Studies on the neural z-Vertex-Trigger for the Belle II Particle Detector

Fernando Abudinén

München 2014

Studien zu dem neuronalen z-Vertex-Trigger für den Belle II Teilchendetektor

Fernando Abudinén

Masterarbeit
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Fernando Abudinén
aus Barranquilla, Kolumbien

München, den 28.01.2014

Gutachter: Prof. Dr. Jochen Schieck

Contents

Zusammenfassung

Der Elektron-Positron Teilchenbeschleuniger KEKB und der dazugehörige Belle Detektor in Tsukuba, Japan werden zur Zeit verbessert. Bisher hielt KEKB mit einer Luminosität von $2.11 \cdot 10^{34}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ den Weltrekord. Der neue Beschleuniger, SuperKEKB, wird eine 40 mal höhere Luminosität als bis jetzt erreichen. In dem neuen Belle II Detektor wird ein hochauflösender Pixel-Vertex-Detektor eingebaut. Damit werden eine höhere Datenproduktion und eine bessere Vertexaufösung erreicht, was das Unitaritäsdreieck weiter einschränken wird. Infolgedessen wird die Suche nach neuer Physik präzisiert.
Auf Grund der neuen Strahlparameter wird sich die Belastung durch ungewollte Streungsprozesse, z.B. Touschek-Streung, um ca. einen Faktor 30 erhöhen. Die gestreuten Teilchen produzieren Untergrundsignale im Detektor. Da die enorme Datenproduktion des Detektors jegliche Datenübertragungskapazität übersteigt, muss der Untergrund im Voraus gefiltert werden.
Von Belle ist bekannt, dass diese Untergrundereignisse nicht vom Kollisionspunkt ausgehen. Eine sehr gute Möglichkeit, den Untergrund zu reduzieren, wäre die Spuren geladener Teilchen zu identifiziern und Ereignisse zu verwerfen die nicht vom Kollisionspunkt stammen. Für eine ausreichende Untergrundunterdrückung ist eine Auflösung von ungefähr 2 cm erforderlich. Da die Entscheidung in Echtzeit erfolgen muss, ist nicht ausreichend Zeit für herkömliche Spurrekonstruktion. Die vielverspechendste Methode ist bisher ein neuronaler z-Vertex Trigger, bestehend aus mehreren Multi Layer Perzeptronen (MLP).
Die bisher durchgeführten Experimente basierten auf idealisierten Simulationsereignissen innerhalb sehr begrenzter Detektorbereiche mit einzelnen Teilchenspuren ohne Untergrund. In dieser Arbeit wurden die MLP für bestimmte Bereiche des Detektors optimiert und das Netz mit realistischeren Daten inklusive Untergrund trainiert und getestet. Die Anzahl der MLP, die notwendig sind, um den ganzen Detektor abzudecken, konnte anhand der Ergebnisse auf $\sim 2 \cdot 10^6$ abgeschätzt werden. Die niedrigste und die höchste Auflösung, die unter Anwendung von simulierten Ereignissen mit Untergrund erreicht wurden, sind

$$\sigma_{\mathrm{max}} = 1.70 \pm 0.02 \text{ cm}$$
$$\sigma_{\mathrm{min}} = 0.85 \pm 0.04 \text{ cm}.$$

Somit wurde gezeigt, dass auch in der Anwesenheit von Untergrund der neuronale z-Vertex Trigger in der Lage ist, die gefordete Auflösung von weniger als 2 cm zu erreichen.

Abstract

The electron-positron particle collider KEKB together with its Belle detector in Tsukuba, Japan are currently being upgraded. So far, KEKB held the world record luminosity of $2.11 \cdot 10^{34}$ cm^{-2}s^{-1}. The new collider, SuperKEKB, will reach a 40 times larger luminosity than its predecessor. Additionally, a new high-resolution Pixel-Vertex-Detector will be built into the new Belle II detector. In consequence, a higher data production and a better vertex resolution will be achieved. This will constrain the unitarity triangle and the search of new physics.
Because of the new beam parameters, the load due to undesired scattering process, e.g. Touschek scattering, will be larger by a factor 30. The scattered particles produce background signals inside the detector. Since the enormous data production of the detector exceeds any data transference capacity, the background has to be filtered out in advance. From the Belle experiment, it is known that these background events do not come from the collision point. Therefore, a good possibility to reduce the background could be to identify the tracks of charged particles and to reject events which do not stem from the collision point. For a sufficient background suppression, a resolution of about 2 cm is needed.Since the decision has to me made in real time, there is not enough time for usual track reconstruction. The most promising method up to now has been a neural z-vertex trigger composed by several multi layer perceptrons (MLP).
Previous experiments were based on idealized simulated events within narrow detector regions using single particle tracks, but without background. In this work, the MLP has been optimized for specific detector regions. The networks were trained and tested with more realistic data samples including background. On the basis of the achieved results, the total amount of required MLPs to cover the whole detector could be approached to $\sim 2 \cdot 10^6$. The lowest and the highest resolution reached, using simulated events under the presence of background, is

$$\sigma_{\max} = 1.70 \pm 0.02 \text{ cm}$$
$$\sigma_{\min} = 0.85 \pm 0.04 \text{ cm}.$$

Thus it could be shown that, also under the presence of background, the neural z-vertex trigger is capable to reach the required resolution of less than 2 cm.

Chapter 1

Introduction

High luminosity colliders and high resolution detectors are required to provide the necessary amount of data, with sufficient accuracy for the search of new physics. B factories are a special kind of collider which produce large amounts of so-called B mesons. B mesons are particles that allow to study a particular violation of symmetry, the Charge-Parity violation, which plays a crucial role in understanding why there is much more matter than antimatter in our universe.
The KEKB collider in the locality of Tsukuba, Japan belongs to this kind of collider. It operated at the highest Luminosity ever achieved of $2.11 \cdot 10^{34}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ before being temporarily suspended in 2010. At the present time, the KEKB collider and its detector Belle are under an upgrade process. The new SuperKEKB factory will achieve a luminosity of $80 \cdot 10^{34}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. In addition, the resolution of the new Belle II detector will be improved through the insertion of a new pixel vertex detector in its innermost part.
Nevertheless, not all the signals that will be registered by the detector belong to interesting physics events: particle detectors have also to deal with background signals caused by scattered particles. In the case of SuperKEKB, the new beam parameters which, on the one hand are required to reach the design luminosity, lead on the other hand to a higher load on scattered particles.
Since the data production rate of Belle II will exceed all possible transference capacities at the current level of technological advance, it would not be possible to store all the detected events. Therefore, the background noise needs to be suppressed as much as possible. This is the task of the so-called detector trigger.
Analyses on the data taken by Belle provide a very good hint: background events do not come from the collision point of the collider beams. Thus, identifying and rejecting those events which do not stem from the collision point will efficiently reduce the amount of background. Nevertheless, for this purpose a resolution of about 2 cm has to be reached. Since the decision has to be made in real time while data taking, there is not enough time to make the usual track reconstruction applied for analysis.
From all the methods considered, the best results so far were achieved by the so-called neural z-vertex trigger, proposed by S. Skambraks [1], which is composed of a special kind

of artificial neural networks. His studies were based on simulation events for idealized single tracks within narrow detector regions, but without including background. They proved that the neural z-vertex trigger can indeed reach the required resolution.
Since the neural z-vertex trigger will need to deal with real detector data it was essential to test the performance of the neural networks under the presence of background, which was the main purpose of the present study. In order to do this, several studies had to be carried out to estimate the attributes of the neural network. To make the results more significant, also more realistic input information was selected to train and test the neural network.
This thesis is composed of four chapters. In chapter 2, a detailed physics motivation for the Belle II experiment is given. The following chapter 3 contains a brief description of the SuperKEKB collider and the Belle II detector. In this chapter also the different background sources, the general tasks of the trigger and the Belle II simulation software will be treated. The chapter number 4 is dedicated to a description of the neural z-vertex trigger, the used neural networks, the training and testing process and the different network parameters. The last chapter 5 presents the procedure description and the results of all studies and experiments done within the scope of this work.

Chapter 2

Physics Motivation

In particle physics, the Standard Model (SM) describes the fundamental building blocks of matter and their properties derived from the interactions among elementary particles. It is one of the best verified physical theories within the explored energy ranges at the actual level of experimental precision.
The main contribution of the experimental tests executed inside the KEKB collider with the Belle detector has been the research concerning the violation of Charge-Parity (CP) symmetry in the quark sector. Together with its companion collaboration BaBar, operating at the PEP-II B-meson factory at SLAC, the Belle collaboration collected numerous results, which could confirm with good precision the predictions of the SM theory[2]. This effort conduced to the award of the Nobel Prize in physics to M. Kobayashi and T. Maskawa for their CP violation theory [3, 4].
Notwithstanding the considerable success of the Standard Model, many fundamental questions still remain open. At cosmological scale for example, the observed matter-antimatter asymmetry in the universe can not be explained by the CP violation theory in the SM[5]. Thus, a search of new physics (NP) is required.
Future approaches searching for NP can be classified into two complementary frontiers based on experiments inside particle accelerators. First, the energy frontier represented principally by the ATLAS and the CMS experiments at the Large Hadron Collider (LHC) at CERN. Second, the rare and precision frontier to which the Belle II experiment at SuperKEKB belongs [6]. The LHCb experiment at the LHC should be also mentioned here as an example.
In what follows, the most relevant theoretical principles related to CP Violation in the SM will be treated. In addition, some motivating aspects will be examined where Belle II can extend the current experimental reach of its predecessor Belle towards the search of NP.

2.1 The Violation *CP* Symmetry

2.1.1 Parity Transformation

The effect of the parity transformation, P, is defined as an inversion of the spatial coordinates at the origin. It can be interpreted also as a reversion of the three spatial axes. A graphic equivalent is shown in Fig. 2.1.

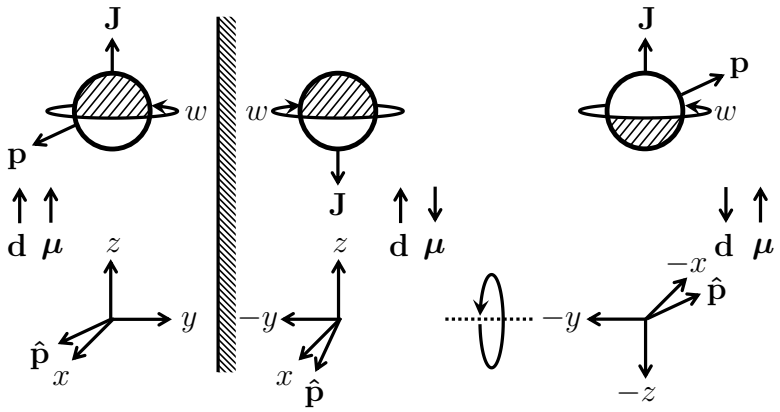


Figure 2.1: A mirror reflection at an arbitrary plane followed by a rotation of 180° around the axis orthogonal to the plane corresponds to the parity transformation P. The graphic illustrates the effect of P on electric and magnetic dipole moments $\mathbf{d}$ and $\boldsymbol{\mu}$ as well as on linear and angular momenta $\mathbf{p}$ and $\mathbf{J}$. The vector $\hat{\mathbf{p}}$ is the normalized vector $\mathbf{p}$.

In quantum mechanics this concept is introduced through the parity operator P, which operates on state vectors. A parity transformation for a wave function $\psi(\mathbf{x}) = \langle \mathbf{x}|\psi\rangle$ is given by

$$\begin{aligned} P\psi(\mathbf{r}) &= \langle \mathbf{x}|P|\psi\rangle \\ &= \langle -\mathbf{x}|\psi\rangle = \psi(-\mathbf{x}) \qquad (2.1) \\ \Rightarrow P^2|\psi\rangle &= |\psi\rangle \qquad (2.2) \end{aligned}$$

The wave function $|\psi\rangle$ is an eigenstate of P in the case that

$$P|\psi\rangle = \eta_P|\psi\rangle. \qquad (2.3)$$

Because of $\eta_P^2 = 1$, only the real numbers $\eta_P = \pm 1$ follow as possible eigenvalues since P is an hermitian observable. Physical Observables can be classified into four types V, A, S and P, according to their transformation properties with respect to P (S. Table 2.1). If the laws of nature would be symmetric under P transformations, the results of a mirror experiment in a reflected reference frame would yield exactly the same results of the original experiment in the original reference frame. Thus, it would not be possible to define "left" and "right" in absolute [7].

Observable	Example	Effect of P
V: polar vector	$\mathbf{p}$	$P\mathbf{v} = -\mathbf{v}$
A: axial vector	$\mathbf{L} = \mathbf{r} \times \mathbf{p}$	$P\mathbf{a} = \mathbf{a}$
S: scalar	$\mathbf{p} \cdot \mathbf{p}$	$Ps = s$
P: pseudoscalar	$\mathbf{L} \cdot \mathbf{p}$	$Pp = -p$

Table 2.1: Transformation properties of some typical observables under P.

2.1.2 Charge Conjugation

The charge conjugation is a sign reversal of all additive quantum numbers such as baryon number, strangeness, lepton number and in particular the electric charge of a particle, i.e. it transforms a particle into his anti-particle. In quantum mechanics this is done by the charge conjugation operator C. Considering a state with definite linear momentum $\mathbf{p}$, angular momentum projection l, spin projection s and charge q, it follows that

$$C|\mathbf{p}|l|s|q\rangle = |\mathbf{p}|l|s|-q\rangle \tag{2.4}$$

leaving quantum numbers related to $\mathbf{p}$, $\mathbf{L}$, $\mathbf{s}$ unaffected. Eigenstates of C are then those who fulfill

$$C|q\rangle = |-q\rangle = |\bar{q}\rangle = \eta_C|q\rangle \quad \Rightarrow q = 0 \tag{2.5}$$

$$C^2|q\rangle = |q\rangle \quad \Rightarrow \eta_C = \pm 1. \tag{2.6}$$

The eigenvalue η_C is commonly called the C-Parity of the system. Since the eigenvalues $|q\rangle$ and $|-q\rangle$ only differ by a phase, they represent the same physical state. For this reason only particles which are their own antiparticles can be eigenstates of the C operator.
Again, C invariance of nature laws would mean that experiments in a world of antiparticles would provide completely identical results to the ones in our world differing only in the fact that all names of particles would be "anti" in reference to ours. In consequence, it would not be possible to define in absolute, whether a particle is composed of matter or antimatter: e.g. the spectral lines emitted by an antiatom, built up of antinucleons and positrons, would be indistinguishable to those of the corresponding atom[7].

2.1.3 *CP* Transformation

The product of the charge conjugation and the parity transformation is referred to as *CP* transformation. Consequently if *CP* Symmetry is conserved, there should be no difference between particles in the real world and antiparticles in the mirrored world[7].
In 1957 the experiment of C.S. Wu [8] could proof for the first time a violation of P Symmetry in weak decays. After this discovery, in the same year, T.D. Lee, R. Oehme and C.N. Yang [9] indicated that charge conjugation symmetry is also violated in nature, since the P and the C symmetries are tightly related within weak interactions. T.D. Lee, and

C.N. Yang were rewarded with the Nobel Prize in 1957 for their investigations regarding violations of parity symmetry in processes involving elementary particles.
These symmetry violations can be visualized without difficulty considering the linear polarization of relativistic fermions produced in weak decays. The linear polarization of a particle with spin $\mathbf{s}$ and momentum $\mathbf{p}$ is commonly characterized by the helicity H, which is a pseudoscalar quantity defined as the projection:

$$H := \hat{\mathbf{s}} \cdot \hat{\mathbf{p}}. \tag{2.7}$$

where $\hat{\mathbf{s}}$ and $\hat{\mathbf{p}}$ are unity vectors parallel to $\mathbf{s}$ and $\mathbf{p}$ respectively. For particles created in weak decays, it can be shown [10] that the helicity expectation value is given by

$$\langle H \rangle = -\beta = \frac{-v}{c} \quad \text{for spin-1/2 fermions and} \tag{2.8}$$

$$\langle H \rangle = \beta \quad = \frac{v}{c} \quad \text{for spin-1/2 antifermions.} \tag{2.9}$$

As it can be seen from this result, the expectation value $\langle H \rangle$ depends in general on the chosen reference frame. To be specific, the direction of $\hat{\mathbf{s}}$ is the same in all reference frames, but this is not the case for $\hat{\mathbf{p}}$. So far $v < c$, the sign of $\hat{\mathbf{p}}$ and thus the sign of its projection to $\hat{\mathbf{s}}$ can change depending on the chosen reference system, since it is possible to find reference frames displacing faster than v. This is not the case for relativistic particles, i.e. in the limit $v \to c$. Relativistic particles with $\langle H \rangle = -1$ are referred to as left-handed and those with $\langle H \rangle = 1$ as right-handed. Therefore, non relativistic particles are not pure left- or right-handed particle states but linear combinations of them.
Complementary to the work of C.S. Wu, in 1958 the experiment of M. Goldhaber *et al.* [11] could confirm that all neutrinos emerging from weak vertices are 100% left-handed. Application of P transforms a left-handed neutrino into a right-handed neutrino. On the other hand C transforms a left-handed neutrino into a left-handed antineutrino. Both states have never been observed. This means that the weak interaction, mediated by W-bosons, do not couple neither to right-handed neutrinos nor to left-handed antineutrinos, violating maximally C and P symmetries (S. Fig. 2.2).
Nevertheless, it can not be deduced from these early experiments that weak interactions break the symmetry under the combined operation CP . The first experiment who proved a violation of CP Symmetry, specifically within flavour mixing in weak interactions, was executed by J.W. Cronin, V.L. Fitch *et al.* [12] in 1965 who received, in turn, the Nobel Prize in 1980. However, it was a very small effect of order $\mathcal{O}(10^{-3})$.
At the time CP violation was first observed, only three quarks had been discovered, namely u, the up quark; d, the down quark; and s, the strange quark. Strange quarks were produced in strong interactions (S. Fig.2.3(a)). Since the strong interaction is strictly flavor conserving, they were always produced in $s\bar{s}$ pairs, a process called therefore "associated production". Though, the decay of hadrons containing s quarks is only possible via the weak interaction (S. Fig. 2.3(b)), which is not conserving flavor. To explain this, it can be argued that the mass eigenstates of the u, d and s quarks are simultaneously eigenstates of

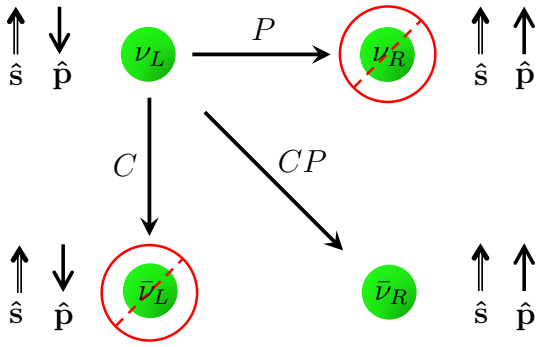
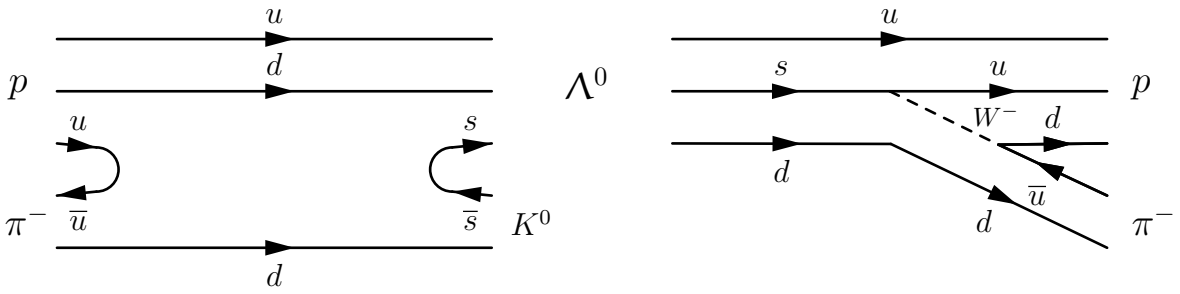


Figure 2.2: The effect of parity P, charge C and the combined CP conjugation operations on a weakly interacting and therefore left-handed neutrino ν_L.

strong interactions but not of weak interactions. As an eigenstate of the weak interaction, N. Cabibbo [13] proposed the left-handed vector doublet

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L = \begin{pmatrix} u \\ d \cdot \cos\theta_C + s \cdot \sin\theta_C \end{pmatrix}_L \tag{2.10}$$

where the mixing angle θ_C, has been called Cabibbo-angle in his honor.



(a) Associated $s\bar{s}$ production

(b) s-quark decay

Figure 2.3: Typical examples of associated $s\bar{s}$ production by the strong interaction and decay of a hadron containing an s-quark.

Nonetheless, the three-quark model led to discrepancies between theoretical predictions and experimental results for higher-order weak interactions. Therefore, in 1970 S. L. Glashow, J. Iliopoulos and L. Maiani [14] introduced a four-quark model, historically called GIM

model, where the eigenstates of the weak interaction were given by

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L = \begin{pmatrix} u \\ d \cdot \cos\theta_C + s \cdot \sin\theta_C \end{pmatrix}_L$$
$$\begin{pmatrix} c \\ s' \end{pmatrix}_L = \begin{pmatrix} c \\ -d \cdot \sin\theta_C + s \cdot \cos\theta_C \end{pmatrix}_L \tag{2.11}$$

with a fourth quark, which they baptized charm quark c.
This model served as starting point to M. Kobayashi and T. Maskawa, who performed the task to investigate conditions to accommodate the violation of CP symmetry in the SM.

2.2 The Six-Quark Model and the CKM Matrix

2.2.1 The Cabibbo Matrix

Flavor mixing arises from the aforementioned mismatch between quark eigenstates of the strong interaction, which correspond simultaneously to particle mass states, and eigenstates of the weak interaction. The latter eigenstates are a certain number of particles folded up into a group called a multiplet. The members of each multiplet are not needed to be identical to single particle species; they are sometimes superpositions of particles. This superposition is then nothing but flavour mixing in a mathematical sense [3].
Starting with the four-quark model, it can be assumed that the eigenstates of the weak interactions are the two doublets given by 2.11. Here, d' and s' are superpositions of the real quark states d and s. This can also be represented as matrix equation through

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \mathbf{V}_C \begin{pmatrix} d \\ s \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} \\ V_{cd} & V_{cs} \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \tag{2.12}$$

where the 2×2 matrix $\mathbf{V}_C$, called Cabibbo matrix, describes the flavour mixing and can be understood mathematically as a rotation in the complex vector space, wherefore it should be unitary.
In general, the elements of a unitary matrix are complex numbers. Some of them, however, can be transformed to real numbers by an adjustment of the own phase factor of the particle state and, at the same time, leaving the physics results intact. Those elements are thus called reducible or otherwise, if they can not be made real, irreducible.
The condition for CP violation in the SM is that irreducible complex numbers appear within the elements of the mixing matrix [3], i.e. at least one complex number should not be possible to remove by adjusting the phase of the particle states. Considering, for example, a 2×2 unitary matrix, the phase adjustment can be done by multiplication with two diagonal matrices whose diagonal elements are pure quark phase factors.
In 1972 M. Kobayashi and T. Maskawa [3, 15] showed that, if an appropriate phase convention is chosen, any 2×2 unitary matrix can be turned into a real matrix by

$$\begin{pmatrix} e^{i\phi_d} & 0 \\ 0 & e^{i\phi_c} \end{pmatrix} \begin{pmatrix} V_{ud} & V_{us} \\ V_{cd} & V_{cs} \end{pmatrix} \begin{pmatrix} e^{-i\phi_d} & 0 \\ 0 & e^{-i\phi_s} \end{pmatrix} = \begin{pmatrix} \cos\phi_C & \sin\phi_C \\ -\sin\phi_C & \cos\phi_C \end{pmatrix} \tag{2.13}$$

without a single irreducible complex number in the final expression, proving consequently, that there is no possibility for CP Violation in the four-quark model.

2.2.2 The CKM Matrix

Recognizing the impossibility to accommodate CP violation in the four-quark model, M. Kobayashi and T. Maskawa considered a six-quark model with the weak doublet eigenstates

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \text{ and } \begin{pmatrix} t \\ b' \end{pmatrix}. \tag{2.14}$$

The quarks in the top row, also called up-type quarks, have electric charge $+\frac{2}{3}e$ while the quarks in the bottom row, i.e. down-type quarks, have electric charge $-\frac{1}{3}e$. Each quark of the six-quark model represent a flavour state. For the flavour mixing they introduced a 3×3 unitary matrix

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \mathbf{V}_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.15}$$

which is the famous Cabibbo-Kobayashi-Maskawa (CKM) matrix. In general, a $n \times n$ unitary CKM matrix contains $n(n+1)/2$ degrees of freedom for complex phases if the doublet quark fields are rewritten in terms of the mass eigenstate quark fields [4]. There are $2n$ quarks and n quark families in this generalized case. Since the effect of one single global phase for all quarks does not change the matrix properties, $2n-1$ degrees of freedom can be eliminated by adjusting quark phase factors. The net number of physical complex phases in the CKM matrix is then [4]

$$n_{CP} = \frac{n\,(n+1)}{2} - (2n-1) = \frac{1}{2}\,(n-2)\,(n-1)\,. \tag{2.16}$$

For $n = 1$ and $n = 2$ this number vanishes, explaining why there is no irreducible CP-violating complex phase in the two- and four-quark models. If $n = 3$, i.e. in the six-quark model, $n_{CP} = 1$. In this case, it is not possible to absorb all complex numbers by the phase convention of the quark states and at least one single complex phase remains [15]. Applying the best adjustment of the phase factors of the quark states, the matrix elements entail four parameters. A widely used parametrization first proposed by [16], but following the notation in [17], is given by

$$\mathbf{V}_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}c_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{2.17}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$ with $i, j = 1, 2, 3$. Here, δ is an irreducible complex phase remaining in the matrix elements and thus violating CP symmetry unless $\delta \in \{0, \pi\}$.

The three angles $\theta_{ij} \in [0, \frac{\pi}{2}]$ are mixing angles, i.e. generalized Cabibbo angles, between the quark families i and j: $\theta_{ij} = 0$ would mean that family i and j are decoupled. The angle θ_{12} corresponds to the aforementioned Cabibbo angle θ_C [7].
Furthermore, the charged weak current responsible for flavour changes is described in the SM via the Yukawa coupling $gW^\mu J^{cc}_\mu$ of the W boson, through the weak coupling constant g, to the current [7]

$$J^{cc}_\mu = \left(\bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau \right) \gamma^\mu \frac{1-\gamma_5}{2} \begin{pmatrix} e \\ \mu \\ \tau \end{pmatrix} + \sum_{r,g,b} \left(\bar{u}, \bar{c}, \bar{t} \right) \gamma^\mu \frac{1-\gamma_5}{2} \cdot \mathbf{V}_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.18}$$

where in the latter term the triplet of up-type quarks is coupled through the V operator γ^μ with the triplet of down-type quarks, which are rotated by the CKM Matrix. Here, the chirality operator

$$P_L \psi = \frac{1-\gamma_5}{2} \psi = \psi_L \tag{2.19}$$

$$\Rightarrow \bar{\psi}_L \gamma^\mu \psi_L = \bar{\psi} \gamma^\mu \frac{1-\gamma_5}{2} \psi \tag{2.20}$$

projects the left-handed components of the triplets, i.e. that in the expression 2.18 only the left-handed components of the fermion triplets participate in weak interactions. Since $\gamma^\mu \gamma^5$ is an operator of type A, the resulting weak operator $\gamma^\mu P_L$ is a V−A superposition. The last two statements mean nothing else, but that the quark weak current is asymmetric under C and P transformations. This is consistent to the experimental results exposed in subsection 2.1.3, where it has been also explained, that only relativistic or, in other words, massless particles can be completely left-handed. Consequently, only these particle states are eigenstates of the P_L operator. A graphic representation of a weak charged current is shown in Fig. 2.4.

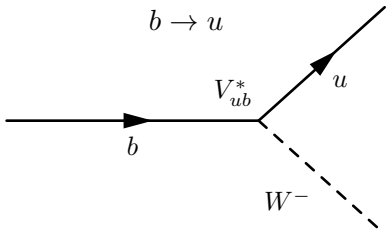


Figure 2.4: A b-quark decays into an u-quark by emmitance of a W^- boson. The weak charged current coupling between $\bar{u}$ and $\bar{b}$ is proportional to $V_{ub} \sim 0.004$ and thus suppressed.

It is yet easier to understand, that an irreducible complex phase in the CKM matrix leads to interference among the different components of the quark weak current [15]. This produces

a coupling asymmetry between quark families and this is exactly how *CP* is accommodated in the SM through the CKM matrix: without any interference, the current in 2.18 would be symmetric under *CP* transformations.

2.2.3 The Wolfenstein Parametrization and the Unitarity Triangles

The CKM matrix parametrization can be arbitrarily chosen, since all them lead to the same physical results. For phenomenological analysis, however, it is convenient to use the parametrization proposed by L. Wolfenstein [18]. He introduced four parameters λ, A, ρ and η, defining simultaneously

$$s_{12} = \lambda \quad \Rightarrow \quad c_{12} = \sqrt{1-\lambda^2} \tag{2.21}$$

$$s_{23} = A\lambda^2 \quad \Rightarrow \quad c_{23} = \sqrt{1-A^2\lambda^4} \tag{2.22}$$

$$s_{13}e^{-i\delta} = A\lambda^3(\rho - i\eta) \quad \Rightarrow \quad c_{13} = \sqrt{1-A^2\lambda^3(\rho^2+\eta^2)} \tag{2.23}$$

and developing afterwards the elements of the CKM matrix with $\lambda = sin(\theta_C) = 0.221 \pm 0.002$ [19] as expansion parameter. The other real parameters are $A, \rho, \eta \in (0,1)$. Originally, he proposed an expansion up to the $\mathcal{O}(\lambda^3)$ order, however, considering higher order weak interactions, it is more convenient to use an approximation up to the $\mathcal{O}(\lambda^6)$ order [20]. The CKM matrix can then be expressed by

$$\mathbf{V}_{CKM} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2-\frac{1}{8}\lambda^4 & \lambda & A\lambda^3(\rho-i\eta) \\ -\lambda+\frac{1}{2}A^2\lambda^5[1-2(\rho+i\eta)] & 1-\frac{1}{2}\lambda^2-\frac{1}{8}\lambda^4(1+4A^2) & A\lambda^2 \\ A\lambda^3[1-(1-\frac{1}{2}\lambda^2)(\rho+i\eta)] & -A\lambda^2+\frac{1}{2}A\lambda^4[1-2(\rho+i\eta)] & 1-\frac{1}{2}A^2\lambda^4 \end{pmatrix} + \mathcal{O}(\lambda^6) \tag{2.24}$$

violating *CP* if $\eta \neq 0$. Taking now the unitarity condition

$$\mathbf{V}_{CKM}^{\dagger}\mathbf{V}_{CKM} = \mathbb{1} \quad \Rightarrow \sum_{j\in\{u,c,t\}} V_{ji}^* V_{jk} = \delta_{ik} \text{ with } i,k \in \{d,s,b\} \tag{2.25}$$

$$\mathbf{V}_{CKM}\mathbf{V}_{CKM}^{\dagger} = \mathbb{1} \quad \Rightarrow \sum_{j\in\{d,s,b\}} V_{ij} V_{kj}^* = \delta_{ik} \text{ with } i,k \in \{u,c,t\} \tag{2.26}$$

into account, twelve distinct complex relations among the matrix elements emerge. Six of them, those with $i \neq k$, can be geometrically represented as triangles in the complex space, whose interior angles represent observable quantities. The six triangles have different shapes by expressing the side lengths in terms of powers of λ. The fact that they are not degenerate into a line, i.e. that it is not possible to make real all the three vectors describing the sides of each triangle, is in addition the geometrical consequence of CP violation [21]. All triangles have the same area, commonly written as $J_{CP}/2$, where $J_{CP} = \lambda^6 A^2 \eta$ is the

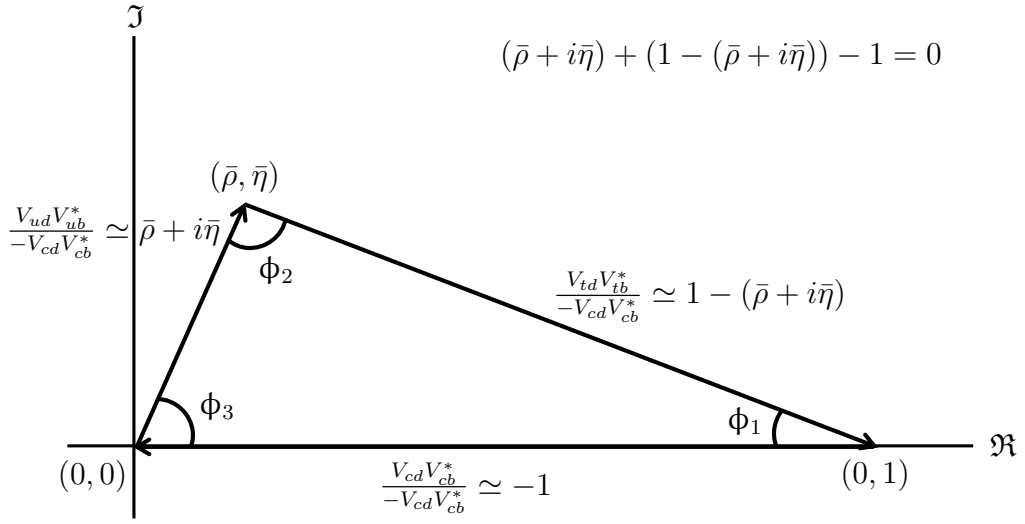


Figure 2.5: Visualization of the unitarity triangle corresponding to equation 2.27.

single *CP* violating parameter if $J_{CP} \neq 0$ [22].
Two of these unitarity triangles,

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* \;\Rightarrow\; \mathcal{O}(\lambda^3)+\mathcal{O}(\lambda^3)+\mathcal{O}(\lambda^3)=0 \tag{2.27}$$

$$V_{td}V_{ud}^* + V_{ts}V_{us}^* + V_{tb}V_{ub}^* \;\Rightarrow\; \mathcal{O}(\lambda^3)+\mathcal{O}(\lambda^3)+\mathcal{O}(\lambda^3)=0 \tag{2.28}$$

have terms of equal order and thus sides of comparable length. In consequence, the interior angles are large enough to be measured [22].
Within the scope of this work the triangle described by 2.27 is more relevant, since it is related to processes involving B Mesons. This means, accordingly, that a relative large *CP* violation is expected in the B system. A simple visualization of this triangle, as shown in Fig. 2.5, can be done dividing the whole equation in 2.27 by $-V_{cd}V_{cb}^*$.
Simplifying equation 2.27 with

$$\bar{\rho} = \left(1-\frac{\lambda^2}{2}\right)\rho \text{ and } \bar{\rho} = \left(1-\frac{\lambda^2}{2}\right)\eta \tag{2.29}$$

the inner angles of the considered triangle, can be then expressed by

$$\alpha \equiv \phi_2 \equiv \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right) \simeq \arg\left(-\frac{1-\bar{\rho}-i\bar{\eta}}{\bar{\rho}+i\bar{\eta}}\right) \tag{2.30}$$

$$\beta \equiv \phi_1 \equiv \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right) \simeq \arg\left(\frac{1}{1-\bar{\rho}-i\bar{\eta}}\right) \tag{2.31}$$

$$\gamma \equiv \phi_3 \equiv \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right) \simeq \arg\left(\bar{\rho}+i\bar{\eta}\right) \tag{2.32}$$

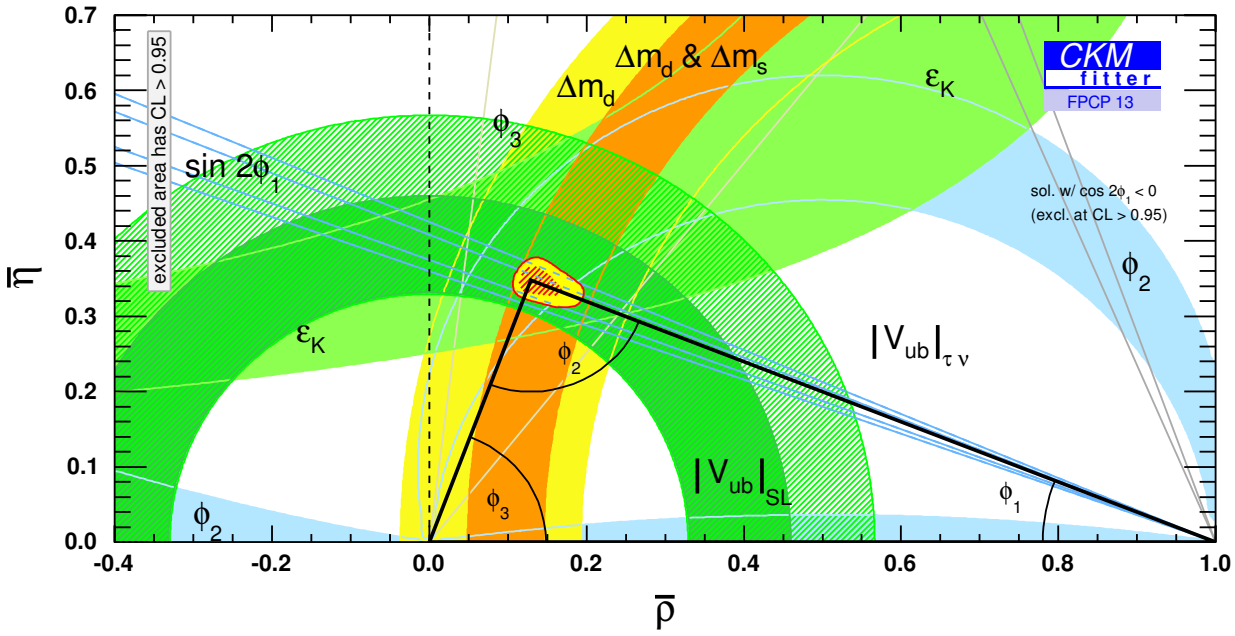


Figure 2.6: Current constraints on the CKM Triangle [23].

where it can be noted that the angle γ also coincides to a good approximation with the irreducible complex phase δ in 2.17. In the SM, therefore, all *CP* violating phenomena are related, since they all stem from the appearance of this single phase [21].
Since the sides of the triangles are products of CKM matrix elements, their lengths can be quantified measuring proper decay rates. Simultaneously, the interior angles can be obtained from them. The measurement problem is evidently over-constrained from the experimental point of view, since each triangle is defined by either three sides; two sides and one angle; or oneside and two angles.
These measurements, however, offer an excellent test of the SM: if the description of *CP* violation in the SM is correct and complete, the unitarity condition of the CKM matrix must be fullfilled, implying that each unitarity triangle closes. Experimental results considerably deviating from the unitarity condition would provide an indubitable hint for new physics (NP) beyond the SM.
At the actual level of experimental precision, as shown in Fig. 2.6 [23], combined results from different measurements agree with the theoretical predictions of SM for the triangle in 2.27 within the errors. However, it is not possible to affirm that there is no scope for NP at the present time. In fact, a reduction of experimental uncertainties is needed to tighten the constraints: a task that can be achieved through new high precision measurements as those expected from the new Belle II experiment.

2.3 Flavoured Neutral Mesons

2.3.1 Quantum-Mechanical Description of Flavour Oscillation

Neutral mesons are unstable subatomic particles composed of a quark and an antiquark with opposite electric charges to each other. Because of their short lifetime, they can not be always observed directly, but through their decays into other final particles.
A neutral meson has a characteristic quantum number, the aforementioned flavour, referring to its quark composition, e.g. "strangeness", related to s quarks; "charm", related to c quarks and "beauty", related to b quarks (S. Table 2.2). These quantum numbers are conserved in strong interactions and thus the flavour eigenstates of mesons are simultaneously eigenstates of strong interactions.

	Strangeness (S)	Beauty (B)
$K^0(d\bar{s})$	+1	0
$\bar{K}^0(\bar{d}s)$	-1	0
$B^0(d\bar{b})$	0	+1
$\bar{B}^0(\bar{d}b)$	0	-1

Table 2.2: Flavour quantum numbers for neutral Kaons and B mesons. Weak decays, e.g. $K^0 \to \pi^+\pi^-$, change flavour numbers by one unit, $|\Delta S| = 1$. A meson oscillation, eg. $|K^0\rangle \to |\bar{K}^0\rangle$, implies a flavour change by two units, $|\Delta S| = 2$.

