performance of the GeForce
GTX 750 Ti is 50 % smaller than the performance of the Tesla K20X. At this point, the Tesla K20X
processes the hits in 15 ms, the GeForce GTX 750 Ti in 29 ms.

Since the GeForce GTX 750 Ti is a consumer-grade GPU, the number of built-in multiprocessors
is much smaller (5, instead of 14 for the Tesla K20X). Also, the card's clock rate is higher
than the Tesla K20X's (1020 MHz compared to 732 MHz). Normalizing the performance of
Figure 4.46(a) by both values, the distribution of Figure 4.46(b) is obtained. It illustrates the
performance difference without accounting for the *size* of the card and gives an estimate of
the raw performance. With this normalization, the performance of the GeForce GTX 750 Ti is
higher for a small and intermediate number of hits. The distribution of the GeForce GTX 750 Ti
reaches a plateau earlier than the distribution of the Tesla K20X. For the full number of hits,
both cards have comparable performances. A different picture emerges when normalizing the
distributions not to cores and frequency, but to the retail prices of the cards; in Figure 4.46(c)
the 30-fold higher price of the Tesla K20X can be observed drastically. Both normalized graphs
do not account for the difference in possibly important features of the Tesla K20X, missing in
the GeForce GTX 750 Ti; especially error detection and correction might be a necessity for a
reliable tracking algorithm. They are a mere illustration of the gross economy of the cards.

Further Optimizations   In Section B further optimizations are presented. The way hits are
associated to track candidates in the last step of the TF is optimized to first define regions of

(a) Absolute performance.

(b) Core-normalized performance.

(c) Price-normalized performance.
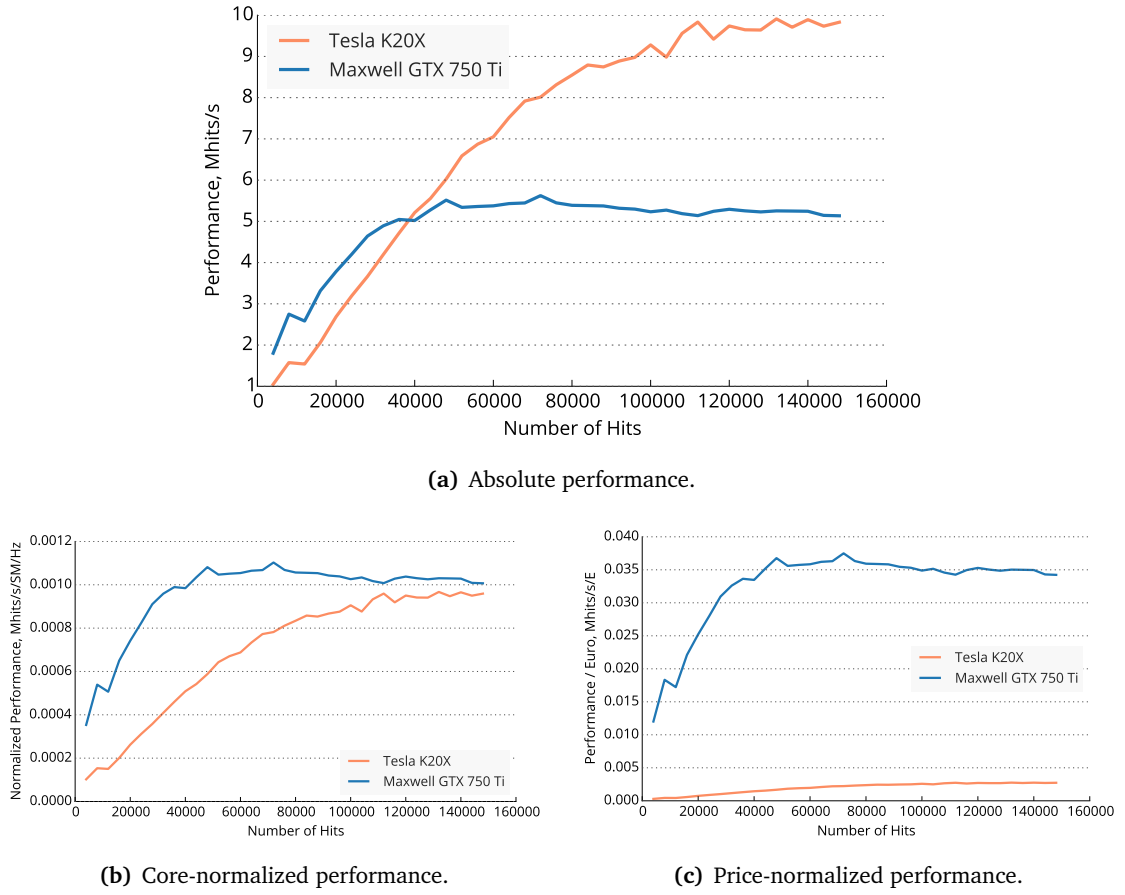
Figure 4.46: Performance comparisons of the Triplet Finder on two GPU devices: A Tesla K20X (orange) and a GeForce GTX 750 Ti (blue). **(b)** shows the distributions normalized to number of multiprocessors and GPU clock rate, **(c)** is normalized to the current retail price of the cards (GeForce GTX 750 Ti: 150 € ; Tesla K20X: 3600 €).

interest before associating hits. The techniques and results are shown in Section B. Three versions of CUDA are benchmarked in Section B. The latest CUDA version, 6.5, performs best compared to the two previous versions, although differences are only small. As already mentioned in this section, also different ways of processing bunches are investigated (Section B) and the two Tesla supercomputing cards benchmarked (Section B).

Summary After Optimizations To summarize the achieved performances, the set of approximately 150,000 DPM hits is processed in 15 ms (Tesla K20X) / 14 ms (Tesla K40) / 28 ms (GeForce GTX 750 Ti). The algorithm still employs parts which are running on the CPU (grouping of hits, hit-tube association). Adding this pre-processing step to the time measurement yields 17.6 ms for the Tesla K20X, or 120 ns/hit. In terms of performance, this is equivalent to 8.3 Mhit/s.

4.5.3   Summary and Outlook

The TF is capable of reconstructing DPM-generated tracks with 74 % efficiency. Using a Tesla K20X, it does so taking 120 ns/hit after all optimizations. The rate is equivalent to a performance of 8.3 Mhit/s. A first, very rough estimate of performance on future GPU architectures is given when run on the GPUBoost-enabled Tesla K40. Here, a performance of 11.5 Mhit/s is reached.[23] Using 120 ns/hit as a processing rate and assuming $\overline{\text{P}}$ANDA's high luminosity mode (2×10^7 event/s) and on average 80 hits/event in the STT, less than 200 Tesla K20X cards would be capable of processing the TF in realtime.[24] The bunching wrapper offers a natural way of parallelizing data processing beyond borders of a single device.

The TF algorithm and its GPU implementation still have potential for improvements, as outlined in the following.

Possible Algorithmic Improvements   Concerning the algorithmic performance, the most important improvement appears to concern the reconstruction efficiency. Benchmarking and careful tuning of the different discrimination parameters is needed.

The algorithm can be extended to cover also tracks originating from secondary vertices – by introducing a fourth pivot layer. The combinatorial complexity is vastly increased, but the capability for secondary-track reconstruction of the TF should be tested.

The algorithm creates tracks without using t_0 information. It should be possible to constrain the t_0 information of hits associated to the track and, by combination of all hits of a track, to give a first t_0 estimate. This is currently not utilized, as the algorithm itself does not need the information.

In addition, the TF currently works in the xy-projection and only reconstructs the transversal component of tracks. A reconstruction of the z-component is needed, especially since information from the skewed straws is already utilized in the process of skewlet association. The z-component can also be used to discriminate track candidates.

Finally, the large fraction of multiply reconstructed tracks illustrates the necessity of a track merger, combining many tracklets into one common track candidate. The earlier the merger can be applied in the chain of the algorithm, the higher the impact on the combinatorial complexity of the algorithm, potentially reducing its run time significantly.

Possible GPU Improvements   A benefit for both the CPU as well as the GPU would be book-keeping of associated hits. Currently, hits are associated to tracks by means of their time compatibility. While this works well, all hits have to be checked against all tracks. The number of comparisons could be reduced if hit book-keeping is introduced and the association status of a hit noted as a meta information. While this adds data to each hit point, possibly in the amount of 4 Byte (`int`), the combinatorial complexity is reduced, potentially outweighing the additional memory overhead.

---

[23]Both performance numbers include pre-processing steps still done on the CPU.
[24]Less than 140 GPUBoost-enabled Tesla K40 cards would be needed to process the same rate of 1600 Mhit/s.

The first attempt of reducing the load of skewlet association was made when introducing skewlet binning. Unfortunately, the results were poor. An alternative approach can be tested as well: The positions on the line, a skewlet can lie on, are not continuous but discrete based on all possible intersections of skewed straws. The possible positions can be computed once at startup of the algorithm and stored to constant memory. The look-up-table-like structure may have beneficial properties in terms of computing time.

While not expected to have a large impact, the pre-processing steps currently done on the CPU should be ported to GPU as well, to constrain the algorithm fully to the GPU. The capability of running with multiple GPU devices at the same time must be shown as well.

4.6 Summary

In the course of this section, algorithms for online track reconstruction on GPUs have been presented and performances measured.

Two Hough transforms using straight lines as representations of tracks are the subject in Section 4.3. Two different GPU implementations are benchmarked. A version, making use of pre-built functions offered by the Thrust package, and a version built directly into CUDA with many advanced programming techniques. The CUDA version outperforms the Thrust in all tested quantities. In addition, the code is better customizable. The CUDA version is the preferable solution for Hough transforming lines. The peak speed reached in the tests is 30 khit/s, not on a HPC-grade card, but on a consumer grade NVIDIA GeForce GTX 750 Ti.

A novel Hough transform is presented in Section 4.4: the Circle Hough Transform. Instead of sampling line parameters as track candidates, the algorithm directly samples circles. This way, 90 % of tracks in a benchmark physics channel can be reconstructed with an average transverse momentum resolution of 4 %. Using a simple online event builder, kaons and pions of the analysis channel of this thesis ($D^\pm \rightarrow K^\mp \pi^\pm \pi^\pm$) are combined to D meson candidates. The achieved mass resolution is about 35 MeV/c^2 and 80 % of D meson candidates are reconstructed. Comparing the results to those obtained when applying the online event builder to tracks reconstructed with PandaRoot's offline reconstruction shows that the CH has both higher efficiency and resolution. The first GPU implementation of the algorithm reaches 30 Mhit/s performance for the actual Hough transformation kernel on a Tesla K40. The current bottleneck has been identified as the data transfer between the host and device, which is subject to the next development of the GPU Circle Hough Transform.

The Triplet Finder is highlighted in Section 4.5. The algorithm initializes track reconstruction by creating fast circle hypotheses based on center-of-gravity hit combinations. The Triplet Finder is extensively optimized on the GPU and exploits several advanced techniques for performance gain. The most notable is the bunching wrapper, which splits the stream of hits into sets which occupy the GPU best. Using all optimizations, the Triplet Finder reaches a peak performance of 11 Mhit/s on a Tesla K40.

The different kinds of algorithms show that online track reconstruction on GPUs is a feasible technique for coping with $\overline{\text{P}}$ANDA's high-performance demands. As the devices are inherently

parallel, they can be used in sets to reconstruct the whole data stream of P̄ANDA. The speed performance is in a order of magnitude which makes GPUs a realistic solution for online track reconstruction, and it is expected to rise further in the future, as new GPU devices are released. The performance in reconstructing tracks is already good, but needs to be improved further. In the setup of the online trigger of P̄ANDA it is essential to lose as few events as possible in the track reconstruction stage. Here, further work is already on its way.

A field not covered in this thesis is the way GPUs can be embedded into P̄ANDA's DAQ system. Especially data transfer to GPUs is a non-trivial challenge. As the devices are always included into a PC, there are no well-established solutions for transferring raw experimental data to the device without passing the bottleneck of the intermediate CPU. There are projects in other experiments which push data directly to GPU memory via the PCI-Express bus. An evaluation of the techniques for P̄ANDA is needed.

Chapter 5

Application to the Reconstruction of $D^\pm \rightarrow K^\mp \pi^\pm \pi^\pm$

In this chapter, the charged decay of D mesons is investigated. The full decay chain is:

$$\bar{p}p \rightarrow D^+D^- \rightarrow K^-\pi^+\pi^+ \, K^+\pi^-\pi^-.$$

The D mesons are produced close to their threshold.

The chapter is structured as follows. First, the investigation is motivated and the over-all parameters of the analysis are outlined. Then, the event simulation is analyzed and compared to the original data. Subsequently, the individual hits in the sub-detectors are used to study the event topology and detector performance. The event is reconstructed, first inclusively, then exclusively, after this. Finally, the physics background is benchmarked and the event reconstruction applied with different track reconstruction algorithms.

5.1 Motivation

The physics program of $\overline{P}ANDA$ (see Section 3.2) focuses strongly on physics at the charm threshold. Charmonium and open charm physics will be investigated. Both spectra, the D meson spectrum of open charm, and the charmonium spectrum, are of intense interest in the particle physics community, as they continue to evolve and surprise. Many of these excited states decay in one, or even two ground state D mesons. Hence, it is important to reconstruct D mesons with a high efficiency and resolution.

The largest decay fraction of all hadronic D decays[1] is $D^\pm \rightarrow K^\mp\pi^\pm\pi^\pm$, as sketched in the Feynman graph in Figure 5.1 for D^+. The current literature value for the branching ratio is $(9.13 \pm 0.19)\,\%$ [1]. Because of its large branching ratio, the decay is often used to normalize other decays.

---

[1]The only channel with an higher decay branching ratio is $D^+ \rightarrow \overline{K}^0 e\nu$ with 9.2 %.
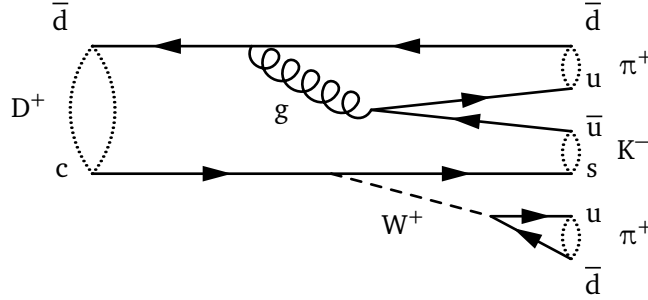
Figure 5.1: Lowest order Feynman graph of the $D^{+} \to K^{-}\pi^{+}\pi^{+}$ decay.

In addition, the final state consists solely of charged particles. Since this thesis benchmarks different tracking algorithm, this decay channel is a perfect test case for the performance of those algorithms.

Parameters of the Simulation Setup

This chapter makes extensive use of the simulation and reconstruction chain provided by PandaRoot. The following list summarizes the default parameters of the software setup used for the event simulation and reconstruction. Individual sections may differ in individual points.

- The PandaRoot overall version is fixed to `scrut14`, which combines with FairRoot external packages of version `apr13`. The Rho classes are updated manually to reflect the changes of PandaRoot's trunk with version `Rev. 27176`. The full detector setup is included in the applied full simulation.

- In the simulation, GEANT3 is used via the Virtual Monte Carlo interface.

- The beam momentum is set to $p_{z,\text{beam}} = 6.5\,\text{GeV}/c$.

- The decay channel is simulated using the Dalitz decay from EvtGen (see Section 5.2.2).

- Track reconstruction uses the default chain included in the specified PandaRoot version[2] (see also Section 3.5.3.4). In the forward part, an ideal track finder is used.

- PID employs `PidAlgoIdealCharged` to identify K and π. The algorithm retrieves the PID information from the MC truth.

5.2 Event Generation

EvtGen is used for the simulation of the studied physics channel. The tool was introduced in Section 3.5.3.1. Here, first the production of the charged D meson pair is outlined, then their decay to $K^{\pm}\pi^{\mp}\pi^{\mp}$ analyzed.

---

[2]In particular, `PndTracking` and not `PndTracking2` is used.
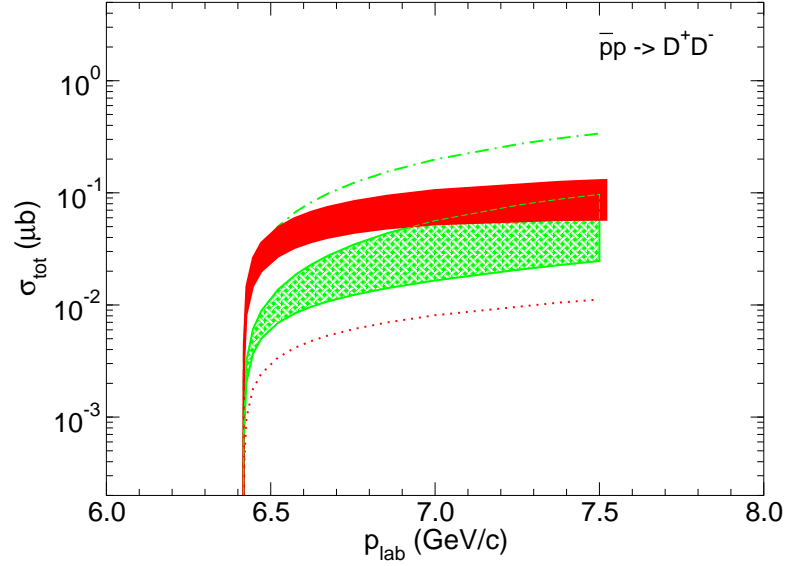
Figure 5.2: Production cross section of charged D mesons in the reaction $\bar{p}p \rightarrow D^+D^-$. The green grid denotes the spectrum obtained by the quark model method, the red solid band refers to results from the baryon-exchange model [148].

5.2.1   Production

D meson pairs are produced using a custom decay file in EvtGen to create only particles of the signal channel. In the experiment, the production cross section determines the ratio of events with D mesons to other events. The cross section is the topic of the following section, then D mesons are analyzed after that.

5.2.1.1   Production Cross Section

Experimental data for the cross section of D^+D^- production from $\bar{p}p$ annihilation is not available. Different theoretical models do exist, but they differ significantly.

Figure 5.2 shows the reaction cross section as a function of the momentum of the $\bar{p}$ beam in the laboratory frame, as obtained by HAIDENBAUER and KREIN in [148]. In the work, two different production models are studied: A baryon-exchange model and a quark model, red and green in the picture, respectively. The first model calculates D^+D^- production from $\bar{p}p$ by possible exchanges of Λ_c and Σ_c baryons. The quark model approach derives a potential for charm production from the constituent quark model, in which two pairs of light quarks (u, d) annihilate and create a pair of heavier quarks (s, c). The results vary from approximately 10 nb to 100 nb in the momentum range from 6.442 GeV/c to 7.5 GeV/c, with the quark-model-based calculations being systematically lower than those based on the baryon-exchange model. Just over threshold, both cross sections rise rapidly, but for greater beam momenta only a limited increase is obtained. During this thesis, a value of $\sigma = 100$ nb is taken.
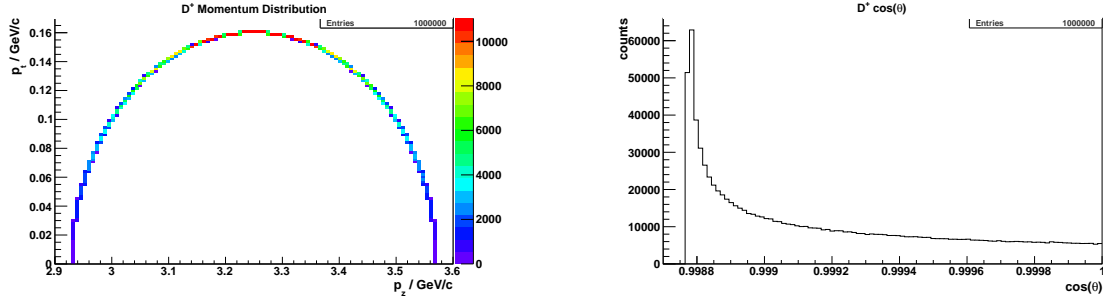
(a) Correlation of transverse momentum versus longitudinal momentum for D^+ mesons. For projections onto the individual axes, see Figure 5.4.

(b) D^+ angular distribution.

Figure 5.3: Momentum and angular distribution of EvtGen-generated D^+ meson.
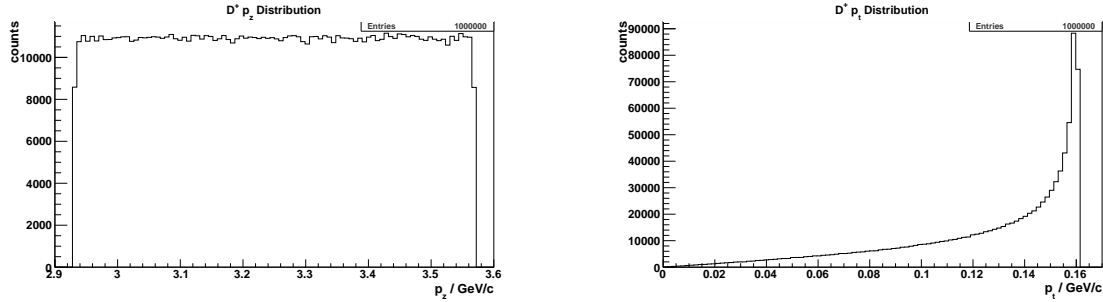
(a) p_z distribution for D^+.

(b) p_t distribution for D^+.

Figure 5.4: Momentum distributions of D^+ from EvtGen.

5.2.1.2 Production in Generator

In this study, the signal and background events are generated separately. The beam momentum has been set to $6.5\,\mathrm{GeV}/c$, just above the production threshold,

$$\vec{p}_{\text{threshold}} = 2m_D \sqrt{\left(\frac{m_D}{m_p}\right)^2 - 1}$$
$$= 6.445\,\mathrm{GeV}/c.$$

The momentum distributions of the D^+ are shown in Figure 5.3 and Figure 5.4. The momentum distributions of the D^- mesons are equivalent. A sharp half-ellipse is obtained for this two-body intermediate state in the two-dimensional plot of Figure 5.3(a), comparing longitudinal momentum (shown in Figure 5.4(a)) versus transverse momentum (shown in Figure 5.4(a)). Longitudinally, a uniformly filled momentum spectrum is generated. The transverse momentum peaks at about $150\,\mathrm{MeV}/c$, defined by the excess energy of the reaction. As seen in Figure 5.3(b), D^+D^- are produced with shallow angles with respect to the beam axis.
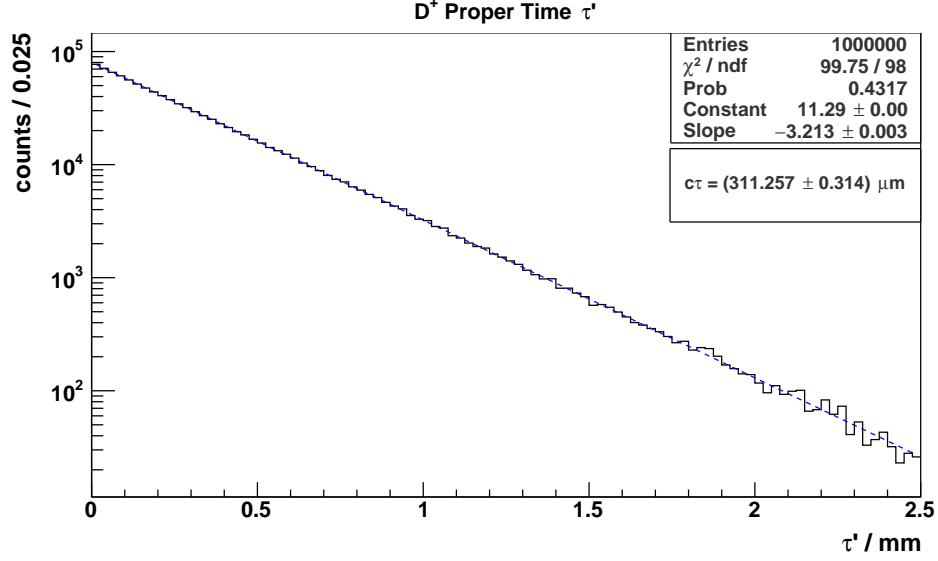
Figure 5.5: Proper time τ' distribution of the decay point of D$^+$. Logarithmic plot with fitted exponential function (dashed line).

The current PDG value for decay lengths of D mesons is $c\tau = 311.8\,\mu$m [1]. To obtain a comparison of the generated decay length to the PDG value, the proper time τ' is calculated and shown in Figure 5.5, with

$$\tau' = \frac{|\vec{x}_{\text{decay}}|}{|\mathbf{P}|/m},$$

using the D mass $m = 1869.62\,\text{MeV}/c^2$, the magnitude of the distance between the interaction point and the decay vertex $|\vec{x}_{\text{decay}}|$ and the D meson four-momentum $|\mathbf{P}|$. A fit with an exponential function yields a decay length of $(311.30 \pm 0.31)\,\mu$m, which is in reasonable agreement with the PDG value.

5.2.2 Decay of D Mesons

The parameters of the decay of D$^+$ to K$^-\pi^+\pi^+$ and D$^-$ to K$^+\pi^-\pi^-$ used in EvtGen are based on data from the CLEO-c experiment, published in »Dalitz plot analysis of the $D^+ \rightarrow K^-\pi^+\pi^+$ decay« [149].

5.2.2.1 Invariant Mass Distributions

The Dalitz plot of this decay is shown in the left plot (»a«) of Figure 5.6. The authors of the paper extracted resonant sub-structures of the decay, which are included in the EvtGen implementation. The resonances that are included are K*(892), K*(1430), K$_2^*$(1430), K*(1680), and an unspecified low mass resonance, κ, at 800 MeV/c^2.

Since the two pions in the final state are indistinguishable from each other, the Dalitz plot is mirrored along the $y = x$ diagonal axis. The two possible Kπ combinations are ordered by
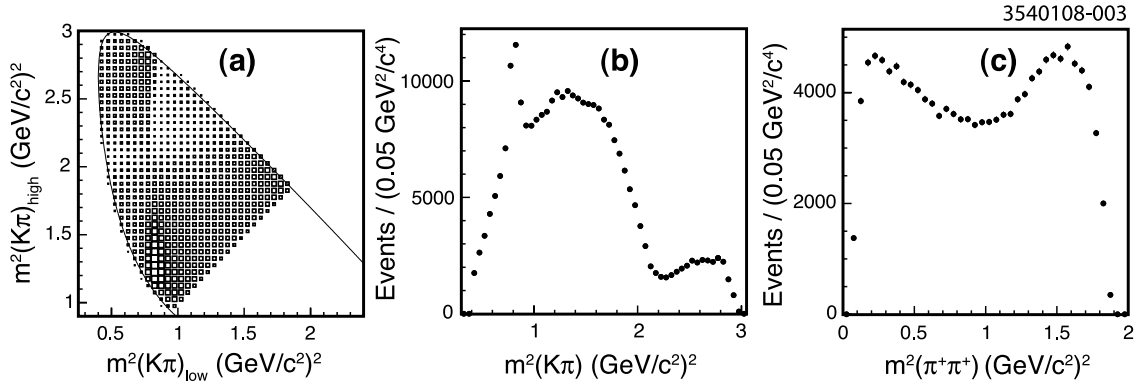
Figure 5.6: Invariant mass distributions of $\bar{p}p \to D^+D^- \to K^-\pi^+\pi^+K^+\pi^-\pi^-$ from CLEO-c. The corresponding plots of the results from the implementation in EvtGen are shown in Figure 5.7, Figure 5.8(a), and Figure 5.8(b) [149].
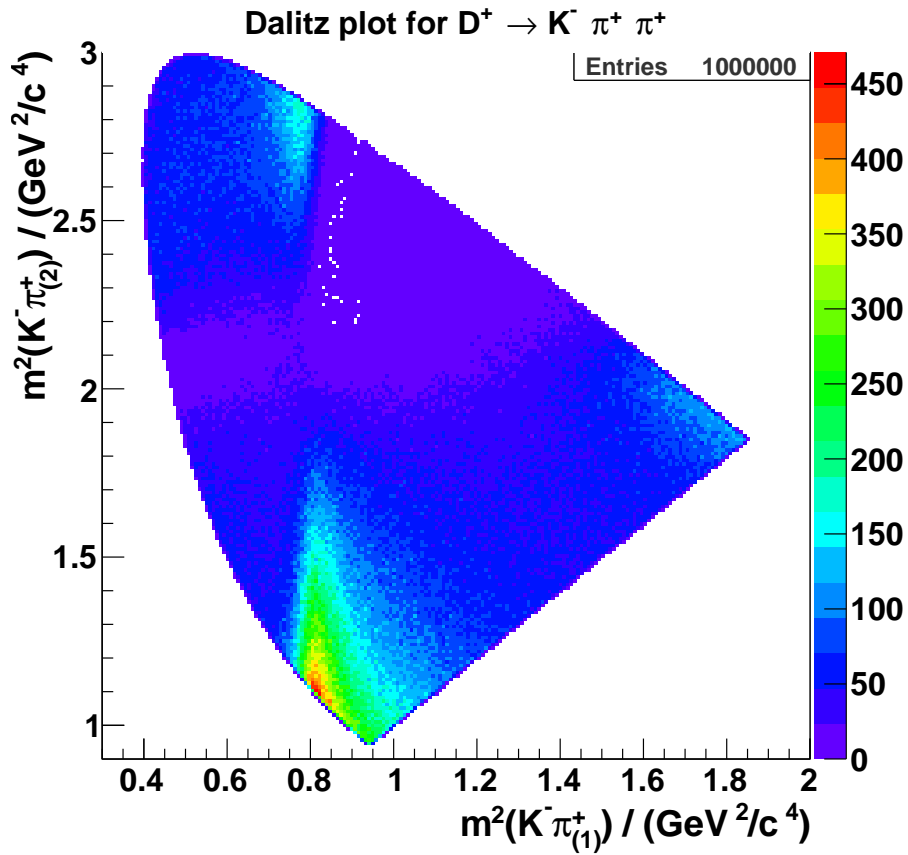
Figure 5.7: Dalitz plot for $D^+ \to K^-\pi^+\pi^+$. Since the two pions in the final state are indistinguishable, the distribution is symmetric around the diagonal. $K^-\pi^+_{(1)}$ is the lighter combination.
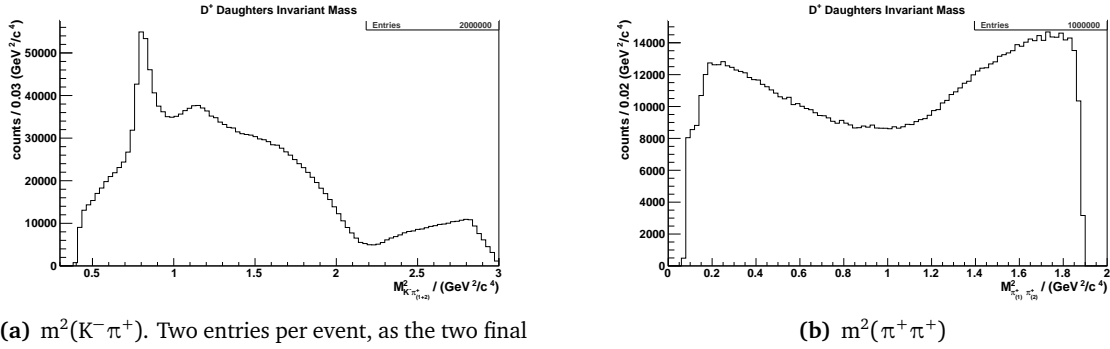
(a) $m^2(K^-\pi^+)$. Two entries per event, as the two final state pions are indistinguishable.

(b) $m^2(\pi^+\pi^+)$

Figure 5.8: Squared masses of the two two-particle-combinations of the D^+ daughter particles.

their invariant mass. The pair with the higher mass combination is plotted on the y-axis and the pair with the lower combination is plotted on the x-axis. Figure 5.7 reproduces the original Dalitz plot from CLEO.

The one-dimensional invariant mass distributions of the two $K\pi\pi$ sub-systems are shown in the center and on the right in Figure 5.6 for the CLEO original data, and in Figure 5.8 for the EvtGen generator data. The $K\pi$ invariant mass distribution has two entries per event, as it includes both possible $K\pi$ combinations. The prominent features of Figure 5.6 can be seen in Figure 5.8 as well, e.g. the peak around $0.8\,\text{GeV}^2/c^4$ and the dip at $2.2\,\text{GeV}^2/c^4$ in the $K\pi$ invariant mass spectrum.

5.2.2.2 Momentum Distributions of Daughter Particles

The momentum distribution of the kaon daughter is shown in Figure 5.9. The distribution of the magnitude of the momentum in Figure 5.9(a) has its maximum around $1\,\text{GeV}/c$. Most of the momentum is carried in the longitudinal direction (see Figure 5.9(c)), which peaks around $0.9\,\text{GeV}/c$. Accordingly, the p_t distribution is highest around $0.2\,\text{GeV}/c$. The comparison of longitudinal and transversal momentum in Figure 5.9(b) resembles the structure seen in the individual projection plots. Most of the kaons have momenta around $p_z = 0.8\,\text{GeV}/c$ and $p_t = 0.2\,\text{GeV}/c$. The *spot* is complemented by a band of particles with larger p_t around the upper left edge of the momentum space. It is created by kaons coming from the resonant sub-structure located top left in the Dalitz plot of Figure 5.7.

The momentum distributions of pions have similar shapes, as shown in Figure 5.10. The p_z momentum plot in Figure 5.10(c) peaks at lower values, around $0.2\,\text{GeV}/c$. The transverse momentum is maximum at approximately $0.5\,\text{GeV}/c$ (Figure 5.10(d)). The momentum comparison of Figure 5.10(b) illustrates the origin of the projections. Most pions are located in the band to the upper left. A small contribution from pions of low momenta can be seen at $p_t = 0.1\,\text{GeV}/c$ and $p_z = 0.3\,\text{GeV}/c$. Indeed, these are the pions complementing the kaons in the border region of Figure 5.9(b) from the resonant structure.

The distribution of the ratio of the transverse momenta of both pions in a D decay is shown in Figure 5.11. For this plot, the momentum of the pion with higher momentum, $p_t(\pi_{(1)})$,

(a) Momentum distribution of the K^-.

(b) p_t vs. p_z distribution of the K^-.

(c) p_z distribution of the K^-.

(d) p_t distribution of the K^-.
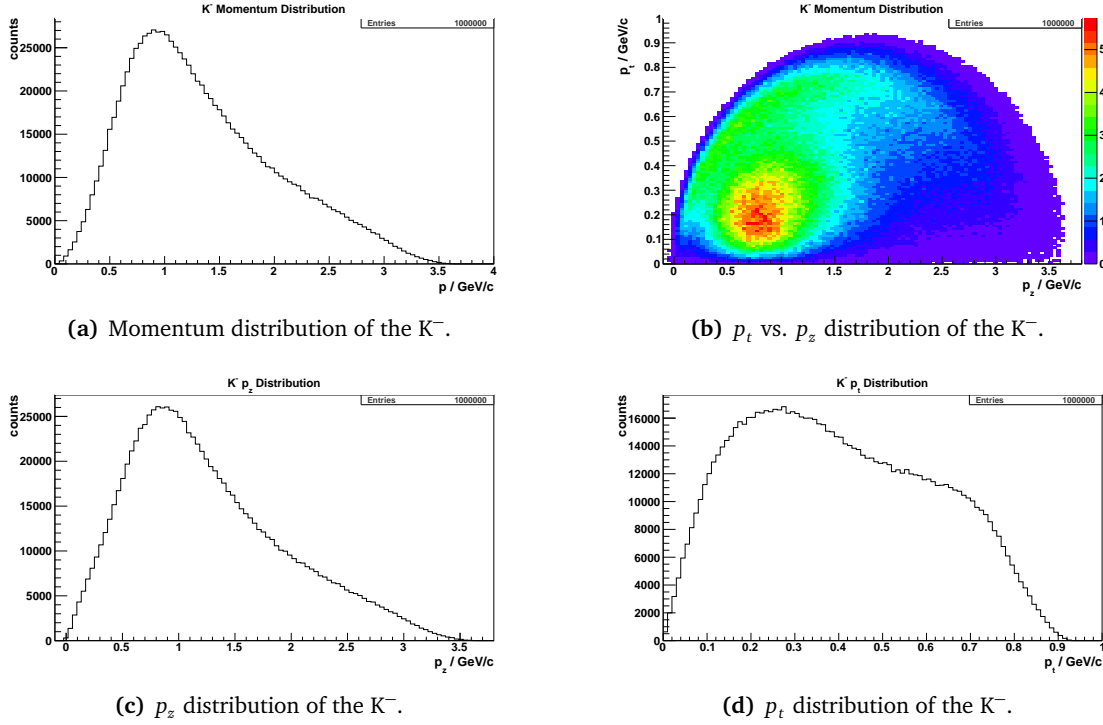
Figure 5.9: Momentum distributions of the K^- coming from the D^+ decay (laboratory frame).

is divided by the momentum of the pion with lower momentum, $p_t(\pi_{(2)})$. The figure shows the ratio for pions with positive charges (blue) and negative charges (red). In most cases, the momentum distributions of the two particles are very similar, as the distributions clearly peaks at 1. However, the tail is non-negligible and spreads out to large ratios. As expected, the distributions of the particles of different charges overlap. Except for statistical fluctuations, the event generator does not discriminate the charge.

5.2.2.3 D Meson Decay Vertex Distance

The decay point of the D mesons is statistically distributed in space. The same is true for the distance between the decay vertices of the D^+ and D^- mesons. Figure 5.12 sketches the definition of $\overrightarrow{\Delta R}$, the three-vector difference between the decay vertices of the D^+ and the D^-.

The distribution of $|\overrightarrow{\Delta R}|$ is shown in Figure 5.13(a). The exponential decay observed is the convolution of the two decay length distributions of the D mesons. Distributions for the longitudinal distance and transverse distance are given in Figure 5.13(b) and Figure 5.13(c), respectively. $(\Delta R)_z$ is distributed around 0; the decay positions in z are symmetrically arranged for the two charged states of the D mesons. The majority of D meson pairs have a longitudinal distance of <1 mm, with many entries >100 μm, the measurement precision of $\overline{P}ANDA$. With proper detector calibration, decay vertex offsets should be measurable in the z-direction. In contrast, the transverse separation is mostly less than 100 μm. The most probable value is approximately 20 μm, well below the resolution of $\overline{P}ANDA$.
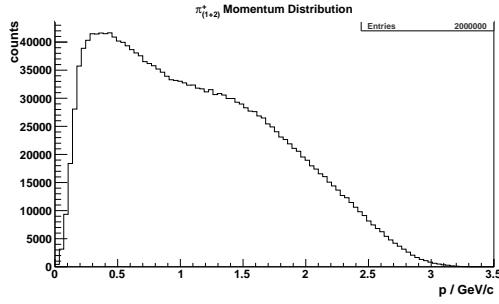
(a) Momentum distribution of the π^+.
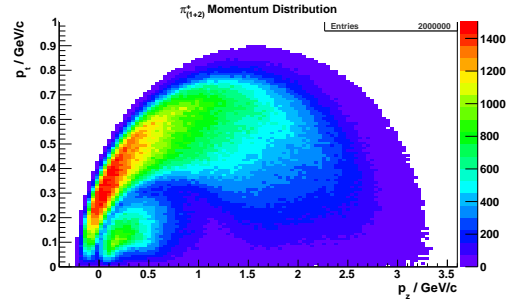
(b) p_t vs. p_z distribution of the π^+.
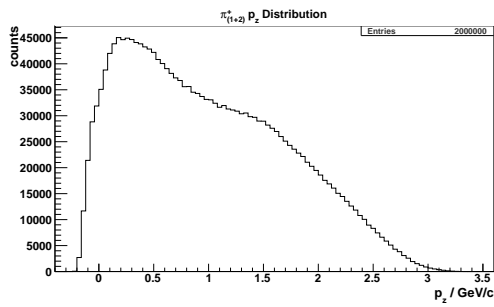
(c) p_z distribution of the π^+.
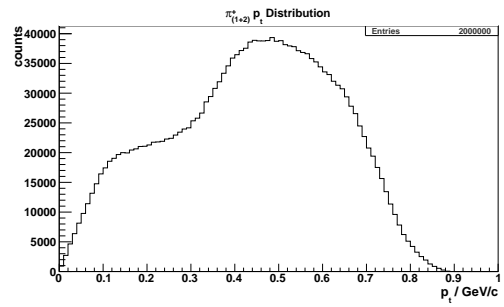
(d) p_t distribution of the π^+.

