# DIPLOMARBEIT

Titel der Diplomarbeit

## Pattern recognition in the Silicon Vertex Detector of the Belle II experiment

Verfasser

## Jakob Lettenbichler

angestrebter akademischer Grad

## Magister der Naturwissenschaften (Mag.rer.nat.)

Wien, 2012

# Contents

# List of Figures

# List of Tables

# 1. Introduction

## 1.1. Overview

The purpose of this thesis is to present a possible way of track finding specialized for low momentum tracks within the Silicon Vertex Detector (SVD) of the Belle II experiment. Track finding algorithms are an essential tool for the reconstruction of tracks of elementary particles moving through the detector.

Chapter 1.2 is a general description of accelerators used for particle collision experiments and detectors used for recording tracks of produced particles. Chapter 1.3 gives an overview of the steps of event reconstruction.

Chapter 2 describes the current situation at Belle and the future Belle II. 2.1 includes some important physical aspects and their effect on the design of the experiment. 2.2 explains the concept of the superKEKB-Accelerator used for this experiment and shows how the design has been adapted to current purposes. 2.3 gives an overview of the detector and explains some details important to pattern recognition.

Chapter 3 starts with a list of requirements for pattern recognition defined by the current situation in 3.1 and 3.2. Furthermore it presents some typical algorithms used for track finding and explains their advantages and disadvantages. In particular, conformal mapping 3.3, Hough transform 3.4, the Kalman filter 3.5 and the Hopfield network 3.6 are discussed. Chapter 3.7 contains our current concept for the low momentum track finder in full detail.

Finally, its performance is compared with common techniques in 4.

## 1.2. The toolbox of particle physicists

Before having a closer look at the details of pattern recognition, one has to know the general conditions of high energy physics experiments, starting with the most important principles of particle physics. The following sections give an overview on some widely used techniques for particle acceleration and detection.

### 1.2.1. Accelerators

For more details about accelerators, see [1], [3], [7], [8] and [5].

There are no man-made machines capable of producing particles with energies as high as those coming from the highest-energetic natural sources (produced by supernovae and other cosmic sources, arriving at the Earth in form of cosmic rays). At the moment, there are machines which are able to accelerate particles to up to 4 TeV, but their primary

advantage is the capability of producing a sufficient number of events (where new particles are produced) per second to be able to get enough data for statistical analysis. The most common types of accelerators are linear and circular colliders. Llinear colliders are not discussed in this thesis; see for example [3] for more details on this as well as on fixed target experiments. Circular colliders are a type of storage rings, which are a subtype of synchrotrons themselves. They accelerate bunches of charged particles via oscillating electric fields provided by cavities; cavities are hollow conductors that carry electromagnetic waves and increase the intensity of certain frequencies (resonant frequencies). Accelerators, however, are expensive equipment; to reduce the cost, the bunch of particles is forced into a circular path by a magnetic field (Lorentz force) passing the same accelerating segments on this path many times. A set of bunches is traveling in an evacuated ring and another set is moving into the opposite direction while using the same or another ring (depending on particle type and other design requests).
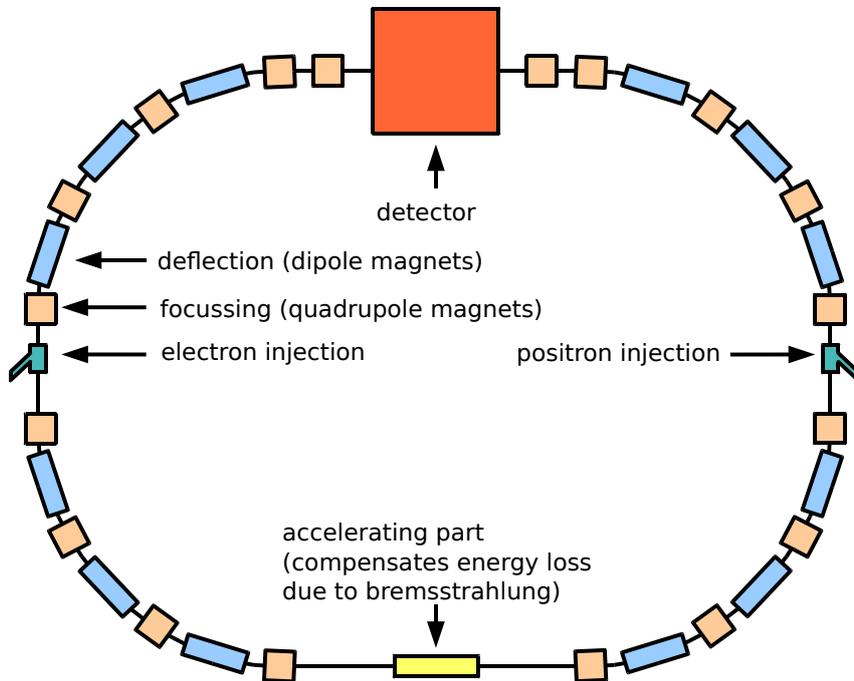


Figure 1.1.: sketch of a lepton storage ring

Some accelerators use the same type of particle for both directions, others use a particle for one direction and its antiparticle for the other direction. Circular accelerators can be subdivided into hadron colliders (using mostly protons and antiprotons, but also heavy-ion-setups are known) and lepton colliders (electrons and positrons). Both have their own advantages and disadvantages: while lepton colliders are much more precise (and therefore used for detailed analysis of special decay channels and for the detailed determination of parameters of known particles), their main disadvantage is

their limitation by the synchrotron radiation which is emitted per turn:

$$\Delta E_{\text{s}} = \frac{4\pi}{3} e^2 \beta^2 \frac{\gamma^4}{R} \sim \frac{E^4}{m^4} \frac{1}{R} \tag{1.1}$$

- $e$ is the elementary charge (of the electron)

- $\beta = \frac{v}{c}$ the velocity

- $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ the Lorentz factor (the product $\beta\gamma$ is often used to describe the velocity of a particle of arbitrary mass, since the mass term disappears when written this way. This is interesting for example checking the energy loss of particles, see (1.2.2.1) for more details.)

- $E$ the initial energy

- $m$ the mass of the particle

- $R$ is the radius of the accelerator

Synchrotron radiation strongly depends on the particle mass. The accelerator has to inject more energy per turn than the amount of energy emitted by the particles in the same time to be able to accelerate them. Lighter particles like the electron/positron (currently the only candidates for lepton colliders) are limited by this. Looking at the mass ratio of electrons and protons reveals the importance of that problem:

$$\frac{m_{\text{e}}^4}{m_{\text{p}}^4} \sim 10^{-13} \tag{1.2}$$

Consequently, electron machines are limited to about LEP energy, which is in the order of $\sim 100$ GeV. For studies of higher energies, hadron colliders are used. Famous examples of hadron colliders are the LHC at CERN (Switzerland) and the now deactivated Tevatron at Fermilab (USA). They allow for collision energies of up to currently 8 TeV, but are less precise because of the fact that the momentum of hadrons can not be determined as precisely as the momentum of elementary particles like leptons. The reason for that is that the momentum is carried by the constituent quarks and gluons which cannot be measured precisely because of the fact that they are interacting with each other all the time and therefore exchange momentum. This also means that there are 3 scatterong centers (at the quarks) instead of one.

A typical setup of an accelerator facility starts with the particle sources. For lepton colliders electrons are easy to get, but positrons have to be produced e.g. by irradiating a tungsten target with preaccelerated electrons and capturing the produced positrons[5]. For hadron colliders, protons are produced by using negatively charged hydrogen ions which are shot through a carbon foil at several MeV, where the electrons are collected and bare protons are left behind. To get antiprotons, some of the protons are shot at a tungsten target and emitted antiprotons can be collected[7]. All these particles are preaccelerated by smaller linear accelerators, and depending on the level of target

energy, a chain of accelerators (linear and/or circular) is used before injecting bunches of particles in the final collider ring. A main reason for requiring this chain of accelerators is that synchrotrons need particles with an initial energy of at least $\Delta E_\text{s}$ (see equation (1.1)). These bunches consist of many particles of the same type (for example about $10^9$ particles per bunch at the KEKB Collider of the Belle experiment [5]), which reveals several problems: first, they are of the same charge, which means that they are repelling each other and therefore exchange transverse to longitudinal momentum. This leads to the Touschek-effect [9], where single particles are kicked out of the bunch and leave the ring. It is one of the main effects contributing to beam lifetime, especially in lepton storage rings. This is one of the major causes of background within the detector at lower energy accelerators ($\sim$MeV/GeV), but rather irrelevant for high energy accelerators like the LHC or the Tevatron. It also means that the particles disperse and have to be refocused constantly. This is done by pairs of quadrupole magnets where the first one is focusing horizontally and the second one is focusing vertically. But not only the focusing in direction has to be corrected, but differences in momentum have to be considered too. Several techniques, like stochastic or electron cooling, are used to reduce the standard deviation of the momentum distribution. Focusing and cooling improve the quality of the beam and are, next to the achieved momenta, the most important features of accelerators. After accelerating the particles to their final energy, the particle beams are crossed at a certain point (*interaction point (IP)* or *interaction region (IR)*) to use their collision energy for the production of new particles. The event (or interaction) rate $R_\text{int}$ is another important characteristic of the accelerator:

$$R_\text{int} = \sigma \cdot L \tag{1.3}$$

- $\sigma$ is the intrinsic cross section $\sigma = \pi \cdot (2r_\text{e})^2$ — where $r_\text{e}$ is the radius of the particle — has dimension area. Measured in barn b $= 10^{-24}\text{cm}^2$

- $L$ is the *luminosity*, which is defined by:

$$L = f \cdot n \cdot \frac{N_1 \cdot N_2}{A} \tag{1.4}$$

  - $f$ is the revolution frequency of the accelerator
  - $n$ is the number of bunches in the ring
  - $N_1, N_2$ are the numbers of particles per bunch in both directions
  - and $A$ is the area of overlapping bunches (the smaller the area, the higher the event rate, affected by the quality of cooling and focusing steps)
  - the unit of the luminosity is $\text{b}^{-1} \cdot \text{s}^{-1}$. Another derivation of it is the *integrated luminosity* (unit $\text{b}^{-1}$) which is the total number of events taken place within a defined amount of time (e.g. total runtime of the experiment).

For obvious reasons it is the goal of every experiment to achieve an $R_\text{int}$ as high as technically possible. Unfortunately, increasing the luminosity (and therefore the event

rate) of an accelerator increases the Touschek effect, which is directly proportional to $N_1, N_2$ and indirectly proportional to $A$. The only way to decrease the Touschek effect while increasing the luminosity is to increase the energy per particle, which is not always compatible with the requirements of the experiment.

A bunch crossing where particles of both bunches are interacting with each other is called an *event*. In such an event, new particles are created from the collision energy of the primary particles. Depending on the result of the collision, there will be two or more secondary particles that have their own individual lifetime until decaying. Most of the known particles have lifetimes far shorter than a second. Because of the fact that the lifetime of a particle is proportional to the distance of flight, technical reasons (e.g. placement of detector layers) prevent us from detecting particles with lifetimes of less than $\sim 10^{-10}$ s directly (see chapter 1.3 for more details). Long lived particles will escape the interaction region and ideally fly through the detector which detects and records the results of the event. This will be described in the next subsection.

## 1.2.2. Detectors

The section *Detectors* and its sub-sections 1.2.2.1, 1.2.2.2, 1.2.2.3 and 1.2.2.4 is a summary of common techniques and detector types used in the field of particle physics. More details about detectors and the behavior of particles within matter can be found in [1], [2], [5], [10], [11], [12] and [6].

To be able to reconstruct what happens in an event, several types of detectors are arranged around the interaction point to record traces of the particles produced in the collision. Detectors use certain types of interactions between charged particles and matter to obtain information about the particle (trajectory, momentum, energy, velocity). Neutral particles interact differently with matter, which requires different methods of reconstruction; taking photons as an example, see page 18, for neutrons see page 25. The quality of a certain detector type can be characterized by several parameters. One of them is the *total detector efficiency* which is limited by the dead time of the detector, its resolution and its sensitivity.

$$\epsilon_{\text{tot}} = \frac{\text{number of events registered}}{\text{total number of events emitted by source}} \tag{1.5}$$

Another parameter is the *intrinsic detector efficiency*,

$$\epsilon_{\text{i}} = \frac{\text{number of events registered}}{\text{total number of events impinging on the detector}} \tag{1.6}$$

where the latter is dependent on the solid angle covered by the detector module. Another important value (especially for event reconstruction) is the *signal to noise ratio*, which should be as good as possible, which means a strong signal versus small background (like thermal noise, unwanted side effects, etc).

### 1.2.2.1. Passage of particles through matter

Particles interact with the electrons and atoms of the material they are traversing, where as a result they change their direction (equation (1.8)) and lose energy (equation (1.2.2.1)).

At moderately relativistic velocities (roughly $0.1 < \beta\gamma < 1000$), the *energy loss* of most charged particles (except electrons, see equation (1.9)) can be described by the Bethe formula which includes effects like ionization (transferring enough energy to bonding electrons to free them from their bound state) and atomic excitation (transferring enough energy to a nucleus to raise it into excited states). The effects mentioned above dominate at these velocities and therefore the behavior of particles can be described well with the Bethe formula [11], which is:

$$-\frac{dE}{dx} = 4\pi N_\mathrm{A} r_\mathrm{e}^2 m_\mathrm{e} c z_\mathrm{p}^2 \frac{Z_\mathrm{matter}}{A_\mathrm{matter}} \frac{1}{\beta^2} \left[ \frac{1}{2} \ln \frac{2 m_\mathrm{e} c^2 \beta^2 \gamma^2 T_\mathrm{max}}{I^2} - \beta^2 - \frac{\delta_\mathrm{corr}}{2} \right] \qquad (1.7)$$

- $-\frac{dE}{dx}$ is the energy loss per unit length

- $N_\mathrm{A}$ is Avogadro's number

- $c$ the speed of light

- $z_\mathrm{p}$ the charge of the particle

- $Z_\mathrm{matter}$ the atomic number of the matter penetrated by the particle and responsible for the energy loss (absorber)

- $A_\mathrm{matter}$ the atomic mass of the absorber

- $T_\mathrm{max}$ is the maximum kinetic energy which is possible to be transferred to a free electron in a single collision

- $I$ is the mean excitation energy of the absorber

- $\delta_\mathrm{corr}$ is the density effect correction to ionization energy loss

Figure 1.2 shows the energy loss for a muon depending on its momentum. Many detector types use the ionization effect to obtain their measurements (which means that the passing particle is transferring a part of its momentum to bonding electrons of the sensitive material). But energy loss means that particles are changing their momentum while passing through the detector. The order of magnitude of the energy loss within the detector depends on the detector type (e.g. low for gaseous detectors or a total energy loss for calorimeters). It also complicates the event reconstruction. But it is not the only effect that represents a complication for the reconstruction. Another important aspect, the multiple scattering within the absorber, can be described by the Highland formula [11]:

$$\sigma_\theta = \frac{13.6 \mathrm{MeV}}{\beta c p_\mathrm{p}} z_\mathrm{p} \sqrt{x/X_0} \left[ 1 + 0.038 \ln\left(x/X_0\right) \right] \qquad (1.8)$$

- $\sigma_\theta$ is the standard deviation of the scattering angle
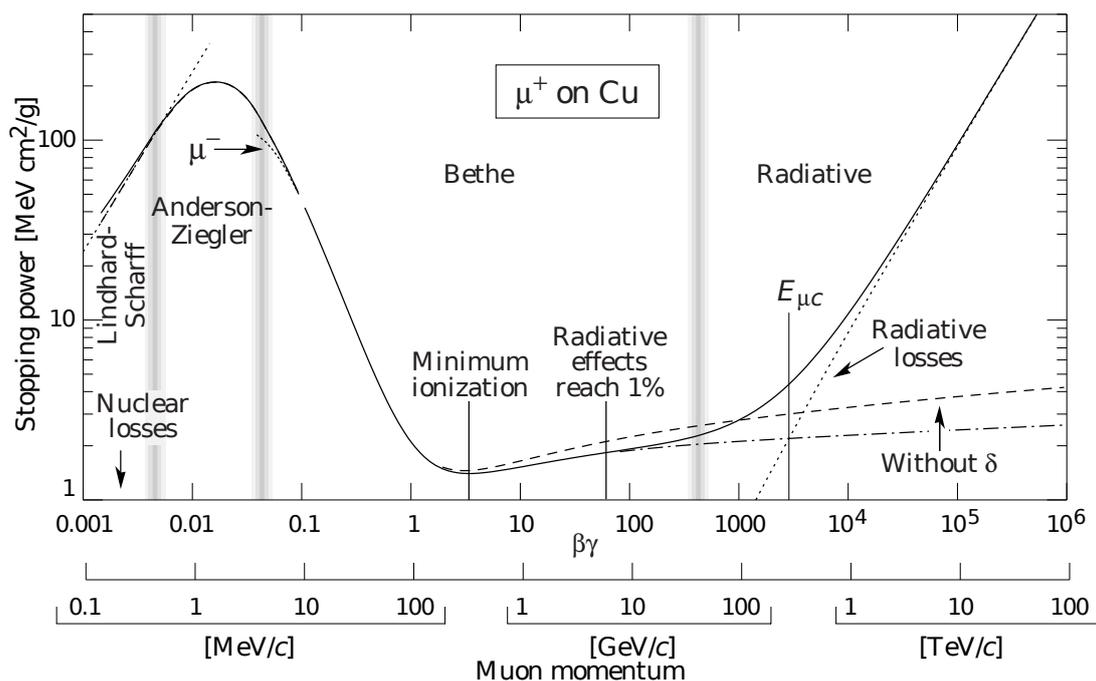
Figure 1.2.: Energy loss within copper for muons over a wide momentum range

- $p_p$ is the momentum of the particle

- $x/X_0$ is the thickness of the medium in radiation lengths $(X_0)$
  - a *radiation length* [11] for a high energy electron is the mean distance over which it loses roughly 62% of its energy by bremsstrahlung (see page 18). For a photon the radiation length is defined as $\frac{7}{9}$ of the mean free path for pair production (see page 18). $X_0$ depends on the type of the material and its density.

The Highland formula describes the *multiple Coulomb scattering* for small deflection angles (not more than a few $\sigma_\theta$), where the Coulomb scattering distribution is Gaussian, an illustration can be seen in figure 1.3. We therefore can use a Gaussian approximation for the central 98% of the projected angular distribution, where the Highland formula delivers the standard deviation of that Gaussian distribution (the mean is 0). The result is valid for the magnitude of trajectory deviations at thicknesses in the range of $10^{-3} < x/X_0 < 100$. For very thin layers of matter, the similarity of the distribution to a Gaussian decreases because of the tails of the distribution are getting larger and therefore this formula starts to differ significantly from experimental results. The reason for that behavior is the fact that this formula is describing the process of a sufficiently large number of Coulomb scattering from nuclei. When there are only a small number of nuclei lying at the trajectory through the layer of matter, single scattering processes will have a more visible impact on the final result and the similarity to a Gaussian distribution disappears
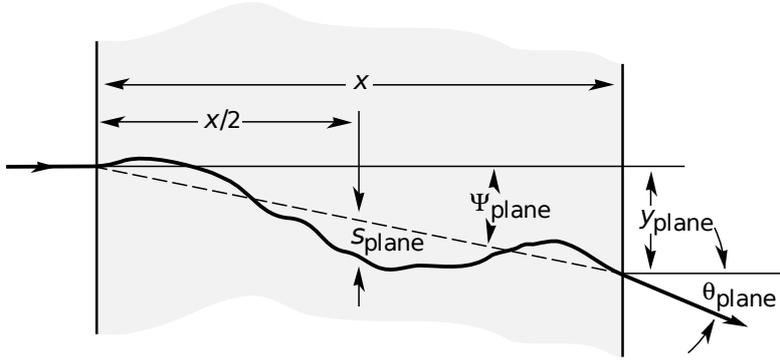
Figure 1.3.: Multiple scattering within matter

(central limit theorem). Very thick layers of matter represent a smaller problem, simply because of the fact that they are avoided anyway for tracking detectors — which are used for position estimation — but only used for energy measurement. For more details about this topic, see 1.2.2.3 and 1.2.2.2. Multiple scattering will be described in more detail in [11].

For electrons in high energy physics, *bremsstrahlung* is the dominating effect of energy loss which adds to collisional energy loss (which is described by a modified Bethe formula (1.2.2.1)): a photon is emitted by a high energy electron. For momentum conservation, this effect needs the presence of a magnetic field (for example of an atomic nucleus).

$$-\frac{dE}{dx} = 4\alpha N_A z_p^2 \frac{Z_{\text{matter}}^2}{A_{\text{matter}}} \left( \frac{1}{4\pi\epsilon_0} \frac{e^2}{m_p c^2} \right)^2 E \cdot \ln \frac{183}{Z_{\text{matter}}^{1/3}} \tag{1.9}$$

- $\alpha$ is the fine structure constant characterizing the strength of the electromagnetic interaction

- $\epsilon_0$ vacuum permittivity

- $m_p$ mass of the particle entering the material

Its strength is $\propto \frac{1}{m_p^2}$ which effectively means that it's quite important for the lightweight electron and much less important for the muon, which is about 200 times heavier. For electrons equation (1.9) can be simplified to:

$$-\frac{dE}{dx} = \frac{E}{X_0} \tag{1.10}$$

where $X_0$, the radiation length, is especially dependent on the nuclear $Z$ charge of the material. A special case of bremsstrahlung is synchrotron radiation (1.1), which is the same effect but with the magnetic field of synchrotrons instead of a nucleus for deflection. Bremsstrahlung is strongly coupled with an effect called *pair production* of high energy photons. Photons with energies $\gtrsim 1.022$ MeV ($\hat{=}$ the energy of a pair of electrons at

rest) can produce a pair of $e^+e^-$. This effect needs the presence of an atomic nucleus for momentum conservation, in contrast to the synchrotron radiation/bremsstrahlung where the field is sufficient (photons have no charge). Starting with an electron or a photon of sufficient energy, a cascade of coupled bremsstrahlung and pair production events can occur (see 1.2.2.3), which is called an *electromagnetic shower*. This means that one high energy particle produces a high number of lower energy particles. This process of pair production happens about once per radiation length. After $n$ radiation lengths the energy of the initial particle is split between $2^n$ created particles. This means that the mean path length of a particle of energy $E_\mathrm{p}$ is proportional to $\ln E_\mathrm{p}$. This is a very important fact for the construction of calorimeters, see 1.2.2.3 for more details.

For hadrons, a comparable effect is the *hadronic shower*, which is a series of inelastic scattering of hadrons and atomic nuclei. At such inelastic scattering processes new particles can be produced, which then decay themselves (which results in a combined hadronic and electro-magnetic shower).

Another important effect is *Cherenkov radiation* which occurs when a charged particle passes through matter with a velocity higher than the speed of light within this material. This effect is comparable to the sonic boom caused by objects moving faster than sound. Its impact on energy loss is only a fraction compared to ionization, but because of the fact that the cone of light emitted by the particle allows the reconstruction of direction and speed of the particle, it is relevant for a special detector type (see 1.2.2.4). The cone angle is directly dependent on the speed of the particle [11]:

$$\cos \Theta_\mathrm{C} = \frac{c}{n\beta c} = \frac{1}{n\beta} \tag{1.11}$$

- $\Theta_\mathrm{C}$ Cherenkov angle of the cone of light emitted while entering the medium.

- $n$, refraction index of the material

Of course there are other interesting effects that occur while particles are traversing matter, but discussing them would go beyond the scope of this thesis. A more detailed look at this topic delivers [11].

### 1.2.2.2. Tracking detectors

Detectors are built around the interaction point of the accelerator. They cover almost the complete area around the interaction point and are therefore called $4\pi$ detectors. The reason for that is simple: the detectors have to be able to record all particles leaving the interaction point after an event. An important step is the geometrical reconstruction of charged particle tracks which is done by tracking detectors; this allows for the determination of several aspects of the observed event. In combination with the presence of a magnetic field within the detector it is possible to determine the charge of the particle ( $\hat{=}$ sign of curvature) and its momentum ( $\hat{=}$ curvature of the track and its

azimuthal and polar angle). Even single hits allow rough momentum estimation which is coupled to the signal strength recorded by the detector and can be used to recognize hits of very low momentum tracks (see (1.2.2.1) for more details). To get even very low momentum particles, the innermost detector parts are normally tracking detectors. At least as important is the fact that the knowledge of the particle's trajectory allows to reconstruct the location of the production point of the particle (the *decay vertex*). Knowing the position of the decay vertex enables drawing conclusions about the type of the mother particle. More precisely, it allows to reconstruct the lifetime of the mother particle when knowing the type of the daughter particles, the position of their production point and their kinematic parameters (especially their total energy which allows the calculation of the primary particles energy). The importance, its benefits and the concept of the reconstruction process called *event reconstruction* will be discussed in more detail in chapter 1.3.

Reconstructing tracks of particles by absorbing them doesn't make sense, so calorimeters aren't up to that task. On contrary, the measurement of the particle track should not affect the trajectory at all. This is of course impossible, but the design of tracking detectors tries to keep their impact on the particle track as small as possible, but high enough to get detectable signals from the detector. A logical step is to use the ionization effect in gaseous volumes, where the radiation length is very long (or thinking a bit further ahead, thin layers of semiconductors can be used too, but more on that part later).

One of the oldest principles is the cloud chamber, where a supersaturated gas (for example a vapor of water or alcohol) is used to make particle tracks visible. The particle ionizes the atoms while passing them and these ions are working as condensation nuclei in the gas, forming a track of mist within the chamber. Photographing them allows for detailed study of the tracks where the gas has to be desaturated between each shot to erase the old tracks. This technique and its derivatives (for example bubble chambers using superheated liquids) is therefore unfortunately far to slow to reach reaction rates needed for the statistical accuracy of current experiments.

Electronic readout is crucial for computer-aided event processing. For this reason the invention of a concept combining the delivery of space points lying on the trajectory of the track and an electronic readout of that information was awarded with the Nobel prize (Nobel Prize in Physics, 1992 won by Georges Charpak for his invention of the multi-wire proportional chamber). Its principle is based on the technique used in Geiger-Müller tubes. There, an electric field is produced between the conducting walls of the tube (forming the negatively charged cathode) and a wire spanned along the tube axis (working as an anode which is positively charged). They are separated by an isolator coating the ends of the tube sealing it. The potential difference between them is several hundred volts (direct current) and a noble gas is filled into the tube at low pressure. A passing charged particle leaves an ionized track behind, where the ions wander to the respective electrodes due to the electric field within the tube. The field is strong enough to move all primary ions (those produced directly by the passing particle) to the electrodes. In that case, measuring the strength of the signal (caused by the arriving

ions) allows conclusions about the energy of the particle which is proportional to the signal strength (see (1.2.2.1) for more details). Depending on the setup (voltage and used gas), higher voltages are possible which allow for the production of secondary ions (the primary ions get enough energy from the field to produce new ions themselves). Increasing the voltage means increasing the amplification of the signal via secondary ions (but there is still a proportionality between the signal strength and the particle energy) until an effect of saturation occurs (more specifically, Geiger-Müller counters are working in this region of voltage called the *Geiger plateau* where only the number of counts per unit time is relevant, while proportional counters are working in the lower region called the *proportional region*). Therefore it is useful to tune the setup searching for the best combination of voltage and used gas. During the experiment the purity and pressure of the filling gas has to be maintained (because of unavoidable leaks), which is a major matter of expense. An important point is deleting the signal (meaning the arrival of new ions) because of the fact that, as long as there is a signal on the wire, it is blind for new passing particles. To achieve this, a resistor is connected in series with the counter to delete the signal. However, there is still dead time of the device in the range of $\sim$100 nanoseconds in gaseous detectors. This limits the maximum number of possible recorded events per second to the range of several MHz, which is still by several orders higher than former techniques.

