# d$E$/d$x$ Particle Identification and Pixel Detector Data Reduction for the Belle II Experiment

## Christian Pulvermacher

## Diplomarbeit

an der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

*Referent: Prof. Dr. M. Feindt*
*Korreferent: Prof. Dr. G. Quast*

*Institut für Experimentelle Kernphysik*

Juni 2012

# Deutsche Zusammenfassung

Diese Diplomarbeit beschäftigt sich mit zwei verschiedenen Themen, bei denen in Detektoren des Belle-II-Experiments verschiedene Teilchen voneinander getrennt werden. Zum einen sollen im Pixeldetektor Cluster von langsamen Hadronen selektiert werden, um zu verhindern, dass diese Daten verlorengehen, zum anderen wird eine allgemeine Methode zur Identifizierung geladener Teilchen in den Tracking-Detektoren beschrieben. In beiden Fällen wird der Energieverlust der Teilchen im Detektormaterial eine besondere Rolle spielen.

Ein Hauptziel der beiden B-Fabriken Belle und BaBar war die Messung der $CP$-Verletzung im B-Meson-System, was 2001 beiden Experimenten gelang und für deren akkurate Vorhersage Makoto Kobayashi und Toshihide Maskawa 2008 den Physik-Nobelpreis erhielten. Zudem eignen sich B-Fabriken aber auch für Präzisionsmessungen physikalischer Parameter und die Suche nach Abweichungen vom Standardmodell der Teilchenphysik. Beide Experimente wurden inzwischen abgeschlossen, zuletzt Belle im Jahr 2010. Neue B-Physik-Projekte sind in Planung, die an ihre Erfolge anknüpfen sollen.

Eines dieser Projekte ist Belle II, ein Upgrade des Belle-Experiments am Beschleuniger KEKB in Tsukuba, Japan. Für das Belle-II-Experiment wird die Luminosität des Beschleunigers um den Faktor 40 erhöht, so dass nach dem voraussichtlichen Start des Experiments in 2016 im Verlauf eines Jahres eine höhere Luminosität aufgezeichnet werden soll als die Gesamtmenge, die von Belle geliefert wurde. Dies erlaubt wesentlich präzisere Messungen als sie mit Belle oder BaBar möglich waren, erfordert aber auch einige Verbesserungen an Detektor und Rekonstruktionssoftware. So wird für die Vertexrekonstruktion bei Belle II eine Kombination aus einem 4-lagigen Siliziumstreifendetektor (SVD) und einem 2-lagigen Pixeldetektor verwendet, so dass eine sehr gute Vertex-Auflösung zu erwarten ist. Auch die Teilchenidentifizierung wurde beträchtlich verbessert, hier wurden die Schwellen-Tscherenkov-Detektoren durch RICH-Detektoren (Ring-Imaging Cherenkov) ersetzt, welche den Tscherenkov-Winkel und damit die Geschwindigkeit der sie durchfliegenden Teilchen messen können.

## Pixeldetektor-Datenreduktion

Aufgrund der sehr hohen Luminosität ist für das Belle-II-Experiment eine wesentlich höhere Belastung durch Untergrundprozesse zu erwarten, die sich zum einen in einer möglicherweise kurzen Lebensdauer der Detektorkomponenten (u. a. durch Absorption von Neutronen), als auch einer hohen Datenmenge, insbesondere für die Tracking-Detektoren, niederschlägt. Der zweilagige Pixeldetektor (PXD) befindet sich direkt außerhalb des Strahlrohrs und ist dem Wechselwirkungspunkt somit am nächsten. Zusammen mit seiner mit 20 $\mu$s recht langen Integrationszeit – innerhalb derer ein Pixel Ladung ansammeln kann, bevor er ausgelesen wird – macht ihn dies besonders anfällig für in Untergrundprozessen erzeugte Teilchen, hauptsächlich durch Streuung innerhalb eines Strahlpakets sowie durch QED-Prozesse. Aktuellen Simulationen zufolge werden durch diese Teilchen bei jedem Auslesen etwa 1 % der

Pixel aktiv sein, was einer Datenmenge von 333 kB je Ereignis entspricht. Da dies auch nach Berücksichtigung der Vorselektion durch Trigger die maximal zu speichernde Datenrate überschreitet, ist eine Reduktion der Daten nötig. Bedenkt man, dass je Ereignis nur eine Handvoll Pixel durch Nicht-Untergrund-Teilchen aktiviert werden, ist dies äquivalent zu einer Reduktion des Untergrunds.

Hauptstrategie zur Untergrundreduktion ist die Extrapolation der Teilchenspuren aus den weiter außen liegenden Tracking-Detektoren, die aufgrund ihrer Position und höheren Ausleserate in der Lage sind, wesentlich sauberere Daten zu liefern. Diese Methode verwirft sämtliche Pixel, die sich nicht in der Umgebung der extrapolierten Spuren befinden. Dies reduziert die finale Datenmenge drastisch, hat allerdings den Nachteil, dass für Spuren niederenergetischer Teilchen, welche möglicherweise nur wenige Treffer in anderen Detektoren erzeugen, eventuell gar keine Pixeldaten zur Verfügung stehen. Insbesondere betrifft dies Spuren von langsamen Pionen $\pi_s$, welche beim Zerfall von $D^*$-Mesonen erzeugt werden. Da die $D^*$-Masse nur wenig höher ist als die Summe der Massen seiner Zerfallsprodukte, ergibt sich für das $\pi_s$ im Zerfall $D^*(2010)^\pm \to \pi^\pm_s D^0 [\to K^\mp \pi^\pm]$ eine ausgesprochen geringe kinetische Energie.

Da Tranversalimpulse $p_t < 60\,\text{MeV}$ nicht mehr ausreichen, um zuverlässig die vierte (und letzte) Lage des Silizium-Vertex-Detektors (SVD) zu erreichen, würden für eine signifikante Anzahl der langsamen Pionen – in Praxis etwa 15 % – keine Pixeldaten zur Verfügung stehen. Da dies die Rekonstruktion der $D^*$-Mesonen sowie anderer Zerfälle negativ beeinflusst, wird in Kapitel 5 eine Methode vorgestellt, um die diesen Teilchen zuzuordnenden Pixel zu retten.

Die Methode verwendet nur die Daten des Pixeldetektors und macht sich dabei zunutze, dass die einzelnen Pixel zu Clustern zusammengefasst werden können. Mithilfe der existierenden Simulationsprogramme für Untergrund- und Signalprozesse sowie der PXD-Digitialisierungs- und Clustering-Software wurden für die von simulierten Teilchen erzeugten Cluster verschiedene Variablen definiert. Diese enthalten unter anderem die Position und Ausdehnung des Clusters in verschiedenen Dimensionen und die minimale, maximale und gesamte Pixelladung innerhalb des Clusters. Diese dienen als Eingabe für ein künstliches neuronales Netzwerk, welches darauf trainiert wird, zwischen Clustern von langsamen Pionen und Untergrundteilchen (hochenergetische Elektronen) zu unterscheiden.

Die resultierende Trennung lässt sich bewerten durch den Anteil bewahrter Signalcluster (Signaleffizienz) sowie den Anteil verworfener Untergrundcluster (Untergrundreduktion). Trägt man diese Größen für verschiedene Schnitte auf die Netzwerk-Ausgabe übereinander auf, erhält man einen Graphen wie in Abbildung 1.

Es ist zu erkennen, dass die Methode sehr gut zwischen langsamen Pionen und Untergrundteilchen unterscheidet, so dass Pixeldaten für die $\pi_s$ aus $D^*$-Zerfällen gerettet werden können, ohne die Gesamtdatenmenge übermäßig zu erhöhen. Verschiedene Robustheitstests bestätigen die Wirksamkeit der Methode auch wenn nur die Ladungsvariablen als Eingabe verwendet werden, oder die einzelnen Pixelladungen breit verschmiert werden.

Dies sind gute Voraussetzungen dafür, das Netzwerk in der Praxis anzuwenden, auch wenn die zu erwartende Genauigkeit der ausgelesenen Pixelladung noch nicht feststeht. Da das Netzwerk hochperformant und möglichst nah am Pixeldetektor sein sollte, liegt es nahe, das Netzwerk auf FPGAs (Field Programmable Gate Arrays) zu implementieren, welche auch für die Tracking-Datenreduktion die Signalcluster auswählen.
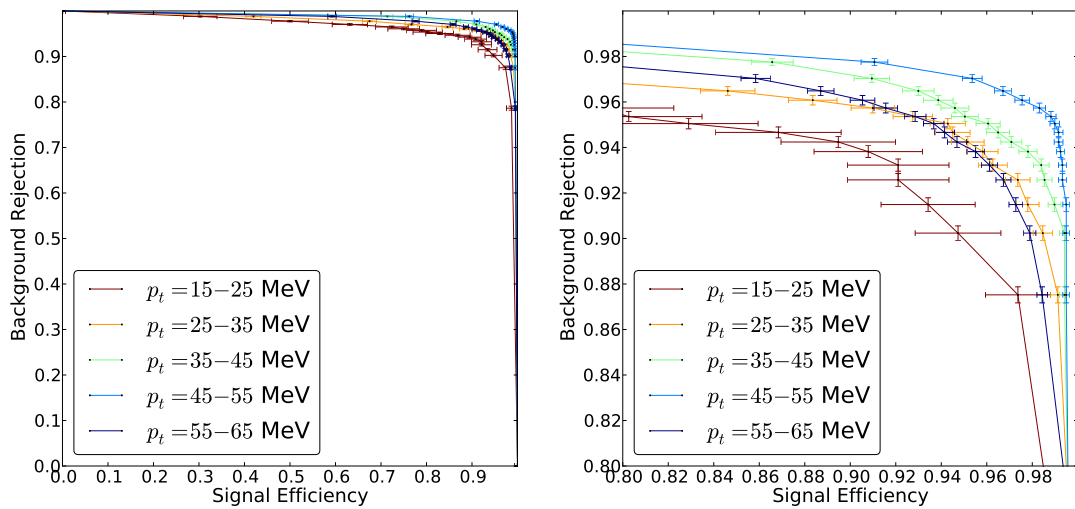
Abbildung 1.: Untergrundreduktion über der Signaleffizienz für verschiedene Schnitte auf die Netzwerkausgabe, mit verschiedenen Farben für unterschiedliche Transversalimpulse der zu selektierenden Pionen. Die rechte Seite zeigt einen Ausschnitt in welchem beide Achsen bei 80 % beginnen.

## Teilchenidentifizierung über den spezifischen Energieverlust

Das zweite Thema dieser Diplomarbeit ist die Teilchenidentifizierung mit den Trackingdetektoren, um so für jede Spur eine möglichst akkurate Vorhersage zu erhalten, ob es sich um ein geladenes Pion, Kaon, Proton, Elektron oder Myon handelt. Dazu lässt sich der Energieverlust der Teilchen im Detektor verwenden, welcher für Teilchen verschiedener Masse und selbem Impuls stark unterschiedlich sein kann. Mithilfe der durch die Tracking-Software bereitgestellten Impulsmessung und den Ladungssignalen in den Detektoren, welche den Energieverlust durch Ionisation je Detektorzelle wiederspiegeln, lassen sich so zweidimensionale Wahrscheinlichkeitsdichtefunktionen konstruieren. Diese lassen sich anschließend mit der Likelihood-Ratio-Methode mit Messwerten einer zu klassifizierenden Spur vergleichen.

Es stellt sich hierbei heraus, dass es vorteilhaft ist, neben den Daten der Driftkammer auch die der Siliziumdetektoren, im Speziellen des SVD, zu verwenden, was zu einer starken Verbesserung für Impulse unterhalb von 1.5 GeV führt. Ebenfalls von Vorteil ist die Verwendung der einzelnen Messwerte anstatt eines getrimmten Mittelwertes (*truncated mean*), welcher einen Großteil der für eine Spur gesammelten Information verwirft.

Abbildung 2 zeigt die resultierende Trennung zwischen den obengenannten fünf Sorten geladener Teilchen für eine Fehlklassifizierungsrate von 5 %. Es ist ersichtlich, dass sich Elektronen über einen großen Impulsbereich beinahe eindeutig identifizieren lassen, aber auch für Impulse unter 1 GeV ist eine Selektion teilweise weiterhin möglich. Für Pionen und Kaonen ergibt sich von etwa 900 MeV bis zu niedrigsten Impulsen eine sehr hohe Effizienz bei der gewünschten Reinheit. Lediglich für sehr niedrige Impulse um 100 MeV sinkt die Effizienz für Kaonen, da ihre Anzahl dort stark abnimmt.
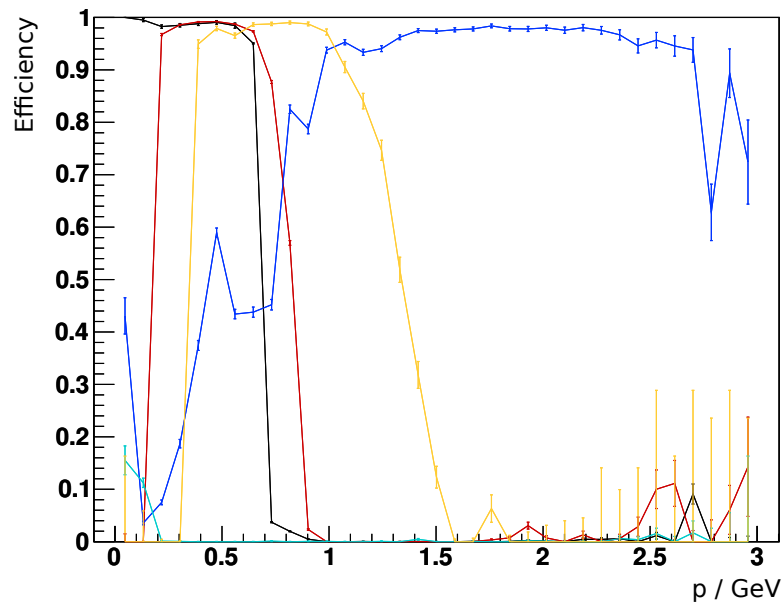
Abbildung 2.: Selektionseffizienz über Teilchenimpuls für eine Klassifikation mit einzelnen d$E$/d$x$-Messungen aus allen drei Tracking-Detektoren, für eine feste Reinheit (Anteil korrekt klassifizierter Teilchen) von 95 %. Farben unterscheiden zwischen Elektronen, Myonen, Pionen, Kaonen und Protonen.

Beim Vorgängerexperiment Belle wurde der spezifische Energieverlust lediglich in der Driftkammer bestimmt und die einzelnen d$E$/d$x$-Messwerte mittels des getrimmten Mittelwertes zusammengefasst. Im Vergleich zu Klassifikationen welche die zusätzliche Information der Einzelmessungen innerhalb der Likelihood-Funktion berücksichtigen, verschlechtert der getrimmte Mittelwert die Selektion deutlich. Auch die Verwendung von Messungen des spezifischen Energieverlustes in den Siliziumdetektoren verbessert die Selektion erheblich gegenüber Klassifikationen, die lediglich Driftkammer-Messungen nutzen. Hauptsächlich erklärt sich dies durch den in Silizium unterschiedlichen Energieverlust von Elektronen, die sich dadurch auch in Impulsbereichen identifizieren lassen, in denen ihr Energieverlust in der Driftkammer mit dem anderer Teilchen überschneidet. Dadurch ergeben sich auch für andere Teilchen mit Impulsen unter 1 GeV höhere Selektionseffizienzen.

# Contents

# 1. Introduction

The visible universe around us is almost completely dominated by matter, and only a very small amount of antimatter can be found in cosmic radiation. Since antimatter annihilates in contact with matter, antimatter accumulations in space would also be quite visible through radiation at the boundaries, and can thus be excluded. But what is the origin of this asymmetry?

One of the necessary ingredients for the observed matter–antimatter asymmetry is the violation of the so-called $CP$ symmetry, or $CP$ violation [1]. This symmetry was observed to be broken in the decay of neutral kaons in 1964 [2], which means that the interactions between matter and antimatter are not identical. However, the effect found in the kaon system is relatively small and not nearly large enough to explain the observed absence of antimatter. Still, measuring $CP$ violating processes may help us further our understanding of the observed asymmetry.

One other such process is the decay of neutral B mesons, which are bound states of a light u or d quark and a heavy b quark, and can be produced in large quantities at *B factories*, asymmetric electron–positron colliders operated at the $\Upsilon(4S)$ resonance, which, at 10.58 GeV, is only a little heavier than twice the $B^0$ or $B^\pm$ mass. Because of this, the $\Upsilon(4S)$, consisting of a $b\bar{b}$ quark pair, decays mostly into $B^0\bar{B}^0$ and $B^+B^-$ pairs. Due to the asymmetric beam energies, the created B mesons are not at rest with respect to the detector, but are boosted in forward direction. Through reconstruction of decay vertices (*vertexing*), this allows accurate measurements of the decay time of each B meson, which is vital for many measurements of $CP$ violation.

In this way, B factories, like PEP-II (BaBar experiment) in Stanford, USA, or KEKB (Belle experiment) in Tsukuba, Japan, can produce and analyse large numbers of B decays. Among the experiments' achievements are the discovery of $CP$ violation in the B system [3][4], verifying the theory for which Makoto Kobayashi and Toshihide Maskawa received part of the 2008 Nobel Prize in Physics [5], as well as the precise measurement of the parameters of the Cabibbo–Kobayashi–Maskawa (CKM) quark mixing matrix [6]. Additionally, since the b quark is the heaviest quark that can be found in bound states, higher order effects like loop corrections can be more readily observed, which makes B physics very interesting to probe for deviations from the standard model of particle physics. For example, a BaBar paper recently submitted for publication shows a $3.4\sigma$ deviation for $B \to D^{(*)}\tau\nu_\tau$ decays [7].

The KEKB collider holds the current world record in instantaneous luminosity [9], and was able to outperform PEP-II in the integrated luminosity recorded (see Figure 1.1). Both experiments have since been shut down, BaBar (at PEP-II) in 2008, Belle (at KEKB) in 2010. As a replacement, new B factory experiments are being planned, like the Belle II experiment in Tsukuba or possibly SuperB near Rome. Additionally, a different approach is to analyse B decays at hadron colliders, the prime example of which is the LHCb experiment at the Large Hadron Collider (LHC) in Geneva, with the first B decays recorded in 2010.
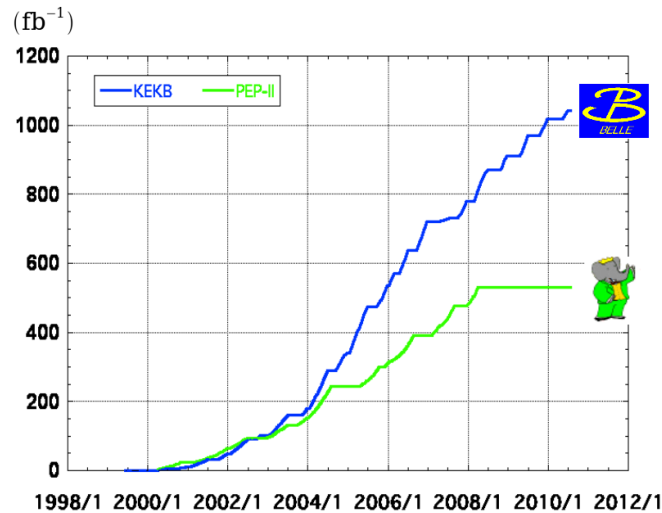
Figure 1.1.: Total integrated luminosity over time at the B factories KEKB (Belle) and PEP-II (BaBar). At the $\Upsilon(4S)$ resonance, $711\,\mathrm{fb}^{-1}$ have been recorded by Belle, $433\,\mathrm{fb}^{-1}$ by BaBar. Adapted from [8].

The successor of the Belle experiment, Belle II, is currently in the final stages of planning with some components already being built, and is slated to begin recording data in 2016. For Belle II, the KEKB collider will be upgraded to about 40 times its former luminosity and should be able to reach an integrated luminosity twice as high as KEKB's within a year of data taking (see Figure 1.2). This alone would allow many measurements with unprecedented statistics and precision. Most parts of the Belle detector will be replaced, and new detectors are added to, e. g., provide a better vertex resolution or improve the particle identification (see Chapter 2).

A good particle identification system is important for physics analyses, and needs to classify charged tracks into one of five particle types (pions, kaons, protons, electrons, and muons), with a low rate of misidentifications. Since, in addition to the upgrade of the detector components, the software framework was also rewritten [11], this is an opportunity to improve upon Belle's particle identification software. Chapter 6 will describe a likelihood ratio-based method to identify charged particles using their energy deposition in the tracking detectors. Different variations, such as using full likelihoods instead of truncated mean values or including silicon detector information, will be evaluated.

As a side effect of the much higher luminosity, the Belle II detector will also have to cope with a much higher background, e. g. from particles not from the $\Upsilon(4S)$ resonance. The pixel detector (PXD), which is closest to the collision point, will be quite strongly affected by this, with most of the data recorded being from non-$B\overline{B}$ events. If current estimations of the amount of background are correct, this data will need to be significantly reduced before it can be stored. This can be achieved by extrapolating particles from the outer detectors (where less background is expected) into the PXD, and discarding pixel data that does not correspond to one of the extrapolated tracks. This approach, however, is in danger of losing a significant amount of low-energy particles that might not propagate far enough into other detectors to allow a meaningful extrapolation. This would likely have a negative effect on physics analyses,
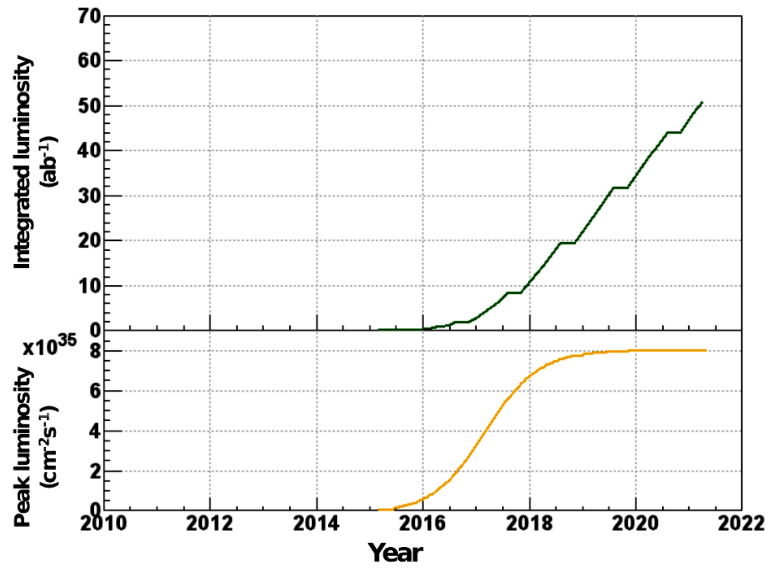
Figure 1.2.: Projected integrated (top) and instantaneous (bottom) luminosity for the Super-KEKB collider. One ab$^{-1}$ equals 1000 fb$^{-1}$. Adapted from [10].

e. g. in the above-mentioned analysis of B $\rightarrow$ D$^*\tau\nu_\tau$ decays, low-energy pions are produced by the decay of the D$^*$ that need to be found for an accurate reconstruction. In Chapter 5, a method will be presented that uses a neural network to distinguish pixel data created by low-energy particles (especially pions) from those created by background processes using only information from the pixel clusters themselves.

# 2. The Belle II Detector

In this chapter, an overview of the components of the Belle II detector will be given, with a special emphasis on both the tracking detectors, and the components used for particle identification. Most of the information contained therein is from the Belle II Technical Design Report [12]. The design of the detector components is not yet finalised, but the working principles are unlikely to change.

## 2.1. Overview

For the Belle II experiment, the KEKB $e^+e^-$ collider will be upgraded to achieve a luminosity of $8 \cdot 10^{35}\,\mathrm{cm}^{-2}s^{-1}$. Since this will be a fortyfold increase compared to what was achieved for the predecessor experiment, Belle, the upgraded collider will be appropriately called SuperKEKB. This gain is to be implemented mostly through a smaller beam size, and, less so, through a doubling of the beam current. To minimise the beam cross-section at the interaction point (IP), focussing magnets have to be positioned fairly close to the IP. At SuperKEKB, this is achieved using a relatively large collision angle of 83 mrad, which separates the beam pipes more quickly. The rest of the detector is built around the interaction point, with a design that takes into account the asymmetric beam energies (4.0 GeV on 7.0 GeV[1]). A cross-section of the resulting geometry, as currently implemented in the Belle AnalysiS Framework 2 (`basf2`), is shown in Figure 2.1.

To be able to reconstruct charged tracks, the three tracking detectors (PXD, SVD, and CDC) are placed in the center of the detector, in a relatively homogeneous 1.5 T magnetic field. Through measurements of the curvature of tracks, the central drift chamber (CDC) provides accurate information about the momentum and charge of most tracks. The pixel and silicon vertex detector (PXD, SVD) enclosed by the CDC are used to reconstruct decay vertices and find low-momentum tracks. Further out, the TOP and ARICH detectors can identify charged particles in the barrel and endcap part of the detector. They are surrounded by the electromagnetic calorimeter (ECL), which is, among other things, used for the detection of photons and identification of electrons. [2] The superconducting coil providing the magnetic field is situated in the outermost part of the Belle II detector, and is enclosed by the KLM detector, which can identify muons and $K_L^0$, and also serves as a flux return.

In the following sections, these components will be described in more detail.

---

[1] The convention $c = 1$ is used, so both energies and momenta will be given in electron volt.
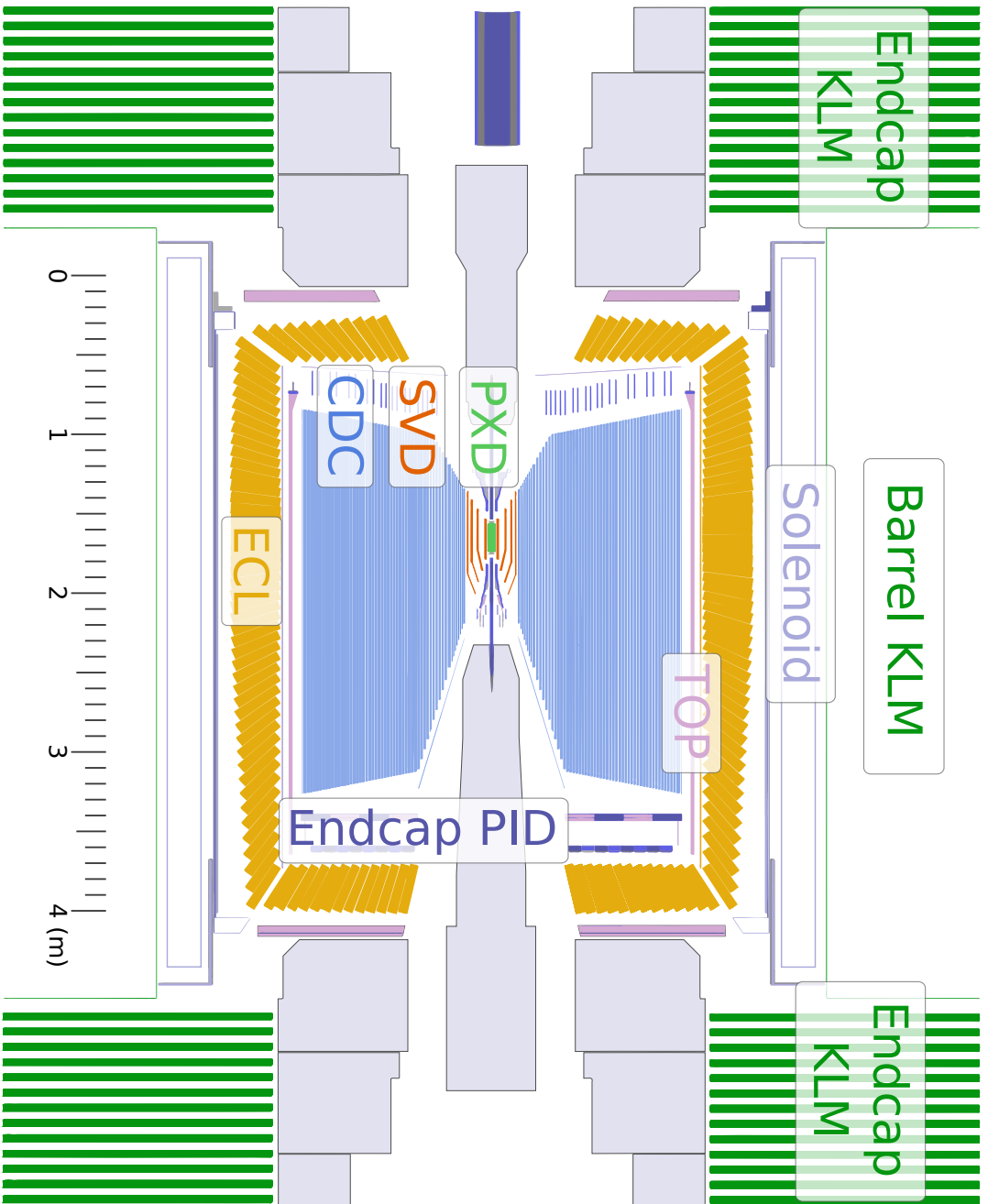[2] The charge conjugate is always implied.

