

Bachelor's Thesis

Predicting injection backgrounds at SuperKEKB using neural networks

Vorhersage vom Injektionsuntergrund am Beschleuniger SuperKEKB mit neuronalen Netzen

prepared by

Lukas Herzberg

at the II. Physikalischen Institut

Thesis number: II.Physik-UniGö-BSc-2022/01

Thesis period: 6th January 2022 until 31st March 2022

First referee: Dr. Benjamin Schwenker

Second referee: Prof. Dr. Stan Lai

Contents

1.	Introduction	1
2.	The Standard Model	2
2.1.	Fermions	2
2.2.	Bosons	3
2.3.	CP-Violation	3
3.	The SuperKEKB collider	5
3.1.	Beam dynamics	6
3.2.	Beta function and emittance	8
3.3.	Top up Injections	9
3.4.	SuperKEKB movable collimator system	10
4.	The Belle II detector	12
4.1.	Pixel Vertex Detector	13
4.2.	Silicon Vertex Detector	14
4.3.	Central Drift Chamber	15
4.4.	Time of Propagation Counter	15
4.5.	Aerogel Ring-Imaging Cherenkov detector	16
4.6.	Electromagnetic Calorimeter	17
4.7.	The Solenoid	17
4.8.	The K_L -Muon Detector	18
4.9.	Global Decision Logic	18
5.	Backgrounds	19
6.	Artificial Neural Networks	21
6.1.	Network architecture	21
6.2.	Network training	22
7.	The Injection Duration Network	23
7.1.	Injection Duration	23
7.2.	Preprocessing	24
7.3.	Network Training	25
7.4.	Prediction and generalization	26

Contents

7.5.	Path Explain feature attribution	27
7.6.	Collimator scans	30
7.7.	Correlation coefficients	31
7.8.	Limitations	32
8.	Conclusion	33
I.	Lower ranking features	37

1. Introduction

Every process that has ever been observed in particle physics respects the conservation of baryon and lepton numbers. That means every time a particle is created or destroyed, a corresponding antiparticle is also created or destroyed at the same time. Following this principle, the difference between the number of particles and the number of antiparticles will always be the same, yet if we look at the observable universe it is clear that this difference is not zero. Matter is abundant while antimatter is absent in comparison. It is hypothesised that this imbalance came to be in the early stages of the universe in a process called baryogenesis, but the exact mechanism can not be explained by modern physics. One possible approach to find an explanation is to study the violation of the charge conjugation parity (**CP**) symmetry. CP-symmetry states that the laws of physics should stay the same upon exchanging a particle with its antiparticle and mirroring its spatial coordinates. In 1964, James Cronin and Val Fitch showed that CP-symmetry is broken in the neutral kaon system [6] for which they earned the Nobel Prize in 1980. This discovery shows that physics is different for matter and antimatter and therefore is a natural starting point to investigate the matter-antimatter discrepancy. For this purpose the KEKB particle collider and Belle detector have been constructed at the High Energy Accelerator Research Organisation known as KEK in Tsukuba, Japan. The experiment ran from 1999 to 2010, where it found evidence for CP violation in B mesons which confirmed the theory established by Cabibbo, Kobayashi and Maskawa of CP violation caused by flavor changing charged current in the quarks sector and led to the Nobel Prize for Kobayashi and Maskawa in 2008. Due to the success of the experiment an upgrade of the accelerator and detector has been commissioned, which has been realized in the form of SuperKEKB and Belle II. The goal of this new experiment is to do precision measurements of rare processes that were limited by the amount of data taken by the Belle experiment, which achieved a total integrated luminosity of 710 fb^{-1} . To do this, Belle II is set to collect around 50 times more data than its predecessor, mostly due to a 40 times increase in luminosity, a quantity that describes the number of collisions in a certain time period, of the KEKB. Increasing the luminosity by such a large margin is not trivial by any means and raises many problems such an increase in undesired effects seen by the detector. These undesired effects are universally referred to as backgrounds. In this thesis a machine learning approach will be introduced to better understand the source of one of these types of backgrounds, so that its impact might be mitigated in the future of Belle II and SuperKEKB.

2. The Standard Model

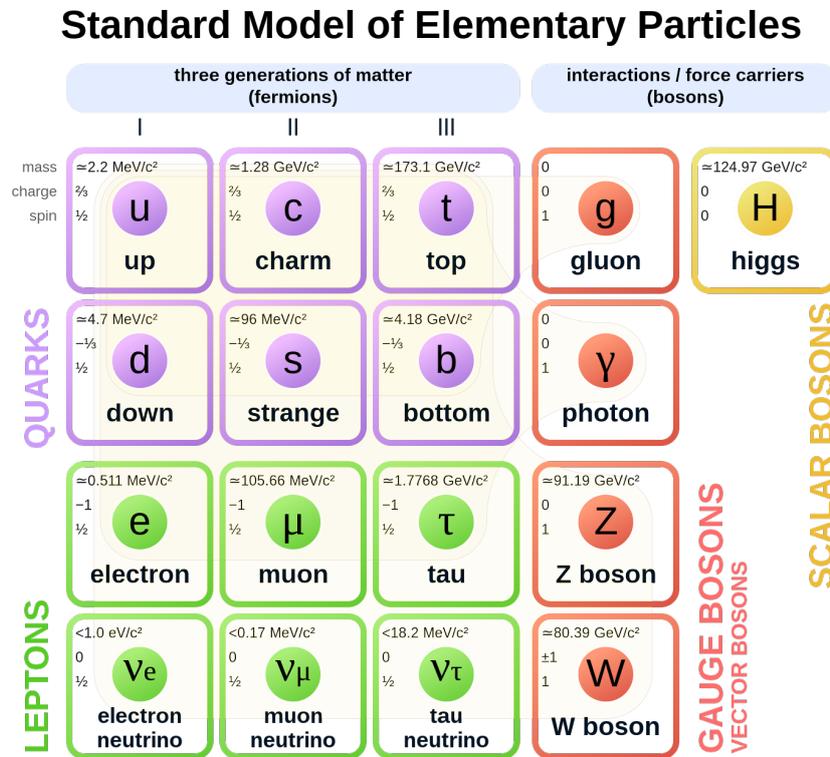


Figure 1.: The Standard Model of particle physics [9].

The Standard Model of particle physics, which can be seen in figure 1, describes the fundamental particles that constitute matter and the interaction forces amongst them, namely the electromagnetic, strong and weak interaction. It consists of fermions, which can be subdivided into quarks and leptons, and bosons, which consist of the vector bosons, as carrier particles of the fundamental forces, and the Higgs boson.

2.1. Fermions

Fermions share the overall property of being spin $1/2$ particles. There are a total of 12 fermions, which can be divided into six quarks and six leptons. Every fermion also has a corresponding antifermion, with opposite charges.

Quarks can interact through all three fundamental forces of the standard model. They carry electric charge and can therefore interact with photons, the carrier particle of the electromagnetic force. Quarks are divided into up- and down-type quarks, with up-type having a charge of $+2/3e$ and down-type quarks having a charge of $-1/3e$. Other than

leptons, quarks also carry colour charge and can interact strongly with gluons. The colour charges are referred to as *red*, *blue* and *green* and antiquarks carry corresponding anticolour charge. In nature, only "colourless" object can exist, meaning quarks have to form baryons consisting of three quarks each carrying a different colour charge thus forming *white*, or mesons consisting of a quark and an antiquark with a colour and an anticolour that cancel each other out.

Quarks also possess weak isospin and can therefore couple to W and Z bosons.

Leptons can be divided into charged and neutral leptons. Charged leptons carry the electromagnetic charge $-1e$ and can therefore interact electromagnetically. Other than quarks, leptons do not possess a colour charge and can consequently not interact strongly. Furthermore, neutral leptons, also called neutrinos, do not possess an electric charge and can thus only couple to W and Z bosons.

2.2. Bosons

Bosons are integer spin particles. The vector bosons have spin 1 and are responsible for the interactions between particles:

- The photon γ is the massless carrier of the electromagnetic force.
- The gluon is also massless and the mediator of the strong force. Gluons also carry colour charge.
- The W^\pm and the Z^0 are responsible for the weak interaction. Both are massive particles but the W^\pm is charged with $\pm 1e$, while the Z^0 is neutral.
- Besides the vector bosons, there also is a single scalar boson, the Higgs boson. Other than the vector bosons it has spin 0 and does not mediate a force. The Higgs boson is an excitation of the Higgs field, which is responsible for particle masses.

2.3. CP-Violation

CP-Violation describes the process of violating CP-invariance, which states, that laws of physics should be the same if all particles within a system are interchanged with their antiparticles and its spatial coordinates are inverted. The existence of CP-violation means

that antimatter behaves differently from matter. There are three types of CP-violation: **Indirect CP-violation** occurs for neutral mesons, that can transform into their antiparticles with Feynman diagrams like the one shown in figure 2. CP-violation occurs if the probability P of a neutral meson M^0 transforming into the corresponding antimeson \bar{M}^0 are different: $P(M^0 \rightarrow \bar{M}^0) \neq P(\bar{M}^0 \rightarrow M^0)$

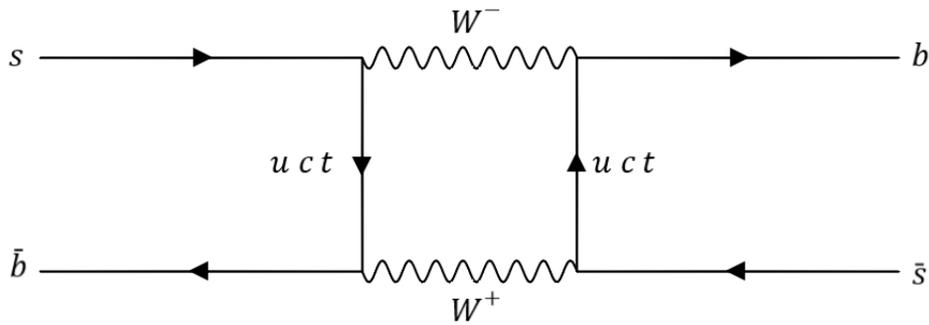


Figure 2.: This box Feynman diagram shows one possible way for mixing of B_s^0 and its antiparticle.

Direct CP-violation occurs when the probability of a particle decay from a particle T into its final state E is different from its antiparticle \bar{T} : $P(T \rightarrow E) \neq P(\bar{T} \rightarrow \bar{E})$
Violation in the interference between mixing and decay can occur when a final state E is accessible both by the decay of a neutral meson M^0 or its antiparticle \bar{M}^0 . This means the particle can either decay directly or by first mixing with its antiparticle. CP violation is present if one process is preferred.

3. The SuperKEKB collider

The SuperKEKB operates at a center of mass energy of 10.58 GeV, which is a slightly above the $\Upsilon(4S)$ resonance. $\Upsilon(4S)$ decays into $B\bar{B}$ over 96% of the time [15], which is why the SuperKEKB is called a B-Factory. A list of cross sections at this center of mass energy can be seen in table 1. The collider consists of two storage rings, one for the electrons at 7 GeV called the High Energy Ring, or HER, and one for the positrons at 4 GeV called the Low Energy Ring, or LER, which can be seen in figure 3. The differences in beam energy cause the center of mass to have a boost in the laboratory system. This boost is necessary to study time dependent phenomena such as certain CP-symmetry violating events.

The collider carries current in form of bunches which collide at the point where the beam pipes intersect, this point is called the Interaction Point (IP). During operation bunches are injected at full energy into both rings by the linear accelerator (Linac) in a process called beam filling. If the Linac stops injections, the current in the storage rings starts to drop in a process called beam decay with the dominant reason for this being the Touschek effect. The beam lifetime is defined as the time it takes for the beam to lose half its current. To counteract this during operation, the Linac will keep the beam currents stable with top up injections.

Table 1.: A list of cross-sections at the center of mass energy $\sqrt{s} = 10.58$ GeV [4].