Geiger-Müller tubes enable electronic readout, but they have only one wire and therefore no spatial resolution. Considering that downside leads to the development of multi-wire chambers, where several wires are spanned within a conducting case. Combining the spatial resolution and the effect of proportionality between the signal strength and the particle energy leads to the *multi-wire proportional chamber* (MWPC). One of the advantageous features of having two plates and equidistant wires spanned in between them along their plane axis is the amplification effect that occurs near the wires. Looking at the following formula clarifies that:

$$|E(r)| = \frac{1}{r} \cdot \frac{V}{\ln(b/a)} \tag{1.12}$$

- $|E(r)|$ is the strength of the electric field dependent of the distance to the charge (which therefore increases near the wires)

- $V$ is the applied voltage

- $a$ is the radius of the wire (which is typically very small)

- $b$ is the radius of the outer tube

The total voltage is already high enough to move all the ions to the nearest electrodes (proportional region), but this effect alone is not enough to produce signal strengths high enough to detect them. That's where the amplification effect becomes important: the electrons wandering to the anode wires get — due to the stronger field near the wires — enough energy for secondary ionization which results in an avalanche of new ions and produces signals strong enough for detection. Using several layers of these

devices allows two-dimensional reconstruction of the particle trajectory with an accuracy of about 100 micrometers. This can be improved by considering the typical drift velocity of the used setup (voltage of the field and the features of the gas used) while evaluating the signal of the nearest wires. Detectors using this technique are called *drift chambers* (DC) and are among the most common high precision tracking detector types nowadays. A comparative illustration of a MWPC and a DC can be seen in figure 1.4. Three
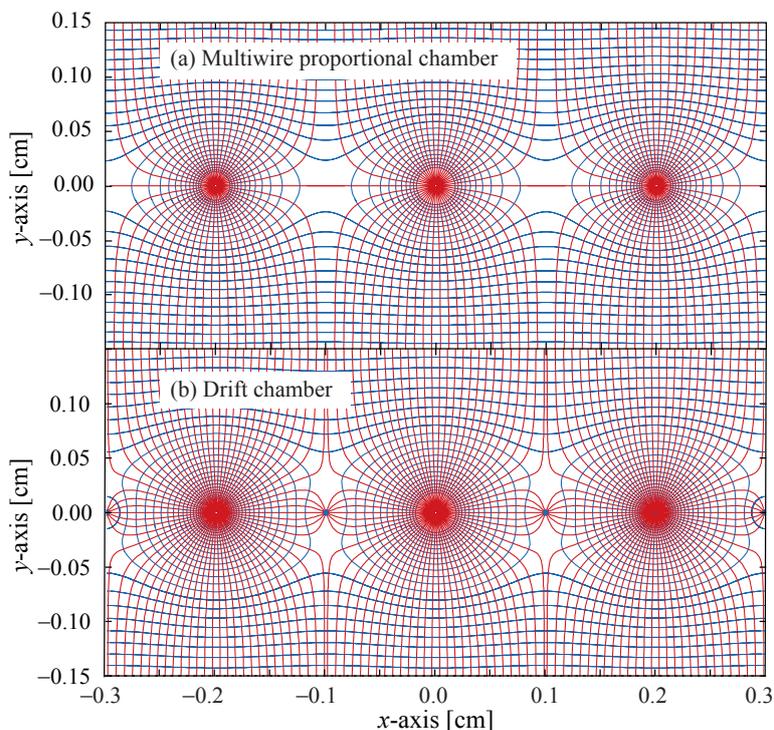


Figure 1.4.: Electric field lines of a MWPC versus DC where cathode wires are added

dimensional reconstruction of the track can be achieved for example by using a *time projection chamber*. This detector uses a big tube filled with gas where the drift chambers (or other tracking detector types) are at the end of the tube. A conducting plate (cathode) orthogonal to the tube is collecting the positive ions while the negative electrons are collected by the two-dimensional end plates. Considering the drift time of the electrons allows reconstruction of the third dimension. Three dimensions and low material budget combined with relatively short dead times makes the time projection chamber an attractive device, but extremely difficult handling during the run of an experiment (maintaining the required purity of the gas used) took care that this device has still not replaced former techniques like the drift chamber although its obvious benefits.

The third dimension (along the beam axis) is the least important for track reconstruction because lifetime measurements mostly need good transversal resolution. This leads to the development of semiconductor detectors which provide higher resolutions but allow for only one- or two-dimensional track reconstruction (due to the higher material budget

and the high production cost, only a small number of layers are used compared to the number of layers within a drift chamber). The most commonly used version is the silicon strip detector. This device is roughly comparable to a single-layer multi-wire chamber, except for the fact that the gaseous volume is replaced by the semiconductor having a depleted zone acting as sensitive matter where ions are produced by passing particles. Strips on the surface of the layer take over the job of the wires. The strips are the anodes and collect the electrons, while the other side works as the cathode. Due to the fact that these strips can be manufactured with smaller distances than wires, higher resolutions can be achieved (in the range of tens of micrometers). But these strip detectors can be upgraded to something akin to two dimensional detectors by using double sided strip detectors (which are much more expensive in production), where the strips on both sides are orthogonal to each other. Reading out the strips deliver intersection points of the strips which can be interpreted as two dimensional space points of the particle. Although these strip detectors are relatively fast (in the range of tens of nanoseconds dead time), they have one major disadvantage: ghost hits. If there is more than one particle passing the sensor module within the same time frame, it is not possible to assign the intersection points reliably to the particles, since the number of intersection points within the layer grows proportional to $n^2$, where $n$ is the number of particles passing. Using smaller detector elements reduces the effect but this is more difficult to handle (readout and mounting of the layers). Other techniques, like decreasing the angle of the double layers to each other instead of using a right angle between them, allows the elimination of ghost hits by placing them outside the physical borders of the detector element. This reduces the number of ghost hits, but decreases the resolution as well.

A way to solve this problem is to build pixel detectors that are roughly comparable to the CCD technique, widely used as photo sensors in digital consumer cameras, where each pixel is read out separately by the readout chip lying directly below the sensor layer, sandwich-style. This allows for a better resolution (in the range of micrometers) and eliminates the problem of ghost hits. But because of the small pixels (compared to long strips of the strip detector), it takes much longer to collect enough electrons for usable signal strengths (in the range of microseconds). Consequently, one readout frame includes the picture of several events, which obviously complicates the track reconstruction. The next problem is the huge amount of information delivered by pixel detectors which results in barely manageable data rates that are not easy to reduce to reasonable amounts of data per second. This makes strip detectors still irreplaceable. For more details about tracking detectors, see [2] or [10].

### 1.2.2.3. Calorimeters

Calorimeters fulfill several important tasks: first, they are needed for the measurement of the total energy of the particle by absorbing it completely. Additionally, they are the only type of detectors which are able to detect neutral particles. Another important feature is their fast response time, which is useful for triggering of events.

There are two subtypes of calorimeters used in particle physics, *electromagnetic calorimeters* and *hadronic calorimeters*. While the first type is used for detection of photons
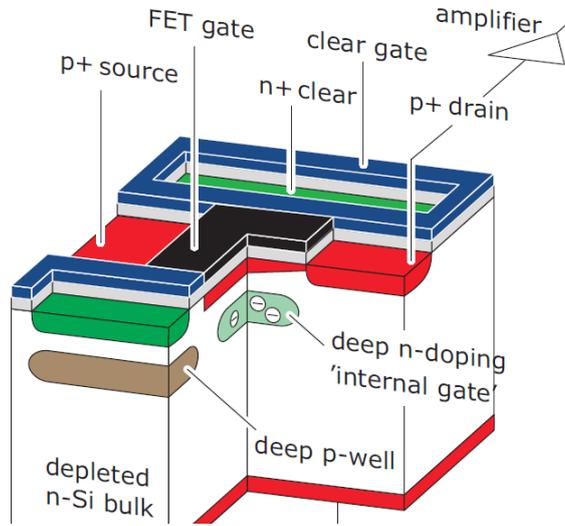
Figure 1.5.: A sketch from a pixel detector using DEPFET technology, provided by the Belle II collaboration

and electrons using electromagnetic interactions between the particle and the detector (electromagnetic shower), the latter type uses the strong interaction for detection of hadrons. An important unit for electromagnetic calorimeters is the *radiation length $X_0$* as mentioned in 1.2.2.1, where the calorimeters are typically of the size of 15-30 $X_0$. $\frac{X_0}{\rho}$ (where $\rho$ is the density of the absorber) is of dimension length in cm. The pendant for hadronic calorimeters is the *characteristic nuclear interaction length $\lambda_I$* which is considerably bigger than $X_0$. Therefore hadronic calorimeters are typically of the depth of 5-8 $\lambda_I$ for economical reasons. Similar to the radiation length the characteristic nuclear interaction length can be written in unit cm when divided by the density of the absorber: $\frac{\lambda_I}{\rho}$. Because of the fact that neutral hadrons don't interact via the electromagnetic force but via the strong force, they can be detected by hadronic calorimeters. More details about this detector type can be fount at [10] (chapter 28.9).

**Electromagnetic calorimeters** are often crystals in which particles produce electromagnetic showers. These showers produce avalanches of low energy photons (visible or ultraviolet light) which can be detected by photodiodes. The amount of emitted light is proportional to the amount of deposited energy. Because of the strong bremsstrahlung effect of electrons (and photons via pair production, where the mean path is proportional to $\ln E_p$) the material can be rather thin, which allows using more expensive material. Such detectors are called *scintillators*. Scintillators have a good energy resolution but no spatial resolution. That's why a big number of smaller scintillators are read out separately to achieve a reasonable spatial resolution. That resolution is needed for the

24

correct assignment of deposited energy to the track reconstructed by other detector types. The radiation length of typical materials used is in the range of centimeters, and only a small number of radiation lengths is needed to stop high energy photons or electrons.

**Hadron calorimeters** are used for the detection of hadrons. Although charged hadrons interact with electromagnetic calorimeters too (neutral hadrons hardly interact with electromagnetic calorimeters as described above), they lose much less energy in EM calorimeters compared to the amount of energy electrons and photons lose within hadronic calorimeters (the bremsstrahlung effect is responsible for the higher energy loss for electrons and photons). That's why a pair of these calorimeter types is always stacked so that the electromagnetic calorimeter comes first (as seen from the particles' point of view). This also means that hadronic calorimeters need much more material to stop hadrons. This makes the ideal materials (as used in the electromagnetic calorimeters) too expensive for reasonable use. There are two ways to solve this problem: the first one uses a cheaper material which has either worse energy resolution or other disadvantages. The second one is to use a sandwich technique with alternating layers of sensitive and stopping material (usually Fe or Cu). The stopping material is only used for reducing the length of the shower by having short interaction lengths. This means that the deposited energy within that matter is not measured and is therefore lost. This of course means careful calibration for the calorimeters to be able to reconstruct the total energy despite having blind layers. A special case is presented by neutrinos, which only interact via the weak force (which has extremely small cross sections and therefore very low probabilities for interaction with matter). This is the reason why they can't be detected this way; one would need huge volumes of detectors for neutrino reconstruction. Consequently, their presence within events can only be determined by the detection of missing energy and missing particles for momentum conservation.

Muons are a special case too: although they interact via the electromagnetic force and therefore can be detected by tracking detectors and and calorimeters, their mass leads to a much lower energy loss within the electromagnetic calorimeters (which use mainly bremsstrahlung and its dependence $\propto \frac{1}{m_{\mathrm{p}}^2}$ for energy loss) compared to electrons. That's why they are normally the only particles which are are able to pass both calorimeters. More about muon detection, see 1.2.2.4.

### 1.2.2.4. Other detector types

**Muon chambers** are needed because of their importance for event reconstruction. Because of their high mass at rest (see 1.2.2.1 for more details) the kinematic energy of muons is not totally absorbed when passing electromagnetic and hadron calorimeters. Muons are quite important for the study of decay channels and therefore important enough to warrant their own detection devices. These detectors are wrapped around the calorimeters. An often-used technique for this are *resistive plate chambers* (RPCs [10]); they are cheap to build, rather fast and can be built with resolutions down to $\sim 250\ \mu\mathrm{m}$. For instance, the muon detector system of the ATLAS experiment (LHC,CERN) uses
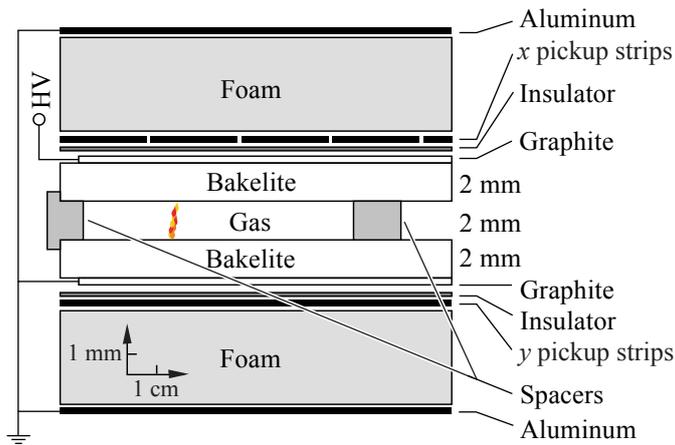
Figure 1.6.: A sketch from a resistive plate chamber

more than 3600 m$^2$ of RPCs with over 300,000 readout channels [12]. This detector type consists of two planar electrodes having a high volume resistivity which are positioned parallel with a distance of the chosen resolution (in the range of several hundred micrometers to several millimeters). The gap between the electrodes is filled with a gas. The efficiency of RPCs (depending on the number of devices connected to a readout channel) is near 100% and their time resolution is below 1 ns (down to tens of ps), which makes them an interesting device for triggering. Between the muon chambers there often are blind layers (see calorimeters). Their first task is to stop other particles which managed to survive while passing through the calorimeters (for example pions). That process is called *punch through*. Their second task is to keep the magnetic flux lines parallel which is called a *return yoke*. Keeping the flux lines parallel is essentially important for tracking since only uniform magnetic fields allow a high precision momentum measurement of the particle.

**Ring imaging Cherenkov detectors** are a newer type of detector. Their first large-scale performance in high energy physics was at the DELPHI experiment at LEP, CERN in the 90's; another very important appearance of that detecting technique was at the *Super-Kamiokande* experiment, which had a leading role in the solution of the solar neutrino problem [13], one of the most obvious signs for physics beyond the Standard model. Their task is an accurate measurement of the particle's velocity. These detectors use the cone angle of the emitted Cherenkov radiation (see equation (1.11)) at the entry of a volume − filled with a liquid or gas − to determine the particle's speed. The refraction index of the volume is known and using nearby tracking detector layers, the entry position of the particle into the Cherenkov detector can be determined too. That determines the tip of the cone. Within the sensitive devices (scintillators coupled with photomultipliers) of the Cherenkov detectors, a ring at the bottom of the cone is detected which allows for the calculation of the Cherenkov angle $\Theta_C$ and using equation (1.11) to

| Detector type | Task | Advantages | Disadvantages |
|---|---|---|---|
| Multi-wire-proportional-chamber | tracking, vertexing, charge and momentum estimation | cheap, easy to handle, fast | only 1D per layer, low resolution |
| Drift chamber | tracking, vertexing, charge and momentum estimation | same advantages as MWPC, better resolution than MWPC | only 1D per layer (measuring 2D at low resolution possible when using stereo readout) |
| Silicon Strip Detector | tracking, vertexing, charge and momentum estimation | high resolution, fast, two dimensional detection via double sided layer technique | ghost hits |
| Silicon Pixel Detector | tracking, vertexing, charge and momentum estimation | very high resolution, two dimensional detection | slow, huge amount of data needs filtering |
| Electromagnetic calorimeter | particle identification, total energy estimation | error shrinks with increasing energy, detects photons too | low resolution (compared to tracking devices), no muon detection |
| Hadronic calorimeter | particle identification, total energy estimation | error shrinks with increasing energy, detects neutrons too | low resolution, expensive, therefore sandwich style |
| Resistive plate chamber | particle identification, tracking, charge and momentum estimation | fast, cheap | low resolution, huge areas needed |
| Ring imaging Cherenkov counter | time-of-flight, speed-estimation | usable for faster particles than TOF by trigger detectors | only possible when particle is faster than light within matter |

Table 1.1.: General overview about different detector types. [2]

finally determine the velocity of the particle. By combining the measurement of velocity and momentum the mass, and hence the type of the particle can be determined, at least in certain momentum ranges.

**Time of flight detectors** are a combination of very fast triggering devices. Fast scintillators or resistive plate chambers are typical tools for this task. The principle is simple: a particle is passing the inner layer of the time of flight detector which is preferably positioned between different types of tracking detectors for reasonable particle assignment. This event delivers the 'tick' signal. After covering some further distance, the particle reaches the second layer of the time of flight detector, which triggers the 'tock' signal. Knowledge of the trajectory between these layers (that's why they are positioned next to tracking detectors) and the amount of time between the triggering events allows the

calculation of the particles velocity.

### 1.2.2.5. Conclusion

There are many different detector types, each with a special task, and together they deliver as complete a picture as possible of the events produced by accelerators. The complete picture of an event includes the total list of participating particles and their kinematic parameters, their tracks, and their production vertices. While obtaining as much information as possible (high detector efficiency due to well designed detectors and their spatial arrangement reducing the number and size of gaps) it is an essential task to keep a high purity of the data (good signal to noise ratio). The cost-efficiency is even more important to keep the total cost of the project in manageable ranges. In the following table, there is a short overview about the detector types and their task:

## 1.3. Event Reconstruction

This section presents an overview about the steps of event reconstruction. A more detailed view on the topic can be found in [8], [14], [15]. The primary goal of current particle physics experiments is to verify/disprove current theories or find hints for new ones. Deviations from current models (especially the Standard Model) used to describe processes in the world of particle physics are quite small and not easy to find. To be able to do that, a huge amount of data from recorded events has to be collected and analyzed to exclude statistical fluctuations and subtract results of interfering effects. In accelerator experiments, a set of detectors records data (hits of passing particles, position, energy deposit, ...) which consists of background tracks coming from beam-gas-interaction, bhabha-scattering or Touschek effect at very high rates (for example, the bunch crossing rate at the Belle II experiment will be 508MHz [5]) and of particle tracks resulting from desired physics events (for example $\Upsilon(4S)$ and $\Upsilon(5S)$ decays resulting in B meson production, whose rate will be about 30kHz at Belle II). Reconstructing what happened in each physics event is called pattern recognition or event reconstruction. The task of event reconstruction is to

- reduce the raw data to processable amounts by suppressing events consisting purely of background effects and enhancing the rate of desired physics events (*triggering*),

- distinguish between real data and background signals within the same event and bundle the hits to track candidates (*track finding*),

- determine the kind of the particle (*particle identification*)

- estimate the kinematic parameters of the partices (*track fitting*), and

- deliver estimates of the locations and parameters of the production points of one or more particles (*vertex finding and fitting*).

Although these are different important steps within the track reconstruction process, they are often combined. For example, track finding already does simple track fitting in order to be able to reconstruct the tracks; particle identification can be also done within track fitting because of type-specific material effects.

Higher demands in the quality of the final dataset ready for analysis and higher luminosity increases the demands on the pattern recognition processes. High speed (especially for on-line reconstruction algorithms) and reliability (to avoid losing important data) are essential demands on the algorithms used.

Track reconstruction can be done on-line as a part of triggering (1.3.1), where simplified algorithms are used for the sake of speed, and off-line where more sophisticated versions can be implemented to get more accurate results.

### 1.3.1. Triggering

More information about triggering can be obtained from [8]. The task of a trigger is to distinguish between good event (interesting physics events) and bad ones (uninteresting physics event or background event). Since uninteresting events occur much more often than interesting ones, a trigger system has to work very fast. Background events can be filtered by setting several conditions which are considering characteristics of such events. Fast tracking algorithms like the conformal mapping or Hough transform (see 3.3 and 3.4 for more details) are used to detect whether tracks are coming from the interaction region or not. Several types of events can be recognized this way, for example showers originating from the vacuum tube or particles from Touschek effect miss the interaction region in in the $r - \phi-$plane and in $z-$direction. The same is valid for cosmic ray background. Beam-gas-interaction on the other hand is located at the IP in the $r - \phi-$plane but miss it in $z-$direction. QED-effects (Bhabha-scattering) are recognized by two colinear tracks pointing at the IP. They can be used to determine the current luminosity since their cross section is well known. In typical trigger setups several detectors are combined to recognize physics events. Especially calorimeters and tracking detectors work together to determine background effects. Without triggering there would be far to much data to store and more sophisticated event reconstruction algorithms would waste much time computing uninteresting events. Therefore is triggering an essential part of the event reconstruction process.

### 1.3.2. Track finding

The task of a track finding algorithm is to determine which hits in a detector belong to a track and which ones can be dismissed as background effects or ghost hits. As described in chapter 1.2 and 1.3.1, several effects originating from the accelerator or from the detectors itself cause interfering signals which have to be recognized. Signal to noise ratios of 1:20, meaning 20 times more background/ghost hits than real hits or worse, have to be dealt with. This means that it is not possible to distill the real hits from the recorded data by checking single hits but only via pattern recognition in several layers of a detector. Knowing the behavior of particles within the detector, therefore, is a crucial

point for recovering the real tracks. To study the particle tracks, several techniques, like using many detector layers, combinations of different detector types, a magnetic field to deflect charged particles (essential for momentum estimation but problematic for track reconstruction) and knowledge of the full geometry during the process of pattern recognition are applied. These techniques allow us to reconstruct the majority of the tracks while suppressing ghost tracks (random combinations of hits which accidentally look like real tracks). This is a main reason why event reconstruction algorithms have to be adapted for each experiment for optimal results. A baseline of such algorithms is the development of a track model which describes the behavior of the particle. Simple versions neglect the third spatial dimension or material effects that cause deviations in the trajectory of the particle (a famous example is conformal mapping [8]). More sophisticated methods consider as much effects as possible (the Kalman filter [8], [16] is a good example for such an algorithm — a recursive implementation of the least squares method). Depending on the situation at the experiment (event rate, on-line or off-line track finding, number of expected tracks per event, level of expected background, ...), a trade-off between cost (in CPU-time) and efficiency has to be found. Further information about track finding can be found in chapter 3 and 4.

In the end the track finding process delivers rough track candidates for further use. Not all of them are real tracks or even pure tracks (meaning that all hits assigned to the track candidate belong to the same real track). Separating good from bad track candidates needs more knowledge about the parameters of the track. These are calculated by track fitting algorithms. Chapter 3 is covering that topic with more detail applied at the case of low momentum track finding in Belle II.

### 1.3.3. Track fitting

As mentioned in the previous section, an algorithm optimized for fast track recovery tends to produce contaminated track candidates (means the majority of the assigned hits are of the same track, but not all of them) or even ghost track candidates in order to lose as few real tracks as possible. Especially online reconstruction algorithms, where the time constraint is quite narrow, tend to reconstruct fake tracks consisting of a random combination of ghost or background hits at that point. Therefore it is a logical step to use specialized track fitting tools to separate the good track candidates from the bad ones. In their mode of operation they are much more complex and therefore need the track finding algorithms to reduce the combinatorial problem which would increase enormously when not prefiltering the information provided by the tracking detectors. The process of track finding and track fitting are not always separated (in fact they are combined quite often, especially for off-line purposes), where on-line finders often use quite simple fitting algorithms and off-line tools can use more complex ones. For the track fitting process, a crucial point is obtaining the track parameters (defined by the track model used), which contain the full information about the particle and its path through the detector. To store the full information available, five different parameters have to be reconstructed at the vertex from where the particle track can be followed through the tracking detectors. Some algorithms use more than five parameters for internal usability reasons, where some
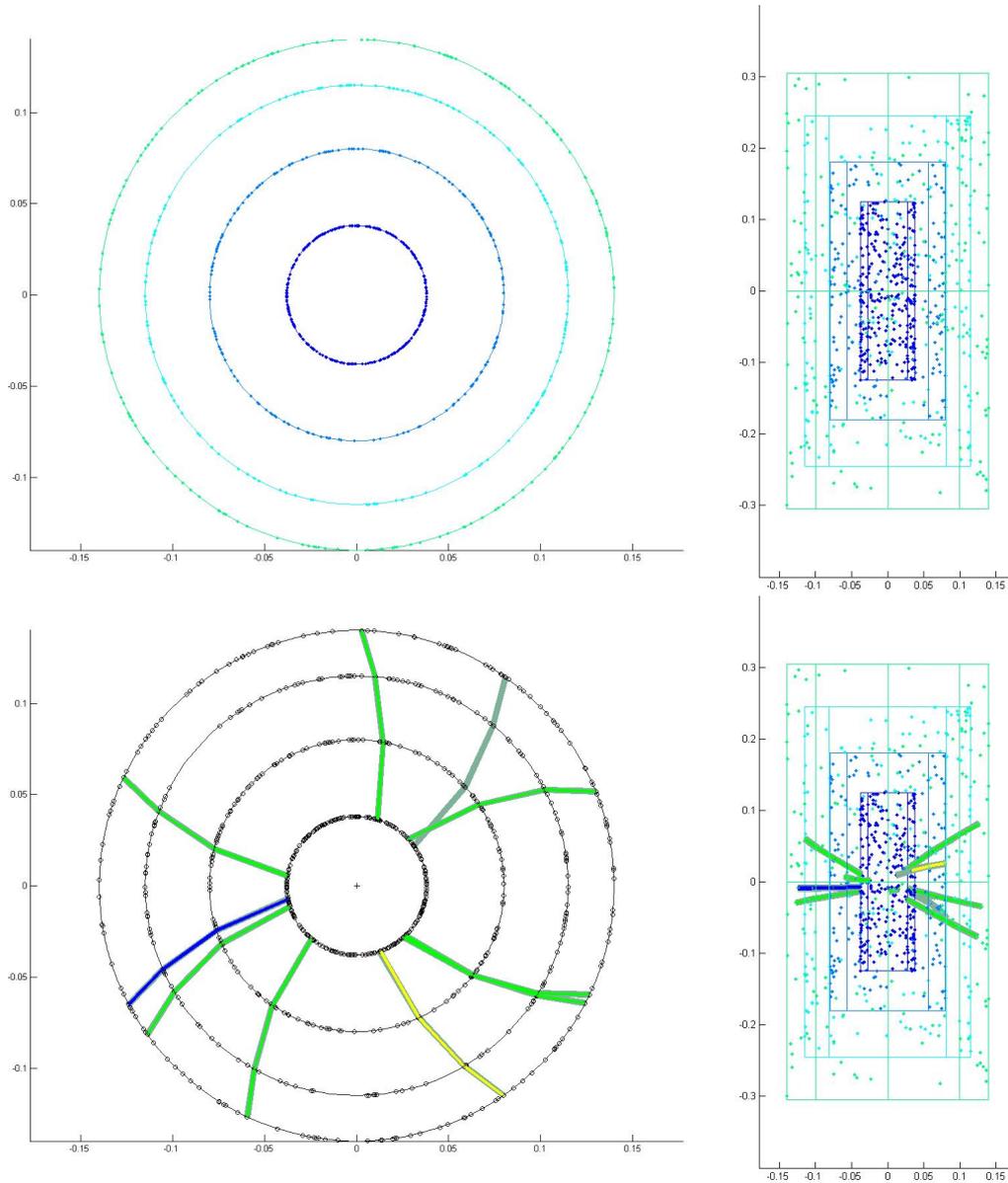
Figure 1.7.: Images of a simplified 4 layer detector geometry showing an event with 10 tracks at a SNR of 1:20. a) is the $x - y-$view, b) the $r - z-$view without reconstructed tracks, c) and d) are the same event with tracks reconstructed. Blue and green tracks are recognized correctly, yellow ones are recognized partially and turquoise ones are lost

of the parameters are partially redundant and therefore can be reduced to five parameters without losing real information. There are many conventions encoding that information, one possible way is:

- $\Phi$ is the azimuth angle of a point on the track

- $z$, is the Cartesian coordinate of the particle parallel to the beam.