Figure 2.1.: Cross-section of the Belle II detector. This drawing reflects the state of the geometry implemented in the Belle II software in February 2012.

## 2.2. Tracking detectors

### 2.2.1. PiXel Detector (PXD)

The pixel detector is a two-layered silicon sensor positioned directly outside the beam pipe, built using DEPleted Field Effect Transistors (DEPFET). Its primary purpose is to provide high-resolution data for the reconstruction of track vertices, which is vital for the determination of B decay lengths.

To this end, two layers of pixel sensors ($r = 14\,$mm and $22\,$mm) are placed directly outside the beam pipe, occupying a total volume that is comparable to that of a small soft drink can. This maximises the possible vertex resolution while reducing the required material budget. Figure 2.2 shows the arrangement of the two layers and the position of the active detector material.
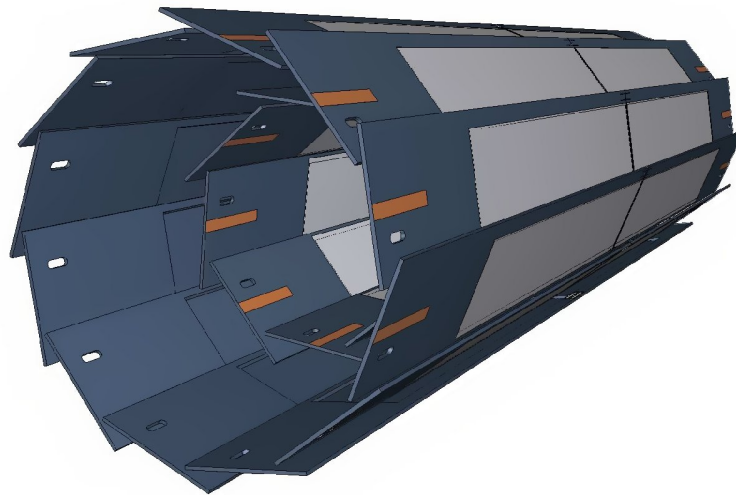
Figure 2.2.: Schematic view of the pixel detector's geometry, with areas containing DEPFET pixels shown in gray. Adapted from [12, p. 77].

The individual sensors consist of a large number of DEPFET pixels that are thinned down to about $75\,\mu$m [13], which significantly reduces the effect of multiple scattering. The individual pixels consist of a $p$-channel MOSFET or JFET with an *internal gate* being implanted through doping, about $1\,\mu$m beneath the transistor channel (see Figure 2.3). When a high voltage is applied to the back side of the substrate, electron–hole pairs created by incident particles in the bulk will be separated. Whereas the holes will drift toward the back contact, electrons will collect in the potential minimum of the internal gate. Since the collected charge will modulate the current in the $p$-channel between source and drain, the charge can be read out—non-destructively—by measuring this current. The internal gate can afterwards be cleared by applying a positive voltage to a nearby contact ($n+$ clear), which will remove the collected electrons.

The pixels are arranged in a dense matrix, with a row-wise readout in which a gate voltage is applied to all pixels in one row and the drain currents of all pixel columns are measured. Afterwards, the accumulated charges are removed by applying a clear voltage. However,
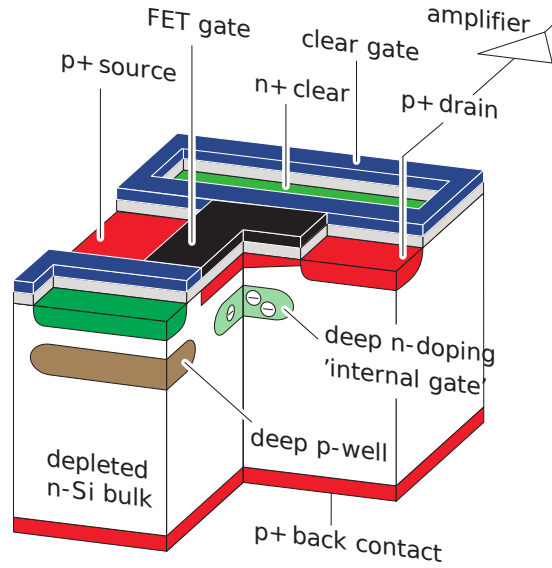
Figure 2.3.: Illustration of the design of a DEPFET pixel. Taken from [12, p. 79].

even with no charge accumulated in the internal gate of the pixel, some *pedestal charge* is measured. These are caused by leak currents through the silicon, which tend to differ between pixels and will be affected by the sustained radiation dosage. One possibility of removing them is double-sampling, where the drain current is measured again after clearing any pixel charges. Because of the additional time required to do so, single-sampling will be used instead, where the average current in a pixel row (as measured through the normal readout) is subtracted from the pixel current measurements. A pedestal correction for individual pixels will then be performed off-line. [13]

Even with some parallelisation, the readout of the PXD's total eight million pixels, arranged in 1600 pixel rows, still takes about 20 $\mu$s. As a result, a relatively large number of pixels in each readout frame will be active, a problem that will be addressed in Chapter 5.

### 2.2.2. Silicon Vertex Detector (SVD)

The silicon vertex detector consists of four layers of double-sided silicon ladders with long *p*-doted strips on one side, and perpendicular *n*-doted strips on the other. The SVD is sandwiched between the pixel detector and the drift chamber, corresponding to inner and outer radii of 38 mm and 140 mm. The geometric arrangement of the silicon vertex detector is reproduced in Figure 2.4.

The SVD covers the full 17°–150° acceptance of the Belle II detector. In the forward region, the three outer layers are slanted toward smaller radii, resulting in wedge-shaped sensors that reduce the material budget and improve the resolution for tracks with high $p_z$.

Particles traversing the *p–n* junction create electron–hole pairs that are then separated by an applied field and collected on the contacted strips. In the (rare) case of only one particle hitting a ladder, its position can be reconstructed without ambiguity from the active strips on both sides. However, with increasing number of active strips, so-called *ghost hits* are created, as illustrated in Figure 2.5. This ambiguity can be resolved by using information from the
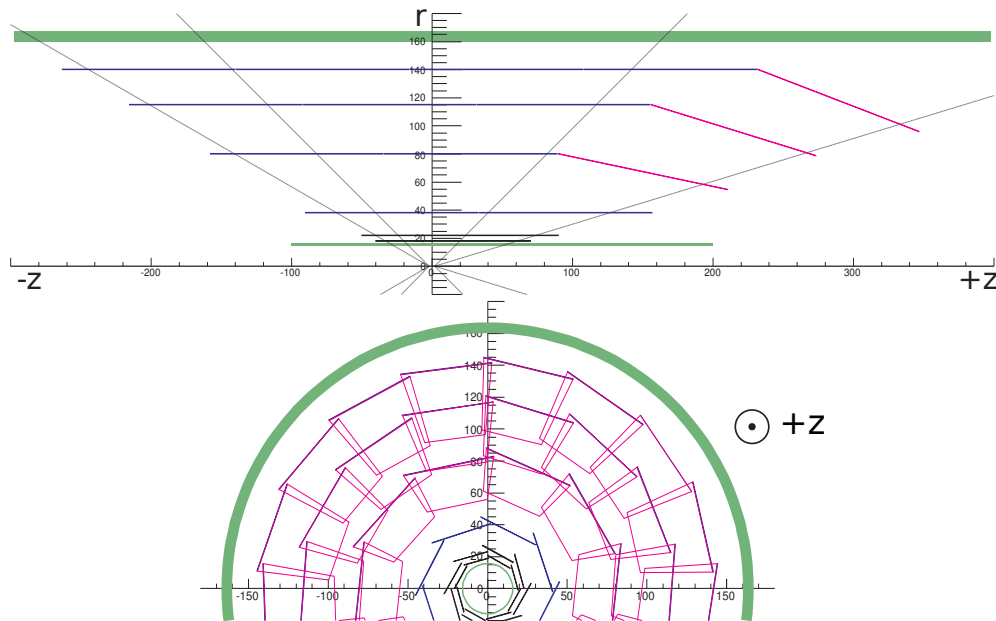
Figure 2.4.: Configuration of the four SVD layers, with the two PXD layers shown directly outside the beam pipe as seen from the side (top) and from the forward direction (bottom). All dimensions in mm. Adapted from [12, p. 142]

other SVD layers, since ghost hits are unlikely to line up into tracks.

Compared to the PXD, the silicon vertex detector is further away from the interaction point and will have a faster readout, making it less prone to background and helping with the extrapolation of tracks from the drift chamber back into the pixel detector.

### 2.2.3. Central Drift Chamber (CDC)

Surrounding the SVD—and with an outer radius of 113 cm the largest tracking detector—is the central drift chamber, consisting of 14,336 sense wires suspended in a 50% helium, 50% ethane gas volume. Charged particles passing through the gas lose some of their energy due to collisions, producing electron–ion pairs. These are separated by an electrical field applied through 42,240 field wires, with the electrons being increased in number by gas amplification. [14] This works by accelerating the primary electrons in the electric field around the sense wire (anode), providing them with the energy to ionise further gas atoms. However, as this energy needs to be gained in between collisions with gas atoms, where the energy would be dissipated, this avalanche effect only occurs in the close vicinity of the anode. The electrons are collected by a sense wire, yielding a signal proportional to the number of electrons originally created through ionisation by the track, and thus to the track's energy loss in the wire cell. The details of the processes through which particles can lose energy in both the drift chamber and the silicon detectors are discussed in Section 3.1.

Further away from the sense wire, before gas amplification can take place, the acceleration in the electric field is compensated by the collisions with the surrounding gas atoms, resulting in a nearly constant drift speed. Using the time stamp of the signal and a reference $t_0$, this
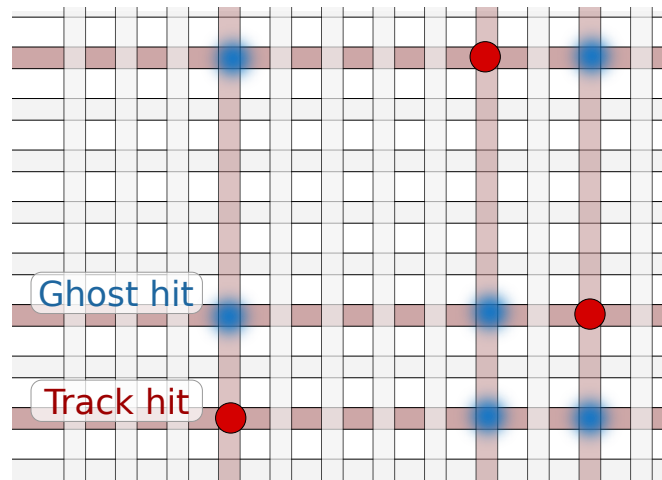
Figure 2.5.: Creation of ghost hits: three particles create hits in an SVD sensor, with the hits and activated strips shown in red. All intersections of active strips are hit candidates, resulting in ghost hits (blue).

can be utilised to determine the *drift time* taken by the electrons on their way to the sense wire. As the electrons quickly approach a constant final drift velocity, this is proportional to the (minimal) distance between a particle's path and the sense wire. This information can be used by the tracking algorithms to easily achieve a resolution orders of magnitude smaller than than the distance between wires. This is illustrated in Figure 2.6, which visualises the drift lengths around CDC hits along a track.
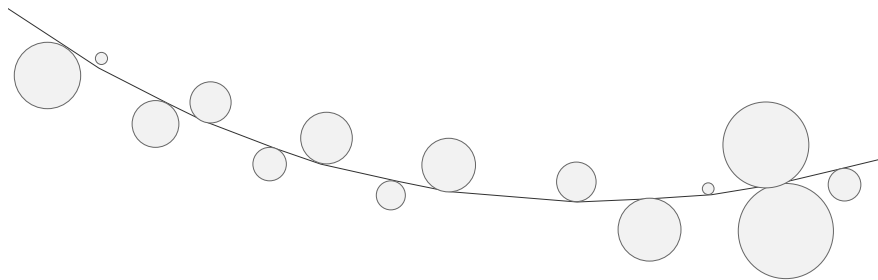


Figure 2.6.: Illustration of a track in the drift chamber with drift lengths shown as dark circles around each wire hit.

Passing through the CDC, a track will typically produce at least 56 wire hits. If all wires were arranged in parallel, this would enable us to reconstruct the track only in the $x$–$y$ plane, $z$ information would have to come from other detectors. This can be avoided by adding layers with a non-zero angle between sense wires and the $z$ axis, so-called *stereo layers*. Layers with wires in parallel to the $z$ axis are called *axial layers*. For the Belle II drift chamber a configuration of alternating superlayers is used, where one superlayer consists of six layers with the same orientation (eight for the innermost superlayer). An axial superlayer

is thus followed by a stereo superlayer, followed by another axial superlayer. The next stereo superlayer then has a negative stereo angle to optimise the *z* resolution. The arrangement of the total 9 superlayers can be represented as AUAVAUAVA, with axial (A) and stereo superlayers with different angles (U and V). A cross-section of the drift chamber can be seen in Figure 2.7, which illustrates the arrangement of the superlayers. To reduce the impact of background, expected to be strongest in the inner parts of the detector, the innermost superlayer is realised with a more dense packing of wires. This so-called small-cell chamber consists of 8 layers instead of the usual 6, and should have a lower occupancy than a superlayer with normal cell configuration.
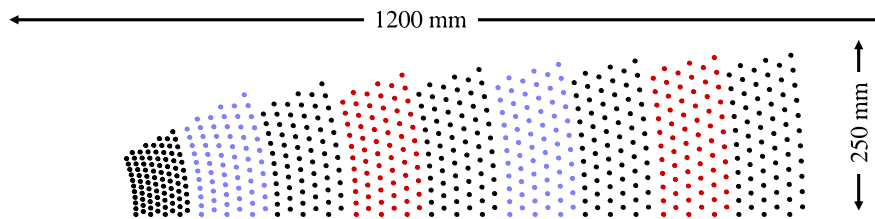


Figure 2.7.: Cross-section of the CDC wire configuration, with axial superlayers in black, and stereo superlayers in violet and red for positive and negative stereo angles, respectively. Adapted from [12, p. 204].

Overall, the central drift chamber provides a very cost-efficient instrumentation of a large volume, and provides—in conjunction with the tracking algorithms—excellent spatial resolution. It also has a fairly low dead time and was also designed for the purposes of particle identification through specific energy loss (see Chapter 6).

## 2.3. Barrel PID: Time-Of-Propagation (TOP) counter

Situated immediately outside the central drift chamber, the time-of-propagation counter aims to identify particles using a combination of time-of-flight and Cherenkov angle measurements. This is achieved by having particles exceed the speed of light in the medium of a quartz radiator, and measuring both the time of arrival and the angle of the produced Cherenkov photons. Using a reference time $t_0$, the time measurement can be used to obtain the velocity of the particle; similarly, particles with different velocities $\beta c$ will produce photons with different angles $\theta_c$ (see Section 3.3). By utilising the internal reflection of the photons in the quartz radiator and combining these two measurements into a single detector, the TOP counter is very compact, with the quartz bar being only 2 cm thick (see Figure 2.8).

The photons are detected using an array of photomultipliers (PMTs), that resolve both $x/y$ and $t$ dimensions. For the time measurements, the resolution was greatly improved, with a PMT $t$ resolution of below 40 ps and an expected resolution for the reference time $t_0$ of around 25 ps. For comparison, the performance of time-of-flight detectors has stagnated at 100 ps since the early 1990s.
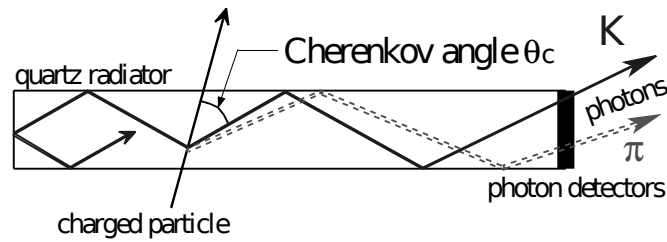
Figure 2.8.: Illustration of the internal reflection of Cherenkov photons in the TOP detector, with pions and kaons of the same momentum producing photons with different Cherenkov angles. Adapted from [12, p. 220].

## 2.4. Endcap PID: Aerogel Ring-Imaging Cherenkov (ARICH) detector

For the endcap particle identification, a pure ring-imaging Cherenkov detector with aerogel radiators is to be installed. The detector consists of two separate layers for the aerogel, where Cherenkov photons are produced, and the photomultipliers, where they are detected, as can also be seen in Figure 2.1. This allows the photons to form Cherenkov rings large enough to be clearly separated by the PMTs, similar to the effect of the internal reflection in the TOP counter's quartz bars.

The ARICH is designed for a good pion/kaon separation over most of the momentum spectrum, as well as the separation of the lighter $\pi, \mu$ and e for momenta below 1 GeV.

## 2.5. Electromagnetic Calorimeter (ECL)

Belle II's electromagnetic calorimeter encloses both TOP and ARICH, and consists of 8736 scintillating CsI(Tl) (caesium iodide doped with thallium) crystals, weighing a total of 43 tons. It has been designed mainly for the detection of photons and measurement of their position and energy, as well as the identification of electrons, and—in conjunction with the KLM—for the detection of $K_L^0$.

Both photons and electrons produce electromagnetic cascades that are quite similar. Photons produce secondary electrons either through pair production for high energetic tracks, or through the photoelectric or Compton effect for photons with lower energies, knocking away electrons from the atoms of the scintillator crystals. Primary or secondary electrons in turn interact with the material, freeing electrons, radiating photons; or, in the case of positrons, annihilating and again producing photons. The result is an electromagnetic cascade that gradually deposits a large portion, or in many cases the entire energy of the incidental photon or electron.

These cascades are accompanied by excitations of the scintillator material, which are dissipated through low-energy (few eV) photons. These are then amplified and detected through photomultipliers mounted at the end of each crystal. Since the number of photons is directly dependent on the energy deposition (with some fluctuations), the PMT signal provides a measurement of the absorbed particle's energy.

Electrons are identified by assigning ECL clusters to charged tracks in the tracking detectors,

clusters without a charged track are correspondingly presumed to have been created by a photon. Since the only indication that an event contains photons is through ECL clusters, a good position resolution can be important for physics analyses. This is achieved through the high number of separate crystals, an approach that has already proven itself in the Belle experiment. As a result, the two photons produced in the decay of $\pi^0$ can be identified separately, allowing positive identification.

Because of the good time resolution of the PMTs, the ECL readout is also used as a trigger signal for the other detectors.

## 2.6. $K_L^0$ **and muon (KLM) detector**

Beyond the calorimeter and the superconducting coil responsible for the detector's magnetic field, the K-Long and Muon Detector is made up of a sandwich structure of 4.7 cm thick iron plates with interleaved resistive plate chambers.

Resistive plate chambers (RPCs) consist of two planar glass sheets that act as high voltage electrodes, separated by a thin gas volume. Particles traversing this volume create ion–electron pairs that are accelerated by the electric field, producing a streamer between the two electrodes. This streamer is conductive, and causes a voltage drop in the electrodes nearby, that, because of the high resistance of glass, is not immediately evened out. This voltage drop is then detected by pick-up strips placed on either side of the chamber. The pick-up strips are a few centimeters wide and are arranged orthogonally on both sides, so the particle track can be localised in $z/\phi$ for the barrel, and $\phi/\theta$ for the endcap in addition to the layer information.

To be able to deal with the background radiation, especially neutrons, that would otherwise reduce the detection efficiency of the RPCs, the inner two layers of the barrel KLM will instead be instrumented with scintillators. [15]

To identify muons, the KLM utilises their high penetration power to distinguish them from hadrons. For hadrons, the KLM and ECL combined provide 4.7 interaction lengths worth of material, which means their energy will be quickly dissipated through hadronic showers. Electrons, on the other hand, fall victim to the short radiation length in iron (1.7 cm [16]), and are usually absorbed by the electromagnetic calorimeter.

Neutral $K_L^0$ mesons can also be identified, as they are able to produce clusters in both ECL and KLM. To this end, ECL/KLM clusters are grouped and associated with charged tracks found in the tracking detectors. Clusters without an accompanying charged track are then taken as $K_L^0$ candidates. Because of the large fluctuations inherent in hadron shower development, the clusters cannot be used to measure the $K_L^0$ energy, this would only be possible with a full hadron calorimeter.

# 3. Interactions of Charged Particles in Matter

This chapter will provide a brief overview of the interactions of charged particles with a surrounding medium, with a special focus on the interactions most relevant to the operation of the Belle II detector. Most significant for the tracking detectors is ionisation (and the related excitation) of atoms, which also describes the creation of electron–hole pairs in the silicon detectors. Since this energy loss can be measured through the charge signal provided by both drift chamber and silicon detectors, it will play a decisive role in Chapters 5 and 6. Bremsstrahlung is crucial for the understanding of the interaction of electrons in the detector, especially in the electromagnetic calorimeter (ECL), and finally, Cherenkov radiation is the underlying process used for Belle II's dedicated particle identification detectors, the TOP and ARICH.

## 3.1. Ionisation and excitation

Low-energy charged particles traversing a detector interact mainly with the atomic electrons of the detector material, and can either excite them into more energetic states, or ionise them, creating electron–ion pairs. Excited atoms can also create further ionisation by transferring their energy to other atoms.

These processes depend strongly on the incident particle's momentum, which determines the maximum kinetic energy that can be transferred to the electrons of the medium. For particles other than electrons ($m_0 > m_e$) and energies low enough to satisfy $2\gamma m_e/m_0 \ll 1$, this maximum energy can be approximated as [14]

$$E_{\text{kin}}^{\text{max}} = 2m_e c^2 \beta^2 \gamma^2.$$

This expression only depends on $\beta\gamma$, not on the particle's mass. For the next-lightest particle, the muon, the latter condition can be written as $\gamma \ll 424$, which should hold for muons up to a few GeV and all heavier particles with the same momentum.

The actual dependence of the energy loss on $\beta\gamma$ is described by the Bethe–Bloch equation

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

which is valid for values of $\beta\gamma$ between about 0.1 and 1,000. [16] Here, $z$ is the charge of the incident particle (in multiples of the elementary charge), $Z$ the atomic number of the absorbing material, $A$ the atomic mass in $\text{g mol}^{-1}$ and $K$ is a shorthand for $4\pi N_A r_e^2 m_e c^2$. $T_{\text{max}}$ is the maximum possible energy transfer in a single collision (in MeV) and $I$ the mean excitation energy in eV. For muons, the Bethe–Bloch shape (with some additional effects for lower and higher energies) is shown in Figure 3.1.

A peak in ionisation can be seen for very low $\beta\gamma$, which quickly falls away for higher particle energies. In practise, particles with a momentum coinciding with the ionisation peak either
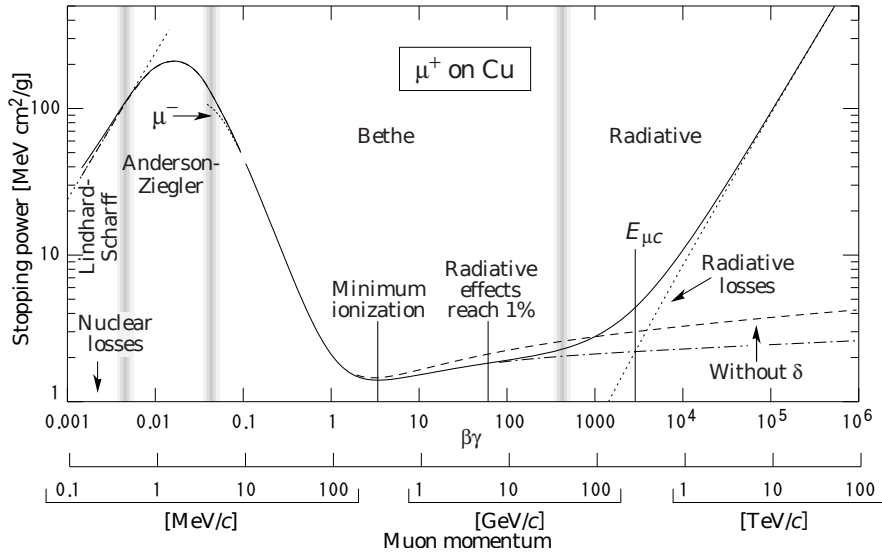
Figure 3.1.: $dE/dx$ (stopping power) for muons in copper over $\beta\gamma = p/(m_\mu c)$. The solid curve shows the total energy loss, with different components highlighted as dashed lines. The arrow labelled $\mu^-$ points to an effect that is specific to negatively charged particles. Taken from [16].

fail to reach the tracking detectors due to insufficient momentum or are absorbed by the beam pipe, so only the downward slope will be seen. Around $\beta\gamma \approx 4$, a minimum at very low ionisation is reached; particles at this minimum are thus called *minimum-ionising particles* (MIPs). Table 3.1 shows the energy loss through ionisation in the active materials of the tracking detectors, which also provides a scale for the total energy loss expected in the silicon detectors and the drift chamber. For higher energies, the energy loss increases again, leading to the logarithmic *relativistic rise*.

Table 3.1.: Energy loss of minimum-ionising particles in the materials used in the tracking detectors. Values in the last column are provided as energy per length, and thus include material densities. Data from [16].

| Material | $\frac{dE}{dx}\Big|_{min} / \frac{\text{MeV}\cdot\text{cm}^2}{\text{g}}$ | $\frac{dE}{dx}\Big|_{min} / \frac{\text{MeV}}{\text{cm}}$ |
|---|---|---|
| Silicon (Si) | 1.664 | 3.876 |
| Helium (He) | 1.937 | $3.222\cdot10^{-4}$ |
| Ethane ($C_2H_6$) | 2.304 | $2.910\cdot10^{-3}$ |

The function $\delta(\beta\gamma)$ describes the so-called *density effect* that suppresses the relativistic rise. For relativistic incident particles, their transverse electric field is to some extent screened by the surrounding atomic electrons, an effect that depends mostly on the density of the detector material. [14] As a result, the probability of interactions is significantly reduced for particles with higher momentum, creating the *Fermi plateau* also visible in Figure 3.1.

For solids, this effect is especially pronounced and almost entirely hides the relativistic

rise. For example for 1 GeV electrons in silicon detectors, the energy loss through ionisation is about the same as for pions or muons with the same momentum. In gases, the energy loss is much higher, creating a quite visible band for electrons (see also Section 6.3).

It should again be emphasised that this formula for the energy loss through ionisation only depends on the particle's velocity through $\beta\gamma$. Consequently, if one plots the energy loss for different particle types over their momentum $p = \beta\gamma mc$, the same curve can be used to describe those particles, only scaled by the particle mass $m$. The difference between these curves forms the basis of the $dE/dx$ particle identification method described in Chapter 6.

Strictly speaking, the assumption of heavy particles inherent in the Bethe–Bloch formula is no longer fulfilled for electrons and ionisation through collision of incident electrons with the atomic electrons needs to be handled separately. The overall deviation from the Bethe–Bloch formula is on the order of 10–20 %. [14] The ionisation for positrons in turn differs from the behaviour of both that for heavy particles and electrons; additionally, positrons will annihilate with electrons in the detector medium, releasing additional energy. This effect, however, only results in a slightly higher energy loss and is only visible at low particle energies. [16]

### 3.1.1. Energy loss distribution

The Bethe–Bloch formula just introduced describes the momentum-dependence of the mean energy loss, with the actual measurements for $dE/dx$ distributed around this value. This distribution characterises the energy transferred to the electrons of the medium (gas, silicon) and is dominated by very large losses that create so-called $\delta$ *electrons* (also $\delta$ rays or knock-on electrons). [14] These energetic secondary electrons can then create further ionisation. $\delta$ electrons are a common sight in bubble chamber pictures, as Figure 3.2 demonstrates.
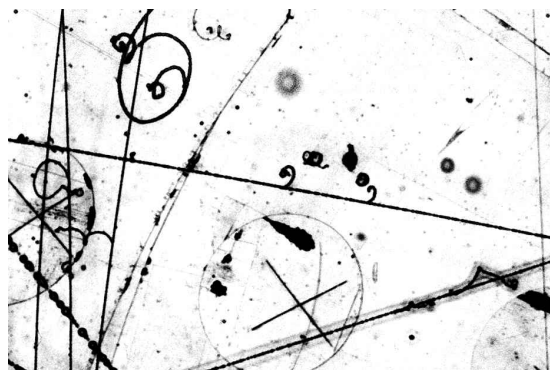


Figure 3.2.: Tracks in a bubble chamber, with curling $\delta$ electrons visible along the horizontal track. Adapted from [17].