$e^+e^- \rightarrow$	$\sigma[\text{nb}]$
e^+e^-	300 ± 3
$u\bar{u}$	1.61
$c\bar{c}$	1.30
$\mu^+\mu^-$	1.148
$\Upsilon(4S)$	1.110 ± 0.008
$\tau^+\tau^-$	0.919
$d\bar{d}$	0.40
$s\bar{s}$	0.38
$\nu\bar{\nu}$	0.25×10^{-3}

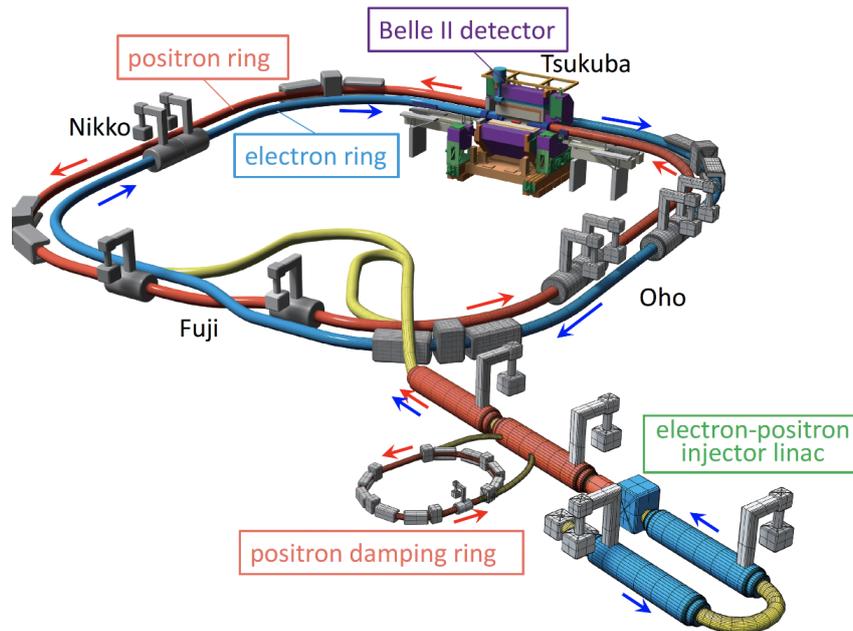


Figure 3.: This image shows the Linac, SuperKEKB and the Belle II detector [8]

3.1. Beam dynamics

Table 2.: A summary of beam parameters [10]

	HER		LER	
	Design	Achieved	Design	Achieved
Energy (LER/HER) [GeV]	7		4	
Lorentz boost $\beta\gamma$	0.28			
Crossing angle 2ϕ [mrad]	83			
Current [A]	2.6	0.82	3.6	0.85
σ_x^* [μm]	11	17	10	18
β_x^* [mm]	25	60	32	80
β_y^* [mm]	0.3	1	0.27	1
ε_x [nm]	4.6	4.6	3.2	4
ε_y [pm]	13	45	8.6	60
Luminosity [$10^{34} \text{ cm}^{-2} \text{ s}^{-1}$]	Design: 80		Achieved: 3.1	

SuperKEKB has two storage rings designed to carry particles along a specific path inside the beam pipe. The ideal trajectory is referred to as the design orbit along which a particle will propagate with exactly the right energy and position. In reality however, particles

deviate from this design orbit. For small enough deviations, they will perform oscillations transverse and longitudinal to the direction of motion. Transverse oscillations are referred to as betatron oscillations, whereas longitudinal oscillations are called synchrotron oscillations. To make sure that these oscillations stay stable, particles are continuously redirected to the design orbit by different types of accelerator elements, which will be discussed in the following.

Bending magnets: The SuperKEKB consist of four straight sections named Fuji, Oho, Tsukuba and Nikko that are connected by curved sections. To guide particles around these curves the accelerator is equipped with bending magnets that create a magnetic field perpendicular to the particles direction. Particles will experience a centripetal Lorentz force which keeps them on the design orbit, but also causes them to loose energy due to synchrotron radiation.

Radiofrequency (RF) accelerating cavities: To counteract the loss of momentum due to synchrotron radiation and to correct the energy of particles with too low momentum, there are RF cavities placed at the sections Nikko and Oho for the HER and Fuji and Oho for the LER that can give an energy boost to particles. The RF cavities create an oscillating electric field with a period tuned to the frequency of bunches passing. A particle that exactly follows the design orbit would get a boost that counteracts the loss due to synchrotron radiation. Particles with too little or too much momentum get corrected by receiving either more or less energy to maintain phase stability.

Focusing Magnets: While the acceleration cavities correct longitudinal deviations from the design orbit, transverse deviation are corrected by focusing magnets. This can partly be achieved by configuring the magnetic field of the dipole bending magnets to be inhomogeneous as seen in figure 4.

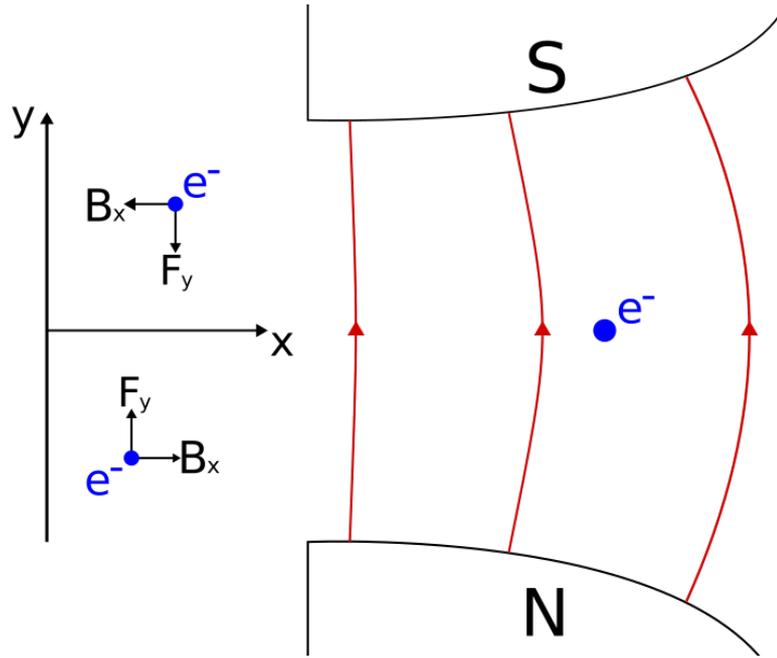


Figure 4.: An illustration of a dipole focusing bending magnet with movement direction of the electron into the plane.

Additionally there are dedicated focusing magnets of higher orders (quadrupole and sextupole) placed around the ring with much higher focusing strength than the dipole magnets. To increase luminosity bunches are "squeezed" by a series of superconducting magnets right before the IP which reduces beam emittance and is referred to as "nano beam scheme".

3.2. Beta function and emittance

The transverse emittance ε describes the particle distribution in the phase space and is defined as the area of the ellipse as seen in figure 5. The area of the ellipse is defined by the Courant-Snyder parameters α , β and γ as follows [3]:

$$\varepsilon = \gamma x^2 + 2\alpha x x' + \beta x'^2 \tag{1}$$

Where x describes the position and x' the velocity of particles.

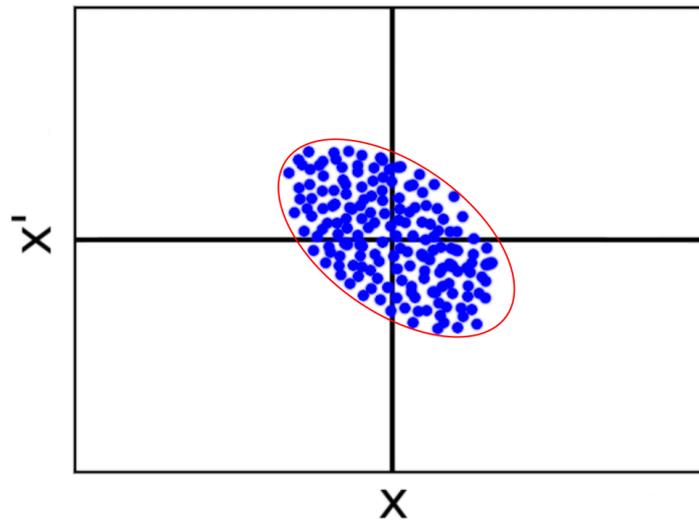


Figure 5.: The blue dots represent particles in the phase space where x is the position axis and x' the velocity axis. The ellipse is the smallest ellipse that envelopes the particles.

The beta function describes the positional distribution of particles by relating the standard distribution σ of a gaussian particle distribution, the emittance ε and the Courant-Snyder β :

$$\sigma = \sqrt{\varepsilon\beta} \quad (2)$$

3.3. Top up Injections

To keep the beam current stable new particles are continuously injected in already stored bunches. For this the stored bunch performs a curve away from the design orbit towards the beampipe into a "bumped orbit". Ideally, this would only affect the bunch being injected into, but the spacing between bunches is so tight, that close bunches are also partially affected. A new bunch is then injected by the injection septum magnet next to the stored bunch. Both bunches are brought back on the design orbit where the injected bunch will start performing betatron oscillations around the stored bunch.

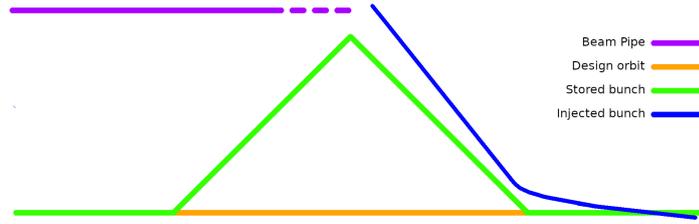


Figure 6.: This figure shows an illustration of the injection process

The particles in the injected bunch will emit synchrotron radiation due to betatron oscillation. The energy loss caused by this, together with the feedback system which corrects bunch orbits individually, will dampen the injected bunch and causes it to eventually merge with the stored bunch.

3.4. SuperKEKB movable collimator system

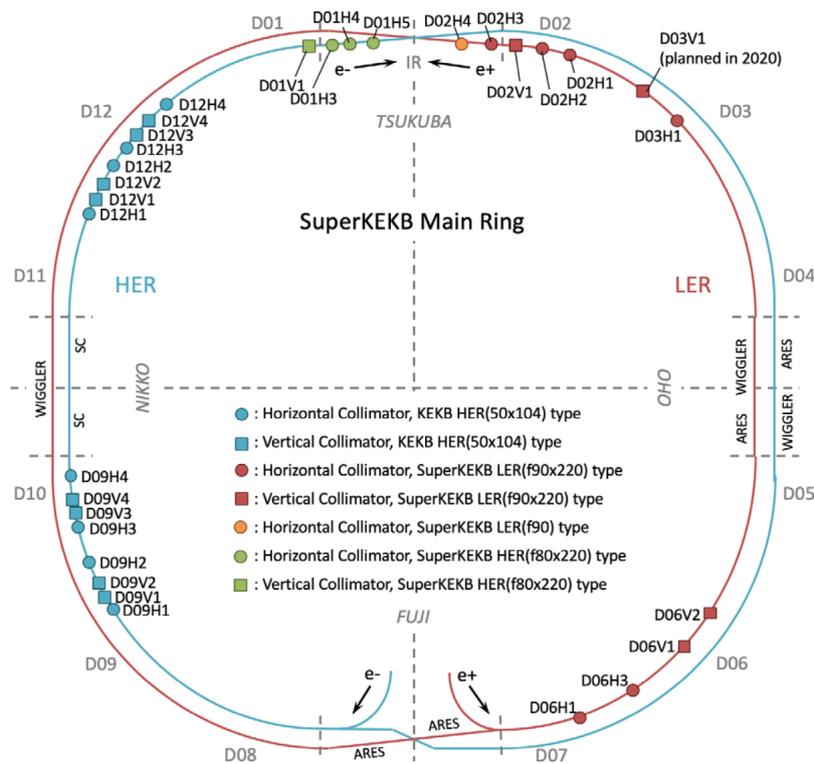


Figure 7.: An overview over the types and positions of collimators around the two storage rings [12]

3. The SuperKEKB collider

The movable collimators placed around both storage rings are a vital part of the beam optics. Collimators act as apertures for the beam. They consist of heavy metal blocks called jaws that can be set to varying distances to the beam center, where close to the center is referred as "closing" and further away as "opening" the collimator. The HER reuses a number of collimators from KEKB which are equipped with only one jaw. With the upgrade to SuperKEKB new collimators have been added with two opposing jaws that can be moved independently from each other. The main purpose of collimators is to cut off stray particles that diverge too far from the beam center, which prevents them from colliding with accelerator elements close to the IP and causing backgrounds. Closing the collimators generally improves the effect of background reduction but comes at the cost of reducing the, already short, beam life time.