Figure 5.10: Momentum distributions of the π^+ coming from the D^+ decay (laboratory frame). Two entries per event.
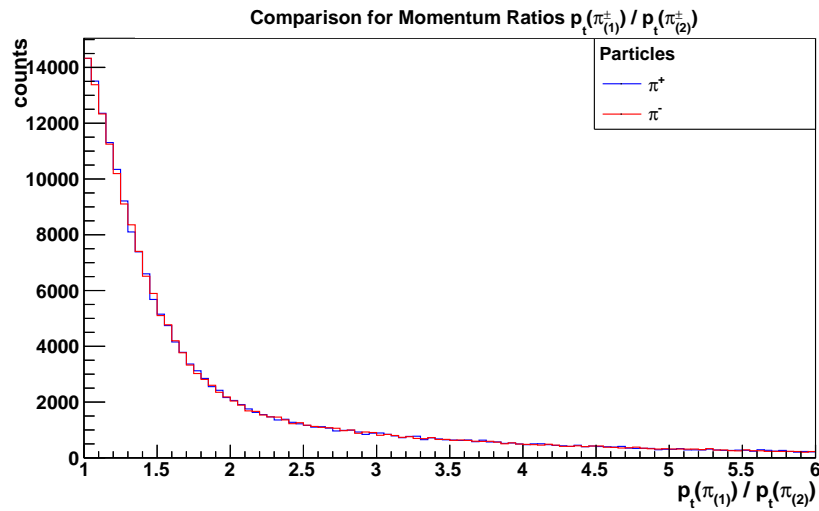
Figure 5.11: Comparison of the ratio of transverse momentum of the first to the second π, with $p(\pi_{(1)}) > p(\pi_{(2)})$. Two overlapping graphs are shown for π^+ (blue) and π^- (red).
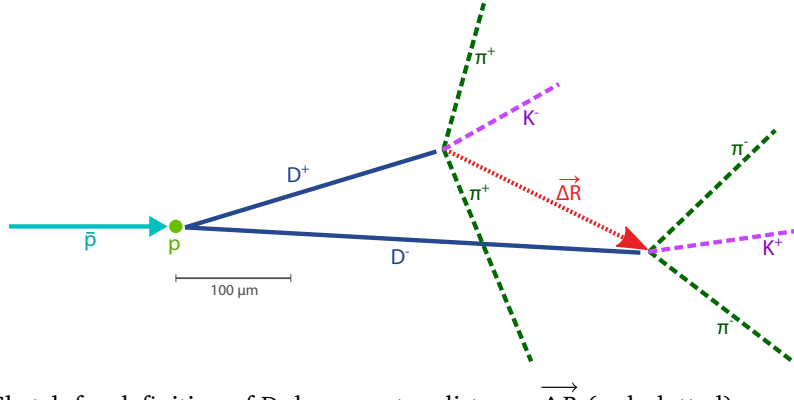
Figure 5.12: Sketch for definition of D decay vertex distance $\overrightarrow{\Delta R}$ (red, dotted), used in Figure 5.13. The D daughter particles are shown in dashed lines in green (π) and magenta (K).
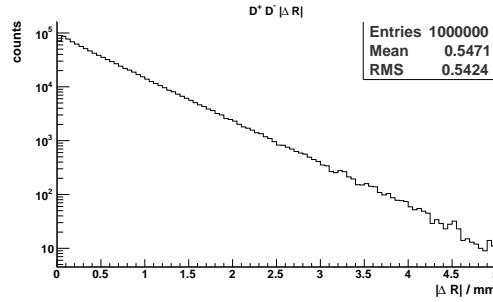
(a) Distribution of the distance between the decay vertices, $|\Delta R|$.
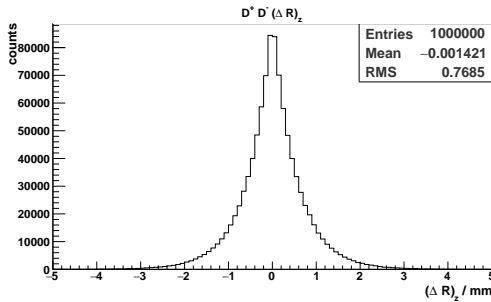
(b) Distribution of the longitudinal distance between the decay vertices, $(\Delta R)_z$.
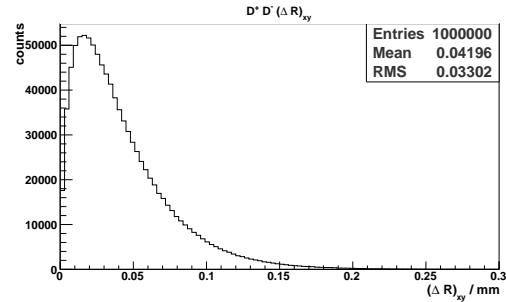
(c) Distribution of the distance between the decay vertices in the xy plane, $(\Delta R)_{xy}$.

Figure 5.13: Distributions of the distances between the decay vertices of the D^+ and D^- in three-dimensional space, longitudinally and in the xy-plane. $\overrightarrow{\Delta R} = \overrightarrow{vtx}(D^+) - \overrightarrow{vtx}(D^-)$, see Figure 5.12.

5.3   Hit Multiplicities in the Detectors

Before considering inefficiencies in the reconstruction of particle candidates, the distribution of hit point multiplicities from the daughter particles of the D mesons is analyzed in this section. The position of the decays of the pions and kaons are also investigated.

5.3.1   Multiplicity of Hits in the Different Sub-Detectors

$\overline{\text{P}}$ANDA has four main sub-detectors dedicated to track reconstruction, the MVD, the STT, the GEM and the FTS. The four detectors are introduced in Section 3.3.2. The MVD[3] has four barrel layers surrounding the interaction point and six disks in the forward direction. The last disk is located at $z = 23$ cm, the outer barrel layer at $r = 13.5$ cm. The STT instruments the volume from $r = 15$ cm to $r = 42$ cm, and z from -55 cm to 110 cm. A particle traversing the full STT on a straight line trajectory can have between 23 and 29 hits. At low p_t, particles can have more hits. The GEM stations are installed at positions with $z = 0.9$ m, 1.1 m, and 1.5 m. A station is comprised of two detecting layers, generating two hit points per station. Finally, the FTS has six stations in the forward part of $\overline{\text{P}}$ANDA, located between $z = 3$ m and $z = 7.5$ m. Each station has straws aligned into four double-layers, a particle on a straight trajectory passing all stations leaves 48 hits.

The following plots are generated with 200,000 D^+D^- events, corresponding to 800,000 pions and 400,000 kaons.

Figure 5.14 shows the distribution of the number of hits per track for pions. Most (75 %) of the particles leave 4 ± 1 hits in the MVD. The low fraction of pions with no hits in the MVD is notable (<4 %), underlining the importance of the MVD for particle reconstruction. Most of the pions generating hits leave 26 hits in the STT detector; 40 % of all particles leave from 24 to 27 hits. Approximately 18 % of pions do not leave any hits in the STT, they are emitted at shallow forward angles or pass the detector through not-instrumented slits (see below). About 2 % of pions have more than 35 hits, as they have low p_t and consequently curl up in the detector. If a pion has a hit in the GEM, which occurs for 33 % of poins, it generally leaves hits in all three double layers (70 %). Only 0.03 % of the pions leave more than 6 GEM hits. 92 % of the pions do not leave any hit points in the FTS, since most of them do not reach this sub-detector system. If they do, the most probable value is located around 42 hits.

A generally comparable picture is obtained for the hit count distributions of kaons, as seen in Figure 5.15. Two difference can be noted. First, for 30 % of the tracks, kaons leave no hits in the STT – nearly twice as often as pions. Secondl, the percentage of 6 hits in the GEM is 34 %, 11 percentage points higher than the comparable number for pions. Both properties show that K are emitted at lower polar angles. Consequently, the forward detectors of the GEM are relatively more important for K than for π. This effect can already be seen in the momentum distribution of the kaons from EvtGen in Figure 5.9: Kaons have higher longitudinal and lower transverse momentum than pions. Figure 5.16 relates the hit count distribution of the STT to that of the GEM. If one of the two sub-detectors is hit, usually the other one is not. A large fraction of kaons leave no hits in either of the two sub-detectors.

---

[3]In this section, the tracking sub-detectors are marked with a color key to distinguish one from another. The colors in the text match the colors of the distributions in the figures.
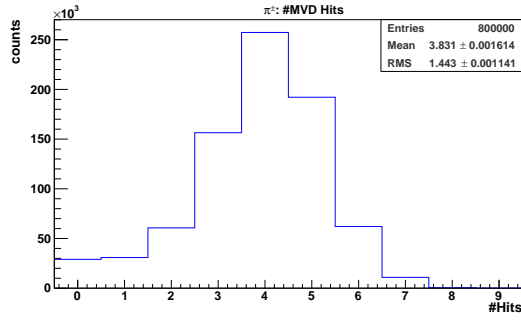
(a) Hit multiplicity distribution in the MVD.
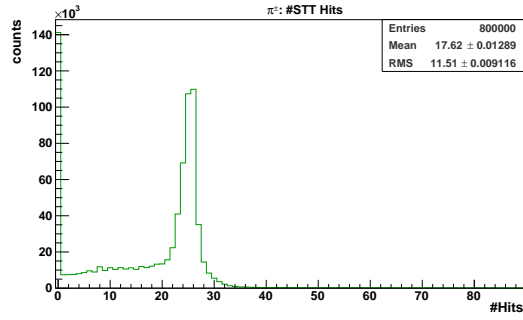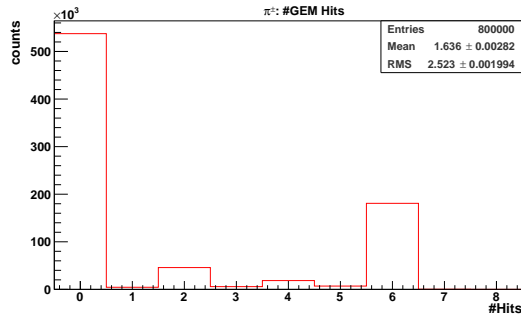
(b) Hit multiplicity distribution in the STT.

(c) Hit multiplicity distribution in the GEM.
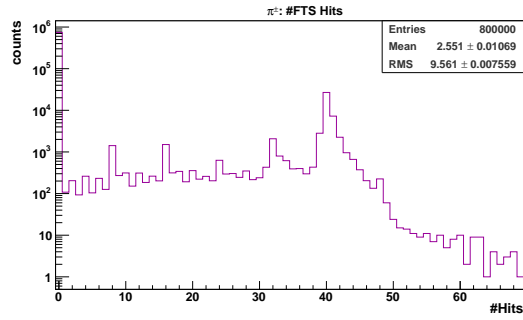
(d) Hit multiplicity distribution in the FTS (n.b. logarithmic y-axis).

Figure 5.14: Hit multiplicity distributions per track for the four main tracking sub-detectors for **pions**.

In order to further analyze the pion tracks, that have no STT hits, the number of FTS hits is plotted versus the number of GEM (Figure 5.17). One important point is shown: If the FTS is hit, the particle almost always hits the GEM – the three stations lie inbetween the interaction point and the FTS. In most cases where no STT hit points are recorded, also no FTS and no GEM hits are produced.

The momentum distribution of pions, leaving no hits in the STT is given in Figure 5.18(a). Many of these pions have large longitudinal momenta and are emitted with small angles. The bulk of the pions have however a small total momentum, located at a hotspot around $p_t = p_z = 0.1 \, \text{GeV}/c$. As shown in Figure 5.18(b), where pions without hits in the STT, GEM, or FTS are selected, these low-momentum pions are particles which reach nothing but the MVD. About 7 % of all pions have hits solely in the MVD. 1.5 % of all pions leave no hit points in any of the tracking sub-detectors. For kaons this number is 2.5 %.

5.3.1.1 Angular Distribution

The angular distribution of hits in the individual sub-detectors is shown in Figure 5.19 for pions. In addition to the four graphs for the MVD, STT, GEM, FTS, the sum of **all** four is given. Different patterns can be noted. The FTS is most important for shallow polar angles below 6°, where it gives up to 35 hit points. For larger polar angles, from 23° to 125°, the STT provides constantly the largest number of hit points (22 to 26). Covering a large polar angular range
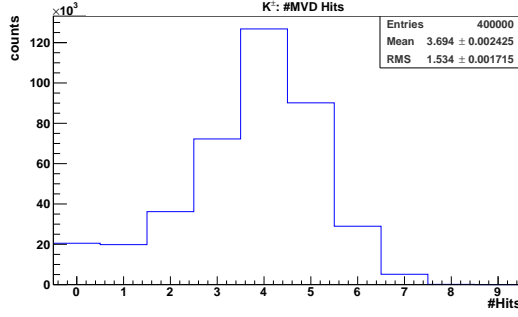
(a) Hit multiplicity distribution in the MVD.
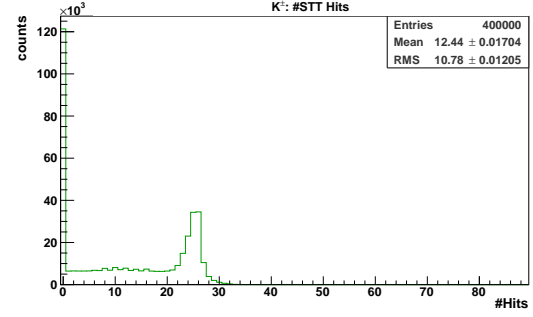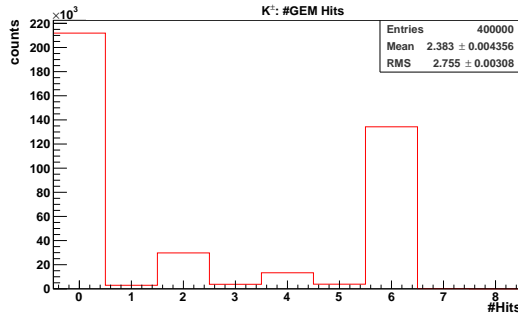
(b) Hit multiplicity distribution in the STT.

(c) Hit multiplicity distribution in the GEM.
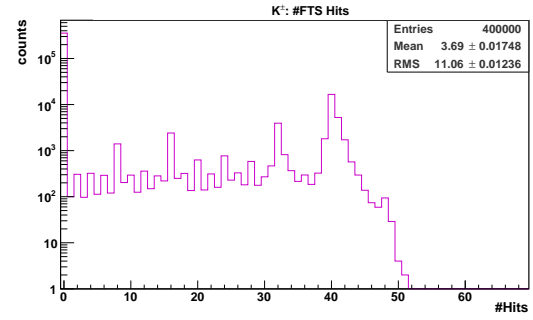
(d) Hit multiplicity distribution in the FTS (n.b. logarithmic plot).

Figure 5.15: Hit multiplicity distributions per track for the four main tracking sub-detectors for **kaons**.
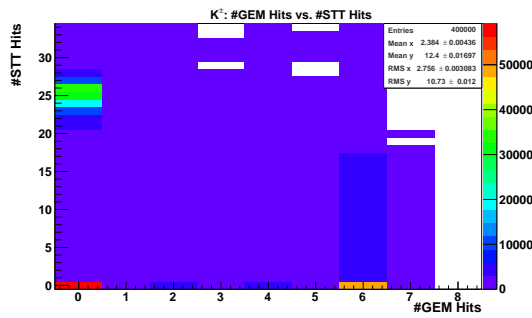
Figure 5.16: K: Number of hits per track in STT versus number of hits per track in GEM.
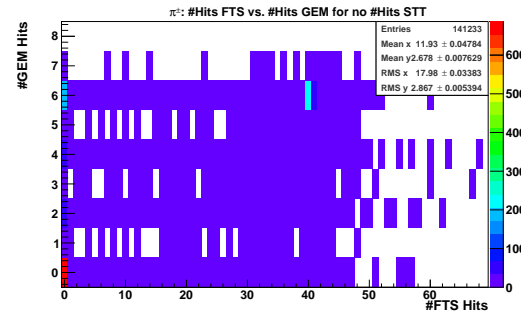
Figure 5.17: π: Number of hits per track in GEM versus number of hits per track in FTS, for tracks having no STT hits.

(a) Distribution for no STT hits.

(b) Distribution for no STT, no GEM, no FTS hits.
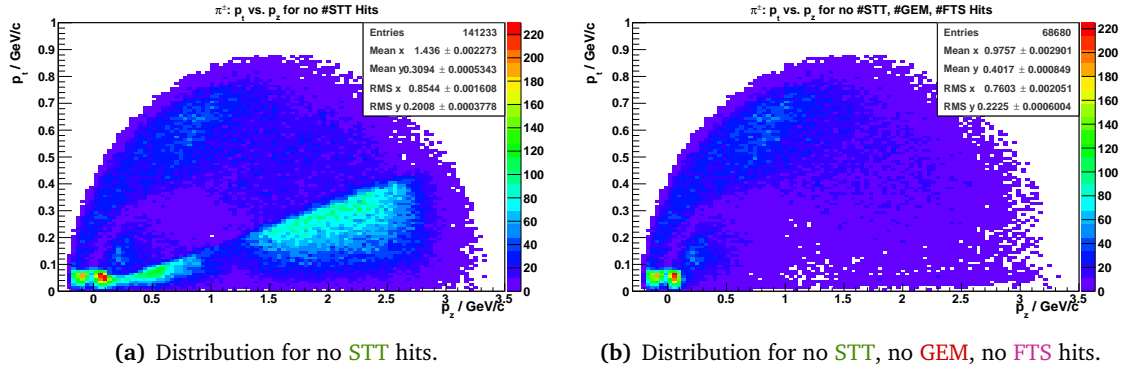
Figure 5.18: π: p_t vs p_z distribution for tracks with no STT hits **(a)** and no STT, no GEM, no FTS hits **(b)**.
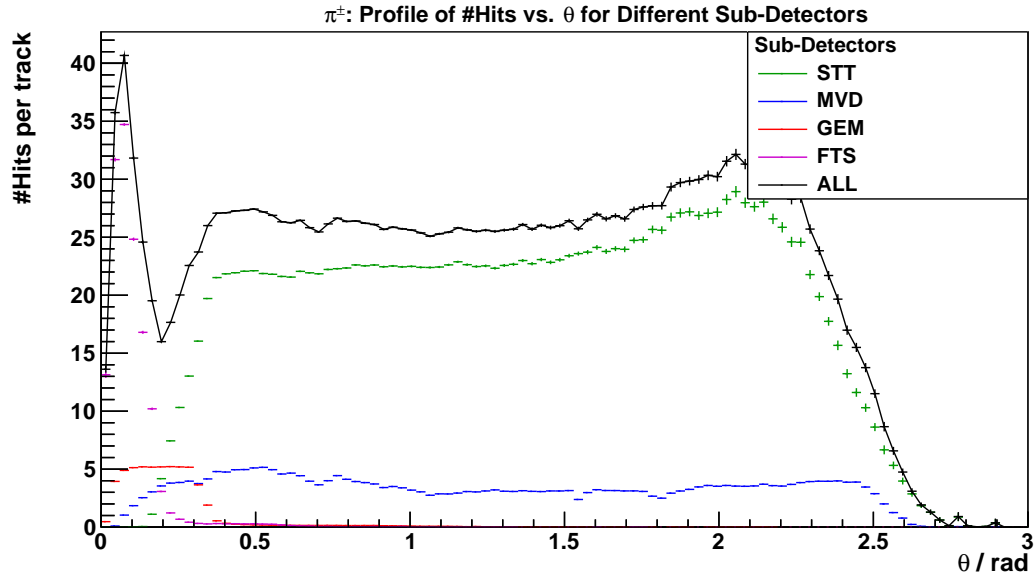
Figure 5.19: π: Profile histogram of the mean number of hits per track for individual tracking sub-detectors as a function of θ. The error bars along the y-axis of each data point represents the uncertainty of the mean at this point. The two-dimensional histograms used for this plot are given in Figure A.9.
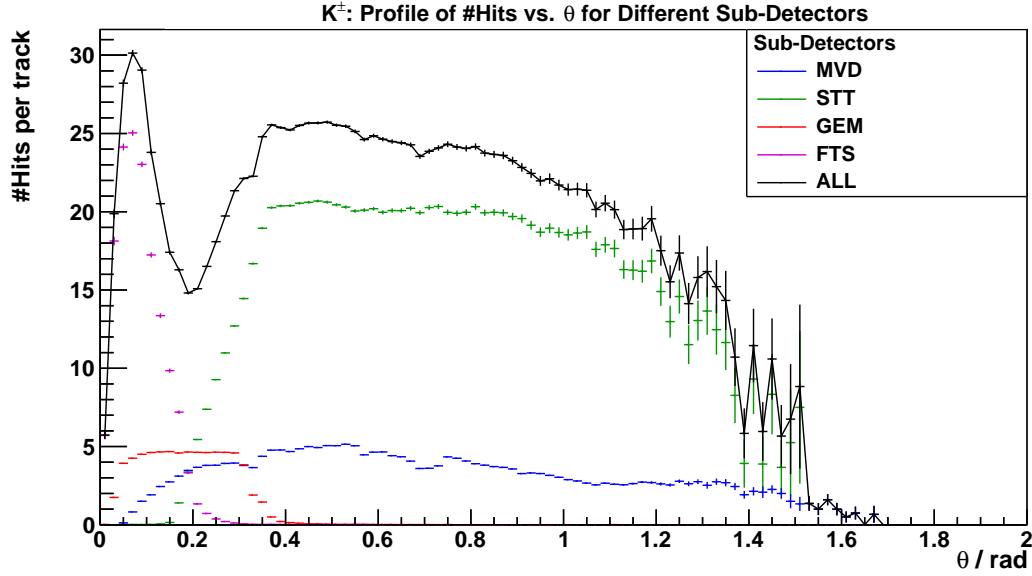
Figure 5.20: K: Profile histogram of the mean number of hits per track for individual tracking sub-detectors as a function of θ. The two-dimensional histograms used for this plot are given in Figure A.10.

from 11° to nearly 140°, the MVD delivers on average 4 hit points. The intermediate region between 3° and 12° is covered by the GEM. In total, the setup of the tracking sub-detectors always provides on average more than 16 hit points over the whole angular range from nearly 0° to 140°. Above 140°, the number of hits per track quickly falls to zero, as the limit of the volume covered by tracking detectors is reached.

The same principle patterns can be observed for kaons, as shown in Figure 5.20. Two differences emerge: the overall number of hits for kaons is smaller over the whole angular range and the hit multiplicity starts dropping already around 60°. This is due to the low lab momentum of the kaon at these polar angles and the life-time of the particle, which is approximately a third that of the pion. Since the particles decay earlier, they do not leave as many hit points in the detectors.

5.3.1.2 Hit Efficiencies

The number of hits that particles leave in the tracking sub-detector directly affects the efficiency and resolution of track reconstruction. The higher the number of hits, the better a track can be reconstructed. Before including any detector response, the efficiency of hits in the tracking sub-detectors can be defined. The *hit efficiency* is the number of hits over a certain threshold divided by the total number of hits. The hit efficiency is shown in Figure 5.21 as a function of θ for three different discrimination values: The number of tracks with more than 4, more than 6, and more than 8 hits.

Comparing the efficiency for pions (Figure 5.21(a)) with the efficiency for kaons (Figure 5.21(b)) shows the known difference in angular coverage. Demanding already >4 hits has

(a) Ratio for π.

(b) Ratio for K.
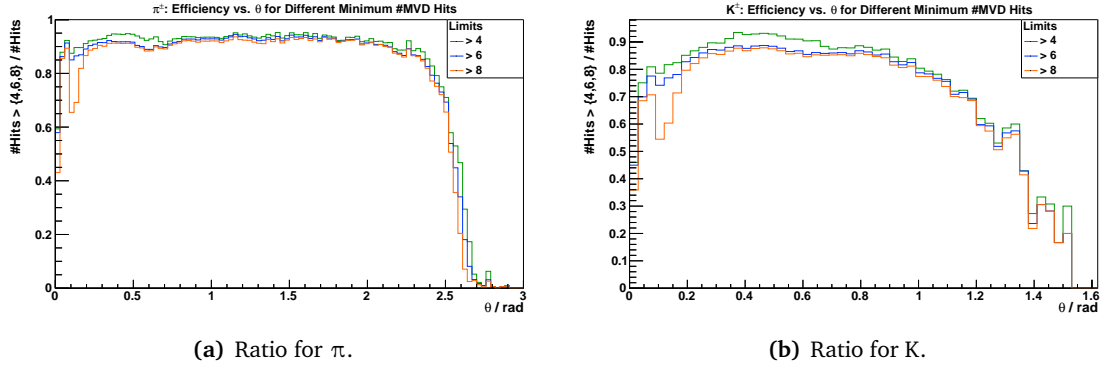
Figure 5.21: Ratio of the number of tracks with more than 4 (green), 6 (blue), and 8 (orange) hits to the number of all tracks as a function of the polar angle θ for π (left) and K (right).

significant impact on the efficiency, as the K efficiency is only over 90 % for the range $\theta \approx 23°$ to $\theta \approx 29°$. For pions, the picture looks differently, as for most of the θ range the efficiency is over 90 %. Demanding more than 8 hits has the largest consequences, independent of the particle type, in the region around $\theta = 0.1$ rad, where the FTS to STT transition is mitigated by the GEM. Since the GEM only has a maximum of 6 hit points, this constraint discriminates a considerable fraction of particles.

Already based on the raw number of hits, without any detector or algorithm specific effects, it can be seen that the reconstruction of kaon tracks is more challenging than the reconstruction of pions.

5.3.2 Pion Charge Asymmetry

In the previous section, the produced hits of kaons and pions have been investigated independent of the charge of the particles. The distribution of the number of hits in the MVD[4] is plotted in Figure 5.22 and Figure 5.23 for π and K, respectively, for both charges. The positively charged particles are shown with blue lines and the negatively charged particles with red lines.

While the distributions for different kaon charges in Figure 5.23 overlap and only deviate statistically, the picture is dramatically different for pions in Figure 5.22. π^- particles have an excess in relative multiplicity for 4 MVD hits. Accordingly, π^+ more frequently have 1 or 2 MVD hits compared to π^-. This **charge asymmetry** is an unexpected effect as both charges should be equal in all stages of the simulation. The only difference between the two charges is the bending direction of the tracks due to the magnetic field. But since the detector is built rotationally symmetric, this should not lead to such an asymmetry. The cause of this asymmetry is explored in the following section.

5.3.2.1 Momentum Dependence

The dependence of the hit multiplicity on the transverse momentum is the subject of Figures 5.24 and 5.25. Figure 5.24 displays the ratio between the number of hits in the MVD for π^+

---

[4]The other sub-detectors are shown in Figures A.11 (π) and A.12 (K).
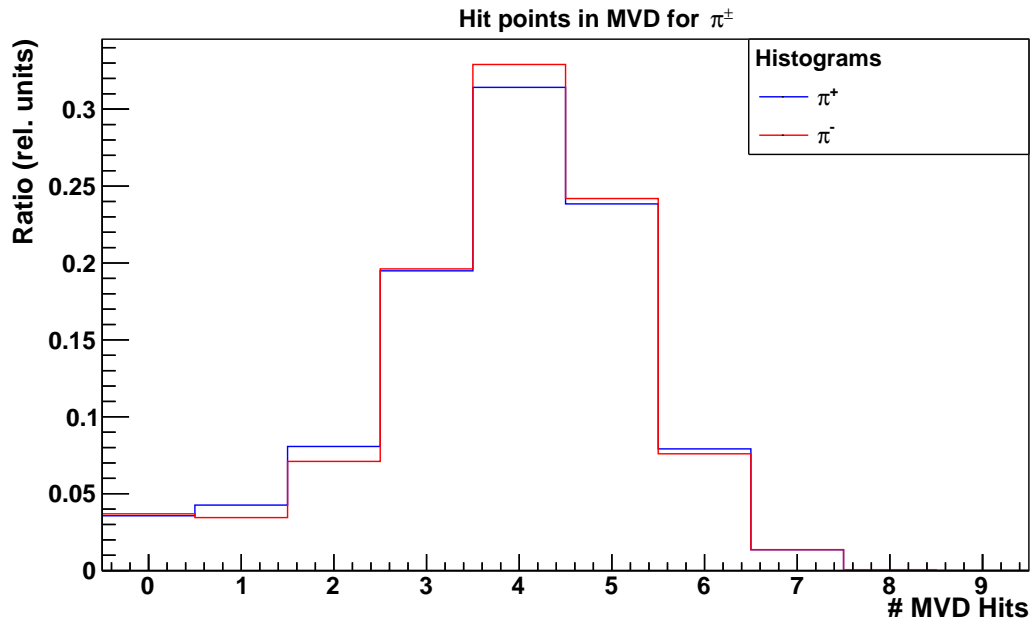
Figure 5.22: π^+ & π^-: Distribution of number of hits per track in the MVD. Graphs for other tracking sub-detectors are given in Figure A.11.
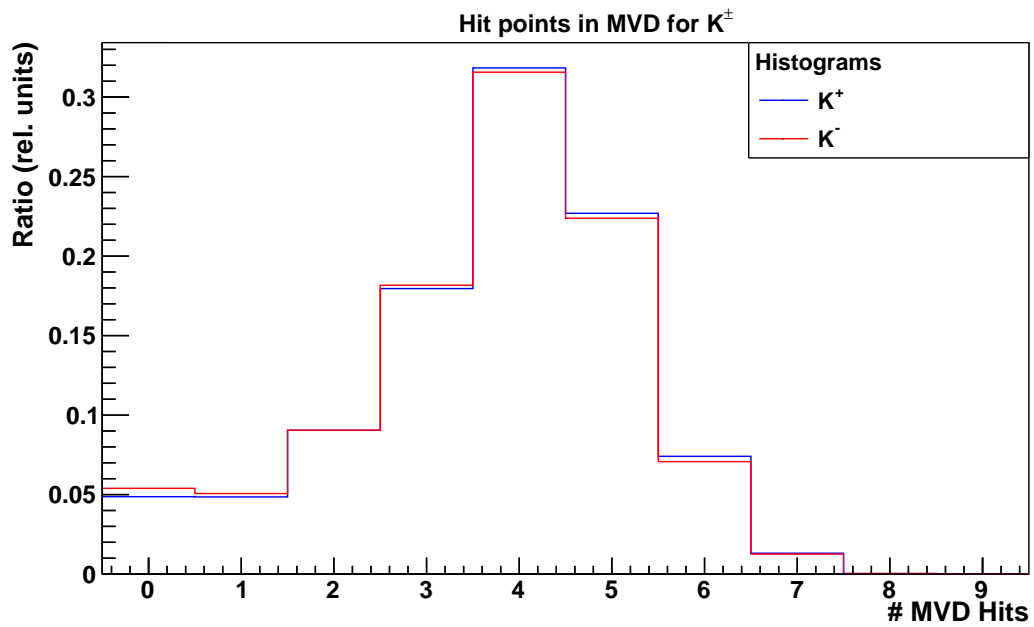
Figure 5.23: K^+ & K^-: Distribution of number of hits per track in the MVD. See Figure A.12 for a comparison with other tracking sub-detectors.
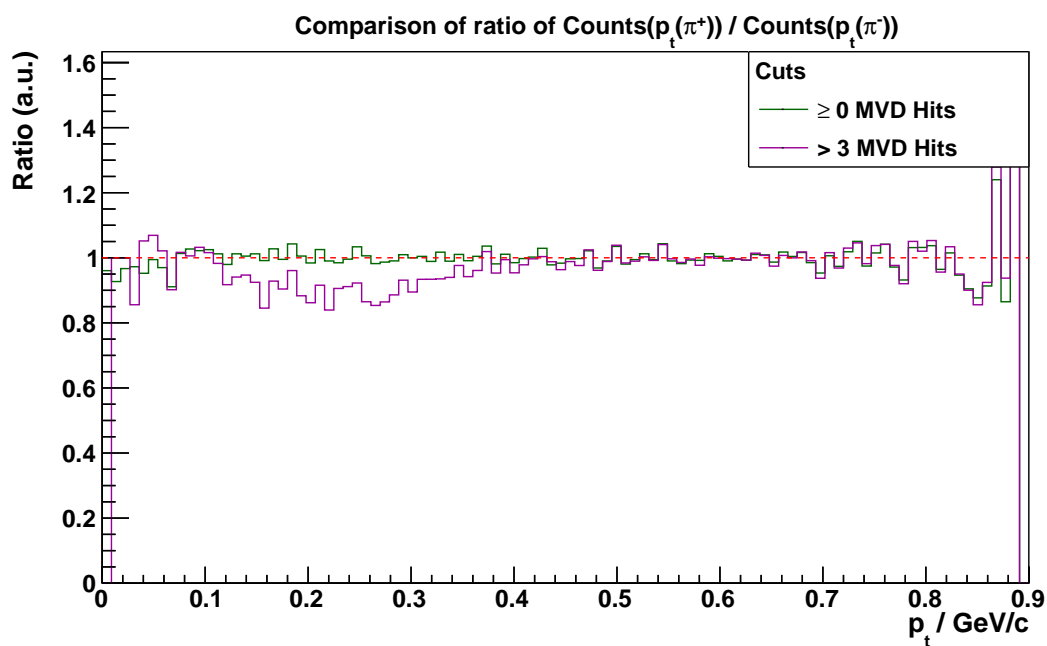
Figure 5.24: π^+ & π^-: Ratio of the number of hits in the MVD for π^+ to π^- as a function of p_t. For all tracks and for tracks with >3 MVD hits. The dashed red line marks the ratio of 1.
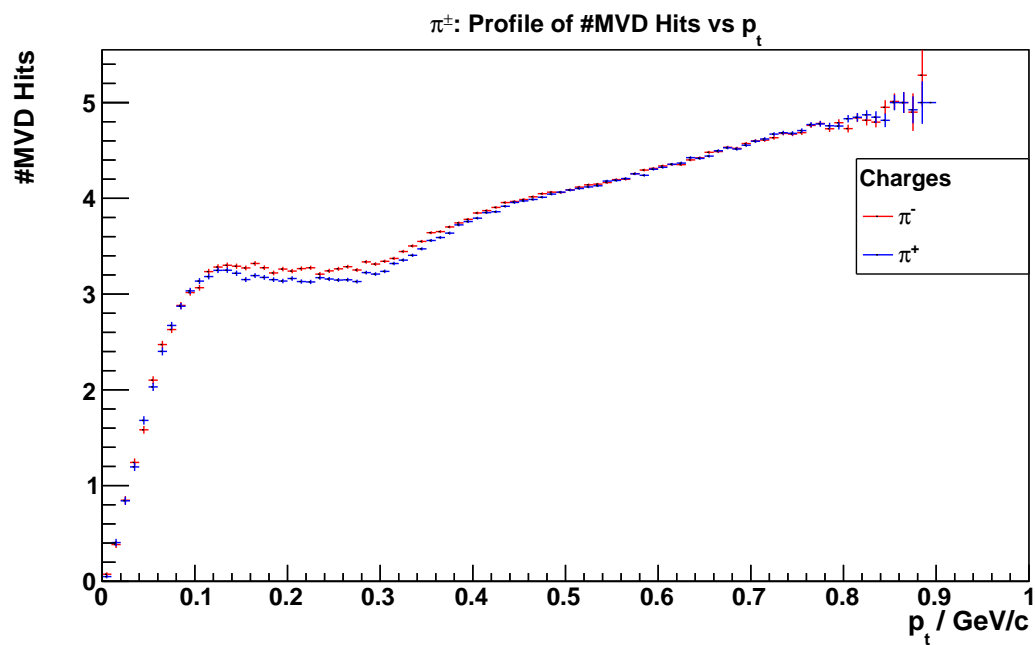
Figure 5.25: π^+ & π^-: Profile histogram of the mean number of MVD hits as a function of p_t.
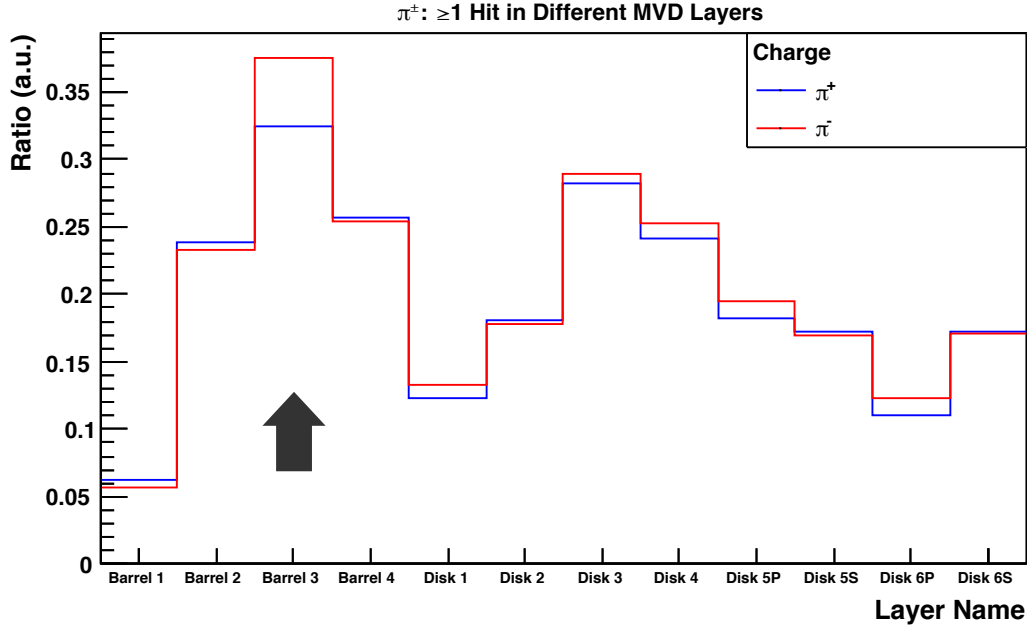
Figure 5.26: π^+ & π^-: Fraction of tracks with at least one hit in the corresponding layers of the MVD for π^+ (blue) and π^- (red). For the last two disks, the addition »S« refers to the strip part of the disk, »P« to the pixel part. The arrow denotes the layer with the largest asymmetry.

to π^-. The green distribution imposes no other constraint, it is symmetric around 1. The magenta-colored graph demands at least four MVD hits. The result is a deficit of π^+ counts in the p_t range of $0.12\,\text{GeV}/c$ to $0.4\,\text{GeV}/c$. The same picture can be seen in the profile histogram of Figure 5.25, which displays the average number of MVD hits on the y-axis versus p_t on the x-axis. The distributions for π^+ and π^- are largely identical, except for the momentum region of $0.12\,\text{GeV}/c$ to $0.4\,\text{GeV}/c$.

Positively charged π^+ seem to leave systematically fewer hits in the MVD in the low-p_t region between $0.12\,\text{GeV}/c$ to $0.4\,\text{GeV}/c$. Particles with lower transverse momenta have tracks with smaller bending radii.

5.3.2.2 MVD Layers

The MVD is composed of four barrel layers and six disks. The inner two barrel layers, as well as the first four disks are built with a pixel segmentation. The outer two barrel layers and the outer part of the last two disks are built with strip sensors (Section 3.3.2.1). Counting the number of hits for the pion charges separately in each layer creates Figure 5.26. It can be seen that the asymmetry in the overall hit count is largely caused by an asymmetry in barrel layer 3. In this layer, 37.5 % of π^- leave at least one hit, but only 32.4 % of π^+. For the other MVD layers, the relation is inverse, but not as pronounced as for barrel layer 3.
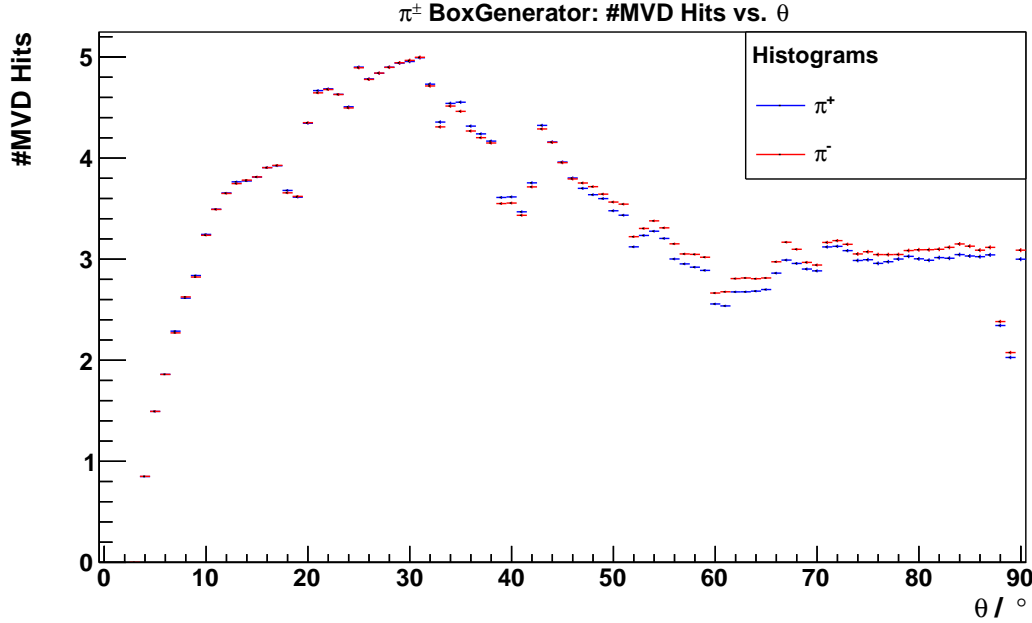
Figure 5.27: π^{+} & π^{-}: Angular distribution of the average number of MVD hits. The two two-dimensional histograms that are the basis for this profile plot are given in Figure A.13.

5.3.2.3 BoxGenerator Study

To investigate the charge asymmetry further, the BoxGenerator of PandaRoot is used. The MC generator can be employed to expose the $\overline{P}$ANDA detector with specific particles of defined momenta and in limited emission angles. The generator is utilize to produce 88 data sets with 4000 events each containing two π^{+} and two π^{-}. The π momentum in each data set is limited to the range of $0\,\text{GeV}/c$ to $1\,\text{GeV}/c$. For each set, the θ angle of the pions was forced to be a defined value. The range $3° < \theta < 90°$ is sampled in $1°$ steps.

