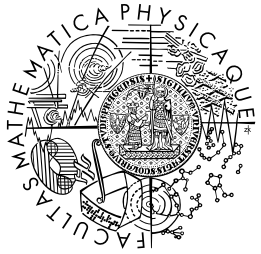


MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

DIPLOMOVÁ PRÁCE

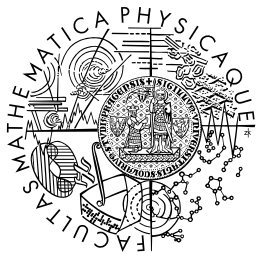
Michal Krištof

Studium rozpadů B-mezonů v experimentu Belle

Ústav částicové a jaderné fyziky

Vedoucí diplomové práce:	doc RNDr. Zdeněk Doležal, Dr.
Studijní program:	Fyzika
Studijní obor:	Částicová fyzika

Praha 2019



FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Michal Krištof

Study of B-meson decays in the Belle experiment

Institute of Particle and Nuclear Physics

Supervisor of the master thesis:	doc RNDr. Zdeněk Doležal, Dr.
Study programme:	Physics
Study branch:	Particle Physics

Prague 2019

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ..... signature of the author

Název práce: Studium rozpadů B-mezonů v experimentu Belle

Autor: Michal Krištof

Katedra: Ústav částicové a jaderné fyziky

Vedoucí diplomové práce: doc RNDr. Zdeněk Doležal, Dr., Ústav částicové a jaderné fyziky

Abstrakt: V této práci se zaměřujeme na studium rozpadu B_0 mezonu na mezony D_s^* a ρ. Práce vysvětluje metody a přístupy k analýze dat na takzvaných B-továrnách, jakou je urychlovač KEKB. Cíl této teze je vypočítat větvící poměr zmíněného rozpadu, upřesnit již změřený větvící poměr změřený na experimentu BaBar pomocí dat získaných z experimentu s vyšší integrovanou luminositou. Výhledy této teze nejsou jenom v rozšíření našeho poznání větvících poměrů rozpadů B_0 mezonu, ale také jako první krok vedoucí k rozšíření poznání narušení CP symetrie v Standardním modelu měřením úhlů Unitárního trojůhelníku.

Klíčová slova: Belle detektor Analýza dat Větvící poměr CP violation Rozpady mezonu B_0

Title: Study of B-meson decays in the Belle experiment

Author: Michal Krištof

Institute: Institute of Particle and Nuclear Physics

Supervisor: doc RNDr. Zdeněk Doležal, Dr., Institute of Particle and Nuclear Physics

Abstract: In the thesis we study the decay of B_0 meson to D_s^* and ρ mesons. The thesis explains the methods and approaches to data analysis in so-called B-factories, simmilar to the KEKB accelerator. The aim of this thesis is to calculate the branching fraction of this decay to further improve the previously measured branching ratio at BaBar experiment with additional data gained from an experiment with higher integrated luminosity. This thesis's prospect is not only broadening our knowledge of branching fractions of B_0 meson decays, but also it is a starting point for further analysis with the goal of broadening our knowledge of CP symmetry violation in the Standard Model by measuring angles of the Unitary Triangle.

Keywords: Belle detector Data analysis Branching fraction CP violation B_0 meson decays

I would like to express my humblest gratitude to my co-supervisor, Mgr. Červenkov, for his guidance and to my supervisor doc. Doležal for the possibility to work at the Belle experiment and also to the analysis group at KEK for their consultation and help during my analysis.

Dedicated to Murat Hunyadi and to Tasa, may you rest in peace.

Contents

Introduction

The 20th century has been crucial for scientific research of our physical world. The industrial revolution of the century prior enabled advances in science not heard of in history of mankind and gave rise to the greatest scientists of all time not only in theoretical fields, but also experimental. From the toil of Curies at the crossing of centuries, through the genius of Einstein and Dirac up until the novel approaches of Feynman; from the study of the largest structures of our universe up to its smallest elements, all fields of physics observed rapid evolution and numerous breakthroughs.

But one of the most influential breakthroughs came not from physics, but from mathematics. It was the Theorem of Emmy Noether, proven in 1915 and published in 1918 by a German algebraist Emmy Noether [38], that helped to formulate a comprehensive model of quantum fields later in the twentieth century. It states that every differentiable symmetry of action corresponds to a conservation law. In other words, the theorem links the various symmetries of our universe with laws of conservation. This is an important finding, because it gives theorists mathematical tools to explore the fundamental laws of nature and provides mathematical explanation for these laws.

While the theoretical world has been busy in building the quantum theory, experimentalists were making discoveries using novel method of particle acceleration, collision and scattering. First scattering experiments date to 1895, when Lenard scattered electrons on gasses. But the next milestone was reached at last in 20th century, in 1919, when Rutherford observed first nuclear reactions induced by natural alpha particles. The first accelerator-enhanced experiments is accounted to Cockroft and Walton in 1931, where the first atoms, atoms of lithium, were split. With these experiments and the rapid rise of the nuclear physics that followed thereafter, a need for novel acceleration techniques and higher energies arose. The first accelerators were mere voltage multipliers, that found use also in electronics, and electrostatic generators that were limited in accumulated voltage either by the dielectric strength of the medium in which they were placed or the input voltage and capacitance of used components, hence a new method of acceleration, one that would not be dependent on simple voltage difference, was needed [15]. The

resonant accelerator was born. Fast forward a few years, and physicists were using resonant particle acceleration to achieve high energies in collision and scattering experiments. As soon as the center of mass system (CMS) energy started rise, new particles emerged. Hundreds of new hadrons were discovered by the end of the 6th decade [20]. To bring order and reason to this "particle zoo", a new model of elementary particles was created, in which hadrons were no longer elementary, but made up from tiny constituents.

The model, now called the Standard Model, contains three fundamental interactions: the strong interaction, binding the constituents of hadrons, quarks, together and its leakage out of those hadrons in turn creates nuclear forces. The electromagnetic interaction of electrical charges that keeps electrons in atoms and its residue creates molecular binding forces. And a third one, in which we observe breaking of symmetries, at least up to date, weak interaction, governing the decay of radioactive isotopes.

The asymmetry of the weak interaction was hinted from the first observation of strange decays involving what was turning out to be the same particle (now called K-meson) decaying with two different decays with final states with two different intrinsic parities - a quantum number that describes the behaviour of the quantum wavefunction under the mirroring, $\vec{x} \to -\vec{x}$ transformation (the famous τ-θ puzzle). The first, who suspected parity violation formally, were T. D. Lee and C. N. Yang, who argued that parity conservation in weak processes has not been verified and proposed a number of experimens to do so [31]. Also thanks to their suggestions, it was finally proven to be violated in the experiment of C. S. Wu in 1956 [46].

To maintain order in the quantum world after the discovery of parity violation, it was soon proposed by a Soviet theorist Lev Landau, that the combined symmetry of charge conjugation, the exchange of particles for antiparticles, and parity (CP), could be an exact symmetry of weak interactions [30]. But this idea didn't age well either. It was found to be violated in 1964 by a collective behind J. W. Cronin and V. L. Fitch, again within the system of neutral kaons [19] and described theoretically by M. Kobayashi and T. Masakawa in 1973 [26]. Today, a combined symmetry of CP and time (CPT) theorized by J. Schwinger in 1951 [43] is believed to hold true [27].

Although for a long time only kaons were confirmed to be CP-asymmetrical, in 2001, on the Belle experiment, the very experiment we are taking data for analysis from, a CP-asymmetry was observed in B-meson decays [7] and 2011 shew indications of CP-asymmetry of neutral D-meson decays at LHCb experiment in CERN [4], which were finally confirmed in May 2019 [5].

That is the evolving history of the Standard Model of elementary particles that we know today. Its complete theory was formulated in mid-70s, but its

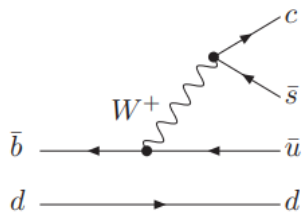


Figure 1: Tree-level Feynman diagram of decay $B^0 \to D_s^{*+}\rho^-$

full experimental confirmation had to wait even longer than the observation of CP-asymmetric decays, up until the discovery of the Higgs Boson in 2012 [18, 3]. Nowadays, the Standard Model is one of the most successfull and well-tested physical theory.

Despite the success of the Standard Model, it is still a theory that leaves some questions about particle physics unanswered. And as every physical theory, it also has its limits. For these reasons there is an incentive to find these limits, push beyond them and try to find answers for the pending questions. For this reason, it is important to look at the unknown not only to test the contemporary theory, but also to try and find glimpses “new physics" that might lay there.

For this reason we've decided to study the decay process of $B^0 \to D_s^{*\pm} + \rho^{\mp}$. Not only is this process, being rare enough, suitable for studying branching fractions (BF) of rare processes governed by the weak interaction, but it is also suitable for studying the CP symmetry breaking. Though CP asymmetries are minute, they have a tremendous impact on our universe. Especially during the period of early universe, when particles were starting to emerge from the primordial substrate. Since the symmetry generates a law of conservation, as discussed above, had our world been completely symmetrical, it would be static and lifeless for everything would be conserved in physical processes and represented in equal amounts. The tiny little differences in some processes not obeying conservation laws due to symmetry breaking of this kind are what ultimately enabled our universe to form to the state we observe today.

Chapter 1

Experimental Setup

The Belle experiment begun data taking in 1999 and ended operation the 30th June 2010. The experiment was part of an international collaboration of 19 countries. It was operated by Japanese High-Energy Accelerator Research Organisation on particle collider KEKB near the town of Tsukuba in Ibaraki prefecture. During its lifetime, it achieved a world record in instanenous luminosity of $2.11 \times 10^{34}\,\mathrm{cm}^{\breve{}2}\,\mathrm{s}^{\breve{}1}$ and also its results helped award Makoto Kobayashi and Toshihide Masakawa Nobel prize for physics in 2008. Its total integrated luminos-

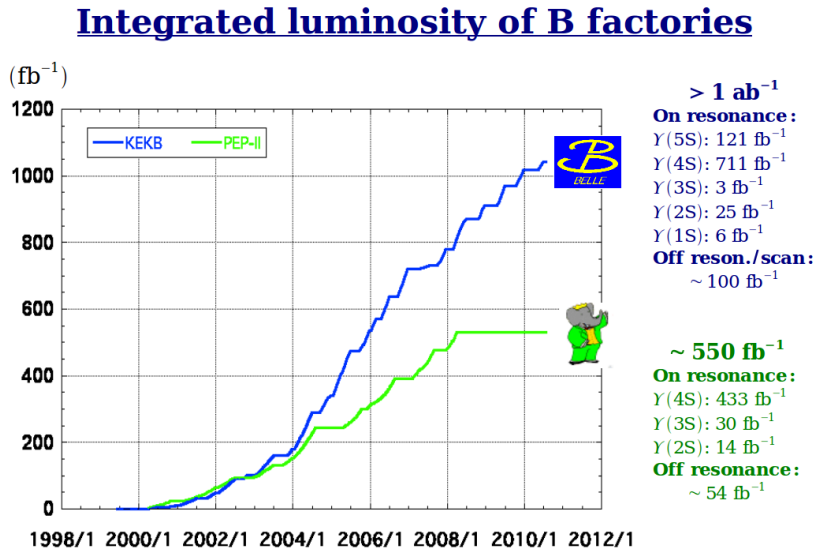


Figure 1.1: Integrated luminosity of the Belle experiment during its life compared to BaBar experiment (sister experiment, located at SLAC in California)[1]

ity exceeded $1000\,\mathrm{fb}^{-1}$[1].

After its shutdown, the experiment, together with the collider, has undergone an upgrade. The currently running Belle-2 experiment on SuperKEKB collider are their successors. A state-of-the-art technology like DEPFET silicon pixel detectors were added to the experiment[34] and the collider to achieve instantenous luminousity 40-times higher, $8 \times 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The upgraded experiment should accumulate an integrated luminosity of $50\,\mathrm{ab}^{-1}$ during its lifetime. The new collider started operation in 2017[2].

1.1 The KEKB Collider

The main goal of the KEKB collider was to probe the decays of the heaviest quark that lives long enough to create bound states, the b quark. To achieve this, it was set to produce the Υ resonance, which is a bound state of $b\bar{b}$ quarks. Most of the time, it run in CMS energy setup of $\sqrt{s} = 10.58$ to achieve production of $\Upsilon(4S)$ resonance, in which mode it took $711\,\mathrm{fb}^{-1}$ of data as can be seen in fig. 1.1.

The importance of $\Upsilon(4S)$ resonance in the case of B-factories is, that $\sim 96\%$ of the decays of this resonance is composed of $B\bar{B}$ meson pair, with $(48.6 \pm 0.6)\%$ being neutral B mesons and $(51.4 \pm 0.6)\%$ charged [45].

The collider design, as can be seen in fig. 1.2, consisted out of two storage rings, 3 km in circumference, and one interaction point (IP), built in a tunnel of former TRISTAN accumulation ring. The two-ring design is needed because the collider was designed to be asymmetric, with one ring, the Low Energy Ring (LER) accelerating positrons to the energy of 3,5 GeV and the High Energy Ring (HER) accelerating electrons to 8 GeV. The collider is fed by a linear accelerator. During its runtime there were 5000 particle bunches with spacing of 59 cm and particle numbers per bunch reached order of 10^{10}. Particles were interacting at the Belle detector IP at a crossing angle of 11 mrad to avoid parasitic collisions and separate incoming and outgoing beams [29].

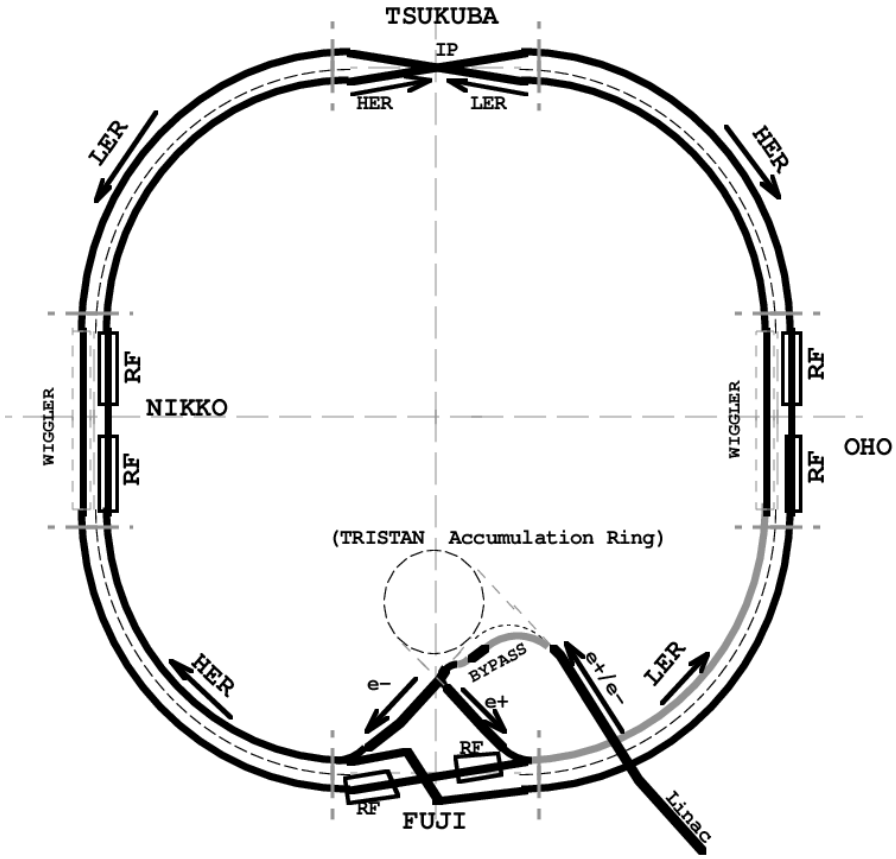


Figure 1.2: A schematic of the KEKB collider [29]

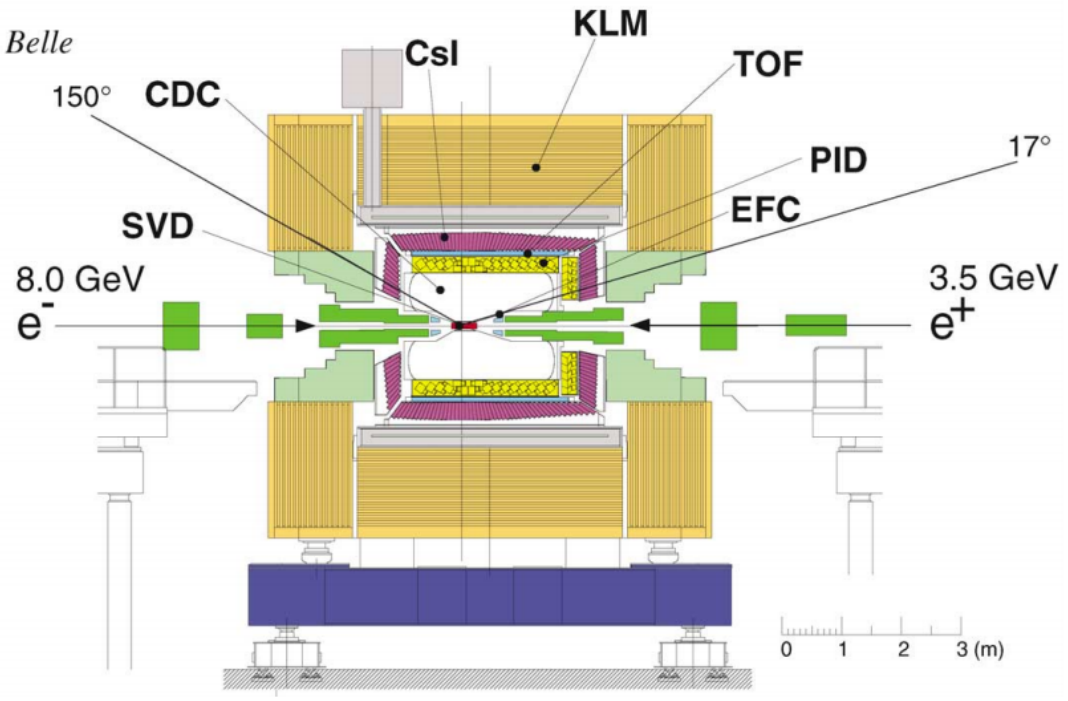


Figure 1.3: Belle detector cross-section [6]

1.2 The Belle Detector

Identifying products of electron-positron collision and measuring their energy and momentum is a complex task. The technology has gone far from the old days of bubble chambers, when people analyzed photos of traversing particles by hand measuring energy, mass and momentum by curvature of visible tracks and relative thickness of ionization track. Today, even if we forget the fact that the needed resolution is much greater, the amount of data alone taken from particle experiments is so big, that manual analysis would be impossible. The Belle detector itself took circa 400GB of raw data daily during its runtime. To identify and measure particles it used advanced electronics in 7 subdetectors which can be seen in fig. 1.3. Providing sufficient electronics is a challenging task for the detector development teams thanks to extreme requirements for radiation and magnetic hardness and reliability together with sufficient resolution. What's more, the innermost detectors ought not to provide exquisite temporal and spatial resolution, but also have to impede paths of traversing particles as little as possible.

Thanks to the asymmetric energies of colliding particles, the main components of the detector are designed to cover angle range between 17° and 150° of polar angle.