4. The Belle II detector

The increase of luminosity and therefore backgrounds due to the upgrade from the KEKB to the SuperKEKB also required an upgrade of the detector Belle. This upgrade is realized with Belle II detector, which can be seen in figure 8. It is a hermetic detector located at the IP with an asymmetric design due to the difference in beam energy and can be divided into three sections: A forward section in the electron beam direction, a backwards section in the positron beam direction and a barrel section in between. Its purpose is to fully reconstruct events at the IP. This is somewhat limited by the fact that the beampipes have to enter the detector on each side making it impossible to detect particles that exit at angles outside of the angular acceptance of $17^\circ < \theta < 150^\circ$. Not all types of collisions are of interest and should be recorded. For example $B\bar{B}$ decays happen at a frequency of about 100 Hz. Bunches on the other hand pass the IP every few nanoseconds, often producing several collisions per passing. For this reason Belle II has a trigger system which uses the data of some of its subsystems to decide if an event is worth recording or not. If the detector is triggered all of the subsystems are used to record the important information of the event: The position of decay vertices of particles produced in collisions, the track of the particles in the detector, their momentum, energy and type of particle.

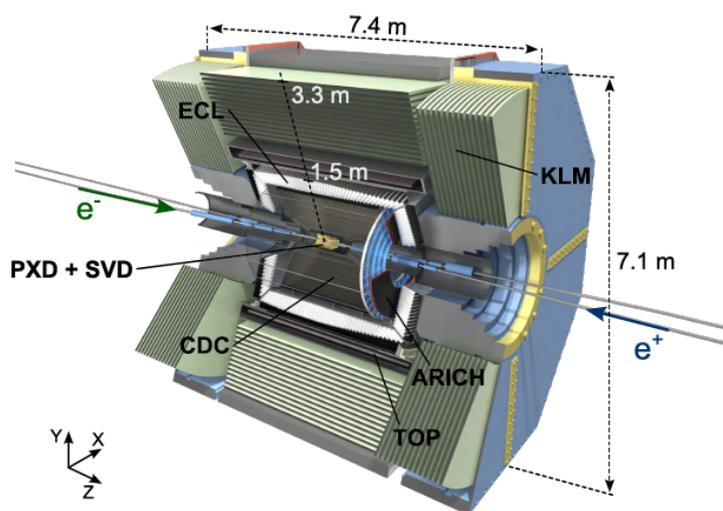


Figure 8.: An illustration that shows Belle II with positions of all subdetectors[1]. The z axis is parallel to the beam pipes at the interaction point, while the y axis is orthogonal to the plane of the accelerator and the x axis points outward, in the plane of SuperKEKB. The polar angle θ describes the angle between z and y components, while the azimuthal angle ϕ describes the angle between x and y components.

4.1. Pixel Vertex Detector

The Pixel Vertex Detector (PXD) makes up the innermost layer of the detector. It consists of silicon chips that are arranged around the beampipe. If a particle passes through these chips they can measure the 2D position of the passthrough and the energy of the particle lost inside the module. Together with the information of the position of the chip, the hit can be reconstructed in 3D space. Since the PXD is the closest subdetector to the beampipe its data is very important for the reconstruction of decay vertices.

The pixels are based on the depleted p-channel field-effect transistor (DePFET) technology. Particles passing through the pixel create electron hole pairs. The positively charged holes generally drift to the p-doped back plate and are not used for signal generation. The electrons on the other hand drift to the n-doped internal gate where they accumulate. With a distance of 1-2 μm to the internal gate, a field effect transistor (FET) can be used to read out the number of electrons in the internal gate by measuring its conductivity which is influenced by the electric field created by the electrons. This is done in a *rolling shutter* readout scheme by turning the FETs of four rows of pixels on for 100 ns every 20 μs . After all the pixels in the row are read out, the collected charge has to be cleared which is done by applying a positive voltage to a n-doped clear gate situated close to the internal gate. Since the current flowing through the FET rises linearly with the collected charge, it is not only possible to register if a pixel has been hit, but also to determine the amount of electrons in the internal gate and thus the energy lost inside the silicon by the particle.

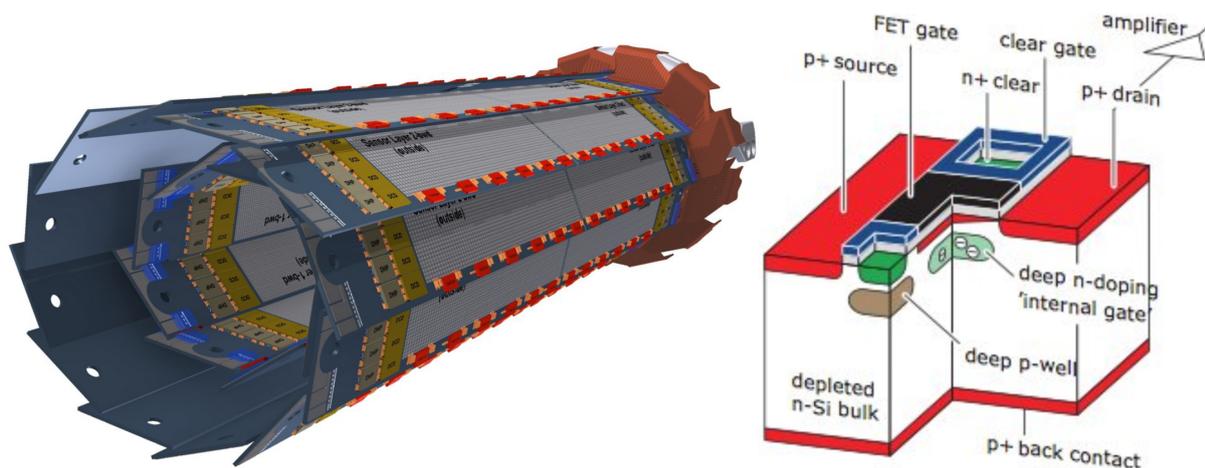


Figure 9.: On the left: An illustration of the PXD. The ladders are positioned in an overlapping manner to make sure the detector area has full coverage [11]. On the right: An illustration of a DePFET pixel [2].

Table 3.: A summary of the characteristics of the individual PXD and SVD layers.

	PXD		SVD			
	L1	L2	L3	L4	L5	L6
Radius [mm]	14	22	39	80	104	135
No. ladders	8	12	7	10	12	16
No. modules/ladder	2	2	2	2+1'	3+1'	4+1'
Thickness [μm]	75		320	320, 300'		
Mat. Budget [X_0]	0.19%		0.70%			
No. channels	3M	4.6M	22k	38k	61k	102k
Angle forward module [$\hat{\text{A}}^\circ$]	-		-	11.9	16.0	21.1

4.2. Silicon Vertex Detector

The Silicon Vertex Detector (SVD) is the second part of the Belle II vertex detector and is located around the PXD. The SVD is a double sided striped detector (DSSD) that forms four additional layers around the PXD as seen in figure 10. Each layer consist of several modules where the most forward module of the three outer layers is angled to allow for greater θ coverage without having to use more modules.

DSSDs consist of a depleted silicon bulk with p-doped silicon stripes on one and n-doped silicon stripes on the other side. The differently doped stripes are angled orthogonally to each other. A particle passing the bulk will create electron-hole pairs. Electrons drift towards the n-doped side while holes drift towards the p-doped side, creating a signal in at least one stripe on each side. Combining the signals from both sides, the 2D location of the hit on the chip can be reconstructed. Unlike the PXD, the SVD can measure the hit time accurately.

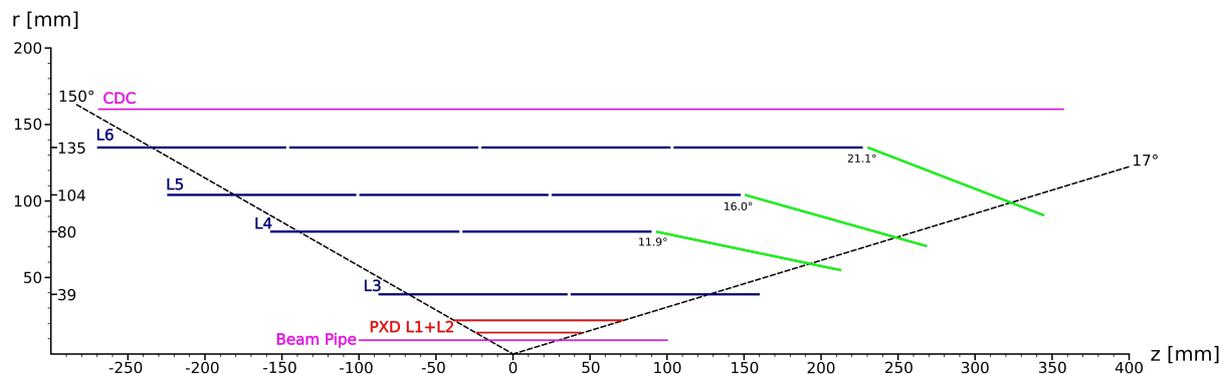


Figure 10.: This illustration shows the arrangement of the PXD and SVD layers.

4.3. Central Drift Chamber

The Central Drift Chamber (CDC) is the main tracking subdetector of Belle II. The CDC has the form of a hollow cylinder with an external radius of 1130 mm and an internal radius of 160 mm. The chamber is filled with a gas mixture of equal parts of He and C₂H₆. A total of 14 336 sense wires span across the inside of the chamber. These wires are divided into alternating layers of axial wires, which are parallel to the beam pipe and stereo wires, which are tilted by 45 mrad to allow measurement of the z component. This can be seen in figure 11. Particles that cross the CDC will ionize the gas and leave tracks of electron ion pairs. These will then be accelerated by the electric field created by the wires and ionize more gas particles causing a cascade. The particles follow the field lines of the wires and create an electric signal in the wire. This signal is used to reconstruct the particle track, identify the type of particle based on their energy loss inside the gas, and trigger the readout of other subdetectors.

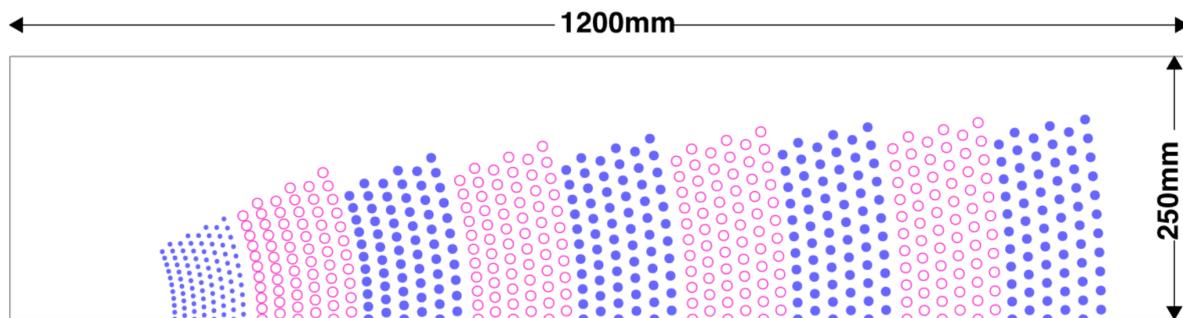


Figure 11.: Schematic view of the CDC. Blue dots represent axial wires while pink dots represent stereo wires [5].

