Unterschrift des Betreuers

DIPLOMARBEIT

Multivariate analysis methods for the optimization of searches for new physics with leptons at Belle II

Ausgeführt am Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften

unter der Anleitung von

Privatdoz. Dipl.-Ing. Dr.techn. **Christoph Schwanda**

und

Dr. **Gianluca Inguglia**

durch

Lukas Goldschmied BSc

Matrikelnummer: 01527387

Ort, Datum

Unterschrift des Verfassers

Abstract

This thesis will cover the use of multiple multivariate analysis (MVA) methods approaches for the identification of Leptons in the Belle II experiment. First, I will compare two machine learning techniques in terms of classification and performance, then I will focus on Lepton Flavour universality (LFU) violation in tau decays using multivariate analysis methods.
LFU requires that the three charged lepton couple to the W bosons in the same way. This means that any ratio between the couplings g_e, g_μ and g_τ , should equal the unity. By measuring those ratios, one can possibly reject the standard model of particle physics. By performing high precision measurements, which where done by the the Belle II detector, located at the SuperKEKB accelerator in Tsukuba, Japan, one can determine the couplings. Multivariate analysis methods help in the Signal/Background classification which leads to a bigger phase space and ultimately to a more precise determination of the coupling ratios.
The investigation of LFU violation is done in $\tau \to \ell\bar{\nu}\nu$decays using 3x1 prong τ -pair events at Belle II. The algorithm is trained on MC14 data and the purity and efficiency metric will then be compared to a similar analysis which uses a cut based approach [1]

Contents

1 Introduction

The Standard model of particle physics (SM) is thought to be the most complete description of elementary processes we have so far. It can explain three of the four fundamental forces, it correctly predicted the Higgs Boson which was found 2012 (at the LHC in Cern) and yet there are many questions that can't be answered by the SM, like the matter-antimatter asymmetry of the universe or the accelerated expansion of the universe due to dark matter. It is therefore believed, that there is an even more fundamental theory or at least an extension to the SM. By experimental disproving the SM predictions like a unify coupling of leptons, we can test theories that goes beyond the SM (BSM). To do so, it is extremely important to have exact measurements of the underlining parameters.
This thesis will focus on a software based improvement of the measurements done by the Belle II collaboration, using MVA methods. The underlining problem for different algorithms will be the same: Distinguish between Signal and Background events. This is a very common binary classification problem, therefore a lot of libraries exist. The analysis was exclusively done in python using the **sklearn** and the BASF2 software packages. Different parts of the work focus on different algorithms, however all of them are based on boosted decision trees since it's a promising approach for those kind of problems.
This thesis is organised in 4 sections: a theoretical overview, the experimental setup, discussion of the work and the discussion of the results.

2 | Theoretical overview

2.0.1 The Standard Model of particle physics

The Standard Model of particle physics is the central theory behind all particle experiments. It's most fundamental elements are the existence of Leptons and Quarks which together are also called matter particles. Although the SM is well proven, some physical phenomena can not be explained by it. For example, even though gravitation can be described by Einsteins general relativity, the SM can't explain the mediation through particles. A boson, the graviton is claimed to exist, yet it still remains uncertain if this hypothetical particle even exists since the SM makes no prediction about it. Another example would be dark matter, which can't be explained through any fundamental theory but still it's existence is not doubted. For those reasons scientists all over the world search for physics BSM or at least for SM extensions.

The construction of the standard model

The SM is a theory of interacting fields, treating the boson fields as classical and the fermion fields as completely anticommuting. The construction of the SM relies on symmetries (based on Noether's theorem) which can be expressed through groups. By requirement, the equations of the model retain the same form after certain transformations. This requirement seems logical in a sense as that the equations should be consist in every inertial frame of reference, so that they are covariant under Lorentz transformation which also is a symmetry transformation [2].
The group structure of the SM is as follows:

$$\underbrace{U(1)}_{\text{em force: 1 boson}} \times \underbrace{SU(2)_L}_{\text{weak force: 3 bosons}} \times \underbrace{SU(3)}_{\text{strong force: 8 bosons}}$$

This Structure arises from the demand of a local gauge invariance of the underlining Lagrangian density. The choice of the groups seem arbitrary and Quantum field theories with many other gauge groups are consistent and valid theories, they just don't describe the data observed in colliders as good as those groups. U(1) gauge invariance of the Dirac Lagrangian together with a electromagnetic (em) field results in the appearance of a gauge field which can be identified with the field of the em Boson,

e.q as the Photon (γ). If we now demand invariance under $U(1) \times SU(2)_L$ Symmetry, the Lagrangian of interests results in 4 fields, one has already been identified with the em field, the other 3 originate from the generator of the group (Pauli Matrices) and can be mapped to the Bosons of the weak interaction ($W^{\pm}, Z$). By inspecting the chiral structure of the weak force, one find that just left handed chiral states couple to the $W^{\pm}$ Boson which is expressed through the L in $SU(2)_L$. Finally, in the same way the $SU(2)_L$ group was introduced, one can show invariance of the underlining Lagrangian under SU(3), resulting in a triplet state (red,green,blue) where each element represents a four-component Dirac spinor. The group is generated by 8 Matrices (Gell-Mann Matrices) and thus results in 8 gluon fields (colour octet).

Fermi's interaction

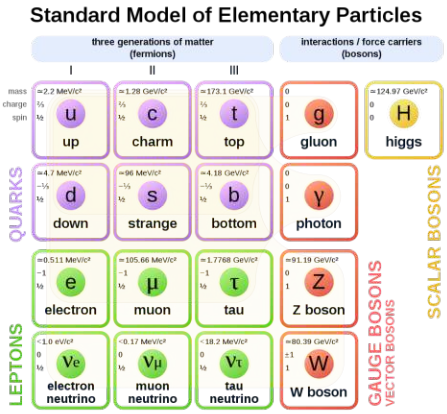
The weak interaction is transited through the W-Boson which has a short lifetime and therefore a short range due to its big mass of around 80 GeV. This fact can be explaind through the Heisenberg uncertainty principal and Einsteins energy-mass equivalent, since $\Delta t \approx \frac{\hbar}{\Delta E}$ which corresponds to a distance of $\Delta s \approx 10^{-18} m$. Together with the coupling constant g_W, one can write down the strength of the weak interaction, also known as Fermis constant.

$$\frac{G_F}{\sqrt{2}} = \frac{g_W^2}{8 m_W^2} \tag{2.1}$$

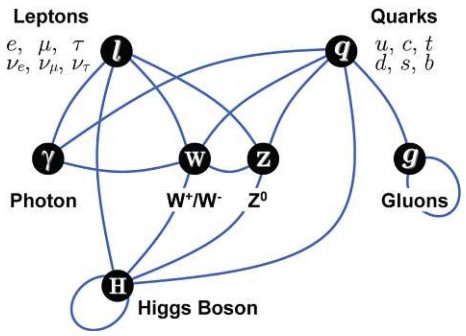
where G_F denotes the Fermi constant, and m_W the mass of the W Boson. Through experimental measurements with muons, the value of G_F (the Fermi constant) could be calculated to $G_F = 1.16638 \times 10^{-5} GeV^{-2}$, and since the mass of the W Boson is also known, one can easily calculate g_w.
For low energies, the propagator of the quantum electro dynamic (QED) can be expressed through $P_{QED} \sim \frac{1}{q^2}$ where q represents the momentum transfer, while the one for the weak interaction results in $P_W \sim -\frac{1}{m_W^2}$. Thus, for low energies ,the probability of observing QED processes are around the factor $\frac{q^4}{m_W^4}$ more likely. For high energies however, m_W is negligible which leads to similar couplings of QED- and weak interaction- processes. Therefore, one should use high energies like in particle colliders to investigate weak interaction physics.

Leptons

Leptons, together with Quarks and Bosons are the most fundamental particles we know so far. They have half integer spin, masses between 0.51 MeV and 1776.86 MeV and do not interact with the strong force. Leptons are divided in 2 classes and 3 generations (flavours), charged (electron like Leptons) and not charged (neutrinos) Leptons and generation one to three. A complete graphic of the SM particles as well as their interactions can be found in Figure 1.



(a) SM particles and their properties [3].



(b) SM Interactions [4].

Figure 1: The SM particles and how they interact

Of the charged leptons, only the electrons are stable. Tauons and muons differ from the electron only in their masses and their finite lifetime, all Leptons should therefore behave equally besides effects due to their difference in mass. The study of lepton decays show, that leptons not only preserve electric charge but also can just change in another of the same type. Thus the conservation of lepton number is apparent in the SM [2].

2.0.2 Lepton flavour universality violation

As seen in equation 2.1, all Leptons couple to the W Boson in the same way. This prediction of the SM is called Lepton Flavour universality, since all Leptons are threatened to be universal in their interaction with the weak force. This means that any ratio between the couplings g_e,g_μ,g_τ should be equal to one. To test if this statement really holds, one can determine the branching fractions, which is a measurement for the fraction of particles decaying in a certain final state, and then compare these to the ration of the $g_{e,\mu,\tau}$.

$$R_\mu = \frac{BF[\tau^- \to \mu^- \bar{\nu_\mu} \nu_\tau]}{BF[\tau^- \to e^- \bar{\nu_e} \nu_\tau]} \; R_{h=\pi,K} = \frac{BF[\tau \to h \nu_\tau]}{BF[h \to \mu \nu_\mu]} \; R_e = \frac{BF[\tau^- \to e^- \bar{\nu_e}] \nu_\tau}{BF[\mu^- \to e^- \bar{\nu_e} \nu_\mu]} \tag{2.2}$$

In this work, the ratio R_μ will be from interest since it can be compared to the current world-leading measurement, done by the BABAR collaboration [5] With R_μ, one can calculate the ratio of the constants $g_{\mu,e}$:

$$\left(\frac{g_\mu}{g_e}\right) = \sqrt{R_\mu \frac{f(\frac{m_e^2}{m_\tau^2})}{f(\frac{m_\mu^2}{m_\tau^2})}} \tag{2.3}$$

where $f(x) = -8x + 8x^3 - x^4 - 12x^2 \log(x)$ assuming mass-less neutrinos [6].
In the BELLE 2 experiment, we created τ particles via the process $e^+e^- \rightarrow \tau^+\tau^-$ to then examine the relevant decay of $\tau \rightarrow \ell^{\pm}\nu_{\ell}\bar{\nu_{\tau}} \quad \ell = e, \mu$, while the other tauon decays into three charged hadrons (typically pions) $\tau \rightarrow 3\pi^{\pm}(n\pi^0)\nu$. A Feynman diagram of the underlining process can be found in Figure 2.

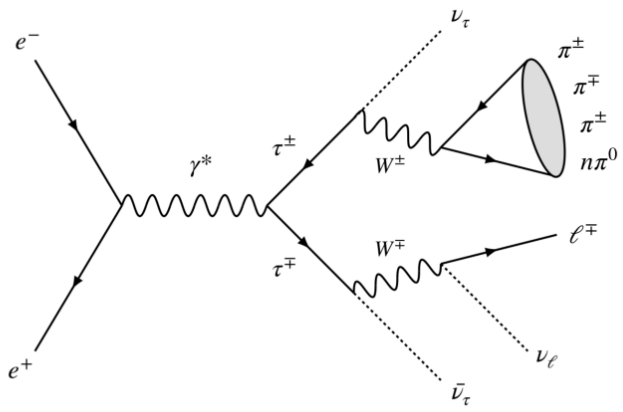


Figure 2: Underlining Feynman diagram for the $e^+e^- \rightarrow \tau^+\tau^-$ process [6].

The current interest in investigating LFU violation arrives from promising measurements, done by the BABAR collaboration [5] and the LHCb collaboration [7] as well as the BELLE II collaboration [8]. Especially the measurement of 3.1 standard deviations (σ) by the LHCb collaboration are a strong hint for new physics. Figure 3 compares the measurement accuracy of the mentioned collaborations in terms of the R ratio,which is the ratio of the hadronic cross section (here the Kaon) to the muon cross section in electron–positron collisions.

Possible new physics explanations for LFU violation

Since the latest investigation of LFU violation brought promising results, it is possible that the five σ criterion can be reached in near future, thus causing need to explain those phenomenons. The most reasonable explanation is, that there is an additional contribution to the decay rate, due to the exchange of a new virtual particle. This hypothetical particle would have to have a coupling strength around unity, so that it is considerable strong but yet small enough to keep all calculations reliable. Many possible candidates for this particle are considered, nevertheless the existence of leptoquarks seems to be the most promising one.
A leptoquark is a theoretical particle that carry both lepton and baryon numbers and is supposed to couple to leptons as well as to quarks. A big difference to a normal particle is the spin number which can vary between 1 and 0. Since they have not been discovered yet, it's assumed that Leptoquarks have high masses (around 1TeV). If for instance a beauty quark decays into a strange quark with production of an lepton (e, μ), this process would occur under the exchange of a leptoquark instead of a electroweak

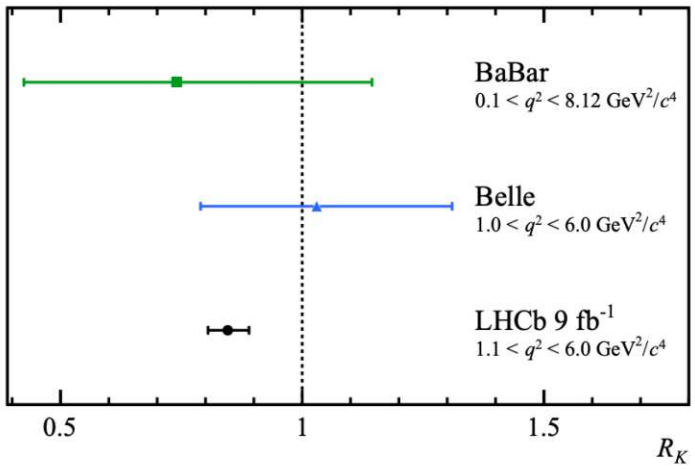


Figure 3: Comparison of the world leading precision measurements of the R ratio [7]. With the smallest error, LHCb currently has the most accurate measurements of the R_K ratio.

or $W^{\pm}, Z^0$ Boson as predicted by the SM. Depending on the lepton, the leptoquark could have different coupling strength and thus resulting in different branching fraction ratios for each lepton.

2.0.3 Multivariate analysis methods

In this thesis, many different machine learning algorithms had been used, most of them are based on gradient boosted decision trees. Gradient boosted decision trees are commonly used in high energy physics because of their tremendous success in the particle identification, but it's also commonly used in other classification problems.
A decision tree represents an ordered structure of decisions based on some predefined rules where nodes represents features(attributes), branch represents decision(rule) and leaf nodes represents outcomes(discrete and continuous). An example tree can be found in Figure 4.

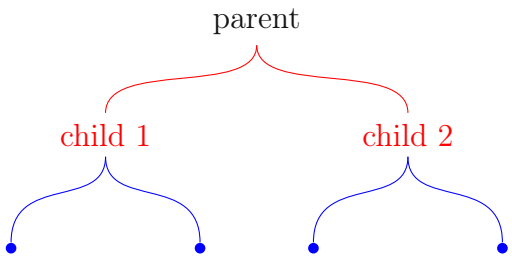


Figure 4: Simple example tree. The first level node is called parent or root node, the second level nodes (child) node, third level (grandchild) node and so on. The last node is usually called leaves.

Usually, the decision is based on the Gini index, which is a measurement of the impurity of the sample and it is defined as

$$i(m) = \sum_{c \neq o} p(o|m)p(c|m) \tag{2.4}$$

where:
$p(c|m)$ is the probability of obtaining decision class c at node m and
$p(o|m)$ is the probability of obtaining decision class o at node m,
or on the entropy of the system, which defined as

$$S = \sum_{j=1} = -\frac{|X_j|}{|X|} * \log_2(\frac{|X_j|}{|X|}) \tag{2.5}$$

where $\frac{|X_j|}{|X|} = p_j$ is the fraction of a single value of set X of the data set and X which can be identified with the frequentative probability of an element[9].
For a multidimensional Dataset X with $x_1, x_2, ...x_n \in X$ features, one can define a target variable y, for instance $y = x_j$. A basic decision tree algorithm works as follows:

1. Initialise the model with a first guess γ, often the mean value of the target vector variables is used. Further, we define a loss function $\mathcal{L}$, often the mean squared error (MSE) is used, as $\mathcal{L}_{MSE} = \frac{1}{n}(y_i - F(x_i))$ where $F(x)$ is a function that gives the predicted value. The goal is to minimize $\hat{F}(x) = arg\,min[\mathcal{L}(y, F(x))]$, so the initial guess results in $F_0(x) = arg\,min[\sum_{i=1} \mathcal{L}(y_i, \gamma)]$.

2. Calculate the pseudo residuals r_{im} of the i-th data point of the m-th tree . Since we want to minimise the loss function, we differentiate $\mathcal{L}$: $r_{im} = -\left[\frac{\partial \mathcal{L}(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$. For the first tree, we can use the constant value $F_0(x)$

3. Fit tree to the pseudo residuals by comparing the Gini impurity coefficient of all data features. Since the r_{im} end up in the leaves (denoted with R_{jm} where j is the leave)of the decision tree, in most cases there will be less r_{im} than x_i (because most trees have a pre defined number of leaves, generally 8-32), so the average of the residuals resulting in the same leave is used. This is only possible if we use the MSE as a loss function.

4. Make a new prediction: $F_m(x) = F_{m-1}(x) + \alpha \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$ where γ_{jm} denotes the output value (the mean of R_{jm}) and J_m the number of leaves. The value α must be between 0 and 1 and is called the learning rate. A small α reduces the effect each tree has on the final prediction. The advantage of a small learning rate is a higher accuracy but with the cost of higher computational resources since the steps towards the minimum are smaller.

5. Use the updated function F(x) to make a new prediction.

6. go back to 2 until the maximum iteration is reached.

A pseudo code can be found in Sketch 1.

Algorithm 1 Gradient boosted decision tree using MSE loss function

1: $F_0 = mean(y)$
2: $\alpha = const \in [0, 1]$
3: **for** m = 1 to M **do**
4: **for** i = 0 to N **do**
5: $r_{im} = -\left[\frac{\partial \mathcal{L}(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$
6: $\gamma_{jm} = mean(R_{jm})$
7: $F_m(x) = F_{m-1}(x) + \alpha \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$

2.1 Metrics

To test the performance of any algorithm, it is important to define some metrics and terminology. In this thesis, the metrics "purity", "efficiency" are used and the receiver operating characteristic will be taken to visualise the performance of the algorithm. The purity in particle physics is defined as:

$$P = \frac{N^{sig}_{selected}}{N^{total}_{selected}} \quad \epsilon = \frac{N^{sig}_{selected}}{N^{signal}_{generated}} \tag{2.6}$$

where $N^{sig}_{selected}$ is the number of selected signal events, $N^{total}_{selected}$ is the total number of selected signal and background events and $N^{signal}_{generated}$ is the number of generated signal events. While this terminology is very specific to particle physics, it is very common to use the metrics precision and recall as synonyms for purity and efficiency. Purity can be understood as the ability of the classifier not to label as positive a sample that is negative and is depended on the a priori probability for an event to be signal or background. In contras, efficiency can be interpreted as the ability of the classifier to find all the positive samples.
The ROC curve is a very common used plot to visualise the diagnostic ability of a binary classifier system. It plots the true positive rate, which is defined as $TPR = \frac{True\ Positive}{True\ Positive + False\ Negative}$, over various thresholds against one minus the False positive rate $FPR = \frac{False\ Positive}{False\ Positive + True\ Negative}$. The ROC can therefore be understood as a visual representation of all confusion matrices for a varying threshold, where the diagonal of the graph shows where the TPR is equal one minus the FPR. An example ROC Curve is shown in Figure 5.

