UNIVERSITÉ
DE FRANCHE-COMTÉ

CENTRE DE PHYSIQUE DES PARTICULES
DE MARSEILLE
UMR CNRS 7346

MASTER PHYSIQUE & PHYSIQUE NUMÉRIQUE

ANNÉE SCOLAIRE : 2019/2020

Master Thesis

Track selection optimisation for τ studies at Belle II

Robin Leboucher

Encadrant :
Justine Serrano

Vendredi 26 juin 2020

Acknowledgements

So long as I can remember, I have always wanted to understand how the world functioning and the course was for me the meaning to learn more about it. The culmination is to understand of hat the universe is fundamentally constitute. To have received the chance to ending my master in particle physics, I would like to present the biggest thanks to Justine Serrano for accepted and piloted me in the group. Your advice, comments and lessons was very precious. Thanks to providing a healthy atmosphere in your work and in your team. I would also acknowledge Laura Zani and Güney Polat for yours helps and particularly Laura to be patient with me and my apprenticeship of ROOfit. I have also a thought to the τ-group around the world. You have provided a feeling of security and of kindness. Thank to everyone for your attention and your comments during the meetings.

During these last months, I had the chance to evolve in a well environment providing by all the CPPM staff. For the hospitality and for the support during these months, I present my deepest thanks to everyone.

I can't finish my acknowledgements without thinking to my relatives. Sure, thanks Laura because you support and listen me like anyone. You have the superpower to rise my self esteem and give me the courage to continue. And especially thanks to follow me at Marseille and to have waiting me in the internship looking, the journey can't have the same taste without you. To my parents, I have to thank you for the good study conditions that you have send to me. We have a real chance to get you for the moral and material support that you provide to your children.

I have conscience that without all of you, the hardship to go through the 2020 lock-down could became a real nightmare.

Glossary

ARICH Aerogel Ring-Imaging CHerenkov counter.

basf2 Belle II analysis framework.

BSM Beyond the Standard Model.

CDC Central Drift Chamber.

ECL Electromagnetic Calorimeter.

IP Impact Parameter.

KLM K_L and Muon detector.

LFV Lepton Flavour Violation.

MC Monte Carlo.

NP New Physic.

POCA Point Of Closest Approach.

PXD silicon PiXel Detector.

SM Standard Model.

SVD Silicon Vertex Detector.

TOP Time Of Propagation counter.

Contents

Introduction

In particle physics many experiments were able to prove the Standard Model (SM) reliability. The fundamental component of the universe and their interactions are described by gauge theory mechanisms. Since 1954 where C. N. Yang and R. L. Mills laid the foundations with the strong interaction description in [1], the model was extended with the electromagnetism and weak interaction combination by S. Glashow in [2] and by adding Higgs mechanism [3] in electroweak interaction by S. Weinberg in [4]. The SM was verified many times since its elaboration and gives accurate predictions. Nevertheless nowadays the SM shows some deficiencies to explain some crucial questions as the incapacity to explain particles predominance over antiparticles, the missing mass and the accelerating expansion of the universe, and the gravitation. These observations answer the requirement to complete the SM and many theoretical models were already proposed. The goal of the modern particle physics is to search these Beyond the Standard Model (BSM) New Physic (NP) models inside the observation given by the last experimentation.

The search of new physic in particle physics has two different strategies. The first one is to provide highest energy to directly produce new heavy particles, it is called energy frontier. This strategy is adopted by the Large Hadron Collider at the CERN by accelerating protons at 6.5 TeV to produce a collision of 13 TeV. The strategy adopted by SuperKEKB [5], and generally by the B-factories is to indirectly search for NP by making precise measurements with the highest luminosity, it is called intensity frontier.

The search of NP is the primary goal of the Belle II experiment located at KEK in Japan. Belle II consists in a collaboration of a thousand of physicists and engineers from 115 institutions in 26 countries since 2017 around the same detector Belle II. Its associated accelerator SuperKEKB provides electron positron collisions to produce B-meson pairs and other particles. This B-factory aims to produce 50 times more events than its predecessor Belle by providing a luminosity 80 times superior over 8 years. The large amount of electron positron collisions provide a good environment for τ studies and lepton flavour violation research. The *Centre de Physique des Particules* of Marseille joins the collaboration in 2019 to provide a branching ratio measurement of the τ decays, as $\tau \to 3\mu$, which violate the lepton flavour. These decays are forbidden in the SM and an observation would be a clear evidence of new physics.

All analyses involving tau leptons use the same first level of selection based on the track properties. The goal of this note is to provide an new definition of this selection based on the latest simulation.

Section 1 presents an overview of the physic used in Belle II analyses. Section 2 details the tools used during the track selection for the τ group study and finally Section 3 details the results of the selection optimization.

1 Belle II physics overview

1.1 Belle II experiment

1.1.1 SuperKEKB accelerator

Belle II is an upgrade of its predecessor Belle, with a more efficient accelerator, SuperKEKB in Fig. 1.1, more precise detector and new software. These better performances should provide 50 times more data to reach a region of the NP phase space not yet probed.

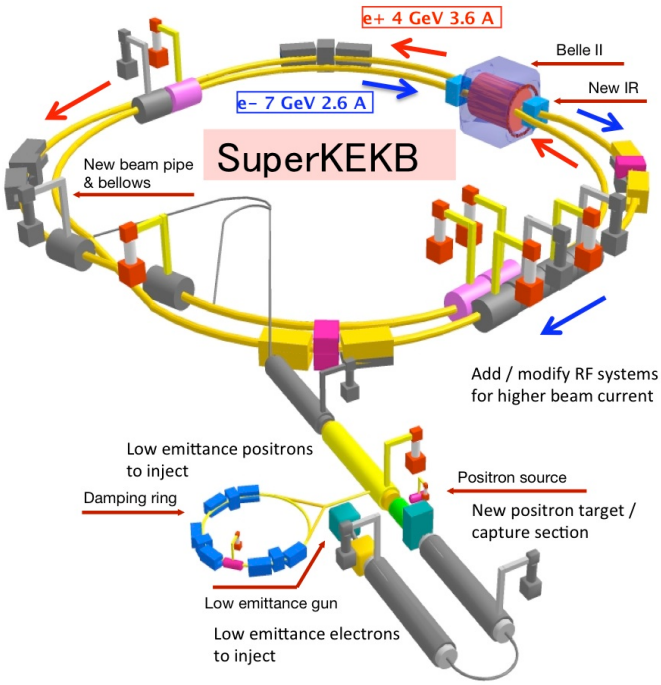

Figure 1.1: Schematic representation of the SuperKEKB accelerator in Japan. It provides asymmetric e^+e^- collision at $\Upsilon(4S)$ to produce B pair mesons. (Source: BelleII Website https://www.belle2.org/project/super_kekb_and_belle_ii)

The B factories is the name given to colliders, which provide large sample of B mesons. Their primary was the study of Charge-Parity violation, Babar and Belle took data from 2000 to 2010. In 2008, after the experiment validation, M. Kobayashi and T. Maskawa was rewarded by the Nobel prize in 2008 for their work on the Cabibbo Kobayashi Maskawa matrix used in Charge-Parity violation understanding. While Belle received an upgrade in Belle II, Babar was stopped. B mesons can be produced in e^+e^- collisions, which provide a clean environment and a well known initial state. Moreover the production of $B\bar{B}$ is amplified by operating at 10.58 GeV, called $\Upsilon(4S)$ resonance which is a bound state of the beauty quark b and its antiquark $\bar{b}$. The accelerator reaches the $\Upsilon(4S)$ energy by an asymmetric acceleration of electrons e^- at 7 GeV and positrons e^+ at 4 GeV [6]. at this energy, the collisions also produce pair of lighter quarks as well as lepton pairs.

The luminosity is an important indicator of accelerator performances. An high luminosity allows to the experience to collect a large amount of data and to observe rare processes. It is proportional to the number of collisions N made in a given time t and to the interaction cross-section σ,

$$L = \frac{1}{\sigma}\frac{\mathrm{d}N}{\mathrm{d}t}. \tag{1.1}$$

The integrated luminosity L_{int} characterizes the amount of data collected during a period:

$$L_{int} = \int L\mathrm{d}t, \tag{1.2}$$

an it is measured in inverse femto barn fb^{-1} with $1\,b = 10^{-24}\,cm^2$. It is expected at SuperKEKB to reach $8 \times 10^{35}\,cm^{-2}\,\mathrm{s}^{-1}$ [7] which is 40 times larger than KEKB recorded peak and 80 times more than KEKB luminosity goal, as visible in Fig. 1.2. In order to ensure this improvement the accelerator was upgraded: the most important improvements have been to reduce the interaction the beam size and increase the beam currents. The first data taking run was undertaken in 2018 with a partial detector. The physics data taking resumed in 2019 with a complete detector.

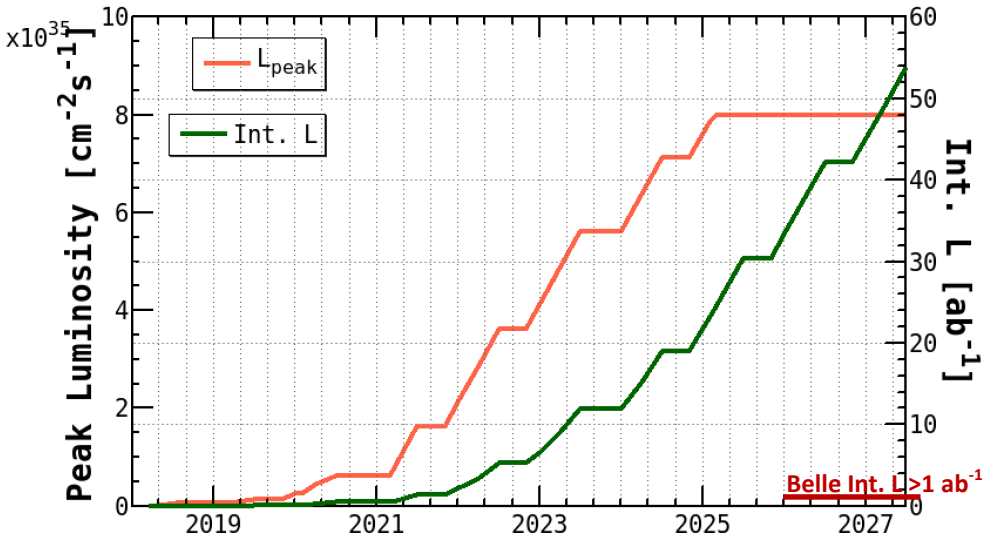

Figure 1.2: Expected instantaneous and integrate luminosity for Belle II and SuperKEKB. Equivalent to the amount of events produced. (Source: BelleII Public Website https://confluence.desy.de/display/BI/Public+GeneralSlides)

1.1.2 Belle II overview

Belle II consists in concentric barrels which supports the different detector layers. It also contains a superconducting solenoid which provides a 1.5 T magnetic field to measure momentum. The new expectations of the accelerator induced to drastically update the detector, indeed, Belle II should work with 40 times more generated events than its predecessor. This requires to improve each components [7, 8].