4.4. Time of Propagation Counter

The Time of Propagation Counter (TOP) is a Cherenkov counter that is located in the barrel section between the CDC and the Electromagnetic Calorimeter and covers the polar angle $32.5^\circ < \theta < 123^\circ$. Its main purpose is particle identification (ID), in particular the separation of K^\pm and π^\pm . TOP consists of 16 2.5 m long and 2 cm thick quartz radiators. If a particle passes through the quartz at a higher velocity than light would have in that medium, it will produce photons due to the Cherenkov principle. These photons are emitted in form of a cone, the Cherenkov cone. The opening of this cone depends on the velocity of the particle and the material's refractive index. The light travels inside of the modules due to total internal reflection and are read out by photo sensors. The photo sensors consist of photomultiplier tubes that are arranged in a 2×16 array located at the

end of each module. An illustration of this principle can be seen in figure 12

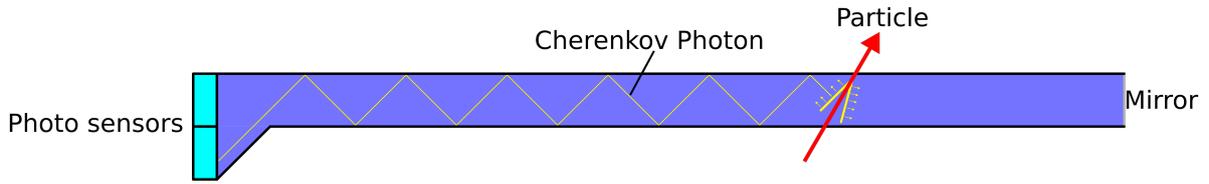


Figure 12.: 2D illustration of a quartz radiator with photo sensors.

4.5. Aerogel Ring-Imaging Cherenkov detector

The Aerogel Ring-Imaging Cherenkov detector (ARICH) is located at the forward side of the Belle II detector covering the range $14^\circ < \theta < 30^\circ$. It has many similarities in use and function to the TOP.

It is also part of the particle ID system and differentiates between K^\pm and π^\pm that do not have enough transverse momentum to pass through the TOP. The Cherenkov light is produced in two 2 cm thick aerogel plates with slightly different refractive indices. The produced photons will then propagate through a 16 cm gap before reaching a two dimensional array of hybrid avalanche photo detectors (HAPD) as seen in figure 13. Since the particle propagated through aerogel plates with different refractive indices, it will produce Cherenkov cones with two different opening angles. This causes the Cherenkov cones to overlap and improve the resolution of the cone opening angle.

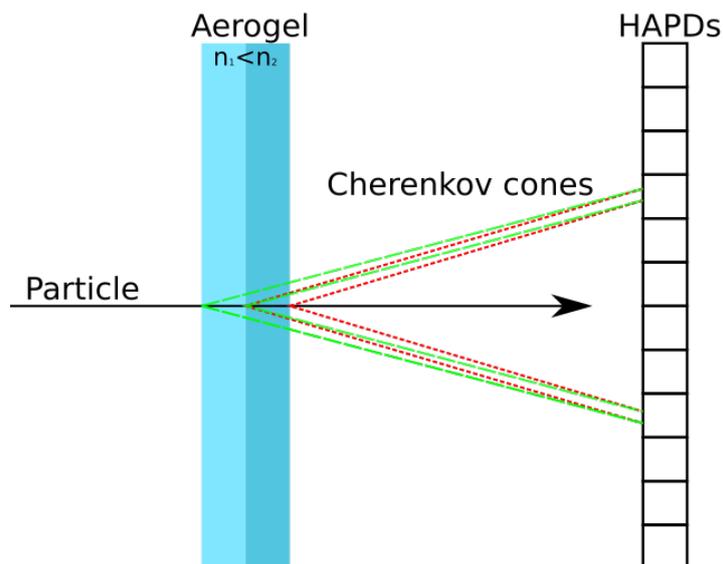


Figure 13.: Illustration of the ARICH that shows how different refractive indices create overlapping cones.

4.6. Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECL) surrounds all of the aforementioned subdetectors. Its main purpose is the detection of photons and measuring their energy. It also plays a vital role in particle identification. It consists of $6\text{ cm} \times 6\text{ cm} \times 30\text{ cm}$ 8736 CsI(Tl) crystals, arranged with the square base facing the IP, as seen in figure 14. High energy charged particles create photons in the material through Bremsstrahlung. These photons can then create electron positron pairs that can again produce photons giving rise to particle showers. Photons hitting the crystals will produce electron positron pairs directly causing the same effect. These showers will eventually stop when particle energy falls below the energy required to create new particles. The remaining energy will instead be dissipated by ionization and excitation of bound electrons which will produce scintillation light that is then detected by photodiodes which are located at the back of each crystal. The signal is used for reconstruction of the particles energy and for particle identification since the shape of the particle shower is different depending on the type.

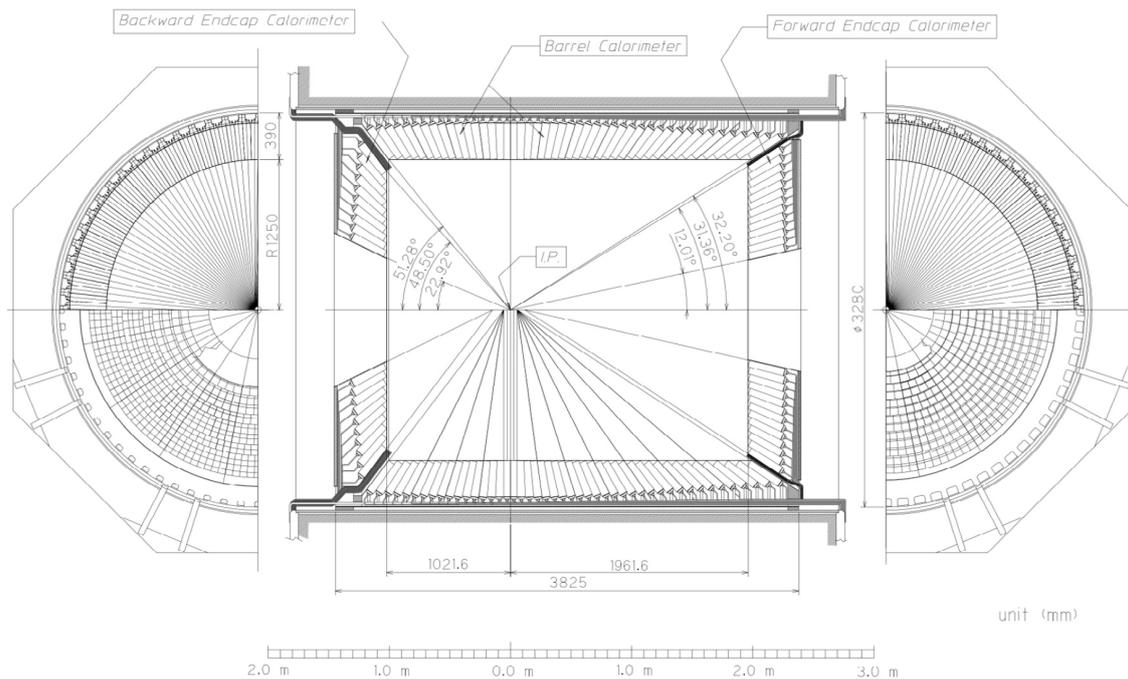


Figure 14.: Schematic view of the barrel section and the endcaps of the ECL [5].

4.7. The Solenoid

The solenoid creates a 1.5T magnetic field inside Belle II, parallel to the beam pipe. This is required to do momentum measurements in the CDC and vertex detectors. The

momentum results from the field strength and the curvature of the track. The solenoid encapsulates all detectors except for the The K_L -Muon Detector.

4.8. The K_L -Muon Detector

The K_L -Muon Detector, or KLM is the outermost layer of Belle II. It is used to detect muons, which typically do not interact much with the other layers of Belle II, and as an extended calorimeter for hadronic shower such as those caused by K_L . The KLM consists of alternating layers of 4.7 cm iron plates and sensors. The KLM provides an hadronic interaction length of $3.9 \lambda_I$ in addition to the interaction length of the ECL $0.8 \lambda_I$. This is sufficient for measurements of hadronic showers caused by K_L . The iron plates also serve as a return yoke for the solenoid and help the structural integrity of the whole detector.

4.9. Global Decision Logic

Since the total data produced from all readout channels combined is too much to handle, Belle II needs a system that can quickly identify interesting events based on a fraction of the available data. This is realized by implementing a hardware trigger called the Level 1 (L1) Trigger and a software based High Level Trigger (HLT). The L1 Trigger is given by the Global Decision Logic and works with subtriggers implemented on a hardware level on the subdetectors CDC, TOP, ARICH, ECL and KLM. The CDC provides first estimation of momentum, charge and track multiplicity. The ECL counts the number of isolated clusters and gives information about clusters that exceed certain energy thresholds. TOP and ARICH send timing information to the GDL while the KLM provides information about muon tracks. The GDL then compares these parameters to a list of predefined conditions to decide if the event gets discarded or sent to the HLT. If the event is kept the data acquisition system will read out the data from all subdetectors except the PXD and perform a full reconstruction of the event. The HLT will then make the final decision if the event gets archived or not. PXD data is treated differently from the data of other subdetectors. While the PXD readout is also triggered by the L1 trigger, its data is sent to a Field Programmable Gate Array instead of the HLT where it is saved for 5 s, awaiting a decision of the HLT. If the HLT decision is positive, part of the buffered PXD data is saved for offline processing.

5. Backgrounds

Backgrounds refer to hits in the Belle II detector that are not caused by signal collisions. They are generally caused by particle losses close to the IP. Particles that hit the beampipe or other elements causing particle showers that can enter the detector. While particle loss close to the IP is especially problematic, all particle loss around the ring will reduce beam life time.

The background sources at the SuperKEKB can be classified into three categories: Beam induced backgrounds which come from the beam interacting with itself or its surrounding environment like residual gas in the beam pipe which can happen anywhere in the storage rings.

Luminosity dependent backgrounds stem from interactions between the beams and are therefore local to the IP. The rate of this type of background scales with the luminosity, hence the name.

Injection backgrounds refer to the increase in background level observed after top up injections.

Beam induced backgrounds:

- **Beam-gas scattering:** Even though the beam pipes are evacuated the vacuum is not perfect and there are always residual gas particles left which can interact with the particles of the beam either through elastic Coulomb scattering or inelastic Bremsstrahlung. These interactions can throw particles too far off the design orbit to be recovered by the beam optics. These particles will then leave the stable orbit and eventually collide with an optics element or the beam pipe itself. The collisions produce secondary radiation that can reach the detector and therefore cause backgrounds. The rate of this background is proportional to the pressure and the beam current.
- **Touschek scattering:** The Touschek effect refers to an energy transfer between two particles in a single bunch through Coulomb scattering. Just like with beam-gas scattering this can cause the particles to diverge from the design orbit by receiving or losing too much energy in the process. The Touschek effect is proportional to the second power of the bunch current and inversely proportional to the beam size and the third power of the beam energy[14].
- **Synchrotron radiation:** Charged particles that experience centripetal acceleration emit photons due to synchrotron radiation. The turn wise energy loss ΔE of

the particle is proportional to [13]:

$$\Delta E \propto \frac{E^4}{\rho}$$

with the energy of the particle E and radius of the turn ρ . Photons emitted this way can cause backgrounds in the detector, especially those emitted right before the IP.