- $\beta = \phi - \Phi$, which is the difference of the angle $\Phi$ and the angle $\phi$ which is the azimuthal angle of the momentum vector at the same point on the track,

- $\theta$ is the polar angle between the momentum vector and the $z$-axis. It includes information about the non-transverse part of the momentum,

- the curvature of the projection of the track on the $x - y$-plane is ($\kappa = \frac{q}{p_T}$) which includes information about the charge of the particle $q$ and its transverse momentum $p_T$ (meaning the momentum within in the $r - \Phi-$plane),

Track fitting algorithms aim to determine these parameters with precisions as high as possible (considerably more than the intrinsic resolution of a single hit). To get a higher resolution of the parameters than the resolution of a single hit, the information of the whole track is combined. Information of the other hits refines the information of the current hit. Better knowledge of these parameters increases the probability to recognize ghost tracks. That step works closely together with another step at pattern recognition, the particle identification (PID).

### 1.3.4. Particle identification

Knowing the type of the particle allows for refinement of the parameter estimations because of the ability to consider type specific behaviors for example material effects (see 1.2.2.1 for more details). Although there are a many (elementary or hadronic) particles, only a small number lives long enough to be able to be detected directly. Those particles are e, ē, μ, μ̄, p, p̄, π, π̄, $K^+$, $K^-$. Neutral particles like neutrons, photons and $K_L^0$ can be detected as well, but only with calorimeters, and therefore no direct tracking is possible. Indirectly there is still some information which can be obtained by finding their place of birth and place of decay. The latter is simply the shower in the calorimeters. The place of birth is the interaction point (often called primary vertex) or a secondary vertex (see 1.3.5 for more details). Neutral particles are not deflected by a magnetic field and therefore their track is an almost straight line (only disturbed by interaction with matter) which can be drawn from the vertex to the calorimeters. But this is already indicating one of the main techniques of particle identification:

*PID by difference in interaction.* As mentioned in 1.2.2.1 and 1.2.2.4, muons are easy to identify simply by the muon system. Photons are only visible in electromagnetic calorimeters, which allows for the recognition of electrons too, by considering the knowledge of their track (which is visible for electrons but not for photons). To determine whether the

current track belongs to a particle or its antiparticle, the determination of the charge is sufficient. This reduces the list above to neutrons and neutral kaons which interact with the calorimeters and another bunch, the protons, charged kaons and pions, which leave a track within the tracking system and are stopped in the calorimeter system. To be able to distinguish them, another step is necessary:

*PID by difference in mass.* Because of the fact that each of those particles mentioned above have their unique mass at rest, it is possible to use that information for particle identification. Unfortunately, the mass at rest cannot be measured directly and therefore has to be reconstructed. To get the mass at rest, looking at the following formula helps to find the solution:

$$p = \gamma m_{\text{rest}} v \;\; \rightarrow \;\; m_{\text{rest}} = \frac{p}{c\beta\gamma} \tag{1.13}$$

The momentum is easily reconstructed: the radius of the helix of the particle of interest at given transverse momentum ($x - y$-plane) is only dependent of the strength of the magnetic field, the momentum in $r - z$-direction is determined by the angle $\theta$ of the particle defined in 1.3.3. The velocity of the particle can be reconstructed in different ways, like time-of-flight measurements, velocity determination using ring imaging Cherenkov detectors (see 1.2.2.4) or considering the energy deposit and its dependency of the velocity of the particle.

More details about particle identification can be found for example in [14].

## 1.3.5. Vertex finding and fitting

As mentioned above, only a small number of particles survive long enough to reach the sensitive detector parts. All of them are already known for at least 50 years and their only task for us is to tell us everything about their 'birth' at a vertex. The vertex is a point in space where short-lived particles decay into other particles. Every particle decays in particular decay channels (with different probabilities) which allows us to recover the identity of the mother particle. But not only the input and outcome of the decay is important, the position in space is important as well — it tells us how long the mother particle lived until its decay (the distance of flight is directly proportional to the lifetime of the particle). An experiment that wants to find evidence for CP violation (e.g. expressed in different lifetimes of particles and their antiparticles) relies on the quality of the reconstructed vertices. Next to the identity of the mother particle, the decay vertex tells us its parameters. Using a huge number of events containing records of the same decay channel allows the determination of the decay-probability of a decay channel.

Vertex finding is done by combining the tracks of several particles and calculating intersection points. If two or more tracks share an interaction point (within the limits of errors), they probably share a production vertex and are the result of the same decay. Knowing the daughter particles allows the determination of the decay channel and the identification of the mother particle. This is possible by using the rules for quantum number conservation (for example conservation of the lepton generation number). Vertex fitting is done by combining the parameters of all tracks of the vertex. This increases the quality of the vertex-parameter-estimation in the same way as the resolution of a

single point of the particle-track can be refined by using the total information of all hits assigned to that track.

# 2. The Belle and Belle II experiments

## 2.1. Physics at the Belle experiments

One of the big questions of mankind is "*Where do we come from?*", and different disciplines try to answer it from a different point of view. The beginning of the universe is therefore a well debated field especially for physicists. One of the aspects not thoroughly understood so far is the fact that with the big bang marking the beginning of our universe a bigger asymmetry between matter and antimatter than initially suggested has to exist, as matter and antimatter annihilate each other to photons but a small fraction of the matter survived the annihilation process. Antimatter is not exactly the mirror of matter but behaves differently in special cases (ergo the transformation in $C$ and $P$ is violated). This *CP Violation* can be described by the Standard Model of particle physics but effects included in the Standard model are not sufficient to explain the observed baryon asymmetry.
The Belle experiments are two in a long row of experiments studying the phenomenon of $CP$ violation, where the Belle experiments concentrate on the $CP$ Violation in the B meson system. This section gives an introduction to the topic of $CP$ violation. More detailed informations can be found for example in [17] and [18].

### 2.1.1. The Standard Model of Particle Physics in a nutshell

In the $20^{\text{th}}$ century a revolution in physics has started after the discovery of quantum mechanical effects. During the $19^{\text{th}}$ century, there were only two fundamental forces known, the gravitational and the electromagnetic force. Two new forces had to be introduced to be able to explain the stability of the positively charged and therefore self-repelling atomic core (strong force) or effects like the $\beta$ decay (weak force). It is possible to calculate coupling constants for each fundamental force in order to be able to compare their relative strength. Normalizing the strong force to 1, the relative strength of the electromagnetic force is $\sim 10^{-2}$ compared to the strong force. The weak force has a relative strength of $\sim 10^{-5}$, and the gravitational force is weak enough ($\sim 10^{-38}$) to be neglected within the Standard Model. The forces are mediated by gauge bosons having spin 1 and hypothetically the graviton, which is a gauge boson with spin 2. This means that each interaction is performed by the exchange of a gauge boson of the current force. Photons and the hypothetic gravitons of the electromagnetic and the gravitational forces, respectively, have no mass and therefore their forces have infinite range. Although the 8 gluons carrying the strong force are massless too, the strength of the strong interaction increases with distance leading to a limited range (confinement). This is the reason why single quarks can't be found in nature. The weak force is the only one having gauge bosons with mass $\neq 0$. This leads to a very short range of this interaction in order to

|  | Quarks | charges | Interactions |
|---|---|---|---|
| up-type | u,c, t | $+\frac{2}{3}$e | strong, weak, electro- |
| masses (MeV) | 1.5-3.3, 1270, 171200 | | magnetic |
| down-type | d, s, b | $-\frac{1}{3}$e | strong, weak, electro- |
| masses (MeV) | 3.5-6, 104, 4200 | | magnetic |
|  | Leptons | charges | Interactions |
| charged | e, μ, τ | $-1$e | weak, electromagnetic |
| masses (MeV) | 0.511, 106, 1780 | | |
| uncharged | $\nu_e, \nu_\mu, \nu_\tau$ | 0 e | weak |
| masses (eV) | $< 10^{-6}, < 10^{-6}$, ? | | |
|  | Gauge bosons | charges | Symmetry |
| electromagnetic force | $\gamma$ | 0 e | U(1), QED |
| mass (eV) | 0 | | |
| weak force | W$^\pm$, $Z^0$ | $\pm 1, 0$ e | SU(2), GSW |
| masses (GeV) | 80.4, 91.2 | | |
| strong force | 8 g's | 0 e | SU(3), QCD |
| masses (eV) | 0 | | |

Table 2.1.: General overview about quarks, leptons and gauge bosons. Data taken from [3] and [4]

conserve the uncertainty principle. The weak force has two gauge bosons carrying an electromagnetic charge. This causes the possibility that an up-type quark can change its flavor into a down-type quark and vice versa.

There are 6 different quarks and 6 different leptons grouped into 3 generations. However only the lightest generation is forming the matter of the universe Each quark and each lepton has an anti matter counterpart with the same properties but an opposite charge. Their spin is 1/2, and therefore they are fermions. All of them obey the weak force, but only the charged particles obey the electromagnetic force, leaving the neutrinos which only interact via the weak force. The strong force affects only quarks which carry a color charge. These charges (red, green, blue) add up to a color singlet state to form bound states. Although there are many way to get such a color singlet state (glueballs, tetra- and pentaquarks, etc.) there are only two observed ways to get a white state, one by using two quarks (color and anti-color) forming a meson and the second one by using three quarks forming a baryon.

An overview of the elementary particles and the mediating gauge bosons can be found in Table 2.1. The following list provides the names of the quarks and leptons:

- the quarks of first generation are u (up quark) and d (down quark)

- the quarks of second generation are c (charm quark) and s (strange quark)

- the quarks of third generation are t (top quark) and b (bottom or beauty quark)

- the leptons of first generation are e (electron) and $\nu_e$ (electron neutrino)

- the leptons of second generation are $\mu$ (muon) and $\nu_\mu$ (muon neutrino)

- the leptons of third generation are $\tau$ (tauon) and $\nu_\tau$ (tauon neutrino)

### 2.1.2. A closer look at *CP* Violation

#### 2.1.2.1. The CKM matrix

After the discovery of antimatter in the 1930's it was assumed that there is a symmetry in $C$ and $P$. $C$ stands for the charge conjugation transformation which exchanges a particle with its antiparticle (same attributes but different charge), while $P$ stands for parity transformation and changes the sign of the spatial coordinates:

$$\boldsymbol{x} = (t, \boldsymbol{a}) \quad \xrightarrow{P} \quad \boldsymbol{x}_P = (t, -\boldsymbol{a}) \tag{2.1}$$

The third important operation in that context is the time reversal, denoted by $T$, changes the sign of the time coordinate:

$$\boldsymbol{x} = (t, \boldsymbol{a}) \quad \xrightarrow{T} \quad \boldsymbol{x}_T = (-t, \boldsymbol{a}) \tag{2.2}$$

$CP$ symmetry says that the laws of physics should not change when performing a transformation in $C$ and $P$. While strong and electro-magnetic interactions do not violate this invariance, there exists a $CP$ violation in weak interactions (discovered 1964 in the $K^0$ meson decay [19], Nobel prize 1980) but no violation of the total combination of $CPT$, which is still an exact symmetry in any local Lagrangian field theory. The weak interaction is the only one where a quark or a lepton can change its flavor. This mechanism is needed for example for neutrino oscillations where neutrinos can change their generation or the beta decay where a down quark in a neutron changes its flavor into an up quark (the neutron becomes a proton), and emits an electron and an electron antineutrino. As mentioned in 2.1.1, the Standard Model defines three generations of quarks where each of them has got its own antiparticle:

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix} \tag{2.3}$$

Right handed fermions transform differently under the gauge group $SU(2)$ than left handed ones. On that account only left handed particles and right handed antiparticles undergo weak interaction. The particle states (d, s, b) described by (2.3) are the mass eigenstates of these particles. However, the weak eigenstates $(d', s', b')$ participating in the weak interaction are slightly different:

$$\begin{pmatrix} u \\ d' \end{pmatrix} \quad \begin{pmatrix} c \\ s' \end{pmatrix} \quad \begin{pmatrix} t \\ b' \end{pmatrix} \tag{2.4}$$

37

and are linear combinations of the mass Eigenstates.

$$|\text{d}'\rangle = V_{\text{ud}}|\text{d}\rangle + V_{\text{us}}|\text{s}\rangle + V_{\text{ub}}|\text{b}\rangle \tag{2.5}$$

$$|\text{s}'\rangle = V_{\text{cd}}|\text{d}\rangle + V_{\text{cs}}|\text{s}\rangle + V_{\text{cb}}|\text{b}\rangle \tag{2.6}$$

$$|\text{b}'\rangle = V_{\text{td}}|\text{d}\rangle + V_{\text{ts}}|\text{s}\rangle + V_{\text{tb}}|\text{b}\rangle \tag{2.7}$$

The parameters $V_{ij}$ describe the probability for a quark $j$ changing into a quark $i$, and every triplet obeys the normalization condition:

$$|V_{id}|^2| + |V_{is}|^2 + |V_{ib}|^2 = 1, \text{ where } i \in \{\text{u}, \text{c}, \text{t}\} \tag{2.8}$$

Therefore, in the weak interaction, the up-type quark form doublets with a superposition of all down-type quarks. These weak eigenstates are rotated by a $3 \times 3$ unitary matrix which is known as the Cabibbo-Kobayashi-Masukawa matrix $V_{\text{CKM}}$ and combines the equations from (2.5), (2.6) and (2.7) into one formula using matrix notation:

$$\begin{pmatrix} \text{d}' \\ \text{s}' \\ \text{b}' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} \text{d} \\ \text{s} \\ \text{b} \end{pmatrix} = \begin{pmatrix} V_{\text{ud}} & V_{\text{us}} & V_{\text{ub}} \\ V_{\text{cd}} & V_{\text{cs}} & V_{\text{cb}} \\ V_{\text{td}} & V_{\text{ts}} & V_{\text{tb}} \end{pmatrix} \begin{pmatrix} \text{d} \\ \text{s} \\ \text{b} \end{pmatrix} \tag{2.9}$$

This matrix was presented first in 1973 by Kobayashi and Masukawa [20] to explain the $CP$ violation in the Standard Model. At that time only four different quarks (u,d,s, c) were known, where the charm quark was proposed only some years ago by Cabbibo. The CKM concept was an expansion of a former concept developed by Cabbibo [21] and introduced a third generation of quarks postulating the existence of six different quarks (which is today well confirmed by various experiments), and allows to introduce a complex phase in that quark mixing process. The CKM matrix has got 9 different parameters which can be reduced to 4 independent parameters by re-phasing the 6 quark fields. Three of them are real parameters and correspond to rotation angles among different generations. The fourth one is complex and the source of the $CP$ violation explained by the Standard Model. To get a useful parameterization for the CKM matrix, the unitarity of the matrix can be used which implies various relations among its elements. The following three are especially interesting for the understanding of the $CP$ violation:

$$V_{\text{ud}}V_{\text{us}}^* + V_{\text{cd}}V_{\text{cs}}^* + V_{\text{td}}V_{\text{ts}}^* = 0 \tag{2.10}$$

$$V_{\text{us}}V_{\text{ub}}^* + V_{\text{cs}}V_{\text{cb}}^* + V_{\text{ts}}V_{\text{tb}}^* = 0 \tag{2.11}$$

$$V_{\text{ud}}V_{\text{ub}}^* + V_{\text{cd}}V_{\text{cb}}^* + V_{\text{td}}V_{\text{tb}}^* = 0 \tag{2.12}$$

Each of these relations requires the sum of three complex quantities to vanish. This allows a geometrical representation in the complex plane as a triangle. Although these relations deliver three triangles (the first one is connected to K decays, the second one to $B_s$ decays), only the triangle of (2.12) is called the "Unitary Triangle". The reason for this is the fact that within the first triangles the $CP$ violating effect is rather small and therefore the triangles almost collapse to a line. The third triangle provides a significant

$CP$ violation which makes it more interesting for the studies of $CP$ violating effects. Choosing a phase convention such that $(V_{cd}V_{cb}^*)$ is real and rescaling the triangle by dividing the sides by $|V_{cd}V_{cb}^*|$ which normalizes the side along the real axis to one deliver the rescaled Unitary Triangle. Two of the vertices are fixed at the coordinates (0,0) and (1,0), the third one is denoted by $(\rho, \eta)$. This leads to the Wolfenstein parameterization of the CKM matrix [22] using the 4 real parameters $(A, \lambda, \rho, \eta)$:

$$
V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4)
\qquad (2.13)
$$

- $\rho, \eta$ are the coordinates of the third vertex of the rescaled Unitary Triangle

- $\lambda = |V_{us}| \simeq 0.23$ is an expansion parameter

- $A = |V_{cb}| \simeq 4 \times 10^{-2}$

This parameterization can be improved by setting $|V_{cd}V_{cb}^*| = A\lambda^3$ ($V_{cd}V_{cb}^*$ is real up to $\mathcal{O}(\lambda^7)$). This leads to a modified apex $(\overline{\rho}, \overline{\eta})$, where

$$
\overline{\rho} = \rho\left(1 - \frac{\lambda^2}{2}\right), \quad \overline{\eta} = \eta\left(1 - \frac{\lambda^2}{2}\right)
\qquad (2.14)
$$

These parameters can be used to get the lengths $(R_1, R_2)$ of the two complex sides of the Unitary Triangle:

$$
R_1 \equiv \sqrt{\overline{\rho}^2 + \overline{\eta}^2} = \frac{1 - \frac{\lambda^2}{2}}{\lambda}\left|\frac{V_{ub}}{V_{cb}}\right|, \quad R_2 \equiv \sqrt{(1 - \overline{\rho})^2 + \overline{\eta}^2} = \lambda\left|\frac{V_{td}}{V_{cb}}\right|
\qquad (2.15)
$$

Finally the three angles of the Unitary Triangle (see figure 3.2) representing the complex phase of the combinations are denoted by:

$$
\phi_1 = \beta = \arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right], \quad \phi_2 = \alpha = \arg\left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right], \quad \phi_3 = \gamma = \arg\left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right]
\quad (2.16)
$$

As shown in (2.16), these angles have got different names in literature.

To understand how to get these angles and how such a concept can be examined by an experiment like the Belle experiments, the following chapter presents a closer look at the $CP$ violation in the B meson system which was predicted by the CKM matrix.

### 2.1.2.2. *CP* Violation in the B meson system

Interesting for the study of $CP$ violation are neutral self conjugate pairs of mesons and the following two involve beauty quarks:

$$
\text{B}^0 = \overline{\text{b}}\text{d}, \; \overline{\text{B}}^0 = \overline{\text{d}}\text{b}
\qquad (2.17)
$$
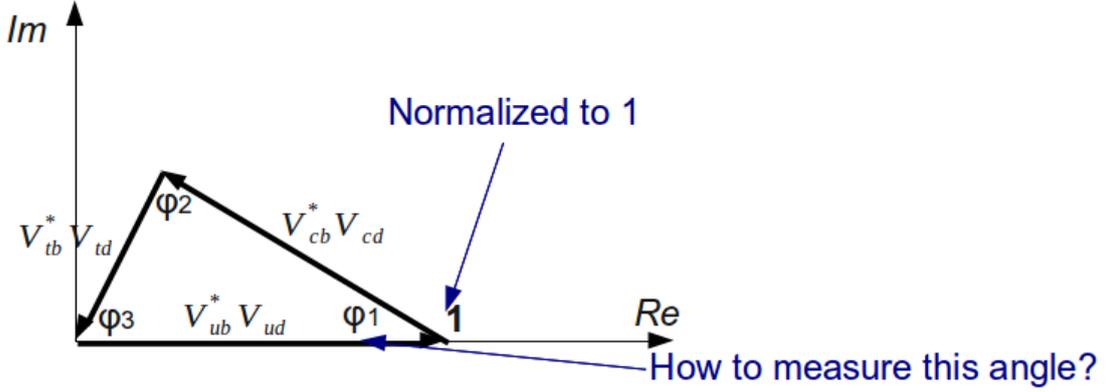
Figure 2.1.: The CKM matrix condition (2.12) presented as a triangle whose hypotenuse is normed to one and transformed to be real

This involves the element of the CKM matrix

$$\mathrm{B_s} = \bar{\mathrm{b}}\mathrm{s}, \ \overline{\mathrm{B}}_\mathrm{s} = \bar{\mathrm{s}}\mathrm{b} \qquad (2.18)$$

These are related to $|V_{td}|$ ((2.17)) and $|V_{ts}|$ ((2.18)), respectively. There is a similar system in the K meson system where $K^0$ contains the $\bar{\mathrm{s}}$ quark, but discussing this would go beyond the scope of this thesis. The D meson system has a decay rate much faster than its mixing rate and therefore the flavor eigenstates are a convenient basis.
Because of the quark mixing, these mesons change their state when propagating through space. The following discussion uses $B^0$ but is working for both B systems. The state $|B^0(t)\rangle$ at time $t$ is a superposition of its two flavor eigenstates:

$$|B^0(t)\rangle = a|B^0\rangle + b|\overline{B}^0\rangle \qquad (2.19)$$

where the time evolution can be described by a time-dependent Schrödinger equation:

$$i\frac{d}{dt}\begin{pmatrix} a \\ b \end{pmatrix} \equiv i\frac{d}{dt}|B^0(t)\rangle = (M - \frac{i}{2}\Gamma)\,|B^0(t)\rangle \qquad (2.20)$$

$M$ and $\Gamma$ are 2 $\times$2 hermitian matrices, where $M$ denotes mass and $\Gamma$ denotes decay width. While $CPT$ invariance guarantees that the diagonal elements are equal ($M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$), the off-diagonal terms are important for $CP$ violation. The off-diagonal elements $M_{12}$ and $M_{21}$ represent the dispersive part and the elements $\Gamma_{12}/2$ and $\Gamma_{21}/2$ the absorptive part of the transition amplitude from $B^0$ to $\overline{B}^0$. The elements of these matrices are $\in \mathbb{C}$. To calculate them within the Standard Model, box diagrams with two W exchanges are used ([18]). The mass eigenstates are called $B_L$ (light) and $B_H$ (heavy) and given by:

$$|B_L\rangle = p|B^0\rangle + q|\overline{B}^0\rangle \qquad (2.21)$$

$$|B_H\rangle = p|B^0\rangle - q|\overline{B}^0\rangle \qquad (2.22)$$

where the coefficients $p$ and $q$ are complex and obey the normalization condition:

$$|q|^2 + |p|^2 = 1 \tag{2.23}$$

Considering the normalization condition $p$ and $q$ can be obtained by solving:

$$\frac{q}{p} = +\sqrt{\frac{M_{12}^* - i\Gamma_{12}^*/2}{M_{12} - i\Gamma_{12}/2}} \tag{2.24}$$

To get the ratio of $q/p$, the *mass difference* $\Delta M_{\mathrm{B}}$ and the *width difference* $\Delta\Gamma_{\mathrm{B}}$ are important. They are defined as $\Delta M_{\mathrm{B}} \equiv M_H - M_L$, which is positive by definition and also defines the $\mathrm{B}^0 - \overline{\mathrm{B}}^0$ mixing frequency, and $\Delta\Gamma_{\mathrm{B}} \equiv \Gamma_H - \Gamma_L$ which is the difference in lifetime. The difference in width in the $\mathrm{B_d}$ system has to be obtained by experiment is rather small and produced by the common decay channels to $\mathrm{B}^0$ and $\overline{\mathrm{B}}^0$ (for $\mathrm{B_s}$ it is more significant). This results in the fact that the element $\Gamma_{12}$ is much smaller than $M_{12}$ which implies that $\Delta\Gamma_{\mathrm{B}} \ll \Delta M_{\mathrm{B}}$. This relation reduces (2.24) for the $\mathrm{B}^0$ system to a much simpler equation:

$$\left(\frac{q}{p}\right)_d = \mathrm{e}^{-2i\phi_1} \tag{2.25}$$

Here, $2\phi_1$ is the CP-ciolating phase difference between $M_{12}$ and $\Gamma_{12}$ and $\phi_1$ itself is part of a graphical interpretation of the CKM-matrix (see figure 3.2). More about the CKM-matrix and its experimental examination can be found in chapter 2.1.2.3.

After the decay of $\mathrm{b\overline{b}}$ meson ($\Upsilon$), the $\mathrm{B}^0$ and $\overline{\mathrm{B}}^0$ mesons produced from that decay are in a coherent $L = 1$ state. This means when they evolve in time and oscillate between the $\mathrm{B}^0$ and $\overline{\mathrm{B}}^0$ states with the frequency $\Delta M$. However this process is in phase, which means that at any time, there is always one $\mathrm{B}^0$ and one $\overline{\mathrm{B}}^0$ present. Once one of them decays, the other continues to evolve and therefore it is possible to get decays with two $\mathrm{B}^0$'s or $\overline{\mathrm{B}}^0$'s. The decay rate of a B meson to a final state $f$ is time dependent ($\Gamma(\mathrm{B}^0(t) \to f)$) because of the fact that the state of the meson is time dependent itself (as mentioned in (2.19)). The decay amplitude of flavor Eigenstates are written as $A_f = \langle f|\mathrm{B}^0\rangle$ (which is the amplitude for a $\mathrm{B}^0$ to decay in the state $f$) and $\overline{A}_f = \langle f|\overline{\mathrm{B}}^0\rangle$ (which is the amplitude for a $\overline{\mathrm{B}}^0$ to decay in the *same* state $f$). This allows to define a parameter $\lambda$ for the state $f$:

$$\lambda_f = \frac{q}{p}\frac{\overline{A}_f}{A_f} \tag{2.26}$$

Next to the $CP$ violation important types of $CP$ violation in B decays are defined as:

$$\left|\frac{\overline{A}_{\overline{f}}}{A_f}\right| \neq 1 \implies CP \text{ violation} \tag{2.27}$$

Which describes $CP$ violation in decay $-$ often called *direct CP violation* $-$ and occurs with both charged and neutral decays, when the amplitude for a decay and its $CP$ conjugate process have different magnitudes.

$$\lambda \neq \pm 1 \implies CP \text{ violation} \tag{2.28}$$

Which describes $CP$ violation in mixing $-$ often called *indirect CP violation* $-$ and occurs when the two neutral mass eigenstates cannot be chosen to be $CP$ eigenstates (interference between decays with and without mixing). Another way to describe these two types of $CP$ violation is via the time dependent asymmetry $a_f(t)$ between $\mathrm{B}^0$ and $\overline{\mathrm{B}}^0$ decays which is defined as follows:

$$a_f(t) = \frac{\Gamma(\overline{\mathrm{B}}^0(t) \to f) - \Gamma(\mathrm{B}^0(t) \to f)}{\Gamma(\overline{\mathrm{B}}^0(t) \to f) + \Gamma(\mathrm{B}^0(t) \to f)} \tag{2.29}$$

Is the final state $f$ a $CP$-eigenstate $CP|f\rangle = \xi_f|f\rangle$ with an eigenvalue $\xi_f = \pm 1$, the asymmetry $a_f(t)$ becomes:

$$a_f(t) = \mathcal{A}_f \cos(\Delta M t) + \mathcal{S}_f \sin(\Delta M t) \tag{2.30}$$

With

$$\mathcal{A}_f = \frac{|\lambda_f|^2 - 1}{|\lambda_f|^2 + 1} \tag{2.31}$$

being the direct $CP$ asymmetry requiring the condition mentioned in (2.27) for not being zero, and

$$\mathcal{S}_f = \frac{2\,\mathrm{Im}\lambda_f}{|\lambda_f|^2 + 1} \tag{2.32}$$

being the indirect $CP$ asymmetry which requires (2.28).