As visible in Fig. 1.3, the detector is built in different layers to measure different properties of a track without altering the next measure. The first layer is composed of the silicon PiXel Detector

(PXD) and the Silicon Vertex Detector (SVD) which form together the vertex detector. They are together composed of 6 concentric layers of silicon sensors located around the interaction region, where particles collide, which provide detection of very low momentum particles, the charged track reconstruction and the momentum measurements [9].

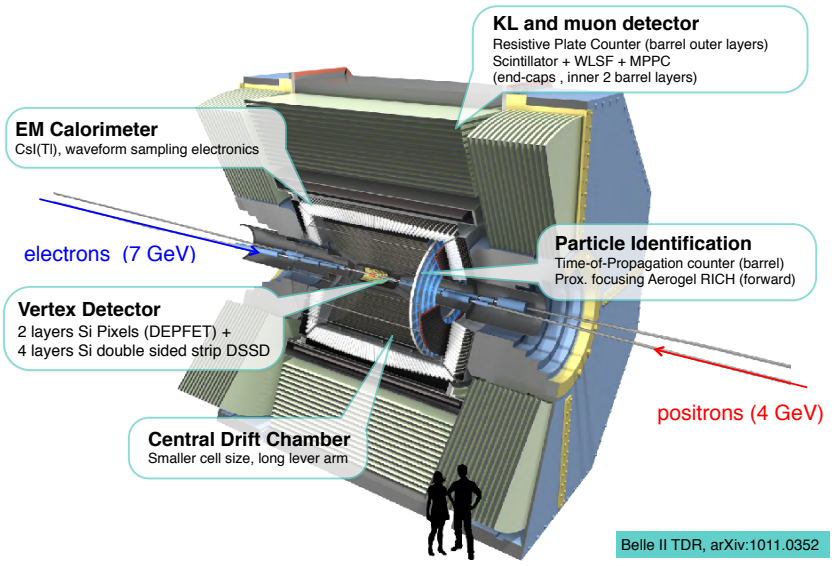

Figure 1.3: Belle II detector open 3D view (Source: BelleII Public Website https://confluence.desy.de/display/BI/Public+GeneralSlides)

After going through the vertex detector, the particles arrive on the Central Drift Chamber (CDC). The CDC extends the function of the PXD and the SVD by providing the track reconstruction and momentum measurement. But it also provide particle identifications by loss energy measurement. The third aim is to provide a signal trigger for charged particle events, in order to identify interesting events to be stored during data taking. It is composed of 56 layers of chambers filled with gas mixture and tungsten sense wires placed into an electric field. When a charged particle passes through the chamber electrons are emitted in cascade. The resulting electric pulse is detected by the wires. The fin wire mesh allows the precise particle location measurement and the redrawing of its trajectory.

The next detectors are the particle identification devices with the Time Of Propagation counter (TOP) and the Aerogel Ring-Imaging CHerenkov counter (ARICH). The TOP is composed of 16 modules of quartz which surrounds the CDC. When a charged particle crosses a quartz bar, Cherenkov photons are emitted, reflected in the bar and finally detected in function of the reflecting angle. From the time propagation of photons measurement, the Cherenkov angle and the velocity are reconstructed. These quantities allow to calculate a likelihood for different mass hypothesis used in the particle identification. The ARICH detector uses the same process of Cherenkov photon detection in an aerogel, after crossing an expansion volume the photon are singly detected on photon detectors.

All of the previous layers enable to detect charged particles but for the case of neutral ones as photons the adopted strategy is to used a calorimeter. Belle II made the choice to reuse the Electromagnetic Calorimeter (ECL) used by Belle, upgrading the electronics to cope with the higher background rate. It is composed on $CsT(Ti)$ crystals where neutral particles are subject to a cascade of collisions which produces lower energy particles. They form a particle shower and ends when all initial energy is dissipated. The particle energy is measured by detecting the

produced light quantity. In addition the Belle II ECL is also able to ensure: electron identification, luminosity measurement and signal trigger.

The K_L and Muon detector (KLM) detector is the outermost and the biggest layer, made of a sandwich of thick iron plates and active materials detectors. In the KLM only long-kaon K_L^0 and muon μ lives, other particles has already interact with the detector in the previous layers. The K_L^0 interact in particle shower which allow the energy measurements. The strategy for muon is to verify its trajectory with respect to an extrapolation from inner detectors.

1.2 The Standard Model

1.2.1 Standard Model

The SM is still the best theoretical model to describe the fundamental bricks of the matter and the fundamental interactions as the electromagnetism, weak and strong interactions [1, 2, 3, 4]. Two kinds of particles appear in the SM, they are differentiated according to their spin. A spin $1/2$ particle corresponds to a fermion which could be seen as the elementary brick of the matter. The second family named bosons gathers the particles of spin 1 and operate as fundamental interaction mediators.

The SM follows a formalism of a unified quantum field theory obeying the gauge group symmetry $SU(3)_c \times SU(2)_L \times U(1)_Y$ with c the color charge, L the chiral component and Y the hypercharge [9]. This model is divided in two parts, a unification of electromagnetic and weak interactions into the electroweak theory and a strong interaction description with the theory of quantum chromodynamics. The electroweak theory which contains the symmetry group $SU(2) \otimes U(1)$.

The SM mathematical description could be summarized under the Langrangian $\mathcal{L}_{SM}$:

$$\mathcal{L}_{SM} = \mathcal{L}_{kin} + \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yuk}, \tag{1.3}$$

where $\mathcal{L}_{kin}$ is associated to the kinematic part of gauge theory and the self-interaction, $\mathcal{L}_{EW}$ models the electroweak theory, $\mathcal{L}_{QCD}$ models the quantum chromodynamics. At these contributions is added the $\mathcal{L}_{Higgs} + \mathcal{L}_{Yuk}$ term which allows to define the Higgs mechanism and the interaction of fermions and leptons with this field. These processes result in the definition of fermions and leptons masses by the Higgs boson.

1.2.2 The bosons

According to the SM, the different matter particles interact with each other through the exchange of virtual bosons. Bosons are viewed as force-carrying particles coming from associated gauge fields excitation. The elementary bosons known in the SM are:

- the photon γ for the electromagnetism,
- the $W^{\pm}$ and Z bosons for the weak interaction,
- the gluon g for the strong interaction.

At this list, the Higgs boson (h), with 0 spin, must be added to define a mass for $W^{\pm}$, Z bosons and fermions by a symmetry breaking.

1.2.3 The fermions

In the other side of the SM reside spin $1/2$ particles called fermions which are the fundamental matter particles. Two kinds of fermions could be differentiated:

- the leptons which interact weakly and electromagnetically if they are charged,

- and the quarks which interact electromagnetically, weakly and strongly.

As seen in Fig. 1.4, leptons and quarks are classified following two properties: the electronic charge and the mass. Leptons could be charged $-e$, or neutral for the neutrinos associated to charged leptons. All of the six quarks present a fractionary electric charge $\frac{2}{3}e$ or $\frac{-1}{3}e$ for respectively *up*- and *down*- type.

An additional charge related to quantum chromodynamic, named color, is added in the quark description. The color can take tree different value (red, green and blue). The strong interaction allows to combine quarks to form bound states called hadrons by the exchange of gluon carrier of the color. In contrast with the electromagnetism, the strong interaction increases with the distance between two quarks. This phenomenon is responsible of the absence of free state quark. The most remarkable hadrons are the proton and neutron respectively composed on two up, one down quark and two down, one up, since they constitute the ordinary matter around us.

Finally, fermions can be classified in three mass families each constructed with one *up*- and one *down*- type quarks, one charged lepton and one neutral lepton. At the fermions and also bosons, it is added their electric charge complementary known as antiparticles [10, 11], some neutral particles can be their own antiparticles such as photon and Z^0 bosons.

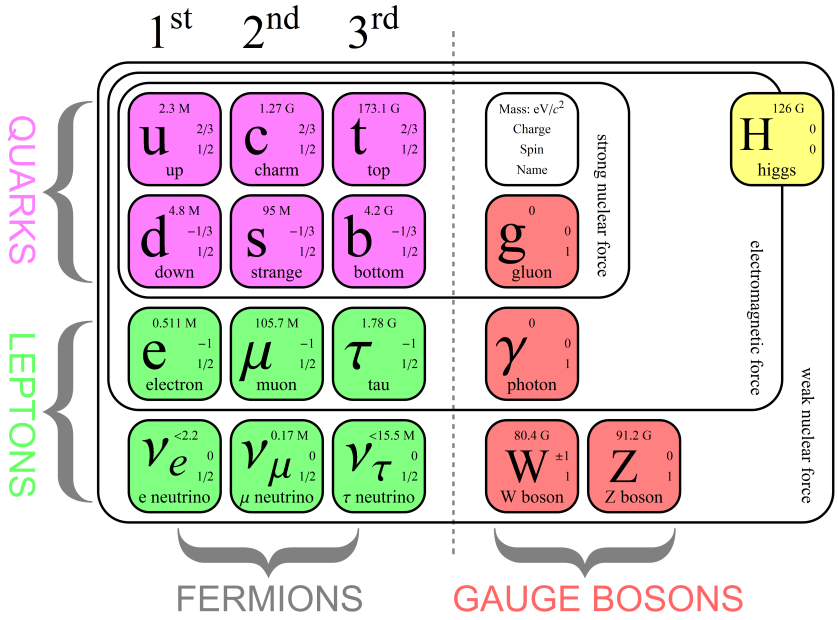

Figure 1.4: Representation of Standard Model classification of particles. (Source: Matic Lubej https://www.physik.uzh.ch/groups/serra/StandardModel.html)

1.2.4 Standard Model limitations

The SM is a really efficient theory to describe the matter and the interactions but it is unable to answer to some fundamental questions. The first regards the description of the fourth fundamental interaction, the gravitation. Indeed the gravitation is well described by the theory of general relativity but this theory is by nature incompatible with the quantum mechanic. Secondly it does not explain origin of dark matter and dark energy which constitute 96% of the Universe. Moreover the big bang should have created the same amount of matter and antimatter but our world is only made of matter. This matter/antimatter asymmetry is also not solved by the SM. For all these reasons, and even more, the SM seems to be an incomplete theory.

1.3 τ physic and Lepton Flavour Violation

1.3.1 Lepton Flavour

The SM reveals unexpected symmetries in observed decays. In order to explain these additional symmetries, extra quantum numbers are inserted in the model. In the case of fermions the numbers are: weak isospin, baryon number B and lepton number $L = 1$ for leptons, -1 for antileptons and 0 otherwise. In refinement of the lepton number, the lepton family numbers was defined:

- L_e the electron number, equal to 1 for electron and electron neutrino, -1 for positron and anti-electron neutrino and 0 otherwise,
- L_μ the muon number, equal to 1 for muon and muon neutrino, -1 for their antiparticles and 0 otherwise,
- L_τ the tau number, equal to 1 for tau and tau neutrino, -1 for their antiparticles and 0 otherwise.