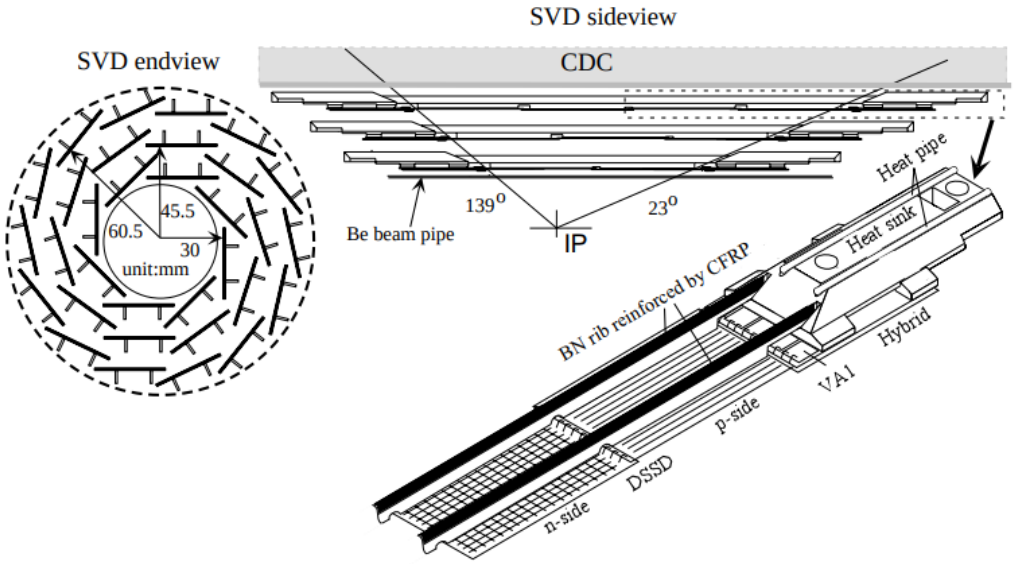


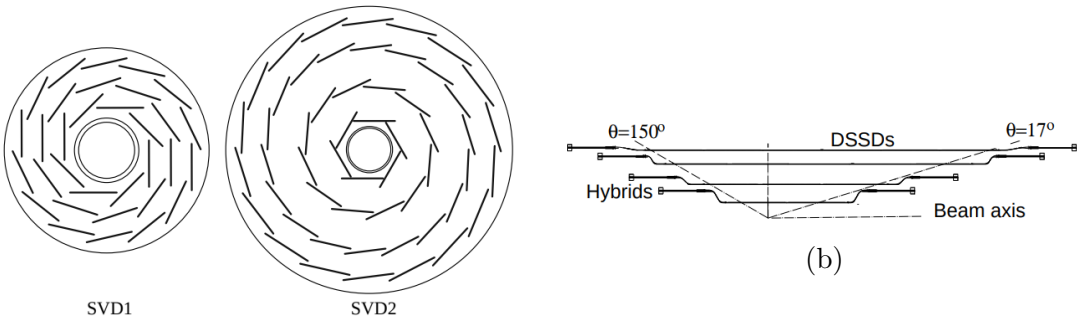
Figure 1.4: SVD detector schematic and design overview [6]

The very heart of the detector transects a double-wall beryllium beampipe 40 mm in diameter that has to be cooled thanks to the heat produced by the radiation of the collision. In order to achieve sufficient cooling without impeding the traversing particles too much, helium-gas cooling technique was adopted. To reduce the X-ray background with energy lowed tha 5 keV from the HER, the outer wall was coated with 20 μm thick gold layer.

The whole complex of the detector, apart fom just the KLM detector, was surrounded by a 1.5 T, 3.4 m diameter and 4.4 m long, 24 t niobium-titanium-copper solenoid embedded in high-purity aluminium stabilizer with the entire detector supporting structure and KLM iron absorber servig for it as a yoke [6].

1.2.1 SVD

The main purpose of the Belle detector, measuring CP violation in B meson decays, requires precise measurement of time difference Δt between B meson decays in single event. To achieve this, Belle detector employs a Silicon Vertex Detector (SVD). The SVD is the innermost subdetector closest to the IP. It measures Δt indirectly by measuring spatial Δz observable (the difference in position in the direction along the beampipe) of vertices of B decays. The Δt is then gained from relativistic formula 1.1 (β denotes here the speed of the particle expressed in



(a) (b)

Figure 1.5: (a) Comparison between layouts of SVD1 and SVD2, (b) Longitudinal cross section of SVD2 [32]

fractions of the speed of light c and γ its Lorentz factor).

$$\Delta t = \frac{\Delta z}{c\beta\gamma} \tag{1.1}$$

It is designed to measure Δz with accuracy better than $\sim 100\,\mu$m needed for observation of time-dependent CP asymmetries [6].

During its lifetime, the Belle SVD detector was updated once. The first version, SVD1, consisted of three layers of double-sided silicon detectors (DSSD) S6936 fabricated by Hamamatsu Photonics originally developed for DELPHI micro-vertex detector. The active region of the detector covered a solid angle from 23° to 139°. Each DSSD consisted of 1280 300 μm thick sense strips read out by 640 pads on opposite sides with z-strip pitch of 42 μm. In order to cover a cylindrical shape with planar detectors, wind-mill design was adopted with each layer composed of planar ladders of DSSDs. The number of ladders in layers from the innermost layer to the outermost one was as follows: 8-10-14 [6]. The wind-mill design as well as the ladder schematic can be seen in fig. 1.4.

In September 2003, the second verison of SVD detector, denoted SVD2, was installed. Asides from the detector gaining one more layer of DSSD ladders, the whole layout of layers was redesigned. As can be seen from fig. 1.5, the inner radius decreased from 30 mm to 20 mm, outer increased from 72 mm to 101 mm and the average distance between layers increased, the composition of radii of layers changed from 30.0-45.5-60.5 mm of SVD1 to 20.0-43.5-70.0-88.0 mm on SVD2. The length of the ladders also increased to achieve better solid-angle acceptance of $17° < \theta < 150°$. For the purpose of SVD detector upgrade, the beampipe diameter had to be decreased from 40 mm to 30 mm and the next subdetector in line, the Central Drift Chamber (CDC), had to be redesigned [32]. The upgraded SVD detector was capable to achieve accuracy in the $r - \phi$-plane of $\sigma_{r-\phi} = 12\,\mu$m and

$\sigma_z = 19\,\mu$m in the z-plane [8].

1.2.2 CDC

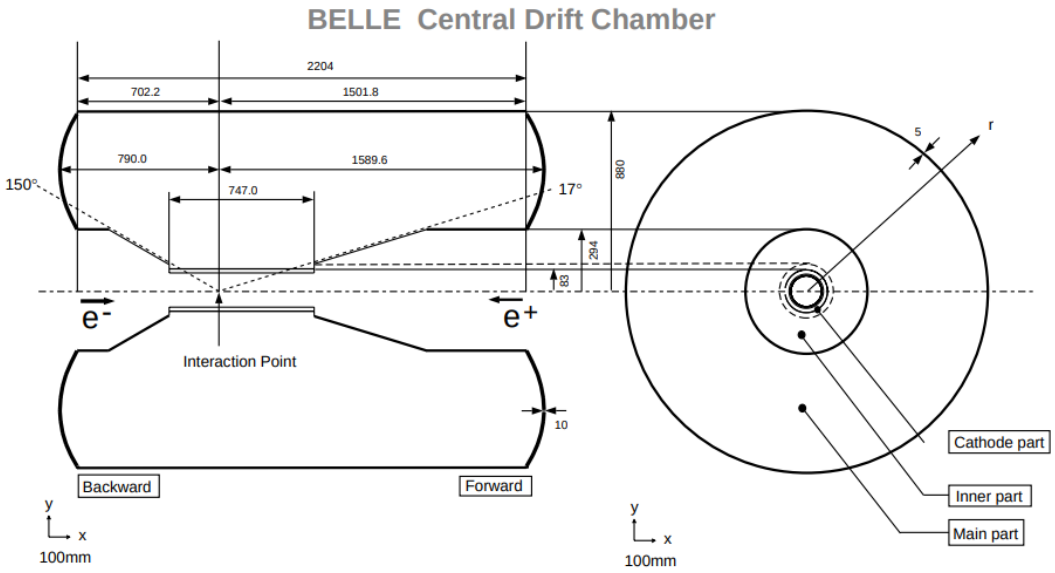
The CDC is a tracking detector. However, its working principle is different from silicon trackers, which could be equivocated to a large camera chip. As the name suggests, it is a chamber filled with gas with wires running through it to provide electric field for the charges produced by ionization from the passing charged particles to drift towards the readout wires and also to amplify the signal by accelerating the ionization electrons, which in turn ionize more charges along the way, generating electron avalanches. This provides advantages in form of a low cost in general, however this is traded off with a bad resolution. The particle tracking is achieved by measuring the drift time of ionization electrons in the gas filling and the signal strength, in another words the number of collected charges, can provide information about the energy loss in the chamber through the Bethe-Bloch formula and from it distinguish between different particle families [13]. The particle curvature in the homogeneous field, in other hand, provides information about its momentum.

The Belle CDC was made of 50 cylindrical layers, each containing between 3 and 6 axial or small-angle stereo layers and 3 cathode strip layers. The longest wires were 2400 mm long and was designed to provide coverage of angle range between 17° and 150°. It contained total of 8400 nearly square drift cells with a maximum drift distance between 8 and 10 mm and radial thickness from 15.5 to 17 mm except from the inner 3 layers, that were smaller. Sense wires were gold-plated tungsten of 30 μm diameter and field wires are from unplated aluminium, 126 μm in diameter.

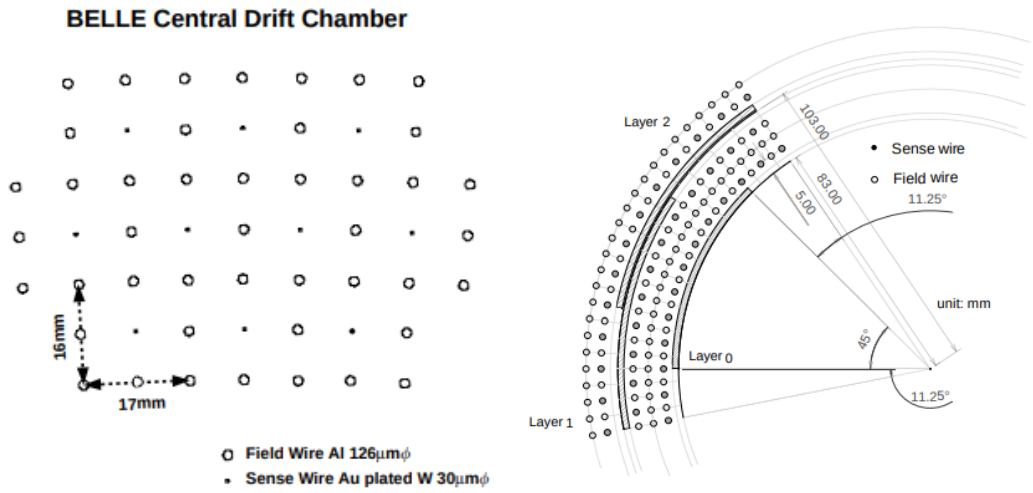
Measurements at the innermost radii were provided by 8 segments (in the ϕ direction) of cathode strips glued on the inner cylinder surface of the chamber and on both sides of a cylinder between second and third anode layers. This setup provided radial deviation from a perfect cylindrical shape smaller than 100 μm.

The chamber was filled with a mixture of helium-ethane gas mixed with ratio of 1 : 1. The gasses were selected specifically to minimize the effects of multiple scattering and small photo-electric cross-section provided small background from synchrotron radiation. This setup provided approximately 130 μm resolution in $r - \phi$ plane and 0.64 mm for Δz.

The nominal value of the magnetic field present was set to be 1.5 T. To correct the effects created by the deviation from perfect uniform homogenous magnetic field of the Belle solenoid, the Kalman filtering method was used [6].



(a)



(b) (c)

Figure 1.6: (a) A schematic view of Belle CDC, the numbers stated are in millimeters (b) CDC drift cell structure (c) Cell structure and cathode sector [6]

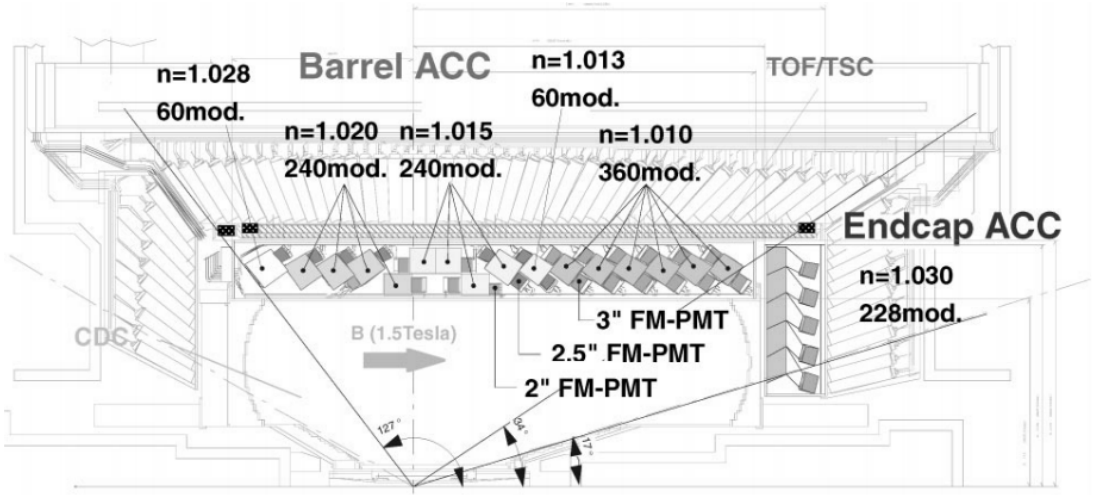


Figure 1.7: Schematic of the Belle ACC [24]

1.2.3 ACC

When a particle travels through a medium at speeds higher than the speed of light in that particular medium, it emits radiation of light in rings. This spectacular effect can be most prominently seen in nuclear reactor cores, where superluminal reaction products in water generate a characteristic bluish glow. This effect is comparable to the sonic boom of a jet flying at supersonic speeds in air. Although the effect can be derived from the classical electrodynamics of uniform moving charges [28], the relevant formulae can be easily derived from basic geometrical picture of the spreading radiation, resulting in relation between the Cherenkov ring's size and particle's speed [41]. This is the basis of operation for all of the Cherenkov detectors.

Cherenkov detectors can either measure this ring's diameter and provide a measurement of traversing particles' speed, in which cases they are called Ring Imaging Cherenkov detectors, or just distinguish between slow and fast particles by choosing a medium with appropriate refraction index to provide a threshold for particle speed ($\beta_{\min} > 1/n$) to emit the cherenkov ratiation. The Belle spectrometer used the latter, while the detector upgrade, Belle-II, uses the more sofisticated, ring imaging detector [35].

Cherenkov detectors generally use either a gas (most popular being perfluorobutane - C_4F_{10}, from major experiments used in DELPHI [12], LHCb [14], HERMES @ DESY [10], COMPASS [21] detectors) with appropriate refractive index, or utilize silica aerogel. Aerogel is advantageous to use thanks to its solid state and very low density and hence its low weight. Gaseous Cherenkov detectors need to fill and maintain purity of the filler gas a complex system of gas pumping and circu-

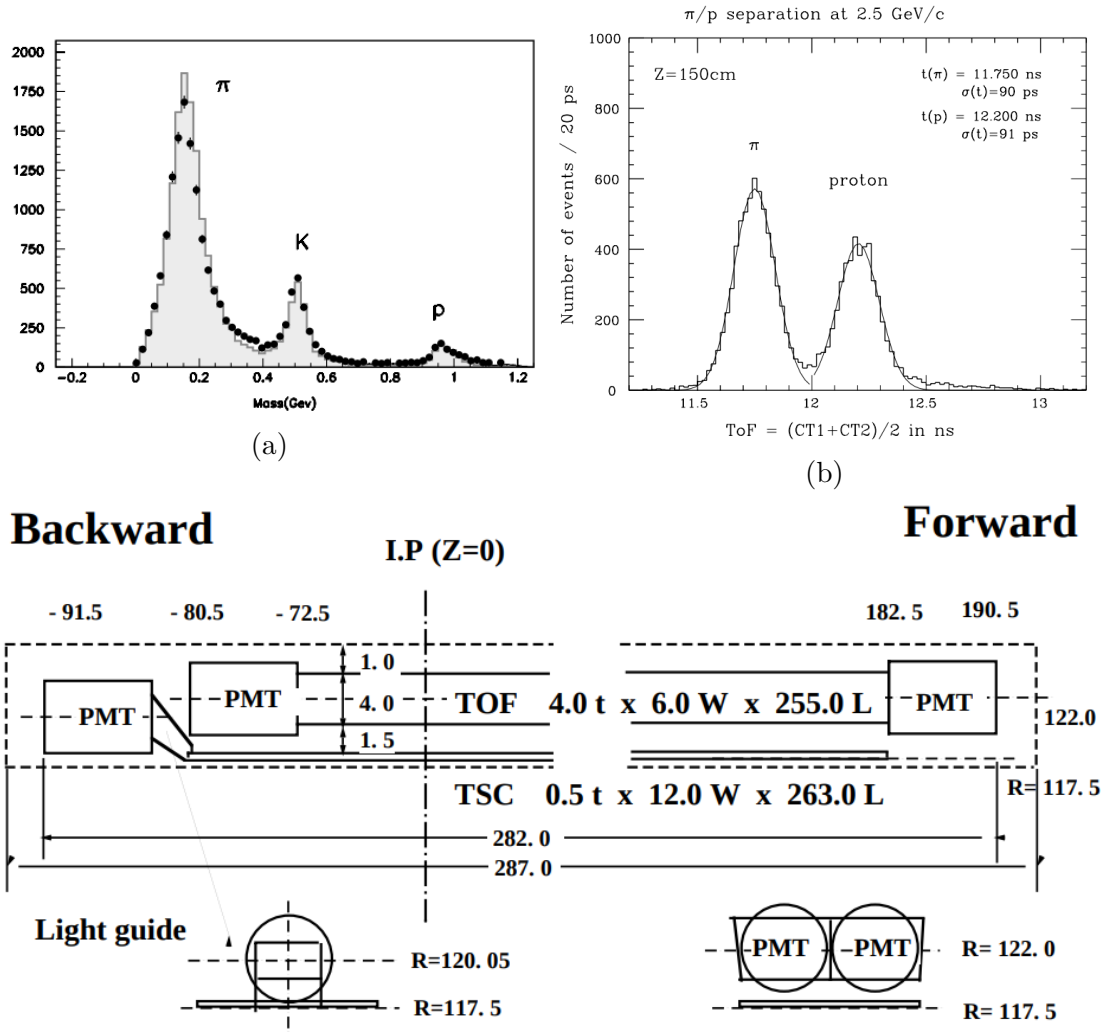
lation system (as an example can be seen gas pumping and circulation system of COMPASS experiment at CERN [11]). The aerogel, on the other hand, is not easy to produce, especially in large quantities (aerogel production process is described, for example, in [9]).