Figure 5.27 shows the distribution of the average numbers of MVD points as a function of the polar angle for both pion charges. While the graphs overlap for small angles, the previously noted count asymmetry emerges for angles greater than $30°$. π^{-} are systematically measured with higher multiplicity than π^{+}.

A data set with high asymmetry ($\theta = 60°$) is the focus of Figure 5.28, where the fraction of events with hits in the given MVD layer is shown. The pattern already seen in Figure 5.26 is repeated here, although only for the sub-set of MVD layers hit by particles with $\theta = 60°$. Again, barrel layer 3 has a large excess of π^{-} hits.

The *BoxGenerator* data set was studied by the event display of PandaRoot. The event display allows for visualization of particles, tracks, and the detector geometries. Figure 5.29 shows an example event where one π^{+} traverses the third MVD layer without leaving a hit point (indicated by a red ellipse). Shown are layer 3 and layer 4 of the MVD barrel. Hit points in the MVD are blue squares and hit points in the STT are purple squares. The xy-projection is
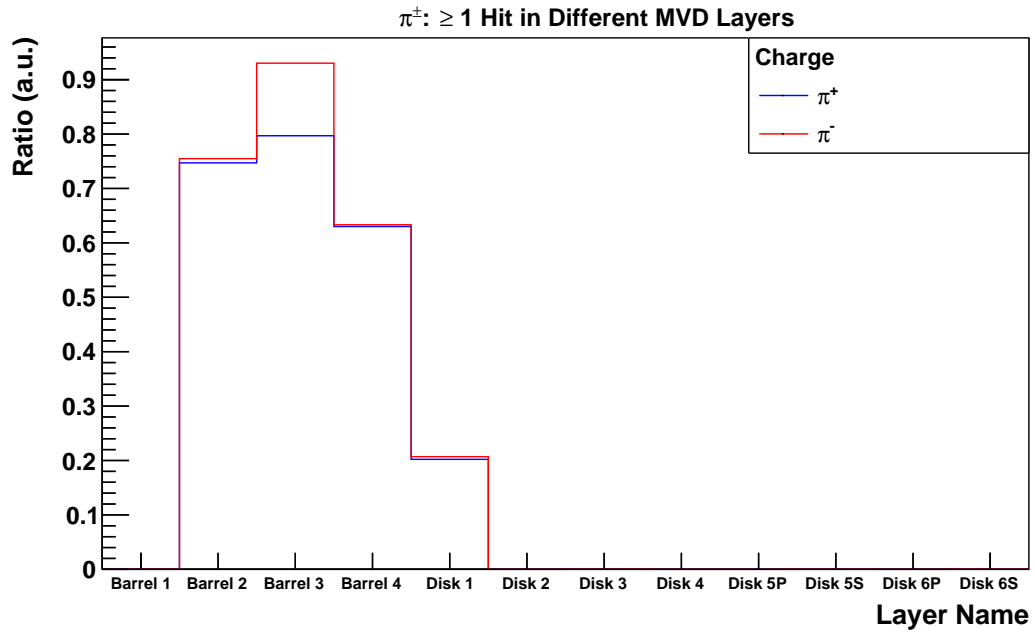
Figure 5.28: π^+ & π^-: Fraction of tracks with a hit in the corresponding MVD barrel and disk layers. The shown data is from a `BoxGenerator` data set with $\theta = 60°$.
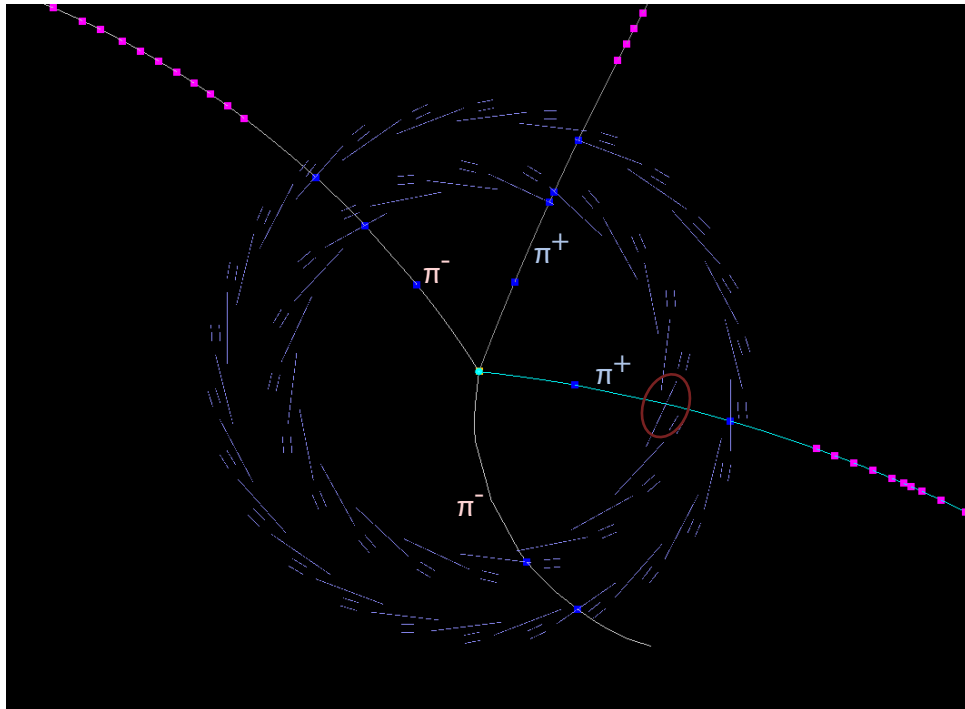
Figure 5.29: Example event in the event display. The point at which a π^+ pierces through the active detector material but produces no hit is outlined with a red ellipse. MVD hit points are represented by blue squares and the STT hit points by purple squares.
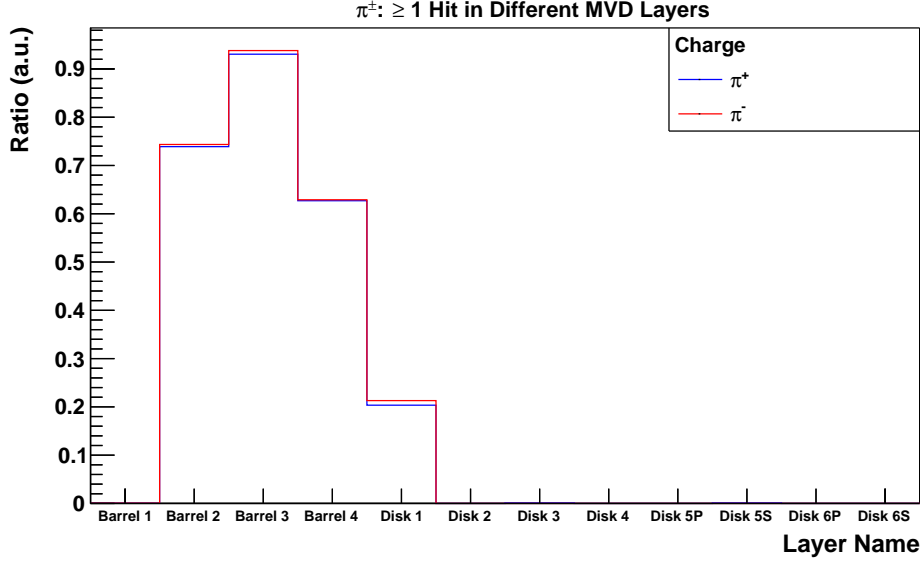
Figure 5.30: π^+ & π^-: Fraction of tracks with a hit in the corresponding MVD barrel or disk layer for BoxGenerator data at $\theta = 60°$, simulated with GEANT4.

shown. Careful analysis shows that the track should produce a hit in the detector material. No uninstrumented slits or inactive material is hit.

5.3.2.4   GEANT4

Further study shows that the reason for not creating a *hit* in the sensor material lies in the chosen MC propagator. During the course of this analysis, GEANT3 is used as the MC generator. Switching it to GEANT4 solves the problem of the pion charge asymmetry. Figure 5.30 shows the same $\theta = 60°$ data point as Figure 5.28, but this time data is simulated using GEANT4.

Different reasons for the wrong performance of GEANT3 can be imagined. One example is the step size for re-evaluation of interaction with active detector material, which can be set to be too large. This could either be a problem for positively charged particles in particular, or for particles incident nearly perpendicular to the traversed material. Because of the fanning setup of layer 3 of the MVD, there are statistically more π^+ which are perpendicular to the sensor. Why the same effect does not occur for π^- in the layer 4 is not really clear. In any case, the problem surely is a simulation software bug, which is in need of further analysis but beyond the scope of this thesis.

Since the cause of this effect is only discovered during the final days of writing this document, the rest of the analysis is performed with GEANT3 and incorporates the inherent pion charge asymmetry.

5.3.3   Decay Positions

Another quantity is correlated to the number of hits a particle leaves when traversing the tracking sub-detectors of $\overline{\text{P}}$ANDA: the point of the particle's decay. Only if pions or kaons have
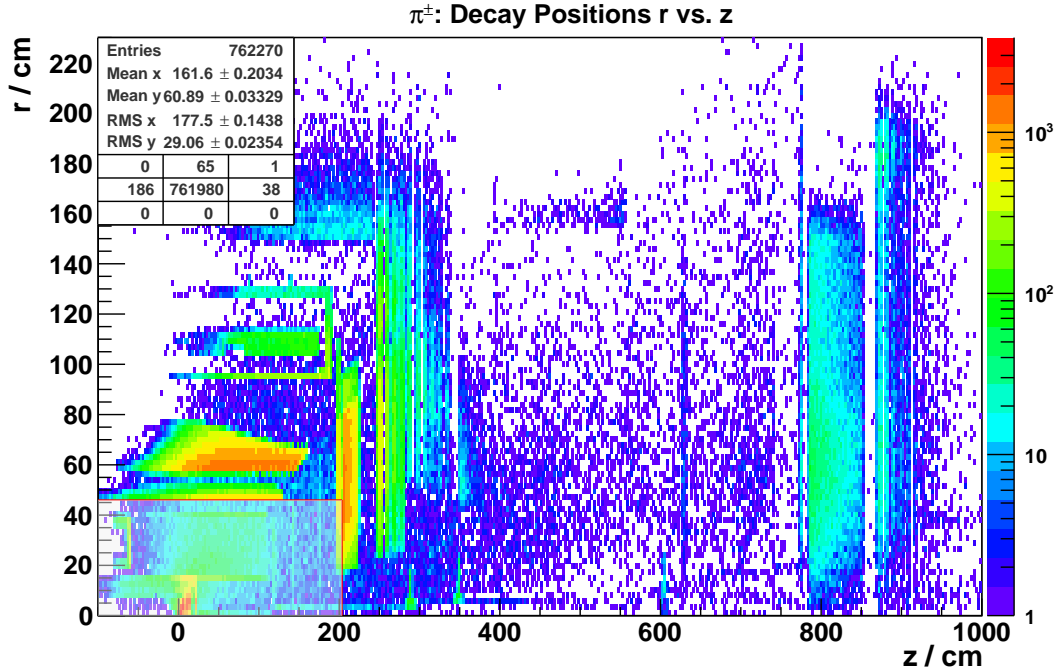
Figure 5.31: π: Decay positions of pions. The red outlined window is chosen to lie inside of the DIRC (radially) and inside of the EMC (longitudinally).

not yet decayed are they measurable directly. Both particles decay mainly into a $\mu\, \nu_\mu$ pair, creating a kink in the track due to their large lifetimes.

The positions of decays of π and K mesons are shown in Figure 5.31 and Figure 5.32, respectively, in the rz-projection. The different structures of the $\overline{\mathrm{P}}$ANDA detector are visible. Especially the material-intensive EMC shows up clearly, as many of the pions and kaons are stopped and then decay in this detector part.

The ratio of the number of particles decaying in the space enclosed by the EMC to the number of all particles gives a measure of the detection inefficiency due to the particle decay. For analyzing this ratio, a sector inside the EMC (in the forward direction) and the DIRC (radially) is defined with $z < 203.5\,\mathrm{cm}$ and $r < 46\,\mathrm{cm}$; the box is outlined in Figure 5.31 and Figure 5.32. Only if particles leave the marked box, the full number of hit points can be measured, resulting in good track resolution.

The ratio of the number of decays inside the box to all tracks is shown in Figure 5.33(a) (pions) and Figure 5.33(b) (kaons). Through a large angular range, about 90 % of the pions decay outside the tracking volume. For very large angles, pions do not leave the DIRC-enclosed area any more. A similar picture is seen for kaons. Their shorter lifetime leads to more decays in the tracking volume. As already seen in the previous parts of this section, kaon reconstruction is limiting factor in the analysis.
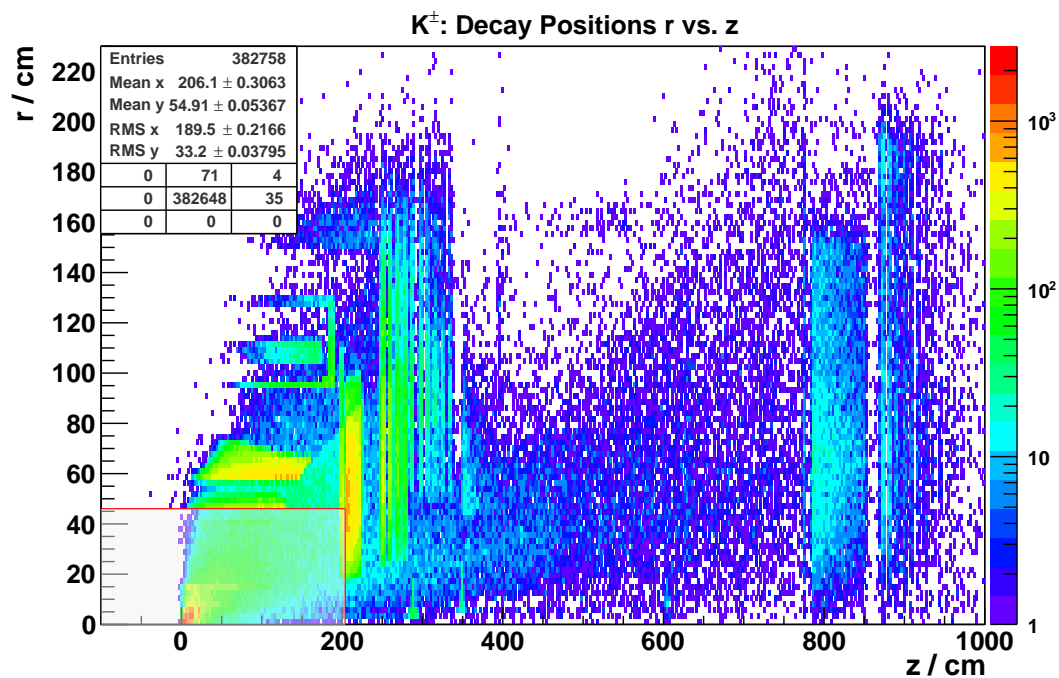
Figure 5.32: K: Decay positions of kaons. The red outlined window is chosen to lie inside of the DIRC (radially) and inside of the EMC (longitudinally).

(a) π.

(b) K.
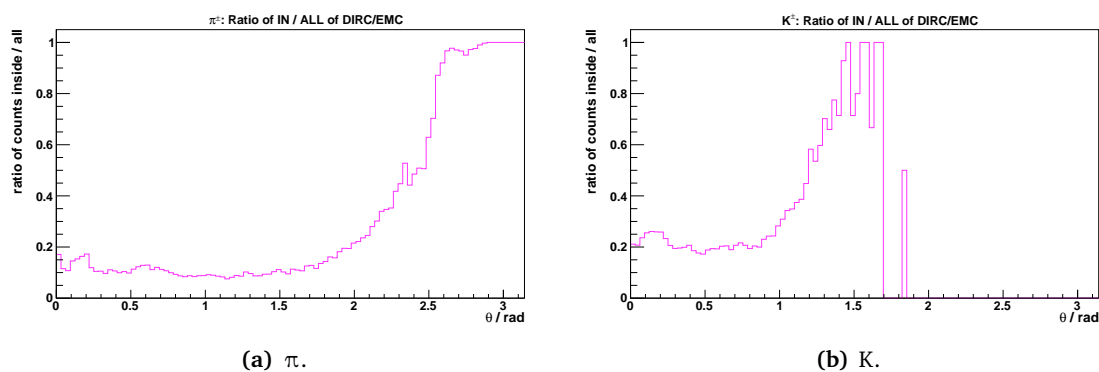
Figure 5.33: Probability for particles to decay inside the box defined in Figure 5.31 as a function of the polar angle.
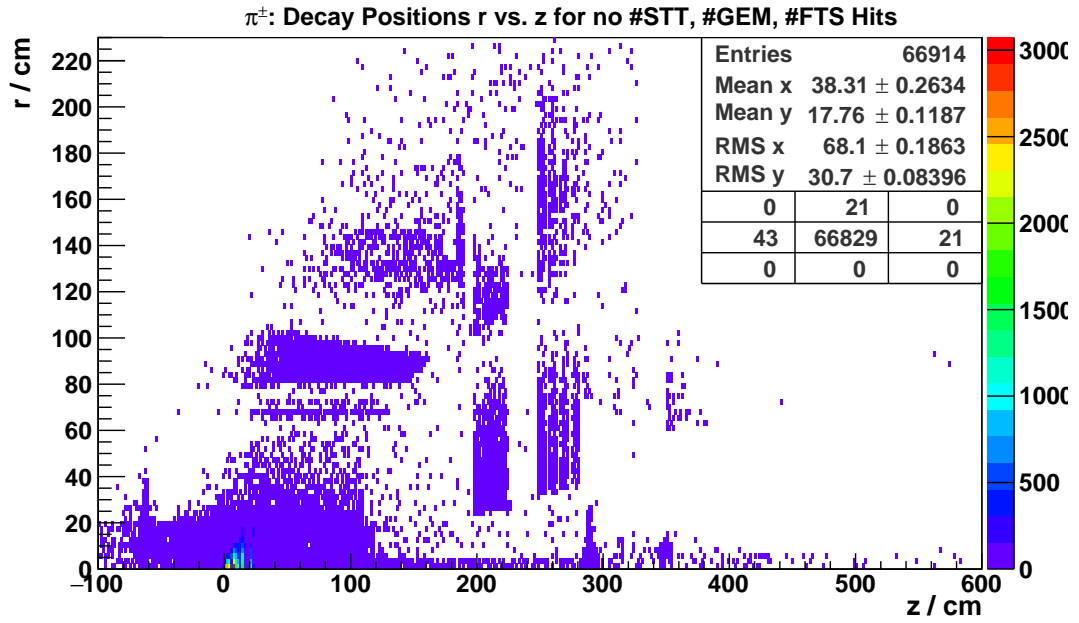
Figure 5.34: π: Decay positions of pions which produce no hits in the STT, GEM, and FTS.
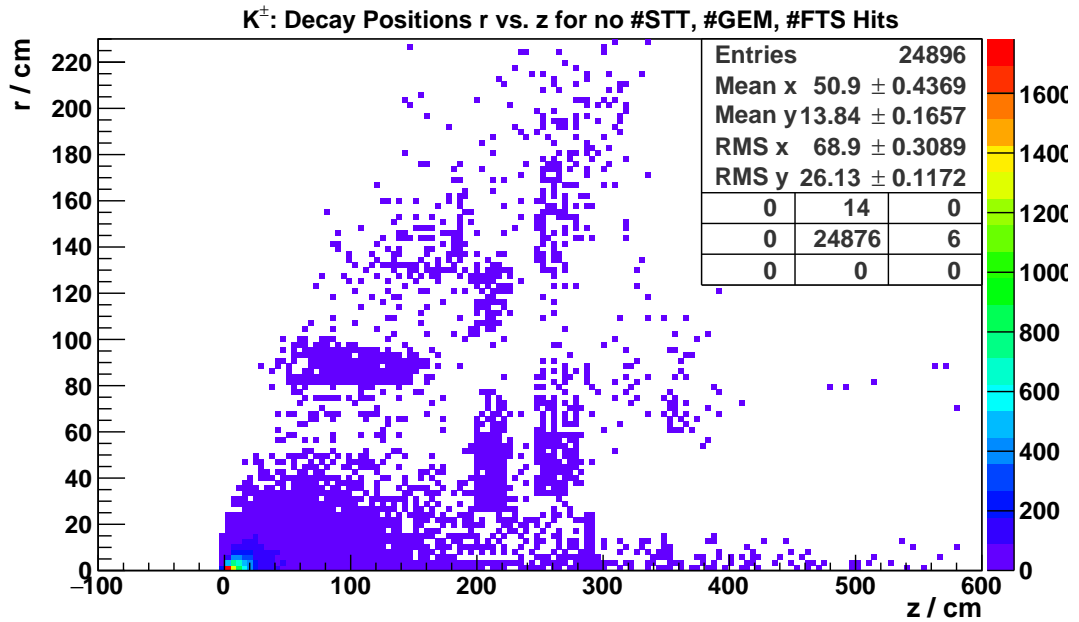
Figure 5.35: K: Decay positions of kaons which produce no hits in the STT, GEM, and FTS.

(a) Longitudinal view with markers.

(b) xy-view for $z < 180\,\text{cm}$.

(c) xy-view for $180\,\text{cm} < z < 230\,\text{cm}$.

(d) xy-view for $z > 230\,\text{cm}$.
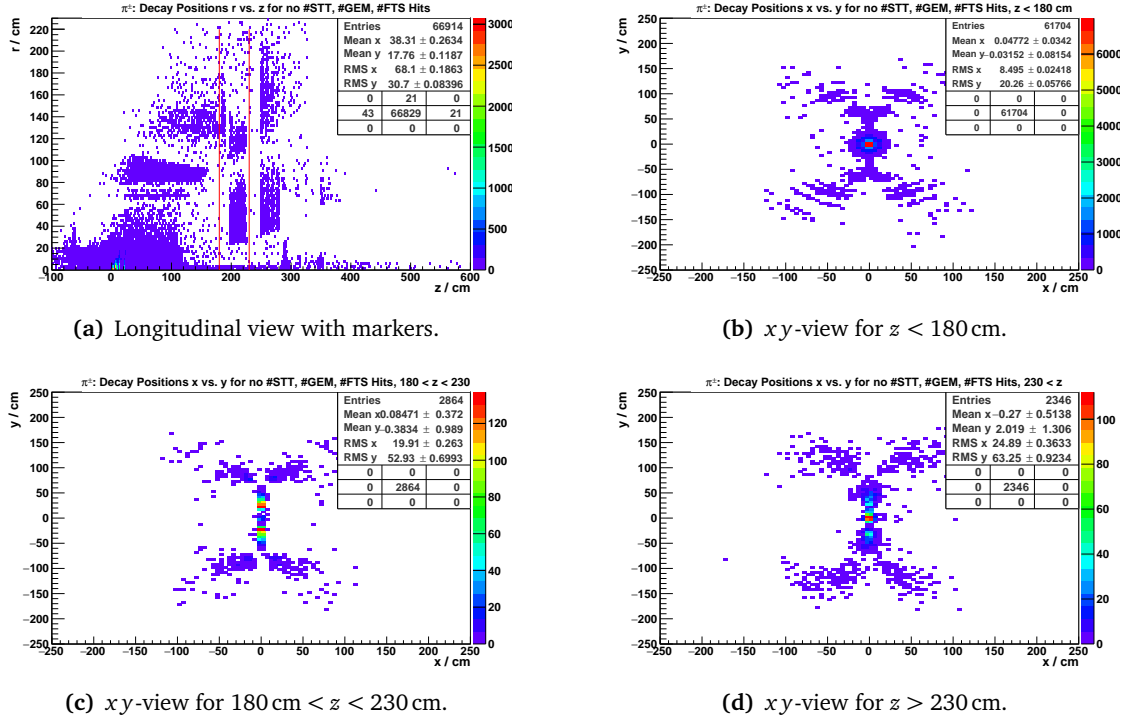
Figure 5.36: xy-position distribution of pion tracks which do not have hits in the STT, GEM, or FTS for several z-ranges.

Tracks With Only MVD Hits It is already been seen in Figure 5.18(b) that a significant fraction of particles produce no STT, no GEM, or FTS hits. The decay positions of these particles are shown in Figure 5.34 for pions and in Figure 5.35 for kaons. For both cases, most of the particles decay within the MVD. For pions, the last MVD forward disk at $z = 23\,\text{cm}$ can be spotted. Apart from this expected behavior, some particles still reach the outer sub-detectors of $\overline{\text{P}}$ANDA (DIRC, EMC, muon system). This happens without any hits in the STT or GEM.

The effect is analyzed further in the graphs of Figure 5.36. The figure shows decay position distributions in the xy-plane for different slices in z, as an example for pions. The analyzed z limits are outlined in the overview in Figure 5.36(a): one range is to $z = 180\,\text{cm}$ and another is at $z = 230\,\text{cm}$. The resulting distributions show how the pions can reach the outer sub-detectors without traversing either the STT or the GEM: they go through the vertical slits of the detector for the target, as their comparably large p_t leads to trajectories which are curved only little and do not pass any detector.

The detection efficiency of these kind of tracks is very limited, as only the low number of MVD hit points can be used to formulate track candidates. Furthermore, a large fraction of the particles decay inside the MVD and can usually not be reconstructed at all.
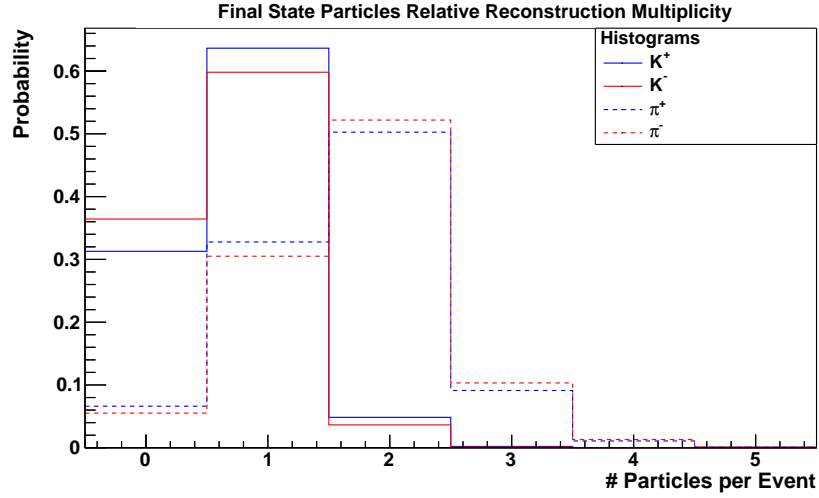
Figure 5.37: Multiplicity distributions of reconstructed π^+, π^-, K^+, K^-. Pions are shown with dashed lines, kaons with solid lines; positively-charged particles in blue, negatively-charged in red.

5.4 Event Analysis

In this section, pions and kaons are combined into the composite D candidates. First, the particles of the final state are analyzed, before the combined D meson candidates are studied extensively, both as single particles and as a combined system. For reconstruction PandaRoot is employed. The analysis parameters are summarized at the beginning in Section 5.1. For event analysis, the Rho package is used (see Section 3.5.3.5).

5.4.1 Final State Particles

The first step of the analysis is to identify $\pi^\pm$ and $K^\pm$ particles of the final state. The multiplicity distributions with which the particles are detected is shown in Figure 5.37. In 69 % of events at least one K^+ and in 64 % at least one K^- can be reconstructed. The number of reconstructed pions has its highest point at two pions per event, as expected for the simulated decay. The fraction of events, which have only one single pion reconstructed, is quite high: One single π^- is found for 30 % of events and 32 % of events contain one π^+. These events with only one pion limit the reconstruction efficiency of D mesons strongly – see also Section 5.5.

The energy and momentum of the particles are well-matched to energy and momentum of the originally simulated particles. The resolutions of energy and mass are given in Table 5.1. An interesting pattern can be seen in the distribution of relative energy and momentum resolution, ΔE and Δp, drawn as a function of the respective MC quantity (E^{MC}, p^{MC}) in Figure A.14 and Figure A.15. For both distributions the reconstructed value for low magnitudes is systematically smaller. The reason lies in the fact that low-energy / low-momentum particles are more strongly affected by multiple scattering, leading to energy / momentum loss.

Table 5.1: Overview of reconstruction efficiencies and quality of K and π mesons before assembly into composite D candidates. The number of mesons, N, is in relation to the full set events. For K, it refers to at least one K present, for π at least two pions are demanded. The resolutions of energy E, σ_{E}, and momentum p, σ_{p}, are taken from double Gauss fits to the distributions.

	N / %	σ_{E} / (MeV)	σ_{p} / (MeV/c)
K^+	69	11.2	13.2
K^-	64	11.5	13.4
π^+	61	7.9	8.3
π^-	64	7.8	8.2
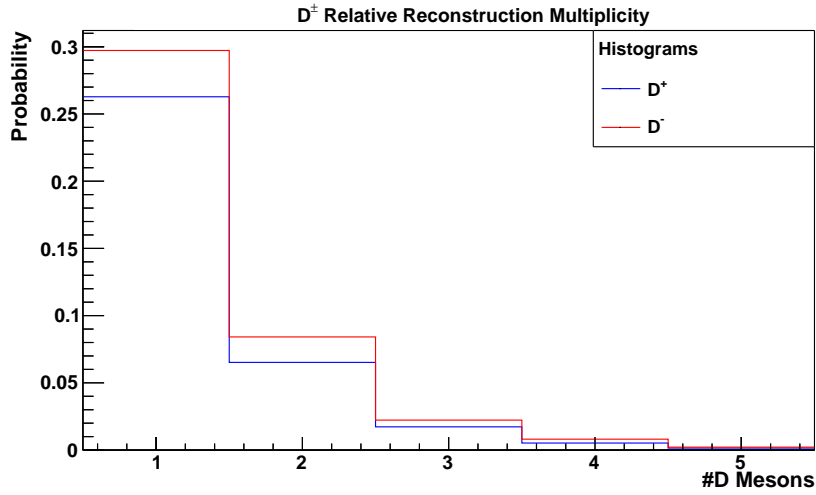
Figure 5.38: Multiplicity spectrum for D meson candidates.

5.4.2 Single D Mesons Reconstruction *(Inclusive Decay)*

If at least one K and two π are reconstructed, a D meson can be assembled. While the clearest signal emerges only when exclusively reconstructing both branches of the D^+D^- decay, insight into many physics topics of interested can already be gained when considering only one single D meson (see Section 3.2.1.1). Reconstruction of single D mesons lays the basis for all subsequent analyses and is, hence, of high importance.

5.4.2.1 Before Cuts

D mesons are combined from the signal data set by applying only a rough mass window around the PDG D mass, $m_{\mathrm{D}} \pm 0.15\,\mathrm{GeV}/c^2$. In some events, more than one $D^{\pm}$ meson candidate is reconstructed. The multiplicity distribution is shown in Figure 5.38. In 25 % of the cases where D^+ mesons can be reconstructed there is more than one candidate. For D^- this is 29 %. Overall, in 35 % of all events at least one D^+ can be reconstructed, and one D^- in 42 %. The difference in higher reconstruction efficiency for D^- lies mainly in π^- leaving more hit points in the tracking sub-detectors. The effect is outlined in Section 5.3.2.

(a) K⁻.

(b) π⁺ (two entries per D meson).
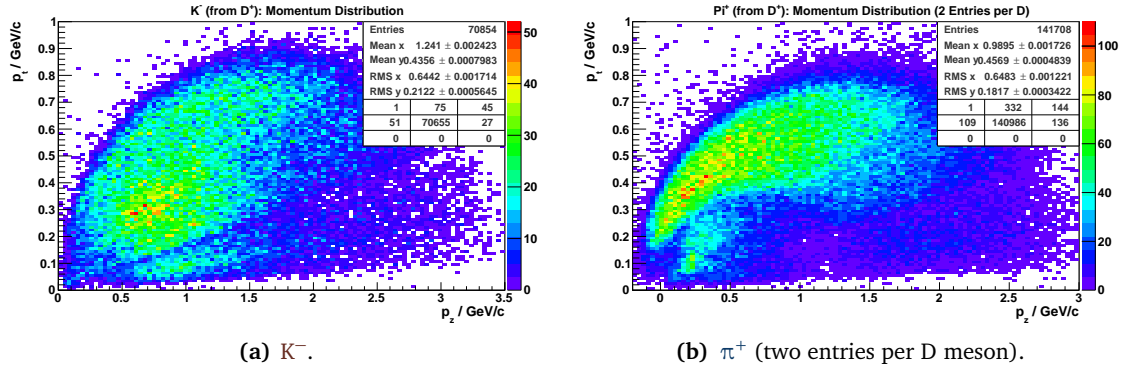
Figure 5.39: p_t vs. p_z distributions of K⁻ **(a)** and π⁺ **(b)**, which are assembled to D⁺ candidates.

(a) D⁺.

(b) D⁻.
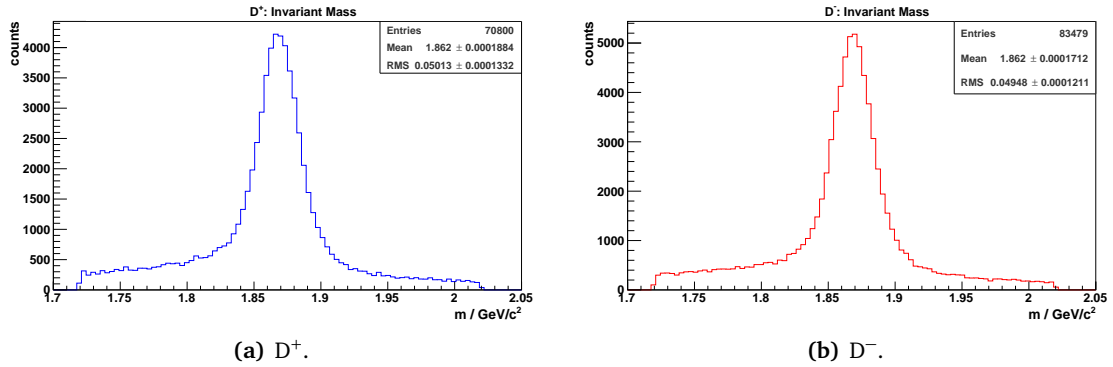
Figure 5.40: Invariant mass distributions for D± candidates.

Daughter Momentum In Figure 5.39, the momentum distributions of the daughter particles of assembled D⁺ meson candidates are compared[5]. The features seen in the data from the event generator (Figure 5.9(b) and Figure 5.10(b)) can be seen, with added reconstruction uncertainties.

Invariant Mass The invariant mass distribution after assembly of the pions and kaons is shown in Figure 5.40 for both charge states of D mesons. The coarse mass selection is seen, as the histograms only have entries in the range of $1.72\,\text{GeV}/c^2$ to $2.02\,\text{GeV}/c^2$. The distribution peaks around the D meson mass. A pedestal is visible under the peak, originating from incorrect combinations of daughter particles to composite D meson candidates. The combinatorial background rises for lower mass values.

5.4.2.2 Vertex Fit

Since both pions and the kaon of a D meson candidate come from the same decay vertex, their tracks can be fit to a common vertex. The fit routine employed is the PndKinVtxFit

---

[5]D⁺ meson candidates are taken in this section as an example. The D⁻ distributions are similar.

(a) χ^2 distribution.

(b) Prob(χ^2) distribution.
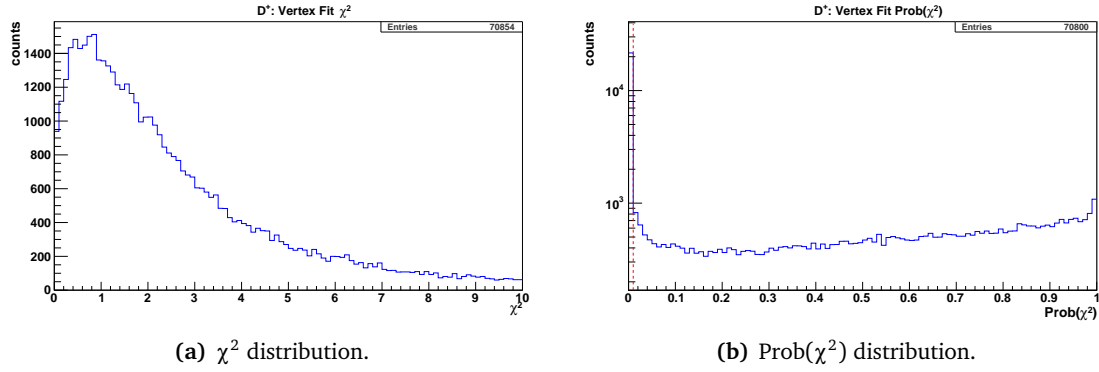
Figure 5.41: Quality plots of the vertex fitter for D^+ meson candidates. In the logarithmic probability distribution, right, the cut$_{\text{vtx}}$ of Prob(χ^2) > 0.01 is marked.

(a) Comparison of Prob(χ^2). Logarithmic plot.

(b) Comparison of Δp_z (on x) versus Δp_t (on y) for $\chi^2_{(1)} - \chi^2_{(2)} < 0.05$.
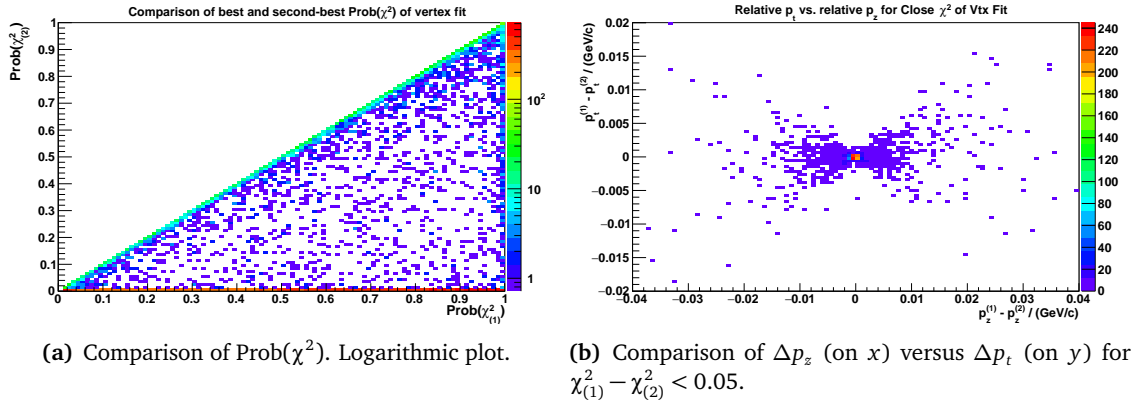
Figure 5.42: Comparisons of χ^2 after the vertex fit, passing cut$_{\text{vtx}}$. The D^+ candidate with the least χ^2 is compared to the candidate with the second-least χ^2. In **(a)**, the probability distribution of the two χ^2 values are shown. If no second candidate is available in the event, the first bin along the y-axis is used to note the $\chi^2_{(1)}$ value. **(b)** shows the difference in p_t versus the difference of p_z for candidates with similar χ^2.

class of PandaRoot, which modifies the track parameters of the daughter particles within their measurement uncertainties to originate from one common vertex.

Figure 5.41 shows the quality of the vertex fits for the D^+ candidates. The χ^2 distributions (Figure 5.41(a)) peaks around 1, but has a tail towards 10. This effect can also be seen in the probability distribution of the χ^2 in Figure 5.41(b)[6]. The distribution has a large peak around 0. To achieve a distribution that is roughly flat over all Prob(χ^2) values, the fits of D meson candidates with Prob(χ^2) < 0.01 are removed, indicated by the red line in Figure 5.41(b). In the following the cut is referred to as cut$_{\text{vtx}}$. In Section 5.4.4.1 the cut is optimized for ideal background suppression.

---

[6]Prob(χ^2) is calculated using ROOT's `TMath::Prob()` method. It refers to the probability that an observed χ^2 exceeds the χ^2 p.d.f. by chance.

(a) First track. **(b)** Second track.
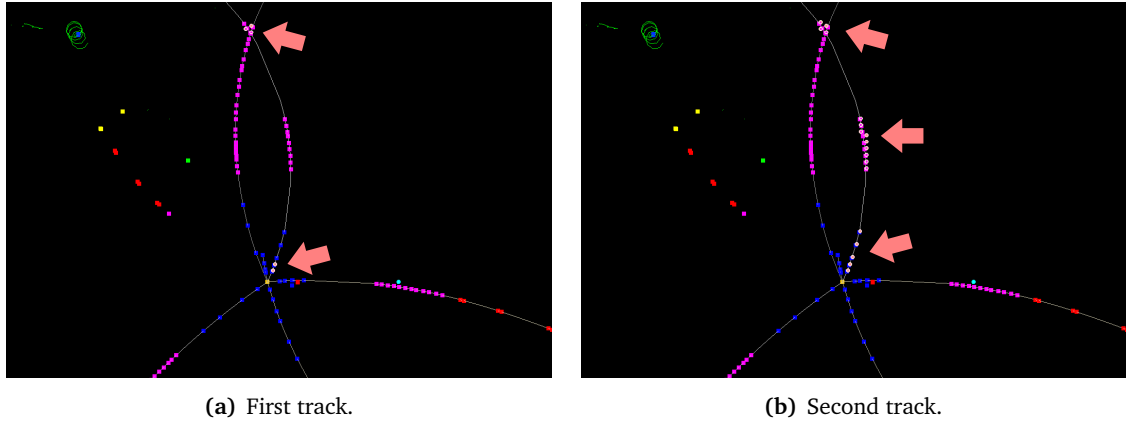
Figure 5.43: Event display of a track that was found twice. The hit points associated to the track are highlighted in white. The red arrows mark the clusters of hits which comprise the tracks.