2.2 Algorithm optimisation

The most common machine learning algorithms need some parameters (hyperparameter) like learning rate etc. as input. Because those parameters have a big effect on the performance, it is crucial to find the ideal ones.
A very common but computational expensive technique to find the optimal hyperpa-

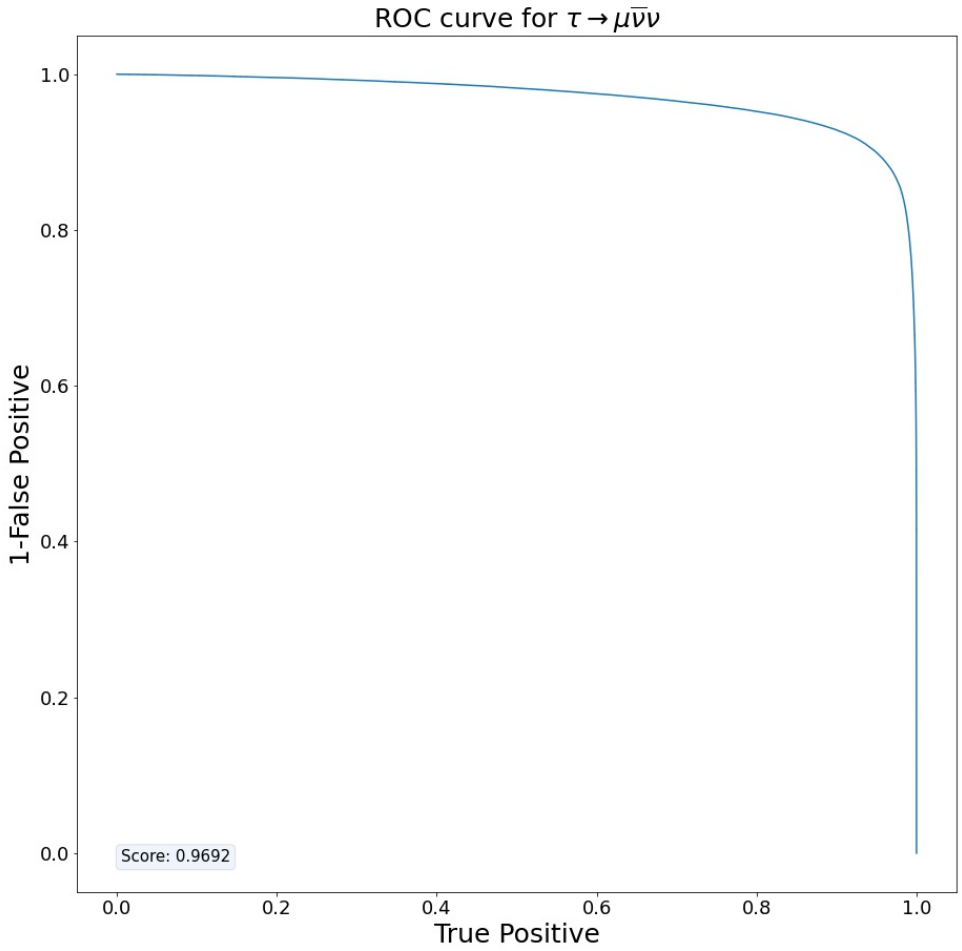


Figure 5: ROC curve of a BDT classifier with an area under the Curve (AUC) of 0.9692.

rameter is the so called grid search. It takes a predefined set of values an performs and exhausting search, to then compare certain metrics (usually AUC). For example, to optimise the hyperparamethers of an BDT, we can define two subsets
$C \in \{2, 8, 16, 32\}$
$\gamma \in [0.1, 0.9]$
where the subset C represents the possible maximum depths of the decision tree and γ the learning rate. By trying every value of every set, we get $4 \times 9 = 36$ possible combinations and therefore 36 metrics that can be compared. The hyperparameter with the highest outcome will then be taken. A illustration of the search is given in Figure 6.

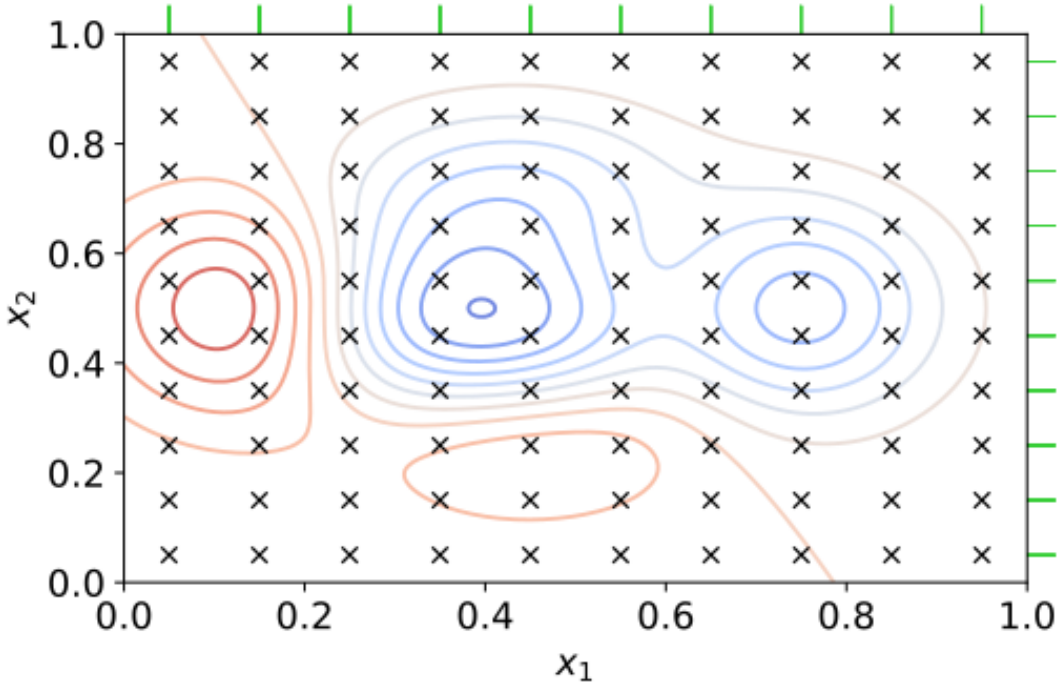


Figure 6: By performing a grid search, we take the Cartesian product of x_1 and x_2. The blue line displays parameters that lead to a favourable outcome, red an unfavourable one [10].

Once the model is trained, it naturally occurs that one needs to know which variable is the most important in the training. To calculate the feature importance, we use a technique called permutation feature importance. The theory behind it is straight forward. The data of a feature vector gets randomly shuffled, the model gets trained and the increase in the model's prediction error can be compared. If the error is big, the feature is said to be important since the model is relies on that feature. Otherwise, if the error is low, the feature is said to be unimportant because shuffling is hardly relevant.

3 Experimental setup

The Belle II experiment is an international collaboration hosted by KEK in Tsukuba, Japan which studies the properties of B mesons. Belle II is the successor to the Belle experiment (which ran from 1999 to 2010), it started taking data in early 2018. Unlike other particle experiments like ATLAS or CMS, located at the LHC in Cern, Belle II is not a high energy experiment, but more a high precision experiment, meaning that physics beyond the standard model is not investigated by directly creating new particles, but by an exact surveying of rare decays from B-mesons. To do so, an accelerator (SuperKEKB) is needed to accelerate particles at energies of 10.58 GeV (COM frame), which corresponds to the mass of the $\Upsilon(4S)$ resonance.

3.1 Luminosity and cross section

Both, Luminosity and cross section are probably the most frequently used parameters in any particle accelerator experiment, therefore they will be shortly introduced. Instantaneous Luminosity can be defined as

$$\mathcal{L}(t) = \frac{1}{\sigma}\frac{dN}{dt} \quad [cm^{-2}s^{-1}] \tag{3.1}$$

where σ denotes the cross section of the experiment and N the number of detected events. By integrating equation 3.1 over the time, we obtain the integrated Luminosity

$$\mathcal{L}_{int} = \int \mathcal{L}(t)dt \quad [cm^{-2}] \tag{3.2}$$

which reflects the size of the data sample ($N = \mathcal{L}_{int} * \sigma$). The dimension of the integrated Luminosity can also be expressed through inverse barn, where $1b \equiv 10^{-24}cm^2$. From a known cross section, one can calculate the number of produced events or, vice versa, measure the cross section from a determined number of produced events. The precise measurement of integrated luminosity is thus fundamental to estimating experimental yields accurately and testing theoretical models precisely [11].
On the other hand, the cross section is a measurement for the probability that particle interact in a certain way. The idea behind that name is, that particles can be imagined as tiny point like balls. If these balls are shot against each other, the probability that they will collide is proportional to the size of the area within a reaction will take place. Therefore the dimension of the cross section is measured in units of area.

3.2 SuperKEB

The SuperKEKB particle collider, located at the High Energy Accelerator Research Organisation (KEK) is an electron positron collider which collides particles at a centre-of-momentum energy close to the mass of the $\Upsilon(4S)$ resonance ($\sqrt{s} = 10.58\text{GeV}$) making it a second-generation B-factory. It is an upgrade to the KEKB accelerator, providing approximately 40 times higher Luminosity which yields the highest luminosity recorded at a particle collider world wide. The target luminosity of SuperKEKB is $8 \times 10^{35} cm^{-2} s^{-1}$, a goal that comes from physics requirements [12]. In 2020, the goal was almost reached with an instantaneous luminosity of $2.4 \times 10^{35} cm^{-2} s^{-1}$
SuperKEKB has an asymmetric beam energy, meaning that the electron beam energy (up to 7GeV), also called HER (high energy ring), differs from the positron beam energy (up to 4GeV), also called LER (low energy ring). As a consequence, the created particles are thrown in the forward direction what brings the advantage of a more precise measurement since the particle location can be better estimated. By shifting the energy in the HER and/or LER, one can create different exited states of the Υ meson. Figure 7 shows the different center of mass energies for different beam energies.

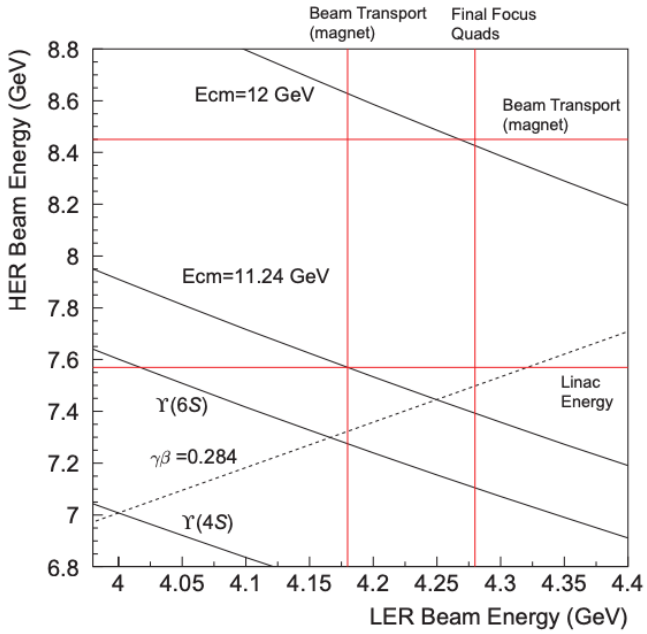


Figure 7: Beam energies to achieve the center of mass energies $\Upsilon(4s)$,$\Upsilon(6s)$, 11.24 GeV, 12 GeV [12].

The SuperKEKB accelerator consists of two 3 km long circumference. Low momentum electrons and their antiparticles are produced and injected in the acceleration ring, where they gain momentum until they nearly reach the speed of light. Once they are fast enough, magnets change their direction and the particles collide, creating new

particles that can be measured through a detector. A sketch of SuperKEKB and the Belle II detector can be found in Figure 8.

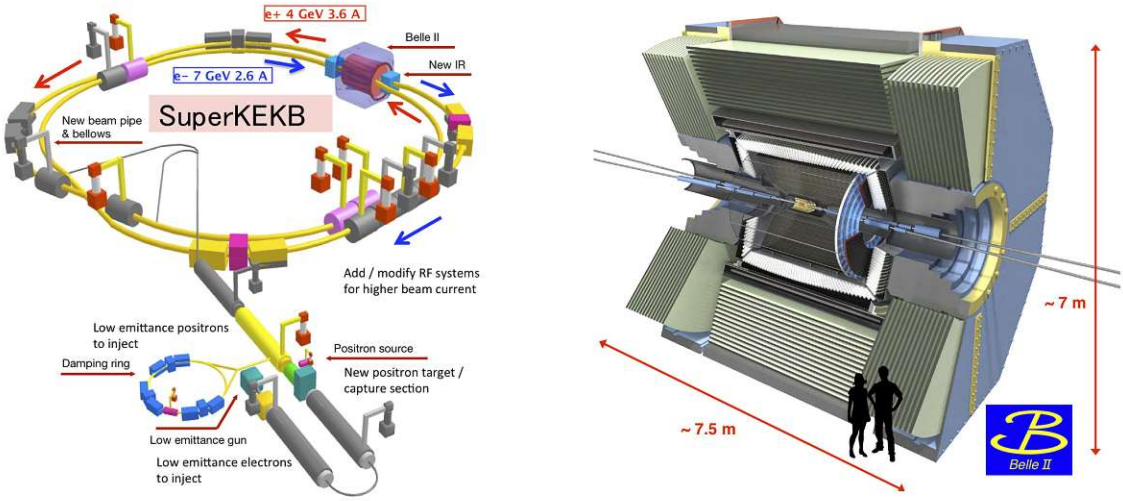


Figure 8: The SuperKEKB accelerator and the Belle II detector [13].

Because the center of mass energy of 10.58 GeV corresponds to the rest mass of the $\Upsilon(4S)$, those particle are favoured in the production. The $\Upsilon(4S)$ state consists of two excited b quarks ($b\bar{b}$) and decays with a probability of 51.4% in B^+B^-:

$$f^{+-} = \frac{\Gamma(\Upsilon(4S) \to B^+B^-)}{\Gamma_{Tot}(\Upsilon(4S))} = 0.514 \pm 0.006 \tag{3.3}$$

and with 48.6% in $B^0\overline{B^0}$:

$$f^{00} = \frac{\Gamma(\Upsilon(4S) \to B^0\overline{B^0})}{\Gamma_{Tot}(\Upsilon(4S))} = 0.484 \pm 0.006 \tag{3.4}$$

which makes it to the mentioned B factory [14].

The whole fraction of created particles can be found in Figure 10 as well as in Table 3.1. A sketch of the B meson production is represented through Figure 9

The reason Belle II uses B mesons for it's research can be explained by the fact that the Belle collaboration was originally establish to examine CP violation and to measure precisely the CKM (Cabbibo-Kobayashi-Maskawa) matrix elements, which determine quark flavour mixing. The amplitude for the direct decay $\overline{B^0} \to J/K_s^0$ is proportional to the $|V_{cb}|$ CKM matrix element. This decay is also possible through $B^0 \to \overline{B^0} \to J/K_s^0$, see Figure 11.
By producing $B^0\overline{B^0}$ pairs via the process $\Upsilon(4S) \to B^0\overline{B^0}$, one can measure the meson decay. Since most B^0 produce flavour specific final states, we can use the final-state particles to determine whether the decaying meson was a B^0 or a $\overline{B^0}$ meson. For example, measuring a K^+ particle in the final state is a strong hint for a B^0 decay because

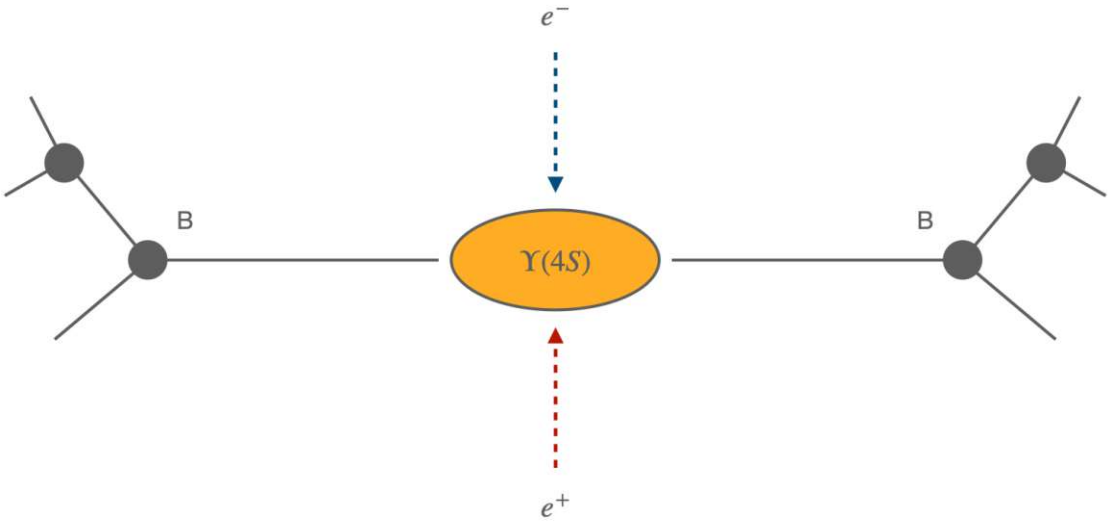


Figure 9: $\Upsilon(4S)$ mesons are produced by the e^+e^- collision with a centre of mass energy of 10.58 GeV, resulting in a B meson decay. Other particles are created through further decay of the B mesons.

Table 3.1: Most dominant processes at Belle II with a center of mass energy of 10.58 GeV [15]

Process	cross-section [nb]
$\Upsilon(4S)$	1.11
$u\bar{u}(\gamma)$	1.61
$d\bar{d}(\gamma)$	0.4
$s\bar{s}(\gamma)$	0.38
$c\bar{c}(\gamma)$	1.3
$\tau^+\tau^-(\gamma)$	0.919
$\mu^+\mu^-(\gamma)$	1.148
$e^+e^-(\gamma)$	$\sim$ 300

the most probable process for this to happen is $B^0 \to \overline{D} \to K^+$. Decays like that are called "flavor-tag" decays. If a B-meson now decays at the time t_1, it's partner meson has to have opposite flavour and propagates in time, meanwhile it's quark flavour content can oscillate from an unmixed state into a mixed one, until it decays (at time t_2). If this product is in a CP eigenstate (like J/K_s^0), the unmixed and mixed flavour components of the mesons interfere, resulting in different decay rates for B^0, $\overline{B^0}$. If the mesons decay in CP eigenstates before the flavour tagging, the phase from the mixing has opposite sign. Thus, events that are B^0 tagged are interfering destructive if $\Delta t = t_2 - t_1$ is negative, and constructive if Δt is positive, while it behaves vice versa for the $\overline{B^0}$ meson[16]. Measuring the live time and decay rates of B mesons therefore allow to investigate CP violation.

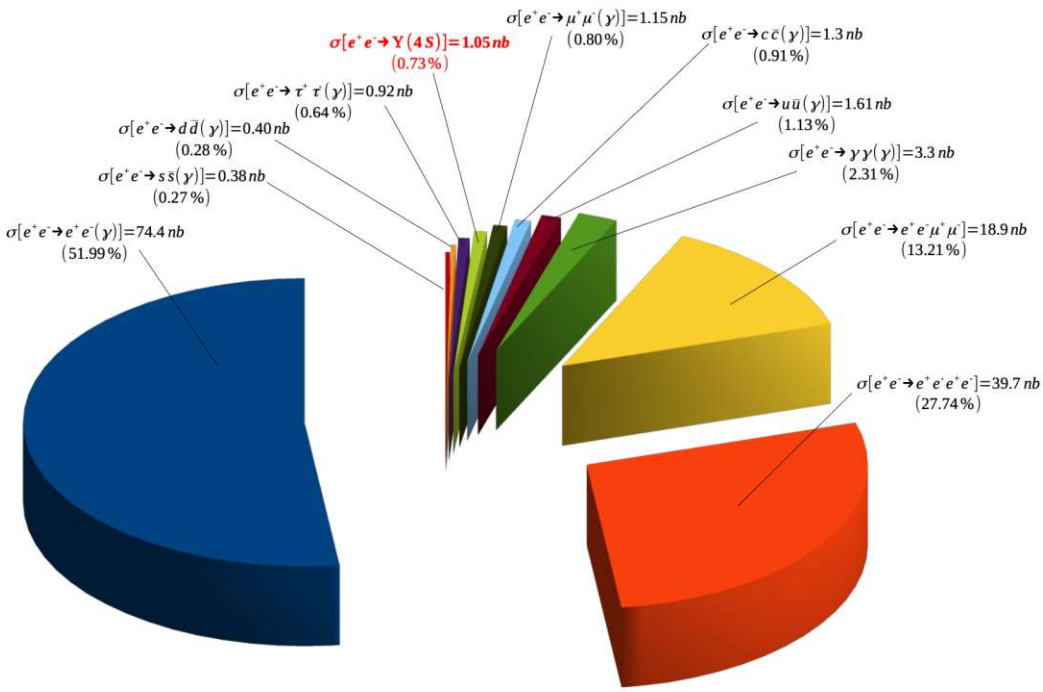
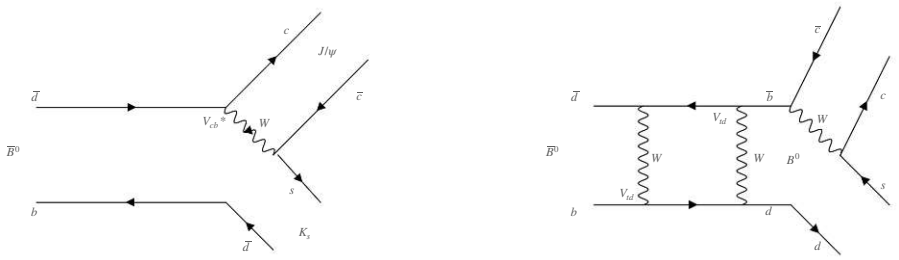


Figure 10: Particle production at the SuperKEKB. Only a small fraction of produced particles are actually $\Upsilon(4S)$.

Besides the production of B-mesons, SuperKEKB is intensively used for tau pair production. While both cross section are very similar (1.05 nb compared to 0.92 nb of the tau pair production), the number of resulting particles is almost the same. A tau particle decays either hadronically with the emission of one neutrino, or leptonically with the emission of two neutrinos, thus resulting in a larg amount of missing energy. The advantage of e^+e^- colliders in the measurement of tau particles results from the precise measurements of missing energy. By colliding elementary particles, we know precisely their momenta and energies while when colliding protons which consist of quarks and gluons, we can not exactly determine which elementary particle interacted and if it was carrying most of the energy or not. This yields to a much more uncertain reconstruction of the collision energy and consequently to a much more uncertain reconstruction of the missing energy.