The Belle ACC was designed to complement CDC and Time-Of-Flight (TOF) measurements and extend momentum coverage for $\pi^{\pm}/K^{\pm}$ separation up to 3.5 GeV/c. It was divided into 960 counters segmented into 60 cells in the ϕ direction in barrel region and 228 counters in 5 concentric layers in the endcap region. To cover the momentum range from 1.2 to 3.5 GeV/c, the refractive index of the aerogel modules was ranging from 1.01 to 1.03. Each module was equipped with one or two fine-mesh photo-multiplier tubes (FM-PMT) [24]. Schematic of the ACC can be seen in fig. 1.7.

1.2.4 TOF

Belle TOF detector is timing the ∼1.2 m path the particle has to travel from the IP to its plastic scintillators. In order to do this for particle momenta of 1.2 GeV/c, 100 ps temporal resolution is needed. The system provides particle identification and trigger for the experiment. It is separated into 128 TOF scintillator counters and 64 trigger scintillation counters ended with FM-PMT, separated by 1.5 cm gap to isolate TOF from photon conversion backgrounds by taking coincidence between TOF scintillators and trigger scintillators. This layout can be seen in fig. 1.8-c as well as example plots of TOF response histograms for three different particles at two different momenta ranges 1.8-a, b. The overall thickness of the whole system reaches 6 cm and probes $34° < \theta < 120°$ of solid angle.

The triggering system of the Belle detector, however, does not rely only upon TOF system. The full trigger schematic can be seen in fig. 1.9. The TOF system only provides timing information (denoted as Level-0 trigger), but the whole background suppression algorithm uses the entire complex of the detector, SVD, CDC, TOF as well as calorimetry data. This provides data complex for so-called Level-1 online hardware trigger that is governed by global decision trigger logic circuit (see fig. 1.9). Asides from this triggering solution, fast track fitting algorithms denoted as Level-3 software trigger was used. The latency of trigger levels were as follows - level-0 trigger, the fastest, utilizing TOF only with a latency of $\sim 0.85\,\mu$s and level-1 trigger had a latency of $2.2\,\mu$s [6].

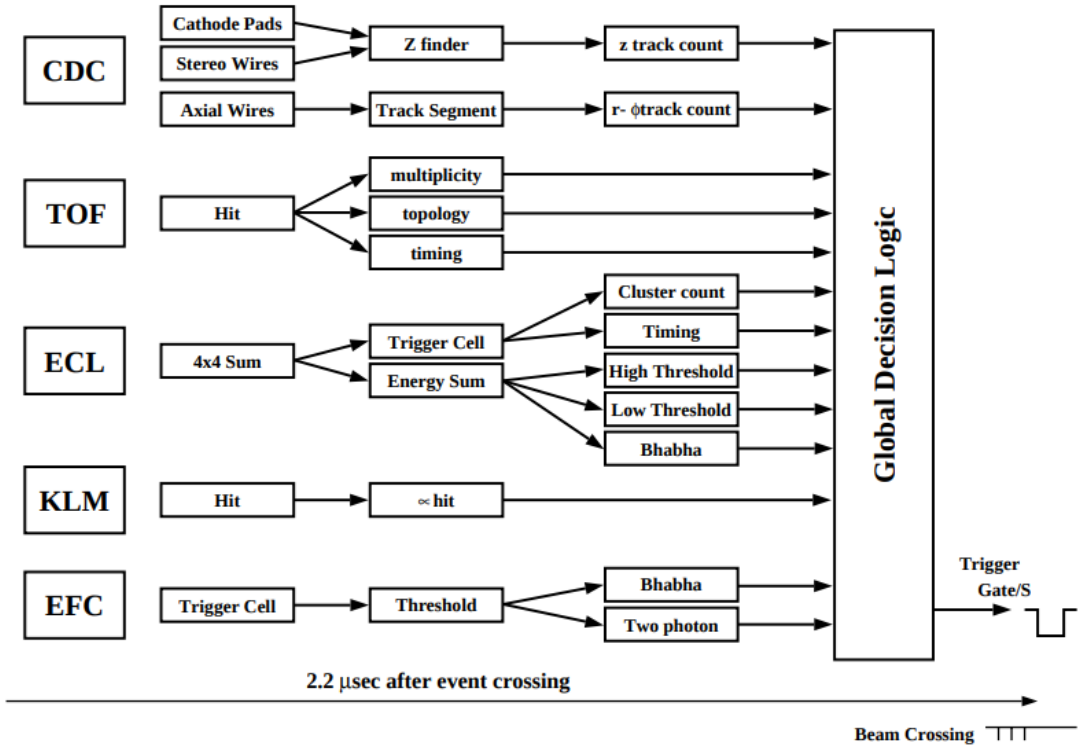


(a)

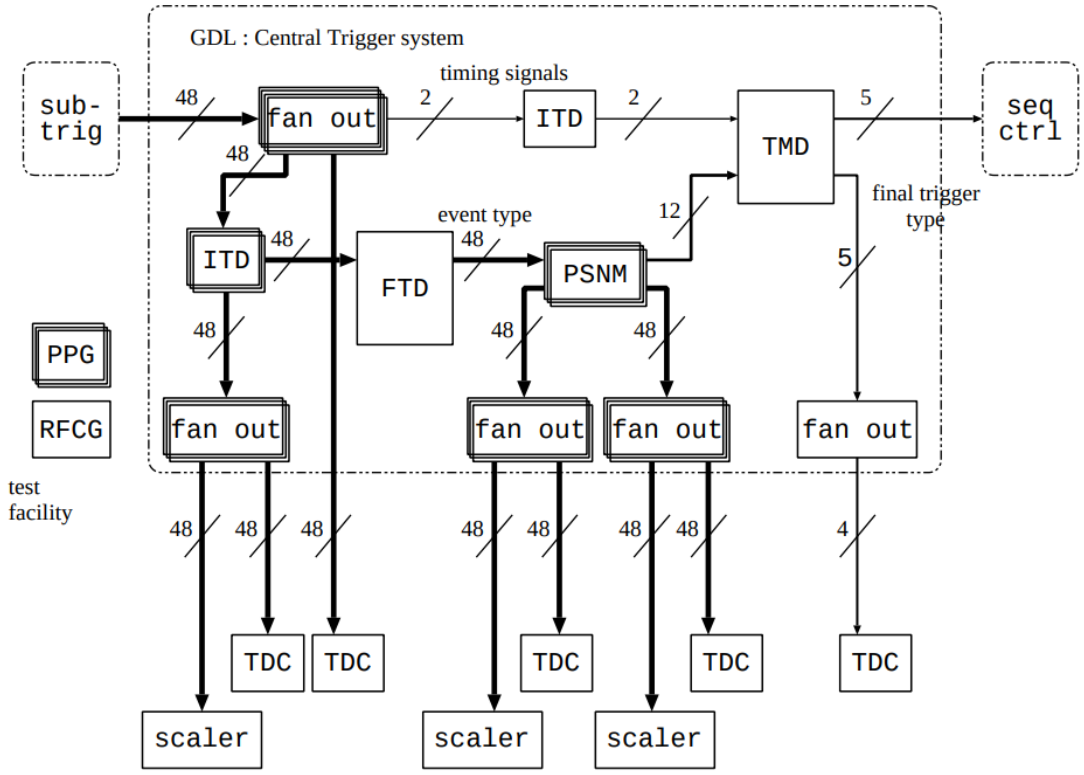
(b)

(c)

Figure 1.8: Separation of particles by TOF (a)$\pi/K/p$ for particle momenta below 1.5 GeV/c (b) π/p separation at 2.5 GeV/c (c) A schematic of the TOF counter [6]



(a)



(b) ITD - adjusts timing of input trigger signals, FTD - correction of sub-detector triggers, global trigger logic, PSNM - prescaler of high rate input triggers, TMD - generates final 2.2 μs trigger signal based on timing from TSC and ECL

Figure 1.9: Triggering schematics: (a) Level-1 trigger schematic, (b) Global decision logic circuit schematic [6]. For the meaning of abbreviations please see the list of abbreviations.

1.2.5 Calorimeters

The purpose of calorimeters is to absorb all of the incident particle's energy, so they have to be placed behind PID, tracking and timing. Belle has three calorimeters: The electron calorimeter (ECL) with Caesium Iodine chrystals as scintillating agents (marked as CsI in fig. 1.3), the K_L and Muon (KLM) detector and Extreme Forward Calorimeter (EFC).

The EFC, closest to the IP from all calorimeters, has multiple roles to fullfil. Its primary design is to further increase the polar angle coverage. It also provides shielding of the CDC from the background generated by the beam. And lastly, it is used for monitoring the beam and luminosity of the KEKB. The calorimeter is designed to look at small polar angles from 6.4° to 11.5° in forward direction and 163.3° to 171.2° to probe processes such as $B \to \tau\nu$ and is attached to the cryostats of solenoids encircling the beampipe. Thanks to its extraordinary closeness from the beampipe and the IP, the calorimeter has to be extremely radiation hardened. For this purpose, bismuth germanate crystal calorimeter design was used. As all of the sub-detector instruments, the EFC is also taking part in the triggering process. Its triggering comprises, as can be seen from the level-1 trigger schematic in fig. 1.9, from two processes: Bhabha trigger is a coplanar forward/backward coincidence of energetic electromagnetic (EM) showers from electron-positron (called Bhabha) scattering, and two-photon trigger that combines a single EM shower together with some CDC tracks or ECL clusters.

The next calorimeter in line is the ECL. Electromagnetic calorimetry focuses on measuring the energy of highly energetic photons and electrons, or positrons respectively. It is designed to absorb energy of electrons. The passing particles generate EM showers in ECL crystals that is then measured. Highly energetic photons generate these showers thanks to them having high cross-section for a conversion into electron-positron pairs that then lose their energy through bremsstrahlung, generating more highly energetic photons and so cascading with energy down to a point where they start ionizing their surroundings. This way the ECL senses $e^{\pm}$ and γ using the same principle. Thanks to the fact, that electrons have lower mass and ionization than hadrons, showers that they cause are quicker to die out than those of hadrons.

Highly energetic photons at the Belle experiment usually form the final state of cascade decays and thus a good performance for $E_\gamma < 500\,\mathrm{MeV}$ is needed, however the calorimeter had to be built with two-body decays like $B \to 2\pi_0$ and detection of photons with E_γ up to $4\,\mathrm{GeV}$ in mind. For the latter, the spatial resolution is also important, to be able to distinguish the photons originating in π_0 decays from random background. For this purpose, towers of thalium-doped CsI crystals were adopted. The ECL contains in total 8736 of these crystals weighting 43 tons. These crystals cover 91% of 4π solid angle. Crystals point almost entirely towards

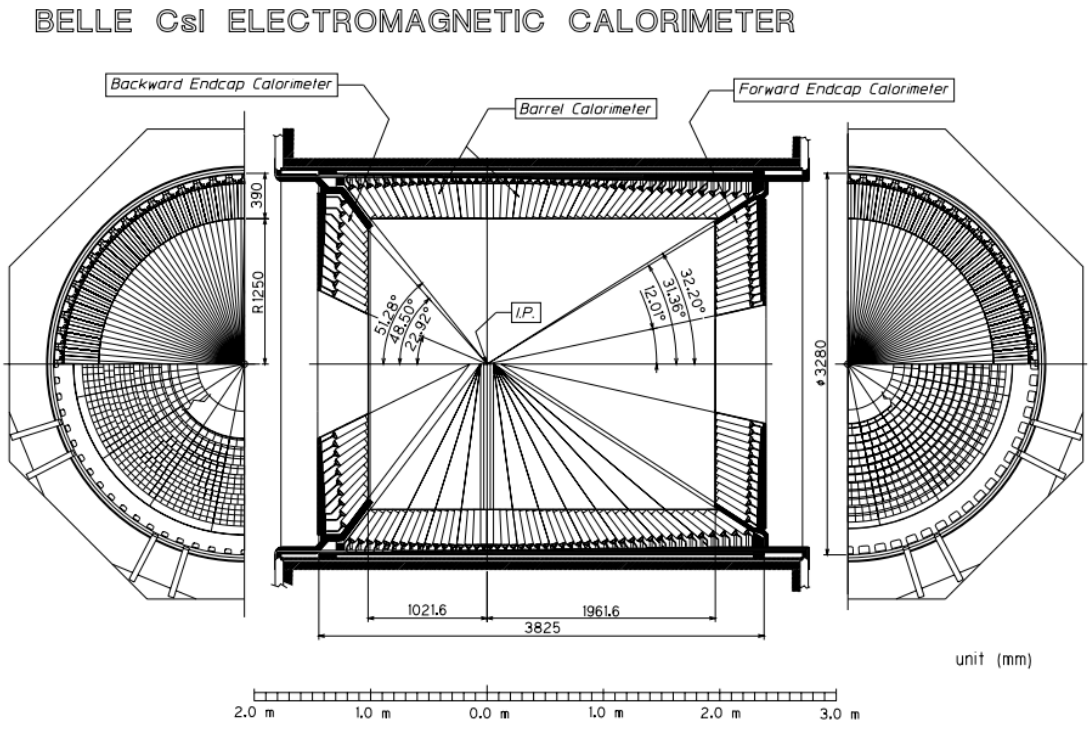


Figure 1.10: Schematic of the ECL [6]

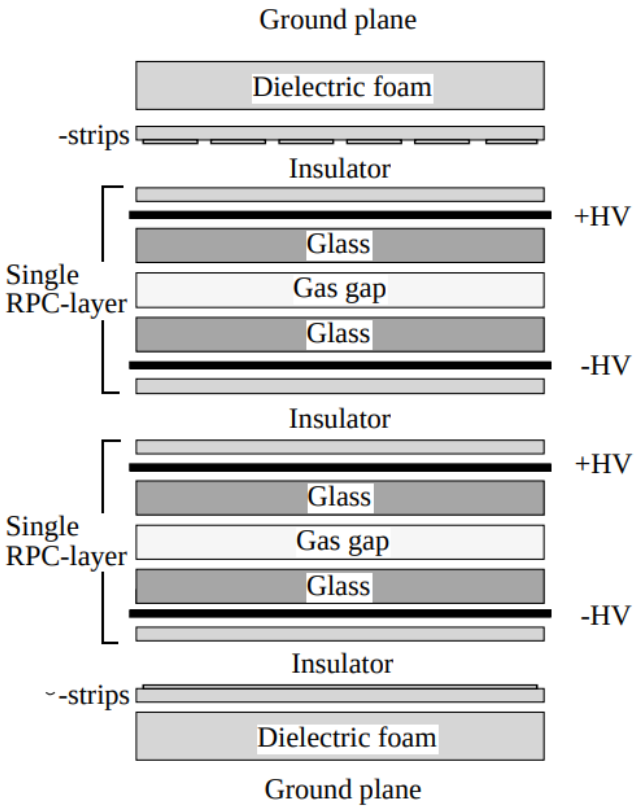


Figure 1.11: A structure of the KLM superlayer. The gas gap contained a gas mixture of 30% Ag, 8% C_4H_{10} and 62% CH_2FCF_3 [6].

the IP, with only a small tilt, $\sim 1.3^\circ$ in θ and ϕ-directions in the barrel and $\sim 1.5^\circ$ and $\sim 4^\circ$ in the θ direction in the enedcap region, to minimize the number of runaway particles through the gaps between crystals. Their dimensions were set by the requirement of 80% energy absorption efficiency of a photon injected into the center of the crystal. To contain the EM shower, the length of the crystals was set to 30 cm. Each crystal is ended by signal-detecting photodiode. The schematic of the ECL can be seen in fig. 1.10.

1.2.6 K_L and Muon System

The last detector of the Belle detector is K_L and muon detection system (KLM). It is designed mainly to convert the invisible K_L particles to showers of ionizing particles. For this purpose, its design provides 3.6 interaction lengts of material for K_L particles incident in normal direction in addition to 0.8 interaction lengths of the ECL.

The KLM had sandwich design with an alternating layers of 4.7 cm thick iron plates and detector superlayers detecting on the principle of discharges in gas in the presence of a charged particle in Resistive Plate Counter (RPC). The design of the KLM superlayers can be seen in fig. 1.11. The whole complex consisted of 15 detector layers and 14 iron layers in barrel (angles between $\sim 32^\circ$ and 129° in the z-plane) and 14 detector layers in endcap regions (other than barrel angles). The detector design also allowed for good discrimination between muons and charged hadrons based upon their range and transverse scattering thanks to the muon's general reluctance to deflection and stopping and high speed [6].

Chapter 2

Data Analysis

To minimize the experimenter's bias to analyze the experimental data the Belle experiment adopted the method of blind analysis. To compile and tune the analysis program without directly observing the actual experimental data, we used Monte Carlo (MC) simulations. To validate these simulations, we used a control data sample and the hidden signal box method [40].

Although the analyzed decay of $B^0/\bar{B}^0 \to D_s^{*\pm}+\rho^{\mp}$ is well-defined, the particles resulting from this decay are not directly visible to the detector, hence a suitable decay cascades had to be chosen. The set of directly observable particles by the Belle detector prompted us to choose three decay cascades each with the same families of final state particles (FSP). The probed decay cascades were as seen in table 2.1. These decays have known branching ratios listed in table 2.2.

To lower the computational time, the Belle collaboration uses data preprocessing, called skimming. This skimming procedure consists of applying selection cuts on experimental data or simulations. Skim module was very simple to allow for fast data discrimination. The skimming selection criteria can be seen in table 2.1. Definition of some of these criteria is given in section 2.1. The criteria were selected to be wide enough to pass signal events even with considerably simpler module than our main analysis module and to enable sidebands validation, which is discussed in section 2.6.2.

2.1 Particle reconstruction

To discriminate between charged particles, in our case kaons and pions, originating from the IP and those originating from background radiation, cuts on the impact parameters dr and dz was used. These impact parameters are radial and z positions of the point of closest approach of charged track to the IP and their value was set to dr $< 0.5\,$cm and $|\mathrm{d}z| < 3\,$cm. Additionally, to obtain a good

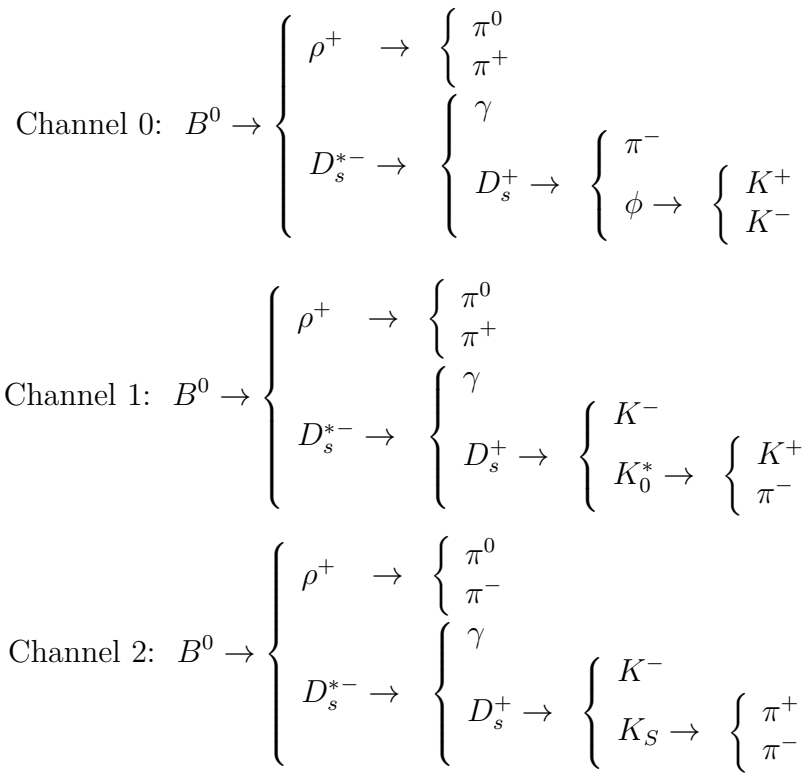


Figure 2.1: Decay channels considered in analysis

reconstructed B vertex, charged tracks with less than 2 SVD hits in $r-\phi$ and z-plane were discarded [44].

