

UNIVERSITÀ DEGLI STUDI DI TORINO

Master Degree in Physics



Data selection tools based on machine
learning for hyperon form factor studies with
the Belle II experiment

Federico Bonaldo

Supervisor:
Prof. Simonetta Marcello

Co-Supervisors:
Prof. Karin Schönning
Dr. Bianca Scavino
Dr. Martina Laurenza

Examiner:
Prof. Elena Botta

Academic Year 2022-2023

Abstract

Study the internal structure of the Λ hyperon at the energy of Belle II may be possible through the channel $e^+e^- \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$. The reconstruction for this channel is fundamental for the scope of the project. Hence, it is crucial to investigate the reconstruction of the γ_{ISR} because including its information on the final state may improve the reconstruction analysis for this channel. Whereas hyperons are unstable particles, we use a reconstruction algorithm to access the Λ particles. Since the reconstruction algorithm collects a lot of background caused by a wrong reconstruction, in my thesis I developed a Λ hyperon selector using machine learning tools aimed to improve the current reconstruction algorithm for the Λ hyperons. The analysis I conducted uses a supervised learning model, trained and tested on Monte Carlo data samples, to study which features, i.e. kinematics variables, are relevant for the selector. Thanks to the Monte Carlo Truth used as a target variable, I evaluated the performance of my Λ selector reaching an efficiency of 80% and a purity of 90% on an independent Monte Carlo sample. The work of my thesis proves the great potential of the application of machine learning tools in high energy physics experiment.

Thesis overview

My work of thesis has been conducted in Uppsala University for an Erasmus Plus Traineeship program. The activity started on September 2023 until March 2024. In this period I had the occasion to study and work together with the hadron physics group of Uppsala, supervised by Prof. Karin Schönning. Since the aim of my thesis is to build a starting tool for the study of the hyperons form factor in Belle II, I investigated on the performance of reconstruction of the γ_{ISR} in the channel $e^+e^- \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$ and developed a Λ hyperon selector for distinguishing the Λ candidates from the background. At the beginning of the thesis I give a brief introduction on the Standard Model of particle physics, and in particular on the Λ hyperon. In Chapter 3 I introduce the Belle II experiment, describing the components of the entire detector. Since a great part of my thesis consists of the development of a machine learning tool for the Λ hyperon selection I explain in Chapter 4 how machine learning works, while Chapter 5 defines the strategy and the analysis plan for the study. My analysis starts from Chapter 6, presenting the Monte Carlo samples I generated and used during the activity. The analysis of the performance of the reconstruction of the γ_{ISR} is described in Chapter 7, while Chapter 8 is completely dedicated on the reconstruction of the Λ hyperons in the Belle II experiment, first presenting a classical approach, i.e. studying the single kinematics variables distributions, and then presenting the machine learning approach comparing the different configurations I evaluated during the analysis. Finally, at the last Chapter, the results of my thesis are summarized and compared to previous analysis.

Contents

1	Introduction	5
1.1	The Standard Model	6
1.2	The Strong Interaction	7
1.2.1	Quantum Chromodynamics	8
1.3	Quarks and Hadrons	8
1.3.1	Quantum Numbers	9
1.3.2	Mesons	11
1.3.3	Baryons	12
1.3.4	Hyperons	12
2	The structure of the Λ hyperon	14
2.1	Form Factors	15
2.2	$\Lambda\bar{\Lambda}$ final state in e^+e^- annihilation	17
2.2.1	Tagged vs untagged ISR	19
3	The Belle II experiment	20
3.1	SuperKEKB	20
3.1.1	Nano-Beam Scheme	21
3.2	Belle II detector	22
3.3	Software tools	26
3.3.1	Analysis chain	26
4	Machine Learning (ML)	28
4.1	Learning types	29
4.2	Model Optimization	30
4.3	Workflow	31
4.4	Evaluation	32
4.5	Decision trees	35

5	Analysis Strategy	37
5.1	Strategy	37
6	$\Lambda\bar{\Lambda}$ Generation	39
6.1	$\Lambda\bar{\Lambda}$ signal samples	39
6.1.1	$e^+e^- \rightarrow \Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$	39
6.1.2	$e^+e^- \rightarrow e^+e^-\gamma_{ISR} \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$	41
6.1.3	Generic Hadronic background	43
7	Topology and reconstruction of the γ_{ISR}	45
7.1	Topology of the γ_{ISR}	45
7.1.1	Energy distribution	45
7.1.2	Angular distribution	46
7.2	γ_{ISR} reconstruction	47
7.2.1	Energy distribution	48
7.2.2	Angular distribution	49
7.3	Efficiency	49
7.3.1	Efficiency dependence on energy	50
7.3.2	Efficiency dependence on angle	51
7.4	Conclusion	52
8	Λ reconstruction	53
8.1	Λ and $\bar{\Lambda}$ reconstruction	53
8.1.1	Kinematics plots	54
8.2	MVA analysis	59
8.2.1	Workflow and variables	60
8.2.2	Configuration 0	62
8.2.3	Configuration 1	66
8.2.4	Configuration 2	70
8.2.5	Configuration 3	74
8.2.6	Comparing the configurations	76
8.2.7	Cut value choice	77
8.2.8	Applying model	80
9	Conclusion and outlook	83

Chapter 1

Introduction

Understanding the very nature of our existence is one of the tasks of physics. Microcosmos, i.e. particles, follow rules described by quantum mechanics and those have a stochastic nature. The measurable quantities, i.e. the observables, follow certain distributions, and sometimes these distributions can give information about the universe itself. A field of research in nuclear physics is how the strong interaction forms hadrons and generates their mass. Hadrons consist mainly in protons and neutrons. Heavy relatives of these abundant particles are hyperons. They are similar to protons and neutrons, in the sense that they contain quarks, but in hyperons, one or several light up or down quarks are replaced by a strange quark. Hyperons are unstable particles and are challenging to reconstruct, which calls for new methods. In the Belle II experiment, one way to produce Λ hyperons is via the Initial State Radiation (ISR) process. In my thesis I have studied the irradiated ISR photon in the Belle II detector and investigate its detection feasibility. In addition, I studied different ways to reconstruct Λ hyperons. I developed an algorithm for the Λ hyperons selection from different MonteCarlo samples used for simulate physical processes. This algorithm aims to improve the current reconstruction algorithm available in the experiment discerning the Λ hyperon from the background. For this purpose I used a modern tool for the event selection i.e. a machine learning model called Fast Boosted Decision Trees.

1.1 The Standard Model

In the quest to understand the fundamental building blocks of the universe, scientists have developed the *Standard Model* of particle physics. This theoretical framework elegantly describes the elementary particles and their interactions, providing a comprehensive picture of the microscopic world [1]. The Standard Model categorizes particles into two main groups: *fermions* and *bosons*. Fermions make up matter and include *quarks* and *leptons*. Quarks combine to form *hadrons*, such as protons and neutrons, while leptons include electrons and neutrinos. Bosons, on the other hand, mediate the fundamental forces in the universe. These fundamental forces rule the elementary particle interactions and are classified as strong, electromagnetic, weak and gravitational forces, which their strengths and ranges are described in Tab. 1.1:

- Electromagnetic force: described by the exchange of photons. This force governs interactions between charged particles, such as electrons and protons.
- Weak force: mediated by $W^\pm$ and Z^0 bosons. The weak force is responsible for processes like radioactive decays and neutrino interactions.
- Strong force: gluons mediate the strong force, binding quarks together to form protons, neutrons, and other hadrons. This force is dominant at very short distances and holds the nucleus of atoms together.

The gravitational force is not explicitly included in the Standard Model. Instead, gravity is described by the theory of general relativity. While the other three forces are well understood within the quantum framework, unifying gravity with the quantum world remains a significant challenge. Experi-

Interactions	Strength (relative to strong)	Range (m)
Strong	1	10^{-15}
Electromagnetic	10^{-2}	∞
Weak	10^{-6}	10^{-18}
Gravitational	10^{-38}	∞

Table 1.1: Four types of interactions in order of strength. Note that the gravitational interaction is not part of the Standard Model.

tal validations of the Standard Model have been numerous and impressive.

The discovery of the $W^\pm$ and Z^0 bosons at CERN in 1983 provided crucial evidence for the weak force predictions. Additionally, the precise measurements at the LHC have confirmed many aspects of the model, including particles as the the Higgs boson. In Fig. 1.1, the fundamental particles described by the Standard Model are represented.

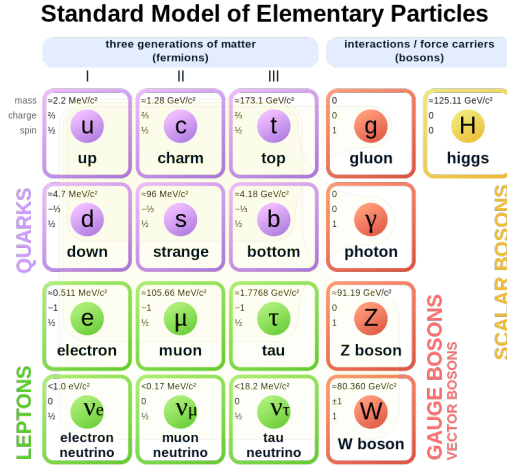


Figure 1.1: The fundamental particles of the Standard Model. Figure from [2].

1.2 The Strong Interaction

The strong interaction plays a crucial role in governing the behavior of quarks and maintaining the cohesion of atomic nuclei. This influential force is responsible for creating stable structures of particles, generating most of their masses. Understanding the mechanisms of the strong interaction poses one of the most challenging questions in contemporary physics. Main characteristics include its short-range dominance, typically on the femtometer scale, making it most impactful within the confines of the atomic nucleus. Additionally, the *binding of quarks* involves the exchange of gluons at a fundamental level, leading to the formation of hadrons.

1.2.1 Quantum Chromodynamics

To comprehend the rules of the strong interaction, one must delve into the theoretical framework known as *Quantum Chromodynamics* (QCD). In QCD, quarks are assigned an unique property called *colour charge*, analogous to electric charge in electromagnetism but distinct in its threefold nature: red r , green g and blue b . *Antiquarks* carry *anticolours*, $\bar{r}$, $\bar{g}$ and $\bar{b}$. Gluons, which serve as the force carriers for the strong interaction, possess colour charge. The gluon exchange between quarks constitutes the mechanism through which the strong force is mediated, resulting in a dynamic interplay of forces that binds quarks together. While photons in electromagnetic interaction are electrically neutral and do not interact mutually, gluons are self-coupling. The interquark interactions are assumed to be invariant under colour interchange, hence, they are described by the symmetry group $SU(3)$ [3]. The gluon-gluon interaction has no analogue in *Quantum Electron Dynamics*, and it can be shown that it leads to properties of the strong interaction which differ from those of the electromagnetic interaction. These referred properties are *colour confinement* and *asymptotic freedom*. Colour confinement is the requirement that observed states have zero colour charges. Asymptotic freedom means that when two colour charges are separated in space, the potential energy between them increases and at some point the energy necessary to separate them further is bigger then to product a new colour-anticolour charges that can team up with the original quark and antiquark to form new colour neutral systems [4]. These concepts refers to the fact that it is impossible to observe individual quarks or gluons. The colour quantum number does not enter in our description of hadrons. It means that baryons and mesons must be colourless [3]. Combining all the three quarks (antiquark), i.e. three colours (anticolours), a colourless baryon (antibaryon) can be formed. A bound state with one quark (colour) and one antiquark (anticolour) can form a bound colourless meson.

1.3 Quarks and Hadrons

Hadrons are composite particles made up of quarks and held together by the strong forces, mediated by gluons. Most of the observed states are baryons, made by three quarks qqq and mesons, made by a quark anti-quark pair $q\bar{q}$. According to QDC, same other colourless states are expected, so-called

exotic hadrons, such as glueballs i.e. bound states of two or three gluons representing states with quarks and gluonic degrees of freedom called hybrids. Other exotic states may be multiquarks states that can contain different combination of quarks and antiquarks for example $qq\bar{q}\bar{q}$ or other as $qqqq\bar{q}$. The description of these states is beyond the scope of this thesis.

1.3.1 Quantum Numbers

Quantum numbers are useful to describe the fundamental interactions and conserved quantities in the dynamics of a quantum system [3]. Hadrons have several quantum numbers which the most important are:

- *Spin*: denoted by S , it is defined as a intrinsic angular momentum of the elementary particles. Particles with half-integer spin following Fermi statistics are called fermions, while particles with integer spin following Bose-Einstein statistics are called bosons. To describe a system of fundamental particles, it is useful to introduce the *total angular momentum*, denoted by J , which is related to the total spin of the system, S , and its total orbital angular momentum L :

$$\vec{J} = \vec{L} + \vec{S} \quad (1.1)$$

The values of J are calculated according to the rules of the additive quantum numbers and start from $J = |L - S|$ to $J = |L + S|$. Since quarks are fermions the spin of the mesons may be $S=0$ or $S=1$, while the baryons spin may be $S=1/2$ or $S=3/2$.

- *Electric charge*: related to the electromagnetic force, the charge quantization is the principle that the charge of any object is an integer multiple of the elementary charge. Each quark carries a fraction of the unit electric charge $e = 1.6022 \times 10^{-19}C$, while leptons have an integer electric charge [3].
- *Parity*: denoted by P and represents a discrete symmetry. It describes the response of a particle's wave function to the sign inversion of its spatial coordinates: $(x, y, z) \rightarrow (-x, -y, -z)$. If, after the transformation, the wave function remains the same, the parity is classified as *even* ($P = +1$) whereas if the sign changes, it is *odd* ($P = -1$). A fermion and an antifermion has always opposite parity: by convention

quarks have $P_q = +1$ and antiquarks $P_{\bar{q}} = -1$. In baryons and mesons the parity is calculated according to the rules of the addition of the angular momentum. For the mesons, the parity P_M is given by:

$$P_M = P_q P_{\bar{q}} (-1)^L = (-1)^{L+1} \quad (1.2)$$

while for baryons, the parity P_B is given by:

$$P_B = P_{q1} P_{q2} P_{q3} (-1)^{L_{12}} (-1)^{L_3} = -P_{\bar{B}} \quad (1.3)$$

where L_{12} is the orbital angular momentum of a pair of quark in their center of momentum frame and L_3 is the orbital angular momentum of the third quark with respect to the center of mass of the quark pair. Parity is conserved in strong and electromagnetic interactions, but may be violated in weak interactions.

- *Baryon number*: denoted by B , it is an additive quantum number conserved in all the SM interactions. It is defined as:

$$B = \frac{1}{3}(n_q - n_{\bar{q}}) \quad (1.4)$$

where n_q is the number of quarks while $n_{\bar{q}}$ is the number of antiquarks. Each quark has $B = 1/3$ while antiquarks have $B = -1/3$. As consequence, baryons have $B = 1$ while antibaryons have $B = -1$ and mesons $B = 0$.

- *Strangeness*: denoted S , it quantifies the net number of strange quarks in a particle. By convention, the s quark has $S = -1$ and $\bar{s}$ has $S = +1$. All the other quarks have $S = 0$. The strangeness of baryons and meson is calculated as:

$$S = -N_s = -[N(s) - N(\bar{s})] \quad (1.5)$$

defining $N(s)$ the number of strange quarks and $N(\bar{s})$ the number of antistrange quarks. In the same way are defined *charm* (C), *bottomness* ($\hat{B}$) and *topness* (T).

- *Hypercharge*: denoted Y , it is related to the quark flavour. The hypercharge is defined as:

$$Y = B + S + C + \hat{B} + T \quad (1.6)$$

where, B is baryon number and S strangeness. The hypercharge is conserved in strong interaction but not in weak interactions.

- *Isospin*: hadrons with the same spin, parity, baryon number, strangeness and similar mass but with different electric charge may be organized in families called *isospin multiplets* [4]. The particles within a multiplet can be distinguished through the third component of isospin I_3 . As quantum number it can also be calculated as:

$$I_3 = Q - \frac{Y}{2} \quad (1.7)$$

with Q the electric charge and Y the hypercharge.

- *Charge Conjugation*: it is denoted C and it represents a discrete symmetry. The charge conjugation turns the particles into their anti-particles and vice-versa inverting all the intrinsic additive quantum numbers such as electric charge and baryon number. Only particles which are their own anti-particles, like π^0 , are eigenstates of C. In addition, charge conjugation is a symmetry of the strong interaction.

Quarks quantum numbers are summarized in Tab. 1.2 :

Quark	Q	B	Y	I	I_3
u	2/3	1/3	1/3	1/2	1/2
d	-1/3	1/3	1/3	1/2	-1/2
s	-1/3	1/3	-2/3	0	0
c	2/3	1/3	4/3	0	0
b	-1/3	1/3	-2/3	0	0
t	2/3	1/3	4/3	0	0

Table 1.2: Quantum numbers of individual quarks: electric charge Q, Baryon number B, hypercharge Y, isospin I and its third component I_3 .

1.3.2 Mesons

Mesons represent bound states composed by a quark-antiquark pair ($q\bar{q}$). The main property for these particles is the integer spin. Pions are the lightest mesons and they form an isospin triplet ($\pi^+\pi^0\pi^-$). Different isospin multiplets may be represented in a spin-parity representation in SU(3), referred to the three flavours taken into account. An example is shown in Fig. 1.2, where isospin multiplets appear on the horizontal lines. The Belle II experiment is

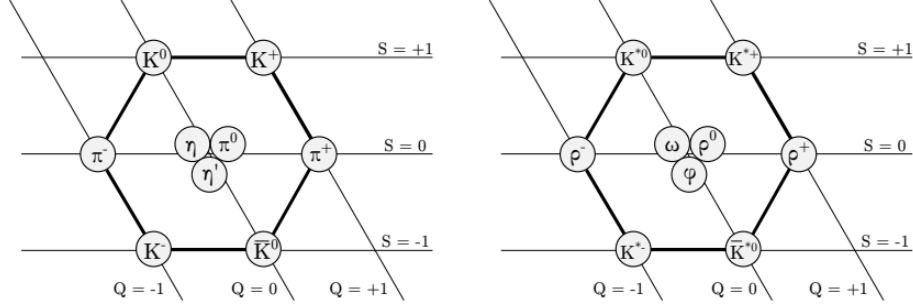


Figure 1.2: pseudoscalar $J^P = 0^-$ mesons on the left and vector $J^P = 1^-$ mesons on the right. Figure from [5].

also known as *B-factory* due to its abundant production of *Bottomonium*, a general name for $b\bar{b}$ vector states. Currently, a large data sample is being collected at the $\Upsilon(4S)$ resonance, that decays with a large fraction into B mesons.

1.3.3 Baryons

Composed by three quark, baryons contain an odd number of fermions and for this reason they are also fermions. The most well-known baryons are proton (uud) and neutron (udd). Like the mesons, baryons can be organised into spin-parity multiplets as a consequence of SU(3) flavour symmetry. Proton and neutron represent an isospin multiplet with $I=1/2$. In Fig. 1.3, the baryon spin $1/2^+$ octet and the baryon spin $3/2^+$ decuplet are represented in diagrams. In the former, we note that the protons and neutron form an isospin doublet.

1.3.4 Hyperons

As we see in Fig. 1.3, there are baryons with strangeness -1, -2 and -3 such as Λ , Σ^0 and Ω^- . Here, one or several up or down quarks in the nucleon have been replaced by a strange s quark or heavier, forming the *hyperons*. Hyperons therefore have properties that set them apart from their more fa-

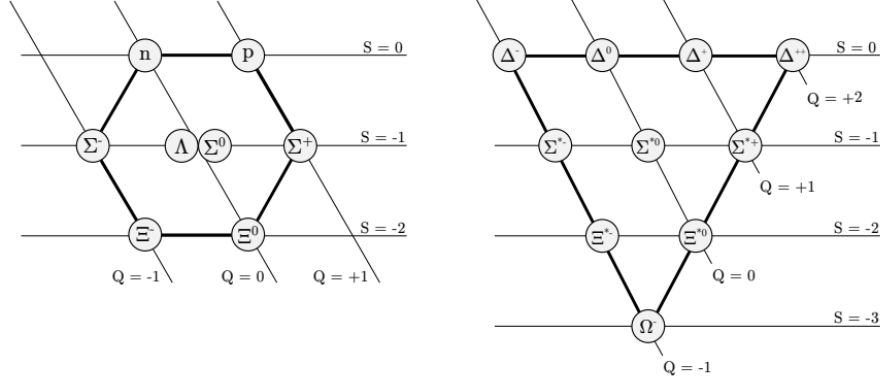


Figure 1.3: The baryon SU(3) octet $J^P = 1/2^+$ to the left and the decuplet $J^P = 3/2^+$ to the right. Figure from [5].

miliar counterparts. These particles are characterized by their strangeness¹, quantum number. By comparing the hyperons with nucleons may be interesting to understand the role of flavour in strong interactions. Through the hyperons decays studies is possible to extract properties related to the conservation of quantum numbers in strong interactions and, as consequence, their symmetries. Due to their nature is hard make hyperons beams: namely their finite lifetime ($< 10^{-10}s$), along with the fact that many of them are neutral. Neutral hyperons can be reconstructed from their decays products. Analysing their decays give access to their spin observable [6] [7]. Since this thesis focuses on the Λ hyperon reconstruction in The Belle II experiment, I am going to introduce in the next chapter more details about the Λ hyperon in particular.

¹hyperons may be characterized also by the charm, bottomness and topness.

Chapter 2

The structure of the Λ hyperon

The Λ hyperon is a baryon, discovered in 1950 in a cosmic ray experiment. Historically, Λ hyperon was denoted as *strange* due to its behavior in production and decay. In fact, though the particle was expected to live for $\sim 10^{-23}s$, it actually survived for $\sim 10^{-10}s$. The property that caused it to live so long was dubbed strangeness and led to the discovery of the strange quark [8]. Hyperons can be produced in strong interactions, electromagnetic interactions and in weak processes but their decay is mediated by the weak interaction, which occurs at a longer time-scale. The Λ hyperon is composed by one up, one down and one strange quark, uds , and the spin-parity $J^P = 1/2^+$. It is electrically neutral, and represents a singlet of isospin, with $I = 0$ and the third component $I_3 = 0$. Some of the Λ properties are summarized in Table 2.1 [9, 10]. Furthermore, the antiparticle of Λ is $\bar{\Lambda}$, composed by $\bar{u}\bar{d}\bar{s}$ system. A lifetime of $\sim 10^{-10}s$ means that a relativistic particle will decay

<i>Particle</i>	<i>Quark Composition</i>	<i>Mass (MeV/c²)</i>	<i>Mean Life-Time (ps)</i>	<i>Q</i>	<i>J^P</i>	<i>S</i>	<i>I</i>
Λ	uds	1115.683 ± 0.006	$261.20 \pm 0.49 \pm 0.63$	0	$1/2^+$	-1	0

Table 2.1: Λ hyperons properties.

after travelling a distance of a few centimeters or decimeters. This makes it short enough to decay inside a detector, that are typically several meters large, but long enough to travel a measurable distance before decaying. So, it is short compared to a charged pion that is typically stopped in the detector and hence completely measured, before it decays. But at the same time, it is long compared to an Υ or a J/Ψ that decay instantly on an ex-

perimental scale and therefore does not travel any distance in the detector before decaying [9]. Through the weak interaction the Λ hyperon can decay through different modes, summarised in Table 2.2 [10]. Most of the decays are two-body decays. Due to its electrical neutrality it does not interact with the detector material, hence leaving no trace in the detector. Therefore it has to be reconstructed by its charged daughters. The reconstruction may be done also through the neutral $\Lambda \rightarrow n\pi^0$ mode decay, by measuring the $\pi^0 \rightarrow \gamma\gamma$ decay in a calorimeter and by strong scattering of the neutron, but it is much more difficult and results in quite poor resolution.