These three numbers allow to define the three lepton flavours and should be separately conserved for electromagnetic and strong interactions.

For example:

$$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau, \tag{1.4}$$

is possible in the SM because in the left side the tau gives:

$$\begin{aligned} L^{left} &= L(\tau^-) = 1, \\ L_\tau^{left} &= L_\tau(\tau^-) = 1, \\ L_\mu^{left} &= L_\mu(\tau^-) = 0, \end{aligned}$$

when the right side presents:

$$\begin{aligned} L^{right} &= L(\mu^-) + L(\overline{\nu}_\mu) + L(\nu_\tau) = 1 - 1 + 1 = 1, \\ L_\tau^{right} &= L_\tau(\mu^-) + L_\tau(\overline{\nu}_\mu) + L_\tau(\nu_\tau) = 0 + 0 + 1 = 1, \\ L_\mu^{right} &= L_\mu(\mu^-) + L_\mu(\overline{\nu}_\mu) + L_\mu(\nu_\tau) = 1 - 1 + 0 = 0. \end{aligned}$$

$L_\tau^{left} = L_\tau^{right}$, $L_\mu^{left} = L_\mu^{right}$ and global lepton number $L^{left} = L^{right}$ are well conserved by the decay, this interactions is allowed in SM. Nevertheless the studied decay:

$$\tau^- \to \mu^- \mu^+ \mu^-, \tag{1.5}$$

gives:

$$\begin{aligned} L^{left} &= L(\tau^-) = 1, \\ L_\tau^{left} &= L_\tau(\tau^-) = 1, \\ L_\mu^{left} &= L_\mu(\tau^-) = 0, \end{aligned}$$

when the right side gives:

$$\begin{aligned} L^{right} &= L(\mu^-) + L(\mu^+) + L(\mu^-) = 1 - 1 + 1 = 1, \\ L_\tau^{right} &= L_\tau(\mu^-) + L_\tau(\mu^+) + L_\tau(\mu^-) = 0, \\ L_\mu^{right} &= L_\mu(\mu^-) + L_\mu(\mu^+) + L_\mu(\mu^-) = 1 - 1 + 1 = 1. \end{aligned}$$

$L_\tau^{left} \neq L_\tau^{right}$, $L_\mu^{left} \neq L_\mu^{right}$, the decay violate the lepton flavour, because lepton flavour are not conserved, although the global lepton number is conserved, $L^{left} = L^{right}$. This decay can not be described by the SM but some BSMs predict a branching ratio, equal to the probability of this particular τ decay to happen, reachable by new generation detector.

1.3.2 Motivation

Are there sources of lepton flavour violation beyond the SM? it is one the question that the τ-group of Belle II collaboration tries answering [7]. Precisely Belle II search for Lepton Flavour Violation (LFV) decays such as $\tau \to \mu\gamma$ or $\tau \to 3\mu$ at a branching ratio, which is the probability of this particular τ decay to happen, reachable, 10^{-8}. The fact that the neutrinos are observed to be massive and to oscillate, changing from one family to the other, can imply a branching fraction of the order of 10^{-54}. Nevertheless BSM predicts LFV decays reachable by Belle II like: supersymmetric standard models, little Higgs models, low-scale seesaw models, leptoquark models, Z' models and extended Higgs models. Many other studies, like characterizing the lepton tau or looking many other decays, are provided in the τ-group.

To verify these predictions the study of tau leptons presents advantages compared to μ, the bigger mass of τ allows it to decay into many final states including muons and electrons, while muons can only decays into gamma and electrons. A second advantage is the fact that other channels with final state containing light mesons allows to test the LFV coupling between quarks and leptons. But the golden channel is $\tau \to 3\mu$ because it has a pure leptonic final state and less background. In these channels it is expected that Belle II overcome all previous studies with a better 90% confidence level upper limit as seen in Fig. 1.5 [7].

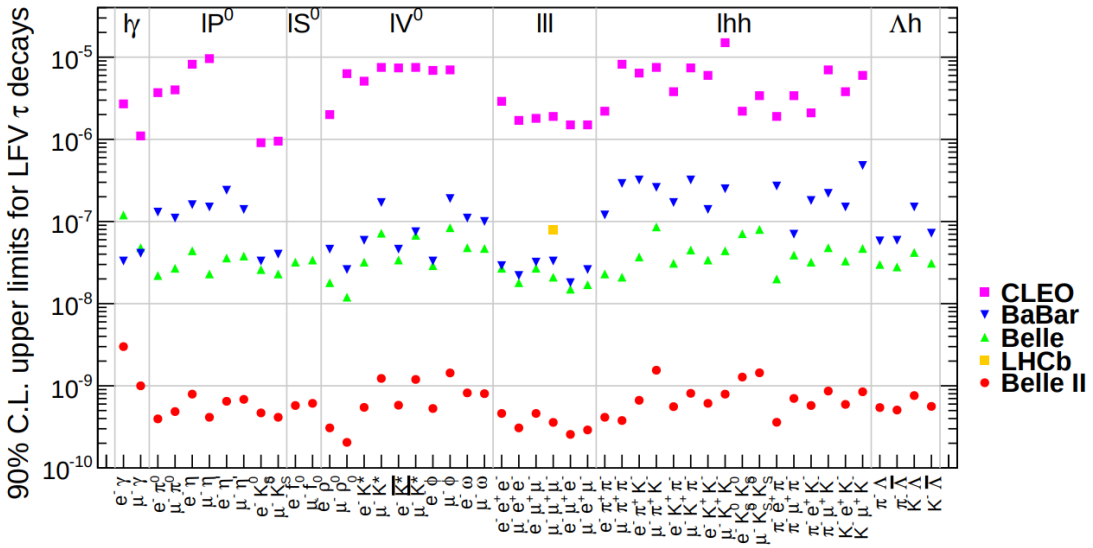

Figure 1.5: 90% confidence level upper limit of current experiments for the τ LFV decay branching ratio for CLEO, BaBar, Belle and LHCb. The Belle II expectations are added by extrapolating Belle results [12] on 50 ab^{-1}. (Source: Belle II physic book [7])

2 Method and tools

2.1 Goal and method

The purpose of this study is to provide a low-level track selection common to all τ analysis by updating the previous one done by F. Tenchini [13]. This requires to well define signal/backgrounds and characterize their Impact Parameter (IP) distributions. The signal could be defined as tracks which present a tau among its ancestors, no matter at which generation. Moreover the tracks coming from K_s^0 are removed from this study, their particularities need a specific selection. Background is then defined to be not come from tau neither K_s^0.

The Monte Carlo (MC) samples used in this study allow to get the truth PDG ID (called in Belle II analysis framework (basf2) *mcPDG*), unique particle identification number given by the Particle Data Group, of the reconstructed particles. These information allows to check the ancestors of a given tracks and to apply a selection.

The second aim is to choose a selection based on IP variables, in Fig. 2.1, that maximize the signal efficiency while keeping low background level. The charged particles, tracks, produced by a collision, describe helical trajectories due to the magnetic field provided in the detector. After reconstruct the helical track trajectory, the software is able to find the Point Of Closest Approach (POCA): the track position coordinates are set to the perigee, i.e. the POCA of the helix to the interaction point, theoretical point of collision [14]. This point could be associated to the point where the track is produced.

The study focused on:

- dz is the distance in respect to the z axis between the POCA and the interaction point,
- dr represents the radial distance.

A track which come from a e^+e^- collision should have a POCA in the area where e^- and e^+ beam overlapping, otherwise the track come from a noisy interaction or a miss reconstruction.

Other variables are probed to decrease background level, and particularly the momentum variables, p_T and θ.

- θ is the angle between particle momentum and beam direction,
- p_T represents the projection of momentum particle on the orthogonal beam axis, $p_T = \left\| \overrightarrow{p} \right\| sin(\theta)$.

2.2 Simulation

2.2.1 Simulation framework

In addition to data taking, it is necessary to produce simulated data. These simulations allow to perform detectors calibration, software updates, performance studies and are also useful to understand the signal and background characteristics. These simulations tools are included in the basf2.

The simulated data are obtained with a MC event generator which provides the primary physics process for electron positron collisions. To follow the various physic studies provided in Belle II, different generators are used. Most of them use the EvtGen one [16], principally for B and

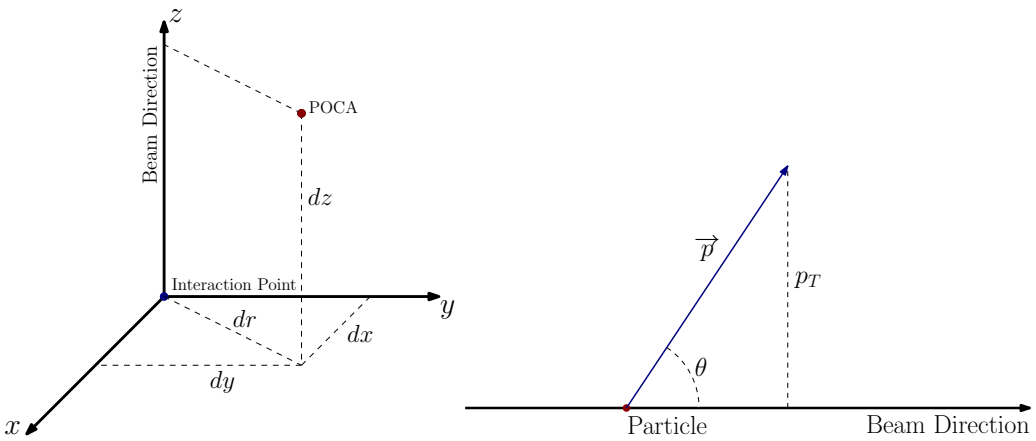

Figure 2.1: **Left scheme** represent the IP variables between the POCA and the interaction point. The cylindrical symmetry of the beam pipe implies to represent the distance between these points in cylindrical coordinates, with the z axis confounded with the beam direction.
Right scheme represent the momentum variables of a tracks. The momentum of a particle is characterize by the angle made with the beam pipe and the traverse norm.

Decay channel	Branching fraction (%)
$\tau^- \to e^- \overline{\nu}_e \nu_\tau$	17.82 ± 0.04
$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$	17.39 ± 0.04
$\tau^- \to \pi^- \nu_\tau$	10.82 ± 0.05
$\tau^- \to \rho^- \nu_\tau \to \pi^- \pi^0 \nu_\tau$	25.49 ± 0.09
$\tau^- \to a_1^- \nu_\tau \to \pi^- \pi^+ \pi^- \nu_\tau$	9.31 ± 0.05
$\tau^- \to K^- \nu_\tau$	$(6.96 \pm 0.10) \times 10^{-3}$
$\tau^- \to K^{*-} \nu_\tau \to K^- \pi^0 \nu_\tau$	$(4.33 \pm 0.15) \times 10^{-3}$
$\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$	4.62 ± 0.05
$\tau^- \to \pi^- \pi^0 \pi^0 \pi^0 \nu_\tau$	1.04 ± 0.07

Table 2.1: Main tau decays provided by simulations with their associated branching ratio [15].