Luminosity dependent backgrounds:

- **Bhabha scattering:** Elastic scattering between positrons and electrons $e^+e^- \rightarrow e^+e^-$ at the IP can cause particles to come off course and collide with the beam pipe or other elements further down the beam line. These collisions can cause scattering of particles back in the direction of the IP which causes hits in the detector. As this is an interaction between the beams its rate gets higher with increasing luminosity. This process can also produce photons in the end state $e^+e^- \rightarrow e^+e^-\gamma$ which is referred to as radiative Bhabha scattering. The rate of this process is proportional to its cross section times the luminosity.
- **Two photon process:** Interactions between positrons and electrons can create another low momentum lepton pair with the two photon process: $e^+e^- \rightarrow e^+e^-\gamma\gamma \rightarrow e^+e^-\ell^+\ell^-$. These new leptons are generally a positron and an electron but it is possible for muons to be created on rare occasions as well. Due to the magnetic field, these particles will then spiral inside the detector causing several hits in the inner subdetectors.

Injection backgrounds: After top up injections, a period of elevated background levels is observed. This is caused by the emitted synchrotron radiation during the merging process of the newly injected bunch and the stored bunch. Also, not all particles end up merging into the stored bunch, a fraction will always be on unstable orbits and collide with the beam pipe or optics elements causing backgrounds. The bunch causing the elevated backgrounds is referred to as noisy bunch. The revolution period of bunches in the SuperKEKB is about 10 μs .

The time it takes for the noisy bunch to stop causing elevated backgrounds, a process called cooling, is in the order of several milliseconds. Understanding what influences this cooling time is the subject of this work.

6. Artificial Neural Networks

Consider two sets I and O and a function f that assigns each element of O to exactly one element of I . Knowledge of the sets and the assignments allows for an approximation of the function f . This approximation can be realized by an artificial neural network (ANN).

6.1. Network architecture

Neural networks consist of artificial neurons which are connected by weights. The neurons are often sorted into layers of which there are three types: The input layer, hidden layers of which there can be any number and one output layer. This structure of an ANN can be seen in figure 15. Parameters that define the structure of a neural network, such as number of hidden layers and neurons per hidden layer, are called hyperparameters.

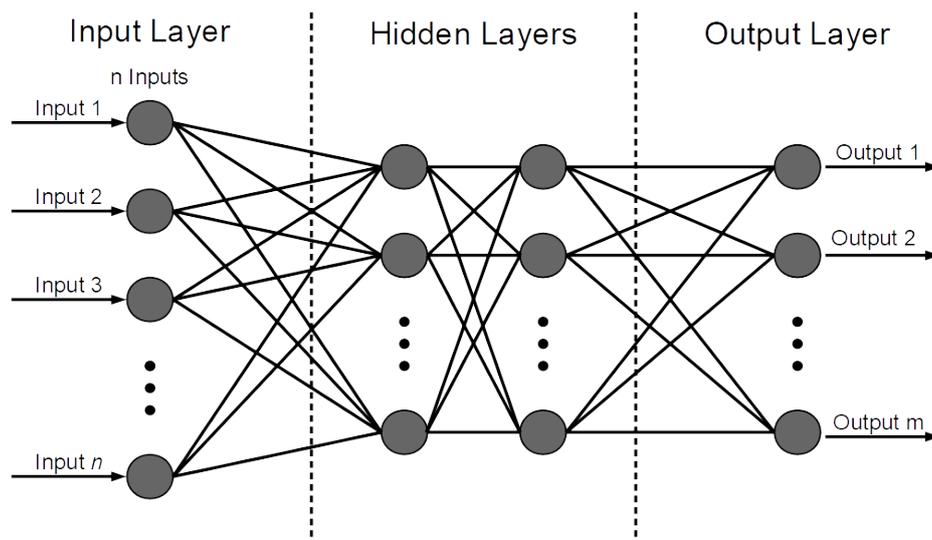


Figure 15.: Illustration of a feed forward artificial neural network.

A model constructed according to the set hyperparameters will then approximate f , in a process referred to as learning, by using a dataset of a finite number of pairs I_i and O_i where I represents inputs and O represents outputs. I_i and O_i are usually multidimensional. The input values enter the neural network through the neurons in the input layer. Each node in the next layer has a value equal to a bias and the sum of all the values of the previous layer multiplied by weights. The new value then gets passed through an activation function $g(z)$ before moving to the next layer. An illustration of this process can be seen in figure 16. Activation functions allow neural networks to be non-linear be-

cause without them, it would be a composition of linear functions and therefore be linear. Networks where information propagate only into one direction are called feed forward neural networks.

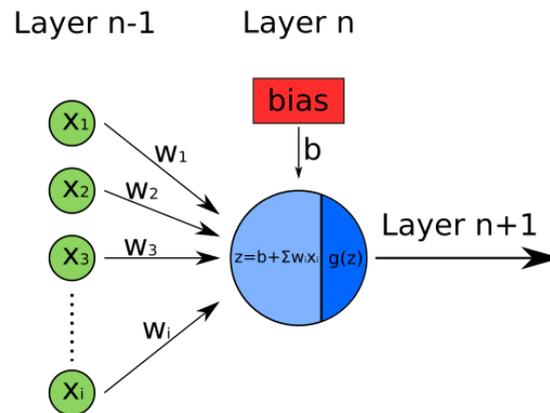


Figure 16.: A visualization of how a neuron works. It sums over the values x of the neurons of the previous layer multiplied with the weights w and adds the bias b . The result z is then inserted into an activation function g before being passed to the next layer where it gets processed as an input for the next set of neurons.

6.2. Network training

A neural networks approximates the function f by changing its weights and biases during a training phase. For a relatively small neural network with 50 inputs, 2 hidden layers at 32 neurons each and one output, the total amount of weights and biases is 2720. This high number of free parameters allows neural networks to approximate complex functions. A network starts training with a set of weights and biases which are generated based on the initialization method used. A common way to initialize a network is to set all biases to zero and generate a uniform distribution of weights in the range $[-limit, limit]$ where $limit$ is a function of the number of neurons in the layers that are connected by the weights. This random network is then applied to the first batch of data and the outputs are used in a loss function to evaluate its performance. A simple example for such a loss function would be the *mean absolute error*, meaning the mean absolute of the difference between the output and the true value of the training target. The weights and biases are then adjusted to minimize the loss function, with the most common method of doing this being stochastic gradient descent. The training will continue in batches until all the training data has been used. This process is then repeated for a number of *epochs*. A fraction of the data will always be withheld from the training and instead be used for validation. This

validation happens at the end of an epoch and evaluates the generalization capabilities of the network. The final output are the weights and biases that resulted in the lowest validation loss.

7. The Injection Duration Network

7.1. Injection Duration

After charge is injected into a bunch during the top up injections process, the Belle II detector experiences short durations of massively elevated backgrounds. During this background increase a veto is applied to the L1 trigger, preventing all data readout. There are three parts to this veto.

The first part is right after the injection, where the L1 trigger is blocked no matter where the noisy bunch is.

The second and third part are localized around the time when the noisy bunch passes the IP. The time it takes for the backgrounds to return to normal is referred to as the Injection Duration. For this thesis the ECL occupancy is used to define the timeframe of increased backgrounds. Keeping this time as low as possible is desirable, because the veto is a big part of the deadtime of Belle II.

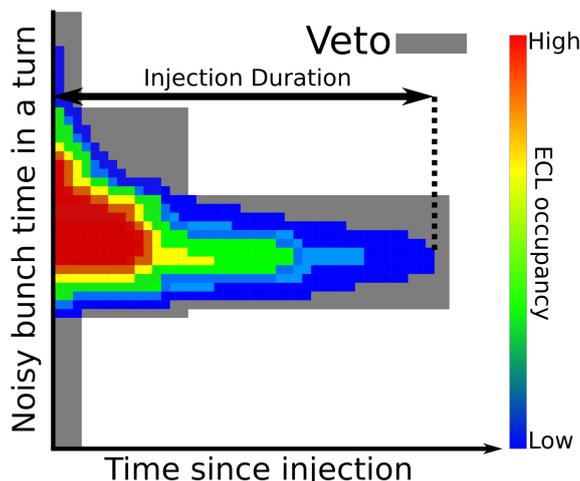


Figure 17.: This figure shows the behaviour of the ECL occupancy after injections. The color gradient denotes the average ECL occupancy over a baseline without injections from red (high) to blue (low). The injection duration is defined as the time from injection to the end of increased ECL occupancy. The three vetos are applied in a way that covers most of the background increase.

The Injection Duration is in the order of 5 ms to 25 ms for the LER and 1 ms to 10 ms for the HER. These numbers show that the values can vary widely depending on machine configurations. The exact dependencies are not yet fully understood. Modeling the injection duration with neural networks can help to understand the dependencies of the injection duration. Since neural network are not intrinsically transparent in how they produce their outputs, tools are required to understand the importance of specific inputs for the outputs.

7.2. Preprocessing

The networks work with 1 Hz data from the EPICS archiver at KEK which stores a wide variety of machine data and measurements from both SuperKEKB and Belle II. The variables that contain the information about machine parameters and measurements are referred to as PV. The PVs that holds the information about the injection duration are called "TRGOSCILLO{0,1}lff{1,h}er" for LER and HER respectively. The input PVs for the models are based on the target. The HER model only gets inputs specific to the HER, the same goes for the LER.

Before training, there are several data selection steps:

- The goal of the networks is to describe the injection duration during stable physics runs. For this the currents of both rings need to be above a defined threshold to mask data where one beam is off and their standard deviation has to be below a certain threshold to mask beam fillings and beam decays. The thresholds can be seen in table 4.
- There are cases of "bad data" where either one of the beam currents or the PV for the injection duration stays at a constant value for a long time. This data gets masked.
- Only a subset of the available PVs is used as input for each network as the total number of PVs is simply too large.

After data selection the data gets resampled from 1 s to 5 min averages. This helps reduce noise in the data which the network could use to overfit. The input data gets scaled with the following formula:

$$x_{\text{scaled}} = \frac{x_{\text{unscaled}} - u}{q_{0.17} - q_{0.83}} \quad (3)$$

where x is the value of the input, u is the mean of that input and $q_{0.17} - q_{0.83}$ are the 17% and the 83% quantile respectively.

This scaling is done because neural networks are better at handling input data that is between -1 and 1 and without large spread between different inputs. Quantiles are used to make scaling robust against outliers.

The target data is scaled similarly, but without subtracting the mean:

$$y_{\text{scaled}} = \frac{y_{\text{unscaled}}}{q_{0.17} - q_{0.83}} \quad (4)$$

7.3. Network Training

Table 4.: A list of the important parameters used for construction and training of the networks and data selection. While these values are used for all the models used in this thesis, they can be changed for different configurations of the networks.

Parameter	Value
Number of hidden layers	2
Neurons per hidden layer	32
Loss function	mean squared error+L2 regularizer
L2 regularizer constant λ	0.001
Optimizer	Adam
Activation function	Hidden Layers: $g(z) = \tanh(z)$ Output Layer: $g(z) = \ln(1 + \exp(z))$
Training epochs	100
Training patience	20
Learning rate scheduler	$0.01 * 0.9^{\text{steps}/10^4}$
Number of inputs for each network	87
Current threshold	50 mA
Current maximal standard deviation	5 mA

After preprocessing the data, the network is created and trained with the parameters seen in table 4. The timeframe of used data usually is one month, but narrower timeframes can be selected if desired. The loss function, which the network will try to minimize during training, consists of the mean squared error and an L2 regularizer:

$$\text{Loss} = \underbrace{MAE}_{\text{Mean squared error}} + \underbrace{\lambda \sum_{i=1}^n w_i^2}_{\text{L2 regularizer}} \quad (5)$$

The regularizer applies an additive term to the loss function that penalized high weightings, which can be a result of overfitting.