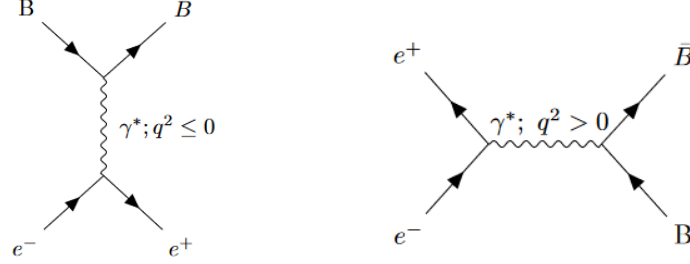
Λ Decay Modes	Fraction (Γ_i/Γ)
$p\pi^-$	$(63.9 \pm 0.5) \%$
$n\pi^0$	$(35.8 \pm 0.5) \%$

Table 2.2: Λ decay modes.

2.1 Form Factors

A tool to study the structure of the hadrons is provided by Electromagnetic Form Factor (EMFF). Form Factors, defined as a function of the momentum transfer squared q^2 , quantify the deviation from the point-like particle. Since the Electromagnetic Form Factor is a function of the momentum squared q^2 transferred to the baryon via virtual photon [6], EMFFs are either *space-like* or *time-like*. Space-like EMFFs refers to $q^2 \leq 0$ and they are studied in electron-baryon ($e^- B \rightarrow e^- B$) elastic scattering, as illustrated in Fig. 2.1(a). Through the space-like electric G_E and magnetic G_M form factors, one can retrieve the charge and magnetization densities [11]. This is because in the so-called Breit-frame, these densities are the Fourier transforms of the form factors. Time-like EMFFs, for which $q^2 > 0$, are accessible in electron-positron annihilation with subsequent baryon production ($e^+e^- \rightarrow B\bar{B}$), as shown in Fig. 2.1(b). Due to the limited lifetime of hyperons, it is hard to use them as beam or target in electron scattering experiments and therefore, their space-like EMFFs are not directly accessible. Instead, the time-like EMFF provides the most feasible structure observables for hyperons [12].

Some information about the structure is accessible also for very small data samples. One can then define a function as a linear combination of



(a) Space-like process

(b) Time-like process

Figure 2.1: Space Like process (left) and Time-like process (right)

$G_E(q^2)$ and $G_M(q^2)$. This is denoted the *effective form factor* for spin 1/2 baryons, and can be defined in the following way, eq. 2.1:

$$|F|^2 = \frac{2\tau|G_M|^2 + |G_E|^2}{2\tau + 1} \quad (2.1)$$

with $\tau = 4M_B^2/q^2$, M_B being the mass of the baryon. To separate the electric and the magnetic form factor, the modulus of their ratio can be measured from the angular distribution of the produced baryon. The ratio is defined in eq. 2.2:

$$R = \frac{|G_E(q^2)|}{|G_M(q^2)|} \quad (2.2)$$

Defining $\eta = \frac{\tau - R^2}{\tau + R^2}$, with $-1 < \eta < 1$, the ratio of the electromagnetic form factor [13] can be determined as in eq. 2.3:

$$R = \sqrt{\tau} \sqrt{\frac{1 - \eta}{1 + \eta}} \quad (2.3)$$

Thanks to these definitions it is possible to express the electric and magnetic form factors as in eq. 2.4 and 2.5:

$$|G_E|^2 = R^2 \frac{2\tau + 1}{2\tau + R^2} |F|^2 \quad (2.4)$$

$$|G_M|^2 = \frac{2\tau + 1}{2\tau + R^2} |F|^2 \quad (2.5)$$

In the time-like region, the electric and magnetic form factors are complex functions:

$$G_E(q^2) = |G_E(q^2)|e^{i\Phi_E} \quad (2.6)$$

$$G_M(q^2) = |G_M(q^2)|e^{i\Phi_M} \quad (2.7)$$

The individual phase of the electric and magnetic form factors seen in eq. 2.6 and 2.7 does not manifest in experimental observables. However, the relative phase, as defined in eq. 2.8, can be measured.

$$\Delta\Phi = \Phi_M - \Phi_E \quad (2.8)$$

The phase $\Delta\Phi$ reflects the fluctuations of the γ^* into a intermediate state of e.g. two pions [14]. The phase is experimentally accessible since it has a polarization effect on the final state baryons, even if the initial state is unpolarized. In hyperon decays, the polarization manifests in the angular distribution of the daughter particles, hence revealing the relative phase. At a certain scale of momentum transfer squared, i.e. when $q^2 \rightarrow \infty$, the time-like and space-like form factor converge to the same real value, hence with zero relative phase. From this asymptotic behaviour of the EMFF phase, it is possible to extract information about the space-like quantities [6, 15, 16]. This information allows to extract both dynamical and static properties of the Λ hyperon such as the charge radius [17], still unknown.

2.2 $\Lambda\bar{\Lambda}$ final state in e^+e^- annihilation

The Belle II experiment offers the opportunity to study the hyperons EMFFs through the ongoing data campaign aiming at an integrated luminosity of $50ab^{-1}$ primarily at a center-of-mass (CMS) energy of 10.58 GeV. The initial state is characterized by the annihilation $e^+e^- \rightarrow \gamma^*$. At this energy range, the virtual photon couples significantly to the $\Upsilon(4S)$ resonance, that has the same quantum numbers as the photon. The $\Upsilon(4S)$ decays subsequently into other final states. The most important decay chain for this analysis is the $\Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$. In this way, the Λ hyperon production can be studied through the channel $e^+e^- \rightarrow \Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$. Since the branching fraction for this decay is $<< 0.1\%$, the resulting data sample is expected to be small.

Fortunately, through electron-positron annihilation, it is still possible to study Λ production in a similar channel, thanks to *Initial State Radiation* (ISR) in the process $e^+e^- \rightarrow \gamma_{ISR}\Lambda\bar{\Lambda}$. Before the collision, either the e^+ or

the e^- can radiate one photon. The radiation reduces the effective center-of-mass energy of the e^+e^- annihilation, thereby opening up for studying $\Lambda\bar{\Lambda}$ production in the continuum, i.e. through non-resonant $\gamma^* \rightarrow \Lambda\bar{\Lambda}$ processes. The probability of a beam particle of radiating the γ_{ISR} is described by the radiator function $W(s, x, \theta_\gamma)$, in eq. 2.9, where x represents the fraction of energy carried by the ISR photon and θ_γ the emission angle in the center of mass system [18] [19].

$$W(s, x, \theta_\gamma) = \frac{\alpha}{\pi x} \left(\frac{2 - 2x + x^2}{\sin^2 \theta_\gamma} - \frac{x^2}{2} \right) \quad (2.9)$$

The angular distribution of the ISR photon is described by :

$$P(x, \theta_\gamma) = \frac{\sin^2 \theta_\gamma - \frac{x^2 \sin^4 \theta_\gamma}{2(x^2 - 2x + 2)}}{(\sin^2 \theta_\gamma + \frac{m^2}{E^2} \cos^2 \theta_\gamma)^2} - \frac{\frac{m^2 (1-2x) \sin^2 \theta_\gamma - x^2 \cos^4 \theta_\gamma}{E^2 (x^2 - 2x + 2)}}{(\sin^2 \theta_\gamma + \frac{m^2}{E^2} \cos^2 \theta_\gamma)^2} \quad (2.10)$$

where m is the electron mass and x the fraction of energy carried away by the ISR [20]. In Fig. 2.2 is shown the expected angular distribution for the ISR photon according to the formula 2.10 after integrating for x from 0 to 1. The ISR process probes to the continuum, i.e. non-resonant $\gamma^* \rightarrow \Lambda\bar{\Lambda}$, which in turn gives access to the form factors.

Comparing the two different processes for the $\Lambda\bar{\Lambda}$ production can help to understand the production mechanism in e^+e^- annihilation.

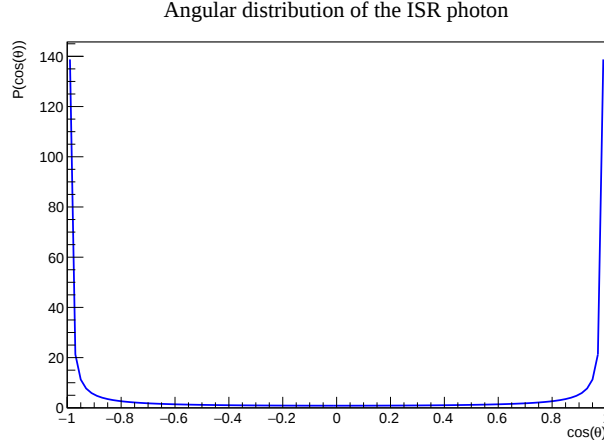


Figure 2.2: expected angular distribution of the ISR photon.

2.2.1 Tagged vs untagged ISR

Thanks to the properties of e^+e^- colliders, the initial state is well-known. This makes it straight-forward to access final state variables through missing kinematics. In ISR processes, such as $e^+e^- \rightarrow \gamma_{ISR}\Lambda\bar{\Lambda}$, the final state may be reconstructed in two approaches. The first consists on tagging the ISR photon. It means to reconstruct the photon in the final state and use it together with one of the two Λ hyperons in the final state to perform a decomposition of the full process. In the second approach, referred to as untagged, the ISR photon is not reconstructed from detector signals. Instead, the full process can be determined by reconstructing both particles in the $\Lambda\bar{\Lambda}$ pair in the final state. Whether the ISR photon can be tagged or not affects the choice of strategy to reconstruct the final state. In Chapter 7, I will discuss the efficiency of reconstruction of the ISR photon in Belle II experiment. In Chapter 8, the focus is on Λ reconstruction, to be used in both the tagged and the untagged case. This is why the aim of this thesis is to build a reconstruction algorithm for the Λ hyperon.

Chapter 3

The Belle II experiment

The Belle II experiment [21] at the SuperKEKB [22] e^+e^- collider represents the second generation of B-Factory. Successor of Belle and BaBar experiments, Belle II collects data on its journey towards and possibly beyond the Standard Model frontiers. To achieve these ambitious goals the experiment is designed to reach a higher luminosity with respect to the predecessor at KEKB [23]. In this chapter I am going to describe the SuperKEKB collider and the Belle II detector.

3.1 SuperKEKB

The new generation of B-Factory is represented by SuperKEKB, the upgraded version of the collider KEKB, located at KEK Laboratory in Tsukuba, Japan. The upgrades carried out allows for a significant improvement of the instantaneous luminosity, from $2.1 \cdot 10^{34} \text{cm}^{-2} \text{s}^{-1}$ (KEKB) to $6.5 \cdot 10^{35} \text{cm}^{-2} \text{s}^{-1}$ (SuperKEKB). This upgrade allows to collect the statistics necessary for the physics purposes of the experiment: searching for new physics (NP) in the flavour sector at the intensity frontier improving the precision of some Standard Model parameters measurement. SuperKEKB is an e^+e^- asymmetric circular collider, composed by a 7.007 GeV High-Energy Ring (HER) for the electron beam, a 4.000 GeV Low-Energy Ring (LER) for the positron beam and an inject linear accelerator (LINAC) with a 1.1 GeV positron Damping Ring. SuperKEKB is designed to have only one Interaction Point (IP). The beam asymmetry produces a Lorentz boost factor of the center-of-mass system equal to $\beta\gamma \approx 0.28$, while KEKB had $\beta\gamma \approx 0.425$. Thus, Belle II

needs a more precise track reconstruction with respect to the one achieved by Belle, in order to have a comparable resolution in the measurement of the decay time of primary particles. The accelerator is designed to collide the electron and positron beams at the center-of-mass energy in the region of the Υ resonances. The center-of-mass energy range available is between $9.46 \text{ GeV} < \sqrt{s} < 11.24 \text{ GeV}$ covering the $\Upsilon(1S)$ and $\Upsilon(6S)$ resonances. The designed peak luminosity of $8 \cdot 10^{35} \text{ cm}^2 \text{ s}^{-1}$ is a goal to achieve over time through the gradual increase of beam currents and the optimization of the accelerator parameters [24]. A schematic view of SuperKEKB collider is shown in Fig. 3.1.

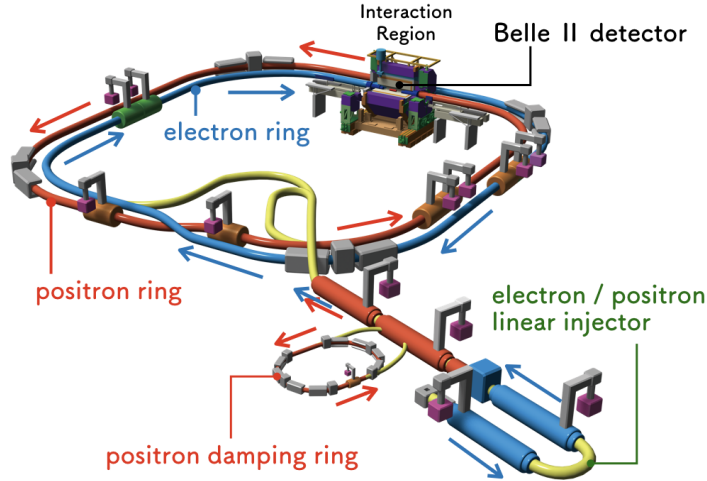


Figure 3.1: Schematic view of the SuperKEKB collider [22].

3.1.1 Nano-Beam Scheme

The main difference between KEKB and SuperKEKB is the improvement of the luminosity thanks to the *Nano-Beam* scheme [25]. The idea was to squeeze the vertical beta function at the IP (β_y^*) by minimizing the longitudinal size of the overlap region of the two beams at the IP, which generally limits the effective minimum value of β_y^* through the *hourglass effect*. The size of the overlap region d , considered as the effective bunch length for the Nano-Beam scheme, is much smaller than the bunch length σ_z . The length d

is determined by the horizontal half crossing angle ϕ and the horizontal beam size at the IP, σ_x^* , via the equation 3.1:

$$d \cong \frac{\sigma_x^*}{\phi} \quad (3.1)$$

The hourglass condition is expressed as in 3.2:

$$\beta_y^* > d \quad (3.2)$$

instead of the usual head-on collision as in eq. 3.3:

$$\beta_y^* > \sigma_z \quad (3.3)$$

To reduce the length d , a relatively large horizontal crossing angle ϕ , an extremely small horizontal emittance σ_x and a horizontal beta function at the IP for the electron and positron beams are required. Assuming flat and equal size in the horizontal and vertical dimension for both beams at the IP, the luminosity of the collider is express by the formula 3.4, where the suffix $\pm$ specifies the positron (+) or the electron (-):

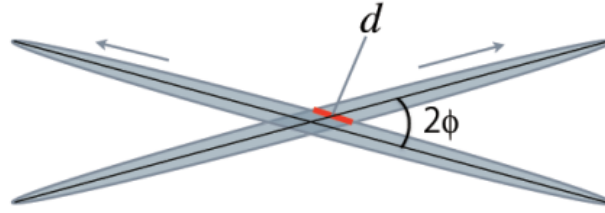
$$L = \frac{\gamma_{\pm}}{2er_e} \left(\frac{I_{\pm}\xi_{y\pm}}{\beta_y^*} \right) \left(\frac{R_L}{R_{\xi_y}} \right) \quad (3.4)$$

with γ as a Lorentz factor, e the elementary electric charge and r_e the electron classical radius. R_L is the reduction factor and R_{ξ_y} is the vertical beam-beam parameter, which arise from the crossing angle and the hourglass effect. Since the ratio between these two factor is not so far from the unit, the luminosity is mainly determined by the three fundamental parameters; i.e. the total beam current I , the vertical beam-beam parameter ξ_y and the vertical beta function at the IP, β_y^* . The choice of these three parameters, the beam energy and the luminosity is shown in Tab. 3.1. For more details see *The Belle II Book* [24]. In Fig. 3.2 the Nano-Beam scheme is summarized.

3.2 Belle II detector

The Belle II detector has been designed as an upgrade of the Belle detector. It is a general purpose experiment, optimized for the reconstruction of $\Upsilon \rightarrow B\bar{B}$ events performing efficiently on the tracking of charged particles. In addition, the detector is composed by a neutral identification system

Parameters	SuperKEKB
Energy (GeV) (LER/HER)	4.0/7.0
ξ_y	0.090/0.088
β_y^* (mm)	0.27/0.41
I (A)	3.60/2.62
Luminosity ($10^{34} cm^{-2} s^{-1}$)	80

Table 3.1: Fundamental parameters of SuperKEKB [25].**Figure 3.2:** Schematic view of beam collision in the Nano-Beam scheme. Figure from [26].

and a multi-detector Particle Identification system (PID). The high resolution momentum and vertex reconstruction allow precise time-dependent measurement. Furthermore, the knowledge of the initial state allows to perform missing energy analysis, useful to reconstruct events when the final state information is not complete. A detailed description of the Belle II detector may be found in [24] and [25]. In Fig. 3.3, Belle II detector is shown. As represented in the Fig. 3.3, the detector features an approximate cylindrical symmetry around the z-axis, defined along the direction of the electron beam. From the innermost to the outermost sub-detector system Belle II is composed of:

- Pixel Detector (PXD): 2 layers of pixel sensors.
- Silicon Vertex detector (SVD): 4 layers of Double-sides Silicon Strip sensors; together with the PXD it forms the Vertex Detector (VXD).
- Central Drift Chamber (CDC): helium-ethane wire drift chamber, composed of 56 layers with "stereo" geometry.

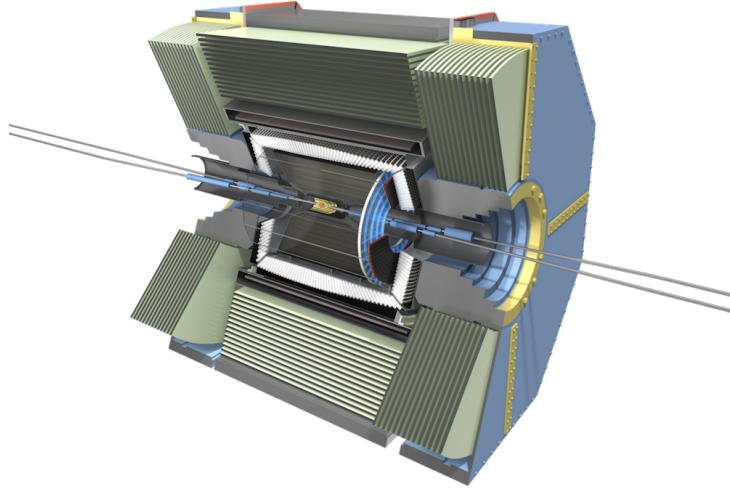


Figure 3.3: Schematic view of the Belle II detector [26].

- Particle Identification System (PID): composed by a Time-Of-Propagation (TOP) counter with a quartz radiator for the barrel region with a Cherenkov quartz radiator, and an Aerogel Ring-Imaging Cherenkov (ARICH) detector, with an areogel radiator for the forward endcap region.
- Electromagnetic Calorimeter (ECL): a homogeneous calorimeter composed of more than 8000 CsI(Tl) crystals that provide 16.1 radiation lengths X_0 .
- Superconducting magnet: a NbTi/Cu magnet that provides a homogeneous magnetic field of 1.5 T parallel to the beam direction in the internal region.
- K_L and Muon Detector (KLM): composed by alternated layers of Resistive Plate Chambers (RPCs) and iron plates in the outermost barrel region and scintillator strips in the innermost barrel region and in both endcaps.

Vertex Detector (VXD)

The vertex detector of Belle II is composed by two devices: the Pixel Detector PXD and by the Silicon Vertex Detector SVD, with altogether six layers

around the beam pipe with radius 10 mm in Beryllium. The first layers at $r = 14$ mm and $r = 22$ mm compose the PXD. They are pixel sensors with monolithic structure, DEPFET type [26] [27]. The remaining four layers at radii of 38 mm, 80 mm, 115 mm and 140 mm form the SVD and are double-side silicon strip sensor. Getting closer to the interaction point the beam pipe and the first two detector layers, the vertex resolution is significantly improved compared to the one of the Belle vertex detector.

Central Drift Chamber (CDC)

The central Drift Chamber is formed by a large volume occupied by small drift cells with radius 1130mm. In total it contains 11336 sense wires arranged in 56 layers in *axial*, i.e. aligned with the solenoidal magnetic field, or *stereo*, i.e. skewed with respect to the axial wires. By combining information from axial and stereo layers we are able to reconstruct a full 3D helix track. Together with the VXD, the CDC makes up the Belle II tracking system with an angular acceptance of $17^\circ < \theta < 155^\circ$.

Particle identification system (PID)

The particle Identification system is made up by the time-of-propagation (TOP) counter and by the Cherenkov ring imaging detector (ARICH)[28]. The former still exploits the principle of a normal Cherenkov detector. It is 2.6 m long quartz bar, at the end of which a photo-detector collect the time of arrival and the impact position of the Cherenkov photons. Measuring these two quantities one can reconstruct the 2D information of a Cherenkov ring image. The latter, placed in the forward endcap region, is a proximity focusing Cherenkov ring imaging detector with aerogel as Cherenkov radiator, employed to identify charged particles. The PID system is of enormous importance for the experiment since it allows to distinguish between final state hadrons and leptons.

Electromagnetic Calorimeter (ECL)

The electromagnetic calorimeter is used to detect gamma rays as well as to identify electrons, i.e. separate electrons from hadrons, in particular pions. The detector region, composed by barrel as well as the forward and backward endcaps, is instrumented with a total of 8736 CsI(Tl) crystals, covering about 90% of the solid angle in the centre-of-mass system. The calorimeter covers

the polar angle region of $12.4^\circ < \theta < 155.1^\circ$, except for two gaps $\sim 1^\circ$ wide between the barrel and endcaps. The intrinsic energy resolution of the calorimeter, as measured in a prototype [29], can be approximated as eq. 3.5:

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{0.066\%}{E}\right)^2 + \left(\frac{0.81\%}{\sqrt{4}E}\right)^2 + (1.34\%)^2} \quad (3.5)$$

where E is the energy in GeV, while the first term represents the electronics noise contribution [25].