D meson decays. However in our case, τ pair production is provided by KKMC generator [17] according to decays handled by TAUOLA [18], in Tab. 2.1. Additional specific simulations can be requested by physicists, to study for example tau LFV decays. In our study, all tau decays are taking into account, without looking the final state decay, to be able to provide a standard track selection for every τ analysis.

These generators encode the SM physics. The second simulation step is to interface the resulting tracks with the detector simulation Belle II based on Geant4 [19]. The particles produced by the event generator should go through the detector and the magnetic field. This is provided by a detector and magnetic field simulation, where particles quantities like hits in tracking system or energy deposition in the calorimeter are simulated. This step results in a similar data architecture between simulated and real ones.

Finally it is possible to add an independent background simulation. The different background contributions are independently simulated in the dedicated accelerator group software. The most important ones are:

- Radiative Bhabha process, is the e^+e^- scattering by virtual photon emission,
- Touschek scattering, which is an intrabunch effect where Coulomb scattering of two particles in the same beam bunch modify the particle energy,

- Beam-gas scattering, induced by scattering between particles and residual gas molecules,
- Synchrotron radiation, corresponding to the emission of photon by a charged particle inside a magnetic field,
- Two photon process, where low momentum e^+e^- are produced by two-photon process.

At this step, the data files contains only detector information, no physical quantities are yet extracted from these information. This architecture implies that data take a large amount of storage space and can't be instantly used by analysts but the modularity is kept.

2.2.2 Used simulation campaigns

The simulations are officially done as campaigns in function of the software progress, each major release gives a new MC production campaign. During the track selection study, the 12th and 13th MC campaigns are used following the different dataset:

- MC12d taupair with pseudo run-dependent background ($10\,\mathrm{fb}^{-1}$),
- MC13a bgx0 taupair without beam background ($100\,\mathrm{fb}^{-1}$),
- MC13a bgx1 taupair with run independent background ($100\,\mathrm{fb}^{-1}$),
- MC13b taupair with pseudo run-dependent background ($10\,\mathrm{fb}^{-1}$).

These different datasets allow to compare in the study the background sources and to update the previous results based only on MC12 campaign. By default, the MC sample used is MC13b, unless explicitly stated. The whole study is performed with the basf2 release 04-01-05.

In each campaign, data are stocked following different samples in function of the studied physics(B, C, τ and so on). The use of official MC campaigns allow to minimise introducing bias in studies. This is why the simulation campaigns give raw data which required a reconstruction of the decays products, in function of the personnal requirements.

2.3 Reconstruction

2.3.1 Reconstruction overview

Simulated and real raw data are stored in the collaboration servers under the form of detector information like track hits and calorimeter clusters. One of the basf2 task is to transform these raw data into physic quantities by constructing high-level objects as charged tracks. The useless raw information is then discarded to reduce event size by a factor of 40. During this process, made before each physic analysis or study, some preselection requirements could be assigned on tracks for a preliminary background decontamination. The resulting high-level objects are then used to determine particle and event information such as the momentum, the energy and so on.

The reconstruction is done in further step:

- The tracking is the reconstruction of charged particles from the primary and secondary decay vertices. The PXD, SVD and CDC hits of a given charged particle are identified from background hits and then the particle trajectory is fitted from hit position.
- The calorimeter reconstruction is used to determine energy and position of charged and neutral tracks, and missing energy in decays involving neutrinos from the ECL information. The ECL is also used for particle identification in case of electrons, muons, charged hadrons, neutral hadrons and photons by interpreting the shower shapes.

- Charged particle identification is made by defining a likelihood on mass particles from TOP and ARICH information.
- Neutral particle identification depends on the kind of particle. Photons and neutral pions can be identified by checking the shower shapes in ECL. Finally K_L^0 mesons are identified by KLM and ECL detectors using classifier which uses all of the available variables.

2.3.2 Used Reconstruction parameters

Tracks are reconstructed in $e^+e^- \to \tau^+\tau^-$ MC sample using a loose selection on the impact parameter (IP) variables:

- $|dz| < 20\,cm$ and $dr < 20\,cm$, for tracks,
- $E > 0.2\,\text{GeV}$ and $clusterNHits > 1.5$ and $-0.8660 < cos\theta < 0.9563$, for photons.

The created ntuple is mainly filled with:

- impact parameters like dz and dr,
- momentum p and transverse momentum p_T,
- number of SVD and CDC hits,
- Monte-Carlo particle and ancestor truth identification.

They are then classified according to their origin by exploiting information from the Monte Carlo Truth (tau, material interaction, K_s^0, or background). The signal efficiency and background rejection are then computed for different cuts on the IP variables. Further variables are also studied in order to improve the background rejection while keeping a high signal efficiency.

2.4 Analysis tools

2.4.1 Measurement of the purity

The different datasets are composed of several kind of tracks. In order to reduce this contamination by selecting tracks on IP parameters, it is required to have a tool to measure the contamination and also the rate of discarded signal and background tracks.

The signal efficiency E of a set of selection is the ratio between the number of signal tracks which verify the selection s and the total number of signal tracks s_{tot}:

$$E = \frac{s}{s_{tot}}. \tag{2.1}$$

This information is the marker of the impact on the signal of the selection. The goal is to keep it as high as possible.

The background rejection R consists on the fraction of rejected background tracks, which is the complement of the kept background tracks b over the total number of background tracks b_{tot}:

$$R = 1 - \frac{b}{b_{tot}}. \tag{2.2}$$

The rejection is used to measure the number of discarded background tracks. A good selection presents a high efficiency coupled with a high rejection.

But these quantities are not enough to completely measure the effectiveness of a selection. This is why the purity is added to the tools. The purity P consists on a quantification of the contamination of dataset. It is calculated by:

$$P = \frac{s}{s+b}. \tag{2.3}$$

The aim of the study is to maximise these three quantities which are calculated for each value of the selection. The purity is not sufficient because a selection which keeps a poor number of signal tracks and reject all the background gives a high result.

2.4.2 ROOT Software

The particle physics domain was rapidly confronted to the processing of a large amount of data. This constraint forced the different collaborations to numerically process data. At the CERN, the solution was to used a Fortran library. But in the mid 1990's in view to the Large Hadron Collider challenge, the CERN had to change their numeric paradigm with the birth of ROOT framework.

ROOT is an oriented object framework provided in C++ and more recently implemented in Python due to its success. The data are stocked inside object constructed such as trees, the trees allow to get a specific attribute faster [20]. ROOT has a real success in physics data analysis from the particle physics to the astrophysics. Indeed ROOT performs in data processing and fitting.

The Python language has gained in popularity in the last 10 years, with its user-friendly approach and high performances reached by Scipy module, Python appears to be a good complement to ROOT. The PyROOT library allows this gateway by enabling the use of ROOT features, such as data processing, fitting and so on, in Python scripts. Nevertheless PyROOT does not allow to directly treat data with other Python modules. In this way I used the small library root_pandas which read a ROOT trees into pandas DataFrame or a DataFrame to a tree, related to effective Python dictionaries. Using pandas allow me to entirely treat the data with Python and its usual libraries like pandas, scipy or matplotlib.

3 Good track selection for tau events

3.1 Signal and background characterization

3.1.1 Signal and background definition

Before probing the different impact parameter variables and searching the best set of cuts, a characterization of the signal and background tracks is required. A track is identified as signal if it has a tau among its ancestors (no matter at which generation). Tracks from K_s^0, expected to be displaced from the IP, are excluded from this study since they require a specific selection. Some signal tracks are found to be matched to a proton or a pion while having a pion as a mother. They form bumps in the dz region around ±6cm and potentially correspond to material interactions, and should be removed from signal. They represent about 3.5% of the total signal tracks. Tracks which are not signal, nor K_s^0, nor material interaction are identified as background.

[%]	MC12d	MC13a bgx0	MC13a bgx1	MC13b
Total	2301304	2149345	2245224	2550725
Signal	92.09	92.99	88.74	87.49
Material	3.33	3.61	3.26	3.17
Bkg total	3.65	2.42	7.10	8.46
$mcPDG = 0$	3.42	2.17	6.87	8.24
$mcPDG \neq 0$	0.23	0.25	0.23	0.22
K_S^0	0.93	0.99	0.90	0.88

Table 3.1: Total number of reconstructed tracks for each sample and their corresponding fraction of signal cleaned from material tracks, material interaction, background and K_S^0 obtained from MC truth matching. A zero is assigned to *mcPDG* when tracks are not matched, *i.e.* when ancestor is unknown.

The composition of the different MC samples is shown in Table 3.1. An additional distinction is made for background tracks according to their MC truth matching, this will be further discussed in the next section. Along the different datasets the fraction of the material interaction, the K_S^0, and the matched background tracks does not vary. Only the not truth matched ones increase with MC13a bgx1 and MC13b. In an other hand, the dataset composition are similar between MC12d and MC13a bgx0, while MC13a bgx1 and MC13b presents an higher background rate. The signal/background composition of the recent MC13b differ from MC12d used in the previous study and imply the requirement to improve our selection.

The IP distributions of the different component are shown in Fig. 3.1. The material interaction tracks present a bump around $\pm 6\,cm$, which justify a tightest cut than $|dz| < 6$ to remove them. Tighter selections on dz drastically improve also the background rejection, especially for what concerns the beam background contribution which is known to dominate at smaller radius closer to the beam pipe.

Fig. 3.2 shows a zoom in the low theta distributions for the different datasets[1]. The background distribution shows spikes which appear mainly in the run-dependent datasets. According to the

[1] By default in this note, histograms are normalized to the same area

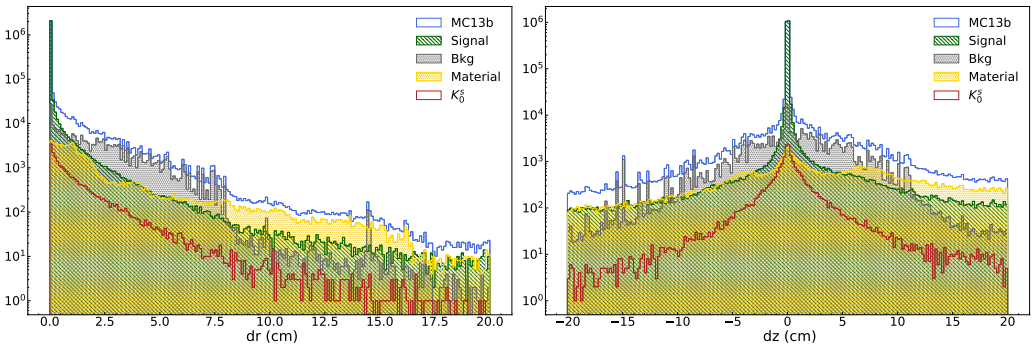

Figure 3.1: Impact parameter distributions for MC13b taupair tracks.