Multiple D Meson Candidates   After the vertex fit and the cut$_\text{vtx}$, often more than one D$^+$ meson candidate per event is present. Sorting the D$^+$ candidates by increasing χ^2, the D$^+$ candidate with the smallest χ^2 is compared to that of the second smallest in Figure 5.42. Figure 5.42(a) shows the distribution of the χ^2 probability of the first candidate against that of the second. If no second candidate is present, the values are filled into the first bin of the y-axis. It can be seen that most of the Prob(χ^2) values are close to the $x = y$ diagonal. The fact is underlined when considering the momenta of the two candidates on the diagonal. Figure 5.42(b) shows the longitudinal momentum difference between the best vertex fit candidate and the second best vertex fit candidate versus the transverse momentum difference. The obtained distribution is symmetric around 0, no candidate is systematically different. In addition, the momentum difference in both directions is very small, most entries are in the bin at $(0,0)$. This suggests as well that there is no definitive *better* candidate among the two.

The reason for two or even more D$^+$ meson candidates per event is that some tracks (usually pions) are reconstructed more than once. The whole track is reconstructed as different tracklets, each one a valid pion track hypothesis. Currently, a track merger capable of combining the two tracks of one particle is not available, leading to multiply reconstructed D$^+$ meson candidates. An example event is shown in Figure 5.43. The red arrows mark the starting points of two different reconstructed tracks from the same pion. None of the pion tracks is discriminated at this stage, thus the combinatorics create multiple D$^+$ meson candidates.

Since both reconstructed D$^+$ meson candidates are representations of the same particle, the particle with the least χ^2 is taken in the following as the best candidate. For the D$^+$, 22 % of the generated mesons are still present, for D$^-$ 25 %.

Vertex Distribution   After fitting the K$\pi\pi$ to a common vertex, the D meson candidate has a non-zero decay vertex position. Figure 5.44 shows the distributions for the three coordinates. The turquoise distribution additionally demands the cut$_\text{vtx}$ to be passed and shows the D candidate with the smallest χ^2. The x and y-distributions have their central values at 0, the z-distribution has the most probable value at 210 μm. The resolutions in the x and y-directions

(a) x-vertex distribution for D meson candidates.

(b) z-vertex distribution for D meson candidates.
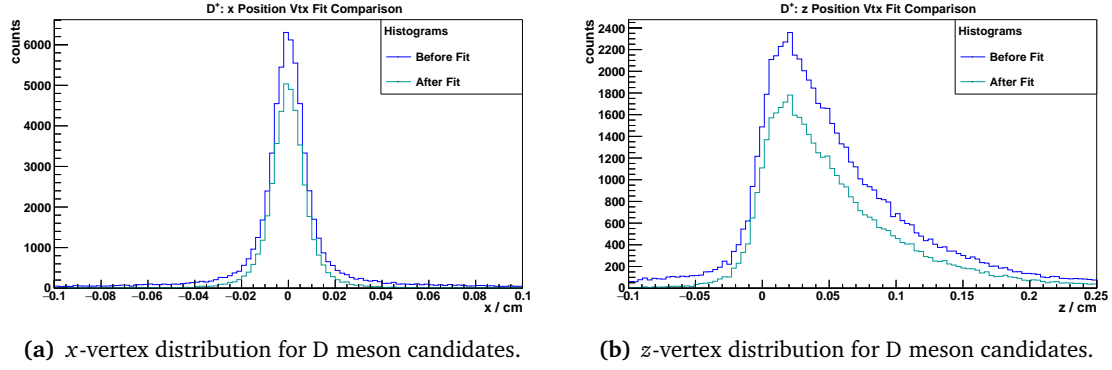
Figure 5.44: Vertex position distribution of D^{+} meson candidates after the vertex fit. The distribution of the best D candidates passing the cut_{vtx} selection is superimposed on all vertex-fitted D meson candidates. The resolutions are displayed in Figure A.16 with fits. The distribution for the y coordinate is not shown as it is similar to the distribution of the x coordinate.
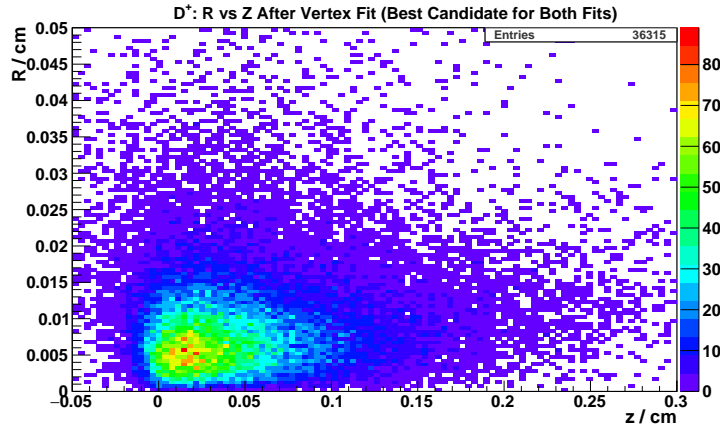
Figure 5.45: Decay vertex distribution of D^{+} meson candidate in the $R-z$ space, for candidates, passing cut_{vtx} and cut_{mass}. $R = \sqrt{x^2 + y^2}$.

are about $65\,\mu m$ and $63\,\mu m$ after the fit, respectively. Taking only the best candidate per event with the smallest χ^2, the resolution is improved to $58\,\mu m$ and $59\,\mu m$, respectively. For z, the resolution is lower ($104\,\mu m$), as the distribution is smeared out by the motion of the D mesons and the finite life-time of $c\tau = 311\,\mu m$. The resolutions are shown in Figure A.16.

In the two-dimensional plot shown in Figure 5.45, the radial decay position is plotted versus the longitudinal position. Most of the D mesons decay within $150\,\mu m$, radially, and $750\,\mu m$ longitudinally.

Momentum Resolution   Figure 5.46 shows a comparison of the transverse momentum distribution of D^{+} candidates before and after the vertex fit. The vertex fit does not change the mean momentum of the particles, but the number of candidates passing the fit and the cut_{vtx} are reduced, resulting in a different shape of the distribution. Taking only the best candidate
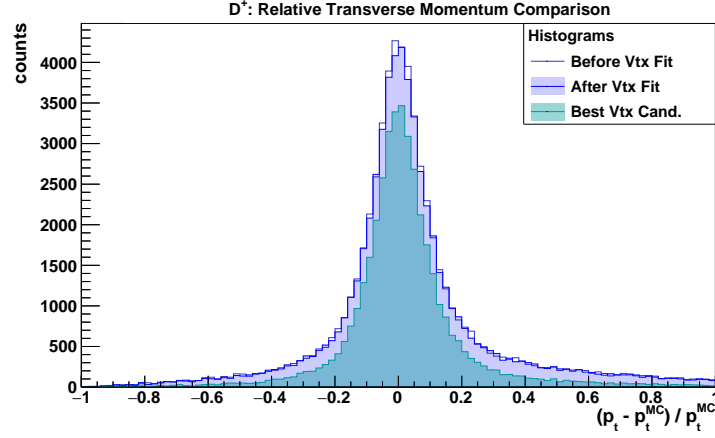
Figure 5.46: Relative transverse momentum resolution for three cases of D^+: Before the vertex fit, after the vertex fit, and after the vertex fit, taking only the best candidate, which passes cut_{vtx}. The individual graphs with Gauss fits are shown in Figure A.17.

(a) χ^2 distribution.

(b) $\text{Prob}(\chi^2)$ distribution.
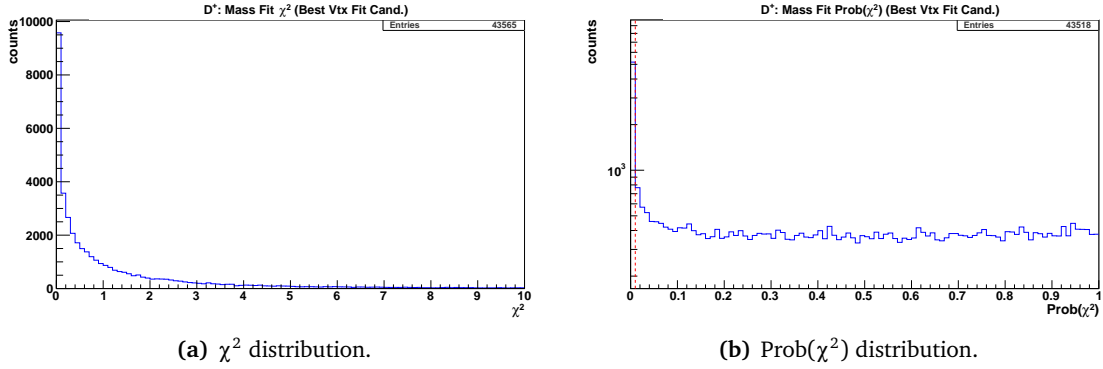
Figure 5.47: Quality plots of the mass constraint fitter for D^+ meson candidates. In the right plot, the preliminary cut_{mass} is marked.

after the vertex fit results in a relative p_t resolution of $\sigma_{p_t}/p_t = 8.1\,\%$. The resolution is $\sigma_{p_t}/p_t = 8.3\,\%$ for all particles passing only the vertex fit.

5.4.2.3 Mass Constraint Fit

To additionally constrain the events to those coming from D mesons, a mass constraint fit is applied. The `PndKinFitter` class of PandaRoot is used to fit the four-vectors of the D candidate to the nominal mass $m_{D^\pm} = 1.869\,\text{GeV}/c^2$. Input to this fitter is the best candidate of the vertex fitter, passing cut_{vtx}. The mass fit constrains the invariant mass of the candidate to the nominal $D^\pm$ mass, as can be seen in the difference plot of Figure A.18.

The quality of the mass constraint fit is shown in Figure 5.47. Like in the case for the vertex fit, a flat $\text{Prob}(\chi^2)$ distribution is desirable. Hence, a cut is applied, as indicated in red in Figure 5.47(b), to accept only D candidates with $\text{cut}_{\text{mass}} = \text{Prob}(\chi^2) > 0.01$, removing the fits with
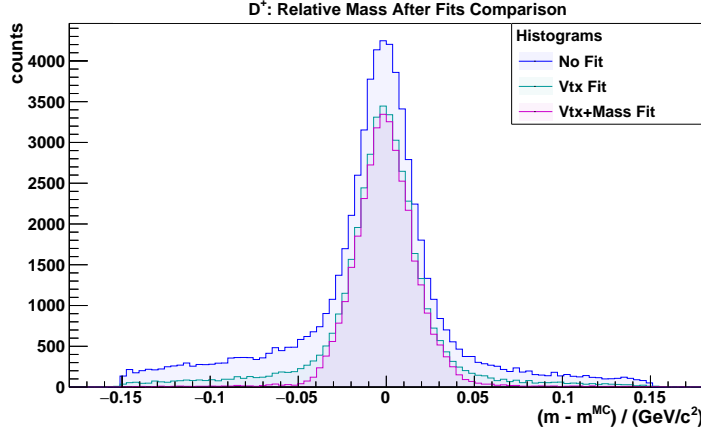
Figure 5.48: Distribution of the mass difference between the D candidate and the PDG value for the two presented fits. The individual plots, including Gaussian fits, are shown in Figure A.19.

very large χ^2. In Section 5.4.4.1 also this cut is optimized to reduce the largest amount of background.

18 % of the generated D^+ and 21 % of D^- candidates remain after combining cut$_{\text{mass}}$ with the best candidate from cut$_{\text{vtx}}$.

5.4.2.4 Invariant Mass After Fits

The progress of the quality of the reconstructed D mesons can be seen in the plot of Figure 5.48. The D^+ mass distribution minus the PDG mass is shown for three stages of the reconstruction chain. The blue distribution displays the mass distribution before any fit, as seen already in Figure 5.40(a). The turquoise distribution shows the mass distribution of the vertex-fitted D^+ candidate after passing cut$_{\text{vtx}}$. Finally, the purple distribution shows the mass distribution of the vertex-fitted candidate, for candidates that pass both cut$_{\text{vtx}}$ and cut$_{\text{mass}}$. The invariant mass resolutions for D^+ and D^- candidates after both cuts are $\sigma^+ = 15.7\,\text{MeV}/c^2$ and $\sigma^- = 15.3\,\text{MeV}/c^2$, respectively – see also Figure A.19.

5.4.2.5 Energy Resolution After Fits

In Figure 5.49, the energy resolution of the D meson candidates is shown as a difference of the reconstructed energy minus the generated energy. The energy resolution after the vertex fit, using cut$_{\text{vtx}}$ and cut$_{\text{mass}}$, is $\sigma^+ = 30.2\,\text{MeV}$ for D^+ and $\sigma^- = 29.0\,\text{MeV}$ for D^-. The individual graphs with superimposed fits are given in Figure A.20. A slight offset of the center of the distribution to $-3\,\text{MeV}$ can be seen over the whole range of E^{MC} (see also Figure A.21). The effect is already assessed at the reconstruction of the final state particles in Section 5.4.1 and propagates to the assembled D meson candidates.

The energy and invariant mass resolutions are summarized together with the percentages of events passing the different cuts in Table 5.2.

(a) ΔE distributions for D$^+$ candidates.

(b) ΔE distributions for D$^-$ candidates.
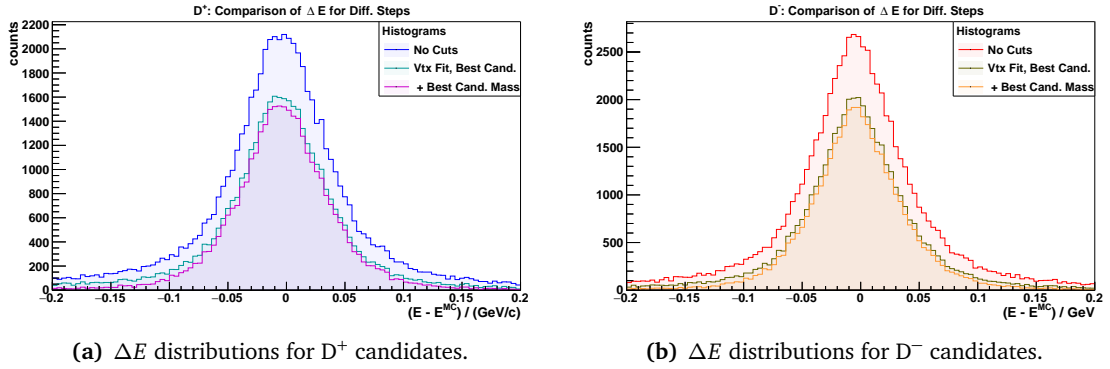
Figure 5.49: Energy resolution, $\Delta E = E^{\mathrm{RECO}} - E^{\mathrm{MC}}$, of the two D meson charges, for candidates of the different stages.

(a) m_{miss} distribution for D$^+$.

(b) m_{miss} distribution for D$^-$.
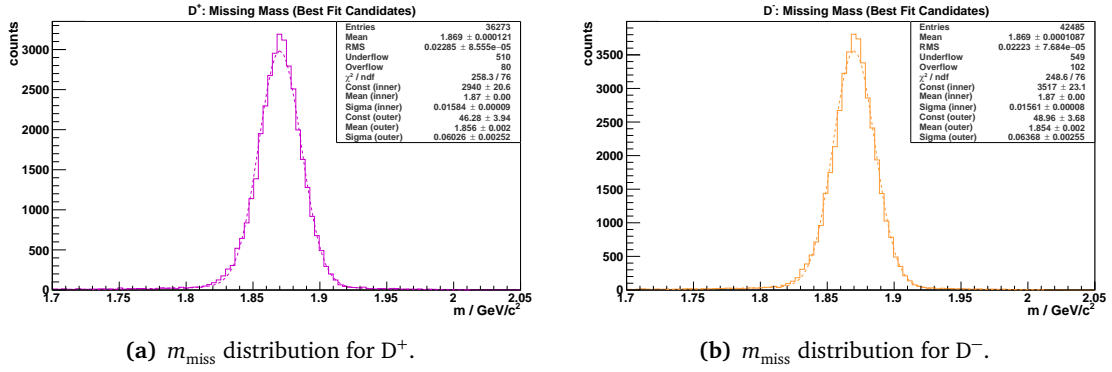
Figure 5.50: Missing mass distributions of the two D meson charges, for candidates passing both cut$_{\mathrm{vtx}}$ and cut$_{\mathrm{mass}}$. Double-Gaussian fits are superimposed.

5.4.2.6 Missing Mass Distribution After Fits

The missing mass distributions of both reconstructed D meson species are shown in Figure 5.50. The distributions peak clearly at $1.869\,\mathrm{GeV}/c^2$; the mass of the respective D meson partner. The distribution is largely similar to the invariant mass distribution shown in Section 5.4.2.4, with the one difference that the vertex and mass fit constraints affect the particle of which the mass is *not* shown. A cross connection is established. The achieved resolution of the missing mass is $\sigma^+ = 16.6\,\mathrm{MeV}/c^2$ (for D$^+$) and $\sigma^- = 16.3\,\mathrm{MeV}/c^2$ (for D$^-$).

5.4.3 D Meson Pairs *(Exclusive Decay)*

D$^+$ and D$^-$ meson candidates, which pass the fits introduced in the previous section (cut$_{\mathrm{vtx}}$ and cut$_{\mathrm{mass}}$) are combined to a common, exclusive D$^-$D$^+$ system. The distributions of momentum and energy before any further fits are shown in Figure A.22. From the data set with 200,000 generated D meson pairs, about 9000 can be combined to a D$^-$D$^+$ system; corresponding to an efficiency of about 4.5 %.

(a) χ^2 distributions.

(b) Prob(χ^2) distributions.
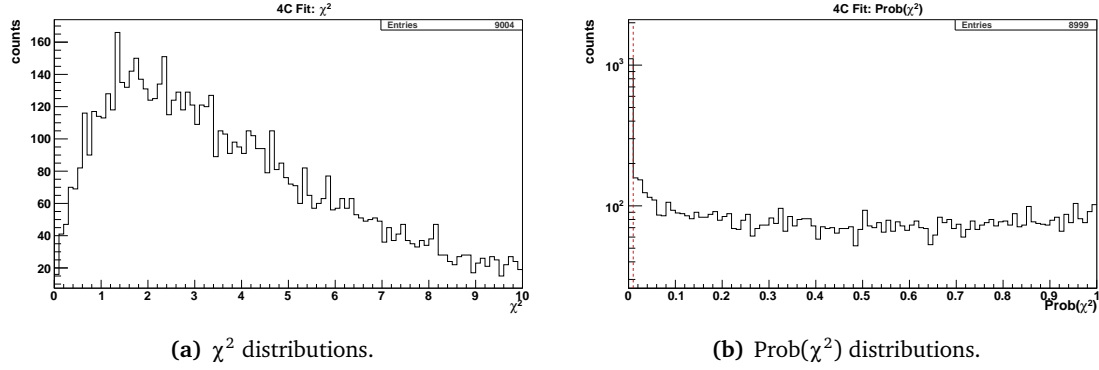
Figure 5.51: Quality plots of the four-constraint fitter for the D^-D^+ system. The preliminary cut of cut$_{4C}$ is marked in the right plot.

(a) Distribution of transverse momentum Δp_t.

(b) Distribution of longitudinal momentum Δp_z.
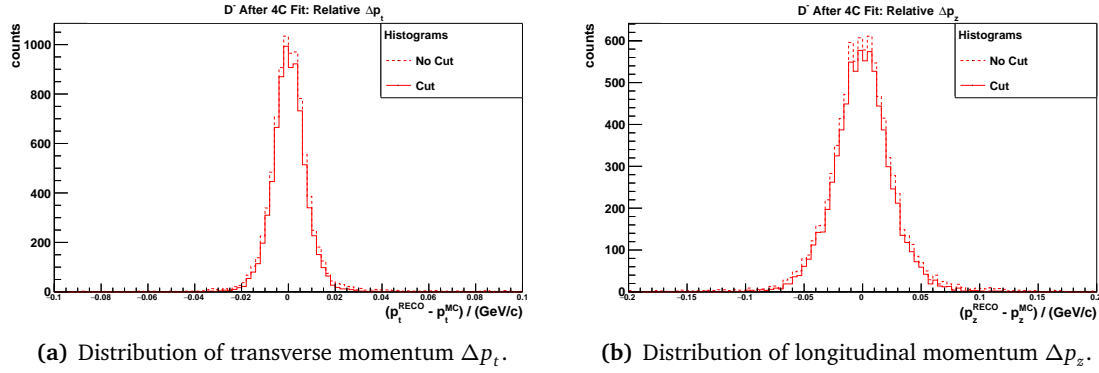
Figure 5.52: Comparison of relative momenta for D^- in the D^+D^- system after four-constraint fit. D mesons, before the cut$_{4C}$, are shown in dashed lines.

5.4.3.1   Four-Constraint Fit

Since both D mesons originate from one common vertex ($\overline{p}p$ annihilation), they can be constrained to this point. The Pnd4CFitter from PandaRoot is used to apply a four-constraint fit, modifying the four momenta in the D^-D^+ system to conserve energy and momentum of the $\overline{p}p$ beam (Figure A.23).

The quality of the fit is shown in Figure 5.51(b). As before, a cut is chosen on the probability of the χ^2 to discriminate against fits with bad quality. Marked in Figure 5.51(b), cut$_{4C}$ = Prob(χ^2) > 0.01 is chosen preliminarily[7]. Approximately 4 % of the initial D^+D^- pairs survive this cut.

5.4.3.2   Momenta of D Mesons After Fit

The four-constraint fit adjusts the four-momentum of the D^-D^+ system as a whole to match that of the beam. The individual $D^{\pm}$ mesons still have momenta which differ from the intitial

---

[7]For better background suppression, the cut$_{4C}$ is benchmarked and chosen more tightly in Section 5.4.4.3.

(a) Pull for p_x.

(b) Pull for p_z.
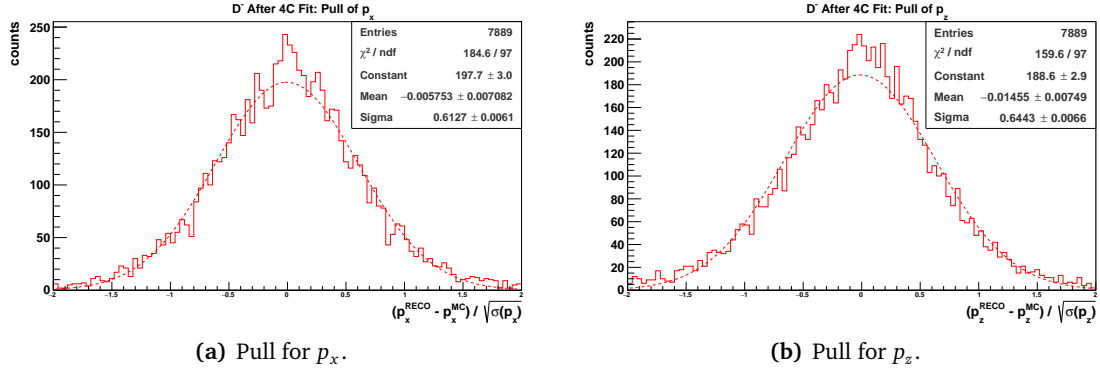
Figure 5.53: Two pull distributions of momenta of D^- after four-constraint fit of the D^+D^- system. The distribution for p_z is not shown since it is similar to that of p_x. Gaussian fits are superimposed.

beam momentum. Figure 5.52 gives an overview of p_t and p_z after the four-constraint fit, as an example for D^-. The resolutions for D^- are $\sigma_{p_t} = 5.6\,\text{MeV}/c$ and $\sigma_{p_z} = 17.4\,\text{MeV}/c$, the results are comparable for D^+. The pull distributions of the momenta in all three directions of the D^- meson are shown in Figure 5.53. The distributions are reasonably centered around 0, but have standard deviations of approximately $\sigma_{\text{pull}} = 0.6$, suggesting that the uncertainties on the momenta are overestimated.

5.4.3.3 Decay Vertex Distance

In Section 5.2.2.3, the distance $\overrightarrow{\Delta R}$ between the D^+ and the D^- decay vertices is measured on MC simulation data. Since the full event topology is known after assembly into the D^+D^- system and fitting it to the beam four-momentum, the plots of Figure 5.13 can be repeated with reconstructed data. The result is shown in Figure 5.54. The pattern already seen in Figure 5.13 can be spotted for the reconstructed data as well: The convolution of two exponential decays in the full three-dimensional representation of Figure 5.54(a); the mean value of 0 along z, and the much smaller width of the transverse distribution of $(R)_{xy}$. The resolution of the detector is noticeable in the distribution of the xy-plane. While the MC simulation data peaks around $20\,\mu\text{m}$, the reconstructed data has its maximum value at about $100\,\mu\text{m}$.

5.4.3.4 Invariant Mass Distributions

The mass difference distributions of the D meson candidates after the four-constraint fit are shown in Figure 5.55. The distributions are centered around 0, the resolutions are $\sigma_m^+ = 10.8\,\text{MeV}/c^2$ and $\sigma_m^- = 8.6\,\text{MeV}/c^2$ for D^+ and D^-, respectively. The mass resolution difference for the two charged states of the D mesons is a result of the higher reconstruction quality for D^- mesons. The higher number of π^- hit points lead to slightly better reconstructed π^- tracks, which result in a more sharply defined mass peak.

5.4.4 Background Studies

To study the performance of the reconstruction of the decay channel under realistic conditions, the non-resonant background at the beam momentum needs to be considered. The background
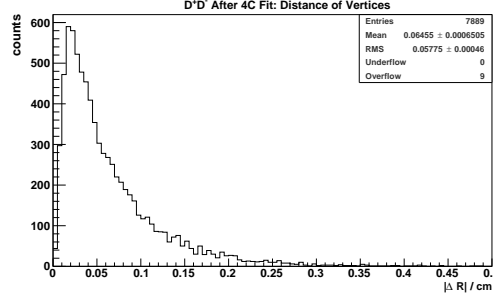
(a) $|\Delta R|$ distribution.
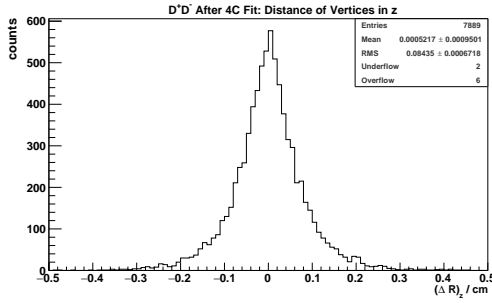
(b) Longitudinal distance distribution $(\Delta R)_z$.
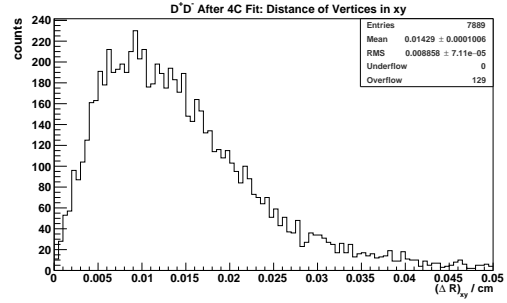
(c) Transverse distance distribution $(\Delta R)_{xy}$.

Figure 5.54: Distances between decay vertices of D^+ and D^-. $\overrightarrow{\Delta R} = \overrightarrow{vtx}(D^+) - \overrightarrow{vtx}(D^-)$, see Figure 5.12.
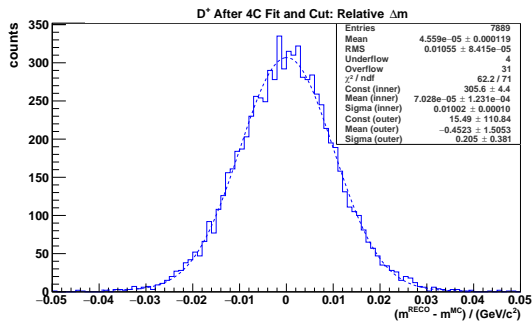
(a) D^+.
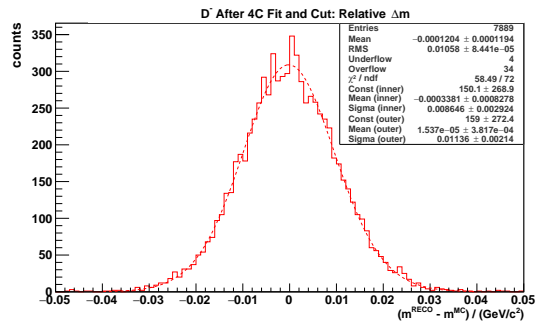
(b) D^-.

Figure 5.55: Invariant mass distributions minus the generated MC mass of the D mesons after the four-constraint fit and cut$_{4C}$. Double Gaussian fits are shown as well.

Table 5.2: Overview of percentages and resolutions in the different stages of the analysis: before and fits, after the vertex fit (including cut$_{vtx}$), after the mass constraint fit, for D meson candidates passing cut$_{mass}$. In the last line, results for exclusive D$^+$D$^-$ events is given after the four-constraint fit, which pass cut$_{4C}$.

	D$^+$			**D$^-$**		
	N	σ_E	σ_m	N	σ_E	σ_m
	%	MeV	MeV/c^2	%	MeV	MeV/c^2
			inclusive			
Before Fits	35.4	31.8	15.5	41.7	31.2	15.3
Vertex Fit	21.8	31.4	15.4	25.3	30.0	15.2
Mass Fit	18.1	30.6	15.7	21.2	29.1	15.3
			exclusive			
4C Fit	3.9	26.7	10.0	3.9	26.8	8.6

events can mimic the signal channel and lead to false-positively reconstructed events. The total hadronic background has a much larger production cross section, making effective background suppression challenging.

Using the `PndDpmDirect` generator included in PandaRoot, 100,000,000 DPM events have been generated. The production cross section for DPM events at $\vec{p}_{\text{Beam}} = 6.5\,\text{GeV}/c$ is $\sigma_{\text{DPM}} \approx$ 45,000 μb.
In the following section, cuts for inclusive events are studied first, then the exclusive system of D$^+$D$^-$ is analyzed.

5.4.4.1 Inclusive Decays

Of the 100 million simulated background-like events, 19,886 contain D$^+$ candidates, passing cut$_{vtx}$ and cut$_{mass}$ (D$^-$: 30,935). Since the final state of the decay channel consists solely of pions and kaons coming from one common vertex, the number of misidentified DPM-D$^\pm$ candidates is comparably high.

Two cuts are applied to further discriminate the background: A cut on the center-of-mass (CM) energy of the D candidate and a cut on the CM momentum. In addition, the preliminary cuts on the vertex fit and the mass constraint fit probability are refined.

To match the produced signal events to those of the background, the number of background events is scaled up by a factor:

$$\mathcal{F} = \frac{N_{\text{sig}}^{\text{gen}} / (\sigma_{\text{sig}} \times \mathcal{BR})}{N_{\text{bkg}}^{\text{gen}} / \sigma_{\text{bkg}}}, \tag{5.1}$$

with the number of generated signal events $N_{\text{sig}}^{\text{gen}} = 200,000$, the production cross section of the signal events $\sigma_{\text{sig}} \approx 100\,\text{nb}$ [148], the signal branching ratio $\mathcal{BR} = 9.13\,\%$ [1], the number

(a) $\mathcal{S}$ distribution for the vertex fit.

(b) $\mathcal{S}$ distribution for the mass constraint fit.
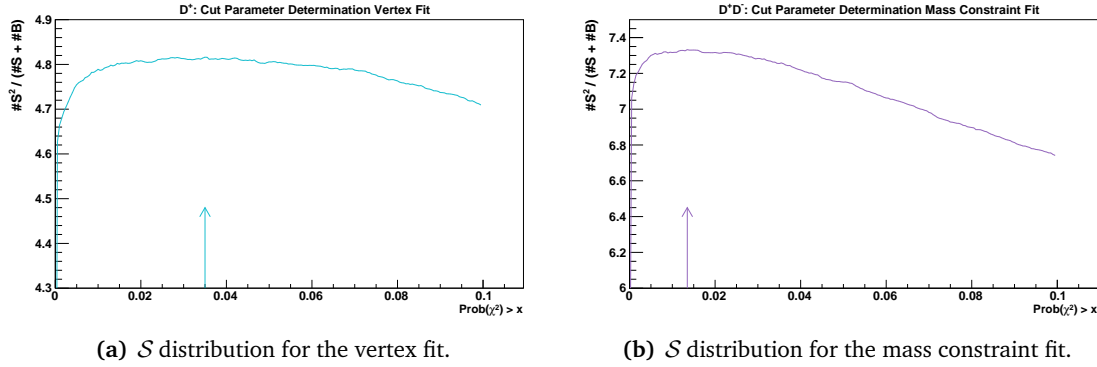
Figure 5.56: Optimization of the $\mathcal{S}$ parameter, $N^2_{\mathrm{sig}}/(N_{\mathrm{sig}} + N_{\mathrm{bkg}})$, for the χ^2 probability of the vertex fit (left) and the mass constraint fit (right). Both distributions show the x range to 0.1, to emphasize the dynamics of low probabilities on this axis. Chosen cut positions are marked by vertical arrows.

of generated background-like events $N^{\mathrm{gen}}_{\mathrm{bkg}} = 1 \times 10^8$, and the cross section $\sigma_{\mathrm{bkg}} \approx 45\,\mathrm{mb}$. In this section the background histograms are already scaled by $\mathcal{F}$.

Vertex and Mass Constraint Fit   In Section 5.4.2.2 and Section 5.4.2.3 the cuts on the vertex fit and mass constraint fit χ^2 probability, $\mathrm{cut}_{\mathrm{vtx}}$ and $\mathrm{cut}_{\mathrm{mass}}$, were preliminarily motivated by the need to achieve a flat $\mathrm{Prob}(\chi^2)$ distribution. The exact position of the cut is studied in the following in relation to the background-like events, in order to obtain an ideal cut parameter.

The optimization in this section, but also in the following cut parameter analyses, determines the signal-to-background ratio for a given cut. The highest point of the function obtained by the formula

$$\mathcal{S}\big(x_0, x_1; \mathrm{val}(x)\big) = \frac{N^2_{\mathrm{sig}}}{N_{\mathrm{sig}} + N_{\mathrm{bkg}}}(x_0, x_1; \mathrm{val}(x)) = \tag{5.2}$$

$$= \frac{\left(\int_{x_0}^{x_1} \mathrm{val}_{\mathrm{sig}}(x)\,\mathrm{d}x\right)^2}{\int_{x_0}^{x_1} \mathrm{val}_{\mathrm{sig}}(x)\,\mathrm{d}x + \int_{x_0}^{x_1} \mathrm{val}_{\mathrm{bkg}}(x)\,\mathrm{d}x} \tag{5.3}$$

is taken as the cut parameter; $\max(\mathcal{S})$. Equation 5.2 assesses the number of particles in a certain range ($[x_0, x_1]$) as obtained by evaluating the function $\mathrm{val}(x)$. With $\mathcal{S}$, events of the signal data set and of the background data set, which are still present when asserting a certain cut parameter, are set into a ratio. For optimizing the vertex and mass constraint fit, the probability of a χ^2, $\mathrm{Prob}(\chi^2)$, is chosen as the test function $\mathrm{val}(x)$. $\mathrm{Prob}(\chi^2)$ is defined in the range from 0 to 1, i.e. $x_0^{\min} = 0$ and $x_1 = 1$.

First, the vertex χ^2 probability is benchmarked without any other cut parameters. The $\mathcal{S}$ distribution is given in Figure 5.56(a). An ideal signal significance is achieved for $\mathrm{cut}^*_{\mathrm{vtx}} = \mathrm{Prob}(\chi^2) > 0.035$; the value is marked with a vertical arrow in the graph. Under the new assumption of $\mathrm{cut}^*_{\mathrm{vtx}}$, also the probability of the mass constraint fit is optimized, with the resulting distribution shown in Figure 5.56(b). The parameter chosen as the new cut value is $\mathrm{cut}^*_{\mathrm{mass}} = \mathrm{Prob}(\chi^2) > 0.0135$.

(a) Signal.
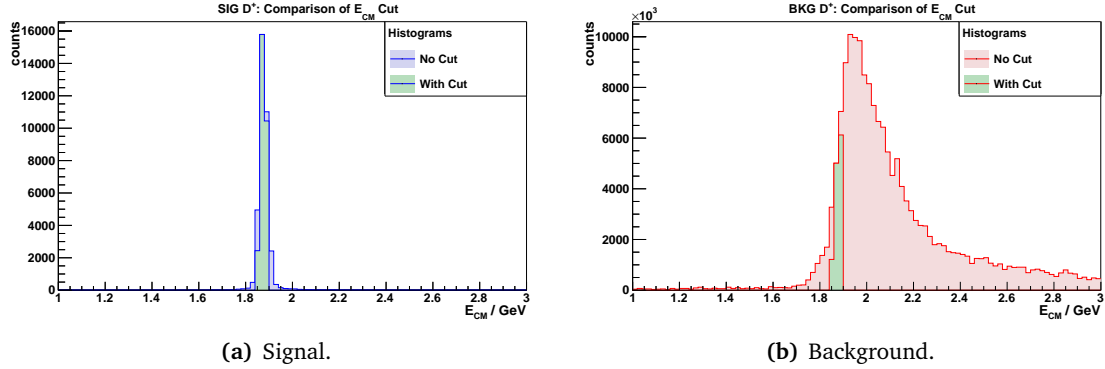
(b) Background.

Figure 5.57: Center-of-mass energy distributions of D$^+$ candidates from the signal data set (left) and the background-like DPM data set. Chosen cut ranges are superimposed.
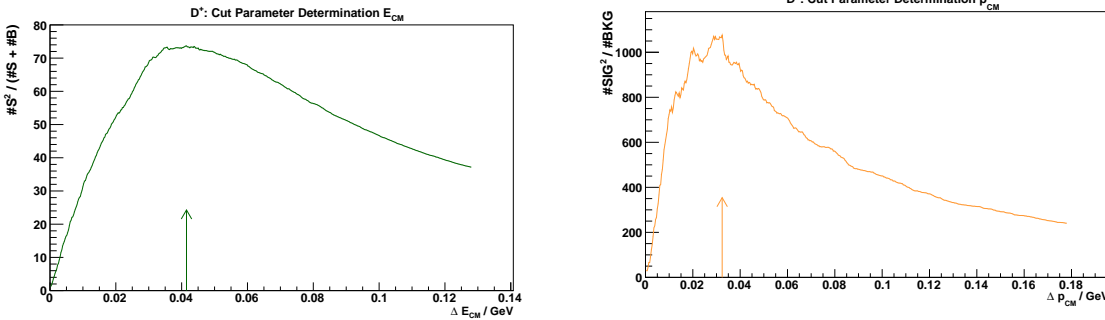
Figure 5.58: Determination of the cut position on the center-of-mass **energy**. The chosen value, cut$_E$, is marked on the x axis.
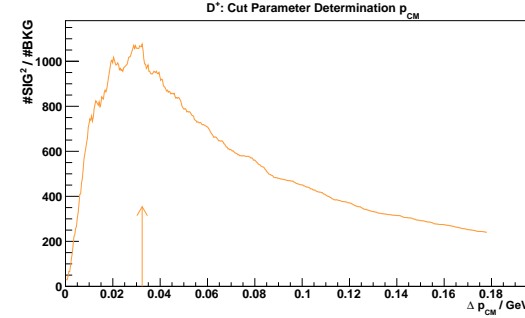
Figure 5.59: Determination of the cut position on the center-of-mass **momentum**. cut$_p$, which is taken as the best value, is marked with a vertical arrow.

Center-of-Mass Energy   In Figure 5.57 the CM energy distribution of the D$^+$ candidate is shown. The signal distribution (Figure 5.57(a)) is clearly centered around the D$^+$ meson mass, while the DPM distribution has its maximum at slightly larger energy. To determine an ideal selection range, different cut ranges are benchmarked using Equation 5.2, under the cut$^*_\text{vtx}$ and cut$^*_\text{mass}$ assumption. The test quantity is the center of mass, E_CM. The value is evaluated starting from the center of the signal distribution and then step-wise extending a symmetrical cut window of the size ΔE_CM, counting the number of signal and the background events in the window. The resulting $\mathcal{S}$-distribution is shown in Figure 5.58. The highest $\mathcal{S}$-ratio is obtained for cut$_E$ = 1.854 GeV < E_CM < 1.898 GeV. The cut is applied to select only events that fall into this region. The region is superimposed in green in Figure 5.57. The signal data set is reduced from 36,273 to 28,674 events (79 % efficiency), while the background set is reduced from $\mathcal{F} \times$ 19,886 to $\mathcal{F} \times$ 1252 events (6 % efficiency).