K_L - Muon detector (KLM)

The K_L and muon detector (KLM) consists of an alternating sandwich of 4.7 cm thick iron plates and active detector elements located outside the superconducting solenoid. The iron plates serve as the magnetic flux return for the solenoid. They also provide 3.9 interaction lengths λ_0 or more of material, beyond the 0.8 interaction lengths λ_0 of the calorimeter, in which K_L mesons can shower hadronically. The KLM covers all three regions of the detector with a total polar angle acceptance of $20^\circ < \theta < 155^\circ$, providing the K_L reconstruction and the muon identification.

3.3 Software tools

The analysis of Belle II experiment is done through basf2 software [30]. Coding in Python, it allows to generate, reconstruct and analyse physical processes of interest. The software is optimized for Belle II experiment and make easier the analysis chain for our purposes.

3.3.1 Analysis chain

The process starts with the generation of the particles inside the detector. According to the particles to produce, we use the appropriate generator to generate the *event*, i.e. the particle of interest. The generation simulates the particle production after the e^+e^- collisions. The geometry of the detector and the propagation of the particles in the system are already implemented in the software. What we get after the generation is a collection of .root files containing raw data about what the entire detector has detected. The

reconstruction is the step after. In this phase, the events of interest are reconstructed through the detected final particles. In basf2 software this procedure is called *steering*. After saving in a .root file some kinematics variables related to the particles of interest, the offline analysis may start. Thanks to these variables, I used Machine Learning tools to distinguish which events from the data sample are *signal* or *background*. Signal are those events correctly reconstructed by the reconstruction algorithm, i.e. that follows the decay rules written for that event. Background are those events wrongly reconstructed. The main task of my thesis is to improve the current reconstruction algorithm for the Λ hyperons increasing the Λ hyperon selection over the background.

Chapter 4

Machine Learning (ML)

Identifying Λ hyperons from detector signals is the purpose for this analysis. By this, we mean selecting those events that contain good Λ candidates and classify them as signal. This is done by a reconstruction algorithm that classifies the events by giving each data point a *label*: signal or background. The signal are the points identified as true i.e. events with Λ hyperons, while the background are the events identified as false [31]. To facilitate the Λ hyperon selection, I have in my analysis utilized a machine learning approach. In this chapter I am going to introduce some central concepts.

Machine Learning is a branch of artificial intelligence that uses statistical techniques to enable machines to recognize specific patterns by learning and improving through experience on a set of provided data [32]. Instead of combining a set of human-created rules with data to create answers to a problem, as is done also in conventional programming, machine learning uses data and labels to discover the rules behind a problem [33] as represented on the figure 4.1. Each data point provides several *features*, set of attributes

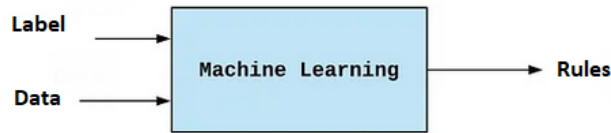


Figure 4.1: Conceptual map of how machine learning works.

often represented as a vector [31]. The dimension of the vector represents the

number of features taken into account. If the algorithm handles data with more than one attribute, the analysis is called *multi-variate analysis*.

In the learning process the machine takes in input data points and their labels to determinate their correlations. The results are rules connecting the data points to their labels. To learn the rules governing a phenomenon, the machines must go through a learning process trying different rules and learning from how well they perform. The learning problem in fact consists of identifying from the set of candidate formulas the hypothesis that best matches the *unknown target function*, representing the ideal formula [33]. The target function has defined as $f : X \rightarrow Y$, where X is the input space containing all the features and Y the output space, represented by the labels.

4.1 Learning types

To understand how machine learning works, it is necessary to explore different machine learning methods and types. *Supervised learning* makes predictions based on labeled *training* data. Training is the phase where the machine learns from data. Each training sample includes an input and a desired output. A supervised learning algorithm analyzes this sample data and makes an inference of the labels for future analysis of unseen data [32]. It is used for classification i.e. it requires to make a choice among two or more classes to be assigned to the data, usually providing a probability for each class, also known as *classifier distribution*. After the training, the model can be applied to new data samples to classify their points. The best configuration for a model is when it is applied to unseen data and it still performs the classification well, hence, the algorithm is able to *generalize* new data. As output, the algorithm gives a classifier distribution that is used to distinguish between two classes. Supervised learning is considered appropriate for classification problem.

There are other two approaches in addition to supervised learning: *unsupervised learning* and *reinforced learning*, but they are outside the scope of this thesis.

4.2 Model Optimization

To reach the best generalization of trained data, it is necessary select a hypothesis as close as possible to the unknown target function. During the learning phase, the algorithm evaluates the *in-sample error*, defined through the misclassification between the prediction of the data points and their labels. The in-sample error takes into account how well the algorithm learns from data. During the training phase, the goal is to find the configuration that minimizes the in-sample error.

To estimate the performances of the model, it is fundamental evaluate the *out-of sample error*. The latter quantifies how well the model has generalized unseen data. The out-of-sample error based on the performance over the entire input space X . It is estimated through a sample of data points that have not used for training [33]. This process constitutes the *testing phase*. Training and testing have the purpose to learn how the algorithm works and performs. The scope of the training phase is to identify the configuration that minimizes the out-of sample error [33] i.e. the best generalization. A configuration is composed by a choice of parameters inside the algorithm, by the features to analyse and by the complexity of the model. Changing a few of the parameters of a configuration can have a large impact on the result.

Choosing the configuration associated with the lowest level of in-sample error can cause worse performances of the out-of sample error. To understand how well the algorithm generalizes unseen data, we need the *learning curves*, quantifying the expected error during the training and the evaluation for the configuration used. Achieving a balance between the two curves is the best strategy to avoid the *overfitting* or the *underfitting*, as in figure¹ 4.2. The former is the configuration where the algorithm has modelled fluctuations and noise on the training data. The latter is verified in the opposite situation i.e. where the model doesn't match to the points.

¹Picture adapted from the University of Turin course "Data Mining, Statistical Modeling and Machine Learning", with professor Sanavia, Panisson, Gauvin.

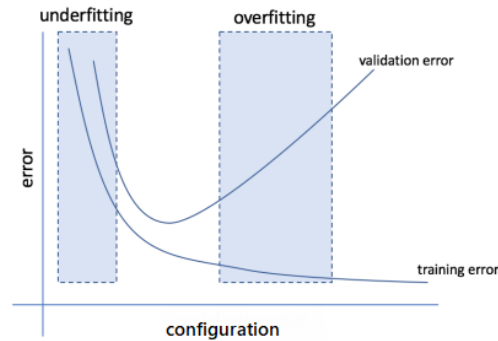


Figure 4.2: The picture shows how the overfitting occurs for the different choices of in-sample Error.

4.3 Workflow

Studying data with machine learning tools needs a clear workflow to follow during the entire process. In the first *data collection* step, the training data are prepared for the algorithm to learn from. A proper preparation of data is crucial for the performance. For this reason the *data preparation* is the second step. It is a pre-processing of the data into the optimal format, extracting the most important features and reducing the number of input features. The step after is the training. In order to optimize the performances of the algorithm there are the *tuning* and the *validation*. They consist to change parameters inside the model and see the performance through the learning curves. The final step is the *evaluation* where we test the model on independent data that it has not seen before. The entire procedure is resumed on the figure² 4.3.

²Picture adapted from the University of Turin course "Data Mining, Statistical Modeling and Machine Learning", with professor Sanavia, Panisson, Gauvin.

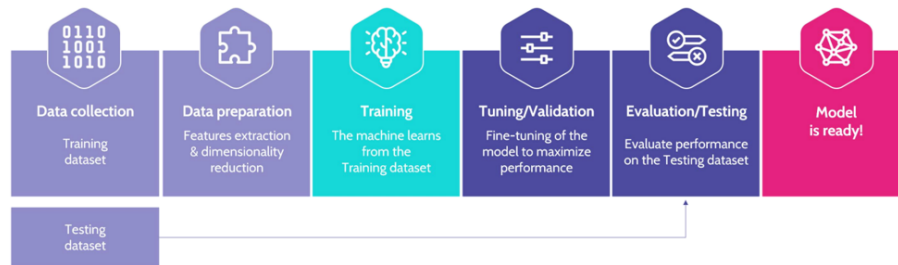


Figure 4.3: Workflow for a MVA analysis using machine learning tools.

4.4 Evaluation

In a classification problem the algorithm provides its prediction for each point in the input data. During the evaluation process, the binary classification predicted by the algorithm is compared with the labels of the provided points. The points classified as 1 are called *positive* and the points classified as 0 are called *negative*. If the predictions and their labels match, the classification will be classified as *True*, if not that will be classified as *False*.

In a classification problem there are four different cases to take into account:

- *True Positive* (TP): the classifier predicted positive is correctly classified.
- *True Negative* (TN): the classifier predicted negative is correctly classified.
- *False Positive* (FP): the classifier predicted positive is not correctly classified.
- *False Negative* (FN): the classifier predicted negative is not correctly classified.

These definitions form the *confusion matrix* [34], showed in the figure 4.4.

		True Class	
		Positive	Negative
Predicated Class	Positive	TP	FP
	Negative	FN	TN

Figure 4.4: Structure of the confusion matrix.

Through the confusion matrix, it is possible to define some relevant variables used in classification problems in physics. Below the definitions of *signal efficiency*, *background rejection* and *purity*.

$$\text{signal efficiency} = \frac{TP}{TP + FN} \quad (4.1)$$

$$\text{background rejection} = \frac{TN}{TN + FP} \quad (4.2)$$

$$\text{purity} = \frac{TP}{TP + FP} \quad (4.3)$$

To get a binary classification from the classifier distribution, one needs to set a threshold, dividing the distribution in two zones. The points over the threshold are classified as positive, while points below the threshold are classified as negative [34]. An example of classifier distribution is illustrated in Fig. 4.5.

To evaluate the performance of the algorithm, tools in binary classifiers called *metrics* are applied. A central metric is a *receiver operating characteristic* (ROC), that represents the background rejection versus the signal efficiency plot [34]. For a purely random classifier, the ROC curve is represented as a line from the top-left corner to the bottom-right corner; a good classifier stay as far away from the diagonal line as possible, toward the top-right corner [34]. Another metric is the *area under the curve* (AUC). A

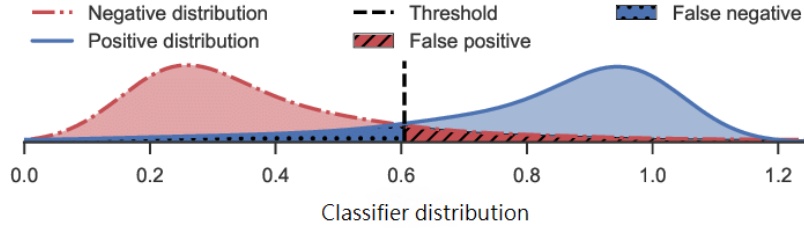


Figure 4.5: How the classifier distribution is represented in output from a model. The threshold has the role to classify a point as positive or negative. Figure taken from [35].

perfect classifier will have a *ROC AUC* score equal to 1, whereas a purely random classifier will have a ROC AUC equal to 0.5. An example of ROC curve and its area is illustrated in Fig. 4.6.

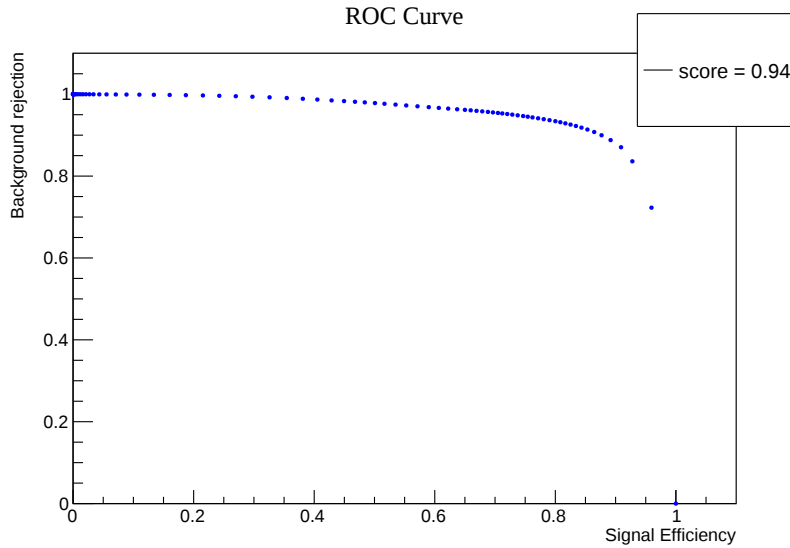


Figure 4.6: ROC curve example. On x-axis we set the signal efficiency, on y-axis the background rejection.

4.5 Decision trees

Decision tree is one of the simplest and most successful forms of machine learning. The algorithm represents a function that takes as input a vector of attribute values and returns a *decision* i.e. a single output value. The name "tree" refers to its structure: root node, branches and leaves. As illustrated in Fig. 4.7, the root node is the node on the top of the tree, represented by the blue node, then the branches are represented as a line that links two subsequent nodes and, when a node has no more branches, it is called leaf. Decision trees classify instances by sorting them down the tree from the root to some leaf nodes, which provide the classification of the instance [36].

A decision tree reaches its decision by performing a sequence of tests. Each internal node in the tree, represented in the Fig. 4.7 in red, green, purple and yellow, corresponds to a test of the value of one of the input features. An instance is classified by starting at the root node of the tree, testing the feature specified by this node, then moving down the tree branch corresponding to the value of the feature in the given example. This process is then repeated for the subtree rooted at the new node [36].

The algorithm splits the population or sample into two or more homogeneous sets based on most significant attributes i.e. features in input variables. There are different ways to split:

- i. *Two-way split*: Divides the data into two subsets.
- ii. *Multi-way split*: Divides the data into many subsets.

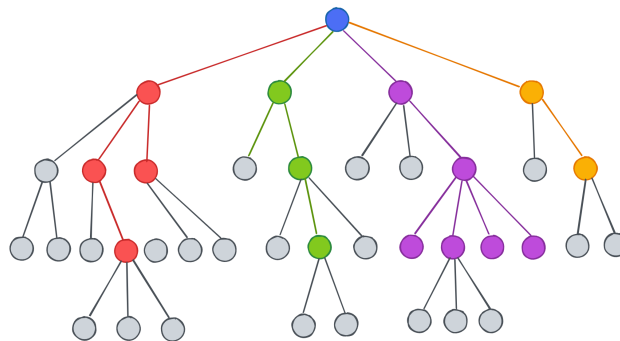


Figure 4.7: Conceptual map of how decision tree appears. Figure taken from [37]

The model is composed by a sequence of selection criteria applied according to a specific order. Every criterion separates the sample in nodes. If the sample contains events of signal or background, the model aims to separate them. Each node represents a fraction of signal event. The nodes with a higher fraction of signal are called leaves and no other splits are done.

A single decision tree has low performance. The performance can be improved by combining a set of decision trees forming a *decision forest*, usually containing hundreds of trees. This procedure is called *boosting*, a process where each individual tree is built sequentially iterating over the previous one. In this approach each model in the sequence is fitted giving more importance to observations in the dataset that were poorly handled by the previous models in the sequence. Each new model focuses its efforts on the most difficult observations obtained at the last iteration. This process helps to decrease the bias, i.e. the difference between the unknown target function and the one obtained through the model [38].

Chapter 5

Analysis Strategy

5.1 Strategy

The purpose of the project is to develop tools for Λ hyperon Form Factors with Belle II. The main challenge is to select Λ hyperons from a data sample where the reaction of interest is relatively rare compared to, for example, B meson events. As a first step we must know how to generate the Λ hyperons in Belle II. Hence, we have to study which generator is appropriate for our purpose. In Chapter 6, $\Lambda\bar{\Lambda}$ Generation, I will discuss the main channels of interest for Λ structure studies with Belle II. The most important channel for studying the form factors is the ISR process: $e^+e^- \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$. This channel is the only one that gives access to the form factors at this energy. The generation and the reconstruction of $e^+e^- \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$ is fundamental for the aim of the project. Understanding how well we can investigate the γ_{ISR} in Belle II detector constitutes a significant part of my work. I will study the performance of reconstruction of the ISR photon, i.e. the efficiency as a function of the γ_{ISR} energy and γ_{ISR} angular distribution. This analysis aims to understand the feasibility of reconstructing the ISR photon in the Belle II detector. The main focus is however to find the optimal way to reconstruct the Λ and $\bar{\Lambda}$ in the detector. I developed a reconstruction algorithm starting from the analysis of the kinematics variables of the Λ and $\bar{\Lambda}$ hyperons. By applying modern tools based on machine learning techniques, I have developed a Λ hyperon selector that can distinguish $\Lambda \rightarrow p\pi^-$ and $\bar{\Lambda} \rightarrow \bar{p}\pi^+$ events from the background in a generic Monte Carlo sample. Here, the identification relies on the reconstructed properties of the charged particles $p\pi^-$ ($\bar{p}\pi^+$). Finally,

I applied the selector to a generic hadronic Monte Carlo sample in order to study its performances on a realistic simulation of real, experimental data.

Chapter 6

$\Lambda\bar{\Lambda}$ Generation

In this chapter, I introduce the samples utilized for the analysis. The samples represent signal events, i.e. Λ hyperon production, and generic background from hadronization of $e^+e^- \rightarrow q\bar{q}$ events.

6.1 $\Lambda\bar{\Lambda}$ signal samples

In Belle II the $\Lambda\bar{\Lambda}$ pairs can be generated for example by:

- i. $e^+e^- \rightarrow \Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$
- ii. $e^+e^- \rightarrow e^+e^-\gamma_{ISR} \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$

In the following, I will discuss how $\Lambda\bar{\Lambda}$ are generated in basf2.

6.1.1 $e^+e^- \rightarrow \Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$

The generator *Evtgen* [39] is designed to simulate decays of hadrons with heavy flavour such as charm and bottom. Therefore, it is suitable for the process $\Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$ and its subsequent decays. The main decay channels of the Bottomonium are in B^+B^- (51%) pair or $B_0\bar{B}_0$ (49%) pair. Since I only want to study the case when $\Upsilon(4S)$ decays into a $\Lambda\bar{\Lambda}$ final state, I set the branching fraction to 1.0 in the generator user option file. Furthermore, I only consider the $\Lambda \rightarrow p\pi^-$ (and charge conjugate) decay since it is the most straight-forward to analyse. Therefore I set the Branching Ratio at 1.0 instead of the *Particle Data Group*(PDG) [10] value of 0.64.

In the generator, the particles decay according to the phase space, which means that only kinematics, and not dynamics, is taken into account. Hence, the $\Lambda/\bar{\Lambda}$ scattering angle is isotropically distributed. In addition, the polarization is not taken into account in the Λ and $\bar{\Lambda}$ decays, since it is not known at this energy.

The sample generated contains 100k events of $\Upsilon(4S) \rightarrow \Lambda\bar{\Lambda} \rightarrow p\pi^-\bar{p}\pi^+$ production. Through the $(\Lambda\bar{\Lambda})$ invariant mass, it is possible to identify the production from the different channels considered. In the same way, a Λ or a $\bar{\Lambda}$ manifests as a peak in the $(p\pi)$ invariant mass distribution. In the Fig. 6.1, the invariant masses $M(\Lambda\bar{\Lambda})$ is shown. In Fig. 6.2 is illustrated the invariant mass $M(p\pi^-)$. For each figure is represented the generation of the events and their reconstruction, in order to emphasize the current performance of the reconstruction algorithm.

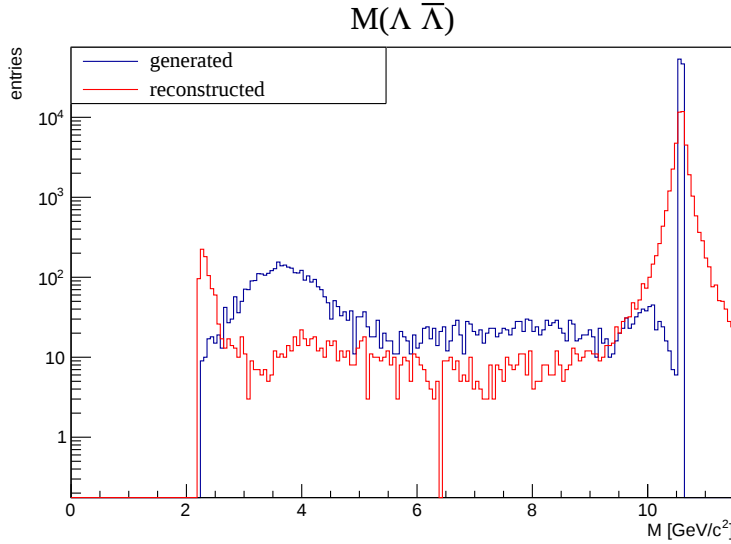


Figure 6.1: invariant mass for the $\Lambda\bar{\Lambda}$ system with a statistics of 100k events for the channel $e^+e^- \rightarrow \Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$.

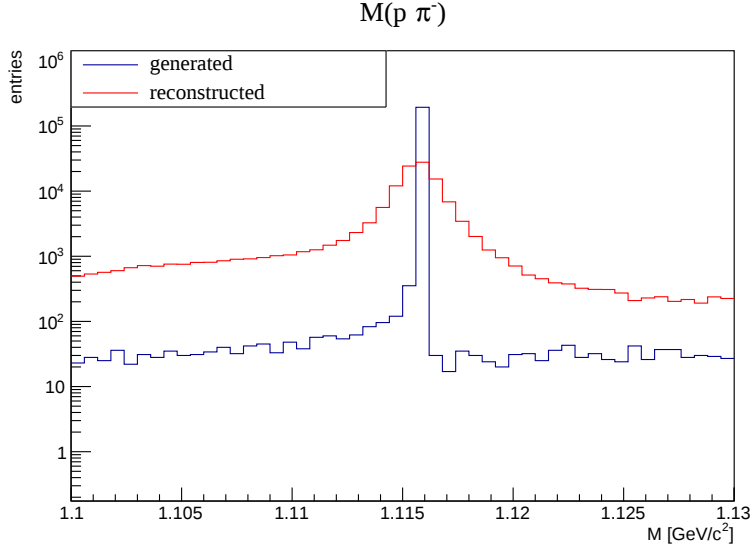


Figure 6.2: invariant mass $M(\Lambda \rightarrow p\pi^-)$ and charge conjugate system with a statistics of 100k events for the channel $e^+e^- \rightarrow \Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$.