After covering the theory of the $CP$ violation in the previous chapters, the next chapter will discuss its application within the Belle experiments.

### 2.1.2.3. *CP* Violation in the Belle experiments

This chapter provides an overview of how to retrieve information regarding the $CP$ violation with an experiment, whereas a deeper view into this topic can be found in [17] and [18]. The indirect $CP$ violation as mentioned in (2.28) is important for methods to measure the angles of the Unitary Triangle (see equation (2.16) and figure 3.2. In this section some of them are listed to summarize the issue. The figures 2.2 and 2.3 show the improvement of the accuracy of the angle measurement over the years.

**The measurement of $\phi_1$** can be performed by using the best example known so far, where the final state $f$ is $\mathrm{J}/\psi\mathrm{K}_s$ [23]. In figure 2.4 the result of the measurement of this angle is shown.

The quark level process of that decay channel is $\overline{\mathrm{b}} \to \overline{\mathrm{c}}\mathrm{c}\overline{\mathrm{s}}$ followed by the $\mathrm{K}^0 - \overline{\mathrm{K}}^0$ mixing. If the decay is dominated by a single amplitude and therefore complicating effects penguin amplitude contributions can be neglected. In that case the ratio of the decay amplitudes is given by:

$$\frac{\overline{A}_{\mathrm{J}/\psi\mathrm{K}_s}}{A_{\mathrm{J}/\psi\mathrm{K}_s}} = -\left(\frac{V_{\mathrm{cb}}V_{\mathrm{cs}}^*}{V_{\mathrm{cb}}^*V_{\mathrm{cs}}}\right)\left(\frac{V_{\mathrm{cs}}V_{\mathrm{cd}}^*}{V_{\mathrm{cs}}^*V_{\mathrm{cd}}}\right) \tag{2.33}$$
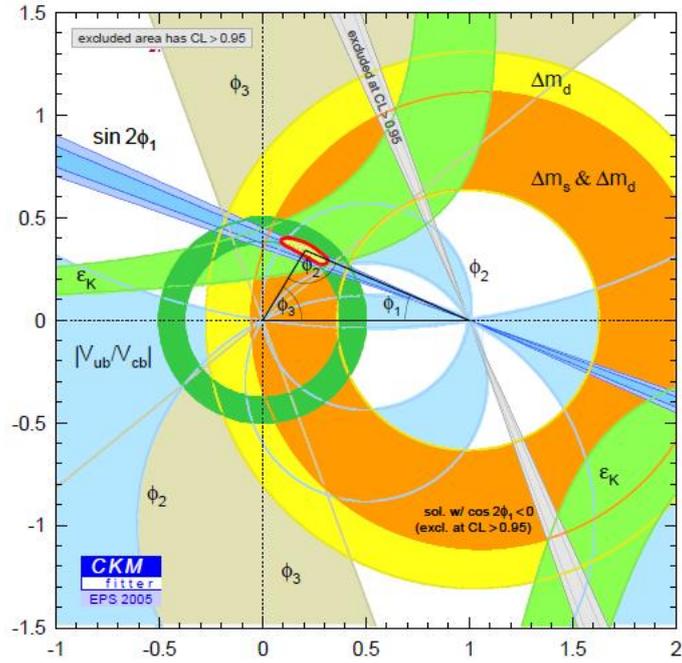
Figure 2.2.: the accuracy of the measurement of the CKM triangle - 2005
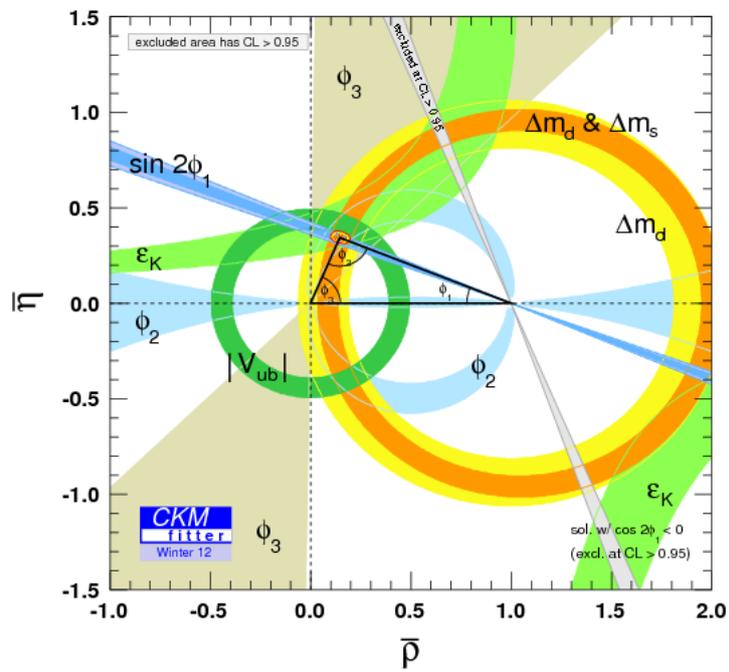


Figure 2.3.: the accuracy of the measurement of the CKM triangle - 2012

43
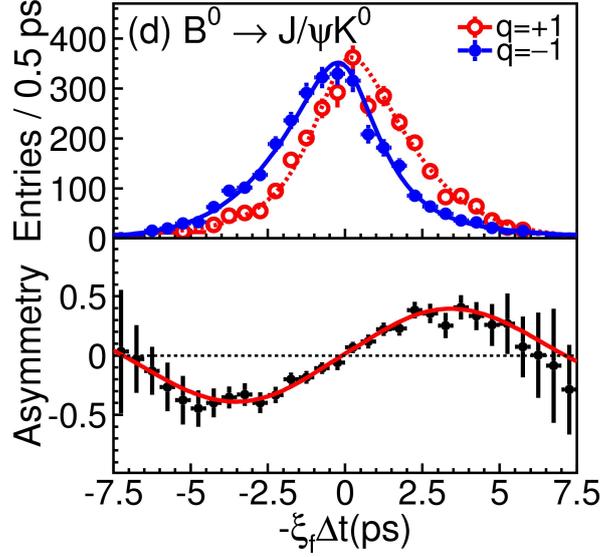
Figure 2.4.: The difference in lifetime between $B^0$ and its antiparticle, this figure is provided by the Belle collaboration

The minus sign appears due to the fact that the $CP$ final state $J/\psi K_s$ is odd. When adding the phase in mixing

$$\frac{q}{p} \simeq \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} \tag{2.34}$$

the entire ratio $\lambda_{J/\psi K_s}$ becomes:

$$\lambda_{J/\psi K_s} = -\left(\frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}\right)\left(\frac{V_{cb} V_{cs}^*}{V_{cb}^* V_{cs}}\right)\left(\frac{V_{cs} V_{cd}^*}{V_{cs}^* V_{cd}}\right) = e^{-2i\phi_1} \tag{2.35}$$

The resulting time dependent asymmetry allows to measure the angle $\phi_1$:

$$a_{J/\psi K_s}(t) = \sin(2\phi_1)\sin\Delta M t \tag{2.36}$$

**The measurement of $\phi_2$** can be done by a time-dependent $CP$ asymmetry analysis using $B^0$ decays to $CP$ eigenstates through a b $\to$ u tree diagram [24]. This allows the measurement of the angle between $V_{ub}^* V_{ud}$ and $V_{tb}^* V_{td}$ ($= \phi_2$). A pure b $\to$ u tree transition (which is a indirect $CP$ violation with an asymmetry $\mathcal{S}$) would allow the calculation of $\sin 2\phi_2$ as easily as the way getting $\phi_1$. A typical example among the decay channels of B mesons is $B^0 \to \pi^+\pi^-$, where the decay amplitude can be parameterized as:

$$\lambda_{\pi^+\pi^-} = -\left(\frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}\right)\left(\frac{V_{ub} V_{ud}^*}{V_{ub}^* V_{ud}}\right) = e^{-2i\phi_2} \tag{2.37}$$

Unfortunately the contribution of the b $\to$ d penguin amplitude can not be neglected this time. As these amplitudes have different weak phases, $\mathcal{S}$ is not equal to $\sin 2\phi_2$ anymore,

but gets a dependence of $\kappa$:

$$\mathcal{S} = \sqrt{1 - \mathcal{A}^2}\sin(2\phi_2 + \kappa) \tag{2.38}$$

The parameter $\kappa$ can be extracted by measuring the branching ratios of $\mathrm{B}^0 \to \pi^+\pi^-$, $\mathrm{B}^0 \to \pi^0\pi^0$, $\mathrm{B}^+ \to \pi^+\pi^0$ and the time dependent asymmetry of $\pi^+\pi^-$.

**The measurement of $\phi_3$** can not be measured by using mixing induced $CP$ violation in $\mathrm{B}^0$ decays as they were used for measuring $\phi_1$ and $\phi_2$ [25]. To get $\phi_3$ − which is the angle between $V_{\mathrm{ub}}^* V_{\mathrm{ud}}$ and $V_{\mathrm{cb}}^* V_{\mathrm{cd}}$ − decay modes involving an interference between $V_{\mathrm{ub}}$ and any element of the CKM matrix except $V_{\mathrm{td}}$ can be used. However there is an interfering effect of strong phase in $CP$ violation which complicates the measurement of $\phi_3$. A promising way to get good results is using pure tree decays which receive no penguin contributions at all and therefore no interfering strong phase interactions. An example for such decays are the combination of $\mathrm{B}^- \to \mathrm{D}^0\mathrm{K}^-\ f_{com}\mathrm{K}^-$ and $\mathrm{B}^- \to \overline{\mathrm{D}}^0\mathrm{K}^-\ f_{com}\mathrm{K}^-$, where both $\overline{\mathrm{D}}^0$ and $\mathrm{D}^0$ have to decay to a common final state ($f_{com}$). If this condition is fulfilled, the amplitudes of these decays interfere, which results in a $CP$ violation related to $\phi_3$.

Another way to determine $\phi_3$ is using $\mathrm{B}_s$ mesons which can be produced by running the accelerator at $\Upsilon(5S)$ energy level. Both experiments BaBar and Belle produced much more data at the $\Upsilon(4S)$ energy level and therefore these decay modes are not as promising as the modes mentioned above. Even after more than ten years of the first generation of B factories, this angle is still difficult to determine and the current value carries a high level of uncertainty.

There are many other decay modes which allow measuring angles of the Unitary Triangle. A deeper view into this topic can be found in [17] and [18].

### 2.1.2.4. *CP* Violation beyond the Standard Model

Although the Standard Model (SM) is the currently most supported model for elementary particles, there are still many open questions. One of them is the characteristic pattern of the mass spectrum of quarks and leptons. No reason has been found within the SM for the fact that the second generation of elementary particles is by several orders of magnitude heavier than the first generation and the third generation tops the second one by another order of magnitude. The hierarchical structure of the CKM matrix, where the values of the diagonal elements are near to one and the mixing angles $\theta_{ij}$ between the $i$-th and $j$-th generation follow the following relation $1 >> \theta_{12} >> \theta_{23} >> \theta_{13}$, can not be explained satisfactorily. Furthermore, the observed neutrino oscillations indicate a rich flavor structure in the lepton sector. All these things and more are free parameters within the Standard model, but should be explained by a worthy successor. Since hadron collider experiments are largely insensitive to flavor violating processes, this area has to be investigated by lepton collider experiments as Belle II for example.

Among the candidates of possible successors for the Standard Model, the following three models are the most promising ones:

- First, the *Minimal supergravity model ($mSUGRA$)* which assumes that supersymmetry (SUSY) is broken in some hidden sector and only mediated to the observable world by gravitational interaction.

- The second one is the $SU(5)$ $SUSY$ $GUT$, which is a model obeying the rules for special unitary groups of 5 degrees. GUT stands for Grand Unification Theory and means that in this theory the gauge couplings of the different interactions can be unified at very high energies. The unification of these interactions assume a relation between flavor mixings of the quark and lepton sector. This theory allows the existence of right handed neutrinos which have not been observed so far.

- Another theory is the $U(2)$ *flavor symmetry*, where the family structure of quarks and leptons could be explained by some flavor symmetry among generations. Because of the Youkawa coupling of the top quark, the natural candidate $U(3)$ is badly broken. $U(2)$ considers a symmetry among the first two generations in the context of SUSY.

But even these models still contain many free parameters. Deviations from predictions of the Standard Model are expected to be found at an integrated luminosity level of $5\text{ab}^{-1}$ which is five times as high as the total integrated luminosity collected by the Belle experiment within 10 years which still holds the world record for the highest luminosity so far.

There are many possibilities for the study of physics beyond the Standard model, but explaining or even listing these would go beyond the scope of this diploma thesis. For more details about that topic, please refer to the sources mentioned at the beginning of this chapter.

## 2.2. SuperKEKB, a second generation B-Factory

The SuperKEKB machine is a second generation B-factory and the successor of KEKB, Tsukuba Japan, details about the upgrade plan for SuperKEKB can be found for example in **??**. The first generation of B-factories proved the existence of $CP$ violation outside the K-meson-system and confirmed the CKM theory for $CP$ violation within the Standard Model. But there are still some open questions (please read 2.1 for more details) which need much more data to answer. Instead of letting the old B-factories collect that data (which would take several decades to achieve the requested amount), it is cheaper − and especially faster − to upgrade the existing ones in order to profit from advances in technology and the experience made with the first generation. The SuperKEKB collider is therefore the upgrade for the successful KEKB, whose task will be to achieve an integrated luminosity about 40 times as high as the current world record hold by its predecessor.

In fact, as many parts as possible are reused from the KEKB-accelerator to keep the production costs low. SuperKEKB will reuse the facility of KEKB and most of the magnets. The main idea will stay the same: using an e$^+$e$^-$-collider (in case of KEKB/SuperKEKB

it has a length of about 3 km) tuned to a collision energy of 10.58 GeV where the resonance peak of the $\Upsilon(4S)$ can be found. Using a lepton machine guarantees low background and high precision, both essential requests for high luminosity. Having an asymmetric setup of a low energy e$^+$-ring (called LER, low energy ring) and a high energy e$^-$-ring (called HER, high energy ring) ensures a boost for the created particles which is essential for the measurement for $CP$ violation. Giving the particles a boost means that they will move farther before decaying into their daughter particles. This allows a distinction between the primary vertex (the interaction point) and the secondary vertex (the point in space of decay).

The beam pipe will be replaced because of the change to a nano-beam scheme which is needed to achieve the higher luminosity. According to (1.4) the luminosity can be increased by increasing the number of bunches ($n$), the number of particles per bunch ($N_1, N_2$ for ) and by decreasing the area of overlapping bunches ($A$). It is planned that all of those parameters will be optimized, which leads to roughly twice the number of bunches, twice the number of particles per bunch and especially a smaller $A$.

Decreasing $A$ requests a better focusing of the beam which is bad for the beam lifetime and increases the Touschek effect resulting in a higher beam background in the detector (please read 1.2.1 for more details).

Increasing $n$ does not reduce the beam quality which makes it an easy way to increase the luminosity without side effects. However the number of bunches per beam can not be increased infinitely. Such a beam is no continuous beam of particles, but cut into bunches to be able to accelerate them by oscillating RF fields. These RF fields provide gaps called *buckets* allowing acceleration which is dependent of the frequency the fields are oscillating. Between these gaps, particles would be decelerated by the field and therefore the number of buckets limit the number of bunches possible within a beam. The doubled number of bunches fills every bucket available, thats why the other parameters mentioned above have to be tuned too, although this means getting worse beam quality. The bunch crossing rate of SuperKEKB will be about 518MHz which means that there is one bunch crossing every 2ns.

To reduce disturbing intra-beam-effects provoked by the increased luminosity, it is necessary to increase the beam energy of the LER because of the fact that these effects decrease dramatically with higher energies. This leads to a lower boost for the created particles and therefore worse resolution of their lifetime, but ensures a good signal to noise ratio essential for event reconstruction.

The higher current needed for higher $n, N_1, N_2$ and the reduced beam lifetime demands an upgrade for the electron and positron source to be able to deliver the requested amount of particles. The preaccelerator has to be adapted for the new energies too (the collider ring does only compensate energy losing effects like bremsstrahlung but does no real acceleration by itself, particles arrive at the collider rings with their target energy). All these improvements to the KEKB accelerator ensure fulfilling the request of higher luminosity and physics rate. More details about the differences between KEKB and SuperKEKB can be obtained from the following table or from [5].
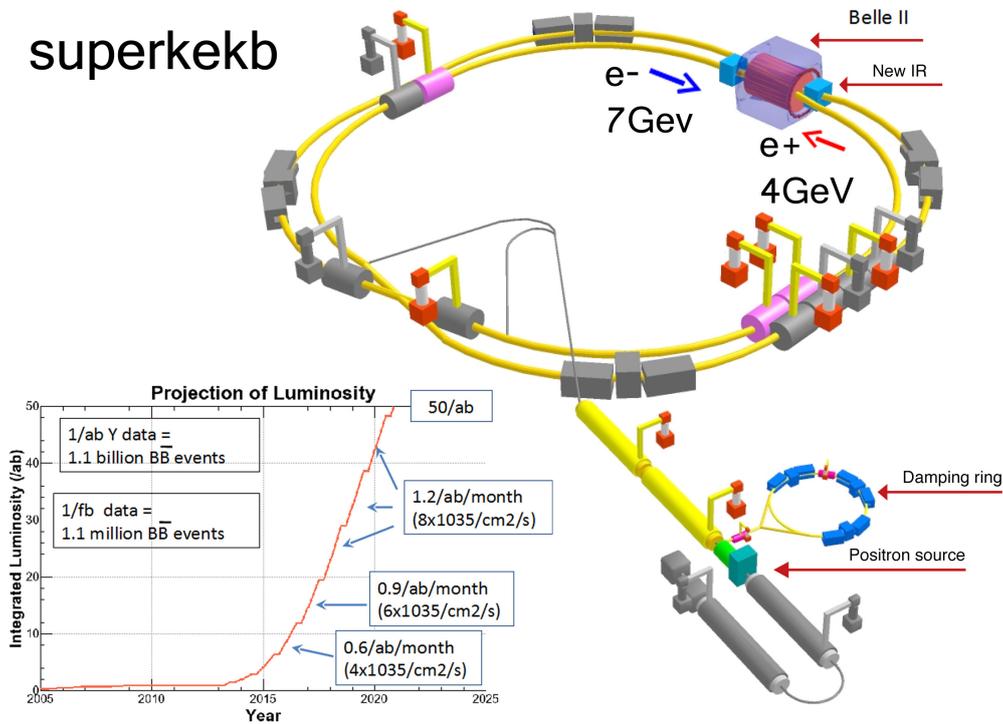
Figure 2.5.: a sketch of the upcoming SuperKEKB collider including the projected time schedule for the integrated luminosity, figure provided by the Belle II colaboration http://belle2.kek.jp

|  | KEKB LER ($e^+$) | KEKB HER ($e^-$) | SuperKEKB LER ($e^+$) | SuperKEKB HER ($e^-$) |
|---|---|---|---|---|
| Beam energy (GeV) | 3.5 | 8.0 | 4.0 | 7.0 |
| Beam lifetime (min) | 150 | 200 | 10 | 10 |
| Beam current (A) | 1.6 | 1.2 | 3.6 | 2.62 |
| Number of bunches | 1293 | 1293 | 2503 | 2503 |
| Luminosity ($10^{34}$cm$^{-2}$s$^{-1}$) | 2.11 |  | 80 |  |

Table 2.2.: some parameters of the KEKB and SuperKEKB accelerators. Data taken from [5] and [6]

## 2.3. Concept design of Belle II

A much more detailed view on the Belle II detector is presented in the technical design report of Belle II ([5]). The major challenge for the new Belle II detector design is the increased physics rate and background level produced by the target luminosity which will be − according to estimations − 20 times higher than the background level of Belle. The old detector design would not be capable of dealing with that, which demands an
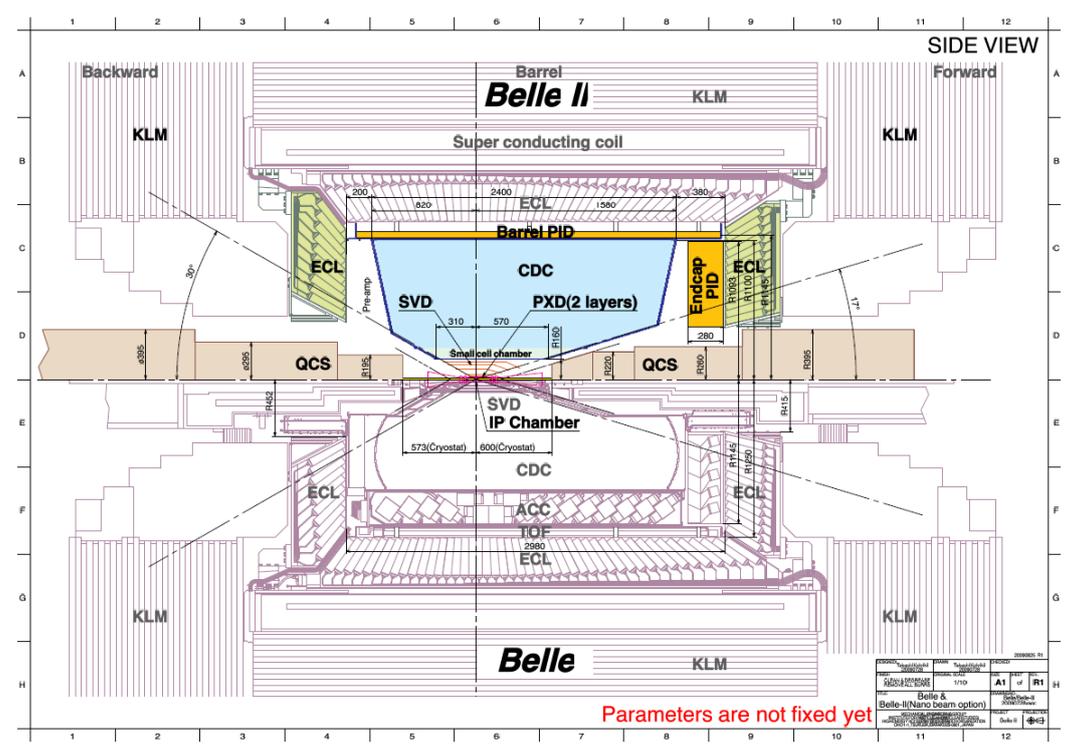


Figure 2.6.: A sketch showing the differences between the Belle II detector (upper half) and the Belle detector (lower half), figure provided by the Belle II colaboration http://belle2.kek.jp

upgrade for the detector too. Although some parts of the old Belle detector can be reused, most of it has to be modified at least because of numerous bottlenecks. To remove these bottlenecks and increase the performance of the new detector despite of the worse environmental conditions the following changes will be made:

- The old 4-layer double sided silicon strip detector (SVD) windmill design will be replaced by a new vertex detector system. The new design consists of 2 layers of silicon pixel detectors based on DEPFET technology (PXD) and 4 layers of a higher resolution double sided silicon strip detector (SVD) including slanted parts for better track reconstruction at small thetas. Compared to the old design the new one extends to a larger radius than in Belle.

- The readout of the SVD has been dramatically improved by reducing the shaping time by several orders of magnitude. This removes a bottleneck which would have lead to an intolerable amount of ghost hits.

- The old central drift chamber will be replaced by a new one (CDC) which has got smaller drift cells improving its resolution and extends to larger radius than in Belle too.

- The electromagnetic Calorimeter will get better electronic readout.

- Barrel and endcap get new particle identification devices.

- The new data acquisition system will be able to deal with the requirements of the increased level of event rate.

The new vertex detector concept increases the vertex resolution by far and combined with the CDC the efficiency for constructing $K_S$ decays into charged pions will be increased. The increased vertex resolution is not only wanted for more precise measurements but also needed due to the reduced momentum boost for the particles of an event. The reduced boost shortens the path-length of the particles roughly by a half.
The Silicon tracking detector gets now 6 instead of 4 layers and since the occupancy at the target luminosity would be to high for the CDC at the old radius of 77mm. Therefore the CDC starts now at about 160mm and allows more space for the silicon detector which is less sensitive to luminosity bound problems.
 Barrel and endcap improve the pion/kaon separation while the new electronics for the electromagnetic calorimeter reduces the noise pile up which is needed for missing energy studies. The following list is a summary of the components found in the new Belle II detector.

- Beam pipe, is a Beryllium double walled tube with an inner radius of 10mm.

- PXD, 2 layers of silicon pixel (DEPFET) detector @ radii of 13 and 22mm.

- SVD, 4 layers of double sided silicon microstrip detector @ radii 38, 80, 115 and 140mm.

- CDC, 56 layers within a small cell drift chamber, where the sense wires are spanned @ radii between 16 and 112cm.

- TOP, a ring imaging Cherenkov detector (RICH) using a quartz radiator of 2cm thickness, situated at r∼120cm coating the barrel.

- ARICH, a RICH using an aerogel radiator of 4cm thickness situated at the endcap in forward direction.

- ECL, an electromagnetic calorimeter using CsI enveloping the barrel and the endcap. At the barrel it is positioned between the radii of 125 and 162cm.
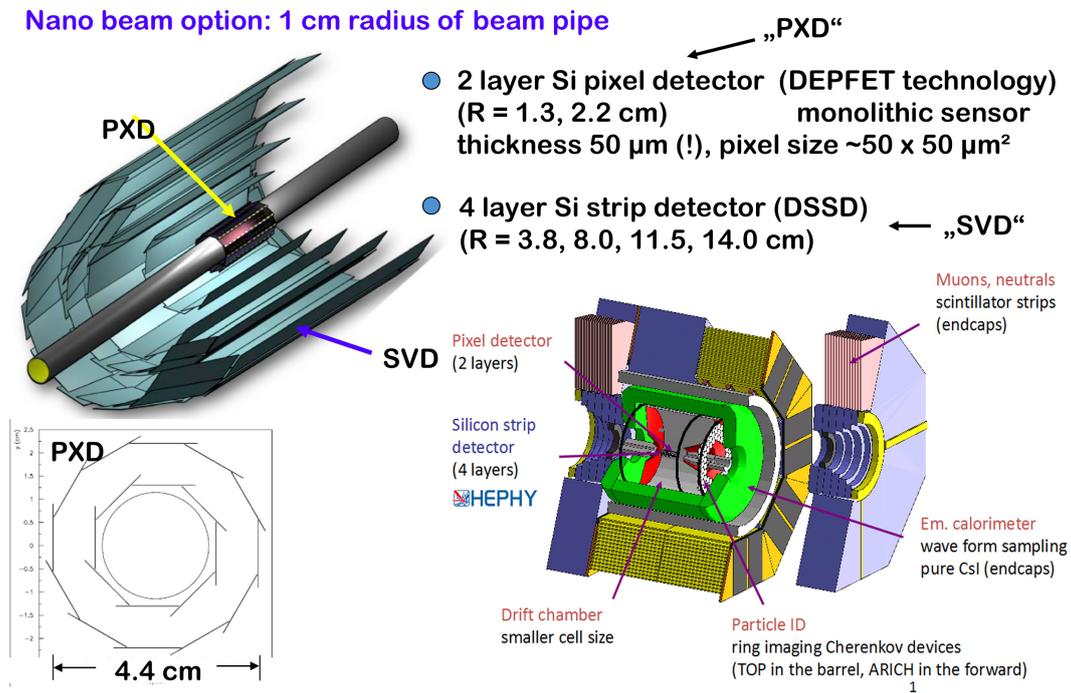
Figure 2.7.: A sketch showing the main parts of the Belle II detector and a closer look at the silicon detector, figure provided by the Belle II colaboration http://belle2.kek.jp

- KLM, at the barrel there are 14 layers (5cm Fe + 4cm gap filled with 2 RPCs per gap) and at the endcap there are 14 layers of scintillator strips.