Center-of-Mass Momentum   In addition to the CM energy, the CM momentum can be used to reduce background events. Figure 5.60 shows graphs before and after cutting on the CM momentum, p_CM. cut$^*_\text{vtx}$ and cut$^*_\text{mass}$ are applied, cut$_E$ is omitted for this optimization. As before,

(a) Signal.
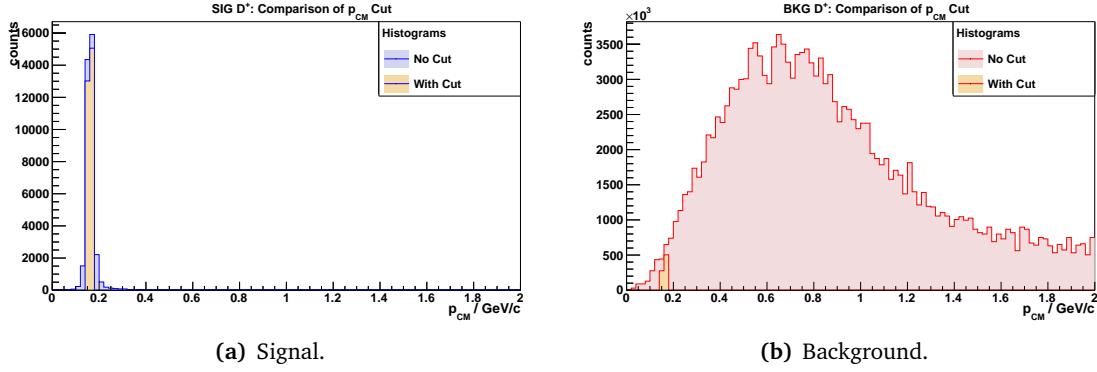
(b) Background.

Figure 5.60: Center-of-mass momentum of D^+ candidates from the signal data set (left) and the background-like DPM data set (right).

(a) Signal.
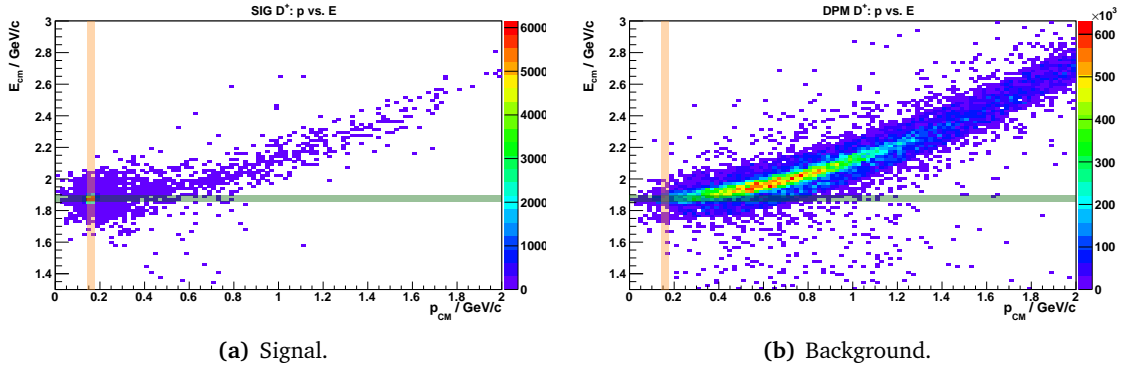
(b) Background.

Figure 5.61: Comparison of CM energy and CM momentum distribution for D^+ signal events and background events.

the cut parameter is determined using Equation 5.2 to count events in various ranges around the p_{CM} peak, see Figure 5.59. The ideal range is $\mathrm{cut}_p = 0.145\,\mathrm{GeV}/c < p_{CM} < 0.177\,\mathrm{GeV}/c$. This range is relatively narrow around the momentum peak, leaving 28,076 from the initial 36,273 signal events (77 %) and reducing the background events to $\mathcal{F} \times 79$ (0.4 %).

Results of Cuts   The two-dimensional plot of Figure 5.61 compares the two cut distributions with each other. Applying both cuts on E_{CM} and p_{cm}, 23,993 (66 %, or 12 % compared to all produced events) of the signal events and $\mathcal{F} \times 28$ (0.14 %) of the background events survive the cuts (compared to the full amount of produced background data, this is 2.8×10^{-5} %). Multiplying the scale factor $\mathcal{F} \approx 9861$ compares approximately 24,000 signal events to approximately 270,000 background events – a factor of approximately 10 times more background than signal remain.

The large combinatorial background which still exists after applying the cuts makes reconstruction of single $D^{\pm}$ mesons very challenging. The exclusive reconstruction (see below) gives promising reconstruction results, but has limited statistics (≈ 4 %).

(a) Signal.
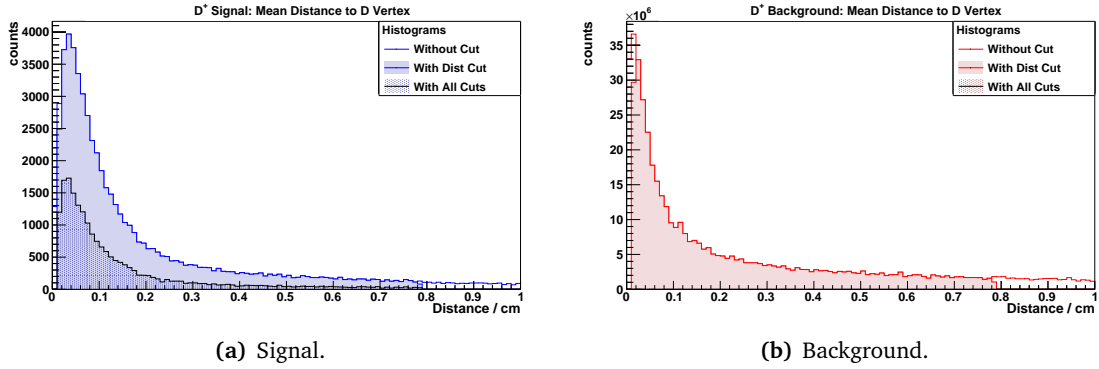
(b) Background.

Figure 5.62: Comparison of distributions of the mean minimal track distances to the D meson candidate vertices for signal events and background events.

5.4.4.2 Semi-Inclusive Analysis

To further suppress background, a first semi-inclusive analysis is done. In this analysis, one branch of the decay is reconstructed, e.g. the D^+, and the remaining particles in the event used for discrimination. The high statistics of the inclusive analysis is combined with the purity of the exclusive reconstruction (see next section).

The analysis starts with the reconstructed decay vertex of a D meson candidate. It considers the tracks from all remaining particles of the event and calculates the mean minimal distance to the decay vertex. The resulting distributions of the distances are shown in Figure 5.62.

As before, using Equation 5.2, the ideal cut positions are determined by maximizing $\mathcal{S}$. Since an upper and an lower cut parameter are needed, the optimization of $\mathcal{S}$ is done iteratively. First, the upper cut parameter is determined. Under this assumption, the lower cut parameter close to 0 is determined. Iteratively, both cut parameters are set until stable conditions are met and the respective $\mathcal{S}$ ratios do not change. The final cut parameters are: $\mathrm{cut}_{\mathrm{dist}}^{\mathrm{low}} = 0.0118\,\mathrm{cm}$ and $\mathrm{cut}_{\mathrm{dist}}^{\mathrm{up}} = 0.786\,\mathrm{cm}$. Only if the average distance of all remaining particles to the vertex is in between $\mathrm{cut}_{\mathrm{dist}}^{\mathrm{low}}$ and $\mathrm{cut}_{\mathrm{dist}}^{\mathrm{up}}$, the event is selected. The discriminated distributions are superimposed in Figure 5.62.

Using only $\mathrm{cut}_{\mathrm{dist}}^{\mathrm{low}}$ and $\mathrm{cut}_{\mathrm{dist}}^{\mathrm{up}}$, 27 % of the produced signal and 0.04 % of the produced background events are present. Combining the two $\mathrm{cut}_{\mathrm{dist}}$ with all previous introduced cuts ($\mathrm{cut}_{\mathrm{vtx}}$, $\mathrm{cut}_{\mathrm{mass}}$, $\mathrm{cut}_{\mathrm{E}}$, $\mathrm{cut}_{\mathrm{p}}$), the D meson candidates from the signal data set are reduced to 9.5 % and the background candidates to 1.4×10^{-5} %; from the DPM data set, $\mathcal{F} \times 14$ events remain. Multiplying the scaling factor compares about 20,000 signal events to 140,000 background events. Still a factor of 6.8 more background events than signal events are present.

This part is only a first semi-inclusive analysis. Further exploitation of event topologies is needed to continue reducing the amount of background events.

(a) Prob(χ^2) of fit.

(b) E_{CM}.

(c) p_{CM}.
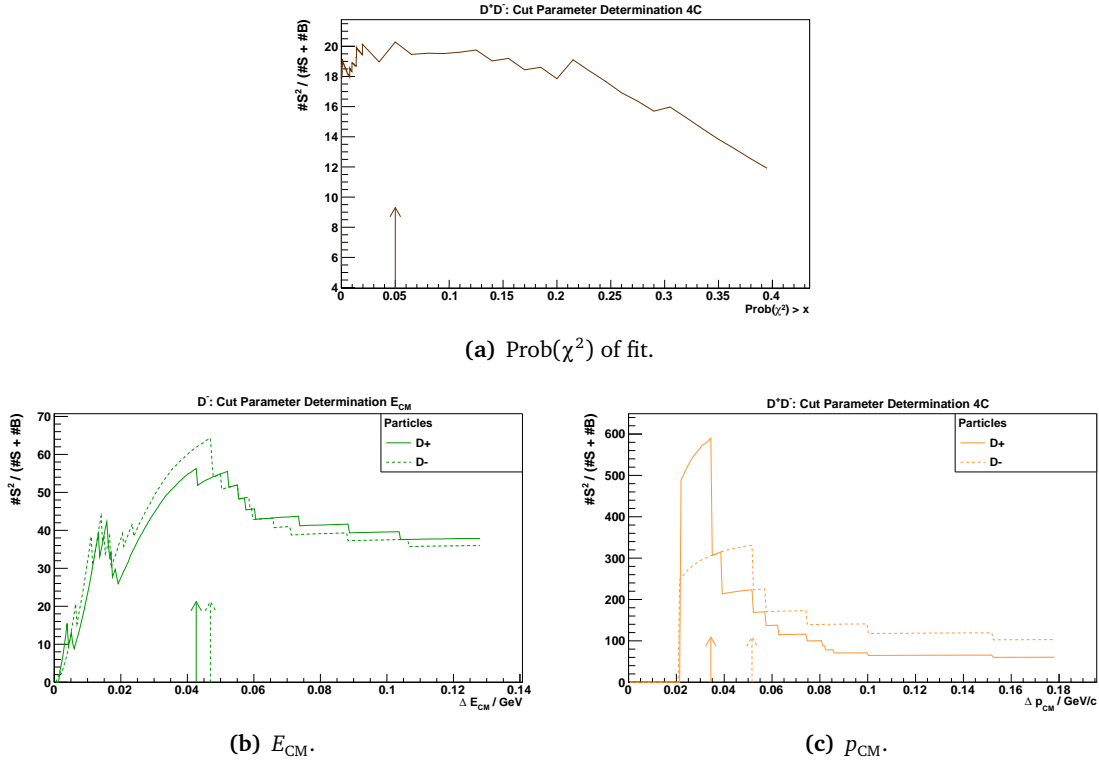
Figure 5.63: Significance $\mathcal{S}$ for optimizing the cut of the four-constraint fit, the CM energy, and CM momentum. The chosen cut values are marked.

5.4.4.3 Exclusive Decays

Using the event reconstruction of both branches of the $D^+D^- \to K^-\pi^+\pi^+ K^+\pi^-\pi^-$ decay, the background events can be suppressed much more efficiently. The four-constraint fit (see Section 5.4.3.1) assures that both D mesons originate from a common position and carry the four-momentum of the beam. The full event topology is known and can be consulted to define cuts.

In practice, the already introduced cut variables of the last sections, are sufficient to reduce the background to zero in the available data sample: A cut on the probability of χ^2 of the four-constraint fit and cuts on the CM energy and momentum, with different values for D^+ and D^-. The five parameters are optimized with respect to the significance ratio $\mathcal{S}$ in Figure 5.63. The scaling factor $\mathcal{F}$ of Equation 5.1 is appended for the exclusive analysis by an additional $\mathcal{BR}$ to incorporate the two $D \to K\pi\pi$ decays happening coincidentally; $\mathcal{BR} \to \mathcal{BR}^2$. The new factor, hence, relates as $\mathcal{F}^{\text{excl}} = \mathcal{F}/\mathcal{BR}$.

The optimization of the cut parameters in Figure 5.63 yields:

Four-Constraint Fit

$\quad$ Prob(χ^2)$_{4C} > 0.05$.

Center-of-Mass Energy

$1854.67\,\text{MeV} < E_{\text{CM}}(\text{D}^+) < 1897.33\,\text{MeV}$ and
$1852.53\,\text{MeV} < E_{\text{CM}}(\text{D}^-) < 1899.47\,\text{MeV}$.

Center-of-Mass Momentum

$143.793\,\text{MeV}/c < p_{\text{CM}}(\text{D}^+) < 178.207\,\text{MeV}/c$ and
$135.190\,\text{MeV}/c < p_{\text{CM}}(\text{D}^-) < 186.810\,\text{MeV}/c$.

While 76 % of the 8999 D meson pairs, which can be assembled into a D^+D^- system, pass the cuts, the full background of 399 fake D^+D^- candidates is reduced to zero. The scaling factor is $\mathcal{F}^{\text{excl}} = 106{,}000$.

For a statistically reliable statement the scaling factor is too large. Would only one single background event remain after applying the cut parameters, a factor of about 15 more background events compared to signal events would be present after scaling. For more reliable measurements, a much larger quantity of background-like events is needed. This is very challenging, as a lot of computation and data storage is necessary for the according simulation.

5.5 Comparison of Other Track Reconstruction Algorithms

The event reconstruction, as outlined in Section 5.4, is performed using the default tracking of PandaRoot. This *default chain* is compared to a reconstruction based on tracks from the Circle Hough Transform (CH), as presented in Section 4.4. In contrast to the simple event building, shown in Section 4.4.2.1, this offline-like chain uses full PID information and also re-fits the tracks from the CH using a Kalman filter. For an upper bound to the achievable reconstruction efficiency, the reconstruction chain is also conducted for tracks from an *ideal trackfinder*. This trackfinder solely relies on MC truth information to assemble tracks.[8] As a fourth option, the same ideal trackfinder as before is used to reconstruct tracks if at least two MVD hit points are present.

The number of particles over different stages, the *count flow*, is shown in Figure 5.64 for the ideal tracking and in Figure 5.65 for the offline PandaRoot tracking and the CH. In addition, the quantities are summarized in Table 5.3. The shown stages of the flow of D meson reconstruction are:

All The full number of simulated events.

> 3 Hits It is checked on the MC level if a kaon and two pions are present, which each leave more than three hits in one of the Target Spectrometer tracking sub-detectors (MVD, STT, GEM). The number of hits in the sub-detectors was subject of Section 5.3.

1 K The event contains at least one reconstructed kaon of the right charge and identified by PID.

---

[8]To emulate detector uncertainties, a Gaussian smearing on the track momentum and track vertices of 5 % is applied.
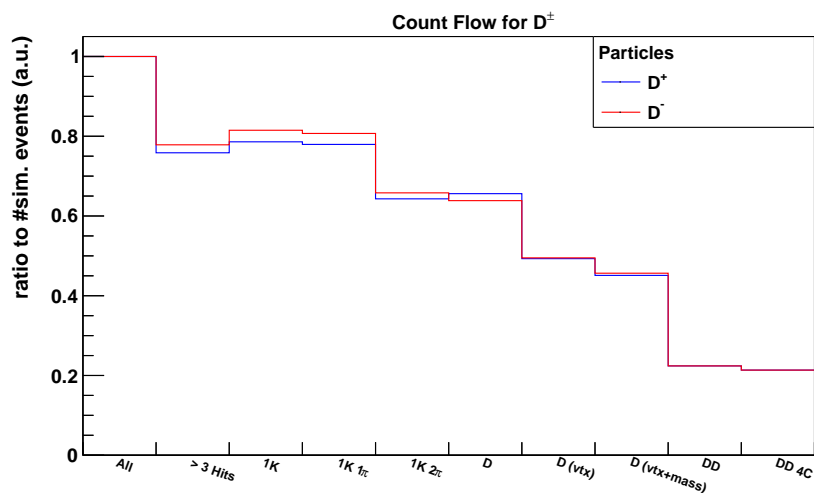
(a) Ideal Tracking.
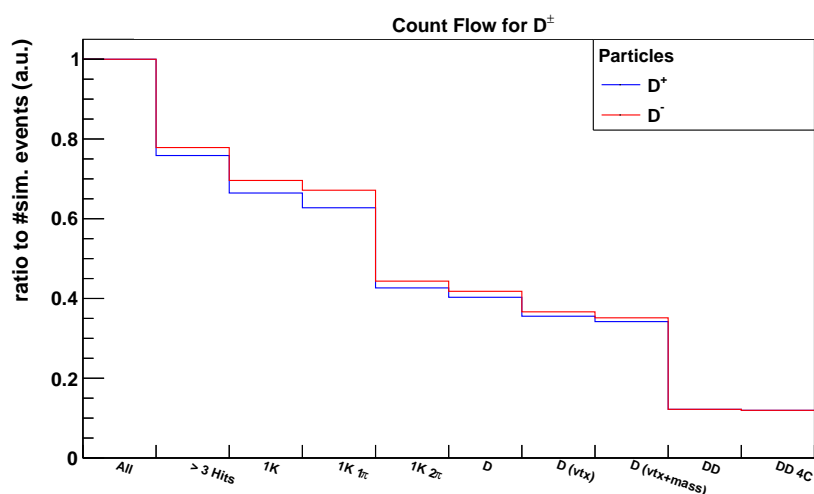
(b) Ideal Tracking ($\geq$ 2 MVD hits).

Figure 5.64: Progression of the relative numbers of D^{+} and D^{-} mesons for the individual steps of the reconstruction chain using ideal track reconstruction. The distribution on the top uses MC information for track reconstruction, the lower distribution uses the same reconstruction, but demands at least 2 MVD hits. The data is shown in Table 5.3.
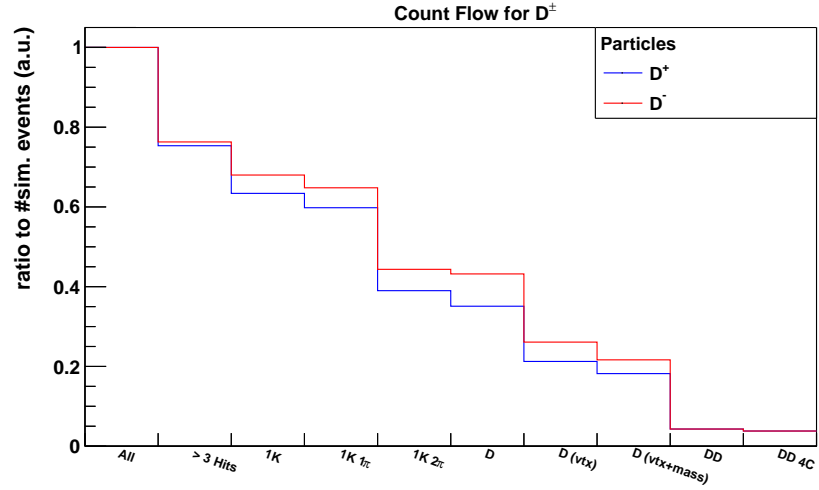
(a) PandaRoot default tracking.
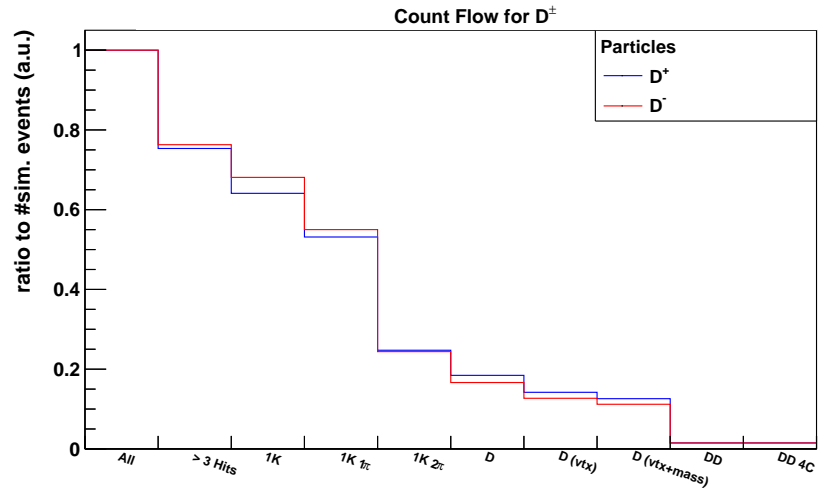
(b) Circle Hough Transform tracking (incl. Kalman).

Figure 5.65: Progression of the relative numbers of D^+ and D^- mesons for the individual steps of the reconstruction chain for the standard PandaRoot tracking (top) and a Kalman-supported CH (bottom). The data is shown in Table 5.3.

1 K 1 π In addition to the previous point, at least one pion of the right charge is reconstructed and identified.

1 K 2 π One kaon and at least two pions, all three with the correct charge, are reconstructed and identified in the event. The analysis of the final-state particles was carried out in Section 5.4.1.

D At least one D meson candidate can be combined using the kaons and pions in the event. The candidates of this stage were investigated in Section 5.4.2.1.

D (vtx) The D meson candidate passes cut_{vtx} (see Section 5.4.2.2).

D (vtx+mass) The D meson candidate passes cut_{vtx} and additionally cut_{mass} (see Section 5.4.2.3).

DD Two D meson candidates, which pass the previous cuts, can be combined into an exclusive D^+D^- system.

DD 4C The system of $D^+ D^-$ mesons passes cut_{4C} (Section 5.4.3.1).

For the case of the ideal track finder, giving an upper limit for realistic reconstruction, about 45 % $D^{\pm}$ candidates and 22 % D^+D^- pairs can be reconstructed. Already without including any detector and track reconstruction inefficiencies, only about half of the single D mesons and about a fifth of the D meson pairs can be reconstructed. The additional constraint of leaving at least two MVD hit points further reduced the amount of reconstructible meson candidates. For single $D^{\pm}$ candidates, the reduction is about 25 %, for D^+D^- pairs it is about 40 %. The constraint (≥ 2 MVD hits) is a coarse criterion to define the class of reconstructible tracks; only tracks which leave at least 2 hits can be used to formulate track hypotheses. Compared to the full number of simulated events, about one third of the D meson candidates can be reconstructed. Already this basic requirement of 2 MVD hits limits the reconstruction capabilities strongly. For the *default* tracking of PandaRoot, about 20 % of the initial D mesons can be reconstructed, excluding any background-related cuts. The reduction, when reconstructing the whole system of a D^+ and a D^- is stronger than for the ideal cases: only 4 % of all initial D^+D^- pairs can be reconstructed. For the CH, the fourth algorithm, the number is even lower with 1.5 %.

The largest reduction for all algorithms happens when demanding a second pion to be present in the event. Especially for the CH, a nearly 50 % reduction compared to the previous stage can be seen. The reason lies in the reconstruction capability of pion tracks: Statistically, one of the two present pion tracks can be reconstructed well enough to formulate a track hypothesis, but the quality of the track reconstruction of the second pion is not good enough to assemble a composite particle.

The second largest reduction can be seen when combining individual D meson candidates to systems of two for the exclusive study. The reduction is largely motivated by statistical arguments of combining the reconstruction probabilities of the individual candidates to a common system. For the default tracking and the CH, the combined reconstruction probability is slightly larger as the product of the individual $D^{\pm}$ probabilities, suggesting correlations between the reconstruction efficiencies of the meson candidates of the two charges.

Table 5.3: Number of events which have the indicated properties for different track reconstruction algorithms. »Ideal« is an ideal trackfinder, using MC information for track finding. »Ideal ≥ 2« is the same ideal trackfinder, but constrained to have at least 2 MVD hits. »Default« is the default PandaRoot track reconstruction algorithm. »CH« is the Circle Hough Transform algorithm. The plots visualizing the numbers are given in Figure 5.64 and Figure 5.65.

	Ideal N / %	Ideal ≥ 2 N / %	Default N / %	CH N / %
		D$^+$		
> 3 Hits	-	75.6	75.3	75.4
1 K	78.6	66.5	63.4	64.1
1 K 1 π	78.0	62.8	59.8	53.2
1 K 2 π	64.3	42.7	39.0	24.8
D	65.6	40.3	35.1	18.5
D (vtx)	49.3	35.6	21.6	14.2
D (vtx+mass)	45.1	34.2	18.2	12.6
		D$^-$		
> 3 Hits	-	77.9	76.3	76.3
1 K	81.5	69.6	68.0	68.1
1 K 1 π	80.7	67.2	64.8	55.0
1 K 2 π	65.8	44.4	44.4	24.4
D	63.9	41.8	43.2	16.6
D (vtx)	49.5	36.7	26.1	12.7
D (vtx+mass)	45.7	35.2	21.7	11.2
		D$^+$D$^-$ System		
DD	22.4	12.2	4.3	1.5
DD 4C	21.4	12.0	3.9	1.5

The charge asymmetry between D$^+$ and D$^-$ (see Section 5.3.2) is present for all four tracking algorithms, but is most pronounced for the default tracking algorithm. The inherent algorithmic properties seem to be especially susceptible to the reduced number of hits for π^+ tracks.

Overall, the default tracking leads to D candidate reconstruction results of about 50 % compared to the constrained ideal tracking for single D mesons, and 35 % for D$^+$D$^-$ pairs. The CH algorithm has a lower efficiency, as only about a third of the constrained *ideal* D meson candidates are found (D$^+$D$^-$ pairs: 12 %). As already noted in Section 4.4.2.3, the number is too low for an actual productive environment. Important improvements are already in the course of implementation.

5.6   Summary and Discussion

In this section, the physics channel of $\bar{p}p \to D^+D^- \to K^-\pi^+\pi^+K^+\pi^-\pi^-$ was studied extensively, using a beam momentum of $6.5\,\text{GeV}/c$.

5.6.1   Summary

In Section 5.2, the MC simulation of the decay of a D meson into $K\pi\pi$ was investigated. The quantities of the simulation, which is based on data from CLEO, were successfully recovered.

The multiplicities of hits from kaons and pions in the four tracking sub-detectors MVD, STT, GEM, and FTS was studied in Section 5.3. The importance of the individual sub-detectors of $\overline{\text{PANDA}}$ was seen in the multiplicity distributions. The MVD has a position close to the interaction point and measures hit points for most of the particles: for 75 % of all pions, 4 ± 1 hit can be measured. The quantity of hit points is comparably low, in particular with respect to the STT. The STT covers a large angular range and provides usually the most hit points per particle, if it is hit: 26 both for pions and kaons. For low polar angles, the FTS can measure up to 35 hit points on average. The intermediate polar range between the FTS and the STT is mitigated by the GEM, which provides usually 6 hit points if it is hit.

Pions leave more hits in the sub-detectors than kaons, since they have a longer life-time and are emitted to larger polar angles. Over the whole polar range to 140°, 90 % of all pions leave more than four hit points. For kaons, the polar range is more limited; the particles only leave four hit points up to 45°. Between 10 % (pions) and 20 % (kaons) of the particles decay in the area enclosed by the DIRC and the forward endcap of the EMC. The higher the probability of decay inside of the sub-detectors is, the less likely is a reconstruction of the full track of the particle. A number of events were found, in which pions do not leave a hit in any of the sub-detectors but the MVD, but still reach the EMC. A study showed that they curve through the uninstrumented STT slit, which is introduced by the target beam-pipe.

A study of the positively and negatively charged states of the particles showed that π^- are detected with a significant larger multiplicity. The problem is most pronounced in the barrel layer 3 of the MVD. The charge asymmetry was back-traced to the MC propagator employed; the effect vanished when switching from GEANT3 to GEANT4. The reason for the unaccounted hits is still unknown and subject for future studies.

In Section 5.4, the event reconstruction was studied. First, the particles of the final state were analyzed. In 65 % of all events, one kaon was found, in about 60 % two pions. The energy and momentum resolutions for pions are approximately $8\,\text{MeV}$ and $8\,\text{MeV}/c$, respectively. For kaons they are slightly worse.

The kaons and pions were taken to assemble D meson candidates. In 42 % of all events, a D^- meson can be combined from the particles of the final state. The number of D^+ is lower (35 %), as this particle incorporates a π^+, which suffers from the charge reconstruction asymmetry. The D meson candidates were fit with a vertex fitter to reconstruct the decay vertex from the pion and kaon tracks. The vertex resolution is about $64\,\mu\text{m}$ in the xy-plane and $100\,\mu\text{m}$ in the z-direction. Already when combining pions and kaons to D meson candidates, in some cases

more than one D candidate can be formed. The D meson multiplicity per event was studied after the vertex fit and it was found that the individual candidates are very similar and based on multiply reconstructed tracks. The D candidate, which has the lower χ^2 from the vertex fit, is picked as the candidate to succeed. The transverse momentum resolution after the vertex fit is 8 %.

A second fit constrains the mass of the D candidate to that of the PDG values. Only candidates from the vertex fit, which pass also the mass constraint fit, were considered further. *Passing* means for both fitters that the χ^2 probability is greater than 1 %. After both fits, about 20 % of the D candidates remain. They have a mass resolution of $\sigma_m = 15\,\mathrm{MeV}/c^2$ and an energy resolution of $\sigma_E = 30\,\mathrm{MeV}$.

Continuing, the individual D^+ and D^- candidates were combined into an exclusive D^+D^- system. For about 4.5 % of the events the combination is successful. A four-constraint fit constrains the system of the D meson candidates to the four-momentum of the beam-target system. Applying the same probability cut as before, 4 % of the initial D mesons remain. In a study of the pull distribution of the momenta in all three spatial coordinates, it was seen that the measurement uncertainties seem to be overestimated ($\sigma_{\mathrm{pull}} = 0.6$). Mass and energy resolutions after the four-constraint fit are $9\,\mathrm{MeV}/c^2$ and $26\,\mathrm{MeV}$, respectively for both D candidates.

The background of the beam momentum was studied using the DPM generator. First, the inclusive events of single D mesons were investigated. For optimal background suppression, the probability cuts of the vertex fit and the mass constraint fit were benchmarked with respect to the significance ratio, $S^2/(S + B)$. Slightly stronger cuts than for the preliminary study previously were found. To suppress the background even further, cuts on the CM energy and CM momentum were introduced. Using ideal cut positions, 2.8×10^{-5} % of the background was still present. With the appropriate scaling factor, the number of background events is 10 times higher than the number of signal events. For inclusive decays, without any further cuts, seeing a signal peak is very challenging.

For the case of exclusive decays, the procedure of optimizing the significance ratio was repeated after the four constraint fit. While 3 % of the initial D mesons are still present after the cuts, all of 1×10^8 simulated DPM background events are removed. A first semi-inclusive analysis was done which uses the remaining particles in an event with a reconstructed D meson candidate for discrimination. Using the average minimum distance of tracks to a fitted D candidate decay vertex and optimizing cut parameters, a reduction to 1.4×10^{-5} % was reached. Further investigation is needed to utilize even more parameters of the event toplogy for suppression.

In the last part of this chapter, Section 5.5, the performance of the analysis of the D decay was studied with other tracking algorithms. An ideal trackfinder, which takes MC truth information for track determination, is compared to the PandaRoot default tracking and the CH algorithm introduced in Section 4.4. The largest efficiency reduction, independent from of employed algorithm, can be seen when demanding an event which has a kaon and *two* pions. Especially for the CH, the number of events which fulfill this criterion is reduced to 25 %, already before assembly to D candidates. The charge asymmetry seems to be most significant for the default tracking, as the difference between reconstructed D^+ and D^- is largest here. For an ideal

trackfinding, which has the additional realistic constrain of incorporating only tracks with at least two MVD tracks, the efficiency for reconstructing a system of D^+D^-, passing the four constraint fit, is 12 %. The comparable number for the default tracking is 4 %; a third of the ideal number.

5.6.2 Discussion and Outlook

During the study of the multiplicity of hits in the four different tracking sub-detectors, the significance of the individual detectors was highlighted and understood. Pions have a higher detection efficiency than kaons, largely dependent in the pion's larger lifetime. Although the efficiency for pions is relatively high, it is still the limiting factor for realistic track reconstruction algorithms to form tracks – and subsequently, for the combination of D meson candidates. A detailed study is needed to investigate the reason underlying this loss in efficiency for secondary pions. It needs to be assessed if this large difference is only a result of statistical reconstruction inefficiencies or the fact has some underlying systematical reasons.

The charge asymmetry was identified to lie in the employed MC generator. Further analysis is needed to exactly find the reason why GEANT3 does not recognize some hits of positive pions.

The event reconstruction has its largest efficiency loss due to the not properly reconstructed second pion. After this, the fitting and event cleaning was straight-forward. Unfortunately, the background could not be suppressed fully in the inclusive decays. Many different discrimination values have been benchmarked during the course of this thesis, the two most effective presented in this work. Still, further research is needed to find more complex event variables to suppress the background more.

The background of the exclusive analysis is reduced to zero from the sample data set. Here, more DPM background data is needed, to make statistically sustainable statements. This is very challenging, as already the produced 100 million events can be realized only with very high computational and storage efforts. A possibility to reduce the demand of the simulation concerning computing time and storage space could be to apply a simple *online* discrimination already during simulation. This way, for example no PID algorithm invocation would be needed for particles of which the tracks have already been identified as non-relevant to the simulation.

A semi-inclusive analysis is currently ongoing, which combines the good background suppression of the exclusive reconstruction with the higher D meson numbers of the inclusive analysis.

The comparison of ideal and realistic track reconstruction algorithms showed that the CH algorithm suffers from the inefficiencies in track reconstruction, already seen in Section 4.4.2.1. For the usage of this algorithm to reconstruct events in an online environment, further work is needed to match the reconstruction quality of the PandaRoot default tracking.

Using a luminosity of $\mathcal{L} = 2 \times 10^{32}/(\text{cm}^2\,\text{s})$, a production cross section of $\sigma_{D^+D^-} = 100\,\text{nb}$, and the efficiency of the inclusive decay, about 1300 single D mesons of each charge can be measured at $\overline{\text{P}}$ANDA, not accounting for background events. Using the exclusive decay with an efficiency of 4 %, D^+D^- pairs can be measured with a rate of about 25 per hour.

Chapter 6

Conclusion

In this thesis, track reconstruction algorithms running on GPUs have been studied, and the decay channel $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ analyzed.

It was shown that the algorithms and the infrastructure employed can match the demands of $\overline{P}ANDA$'s online event reconstruction. The decay channel was reconstructed and extensively tested in an offline environment. An algorithm for online track reconstruction was used for the same reconstruction, also in a simplified online reconstruction.

6.1 Online Track Reconstruction

Online track reconstruction is a vital part in the online event reconstruction of $\overline{P}ANDA$. In Chapter 4, a first basis of track reconstruction algorithms was laid. The algorithms employ the high-performance devices of GPUs to reconstruct tracks with great speed. Three different algorithms were presented; the Line Hough Transform, the Circle Hough Transform, and the Triplet Finder.

6.1.1 Summary

Line Hough Transform   The Line Hough Transform (LHT) is based on an approach, which finds application in many different track reconstruction setups (e.g. the offline track reconstruction of $\overline{P}ANDA$). For sets of hit points a large number of possible straight track parameters is sampled and the most probable parameters extracted. The approach is inherently parallel and thus very suitable to compute on GPUs.

Two different implementations of the LHT were tested. One uses fast, pre-programmed routines of the package Thrust; one is tailored deeply for $\overline{P}ANDA$ and makes use of advanced acceleration techniques (the *CUDA approach*). The peak performance for the CUDA approach is 30 khit/s, significantly outperforming the inflexible Thrust-based approach.

Circle Hough Transform   The Circle Hough Transform (CH) builds on the idea of the LHT and extends it greatly. The algorithm has been developed from scratch during the work on this

thesis. Instead of sampling line parameters with needed pre-processing steps, the CH directly Hough transforms circles as track representations.

The transform achieves a track momentum resolution of 4 %, comparable to that of the offline tracking in $\overline{\text{P}}$ANDA. A simple online event builder was programmed to combine kaon and pion tracks to D meson candidates. An efficiency of 80 % and a mass resolution of about 35 MeV/c^2 was achieved. Slightly better than the corresponding numbers obtained using the offline track reconstruction algorithms from PandaRoot. The algorithm was implemented and studied on GPUs. The best performance attained was 30 Mhit/s for the actual Hough transformation part of the algorithm. The current bottleneck of the algorithm are parts, which still rely on the CPU for processing. Work is on the way to convert the CH algorithm fully to the GPU device and decouple it from any CPU processes. Also, algorithmic improvements are investigated, as the track reconstruction efficiency of the algorithm (90 %) is still not ideal.

Triplet Finder   This algorithm exclusively uses STT hits to reconstruct tracks and is built from the beginning to run time-wise without event structures. The Triplet Finder (TF) is completely implemented on the GPU and extensively optimized. Performance measurements show that the algorithm reaches peak performance of 11 Mhit/s for the whole algorithm on a HPC-grade Tesla K40 card. Promising results are obtained when the algorithm is run on a consumer-grade GeForce GTX 750 Ti card: a performance of nearly 5 Mhit/s is reached. This is significant as this card costs only a small fraction of the HPC-grade cards.

6.1.2   Remarks and Outlook

When studying the different algorithms for online track reconstruction, one important fact emerged: Employing GPUs for online track reconstruction in $\overline{\text{P}}$ANDA seems feasible. The devices can be a solution to realize the immense computational requirements of the experiment's novel DAQ scheme in terms of reconstruction speed.

Although it was shown that the well-known Line Hough Transform algorithm can have the necessary performance on GPUs, the Circle Hough Transform seems to be the more promising algorithm. Specific track hypotheses of chosen transverse momenta can be sampled directly. The algorithm has a good resolution and achieves close to 100 % reconstruction efficiency in first studies of enhanced algorithmic routines. At the same time, the algorithm is massively parallel. It can match the needed reconstruction efficiency, resolution, and performance.

During the optimization of the Triplet Finder a multitude of optimization parameters for GPU devices were explored and studied in detail. A high speed for the Triplet Finder was reached, but at the same time techniques were developed which can also be applied to other algorithms, e.g. the Circle Hough Transform.

The next steps for *online track reconstruction on GPUs for* $\overline{\text{P}}$*ANDA* are the algorithmic and performance optimization of the CH and research on embedding the GPU devices into the DAQ data stream of $\overline{\text{P}}$ANDA.

6.2   Event Reconstruction of $D^{\pm} \rightarrow K^{\mp}\pi^{\pm}\pi^{\pm}$

In Chapter 5 of this thesis, the reaction $\bar{p}p \rightarrow D^+D^- \rightarrow K^-\pi^+\pi^+K^+\pi^-\pi^-$ was extensively reconstructed and analyzed.

6.2.1   Summary

The data obtained from the simulation with PandaRoot was verified to match the producing parameters from the input Dalitz decay.

In a detailed study of the multiplicity of hits through the four main tracking sub-detectors of $\overline{P}$ANDA (MVD, STT, GEM, FTS), different features of the decay were assessed. The STT, as $\overline{P}$ANDA's central tracking detector, provides the bulk of all hit points. The MVD, close to the interaction point, measures particle hit points in most of the cases and covers a large angular range. It was found out that pions have a better reconstruction probability as kaons, since they leave more hit points over a larger polar angle range.

D^+ and D^- candidates were combined from reconstructed π and K. Even before applying further steps of reconstruction, only in 42 % and 32 % of the events a D^- and D^+ meson candidate was assembled, respectively. To reconstruct the topology of the event and discriminate the D meson candidates from false combinations, two fits were employed in the inclusive reconstruction: A vertex fit, to reconstruct the D decay vertex, and a mass constraint fit, to match the D candidate mass to the kinematics of the event. Applying cuts on the probability of the χ^2 of the fits, about 20 % of the D meson candidates were still present. The energy resolution of the D mesons is 30 MeV, the mass resolution 15 MeV/c^2.

In the next part of the event reconstruction, two D meson candidates were combined to a D^+D^- system. A four-constraint fit was applied to match the four-momenta of the beam and the target. Requiring the χ^2-probability cut as before, in about 4 % of the cases a pair of D mesons was reconstructed.

With the high-luminosity mode of $\overline{P}$ANDA, about 1300 single D mesons of each charge can be measured per hour, assuming a cross section of $\sigma_{D^+D^-} = 100$ nb. For the case of the exclusive reconstruction of both D^+D^- mesons, about 25 meson pairs can be measured per hour.

The ability to reconstruct the signal decay channel under realistic experiment conditions was studied with the DPM background generator. In the inclusive decay, the cuts of the vertex fit and the mass constraint fit were optimized for the best $\mathcal{S} = S^2/(S+B)$ ratio, with S being the number of signal events and B the number of background events. Two additional cuts were introduced: Cuts on the center-of-mass energy and the center-of-mass momentum. Incorporating all cuts, the background of 1×10^8 events was reduced to 28; about 1.4×10^{-5} %. Applying the appropriate scaling factors and assuming an optimistic $\bar{p}p \rightarrow D^+D^-$ production cross section of 100 nb, still a factor 10 more background events than signal events are present.