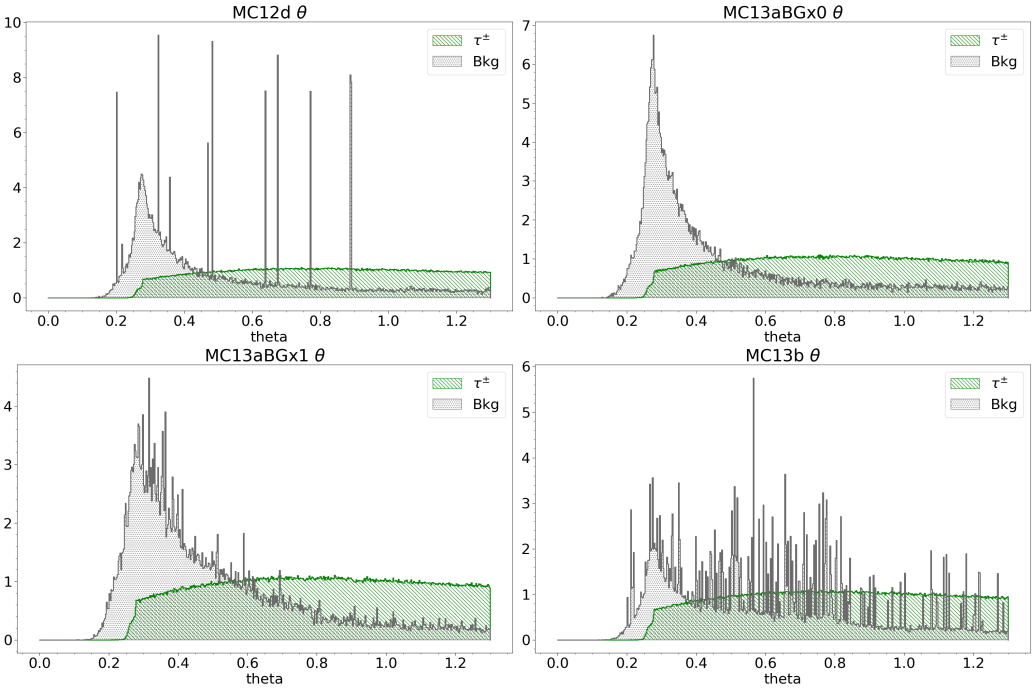

Figure 3.2: θ distribution of the different samples. Spikes appear in the MC13b background.

dataproduction group, these spikes come from background tracks generated from low statistics runs. Future simulations will use higher statistics runs and should not have these artifacts anymore.

3.1.2 Background composition

The background is composed of three main sources: beam background track, which are expected not to be matched at the MC truth level, fake track and electrons or positrons from photon conversion. The origin of background tracks is shown in Table 3.2, which lists the PDG ID of the track and of its four ancestors. In addition, the photon daughters are given in parenthesis. This highlights the importance of photons in matched background tracks. The fraction of not matched background tracks is 97.4%, while 2.5% are matched to $e^{\pm}$ from photon conversion. Between MC12d and MC13b the main contamination source is the not truth matched tracks but MC13b samples present an higher contamination from these tracks which force us to revise the selection.

The datasets signal and background IP distributions are compared in Fig. 3.3. The signal purify from materail interaction still presents bumps in the region $\pm 6 cm$ implying some other material interaction sources have been forgot. The backgrounds present similarities between MC12d and MC13a bgx0 and in other parts between MC13a bgx1 and MC13b. These are more relevant on dz distribution and below $6 cm$ on dr. This is consistent with the datasets composition seen in

PDG ID	Tracks	Anc 1	Anc 2	Anc 3	Anc 4
not matched (0)	210130	210130	210130	215721	215742
e^+ (-11)	2669 (2662)	5	0	0	0
e^- (11)	2817 (2814)	5	5591	21	1
μ^+ (-13)	4 (2)	0	0	0	0
γ (22)	0	5599	21	1	0
ρ^0 (113)	0	2	0	0	0
π^- (-211)	21 (20)	0	0	0	0
π^+ (211)	21 (20)	2	0	0	0
η (221)	0	0	1	0	0
K^+ (321)	1 (1)	0	0	0	0
p (2212)	79 (79)	0	0	0	0

Table 3.2: Background track sources. The first column gives the number of tracks matched to each PDG ID, number given by the Particle Data Group to represent each particles, (PDGID is 0 in case it is not matched), the next columns show the number of candidates for each PDG ID for the different level of ancestors, e.g Anc 1 is the track mother. Numbers in parenthesis correspond to the case where the first ancestor is a photon

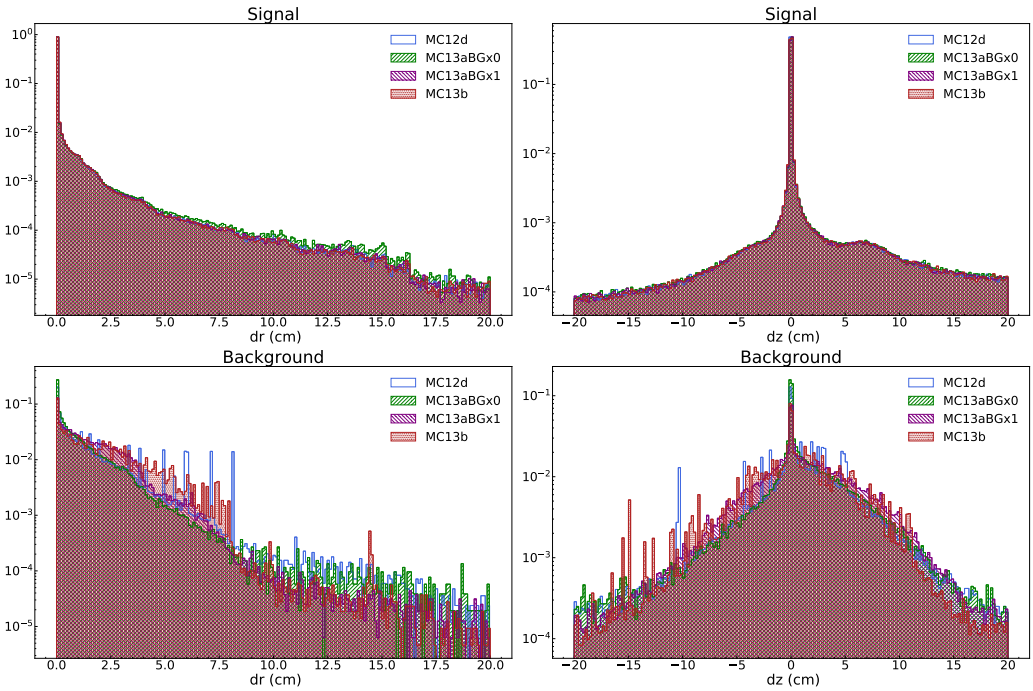

Figure 3.3: Comparison of the Signal (first row) and the Background (second row) IP distributions of the different samples.

Section 3.1.1.

Figure 3.4 shows the same IP distributions for background tracks separating matched and non-truth matched tracks. The distributions for matched background tracks gives similar results and they mainly correspond to the photon conversion, as seen in Table 3.2.

Nonetheless the not-truth matched tracks present the same similarities seen previously between MC12d/MC13a bgx0 and MC13a bgx1/MC13b. As a result the backgrounds of MC13b and MC13a bgx1 present the same behaviour and can be treated similarly.

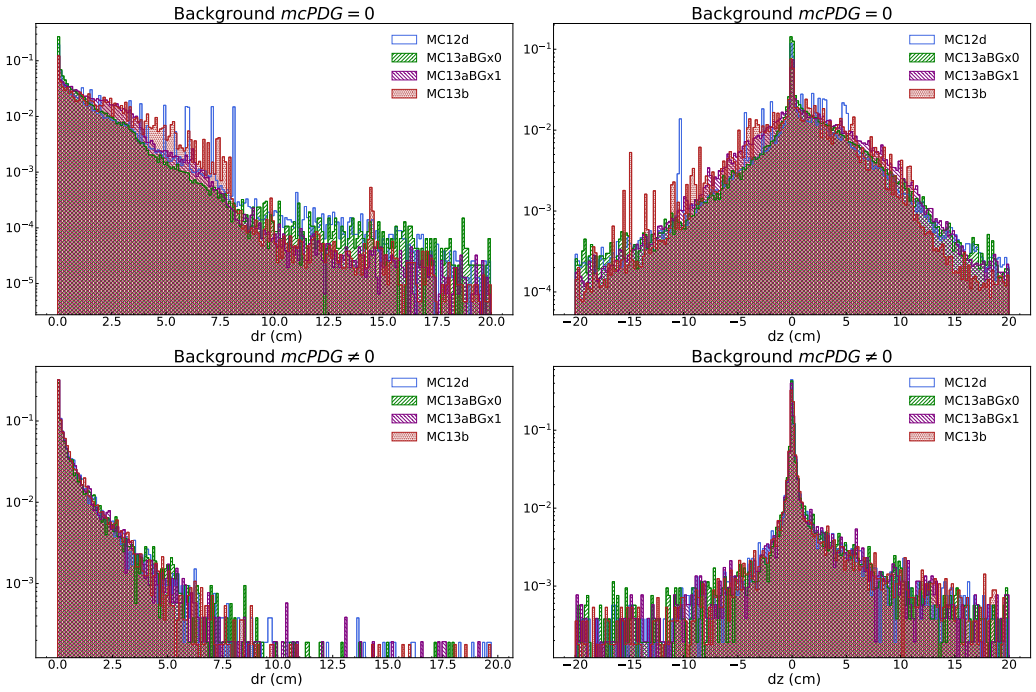

Figure 3.4: Comparison of the Background sources IP distributions of the different samples.

3.2 Selection optimization

3.2.1 dr and dz selection

An improvement of this study is to consider the dr and dz selection as a bi-dimensional problem. The adopted solution is to iterate simultaneously on both cuts. The dz boundaries are coupled by construction to ensures to keep the same amount of tracks in each dz side or to be symmetric around 0, this coupling allows to reduce the problem to two dimensions. Then we iterate on dr and dz boundaries and compute at each step the signal efficiency, the background rejection, and the purity, defined respectively as the fraction of surviving signal tracks, the fraction of rejected background tracks and the ratio between the surviving signal (S) and the sum of surviving signal and background tracks ($S+B$) after each cut.