On the other hand, a neutral meson can not be its own antiparticle, since the flavour eigenstate $|M\rangle$ of a neutral meson is not an eigenstate of the charge operator:

$$C|M\rangle = |\bar{M}\rangle. \tag{2.33}$$

Nevertheless, weak interactions do not conserve flavour quantum numbers and, consequently, a neutral meson $|M\rangle$ can undergo a transition to its antiparticle $|\bar{M}\rangle$ or vice versa. Ergo, neutral mesons are subjected to oscillations in the flavour space during their whole lifetime and, therefore, what can be measured in experiments, are not the pure flavour eigenstates $|M\rangle$ and $|\bar{M}\rangle$, but superpositions of them.
A neutral meson can be thus considered as a two-state quantum system with the generic wave function:

$$|\psi(t)\rangle = \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = a(t)|M\rangle + b(t)|\bar{M}\rangle. \tag{2.34}$$

This approach, first proposed by Weisskopf and Wigner [24, 25], is valid if the time evolution of the coefficients is studied for time scales which are much larger than typical time scales of strong interactions. The effect of weak interactions can be then considered as a perturbation of the strong interactions [21]. Indeed, weak interactions in first order only change flavour by one unit: for a transition $|M\rangle \to |\bar{M}\rangle$ a flavour change by two units is needed. This means that the mixing of neutral flavoured mesons is an effect of second order

weak interactions, i.e. it is caused by a perturbation of second order. The time evolution is described by the time-dependent Schrödinger equation

$$i\hbar\frac{d}{dt}|\psi\rangle = \mathcal{H}|\psi\rangle \tag{2.35}$$

whose 2×2 hamiltonian matrix can be decomposed into a mass and a decay matrix

$$\mathcal{H} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} \tag{2.36}$$

where $\mathbf{M}$ and $\mathbf{\Gamma}$ are hermitian matrices. The hermicity condition imposes that $M_{12} = M_{21}^*$ and $\Gamma_{12} = \Gamma_{21}^*$. In addition, the so called CPT theorem [26] requires that $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$. This diagonal elements are determined by quark masses together with strong and electromagnetic (EM) interactions. Subsequently, the diagonal elements of $\mathcal{H}$ can be identified as the strong and EM hamiltonian $\mathcal{H}_0$ [7, 21]. The not diagonal elements are then those involving weak interactions. These elements can be merged in a second order perturbation hamiltonian $\mathcal{H}_W$. It follows that

$$\mathcal{H} = \begin{pmatrix} H_0 & 0 \\ 0 & H_0 \end{pmatrix} + \mathcal{H}_W = \begin{pmatrix} M - \frac{i}{2}\Gamma & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M - \frac{i}{2}\Gamma \end{pmatrix}. \tag{2.37}$$

Since the phase factors of $|M\rangle$ and $|\bar{M}\rangle$ can be arbitrarily adjusted without changing the physics results, as argued before in subsection 2.2.2, the phases of M_{12} and Γ_{12} can be chosen arbitrarily and only the difference between them is significant. This phase difference can be quantified by a complex weak phase ϕ_{12} [27]. Consequently, for the mixing physics the relevant quantities are

$$|M_{12}|,\ |\Gamma_{12}| \text{ and } \phi_{12} = \arg\left(-\frac{M_{12}}{\Gamma_{12}}\right). \tag{2.38}$$

By diagonalizing the hamilton operator in 2.37, the time evolution of the meson-antimeson system can be described in terms of its eigenvalues and eigenstates [19]. For the eigenvalues $\omega_\pm$, it can be obtained

$$\omega_\pm = M - \frac{i}{2}\Gamma \pm \sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)} = M - \frac{i}{2}\Gamma \pm pq. \tag{2.39}$$

The two associated eigenstates $v_\pm = (1, \pm\frac{q}{p})^T$ represent the physical mass eigenstates

$$\begin{aligned} |M_+(t)\rangle &\equiv p|M(t)\rangle + q|\bar{M}(t)\rangle, \\ |M_-(t)\rangle &\equiv p|M(t)\rangle - q|\bar{M}(t)\rangle \end{aligned} \tag{2.40}$$

where p and q are complex numbers which satisfy

$$\frac{q}{p} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}} = e^{-i\phi_M}\sqrt{\frac{|M_{12}| + \frac{i}{2}|\Gamma_{12}|e^{i\phi_{12}}}{|M_{12}| + \frac{i}{2}|\Gamma_{12}|e^{-i\phi_{12}}}} \text{ with } |p|^2 + |q|^2 = 1. \tag{2.41}$$

Here, the absolute mixing weak phase $\phi_M \equiv \arg(M_{12})$ has been introduced. It should be noted at this point that the mass eigenstates $|M_\pm(t)\rangle$ correspond to pure *CP* eigenstates if $\phi_{12} \in \{0, \pi\} \Leftrightarrow |q/p| = 1$. On the other hand, the absolute phase ϕ_M does not violate *CP* symmetry in the mixing, since it cancels out building the absolute value of $|q/p|$.
In the case $\phi_{12} \in \{0, \pi\}$, *CP* invariance would be fulfilled and the two eigenstates in 2.40 (even and odd) would be orthogonal to each other. Otherwise, if they are not orthogonal, the overlap between them is given by

$$\langle M_+|M_-\rangle = \frac{1 - \left|\frac{q}{p}\right|^2}{1 + \left|\frac{q}{p}\right|^2}. \tag{2.42}$$

The masses $m_\pm$ and the lifetimes $\Gamma_\pm$ for the mass eigenstates states can be derived from the eigenvalues

$$m_+ = \Re(\alpha_+),\ \Gamma_+ = -2\Im(\alpha_+), \tag{2.43}$$
$$m_- = \Re(\alpha_-),\ \Gamma_- = -2\Im(\alpha_-) \tag{2.44}$$

which are commonly recasted in terms of the observables

$$m \equiv \frac{1}{2}(m_+ + m_-) = M, \tag{2.45}$$
$$\Gamma \equiv \frac{1}{2}(\Gamma_+ + \Gamma_-) = \Gamma, \tag{2.46}$$
$$\Delta m \equiv m_- - m_+ = -2\Re\left(pq\right), \tag{2.47}$$
$$\Delta\Gamma \equiv \Gamma_- - \Gamma_+ = 4\Im\left(pq\right). \tag{2.48}$$

Subsequently the time evolution of $|M_\pm\rangle$ is given by

$$\begin{aligned} |M_+(t)\rangle &= e^{-i\alpha_+ t}\left(p|M\rangle + q|\bar{M}\rangle\right), \\ |M_-(t)\rangle &= e^{-i\alpha_- t}\left(p|M\rangle - q|\bar{M}\rangle\right) \end{aligned} \tag{2.49}$$

where both mass eigenstates decay as $e^{-\Gamma_\pm t}$ with a modulation generated by the complex phase $e^{im_\pm}$. As a further matter, the time evolution of the pure flavour states is needed for a deeper understanding of *CP* violation. From the expressions in 2.40 it follows that

$$\begin{aligned} |M(t)\rangle &\equiv \frac{1}{2p}\left(|M_+(t)\rangle + |M_-(t)\rangle\right), \\ |\bar{M}(t)\rangle &\equiv \frac{1}{2q}\left(|M_+(t)\rangle - |M_-(t)\rangle\right) \end{aligned} \tag{2.50}$$

where inserting the time evolution in 2.49 yields

$$|M(t)\rangle = \frac{1}{2}\left(\left(e^{-i\alpha_+ t} + e^{-i\alpha_- t}\right)|M\rangle + \frac{q}{p}\left(e^{-i\alpha_+ t} - e^{-i\alpha_- t}\right)|\bar{M}\rangle\right)$$
$$= \frac{1}{2}e^{-i\alpha_+}\left(\left(1 + e^{-i\left(\Delta m - \frac{i}{2}\Delta\Gamma\right)t}\right)|M\rangle + \frac{q}{p}\left(1 - e^{-i\left(\Delta m - \frac{i}{2}\Delta\Gamma\right)t}\right)|\bar{M}\rangle\right), \tag{2.51}$$

$$|\bar{M}(t)\rangle = \frac{1}{2}\left(\frac{p}{q}\left(e^{-i\alpha_+ t} - e^{-i\alpha_- t}\right)|M\rangle + \left(e^{-i\alpha_+ t} + e^{-i\alpha_- t}\right)|\bar{M}\rangle\right)$$
$$= \frac{1}{2}e^{-i\alpha_+}\left(\frac{p}{q}\left(1 - e^{-i\left(\Delta m - \frac{i}{2}\Delta\Gamma\right)t}\right)|M\rangle + \left(1 + e^{-i\left(\Delta m - \frac{i}{2}\Delta\Gamma\right)t}\right)|\bar{M}\rangle\right). \tag{2.52}$$

With non vanishing Δm and $\Delta\Gamma$, which are related to the weak interaction components of the hamiltonian, an initial pure flavour state will develop and interfere with its opposite flavour state while it simultaneously decays. In fact, decay rates can be only properly defined, if the initial state is specified either as $|M\rangle$ or as $|\bar{M}\rangle$ [28]. Starting with a specific flavour, e.g. $|M\rangle$, the probability to observe a $|\bar{M}\rangle$ after a time t' can be calculated as

$$|\langle\bar{M}|M(t')\rangle|^2 = \frac{1}{4}e^{-2i\alpha_+ t'}\left|\frac{q}{p}\right|^2\left|1 - e^{-i\left(\Delta m - \frac{i}{2}\Delta\Gamma\right)t'}\right|^2$$
$$= \frac{1}{2}e^{-2i\alpha_+ t'}\left|\frac{q}{p}\right|^2\left(1 - \cos\left(\left[\Delta m - \frac{i}{2}\Delta\Gamma\right]t'\right)\right)^2 \tag{2.53}$$

using the orthogonality between flavour eigenstates. In the B meson system, which are mesons containing b quarks, the mass eigenstates $|P_\pm\rangle$ are commonly called $|B_{H,L}\rangle$, where H stand for heavy and L for light. In the Kaon system, which are mesons containing s quarks, the convention light and heavy is interchanged with short and long, i.e. $|K_{S,L}\rangle$. This description of flavour oscillations in neutral mesons was necessary, since there is a rich exhibition of CP Violation in these systems and, indeed, they have been the only observed source of this phenomenon so far.

2.3.2 Types of *CP* Violation

After the previous introduction to flavour mixing, it is possible to classify the different CP violation mechanisms into three types depending on the way, how the complex phase appears. For this purpose, it is needed to consider four amplitudes for the transition of a flavoured neutral meson $|M\rangle$ together with its antiparticle $|\bar{M}\rangle$ into a final state $|f\rangle$ or rather $|\bar{f}\rangle$

$$A_f = \langle f|H|M\rangle = A\left(|M\rangle \to |f\rangle\right), \tag{2.54}$$
$$\bar{A}_f = \langle f|H|\bar{M}\rangle = A\left(|\bar{M}\rangle \to |f\rangle\right), \tag{2.55}$$
$$A_{\bar{f}} = \langle \bar{f}|H|M\rangle = A\left(|M\rangle \to |\bar{f}\rangle\right), \tag{2.56}$$
$$\bar{A}_{\bar{f}} = \langle \bar{f}|H|\bar{M}\rangle = A\left(|\bar{M}\rangle \to |\bar{f}\rangle\right) \tag{2.57}$$

with the hamiltonian operator H. The different types of CP violation are then:

a) ***CP* violation in decay** also called direct CP violation occours in decays if

$$\left| \frac{A_f}{\bar{A}_{\bar{f}}} \right| \neq 1 \text{ or } \left| \frac{\bar{A}_f}{A_{\bar{f}}} \right| \neq 1. \tag{2.58}$$

This is the only kind of CP violation that is not only possible for neutral mesons, but also for charged mesons. Other types of CP violation exclude charged mesons, since they do not mix with their antiparticles. The phase difference between the decay amplitudes entailed in 2.58 can be quantified by a complex phase ϕ_D, which is also a superposition of complex weak phases stemming from the CKM matrix components and the own not CP violating strong phases of the involved particle states.

b) ***CP* violation in mixing** or indirect CP violation means, by definition, that

$$\left| \frac{q}{p} \right| \neq 1 \;\Rightarrow\; \epsilon \neq 0. \tag{2.59}$$

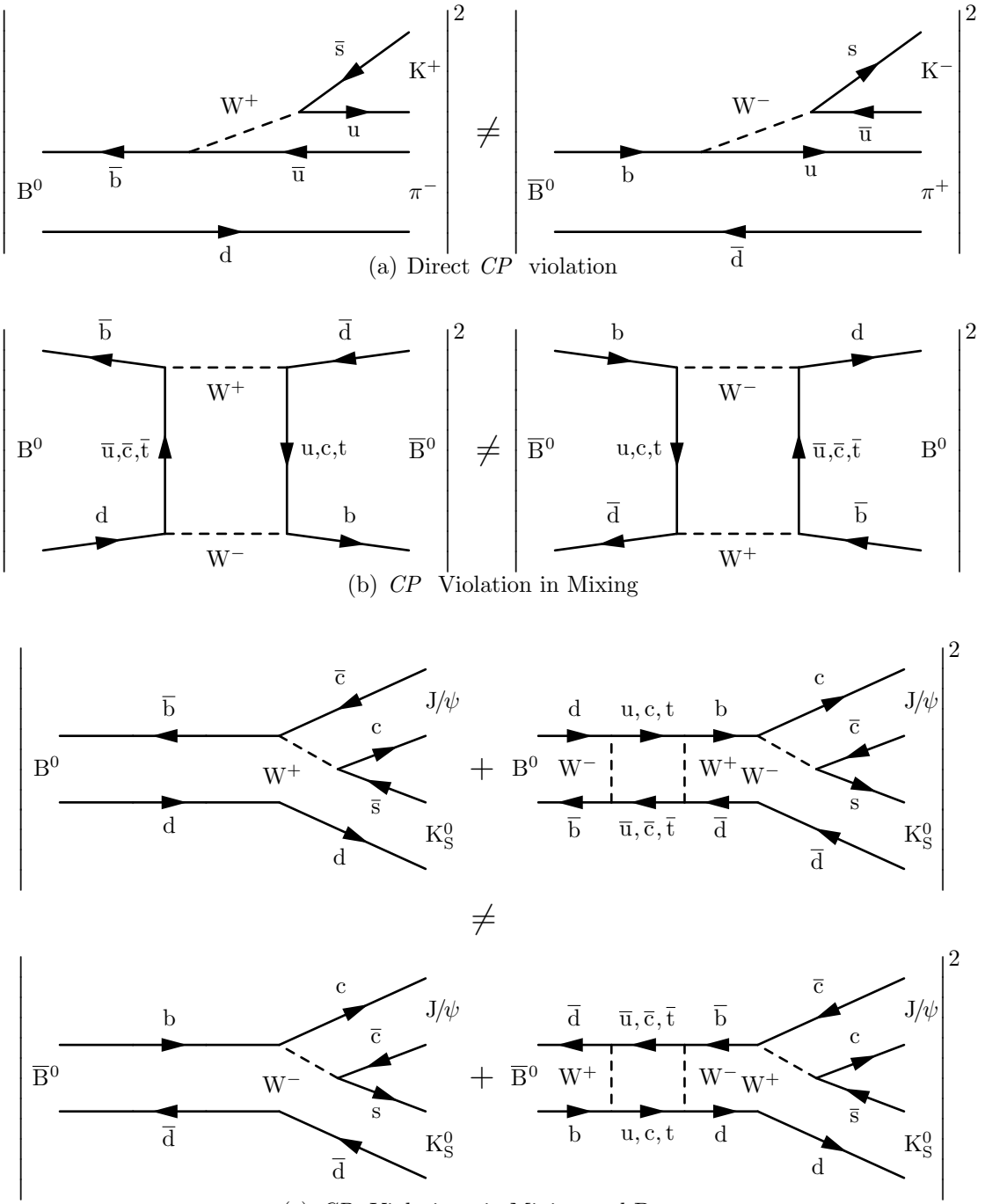
In this case CP violation arises as a consequence of an asymmetry in the flavour oscillation $|M\rangle \to |\bar{M}\rangle \to |M\rangle$, what means that physical mass eigenstates are not CP eigenstates, as explained in section 2.3.1. In other words, there is a complex weak phase ϕ_{12} between M_{12} and Γ_{12}.

c) ***CP* violation by interference of mixing and decay** is possible to be observed when the neutral mesons $|M\rangle$ and $|\bar{M}\rangle$ have both a common final state $|f\rangle$, preferentially a pure CP eigenstate $|f\rangle = |\bar{f}\rangle = |f_{CP}\rangle$. Even if there is no CP violation in mixing and decay respectively, i.e. if $|\bar{A}_{f_{CP}}/A_{f_{CP}}| = |q/p| = 1$, the superposition of ϕ_D and ϕ_M can produce a phase difference and thus an interference between these two processes, generating consequently a violation of CP symmetry. The condition for this kind of CP violation, introducing a new complex quantity λ_{CP}, can be written as

$$\Im\left(\lambda_{CP}\right) \neq 0, \quad \text{where} \quad \lambda_{CP} \equiv \frac{q}{p} \cdot \frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}} = \left| \frac{q}{p} \right| \cdot \left| \frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}} \right| e^{-i(\phi_M + \phi_D)} \,. \tag{2.60}$$

This phenomenon was observed for the first time by the Belle [29] as well as by the BaBar [30] collaboration within the so called golden channel $B \to J/\psi K_S^0$ and related channels.

Typical Feynman diagrams for the three aforementioned CP Violation types are presented in Fig. 2.7[31].



(a) Direct *CP* violation

(b) *CP* Violation in Mixing

(c) *CP* Violation via Mixing and Decay

Figure 2.7: Examples for the three types of *CP* violation in neutral B meson decays. *CP* violation is observed through assymmetries in the decay rates and, therefore, in the squared Feynman amplitudes for B^0 and $\overline{B}^0$ [31].

2.4 *CP* Violation in the B-Sector

2.4.1 B Mesons phenomenology

In 1974 the J/ψ meson was discovered and it was identified as a bound state of the fourth quark c, proposed by the GIM mechanism, and its antiparticle $\bar{c}$. Subsequently, the discovery of the Upsilon meson Υ in 1977, which turned out to be a bound state of a bottom-anti-bottom quark pair, had a great impact on the six-quark model proposed by the CKM mechanism which began then to be accepted as a standard one. The discovery of the last quark t, however, had to wait until 1995. Both b and t quarks were discovered at the Fermilab in Illinois. Fig. 2.8 illustrates the actual SM classification of elementary particles into three generations.
Besides the existence of the six quarks, the most significant predictions of the CKM mechanism are those related to violations of CP symmetry. As argued in subsection 2.2.3, the B meson system offers the possibility to test many of these predictions since a relative large CP violation is expected therein.
B mesons consist of a heavy b or $\bar{b}$ quark and a lighter quark. Depending on the lighter quark, B mesons can be classified into two families, namely, a neutral and a charged one. The neutral B mesons are: B_d, represented by $B^0(d\bar{b})$ and $\bar{B}^0(\bar{d}b)$ with a mass $m_{B^0} = 5.2795 \pm 0.0003$ GeV; and B_s, represented by $B_s^0(s\bar{b})$ and $\bar{B}_s^0(\bar{s}b)$ with a mass $m_{B_s^0} = 5.3663 \pm 0.0006$ GeV. Analogously, the charged B mesons are: $B^{\pm}$, represented by $B^+(u\bar{b})$ and $B^-(\bar{u}b)$ with a mass $m_{B^{\pm}} = 5.27917 \pm 0.00029$ GeV; and $B_c^{\pm}$, represented by $B_c^+(c\bar{b})$ and $B_c^-(\bar{c}b)$ with a mass $m_{B_c^{\pm}} = 6.277 \pm 0.006$ GeV.

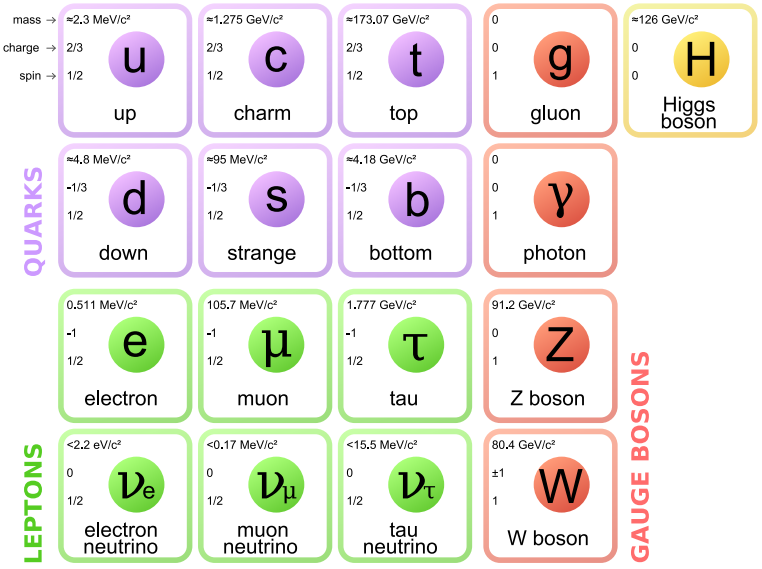


Figure 2.8: The three particle generations and the gauge bosons in the SM [32].

Additionally, B mesons have relatively long lifetimes. This property makes a little easier to study $B^0 - \bar{B}^0$ oscillations as well as B decays. With the exception of the heavier $B_c^{\pm}$ mesons with a lifetime of $\tau \backsim 0.5$ ps, all other B mesons have lifetimes around 1.5 ps, e.g. $\tau_{B^0} = (1.525 \pm 0.009) \times 10^{-12}$ s. All particle data in this work have been taken from the Particle Data Group (PDG) [22].
The comparably long lifetimes of B mesons are a consequence of small couplings through the CKM matrix between the third quark family (t, b) and the other two (c, s) and (u, d). Furthermore, the fact that decays via the $b \to c$ transition dominate over those via $b \to u, d, s$ transitions points out directly the "hierarchical" structure of the CKM matrix.

2.4.2 Flavour Oscillation of Neutral B Mesons

In the case of neutral B mesons the time evolution description expressed in the equations 2.51 and 2.52 can be simplified taking the following features into account. First, the lifetime difference between the two physical eigenstates B_H and B_L is practically negligible since [22]

$$\frac{|\Delta\Gamma|_B}{\Gamma_B} = -0.001 \pm 0.014 \tag{2.61}$$

and thus it can be ignored, i.e. $\Delta\Gamma = 0$. The lifetime is then redefined as $\Gamma \equiv \Gamma_H = \Gamma_L$. Second, the *CP* violation in $B^0 - \bar{B}^0$ mixing is a very small effect [22]

$$\left| \frac{q}{p} \right|_{B_d} = 1.0002 \pm 0.0028. \tag{2.62}$$

Therefore, the eigenstates B_H and B_L can be considered as pure flavour eigenstates in a good approximation. This actual experimental result was already predicted by Bander, Silverman and Soni in 1979 [33]; considering cancellations due to the GIM mechanism and the fact that B^0-oscillations are dominated by t-quark contributions, i.e. $V_{tb} \simeq 1 \gg V_{cb} \gg V_{ub}$, it follows from equation 2.41 that

$$\left. \frac{q}{p} \right|_{B_d} = \sqrt{\frac{M_{12}^*}{M_{12}}} + \mathcal{O}\left(\frac{\Gamma_{12}}{M_{12}}\right) \Bigg|_{B_d} \simeq \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}, \tag{2.63}$$

$$\Rightarrow \left| \frac{q}{p} \right|_{B_d} \backsim 1 \tag{2.64}$$

since $\Gamma_{12}/M_{12} \simeq m_b^2/m_t^2 \backsim \mathcal{O}(10^{-2})$ [19]. Ergo, the time evolution for neutral B mesons, starting as a pure flavour state, is given by

$$|B^0(t)\rangle = \frac{1}{2} e^{-\frac{\Gamma}{2}t} \left(\left(1 + e^{-i\Delta m t}\right) |B^0\rangle + \frac{q}{p} \left(1 - e^{-i\Delta m t}\right) |\bar{B}^0\rangle \right), \tag{2.65}$$

$$|\bar{B}^0(t)\rangle = \frac{1}{2} e^{-\frac{\Gamma}{2}t} \left(\frac{p}{q} \left(1 - e^{-i\Delta m t}\right) |B^0\rangle + \left(1 + e^{-i\Delta m t}\right) |\bar{B}^0\rangle \right) \tag{2.66}$$

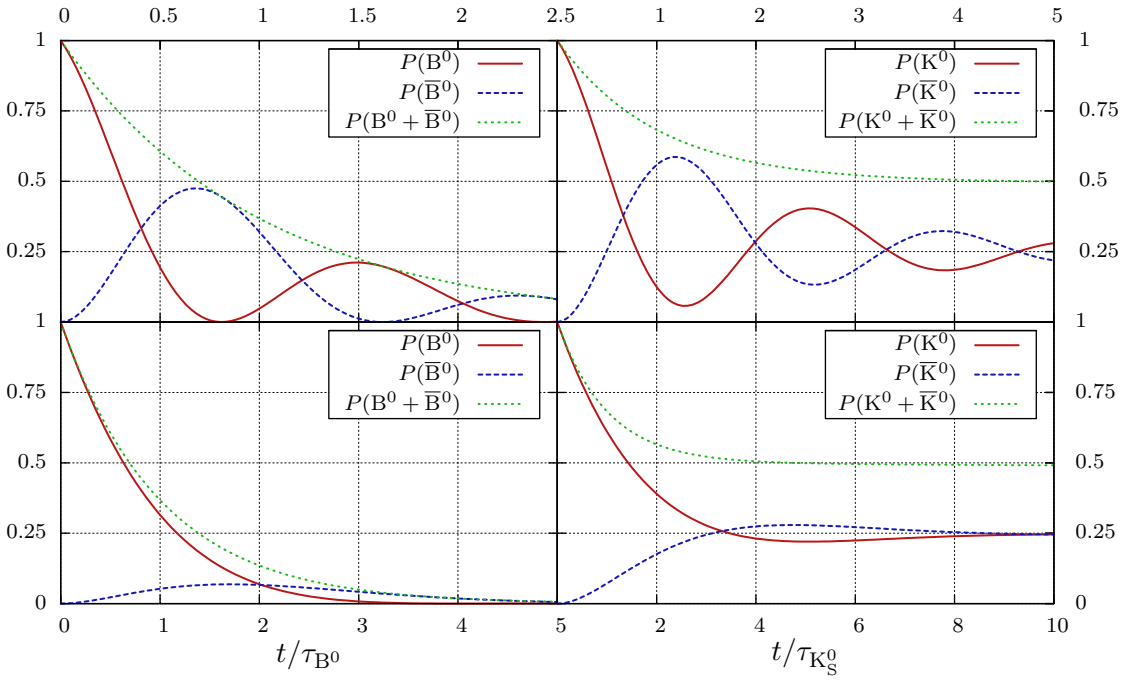


Figure 2.9: Mixing probabilities for initial pure B^0 (left) and K^0 (right) mesons as functions of their lifetimes. The bottom row shows the probabilities using the values $\Delta m/\tau$ from experimental results. The top row shows the probabilities for a Δm 5 times the measured value [31].

where $e^{-im_H t}$ have been removed by phase convention. The probabilities of observing a B^0 or a $\bar{B}^0$ after a time t', assuming that the initial state is a pure B^0, can be calculated as

$$P(B^0) = |\langle B^0|B^0(t')\rangle|^2 = \frac{1}{2}e^{-\Gamma t'}\left(1+\cos\left(\Delta m t'\right)\right)^2, \tag{2.67}$$

$$P(\bar{B}^0) = |\langle \bar{B}^0|B^0(t')\rangle|^2 = \frac{1}{2}e^{-\Gamma t'}\left(1-\cos\left(\Delta m t'\right)\right)^2. \tag{2.68}$$

In Fig. 2.9[31], these probabilities have been plotted using values from experimental results, $\Delta m_{B^0}/\tau_{B^0} = 0.774 \pm 0.008$ [22]. As it can be seen there, the most B^0 mesons decay before they change flavour, but there is a non vanishing probability that an initial B^0 develops into a pure $\bar{B}^0$. In the neutral K meson system an initial K^0 will never convert into a pure $\bar{K}^0$.
Moreover, in the neutral B meson system, the complex quantity λ_{CP} can be written as

$$\lambda_{CP} = e^{-i\phi_M(B)} \cdot \frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} \cdot \frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}}, \tag{2.69}$$

$$\Rightarrow |\lambda_{CP}| = \left|\frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}}\right|. \tag{2.70}$$

2.4.3 Time Dependent *CP* Violation

Now, the case where both B^0 and $\bar{B}^0$ decay into the same CP eigenstate f_{CP} is considered again. The corresponding decay amplitudes are given by

$$A_{CP} = \langle f_{CP}|H|B^0\rangle, \quad \bar{A}_{CP} = \langle f_{CP}|H|\bar{B}^0\rangle. \tag{2.71}$$

If the time evolution in equations 2.65, 2.66 is used, the time dependent decay amplitudes can be expressed through

$$\begin{aligned} A_{CP}(t) = \langle f_{CP}|H|B^0(t)\rangle &= \frac{1}{2}e^{-\frac{\Gamma}{2}t}\left(\left(1+e^{-i\Delta mt}\right)A_{CP} + \frac{q}{p}\left(1-e^{-i\Delta mt}\right)\bar{A}_{CP}\right) \\ &= \frac{1}{2}e^{-\frac{\Gamma}{2}t}A_{CP}\left[\left(1+e^{-i\Delta mt}\right) + \lambda_{CP}\left(1-e^{-i\Delta mt}\right)\right], \end{aligned} \tag{2.72}$$

$$\begin{aligned} \bar{A}_{CP}(t) = \langle f_{CP}|H|\bar{B}^0(t)\rangle &= \frac{1}{2}e^{-\frac{\Gamma}{2}t}\left(\frac{p}{q}\left(1-e^{-i\Delta mt}\right)A_{CP} + \left(1+e^{-i\Delta mt}\right)\bar{A}_{CP}\right) \\ &= \frac{1}{2}e^{-\frac{\Gamma}{2}t}A_{CP}\frac{p}{q}\left[\left(1-e^{-i\Delta mt}\right) + \lambda_{CP}\left(1+e^{-i\Delta mt}\right)\right]. \end{aligned} \tag{2.73}$$

In turn, the time dependent decay rates are

$$\begin{aligned} \Gamma(B^0(t)\to f_{CP}) &= \left|\langle f_{CP}|H|B^0(t)\rangle\right|^2 \\ &= \frac{1}{2}e^{-\Gamma t}|A_{CP}|^2 \\ &\cdot\left[1+\cos(\Delta mt) - 2\mathfrak{I}(\lambda_{CP})\sin(\Delta mt) + (1-\cos(\Delta mt))\,|\lambda_{CP}|^2\right], \end{aligned} \tag{2.74}$$

$$\begin{aligned} \Gamma(\bar{B}^0(t)\to f_{CP}) &= \left|\langle f_{CP}|H|\bar{B}^0(t)\rangle\right|^2 \\ &= \frac{1}{2}e^{-\Gamma t}|A_{CP}|^2 \\ &\cdot\left[1-\cos(\Delta mt) + 2\mathfrak{I}(\lambda_{CP})\sin(\Delta mt) + (1+\cos(\Delta mt))\,|\lambda_{CP}|^2\right]. \end{aligned} \tag{2.75}$$

If there is CP violation by interference of mixing and decay, a difference between the above written decay rates can be measured. Though, these effect is expected to be small and thus the asymmetry between these decays is more useful in practice, since many systematic errors cancel each other reciprocally in this measurement. This time-dependent CP asymmetry between decay rates is defined as

$$\begin{aligned} a_{CP}(t) &\equiv \frac{\Gamma(\bar{B}^0(t)\to f_{CP}) - \Gamma(B^0(t)\to f_{CP})}{\Gamma(\bar{B}^0(t)\to f_{CP}) + \Gamma(B^0(t)\to f_{CP})} \\ &= \frac{(|\lambda_{CP}|^2-1)\cos(\Delta mt) + 2\mathfrak{I}(\lambda_{CP})\sin(\Delta mt)}{1+|\lambda_{CP}|^2} \\ &= \mathcal{A}_{CP}\cos(\Delta mt) + \mathcal{S}_{CP}\sin(\Delta mt) \end{aligned} \tag{2.76}$$

where the CP parameters

$$\mathcal{A}_{CP} \equiv \frac{|\lambda_{CP}|^2 - 1}{|\lambda_{CP}|^2 + 1} \quad \text{and} \quad \mathcal{S}_{CP} \equiv \frac{2\Im(\lambda_{CP})}{|\lambda_{CP}|^2 + 1} \tag{2.77}$$

have been introduced. Since $|\lambda_{CP}| = |\bar{A}_{CP}/A_{CP}|$ in this case, the $\mathcal{A}_{CP}$ parameter is corresponds to direct violation of CP symmetry, i.e. $\mathcal{A}_{CP} \neq 0$ if $|\lambda_{CP}| \neq 1$. Respectively, $\mathcal{S}_{CP}$ is related to mixing-induced CP violation, viz. to CP violating effects by interference between mixing and decay.
The task is then to measure the time-dependent rates of B^0 and $\bar{B}^0$ decaying to f_{CP}, which has represented a challenge for experimentalists, since high precision and rare decays are both required.

2.4.4 Time Dependent *CP* Violation Measurement

Since B mesons are relatively heavy, there is a large variety of possible decay modes. At the same time, those decay modes which are usable for CP violation studies are very rare with generally small branching ratios. For significant measurements of CP violation a huge amount of B meson events $\backsim \mathcal{O}(10^8)$ is thus required [34]. Such an amount of events, measured simultaneously with high precision, can be produced by so called B-factories. The KEKB accelerator at KEK in Japan is an example of such a machine.
To create B mesons with low background, i.e. with a low amount of not desired particles, $B\bar{B}$ pairs are produced at the $\Upsilon(4s)$ resonance in an e^+e^- collider. As it can be seen in Fig. 2.10, the $\Upsilon(4s)$ [35], the fourth radially excited ${}^3S_1$ $b\bar{b}$ bound state, is the first hadronic $b\bar{b}$ resonance above the $B\bar{B}$ production threshold. This is the lowest energy at which $B\bar{B}$ pairs can be produced and for this reason they are essentially at rest in the center of mass frame. $\Upsilon(4s)$ decays to more than 95% into $B\bar{B}$ mesons and roughly equally into charged and neutral pairs, namely $B^+B^-(51.6 \pm 0.6)\%$ and $B^0\bar{B}^0(48.4 \pm 0.6)\%$ respectively [22].
The quantum numbers of $\Upsilon(4s)$ can be also written as $J^{PC} = 1^{--}$. Since P and C are conserved in strong interactions, the quantum numbers of the $B\bar{B}$ pair must be the same. Each B meson is a pseudo-scalar particle with $J^P = 0^-$. Therefore, the $B\bar{B}$ system has to be in a p-wave configuration, viz. with orbital angular momentum $L = 1$, in order to conserve the total angular momentum J.
On the other hand, the wave function of $\Upsilon(4s)$ is anti-symmetric and, accordingly, the state of the $B\bar{B}$ system has to be anti-symmetric too. This condition forbids the states B^0B^0 and $\bar{B}^0\bar{B}^0$, which must be symmetric by reason of Bose-Einstein statistics. If the two B mesons in the system are allowed to oscillate independently, there is a non vanishing probability to find the forbidden B^0B^0 and $\bar{B}^0\bar{B}^0$ at a given time. In consequence, the $B^0\bar{B}^0$ pair must oscillate coherently [31]. The quantum mechanical entangled state is given by

$$|B_1(t_1), B_2(t_2)\rangle = \frac{1}{2}\left(|B_1^0(t_1)\rangle|\bar{B}_2^0(t_2)\rangle - |\bar{B}_1^0(t_1)\rangle|B_2^0(t_2)\rangle\right). \tag{2.78}$$

The coherence of the $B^0\bar{B}^0$ pair prevails until one of the two mesons decays at a time t_1. After this, the other one is free to oscillate up to its decay at a time t_2. Thus, it is better

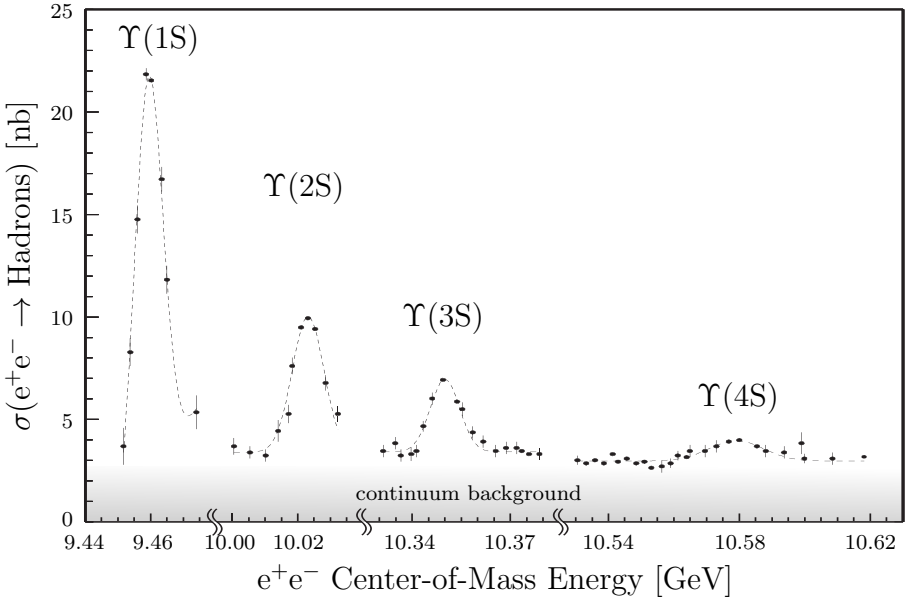


Figure 2.10: Hadronic cross section for e^+e^- collisions as a function of the e^+e^- center-of-mass energy in the region at and above the $\Upsilon(1S)$ resonance [35].

to use the time difference $\Delta t = t_2 - t_1$. Substitution with equations 2.65, 2.66 yields

$$\begin{aligned}
|B_1(t_1), B_2(t_2)\rangle =& \frac{1}{2\sqrt{2}} e^{-\frac{\Gamma}{2}(t_1+t_2)} \Bigg[\left(1 + e^{-i\Delta m \Delta t}\right) |B_1^0\rangle|\bar{B}_2^0\rangle + \frac{q}{p}\left(1 - e^{-i\Delta m \Delta t}\right) |\bar{B}_1^0\rangle|B_2^0\rangle \\
& - \frac{p}{q}\left(1 - e^{-i\Delta m \Delta t}\right) |B_1^0\rangle|\bar{B}_2^0\rangle - \left(1 + e^{-i\Delta m \Delta t}\right) |\bar{B}_1^0\rangle|B_2^0\rangle \Bigg] \\
=& \frac{1}{2\sqrt{2}} e^{-\frac{\Gamma}{2}(t_1+t_2)} \Bigg[\left(1 + e^{-i\Delta m \Delta t}\right) \left(|B_1^0\rangle|\bar{B}_2^0\rangle - |\bar{B}_1^0\rangle|B_2^0\rangle\right) \\
& + \left(1 - e^{-i\Delta m \Delta t}\right) \left(\frac{q}{p}|\bar{B}_1^0\rangle|B_2^0\rangle - \frac{p}{q}|B_1^0\rangle|\bar{B}_2^0\rangle\right) \Bigg]
\end{aligned} \tag{2.79}$$

If the flavour of one of the two B's can be determined by a flavour specific decay, the flavour and the time evolution of the other B is known with exactitude. In the special case that the flavour of one B meson is tagged by a specific flavour decay, B_{tag}, and the other

B decays into a CP eigenstate, B_{CP}, the time dependent decay rates can be written as

$$\begin{aligned}\Gamma(f_{CP}, f_{\text{tag}}) =& |\langle f_{CP}, f_{\text{tag}}|H|B_{CP}(t_{CP}), B_{\text{tag}}(t_{\text{tag}})\rangle|^2 \\ =& \frac{1}{4} e^{-\Gamma(t_{CP}+t_{\text{tag}})} |A_{CP}|^2 |A_{\text{tag}}|^2 \\ & \cdot \left[1 + \cos(\Delta m \Delta t) - 2\Im(\lambda_{CP}) \sin(\Delta m \Delta t) + (1 - \cos(\Delta m \Delta t)) \, |\lambda_{CP}|^2\right], \end{aligned} \tag{2.80}$$

$$\begin{aligned}\Gamma(f_{CP}, \bar{f}_{\text{tag}}) =& \left| \langle f_{CP}, \bar{f}_{\text{tag}}|H|B_{CP}(t_{CP}), B_{\text{tag}}(t_{\text{tag}})\rangle \right|^2 \\ =& \frac{1}{4} e^{-\Gamma(t_{CP}+t_{\text{tag}})} |A_{CP}|^2 |A_{\text{tag}}|^2 \left| \frac{p}{q} \right|^2 \\ & \cdot \left[1 + \cos(\Delta m \Delta t) + 2\Im(\lambda_{CP}) \sin(\Delta m \Delta t) + (1 - \cos(\Delta m \Delta t)) \, |\lambda_{CP}|^2\right]. \end{aligned} \tag{2.81}$$

introducing the decay amplitude $A_{\text{tag}} = \langle f_{\text{tag}}|H|B^0\rangle = \langle \bar{f}_{\text{tag}}|H|\bar{B}^0\rangle$. The respective time-dependent CP asymmetry results in

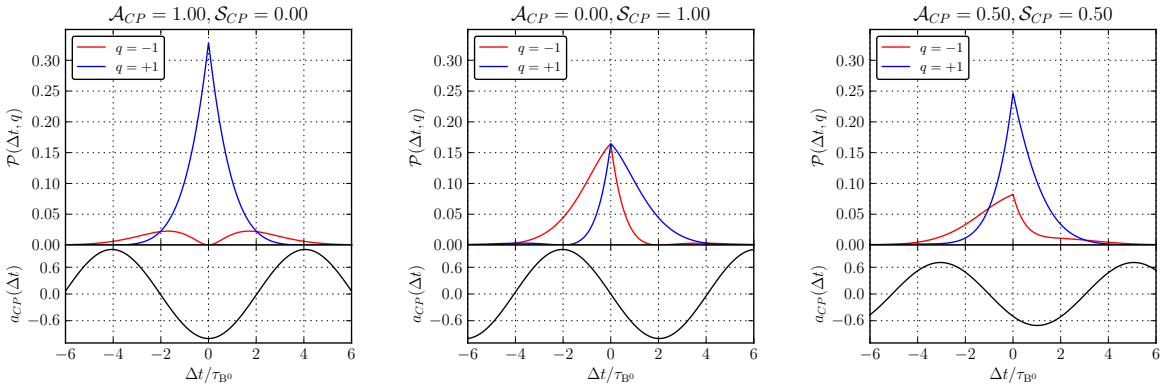
$$\begin{aligned} a_{CP}(t) \equiv & \frac{\Gamma(f_{CP}, \bar{f}_{\text{tag}}) - \Gamma(f_{CP}, f_{\text{tag}})}{f_{CP}, \bar{f}_{\text{tag}}) + \Gamma(f_{CP}, f_{\text{tag}})} \\ = & \mathcal{A}_{CP} \cos(\Delta m \Delta t) + \mathcal{S}_{CP} \sin(\Delta m \Delta t) \end{aligned} \tag{2.82}$$

conserving the same form as in equation 2.76 but in Δt. Therefore, the measurement of the CP violating parameters at a B Factory includes three ingredients: flavour tagging of B_{tag}, reconstruction of the B_{CP} decay to a specific f_{CP} final state, and measurement of their decay time difference Δt.
The probability of observing f_{CP} for a given flavour $q = +1(-1)$ of $B_{\text{tag}} = B^0(\bar{B}^0)$ at time Δt can be obtained normalizing equations 2.80, 2.81 to unity in the region $-\infty < \Delta t < \infty$,

$$\mathcal{P}(\Delta t, q) = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \left[1 + q \left(\mathcal{A}_{CP} \cos(\Delta m \Delta t) + \mathcal{S}_{CP} \sin(\Delta m \Delta t)\right)\right] \tag{2.83}$$

with the B^0 lifetime τ_{B^0}. In Fig. 2.11 [31], this probability is plotted together with the resulting CP asymmetry a_{CP} for three different values of the CP violating parameters $\mathcal{A}_{CP}$ and $\mathcal{S}_{CP}$.

Although the lifetime of the B mesons is relatively large, it is impossible to measure its value directly as precise as needed for the study of CP violation, since a resolution in order of some pico seconds cannot be reached with current technology. A clever experimental trick in B-factories gets rid of this drawback by using asymmetric beam energies: the $\Upsilon(4s)$ system is produced with a Lorentz boost $\beta\gamma$ in the direction of the high energy beam, i.e. in positive z-direction if the convention of Belle is used. Because of the low momentum of the B mesons in the $\Upsilon(4s)$ center-of-mass frame, they fly practically parallel to the boost direction. Thus, the Lorentz boost amplifies the separation Δz between decay vertices and



(a) Direct *CP* violation (b) Mixing-Induced *CP* violation (c) Combination of a) and b)

Figure 2.11: Three different examples how *CP* violation can manifest in the probability $\mathcal{P}(\Delta t, q)$ and the related *CP* asymmetry a_{CP}. No *CP* violation means no asymmetry, i.e. perfect matching of the distributions for $q = +1$ and $q = -1$ [31].

makes it detectable. The proper decay time difference is then,

$$\Delta t = \frac{\Delta z}{\langle \beta\gamma \rangle c}. \tag{2.84}$$

For a boost of $\beta\gamma \simeq 0.284$ in SuperKEKB [6], the mean flight distance of the B^0 amounts only 130 µm (S. Fig. 2.12). Such a spatial resolution for vertexing can be only achieved using high precision vertex detectors like those integrated in Belle II .

