UNIVERSITÀ DI PISA

Dipartimento di Fisica E. Fermi
Corso di Laurea Magistrale in Fisica

Search for an Axion-Like Particle in B decays at the BABAR experiment and projections to the Belle II data sample

Candidate:
Michael De Nuccio

Supervisor:
Prof. Francesco Forti

Academic Year 2016/2017

Contents

Introduction

The goal of this thesis is the search for an Axion-Like Particle (ALP) A in the decay $B^{\pm} \to K^{\pm}A$, $A \to \gamma\gamma$ in the data collected by the *BABAR* experiment from 1999 to 2007, and the projections to the Belle II experiment, that will soon start collecting data (April 2018).
The ALPs are neutral, massive, pseudoscalar particles constituting an extension of the axionic model. The axion was originally hypothesized by Peccei and Quinn to solve the problem of the non-violation of CP, i.e. the symmetry under charge conjugation and spatial parity, by the strong interactions. The strong interaction Lagrangian includes CP-violating terms controlled by parameters that can be interpreted as angles. Without constraints they can assume any value in $[0, 2\pi)$, while the fact that no CP violation has ever been observed in strong processes implies that these parameters must be very close to zero, with need of a fine-tuning of their values.
To solve this problem, Peccei and Quinn hypothesized the existence of the axion, a particle that, thanks to a spontaneous symmetry breaking mechanism, would have naturally made these parameters null. Although all the experimental searches carried out to find the axion have been unsuccessful, the axionic model has lead to the creation of a whole family of successors, predicting the existence of Axion-Like Particles, possible candidates to solve other open problems in Physics, as the dark matter.

The ALP searched in this analysis would be produced in flavor changing neutral currents processes, rare in the Standard Model, and presents a characteristic signature, with good prospects for discovery. Since there are no constraints from the theory about the ALP mass, the search is carried out up to the kinematical limit, given by $m_A^{max} = m_{B^{\pm}} - m_{K^{\pm}} = 4786$ MeV/c^2.
A B-factory is an accelerator optimized to generate a high number of B mesons: the PEP II accelerator, SLAC, California, hosted the *BABAR* experiment, while SuperKEKEB, KEK, Japan, hosts the Belle II experiment. The *BABAR* experiment has collected 424.2 fb^{-1} of integrated luminosity at the energy of the $\Upsilon(4S)$ meson, and this is the data sample available for my analysis.

The objective of this analysis is to reconstruct the decay of a charged B meson into a charged K and two photons, and to search resonances in the diphoton invariant mass spectrum. Various multivariate analysis have been tested to discriminate the signal from the background, and different selection strategies have been taken into account to reject the two main backgrounds, i.e. the one related to B decays and the one related to light quarks production (uds) and quark c.

In order to maximize the selection efficiency the rest of the event is not reconstructed.
The uds background is jet-like, while the signal has the spherical topology of a B decay, hence the shape variables, that describe the topology of the event, are particularly useful to reject this source of background. The rejection of the B-like background is more complex because of the similarities with the signal.

The discrimination is performed with two Boosted Decision Trees, one optimized to reject the B-like background, the other to reject the uds background. Twelve discriminating variables have been used, including shape variables, kinematic variables like m_{ES} and ΔE, and a variable related to the particle identification of the K candidate. The multivariate selection is preceded by a loose pre-selection, finalized to reduce the quantity of data to process without losing signal events, to train the selection on a reduced and simplified sample.
The selection has been trained on a signal sample produced with Monte Carlo simulation, and modified to have a flat diphoton mass distribution in the considered range: this has been done to avoid the introduction of biases in the training of the selection procedure, caused by the overspecialization over a specific signal ALP mass. In this way it is also possible to include in the selection variables that are dependent on the ALP mass, like the energy of the two photons.

The signal yield is finally extracted from a fit to the diphoton invariant mass distribution.

The efficiency of the entire selection (pre-selection and BDTs) is between 17% and 19% for the signal (depending on the ALP mass taken into consideration), while the efficiency for the background events of type B and uds is, respectively, 5.0×10^{-6} and 3.9×10^{-6}.

The *BABAR* experiment collected data during six main Runs: Run 3 has been used as a control sample to verify the validity of the background Monte Carlo samples. Only the Run 3 has been observed as this is a blinded analysis, and the analysts do not have access to the entire data sample before having completed the selection procedure, to avoid the introduction of any bias due to the optimization on the analyzed sample.
The diphoton invariant mass distributions for the Run 3 data and for the simulated data are in good agreement, nonetheless the simulation underestimates the total number of events with respect to the Run 3. To avoid biases deriving from an improper modelization of the Monte Carlo, the probability distribution function that models the background has been extracted directly from the control sample.

I performed a scan on the control sample, locally fitting the data with a function obtained as a sum of the background and the signal functions, for different ALP mass hypotheses. Assuming that the signal strength is negligible, we expect to find a signal yield distribution that is approximately Gaussian and centered in zero.

The next analysis steps will consist in the generation of Toy Monte Carlo to evaluate the performances of the fit procedure and in the quantitative evaluation of the systematic uncertainties.

My thesis activity did not only revolve around the analysis for the ALP search, but given the possibility to extend this search also to the upcoming Belle II data sample, it included a

collaboration with this experiment.
During the course of my thesis I participated to the quality tests performed on the double-sided silicon strip detector modules of the Belle II Silicon Vertex Detector, which were assembled in Pisa.

I also contributed to the Silicon Vertex Detector offline reconstruction software, developing and implementing the algorithm to extract the hit time of the particles crossing the detector. Using this time information it is possible to reduce the off-time background and thus improve the tracking performances.

The Belle II experiment is foreseen to collect an integrated luminosity equal to 100 times the one collected by *BABAR*; I also roughly evaluated the sensitivity on the branching ratio of the analyzed channel using the *BABAR* and the Belle II data samples.

Chapter 1

Physics Motivation

The Standard Model (SM) of particle physics is a well-established quantum field theory that coherently explains almost all the experimental results obtained in particle physics. It has been developed in the early 1970s, and it has repeatedly proved its solidity, having been able to predict particles and processes later experimentally confirmed. It incorporates all the known fundamental particles and describes three of the four known fundamental forces: electromagnetic, weak, and strong.

Even if the SM is currently the best theory describing the fundamental particles and the way they interact with each other, it still fails to provide a complete explanation of many known phenomena: the gravitational force, which is not yet included in the model; the dark matter (DM), a kind of matter that up to now has only been detected due to its gravitational effects; the dark energy, considered responsible for the accelerated rate of the expansion of the Universe; the baryon asymmetry, i.e. the different abundance of matter and anti-matter; the neutrino oscillation and their mass; the number of families of leptons and quarks and their important mass differences. It is to answer these questions that a large effort is devoted to search for physics beyond the Standard Model (BSM).

Section 1.1 briefly reviews the main aspects of the SM. In Section 1.2 an overview on the *axions* is provided, including the reason for their postulation, the current experimental situation, and the extension of the model to the *axion-like particles*. The axion-like particle object of my analysis is described in Section 1.3.

1.1 Standard Model Brief Overview

The SM description of Nature is based on fundamental particles and their interactions. A particle is considered fundamental if it has no internal structure. Particles can be classified depending on their spin, that can be integer or half-integer (in terms of $\hbar$): *bosons* possess integer spin, and *fermions* possess half-integer spin. Each particle has its own anti-particle, that in some cases coincides with the particle itself. Figure 1.1 shows the classification of the SM fundamental particles.

The known fundamental fermions all have spin equal to $\frac{1}{2}$. They are classified in two

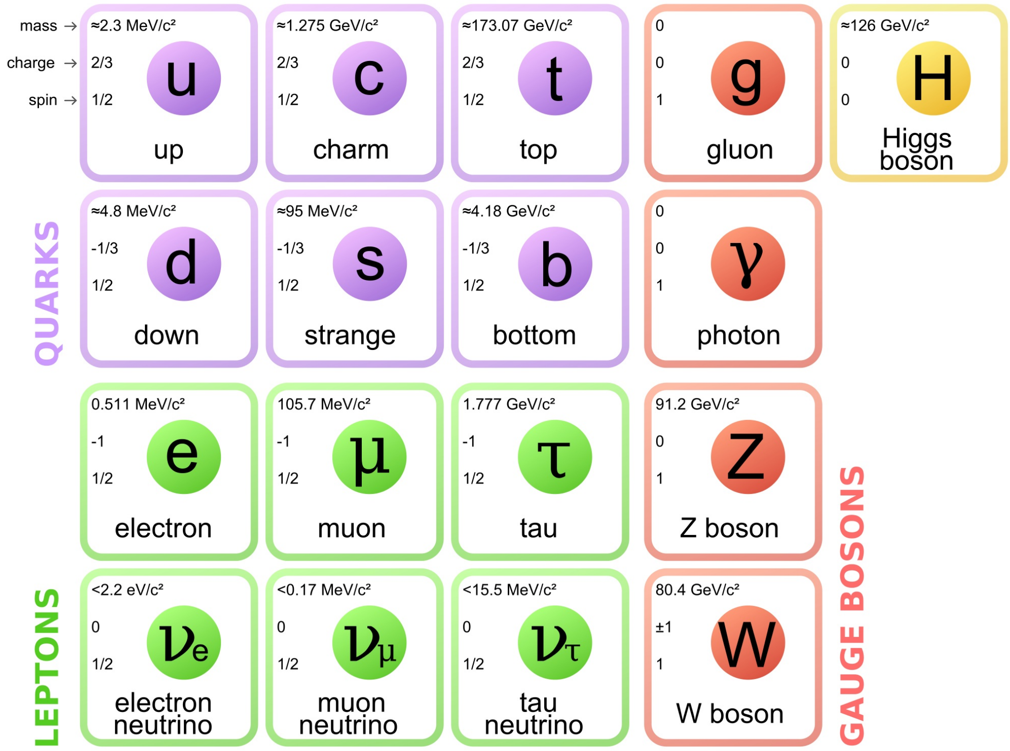


Figure 1.1: Fundamental particles included in SM model. For each particle its mass, electric charge, and spin are reported. The three generations of fermions are collocated in the first three columns: quarks are on the top, leptons on the bottom. The force carriers are placed in the fourth column. Quarks are subject to all the three forces; charged leptons are subject only to electromagnetic and weak forces; neutrinos are subject uniquely to the weak force. The hypothetical graviton is not included. Image from [1].

categories: *quarks* take part to all the interactions, while *leptons* do not interact strongly. Both leptons and quarks are grouped into three doublets, called *generations*. There is a great difference in mass between different generations, what is called a mass *hierarchy*.

The six different quarks (u, d, s, c, b, t) have a different *flavor*. Each of them also comes in three different *colors*, that represent the strong charge. Particles that do not interact strongly are *colorless*, like the leptons. The strong interaction is described by *quantum chromodynamics* (QCD). QCD includes a phenomenon known as color confinement, that implies that no free quarks can be observed: this means that quarks always form bound states. These can be made by two quarks (a quark and an anti-quark), and are called *mesons*, e.g. the π, or by three quarks (three quarks *or* three anti-quarks), and are called *barions*, e.g. the proton. Pentaquark states (i.e. baryons formed by five quarks) can in principle exist; there are some claims, but no fully convincing evidence of their existence. Quarks posses a fractionary electric charge: in each generation, one quark has an electric charge of $+\frac{2}{3}$ while the other has $-\frac{1}{3}$ (in terms of the elementary charge e), such that the difference in electric charge inside a generation is equal to 1.

The six leptons (e, ν_e, μ, ν_μ, τ, ν_τ) are colorless, so they can only interact weakly and/or electromagnetically. In each of the three generations, one lepton has unitary electric charge, and the other is neutral: like in the case of quarks, the intra-generation electric charge

difference is unitary. The neutral leptons can only interact weakly, and are known as *neutrinos* (ν).
Leptons posses two conserved quantum numbers: the *lepton number*, equal to +1 [-1] for leptons [anti-leptons] and 0 for particles that are not leptons, and a *lepton family number* for each of the three generation. The lepton family number is equal to +1 [-1] for the leptons [anti-leptons] of the given generation.
In experiments only left-handed [right-handed] neutrinos [anti-neutrinos] have been observed. If this is the case, neutrinos should be massless, because both helicities are required to have a massive fermion. The mass of the neutrinos has not yet been directly measured, but the fact that a neutrino can oscillate into a neutrino of another generation implies that it must be non null for at least two of them. Neutrino oscillation is the only way the lepton family number conservation principle is violated, as neutrinos do not decay, and it is part of the BSM physics.

All the known stable matter is composed uniquely by u and d quarks and electrons e, i.e. by the charged first-generation fermions.

Each of the three fundamental forces described by the SM is associated with a *force carrier*. All of these carriers are vector (spin 1) bosons.

The gluon g is the mediator of the strong interaction and is the only colored boson (it can come in eight colors); the photon γ is the mediator of the electromagnetic force; the Z and the $W^{\pm}$ are the mediators of the weak interactions.
The gluon and the photon are massless, while the Z and the W bosons are massive. The only charged boson is the W.
The *Higgs boson* H is the scalar (spin 0), massive boson which is responsible for the acquisition of mass by the otherwise massless particles through the mechanism of spontaneous symmetry breaking.
Finally, a hypothetical tensorial (of spin 2) massless boson would be the responsible for the gravitational interaction, known by the name of *graviton*, but up to now there are no evidences of its existence.

1.2 Axions

While the SM has repeatedly proved itself a very strong theory capable of accurate predictions, it still presents a number of open questions and shady areas. One of these is the *strong CP problem*.

The operator C represents the charge conjugation, the operator P is the inversion of spatial coordinates, and the operator T is the application of time-reversal. The CPT theorem states, under reasonable and well-tested hypotheses such as Lortentz invariance, that CPT-symmetry is an exact symmetry, meaning that all processes are invariant under a CPT transformation. It is known that C, P, and T symmetries can be individually violated, for example in weak interactions. Also the CP-symmetry is not preserved by the weak interactions: the first experiment showing this was carried out by Cronin and Fitch in 1964 and examined the decays of neutral kaons [2], while the predicted CP violation (CPV) in the B

meson decays was first observed in 2001 by the first generation of B factories.

The strong CP problem arises from the fact that such kind of violation is actually allowed in the strong processes, although it has never been experimentally observed: the QCD Lagrangian can be written such that terms allowing CPV are present:

$$\mathcal{L}_{QCD} = -\frac{1}{4}G_{\mu\nu}G^{\mu\nu} - \frac{g_s^2\theta}{32\pi^2}G_{\mu\nu}\tilde{G}^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - me^{i\theta'\gamma_5})\psi \tag{1.1}$$

Providing a detailed description of the QCD Lagrangian is beyond the purpose of this section; the interested reader can be found more details in [3, 4].

The term $G_{\mu\nu}\tilde{G}^{\mu\nu}$ is CP violating: an analogy can be carried out with the similar tensorial term present in the electromagnetic Lagrangian, $F_{\mu\nu}\tilde{F}^{\mu\nu}$, as they are built in a similar way. It can be shown that $F_{\mu\nu}\tilde{F}^{\mu\nu} = -4\vec{E}\cdot\vec{B}$, where $\vec{E}$ is odd (i.e. it changes sign) under C and P and even (i.e. it does not change sign) under T; $\vec{B}$ is odd under C and T and even under P. This implies that the term $\vec{E}\cdot\vec{B}$ is odd under CP. If the parameter θ is non null, CPV occurs.

The parameter $\theta\prime$ is manifestly CP violating since it appears in a complex exponential, which changes sign under CP.

The parameters θ and $\theta\prime$ can indeed be interpreted as angles, so a priori, their values should equiprobably range in $[0, 2\pi)$. $\theta\prime$ can be converted into a contribution for θ to obtain a total effective angle $\bar{\theta}$. There is no experimental evidence of CPV by the strong interactions (the strongest limit coming from the null electric dipole moment of the neutron [5]), hence the angular parameters of the QCD Lagrangian must be very close to zero ($|\bar{\theta}| < 10^{-10}$): this is a so-called *fine-tuning* problem, that is a situation where a theoretically allowed parameter must be finely tuned to maintain consistency with the experiments.

The most famous and promising solution to the strong CP problem has been proposed by Peccei and Quinn (PQ) [6, 7]. They promoted the parameter $\bar{\theta}$ to a dynamic field, by adding a new symmetry that is spontaneously broken, hence resulting in a new particle, called *axion*. This particle, sitting at the bottom of its field potential, would naturally make the CP violating parameter null, without requiring fine-tuning.

In this model, the coupling constant f_A for the axion-photon-photon interaction vertex and the mass m_A of the axion are not independent, as an inverse proportional relation exist:

$$f_A m_A = f_\pi m_\pi \tag{1.2}$$

with $f_\pi = 92$ GeV and $m_\pi = 135$ MeV.

1.2.1 Experimental Searches for Axions

Given Equation 1.2, no theoretical prediction can be obtained on the mass of the axion. In terms of interactions, the axion can couple to gluons and fermions, but also presents a characteristic two-photons vertex, that represents the most interesting channel experimentalists can exploit to search for these particles. This vertex makes the oscillation of an axion into a photon and vice versa possible, for instance in the presence of a strong magnetic field. Some ongoing or upcoming experiments searching for the axion, with different methods, are reported below:

- **QUAX** [8]: the QUest for AXion experiment aims to detect axions thanks to their interaction with the spin of fermions. The assumption is that axions are a component of the DM, hence the motion of the Solar System in the Milky Way would cause an effective cosmological "axion wind" on Earth. A material magnetized with an external magnetic field B_0 acts as the detector, as the effect of the axion wind on such a material would be a spin oscillation with a frequency given by the axion mass and an amplitude given by the coupling constant between axions and fermions. The observable would be the variation of the magnetization of the material. The sensitivity to the axion mass is given by the strength of B_0, that determines the oscillation Larmor frequency. QUAX can search axion with a mass $50\mu\text{eV} < m_A < 1.5\text{meV}$;

- **AXIOMA** [9]: the working principle of this experiment is the same as for QUAX, the difference being the detection procedure. AXIOMA uses a laser with a frequency ν to hit the target: the material, in the ground state L_0, is transparent to the laser frequency, as the energy gap between the ground state and an excited level L_2 is larger than the energy $h\nu$ of the photons of the laser. The interaction of an axion from the axion wind with the material could excite an electron from the ground state to an excited state L_1. The laser frequency is tuned such that it allows the transition between L_1 and L_2: the electron reaches the excited level L_2, and decaying to a lower level it emits a fluorescence photon than can be detected;

- **ADMX** [10]: the Axion Dark Matter eXperiment uses a resonant microwave cavity permeated with a strong magnetic field ($\sim$ 8 T). The magnetic field enhances the conversion of axions from the galactic halo of DM into a microwave photon, that is resonantly multiplied by the tuned cavity, and the very tiny energy deposited measured. The cavity can be tuned to different frequencies, hence to different axion masses: the ADMX experiment has excluded the $1.9 \div 3.5$ μeV axion mass region;

- **STAX** [11]: this experiment exploits the two-photons vertex of axions, aiming to convert a photon to an axion using a strong magnetic field, make the axion pass through a barrier (a *wall*), and re-convert it into a photon afterward. The effect is that a photon seems to have passed through a wall, hence the naming of Light-Shining-Through-Wall experiments. These sort of experiments have to face the low rate due to the double oscillation required, hence they maximize the luminosity using a powerful laser source ($\sim 10^{29}\gamma$/s) in a controlled laboratory environment. Another key point is to use a very strong magnetic field ($\sim$ 15 T) to augment the transition rate;

- **Astronomical** experiments: the axion-photon conversion induced by strong magnetic fields is a process exploited in astronomical experiments too. Experiments such as CAST [12] and IAXO [13] search for axions emitted by the Sun, and try to convert them into photons using strong magnetic field close to the photon detectors. Galactic and extragalactic experiments try to look at the difference in the early Universe opacity deriving from the reduced amount of photons that oscillate into axions due to interstellar magnetic fields.

All experiments have given negative results up to now. The current exclusion limits for PQ axions are shown in Figure 1.2.

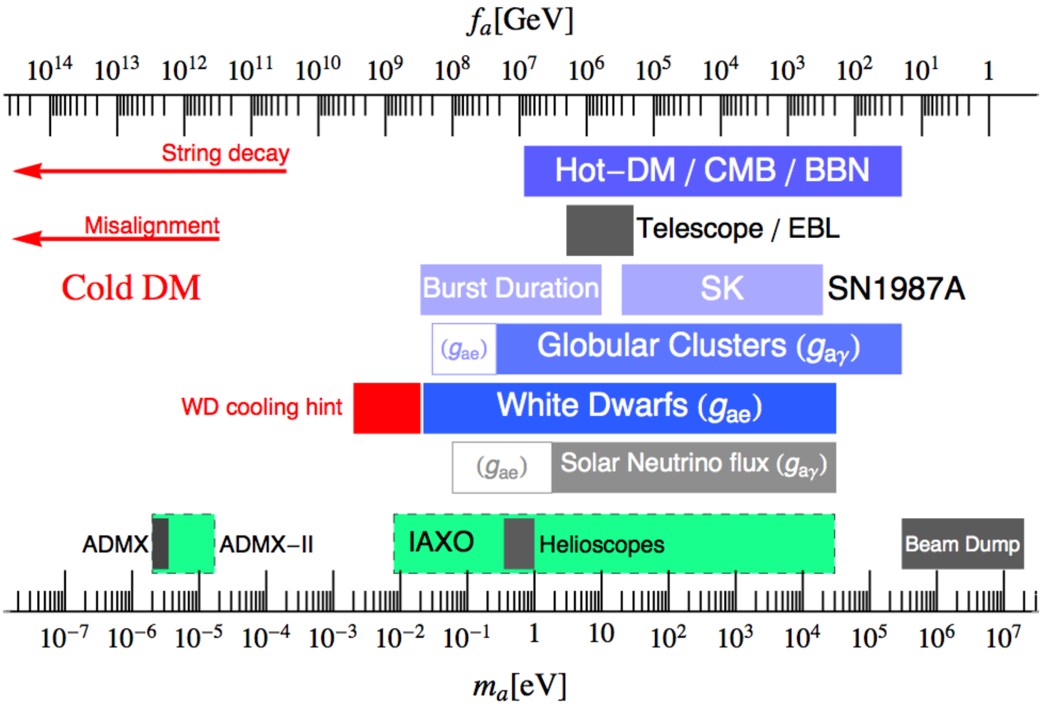


Figure 1.2: Exclusion limits of the parameter space for PQ axions. Blue and gray areas respectively correspond to constraints from cosmological or astrophysical arguments and constraints from astronomical observations. Light green areas are the sensitivity of planned experiments. Plot from [14].

1.2.2 Axion-Like Particles (ALPs)

The physical mechanism that would allow the existence of the axion, i.e. the spontaneous breaking of an approximate symmetry, can also generate other *axion-like particles* (**ALP**s, I will refer to them simply as A). Unlike axions, which are linked to the strong interactions and whose mass and coupling are determined by a single new parameter f_A, ALPs are much less constrained, and their masses and couplings to photons are independent parameters. ALPs are pseudoscalar neutral massive particles that couple predominantly to gauge bosons, hence ALPs can couple to pairs bosons as gg, $\gamma\gamma$, ZZ, γZ, W^+W^-, etc, the specifics depending on the model.

Many models have been developed about ALPs. As any hypothetical, feebly interacting particle, they are DM candidates. Appropriately tuned ALPs could help solve astrophysical puzzles such as the anomalies in the energy loss of white dwarf stars [15] and in the transparency of the Universe to TeV gamma rays [16].

There exist ALP models that are extensions of the original PQ model, as the *axiflavon* [17]. In this model the symmetry breaking also explains the *flavor puzzle*, i.e. why the fermion masses are hierarchical. This ALP would also constitute a DM candidate. All the constraints coming from having imposed the solution of these three issues at the same time makes this ALP model quite predictive.

Other ALP models predict ALP coupling with SM fermions, and emission of an ALP in flavor changing neutral current (FCNC) processes, or coupling with Higgs and emission in Higgs loops [18].

Current exclusion limits for ALPs with direct coupling with photons are shown in Figure 1.3.

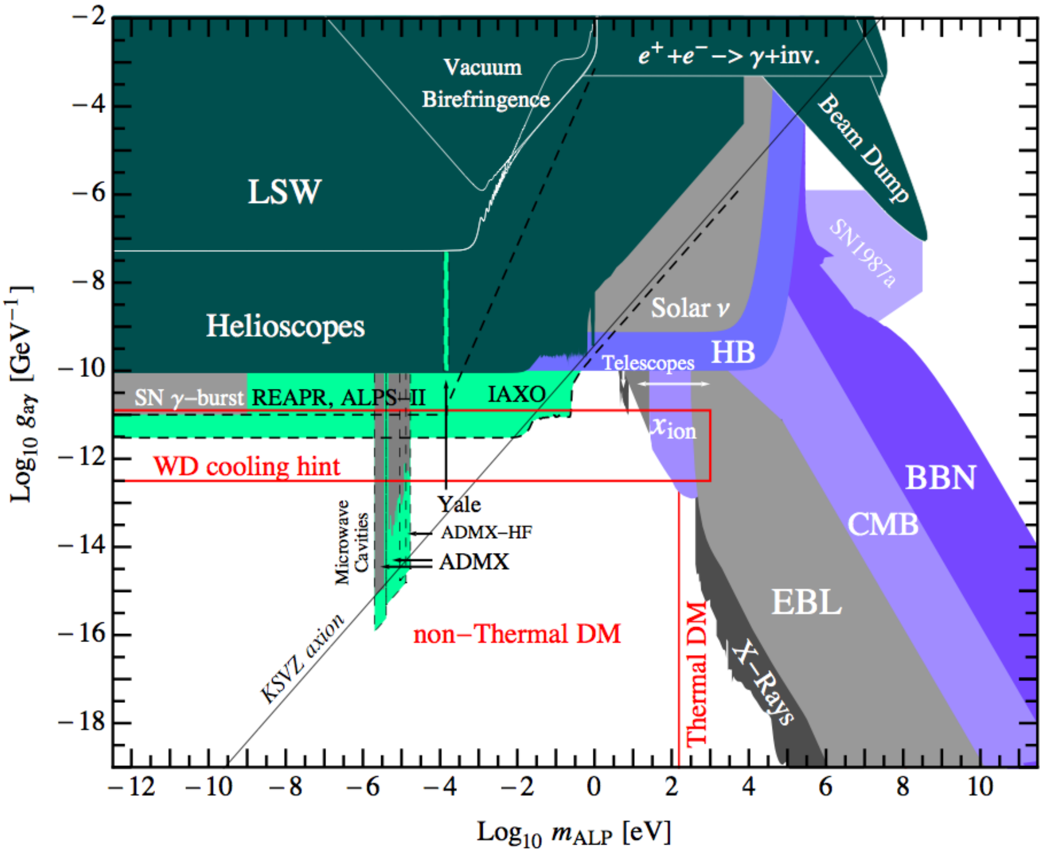


Figure 1.3: Exclusion limits of the parameter space for ALPs (mass and coupling constant). The QCD axion model lies around the "KSVZ" line (a particular axion model). Blue, gray, and dark green areas respectively correspond to constraints from cosmological or astrophysical arguments, constraints from astronomical observations, and experimentally excluded regions. Light green areas are the sensitivity of planned experiments. Plot from [14].

1.3 The $B \to K^{(*)}A$, $A \to \gamma\gamma$ Process

My analysis revolves around the search of an ALP; the model I refer to is described in detail in [19].

This model is characterized by the fact that the coupling of the ALP to the weak gauge bosons

$W^{\pm}$ gives rise to observable signatures, differently from most of the other ALP models where the main effective coupling is with photons and/or gluons. This ALP model has a zero coupling with gluons. The effective Lagrangian is given by:

$$\mathcal{L} = (\partial_\mu a)^2 - \frac{1}{2} m_a^2 a^2 - \frac{g_{aW}}{4} W_{\mu\nu} \tilde{W}^{\mu\nu} \tag{1.3}$$

where the g_{aW} is the coupling between the ALP field a and the electroweak gauge boson field W ($\tilde{W}^{\mu\nu} \equiv \epsilon^{\mu\nu\alpha\beta} W_{\alpha\beta}/2$).

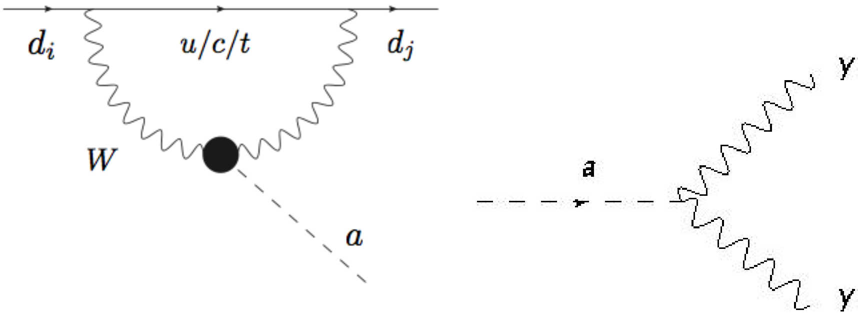


Figure 1.4: Diagrams for production and decay channels of the ALP a. The ALP is emitted from a W during a FCNC process. In case of $B \to K^{(*)} A$, the d_i down-type quark on the left is a b [$\bar{b}$] quark, and the d_j down-type quark on the right is a s [$\bar{s}$] quark. Left image from [19].

ALPs like this, with AWW coupling, can be emitted in FCNC processes (see Figure 1.4). Since FCNC processes are forbidden at the tree level in the SM, promising channels for ALP detection are the following:

$$B \to K^{(*)} A,\ A \to \gamma\gamma \tag{1.4}$$

or

$$K \to \pi A,\ A \to \gamma\gamma \tag{1.5}$$

As stated in Section 1.2.2, there are no constraints from the theory on the ALP mass m_A, so the only limits come from previous experiments and from the kinematic of the analyzed process. Roughly speaking, the analyzable ALP mass range goes up to 5 GeV for the B decay process, the exact values depending on the process under examination. In these conditions we assume that the only relevant decay channel for the ALP is $A \to \gamma\gamma$.

Channel 1.4 is of particular interest because the decaying particle is a B meson: this is relevant because of the existence of accelerators and detectors optimized to generate and detect B mesons, known as *B-factories*; Chapter 2 is dedicated to examine their features. What is relevant to the current discussion is that large data samples of B decays, obtained in a well-known and clean environment, are available for this kind of analysis, making it worth to be carried out. Channel 1.4 includes the decays of both the neutral and the charged B mesons. No search has been carried out for such decay, and the SM continuous process $B \to K^{(*)}\gamma\gamma$ has never been measured.

Figure 1.5 shows the current exclusion limits for this process, as well as the projections for ALP searches for two branching ratio (BR) sensitivities (10^{-6} and 10^{-8}). The regions around the masses of π^0, η, and $\eta\prime$, as will be specified later in the text, have been excluded, as they represent an irreducible background. It has been requested that the ALP decays within 30 cm from the interaction point, in order to be observable.

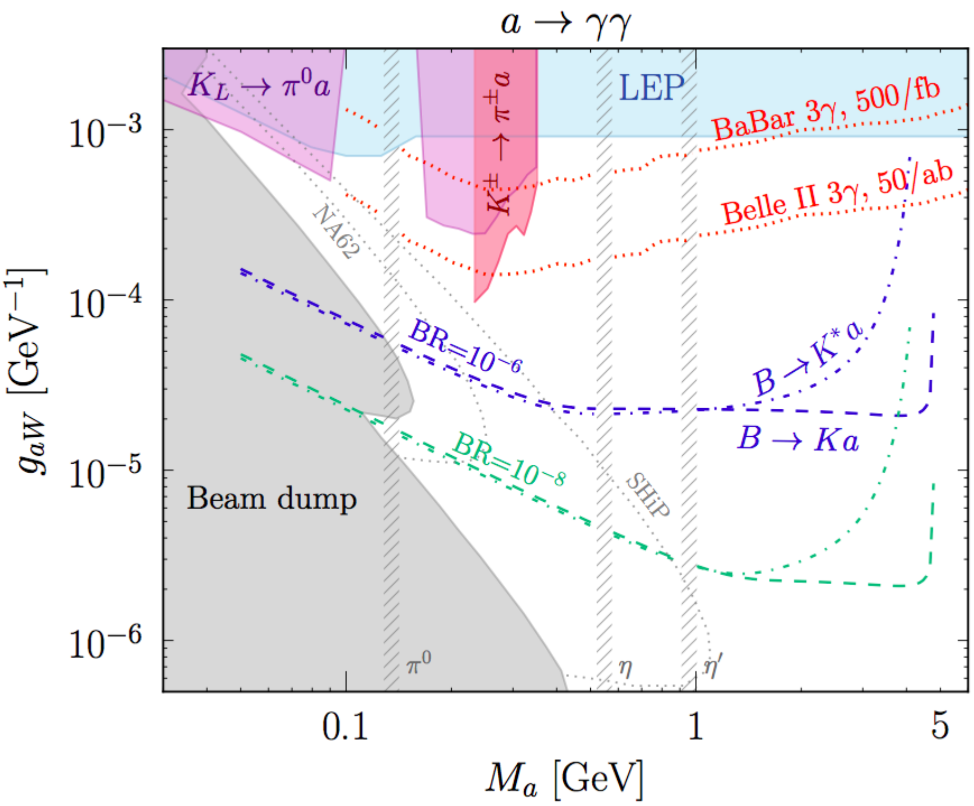


Figure 1.5: Sensitivity estimation for the $B \to K^{(*)}A$ process, in the ALP parameters space, assuming that the ALP decays into two photons. Given a branching ratio of $10^{-6[-8]}$, the blue [green] lines represent the estimated sensitivity. The dashed lines are for the $B \to KA$ process, while the dot-dashed lines are for the $B \to K^*A$ process. Dotted red lines represent a process that is not taken into consideration in this thesis, i.e. $e^+e^- \to A\gamma$, $A \to \gamma\gamma$. Grey dotted lines indicate projections relative to proposed beam-dump experiments. Parameters regions excluded from beam-dump experiments and from LEP are shown in shaded grey and azure areas respectively. The shaded purple and red areas are relative to the $K \to \pi A$ decays, that are beyond the purposes of this thesis. Plot from [19].

The red dotted lines in Figure 1.5 are related to another model where the ALP is produced in association with a photon from a e^+e^- annihilation:

$$e^+e^- \to A\gamma,\ A \to \gamma\gamma \tag{1.6}$$

The limits from *BABAR* and projection on Belle II are reported in the Figure.

The channel $B \to K^{(*)}A$, $A \to \gamma\gamma$ is predicted also by other ALP models where the ALP only possess a direct coupling with quarks; in these scenarios, though, the diphoton decay mode of A is dominant only below the π mass threshold, while in my case the decay $A \to \gamma\gamma$ is dominant throughout all the allowed diphoton mass spectrum.

Chapter 2

B Factories

Before the experiment by Cronin and Fitch it was known that the Universe is not symmetrical under C nor P, but there were no evidences for violations of the composed operator CP. This experiment provided the first evidence for the CPV and therefore that Physics is not the same for matter and antimatter.
In 1973 Kobayashi and Maskawa demonstrated that CPV could be included in the SM assuming the existence of 6 different quark flavors, because of an irreducible complex phase in the weak interaction quark-flavor-mixing matrix, called Cabibbo-Kobayashi-Maskawa (CKM) matrix [20].
At that time, though, only 3 quarks (u, d and s, the lightest ones) were known. By 1980, after the discoveries of the c and the b quark, the CKM matrix was included in the SM, and predictions about the magnitude of the CPV in charm and beauty sectors were performed. Large CPV was expected in neutral B mesons decaying to CP eigenstates (*e.g.* $B^0 \to J/\psi K_S^0$). The CLEO experiment tried to look at it, but it failed because of the low BR of the decay, order of 0.1%.

The goal of measuring the CP violation in the $B^0\bar{B}^0$ system could only be achieved with a very large amount of B mesons, and was because of this that the first B factories were conceived and the first related experiments, *BABAR* (1999-2008) and Belle (1999-2010), were built. Their successor is the Japanese experiment Belle II, whose data taking will start in 2018.

The main features of those detectors were modeled with the goal of measuring CP asymmetries, in particular exploiting the time-dependent method. In addition they also performed precise measurements of the CKM matrix elements and of several branching ratios of rare B meson decays, putting constraints on SM parameters. These experiments also allow the study of the charm and tau lepton physics, as well as two-photon physics.
To reach such goals a B factory needs to achieve a *high luminosity*, as the BR for the most interesting process to study CPV ($B^0 \to J/\psi K_S^0$, $J/\psi \to l^+l^-$, $l = e, \mu$) is $\sim 5 \times 10^{-5}$. It is also necessary to *boost* the produced $B\bar{B}$ pairs to increase their respective displacement and distinguish between them, as will be explained in Section 2.1.

This Chapter provides a description of the features of the B factories and related detector.

Section 2.1 explains the reasons of the asymmetric beam energies and the choice of a particular center of mass (CM) energy, and introduces analysis techniques which exploit the features of the B factories. Section 2.2 provides an overview on the characteristics of the accelerator supplying the beams to the detectors, together with a description of the beams characteristics and a general introduction on the needs of the detectors.
Section 2.3 focuses on the features of *BABAR* detector, the one that collected the data used for the analysis I performed, while Section 2.4 describes the Belle II detector.

2.1 B Mesons Production at the $\Upsilon(4S)$ Energy

B mesons can be produced both in e^+e^- and in hadronic interactions. The production in hadronic interactions at high energy (for instance in $p\bar{p}$ collisions at 13 TeV at LHC) offers a large cross-section and a large spectrum for beauty mesons and baryons. On the other hand, electron-positron colliders provide a clean environment and a well-known initial state.

The *BABAR* and the Belle experiments were installed on the high luminosity e^+e^- colliders (the B factories) PEP-II and KEKB respectively [21], the first one at SLAC, California, and the latter at KEK, Japan. They collected data for a decade, and their successor will be Belle II, still located at KEK, using beams from the upgraded B factory SuperKEKB [22].

To obtain the required high number of $B\bar{B}$ pairs, the colliders run at CM energy of 10.58 GeV, corresponding to the invariant mass of the $\Upsilon(4S)$ resonance: this is a bound state of quark b and antiquark $\bar{b}$ (*bottomonium*), and its quantum numbers are $J^{CP} = 1^{--}$. This resonance is of particular interest because it is the first bottomonium state whose mass is high enough to allow it to decay into a B mesons pair ($m_{B^\pm} = 5.279$ GeV/c^2, $m_{B^0} = 5.280$ GeV/c^2). Since the difference in mass between a $\Upsilon(4S)$ and 2 B mesons is approximately 20 MeV/c^2, the two B are produced almost at rest in the CM frame.
B mesons have a short lifetime ($\tau_{B^\pm} = 1.638 \times 10^{-12}$ s, $\tau_{B^0} = 1.519 \times 10^{-12}$ s), hence if the CM frame and the laboratory frame are the same, the relative displacement is of the order of 60 μm. This was the situation for the CLEO detector as CESR, New York.

This small vertex distance is barely distinguishable, given the state-of-the-art of the vertex detectors. This issue pushed toward the conception of *asymmetric* accelerators, i.e. the usage of electronic and positronic beams with different energies, so that the resulting products are boosted in the laboratory frame. If the Lorentz boost $\beta\gamma$ is large enough, the vertex separation $\Delta z = c\tau\gamma\beta$ can be measured, allowing the determination of the proper time τ.

The high luminosity required for this kind of experiments, $\sim 10^{33}$ cm^{-2} s^{-1}, is the reason why these machines are called *B factories*.

Using e^+e^- B factories brings a number of advantages:

- High signal-over-background ratio, as it can be seen from the Table 2.1;
- A low rate ($\sim$ 10 Hz for *BABAR* and Belle), together with the low background, allows relatively simple trigger strategies;

- The detectors have an almost complete angular acceptance, $\sim 260°$ in θ and $360°$ in ϕ;
- The initial state and the energies of the B mesons in the CM are completely known, thus it is possible to set up kinematical constraints (see Section 2.1.1) and exploit the quantum mechanical correlation between the two B mesons;
- The total number of charged tracks is less than 15, as shown in Figure 2.1, reducing the combinatorial background and often allowing a full event reconstruction.

$e^+e^- \to$	**Cross-section** (nb)
$b\bar{b}$	1.05
$c\bar{c}$	1.30
$s\bar{s}$	0.35
$u\bar{u}$	1.39
$d\bar{d}$	0.35
$\tau\bar{\tau}$	0.94
$\mu\bar{\mu}$	1.16
e^+e^-	~ 40

Table 2.1: Production cross-sections at $\sqrt{s} = M\left(\Upsilon\left(4S\right)\right)$. The $e^+e^- \to e^+e^-$ cross-section is the effective one, expected within the experimental acceptance. Data from [23].

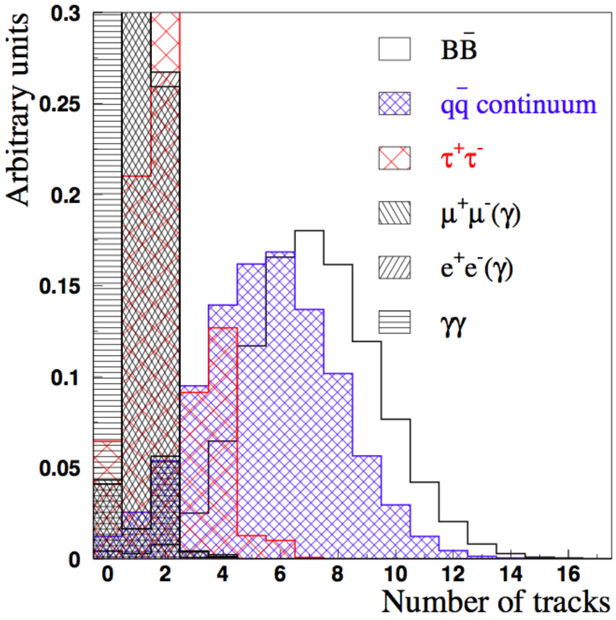


Figure 2.1: The number of charged tracks per event, for different processes. Image from [21].

2.1.1 Kinematical Constraints on $B\bar{B}$ Events from $\Upsilon(4S)$

In the CM frame of the $\Upsilon(4S)$, the two B mesons are produced back-to-back. Each one of them takes half of the total energy in the CM frame, that is equal to $\sqrt{s}$. The following relation holds:

$$E(e^{\pm})^{*}_{rec} = \frac{\sqrt{s}}{2} = E^{*}_{beam} \tag{2.1}$$

where the quantities with the * superscript are computed in the CM frame.
The two B mesons are produced almost at rest in the CM of the $\Upsilon(4S)$, having a kinetic energy of ~ 10 MeV.

Two useful variables can hence be defined: ΔE and m_{ES}, respectively the *energy difference* and the *beam-energy substituted mass.*
The energy difference is defined as:

$$\Delta E = E_B^* - E_{beam}^* \tag{2.2}$$

E_B being the energy of the B meson. Its expected mean value is 0 (MeV) for $B\bar{B}$ events. Its uncertainty depends on how well E_B^* and E_{beam}^* are known. Generally E_B^* provides the dominant spread, coming from the detector resolution, and can be negatively affected by Bremsstrahlung photons carrying away part of the energy of the decay products of the B.

The beam-energy substituted mass is derived from the B meson mass m_B:

$$m_B = \sqrt{E_B^2 - |\vec{p_B}|^2} = \sqrt{{E_B^*}^2 - |\vec{p_B^*}|^2} \tag{2.3}$$

The *BABAR* definition is the following:

$$m_{ES}^{BaBar} = \sqrt{\left(\frac{\frac{s}{2} + \vec{p}_B \cdot \vec{p}_0}{E_0}\right)^2 - |\vec{p}_B|^2} \tag{2.4}$$

where $(E_0, \vec{p}_0)$ and $(E_B, \vec{p}_B)$ are the four-momenta of the e^+e^- system and of the B meson candidate, in the laboratory frame.

Belle used a similar variable, called *beam-energy constrained mass*:

$$m_{BC}^{Belle} = \sqrt{(E_{beam}^*)^2 - |\vec{p}_B^*|^2} \tag{2.5}$$

Both expressions are expected to yield the B meson mass, but they exploit the kinematical constraint on the B energy in the CM to improve the resolution and reduce the correlation with ΔE.
The two equations are exactly equivalent only for symmetric colliders, but in practical terms they are very similar one to each other, therefore it is possible to switch between the two of them. I will henceforth refer only to the *BABAR* definition, because I have performed my analysis with *BABAR* data (so $m_{ES} \equiv m_{ES}^{BaBar}$).
The uncertainty on m_{ES} is dominated by the spread in the beam energy. The signal resolution in m_{ES} is much less affected by the uncertainty in the measured B meson four-momentum compared to ΔE. For signal events, m_{ES} is peaked at the mass of the B meson, while for continuum events, composed of light quarks, the distribution is flat, with an endpoint at $\sim \frac{\sqrt{s}}{2}$.

The m_{ES} resolution is about 3 MeV/c^2 when no neutral particles contribute to the final state. The resolution for ΔE more strongly depends on the B meson decay mode: it is larger for low mass final states such as $\pi^+\pi^-$ (order of 29 MeV) than for high mass final states such as $D^{(*)}\bar{D}^{(*)}K$ (between 6 and 14 MeV for modes with zero or one D^{0*} mesons in the final state).

Since both m_{ES} and ΔE are computed starting from the beam energy a correlation is expected, and this can be seen in Figure 2.2. The correlation becomes strong for final state with heavy particles because the spread in the beam energy dominates in both variables. This also happens if the final state comprehends hard photons. In other situations the correlation is weak, generally negligible.

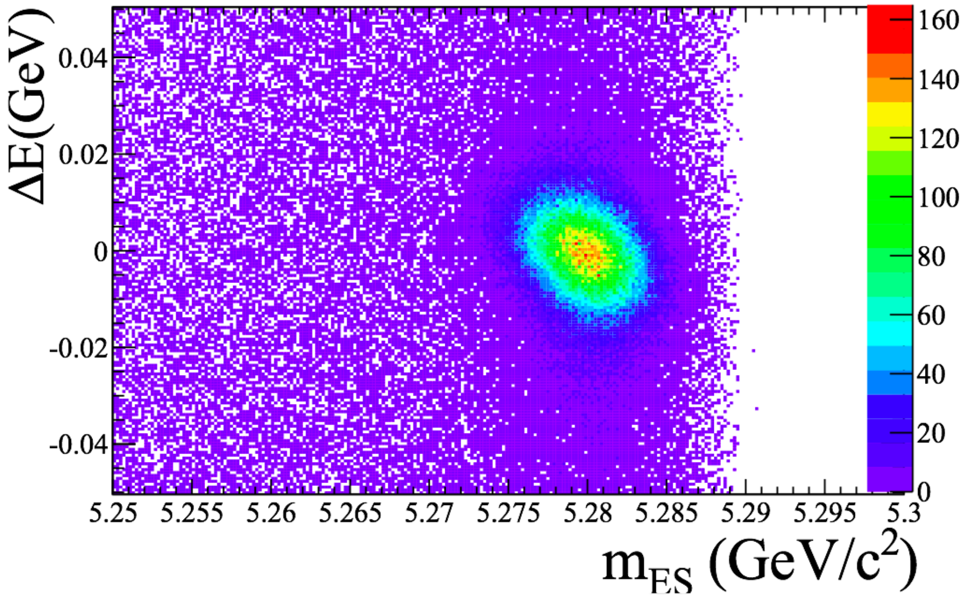


Figure 2.2: Distribution of m_{ES} and ΔE for $B^0 \rightarrow \Lambda_c^+ \bar{p} \pi^+ \pi^-$ events (Monte Carlo). The plot shows a strong correlation. Image from [21].

2.2 *B* Factories and Detectors

To provide adequate beams to the detector, an accelerator and a storage ring are required: *BABAR* was located on PEP-II and Belle on KEKB. A schematic view of the two rings is in Figure 2.3.

The accelerator needs to provide high instantaneous luminosity ($\sim 10^{33}$ cm^{-2} s^{-1}) and a boost of the CM frame sufficient to being able to study the time evolution of the B decays.

Asymmetric beams means dedicated storage ring for each of them. To preserve the high luminosity, one single interaction region (IR) was present. The lower energy beam was the positron one: the two rings were called High Energy Ring (HER) for electrons, and Low Energy Ring (LER) for positrons. These features were the same for both accelerators.

The luminosity of a e^+e^- storage ring can be written as the following:

$$\mathcal{L} = \frac{N_b n_{e^+} n_{e^-} f}{A_{eff}} \tag{2.6}$$

where $n_{e^\pm}$ is the number of $e^\pm$ inside a bunch, N_b is the number of bunches itself, f is the circulation frequency, and A_{eff} is the effective cross-sectional overlapping transverse area of

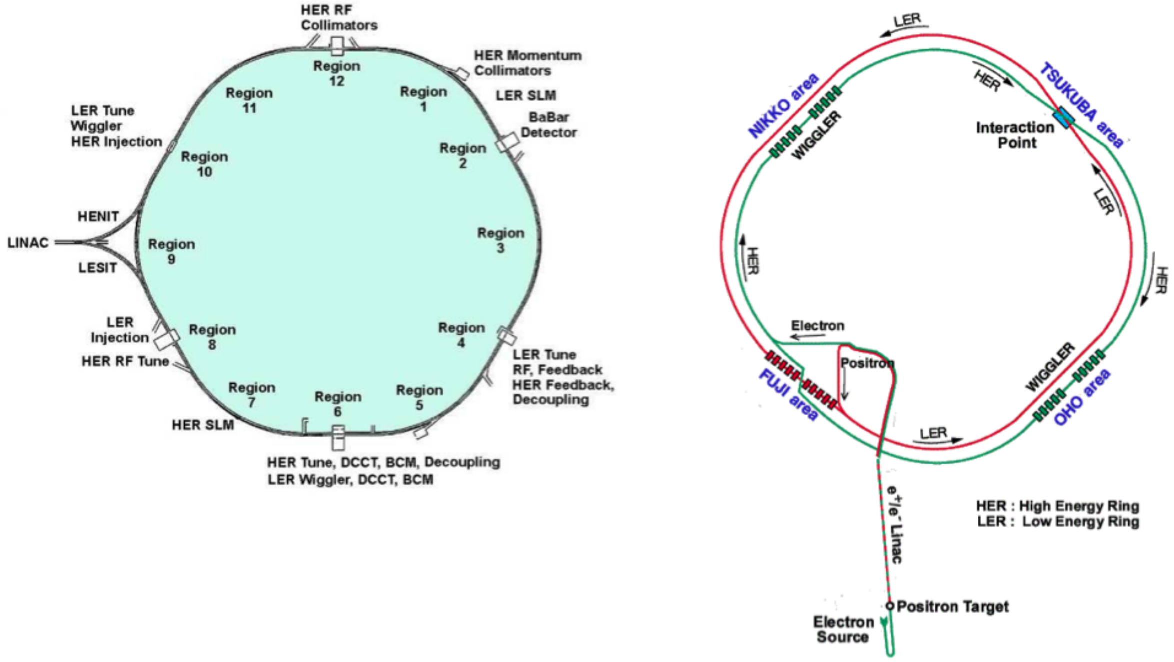


Figure 2.3: Schematic view of the PEP-II (left) and KEKB (right) rings. At PEP-II the two beams are stacked one on top of the other; the BABAR experiment is located in an experimental hall at the interaction region. At KEKB the two beams are side-by-side and intersect in the Tsukuba area experimental hall where the Belle detector was placed. Image from [21].

the beams at the interaction point (IP). A_{eff} increases when the product $N_b n_{e^+} n_{e^-}$ increases, hence limiting the luminosity at high currents.

Trying to augment the luminosity increasing N_b implies another problem, i.e. the non-linear interactions with the residual electromagnetic fields generated by the previous bunches, that can bring beam instabilities. The same problem arises after the crossing of the beams at the IP, so a beam-separation scheme is needed to divert the beams immediately after they leave the IP, to avoid parasitic interactions. This was achieved in different ways at the two experiments: PEP-II used a head-on collision scheme with near-IP bending magnets to steer the e^+ and e^- beam bunches away from each other as soon as possible after the collision, while KEKB opted for a scheme in which the two beams collided with a small crossing angle (± 11 mrad).

This last scheme allowed shorter bunch spacing and more available space for the detector components near the IP, but there were issues related to beam instabilities, that were overcome using so-called superconducting *crab cavities* that realigned the directions of the beam bunches, finally allowing KEKB reaching a peak luminosity of 2.1×10^{34} cm^{-2} s^{-1}, more than twice the original design goal.

A high quality vacuum is required to reduce beam losses caused by interactions with residual gasses. A highly efficient cooling system is also required due to the heating from synchrotron X-rays.

2.2.1 Integrated Luminosity and Working Parameters

Both experiments mainly ran at the $\Upsilon(4S)$ energy, collecting so-called *on-peak* events; a small but not negligible fraction of the data was taken at different energies, centered at other Υ resonances or *off-peak*, so collecting *continuum* events.
In particular, for what concerns *BABAR*, Runs from 1 to 6 (and part of Run 7) were taken at $\Upsilon(4S)$ energy; amongst them 90% of data was taken on-peak, while the remaining was taken ~ 40 MeV away from peak to study non-B backgrounds.
The remaining main part of Run 7 was taken at different Υ on-peak energies and some scans were performed off-peak.

The size of the data samples collected at various Υ and off-peak is shown in Table 2.2.

Experiment	**Resonance**	**On-resonance** Luminosity (fb^{-1})	**Number of Υ**	**Off-resonance** Luminosity (fb^{-1})
BABAR	$\Upsilon(4S)$	424.2	$(471.0 \pm 2.8) \times 10^6$	43.9
	$\Upsilon(3S)$	28.0	$(121.3 \pm 1.2) \times 10^6$	2.6
	$\Upsilon(2S)$	13.6	$(98.3 \pm 0.9) \times 10^6$	1.4
	Scan $> \Upsilon(4S)$	n/a	n/a	~ 4
Belle	$\Upsilon(5S)$	121.4	$(7.1 \pm 1.3) \times 10^6$	1.7
	$\Upsilon(4S)$ - SVD1	140.0	$(152 \pm 1) \times 10^6$	15.6
	$\Upsilon(4S)$ - SVD2	571.0	$(620 \pm 9) \times 10^6$	73.8
	$\Upsilon(3S)$	2.9	$(11 \pm 0.3) \times 10^6$	0.2
	$\Upsilon(2S)$	24.9	$(158 \pm 4) \times 10^6$	1.7
	$\Upsilon(1S)$	5.7	$(102 \pm 2) \times 10^6$	1.8
	Scan $> \Upsilon(4S)$	n/a	n/a	27.6

Table 2.2: Summary of the integrated luminosity of *BABAR* and Belle, sorted by CM energy, and number of produced Υ particles.

In Figure 2.4 *BABAR* and Belle integrated luminosities as a function of time are shown.

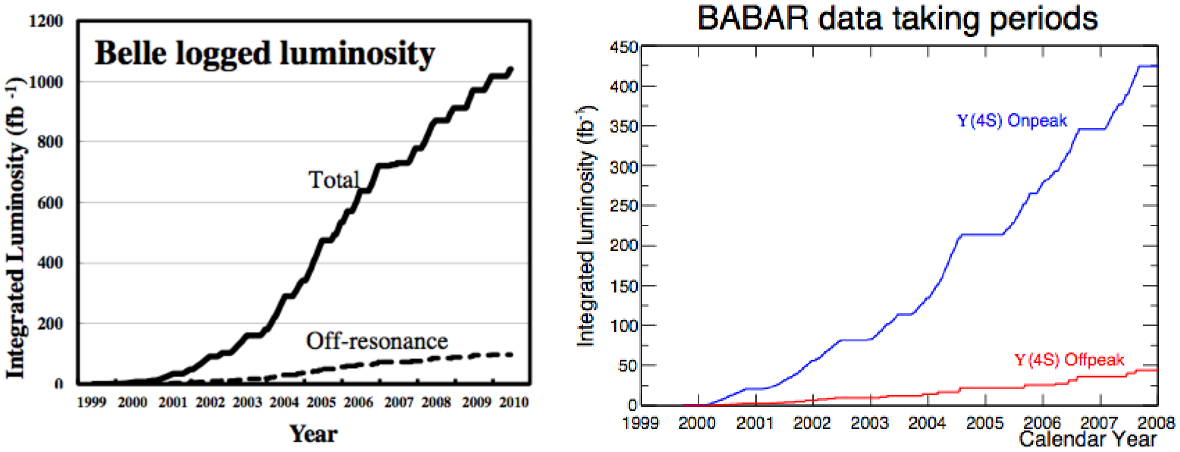


Figure 2.4: On the left [right]: evolution of Belle [*BABAR*] integrated luminosity. Image from [21].

The main parameters of the two accelerators are reported in Table 2.3; these are referring to the last stage of their operation.