6.1.2 $e^+e^- \rightarrow e^+e^-\gamma_{ISR} \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$

When the production of the $\Lambda\bar{\Lambda}$ pair occurs after the emission of a ISR photon, denoted γ_{ISR} , it is applied *PHOKHARA* generator [40]. Since the channel is characterized by the electromagnetic radiation in the initial state with the subsequent production of a hyperon-antihyperon pair, I use the combination of *PHOKHARA* and *Evtgen*. *Phokhara* is acting as ISR generator by generating $e^+e^- \rightarrow \mu^+\mu^-$, the muon pair being replaced by a virtual photon γ^* decays into a $\Lambda\bar{\Lambda}$ pair in *Evtgen*. As in the case of the $\Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$ process, I set the branching fraction of $\gamma^* \rightarrow \Lambda\bar{\Lambda}$ to 1.0 and the same for $\Lambda \rightarrow p\pi^-$ and $\bar{\Lambda} \rightarrow \bar{p}\pi^+$. Hence, in the final state, the particles detected are the $p\pi^-$ or $\bar{p}\pi^+$ and the γ_{ISR} . The $\Lambda\bar{\Lambda}$ pair is generated through phase space, with the Λ and the $\bar{\Lambda}$ produced isotropically in the rest system of the virtual photon. As in $\Upsilon(4S) \rightarrow \Lambda\bar{\Lambda}$, we don't take into account the polarization of Λ and $\bar{\Lambda}$ in their decays. The sample generated contains 100k events of $\Lambda\bar{\Lambda}$ production. As shown in the last section, the invariant masses $M(\Lambda\bar{\Lambda})$ and $M(p\pi^-)$ are useful to identify the events and to study the differences between the two channels. It is noteworthy that the presence of the γ_{ISR} photon carries away some of the energy from the beam.

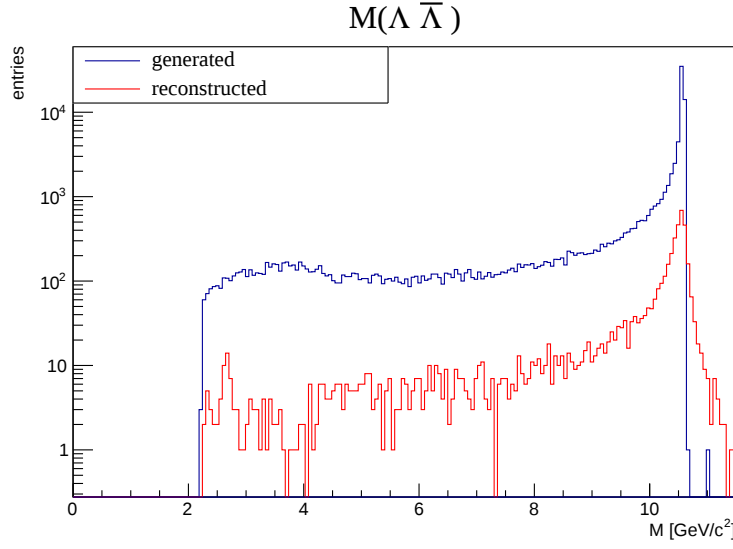


Figure 6.3: invariant mass for the $\Lambda\bar{\Lambda}$ system with a statistics of 100k events for the channel $e^+e^- \rightarrow e^+e^-\gamma_{ISR} \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$.

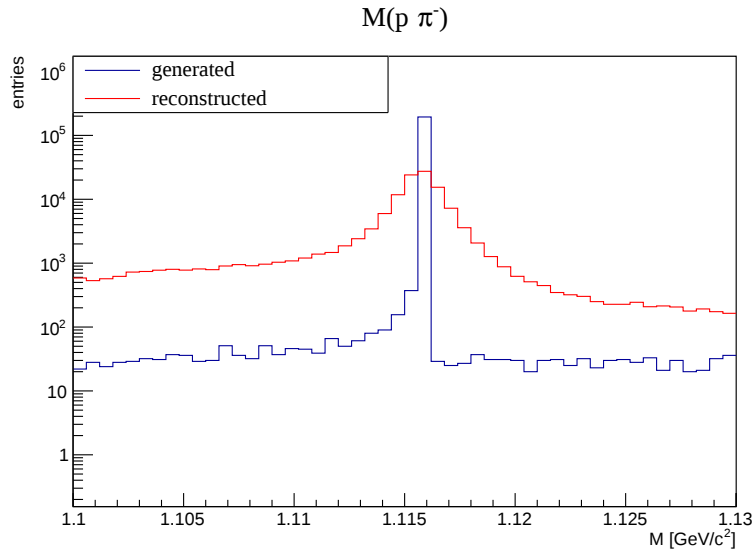


Figure 6.4: invariant mass $M(\Lambda \rightarrow p\pi^-)$ and charge conjugate system with a statistics of 100k events for the channel $e^+e^- \rightarrow e^+e^-\gamma_{ISR} \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$.

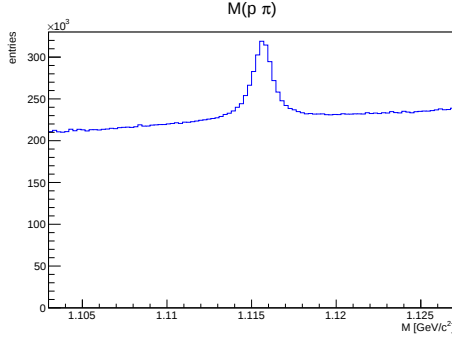
As it is shown in Fig. 6.3, the γ_{ISR} photon gives rise to a long tail to the left of the peak of the invariant mass $M(\Lambda\bar{\Lambda})$. This is the consequence of the γ_{ISR} photon reducing the collision energy. Therefore, we can effectively scan a large range of energies away from the Bottomonium production. In Fig. 6.4 the invariant mass $M(p\pi)$ for this channel is shown.

6.1.3 Generic Hadronic background

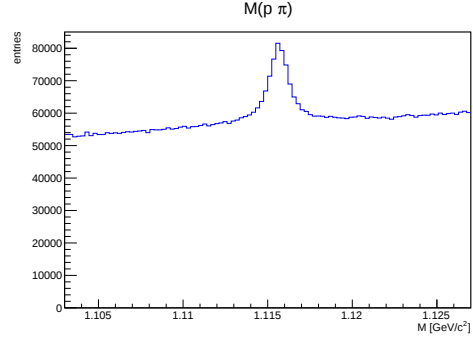
The Belle II experiment has produced generic MC samples for studies of Bottomonium physics. In my hyperon analysis, I have utilized several of these samples. Here, generic refers to the fact that the samples include everything that can be produced from the hadronization of a quark-antiquark pair produced from the electron-positron annihilation, taking into account our best knowledge on branching fractions and cross sections. These samples are produced using different generators: Bottomonium is generated through Evtgen, while the continuum, i.e. the events where particles were not created through the decay chain of a Bottomonium resonance, are generated in PYTHIA according to the Lund String model [41]. Each sample corresponds to an integrated luminosity of $L = 10fb^{-1}$. The samples considered are:

- *continuum* $q\bar{q}$, with $q = u, d, s, c$
- *charged*, i.e. B^+B^- production
- *mixed*, i.e. $B_0\bar{B}_0$ production

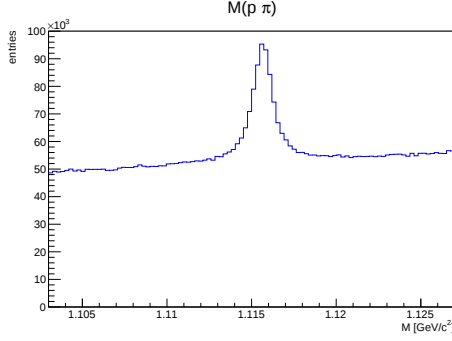
In all the aforementioned cases, hyperons and antihyperons can be produced either in the hadronization process or in the decay of a heavy flavour hadron. To compare the datasets, I plot the invariant mass of the $(p\pi)$ and c.c. system. Continuum samples and Bottomonium samples may contain background that do not correspond to Λ hyperons. Fig. 6.5 represents the Λ production through the decay from continuum, mixed and charged channel. Since the luminosity taken for each sample is the same, with $L = 10fb^{-1}$, the number of entries in each histogram is correlated to the cross section of the Λ production through the channel considered. Hence, the number of events is directly proportional to the luminosity of the collider, to the cross section of the channel and to the detector efficiency. In Fig. 6.5 we discern that the Λ hyperon production through $c\bar{c}$ hadronization is advantaged than $d\bar{d}$ due to a higher cross section.



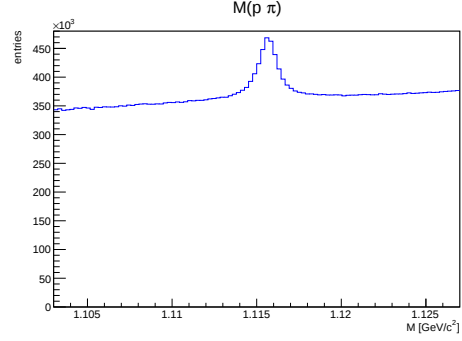
(a) $u\bar{u}$ sample



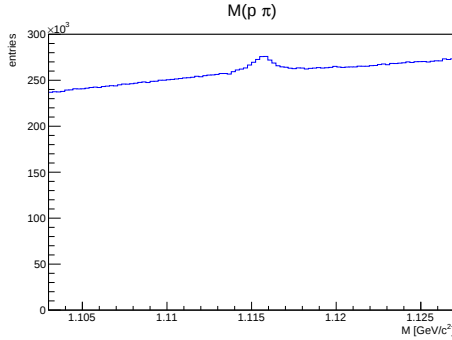
(b) $d\bar{d}$ sample



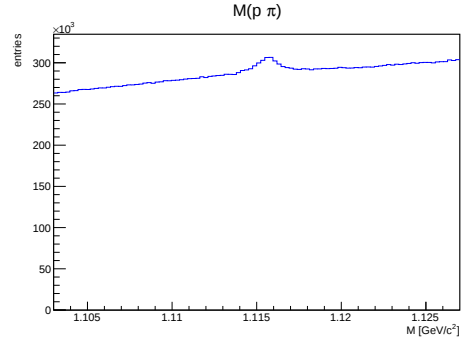
(c) $s\bar{s}$ sample



(d) $c\bar{c}$ sample



(e) *mixed* sample



(f) *charged* sample

Figure 6.5: invariant mass $M(p\pi^-)$ and c.c. system from generic Monte Carlo samples with an integrated luminosity of $L = 10fb^{-1}$

Chapter 7

Topology and reconstruction of the γ_{ISR}

Considering the Λ production through the channel $e^+e^- \rightarrow e^+e^-\gamma_{ISR} \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$, a complete reconstruction of the event means to be able to reconstruct all the particles in the final state, including the γ_{ISR} . In this chapter I present the analysis of the γ_{ISR} . The γ_{ISR} photon can be emitted by both the electron and the positron beam. The aim is to evaluate the feasibility of reconstructing the ISR photon with the Belle II detector to learn if it is possible to tag the ISR events in the final state.

7.1 Topology of the γ_{ISR}

In the previous chapter I discussed the $\Lambda\bar{\Lambda}$ generation through the two main channels focusing on the invariant mass of the final state hadrons. The question now is: how efficiently can we reconstruct the γ_{ISR} ? By comparing what the detector is able to reconstruct to what the generator has produced we can calculate the reconstruction efficiency and thereby evaluate the feasibility of hyperon studies with γ_{ISR} .

7.1.1 Energy distribution

The energy distribution of the γ_{ISR} photon is of fundamental importance, since it defines the energy of the recoiling hadronic $\Lambda\bar{\Lambda}$ system. The electron beams at SuperKEKB has an energy of 7.007 GeV while the positron beam

is produced at an energy of 4.000 GeV [24]. Through the information on the details of the generated particles, referred to as Monte Carlo Truth, it is possible to study the γ_{ISR} photon and its energy and angle. The energy distribution of the γ_{ISR} photon from the $e^+e^- \rightarrow e^+e^-\gamma_{ISR} \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$ is shown in Fig. 7.1.

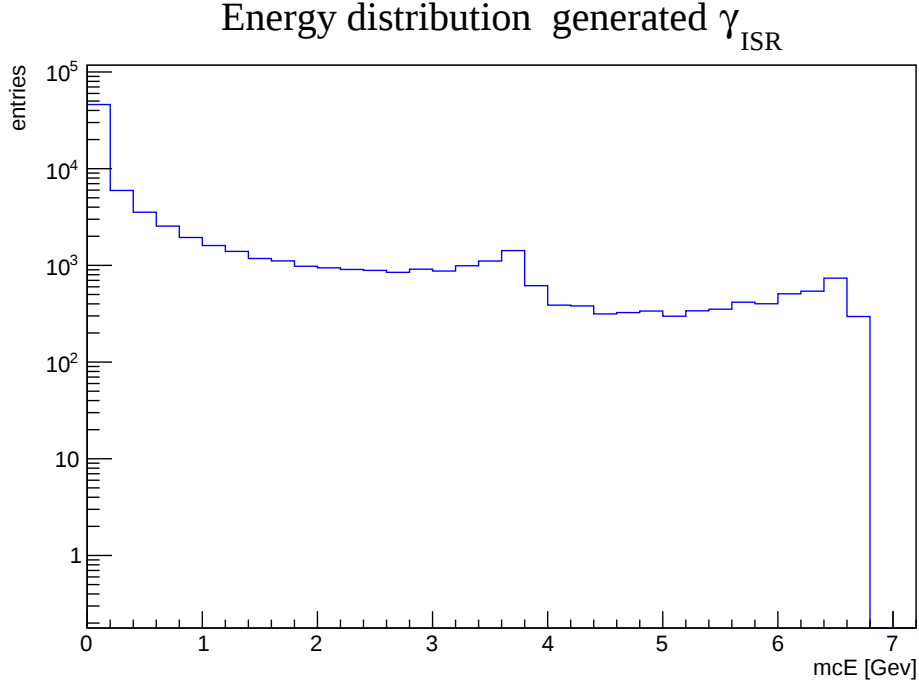


Figure 7.1: Energy distribution of the generated γ_{ISR} photons.

We can discern two different peaks: slightly below 4 GeV and 7 GeV. These correspond to the maximum energy that an ISR photon can have when emitted from the positron and electron beam, respectively.

7.1.2 Angular distribution

The next step is to study how well the detector is able to reconstruct among the polar angle: as defined between the momentum direction of γ_{ISR} and the z-axis defined by the electron beam. In particular, we would like to investigate whether the loss of events around 4 GeV in Fig. 7.1 is due to

an effect in the angular acceptance. The efficiency dependence on the polar angle gives valuable information about this.

In the Belle II experiment, there are no detectors in the forward and backward direction along the beam pipe. Hence, we expect that the reconstruction efficiency will drop for photons emitted at very small or very large angles with respect to the beam line. Starting with the generated angular

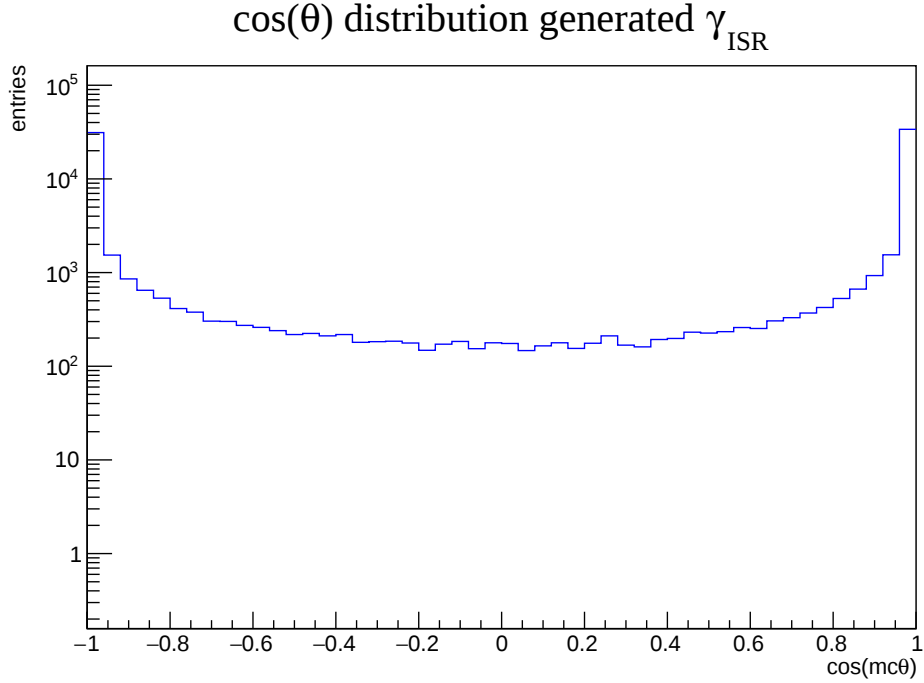


Figure 7.2: Polar angle distribution of the generated γ_{ISR} photons.

distribution of the γ_{ISR} photon, shown in Fig. 7.2, we conclude that the most of the γ_{ISR} are emitted at very small (large) angles. This is explained by the angular distribution of the ISR photon, already described in Fig.2.2.

7.2 γ_{ISR} reconstruction

The performance of the detector and its reconstruction algorithm to reconstruct the ISR photon can be evaluated comparing the MC truth and the reconstructed information. In the next section I will present the reconstruction analysis of the ISR photon.

7.2.1 Energy distribution

First, we reconstruct the energy distribution of the detected γ_{ISR} . From this, we can learn at which energies of the ISR photons the detector is missing more events. In Fig. 7.3 the energy distribution of the all ISR photons generated and reconstructed. It is evident that the two peaks seen in the

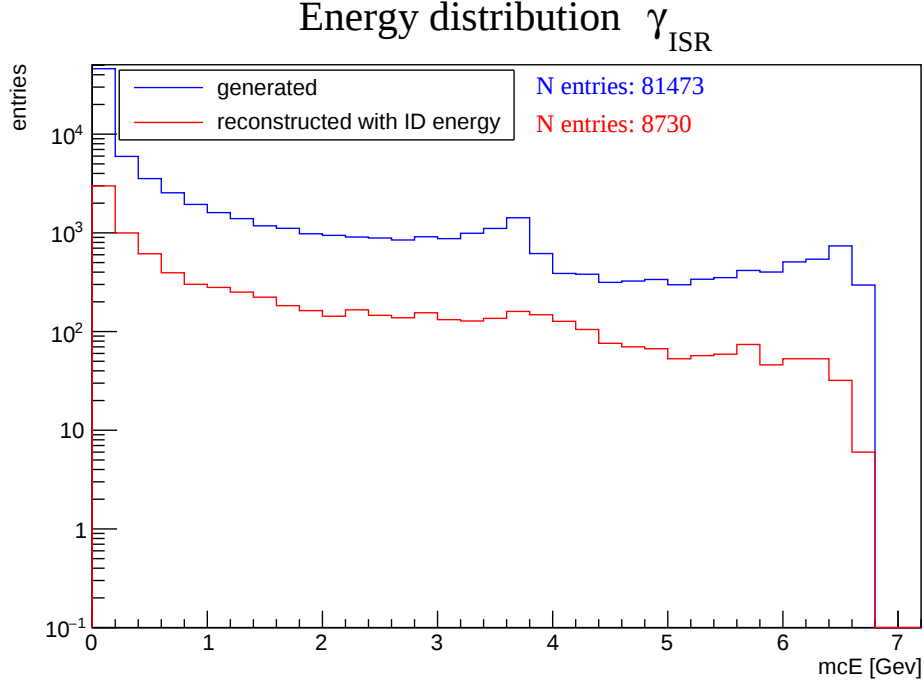


Figure 7.3: energy distribution of generated and reconstructed ISR.

MC truth distribution are smoother than in the reconstructed distribution. Moreover, the number of detected photons is smaller than the number of generated photons. This is partly because of the finite acceptance of the Belle II detector, that does not completely cover the 4π solid angle, but also because the detection and reconstruction efficiency is not perfect. The absence of detectors in the forward or backward direction due to the beam pipe seems to have a moderate impact on the shape of the energy distribution, though there are more detected photons at lower energies.

7.2.2 Angular distribution

We expect the lack of detectors in the forward direction has a strong negative impact on the performance on the reconstruction at extreme angles. The generated and reconstructed angular distribution of the ISR photons are shown in Fig. 7.4.

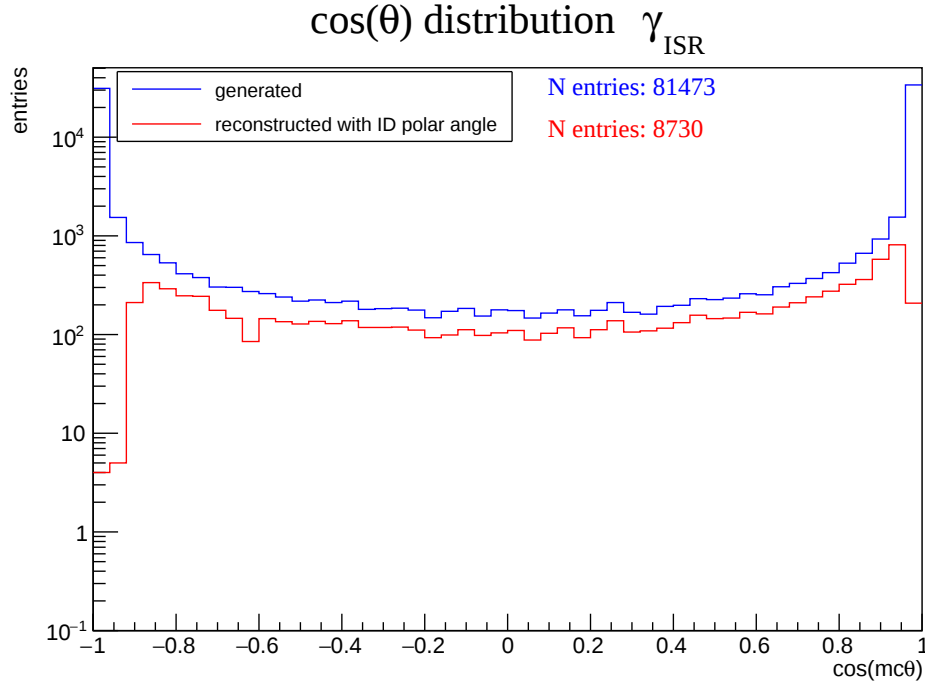


Figure 7.4: $\cos(\theta)$ distribution of generated and reconstructed ISR.

As anticipated, there are missing events at extreme angles due to the limited coverage of the solid angle of the detector. Apart from this, the angular dependence follows the shape as the generated distribution, but with much less events.