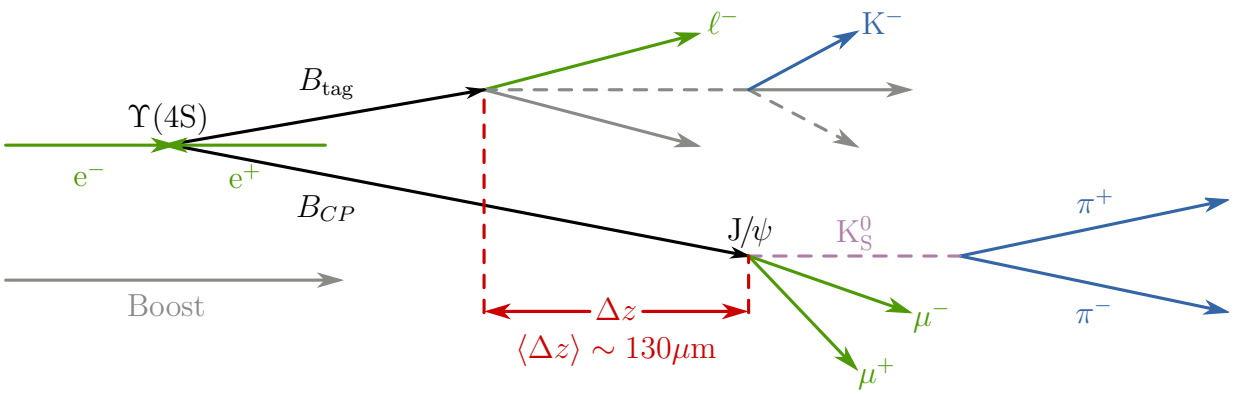


Figure 2.12: Simple schematic of the $B^0\overline{B}^0$ decay in the case of the golden channel $J/\psi K^0_S$. The electric charge of the high momentum lepton l^- reveals the flavour of B_{tag} [31].

2.5 Achievements of Belle

The Belle detector was designed and optimized in order to observe time dependent *CP* violation in the B meson system running at the KEK e^+e^- collider. Nevertheless, its excellent detector capabilities and its record breaking high luminosity allowed Belle not only to make important discoveries in B physics, apart from the *CP* violation measurements, but also in c-quark physics, τ-lepton physics, hadron spectroscopy and $\gamma\gamma$-photon physics. A detailed conglomeration of the accomplishments by the Belle experiment can be found in [36] and [6].
However, the measurement of time-dependent *CP* asymmetries in decay modes such as $B \to J/\psi K_S^0$ had the greatest impact on particle physics as it could provide a theoretically clean measurement of one of the unitarity triangles, ϕ_1. This results, together with those done in friendly competition by the BaBar experiment, demonstrated that the proposal done by M. Kobayashi and T. Maskawa was correct.
In the following, this measurement of the ϕ_1 angle will be briefly described, since it is the best example of the achievements of Belle and, simultaneously, it illustrates how the upgrade to Belle II can extend the actual experimental reach. Therefore, the next subsections are dedicated to two processes which can provide data to measure ϕ_1 via quark transitions of the form $\bar{b} \to \bar{q}q\bar{s}$, namely $B \to J/\psi K_S^0$ via $\bar{b} \to c\bar{c}s$ and $B \to \phi K_S^0$ via $\bar{b} \to s\bar{s}s$.

2.5.1 $\bar{b} \to c\bar{c}s$ Transitions and the Golden Chanel $B \to J/\psi K_S^0$

Neutral B meson decays into $(c\bar{c})K^0$ are induced via $b \to c\bar{c}s$ transitions. The Feynman diagrams for these decays at the tree and penguin levels are shown in Fig. 2.13. In the SM, the contributions of the penguin diagram can be neglected to an approximation that is better than 1% [22]. Since the only possible physical final states are $(c\bar{c})K_L$ or $(c\bar{c})K_S^0$ due to the neutral K meson mixing, violation of *CP* symmetry can occur by interference between B^0 meson mixing, B^0 decay and K^0 mixing (Fig. 2.14).

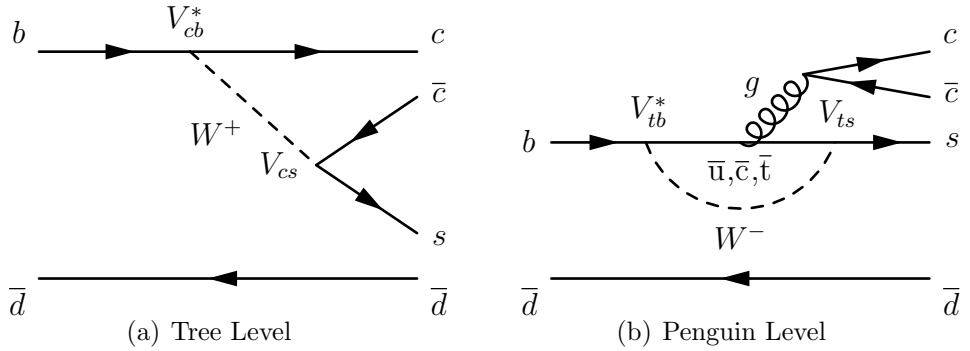


Figure 2.13: Feynman diagrams contributing to $B \to (c\bar{c})K^0$ via $b \to c\bar{c}s$. The contribution of the penguin diagram can be neglected to an approximation that is better than one percent [22].

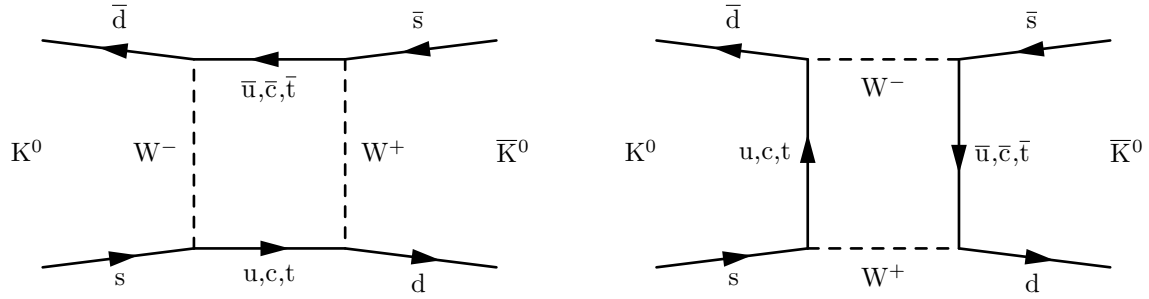


Figure 2.14: Leading Feynman diagrams contributing to neutral Kaon K^0 mixing. K^0 oscillations are dominated by c-quark contributions since $V_{cs} \simeq 1 > V_{us} > V_{ts}$.

In consequence, the *CP* violating complex quantity $\lambda_{(c\bar{c})K^0}$ becomes a product of three phases

$$\begin{aligned}\lambda_{(c\bar{c})K^0} &= -\left(\frac{q}{p}\right)_{B^0} \cdot \left(\frac{\bar{A}_{(c\bar{c})K^0}}{A_{(c\bar{c})K^0}}\right) \cdot \left(\frac{q}{p}\right)_{K^0} \simeq -\frac{V_{tb}^* V_{td}}{V_{tb} V_{td}*} \cdot \frac{V_{cb} V_{cs}^*}{V_{cb} V_{cs}^*} \cdot \frac{V_{cs} V_{cd}^*}{V_{cs} V_{cd}^*} = -\frac{V_{tb}^* V_{td} V_{cb} V_{cd}^*}{V_{tb} V_{td}^* V_{cb}^* V_{cd}} \\ &= \left(-\frac{V_{cd} V_{cb}^*}{V_{td} V_{tb}^*}\right)^{-1} \cdot (-1) \left(-\frac{V_{cd} V_{cb}^*}{V_{td} V_{tb}^*}\right)^* = \left(e^{i\phi_1}\right)^{-1} \cdot (-1) \left(e^{i\phi_1}\right)^* = -e^{-2i\phi_1} \end{aligned} \quad (2.85)$$

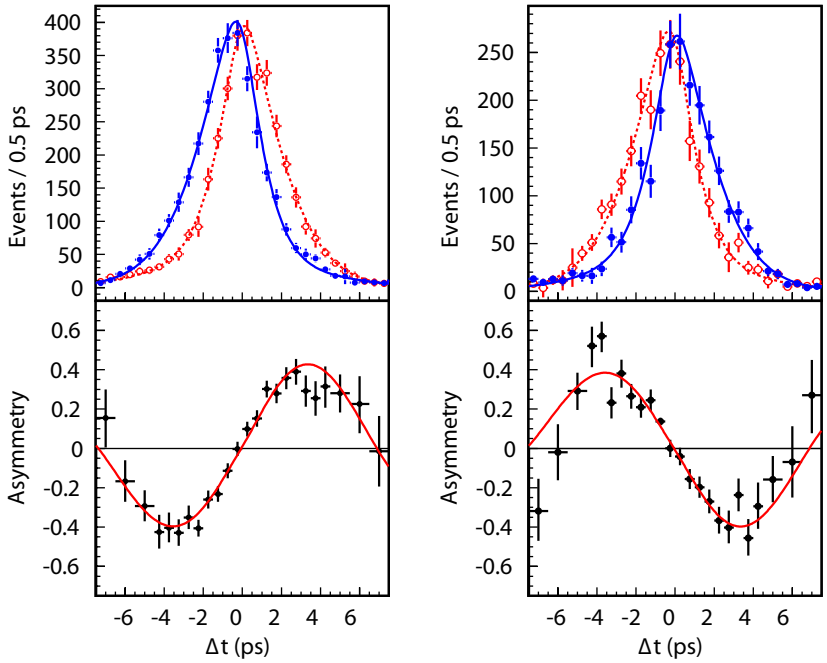
where the definition of the ϕ_1 angle 2.31 has been used. It follows then for the *CP* violating parameters [22]

$$\lambda_{CP} = -\eta_{CP} \lambda_{(c\bar{c})K^0} \quad \Rightarrow \quad \mathcal{A}_{CP} = 0, \quad \mathcal{S}_{CP} = -\eta_{CP} \sin(2\phi_1) \quad (2.86)$$

where η_{CP} is the *CP* eigenvalue of f_{CP}. In 2002 Belle observed this kind of time-dependent *CP* violation for the first time [37] after reconstruction of neutral B decays into the odd($\eta_{CP} = -1$) *CP* eigenstates $J/\psi K_S^0$, $\psi(2s)K_S^0$, $\chi_{c1}K_S^0$, $\eta_c K_S^0$ and the even($\eta_{CP} = +1$) $J/\psi K_L$. Because of the low theoretical uncertainties, the large branching fraction and the high reconstruction efficiencies in comparison with the other decay modes, $B \to J/\psi K_S^0$ has been called the "Golden Channel" for the measurement of time dependent *CP* violation [38].
As in Fig. 2.12[31], the flavour of B_{tag} is identified using semileptonic measurements: by charge conservation if B_{tag} emits a single lepton carrying positive electric charge then $B_{\text{tag}} = B^0$, otherwise if the lepton is negative $B_{\text{tag}} = \bar{B}^0$. The actual values obtained by the Belle collaboration for the *CP* violating parameters are [39]

$$\mathcal{A}_{CP}^{(c\bar{c})K^0} = 0.006 \pm 0.016(\text{stat}) \pm 0.012(\text{sys}),\ \sin(2\phi_1)^{(c\bar{c})K^0} = 0.667 \pm 0.023(\text{stat}) \pm 0.012(\text{sys}). \quad (2.87)$$

The corresponding plots for $a_{CP}^{(c\bar{c})K^0}(\Delta t)$ and $\mathcal{P}^{(c\bar{c})K^0}(\Delta t, q)$ can be found in Fig. 2.15[39].



(a) All CP modes combined (b) CP even mode $J/\psi K_L^0$

Figure 2.15: In the top row: background-substracted Δt event distribution for $q = +1$ (red) and $q = -1$ (blue). In the bottom row: $a_{CP}(\Delta t)$. Both for good tag quality events [39].

2.5.2 $\bar{b} \to s\bar{s}s$ Transitions and $B \to \phi K_S^0$

Neutral B meson decays into $(s\bar{s})K^0$ are only possible via $b \to s\bar{s}s$ transitions through loop processes like that in the penguin diagram showed in Fig. 2.16. Violation of CP asymmetry can occur again by interference between B^0 mixing, B^0 decay and K^0 mixing. In the SM, the subdominant u and t quark contributions to the decay amplitude $A_{(s\bar{s})K^0}$ can be neglected to an approximation that is good on the order of a few percent [22]. The CP violating complex $\lambda^{(s\bar{s})K^0}$ quantity becomes then

$$\lambda_{(c\bar{c})K^0} = -\left(\frac{q}{p}\right)_{B^0} \cdot \left(\frac{\bar{A}_{(s\bar{s})K^0}}{A_{(s\bar{s})K^0}}\right) \cdot \left(\frac{q}{p}\right)_{K^0} \simeq -\frac{V_{tb}^* V_{td}}{V_{tb} V_{td}*} \cdot \frac{V_{cb} V_{cs}^*}{V_{cb} V_{cs}^*} \cdot \frac{V_{cs} V_{cd}^*}{V_{cs} V_{cd}^*} = -e^{-2i\phi_1} \quad (2.88)$$

which is the same approximate expression for $\lambda_{(c\bar{c})K^0}$. The CP violating parameters can be thus written again as

$$\lambda_{CP} = -\eta_{CP} \lambda_{(s\bar{s})K^0} \quad \Rightarrow \quad \mathcal{A}_{CP} = 0, \quad \mathcal{S}_{CP} = -\eta_{CP} \sin(2\phi_1) \quad (2.89)$$

with the η_{CP} eigenvalue. This kind of time-dependent CP violation was measured by Belle for the first time in 2003 after reconstructing neutral B decays into $K^+K^-K_S^0$ [40]. The final state $K^+K^-K_S^0$ is primarily even, but it is not a pure CP eigenstate. However, there are several intermediate resonant contributions to this final state; some of them

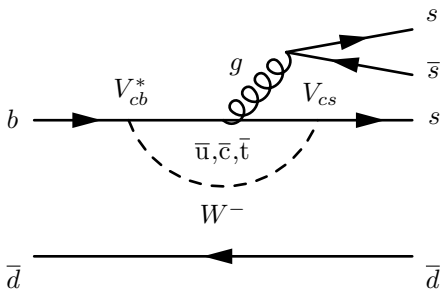
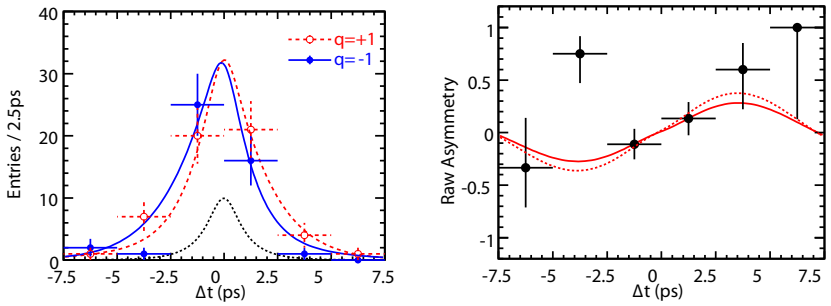


Figure 2.16: Feynman diagram for the penguin amplitude contributing to $B \to (s\bar{s})K^0$. The process sub-dominant contributions can be neglected to a good approximation that is better than a few percent [22].

are CP eigenstates like the odd ϕK_S, $\eta' K_S$ and the even $f_0 K_S$ [41]. With a so called time-dependent Square Dalitz plot analysis [42, 43], it is possible to measure CP violating parameters for neutral B decays into the intermediate CP eigenstates. Among them, the highest reconstruction efficiency and the lowest systematic uncertainties are reached for the decay channel $B \to \phi K_S^0$ [40, 44]. The actual values for the CP violating parameters obtained by the Belle collaboration amount [44]

$$\mathcal{A}_{CP}^{\phi K_S^0} = -0.04 \pm 0.20(\text{stat}) \pm 0.10(\text{sys}),\ \sin(2\phi_1)^{\phi K_S^0} = 0.90 \pm 0.31(\text{stat}) \pm 0.09(\text{sys}). \quad (2.90)$$

The corresponding plots for $a_{CP}^{\phi K_S^0}(\Delta t)$ and $\mathcal{P}^{\phi K_S^0}(\Delta t, q)$ can be found in Fig. 2.17[44].



(a) Δt event distribution (b) $a_{CP}(\Delta t)$

Figure 2.17: Measurement results for $B^0 \to K^+K^-K_S^0$ candidates in the ϕK_S^0 region, i.e. $|m_{K^+K^-} - m_\phi| < 0.01\text{GeV}/c^2$, with good tags. In (a) the blue solid and red dashed curves show the fitted results with B^0 and $\bar{B}^0$ tags, respectively. The dotted black curve shows the backbround component with B^0 and $\bar{B}^0$ tags. In (b), the solid curve shows the fit projection and the dashed curve shows the SM expectation from the time-dependent CP asymmetry measurement in decays via transitions $b \to (c\bar{c})s$ [44].

2.6 New Physics in Belle2

The new detector Belle II will improve the performance of its predecessor Belle and provide larger data sets. Very remarkable are the new high precision pixel detector (PXD) in the innermost part of the Belle II detector and the expected luminosity of the SuperKEKB factory nearly 40 times higher than that reached by KEKB. Because of its unprecedented sensitivity, the Belle II experiment is expected to make a tremendous contribution to the search of NP effects at the rare and high precision frontier. There is a broad physics program with the purpose to bring the search of NP, performed in B-factories, to a new level of sensitivity: the SuperKEKB factory, together with the Belle II detector, will be able to execute measurements not only related to B physics and CP violation in the B meson sector, but also measurements in all fields of heavy flavour physics, τ-lepton physics, c-quark physics, spectroscopy, and pure electroweak processes. The vast amount of planned measurements will over-constrain the SM parameter space as well as its extensions. Complementary, Belle II measurement results should help at the identification of NP signatures that may be directly observed at high energy frontier experiments.
In this subsection only one example of the large number of NP sensitive channels at the rare and high precision frontier will be treated, namely the measurement of ϕ_1 through decays via $b \to s\bar{s}s$ transitions. A comprehensive study of expected NP output related to observables and sensitivities in Belle II can be found in [45].

2.6.1 New Physics in $\bar{b} \to s\bar{s}s$ Decays

Comparing the values for $\sin(\phi_1)$ in 2.87 and 2.90, a slight difference between both values can be recognized within 1σ. Although $\sin(\phi_1)$ is extracted in both cases doing approximations from the respectives CP violating quantities $\lambda_{(c\bar{c})K^0}$ and $\lambda_{\phi K^0_S}$, in the SM, the difference

$$\Delta S(\phi K^0_S) \equiv \sin(2\phi_1)^{\phi K^0_S} - \sin(2\phi_1)^{(c\bar{c})K^0} \tag{2.91}$$

is expected to disappear within small corrections, i.e $\Delta S(\phi K^0_S) = 0.03^{+0.01}_{-0.04} \pm 0.01$ [46].
Since the irreducible CP violating phase δ in the SM can not alone explain the origin of the matter-antimatter asymmetry of the universe, additional CP violating phases are unavoidable in most of the NP theories [45]. In the presence of NP, decay amplitudes for processes via $b \to c\bar{c}s$ transitions are likely to remain being dominated by contributions at the tree level, but with small corrections for the mixing amplitudes [22]. Otherwise decays trough processes via $b \to s\bar{s}s$ transitions are only possible at the penguin level; NP particles in the loop of the penguin diagram would make new contributions to the decay amplitudes of such decays and, thus, change the expectation value of $\lambda_{\phi K^0_S}$ by an extra CP violating contribution. An example of a possible contribution by a new NP virtual particle is shown in Fig. 2.18.
Nevertheless, the value of $\sin(2\phi_1)^{\phi K^0_S}$ is still statistically limited with a relative wide uncertainty. To perform a more precise Dalitz plot analysis, more data as well as higher accuracy in vertex determination and particle identification are required. These will be

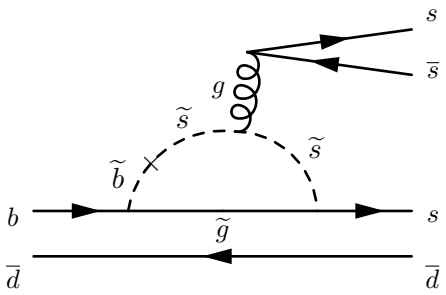


Figure 2.18: A super symmetric gluino-down squark contribution to the $b \to s\bar{s}s$ transition [6].

achieved by the new PXD and the silicon vertex detector (SVD) in Belle II together with its particle identification system; the greater sensitivity of Belle II will allow to discern if NP effects are present in the corresponding penguin modes [6].
In Fig. 2.19[45] some ΔS differences in dependence on the integrated luminosity are plotted by extrapolating the present experimental results of Belle.

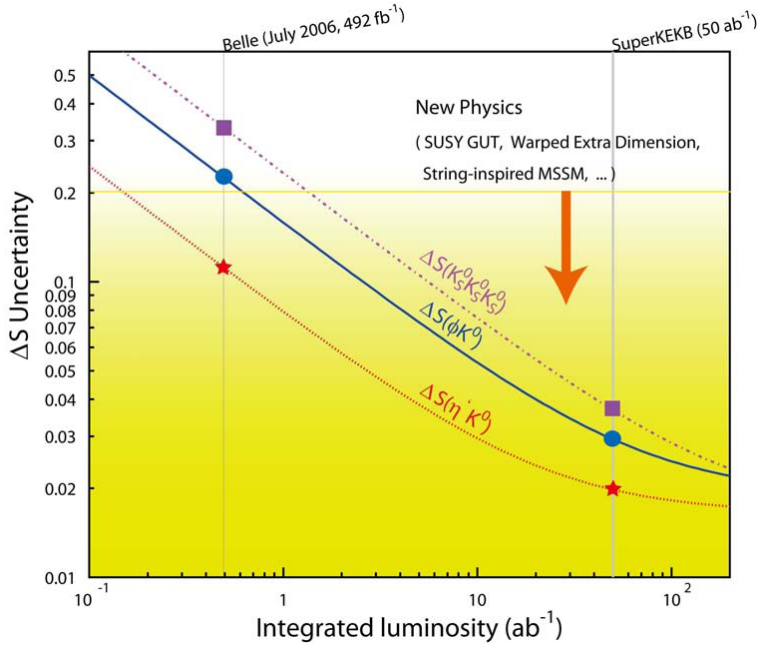


Figure 2.19: Expected total errors on ΔS for different intermediate CP resonances of $B^0 \to K^+K^-K^0_S$ as functions of the integrated luminosity [45].

Chapter 3

The Belle II Experiment

The KEKB accelerator complex, together with its Belle detector, was operating from 1999 to 2010. After a decade of running, the tremendous success of the measurement results, which confirmed amongst other achievements the theoretical predictions of the SM related to *CP* violation, motivated the actual upgrading process. In 2016 the new Belle II detector at SuperKEKB is scheduled to start taking data.
The higher Luminosity of SuperKEKB will be maximally exploited with the new pixel vertex detector in Belle II; the physics event rate is predicted to be 50 times larger [6]. With much larger data samples the planned measurements are expected to provide a new insight in the heavy flavour physics research. Nevertheless, there will be also a huge amount of data produced by the Belle II detector caused by undesired physical processes, which are referred to as background processes. With the higher currents and the smaller beam sizes in SuperKEKB, the background in Belle II will considerably increase that of Belle.
Within the scope of this work, the relevant attributes of the new Belle II detector are those related to the triggering task. Since the detector is not operating right now, event simulation is required to study the features of the proposed neural z-vertex trigger, which will be introduced in the next chapter.
Therefore, a brief description of the new SupperKEKB and Belle II together with the most important aspects of the event simulation will be presented in this chapter. In addition, different sources of background processes will be examined.

3.1 The SuperKEKB Accelerator and the IR design

As argued in subsection 2.4.4, the SuperKEKB factory uses asymmetric e^+e^- beam energies. The accelerator will be placed in the same tunnel as KEKB in the locality of Tsukuba, Japan. The positrons are accelerated up to 4 GeV in the so called Low Energy Ring (LER) and the electrons up to 7 GeV in the High Energy Ring (HER). The two beams intersect each other at the so-called Interaction Point (IP) by a relatively large crossing angle $2\phi = 83$ mrad$\sim 4.76°$ (about 4 times larger than in KEKB). The ± 2 m region around the IP is denominated as the interaction region (IR).

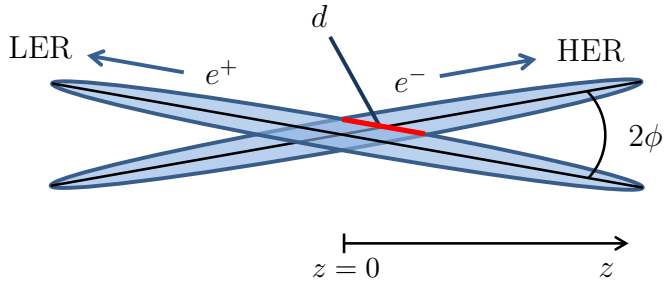


Figure 3.1: Schematic view of the beam intersection in the Nano Beam scheme.

The particle beams are characterized by several parameters. If an ideal particle in one of the beams is considered, it follows a particular trajectory, which closes on itself after one revolution inside the accelerator ring. This ideal trajectory is referred to as the closed orbit. In reality, beam particles undergo oscillations around this closed orbit. The emittance ε measures how much the particles depart from the ideal closed trajectory and it depends on the initial beam conditions. The so-called β function can be thought as an envelope around all the possible particle trajectories inside the beam. The beam size in horizontal and vertical direction can be then written as

$$\sigma_{x,y}(s) = \sqrt{\varepsilon_{x,y}\beta_{x,y}(s)} \tag{3.1}$$

where s is a point along the ideal closed orbit. The β function depends also strongly on the quadrupole magnets. Since a single quadrupole magnet focuses the beam in one direction and defocuses it in the other, doublets of quadrupole magnets are needed to focus the beam in vertical and horizontal direction. The beam deflection inside the accelerator rings is done by dipole magnets.
The Luminosity of a collider at the IP can be expressed by [34]

$$L = \frac{\gamma_\pm}{2er_e}\left(1 + \frac{\sigma_y(s=\mathrm{IP})}{\sigma_x(s=\mathrm{IP})}\right)\left(\frac{I_\pm \xi_\pm}{\beta_y(s=\mathrm{IP})}\right)\left(\frac{R_L}{R_{\xi_y}}\right) \tag{3.2}$$

where $\gamma_\pm$ is the Lorentz factor, r_e is the classical electron radius, e is the electric charge, $I_\pm$ are the beam currents and $\xi_\pm$ denotes the so-called beam-beam parameters. The suffix $\pm$ denotes the positron $(+)$ or the electron $(-)$ beam. The parameter R_L is a reduction factor for the luminosity and R_{ξ_y} is a reduction factor for the beam-beam parameter, which stem from the crossing angle and the so-called "hourglass effect" [47].
The main new feature, in order to achieve the design luminosity of $8 \cdot 10^{35}$ cm^{-2}s^{-1}, is the upgrade based on the so-called "Nano Beam" scheme [48], which was first proposed for the Super B factory in Italy [49]. The concept of the nano beam scheme consists in reducing maximally the overlap d of the beams at the IP by squeezing the vertical β_y function of the beams (S. Fig. 3.1). New doublets of superconducting quadrupole magnets close to the IP will allow the squeeze of β_y. To obtain the design value of ξ_y, also extremely small horizontal and vertical beam sizes are required. The most relevant machine parameters of SuperKEKB in comparison with those in KEKB are shown in Table 3.1[6].

		KEKB Achieved	SuperKEKB
Energy(GeV) (LER/HER)	E	3.5/8.0	4.0/7.0
Vertical Beta Function at IP(mm)	$\beta_y(IP)$	5.9/5.9	0.27/0.41
Beam-Beam Parameter	ξ_y	0.129/0.090	0.090/0.088
Horizontal Beamsize at IP(µm)	$\sigma_x(IP)$	10.2/11.2	10.2/7.75
Vertical Beamsize at IP(nm)	$\sigma_y(IP)$	48.3/61.8	59/59
Beam current (A)	I	1.64/1.19	3.60/2.62
Luminosity $10^{34}cm^{-2}s^{-1}$	L	2.11	80

Table 3.1: Fundamental parameters of SuperKEKB and present KEKB [6].

With the design luminosity in SuperKEKB and the total cross section $\sigma_{B\bar{B}} = 1.2$ nb from the process $\Upsilon(4S) \to B\bar{B}$, the expected number of $B\bar{B}$ events

$$N_{B\bar{B}} = \sigma_{B\bar{B}} \int Ldt \tag{3.3}$$

will amount approximately $30 \cdot 10^9$ around 2021-2022 [6, 34]. Although the nano-beam scheme allows to achieve the design luminosity, it shortens the Touschek beam lifetime and produces a growth of the beam emittance. The Touschek scattering is mitigated by the increase of the beam energy in the LER. This kind of scattering phenomenon will be examined in subsection 3.3.1. On the other hand, the decrease of the beam energy in the HER will help to maintain a lower emittance. Nevertheless, the reduction of the beam energy asymmetry will lead to a smaller Lorentz boost $\gamma\beta = 0.284$, about $\sim 2/3$ that of KEKB, augmenting the resolution exigencies on the vertex detector. Therefore, the Belle II detector has been designed not only to maximally exploit the high luminosity, but also to fulfill the high vertexing requirements.

3.2 The Belle II Detector

The Belle II detector will be an upgraded version of the Belle detector, but retaining the basic layout. The most remarkable change in the upgrade process, will be the introduction of a completely new pixel vertex detector (PXD) in the innermost part of the vertex detector. The preexisting silicon vertex detector (SVD) will be also renewed, in order to free space for the new PXD. In consequence, the adjacent central drift chamber (CDC), responsible for the tracking of charged particles, will be extended to a larger radius while being subjected to other changes for the sake of a higher momentum resolution.
The particle ID systems TOF and ACC used in Belle will be replaced by a Time Propagation Counter (TOP) in the barrel region and an Aerogel Ring-Imaging Čereknov (ARICH) detector device in the forward endcap. These are enclosed within the Electromagnetic Calorimeter (ECL), the 1.5 T superconducting solenoid and the K_L^0/μ detector (KLM). This detector parts, the ECL, the solenoid and the KLM will be reused under some significant upgrades. An overview of all sub-detectors is visualized in Fig. 3.2[31].

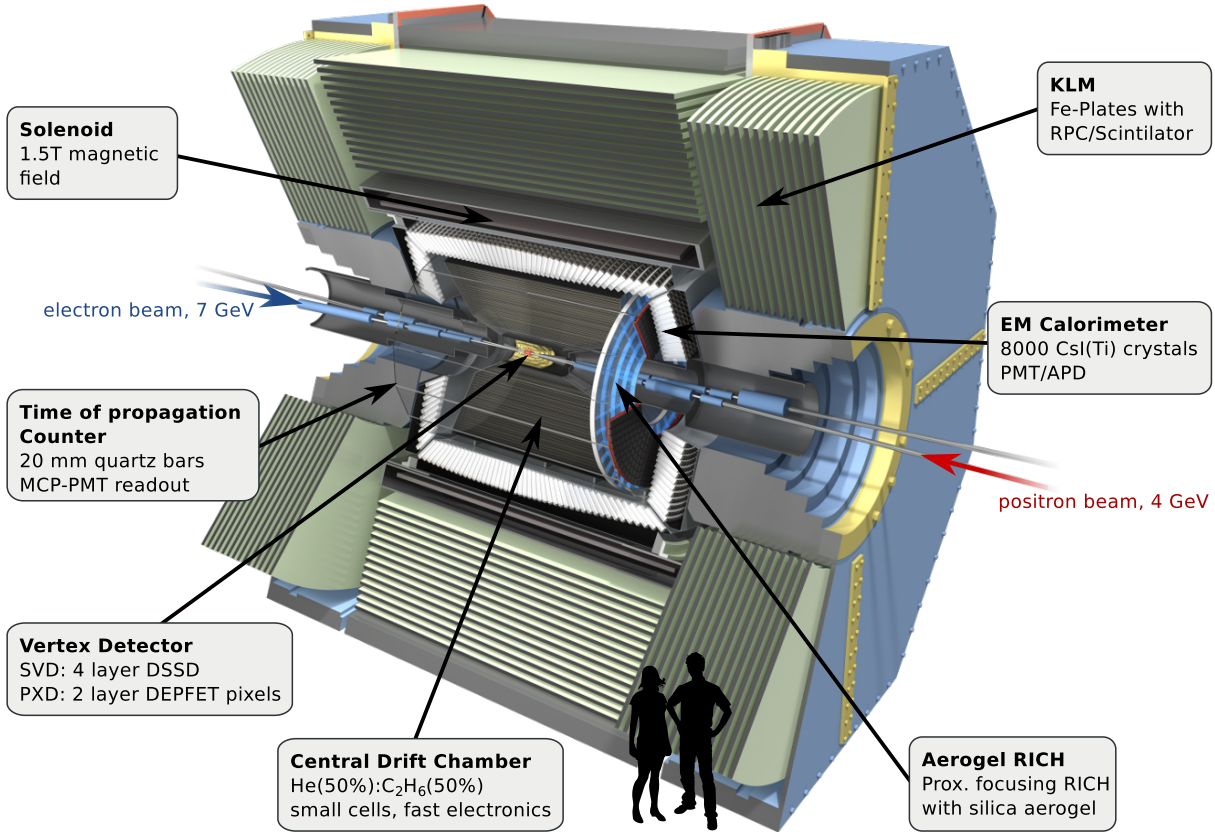
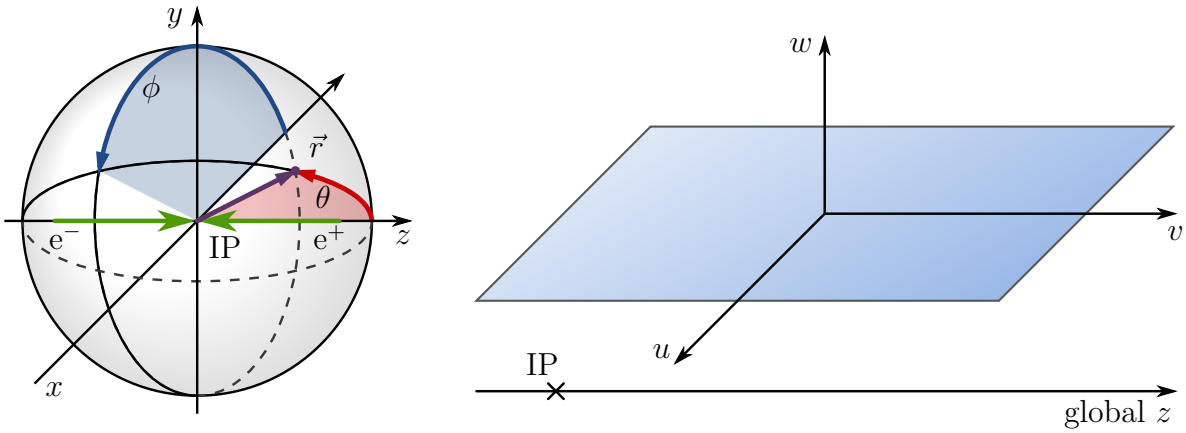


Figure 3.2: Overview of the Belle II detector [31].

In the following, the focus will be set on the vertex detectors, which are the most interesting detector part for the NP research motivated in the first chapter, and the CDC together with the Belle II trigger system, which generate the input data for the presented studies in this work. The description of some detector properties including the vertex detectors has been done on the basis of the PHD thesis of M. Ritter [31]. A comprehensive description of the new features and the upgrading process of the SuperKEKB accelerator ring as well as of all Belle II detector components can be found in the Belle II Technical Design Report [6].

3.2.1 The Coordinate System

The global coordinate system of Belle II is defined as a right-handed (x, y, z)-cartesian coordinate system with its origin situated at the nominal IP. The z-axis lies parallel to the magnet field of the solenoid. As mentioned before, the positive z-direction is the flight direction of the high energy beam; $z > 0$ denotes the forward side of the detector and $z < 0$ the backward side. The y-axis is the vertical coordinate pointing upwards while the x-axis is the horizontal coordinate, placed on the accelerator plane and pointing away from the



(a) Global coordinates (b) Local coordinates

Figure 3.3: Belle II coordinate system [31].

center of the accelerator ring.
For the definition of the transverse $\mathbf{p}_T$ momentum and the curvature κ of charged particle tracks, which will be introduced in subsection 3.5, it is useful to define the radial distance $r = \sqrt{x^2 + y^2}$ to the origin, projected on the xy-plane. Similarly, the polar angle θ to the z-axis and the azimuthal angle ϕ with respect to the positive x-axis can be introduced. In turn, the detector acceptance is

$$17° \leq \theta \leq 150° \tag{3.4}$$

what means, that particles flying outside the acceptance region will not be detected.
For the PXD and SVD sensors a local coordinate system is defined. By convention, it is a right-handed coordinate system with axes (u, v, w), where u points along the global $r\phi$ direction, v points in the direction of the global z-axis and w points away from the IP. A visualization of global and local coordinate systems is shown in Fig. 3.3[31].

3.2.2 Beampipe

The beampipe in Belle II has a similar design of that used in Belle. It consists of two layers of Beryllium cylinders with a cooling gap between them. The inner side of the beam pipe is plated with a 10µm gold layer to shield the detector from low energy X-ray background. The inner radius of the beampipe is 10 mm and the outer radius 12 mm.

3.2.3 Pixel Detector PXD

As mentioned before, the vertex pixel detector is the completely new detector component in Belle II. It consists of two layers of pixel sensors based on Depleted Field-Effect Transistor (DEPFET) technology [50, 51].

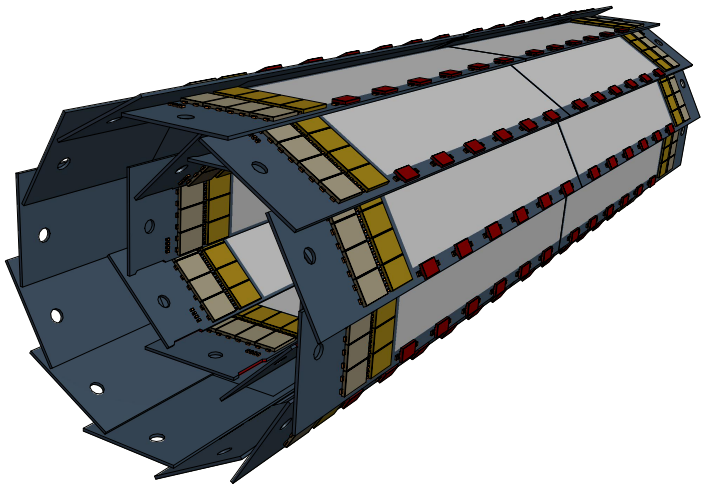


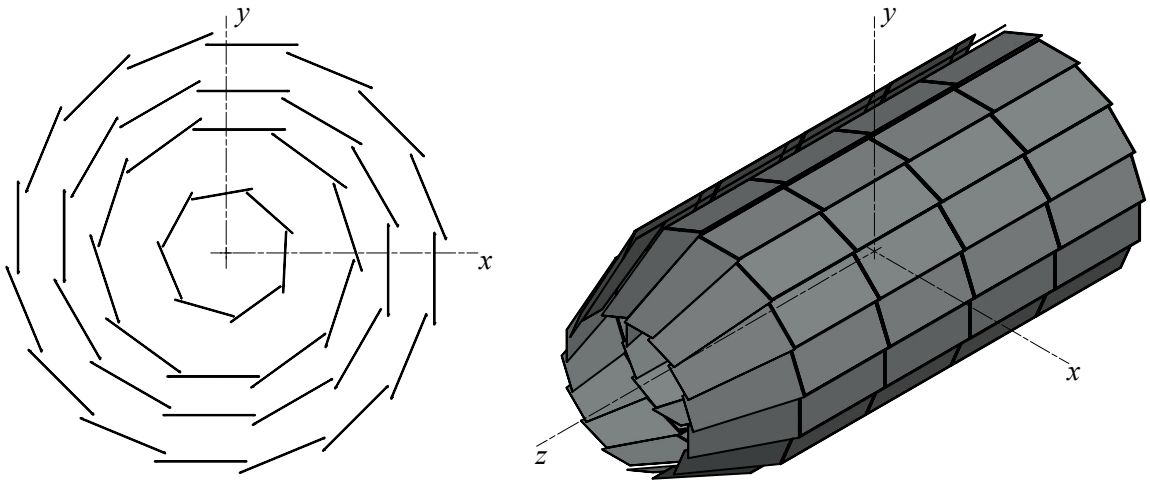
Figure 3.4: Schematic view of the geometrical arrangement of the sensors for the PXD. The light gray surfaces are the sensitive areas of the DEPFET sensors; a ladder is conformed by tow jointed sensors. The yellow structures are the readout chips. The red structures are switcher chips to control the readout process [31].

The inner PXD layer has a radius of 14 mm and the outer layer 22 mm. Fig. 3.4[31] shows a scheme of the geometrical sensor arrangement.
The extremely high luminosity, the Nano Beam scheme and the new beam parameters in SuperKEKB are good news for physics concerning vertex reconstruction. Nonetheless, for a high precision vertex detector placed close to the beam pipe, i.e. at radii between 14-40 mm, the background due to the new features is expected to cause very high hit rates. Consequently, there is no possibility to build the innermost layers of the vertex detector based on silicon strips; the large strip occupancy because of the high hit rates would make it impossible to reconstruct B-decays. The new PXD is the solution to this problem: pixel sensors have a notably larger number of channels and thus a much smaller ocupancy.

The PXD is made of 40 sensors with a total amount of 7.68 million pixels. The DEPFET pixel sensors can be built very thin, due to their internal signal amplification [6]. In $r\phi$ direction, the pixel size is 50 µm for both layers. In z direction, the sensors are segmented into two pieces. The size of the pixels in z direction varies from 55 µm to 85 µm; smaller pixels are placed closer to the detector center. The different pixel sizes together with other PXD properties can be found in Table A.1[31] of Appendix A. The resulting resolution of the PXD detector fulfills, therefore, the vertexing resolution requirements($\sim$ 130 µm) set in subsection 2.4.4.