For purpose of distinguishing kaons from pions, a likelihood cut was imposed. The probability likelihood ratio, calculated out of data from CDC, ACC and TOF detectors, for particle type i (j) is defined as $\mathrm{PID}(i/j) = L_i/(L_i + L_j)$. These cuts were set to $\mathrm{PID}(K/\pi) > 0.6$. This cut provides the efficiency of 86.3% (85.2%) for kaons in SVD1 (SVD2) experiment and 9.8% (10.5%) misidentification rate for pions [37].

Reconstruction of neutral particles K_S and π_0 is done by combining their most probable decays. In case of K_S, it is a pair of charged pions, in case of π_0, they are two photons. In case of K_S, a goodKs selection function developed by F. Fang [22] was applied. For π_0 reconstruction, multiple cuts were imposed: mass constraint $115 < M_{\gamma\gamma} < 150\,\mathrm{MeV/c}$, which is roughly 3.5FWHM of π_0 mass distribution and photon energy cut $E_\gamma > 50\,\mathrm{MeV}$ in barrel region, $E_\gamma > 100\,\mathrm{MeV}$ cut in endcap region and a cut on χ^2 of photon mass-constrained vertex fit $\chi^2 < 17$.

The other particles were combined from their decay products as shown in table

Selection criterion	Value
PID	PID(K/π)>0.6
	PID(π/K)<0.4
Impact	dr<0.5 cm
parameters	$\lvert$d$z\rvert$ < 3 cm
M_{bc} cut	$M_{bc} > 5.24$
ϕ mass	1.0 GeV < M_ϕ < 1.1 GeV
K^* mass	0.7 GeV < M_{K^*} < 1.0 GeV
D_s mass	1.8 GeV < M_{D_s} < 2.1 GeV
Mass difference cut	0.09 GeV < $\Delta M_{D_s^*,D_s}$ < 0.26 GeV
Rho mass	0.27 GeV < M_ρ < 1.25 GeV
Photon energy	$E_\gamma > 50$ MeV
ΔE window	$\lvert\Delta E\rvert$ < 0.25 GeV

Table 2.1: Selection criteria used for skimming.

2.1. For each channel, a signal-only MC with 10^5 events was generated through EvtGen and Gsim package provided with Belle Analysis Framework (BASF) [42].

Invariant masses of combined particles were then fitted by Gaussian distribution and a cut on 3σ was applied. When reconstructing B^0 and D_s^* particles, however, a different choice of observables was fitted. Instead of fitting and cutting on invariant mass, mass difference between D_s^* and D_s was used and in case of B^0, beam-constrained mass M_{bc} was used as one of the cut parameters together with energy difference ΔE:

$$M_{bc} = c^{-2}\sqrt{E_{\text{beam}}^2 - |\vec{p}_B|^2 c^2} \tag{2.1}$$

Decay	BF
$B^0 \to D_s^{*-}\rho^+$	$(4.1 \pm 1.3) \times 10^{-5}$
$\rho^\pm \to \pi^\pm\pi_0$	$\sim (100)\%$
$D_s^{*\pm} \to \gamma D_s^\pm$	$(93.5 \pm 0.7)\%$
$D_s^\pm \to \phi\pi^+$	$(4.5 \pm 0.4)\%$
$D_s^\pm \to K^\pm K_S$	$(1.50 \pm 0.05)\%$
$K_S \to \pi^+\pi^-$	$(69 \pm 0.05)\%$
$\phi \to K^+K^-$	$(49.2 \pm 0.5)\%$
$D_s^- \to K_0^* K^-$; $K_0^* \to \pi^- K^+$	$(1.9 \pm 0.4) \times 10^{-3}$

Table 2.2: Branching ratios of used decays obtained from PDG [45]

Cut quantity	**Low cut[GeV]**	**High cut[GeV]**
M_ϕ	1.006	1.030
M_{K^*}	0.818	0.965
M_{D_s}	1.954	1.982
$\Delta M_{D_s^*,D_s}$	0.130	0.160
ΔE	-0.25	0.25
M_{bc}	5.27	-

Table 2.3: Cuts before optimization.

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

These cuts can be seen in table 2.3. Each cut was later optimized.

Also cut on daughter photon energy of D_s^* was applied to $E_{\gamma,D} > 50\,\text{MeV}$ which was calculated as an approximate lower bound for this decay. This cut was later optimized also. To obtain a more accurate D_s decay vertex, a mass-constrained fit was performed with its daughters.

The reconstruction of B candidate from the set of available candidates, should

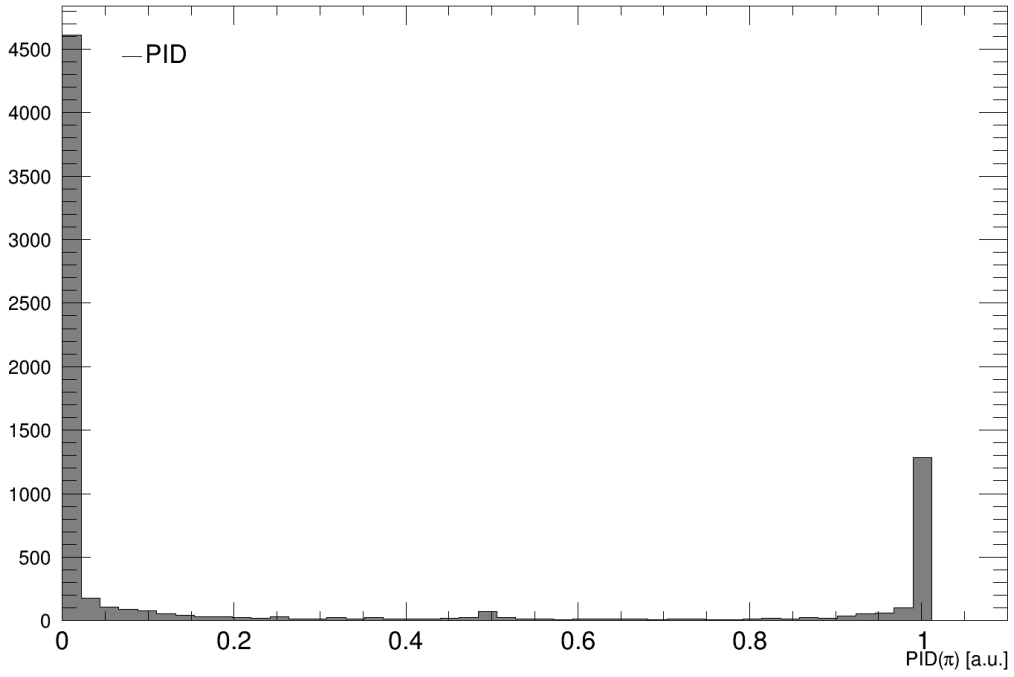


Figure 2.2: π/K likelihood function

there be multiple candidates fo B-meson, called best candidate selection (BCS), was performed by minimizing the χ^2 of $M_{bc} - \chi^2_{M_{bc}} = ((\bar{M}_{bc} - M_{bc})/\sigma_X)^2$. In this formula, $\bar{M}_{bc}$ is the mean M_{bc} and $\sigma_X = \chi^2(M_{\pi_0}) + \chi^2(M_{bc})$. We obtained the value of $\chi^2(M_{bc})$ from signalMC fit and the value of $\chi^2(M_{\pi_0})$ from PDG [45].

2.2 Cut Optimization

Each cut was optimized using the Figure-of-Merit (FOM) optimization. The FOM is defined as:

$$FOM = \frac{n_S}{\sqrt{n_S + n_B}} \tag{2.3}$$

where n_S and n_B are the numbers of signal or background respectively, and provides a measure of performance. The methodology of optimization consisted of tightening cuts on each single variable and calculating the FOM for each cut. The cut with the best FOM was chosen in the end.

However, the cuts considered were heavily correlated, changing a single cut value changed the numbers of signal and background for each successive optimization procedure. Hence a suitable variable order for cut optimization had to be found. It is also important to mention, that an ideal optimization procedure would require multiple optimization iterations over each variable. Four methods of finding the desired order were explored:

1. The first and simplest variable order was linear and starting from the “bottom" to the “top" - from the invariant mass of the very first combined particle (in this case $\phi/K_S/K_0^*$).

2. Using the “bottom-to-top" ordering and iterating over it multiple times.

3. Ordering variable cuts by the FOM and iterating multiple times.

4. Finding the best order from all of the possible variable orders.

The last method would necessitate performing cut optimization technique on each variable order, which would be too computationally demanding, hence it was abandoned first.

The rest of the optimization techniques were explored. The results of the study can be seen in table 2.5. The dataset used for cut optimization consisted of 10^6 signal data and 6 streams of generic MC data normalized through expected branching ratios and stream numbers respectively to expected levels of signal and background. However, it is important to mention, that these datasets that the optimization study was performed on, were test datasets, that had loose enough

Variable	**Cut low [GeV]**	**Cut high [GeV]**
M_ϕ	1.006	1.030
M_{K*}	0.818	0.965
$M_{K_S}0.4$	0.6	
M_ρ	Ch0: 0.688 Ch1: 0.425 Ch2: 0.425	0.950
M_{D_s}	1.954	1.982
$\Delta M_{D_s^*,D_s}$	0.1308	0.160
E_{γ,D_s^*}	0.160	-
M_{bc}	5.276	-
$BDTG$	Ch0: 0.8 Ch1: 0.8 Ch2: 0.6	-

Table 2.4: Cuts after optimization procedure. For the explanation of the $BDTG$ classifier variable, see section 2.3. Ch0 is channel with ϕ decaying into FSP, Ch1 channel with K^* and Ch2 with K_S.

cuts and contained numbers of events similar, but not exactly identical to the actual working datasets.

The results of the study favour the method of ordering from the FOM-ordered method, which although comparable in FOM with the “bottom-to-top" approach with iterations in channel 1 (K^* mode) and 2 (K_S mode), is a little lacking in FOM in channel 0 (ϕ mode). Where the FOM-ordering wins over the “bottom-to-top" approach with iterations, however, is its computational time, which is considerably better, converging after only 2 iterations.

Finally, the FOM-ordering method was applied to the working dataset. The results of the cut-optimization procedure using FOM-ordered iterative method can be seen in table 2.4. During this cut optimization, the rareMC (see section 2.4) was considered.

We also used 6 streams of genericMC and 2.2 million signalMC events containing $B^0 \to D_s^{*-}\rho^+$. The non-resonant $B^0 \to D_s^{*-} + \pi^+ + \pi_0$ decays were not considered for the cut optimization (they weren't included in signal, since it is indistinguishable, yet unwanted decay mode and not included in background, since they do not contribute into background) except for its residue (for the definition of $B^0 \to D_s^{*-} + \pi^+ + \pi_0$ decay residue, please see the section 2.4). This was done because the two decays are not distinguishable by the method we are using for calculating the yield (see section 2.5). All data were normalised appropriately.

	BTT 1	**FOM**	**BTT 2**
$\boldsymbol{n_I}$	1	2	3
$\boldsymbol{n_{\mathrm{S,0}}}$	22	20	19
$\boldsymbol{n_{\mathrm{B,0}}}$	228	154	155
$\boldsymbol{FOM_0}$	1.420	1.479	1.452
$\boldsymbol{n_{\mathrm{S,1}}}$	15	15	15
$\boldsymbol{n_{\mathrm{B,1}}}$	24	41	41
$\boldsymbol{FOM_1}$	1.942	1.948	1.948
$\boldsymbol{n_{\mathrm{S,2}}}$	13	15	15
$\boldsymbol{n_{\mathrm{B,2}}}$	28	34	34
$\boldsymbol{FOM_2}$	2.061	2.199	2.199
$\boldsymbol{FOM_{\mathrm{tot}}}$	2.669	2.993	2.934

Table 2.5: Comparison of optimization methods for thre channels. BTT 1 method signifies “bottom-to-top" approach without iterating, FOM signifies FOM ordering of cuts and BTT 2 signifies “bottom-to-top" method with iterations. n_I signifies the number of iterations and the integer subscript channel number. Please note that the n_{B} and n_{S} numbers are rounded, while FOM was obtained from unrounded numbers

As one can spot in table 2.4, much of the cuts were unchanged by the cut optimizer. A decision has been made to not widen the base cut range and use the basic, physical, 3σ cuts.

2.3 Continuum Suppresion

The e^+e^- collisions produce except from $B\bar{B}$ pairs (charged or neutral) also a nontrivial number of events that do not contain any B-meson pair or contain a random background. This random background predominantly originates in processes of $e^+e^- \to \bar{q}q$, where $q \in \{u, d, s, c\}$ is a quark. Thanks to the nature of the experimental setup generating $\Upsilon(4S)$ resonance peaking at (10,579.5±1.2) MeV and the high mass of B-mesons: (5.279.63±0.15) MeV for B^0 [45], only circa 20 MeV is left for the B^0 momentum, so it is safe to suppose that these events exhibit a different topology from the undesirable background, since $q\bar{q}$ pairs are in general much lighter and the transfered momentum is much higher. For events not containing a B-meson pair, the expected distribution of observed particles is jet-like, whereas the B-pair events ought to have FSPs almost uniformly spherically distributed.

To identify those events that contribute to the generic uniform background, called continuum background, a powerful tool was developed in form of Fox-

Wolfram moments, or better yet, the improved Kakuno-Super-Fox-Wolfram (KSFW) moments. The original Fox-Wolfram $k-$th moment is defined as follows:

$$H_k = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j P_k(\cos\theta_{ij})|}{E_{\text{vis}}^2} \tag{2.4}$$

where P_k is the $k-$th Legendre polynomial, θ_{ij} the opening angle between $i-$th and $j-$th particle, E_{vis} the sum of measured energy in the event and $\vec{p}_i/\vec{p}_i$ the momentum of i-th or j-th FSP respectively.

It is helpful to normalize the moment:

$$R_k = \frac{H_k}{H_0} \tag{2.5}$$

which in continuum events with two jets have value close to 0 (1) if k is odd (even) and in events with spherical topology tend to have different values.

KSFW momets are 17 different variations of moments with the same definition as eq. 2.4, however the summation set is different and contains specific particles. There are three groups of KSFW moments - R_k^{so}, R_k^{oo} and R_k^{ss}. The summation is over particles coming just from the reconstructed signal B-meson (denoted by superscript ss), the other B-meson, rest-of-event particles respectively (denoted by superscript oo) or the combination of both (superscript so).

To obtain better continuum classification, also the $\cos\theta_{\text{thr}}$ was used, where θ_{thr} is a thrust angle, defined as an angle between the thrust axis of the momenta of the B candidate and the thrust axis of the rest of event, while the thrust axis is defined as a unit vector $\vec{T}$ maximizing an expression:

$$T = \max \frac{\sum_i |\vec{p}_i \cdot \vec{T}|}{\sum_i |\vec{p}_i|} \tag{2.6}$$

The value of $|\cos\theta_{\text{thr}}|$ is expected to peak at 1 for continuum and be flat for $B\bar{B}$ events [16].

In order to formulate a sensible prescription for continuum suppression, Multivariate Analysis (MVA) was used, or particularly Toolkit for Multivariate Data Analysis of ROOT data analysis framework. Not only to enable multivariate analysis for continuum suppression, Belle collaboration has generated a complex MC simulation (denoted as genericMC) with 6 times the real experimental data size (we call it 6 streams from now on). The total luminosity of each stream of the MC dataset is set to match the real luminosities that can be seen in table 1.1.

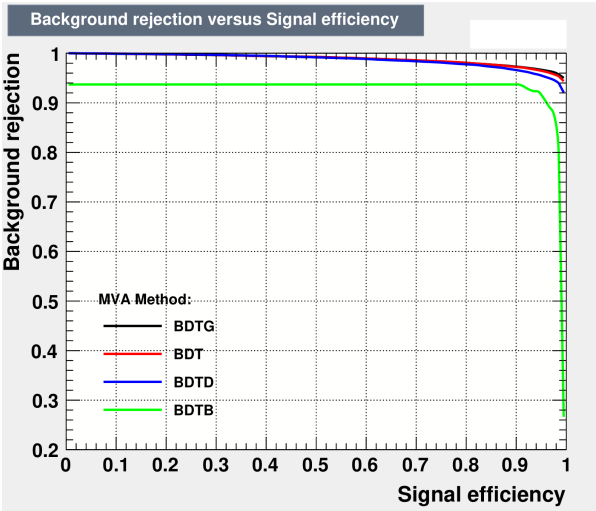
ROOT TMVA Toolkit contains methods listed in table 2.6. Firstly, those methods were trained with default settings and their performance characterized by their receiver operating characteristic (ROC) curve was assesed. Methods with

biggest area under their ROC curve were selected for a final comparison between groups of methods as defined in table 2.6. ROC curves for single groups of methods can be seen on figure 2.3. As we can see from figure 2.3, methods selected from their groups were:

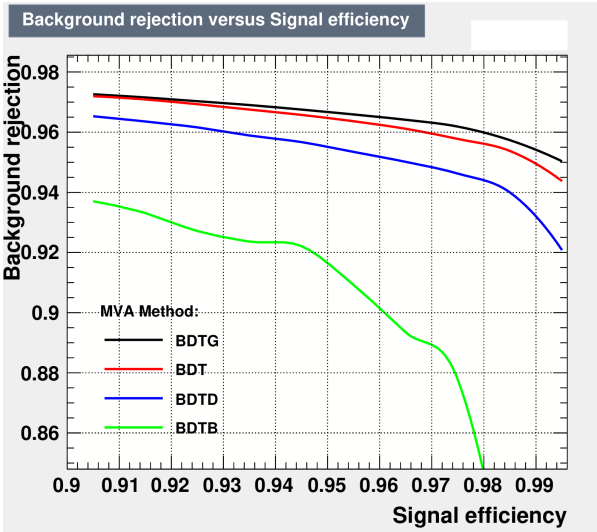
- CutsGA from Cuts
- FDA MC from function discriminant (FDA)
- BoostedFisher from linear discriminant (LD)
- LikelihoodKDE from likelihood
- MLP from artificial neural networks
- BDTG from boosted decision trees

The comparison of ROC curves between the best methods out of their respective method groups can be seen on figure 2.4

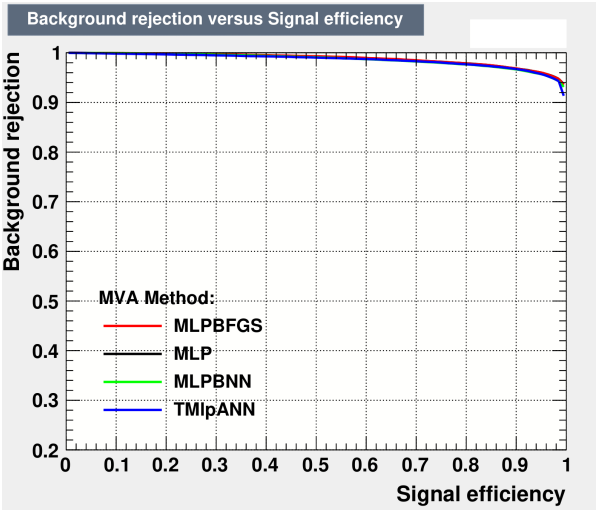
The detailed description of the simpler methods are available in [23], we will describe the basics of boosted decision trees and artificial neural networks here, since they've become very popular in the past few years.