In 1944 Lev Landau published a theoretical description of these processes and the resulting energy loss distribution. [18] The so-called Landau distribution is strongly asymmetric and exhibits a long tail corresponding to the aforementioned high energy transfers (see Figure 3.3). As a result of these tails, the integrals defining the moments of the distribution (mean, variance, etc.) diverge, and the associated quantities are thus undefined. [19] The Landau distribution is commonly parametrised using the most probable value (MPV) and a scale factor describing the width in $x$.
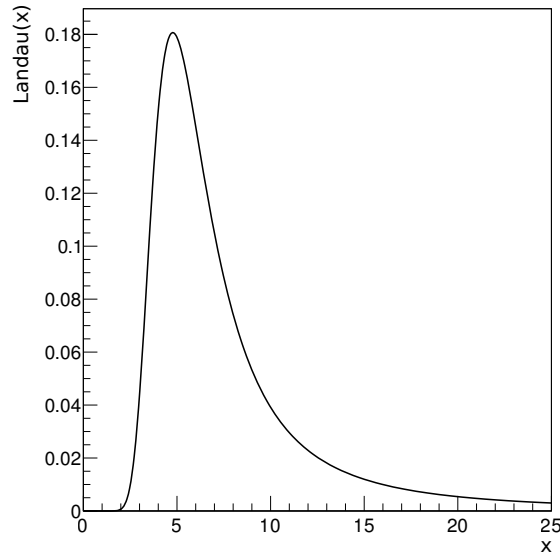
Figure 3.3.: Plot of a normalised Landau distribution with MPV=5 and scale=1.

It should be noted that, while the central limit theorem cannot be applied to the Landau distribution since neither mean nor variance have finite values [19], for very thick absorbers the energy loss distribution will have less pronounced outliers [16], which merely indicates that there is an upper bound on the energy transfer. For the amount of material traversed in the tracking detectors, i. e. thin silicon sheets and a few meters of fairly thin gas, this effect will not be visible.

The derivation by Landau allows infinite momentum transfers and thus cannot be an accurate description of the actual distribution. Improved parametrisations have been suggested by, e. g., Vavilov [20] and Bichsel [21].

## 3.2. Bremsstrahlung

For higher energies, radiative losses become more important, where the incident particles interact with the Coulomb field of the **nucleus** and lose some of their energy through emission of photons. The energy loss through bremsstrahlung is directly proportional to the energy and inversely proportional to the square of the mass of the incident particle. [14] Because of this, the effect of bremsstrahlung for heavy particles (i. e. other than electrons) can usually be neglected. For the next-heaviest particle, the muon, with a mass of about $106\,\text{MeV} \approx 212 m_e$, bremsstrahlung is attenuated by a factor of $(m_\mu/m_e)^2 \approx 45{,}000$.

It is thus useful to consider the special case of electrons, where the energy loss through bremsstrahlung can be written as:

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{E}{X_0}$$

with the particle energy $E$ and the material-specific radiation length for electrons $X_0$. [14] Solving this simple differential equation yields

$$E = E_0 e^{-x/X_0}.$$

Thus, after traversing one radiation length worth of material, the electron's energy will be reduced to $1/e$ of the original value. Additionally, the same $X_0$ is also equal to 7/9 of the mean free path for pair production of photons. [16] Consequently, both the photon and electron part of electromagnetic showers can be described by a single quantity, which, together with the linear dependence on the particle energy is also the underlying principle of the energy-loss sampling used in calorimeters.

Since energy loss through bremsstrahlung only becomes significant for high energies, the energy at which the loss through ionisation and bremsstrahlung are equal is commonly called *critical energy*. For muons in copper, $E_{\mu C} \approx 300\,\text{GeV}$ is also labelled in Figure 3.1. For electrons, the critical energy is much lower, for example for the cesium iodide (CsI) crystals used in the Belle II calorimeter, it is about 11.2 MeV for e$^-$ and 10.8 MeV for e$^+$. [16]

## 3.3. Cherenkov radiation

If a charged particle exceeds the phase velocity of light in a medium, Cherenkov photons are emitted. For slower particles, the dipoles both before and after the incident particle align themselves toward it, generating a spherically symmetric field whose higher multipole moments vanish. This is no longer the case once the particle exceeds the local speed of light: only the dipoles behind the incident particle can be aligned, which results in a no longer symmetric field shape that follows the particle. As a result, a cone-shaped electromagnetic wave front extends behind the particle. [14]
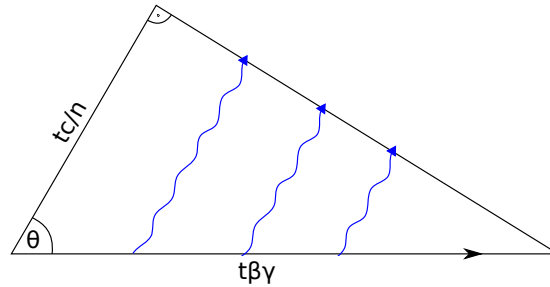


Figure 3.4.: Illustration of emitted Cherenkov photons along the track of an incident particle (black arrow), with the Cherenkov angle $\theta$.

Here $\theta$ is called the Cherenkov angle that can be calculated as follows: In the same time needed by the electromagnetic wave to travel a distance $t v_{\text{ph}}$, the incident particle covers a distance $t\beta c$. Since the wave fronts created along the track must interfere constructively to actually result in Cherenkov photons, the track, wave front and direction of emitted photons must create the right triangle shown in Figure 3.4. Thus, using $v_{\text{p}} = c/n$, the Cherenkov angle can be calculated as:

$$\cos\theta = \frac{tc/n}{t\beta c} = \frac{1}{n\beta}$$

Since the Cherenkov angle of the emitted photons in a certain material depends only on the incident particle's velocity, Cherenkov radiation is well suited to be used for particle identification (in combination with momentum measurements), as is done for Belle II with the TOP and ARICH detectors.

# 4. Statistical Classification

Since both of the following chapters are concerned with the classification of data into different categories, the underlying principles will be explained here.

Statistical classification is very closely related to hypothesis testing, where one wants to quantify how well a given data set matches a hypothesis $H_0$, as compared to an alternative hypothesis $H_1$. In general, this works by mapping a many-dimensional data set $\vec{x}$ onto a scalar test statistic $t(\vec{x})$ in which the different hypotheses should have a sufficiently disparate distribution. Cuts $t > t_c$ on the test statistic then should separate between the two hypotheses, though in most cases, misclassifications will occur. These can be discerned into false positives, where a data point has been selected as belonging to a hypothesis $H_0$ but actually belongs to a different hypothesis, and false negatives, where $H_0$ has been falsely discarded by the cut. This is illustrated by Figure 4.1, which shows normalised probability densities for two hypotheses and their distribution in $t$. The two types of errors are shown for an arbitrary cut $t_c$, which separates the curves into regions of correct and false assignments. The fraction of events selected by the cut $t > t_c$ where the hypothesis $H_0$ was falsely discarded is called the *significance* $\alpha$, the fraction of data points belonging to the other hypothesis and mistakenly selected is named $\beta$, with $1 - \beta$ being called *power*.
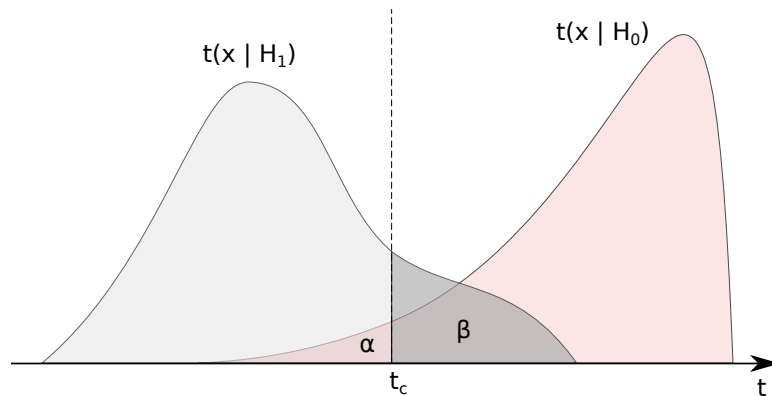


Figure 4.1.: Distributions of a test statistic $t$ for two hypotheses $H_0$ and $H_1$. For a given cut $t > t_c$ for the selection of data points belonging to $H_0$ the significance $\alpha$ and power $1 - \beta$ are shown.

For practical purposes, it may be more useful to define quantities that describe the data set selected with $t > t_c$. One can then define the *efficiency* as

$$\epsilon = 1 - \alpha = \int_{t_c}^{\infty} t(x|H_0)\mathrm{d}x,$$

and the *purity*

$$p = \frac{\int_{t_c}^{\infty} P_0 \, t(x|H_0) \mathrm{d}x}{\int_{t_c}^{\infty} (P_0 \, t(x|H_0) + (1 - P_0) \, t(x|H_1)) \, \mathrm{d}x},$$

where $P_0$ is the prior probability for $H_0$ [19]. These quantities for the efficiency and purity of the selected data sample can be used to easily compare the effect of different cuts $t_c$, but can also serve as a quantification of how well a certain mapping of the data space to the test statistic can separate the hypotheses $H_0$ and $H_1$ if considered for the entire range of possible cuts.

For more than two hypotheses, a one-dimensional test statistic is usually not sufficient and a higher-dimensional one will have to be used instead.

## 4.1. Likelihood ratio

For two hypotheses $H_0$ and $H_1$ a simple test statistic can be constructed using the conditional probability $\mathscr{L}(\vec{x}|H_k)$ of the data $\vec{x}$ under the assumption that $H_k$ is correct. This conditional probability, or *likelihood*, can be compared with that for another hypothesis by considering their ratio

$$t = \frac{\mathscr{L}(\vec{x}|H_0)}{\mathscr{L}(\vec{x}|H_1)}.$$

The difference between the likelihood and a simple probability density function is that while the latter is an analytical quantity, the former depends on the composition of the data sample and is thus a random variable.

If the individual data points $x_i$ are uncorrelated, the likelihood can be written as

$$\mathscr{L}(\vec{x}|H_k) = \prod_i p_i(x_i|H_k).$$

This becomes most useful when $\vec{x}$ consists of independent measurements of the same physical quantity $x$: Then $\mathscr{L}$ becomes a product of repeated evaluations of the same probability density function (PDF) $p(x|H_k)$, which makes the computation of this statistic remarkably simple and illustrates its applicability to uncorrelated data.

The Neyman–Pearson lemma states that—provided the hypotheses do not contain free parameters—this *likelihood ratio* test statistic provides the most powerful test (smallest $\beta$) at a given significance level $\alpha$, which is equivalent to the highest purity at a given efficiency [19].

While the test itself is simple, constructing the PDFs $p_i(x_i|H_k)$ can be more involved and can be done using simulations (Monte Carlo) or data, if the correct hypothesis can be determined through some other process. This should generally be done with an amount of data high enough to avoid large statistical uncertainties on the created PDFs. When correlations between some of the $x_i$ have to be taken into account, these PDFs will have to be multi-dimensional (e. g. two-dimensional for repeated measurements of two correlated variables). The required amount of data rises exponentially with the number of correlated dimensions $n$. Because of this, classifications using the likelihood ratio method are usually unsuitable for problems involving many correlated variables.

## 4.2. Neural networks

### 4.2.1. Overview

Neural networks, on the other hand, deal especially well with correlations. In the biological world, a neural network like the brain consists of neurons, which usually possess multiple inputs (called dendrites), a cell body and a long protrusion, the axon, that eventually branches out and connects to other neurons. The dendrites receive stimuli from other neurons that may elicit the neutron's activation (firing). When fired, an electrical pulse travels along the axon to the nerve endings and, depending on the type of the connections, can either facilitate or inhibit the firing of the connected neurons.

Artificial neural networks emulate this structure, with nodes and the connections between them as their basic constituents. A node can have any number of inputs $i_k$ that are transformed into an output $o$ by adding the inputs and feeding them into a so-called *activation function g*:

$$o = g\left(\sum_k i_k\right).$$

Often the S-shaped sigmoid is used for the activation function, but other options, e.g. a step function, exist. One can also add *bias nodes* with no input and an output set to a fixed value. Connections link inputs and outputs of nodes together, and can carry an arbitrary real-valued weight that the output is multiplied with before being handed to the next node.

In principle, these networks could become arbitrarily complicated, and might, for example, include cycles. In practise, a network arranged in layers, with connections only from layer $n$ to layer $n+1$ is more useful and greatly reduces the computational costs. This type of network is appropriately called *feed-forward network*.
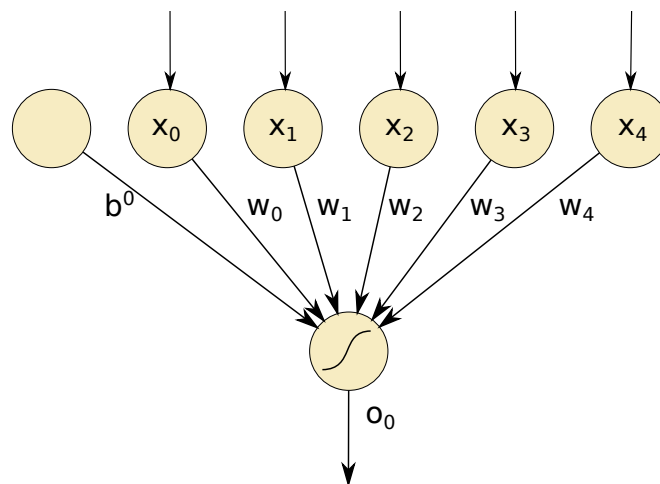


Figure 4.2.: A single-layer perceptron with a single output node.

The simplest type of feed-forward network only has a single layer of nodes, which directly provide the output, plus the input variables $x_i$ and a bias node $b^0$. For a single output node, this *single-layer perceptron* is illustrated in Figure 4.2. The output value of this simple network

is given by

$$o_0 = g\left(b^0 + \sum_k w_k x_k\right),$$

which, for a monotonous activation function $g$, corresponds to a linear test statistic, i. e. the most complicated separation boundary that could be described by the network is a hyperplane (Fisher's discriminant) [19][22].

When another layer is added, the network becomes much more expressive, and it can be shown that given enough nodes in the *hidden layer*, any continuous function can be parametrised [23]. Such a two-layer perceptron is shown in Figure 4.3 and produces an output of the form

$$o_0 = g\left(b^1 + \sum_j w_j^1 g\left(b_j^0 + \sum_k w_{kj}^0 x_k\right)\right),$$

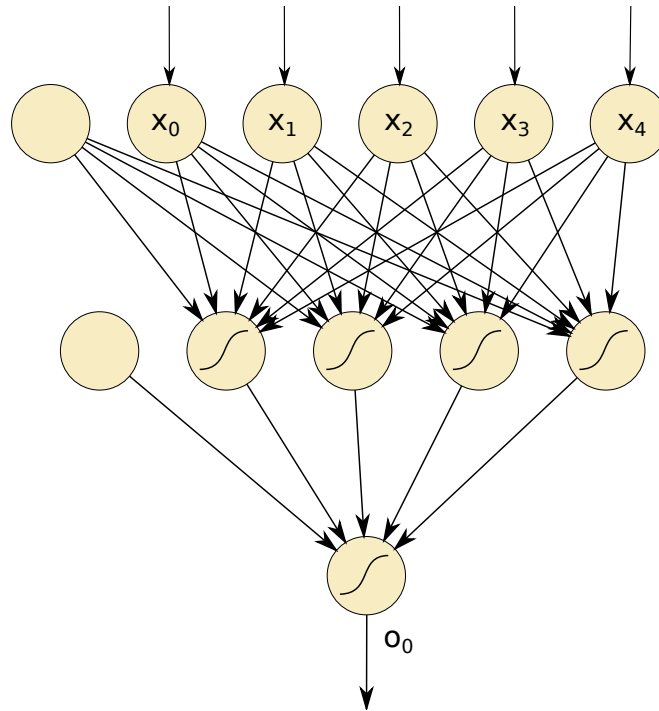where $k$ iterates over the input nodes and $j$ over the hidden nodes.



Figure 4.3.: A two-layer perceptron with four hidden nodes and a single output node.

Since the nodes themselves are static, the entire information of the network is contained in the connections between them and in their weights. Finding an appropriate set of weights that produces the desired mapping of input to output is called *training* and involves presenting the input data to the network and adjusting the weights to reduce the deviation of the output from the true value. One common training method for these layered networks is called *error back-propagation*, which uses a gradient descent to minimise the output errors [22]. Simply stopping the training when the error on the data sample does not decrease further can however lead to undesired results. Especially for relatively small samples, the network

might then simply have learned the statistical fluctuations inherent in the data sample, an effect called *overtraining*. This can be avoided by reducing the degrees of freedom of the network (i. e. lowering the number of nodes) and/or dividing the data into a training set presented to the network for back-propagation, and a validation set that is only used to track the total error.

For the background reduction discussed in Chapter 5, a software package for neural networks called NeuroBayes [24] will be used. In addition to the implementation of the neural network itself, NeuroBayes uses sophisticated methods to improve the network's performance and robustness. Some of these methods, in so far as understanding them is relevant to the study at hand, will be described in the following paragraphs.

### 4.2.2. Preprocessing

Before the input variables are fed into the network, they can be preprocessed to transform them into the most suitable inputs for a neural network. As a first step, NeuroBayes flattens real-valued variables to potentially avoid giving bins with low statistics undue weight. This results in a histogram as shown in Figure 4.4, with an approximately equal number of entries in each bin.

Afterwards, the purity $N_S/(N_S + N_B)$ in each bin is considered and, if applicable, spline-fitted, which again reduces the effect of statistical fluctuations (see Figure 4.5). This is transformed into the final network input, as shown in Figure 4.6. The resulting distribution is centered around 0 with a standard deviation of 1, with bins with a better signal–background ratio being assigned higher values.

Finally, linear correlations between variables are determined, and the variables are linearly combined into a new set of variables that is uncorrelated. This is not strictly necessary for the training, as the network could learn the correlations, but simplifies the minimisation of errors. In fact, non-linear correlations will still have to be learned by the network itself.

### 4.2.3. Pruning

During training, the input variables are ranked by their importance by individually removing them from the network. Considering the sample statistics, variables are then permanently removed from the network when their added significance is below a user-defined cut (e. g. $3\sigma$). Similarly, if individual weights of network connections are statistically indistinguishable from zero, they can be removed. Both methods are useful precautions against overtraining by decreasing the network's size, and consequently its degrees of freedom.

### 4.2.4. Network output

With a sigmoid activation function, as used by NeuroBayes, the final output will have values in the interval $[-1, 1]$ and should, for a fully trained network, be linear in the per-bin purity [24]. That is, the signal fraction for a network output $o$ should be given by

$$p = \frac{o+1}{2}.$$

If the a priori distribution of the input data is the same as for the training data set, one can thus directly interpret the output as a probability.
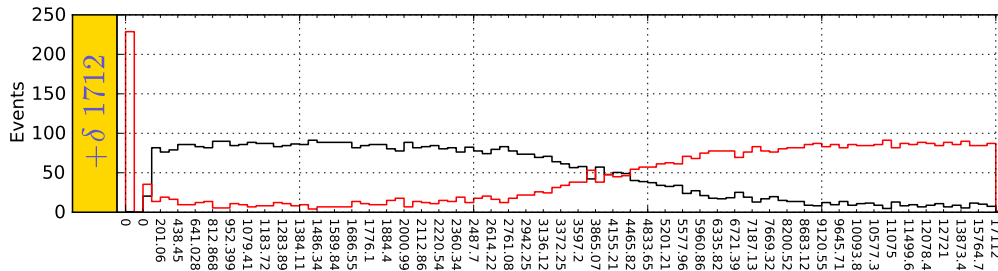
Figure 4.4.: Flattened histogram created by NeuroBayes' preprocessing for an input variable. Different colours show signal and background. For the 1712 events in the yellow box this variable is not defined, and they are treated separately (e. g. not included in the histogram flattening).
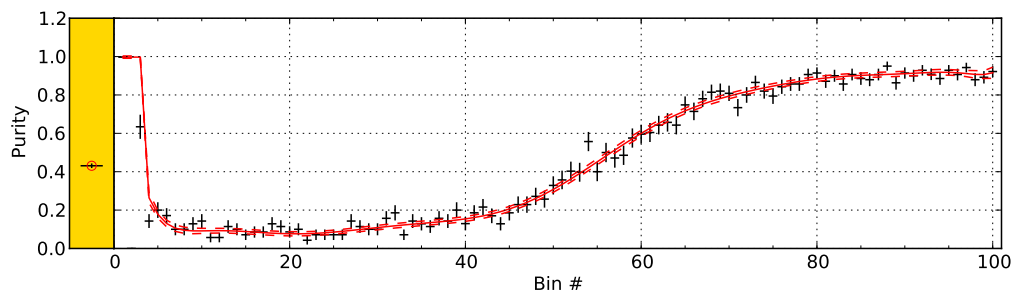


Figure 4.5.: Spline fit to the per-bin purity of the histogram shown in Figure 4.4.
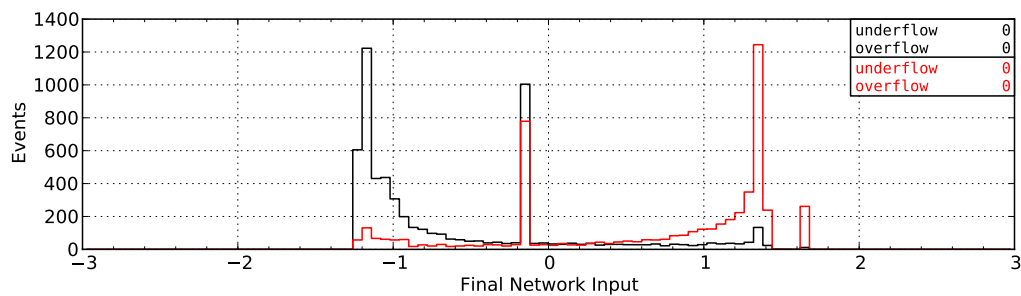


Figure 4.6.: Distribution of the final network input for the same variable, with a mean of 0 and a standard deviation of 1.

# 5. Pixel Detector Background Reduction

Because of the higher luminosity and decreased beam cross section at SuperKEKB compared to KEKB, the detectors of the Belle II experiment will have to cope with a much higher background, possibly 20 times as much as for Belle. While all detectors will be affected by this, the pixel detector with its relatively large integration time of 20 μs will suffer from an especially large event size. Assuming a worst case occupancy of 3 %—corresponding to 240,000 of the PXD's total eight million pixels being active—and a size requirement of 4 bytes for each active pixel, one arrives at an event size of 1 MB. This can be compared to only about 100 kB for all other detectors combined.

For a bandwidth of 1.2 GB/s to storage, the total event size must not exceed 200 kB for an effective trigger rate of 6 kHz. This means that there must be some kind of data reduction that can select pixel data from physics events and discard the rest. To avoid discarding PXD data, a background reduction of 90 % would be necessary for the worst case. For Belle II, this will be done by finding tracks in the silicon vertex detector and central drift chamber and extrapolating them into the PXD. This will be described in 5.2.

In the scope of this diploma thesis, a different approach based on the analysis of individual PXD clusters will be examined that addresses some of the shortcomings of the tracking-based data reduction.

## 5.1. Backgrounds at Belle II

In the following paragraphs, a brief overview of the different types of backgrounds that are expected at Belle II will be given.

**Touschek effect:** The Touschek effect is a form of intra-beam scattering, i. e. electrons or positrons in one bunch interact with each other, resulting in particles with slightly lower and higher momentum than the bunch average. It was first observed at the $e^+e^-$ storage ring AdA [25] and was found to have a strong dependency on the beam energy. The effect is proportional to $E^{-3}$ (with the beam energy $E$), so it is especially large for low energy beams. [26] For this reason, SuperKEKB will use slightly less asymmetric beam energies than KEKB, with 7.0 and 4.0 GeV instead of 8.0 and 3.5 GeV for the high and low-energy rings, respectively.

The scattered electrons and positrons eventually escape the bunch and will hit the beam pipe. Because of beam focussing and a smaller beam pipe cross-section, this happens more often in the vicinity of the detector, where secondary particles may produce background hits in the tracking detectors. These generally have a strong asymmetry: Touschek particles will have a small incidence angle (originating from somewhere along the beam pipe) and mostly go in backward direction, as the majority originate in the low energy ring. This also extends to the secondary particles produced

in the beam pipe. The high energy ring will contribute only a very minor fraction of total background events.

**Two–photon QED processes:** These are luminosity dependent QED processes in which a positron and an electron interact via two photons and producing a new $e^+e^-$ pair, as shown in Figure 5.1.
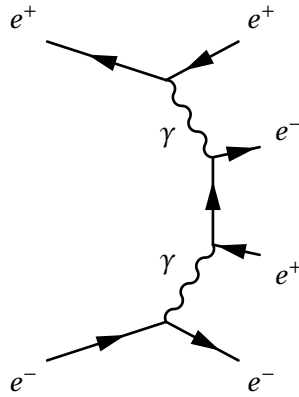


Figure 5.1.: Feynman graph for a two–photon QED process.

The result is a background that originates from around the interaction point, with very low-energy particles being produced.

Current simulations indicate that this will be the largest background for the pixel detector at Belle II. [26] Additionally, the produced particles will look similar to physics events, which makes them harder to pick out.

**Beam–gas Coulomb scattering:** While an ultra-high vacuum of less than $10^{-7}$ Pa is maintained around the interaction region, and a pressure $< 10^{-6}$ in the rest of the beam pipe [12], a few gas molecules will always remain. The beam particles can scatter with the remaining gas, producing high energy particles from both the LER and HER. This background only includes the first order, elastic scattering (Coulomb). Effects of inelastic scattering are expected to be smaller and are consequently neglected.

There is currently no accurate simulation of this background available, but it is expected to be similar in structure to the Touschek background.

**Radiative Bhabha scattering:** Bhabha scattering is an extremely common process at $e^+e^-$ colliders, where the input particles either annihilate into a $\gamma$, which then creates a new electron–positron pair, or simply exchange momentum via one intermediary $\gamma$. In some cases, an additional photon can be generated, whose daughter particles may end up hitting the detector. Feynman graphs for Bhabha scattering with the radiated photons highlighted are shown in Figure 5.2.

This process is luminosity dependent (normal Bhabha scattering is in fact often used to monitor luminosity), and is also dependent on the geometry of the magnetic field. It especially affects the outer detectors, where according to current studies the time of propagation (TOP) counter would suffer from a background dose much higher
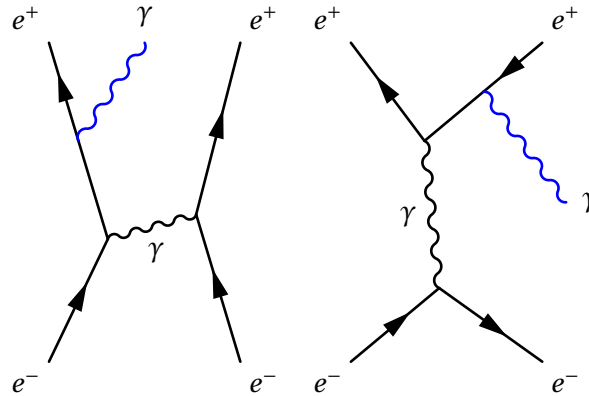
Figure 5.2.: Example Feynman graphs for radiative Bhabha scattering.

than design specifications. The pixel detector however is relatively free from particles originating from radiative Bhabha scattering, and most particles that do reach the PXD are the result of backscattering in the outer detectors.

Recent studies of the aforementioned backgrounds for the PXD show the QED processes to be the largest background, with 0.64 % occupancy in the first and 0.23 % in the second layer, followed by the LER Touschek background with 0.12 % and 0.09 % in the first and second layers, respectively. The other backgrounds do not contribute significantly to the occupancy, with both beam–gas and radiative Bhabha scattering creating only a handful of hits per readout frame. Consequently, only the Touschek and QED backgrounds will be used for the following study, since the other processes do not contribute significantly to the total data volume. The total occupancy in both PXD layers according to the simulation is thus less than one per cent, which would necessitate a background reduction of only 70 %.

The simulation of the QED processes can be assumed to be fairly accurate, as these processes are confined to the interaction point and do not depend on the detector geometry or the shape of the magnetic fields in the beam pipe. The inverse holds for the other background processes, where the $B$ field geometry and shielding can have a large impact. Both are likely to change in the future, in an effort to reduce the background rate suffered by the detectors, which is necessary in particular for the TOP detector. One example for the effect of geometry changes is the addition of vertical collimators, which reduced the estimated rate for LER Touschek—at the time expected to be the largest background—from 0.89 to 0.16 GHz. [27]

Figure 5.3 shows the pixels activated during one readout frame on a typical sensor, illustrating the different asymmetries of the main background processes: Touschek particles tend to produce longer clusters parallel to the $z$-axis, whereas particles from QED background result in a somewhat larger variety of clusters. This amount of background per sensor has to be compared with very few clusters from actual physics events (not shown here).