7.3 Efficiency

To study the performance of reconstruction of the γ_{ISR} we calculate the reconstruction efficiency of the *ISR* photon as a function of the energy and as a function of the angular distribution.

7.3.1 Efficiency dependence on energy

For histogrammed data defining the efficiency as the number of reconstructed events in bin i divided by the number of the generated events in the same bin:

$$\epsilon^i = \frac{N_{rec}^i(E_{gen})}{N_{gen}^i(E_{gen})} \quad (7.1)$$

The uncertainty of the efficiency can be calculated assuming that the events are binomially distributed [42] [43]:

$$\sigma_\epsilon^i = \sqrt{\frac{(N_{rec}^i + 1)(N_{gen}^i - N_{rec}^i + 1)}{(N_{gen}^i + 2)^2(N_{gen}^i + 3)}} \quad (7.2)$$

This definition takes into account the ratio of reconstructed and generated photons at the same energy. This is defined as an efficiency *without migration* i.e. without any energies distortion, along the detector. In Fig. 7.5 the

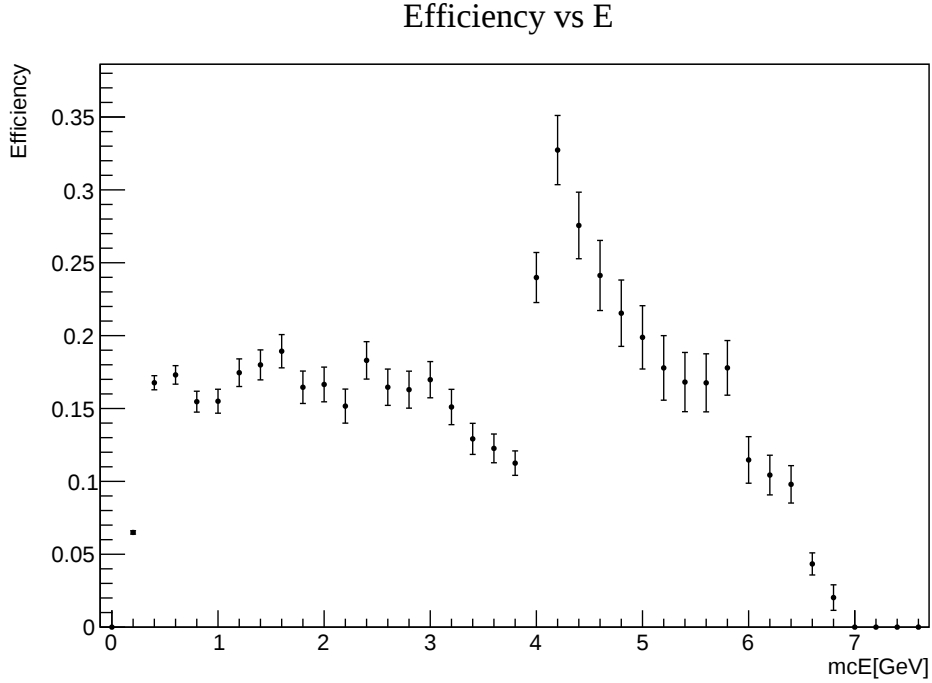


Figure 7.5: Efficiency vs energy plot.

distribution of the efficiency as a function of the energy. The efficiency level is below 30% over almost all the energy range.

7.3.2 Efficiency dependence on angle

As the next step, it is relevant to calculate the efficiency of the reconstruction as a function of the polar angle. The definition for the efficiency in bin i is:

$$\varepsilon^i = \frac{N_{rec}^i(\theta_{gen})}{N_{gen}^i(\theta_{gen})} \quad (7.3)$$

For the uncertainty, we use the same method as before. In Fig. 7.6, the distribution of the efficiency as a function of the reconstructed ISR photon angle is shown. Most of the missing events are at the edge of the distribution.

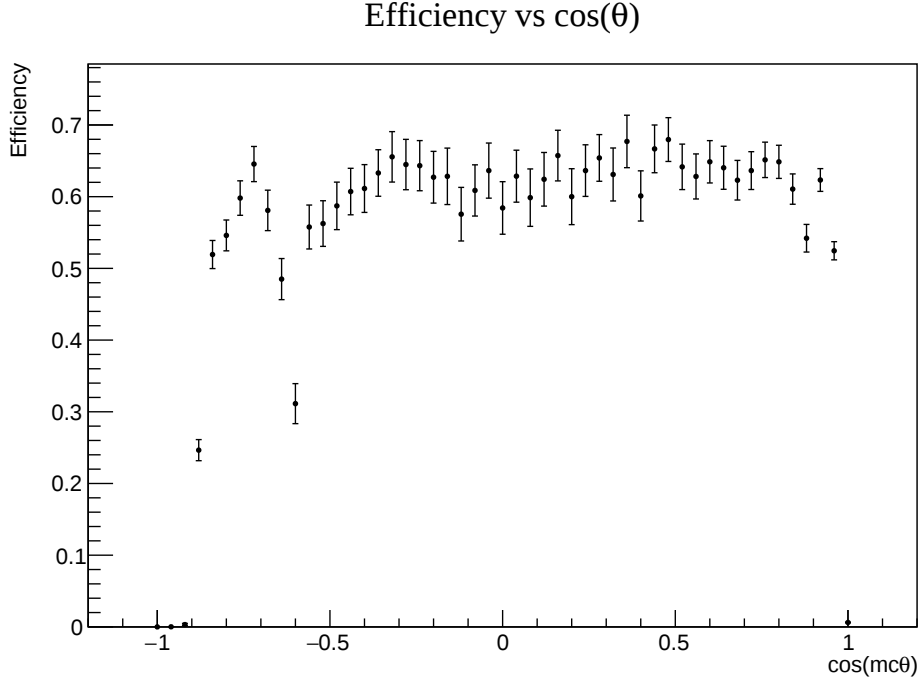


Figure 7.6: Efficiency vs polar angle $\cos \theta$.

This means that the efficiency is very low at extreme angles near $\cos \theta = \pm 1$. This result was expected due to the configuration of the experiment apparatus. In the remaining angular range, the efficiency is almost at 70% indicating the feasible range to study the ISR photons. In addition, we discern a drop at -0.6 . This is caused by the gap of 2° in the calorimeter between the forward direction and the barrel in the angular range $[128.7^\circ, 130.7^\circ]$ [25].

7.4 Conclusion

This chapter ends showing the plots of the efficiency of the reconstruction of the *ISR* photon as a function of the energy and of the polar angle. From this study we know that the integrated efficiency of reconstruction for the *ISR* photon is around the 10%. From the energy plot, Fig. 7.5, we learn the different performance of reconstruction as a function of the energy of the *ISR* photon. We discern the wide range of emission of the *ISR* and its feasible reconstruction. As regard the angular plot we know in which angular region is the highest efficiency of reconstruction. From Fig. 7.6 we learn that, despite the integrated efficiency is around 10%, in a smaller angular range as $-0.9 < \cos \theta < 0.9$ the efficiency increases a lot reaching values over the 50%. These results related to the angular distribution are summarized in Tab. 7.1.

full-range efficiency $0^\circ < \theta < 180^\circ$	barrel region efficiency $12.4^\circ < \theta < 155.1^\circ$
10 %	55%

Table 7.1: *ISR* photon angular reconstruction results.

Chapter 8

Λ reconstruction

In the previous chapter, I concluded the discussion on the reconstruction of the γ_{ISR} photon. From the $e^+e^- \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$ process the next step is to investigate the reconstruction of the Λ and $\bar{\Lambda}$.

The branching ratio of the $\Lambda \rightarrow p\pi^-$ decay is large (64 %). The charged decay products are straight-forward to be measured in a high-acceptance detector, and enable the complete reconstruction of the Λ and $\bar{\Lambda}$. Hence, reconstructing the γ_{ISR} photon may not be necessary, and we may recover some events with very forward or backward-going γ_{ISR} . I will therefore investigate whether it would be more suitable from an efficiency perspective to reconstruct the $\Lambda\bar{\Lambda}$ pair directly while inferring the γ_{ISR} from missing kinematics. The goal of this study is therefore to develop an algorithm that tags a Λ or a $\bar{\Lambda}$ from the $e^+e^- \rightarrow \gamma_{ISR}\Lambda\bar{\Lambda}$ process.

8.1 Λ and $\bar{\Lambda}$ reconstruction

Λ hyperons have many well-known properties such as mass, decay modes and spin. The reconstruction algorithm available for the experiment is able to reconstruct Λ hyperons selecting those events where a pion and a proton come from a common point, or vertex, in 3D space, and where they have an invariant mass close to that of a Λ hyperon. The former is achieved by a successful vertex fit and the latter by requiring the invariant mass to be within the range $1.10 < M < 1.13$ GeV. Using kinematics variables is possible to study some other features for these particles. One way to improve the reconstruction algorithm is to apply a selection criteria, or cuts, on these variables

in order to reduce the background otherwise wrongly classified as hyperons. Assessing the performance is done in terms of efficiency and precision, or purity. A high efficiency means that good events, or events of interest, are kept in the sample but it may come at the price of a larger background, hence decreasing the purity of the sample. From Scavino's PhD thesis [5] it is demonstrated that the reconstruction algorithm for the Λ hyperons can be significantly improved by a proper choice of cuts.

8.1.1 Kinematics plots

In order to identify the most relevant kinematic variables related to the Λ hyperon from a generic MC sample, I will analyse the kinematic distributions of the involved particles. Inspecting those distributions is instrumental to understand whether applying cuts on those variables can separate signal from background. The distributions shown in Figures 8.1-8.7 display the signal distribution over the generic sample. The signal distribution shows the Λ candidates correctly reconstructed while the generic sample contains all the Λ candidates. The sample I considered for the distributions is a generic hadronic MC sample producing an $s\bar{s}$ pair in the continuum with subsequent hadronization. Since the main decay channel for the Λ ($\bar{\Lambda}$) hyperon is a $p\pi^-$ ($\bar{p}\pi^+$) pair, the first variable I considered is $p\pi^-/\bar{p}\pi^+$ momentum. In Fig. 8.1 the two momentum distributions are reported. From these two distributions

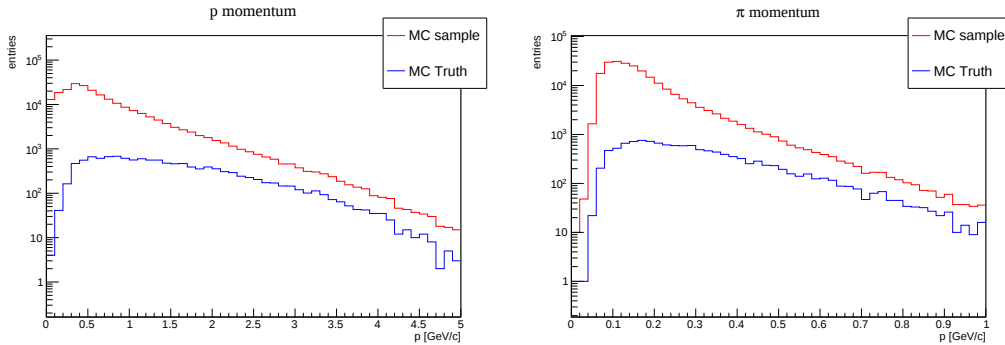


Figure 8.1: Momenta of pions and protons (antiprotons) from $\Lambda \rightarrow p\pi^-$ ($\bar{\Lambda} \rightarrow \bar{p}\pi^+$). Signal events are correctly reconstructed $\Lambda/\bar{\Lambda}$ and shown in blue. The red curve represents both the case when the $p\pi^-$ ($\bar{p}\pi^+$) pair comes from a Λ ($\bar{\Lambda}$) and when they do not.

we can compare the true signal to all the reconstructed events. Signals are all events containing a $p\pi^-$ ($\bar{p}\pi^+$) from a Λ ($\bar{\Lambda}$) decay, and background all events where no Λ or $\bar{\Lambda}$ are involved. Figure 8.2 shows the momentum distribution for the Λ candidate, where Λ from now on refers to both Λ and/or $\bar{\Lambda}$ without distinguishing between them. In Fig. 8.2, the momentum distribution of

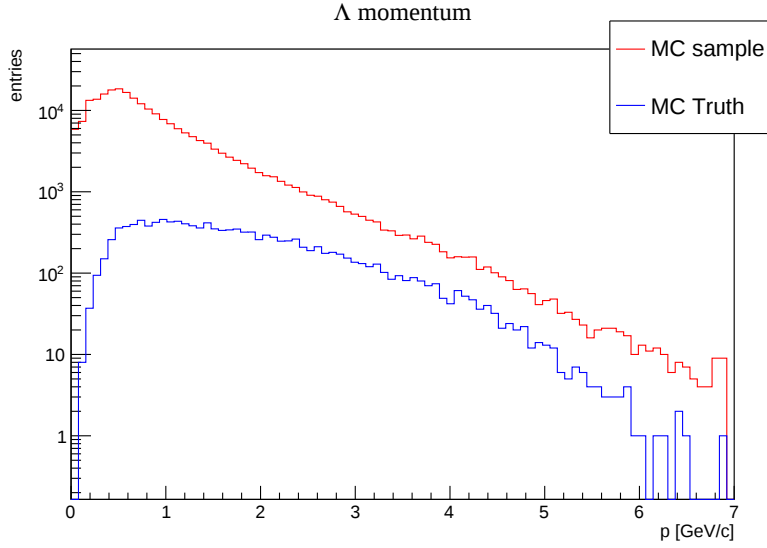


Figure 8.2: Λ momentum distribution.

a Λ candidate is shown. Since the signal and background distribution in Figs. 8.1 and 8.2 overlap almost completely, it makes no sense to apply a cut on these variables. Instead of applying cuts on momenta, variables such as angles and angular correlations are often useful to distinguish signal from background. In particular, variables related to the Λ daughters can give valuable information. I studied the angular distribution of the protons and pions, shown in Fig 8.3 and the open angle in the lab system frame between protons and pions, shown in Fig. 8.4.

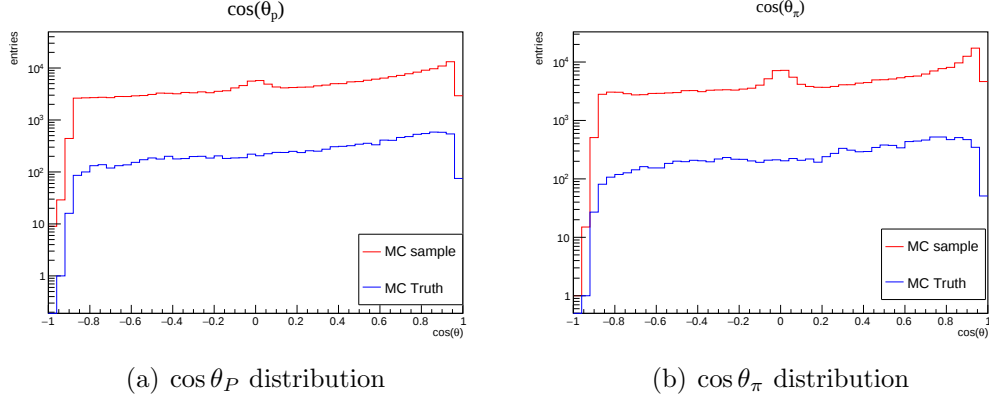


Figure 8.3: Proton and pion angular distribution. In red all the events, in blue only the MC truth.

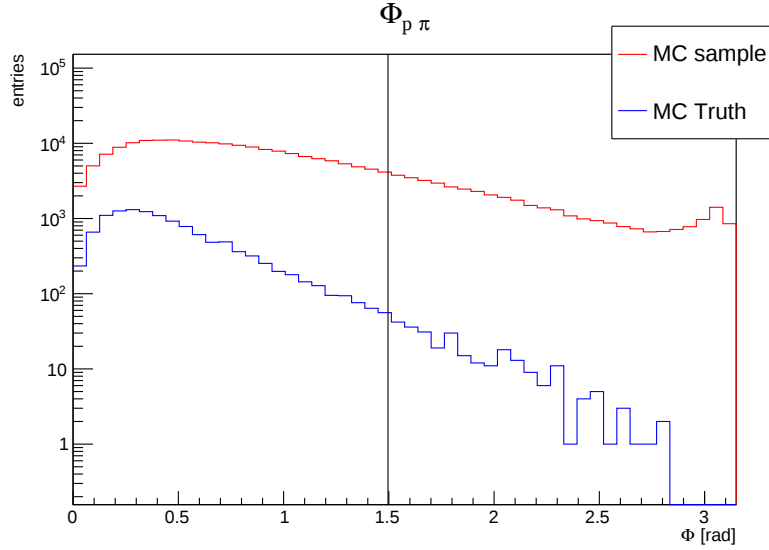


Figure 8.4: Opening Angle distribution between proton and pion.

The opening angle distribution between the proton and the pion is strikingly different for signal compared to the entire sample, as shown in Fig. 8.4. Hence, requiring an opening angle smaller than 1.5 rad could be a good strategy to reduce the background.

Another variable to take into account for the Λ selection is the cosine of the angle between the vertex vector and the Λ momentum. The vertex

vector starts from the production vertex and ends to the Λ vertex decay, where the Λ hyperon decays producing the final state particles. In basf2 analysis software, the Λ vertex decay is reconstructed through the charged tracks of the decays products. Since the track parameters have uncertainties, the reconstructed vertex will be affected. This means that the vertex vector and the Λ momentum vector, as reconstructed from summing the proton and pion momentum vectors, will not be exactly the same but differ by a small angle ξ . In Fig. 8.5 this angle is represented in red, and since it quantifies the difference between the Λ direction as reconstructed from its vertices, and the Λ direction as reconstructed from the momenta of its daughter, it should be very small for true Λ events while take any value for a false Λ . In Fig. 8.5 the distribution for the cosine of the angle between the Λ momentum and its vertex vector is shown. We conclude that this variable is very differently

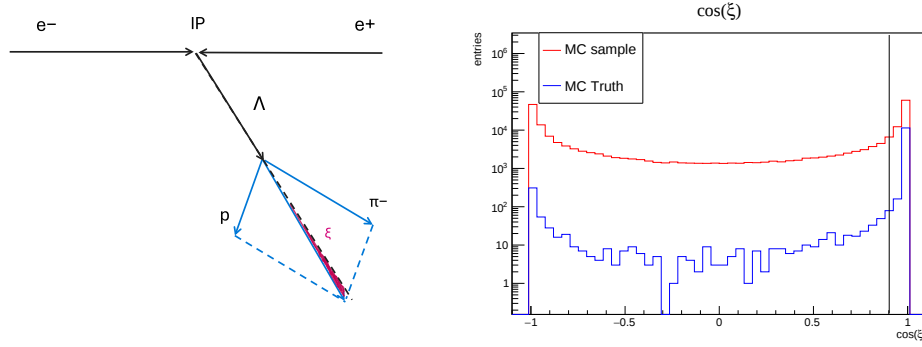


Figure 8.5: cosine of the angle between momentum and vertex vector of the Λ hyperon.

distributed for signal and background, as expected. Therefore, applying a cut on this variable can significantly reduce the background and increase the purity of the sample.

Relevant topological variables related to Λ hyperons are the vertex distance and the flight distance significance. The former is defined as the distance between the IP and the Λ decay vertex position, while the latter is defined as the ratio of the flight distance divided by its uncertainty, defined in eq. 8.1:

$$FD = lx \cdot \frac{p_x}{p} + ly \cdot \frac{p_y}{p} + lz \cdot \frac{p_z}{p} \quad (8.1)$$

with lx , ly , lz , defined as the distance between the production and decay vertex of the Λ particle in x, y, z coordinates. The vertex distance and the flight distance significance for signal and background are shown in Fig. 8.6 panel (a) and (b) respectively. In both cases, it is possible to identify a

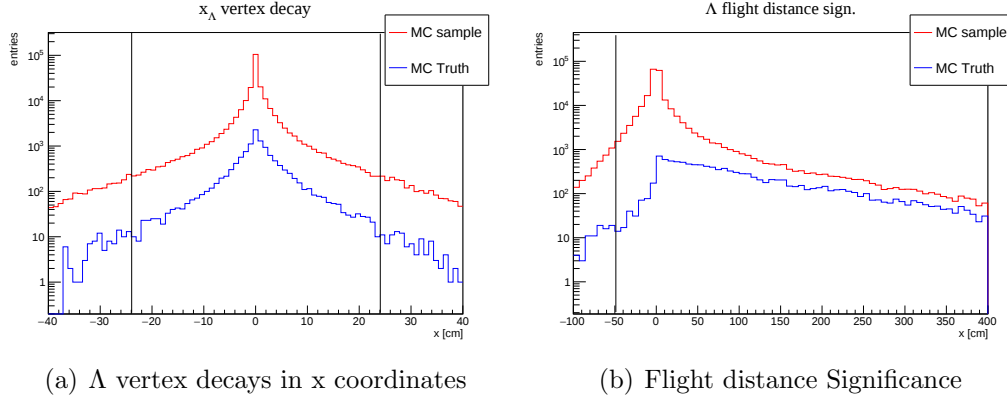


Figure 8.6: Λ vertex related distributions. In red all the events, in blue only the true signal.

suitable cut and thereby increase the purity of the sample.

Finally, the invariant mass of the $p\pi^-/\bar{p}\pi^+$ system, i.e. $M(p\pi^-)$ is shown in Fig. 8.7. From the entire generic hadronic MC sample only a small fraction of the sample is signal.

We now have a better idea of which variables can be used to apply on the selection criteria and thereby improve the reconstruction algorithm. From Scavino's Phd Thesis [5], we learn that there is room for further improvement on the optimization of the event selection criterion considered in the analysis. With the best configuration of her work, the performance on the Λ selection reached an efficiency of 65% and a purity of 86% [5]. However, selecting which variables to take into account for the analysis and where to apply cuts could be hard and a long process. The optimization is done applying a cut on a variable and evaluating again the reconstruction algorithm in order to evaluate the efficiency. Then, the step after is moving to another variable, apply a cut on it and evaluate again the efficiency and the purity. This process must be repeated for all the variables taken into account. Each cut may also be correlated with other cuts in a way that may be difficult to properly account for in the analysis of the systematic uncertainties. In addition, with

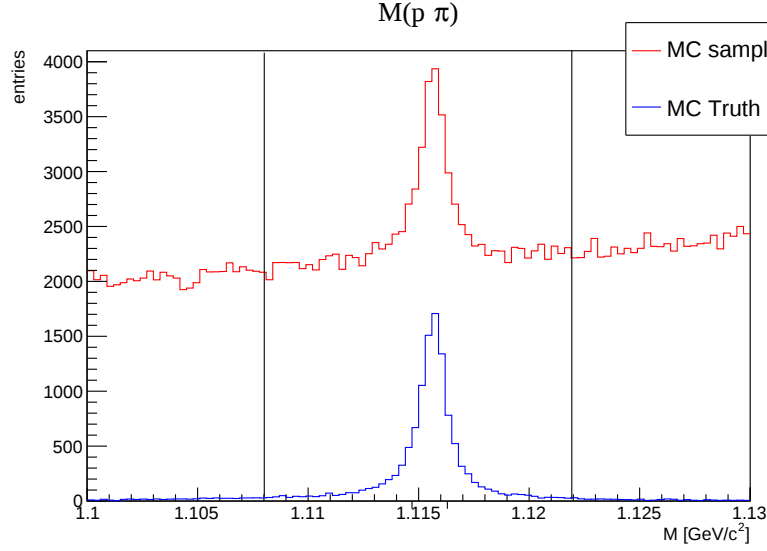


Figure 8.7: invariant mass $M(p\pi^-)$.

this classical approach, it is difficult to study the correlation between more than two variables and what impact a cut has on the other variables. The best configuration is a the trade-off between efficiency and purity. Can we instead use a different approach, where this multiple trial-and-error is not necessary and where correlations are accounted for?