In the exclusive analysis, which reconstructs both branches of the decay, the optimization of the previous cut parameters was repeated, individually for both D mesons. While about 3.5 % of the signal D candidates are still present after the cuts, no background event remains.

A first effort was made to combine the good background suppression of the exclusive reconstruction with the high efficiency of the inclusive reconstruction: a first semi-inclusive reconstruction was able to reduce the background by a factor of 2, compared to the final number in the inclusive reconstruction.

Finally, three types of track reconstruction algorithms were compared through the steps of the event reconstruction. An ideal trackfinder gave an upper limit ony the achievable performance, the default reconstruction from PandaRoot provided the well-known characteristics, and the introduced Circle Hough Transform was supplemented by a Kalman filter to benchmark the current status of the developed algorithm in the context of a full reconstruction. Individual D mesons can be reconstructed with an efficiency of about 35 % for an ideal trackfinder. The efficiency for the offline reconstruction is, as presented, 20 %. The performance of the Circle Hough Transform is 12 %; slightly more than half of it.

6.2.2 Remarks and Outlook

The principle ability to reconstruct D mesons in $\overline{P}ANDA$ was confirmed in this thesis. A mass resolution of better than $10\,\mathrm{MeV}/c^2$ was achieved for exclusive decays.

A challenge in the reconstruction is the large background of $\overline{p}p$ annihilations at $\overline{P}ANDA$'s beam momentum. Especially in the case of the inclusive decay, a reconstruction is at best challenging. Further discrimination parameters need to be found to see a signal peak on top of the large amount of background.

For exclusive decays the reconstruction of D^+D^- events seems feasible. Here, further background data is needed to make statistically sound statements. The utilized data set consisted already of 1×10^8 events. Event numbers of an additional magnitude would be challenging from a simulation point of view. A possible solution is the analysis of the type of DPM events which are still present after the different cuts to target those specifically in the simulation.

Using the Circle Hough Transform algorithm for reconstruction of D meson candidates was successful. This suggests that also in an online event reconstruction environment, the Circle Hough Transform is capable of reconstructing the decay channel for possible triggering. The only drawback is the low efficiency, also when used as an offline algorithm. Here, the algorithmic enhancements already discussed can have a large impact.

Appendix A

Additional Figures
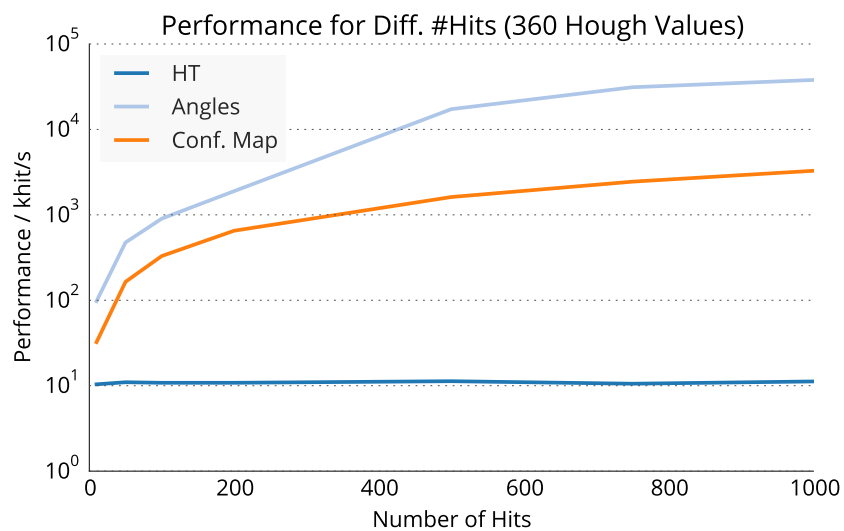
Figure A.1: Line Hough Transform: Logarithmic plot of performance of the Thrust-based Line Hough Transform as a function of the number of hits. The number of angles is fixed to 360. The corresponding time measurement plot is Figure 4.10.
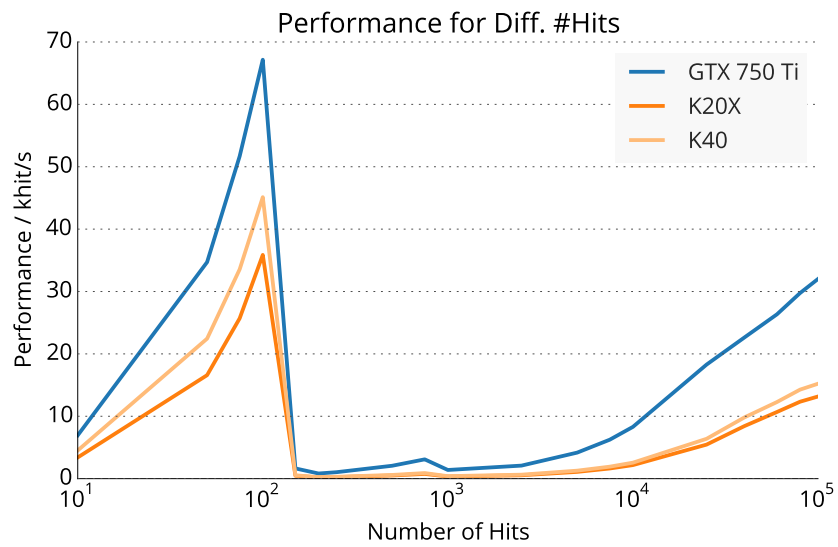
Figure A.2: Line Hough Transform: Logarithmic plot of performance of the adaptive Line Hough Transform for different GPU devices. See Figure 4.13(b) for the linear plot. The peak in performance at $\mathcal{O}(100)$ hits is due to an ideal utilization of GPU cores and memory.

(a) $\Delta p_{t,\mathrm{rel}}$ distribution for kaons.

(b) $\Delta p_{t,\mathrm{rel}}$ distribution for pions.
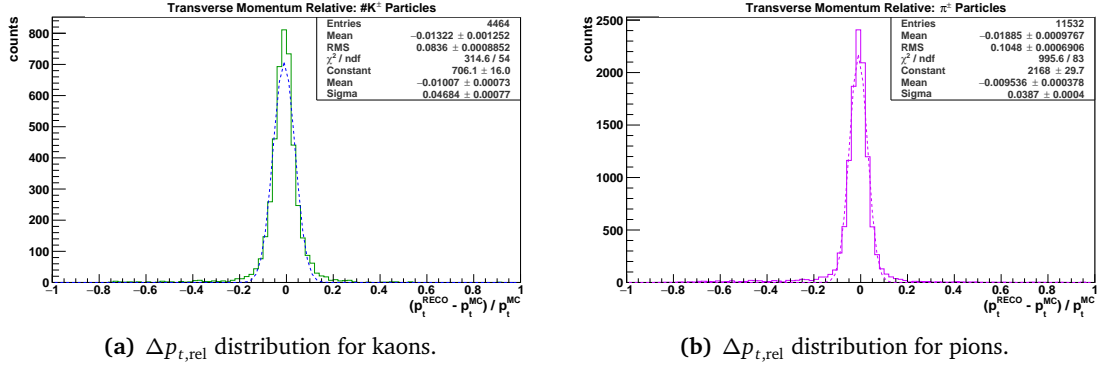
Figure A.3: Circle Hough Transform: Relative track reconstruction quality split by particle species into K and π. This is the MC-normalized plot of Figure 4.24.
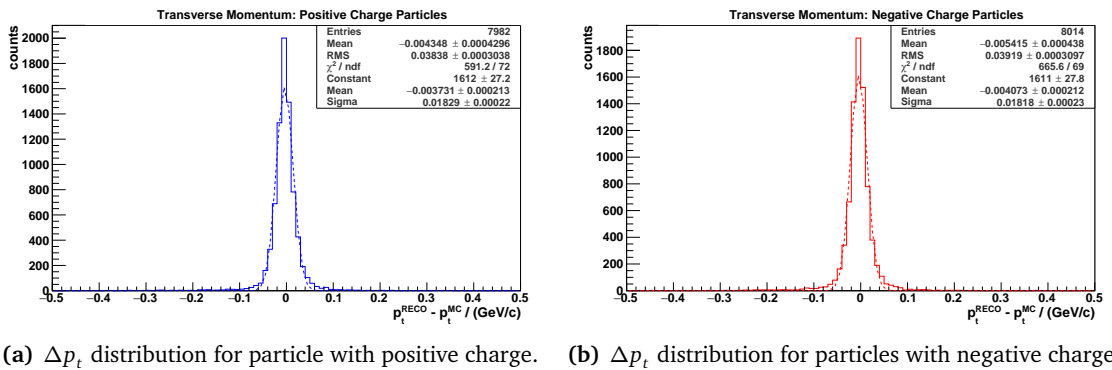
(a) Δp_t distribution for particle with positive charge.

(b) Δp_t distribution for particles with negative charge.

Figure A.4: Circle Hough Transform: Track reconstruction quality for both charge states.
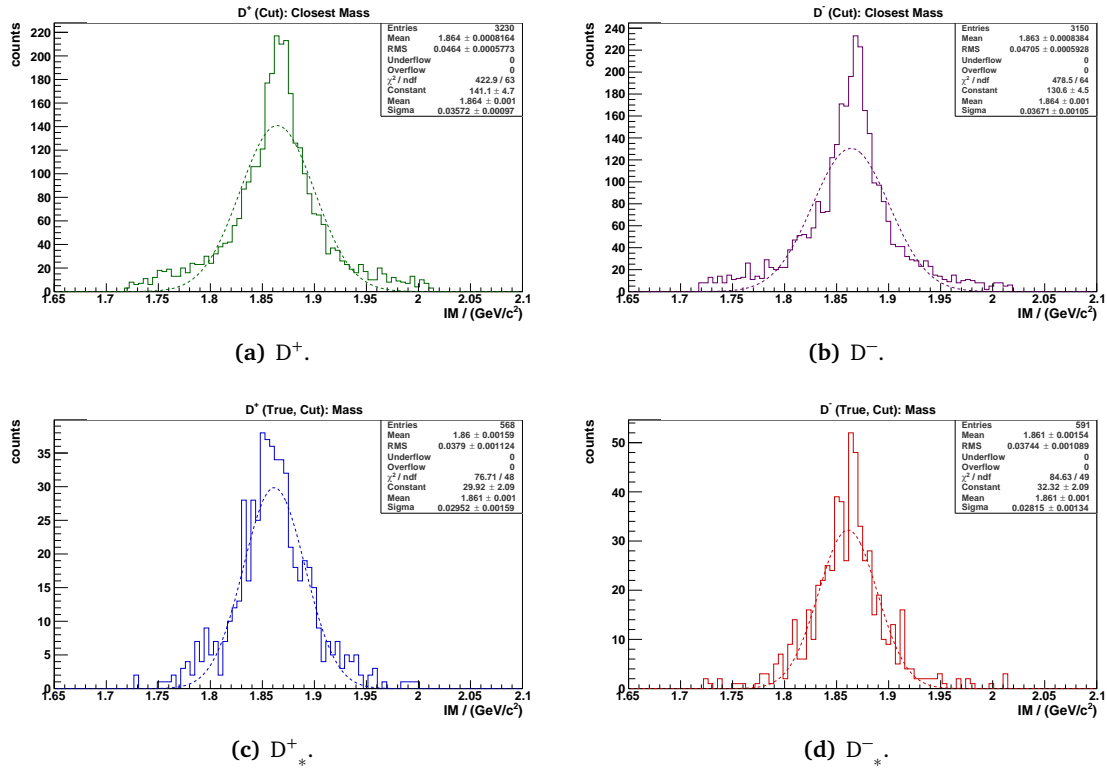
Figure A.5: Circle Hough Transform: Invariant mass distributions of D candidates assembled in the simple online event reconstruction for particles with tracks reconstructed in the Circle Hough Transform. D$_*$ refers to the particles which were matched to the original track, as outlined in Section 4.4.2.1.

(a) D$^+$ with Kalman Filter.

(b) D$^-$ with Kalman Filter.

(c) D$^+$ without Kalman Filter.

(d) D$^-$ without Kalman Filter.
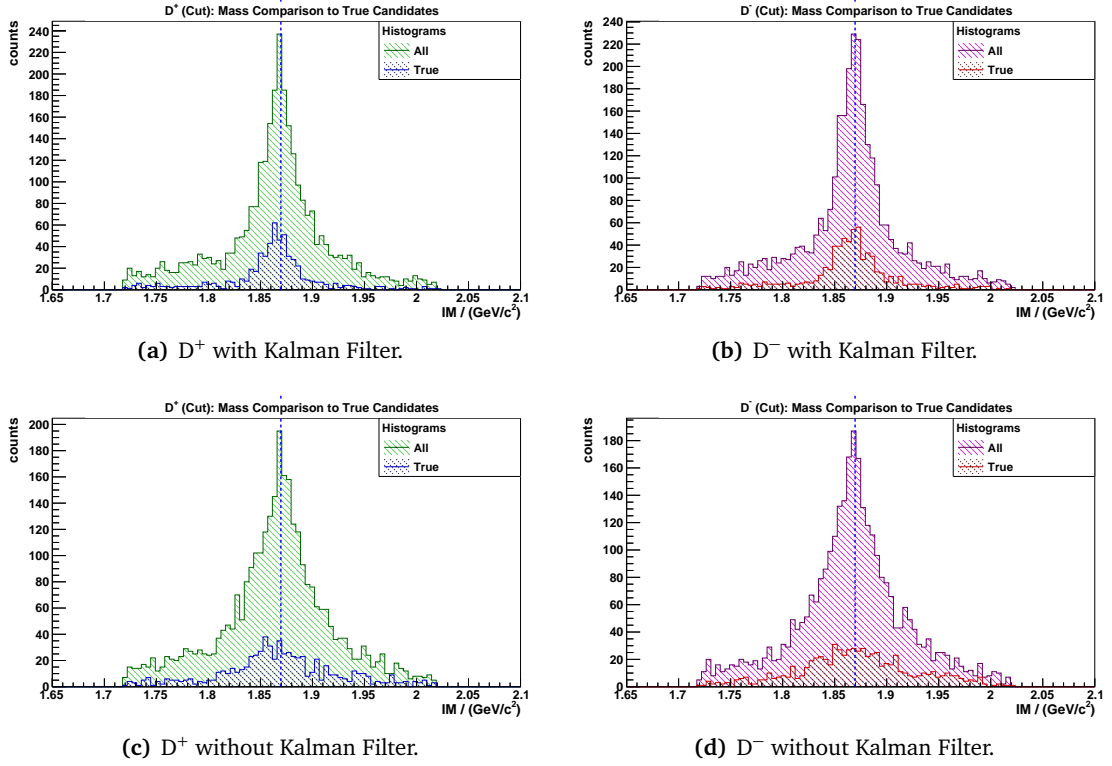
Figure A.6: Circle Hough Transform: Invariant mass distributions of D candidates assembled in the simple online event reconstruction for particles with tracks reconstructed with the offline track reconstruction of PandaRoot. In the lower row, the Kalman filter is turned off. Each graph compares the reconstructed D mass with the mass of the MC-matched candidate, as outlined in Section 4.4.2.1.

(a) With Kalman Filter.

(b) Without Filter.
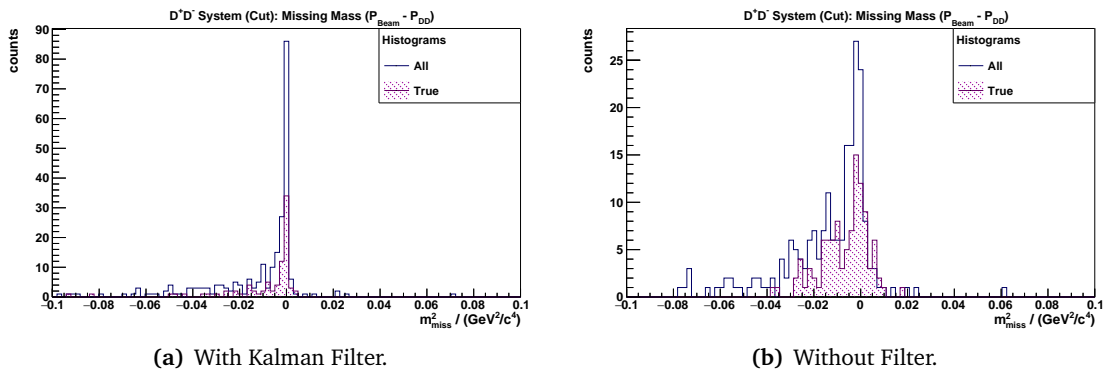
Figure A.7: Circle Hough Transform: Distributions of the missing mass squared of the D$^+$D$^-$ system after cuts for particles from the $\overline{\text{P}}$ANDA offline track reconstruction. $\Delta m^2 = (E_{\text{beam}} - E_{\text{D}^+ \text{D}^-})^2 - (p_{\text{beam}} - p_{\text{D}^+ \text{D}^-})^2$.
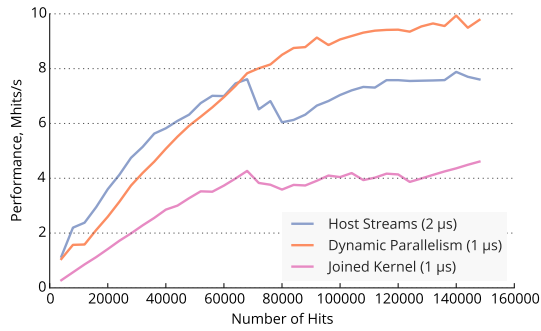
(a) Bunching Schemes (Tesla K20X).
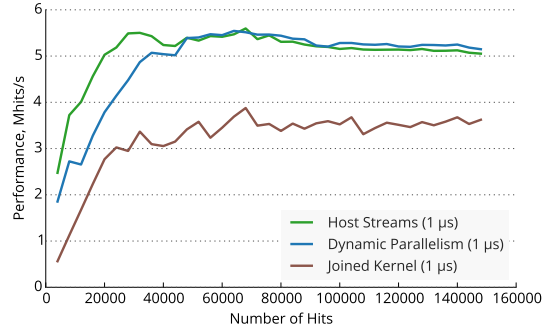
(b) Bunching Schemes (GeForce GTX 750 Ti).
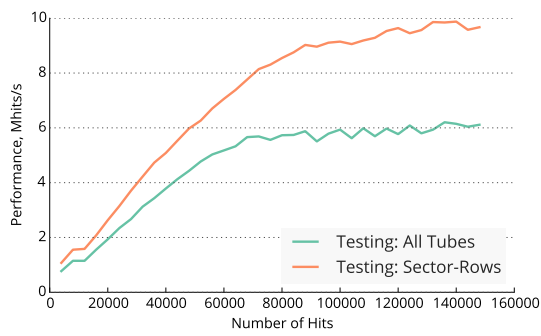
(c) Sector-Row Testing (Tesla K20X).
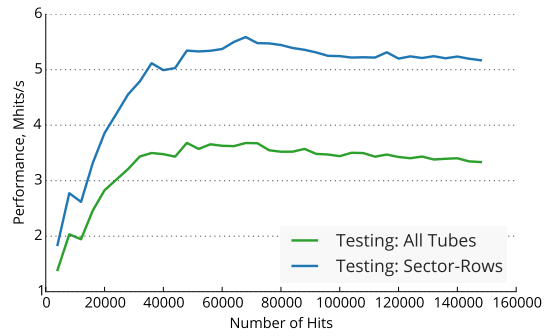
(d) Sector-Row Testing (GeForce GTX 750 Ti).
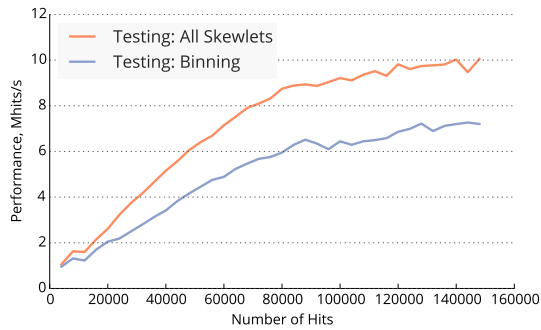
(e) Skewlet Binning (Tesla K20X).
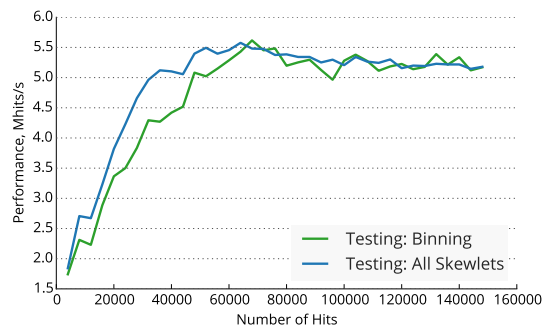
(f) Skewlet Binning (GeForce GTX 750 Ti).
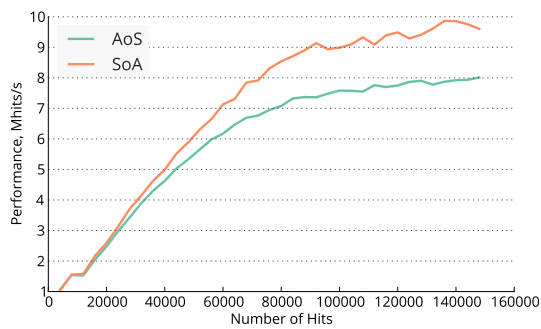
(g) SoA and AoS (Tesla K20X).
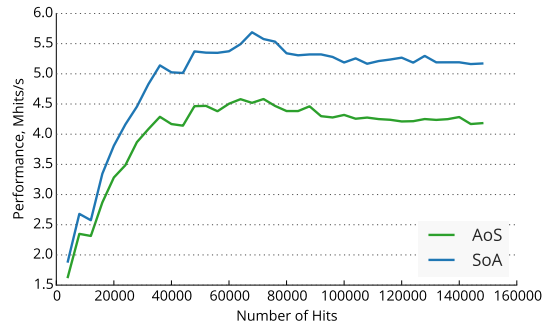
(h) SoA and AoS (GeForce GTX 750 Ti).

Figure A.8: **Triplet Finder**: Comparison of performance optimizations of the Triplet Finder on Tesla K20X and GeForce GTX 750 Ti, as presented in Section 4.5.2.2. Left plots: Tesla K20X, right plots: GeForce GTX 750 Ti.

(a) MVD.

(b) STT.

(c) GEM.

(d) FTS.
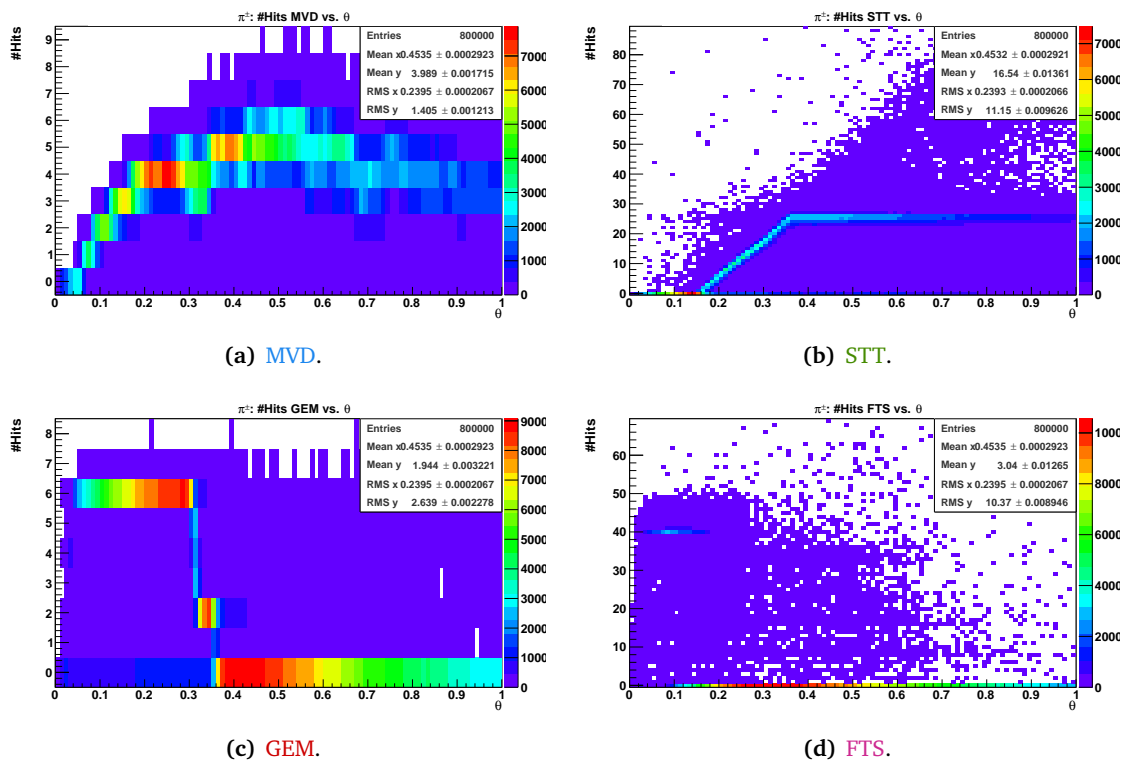
Figure A.9: Hit Multiplicity Analysis: Hit multiplicity distribution of **pions** for four tracking sub-detectors. Hits per track vs. θ. These plots are the basis for the profile graph shown in Section 5.3.

(a) MVD.

(b) STT.

(c) GEM.

(d) FTS.
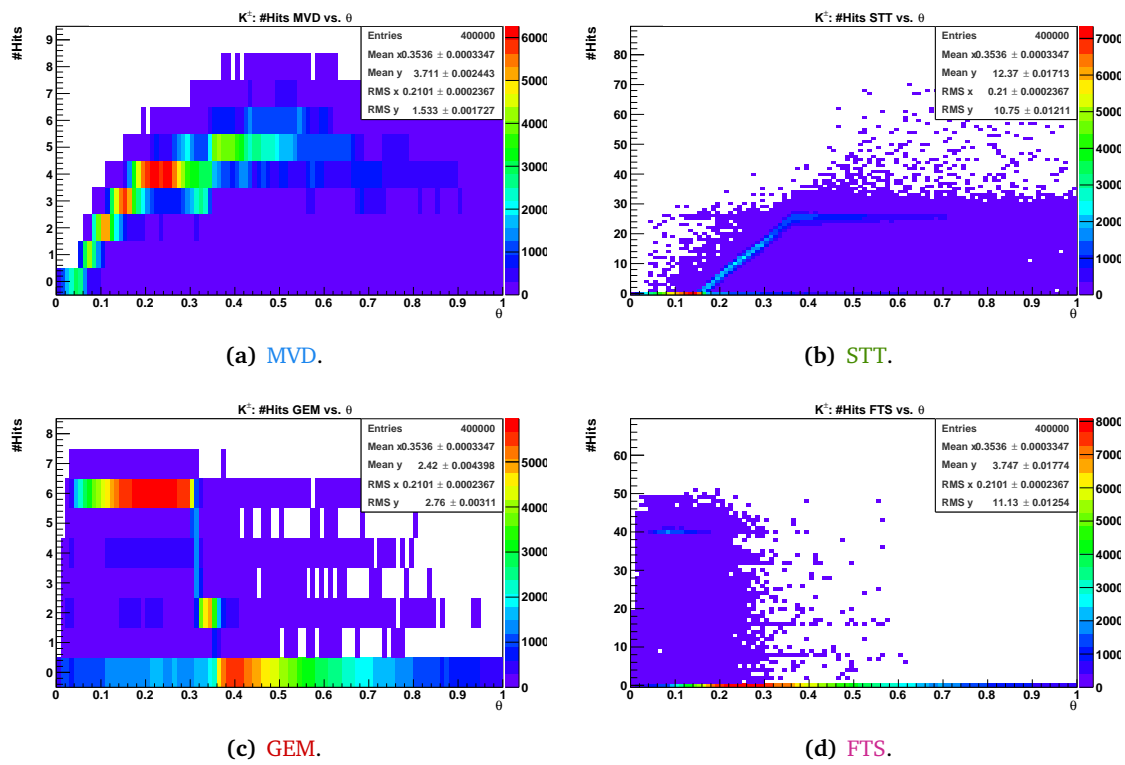
Figure A.10: Hit Multiplicity Analysis: Hit multiplicity distribution of **kaons** for four tracking sub-detectors. Hits per track vs. θ. These plots are the basis for the profile graph shown in Section 5.3.

(a) MVD.

(b) STT.

(c) GEM.

(d) FTS.
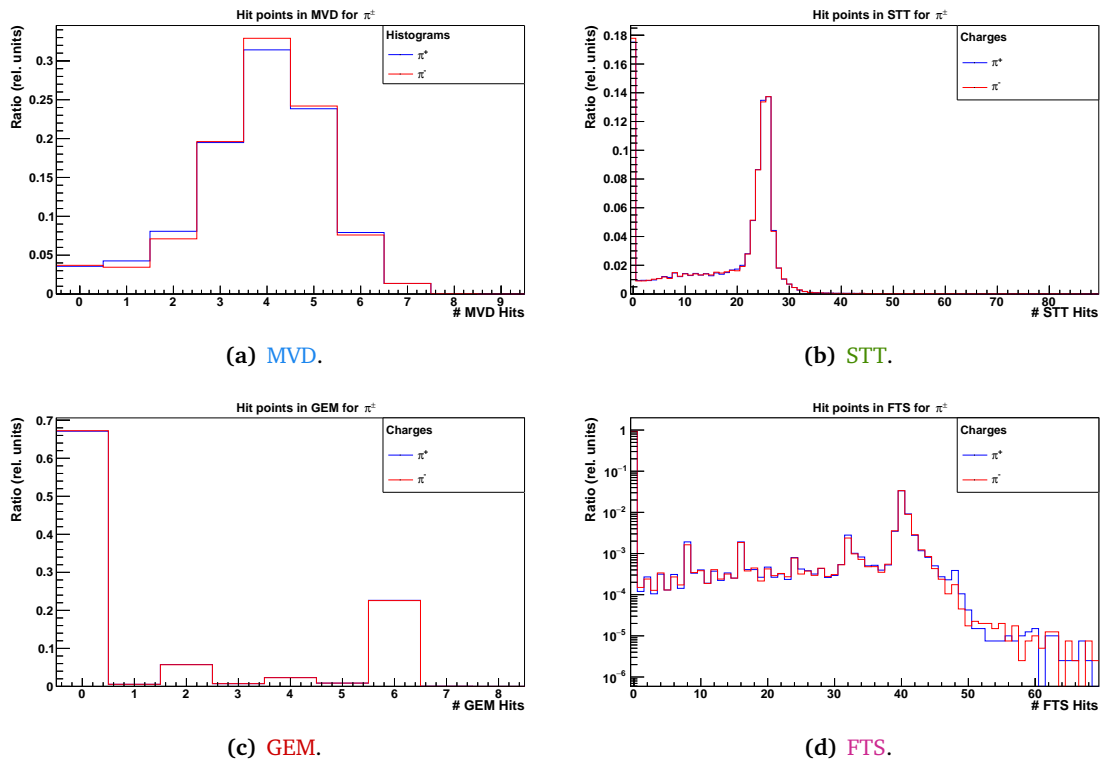
Figure A.11: Hit Multiplicity Analysis: Hit multiplicity distributions of π^+ and π^- for different tracking sub-detectors.

(a) MVD.

(b) STT.

(c) GEM.

(d) FTS.
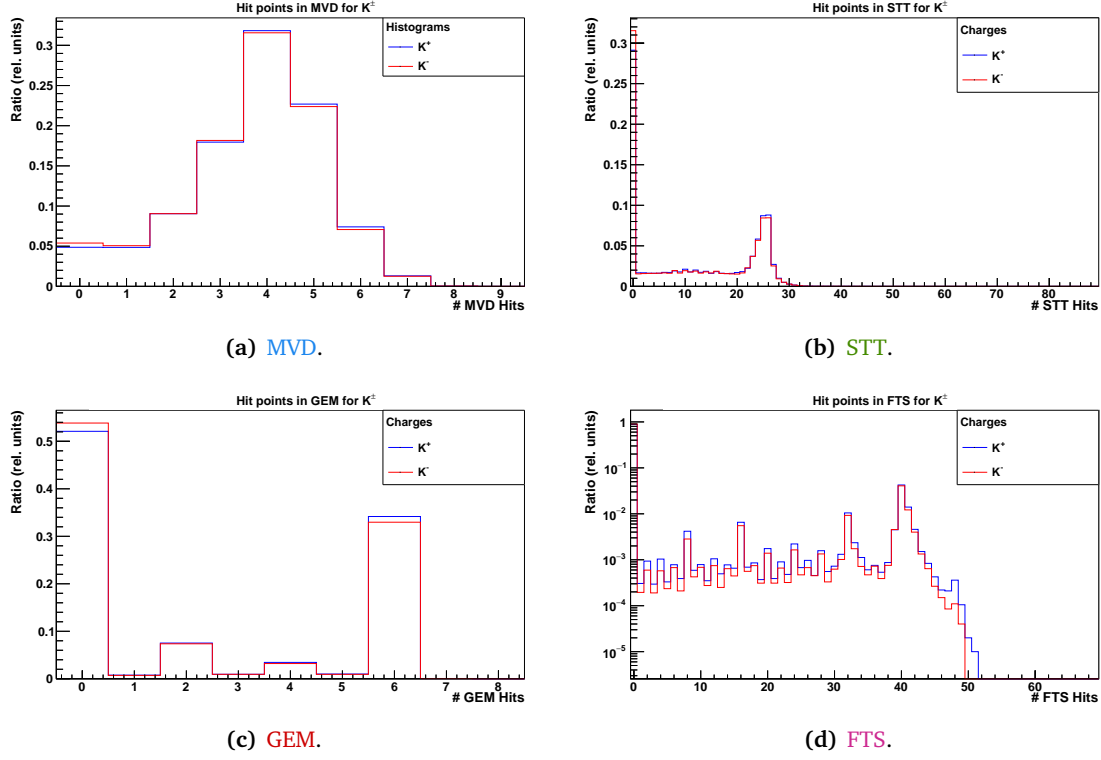
Figure A.12: Hit Multiplicity Analysis: Hit multiplicity distributions for K$^+$ and K$^-$ for different tracking sub-detectors.

(a) π$^+$.

(b) π$^-$.
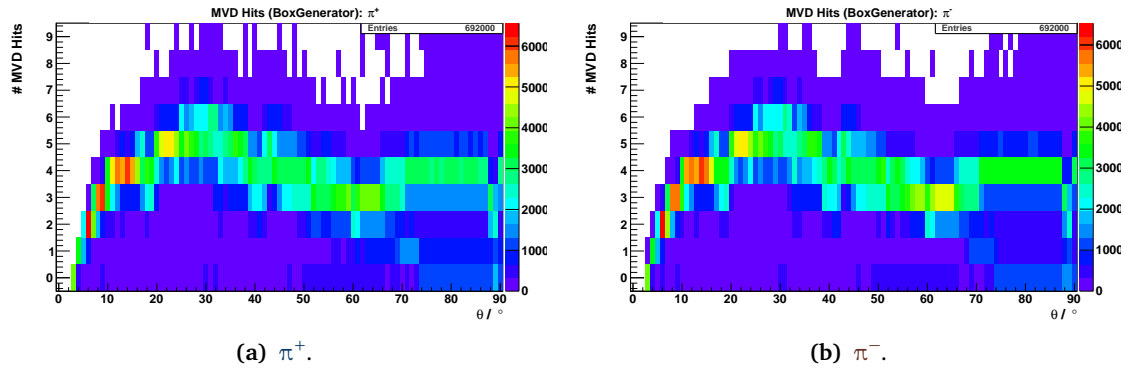
Figure A.13: Hit Multiplicity Analysis: Hit multiplicity distributions vs. θ for **poins** from the BoxGenerator.

(a) K$^{\pm}$.

(b) $\pi^{\pm}$.
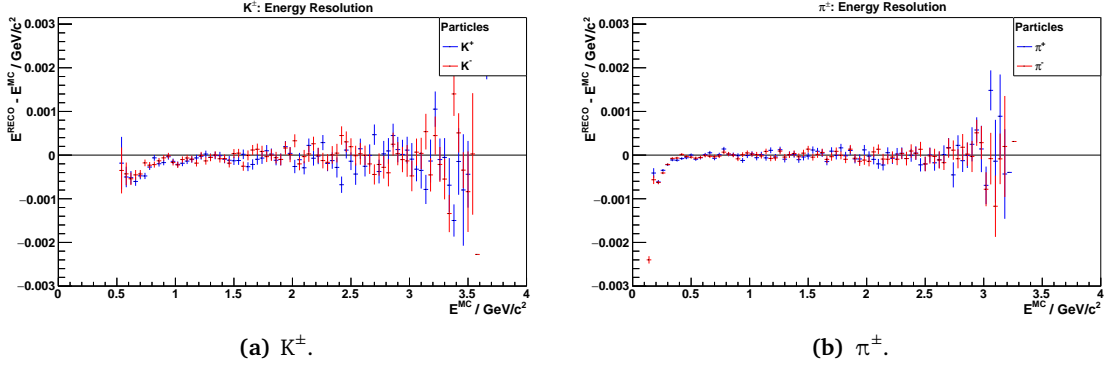
Figure A.14: Event Reconstruction (Inclusive): Systematic shift of the reconstructed energy for K and π as a function of E^{MC} before assembly into D candidates. The energy difference is defined as $\Delta E = E^{\mathrm{RECO}} - E^{\mathrm{MC}}$. Shown are only K and π which are MC-truth-matched.

(a) K$^{\pm}$.

(b) $\pi^{\pm}$.
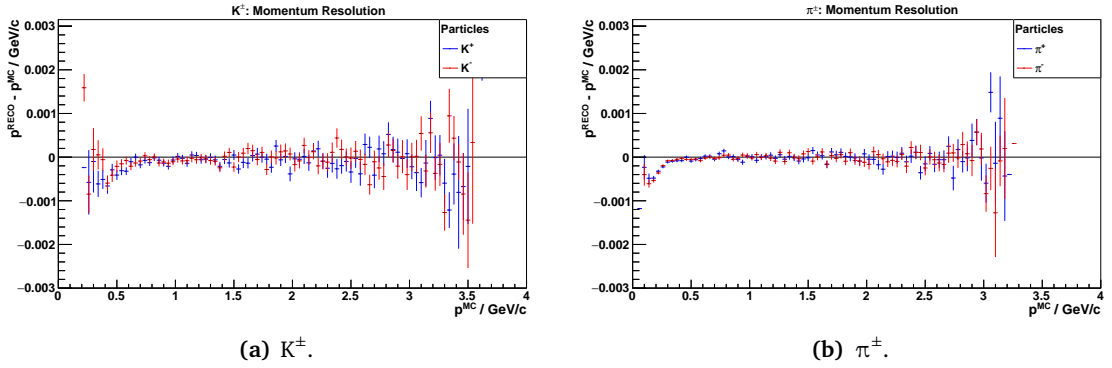
Figure A.15: Event Reconstruction (Inclusive): Systematic shift of the reconstructed momentum, $\Delta p = p^{\mathrm{RECO}} - p^{\mathrm{MC}}$, as a function of the MC momentum for K and π before assembly into D candidates. Shown are only MC-true-matched particles.

(a) x (all).

(b) x, best candidate.

(c) y (all).

(d) y, best candidate.

(e) z (all).