3.2.1 Nano beam optics

The 40 times higher target luminosity in SuperKEKB could only be realised through nano beam optics, a technique that was first used by P. Raimondi for the Italian super B factory. This technique squeezes the vertical beta function at the interaction point by minimizing the longitudinal size of the overlap region of the two beams at the interaction point, which generally limits the effective minimum value of β_y* through the "hourglass effect" [17]. Since the Luminosity is inverse proportional to the beta



(a) $\overline{B^0} \to J/K_s^0$ **(b)** $B^0 \to \overline{B^0} \to J/K_s^0$

Figure 11: The interfering Feynman diagrams.

function, a smaller β_y* leads to a higher luminosity. A schematic overview is given in Figure 12. The nano beam scheme tries to minimize the effective bunch length d, which is determined by the horizontal half crossing angle ϕ. The hourglass condition in the nano-beam scheme then requires that $\beta_y* > d$ what leads to a smaller beta function if d can be made sufficiently small which can be realised by a relatively large horizontal crossing angle together with an extremely small horizontal emittances and horizontal beta functions at the interaction point for both beams.

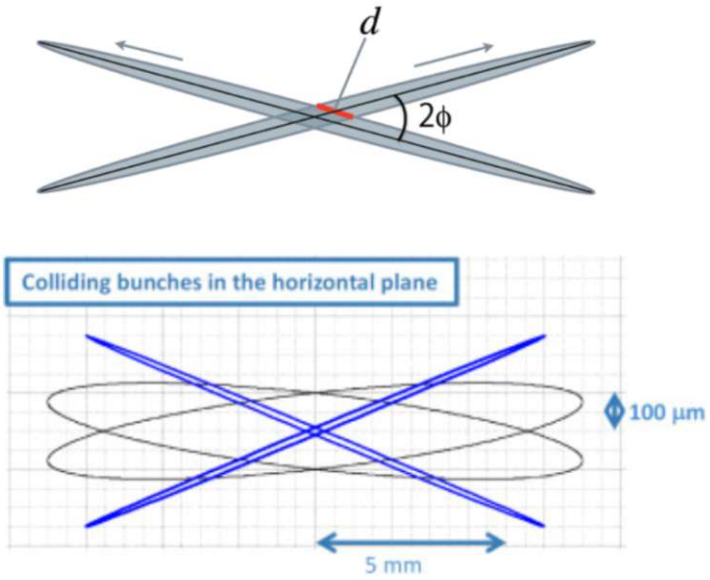


Figure 12: Nano beam collision scheme [18], [17].

3.3 The Belle II Detector

To fully understand the measured data, one must not only have an understanding of how it was produced, but also of how it was captured. Thus, making the understanding of the detector vital for every particle experiment. With a height and length of 10 meters and a weight of 1.500 tons, the Belle II detector consists of 7 cylindrical layers around the beam pipe including a superconducting magnet to bend the particle tracks. An overview of the detector is given in Figure 13.
Please note that the following descriptions follow the Belle 2 technical report [17] and the Belle II Physics Book [15]. For a more detailed overview, please visit the sources.

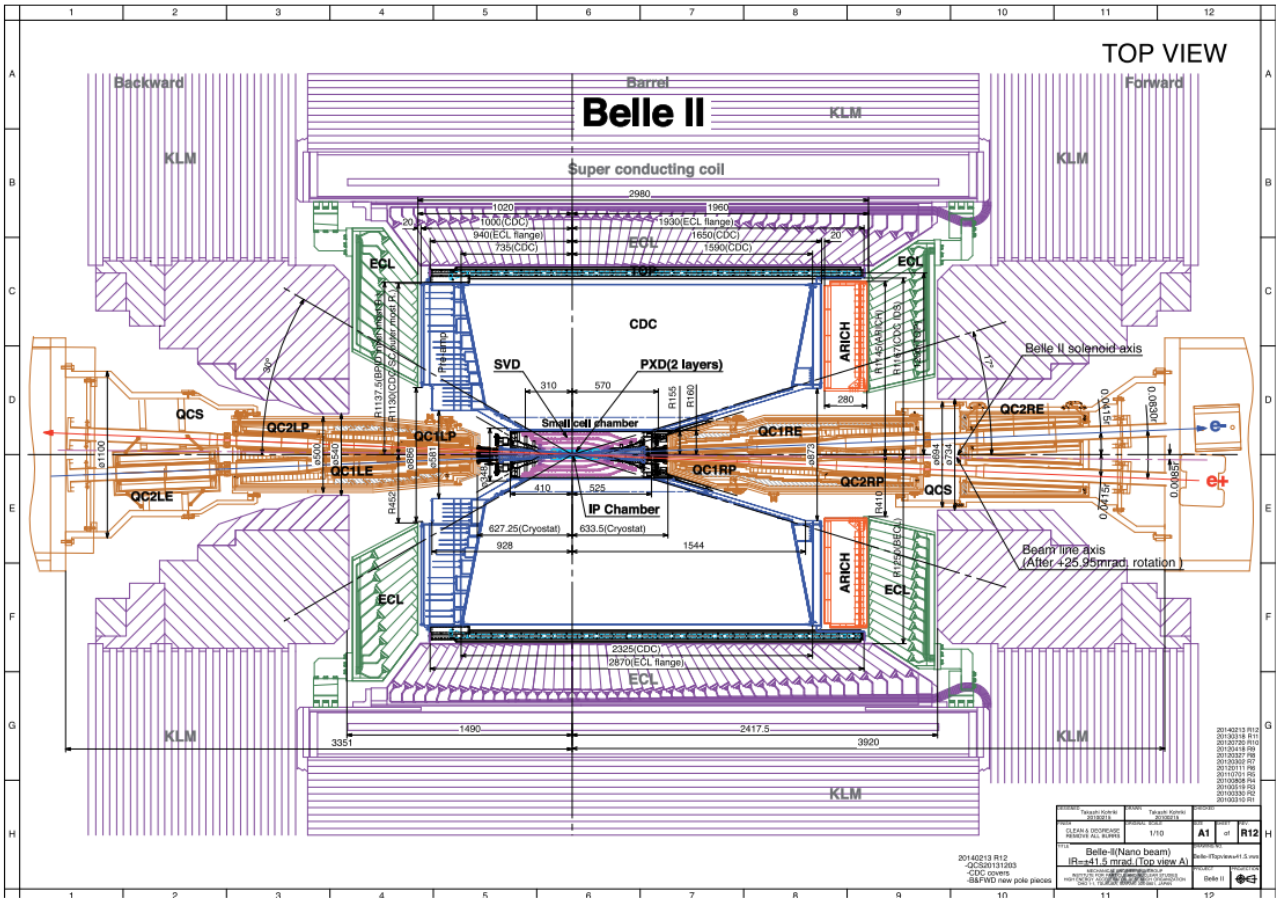


Figure 13: Belle II top view [15].

- **Pixel Detector (PXD):**
  Together with the Silicon Vertex Detector, it is located closest to the particle interaction and is therefore faced with extremely high hit rates, caused by beam-related background which will be further explained in section 3.3.1. As mentioned in section 3.2.1, the SuperKEKB machine uses nano beam optics with a beampipe

radius in the interaction region of only 10 mm. Because the background increases roughly with the inverse square of the radius, the innermost layer can no longer be accomplished through strip detectors due to the large occupancy which is defined as the fraction of channels hit in each triggered event. Thus, to avoid strips, Belle II uses pixel sensors for the innermost layers because they have a much larger number of channels which correspond to much smaller occupancy. Beyond a radius of 40 mm, the large occupancy of stripe detectors have no longer to be taken into account, leading to a pixel-stripe scheme. This technique was also used when building the LHC, however the lower energies of the SuperKEKB require a different semiconducting technology, named DEPFET (DEPleted Field Effect Transistor), which enables us to use very thin sensors in the range of 50 μm.
If a charged particle hits one of the many DEPFET detectors, an electron will be created in the silicium layer which leads to a conductivity increase that can be measured. The track of the particle which bends due to the strong magnetic field can be determined through the huge amount of pixels ($\sim$ 8 million in two layers). A sketch of the PXD can be found in Figure 14.

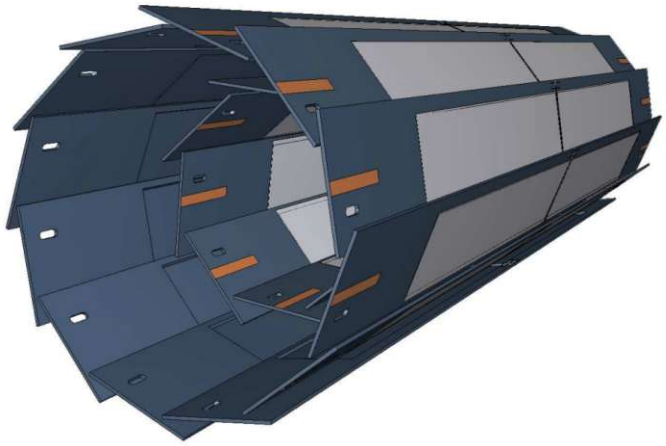


Figure 14: PXD detector overview [17] .

- **Silicon Vertex Detector (SVD):**
  The SVD consists of four layers of double-sided silicon strip detectors (DSSDs) and it covers the full forward boosted polar acceptance angel from 17° to 150°. To do so, it uses slanted sensors in the forward region, resulting in a lantern shape (see Figure 15). Together with the PXD and CDC (Central drift chamber), it's main purpose is to collect data to extrapolate tracks from the Pixel detector to reconstruct in the drift chamber. This strategy is necessary to determine the vertex of a particle. The SVD together with the PXD are also able to reconstruct low momentum particles (down to $\sim$ 10 Me/V) that do not leave enough hits in the CDC.
  In contrast to the PXD, the Silicon Vertex detector uses silicon stripes, rather

than pixels. This brings the advantage of a two dimensional measurement of the particle, rather than a binary signal and since the SVD is 39 mm, 80 mm, 104 mm, 135 mm away from the interaction point, the background dependency of the radius is no longer a concern.

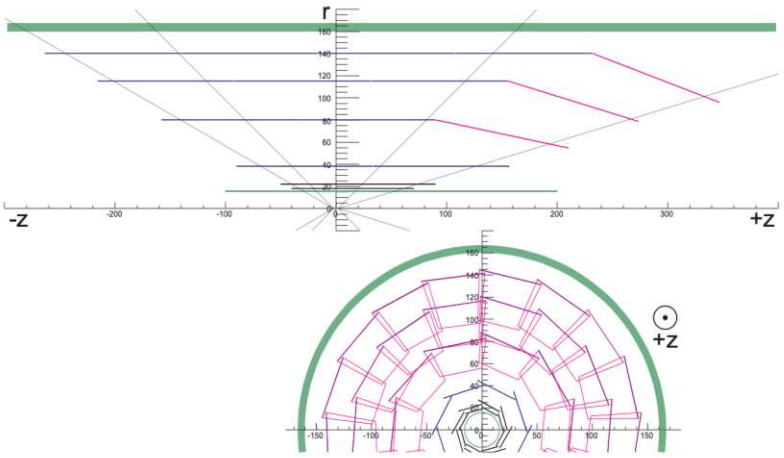


Figure 15: Configuration of the slanted layers of the SVD [17].

- **Central Drift Chamber (CDC):**
  The Central Drift Chamber has three main purposes. It reconstructs charged particle tracks and enables a precise measurement of its momentum. It provides particle identification information by measuring the energy loss within the gas volume and third, it provides efficient and reliable trigger signals for charged particles. An overview of the CDC properties is given in table 3.2, a semantic view in Figure 16.

Table 3.2: Main parameters of the Belle II CDC [17].

Parameter	Proportion
Radius of inner cylinder (mm)	160
Radius of outer cylinder (mm)	1130
Radius of innermost sense wire (mm)	168
Radius of outermost sense wire (mm)	1111.4
Number of layers	56
Number of sense wires	14.336
Gas	$H_e - C_2H_6$
Diameter of sense wire μm	30

The layers are grouped into axial layers (aligned with the magnetic field) or in stereo orientation (skewed to the magnetic field), thus enabling to observe the z-position of the particle which leads together with the PXD and CDC to a three dimensional particle track. If a particle hits the CDC, it ionises the gas.

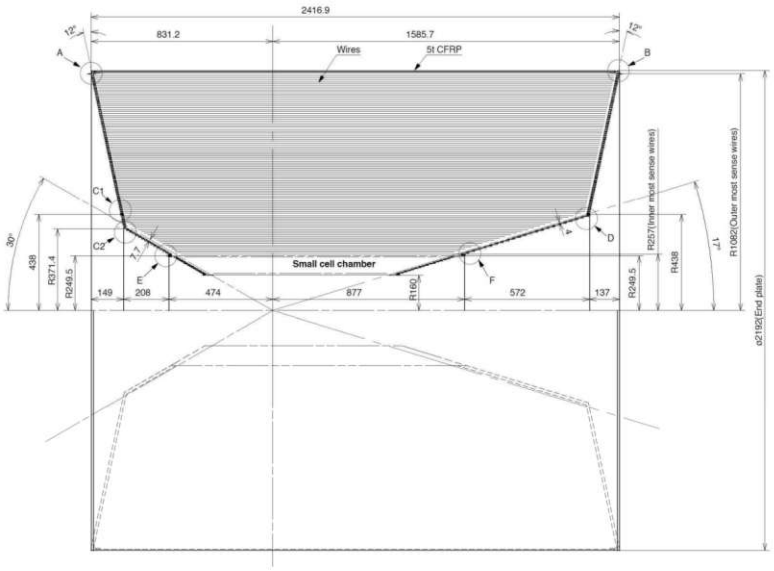


Figure 16: Structural overview of the CDC [17].

Depending on the loss of energy with respect to its track, we can identify the particle. Figure 17 shows the different loss of energy for different particles.

- **Particle Identification - Barrel:**
  The Particle Identification - Barrel is necessary to improve the $K\backslash\pi$ separation. It is split into a time of propagation (TOP) and an Aerogel ring imaging (ARICH) counter region.
  Both techniques rely on Cherenkov radiation. If a charged particle travels trough a dielectric medium of refractive index n and thereby excites the molecules of that medium, photons are emitted if the molecules fall back in the ground state. If that particle travels faster than light in that medium, the photons interfere positive, thus leading to coherent radiation at a fixed angel θ_C in dependency of the particles velocity $\cos(\theta_C) = \frac{vc}{n}$ as shown in Figure 18.
  TOP: The time of propagation counter measures the Cherenkov angel and together with the particle momentum, one can calculate the particle mass which is a definite parameter for particle identification. Charged particles pass through a dielectric medium, radiating Cherenkov photons and then internally reflected towards the end of the bar and registered by a fast position sensitive sensor of single photons. Different particles with the same momentum emit Cherenkov Photons at a different θ_C, leading to a variation of the photon path length and ultimately to a different time of propagation as shown in Figure 19a. In the Figure 19b the arrival time of photons versus the position on the photon detector is plotted for photons from pions and kaons, which enables a clear distinction between the patterns [21].

  ARICH: The aerogel ring imaging counter sits between the central drift chamber

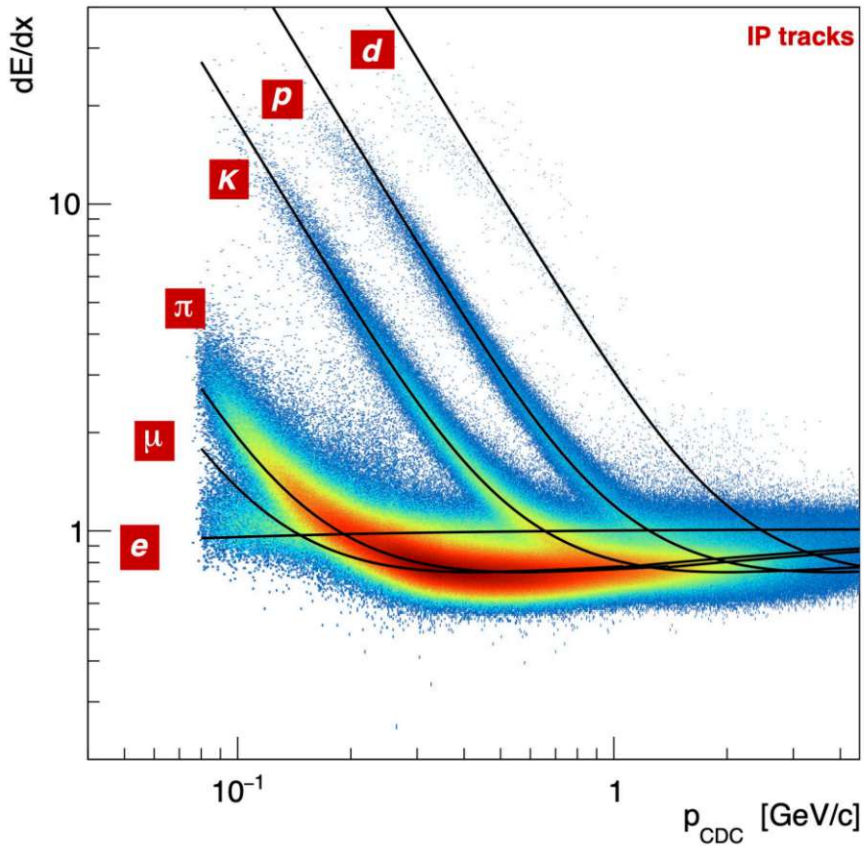


Figure 17: Ionisation energy loss with respect to the momentum of π, K, P, e in the Belle II CDC [19]. Low momentum particles can be better separated due to the big differences in $\frac{dE}{dx}$ while the distribution for high momentum particles overlap.

and the electromagnetic calorimete inside the magnetic field. It consists of two aerogel radiators with different refracting indices ($n_1 = 1.045$, $n_2 = 1.055$) which increases the number of detected photons without the loss of resolution, where Cherenkov photons are emitted if a charged particle hits the device. The ray of light is refracted and forms a cone, depending on θ_C. The size of the ring which corresponds to the number of hits is then detected as shown in Figure 20a and 20b.

- **The Electromagnetic Calorimeter (ECL):**
  As approximately one third of the B-decay products are π^0 particles that provide photons in a wide energy range, a ECL is from great need. It is used for the detection of photons and their parameters like angel and energy, but also for electron and π^0 detection. It consists of a 3 m long barrel section with an inner radius of 1.25 m and covers the polar angle region of $12.4° < \theta < 155.1°$, except for two gaps $\sim 1°$ wide between the barrel and endcaps. It consists of 8736 thalliumdoped caesium iodide CsI(Tl) crystals. Electrons or photons that hit the ECL deposit most of their energy in those crystals, producing new free electrons.

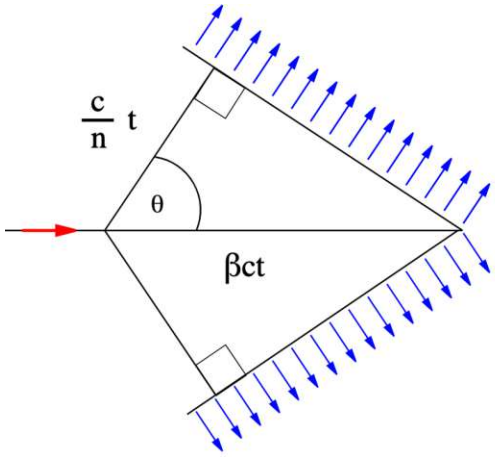
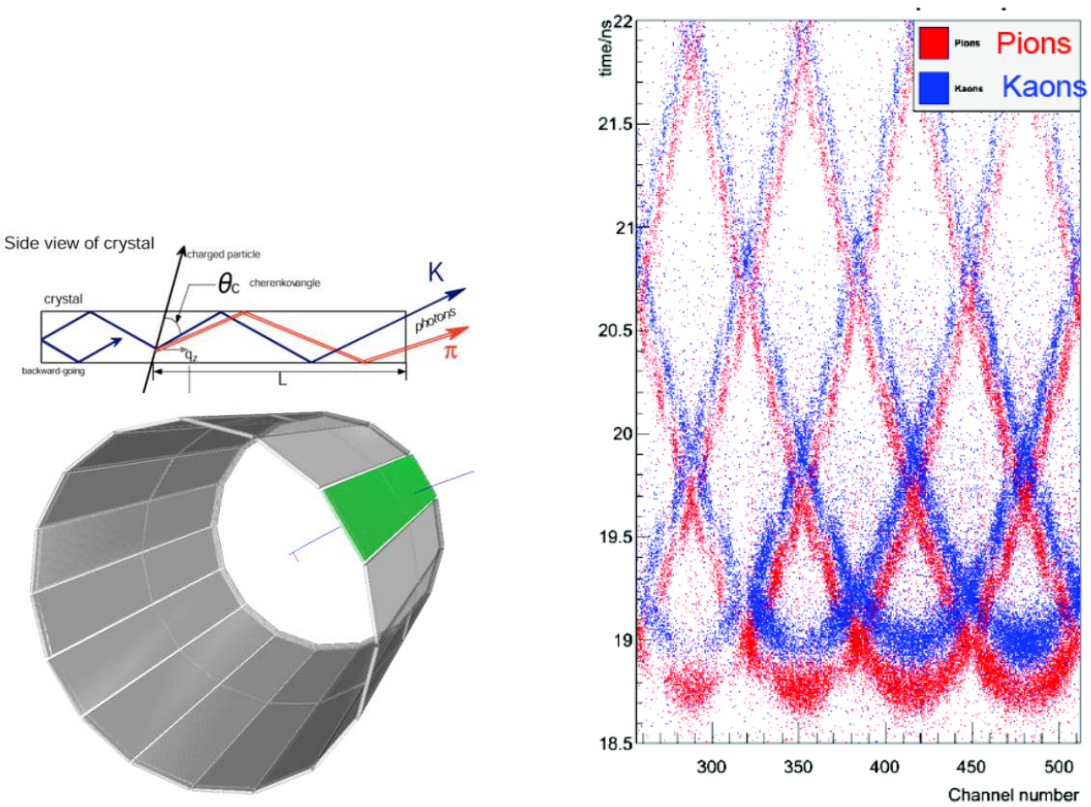


Figure 18: The geometry of Cherenkov radiation[20].