8.2 MVA analysis

The MC samples are utilized to develop a selection algorithm for the selection of Λ hyperons and $\bar{\Lambda}$ antihyperons. The classical approach presented in the previous section is not able to separate events containing Λ hyperons from other events with the same $p\pi$ invariant mass. Instead, we would like to identify a more complicated pattern for the production of the Λ hyperons and classify the event as true. For this reason, Machine Learning tools already implemented in the basf2 software can serve this purpose. The aim of the analysis is to find an algorithm for the selection of Λ hyperons from the MC sample. Those Λ hyperons classified as true will be used for future analysis at the energy of Belle II.

Strategy

Simple models of Machine Learning can be useful in the analysis of the Λ production. Supervised learning, explained in section 4.1, is the tool that allows the model to learn from the sample which Λ hyperons come from the right channel. During the training, the Monte Carlo truth for the events of hyperon production is used as target variable. The Λ candidates that are correctly reconstructed are classified as true. Here, correctly reconstructed means the Λ hyperon reconstruction with the proper charged tracks, i.e. proton and pion, and an invariant mass $M(p\pi)$ in the correct mass region. False are those events not produced according to the decay chain. Those events form the background.

8.2.1 Workflow and variables

The analysis starts giving in input to the supervised model the proper data sample. A generic Monte Carlo sample contains not only hyperons but other particles produced at Belle II energy. In Belle II collaboration, generic simulated samples are provided by the DESY centre. The samples used in the analysis are presented in Appendix A chapter 9. The size of the samples correspond to a luminosity of $L = 100 fb^{-1}$. Therefore, in order to use a Machine Learning approach, I split the sample using the 90% for the training phase and the 10% for the testing phase to evaluate the performance. The model has been trained and evaluated with the $s\bar{s}$ MC sample. After the optimization, the model has been applied on the rest of the MC samples. After this step, it is time to choose an appropriate model.

Fast Boosted Decision Trees (FastBDT) is the one I implemented, due to its straight-forward tuning and good performance in classification problems [44]. The hyperparameter available to be set is the number of trees. During the analysis, I kept the same number of trees, 200.

The **Feature extraction** is one of the most important step during the analysis with ML tools. Choosing which variables are relevant and which not is fundamental for the aim of this analysis. To study the best performance of the model, I evaluated four different configurations with different lists of kinematics variables used during the training phase. These configurations are related to the most important kinematics variables identified in the analysis by Scavino [5] and in this work at subsection 8.1.1. The evaluation and its results will provide guidance on which variables are most powerful in

distinguishing signal from background. In Table 8.1, the variables in the different configurations studied in this work, are listed. In the following, a

Configuration 0	Configuration 1	Configuration 2	Configuration 3
	<i>Proton ID</i>	<i>Proton ID</i>	<i>Proton ID</i>
	<i>Pion ID</i>	<i>Pion ID</i>	<i>Pion ID</i>
Φ	Φ	Φ	Φ
$\cos \xi$	$\cos \xi$	$\cos \xi$	$\cos \xi$
<i>FDS</i>	<i>FDS</i>	<i>FDS</i>	
p_π	p_π	p_π	p_π
p_{π_x}	p_{π_x}	p_{π_x}	p_{π_x}
p_{π_y}	p_{π_y}	p_{π_y}	p_{π_y}
p_{π_z}			
p_p	p_p	p_p	p_p
p_{p_x}			
p_{p_y}			
p_{p_z}			
$\cos \theta_p$			
$\cos \theta_\pi$			
p_Λ			
x_Λ			
y_Λ			
z_Λ			
$M(p\pi)$			

Table 8.1: List of variables used during the training. These configurations have been evaluated in order to establish which has the best results.

short description of these variables, with basf2 labels in parentheses, is given:

- *Proton ID* (Pprot): the identification probability defined as the ratio between the proton Likelihood on the sum of likelihoods of all the detected particles.
- *Pion ID* (Pipio): the pion identification probability defined as the ratio between the pion Likelihood on the sum of likelihoods of all the detected particles.
- Φ (betwe): the angle between Λ daughters: proton and pion.
- $\cos \xi$ (cosVe): the cosine of the angle between momentum and vertex vector of Λ particle.

- FDS (fligh): the flight distance significance, the ratio between the flight distance and its error, seen in the eq. 8.1.
- p_π (pPi, pxPi, pyPi): the pion momentum.
- p_p (pProt): the proton momentum.
- $\cos \theta_p$: the cosine of the polar angle of the proton in the lab system.
- $\cos \theta_\pi$: the cosine of the polar angle of the pion in the lab system.
- p_Λ : the Λ momentum.
- x_Λ (y_Λ, z_Λ): the Λ 3D vertex position in x (y, z).
- $M(p\pi)$ (MLamb): the invariant mass of the system ($p\pi$) and charge conjugate.

These variables will be used as features. Every single data point has as information the list of these features that characterizes the event. Through the Monte Carlo Truth it is possible setting the target variable. The events correctly reconstructed are classified as **signal**, those not well reconstructed are classified as **background**.

8.2.2 Configuration 0

Variable Ranking

The evaluation is done to understand whether the FastBDT model is good to generalize a set of independent data or not. There are several criteria to study its performance. The first one is the **variable ranking** i.e. a scale running from 0 to 100, where large numbers indicate high importance. For the ranking of the variables in Configuration 0, in Tab. 8.2, we conclude that $\cos \xi$, i.e. cosine of the angle between momentum and vertex vector, is the most relevant variable for this configuration in Tab. 8.1 and other variables e.g. vertex decay positions and $\cos \theta_p$ not.

variable ranking	importance
$\cos \xi$	100
Φ	23
p_{π_x}	21
p_{π}	21
p_{π_y}	14
p_P	6
FDS	2
$\cos \theta_p$	0
$\cos \theta_{\pi}$	0
x_{Λ}	0
y_{Λ}	0
z_{Λ}	0
p_{p_x}	0
p_{p_y}	0
p_{p_z}	0
p_{π_z}	0

Table 8.2: variable ranking for the Configuration 0.

ROC-curve

We study now the ROC-curve, discussed in section 4.4, for Configuration 0. Here, the score is between 0 and 1, where 1 implies perfect generalization of the model on an independent sample. In Fig. 8.8 the ROC-curve is shown for an independent sample generated from $s\bar{s}$ hadronization, with a size corresponding to $L = 10fb^{-1}$. The score obtained is 0.94. This is a good starting point, but Table 8.2 indicates that there is room for improvement by excluding variables with low impact.

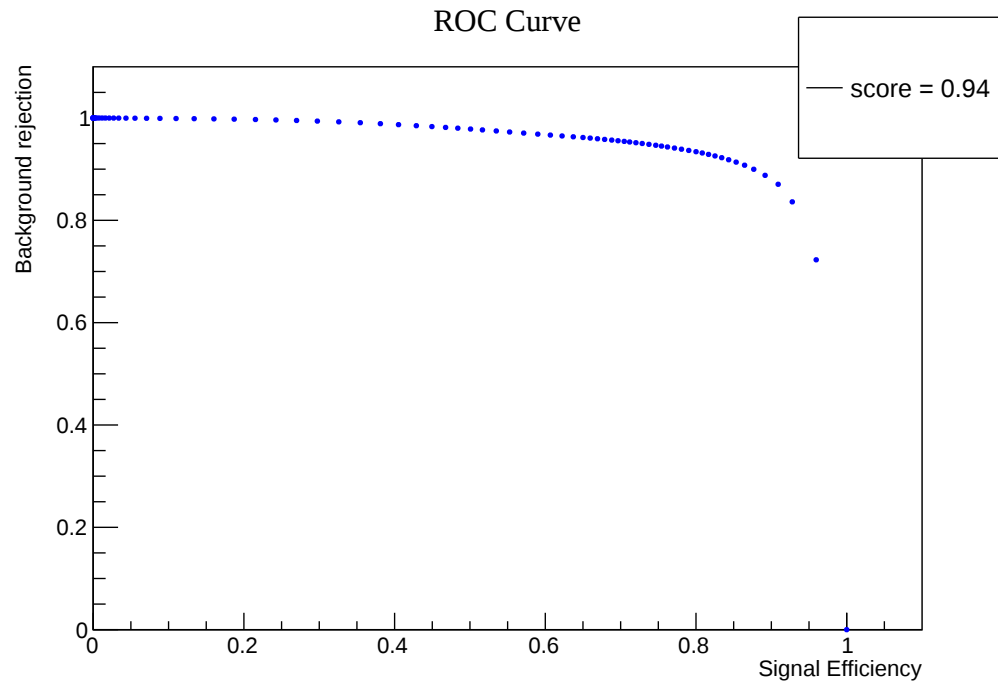


Figure 8.8: ROC-curve for Configuration 0.

Classifier Output distribution

The evaluation gives as output the classifier distribution i.e. the probability for each point to be signal according to the model, ranging from 0 to 1. A classification closer to 1 means a bigger probability to be signal and viceversa with background. To classify the event I must choose a cut-value to separate signal from the background.

In Fig. 8.9 the Classifier Distribution for the Configuration 0 on the evaluation is shown. In red is represented the background, in blue the signal. From this plot it is hard to select a cut-value to separate signal from back-

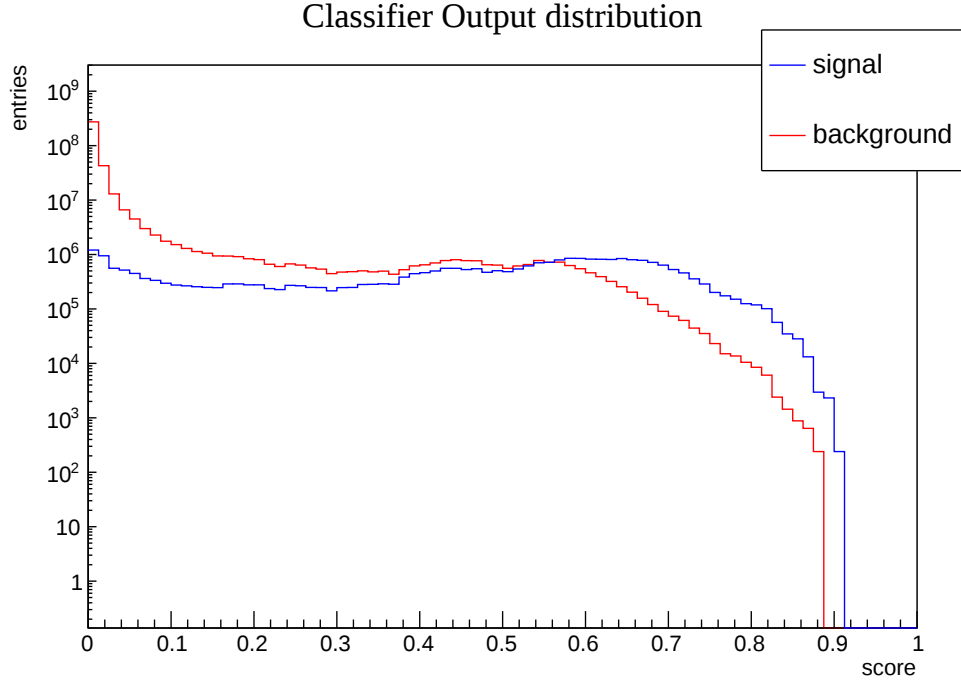


Figure 8.9: Classifier output distribution, the red line represents the background, the blue line the signal.

ground. In addition, there is not a clear peak next to 1. It means the model fails to predict in reliable way which events are signal. Optimally, one should be able to discern two distinct peaks, with the blue closer to 1 and the red closer to 0.

Efficiency and Purity

Other useful plots that make easier studying this configuration are the **Efficiency vs cut-value** and **Purity vs cut-value** distributions. From these it is possible identify a range of cut-values, within which one can discriminate signal from background from the classifier output distribution. In Fig. 8.10 the efficiency is shown in blue and the purity in red, both as a function of

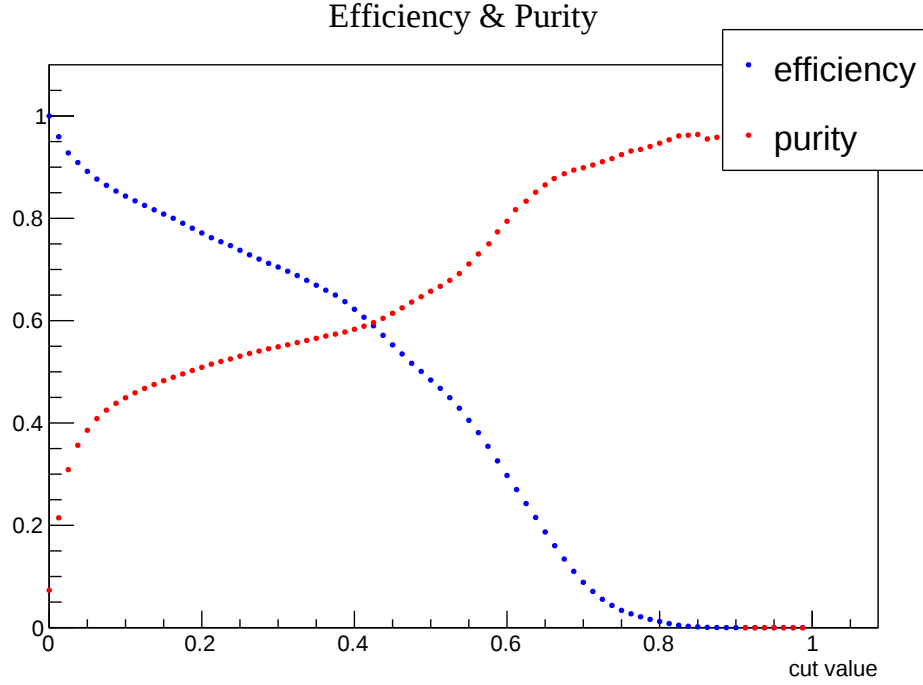


Figure 8.10: Efficiency and purity as function of the cut-value for Configuration 0.

the cut-value. The best cut value is the one that allows a high efficiency and at the same time, keeps the purity high. By increasing the cut values the efficiency decreases while the purity increases.

We have learned few useful lessons by studying Configuration 0, from which we can outline a strategy for modifications to find better results. First, reduce the number of features taking off the ones ranked as not relevant. Then, add other variables such as the particle ID. In this analysis the fine tuning consists in feature reduction and in feature extraction from the list of variables.

8.2.3 Configuration 1

Variable ranking

By removing some variables with low impact from Configuration 0 may increase the performance of the model. Here, we add the Proton ID and Pion

ID to test their impact on the performance. The evaluation is carried out in the same way as before: after the training with these variables, I tested the model on the same independent sample as Configuration 0.

As before, in Tab. 8.3, the variable ranking is presented. The first result

variable ranking	importance
<i>Proton ID</i>	100
$\cos \xi$	21
<i>FDS</i>	14
<i>Pion ID</i>	8
p_π	7
p_{π_x}	6
p_{π_y}	4
p_p	3
Φ	0

Table 8.3: ranking variable for the configuration 1.

is the importance of ProtonID respect to the other variables. According to the model, it is the most important variable from the list following by $\cos \xi$ with a percentage of 21% . The flight distance significance is the third most relevant.

ROC-curve

In Fig. 8.11 the ROC-curve is shown. We see, that after removing the low-impact variables from Configuration 0 the performance of the model improves. As expected, the score increases with respect to Configuration 0 and is found to be 0.99. The model has an high background rejection and a high signal efficiency.

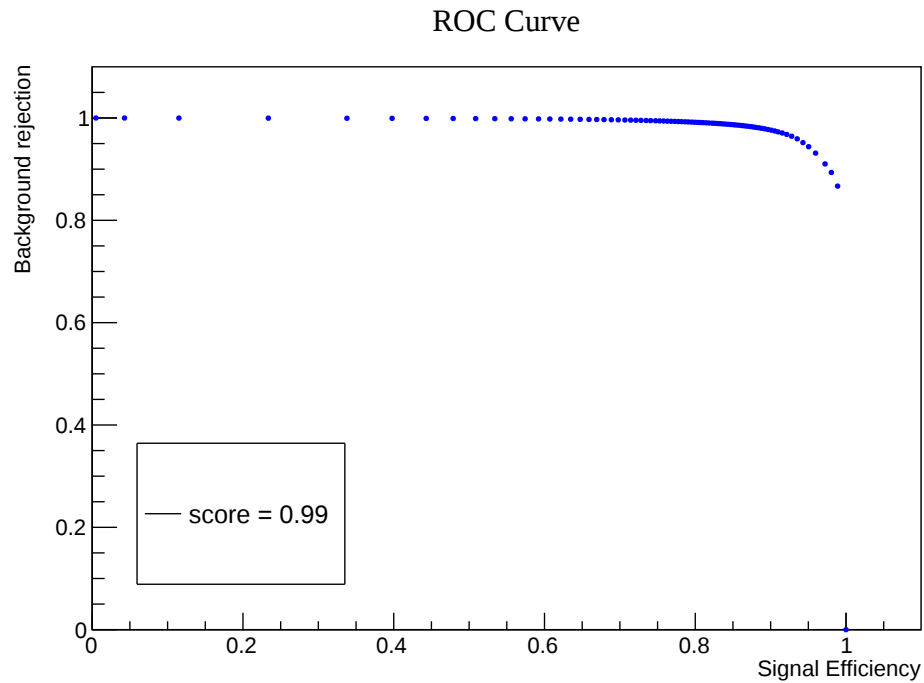


Figure 8.11: ROC-curve for Configuration 1

Classifier Output distribution

In Fig. 8.12 the Classifier distribution for the Configuration 1 is shown. In this case the separation between the background and signal is more evident. It is possible to discern a signal peak to the right on the blue curve and a background peak to the left on the red curve. For Configuration 1 the effect of a cut value in the classifier is therefore significant.

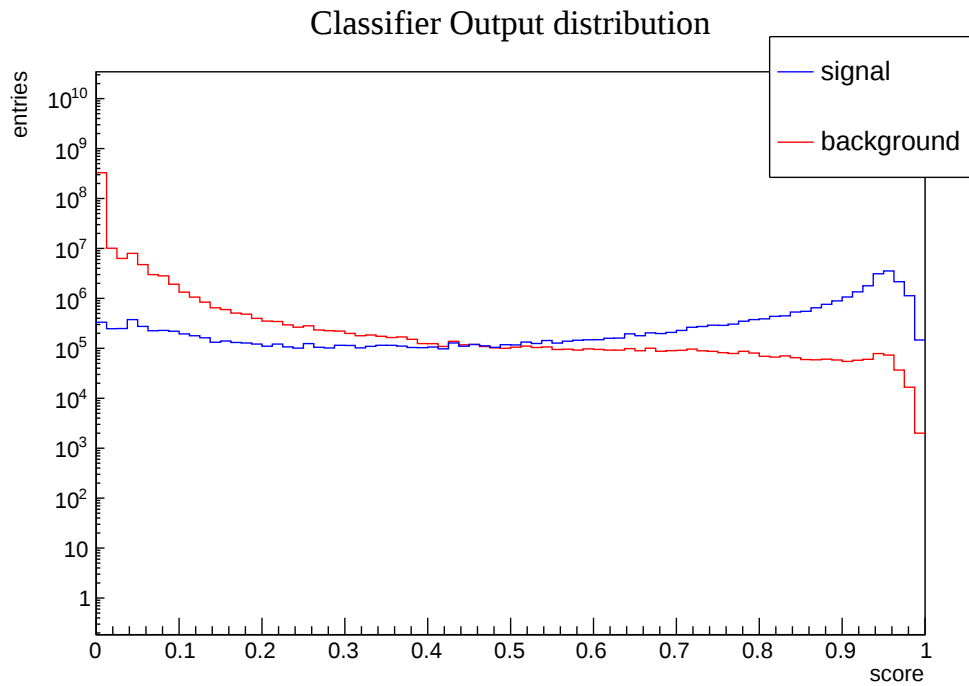


Figure 8.12: Output classifier distribution for the configuration 1.

Efficiency and Purity

The distributions of Efficiency and Purity can help to understand the best range of cut-value. In Fig. 8.13, the efficiency and purity are shown as a function of the classifier distribution.

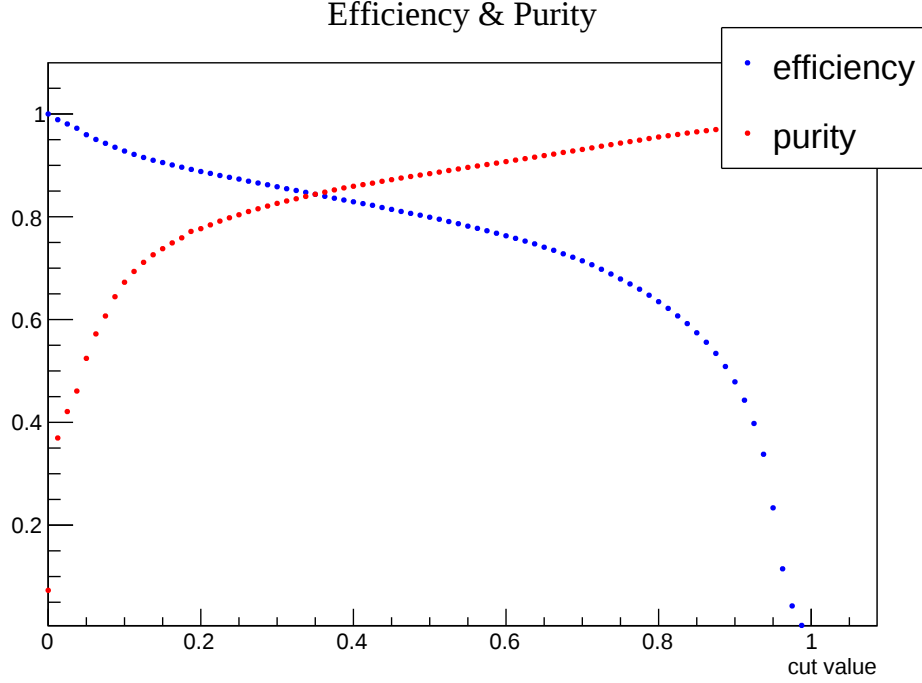


Figure 8.13: Efficiency and purity as function of the cut-value.