More detailed information about the detectors of Belle and Belle II can be found in [5] [6] and [26].

## 2.3.1. SVD + PXD of Belle II

A closer look at the new vertex detector system (VXD) consisting of the PXD and the SVD will be enabled because of the importance for this thesis. The VXD will cover a theta range of $17^o < \theta < 150^o$. All layers consist of several ladders positioned in a windmill design to prevent dead spots and increase the number of hits for some tracks (a higher number of hits result in a better vertex resolution which is increasing with $sqrtn$, where $n$ is the number of hits).
 Due to the higher luminosity of the accelerator, especially the time needed for readout is being reduced by far. The SVD2 of Belle needed about 2000ns for one readout frame.
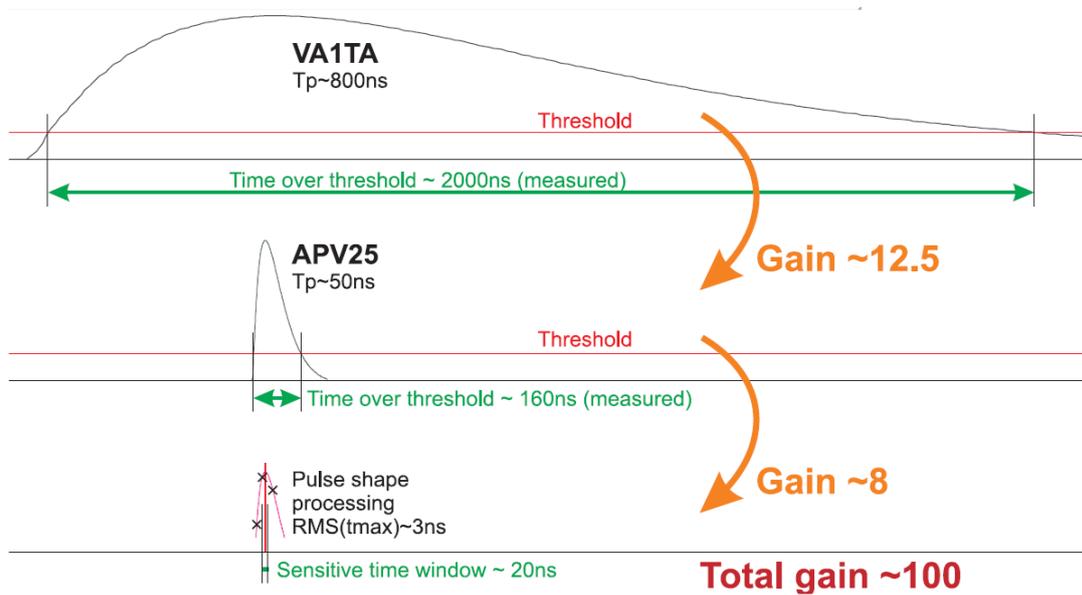
Figure 2.8.: This figure shows the steps needed for reducing the readout time frame by a factor of 100. This can be achieved by using a faster readout chip and pulse shape processing
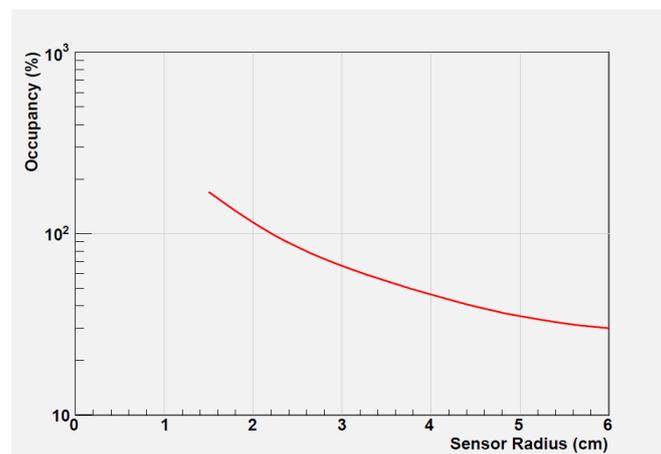


Figure 2.9.: This plot illustrates the theoretical occupancy for the SVD2 detector of Belle asuming the luminosity of Belle II

This would lead to an occupancy of more than 100% at radii below 2.3 cm. Therefore the SVD of Belle II is optimized for faster readout which results in an at least 100 times faster readout with 20ns. But still the occupancy is too high for a SVD-only design because of the ghost-hit problem which is typical for strip detectors. This and the substantially higher resolution in z-direction makes the new pixel detector an essential add-on for tracking.

The PXD will dramatically improve the vertex resolution but has got a major downside: Because of the thinness of the detector layers (the sensitive part has got a thickness of $50\mu m$ which is needed for low energy loss) and the small pixel size (needed for good vertex resolution) the PXD is rather slow compared to the SVD and the CDC, resulting about $20\mu s$ per readout frame (but still only a factor of ten slower than the old SVD2 setup. This means that one readout frame of the PXD will contain the information of
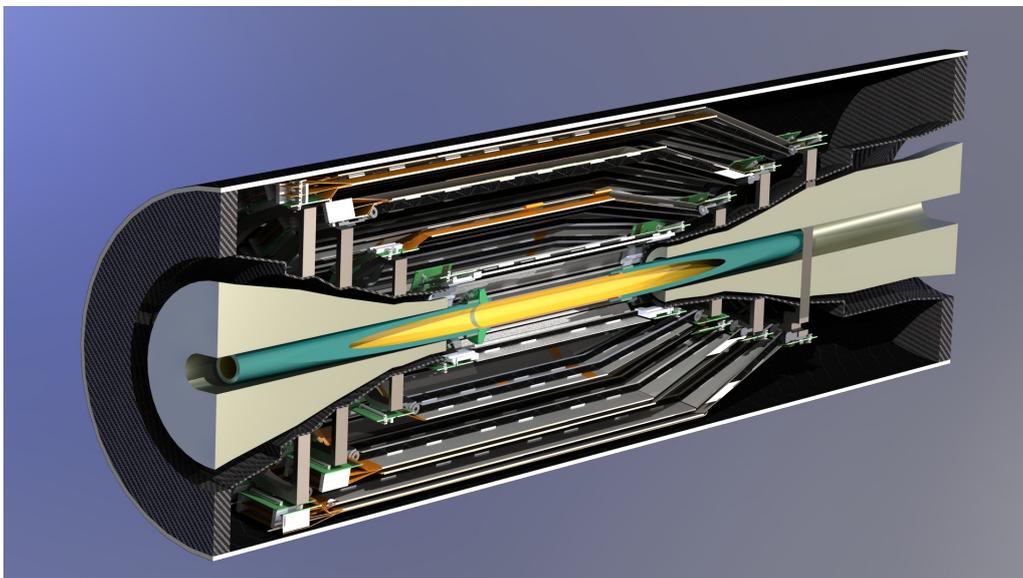


Figure 2.10.: A rendered view of the silicon detector including beampipe and support structure, plot provided by Immanuel Gfall (Belle II collaboration)

roughly 1000 readout frames of the SVD. Although about only one interesting physics event will occur during a PXD readout frame, 1000 bunch crossing events will happen in the same time and produce noise. QED-Background, Touschek effect and partially Bhabha scattering will be the main causes for background within the PXD. However using a faster silicon strip detector instead of the PXD is not an option. The reason for that is that the downside of microstrip detectors takes effect at the assumed level of background: The high background results in high occupancy and therefore leads to a high number of ghost hits (please read 1.2.2.2 for more details). The number of ghost hits rise quadratically and therefore an occupancy of about 10% would already reach the limits of reconstruction techniques. This makes the PXD still far better for high occupancy situations although having a slow readout. The occupancy level for the PXD

at the innermost layer is expected to be about 1%, where the occupancy level for the innermost SVD layer is expected to be about 7% and therefore manageable. The lower occupancy in the PXD arises because of the high number of pixels at the detector layers compared to the number of strips within the SVD.

The SVD introduces slanted parts within the forward region to increase the efficiency of track reconstruction at small thetas. These slanted parts reduce the radii of the layers in that area and therefore increase the possibility of recording low $p_T$ tracks. The decreased shaping time of the SVD readout electronics makes the SVD fast enough to allow the recognition of curling tracks even at thetas near $90^o$. Relevant for both subdetectors is the energy resolution of the recorded hits. For example the SVD will have 7-8 bit resolution for energy deposit within its layers. This allows distinguishing between hits of high energy tracks (small energy deposit) and low energy track hits (high energy deposit) which supports the event reconstruction.

# 3. Pattern recognition in theory and practice

## 3.1. Motivation

The Belle experiment was a great success for the B factories. Due to studying the high number of events including B mesons, the Standard Model could be confirmed. However, there are several hints that the Standard Model can not be the final theory for particle physics, see 2.1.2.4 for more details. To find deviations from this theory, much more data is needed than the two B factory experiments BaBar and Belle have recorded together. Therefore Belle is being upgraded to Belle II which will record about 50 times more events than Belle. A detailed description of the changes in detector design due to the new environmental conditions can be found in [5] and chapter 2.3.

However some aspects especially relevant for low momentum track finding will be illustrated here. There exist decay channels with low momentum charged pions which could not be analyzed satisfactorily at the first generation experiment Belle.

$$D^{*+} \rightarrow D^0 \pi^+ \tag{3.1}$$

In the old tracking detector, tracks were found in the CDC and extrapolated into the SVD. This method is applied in the new detector too. However, the old track reconstruction system was not capable of tracking particles with transverse momenta less than 100Mev/c. The new inner radius of the CDC being 160mm instead of 77mm increases the effective minimum momentum value for this type of track finding, which in turn increases the urgency for a specialized low momentum track finding technique. This track finding shall be done by using only the silicon detector and aims to find tracks of particles with transverse momenta down to 50MeV/c.

The next important reason for low momentum track finding is the amount of data recorded by the PXD. This is roughly 10 times more than the amount that can be stored, which demands on-line data reduction.

Next to typical triggering techniques there is a plan to include a more complex track finding tool for on-line data reduction. Although the current low momentum track finder, which is described in this thesis in chapter 3.7, is developed for off-line track finding, the demands for on-line track finding are already considered. This allows us to adapt the off-line track finder for on-line purposes in future steps.

## 3.2. General requirements for the low momentum Track Finder of Belle II

The first two chapters *Introduction* 1 and *The Belle and Belle II experiments* 2 illustrate typical subtleties that have to be considered when developing a track finding tool. General restrictions are described in chapter 1, while Belle II-specific aspects are mentioned in chapter 2. The purpose of this thesis is developing a track finder which is up to the given task of finding low momentum tracks in the silicon part of the tracking detectors of the Belle II experiment. The following list gives a summary of the most important requirements for the low momentum trackfinder of Belle II:
The track finder should be able to deal with

- the specific situation in general (presence of magnetic field, arrangement and total number of the detecting layers, ...),

- the expected rate of real tracks (expected to be about 10 tracks per event) per timeframe,

- missing hits (due to inefficiencies in the detecting layers),

- ghost hits (a detector-specific problem, causing multiple possible hits for one true hit. Effect increases quadratically with the number of hits per sensor),

- background hits (accelerator-specific effects like *Bhabha-scattering* as the dominating beam-beam-scattering effect and the *Touschek effect* as the dominating intra-beam effect are the most important contributions to background),

- energy loss and multiple scattering (especially relevant for low momentum tracks) changing the trajectory of the particle and

- time constraints (only a short amount of time can be reserved for reconstructing an event).

The higher projected luminosity of Belle II compared to its predecessor will not only increase the physics rate per second but also the side effects like background and ghost hits, where the latter increase with the second power of the number of hits per sensor. The developers of the tracking detectors are aware of that fact and have adapted the design to the more difficult setting. The PXD(2.3.1) and the dramatically reduced time consumption for a readout frame for the SVD will help to suppress negative effects but will not be able to eliminate them completely. Therefore, a major design goal for the low momentum track finder will be to be able to deal with signal to noise ratios of up to about $1 : 20$.
Another problem is the energy loss within the detector layers. As mentioned in 1.2.2.1, the Bethe formula ((1.2.2.1)) indicates the lowest energy loss at $\beta\gamma = 3$, rising rapidly when going to lower momenta which means bigger changes in trajectory after each layer. Figure 3.1 shows the energy loss for the interesting momentum region for the most
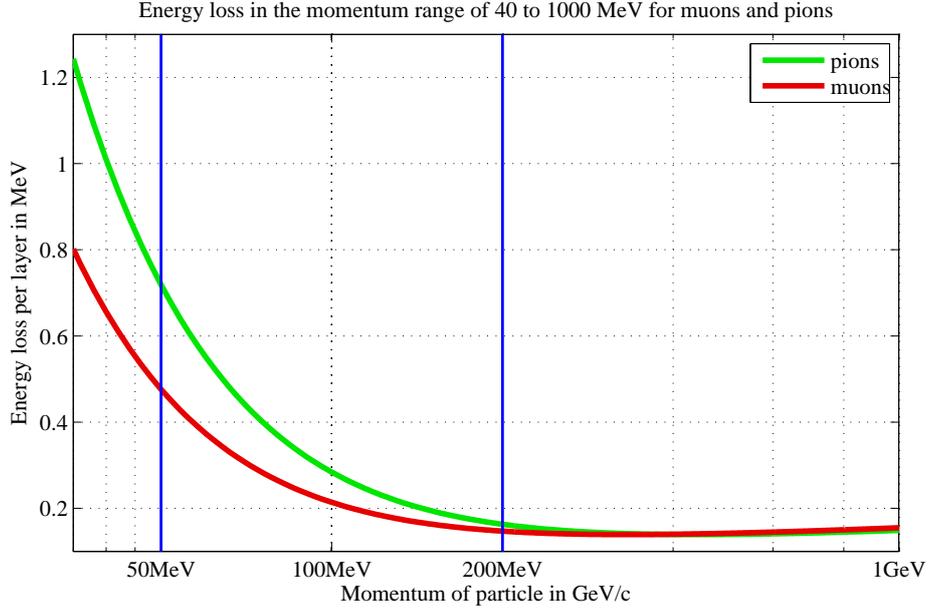
Figure 3.1.: Energy loss within a silicon layer of the tracking detector according to Bethe formula for muons and pions at low momenta. The interesting rangfor the low momentum track finder is $\sim 50\text{MeV} < p_\text{T} < 200\text{MeV}$

common particles detected in Belle II, muons and pions. For the most common particles — pions and muons — this means a momentum of $\sim 300\text{MeV}$. The operation area of the low momentum track finder will be within the range of $\sim 50\text{MeV} < p_\text{T} < 200\text{MeV}$ because of the fact that tracks with higher momentum can be found by extrapolating tracks reconstructed in the *Central Drift Chamber* (CDC).

The following sections show how simple and sophisticated track finding concepts deal with these problems and keep the time consumption within acceptable limits.

## 3.3. A basic technique: Conformal mapping

A very simple and fast method is finding tracks by histogramming the hits over the angle $\phi$ of the transversal plane. This technique, in its simplest form, uses the fact that the interaction point in the transversal plane is roughly at the origin of the system. In the simplified case without a magnetic field, particles propagate through space in straight lines starting from the origin, if material effects are neglected. This means that each hit from the same track will have the same polar angle and will therefore produce a peak in the $\phi$ histogram. Even when introducing a magnetic field (bending the tracks into circles in the transversal plane), the case is still very easy to treat: Applying the inverse conformal transformation

$$u = \frac{x}{x^2 + y^2}, \tag{3.2}$$

$$v = -\frac{y}{x^2 + y^2} \tag{3.3}$$

on each hit defines the *conformal mapping*. It transforms points lying on a common circle passing through the origin into points on a straight line. For high momenta, these lines pass very near the origin since the minimal distance to the origin is $1/2R$, where $R$ is the radius of the circle. Likely tracks can now again be found by investigating peaks in a histogram.

This technique provides many advantages: apart from its simplicity, a very important aspect is that each hit can be treated independently from every other, which allows for an easy adaption to parallel processing. Such a concept is called a *global method* [8]. Even more important is the second point that the time consumption increases *linearly* with the number of hits per event. This makes this technique a good candidate for track finding in drift chambers and other tracking detectors with a large number of (super-)layers, because time consumption is not a limiting factor for this technique. Another advantage of conformal mapping is the fact that missing hits due to detector inefficiencies are easy to compensate for. In the case of 6 layers, a standard track would lead to a bin height of 6, whereas a track with a missing hit would lead to a height of 5, which is still easy to recognize.

However, there are drawbacks that disqualify conformal mapping in several situations (especially in the low momentum track finding of Belle II, see chapter 4 for more details). First of all, this technique is only capable of finding circles with high radii (and therefore high momenta) because of the $1/2R$-dependency of the distance to the origin. Circles with small radii miss the origin by far after the transformation which results in the following Situation: low momentum tracks lead to different polar angles for each hit, and therefore performance can only be recovered by reducing the number of bins in the histogram which leads to a loss in reconstruction resolution. To compensate for this, more sophisticated techniques like the *Hough transformation* have been developed (see 3.4 for more details). The second drawback is that this algorithm only works in 2D: two tracks going in the same direction on the transversal plane, but in different directions in the $r$-$z$-plane cannot be distinguished this way. In low occupancy situations this is not very problematic, but combined with the next drawback, it can be a serious problem: the method is highly sensitive to track deflecting effects like an inhomogeneous magnetic field, energy loss and multiple scattering. These effecte are especially relevant for silicon detectors and low momentum situations (see 1.2.2.1 and (1.2.2.1) for more details), while in gaseous detectors it is easier to deal with, since there the density of the fiducial volume is lower by several orders of magnitude. Such deflecting effects demand a more loose binning in the histogram which leads to worse resolution in track recognition. Combined with background hits and multiple real tracks, these disadvantages are severe enough to disqualify conformal mapping completely for low momentum track finding in the SVD of Belle II (more information can be found in 4.3).

## 3.4. Hough transform

The *Hough transform* uses a more generalized approach to recognize tracks. While conformal mapping is only working with circles passing the origin and in 2D, the Hough transform searches for arbitrary analytical functions in 2D or 3D. The only condition is that the track can be described with a model which can be parameterized.

One of the simplest models is a straight line. The idea is that the equation of a straight line $y = cx + d$ in the $x$-$y$ plane (called *image space*) can be transformed into another straight line in the $c$-$d$ plane (called *parameter space*). In the image space $(x, y)$ are the coordinates of a hit and $c, d$ the parameters for the line equation. Staying in the image spaces would lead to an infinite number of lines fulfilling that equation for the current coordinates. But transforming them into parameter space leads to the situation where $(c, d)$ are the coordinates and $x, y$ the parameters for the new line equation $d = y - cx$. Hits along the same line (means having same parameters $c, d$) in image space will intersect at the same point $(c, d)$ in parameter space. Finding these intersection points (by histogramming over the $c, d$ space) allows the determination of the parameters for the lines connecting hits of a straight track since they form a peak in the parameter space.

Although the transformation itself is linear with the number of hits, the two-dimensional histogramming is not. The number of parameters needed to describe the model determines the numbers of dimensions of the parameter space. Unfortunately this parameter based approach is paired with its major drawback, time consumption. While conformal mapping has a linear increase of time consumption with the number of hits, the amount of time needed for Hough transform goes with the number of hits squared when using a 2D model and with the number of hits cubed when using a 3D model. This aspect practically disqualifies the usage of a 3D model or more complex ones in situations of high occupancy, while low occupancy situations can be processed equally well by simpler techniques like conformal mapping. Therefore, the Hough transform is effectively a 2D model as well, which results in the same drawbacks as for conformal mapping.

An optimized version called *fast Hough transformation* searches for peaks in parameter space without the need for binning the whole parameter space, by using an algorithm of the divide-and-conquer type.

There exist many different implementations of the Hough transformation using models which describe the track more accurately. Using constraints reduces the number of parameters needed for the track model (the ideal model would need 5 parameters which demands binning of 5 dimensions).

## 3.5. Sophisticated technique: Kalman filter

This technique is planned for implementation within the low momentum trackfinder (see section 3.7), which demands a closer look at the principle. The Kalman filter was introduced to the field of event reconstruction in particle physics in 1987 [16] and is now a standard tool for track fitting. It can be regarded as a statistically optimal track following procedure which takes a track seed as a starting point.

A track seed is a combination of hits which can be chosen randomly or delivered by pre-filtering techniques using boundary conditions defined by the current detector layout. Statistically optimal means that — when using a linear model where the errors have a mean of zero, are uncorrelated and have equal variances — the Kalman filter is the best linear unbiased estimator of the parameters describing the current state of the system. This feature is inherited from the least-squares method, of which the Kalman filter is a recursive version.

Recursive means that instead of calculating the total track properties at once — which would include the inversion of matrices of considerable size (which is always a time consuming process) — the algorithm propagates from one measurement of the track to the next. The underlying model is the algorithm operating on a Markov chain, where each state is only dependent of its predecessor but carries the information of the full chain up to the current state. Following the the full chain means that the algorithm has gathered the full information stored in the chain, which delivers the same quality of result as using the normal least squares method. However, the Kalman filter is faster due to the fact that for $n$ measurements and $a$ parameters, inverting $n$ $a \times a$ matrices is much faster than inverting one $an \times an$ matrix where many digits are zero.

As indicated above, this technique needs a model describing the state of the system to be able to follow a potential track from the track seed. This model describes the state of the particle track at each measurement and allows for propagation from one measurement to the next (e.g. propagating from layer 4 to layer 5). In our case, a track model has to be chosen, appropriate to the experimental situation and using the smallest possible number of parameters to describe it. A model using more than the minimum number of parameters would lead to linear dependent rows in the covariance matrix. This means that the matrix can not be inverted any more. The number of parameters are determined by the state of the system (the track) which has to be described sufficiently. At each point along the trajectory it is sufficient to describe the state of the particle by its position $(x, y, z)$ and its momentum $(p_x, p_y, p_z)$, leading to a total of 6 parameters. Constraining the position to a plane (such as a sensor layer), the position can be reduced to 2 parameters; therefore, the full track description has only 5 components (see 1.3.3), which are stored in the state vector for the $k$-th measurement $\mathbf{q}_k$. In the index of this value, additional information is coded: in case of $\mathbf{q}_{k|j}$, when $k > j$, the filter is in prediction mode, and when $k = j$, it is in filter or update mode. The third case is $j > k$, where the filter runs backwards. This is often used as smoothing step: after doing the fit in one direction, the Kalman filter can revisit the hits in the other direction again and update the errors along each step, using the full information from the complete fit (forward and backward). This improves the errors, but is relatively time-consuming, which is why smoothing is generally used in track fitting, but not track finding. Propagation from one layer to the next requires the knowledge of the magnetic field and the layout of the detector in order to consider material effects.

Algorithms such as the Runge-Kutta-method are a common choice for the solution of the ordinary differential equation describing the propagation from one hit to the next.

The Kalman filter consist of two steps. First there is a *prediction step*, in which it

calculates a prediction of the next state based on the information of the current one and the track model used. In the reconstruction process, the track at each measurement is described by the state vector:

$$\mathbf{q}_{k|k-1} = \mathbf{f}_{k|k-1}(\mathbf{q}_{k-1|k-1}) \tag{3.4}$$

where the notation $k|k-1$ indicates the predicted state vector for the $k$-th step based on the information from $k-1$ and $k-1|k-1$ indicates the situation at $k-1$. $\mathbf{f}_{k|k-1}$ is the track propagator from surface $k-1$ to $k$. And the covariance matrix of the state:

$$\mathbf{C}_{k|k-1} = \mathbf{F}_{k|k-1}\mathbf{C}_{k-1|k-1}\mathbf{F}_{k|k-1}^{\mathrm{T}} + \mathbf{Q}_k \tag{3.5}$$

It contains the derivative of the track propagator, the Jacobian $\mathbf{F}_{k|k-1}$, the covariance matrix of the last state, $\mathbf{C}_{k-1|k-1}$, and the covariance matrix of the material effects after layer $k-1$ up to and including layer $k$, denoted by $\mathbf{Q}_k$.

After finishing the prediction step, the *update step* combines the predicted state with the measurement on the next layer to define the next state:

$$\mathbf{q}_{k|k} = \mathbf{q}_{k|k-1} + \mathbf{M}_k\Big[\mathbf{m}_k - \mathbf{h}_k(\mathbf{q}_{k|k-1})\Big] \tag{3.6}$$

which includes the gain matrix $\mathbf{M}_k$ and is belonging to $\mathbf{m}_k$ the measurement vector of the hit at layer k. $\mathbf{M}_k$ is given by:

$$\mathbf{M}_k = \mathbf{C}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}(\mathbf{V}_k + \mathbf{H}_k\mathbf{C}_{k|k-1}\mathbf{H}_k^{\mathrm{T}})^{-1} \tag{3.7}$$

The Matrix $\mathbf{V}_k$ is the covariance matrix of the measurement errors and is expected to be known. $\mathbf{h}_k$ denotes the measurement model function, where $\mathbf{H}_k$ is the Jacobian of the transformation of the state vector into the local coordinate system; in most cases, it can be calculated analytically. In the calculation of the gain matrix, a matrix inversion has to be done. Since the matrix is of rank 5 (the free parameters in our track model), this calculation can be done relatively fast.

Finally, the update of the covariance matrix is given by:

$$\mathbf{C}_{k|k} = (\mathbb{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{C}_{k|k-1} \tag{3.8}$$

where $\mathbb{I}$ is the identity matrix.

These two steps are repeated until the last measurement is reached or the track quality gets too bad. The calculation of the track quality is done in each update step to find out whether the newly added measurement was a good choice or a bad one. To define the quality of a hit and to make the filter robust against outliers, the residuals are used. The residual of the measurement in layer k in respect to the updated state vector is

$$\mathbf{r}_{k|k} = \mathbf{m}_k - \mathbf{h}_k(\mathbf{q}_{k|k}) \tag{3.9}$$

The covariance matrix of this residual is

$$\mathbf{R}_{k|k} = \mathbf{V}_k - \mathbf{H}_k\mathbf{C}_{k|k}\mathbf{H}_k^{\mathrm{T}} \tag{3.10}$$

This allows the calculation of the chi-square increment in layer k, which is

$$\chi^2_{k|k} = \mathbf{r}^\mathrm{T}_{k|k} \mathbf{R}^{-1}_{k|k} \mathbf{r}_{k|k} \mathbf{H}^\mathrm{T}_k \tag{3.11}$$

To get a track *quality indicator*, all chi-square increments for all measurements of the track candidate are summed up. Outliers can be found that way, due to the fact that too large values for the chi-square indicate that their corresponding hits do not originate from the same particle as the others. The theoretical mean of the total chi-square is calculated the following way:

$$\overline{\chi}^2 = np - f \tag{3.12}$$

where $n$ represents the number of hits in the track candidate, $p$ stands for the dimensionality per hit (in the VXD of Belle II, this is 2) and $f$ is the number of parameters needed to define the track (in our case, $f = 5$).