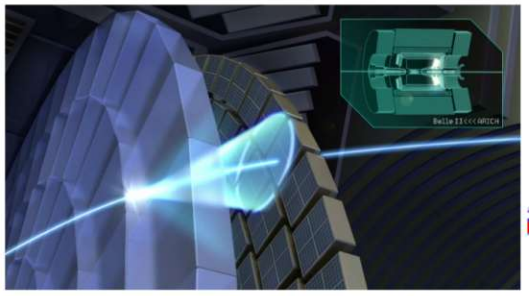


(a) Time of propagation counter setup [21].

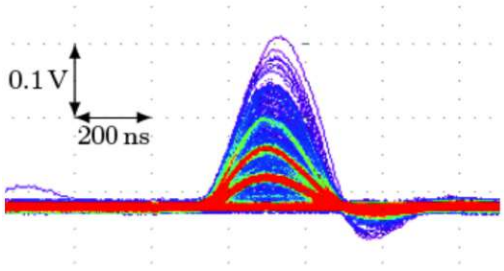
(b) Time of arrival of photons as a function of position[21].

Figure 19: TOP counter.

Because of Bremsstrahlung, photons are radiated producing new electrons and continuing the process until the photon energy is below the threshold necessary to produce further electrons. This cascade of photons and electrons is called



(a) Schematic view of the ARICH detector [21].



(b) Photon counts for different charged particles - clear separation between different number of hits is seen in the analog output signal [21].

Figure 20: ARICH counter.

electromagnetic shower. By analysing the shape of the shower, we can distinguish between electrons and other leptons, since heavier leptons do not fully deposit their energies. By just observing showers without the presence of any charged particle, we can identify photons.

- **Superconducting magnet:**
  The superconducting magnet is placed between the ECL and KLM and provides a 1.5T magnetic field. By observing a bent particle track, together with other particle informations we can calculate its momentum.

- **K_L and μ detector (KLM):**
  The KLM consists of a 14-15 alternating layers of 4.7 cm-thick iron plates and active detector elements located outside the superconducting solenoid around the barrel and at the endcaps (see Figure 13). The iron plates serve the purpose of a flux return produced by the superconducting coil, but they also allow the K_L to shower hadronically which can then be detected by the Resistive Plate Chambers (RPCs) in the barrel region. In the endcaps region, scintillator stripes are used instead of the RPCs, because of the high background rates. Particles that continue from the ECL in the KLM and do not shower hadronically can be identified as muons.

3.3.1 Background sources

When performing high precision measurements it is absolutely crucial to distinguish between fake and real signals. To do so, one has to understand where possible fake signals come from. Thus, limiting the background sources to a minimum and taking the rest of it into account. While there are many sources, I will explain the most common ones in this section.

- **Touschek Effect:**

The Touschek Effect arises due to the Coulomb scattering of two particles in the same beam bunch. The particles scatter, leading to a particle with energy higher than the nominal beam bunch energy, and one with lower energy. Those particles eventually hit the inner surface of the beam pipe, generating showers which can propagate to the detector if close enough to the interaction region, causing damage to silicon devices and contributing to generate fake hits that affect the performance of the detector. The scattering rate can be estimated by

$$R_{Tou} \propto \frac{1}{\sigma} E^3 n_b I^2 \tag{3.5}$$

where σ is the beam size, E the beam energy, n_b the number of bunches and I the beam current [22].
Due to the nano beam scheme, the Touschek background is $\sim$ 20 times higher at SuperKEKB than at KEKB. To mitigate Touschek background, movable collimators and metal shield are used, so that particles that deviate from nominal trajectories and prevent them from reaching Belle II.

- **Beam-gas Scattering:**
  The beam gas scattering occurs due to the residual gas molecules in the beam pipe. There are two possible scattering processes: Coulomb scattering which changes the particle tracks and Bremsstrahlung scattering which decrease the particles energy. We can estimate the rate of the beam-gas scattering with

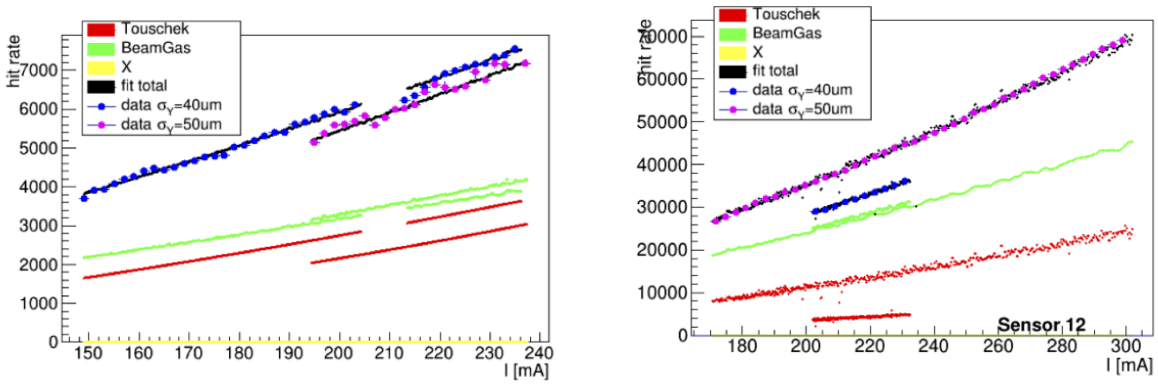
$$R_{bg} \propto IP \tag{3.6}$$

  where P is the residual pressure inside the beam pipe. While the beam current at SuperKEKB is approximately twice as high as in KEKB, the residual pressure is almost the same, leading to a $\sim$ 100 times higher beam-gas scattering. As seen in the Touschek effect, collimators and metal shield can be used to reduce background caused by Coulomb scattering and Bremsstrahlung effects. However, vertical collimators can lead to transverse mode coupling (TMC) instabilities because the vertical beta function is larger than the horizontal one. To avoid this, the collimator width must be narrow enough to avoid beam loss in the detector but also wide enough to avoid TMC instability. This is accomplished by vertical collimators with $\sim$ 2 mm width in locations where the vertical beta function is relatively small. Figure 21 shows the background contribution of the Tuschek effect and the beam gas scattering for the HER (Figure 21a) and the LER (Figure 21b) ring.

- **Synchrotron Radiation (SR):**
  Synchrotron radiation is very similar to Bremsstrahlung and is emitted when particles close to the speed of light change their direction of motion. The rate of emission is proportional to

$$R_{bg} \propto E^2 B^2 \tag{3.7}$$

  where B is the magnetic field intensity. As a consequence, the main contribution to the Synchrotron radiation background will originate from the HER ring, with



(a) Data of beam size study for HER with Touschek and beam-gas components. [22].

(b) Data of beam size study for LER with Touschek and beam-gas components [22].

Figure 21: HER and LER beam size study.

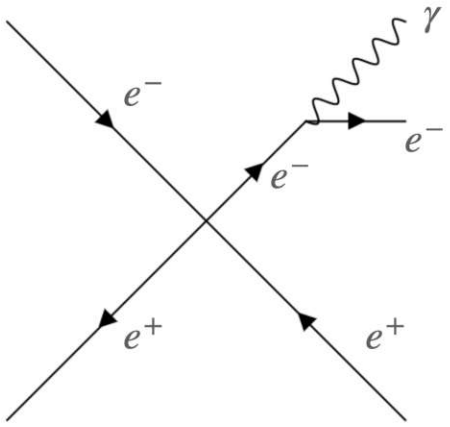
photon energies going from a few keV to tens of keV. To protect the detector and to avoid an additional background source the beam pipe is coated with an Au layer and the interaction region beampipe is shaped in a way to avoid direct SR hits at the detector.

- **Radiative Bhabha process:**
  Radiative Bhabha scattering occurs via the process $e^+e^- \rightarrow e^+e^-$ and can produce energetic photons. The underlining Feynman diagram can be found in Figure 22a. The produced photons can lead to showers in the detector and influence the measurements. Bhabha photons can also propagate and interact with the iron of the magnets, thus leading to a very large production rate of neutrons via the photo-nuclear resonance mechanism which is the largest background of the KLM. We can minimize this mechanism with shielding in the accelerator tunnel.

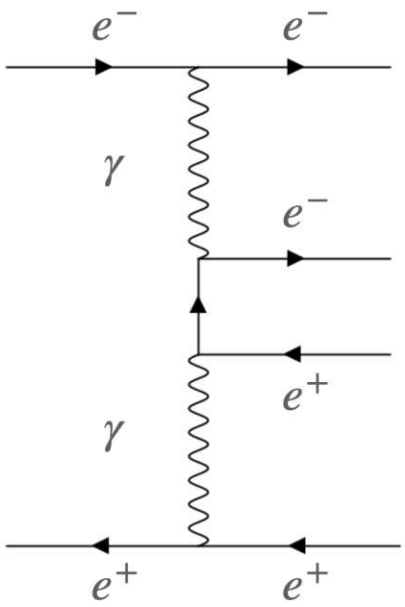
- **Two-photon process:**
  Via the process $e^+e^- \rightarrow e^+e^-e^+e^-$ shown in Figure 22b, low momentum electron-positron pairs can be produced, which then can spiral around the solenoid field lines and leave multiple hits in the inner Belle II detectors [15]. Radiative Bhabha scattering as well as two-photon processes increase with the Luminosity (therefore called luminosity background), making it to a bigger problem in Belle II than its predecessor.

3.3.2 Trigger

Every 8 ns beam bunches cross, creating a huge amount of data. Because of the significantly increased beam backgrounds in Belle II, a stringent system to decide which data is recorded is implemented, in the trigger system. It is responsible for identifying



(a) Leading order Feynman diagram for Bhabha scattering.



(b) Two photon process Feynman diagram.

Figure 22: Feynman diagrams for luminosity backgrounds.

events for physics interest. In Belle II, particles must first pass the hardware based level 1 (L1) trigger and are then again evaluated by the software based high level trigger (HLT).

- **L1 trigger:**
  The level 1 trigger must decide within 5 μs if an event is kept or not while still being able to separate two events within 200 ns. It has an output rate of 30 kHz, limited by the read-in rate of the data acquisition (DAQ) system. It takes data from the various sub detectors like charged track information from the CDC, number and properties from clusters or muon hits in the KLM and merges them to make a decision whether or not to pull the trigger. Over 99.9 % of the $B\overline{B}$ and $e^+e^- \to q\bar{q}$ data is retained.

- **HLT trigger**
  The high level trigger system uses advanced reconstruction algorithms, together with information from all sub detectors (except the PXD) it makes a decision. In order to make a storage of the data possible, it must reduce the event rates to 10 kHz. Many trigger lines exist and they hold information like how many layers of a detector are traversed, how the particle is reconstructed or information about the track angel of the particle.

3.4 Data taking periods

The data taking periods at the Belle II experiment have been split into 3 phases. The first collision data-taking phase, also called phase 2 since the accelerator commissioning phase in 2016 is denoted as phase 1, was taking place from February 2018 until July 2018 with a peak luminosity of $0.5\dot{1}0^{34}cm^{-2}s^{-1}$ where a dataset of order $0.5fb^{-1}$ was collected at the $\Upsilon(4S)$ resonance. At that time, the silicon inner detectors haven't been installed yet, collecting data only for the use of the searches of dark sectors that were previously limited by a lack of efficient triggers. The second collision period, also called (early) phase 3 started on March 25^{th} 2019 and had the full detector to its use. The prefix early was given because some parts of the PXD are still missing and the full PXD will only be installed during a shutdown in 2022. Further, the data taking periods are divided into consecutively numbered experiments, where only some experiments are physics relevant while others are flagged as "bad", for example if one of the detector's subsystems is not working as expected. The weekly recorded luminosity for experiment 7 to 22 can be found in Figure 23.

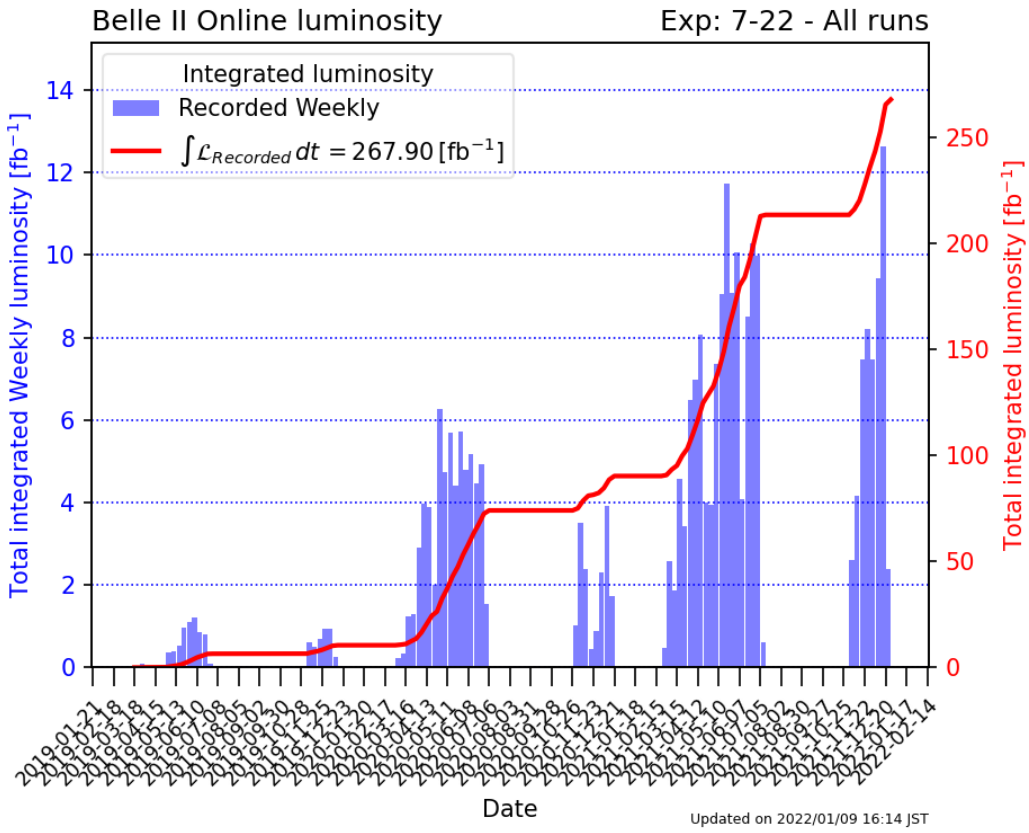


Figure 23: Weekly luminosity recorded in the span from January 21^{st} 2019 until January 9^{th} 2022 together with the integrated luminosity with a peak at $267.9fb^{-1}$ [23].

3.5 Software

The Belle II software package provides the foundation for the development of complex algorithms and their efficient application on large data sets. It is organised in the Belle II Analysis Software Framework also called basf2, contains third parties codes and tools containing scripts for the software installation. This thesis will focus on basf2 since it's been used by this analysis.
The Framework is partitioned into around 40 packages, one for each detector component, base-level framework and many others. Those packages are written in C++, their content is treated in pre-defined subdirectories like modules, tools, dataobjects, scripts, data, tests, validation and examples. The data collected by the Belle II detector is organized into runs, each run containing a sequence of independent events. Each event records the by-product of an e^+e^- collision. A set of runs with similar hardware state and operational characteristics is classified as an experiment. To process a collection of events, basf2 executes dynamical loaded modules, which are arranged linearly in a *path*. The selection of modules, their configuration, and their order of execution are defined via a Python interface.
For a more detailed insight in the Belle II software package and to visit the source of the stated text above, please visit [24].

3.6 Event simulation

Monte Carlo (MC) simulations of particle decays play a big role in every collider experiment. By knowing decay probabilities, we are able to simulate processes that take place during the experiment, however unlike real collisions where the known origin of an event is unknown, we can tell exactly which simulated particle was created through a certain process.
Most studies relay on three event generators:
`EvtGen 1.3` [25] originally developed for BaBar and CLEO works with decay amplitudes instead of probabilities and is used to simulate the decay of B and D mesons into exclusive final states. `PYTHIA 8.2` [26], is used to simulate inclusive decay final states and for the continuum production of light quark pairs. `KKMC 4.15` [27] is used for the τ pair production, while `TAUOLA` handles the τ decays [28]. Additional packages have to be used for QED background processes.
The mentioned simulations are just used to create particle decays but additionally we are also interested in how they interact with the detector. This is based on a software called `Geant4` [29] which is implemented in basf2. It uses Monte Carlo methods to generate realistic trajectories, secondary particles and hits (energy deposits) in the detector and can be adjusted for the Belle II detector in terms of geometry and composition.
The through MC simulation gathered data can then be analysed just like real data with the only difference of knowing which process or particle corresponds to the signal in the detector. This is a huge advantage because models can be tested, algorithms

to separate signal from background can be verified and theories can be investigated. By using the mentioned software packages, one has to be aware that no simulation can imitate real world problems totally accurate, as there will always be a discrepancy between real data and simulated data.

4 Data analysis and particle identification

This section will focus on the methodology of how the collected data from the Belle II detector is used to reconstruct particle decays, with a special focus on the process $e^+e^- \to \tau^+\tau^-$. The following methodology in this chapter is closely related to [1] since the analysis in section 6 is based on the mentioned reference.

4.1 Event reconstruction

For the analysis following in chapter 6, we do not aim to reconstruct B meson decay but rather tau decays. However, similar methods can be used to reconstruct any other particle decay. To observe τ decay as shown in Figure 24, we can search for π^0 production.

Charged tracks that originate from the interaction region, which means they have the properties:

- -3.0 cm $<$ dz $<$ 3.0 cm
- dr $<$ 1.0 cm

are called "good tracks". Events that pass that criteria are most likely originated from the collision. To further reconstruct neutral pions we are combining two photons satisfying the requirements:

- $E(\gamma) > 100MeV$
- -0.8660 $<$ cos(θ) $<$ 0.9565 (in the CDC acceptance)
- clusterNHits $>$ 1.5
- 0.115 GeV $<$ $M_{\gamma\gamma}$ $<$0.152 GeV

where $M_{\gamma\gamma}$ denotes the invariant photon mass. Photons that pass the above requirements but have higher Energy $E(\gamma) > 200MeV$ are considered separately and denoted as "additional photons".

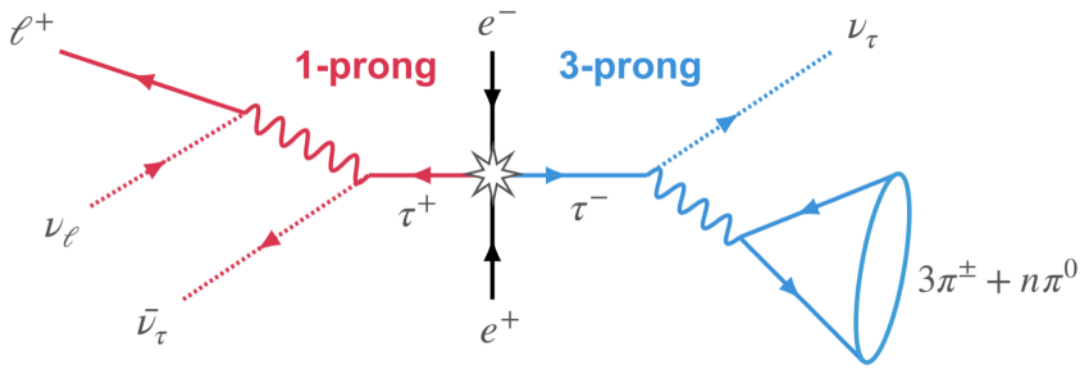


Figure 24: Illustration of the 3x1 prong topology for a $e^+e^- \rightarrow \tau^+\tau^- \rightarrow (\ell^\pm\nu\bar{\nu})(3\pi^\pm n\pi^0)$ event. The particles on the signal side (1-prong) are shown in red, while those in the tag side (3-prong) are shown in blue [1].