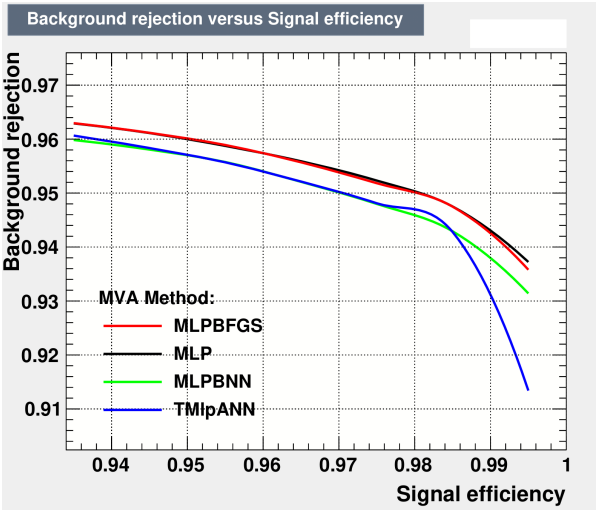
(a) BDT methods



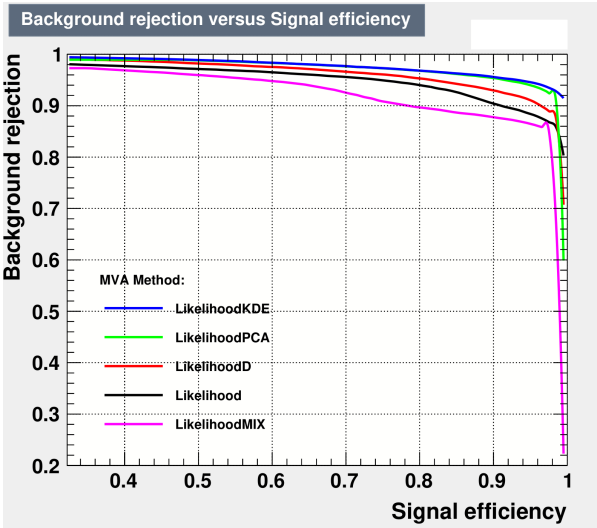
(b) Zoomed BDT methods



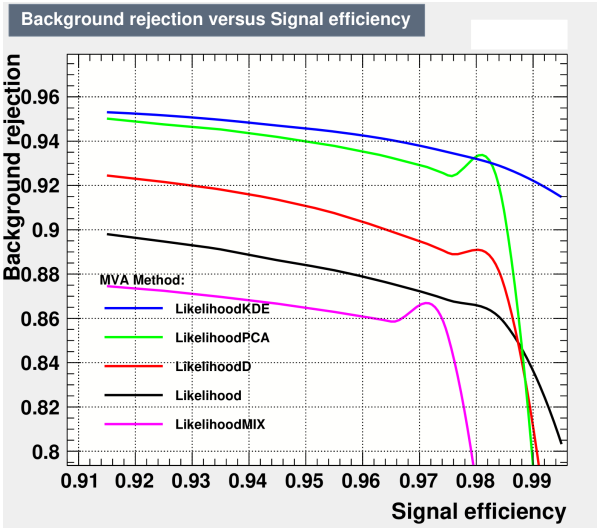
(c) ANN methods



(d) Zoomed ANN methods

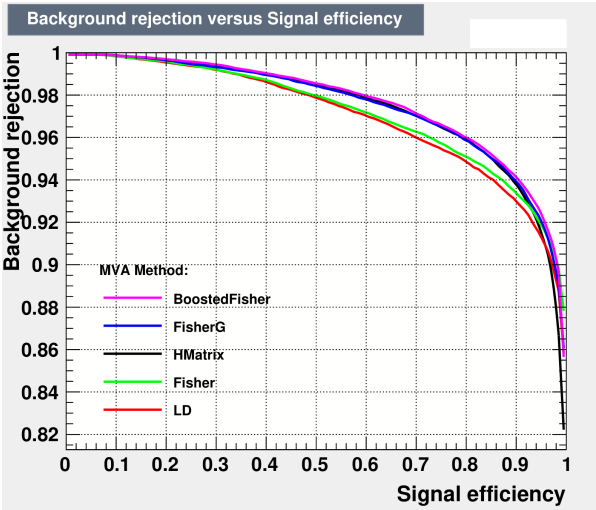


(e) Likelihood methods

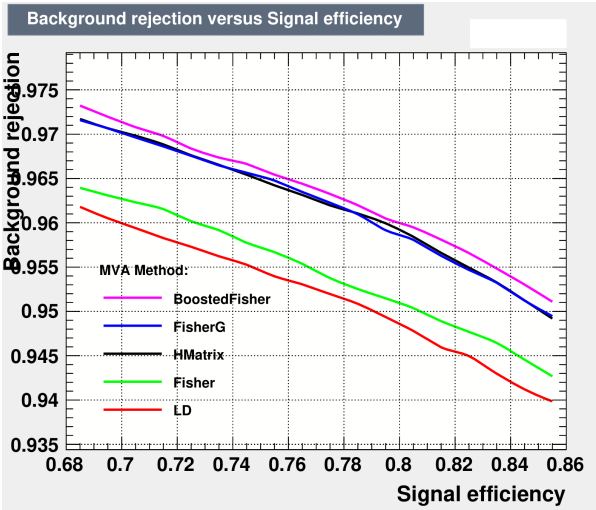


(f) Zoomed likelihood methods

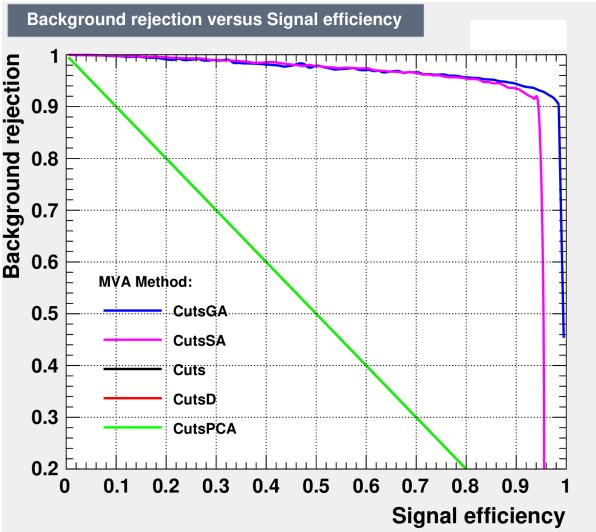
Figure 2.3: TMVA method comparison between groups of methods - ANN, BDT and likelihood methods



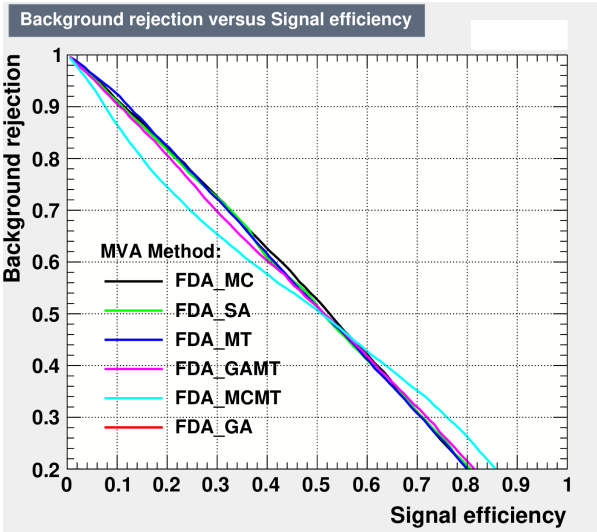
(g) Linear discriminant methods



(h) Zoomed linear discriminant methods

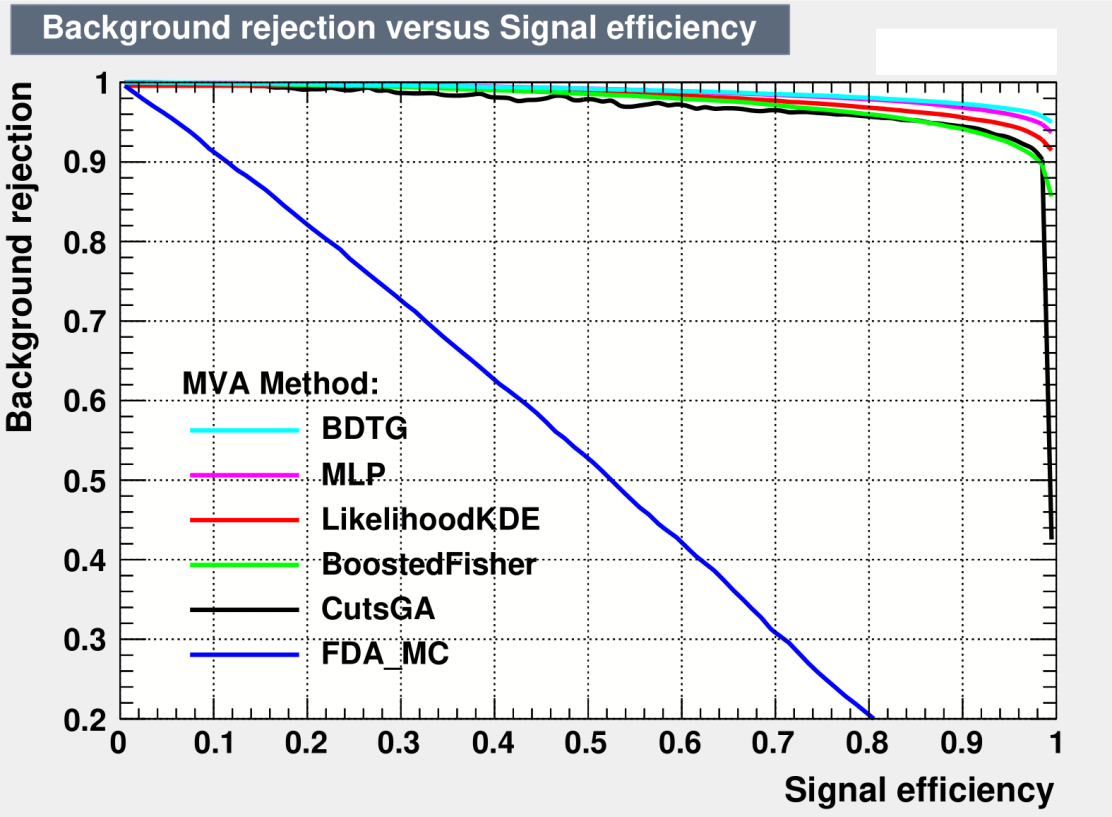


(i) Cuts methods

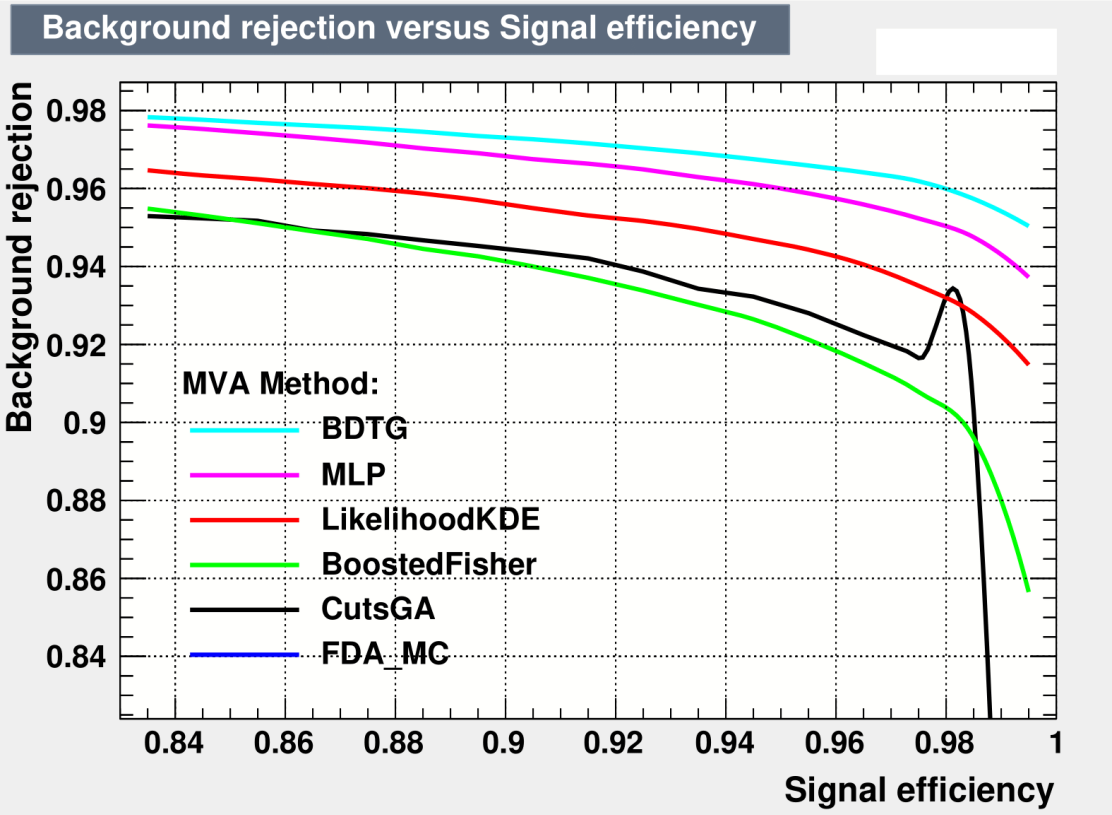


(j) Function discriminant methods

Figure 2.3: TMVA method comparison between groups of methods - Linear discriminant and function discriminant methods



(a)



(b)

Figure 2.4: Comparison between the best methods of their groups defined in table 2.6 (a) and zoomed right top corner (b)

Method	Converged	Group
Cuts	No	Cuts
CutsD	No	
CutsPCA	Yes	
CutsGA	Yes	
CutsSA	Yes	
Likelihood	Yes	Likelihood
LikelihoodD	Yes	
LikelihoodPCA	Yes	
LikelihoodKDE	Yes	
LikelihoodMIX	Yes	
PDERS	No	Nearest neighbourhood
PDERSD	No	
PDERSPCA	No	
PDEFoam	No	
PDEFoamBoost	No	
KNN	No	
LD	Yes	Linear discriminant
Fisher	Yes	
FisherG	Yes	
BoostedFisher	Yes	
HMatrix	Yes	
FDA GA	Yes	Function discriminant
FDA SA	Yes	
FDA MC	Yes	
FDA MT	Yes	
FDA GAMT	Yes	
FDA MCMT	Yes	
MLP	Yes	Artificial neural networks
MLPBFGS	Yes	
MLPBNN	Yes	
CFMlpANN	No	
TMlpANN	Yes	
DNN GPU/CPU	No	
BDT	Yes	Boosted decision trees
BDTG	Yes	
BDTB	Yes	
BDTD	Yes	
BDTF	No	
RuleFit	No	Other
SVN	No	

Table 2.6: List of methods available in ROOT TMVA Toolkit

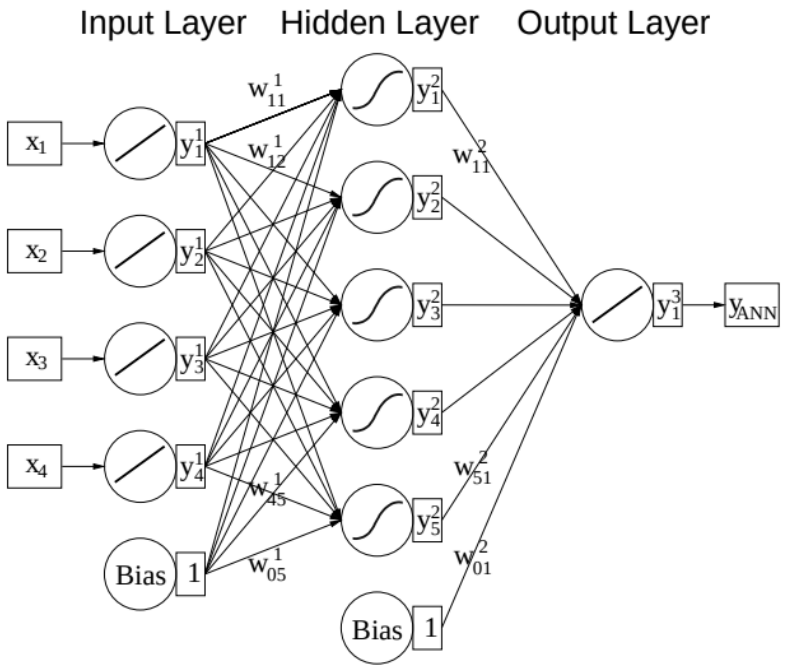


Figure 2.5: A multi-layer perceptron (MLP) ANN with one hidden layer [23]

2.3.1 Artificial Neural Networks

Artificial Neural Networks (ANN) loosely model the network of neurons in a human brain. It consists of multiple “nodes" that model biological neurons, connected via multiple connections, artificial synapses. These nodes are grouped in layers with the first layer being an input layer, the last an output layer and each layer in between called a hidden layer. The behaviour of ANN differs with differing layout of layers, number of neurons and connections between them (an example schematic of an ANN is in figure 2.5). On the input of each node, there is a “signal" represented by a real number and each node contains a function that processes each signal. Nodes are where the computation in the neural network takes place. Each connection to each node is given a weight, that is set during the ANN training. The training consists of a set of data with a known answer that the ANN tries to replicate by setting appropriate weights [23, 16].