## 5.2. Background reduction via tracking

To reduce the PXD event size, the majority of which is caused by these backgrounds, one can use data from the silicon vertex detector and the central drift chamber. Both detectors add
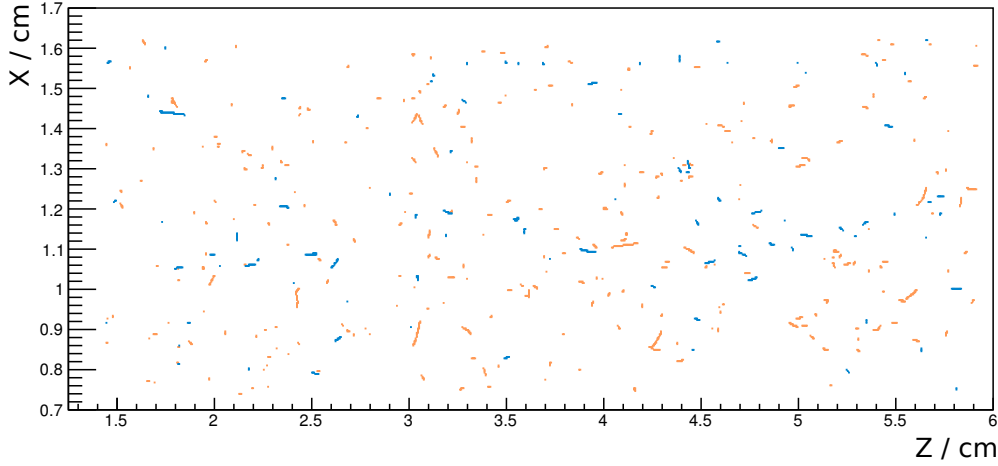
Figure 5.3.: Simulated background hits during one readout frame $(20\,\mu s)$ for a single PXD sensor, separated into low energy ring Touschek and 2–photon QED background.

possibly much cleaner data—through a faster readout and their increased distance from the interaction point—which can be used by the High Level Trigger (HLT) to reconstruct tracks and extrapolate them into the pixel detector. The intersections of the tracks with the PXD sensors then define *regions of interest* (ROI): pixels inside the ROI are assumed to be part of a physics event, whereas all other pixels are assumed to originate from background and discarded. It is also possible to use only part of the data, e. g. from the silicon vertex detector, which simplifies the track finding. While the HLT track extrapolation with the full data will be the dominant method, drift chamber data is disregarded in the following paragraphs. However, as the tracks considered rarely produce CDC hits, the arguments are fairly general and apply to both methods.

As the chance of random background hits in 4 layers lining up to create a track that looks like it came from a physics event is rather small, this approach should lead to a very high background reduction. The efficiency, i. e. the ratio of found tracks to the total number of tracks, depends strongly on the tracking algorithm, but should be relatively high. However, particles with low momentum may not reach all SVD layers and have a higher probability for multiple scattering, which diminishes tracking efficiency.

Figure 5.4 shows the momentum distribution of pions from simulated $B\overline{B}$ events that are stopped before the fourth SVD layer: plotting the transverse momentum $p_t$ of particles over their relative $z$ momentum $\cos(\theta) = p_z/p$, one notices that a $p_t$ of 60 MeV may already be insufficient to reach all SVD layers. Naively, one would expect the transverse momentum $p_t$ (right plot) to be largely independent of the angle $\theta$, since the sensors—with the exception of the three outer SVD layers in forward direction—are parallel to the $z$ axis. The large dependence on $\cos\theta$ can be explained by considering the traversed material for different angles: for particles with $\theta = \frac{\pi}{2}$ ($p_z = 0$), the particle only crosses a fairly short section of beam pipe, whereas the amount of traversed material increases for smaller or larger angles. While the beam pipe consist of a fairly thin gold-coated beryllium $(Z = 4)$ pipe, at the low particle energies involved it can still stop particles with some efficiency. Thus, since particles with $\theta \approx \frac{\pi}{2}$ have a lower total momentum $p = p_t/\sin\theta$ for the same $p_t$, they are far more likely to

be absorbed by the beam pipe.

This results in the shape seen for $p_t(\cos\theta)$, with particles in forward or backward direction requiring less $p_t$ to reach the detector layers, and getting close to the 20 MeV required to reach the 3rd SVD layer. The plot also shows a slight asymmetry resulting from the slanted SVD parts in the forward direction, as well as particles of even high momenta being lost at the outer limits of the acceptance in $\theta$.
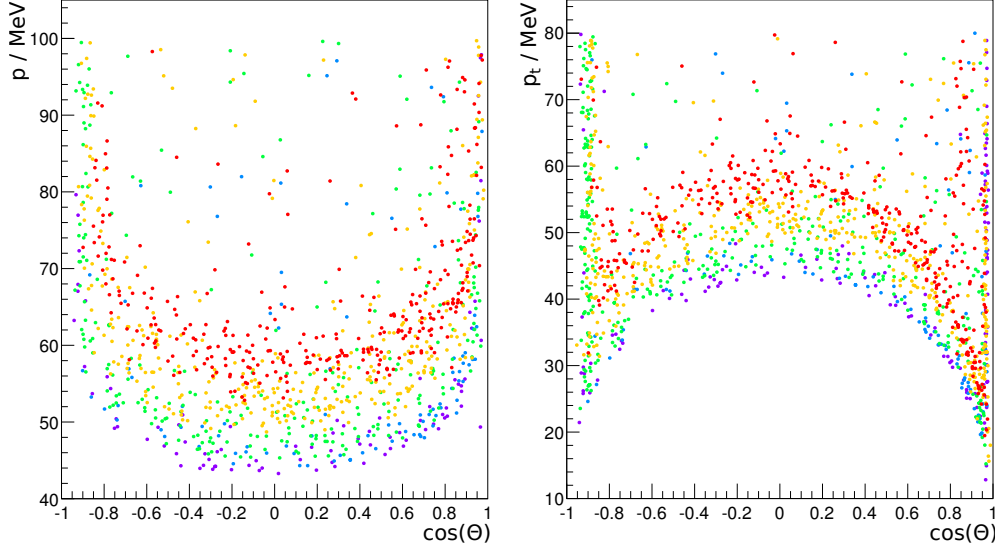


Figure 5.4.: Total and transverse momentum of simulated pions that do **not** reach the last SVD layer, over $\cos\theta = p_z/p$. Colours indicate the outermost layer reached: violet and blue for PXD layers 1 and 2; green, yellow and red for SVD layers 1, 2, and 3, respectively.

One important process $(\mathrm{BR}(\mathrm{B} \to \mathrm{D}^*(2010)^\pm \mathrm{anything}) \approx 22.5\,\%$ [16]) is the decay of a B meson into a D$^*$, an orbitally excited D meson, which can then decay as $\mathrm{D}^*(2010)^\pm \to \pi^\pm{}_s \mathrm{D}^0 [\to \mathrm{K}^\mp \pi^\pm]$, where $s$ stands for slow. For example, for many analyses it is useful to do a full reconstruction of the decay products of one B meson (*tag side* B), which, as the B mesons are always produced in pairs, also fixes the four-momentum of the other B. This greatly simplifies the reconstruction of the so-called *signal side* B, which may also include neutrinos in its decay. Because of the large branching fraction for the production of $\mathrm{D}^*(2010)^\pm$, it is vital that the D$^*$ is correctly reconstructed. This entails finding and identifying both the D$^0$ decay products (a kaon and a pion), and the pion coming from the D$^*$ itself. This decay also can be used to calibrate or verify particle identification algorithms, as described in Section 6.5.

Because of the small mass difference between the $\mathrm{D}^*(2010)^\pm$ and the D$^0$ of only 146 MeV, the produced pion with a mass of 140 MeV will be of very low energy, earning it the name *slow pion*, or $\pi^\pm{}_s$. The resulting $p_t$ distribution for B$\overline{\mathrm{B}}$ events is shown in Figure 5.5. It is apparent that below 60 MeV $p_t$, the majority of pions can be attributed to this one decay. This, however, means that the data reduction scheme will lose some data from important physics events: for the slow pions from D$^*$ decays, about 15 % will not reach the last SVD layer. If the tracking algorithm then fails to find them, no PXD hits would be saved, and only hits in the first three SVD layers would be available for later reconstruction. Of course, for

many slow pions there is even less information available.

For the full reconstruction algorithm, this means that a fairly large portion of the $D^*$ decays might be incorrectly reconstructed. If the slow pion is not found at all, its charge information is also lost, which for example might result in a $B^0$ being reconstructed as a $B^+$. This would negatively affect physics analyses by increasing the amount of background events.
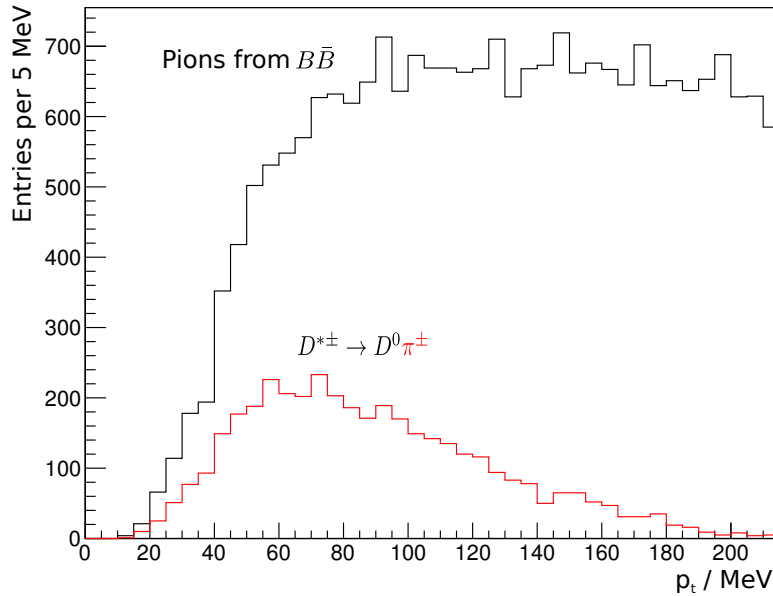


Figure 5.5.: Distribution of the transverse momentum of all pions in $B\overline{B}$ events and of only pions from $D^*$ decays (slow pions), in black and red, respectively. Particles that do not reach the tracking detectors are omitted.

## 5.3. Rescuing slow hadron hits via cluster analysis

A particle traversing the detector will in many cases hit more than one pixel. For this reason, adjacent active pixels will be grouped into clusters. This can be used to further improve the resolution, for example in the case of two active pixels, the particle's position is not only known to be between both pixels, but can be weighted according to the deposited charge.

In the Belle II analysis framework (`basf2`), the `PXDClustering` module finds adjacent pixels in each sensor and groups them into clusters, which are used in the following analysis. Because of different behaviours of the background and signal particles, one should be able to distinguish them to some degree by looking at individual clusters. Variables particularly suited for separating slow hadron clusters from background are given in the following, with distributions for signal and background for some of them shown in Figure 5.6.

**Charge:** The charge depositions in the pixel detector are rarely caused by the primary particles produced directly in the background process, but rather from secondaries that are created when the primary particles hit the beam pipe. For secondary electrons or positrons that have enough transverse momentum to reach the PXD, the average

momentum for particles from Touschek effect is about 57 MeV, and 390 MeV from two–photon QED processes. For electrons, these are fairly high energies, making them essentially minimum-ionising particles, since the relativistic rise for high momenta is small in silicon sensors (see Section 3.1). One thus expects a relatively small charge deposition from background, whereas pions or heavier hadrons with a momentum below 100 MeV—corresponding to a $\beta\gamma < 0.7$ even for pions—should have an energy loss many times as large. For a cluster, one can simply use the **total charge**, but the **minimal** and **maximal pixel charge**, as well as the **variance of pixel charges** (or its square root) in a cluster can be used.

The peaks visible in the total charge histogram in Figure 5.6 are caused by a cutoff at 25,000 electrons on the pixel charge, resulting in a large number of slow pions being sorted into the corresponding bin, or into 50,000 if two pixels reach the highest value. The effect of this cutoff on the network performance is discussed further in Section 5.5.

**Position:** Since background particles mostly originate from within both beam pipes, whilst signal events are created with a boost in forward direction, a noticeable asymmetry in the position of signal and background clusters is expected. In addition to the position in $\theta$ and $\phi$, the **layer** is also used, which is useful mostly for distinguishing background components from each other.

**Cluster length:** Similar reasoning applies to the length of clusters. As already illustrated in Figure 5.3, Touschek clusters tend to be elongated in $z$ direction. Particles coming from the interaction point, e. g. from physics events or 2–photon QED background, usually produce more point-like clusters. However, depending on the particle's momentum, the track may curl in the magnetic field, producing longer clusters perpendicular to the $z$-axis. One can thus define the **cluster length in $z$** and $r/\phi$, the **total length** and the **number of pixels per cluster**.

For background, simulations corresponding to one readout frame, with 23 % LER Touschek and 77 % QED background, were used. The QED background was generated using the `KoralWInput` module included in the Belle II software, Touschek background files were provided by the collaboration. Other available backgrounds do not contribute significantly to the detector's occupancy and have been neglected. For the signal component, B$\overline{\text{B}}$ events have been simulated using EvtGen with the standard decay file used by Belle, and clusters from low-energy pions selected: The pion producing the cluster must have a $p_t < 65$ MeV and less than half of a cluster candidate's charge must be assigned to secondary particles. In the forward direction ($\theta < 30°$, corresponding to $\cos\theta > 0.86$), the SVD sensors are slanted towards the beam pipe to increase acceptance. As this means that particles with lower transverse momentum can reach the outer SVD layers, the selection criterion is lowered to $p_t < 40$ MeV. The effect of this changed signal definition can be seen in the distribution of $\theta$, corresponding to the $z$ position, as shown in Figure 5.6. Below $30° \approx 0.52$ rad, the amount of signal suddenly drops to about half its former value, which is likely to influence measurements of the performance of the network. However, even without this cut, variations of the network performance for different angles are expected, since the distribution of signal and background differs along $\theta$ (and also $\phi$), for example a somewhat softer drop of similar magnitude can be seen in backward direction.
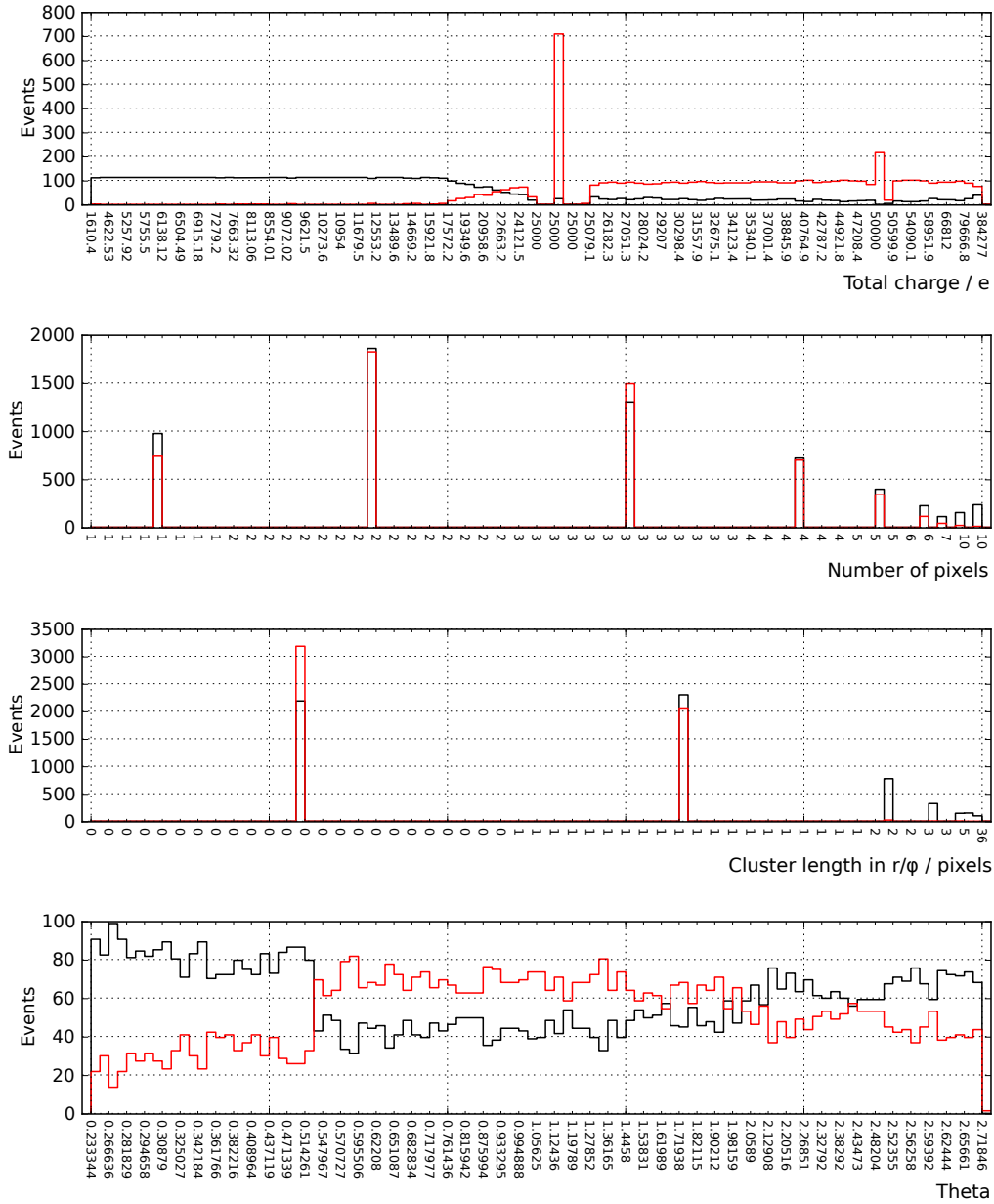
Figure 5.6.: Flattened distributions of a subset of cluster variables for slow hadrons (red) and background clusters (black).

This signal definition is then used together with the cluster variables defined above to train a neural network, as described in Section 4.2. The employed neural network software, NeuroBayes, will then determine correlations between variables and rank them by their importance for the network's performance.

Variables whose added significance is compatible within three sigma with fluctuations from random chance are discarded. This helps in keeping the network from learning the statistical fluctuations present in all data, which would reduce its performance when applied to a different data set. In this case, the layer, cluster length and charge variance variables are affected. For the charge variance and cluster length, this is probably caused by their high correlation to other variables, i. e. their information content can easily be restored by using a combination of other variables.

For binary classification, a single output node in the network combines all available information and will serve as the final discriminator between signal and background clusters. Figure 5.7 shows this network output for signal and background clusters. While a good separation between both categories is already apparent, the next section will attempt to quantify this by analysing the effect of different cuts on the output.
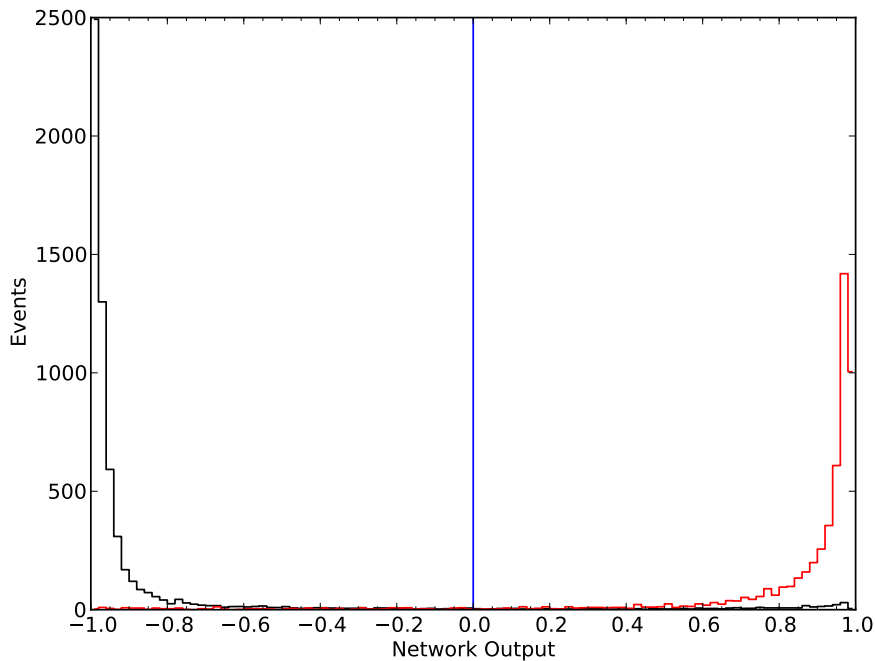


Figure 5.7.: Distribution of the network output for signal (red) and background (black) clusters.

## 5.4. Results

To analyse the performance of the network, the most important quantities for a specific network output cut are the background reduction and the signal efficiency, defined as:

$$\text{background reduction} = \frac{\text{discarded background clusters}}{\text{total background clusters}},$$

$$\text{signal efficiency} = \frac{\text{kept signal clusters}}{\text{total signal clusters}}.$$

Statistical errors on both signal efficiency and background reduction are calculated using the variance of the associated probability density function, as detailed in [28]. It should be noted that since only the numerator changes slightly for adjacent network cuts, these errors will be highly correlated for similar values of efficiency or background reduction.

As only a certain amount of data can be retained, the background reduction variable will be used to determine the final network cut, which will depend on the actual detector occupancy and the amount of bandwidth remaining after the tracking data reduction has selected interesting clusters. An ideal network would retain most signal clusters while discarding a high number of background clusters.

Figure 5.8 shows the background reduction over the signal efficiency for different cuts on the network yields. While the background reduction is independent of the selected signal clusters, the signal efficiency may vary for different types of signal. Here this is shown for different transverse momenta $p_t$ of the slow hadrons: For all $p_t$ values, a background reduction of over 90 % is possible with > 95 % signal efficiency, with the highest efficiency being reached for pions around 50 MeV $p_t$.

One can also look at the performance of the network for different values of $\theta$, as shown in Figure 5.9. As is expected from the different distributions of signal and background clusters in $\theta$, the performance is not uniform, and drops significantly in forward direction. For $\theta < 30°$ this is most likely caused by our signal definition, which causes the fraction of signal clusters to drop in this region.

## 5.5. Robustness tests

As the details of some components of the pixel detector electronics are not yet finalised, including what the maximal charge of a pixel can be before the data type overflows, it is still unclear what the precision of the data available to the network eventually will be. Similarly, the background simulation is still incomplete: Synchrotron radiation is not yet included, and not all background components have a physically accurate simulation. For this reason, some tests of the network's robustness will be described in this section.

As a general consideration, a changing background composition, with e. g. a larger amount of background from radiative Bhabha scattering, should not have a large effect on the network. Of all network variables, the pixel/cluster charges have the highest discrimination power, meaning that the energy loss of slow hadrons is—even with some fluctuations—much larger than that of practically minimally ionising electrons and positrons.

To verify that the network in fact does not rely overly on the highly background-dependent distribution of the position and length of clusters, the network was retrained with only the

Figure 5.8.: Background reduction over signal efficiency for different cuts on the network output, colours indicate different transverse momenta of pions. The help distinguish the curves, the second plots shows the same data with both axes fixed to values larger than 80 %.
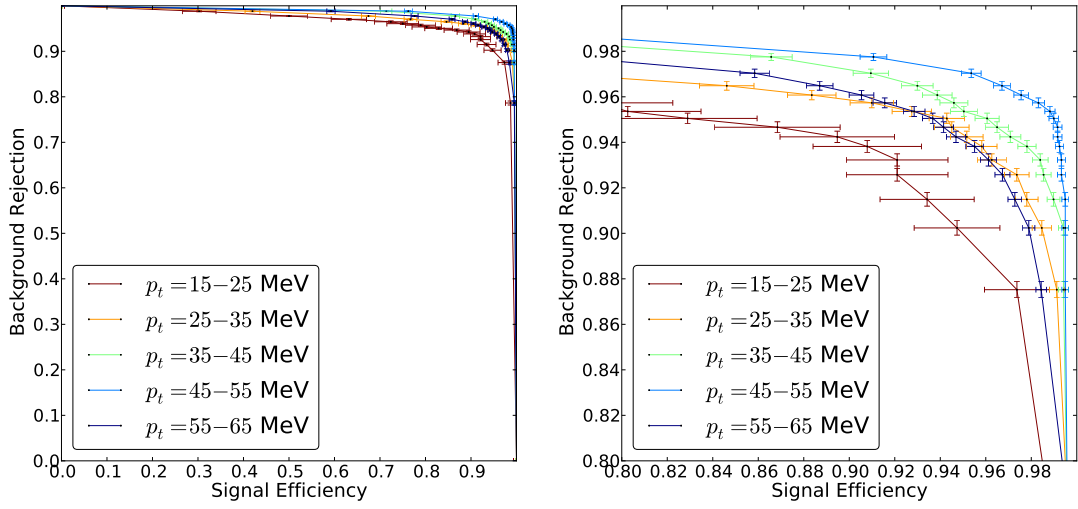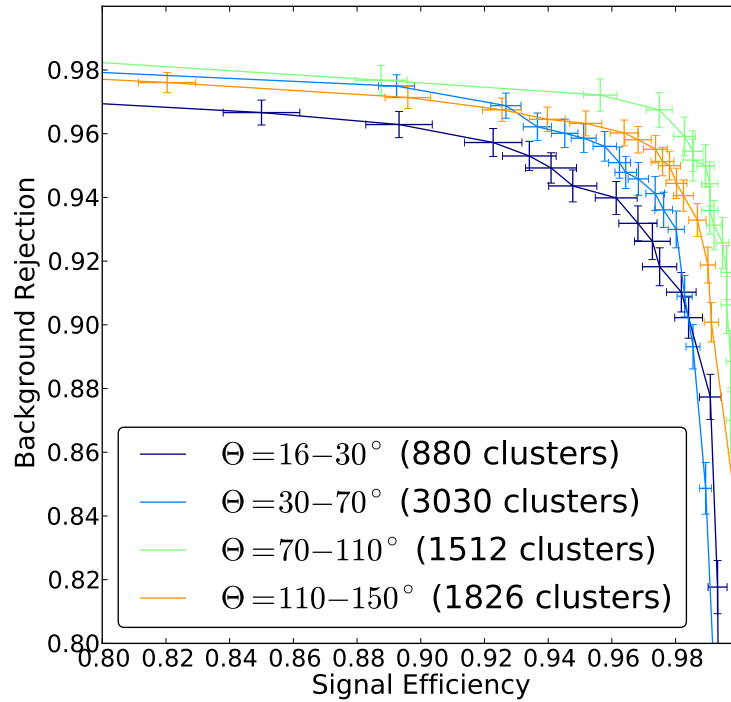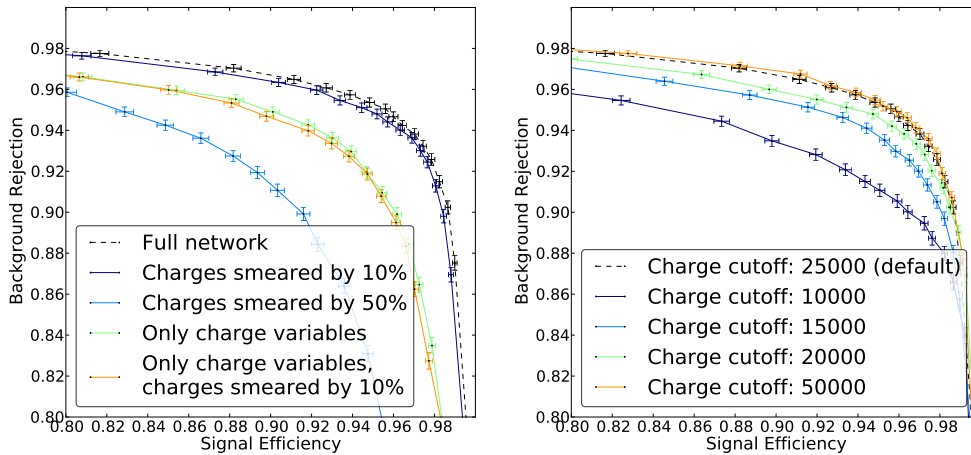


Figure 5.9.: Background reduction over signal efficiency for different cuts on the network output, colours indicate different $\theta$ values of clusters.

charge variables as input. The only network inputs are then the minimal, maximal and total charge, plus the charge variance in a cluster. Also tested was the effect of adding Gaussian smearing to pixel charges, with the width set to 10 % of their value. The smearing was done independently for training and evaluation of the network, and would thus make visible a possible overtraining on the charge variables. As can be seen from Figure 5.10a, the network is quite robust against a slightly lowered charge resolution; overtraining can also be ruled out. Additional tests with a very generous smearing by 50 % show a signal efficiency decrease by only 7 % for 90 % background reduction. Using only charge variables reduces the attainable efficiency by a few per cent, so the performance should not be unduly dependent on the background composition, which differs primarily in the position and direction of clusters. Conversely, for a network that excludes the charge variables, only about 30 % background rejection can be achieved with reasonable signal efficiencies.



(a) using only charge variables as input, and with pixel charges smeared by 10 %

(b) for different pixel charge cutoffs

Figure 5.10.: Background reduction over signal efficiency for networks with some additional limitations.

Some of the attributes of the readout electronics also can adversely affect the network's performance, possibly the most important one being the pixel charge cutoff. Most pixels have charges below 10,000 electrons, with a peak at 1,000, so keeping the maximal pixel charge relatively low would increase the energy resolution for most pixels, while not seeming overly harmful. To distinguish slow hadrons from background, the network makes use of the fact that those clusters tend to have pixels with more than 20,000 electrons deposited, resulting in very high purities. This purity would be unfavourably affected by a too stringent cutoff, as is shown in Figure 5.10b. For the network, a maximal pixel charge of 25,000 seems reasonable, with only very minimal improvements gained by doubling this to 50,000.