Parameter	**Unit**	**PEP-II**	**KEKB**
e^- beam energy	GeV	9.0	8.0
e^+ beam energy	GeV	3.1	3.5
e^- beam current	A	1.8	1.2
e^+ beam current	A	2.7	1.6
Lorentz factor	$\beta\gamma$	0.56	0.425
	x μm	140	80
Beam size at IP	y μm	3	1
	z mm	8.5	5
Luminosity	cm^{-2}s^{-1}	1.2×10^{34}	2.1×10^{34}
Number of beam bunches		1732	1584
Bunch spacing	m	1.25	1.84
Bunch crossing angle	mrad	0 (head-on)	±11 (crab-crossing)

Table 2.3: Machine parameters for PEP-II and KEKB during the last stage of their operation.

2.2.2 B Factories Detectors

BABAR and Belle were conceptually similar, with minor differences deriving from the differences in the accelerator design and on some decisions of the two collaborations.

They were constituted by a number of subdetectors, arranged in a concentric structure where the most massive ones were placed in the outer layers. The main design features are the following:

- **Light material for the inner detectors** to reduce multiple scattering. In particular the beam pipe, in the active region, was made of berillium;
- **Vertexing capability** to distinguish the decay vertexes of the two B mesons. Silicon-strips-vertex detectors were used;
- **Particle Identification (PID)** to separate kaons from pions, and to identify K_L^0, K_S^0, and μ, necessary to properly measure CPV. Multiple methods were used to perform a good PID;
- **Electromagnetic calorimetry** to accurately measure the energy of the final states products;
- **Data handling capability** as huge amount of data was produced, so powerful computing infrastructures were needed.

In Section 2.3 a more detailed examination at the structure of the *BABAR* detector is provided. Since I have not used data from Belle (and it was not significantly different from *BABAR*), no extra details will be provided about it. An insight about Belle II will be provided in Section 2.4. All the detector located on an asymmetric B factory have a better coverage in the forward region (due to the Lorentz boost in that direction), while they are symmetrical around the beam axis. All the following images are taken from [21], unless otherwise specified.

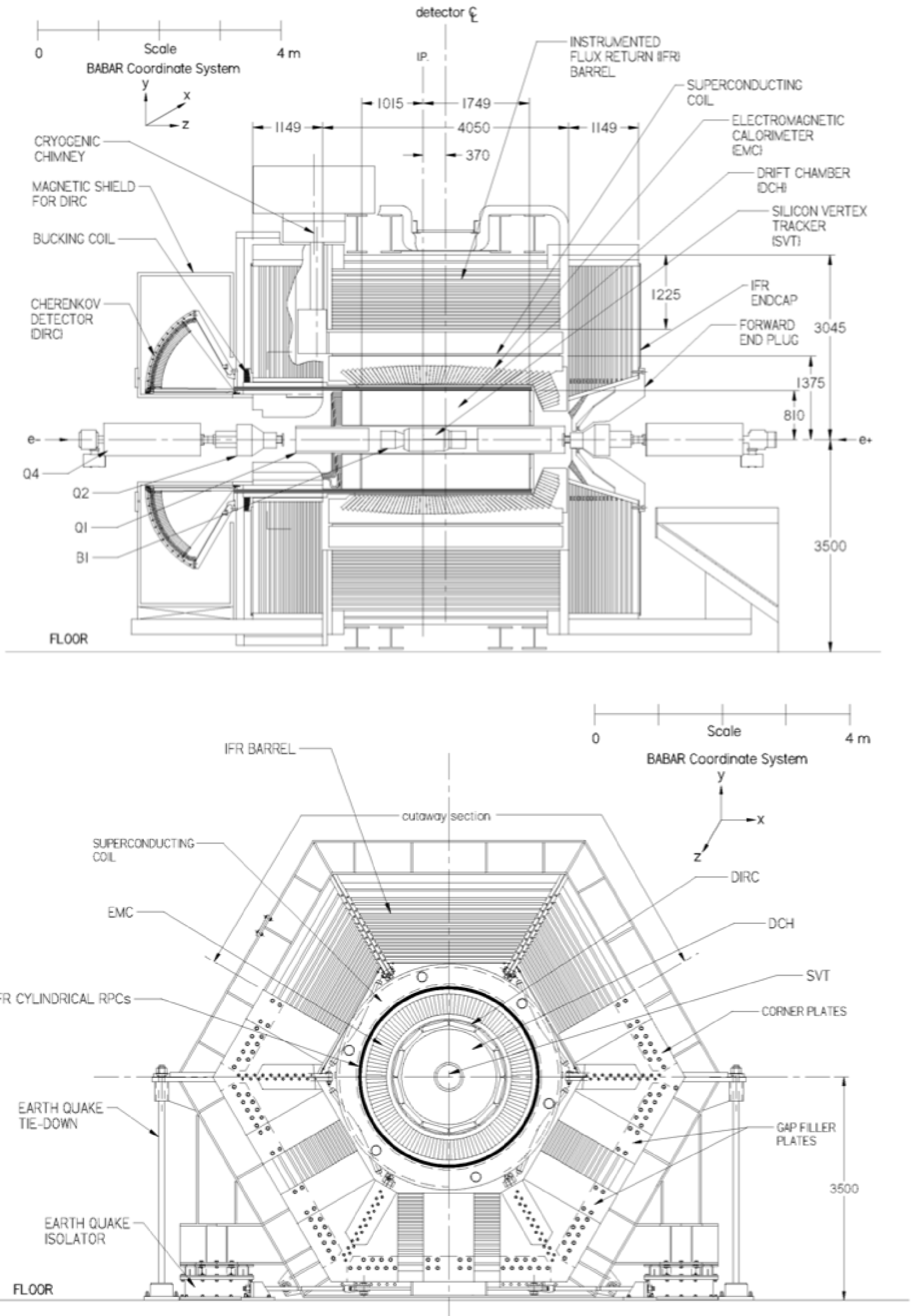


Figure 2.5: Top: longitudinal view of *BABAR* detector. Bottom: end view of *BABAR* detector.

2.3 The *BaBar* Detector

Longitudinal and end views of *BABAR* detector are shown in Figure 2.5.

Here is the list of the subdetector which *BABAR* was composed of, that will be explained in the following sections. More information can be found in [24].

- **Silicon Vertex Tracker (SVT)**: the innermost detector, whose main purposes were to perform vertexing and participate to tracking, particularly for low-momentum particles;
- **Drift Chamber (DCH)**: performed momentum and position reconstruction and participated to tracking and PID;
- **Detector of Internally Reflected Cherekov Light (DIRC)**: contributed to PID, particularly on discrimination between π and K;
- **Electromagnetic Calorimeter (EMC)**: performed energy measurement for electrons and photons, as well as K, p, π, μ, and contributed to PID;
- **Instrumented Flux Return (IFR)**: performed measurement of μ and K_L^0 energy.

2.3.1 Silicon Vertex Tracker (SVT)

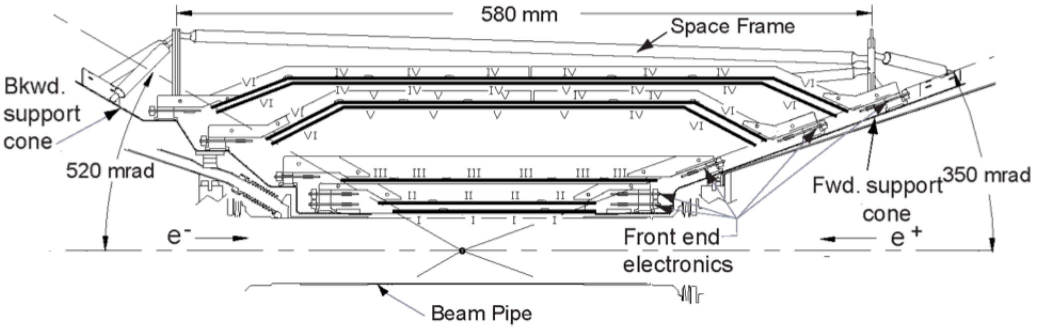


Figure 2.6: Longitudinal view of the *BABAR* SVT.

The SVT, whose scheme is shown in Figure 2.6, was the innermost *BABAR* subdetector and performed precise vertex measuring and contributed to tracking, particularly for particles with low momentum, because only particles with a momentum greater than 120 MeV/c could be reliably reconstructed in the DCH.

SVT was structured in 5 layers of silicon-strips sensors: three close to the beryllium pipe to perform vertexing and two at a larger radius to help pattern recognition. The inner layers were parallel to the beam axis, while the outer were arc-shaped to minimize the amount of material the particles had to pass through in the backward and forward regions.

Angular coverage was 100% in ϕ and was from 20° to 150° in the θ angle (laboratory frame), for a total coverage in solid angle of 90%.

Each of the 340 SVT sensors was double sided: on the two sides the silicon strips were parallel [transverse] to the beam and so measured the ϕ [z] coordinate.

Each module was electronically divided into 2 half-modules. For each of them the tracking efficiency was computed by dividing the number of hits associated to tracks with the number of tracks crossing that half-module: that resulted in an average efficiency above 95% over the nine years of operation. The z and $r\phi$ position resolution ranged between 15 and 40 μm, having their maximums for $\theta = 90°$ and degrading in the forward and backward directions.

The SVT also contributed to PID through the measurement of $\mathrm{d}E/\mathrm{d}x$: that allowed a 2σ separation between kaons and pions up to a momentum of 500 MeV/c.

2.3.2 Drift Chamber (DCH)

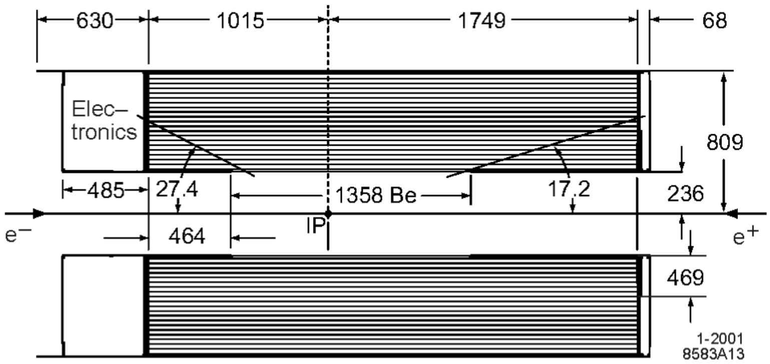


Figure 2.7: Longitudinal view of the DCH, dimensions are in mm.

The DCH, shown in Figure 2.7, was the main tracking detector of *BABAR*; in addition, it contributed to the PID (thanks to measurement of track ionization losses, $\mathrm{d}E/\mathrm{d}x$). Since the DIRC was only reached by high momentum particles and covered only the *BABAR* barrel section, the DCH and the SVT were the only subdetectors able to perform PID for forward and/or low momentum particles.

The DCH was constituted of 40 layers of hexagonal cells. Wires were made of aluminum and were immersed in a gas mixture of 80:20 Helium:Isobutane.
The 40 layers were organized in 10 superlayers, in which all layers had the same orientation. The orientation of the superlayers was studied to optimize the performance of the DCH.

The DCH was kept under constant control, in particular for what concerned the gas mixture composition, gas leakage and the high-voltage settings of each group of wires.
The voltage, in particular, was reduced in case the current of the corresponding wire was found to be too high, to prevent damaging.

DCH importantly contributed to PID thanks to a measured $\mathrm{d}E/\mathrm{d}x$ resolution of about 8%, very close to the design project of 7%.

2.3.3 Detector of Internally Reflected Cherekov Light (DIRC)

The DIRC was the main subdetector appointed to PID, in particular for π/K separation at high momentum; Figure 2.8 shows its scheme.

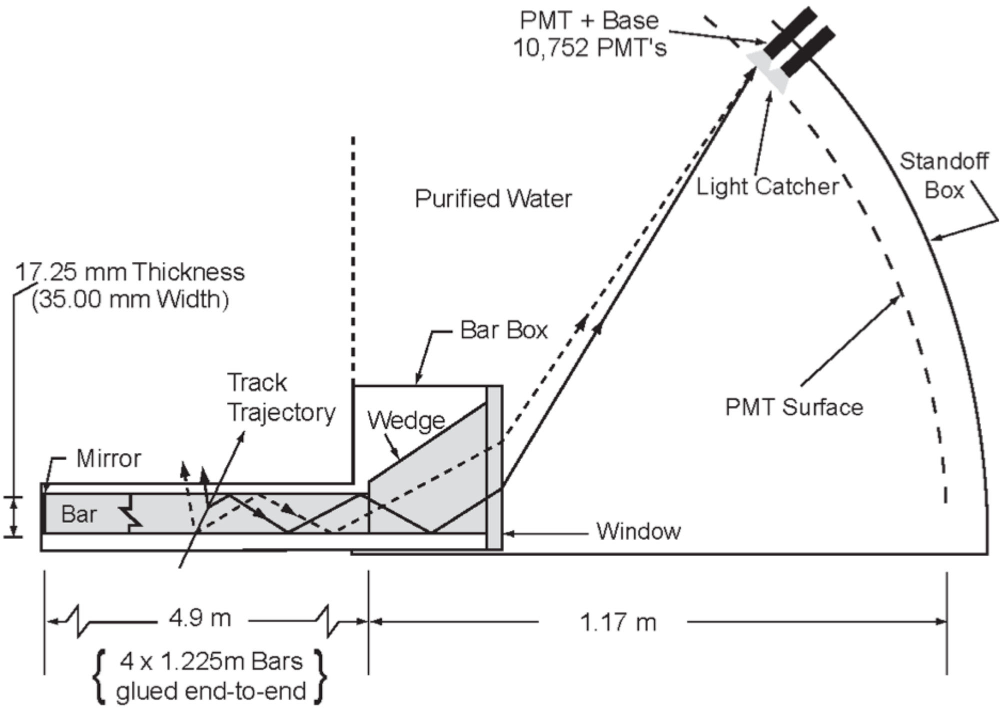


Figure 2.8: Longitudinal view of *BABAR* DIRC. Backward side, differently from top of Figure 2.5, is on the right side of the image.

The physical principle behind this detector is the Cherenkov light emission. This light was emitted by charged fast particles passing through a quartz bar radiator and was reflected inside of it thanks to total internal reflection; part of these emitted photons were emitted toward the backward direction, while the forward section was equipped with a mirror to reflect them to the backward region and enhance the number of collected photons.
The photons that came out of the quartz bar were then deflected by the quartz edge which reflected them at a large angle with respect to the bar axis. They finally traveled through a toroidal tank filled with ultra-pure water, and they could finally be detected by one of the 10752 photo multiplier tubes (PMTs) located beyond the backward end of the magnet.
Both position and arrival time of the photons was used to reconstruct the Cherenkov angle at which they were emitted.

The DIRC only covered the barrel section and due to the Cherenkov threshold was only able to detect particle traveling at a speed higher than c/n, n being the refraction index of the quartz. It also suffered from neutronic background: interactions between neutrons and H_2O could generate significant background photons.

Since the signal photons underwent an unknown number of reflections inside the quartz bar, the exact position and time the particle passed through the quartz bar were not precisely known. This uncertainty was partially overcome matching the photons to the extrapolation of charged tracks reconstructed in the DCH; the parameters were passed to an algorithm that, through a maximum likelihood formalism, produced as outputs 5 likelihood values, one for each of the 5 charged stable[1] particles (e, μ, π, p, K), an estimation of the Cherenkov

---

[1] These charged particles are considered stable as they live long enough to reach the tracking subdetectors.

angle and of the number of signal and background photons.

The achieved angular resolution was order of 10 mrad per photons and 2.5 mrad per tracks, 10% higher than design project. That led to a separation power between pions and kaons of more than 4 σ at a momentum of 3 Gev/c.

2.3.4 Electromagnetic Calorimeter (EMC)

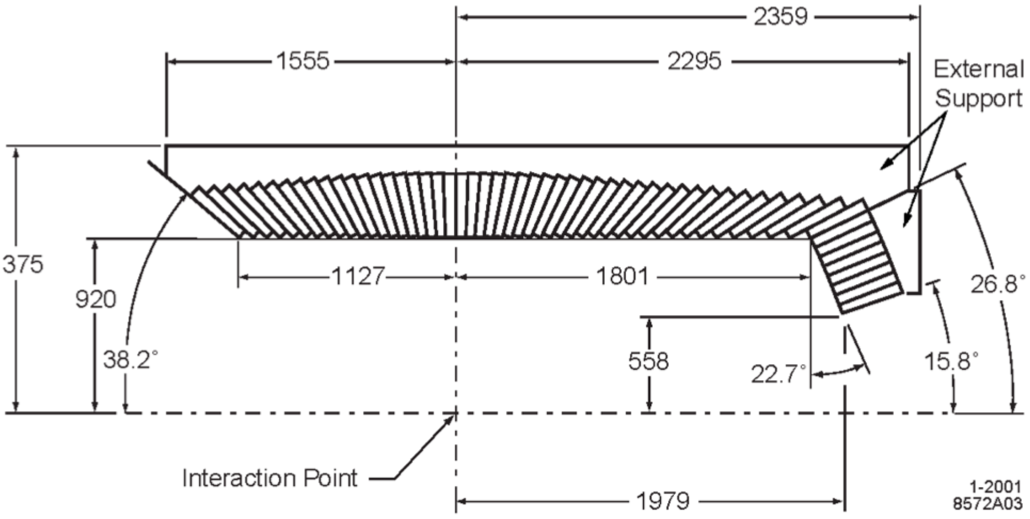


Figure 2.9: Longitudinal view of the *BABAR* EMC. Dimensions are in mm.

The EMC, shown in Figure 2.9, measured the energy of particles mainly interacting via electromagnetic interaction, i.e. photons and electrons.
The *BABAR* EMC was constituted by a total of 6580 CsI(Tl) crystals, organized in a barrel and a forward section. All of these crystals were arranged in a quasi-projective geometry, pointing to the IP. The slight offset was set up to avoid particles going through the gaps between the crystals, and passing undetected.

The crystals were supported on the external radius to reduce the interactions with material prior to the crystals themselves. The amount of material between the IP and the EMC ranged between 0.3 and 0.6 X_0, except for the most forward parts that saw the extra material of a longer beampipe and SVD readout systems, bringing the seen material up to 3 X_0.

Calibration of the EMC was performed with 2 methods: a low-energy one, that used a 6.12 MeV radioactive photon source, and a high-energy one, that used reconstructed Bhabha events.

The energy resolution (σ_E/E) varied from 5% to 2% going from 6.13 MeV to 7.5 GeV. The angular resolution ranged from 12 mrad at low energies to 3 mrad at high energies.
The π^0 mass was reconstructed with a value compatible with PDG reported one, with a resolution of 7 MeV/c^2.

2.3.5 Superconducting Solenoid

A strong magnetic field is needed to bend charged particles and so measure their momentum. It was achieved with a superconducting solenoid that provided a 1.5 T longitudinal magnetic field in the volume of the tracking detectors (SVT, DCH).
A cryostat was necessary to maintain the working temperature of the superconductor.

2.3.6 Instrumented Flux Return (IFR)

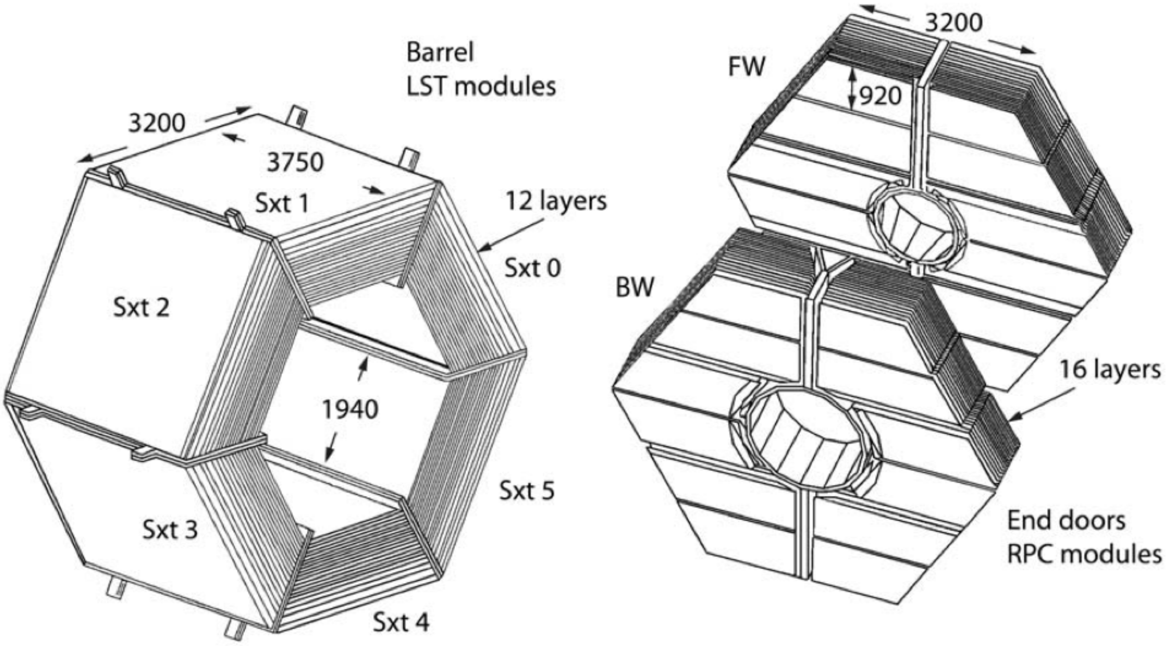


Figure 2.10: View of the *BABAR* IFR at the end of data taking period. Dimensions are in mm.

The IFR, see Figure 2.10, had two functions: it was used as a "Flux Return"er because it channeled the magnetic field generated by the solenoid, reducing leakage, and it was "Instrumented" because it exploited the steel as a absorber for μ identification and to generate showers from K_L^0. Active detectors sandwiched between the steel plates provided information on the penetration and shower development. The width of these steel plates ranged from 2 cm for the inner plates to 10 cm for the outer ones.

At the beginning, Bakelite-based Resistive Plate Chambers (RPCs) were used, but it became soon clear that they were inadequate, because of aging problems, so they were substituted with a system of Limited Streamer Tubes (LSTs). $6 \div 17\%$ of μ were lost because of this issue.

The LST efficiency was tested using dimuon events; on average it resulted being 88% compared to a geometrical acceptance of 92%, the difference being mainly due to malfunctioning or broken channels.

2.3.7 Triggers

BABAR used a two-level trigger to choose the events.

The first of the two triggers, Level-1 (L1), was hardware-based and accepted a rate of $\mathcal{O}$(kHz), while the second, the Level-3 (L3), was software-based and accepted a rate of $\mathcal{O}$(100 Hz) and had a few percent of dead time. These parameters have evolved during the period of data taking, following the increase in luminosity of the accelerator. The decision to accept or reject the event was taken by the Global Trigger (GLT), a central trigger system.

The L1 trigger used information coming from DCH, EMC and IFR; each of the three subdetectors provided its own trigger information. The information from these three subtriggers were sent to the GLT that processed them and sent the output to the central *BABAR* DAQ system.

The L3 trigger could be programmed with a variety of algorithms, based on simple logical OR. It was upgraded to handle the events coming from $\Upsilon(2S)$ and $\Upsilon(3S)$, that are different from the ones coming from the usual $\Upsilon(4S)$, having a lower multiplicity and lower visible energy than these last ones.

The *BABAR* trigger efficiency at the $\Upsilon(4S)$ energy was more than 99% for $B\bar{B}$ events, more than 95% for continuum events (i.e. $q\bar{q}$ events, $q = u, d, c, s$), and around 92% for $\tau^+\tau^-$ events.

2.4 The Belle II Detector

The Belle II experiment is an upgrade of Belle, both in terms of luminosity and precision. In the following sections its subdetectors will be examined. More details can be found in the Technical Design Report (TDR) [22].

A scheme with the top view of Belle II can be seen in Figure 2.11.

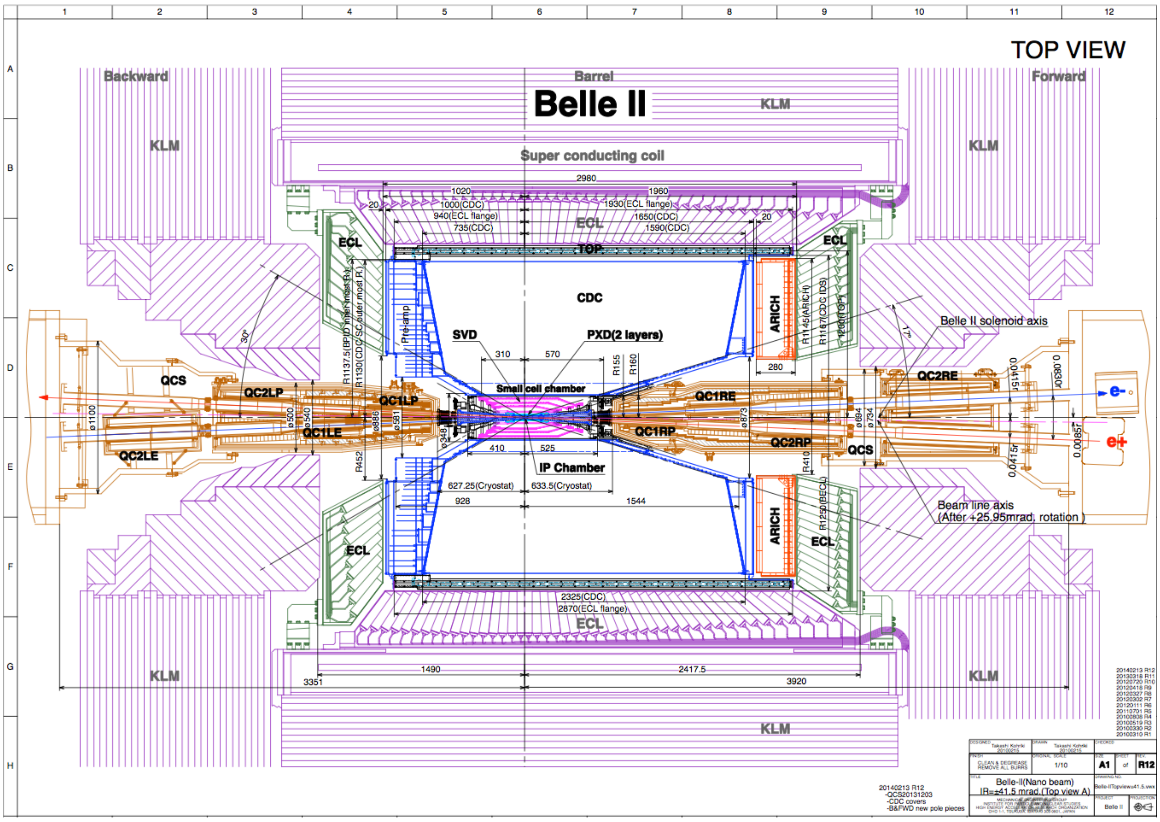


Figure 2.11: Scheme of Belle II detector.

2.4.1 Pixel Detector (PXD)

The Belle II experiment is designed to receive a ∼ 50 times higher instantaneous luminosity than Belle and *BABAR*, so a strip-based detector is not feasible at small distance from the IP because of the high occupancy deriving from the high background, that roughly increases with the inverse square of the distance from the IP. The beampipe radius in the interaction region is only ∼ 10 mm.
Another technology has hence to be used in place of the strip-based one: the solution is to use a pixel-based detector, that has a much larger number of channels and therefore a much smaller occupancy. The outermost layers can be based on strips, and so is the Silicon Vertex Detector (SVD), examined in Section 2.4.2.

The Large Hadron Collider has successfully implemented this scheme (pixel detector followed by strip detector) for its detectors, but it works at much higher energies than Belle II, therefore its sensors are much thicker than the ones Belle II can afford in terms of multiple

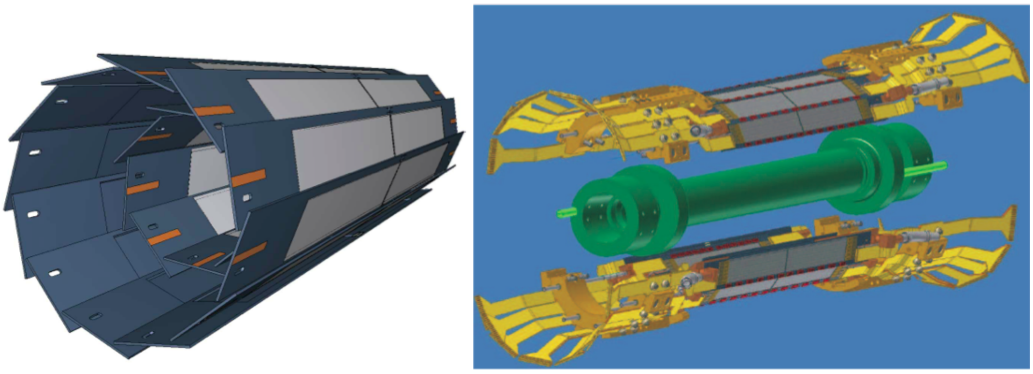


Figure 2.12: Left: scheme of the positions of the PXD sensors. The grey surface is the active one, that fully covers the ϕ angle. The PXD length is 174 mm. Right: scheme of the support structure. The two half shells are directly supported on the beampipe, in the Figure they are are displaced for the sake of comprehensibility.

scattering. The solution, for the Belle II PXD, is to use the DEPleted Field Effect Transistor (DEPFET) technology with very thin sensors, no thicker than 50 μm. Most of the readout electronics and its cooling system are located outside the acceptance region, hence not contributing to the multiple scattering.

The PXD (see Figure 2.12) is constituted by 2 layers of sensors, with radii of 14 and 22 mm from the IP. The inner layer consists of 8 planar ladders, each with an active surface of 15×90 mm^2; the pixels in this layer have an area of 50×50 μm^2. The outer layer is made of 12 ladders with an active surface of 15×123 mm^2; here the pixels have an area of 50×75 μm^2. In each ladder the pixels are assembled in rows, each of them consisting of 1600 pixels. The active surface covers all the angular acceptance, between 17° and 150° in the θ angle.

2.4.2 Silicon Vertex Detector (SVD)

Since my thesis work also included activities related to the SVD, some of its features are explained with greater details in Section 3.

At higher radii from the IP it is no longer necessary to adopt a pixel-based strategy for the tracking and vertexing detectors. The SVD, shown in Figure 2.13, uses Double-sided Silicon Strip Detector (DSSD, see Section 3.1.1) coupled to a fast readout with 50 ns shaping time, to efficiently reduce the high occupancy from machine background.

The combined PXD-SVD detector provides precision vertexing, as well as stand-alone tracking of low momentum particles. The SVD tracking capabilities are also used to suppress background hits in the PXD, thus reducing its readout data rate.

The SVD must be background resistant (background rate will be 30 times higher than the one of Belle, and the expected trigger rate is order of 30 kHz). It also must have low mass to minimize the amount of material a particle has to pass through to reduce multiple scattering, but at the same time it is necessary to have a cooling system in the active area to

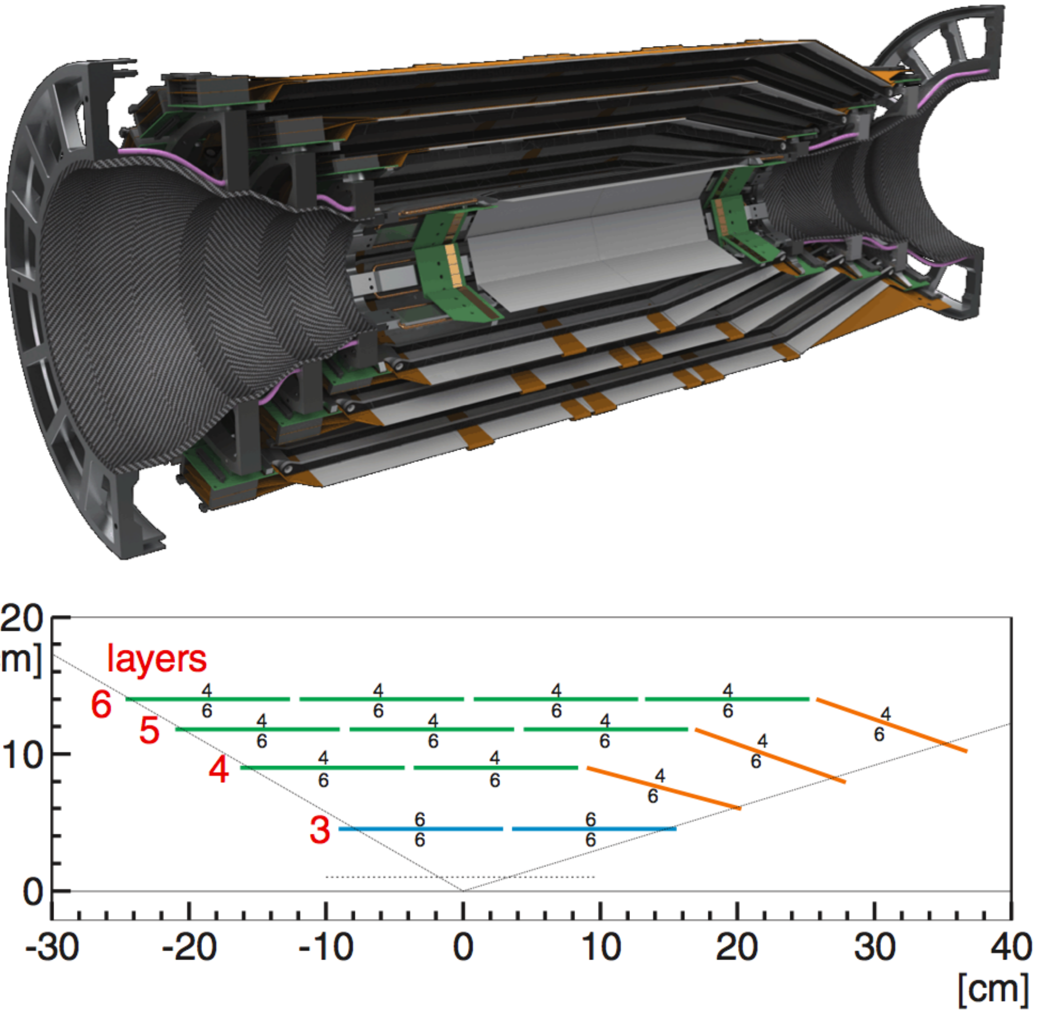


Figure 2.13: Longitudinal view of the Belle II SVD. The four layers of sensors can be seen. In the bottom figure, azure lines correspond to the small rectangular sensors of layer 3, the green ones are the large rectangular sensors of the outer layers, and the orange ones are the slanted FW sensors.

dissipate the heat generated by the electronics, that needs to be positioned in the active area too (more details are in Section 3.1.3). To ensure a good resolution it must be mechanically stable [25].

The SVD is constituted by 4 layers of sensors. Each layer is made by multiple ladders, and each ladder contains several sensors (the exact numbers can be found in Table 2.4). A photo of a ladder can be see in Figure 2.14.

Each sensor is a DSSD, where, similarly to *BABAR* SVT, the strips are perpendicularly oriented on the two sides of the sensors to provide 2D position information. Differently from *BABAR*, the backward (BW) sensors are not trapezoidal: only the slanted forward (FW) sensors of the three outermost layer are trapezoidal, all the other ones are rectangular (*barrel* sensors); layer 3 does not have FW sensors. The barrel sensors have not all the same characteristics: the ones of layer 3 are smaller and have an higher number of strips. The FW sensors have a trapezoidal shape to reduce the amount of material a forward particle has to pass through. More information on the SVD sensors can be found in Section 3.1.3.

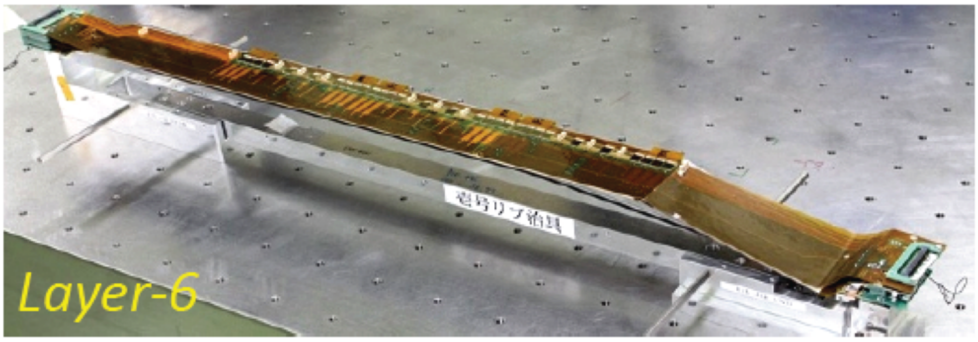


Figure 2.14: A ladder of the layer 6. It is composed of four rectangular barrel sensors and one slanted FW sensor. Image from [26].

All sensors, on each side, host a number of strips that is an integer multiple of 128, and every block of 128 strips is read by an APV25 chip [27], originally designed for CMS.
The APV25 chip is made by 128 channels of low-noise preamplifiers followed by a shaper stage with a peaking time of 50 ns. More information concerning the resulting waveform and the overall readout electronics are provided in Section 3.1.2.

The sensors are slightly tilted with respect to the radial direction, to compensate for the Lorentz drifts that bend the direction of electrons and holes generated by the passage of a particle. It is also present a small overlap between the edges of the sensors, to ensure total coverage in ϕ angle.

Layer	**Radius** (mm)	**Slant angle** °	**#Ladders**	**Sensors/Ladder** S	L	T	**APV25/Sensor**
3	39	0	7	2	0	0	12
4	80	11.9	10	0	2	1	10
5	104	17.2	12	0	3	1	10
6	135	21.1	16	0	4	1	10

Table 2.4: Geometrical parameters of the SVD. S indicates the small sensors of layer 3, L indicates the barrel sensors of the outer layers, F indicates the FW sensors. The numbering scheme treats the PXD and the SVD as an only object, hence the layer numbers 1 and 2 are taken by the two PXD layers.

2.4.3 Central Drift Chamber (CDC)

The Belle II CDC (see Figure 2.15) has to satisfy three important requirements: reconstruct tracks of charged particles and precisely measure their momenta; provide PID information using measurements of energy loss dE/dx within its volume; provide efficient and reliable trigger signals for charged particles.

The Belle II CDC is based on square cells of wires immersed in a helium-ethane gas mixture and an alternate configuration of differently oriented superlayers, each constituted by a certain number of layers.

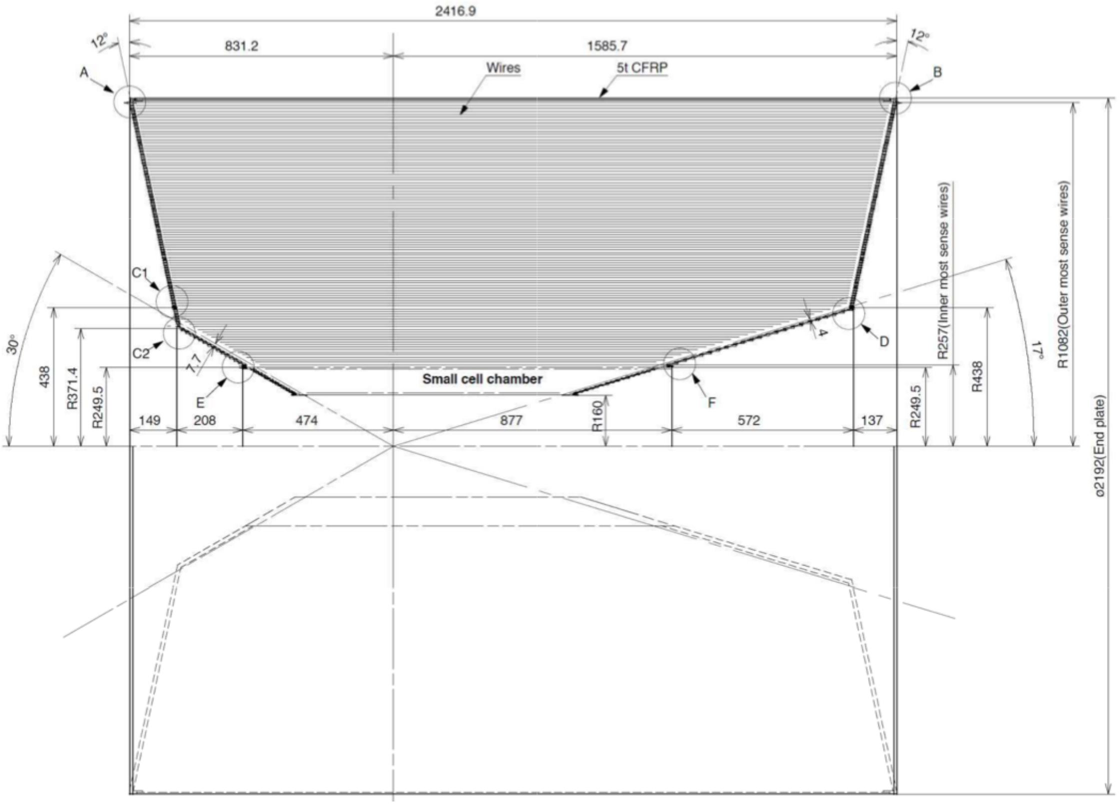


Figure 2.15: Scheme of the the Belle II CDC. Measures are in mm.

The radial cell size is about 18.2 mm for all the superlayers except the innermost, where it is 10 mm. Each layer is constituted by a number of cells that is an integer multiple of 32 because of the requirements of the readout electronics.
Numerical details about the configuration of the Belle II CDC, as well as a comparison with Belle CDC, can be found in Table 2.5, while information about the characteristic of the wires are in Table 2.6.

Quantity	Belle	Belle II
Radius of inner cylinder (mm)	77	160
Radius of outer cylinder (mm)	880	1130
Radius of innermost wire (mm)	88	168
Radius of outermost wire (mm)	863	1111.4
Number of layers	50	56
Number of sense wires	8400	14336
Gas	He-C_2H_6	He-C_2H_6

Table 2.5: Comparison between characteristics of Belle and Belle II CDC.

There are three main differences between the Belle CDC and the Belle II CDC: the new detector has to face a higher luminosity, so a higher trigger rate, and needs to do that with less deadtime; to avoid the high background close to the IP, the inner radius is extended from 77 to 160 mm, but also the external radius is augmented from 880 to 1130 mm; finally, a z trigger information for charged particles is provided, and this 3D information, provided with

Quantity	**Sense**	**Field**
Material	Tungsten	Aluminum
Plating	Gold	none
Diameter (μm)	30	126
Tension (g)	50	80
Number of wires	14336	42240

Table 2.6: Characteristics of the Belle II CDC wires.

the combined usage of axial and stereo wires and charge division between the two ends of the wires, helps in the reduction of background without losing any physical-relevant event.

The energy loss $\mathrm{d}E/\mathrm{d}x$ resolution has been found to be of 12% or better, depending on the incident angle. Spatial resolution is similar or better than the one of Belle, so $\lesssim 130\ \mu$m.

2.4.4 Particle Identification (PID)

Also the PID system has to be upgraded to cope with the higher background rate of Belle II with respect to Belle.
Being able to improve separation power between K and π while, at the same time, reducing the amount of material a particle has to pass through before reaching the electromagnetic calorimeter, is the goal leading the improvements of the PID systems, that in Belle II is divided into 2 different subdetectors: the TOP in the barrel region and the ARICH in the end cap zone.

2.4.4.1 Time of Propagation (TOP) counter

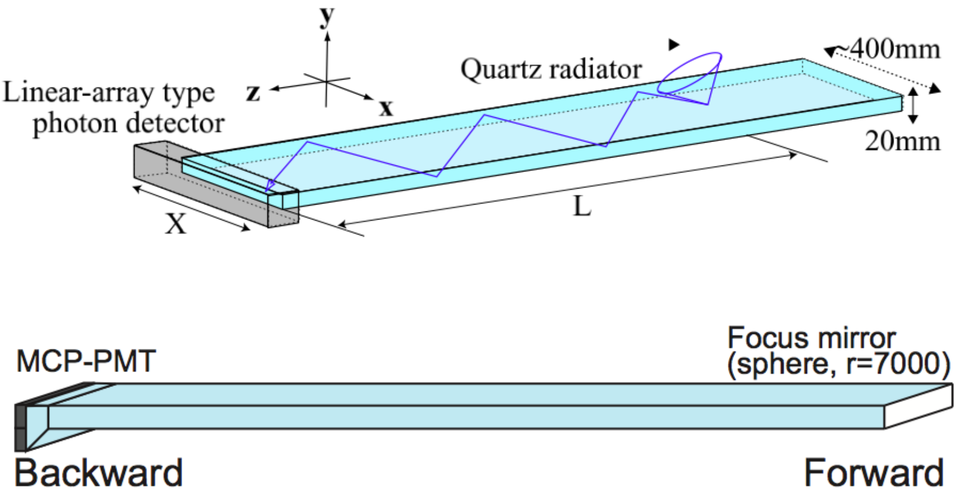


Figure 2.16: Conceptual sketches of the Belle II TOP.

The concept behind the Belle II TOP (see Figure 2.16) is Cherenkov radiation, as well as it was for *BABAR* DIRC. Exploiting the total internal reflection inside the quartz bar, the emitted photons are collected by PTMs collocated in the backward end of the detector, that can be seen in Figure 2.16, where it is also indicated the focusing spherical mirror that is used to focus toward the backward-located readout the photons that have been emitted toward the forward region. The Cherenkov image is reconstructed and the time of propagation of the photons is measured, obtaining so a 3D information: time, x, and y.

The TOP is located immediately outside of the CDC, and it is constituted of 16 modules covering the barrel sector. Each module is made by multiple components: the quartz bar, where particles pass and emit photons; the expansion prism between the PMTs and the quartz bar, dedicated to clarify the ring image and improve the discrimination of wavelengths; the PMTs to detect the photons; the spherical mirror.

Time resolution tests performed on a prototype with a 2 GeV/c electron beam determined a time resolution of $\sim$ 77 ps, a result compatible with the simulations. The time information is combined with the reconstructed Cherenkov ring, i.e. the spatial information, to provide a likelihood through an multivariate analysis (MVA) or an analytic PDF.
The gain in the PMT is order of 10^6 and is stable even with an applied magnetic field of 1.5 T.

2.4.4.2 Aerogel Ring-Imaging Cherenkov (ARICH)

The Belle II ARICH detector (see Figure 2.17) also is based on Cherenkov radiation emission. The ARICH, located in the forward region, is constituted by the following elements: an aerogel radiator where the radiation is emitted; an expansion volume that let the emitted photons diverge to form a distinguishable ring; arrays of photon detectors, capable of detecting single photons in a 1.5 T magnetic field; readout system for the photon detectors.

The ARICH has to be able to discriminate between pions and kaons up to 4 GeV/c and discriminate electrons up to 1 GeV/c.

To achieve the desired precision it is necessary that enough photons ($\sim$ 10) are collected, and they have to be spatially separated: these conditions imply that the radiator region has to be thick enough to allow the emission of enough photons, and that the expansion volume too has to be wide enough to let the photons separate. The needed dimensions are approximately 2 cm and 20 cm respectively, as can be seen on the left of Figure 2.17.

Tests have been performed to choose among two possible configurations of the radiation region: homogeneous or inhomogeneous aerogel. While a longer radiation region provides a higher number of photons and hence a higher precision in the reconstruction of the Cherenkov ring, on the other hand the resolution degrades due to the increased uncertainty in the emission position. The inhomogeneous configuration produces two different rings with slightly different emission angles: the effect is equivalent to focus the photons inside the radiator. This can be achieved adequately tuning the refractive indexes.
The inhomogeneous configuration proved itself having a better angular resolution on the single photon: $\sigma_\theta^{inhomo} \simeq 14.3$ mrad versus $\sigma_\theta^{homo} \simeq 20.7$ mrad, while maintaining the same number of photons in both cases.

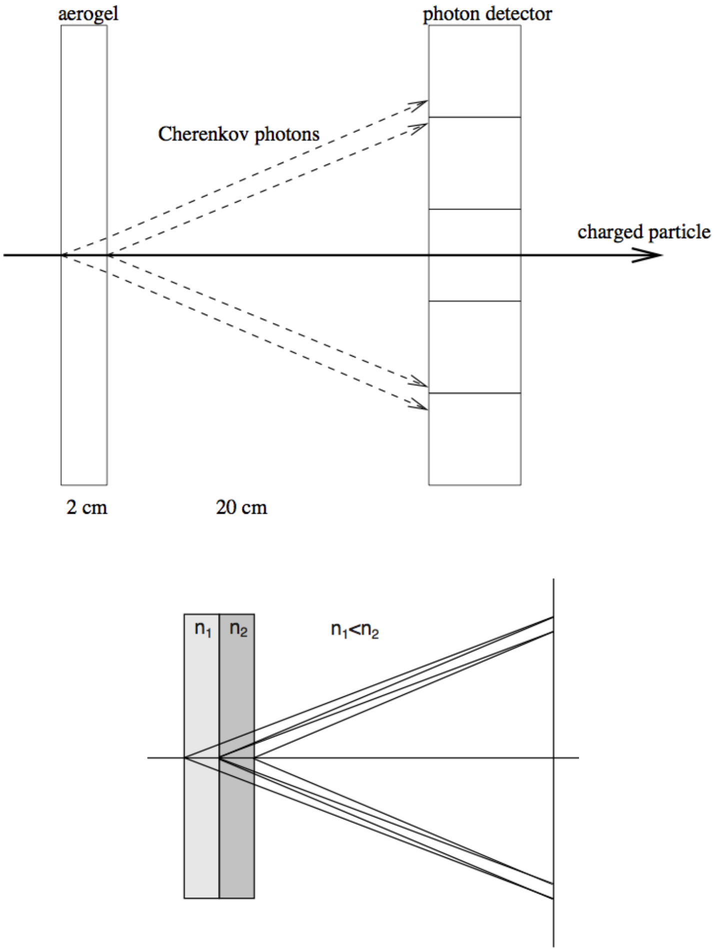


Figure 2.17: Conceptual sketches of the Belle II ARICH. The left image shows the general working principle; the right image shows the inhomogeneous radiator solution that is discussed in the text.

The aerogel radiator region is made with 300 hexagonal pieces for each layer. The readout is composed by $\sim$ 80000 channels; the detectors are organized in concentric rings around the beam pipe.

Refracting indexes of the two radiator layers are 1.054 and 1.065 and both are 20 mm thick. The expected average number of detected Cherenkov photons from both layers is 20. With this number of photons ($N = 20$) the measurement precision is $\sigma_\theta^N = \frac{\sigma_\theta}{\sqrt{N}} = 3.1$ mrad. This precision leads, in absence of background, to a separation higher than 5σ for pions vs kaons and of 4σ for electrons (up to 1 GeV/c) versus pions. $\sim$ 0.7 smoothly distributed background hits are expected within a $\pm 3\sigma_\theta$ annulus of the 4 GeV/c pion ring.

2.4.5 Electromagnetic Calorimeter (ECL)

One third of all B decays include π^0 or other neutral particles in the final state, hence a high resolution calorimeter is mandatory in a B factory.
The Belle II calorimeter, shown in Figure 2.18, is based upon the Belle calorimeter, using the same crystals and photodetectors but improving the readout electronics, to cope with the

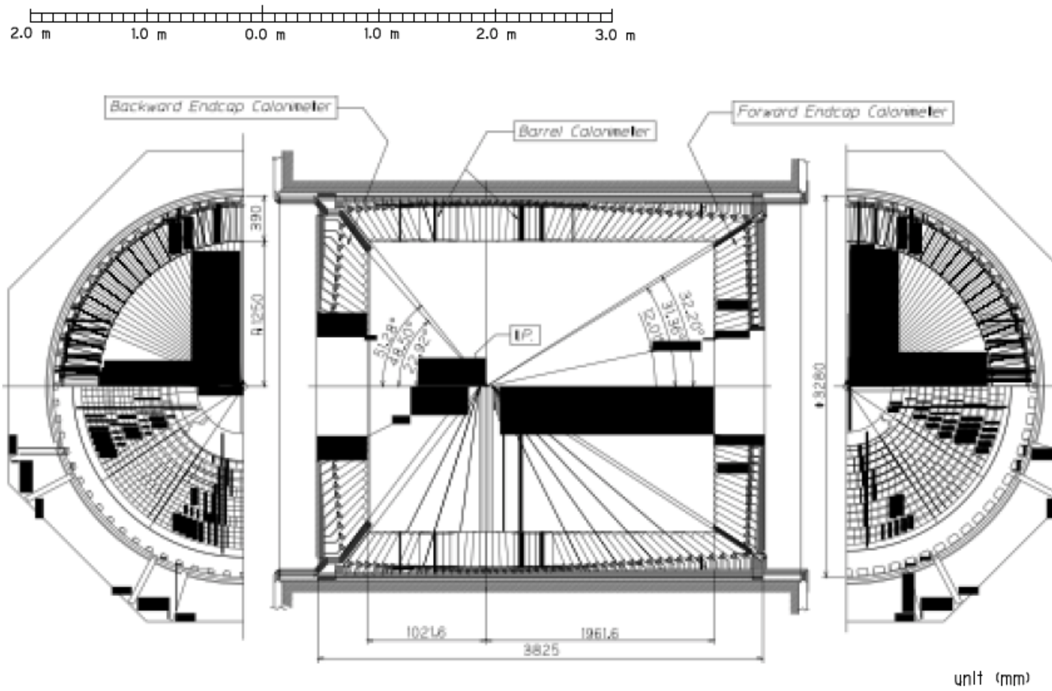


Figure 2.18: Scheme of the Belle II ECL. Measures are in mm.

higher background rate.

The ECL must be able to efficiently measure photon energy between 20 MeV and 4 GeV; it must provide angular coordinates of the detected photons; it has to perform electron identification; it must generate a proper trigger signal; it has to be able to perform measurement of the luminosity, both online and offline; together with the KLM (see Section 2.4.7), it performs the K_L^0 detection.

The ECL is constituted by three main components: the barrel section (3 m long, inner radius of 1.25 m) and the two endcaps, in the forward ($z = 1.96$ m) and in the backward ($z = -1.02$ m) regions. It covers the polar angle between 12.4° and 155.1°, with the exceptions of two gaps approximately 1° wide between the barrel and the endcaps. The barrel region is built with a quasi-projective structure, to avoid leakages of particles between the crystals.
The ECL barrel is built with 6624 CsI(Tl) truncated pyramid crystals, appearing in 29 different shapes; the average surface and length are about 6×6 cm^2 and 30 cm (16.1 X_0) respectively. The endcaps consist of 2112 CsI(Tl) crystals, in 69 shapes. Each crystal is wrapped in a 200 μm Gore-Tex teflon and is covered with two 25 μm layers, one in aluminum and one in mylar. Each crystal is read by 2 photodiodes with a surface of 10×20 mm^2, place at the back of the crystal.

The intrinsic resolution of the calorimeter for a photon of energy E can be approximated as follows:

$$\frac{\sigma_E}{E} \sim \sqrt{\left(\frac{0.066\%}{E}\right)^2 + \left(\frac{0.81\%}{\sqrt[4]{E}}\right)^2 + (1.34\%)^2} \qquad (2.7)$$

where all energies are given in GeV.
The resolution of the Belle calorimeter for the $e^+e^- \to \gamma\gamma$ process, averaged over all the

calorimeter, was $\frac{\sigma_E}{E} = 1.7\%$, and the azimuthal angular resolution was of 0.23°. The $\gamma\gamma$ invariant mass for π^0 and η had a resolution of 4.8 and 12 MeV/c^2 respectively. The light output loss due to the absorbed dose is not of concern, since it has been 7% [13%] in the barrel [endcap] region, in accordance with the estimations. After all the dose foreseen for the Belle II runs, it is estimated to be reduced by a total of 30%.

2.4.6 Detector Solenoid

The 1.5 T magnetic field needed to bend the charged particles, and hence extracting their momenta, is provided by a superconducting solenoid magnet, placed outside the ECL.
The cylindrical volume filled with magnetic field is 3.4 m in diameter and 4.4 m in length.

Most of the hardware is reused from Belle's structure.

2.4.7 K_L & μ Detector (KLM)

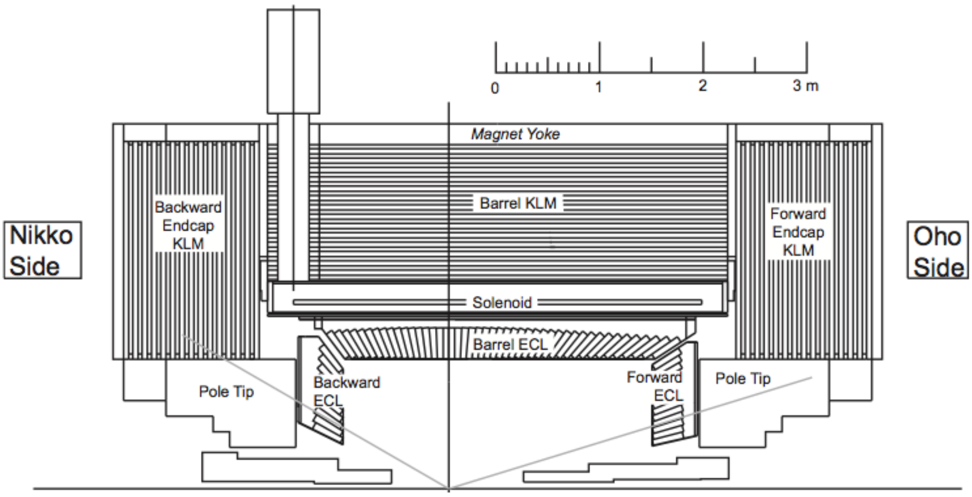


Figure 2.19: Scheme of the the Belle II KLM.