With classifier distribution at 0.35, the efficiency is almost 95% and the purity 90%. Otherwise, below 0.2 the purity is poor, and about 0.8 the efficiency drops quickly. Hence, a feasible region to choose a cut-value is between 0.2 and 0.8 .

8.2.4 Configuration 2

In the final configuration it is not convenient using the invariant mass as variable because MC sample could have too many proton-pion combinations that fall inside the invariant mass window.

Variable Ranking

In order to investigate the correlation between the invariant mass of the $(p\pi)$ system and the other variables, we now add the former to the list of variables. The model was trained using the same data samples as for the other configurations. In Tab. 8.4 the variable ranking is shown. Comparing

variable	ranking	importance
<i>Proton ID</i>		100
$\cos \xi$		19
<i>FDS</i>		13
<i>Pion ID</i>		9
$M(p\pi)$		4
p_p		4
p_π		3
p_{π_x}		3
p_{π_y}		3
Φ		0

Table 8.4: ranking variable for the configuration 2.

with Configuration 1, shown in Tab. 8.3, the result is not so far. According to the model the invariant mass as a variable has not a crucial role on the Λ selection. It has a relevant importance of 4%, ranked only as the fifth most important form the list. It means that for the model this additional variable has not a relevant importance.

ROC-curve

With only one variable more from the Configuration 1, the ROC curve expected is not so different from configuration 1. Adding only one variable in the list, i.e. the invariant mass $M(p\pi)$, can not involve in a ROC-curve so far from the one obtained in the Configuration 1. In Fig. 8.14 the ROC-curve is shown. The score obtained for this curve is 0.99, as the same as the Configuration 1.

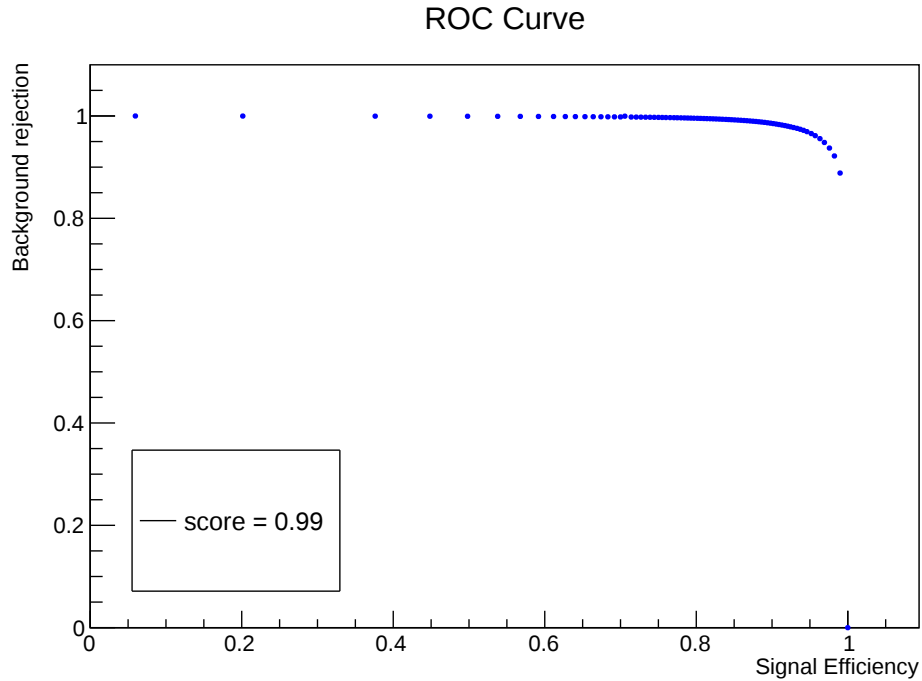


Figure 8.14: ROC-curve for Configuration 2.

Variable Correlation

The basf2 MVA package offers a straight-forward way to study the correlations of the variables used in the model. The invariant mass has another role, understanding which other variables from the list is related to it. In Fig. 8.15 the variable correlation plot is shown. The figure has a scale normalized to 100 for the positive correlation and -100 for the negative correlation.

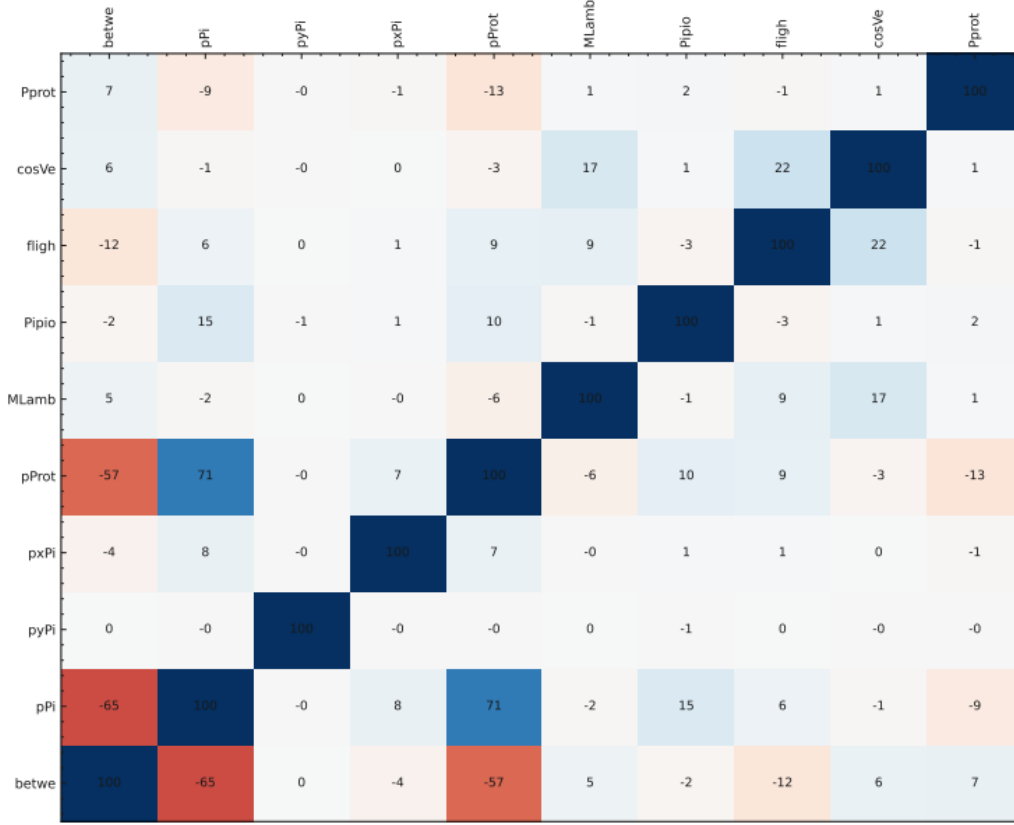


Figure 8.15: variable correlation plot for Configuration 2. In red is represented the negative correlation, in blue the positive.

From Figure 8.15, we conclude that the $M(p\pi)$ invariant mass is weakly correlated to two variables, i.e. flight distance significance and $\cos \xi$ (Table 8.2.1).

Classifier Output distribution

Since the impact of the $M(p\pi)$ variable was found to be small, we do not expect the classifier output to differ significantly compared to Configuration 1. In Fig. 8.16 a clear signal peak is discernible, with a good separation between background and signal. As expected, Configuration 2 yields almost the same output classifier distribution of the Configuration 1.

From Configuration 2, we have learned about the correlation between the proton-pion invariant mass and the other variables from the list. There is

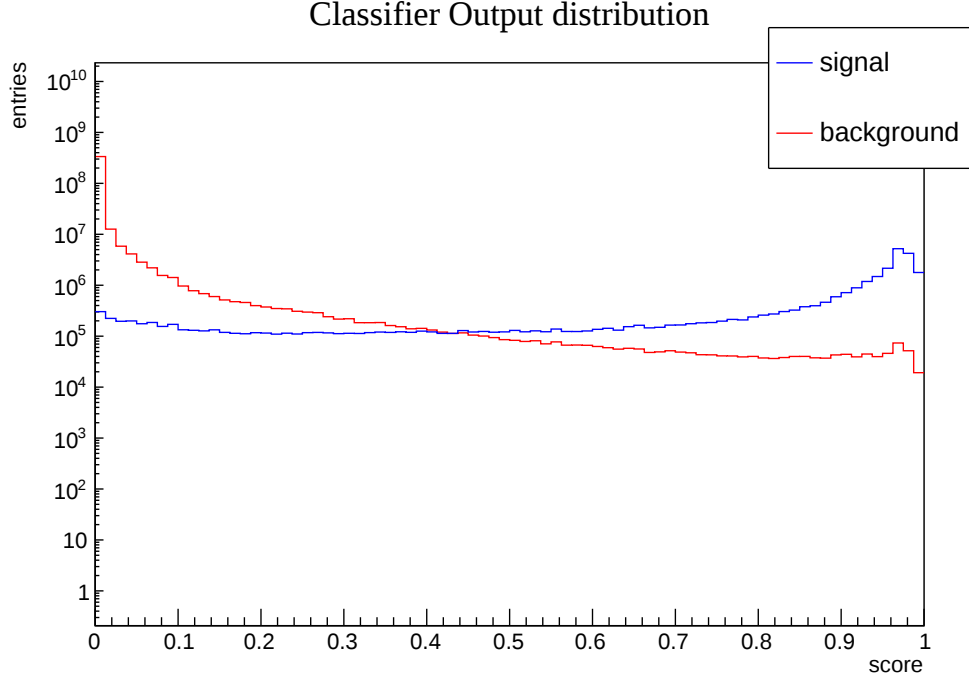


Figure 8.16: Output classifier distribution for the configuration 2 .

only a weak correlation with two variables, which means that there is a very small risk of creating an artificial Λ peak when using Configuration 1. This is reassuring.

8.2.5 Configuration 3

Variable ranking

With Configuration 2, we investigated the impact of the flight distance significance as a variable. By excluding the flight distance significance as variable in Configuration 3, we gain further information of its impact. Furthermore, we remove also the pion-proton invariant mass as a variable. From Table 8.5, we conclude that the variable ranking is similar to that of Configuration 0-2, except from a rescaling of the importance.

variable ranking	importance
<i>Proton ID</i>	100
$\cos \xi$	31
<i>Pion ID</i>	10
p_{π_x}	8
p_{π_y}	8
p_{π}	4
p_p	3
Φ	0

Table 8.5: variable ranking for the configuration 3.

ROC-curve

In Fig. 8.17 the ROC-curve for Configuration 3 is shown. By calculating

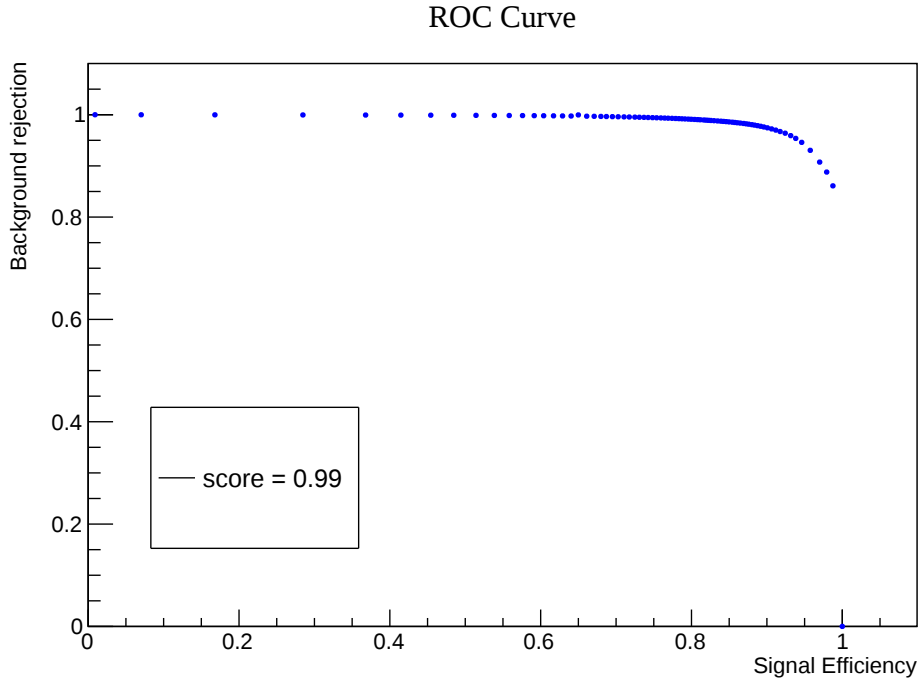


Figure 8.17: ROC-curve for configuration 3.

the integral, we obtain a score is 0.99 as for Configurations 1 and 2, seen in section 8.2.3 and section 8.2.4 respectively, where there is a good signal

efficiency and background rejection.

Classifier Output distribution

In Fig. 8.18 is shown the classifier output distribution. Its distribution is similar to the ones seen in Configurations 1-2. This is motivated by removing few variables from the list in Tab. 8.1: flight distance significance and invariant mass $M(p\pi)$. In Fig. 8.18, it is straight-forward to discern the signal peak and the background peak.

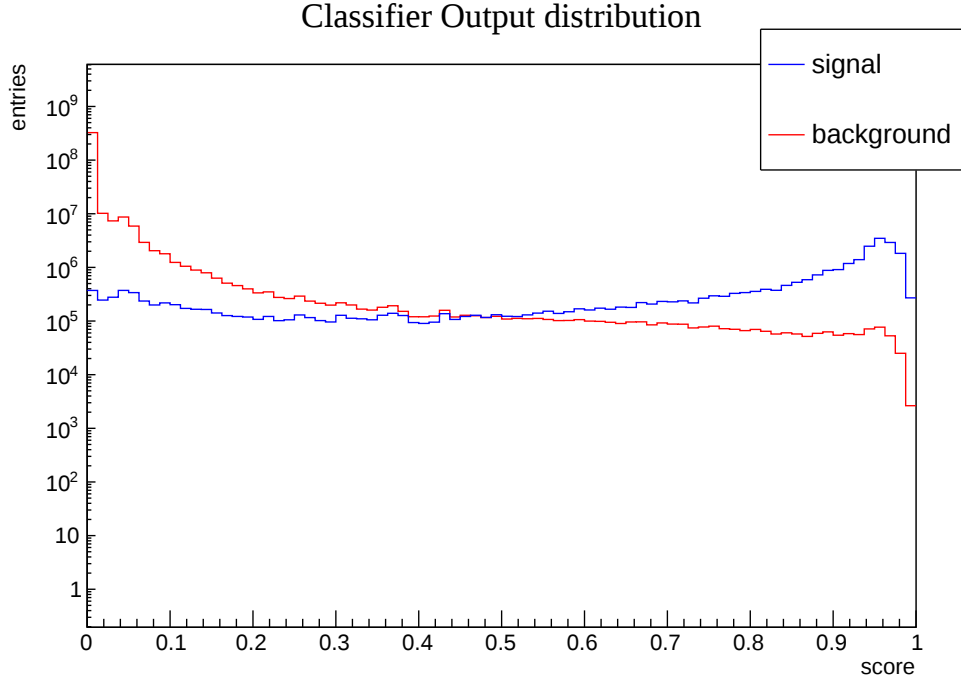


Figure 8.18: classifier output distribution for configuration 3.

8.2.6 Comparing the configurations

The results obtained in this chapter show that Configuration 1, 2, 3 have the best performance on the Λ selection from a generic hadronic MC sample. In Configuration 1, Configuration 2 and Configuration 3 the ROC-score on an independent dataset is 0.99, while the Configuration 0 has a ROC-score

of 0.94. A higher ROC-score means a better generalization of the data in our sample. Furthermore, the Classifier Output distribution in Configuration 0 is not able to discern clearly the signal peak from the background peak. For this reason we can exclude the Configuration 0 from the candidates for the Λ selection. Hence, the recommendation is to use either Configuration 1 or 3, since the Configuration 2 is essentially the same as Configuration 1 but with the invariant mass added. Configuration 2 was used to understand the correlation between the proton-pion invariant mass and other variables. This correlation was found to be small. Configuration 3 is without flight distance significance and mass. As it is possible to see in the Variable ranking of Configuration 1 and Variable Ranking of Configuration 2, the flight distance significance is found to be the third most relevant variable in terms of impact on the final results. Not including that variable could have a worse influence on the generalization of the model on an independent MC sample. Therefore, taking all available information into account, Configuration 1 should be the most suitable one for the purpose of selecting Λ hyperons events.

Once that we have chosen the Configuration 1 for the Λ selection, the analysis continue choosing the right cut-value to apply to the classifier output distribution to separate the signal from the background.

8.2.7 Cut value choice

After concluding that Configuration 1 is the most suitable one, we proceed by optimizing the cut value on the classifier output distribution. Recalling section 8.2.3, the classifier output gives the probability for each event to be signal. Events are binary classified either as signal or background. It is essential choosing the proper cut-value to apply on the distribution and separate the events in signal or background. For this purpose a *Figure of Merit* (FoM) has been evaluated. The best cut-value is the one that make higher the Figure of Merit.

Figure of Merit

Figure of merit is a useful metric to evaluate the performance of reconstruction of the signal. We use the standard definition, which is the formula 8.2:

$$FoM = \frac{N_{signal}}{\sqrt{N_{signal} + N_{background}}} \quad (8.2)$$

where N_{signal} is the number of events that the algorithm classifies as true while $N_{background}$ are the number of events classified as false. A higher FoM means a higher performance in the reconstruction of the signal. In order to keep the efficiency high, the cut value should be chosen as low as possible. As it has been shown in Figures 8.10 and 8.13 the higher cut-value, the lower is the efficiency. The FoM has been calculated on a set of independent MC sample. The best cut-value is the one that maximize the Figure of Merit for the generic hadronic MC sample. Hence, I implemented a grid of one hundred cut values, from zero to one, to use as a threshold. Consequently, since the model is not able to distinguish between the True Positive and False Positive, each event over the threshold is classified as true. Finally, the threshold is applied on the classifier output distribution and, with the events classified as signal, i.e. true, we can study the invariant mass of the $(p\pi)$ system. Since the Λ selector we are building will be applied on Belle II data, we must choose the cut value that maximize the FoM on a generic hadronic MC sample. The generic hadronic sample may be composed by a sum of all the different samples available for the experiment, as explained in 6.1.3. In Appendices are reported the different data samples used for composing the generic MC sample. Hence, the sample I used is composed by the sum of continuum, mixed and charged, in equal luminosity, with a total of $L = 60fb^{-1}$. In Fig. 8.19 is shown the invariant mass $M(p\pi)$ for this generic hadronic MC sample. The histogram in Fig. 8.19, representing $M(p\pi)$, can be fit as a sum of two different functions in order to separate the True Positive (signal) from the the False Positive (background). The signal is represented as the integral of the events below the peak while the background as the integral for those events further to the peak. To fit the histogram I summed the Voigt function with a second order polynomial. The former is defined as a convolution of the gaussian function 8.3 and the lorentzian function 8.4:

$$gauss(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (8.3)$$

$$lorentz(x) = \frac{1}{\pi} \frac{\frac{\Gamma}{2}}{x^2 + \frac{\Gamma^2}{4}} \quad (8.4)$$

Step by step along the grid, the cut value is applied on the classifier distribution. Those events with prediction over the threshold compose the invariant mass histogram. Hence, using the fit function I apply the fit on the $M(p\pi)$ distribution. As a consequence, N_{signal} is calculated as the difference between

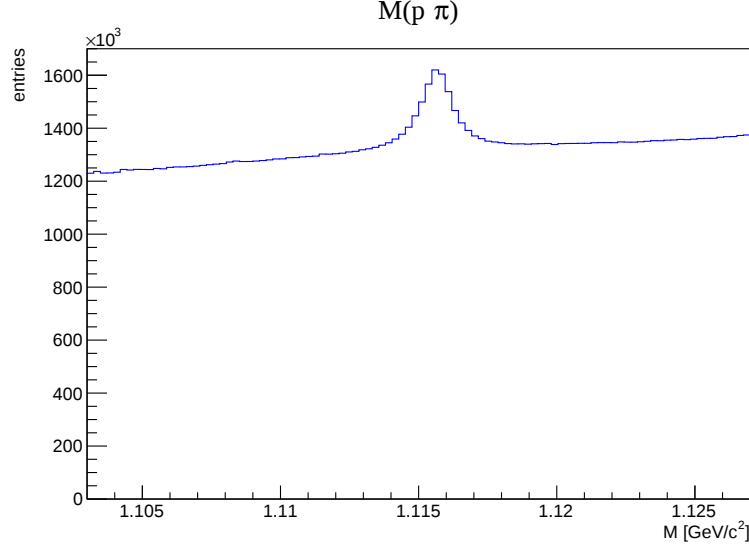


Figure 8.19: invariant mass $M(p\pi)$ for the generic hadronic MC sample.

the integral of the fit function in the range of interest minus the integral of the second order polynomial, defined thanks to the parameters of the fit function. The integral of the second order polynomial is used to calculate $N_{background}$. The procedure is explained in the following equations 8.5 and 8.6:

$$N_{signal} = \int_{1.107\text{GeV}/c^2}^{1.123\text{GeV}/c^2} (Voigt(x) + pol2(x))dx - \int_{1.107\text{GeV}/c^2}^{1.123\text{GeV}/c^2} pol2(x)dx \quad (8.5)$$

$$N_{background} = \int_{1.107\text{GeV}/c^2}^{1.123\text{GeV}/c^2} pol2(x)dx \quad (8.6)$$

In Fig. 8.20 it is represented the fit on the $M(p\pi)$ histogram at two different cut-values. Calculating the signal and the background, the FoM has been valuated for all the one hundred different values of threshold. As a consequence, we obtain the FoM plot as a function of the cut value. In Fig. 8.21 is shown the plot of the FoM for the generic hadronic MC sample. From Fig. 8.21, we choose the cut-value that maximize the FoM. The max is at 0.55 . This is the cut value to apply on the classifier distribution of the independent generic hadronic sample.