Figure 5.11 shows the results for two networks where one of the two background components has been removed entirely. This reveals some slight differences in how well these backgrounds can be separated from the slow pions, but overall, the discrepancy is tiny.
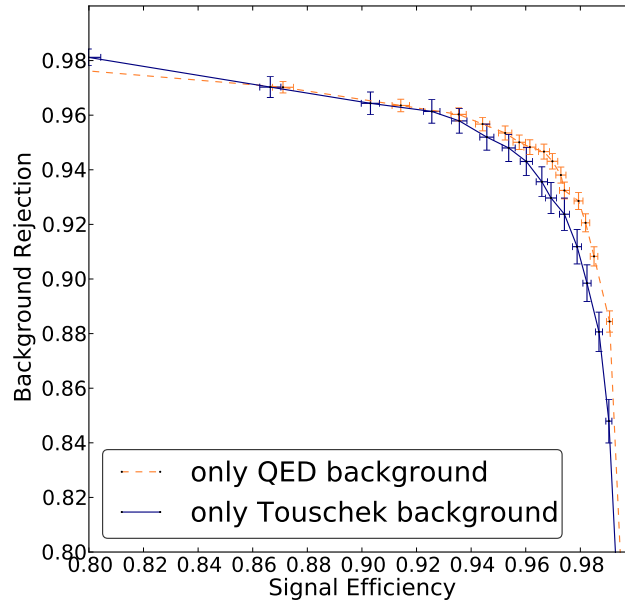
Figure 5.11.: Background reduction over signal efficiency, assuming only one of the considered backgrounds is present.

## 5.6. Summary and outlook

In summary, analysis of PXD clusters should be able to improve upon the tracking data reduction by rescuing slow hadron clusters, with only a relatively small amount of information. One of the main advantages is the ability of choosing a working point for the network to match the available bandwidth. If the necessary background rejection is, as expected from simulation, below the worst case of 90 %, one would profit from the increasing signal efficiency, which quickly reaches values of 99 % and higher.

The network should be fairly robust against changes in the background composition because of the minimally ionising background particles, though simulations of the effect of synchrotron radiation in the PXD are still missing. It is also likely that the estimate for beam–gas Coulomb scattering could be improved, as it currently is simulated using particles shot from two single points. A more realistic simulation might increase the total background, and is necessary for an accurate understanding of the network's performance.

To actually make use of the network, it would need to run relatively close to the readout and needs to be able to cope with the high (unreduced) data rate. Both objectives could be accomplished by implementing the network in a Field Programmable Gate Array (FPGA), which can be found on the ONSEN box slated to perform the selection of PXD clusters in regions of interest (ROIs) for the tracking data reduction. As the ROI selector will remove a sizable number of slow hadron clusters, the network will have to use the same raw data as input. The network's signal output would then be added to the ONSEN's output, in accordance with the total amount of available bandwidth to storage. This setup is unlikely to require additional hardware, so it should be fairly low-cost. A schematic of the data flow in

this setup is shown in Figure 5.12.
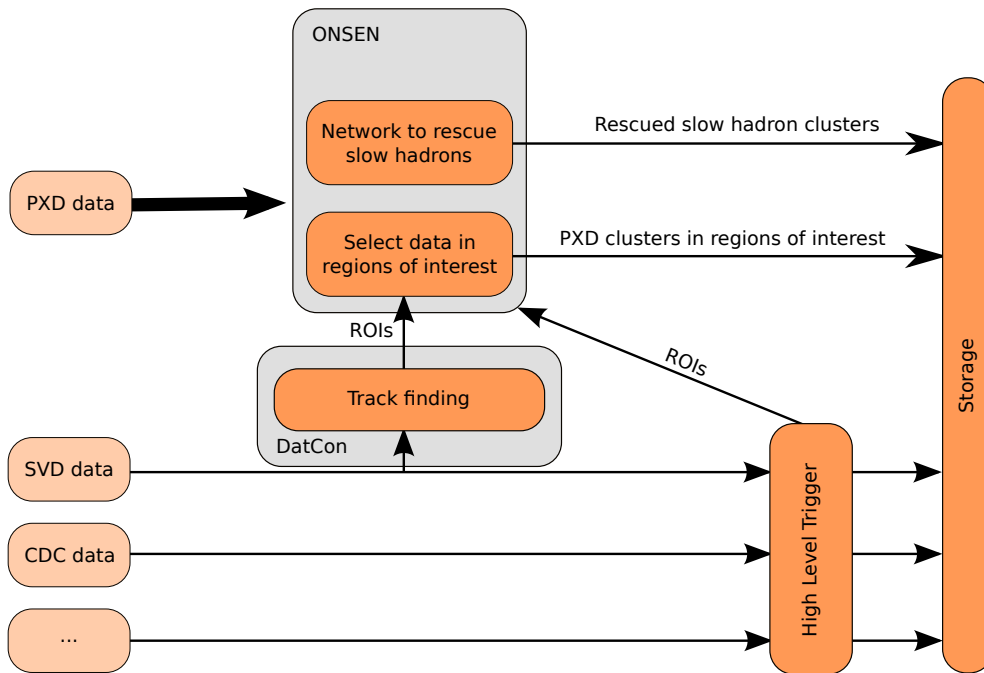


Figure 5.12.: Visualisation of data flow from detector to storage.

Implementing the network itself on an FPGA requires mostly a few additions and multiplications, plus the transformation of the output of network nodes with a sigmoid function, which could possibly be implemented as a lookup table. The decorrelation of variables—requiring a matrix multiplication of the input variables with a transformation matrix to get new decorrelated variables—as well as the flattening and other preprocessing of individual variables should likewise present no large problems, but could also be (at least in part) omitted, as shown in Figure 5.13.

Figure 5.13.: Background reduction over signal efficiency for the full network and a network with no decorrelation and only flattening of variables.

# 6. Particle Identification using $\mathrm{d}E/\mathrm{d}x$

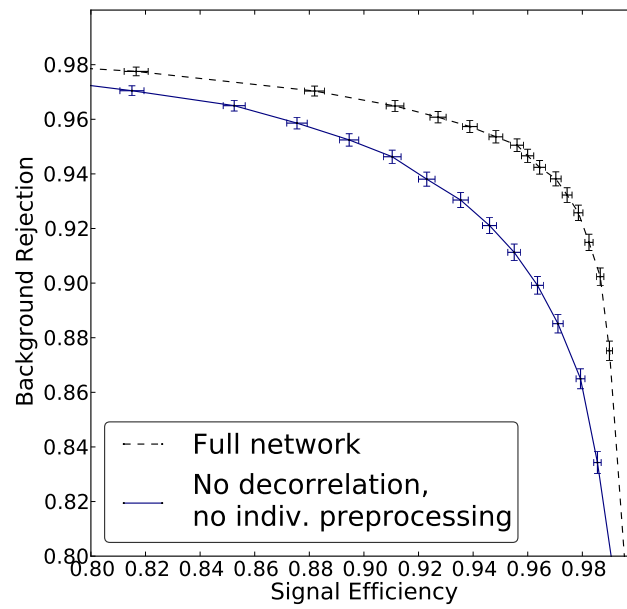The main objective of a particle identification (PID) system is to classify tracks into pions, kaons, protons, electrons, or muons. As shown in Chapter 3, there exist a number of differing interactions for these particles, allowing one to distinguish between these classes. In the scope of this diploma thesis, the ionisation from particles in the tracking detectors will be used for this purpose.

## 6.1. Used data sample

Within the Belle II software framework [11], two basic strategies for the generation of particles exist. One option is to just shoot particles of various types into the detector at different momenta, which can be done using the aptly named `ParticleGun` module. A more realistic approach is to simulate the decay products of the $\Upsilon(4S)$ resonance produced at B factories. To this end, $B^0\overline{B}^0$ and $B^+B^-$ pairs are created using EvtGen [29], an event generator built for the accurate simulation of B decays. The subsequent generic decays are generated using the standard decay file used by the Belle experiment, and should reasonably reflect the particle decays at a B factory. Afterwards, PXD, SVD, and CDC hits are assigned to the Monte Carlo particle they were created by and stored in a track candidate. Realistic track finding is for the moment not used, because in its current state it only works for the drift chamber, and has issues with curling low-momentum tracks. The track candidate is then processed with GENFIT [30], the track fitting toolkit used by Belle II, resulting in track objects which include information about the tracks reconstructed momentum and path. The generated data sample corresponds to 125,000 $B\overline{B}$ events, with a total of around 958,000 reconstructed particle tracks.

This produces a more realistic distribution of the particles, with pions clearly dominating over a large momentum range, as can be seen in Figure 6.1. Kaons also contribute a sizeable fraction of tracks, mostly below 2 GeV, whereas for higher momenta muons and electrons make up almost half of the sample. At around 1 GeV, a few per cent of the whole sample are protons. There is an additional very small contribution from hyperons, which are baryons with strangeness, amounting to about 4 per mill; this component is neglected. The absolute numbers also show that most particles have a momentum around 500 MeV, which quickly diminishes for higher momenta. The particle identification will be evaluated in the momentum range from 0 to 3 GeV, where almost all particles can be found.

Tracks with a transverse momentum of around 250 MeV and curved by the 1.5 T magnetic field have a diameter equal to the outer radius of the central drift chamber. In the used version of the track fitting software, this seems to cause problems, since tracks graze the outer cells of the drift chamber and may be harder to fit to the now larger drift cylinders. As a result, the track fitting efficiency drops for momenta from around 250–420 MeV, causing the double-peak structure seen on the left side. Additionally, the track fitting efficiency also

decreases for low momenta, but in both cases, improvements are expected in the near future.
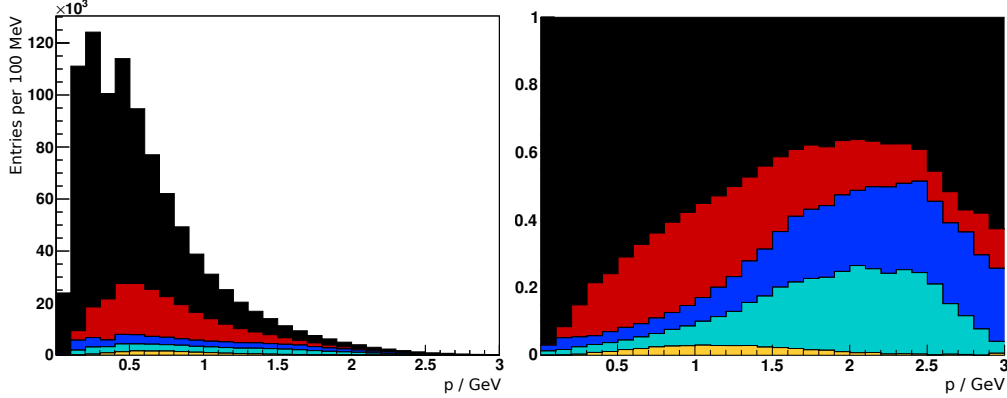


Figure 6.1.: Stacked momentum distributions of the reconstructed tracks by particle type, in absolute and relative numbers. Colours denote, from top to bottom, pions, kaons, electrons, muons, and protons.

## 6.2. Reconstructing d$E$/d$x$ via track information

As described in Chapter 3, the energy loss in the detector creates a corresponding charge signal, which will be available for each tracking detector hit. However, the ionisation is proportional to the traversed distance in the detector medium, so the charge divided by the distance is used instead. Although this is not equal to the specific *energy* loss, they are related by a detector-dependent constant. In the following, the expression d$E$/d$x$ will be used to refer to both quantities.

Calculating d$E$/d$x$ requires reconstructing the traversed length of detector medium, to which end information about the tracks path and detector geometry needs to be combined. Both position and momentum at each hit are available through the track representation of fitted tracks. As the track representation uses a rather time-consuming Runge–Kutta algorithm to extrapolate the track to each hit, taking into account the particle's interaction with the detector material, it may be worthwhile to take the track parameters at the origin and assume and ideal helical track instead. This is discussed in Section 6.4.4.

The charge information for each hit is provided by the hit classes assigned (directly or indirectly) to the track: `PXDClusters`, `SVDTrueHits`, and `CDCHits`. The `PXDClusters` include charge digitisation performed using the `PXDDigitizer` module, which, among other things, simulates the drift of the collected charge through the silicon. A digitisation module for the drift chamber is available, but does not provide charge digitisation, whilst for the SVD, no tested digitisation is available.

For the silicon detectors, the track's momentum at the hit position is used to calculate the approximate angle $\phi$ between the track and the sensor's normal vector. The traversed distance in the detector medium where charge can be accumulated is then equal to $s = \frac{d}{\cos\phi}$, with the sensor thickness $d$. As this diverges for extremely flat incidence angles ($\phi \approx \frac{\pi}{2}$), values of $s$ are limited to the width of the sensor. This approximates the track as a straight

line in each sensor, which is for most cases accurate, and also assumes at most one hit per sensor. Multiple hits per layer, e. g. from overlapping windmill structure, or curling tracks reaching the same sensor twice have no negative impact on the distance reconstruction.
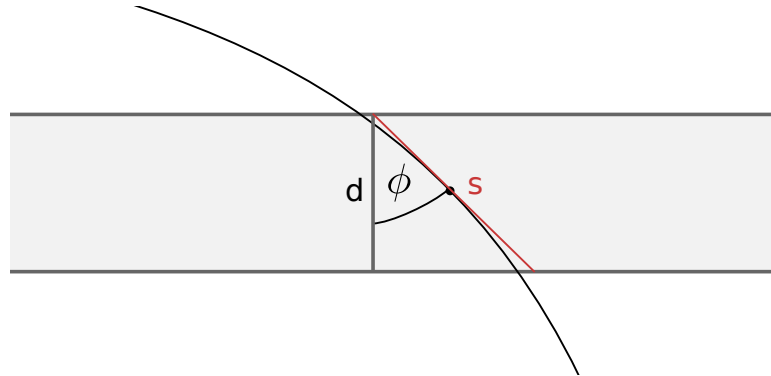


Figure 6.2.: Illustration of the path of a track and the calculated distance *s* in a silicon sensor.

For the drift chamber, the traversed distance for a hit is less well defined, as practically the whole volume of the CDC is active medium. While it is possible to divide the CDC into block-shaped cells, there is no guarantee that all cells crossed by a track will also contain a hit. For this reason, a different construction will be used that aims to avoid this problem: When a track traverses a layer, all hit charges in this layer assigned to the track are summed. The distance is then calculated as the straight line distance to the next layer, using the momentum vector of the track, averaged over the included hits.

This also works for curling tracks, provided the hits are sorted by their position along the track, as only consecutive hits in one layer will be grouped into a single charge/distance measurement. As the distance diverges at the point a track returns to a lower layer, it it limited to a maximum of the distance between consecutive hits in this layer. While this introduces some error at the boundaries to the lower layer, this is unlikely to occur often. The resulting path approximation can be seen in Figure 6.3, where the charges of adjacent hits between layer boundaries are summed and divided by the length of the corresponding straight line segment.
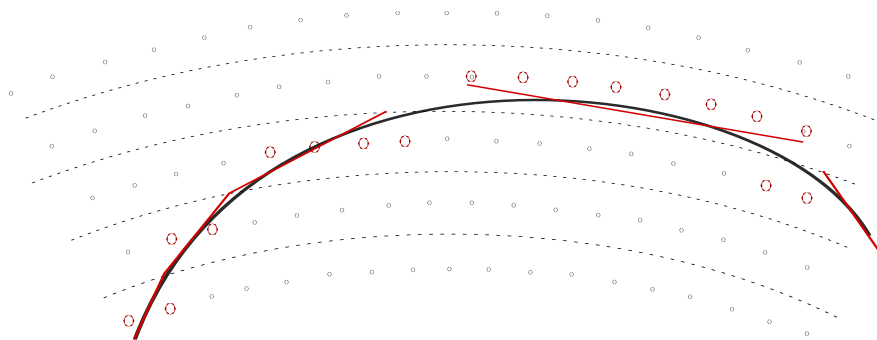


Figure 6.3.: Illustration of the path of a track in the drift chamber, with active wires and approximated distances in red. Dashed lines show layer boundaries.

## 6.3. Likelihood ratio

As explained in Chapter 3, the reconstructed d$E$/d$x$ values follow an approximate Landau distribution. If one wants to combine these individual measurements into a single value, a simple arithmetic mean will not suffice, as the mean of a Landau distribution is undefined. To avoid giving undue influence to values in the distribution's tail, one can instead use a *truncated mean*, where a fixed percentage of the highest and/or lowest values are discarded before the mean is calculated. This results in a high robustness against outliers in a distribution and is commonly used when estimating d$E$/d$x$ values for a track. [14, p. 217]

For these measurements to be useful for the purpose of particle identification (PID), they need to be combined with an estimate of the track's momentum. In this case, the track's momentum at the origin is provided by GENFIT. The effect of energy losses on the momentum is usually small, and thus neglected.
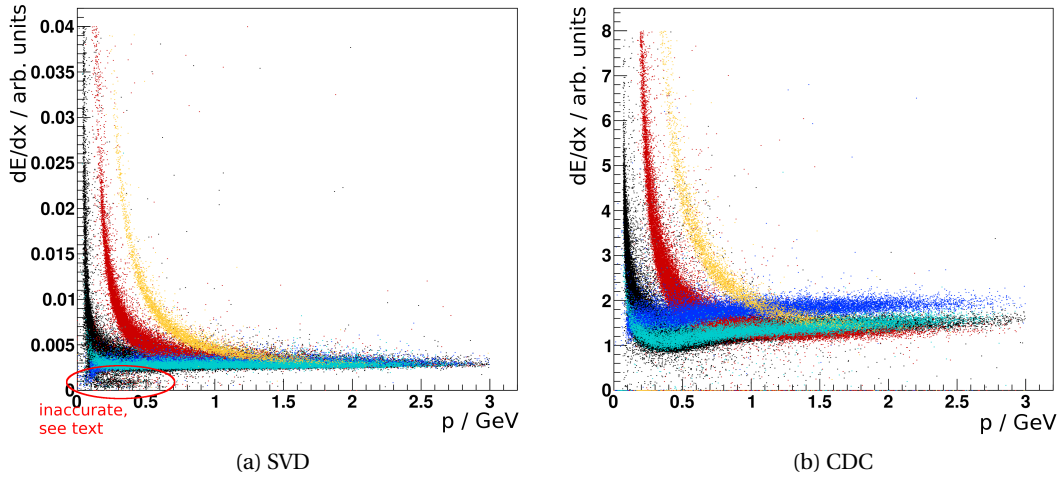


(a) SVD

(b) CDC

Figure 6.4.: Scatter plot of the truncated mean of d$E$/d$x$ values (lowest 5 % and highest 25 % removed) over the track momentum for SVD and CDC. Colours denote electrons, muons, pions, kaons, and protons.

In Figure 6.4 the truncated mean of all d$E$/d$x$ values for a track in one detector is plotted over the track's reconstructed momentum for the silicon vertex detector and the central drift chamber. The corresponding data for the pixel detector has been omitted because the particle behaviour is identical in both silicon detectors and with the PXD's two layers there is not enough data to build a meaningful truncated mean. For the truncated mean used in the plots, the lowest 5 % and the highest 25 % of data points were discarded, and the average of the remaining data was taken. This helps suppress the outliers inherent in the individual d$E$/d$x$ measurements. While it is possible and quite common to use the truncated mean for PID classifications, the truncated mean, at least in theory, discards a lot of information that might be useful for a classification, producing one value from around 50 d$E$/d$x$ measurements on average. Because of this, a classification that uses the information contained in the individual measurements will be considered for this thesis. Differences in performance to the approach using a truncated mean are discussed in Section 6.4.2.

The plots clearly show different behaviour for different particle types: the mass hierarchy of the hadrons (pions, kaons, and protons) is reflected by the position of the rise in energy loss for low momenta. As explained in Section 3.1, the relativistic rise of the energy loss is suppressed by the density effect in silicon, resulting in flat curves once the particle reaches the threshold for minimum-ionising particles. As a result, electrons in the PXD and SVD show about the same energy loss as muons, because even with their much higher $\beta\gamma >$ 100, corresponding to 10 GeV muons, the loss from ionisation is not significantly higher. The cluster of measurements in the SVD circled in red with a $\mathrm{d}E/\mathrm{d}x$ about half of that for minimum-ionising particles is caused by more particles entering the sensor with an angle $\phi$ or $\theta$ between the track and the sensor normal of almost $\frac{\pi}{2}$ at low momentum. This translates to a reconstructed distance many times larger than the thickness of the sensor, which is no longer accurate.

It is evident from looking at these $\mathrm{d}E/\mathrm{d}x$ curves that at low momenta a good separation between the different hadrons should be possible. Additionally, one can expect a good performance for electrons over almost the entire momentum range, particularly if both silicon and drift chamber measurements are used. To actually make use of these distributions for classification, a likelihood ratio method as described in Section 4.1 will be used. For this, one needs to build a likelihood function $\mathscr{L}$ for each particle type hypothesis $m$:

$$\mathscr{L}_m\left(\frac{\mathrm{d}E}{\mathrm{d}x}, p\right) = \prod_i p_m^d\left(\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right)_i^d, p\right), \quad m = \pi, \mathrm{K}, \mathrm{e}, \mu, \mathrm{p}, \quad d = \mathrm{PXD}, \mathrm{SVD}, \mathrm{CDC}$$

Here $i$ loops over all $\mathrm{d}E/\mathrm{d}x$ values assigned to the track, $p_m^d$ is the two-dimensional probability density function (PDF) for the ($\mathrm{d}E/\mathrm{d}x$, $p$) pair. It should be noted that $p_m$ is detector-specific, as the $\mathrm{d}E/\mathrm{d}x$ values vary greatly between the different subdetectors. Because of the small individual factors involved and the limited range of floating point numbers, the product is handled as an addition of log-likelihoods, which should avoid the numerical inaccuracies.

In Section 4.1 likelihoods for two hypotheses $H_0$ and $H_1$ were combined into a test statistic useful for distinguishing them from each other by taking the ratio $R = \mathscr{L}_0/\mathscr{L}_1$. For five mutually exclusive hypotheses that need to be combined, pairwise likelihood ratios like these are not very useful. Instead, one can use the sum of the likelihoods for all hypotheses in the denominator, including the one already in the numerator. Of course, the considered hypotheses are not all equally likely, and differ in their momentum distribution (cf. Figure 6.1). This is taken into account by multiplying the likelihood $\mathscr{L}_m$ with a *momentum prior* $P_m(p)$, which is normalised so that $\sum_m P_m(p) = 1$ for all values of $p$.

$$R_m\left(\frac{\mathrm{d}E}{\mathrm{d}x}, p\right) = \frac{P_m(p)\mathscr{L}_m\left(\frac{\mathrm{d}E}{\mathrm{d}x}, p\right)}{\sum_k P_k(p)\mathscr{L}_k\left(\frac{\mathrm{d}E}{\mathrm{d}x}, p\right)}$$

This test statistic has the advantage of corresponding directly to the probability of a track being of a certain type $m$, assuming the PDFs are accurate.

The two-dimensional probability density functions $p_m(\mathrm{d}E/\mathrm{d}x, p)$ are constructed directly from the Monte Carlo sample as *template PDFs*, which are simple histograms, using a rather fine binning of $100 \times 100$ bins to ensure a sufficiently high resolution in areas where the $\mathrm{d}E/\mathrm{d}x$ measurements of some particles overlap. While the histograms have a high total number of entries, e. g. about three million for the pion histogram for the SVD, most entries

can be found in a rather narrow band around the theoretical value given by the Bethe–Bloch equation. To avoid giving undue influence to bins with a very low number of entries (i. e. outliers), bins with less than four entries are discarded. When the PDF is later evaluated in a bin that is empty, a very low probability is taken instead of zero ($10^{-5}$). These somewhat arbitrary changes to the PDFs give an impression of the extent of the statistical uncertainties involved in the particle identification: The modifications mostly affect the high momentum region above 2 GeV, where the number of particles drops off rapidly (only about 1.6 % of all particles have a momentum that high). In other momentum regions, no significant changes are observed.

To build a correct test statistic, it is vitally important that the PDFs are correctly normalised. Since the momentum distribution of particles is already included in $R_m$ through the momentum prior, this will be done independently for each momentum bin. Then, for each particle type and each detector, the content of every $\mathrm{d}E/\mathrm{d}x$ bin will be divided by the number of all entries with this particle type in the current momentum bin.

## 6.4. Evaluation

To evaluate the performance of the resulting classification, purity–efficiency curves can be used. For each cut $c$ on the test statistic $R_m$, these quantities are defined as

$$\mathrm{purity}_m(c) = \frac{\text{particles with } R_m > c \text{ of type } m}{\text{particles with } R_m > c},$$

$$\mathrm{efficiency}_m(c) = \frac{\text{particles with } R_m > c \text{ of type } m}{\text{particles of type } m}.$$

As before, statistical uncertainties are calculated according to [28].

Because of the variations in the possible separation by momentum, the purity and efficiency are determined using only particles in a small momentum range, as is shown in Figure 6.5 for momenta $p$ just below 1 GeV. All curves show that there is a trade-off between the purity of the selected sample and the corresponding efficiency: users of the particle identification have to select a *working point* on each curve that yields a purity and efficiency sufficient for the purposes of their analysis. In general, purity–efficiency curves should be monotonous, i. e. a higher efficiency should mean worse (or at best, equal) purity for the selection.

In the example, the purity–efficiency curves for both protons and electrons show a very good separation, whereas for pions and kaons it is not possible to simultaneously achieve both high efficiencies and purities. This can be understood by comparing with the region around 0.9 GeV in Figure 6.4, where pions and muons almost entirely overlap each other, whereas some kaons can be seen as clearly separate. For muons, any selection with reasonable purity requires sacrificing most of them, which is typical due to their pion-like energy loss and their low rate. Protons and electrons, on the other hand, show a much higher energy loss in this momentum region and are easily separated from the other particles. Adding information from the silicon detector, as was done here, electrons and kaons can also be distinguished in this momentum region. For efficiencies near 1, the purity reflects the a priori probability, i. e. the fraction of all particles that are of the considered type. This explains why the electron and proton curves start at significantly lower purities than pions or kaons.
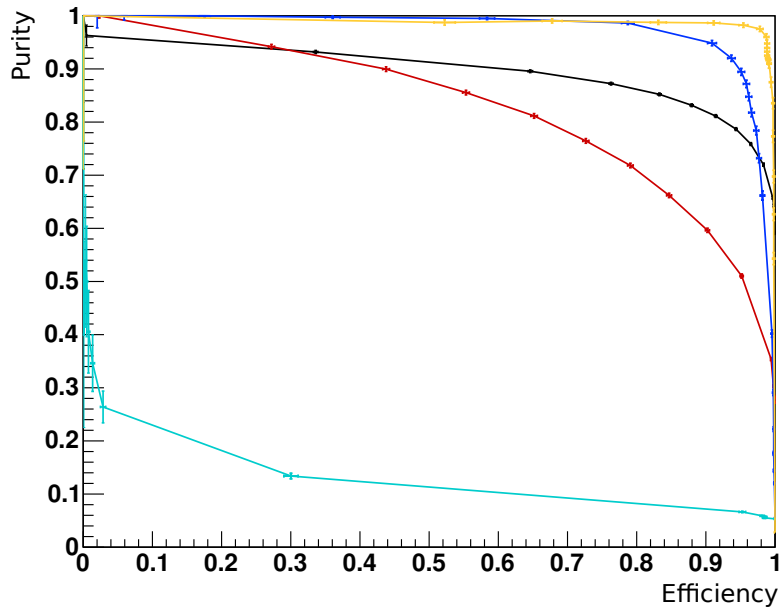
Figure 6.5.: Example purity–efficiency plot for tracks with $p$ in [860 MeV, 940 MeV], using information from individual PXD, SVD, and CDC hits. Colours denote electrons, muons, pions, kaons, and protons.

To show the variations for different momenta explicitly, it makes sense to fix the purity to some value, and plot the efficiency over the momentum. This representation hides most of the information contained in the individual purity–efficiency plots, but can be used to easily compare the performance of the separation for different particle types and momenta. In momentum regions where the attainable purity drops below the requirement, the efficiency is zero. This indicates a bad separation, but does not mean there is no information in the sample. Figure 6.6 demonstrates this for the same classification already shown in Figure 6.5: for electrons above 1 GeV, efficiencies above 90 % are reached, for lower momenta this is reduced by the overlap with pions. The efficiency for pions and kaons is very high for momenta below 700 MeV, with a slight dip for 100 MeV kaons. For protons, the separation is best around 1 GeV, the same momentum region that also contains the majority of protons. The purity for the selection of muons almost never reaches the desired 95 per cent, which illustrates the need for other detectors to help distinguish them.

It should be noted that the efficiency values for fixed purity are taken directly from the purity–efficiency plots, by taking the highest efficiency which has a purity at least as high as required. This means there is no interpolation between different points and the resulting efficiency values are likely to be an underestimate, in particular for high purities. The error bars show statistical errors on the determined efficiency only, and do not reflect binning effects.