The KLM (see Figure 2.19), the K_L and μ detector, partially exploits the iron needed for the flux return of the magnet, that serve also as 3.9 interaction lengths (or more) of material, in addition to the 0.8 interaction lengths provided by the ECL, and that allow the K_L to hadronically shower. It is split in the barrel and in the endcaps (FW and BW) regions. The barrel region alone covers the polar angle between 45° and 125°; the endcaps extend the acceptance to the angular interval $20° < \theta < 155°$.

In the KLM, interaction material (4.7 cm thick iron plates) and detectors are alternated: 14 iron plates and 14 detectors in each of the three regions, except the barrel one, where there are 15 detectors.
For the barrel region, glass-electrode RPC are used, the same way they were used in Belle. In the endcaps, due to the higher background, the RPC efficiency would be estimated to drop to $\sim$ 50%, thus different detectors are required in their place, i.e. scintillators with a double readout, at both end, performed with two Silicon PhotoMultiplier (SiPM), used instead of the usual PMT because of the limited space and the intense magnetic field. Tests performed with cosmic trigger resulted in a time resolution of $\sigma_t \sim 0.7$ ns, obtained from the

time difference dispersion.

The μ identification method is based on extrapolating a CDC track to the KLM and search for close hits in the KLM itself, and try to match them with the extrapolated track. For what concerns the K_L, clusters in the ECL that can be spatially correlated with the hits in the KLM are searched.

The μ reconstruction efficiency is 89% for muons above 1 GeV/c.
The K_L efficiency rises almost linearly from 0 to 3 GeV/c, from where on it presents a plateau at 80%. Angular resolution for K_L is 3° for KLM-only candidates and 1.5° for KLM+ECL candidates.

2.4.8 Trigger

To deal with the high luminosity expected in Belle II ($\sim 8 \times 10^{35}$ cm^{-2} s^{-1}) a selection of the events is necessary; in particular, Bhabha and diphoton events contribute to the high background.
The requirements of the trigger are the following:

- Maximum average trigger frequency of 30 kHz;
- Trigger latency $\lesssim 5\mu$s;
- Timing precision $\lesssim$ 10 ns.

Also a $\sim$ 100% efficiency for physics processes ($\Upsilon(4S) \to B\bar{B}$ and continuum) is required, as well as a trigger configuration that is flexible and robust.
The trigger follows the criteria used in Belle (with improved technologies), so a hardware Level 1 (L1) Trigger and a software High Level Trigger (HLT).
After appropriate prescaling, we expect a rate of physics event $\mathcal{O}(15\div20$ kHz).

2.4.8.1 L1 Trigger

The L1 Trigger is comprised of various sub-trigger systems, and each summarizes the trigger information coming from its related subdetector. These information are sent to the final-decision logic, made by the Global Reconstruction Logic (GRL) and Global Decision Logic (GDL), the first being an input for the second.

The main technological improvements with respect to Belle are the introduction of high speed serial link cables instead of parallel ones and the usage of Field Programmable Gate Array (FPGA) for each component, so that the trigger logic is configurable rather than hard-wired.

The sub-trigger systems are the following: CDC, ECL, BPID (Barrel PID, so from the TOP subdetector), KLM; the CDC provides charged track information (momentum, position, charge, multiplicity); the ECL provides energy deposit and cluster information, and Bhabha and cosmic ray identification; the BPID provides precise timing and hit topology information; the KLM provides muon track information.

The two main sub-trigger systems are the CDC and the ECL ones. The CDC sub-trigger finds and characterizes the charged tracks detected in the CDC, both with 2D and 3D tracking, the last one being particular useful to reject beam-gas and beam-wall background thanks to the z_0 information.

The ECL sub-trigger uses two schemes: total energy and isolated-cluster counting triggers. The first is sensitive to physics events with high EM energy deposit, while the second is sensitive to multi-hadron events with low-energy clusters and/or minimum ionizing particles (MIPs). The ECL also has to identify Bhabha and $\gamma\gamma$ events, needed to measure luminosity.

2.4.8.2 High Level Trigger (HLT)

The HLT consists of two parts: the Level 3 (L3) trigger and the physics-level trigger.

The L3 trigger performs a fast, rough reconstruction of the event using CDC data only and a quick ECL cluster reconstruction. After that it performs a cut based on track multiplicity, vertex position, and the total energy deposition. Its reduction power is approximately 50%.

The physics-level performs a complete reconstructions of the event, using the same software used in offline reconstruction, and classifies them with criteria similar to the ones used to produce skims (hadronic selection, low-multiplicity selection) and to select a small fraction ($\sim$ 1%) of Bhabha and $\mu\mu$ for monitoring purposes; this trigger has a reduction power of about 25%.

The total reduction power of HLT is therefore $\sim$ 12.5%.

2.4.9 Belle II Software

The Belle II software has to take account of the control of the detector, data acquisition, reconstruction and transfer, database handling, physics analysis, Monte Carlo (MC) simulation etc. Except for the tasks related to the direct hardware interface, Belle II software is handled by **basf2**, i.e. Belle Analysis Software Framework 2, that is the software framework for Belle II. It is used for online (e.g. HLT) as well as for offline data handling. The offline usage of basf2 spans the whole range of applications, from detector optimization to physics analyses.

The main components of basf2 are modules, dataobjects, and scripts. It relies on a database (DB) to retrieve parameters.

- **Modules**: modules are written in C++, and they are the "bricks" of the Belle II software. Their tasks vary from simple ones like reading data from a file to complex tasks like full detector simulation or tracking. In basf2 all work is done in module. They are arranged in so-called *path*, handled by *script* files;

- **Script files**: script files are written in Python, and they handle the execution of modules. In script files the user can call the desired modules, set their parameters, and arrange them in a path, which corresponds to a container where the modules are arranged in a linear order. When processing data, the framework executes the modules of a path, one at a time, starting with the first one and proceeding with the module next to it. The data, to be processed by the modules, is stored in a common storage, the so-called DataStore, to whom each module has reading and writing access;

- **Dataobjects**: dataobjects are classes written in C++, and they correspond to the elementary objects of basf2. Dataobjects are the pieces of code written and read by modules, containing information from the detector or produced by other modules, or from elsewhere. They are stored in the DataStore;
- **DB**: the database contains all sort of constants needed by the experiment. It can be accessed and written by the users, which can upload and retrieve parameters as the constants extracted from calibration procedures, detector geometry, accelerator parameters and so on.

Chapter 3

SVD Modules Overview & Software Development

The Pisa Belle II group has a strong participation in tracking and SVD related activities. For what concerns the latter, the activities revolve around both hardware and software aspects. The SVD-hardware activities center around the assembling and the testing of the forward (FW) and backward (BW) modules (see Section 2.4.2) of layers 4, 5, and 6, performed in the clean room at Istituto Nazionale di Fisica Nucleare (National Institution of Nuclear Physics, INFN) High Technology Laboratories. The software activities are focused on the offline part of the SVD software.

My thesis activity does not only revolve around the ALP search in the $B^{\pm} \to K^{\pm}A$, $A \to \gamma\gamma$ decay, but also concerns the development of the SVD offline software, in particular the construction of a hit time estimator basf2 module, called Centre of Gravity (**CoG**), whose main purpose is the reduction of the off-time background. I also took part at the tests of the SVD sensor modules assembled in Pisa.

This Chapter provides a description of the SVD modules and software related to my thesis activity.

Section 3.1 describes the SVD sensor modules, their working principle, the read-out electronics, and the procedure for the electrical test, while Section 3.2 provides a picture of the status of the SVD software, with particular focus on the offline reconstruction of the events. In Section 3.3 there is a detailed description of my contribution to the offline software, i.e. the CoG module, dedicated to extract the hit time information from the strip data. Performances are shown in Section 3.4.

3.1 SVD Modules Description

In Section 2.4.2 there is an overall description of the SVD. This section focuses on the working principle of the DSSD sensors, the structure of the FW and BW modules, and the readout electronics, providing also information about the electrical tests performed on the modules assembled in Pisa.

3.1.1 Sensor Working Principle

As previously mentioned, Belle II SVD is constituted by DSSD, that are double-sided silicon microstrip sensors. Their task is to provide 2D spatial information on the crossing position of the particles, and to do so they exploit the energy released by the particle interacting with the matter of the sensors themselves. The energy electromagnetically released from the crossing particle can excite the material generating the so-called *e-h-pairs*, i.e electron-hole pairs. The e-h-pairs can travel through the material if an external electric field is applied, and this currents can be detected, providing the desired signal.
In fact, a crossing particle generates a number of e-h-pairs that is negligible in comparison to the intrinsic charge carrier density of silicon. To estimate this number we need the following quantities: the mean energy loss in silicon $\left(\frac{dE}{dx}\right)_{mean}$ for a MIP, the minimum energy E_{pair} needed to create an e-h-pair, and the thickness d_{sensor} of the sensor. With the actual values ($\left(\frac{dE}{dx}\right)_{mean} = 388$ eV μm^{-1} [4], $E_{pair} = 3.63$ eV [28], $d_{sensors} \simeq 300$ μm) we obtain:

$$\frac{\frac{dE}{dx}_{mean} \times d_{sensor}}{E_{pair}} = \frac{388 \times 300}{3.63} \simeq 32 \times 10^3 \text{ e-h-pairs} \tag{3.1}$$

This number is indeed negligible with respect to the number of intrinsic charge carrier: this can be obtained knowing their density n_i at a given temperature and the volume of the material. For silicon at $T = 300$ K we have $n_i = 1.45 \times 10^{10}$ cm^{-3}, and for a sensor with an area of 1 cm^2 and thickness ~ 300 μm we obtain:

$$n_i \times V_{sensor} = 1.45 \times 0.03 \times 1 \times 10^{10} \simeq 44 \times 10^7 \text{ e-h-pairs} \tag{3.2}$$

Hence, the generated e-h-pairs are four order of magnitudes smaller than the intrinsic ones. In order to be able to appreciate the number of generated pairs over the total number it is mandatory to reduce the latter, and the best way to do that is to use a reversely polarized *p*-*n*-junction.

A *p-n-junction* is the contact region between two oppositely doped semiconductors. Here the so-called depletion region develops: in this regions the extra electrons from the *n*-doped side and the extra holes from the *p*-doped side recombine, producing an area deprived of free charges. The thickness of this area can be enhanced applying a reverse bias voltage.
In a double- or single- sided silicon detector this junction is very asymmetrical (i.e. the doping of the two sides is very different) and is created by heavily doping a small layer at the surface (*p*-type) inside a bulk material *n*-doped.
Applying a sufficiently high reverse bias voltage the intrinsic charge carriers are removed from the bulk, so the e-h-pairs generated by the passage of a particle become detectable. The applied voltage does not only deplete the bulk, but also induces the drifting of the generated electrons and holes, as they move along the lines of the electric field associated with the voltage, toward the electrodes. This movement produces a current that induces a signal on the electrodes (thanks to Ramo theorem [29], that models the current signal induced on electrodes by a charge drift). This current signal allows a 1D position measurement, if the electrode is segmented. If the segmentation of the electrode is orthogonally applied to both

sides of the sensor, a 2D position measurement can instead be performed. Figure 3.1 depicts this structure.

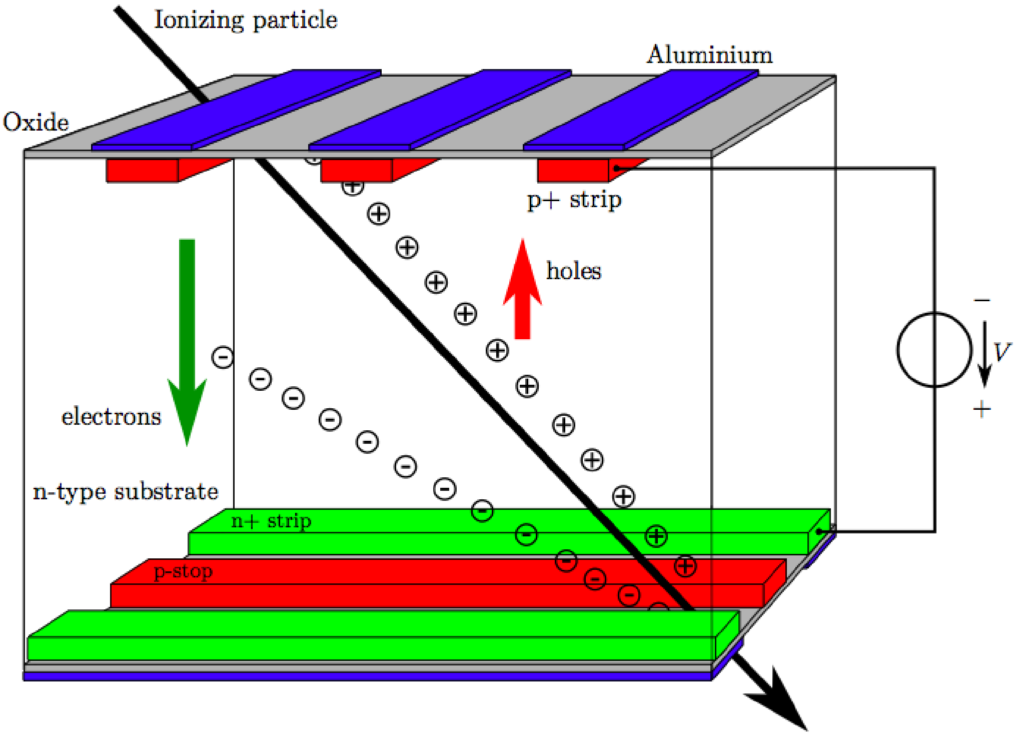


Figure 3.1: Working principle and scheme of a DSSD. Image from [30].

The electrode is divided in segments called *strips* (with a width order of some tens of μm), and the distance between two of them is called *pitch* (50 to 300 μm, see Table 3.1 for precise values for Belle II). Each implanted strip is read by an aluminum strip that is electrically isolated with an oxide (SiO_2) layer to avoid direct current leakage because, as mentioned, the signal is constituted by the effect of the moving charges rather then the charge accumulating on the metal strip. This signal is read by an APV25 chip (see Section 3.1.2 for more information), that is very sensitive to the DC fraction of the current, so this kind of insulating coupling is necessary.

The oxide layer intrinsically contains fixed positive charges that can attract electrons from the n-type bulk, resulting in an electron accumulation layer at the Si-SiO_2 surface that could connect two n-type strips. Since this layer cannot be depleted even with a very high reverse bias voltage applied, a p-stop implanted strip is collocated between two adjacent n-type strips, to interrupt this electronic layer connection and ensure a good isolation between the strips on the n-side.
In Figure 3.1 a p-stop strip can be seen.

To identify the strips interested by the passage of a particle, a selection based on the collected signal is applied: the signal has to be above a certain threshold, defined with a signal-to-noise ratio (SNR), the noise being the one typical of the strip under examination.
It is also possible that several adjacent strips are activated by the same particle, and in this

case the released charge is distributed among them according to the position of the crossing.

3.1.2 Readout Electronics

As seen in Sections 3.1.1 and 2.4.2, a crossing particle induces a signal in the APV25 readout channels. The APV25 chip is a low noise front-end amplifier, whose noise is directly proportional to the capacitance seen at the input by the channel. The components of main interest of the APV25 channel are the following:

- **Low noise charge pre-amplifier**: picks the raw current signal of the strip and transforms it into a voltage usable from the subsequent stages;
- **Shaper**: an analog CR-RC shaper with 50 ns shaping time;
- **Analog pipeline**: the shaper output is sampled at the clock frequency, and then written on the ring buffer;

The output of the shapers is a curve with a steep rise and a slow decrease. An example can be seen in Figure 3.2.

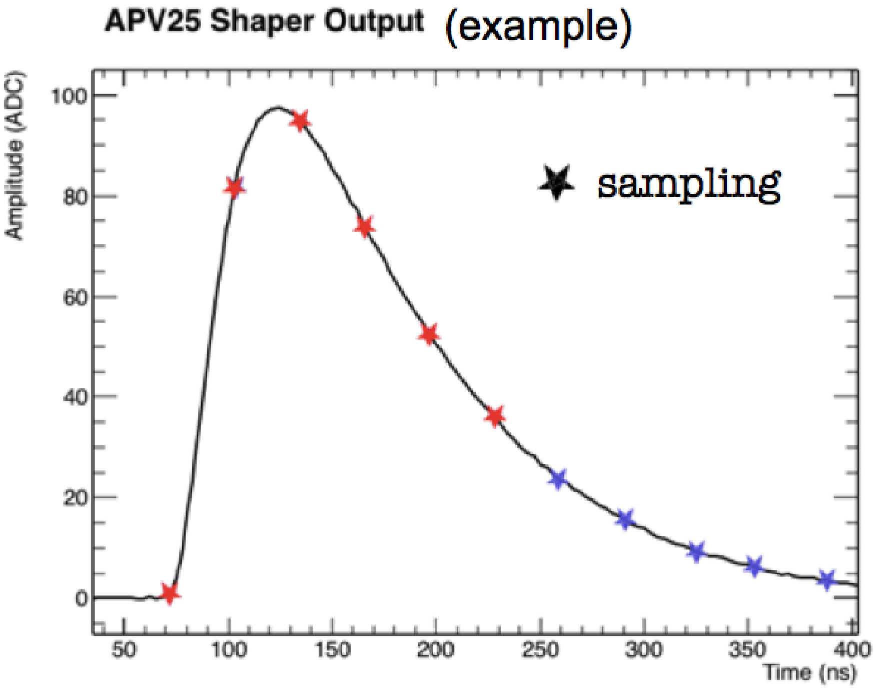


Figure 3.2: Example of the output of the shaper of the APV25, taken from a calibration. The stars indicate the time position of the APV25 sampling clock. The 6 red stars are an example of a good sampling (in 6-samples mode). Image from [31].

This waveform is sampled at the clock frequency (31.8 MHz, a frequency equal to 1/16 of the radiofrequency SuperKEKB works at), and the samples are recorded in the ring-buffer. The ring-buffer hosts 192 cells, with 32 indexes to label the cells pending to be read, so the total number of available cells able to be written is 160. Only the samples close to the peak of the curve are relevant to reconstruct the curve itself, and so to retrieve charge and time information. It is therefore necessary to decide how many samples to read and where to start reading them.

The number of read samples can vary: the current default is 6, but in the future, in a high-luminosity regime, we expect to reduce it to 3 to reduce the amount of data to be transferred and so reduce the dead time of the chip associated with very high trigger rate. Given the clock frequency, the time between two samples is $\Delta T \simeq 31.4$ ns, so in the 6-samples regime the waveform is sampled for ~ 150 ns. In Figure 3.2 the 6 recorded samples (in an ideal case) are marked with red stars.

The information about the the timing of the event is given by the Trigger (see Section 2.4.8), that is is provided by the outer subdetectors. Since it is delayed with respect to the physical event in the SVD, it is necessary to know this delay to be able to come back of the right number of cells in the ring buffer and retrieve the relevant samples.
The knowledge of this delay, known as *latency*, derives from calibrations performed with cosmic rays. Currently the latency is set at 158 cells, so ~ 4.97 μs; the latency has to be smaller than 160×31.4 ns $\simeq 5030$ ns $= 5.03$ μs, otherwise cells are overwritten and no data can be saved and retrieved.
When the latency L is known (in ns) and a Trigger arrives at a time corresponding to the cell number N, it is sufficient to go back to the cell number $N - \frac{L}{31.4}$ and look at that sample and the following 5 (or 2, depending on the recording mode), that will be recorded and analyzed by the SVD reconstruction software (see Section 3.2).
A graphical example of this mechanism can be seen in Figure 3.3.

The output of the APV25 is analog; before being passed to the SVD reconstruction software it is digitized and processed by the FADC (Flash Analog-to-Digital Converter), that performs the analog-to-digital conversion and low-level data processing, as noise subtraction and zero suppression (i.e. the elimination of strips whose SNR is below threshold).
In the end, each of the 6 sample amplitude is converted into a 8-bit integer (so going from 0 to 255); these are the amplitudes used by the SVD reconstruction software. More details about the read-out chain of the SVD can be found in [32].

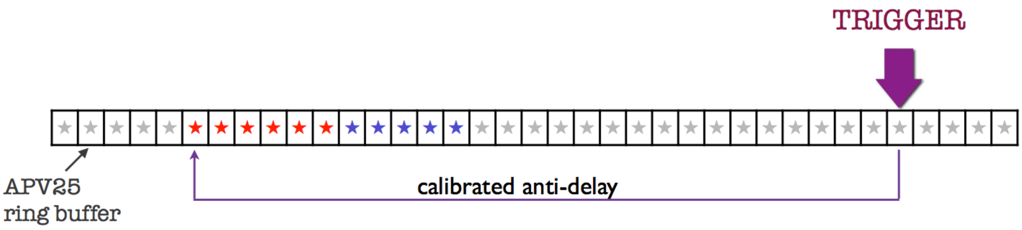


Figure 3.3: Ring buffer of the APV25, showing the arrival of the Trigger, the latency (or *anti-delay*) and the read 6 samples (in red). Image from [31].

The Trigger can be imprecise, in the sense that does not provide the correct timing, so going back in the ring buffer by the set latency would mean to register samples that are not adequately related to the event. This uncertainty is called Trigger Jitter (TJ). The TJ is estimated to be order of $\sim 20 \div 40$ ns at the early stage of data taking and should be reduced down to ~ 10 ns when the detector will be fully operative.

The clock frequency of the FADC, appointed to save the Trigger information, is 4 times higher than the frequency of the sampling clock of the APV25. We also record the information about the arrival time of the Trigger, so we have an additional information about the real time of the event, consisting on which fourth of the sampling window of 31.4 ns we can expect to actually have the start of the event, assuming that the Trigger works well (i.e. low TJ). This information is stored in a variable called Trigger-Bin (TB). Figure 3.4 shows this structure.

The APV25 chips are mounted on the *hybrid boards*, that are electronic boards which house the readout chips, supply power to them, and route their output to the subsequent data cables.

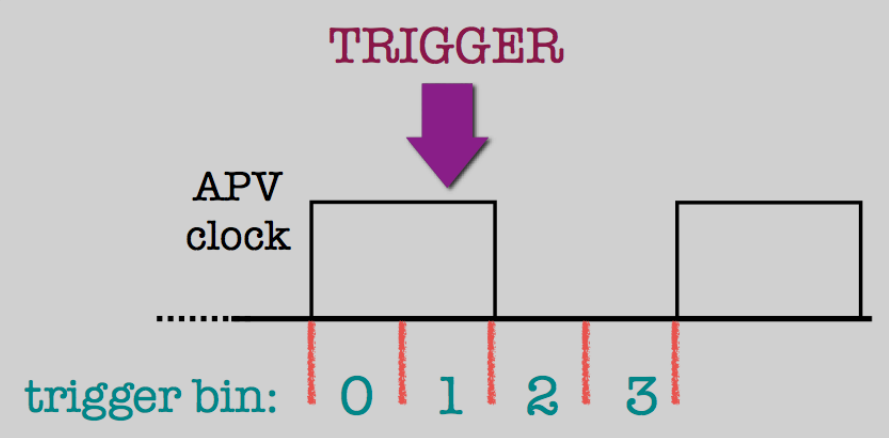


Figure 3.4: Arrival of the Trigger and TB information. The clock of the FADC, that saves the Trigger arrival information, has a frequency that is 4 times higher than the frequency of the APV25 clock, so it is possible to store the information of which of the four fourth of the APV25 time window the Trigger has arrived.

3.1.3 Forward and Backward Modules

The fast shaping time of the APV25 implies great susceptibility to noise, that is mainly influenced by the load capacitance seen in input by the chip, therefore the APV25 has to be mounted as close as possible to the sensor strips, to reduce the capacitance due to long connections.

This necessity has brought to two different solutions for the modules at the end of the acceptance region and for the internal ones. Figure 3.5 shows a ladder scheme with the two different approaches, that are described below.

Since the BW and FW modules (modules are also known as *sub-assemblies*) are located at the edge of the acceptance region, the readout electronic can be placed externally, outside the sensitive volume, as shown in Figure 3.5.

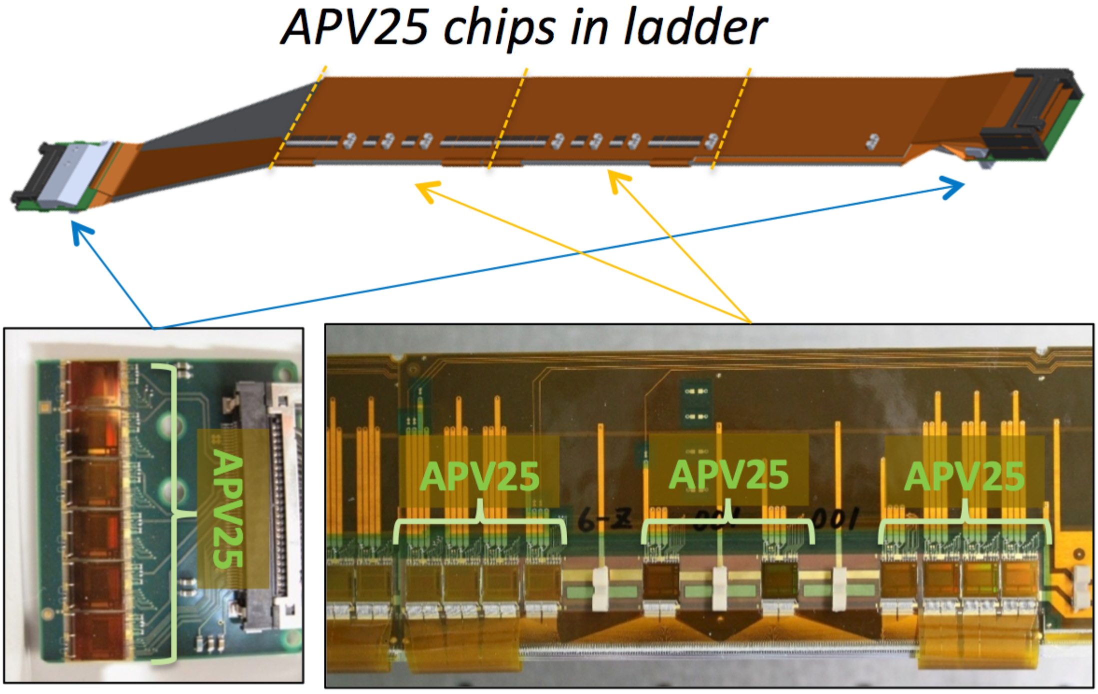


Figure 3.5: Position of the APV25 chips on the different modules of a given ladder. The FW and BW modules have the readout electronic located outside the active regions (azure arrows); the internal modules have the readout electronic located on the sensor itself, and are characterized by the Origami-on-chip solution (yellow arrows). Image from [26].

The FW and BW modules have a different read-out layout than the internal modules, but they are similar to each other, the only difference being the form of the sensors, as can be seen in Figure 3.6.

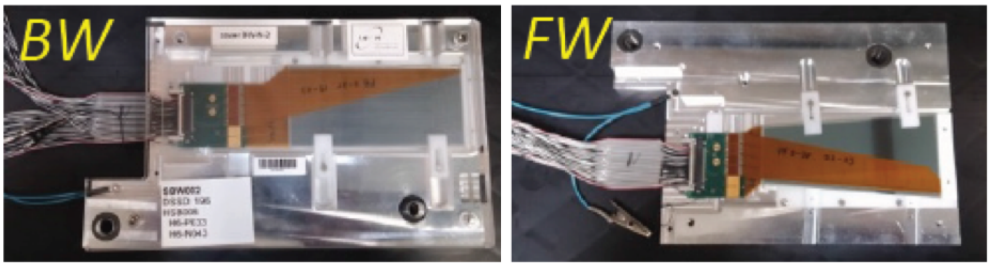


Figure 3.6: SVD BW and FW modules. As can be seen from the photos, the readout electronic is located outside the sensitive region of the sensor. Image from [33].

Each FW and BW module is composed of three main components:

- **DSSD**: its working principle has been described in Section 3.1.1. As mentioned, the two kind of sensors have a different shape.

  The sensors are mounted, in the frame of the SVD, such that the p-side is internal (i.e. it faces the IP) and measures the $r\phi$ coordinate; the n-side, external, measures the z coordinate.

- **Hybrid Sandwich (HS)**: it is composed of two hybrid boards, each of them housing a number of APV25 chips equal to 1/128 the number of strips that that side has to read, so 6 APV25 for the p-side and 4 APV25 for the n-side. The two boards are glued together to form the HS; the signal is brought from the sensor to the chips through the Pitch Adapter.

- **Pitch Adapter (PA)**: it is the flexible circuit that connects the APV25 chips to the sensor through copper lines. There are different PA shapes and dimensions, depending on the type of module and on the side. The PA is made of a kapton layer $\sim 80\ \mu$m thick.

The schemes and the geometrical and electrical characteristics of the sensors can be found in Figures 3.7 and in Table 3.1.

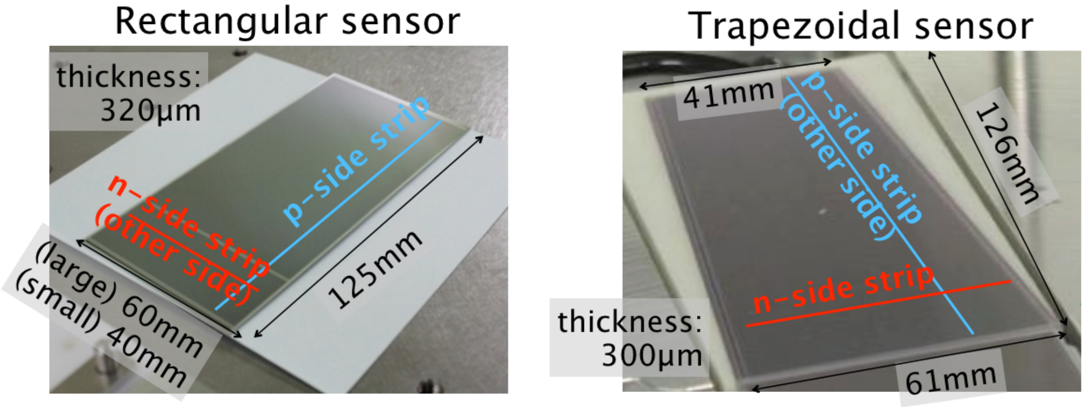
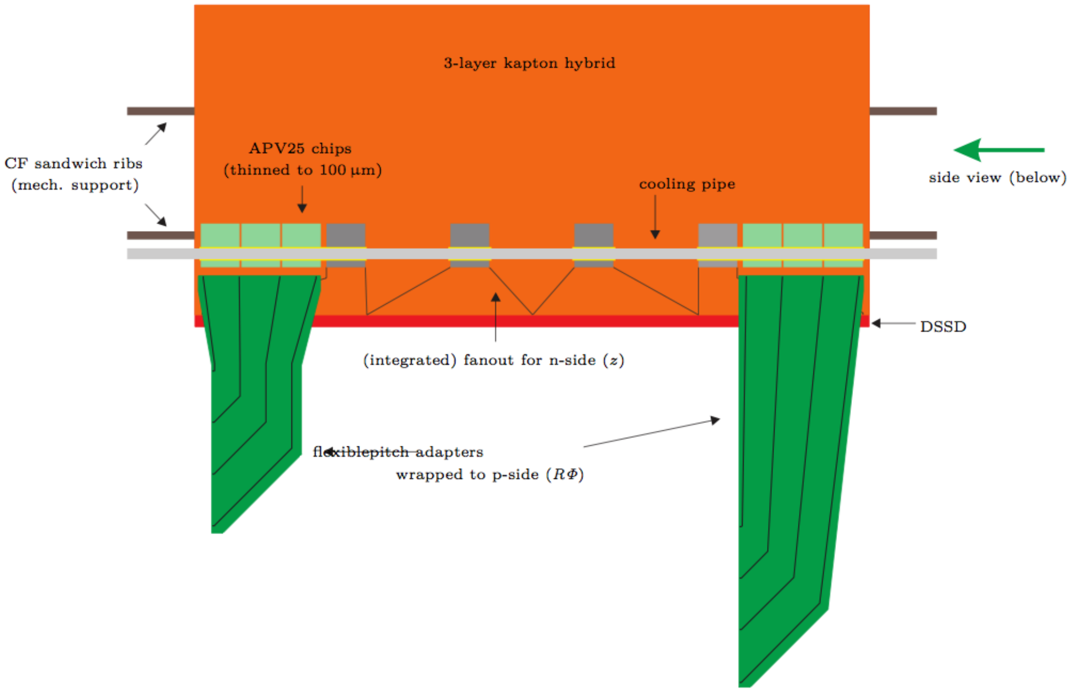


Figure 3.7: Photos of the SVD sensors. Adapted from [26].

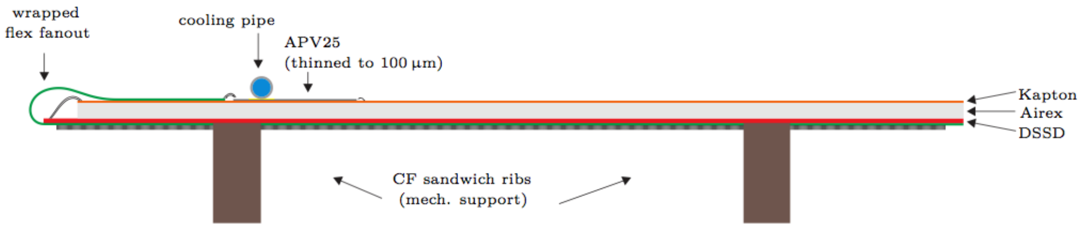
Quantity	**Large rectangular Barrel & BW**	**Small rectangular Layer 3**	**Trapezoidal FW**
# strips p-side	768	768	768
# strips n-side	512	768	512
Readout pitch p-side	75 μm	50 μm	50 to 75 μm
Readout pitch n-side	240 μm	160 μm	240 μm

Table 3.1: Geometrical parameters of the SVD sensors.

The *Origami chip-on-sensor* design has been developed for the inner sensors: the chips are directly located on the sensor, mounted on a flexible circuit "Origami flex" glued on the sensor with an electrical and thermal insulator foam in between. All of the ten chips are mounted on the n-side, as shown in Figure 3.8 and on the bottom right part of Figure 3.5. The long p-side strips are connected by small, flexible pitch adapter wrapped around the edge of the sensor (this is the origin of the name of this scheme). The short n-side strips are connected by a pitch adapter integrated into the hybrid itself.
All APV25 chips are arranged in a row, hence all the chips can be cooled at once using one single cooling pipe, thus minimizing the amount of material for cooling in the active region.



(a) Top view



(b) Side view

Figure 3.8: The Origami chip-on-sensor concept. (a) Top view: The six APV25 chips reading out the strips on the opposite side are displayed in pale green, the flexible double-layer pitch adapters to be wrapped around the sensor are unfolded. The four APV25 chips reading out the top side are shown in grey, with the integrated pitch adapter next to them. (b) Side view: The flexible pitch adapter connecting to the bottom side strips is located at the left side. Image from [30].

3.1.4 Electrical Tests

During my thesis I have participated to the electrical tests performed on HS and modules. These tests are done to check the quality of the components, after the assembling and before the shipping to the ladder mounting sites, to identify and possibly fix any new defect possibly introduced during the assembling, with respect to the known ones detected during the tests performed at the sensors production site.

A dedicated testing station, known as *test stand*, has been prepared in the clean rooms in Pisa. It is composed of:

- APVDAQ data acquisition system and its software (developed at HEPHY);
- Low voltage to supply power to APV25 chips;
- High voltage (up to 200 V) to supply the reverse bias voltage to the sensors;

- Laser scan device, composed of an infrared laser emitter to simulate crossing particles, and a motor to move it over the sensor and test different areas of it;
- Metal testing box, where the modules are put inside to test them, shielded from external electromagnetic fields and light.

Two computers accompany the test stand, one dedicated to measure the temperature of the testing box, one to perform data acquisition through a LabView software interface. Every test run is preceded by a test on a so-called *Origami test hybrid*, a well-known and many-times tested couple of hybrids used as reference to make sure that no bias or error is due to test stand issues.

The following tests are performed on hybrids, HS and modules:

- **Passive measurement**: resistance and capacitance of the HS are measured and compared to the standard one;
- **Internal calibration run**: to measure the gain of each preamplifier channel. A known charge pulse is injected into each chip channel, via a calibration capacitance, to simulate the signal response to the passage of a particle (there is no need for the presence of the sensor), and the response is measured, registered and fitted, to extract the height and the time of the peak;
- **Pedestal run**: to measure the pedestal and the noise of each strip of the sensor. During this test the response of each readout channel is measured with no signal injection in the calibration capacitance, and applying the standards bias voltage $V_{bias} = 100$ [200] V for the FW [BW] modules. The mean signal level (the pedestal) and its variance (the noise) are recorded for each strip.

The following tests are performed only on assembled modules, since they require the presence of the sensor:

- **I-V curve**: varying the V_{bias} between 0 V and the standard value (100 V for FW, 200 V for BW), the total leakage current is measured and the curve is saved and compared with the one recorded at the sensor production site;
- **V_{sep} scan**: for each strip the gain is measured, applying a separation voltage V_{sep} between the chip power supply and the sensor reference voltage (i.e. between the preamplifier and the strip, separated by the decoupling capacitance constituted by the SiO_2 insulant layer); V_{sep} varies between -5 V and +5 V. In case the oxide is broken, the current flowing across it due to the V_{sep} saturates the preamplifier, resulting in a low gain and hence allowing the detection of such defects;
- **Laser scan**: using the laser mounted on the motor, the entire sensor is scanned with an infrared laser pulse, moving in the two perpendicular directions to examine both sides; the hits for each strip are counted to search for any anomaly (both excess or deficit of counting).

The data are analyzed in order to identify the following defects:

- **Open**: a broken metal strip or bond. The defect is characterized by high noise, anomalous peak value and peak time, lower number of hits in the laser scan;
- **Short**: a contact between two (or more) adjacent metal strips; the collected charge is shared among them and the outputs show lower gains;
- **Pinhole**: a break in the SiO_2 insulator between the aluminum strip and the strip implanted in the bulk. This produces a current between the strip and the preamplifier; it is characterized by the fact that the strip has a low gain at $V_{sep} = -5$ V and $V_{sep} = 5$ V, but the usual gain is restored at $V_{sep} \simeq -0.75$ V, that compensates the voltage difference between the preamplifier input and the sensor strip, preventing the current from flowing between the two of them and hence saturating the preamplifier.

3.2 SVD Software

The SVD software is divided into two main areas: *online* and *offline*. The SVD online software is dedicated to run, configure, and monitor the SVD; it is beyond the purposes of this thesis to go into further details, as it is not related to my activities. For further information, refer to the TDR [22].

The main purposes of the SVD offline software are to simulate the response of the detector to charged particles and to reconstruct the corresponding hit. The latter is explained in detail in Section 3.2.1.

The simulation software drifts the e-h-pairs, generated by the passage of a charged particle in the sensor, toward the electrodes, and includes the diffusion effects. It then simulates the strip signal generation, taking into account also charge sharing between adjacent strips, and the corresponding APV25 signal. It finally samples the output of the APV25 shaper and digitizes it.

3.2.1 SVD Reconstruction Software

The SVD offline reconstruction software purpose is the reconstruction of the raw data coming from the detector, to build the space points needed to reconstruct the trajectories of charged particles. The raw data from the detector are processed to determine the charge and the hit time for each strip. Adjacent strips are united in clusters, and clusters from the opposite sides are combined to form the space points used in tracking.

Figure 3.9 shows the workflow of the reconstruction software; the steps are explained below.

Raw data come from the detector in the form of a RawSVD dataobject: it contains low-level information as the channel number of the activated strip and the amplitude of each of the APV25 shaper output samples, as well as the number of read samples. This object needs to be elaborated in a more user-friendly form, and this is done by the SVDUnpacker module, which produces SVDShaperDigit as output. The position of the strip under examination is

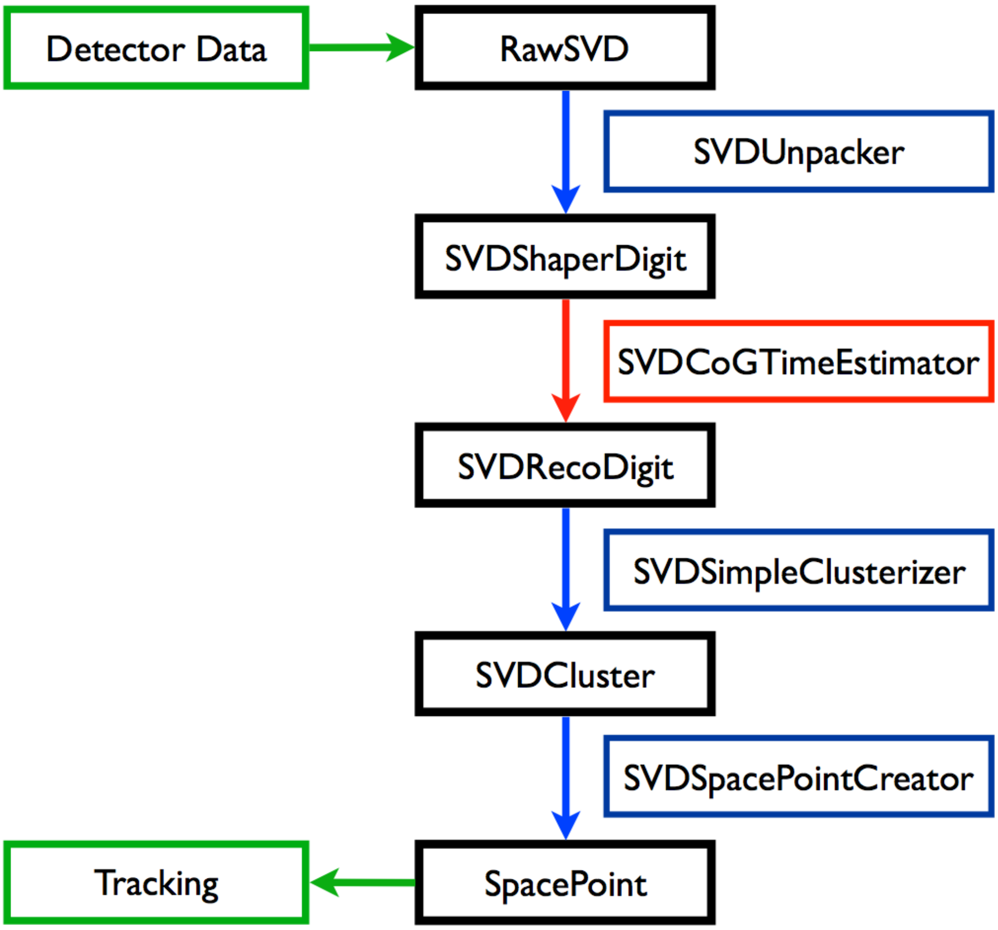


Figure 3.9: SVD offline reconstruction workflow. The image shows the steps needed to pass from the raw information coming from the detector to the so-called SpacePoints, the objects used by the tracking. In the leftmost column there are the input and output objects; in the middle column there are the dataobjects that are elaborated from the software, while on the rightmost column there are the basf2 modules performing these elaborations. The red step corresponds to the module I wrote, SVDCoGTimeEstimator, that takes the SVDShaperDigit as input and returns the SVDRecoDigit as output.

now given by the index of the layer, ladder, sensor and side, while the sampled amplitudes are stored as 8-bit integer numbers. The RawSVD also contains diagnostic information.

SVDShaperDigits are processed by my module, the **SVDCoGTimeEstimator**, that produces SVDRecoDigits as output: it preserves all the information about the position of the analyzed strip, and processes the amplitudes of the samples to obtain an estimation of the hit time and strip charge; the details of how this is done are explained in Section 3.3.

Adjacent activated strips, as mentioned at the end of Section 3.1.1, can be produced by one single crossing particle, and it is the duty of the SVDSimpleClusterizer to merge them together into a single dataobject called SVDCluster. To prevent the inclusion of noisy and/or background strips in the cluster, some constraints on their SNR are applied (there must be at least one seed strips with SNR ≥ 5, and each strip must have SNR ≥ 3). Both the position and the time of the cluster are computed with weighted means, the weights being the charge of the strips composing the cluster:

$$X_{\text{cluster}} = \frac{\sum_i X_i \cdot A_i}{\sum_i A_i} \tag{3.3}$$

$$T_{\text{cluster}} = \frac{\sum_i T_i \cdot A_i}{\sum_i A_i} \tag{3.4}$$

where X_i, T_i, and A_i are the position, the time, and the charge of the i-th strip.
The above definitions stand true until a certain cluster size limit, beyond which a head-tail algorithm [34] is applied to take care of the wide SVDClusters.

The 2D on-sensor position of a crossing particle is obtained combining the 1D clusters from the two opposite sides of the sensor: this is done by the SVDSpacePointCreator, that combines the SVDClusters to produce the SpacePoints, used by the track finding algorithm to derive the trajectories of the particles.
One of the main problems that the SVDSpacePointCreator has to face is the presence of *ghost hits*, i.e. combinations of perpendicular activated strips where at least one does not correspond to any physical particle. An application of the time estimation provided by the SVDCoGTimeEstimator is to reduce the ghost hits composed also by background, eliminating combinations whose timing is incompatible. SVDSpacePointCreator provides the 3D position in the global coordinate system.

The SVD also contributes to PID thanks to measurement of the energy loss $\frac{dE}{dx}$ in the silicon, but going into the specifics for this is beyond the purposes of this thesis.

3.3 Hit Time Estimator

As earlier mentioned, the main purpose of extracting a reliable hit-time information from the signal data is to reduce off-time background.

The importance of this rejection can be roughly evaluated as follows. The expected occupancy from machine background particles on the innermost SVD layer is expected to be 1%: in each event, 1% of the strips of a sensor are active, i.e. their SNR is above threshold. Layer 3 sensors have 768 strips for each side (see Table 3.1), resulting in approximately 8 active strips per side due to background. If the strips are not adjacent, this also implies 8 clusters per side.
Since a $\Upsilon(4S)$ event has ~ 10 tracks per event and the layer 3 has 14 sensors, we can roughly estimate 1 signal-related SpacePoint per sensor, hence 1 cluster per side.
The total number of cluster per side per sensor per event is thus 9 (8 from background + 1 from signal), resulting in ~ 80 SpacePoints, of which only 1 correspond to true signal.

Since machine background hits are uniformly distributed in time, most of these background SpacePoints are off-time with respect to the signal, therefore the hit time is a powerful observable to significantly improve the tracking performance.

The background-signal discrimination, based on the reconstructed hit time, can be performed in various ways at different levels in the reconstruction; for example:

- Adding a time difference requirement on the RecoDigits constituting a Cluster;
- Applying a cut on the time when building the SpacePoints (both in terms of absolute time and in difference of the times of the SVDClusters on the two sides);
- Introducing a quality estimator based on the difference in time between the Clusters when combining them to form a SpacePoint. This is particularly useful to reduce the ghost hits;
- Utilizing the time information at the level of track finding, using it to train an MVA selector.

Some of these possibilities are under investigation; currently the track finding algorithm uses cuts on the times of the SpacePoints when combining them to create a track candidate, and later the timing information is used as an input for an MVA selector to compute a quality estimator for the track.

The current default method to compute the hit time is the CoG (see Section 3.3.1); another method, based on a neural network, is under development by the collaboration, but has up to now proved itself less robust and reliable than the CoG.

3.3.1 Center of Gravity (CoG) Definition

The *hit time* (also called *crossing time*) corresponds to the start of the rise of the APV25 curve (i.e. the first red star in Figure 3.2).

The main idea at the base of the CoG time estimator is to first evaluate the *peak time*, i.e. the time of the maximum of the curve, and then subtract the *peaking time*, i.e. the time needed by the curve produced by the APV25 to rise till the peak, starting from the hit time. In this way we can retrieve the actual hit time. The peaking time is characteristic of each strip and is measured in calibration runs (see Section 3.3.2.1).

The *peak time* is estimated using the center of gravity of the charge-vs-time distribution, as the weighted mean of the times T_n of the recorded samples, using their amplitudes A_n as weights. In formula:

$$T_{CoG_{raw}} = \frac{\sum\limits_{n=0}^{N} A_n \cdot T_n}{\sum\limits_{n=0}^{N} A_n} = \Delta T \cdot \frac{\sum\limits_{n=1}^{N} A_n \cdot n}{\sum\limits_{n=0}^{N} A_n} \tag{3.5}$$

where $\Delta T \simeq 31.4$ ns is the period of the APV25 clock and n runs over the N recorded samples. $N = 6$ by default, but it will probably become 3 in high-luminosity regime.

The above defined raw CoG time does not correspond exactly to the peak time. Some corrections are applied to take care of the introduced bias, as it is described in Section 3.3.2.

For what concerns the charge estimation of the strip, the CoGTimeEstimatorModule computes it as the maximum of the charge of the six samples, coherently with other modules previously present in basf2.

3.3.2 Corrections

Equation 3.5 shows the definition for the first step in the computation of the CoG time estimator (called simply CoG where there is no ambiguity). Some corrections need to be applied to achieve a better hit time estimation; some of them are under investigation, and some will not be included in the early period of running of Belle II, while others are default.

The main purpose is to have a reliable output for the hit time, that is conventionally considered to correspond to time 0 in the SVD reference frame. This is done because we want to associate the same hit time to all of the hit corresponding to the same physical event. The choice of considering this time equal to 0 is not only natural but is also driven by the fact that the neural network method, whose development started before the one of the CoG, returns a hit time 0 for signal events.

In the following I describe the corrections applied to the raw CoG; their effects are shown in Figure 3.11.

3.3.2.1 Strip Calibration Peaking-Time

The peaking time for each strip is a known information, stored in the database. The peaking time is different for the two sides (p, also known as U, and n, also known as V) of the sensor, because it is dependent on capacitance seen by the preamplifier, hence by the length of the strips. This value can also change for the different strips of the slanted (FW) sensors, as their lengths varies depending on their position on the sensor. It is, generally speaking, a strip-dependent quantity.

The first correction hence consists in the subtraction of the peaking time from the $T_{CoG_{raw}}$: even if the latter was an optimal estimator of the peak time, it still would not be a good estimator of the hit time.

The peaking time stored in the database includes a non-exactly known offset coming from the procedure followed to extract these calibrations. The calibration procedure consists in injecting a charge in the channel through a calibration capacitance, and the offset is related to the time passing from the starting of the rise of the voltage applied to the capacitance to the actual injection; it is a constant term, common to all the strips, hence it does not represent an issue because of the presence of the shift-mean-to-zero correction described in Section 3.3.2.3.

3.3.2.2 Trigger-Bin Correction

As explained in Section 3.1.2, the information about the moment of arrival of the Trigger with respect to the APV25 clock is saved, expressed in the form of which fourth of the APV25 clock window the Trigger has arrived in. Assuming that the Trigger information is reliable

(that is equivalent to state that the TJ is small with respect to the CoG time resolution), this tells us with more precision the moment which the curve has started to rise. The effects of different TJs are shown in Section 3.4.1.
If we take into account this information, it is possible to move the shaper outputs for different TBs back to one single starting point (that is defined as the central point of the first fourth of the APV25 clock window): this would reduce the width of the distribution of the estimated hit time, because it would be equivalent to overlay four distributions, each of them approximately having the same shape, but translated in time by a quantity that is in first approximation equal to 1/4 of the clock time window ΔT.

The CoG time distribution relative to the four different TBs, as well as the effect of this correction, are shown in Figure 3.13.

Operatively, we subtract from $T_{CoG_{raw}}$ the following quantity:

$$\Delta T_{TB} = \Delta T \left(\frac{1}{8} + \frac{1}{4} \times \mathrm{TB}_{\mathrm{Index}}\right) \tag{3.6}$$

where $\mathrm{TB}_{\mathrm{Index}}$ ranges from 0 to 3 (see Figure 3.4). The constant term $\Delta T\frac{1}{8}$ is the mean value of 1/4 of the time window, because we can assume that the actual starting point of the curve is uniformly distributed in time in the considered sub-window.

3.3.2.3 Shift Mean to Zero

This correction is quite straightforward, as it simply consists in the subtraction of the mean value of the hit time distribution from the hit time of each strip. This forcibly move the mean of the distribution to 0. This correction takes care of the bias introduced by the CoG when estimating the peak time, as well as the offset in the knowledge of the peaking time deriving from the calibration procedure.
Operatively speaking, the calibration procedure for this correction consists in extracting the mean of the CoG time distribution from a calibration-dedicated run of Belle II, and save this number in the database. Later, during data taking, this correction is subtracted from the estimated CoG time.

This operation can be performed extracting the mean of the CoG time from different subsets of sensor modules of the SVD:

- **Single sensor**: the mean of the CoG distribution for each sensor is saved;
- **Integration in ϕ**: this correction exploits the symmetry in ϕ that the detector presents. On a given layer, all the sensors with the same sensor index (i.e. the same θ angle) are evaluated together; the extracted mean is the one resulting from 7, 10, 12, or 16 sensors, one for each ladder (for the layer 3, 4, 5, or 6; see Table 2.4);
- **Sensor type**: layer 3 sensors are different from the barrel sensors of the outer layer (see Section 2.4.2) and from the FW sensors, hence these three different kinds of sensors can be grouped together when performing the extraction of the mean of the CoG.

Whichever grouping is chosen, the procedure is separately applied for the two sides, U and V, because having different strip length their behavior is different.

This correction is currently under investigation to be applied in the early stage of data taking.

3.3.2.4 Trigger-Bin Dependent Shift

The fourth and final correction that can be applied is similar to the previous one, but it depends on the TB of the event: the procedure described in Section 3.3.2.3 is repeated for each of the four TB index, and four different corrections are saved.

This correction is being implemented because the Trigger-Bin correction does not make the four distributions perfectly overlap. This happens because the CoG estimation, given the shape of the curve of the APV25, is not linear, and the introduced bias depends on the actual start of the sampling, as can be seen in Figure 3.13.

This correction still has to be properly examined, and is currently not take into consideration to be applied at the beginning of data taking.

3.4 Performances

A variety of tests has been carried out to check the reliability of the CoG hit time estimator, using simulated events and Test Beam data.

In the test performed on simulated data the entire SVD structure has been simulated, and $\Upsilon(4S)$ events have been generated.
The Test Beam data have been collected at DESY, in 2017, with a 5 GeV/c electron beam in 1 T magnetic field. In this test the SVD did not present its complete structure, as only one ladder per layer and two rectangular sensors per ladder were mounted. These data are therefore characterized by a low multiplicity and low bending (when the magnet was on), so they are somehow a simplified version of the data that we will collect in actual run, but they are precious data to perform numerous tests.
Test on Cosmic Ray runs are currently being carried out.

3.4.1 Test on Simulation

To study the performance of the CoG in a controlled environment, I have used the official Belle II simulation. It has to be kept in mind, however, that there are the uncertainties deriving from the models used for the simulation.
Unless otherwise specified, the following plots have been produced in the simulation of $\Upsilon(4S)$ events with full SVD structure, with no background, TJ = 0 ns.

The first thing that can be controlled is that the sampled shape of the curve from the APV25 matches the expected one. Some shapes have been plotted for both the U and the V side; the results are visible in Figure 3.10, and they have to be compared with the example curve in Figure 3.2. Given that the latter is a reference curve obtained from calibrations, the correspondence is very good, in particular for the V side. The different peaking time of the two sides is due to the different length of the strips and the capacitance they see.