There are some implementations of the Kalman filter which improve its numerical robustness (the *information filter*) or its behavior in high occupancy situations (the *combinatorial Kalman filter*). The combinatorial Kalman filter clones itself whenever there are more than one good hit in reach. This ensures that the seed does not die because of one "well placed" background hit, as it could happen in the standard Kalman filter. No matter which implementation is used, without the reduction of possible next hits to include, the Kalman filter has to check every one in the next layer. This is a very time-consuming process because of the complexity of the calculation and especially the combinatorial problem which arises from the increasing number of possible combinations with each new layer. Therefore it is important to combine the Kalman filter with a pre-filtering method to reduce the bottleneck of time consumption when searching for tracks.

## 3.6. Neural network of Hopfield type

This filter is implemented in the low momentum trackfinder (see 3.7) and therefore discussed in more detail. The Hopfield network is a widely used technique for combinatorial optimization problems. Event reconstruction processes can be formulated as such a problem, which is described in detail in [15], [27] and [8]. In the current version it is not used as track finding algorithm but as a tool for chosing the best subset among overlapping tracks (see [28] for more details).

A Hopfield network consists of binary neurons where each neuron is connected to all other neurons with a symmetric connection weight $W_{ij} = W_{ji}$, which can be stored in a weight matrix $W$. Neurons are updated according ot the following rule:

$$S_i = \mathrm{sgn}\left(\sum_{i=1}^{n} W_{ij} S_j\right) \ , \ i = 1, ..., n \tag{3.13}$$

using the sign function

$$\mathrm{sgn}(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ -1, & \text{if } x < 0. \end{cases} \tag{3.14}$$

Applying rule (3.13) asynchronously (meaning single neurons can calculate their new state in a random order) leads to a minimization of the energy function of the network:

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}S_iS_j \tag{3.15}$$

This process leads to a local minimum of the enrgy function whereas the global minimum represents the solution to the problem. Which local (or global) minimum is reached strongly depends of the starting point. Since the solution of the problem is not known beforehand, an arbitrary starting point can be chosen. Finding the global minimum can be done by restarting the process several times at different positions and collecting the results to chose the one with the lowest final energy. This approach is very time consuming, therefore the technique of *mean-field annealing* is preferred. It introduces an additional variable acting as temperature $T$, which allows leaving local minima due to smoothing out the area by the additional energy of the temperature. The temperature is decreased after every round. Although this technique does not guarantee to find the global minimum, it usually finds at least a good local one.

In the current problem of finding a clean subset of overlapping track candidates, where a tradeoff between the highest number of chosen track candidates or the highest quality of the track candidates has to be made, each neuron represents a track. The information whether two neurons are compatible, which means that coresponding tracks do not share hits, or incompatible (they share at least one hit) is stored in matrix $W$. Its element at position $ij$ is set to 0 when track $i$ and $j$ are incompatible, the element is set to one when this is not the case. The weight matrix $W$ consists of the elements:

$$w_{ij} = \begin{cases} -1 & \text{if } i \text{ and } j \text{ are incompatible} \\ (1-\omega)/N & \text{if } i \text{ and } j \text{ are compatible} \end{cases} \tag{3.16}$$

where $\omega$ is a weight steering whether large number of track candidates or large total value of quality indicators are favored. The quality indicators for each track candidate are provided by the Kalman filter. With the activation funciton using the temperature $T$

$$g(x_i) = \frac{1}{2}(1 + \tanh(x_i/T)) \tag{3.17}$$

the new state $g(x_i)$ will be calculated for each neuron sequentially. It will be calculated summing over all inputs of the current neuron. If the result is above a threshold, the neuron gets activated, if the value is below, the neuron gets deactivated. The initial states are chosen randomly between [0, 0.1] (uniformly distributed) and are changed by the activation function within each iteration for each track once. It is important that the sequential arrangement of the tracks be chosen randomly for each iteration and not updated in parallel, since this could cause the network to cycle between two states.

The temperature starts at $T_1 = 2.1$ and is lowered for each iteration using the following formula:

$$T_{n+1} = \frac{1}{2}(T_n + T_\infty) \quad \text{with } T_\infty = 0.1 \tag{3.18}$$

The algorithm repeats applying the activation function for each track until the states change by no more than 0.01. All track candidates whose states are above a threshold of 0.7 are allowed for the final set of track candidates.

The major downside of this approach is its incompatibility with parallel computing. However, stable parallel versions of the Hopfield network have been developed.

## 3.7. Low momentum trackfinder for the SVD at Belle II

In this section, the development of a track finder specialized on low momentum situations and only 3 to 4 layers to find tracks is sketched. The requirements are mentioned in 3.2 and will be checked in 4.

The track finder will run as an off-line version in the basf2 framework, which is the software framework for the upcoming Belle II experiment at KEK, Japan. The future task of that framework will be a range of applications from detector optimization (e.g. fine-tuning the detector layout by simulation studies) to physics analyses. Its core for simulation is the Geant4 engine [29], which is a toolkit simulating the interaction of elementary particles with matter. Within the framework, all applications are structured in modules, sharing a common datastore.

The current track finder consists of two modules and an external converter tool. The first module is called *FilterCalculator*, the second one *VXDCATF*, which stands for *VerteX Detector Cellular Automaton TrackFinder*.

The task of the FilterCalculator module is to streamline the cellular automaton for its given task. It delivers a lookup table of cutoffs for special filters applied by the cellular automaton as well a set of sectors and their relations (more details on this in the next section).

The TrackFinder itself consists of several parts. The first one applies the filter rules to the current event. Its task is to reduce the combinatorics to improve the speed of the following filter steps. The next part is the cellular automaton, searching for usable chains of hits that represent promising track candidates. These are collected and provided with a quality indicator for the neuronal network, which then chooses a compatible subset of final track candidates.

It has to be mentioned that the current state of the track finder is under construction and its development will take about another year after the completion of this thesis. So the current state is merely a snapshot of the track finder and many parts will change before completion; major parts like the quality indicator calculator (Kalman filter) or features like curling track recognition are missing, and therefore the final performance cannot yet be determined. Still, some of the planned aspects are briefly mentioned here to complete the concept of the track finder. Furthermore, parts of the framework relevant to the track finder (for example the digitizers for the detector which produce realistically smeared hits and ghost hits) have been committed to the framework very recently, and therefore are not yet implemented in the track finder. Since features like simple or realistic background are still partially missing or have been added recently, performance studies concerning these effects do not exist yet.

**Schematic view of the low momentum track finder in Belle II**

Unsorted hits from tracks, background, ghost coming from an event

**O**     *Sector setup - 1-hit filter*
filters by set of compatible sectors, allows momentum dependent setups

**O**     *Segment finder - 2-hit filter*
filters by distance, min&max, including virtual Segment

**O**     *Neighbour finder - 3-hit filter*
filters by angle and Δ-distance min&max

*Cellular Automaton*
evolving states, includes TC-collector

**O**     *Post 4-hit filter*
filters by zigZag, ΔpT

*Kalman filter*
not implemented yet

*Hopfield Network*
uses QI's to find best subset among overlapping TC's

*Clean TC's*

- *The arrows represent a schematic interpretation of the possible number of combinations of hits at that point*
- *Filters marked with an* **O** *use external information generated by simulation*

Figure 3.2.: A schematic view representing the parts of the track finder and their use in reducing combinatorics

However, a proof of concept exists, using a simplified detector geometry; in this case the combination of cellular automaton, Kalman filter, and Hopfield network produces promising results (more details can be found in 4).

### 3.7.1. Chosen approach to low momentum track finding

The combinatorial problem and the resulting time consumption is the bottleneck of practically every track finding algorithm. It is therefore a desirable goal to reduce the combinatorics as much as possible. Our approach is to combine the advantages of a slow but statistically optimal track following algorithm like the Kalman filter and the advantages of fast but rough filters.

The sensible way to do this is starting the track finding process with the fastest filtering technique, which reduces the possible number of hits for the second iteration where a sightly more complex filter step is executed. This chain of increasingly complex filter steps ends in the Kalman filter as the optimal estimation algorithm, whose pool of possible hits at that point is so small that the combinatorics are not an issue anymore. The flow chart describes the general concept of filters in the current implementation of the track

finder (kommt noch). The following list summarizes some aspects of it as follows:

- The simplest filter, the *sectormap*, can be regarded as a one-hit-filter whose time consumption increases linearly with the number of hits. The concept of sectors, the sectormap, and its pros and contras will be described in 3.7.2.

- 2-hit filters that combine two compatible hits to segments (compatibility defined by sectormap and filter cut-off) are bundled into the *segfinder*. These segments are of central relevance for a later filter (the cellular automaton) and are described in 3.7.3.

- The *neighborfinder* contains all 3-hit filters which are applied to two segments sharing a node (a hit in the center of that two-segment-chain). If the segments are compatible, they become neighbors; this is a central aspect needed by the cellular automaton. Can be found in 3.7.4.

- The *cellular automaton* uses the neighbors created by the previous filters to find chains of segments which form track candidates. Its working principle will be shown in 3.7.5.

- The *superneighborfinder* is the set of 4-Hit-Filters that filter out the last remnants of physically incorrect track candidates. The filters used and some other aspects are discussed in 3.7.6.

- The *Kalman filter* is not implemented yet, but since it is an integral component of the concept, some aspects will be explained in 3.7.7.

- Finally, the *Hopfield network* takes the set of overlapping track candidates and uses their quality indicators to find a clean subset of final track candidates. It will be discussed in 3.7.8.

It should be noted that in order to avoid confusion, in the following sections two compatible hits or segments are called *neighbors*, whereas two compatible sectors are called *friends*. Friends or neighbors are always the *inner* partners of the pair since the track following algorithms within this track finder are moving inwards.

### 3.7.2. 1-Hit filter: sectormap

Looking at the windmill design of the VXD reveals a problem for track following techniques: since the sensors of each layer overlap, there can be more than one hit per layer of the same track (actually the same track can have many hits on the same layer when also considering curling tracks). This means that a track following technique does not only have to consider every hit of the following layer, but (when not applying further constraints) all of the hits of the current layer as well. The number of possible two hit combinations per hit doubles and therefore the total number of combinations rises with $(2n)^2$ instead of $n^2$ (where $n$ is the number of hits). In light of the total number of combinations, even simple distance calculations of two hits can become very time consuming. To reduce that problem, one

could calculate the radius of each hit and only consider hits of smaller radii. But even this technique requires $(2n)^2$ total comparisons.

Another concept is subdividing each sensor into at least two sectors in the $x - y$ plane. One forms the upper sector (in terms of mean radius) and can therefore have hits on the same layer while the other one forms the lower sector which is not allowed to search for partner hits on the same layer. The linear operation of sorting each hit into such sectors reduces the number of combinations to about $(1.5n)^2$.

This idea can be developed further by defining a set of compatible sectors for each sector separately; the sectors in this list are called *friend sectors* of the current sector. A track following algorithm using a hit on a sector is therefore only allowed to search for neighbors in the friend list of the sector.

In a basic implementation, the set of friends of an upper sector of layer $i$ contains all lower sectors of the neighboring ladder in layer $i$ and all sectors of layer $i - 1$.

Since the geometry of the detector and the position of the interaction point will roughly stay the same over a long time, the friend list for each sector can be defined by simulating a sufficient number of tracks with different parameters, but all passing the same sector of interest. This technique reduces the number of friends for each sector to the physically relevant ones. The reason for this is that no sectors are stored in the friend list of the sector of interest when there was no track passing both sectors during simulation.

A rough estimation shows the advantage of this technique: since the total area of the friend sectors combined for each sector will be much smaller than the total area of any layer, the total number of hits can be calculated via the hit density per unit area. Consequently, when the total area of the friend list is e.g. $\sim 20\%$ of the total area of a layer, only $\sim 20\%$ of the hits will be in the friend list compared to the list of the whole layer (which would have to be checked without sectors). This greatly reduces the total number of combinations.

The cost of such a sectormap, where each sector and its friends are stored, is the time-consuming simulation of sufficiently many cases for each sector of interest and the sorting of the hits at the beginning of each event. However, the simulation can be done in advance without the need for real data events, so computing time is not an issue; and sorting is a task the time consumption of which grows linearly with the number of hits. The sectormap is calculated by the FilterCalculator module and is stored in a xml file. At the beginning of each run, the sectormap is imported into the track finder, whose first task is to sort every hit at the beginning of the current event into the sectormap.

The ideal number of sectors per sensor is a question of fine tuning; the more finely graded sectors reduce the total area of friends per sector (and therefore the number of possible hit combinations), but increase the memory consumption. This fine tuning will be part of future studies.

Another important point is the possibility of defining different settings for different momentum ranges which can be used for specialized searches for tracks of $\sim 50$MeV. The idea of different settings can be applied to special settings for curling tracks too, where the innermost and outermost sectors can be connected by curlers. There are some other advantages due to the introduction of a sectormap concerning the 2-, 3- and 4-hit filters
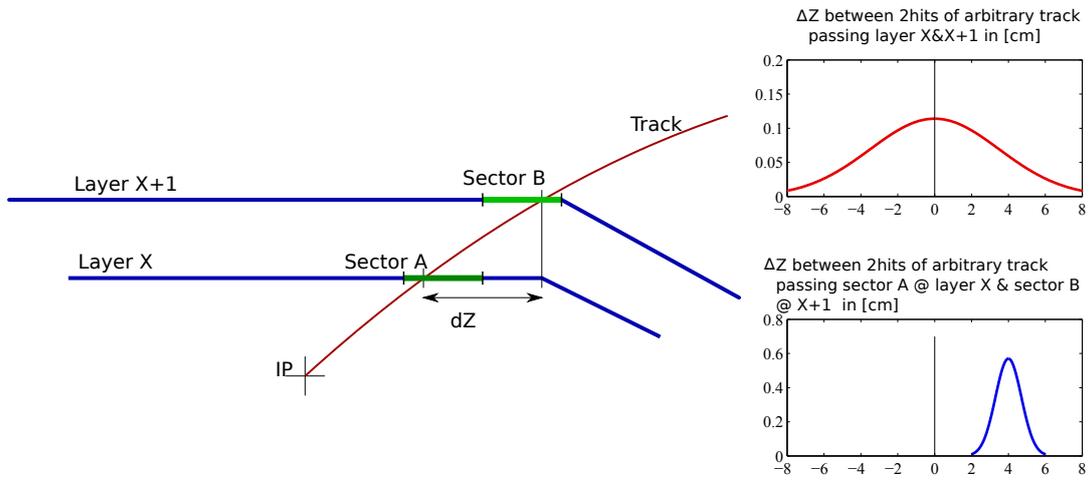
Figure 3.3.: This figure shows the difference between layer-wise and sector-wise cutoffs

and the cellular automaton as well. These will be discussed in the following sections.

### 3.7.3. 2-Hit filter: segfinder
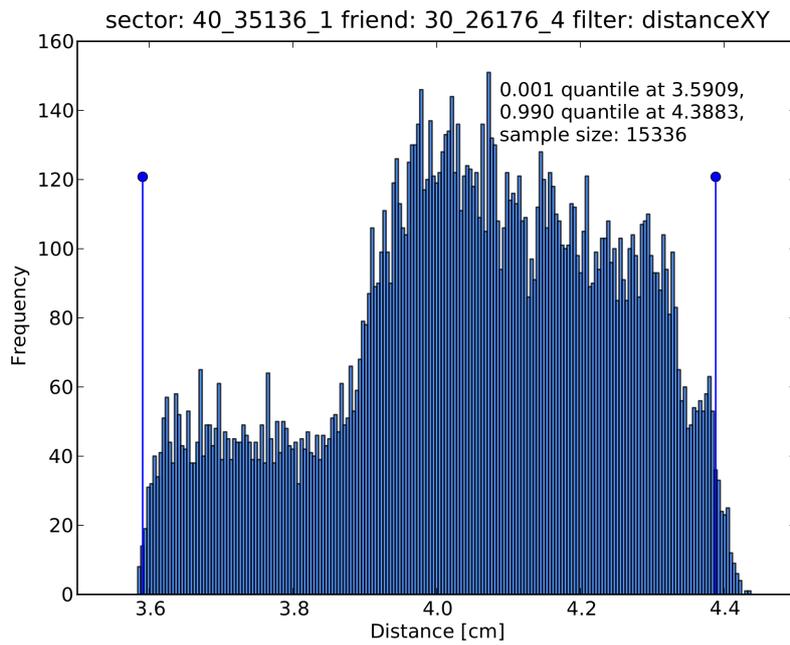


Figure 3.4.: This plot shows the result of the analysis of two hits passing both sectors. The quantiles are calculated and stored in the sectormap

After the sorting of hits into sectors, each hit in a sector gets a subset of possible
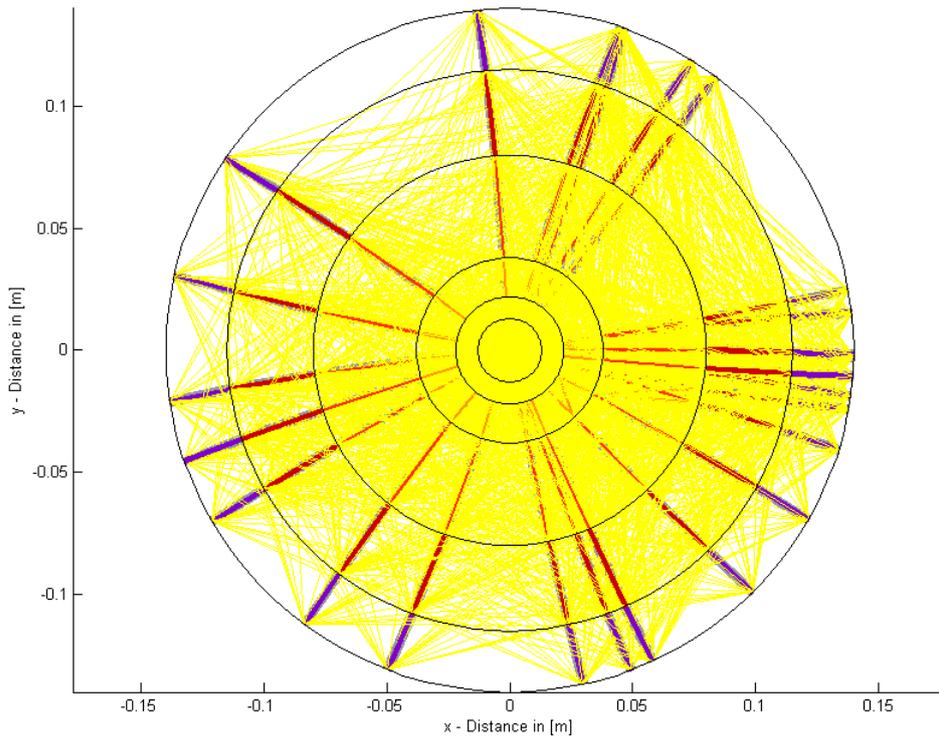
Figure 3.5.: an event with 25 tracks without applying segfinder filters before generating
segments. The yellow lines are arbitrary combinations of 2 hits among
neighboring layers

neighbor hits defined by the friendlists. These possible neighbors can now be checked by
tests combining two hits.

The same technique as for the sectormap is used to find definitions for useful two-hit
combinations: simulate a sufficiently high number of tracks and analyze the properties of
pairs of neighboring hits.

A possible filter of a 2-hit type is the distance between two hits. The analysis module
(FilterCalculator) calculates the distance between each neighboring pair of hits and stores
that information in a list. This list can be sorted and its quantiles can be used as cutoff
values for a distance filter. Only pairs of hits whose distance is between the lower and the
upper cutoff are allowed to form a segment, which is then stored for the following filter
steps. These cutoffs have to be stored in a lookup table, which is imported by the track
finder. There is another advantage to the sectormap method here: instead of storing
the filtering information for each filter layer-wise, that information is stored sector-wise,
which allows for a much more fine grained definition of the cutoff. Figure 3.5 and 3.6
illustrate the difference in combinatorics when applying segfinder-filters (3.6) or not (**??**).
Neighborfinder-filters and CA still applied (segments not passing the neighborfinder tests
are yellow, other colors indicate promising segments) However, the friendlist for each
sector makes it possible to store the cutoffs independently for each sector-friend-pair. The

69
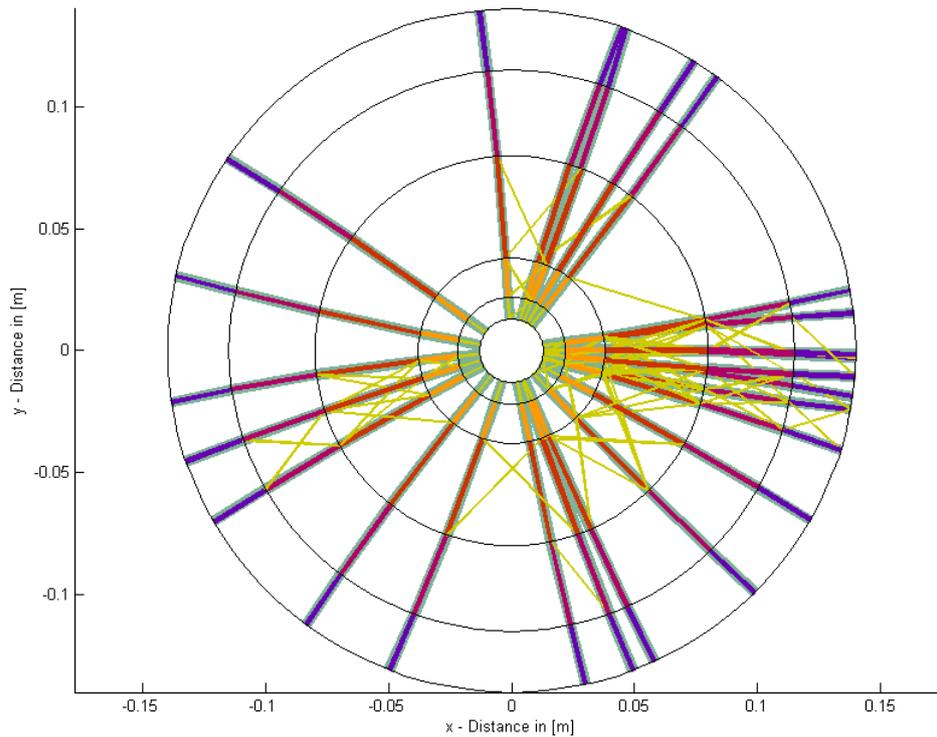
Figure 3.6.: Same event like 3.5 with normal filtering conditions, where hit-combinations to long or to short for the cutoffs are neglected

drawback of this method is a rapidly increasing memory consumption when increasing the number of sectors per sensors, so this is another point for fine tuning analysis to keep the memory usage within reasonable ranges. There are many ways to define a 2-hit-filter. The currently implemented versions can be found in the following table 3.1. If there are hits having no connected segment after this step, they can be removed since they are marked as background hits this way.

### 3.7.4. 3-Hit filter: neighborfinder

The main principle is the same as for the 2-hit filters, except that here neighboring segments are defined instead of neighboring hits. As mentioned above, only segments sharing a hit as a node will be compared to each other. Since not all pairs of hits within a pair of sector and friend become segments, the more complex calculation of e.g. an angle is less likely to become a bottleneck. Segments without friends in any direction after this step can be deleted since they will not be needed anymore. In addition to the angle between two segments, some other filters can be defined when using 3 hits. Their description can be found in the following table 3.2. The cutoffs used for this filter step are calculated by the FilterCalculator-module the same way as in the case of 2-hit filters.

| summary of 2-hit filters | |
|---|---|
| type | explanation |
| distance3D | measures the distance in 3D and filter filters the result by sector-dependent lower and upper cutoffs |
| distanceXY | the same as distance3D with using only 2D ($x - y$-plane) as dimensions |
| distanceZ | the same again, only distance in z-direction is measured and normed to its total length in $r - z$-direction |
| normedDist3D | distance3D/distanceXY reduces the fringe effects which penalize tracks passing the sector near the border |

Table 3.1.: This table summarizes the implemented 2-hit filters.
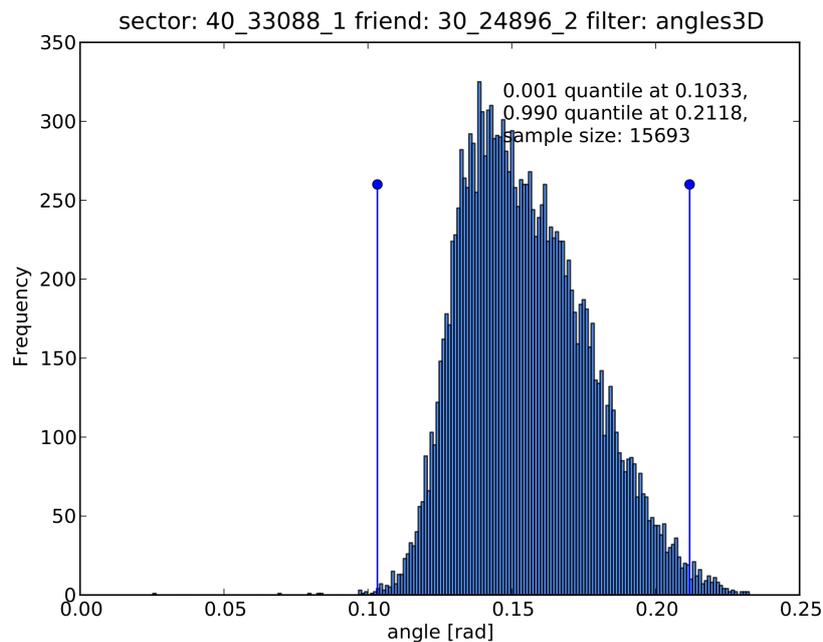


Figure 3.7.: This plot shows the result of the analysis of two neighboring segments. The quantiles are calculated and stored in the sectormap

71

| summary of 3-hit filters | |
|---|---|
| type | explanation |
| angles3D | measures the angle between two connected segments in 3D and filters the result by sector-dependent lower and upper cutoffs |
| anglesXY | the same as angles3D with using only 2D (x- and y-plane) as dimensions |
| anglesRZ | the same again, only angles in r-z-direction is measured |
| $\Delta$distanceZ | calculates the difference in distanceZ (see table 3.1) of both segments, only max cutoff is used |
| dist2IP | takes the three hits of the two segments to calculate the radius and the center of the circle described by them and checks the distance of the center to the IP. Filters hits away, where their corresponding circles miss the origin to far (only max cutoff is used) |

Table 3.2.: This table summarizes the implemented 3-hit filters.

### 3.7.5. Full chain filter: cellular automaton + TCC

A cellular automaton (CA) defines track segments as cells and calculates their state when propagating in time depending on the neighborhood of each cell. Mathematically, it can be described as a directed graph without loops with discrete space and discrete time evolution.

*Discrete space* means that the CA consists of segment-based cells whose state is defined by the state of its neighborhood. The neighborhood is not static but changes with each time step.

*Discrete time evolution* means that all cells start with state 0 and within each time step the same rule is applied to every cell: if there is an inner neighbor (found by 3-hit filters) with the same state, the current cell is allowed to wait for a state-upgrade in the upgrade step following the time step. If there is no neighbor that meets this condition, the state stays the same and the cell gets inactive and therefore can be removed from the checklist, which is checked each time step. However getting deactivated does not mean to be erased from the neighborhood-map since this information is needed for track candidate collecting.