(f) z, best candidate.
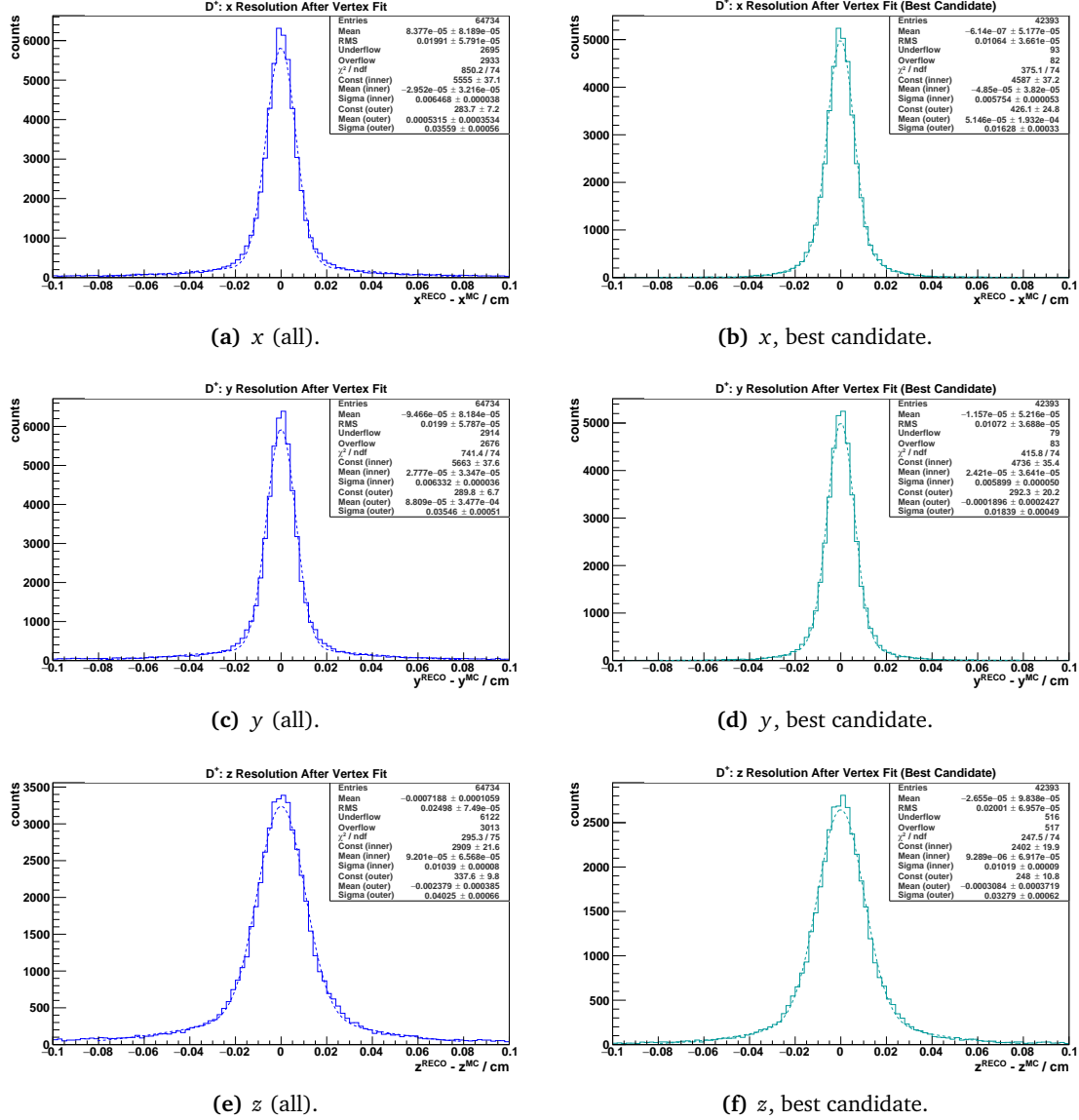
Figure A.16: Event Reconstruction (Inclusive): Relative decay vertex position distributions for all three coordinates of D^+. The left column shows distributions after the vertex fit for all D^+ candidates (with $\mathrm{Prob}(\chi^2) > 0.01$), the right column shows only the vertices of the *best* candidates. Double-Gaussian fits are applied to the distributions.

(a) After vertex fit.

(b) Best candidate after vertex fit and passing cut_vtx.
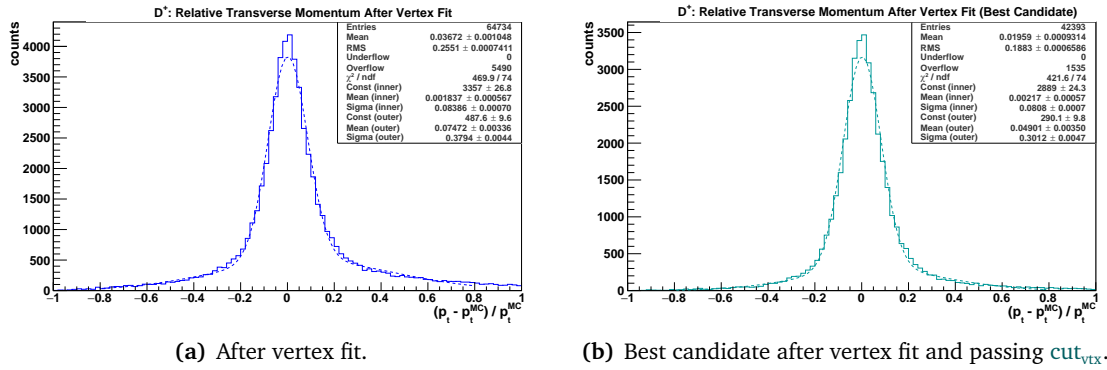
Figure A.17: Event Reconstruction (Inclusive): Transverse momentum resolution after vertex fit for D^+. Double-Gauss fits are superimposed.
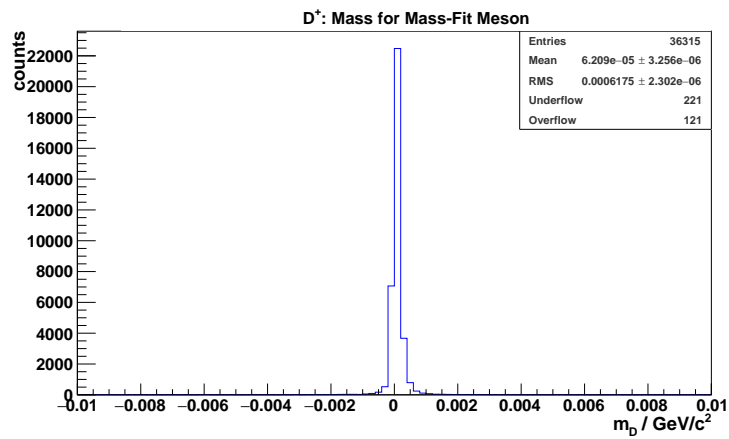
Figure A.18: Event Reconstruction (Inclusive): Distribution of the difference between the D^+ meson candidate invariant mass after the mass constraint fit minus the PDG mass.

(a) Δm_{D^+} distribution without vertex or kinematic fits.

(b) Δm_{D^+} distribution after vertex fit and cut$_{vtx}$.

(c) Δm_{D^+} distribution after vertex and mass constraint fit and cut$_{vtx}$ and cut$_{mass}$.

(d) Δm_{D^-} distribution after vertex and mass constraint fit and cut$_{vtx}$ and cut$_{mass}$.
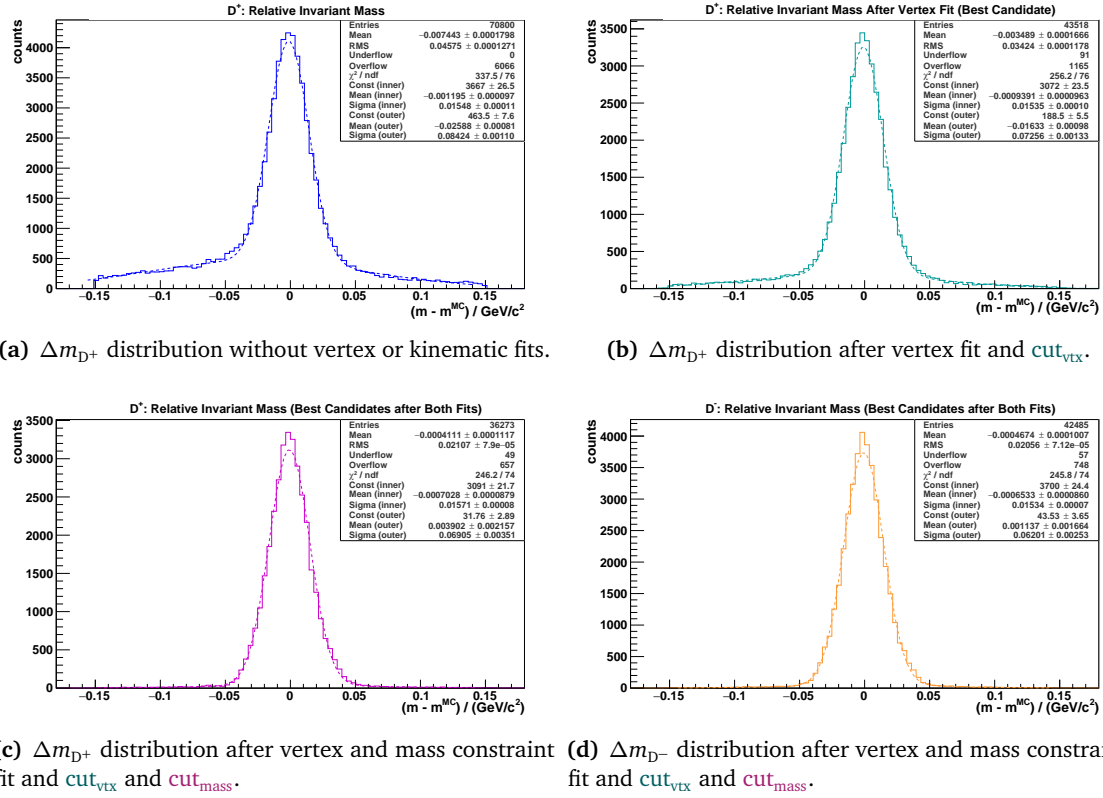
Figure A.19: Event Reconstruction (Inclusive): Invariant mass distribution for various stages of the reconstruction for D^+ candidates. After assembling pions and kaons to a D candidate **(a)**. After the vertex fit, taking only the best candidate passing cut$_{vtx}$ in **(b)**. After vertex and mass constraint fit, taking only the best candidate passing the cut$_{vtx}$ and also cut$_{mass}$ in **(c)**. In **(d)**, the same quantity is shown for D^-.

(a) ΔE_{D^+} distribution without fits.

(b) ΔE_{D^-} distribution without fits.

(c) ΔE_{D^+} distribution after vertex fit and cut$_{\text{vtx}}$.

(d) ΔE_{D^-} distribution after vertex fit and cut$_{\text{vtx}}$.

(e) ΔE_{D^+} distribution after vertex fit and cut$_{\text{vtx}}$ and cut$_{\text{mass}}$.

(f) ΔE_{D^-} distribution after vertex fit and cut$_{\text{vtx}}$ and cut$_{\text{mass}}$.
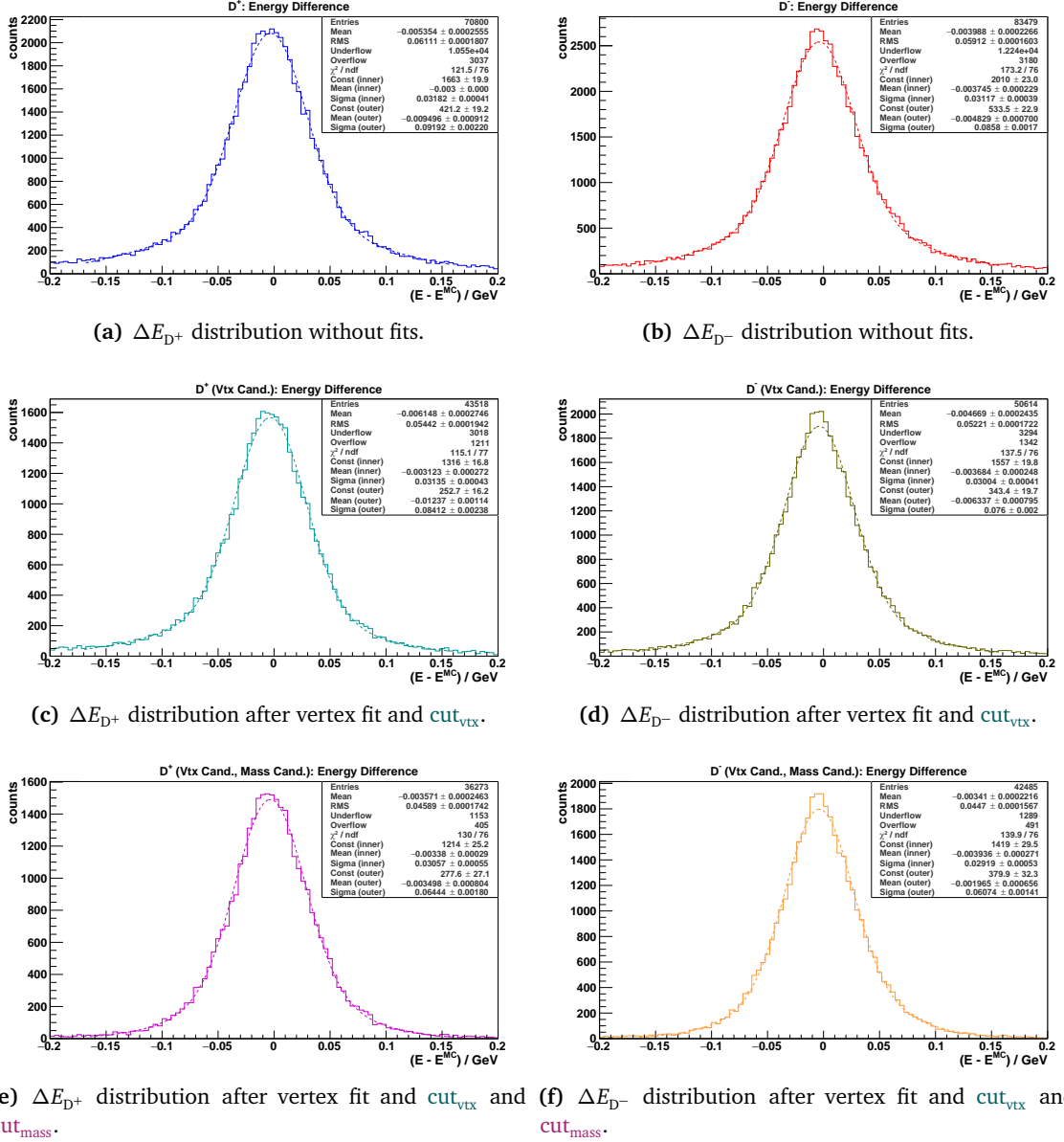
Figure A.20: Event Reconstruction (Inclusive): Progression of the energy resolution for the vertex and mass constraint fit. Left: D^+; right: D^-. Double-Gaussian fits are superimposed. The energy difference is defined as $\Delta E = E^{\text{RECO}} - E^{\text{MC}}$.
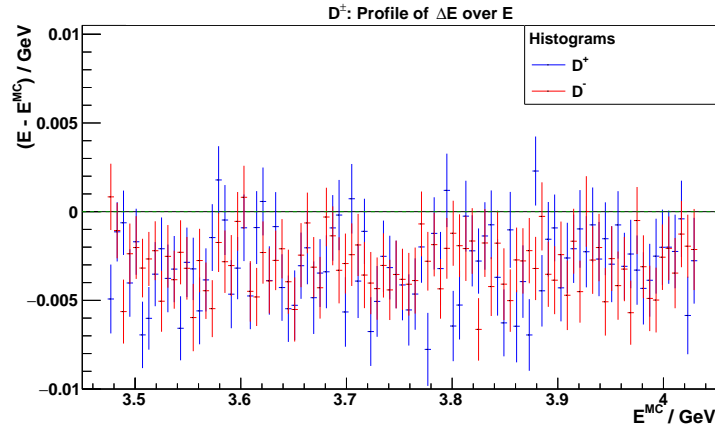
Figure A.21: Event Reconstruction (Inclusive): Mean offset of the energy as a function of the MC true energy for $D^\pm$.

(a) Momentum p.
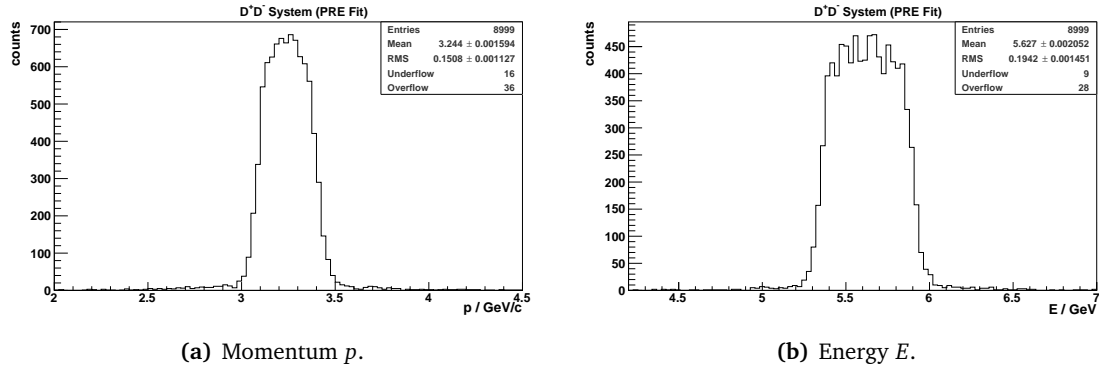
(b) Energy E.

Figure A.22: Event Reconstruction (Exclusive): Momentum and energy distribution after assembly to a combined system. Both D mesons passed the vertex and mass fit cuts before.

(a) Momentum p.
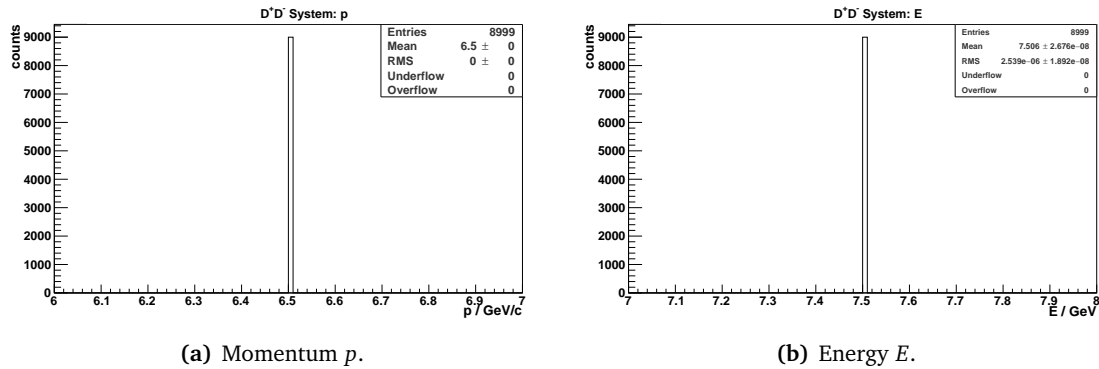
(b) Energy E.

Figure A.23: Event Reconstruction (Exclusive): Total momentum and energy after the four-constraint fit and $\mathrm{cut_{4C}}$.

(a) Transverse momentum p_t.

(b) Longitudinal momentum p_z.
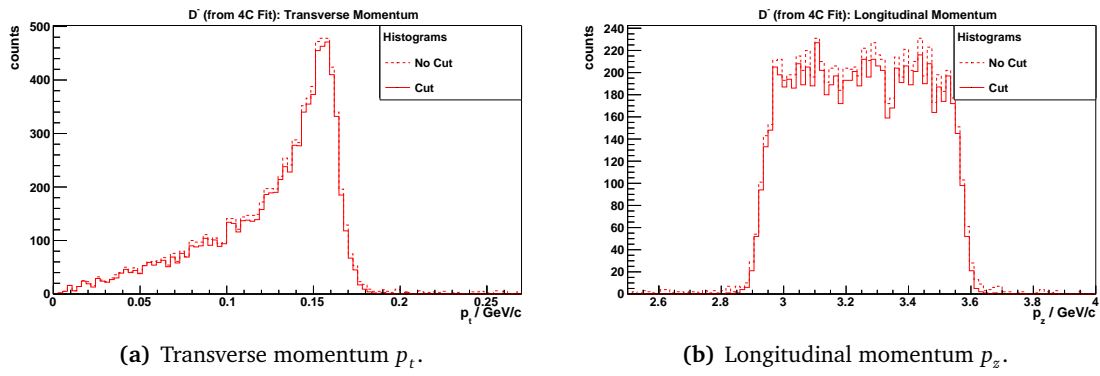
Figure A.24: Event Reconstruction (Exclusive): Momentum distributions for D$^-$ candidates after the four-constraint fit. D meson candidates before the cut$_{4C}$ are shown by the dashed lines.

Appendix B

Additional Description

B.1 GPU-based Tracking Algorithms

B.1.1 Triplet Finder

This section extends the optimizations given in Section 4.5.2.2.

B.1.1.1 Bunch Processing Schemes

Bunching, the scheme for grouping hits into sets of size, which occupy the GPU device best, is explained in Section 4.5.2.2. Different techniques for initiating computation on the individual bunches are tested, sketched in Figure B.1.

(a) Dynamic Parallelism. **(b)** Joined Kernel **(c)** Host Streams.
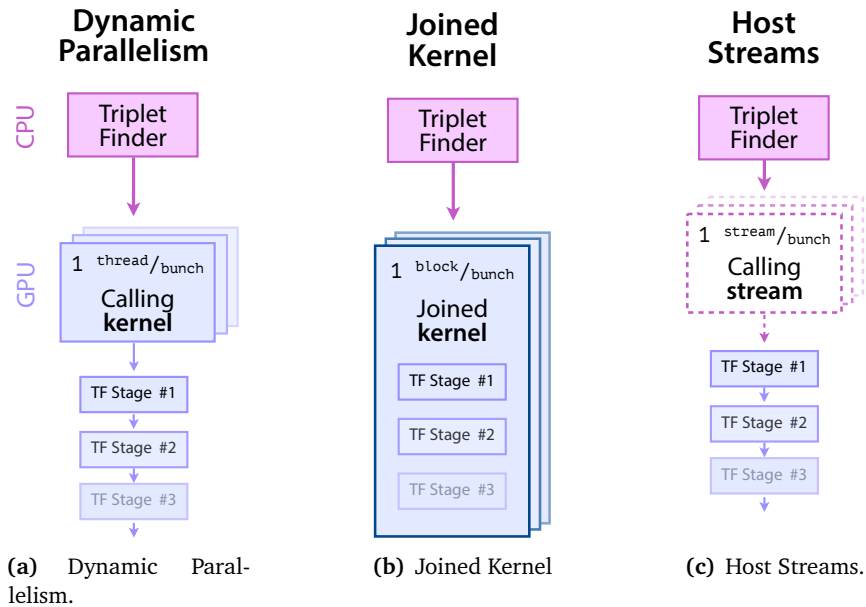
Figure B.1: Different schemes for calling the kernels. The magenta-colored boxes indicate processes on the CPU, blue-colored boxes processes on the GPU.
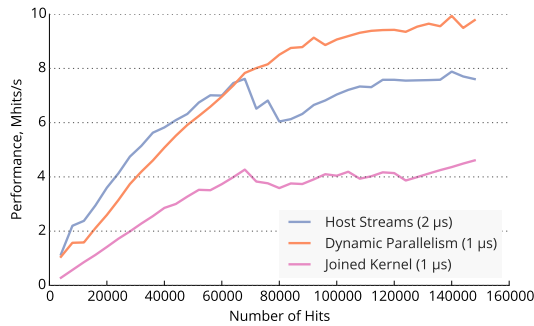
Figure B.2: Determination of ideal bunch sizes for the different bunching schemes.
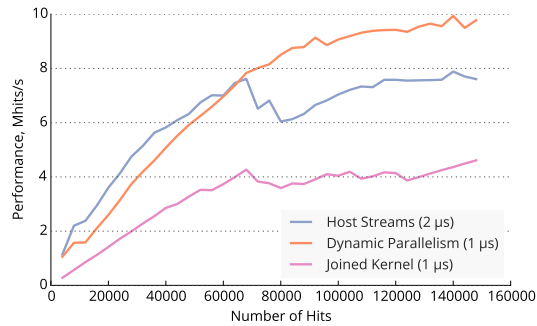
Figure B.3: Performance measurements of the different bunching schemes.

Dynamic Parallelism The initial approach. The approach initializes a one-threaded kernel per bunch. The kernel uses CUDA's possibility for calling GPU-side kernels from inside other kernels and starts the individual stages of the TF directly form the GPU. Shown in Figure B.1(a).

Joined Kernel For this technique, the individual kernels of the different stages of the TF are fused into one single kernel. Per bunch, the joined kernel is invoked with one CUDA block. See Figure B.1(b).

Host Streams Like in the first approach, the TF has an individual kernel for each stage. The stages are not called from a common GPU-side kernel, but from a host-side CUDA stream. One stream per bunch is used. See Figure B.1(c).

Bunch Size Determination   To analyze the individual performances of the different approaches, it is necessary to find the optimal bunch size for each. The ideal sizes are the high points in the plots of performance as a function of the number of processed hits. The distributions are shown in Figure B.2 as functions of the bunch size.[1] The curve for the dynamic parallelism and the joined kernel both have their respective maximum at approximately $1\,\mu s$ to $1.2\,\mu s$. The host streams approach reaches a plateau at $2\,\mu s$, before its performance starts dropping after around $3\,\mu s$.

Performances of Approaches   Using $2\,\mu s$ for the bunch size in the host streams approach and $1\,\mu s$ for both the joined kernel and the dynamic parallelism approach yields the performance graph of Figure B.3. The slowest approach over the whole range of bunch sizes is the joined kernel approach. Since the individual methods of the TF are fused into one common, large kernel, it also consumes a large amount of registers on the GPU. The occupancy is not optimal.

The performances of the host streams approach and the dynamic parallelism approach are similar in a large range of processed hits. At approximately 65,000 processed hits, the performance of the host streams approach drops and subsequently recuperates again slowly, while

---

[1]The bunch size is given as a time. The time stamps of hits are checked to lie in the current bunch size window, as illustrated in Figure 4.43.

(a) Sector-Row Testing.   **(b)** Skewlet Binning.
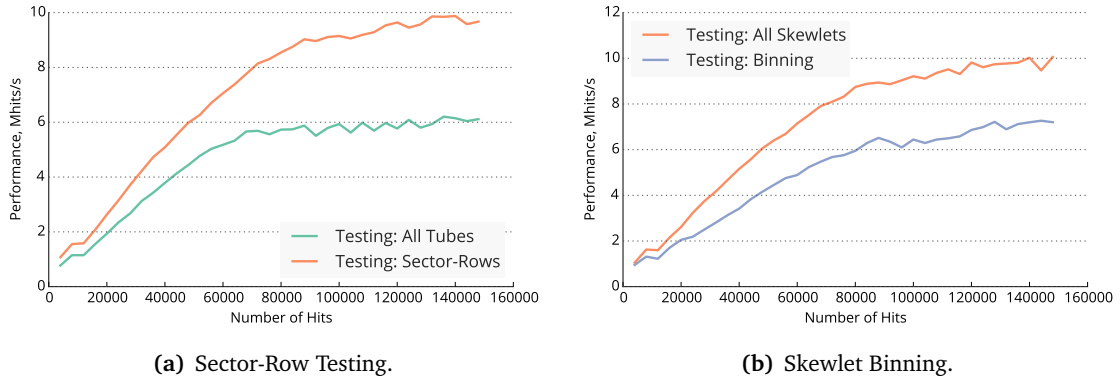
Figure B.4: Two optimization attempts for optimization of hit association in the Triplet Finder.

the dynamic parallelism approach continues gaining performance beyond 65,000, although the rate is flattening out. The reason for the performance difference lies in the location of the steering processes of the two approaches. For the dynamic parallelism approach, all kernel calls after the initial kernel call happen directly from the GPU, while the host streams approach invokes kernels from the CPU side. The latency of kernel launches and memory copy are limiting factors. The host streams approach transfers the kernel configurations via the PCI-E bus of the machine, adding to latency of the kernel call. GPU kernels are launched directly on the device in the dynamic parallelism approach, leading to a higher launch throughput. Another reason for the performance difference is the avoidance of false dependencies for the dynamic parallelism approach. False dependencies arise for the host streams approach, as one single host-side process handles all CUDA streams. In addition, further false dependencies are created when more device connections are requested by the host streams approach as available [146].

B.1.1.2 Hit Association Optimizations

In the scheme of the algorithm, the last step is associating hits in a compatible time frame to the track candidate. The association has been optimized for the GPU version, although the techniques employed are applicable to a CPU version as well.

Axial Straws: Sector-Row Testing   In the original CPU implementation, hit association is done by testing all hits in the STT against the Triplet-based track candidate. Most of the hits belong to other tracks and are tested unnecessarily. Geometrical information can be used to limit the scope of the tested hits and with it, the number of computations.

In a process dubbed *sector-row testing*, track candidates are evaluated for intersections with *sector-rows*, a layer of axial straws belonging to one of the six STT sectors. The intersection of the curve of the track candidate with the line of a sector row directly gives the location to search for a hit to associate to a track candidate. To incorporate the straw diameter and additional safety margins, the curved line of a track candidate is artificially thickened to a ring.

Figure B.4(a) compares the performance of the TF with and with out sector-row testing. The sub-routine of the TF responsible for hit association is a factor of 5 faster when using the sector-row approach. A speed-up of approximately 60 % is obtained for the whole TF algorithm.

Skewed Straws: Skewlet Binning   The central eight layers of straws in the STT are tilted with respect to the beam by ±2.9°. These skewed straws are as well included in the algorithm, but treated differently due to their inclination. First, *half-skewlets* are formed as combinations of two or three hits in each tilted double layer. A half-skewlet lies on a straight line in the center of a double layer. A *skewlet* is computed when combining two half-skewlets: The point exactly in the middle at the closest approach of the half-skewlet is taken as the position of the skewlet, see Figure B.5. The skewlet is then checked during hit association stage.

Figure B.5: Sketch for determination of *half-skewlet* (red tubes) and *skewlets* (black point with red outline). Hit straws (red circles) are combined to half-skewlet. The point of closest approach is found (black line), in its center the skewlet is located. For *skewlet binning*, the line between two double layers is divided into bins (green cells).

After optimization of hit association in axial straws, testing of skewlets of the skewed layers is the method with highest time consumption of the TF. A similar approach as chosen for sector-row testing is implemented, limiting the number of skewlets a track is tested against. For skewed straws, the line to be intersected with the track candidate does not lie in the middle of a straw layer. The line, all skewlets are lying on in the xy-projection, is located between two double layers of skewed straws. The eight double layers, hence, create three *skewlet layers* for each sector of the STT. As for the depth information obtained by the tilt of the straws, the positions on this *skewlet line* are not as defined as they are for the axial straws. Varying skewed straw combinations result in different lateral locations on the skewlet lines. To give a rough estimate about the location on the skewlet line, the skewlet line is divided into sub-parts and each skewlet associated to one part (*bin*). The track, as before, is then tested for overlaps with the bins of the skewlet line.

The results of the performance optimization attempt are shown in Figure B.4(b). Against expectations, the approach of skewlet binning is slower than testing all skewlets. A reduction of 30 % in the total TF performance can be seen. The reason for this behavior is the number of skewlets to test per bin and the number of bins to test per sector-row. Both are comparably high and lead to many computations. Skewlet binning is not used in any of the other performance tests and optimizations.

B.1.1.3   CUDA Versions

The TF is programmed in CUDA, NVIDIA's GPU infrastructure. CUDA not only covers the programming model on NVIDIA GPUs but also optimizes code for the cards. With newer versions, more optimizations are introduced and features of the GPUs exploited.

Comparing the TF running under the three CUDA versions, without any other optimizations or modifications, yields the performance graph of Figure B.6. The performance for CUDA 6.0 and 6.5 are comparable, while the performance of CUDA 5.5 is slightly worse.

B.1.1.4   GPU Chipset and Core Clock Rate

All previous performance measurements are obtained on a Tesla K20X GPU. The TF is also

Figure B.6: Performance for different CUDA versions.

Figure B.7: Performance of the Triplet Finder on different GPUs, the Tesla K20X (*K20X*) and Tesla K40 (*K40*). Two distributions per card are show: The lower value is the core GPU clock rate without modification, the higher value is the value as extended by GPUBoost.

tested on a slightly newer card, a Tesla K40 with 2880 streaming processors and 12 GB RAM. In addition to measuring performances on the different cards with their respective properties, also the core clock rate is opened up for manipulation, both for the Tesla K20X and the Tesla K40. An NVIDIA feature called *GPUBoost* enables modifying the core clock frequency, to allow GPU processes to run with higher speed if the thermal budget of the cards allows for it [150]. The results are shown in Figure B.7.

While the transition from a Tesla K20X to a Tesla K40 improves the best performance by 10 %, the application of GPUBoost, especially on the Tesla K40, increases the best TF performance from 9.8 Mhit/s to 12.5 Mhit/s – a 27.5 % increase. The algorithm appears to profit of clock rate elevations.

The whole set of 150,000 hit points is processed in 15.8 ms on the Tesla K20X, or in 12.0 ms for the *overclocked* Tesla K40.

Bibliography

[1]   K.A. OLIVE *et al.*: **Review of Particle Physics**. In: *Chin. Phys. C*, **38**:9 (090001) (2014). DOI: `10.1088/1674-1137/38/9/090001` (see pp. 4, 7–9, 14, 31–32, 141, 145, 179).

[2]   PETER W. HIGGS: **Broken Symmetries and the Masses of Gauge Bosons**. In: *Phys. Rev. Lett.*, **13**: (16 Oct. 1964), pp. 508–509. DOI: `10.1103/PhysRevLett.13.508` (see p. 5).

[3]   T. W. B. KIBBLE: **Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism**. In: *Scholarpedia*, **4**:1 (2009). revision 91222, p. 6441. DOI: `doi:10.4249/scholarpedia.6441` (see p. 5).

[4]   T. W. B. KIBBLE: **Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism (history)**. In: *Scholarpedia*, **4**:1 (2009). revision 137393, p. 8741. DOI: `doi:10.4249/scholarpedia.8741` (see p. 5).

[5]   CMS COLLABORATION: **Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC**. In: *Physics Letters B*, **716**:1 (2012), pp. 30–61. URL: `http://arxiv.org/abs/1207.7235` (see p. 5).

[6]   ATLAS COLLABORATION: **Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC**. In: *Physics Letters B*, **716**:1 (2012), pp. 1–29. DOI: `10.1016/j.physletb.2012.08.020` (see p. 5).

[7]   **The Nobel Prize in Physics 1969**. URL: `http://www.nobelprize.org/nobel_prizes/physics/laureates/1969/` (see p. 6).

[8]   **The Nobel Prize in Physics 2004**. URL: `http://www.nobelprize.org/nobel_prizes/physics/laureates/2004/` (see p. 6).

[9]   G.M. PROSPERI, M. RACITI, and C. SIMOLO: **On the running coupling constant in QCD**. In: *Progress in Particle and Nuclear Physics*, **58**:2 (2007), pp. 387–438. DOI: `10.1016/j.ppnp.2006.09.001` (see p. 8).

[10] NICUŞOR TIMNEANU: **The colour of gluon interactions: Studies of Quantum Chromodynamics in soft and hard processes**. PhD Thesis. Uppsala University, 2002 (see p. 8).

[11] R. KEITH ELLIS, W. JAMES STIRLING, and BRYAN R. WEBBER: **QCD and collider physics**. Vol. 8. Cambridge University Press, Dec. 2003 (see p. 9).

[12] **Picture by user E2m at Wikimedia Commons**. URL: `https://commons.wikimedia.org/wiki/File:Meson_nonet_-_spin_0.svg` (see p. 10).

[13] **Picture by user E2m at Wikimedia Commons**. URL: https://commons.wikimedia.org/wiki/File:Meson_nonet_-_spin_1.svg (see p. 10).

[14] **Picture by user E2m at Wikimedia Commons**. URL: https://commons.wikimedia.org/wiki/File:Baryon_octet.svg (see p. 10).

[15] **Picture by user Wierdw123 at Wikimedia Commons**. URL: https://commons.wikimedia.org/wiki/File:Baryon_decuplet.svg (see p. 10).

[16] J.-E. AUGUSTIN, A. M. BOYARSKI, M. BREIDENBACH, *et al.*: **Discovery of a Narrow Resonance in e^+e^- Annihilation**. In: *Phys. Rev. Lett.*, **33**: (23 Dec. 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406 (see p. 11).

[17] J. J. AUBERT, U. BECKER, P. J. BIGGS, *et al.*: **Experimental Observation of a Heavy Particle J**. In: *Phys. Rev. Lett.*, **33**: (23 Dec. 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404 (see p. 11).

[18] S. W. HERB, D. C. HOM, L. M. LEDERMAN, *et al.*: **Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions**. In: *Phys. Rev. Lett.*, **39**: (5 Aug. 1977), pp. 252–255. DOI: 10.1103/PhysRevLett.39.252 (see p. 11).

[19] G. GOLDHABER, F. M. PIERRE, G. S. ABRAMS, *et al.*: **Observation in e^+e^- Annihilation of a Narrow State at 1865 MeV/c^2 Decaying to Kπ and Kπππ**. In: *Phys. Rev. Lett.*, **37**: (5 Aug. 1976), pp. 255–259. DOI: 10.1103/PhysRevLett.37.255 (see p. 11).

[20] I. PERUZZI, M. PICCOLO, G. J. FELDMAN, *et al.*: **Observation of a Narrow Charged State at 1876 MeV/c^2 Decaying to an Exotic Combination of Kππ**. In: *Phys. Rev. Lett.*, **37**: (10 Sept. 1976), pp. 569–571. DOI: 10.1103/PhysRevLett.37.569 (see p. 11).

[21] P. DEL AMO SANCHEZ *et al.*: **Observation of new resonances decaying to $D\pi$ and $D^*\pi$ in inclusive e^+e^- collisions near $\sqrt{s}$ =10.58 GeV**. In: *Phys. Rev.*, **D82**: (2010), p. 111101. DOI: 10.1103/PhysRevD.82.111101. arXiv: 1009.2076 [hep-ex] (see p. 11).

[22] STEPHEN GODFREY and NATHAN ISGUR: **Mesons in a relativized quark model with chromodynamics**. In: *Phys. Rev. D*, **32**: (1 July 1985), pp. 189–231. DOI: 10.1103/PhysRevD.32.189 (see pp. 11–12).

[23] M. DI PIERRO and E. EICHTEN: **Excited heavy-light systems and hadronic transitions**. In: *Phys. Rev. D*, **64**: (11 Oct. 2001), p. 114004. DOI: 10.1103/PhysRevD.64.114004 (see p. 12).

[24] PAOLA GIANOTTI: **Results and perspectives in hadron spectroscopy**. In: *Physica Scripta*, **2012**:T150 (2012), p. 014014. DOI: 10.1088/0031-8949/2012/T150/014014 (see pp. 12–13).

[25] QIN-TAO SONG, DIAN-YONG CHEN, XIANG LIU, and TAKAYUKI MATSUKI: **Charmed-strange mesons revisited: mass spectra and strong decays**. In: *arXiv Preprint*, (Jan. 2015). arXiv: 1501.03575 [hep-ph] (see pp. 11–12).

[26] STEPHEN GODFREY and STEPHEN L. OLSEN: **The Exotic XYZ Charmonium-Like Mesons**. In: *Annual Review of Nuclear and Particle Science*, **58**: (Nov. 2008), pp. 51–73. DOI: 10.1146/annurev.nucl.58.110707.171145. arXiv: 0801.3867 [hep-ph] (see p. 12).

[27] STEPHEN GODFREY: **Topics in Hadron Spectroscopy in 2009**. In: *2009 Meeting of the Division of Particles and Fields of the American Physical Society*. Oct. 2009. arXiv: 0801.3867 [hep-ph] (see p. 14).

[28] GSI Helmholtzzentrum für Schwerionenforschung: **FAIR Accelerator Chain**. Private communication (see p. 17).

[29] BIOMAT Project Group: **Technical Proposal for the Design, Construction, Commissioning and Operation of the High-Energy Irradiation Facility for Biophysics and Materials Research**. Tech. Rep. *FAIR-PAC*, 2005 (see p. 15).

[30] M. Durante: **The BIOMAT facility at FAIR: a new tool for ground-based research in space radiation biophysics**. In: *37th COSPAR Scientific Assembly*. Vol. 37. COSPAR Meeting. 2008, p. 770 (see p. 17).

[31] E. Widmann: **FAIR - Facility for Antiproton and Ion Research**. *Talk* at Stefan Meyer Institute for Subatomic Physics, Wien. Nov. 2007. URL: http://www4.ffg.at/veranstaltungen/Downloads/7B22A1D1.pdf (see p. 18).

[32] The FLAIR community: **FLAIR: A Facility for Low-energy Antiproton and Ion Research Letter of Intent for the Future Accelerator Facility for Beams of Ions and Antiprotons at Darmstadt**. Tech. Rep. *APPA-PAC*, 2005 (see p. 17).

[33] The FLAIR Collaboration: **FLAIR: A Facility for Low-energy Antiproton and Ion Research**. Tech. Rep. *FAIR Baseline Technical Report*, 2006 (see p. 18).

[34] HEDgeHOB Collaboration: **Technical Proposal for Design, Construction, Commissioning and Operation of the: HEDgeHOB (High Energy Density Matter Generated by Heavy Ion Beams) - Studies on High Energy Density Matter with Intense Heavy Ion and Laser Beams at FAIR**. Tech. Rep. *FAIR-PAC*, 2005. URL: http://hedgehob.physik.tu-darmstadt.de/Docs/HEDgeHOB-TP2005.pdf (see p. 18).

[35] HEDgeHOB Collaboration: **Baseline Technical Report for HEDgeHOB: Studies on High Energy Density Matter with Intense Heavy Ion and Laser Beams at FAIR**. Tech. Rep. *FAIR Baseline Technical Report*, 2006 (see p. 18).

[36] The SPARC Collaboration: **Technical Report for the Design, Construction, Commissioning and Operation of the *SPARC* Project: Stored Particle Atomic Physics Collaboration at the FAIR Facility**. Tech. Rep. *FAIR-PAC*, 2005. URL: https://www.gsi.de/fileadmin/SPARC/documents/06_01_06_SPARC_TR.pdf (see p. 18).

[37] Thomas Stöhlker, Yuri A. Litvinov, Angela Bräuning-Demian, *et al.*: **SPARC collaboration: new strategy for storage ring physics at FAIR**. In: *Hyperfine Interactions*, **227**:1-3 (2014), pp. 45–53. DOI: 10.1007/s10751-014-1047-2 (see pp. 18, 27).

[38] Thomas Stöhlker, Reinhold Schuch, Siegbert Hagmann, and Yuri A. Litvinov: **SPARC Experiments at the HESR: A Feasibility Study**. Tech. Rep. Jan. 2012. URL: http://lxcms1.gsi.de/typo3/start/fair/fair_experimente_und_kollaborationen/sparc/dokumente.htm (see p. 18).