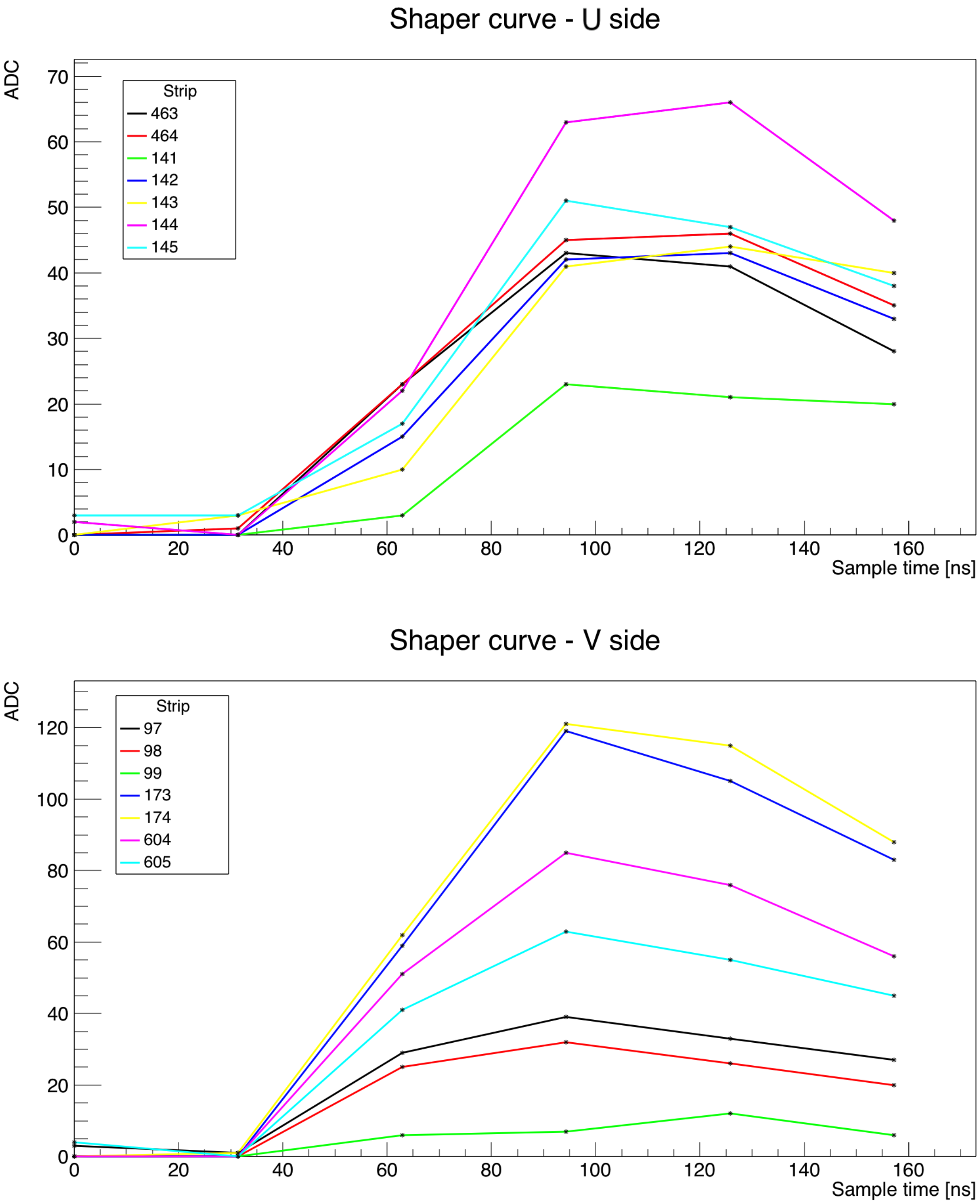


Figure 3.10: Reconstructed shaper curves for U (top) and V (bottom) sides (the digital unit are called ADC). Different heights correspond to different amounts of e-h-pairs generated in the DSSD, hence different energy released by the passing particle. This can be due to particles having different energy, statistical fluctuations of the energy release, or because the charge generated by a particle is collected by more than one strip.

The effects of the peaking-time correction and of the Trigger-Bin correction on the CoG distribution are shown in Figure 3.11.

Since the peaking-time correction is strip-dependent, it could modify the structure of

the distribution. In practice it behaves like a shift, because the peaking time for the strips belonging to the same side are almost always identical. The Trigger-Bin correction, on the other hand, actually reduces the width of the distribution.
The shown results are obtained evaluating all the SVD strips of the same side, since the performance is very slightly dependent on the particular sensor.

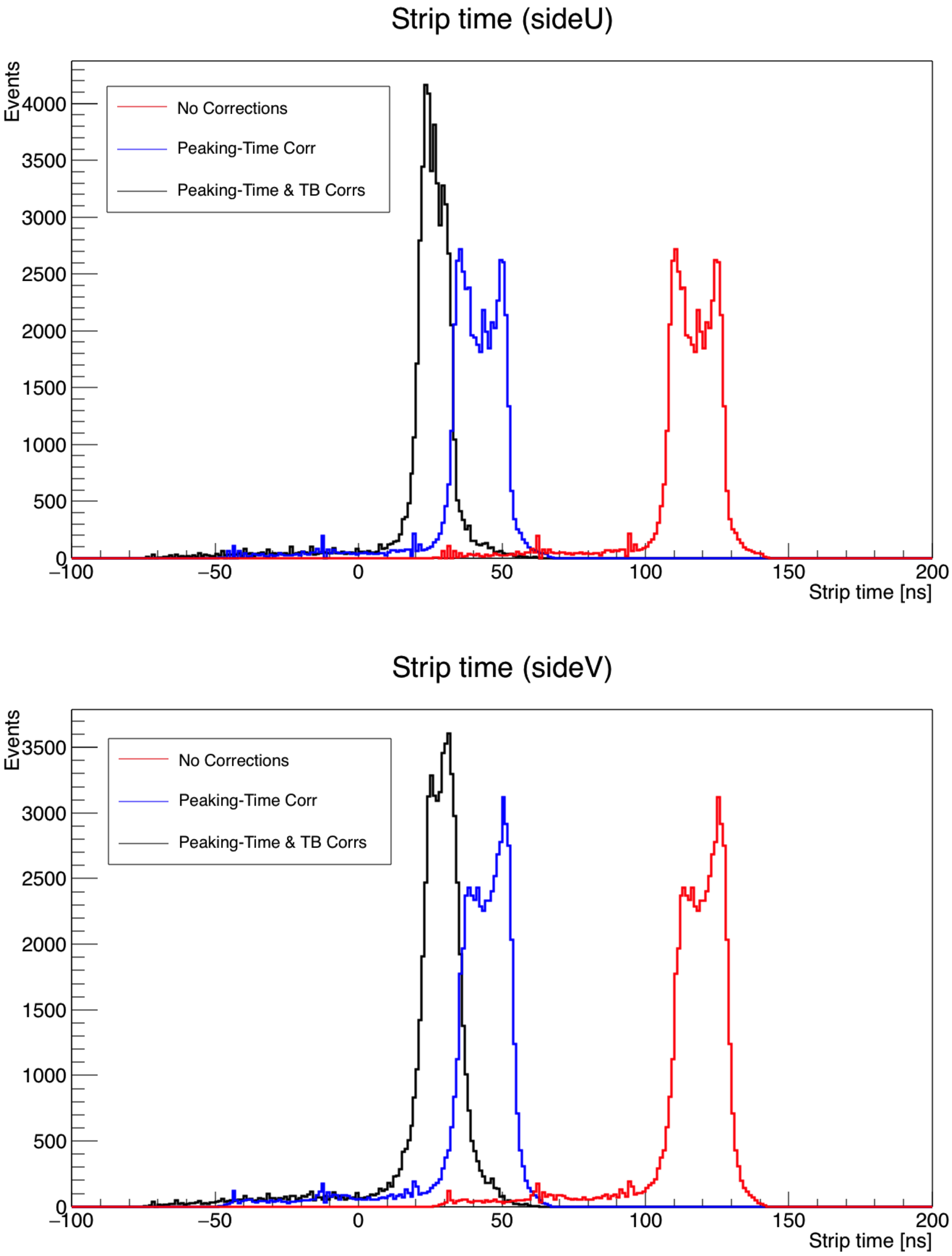


Figure 3.11: CoG time distribution, for the U (top) and V (bottom) sides, applying different corrections.

Using the truth information it is possible to plot the strip time resolution, i.e. the difference between the reconstructed CoG time and the MC time of the hit. In this case we are not interested in the central value of the distribution, because it depends on the bias of the

CoG and on the definition of hit time of the simulation. We are instead interested in the width of this distribution, because it tells us how good we are in reconstructing the hit time, regardless of the overall bias.

In Figure 3.12 the strip time resolutions are reported for two different sets of applied corrections. In the first case I used only the peaking-time correction, to take into account the peaking time of the strips. In the other case I also applied the Trigger-Bin correction, taking into account the TB information and hence reducing the width of the distribution.

The resolution is $7 \div 8$ ns in case the Trigger-Bin correction is not applied, otherwise it improves up to $5 \div 6$ ns. We see again that the the V side behaves better than the U side, having a resolution ~ 1 ns smaller.

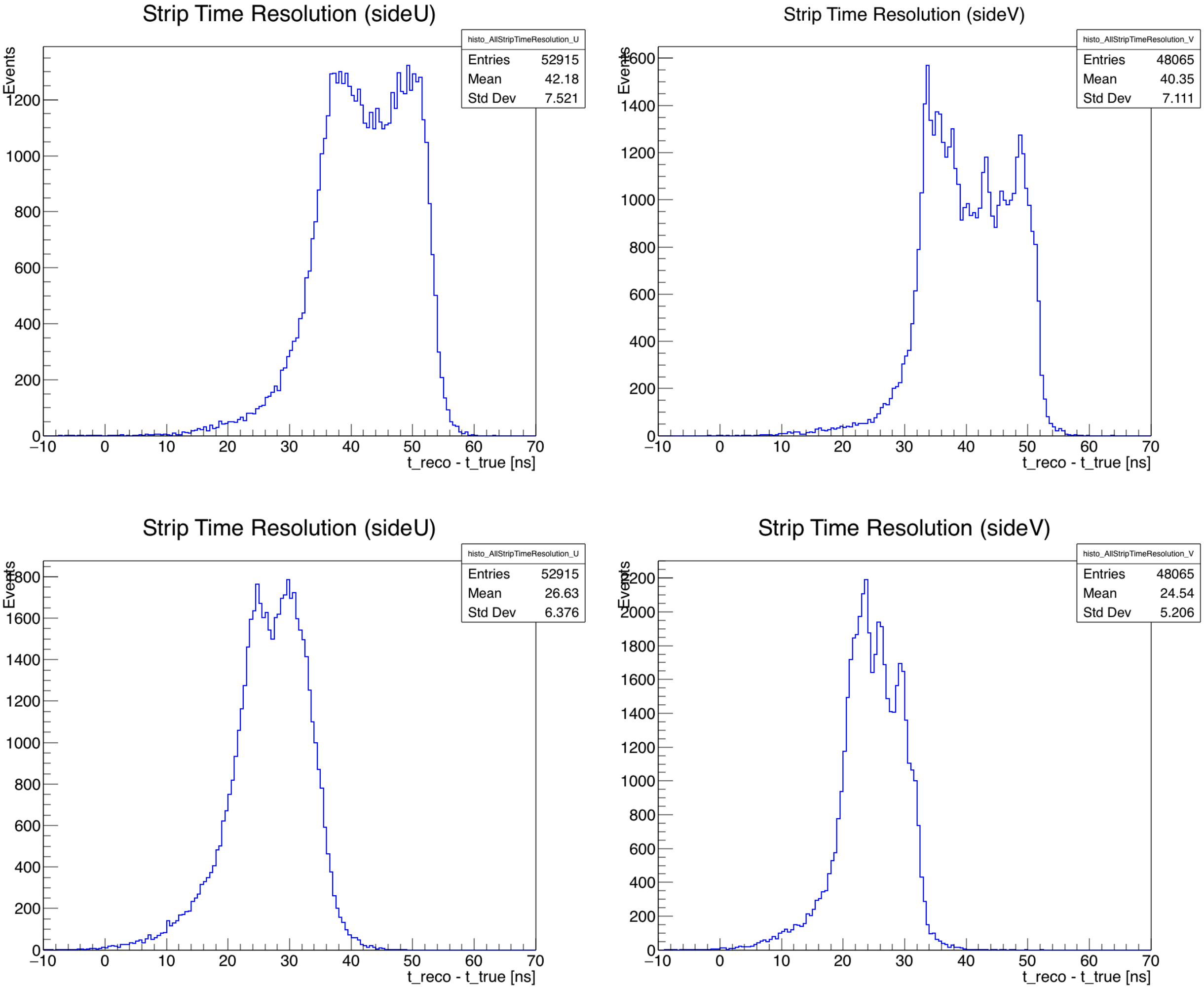


Figure 3.12: Strip time resolution, calculated as the difference between the CoG reconstructed time and the MC hit time. On the top there are the results obtained with only the peaking-time correction applied, while on the bottom also the Trigger-Bin correction is applied; on the left there are the distributions for the U side, on the right for the V side. Only the events that have a corresponding MC particle have been plotted.

Figure 3.13 shows, separately, the CoG distributions for events classified with the four different TBs, for two different sets of applied corrections. In the first case I only used the

peaking-time correction, in the second I also used the Trigger-Bin correction.
We can see that the distributions are more separated when only the peaking-time correction is applied. It is evident that the events classified with different TBs have a different average time. This effect derives from the way the CoG is computed, and from the shape of the curve: if the events have a higher TB, it means that the particle is passed late in the first APV25 sampling window, so the overall timing is shifted onward, because the CoG time is computed as a weighted average of the times of the samples.
If also the Trigger-Bin correction is applied, each of the four distributions is shifted according to Equation 3.6. The result is slightly different than the expected one, as the four distributions are not shifted such that their averages become the same, but the Trigger-Bin correction overcorrects by ~ 3 ns each distribution. This effect derives from the shape of the APV25 shaper output: if the sampling starts 8 ns later, the CoG time does not increase by exactly 8 ns, but by a smaller time. The details of this effect are currently under investigation, but can phenomenologically be absorbed in the fourth correction, the Trigger-Bin dependent shift (see Section 3.3.2.4).

For the sake of readability, only the results regarding the V side are reported.

It is important to verify that the CoG has the capability to discriminate between signal and background. Using MC information it is possible to distinguish between activated strips associated with a signal particle, i.e. associated with a True Hit (TH), and activated strips not associated with a TH (that can be activated by electric noise or physical background).
The two CoG time distributions can be seen in Figure 3.14.
It is evident that using the CoG and applying an appropriate cut it is possible to eliminate a huge percentage of background (order of $\sim 75\%$ even in a conservative case), losing very few or no signal at all. This cut is actually meant to be applied at the level of the SpacePoint creation, also because the time of the SVDClusters is a more precise quantity than the time of the single strips: the cluster time is computed as the weighted mean of the times of the strips composing the cluster, weighted with their charge, so it naturally reduces the importance of the strips with a low signal, whose shaper output is more disturbed by noise.
The structure at low times in the distribution of the CoG for background events derives from the way the CoG is calculated and the typical shape of the APV25 curve: background events whose curve peak is before the start of the sampling have a low CoG associated, as only the slow decreasing tail is used for the computation of the CoG. Since the background is almost continuously distributed in time, the curve peak is narrow with respect to the tail, and the sampling window is narrow with respect to the whole curve, most background events present a low CoG associated. Further studies are being performed to fully understand this feature.

Another important test that can be carried out exploiting the potential of the simulation is to verify the effects of the Trigger Jitter on the CoG time distribution. As already mentioned, the higher the TJ is, the lower is the precision of the Trigger in providing the correct start of the event, so, going back in the ring buffer by the amount of cells given by the fixed latency, we could end up reading samples that are low or not related with the passage of the particle.

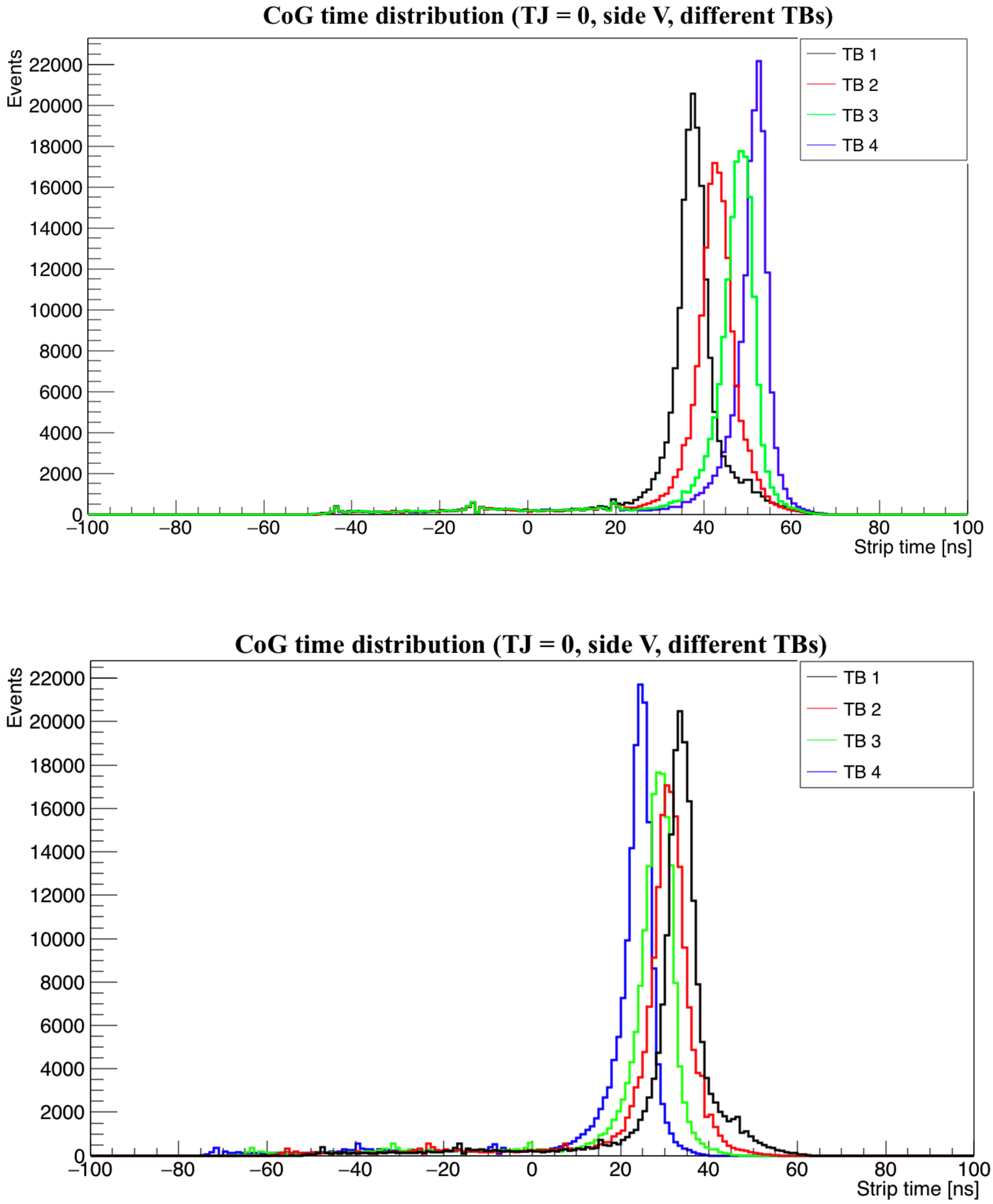


Figure 3.13: CoG time for different Trigger Bin, for the V side only. On the top there are the results obtained with only the peaking-time correction applied, while on the bottom also the Trigger-Bin correction is applied. The Trigger-Bin correction overcorrects the distributions by ~ 3 ns.

In the simulation, the uncertainty on the Trigger has been modeled with a Gaussian having a width equal to the TJ and a null mean.

Six different TJs, spanning from 0 to 60 ns, have been simulated and applied. The resulting CoG distributions can be seen in Figure 3.15.

The values of the standard deviation of the CoG distribution as a function of the TJ are reported in Figure 3.16. It is interesting to note that the V side behaves better than the U

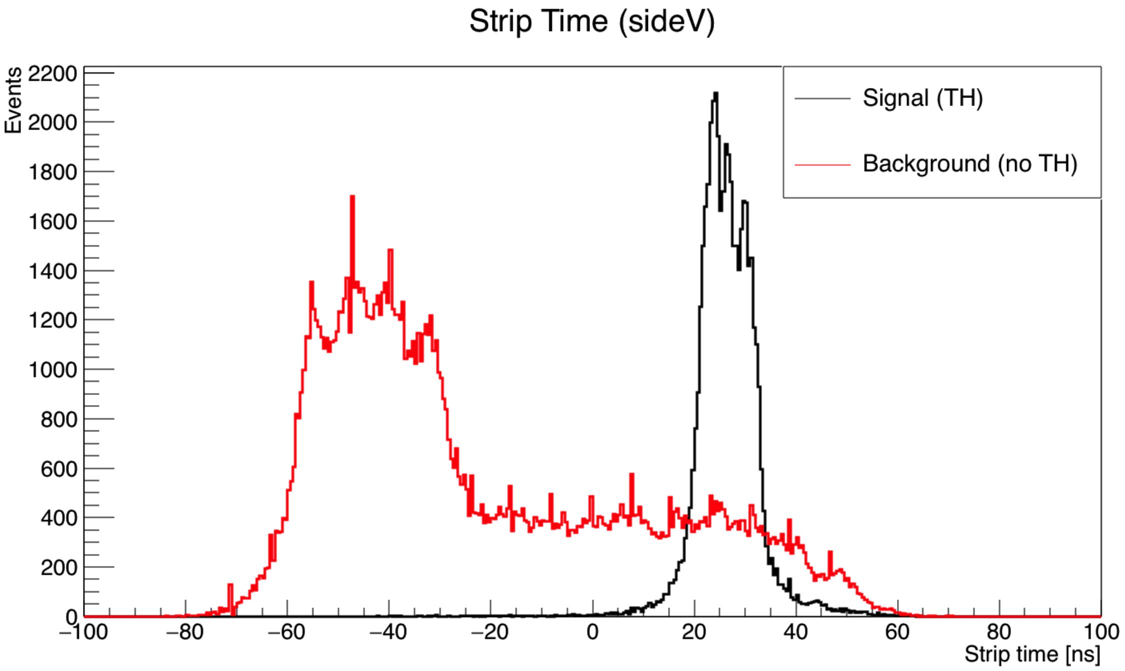


Figure 3.14: CoG time distribution, for the V side only, for signal and background strips. High discrimination power can be achieved with a simple cut in time. Both the peaking-time and the Trigger-Bin correction are applied.

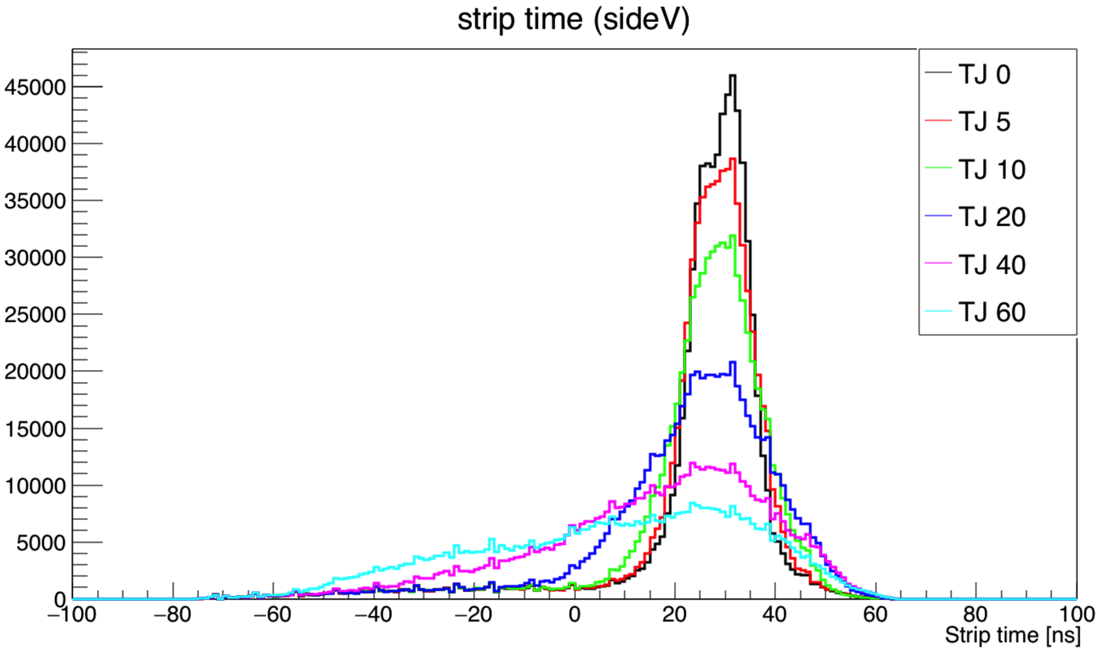


Figure 3.15: CoG time distribution, for the V side only, applying different TJ. For TJ $\gtrsim 20$ ns the distribution becomes appreciably broader. Both the peaking-time and the Trigger-Bin corrections are applied. The TJs are expressed in ns.

side for low values of the TJ, while arriving at very high values (the expected TJ in the first running period of Belle II is approximately $\sim 20 \div 40$ ns, while the TJ for late running is expected to be about 10 ns) the U side presents a better resolution than the V side.

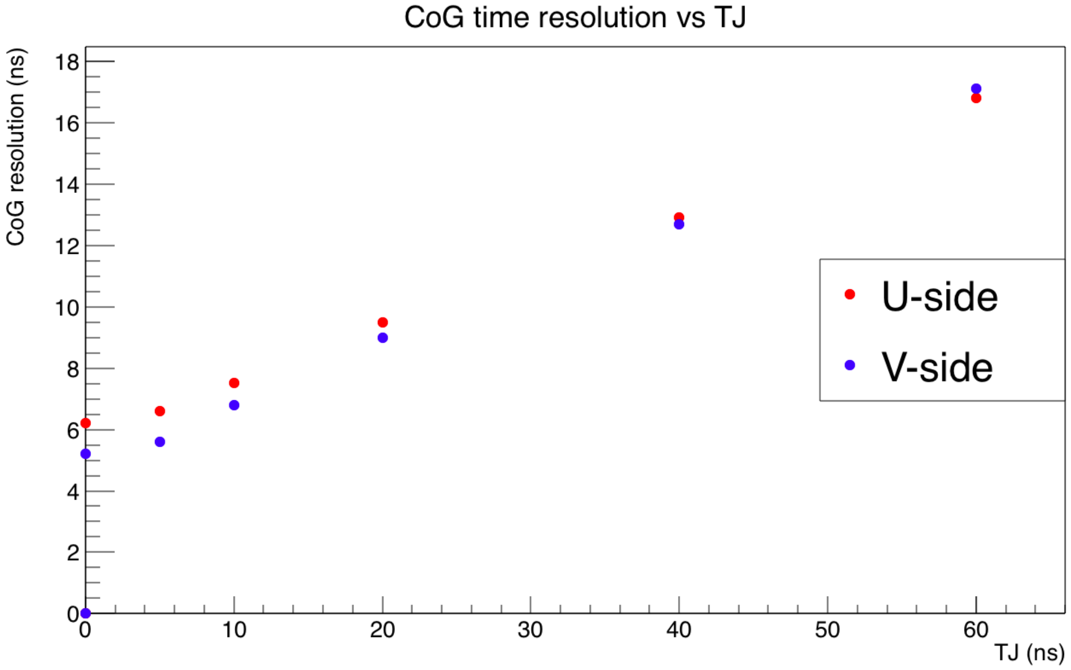


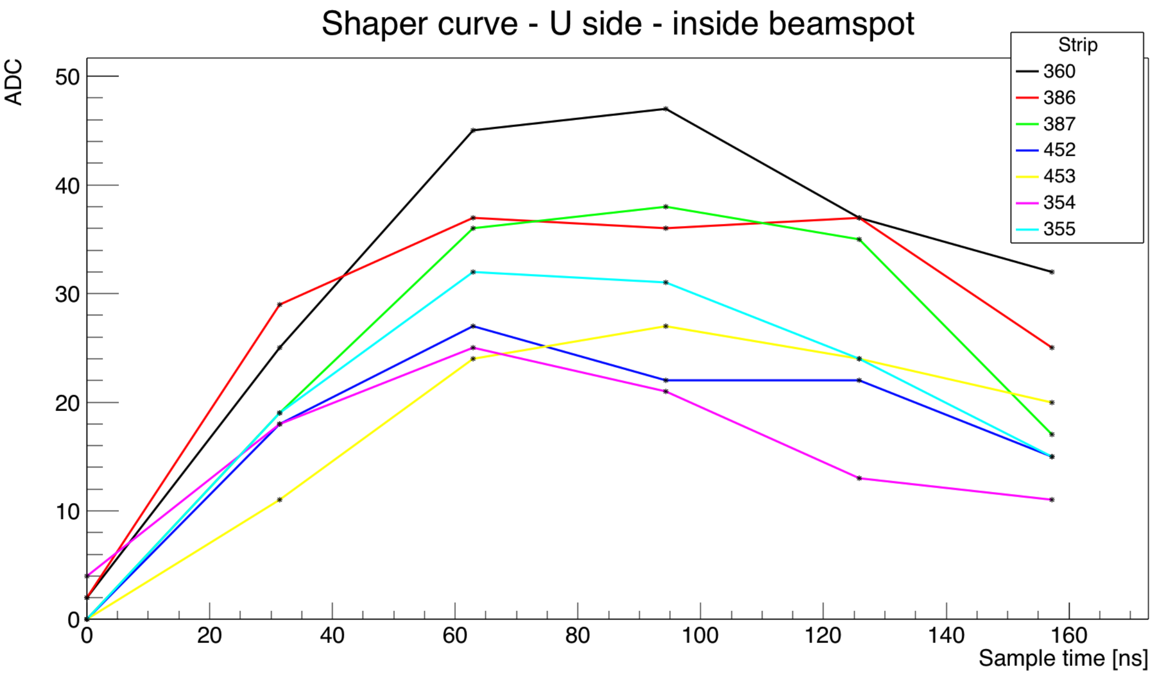
Figure 3.16: Dependence of the CoG resolution as a function of the TJ, both U and V sides. Both the peaking-time and the Trigger-Bin corrections are applied.

3.4.2 Test on Test Beam Data

Test Beam data allow to check the behavior of the CoG estimator on real data, even if with a reduced number of SVD sensors and with only one species of particle, i.e. high-energetic electrons.

Similarly to what has been done with the simulated data, we can plot the curves produced from the APV25. We do not expect them to be exactly equal to the ones previously seen in Figure 3.10, because their shape depends on the Trigger, and simulated Trigger for real Belle II can be quite different from the actual Trigger used during the Test Beam.
What can be seen from Figure 3.17 is that apparently there is a different set Trigger latency, as the start of the rise of the curve happens earlier than what is seen in simulated data.

Test Beam have been used to observe the distribution of the CoG and compare it with the one seen in the simulated data. I did that for two different kind of strips, the ones being part of the beamspot (i.e. the ones illuminated by the electron beam) and the hot strips, that are strips which fire noise with high rate. We expect a defined structure in the first case, and a more flat one in the second case.
The results are shown in Figure 3.18. These distributions correspond to the $T_{CoG_{raw}}$, i.e. no correction had been applied. It is evident that the two distributions are very different, and this can be used to discriminate between signal events and, at least, hot strips. The performance of the U side is slightly worse than the V side, and this mainly due to the higher length of the strip on the U side.



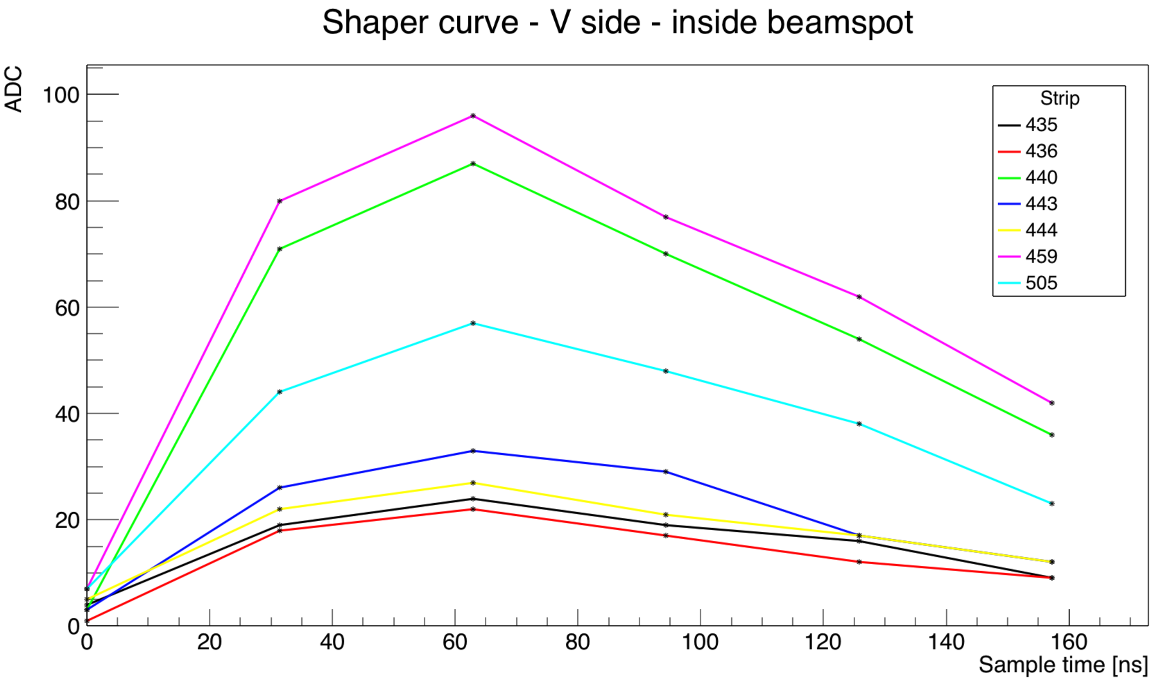


Figure 3.17: Reconstructed sampled shaper curves for U (top) and V (bottom) sides; only events inside the beamspot of the electron beam are taken into account, to ensure to look at good signal events and not noise. The different raising time between the two families of curves is evident, so we expect a different time distribution for the two of them.

Since 5 GeV electrons have a negligible flight time between the different layers, if we were able to follow a given particle through the different layers, it would be possible to verify that the CoG reconstructed on the four layers is compatible; though, no reliable tracking was available to analyze these data. To overcome this problem I plotted all the reconstructed

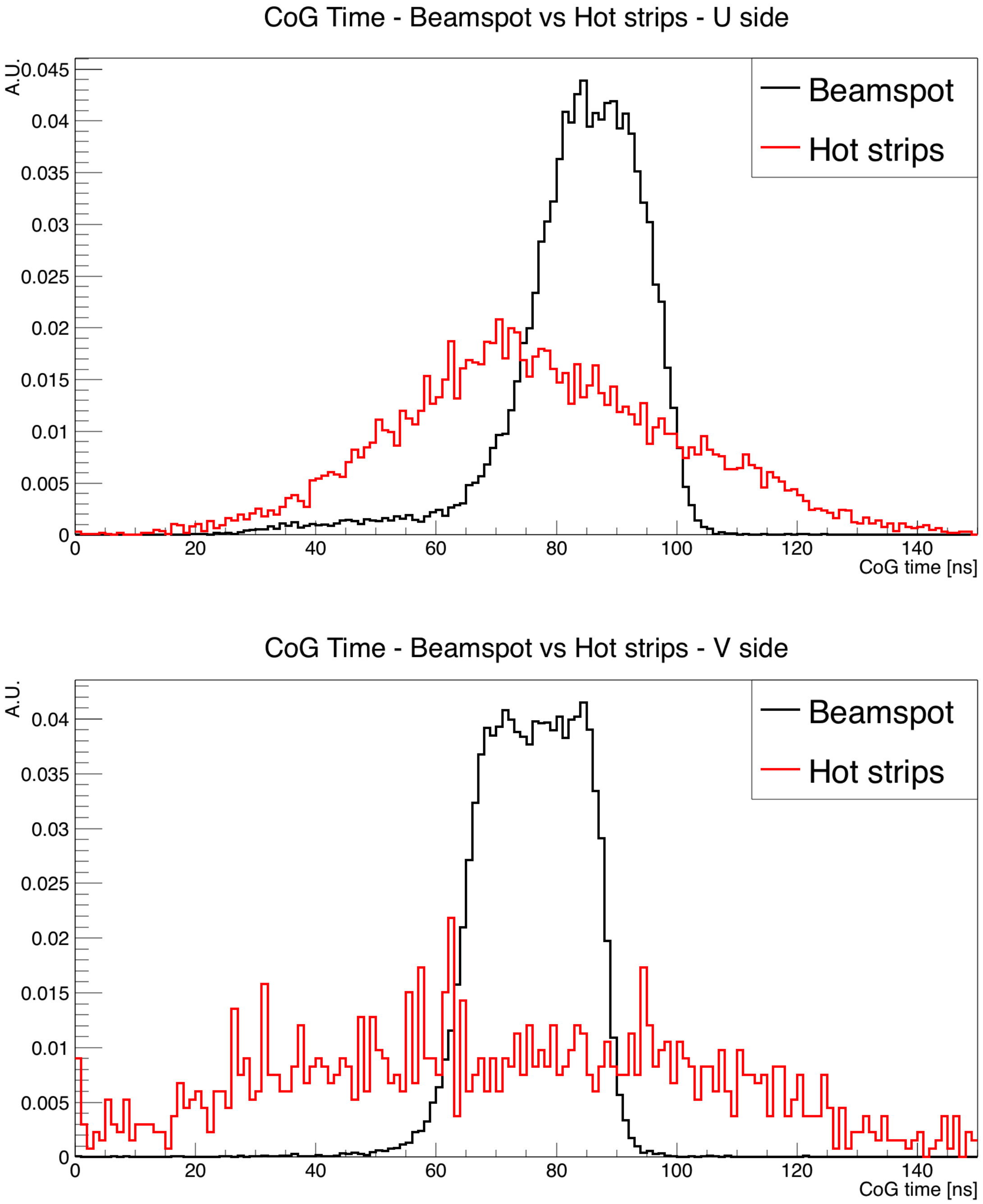


Figure 3.18: CoG time distribution for U (top) and V (bottom) sides. The plots show the distributions of the CoG for the strips in the beamspot, so where the probability of having a real signal event is high, and for the hot strips. The ordinate axis is in arbitrary units, as the two distributions have been normalized to 1 to help the comparison. No correction has been applied.

CoG from a layer against the ones of another layer, event per event. The downside of this approach is the high combinatorial background, i.e. the combination of times that are not related to the same physical particle.

It is important to note that also physical background is present, because the electron beam had to pass through the material of the magnet, hence some particles were created due to the interaction of the electrons with the material.

The results of this study can be seen in Figure 3.19. As expected, regardless the huge combinatorial background, a linear correlation between the two layers can be clearly seen. A linear correlation between the times of hits on different layers coming from the same physical particle is indeed what we expected: all the layers of the SVD are synchronized and the Trigger Bin information is the same for all of them. So, for example, if the CoG time on a layer is lower than the average, we expect that the same happens on the other layer, with approximately the same magnitude, because most of the causes that can result in a different time estimation than the average affects all the layers, such as a bad Trigger information (that implies that the curve is not sampled in the optimal region, so the computed weighted average is not the expected one).

The difference between the means for layer 4 and layer 5 is due to the distortion of the distribution caused by the background, since the flight time between two layers of the SVD (whose distance is order of 10^{-1} m, see Table 2.4) for an relativistic electron is less than 1 ns. We see again here that the U side presents a broader distribution than the V.

3.4.3 Effects on Track Finding Efficiency and Fake Rate

Finally, what I have studied is the effect of the timing information on the performances of the SVD-only tracking.

To evaluate the tracking performance, MC information are used.
A MC track candidate is the collection of the SVDClusters, i.e. hits, related to a given MC particle.
A reconstructed track candidate, found by the track finding algorithm, is considered to be Truth-Matched (TM) if at least 66% of the SVD hits it is composed of correspond to MC hits of the same MC track. In most of the cases a track is comprised of 8 SVD hits (2 per layer, i.e. 1 for each side, for each of the 4 layers), but in some situations there could be less, e.g. if the particle decays between two SVD layers or exits acceptance region, or more, if it passes through the small overlap regions that are present in the SVD.
The quantities to look at are the following:

- **Finding efficiency**: number of TM tracks over all MC tracks;
- **Fake rate**: number of non-TM tracks over all reconstructed tracks;
- **Hit efficiency with matched tracks**: number of correctly reconstructed hits over all MC hits, for each track.

These quantities can be evaluated as a function of different observables, such as the transverse momentum, the ϕ or θ angle and so on. In Figures 3.20, 3.21, and 3.22 I chose to show them as a function of ϕ angle. The angle used on the horizontal axis of the plots is the *seed* ϕ angle, i.e. the angle ϕ of the MC track, evaluated at the IP.
All the graphs have been obtained in a simulation of 1000 $\Upsilon(4S)$ events with full SVD structure, with the nominal background, applying both the peaking-time and the Trigger-Bin correction. The track finding algorithm currently uses the time information applying cuts on

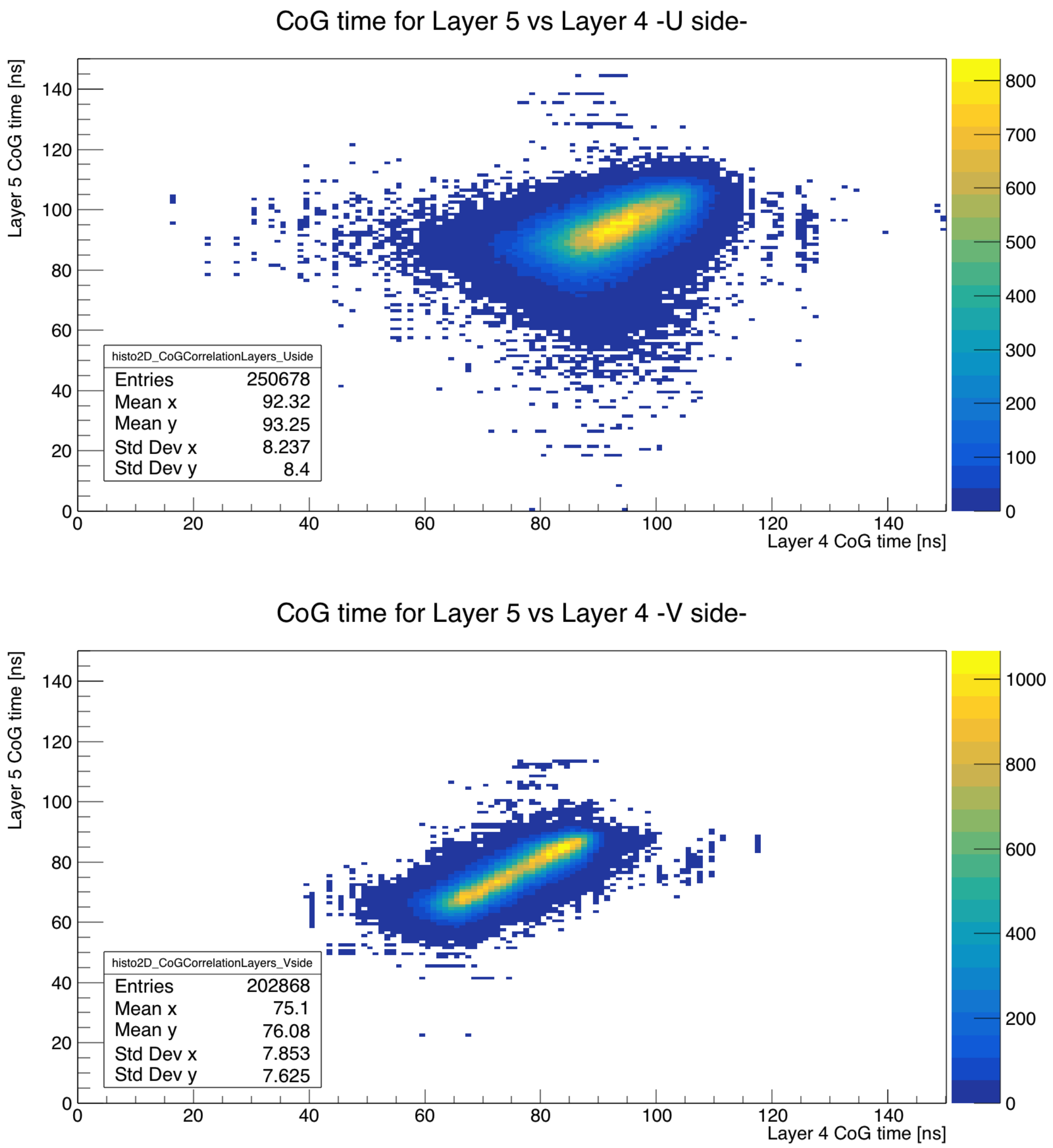


Figure 3.19: Scatter plot of the CoG for layer 4 and layer 5. The combinatorial background coming from random combinations of unrelated times is high. Nonetheless, a linear correlation between the CoG times of the two layer is clearly visible, indicator that the CoG behaves as expected. The background comes also from physical particles, generated from the interaction of the electron beam with the magnet used in the Test Beam. No correction has been applied.

the time difference of the SpacePoints when combining them to create a track candidate; it is also used as an input for an MVA selector that returns a quality estimator for the track candidate. It is worth to note that a CoG correction that is the same for all the hits of the same event has almost no effect on the tracking performances, because only the time differences between the hits are used by the algorithm. The only effect is on the width of the possible cut on the SpacePoints, that is currently very loose, to not lose any signal hit.

The timing information, provided by the CoG, is expected to help reducing the off-time background, such as the beam background, that is particularly high in the horizontal plane

(this is the reason why I chose to show the tracking performances as a function of ϕ). On the other hand, we do not expect that using time information the track finding efficiency will improve.
This is what we can see from Figures 3.20, 3.21, and 3.22. The hit finding efficiency and the track finding efficiency are statistically unchanged with or without timing information, while the fake rate is reduced by 80% (it passes from 25.8% to 5.8%, averaged over all ϕ), also it is basically flat in ϕ, while without timing information it presents a peak at 0 degrees.

We can thus conclude that the CoG time estimator provides a very useful information characterizing the hits that will play a significant role in improving the tracking performances in Belle II.

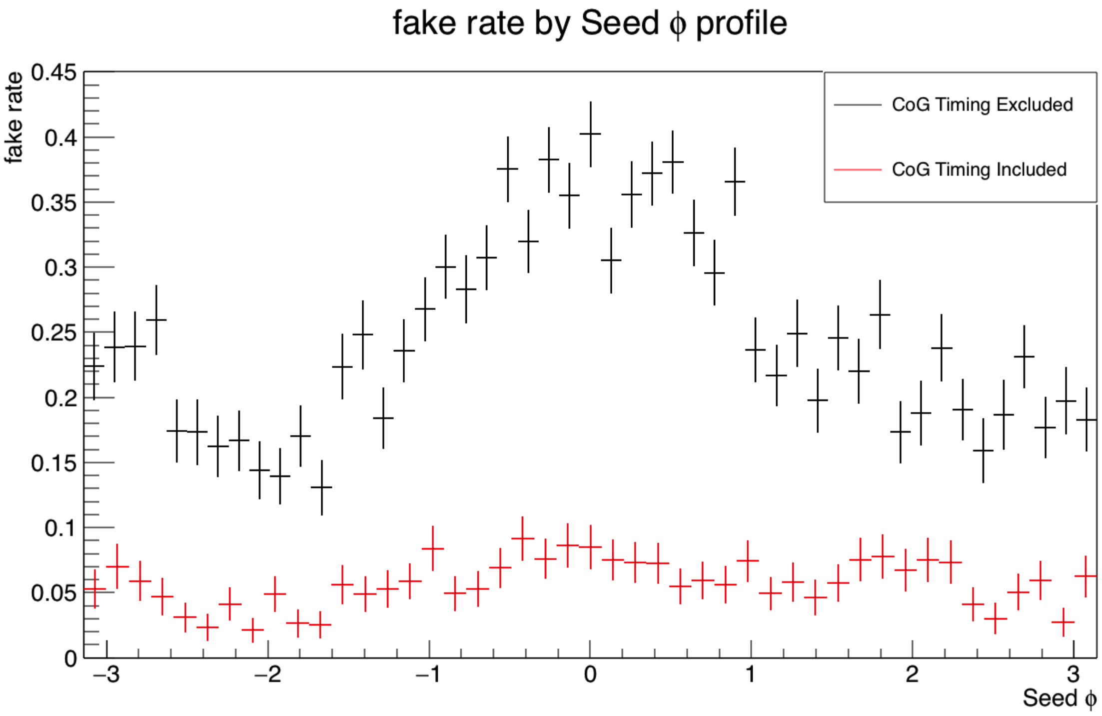


Figure 3.20: Fake rate with and without CoG timing information, as a function of the seed angle ϕ.

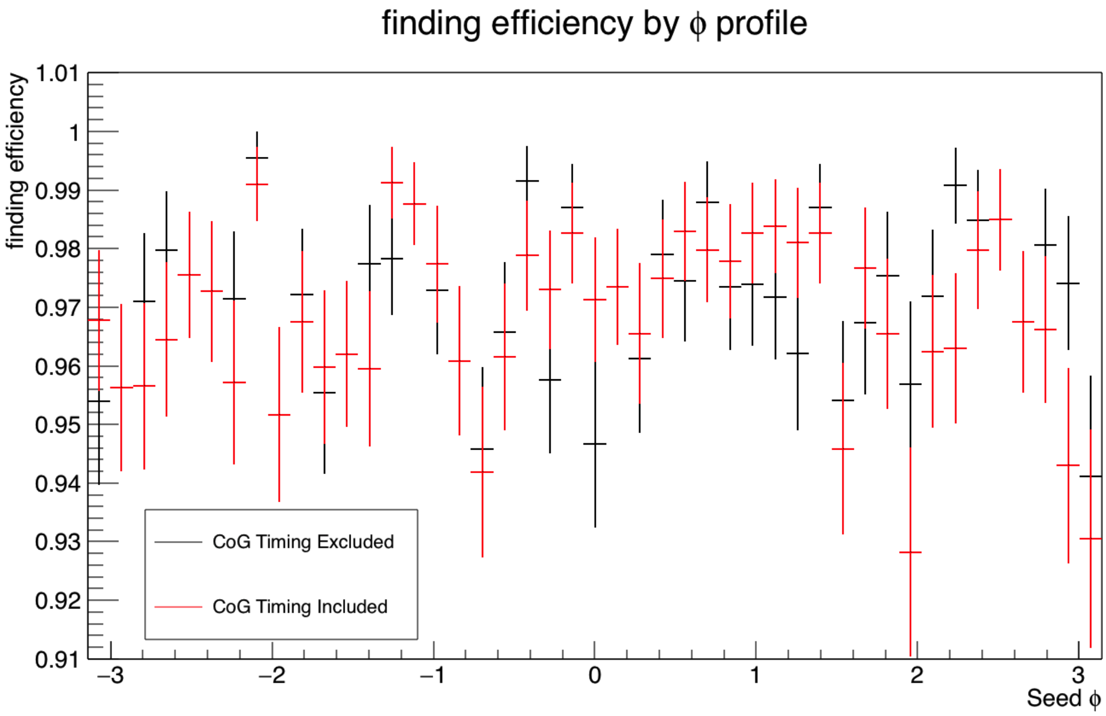


Figure 3.21: Finding efficiency with and without CoG timing information, as a function of the seed angle ϕ.

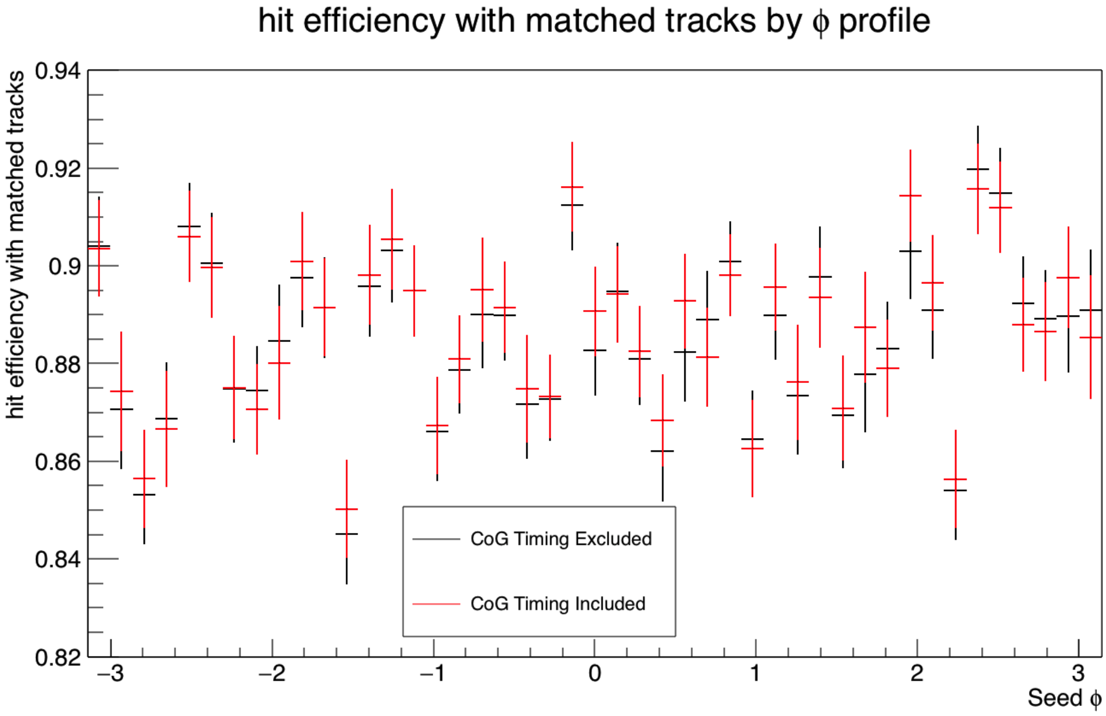


Figure 3.22: Hit efficiency with and without CoG timing information, as a function of the seed angle ϕ.

Chapter 4

Analysis Overview and Data Samples

This Chapter is dedicated to examine the characteristics of the $B \to K^{(*)}A$, $A \to \gamma\gamma$ channel, described in Section 1.3, and the strategies used to search for it in the *BABAR* dataset. The search is performed only in the charged mode $B^{\pm} \to K^{\pm}A$ with un-excited K, since the higher background in the neutral mode significantly reduces the sensitivity.

Section 4.1 provides a summary of the analysis steps, which are explained in detail in Chapters 5 and 6.
In Sections 4.2 and 4.3 the main characteristics of the signal and of the background are described, with particular attention to the features that can be exploited to discriminate between the two. Section 4.4 specifies the used data samples, their characteristics and the preparations needed to process them.

4.1 Analysis Method Outline

This analysis studies the process $B^{\pm} \to K^{\pm}A$, $A \to \gamma\gamma$, which is a particular case of the more general $B \to K^{(*)}A$ decay, illustrated in Figure 4.1.

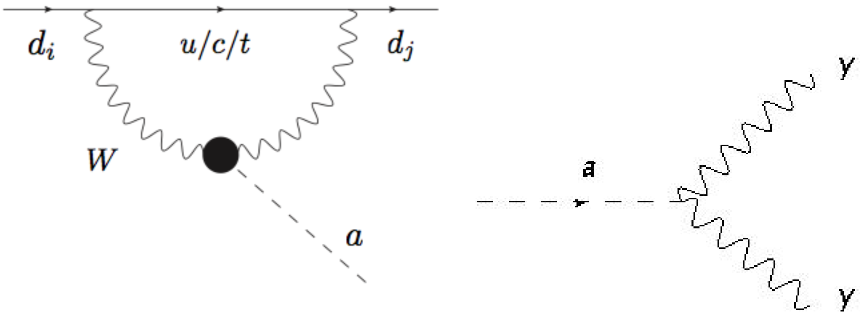


Figure 4.1: Diagrams for production and decay channels of the ALP. The same images can be seen in Figure 1.4.

The analysis is based on a search for diphoton resonance structures, performed by seeking

an excess in the $\gamma\gamma$ mass spectrum of events pre-selected with a specific topology ($K^{\pm}\gamma\gamma$) and with kinematical constraints on the mass of the B candidate.
Only the signal side $B^{\pm} \rightarrow K^{\pm}\gamma\gamma$ is reconstructed, with no additional constraints on the rest of the event. This constrains the ALP mass to be lower than:

$$m_A^{max} = m_{B^{\pm}} - m_{K^{\pm}} = 5279.26 - 493.68 \text{ MeV/c}^2 = 4785.58 \text{ MeV/c}^2 \tag{4.1}$$

The choice of using only charged decays is motivated by the need of reducing backgrounds, while not imposing constraints on the rest of the event maximizes the signal selection efficiency.
The main signal and background characteristics are described in the following.

The first step is the event selection. Several variables can be used to discriminate between signal and background, the most relevant being kinematic and shape variables, and PID quality requirement.
Because of the high number of possible discriminating variables, a simple rectangular cut & count procedure is not suitable; we therefore decided to adopt a more complex MVA (see Section 5.2.2), where machine learning is used to train the algorithm and optimize the selection process. The first main step is hence to find the optimal set of variables and training parameters for the MVA. This training step is performed on simulated data.

The optimized event selection is checked for consistency with a real data control sample, to verify that the MC simulation is a good model of the actual data.
Since the signal yield will be determined by fitting the diphoton mass spectrum, the next main step is to find a proper Probability Distribution Function (PDF) describing the background and the signal. The PDF of the signal is obtained from the MC samples, while the model for the background is extracted from the real data of the control sample, to ensure a more reliable description and to overcome possible biases coming from an imperfect modelization of the MC. The drawback of this choice is the risk of overtraining the modelization over statistical fluctuations of the control sample, as well as including a possibly present signal in the background model. The solution to overcome these dangers will be discussed in more details in Chapter 6, but the basic idea is to adopt a flexible background model and to exploit the narrowness of the signal mass distribution.