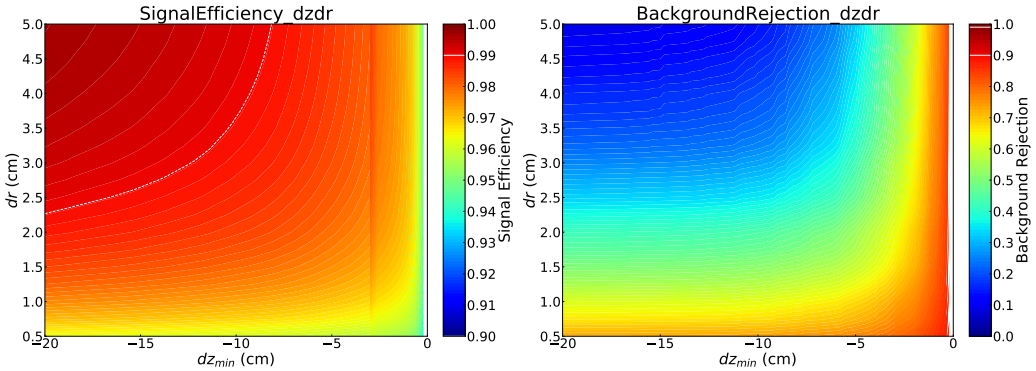

Figure 3.5: Efficiency and rejection trough the impact parameter.

Figure 3.5 represents the 2D efficiency and background rejections. The efficiency decreases significantly close to the beam pipe ($2\,cm$ diameter) but further it is quite flat. In the interesting region $|dr| < 2$ and $< -5 < dz < 5$, which is given as general recommendation for B-physics analysis, the efficiency reaches around 97%. Outside this region, the rejection is very low and it increases while getting closer to the beam pipe. In fact with the previous recommended IP selection, $-3 <$

$dz < 7$ and $dr < 1$, our rejection is limited to 54%, which explains our need to have a tighter one. We see from these plots that the rejection can be improved to reach 70% with an efficiency around 97%.

Signal Efficiency [%]				Background Rejection [%]			
$dz\backslash dr$	2.0	1.0	0.5	$dz\backslash dr$	2.0	1.0	0.5
[−5.0, 7.3]	98.09	97.12	96.00	[−5.0, 7.3]	50.19	65.78	75.32
[−3.0,5.42]	97.82	96.94	95.89	[−3.0,5.42]	57.91	70.22	78.23
[−2.0,4.15]	97.60	96.79	95.80	[−2.0,4.15]	63.58	73.09	79.51
[−1.0,2.55]	97.24	96.52	95.62	[−1.0,2.55]	71.13	77.69	81.66
[−3.0, 7.0]	97.94	97.02	95.94	[−3.0, 7.0]	55.06	68.71	77.22

Purity [%]

$dz\backslash dr$	2.0	1.0	0.5
[−5.0, 7.3]	93.18	93.18	93.18
[−3.0,5.42]	94.32	94.32	94.32
[−2.0,4.15]	95.21	95.21	95.21
[−1.0,2.55]	96.40	96.40	96.40
[−3.0, 7.0]	93.91	93.91	93.91

Table 3.3: IP selection for asymmetric dz combination.

The Table 3.3 presents in details some cut combinations and the previous track selection for tau analyses, $-3 < dz < 7$ and $dr < 1$. These tables confirm the slow efficiency variations along dz. dr appears to be a promising parameter to gain in efficiency, however with looser selection the rejection drops too much and a compromise between higher efficiency and the rejection rate has to be optimized. Two combinations with an efficiency around 97% and a background rejection higher than 70% can be distinguished: $-3 < dz < 5.42$, $dr < 1$ and $-2 < dz < 4.15$, $dr < 1$,

Signal Efficiency [%]				Background Rejection [%]			
$dz\backslash dr$	2.0	1.0	0.5	$dz\backslash dr$	2.0	1.0	0.5
[−5.0,7.22]	98.18	97.41	96.46	[−5.0,7.22]	47.99	62.97	72.23
[−3.0,5.26]	97.91	97.22	96.36	[−3.0,5.26]	55.57	67.31	75.04
[−2.0,4.01]	97.69	97.07	96.26	[−2.0,4.01]	61.01	70.07	76.25
[−1.0,2.39]	97.30	96.77	96.07	[−1.0,2.39]	68.47	74.50	78.32
[−3.0, 7.0]	98.03	97.30	96.40	[−3.0, 7.0]	52.63	65.76	74.02

Purity [%]

$dz\backslash dr$	2.0	1.0	0.5
[−5.0,7.22]	95.03	95.03	95.03
[−3.0,5.26]	95.84	95.84	95.84
[−2.0,4.01]	96.46	96.46	96.46
[−1.0,2.39]	97.28	97.28	97.28
[−3.0, 7.0]	95.53	95.53	95.53

Table 3.4: IP selection for asymmetric dz combination for the 3x1 topology.

The topologies is a decay characterization by the number of charged particles in the final state, 3×1 means that one tau decays into three charged particle while the other decay into one charged particle. Originally track selection was performed in function of the topology, in order to verify the different impact of the new track selection the study is repeated in function of the topologies: 3×1 in Table 3.4 and 3×3 topology in Table 3.5. The dz distributions of each topology give similar results as visible in upper dz limits. The different selections give similar results and are in

Signal Efficiency [%]				Background Rejection [%]			
$dz\backslash dr$	2.0	1.0	0.5	$dz\backslash dr$	2.0	1.0	0.5
$[-5.0, 7.99]$	98.33	97.64	96.79	$[-5.0, 7.99]$	44.22	58.68	67.66
$[-3.0, 5.78]$	98.02	97.44	96.66	$[-3.0, 5.78]$	50.92	62.20	69.80
$[-2.0, 4.23]$	97.78	97.27	96.56	$[-2.0, 4.23]$	56.68	65.23	71.57
$[-1.0, 2.47]$	97.41	96.98	96.36	$[-1.0, 2.47]$	63.94	69.61	73.49
$[-3.0, 7.0]$	98.12	97.50	96.69	$[-3.0, 7.0]$	49.68	61.60	69.48

Purity [%]			
$dz\backslash dr$	2.0	1.0	0.5
$[-5.0, 7.99]$	95.88	95.88	95.88
$[-3.0, 5.78]$	96.45	96.45	96.45
$[-2.0, 4.23]$	96.97	96.97	96.97
$[-1.0, 2.47]$	97.62	97.62	97.62
$[-3.0, 7.0]$	96.33	96.33	96.33

Table 3.5: IP selection for asymetric dz combination for the 3x3 topology.

agreement with the Table 3.3. The main difference between the two topologies is the rejection rate which can gain 5% in the 3×1 case.

Signal Efficiency [%]				Background Rejection [%]			
$dz\backslash dr$	2.0	1.0	0.5	$dz\backslash dr$	2.0	1.0	0.5
$[-5.0, 5.0]$	97.92	97.01	95.93	$[-5.0, 5.0]$	54.03	67.88	76.57
$[-3.0, 3.0]$	97.59	96.78	95.80	$[-3.0, 3.0]$	63.30	73.33	79.63
$[-2.0, 2.0]$	97.34	96.60	95.68	$[-2.0, 2.0]$	69.45	76.31	81.23
$[-1.0, 1.0]$	96.90	96.27	95.45	$[-1.0, 1.0]$	78.48	81.61	84.00

Purity [%]			
$dz\backslash dr$	2.0	1.0	0.5
$[-5.0, 5.0]$	93.75	93.75	93.75
$[-3.0, 3.0]$	95.26	95.26	95.26
$[-2.0, 2.0]$	96.21	96.21	96.21
$[-1.0, 1.0]$	97.43	97.43	97.43

Table 3.6: IP symmetric dz combination.

For the finalization of the impact parameters selection, the symmetric dz cuts is also studied in table 3.6. No significant difference in the signal efficiency has been observed between the symmetric and asymmetric cases, while the symmetric case showed a better background rejection. The symmetric combination $|dz| < 3$ and $|dr| < 1$ allows to reach 96.78% of efficiency with rejection of 73.33%. The symmetric cuts present a better rejection with a similar efficiency than asymmetric ones.

To summarize, we identify three possible IP selections, two asymmetric and one symmetric:

- $-3 < dz < 5.42$ and $dr < 1$,
- $-2 < dz < 4.15$ and $dr < 1$,
- $-3 < dz < 3$ and $dr < 1$.

Furthermore in Fig. 3.6 the dz distribution is compared before and after the symmetric IP combination, $-3 < dz < 3$ and $dr < 1$. This allows to verify that all these cuts allow to remove bumps by reducing the material interaction tracks. The impact on the material interactions rejection is

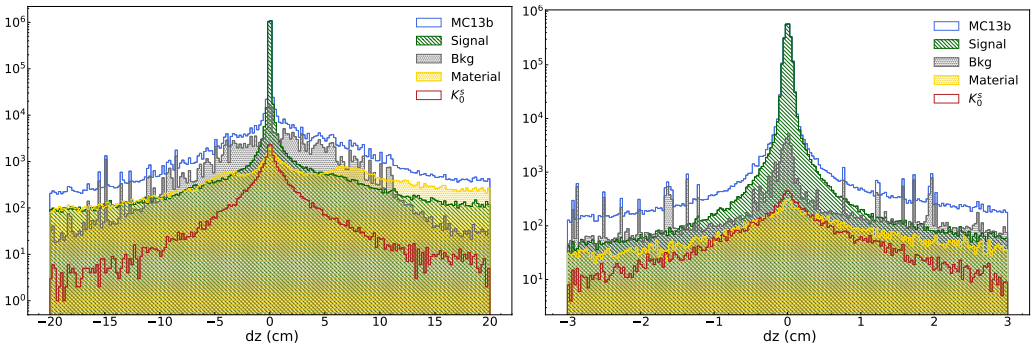

Figure 3.6: Comparison of dz distribution before (lepft) and after (right) $-3 < dz < 3$ and $dr < 1$ cut in MC13b, "bumps" structures disappeared with the cut.

	Material rejection [%]	Material rejection [%] after $pT > 0.075$ cut
$-2 < dz < 4.15$ and $dr < 1$	80.7	80.9
$-3 < dz < 5.42$ and $dr < 1$	77.5	77.7
$-3 < dz < 3$ and $dr < 1$	81.0	81.2

Table 3.7: Rejection of material interaction in MC13b sample.

summarized in Table 3.7. These cuts allow to reduce the identified material interactions by 80%, the symmetric selection presents the best rejection.

3.2.2 Additional selection

The first conclusion about dr and dz selection is the small variation of signal efficiency in our interesting region. This is why we want to optimize the background rejection for a fixed efficiency. However the background rejection achievable by applying selections on the impact parameters is limited by the large amount of background which also peaks in the proximity of the beam pipe. Nevertheless other variables, in particular kinematics variables, present some important differences between signal and background. In this section, these variables are probed in addition to the IP selection with the goal of improving the background rejection.