3.2.4 Silicon Vertex Detector SVD

The silicon vertex detector (SVD) is the second part of the vertex detector in Belle II. As argued before, the main task of the first detector part, the PXD, is to measure the separation between B and $\bar{B}$ decay vertices for the study of mixing-induced CP violation. Complementary to the PXD, the SVD is able to measure vertex information of other interesting decay channels involving D-mesons and τ-lepton decays [6]. The SVD will also provide the data needed to extrapolate the tracks detected in the CDC to the PXD, in order to improve the vertex reconstruction.
The Belle II SVD design shares the good characteristics of the SVD in Belle. It is a low mass, high precision sub-detector and consists of four layers of silicon sensors. The silicon sensors are based on a double sided strip technology, in order to avoid the count of huge pixel numbers. They also tolerate the radiation levels expected for the SVD region [6].
The innermost and the outermost SVD layers have the radii 38 mm and 140 mm, respectively. Except for the innermost layer, the forward region towards the z-axis is slanted to reach a better precision in forward (HER) direction and, simultaneously, to reduce the required number of sensors. A 3D visualization of the SVD is presented in Fig. 3.5[31]. Different SVD parameters are summarized in Table A.1[31] of Appendix A.



(a) front view without slanted sensors (b) isometric projection

Figure 3.5: Schematic view of the geometrical arrangement of the sensors for the SVD. Only the silicon sensors are shown, all other components have been omitted [31].

3.2.5 Central Drift Chamber CDC

The central drift chamber detects charged particle tracks within the detector acceptance region. The reconstruction of the particle tracks allow to accomplish two significant tasks: first, to measure the particle momentum from the track curvature and second, to contribute

to the particle identification from the energy loss within the CDC gas volume. Nevertheless, the most important attribute of the CDC, with respect to the present studies, is that it provides efficient and reliable trigger signals for charged particles, since the PXD and SVD are not available for this purpose.
The CDC in Belle II inherits the same global structure of the CDC in Belle, but with increased radii and wire density, in order to improve the momentum resolution. A comparison of the main CDC parameters in Belle II and Belle are summarized in Table 3.2[6].

	units	Belle	Belle II
Inner Cylinder Radius	cm	7.70	16.00
Outer cylinder Radius	cm	8.80	113.00
Innermost Sense Layer Radius	cm	8.80	16.80
Outermost Sense Layer Radius	cm	86.30	111.14
Sense Wire Diameter	µm	30	30
Number of Layers		50	56
Number of Sense Wires		8,400	14,336
Gas		He-C_2H_6	He-C_2H_6

Table 3.2: Main parameters of the Belle CDC and the upgraded Belle II CDC [6].

The CDC consists of an inner cylinder made up of thin carbon-fiber reinforced plastic (CFRP), two aluminum endplates and a CFRP outer cylinder. The CDC volume is filled with a 50% helium and 50% ethane gas mixture. There are 42,240 field wires and 14,336 sense wires spanned between the endplates. The properties and the amount of the two different wire types are listed in Table 3.3[6].

	Sense Wires	Field Wires
Material	Tungsten	Aluminum
Plating	Gold	No
Diameter (µm)	30	126
Tension (g)	50	80
Total Number of Wires	14,336	42,240

Table 3.3: Wire parameters in the Belle II CDC [6].

Each sense wire has an unique wire identification number (ID). Together with the surrounding field wires, each sense wire also builds an unique drift cell. The sense wires are arranged in radial layers (S. Fig. 3.6). In each layer the number of sense wires, or rather the number of drift cells, is required to be a multiple of 32, in order to match the number of electronic channels and trigger track segments.
Six contiguous wire layers in the same orientation form a so-called superlayer. There is a total of 9 superlayers. The innermost superlayer contains two additional layers of guard

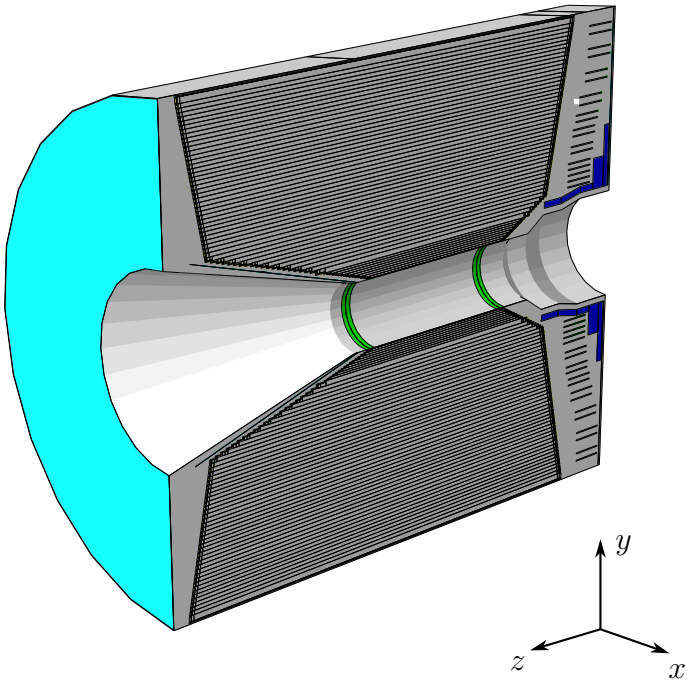
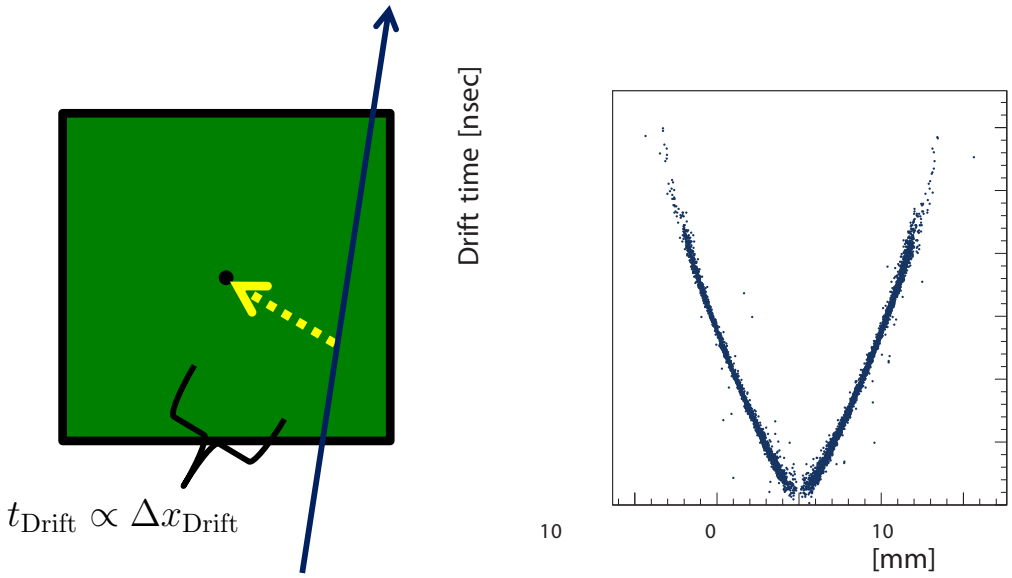


Figure 3.6: Schematic view of the CDC. The opening angle at the left side is 17° and at the right side 30° for the total θ acceptance of ($17° < \theta < 150°$). The plotted coordinate system only points to the correct directions.

sense wires, giving a total of 8 layers. The cell size in this superlayer is small (10mm), in order to reduce the occupancy caused by the high background rates at smaller radii. In the outer superlayers the cell size is $\sim$ 18.2mm.
The superlayers alternate between Axial ("A") orientation, paralell to the z-axis and stereo ("U", "V") orientations, which are slanted by a small, so-called stereo angle between $\sim \pm(2.6 - 4.2)°$ with respect to the z-axis. The stereo angles help to provide a better z resolution. The total configuration of all superlayers is AUAVAUAVA. Table 3.4[6] presents a list of the main configuration parameters.
A charged particle passing through the CDC ionizes the volume gas. The electrons lost by the ionized gas molecules drift towards the positively charged sense wires while ionizing further gas atoms. The result is an avalanche of electrons producing a measurable current. A particle drifting to a sense wire needs a certain time to reach it, the so-called drift time (S. Fig. 3.7(a)). The drift time depends on the distance between ionization point and sense wire. As it can be seen in Fig. 3.7(b), this relation is almost linear: the drift velocity can then be determined from the slope after making a linear fit. The measured CDC spatial point resolution is about 100 µm. For a wide range of voltages between field and sense wires, the drift speed in the CDC gas mixture remain almost constant at $\sim$ 40µmm/ns. Furthermore, the field wires are situated such that there are almost circular isochrones in each drift cell.



(a) (b)

Figure 3.7: a) A charged particle ionizes the gas in a CDC drift cell (green box); the drift time t_{Drift} is proportional to the distance Δx_{Drift} between the particle track (blue line) and the sense wire (black dot). b) The measured relation between x_{Drift} and Δx_{Drift} [6]: electrons from ionized particles are accelerated to the sense wires because of the electric field; coulomb scattering on further volume gas molecules counteracts this effect. This results in a motion with an effective drift velocity.

SL No.	Orientation	Type	No. of layers	Drift cells per layer	Radius (cm)	stereo angle (mrad)
1	Axial	A	8	160	16.80-23.80	0.0
2	Stereo	U	6	160	25.70-34.80	45.4 – 45.8
3	Axial	A	6	192	36.52-45.57	0.0
4	Stereo	V	6	224	47.69-56.69	-55.3 – -64.3
5	Axial	A	6	256	58.41-67.41	0.0
6	Stereo	U	6	288	69.53-78.53	63.1 – 70.0
7	Axial	A	6	320	80.25-89.25	0.0
8	Stereo	V	6	352	91.37-100.37	-68.5 – -74.0
9	Axial	A	6	384	102.09-111.14	0.0

Table 3.4: Main configuration parameters of the sense wire superlayers (SL) in the Belle II CDC [6].

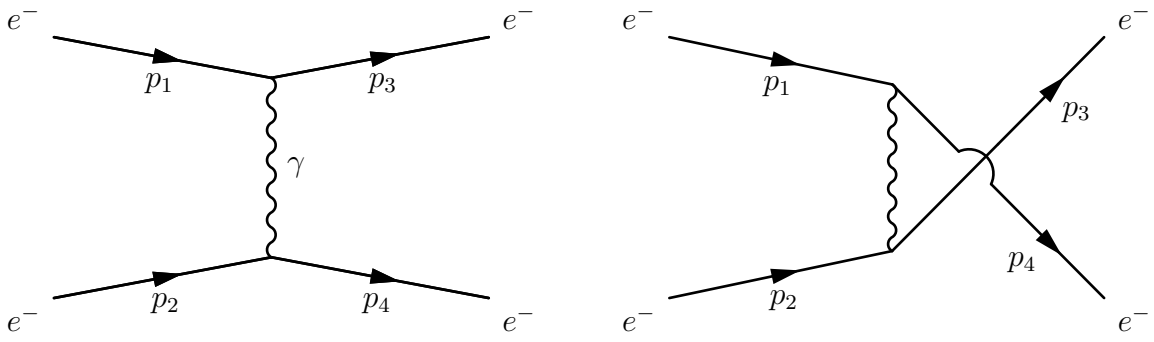
The CDC registers the wire ID, of the wires which have been hit, together with the respective drift time values t_i, i.e. the time at which each wire hit occurred. The readout of the CDC is done by a digitizer board contained in a Field Programmable Gate Array (FPGA), which is placed on the CDC front endplate. The FPGA read out rate is 1 GHz, which results in a drift time resolution of 1 ns. The CDC information, i.e. wire IDs and drift times, will be then transferred to the first level trigger within a lapse of about 1µs.

3.3 The Background

In a general sense, all particle tracks and signals in the detector stemming from undesired physical processes can be defined as background. They are undesired, not only because they do not contribute to the search of NP, but also because they can damage expensive detector components creating dangerous situations at the same time. The background can be classified into luminosity-dependent background, generated by quantum-electrodynamical (QED) processes between electrons and positrons, and machine background, which depends on the specific beam settings. In Belle II the possible sources of machine background are beam synchrotron radiation, scattering of the beam on residual gas and Touschek scattering. Luminosity-dependent background sources are Bhabha scattering and $e^+e^- \rightarrow e^+e^-e^+e^-$ processes via two $\gamma\gamma$ photons.

3.3.1 Touschek Scattering

Due to particle losses caused by intra-bunch scattering, each Beam in the accelerator ring has its own independent lifetime, also called Touschek lifetime. As shown in Fig. 3.8), by elastic Møller scattering between two electrons in the same bunch, a transfer Δq of radial momentum into longitudinal momentum can be produced. If $\Delta q > \Delta p/\gamma$, were Δp is the central momentum acceptance in the lab system of the accelerator ring, then one of the particles will become too much energy while the other is left with too little energy in order to remain in the stable region of the closed beam orbit. Subsequently, after the scattering process, both particles get lost. This effect is called Touschek scattering [52].
Touschek scattered particles hit the inner wall of the beam pipe propagating around the ring. If their loss position is close enough to the detector, they will hit the vacuum chamber and the magnet walls near the IP, generating a large number of particle showers and also neutrons, which will be registered by the detector. This, of course, produces noise to the interesting detector events, deteriorating the physics resolution.
The scattering rate of the Touschek effect is inversely proportional to the beam size. In the case of SuperKEKB it will be reduced to $\sim 1/20$ in comparison with KEKB, because of the nano-beam scheme. Therefore, it will be the most dangerous source of Background at SuperKEKB. Nonetheless, the Touschek scattering rate is also proportional to the bunch current density and to E^{-3}, were E is the beam energy. Thus, only the Touschek background contribution of the LER matters; one of the reasons why the LER energy will be increased from 3.5 to 4 GeV, was indeed to reduce the Touschek background [53].



(a) t-channel (b) u-channel

Figure 3.8: Møller scattering: $e^-e^- \to e^-e^-$. The final momenta of the electrons differ from those in the initial state.

Studies of the Touschek effect based on simulations predict an amount of Touschek Background for SuperKEKB 20-30 times higher than in KEKB [6].

3.3.2 Beam-Gas Scattering

Since ideal vacuum conditions inside the beam-pipe are impossible to reach, charged beam particles can interact with the residual gas molecules through bremsstrahlung and Coulomb scattering [54–56]. This produces a momentum transfer of the beam particles, and thus a deviation from their original path. A considerable part of these particles hit the walls of vacuum chambers and magnets creating shower particles which create then background events. The size of the beam-gas scattering depends on the beam current, the vacuum pressure in the accelerator ring and the strength of the magnets [6].
The beam currents in SuperKEKB will be about two times higher than in KEKB and also the magnet fields of the final focus magnets will significantly increase. Due to the limited space near the IP, the thickness of the chamber mask around the IP will be decreased. Additionally, the vacuum level at the IR, about ± 2 m from the IP, will be 100 to 1000 times higher than in Belle. Therefore, beam-gas background can be considered as one of the major sources of beam induced background. Nevertheless, outside from the IR the vacuum level in SuperKEKB will be the same as in KEKB. Particles scattered at the IR will be lost far enough from the IP and therefore beam-gas scattering is not as dangerous as Touschek scattering. In SuperKEKB, approximately the same amount of beam-gas background as in KEKB is expected.

3.3.3 Synchrotron Radiation

Synchrotron radiation (SR) is emitted when a charged particle is radially accelerated ($\dot{\mathbf{v}} \perp \mathbf{v}$); in the case of high energetic particles, photons are radiated in a narrow cone, of opening angle $\sim 1/\gamma$, centered around the instantaneous tangent of the particle trajectory

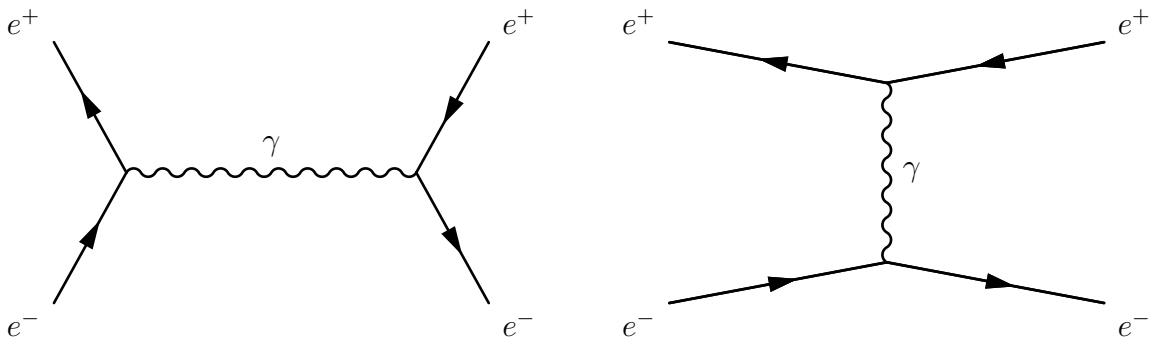
[57]. SR occurs when the particle beam is being bended and its size depends on the beam current and the beam optics. In accelerator rings, particles have to be involved into a circular motion and therefore this phenomenon limits their efficiency, especially for e^+e^- accelerators. In SuperKEKB, the main SR source is the HER beam because of the beam optics.
In Belle II the incoming beam direction, for an observer at the IP, is called upstream direction. Similarly, the outgoing beam direction is called downstream. With this convention it is possible to classify between SR coming from upstream and SR coming from downstream after being scattered back at ferromagnetic materials:

i) **SR Upstream:** To protect the inner PXD and SVD detectors, the design of the IP chamber has been developed to avoid direct SR hits from the HER. With the planned HER beam current and beam optics (magnet positions, magnetic field strength, bending radii and beam orbits) together with the IR components geometry, GEANT4-based simulations showed no direct SR hits, not even in the PXD [6].

ii) **SR Upstream:** SR Backscattering: In KEKB the final focusing magnet for the HER also served as bending magnet to separate the outgoing HER beam from the LER beam. This bending generated a lot of SR photons (with ~ 38 keV [58]) which were scattered back to the detector after hitting the beam pipe. In SuperKEKB, such a bending does not exist anymore, since two separate quadrupole magnets are located at each side of the IR in order to focus each beam independently. Also the beam orbits of both incoming and outgoing beams are on the center in the magnets. Although optics calculations for SuperKEKB predict a 1/800 reduction of the SR backscattering in comparison with KEKB, this result still needs to be supported by simulation studies [6].

3.3.4 Radiative Bhabha Scattering

Bhabha scattering is a QED process between electrons and positrons. The leading-order Feynman diagrams contributing to this interaction are annihilation (s-channel) and scattering (t-channel), which can be seen in Fig. 3.9. Photons radiated by the Bhabha scattered particles propagate along the beam axis direction. When these photons interact with the iron of the magnets, neutrons are copiously produced by effect of the so-called giant photonuclear resonance mechanism [59]. Neutrons produce background hits in the outermost detector, namely in the K_L and μ (KLM) detectors. Hence, Bhabha scattering is the main background source for this detector parts.
Scattered electrons and positrons lose energy. Because of their lower momentum they do not follow the ideal path through the focusing magnets and they are said to be over-bent. In turn, they hit the wall of the magnets generating electromagnetic showers.
The rate of radiative Bhabha events is proportional to the luminosity, which will be much higher in SuperKEKB, though the estimated cross section for Bhabha scattering is very small, ~ 120 nb for the t-channel, within the acceptance ($17° < \theta < 150°$) of the Belle II detector [34]. As mentioned before, in SuperKEKB, the focusing of each beam is done



(a) s-channel: annihilation (b) t-channel: scattering

Figure 3.9: Bhabha scattering: $e^+e^- \rightarrow e^+e^-$. The scattered electrons and positrons radiate photons.

independently by doublets of quadrupole magnets at each side of the IR. The orbits of the incoming and outgoing beams are also centered to the quadrupole magnets. In consequence, the radiative Bhabha background due to over-bent electrons should be small. Based on beam optics calculations, 1/40 of the Background at KEKB is estimated for SuperKEKB [6].

3.3.5 Electron-Positron Pair Production via two Photon Process

The luminosity-dependent QED process $e^+e^- \rightarrow e^+e^-e^+e^-$ via two $\gamma\gamma$ photon has an estimated cross section of $\sim \mathcal{O}(10^7)$ nb [34], within the acceptance of Belle II . The leading-order Feynman diagram contributing to this interaction is presented in Fig. 3.10. Since the new PXD detector is placed very close (1.3 cm) to the IP, also low momentum particles can reach the first PXD layer and produce damages.
Studies on the low energy secondary e^+e^- pairs have never been done experimentally [34]. Therefore, to predict the $\gamma\gamma$ background situation, Monte Carlo (MC) simulations has been done by E. Nedelkovska [34] using the MC generators BDK[60] and KoralW[61]. The MC Simulation result comparing the event kinematics of the two MC generators is shown in

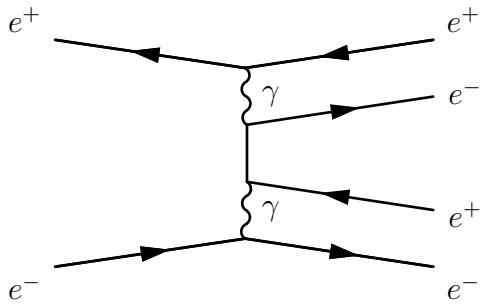


Figure 3.10: Multipheral $e^+e^- \rightarrow e^+e^-e^+e^-$ production via two γ photons.

Fig. 3.11[34]. The energy spectra of both MC generators agree in a good way.
Due to the magnetic field inside the Belle II detector cylinder, at least a minimum transverse momentum of 3.5 MeV is needed to reach the innermost layer of the PXD. As it can be seen in Fig. 3.11 only a very small fraction of the tracks overcome this energy barrier. Thus, no hits are expected for further detector components, including the CDC.

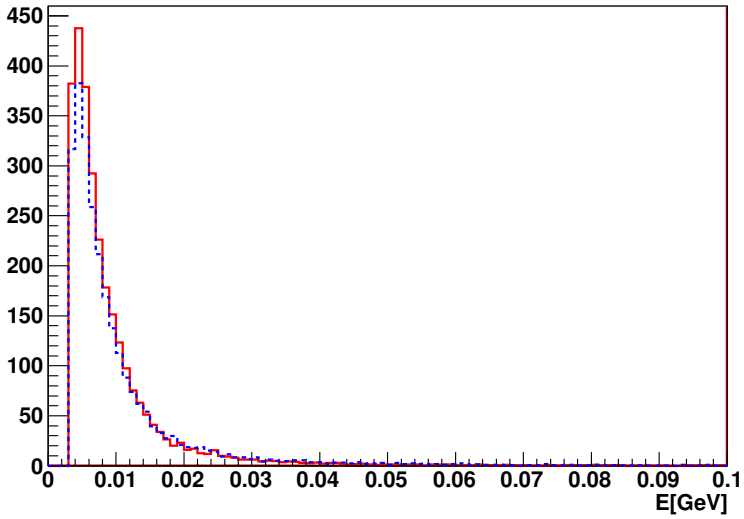


Figure 3.11: KoralW (dashed blue) and BDK (solid red) simulation for the energy spectrum of e^+e^- pairs produced via $\gamma\gamma$ processes [34].

3.3.6 The z-Background-Distribution in Belle

After ten years of operation in KEKB, the distribution of the track vertices in z-axis direction, stemming from real data, was recorded by a random trigger in the Belle experiment. This distribution can be seen in Fig. 3.12[6]. The narrow peak at $z = 0$ correspond to particle tracks whose vertex was located at the IP. Considering the understanding of background presented in this section, together with further simulation studies [6], these tracks can be confidently associated to physical interesting reactions from the e^+e^- collisions. The broad distribution outside the $(-2, 2)$ cm region belongs then to particle tracks generated via Touschek and beam-gas scattering processes.
This Background events could have been removed giving a veto on those particle tracks whose z vertex position is not located inside the $(-2, 2)$ cm region. In order to do this, a fast detection of the z vertex is required while detector operation. This is exactly the task of a so-called z-vertex trigger: the Belle experiment could not reject these Background events, because it had no such a z-vertex trigger at its disposal.
In Belle II , the background is expected to be similarly distributed, i.e. around the $z = 0$ peak, but with a much more larger background-to-physics proportion, due to the higher Touschek background rates. The most worse background level will be at the beginning of the accelerator operation, when the vacuum condition together with several other detector parameters will not have been optimized.

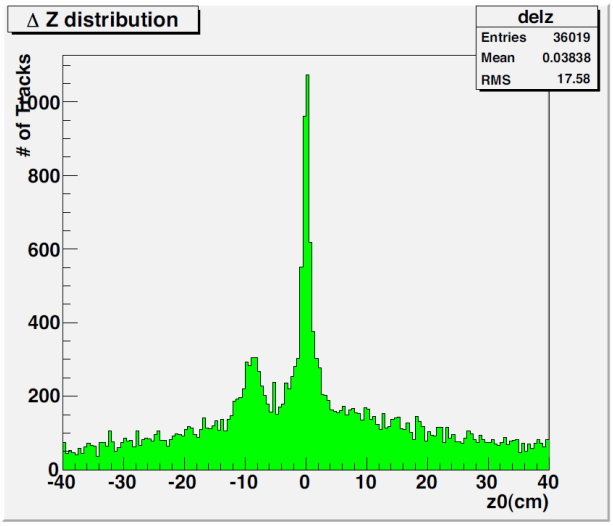


Figure 3.12: Distribution of the z-position of vertices belonging to reconstructed particle tracks in Belle. The peak at $z = 0$ corresponds to $\Upsilon(4S)$-decays. The broad background is due to Touschek and beam-gas scattering. The lower peak at $z \sim -10$ cm is an artefact caused by the beam bunch structure [6].

3.4 Trigger System

The high data production rate in Belle II, due to the high design luminosity and the high background rates, overwhelms any transfer and recording capacity at the current levels of technological advance. Therefore, it is absolutely necessary to filter out the background events in real time, i.e. while taking data from the detector, and to record only those events which are relevant for the search of NP. This is, in principle, the task of the Level 1 (L1) trigger system.
The trigger consists of several sub-trigger components and one global decision logic (GDL). Each sub-trigger collects information upon its sub-system and transfers it to the GDL, which makes the final decision combining all sub-trigger votes. The whole trigger process is required to be performed within a fixed latency of $\sim 5\mu$s. The total data rate in Belle II is expected to be ~ 20 kHz and thus the maximum average trigger rate is set to be 30 kHz. Further requirements on the trigger system are [6]:

- high efficiency for hadronic events from $\Upsilon(4S) \to B\bar{B}$ and from continuum;
- a timing precision of less than 10 ns;
- a minimum separation of 200 ns between two succeeding events; and
- a trigger configuration that is flexible and robust.

A schematic view of all sub-trigger components is shown in Fig. 3.13[6]. Each independent sub-trigger contributes, within its capabilities, to the accomplishment of the requirements

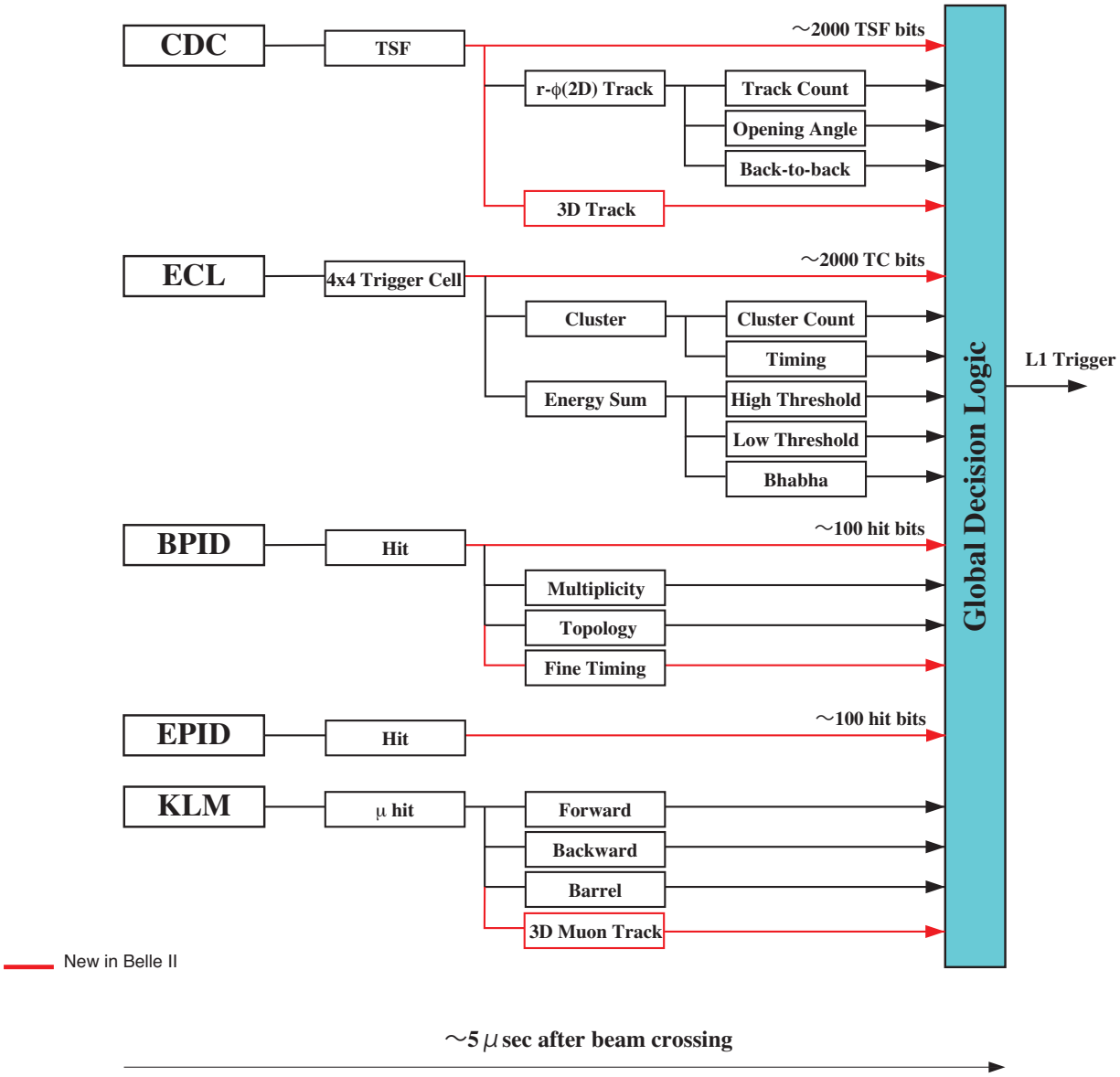


Figure 3.13: Schematic overview of the trigger system. the five sub-trigger systems send their outputs to the Global Decision Logic GDL, where the final decision is made. The red lines represent the new information paths for the Belle II trigger [6].

above. The CDC and ECL sub-triggers are used in the interest of achieving high efficiency for hadronic events. The ECL provides information about energy deposition and energy clustering, as well as identification of Bhabha radiation and cosmic rays. Notwithstanding, within the scope of this work the focus is set on the CDC sub-trigger. The CDC sub-trigger supplies information derived from charged particle tracks detected therein. It includes itself three different sub-components: the track segment finder (TSF), the 2D CDC trigger and the 3D CDC trigger.

3.4.1 The Track Segment Finder (TSF)

The whole data from the 1 GHz CDC readout board is too large and is thus preprocessed by the track segment finder (TSF). It is the first step in reducing data and suppressing background hits. The track segments (TS) are groups of several wires, or rather drift cells, arranged in pre-defined geometrical shapes within each superlayer [6].
Each TS overlaps with the neighboring TS except for the so-called "priority cell". In the case of the first super layer, the priority cell is the innermost drift cell of the TS and, in the case of all other superlayers, it is the central drift cell. There is a total of 2336 TS, each one of them has a unique identification number (TS-ID).

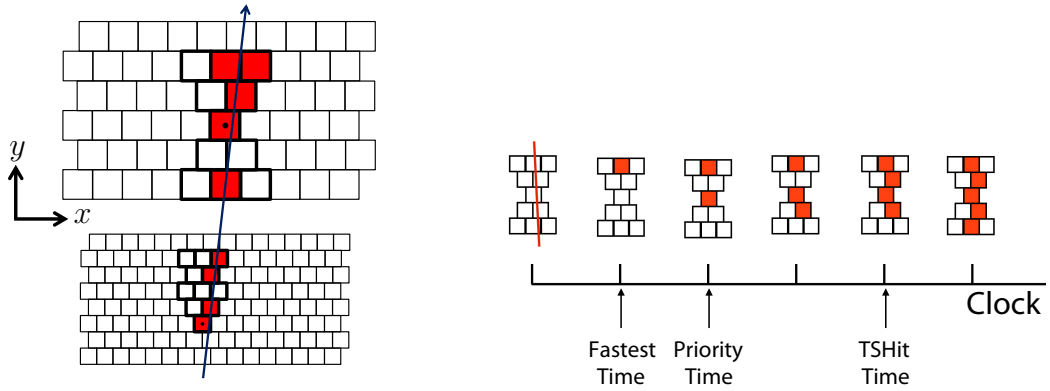


Figure 3.14: a) Each track segment (TS) consists of an arrangement of 5 layers containing several wires within a specific geometrical shape. In the innermost superlayer the first two and the last sense wire layers are not taken into account. In the outer superlayers, only the information of the last sense wire layer is neglected. b) A TS is active if at least one wire in four of its layers have fired. In that case, the TSF registers the position and the clock time information of the priority cell in the TS (dot in (a)) [62].

There is an incorporated clock in the TSF with a time resolution of 2 ns. A lapse of 32 ns is referred to as clock cycle. The TSF determines which TS are active at each cycle tick of the clock. If a CDC wire fires, the hit is memorized for a period of 16 clock cycles. A TS is then found active if at least four of its five layers get a hit within a clock cycle. In that case, the TSF registers only the clock time at which the priority cell in the TS was hit. This "priority time" is just the drift time value of the priority cell, scaled to the coarser 2 ns clock resolution. In the case that four layers in the TS fire within a clock cycle, but not including the priority cell, then the clock time of one of the two drift cells above is used as second priority time [62]. In consequence, the information of all layers in each superlayer is mapped to a single layer of track segments. An illustration of the TSF operating mode is shown in Fig. 3.14.
Within the period of 16 clock cycles, the TSF sets the priority time of the first active TS as zero point for further active TS. Since there is always a time lapse between ionization

and wire hit, setting the priority time as zero point means that the clock time information of all TS in an event is delayed for an additional unknown time interval, which is called random offset [62].
The final operation concept of the TSF [63] is not implemented yet in the current version of the Belle II simulation software, which produced the input data for these studies. In the current software version of the TSF, the priority cell for all TS is the central drift cell. It gives only the drift time values of these cells as output, without random offset. Therefore, in this work, some experiments have been done in order to study the effect of the random offset on the performance of the neural z-vertex trigger.

3.4.2 2D and 3D CDC Trigger

The further components of the CDC sub-trigger are the 2D and the 3D CDC triggers, which run parallel to each other. Because of the maxmimal latency of $\sim 5\mu s$, for the whole L1 trigger, and the latencies associated with the TSF and GDL, only a time window of $\sim 2\mu s$ is left for the 2D and 3D CDC triggers. Since this requirement exceeds the capabilities of conventional CPUs, the 3D and 2D triggers are implemented on FPGAs.
The 2D trigger uses a conformal transformation and a Hough transformation [6, 64]. It performs a track fit on the (r, ϕ)-plane using the TS hit information provided by the TSF. From this fit, the 2D trigger extracts the transverse momentum p_T, and the azimuthal ϕ angle.
The 3D CDC trigger, on the other hand, performs a two-step process: a track fit on the (r, ϕ)-plane, similar to that done by the 2D trigger, followed by a fit on the (r, z)-plane. The output of the 3D trigger gives information about all track parameters (p_T, ϕ, θ, z). The

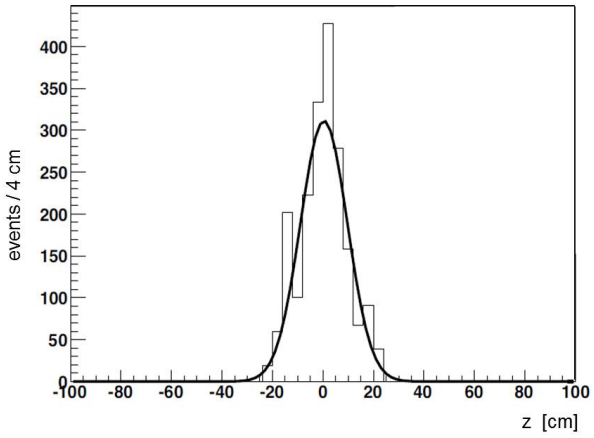


Figure 3.15: Distribution of the difference $z_{\text{fitted}} - z_{\text{Sim}}$ between the z value from the 3D fit and the z value from the simulation. The standard deviation, i.e. the resolution achieved, is $\sigma(z_{\text{Fit}} - z_{\text{Sim}}) = 9.309 \pm 0.250$ [63].

convention used for track parametrization will be presented in the next section 3.5.
Notwithstanding, the resolution in z achieved by the 3D CDC trigger, of about $\sigma = 9.3$ cm (S. Fig. 3.15), is not enough to filter out the background shown in Fig. 3.12. The neural z-vertex trigger approach, on which the present studies are based, reaches a much higher z vertex resolution by almost a factor of 10. Nevertheless, the neural z-vertex trigger needs a coarse prediction of the track parameters. Therefore, the information supplied by the 2D and 3D trigger could be used as prediction for the the neural z-vertex trigger.
In comparison, an offline track reconstruction can reach a z-vertex resolution of some millimeters combining the information of axial and stereo wires provided by the CDC, though there is no possibility to do such a reconstruction within the L1 trigger latency.

3.5 Track Parametrization

In general, particle tracks in the detector can be described by a helix. Since the magnetic field is oriented parallel to the z-axis, the helix curvature is located on the $r\phi$-plane due to the Lorentz force. Inside the detector, the magnetic field strength can be considered as a constant. Similarly, energy losses caused by the particle interaction with the detector material are neglected. Thus, a specific track can be unambiguously determined if one point of the track is known together with the momentum thereat. In this work each track is characterized by the primary vertex and the initial momentum. For the studies with neural networks, only single track events have been used and in consequence, there are no secondary vertices. Furthermore, for the net optimization, the (x, y)-position of the vertex was set to $(x_0, y_0) = (0, 0)$.
With the Belle II coordinate system definition, a single track is described by four parameters which can be merged into a track parameter vector

$$\mathbf{q}_0 = (p_\mathrm{T}, \phi, \theta, z_0)^T \tag{3.5}$$

where $p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}$ is the transverse momentum, ϕ, θ are the track angles at the vertex and z_0 is the z-vertex position. For a given track parameter vector, the respective helix $\mathbf{t} = (x, y, z)^T$ can be described as a function of a single free variable μ [65],

$$\mathbf{t}(\mu) = \begin{pmatrix} \frac{1}{\kappa}\left[\sin(\phi + \kappa\mu) - \sin(\phi)\right] + x_0 \\ \frac{1}{\kappa}\left[-\cos(\phi + \kappa\mu) + \cos(\phi)\right] + y_0 \\ \zeta\mu + z_0 \end{pmatrix} \tag{3.6}$$

where ζ is the extension in z-direction of the track in dependence of the z-component of the total momentum $\mathbf{p} = (p_x, p_y, p_z)^T$, i.e.

$$\zeta = \frac{p_z}{|\mathbf{p}|} = \cos(\theta) \tag{3.7}$$

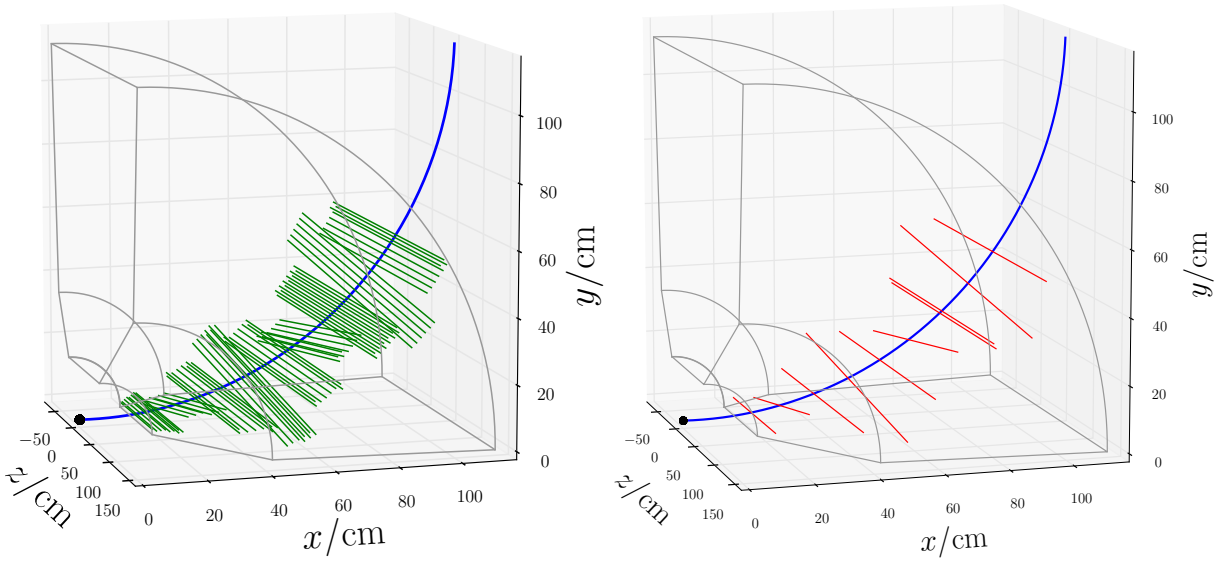


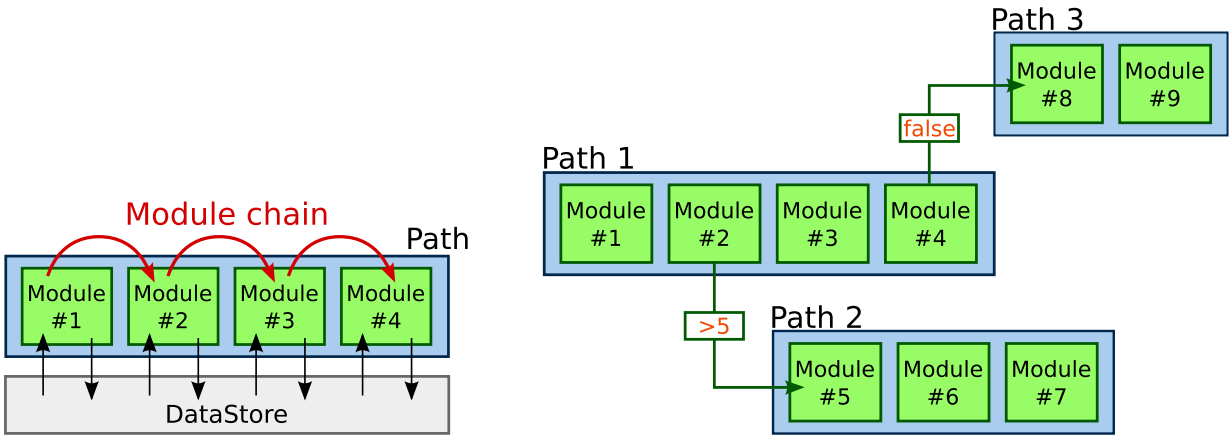
Figure 3.16: Hit pattern of a simulated single track event. The Blue line is the particle track. The green lines in a) represent CDC wires who have fired; stereo wires can be easily recognized since they are not parallel to the z-axis. The red lines are the triggered track segments (TS). The track parameters are $\mathbf{q}_0 = (p_\mathrm{T}, \phi, \theta, z_0) = (0.57\ \mathrm{GeV}, 0.83°, 61.15°, 1.53\ \mathrm{cm})$.

and κ is the curvature of the helix. The curvature can be calculated considering the effect of Lorentz force by the magnetic field B inside the solenoid

$$p_\mathrm{T} = |\mathbf{p}| \sin(\theta) = q \cdot B \cdot r \tag{3.8}$$

$$\Rightarrow \kappa = \frac{1}{r} = \frac{q \cdot B}{|\mathbf{p}| \cdot \sin(\theta)} = \frac{(\pm 1)(2.99 \cdot B[\mathrm{T}] \cdot 10^{-3})}{p[\mathrm{GeV}/c] \sin(\theta)} \tag{3.9}$$

where q is the charge of the particle; $\pm$ indicates if the particle is positive or negative charged. The magnetic field strength in Belle II is $B = 1.5$ T. The factor 2.99 comes from the speed of light, since B is an indirect physical quantity.
This parametrization served only to characterize the phase-space region for the neural network optimization done in these studies. In Fig. 3.16 an example of a simulated single track is shown together with the hits produced by the CDC and active TS.
However, for the track analysis in the Belle II software simulation, another parametrization [66] is used. The difference is that the reference point is the position of the first wire that hit. An additional parameter is then used to measure the distance of the helix to this wire in the xy-plane.