The third main step is to actually perform a fit to extract the signal yield. This is done after various checks, as the reliability of the background model and the stability of the signal+background fit procedure, as well as the capability of the fitter to extract the correct number of events, and the consistency of the background-only fit against the signal+background fit in case of absence of signal.

Last but not least, the final step is the estimation of the systematic errors.

In this thesis work I completed the first two steps, which will be described in detail, and started to work on the last two steps, which will be only outlined.

This analysis is a blind analysis. An analysis is blind when the analysts are not allowed to access the experimental results until the selection has been finalized, to avoid over-optimization of the selection procedure on statistical fluctuations and introduction of unconscious experimenter's biases. A variety of techniques can be used to blind an analysis; in my case, the selection procedure is trained using MC samples and it is validated only on a small subset of the real data. In practice we have used Run 3 (about 8% of the total data) as a *control sample* to compare it with MC backgrounds and check that there is a good agreement after the applied selections. In this way we are able to optimize the background PDF extraction over real data, avoiding inadequate modeling of the MC. In the end the control sample will be discarded from the full dataset of the analysis to ensure independence of the sample used for validation and PDF extraction from the measurement sample.

To use a real data control sample we need to assume that the signal width is narrow and/or that the signal strength is small, otherwise the presence of the signal in the control sample and the absence in the MC background would lead to poor agreement.

The narrowness of the signal can be checked with the MC-generated signal samples; its small strength is a reasonable requirement since this is an analysis meant to find an unknown particle (whose production branching ratio must hence be small, otherwise it would have already been seen in non-dedicated searches) or to exclude its presence.

4.2 Signal Characteristics

Since the signal is a B decay process, the kinematic is well known, and all the assertions made in Section 2.1 stand true.

The process has a number of features that we can exploit to select the signal. The most relevant are:

- Amongst the final products there are 2 photons, emitted by the decaying ALP;
- A kaon is the only other particle in the finals state of the signal; we are only looking at charged K, so we can include a quality requirement on the PID of the track, asking for good kaons;
- Since the process is a B decay, the variables ΔE and m_{ES} should peak at 0 and at the B mass respectively;
- Shape variables should be compatible with a spherical B event.

Due to the absence of neutrinos in the final state there is no missing transverse momentum in the signal itself, but since the other B, which we do not reconstruct, can instead produce neutrinos, missing transverse momentum/energy/mass are not useful quantities.

See Section 5.1 for details about the variables used for the selection.

4.3 Backgrounds Characteristics

Using the *BABAR* naming scheme, the following physics processes have been taken into account for the simulation of the actual sources of background:

- **QCD**, or **uds**, or **continuum**: production of light quarks (u, d, s);
- $\boldsymbol{c\bar{c}}$, or **ccbar**: production of charm quarks;
- $\boldsymbol{B^+B^-}$, or **BpBm**: production of charged B mesons;
- $\boldsymbol{B^0\bar{B}^0}$, or **B0B0bar**: production of neutral B mesons;
- **Bhabha**: e^+e^- elastic scattering;
- $\boldsymbol{\tau^+\tau^-}$, or **Tautau**: production of pairs of τ leptons;
- $\boldsymbol{\mu^+\mu^-}$, or **Mumu**: production of pairs of μ leptons.

$B\bar{B}$ can mean both or either of $B^0\bar{B}^0$, B^+B^- backgrounds; if the actual meaning is not clear from the context, it will be specified.
The cross section for each of these processes can be found in Table 2.1.

There are almost no Standard Model channels identical to the one searched; the exceptions are reported below. The branching ratios are from the Particle Data Group (PDG) [4].

- $B \to K\pi^0$ (BR $= 1.3 \times 10^{-5}$);
- $B \to K\eta$ (BR $= 2.4 \times 10^{-6}$);
- $B \to K\eta\prime$ (BR $= 7.1 \times 10^{-5}$).

It is also necessary to take into account the BR of the particle emulating the ALP of going into 2 photons; for the three particles above we have:

- BR $\pi^0 \to \gamma\gamma = 98.8\%$;
- BR $\eta \to \gamma\gamma = 39.4\%$;
- BR $\eta\prime \to \gamma\gamma = 2.2\%$.

These channels cannot be eliminated with any variable cut, because they have the exact same structure of the signal; the only solution is to cut the diphoton mass distribution in the mass regions of π^0, η, and $\eta\prime$, so these will be areas where we will have no sensitivity.
While there are no other known π^0 decay channels with two or more photons, the same is not true for η and $\eta\prime$. Both of them can decay emitting 2 ore more photons and/or π^0. These other channels are not troublesome because they cannot completely mimic the signal: the extra particles generated in the process take away an amount of energy and momentum such that the requirements on m_{ES} and ΔE are no longer satisfied.

The main background is combinatorial, arising from the random combination of a charged track and a pair of photons; any track and any two photons can in principle mimic the signal: we expect that an important part of the photons comes from real π^0 decays, and an important part of the tracks comes from real charged pions.

Most of the background arises from uds, ccbar, or $B\bar{B}$ channels: Bhabha, Mumu, and Tautau backgrounds are far less troublesome, because they present a very different behavior.

Bhabha background consists almost only in two highly energetic $e^{\pm}$ emitted back-to-back in the CM of the $\Upsilon(4S)$, close to the beam direction. A high number of these events do not enter the angular acceptance. For a Bhabha event to mimic the signal it is necessary that two bremsstrahlung photons are emitted by the $e^{\pm}$ in the acceptance region; also the electron has to be misidentified with a kaon, and the latter is highly improbable given the PID requirement asked on the track (see Section 5.1.7). Also the invariant mass of an electron and its emitted photons is ~ 5 GeV lower than the mass of a B, so it is very unlikely to meet the kinematic requirements.

Mumu background is similar to Bhabha as two high-energetic $\mu^{\pm}$ are emitted back to back. As for Bhabha, a muon should radiate two photons and it should be misidentified as a kaon to be able to mimic the signal, and this is highly unlikely, as well as the possibility that these three particles can reconstruct the invariant mass of a B.

Tautau background is not as different as the previous two from the signal, nonetheless its topology is not spherical, so shape variables are powerful to reject it, and it has a mass less than a half than the one of a B.

It is therefore easy to reject these background. This assertion can be verified at posteriori (see Section 5.3.3).

uds and ccbar backgrounds can be discriminated using shape variables and requirements on m_{ES} and ΔE because they have a jet-like behavior, different from the one typical of the B-like events.

Most of the $B\bar{B}$ backgrounds, arising from random combination of a track and two photons, does not reconstruct a real B, so it can be eliminated with requirements on m_{ES} and ΔE. Almost all the events different from the ones that perfectly mimic the signal ($B \to XA$, $X \to \gamma\gamma$, $X = \pi^0, \eta, \eta\prime$) can be easily rejected, hence we expect also the B0B0bar background being negligible.

4.4 Data Sample Preparation

The analysis optimization requires three kinds of data: the signal MC, the background MC, and the real data control sample.

In the *BABAR* analysis framework, all data samples are organized in ROOT files. To reduce CPU time requirements and simplify the analysis, the data samples undergo a loose pre-selection, meant to reduce the amount of data, eliminating events that are very unlikely to be signal.

In the pre-selection process, the $K\gamma\gamma$ kinematical combination is formed, and the resulting quantities, like m_{ES} and ΔE, as well as the lists of ALPs and B candidates, are stored in a ROOT ntuple structure to be easily accessed by the analyst.

MC ntuples contain the same reconstructed event information available in real data ntuples, with the addition of the MC information relative to each particle of the event; in particular, for each reconstructed particle, it is known if it corresponds to a MC particle (i.e. it is Truth-Matched, TM) or not, and for each MC particle, the information about its decay products (*daughters*) and the particle it has decayed from (*mother*) are provided.

4.4.1 Pre-selection Procedure

The pre-selection procedure has been applied both to data (in our case, before unblinding, only the control sample, that is Run 3) and MC samples (both signal and backgrounds).
The pre-selection is the following:

- Photons that are candidate as being ALP's decay products have to satisfy the following requirements [35]:
  - Minimum energy: 0.030 GeV;
  - Max lateral moment (a quantity defining how spread is the photon cluster in the EMC [35]): 0.8.
- $K^{\pm}$ candidates that, together with the ALP candidates, reconstruct the B meson, have to satisfy a minimum PID requirement; in particular, they have to be part of the KCombinedSuperLoose list, that is the loosest PID list for kaons [35] (in the *BABAR* reconstruction software, particles are classified according to hierarchical set of identification classes, called lists; see Section 5.1.7 for more details);
- The B candidate given by the combination $K\gamma\gamma$ must satisfy the following requirements:
  - $m_{ES} \in [5.0, 5.4]$ GeV/c^2;
  - $\Delta E \in [-0.3, 0.3]$ GeV;
  - $m_{K\gamma\gamma} \in [4.5, 6]$ GeV/c^2.

4.4.2 Background ntuples

The backgrounds listed in Section 4.3 have been generated. Table 4.1 lists the number of background events that have been generated and that have passed the pre-selection.

4.4.3 Signal ntuples

Different signal mass samples have been produced. For each of them, a ntuple has been created, obtained from $\sim$ 30000 (30×10^3) initial events (before applying the pre-selection, that has almost no effect on the signal). Each signal has been produced with a small intrinsic width, 1 MeV/c^2, so the final signal width is due to the detector resolution.
The generated and analyzed mass points are the following (in GeV/c^2):
0.15; 0.30; 0.50; 0.75; 1.0; 1.5; 2.0; 3.0; 4.0; 4.5.

Process	**Generated** ($\times 10^7$)	**Passed pre-selection**
B^+B^-	5.59	14’352
$B^0\bar{B}^0$	5.79	7’831
uds	5.50	1’717’570
$c\bar{c}$	3.56	410’379
Bhabha	6.62	110’403
$\tau^+\tau^-$	5.87	287’549
$\mu^+\mu^-$	7.61	1’944

Table 4.1: Number of generated background events and the corresponding number after the application of the pre-selection.

These values have been chosen to span over the whole mass range allowed by the kinematic constraint, i.e. up to 4786 MeV/c^2 (see Equation 4.1). The higher density at low mass values is because we expect a higher background there; mass point close to the π^0 and η resonances have been produced to check the performances of the selection in those zones; the final mass point at 4.5 GeV/c^2 is present to verify the performances close to the end of the spectrum.

4.4.4 Signal Broadened ntuples

To avoid the introduction of biases toward a particular ALP mass value, a signal sample with a *flattened* mass distribution has been used for the training of the discrimination procedure; more details will be presented in Section 5.2.3. The principle to obtain this flattened mass sample is to merge together a number of single signal *broadened* mass samples.

As earlier mentioned, the natural intrinsic ALP width in the MC signal samples is 1 MeV/c^2.

Another set of MC signal samples, independent from the previous ones, have been generated. The intrinsic ALP width for these samples is 100 MeV/c^2 if the central mass of the ALP is less than 1 GeV/c^2, otherwise the it is 200 MeV/c^2. Figure 4.2 shows the ALP mass distribution for some of the natural samples and for the broadened samples with the same central mass. All the following plots have been obtained selecting only events where the B candidate was Truth-Matched (i.e. corresponded to a MC B meson).

The number of broadened signal samples that have been generated is large enough to ensure a complete coverage of the entire allowed ALP mass spectrum. 30 samples have been generated: 5 samples have been produced in the range $[100, 300]$ MeV/c^2, with a constant separation step of 50 MeV/c^2 between them; 8 samples have been produced in the range $[300, 1000]$ MeV/c^2, with a constant separation step of 100 MeV/c^2 between them; finally, 19 samples have been produced in $[1.0, 4.6]$ GeV/c^2, with a constant separation step of 200 MeV/c^2 between them.

The step is different because smaller ALP masses have a smaller width, so it is necessary that the distributions are closer to each other to allow a sufficient overlay between them. Figure 4.3 shows the overlap present between some different mass points, necessary to allow the extraction of a sample with a flattened mass distribution.

These distributions are merged together, and the resulting distribution can be seen in

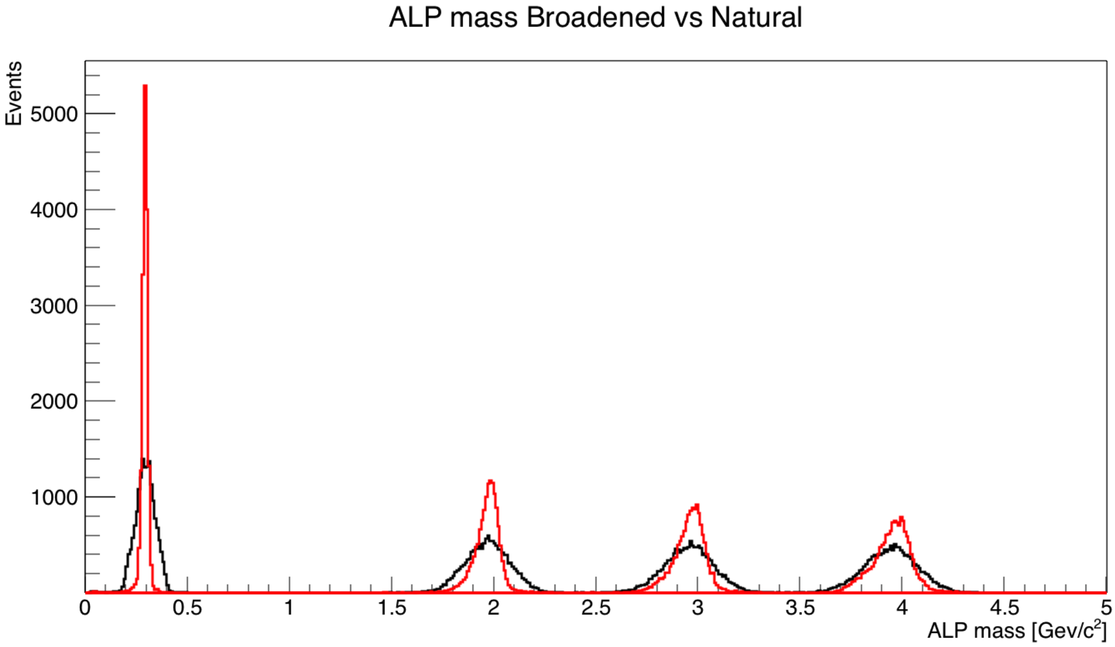


Figure 4.2: ALP mass distribution for the natural (red) and broadened (black) mass points of 0.3, 2.0, 3.0, and 4.0 GeV/c^2.

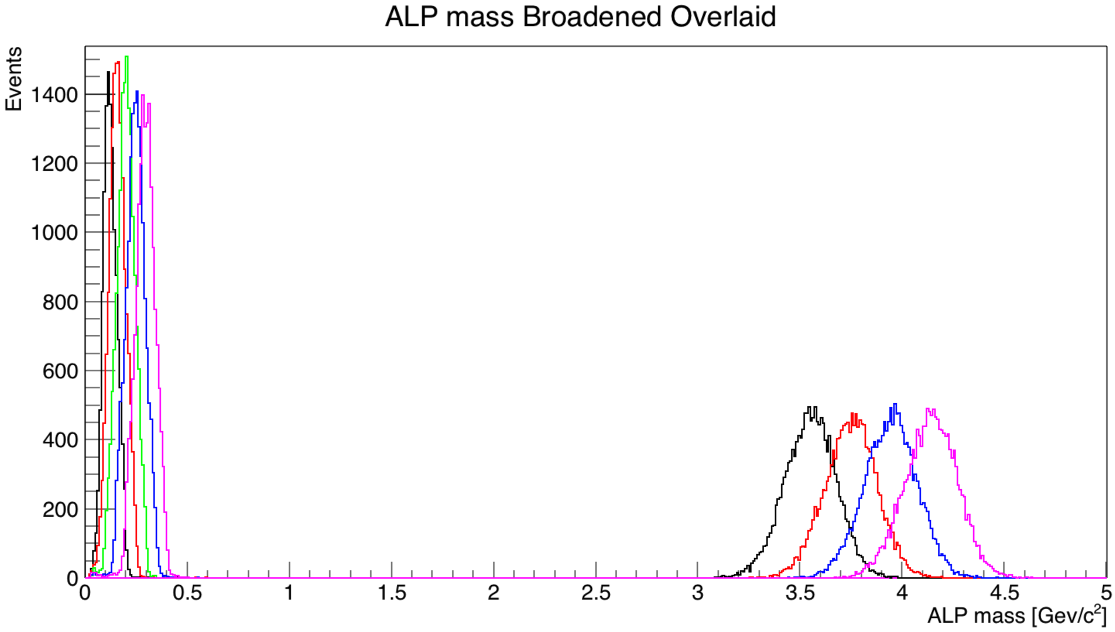


Figure 4.3: ALP mass distributions for some adjacent broadened signal mass points. The overlap is evident. The mass points of 0.10, 0.15, 0.20, 0.25, 0.30, 3.6, 3.8, 4.0, and 4.2 GeV/c^2 are shown.

Figure 4.4. From this distribution we want to extract a flattened distribution, and this is done randomly eliminating events from each mass bin until in average T events are left in each.

The details of the flattening procedure are the following: all the events of all the broadened signal samples are examined again, and for each of them the ALP mass is evaluated; given this

mass, the corresponding bin in the merged distribution is retrieved, as well as its population N. A random number uniformly distributed between 0 and N is generated, and if it is lower than T, then the event is saved, otherwise is discarded. This ends up generating a new sample with a flat ALP mass distribution, each bin being populated, in average, with T events.
I have set $T = 500$. The resulting ALP mass distribution can be seen in Figure 4.5. The corresponding events constitute the flattened mass signal sample used for the training of the selection procedure.

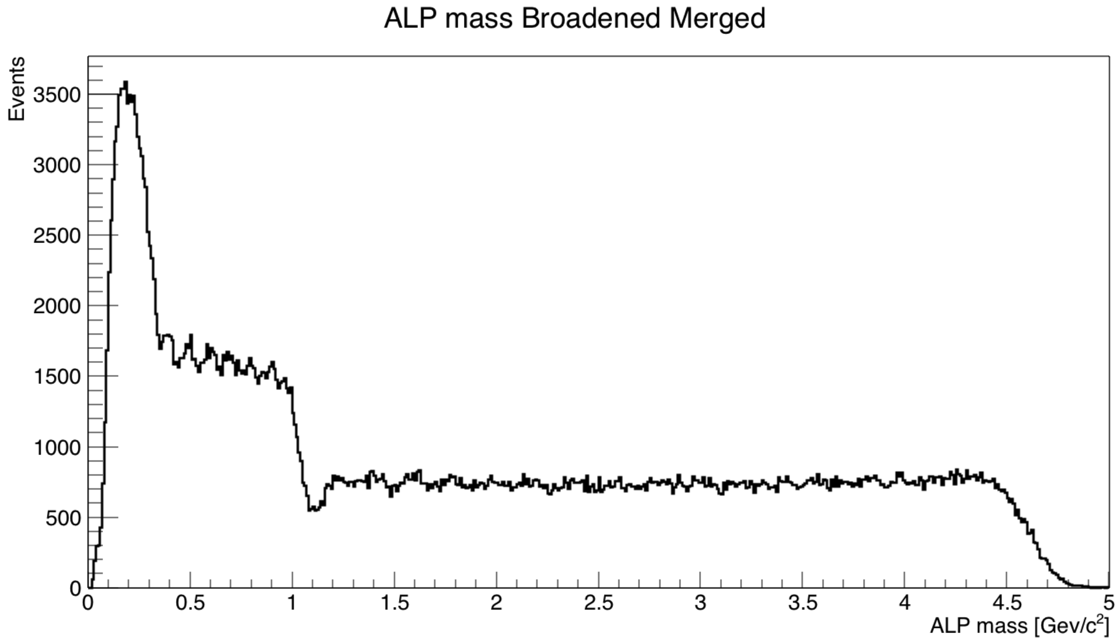


Figure 4.4: ALP mass distribution deriving from the merging of all the used broadened signal mass samples listed in the text.

4.4.5 Analysis and Control Sample

As mentioned in Section 2.2.1, *BABAR* performed 6 Runs at the $\Upsilon(4S)$ peak energy. The integrated luminosity of each Run, as well as the number of produced $B\bar{B}$ pairs, is reported in Table 4.2.

Both for the control sample and for the analysis dataset, we are interested only in on-peak data.
Control sample is constituted by Run 3 data. We hence have:

- 391.91 fb^{-1} of blinded integrated luminosity;
- 32.28 fb^{-1} of unblinded integrated luminosity (control sample, Run 3).

There are 501’277’316 events in Run 3 on-peak data before applying the pre-selection; the control sample is composed of 3’528’336 events.

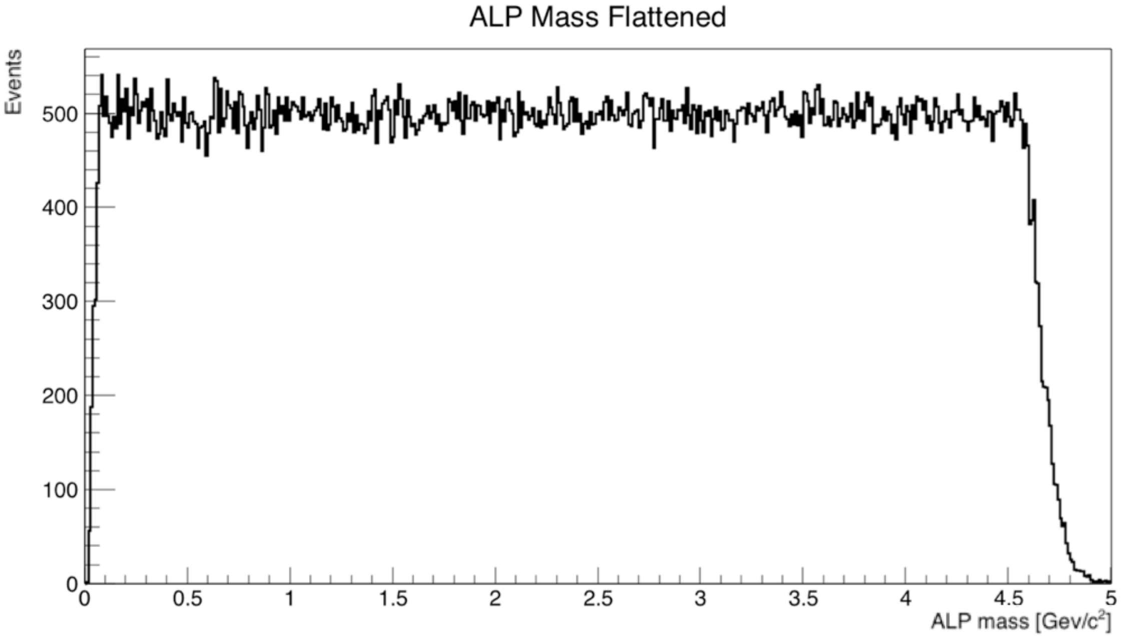


Figure 4.5: Flattened ALP mass distribution. The average population of each bin (that has a width of 10 MeV/c^2) is 500 events, except for the less populated tails.

Run	$\mathcal{L}_{on}$ (fb^{-1})	# $B\bar{B}$ **pairs** ($\times 10^4$)
1	20.37 ± 0.01 ± 0.09 (0.44)	2256 ± 3 ± 14 (0.62)
2	61.32 ± 0.01 ± 0.26 (0.42)	6844 ± 4 ± 41 (0.60)
3	32.28 ± 0.01 ± 0.13 (0.40)	3576 ± 3 ± 21 (0.62)
4	99.58 ± 0.02 ± 0.41 (0.41)	11143 ± 6 ± 67 (0.60)
5	132.33 ± 0.02 ± 0.59 (0.45)	14762 ± 6 ± 89 (0.60)
6	78.31 ± 0.02 ± 0.35 (0.45)	8519 ± 5 ± 51 (0.60)

Table 4.2: The on-peak integrated luminosities $\mathcal{L}_{on}$ of the individual $\Upsilon(4S)$ Runs of *BABAR*, and the number of produced $B\bar{B}$ pairs. For each entry, the first uncertainty is statistical, the second uncertainty is systematic, and the total relative uncertainty in percent is given in parentheses. The luminosities are from [36], the number of $B\bar{B}$ pairs is retrieved from the *BABAR* database.

Chapter 5

Event Selection

This Chapter provides a detailed description of the selection procedure performed to distinguish between the signal and the background in the $B^{\pm} \to K^{\pm}A$, $A \to \gamma\gamma$ process; a summary of this can be found in Section 4.1.

In Section 5.1 all the discriminating variables used in this analysis are listed, together with their descriptions and their distributions for the used training samples. Section 5.2 explains the characteristics of the tested selection methods, and the quantitative way to choose between them, while Section 5.3 details the procedure implemented to find the best performing method and shows its features.

5.1 Discriminating Variables

To properly perform the discrimination between signal and background, it is mandatory to study the features of the signal and of the important backgrounds to find a number of variables to use to recognize the first from the seconds.
The way these variables are used depends on the discrimination method actually implemented (see Section 5.2).

We have already seen the characteristics of the data samples in Section 4.4; this section is dedicated to explain each discriminating variable that has been used in the selection procedure.

For each variable its description is provided, as well as its distribution for the flattened signal sample (described in Section 4.4.4), and for the two backgrounds, uds and BpBm. The training has been performed only against these two sources of background because, as explained in Section 4.3, the Bhabha, Mumu, and Tautau backgrounds are negligible with respect to the others. Also, the characteristics (shape and kinematic variables) of the B0B0bar events are very similar to the ones of the BpBm sample; in addition, B0B0bar does not show SM channels that exactly mimic the signal (because we are only searching for decays of charged B in charged K), so we expect this background to be very small. Finally, even if ccbar background presents some differences with respect to the uds because of the higher mass of the c quark with respect to the u, d, and s, the two samples are similar to each

other when compared to the signal, hence an MVA trained to recognize uds has a similar discriminating power also against ccbar.
A signal event is considered TM only if the reconstructed B candidate corresponds to a MC generated B meson. Only TM signal events have been used to train the selection procedure, as we do not want it to take into account the combinatorial component of the signal. Coherently, only the TM signal event of the flattened sample are reported in the following plots.

For each of the three samples, the variables distributions have been normalized to 1, hence the ordinate axes show arbitrary units (A.U.). This has been done to take into account the different populations of the training samples, as here I am interested in showing the different shapes of the distributions, to underline their discriminating power.

5.1.1 m_{ES}

As claimed in Section 2.1.1, m_{ES} (also referred to as *mes*) is the beam energy substituted mass of the B meson candidate. As its characteristics have already been explained, here only its distributions for the training samples will be shown.

The distribution of m_{ES} for the training samples can be seen in Figure 5.1. As expected, the signal and the BpBm background have a peak at the mass of the B meson, while the uds background has a flat distribution with a kinematical limit. The BpBm background is less peaked than the signal as it contains also combinatorial background, while only TM signal events are plotted.

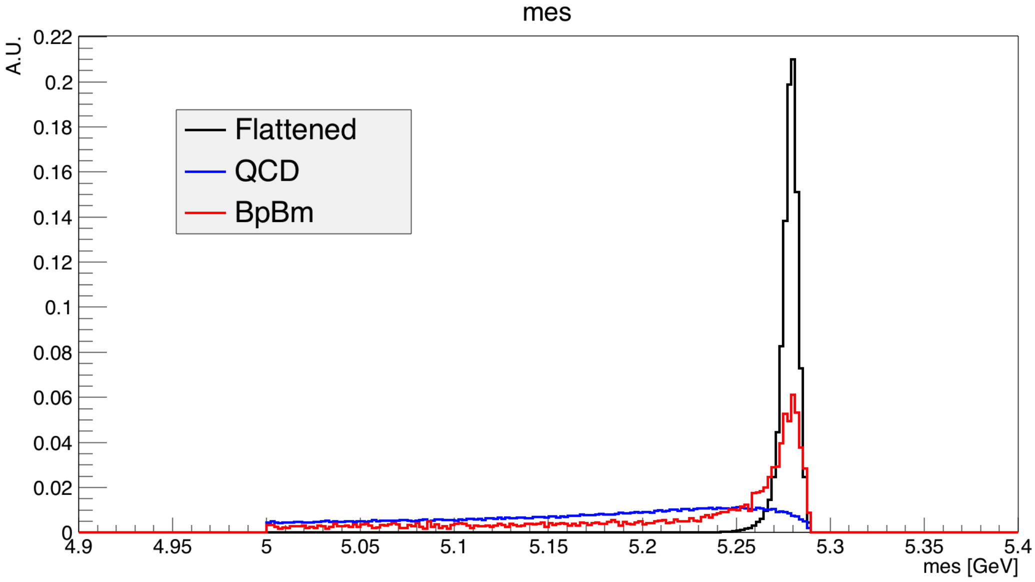


Figure 5.1: Distributions of m_{ES} for the flattened mass signal sample and the two training background samples, uds and BpBm.

5.1.2 ΔE

Similarly to m_{ES}, also the ΔE variable (also referred to as *DeltaE*) has already been discussed in Section 2.1.1, so here only its distributions for the training samples will be shown.

The distribution of ΔE for the training samples can be seen in Figure 5.2. Differently from m_{ES}, the BpBm background is not peaked at 0, as the signal is. This is because ΔE depends on the energies of the particles composing the B candidate, hence a misidentified particle would make its value different from 0, as the energy is computed from the hypothesized mass and the reconstructed momentum.

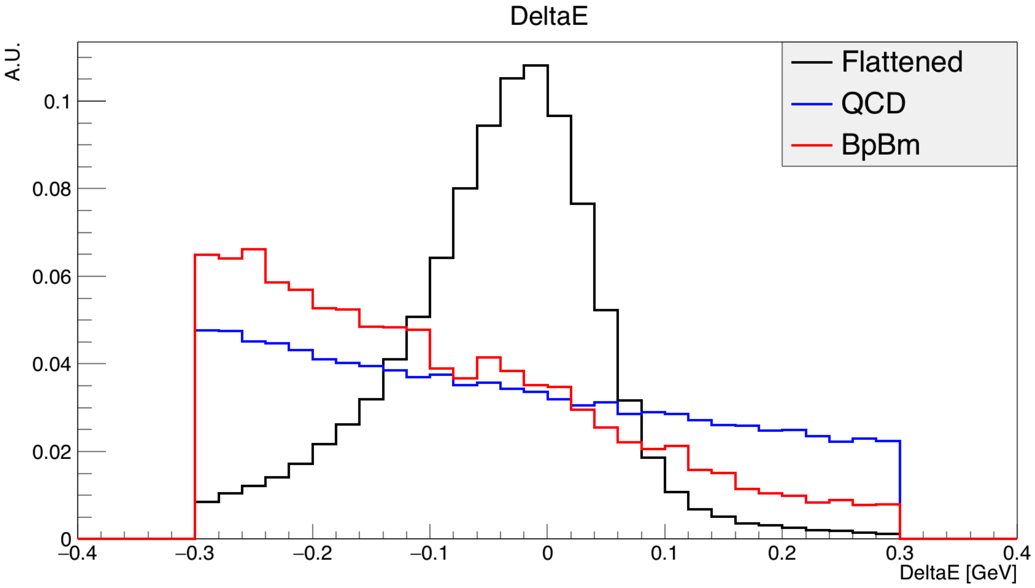


Figure 5.2: Distributions of ΔE for the flattened mass signal sample and the two training background samples, uds and BpBm. The effects of the pre-selection (see Section 4.4.1) are clearly visible here.

5.1.3 Helicity Angle of K

In a decay chain like $A \rightarrow B + C$, $C \rightarrow D + F$, it is possible to define the *helicity angle* for the particles D as the angle between the direction of the particle D and the direction of the particle A, measured in the rest frame of C. A graphic example can be seen in Figure 5.3, where θ_h indicates the helicity angle of D.

This is a quantity that can provide information about the angular properties of the particles involved in the decay: if the particle C is scalar, the distribution of the cosine of its helicity angle is flat; if the particle C has spin, the distribution can be modified according to selection rules.

In this analysis we have included the helicity angle of the kaon candidate (with A identified with the $\Upsilon(4S)$, C as the B candidate, and D as the kaon). We expect it being distributed as a cosine, because the B meson has spin 0, while the same could not be true for combinatorial backgrounds.

The distribution of the helicity angle of K (also referred to as *KHelicityAngle*) for the training samples can be seen in Figure 5.4.

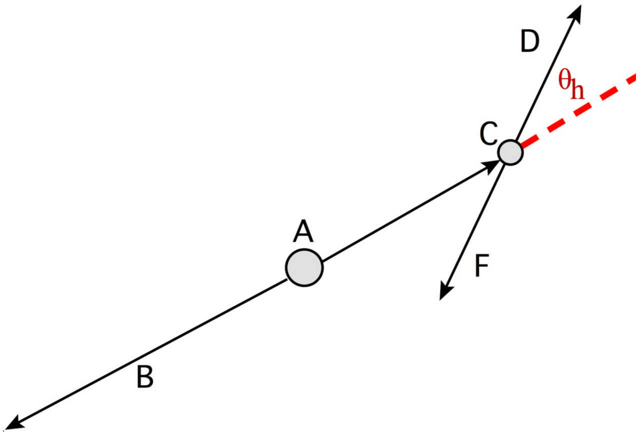


Figure 5.3: Graphic example, with the same names adopted in the text, showing the directions used for the definition of the helicity angle θ_h of the particle D. Image from [37].

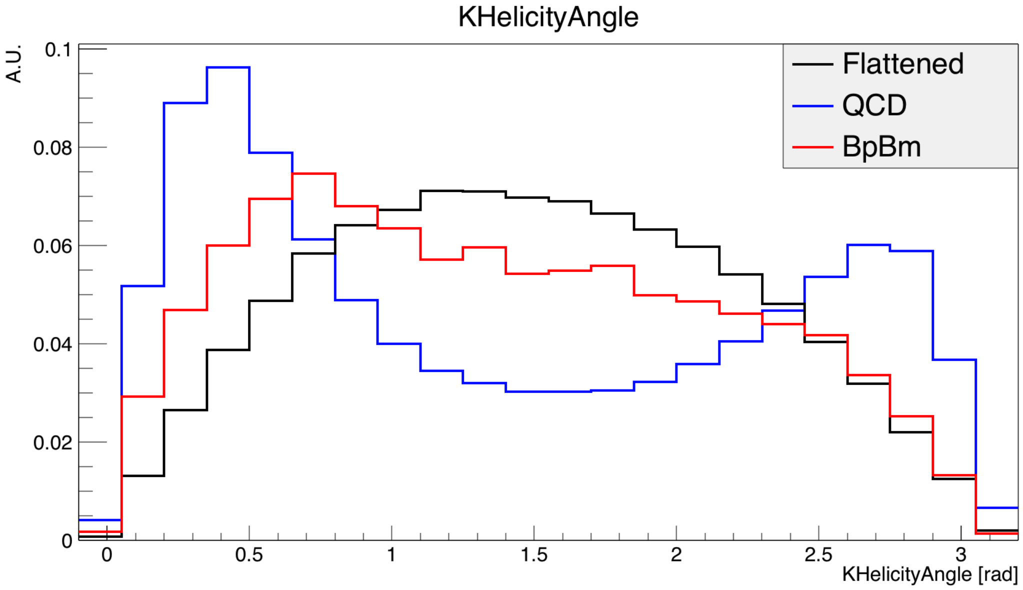


Figure 5.4: Distributions of the helicity angle of K for the flattened mass signal sample and the two training background samples, uds and BpBm.

5.1.4 Thrust Angle

Given a collection of N particles, the thrust axis $\vec{T}$ is the unit vector along which the projection of the total momentum of the group of particles is maximized.

The *thrust angle* θ_T is the angle between the thrust axis of the B candidate (a B candidate here is always defined as the combination of two photons and a kaon) and the thrust axis of the rest of the event (ROE), i.e. all the particles not used to build the B candidate under examination; all of these quantities are evaluated in the rest frame of the $\Upsilon(4S)$.

For a $B\bar{B}$ event the B mesons are produced almost at rest in this frame, so their decay products are isotropically distributed, the corresponding thrust axis is uniformly distributed and hence the distribution of $\cos(\theta_T)$ if flat. This does not stand true for quark-antiquark events, that have a jet-like shape due to the high boost they present, so the thrust axes of the B candidate and the one of the ROE are highly collimated, and thus peaks at ±1.

The distribution of $\cos{(\theta_T)}$ (also referred to as $CosThetaT$) for the training samples can be seen in Figure 5.5.

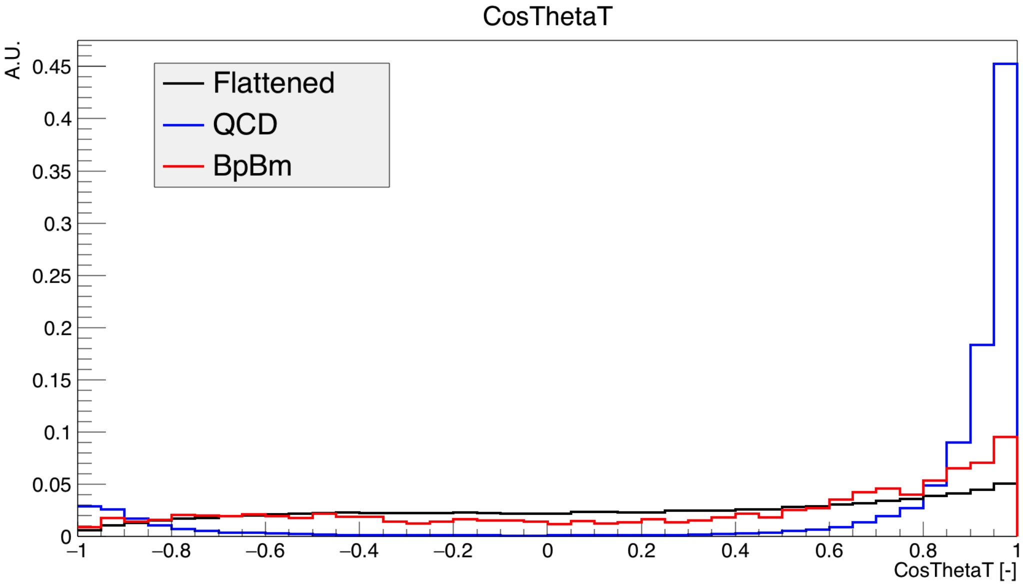


Figure 5.5: Distributions of $\cos{(\theta_T)}$ for the flattened mass signal sample and the two training background samples, uds and BpBm.

5.1.5 Sphericity Angle

The sphericity tensor is given by:

$$S^{\alpha,\beta} = \frac{\sum_{i=1}^{N} p_i^{\alpha} p_i^{\beta}}{\sum_{i=1}^{N} |\vec{p_i}|^2} \tag{5.1}$$

where $\vec{p_i}$, $i = 1, ..., N$ are the momentum of the N particles of the group under examination (the ones constituting the B candidate or the ROE), and $\alpha, \beta = x, y, z$ coordinates. This tensor has 3 eigenvalues λ_j and 3 corresponding eigenvectors. If the distribution of the decay products is isotropic, the eigenvalues have all similar magnitude; if the distribution is planar, one of them is significantly smaller than the others; if the distribution is jet-like, so very directional, two of the three are negligible with respect to the third.
The *sphericity axis* is the unit vector of the eigenvector with the highest eigenvalue, and similarly to the case of θ_T, the *sphericity angle* θ_S is defined as the angle between the sphericity axes of the B candidate and of the ROE. This quantity is conceptually similar to θ_T, indeed the reasoning done about the shape of its distribution depending on the kind of event (B decay or jet-like) are basically the same.

The distribution of $\cos{(\theta_S)}$ (also referred to as $CosThetaS$) for the training samples can be seen in Figure 5.6. Although $\cos{(\theta_T)}$ and $\cos{(\theta_S)}$ provide similar information, it has been empirically observed that including both of them actually improves the discriminating power.

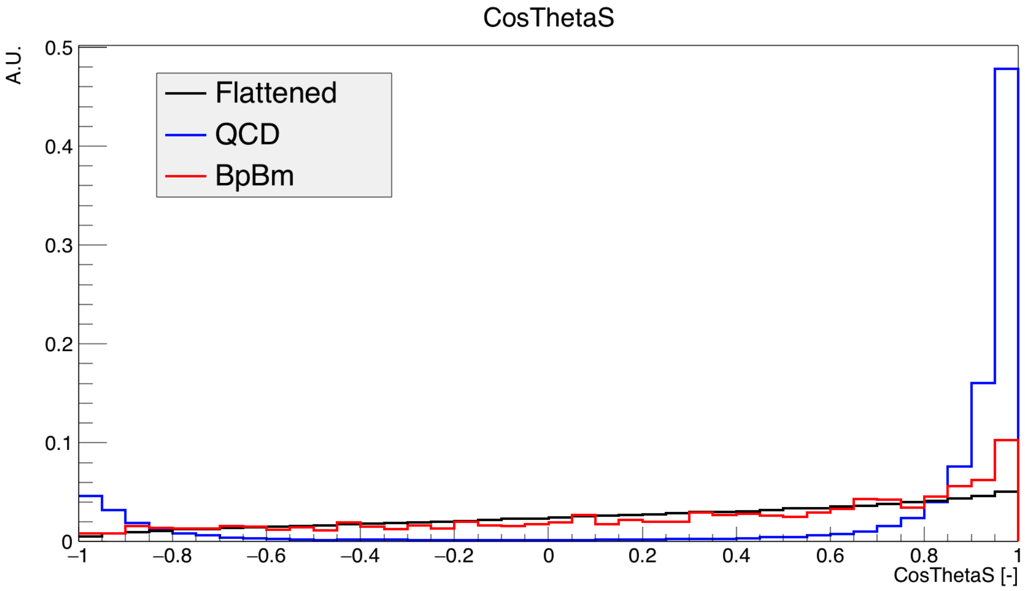


Figure 5.6: Distributions of $\cos(\theta_S)$ for the flattened mass signal sample and the two training background samples, uds and BpBm.

5.1.6 Legendre Moments

The Legendre moment of order k is defined in the following way:

$$L_k = \sum_{j \in ROE} p_j \cos^k(\theta_j) \tag{5.2}$$

where p_j is the magnitude of the momentum of the jth particle being part of the ROE, and θ_j is the angle between the B thrust axis and $\vec{p_j}$.

The distribution of $L0$ [$L2$] (also referred to as *Leg0* [*Leg2*]) for the training samples can be seen in Figure 5.7 [5.8]. The Legendre moments provide information about the sphericity of the event, and peak at higher values for jet-like events. Even if $L0$ alone is poorly discriminant, is has been observed that including it improves the performances of the selection procedure, thanks to its correlation with $L2$.

5.1.7 Maximum Selector of K

Particle lists are a way to categorize particle candidates according to PID criteria: a particle list can be considered as a group of particles that have satisfied some PID requirement. For each of the five stable charged particle (e, μ, π, K, p) there exist multiple lists, each requiring more or less strict requirements to allow a particle to be considered part of that list.
For each kind of charged particle there exist multiple methods (called *selectors*) to classify a particle candidate, and each of these selectors has its set of lists. A particle candidate can in principle enter any of these lists, even ones related to different kind of charged particles.
The purpose of this section is not to go into the details of the definitions of the various *BABAR* particle lists, but make the reader aware of their existence and their working principle.

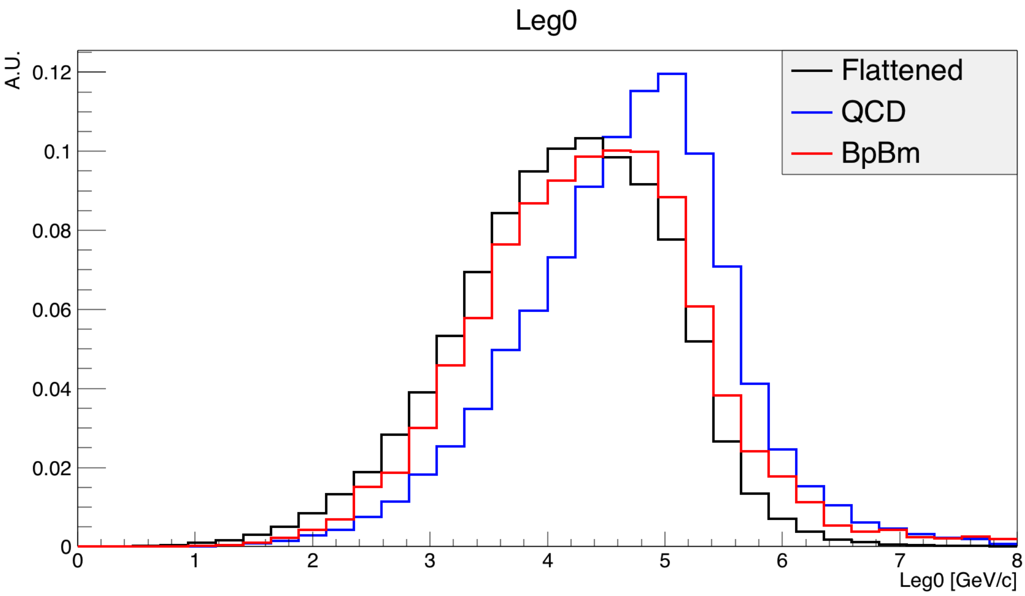


Figure 5.7: Distributions of *Leg*0 for the flattened mass signal sample and the two training background samples, uds and BpBm.

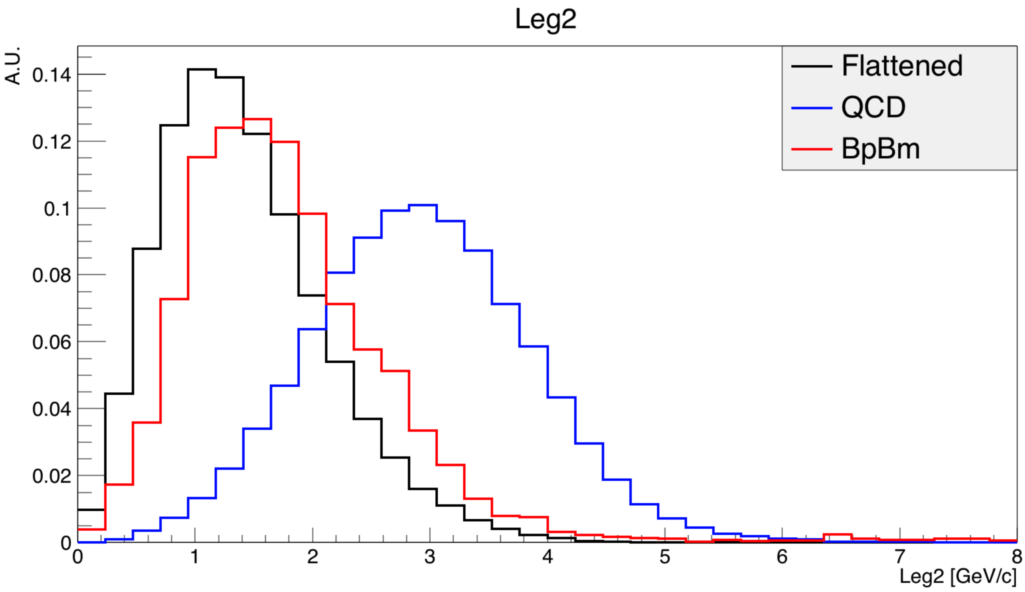


Figure 5.8: Distributions of *Leg*2 for the flattened mass signal sample and the two training background samples, uds and BpBm.

In this analysis I only use the lists of the kaons, to operate a quality PID requirement on the K candidate and reducing contaminations from other kind of particles. To do so, I have used the most recent selector for the kaons, that is called *PidKMSelector* [35].
This method offers six lists: SuperLooseKMKaonMicroSelection, VeryLooseKMKaonSelector, LooseKMKaonSelector, TightKMKaonSelector, VeryTightKMKaonSelector, SuperTightKMKaonSelector.
As the name suggests, particles in the SuperTight list have a very high probability of actually

being kaons (high *purity*), but at the same time, if we only use particle candidates in that list, we will lose some true kaons that have not been able to pass the strict quality requirements to enter that list (low *efficiency*).
If a particle is part of a certain list, it is also part of all the looser lists, so is not possible for a particle candidate being part of the VeryTight list and not being part of the Tight list, and so on.

The following assertions are generally true for all the five kinds of particle lists, while the numerical details are specific for the kaon lists.

Each list has an integer positive number associated to it; for the kaons, these numbers go from 5 to 29. The first 23 lists are associated with the older selectors, so we are not interested in them; lists numbered from 24 to 29 correspond to the used selector, and a higher number corresponds to a tighter list (so the SuperTightKMKaonSelector list is the list number 29).
To evaluate the likelihood of the kaon candidate of actually being a kaon, I have used a variable that is the number associated to the tightest lists of the PidKMSelector in which the kaon candidate has entered (referred to as *MaxKSelectorsOfK*); for example, for a kaon candidate which has entered up to the Tight list, this variable gets the value of 27. If a candidate has entered not even the SuperLoose list, the variables assumes the default value of 22 (22 and 23 would be related to another selector, but we ignore it).

This variable is currently the only one that is not included in the MVA as a discriminating variables but is instead used as a cut before applying the MVA: we only accept events where the kaon candidate is part of the tighter list, the SuperTightKMKaonSelector [35], so only if:

$$MaxKSelectorsOfK = 29 \tag{5.3}$$

The distribution of *MaxKSelectorsOfK* for the training samples can be seen in Figure 5.9. The choice of using this variable as a cut has been motivated by the observation that its inclusion in the MVA deteriorated the discrimination performances, as some MVAs can have issues with discrete variables.

5.1.8 Photon Veto

The Photon Veto (also referred to as *Pi0Veto*) is a variable created to reduce the photonic contamination from real π^0, η, and $\eta\prime$.
To compute this variables it is necessary to run over all the possible pairs of photons in the event and calculate their invariant mass: given N photons in the events, $\frac{N^2-N}{2}$ couples do exist, and the invariant mass of each of them is compared with the mass of π^0, η, and $\eta\prime$: if the difference is low enough, the two photons constituting the couple are marked. The minimum accepted distance from the central mass of π^0, η, and $\eta\prime$ is, respectively, 25, 30 and 35 MeV/c^2.
An ALP candidate is made up with two photons, γ_1 and γ_2; if one or both of the these two photons are marked, then the ALP candidate's *Pi0Veto* variables assumes the value of 1; otherwise, if neither γ_1 nor γ_2 are marked, *Pi0Veto* gets the value 0.

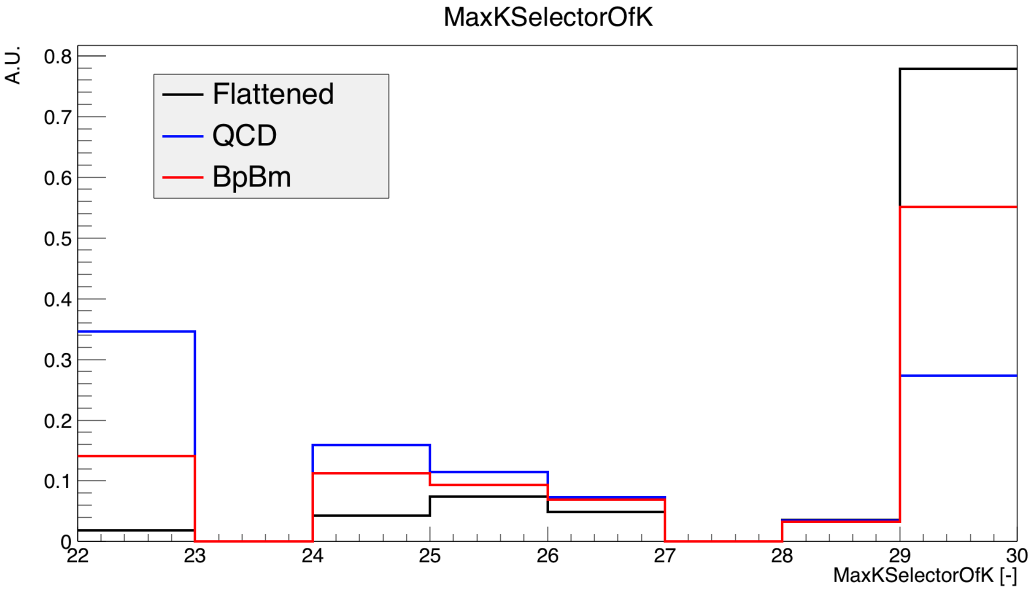


Figure 5.9: Distributions of *MaxKSelectorsOfK* for the flattened mass signal sample and the two training background samples, uds and BpBm. The empty bin at 27 is due to the fact that the Tight list has requirements very similar to the VeryTight list, so if a particle is able to enter the Tight list, it is very likely to enter the VeryTight too, hence the corresponding *MaxKSelectorOfK* becomes 28.

Given the high number of photons in an event, it is quite frequent that at least one among γ_1 and γ_2 can be coupled with another photon to make an invariant mass close to one of the three resonances we are vetoing. This happens more frequently for the backgrounds than for the signal, so having a value of 0 in this variable is an indication of a signal event, while a value of 1 does not provide much information.

The distribution of *Pi0Veto* for the training samples can be seen in Figure 5.10. Despite its discrete nature and apparently low discriminating power, it has been observed that the inclusion of this variable in the MVA does not produce any issue and improves the discrimination performances.

5.1.9 Photons Energies

The energies of the two photons γ_1 and γ_2 constituting the ALP candidate have been included. The usage of the signal flattened mass sample is meant to avoid to bias the training over this kind of variables, that are mass-dependent.
The two energies have been categorized simply as "High" and "Low" for the higher and the lower ones, computed in the laboratory frame.

The distribution of the energy of the most [least] energetic photon among the two (also referred to as *GammaHighEnergy* [*GammaLowEnergy*]) for the training samples can be seen in Figure 5.11 [5.12].

GammaHighEnergy alone is not a powerful discriminant, but it has been seen that including it that improves the performances of the selection procedure, because of its correlation

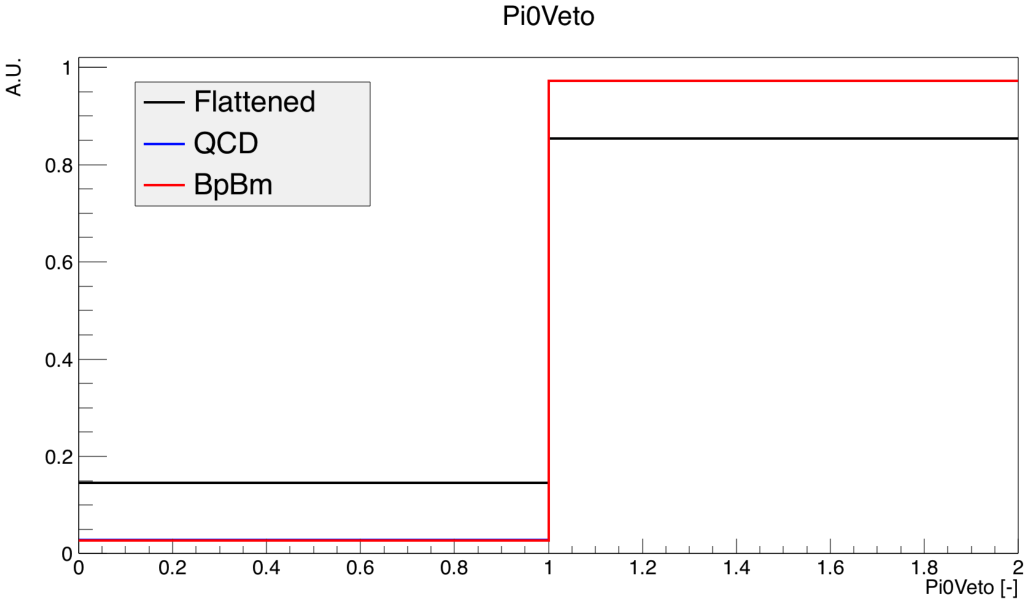


Figure 5.10: Distributions of *Pi0Veto* for the flattened mass signal sample and the two training background samples, uds and BpBm. The distributions of the two backgrounds are overlaid.

with the other variables (see Figure 5.18).

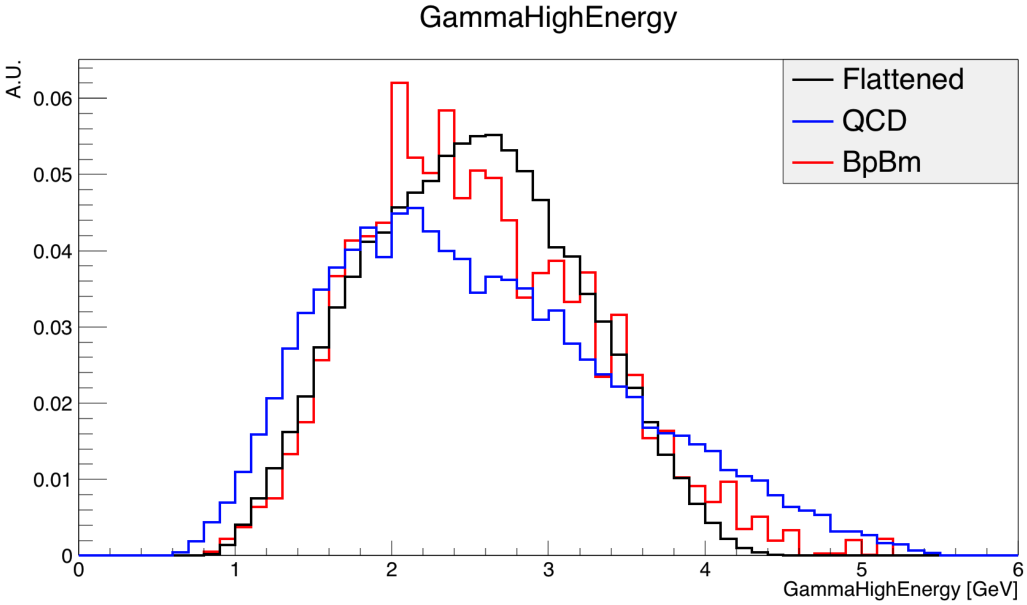


Figure 5.11: Distributions of *GammaHighEnergy* for the flattened mass signal sample and the two training background samples, uds and BpBm.