Interesting signal/background discrimination is observed in the distributions of the numbers of SVD and CDC hits. In the upper plots of Fig. 3.7, we can see that the background tracks population peaks at low SVD(CDC) hits. These two variables allow a much better rejection, but with a considerable loss in the signal efficiency. Due to the signal efficiency decrease, no tighter selection on the number of SVD and CDC hits is considered, except for $nCDCHits > 0$. However, we noticed that SVD and CDC hits selections do not keep all the non-zero SVD hits tracks. Indeed, among the zero-CDC-hit tracks, there is a considerable fraction of SVD-only tracks (which present only hits in the SVD layers) visible in Fig.3.8. Therefore, not to loose this low momenta SVD-only tracks, possible selections based on the numbers of SVD and CDC hits are discarded.

The Fig. 3.9 presents the distribution and the scan after IP selection for the theta angle and the transverse momentum. In theta distribution, the background tracks peak in the endcap regions, we thus try to remove them by a symmetric cut with respect to the parallel x axis passing through $\theta = 1.5\ rad$. Since the rejection grows slower than the efficiency decreases, theta is not a good discriminating variable.

Finally, the most promising additional selection is on the transverse momentum. Indeed, as showed in Fig. 3.9, the background tracks correspond to very low momentum tracks and a loose cut can reject them. This idea of loose cut is verified by the efficiency and rejection plot as function of p_T cut value. The rejection grows fast for low momentum, $p_T < 0.1$, while efficiency

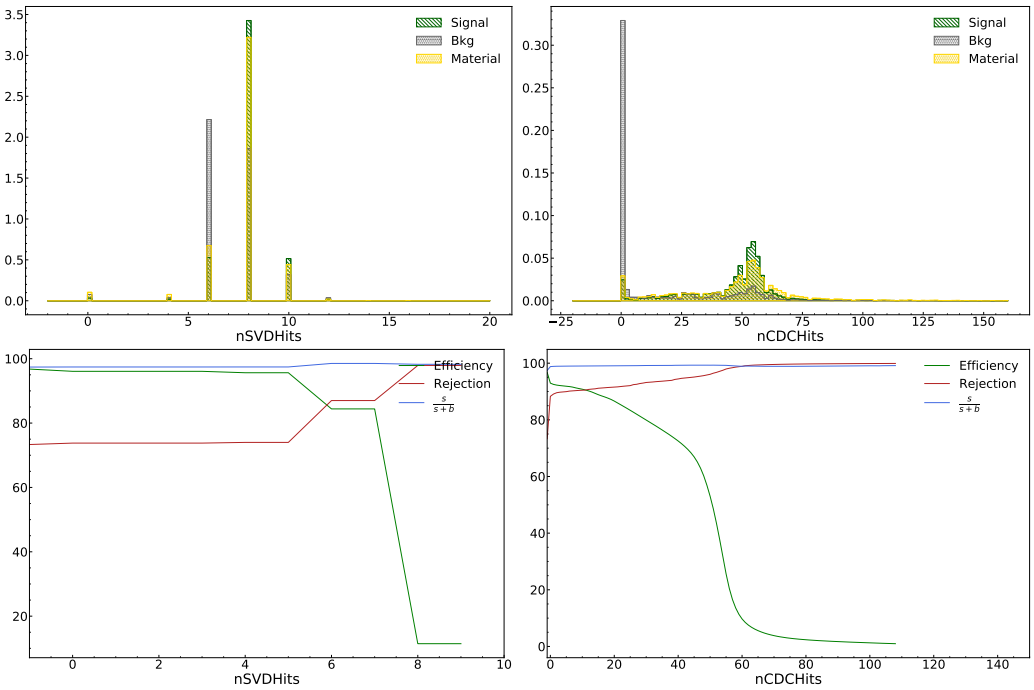

Figure 3.7: Upper plots: $nSVDHits$ and $nCDCHits$ distributions. Lower plots: efficiency, rejection and purity as function of a cut on the number of hits, after "$-3 < dz < 3$, $dr < 1$" cuts.

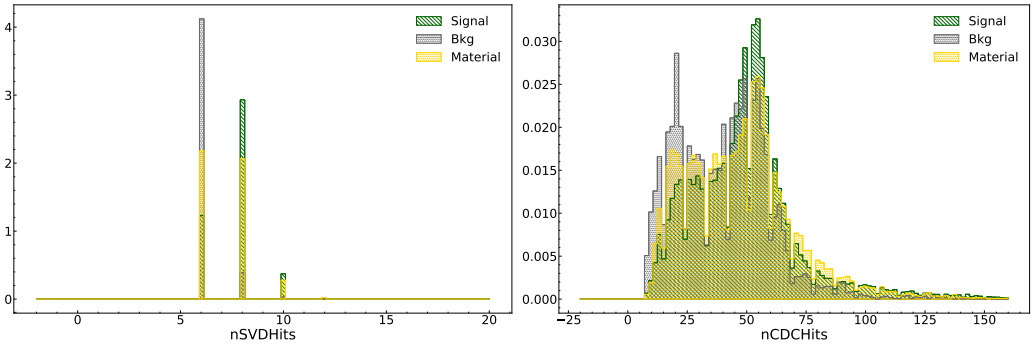

Figure 3.8: Left plot: $nSVDHits$ distribution under $nCDCHits = 0$ condition.
Right plot: $nCDCHits$ distribution under $nSVDHits = 0$ condition.

varies slowly. A cut on $p_T > 0.075$ GeV corresponds to the start of a plateau for the purity. This additional selection can increase by 10% the rejection and keep an efficiency higher than 96%.

In order to verify the feasibility of a p_T cut at 0.075 GeV it is require to check the experimental error on the transverse momentum measurement done by Belle II, it is called resolution. The resolution could be computed by performing a fit on the difference between the measured pT and the truth one $mcPT$, given by the simulation. On Fig. 3.10 the fit model PDF is provided by adding three gaussian distributions PDF_i, with respective weight coefficient f_i, with the same mean but with different standard deviation:

$$PDF = f_1 PDF_1 + f_2 PDF_2 + (1 - f_1 - f_2) PDF_3. \tag{3.1}$$

The fitting parameters are:

- *mean* represent the mean of the three gaussians and of the general fit model,
- σ_1 the standard deviation of the narrowest gaussian (green curve) used to fit the core of the distribution,

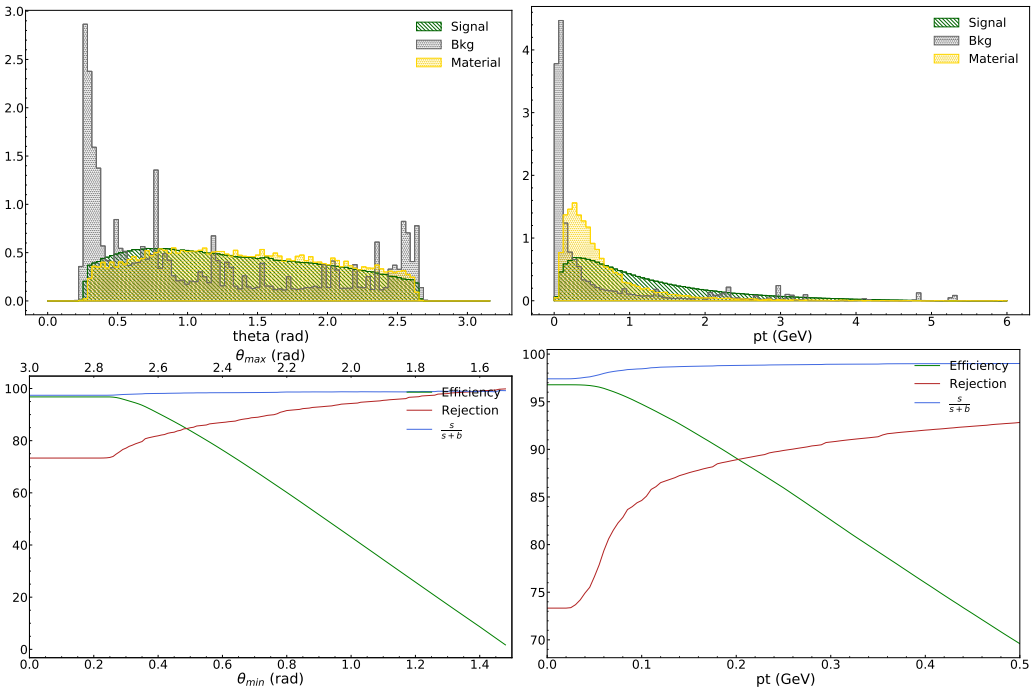

Figure 3.9: Upper plots: θ and p_T distributions. Lower plots: efficiency, rejection and purity as function of the cut value, after "$-3 < dz < 3$, $dr < 1$ selection" cut.

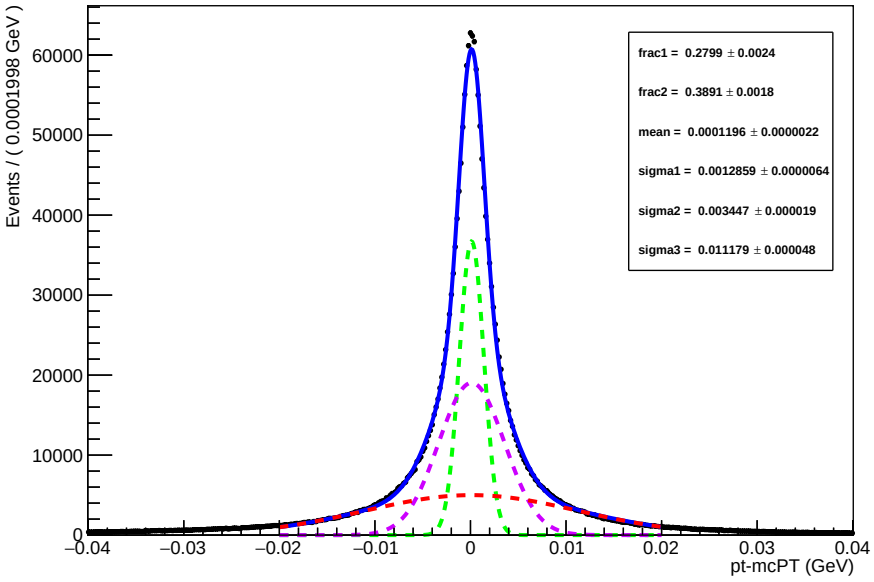

Figure 3.10: Transverse momentum resolution.

- σ_2 is the standard deviation of the middle one (purple curve), its aim is to properly fit the transition between the core and the tails,
- σ_3 is the standard deviation of the largest distribution (red curve) to model the distribution tails,
- f_1 and f_2 (or *frac1* and *frac2*) are the weight coefficient used to built the fit model.

The fit is made in the range $[-0.02, 0.02]$, this allows to do not insert a fourth gaussians with a low

impact on the global model.

The resolution r could be obtained by extrapolating the standard deviation of the fit model *PDF* by the weighted σ_i quadrature summing from fitting results:

$$r = \sqrt{f_1\sigma_1^2 + f_2\sigma_2^2 + (1 - f_1 - f_2)\sigma_3^2}$$
$$= 6.816 \pm 0.007\,\text{MeV}.$$

The resolution is enough small to consider a transverse momentum selection at 75 MeV.