An example of the reduction of the loss function during training can be seen in figure 18

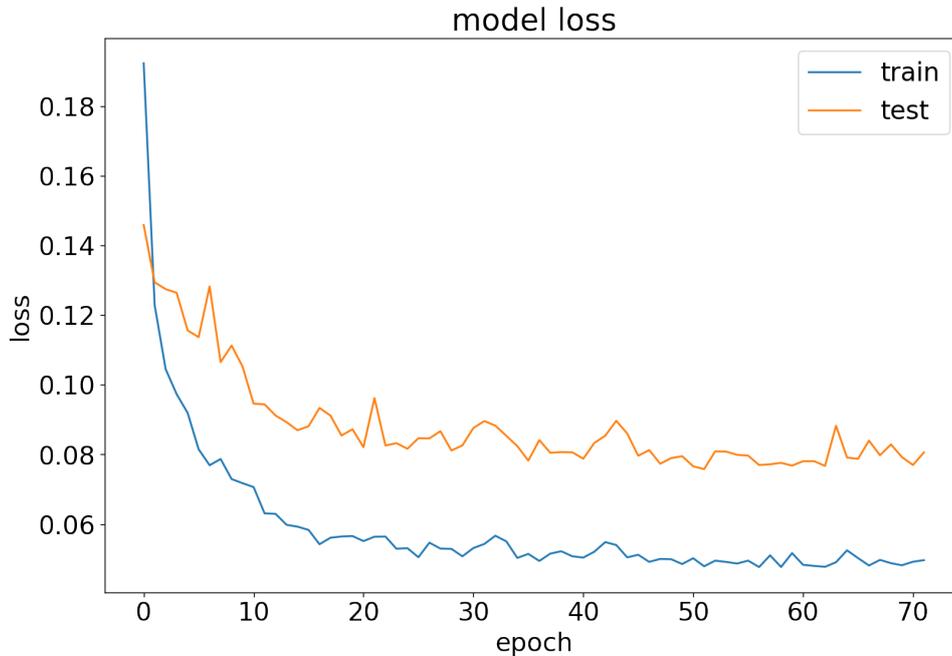


Figure 18.: The trainings loss is the average of all the batch losses calculated during an epoch. The validation loss is calculated once at the end of an epoch. The training stops either after the maximum number of epochs is reached or once there is no reduction in the validation loss for a number of epochs specified by the training patience.

7.4. Prediction and generalization

A trained model can then be used to predict the injection duration. Using data that has been used in training as input will almost always lead to good results. Results based on data that has not been used in training depend on the networks generalization capabilities. Figure 19 shows the generalization capabilities of a network which has been trained on a part of December 2021.

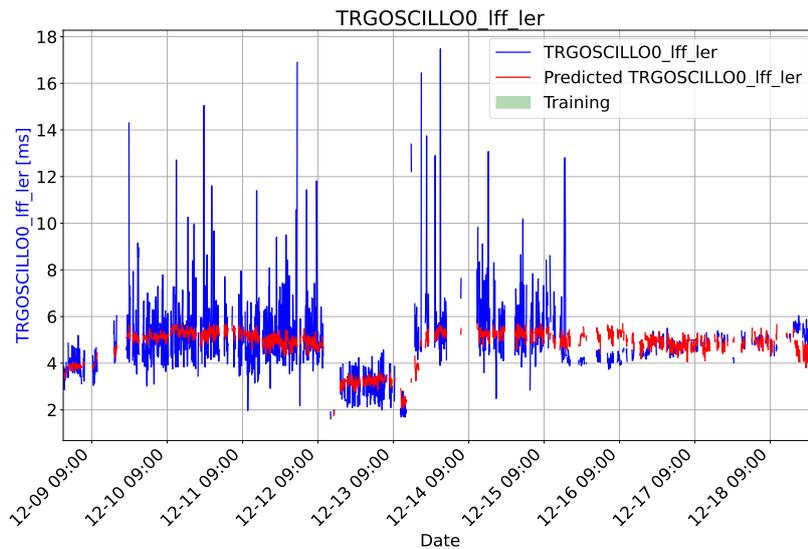


Figure 19.

A network's generalization capabilities can further be hindered by a process called overfitting. This happens when a network fits to noise or statistical fluctuations in the training target. By doing this, the network will learn undesirable relations between the input and target data. An example of this is using the noise of input data to produce results in the target data.

7.5. Path Explain feature attribution

Path explain is an open source tool that can help understand the importance of specific inputs for changes in the output of a neural network [7]. To do this, a test and a reference dataset have to be defined. Path explain will then calculate the attributions of each feature (input of the neural network) at every point in the test dataset using expected gradients:

$$\text{ExpectedGradients}_i(x) := \int_{x'} \left((x_i - x'_i) \times \int_{\alpha=0}^1 \frac{\delta f(x' + \alpha(x - x'))}{\delta x_i} d\alpha \right) p_D(x') dx' \quad (6)$$

x (vector) is a point in the test dataset consisting of the features x_i . x' (vector) is a point in the reference dataset with the density p_D . $f(x)$ is a scalar valued smooth function of x , in this case the neural network. The sum over all the attributions calculated in this way is equal to the difference $f(x) - \langle f(x') \rangle$ where $\langle f(x') \rangle$ denotes the average value of $f(x')$ over the reference set.

In the following the path explain feature attribution method will be demonstrated on data

from June 2021. Oscillations in the HER injection duration were observed between the 12.06.21 and the 14.06.21. To find an explanation for the oscillation the HER injection network is first trained on the data from June 2021. The reference dataset is defined as the data around the minima of the oscillation and the test dataset are chosen around the maxima as seen in figure 20.

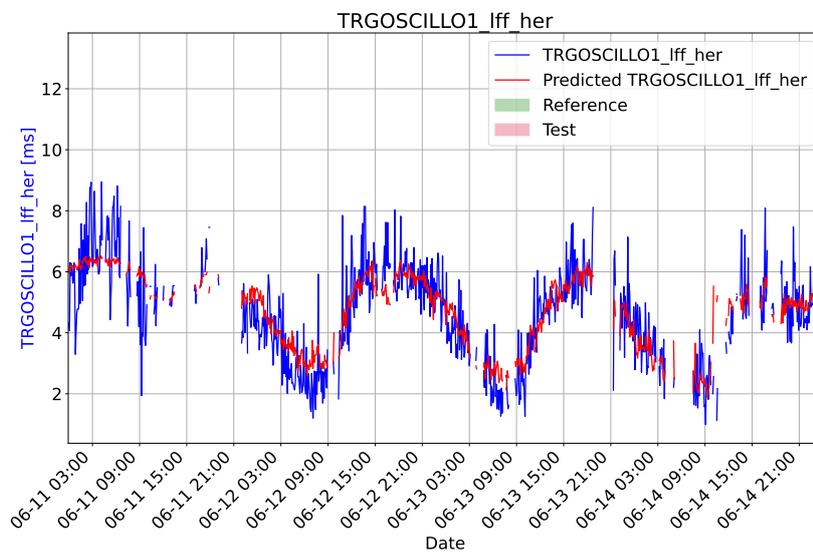


Figure 20.: The blue line represents the observed value of the injection duration. The red line describes the output of the model. The highlighted areas are the timeframes used as the test and reference datasets for path explain. The gaps in the graph are results of the preprocessing as described in section 7.2.

Path explain will then calculate attribution values based on the datasets and the model outputs. The ranking of the features is based on the mean absolute value of the attributions and can be seen in figure 21

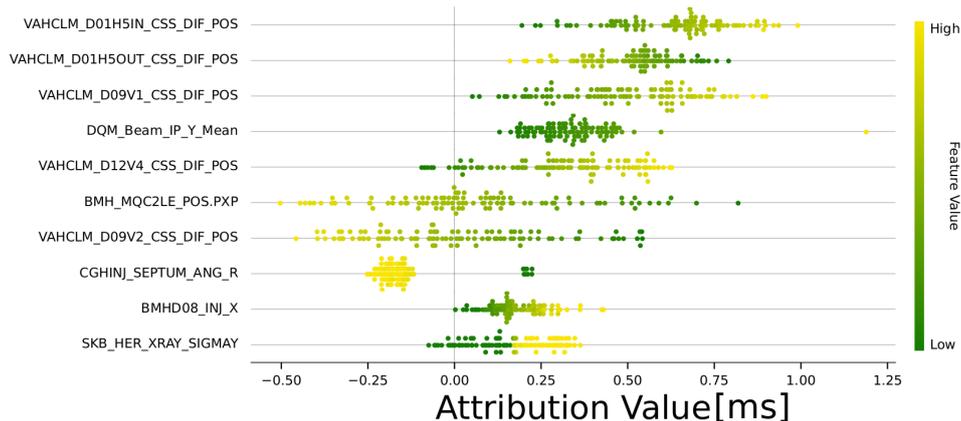


Figure 21.: The attributions of the top 10 features according to the path explain feature attribution method. Each dot represents one point in the test dataset. The colour of the dot shows the relative value of the feature at that point. The position of the dot on the x-axis shows the attribution of that feature to the difference in the model output between the test point and the reference dataset.

The attribution ranking shows what the model deems as important features that drive the changes from the reference data to the test data. The ranking can be used as an indication to what PVs are worth looking at. If we plot the top four features as seen in figure 22 we can see that they all follow the same kind of oscillation. The top three PVs describe collimator positions relative to the beam position. These PVs adhere to the following naming convention: "VA{H,L}CLM_D{ii}{V,H}j_CSS_DIF_POS" where {H,L} stand for HER or LER, D_{ii} denotes the ring section where the collimator is located and {V,H}j describe vertical and horizontal collimators with j differentiating collimators of the same type within a section. The optional parameters "IN", "OUT", "TOP" and "BTM" are used for collimators with two jaws as each jaw has its own PV. The fourth highest PV shows the y position of the interaction point. This information indicated, that the oscillations are caused by changes in the beam position.

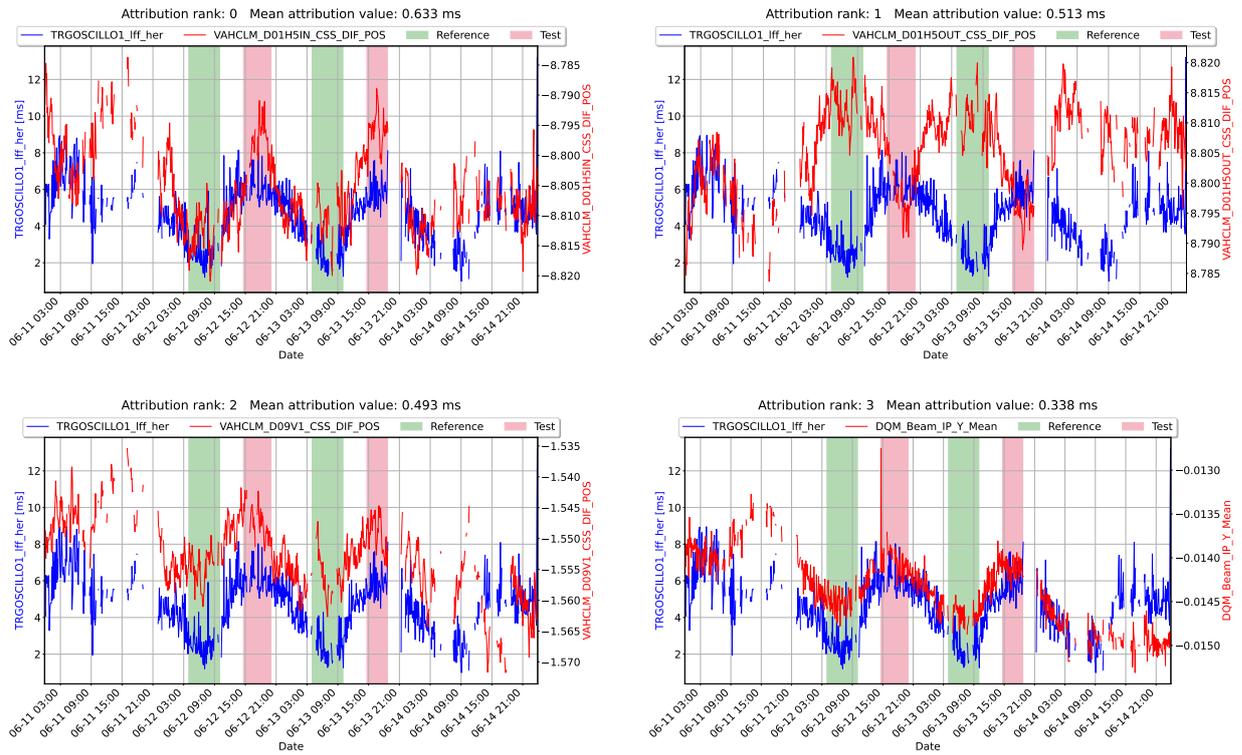


Figure 22.: The top 4 features according to the feature attribution also show an oscillation.