5.1.10 Photon Helicity Angle

The helicity angle (see the definition in Section 5.1.3) of the most energetic photon among the two constituting the ALP candidate is used in the selection. In this case A is identified with the B meson candidate, C is the ALP candidate, and D is the most energetic photon.

The distribution of the helicity angle of the photon (also referred to as *GammaHighHe-*

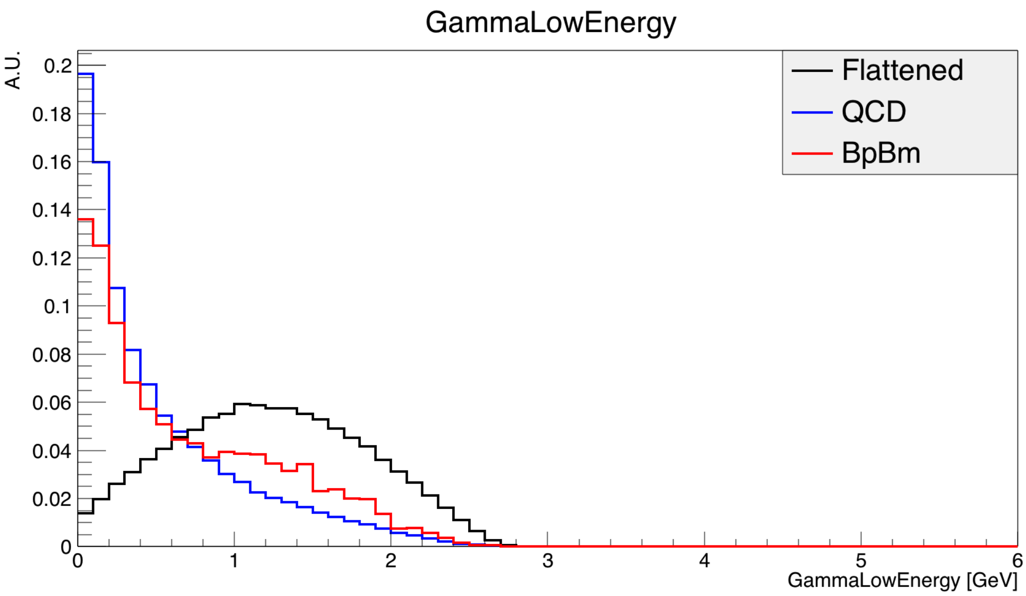


Figure 5.12: Distributions of *GammaLowEnergy* for the flattened mass signal sample and the two training background samples, uds and BpBm.

licityAngle) for the training samples can be seen in Figure 5.13.

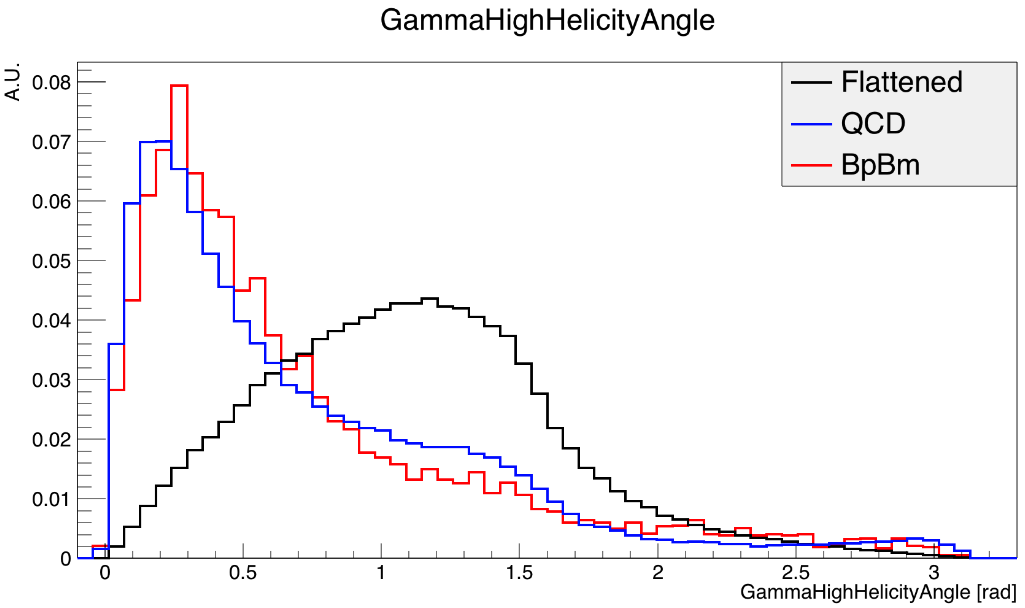


Figure 5.13: Distributions of *GammaHighHelicityAngle* for the flattened mass signal sample and the two training background samples, uds and BpBm.

5.2 Selection Criterion and Methods

The purpose of the selection procedure is to maximize the number of signal events while minimizing the number of background events, exploiting the features of the discriminating variables.

Two useful quantities are the *purity* p and the *efficiency* ϵ. Calling S and B the number of signal and background events that have passed the selection, these two quantities are defined as follows:

$$p = \frac{S}{S+B} \tag{5.4}$$

$$\epsilon = \frac{S}{S^{gen}} \tag{5.5}$$

where S^{gen} is the number of produced signal events. Applying a tighter selection implies to increase the purity p at the expense of the efficiency ϵ, and vice versa.

In this section the quantitative criterion to choose the best performing selection is described, and the working principles of the tested selection methods are explained.

5.2.1 Figures of Merit (FoM)

The goal of a selection is to maximize S (so maximize ϵ) while minimizing B (so maximize p). It is generally impossible to maximize p and ϵ at the same time, so a compromise is required: this is the maximization of the expected value of a properly defined Figure of Merit (FoM).

If we are interested in observing a new effect, we are generally willing to optimize over the significance of this over the known phenomena, here behaving as backgrounds.

The number B of background events passing the selection can be assumed distributed with a Poisson distribution, hence the fluctuation of B is $\sigma_B = \sqrt{B}$. In case we observe an excess of events with respect to the expected ones, to be able to claim the discovery of a new phenomenon we want to be reasonably sure that the difference is not caused by a statistical fluctuation. So, in a region where we expect a number B of background events, we have to compare the number S of passed signal events, obtainable from the number N of total observed events as $S = N - B$, with the fluctuation of the background itself; in formula, we want to maximize the ratio:

$$G = \frac{S}{\sqrt{B}} \tag{5.6}$$

If we are instead interested in performing precision measurement of a known phenomenon, the FoM to maximize is slightly different, because the signal strength is known as being different from zero, so the number of expected events in the studied region is $N = B + S$; we want therefore to maximize the significance of the measurement, that is:

$$G = \frac{S}{\sqrt{B+S}} \tag{5.7}$$

The difference arises from the fact that the fluctuation of the number of events is now $\sqrt{B+S}$ and no more only $\sqrt{B}$, because the signal itself contributes to the amount of observed events.

This analysis aims to look for an unknown process, so the case we are interested in is the first. If we explicitly write down the expression for $\frac{S}{\sqrt{B}}$, we can take out some constant factors that are irrelevant for the maximization procedure.

Being $G = S/\sqrt{B}$ the FoM we want to maximize, we need to connect the number of

signal and background events we can obtain in real data, after the selection procedure, with the MC signal and background events that have been generated and that have passed the selection. We denote N_S^{gen} the number of generated MC signal events and N_S^{pas} the number of MC signal events that have passed the selection procedure; a similar notation goes for the background events, with the only difference that there is more than one source of background, so each of them is labeled with an index j.
We can hence write the following:

$$S = \sigma_S \mathcal{L} \epsilon_S = \sigma_S \mathcal{L} \frac{N_S^{pas}}{N_S^{gen}} \tag{5.8}$$

and

$$B = \mathcal{L} \sum_j \sigma_j \epsilon_j = \mathcal{L} \sum_j \sigma_j \frac{N_j^{pas}}{N_j^{gen}} \tag{5.9}$$

where σ_S and σ_j are the cross sections of the signal channel and of the jth background channel (see Section 4.3 for the list of backgrounds of interest) and $\mathcal{L}$ is the integrated luminosity of the analyzed data sample.
We henceforth have:

$$G = \sqrt{\mathcal{L}} \sigma_S \frac{N_S^{pas}}{N_S^{gen}} \sqrt{\sum_j \sigma_j \frac{N_j^{pas}}{N_j^{gen}}}^{-1} \tag{5.10}$$

Now we factor out the constant terms (i.e. the ones that are not dependent on the selection procedure) and redefine a new FoM, χ, directly proportional to G:

$$\chi \equiv N_S^{pas} \sqrt{\sum_j \sigma_j \frac{N_j^{pas}}{N_j^{gen}}}^{-1} \tag{5.11}$$

(The unit of χ is length^{-1}, but this is totally irrelevant for our purposes, so it will not be stated elsewhere)
So, the criterion to choose the best selection procedure is to take the one that provides the highest value of χ.

5.2.2 MultiVariate Analysis (MVA)

The MultiVariate Analysis is the statistical analysis of a problem characterized by many variables, that may or may not be correlated. There exist many MVA methods (some of them will be explored below), but all of them are characterized by some common requests, as the need for known and differently classified samples, and some common procedures.

The common idea behind every MVA method is to obtain a system able to classify the events provided to it by the user. From the point of view of high energy physics the interesting classifications are *signal* and *background*.
The MVA method (or simply MVA, where there is no ambiguity) has to learn to recognize the events and classify them as signal or background: to do so it is necessary to provide to the MVA two samples, one constituted only by signal events, and the other constituted only by background events. The other mandatory input is the list of *discriminating variables* that

we want to use to perform our discrimination. Operatively speaking, the user has to set the MVA parameters (depending on the method) and to provide the training samples and the lists of variables to be used; after that, the MVA learns how to differentiate the two samples, based on the distribution of the variables and their correlations.

There are two main steps in the learning procedure of an MVA: training and testing. During the training phase the MVA learns, in a way that depends on the actual method implemented, how to distinguish signal from background. The training output is a set of numbers (cuts and/or *weights*); the method, having these numbers, is now trained.
The next step is to test the trained MVA, using a statistically independent sample of events (both for signal and background): this is done to check if *overtraining* has occurred. Overtraining happens if the MVA has learned too well how to recognize the particular features of the training sample, i.e. if it was trained too much on the unavoidable statistical fluctuations present in the samples. In this case the MVA has an excellent discriminating power for what concerns the training samples, but does not work well on other ones. Overtraining is mainly due to two factors: limited statistic, or excessive deepness of the training. In the first case the sample is dominated by statistical fluctuations and the MVA learns to recognize them instead of the true features of the sample itself; in the second case, even if the number of provided events is sufficiently large, the training is so deep that the MVA still is able to grasp little fluctuations and, becoming too specialized in recognizing them, does not work well on another sample.
The testing step consists in running the MVA on the test samples (both for signal and background) and compare the results with the ones obtained from the training samples: if the output distributions (that depends on the method) are statistically not compatible, overtraining has occurred.

Below a description of the working principles of different MVA methods is provided, with particular attention to the ones tested for this analysis. More details can be found in [21, 38]. In Figure 5.14 there is a comparison scheme on the classification operated by different methods, in a 2-dimensional space of variables.

The analysis framework ROOT [39], used for this analysis, hosts TMVA [38], a toolkit which comprehends a large variety of MVA methods, and allows to perform all the procedures needed to train an MVA.

For this analysis, coherently with what is currently available at *BABAR*, I used ROOT version 5.34/10 and TMVA version 4.2.0.

5.2.2.1 Rectangular Cuts

The rectangular cuts is the simplest MVA method. The idea is simply to look at each variable and apply one or more cuts, defining one region in the N-D hyper-space of the N variables. Among the advantages of this method there is its intuitiveness and simplicity, but on the

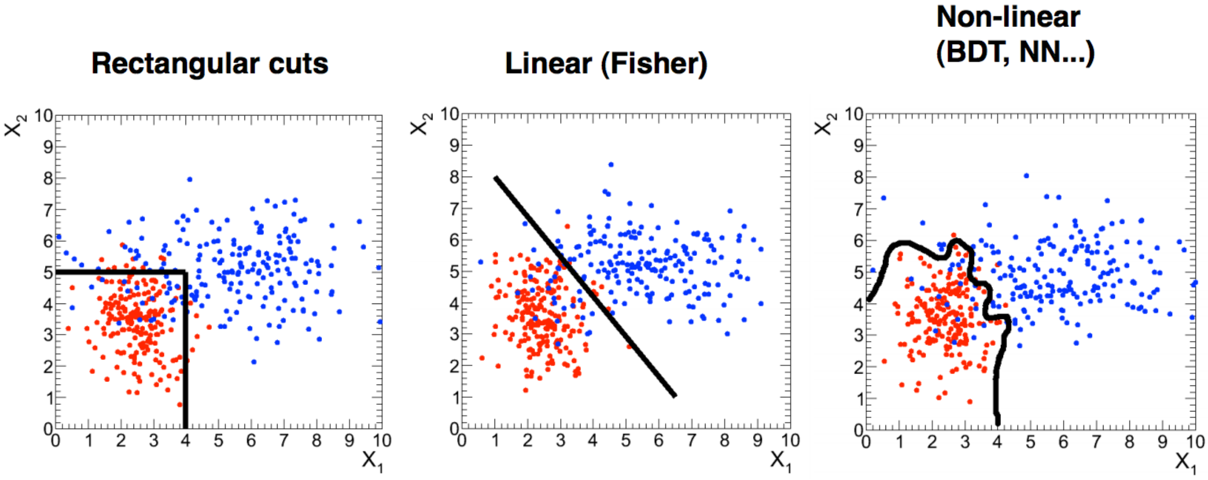


Figure 5.14: Effects of different MVA classification methods applied to events characterized by 2 variables. Different colors correspond to different classes of events (signal or background). The details of the procedures can be found in the dedicated sections. Image from [40].

other hand it suffers high-dimensional problems: exploring a highly dimensioned space can be consuming in terms of computing times; also it is inefficient when the variables are correlated.

A 2D example of the application of the rectangular cuts method can be seen on the left side of Figure 5.14.

5.2.2.2 Fisher Linear Discriminant

The principle at the base of a linear discriminant is to look at the distance of the mean values of the distributions of the two samples. If this is done without proper preparations, little discrimination power could be achieved, because the projections of the distributions on the variables' hyper-planes can largely overlap. A Fisher linear discriminant aims to enhance the separation power projecting the classes' distributions on a particular axis, reducing dimensionality while preserving as much class information as possible; this axis is defined such that the distance between the projected means of the two distribution, weighted with their internal spread, is maximized, while the internal spread itself is minimized; a 2D example can be seen in Figure 5.15.

The final purpose is to obtain a scalar in the form of:

$$S = A + \vec{B} \cdot \vec{x} \tag{5.12}$$

that is the projection of the vector $\vec{x}$ of the event variables on the vector $\vec{B}$, plus possibly an offset A.

Linear discriminant can have high performances until the discrimination power is held by the means of the distributions and not by their variance. Despite being relatively simple, it presents quite good results until the correlations are linear, and it is particularly efficient when the input variables are Gaussianly distributed.

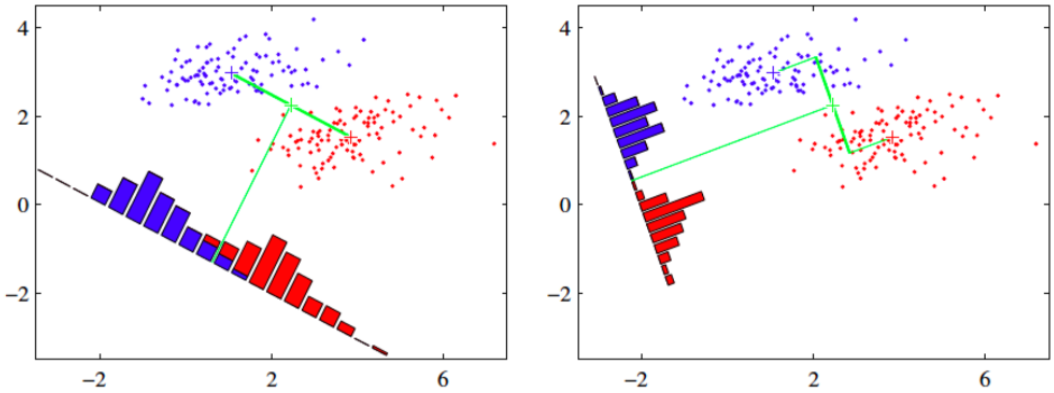


Figure 5.15: Example of linear discriminant method applied on a 2D variable space: on the left, the two distributions are projected on an axis parallel to the segment between their means; on the right, they are projected on the axis that induces the projection such that the distributions are pushed as far as possible away from each other, while events of a same class are confined in a close vicinity. Image from [41].

More details about the math involved in the actual calculation can be found in [42].

This method presents no parameters to be set by the user.

5.2.2.3 Multi Layer Perceptron (MLP)

Artificial neural network (ANN) are based on biological neurons behavior: roughly speaking, a neuron produces an output only if the weighted sum of the input exceeds a certain threshold. Taking inspiration from this, ANNs are constituted by a number of artificial neurons, called *perceptrons*, that are elements that receive k inputs, and if their weighted sum is sufficiently high, a constant output is fired. In practice the behavior is not discontinuous, but is smoothed by a step-like function; typical examples are the *sigmoid*:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{5.13}$$

or the hyperbolic tangent:

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{5.14}$$

In our case, x is the weighted sum of the inputs, in the form of:

$$x = \sum_{i}^{k} w_i s_i + c \tag{5.15}$$

where w_i are the weights, s_i the inputs, and c an offset.
The purpose of training an ANN is to find the best values of w_i and c such that the discrimination power is maximized.

Since the performances of a single perceptron are limited, every actual ANN is constituted by multiple perceptrons, that can be organized in different structures; one of the simplest, but nonetheless very powerful, takes the name of *multilayer perceptron* (MLP). The default configuration in TMVA is made by 3 layers of perceptrons: input layer, hidden layer, and output layer. This is a *feed-forward* NN because the information goes only and always forward and there are no feedback processes.

In the default configuration, the input layer is constituted by the k inputs, i.e. the discriminating variables, plus a constant term, a bias. The hidden layer can in principle be made by any number of perceptrons, the default number being $k+1$. The output layer consists in 1 or 2 simplified perceptron (in our utilization, we only have 1 output) that directly perform a linear combinations of the inputs (i.e. the outputs of the hidden layer) instead of processing that combination with a smoothed step function. In Figure 5.16 there is an example of the structure of a MLP.

A perceptron is also called a *node* of the net. After having properly trained a neural network, when the user provides it an event, the output is a scalar: the larger this scalar is, the more probable is that the event is a signal event. Performing a proper training consists in finding the optimal weights and biases used by all the nodes of the net, that is the set of values that minimizes the distance between the desired output Y_{exp} (e.g. 1 for signal, 0 for background) and the actual outputs of the MLP Y_{MLP}:

$$\Sigma = \sum_{n=1}^{N} \frac{1}{2} \left(Y_{exp} - Y_{MLP}\right)^2 \tag{5.16}$$

where N is the total number of events; Σ is a measure of the classification error rate.

MLP takes advantages from the correlation between the variables, and they are quite stable against overtraining and an increasing number of variables.

The main parameters that can be set for this method are:

- The kind of smoothed step function;
- Number of training cycles the MLP runs over to optimize the search of the weights;
- The number and structure of the hidden layer(s).

5.2.2.4 Boosted Decision Tree (BDT)

The working principle of a decision tree is simply to apply cuts in sequence to the discriminating variables, to separate in the best way possible signal and background events. A scheme depicting the behavior of a decision tree can be seen in Figure 5.17. At each step (node) the method looks at the events and searches the variable, and the cut to apply to it, such that the discrimination between signal and background is the highest: this generates two branches. Each of them now contains a subset of the initial set of events, and for each of them the procedure is repeated, so the variable and the corresponding cut that maximize

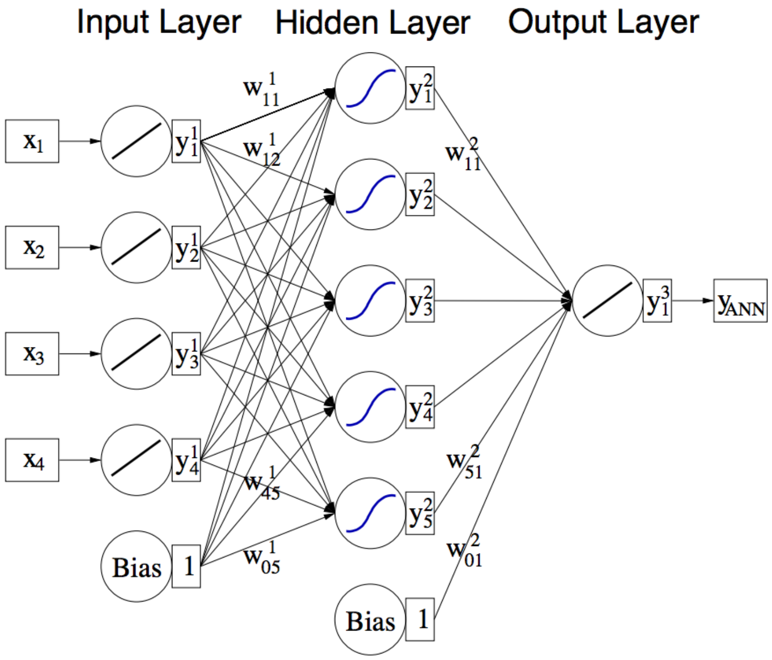


Figure 5.16: Example of the structure of a multilayer perceptron neural network. In this example there is 1 hidden layer, 1 output value, and 4 input discriminating variables. Both the input layer and the hidden layer have an additional bias value. Image from [43].

the separation power are searched. The procedure goes on until each terminal node (now called *leaf*) contains a minimum amount of events, or until the increment of a certain figure of merit, computed from the mean purity of the leaves, reaches a minimum threshold value. Each leaf is labeled as signal or background; subsequently every processed events undergoes the selection criteria, following a different path down the tree from the root to the leaves depending on the values of its variables, until it arrives to a leaf, and there it is classified as background or signal.

Decision trees can be seen as an evolution of the rectangular cut procedure, because they generate a number of hyper-cubes in the variables space, each of them classified as signal or background, exactly as rectangular cuts do, but the latter only generates one single labeled hyper-cube.

Decision trees can be, a priori, trained until they reach a perfect separation power, i.e. every single event is correctly classified as signal or background: this is absolutely to be avoided, because it is a clear example of overtraining, in the sense that the tree has been trained to recognize all the events of the training sample, but its performances on an another sample would be little better than a random choice, due to the excessive specialization it has on the statistical fluctuation of the training sample. Decision trees particularly suffer from overtraining, being highly sensitive to the characteristics of the training sample.

To overcome this issue, a *boosting* can be implemented: the principle of boosting is to gather many weak learners, as a single decision tree is, to create a strong learner. When applied to decision trees, this concept transforms them into a *forest*, and the method is called boosted decision tree (BDT). There are several boosting implementations, the one I used is the Adap-

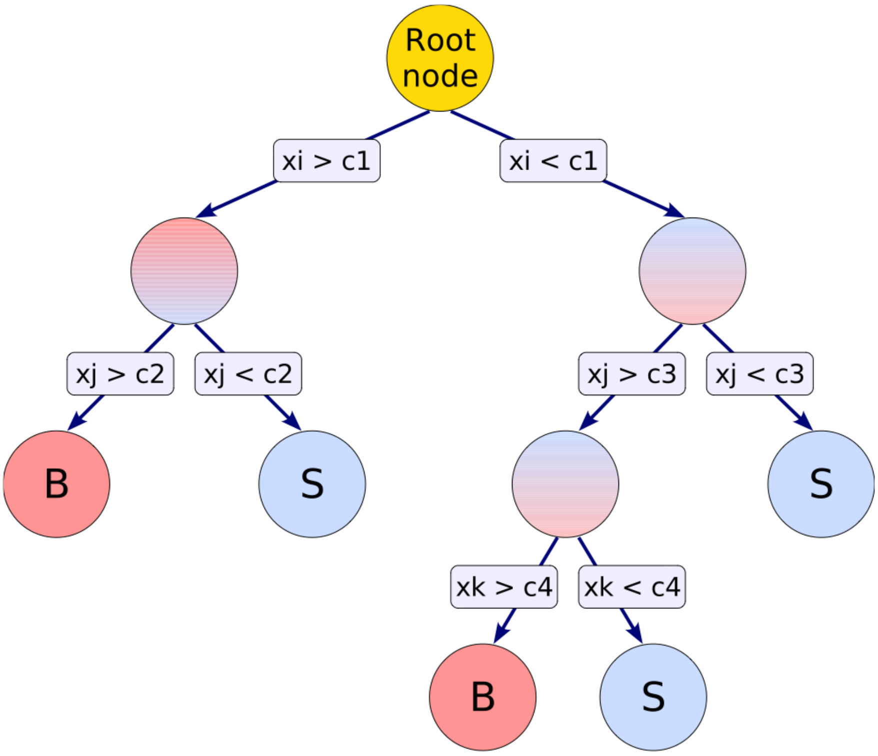


Figure 5.17: Example of the structure of a decision tree. Different colors mean different event classes; in the example 3 variables are taken into account. Image from [43].

tive Boost (AdaBoost), that is the TMVA default: the BDT starts with a normal single tree, whose training is not deep to leave it a weak learner and avoid overtraining, so few branches are developed. After the training of the first tree is completed, all the misclassified events are re-weighted, according to the error E committed by the tree:

$$E = \frac{\text{Misclassified events}}{\text{Total events}} \tag{5.17}$$

and a weight is defined as follows:

$$\alpha = \frac{1-E}{E} \tag{5.18}$$

After the re-weighting of the events, another tree is trained (using the entire sample), and the process is repeated, re-weighting again the misclassified events and so on. The process is repeated for an arbitrary number of trees. After the entire BDT has been trained, when an event is provided to it, a scalar output is returned: this value is obtained as a weighted sum of the outputs of all the trees, where the output of a single tree is binary in the two classes (i.e. -1 for background and +1 for signal) and the weight of each tree is the logarithm of the value α associated to it.

BDTs are very resistant to the inclusion of poorly discriminating discriminating variables, as the trees simply ignore them; also, the tuning of the optimal set of parameters is a task generally easier for a BDT than for a MLP, so, in many practical terms, they can behave better than more complex procedures as the ANNs.

The main parameters that can be set for this method are:

- The number of trees the forest is composed of;
- The max allowed depth of the single decision tree.

5.2.3 Training Strategies

Two training strategies have been taken into account to deal with the two main sources of background (uds and BpBm; I recall here that the training has been carried out only against these two categories of background): a *1-step strategy* and a *2-steps strategy*.
In both cases the used signal sample is the flattened mass sample described in Section 4.4.4, and this is done to avoid the introduction of any bias due to the training with a specific mass sample. This also allows to include some mass-dependent variables, like the energies of the daughters of the ALP candidate, as well as handling variables whose dependence from the ALP mass is not evident.

The 1-step strategy aims to train a *single* MVA to be able to discriminate between the signal and the two backgrounds at the same time. To do so it is also necessary to provide to the MVA adequate weights of the different sources of backgrounds, in this case uds and BpBm, taking into account the number of used ntuples and their cross sections, so that they represent the actual data.
This approach is potentially powerful, but at the same time is less stable, as finding the correct training parameters could result hard.

With the 2-steps strategy, instead, *two* different MVAs are trained: one is optimized to distinguish between signal and uds, and the other is optimized against BpBm. Both of them use the exact same training sample for the signal (the flattened one). In principle the set of discriminating variables could be different, so one could use only the most fitted variables for the two backgrounds separately, and this is a useful feature in case of utilization of an MLP, that is quite weak to poorly discriminating variables, while a BDT, as explained in Section 5.2.2, is resistant to variables of little use. The 2-steps strategy is more complex in terms of coding and is more expensive in terms of computational time; nonetheless, the easier calibration of the MVA parameters makes it actually competitive.

5.3 Choice and Training of the Selection Method

As explained in Section 5.2.1, the criterion to choose the best MVA method (Section 5.2.2) and the best strategy (Section 5.2.3) is to take the one that maximizes the quantity χ defined in Equation 5.11. In practice there are 6 candidates, because each strategy (1-step, 2-steps) can be carried out with any method (Fisher, MLP, BDT).
χ has been computed using BpBm and uds backgrounds only, so the ones used for the training.

5.3.1 Search of the Best Method

The procedure to choose the best performing training method and strategy is the following. The MVAs are trained, using the signal flattened mass sample (only TM events) and the two sources of backgrounds, uds and BpBm. As explained in Section 5.2.2, the training step is always followed by the testing step, meant to check if overtraining has occurred. These two steps are automatically performed one after the other in TMVA, that also provides plots to check the goodness of the training (see Section 5.3.2).
The possible cuts are applied on the ntuples before using them for training the MVAs. Currently the only applied cut is on *MaxKSelectorsOfK*, but all ntuples are deprived of the ALP candidates whose masses are close[1] to the ones of π^0, η, and $\eta\prime$, so the MVAs do not train to recognize them.

Each method is such that, after having been trained, if it is provided with an event, it returns a real value in a certain range (the range depends on the used method), and the higher the value is, the more probable is that the provided event is of signal kind. Our purpose is to find the cut(s) on the MVA(s) output(s) that maximize the quantity χ, so the next step is to perform a scan of the 1D or 2D space of the possible cuts (it depends if we are examining the 1-step or 2-steps strategy), apply these cuts to the known MC samples for signal and backgrounds and compute the corresponding χ, so to have a correspondence between the MVA cuts and the resulting χ. This is called the *evaluation* step.
The evaluation is performed on the original not-flattened mass signal samples (see Section 4.4.3), because we want to analyze the effects of the selection on the possibly real signal, and using only TM events. For what concerns the MC backgrounds, the evaluation is not performed on the same samples used for the training but on different, statistically independent samples, to ensure that no bias is introduced. Even during the evaluation step, the ALP candidate with a mass close to the ones of π^0, η, and $\eta\prime$ are rejected.
The signal and the uds and BpBm backgrounds events are provided to the trained MVA, that returns an output for each of them. If the output is greater than the set cuts, then the event is considered to have passed the MVA selection; the numbers N_S^{pas}, N_{uds}^{pas}, and N_{BpBm}^{pas} are saved. Given these numbers, it is possible to compute the value χ. This is done scanning all the space of the MVA output(s), with a step depending on the method output range[2].

---

[1] With the same rejection width used for creating the *Pi0Veto* described in Section 5.1.8.

[2] The MVA method output ranges are, conservatively speaking, the following: Fisher from -2.8 to 2.8; MLP from -0.4 to 1.2; BDT from -1.1 to 0.9. The scan step is equal to 1/1000 of the considered range.

The following step consists in finding the cuts that maximize χ. Care has been taken in case very few background events have passed a particularly tight set of cuts: if the number of passed events was lower than a certain threshold, it was forced to become that threshold, set at 5 events, and the FoM χ has been computed with these values.
The set of cuts that provides the highest FoM is saved, together with the FoM itself and the number of passed signal and backgrounds events (see Tables 5.1, 5.2, 5.3).

The evaluation procedure, together with the search for the maximum χ, is repeated for all the mass samples listed in Section 4.4.3, even if most plots will not show the values corresponding to the masses of 0.15 GeV/c^2 and 1 GeV/c^2 because, being very close to the masses of the π^0 and the $\eta\prime$ respectively, their efficiency is low. The sample with mass of 0.5 GeV is not as close as the other two to the mass of the η, nonetheless the performance of the MVA on that sample are lower than the others, because it suffers from the applied cut in mass.

Finally, this entire procedure is repeated for all the six combinations of method and strategy, and the results are compared.

During the first phases of the analysis, this procedure was carried out only on the 2 GeV/c^2 mass sample, taken as a reference example being in the middle of the allowed spectrum and sufficiently far away from any forbidden resonance. The results are visible in Table 5.1. The training was performed with the set of discriminating variables in use at the time. For the 1-step strategy were used: *MaxKSelectorsOfK*, *Leg0*, *Leg2*, *CosThetaS*, *CosThetaT*, *DeltaE*, *mes*, *Pi0Veto*, *KHelicityAngle*. For the 2-steps strategy the two MVAs were trained with different sets of variables: *CosThetaS*, *CosThetaT*, *DeltaE*, *mes* were in common, while *Leg0*, *Leg2*, and *MaxKSelectorsOfK* were specific for the MVA dedicated to reject uds, and *Pi0Veto* and *KHelicityAngle* were specific for the MVA dedicated to reject BpBm. The Fisher method proved to be largely inferior to the others; the 1-step procedure was less performant than the 2-steps one; BDT method was slightly better than MLP in both cases.

A check has been done with the current set of discriminating variables, and has been carried out on all the available signal masses, to verify that the chosen strategy and method are still the best performing ones. The results are visible in Table 5.2. The BDT method with the 2-steps strategy still returns the highest value of χ; it is also possible to appreciate the improvement in χ that has been achieved thanks to the new discriminating variables adopted in the meanwhile. All the three MVAs (the one for the 1-step strategy and the two for the 2-steps strategy) have been trained with the complete set of variables discussed in Section 5.1.

In all reported cases, the low-event correction has never been actually applied, i.e. $N^{pas} \geq 5$ for each evaluation sample.

The best combination results being to use the **2-steps strategy** with the **BDT**.

Method	1-step strategy				**2-steps** strategy			
	χ	N_S^{pas}	N_{BpBm}^{pas}	N_{uds}^{pas}	χ	N_S^{pas}	N_{BpBm}^{pas}	N_{uds}^{pas}
MLP	3.22	2242	156	13	4.09	1832	74	5
BDT	3.85	2917	168	16	**4.37**	2058	96	5
Fisher	1.36	3994	780	310	1.30	3922	666	331

Table 5.1: Comparison of the best performances (in terms in highest χ) for different training strategies and MVA methods to search for the optimal combination. The number of passed signal events N_S^{pas} has been evaluated for the 2 GeV/c^2 mass sample. The best combination, and its corresponding χ, is enlightened in bold.

Method	Signal Mass	1-step strategy				**2-steps** strategy			
	[GeV/c^2]	χ	N_S^{pas}	N_{BpBm}^{pas}	N_{uds}^{pas}	χ	N_S^{pas}	N_{BpBm}^{pas}	N_{uds}^{pas}
MLP	0.30	2.53	4196	360	94	2.85	4290	335	76
	0.50	2.26	3755	360	94	2.55	3832	335	76
	0.75	2.45	3758	312	80	2.74	4126	335	76
	1.5	2.48	3796	312	80	2.83	4256	335	76
	2	2.39	3961	360	94	2.69	4052	335	76
	3	2.70	4477	360	94	3.16	4337	154	68
	4	2.84	4357	312	80	3.14	4130	128	63
	4.5	2.82	4315	312	80	2.98	3987	135	65
BDT	0.30	4.19	2513	55	12	**6.05**	2360	25	5
	0.50	3.81	2285	55	12	**5.43**	2117	25	5
	0.75	4.18	2508	55	12	**5.88**	2294	25	5
	1.5	4.61	2535	48	10	**6.04**	2355	25	5
	2	4.66	2311	43	8	**5.67**	2212	25	5
	3	6.19	2614	26	6	**6.43**	2763	32	6
	4	7.03	2969	26	6	**7.26**	2821	24	5
	4.5	7.20	3043	26	6	**7.69**	3995	67	8
Fisher	0.30	1.23	4262	672	443	1.24	4054	536	400
	0.50	1.11	3828	672	443	1.11	3935	564	469
	0.75	1.20	4161	674	448	1.20	4233	564	469
	1.5	1.21	4184	672	443	1.23	3916	471	382
	2	1.16	3879	631	412	1.186	3666	415	359
	3	1.31	4400	634	416	1.40	4210	361	340
	4	1.51	3597	409	206	1.70	3138	84	131
	4.5	1.74	3265	296	126	2.26	1680	16	21

Table 5.2: Comparison of the best performances (in terms in highest χ) for different training strategies and MVA methods to search for the optimal combination. The evaluation has been performed on different signal mass samples.

5.3.2 Details of the Chosen Training

After having found the best strategy and method, there is still one last step to be done before having completed the set up of the event selection procedure, and it is to find the working set of MVA cuts (one cut for the BpBm-optimized BDT and the other for the uds-optimized BDT). Applying the procedure previously described, I obtained Table 5.2, but the values of the χ obtained for each signal mass come from the MVA cuts that maximize each of them. When running on real data we want, instead, to have a single couple of cuts to apply to all events.

The final step is therefore to interpolate between the various couples of cuts to obtain a *single couple of cuts* ($\mathrm{cut}_{\mathrm{MVA}_{\mathrm{BpBm}}}$, $\mathrm{cut}_{\mathrm{MVA}_{\mathrm{uds}}}$) to use on all the real dataset. I did this by looking at the values of the couples of cuts optimized for the single mass samples, trying all the possible combinations of the two cuts, and taking the one that provides the highest χ on the different mass samples. In principle there are N^2 combinations to analyze, where $N = 8$ is the number of mass samples under investigation, but in practice most of the cuts are the same for different signal masses.

Table 5.3 lists the couples of cuts that maximize χ for each signal mass. In Table 5.4 there is the χ for each signal mass deriving from the chosen interpolated couple of cuts. A study on the stability of the chosen selection is reported in Appendix A.

Mass [GeV/c^2]	$\boldsymbol{\chi}$	$\mathbf{cut}_{\mathrm{MVA}_{\mathrm{BpBm}}}$	$\mathbf{cut}_{\mathrm{MVA}_{\mathrm{uds}}}$
0.30	6.05	0.375	0.397
0.50	5.43	0.375	0.397
0.75	5.88	0.375	0.397
1.5	6.04	0.375	0.397
2	5.67	0.375	0.397
3	6.43	0.263	0.472
4	7.26	-0.155	0.582
4.5	7.69	0.095	0.472

Table 5.3: Maximized FoM for different signal mass samples and corresponding MVA cuts used.

TMVA proposes some useful plots to check, providing information about:

- The correlation between the discriminating variables (Figure 5.18). In all the three samples there is a strong correlation between the energy of the most energetic photon and its helicity angle ($\sim 50\div60\%$) and between the two Legendre moments ($\sim 45\div70\%$). Other relevant correlations are typical of each sample. The backgrounds show more correlated variables than the signal;

- The extent of the overtraining, evaluated through the comparison between the distributions of the MVA output, both for signal and background, for the training and the test samples (Figure 5.19). For what concerns the $\mathrm{MVA}_{\mathrm{uds}}$, there is an excellent agreement between the training and the test curves for the signal and a very good agreement for the background, so this BDT has not suffered from overtraining. The $\mathrm{MVA}_{\mathrm{BpBm}}$ still

Mass [GeV/c^2]	$\boldsymbol{\chi}$	$\boldsymbol{N_S^{pas}}$	$\boldsymbol{N_{BpBm}^{pas}}$	$\boldsymbol{N_{uds}^{pas}}$
0.3	6.05255	2360	5	25
0.5	5.42934	2117	5	25
0.75	5.88328	2294	5	25
1.5	6.03972	2355	5	25
2	5.67298	2212	5	25
3	6.10897	2382	5	25
4	6.41672	2502	5	25
4.5	6.86554	2677	5	25

Table 5.4: FoM for different evaluation signal mass samples, obtained setting: $\mathrm{cut}_{\mathrm{MVA}_{\mathrm{uds}}}$ = **0.397**, $\mathrm{cut}_{\mathrm{MVA}_{\mathrm{BpBm}}}$ = **0.375**. The number of backgrounds events is the same for each of the signal mass sample because the same background evaluation ntuples have been used for each of them.

presents a very good agreement for the signal distributions, but the background ones are not as good, showing some discrepancies, in particular at high value of the output; a work on the parameters of the BDT was unsuccessful in completely fixing this issue, and currently we have decided to postpone a more detailed study of this since it only affects a few % of the sample;

- The Receiver Operating Characteristic (ROC) curve, i.e. the curve that gives the background rejection as a function of the efficiency for accepting signal (Figure 5.20). A tighter requirement, i.e. a larger cut on the MVA output, increments the purity of the sample but reduces the signal efficiency, and vice versa. The ROC curve for the $\mathrm{MVA}_{\mathrm{uds}}$ shows that even requiring a very high purity we still are able to retrieve a huge number of signal events (for example, even requiring a cut tight enough to have only a 1% contamination from the background, it is possible to achieve a signal efficiency of 95%). The $\mathrm{MVA}_{\mathrm{BpBm}}$ is instead weaker, not being able to reject the BpBm with the same efficiency as the $\mathrm{MVA}_{\mathrm{uds}}$ rejects the uds background.

These plots are provided for both of the BDT used (the one optimized to reject the uds background, $\mathrm{MVA}_{\mathrm{uds}}$, and the one optimized to reject the BpBm background, $\mathrm{MVA}_{\mathrm{BpBm}}$).

The depicted plots show that the $\mathrm{MVA}_{\mathrm{BpBm}}$ is a weaker discriminator than the $\mathrm{MVA}_{\mathrm{uds}}$. This is understandable, since the B^+B^- background events that pass the pre-selection are more difficult to reject than the uds, because they are more similar to the signal, that is a B decay event. This can be seen in particular looking at the distributions of the discriminating variables in Section 5.1: for each variable, the distribution of the BpBm sample is closer than the one of the uds to the signal's.

Finally, we can see the efficiency of this selection for the MC signals and for the MC backgrounds. The efficiency is defined as the number of events that have passed all the selection procedure over the number of generated events (i.e. the number of events before applying the pre-selection).

Figure 5.21 shows the results for the signals samples. The higher efficiency at the higher

masses is due to the fact that the background is more important at low masses, so the MVA cut had to be tighter at low masses. Interpolating the cuts had hence brought to a slightly smaller efficiency at the beginning of the spectrum than at the end.
Table 5.5 shows the results for the MC backgrounds. As predicted, Mumu, Tautau, and Bhabha backgrounds are far less important than the $B\bar{B}$ and the quark-like ones. B0B0bar efficiency is far smaller than BpBm, while ccbar is comparable with uds.

Bkg process	**# events before selection**	**# events after selection**	**Efficiency** $(\times 10^{-7})$
B^+B^-	5.59×10^7	284	51 ± 3
$B^0\bar{B}^0$	5.79×10^7	18	3.1 ± 0.7
uds	5.50×10^7	206	37 ± 3
$c\bar{c}$	3.56×10^7	108	30 ± 3
$\mu^+\mu^-$	7.61×10^7	0	0 ± 0
$\tau^+\tau^-$	5.87×10^7	7	1.2 ± 0.5
Bhabha	6.62×10^7	0	0 ± 0

Table 5.5: Number of events before and after all the selection procedure, for each MC background source, together with the corresponding selection efficiency. The events in the π^0, η, and $\eta\prime$ peaks have been counted.

5.3.3 Comparison with Control Sample

Since this analysis is a diphoton resonance search, the variable we are most interested in is the invariant mass of the photon pair considered as the ALP candidate. The comparison between MC backgrounds and Run 3 control sample is hence performed looking at the distribution of this variable.

To perform such a comparison, it is necessary to adequately weight the MC contributions. Formula 5.9 provides the number of (physical) background events that have passed all the selections (the pre-selection, see Section 4.4.1, and the 2-steps BDT procedure) as a function of the cross section of each process and the number of MC events that have been generated and that have passed the selections.
Therefore, we need the integrated luminosity of the control sample (see Section 4.4.5), the cross section of each background process (see Table 2.1), and the number of MC background events before and after all the selections, that are listed in Table 5.5.
As predicted in Section 4.3 and 5.1, the B0B0bar, Mumu, Tautau, and Bhabha backgrounds are indeed negligible with respect to the other ones.

Figure 5.22 shows this comparison. The agreement is very good over all the spectrum, as for all the bins the MC and the control samples are in agreement within 1 or at most 2 sigmas. The peaks coming from π^0 and η ($\eta\prime$ peak is much smaller than the other two), here constituting an irreducible background, are well modeled. This will be precious to validate the fit procedure and compare the results with known SM processes.

The main discrepancy between the MC samples and the control sample is the total number of events: **563** events from the control sample pass the entire selection procedure, while only **479** weighted MC events survive. The difference is compatible with 0 only within 2.6 standard deviations[3]. Though, this is not surprising, because we are aware that the MC samples are not perfect representation of the really produced events. In particular what is known to be missing are the Initial State Radiation (ISR) processes, i.e. events where the electrons or positrons of the beams emit one or more hard photons before colliding in the IP.

This non-negligible deficit indicates that, even if the comparison is overall very good, it cannot be completely trusted. This is the reason why the background PDF is modeled from the real data of the control sample and not from the MC, as mentioned in Section 4.1; the details about the modeling of the background are in Section 6.2.

---

[3] This can be obtained assuming that the number of events passing the selection procedure is distributed as a Poisson variable, both for the control sample and for the MC background samples. This is a safe assumption as the number of events before the selection is big and the probability of being selected is small.

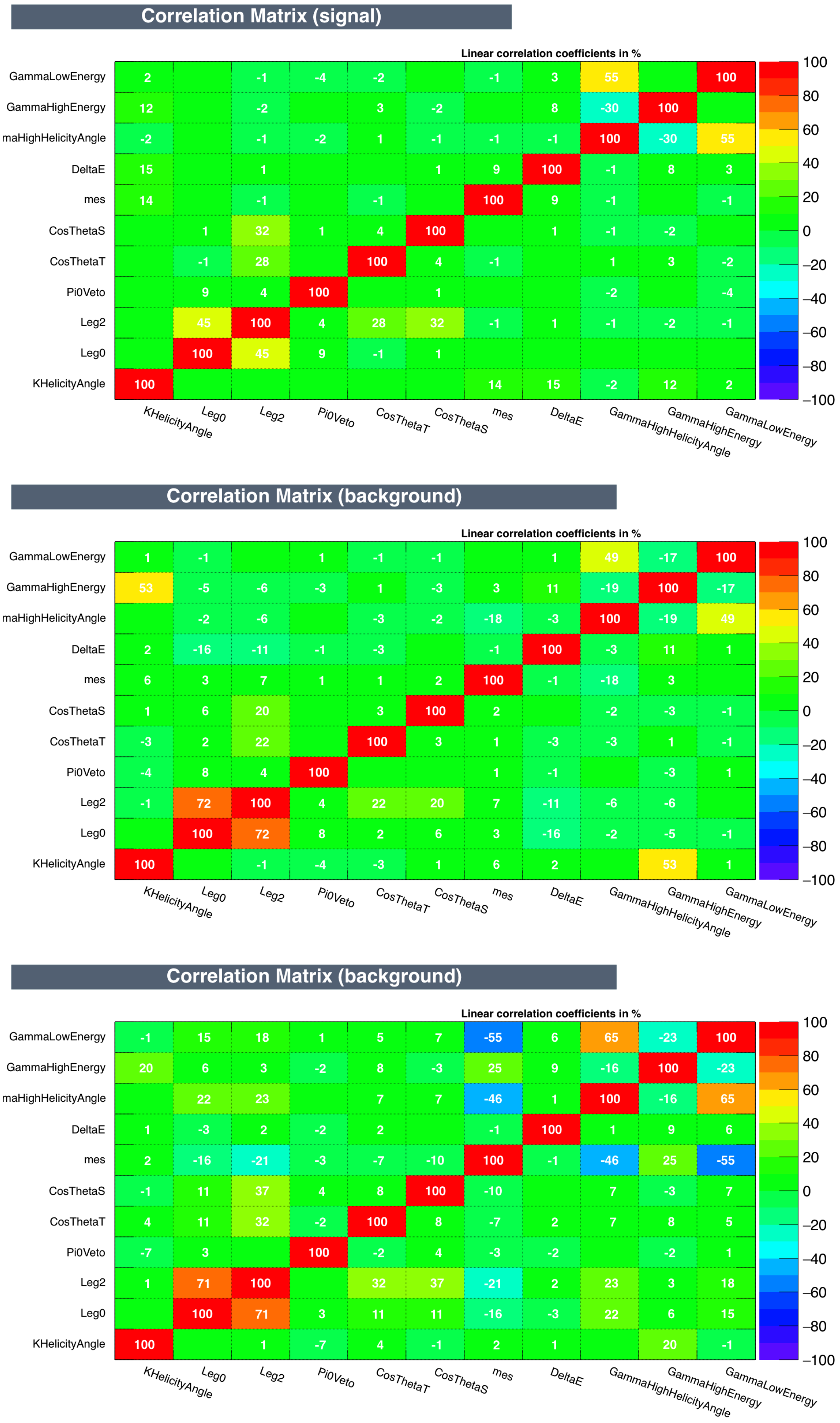


Figure 5.18: Linear correlation between the discriminating variables for the three training samples. The top plot refers to the flattened signal sample, the middle one refers to the uds background sample, the bottom one refers to the BpBm background sample.

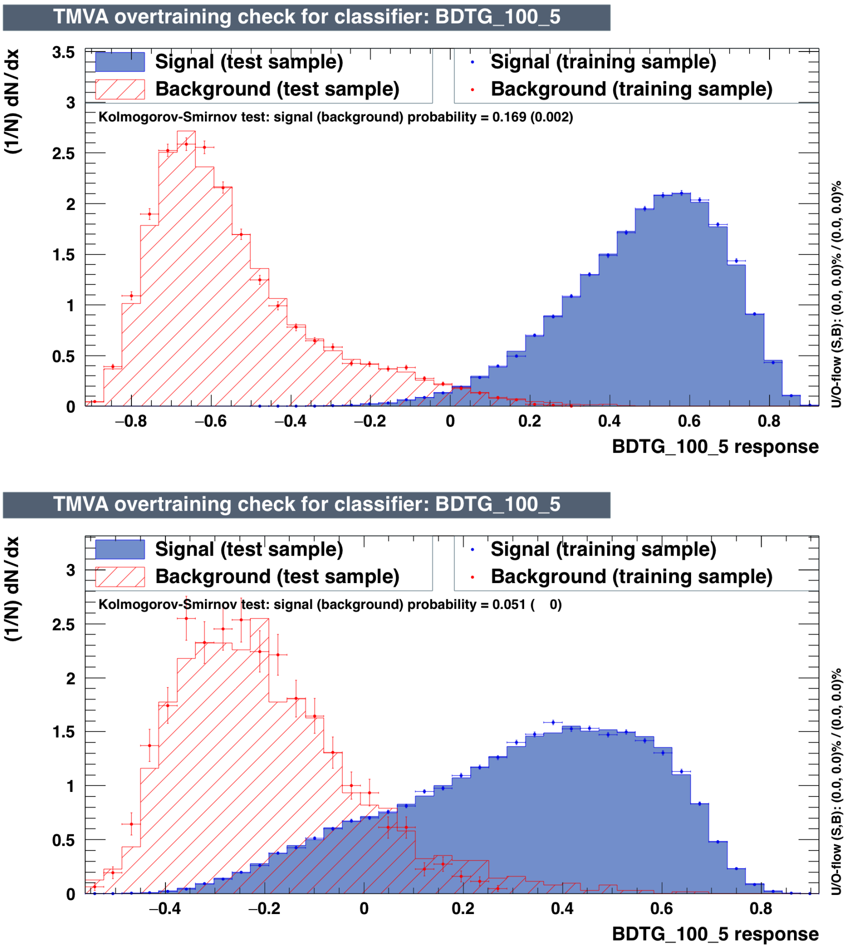
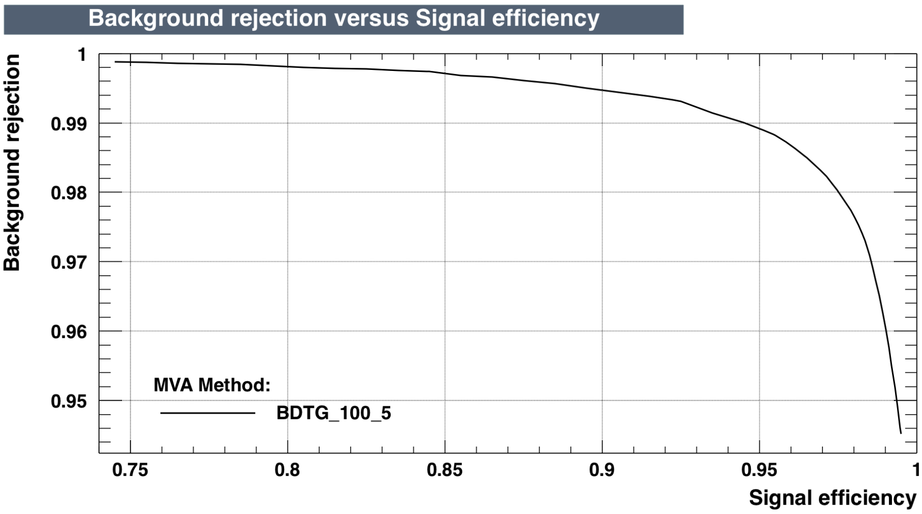


Figure 5.19: Overtraining check. On the top there is the plot about the MVA$_{\text{uds}}$, on the bottom the one about the MVA$_{\text{BpBm}}$. The distributions of the output of the BDT for both signal and background and for both training and test samples are plotted.



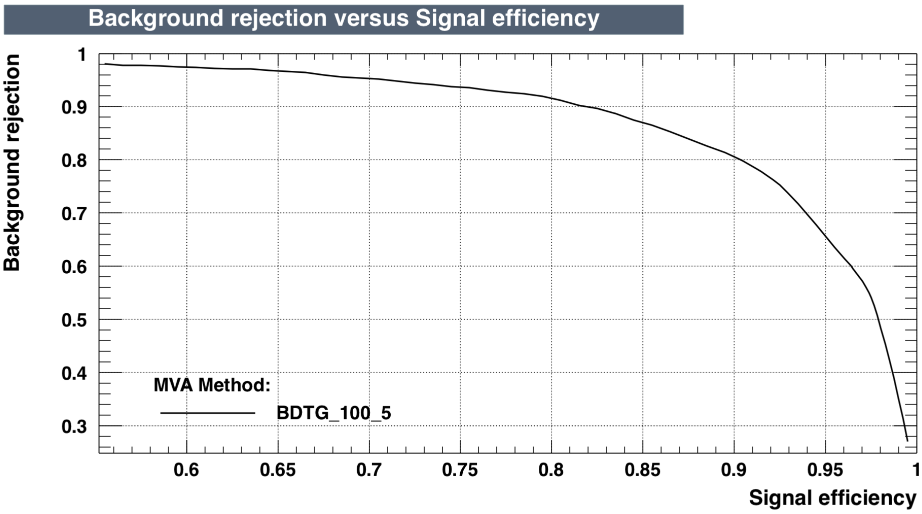


Figure 5.20: ROC curves. On the top there is the plot about the $\mathrm{MVA}_{\mathrm{uds}}$, on the bottom the one about the $\mathrm{MVA}_{\mathrm{BpBm}}$. The ROC curves show the background rejection as a function of the signal efficiency.

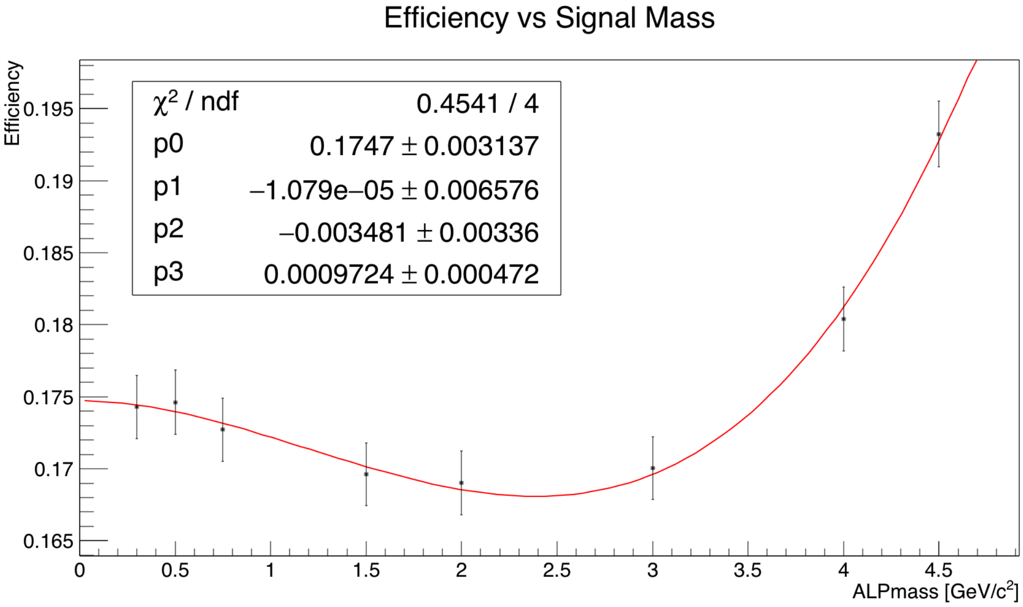


Figure 5.21: Selection efficiency for the different signal masses. The points have been fitted with a 3rd degree polynomial.