One danger of using ANNs is the phenomenon of overtraining. Much like our brain can search for patterns in things that contain none, an ANN with an inappropriate structure and setting can see such non-existent patterns too. A great example of ANN overtraining comes from the field of pattern-sensing on visual data, particularly the DeepDream project created by A. Mordvinstev, that uses convolutional ANN to enhance patterns in images. In this field, the biological,

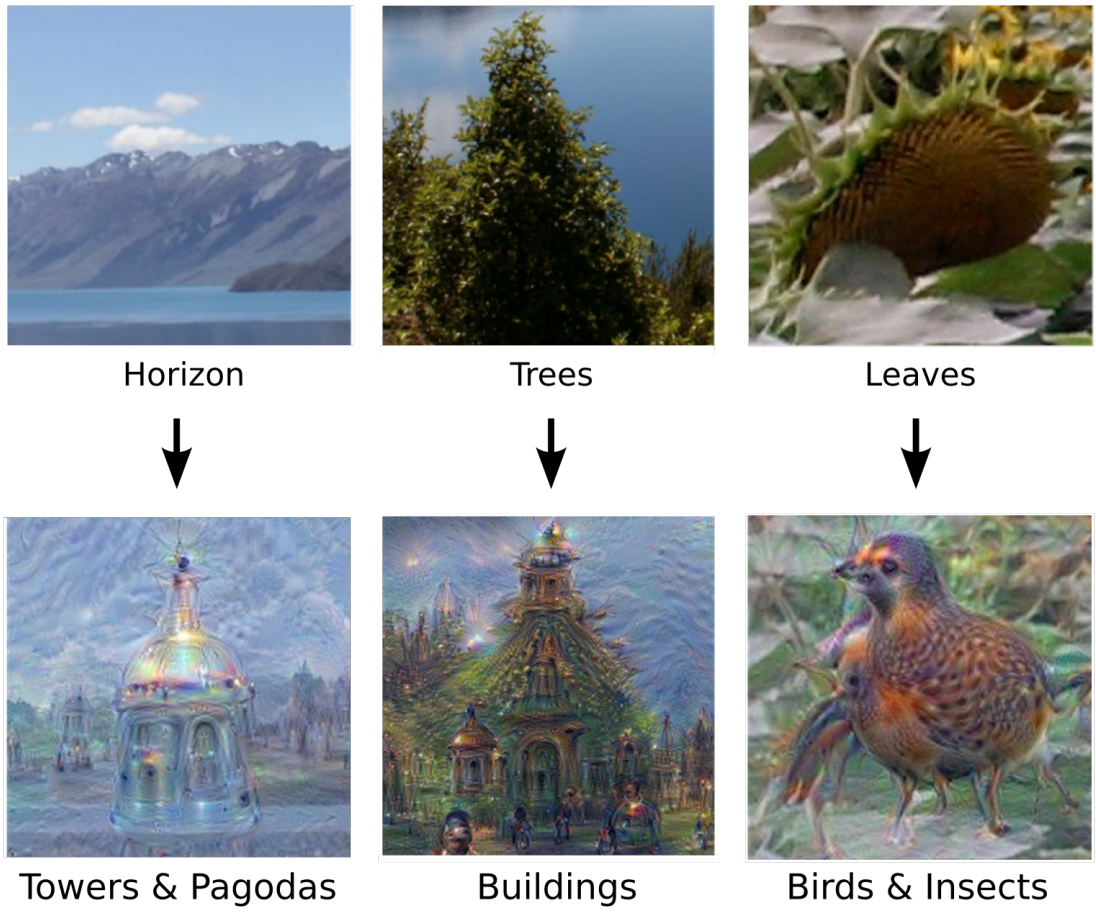


Figure 2.6: Extreme algorithmical pareidolia from DeepDream ANN [33]

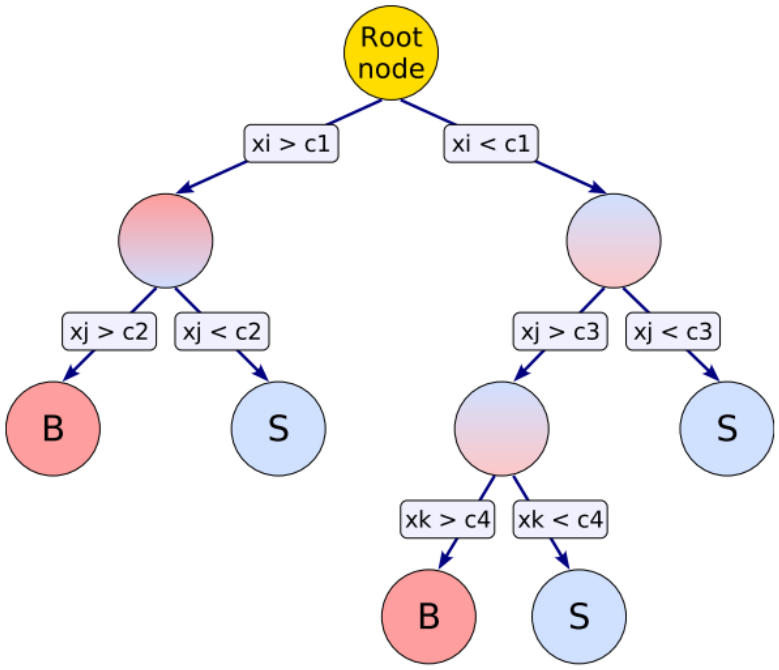


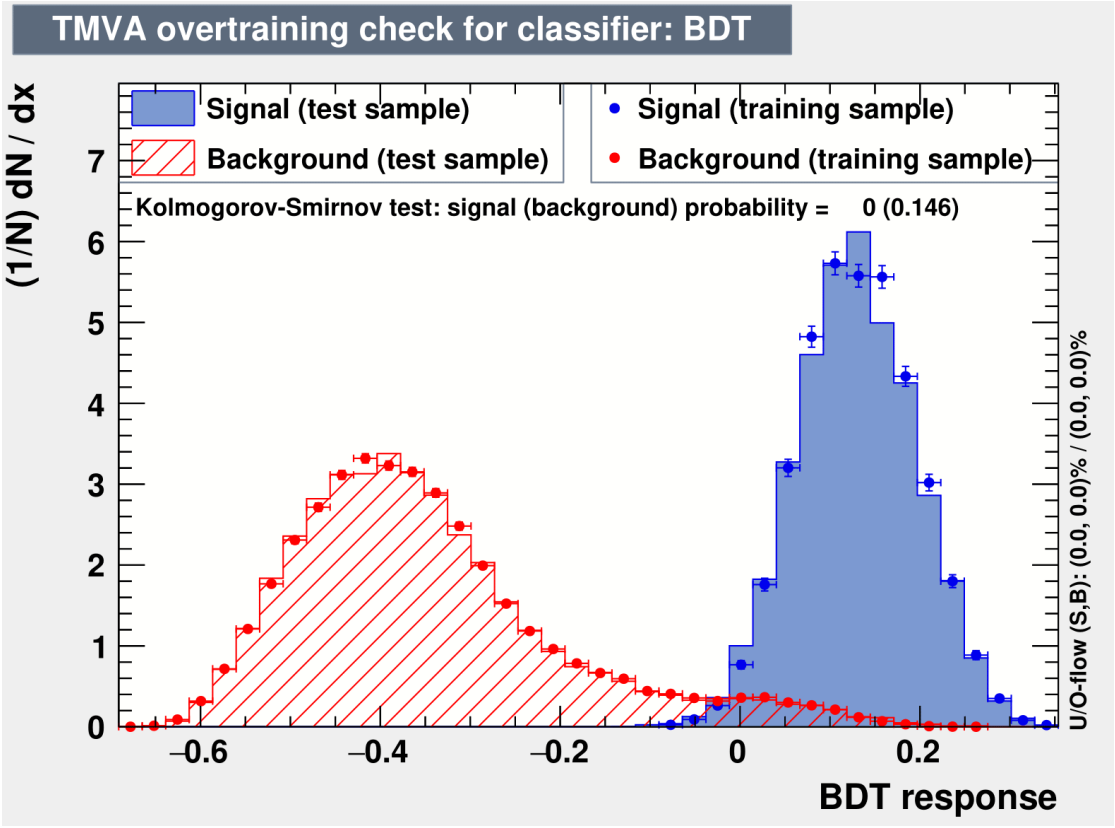
Figure 2.7: A schematic of a binary decision tree [23]

as well as ANN, effect is called pareidolia. An example of extreme algorithmic pareidolia can be seen in figure 2.6 [33].

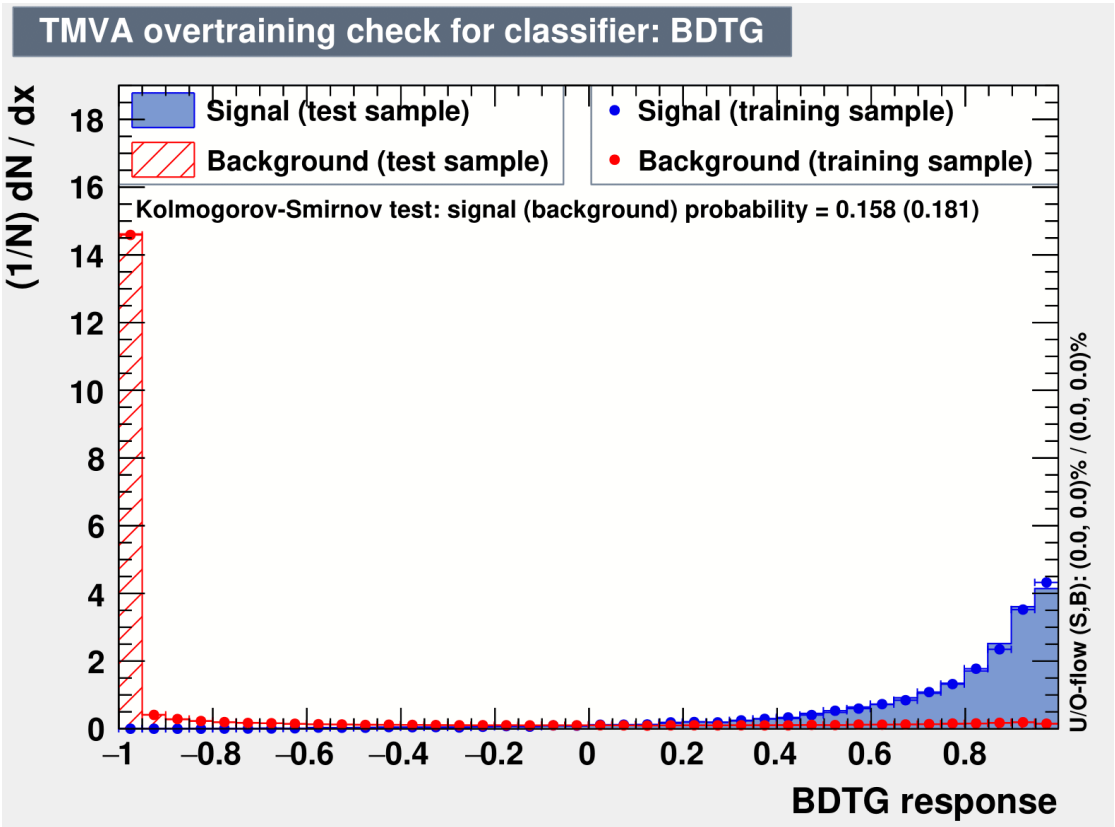
In our case, overtraining leads to a seeming increase in the classification performance (misclassification by overfitting). It results in an effective performance decrease on an indepedent test sample. The TMVA chooses automatically and randomly a test sample from an input training data and provides Kolmogorov-Smirnov test for overtraining check [23].

2.3.2 Boosted Decision Trees

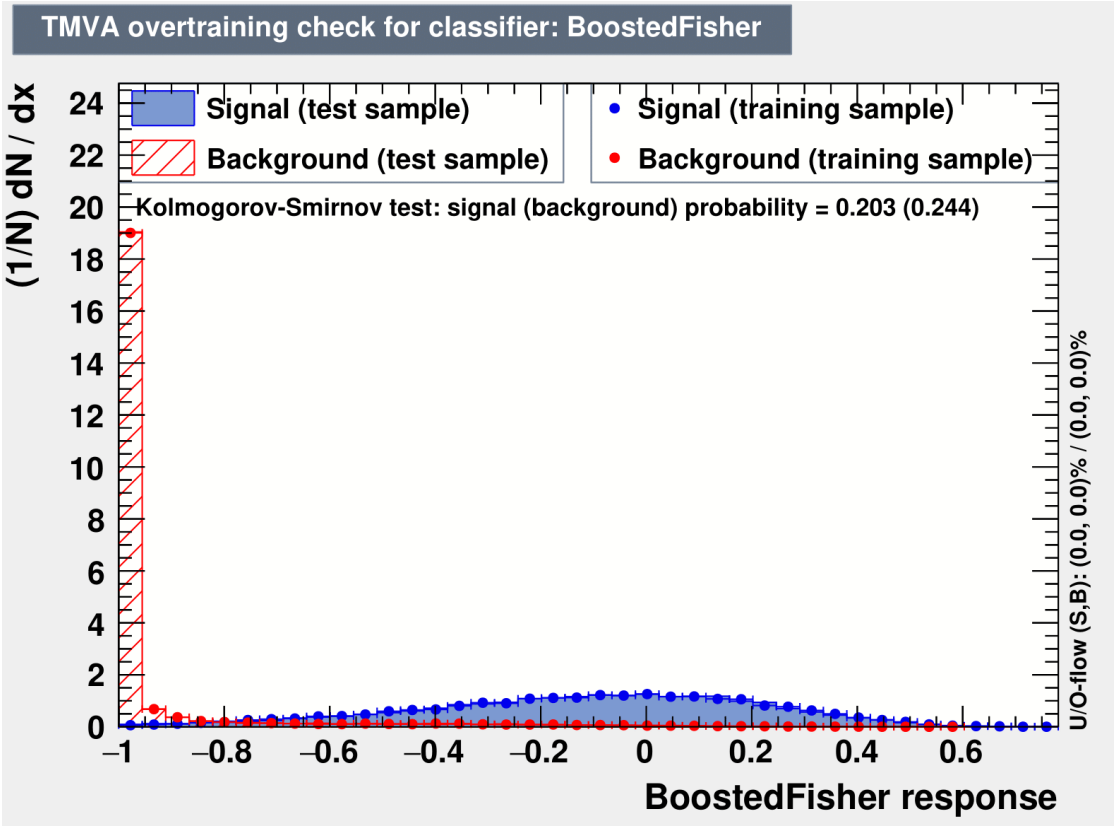
Another form of machine learning are Boosted Decision Trees (BDT). The concept of BDT is based on the concept of binary decision tree sketched in figure 2.7. A binary decision tree asks a question with binary answer (yes/no) at every node of the tree. Each node, or split, uses the variable that at this node gives the best separation between signal and background of the training data. The final nodes, called leaf nodes, provide the final classification of the input data as a signal or background. The concept of boosting is introduced to improve the response of the decision tree. It creates many decision trees by reweighting events and groups them into a “forest". The final single classifier output of the BDT is then given as a weighted average of the individual binary decision trees [39, 23].


TMVA overtraining check for classifier: BDT
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0 (0.146)
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
BDT response

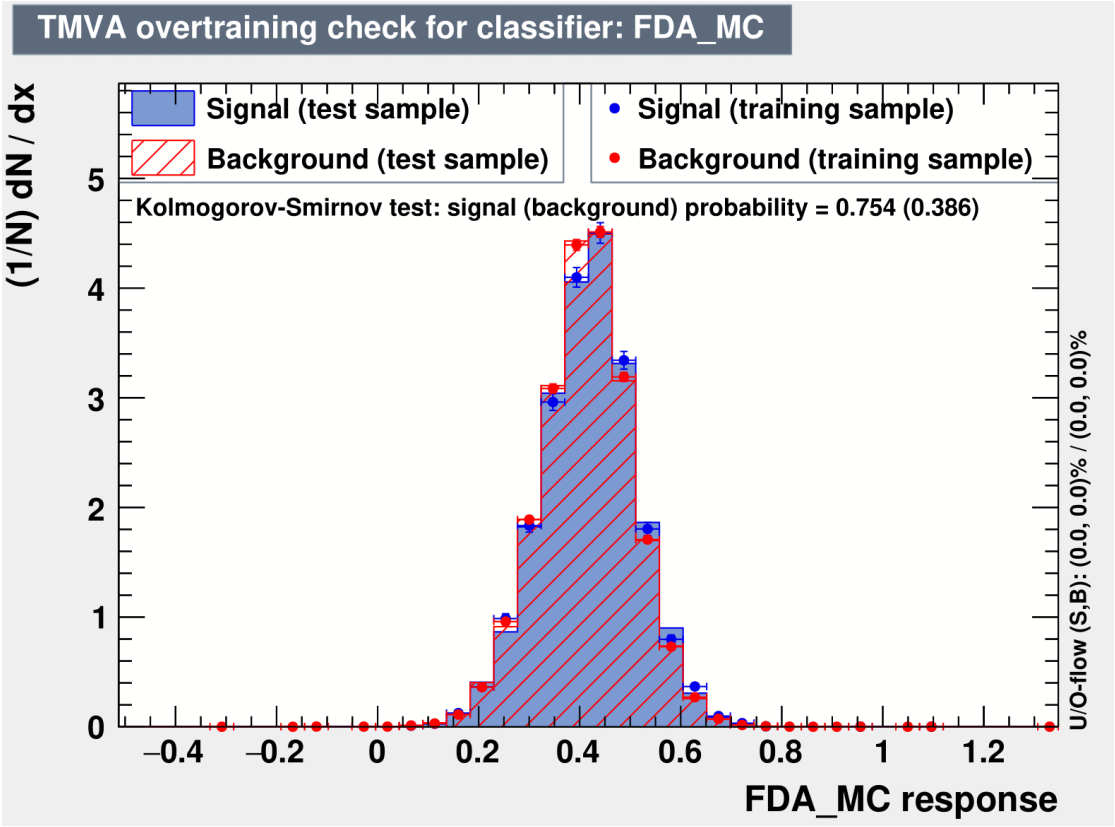
(a)


TMVA overtraining check for classifier: BDTG
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.158 (0.181)
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
BDTG response

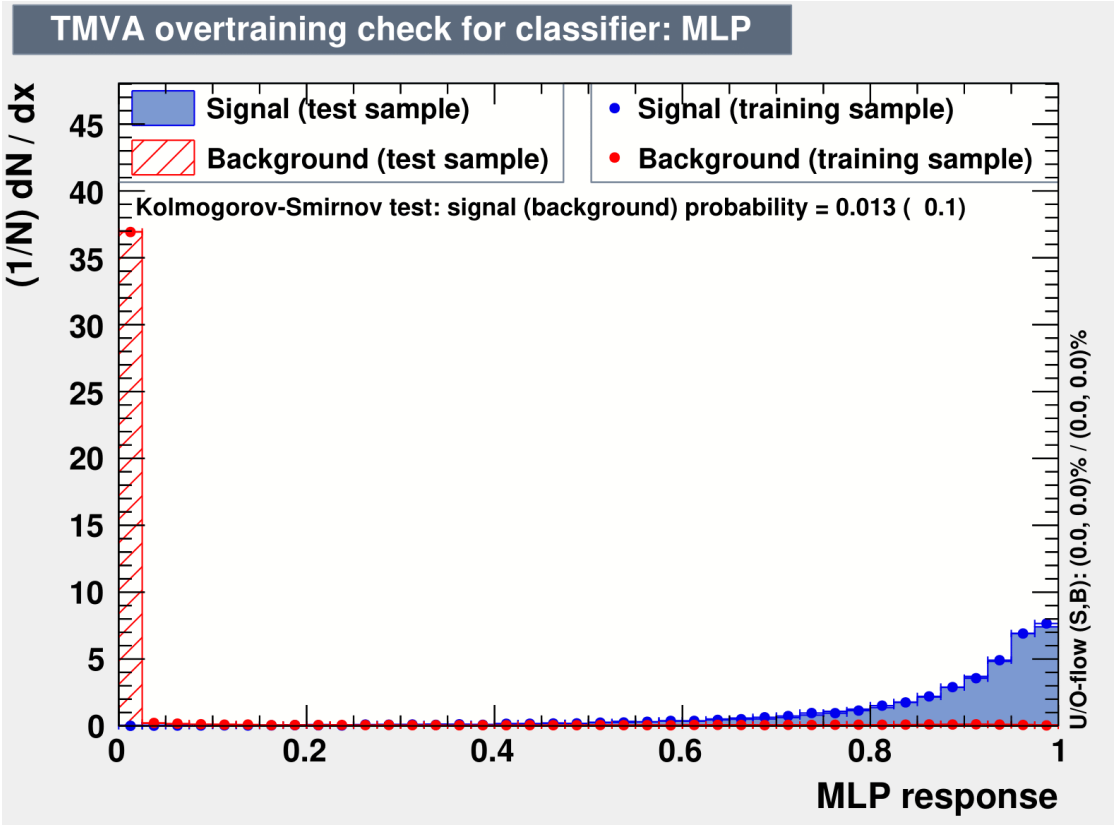
(b)