As shown in Table 3.7, the rejection of material interaction slightly increases with the transverse momentum cut, in a similar way for the different IP selections.

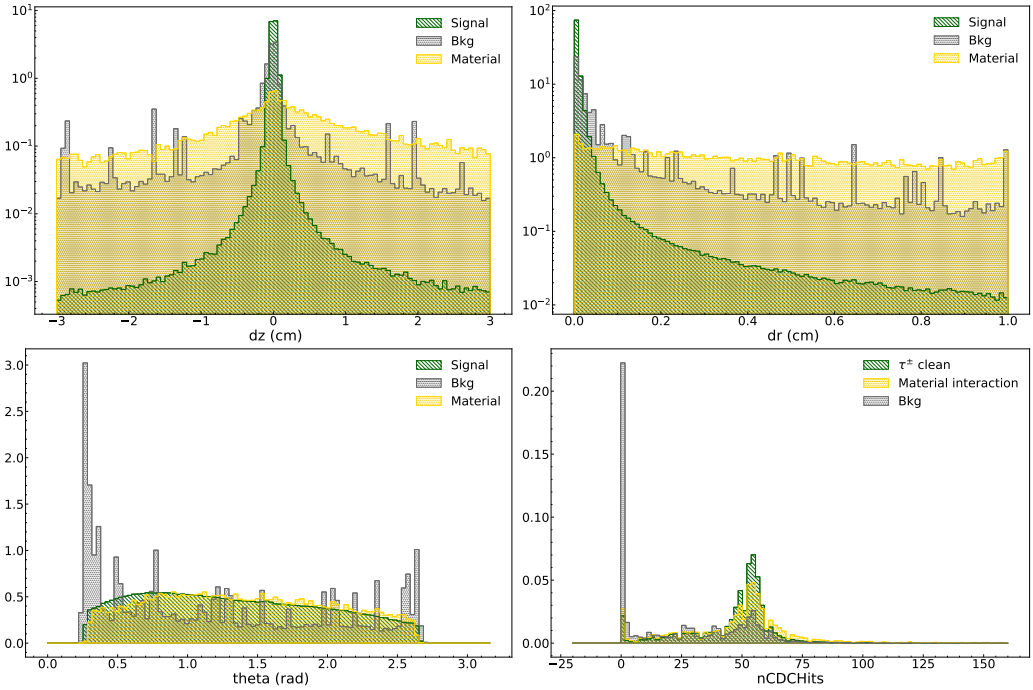

Figure 3.11: Impact of the "$-3 < dz < 3$, $dr < 1$ and $p_T > 0.075$" cut in MC13b distributions.

Fig. 3.11 represents the impact parameters, θ angle and CDC hits distributions after applying the $-3 < dz < 3$ $dr < 1$ and $pt > 0.075$ cuts. The low impact in signal $\simeq 3\%$ do not impact IP, θ and *nCDCHits* distributions.

Conclusion

This study presents an update the previous good track selection recommendations for tau physics with the lastest Monte-Carlo simulation campaign, MC13b. An increase of background tracks due to non-truth matched tracks appears in MC13b compared to previously used MC12d. The goal of the selection is to minimize the background contamination while keeping the maximum of signal tracks.

	Efficiency [%]	Rejection [%]	Purity [%]
$-2 < dz < 4.15$ and $dr < 1$	96.79	73.09	95.21
$-3 < dz < 5.42$ and $dr < 1$	96.94	70.22	94.32
$-3 < dz < 3$ and $dr < 1$	96.78	73.33	95.26

Table 3.8: Selected IP combination

The impact parameter selection aims at cleaning the dataset by having the best rejection with at least 96% efficiency. These requirements allow to keep three possible selection seen in Table 3.8. These candidates show different advantages, a loose selection allows to have the best efficiency at a price of a lower rejection. However the efficiency increase is smaller than the rejection one. The tight cuts $-2 < dz < 4.15$ and $dr < 1$ or $-3 < dz < 3$ and $dr < 1$ present a better purity than the loose $-3 < dz < 5.42$ and $dr < 1$ cut. The asymmetric dz selection shows no evident improvement compared to the symmetric case. Indeed the largest right handed tail increases the efficiency but the rejection decreases drastically. This is why a symmetric selection looser than asymmetric one allows to reach an equivalent or better selection. The symmetric selection also presents the advantage to be more convenient with the general recommendation $-5 < dz < 5$. In addition Table 3.8 shows that the selection $-3 < dz < 3$ and $dr < 1$ has the best purity, it is thus the preferred option, when we exclude p_T selection.

	Efficiency [%]	Rejection [%]	Purity [%]
$-2 < dz < 4.15$ and $dr < 1$ and $p_T > 0.075$	95.86	82.60	98.28
$-3 < dz < 5.42$ and $dr < 1$ and $p_T > 0.075$	95.99	81.17	98.14
$-3 < dz < 3$ and $dr < 1$ and $p_T > 0.075$	95.87	82.28	98.24

Table 3.9: Selection efficiency, rejection and purity when applying IP and p_T cuts.

Finally several additional variables have been probed in order to improve the background rejection. The transverse momentum appears to be promising with a rejection rate increase of 10%, with a slight efficiency decrease as showed in Table 3.9, when selecting $p_T > 0.075\,\text{GeV}$. Nevertheless, this selection raises several questions regarding topology migration and further studies, beyond the scope of this note, are required before adopting it.

Because it presents the best purity and it is symmetric, the $-3 < dz < 3$ and $dr < 1$ combination was finally adopted and used by the τ-group after presenting our results. This work is also documented in a note internal to the Belle II collaboration.

Bibliography

[1] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191.

[2] S. L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2.

[3] P. W. Anderson. "Plasmons, Gauge Invariance, and Mass". In: *Phys. Rev.* 130 (1 Apr. 1963), pp. 439–442. DOI: 10.1103/PhysRev.130.439.

[4] S. Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[5] Y. Ohnishi et al. "Accelerator design at SuperKEKB". In: *Progress of Theoretical and Experimental Physics* 2013.3 (Mar. 2013). DOI: 10.1093/ptep/pts083.

[6] P. Oddone. "An Asymmetric B-Factory Based on PEPa". In: *Annals of the New York Academy of Sciences* 578.1 (Dec. 1989), pp. 237–247. DOI: 10.1111/j.1749-6632.1989.tb50614.x.

[7] E. Kou et al. "The Belle II Physics Book". In: *Progress of Theoretical and Experimental Physics* 2019.12 (Dec. 2019). ISSN: 2050-3911. DOI: 10.1093/ptep/ptz106.

[8] T. Abe et al. *Belle II Technical Design Report*. 2010. arXiv: 1011.0352 [physics.ins-det].

[9] Laura Zani and Professor Francesco Forti. "Search for an invisible Z' in $\mu^+\mu^-$ plus missing energy events at Belle II". PhD thesis. University of Pisa, Physics Department E. Fermi, 2020.

[10] R. Zitoun. *Introduction à la physique des particules*. 2e édition. Sciences sup Physique. Paris: Dunod, 2004.

[11] B. Clément. *Physique des particules : introduction aux concepts et au formalisme du modèle standard*. Sciences sup Physique. Paris: Dunod, 2013.

[12] K. Hayasaka et al. "Search for lepton-flavor-violating τ decays into three leptons with 719 million produced $\tau^+\tau^-$ pairs". In: *Physics Letters B* 687.2-3 (Apr. 2010), pp. 139–143. DOI: 10.1016/j.physletb.2010.03.037.

[13] F. Tenchini. "Track selection for $\tau^+\tau^-$ analysis". In: *BELLE2-NOTE-TE-2020-012* (Mar. 2020).

[14] J.-F. Krohn et al. *Global Decay Chain Vertex Fitting at B-Factories*. Jan. 2019. arXiv: 1901.11198 [hep-ex].

[15] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[16] D.J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (July 2001), pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[17] S. Jadach, B.F.L. Ward, and Z. Was. "The precision Monte Carlo event generator for two-fermion final states in collisions". In: *Computer Physics Communications* 130.3 (Aug. 2000), pp. 260–325. DOI: 10.1016/s0010-4655(00)00048-5.

[18] N. Davidson et al. "Universal interface of TAUOLA: Technical and physics documentation". In: *Computer Physics Communications* 183.3 (Mar. 2012), pp. 821–843. DOI: https://doi.org/10.1016/j.cpc.2011.12.009.

[19] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506.3 (July 2003), pp. 250–303. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.

[20] R. Brun and F. Rademakers. *ROOT documentation*. 2018. URL: https://root.cern.ch/root/htmldoc/guides/users-guide/ROOTUsersGuide.html.

Abstract

Being forbidden in the Standard Model, lepton flavour violating decays appears to be a good probe for searching physics beyond the standard model. In particular, the tau lepton, thanks to its high mass enables more lepton flavour violating decays: $\tau \to \ell\gamma$ and $\tau \to 3\ell$ with $\ell = e,\ \mu$. In addition the interest for τ decays is reinforced by several new physics models prediction on branching fractions just below current experimental limits.

The data taking of Belle II, located at KEK, Japan, started in 2019 and should end by 2027, collecting 50 times more data than its predecessor, Belle. The large amount of e^+e^- collisions will provide a huge number of $\tau^+\tau^-$ pairs, ideal for tau studies.

The τ-group of Belle II needs differ from other B and C mesons analysis, in particular because of the signal contamination by beam backgrounds. The use of latest simulated samples and backgrounds implies the requirement to update the "good track" definition by optimizing the track selection on impact parameters.

Résumé

Interdites dans le modèle standard, les désintégrations violant la saveur leptonique sont des sujets intéressants pour la recherche de physique au delà du modèle standard. En particulier les désintégrations du lepton τ, par sa masse élevée, permettent un plus grand nombre de désintégrations interdites comme: $\tau \to \ell\gamma$ et $\tau \to 3\ell$ avec $\ell = e,\ \mu$. D'autre part l'intérêt porté aux désintégrations du τ a été accru par les probabilités de désintégration juste en dessous des limites expérimentales actuelles prédites par des modèles de nouvelle physique.

La prise de donnée de l'expérience Belle II, à KEK au Japon, a commencé en 2019 et prévoit pour 2027 50 fois plus de données que son prédécesseur Belle. Avec son grand nombre de collisions e^+e^-, un grand taux de paire $\tau^+\tau^-$ sont à prévoir et sont idéales pour l'étude du tau.

Les besoins du groupe d'analyse du τ à Belle II diffèrent des autres B et C mesons analyses, en particulier pour nettoyer le signal du bruit de fond venant du faisceau. L'utilisation des simulations les plus récentes a notamment remis en question la définition des "bonnes traces", et a soulevé la nécessité de faire évoluer la sélection des traces selon les paramètres d'impact.