Within the first time iteration, the innermost cells are getting deactivated since they don't have any inner neighbors. This means that within the second iteration, the active cells next to the innermost cells cannot find any cells with the same state and therefore get deactivated themselves. These iterations are repeated until no active cell is left. The final situation shows the following pattern: the outermost cells have the highest states; the state indicates the number of compatible inner neighbors forming a chain (which is equivalent to the length of the chain).

This allows for an easy way of collecting physically plausible track candidates by simply taking the outermost cells that have sufficiently high states as seeds and follow their

chain inwards to store them when the next inner neighbor has the state of the current cell minus one. When there is more than one acceptable neighbor, the chain is duplicated and followed independently.

An important point is that the cell determines which cell is an inner neighbor, not the hit. There can be situations when two cells of compatible states seem to form a chain with the same hit in the middle. However the state alone is not sufficient to be a neighbor. To fulfill this condition, the pair of segments has to get approved by the filters too. An illustration explaining this is figure 3.8. Considering the case when two outer cells share



layer
cell of state 2
cell of state 1

D is neighbour of A

C is **not** neighbour of A (wrong angle)

C & D are neighbours of B (angle and state okay
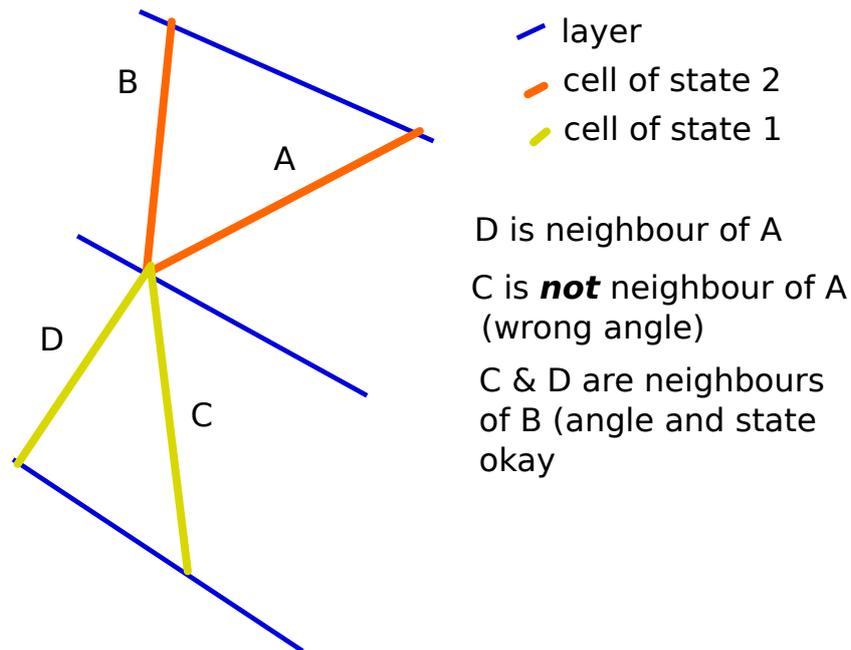
Figure 3.8.: The cellular automaton is more than just colored segments. The neighbourhood calculated by filters determines which segments are connected, not the shared hit.

an inner hit: the same hit has now two inner cells and forms the center of a cross. Only the nearly straight combinations of cells are neighbors, even if all the cells have the right state to be able to combined, since the kink of the non-straight connections would lead to physically unrealistic tracks. (kommt ein Bild dazu)

The working principle of a cellular automaton is derived from [30] and [31]. In our version, several changes have been made: first, in our case, the length and angle measurements have been decoupled from the CA since they can be calculated separately.

Next, since we have only 3 to 4 layers for track reconstruction, the efficiency of the CA is reduced since the outermost cells can only reach lower states. The higher the state of the outermost cell, the more probable is the case that the cell is the seed of a real track. Therefore, we use the fact that the position of the interaction point is roughly known

and can be treated as another track hit. Hits can form a special segment, called *virtual segment*, which enlarges the chain by one. Within the CA, these virtual segments are treated as regular ones, but are discarded when collecting the track candidates where only the normal segments are stored. The interaction point lies on a virtual sector which is stored as friend for compatible sectors in the innermost layer; this is needed for the storage of the cutoffs. Furthermore it is possible to define additional virtual segments carrying other versions of center hits, which could help reconstruct curling tracks by defining only two dimensional center hits, where the $z$ component is left arbitrary. Since curling tracks are not supported yet, this concept has not been implemented yet.

The logic of the CA can already be applied in the 3-hit-filters, where segments having at least one neighbor after that step can get a state upgrade to one. This reduces the number of total time steps needed by the CA by one and has been implemented in the 3-hit filter step.

### 3.7.6. 4-Hit filter: zigzag & $\Delta p_{\text{T}}$
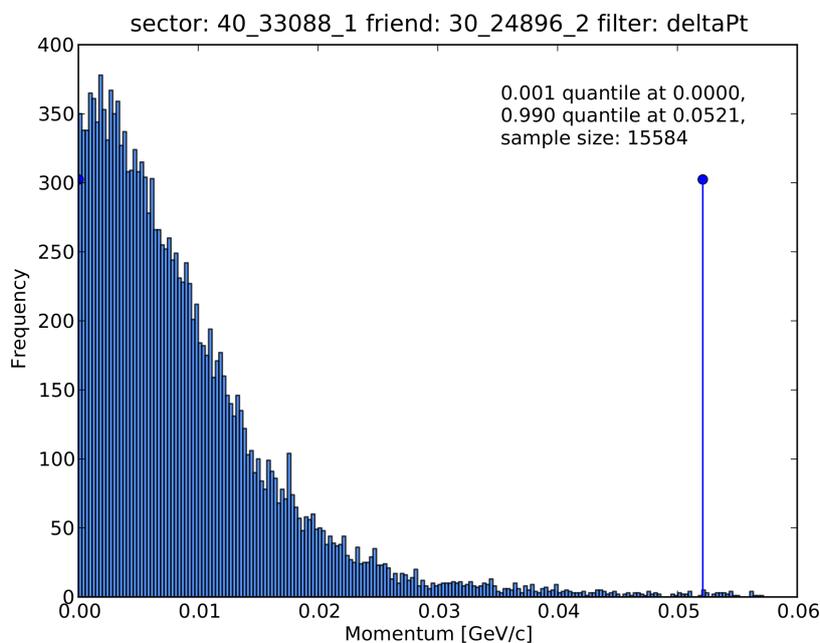


Figure 3.9.: This plot shows the result of the analysis of three segments in a chain. The momentum of two combined sectors is caculated and willThe quantiles are calculated and stored in the sectormap

After finishing the collection of track candidates, some other simple filters can be applied before processing the candidates in the more complex filters. The cutoffs are again derived using the FilterCalculator module, which stores the cutoffs in the sectormap. There are two useful possibilities of positioning this filter set among the chain of filter steps: it can either come after the CA, or between the 3-hit filters and the CA as a

| summary of 4-hit filters | |
| --- | --- |
| type | explanation |
| zigzag | takes three segments in a row and calculates the sign of the curvature for the two triplets. If they are different, they cannot belong to the same track |
| $\Delta p_{\mathrm{T}}$ | same method but calculates the whole curvature. To big curvature changes are highly improbable and can therefore be discarded |

Table 3.3.: This table summarizes the implemented 4-hit filters.

"superneighborfinder". The advantage of the first version is that it is compatible with the classic definition of cells in the cellular automaton. This has the drawback that the CA has to operate a higher number of cells since the track candidates that do not pass the 4-hit filter set are calculated needlessly by the CA. However the CA itself is quite fast (especially since the filters are decoupled from the CA algorithm) and the combinatorial problem on that point is not very pressing any more.

The superneighborfinder version in front of the CA would need an adaption of the definition of cells in order to work; this, however, has been successfully developed by R. Glattauer in [32]. This cell definition introduces 3-hit-cells (and higher ones) as a logical consequence of hits and 2-hit-cells. The advantage of the superneighborfinder version is that the combinatorial problem is ideally reduced through applying the simplest filters first and the most complex ones at last.

The following table 3.3 explains the filters implemented.

### 3.7.7. Full chain filter: Kalman filter

The Kalman filter is not included yet. Its working principle is discussed in 3.7.7 and its use within the track finder will be to reject more unrealistic track candidates and to calculate quality indicators needed for the neuronal network at the end of the track finding chain. It is planned to use the GENFIT module [33], which is already part of the basf2 framework and contains a Kalman filter algorithm.

### 3.7.8. Full trackset filter: neuronal network

There is a neuronal network of Hopfield type implemented at the end of the chain of filters. Its principle has been discussed in detail in the chapter 3.6 which is implemented as discussed there. It takes the quality indicators (QIs) of each track and their relation to each other to derive the optimal subset of non-overlapping track candidates. Since there is no Kalman filter providing the QIs included so far, a simpler definition of QI is used so far, namely the track length; longer tracks have higher quality than shorter ones).

The Hopfield network is activated whenever there are more than two tracks sharing at least one hit with each other. The case of two overlapping track candidates is solved by comparing their quality indicators; this can be done more quickly without the neuronal

network.

After the completion of that step, there is a final set of non-overlapping track candidates ready for further processing in other modules.

# 4. Performance and outlook

This chapter will give a summary of the current results produced by the low momentum track finder. The current state of the tool will be presented and further steps will be illustrated.

## 4.1. Definition of Efficiencies

Before starting with the interpretation of the results, a consistent definition of track and track candidate efficiency has to be made. Although similar, these two efficiencies are not the same. While the efficiency based on tracks (the real ones) allows us to see what rate of real tracks could be recovered, the track candidate efficiency allows us to see the ratio between real recovered tracks and ghost tracks. The track efficiency is in most cases more important, but for on-line track finding, the ratio ghost tracks to real ones can be important too. In the situation of Belle II, the track candidates of the CDC and of the low momentum track finder are used to extrapolate the trajectories to the PXD. This allows us to define regions of interests (ROI) which are stored since 90% of the data recorded by the PXD has to be deleted after a few seconds. Therefore the quality of the track candidates is important. If the ghost rate among the track candidates is too high, too many regions of interests are defined which forces us to reduce the size of the ROIs and therefore increases the probability to delete the information of the real hit. For this reason not only the total number of recovered tracks are important, but also the number of track candidates. In in table 4.1, these two points of view are summarized.

## 4.2. Simulation setup

For the generation of the events, geant4 [29] is used, in which preliminary definitions of the Belle II detector are implemented. The position of the windmill is not final, the radius and exact position of the ladders are still object of discussions. The version of the framework used for this thesis is revision 2777 from the march $17^{th}$, 2012. Since only low momentum tracks are studied, outer detector parts are not included in the simulation. Enabled are only the beam pipe, the PXD and the SVD. Pions, muons and electrons will be the most common particles within the tracking detectors. Electrons because of Bhabha-scattering and other background effects, muons and especially pions because of more interesting events. For this simulation muons are used as primary particles since their probability of decay within the silicon detector is rather small. To prevent curling tracks, the magnetic field is deactivated for radii $> 12.5$cm for tracks with $60$MeV/c$< p_T <70$MeV/c and for radii $> 15$cm for tracks with $70$MeV/c$< p_T <100$MeV/c. Note that the lower momentum

| track based efficiency | |
| --- | --- |
| classification | explanation |
| clean | means that a track candidate of at least 3 hits has been totally assigned to the same track |
| contaminated | means that at least 3 hits of a track candidate are assigned to the same track and less than 30% of the hits may be background hits or coming from other tracks |
| lost | a track whose hits did not form a clean or contaminated track candidate is defined as lost |

| track candidate based efficiency | |
| --- | --- |
| classification | explanation |
| clean | means that a track candidate of at least 3 hits has been totally assigned to the same track |
| contaminated | means that at least 3 hits of a track candidate are assigned to the same track and less than 30% of the hits may be background hits or coming from other tracks |
| ghost tracks | a track candidate where less than 70% of the hits could be assigned to the same track is defined as ghost track |

Table 4.1.: These tables list the definitions of track and track candidate based efficiencies. The difference lies in two points. First, two track candidates assigned clean or contaminated to the same track count only once for recovering the track in track based efficiency. In track candidate based efficiency, they count twice. Next, the total rate of clean, contaminated and lost tracks add up to the total number of tracks in the track based efficiency. In the track candidate based efficiency, the total number of track candidates can be higher than the total number of tracks, even when less than 100% of the tracks could be recovered. The ratio between ghost tracks and others is the interesting information at that point

| Key points from the detector setup in used framework version | | | | |
|---|---|---|---|---|
| layer # | radius | $u$ resolution | $z$ resolution | thickness |
| 1(PXD) | 14mm | 14.4μm | 15.9μm | 75μm |
| 2(PXD) | 21mm | 14.4μm | 24.5μm | 75μm |
| 3(SVD) | 38mm | 14.5μm | 45.1μm | 320μm |
| 4(SVD) | 80mm | 21.7μm | 69.3μm | 320μm |
| 5(SVD) | 115mm | 21.7μm | 69.3μm | 320μm |
| 6(SVD) | 140mm | 21.7μm | 69.3μm | 320μm |

Table 4.2.: Mean radii, combined hit resolution and sensor thickness of the used basf2 framework revision are listed here. $u$ resolution is the resolution in the $r$-$\phi$-plane.

setup means that the outermost layer can not be included in the track search since its radius of about 14cm lies beyond the magnetic field and therefore tracks become straight lines out there. Table 4.2 summarizes some of the relevant settings for the PXD and SVD.

## 4.3. Performance
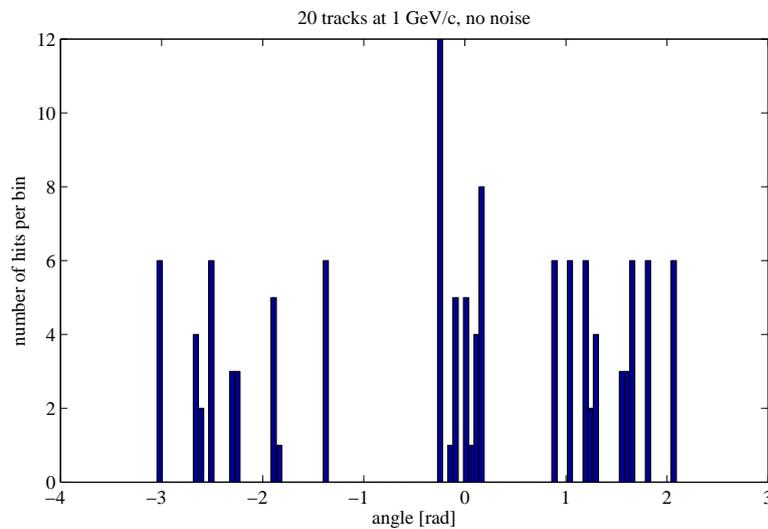


Figure 4.1.: Histogram of 20 tracks with 1 GeV/c, no noise. Bins with a height of 6 can cleanly recognized as hits belonging to the same track-

As an introductory example the figure 4.1 presents 20 tracks within a simplified 6 layer detector of 1 GeV/c each. Histogramming the results after a conformal transformation
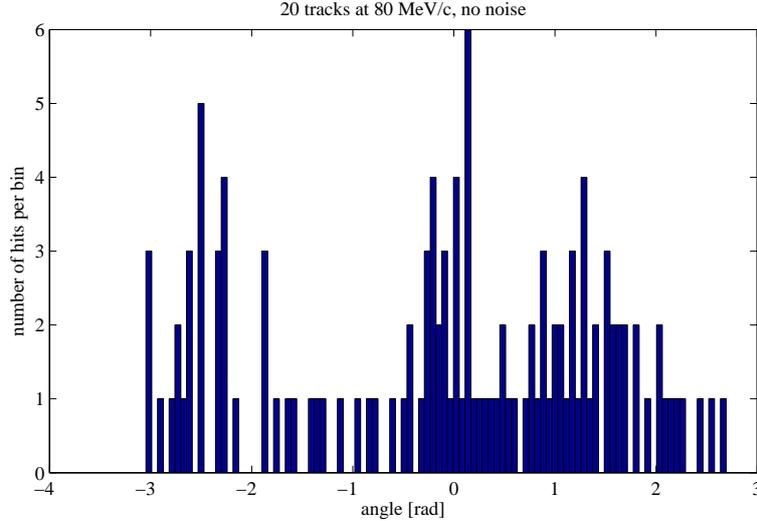
Figure 4.2.: Histogram of 20 tracks with 80 MeV/c, no noise. In this event histogramming does not work any more, since hits belonging to the same track are spread over several bins.

allows us to reconstruct most of the tracks by checking the height of the bin. While 1 GeV-tracks are still easy to find, 80 MeV tracks are much more complicated since the small radius leads to lines missing the origin farther. Looking at the figure 4.2 reveals the problem, the single tracks are not separable any more, which makes the method unusable for low momentum track finding.

Tables 4.3, 4.4, 4.5 and 4.6 summarize the results of the low momentum track finder. Every simulation included 1000 events, with 10 or 20 tracks, respectively. The theta range is $20° < \theta < 145°$, neither random nor realistic background is enabled. Since there is no Kalman filter implemented yet, the Hopfield network uses the track length as quality indicator. This choice for the quality indicator is not optimal but will be used until - Kalman filter is working.

The track finder has no difficulties dealing with that low occupancy situation. This can be seen when comparing the results using 10 or 20 tracks per event, where no relevant drop between the track based efficiency of both cases can be seen. However the total number of track candidates after TCC indicates that the cutoffs can be chosen more strictly since about 40–50% more track candidates are produced than there are real tracks. This conclusion is supported when normalizing the number of track candidates to the number of real tracks, which reveals a slight increase of roughly 15% when comparing the 10 to the 20 track case. Analyzing table 4.3 reveals that at least 4 Layers are needed to have the chance of recovering nearly 100% of the tracks again. This can be explained by the settings of the track candidate collector. Even random combinations of 2 segments have considerable chances of being accepted by the neighbourfinder, especially when the

80

| Momentum range (# of layers) | # of tracks | results post TCC, track based | | | |
|---|---|---|---|---|---|
| | | clean | cont. | lost | $\text{rec}_{tot}$ |
| 60MeV/c< $p_{\text{T}}$ <70MeV/c (3L) | 10T | 88.9% | 0.48% | 10.7% | **89.3%** |
| | 20T | 88.2% | 1.1% | 10.8% | **89.2%** |
| 60MeV/c< $p_{\text{T}}$ <70MeV/c (5L) | 10T | 99.6% | 0.1% | 0.3% | **99.7%** |
| | 20T | 99.1% | 0.3% | 0.6% | **99.4%** |
| 70MeV/c< $p_{\text{T}}$ <100MeV/c (4L) | 10T | 99.6% | 0.1% | 0.4% | **99.6%** |
| | 20T | 99.5% | 0.1% | 0.4% | **99.6%** |
| 70MeV/c< $p_{\text{T}}$ <100MeV/c (6L) | 10T | 99.6% | 0.1% | 0.3% | **99.7%** |
| | 20T | 99.4% | 0.2% | 0.4% | **99.6%** |

Table 4.3.: results after collecting track candidates, track based efficiency. Please note that if there is a clean and a contaminated TC assigned to the same track, this track is recognized as clean recovered

| Momentum range (# of layers) | # of tracks | results post TCC, TC based | | | |
|---|---|---|---|---|---|
| | | clean | cont. | garbage | total # tracks |
| 60MeV/c< $p_{\text{T}}$ <70MeV/c (3L) | 10T | 92.1% | 1.1% | 6.8% | 11,025 |
| | 20T | 82% | 2.3% | 15.7% | 24,469 |
| 60MeV/c< $p_{\text{T}}$ <70MeV/c (5L) | 10T | 85.9% | 7.4% | 6.7% | 14,120 |
| | 20T | 70.4% | 14% | 15.6% | 34,221 |
| 70MeV/c< $p_{\text{T}}$ <100MeV/c (4L) | 10T | 96.5% | 2.6% | 0.9% | 14,451 |
| | 20T | 92.5% | 4.7% | 2.9% | 29,687 |
| 70MeV/c< $p_{\text{T}}$ <100MeV/c (6L) | 10T | 89.2% | 6% | 4.9% | 14,568 |
| | 20T | 78.3% | 11.2% | 10.5% | 32,853 |

Table 4.4.: results after collecting track candidates, track candidate based efficiency

cutoffs are not chosen in a strictly enough. Therefore it would lead to a very high ghost rate, when allowing the track candidate collector to collect chains where the seed has only state 1. But increasing the minimal state for seeds to state 2 means, that a chain of 4 hits have to be found. This is not impossible since the interaction point counts as a track hit. However the quality of that estimation is not very good, and therefore some of these virtual segments connecting the interaction point with a hit of the innermost layer get dropped and reduce the maximally reconstructible number of tracks. So at very low momenta there is the choice between strict cutoffs, high minimal states for seeds or lower ones. The first two reduce the combinatorics, the latter one increases it. Combining strict cutoffs with lower seed states could be an interesting option and will be tested in future runs.

The following list lits the filter setups used for table 4.3, 4.4, 4.5 and 4.6.

- filters activated for 3L setup: distZ,distNorm3D, distDeltaZ, anglesRZ, deltaPt, zigZag

- filters activated for 4L setup: distXY, distZ, distNorm3D, distDeltaZ, angles3D,

| Momentum range (# of layers) | # of tracks | results final, track based | | | |
|---|---|---|---|---|---|
| | | clean | cont. | lost | $rec_{tot}$ |
| 60MeV/c< $p_\mathrm{T}$ <70MeV/c (3L) | 10T | 85.7% | 0.4% | 13.9% | **86.1%** |
| | 20T | 84.6% | 1% | 14.4% | **85.6%** |
| 60MeV/c< $p_\mathrm{T}$ <70MeV/c (5L) | 10T | 94.9% | 2.2% | 3% | **97%** |
| | 20T | 93% | 1.1% | 5.9% | **94.1%** |
| 70MeV/c< $p_\mathrm{T}$ <100MeV/c (4L) | 10T | 94.6% | 0.5% | 4.8% | **95.2%** |
| | 20T | 96.9% | 1% | 2.1% | **97.9%** |
| 70MeV/c< $p_\mathrm{T}$ <100MeV/c (6L) | 10T | 88% | 0.7% | 11.3% | **88.7%** |
| | 20T | 96% | 1.3% | 2.7% | **97.3%** |

Table 4.5.: final results, track based

| Momentum range (# of layers) | # of tracks | results final, TC based | | | |
|---|---|---|---|---|---|
| | | clean | cont. | garbage | total # tracks |
| 60MeV/c< $p_\mathrm{T}$ <70MeV/c (3L) | 10T | 98.1% | 0.5% | 1.4% | 8,734 |
| | 20T | 95.7% | 1.1% | 3.2% | 17,674 |
| 60MeV/c< $p_\mathrm{T}$ <70MeV/c (5L) | 10T | 98.2% | 1.2% | 0.7% | 9,470 |
| | 20T | 96.3% | 2.2% | 1.5% | 19,711 |
| 70MeV/c< $p_\mathrm{T}$ <100MeV/c (4L) | 10T | 99.4% | 0.6% | 0.1% | 9,523 |
| | 20T | 98.7% | 1.1% | 0.2% | 19,620 |
| 70MeV/c< $p_\mathrm{T}$ <100MeV/c (6L) | 10T | 98.7% | 0.7% | 0.6% | 8,923 |
| | 20T | 97.7% | 1.3% | 1% | 19,651 |

Table 4.6.: final results, track candidate based.

anglesxY, anglesRZ, deltaPt, zigZag

- filters activated for 5 & 6L setup: dist3D, distXY, distZ, distNorm3D, distDeltaZ, angles3D, anglesxY, anglesRZ, deltaPt, zigZag

## 4.4. Conclusion and outlook

The low momentum track finder for the silicon vertex detector of Belle II is under construction. Since major parts like the Kalman filter are still missing, neither background (neither random nor physically realistic) nor proper digitizers (reducing the hit resolution and introducing ghost hits) are considered so far, the results are of limited use. However, the proof of concept presented above, using a simplified layer setup, promises some robustness against high occupancy situations. This will be hte subject of further studies within the next view month after implementing the Kalman filter. More crucial is the fact that realistic readout frames of the PXD have not been tested so far, as substantiated research for background studies in the PXD has been started only recently and a relatively small amount of data is available yet. However, a recovery rate of nearly 90% in a low momentum 3 layer setup and nearly 100% at any other tested setup are promising. It

also shows that low momentum track finding with only three layers can be done without danger of losing the majority of the tracks. The project of off-line low-momentum track finding will take several months to finish. The knowledge won during that process will help to develop a similar track finding algorithm for on-line purposes, which will be the real challenge for that concept since an optimal compromise between speed and quality will have to be found.
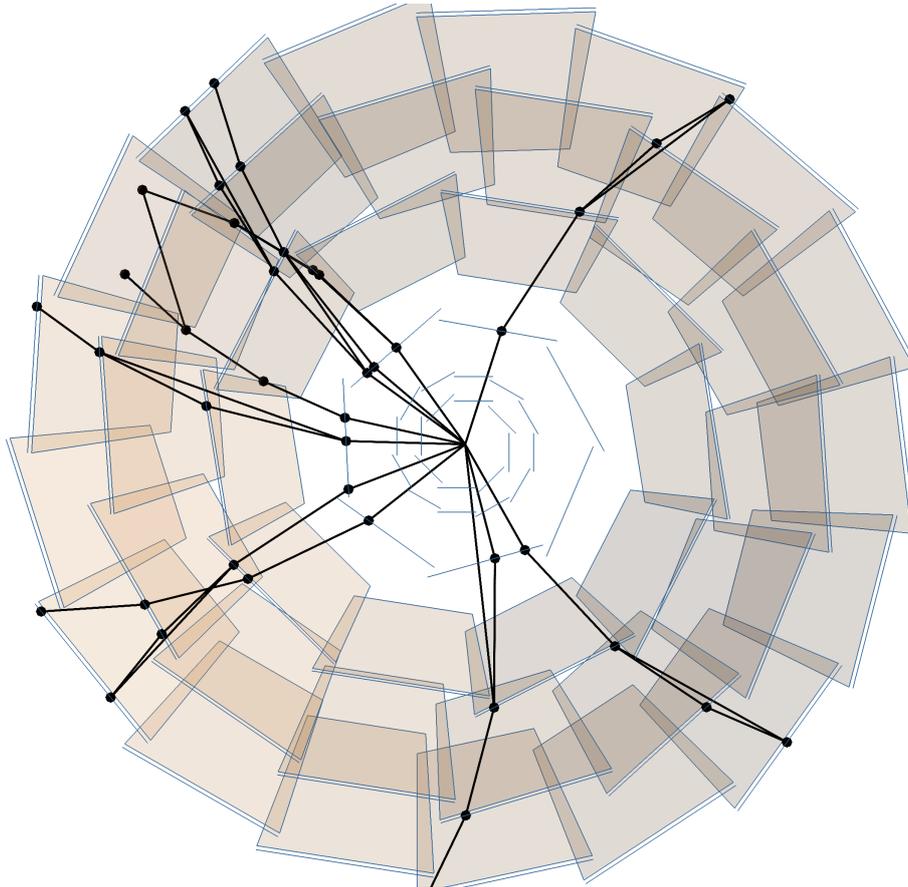