[39] FAIR: **FAIR Baseline Technical Report**. Tech. Rep. 2006. URL: http://www.fair-center.eu/for-users/publications/fair-publications.html (see pp. 18–19, 22–23).

[40] H. Geissel, H. Weick, M. Winkler, *et al.*: **The Super-FRS project at GSI**. In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, **204**: (2003). 14th International Conference on Electromagnetic Isotope Separators and Techniques Related to their Applications, pp. 71–85. DOI: dx.doi.org/10.1016/S0168-583X(02)01893-1 (see p. 19).

[41] NuSTAR COLLABORATION SUPER-FRS WORKING GROUP: **Technical Design Report on the Super-FRS**. Tech. Rep. *FAIR-PAC*, Mar. 2009. URL: https://edms.cern.ch/document/994387/1 (see p. 19).

[42] HISPEC COLLABORATION and DESPEC COLLABORATION: **Technical Proposal for the Design, Construction, Commissioning and Operation of the HISPEC/DESPEC experiment at the Low-Energy Branch of the Super-FRS facility**. Tech. Rep. *FAIR-PAC*, 2005 (see p. 20).

[43] A.N. ANTONOV, M.K. GAIDAROV, M.V. IVANOV, *et al.*: **The electron–ion scattering experiment ELISe at the International Facility for Antiproton and Ion Research (FAIR) — A conceptual design study**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **637**:1 (2011), pp. 60–76. DOI: 10.1016/j.nima.2010.12.246 (see pp. 20, 25).

[44] THE EXL COLLABORATION: **Update of the Technical Proposal for the Design, Construction, Commissioning and Operation of the EXL Project (Exotic nuclei studied in light-ion induced reactions at the NESR storage ring)**. Tech. Rep. *Technical Proposal*, 2006. URL: http://www.rug.nl/kvi-cart/research/hnp/research/exl/tech_proposal_exl_update_080106.pdf (see p. 20).

[45] THE EXL COLLABORATION: **Exotic nuclei studied in Light-ion induced reactions at the NESR storage ring (EXL)**. Tech. Rep. *Executive Summary*, 2006 (see p. 20).

[46] ILIMA COLLABORATION: **Technical Proposal for the ILIMA Project**. Tech. Rep. *FAIR-PAC*, 2005 (see p. 20).

[47] D. RODRIGUEZ, K. BLAUM, W. NÖRTERSHÄUSER, M. AHAMMED, and A. ALGORA: **MATS and LaSpec: High-precision experiments using ion traps and lasers at FAIR**. In: *The European Physical Journal Special Topics*, **183**:1 (2010), pp. 1–123. DOI: 10.1140/epjst/e2010-01231-2 (see p. 20).

[48] THE R3B COLLABORATION: **R3B A universal setup for kinematical complete measurements of Reactions with Relativistic Radioactive Beams**. Tech. Rep. *FAIR Technical Report*, 2005 (see p. 20).

[49] VOLKER FRIESE and CHRISTIAN STURM: **CBM Progress Report 2013**. Tech. Rep. Apr. 2013. URL: https://www-alt.gsi.de/documents/DOC-2014-Mar-16-1.pdf (see p. 21).

[50] JOHANN M. HEUSER: **The Compressed Baryonic Matter Experiment at FAIR: Progress with feasibility studies and detector developments**. In: *Nuclear Physics A*, **830**:1–4 (2009). Quark Matter 2009 The 21st International Conference on Ultrarelativistic Nucleus-Nucleus Collisions, pp. 563c–566c. DOI: 10.1016/j.nuclphysa.2009.09.048 (see p. 21).

[51] GERMANY GSI DARMSTADT: **FAIR CDR - An International Accelerator Facility for Beams of Ions and Antiprotons, Conceptual Design Report - Section 3, The Facility**. https://www-alt.gsi.de/documents/DOC-2004-Mar-201.html. Nov. 2001 (see p. 22).

[52] O KESTER, W BARTH, O DOLINSKYY, *et al.*: **Status of the FAIR accelerator facility**. In: *Proceedings of IPAC2014, Dresden, Germany*. 2014, pp. 2084–2087 (see p. 23).

[53] COLLECTOR RING WORKING GROUP: **Technical Design Report on the Collector Ring**. Tech. Rep. Nov. 2013 (see p. 24).

[54] I.A. Koop, Yu.M. Shatunov, V.G. Nedorezov, A.S. Fomichev, and L.V. Chulkov: **Electron-ion collider at the fair facility for antiproton and ion research (ELISe Project)**. In: *Atomic Energy*, **112**:2 (2012), pp. 110–118. DOI: `10.1007/s10512-012-9532-7` (see p. 25).

[55] K. Knie, A. Dolinskii, B. Franzke, *et al.*: **Concept for the Antiproton Production Target at FAIR**. In: *Proceedings of IPAC2012, New Orleans, Louisiana, USA*, **C1205201**: (2012), WEPPD030. 3 p (see p. 25).

[56] R. Tölle: **HESR Picture**. Private communication. Oct. 2014 (see p. 26).

[57] R. Toelle, K. Bongardt, J. Dietrich, *et al.*: **HESR at FAIR: Status of technical planning**. In: *Proceedings of PAC07, Albuquerque, New Mexico, USA*. 2007, pp. 1442–1444 (see p. 26).

[58] R. Maier *et al.*: **The High Energy Storage Ring (HESR)**. In: *Proceedings of PAC2011, New York, USA*. 2011, pp. 2104–2106 (see pp. 26–27).

[59] N. Alinovskiy, T. Bedareva, E. Bekhtenev, *et al.*: **2 MeV Electron Cooler For COSY and HESR – First Results**. In: *Proceedings of IPAC2014, Dresden, Germany*. 2014, pp. 765–767 (see p. 27).

[60] R. Stassen, R. Greven, R. Maier, G. Schug, and H. Stockhorst: **The Main Stochastic Cooling System of the HESR**. In: *Proceedings of IPAC2013, Shanghai, China*. 2013, pp. 100–102 (see p. 27).

[61] Dieter Prasuhn: **Status HESR**. PANDA XLIX. Collaboration Meeting (*Talk*). Darmstadt, June 2014 (see p. 27).

[62] H. Stockhorst, R. Maier, D. Prasuhn, R. Stassen, T. Katayama, *et al.*: **Feasibility Study of Heavy Ion Storage and Acceleration in the HESR with Stochastic Cooling and Internal Target**. In: *Proceedings of IPAC2013, Shanghai, China*. 2013, pp. 106–108 (see p. 27).

[63] O. Kovalenko, A. Dolinskii, T. Katayama, *et al.*: **Ion Optics of the HESR Storage Ring at FAIR for Operation with Heavy Ions**. In: *Proceedings of IPAC2014, Dresden, Germany*. 2014, pp. 1117–1119 (see p. 27).

[64] M. Steck, C. Dimopoulou, A. Dolinskii, *et al.*: **Operation of the HESR Storage Ring of the FAIR Project with Ions and Rare Isotopes**. In: *Proceedings of IPAC2012, New Orleans, Louisiana, USA*. 2012 (see p. 27).

[65] FAIR Joint Core Team: **Green Paper - FAIR Modularized Start Version**. Tech. Rep. Nov. 2009. URL: `http://www.fair-center.eu/fileadmin/fair/publications_FAIR/FAIR_GreenPaper_2009.pdf` (see pp. 28–29).

[66] H. Stoecker and C. Sturm: **The FAIR start**. In: *Nuclear Physics A*, **855**:1 (2011). Proceedings of the 4th International Conference on Hard and Electromagnetic Probes of High-Energy Nuclear Collisions – HP2010, pp. 506–509. DOI: `10.1016/j.nuclphysa.2011.02.117` (see p. 29).

[67] The PANDA Collaboration: **Technical Progress Report for: PANDA (AnitProton Annihilation at Darmstadt) Strong Interaction Studies with Antiprotons**. Tech. Rep. Feb. 2005. URL: `http://www-panda.gsi.de/archive/public/panda_tpr.pdf` (see pp. 29, 58–60).

[68]   PANDA COLLABORATION, W. ERNI, I. KESHELASHVILI, *et al.*: **Physics Performance Report for PANDA: Strong Interaction Studies with Antiprotons**. In: (Mar. 2009). arXiv: 0903.3905 [hep-ex]. URL: http://arxiv.org/abs/0903.3905 (see pp. 30, 32–34).

[69]   THE BABAR COLLABORATION: **Observation of a Narrow Meson State Decaying to** $D_s^+\pi^0$ **at a Mass of** 2.32 GeV/c^2. In: *Phys. Rev. Lett.*, **90**: (24 June 2003), p. 242001. DOI: 10.1103/PhysRevLett.90.242001. arXiv: 0304021 [hep-ex] (see p. 30).

[70]   MARIUS CHRISTIAN MERTENS and THE PANDA COLLABORATION: **Determination of the** $D_{s0}^*(2317)$ **width with the PANDA detector**. In: *Hyperfine Interactions*, **209**:1-3 (2012), pp. 111–115. DOI: 10.1007/s10751-012-0565-z (see p. 30).

[71]   A. ABELE, J. ADOMEIT, C. AMSLER, *et al.*: **Exotic** $\eta\pi$ **state in** $\bar{p}d$ **annihilation at rest into** $\pi^-\pi^0\eta p_{\mathbf{spectator}}$. In: *Physics Letters B*, **423**:1–2 (1998), pp. 175–184. DOI: 10.1016/S0370-2693(98)00123-3 (see p. 31).

[72]   J. REINNARTH and CRYSTAL BARREL COLLABORATION: **Evidence for an exotic partial wave in** $\pi\eta'$. In: *Nuclear Physics A*, **692**:1–2 (2001). Sixth Biennial Conference on Low-Energy Antiproton Physics, pp. 268–274. DOI: 10.1016/S0375-9474(01)01186-1 (see p. 31).

[73]   EBERHARD KLEMPT and ALEXANDER ZAITSEV: **Glueballs, hybrids, multiquarks: Experimental facts versus QCD inspired concepts**. In: *Physics Reports*, **454**:1–4 (2007), pp. 1–202. DOI: 10.1016/j.physrep.2007.07.006 (see pp. 31–32).

[74]   Y. CHEN, A. ALEXANDRU, S. J. DONG, *et al.*: **Glueball spectrum and matrix elements on anisotropic lattices**. In: *Phys. Rev. D*, **73**: (1 Jan. 2006), p. 014516. DOI: 10.1103/PhysRevD.73.014516 (see p. 32).

[75]   S. YASUI and K. SUDOH: $\overline{\mathbf{D}}$ **and B mesons in a nuclear medium**. In: *Phys. Rev. C*, **87**: (1 Jan. 2013), p. 015202. DOI: 10.1103/PhysRevC.87.015202 (see p. 33).

[76]   A. B. KAIDALOV and P. E. VOLKOVITSKY: **Binary reactions in** $\bar{p}p$ **collisions at intermediate energies**. In: *Zeitschrift für Physik C Particles and Fields*, **63**:3 (1994), pp. 517–524. DOI: 10.1007/BF01580332 (see p. 34).

[77]   PANDA COLLABORATION: **Technical Design Report for the PANDA Internal Targets: The Cluster-Jet Target and Developments for the Pellet Target**. Tech. Rep. Mar. 2012. URL: http://www-panda.gsi.de/archive/TargetTDR/Targets_TDR.pdf (see pp. 37–38).

[78]   ÖRJAN NORDHAGE: **On a Hydrogen Pellet Target for Antiproton Physics with PANDA**. PhD Thesis. Uppsala University, Teknisk-naturvetenskapliga vetenskapsområdet, Physics, Department of Nuclear and Particle Physics, Sept. 2006, p. 105. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-7137 (see pp. 38–39).

[79]   C. EKSTRÖM, C.-J. FRIDÉN, A. JANSSON, *et al.*: **Hydrogen pellet targets for circulating particle beams**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **371**:3 (1996), pp. 572–574. DOI: 10.1016/0168-9002(96)00009-5 (see p. 39).

[80]   THE PANDA COLLABORATION: **Technical Design Report for the: PANDA Micro Vertex Detector**. Tech. Rep. Aug. 2012. arXiv: 1207.6581 [physics.ins-det] (see pp. 39–42).

[81] A. GOERRES, R. BUGALHO, A. DI FRANCESCO, *et al.*: **A free-running, time-based readout method for particle detectors**. In: *Journal of Instrumentation*, **9**:03 (Mar. 2014), p. C03025. DOI: 10.1088/1748-0221/9/03/C03025 (see p. 42).

[82] PANDA COLLABORATION, W. ERNI, I. KESHELASHVILI, *et al.*: **Technical design report for the PANDA (AntiProton Annihilations at Darmstadt) Straw Tube Tracker**. In: *The European Physical Journal A*, **49**:2 (2013), pp. 1–104. DOI: 10.1140/epja/i2013-13025-8 (see pp. 42–44).

[83] **DESY FLC-TPC**. URL: http://www-flc.desy.de/tpc/basicsgem.php (visited on 11/22/2014). Saved at: http://web.archive.org/web/20141122124357/http://www-flc.desy.de/tpc/basicsgem.php (see p. 44).

[84] **CERN Gas Detectors Development**. URL: http://gdd.web.cern.ch/GDD/ (see p. 44).

[85] MARTIN J. GALUSKA, JIFENG HU, WOLFGANG KÜHN, *et al.*: **Hough Transform Based Pattern Recognition for the PANDA Forward Tracking System**. In: *Proceedings of Science (Bormio 2013)*. 2013 (see pp. 46, 70).

[86] THE PANDA COLLABORATION, W. ERNI, I. KESHELASHVILI, *et al.*: **Technical Design Report for the PANDA Solenoid and Dipole Spectrometer Magnets**. Tech. Rep. 2009. eprint: arXiv:0907.0169. URL: http://arxiv.org/abs/0907.0169 (see pp. 46–47).

[87] THE PANDA COLLABORATION: **Proposal for a Scintillator Tile Hodoscope for PANDA**. Internal Report. 2011 (see p. 48).

[88] **PANDA Website: Time Of Flight System**. URL: http://www-panda.gsi.de/framework/det_iframe.php?section=Particle%20ID (visited on 12/17/2014) (see p. 49).

[89] S. BELOSTOTSKI: **Status of FTOF detectors**. PANDA Collaboration Meeting (*Talk*). Mar. 2014 (see p. 49).

[90] G. KALICY, H. KUMAWAT, J. SCHWIENING, *et al.*: **Status of the PANDA Barrel DIRC**. In: *Journal of Instrumentation*, **9**:05 (2014), p. C05060. URL: http://stacks.iop.org/1748-0221/9/i=05/a=C05060 (see p. 50).

[91] O. MERLE, R. DZHYGADLO, A. GERHARDT, *et al.*: **Development of an Endcap DIRC for PANDA**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **766**: (2014). RICH2013 Proceedings of the Eighth International Workshop on Ring Imaging Cherenkov Detectors Shonan, Kanagawa, Japan, December 2-6, 2013, pp. 96–100. DOI: 10.1016/j.nima.2014.04.016 (see p. 50).

[92] C. SCHWARZ, G. AHMED, A. BRITTING, *et al.*: **Particle identification for the PANDA detector**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **639**:1 (2011). Proceedings of the Seventh International Workshop on Ring Imaging Cherenkov Detectors, pp. 169–172. DOI: 10.1016/j.nima.2010.10.116 (see p. 49).

[93] J. SCHWIENING: **The barrel DIRC detector for the PANDA experiment at FAIR**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **639**:1 (2011). Proceedings of the Seventh International Workshop on Ring Imaging Cherenkov Detectors, pp. 315–318. DOI: 10.1016/j.nima.2010.10.061 (see p. 50).

[94] E.N. COWIE, A. BRITTING, A. CLARKSON, *et al.*: **A focussing disc DIRC for PANDA**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **639**:1 (2011). Proceedings of the Seventh International Workshop on Ring Imaging Cherenkov Detectors, pp. 181–184. DOI: 10.1016/j.nima.2010.09.132 (see p. 50).

[95] **PANDA Website: Aerogel Ring Imaging Cherenkov Counter**. URL: http://www-panda.gsi.de/framework/det_iframe.php?section=RICH (visited on 12/27/2014) (see p. 51).

[96] THE PANDA COLLABORATION: **Technical Design Report for the: PANDA Muon System**. Tech. Rep. 2012. URL: http://www-panda.gsi.de/archive/MuonTDR/Muon_TDR.pdf (see pp. 51–52).

[97] THE PANDA COLLABORATION: **Technical Design Report for: PANDA Electromagnetic Calorimeter (EMC)**. Tech. Rep. Oct. 2008. eprint: arXiv:0810.1216. URL: http://arxiv.org/abs/0810.1216 (see pp. 53–54).

[98] THE PANDA COLLABORATION: **Technical Design Report for: PANDA Forward Spectrometer Calorimeter (FSC)**. Tech. Rep. *In preparation*, 2015 (see pp. 55–56).

[99] THE PANDA COLLABORATION: **Technical Design Report for the PANDA Luminosity Detector**. Tech. Rep. *In preparation*, 2015 (see p. 57).

[100] K. GÖTZEN, D. KANG, R. KLIEMT, and F. NERLING: **Present Status of the PANDA Software Trigger**. PANDA Internal Note. Apr. 2014 (see pp. 59, 65).

[101] I. KONOROV, H. ANGERER, A. MANN, and S. PAUL: **SODA: Time distribution system for the PANDA experiment**. In: *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*. Oct. 2009, pp. 1863–1865. DOI: 10.1109/NSSMIC.2009.5402172 (see p. 60).

[102] **PANDA Wiki: Panda DAQ**. URL: https://panda-wiki.gsi.de/foswiki/bin/view/Daq/WebHome (visited on 03/11/2015) (see p. 61).

[103] RENE BRUN and FONS RADEMAKERS: **ROOT — An Object Oriented Data Analysis Framework**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **389**:1 (1997), pp. 81–86 (see p. 66).

[104] M. AL-TURANY, D. BERTINI, R. KARABOWICZ, *et al.*: **The FairRoot framework**. In: *Journal of Physics: Conference Series*, **396**:2 (2012), p. 022001. DOI: doi:10.1088/1742-6596/396/2/022001 (see p. 67).

[105] M. AL-TURANY, D. KLEIN, A. MANAFOV, A. RYBALCHENKO, and F. UHLIG: **Extending the FairRoot framework to allow for simulation and reconstruction of free streaming data**. In: *Journal of Physics: Conference Series*, **513**:2 (2014), p. 022001. DOI: 10.1088/1742-6596/513/2/022001 (see p. 67).

[106] STEFANO SPATARO and THE PANDA COLLABORATION: **The PandaRoot framework for simulation, reconstruction and analysis**. In: *Journal of Physics: Conference Series*, **331**:3 (2011), p. 032031. DOI: 10.1088/1742-6596/331/3/032031 (see p. 68).

[107] DAVID J. LANGE: **The EvtGen particle decay simulation package**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **462**:1 (2001), pp. 152–155. DOI: `doi:10.1016/S0168-9002(01)00089-4` (see p. 68).

[108] **EvtGen Project**. URL: `http://evtgen.warwick.ac.uk/` (visited on 02/20/2015) (see p. 68).

[109] A. CAPELLA, U. SUKHATME, C.-I. TAN, and J. TRAN THANH VAN: **Dual parton model**. In: *Physics Reports*, **236**:4-5 (Jan. 1994), pp. 225–329 (see p. 68).

[110] I. KAWRAKOW, H.J. MOHRING, and J. RANFT: **The Monte Carlo dual parton model DTUNUC**. In: *Rencontres de Moriond, Les Arcs (*Proceedings*)*, (Mar. 1992), pp. 449–452 (see p. 68).

[111] S. ROESLER, R. ENGEL, and J. RANFT: **The Monte Carlo Event Generator DPMJET-III**. In: *Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications*. Springer Berlin Heidelberg, 2001, pp. 1033–1038. DOI: `10.1007/978-3-642-18211-2_166` (see p. 68).

[112] A. GALOYAN, J. RITMAN, A. SOKOLOV, and V. UZHINSKY: **Parametrization of the $\bar{P}P$ Elastic Scattering Differential Cross Section Between 2 GeV/c $\leq P_{lab} \leq$ 16 GeV/c**. Sept. 2008. arXiv: `0809.3804 [hep-ex]` (see p. 69).

[113] TORBJÖRN SJÖSTRANDA, STEFAN ASKB, JESPER R. CHRISTIANSENA, *et al.*: **An introduction to PYTHIA 8.2**. In: *Computer Physics Communications*, (Feb. 2015). In Press, Accepted Manuscript. DOI: `10.1016/j.cpc.2015.01.024` (see p. 69).

[114] STEFFEN A. BASS, MOHAMED BELKACEM, MARCUS BLEICHER, *et al.*: **Microscopic models for ultrarelativistic heavy ion collisions**. In: *Progress in Particle and Nuclear Physics*, **41**: (1998), pp. 255–369 (see p. 69).

[115] I. HRIVNACOVA, D. ADAMOVA, V. BEREJNOI, *et al.*: **The Virtual Monte Carlo**. In: *Proceedings to Computing in High Energy and Nuclear Physics*, (2003). arXiv: `0306005v1 [cs.SE]` (see p. 69).

[116] S. AGOSTINELLI, J. ALLISON, K. AMAKO, *et al.*: **Geant4 – a simulation toolkit**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **506**:3 (2003), pp. 250–303. DOI: `10.1016/S0168-9002(03)01368-8` (see p. 69).

[117] C. HÖPPNER, S. NEUBERT, B. KETZER, and S. PAUL: **A novel generic framework for track fitting in complex detector systems**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **620**:2–3 (2010), pp. 518–525. DOI: `10.1016/j.nima.2010.03.136` (see p. 70).

[118] RUDOLPH EMIL KALMAN: **A New Approach to Linear Filtering and Prediction Problems**. In: *Transactions of the ASME–Journal of Basic Engineering*, **82**:Series D (1960), pp. 35–45 (see p. 70).

[119] R. FRÜHWIRTH: **Application of Kalman filtering to track and vertex fitting**. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **262**:2–3 (1987), pp. 444–450. DOI: `10.1016/0168-9002(87)90887-4` (see pp. 70, 84).

[120]   **Rho: A Framework for Particle Physics Analysis**. URL: https://code.google.com/p/rhoframework/ (visited on 02/20/2015) (see p. 71).

[121]   RALF KLIEMT: **Simulations with the PANDA Micro-Vertex-Detector**. PhD Thesis. Friedrich-Wilhelms-Universität Bonn, 2012 (see p. 71).

[122]   STANFORD UNIVERSITY GRAPHICS LAB: **BrookGPU**. URL: http://graphics.stanford.edu/projects/brookgpu/ (see p. 74).

[123]   NVIDIA: **CUDA Zone**. URL: https://developer.nvidia.com/cuda-zone (see p. 74).

[124]   KHRONOS GROUP: **OpenCL – The open standard for parallel programming of heterogeneous systems**. URL: https://www.khronos.org/opencl/ (see p. 74).

[125]   NEW YORK UNIVERSITY PROF. MOHAMED ZAHRAN: *Lecture*: **Graphics Processing Units (GPUs): Architecture and Programming, 2012, Lecture 2: History of GPUs**. URL: http://cs.nyu.edu/courses/spring12/CSCI-GA.3033-012/ (see p. 74).

[126]   RUIGANG YANG and CHRISTOPHER ZACH: *Lecture*: **GPGPU: General Purpose Programming on the Graphics Processing Unit, 2011, Week 3: GPU History and Architecture**. URL: https://www.cvg.ethz.ch/teaching/2011spring/gpgpu/Spring2011Course.php (see p. 74).

[127]   WIKIPEDIA: **General-purpose computing on graphics processing units – Wikipedia, The Free Encyclopedia**. [Online; accessed 14-March-2015]. 2015. URL: http://en.wikipedia.org/w/index.php?title=General-purpose_computing_on_graphics_processing_units&oldid=651303266 (see p. 74).

[128]   OAK RIDGE NATIONAL LABORATORY: **History of CUDA, OpenCL, and the GPGPU**. URL: https://www.olcf.ornl.gov/kb_articles/history-of-the-gpgpu/ (see p. 74).

[129]   JASON SANDERS and EDWARD KANDROT: **CUDA by Example: An Introduction to General-Purpose GPU Programming**. 1st. Addison-Wesley Professional, 2010 (see p. 74).

[130]   KAMRAN KARIMI, NEIL G. DICKSON, and FIRAS HAMZE: **A Performance Comparison of CUDA and OpenCL**. In: *CoRR*, **abs/1005.2581**: (2010). arXiv: 1005.2581 [cs.PF] (see p. 74).

[131]   CHING-LUNG SU, PO-YU CHEN, CHUN-CHIEH LAN, LONG-SHENG HUANG, and KUO-HSUAN WU: **Overview and comparison of OpenCL and CUDA technology for GPGPU**. In: *Circuits and Systems (APCCAS), 2012 IEEE Asia Pacific Conference on*. Dec. 2012, pp. 448–451. DOI: 10.1109/APCCAS.2012.6419068 (see p. 74).

[132]   JIANBIN FANG, A.L. VARBANESCU, and H. SIPS: **A Comprehensive Performance Comparison of CUDA and OpenCL**. In: *Parallel Processing (ICPP), 2011 International Conference on*. Sept. 2011, pp. 216–225. DOI: 10.1109/ICPP.2011.45 (see p. 74).

[133]   NVIDIA: **CUDA C Programming Guide**. URL: http://docs.nvidia.com/cuda/cuda-c-programming-guide/ (see pp. 74, 77).

[134]   UNIVERSITY OF OXFORD PROF. MIKE GILES: *Lecture*: **Course on CUDA Programming on NVIDIA GPUs, 2014, Lecture 1: An introduction to CUDA**. URL: http://people.maths.ox.ac.uk/gilesm/cuda/ (see p. 77).

[135] R. AMMENDOLA, A. BIAGIONI, O. FREZZA, *et al.*: **APEnet+: a 3D Torus network opti-**
       mized for GPU-based HPC Systems. In: *Journal of Physics: Conference Series*, **396**:4
       (2012), p. 042059. DOI: 10.1088/1742-6596/396/4/042059 (see p. 84).

[136] R. AMMENDOLA, A. BIAGIONI, O. FREZZA, *et al.*: **Analysis of performance improve-**
       ments for host and GPU interface of the APENet+ 3D Torus network. In: *Jour-*
       nal of Physics: Conference Series, **523**:1 (2014), p. 012013. DOI: 10.1088/1742-
       6596/523/1/012013 (see p. 84).

[137] D. H. JONES, A. POWELL, C. BOUGANIS, and P. Y. K. CHEUNG: **GPU Versus FPGA for**
       High Productivity Computing. In: *Field Programmable Logic and Applications (FPL),*
       2010 International Conference on. Aug. 2010, pp. 119–124. DOI: 10.1109/FPL.2010.32
       (see p. 85).

[138] CHRISTOPHER CULLINAN, CHRISTOPHER WYANT, TIMOTHY FRATTESI, and XINMING
       HUANG: **Computing performance benchmarks among CPU, GPU, and FPGA**. In:
       (2013). URL: https://www.wpi.edu/Pubs/E-project/Available/E-project-030212-
       123508/unrestricted/Benchmarking_Final.pdf (see p. 85).

[139] P. V. C. HOUGH: **Machine Analysis Of Bubble Chamber Pictures**. In: *International*
       Conference on High Energy-Accelerators. Vol. 73. Sept. 1959, pp. 554–558 (see p. 87).

[140] RICHARD O. DUDA and PETER E. HART: **Use of the Hough Transformation to Detect**
       Lines and Curves in Pictures. In: *Commun. ACM*, **15**:1 (Jan. 1972), pp. 11–15. DOI:
       10.1145/361237.361242 (see pp. 87, 92, 96).

[141] DAVID MUENCHOW, MARTIN GALUSKA, THOMAS GESSLER, *et al.*: **An FPGA helix track-**
       ing algorithm for PANDA. In: *Verhandlungen der Deutschen Physikalischen Gesellschaft*,
       (2010) (see p. 99).

[142] PANDA COLLABORATION, W. ERNI, I. KESHELASHVILI, *et al.*: **Technical Design Report**
       for the: PANDA Straw Tube Tracker. Tech. Rep. arXiv:1205.5441. Comments:
       accdpted for publication on EPJA. May 2012. DOI: 10.1140/epja/i2013-13025-8 (see
       p. 99).

[143] LUKE OLSON STEVEN DALTON NATHAN BELL and MICHAEL GARLAND: **Cusp: Generic**
       Parallel Algorithms for Sparse Matrix and Graph Computations. Version 0.5.0.
       2014. URL: http://cusplibrary.github.io/ (see p. 99).

[144] M. C. MERTENS FOR THE PANDA COLLABORATION: **Triplet based online track finding**
       in the PANDA-STT. In: *Journal of Physics: Conference Series*, **503**:1 (2014), p. 012036.
       DOI: 10.1088/1742-6596/503/1/012036 (see pp. 128, 131–132).

[145] MARIUS C. MERTENS, KAI-THOMAS BRINKMANN, JAMES RITMAN, and PETER WINTZ:
       Triplet based online track finding in the PANDA-STT. In: *Hyperfine Interactions*,
       229:1-3 (Oct. 2014), pp. 153–158. DOI: 10.1007/s10751-014-1062-3 (see pp. 128–
       129, 131).

[146] ANDREW ADINETZ, ANDREAS HERTEN, JIRI KRAUS, *et al.*: **Triplet Finder: On the Way to**
       Triggerless Online Reconstruction with GPUs for the PANDA Experiment. In: *Pro-*
       cedia Computer Science, **29**: (2014). 2014 International Conference on Computational
       Science, pp. 113–123. DOI: 10.1016/j.procs.2014.05.011 (see pp. 132, 217).

[147] ANDREW ADINETZ, ANDREAS HERTEN, JIRI KRAUS, *et al.*: **Triplet Finder: On the Way to Triggerless Online Reconstruction with GPUs for the PANDA Experiment**. In Publication in *Journal of Computer Science*. 2015 (see p. 132).

[148] J. HAIDENBAUER and G. KREIN: **Production of charmed pseudoscalar mesons in antiproton-proton annihilation**. In: *Phys. Rev. D*, **89**: (11 June 2014), p. 114003. DOI: 10.1103/PhysRevD.89.114003 (see pp. 143, 179).

[149] CLEO COLLABORATION: **Dalitz plot analysis of the $D^+ \to K^- \pi^+ \pi^+$ decay**. In: *Phys. Rev. D*, **78**: (5 Sept. 2008), p. 052001. DOI: 10.1103/PhysRevD.78.052001 (see pp. 145–146).

[150] NVIDIA: **Application Note – NVIDIA GPU BOOST FOR TESLA**. DA-06767-001_v02. URL: http://www.nvidia.com/content/PDF/kepler/nvidia-gpu-boost-tesla-k40-06767-001-v02.pdf (see pp. 219, 234).

[151] JARED HOBEROCK and NATHAN BELL: **Thrust: A Parallel Template Library**. 2010. URL: http://thrust.github.io/ (see p. 235).

Glossary

AAC Antiproton Accumulation Complex

AD Antiproton Decelerator

ADC Analog-to-Digital Converter

ALU Arithmetic Logic Unit

API Short for Application Programming Interface. A standardized set of routines for accessing a certain software.

APPA Atomic, Plasma Physics and Applications

ASIC Application-Specific Integrated Circuit

ATLAS One of the two multi-purpose experiments at LHC.

BaBar A particle detector located at SLAC in Stanford, USA, detecting events of e^-e^+ interactions.

Belle A particle detector located at the KEK laboratory in Tsukuba, Japan, detecting events of e^-e^+ interactions.

BIOMAT Biology and Material Science

Burst HESR's beam is structured into a burst of antiprotons, followed by a gap without particles.

CBM Compressed Baryonic Matter

CERN Conseil Européenne pour la Recherche Nucléaire

CH Circle Hough Transform

CLEO A particle detector located at the Cornell Electron Storage Ring at the Cornell University in Ithaca, USA. The detector was in operation until 2008.

CM center-of-mass

CMS Compact Muon Solenoid

Conformal Map A mathematical mapping function preserving angles between curves going through a mapped point u_0.

COSY Cooler Synchrotron

CPU Central Processing Unit

CR Collector Ring

CRYRING Antiproton and ion storage ring located at FLAIR.

Crystal Barrel Experiment A particle detector first used at CERN's LEAR ring. Mesons were detected using a large calorimeter. After closing of LEAR, the experiment is now operated at ELSA in Bonn, Germany.

DAQ Data Acquisition

DESPEC/HISPEC Decay Spectroscopy/High-Resolution Spectroscopy

DIRC Detection of Internally Reflected Cherenkov Light

DPM Literally: Dual Parton Model – a model describing background events in antiproton-proton annihilation. Often colloquially used to refer to the event generator based on DPM.

DUT Device Under Test

eA Collider Electron Ion Collider

ECC Short for Error-Correcting Code. The phrase covers techniques for encoding digital message redundantly to enable the possibility for data recovery of erroneous messages.

ELISe Electron-Ion Scattering experiment

EMC Electromagnetic Calorimeter

ESR Experimental Storage Ring

EvtGen An event generator used in PandaRoot.

EXL Exotic nuclei studied in light-ion induced reactions at the NESR storage ring

FAIR Facility for Antiproton and Ion Research

FairMQ A message-based data transport system in FairRoot.

FairRoot A common software framework for FAIR, based on CERN's ROOT.

Fermilab Fermi National Accelerator Laboratory in Illinois, USA.

FLAIR Facility for Low-Energy Antiproton and Heavy Ion Research

Forward Spectrometer The part of the $\overline{\text{P}}$ANDA detector measuring at shallow angles in the forward direction.

FPGA Field-Programmable Gate Array

FRS Fragment Separator

FSC Forward Spectrometer Calorimeter

F-TOF Forward TOF

FTS Forward Tracking System

FZJ Forschungszentrum Jülich

GEANT3 A Monte Carlo event generator and propagator.

GEANT4 A Monte Carlo event generator and propagator.

GEM Gas Electron Multiplier

GPD Generalized Parton Distribution

GPGPU General Purpose Graphics Processing Unit

GPU Graphics Processing Unit

GPUBoost A feature available on NVIDIA GPUs, enabling GPU processes to use unused »power headroom«. See »Application Note – NVIDIA GPU BOOST FOR TESLA« [150].

GSI GSI Helmholtz Centre for Heavy Ion Research

HADES High-Acceptance Dielectron Spectrometer

HEDgeHOB High Energy Density Matter generated by Heavy Ion Beams

HESR High-Energy Storage Ring

HITRAP A device for trapping low energy antiprotons, part of FLAIR.

Hough Transform A method for extracting features from images. Used in this work to find straight lines connecting hit points to form track candidates.

HPC High-Performance Computing

HT Hough transform

HV-MAPS High Voltage Monolythic Active Pixel Sensor

ILIMA Isometric Beams, Lifetimes and Masses

LaSpec Laser Spectroscopy

Lattice QCD A theoretical discipline describing QCD dynamics using interactions on a four-dimensional lattice. Abbreviated LQCD.

LEAR The Low Energy Antiproton Ring was a antiproton storage ring located at CERN, operated from 1982 until 1996.

LHC Large Hadron Collider

LHT Line Hough Transform

LMD Luminosity Detector

LSR Low Energy Storage Ring

Magnetic Rigidity A quantity describing how much a track is bent by an external magnetic field. Important for quantifying the strength of an accelerator, where it is usually given in Bρ, the product of the magnetic field and the gyroradius (Larmor radius).

MATS Precision Measurements of very short-lived nuclei with Advanced Trapping System

MC Monte-Carlo

MDT Mini Drift Tube

Modularized Start Version The initial part of FAIR's staged starting version.

MSV Modularized Start Version

MVD Micro Vertex Detector

NESR New Experimental Storage Ring

NUSTAR Nuclear Structure, Astrophysics and Reactions

OpenCL The Open Computing Language, a framework for writing parallel programs.

P̅ANDA Antiproton Annihilation at Darmstadt
PandaRoot P̅ANDA's computing framework.
PASTA PANDA Strip ASIC
PDG Particle Data Group
Penning Trap A device for trapping charged particles (e.g. ions) using a magnetic field, named after F. M. Penning.
PHELIX Petawatt High-Energy Laser for Heavy Ion EXperiments
PID Particle Identification
p-LINAC The proton linear accelerator for FAIR.
PMT photomultiplier tube

QCD Quantum Chromo Dynamics
QED Quantum Electrodynamics
QFT Quantum Field Theory

R3B Reactions with Relativistic Radioactive Beams
Race Condition Race conditions can occur when the output of one computing process is input to another. Especially in parallel programming, race conditions can become critical and need to be considered carefully.
Radiation Length A property of a specific material, denoted X_0. The mean distance an electron travels through this matter, until the electron's energy is reduced to $^1/e$ (by means of bremsstrahlung).
RESR Recuperated Experimental Storage Ring
RF Cavity Radiofrequency cavities are used to accelerated charged particles by means of applied electromagnetic fields.
RICH Short for Ring-Imaging Cherenkov. A detector using rings of Cherenkov light for PID.
ROOT A data analysis framework from CERN.

SciTil Scintillating Tiles – usually referring to P̅ANDA's time of flight detector.

Shashlik Electromagnetic Calorimeter The part of the electromagnetic calorimeter in the Forward Spectrometer. Named after its characteristic structure.
SIMD Single Instruction, Multiple Data
SIMT Single Instruction, Multiple Threads
SiPM Silicon Photo Multiplier
SIS18 The first accelerating synchrotron of FAIR.
SIS100 The main synchrotron of FAIR's accelerator chain.
SIS300 Synchrotron in the same tunnel as SIS100.
SLAC Stanford Linear Accelerator Center
SM Standard Model
SMM Streaming Multiprocessor
SODA Synchronization Of Data Acquisition
SPARC Stored Particles Atomic Research Collaboration
STL Standard Template Library
STT Straw Tube Tracker
Super-FRS Super Fragment Separator

Target Spectrometer The part of the P̅ANDA detector around the interaction region, instrumenting nearly a 4π solid angle.
μ**TCA** Micro Telecommunications Computing Architecture
TF Triplet Finder
Thread A computing process on a GPU.
Thrust A template library providing functions for often-used GPU-computing tasks, resembling the C++ Standard Template Library (STL). Thrust adds a layer of abstraction on top of CUDA and supplies pre-programmed functions for many tasks, e.g. reduction and sorting. [151]
TOF Time Of Flight
ToPix The read-out ASIC for the pixel part of P̅ANDA's MVD.

UNILAC Linear Accelerator for ions at GSI.
USR Ultra Low Energy Storage Ring

VMC Virtual Monte Carlo

Warp A set of 32 threads executed at the same time on a GPU.

WASA Wide Angle Shower Apparatus

WDM Warm Dense Matter

WSF Wavelength-Shifting Fibres

xTCA Extended Telecommunications Computing Architecture

List of Figures

List of Tables

Acknowledgments

Thanks to everyone, who made this work possible and the time spend on it very enjoyable.

First and foremost James Ritman. Jim, thank you for the position in your group and for the extensive discussions and continuous input during the course of this thesis.

Second, Tobias Stockmanns. Thank you for being just one room away to answer all the questions I constantly had.

Thanks to all the **IK-P**ANDAs who made the last three years as great as they were: Qiang, Ludovico, Lu, Marius, Huagen, Dariusch, André, Simone, Frank, Albrecht, Peter, Elisabetta, Harout, Jonathan. And of course all the other members of the IKP! Artur, Farha, Florian, Daniel, Günter, Maria, Rene. And all the ones I forgot.

Thanks to Dirk, Jiri, and Andrew from the Application Lab. Thank you for the possibility to be included into the studies of your lab and all the work you did for $\overline{\text{P}}$ANDA.

Thanks to the members of $\overline{\text{P}}$ANDA, but especially the PandaRoot core developers. Without you, I would not have had the tools to study all the things I presented here. Just to name three persons, which were always happy to answer my questions: Thanks Klaus, Ralf, and Stefano. Thank you, Mohammad, for giving me the chance to learn about Hough transforms and GPUs in your group.

Thank you, Andrea, for being there for me and supporting me during daily and nightly work on this thesis. Danke euch, liebe Eltern, dass ihr mir das Physikstudium ermöglicht und mich auf meinem Weg immer unterstützt habt.

Lebenslauf *Curriculum Vitae*

Persönliche Daten

Name: Andreas Herten
Geburtsdatum: 12.10.1984
Geburtsort: Eschweiler

Schulbildung

1991 - 1995   Kath. Grundschule Langerwehe
1995 - 2004   Rurtal-Gymnasium Düren
                   Allgemeine Hochschulreife

Studium

2005 - 2010   Studium der Physik, RWTH Aachen
                   Diplomarbeit: »Bestimmung der Masse des Top-Quarks
                   anhand der Zerfallslängen von B-Hadronen im CMS-Experiment«

Promotion

2011 - 2015   Promotion im Fach Physik an der Ruhr-Universität
                   Bochum in Kooperation mit dem Institut für
                   Kernphysik am Forschungszentrum Jülich
                   Doktorarbeit: »GPU-based Online Track Reconstruction
                   for $\bar{\text{P}}$ANDA and Application to the Analysis of $D \rightarrow K\pi\pi$«