7.6. Collimator scans

Collimator scans describe the process of moving the collimator jaws to test its effect on parameters like the beam life time, background levels and injection duration. Collimator scans that reduce the injection duration offer a opportunity to test the network as the cause of the reduction is known. Figure 24 shows the injection duration during the 23.03.22 and 24.03.22. On the 24.03 collimator scans have been done from 10:22 to 11:25. As a result of these scans the jaw positions of the D06V1 and D06V2 collimators have changed, resulting in a drop in the injection duration. By choosing the reference and test datasets just before and after the scans, it is possible if the network correctly attributes the drop to the collimator positions.

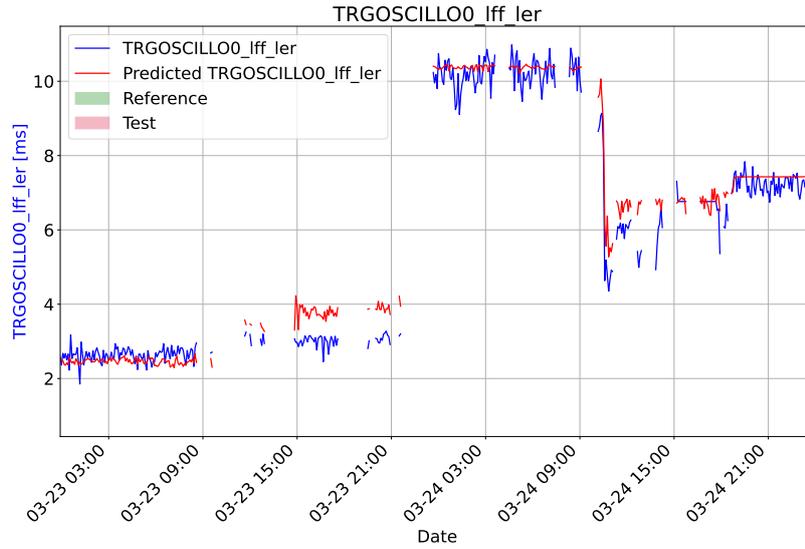


Figure 23.: The observed and predicted injection duration. The highlighted areas show what data is used in the test and reference datasets.

As seen in the attribution plot in figure 24 three of the top four features consist of collimator jaw positions, with the top jaw of D06V1 having the by far highest attribution.

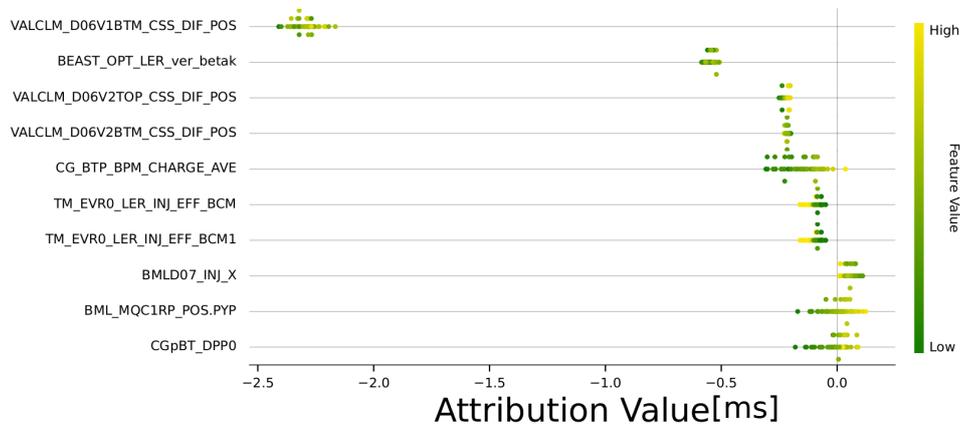


Figure 24.: The top 10 features according to the path explain feature attribution.

7.7. Correlation coefficients

The method that currently is most commonly used is calculating correlation coefficients. Calculating correlation coefficients for PVs and the injection duration assigns every PV a number between 1 and -1 based on how strongly that feature correlates with the injection duration. Creating a ranking based on the absolute value of these coefficients it is possible

to compare the neural network approach to the correlation coefficient approach. Table 5 shows such a ranking based on the same data that was used in test and reference datasets used for the collimator scan attribution seen in figure 24. The collimator positions are ranked lower in the coefficient ranking than in the attribution ranking. The four highest ranked PVs in the coefficient ranking are related to the injection process. These PVs are generally very important for the injection duration but not the cause for the reduction in this case. This comparison shows that the attribution method allows for more control over what specific changes in the injection duration should be analyzed compared to the more general approach of the correlation coefficients.

Table 5.: Ranking of the 8 PVs with the highest correlation to the injection duration.

PV name	Absolute value of correlation coefficient
CGLINJ_SEPTUM_ANG_R	0.832
CGLINJ_SEPTUM_POS_R	0.820
CGLINJ_KICKER_JUMP_R	0.737
CGLINJ_KICKER_2_ANGLE	0.737
VALCLM_D06V1BTM_CSS_DIF_POS	0.718
TM_EVR0_LER_INJ_EFF_BCM	0.616
VALCLM_D06V2TOP_CSS_DIF_POS	0.428
VALCLM_D06V2BTM_CSS_DIF_POS	0.305

7.8. Limitations

While the prediction of the model is generally accurate on training data, it can not generalize indefinitely beyond it. The network works with a very limited subset of all PVs that describe the SuperKEKB configuration. Gradual changes to PVs that are not used as an input can deteriorate the networks performance. Such deterioration can occur over weeks of machine operation or more quickly due to beam aborts or maintenance days. During maintenance days a lot of machine parameters are changed at once which makes it hard for the network to generalize beyond it if only data from before or after the maintenance day is used. An example for this can be seen in figure 25. If a maintenance day has an effect on the injection duration, it can be difficult to determine what changes caused this effect as a lot of PVs change at once.

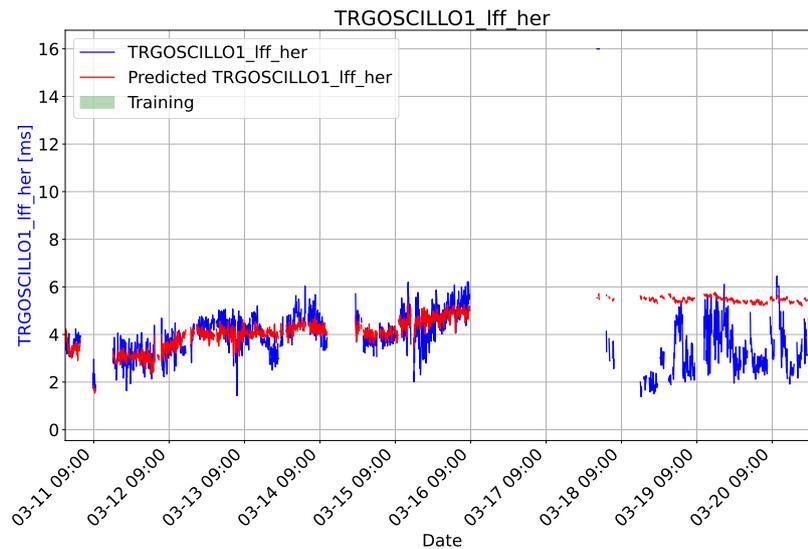


Figure 25.: This figure shows a models loss of generalization over a maintenance day. The model was trained with data from the beginning of March 2022 up to the maintenance day on the 16.03.22

8. Conclusion

In this Bachelors thesis a machine learning based approach to better understand particle collider backgrounds has been introduced. Neural networks have the ability to find dependencies between data but are usually very intransparent in doing so. While a neural network might provide a very accurate output based on a set of inputs, it is not intrinsically clear which inputs are important for the output. To solve this issue the path explain feature attribution method has been introduced. This method allows to find out what inputs are important for changes in the output. This is done by calculating attribution values for each input based on which the importance of inputs can be ranked. This information can be relayed to experts to help find causes of increases in the injection duration and assist in mitigating these. This offers an alternative to the currently used method of calculating correlation coefficients. While the network currently works with offline data, it is possible to integrate network training and creation of attribution rankings in a pipeline that works on online data directly at KEK. This would allow for an instantaneous feedback that with better accessibility. While this thesis focuses on the injection duration as the training target, other training targets are also feasible. Examples for this include the beam lifetime or the luminosity. Together with a new training target, the network would need a reconfiguration of hyperparameters and inputs to be optimized.

Bibliography

- [1] Super KEKB and Belle II. URL https://www.belle2.org/project/super_kekb_and_belle_ii. aufgerufen am: 27.2.2022.
- [2] Kodyš et al Andricek, L. Advanced testing of the DEPFET minimatrix particle detector. *Journal of Instrumentation*, 7(1), January 2012.
- [3] E.D Courant and H.S Snyder. Theory of the alternating-gradient synchrotron. *Annals of Physics*, 3(1):1–48, 1958. ISSN 0003-4916.
- [4] E. Kou et al.. *The Belle II Physics Book. Progress of Theoretical and Experimental Physics, 2019*.
- [5] I. Adachi et al. *Detectors for extreme luminosity: Belle II*.
- [6] J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay. Evidence for the 2π decay of the k_2^0 meson. *Phys.Rev.Lett.* 13, 1964.
- [7] Joseph D. Janizek, Pascal Sturmfels, and Su-In Lee. Explaining explanations: Axiomatic feature interactions for deep networks. *Journal of Machine Learning Research*, 22(104):1–54, 2021. URL <http://jmlr.org/papers/v22/20-1223.html>.
- [8] KEK. Kek press kit. URL <https://www.kek.jp/ja/imagearchive/2018/03/22/0900/>. aufgerufen am: 25.2.2022.
- [9] MissMJ. Standard model of elementary particles. Licensed under Creative Commons Attribution 3.0.
- [10] Yuki Yoshi Ohnishi and Abe et al. Accelerator design at SuperKEKB. *Progress of Theoretical and Experimental Physics*, 2013(3), 03 2013. ISSN 2050-3911. 03A011.
- [11] Srebre, Matej, Schmolz, Pascal, Hashemi, Hosein, Ritter, Martin, and Kuhr, Thomas. Generation of belle ii pixel detector background data with a gan. *EPJ Web Conf.*, 245:02010, 2020.

Bibliography

- [12] S. Terui et al. T. Ishibashi. Movable collimator system for superkekb.
- [13] R.P. Walker. Synchrotron radiation. URL <https://cds.cern.ch/record/398429/files/p437.pdf>. aufgerufen am: 26.2.2022.
- [14] A. Xiao and M. Borland. Touschek effect calculation and its application to a transport line. URL <https://accelconf.web.cern.ch/p07/PAPERS/THPAN098.PDF>. aufgerufen am: 25.2.2022.
- [15] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020. doi: 10.1093/ptep/ptaa104.