Thrust Vector

Thrust is a event shape variable and is used to select the axis that maximizes the sum of the longitudinal momentum components. We use the thrust vector to separate the events into signal and tag hemispheres. The thrust axis $\hat{n}_{trust}$ is defined such that the value V_{thrust}

$$V_{thrust} = \frac{|\overrightarrow{p_i}^{CMS}\hat{n}_{trust}|}{\sum \overrightarrow{p_i}^{CMS}} \tag{4.1}$$

is maximised, whereas $\overrightarrow{p_i}^{CMS}$ is the momentum of each particle in the event. This implicates that for $V_{thrust} = 1$ the event is perfectly diametrically opposed while for $V_{thrust} = 1/2$ the event is spherically symmetric. The hemisphere corresponding to the signal side, $\tau \rightarrow \ell\bar{\nu}\nu$ ($\ell = e, \mu$), should contain a single track ("1-prong"), while the three tracks ("3-prong") composing the tag side are expected to be contained in the opposite hemisphere. The 3x1 topology can be selected by requiring:

$$|\overrightarrow{p}^{CMS}_{signal}\hat{n}_{trust}||\overrightarrow{p}^{CMS}_{tag,i}\hat{n}_{trust}| < 0, \ \forall i \in tag \tag{4.2}$$

According to the origin of the track on the signal side, we define

- *electron channel*, the 1-prong track originates from an electron
- *muon channel*, the 1-prong track originates from a muon

4.2 Particle identification (PID)

Each particle is assigned an identification (ID), corresponding to the likelihood that a track can be identified as a certain particle. For every sub detector, a likelihood is defined based on a probability density function (PDF) that can be derived from

simulations, analytical calculation or directly from data.
The total PID likelihood is given as the product of all sub detectors (denotes as *det* in equation 4.3) likelihood for a given hypothesis i:

$$\mathcal{L} = \prod_{det} \mathcal{L}_{det}(i) \tag{4.3}$$

In general, i can denote electrons, muons, pions, kaons, protons and deuterons which are the most common charged particles. With the information of equation 4.3, we can state the global PID, which is defined as the ratio:

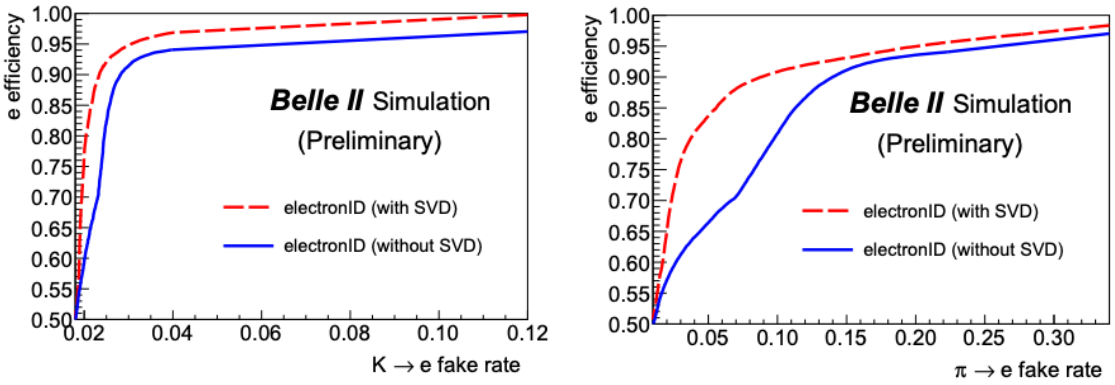
$$globalID(i) = \frac{\prod_{det} \mathcal{L}_{det}(i)}{\sum_j \prod_{det} \mathcal{L}_{det}(j)} \tag{4.4}$$

where the sum is over all possible particles. In other words, equation 4.4 divides the likelihood of obtaining particle i, by the sum of the likelihoods of all particles. In practice, the globalID is named according the hypothesis that is tested (e.g muonID, electronID, ect.) and ranges from 0 to 1.
Additionally, as explained in section 3.3, the ionisation energy loss in the CDC and ECL can be used to introduce an additional likelihood depending on the function shown in Figure 17:

$$\mathcal{L}(\frac{dE}{dx}, p) = \prod_q \mathcal{P}_i^{det}(\frac{dE}{dx}_q, p) \quad det \ = \ PXD,\ SVD,\ CDC \quad i \ = \ \pi,\ K,\ e,\ \mu,\ p \tag{4.5}$$

where q loops over all values of $\frac{dE}{dx}$ and $\mathcal{P}$ is the two dimensional PDF of the ionisation function. It should be mentioned that $\mathcal{P}$ varies between all subdetectors.
By considering the likelihood obtained in equation 4.5, we can claim significant improvement in the particle identification as shown in Figure 25, where only the ionisation energy loss in the PXD is included.

To correctly reconstruct events, we use tagged and untagged reconstruction. Untagged analysis makes use of the hermetical coverage of the detector, leading to an relatively exact measurement of the 4-momentum of one of the two mesons. One meson can be reconstructed by measuring the daughter particles, called signal B meson (B_{sig}). The 4-momentum of the other B meson, called companion B meson (B_{comp}), is determined by adding up the 4-momenta of all the remaining charged tracks and neutral clusters in the event. This yields to the calculation of the missing momentum $p_{miss} = p_{\Upsilon(4s)} - p_{B_{sig}} - p_{B_{comp}}$ up to a efficiency of $\mathcal{O}(10\%)$. It is therefore recommended to use untaged analysis for smaller data sets.
In tagged analysis, we first try to fully reconstruct the companion B meson. Once a candidate is found, we require that the rest of event is consistent with the signature of the signal decay and thus, the p_{miss} can be calculated as in the untagged case. This method yields to a higher precision with a trade of in efficiency which falls to $\mathcal{O}(0.1\%)$ which makes it more usable for bigger data sets [31].
To correctly classify particles in the decay shown in Figure 24, we identify particles with the criterion:



(a) Electron efficiency vs $\pi \rightarrow e$ fake rate. **(b)** Electron efficiency vs $K \rightarrow e$ fake rate..

Figure 25: Electron efficiency vs fake rate for different criteria on the total PID variable with the fake rate being defined as $j \rightarrow i\ fake\ rate = \frac{No.\ of\ tracks\ identified\ with\ PID\ under\ the\ hypothesis\ i}{No.\ of\ tracks\ kinematically\ identified\ under\ the\ hypothesis\ j}$.[30].

- $clusterE/p \leq 0.8$

on the tag side, where "clusterE" denotes the energy deposited in the ECL. By this requirement we specifically search for hadrons which deposit less energy in the ECL than an electron (because the ratio $clusterE/p \sim 1$ for electrons). In the electron channel, the 1-prong track is required to satisfy one of the following two criteria:

- electronID >0.9 (excluding TOP) , or
- BDT electron D > 0.99

and in the muon channel:

- muonID >0.9 , or
- BDT muonID > 0.9

where BDT (boosted decision tree) is the output of a machine learning algorithm, explained in section 2.0.3.

4.3 Trigger requirements

As explained in section 3.3.2, the trigger system is used to gather additional information. In general, one can use trigger information of every subdetector if available, here we will focus on the ECL trigger which provides total energy, cluster information, and Bhabha identification information.
Only events in the muon and electron channel that fire the logical OR of several low-multiplicity ECL (lml) triggers are considered:

- **lml0:** ≥ 3 clusters with at least one having $E^* > 300$ MeV, $12.4° < \theta < 154.7°$ and not an ECL Bhabha.
- **lml1:** exactly 1 cluster with $E^* > 2$ GeV and $32.2° < \theta < 124.6°$
- **lml2:** ≥ 1 cluster with $E^* > 2$GeV, $18.5° < \theta < 32.2°$ or $124.6° < \theta < 139.3°$ and not an ECL Bhabha.
- **lml4:** ≥ 1 cluster with $E^* > 2$ GeV, $12.4° < \theta < 154.7°$ and not an ECL Bhabha.
- **lml6:** exactly 1 cluster with $E^* > 1$ GeV, $32.2° < \theta < 128.7°$, full ECL barrel and no other cluster with $E > 300$ MeV anywhere.
- **lml7:** exactly 1 cluster with $E^* > 1$ GeV, $18.5° < \theta < 31.9°$ or $128.7° < \theta > 139.3°$ and no other cluster with $E > 300$ MeV anywhere.
- **lml8:** cluster pair with $170° < \Delta\phi < 190°$, both clusters with $E^* > 250$ MeV and no 2 GeV cluster in the event.
- **lml9:** cluster pair with $170° < \Delta\phi < 190°$, one cluster with $E^* < 250$ MeV with the other having $E^* > 250$ MeV, and no 2 GeV cluster in the event.
- **lml10:** cluster pair with $160° < \Delta\phi < 200°$,$160° < \sum\theta < 200°$ and no 2 GeV cluster in the event.
- **lml12:** ≥ 3 clusters with at least one having $E^* > 500$ MeV, $18.5° < \theta < 139.3°$, full ECL and not an ECL Bhabha

The resulting logical expression is therefore:
lml0 or **lml1** or **lml2** or **lml4** or **lml6** or **lml7** or **lml8** or **lml9** or **lml10** or **lml12**.
In addition to the mentioned methods in this chapter, one can use kinematic and event shape properties to correctly reconstruct particle decays.

5 Lepton Identification using BDT and FastBDT

The analysis in this thesis is split into two parts: The first part concerning lepton identification using two different but similar machine learning techniques, the second part using the knowledge gained in the first part to improve signal identification for LFU violation in tau decays.
Lepton identification plays a big role in most analysis done at Belle II because this group of particles seem to behave in a similar way. All leptons have the same charge and interact via the weak force, making it hard to differ between them. While they all differ in mass and lifetime, there are other criteria that can help to distinguish, for example muons hardly interact with matter. In addition, hadrons especially pions can behave similar to leptons and thus are easily miss classified.

5.1 Sample generation and reconstruction

We generate one million samples per particle ($e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$) with ParticleGun Module in Basf2 in three momentum range, uniform $p_{low} \in$ (0.2,0.6) GeV/c, $p_{medium} \in$ (0.6,1.0) GeV/c, $p_{high} \in$ (1.0,5.0) and uniform $\theta \in$ (0°, 180°), uniform $\phi \in$ (0°, 360°). The background was simulated using the background overlayer method prerelease-05-00-00a.
Following reconstruction criteria have been used:

- inCDCAcceptance = 1
- clusterTrackMatch = 1
- $|dr| < 2.0$
- $|dz| < 5.0$

For electrons, Bremsstrahlung correction has been done using "*correctBrems*" module in basf2. For that we are using

- γ Energy(E) $<$ 0.1 GeV

- Truth matched
- inTOPAcceptance
- inARICHAcceptance
- inKLMAcceptance according to training region.

5.2 Variables and parameters

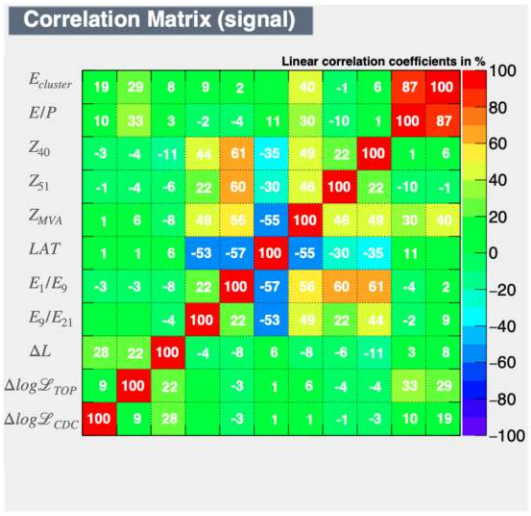
The listed variables below have a strong correlation as shown in Figure 26 and have therefore been used for the training.

- E_9/E_{21}
- $E_{cluster}$
- E/P
- Z_{MVA}
- $\Delta log\mathcal{L}_{CDC}$
- ΔL
- $\Delta log\mathcal{L}_{TOP}$
- Z_{40}
- Z_{51}
- LAT
- E_1/E_9

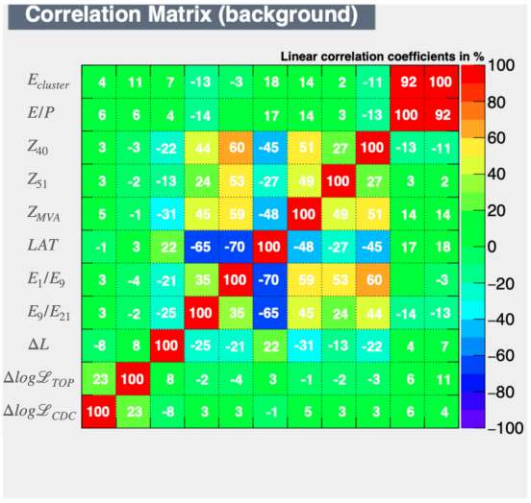
In the $\mathcal{L}_{CDC}$ and $\mathcal{L}_{TOP}$ feature vector occurred NaN values which could not been processed by the algorithm. We therefore set those values to 0.
We used 500 trees and a learning rate of 0.2 together with a maximum tree depth of 2. The data set was then split into 80 % training data and 20% evaluation data.

5.3 Results

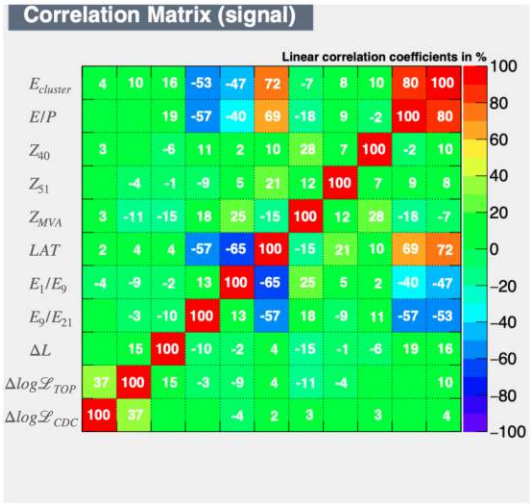
We compared two algorithms, a gradient boosted decision tree built in the *sklearn* library and a FastBDT, to see which can better separate signal from background. In one scenario, we define signal as leptons (either electrons or muons) and pions as background, in the other scenario we tried to differentiate between electrons and



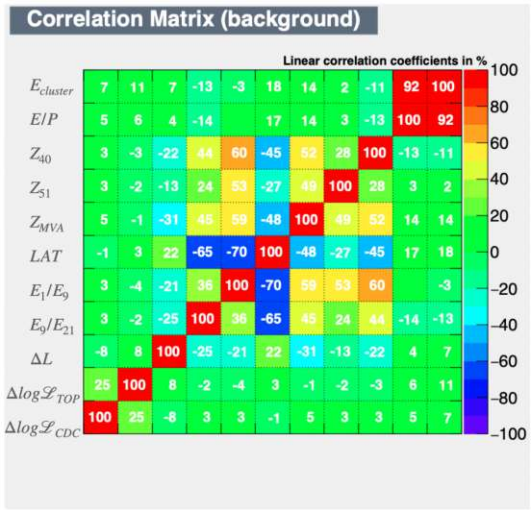
(a) Confusion Matrix of e/π Signal events



(b) Confusion Matrix of e/π Background events



(c) Confusion Matrix of e/mui Signal events



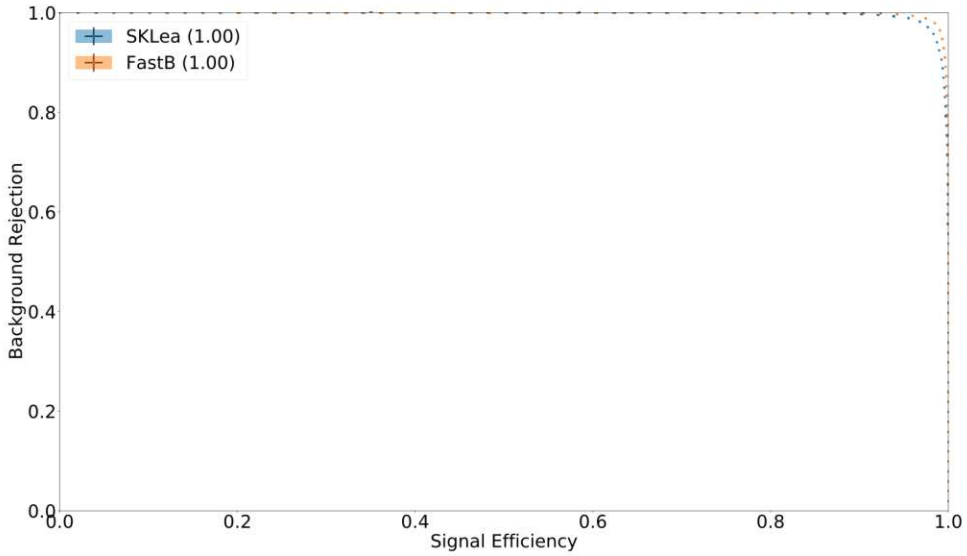
(d) Confusion Matrix of e/μ Background events

Figure 26: Correlation in percentage of e/π 26a,26b and e/μ separation 26c,26d. The order of the x-axis labels is the same as for the y-axis.

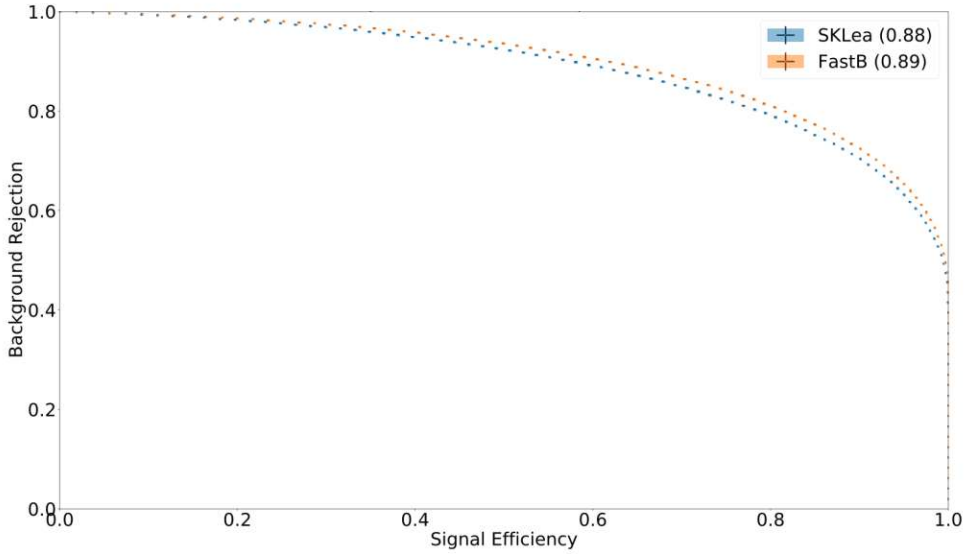
muons. The analysis was done for the three momentum ranges, mentioned in section 5.1. The performance will be evaluated by their respective ROC curves but also in terms of execution time.

The resulting execution times can be found in Table 5.1,5.2,5.3.

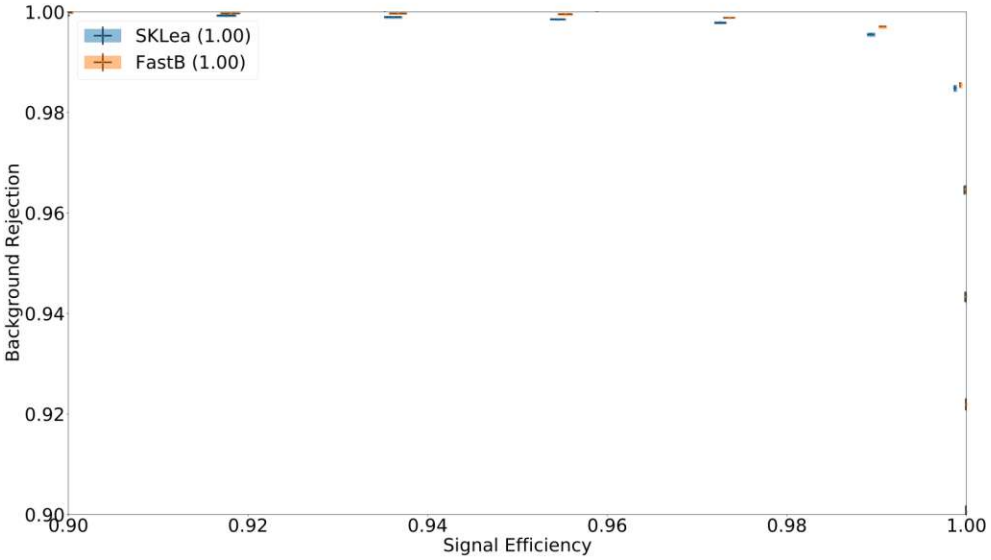
While the FastBDT does not improve the classification significantly, it definitely performs better in terms of execution time. This can be especially useful for big data sets. Additionally, we investigated how well the BDT performs if we choose to classify either muons/pions as background and electrons as signal, electron pions as background and