As we have already seen, the d$E$/d$x$ method is especially useful for identifying hadrons with a momentum below their minimum of ionisation. Of particular importance for physics analyses is the performance for pions and kaons, which will be the main focus of the subsequent considerations.
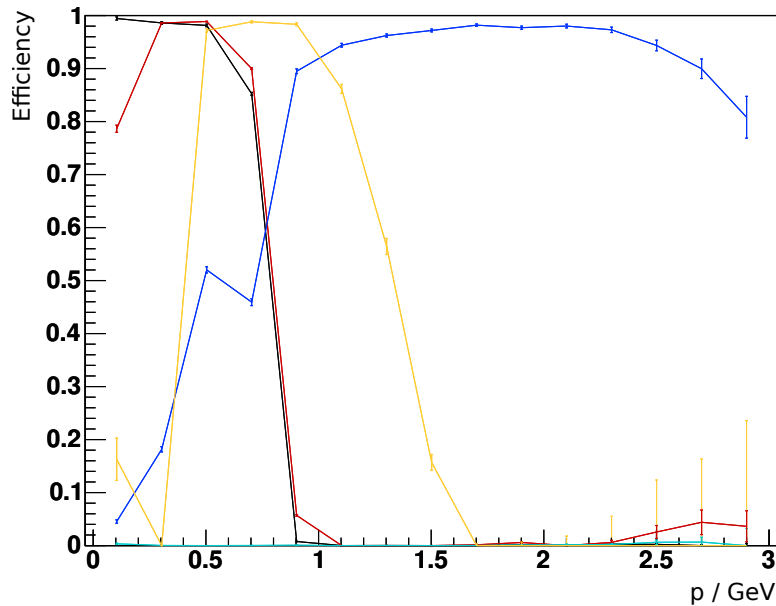
Figure 6.6.: Example efficiency–momentum plot for a fixed purity of 95 %. Colours denote electrons, muons, pions, kaons, and protons.

In the following, a number of different approaches to the d$E$/d$x$ likelihood ratio method will be considered. Differences may occur not only in their resulting PID performance, but also in their robustness against "overtraining" and their computational efficiency. The classification using PDFs from all three tracking detectors, using individual d$E$/d$x$ measurements, will be used as a point of reference for these comparisons, as removing information—either by omitting detectors or by using mean values of measurements—is not expected to yield a better separation.

### 6.4.1. Using hits from silicon detectors

It is clear from the d$E$/d$x$–momentum plots for SVD and CDC (Figure 6.4) that the information from energy loss in silicon sensors and a gas volume is at least somewhat complementary. This is very visible in the case of electrons, but the silicon detectors may also provide additional information for very low-$p_t$ tracks that do not produce a large number of hits in the drift chamber.

To evaluate how large the effect of adding SVD and/or PXD data is, three classifications were made using data from (a) only CDC, (b) SVD and CDC, and (c) all three tracking detectors. The results can be seen in Figures 6.7–6.9, again for purity fixed to 95 %, but with a somewhat finer binning.

The efficiency–momentum plot for a classification that only uses CDC data shows problems with the separation in regions where the d$E$/d$x$ measurements for the particle in question overlap with the electron band. This is very visible for kaons with a momentum of 0.5–1 GeV, but also for protons just above 1 GeV. For lower momenta, the separation is decent, especially the pion–kaon separation below 500 MeV seems to be reasonable. It is quite possible that the observed deficiencies caused by the electrons in the sample can
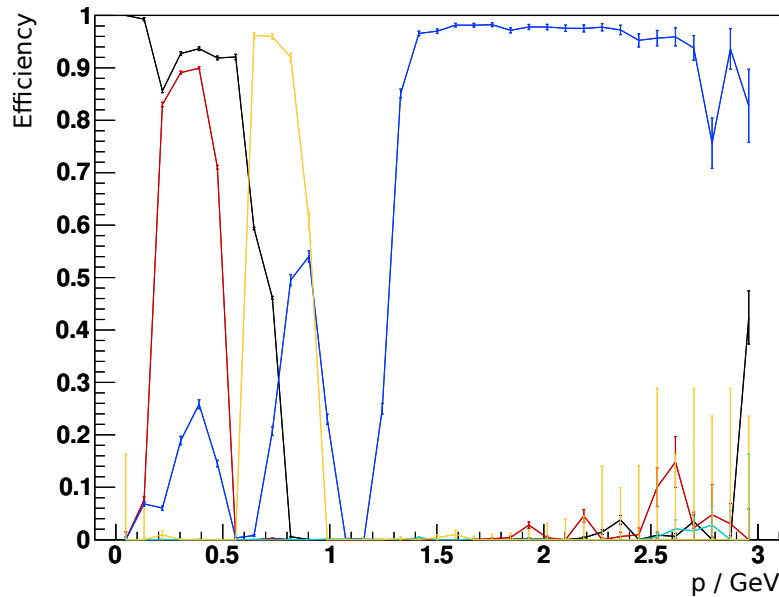
Figure 6.7.: Efficiency–momentum plot for a classification using only CDC data, with a fixed purity of 95 %. Colours denote electrons, muons, pions, kaons, and protons.

be solved by adding information from the outer detectors, especially the electromagnetic calorimeter (ECL).

Adding information from SVD hits to the final likelihood for each hypothesis yields results that are significantly better than with drift chamber data alone, as Figure 6.8 shows. The improved electron separation below 1 GeV greatly increases the efficiency of the selection of hadron candidates, including pions. In particular, kaons can now be selected with high purity and an efficiency of up to 98 % for momenta from 200 to 800 MeV; the proton separation also works fine in a fairly broad momentum region.

The effect of also adding PXD data, i. e. using information from all three tracking detectors, can be seen in Figure 6.9. There is no additional large effect, with only minor changes in the attainable efficiency. In some cases, the efficiency actually seems to decrease, e. g. for pions with $p \approx 700$ MeV. Looking at the underlying purity–efficiency plots, which are shown overlaid in Figure 6.10, the selection efficiency actually appears to improve by a few per cent for electrons, kaons, and also pions. As mentioned before, the fixed-purity plots can underestimate efficiency through binning effects, which in this case is caused by a tiny variation in the pion selection. For moderate purities (70–80 %) and track momenta < 1 GeV, adding data from the pixel detector actually improves the selection efficiency a bit, similar to what is seen in Figure 6.10.

It appears that using the SVD data is quite worthwhile for a good performance, as it restores separation for low momenta with only 6 layers of data, i. e. low computational expense. The pixel detector does not improve things by much, and it is doubtful whether it makes sense to include it.

Figure 6.8.: Efficiency–momentum plot for a classification using SVD and CDC data, with a fixed purity of 95 %. Colours denote electrons, muons, pions, kaons, and protons.
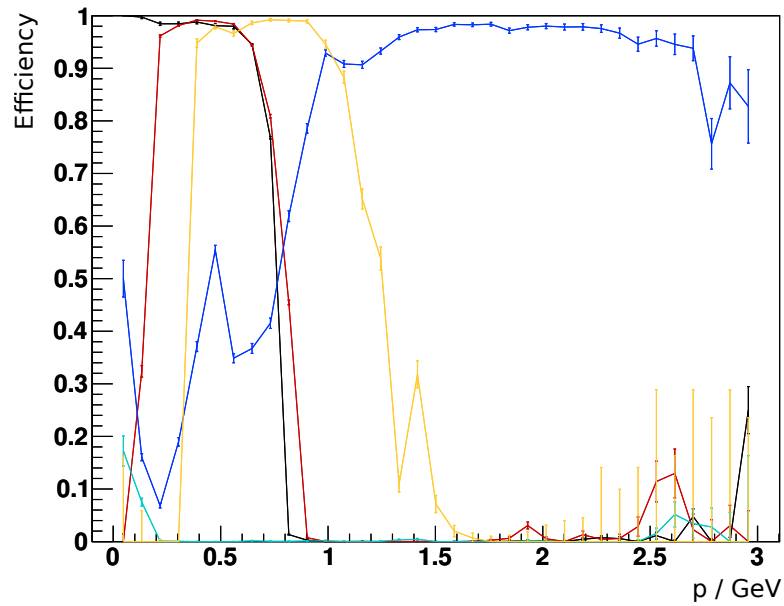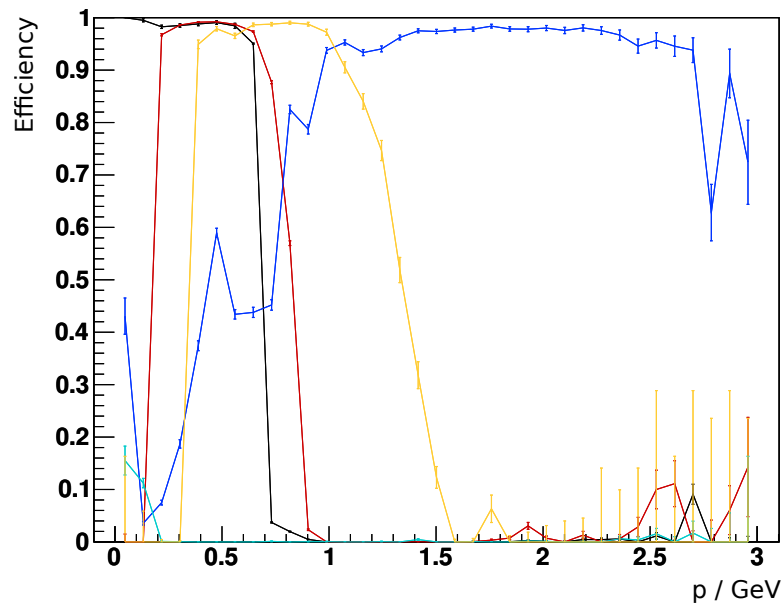


Figure 6.9.: Efficiency–momentum plot for a classification using PXD, SVD, and CDC data, with a fixed purity of 95 %. Colours denote electrons, muons, pions, kaons, and protons.
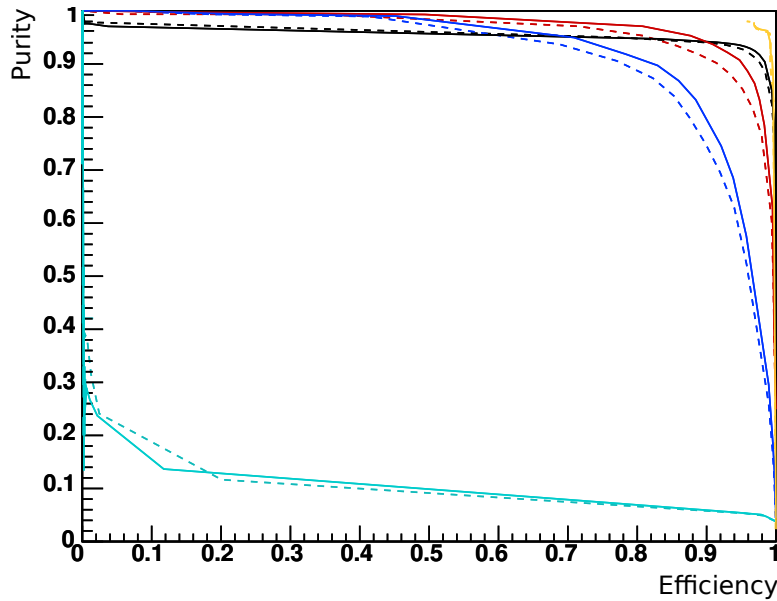
Figure 6.10.: Purity–efficiency plot for $p$ in [690 MeV, 770 MeV], for a classification including PXD data (solid) and with only SVD and CDC (dashed). Errors have been omitted for legibility, but are about 1 percentage point for electrons, muons, and protons; less for pions and kaons. Colours denote electrons, muons, pions, kaons, and protons.

## 6.4.2. Comparison with truncated mean

As mentioned earlier, the truncated mean is a commonly used method to suppress the outliers of the Landau distribution of d$E$/d$x$ measurements. It is also regularly used directly for classifications instead of the individual data points. For example, the predecessor of Belle II, the Belle experiment, used a truncated mean with the top 20 % cut away for its d$E$/d$x$ particle identification [31]. But how does this actually compare to a classification making use of individual d$E$/d$x$ measurements?

For the truncated mean, the likelihood function does not contain factors for each element of the full d$E$/d$x$ vector, but is equal to the product of the values of the PDFs for the measured $\big((\mathrm{d}E/\mathrm{d}x)_{\mathrm{trunc}}, p\big)$ pairs for each included detector. While this can very slightly speed up the classification, most of the run time is taken up by the reconstruction of distances.

Since the truncated mean can be adjusted to cut away more or less data points on both sides, it includes a whole class of estimators, that ranges from no truncation at all (arithmetic mean) to the symmetrical removal of all data points save one (median). Because of this, classifications with different cut-offs will have to be compared. For the underlying landau distribution, asymmetric cutoffs are appropriate, as were used for Figure 6.4 with the lowest 5 % and the highest 25 % removed. Here, successively harder cuts of 3 % and 15 %, 5 % and 25 %, as well as 8 % and 40 % for the lower and higher bound will be compared. This of course does not cover the entire parameter space, but should give an impression of the effect of these variations and the possible separation. For the pixel detector, where more than two hits are quite rare, the average will be used instead of a truncated mean.

Comparing Figure 6.11 with a classification that includes individual d$E$/d$x$ measurements in the likelihood functions (Figure 6.9), one notices a harsh drop in reachable efficiency over almost the entire momentum range. For pions and kaons, the separation is comparable to that attainable using only CDC hit data, with slightly higher purity. The fluctuations in the selection efficiency of protons point to a purity–efficiency curve that almost falls short of the necessary purity, due to still present d$E$/d$x$ outliers from the other particles. The relatively low efficiencies for electrons may be somewhat surprising, given that the electron band in Figure 6.4 (b) seems to have no noticeable overlap above 1.5 GeV momentum. However, as we require a 95 % purity for our electron selection, this merely indicates that the region of the electron band still contains contributions of at least 5 % from other particles at least in some regions. When individual hits are used, it is quite likely that a reasonable fraction of the d$E$/d$x$ measurements for a non-electron track lie below the electron band, which, through being included in the final likelihood, would make the electron hypothesis much less likely.
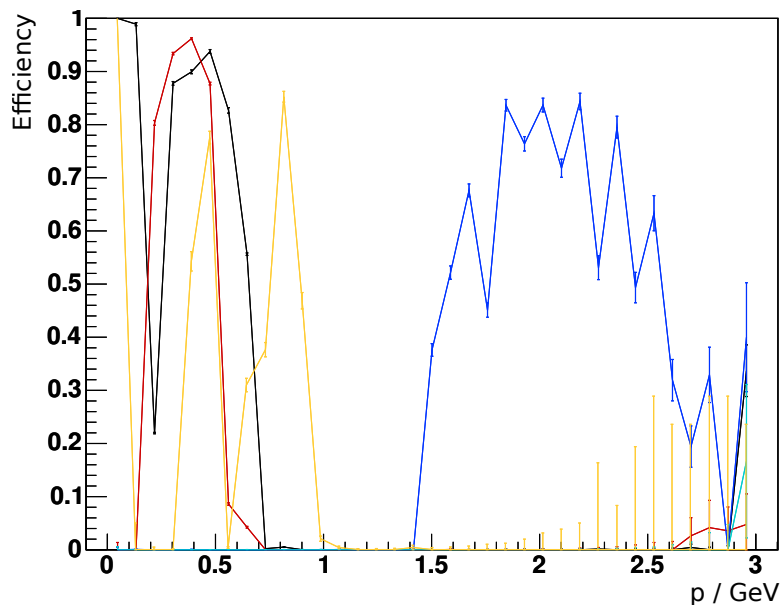


Figure 6.11.: Efficiency–momentum plot for a classification using a 3 %/15 % truncated mean, with a fixed purity of 95 %. Colours denote electrons, muons, pions, kaons, and protons.

Figures 6.12 and 6.13 show further classifications using the truncated mean, with an even larger portion of measurements excluded. The observable differences tend to be confined to single momentum bins, and are caused by relatively minor fluctuations in the underlying purity–efficiency plots. Only for the electron selection efficiency does there seem to be a common trend, where the efficiency over a broad range seems to improve with more measurements being excluded.

### 6.4.3. Fitting distributions

Until now, all 15 d$E$/d$x$ probability density functions (five particle types × three detectors) were saved as two-dimensional (template) histograms with a fairly high number of bins.
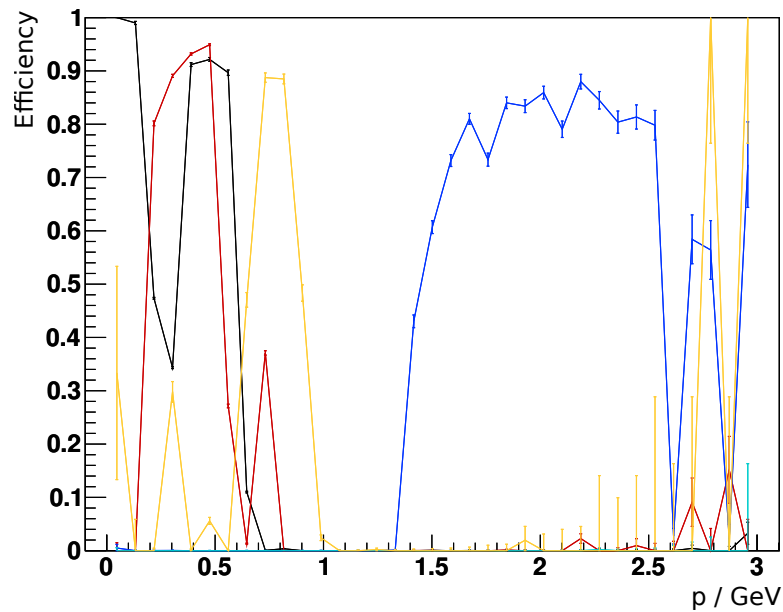
Figure 6.12.: Efficiency–momentum plot for a classification using a 5 %/25 % truncated mean, with a fixed purity of 95 %. Colours denote electrons, muons, pions, kaons, and protons.
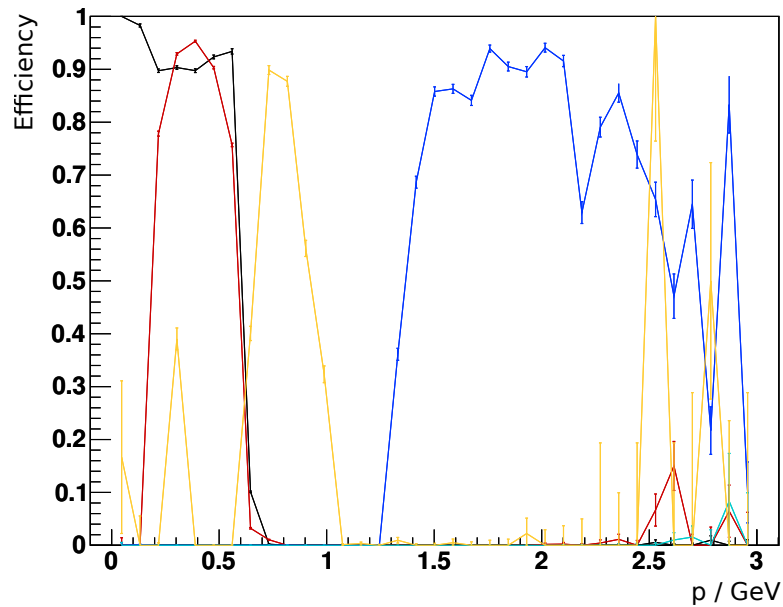


Figure 6.13.: Efficiency–momentum plot for a classification using a 8 %/40 % truncated mean, with a fixed purity of 95 %. Colours denote electrons, muons, pions, kaons, and protons.

Of those, a good portion are empty or contain only very few entries, which may lead to the classification learning more about the statistical fluctuations of the data sample than is acceptable. This can be avoided by parametrising the histograms using their known shape and generating simplified PDFs from these parametrisations. The procedure and its effect on the classification will be discussed in this section.

For our purposes, the two-dimensional PDFs can be described by the Bethe–Bloch function for the d$E$/d$x$–$p$ shape, and something resembling a Landau distribution for the spread of the d$E$/d$x$ values. To avoid the complexity of a full two-dimensional fit, the distribution of d$E$/d$x$ values will be fitted in a number of small momentum regions, for each detector. For the shape, a Landau distribution convolved with a Gaussian for resolution effects will be used. The fit is performed using the RooFit [32] toolkit using a likelihood fit.

For a track momentum around 1 GeV, Figure 6.14 shows the resulting fitted shape plus data points for kaons and electrons. The distributions and fit results for those two particles are typical and clearly show the differences between the different detectors: for the PXD the fit seems to be perfectly fine, but for SVD and CDC some deviations can be seen. The description of the energy loss peak in the SVD appears to be accurate, but higher d$E$/d$x$ values seem to be less common than expected. For the CDC, there are also some deviations visible in the tail. This becomes clearer when looking at the distribution of the pulls, defined as

$$\text{pull} = \frac{x_{\text{data}} - x_{\text{fit}}}{\sigma_{\text{data}}}.$$

These show nearly perfect agreement for the PXD, whereas for the SVD divergences of over three sigma can be seen above d$E$/d$x$ values of 0.02, where the amount of energy loss is overestimated. The deviations in the drift chamber, however, are much larger and clearly visible even without considering the pull distributions. The cause of these differences is unknown, but could arise from inaccuracies in the d$E$/d$x$ reconstruction, or the higher statistics of the SVD and CDC.

Between different momentum bins or particles, the width of the resolution term does not change significantly and is set to $10^5$, $2 \cdot 10^{-4}$, and $4.4 \cdot 10^{-7}$ for the PXD, SVD, and CDC, respectively. The mean value of the Gaussian is set to zero. This stabilises the fit and halves the number of parameters. The remaining two parameters, the most probable value (MPV) and width of the Landau distribution, vary with momentum and can also be parametrised in a separate fit. Note that both are parameters of the fitting function and do not correspond directly to the maximum of the distribution or the full width at half maximum (FWHM). [33] The momentum dependency of the most probable value can be described a Bethe–Bloch curve, as was expected. While the width parameter shows a similar behaviour, this is not caused by any physical spread of the energy loss, but an effect caused by the momentum binning. In regions where the most likely d$E$/d$x$ measurement changes rapidly with the momentum, the measurements in one momentum bin will show a much larger spread, leading to an apparent increase of the width of the Landau distribution. Because of the similar shape, this is also parametrised using a Bethe–Bloch function, but without any linear term.

The momentum dependency of both parameters is fitted using two independent $\chi^2$ fits for every PDF, as is shown for kaons in the SVD and electrons in the PXD in Figure 6.15. Each PDF can then be constructed from only 7 fit parameters (plus the 3 resolution parameters) by calculating the Landau MPV and width and saving the resulting one-dimensional PDF in the
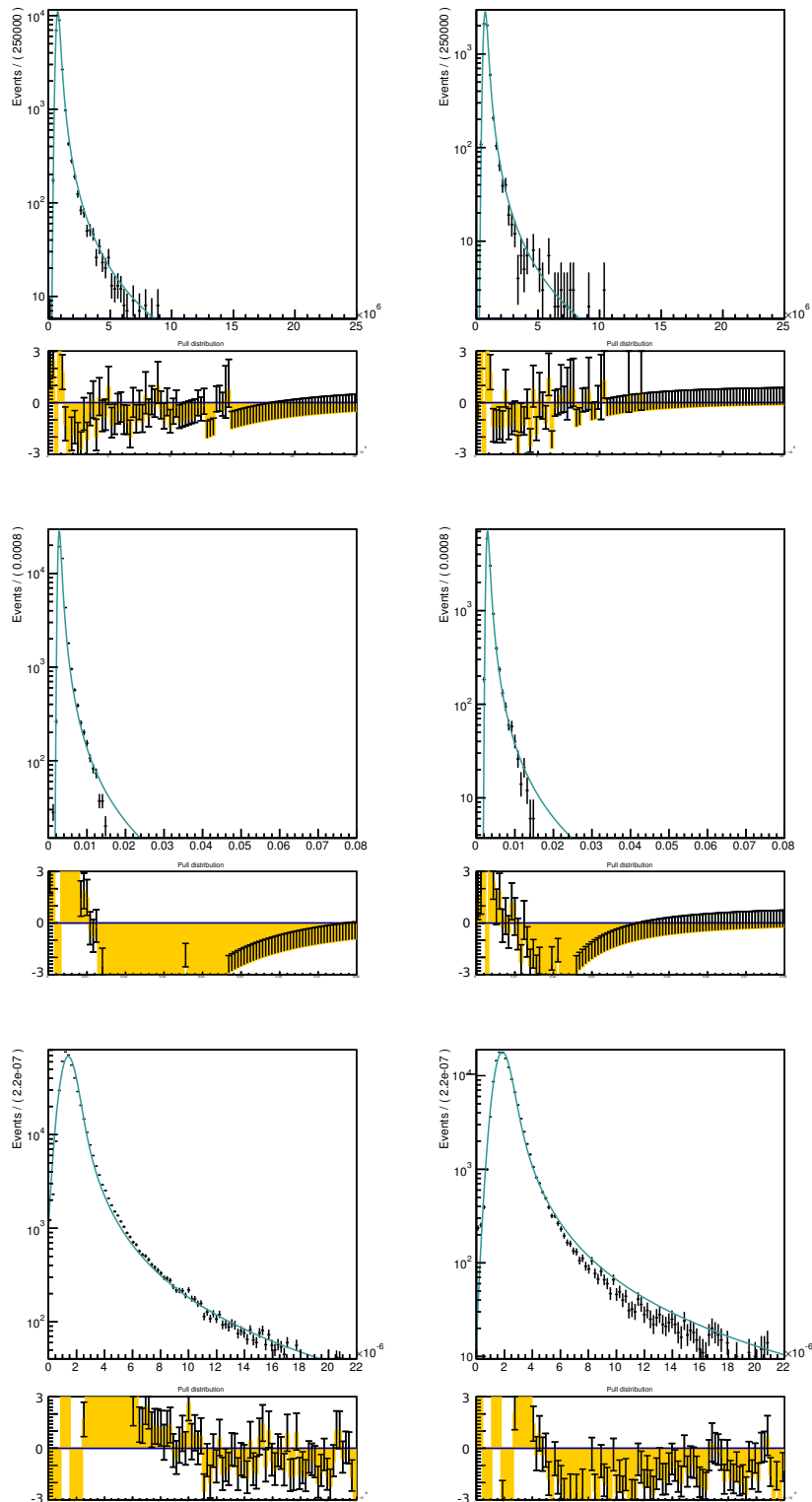
Figure 6.14.: Fits to the d$E$/d$x$ distributions of kaons (left) and electrons (right) in the PXD (top), SVD (middle), and CDC (bottom), shown with a logarithmic scale on the vertical axis. The pull distributions shown below each plot indicate the deviation of data points from the fitted shape.

required number of momentum bins. Since the repeated calculation of the value of the 2D PDFs is rather expensive—the Bethe–Bloch shapes must be evaluated at the given momentum and the Landau and Gaussian shapes need to be convoluted, which involves a Fourier transform—the values are precalculated and saved in a 100×100 histogram. Figure 6.16 compares some of the probability density functions constructed using the fit results with those obtained directly from data. The main difference between them is the graininess of the data PDFs in regions with low probability, whereas the fit PDFs show a smooth distribution. The white regions in both diagrams are for the most part an artifact of the chosen logarithmic scale, and still contain non-zero entries. Of course some regions of the data PDFs, mostly with high momenta, are actually empty. Yet, the PDFs from fit results also cover these low-statistics areas and provide sane probability densities even for very high d$E$/d$x$ values. Some deviations occur for values below the main band, where the data PDFs contain a few entries, but the Landau distribution drops to almost zero, with values much lower than in the high d$E$/d$x$ tail.

It should be noted, however, that these Bethe–Bloch fits are rather unstable and a considerable amount of fine-tuning is required to get decent fit results for all fifteen PDFs. For this reason the PDFs created using the results of a successfully fitted generic sample of 66,000 B$\overline{\text{B}}$ decays were used for the classification shown in Figure 6.17. Compared to a classification using template PDFs (see Figure 6.9), the selection of kaons is slightly less efficient, while the efficiency for electrons shows an obvious increase for momenta between 500 MeV and 1 GeV. However, this latter increase is caused by the different sample, which used a previous version of the tracking software that was less efficient for low-energy pions.

Overall, the observed differences are not large, indicating that with a Monte Carlo sample of the employed size, the template PDFs do not contain too much information about the sample's statistical fluctuations. Additionally, the templates can be constructed without manual intervention and can include features of the data that are not easily parametrised, such as the deviations for lower d$E$/d$x$ values.

### 6.4.4. Discussion of track extrapolation

For the classifications shown previously, the track direction information necessary for the d$E$/d$x$ reconstruction was obtained by assuming a helical track initialised with the fitted track parameters at the origin (provided by GENFIT). This introduces some inaccuracies as possible energy losses and multiple scattering effects are neglected. GENFIT's track representation itself is also capable of providing the track momentum at a specific hit, which for the employed `RKTrackRep` involves a Runge–Kutta algorithm stepping through the detector from the last known track state, taking into account interactions with the material, and updating track position, momentum and the associated error matrix. [30] As this needs to be done for each hit in the tracking detectors, the performance impact of the Runge–Kutta extrapolation can be quite large.