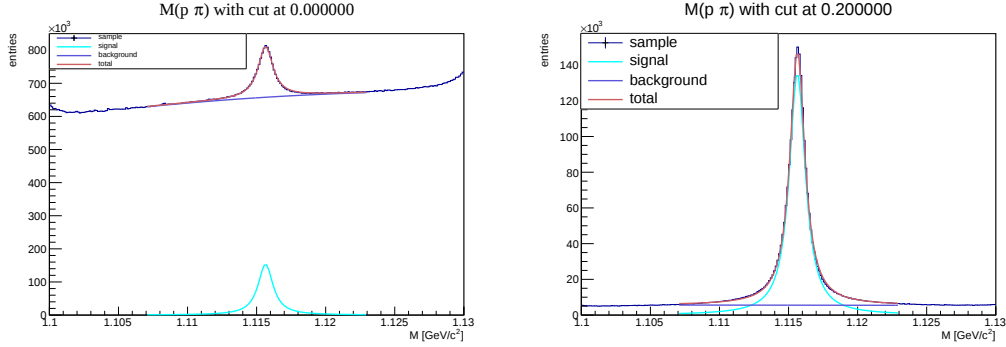


Figure 8.20: On the left it is represented the fit of the $M(p\pi)$ of the generic hadronic MC sample with cut value at 0.0, on the right the fit on a distribution with cut value at 0.2.

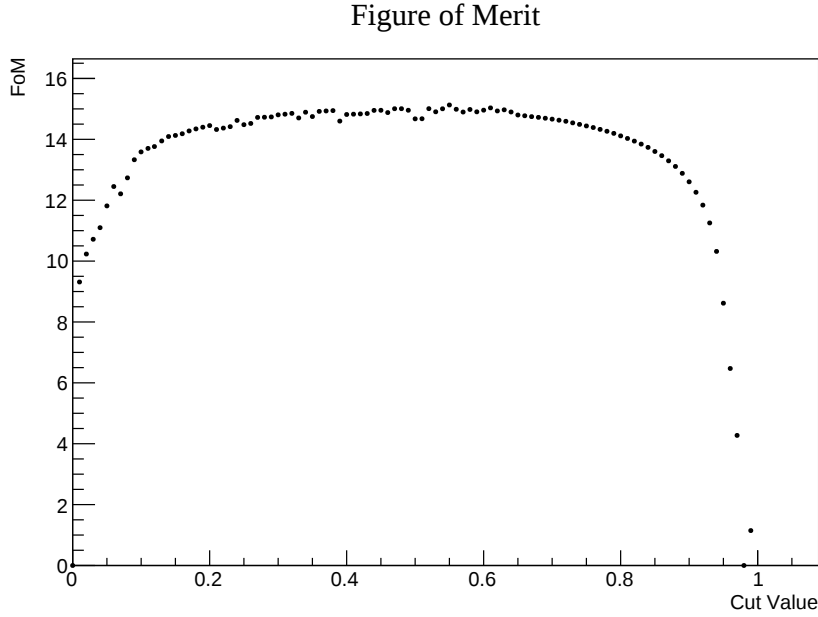


Figure 8.21: Figure of Merit calculated from a generic hadronic sample.

8.2.8 Applying model

The final step for this analysis is to apply the model on a independent generic hadronic sample and, on the classifier output distribution, apply the cut value

to separate the signal from the background. The model has been applied in all the MC sample seen in section 6.1.3 with a luminosity of $L = 10fb^{-1}$. In output from the model we have the classifier output distribution for each sample and the invariant mass distribution. To create a generic hadronic sample, we add the distributions of all the different samples together, creating a generic sample with luminosity $L = 60fb^{-1}$. Finally, on the generic classifier output distribution we apply the cut value chosen in the last section 8.2.7, at 0.55. Hence, all the events with prediction over 0.55 are classified as signal from the model. For those event we can show the invariant mass $M(p\pi)$, as in Fig. 8.22. Thanks to the MC truth of the invariant mass we

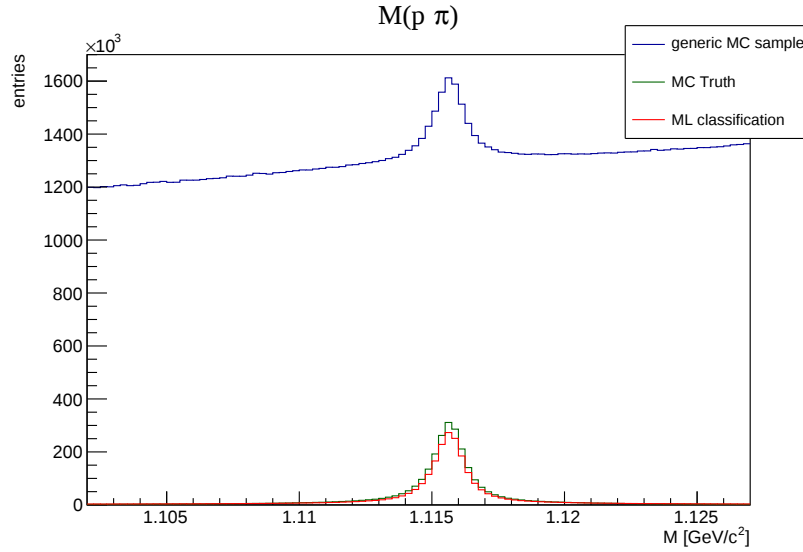


Figure 8.22: invariant mass $M(p\pi)$ for the generic independent MC sample. In the plot in red is represented the classification from the model, in green the MC Truth and in blue the generic MC sample.

can compare the prediction of the model with the classification, showed in Fig. 8.23. In addition, since we are applying the model on a MC sample, we can estimate the performance of the model through the efficiency and the purity.

The performance are summarized in Tab. 8.6 and represent the quality of the Λ hyperon selector using a machine learning tool, FastBDT model. From Fig. 8.23 we realize that the model can fail on the classification on the events. Comparing the MC truth with the classifier it is possible discerning

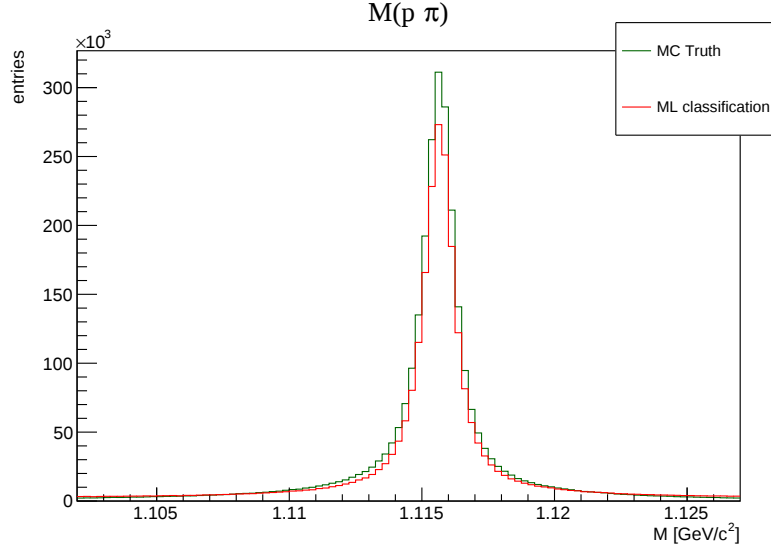


Figure 8.23: zoom on the invariant mass $M(p\pi)$ for the generic independent MC sample. In the plot are represented the classification from the model in red and the MC truth in green.

performance	
<i>cut value</i>	0.55
<i>efficiency</i>	77%
<i>purity</i>	88%

Table 8.6: performance of the Λ selection using a Machine Learning model.

the not perfect matching between the two distributions. The classification through the ML model is strict on the Λ candidate selection, ensuring a good purity of the sample. As a consequence, it causes a limit on the efficiency.

Chapter 9

Conclusion and outlook

The analysis done in this thesis constitutes an initial step for studying the Λ hyperons form factor in the Belle II experiment. The form factors are accessible through the channel $e^+e^- \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$. The analysis has two crucial components: the reconstruction of the γ_{ISR} and the selection of the Λ hyperons and $\bar{\Lambda}$ antihyperons. In the study of the γ_{ISR} , I have focused on the reconstruction efficiency as a function of the γ_{ISR} energy and the γ_{ISR} emission angle. Thanks to the precise knowledge of the initial state on the lepton colliders, it is sufficient to reconstruct γ_{ISR} and either Λ or $\bar{\Lambda}$, while the untagged hyperon or antihyperon can be identified through the missing mass technique. In a time-like form factor analysis it could be useful combining different options of the considered data sample, selecting samples with Λ and $\bar{\Lambda}$ and no γ_{ISR} and samples with γ_{ISR} and one Λ ($\bar{\Lambda}$).

The main focus of my analysis has been to develop a hyperon selector that improves the current standard reconstruction algorithm. After examining the kinematics variables related to the Λ hyperons, and realizing that the classical approach of applying cuts on these was not appropriate for my purpose, I instead adopted a selection algorithm based on machine learning technique called Fast Boosted Decision Trees. In my analysis, I identified the most relevant kinematics variables and correlations related to the Λ and its decay products. I investigated four different configurations of variables, performed training and finally benchmarked the model on a generalized and independent data sample. My analysis revealed that a configuration based on variables such as the ProtonID and $\cos\xi$, has a much higher impact on the classification. It was particularly surprising that the invariant mass has a rather low impact on the final score, reaching an importance score of only 4%.

The most impactful configuration, based on the variables such as ProtonID, $\cos \xi$, flight distance significance, PionID, pion and proton momenta, was chosen for a more rigorous benchmark study. Here, I studied how to optimize the figure of merit for the generic hadronic MC sample. This resulted in an optimal cut value for the classification variable from the Fast Boosted Decision Tree. With this cut, the performance of the Λ hyperon selector I developed has an efficiency of 77% and a purity of 88%. This can be compared to the PhD thesis by Scavino [5], in which she developed a Λ selector based on a classical approach with cuts on different kinematic variables. This selector was optimised using a different data sample at a different energy, and is therefore not directly comparable to my selector. Nevertheless, the selector by Scavino yields an efficiency of 65% and a purity of 86%. In order to have an idea of the merit of the Λ hyperon selector I developed, we can discern that the performance are not so different, testifying the utility of the machine learning approach on the Λ selection. Furthermore, thanks to this approach we have a better idea in which kinematics variables are the most important and which is their correlation.

The long-term goal is to apply my algorithm more generally on experimental data from Belle II. The next step is to separate Λ hyperons from $\bar{\Lambda}$ antihyperons. This will be done investigating on the electric charge of the decay products, $(p\pi^-)$ or $(\bar{p}\pi^+)$. An important benchmark will be to compare the performance of the selector for both particles Λ or $\bar{\Lambda}$; an unbiased algorithm should give similar results for particles and antiparticles. Then, reconstruction algorithm can be combined with the tagging of a γ_{ISR} to select events from $e^+e^- \rightarrow \Lambda\bar{\Lambda}\gamma_{ISR}$ containing exactly one Λ and one $\bar{\Lambda}$. Finally, time-like form factor will be studied through the complete reconstruction of the channel of interest. In Fig. 9.1 I summarize the workflow done and to do for studying the Λ hyperons form factor in Belle II experiment.

Λ hyperons Form Factor in Belle II

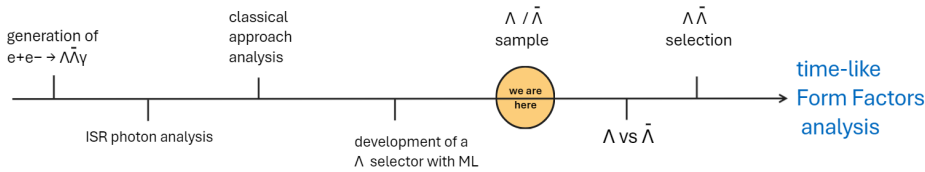


Figure 9.1: Schematic flow of the Λ hyperons form factor studies in Belle II experiment.

Appendices

Appendix A: data sample

Only accessible for DESY members, the following paths provide the data sample I used in the thesis.

In $\Lambda\bar{\Lambda}$ Generation I presented generic hadronic Monte Carlo sample, in the following lines I report their paths.

- $u\bar{u}$ data sample ($10fb^{-1}$):

/pnfs/desy.de/belle/local/belle/MC/release-06-00-08/DB00002100/MC15ri_b/
prod00024801/s00/e1003/4S/r00000/uubar/mdst/sub00/

- $d\bar{d}$ data sample ($10fb^{-1}$):

/pnfs/desy.de/belle/local/belle/MC/release-06-00-08/DB00002100/MC15ri_b/
prod00024791/s00/e1003/4S/r00000/ddbar/mdst/sub00/

- $s\bar{s}$ data sample ($10fb^{-1}$):

/pnfs/desy.de/belle/local/belle/MC/release-06-00-08/DB00002100/MC15ri_b/
prod00024797/s00/e1003/4S/r00000/ssbar/mdst/sub00/

- $c\bar{c}$ data sample ($10fb^{-1}$):

/pnfs/desy.de/belle/local/belle/MC/release-06-00-08/DB00002100/MC15ri_b/
prod00024786/s00/e1003/4S/r00000/ccbar/mdst/sub00/

- *mixed* data sample ($10fb^{-1}$):

/pnfs/desy.de/belle/local/belle/MC/release-06-00-08/DB00002100/MC15ri_b/
prod00024821/s00/e1003/4S/r00000/mixed/mdst/sub00/

- *charged* data sample ($10fb^{-1}$):

/pnfs/desy.de/belle/local/belle/MC/release-06-00-08/DB00002100/MC15ri_b/
prod00024816/s00/e1003/4S/r00000/charged/mdst/sub00/

During the training and the test in MVA analysis I used $s\bar{s}$ data sample:

- $s\bar{s}$ data sample train ($90fb^{-1}$):

/pnfs/desy.de/belle/local/belle/MC/release-06-00-08/DB00002100/MC15ri_b/
prod00024796/s00/e1003/4S/r00000/ssbar/mdst/sub00/

- $s\bar{s}$ data sample test ($10fb^{-1}$):

/pnfs/desy.de/belle/local/belle/MC/release-06-00-08/DB00002100/MC15ri_b/
prod00024796/s00/e1003/4S/r00000/ssbar/mdst/sub00/

Bibliography

- [1] B. Povh et al. *Particles and nuclei: An Introduction to the physical concepts*. Graduate Texts in Physics. Springer, 1995. ISBN: 978-3-662-46320-8, 978-3-662-46321-5. DOI: 10.1007/3-540-36684-9.
- [2] Standard model pics. URL: https://it.wikipedia.org/wiki/Modello_standard.
- [3] D. H. Perkins. *Introduction to high energy physics*. 1982. ISBN: 978-0-521-62196-0.
- [4] Brian Robert Martin and Graham Shaw. *Particle physics*. 2008. ISBN: 978-0-470-03294-7.
- [5] Bianca Scavino. “Development of Λ baryons reconstruction and its application to the search for a stable hexaquark at Belle II”. PhD thesis. University of Mainz, 2022.
- [6] Karin Schönning et al. “Production and decay of polarized hyperon-antihyperon pairs”. In: *Chinese Physics C* 47.5 (2023), p. 052002. DOI: 10.1088/1674-1137/acc790. URL: <http://hepnp.ihep.ac.cn/en/article/doi/10.1088/1674-1137/acc790>.
- [7] Bernard Aubert et al. “Study of $e^+e^- \rightarrow \Lambda\bar{\Lambda}$, $\Lambda\bar{\Sigma}^0$, $\Sigma^0\bar{\Sigma}^0$ using initial state radiation with BABAR”. In: *Phys. Rev. D* 76 (2007), p. 092006. DOI: 10.1103/PhysRevD.76.092006. arXiv: 0709.1988 [hep-ex].
- [8] M. Gell-Mann. “STRANGENESS”. In: *Journal de Physique Colloques* 43.C8 (1982), pp. C8-395-C8-408. DOI: 10.1051/jphyscol:1982825. URL: <https://hal.science/jpa-00222385>.
- [9] ALICE Collaboration. “Measurement of the Λ hyperon lifetime”. In: (Mar. 2023).
- [10] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

- [11] C.F. Perdrisat, V. Punjabi, and M. Vanderhaeghen. “Nucleon electromagnetic form factors”. In: *Progress in Particle and Nuclear Physics* 59.2 (Oct. 2007), pp. 694–764. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2007.05.001. URL: <http://dx.doi.org/10.1016/j.ppnp.2007.05.001>.
- [12] Rinaldo Baldini Ferroli, Simone Pacetti, and Egle Tomasi-Gustafsson. “Threshold phenomenology of nucleon form factors”. In: *J. Univ. Sci. Tech. China* 46.4 (2016), pp. 308–315. DOI: 10.3969/j.issn.0253-2778.2016.04.008.
- [13] Viktor Thoren. “Hadron Physics in a Polarized World : Exploring Electromagnetic Interactions with Spin Observables”. PhD thesis. Uppsala University, Nuclear Physics, 2022, p. 174. ISBN: 978-91-513-1415-0.
- [14] Elisabetta Perotti. “Electromagnetic and Spin Properties of Hyperons”. PhD thesis. Uppsala University, Nuclear Physics, 2020, p. 69. ISBN: 978-91-513-0956-9.
- [15] A. Z. Dubnickova, S. Dubnicka, and M. P. Rekalo. “Investigation of the nucleon electromagnetic structure by polarization effects in $e^+e^- \rightarrow N \text{ anti-}N$ processes”. In: *Nuovo Cim. A* 109 (1996), pp. 241–256. DOI: 10.1007/BF02731012.
- [16] V. A. Matveev, R. M. Muradyan, and A. N. Tavkhelidze. “Automodality in strong interactions”. In: *Teor. Mat. Fiz.* 15 (1973), pp. 332–339.
- [17] Alessio Mangoni, Simone Pacetti, and Egle Tomasi-Gustafsson. “First exploration of the physical Riemann surfaces of the ratio G_E^Λ/G_M^Λ ”. In: *Phys. Rev. D* 104 (11 Dec. 2021), p. 116016. DOI: 10.1103/PhysRevD.104.116016. URL: <https://link.aps.org/doi/10.1103/PhysRevD.104.116016>.
- [18] “An exclusive event generator for e^+e^- scan experiments”. In: *Chinese Physics C* 38.8 (Aug. 2014), p. 083001. DOI: 10.1088/1674-1137/38/8/083001. URL: <https://dx.doi.org/10.1088/1674-1137/38/8/083001>.
- [19] Bo Cao. “Study of $e^+e^- \rightarrow \pi^+\pi^-\pi^0$ process using initial state radiation at KLOE”. In: Sept. 2020, p. 522. DOI: 10.22323/1.364.0522.
- [20] J. Pettersson. “From Strange to Charm: Meson production in electron-positron collisions”. PhD thesis. Uppsala University, 2019.

- [21] Belle II official website. URL: <https://www.belle2.org/>.
- [22] Scheme of SuperKEKB collider. URL: <https://www.bnl.gov/newsroom/news.php?a=117285>.
- [23] S. Kurokawa and E. Kikutani. “Overview of the KEKB accelerators”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 499.1 (2003). KEK-B: The KEK B-factory, pp. 1–7. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(02\)01771-0](https://doi.org/10.1016/S0168-9002(02)01771-0). URL: <https://www.sciencedirect.com/science/article/pii/S0168900202017710>.
- [24] 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. URL: <http://dx.doi.org/10.1093/ptep/ptz106>.
- [25] T. Abe et al. *Belle II Technical Design Report*. 2010. arXiv: 1011.0352 [physics.ins-det].
- [26] Belle II detector scheme. URL: <https://belle2.jp/detector/>.
- [27] O. Alonso et al. “DEPFET active pixel detectors for a future linear e^+e^- collider”. In: *IEEE Trans. Nucl. Sci.* 60 (2013), p. 1457. DOI: 10.1109/TNS.2013.2245680. arXiv: 1212.2160 [physics.ins-det].
- [28] M. Akatsu et al. “Time-of-propagation Cherenkov counter for particle identification”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 440.1 (Jan. 2000), pp. 124–135. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(99)00819-0. URL: [http://dx.doi.org/10.1016/S0168-9002\(99\)00819-0](http://dx.doi.org/10.1016/S0168-9002(99)00819-0).
- [29] H. Ikeda et al. “A detailed test of the CsI(Tl) calorimeter for BELLE with photon beams of energy between 20-MeV and 5.4-GeV”. In: *Nucl. Instrum. Meth. A* 441 (2000), pp. 401–426. DOI: 10.1016/S0168-9002(99)00992-4.
- [30] basf2 software documentation. URL: <https://software.belle2.org/sphinx/release-05-02-17/index.html#>.
- [31] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. The MIT Press, 2012. ISBN: 026201825X.

- [32] Kevin P. Murphy. *Machine learning : a probabilistic perspective*. Cambridge, Mass. [u.a.]: MIT Press, 2013. ISBN: 9780262018029 0262018020. URL: https://www.amazon.com/Machine-Learning-Probabilistic-Perspective-Computation/dp/0262018020/ref=sr_1_2?ie=UTF8&qid=1336857747&sr=8-2.
- [33] Yaser S. Abu-Mostafa, Malik Magdon-Ismael, and Hsuan-Tien Lin. *Learning From Data*. AMLBook, 2012.
- [34] Aurlien Gron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 1st. O'Reilly Media, Inc., 2017. ISBN: 1491962291.
- [35] Denis Reis, André Maletzke, and Gustavo Batista. “One-class Quantification”. In: Sept. 2018.
- [36] Tom M. Mitchell. *Machine Learning*. <http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html>. McGraw-Hill Science/Engineering/Math; (March 1, 1997), 1997. ISBN: 0070428077.
- [37] Decision trees pics. URL: <https://botpenguin.com/glossary/decision-trees>.
- [38] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction*. 2nd ed. Springer, 2009. URL: <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>.
- [39] David J. Lange. “The EvtGen particle decay simulation package”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(01\)00089-4](https://doi.org/10.1016/S0168-9002(01)00089-4). URL: <https://www.sciencedirect.com/science/article/pii/S0168900201000894>.
- [40] H. CzyŻ, M. Gunia, and J. H. Kühn. “Simulation of electron-positron annihilation into hadrons with the event generator PHOKHARA”. In: *Journal of High Energy Physics* 2013.8 (Aug. 2013). ISSN: 1029-8479. DOI: 10.1007/jhep08(2013)110. URL: [http://dx.doi.org/10.1007/JHEP08\(2013\)110](http://dx.doi.org/10.1007/JHEP08(2013)110).
- [41] Christian Bierlich et al. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. 2022. arXiv: 2203.11601 [hep-ph].

- [42] J. Bernhard. PhD thesis. JGU Mainz, 2014.
- [43] T. Ullrich and Z. Xu. *Treatment of Errors in Efficiency Calculations*. 2012. arXiv: `physics/0701199` [`physics.data-an`].
- [44] Thomas Keck. “FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification”. In: *CoRR* abs/1609.06119 (2016). arXiv: `1609.06119`. URL: <http://arxiv.org/abs/1609.06119>.