Figure 4.3.: This is a snapshot of an event after the segFinder, 51 segments have been found.
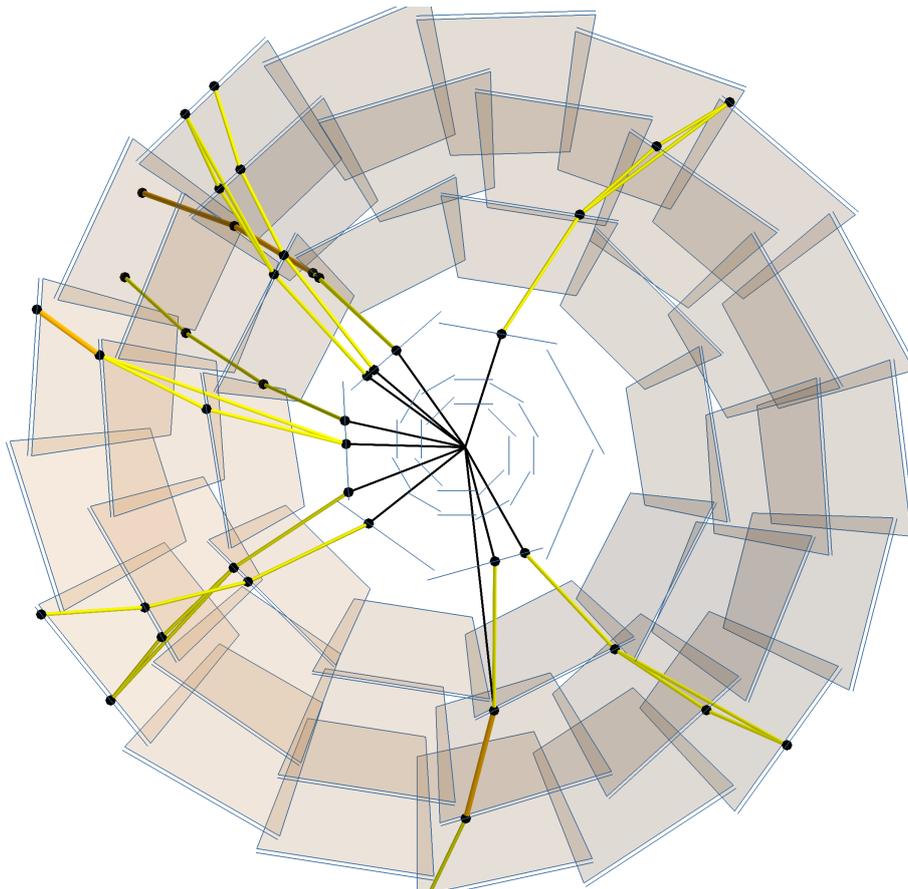
Figure 4.4.: This is a snapshot of an event after the neighbourFinder, which has reduced the number of segments to 49. The color of the segments represents the number of inner neighbours directly connected to the cell. black = no neighbour, yellow, = 1 nb, orange = 2 nbs, red = 3 nbs.
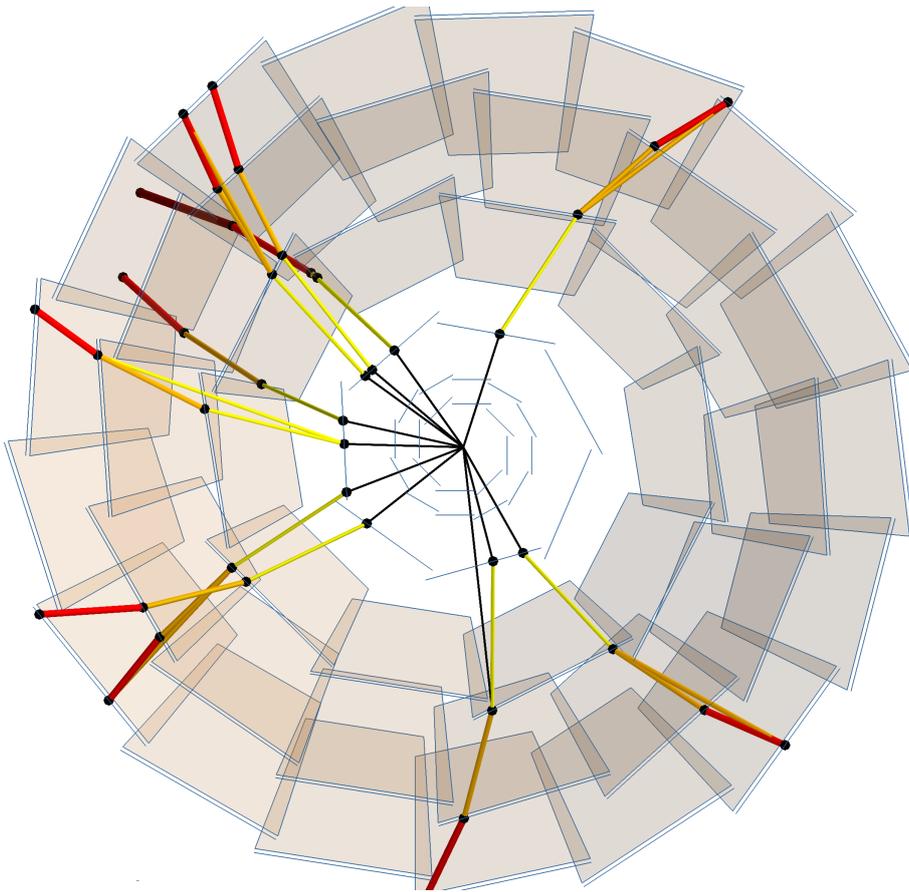
Figure 4.5.: This is a snapshot of an event after the cellular automaton. The color of the cells represents the state of the cell. black = 0, yellow, = 1, orange = 2, red = 3, dark red = 4, violet = 5
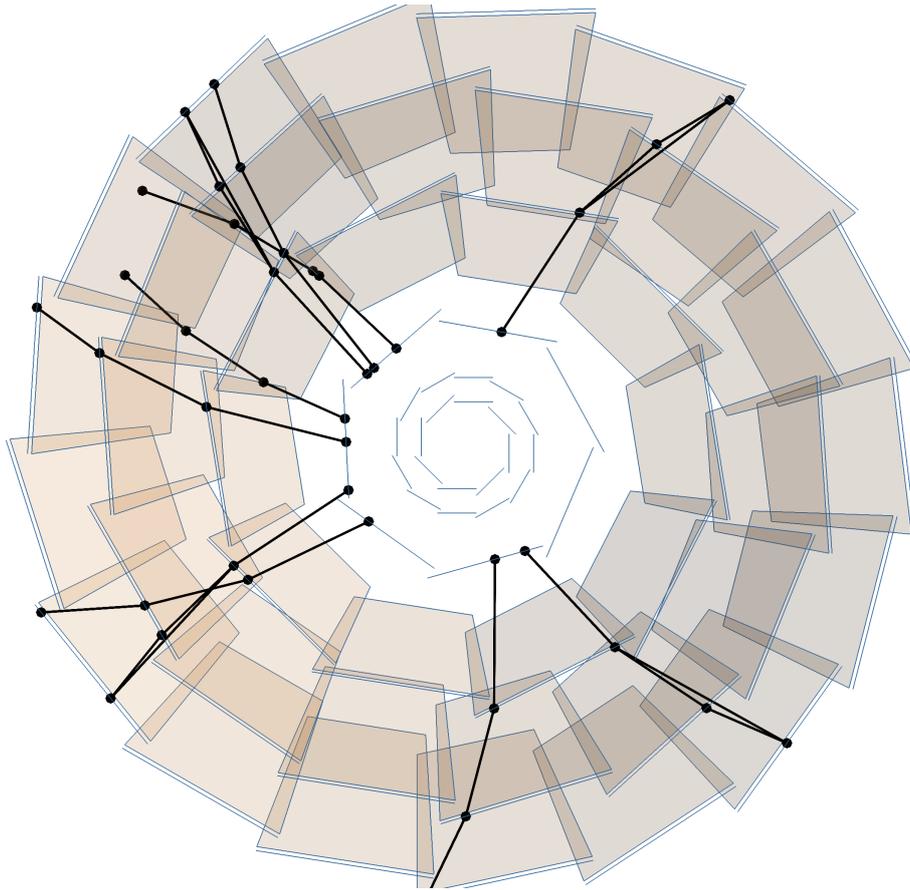
Figure 4.6.: This is a snapshot of an event after the track candidate collector (TCC), some of the 14 track candidates are overlapping and therefore the neuronal network has to chose the optimal subset.
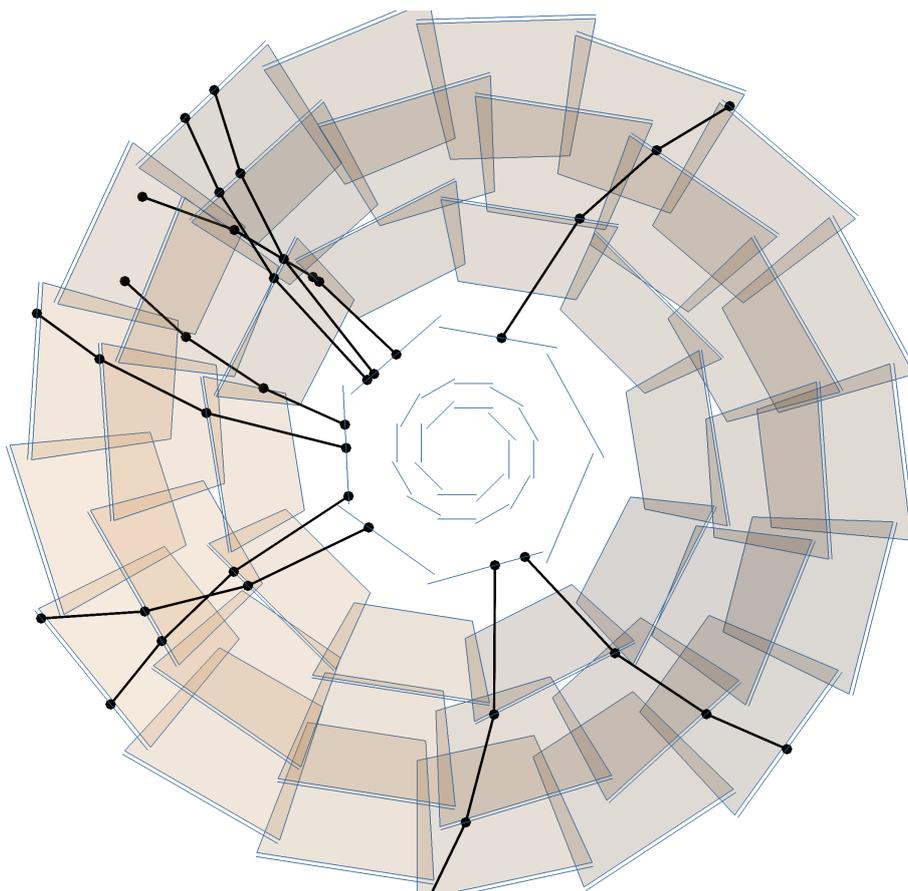
Figure 4.7.: This is a snapshot of an event after the Hopfield network. This is the final result of the track finder showing only 10 nonoverlapping tracks. In this case all tracks have been reconstructed.

# Bibliography

[1] K. Nakamura *et al.* (Particle Data Group), "2011 Review of Particle Physics," *J. Phys. G 37, 075021 (2010)*, 2011.

[2] C. Grupen *et al.*, *Handbook of Particle Detection and Imaging.* Springer, Berlin, 2011.

[3] Bergmann and Schaefer, *Lehrbuch der Experimentalphysik, Band 4, Teilchen.* Walter de Gruyter Verlag, Berlin/New York, 1992.

[4] K. Nakamura *et al.* (Particle Data Group), "2011 Review of Particle Physics, particle detectors of accelerators," *J. Phys. G 37, 075021 (2010)*, 2011.

[5] Z. Doležal and S. Uno (editors), *Belle II Technical Design Report*, Belle II collaboration, KEK, Oct, 2010, http://arXiv.org/abs/1011.0352v1. [Online]. Available: http://arXiv.org/abs/1011.0352v1

[6] A. Abashian *et al.*, "The Belle Detector," *Nuclear Instruments and Methods in Physics Research A 479 (2002) 117*, 2002.

[7] Group, TeVI, *Design Report Tevatron 1 project*, Tevatron Collaboration, Fermilab, 1984, http://inspirehep.net/record/211983. [Online]. Available: http://inspirehep.net/record/211983

[8] Frühwirth, Regler, Bock, Grote, Notz, *Data Analysis Techniques for High-Energy Physics.* Cambridge unversity press, Second edition, 2000.

[9] A. Xiao *et al.*, "Touschek effect calculation and its application to a transport line." Proceedings of PAC07, Albuquerque, New Mexico, USA, 2007. [Online]. Available: http://accelconf.web.cern.ch/AccelConf/p07/PAPERS/THPAN098.PDF

[10] K. Nakamura *et al.* (Particle Data Group), "2011 Review of Particle Physics, particle detectors of accelerators," *J. Phys. G 37, 075021 (2010)*, 2011. [Online]. Available: http://pdg.lbl.gov/2011/reviews/contents_sports.html

[11] ——, "2011 review of particle physics, passage of particles through matter," *J. Phys. G 37, 075021 (2010)*, 2011. [Online]. Available: http://pdg.lbl.gov/2011/reviews/contents_sports.html

[12] *ATLAS: Detector and physics performance technical design report. Volume 1*, ATLAS collaboration, Cern, 1999.

[13] C. Fabjan, "Teilchenphysik: Konzepte und experimentelle Tests," lecture notes, autumn 2010.

[14] C. Lippmann, "Particle identification," Ph.D. dissertation, GSI Helmholtzzentrum für Schwerionenforschung, Jun. 2011, http://arxiv.org/pdf/1101.3276. [Online]. Available: http://arxiv.org/abs/1101.3276

[15] Fabjan, C. W., Schopper, H. (editors), *Landolt-Börnstein New Series, Group I: Detectors for Particles and Radiation. Part 1: Principles and Methods.* Springer-Verlag Berlin Heidelberg, 2011, vol. Volume 21B1.

[16] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting," *Nucl.Instrum.Meth.*, vol. A262, pp. 444–450, 1987.

[17] T. Aushev, W. Bartel, A. Bondar, J. Brodzicka, T. Browder *et al.*, "Physics at Super B Factory," Belle II collaboration, Tech. Rep. KEK-REPORT-2009-12, 2010. [Online]. Available: http://arXiv.org/abs/1002.5012v1

[18] P.F. Harrison and H.R. Quinn, Editors, *The BABAR Physics Book.* SLAC Report 504,the BaBar Collaboration , October 1998.

[19] J. Christenson, J. Cronin, V. Fitch, and R. Turlay, "Evidence for the 2 pi Decay of the k(2)0 Meson," *Phys.Rev.Lett.*, vol. 13, pp. 138–140, 1964.

[20] M. Kobayashi and T. Maskawa, "CP Violation In The Renormalizable Theory Of Weak Interaction," *Prog. Theor. Phys. 49, 652 (1973)*, 1973.

[21] N. Cabibbo, "Unitary Symmetry And Leptonic Decays," *Phys. Rev. Lett. 10, 531 (1963)*, 1963.

[22] L. Wolfenstein, "Parametrization Of The Kobayashi-Maskawa Matrix," *Phys. Rev. Lett. 51, 1945 (1983)*, 1983.

[23] K. F. Chen *et al.*, "Observation of time-dependent $CP$ violation in $B^0 \to \eta' K^0$ decays and improved measurements of $CP$ asymmetries in $B^0 \to \phi K^0$, $K_S^0 K_S^0 K_S^0$ and $B^0 \to J/\psi K^0$ decays," *Phys. Rev. Lett. 98, 031802 (2007)*, 2007.

[24] K. Abe *et al.*, "Measurement of CP-violating asymmetries in $B^0 \to \pi^+ \pi^-$ - decays," *Phys. Rev. Lett. 89, 071801 (2002)*, 2002.

[25] K. Abe *et al.*, "Updated Measurement of $\phi_3$ with a Dalitz Plot Analysis of $B^+ \to D^* K^+$ Decay," 2008, 8 pages, 4 figures, 4 tables. Contributed to Rencontres de Moriond EW 2008. [Online]. Available: http://arXiv.org/abs/0803.3375

[26] G. Taylor, "The Belle silicon vertex detector: present performance and upgrade plans," *Nuclear Instruments and Methods in Physics Research A 501 (2003) 22–31*, 2003.

[27] Strandlie, Are and Frühwirth, Rudolf, "Track and vertex reconstruction: From classical to adaptive methods," *Rev.Mod.Phys.*, vol. 82, pp. 1419–1458, 2010.

[28] R. Frühwirth, "Selection of optimal subsets of tracks with a feedback neural network," *Comput.Phys.Commun.*, vol. 78, pp. 23–28, 1993.

[29] A. R. *et al.*, "GEANT 4 – a simulation toolkit," *Nuclear Instruments and Methods in Physics Research A 506 (2003) 250-303*, 2003.

[30] I. Abt, I. Kisel, S. Masciocchi, and D. Emelyanov, "CATS: A cellular automaton for tracking in silicon for the HERA-B vertex detector," *Nucl.Instrum.Meth.*, vol. A489, pp. 389–405, 2002.

[31] I. Abt, D. Emelyanov, I. Gorbunov, and I. Kisel, "Cellular automaton and Kalman filter based track search in the HERA-B pattern tracker," *Nucl.Instrum.Meth.*, vol. A490, pp. 546–558, 2002.

[32] R. Glattauer, R. Frühwirth, J. Lettenbichler, and W. Mitaroff, "Forward Tracking in the ILD Detector," 2012, proc. 13th Int. Linear Collider Workshop (LCWS '11), Granada, Spain, 26-30 Sept. 2011.

[33] C. Hoppner, S. Neubert, B. Ketzer, and S. Paul, "A Novel Generic Framework for Track Fitting in Complex Detector Systems," *Nucl.Instrum.Meth.*, vol. A620, pp. 518–525, 2010.

# A. Abstract - german version

Die Diplomarbeit hatte zum Ziel eine Methode zur Rekonstruierung von niederenergetischen Teilchenspuren im Silicon Vertex Detector des Belle II Experiments zu entwickeln. Das Belle II Experiment ist der Nachfolger des Belle Experiments, welches zur Aufgabe hatte, die CP-Verletzung im B-Meson-System nachzuweisen. Die Experimente BaBar und Belle waren daher als "B factories" designed, um eine zum Nachweis ausreichende Menge an B-Meson-Zerfällen messen zu können. Der Nachweis war erfolgreich, die vom Standardmodell der Teilchenphysik vorhergesagte CP-Verletzung konnte nachgewiesen werden. Doch die CP-Verletzung des Standardmodell der Teilchenphysik erklärt nicht die Asymmetrie zwischen Materie und Antimaterie, die im Universum beobachtet werden kann. Es gibt mehrere Theorien welche einige der Schwachpunkte des Standardmodells nicht haben. Beispiele dafür sind Supersymmetry oder die String Theorie. Doch ihre prognostizierten Abweichungen vom Standardmodell in der CP-Verletzung im B-Meson-System ist so klein, dass die vorhandene Datenmenge von BaBar und Belle nicht ausreicht, um diese nachzuweisen. Belle II wird nun etwa 40 mal so viele B-Meson-Zerfälle produzieren, wie Belle produziert hat. Dies erlaubt, Abweichungen vom Standardmodell zu finden. Obwohl sich die Kollisionsenergie von 10,58 GeV (Produktionsenergie des $\Upsilon 4(S)$) beim Beschleuniger vom Upgrade des Belle Beschleunigers − KEKB − nicht geändert hat, wird dennoch ein neuer Detektor benötigt, welcher mit der drastisch erhöhten Kollisionsrate umgehen kann. Der neue Detektor erlaubt auch die Rekonstruktion von sehr niederenergetischen Teilchenpuren, welche nur einen Transversalimpuls von 50 MeV/c oder mehr haben. Es gibt einige Zerfallskanäle des B-Mesons welches ein niederenergetisches Pion erzeugen, welche mit dem alten Detektor kaum rekonstruiert werden konnte. Spuren mit diesem Impuls haben einen Krümmungsradius von nur 11cm und erreichen daher nicht die Central Drift Chamber, welche in Belle zur Rekonstruktion von Spuren benützt wurde. Daher ist eine Rekonstruktionsmethode notwendig, welche im Silizium-Teil des Tracking Detektors arbeitet. Der Silicon Vertex Detector hat 4 Lagen von Silizium Streifen Detektoren, welche für die Rekonstruktion zur Verfügung stehen. Der in dieser Arbeit vorgestellte Trackfinder ist darauf optimiert Teilchenspuren mit nur 3 oder 4 Hits zu finden (um eine durch das Magnetfeld gekrümmte Spur zu finden, sind aus Mathematischer Sicht mindestens 3 Hits notwendig) − auch unter Situationen mit beträchtlichem Hintergrund. Zu diesem Zweck werden mehrere Filtermethoden kombiniert und im Rekonstruktionsprozess aneinander gereiht, da komplexere Techniken mit der zu erwartenden Hitmenge zu lange brauchen würden. Daher werden einfache, aber schnelle Algorithmen zur Hitreduktion verwendet, um das kombinatorische Problem für die aufwändigeren Algorithmen zu reduzieren. Der Trackfinder ist folgendermaßen aufgebaut:

Zuerst werden Hits in eine Sectormap einsortiert. Dazu werden die Sensoren des Detektors

in Sektoren unterteilt und durch Simulationen festgestellt, welche Kombinationen von Sektoren in einem Track vorkommen können. Die Sectormap enthält für jeden Sektor die Information, welche weiteren Sektoren zu ihr kompatibel sind, also mit ihr kombiniert werden können. Durch das Einsortieren der Hits wird die Anzahl an möglichen Kombinationen für die Hits bereits beträchtlich reduziert, da nur Hits aus kompatiblen Sektoren kombiniert werden.

Der nächste Schritt sind dann einfache Filter, in denen 2 Hits miteinander kombiniert werden. Dabei werden solche nach Länge gefiltert. Hitkombinationen welche diese Tests bestehen, werden als Segmente gespeichert Danach folgen 3-hit-Filter, welche Winkel und vergleichbare Regeln an miteinander verknüpfte Sektoren anwenden. Sektorkombinationen welche diese Filter bestehen werden auch vermerkt. Diese Information benützt dann der Zelluläre Automat, welcher den Segmenten Zustände nach interner Logik vergibt. Diese Zustände sagen dann aus, wie viele kompatible Segmente in einer Kette zusammen hängen Haben diese eine bestimmte Mindestlänge, werden sie als Track Candidates gespeichert. Track Candidates werden dann noch durch weitere Filter (4-Hit-Filter, Kalman Filter, noch nicht implementiert) sowie einem Hopfield Netzwerk geschickt. Alle Track Candidates, die diese Tests überstehen werden dann abgespeichert für die weitere Verarbeitung.

Diese Arbeit ist Teil eines größeren Projektes, welches neben einem Offline Trackfinder auch eine Online Version vorsieht. Diese wird dann zur Datenreduktion eines weiteren Detektors benützt, wo eine hochqualitative Track Candidate-Bewertung von Vorteil ist.

# B. Abstract - english version

The purpose of this thesis was to develop a pattern recognition tool for recognizing low momentum particle tracks within the silicon vertex detector of the Belle II experiment. The Belle II experiment is the successor of the Belle experiment and is a B factory of the second generation. B Factories are particle accelerators specialized to produce B mesons at high rates using the resonance energy of 10.58 GeV of the $\Upsilon 4(S)$ Meson. Since the amount of data produced by the first generation of B factories is enough to confirm the Standard model of particle physics, it is not enough to show discrepancies to possible successors of the Standard model, namely Supersymmetry or String Theory. These successors are needed since the Standard model is not explaining the currently known level in the baryon asymmetry. To refine the measurements enough to see these discrepancies, the Belle II experiment will deliver 40 times more data since the first generation which would need several decades to achieve this amount The new detector is able to deal with the increased physics rate of the upgraded accelerator. One of the features will be the possibility to reconstruct low momentum tracks down to 50 MeV/c. This low momentum track finder is presented in this thesis. Its approach is to combine filters with different levels of complexity. The simple ones will reduce the level of combinatorics to allow the complex ones to remove realistically looking random combination of hits. This allows us to reconstruct tracks with only three hits within the detector − even when there is considerable background. The low momentum track finder presented in this thesis is a part of a greater project in which next to the off-line track finder there will be an on-line version too. The on-line track finder will be used for data reduction since the innermost detector produces too much data to be able to store it. Using more sophisticated filter algorithms than typical triggering techniques improves the success rate.

# C. Acknowledgements

First of all, I want to thank Christian Fabjan, Rudolf Frühwirth and Christoph Schwanda for providing me with the opportunity to work on such an interesting project like the Belle II experiment. They helped me with their guidance and untiring support to complete this thesis and I am looking forward to a fruitful collaboration during this ongoing project (and beyond). Furthermore I want to thank Christian Thomay for the help and support he gave me during my studies, without him I wouldn't have ever managed to finish them. I also want to thank Moritz Nadler, Monika Madl, Robin Glattauer and Felicitas Thorne for proofreading my thesis and their ongoing help during the last few months, they were a great support.

I want to thank my family for without their environment I wouldn't be what I am now. Foremost I want to thank my father Anton Lettenbichler since he sparked my interest in natural sciences. All over the years he has been a great friend and mentor. I also want to thank my Cousin for supporting me when I needed him most. At last I also want to thank my sister Elisabeth for being a perfect example how I don't want to be.

# Jakob Lettenbichler

## Education

| | |
|---|---|
| 2010 – 2012 | **Diploma thesis**, *Institute for high energy physics*, Vienna. |
| 2002 – 2012 | **Diploma study of Physics**, *University of Vienna*. |
| 1994 – 2002 | **Matura**, *Bundesrealgymnasium Schloss Wagrain*, Vöcklabruck. |
| 1990 – 1994 | **Volksschule**, *Vöcklamarkt*. |

## Diploma thesis

| | |
|---|---|
| title | *Pattern Recognition at the Silicon Vertex Detector of the Belle II experiment* |
| supervisor | Rudolf Frühwirth |
| description | Finding tracks in realtime of low energy particle tracks in a detector of a electron-positron collider can be a challenging task. Our approach for developing a low momentum track finder, is to review, adapt and combine several promising techniques. |

## Nonscientific experience

### Vocational

| | |
|---|---|
| Nov 2006–Jan 2010 | **shop assistant**, *Saturn Electro-Handelsges.m.b.H Wien Gerngross*, Vienna. Consulting and selling of TV's, beamers and satellite recievers |
| Feb 2010 | **shop assistant**, *Ski Sport Salvista*, Itter. Consulting, preparing and selling skis, boots and snowboards. |
| Jan 2011–Feb 2011 | **shop assistant**, *Ski Sport Salvista*, Itter. Consulting, preparing and selling skis, boots and snowboards. |

## Languages

| | | |
|---|---|---|
| German | **first language** | |
| English | **fluent** | |
| French | **basic knowledge** | *6 years of active practicing at school, idle since then* |

## Computer skills

| | | | |
|---|---|---|---|
| Programming | C, C++, Python & Octave (Matlab) | Markup | LaTeX & HTML |
| Office Tools | MS Word, Excell or similar | OS | Windows, Linux |
| Multimedia | Gnuplot, Gimp, Adobe Photoshop & Premiere | | |

## Interests

| | |
|---|---|
| stories | includes reading, watching movies, storytelling, writing (short and long stories, screenplays, ...) and movie making |
| sports | Volleyball, climbing, jogging, ... |
| food | eating and cooking good meals |