TMVA overtraining check for classifier: BoostedFisher
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.203 (0.244)
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
BoostedFisher response

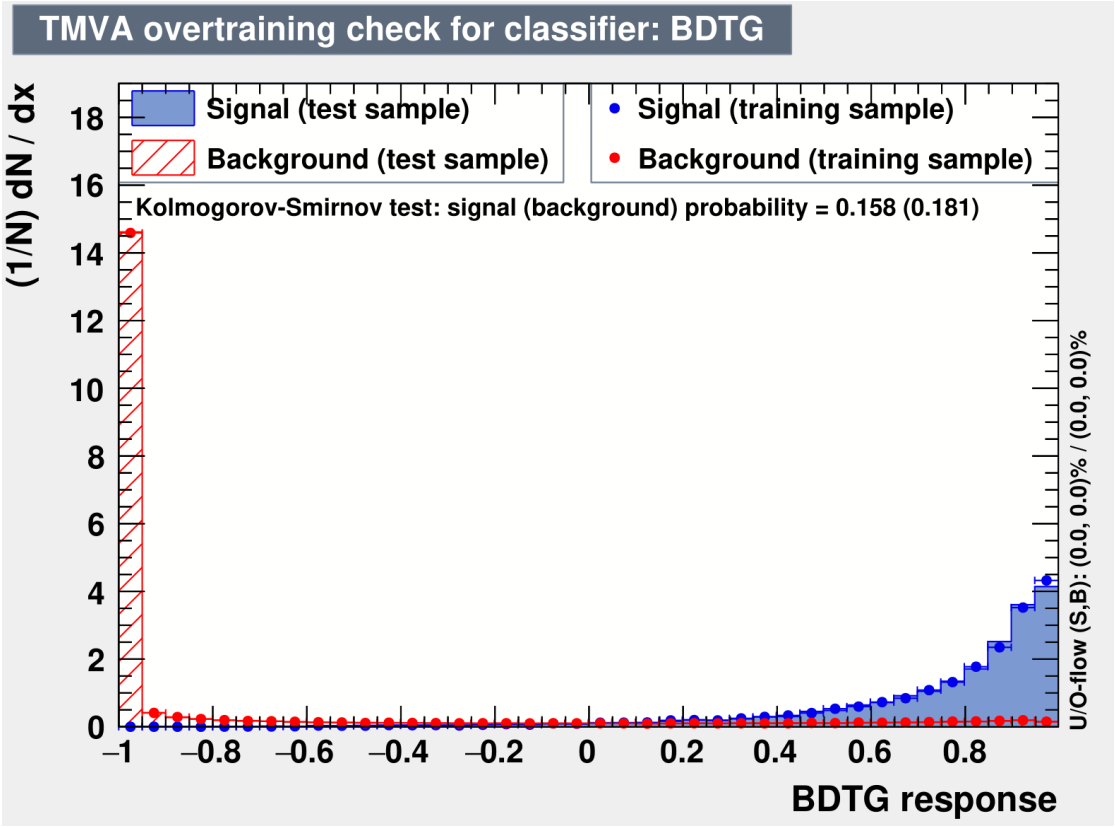
(c)


TMVA overtraining check for classifier: FDA_MC
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.754 (0.386)
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
FDA_MC response

(d)



(e)



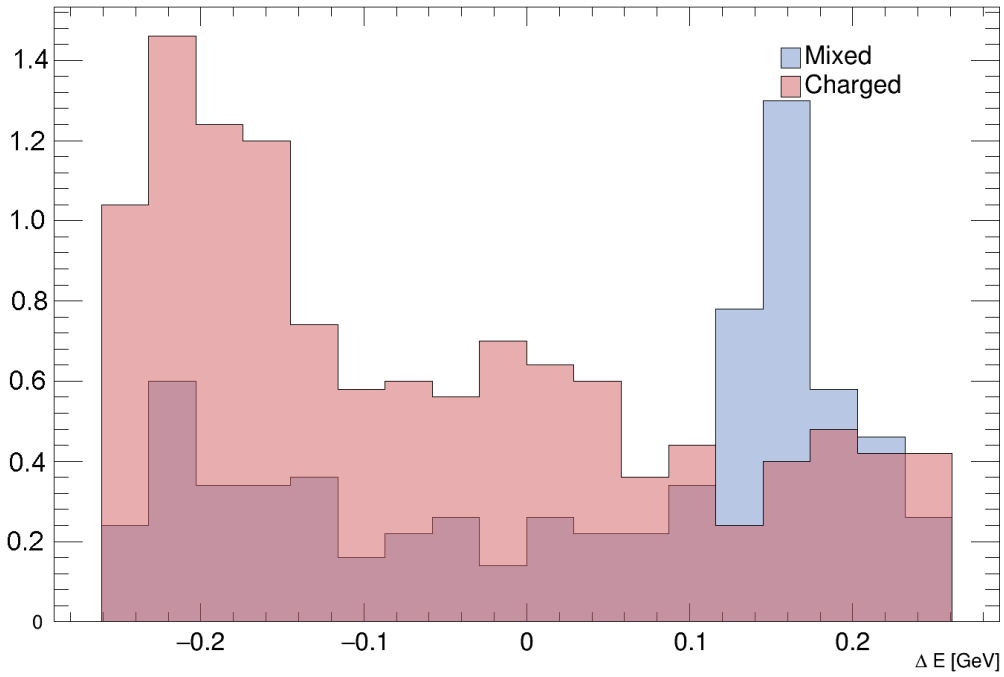
(f)

Figure 2.8: Preview of classifier distributions of test data sample and overtraining checks for multiple considered (a-d) BDT, (e-h) ANN/MLP methods

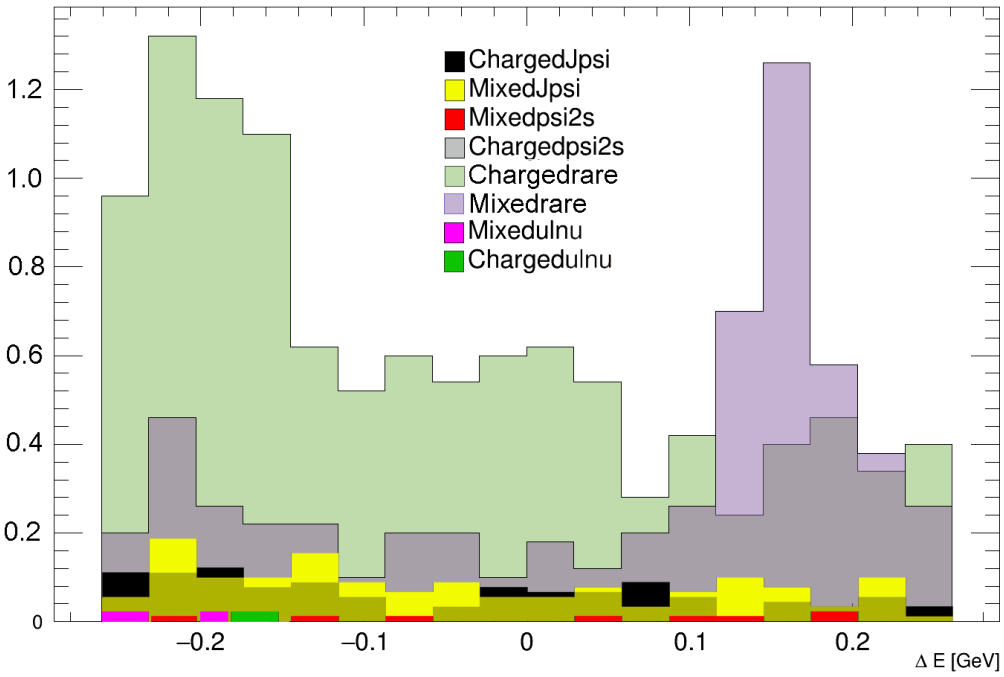
2.3.3 Application of particular method

Another factor that played a role in the final decision of application of a MVA method, was the overtraining. Overtraining check in form of the Kolmogorov-Smirnov test can be seen in figure 2.8. Methods had to be tunable in reasonable time so that Kolmogorov-Smirnov test result was not hinting overtraining.

From the analysis of multiple methods, the final choice was made to apply the BDTG classifier method for continuum suppression trained on dataset comprised of combined data from all channels to improve classification performance with bigger statistical ensemble. The training dataset consisted of signalMC (10^6 events per channel) and selected modes of genericMC (6 streams) - charm and uds, with charm being a component in which a charmed quark-antiquark pair is produced and uds a component in which one of the up, down or strange quarks is produced in quark-antiquark pair. However, the cut on the BDTG classifier was obtained by cut optimization on the whole dataset containig 10^6 signalMC events, all streams of genericMC and all streams of rareMC (for the definition of rareMC see section 2.4).

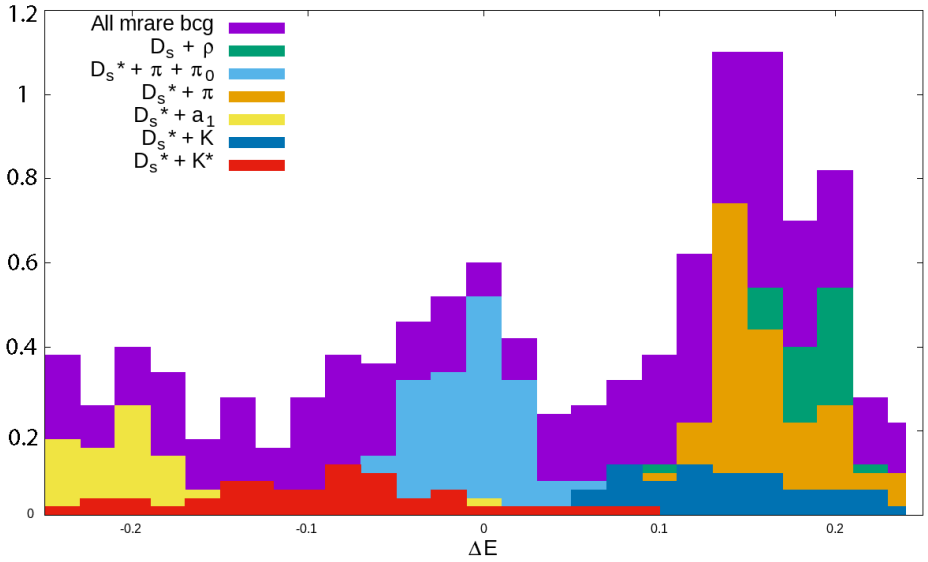


(a)

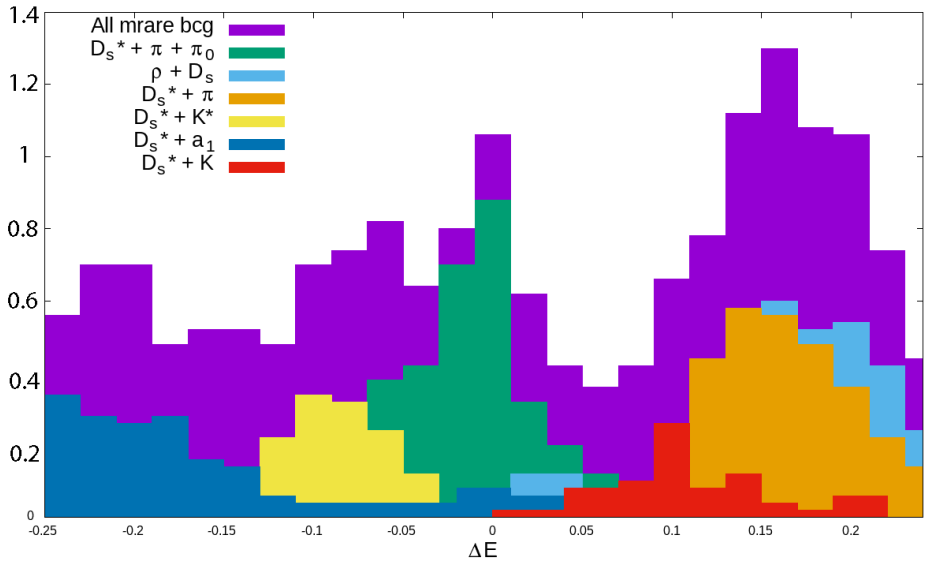


(b)

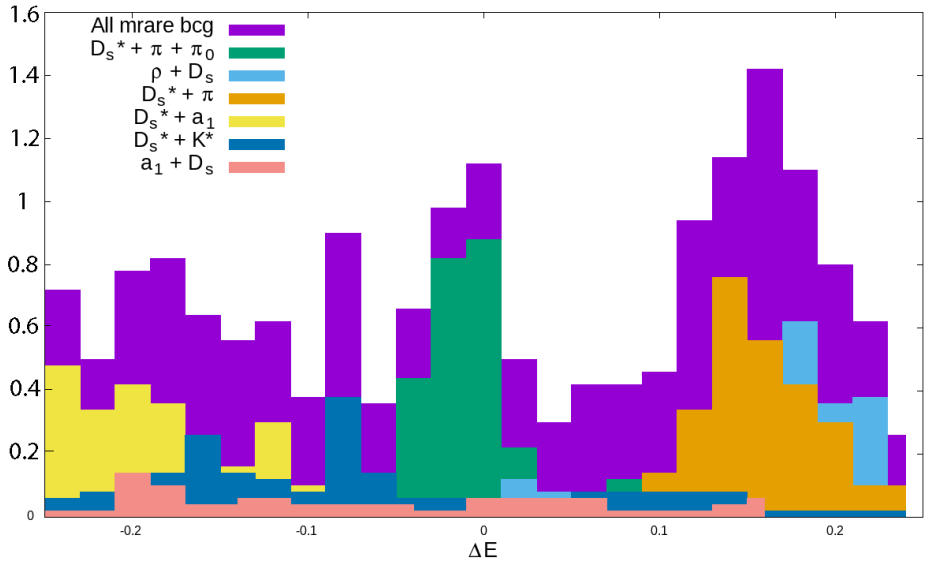
Figure 2.9: Contents of rareMC



(a) Channel 0 (ϕ)



(b) Channel 1 (K^*)



(c) Channel 2 (K_S)

Figure 2.10: Components of *mixedrare* background peaks for every channel

Decay	Ch0 #ev	Ch1 #ev	Ch2 #ev	Simulated BF	Expected BF
$\rho + D_s$	138	177	165	0.000057	<0.000024
$D_s^* + \pi + \pi_0$	123	206	202	0.000058	?
$D_s^* + \pi$	122	154	143	0.0000115	0.000021
$D_s^* + a_1$	59	105	131	0.000094	<0.0017
$D_s^* + K^*$	34	112	95	0.000068	0.000032
$D_s^* + K$	16	42	33	0.0000198	0.0000219
$a_1 + D_s$	19	33	41	0.000043	<0.0021
$K^* + D_s$	10	24	24	0.00008	0.000035
$D_s^* + K + \pi_0$	5	10	10	0.000026	?
$D_s^* + K_0 + \pi$	5	5	5	0.000054	<0.00011
$K^* + K^*$	0	0	6	0.00014100	?
$D_s + \pi + \pi_0$	1	6	2	0.000008	?

Table 2.7: Contents of *mixedrare* background represented in analysis by more than 4 events(#ev means number of events) in at least one channel with simulated and expected BF obtained from PDG [45]. Ch0 is channel with ϕ decaying into FSP, Ch1 channel with K^* and Ch2 with K_S.

2.4 Rare MC Study

The genericMC does not contain only $e^+e^- \to q\bar{q}$ processes. There are 4 genericMC types that contain different types of events: *uds, charm, charged, mixed*. *uds* and *charm* contain the mentioned $q\bar{q}$ processes, but *mixed* contains all of the common $B^0\bar{B}^0$ processes. Similar goes with *charged* genericMC that contains common B^+B^- processes. Our process being rare, does not show up in those MC sets. However, to provide an analysis framework for the rare processes, another MC dataset was generated. It is called rareMC and has components and numbers of streams as listed in table 2.8.

The content of each rareMC component without the signal decay passing the reconstruction can be seen in figure 2.9. As can be seen, the background other than *mixedrare* is slightly decreasing, as expected. The *mixedrare* background is showing multiple peaks. All of the decays contained in *mixedrare* background are listed in table 2.7.

Contents of peaks in *mixedrare* background are shown in figure 2.10 and listed in table 2.7. This figure shows the first 6 most prevalent decays. In this figure, the non-resonant decay $B^0 \to D_s^{*\mp} + \pi_0 + \pi^\pm$ with the decay cascade of D_s^* same as in table 2.1 are suppressed (it does not contain non-resonant decays with the same final state particles as our channels do). However, as one can see, there is still some residue of the non-resonant decay present in table 2.7 and figure

MC component	#streams
mixedrare	50s
mixedJpsi	100s
mixedpsi2s	100s
mixedulnu	20s
chargedrare	50s
chargedJpsi	100s
chargedpsi2s	100s
chargedulnu	20s

Table 2.8: Number of streams (# signifies number) per rareMC component

2.10. This residue of the non-resonant decay is comprised of final state particles not present in our reconstruction which makes it impossible to include into the signal shape with our methodology of non-resonant decay classification. Note that charged particles/antiparticles are marked without their charges on this plot, neutral particles are shown with zero subscript.

2.5 Yield Calculation

The final part of analysis consists of the signal yield calculation. This is done by building a model using MC simulation and then fitting it to ΔE distribution. This is why we did not tighten/optimize ΔE cut in Section 2.2. For fitting, RooFit toolkit of ROOT data analysis framework was used. The final model's probability density function (PDF) comprised of 4 components:

- Signal - obtained from 10^6 events of signalMC (for each channel)
- Selfcrossfeed - events, in which the reconstruction resulted in reconstructing the right decay but either from switched particles (for example the photon originating from neutral pion decay switched for the proton originating from D_s^*) or from particles of the other B-meson decay
- Background - a combinatorial background obtained from 6 streams of genericMC
- Mixedrare - obtained from 50 streams of *mixedrare* component of rareMC without signal events

For Signal, a combination of Crystal Ball distribution and Bifurcated Gaussian was used, for Self-crossfeed (SCF or CF), a Gaussian model with an Exponential

Component	PDF
Signal	Crystall ball with bifurcated Gaussian
Crossfeed	Gaussian with exponential
Background	2nd deg 1st kind Chebyshev polynomial
Mixedrare	2 Gaussian peaks with Landau

Table 2.9: Components and their corresponding PDFs used for yield fit.

was used and the Background was fitted with the second degree Chebyshev polynomial of the first kind. For the Mixedrare model, several different distributions were considered, but we finally settled with two Gaussian peaks and a Landau distribution for the final, rightmost peak thanks to its assymetricity and high tail. The fitted Signal, SCF, Signal+SCF, background and Rare models can be seen in figure 2.11 together with the fit results of model parameters calculated by Minuit2 minimizer through unbinned likelihood method.

The shape of signal, however, was not fitted, but rather set firm. This was done because of the low number of signal events, which would render the confirmation of the signal shape from the data impossible. We chose the shape of the signal PDF to be the same as the shape of $B^0 \to D_s^{*\mp} + \rho^{\pm}$ decay [17]. This decision was based on the fact that [17] studied similar decay $B^0 \to D^{*\mp} + \rho^{\pm}$ and our decay's signal PDF should be dominated by a similar shape. The exact values of parameters used in signal PDFs, can be seen in table 2.11.