The alternative helix extrapolation employed so far is somewhat simpler: for each hit, the closest point on the helix is searched using a one-dimensional minimisation, where the position on the helix is determined using simple geometry for each minimisation step. It is also possible to combine both methods: to make sure the helix approximation remains accurate during the track's passage through the CDC, the distance of the reconstructed hit
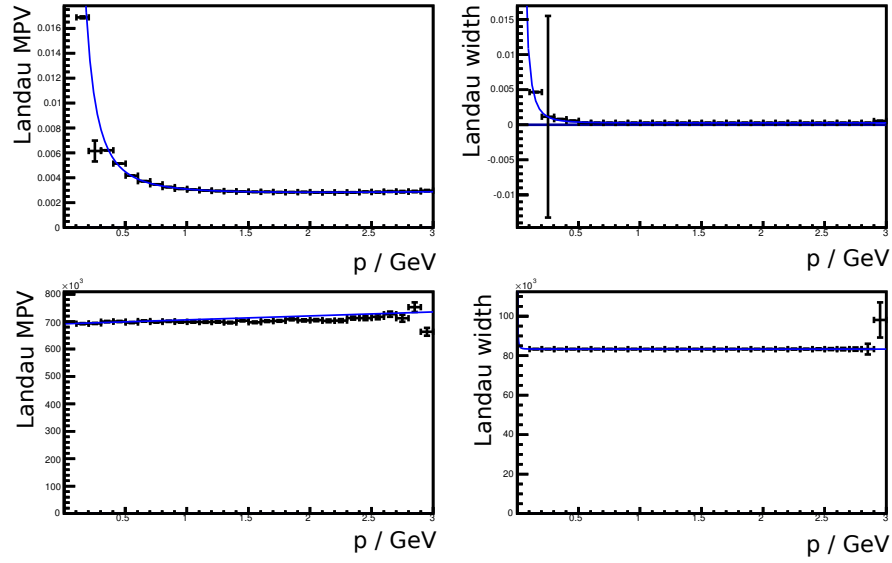
Figure 6.15.: Successful fits of the momentum dependence of the MPV (left) and width (right) parameters of the d$E$/d$x$ distributions, shown here for kaons in the SVD (top) and electrons in PXD (bottom). Dots with error bars show the results of d$E$/d$x$ distribution fits in individual momentum bins with their fit errors, solid lines the fitted Bethe–Bloch-like shape.
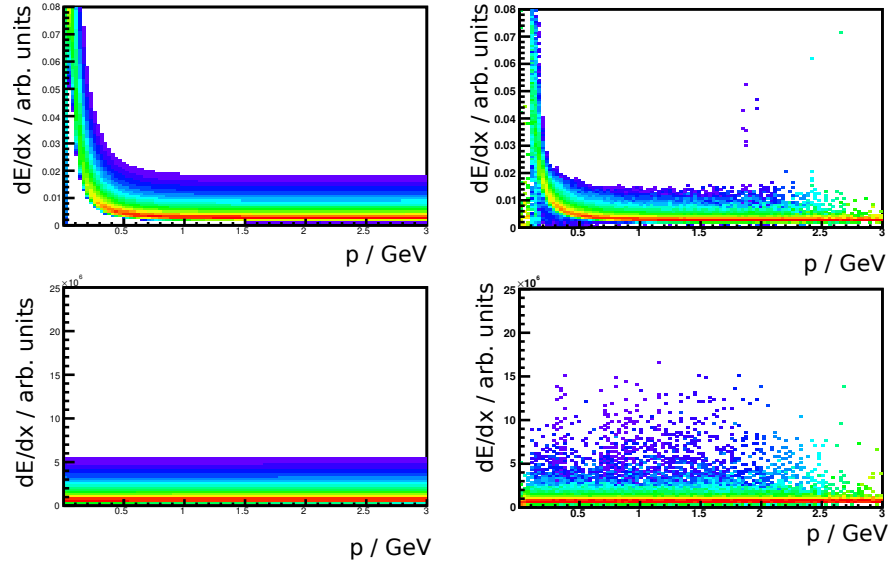


Figure 6.16.: PDFs created using the fit results of Figure 6.15 (left), compared with the corresponding PDFs from data (right). The top row shows probability density functions for kaons in the SVD, the bottom row for electrons in the PXD. Colours show probability densities on a logarithmic scale ranging from close to 1 (red) over $10^{-1}$ (yellow) and $10^{-2}$ (blue) to $10^{-3}$ (violet).

Figure 6.17.: Efficiency–momentum plot for a classification using PDFs created from fit results, using information from individual hits in all three tracking detectors. For this cla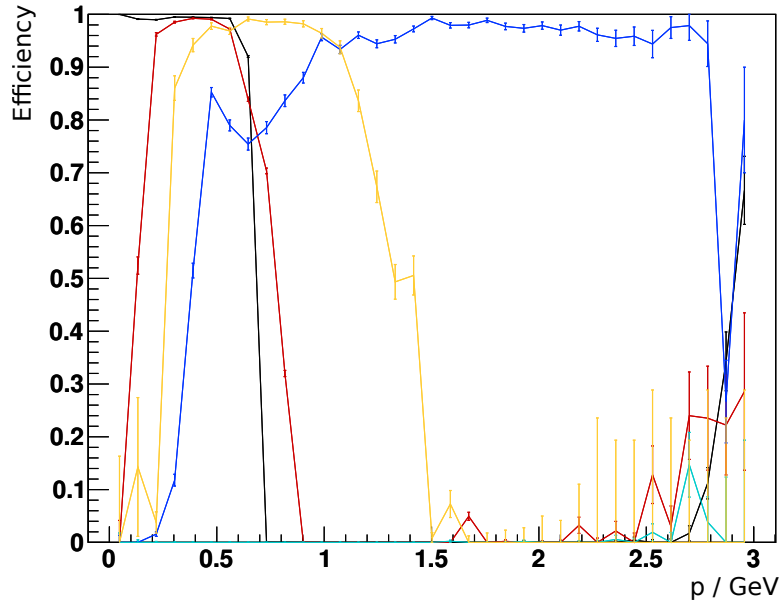ssification, a different B$\overline{\text{B}}$ sample was used (see text). Colours denote, from top to bottom, pions, kaons, electrons, muons, and protons.

to the closest point on the helix is calculated for each hit. Should the distance exceed some fixed limit, the `RKTrackRep` extrapolation is then used to reinitialise the helix parameters.

Table 6.1.: Comparison of the required CPU time needed by the d$E$/d$x$ reconstruction. The `RKTrackRep` extrapolation at the origin is not included in the final count. One reconstructed event contains 8 tracks on average.

|  | Time per event / ms | `RKTrackRep` extrapolations per event |
| --- | --- | --- |
| RKTrackRep | 1429 | 399.1 |
| Helix | 50.0 | 0 |
| Hybrid approach | 136.6 | 27.7 |

Table 6.1 shows the results of a sample of 125,000 Monte Carlo events being processed using the d$E$/d$x$ reconstruction module, either using `RKTrackRep`, a helix, or a hybrid approach for the track extrapolation. As can be seen, the `RKTrackRep` extrapolation is about 28 times slower than the helix approximation and takes more than one second per event (or 50 CPU hours for the whole sample). Clearly this is unacceptable performance for an application that is supposed to run on practically all of the data taken. The hybrid approach, which reinitialises the helix parameters using the track representation once the helix more than 4 cm away from a CDC hit, is almost three times slower than the helix approximation, but still more than ten times faster than the pure `RKTrackRep` extrapolation. In principle, this distance cut can be adjusted to gain acceptable performance without excessively sacrificing
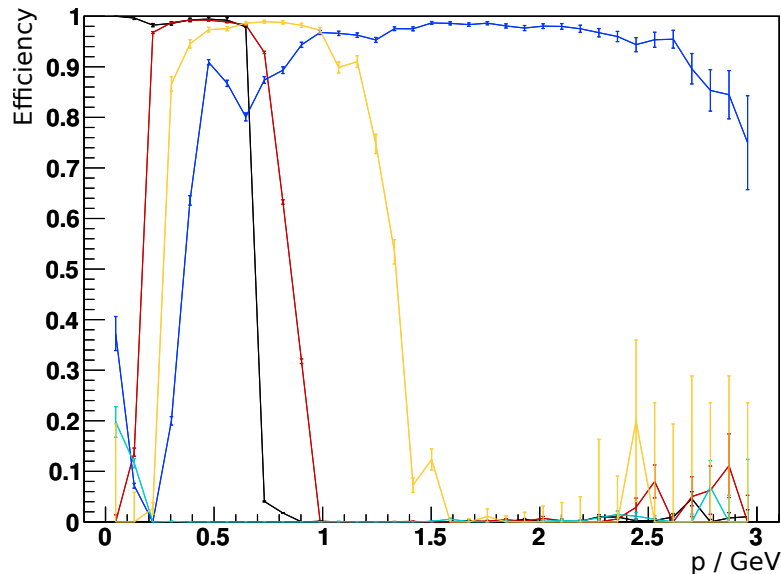
accuracy.



Figure 6.18.: Efficiency–momentum plot for a classification using individual hit information from all three tracking detectors, reconstructed using `RKTrackRep` extrapolation only, with a fixed purity of 95 %. Colours denote electrons, muons, pions, kaons, and protons.

One, however, still needs to compare their respective classification performance to judge whether the speedup of the helix or the hybrid approach can be justified. A classification that uses $dE/dx$ values reconstructed using `RKTrackRep` at each step is shown in Figure 6.18. Compared to a classification using one helix for the whole track (see Figure 6.9), the most striking difference is an improved efficiency for the selection of electrons, particularly around 500 MeV as well as a somewhat broader momentum range for the identification of protons. In the same region, the selection efficiency for pions and kaons also improves very slightly.

Figure 6.19 shows the results of a classification that uses the `RKTrackRep`/helix hybrid approach for its $dE/dx$ reconstruction, individual hits from all three tracking detectors were included. As one would expect, the resulting efficiencies can be found somewhere between the pure helix and pure `RKTrackRep` approach. However, most of the efficiency improvements for hadrons can also be achieved through the hybrid extrapolation. It thus might be worthwhile to further improve the computational performance of this mode, for example by limiting the number of `RKTrackRep` extrapolations per track. Otherwise it is possible a large number of CPU cycles is wasted on tracks that were incorrectly fitted, and where the track representation actually does not provide a better accuracy than a mere helix.

Regarding the performance of the $dE/dx$ reconstruction, it should be noted that the per-event times in Table 6.1 are not final, and are meant to show the relative complexity of the involved algorithms. With minor effort, such as compiling ROOT and GENFIT with optimisation flags and removing some debugging output, these can be reduced to 460 ms and 15 ms for the pure `RKTrackRep` and pure helix extrapolation, respectively. As can be seen, the relative run time does not change much, and the helix extrapolation is still about 30
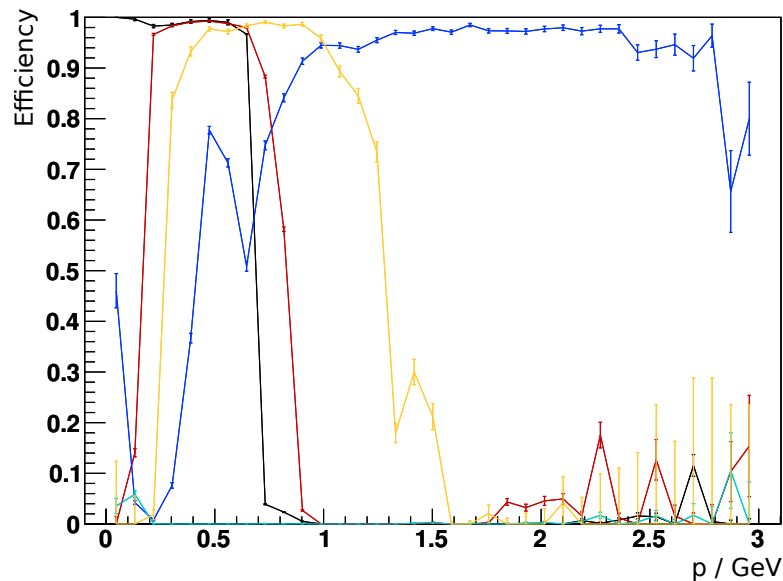
Figure 6.19.: Efficiency–momentum plot for a classification using individual hit information from all three tracking detectors, reconstructed using the hybrid approach, with a fixed purity of 95 %. Colours denote electrons, muons, pions, kaons, and protons.

times faster.

## 6.5. Calibration from $\mathrm{D}^*$ decays

As already mentioned in Chapter 5, the self-tagging decay $\mathrm{D}^*(2010)^\pm \rightarrow \pi^\pm_s \mathrm{D}^0 \left[ \rightarrow \mathrm{K}^\mp \pi^\pm \right]$ can be used for the verification and calibration of particle identification methods. For the Belle experiment, $\mathrm{D}^*$ decays were used to study the performance and systematics of the finished particle identification, which combined data from the CDC, TOF, and ACC detectors. [34] This section, on the other hand, will describe the calibration of the PID method, so data can be used instead of Monte Carlo simulations to construct probability density functions. Both methods may prove useful to either verify the data do not contain detector or electronic effects that have not been understood, or, at least for pions and kaons, avoid the generation of Monte Carlo PDFs entirely. While the other detectors may also be used to independently identify particles, possibly with more accuracy than possible here, the pion/kaon tagging from $\mathrm{D}^*$ decays can be done using only the tracking detectors. For this method, it is first necessary to identify this specific decay using the invariant masses of the daughter particles. Then, the two final state particles with the same charge must be pions, the other is a kaon. This knowledge can then be used to build PDFs for pions and kaons directly from data.

For the first step, all combinations of three tracks with two tracks of negative and one of positive charge in an event (and vice versa) are considered. One of the two tracks with same charge is taken as a slow pion, assuming it passes a very loose cut of $p < 0.5\,\mathrm{GeV}$ that only reduces the number of wrong candidates. The two remaining tracks (of opposite charges) are assumed to be the kaon and pion produced by the decay of the $\mathrm{D}^0$, where the

pion should have the same charge as the slow pion candidate. Taking the known masses of the pion and kaon from [16], one can convert the three-momentum $\vec{p}$ provided by the track fitting algorithm into a four-momentum $\mathbf{P} = (E/c, \vec{p})$ using $E^2 = |\vec{p}|^2 c^2 + m^2 c^4$. The four-momentum of the $D^0$ can then be obtained by adding the four-vectors of the daughters:

$$\mathbf{P}_{D^0} = \mathbf{P}_K + \mathbf{P}_\pi$$

If the hypothesis is correct, that those two tracks are the decay products of a $D^0$ and the masses have not been switched around through combination with a wrong slow pion candidate, the *invariant mass* of the $D^0$ should be equal to the $D^0$ mass of around 1.86 GeV.

$$M(D^0) = \sqrt{\mathbf{P}_\mu \mathbf{P}^\mu} = \sqrt{\left(\frac{E}{c}\right)^2 - |\vec{p}|^2} = mc^2$$

Of course, in reality the fitted momentum vectors are afflicted with errors and the resulting invariant mass distribution shows a peak with some width (see Figure 6.20). The correctly reconstructed $D^0$ daughters are selected by taking a fairly broad region around the peak, from 1.84 to 1.89 GeV. The rest of the distribution's shape is dominated by wrong track combinations, which are discarded.
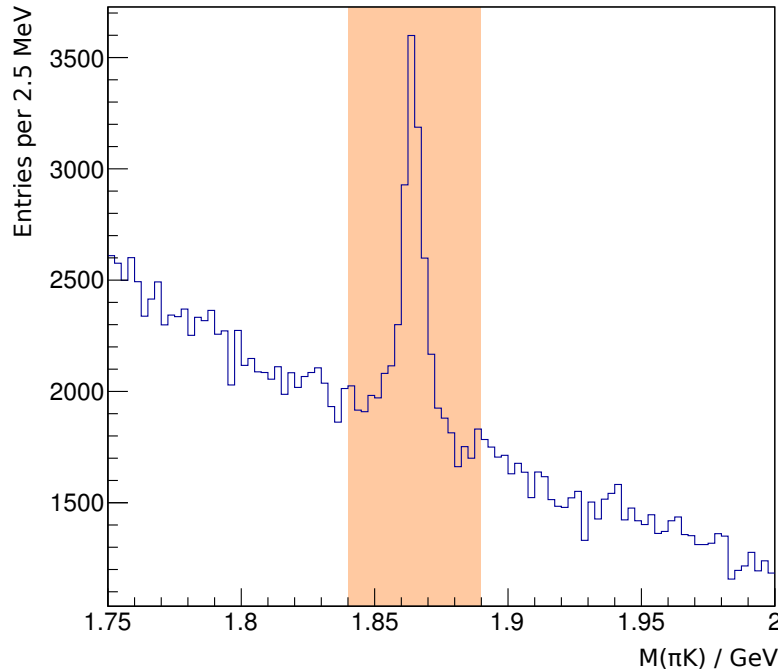


Figure 6.20.: Plot of the invariant mass of the $D^0$ daughter candidates with the cut region highlighted.

After candidates have been selected for each event, one can use the difference $M(\pi\pi K) - M(\pi K)$ of the invariant masses of the $D^*$ and of the $D^0$ daughters to look for a peak around $m_{D^*} - m_{D^0} = 145.4$ MeV. Since there are a number of misreconstructed decays in the region of the peak, just using those tagged decays to construct PDFs would result in a suboptimal

classification. To avoid this, one can fill PDFs using the data in the signal region and subtract the d$E$/d$x$–$p$ distribution of background data, normalised to the amount of background in the signal region. The signal region is defined from 144.5 to 147 MeV, and includes most of the peak, the background distribution is taken from 150 to 160 MeV. The amount of signal and background in the peak region is obtained via an extended likelihood fit of the distribution, where the background is described using an empirical parametrisation of the D$^*$(2010)$^\pm$ − D$^0$ phase space given by the following formula:

$$f(x) = a\sqrt{x - m_{\pi^\pm}} + b\sqrt{(x - m_{\pi^\pm})^3}$$

The signal shape is described by the sum of a Gaussian and a broader bifurcated Gaussian. The resulting fit can be seen in Figure 6.21 and is in good agreement with the data points. The fit separates the signal region into 603.9 signal events that should be correctly reconstructed and 72.1 misreconstructed background events, yielding a signal ration ($N_S/(N_S + N_B)$) of 89 %. From the Monte Carlo truth, the signal box actually contains 610 signal and 66 background candidates. Assuming Poisson statistics for the fitted number of signal/background events, the true and fitted values agree nicely.
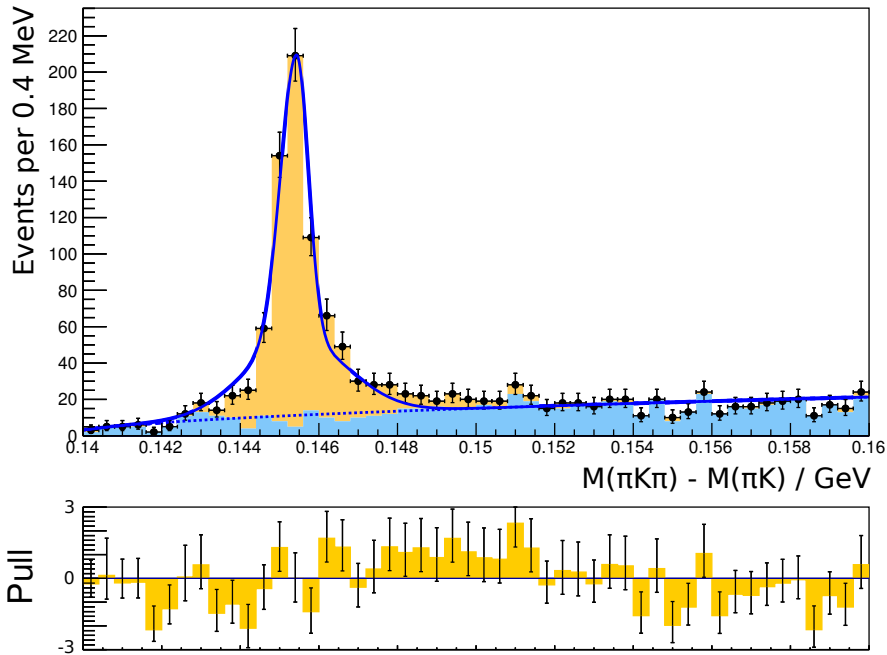


Figure 6.21.: Plot of the difference of the D$^*$ and D$^0$ candidate invariant masses. Dots with error bars show data points, with true signal and background candidates shown in the underlying histogram. The dotted line shows the fitted background shape, the solid line the fitted distribution of signal plus background.

Now that probability density functions for both pions and kaons have been constructed, we can compare the resulting classification with that for PDFs constructed directly using particle type Monte Carlo truth. As no PDFs for leptons or protons can be obtained from the D$^*$ decays—at least not directly—we will only look at the separation of pions and kaons. All

other particles in the used $B\overline{B}$ data sample are ignored and thus not counted as either signal or background in any selection. The momentum distributions for pions and kaons are taken directly from Monte Carlo.

One can compare the classifications for both types of PDFs in Figure 6.22. For both pions and kaons, the probability density functions constructed from $D^*$ decays (classification (b)) seem less accurate. This results in an efficiency that is a few per cent lower for kaons, whereas the selection of pions seems to be about 1 % less efficient below 1 GeV. For higher momenta, pions can no longer be separated with the required purity, which can be understood from the momentum distribution of the pions used to build the PDFs. Slow pions are concentrated at momenta below 200 MeV, while the $D^0$ daughter pions are distributed around 1 GeV. Above 2 GeV, the PDFs built from $D^*$ decays are practically empty. In contrast to classifications including all five particle types, both (a) and (b) show a relatively high selection efficiency for pions above 2 GeV, because there is no longer any need to distinguish them from muons.

## 6.6. Summary and outlook

In this chapter, a likelihood ratio based method for $dE/dx$ particle identification was proposed and evaluated. It was shown that adding charge signals from the silicon detectors, instead of using only drift chamber data, significantly improves the selection efficiencies of all particle types, especially below 1 GeV. Also demonstrated were improved selection efficiencies when including individual $dE/dx$ measurements in the likelihood functions, instead of using the truncated mean. Fitting the $dE/dx$–$p$ distributions is possible, and reduces the number of parameters to 7 per PDF, plus fixed resolution terms for each detector. The resulting classification, however, does not show any large changes that would indicate overtraining. Since it is easy to generate Monte Carlo samples of the size required to produce relatively smooth probability density functions, whereas the fits require a large amount of fine-tuning and cannot describe some of the features seen in the data, it is unclear whether the additional effort is worthwhile.

Regarding the computational performance of the method, it was shown that while the `RKTrackRep` extrapolation yields a better separation, mostly for electrons, it increases the CPU time by a factor 30 compared to a track extrapolation assuming a helical track. If the improved reconstruction is deemed worthwhile, using a hybrid approach that updates the helix parameters when inaccuracies are detected is recommended. This retains most of the improvements and can be tuned to limit the CPU time required for the $dE/dx$ reconstruction.

It is also possible to validate the classification or construct PDFs for pions and kaons using $D^*$ decays, where the decay products can easily be distinguished by their charge. Similar calibrations are possible for the other particles, for example through $J/\psi \rightarrow \mu^+\mu^-$ and $J/\psi \rightarrow e^+e^-$ for muons and electrons. [35][36]

Of course, looking at the specific energy loss in the tracking detectors is not the only method of particle identification, and for Belle II, the TOP, ARICH, KLM, and ECL all provide information that aids the identification of charged particles (cf. Chapter 2). The ECL and KLM are quite useful for the identification of leptons, and can identify electrons (ECL) and muons (KLM) over a large momentum region. The time-of-propagation (TOP) counter and the end cap ARICH, on the other hand, yield a more general particle identification using Cherenkov angles. However, because the particles need to exceed the speed of light in the
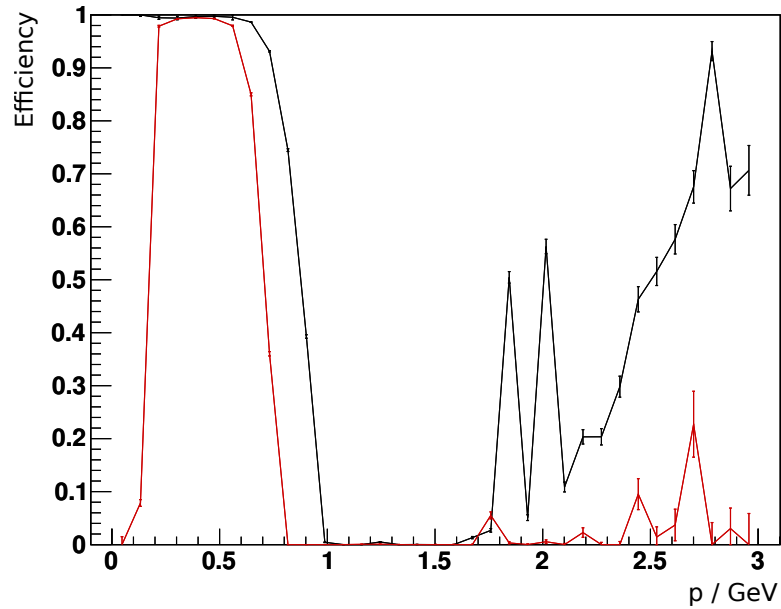
(a) using PDFs from pure Monte Carlo
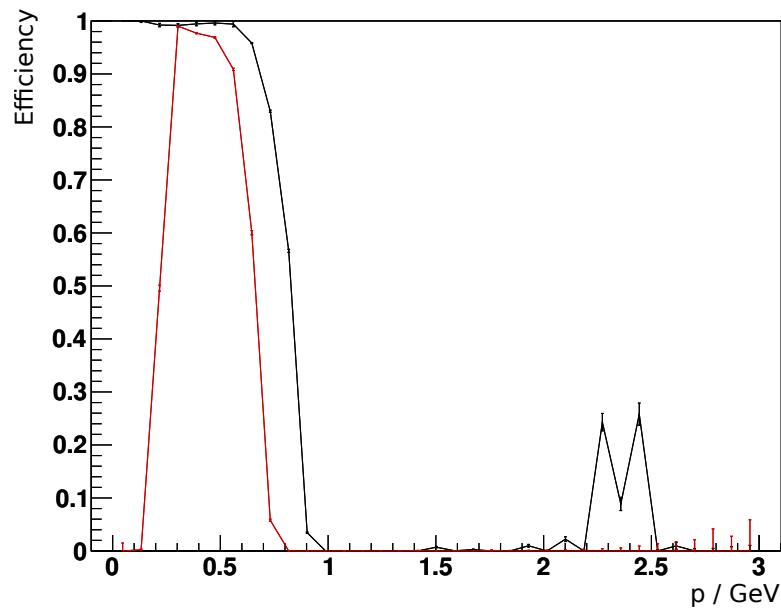


(b) using PDFs from reconstructed D* decays

Figure 6.22.: Efficiency–momentum plot for the selection of pions and kaons from a B$\overline{\text{B}}$ sample with a fixed purity of 95 %. Both classifications use likelihood functions including individual hits from all three tracking detectors.

detector medium to generate Cherenkov photons, both TOP and ARICH show a threshold behaviour. Particles with a too low velocity do not generate signals, but this in turn gives a lower limit on their mass, which helps identify them. Figure 6.23 shows the $\pi$/K separation in the ARICH from simulations, where selection efficiency drops off sharply for momenta below 1 GeV.
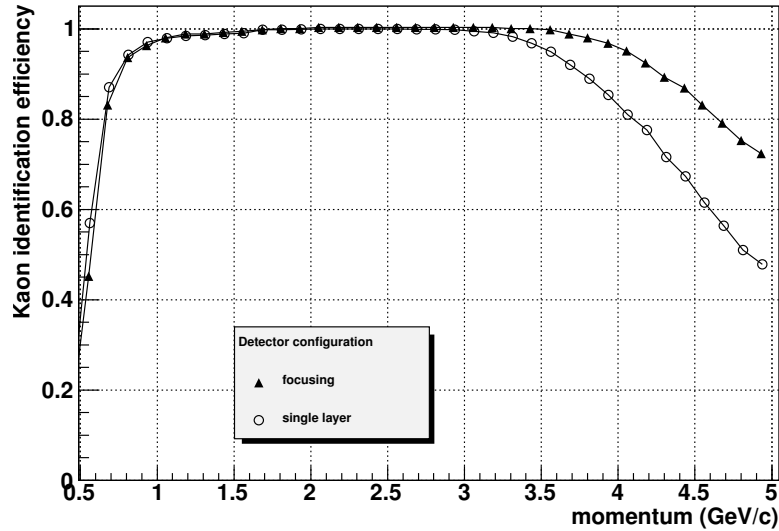


Figure 6.23.: Kaon selection efficiency for 1 % pion misidentification probability over particle momentum for the ARICH detector. Taken from [12, p. 277].