(a) Module execution (b) Path conditions

Figure 3.17: Schematic view of the execution flow in the Belle II software framework. Code is divided into modules which can be placed in Paths. Modules can (a) communicate by reading from and writing to a common DataStore and (b) return a result to change the execution flow between different paths [31].

3.6 Event Simulation

The event simulation, to generate the data sets needed for the studies in this work, was done using the Belle II software framework, which is called basf2 [67] and is currently under development by the Belle II collaboration. This software combines several Belle II projects in a modular C++ framework using the recent C++11 standards [68]. The different tasks such as reading and writing data files, simulation steps, reconstruction or analysis are done by so-called "Modules". A scheme of the basic data flow during computation is presented in Fig. 3.17 [31]. The Modules can write to, or read objects from a common data store. To define the execution order of the different Modules, so-called "Paths" are used. Desired simulation conditions are possible to be set by switching between Paths depending on the intermediate result in each Module [31].
The basf2 software allows to simulate elementary particle events in the simulated Belle II detector. To simulate physical interaction processes between the particles and the detector material, an implementation of the precise description of the geometry and the detector material in Belle II is needed. This is provided by the Geant4 [69] toolkit. The geometry implementation is done by a "Geometry" Module, which determines the materials to be simulated and the sub-detector specific code to be called after reading the configuration parameters [31].
The events needed for the studies in this work were done using different specific basf2 Modules. The "Particle-Gun" module generates simple particle tracks depending on specific input parameters: number of tracks (multiplicity), particle ID (PDG-Codes), x, y, z-vertex position, vertex position distribution in the x, y, z plane, particle transverse momentum

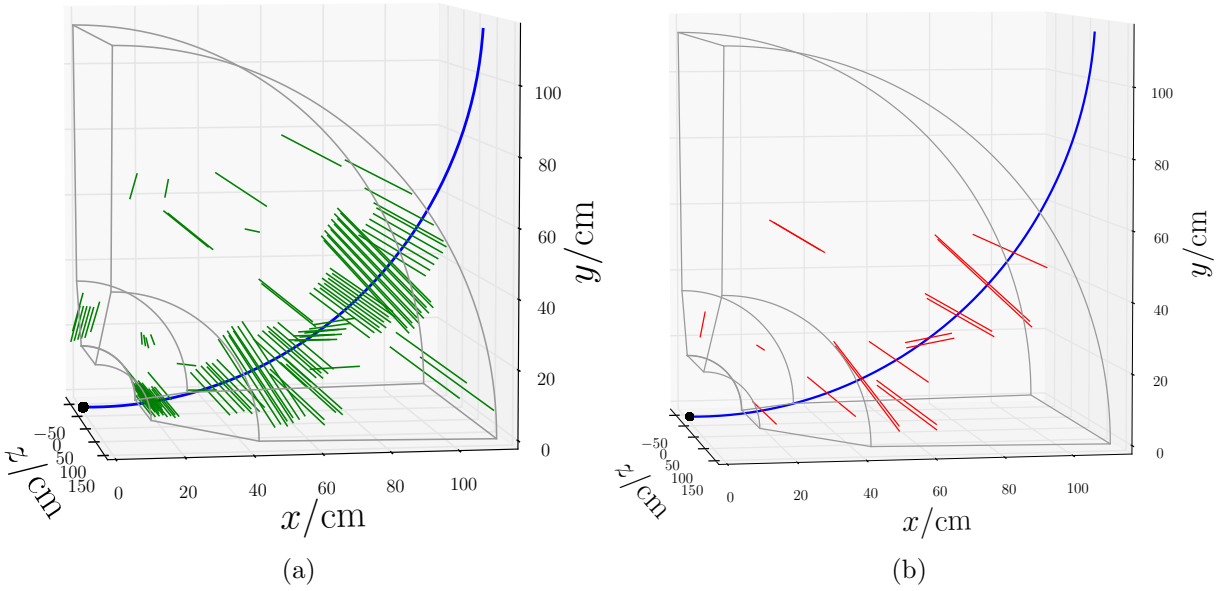


Figure 3.18: A simulated single track with background: the 3D visualization was done using the output of the basf2 Modules a) CDC-Digitizer b) TRG-CDC-TS-Stream. As it can be recognized, the number of active TS is much less than the number of wires that have been hit. The track parameters here are $\mathbf{q}_0 = (p_\mathrm{T}, \phi, \theta, z_0) = (0.54\ \mathrm{GeV}, 0.22°, 57.30°, 1.76\ \mathrm{cm})$.

p_T, transverse momentum distribution, (ϕ, θ) track direction and (ϕ, θ) distribution. Distributions are used to define the distribution of a respective parameter for a large sample of events. The available distributions in the Particle-Gun Module are; fixed, for a fixed value; uniform, for a uniform distribution within a specific region; or normal, for a normal distribution with a specific mean and standard deviation.
The "CDC-Digitizer" Module creates CDC hits from CDC simulation hits (CDC-SimHits), i.e. it generates wire hits in the CDC after the simulation of the particle interaction with the CDC wire material. Each CDC hit in a simulated event gives information about the Wire 3D position, the respective wire-ID, and the drifttime of the hit. The "TRG-CDC" module does the tasks of the trigger system reading the information stored by the CDC-Digitizer Module, e.g. track segment finding (TSF), 2D track finding, the CDC 3D trigger, etc. Using a modified version of the module "TRG-CDC-TS-Stream", the position of the middle wire in the TS, the TS-ID and the clock cycle time for each hit in the event are printed out. This module version was adapted for the special use in these studies and in those of S. Skambraks [1].
The CDC background can be included into the simulation using the "MixBkg" Module, which combine background hits to the simulation hits. The background hits stem from so-called read-out-frame files, which contain a large $\sim \mathcal{O}(10^6)$, but finite number of background hit patterns for each background type. The read-out-frame files used for the present studies contain background event patterns for Touscheck scattering (for the LER and HER), Coulomb scattering (beam-gas scattering for the LER and HER) and radiative

Bhabha. If simulation events with background are desired, the MixBkg Module have to be executed before the CDC-Digitizer Module. An example of a typical single track event with background is shown in Fig. 3.18. Also pure background events can be generated deactivating the Particle-Gun Module.
In this work, also some analyses of expected physical events have been done, in order to find out the relevant phase-space regions for the neural network studies. These physical events have been generated with the "Evt-Gen-Input" Module.

Chapter 4

The neural z-Vertex Trigger

Background events which are not generated at the IP can easily satisfy the conditions set by the ECL sub-trigger, e.g. sufficient energy deposited, as well as the conditions set by other L1 sub-trigger components. The distribution of the z-position of the reconstructed vertices in Belle (S. Fig. 3.12) shows that a considerable amount of such background events passed the L1 trigger of this detector.
Therefore, a z-vertex trigger is needed in order to satisfy the urgent demand of limiting the data transference rate in Belle II, but without rejecting events relevant for NP. As argued before, the requirement on such a trigger is a resolution in order of $1-2$ cm, which has not been reached by the present Belle II 3D trigger. Different studies done by S. Skambraks [1] showed that, in principle, a z-vertex resolution even higher than 1 cm can be achieved by a special kind of neural network if it is trained for single particle tracks within small $(p_{\mathrm{T}}, \phi, \theta, z_0)$ regions.
The success of this neural network approach motivated the proposition of a neural z-vertex trigger for the Belle II detector. Nevertheless, there are several questions thereon which need to be answered by further investigations. Some of them, namely those regarding the optimization of the neural network, have been the target of the present studies.
In this chapter the general concept of the neural z-vertex trigger will presented. Subsequently, the special kind of neural network on which the neural trigger approach is based, will be described together with the used software implementation. At the end, the hardware implementation for the final trigger will be shortly considered.

4.1 Overview of the neural z-Vertex Trigger

The neural z-vertex trigger is composed by a large amount of neural networks and a predictor. The special type of neural network is the so-called Multi-Layer Perceptron (MLP). Each MLP is specialized for a small track parameter region, i.e. a small sector in $(p_{\mathrm{T}}, \phi, \theta, z_0)$, since they reach a much better accuracy than a single MLP for a broad region [1]. The MLPs use the information supplied by the TSF as input and give the z-vertex position as output.

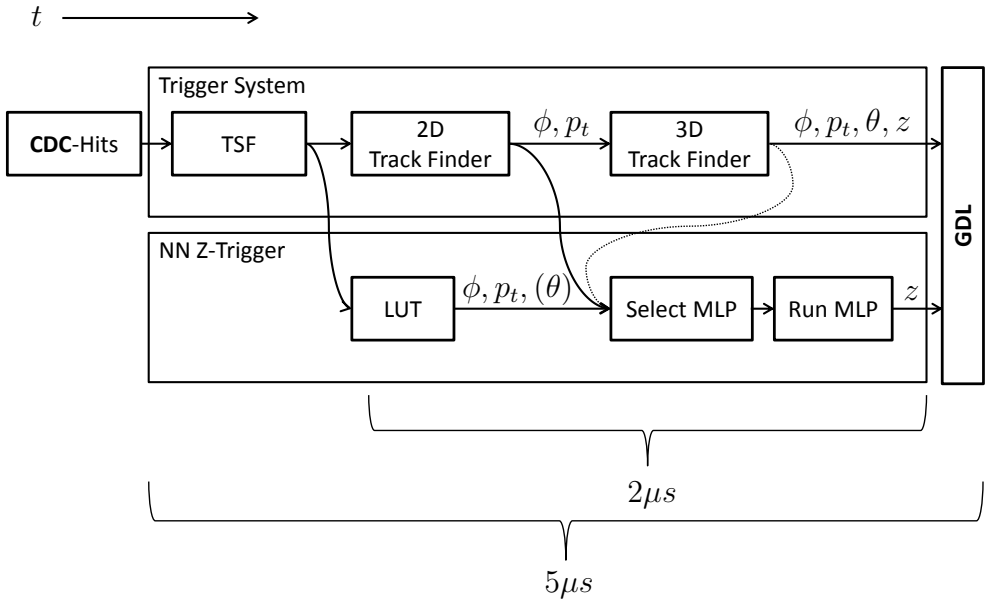


Figure 4.1: A scheme of the neural z-vertex trigger operating mode.

In a general event, the sector in which the track is located is not known a priori. This information is required though to select the specialized net and thus a pre-processing step, or rather predictor, that provides the relevant sectors for a given event is essential. The first option would be to use the output of the 2D and the 3D CDC trigger as prediction. Nevertheless, it should not be forgotten that the z-vertex trigger will only have a time window of $\sim$ 2µs and it is not assured that the 2D and 3D trigger will supply a prediction as fast as needed.
A further solution is to identify the sector of a track candidate using a so-called look-up table (LUT) [1]. The problem here is that look-up tables require a large amount of memory. An illustration of the neural z-vertex trigger operation concept is shown in Fig. 4.1 Furthermore, the neural z-vertex trigger is subjected to all other requirements set on the L1 trigger (S. subsection 3.4).
Future studies will consider the different pre-processing options in depth. For this study the focus was set on the optimization of the specialized MLPs and on the test of their performance under the presence of background.

4.2 The Multi-Layer Perceptron MLP

In computer science, neural networks are constructs that pretend to mimic the function of the real biological neural networks which constitute the brain and the nervous system of human beings and animals. Biological neural networks are composed of a special kind of cells called neurons which are interconnected among themselves. The attempt to simulate the different structures and behaviors observed in nature conduced to several types of artificial neural networks for computing purposes [1, 70].

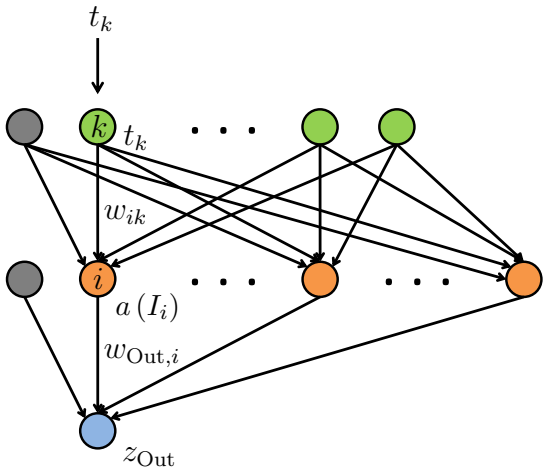


Figure 4.2: Schematic view of a three layer MLP. The nodes represent neurons and the vertices the synaptic connections. The bias neurons are the grey nodes which have a constant input value equal to 1.

A Multi-Layer Perceptron (MLP) is a less biological model of neural network. It consists of several neurons arranged in layers, which are interconnected in a feedforward structure, i.e. the neurons in one layer are only connected to neurons in the following layer. In this kind of network, there are no recursive connections and, consequently, the execution time of a single network is a deterministic constant [1].
For the z-vertex trigger the used MLPs consist of three layers: an input, a hidden and an output layer, where there is only one output neuron (S. Fig. 4.2). Three layer MLPs are very powerful. Indeed, a superposition theorem by Kolmogorov [71] implies that all continuous functions on the hypercube $[0,1]^n$ can be approximated by a three layer perceptron if $n > 2$.
Each neuron in the MLP is simply a node which receives an input and gives an output value. In the input layer, each neuron k represents an input channel: the input channels are the priority cells of the relevant TS located in the region for which the net is specialized. As argued before, each MLP is specialized for a small track parameter region $\Delta(p_{\mathrm{T}}, \phi, \theta, z_0)$. The detector geometry, i.e. (ϕ, θ, z_0), and the track curvature due to p_{T} determine which TS are located in a given specific region. If all TS therein are taken as input channels, the size of the input layer increases to much, even for very small regions. For that reason, only the most relevant of them are selected. In this work, different experiments have been done varying the selection criterion for the TS. The results will be presented in the next chapter.
The input values t_k are the drift times of the priority cells. Additionally, a bias neuron is included with a constant value $t_0 = 1$. The input layer just present the data to the hidden layer. A neuron i in the hidden layer receives in turn the weighted input

$$I_i = \sum_{k=1}^{n_{\mathrm{input}}} w_{ik} t_k + w_{i0} \tag{4.1}$$

where w_{ik} are the weights of the connections between the neurons. For a number of input neurons n_{input} and a number of hidden neurons n_{hidden}, the total number of weights connecting input and output layer is

$$n_{\text{hidden}}^{\text{weigths}} = (n_{\text{input}} + 1)n_{\text{hidden}}. \tag{4.2}$$

The summand +1 arrives from the bias input neuron, whose weights can be also trained by the usual training procedure. Each hidden neuron evaluates the weighted input by an activation function a and transfers this value to the output neuron. Again, a constant bias neuron is added to the hidden layer, such that the input of the neuron in the third layer becomes

$$I_{\text{Out}} = \sum_{i=1}^{n_{\text{hidden}}} w_{\text{Out},i} \cdot a\left(I_i\right) + w_{\text{Out},0}. \tag{4.3}$$

The total number of weights between hidden and output layer is

$$n_{\text{output}}^{\text{weights}} = (n_{\text{hidden}} + 1). \tag{4.4}$$

At the end, the output neuron evaluates its input by the activation function. Therefore, the function that is calculated by the three Layer MLP can be then obtained inserting 4.1 in 4.3

$$z_{\text{Out}} = a\left(\sum_{i=1}^{n_{\text{hidden}}} w_{\text{Out},i} \cdot a\left(\sum_{k=1}^{n_{\text{input}}} w_{ik}t_k + w_{i0}\right) + w_{\text{Out},0}\right). \tag{4.5}$$

4.2.1 MLP Software Implementation

The implementation of the MLP for the neural z-vertex trigger project is provided by the Fast Artificial Neural Network (FANN) library, which is an open source project [72]. This library offers the MLP structure an the training algorithm. The mostly used algorithm from this library, the improved Resilent backPropagation (iRPROP$^-$) [73] algorithm , has been chosen for this work since it conduced to very good results in previous studies for this project [1]. The advantage of this algorithm is that it adjusts the so-called "learning rate" automatically, i.e. there is one parameter less to be chosen before the training process starts [1].

4.2.2 The Choice of the Activation Function

In order to make the MLP trainable, the activation function has to fulfill several conditions: it has to be smooth, easy evaluable and easy differentiable with a smooth derivative [70]. For the neural z-vertex trigger project the activation function of all neurons in the MLP is the hyperbolic tangent

$$a(x) = \tanh(s \cdot x) \quad | \quad \tanh : \mathbb{R} \to [-1, 1] \tag{4.6}$$

where s is the so-called "steepness parameter". In this study the default value $s = 0.5$ of the FANN library is used. The output range of the MLP is the range of the activation

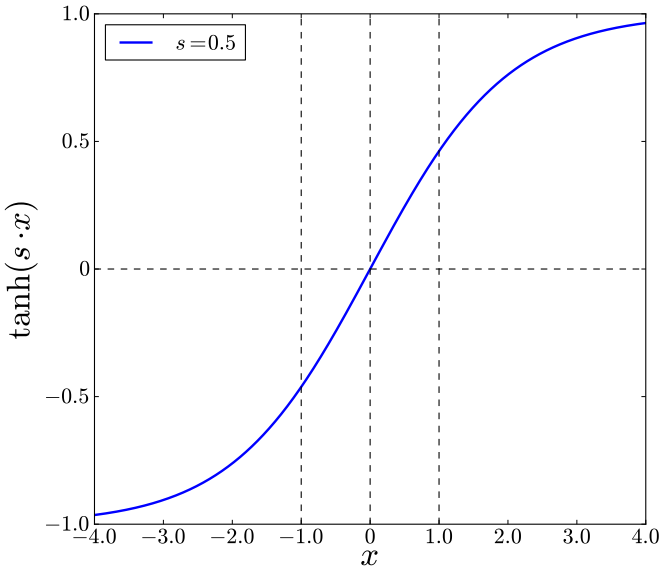


Figure 4.3: The hyperbolic tangent function. In the region $[-1, 1]$ (between the vertical grid lines) the output to input behavior is approximately linear.

function of the output neuron. In consequence, the domain of the MLP can be the whole $\mathbb{R}$ space, though its result z_{Out} is confined to $[-1, 1]$ (S. Fig. 4.3).
Some studies done by P. Sibi et al. [74] investigated the impact of different activation functions for neural networks trained with the RPROP algorithm. The results show that every activation function has approximately the same effect on the performance of the network, if it gets trained up successfully. The authors in [74] also emphasize that the most important tasks to optimize the performance of a neural network is the choice of a proper training algorithm together with the training parameters and the network size. As argued before the selection of the iRPROP$^-$ algorithm bases on studies already done for the neural z-vertex trigger project [1]. Therefore, the focus of the present studies has been set on the size of the MLP and the choice of the training parameters.
Considering again the shape of the tanh function, it can be seen that it is nearly linear around $x = 0$ for small $|x|$ input values and degenerates to a step function for large $|x|$ input values. In order to make full use of the region where the tanh function behaves linearly with respect to the input, the drift time values t_k have been scaled. Experimental results for different treatments of the time information together with the scaling will be presented in the next chapter.

4.2.3 Training Process and Cost Function

Training means nothing else but to iteratively change the weights contained in equation 4.5 using the training algorithm, such that z_{Out} reaches a desired value after the training process. For the neural trigger this value is the z-vertex position z_{Sim} for simulated particle tracks. Since z_{Out} can only cover the range $[-1, 1]$, z_{Sim} needs to be scaled to this range.

The training algorithm uses a cost function to evaluate the level of misclassification for the current weight configuration. For this project the cost function is the Mean Squared Error (MSE) [70]

$$E(\mathbf{w}) \equiv \frac{1}{N_{\text{train}}} \sum_{j=1}^{N_{\text{train}}} \left(z^j_{\text{Sim}} - z^j_{\text{Out}} \right)^2 \tag{4.7}$$

where N_{train} is the number of training events and $\mathbf{w}$ is a vector in which all current weights are merged. During the realization of these studies, very good results were achieved using training samples of $N_{\text{train}} = 20,000$. Therefore, this amount of training events have been used for all experiments, whose results will be presented in the next chapter.
In general, training algorithms update the weights vector at each iteration step $n+1$ by an amount [70]

$$\Delta\mathbf{w} = -\eta \frac{\partial E}{\partial \mathbf{w}} \quad \Rightarrow \quad \mathbf{w}_{n+1} = \mathbf{w}_n + \Delta\mathbf{w} \tag{4.8}$$

where η is the aforementioned learning rate. At this point, it can be clearly recognized that the task of the training algorithm is just to minimize the cost function. Choosing too large η values results in very large step sizes and, subsequently, the algorithm will never find an optimum. On the other hand, if η is too small the algorithm possibly remains fixed in a local minimum.
This dilemma is cleverly solved by the iRPROP$^-$ using only the sign of the derivative of the cost function and a kind of momentum $\Delta_{i,k}$ for each weight $w_{i,k}$ instead of the common learning rate. The iRPROP$^-$ itself is an improved version of the RPROP [75] algorithm. A detailed description can be found in [73].

4.2.4 Overfitting Effect and Stop of the Training Process

If the only criterion to stop the training process is to find the minimum of the cost function E on the training sample, it may occur during the training process that the cost value on this specific data set decreases such that the MLP loses its general purpose. In other words, the MLP becomes specialized on the training sample losing the capability to be used for other data samples representing the same problem. This effect is referred to as overfitting. In this project, overfitting is avoided using a cost function on a validation set in addition to the cost function on the training set. The validation data set is an amount of 1000 events generated in the same track parameter region of the training set.
At each iteration step, which is also called training epoch, the current state of the MLP, i.e. its weights values, are saved in a file with the respective step number. Simultaneously, the current cost value on the validation set is compared with the cost value 500 steps before. If there is an increase of the validation cost, then the training process is stopped and the step at which the cost value reached a minimum is searched. The corresponding state file provides then all the information of the trained MLP. An example of a training process is shown in Fig. 4.4.
The value of 500 steps is an additional parameter chosen arbitrarily. It should not be too

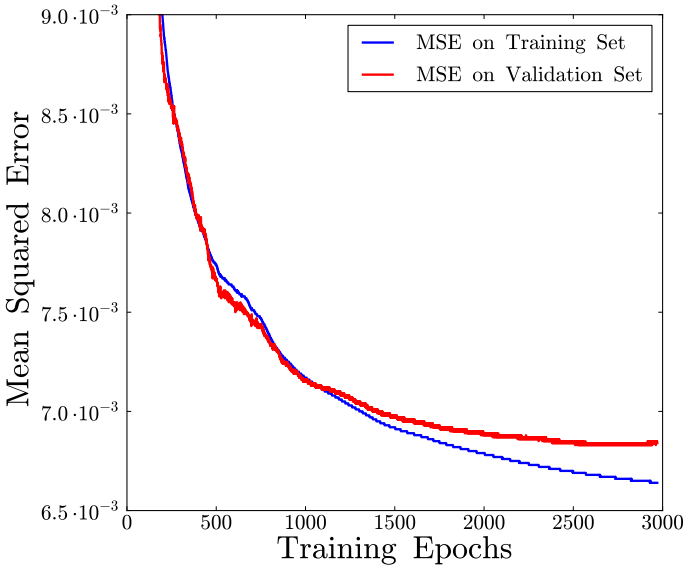


Figure 4.4: Example of a training process. The minimum of the cost function on the validation set was reached at step number 2509. The training process was stopped at step number 2970, while the cost function on the training set could have continued decreasing.

small, to prevent fixing in a fluctuation minimum and not too large, to avoid an unnecessary prolongation of the training process. However, the used amount of 500 steps has conduced to very good results.

4.2.5 Testing Process

For all experiments in this work, only single track events are used. The z_0 vertex position of all event samples is normally distributed with $\sigma = 5$ cm and mean $\mu = 0$ cm. This distribution can be seen in Fig. 4.5. For the training and the testing process in all experiments, only those events with z-vertex in the interval $z_{\text{Sim}} \in [-10, 10]$ cm were selected, in order to focus the present study on the effect of different region sizes in $(p_{\text{T}}, \phi, \theta)$. The MLP is tested always with a sample of $N_{\text{test}} = 1000$ events after each training process. The testing events were produced for the same track parameter region as the respective training and validation sets. To compare z_{Out} with z_{Sim}, both values have to be on the same scaling order. For the training of the MLP, z_{Sim} is scaled to $[-1, 1]$, but for the testing, z_{Out} is scaled to $[-10, 10]$. The performance of the MLP is then evaluated using the mean and the standard deviation of the $z_{\text{Out}} - z_{\text{Sim}}$ distribution, i.e.

$$\mu_l(z_{\text{Out}} - z_{\text{Sim}}) = \frac{1}{N_{\text{test}}} \sum_{j=1}^{N_{\text{test}}} \left(z^j_{\text{Out}} - z^j_{\text{Sim}}\right) \tag{4.9}$$

$$\sigma_l(z_{\text{Out}} - z_{\text{Sim}}) = \sqrt{\frac{1}{N_{\text{test}}} \sum_{j=1}^{N_{\text{test}}} \left(\left(z^j_{\text{Out}} - z^j_{\text{Sim}}\right) - \mu\right)^2} \tag{4.10}$$

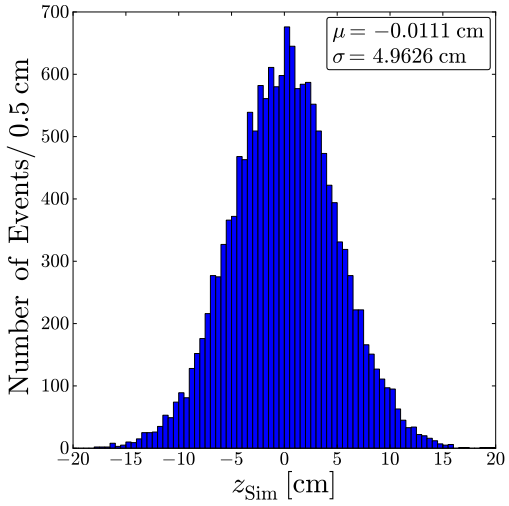


Figure 4.5: The z-vertex distribution used for all experiments. Only events with $z_{\text{Sim}} \in [-10, 10]$ cm are selected for all training, validation and test samples, such that the MLPs are specialized for this region in z_0.

where l denotes the number of a concrete testing process, since they can repeated several times. The standard deviation σ is the fluctuation of the MLP output value around the target value and it corresponds to the z-vertex resolution reached by the MLP. A large σ value means a low resolution; vice versa a small σ value means a high resolution. The mean μ measures the deviation between the output of the net and the target value. If μ differs considerably from zero, it means that the output of the MLP is biased with respect to the target value.

4.3 Hardware Implementation of the neural z-Vertex-Trigger

Considering the requirements for the neural z-vertex trigger, the most preferable approaches for the hardware implementation are the FPGA and the graphics processing unit GPU, because of their parallelizability and their data bandwidths. Both methods have their strengths and weaknesses which need to be weighted up by further studies. The currently preferred solution is the FPGA.

Chapter 5

Experimental Results

The research that has been carried out within the scope of this work is based on several experiments whose aim was to extend the results achieved by S. Skambraks [1], in order to consolidate the neural z-vertex trigger approach for the Belle II detector. As argued before, there are two main issues in this project: the choice, or rather the development of a good predictor and the optimization of the specialized MLPs. The focus of the present studies was centered on the latter. The culmination of the whole study was to test the optimized MLPs under the presence of background and under a light vertex displacement on the (x, y)-plane.
Each MLP in the z-vertex trigger is specialized for a narrow track parameter region. Therefore, the first studies aimed to make an approach of the size of the different $(p_{\mathrm{T}}, \phi, \theta)$ regions before starting with the optimization process since, as argued before, the MLP resolution depends directly on it. These approach also helped to answer another question, namely how many MLPs are needed at all to cover the whole Belle II detector acceptance and the whole transverse momenta space of particles reaching the CDC. The detector acceptance is determined by its well known geometry, but the space of all possible transverse momenta is determined by physics.
Since the training process depends on several parameters, it is not possible to recognize the correlation between each one of these parameters and the performance of the MLP if they are altered simultaneously. Therefore, the optimization process had to be performed stepwise. In consequence, the results will be presented in the same chronological order as the experiments have been executed. This will allow to recognize how each optimization step brought the MLP to a higher resolution level. Additionally, in this work also some physical analyses have been done to determine the track parameters distribution of charged particles stemming from B meson decays.

5.1 Estimation of the θ-Binwidth

From previous studies for the neural z-vertex trigger project [1], it is known that the worst prediction accuracy reached by the 3D CDC trigger as well as by the look-up table approach is the prediction for θ. Thus, the MLPs should be trained as far as possible for wide θ regions.

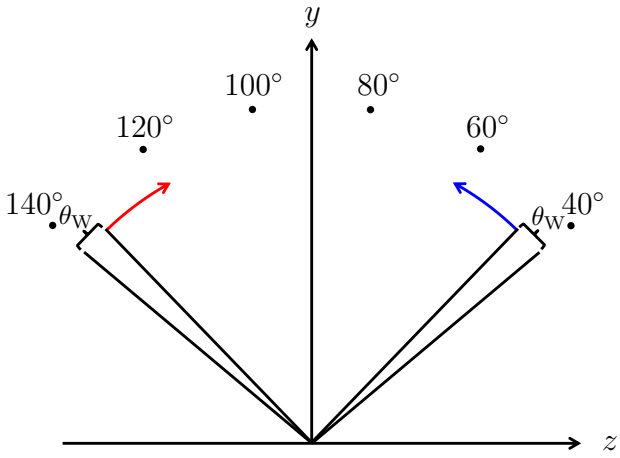


Figure 5.1: Illustration of the different θ sectors. The regions on the backward side of the detector ($z < 0$) are the mirror-inverted regions on the forward side ($z > 0$).

As described in subsection 3.2.1, the Belle II detector is built asymmetric in θ since particles are boosted in positive z-direction by cause of the asymmetric beam energies. To study the effect of training the MLP for different region widths in θ, the whole acceptance region has been divided in sectors of 20°. The first sector was $\theta \in [17, 40]°$, the second $[40, 60]°$, the third $[60, 80]°$ and so on till $[120, 140]°$. The region in $[140, 150]°$ was not covered. To compare the result of equal θ-widths, the regions in the backward side of the detector are the mirror-inverted regions of the forward side of the detector. This convention is illustrated by Fig. 5.1. The purpose was to find the largest θ width at each sector such that the resolution of the MLP do not exceeds the desired 2 cm limit.
For this experiment, single e^--tracks (PDG-Code=11) have been generated. The θ angle of the tracks was uniformly distributed within regions of widths $\{2°, 4°, \dots, 22°\}$ at each sector. The transverse momentum and the azimuthal angle were uniformly distributed in the regions $p_T \in [5.0, 5.1]$ GeV/c and $\phi \in [0, 1]°$. From the simulated events the 20 TS with the highest amount of hits were selected as input channels, i.e. the size of the input layer was 20 neurons for all MLPs. The hidden layer size was always 40 neurons.
For the first experiments, which aimed to estimate the widths of the (p_T, ϕ, θ) regions, the drift times $t_{\text{Drift},k}$ of the priority cells in in the selected TS was used as input information. The drift time values were scaled as in previous studies for the neural z-vertex trigger [1]: only drift times lower than 100 ns are used and scaled to the region $[-0.7, 0.7]$ using the formula

$$t_k = \begin{cases} 1.4 \cdot \left(0.5 - \frac{t_{\text{Drift},k}}{100\text{ ns}}\right) & t_{\text{Drift},k} \leq 100\text{ ns} \\ -0.7 & t_{\text{Drift},k} > 100\text{ ns} \end{cases} . \tag{5.1}$$

This means that short drift times correspond to the highest input values while large drift times to the lowest input values. For inactive TS within an event, the value of -0.7 is assigned such that they are similar to very large drift times.
The MLPs were trained and tested for each width at each sector. The testing process provides a $z_{\text{Out}} - z_{\text{Sim}}$ distribution from which the standard deviation σ and the mean

μ have been calculated. Some examples for different θ widths at different sectors are presented in Fig. 5.2.
The training and testing procedure for all experiments in these studies was repeated 10 times providing consequently 10 values for σ and μ. The mean and the standard deviation for these 10 values has been computed again using

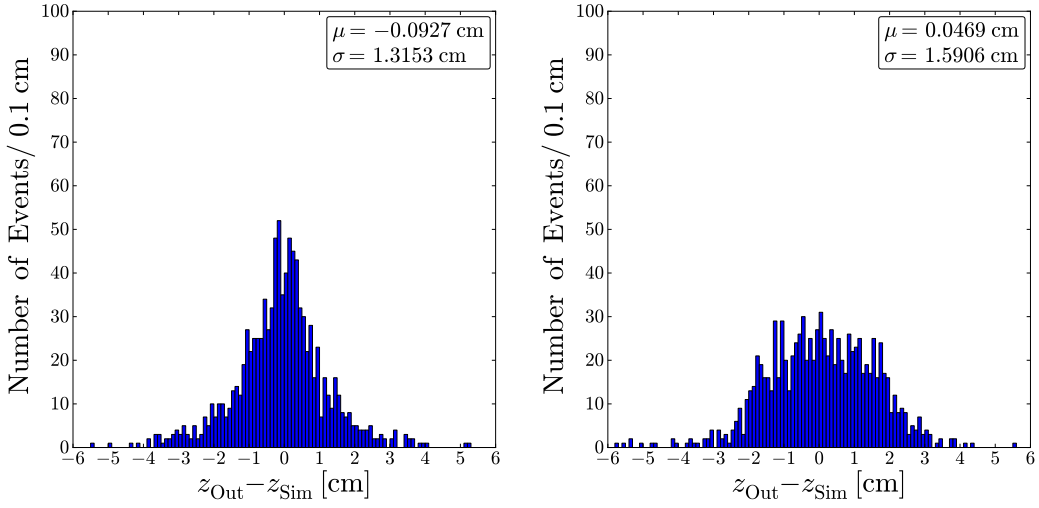
$$\mu(z_{\mathrm{Out}} - z_{\mathrm{Sim}}) = \frac{1}{10}\sum_{l=1}^{10}\mu_l \tag{5.2}$$

$$\sigma(\mu) = \sqrt{\frac{1}{10}\sum_{l=1}^{10}\left(\mu_l - \mu(z_{\mathrm{Out}} - z_{\mathrm{Sim}})\right)^2} \tag{5.3}$$

$$\sigma(z_{\mathrm{Out}} - z_{\mathrm{Sim}}) = \frac{1}{10}\sum_{l=1}^{10}\sigma_l \tag{5.4}$$

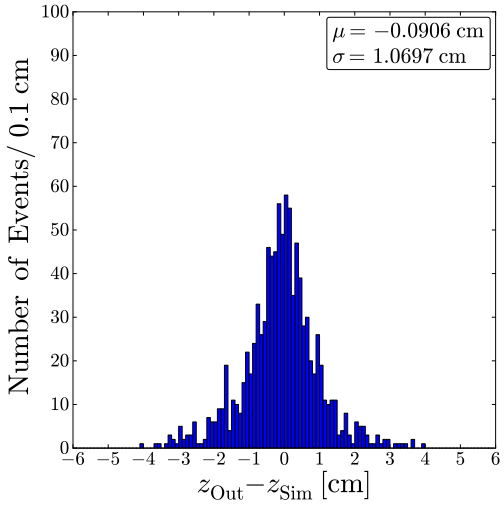
$$\sigma(\sigma) = \sqrt{\frac{1}{10}\sum_{l=1}^{10}\left(\sigma_l - \sigma(z_{\mathrm{Out}} - z_{\mathrm{Sim}})\right)^2} \tag{5.5}$$

The results for the different θ-sectors and θ-widths can be found in Fig. 5.3. The left column shows $\sigma(z_{\mathrm{Out}} - z_{\mathrm{Sim}})$ and the right columns $\mu(z_{\mathrm{Out}} - z_{\mathrm{Sim}})$. The errorbars on the left and on the right column plots correspond to $\pm\sigma(\sigma)$ and $\pm\sigma(\mu)$ respectively. This representation scheme will be used for all other experiments of the same kind in this work. As it can be recognized in the plots on the left columns, $\sigma(z_{\mathrm{Out}} - z_{\mathrm{Sim}})$ increases with the with of the θ region for all sectors. There is a nearly linear dependence at the beginning succeeded by a kind of saturation, or rather a less steep linear behavior. In the graphs 5.3(d) and 5.3(f) the lines for congruent sectors overlap as one may expect. This is not the case in 5.3(b), where the MLP in the sector $[40, 60]^\circ$ reaches a better resolution than in $[140, 120]^\circ$. For the sector in $[17, 40]^\circ$ the resolution limit of 2 cm was exceeded even for the smallest θ width. On the other hand, the graph of $\mu(z_{\mathrm{Out}} - z_{\mathrm{Sim}})$ in the right column shows that there is no relevant bias in all cases.
The reason why σ values are much larger for the sectors which lie closer to the z-axis has been understood afterwards: for particles flying at θ angles closer to 17°, the distance between the wire layers is larger than for those flying at θ angles near 90° (S. Fig. 5.4). Larger distances between wires result in larger intervals between drift times. Because of the threshold of 100 ns, more information is lost for those tracks, than for tracks whose θ angles lie near 90°.
After the first experiments for different region widths in $(p_{\mathrm{T}}, \phi, \theta)$, a drift time study, whose results will be presented in subsection 5.6, conduced to reach higher resolutions. However, the aim of this first experiments was not to determine the final region widths, but to get more realistic values for the track parameters, in order to proceed with the optimization of the MLP training parameters. The final region widths will depend on the accuracy of the predictor, which is the target of future investigations.

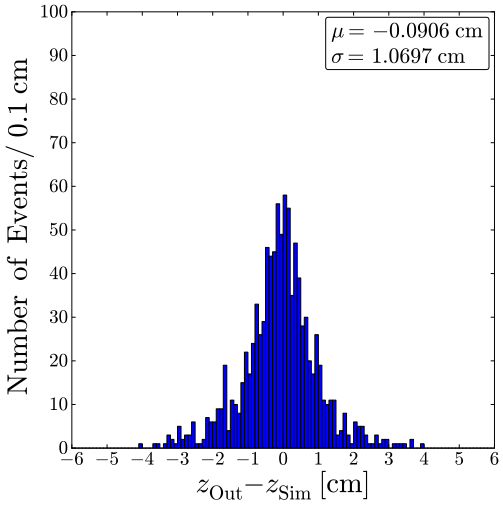


(a) $\theta \in [40°, 44°]$

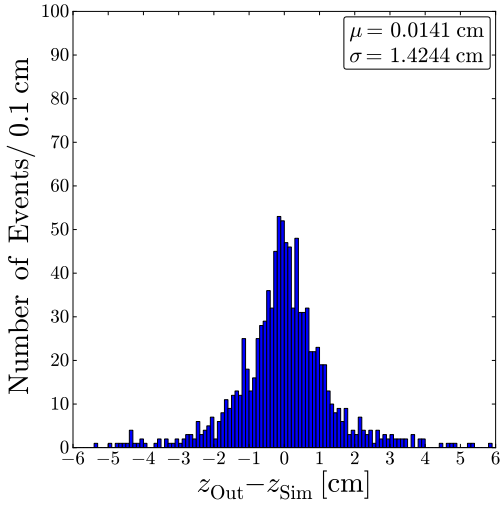
(b) $\theta \in [136°, 140°]$



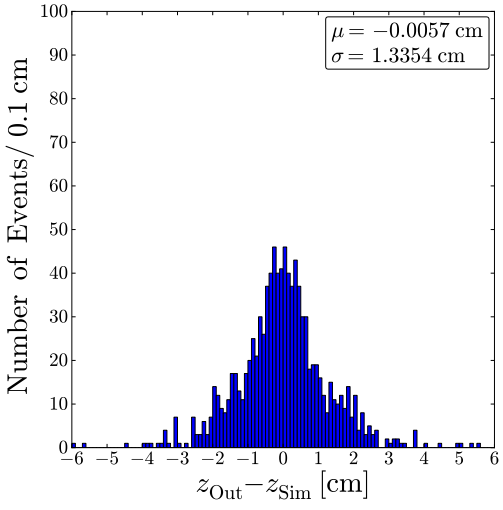
(c) $\theta \in [60°, 66°]$



(d) $\theta \in [114°, 120°]$



(e) $\theta \in [80°, 90°]$



(f) $\theta \in [90°, 100°]$

Figure 5.2: $z_{\text{Out}} - z_{\text{Sim}}$ distributions for different θ regions.