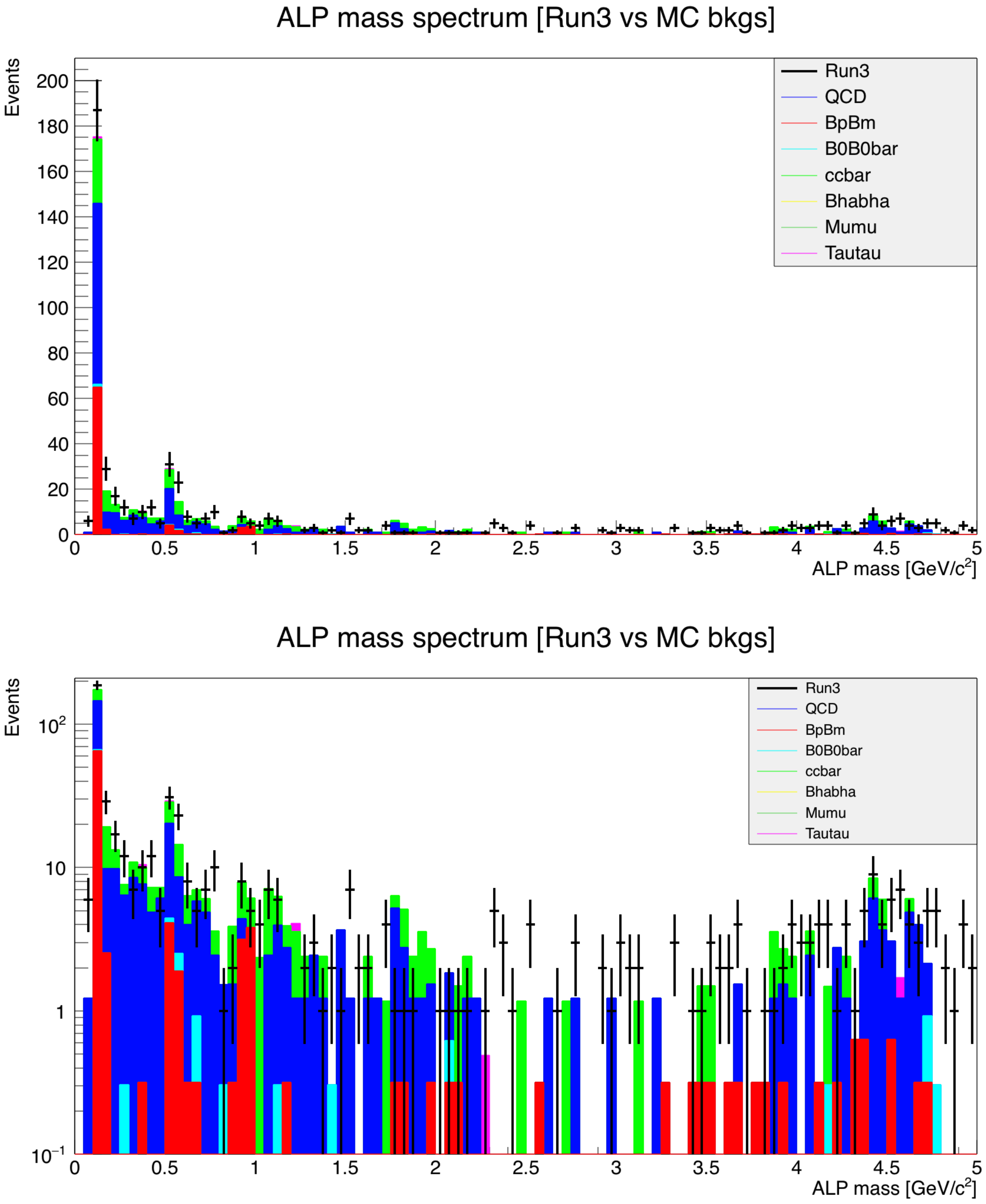


Figure 5.22: ALP mass spectrum, comparison between the MC backgrounds sources and the control sample (*BABAR* Run 3). Linear and logarithmic representation are shown, the latter to underline the behavior of the low-populated bins. Mumu and Bhabha backgrounds are completely absent.

Chapter 6

Signal Fit and Yield Extraction

As outlined in Section 4.1, the signal yield is extracted from a fit to the diphoton mass spectrum of the events that have passed the selection described in Chapter 5. It is hence necessary to find the proper PDFs describing the signal and the background throughout all the spectrum, and verify the reliability of the fit procedure.

Section 6.1 describes the procedure to obtain a model for the signal PDF, extrapolated from the MC signal samples, while Section 6.2 examines the different approach used to obtain a model for the background PDF, extrapolated from the control sample.
Finally, in Section 6.3 the simultaneous fit of signal+background is performed on the control sample to verify the reliability of the fit procedure.

6.1 Signal PDF Determination

Figure 6.1 shows the signal mass distributions, after having applied the entire selection procedure, for the various signal mass samples. It is evident that the distributions differ not only for their mean values, but also for their variance, and possibly even for their asymmetry.

Our purpose is to find a way to model the signal throughout the entirety of the allowed spectrum. Since we want a continuous description, valid in all the possible mass points, but we cannot produce an infinite number of MC samples and fit each of them, what we need to do is to *interpolate* between the available mass points.
The first thing to do is to find a PDF family able to properly fit the MC signal at all the different mass points. This family will be characterized by some parameters $\vec{P}$; the signal fit performed at each mass sample will return a different set of parameter $\vec{P}\left(M_j\right)$, where M_j is the central mass of the signal sample, $j = 1, \ldots, N$. To allow using the PDF for an arbitrary value of the ALP mass, each of these parameters has to be interpolated to all the spectrum to find $\vec{P}\left(M\right)$. This is done using a simple function (as a low-degree polynomial) and fitting each of the parameters P_i as obtained at the mass points M_j.
For each parameter of the PDF family we therefore have a function $P_i\left(M\right)$, whose dominion is the ALP mass spectrum.

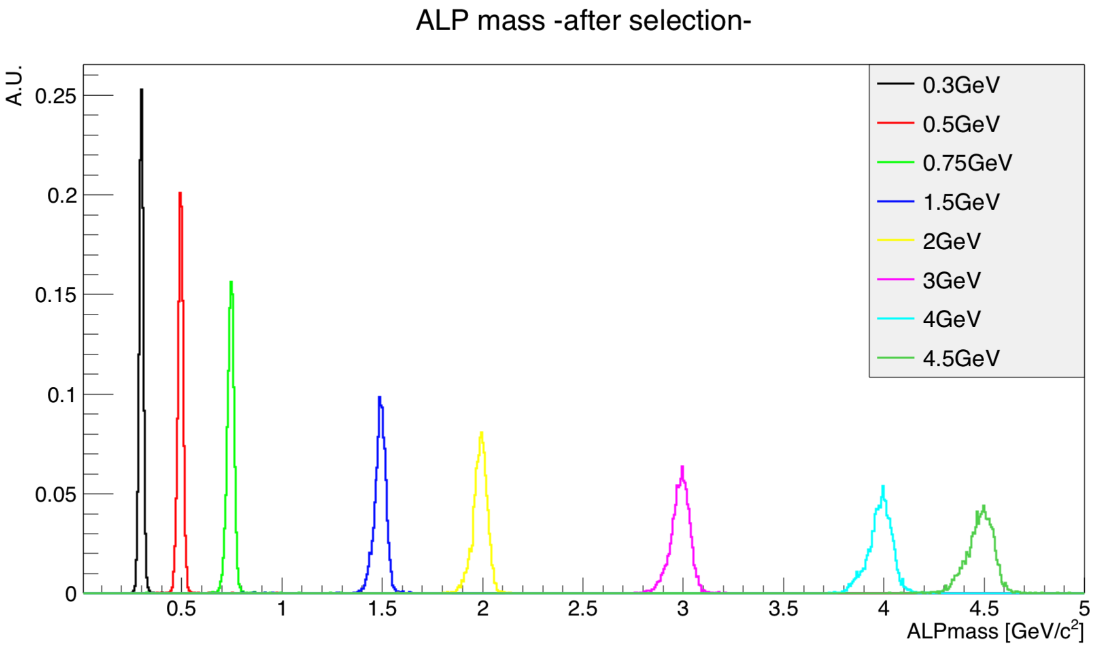


Figure 6.1: Mass distribution of the ALP for different signal mass samples, for events that have passed the entire selection. The distributions are normalized.

This procedure has been carried out using **RooFit** [43], a ROOT library that provides a useful toolkit to perform binned and unbinned maximum likelihood fits, produce plots, and allow the generation of events following the desired PDF, as well as combining different PDF (addition, convolution, etc).

It is important to recall that a PDF, being a function describing a probability, is normalized to 1. The normalization is automatically performed by RooFit, so it is not necessary to manually normalize the PDF we use, not even when we combine more of them. The fact that a PDF describes a probability also implies that it can never be negative.
When we *add* two or more PDFs and perform a simultaneous fit, we do not directly retrieve the number of events for each PDF, but we get the relative fraction of each PDF, with the constraint that the sum of the fractions has to be equal to 1.

For this analysis I used RooFit version 3.57.

6.1.1 Analyzed PDFs

The first step is to find a PDF family that can model the mass distributions of all the signal samples. The first candidate we can think of is the *Gaussian* PDF; another possibility is the *Crystal Ball* (CB) PDF [44]. In neither case we are interested in the normalization, as it is automatically handled by RooFit, so it will be always neglected in the definition of the used functions.
For what concerns the Gaussian, we have the usual definition:

$$G\left(x;\mu,\sigma\right) = e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{6.1}$$

Here the parameters $\vec{P}$ previously introduced are μ and σ.

The CB PDF is a semi-empirical function, common in high-energy physics, used to model lossy processes. Its core structure is a Gaussian, while its lowermost tail, from a certain threshold downward, is a power-law, so it looks like a Gaussian with a slowly decreasing left tail. It is defined as follows:

$$CB\left(x;\mu,\sigma,\alpha,n\right)=\begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, & \text{for } \frac{x-\mu}{\sigma}>-\alpha \\ A\cdot(B-\frac{x-\mu}{\sigma})^{-n}, & \text{for } \frac{x-\mu}{\sigma}\leq-\alpha \end{cases} \tag{6.2}$$

where A and B are defined as:

$$\begin{aligned} A &= \left(\frac{n}{|\alpha|}\right)^n\cdot\exp\left(-\frac{|\alpha|^2}{2}\right) \\ B &= \frac{n}{|\alpha|}-|\alpha| \end{aligned} \tag{6.3}$$

In this case the parameters $\vec{P}$ are μ, σ, α, and n.

The parameters μ and σ have the same meaning that they have in a Gaussian. n is the parameter defining how quickly the left tail decreases, and α expresses the distance from μ, in terms of σ, from which the function ceases to behave as a Gaussian and adopts the power-law behavior. The terms A and B are determined to match the junction point between the two parts of the function and ensure the continuity of both the CB and its derivative.

These two functions have been tested to fit the eight signal mass points (as described in Section 5.3.1, the mass samples close to the cuts in mass around the π^0, η, and $\eta\prime$ have been eliminated). Since the CB is a more complex function than the Gaussian, each data sample is first fit with a Gaussian, then the values of μ_G and σ_G are retrieved and are used as starting points for the fit with the CB.

The parameter n of the CB has shown some stability issues. A small variation in its starting point and/or in its allowed range can easily bring to failed calculation of the uncertainties of the fitted parameters, or to uncertainties greater than the parameters themselves. Even when all the fits converge without failures, the behavior of the parameter n as a function of the ALP mass M shows abrupt changes at the beginning and at the end of the spectrum, together with very high uncertainties, as can be seen in Figure 6.2.
The reason why n happens to be so problematic is that it appear as an exponent in a power-law, and in both A and B, hence the dependence of the CB from it is often critical.

To overcome these issues, I fixed the parameter n at a constant value. I have studied the effect of fixing n at different values, and I observed that the fit procedure is almost insensible to it, as long as it is in the range $[\sim 20, \sim 120]$, as the fitted values of the other parameters are not influenced (see Appendix B for more details).
To simplify the fit I hence fixed it at the following value:

$$n_{CB}=80 \tag{6.4}$$

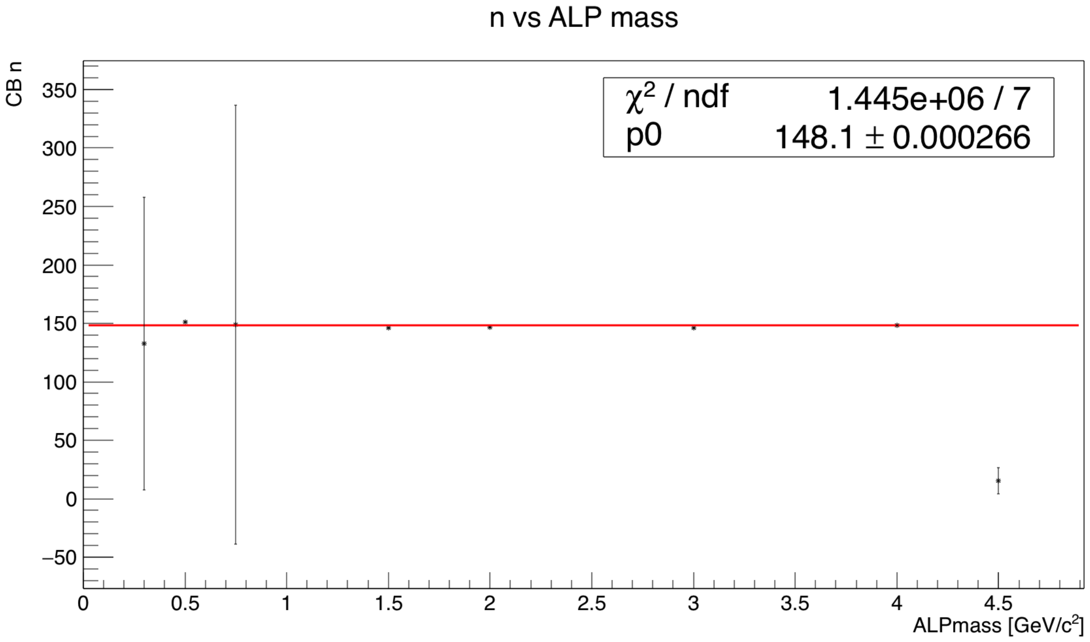


Figure 6.2: Parameter n of the CB, retrieved from the fits performed on each single signal mass point, as a function of the ALP mass. The points have been fitted with a constant. See the text for more comments.

The comparison between the fits performed with the Gaussian and the CB PDFs, for each signal mass sample, can be seen in Figure 6.3. In Table 6.1 are reported the values of the corresponding reduced chi square $\chi^2_{Red} \equiv \frac{\chi^2}{DoF}$, where DoF is the number of degrees of freedom of the fit. The CB provides a better fit in each case, as its χ^2_{Red} is always lower than the for the Gaussian, and having used the caution of fixing $n_{CB} = 80$ there are no stability issues or non-physical uncertainties associated to any of the other fit parameters.
The CB PDF has hence been chosen over the Gaussian to fit the ALP mass signal.

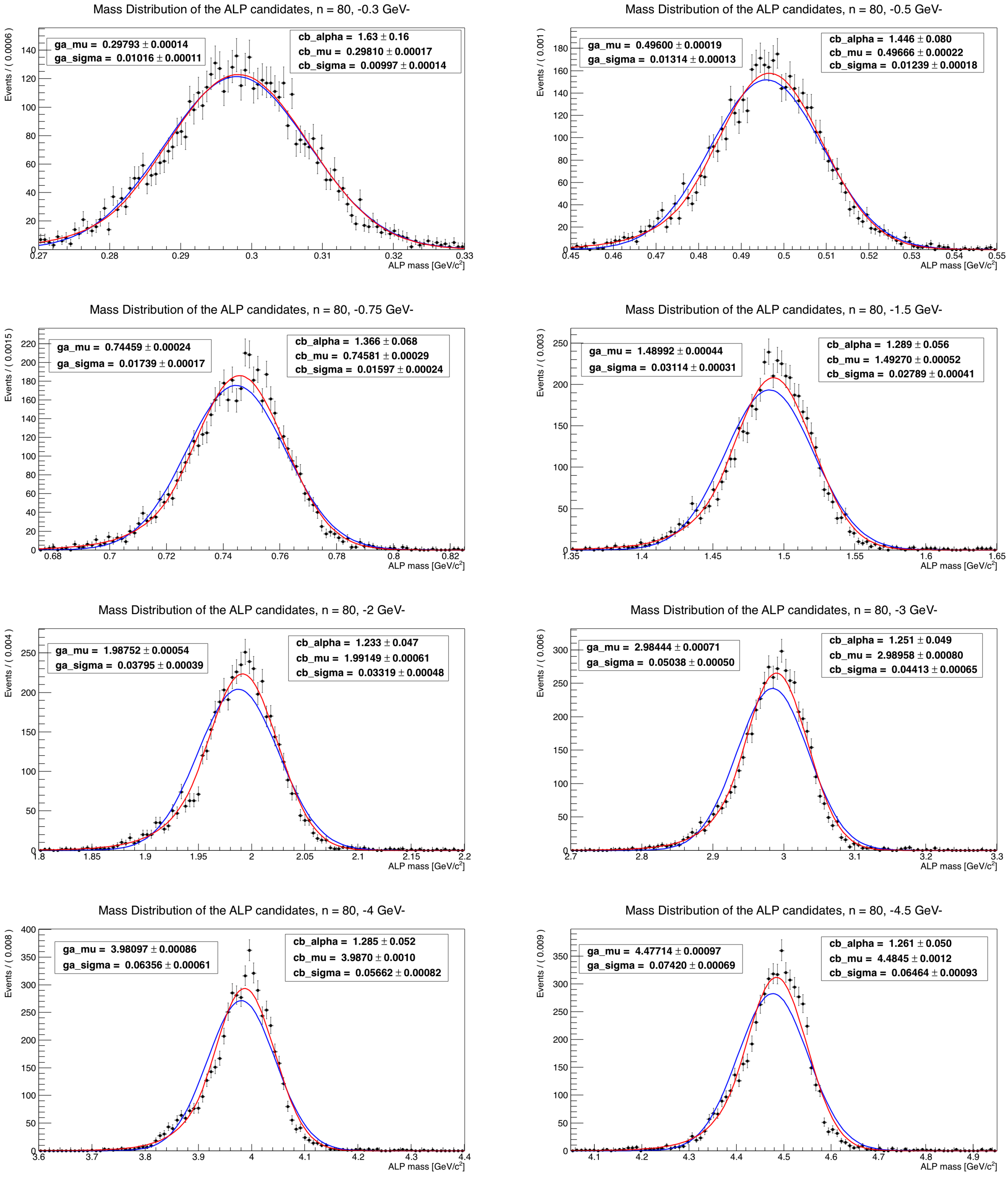


Figure 6.3: Comparison fits between the Gaussian PDF and the CB PDF, performed on the eight signal mass samples. The CB parameter n has been fixed at 80.

ALP mass (GeV/c^2)	χ^2_{Red} **Gaussian PDF**	χ^2_{Red} **CB PDF**
0.3	0.92	0.86
0.5	1.37	0.91
0.75	1.83	0.98
1.5	3.09	1.41
2	3.56	1.24
3	3.94	1.30
4	4.65	2.33
4.5	5.76	2.79

Table 6.1: χ^2_{Red} of the two analyzed functions, Gaussian and Crystal Ball, for all the signal mass points. In all cases the fit with the CB returns a lower χ^2_{Red} than the fit with the Gaussian. The number of bins used to evaluate the χ^2 is 100, hence the DoF for the two fits is 98 and 97 respectively.

6.1.2 Crystal Ball Parameters

The behavior of the CB parameters σ, α, and μ as a function of the ALP mass M is shown in Figure 6.4.

The trend of these parameters as a function of the ALP mass has been fitted with a simple function, i.e. a low-degree polynomial, to avoid overfitting: a function with many free parameters gets naturally more adapted to the set of points, but at the same time it is more sensible to statistical fluctuation of the MC signal samples. It is also important to underline that I performed the fits on simulated MC signal sample, and a priori we cannot know how much this simulation is a good model of the possibly real signal.

The purpose is to find a function that is loosely able to fit the set of points. We hence prefer a simple function with few free parameters. The three free parameters of the CB are hence fitted with polynomials of degree 1, 2, and 3.

The only exception is μ, because it is supposed to represent the mean value of the Gaussian core, hence we expect it to be strictly linear in the ALP mass, as it is actually confirmed by the fit.

The χ^2_{Red} of the fit for $\sigma(M)$ does not get sensibly smaller increasing the degree of the polynomial, as can be seen from the top plot of Figure 6.4.

On the contrary, a great improvement can be seen for $\alpha(M)$ when using a higher degree polynomial, as shown in the middle plot of the same figure.

I thus chose a linear function to model $\mu(M)$ and $\sigma(M)$ and a cubic function for $\alpha(M)$.

The three functions are used to interpolate the values of the three free parameters of the CB in all the allowed ALP mass range, and they are used to fix the CB parameters during the simultaneous fit of signal+background.

All the functions describing the parameters of the CB are explicitly reported below:

$$\begin{aligned} n\,(M) &= 80 \\ \sigma\,(M) &= 0.0131 \cdot M \\ \alpha\,(M) &= 1.65 - 0.46 \cdot M + 0.162 \cdot M^2 - 0.017 \cdot M^3 \\ \mu\,(M) &= 0.9967 \cdot M \end{aligned} \tag{6.5}$$

where the ALP mass M, σ, and μ are expressed in GeV/c^2.

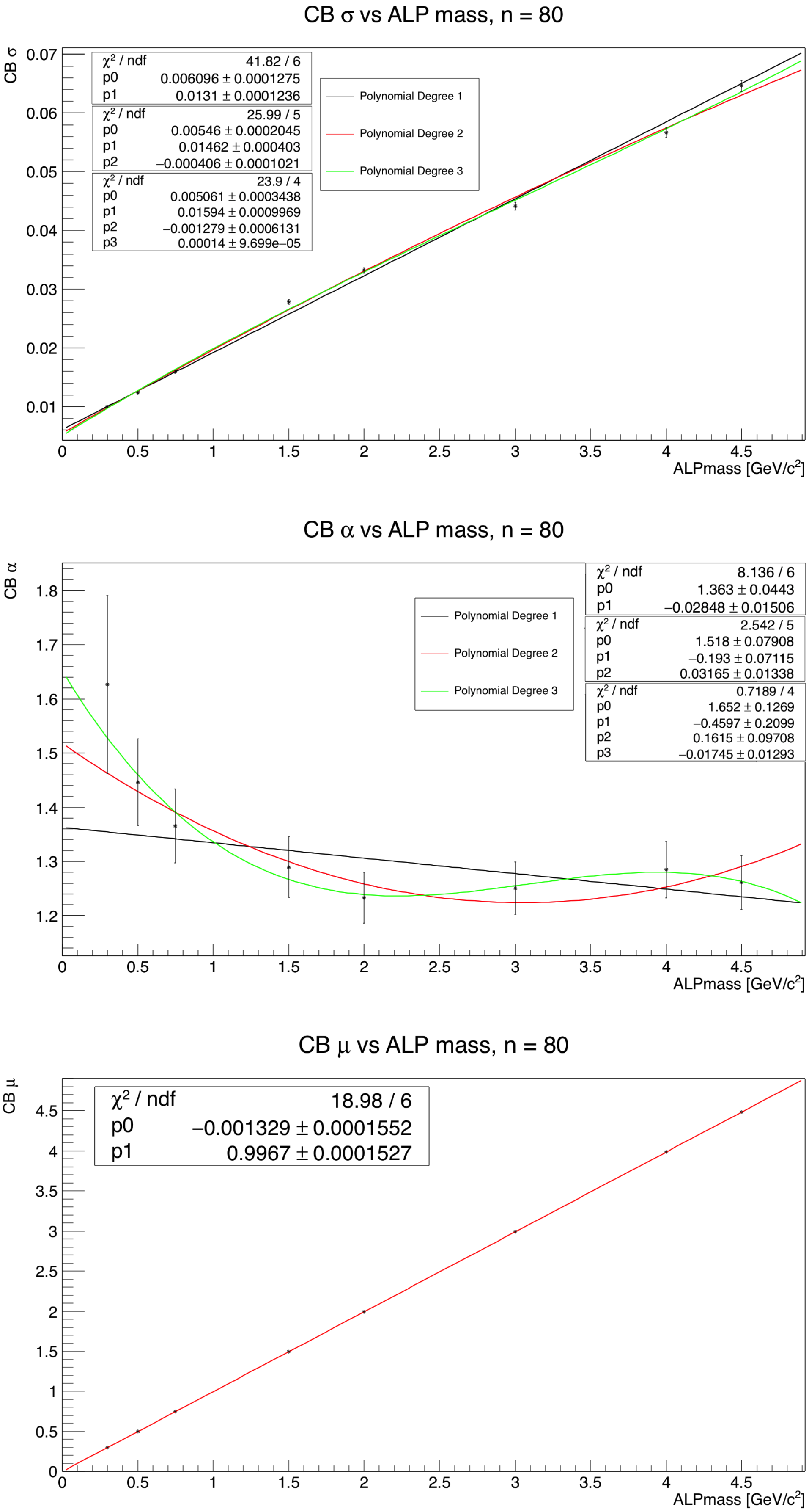


Figure 6.4: Fits on the CB free parameters α, σ, and μ, with polynomials of different degrees. The result of each fit is reported in the corresponding graphic.

6.2 Background PDF Determination

As anticipated in Section 4.1 and explained in Section 5.3.3, the procedure to extract the PDF model for the background is quite different from the one adopted for the signal. The first difference is that I used the data from the control sample instead of the MC simulated data, and this is done because we cannot completely trust the simulation, that has shown some deficit in the number of events with respect to the real data.

Using real data has two main drawbacks: the risk of overfitting over the statistical fluctuations of the control sample, and of including actual signal events and treating them as background.

For what concerns the second issue, we can first of all starting from the assumption that the signal is absent or almost negligible, at least within the small integrated luminosity of the Run 3 (that is the 7.6% of the total on-peak integrated luminosity collected by *BABAR*). This is a reasonable assumption, but even if that was not true, we can rely on the narrowness of the signal with respect to the width of the mass window used for the fit. As it will be explained in detail in Section 6.3, the simultaneous fit of signal+background to extract the yield will be performed on sub-windows of the available mass spectrum. The width of these windows is defined as a multiple of the expected width of the signal at that given mass. In this way, we can be sure that the signal, even if present, is narrow in comparison with the whole analyzed mass window, therefore it would not significantly affect the description of the background. If a non-negligible amount of signal events was present, this still can be checked a posteriori and accordingly modify the background PDF extraction.

The way to overcome the first issue, that is the risk of overfitting on the Run 3, is to adopt a flexible PDF, loosely constrained. This means that we want to use the lowest degree polynomial possible, to be able to grasp the main trend of the background mass distribution without specializing on the fluctuation of the given data sample. In addition, the parameters of the PDFs used to fit the background in the different windows are *not* fixed during the fit to data, as it was done for the parameters of the CB modeling the signal: during the simultaneous fit of signal+background to extract the signal yield, the background PDF is let free to loosely adapt to the background shape, depending on the actual data sample available in that mass window. When we will run on the full *BABAR* dataset we will use the same families of polynomials chosen here, but the particular parameters will not be externally fixed.

If a given invariant diphoton mass window includes the resonances of π^0, η, or $\eta\prime$, they have to be fitted as well. Since we will veto around these masses, there is no need to perform a high-precision fit: the fit is performed to improve the description of the background in the vicinity of the peaks. Also, using a simple function to fit these peaks makes the whole fit procedure more stable. Because of these reasons, each peaks is fitted with a Gaussian, whose parameters are constrained in small ranges. Also the fraction of the Gaussian with respect to the polynomial background is fitted from the data, its value being constrained in a small range.

6.2.1 Candidate Background PDFs

The purpose of this study is to find the optimal PDF families describing the control sample in different mass windows. It is not necessary that the same PDF family is used for all the spectrum.

When performing a fit to the data of the control sample, two things have to be taken into consideration. First, if the window includes a resonance (π^0, η, or $\eta\prime$), the overall PDF must be the sum of a polynomial, describing the continuous background, plus a Gaussian to model the peak. This has already been discussed and does not represent an issue.

The other thing that has to be taken into account is how the normalization of the PDFs is performed by RooFit. Since each PDF has to be normalized to 1, one of its free parameters is actually constrained. A polynomial is usually defined in the following way:

$$P_n(x) = \sum_{k=0}^{n} a_k x^k \tag{6.6}$$

Because in the process of normalization one degree of freedom of the function is lost, when we call a polynomial PDF in RooFit, the program automatically sets its constant term to 1, obtaining a polynomial in the form of:

$$P_n(x) = 1 + \sum_{k=1}^{n} a_k x^k \tag{6.7}$$

This is generally irrelevant, except when there are no data in the vicinity of the value 0 of the independent variable (x in the case of Equation 6.7, the ALP mass in my case). If such a situation occurs, a fit performed with the default polynomials available in RooFit cannot properly converge, or returns unacceptable parameters, because it is impossible to make a function like 6.7 pass through the origin.

This is the situation in this analysis. As we can see in Figure 5.22, there are no events in the vicinity of the origin: a photon pair has a mass close to zero only if the two photons are almost collinear, but in that case the crystals of the EMC would not be able to resolve them as two distinct photons, hence they would be recorded as a single photon, and their mass would not contribute to the reported spectrum.

This is the reason why a self-defined polynomial has to be used in the first mass window, that includes the origin, as it is necessary to define a polynomial deprived of its constant term. Elsewhere the default RooFit polynomials can be used, in particular the polynomials from the RooChebychev class, whose use is recommended [43].

Four different kind of polynomials have been tested in all the mass windows:

- $1 + a \times M$
- 1
- M
- $M + b \times M^2$

where the normalization procedure has already been taken into account, hence one parameters has been neglected. The last two polynomials are the self-defined, without the constant term, while the first two polynomials are members of the RooChebychev class, whose definition is equivalent to the usual polynomials' one for degrees 0 and 1.

6.2.2 Chosen Background PDFs

In this section the results of the fits performed to decide the optimal background PDFs are reported. The fits have been carried out on mass sub-windows of the control sample.

During the simultaneous fit on data of signal+background the width of the mass windows is a multiple of the signal width (the actual windows size for the simultaneous fit is reported in Section 6.3). While performing the studies to find the optimal PDF families to describe the background, instead, the width of the mass windows has been fixed to a constant value. Different widths are shown here.
A wider window contains more events, and in a situation with low-statistics like this analysis, this is an advantage. On the other hand, a larger window can sometimes bring instabilities in the fit, particularly if there are sudden changes in the distribution and the fit function does not have enough parameters to adapt to these changes.

In Figures 6.5 and 6.6 are respectively reported the results of the fits performed on mass widows with a width of 300 and 500 MeV/c^2. For each of the four fit polynomial PDFs tested, the fit outputs are reported in the plots. If the window does not include one of the resonances π^0, η, or $\eta\prime$, then only the free parameters of the polynomial PDFs are reported, otherwise they are reported together with the parameters and the relative fraction of the Gaussian PDF.

It can happen that in windows with low statistic the fit function would go negative, but this is forbidden as a PDF represents a probability. RooFit handles these situations forcing the value of the PDF in these regions to 0, but this situation should be avoided. Comparing the plots around the mass of 1.8 GeV/c^2, it can be seen that enlarging the width of the window is a solution to this issue, due to the increase in statistic. This reflects in the definition of the windows width, as function of the signal sigma, that has to be used for the simultaneous fit of signal+background.

From the fit performed in the first window (that includes the origin), regardless its width, the customized 2nd degree polynomial deprived of its constant term, $M + b \times M^2$, is the one most suited to fit the data, while elsewhere the $1 + a \times M$ function is the best fitting. This can be seen from Figure 6.7, where the χ^2_{Red} for the fits performed with the four tested polynomials is reported, as a function of the starting point of the analyzed window, for a window width of 0.5 GeV/c^2.
In the first window the fit with the $M + b \times M^2$ PDF is the one with the lowest χ^2_{Red}. In all the other windows the fits performed with the $1 + a \times M$ PDF give a result better than the ones obtained with the other polynomials.

The chosen polynomial PDFs hence are:

- $M + b \times M^2$ in the mass window including the origin ($M = 0$);

- $1 + a \times M$ elsewhere.

As already discussed, the parameters a and b will not be determined on the control sample, but will be left floating and fitted on the data.

6.3 Signal Fitting and Yield Extraction

I have carried out a test on the data of the control sample to verify the stability and the possible performances of the simultaneous fit of signal+background. This test consists in fitting the Run 3 data with a composite PDF, obtained as the sum of the signal PDF and the background PDF. As mentioned in Section 6.2, the fit is performed on various sub-windows of the whole mass spectrum. In each of these windows, centered around the analyzed ALP mass hypothesis, the simultaneous fit is performed and the yield, i.e. the number of signal events, is extracted.

The goal of this test is to extract the distribution of the number of signal events. Assuming that the number of signal events in this data sample is negligible, the distribution is expected to be a Gaussian centered in zero if the statistic is large enough. This is because when performing an unbinned fit we are counting the number of signal events in the analyzed region, and this number follows a Poisson distribution, that tends to a Gaussian for large statistic. That could be not true for the Run 3 data sample, but with the 15 times higher statistic of the whole data sample we expect this assumption to be true.

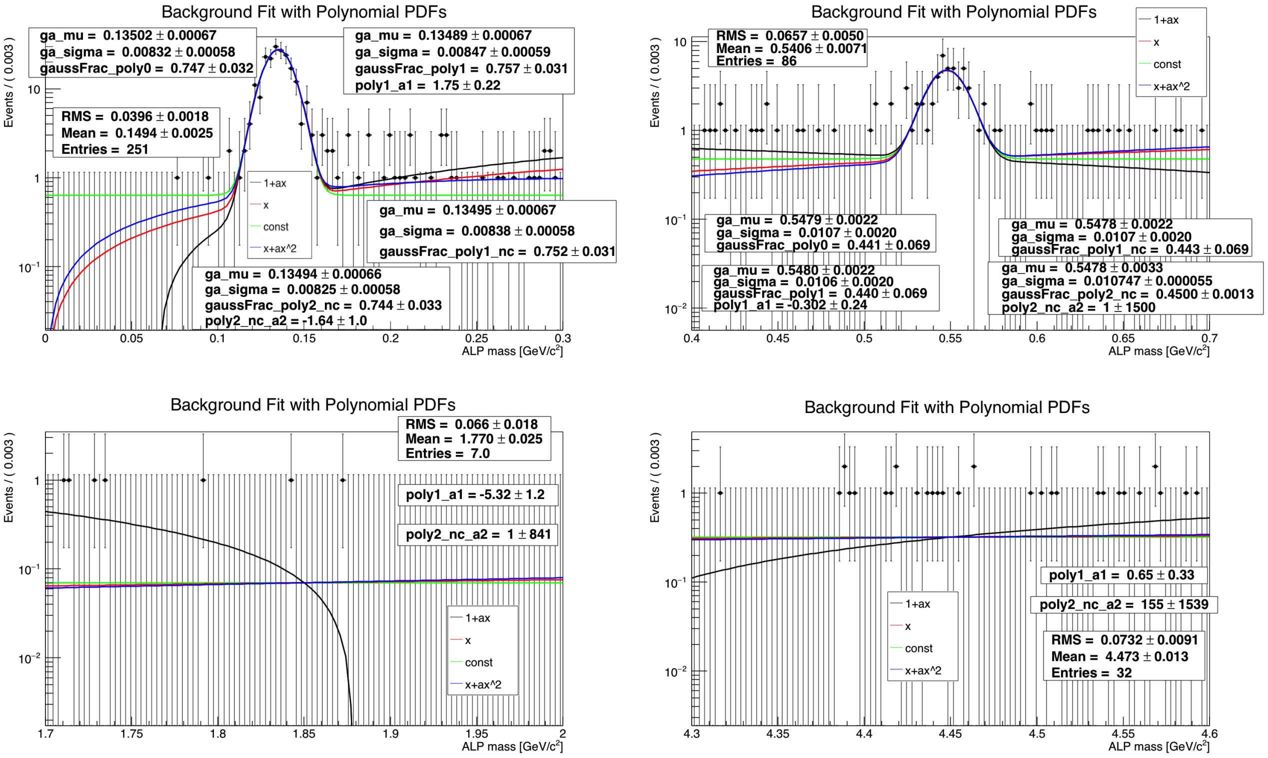


Figure 6.5: Events of the control sample, after the entire selection procedure, fitted with four different polynomial PDFs. The peaks in the first two plots correspond, respectively, to the π^0 and the η resonances. The mass window has a constant width of 300 MeV/c^2.

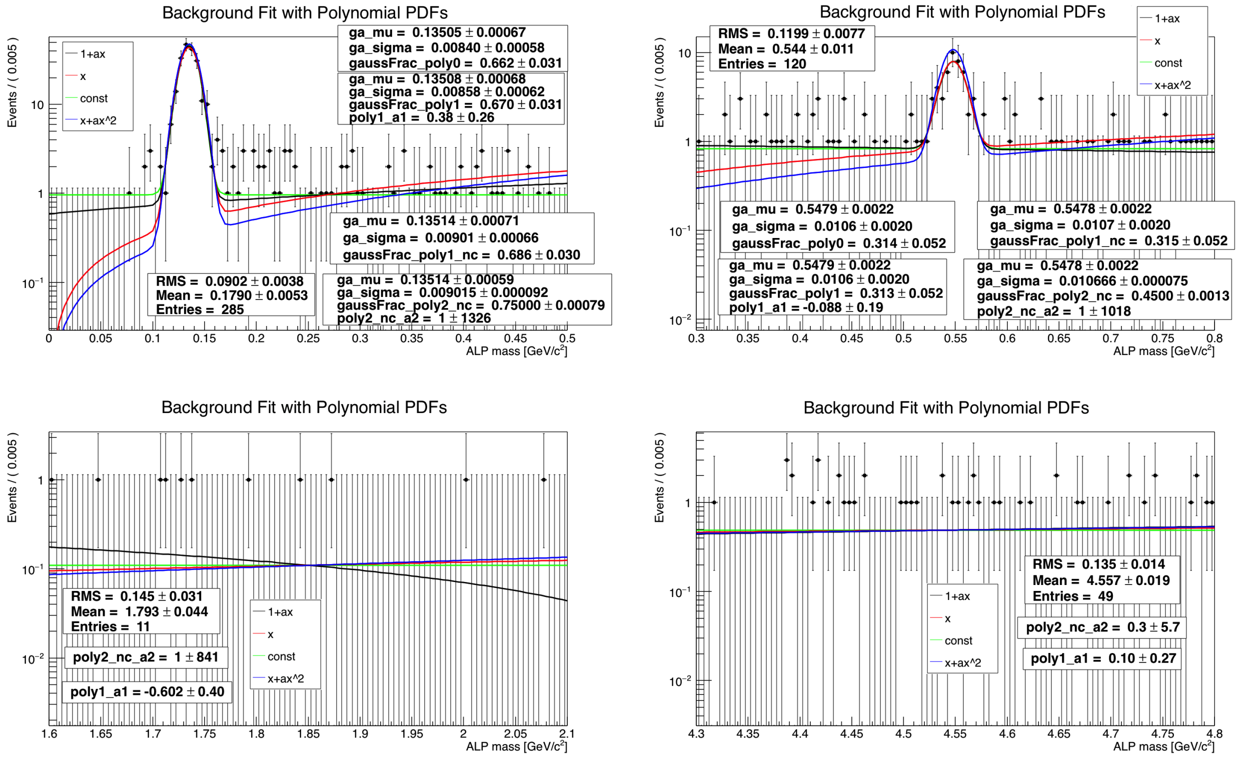


Figure 6.6: Events of the control sample, after the entire selection procedure, fitted with four different polynomial PDFs. The mass window has a constant width of 500 MeV/c^2.

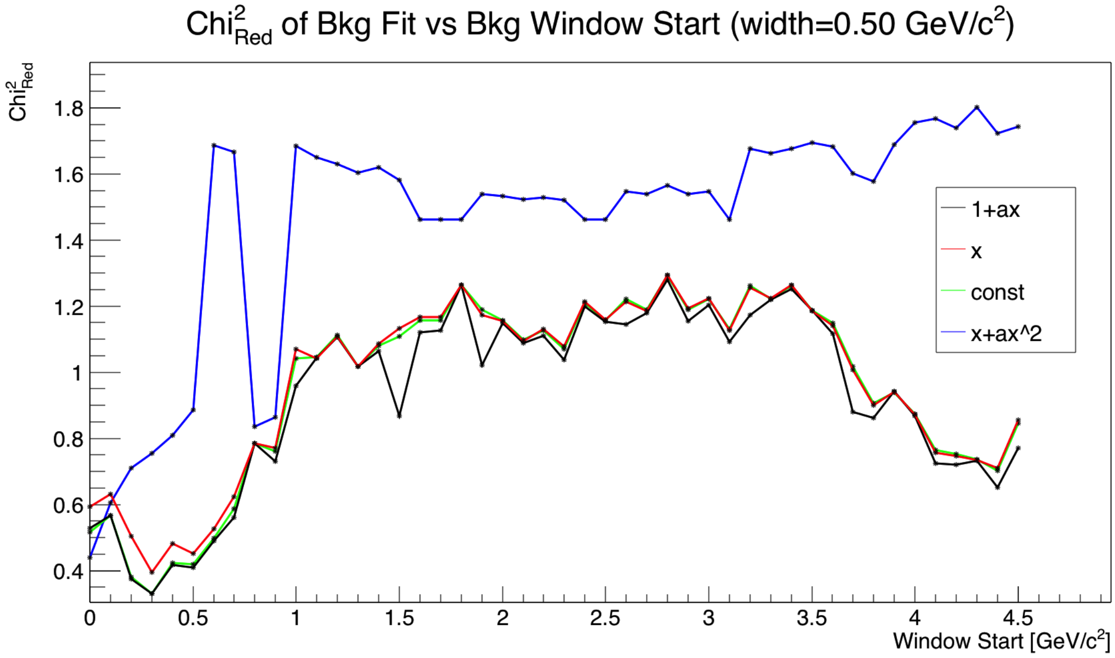


Figure 6.7: χ^2_{Red} for the fits performed with the four tested polynomial background PDFs, as a function of the starting point of the mass window. The mass window has a constant width of 500 MeV/c^2.

6.3.1 PDF for the Simultaneous Fit

The background PDF, obtained with the method described in Section 6.2, is composed by a low-degree polynomial and, in case the mass window includes one or more of the resonances of π^0, η, or $\eta\prime$, a Gaussian for each of them. The parameters of the polynomials and of the Gaussian are left floating.
The signal PDF, obtained with the method described in Section 6.1, is represented by a Crystal Ball whose parameters are functions of the analyzed mass hypothesis, and are fixed before performing the fit.

These two PDFs are added to form the PDF used for the simultaneous fit of signal+background.

When N PDFs are added, $N-1$ fraction are returned by the fitter, indicating the relative fraction of each component of the sum PDF. Generally these fractions are bounded to be between 0 and 1.
In this situation, instead, I constrained the signal fraction between -1 and 1. This is because I have assumed that the signal is represented only by statistical fluctuations of the actual number of events with respect to the number of events I would obtain from the background-only model, therefore I want to be able to grasp also the random deficit and not only the random excess of events.

Having allowed the signal fraction to range between -1 and 1, is possible to extract a negative number of events as well as a positive, and as said, we expect that, on average, the number of signal-interpreted events be 0, if it only comes from statistical fluctuation.

It is important to point out that a possible drawback of this procedure is that the fitter could try to make the sum PDF assume negative values. This can happen if the signal is sitting in a sufficiently large region devoid of events (i.e. a region larger than the width of the signal in that mass hypothesis): in this kind of situation, of which an example is visible in Figure 6.8, the overall PDF would go to negative values in that region, and this has to be avoided, because we cannot trust the returned signal fraction, and it is an issue related to the low statistic. The solution to overcome this problem is to iteratively increase the lower bound of the signal fraction (initially set to -1) and repeat the fit until the PDF is positive everywhere. The lower bound is increased by 0.02 at every cycle.

6.3.2 Mass Windows Selection

This section describes the procedure adopted to choose the sub-windows used for the simultaneous fits.

The main idea is to center the window on the signal mass hypothesis we are examining at that moment, perform the fit, retrieve the yield, increase the signal mass hypothesis and repeat the fit and the yield extraction, till the whole ALP mass range is scanned. This procedure is called *sliding*.
At each step the mass hypothesis is increased by quantity that is a fraction $1/d$ of the width

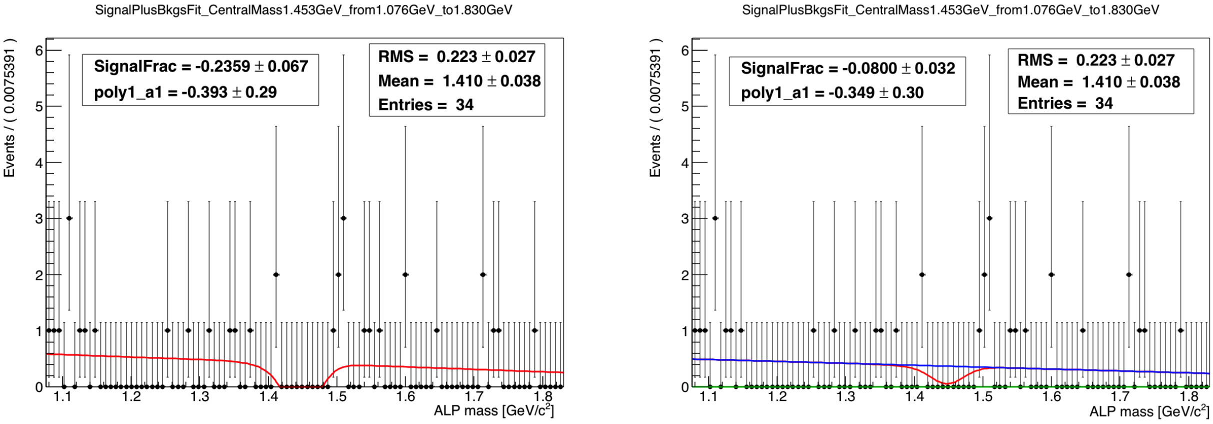


Figure 6.8: Example of a situation where, due to a large area devoid of events, the fraction of the signal PDF would want to go that negative that the sum PDF goes below zero. On the left it is depicted the situation without any correction, while on the right there is the same situation, with the method described in the text implemented. The red curve is the overall PDF; on the right plot, the blue curve is the polynomial-only PDF.

of the signal. If the increase was greater than the signal width, we could miss a signal peak while sliding through the spectrum. The increase is given by:

$$\Delta M = \frac{\sigma(M)}{d} \tag{6.8}$$

where $\sigma(M)$ is given by Equation 6.5. The parameter d has to be chosen such that there is no risk to miss a peak.

The width of the window is set to be equal to a multiple of the signal width in that point. In formula, given a mass hypothesis M, the mass interval W of the window is given by:

$$W = [M - K \cdot \sigma(M), M + K \cdot \sigma(M)] \tag{6.9}$$

where K is a parameter that has to be chosen to maximize the stability of the fit. As explained in Section 6.2.2, a larger window includes more events, and this can help in case of low statistics. It is not necessary to determine K in a strict and definitive way: it is possible that it will be changed when running on the complete sample, in order to enhance the stability of the fit on the actual data sample.

There are two situations where this definition of window is not feasible, i.e. at the beginning and at the end of the mass spectrum. If the mass hypothesis is close to the edges, the interval given by Equation 6.9 would include negative masses or masses beyond the allowed one; we are thus outside from the *sliding region*.

In these situations I used two *fixed* windows. For the case close to the beginning and to the end of the spectrum, respectively:

$$W_{beg} = [0, M_L] \ \text{GeV}/c^2 \tag{6.10}$$

$$W_{end} = [M_R, 5] \text{ GeV}/c^2 \tag{6.11}$$

Even if 5 GeV/c^2 is beyond the maximum achievable mass for the ALP (4786 MeV/c^2, see Equation 4.1), this conservative value has been chosen to include some combinatoric background and allow a smoother description of the background shape. As well as K, also M_L and M_R are parameters that are chosen to stabilize the fit. A loose constraint on M_L is that, if no other reasons prevents that, it is desirable to avoid to include two or more resonance peaks in the same window, both because adding another component to the background PDF makes the whole fit less stable, and because adding too many events, as the ones present in the peaks, in the same window makes the signal fraction became even tinier than it would be, possibly increasing the uncertainty over it.

There are two other parameters that have to be chosen, i.e. the values of the ALP mass M_{CL} where to exit from the fixed region defined in Equation 6.10 and enter in the sliding region, and similarly, the mass M_{CR} where to exit the sliding region and enter in the fixed region at the end of the spectrum, with the window given by Equation 6.11.
There are no strict constraints on the values M_{CL} and M_{CR}, except that we do not want to sharply pass from the first fixed region to the sliding one and from it to the second fixed region, i.e. we want some overlap between the first windows of the sliding region and the fixed regions. This is achieved requiring that $M_{CL} < M_L$ and $M_{CR} > M_R$.

As mentioned at the end of Section 6.2.2, narrow windows can be troublesome, because the background PDF could try to go negative in case of reduced statistic close to the edges of the window. This implies that K has to be big enough, order of $10 \div 20$.
M_{CL} and M_L have been chosen such that the first window includes the π^0 resonance but does not include the η peak, also such that the π^0 peak is not included in the sliding region. The full list of the chosen parameter is provided below:

- $M_L = 0.50$ GeV/c^2;
- $M_R = 3.20$ GeV/c^2;
- $M_{CL} = 0.35$ GeV/c^2;
- $M_{CR} = 4.00$ GeV/c^2;
- $K = 15$;
- $d = 2$.

These values have been used to perform the validity test of the fit on the control sample. It is worth to recall that they are not meant to be used also during the fit on the whole data set available for the analysis, as the fit on the complete data sample could present unexpected issues that could be solved slightly changing these values.

In Figure 6.9 are reported some fitted windows, obtained with the parameters listed above. In particular the width of the windows in the sliding region is equal to $\pm 15\sigma\,(M)$.

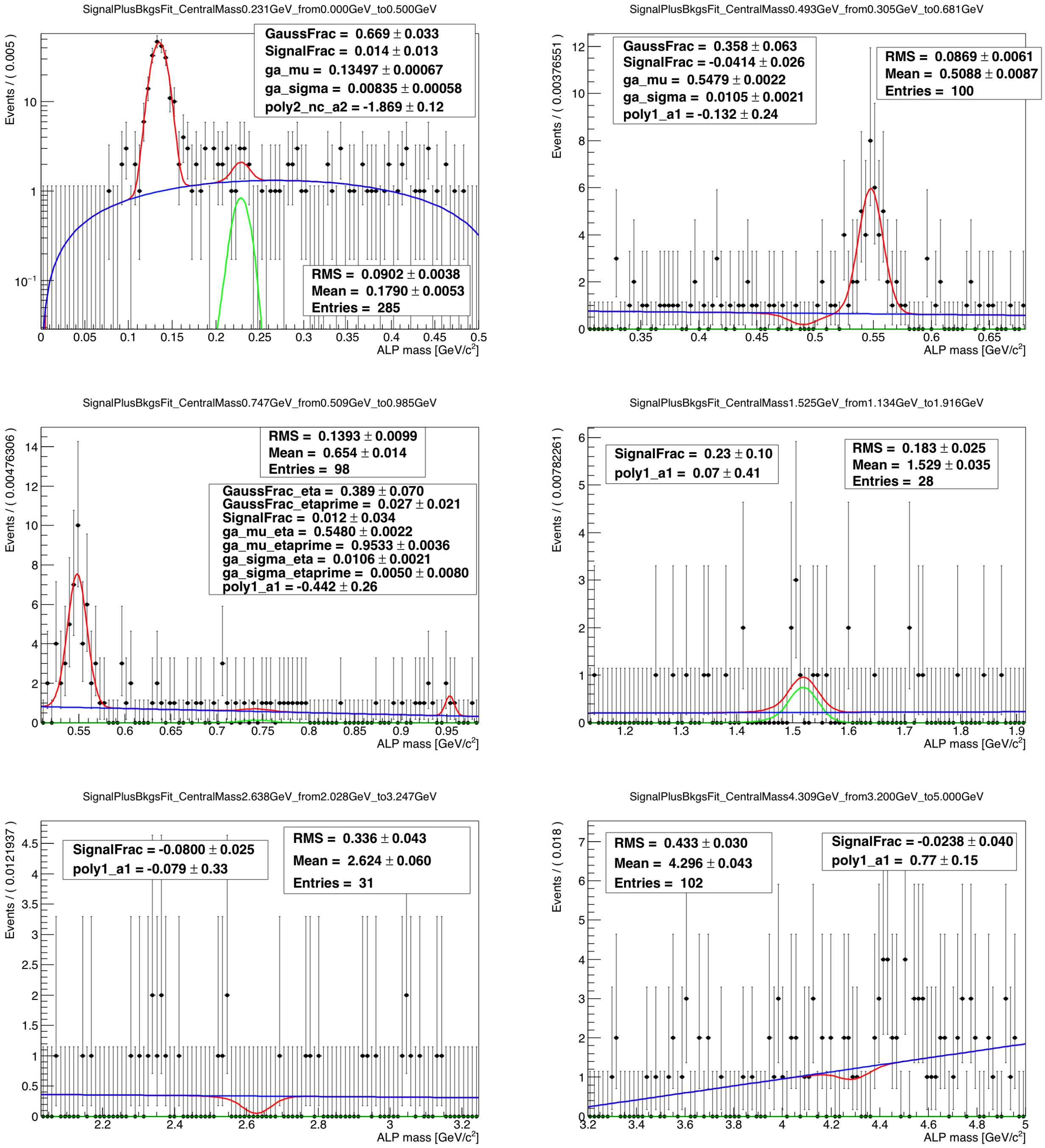


Figure 6.9: Simultaneous fits of signal+background on control sample data, performed on various windows, with the windows parameters described in the text. The blue, green, and red curves represent, respectively, the signal PDF, the polynomial background PDF, and the sum PDF. The signal curve is not visible in case of a negative signal fraction. The number of entries and the PDF parameters returned by the fit are reported for each shown window. The ALP mass hypothesis is in the center of the window, if we are in the sliding region.

6.3.3 Extracted Yield

The simultaneous fit procedure has been performed and the corresponding yield distribution extracted. The whole procedure has been repeated with different sets of windows parameters, to underline that they do not affect the results in a statistically significative way and they

are only relevant for what concerns the fit stability. For each set of parameters, the yield distribution has been fitted with a Gaussian. The results are reported in Figure 6.10.
The yield is computed as the PDF signal fraction multiplied by the number of events in the analyzed mass window.
With regard to the shown plots, the most distant one from the others, in terms of the parameters of the Gaussian, is the one obtained with a mass window of $\pm 20\sigma\,(M)$ in the sliding region: as mentioned, a wider window could bring fit instabilities.

We can see that, for each set of windows parameters, the fit at the yield distribution is compatible with a Gaussian, and its mean value is compatible with zero. This suggests that the fitting procedure, in the assumption of a negligible number of events present in the control sample, behaves correctly. It is not possible to further quantify the goodness of the procedure without having performed studies on Toy MC (see Section 7.1), because we cannot know if the fit introduces some kind of bias, and how to quantitatively evaluate the effect of the presence of a signal in the yield distribution, in terms of shape and mean value.