The d$E$/d$x$ method in the tracking detectors can fill this low momentum gap and provides very good separation for pions and kaons between 200 and 800 MeV. Additionally, electrons can be distinguished over a broad momentum range, as well as protons below 1.4 GeV. In the end, the information from the different PID methods will be combined to provide a nearly uninterrupted separation between particles for the entire momentum range.

Compared to the predecessor, Belle, the Belle II experiment will have two dedicated PID detectors that greatly improve upon their equivalents at Belle. This is demonstrated by a physics impact study in the technical design report [12] that compares the relative amounts of signal and background in $B^0 \rightarrow \rho^0\gamma$ decays, which can be used to measure CP violation. [37] The $\rho^0$ meson decays with almost 99 % branching ratio into two charged pions [16], whereas the more common background process $B^0 \rightarrow K^*_0\gamma$ produces a kaon and a pion. In the study, the pion/kaon separation in this decay is compared between the Belle (EACC, TOF, and d$E$/d$x$ and the Belle II configuration (ARICH and TOP). The barrel and forward end cap regions are compared independently, with the results shown in Table 6.2. It is evident that Belle II's PID significantly improves upon the results obtainable with Belle, with the figure of merit $N_S/\sqrt{N_S + N_B}$ improving by 35 % when switching from a pure Belle configuration (B1+F1) to Belle II's PID detectors (B2+F2), which corresponds to a gain in luminosity of about 80 %.

How does the d$E$/d$x$ method studied in this diploma thesis compare to the PID used in Belle? Belle note 321 [31] analyses the performance of the kaon identification, i. e. the separation of pions and kaons, using the ACC, TOF, and d$E$/d$x$ methods. Figure 6.24 shows

Table 6.2.: Number of signal and background events ($N_\mathrm{S}$, $N_\mathrm{B}$), figure of merit (FOM) and the lower limit for $\Delta E$ for a $B^0 \to \rho^0\gamma$ at 7.5 ab$^{-1}$. Belle (1) and Belle II (2) configurations of barrel (B) and forward (F) PID are compared, i. e. B1 includes Belle's ACC, TOF, and $\mathrm{d}E/\mathrm{d}x$ PID, F1 ACC and $\mathrm{d}E/\mathrm{d}x$; B2 is Belle II's TOP counter, F2 the ARICH detector. Taken from [12, p. 279].

| Barrel | Forward | $N_\mathrm{S}$ | $N_\mathrm{B}$ | FOM | $\Delta E_{\mathrm{min}}$ [ GeV] |
|--------|---------|------|------|------|--------|
| B1 | F1 | 987 | 5242 | 12.5 | $-0.25$ |
| B1 | F2 | 1032 | 5026 | 13.3 | $-0.25$ |
| B2 | F1 | 982 | 2865 | 15.8 | $-0.30$ |
| B2 | F2 | 1027 | 2651 | 16.9 | $-0.30$ |

the results of a fixed cut on the probability (assuming a 1:1 ratio between pions and kaons):

$$\mathrm{Prob}(K:\pi) = \frac{\mathscr{L}_K}{\mathscr{L}_\pi + \mathscr{L}_K}$$

The red dots show the selection efficiency for kaons, the blue dots the pion fake rate, defined as

$$\pi \text{ fake rate} = \frac{\pi \text{ selected as K}}{\text{total number of } \pi} = 1 - \text{purity for } \pi.$$

The plots, divided into barrel and forward end cap, show the complementarity of the different detectors: the ACC provides reasonable separation for relatively high momenta, $\mathrm{d}E/\mathrm{d}x$ works well below 1 GeV and for higher momenta, while the TOF provides decent separation over a broad momentum range in the barrel. This can be compared to Figure 6.25, which shows kaon efficiency and pion fake rate for the same cut using the proposed $\mathrm{d}E/\mathrm{d}x$ method for Belle II, using a sample of about 200,000 pions and kaons in equal proportion, equally distributed in a momentum range of 50 MeV to 3 GeV. The main difference is that the kaon efficiency remains above 40 %, even in the region where Belle's $\mathrm{d}E/\mathrm{d}x$ PID drops below 10 %. Since values of 0.6 or higher are more often reached, the pion fake rate also rises, though more slowly, and assumes a more regular shape. For lower and higher momenta, the efficiency approximates that achieved by Belle within the statistical uncertainties. In the Belle note, no values are provided below 600 MeV, but are expected to be high.

Figure 6.24.: Plot of the kaon selection efficiency (red) and pion fake rate (blue) for a fixed cut Prob(K : π) ≥ 0.6 over the track momentum for different PID detectors of the Belle experiment. The left hand side shows the performance of the subdetectors in the barrel region, the right hand side that in the forward end cap. Taken from [31].



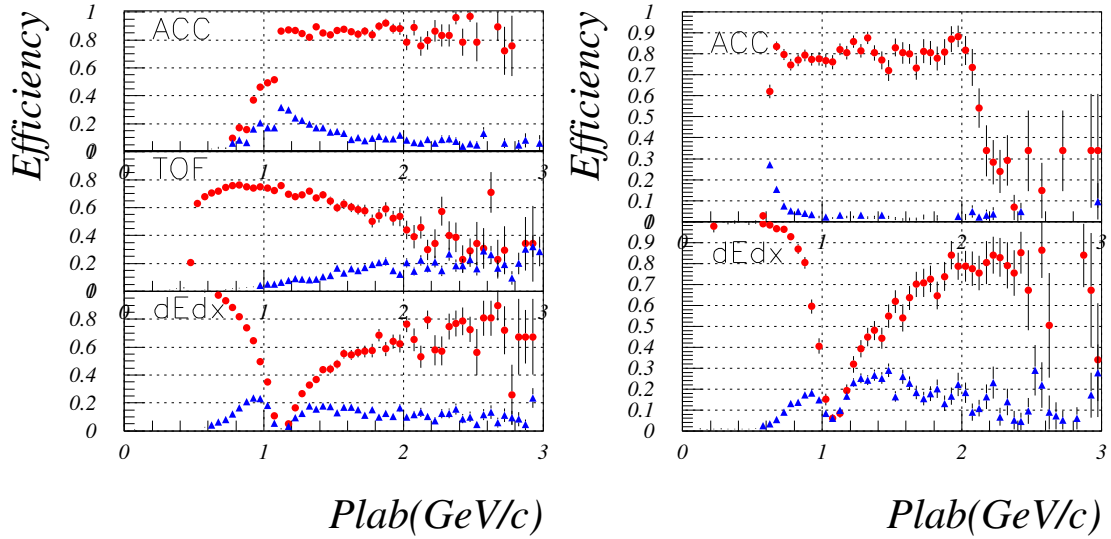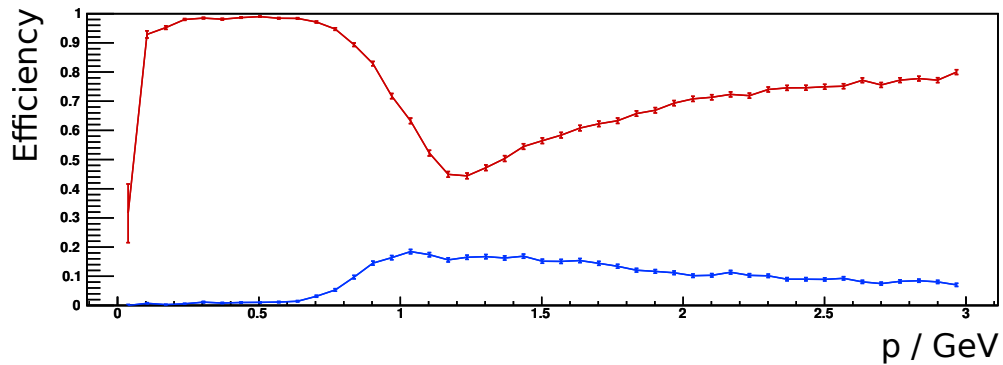Figure 6.25.: Plot of the kaon selection efficiency (red) and pion fake rate (blue) for a fixed cut Prob(K : π) ≥ 0.6 over the track momentum for the PID method studied in this thesis. The likelihood functions include factors for individual hits in all three tracking detectors.

# 7. Conclusions

In Chapter 5 clusters in the pixel detector were analysed using a neural network, using only information from the clusters themselves. It was shown that pixel clusters created by slow hadrons can be efficiently selected from clusters produced by background processes. Primarily, this can be attributed to the much larger energy loss of pions with a total momentum below 100 MeV, whereas the electrons created by the background processes are practically minimum-ionising particles. Consequently, the separation should be fairly robust against changes of the background composition, since the distribution of the pixel charges will be affected only marginally. To verify the robustness of the neural network, tests such as removing additional variables, different charge cutoffs and artificial smearing of the collected pixel charges were performed.

The neural network complements the tracking-based data reduction scheme and is able to rescue 95 % of slow pions while only selecting 10 % of all background clusters. The network and preprocessing are simple enough to transfer them to an FPGA and could thus be used to rescue pixel data for slow pions that would be lost when using the tracking-based data reduction exclusively.

In Chapter 6 a particle identification method for the tracking detectors of Belle II was developed, and the effect of different variations analysed. It was shown that both the use of individual $dE/dx$ measurements in the likelihood functions and the addition of information from the silicon detectors (PXD and SVD) can greatly improve the separation between particles. A likelihood function containing factors for every $dE/dx$ value makes use of the full information contained in a set of measurements, whereas the truncated mean reduces this to a single value. For the added silicon data, this is particularly visible for low-momentum particles, where the different energy loss for electrons in silicon resolves some overlaps present in the CDC measurements, and for tracks with only a small number of associated drift chamber hits. Together, these changes raise the selection efficiencies for pions and kaons to well above 98 % for momenta below 800 MeV, for a fixed purity of 95 %. Electrons can be selected with high efficiency ($> 95\%$) above $p = 1.1$ GeV, and with about 50 % efficiency down to 500 MeV.

In direct comparisons of the pion/kaon separation with the separation achieved by the Belle kaon ID, these improvements are less visible, since particles other than pions and kaons are ignored. Still, in the momentum region around 1.2 GeV where both particles have similar energy loss, kaon selection efficiencies of about 45 % are reached, an increase of 40 percentage points compared to Belle. For the electron identification, improvements are expected mainly for momenta below 1 GeV.

Overall, the particle identification of Belle II is expected to outperform that of the Belle experiment, and the improved method for $dE/dx$ PID outlined in this thesis will contribute to that. The method was implemented as a Belle II software framework (`basf2`) module.

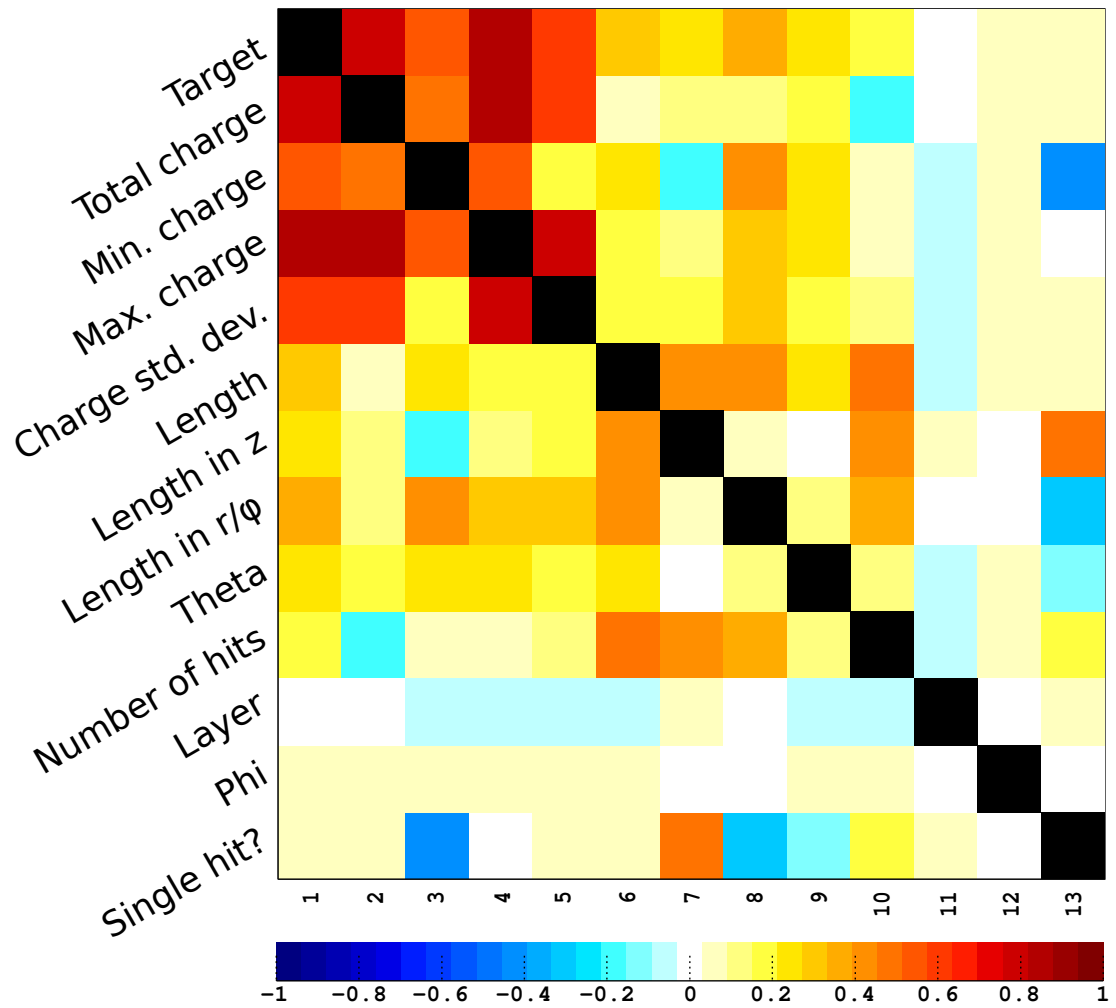# A. Appendix: Technicalities of the Datareduction Network



Figure A.1.: Plot of the linear correlations between different input variables and the target (signal or background cluster).

Table A.1.: List of all input variables with the preprocessing flags used by NeuroBayes. Here 14 is used for variables that can benefit from a spline-fit of the purity, 34 is similar but adds a $\delta$-function for undefined values. 19 is used for ordered classes (integers) that are not to be spline-fitted. See [38, p. 29] for a detailed explanation.

| Variable | Preprocessing flag | Added significance / $\sigma$ | Only this variable / $\sigma$ | Loss when removed / $\sigma$ |
|---|---|---|---|---|
| Total charge | 14 | 85.53 | 85.53 | 21.80 |
| Minimum charge | 14 | 2.92 | 55.29 | 3.41 |
| Maximum charge | 14 | 13.11 | 89.74 | 9.15 |
| Charge standard deviation | 34 | 0.61 | 67.31 | 0.61 |
| Cluster length | 14 | 0.69 | 29.17 | 0.72 |
| Cluster length in $z$ | 14 | 3.90 | 22.85 | 2.56 |
| Cluster length in $r/\phi$ | 14 | 16.35 | 39.46 | 10.82 |
| $\theta$ position | 14 | 6.45 | 28.38 | 5.99 |
| Number of hits | 19 | 35.00 | 19.78 | 12.98 |
| Layer no. | 19 | 2.63 | 0.26 | 2.63 |
| $\phi$ position | 14 | 3.12 | 9.49 | 3.06 |
| Single hit? ($N_{\text{hits}} = 1$) | 19 | 1.37 | 3.41 | 1.29 |

# B. Appendix: Tabulated Results of $\mathrm{d}E/\mathrm{d}x$ Particle ID

Table B.1.: Tabulated selection efficiencies (in %) for different particles and purity requirements. The used classification includes information from individual d$E$/d$x$ measurements, from all three tracking detectors. Because of low statistics, the bin width was doubled above 2 GeV.

| $p$ / GeV | π 68% | π 95% | K 68% | K 95% | e 68% | e 95% | μ 68% | μ 95% | p 68% | p 95% |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.1 − 0.2 | 100.0±0.3 | 99.5±0.3 | 97.5±0.3 | 74.4±0.8 | 83.1±0.6 | 2.2±0.2 | 39.3±1.2 | 2.7±0.4 | 73.6±5.1 | 11.1±3.8 |
| 0.2 − 0.3 | 100.0±0.3 | 97.6±0.3 | 99.3±0.1 | 97.9±0.1 | 27.8±0.7 | 16.0±0.6 | 0.0±0.1 | 0.0±0.1 | 91.6±1.5 | 40.8±2.6 |
| 0.3 − 0.4 | 100.0±0.3 | 98.7±0.3 | 99.5±0.1 | 98.9±0.1 | 58.1±1.0 | 29.7±0.9 | 0.0±0.0 | 0.0±0.0 | 98.7±0.4 | 91.7±1.0 |
| 0.4 − 0.5 | 100.0±0.4 | 99.0±0.4 | 99.7±0.0 | 99.3±0.1 | 85.1±0.6 | 42.1±0.8 | 0.0±0.0 | 0.0±0.0 | 99.5±0.2 | 97.8±0.4 |
| 0.5 − 0.6 | 100.0±0.4 | 99.6±0.4 | 99.6±0.0 | 98.9±0.1 | 89.7±0.5 | 46.4±0.8 | 0.0±0.0 | 0.0±0.0 | 99.6±0.2 | 97.0±0.4 |
| 0.6 − 0.7 | 99.9±0.5 | 90.6±0.6 | 99.5±0.1 | 97.3±0.1 | 90.1±0.5 | 43.2±0.9 | 0.0±0.0 | 0.0±0.0 | 99.7±0.1 | 98.5±0.3 |
| 0.7 − 0.8 | 99.9±0.0 | 3.5±0.1 | 98.0±0.1 | 76.4±0.3 | 94.2±0.4 | 48.4±0.9 | 0.1±0.1 | 0.0±0.0 | 99.6±0.2 | 98.7±0.3 |
| 0.8 − 0.9 | 99.6±0.0 | 1.3±0.1 | 89.2±0.3 | 46.0±0.4 | 96.5±0.4 | 85.7±0.7 | 0.1±0.1 | 0.1±0.1 | 99.8±0.1 | 98.9±0.3 |
| 0.9 − 1.0 | 96.9±0.1 | 0.0±0.0 | 63.5±0.5 | 0.4±0.1 | 98.7±0.2 | 93.1±0.5 | 0.0±0.1 | 0.0±0.1 | 99.5±0.2 | 97.8±0.4 |
| 1.0 − 1.1 | 83.7±0.3 | 0.0±0.0 | 25.7±0.5 | 0.0±0.0 | 98.6±0.3 | 94.0±0.5 | 0.0±0.0 | 0.0±0.0 | 99.6±0.2 | 95.4±0.7 |
| 1.1 − 1.2 | 62.0±0.4 | 0.0±0.0 | 8.9±0.3 | 0.0±0.0 | 99.0±0.2 | 93.6±0.5 | 0.0±0.0 | 0.0±0.0 | 97.9±0.5 | 78.1±1.5 |
| 1.2 − 1.3 | 13.1±0.3 | 0.0±0.0 | 11.4±0.4 | 0.0±0.0 | 99.0±0.2 | 94.2±0.5 | 0.8±0.2 | 0.1±0.1 | 94.2±1.0 | 74.6±1.8 |
| 1.3 − 1.4 | 14.0±0.4 | 0.0±0.0 | 19.6±0.6 | 0.0±0.0 | 99.4±0.2 | 96.8±0.4 | 0.9±0.2 | 0.0±0.0 | 79.7±1.8 | 48.5±2.3 |
| 1.4 − 1.5 | 13.1±0.4 | 0.0±0.0 | 29.3±0.8 | 0.0±0.0 | 99.1±0.2 | 97.0±0.4 | 1.2±0.2 | 0.0±0.0 | 56.6±2.6 | 23.3±2.3 |
| 1.5 − 1.6 | 13.5±0.5 | 0.0±0.0 | 36.6±1.0 | 0.0±0.0 | 99.7±0.1 | 97.4±0.4 | 2.1±0.3 | 0.3±0.1 | 21.1±2.5 | 15.6±2.3 |
| 1.6 − 1.7 | 6.4±0.4 | 0.1±0.1 | 37.1±1.1 | 0.1±0.1 | 99.7±0.1 | 98.5±0.3 | 5.7±0.5 | 0.1±0.1 | 18.7±2.8 | 0.0±0.5 |
| 1.7 − 1.8 | 7.0±0.5 | 0.0±0.0 | 48.6±1.3 | 0.3±0.2 | 99.7±0.1 | 98.3±0.3 | 6.8±0.6 | 0.0±0.1 | 8.2±2.5 | 2.5±1.6 |
| 1.8 − 1.9 | 11.4±0.6 | 0.1±0.1 | 47.3±1.5 | 0.7±0.3 | 99.8±0.2 | 97.9±0.4 | 14.0±0.9 | 0.0±0.1 | 0.0±1.0 | 0.0±1.0 |
| 1.9 − 2.0 | 15.9±0.9 | 0.2±0.1 | 47.4±1.8 | 4.2±0.7 | 99.7±0.2 | 97.5±0.5 | 22.6±1.2 | 0.3±0.2 | 1.9±2.5 | 1.9±2.5 |
| 2.0 − 2.2 | 10.7±0.6 | 0.0±0.1 | 47.3±1.6 | 0.0±0.1 | 99.8±0.1 | 98.0±0.4 | 14.9±0.8 | 0.2±0.1 | 0.0±1.8 | 0.0±1.8 |
| 2.2 − 2.4 | 51.5±1.3 | 0.3±0.1 | 45.7±2.2 | 0.6±0.4 | 99.6±0.2 | 97.3±0.5 | 30.9±1.5 | 0.2±0.2 | 0.0±5.5 | 0.0±5.5 |
| 2.4 − 2.6 | 74.4±1.5 | 0.2±0.2 | 41.6±3.5 | 2.5±1.2 | 99.5±0.4 | 94.4±1.0 | 29.7±2.1 | 0.6±0.4 | 0.0±12.4 | 0.0±12.4 |
| 2.6 − 2.8 | 97.6±0.6 | 0.3±0.3 | 35.9±4.9 | 4.3±2.3 | 95.2±1.3 | 90.0±1.8 | 0.7±0.9 | 0.7±0.9 | 0.0±16.3 | 0.0±16.3 |
| 2.8 − 3.0 | 99.6±0.5 | 0.4±0.5 | 14.5±4.8 | 3.6±2.9 | 84.8±3.6 | 80.8±3.9 | 0.0±3.2 | 0.0±3.2 | 0.0±23.6 | 0.0±23.6 |

Required purity for selection of this particle

# Bibliography

[1] A. Sakharov, "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe," *JETP Letters* **5** no. 24, (1967). `http://www-itp.particle.uni-karlsruhe.de/~hofmann/lehre/GR/semSS2007LIT/Sakharov.pdf`.

[2] J. Christenson, J. Cronin, V. Fitch, and R. Turlay, "Evidence for the $2\pi$ decay of the $K_2^0$ meson," *Phys. Rev. Lett.* **13** no. 4, (1964) 138.

[3] **Belle Collaboration**, K. Abe *et al.*, "Observation of large CP violation in the neutral B meson system," *Phys. Rev. Lett.* **87** no. 9, (2001) 91802.

[4] **BaBar Collaboration**, B. Aubert *et al.*, "Observation of CP violation in the $B^0$ meson system," *Phys. Rev. Lett.* **87** no. 9, (2001) 91801.

[5] The Royal Swedish Academy of Sciences, "Scientific background on the Nobel Prize in Physics 2008." `http://www.nobelprize.org/nobel_prizes/physics/laureates/2008/advanced-physicsprize2008.pdf`.

[6] **Belle Collaboration**, I. Adachi *et al.*, "Precise measurement of the CP violation parameter $\sin 2\phi_1$ in $B^0 \rightarrow (c\bar{c})K^0$ decays," *Phys. Rev. Lett.* **108** no. 17, (Apr, 2012) 171802.

[7] **BaBar Collaboration**, "Evidence for an excess of B → D(*) Tau Nu decays," `arXiv:1205.5442 [hep-ex]`.

[8] "Integrated luminosity of B factories." `http://belle.kek.jp/bdocs/lumi_belle.png`.

[9] "Using crab cavities, KEKB breaks luminosity world record." `http://legacy.kek.jp/intra-e/press/2009/KEKBluminosity2.html`.

[10] Y. Ohnishi, "Projected luminosity (revised after the national disaster)." `http://kds.kek.jp/conferenceDisplay.py?confId=8102`. Belle internal.

[11] A. Moll, "The software framework of the Belle II experiment," in *Journal of Physics: Conference Series*, vol. 331, p. 032024, IOP Publishing. 2011. `http://stacks.iop.org/1742-6596/331/i=3/a=032024`.

[12] **Belle II Collaboration**, T. Abe *et al.*, "Belle II Technical Design Report," tech. rep., 2010. `arXiv:1011.0352 [physics.ins-det]`.

[13] M. Heck. private communication.

[14] C. Grupen, A. Böhrer, and L. Smolík, *Particle Detectors*. Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge University Press, 1996.

*Bibliography*

[15] L. Piilonen, "B-KLM Summary Talk." 11th B2GM, Mar., 2012. `http://kds.kek.jp/contributionDisplay.py?sessionId=29&contribId=17&confId=8895`. Belle II internal.

[16] **Particle Data Group**, K. Nakamura *et al.*, "Review of particle physics," *J. Physics G* **37** (2010) 075021.

[17] Fermi National Accelerator Laboratory, "Negative No. 80-0271CN," 1980.

[18] L. Landau, "On the energy loss of fast particles by ionization," *J.Phys.(USSR)* **8** (1944) 201–205.

[19] G. Cowan, *Statistical Data Analysis.* Oxford Science Publications. Clarendon Press, 1998.

[20] P. Vavilov, "Ionization losses of high-energy heavy particles," *Soviet Phys. JETP* **5** (1957) .

[21] H. Bichsel, "A method to improve tracking and particle identification in TPCs and silicon detectors," *Nucl. Instrum. Meth.* **A562** no. 1, (2006) 154–197.

[22] R. Duda, P. Hart, and D. Stork, *Pattern classification.* Wiley, 2001.

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

[24] M. Feindt, "A Neural Bayesian Estimator for Conditional Probability Densities," *ArXiv Physics e-prints* (Feb., 2004) , `arXiv:physics/0402093`.

[25] C. Bernardini, G. F. Corazza, G. Di Giugno, G. Ghigo, J. Haissinski, P. Marin, R. Querzoli, and B. Touschek, "Lifetime and Beam Size in a Storage Ring," *Phys. Rev. Lett.* **10** (May, 1963) 407–409.

[26] A. Moll, "Backgrounds at Belle II." F2F Tracking Meeting, Mar., 2012. `http://indico.mppmu.mpg.de/indico/getFile.py/access?contribId=10&sessionId=0&resId=0&materialId=slides&confId=1659`.

[27] H. Nakayama, "Background estimation status." Beast II EVO meeting, Jan., 2012. `http://kds.kek.jp/getFile.py/access?contribId=5&resId=0&materialId=slides&confId=8960`. Belle II internal.

[28] T. Ullrich and Z. Xu, "Treatment of Errors in Efficiency Calculations," `arXiv:physics/0701199 [physics.data-an]`.

[29] D. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth.* **A462** (2001) 152–155.

[30] C. Höppner, S. Neubert, B. Ketzer, and S. Paul, "A Novel Generic Framework for Track Fitting in Complex Detector Systems," *Nucl. Instrum. Meth.* **A620** (2010) 518–525, `arXiv:0911.1008 [hep-ex]`.

[31] H. Hamasaki, T. Iijima, *et al.*, "Kaon Identification in Belle." Belle Note 321, July, 2000. Belle internal.

[32] W. Verkerke and D. Kirkby, "The RooFit Toolkit for Data Modeling." Proceedings of PHYSTAT 2005. `http://roofit.sourceforge.net/docs/roofit_phystat05.pdf`.

[33] ROOT authors, "Online documentation of `TMath::Landau`," Nov., 2011. `http://root.cern.ch/root/html/TMath.html#TMath:Landau`.

[34] S. Nishida, "Study of Kaon and Pion Identification Using Inclusive $D^*$ sample." Belle Note 779, Jan., 2005. Belle internal.

[35] A. Abashian, K. Abe, P. Behera, F. Handa, T. Iijima, Y. Inoue, H. Miyake, T. Nagamine, E. Nakano, S. Narita, *et al.*, "Muon identification in the Belle experiment at KEKB," *Nucl. Instrum. Meth.* **A491** no. 1, (2002) 69–82.

[36] K. Hanagaki, H. Kakuno, H. Ikeda, T. Iijima, and T. Tsukamoto, "Electron identification in Belle," *Nucl. Instrum. Meth.* **A485** no. 3, (2002) 490–503.

[37] **Belle Collaboration**, Y. Ushiroda, Sumisawa, *et al.*, "Time-Dependent $CP$-Violating Asymmetry in $B^0 \rightarrow \rho^0\gamma$ Decays," *Phys. Rev. Lett.* **100** (Jan, 2008) 021602.

[38] <phi-t> Physics information technologies, "NeuroBayes user guide." `http://neurobayes.phi-t.de/nb_doc/NeuroBayes-HowTo.pdf`.