I. Lower ranking features

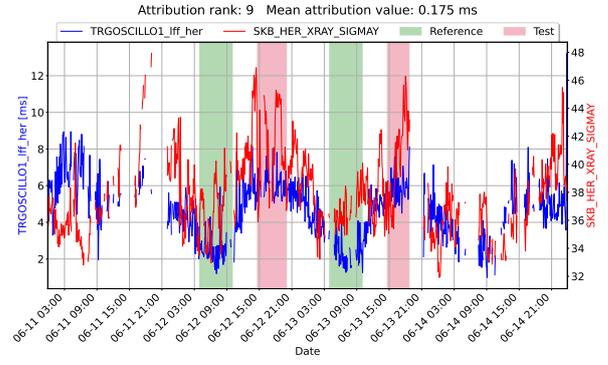
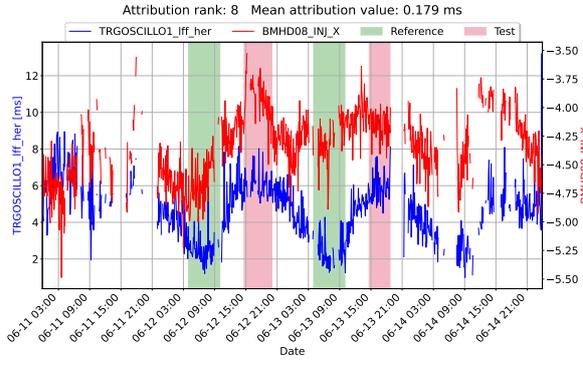
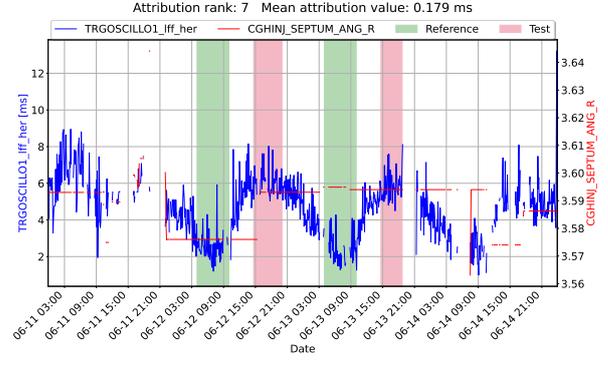
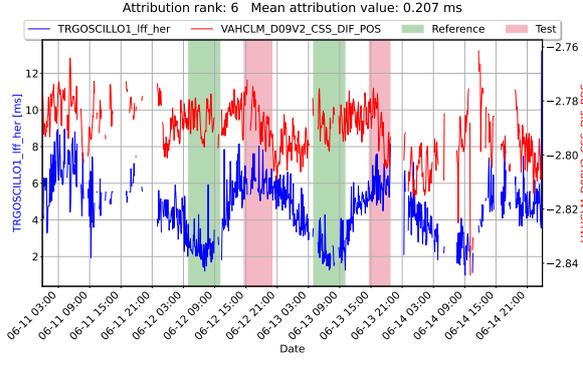
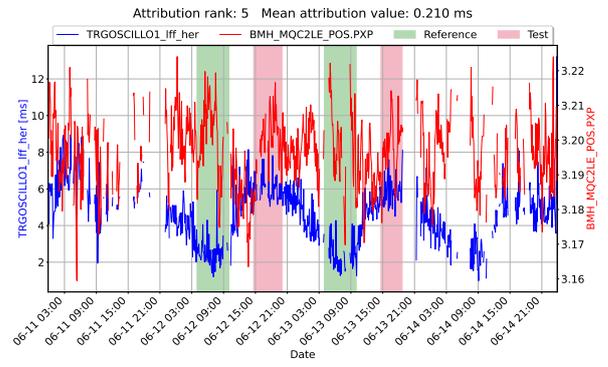
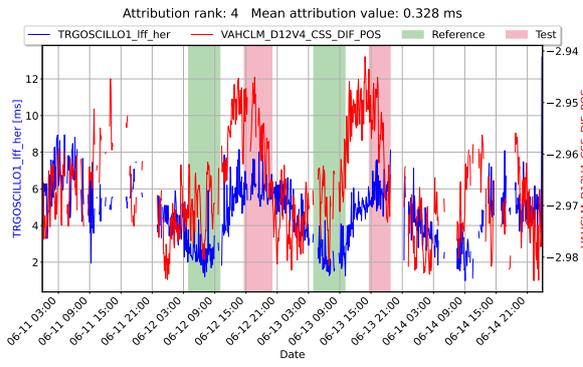


Table 6.: Features that are ranked 5 to 10 in the feature attribution 21

Bibliography

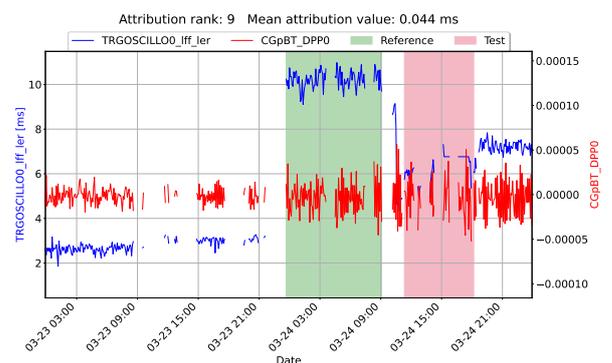
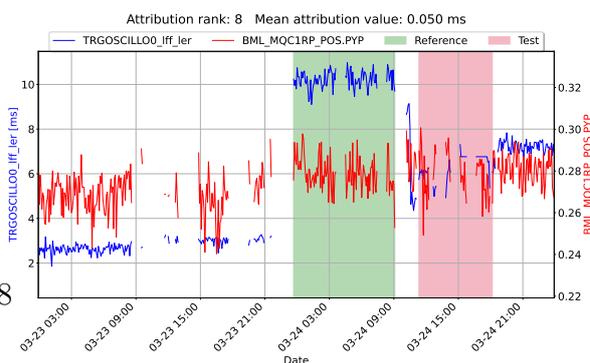
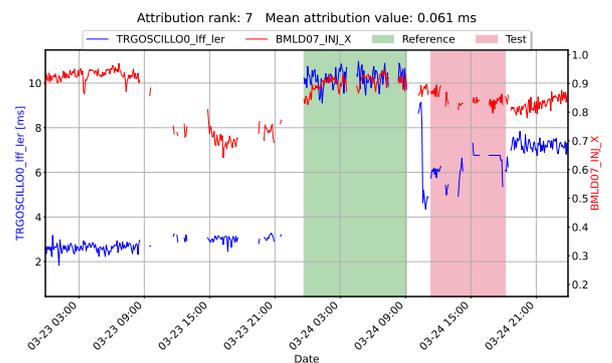
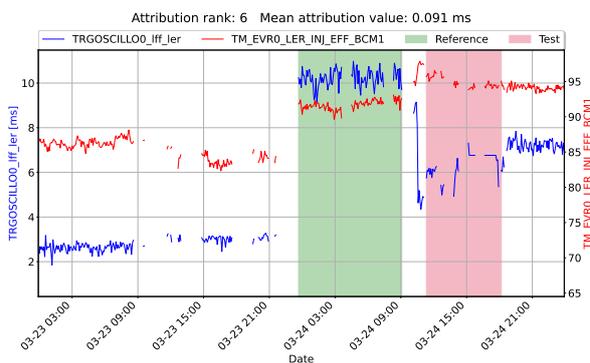
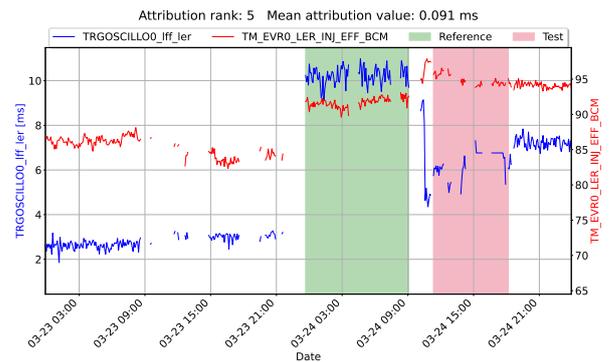
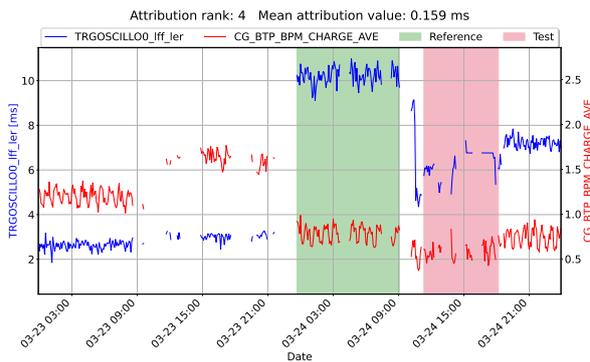
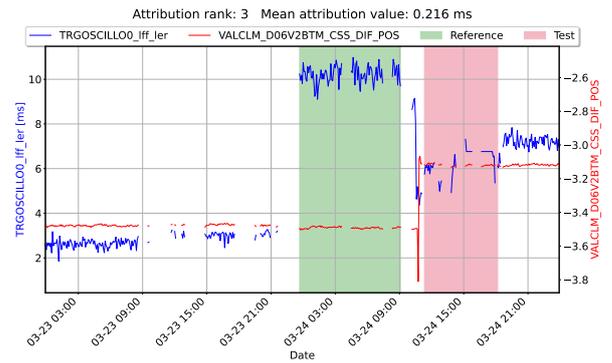
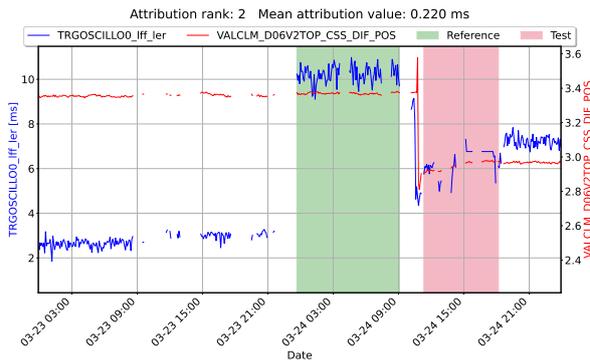
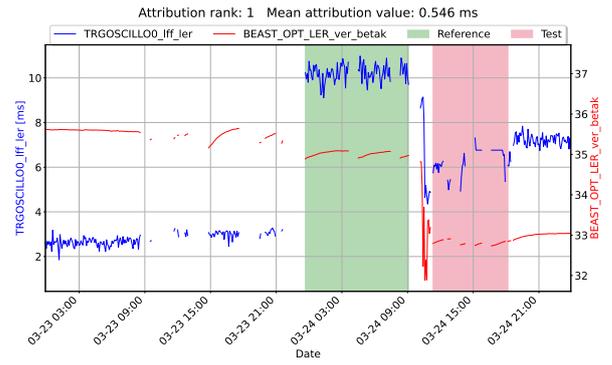
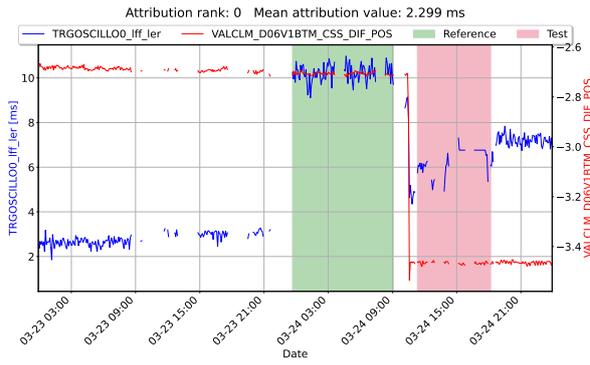


Table 7.: Features ranked 1 to 8 in the feature attribution 24

Erklärung

nach §13(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen: Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 13. Juni 2022

(Lukas Herzberg)