(a) e/π separation in p_{low}

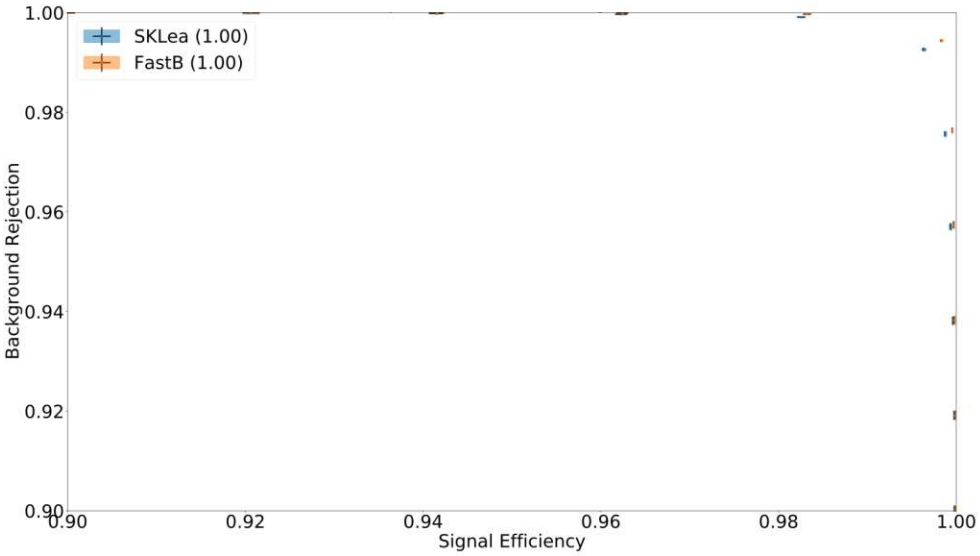


(b) μ/π separation in p_{low}

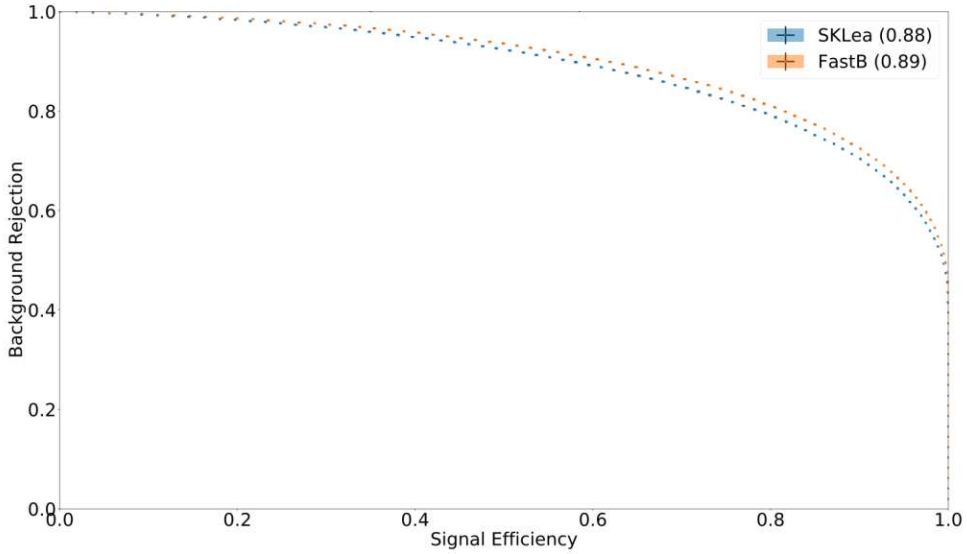


(c) μ/e separation in p_{low}

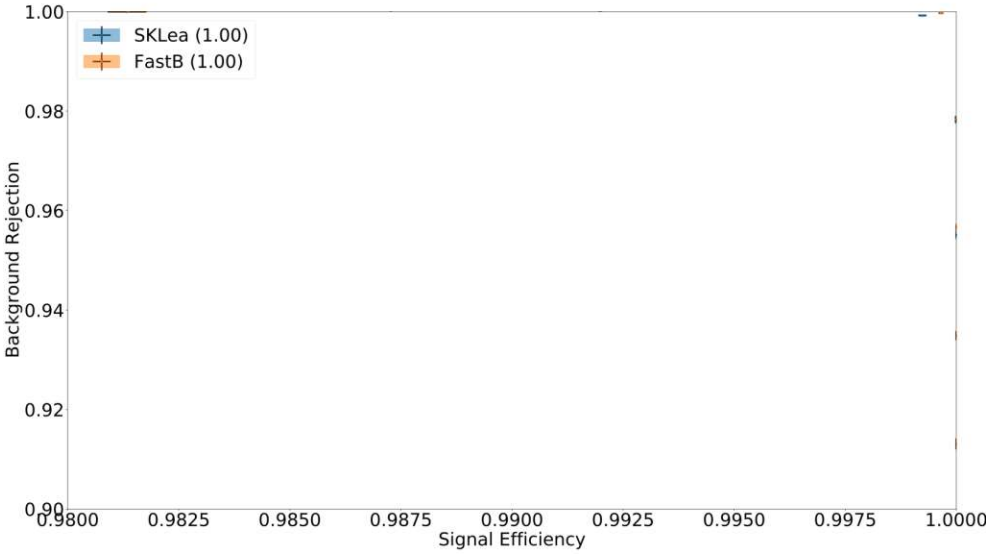
Figure 27: ROC rejection plot in momentum range $p_{low} \in (0.2,0.6)$.



(a) e/π separation in p_{medium}

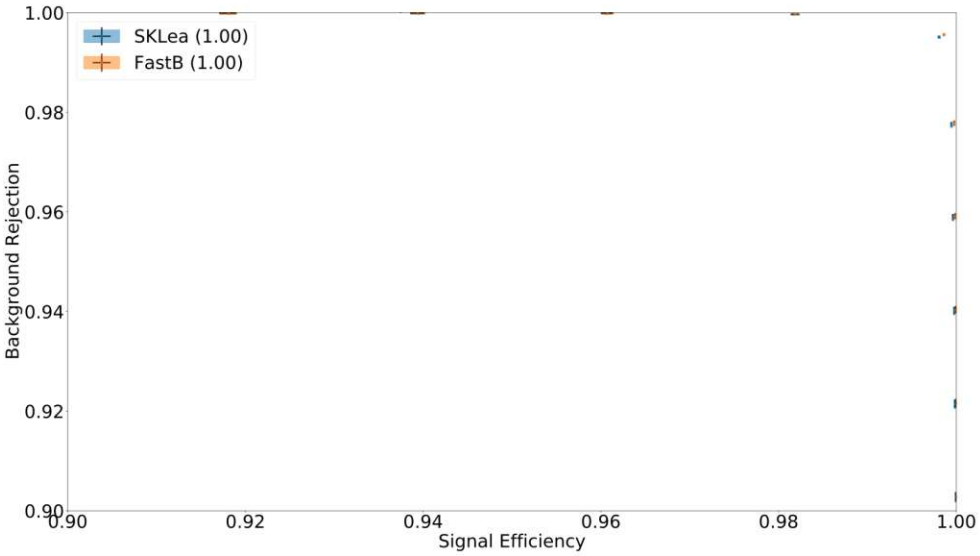


(b) μ/π separation in p_{medium}

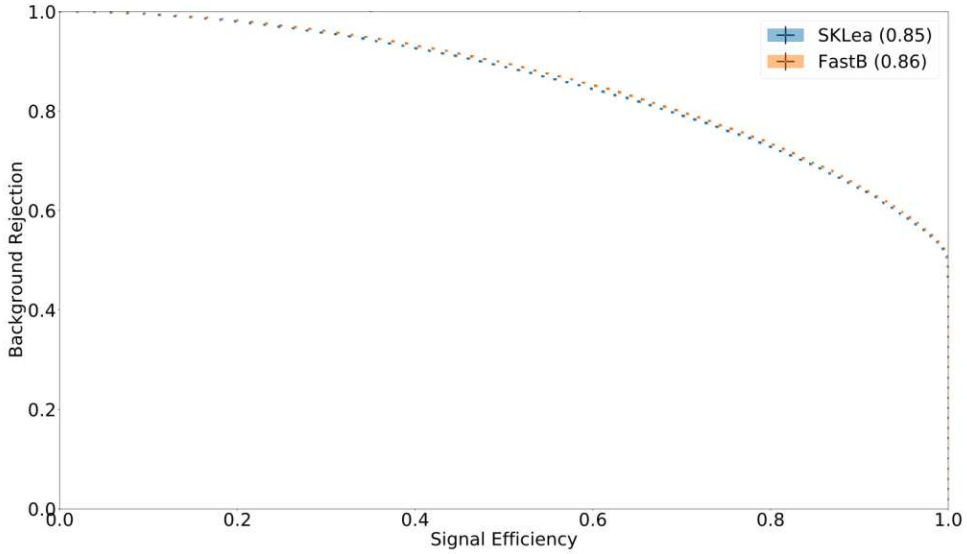


(c) μ/e separation in p_{medium}

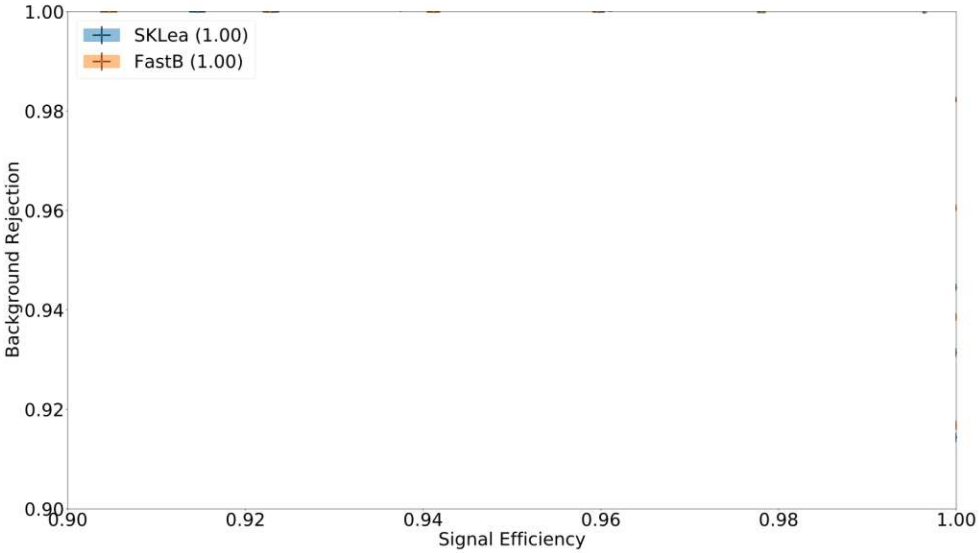
Figure 28: ROC rejection plot in momentum range $p_{medium} \in (0.6,1.0)$.



(a) e/π separation in p_{high}



(b) μ/π separation in p_{high}



(c) μ/e separation in p_{high}

Figure 29: ROC rejection plot in momentum range $p_{high} \in (1.0,5.0)$.

Table 5.1: Execution time for p_{low}. The mean time represents the mean of the execution time of all three separation types. Min and Max time give the time and the separation type that executed the fastest/slowest.

Algorithm	Mean time [s]	Min time [s]/separation	Max time [s]/separation
sklearn	163.7	156.3/ e/π	169.7/ μ/π
FastBDT	22.2	21.3/ e/π	22.7/ μ/π

Table 5.2: Execution time for p_{medium}. The representation is the same as for the low momentum range

Algorithm	Mean time [s]	Min time [s]/separation	Max time [s]/separation
sklearn	286.4	271.3/ e/π	306.9/ μ/π
FastBDT	36.3	34.9/ e/π	38.0/ μ/π

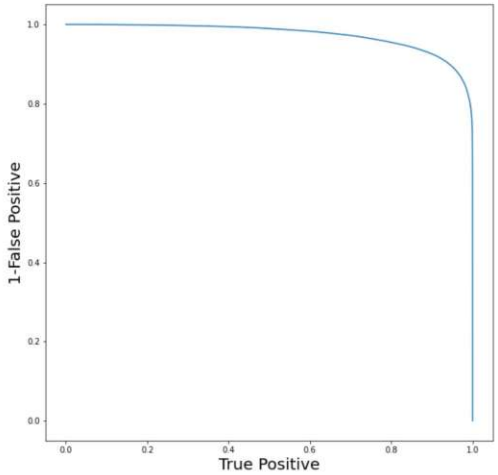
Table 5.3: Execution time for p_{medium}. The representation is the same as for the low momentum range

Algorithm	Mean time [s]	Min time [s]/separation	Max time [s]/separation
sklearn	303.0	293.2/ e/π	320.1/ μ/π
FastBDT	39.0	36.8/ e/π	42.5/ μ/π

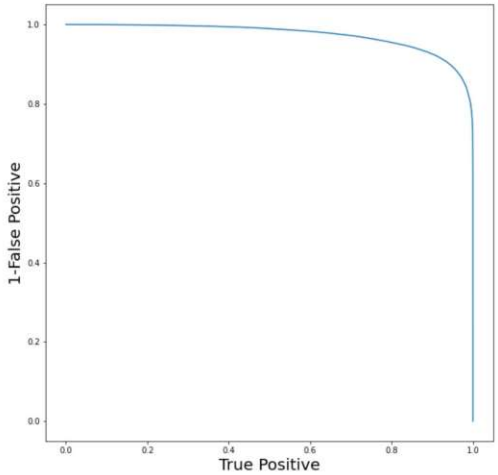
muons as signal or electrons/muons as background and pions as signal. The ROC curve for the three momentum range are given in Figure 30, 31, 32. The curves will be discussed in detail in section 5.4.

Table 5.4: Grid search output values for all momentum ranges.

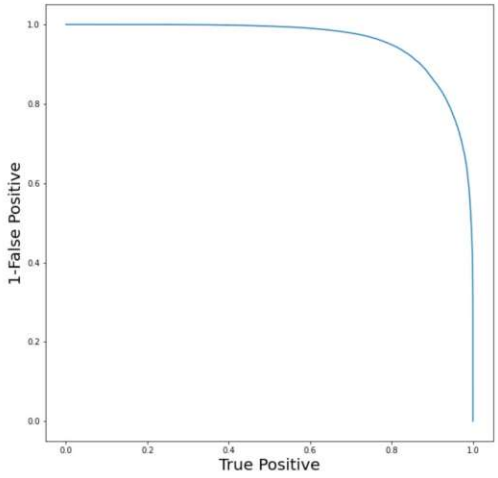
Separation	Max depth	Learning rate	momentum range
e vs π	32	0.9	p_{low}
μ vs π	10	0.3	p_{low}
μ vs e	32	0.9	p_{low}
e vs π	32	0.8	p_{medium}
μ vs π	16	0.1	p_{medium}
μ vs e	32	0.8	p_{medium}
e vs π	8	0.3	p_{high}
μ vs π	8	0.4	p_{high}
μ vs e	8	0.2	p_{high}



(a) Multiclassificatoin ROC curve in momentum range p_{low}. Electrons are tagged as signal, μ and π as background.

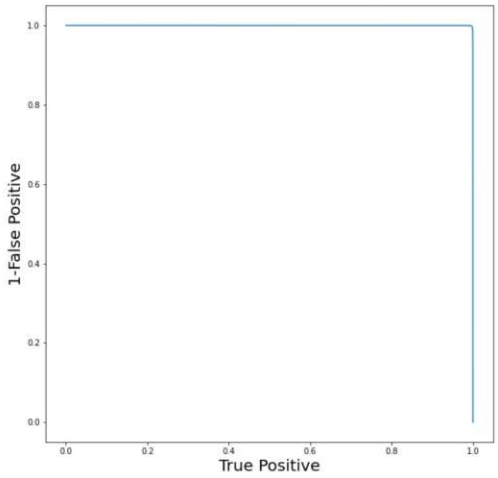


(b) Multiclassificatoin ROC curve in momentum range p_{low}. Muons are tagged as signal, e and π as background.

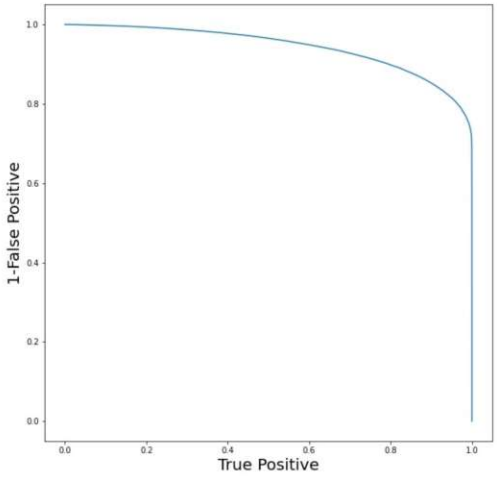


(c) Multiclassificatoin ROC curve in momentum range p_{low}. Pions are tagged as signal, e and μ as background.

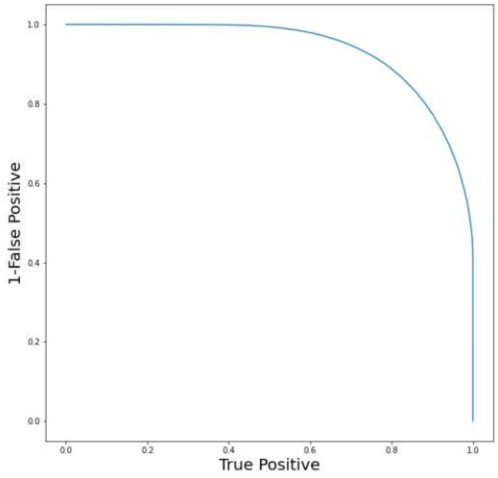
Figure 30: ROC rejection plot in momentum range $p_{low} \in (0.0,0.6)$



(a) Multiclassificatoin ROC curve in momentum range p_{medium}. Electrons are tagged as signal, μ and π as background.

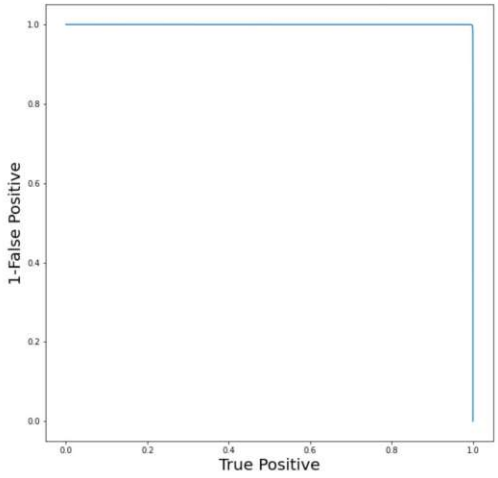


(b) Multiclassificatoin ROC curve in momentum range p_{medium}. Muons are tagged as signal, e and π as background.

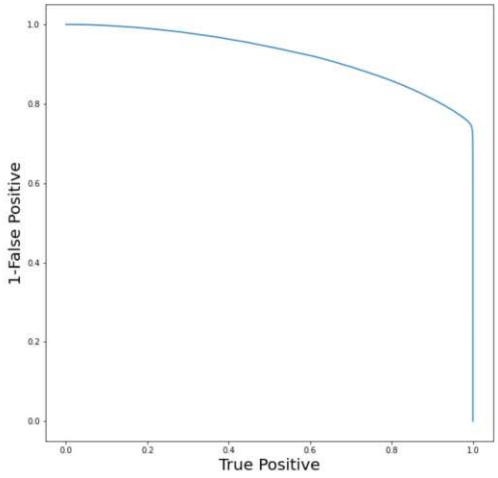


(c) Multiclassificatoin ROC curve in momentum range p_{medium}. Pions are tagged as signal, e and μ as background.

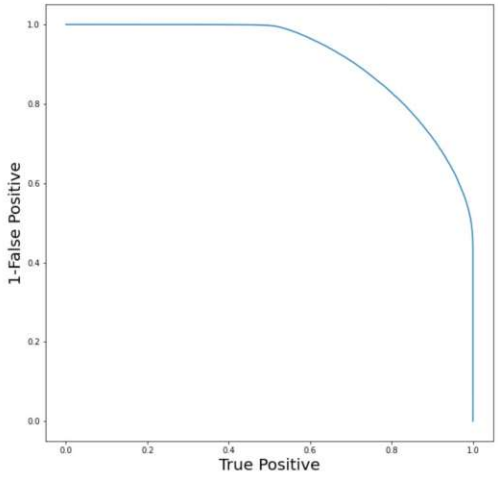
Figure 31: ROC rejection plot in momentum range $p_{medium} \in (0.6,1.0)$.



(a) Multiclassificatoin ROC curve in momentum range p_{high}. Electrons are tagged as signal, μ and π as background.



(b) Multiclassificatoin ROC curve in momentum range p_{high}. Muons are tagged as signal, e and π as background.



(c) Multiclassificatoin ROC curve in momentum range p_{high}. Pions are tagged as signal, e and μ as background.

Figure 32: ROC rejection plot in momentum range $p_{medium} \in (1.0,5.0)$.

As a final step, we tried to optimise the e/π,e/μ,μ/π separation by finding the best parameters, thus performing a grid search. The results of that search can be found in table 5.4.

5.4 Conclusion

As could be shown in the previous section, the FastBDT performs slightly better than the algorithm from *sklearn*. This could be a result of the used calculation method.

FastBDT is a Stochastic Gradient Boosted Decision Trees, meaning that the algorithm is using a randomly drawn sub-sample of the given training data. This yields to a greater robustness against over-fitting. The *sklearn* algorithm however is based on gradient boosted decision tree. This method uses gradient-decent in each boosting step to re-weight the original training sample and thus, data-points which are hard to classify gain influence during the training [32].
Besides the minimal effects on the resulting ROC curves, the FastBDT is considerably faster. It executes on an average of 218.53 seconds faster than *sklearn* algorithm and that number becomes even higher for larger data sets.
On the other hand, the variables used in section 5.2 are highly efficient for the separation of electrons, muons and pions. While it gets usually harder to distinguish between muon or pion signal for higher momentum particles, it's very accurate for electron identification in all momentum ranges. The pion classification is the most restricted one as it suffers the most from detector effects. Especially the high momentum range in Figure 32c but also Figure 32b show a big drop in the ability to classify true negative samples.
Nevertheless, a BDT with the stated variables can efficiently be used to separate electrons from muons and pions, muons from electron and pions and pions from electrons and muons. Algorithms like that will help Belle 2 in the future to gain higher detector resolutions without the need of additional hardware. They will be used in a variety of task and gaining experience with such models will also help in future experiments.

6 Lepton flavour universality violation in τ decays

As shown in the previous chapter, gradient boosted decision trees are successfully used in High energy physics because of their tremendous accomplishments in particle identification. It is therefore of great interest, if a cut based analysis can be improved by an BDT approach. To directly compare these two methods, one can use the purity and efficiency metric, explained in section 2.1. One analysis is said to perform better than the other in terms of efficiency/purity, if the efficiency at a given level of purity is higher. We will therefore compare the efficiency of the BDT based analysis to two cut based analysis: One done by the Belle II collaboration [6] and one done by the BABAR collaboration [5]. The analysis will focuses on the $e^+e^- \to \tau^+\tau^-$ events displaying a 3x1 topology. The comparison will be done both in the electron-, as well as in the muon -channel.