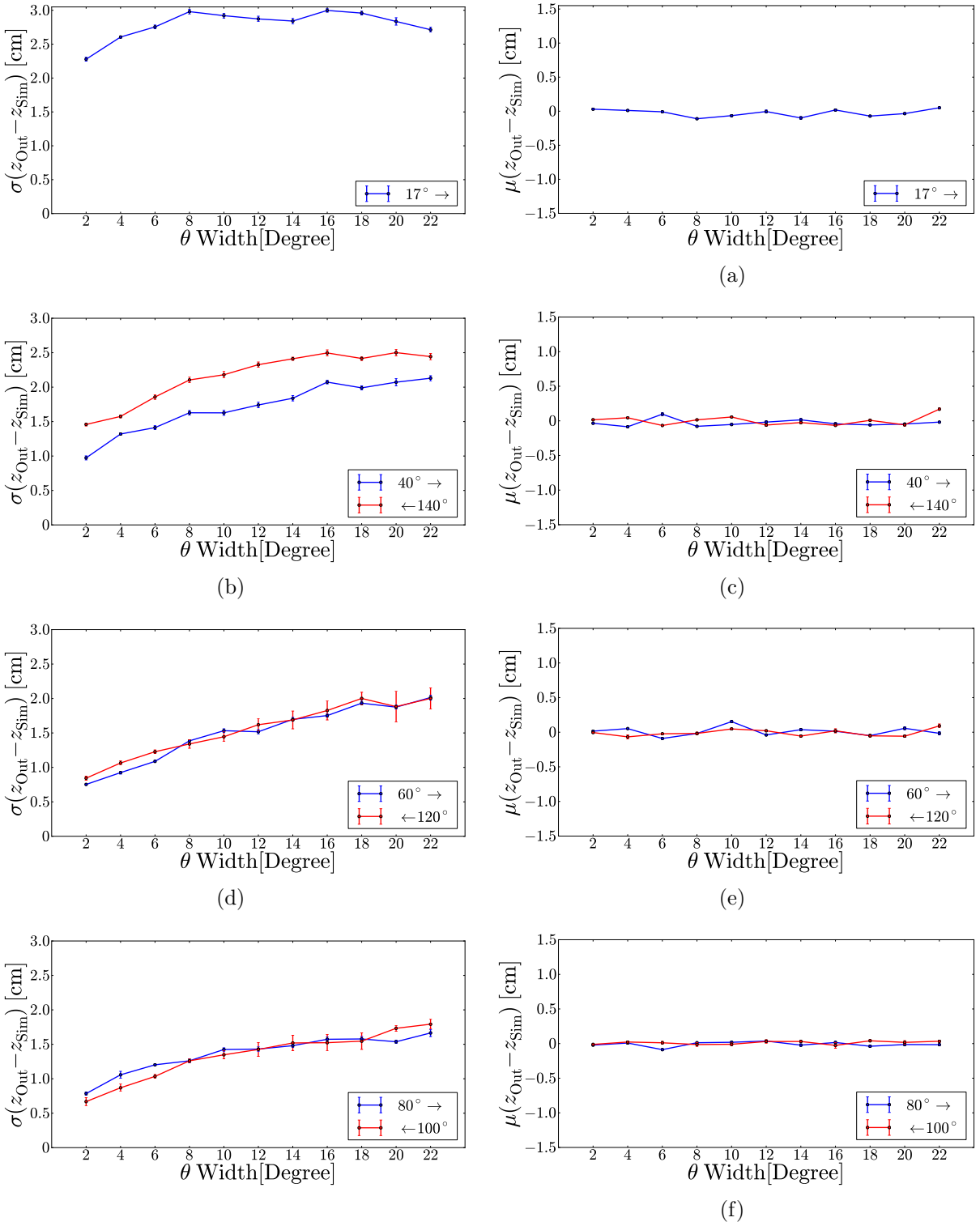


Figure 5.3: Standard deviation and mean value for different θ-widths at different θ-sectors. Each point in the graphs is the mean over 10 experiments, the error bar corresponds to the respective standard deviation.

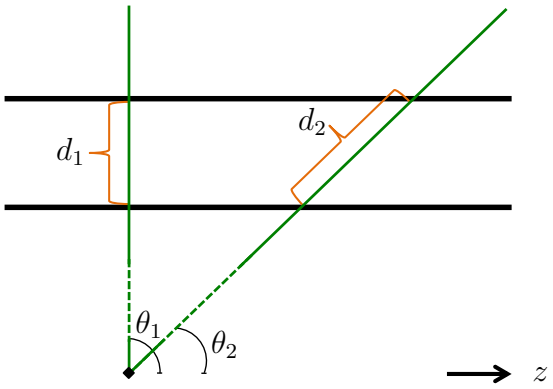


Figure 5.4: Distance between wires for particles flying at different θ angles. For angles closer to the z-axis the distance is larger resulting in larger time intervals between wire hits.

The region for which each MLP is specialized should overlap with the neighboring regions at the half width to improve the accuracy of the final z-vertex trigger (S. Fig. 5.5). If the width θ_{W} is known for a specific sector $[\theta_{\mathrm{min}}, \theta_{\mathrm{max}}]$, then the n_θ number of regions, or rather bins in this sector can be calculated as

$$n_\theta = 2 \cdot h_\theta = 2\frac{\theta_{\mathrm{max}} - \theta_{\mathrm{min}}}{\theta_{\mathrm{W}}} \tag{5.6}$$

On the basis of the results in Fig. 5.3, an approach for the total amount of MLPs, which are needed to cover the whole detector acceptance was done selecting the widest θ widths such that σ is not larger than 1.5 cm, except for the $[17, 30]^\circ$ sector where the smallest width has been selected. The details are summarized in Table 5.1. This approach leads to a number of 57 bins in θ.

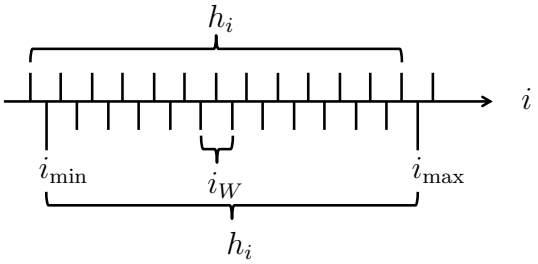


Figure 5.5: Overlap of the MLP bins for a general track parameter $i \in \{p_{\mathrm{T}}, \phi, \theta\}$ and the total bin number within the sector $[i_{\mathrm{min}}, i_{\mathrm{max}}]$.

5.2 Track Parameters for charged Particles

As argued before, the total transverse momentum space of the particles generating tracks in the CDC is determined by physics. Because of the magnetic field of the solenoid only

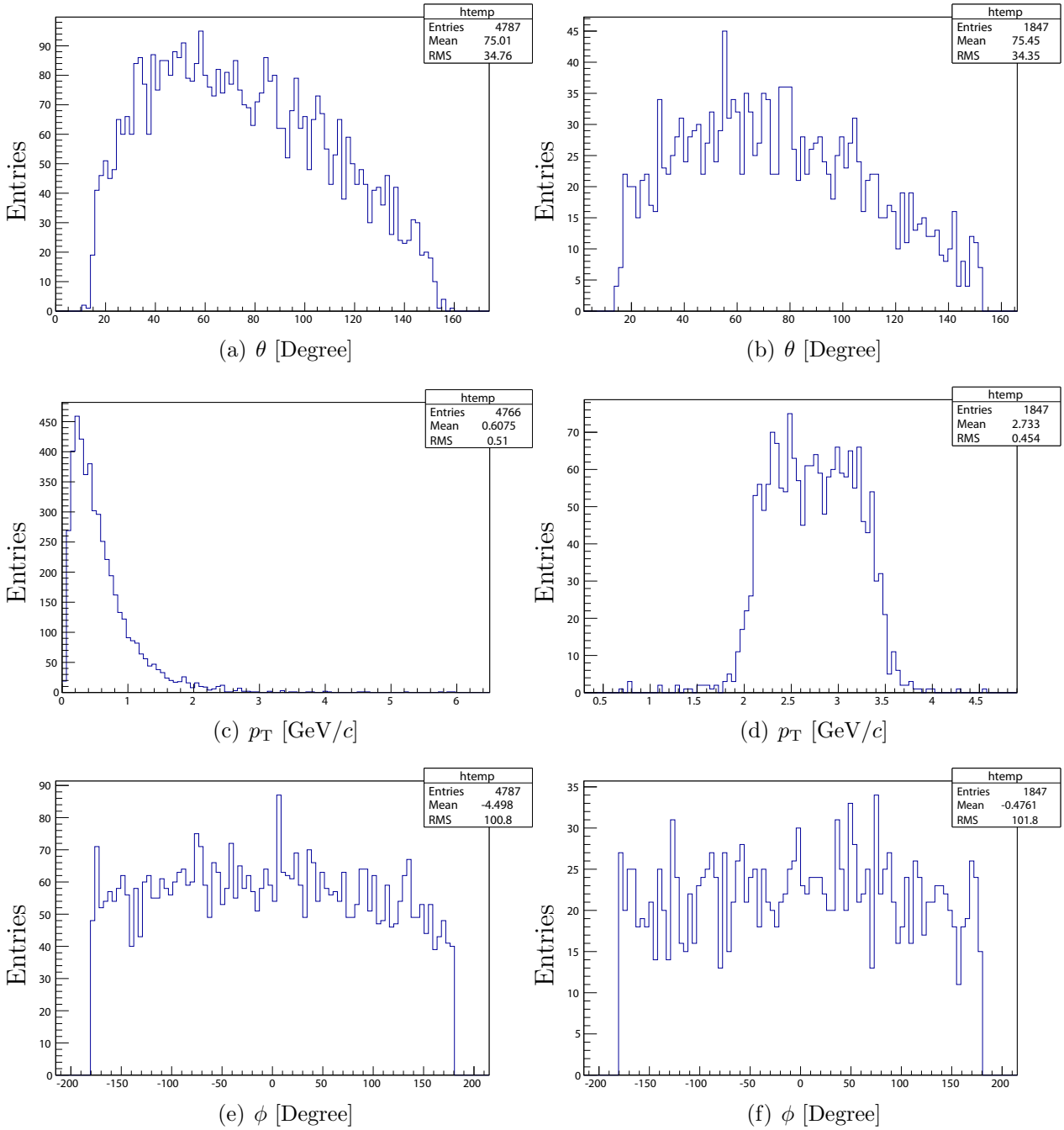
θ Sector		θ Width	Amount of Nets
17° - 30°		2°	13
30° - 50°	150° - 130°	4°	2·10
50° - 80°	130° - 100°	6°	2·10
80° - 90°	100° - 90°	10°	2·2
		Total	57

Table 5.1: Proposal for different θ widths at different sectors.

particles with at least $p_T \sim 50$ MeV can reach the innermost CDC layer. Nervetheless, for the neural z-vertex trigger project only particles with $p_T \geq 0.2$ GeV will be considered.
For the optimization studies done within the scope of this work it was essential to select the most relevant track parameter regions. Therefore, some physical analysis have been done in order to study the track parameter distribution for charged particles stemming from simulated B decays.
The analysis was done using the "Evt-Gen-Input" Module of the basf2 software: an amount of 1500 B decay events were generated and all charged particle tracks were selected. Subsequently, histograms for each track parameter (p_T, ϕ, θ) were produced. In the first analysis all charged particle tracks from all possible B decay channels were considered. In the second analysis only the decay channel $B^0 \to \pi^+\pi^-$ was taken into account because these are the pions from B decays with the highest momentum. Other reason motivating the use of pions is that they will produce single tracks in the CDC. The results can be found in Fig. 5.6.
The histograms in 5.6(a) and 5.6(b) show an asymmetric distribution in θ which is expected because of the aforementioned boost caused by the asymmetric beam energies. This distribution together with the approach for the θ width in 5.1 motivated the use of the region $[56, 62]°$ for all further experiments in this study.
The distribution of p_T reveals that the most larger amount of charged particle tracks are expected in the range $p_T \in [0, 2]$ GeV/c (S. Fig. 5.6(c)) and that the highest transverse momentum for pions is ~ 4 GeV/c.
On the other hand, ϕ is the only variable being distributed uniformly for the whole possible space $[0, 360]°$. This was expected since the detector, as well as the physics in Belle II, is symmetric in ϕ.

5.3 Estimation of the p_T-Binwidth

The next step was the estimation of the region widths for p_T. The procedure, the scaling and the training parameters for the MLP were the same as those of the experiment described in subsection 5.1. The difference was that the events were generated for π^- (PDG-Code=-211) with a uniformly distributed vertex in the region $\theta \in [56, 62]°$ and, that p_T was varied instead of θ. The momentum space was divided into 6 sectors: $p_T \in [0.5, 1.2]$ GeV/c, $p_T \in [1.2, 2.2]$ GeV/c, $p_T \in [2.2, 3.2]$ GeV/c, $p_T \in [3.2, 4.2]$ GeV/c, $p_T \in [4.2, 5.2]$ GeV/c and $p_T \in [5.2, 10.0]$ GeV/c.



(a) θ [Degree] (b) θ [Degree]

(c) p_{T} [GeV/c] (d) p_{T} [GeV/c]

(e) ϕ [Degree] (f) ϕ [Degree]

Figure 5.6: Monte Carlo distributions for the track parameters of charged particle tracks stemming from generic B decays (left columns) and from $B^0 \to \pi^+\pi^-$ decays (right column).

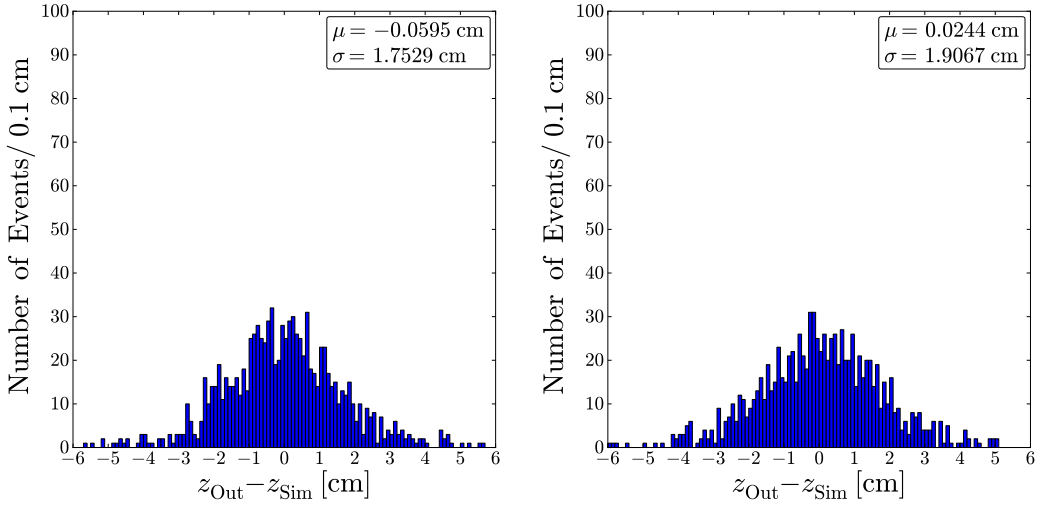
p_T-Sector (GeV/c)	p_T-Width (GeV/c)	Amount of Nets
0.2 - 1.2	0.05	40
1.2 - 2.0	0.4	4
2 - 5	0.6	10
5 - 10	5	1
	Total	55

Table 5.2: Proposal for different p_T widths at different sectors.

The upper limit of 10 GeV/c is nearly the cms momentum. Although it is improbable to find a particle with a transverse momentum of 10 GeV/c in Belle II , high momentum tracks with $p_T \gtrsim 7$ GeV/c have a nearly vanishing curvature and thus they degenerate to straight tracks.
In the first sector the considered p_T widths were $\{0.05, 0.10, 0.15\}$ GeV/c. They are very small in this sector since larger regions lead to larger differences in track curvature and thus to larger numbers of active TS for the input layer of the MLP. For all other sectors the p_T widths were $\{0.1, 0.2, 0.3, ..., 1.0\}$ GeV/c. The tracks were uniformly distributed within these regions.
Some examples of $z_{Out} - z_{Sim}$ distributions after the testing process can be found in Fig. 5.7. The results over 10 training and testing experiments are shown in 5.8.
As it can be seen in the plots, $\sigma(z_{Out} - z_{Sim})$ does not increase considerably for larger p_T widths except for the 1.2 GeV/c sector (S. Fig. 5.8(c)). On the other hand, the $\mu(z_{Out} - z_{Sim})$ plots show that there is no bias.
Again, using the same criterion as for the θ widths, an approach for the different bin widths in p_T has been done selecting the widest p_T widths such that the limit of $\sigma \leq 1.5$ cm is not exceeded, excepting the lowest momentum sector where the smallest width has been selected (S. Fig. 5.8(a)). In the case of the highest momentum, the MLP reaches a good resolution also for a bin size as wide as the sector (S. Fig. 5.8(f)). The details are listed in Table 5.2. The total amount of bins in the approach for p_T is 55.

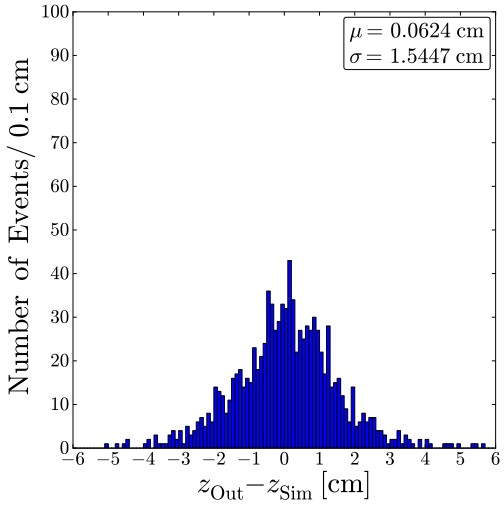
5.4 Estimation of the ϕ-Binwidth and Symmetry Check

The procedure in the last subsection 5.3 was repeated again with the same scaling and training parameters, but varying ϕ instead of p_T. The training and testing events were generated with a uniform distribution in the region $p_T \in [1.6, 2.0]$ GeV/c. The effect of the ϕ widths $\{0.2, 0.4, 0.6, ..., 2.0\}$ was studied at 4 different points of the whole ϕ space: $\{0°, 90°, 180°, 270°\}$.
Examples of $z_{Out} - z_{Sim}$ distributions after the testing process are shown in Fig. 5.9 and the results over 10 experiments in 5.10. The graphs of $\sigma(z_{Out} - z_{Sim})$ remain more or less constant within the considered ϕ widths at each of the 4 points and they practically overlap, as one may expect by reason of the detector and the physics symmetry in ϕ. The $\mu(z_{Out} - z_{Sim})$ plot shows that the MLP output was not biased.

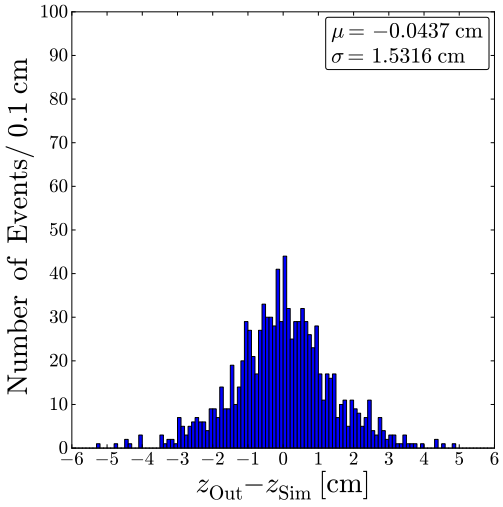


(a) $p_{\mathrm{T}} \in [0.50, 0.55]$ GeV/c

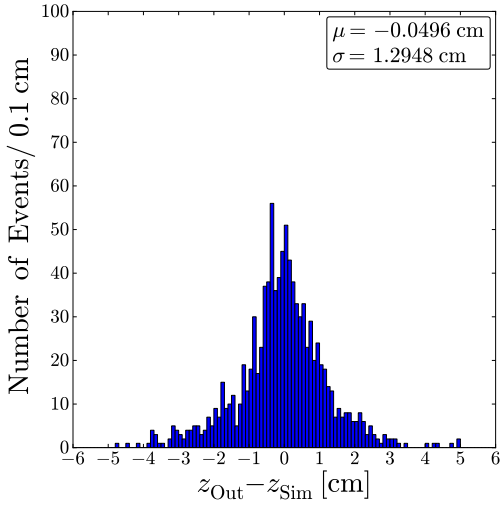
(b) $p_{\mathrm{T}} \in [0.50, 0.65]$ GeV/c



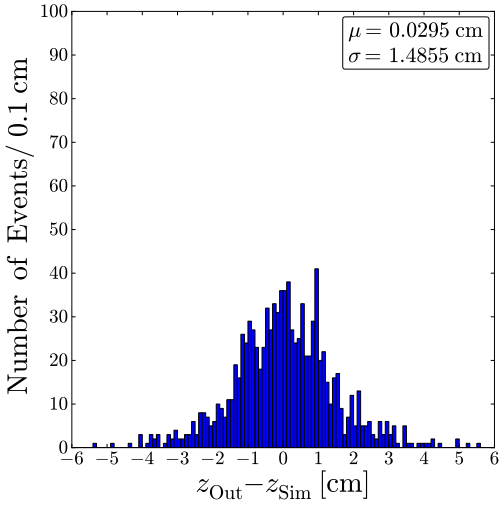
(c) $p_{\mathrm{T}} \in [1.20, 1.60]$ GeV/c



(d) $p_{\mathrm{T}} \in [2.20, 2.80]$ GeV/c



(e) $p_{\mathrm{T}} \in [4.20, 4.80]$ GeV/c



(f) $p_{\mathrm{T}} \in [5.0, 10.0]$ GeV/c

Figure 5.7: $z_{\mathrm{Out}} - z_{\mathrm{Sim}}$ distributions for different p_{T} regions.

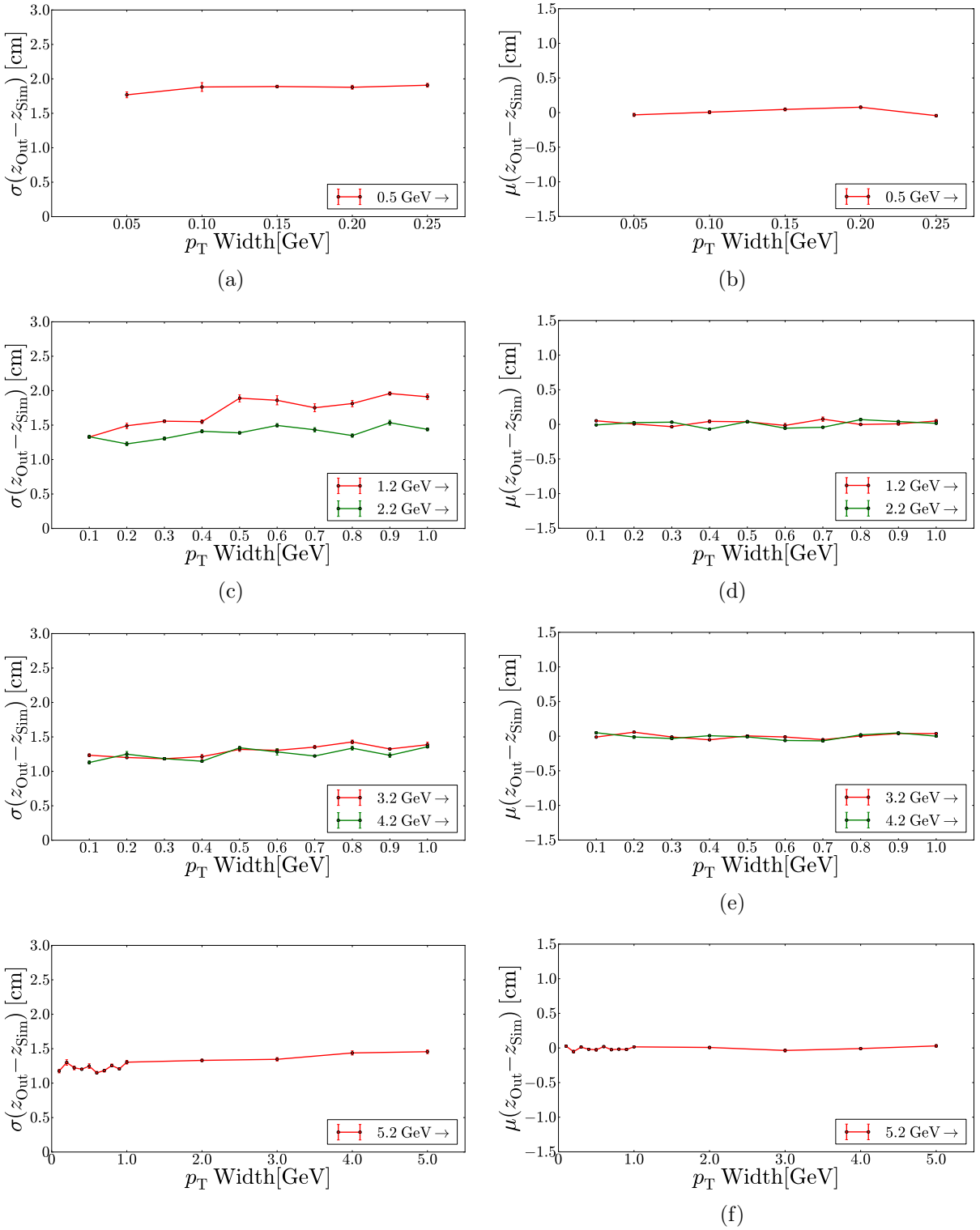
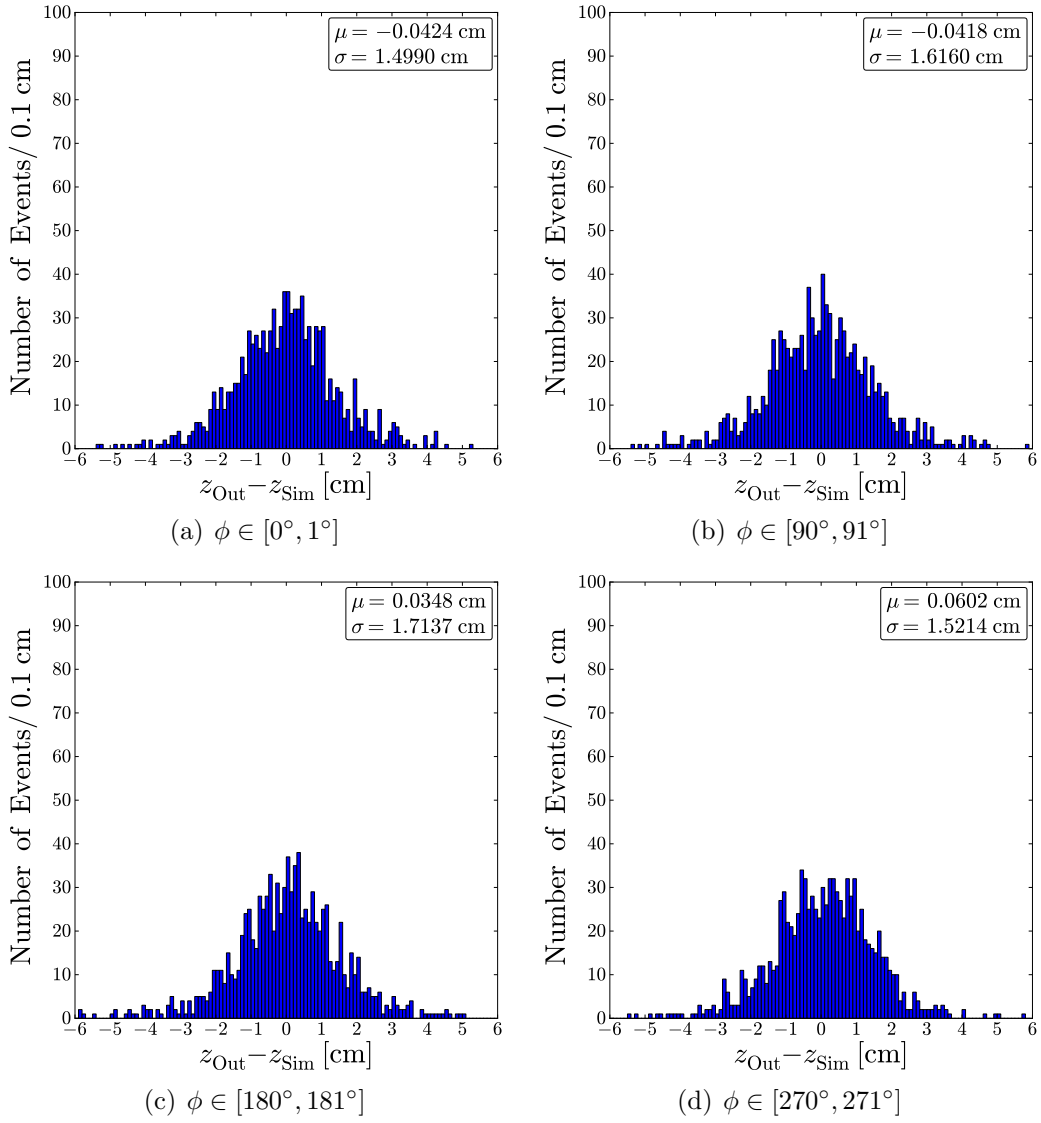


Figure 5.8: Standard deviation and mean value for different p_{T}-widths at different p_{T}-sectors. Each point is the mean over 10 experiments.



(a) $\phi \in [0°, 1°]$ (b) $\phi \in [90°, 91°]$

(c) $\phi \in [180°, 181°]$ (d) $\phi \in [270°, 271°]$

Figure 5.9: $z_{\text{Out}} - z_{\text{Sim}}$ distributions for different ϕ regions.

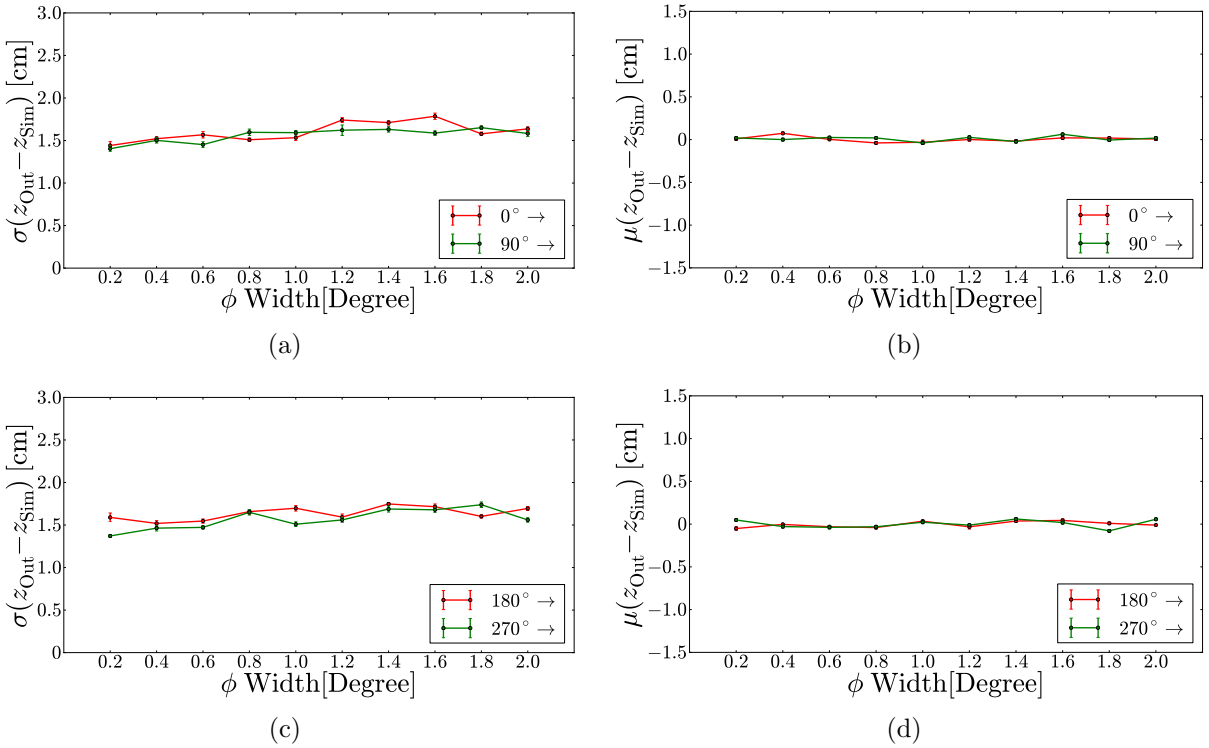


Figure 5.10: Standard deviation and mean value for different ϕ-widths at different ϕ-sectors. Each point is the mean over 10 experiments.

Previous studies for this project [1] show that a high prediction accuracy in ϕ can be achieved by the 3D Trigger and the look-up table approach. Thus, a ϕ bin width of 1° for all MLPs is a good approach and it conduces to an amount of 720 bins.

5.5 Estimation of the total Net-Hedgehog Size

The total amount of networks, i.e. MLPs, can be referred too as "Net-Hedgehog" since one can imagine each MLP, which is specialized for a narrow region, as a spine of a hedgehog. Combining the results presented in the subsections 5.1, 5.3 and 5.4 the total amount n_{Total} of nets to cover the whole $(p_{\text{T}}, \phi, \theta)$ space can be estimated:

$$n_{\text{Total}} = 57 \cdot 55 \cdot 720 = 2,257,200 \sim 2 \cdot 10^6. \tag{5.7}$$

Considering that negatively charged particles have a different curvature as those charged positively, this value has to be calculated by 2. It should be also not forgotten that in these studies only the range $z_0 \in [-10, 10]$ cm is covered. For more z_0 bins the n_{Total} value gets multiplied by the number of bins in z_0.
This approach is relevant for the hardware implementation too. Since a single MLP requires approx. 5 kBytes storing space [1], the total required space for the complete Net-Hedgehog will be about $2n_{\text{Total}} \cdot 5$ kB$\sim$ 20 GB.

5.6 Drift time and Random Offset

After making a histogram of the priority drift time provided by the "TRG-CDC-TS-Stream" Module, it came to light that the threshold of 100 ns included in the scaling method described in subsection 5.1 was too low, causing in consequence a significant loss of information. Fig. 5.11(a) shows clearly how large this information loss was.
For the next experiments, the drift time has been treated such that the MLP receives an input information as similar as possible to the information which will be provided by the final version of the TSF described in subsection 3.4.1. In the simulated single track events, the absolute zero point in time is known and the drift time t_0 of the first hit corresponds to the exact time interval between $t = 0$ and the time point of the first hit $t = t_0$. The drift time information is also supplied with a 1 ns resolution.
In the real case of the clock incorporated in the TSF, the zero point is the priority time of the first active TS $t_{\text{Clock},0}$ and not the absolute zero point $t = 0$. Therefore, the clock times get a delay of the size of the interval between $t = 0$ and $t_{\text{Clock},0}$. Since there is no information about the size of this interval it can be considered as a random value and called thus random offset. Additionally, the clock time information has a 2 ns resolution.
In order to give the MLP a more realistic time input, the 1 ns resolution of drift times supplied by the "TRG-CDC-TS-Stream" Module was reduced to a 2 ns resolution. This was done dividing the drift time values $t_{\text{Drift},k}$ by 2 ns and rounding the fraction to an integer such that the time information corresponds to that of the TSF clock, i.e.

$$t_{\text{Clock},k} = \text{round}\left(\frac{t_{\text{Drift},k}}{2\ \text{ns}}\right). \tag{5.8}$$

A Histogram of the times $t_{\text{Clock},k}$ is shown in Fig 5.11(b); it has the same shape as that of $t_{\text{Drift},k}$ but with a wider bin width of 2 ns. Subsequently, the value of the first clock time $t_{\text{Clock},0}$ is subtracted and a random offset t_{Offset} is added. The distribution of $t_{\text{Clock},0}$ and t_{Offset} can be seen in Fig. 5.11(c). To generate the random offset a uniform distribution was chosen, such that its mean value is in the same order of the mean of the $t_{\text{Clock},0}$ distribution.
To produce the input for the MLP the scaling was done using a similar formula to that used in subsection 5.1, but with a higher threshold of 175 clock bins and a range of $[-1, 1]$, it is given by

$$t_k = \begin{cases} 2 \cdot \left(0.5 - \frac{t_{\text{Clock},k} - t_{\text{Clock},0} + t_{\text{Offset}}}{175\ \text{ns}}\right) & t_{\text{Clock},k} \leq 175\ \text{bins} \\ -1.0 & t_{\text{Clock},k} > 175\ \text{bins} \end{cases}. \tag{5.9}$$

An example of a histogram for t_k is shown in Fig. 5.11(d). For inactive TS within an event, the value -1.0 was also assigned. This treatment and scaling of the time information was applied for all subsequent experiments in this work.

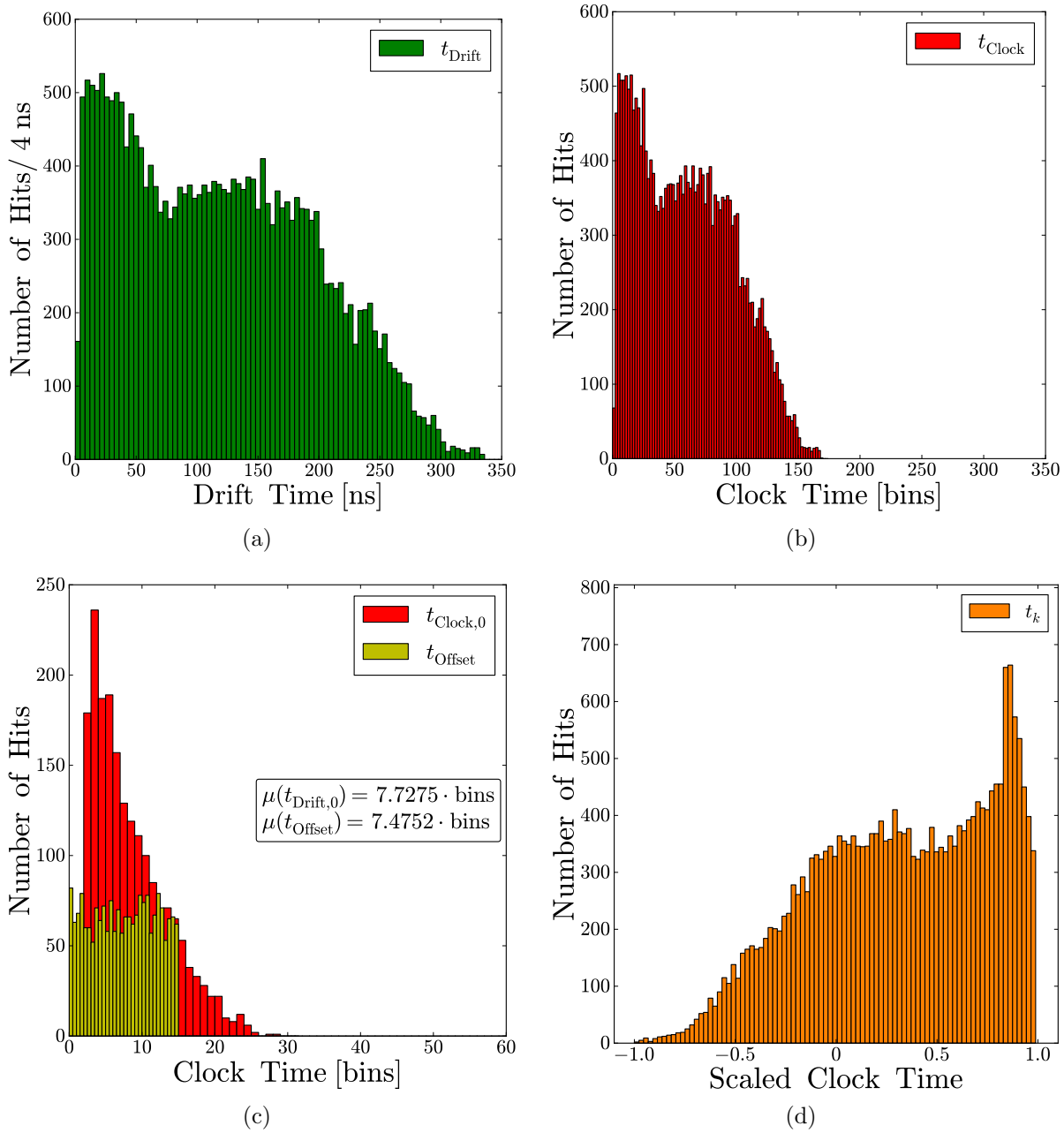


Figure 5.11: Different time distributions: a) drift time provided by the "TRG-CDC-TS-Stream" Module; b) drift time with the resolution of the TSF clock, a bin correspond to 2 ns; c) the first priority time and the random offset; and d) the time input for the MLP.