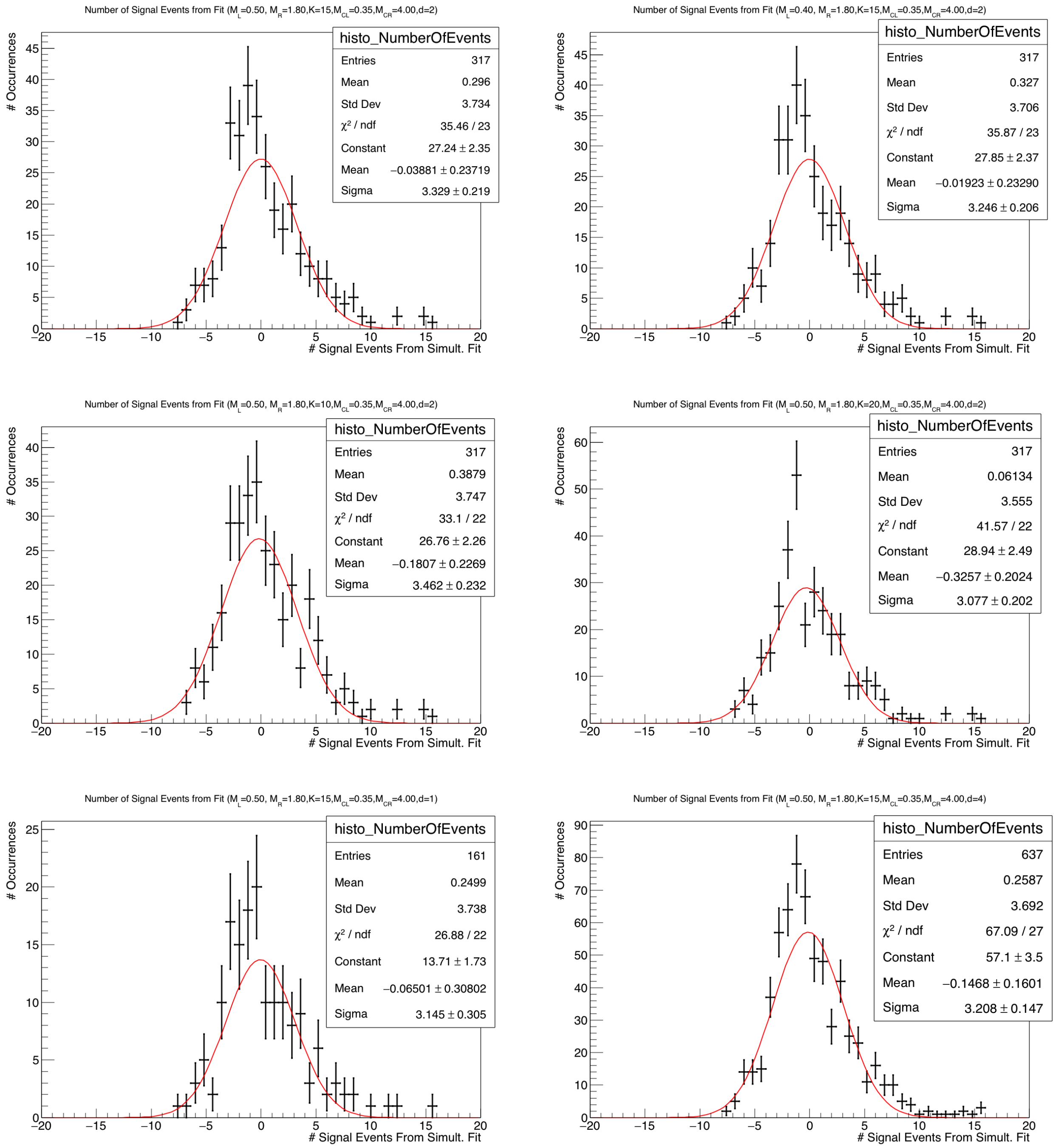


Figure 6.10: Signal yield distribution from the fits on the control sample, performed with different windows parameters. The distributions are fitted with a Gaussian. All the fit results are compatible with each other. On the top left there are the results obtained with the set of parameters described in Section 6.3.2; the top right derives from $M_L = 0.40$ GeV/c^2; in the middle row are shown the results for $K = 10, 20$; in the bottom line there are the distribution obtained with $d = 1, 4$.

Chapter 7

Further Studies and Future Prospects

As anticipated in Section 4.1, in this thesis work I completed the event selection and the signal fit procedures. In this Chapter the studies that still need to be thoroughly carried out to finalize the analysis are outlined, and an estimation on the sensitivity on the BR of the analyzed process is provided, for both the *BABAR* and the Belle II data samples.

Section 7.1 summarizes the studies that need to be performed with Toys MC to validate the fit, while a discussion about the main sources of systematic uncertainties is done in Section 7.2. A basic study on the sensitivity on the BR for the $B^{\pm} \to K^{\pm}A$ process with the *BABAR* data sample is reported in Section 7.3. This result is expected to significantly improve when using the data that will be collected by the Belle II experiment. An evaluation of the improvement is provided: Section 7.4 gives an insight on the characteristics of the data sample that will be available with Belle II, while in Section 7.5 the projection of my analysis on this sample will be estimated.

7.1 Fit Validation with Toy Monte Carlo

A Toy MC is a simulation aiming to study the behavior of a specific portion of a statistical analysis procedure.

In my case, the Toy MC studies are needed to validate the fitting procedure. As discussed at the end of Section 6.3.3, completely reliable quantitative information cannot be extracted from the mere signal yield distribution, as the exact response of the fit in the presence of a signal is not known. Also, it is necessary to validate that the fitted yield is not biased.

With a Toy MC it is possible to generate events distributed according to a given PDF. This procedure is already implemented in RooFit. What I need to do is to generate a number of events equal to the number of events that I expect to pass the selection procedure in the whole data sample (or in Run 3, if I want to validate the procedure carried out in Section 6.3). The PDF to be used is the sum of the background PDF and the signal PDF. This procedure has to be repeated for different mass hypotheses and different signal strength; the latter is expressed in term of the signal fraction.

Each of these simulated data samples undergoes the fit procedure, and the corresponding yield distribution is recorded for each of them. In this way it is possible to study the introduction of possible biases during the fit to the data. In particular the studies I am interested in are:

- Simulate a signal fraction X and verify whether the same fraction X is obtained from the fit, or a bias is introduced;
- For $X = 0$, i.e. the null hypothesis, I also want to record the maximum number of signal events I get from the scan over the whole spectrum and plot its distribution, obtained from re-generating the data sample a large number of times ($\sim 100 \div 1000$). This has to be done because observing a high number of signal events (i.e. 3 or more sigmas above background) in a single mass window is pretty significant, while observing the same 3 sigma discrepancy while scanning the entire mass spectrum is not so unlikely, since the number of analyzed mass windows is $\mathcal{O}(10^2 \div 10^3)$. This is called the *look-elsewhere effect*. The purpose of this study is to quantify this effect;
- Adopt different PDF models (like a constant function or a second degree polynomial even in the sliding region, and different parameters for the CB modeling the signal) and perform the fits again. The difference between the fit results obtained with the Toy PDFs and with the chosen PDFs is part of the systematic uncertainties (see Section 7.2).

7.2 Systematic Uncertainties

In this section an overview of the main sources of systematic uncertainties for this analysis is provided, as well as methods to study them. A quantitative evaluation is provided where possible.

- **Background PDF**: the choice of a particular function over another introduces a systematic uncertainty that can be evaluated, as anticipated in Section 7.1, with the help of Toys MC: new data samples are produced using different background PDFs than the ones described in Section 6.2. The difference between the yield extracted with the chosen PDFs and the yield extracted with the Toy PDFs is part of the systematic uncertainties. A similar evaluation can be performed directly on data, without generating the events with the Toys MC;
- **Signal PDF**: also the particular choice about the signal PDF is part of the systematic uncertainties. To estimate this contribution, other functions can be used (as a Gaussian, a double CB etc), and the parameters of the polynomial functions describing the trend of the parameters of the CB (described in Section 6.1.2) can be varied within their fit uncertainties. Also in this case, the difference in the results obtained with the chosen PDF and with the modified PDFs takes part of the systematic uncertainties;
- **Signal efficiency**: we extract the signal efficiency from the MC, but we cannot be sure of how good the MC models the signal; hence, parameters as the photon identification efficiency, tracking efficiency, kaon PID, and the efficiency of passing the MVA selection

will be part of the systematic uncertainties. These can be evaluated using the SM process perfectly mimicking the signal, i.e. looking at the peaks of π^0, η, and $\eta\prime$, and comparing the prediction of the MC for these channels with the actual data;

- **Luminosity**: the luminosity of each Run of *BABAR* has been measured with a certain uncertainty, between 0.40% and 0.45%, dominated by the systematic component (see Table 4.2 for the exact values).

7.3 Sensitivity Estimation

Given the current status of the analysis, a rough estimation of the sensitivity on the BR of the searched channel can be performed.
A more refined procedure will be carried out using Toy MC studies. MC dataset with different signal strength will be generated, and each will be fitted twice, once with the background only PDF and once with the signal+background PDF. The ratio of the two likelihood L will be computed as $LR_{Toy} \equiv \frac{L^{S+B}_{Toy}}{L^{B}_{Toy}}$ and its distribution built. This procedure will be repeated for each signal mass hypothesis. The data will be fitted both with the background only PDF and with the signal+background PDF and the corresponding LR_{Data} computed, for each mass hypothesis. The location of the value of LR_{Data} in the distributions of LR_{Toy} for different signal strength provides an exclusion limit for that given signal strength, and the value that gives a confidence level of 90% will be found.

For the basic estimation possible given the current status of the analysis, I assume the *null hypothesis*, i.e. no events are present in the data sample. In this situation, I want to estimate an upper limit on the signal strength, within a certain confidence level, and then convert it into a BR upper limit.

Let's assume to have a Poisson counting experiment where the mean number of background events is F. If n events are observed, we can construct an upper limit S_{up} on the mean number of signal event S within a certain *credibility interval* CI. This is given by [45]:

$$CI = \frac{\int\limits_0^{S_{up}} \left[(S+F)\, e^{-(S+F)}/n! \right] P(S)dS}{\int\limits_0^{\infty} \left[(S+F)\, e^{-(S+F)}/n! \right] P(S)dS} \tag{7.1}$$

where $P(S)$ is the prior probability for S. If $P(S)$ can be considered constant for each S, then it can be demonstrated that the above formula becomes:

$$1 - CI = e^{-S_{up}} \frac{\sum\limits_{m=0}^{n} (S_{up}+F)^m/m!}{\sum\limits_{m=0}^{n} F^m/m!} \tag{7.2}$$

The equations reported above are obtained in a Bayesian assumption, and this implies that the possible number of background events cannot be greater than the total number of events observed in this measurement. S_{up} can be interpreted as the limit on the number of average signal events, present in the experiment, that make the integrated probability of observing n events no greater than CI, when F background events are present in average.

To get a rough estimation of the BR sensitivity, I treated this analysis as a counting experiment in the mass region of interest. I looked at the events of the control sample that have passed the entire selection procedure, and for a given ALP mass hypothesis M_0 I counted the number of events in a range $[M_0 - 3\sigma(M_0), M_0 + 3\sigma(M_0)]$ (the function $\sigma(M)$ is the one described in Section 6.1.2), and I assumed that all of them are background events, i.e. $n = F$. I repeated this procedure scanning the ALP mass spectrum with the the same mass step ΔM reported in Section 6.3.2, i.e. $\frac{\sigma(M)}{2}$. I excluded the events close to the resonances of π^0, η, and $\eta\prime$.
The number of detected signal events X can be calculated as follows:

$$X = 2N_{B^+B^-}\mathrm{BR}(B^\pm \to K^\pm A)\epsilon_S \tag{7.3}$$

where ϵ_S is the signal efficiency of the selection (see Figure 5.21), the factor 2 is due the fact that each of the two produced B mesons can decay into KA, and $N_{B^+B^-}$ is the number of B^+B^- pairs in the given data sample. $N_{B^+B^-}$ can be computed as

$$N_{B^+B^-} = \mathcal{L}\sigma_{\Upsilon(4S)}\mathrm{BR}\left(\Upsilon\left(4S\right) \to B^+B^-\right) \tag{7.4}$$

The $\Upsilon(4S)$ BRs in B^+B^- and $B^0\bar{B}^0$ are, respectively, 51.4% and 48.6%, but they will be approximated to 50% for the purpose of this estimation, hence $N_{B^0\bar{B}^0} \simeq N_{B\bar{B}}/2$.
I retrieved the number of $B\bar{B}$ pairs from the *BABAR* database (using former estimations from data, see Table 4.2).

To connect the number of events found in the control sample (the Run 3) with the number of events expected to be found in the whole analysis dataset (Runs 1, 2, 4, 5, 6), I scaled them with the ratio of the numbers of $B\bar{B}$ events in the two samples:

$$f = \frac{N_{B\bar{B}}^{\text{Runs 12456}}}{N_{B\bar{B}}^{\text{Run 3}}} = 12.17 \tag{7.5}$$

The drawback of this choice is the amplification of the statistical fluctuations of the control sample. The other possibility was to use MC simulated data, but I preferred using real data because, as said, the MC underestimates the number of events.

The limit for the BR at a mass hypothesis M_0, assuming the null hypothesis, is hence given by:

$$\mathrm{BR}(M_0) = \frac{S_{up}\left(f \cdot F^{\text{Run 3}}, CI\right)}{fN_{B\bar{B}}^{\text{Run 3}}\epsilon_S} \tag{7.6}$$

where $S_{up}\left(F, CI\right)$ is the function implicitly defined in Equation 7.2, setting $n = F$, and evaluated with a numerical method. Figure 7.1 shows the trend of BR setting $CI = 90\%$.

The estimated sensitivity on the BR is between $\mathbf{10^{-7}}$ and $\mathbf{5\times10^{-7}}$. This can be compared

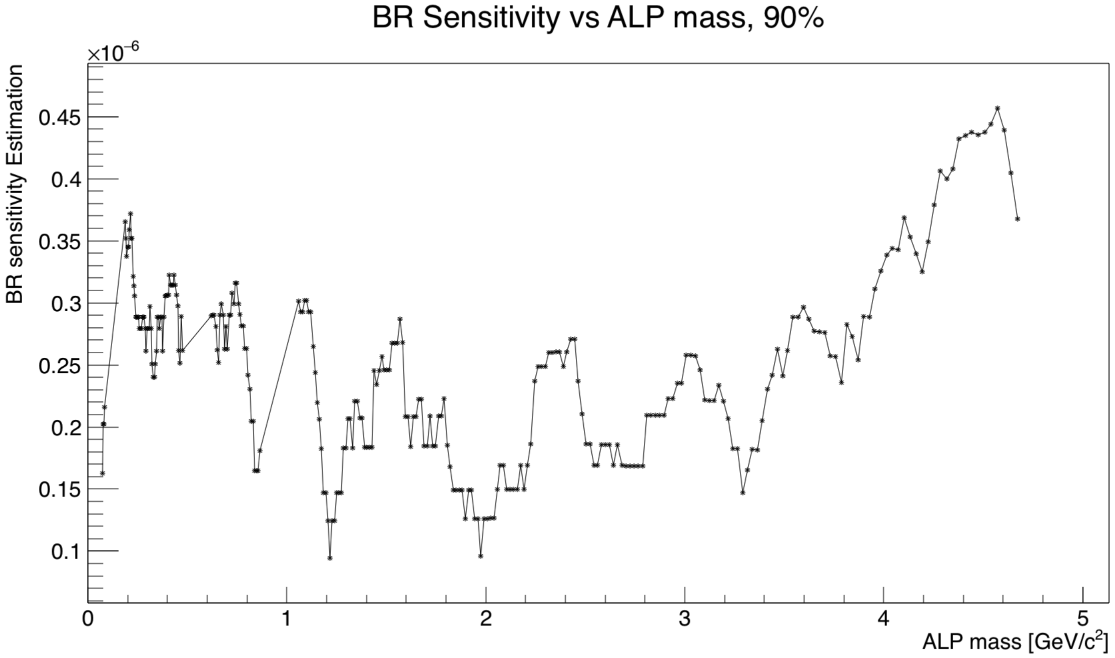


Figure 7.1: BR limits estimation, with a credibility interval of 90%, obtained with the procedure described in the text, projected on the whole analysis data sample available in *BABAR*. Note the holes corresponding to the vetoed masses of the three resonances π^0, η, and $\eta\prime$. The large fluctuations are a consequence of the low statistic of the control sample.

with the exclusion plot in Figure 1.5, expressed in terms of g_{AW}.
This result can be extrapolated also to the upcoming Belle II data sample, as described in Section 7.5.

7.4 Data Samples with SuperKEKB

Belle II will have two main data taking periods, called *Phase 2* and *Phase 3*.

Phase 2 will be mainly dedicated to study the response of the accelerator and of the detectors. The beams will have to be tuned to reach the design luminosity. The response of all the subdetectors of Belle II, as well as the response of the entire detector, will be studied. The data collected during this phase will provide fundamental information to validate the simulation and the luminosity estimation, to study the systematic uncertainties and the actual background we will have to face, and to perform calibration and material mapping.
The luminosity will be raised till the peak value of KEKB, 2×10^{34} cm^{-2} s^{-1}, with the beam currents that will go up to $\sim 50\%$ of the maximum value registered at KEKB (see Table 2.3). The Belle II detector will not be complete during this phase: PXD (Section 2.4.1) and SVD (Section 2.4.2) will only have one ladder each. This will only provide a $\sim 10\%$ angular coverage in the angle ϕ. Additional detectors will be placed inside Belle II to monitor the radiation levels.

This phase will also be precious for physics analyses, as the reduced background with respect to the nominal expectation for Phase 3 will allow to adopt a single-photon trig-

ger, enabling low-multiplicity analyses, particularly interesting for DM and invisible decays searches. It is currently estimated that approximately 20 fb^{-1} of collision data at the $\Upsilon(4S)$ energy will be collected.
Phase 2 will last from April to July 2018.

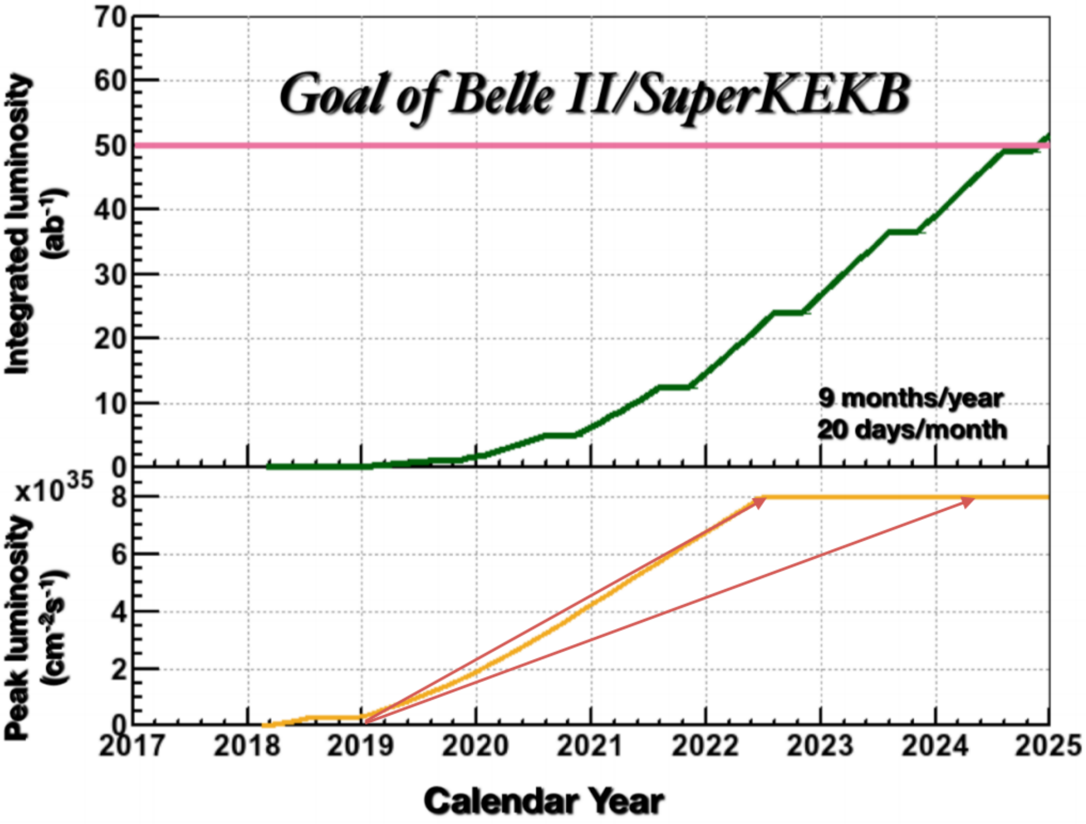


Figure 7.2: Integrated and instantaneous luminosity of SuperKEKB as a function of the data taking time. The goal of 50 fb^{-1} should be reached within 2024. Image from [46].

Phase 3 will start in early 2019. The complete detector will be mounted. The luminosity will increase up to the nominal value of 8×10^{35} cm^{-2} s^{-1}, almost 70 times the value of *BABAR*. Belle II is planned to reach the integrated luminosity of $\sim$ 50 ab^{-1} within the end of the data taking, that will last for 6 years (see Figure 7.2). At least 70% of these data will be collected on-peak at the energy of the $\Upsilon(4S)$.

Their main features of the accelerator for the two phases are reported in Table 7.1.

7.5 Projections on Belle II

As reported in Section 4.4.5, the integrated luminosity available at *BABAR* for this analysis is equal to 391.91 fb^{-1}. This means that the data from Belle II Phase 2 will be irrelevant, as the integrated luminosity for this phase is planned to be 1/20 of the one already available, so the gain in terms of statistic is negligible; also, given the partial angular coverage of the PXD and the SVD, the tracking performances are reduced, hence the identification efficiency of the charged kaon will be negatively affected, in particular for the low-momentum K, associated

Parameter	**Unit**	**Phase 2**	**Phase 3**
Bunch crossing angle	mrad	± 41.5	
Lorentz factor	$\beta\gamma$	0.28	
e^- beam energy	GeV	7.0	
e^+ beam energy	GeV	4.0	
e^- beam current	A	0.8	2.62
e^+ beam current	A	1.0	3.60
Luminosity	$\mathrm{cm}^{-2}\mathrm{s}^{-1}$	1.2×10^{34}	8×10^{35}
Integrated Luminosity	fb^{-1}	20	50×10^3
	x μm	n/a	10.2 [7.75]
e^+ [e^-] beam size	y nm	n/a	59 [59]
	z mm	n/a	6.0 [5.0]

Table 7.1: Machine parameters for SuperKEKB during Phase 2 and Phase 3.

with high mass ALPs.

On the other hand, the data from Phase 3 will be highly valuable, and we can hope to significantly extend the exclusion region.

The main advantage comes from the higher statistic. Assuming the pessimistic lower estimation that only 70% of the 50 ab^{-1} will be collected at the on-peak energy of the $\Upsilon(4S)$, we obtain that at least 35 ab^{-1} of data will be usable for the analysis described in this thesis.

The lower limit on the improvement in statistic with respect to *BABAR* data sample is hence given by:

$$\frac{\mathcal{L}^{\mathrm{BelleII}}_{\Upsilon(4S)}}{\mathcal{L}^{\mathrm{BaBarR12456}}_{\Upsilon(4S)}} \simeq \frac{35000}{392} \simeq 90 \tag{7.7}$$

The fluctuation on the number of events in a given mass region is given by the square root of the number of events itself, as it follows a Poisson distribution. This means that using the complete Belle II data sample we can gain at least a factor $\sqrt{90} \simeq 9.5$ in precision due to the increased statistic.

This could be verified applying the exact same procedure described in Section 7.3, with the only change being the factor f, that for this evaluation would be equal to:

$$f = \frac{\mathcal{L}^{\mathrm{BelleII}}}{\mathcal{L}^{\mathrm{BaBarR3}}} \simeq \frac{35000}{32} \simeq 1100 \tag{7.8}$$

The definition is equivalent to the one used in Equation 7.5.

I have not performed this direct check because the calculation of the upper limit on the number of events S_{up} through the Equation 7.2 quickly becomes computationally very expensive when the number of background events F increases. An indirect, quicker check can be performed verifying that the number S_{up}, as a function of the number of background events F, grows as a square root. This can be seen in Figure 7.3.

The limit on the BR of the process $B^\pm \to K^\pm A$ can therefore be reduced by a factor 9.5 with respect to the limit obtainable from the *BABAR* dataset, taking into account only the

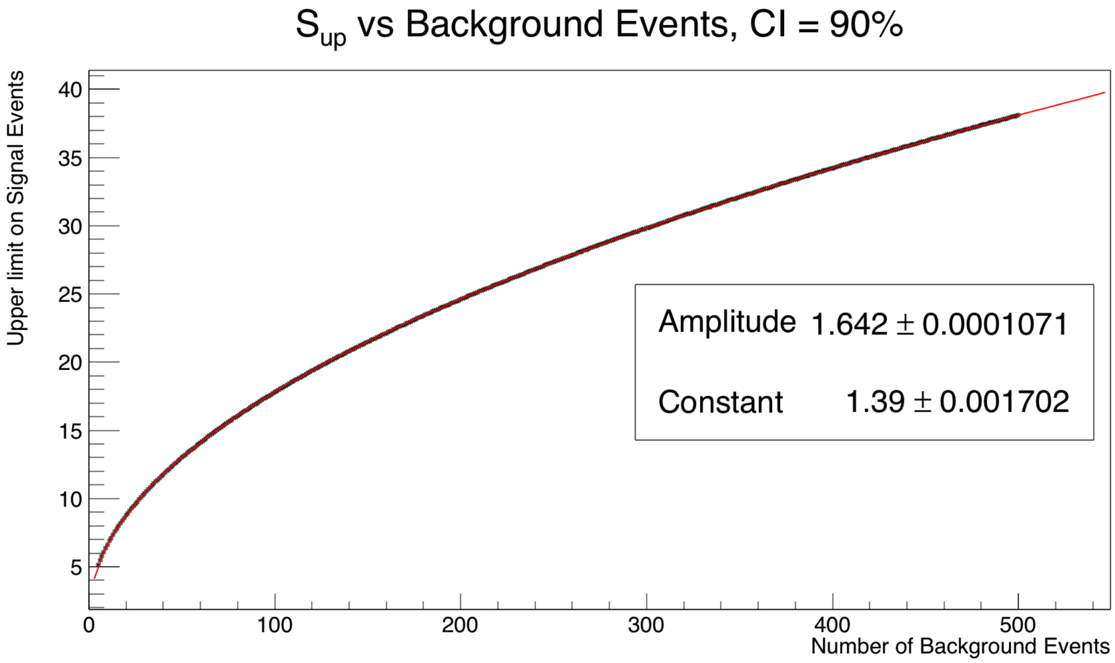


Figure 7.3: Trend of S_{up} as a function of the number of the background events, as retrieved from Equation 7.2, requiring $CI = 0.9$. The fit function is $C + A\sqrt{F}$, where F is the number of background events. The uncertainties on the parameters C and A returned from the fit are not meaningful as each point is errorless.

improvement deriving from the increased statistic, reaching the limit of:

$$\text{BR}^{\text{BelleII}} \sim 3 \times 10^{-8} \tag{7.9}$$

The main improvement in the analysis' results will come from the increased statistics, but also the higher quality of the Belle II data sample with respect to its predecessors *BABAR* and Belle could play in role in extending the BR sensitivity. The main features that could help extending the excluded parameter region of the ALP are the following:

- **Photon energy resolution**: *BABAR* EMC photon energy resolution was such that the resolution on the π^0 mass was 7 MeV/c^2; Belle achieved a resolution of 4.5 MeV/c^2. Belle II will use the same ECL of Belle, with an improved electronics, but on the other hand will have to face a much higher background. Proper evaluations are not yet available, but it is reasonable to have a resolution at least not worse than the one achieved by Belle. A better resolution on the diphoton energy would reduce the width of the signal and hence lead to a greater signal-over-background ratio in the analyzed mass region;

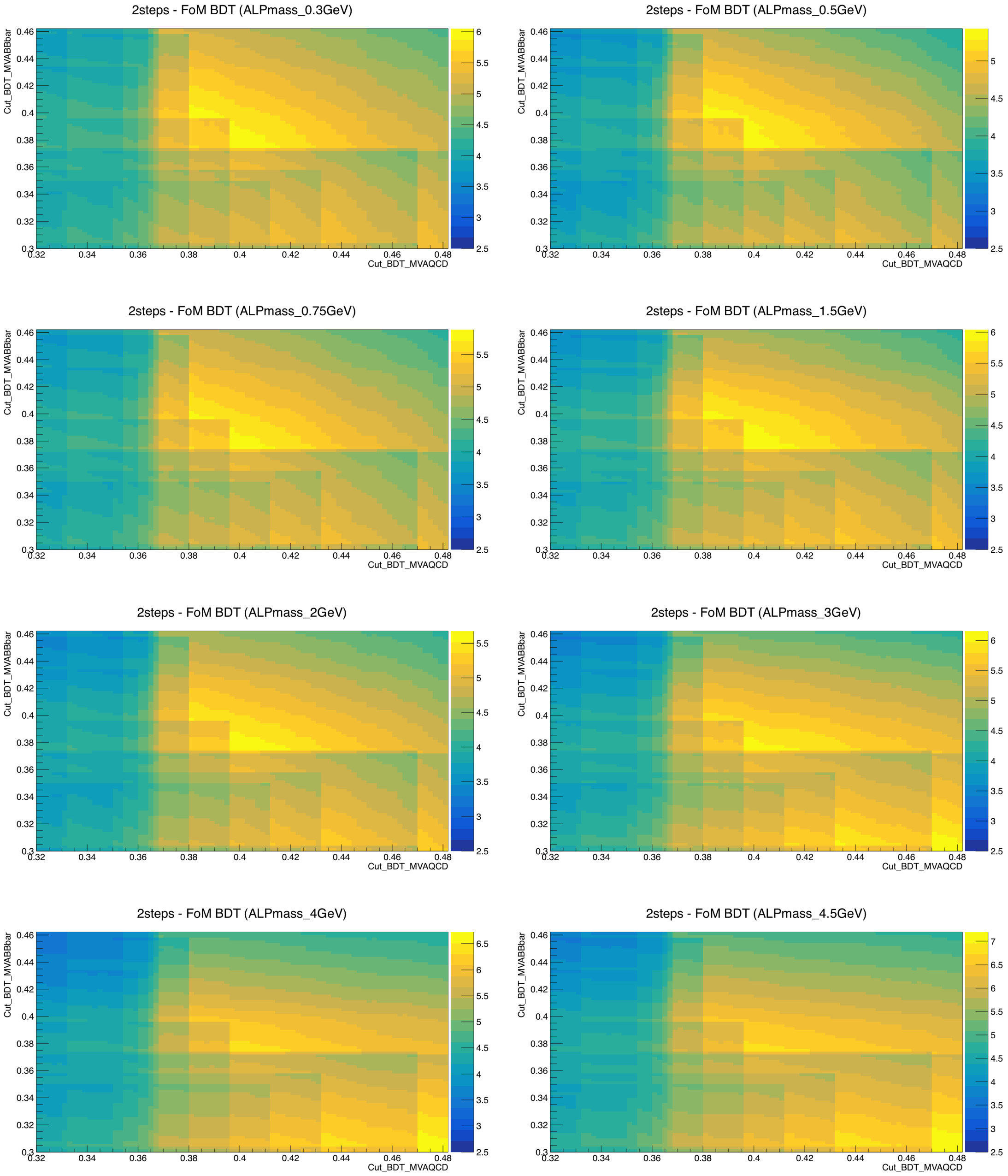
- **Angular coverage**: the *BABAR* angular coverage in angle θ was from 20° to 150°, while Belle II will cover the angles from 17° to 150°. The halved Lorentz boost factor (from 0.56 to 0.28) will also increment the effective angular coverage;

- **Particle identification**: the *BABAR* DIRC detector (see Section 2.3.3) only covered the barrel region, while Belle II will have two different subdetectors to perform PID both in the barrel and in the endcap region (see Section 2.4.4). Also, the different particle classification scheme could help improving the discrimination between charged K and π, even if the studies on the performance of the Belle II lists still have to be completed.

Appendix A

Stability of the Selection in the Cuts Space

In Section 5.3.1 the applied selection procedure is described. It is opportune to verify the stability of the FoM χ (defined in Section 5.2.1) in the MVAs cuts space, i.e. that the chosen maximum is not too narrow. This can be done plotting χ as a function of the MVA cuts for each evaluated signal mass. The results can be seen in Figure A.1, and we can see that the maximum is sometimes just a local maximum (as expected since we used an interpolated couple of cuts), but at the same time the area with high FoM is not narrow.



2steps - FoM BDT (ALPmass_0.3GeV)

2steps - FoM BDT (ALPmass_0.5GeV)

2steps - FoM BDT (ALPmass_0.75GeV)

2steps - FoM BDT (ALPmass_1.5GeV)

2steps - FoM BDT (ALPmass_2GeV)

2steps - FoM BDT (ALPmass_3GeV)

2steps - FoM BDT (ALPmass_4GeV)

2steps - FoM BDT (ALPmass_4.5GeV)

Figure A.1: Distribution of the FoM χ in the MVA cuts space for the different evaluation mass samples. The used interpolated couple of cuts is $\text{cut}_{\text{MVA}_{\text{uds}}} = \mathbf{0.397}$, $\text{cut}_{\text{MVA}_{\text{BpBm}}} = \mathbf{0.375}$, and is the central point of each plot.

Appendix B

Independence of the Crystal Ball Fit from the Parameter n

In this Appendix are reported the studies on the independence of the Crystal Ball parameters α and σ from the parameter n (see Section 6.1.2).
I show the trends of the two studied parameters as a function of the ALP mass M, for four different settings of the parameter n:

- $n = 50$;
- $n = 80$;
- $n = 120$;
- n free.

This study is shown in Figure B.1 for σ and Figure B.2 for α. The values of the fitted polynomial coefficients are shown on the Figures.
No statistically significant effect is seen with the different choices of n.

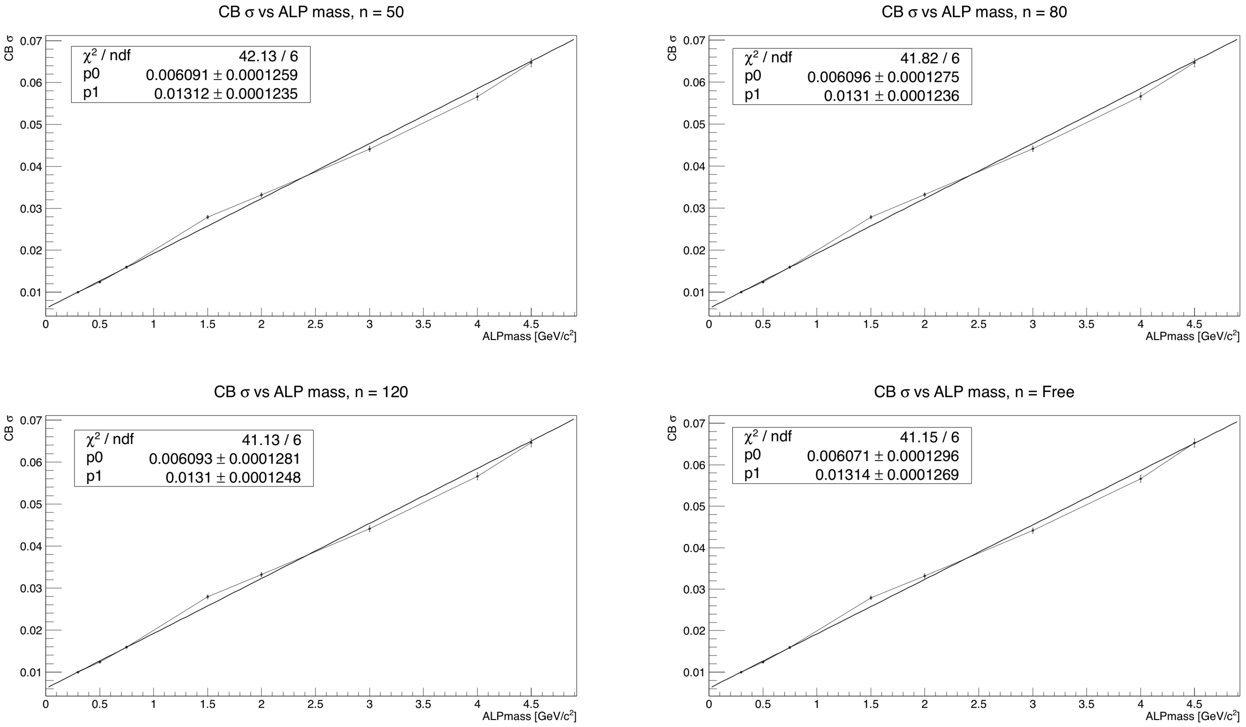


Figure B.1: Trend of the CB parameter $\sigma\,(M)$, as a function of the ALP mass M, extracted from fits performed with different fixed n, or without fixing it. The points have been fitted with a 1st degree polynomial. No statistically relevant difference is present between the different curves.

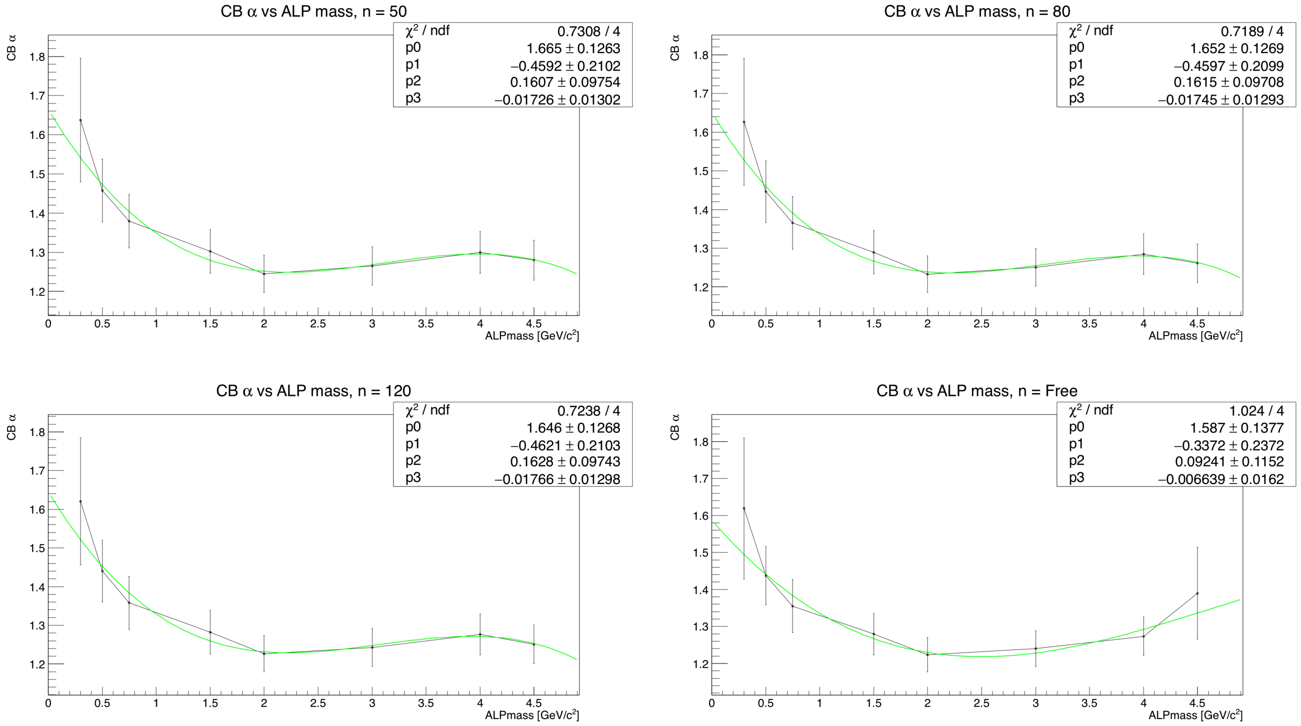


Figure B.2: Trend of the CB parameter $\alpha\,(M)$, as a function of the ALP mass M, extracted from fits performed with different fixed n, or without fixing it. The points have been fitted with a 3rd degree polynomial. No statistically relevant difference is present between the different curves. The last case, obtained from fits performed without fixing n, show a different trend for the ending part of the curve $\alpha\,(M)$, being the last point different from the ones obtained fixing n. Nonetheless, the points are statistically compatible, as well as the curves for α.

Bibliography

[1] P. Gagnon, *The Standard Model: a beautiful but flawed theory* `https://www.quantumdiaries.org/2014/03/14/the-standard-model-a-beautiful-but-flawed-theory/`.

[2] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π decay in the* K_2^0 *meson* Phys. Rev. Lett. 13 (1964) 138-140 `https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.13.138`.

[3] C.T. Chan, *On the Symmetry Constraints of CP Violations in QCD*, arXiv:hep-ph/9704427 `https://arxiv.org/pdf/hep-ph/9704427.pdf`.

[4] C. Patrignani et al., (Particle Data Group), *Chinese Physics C*, 40, 100001 (2016) `http://pdg.lbl.gov/2017/download/rpp2016-Chin.Phys.C.40.100001.pdf`.

[5] B. Yoon, T. Bhattacharya, and R. Gupta, *Neutron Electric Dipole Moment on the Lattice*, arXiv:1712.08557 [hep-lat] `https://arxiv.org/pdf/1712.08557.pdf`.

[6] R.D. Peccei and H. Quinn, *CP Conservation in the Presence of Pseudoparticles*, Phys. Rev. Lett. 38(25) 1440-1443 `https://www.researchgate.net/publication/248549883_CP_Conservation_in_the_Presence_of_Pseudoparticles`.

[7] R.D. Peccei and H. Quinn, *Constraints imposed by CP conservation in the presence of pseudoparticles*, Phys. Rev. D 16, 1791 `https://journals.aps.org/prd/pdf/10.1103/PhysRevD.16.1791`.

[8] R. Barbieri et al., *Searching for galactic axions through magnetized media: the QUAX proposal*, arXiv:1606.02201 [hep-ph] `https://arxiv.org/pdf/1606.02201.pdf`.

[9] G. Carugno (responsible), *Development of spectroscopy techniques on cold system for the detection of Cosmological Axions*, INFN document `http://www.pd.infn.it/segreterie/segre1/CALLG5AXIOMA20luglioOre1411.pdf`.

[10] I. Stern, *ADMX status*, arXiv:1612.08296 [physics.ins-det] `https://arxiv.org/pdf/1612.08296.pdf`.

[11] J. Ferretti, *DSTAX. An Axion-like Particle Search with Microwave Photons*, arXiv:1609.05105 [hep-ph]
`https://arxiv.org/pdf/1609.05105.pdf`.

[12] Konstantin Zioutas (spokesperson), *CERN Axion Solar Telescope*, CAST website
`http://cast.web.cern.ch/CAST/`.

[13] J. K. Vogel et al., *IAXO - The International Axion Observatory*, arXiv:1302.3273 [physics.ins-det]
`https://arxiv.org/pdf/1302.3273.pdf`.

[14] S. Aefsky et al., *Fundamental Physics at the Intensity Frontier*, arXiv:1205.2671 [hep-ex]
`https://arxiv.org/pdf/1205.2671.pdf`.

[15] J. Isern, E. Garcia–Berro, S. Torres, and S. Catalan, *Axions and the cooling of white dwarf stars*, Astrophys J. 682 (2008) L109
`https://arxiv.org/pdf/0806.2807.pdf`.

[16] A. De Angelis, M. Roncadelli, and O. Mansutti, *Evidence for a new light spin-zero boson from cosmological gamma-ray propagation?*, Phys. Rev. D 76, 121301(R)
`https://arxiv.org/pdf/0707.4312.pdf`.

[17] L. Calibbi, F. Goertz, D. Redigolo, R. Ziegler, and J. Zupan, *The axiflavon*, arXiv:1612.08040 [hep-ph]
`https://arxiv.org/pdf/1612.08040.pdf`.

[18] K. Choi, S. Hui Im, C. Beom Park, and S. Yun, *Minimal Flavor Violation with Axion-like Particles*, JHEP 11 (2017) 070
`https://link.springer.com/article/10.1007/JHEP11(2017)070`.

[19] E. Izaguirre, T. Lin, and B. Shuve, *A New Flavor of Searches for Axion-Like Particles*, Phys. Rev. Lett. 118, 111802
`https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.118.111802`.

[20] Makoto Kobayashi and Toshihide Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics, Vol 49, Issue 2, Feb 1973, 652–657
`https://doi.org/10.1143/PTP.49.652`.

[21] Ed. A.J. Bevan, B. Golob, Th. Mannel, S. Prell, and B.D. Yabsley, *The Physics of the B Factories* Eur. Phys. J. C74 (2014) 3026, SLAC-PUB-15968, KEK Preprint 2014-3.
`https://arxiv.org/pdf/1406.6311.pdf`.

[22] T. Abe et al., *Belle II Technical Design Report*, arXiv:1011.0352 [physics.ins-det]
`https://arxiv.org/pdf/1011.0352.pdf`.

[23] D. Boutigny et al., *The BABAR physics book: Physics at an asymmetric BB factory*, SLAC-R-504
`http://inspirehep.net/record/479871/files/slac-r-504.pdf`.

[24] A. Bernard et al., *The BABAR detector*, Nucl. Instrum. Meth., A479, 2002, 1-116 `https://www.sciencedirect.com/science/article/pii/S0168900201020125`.

[25] K. Nakamura et al., *The Belle II SVD detector*, 25th International Workshop on Vertex Detectors (Vertex 2016), La Biodola, Elba Island, Livorno, Italy, 26-30 September 2016 `http://inspirehep.net/record/1615382?ln=en`.

[26] K. Nakamura, *The Belle II SVD Detector* 25th International Workshop on Vertex Detectors (Vertex 2016), La Biodola, Elba Island, Livorno, Italy, 26 September 2016 `https://indico.cern.ch/event/452781/contributions/2297519/attachments/1342772/2024059/160925-BelleIISVD_VERTEX2016_Katsuro.pdf`.

[27] M. L. French et al., *Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker*, Nucl. Instrum. Meth. A466 (2001) 359-365 `http://enoah.home.cern.ch/enoah/Publications/Named-Contributing-Author/2001_French_NIM_A_2001.pdf`.

[28] G. Lutz, *Semiconductor Radiation Detectors: Device Physics*, Springer, Berlin (2007), ISBN 3540716785.

[29] S. Ramo, *Currents Induced by Electron Motion*, Proceedings of the IRE, 27(9) (1939) 584–585 `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1686997`.

[30] M. Valentan, *The Silicon Vertex Detector for b-tagging at Belle II*, PhD thesis `http://repositum.tuwien.ac.at/obvutwhs/download/pdf/1635100?originalFilename=true`.

[31] G. Casarosa, *Readiness for Phase 2 and preparation status for Phase 3*, BPAC focused review on VXD, KEK, 15-17 October 2016 `https://kds.kek.jp/indico/event/24363/session/5/contribution/11/material/slides/0.pdf`.

[32] M. Friedl, T. Bergauer, A. Frankenberger, I. Gfall, C. Irmler, and M. Valentan, *The Belle II Silicon Vertex Detector readout chain*, Journal of Instrumentation, Vol 8, Feb 2013 `http://iopscience.iop.org/article/10.1088/1748-0221/8/02/C02037/pdf`.

[33] G. Mohanty, *Belle II Silicon Vertex Detector*, 26th International Workshop on Vertex Detectors, Las Caldas, Asturias, September 10-15 2017 `https://indico.cern.ch/event/627245/contributions/2676702/attachments/1522193/2378529/Vertex2017-Gagan.pdf`.

[34] R. Turchetta, *Spatial resolution of silicon microstrip detectors*, Nucl. Instrum. Meth. A 335 (1993) 44-58.

[35] The following links can only be accessed by *BABAR* members. They are meant to provide insight to the reader willing to perform a similar analysis. `https://bbr-wiki.slac.stanford.edu/bbr_wiki/index.php/Physics/Neutrals`; `https://bbr-wiki.slac.stanford.edu/bbr_wiki/index.php/Detector_Info`; `https://bbr-wiki.slac.stanford.edu/bbr_wiki/index.php/Physics/PID`; `https://bbr-wiki.slac.stanford.edu/bbr_wiki/index.php/Physics/PID/PID_Selector_List`; `http://www.slac.stanford.edu/BFROOT/www/Physics/Tools/Pid/Selectors/r24c/SuperTightKMKaonMicroSelection_details.html`.

[36] J.P. Less et al., *Time-Integrated Luminosity Recorded by the BABAR Detector at the PEP-II* e^+e^- *Collider*, Nucl. Instrum. Meth. A726 (2013) 203-213 `https://arxiv.org/pdf/1301.2703.pdf`.

[37] A. Davis, *Looking Forward to 2015: Analysis Techniques* `http://www.quantumdiaries.org/tag/helicity/`.

[38] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *TMVA - Toolkit for Multivariate Data Analysis with ROOT*, PoS ACAT 040 (2007) `http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf`.

[39] ROOT homepage `https://root.cern.ch/`.

[40] N. Chanon, *Multivariate analysis in high energy physics - Lecture 2*, lecture slides `https://people.phys.ethz.ch/~pheno/Lectures2012_StatisticalTools/slides/Chanon2.pdf`.

[41] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, Singapore, 2006 `http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf`.

[42] A. Fisher, *The use of multipole measurement in taxonomic problems*, Annals of Eugenics, 7: 179-188 (1936) `https://digital.library.adelaide.edu.au/dspace/bitstream/2440/15227/1/138.pdf`.

[43] W. Verkerke and D. Kirby, *RooFit Users Manual v2.07* `http://roofit.sourceforge.net/docs/RooFit_Users_Manual_2.07-29.pdf`.

[44] M. J. Oreglia, *A study of the reactions* $\psi\prime \rightarrow \gamma\gamma\psi$, 10.1103/PhysRevD.25.2259 `http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-236.pdf`.

[45] S. D. Biller and S. M. Oser, *Another Look at Confidence Intervals: Proposal for a More Relevant and Transparent Approach*, Nucl.Instrum.Meth. A774 (2015) 103-119 `https://arxiv.org/pdf/1405.5010.pdf`.

[46] P. Urquijo, *Physics*, 29th B2GM, KEK, 5-9 February 2018 `https://kds.kek.jp/indico/event/26522/session/47/contribution/35/material/slides/0.pdf`.

Ringraziamenti

Mi risulta non banale scegliere l'ordine col quale vergare questa sezione, piccola anomalia solitaria nella quale posso permettermi un gusto più letterario e, dettaglio non trascurabile, l'utilizzo di una lingua nella quale ancora mi destreggio meglio che non l'inglese, per quanto apprezzi quest'ultimo.

Credo che opterò quindi per una sequenza approssimativamente dettata da ragioni cronologico-categoriche, senza però impormi di esservi rigidamente ligio.

Comincio pertanto ringraziando i miei genitori, coloro i quali per primi mi hanno trasmesso l'amore per la conoscenza, per l'analisi critica, per la ricerca dell'ignoto e che mi hanno insegnato a non fermarmi alle apparenze, alle risposte facili, ma a scavare sempre più in fondo... e me l'hanno insegnato così bene, ma così bene, che non solo l'allievo ha superato i maestri, ma risulta ad essi pure a stento sopportabile per il livello di petulanza e pignoleria raggiunto! Grazie per le gite ai musei, grazie per i robottini saldati insieme, grazie le novelle, grazie per le spiegazioni idrauliche dell'elettrostatica. E grazie anche per non avermi decapitato prima di ogni esame, e soprattutto prima delle discussioni di tesi!

Sperando che non me ne abbiano a male, vorrei poi ringraziare in blocco tutti i parenti che mi hanno supportato in questi anni, credendo in me e facendo il tifo per il loro sciamannato (mi scuso con tutti i non toscani: quasi sempre faccio il possibile per evitare locuzioni regionali, ma questa in particolare rende troppo bene, assumendo la sfumatura sia di disordinato che di scapestrato, in un'accezione comunque bonaria e paziente, come un brontolio con un sorriso) nipote/bisnipote/biscugino/inserire il grado di parentela desiderato che telefona e si fa sentire quasi di frequente.

Voglio poi ringraziare i miei amici, quelli di vecchia data e quelli di nuova acquisizione, quelli che vedo e sento di rado e quelli coi quali sono in contatto quasi quotidianamente. Ognuno di voi, ovviamente nel vostro modo squisitamente personale, mi ha sempre s-u/o-pportato, e non soltanto quando andavo nel panico prima di un esame che poi sarebbe andato meglio di quanto avrei sperato (e che mi avrebbe condannato ad una gragnuola di "Cosa ti avevo detto?!"), ma anche e soprattutto nel corso e negli aspetti tutti di questa cosa così stupida e meravigliosa che si rivela essere la vita. "Squisitamente personale", dicevo, perché dai vostri approcci certamente diversi all'esistenza ho sicuramente imparato tanto, e molto vi devo per la mia crescita come persona.
Mi azzarderò in questa cosa pericolosissima che è l'elencazione, dove anche una singola dé-

faillance può dare origine ai più atroci dei risentimenti: Niccolò (N), Gianfranco (Janz) e Sandro (S), per i quali mi azzardo a ripescare, con un'ombra d'imbarazzo, un'espressione tipica dei miei temi alle elementari, e che spero riesca a rendere appieno ciò che essi sono per me, cioè amici per la pelle; Nicola (aka Andrea), la felice aggiunta al gruppo tramite il canale S (così facendo ho reso sia ai fisici che ai non-fisici questa frase incomprensibile); Davide B., il mio budino bolognese preferito; Davide F., a cui devo de facto la mia introduzione al mondo del metal; Daniele, compagno d'esperienza universitaria, di lunghi discorsi filosofici e di profonda crescita personale; Claudio, quasi un'anima gemella col quale so di poter sempre condividere non soltanto pareri, esperienze ed emozioni relativi all'universo videoludico, fumettistico, d'animazione e, sintetizzando all'estremo, nerd, oltre che a quello della nostra amata Fisica, ma fors'anche più importantemente turbinii più intimi dell'anima; Andrea (no, non è vero, si chiama Baso), conosciuto grazie al mondo di Yu-Gi-Oh! (parlando del quale devo quantomeno citare Laura e Davide C.) e divenuto più che un semplice compagno di Duelli; Carlo, col quale ho strettamente condiviso un mese nella casa della vecchia ladra-di-identità (e col quale ho testato la meraviglia dei wurstel impanati e fritti avvolti nel bacon impanati e fritti avvolti nel bacon impanati e fritti) durante uno dei periodi più belli della mia vita, quello a SLAC, periodo che ha innescato la serie di eventi che mi ha portato a scrivere questi ringraziamenti di cui ancora non si scorge la fine.

Questa cosa è sicuramente già diventata sin troppo lunga, ma voglio ancora ringraziare i miei "colleghi", i compagni di lavoro e di tesi a Pisa e in Belle II, che mi hanno permesso di vivere quest'ultimo caotico, sfrenato, meraviglioso, d-istruttivo anno di Fisica e di Vita. Prima di loro devo però ringraziare Tina, che ha contribuito in maniera sostanziale a rendere non solamente possibili ma pure a meravigliosi i miei due mesi in quel della California, con una gentilezza e disponibilità che non credevo possibile da una sola persona; I also wish to thank Brian and Bertrand, which tutored me through my analyses in *BABAR*, both during the Summer School and during my thesis, without which this all wouldn't have been possible since the very beginning.
Grazie quindi a Francesco, che oltre a redarguirmi in quanto ringrazio troppo e sono troppo prolisso (e che quindi alla fine di questa sezione di ringraziamenti vorrà doppiamente scuoiarmi) mi ha permesso di entrare a far parte di questa collaborazione e conoscere tutte le altre persone che ne fanno parte, mi diede lo sprone nei momenti iniziali del secondo anno di magistrale per tornare sulla retta via della monotonia dei voti, e che mi ha seguito nel corso di quest'anno, evidenziando i miei punti deboli su cui lavorare e i punti di forza su cui far leva; Thomas, that is not only always right, but has been a wonderful office-mate, so willing to tutor me against all the issues I have had with my analysis, my codes, and Physics in general, thanks to which I spent a great week back in California (also eating exaggeratedly big burritos), often briefly telling me I'm far too talkative (and this final section is a confirmation of that...), helping me so so much with my thesis and my development as a scientist, and that is to me more than a colleague; Giulia, l'altra mia office-mate, alla quale forse un giorno riuscirò a far apprezzare i piccioni, colei che mi ha introdotto agli inaspettati orrori e ai turbolenti incanti del software di Belle II e con la quale ho creato la mia piccola creatura, il CoG, e che mi ha paziente-

mente fatto da guida quando, esausto ed estasiato, atterrai nella Terra del Sol Levante per la prima volta in vita mia, oltre che a forzarmi nei primi talk pubblici, a sopportare i miei opinabili puns e a ricordarmi come mi chiamo; Laura e Luigi, che prendendo in prestito la terminologia giapponese potrei definire (così riuscirò a far arrabbiare due persone con una frase sola!) come senpai e kohai; Daniele G., col quale ho riesumato la cara vecchia Lumaca di Metallo; Stefano, Giuliana, Eugenio, e tutte le persone di cui mi sto scordando e che spero sinceramente non mi odieranno troppo per ciò.

Vorrei infine ringraziare mio nonno, per tutte le storie di scienza con le quali ha affascinato (una per tutte, il funzionamento di una bomba atomica, per la mia tesina di terza media), a gambe accavallate, nel corso di una vita che avrei voluto fosse stata un po' più lunga, così da poter vedere suo nipote diventare un dottore magistrale in Fisica.
Grazie anche per le lacrime che mi stai facendo versare in questo momento. Mi manchi.