6.1 Data Sets

The Data used in this note originates from the the 14^{th} official campaign of run independent MC (MC14ri a). They were produced using release-05-02-00 with early Phase 3 geometry and nominal beam background conditions (BGx1). To be more precise:

- generic τ -pair sample, split into signal and background according to the $\tau \to \ell\bar{\nu}\nu$ ($\ell = e, \mu$) decay channel using MC truth information
- $q\bar{q}$ samples, including $u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$ and $b\bar{b}$ (both $B^0\bar{B}^0$ and $B^+\bar{B}^-$)
- radiative dilepton background samples, including $ee(\gamma)$ and $\mu\mu(\gamma)$
- two photon and other low-multiplicity background samples, including $eeee$, $ee\mu\mu$, $ee\pi\pi$, $eeKK$, $eepp$, $\mu\mu\mu\mu$

For a more detailed description regarding the cross section and luminosity, see table 6.1. Please note that the different Luminosity will be respected during the training of the BDT in terms of a weight array, which will then be passed to the algorithm.

Table 6.1: : The simulated processes with their cross section and the corresponding luminosity of the generated samples (MC14ri a). If not suggested otherwise all particles are to be understood as charged particles such that the overall charge sums to zero.

Process	cross-section [nb]	Luminosity [fb^{-}1]	Number of events
$ee \to \tau\tau$	0.919	100	92M
$ee \to u\bar{u}$	1.605	100	161M
$ee \to d\bar{d}$	0.401	100	40M
$ee \to s\bar{s}$	0.383	100	38M
$ee \to c\bar{c}$	1.329	100	133M
$ee \to B^+B^-$	0.54	100	54M
$ee \to B^0\bar{B}^0$	0.51	100	51M
$ee \to ee(\gamma)$	2.958	10	2958M
$ee \to \mu\mu(\gamma)$	1.148	100	115M
$ee \to eeee$	39.55	100	3955M
$ee \to ee\mu\mu$	18.83	100	1883M
$ee \to ee\pi\pi$	1.895	100	190M
$ee \to eeKK$	0.0798	1000	80M
$ee \to eepp$	0.0117	1000	11M
$ee \to ee\mu\mu$	0.3512×10^{-3}	1000	0.35M

6.2 Event Selection and Set up

Events required to contain exactly 4 good quality tracks. Use thrust vector to separate events into signal (1-prong) and tag (3-prong) hemispheres. Events that pass the following criteria are tagged as Signals:

- |track1_prong_mcPDG| == 11 ,13 (e, μ -channel)
- |dmID_1prong| == 11,13 (e, μ) -channel).

As shown in section 4.2, in the electron channel the 1-prong track is required to satisfy the lepton identification (ID) condition:

- Likelihood electron ID > 0.9 (excluding TOP)

and in the muon channel:

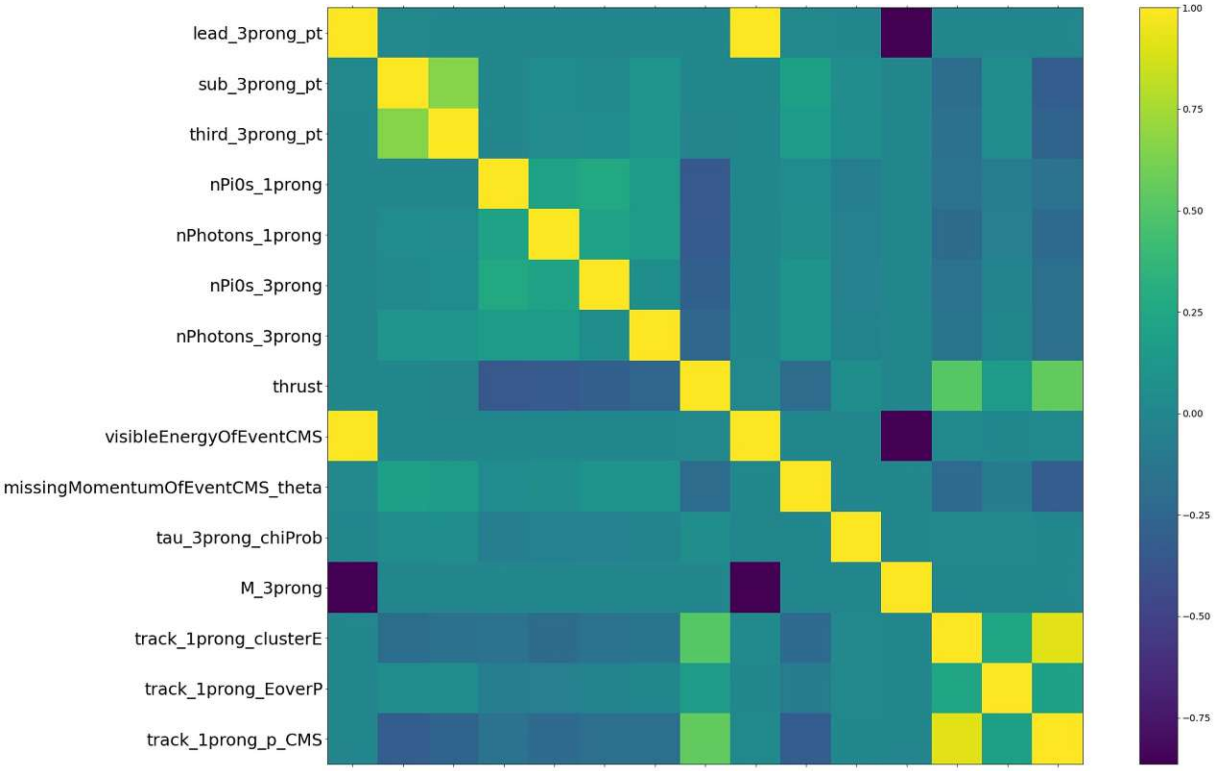
- Likelihood muon ID > 0.9.

6.2.1 Variables and algorithm

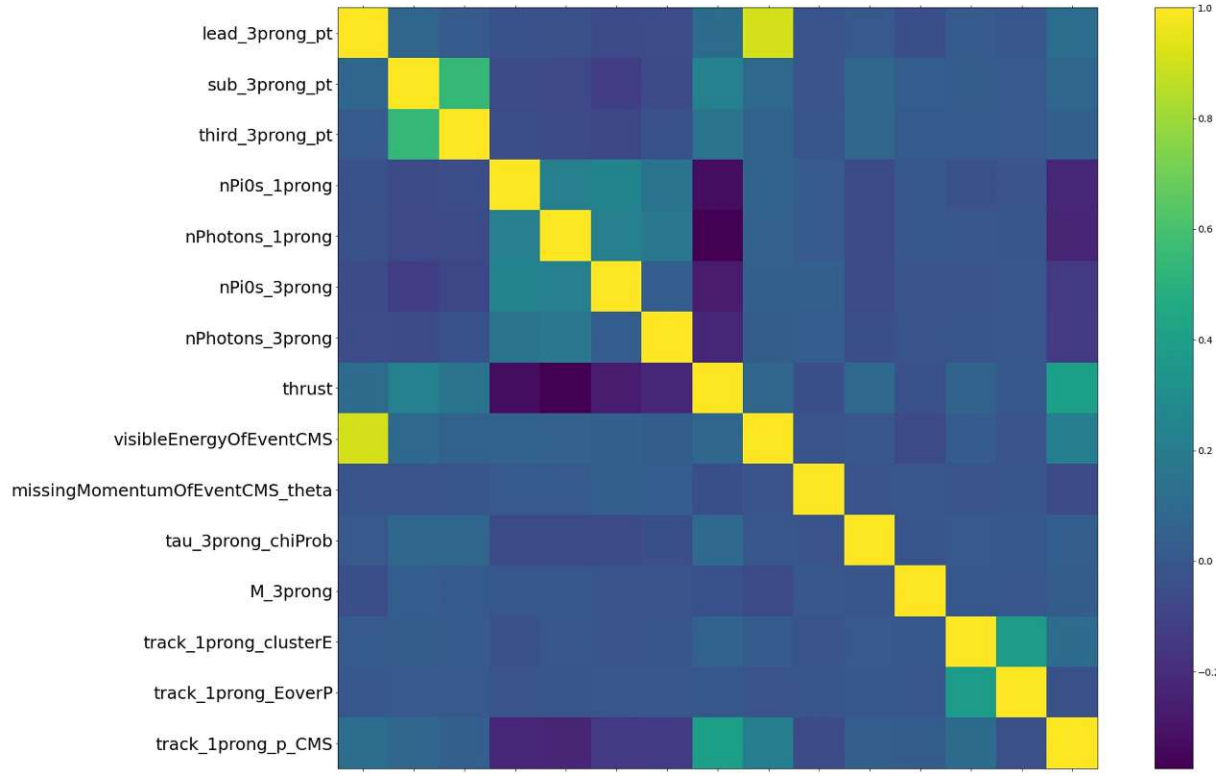
The following Variables have been used for the training:

- lead_3prong_pt
- sub_3prong_pt
- third_3prong_pt
- nPi0s_1prong
- nPhotons_1prong
- nPi0s_3prong
- nPhotons_3prong
- thrust
- visibleEnergyOfEventCMS
- missingMomentumOfEventCMS_theta
- tau_3prong_chiProb
- M_3prong
- track_1prong_clusterE
- track_1prong_EoverP
- track_1prong_p_CMS.

It was chosen to implement the *sklearn* Histogram Gradient boosting classifier due to its fast performance, with a learning rate of 0.2 and a maximum of iterations of 500. The data was randomly split into a subset of 80 % training data and 20 % evaluation data. The correlation matrix of the underlining data can be found in Figure 33.



(a) Variable correlation in the electron channel



(b) Variable correlation in the Muon channel

Figure 33: Correlation matrix of the underlining data. The x-Axis represents the same columns in the same order as seen in the y-Axis.

6.3 Results

Once the training is completed, one can obtain the Probability distribution shown in Figure 34.

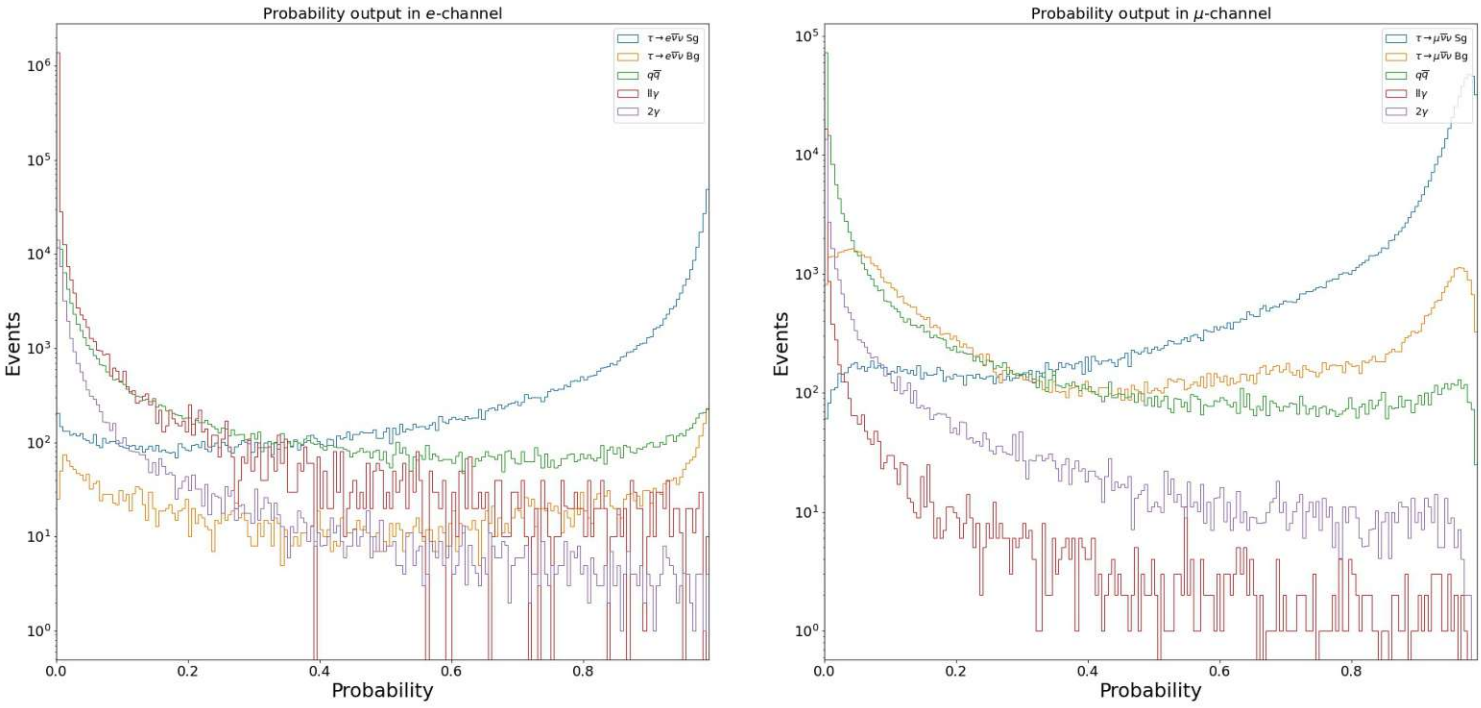


Figure 34: Probability output distribution after the BDT was applied. The signal events dominate in the region with higher probability.

In the following, we compared the most important variables in the training shown in table 6.2 at a high probability. Since the background does not completely vanish for high probabilities, we expect a dominant signal distribution with little background which can be seen in Figure 35 and 36 for the five most import variables.

Table 6.2: Five most important variables during the training calculated with the Permutation feature importance technique.

Rank	e-channel	μ-channel
1.	thrust	nPi0s_1prong
2.	track_1prong_p_CMS	nPhotons_1prong
3.	lead_3prong_pt	M_3prong
4.	nPi0s_1prong	track_1prong_p_CMS
5.	M_3prong	visibleEnergyOfEventCMS

The distribution after the BDT was applied show a similar signal function but the background is strongly suppressed with only around 10 events passing the BDT. This also holds for the muon channel and the background can be even more suppressed if one decides to choose an even higher output probability.

6.3.1 Efficiency and Purity

The performance of the aforementioned event selections are given in terms of efficiency and purity. These are defined as:

$$\varepsilon = \frac{N^{signal}_{selected}}{N^{signal}_{generated}} \quad P = \frac{N^{signal}_{selected}}{N^{total}_{selected}} \tag{6.1}$$

where $N^{signal}_{selected}$ is the number of selected signal events, $N^{signal}_{generated}$ is the number of generated signal events,

$$N^{signal}_{generated} = 2\,\sigma_{\tau\tau}\,\mathcal{L}\,\mathcal{B}_{\tau\to\ell\bar{\nu}\nu}\,\mathcal{B}_{Tag} \tag{6.2}$$

and $N^{total}_{selected}$ is the total number of selected signal and background events. On the other hand, the efficiency and Purity of the BDT is defined as:

$$\varepsilon_{BDT} = \frac{True\ positive}{True\ positive + False\ negative} \quad P_{BDT} = \frac{True\ positive}{True\ positive + False\ positive} \tag{6.3}$$

which leads to an total efficiency of:

$$\varepsilon_{Total} = \varepsilon_{BDT} * \varepsilon \tag{6.4}$$

The achieved values for ε and P are compared to the ones from the cut based analysis done by the BELLE II collaboration [6] and by the BABAR collaboration [5]. The results can be found in table 6.3

Table 6.3: Comparison between efficiency and purity of the cut based analysis, BDT based analysis and world-leading measurements from BABAR.

e **channel**		BELLE II (cut based)	BELLE II (BDT based)	BABAR
	$\varepsilon(\%)$	13.21 ± 0.02	25.4	3.8
	$P(\%)$	99.76 ± 0.01	99.76	99.7 ± 0.06
μ **channel**				
	$\varepsilon(\%)$	14.69 ± 0.02	22.55	3.22
	$P(\%)$	97.91 ± 0.02	97.81	97.3

As seen in table 6.3, the results show an major improvement to the cut based analysis (∼double the efficiency in e channel, ∼1.5 times in μ channel) and to the results of the BABAR experiment (∼ 6.5 times better efficiency in e-channel, ∼ 7 times in μ-channel).
A classification plot of both channels can be found in Figure 37.

6.3.2 ROC curve

The Receiver operating characteristic is a very common used metric in machine learning. It is defined as the true positive rate on the vertical axis and false positive rate on the horizontal axis, as the classification threshold varies.
The ROC curve of the underlining data can be found in Figure 38.

The score in Figure 38 represents the area under the curve and it tends toward 1 for both channels. This is a very good result since a score of 1 would reflect a perfect separation.

6.4 Conclusion

The BDT approach with the used variables could effectively improve the cut based analysis from both collaborations. We gained a higher level of efficiency which ultimately lead to a restriction of the phase space. This is a very important result because we can now sacrifice efficiency to restrict the momentum range and get a more accurate result. For a better understanding, see Figure 39. In some of the bins, the uncertainty is large. By excluding some bins or regions in $\theta / \ p$ to reduce the overall systematics. The efficiency would then of course drop, but it would be compensated for by the increased efficiency from the BDT.

Future analysis can repeat the R_μ sensitivity study (see equation 2.2) after including the BDT and restricting the phase space to reduce systematics.

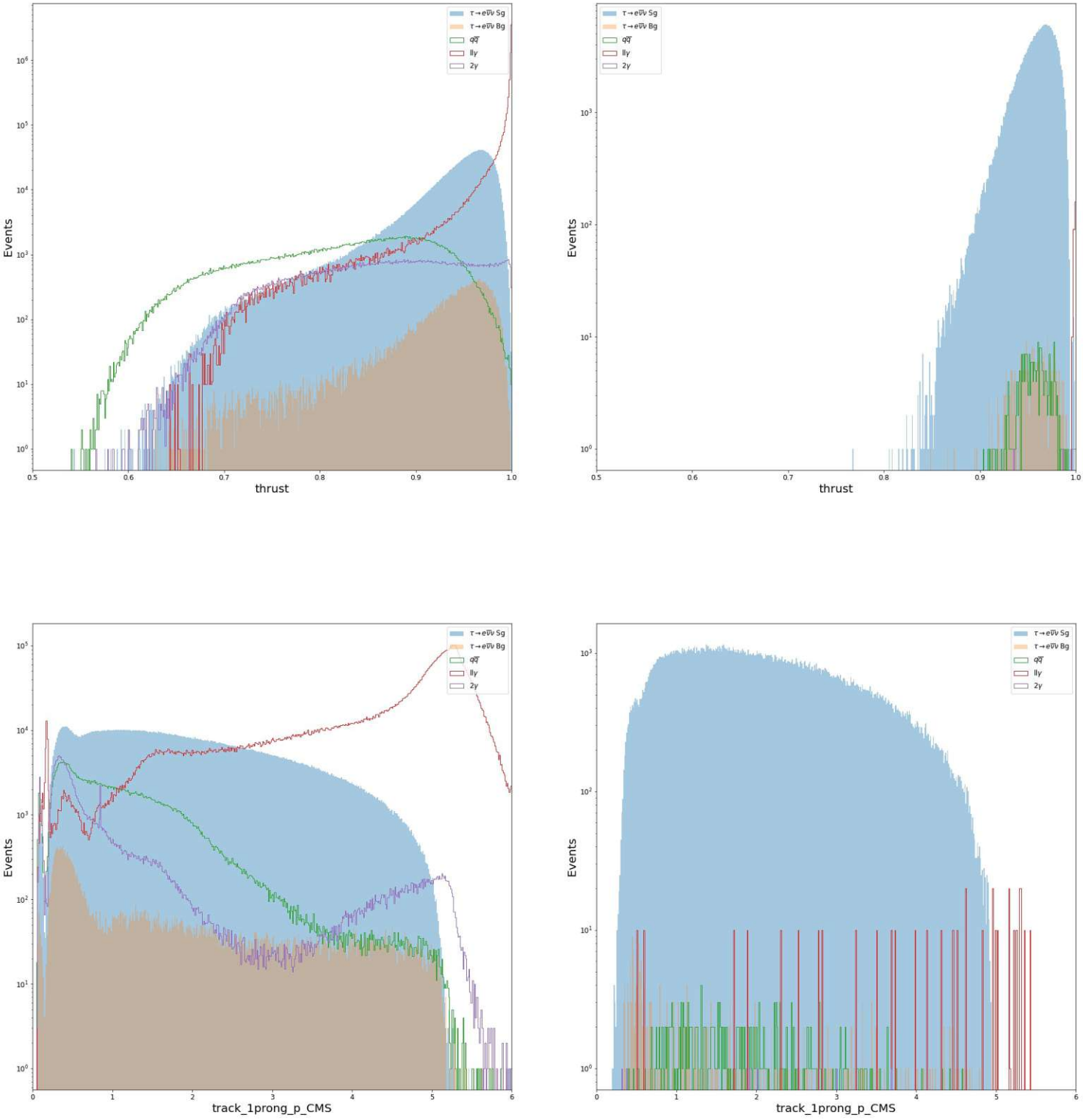


Figure 35: Distribution of the two most important variables during the training in the electron channel. Left are the distributions before the training, right are the distributions after the BDT was applied at a probability level of more than 99.2% which corresponds to a purity of 99.76 %.