5.7 Optimization of the Input Layer Size

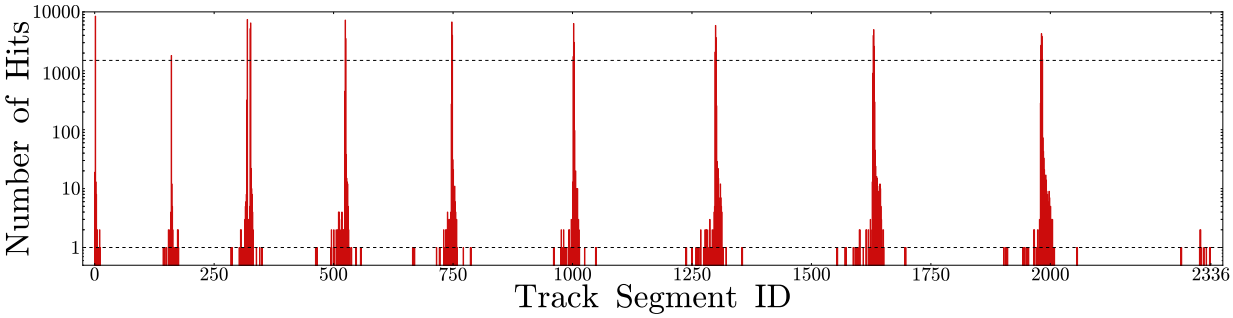
As it has been explained in subsection 4.2 the input and hidden layer size in a MLP, i.e. the number of neurons contained in each layer, determine the total amount of weights $w_{i,k}$ and $w_{\mathrm{Out},i}$ which can be calculated as

$$n_{\mathrm{total}}^{\mathrm{weights}} = n_{\mathrm{output}}^{\mathrm{weights}} + n_{\mathrm{hidden}}^{\mathrm{weigths}} = (n_{\mathrm{input}} + 1)\, n_{\mathrm{hidden}} + n_{\mathrm{hidden}} + 1. \tag{5.10}$$

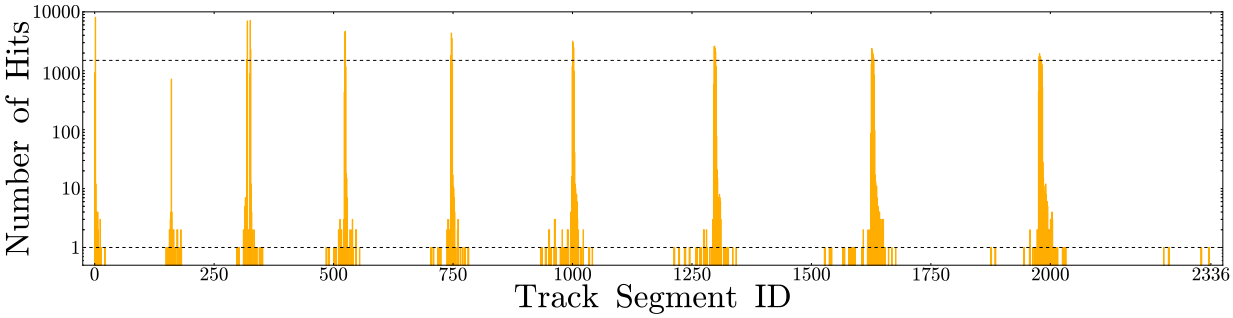
For the neural z-vertex trigger approach the number of nodes in the hidden layer are set to $n_{\mathrm{hidden}} = f_{\mathrm{hidden}} \cdot n_{\mathrm{input}}$, where f_{hidden} is the so-called hidden factor. Replacing this in equation 5.12 yields

$$\begin{aligned} n_{\mathrm{total}}^{\mathrm{weights}} &= (n_{\mathrm{input}} + 1)\, f_{\mathrm{hidden}} n_{\mathrm{input}} + f_{\mathrm{hidden}} n_{\mathrm{input}} + 1 && (5.11)\\ &= f_{\mathrm{hidden}} \left(n_{\mathrm{input}}^2 + 2 n_{\mathrm{input}}\right) + 1. && (5.12) \end{aligned}$$

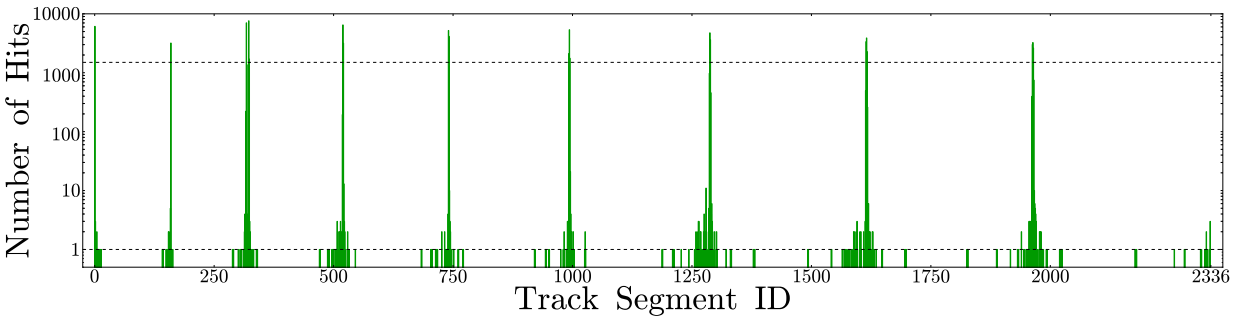
This means that an increase of n_{input} conduces to a quadratic increase of the total amount of weights. In turn, a large amount of weights results in a long training time, because there are more operations to be done at each iteration step. On the other hand, larger amounts of weights bring the MLP to require more space to be stored, since most of the memory is used to store the weight values. Long training times can be mitigated with higher computing power, but larger storing requirements worry for the hardware implementation since the memory space available in FPGAs is limited.
The question for this experiment was: how small can the input layer size be such that a resolution of at least 1.5 cm can still be reached by the MLP? As argued before the input neurons correspond to the relevant TS in the region for which the MLP is specialized. In consequence, the aim of this study was to determine a good selection criterion such that the input layer size is as small as possible, but at the same time exploiting maximally the provided TS information.
For the next experiments 8 event samples were generated to train and test the MLPs covering the whole transverse momentum space, but using the widths approaches proposed in section 5.3 and 5.1. For the lowest momentum sector three widths have been explored: $p_{\mathrm{T}} \in [0.50, 0.55]$ GeV/c, $p_{\mathrm{T}} \in [0.50, 0.60]$ GeV/c and $p_{\mathrm{T}} \in [0.50, 0.65]$ GeV/c. For the other sectors only one region has been considered: $p_{\mathrm{T}} \in [1.20, 1.60]$ GeV/c, $p_{\mathrm{T}} \in [2.20, 2.80]$ GeV/c, $p_{\mathrm{T}} \in [3.20, 3.80]$ GeV/c, $p_{\mathrm{T}} \in [4.20, 4.80]$ GeV/c and $p_{\mathrm{T}} \in [5.20, 10.0]$ GeV/c. Single π^- tracks were generated uniformly distributed in these regions as well as in $\theta \in [56, 62]^\circ$ and $\phi \in [0, 1]^\circ$. The centering of the research focus on the low momentum sector as well as the choice of the θ region were motivated by the physical analysis presented in section 5.2.
The first step was to generate hit histograms for all TS in each one of the regions, in order to study the firing patterns. Four examples are shown in Fig. 5.16. At the first glance, a kind of periodic peak pattern can be recognized, which is the consequence of the super layer arrangement. There are several TS who fired only a few times, but only a small amount of TS received the most of the hits. These TS correspond to the relevant input channels which need to be selected.



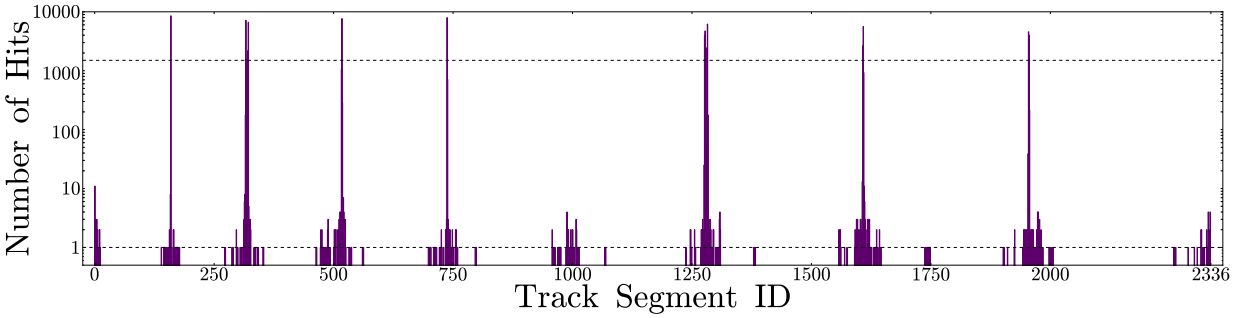
(a) $p_{\mathrm{T}} \in [0.50, 0.55]$ GeV/c



(b) $p_{\mathrm{T}} \in [0.50, 0.65]$ GeV/c

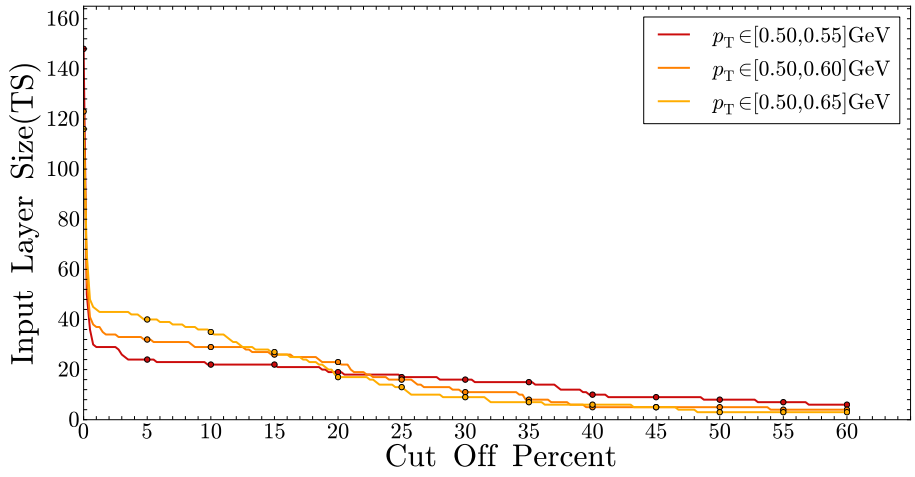


(c) $p_{\mathrm{T}} \in [1.20, 1.60]$ GeV/c

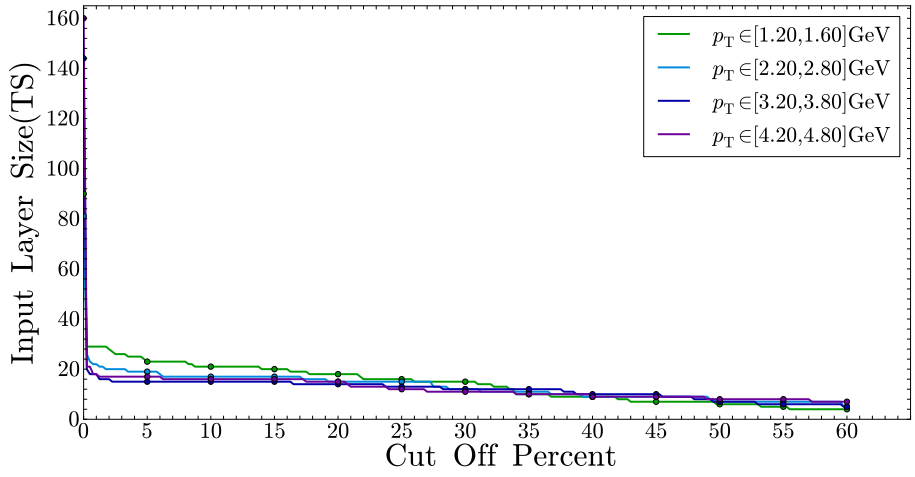


(d) $p_{\mathrm{T}} \in [5.2, 10.0]$ GeV/c

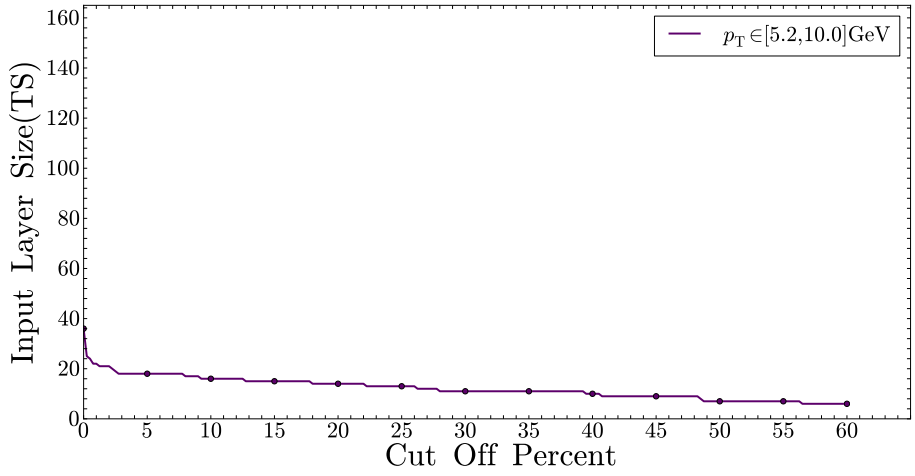
Figure 5.12: Active track segments in 10,000 events. The lower grid line represents a Cut Off percent of 0.01% and the upper grid line a Cut Off percent of 15.00%.



(a)



(b)



(c)

Figure 5.13: Number of active TS, i.e. Input-Layer size, as a function of the Cut Off Percent.

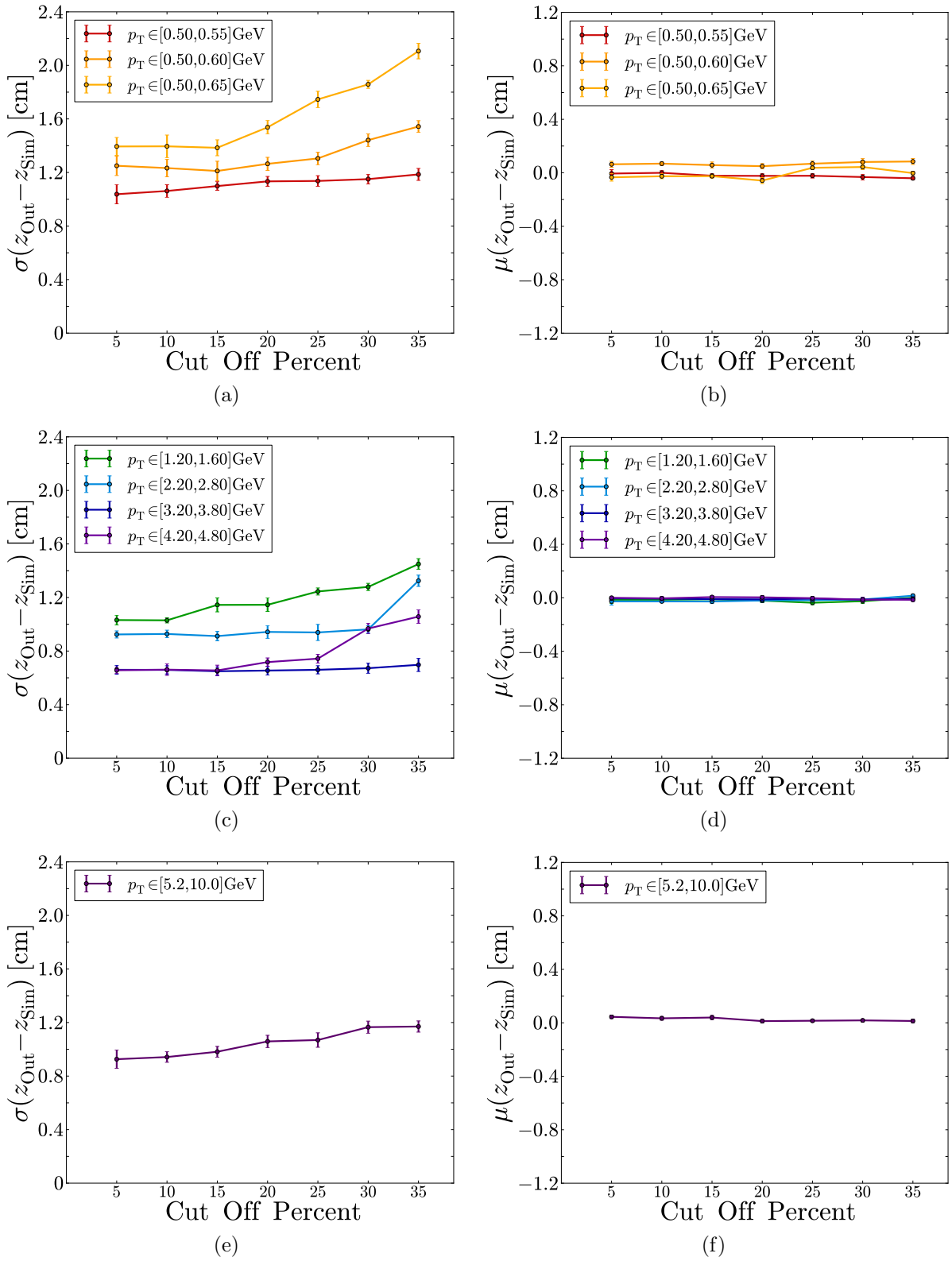


Figure 5.14: Standard Deviation (Resolution) and Mean for different p_T-regions as a function of the Cut Off Percent.

The used selection criterion was to make a cut off, neglecting TS which fired less often than in a given percent of all events. This cut off percent can be then represented as a horizontal line in the hit histogram with a specific height. The TS which are selected are those whose peak is higher than the line. The number of input neurons is then the amount of wires who exceeded the cut of percent. A graph of the input layer size in dependence of the cut off percent value can be seen in Fig. 5.13. As argued before, taking all TS in each region without selection criterion results in very large input layer sizes of about 150 neurons.
To evaluate the performance of the MLP, a training and testing process was carried out 10 times for the cut off percents $\{5, 10, 15, ..., 35\}\%$, but fixing always the number of hidden neurons to 40. The results for $\sigma(z_{\text{Out}} - z_{\text{Sim}})$ and $\mu(z_{\text{Out}} - z_{\text{Sim}})$ are presented in Fig. 5.14. These values were calculated and plotted making use of the same procedure and representation scheme described in section 5.1.
In the lowest momentum sector (S. Fig. 5.14(a)) the performance of the MLP is more sensible to the cut off percent for the wide p_{T} regions. The yellow curve increases nearly linear starting at 15%. This can be explained as follows: in the low momentum sector, a large region width leads to a considerable amount of possible track curvatures and thus larger regions include many possible TS. Subsequently, the peaks become broader and less high, such that they do not overcome the cut off. This effect can be observed comparing the histograms 5.12(a) and 5.12(b). For high momentum regions or for narrow defined low momentum regions the relevant TS are always selected even for high cut off percents since they have very high narrow peaks. This results in a more stable MLP performance for those regions.
For all 8 p_{T} regions, very good σ values were reached for cut off percents between 5% and 15%. Since the aim is to chose the highest possible cut off percent with good results, the value of 15% was utilized for the succeeding experiments. This leads to input layer sizes between 16 and 28 neurons depending of the region. (S. again Fig. 5.13). Additionally, the right column of Fig. 5.14 shows that there is also no significant bias for all results.

5.8 Optimization of Hidden-Layer Parameter

The purpose of this experiment was to find an optimal hidden layer factor f_{hidden}. The same event samples used for the optimization of the input layer size described in the previous section served again as training and testing set.
The same procedure of training and testing over 10 times was repeated. The difference here is that a cut off percent of 15% was fixed and the hidden layer factor was varied within the values $\{1, 1.5, 2.5, ..., 5\}$. Since $n_{\text{hidden}} = f_{\text{hidden}} \cdot n_{\text{input}}$ must be an integer, the value had to be rounded in some cases. The results of this experiment are shown in Fig. 5.15 using the usual representation. It can be recognized that the different f_{hidden} values had no significant repercussion on the performance of the MLP and that the output was not biased. The relevant outcome is that the best results were achieved for $f_{\text{hidden}} \geq 2$. Notwithstanding, for training factors $f_{\text{hidden}} \geq 4$ the training process took much longer than for other experiments. Thus, for the last experiment, the hidden factor was $f_{\text{hidden}} = 3$.

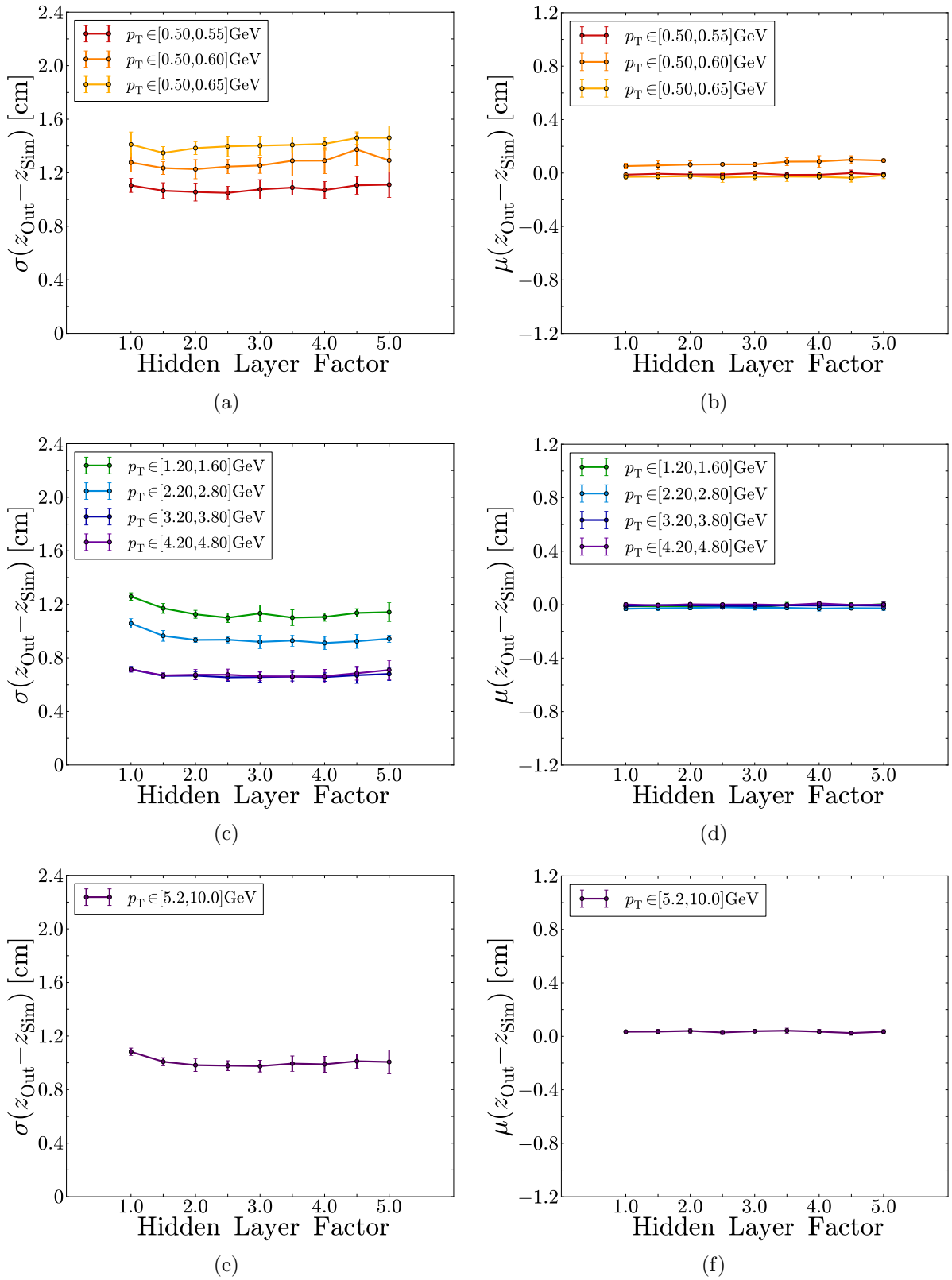
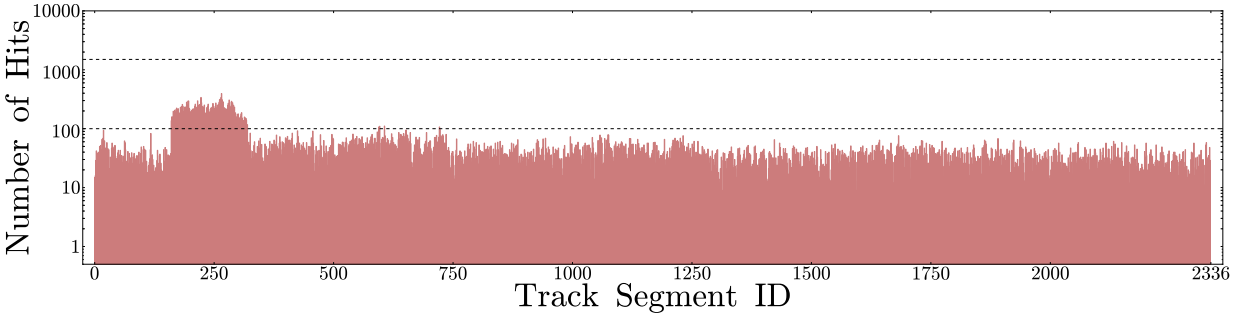


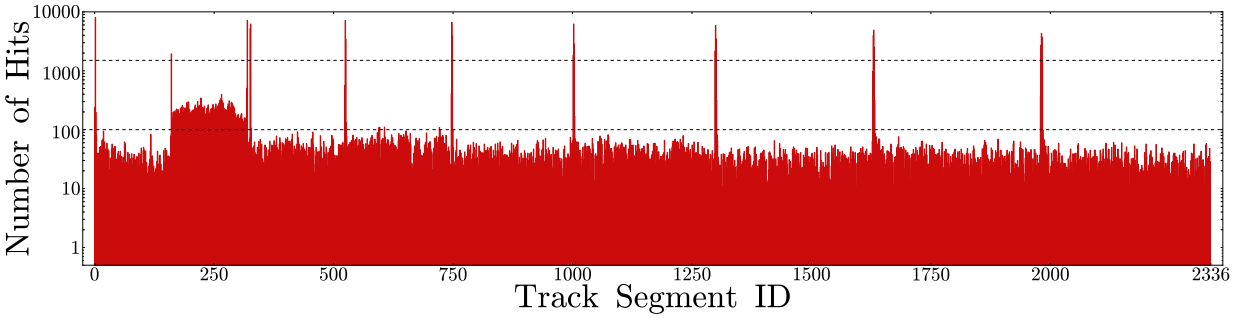
Figure 5.15: Standard Deviation (Resolution) and Mean for different p_T-regions as a function of the hidden factor f_{hidden}.

5.9 The Effect of Background and of a light (x, y)-Displacement

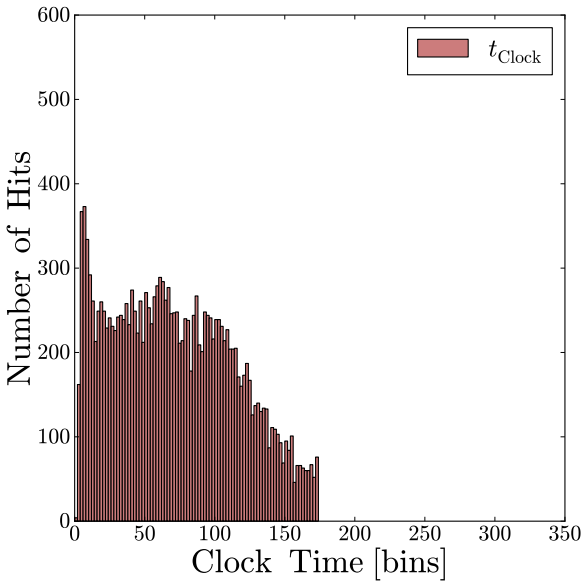
So far in this project, the MLP have been trained and tested using single track events without background, .i.e. all the hit information in each event belongs completely to the particle track. As explained before in section 3.3, there is a large amount of background expected in Belle II, mainly due to Touschek and beam gas scattering. Since the neural z-vertex trigger will operate under real detector conditions, it was imperative to study the effect of Background on the performance of the MLP. This was exactly the main object of this work.
Nevertheless, to produce the most significant results as possible, this study had to be carried out for MLPs trained for realistic track parameter regions, using realistic training parameters. The previous studies in this project [1] provided a proof of the capability of the MLPs using narrow track parameter regions and making assumptions for many training parameters and scaling methods. Thus, it was needed to make an approach for the final attributes of the MLP, before training and testing the MLP with background. That is the reason, why the previous experiments were executed in advance: the approach for the region widths in the whole $(p_{\mathrm{T}}, \phi, \theta)$ space was presented in sections 5.1−5.4, the choice of a good scaling method for a more realistic time input was explained in section 5.6, the choice of a good selection criterion for the input channels was made in section 5.7 and the size factor for the MLP hidden layer was determined in section 5.8.
After defining all these properties and methods, the next step was to generate events with background. This was done using the MixBkg Module of the basf2 software, whose operating mode was explained in subsection 3.6. Again, 8 samples with the same track parameters as those used to optimize the MLPs (section 5.7) were generated, but turning the MixBkg module on.
The hit distribution for $10,000$ pure background events is shown in Fig. 5.16(a). As it can be recognized, a cut off percent of 1% filters out almost all background hits. Since the used cut off percent is 15% there will be no selected input channels supplying pure background. On the other hand, for single track event samples with background, the TS which overcome the cut off percent will also receive background hits because of the nearly uniform distribution of the background (S. Fig. 5.16(b)).
The drift time distribution of pure background events can be observed in Fig. 5.16(c) and that of the first TS hit in Fig. 5.16(d). The background drift time distribution is not uniform at all, but it is more uniform than that of single tracks without background (Compare with Fig. 5.11(b) in section 5.6). Because of this lightly uniform behavior, selecting always the first hit if a TS fires two times in an event will also reject some background hits. Indeed, the final version of the TSF will select the first hit if a wire fires many times within the period of 16 clock cycles since each hit is widened during this period. This was explained in subsection 3.4.1.



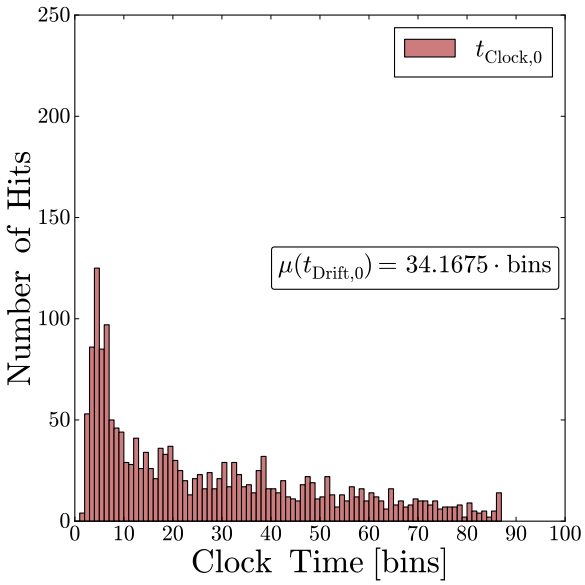
(a) Particle-Gun off, MixBkg on



(b) $p_{\mathrm{T}} \in [0.50, 0.55]$ GeV/c, MixBkg on



(c) Drift time with TSF resolution



(d) First priority time

Figure 5.16: a)&b) Active track segments in 10,000 events. The lower grid line represents a Cut Off percent of 1.0% and the upper grid line a Cut Off percent of 15.0%. c)&d) Time distributions for pure background events, a bin corresponds to 2 ns.

Another issue that needed to be studied is the effect of a light vertex displacement on the (x, y)-plane, i.e. a light displacement perpendicular to the z-axis. The previous studies [1] have only considered the case where all track vertices lie on the z-axis, that is $(x, y) = 0$ in all events. Nevertheless, since the beam focusing will have not been fully optimized in the early operation stages of SuperKEKB [6], the neural z-vertex trigger will also have to deal with light vertex displacements on the (x, y)-plane. Therefore, 8 data samples, within the same regions mentioned above, were generated with a normally distributed displacement around $(x, y) = 0$. The standard deviation of the used distribution is $\sigma_x = \sigma_y = 0.7$ cm. This means that about 15% percent of all track vertices are outside of the beampipe (inner radius=1.0 cm). Since a vertex is produced at most on the beam pipe, a distribution with $\sigma_{x,y} = 0.7$ represents a situation worse than possible.
The performance evaluation of the MLP was done as usual with the same procedure and representation scheme explained in section 5.1: the optimized MLPs were trained and tested 10 times for each parameter region without background, with background and with background and the normally distributed (x, y)-displacement setting $\sigma_{x,y} = 0.7$. Some examples of testing processes are presented in Fig. 5.17 and Fig. 5.18. The average results for the 10 experiments are presented in Fig. 5.19, where the errorbars correspond to the standard deviations for each measurement.
As it can be seen $\sigma(z_{\text{Out}} - z_{\text{Sim}})$ increases with background and much more with background and the light displacement in (x, y). The background leads to an average resolution loss of about 25%: in all cases the resolution with background is below 2 cm. The light (x, y)-displacement leads in turn to a resolution which is in average 2 times lower than that without background and, if compared with the case where there is only background, to an average resolution loss of about 60%. In the worst case a resolution of 2.38 cm is achieved. The lowest and the highest average resolution achieved are: without background

$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 1.40 \pm 0.07 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [0.50, 0.65]\,\text{GeV}/c \tag{5.13}$$

$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 0.66 \pm 0.04 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [3.20, 3.80]\,\text{GeV}/c; \tag{5.14}$$

with background

$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 1.70 \pm 0.02 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [0.50, 0.65]\,\text{GeV}/c \tag{5.15}$$

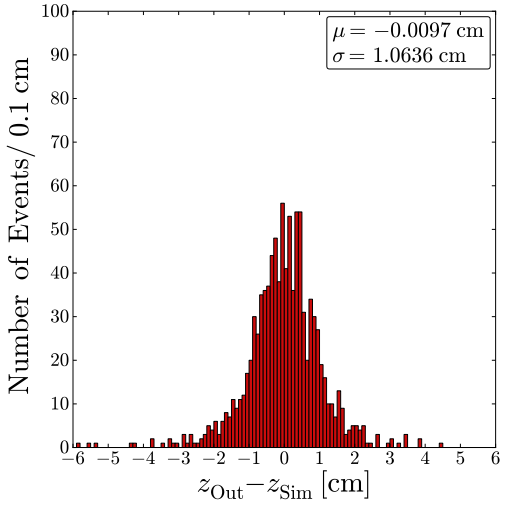
$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 0.85 \pm 0.04 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [3.20, 3.80]\,\text{GeV}/c; \tag{5.16}$$

and with background and a $\sigma_{x,y} = 0.7$ displacement

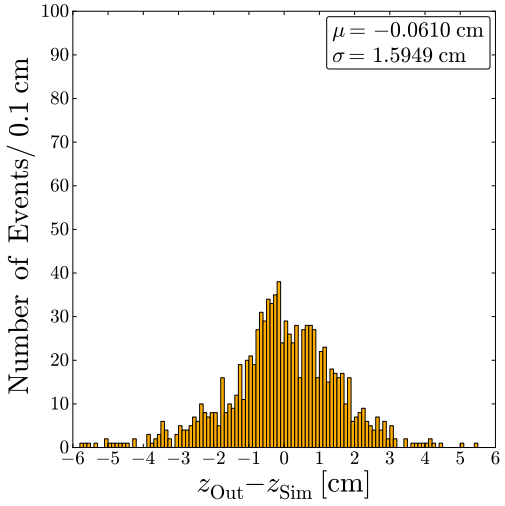
$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 2.38 \pm 0.06 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [0.50, 0.65]\,\text{GeV}/c \tag{5.17}$$

$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 1.75 \pm 0.04 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [3.20, 3.80]\,\text{GeV}/c. \tag{5.18}$$

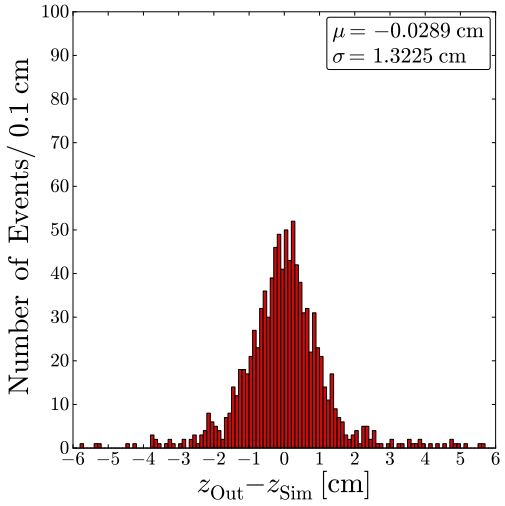
On the other hand the graphs for $\mu(z_{\text{Out}} - z_{\text{Sim}})$ show that the MLP input was not considerably biased.



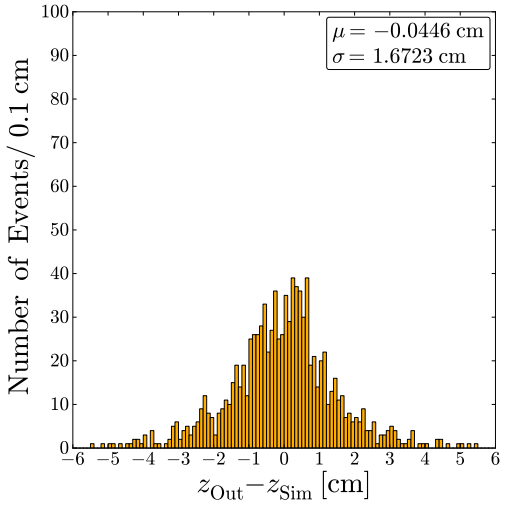
(a) MixBkg off $\sigma_{x,y} = 0.0cm$



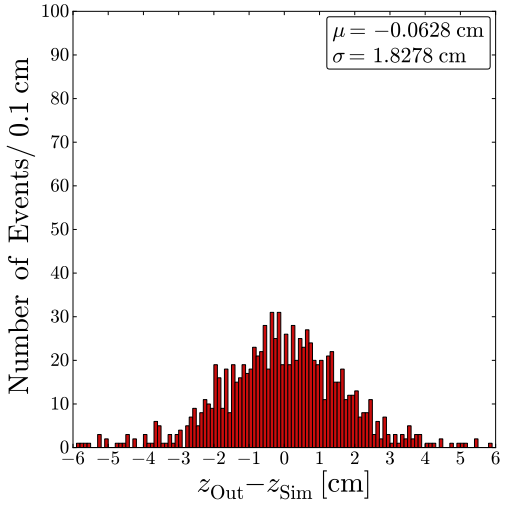
(b) MixBkg off $\sigma_{x,y} = 0.0cm$



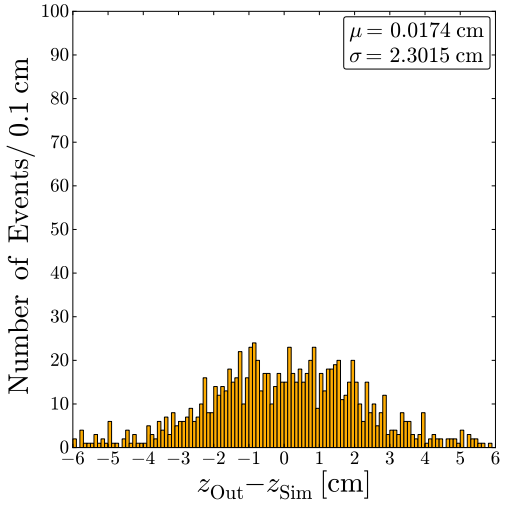
(c) MixBkg on $\sigma_{x,y} = 0.0cm$



(d) MixBkg on $\sigma_{x,y} = 0.0cm$

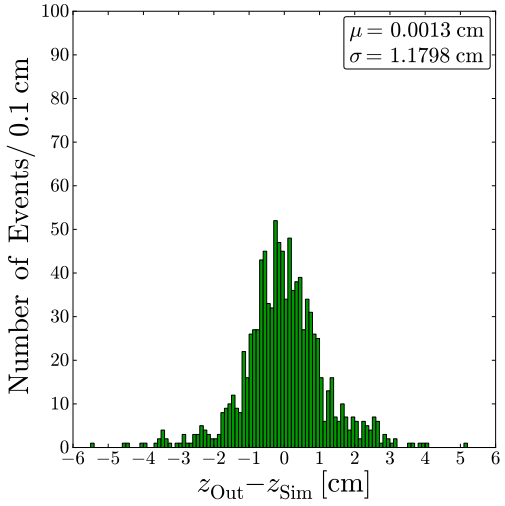


(e) MixBkg on $\sigma_{x,y} = 0.7cm$

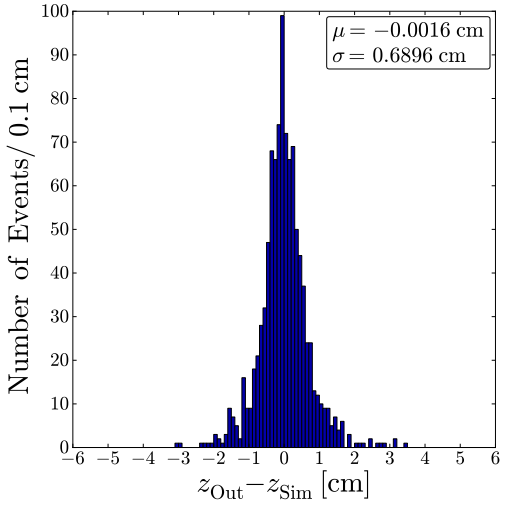


(f) MixBkg on $\sigma_{x,y} = 0.7cm$

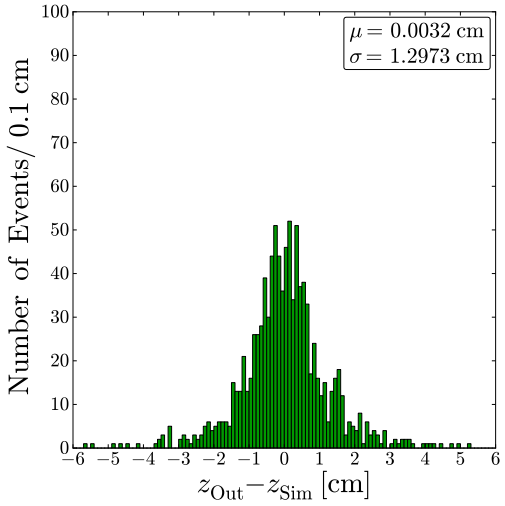
Figure 5.17: $z_{\text{Out}} - z_{\text{Sim}}$ distributions for $p_{\text{T}} \in [0.50, 0.55]$ GeV/c (left) and $p_{\text{T}} \in [0.50, 0.65]$ GeV/c (right).



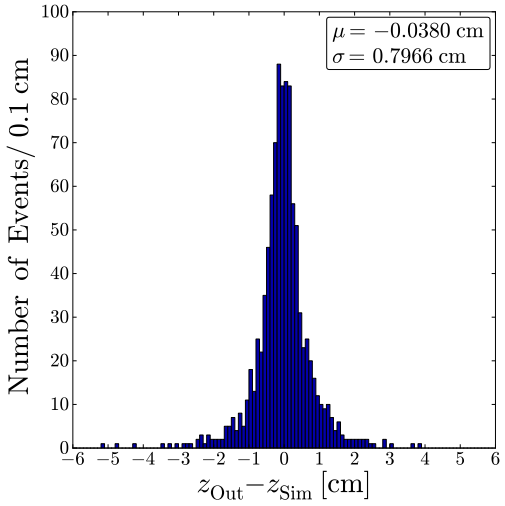
(a) MixBkg off $\sigma_{x,y} = 0.0cm$



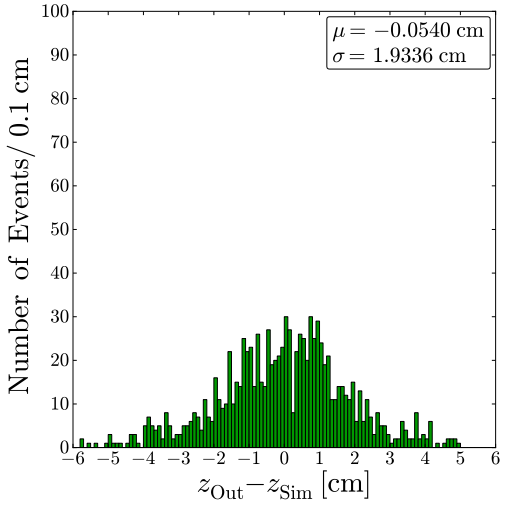
(b) MixBkg off $\sigma_{x,y} = 0.0cm$



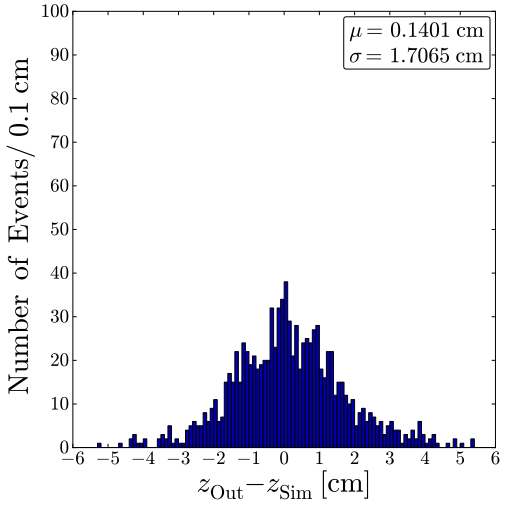
(c) MixBkg on $\sigma_{x,y} = 0.0cm$



(d) MixBkg on $\sigma_{x,y} = 0.0cm$



(e) MixBkg on $\sigma_{x,y} = 0.7cm$



(f) MixBkg on $\sigma_{x,y} = 0.7cm$

Figure 5.18: $z_{\text{Out}} - z_{\text{Sim}}$ distributions for $p_{\text{T}} \in [1.20, 1.80]$ GeV/c (left) and $p_{\text{T}} \in [3.20, 3.80]$ GeV/c (right).