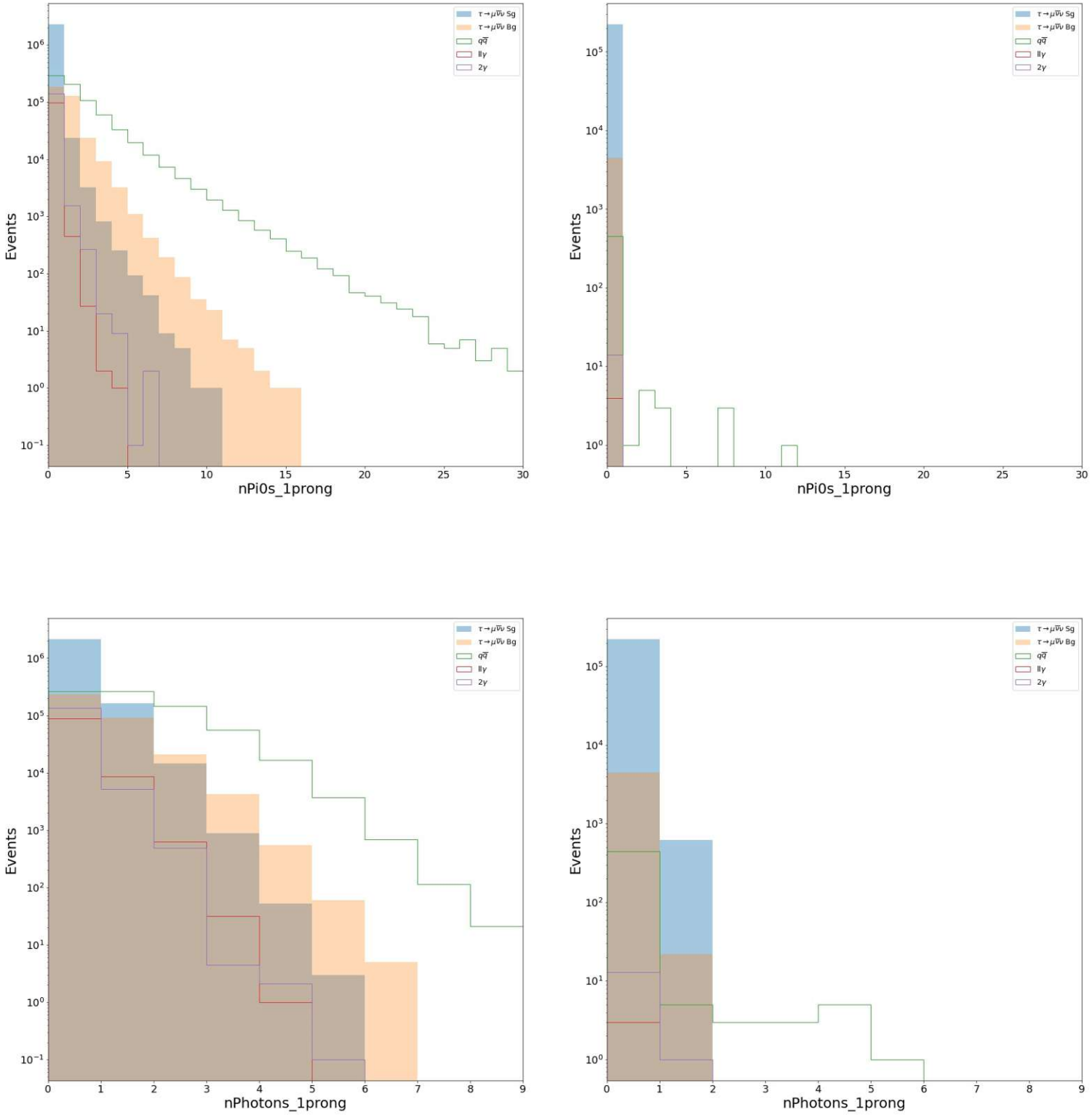


Figure 36: Distribution of the two most important variables during the training in the muon channel. Left are the distributions before the training, right are the distributions after the BDT was applied at a probability level of more than 96.39% which corresponds to a purity of 97.81 %.

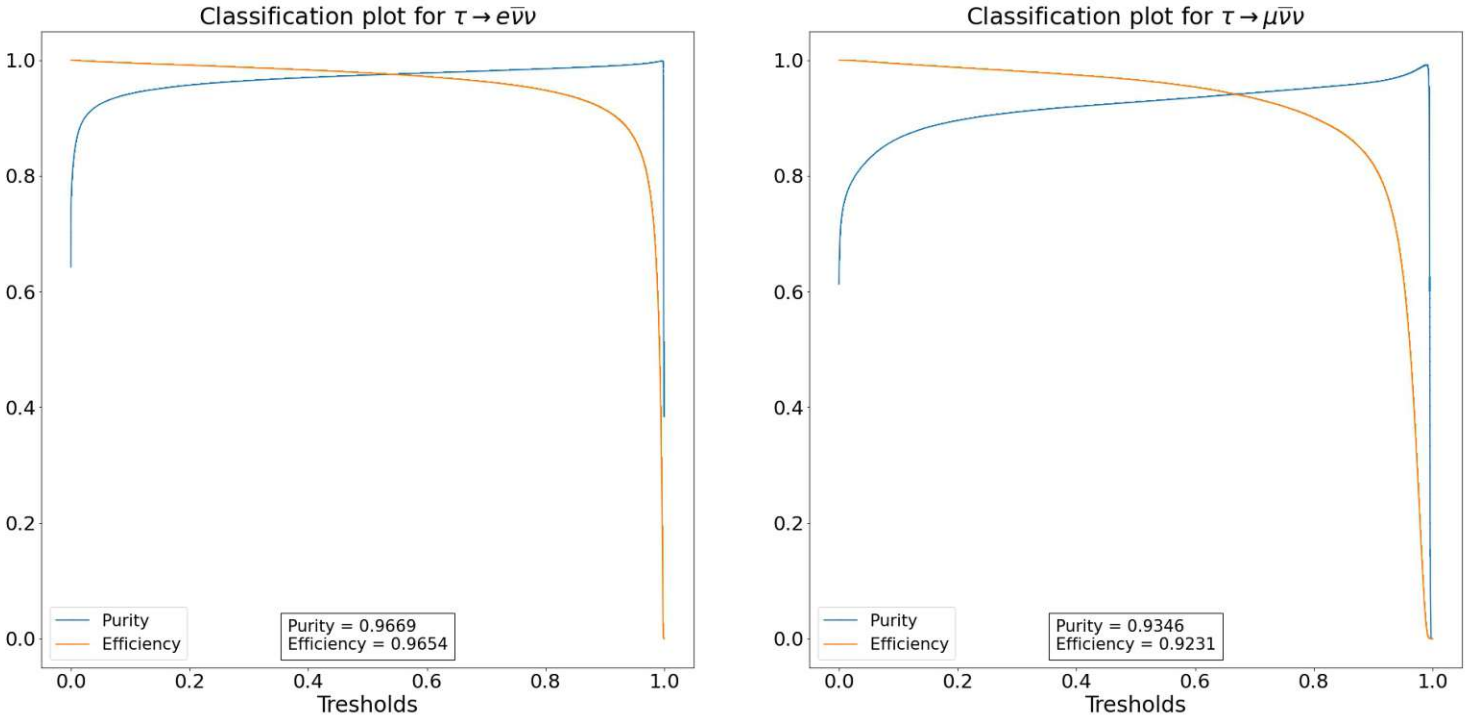


Figure 37: BDT Classification for both channels.

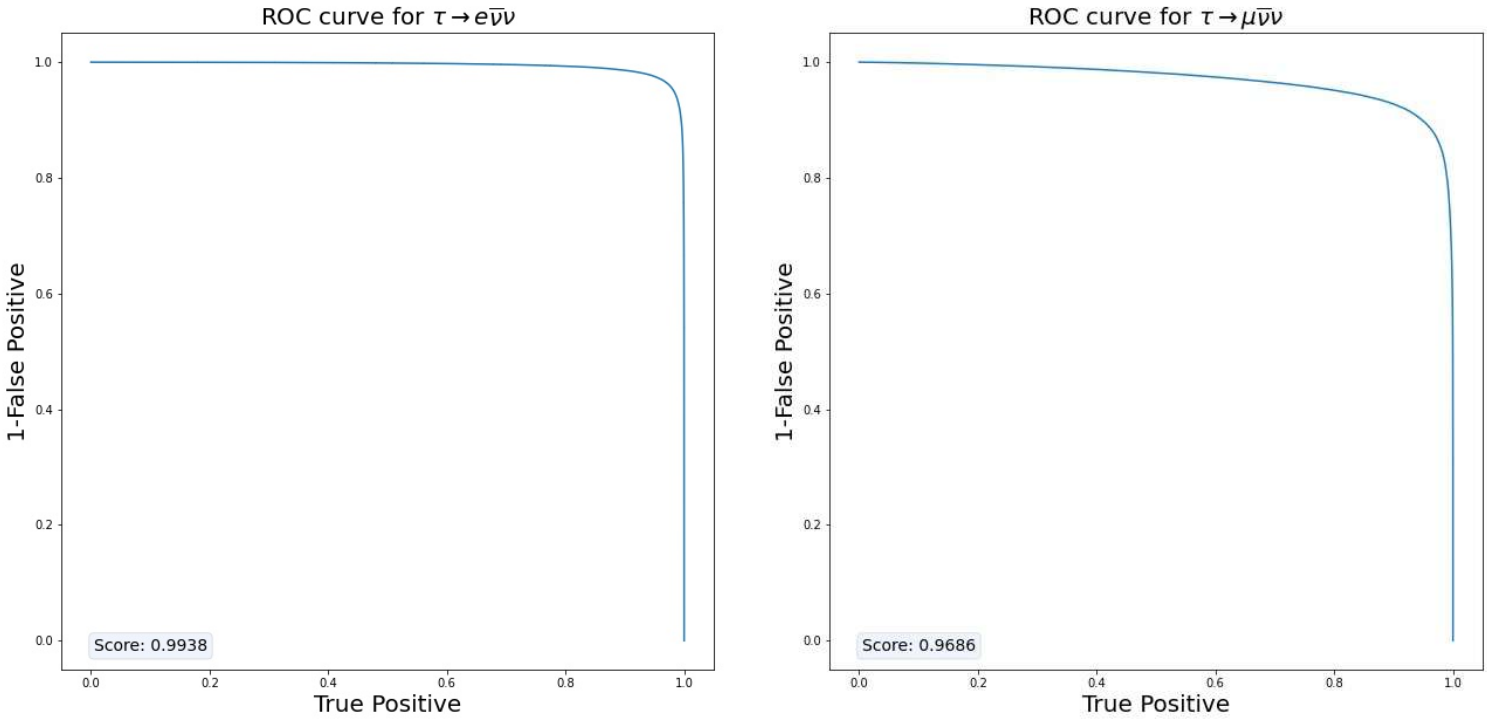


Figure 38: ROC Curve for e- and μ- channel.

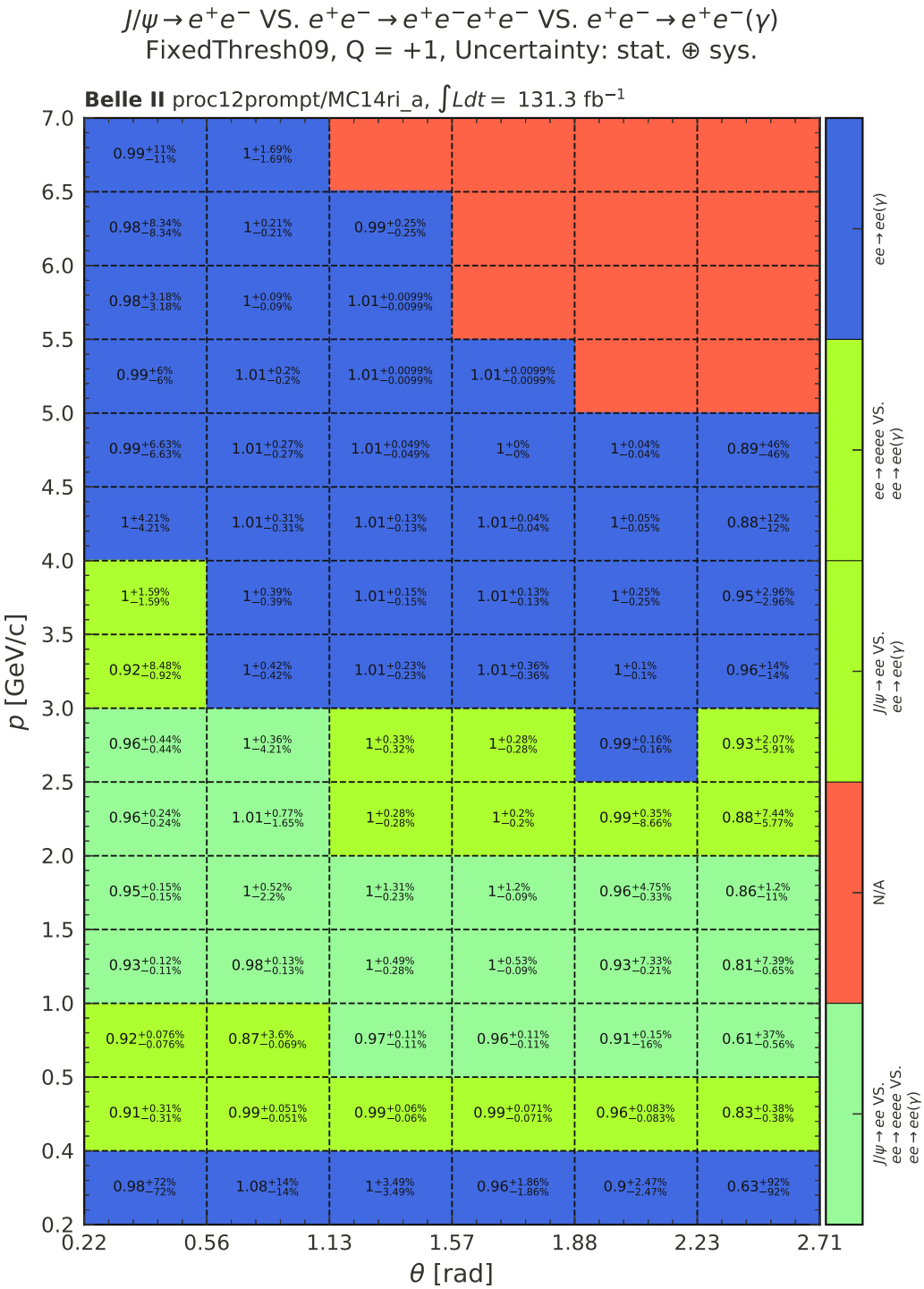


Figure 39: Electron ID, excluding TOP. The error in each bin varies with θ/p.

7 Outlook

The Belle 2 experiment started taking data in 2018 and accomplished a lot since then. In 2022 the detector will be upgraded to include a full, 2-layer PXD, hopefully leading to improved precision. The ongoing experiment will try to find answers for the biggest problems in particle physics:

- finding dark matter
- explain gravity through quantum mechanics
- finding physics beyond the SM

with it's world leading luminosity.
In this work it was shown that machine learning methods can be used for lepton identification. Further we used that knowledge to improve our accuracy in the search for LFU violation. This can ultimately lead to a more accurate measurement of the R_μ value and hopefully together with other improvements to a 5σ measurement of LFU violation. If this happen, we would have successfully found physics beyond the SM, leaving room for theorist to come up with a new model or an extension to the SM. Because of this importance, LFU violation is heavily tested by multiple collaborations (LHC, BaBar) and will hopefully soon reach a precise enough measurement to confirm it.

List of Figures

List of Tables

Bibliography

[1] A. Martini, A. Rostomyan, G. Rauber, G. Inguglia, M. H. Villanueva, P. Feichtinger, P. Rados, and Z. Gruberova, "Search for lepton flavor universality violation in leptonic τ decays with 3x1 topologies," Jun 2021.

[2] W. N. Cottingham and D. A. Greenwood, *The groups of the Standard Model*, 2nd ed. Cambridge University Press, 2007.

[3] Cush. Standard model of elementary particles. https://commons.wikimedia.org/w/index.php?curid=4286964.

[4] TriTertButoxy. Summary of interactions between certain particles described by the standard model. https://commons.wikimedia.org/wiki/File:Elementary_particle_interactions.svg#/media/File:Interactions.png.

[5] B. Aubert, Y. Karyotakis, J. P. Lees, V. Poireau, E. Prencipe, X. Prudent, V. Tisserand, J. Garra Tico, E. Grauges, M. Martinelli, and et.al, "Measurements of charged current lepton universality and $|V_{us}|$ using tau lepton decays to $e^-\overline{\nu}_e\nu_\tau$, $\mu^-\overline{\nu}_\mu\nu_\tau$, $\pi^-\nu_\tau$, and $K^-\nu_\tau$," *Phys. Rev. Lett.*, vol. 105, p. 051602, Jul 2010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.105.051602

[6] A. Martini, A. Rostomyan, G. Rauber, G. Inguglia, M. H. Villanueva, P. Feichtinger, P. Rados, and Z. Gruberova, "Search for lepton flavor universality violation in leptonic τ decays with 3x1 topologies," Jun 2021.

[7] L. collaboration and et.al, "Test of lepton universality in beauty-quark decays," 2021.

[8] J. P. Lees and et.al, "Measurement of branching fractions and rate asymmetries in the rare decays $b \to K^{(*)}\ell^+\ell^-$," *Phys. Rev. D*, vol. 86, p. 032012, Aug 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.86.032012

[9] J. Kozak, *Decision tree and ensemble learning based on ant colony optimization*. Springer, 2019.

[10] A. Elvers. Hyperparameter optimization using grid search. https://commons.wikimedia.org/wiki/File:Hyperparameter_Optimization_using_Grid_Search.svg#metadata.

[11] F. Abudinén, I. Adachi, P. Ahlburg, H. Aihara, N. Akopov, A. Aloisio, F. Ameli, L. Andricek, N. Anh Ky, D. M. Asner, and et al., "Measurement of the integrated luminosity of the phase 2 data of the belle ii experiment," *Chinese Physics C*, vol. 44, no. 2, p. 021001, Feb 2020. [Online]. Available: http://dx.doi.org/10.1088/1674-1137/44/2/021001

[12] Y. Ohnishi and et.al, "Accelerator design at SuperKEKB," *Progress of Theoretical and Experimental Physics*, vol. 2013, no. 3, 03 2013, 03A011. [Online]. Available: https://doi.org/10.1093/ptep/pts083

[13] B. I. web page. Superkekb and belle ii. ,https://www.belle2.org/project/super_kekb_and_belle_ii, visited 2020-12-26.

[14] Y. Amhis, S. Banerjee, E. Ben-Haim, F. U. Bernlochner, M. Bona, A. Bozek, C. Bozzi, J. Brodzicka, M. Chrzaszcz, J. Dingfelder, and et al., "Averages of b-hadron, c-hadron, and

$$\tau$$

-lepton properties as of 2018," *The European Physical Journal C*, vol. 81, no. 3, Mar 2021. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-020-8156-7

[15] K. et. al, "The Belle II Physics Book," *Progress of Theoretical and Experimental Physics*, vol. 2019, no. 12, 12 2019, 123C01. [Online]. Available: https://doi.org/10.1093/ptep/ptz106

[16] A. J. Bevan, B. Golob, T. Mannel, S. Prell, B. D. Yabsley, H. Aihara, F. Anulli, N. Arnaud, T. Aushev, M. Beneke, and et al., "The physics of the b factories," *The European Physical Journal C*, vol. 74, no. 11, Nov 2014. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-014-3026-9

[17] T. A. et al, "Belle ii technical design report," 2010.

[18] K. Akai, K. Furukawa, and H. Koiso, "Superkekb collider," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 907, pp. 188–199, 2018, advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900218309616

[19] T. Browder. Belle ii overview,. https://indico.bnl.gov/event/9843/contributions/42496/attachments/31088/49091/B2_Ops_Review_Oct2020_teb_v8.pdf.

[20] A. Horvath. Cherenkov. https://commons.wikimedia.org/wiki/File:Cherenkov.svg.

[21] R. Pestotnik", ""particle identification devices at the belle ii experiment"," Oct 2015.

[22] A. Paladino, "Beam background at superkekb during phase 2 operations," *Energy [GeV]*, vol. 4, pp. 7–007, 2018.

[23] B. I. confluence page. Total recorded integrated luminosity,. https://confluence.desy.de/display/BI/Belle+II+Luminosity.

[24] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, "The belle ii core software," *Computing and Software for Big Science*, vol. 3, no. 1, Nov 2018. [Online]. Available: http://dx.doi.org/10.1007/s41781-018-0017-9

[25] D. J. Lange, "The evtgen particle decay simulation package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 462, no. 1, pp. 152–155, 2001, bEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900201000894

[26] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to pythia 8.2," *Computer Physics Communications*, vol. 191, p. 159–177, Jun 2015. [Online]. Available: http://dx.doi.org/10.1016/j.cpc.2015.01.024

[27] S. Jadach, B. Ward, and Z. Was, "The precision monte carlo event generator for two-fermion final states in collisions," *Computer Physics Communications*, vol. 130, no. 3, p. 260–325, Aug 2000. [Online]. Available: http://dx.doi.org/10.1016/S0010-4655(00)00048-5

[28] N. Davidson, G. Nanava, T. Przedziński, E. Richter-Was, and Z. Was, "Universal interface of tauola: Technical and physics documentation," *Computer Physics Communications*, vol. 183, no. 3, p. 821–843, Mar 2012. [Online]. Available: http://dx.doi.org/10.1016/j.cpc.2011.12.009

[29] S. A. et al., "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900203013688

[30] R. Tiwary, "Lepton identification using the belle ii silicon-strip vertex detector."

[31] M. Lubej, "Determination of $|v_{ub}|$ at belle ii," 2017.

[32] T. Keck, "Fastbdt: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification," 2016.