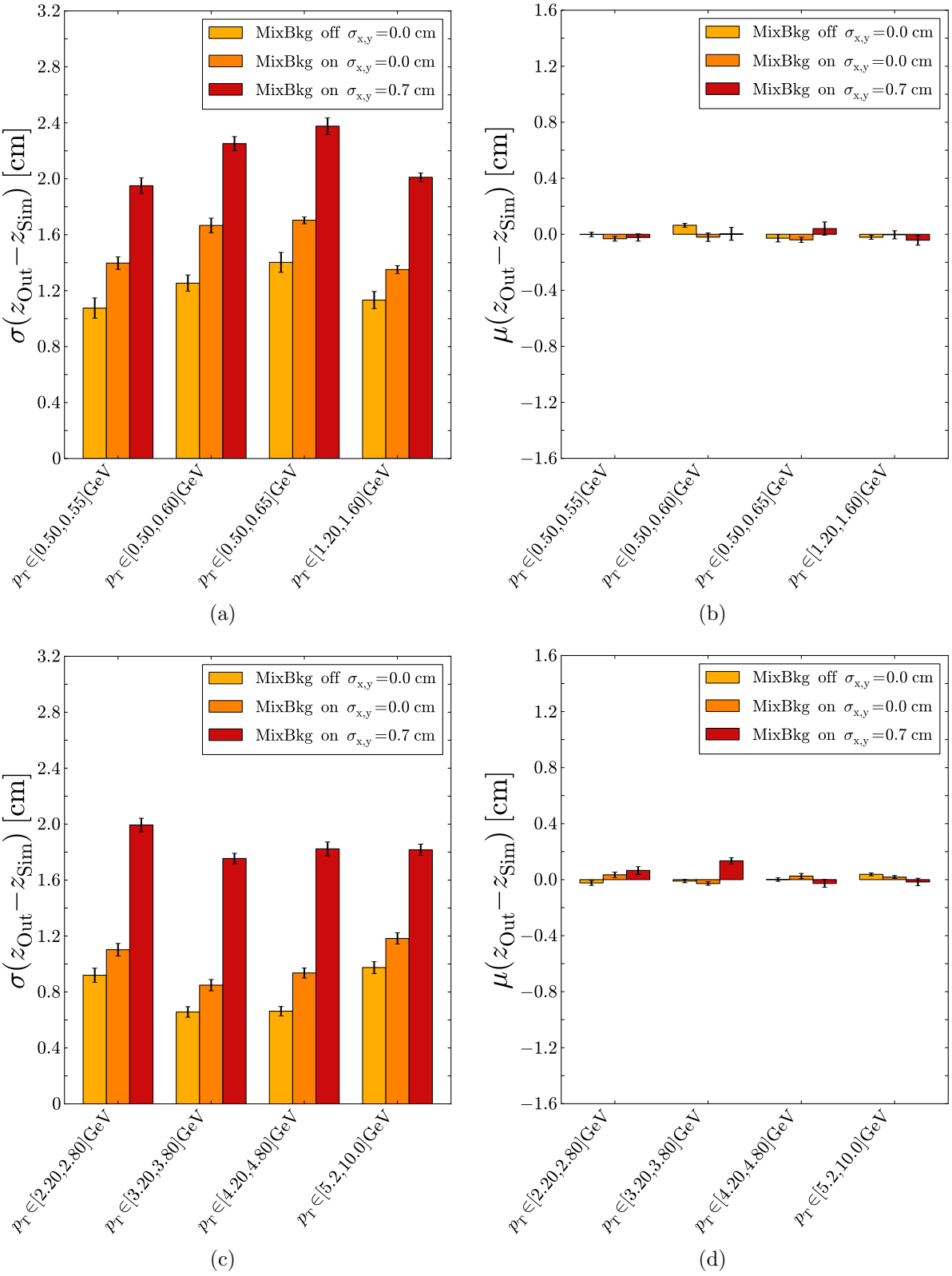


Figure 5.19: Standard Deviation (Resolution) and Mean for different p_T-regions. The MLPs have been trained and tested using events generated without and with Background as well as with Background and a normally distributed vertex on the (x, y) plane.

Chapter 6

Conclusion and Outlook

The main purpose of this study was to evaluate the performance of the MLP under the presence of background. Since the choice of the training parameters in the previous studies was done making assumptions, they had to be optimized in order to increase the significance of the results as much as possible.
Before the optimization process, it was necessary to estimate the track parameter regions for which each MLP is specialized. Based on the performance results, an approach of the region widths was made for all detector sectors. The total number of bins in θ was 57, in p_{T} was 55 and in ϕ was 720. This leads to a total number of bins

$$n_{\mathrm{Total}} = 57 \cdot 55 \cdot 720 = 2,257,200 \sim 2 \cdot 10^6$$

in order to cover the whole detector space. If positively and negatively charged particles are both considered, this number needs to be multiplied by a factor 2 because of the different track curvatures. Also the bin in z_0 has to be taken into account: in these studies only the range $z_0 \in [-10, 10]$ was covered. For more bins in z_0, n_{Total} has to be multiplied by the number of bins. The total z_0 range which has to be covered by the neural z-vertex trigger has also not been determined yet.
The number of $\sim 2 \cdot 10^6$ MLPs is quite large: future studies will require high computing power to test the possibility of training such big amounts of MLPs. This number is also relevant for the hardware implementation since the storing space in FPGAs is very limited. A single MLP needs about 5 kB storing space. Consequently, the whole Net-Hedgehog will require a total storing space of about $\sim 2 \cdot n_{\mathrm{Total}} \cdot 5$ kB ~ 20 GB.
A possibility to reduce the amount of MLPs would be to cover only the region $\theta \in [40, 140]$ because of the low resolution reached in the sectors lying close to the z-axis. The bin size in ϕ could be also increased by a factor 2 without a large resolution loss. In the p_{T} space, the largest amount of MLPs is needed for the low momentum sector. The region widths for this sector should be as narrow as possible because of the larger curvature differences and the large amount of particles expected therein, as it was revealed by the physics analysis. However, the final region size will be determined by the accuracy of the pre-processing step to select the specialized MLPs for a specific event. Further studies will need to investigate this issue in depth, since it is essential for the performance of the whole z-vertex trigger.

So far, only the look-up table approach and the 3D CDC trigger have been considered as predictor. It is not assured if the 3D CDC trigger will supply a prediction as fast as needed since the z-vertex trigger will have only a time window of 2µs. On the other hand the look-up table approach requires large amounts of memory.
In this work the focus was set only on the optimization and the the performance evaluation for the single MLPs. For the present experiments the MLPs were trained with more realistic input information than in previous studies: the random offset expected in the final version of the TSF was taken into account.
Furthermore, the input layer size of the MLPs was optimized using the cut off criterion to select only the smallest amount of relevant input channels, such that the resolution requirements on the MLPs were fulfilled. The size of the MLP hidden layer was set to $n_{\text{hidden}} = f_{\text{hidden}} \cdot n_{\text{input}}$. It was shown that the hidden factor has little impact on the resolution. For large hidden factors the training time was much longer than usual since there were more weights to be trained. The results show that a factor $f_{\text{hidden}} = 3$ conduces to very good results. The total number of weights for the used MLP can be calculated as follows

$$n_{\text{total}}^{\text{weights}} = f_{\text{hidden}} \cdot \left(n_{\text{input}}^2 + 2n_{\text{input}}\right) + 1.$$

After optimizing f_{hidden} and n_{input}, also the number of weights was optimized indirectly.
The culmination of this work was the evaluation of the MLP performance under the presence of background and under a light vertex displacement normally distributed over the (x, y)-plane with $\sigma_{x,y} = 0.7$. The Background led to an average resolution loss of about $\sim 25\%$. The lowest and the highest average resolution achieved are: without background

$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 1.40 \pm 0.07 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [0.50, 0.65]\,\text{GeV}/c$$
$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 0.66 \pm 0.04 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [3.20, 3.80]\,\text{GeV}/c;$$

with background

$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 1.70 \pm 0.02 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [0.50, 0.65]\,\text{GeV}/c$$
$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 0.85 \pm 0.04 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [3.20, 3.80]\,\text{GeV}/c;$$

and with background and a $\sigma_{x,y} = 0.7$ displacement

$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 2.38 \pm 0.06 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [0.50, 0.65]\,\text{GeV}/c$$
$$\sigma(z_{\text{Out}} - z_{\text{Sim}}) = 1.75 \pm 0.04 \text{ cm} \quad \text{for} \quad p_{\text{T}} \in [3.20, 3.80]\,\text{GeV}/c.$$

In the case of the light vertex displacement, the MLP exceeded the limit of 2 cm for some regions, but as it can be observed, the lowest resolution (2.38 cm) is not too far from the limit. The study of the effect of a vertex displacement on the (x, y)-plane was motivated by the fact that the beam focusing will not have been fully optimized at the early stages of SuperKEKB. However the value of $\sigma_{x,y} = 0.7$ cm represented a situation much worse than possible since in the real case there will be no vertices outside of the beampipe, which has a radius of 1 cm.
With this results it has been proved that, also under the presence of background, the MLPs are able to reach the required accuracy of less than 2 cm.

Appendix A

The Belle II Vertex Detector

The vertex detector in Belle II consists of two parts, the pixel vertex detector PXD and the silicon vertex detector SVD. Both were introduced in section 3.2. In the following, further interesting features of these detector components can be found.

layer		radius (mm)	slant angle	ladders	sensors /ladder	pitch $r\phi$ (µm)	pitch z (µm)	width (mm)	length (mm)
PXD	1	14	-	8	2	50	55, 60	12.5	44.80
	2	22	-	12	2	50	70, 85	12.5	61.44
SVD	3	38	-	7	2	50	160	38.52	120.02
	4	80	11.9°	10	3	75	240	57.72	122.90
	5	105	16.0°	12	4	75	240	57.72	122.90
	6	135	21.1°	16	5	75	240	57.72	122.90

Table A.1: Belle II vertex detector configuration. Widths and Lengths are only for the sensitive area. The slanted sensors have trapezoidal shapes with a variable $r\phi$ pitch from 50 µm to 75 µm, a sensitive width from 38.42 mm to 57.59 mm and a sensitive length of 122.76 mm[31].

A.0.1 Readout and Cooling System of the PXD

The DEPFET pixel sensors constituting the PXD are monolithic structures. Two DEPFET sensors are glued together with adhesive joint to build a ladder. Apart from small ceramic inserts in the joints to provide mechanical stability, the DEPFET sensors support themselves without use of additional mechanical structures within the acceptance region (Fig. A.1[31]). A more detailed description of the whole PXD with all its components can be found in [76].
The Depfets are read out row by row from one side with different kinds of current-digitizing chips: there are 8 chips at the end for the actual readout and 6 chips called switchers along

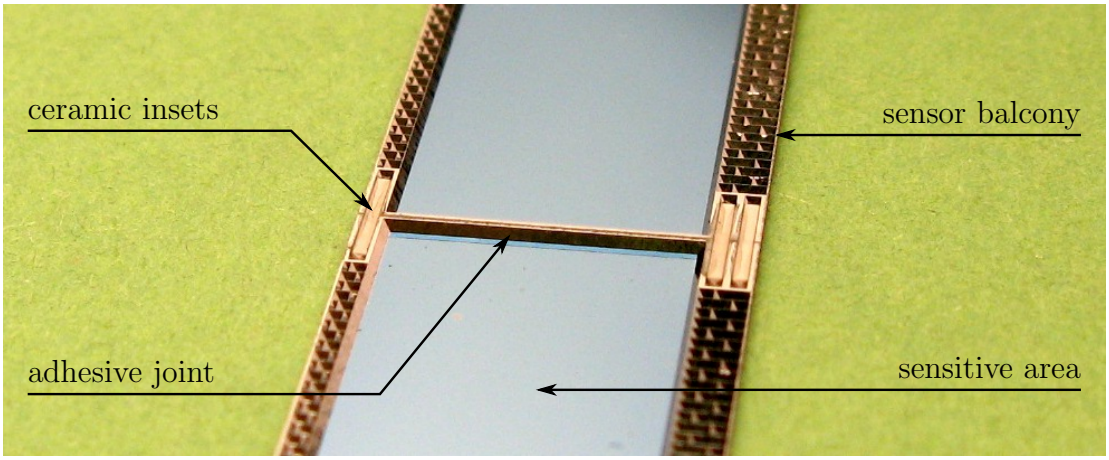


Figure A.1: Back view of the joint between two PXD sensors. The connection is done with an adhesive joint, reinforced using three small ceramic inserts. In addition, one can see the thinning of the active area compared to the sensor border [31].

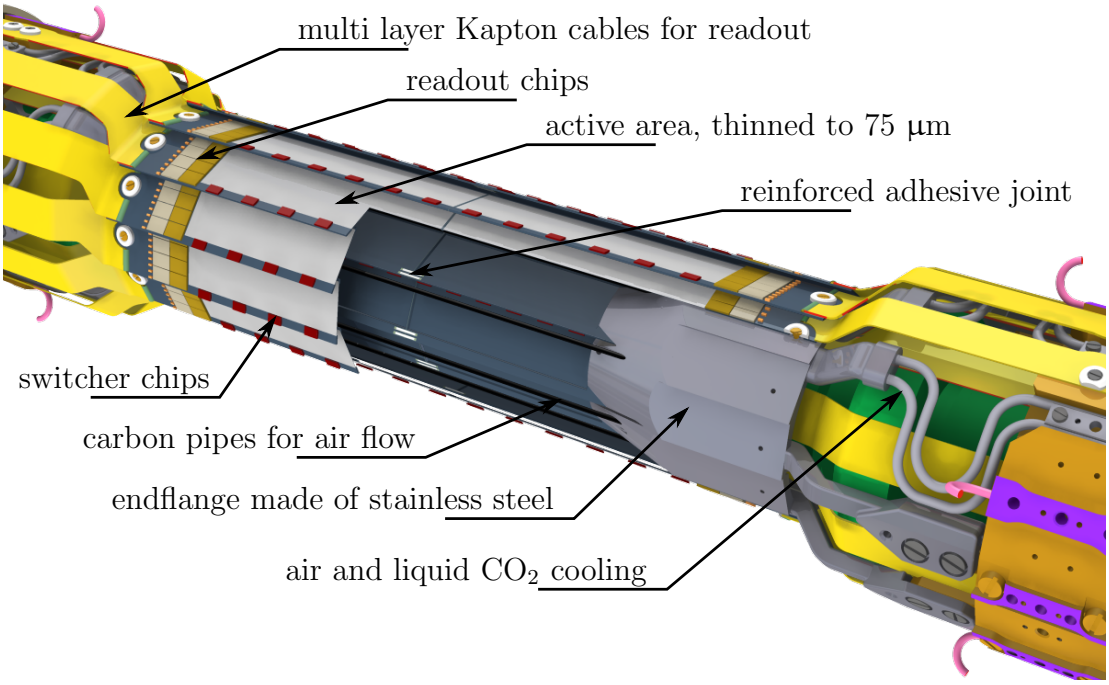


Figure A.2: Rendering of the mechanic design of the PXD. A part of the the sensors in the outer layer has been cut away to show the endlange and the carbon tubes for air cooling [31].

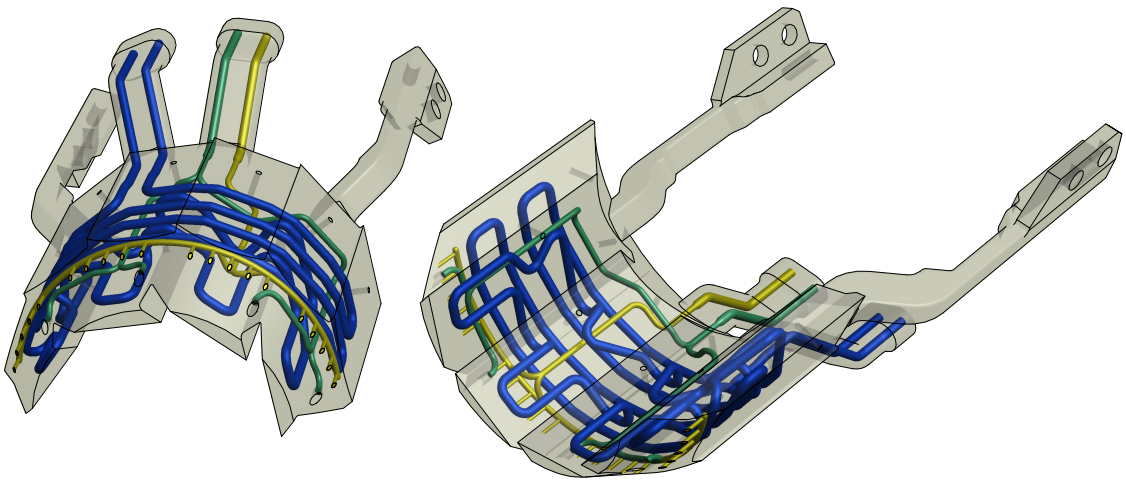


Figure A.3: Mechanical design of the PXD endflange. The internal blue lines are the for liquid CO_2 cooling and green and yellow lines are for airflow cooling[31].

the sensor side, in order to control the read out of the different rows.
The sensor ladders are mounted with screws, in forward and backward direction, on two endflanges outside the acceptance region. These endflanges do not only provide the mechanical support to the sensor ladders, they contain a cooling structure with flowing CO_2 for the readout chips. The cooling of the switcher chips, which are placed on the acceptance region, is done with small carbon tubes blowing cold air (-5° C at speed of $\sim$ 1 m/s) through small holes along their sides(S. Fig. A.2[31]). The described cooling system is sufficient since the total amount of thermal power of the PXD is very little ($\sim$ 360 W).

A.0.2 Readout and Cooling System of the SVD

For the read out of the central sensors in each ladder, the so-called "Origami-chip-on-sensor" concept is employed (S. FigA.4[6, 31]). The readout of the outer sensors is done by readout boards outside of the acceptance region. The readout board and the detector ladders are mounted on stainless steel endrings, which provide cooling for the readout boards at the same time. It consists on readout chips within a flexible circuit, which is mounted on the top of the sensors. The cooling is done with a stainless steel cooling pipe (1.5 mm diameter and 50 µm wall thickness) with flowing liquid CO_2 inside.

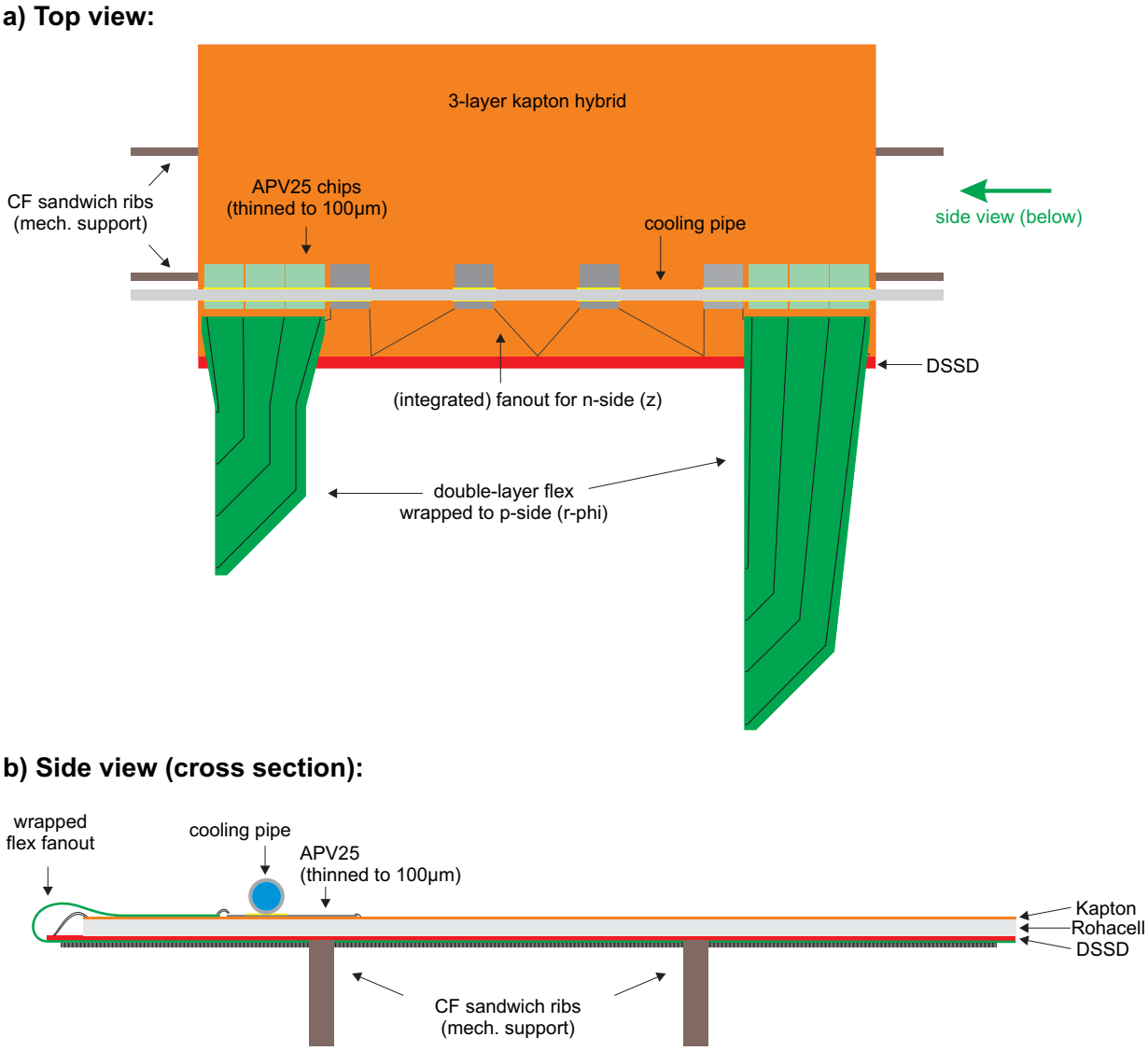


Figure A.4: Top and side views of the Origami chip-on-sensor concept. The flex pieces to be wrapped around the sensor edge are unfolded in the top view[6, 31].

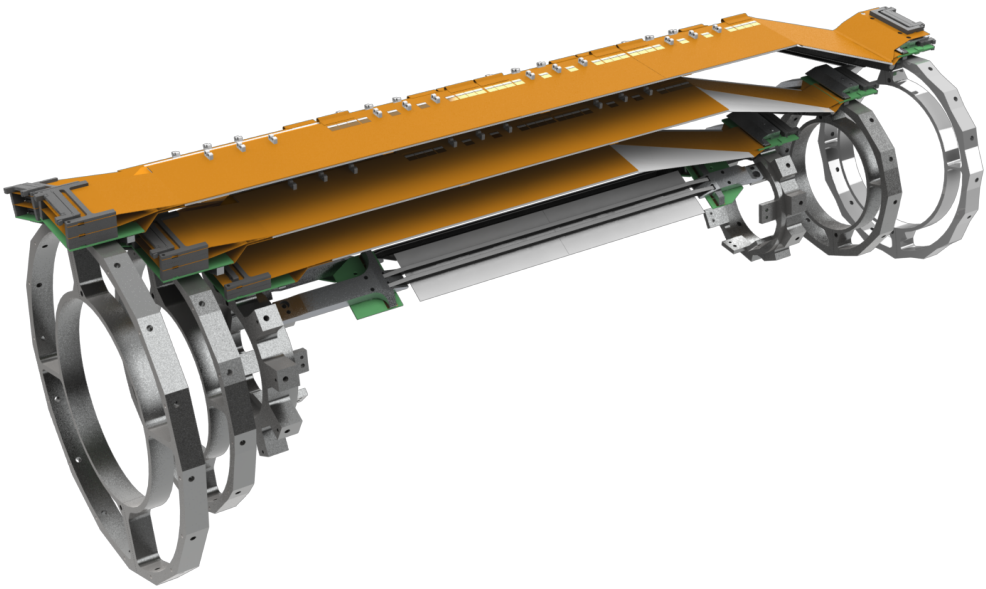


Figure A.5: Rendering of endring stucture for the SVD. The readout boards (green) for the outer modules are places directly on these endrings[31].

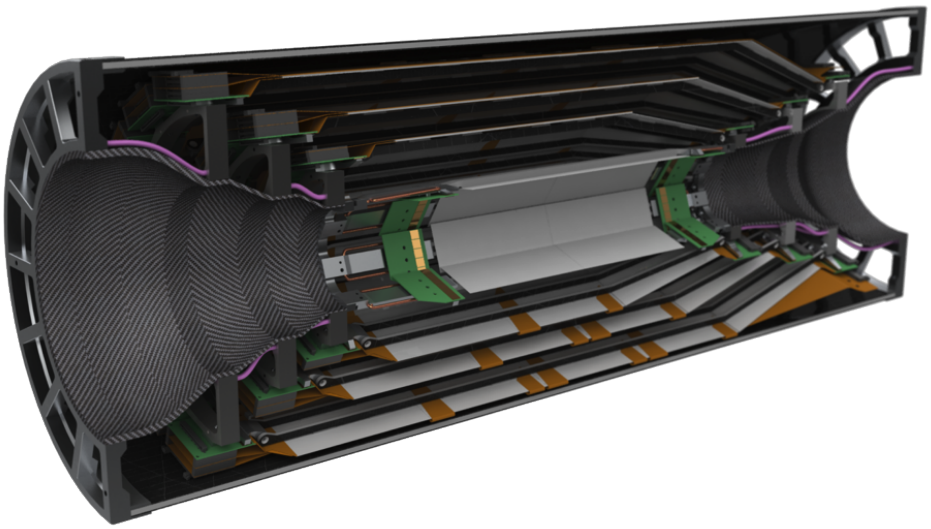


Figure A.6: Rendering of the mechanical design of half the SVD including all components[31].

List of Figures

List of Tables

Bibliography

[1] S. Skambraks. "Use of Neural Networks for Triggering in PARTICLE Physics". MA thesis. Ludwig-Maximilians-Universität, 2013.

[2] B. Aubert et al. "Measurement of the CP Asymmetry Amplitude $\sin 2\beta$ with B^0 Mesons". *Physical review letters* 89.20 (2002), p. 201802.

[3] M. Kobayashi. "Nobel Lecture: *CP* violation and flavor mixing". *Rev. Mod. Phys.* 81 (3 July 2009), pp. 1019–1025. DOI: 10.1103/RevModPhys.81.1019. URL: http://link.aps.org/doi/10.1103/RevModPhys.81.1019.

[4] T. Maskawa. "Nobel Lecture: What does *CP* violation tell us?" *Reviews of Modern Physics* 81.3 (2009), p. 1027.

[5] A. D. Sakharov. "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe". *Soviet Journal of Experimental and Theoretical Physics Letters* 5 (1967), p. 24.

[6] T. Abe et al. "Belle II Technical Design Report". *ArXiv e-prints* (Nov. 2010). arXiv: 1011.0352 [physics.ins-det].

[7] M. Beyer. "CP violation in particle nuclear, and astrophysics". *CP Violation in PARTICLE, Nuclear, and Astrophysics*. Vol. 591. 2002.

[8] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: http://link.aps.org/doi/10.1103/PhysRev.105.1413.

[9] T. D. Lee, R. Oehme, and C. N. Yang. "Remarks on Possible Noninvariance under Time Reversal and Charge Conjugation". *Phys. Rev.* 106 (2 Apr. 1957), pp. 340–345. DOI: 10.1103/PhysRev.106.340. URL: http://link.aps.org/doi/10.1103/PhysRev.106.340.

[10] A. Bettini. *Introduction to elementary particle physics*. Cambridge University Press, 2008.

[11] M. Goldhaber, L. Grodzins, and A. Sunyar. "Helicity of neutrinos". *Physical Review (US) Superseded in part by Phys. Rev. A, Phys. Rev. B: Solid State, Phys. Rev. C, and Phys. Rev. D* 109 (1958).

[12] J. H. Christenson et al. "Regeneration of K_1^0 Mesons and the $K_1^0 - K_2^0$ Mass Difference". *Phys. Rev.* 140 (1B Oct. 1965), B74–B84. DOI: 10.1103/PhysRev.140.B74. URL: http://link.aps.org/doi/10.1103/PhysRev.140.B74.

[13] N. Cabibbo. "Unitary Symmetry and Leptonic Decays". *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: http://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[14] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285. URL: http://link.aps.org/doi/10.1103/PhysRevD.2.1285.

[15] M. Kobayashi and T. Maskawa. "CP-violation in the renormalizable theory of weak interaction". *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657.

[16] L.-L. Chau and W.-Y. Keung. "Comments on the Parametrization of the Kobayashi-Maskawa Matrix". *Phys. Rev. Lett.* 53 (19 Nov. 1984), pp. 1802–1805. DOI: 10.1103/PhysRevLett.53.1802. URL: http://link.aps.org/doi/10.1103/PhysRevLett.53.1802.

[17] H. Harari and M. Leurer. "Recommending a standard choice of Cabibbo angles and KM phases for any number of generations". *Physics Letters B* 181.1 (1986), pp. 123–128.

[18] L. Wolfenstein. "Does the Kobayashi-Maskawa matrix violate *CP* maximally?" *Physics Letters B* 144.5 (1984), pp. 425–426.

[19] P. Vanhoefer. "Measurement of the Branching Fraction of the Decay $B^0 \to \rho^0\rho^0$". MA thesis. Max-Planck Institut für Physik München, 2010.

[20] Z.-z. Xing. "Effect of $K^0 - \bar{K}^0$ mixing on *CP* asymmetries in weak decays of D and B mesons". *Physics Letters B* 353.2 (1995), pp. 313–318.

[21] M. Sozzi. *Discrete Symmetries and CP Violation: From Experiment to Theory: From Experiment to Theory*. Oxford university press, 2008.

[22] K. Nakamura et al. "Review of particle physics". *Journal of Physics G: Nuclear and PARTICLE Physics* 37.7A (2010), p. 075021.

[23] J. Charles et al. "CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories". *Eur.Phys.J.* C41 (2005). Updated results and plots available at: http://ckmfitter.in2p3.fr, pp. 1–131. DOI: 10.1140/epjc/s2005-02169-1. arXiv: hep-ph/0406184 [hep-ph].

[24] E. Wigner and V. Weisskopf. *Z.Phys.* 63.2 (1930), p. 54.

[25] E. Wigner and V. Weisskopf. *Z.Phys.* 65.2 (1930), p. 18.

[26] W. D. Hulsbergen. *A study of track reconstruction and massive dielectron production in Hera-B*. 2002.

[27] W. Hulsbergen. "CONSTRAINING NEW PHYSICS IN MESON MIXING". *Modern Physics Letters A* 28.27 (2013).

[28] R. Lewis and C. Enz. "Phys. Letters 16, 73 (1965); RR Lewis". *Phys. Rev. Letters* 14 (1965), p. 749.

[29] K. Abe et al. "Observation of large *CP* violation in the neutral B meson system". *Physical Review Letters* 87.9 (2001), p. 091802.

[30] B. Aubert et al. "Observation of *CP* Violation in the B^0 Meson System". *Physical Review Letters* 87.9 (2001), p. 091801.

[31] M. Ritter. "Measurement of the Branching Fraction and Time Dependent *CP* Asymmetry in $B^0 \to D^{*-}D^{*+}K^0_S$ Decays at the Belle Experiment". To be pubished. PhD thesis. Max-Planck Institut für Physik München, 2013.

[32] Wikipedia. *Quark — Wikipedia, The Free Encyclopedia*. [Online; accessed 2013-08-06] http://en.wikipedia.org/w/index.php?title=Quark&oldid=315864017. 2013.

[33] M. Bander, D. Silverman, and A. Soni. "*CP* Noninvariance in the Decays of Heavy Charged Quark Systems". *Phys. Rev. Lett.* 43 (4 July 1979), pp. 242–245. DOI: 10.1103/PhysRevLett.43.242. URL: http://link.aps.org/doi/10.1103/PhysRevLett.43.242.

[34] E. Nedelkovska. "Measurement of the Branching Fraction of the Decay $B^0 \to \psi(2s)\pi^0$ and Studies of the Luminosity-dependent Background for the Belle II experiment at the future accelerators SuperKEKB". PhD thesis. Max-Planck Institut für Physik München, 2013.

[35] T. E. Browder and K. Honscheid. "B mesons". *Progress in PARTICLE and Nuclear Physics* 35 (1995), pp. 81–219. ISSN: 0146-6410. DOI: 10.1016/0146-6410(95)00042-H.

[36] J. Brodzicka et al. "Physics achievements from the Belle experiment". *Progress of Theoretical and Experimental Physics* 2012.1 (2012).

[37] K. Abe et al. "Observation of mixing-induced CP violation in the neutral B meson system". *Phys. Rev. D* 66 (3 Aug. 2002), p. 032007. DOI: 10.1103/PhysRevD.66.032007. URL: http://link.aps.org/doi/10.1103/PhysRevD.66.032007.

[38] K.-F. Chen et al. "Observation of Time-Dependent *CP* Violation in $B^0 \to \eta' K^0$ Decays and Improved Measurements of *CP* Asymmetries in $B^0 \to \phi K^0$, $K^0_S K^0_S K^0_S$ and $B_0 \to J/\psi K_0$ Decays". *Phys. Rev. Lett.* 98 (3 Jan. 2007), p. 031802. DOI: 10.1103/PhysRevLett.98.031802. URL: http://link.aps.org/doi/10.1103/PhysRevLett.98.031802.

[39] I. Adachi et al. "Precise Measurement of the *CP* Violation Parameter $\sin 2\phi_1$ in $B^0 \to (c\bar{c})K^0$ Decays". *Phys. Rev. Lett.* 108 (17 Apr. 2012), p. 171802. DOI: 10.1103/PhysRevLett.108.171802. URL: http://link.aps.org/doi/10.1103/PhysRevLett.108.171802.

[40] K. Abe et al. "Measurement of Time-Dependent CP-Violating Asymmetries in $B^0 \to \phi K^0_S$,$K^+K^-K^0_S$, and $\eta' K^0_S$ Decays". *Phys. Rev. Lett.* 91 (26 Dec. 2003), p. 261602. DOI: 10.1103/PhysRevLett.91.261602. URL: http://link.aps.org/doi/10.1103/PhysRevLett.91.261602.

[41] Y. Amhis et al. "Averages of b-hadron, c-hadron, and tau-lepton properties as of early 2012". *arXiv preprint arXiv:1207.1158* (2012).

[42] R. H. Dalitz. "Decay of τ Mesons of Known Charge". *Phys. Rev.* 94 (4 May 1954), pp. 1046–1051. DOI: 10.1103/PhysRev.94.1046. URL: http://link.aps.org/doi/10.1103/PhysRev.94.1046.

[43] J. Dalseno et al. "Time-dependent Dalitz plot measurement of CP parameters in $B_0 \to K^0_S\pi^+\pi^-$ decays". *Phys. Rev. D* 79 (7 Apr. 2009), p. 072004. DOI: 10.1103/PhysRevD.79.072004. URL: http://link.aps.org/doi/10.1103/PhysRevD.79.072004.

[44] Y. Nakahama et al. "Measurement of CP violating asymmetries in $B^0 \to K^+K^-K^0_S$ decays with a time-dependent Dalitz approach". *Phys. Rev. D* 82 (7 Oct. 2010), p. 073011. DOI: 10.1103/PhysRevD.82.073011. URL: http://link.aps.org/doi/10.1103/PhysRevD.82.073011.

[45] T. Aushev et al. "Physics at super B factory". *arXiv preprint arXiv:1002.5012* (2010).

[46] H.-Y. Cheng, C.-K. Chua, and A. Soni. "Effects of final-state interactions on mixing-induced CP violation in penguin-dominated B decays". *Phys. Rev. D* 72 (1 July 2005), p. 014006. DOI: 10.1103/PhysRevD.72.014006. URL: http://link.aps.org/doi/10.1103/PhysRevD.72.014006.

[47] W. Herr and B. Muratori. "Concept of luminosity" (2006).

[48] P. Raimondi. *Status on SuperB effort*. talk given at the 2nd SuperB workshop, Frascati, 2006. URL: http://www.lnf.infn.it/conference/superb06/talks/raimondi1.ppt.

[49] SuperB Collaboration et al. "SuperB Technical Design Report". *ArXiv e-prints* (June 2013). arXiv: 1306.5655 [physics.ins-det].

[50] J. Kemmer and G. Lutz. "New detector concepts". *Nucl. Instrum. Methods Phys. Res. Sect. A* 253.3 (1987), pp. 365–377. ISSN: 0168-9002. DOI: 10.1016/0168-9002(87)90518-3.

[51] R. Richter et al. "Design and technology of DEPFET pixel sensors for linear collider applications". *Nucl. Instrum. Methods Phys. Res. Sect. A* 511.1–2 (2003), pp. 250–256. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01802-3.

[52] C. Bernardini et al. "Lifetime and beam size in a storage ring". *Physical Review Letters* 10.9 (1963), pp. 407–409.

[53] H. Nakano et al. "BEAM BACKGROUND SIMULATION FOR SUPERKEKB/BELLE-II". *Energy [GeV]* 3 (), pp. 8–.

[54] J. Haba. *Beam Background in the Belle Detector*. BELLE Internal Note 277, accessible by Belle members only. 1999.

[55] A. Bozek et al. *Simulation Study of Beam-backgrounds on Silicon Vertex Detector at KEK B-Factory*. BELLE Internal Note 174, accessible by Belle members only. 1997.

[56] S. Sahu. *A Generator for Studies of Background due to Beam-Gas Interactive at KEK B-Factory*. BELLE Internal Note 172, accessible by Belle members only. 1997.

[57] P. Schmüser, M. Dohlus, and J. Rossbach. *Ultraviolet and soft X-ray free-electron lasers: introduction to physical principles, experimental results, technological challenges*. Vol. 229. Springer, 2008.

[58] T. Browder et al. *Synchrotron Radiation Background Simulation Study for Belle*. BELLE Internal Note 361, accessible by Belle members only. 2007.

[59] M. Harakeh and A. Van der Woude. *Giant Resonances: Fundamental High-Frequency Modes of Nuclear Excitations*. 2001.

[60] F. A. Berends, P. H. Daverfeldt, and R. Kleiss. "Monte Carlo Simulation of Two-Photon Processes II: Complete Lowest Order Calculations for Four-Lepton Production Processes in Electron-Positron Collisions". *Computer Physics Communications* 40 (1986), pp. 285–307. DOI: 10.1016/0010-4655(86)90115-3.

[61] S. Jadach et al. "The Monte Carlo Program KoralW version 1.51 and The Concurrent Monte Carlo KoralW and YFSWW3 With All Background Graphs and First Order Corrections to W Pair Production". *Computer Physics Communications* 140 (2001), pp. 475–512. DOI: 10.1016/S0010-4655(01)00296-X. arXiv: 0104049 [hep-ph].

[62] Y. Iwasaki et al. "Status of TSFinder". *16th B2GM*. Nov. 2013.

[63] Y. Iwasaki et al. "Level 1 trigger system for the Belle II experiment". *Real Time Conference (RT), 2010 17th IEEE-NPSS*. IEEE. 2010, pp. 1–9.

[64] P. V. Hough. "Machine analysis of bubble chamber pictures". *International Conference on High Energy Accelerators and Instrumentation*. Vol. 73. 1959.

[65] M. Ohlsson. "Extensions and explorations of the elastic arms algorithm". *Computer physics communications* 77.1 (1993), pp. 19–32.

[66] Y. Ohnishi et al. "Track parametrization". *Belle note* 148 (1996), p. 1996.

[67] A. Moll. "The Software Framework of the Belle II Experiment". *Journal of Physics: Conference Series* 331.3 (2011), p. 032024. DOI: 10.1088/1742-6596/331/3/032024.

[68] ISO. *ISO/IEC 14882:2011 Information technology — Programming languages — C++*. Geneva, Switzerland: International Organization for Standardization, Feb. 28, 2012, 1338 (est.) URL: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=50372.

[69] S. Agostinelli et al. "GEANT4: A Simulation toolkit". *Nucl.Instrum.Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[70] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, 2012.

[71] A. N. Kolmogorov. "On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition". *Dokl. Akad. Nauk SSSR*. Vol. 114. 5. 1957, pp. 953–956.

[72] S. Lenissen. "Implementation of a fast artificial neural network (FANN)". MA thesis. University of Copenhagen, 2003.

[73] C. Igel and M. Hüsken. "Improving the Rprop learning algorithm". *Proceedings of the second international ICSC symposium on neural computation (NC 2000)*. Citeseer. 2000, pp. 115–121.

[74] P. SIBI, S. A. JONES, and P. SIDDARTH. "ANALYSIS OF DIFFERENT ACTIVATION FUNCTIONS USING BACK PROPAGATION NEURAL NETWORKS". *Journal of Theoretical and Applied Information Technology* 47.3 (2013).

[75] M. Riedmiller and H. Braun. "A direct adaptive method for faster backpropagation learning: The RPROP algorithm". *Neural Networks, 1993., IEEE International Conference on*. IEEE. 1993, pp. 586–591.

[76] The DEPFET Collaboration. "The PXD Whitebook". unpublished. 2012. URL: `http://twiki.hll.mpg.de/twiki/bin/viewfile/DepfetInternal/WhiteBookStructure?rev=1;filename=PXDwb20130712.pdf`.

Acknowledgments

In the first place, I would like to express my sincere gratitude to Prof. Dr. Kiesling and Prof. Dr. Schieck, my research supervisors, for their enthusiastic encouragement concerning the neural z-vertex trigger project, their friendly guidance and their interest in my work. I appreciate their willingness to give their time and to help me to solve any kind of questions. I am also very grateful to Prof. Dr. Schieck for his assistance and enthusiasm during his lecture on heavy quarks physics and for having encouraged me to join this project.
I would also like to offer my special thanks to my college Sebastian Skambraks for having helped me to become familiarized with the multi layer perceptron, for his continuous constructive suggestions and his useful critiques.
For the friendly atmosphere in the Belle group, I thank Veronika Chobanova, Elena Nedelkovska, Pit Vanhoefer, Vladimir Chekelian and especially Martin Ritter and Luigi Li Gioi: to Martin for offering his knowledge and his experience to help me solve any kind of problems; to Luigi Li Gioi for his help with the physics analysis and for his friendly disposition to help me answering all kind of questions regarding to physics and to analysis.
I would also like to thank Sara Neuhaus, who recently joined this project, for her recommendations during the writing process. Further I acknowledge the Max-Planck Institute and the help of many of his members.
My deepest thank to all my friends for their support and all the fun I have had during my studies in Munich, specially to Thurid Piehler-Braun, Peter Müller and Beate Cochlovius.
Thanks to my mother and to my family for their support and encouragement. Last but not least, thank you Aleksa for strengthening me with your love.

Einständigkeitserklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig verfasst habe, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, den 28. Januar 2014

Fernando Abudinén.