Also thanks to low numbers of background events in monte carlo simulations, the three datasets for three channels were merged. They were then fitted by the PDF from table 2.9 and this PDF shape was then used to construct the final PDF in the final yield fits of all three channels.

Fitter validation

To validate the fitting program, toy Monte Carlo method was implemented. This method is purely mathematical method, which comprises of generating random events based solely on input PDF (similar to monte carlo integration). These events are then fitted and the level of accuracy in fitting and error estimation is

	Signal+crossfeed	Background	Mixedrare
Ch0 (ϕ)	15	53	6
Ch1 (K^*)	18	320	10
Ch2 (K_S)	13	164	13

Table 2.10: Generated numbers of events for toyMC.

Variable	Value
μ_{g}	0.0
σ_{gl}	0.012
σ_{gr}	0.016
$t_{0,\mathrm{cb}}$	-0.014
σ_{cb}	0.034
$\mathrm{cut}_{\mathrm{cb}}$	0.904
$\mathrm{power}_{\mathrm{cb}}$	3.687
$f_{\mathrm{g/cb}}$	0.594

Table 2.11: Parameters of signal shape used for fitting. Subscript g signifies the parameter belongs to bifurcated Gaussian and subscript cb signifies that parameter belongs to crystal ball distribution.

checked. This is done by so-called pull plots, the plots of a value $\chi = \frac{x-\mu}{\sigma}$ (not to be confused with χ-distribution), where x is fitted and μ is generated value of some variable and σ its error. If the fitter works properly, the pull distribution should be unit Gaussian.
Our fitted variables were n_{sig}, n_{bkg} and n_{rare}, signifying the numbers of events in each PDF component. Generated numbers of events for each channel can be seen in 2.10. These were based on expected numbers of signal, background and mixedrare components calculated by either dividing the numbers of events passing the analysis by the number of MC streams or by the expected branching fraction of respective channel obtained from PDG [45]. Besides fitting Gaussian distribution to pull plots, plots of numbers of events in each component were created and fitted. This served as an extra layer of validation. $n_{\mathrm{sig+cf}}$ plots were fitted by Gaussian convoluted with Landau distribution with Landau most probable value set to 0. This was done because the distribution of number of signal events, thanks to non-negative numbers of events generated, had a bias towards positive numbers making the Gaussian asymmetrical with high right tail. This fit might not provide any information, since the choice of the fitted distribution might seem random, however, it can at least provide an information about convergence of fits of the toyMC dataset. Off course, Gaussian fits do provide some information about the mean number of events and a spread of each variable that can help set its range before fitting. These can be seen in figures 2.12 for channel 0 (ϕ), 2.13 for channel 1 (K^*) and 2.14 for channel 2 (K_S mode).

To obtain reasonable behaviour of the yield fitter, fits on negative numbers of events had to be allowed. Constraining the fitter to positive numbers tended to result in accumulation of events at 0, skewing the pull plots of channel 1 (K^*) and channel 2 (K_S). This effect can be seen in figure 2.15-c,e,f. As one can see, the

fits on that plots have very high χ^2.

3 methods were explored to solve this issue:

1. Cut the dataset on $n_{\text{sig}} < 1$
2. Do not fit mixedrare component - because of high random fluctuations of background and low numbers in mixedrare, try to not include mixedrare component and fit only signal+crossfeed and background
3. Allow for negative numbers for all fitted variables

The result of method 3 we've already shown on figures 2.12, 2.13 and 2.14. The results of method 1 can be seen on figure 2.16. As we can see, the improvement is not great. The issue with low σ of pull Gaussian still persists as well as the split of mixedrare pull into two peaks.

Similarly for the second method. That can be seen on figure 2.17. As there is no mixedrare component, there is no mixedrare peak splitting, however, the issue with low Gaussian σ persist, with the signal pull's mean being shifted towards positive numbers. This is a result of the peak close to 0 in mixedrare component in channel 2 (K_S) as can be seen in figure 2.11-k.

As we can see, a result in which the fit to the pull plot resembles a unit Gaussian the most provides the last method, unconstraining the fitted variables.

A fourth method was suggested by the analysis group of the Belle experiment, to fix mixedrare number of events to zero when appropriate and obtain errors by variational methods. This solution, however, might be too simple and produce positive bias although it is possible the positive bias will be negligible. It is also too time consuming and for this reason was not explored.

Study of these last two methods might be a topic for further improvement of the fitting procedure. The unconstrained variable method might provide in extreme cases more fitted signal events than the total number of events in datasets. In this case we are not sure how it will manifest itself in estimating the yield error.

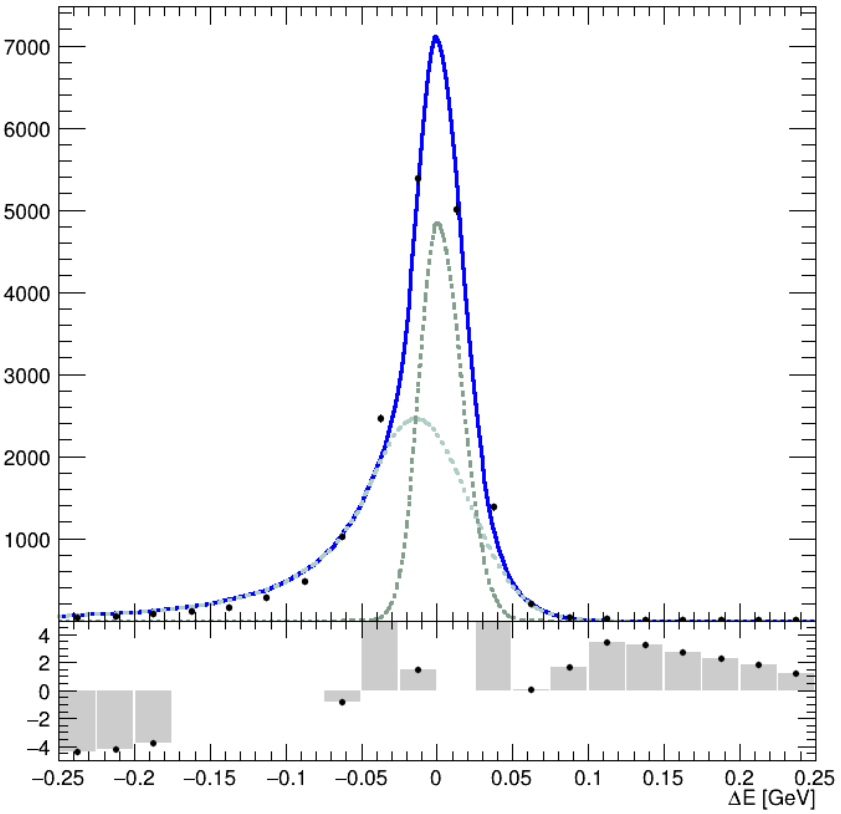
Test datasets

6 different test datasets were then generated, each containing 1 stream of background from genericMC, 1 stream of rareMC and an expected number of signal events calculated from the Belle integrated luminosity and expected BF. These were then fitted with the model created from the MC. Results for unconstrained number of events can be seen in figure 2.18 for channel 0 (ϕ), 2.19 for channel 1 (K^*) and 2.20 for channel 2 (K_S).

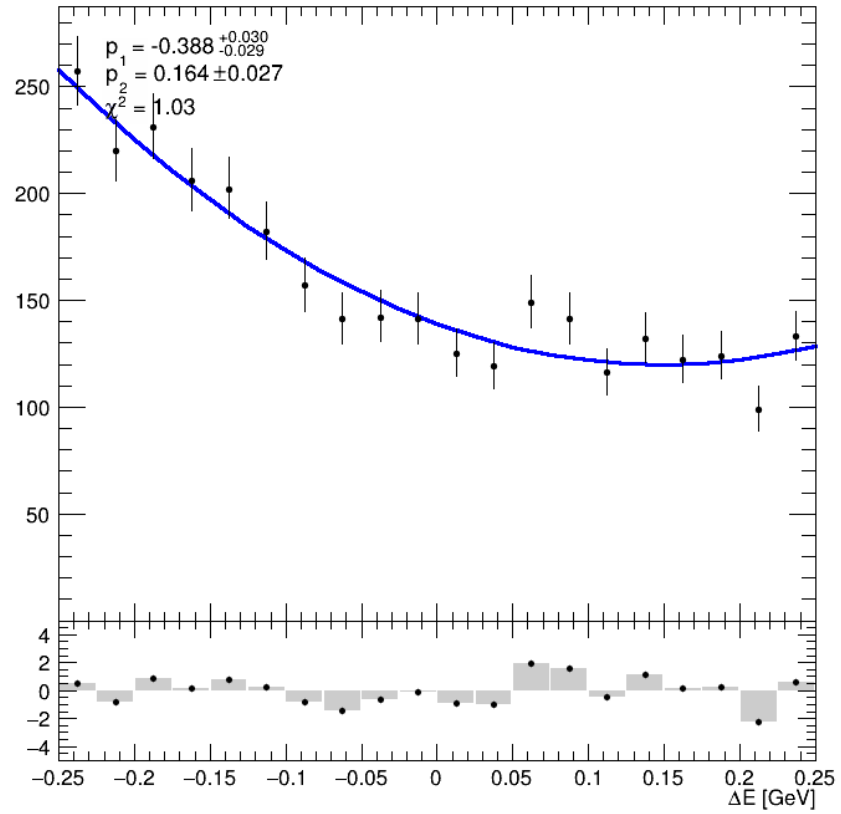
The comparison between constrained and unconstrained fit parameters (numbers of events) was made also with these six test datasets. Similar plots like those on

figures 2.18, 2.19 and 2.20 were made for constrained variables. Since some of them are well in the positive range and didn't change with constraining fitted variables, only comparison between changed fits are shown here. They can be seen in 2.21 for channel 0 (ϕ), 2.22 for channel 1 (K^*) and 2.23 for channel 2 (K_S).

Although plots for unconstrained fits look too much like overfitting, the fitter compensates for this with bigger errors. From the plots with constraints can be seen, that the fitter might have problems converging when trying to fit constrained parameters. Hence either fixing the value of number of mixedrare events to zero or unconstraining the parameter is in order for the fitter to properly converge in these occasions. Hereby we chose the latter. A summary of all six unconstrained test fits can be seen on table 2.12.

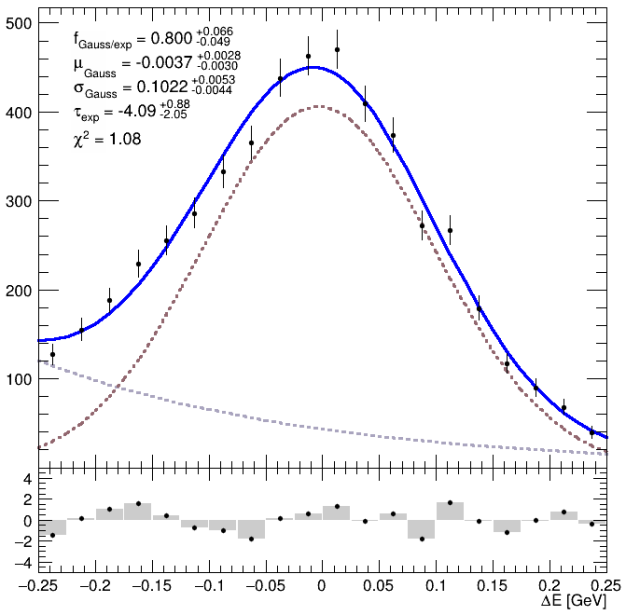


(a) Signal Model

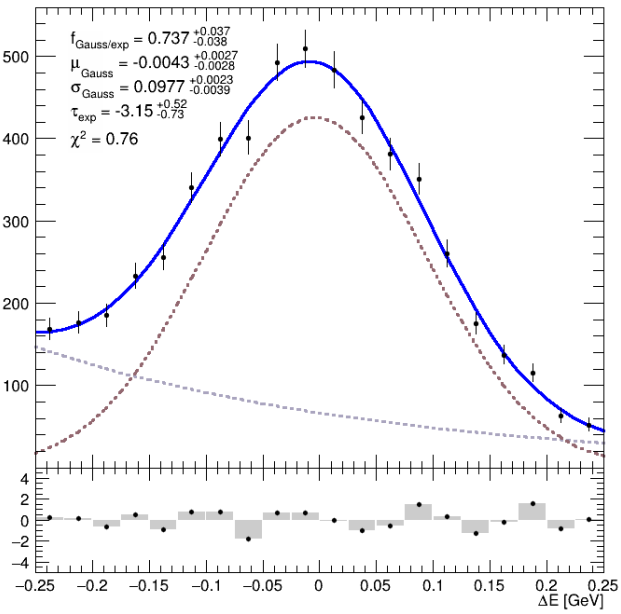


(b) Background model

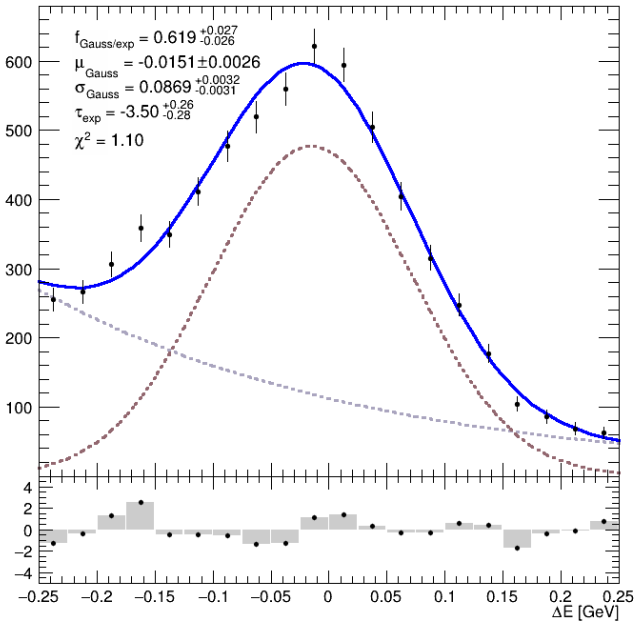
Figure 2.11: Models and their fits to MC data used in analysis: (a) signal model, (b) background model



(c) Channel 0 (ϕ)

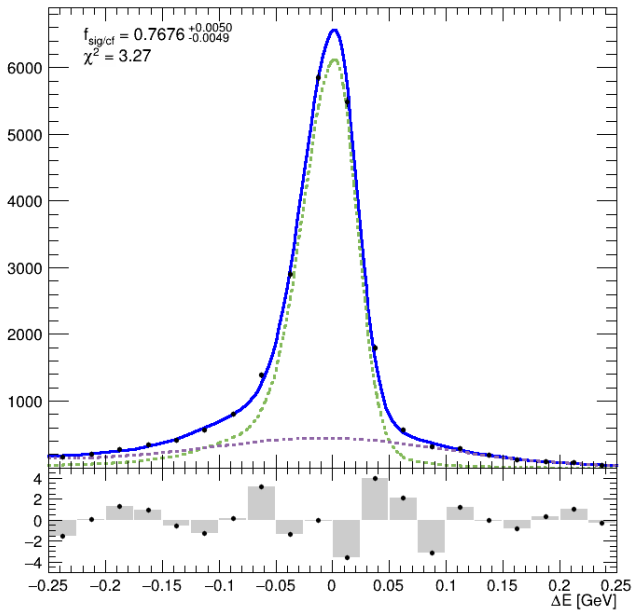


(d) Channel 1 (K^*)

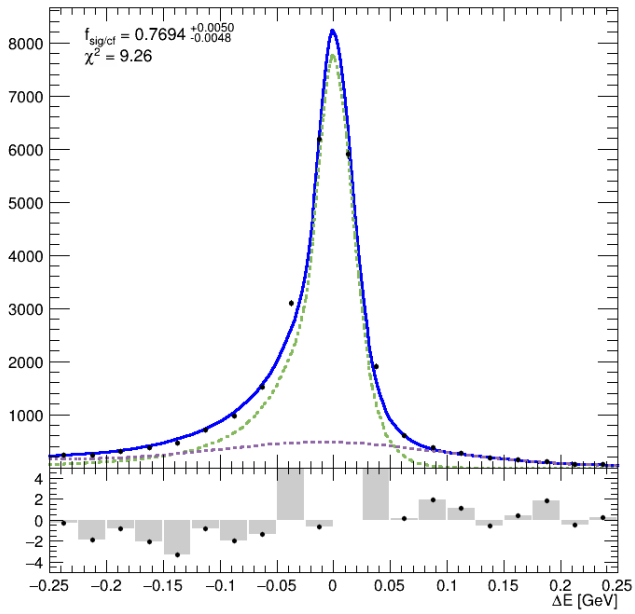


(e) Channel 2 (K_S)

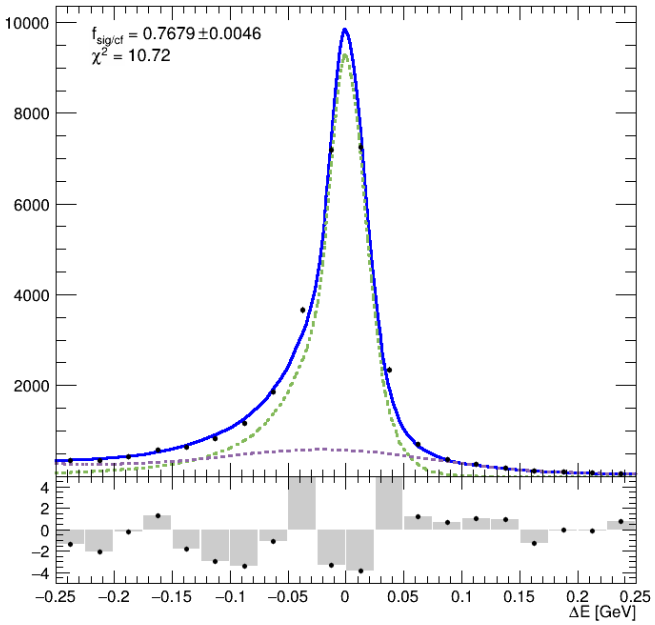
Figure 2.11: Models and their fits to MC data used in analysis: (c)-(e) crossfeed models



(f) Channel 0 (ϕ)

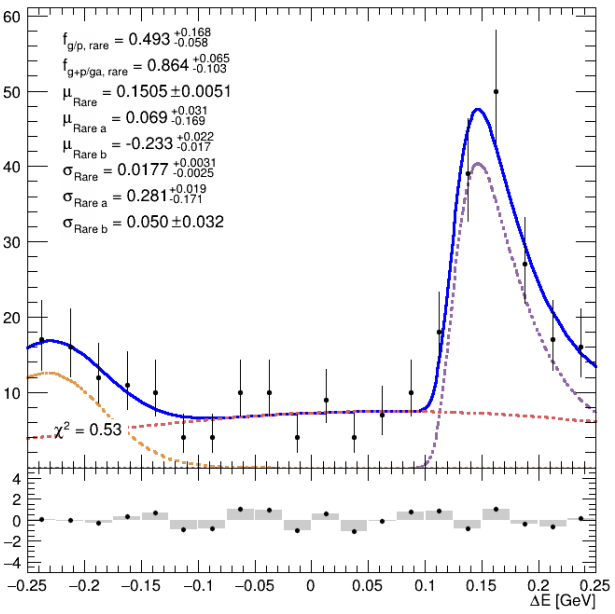


(g) Channel 1 (K^*)

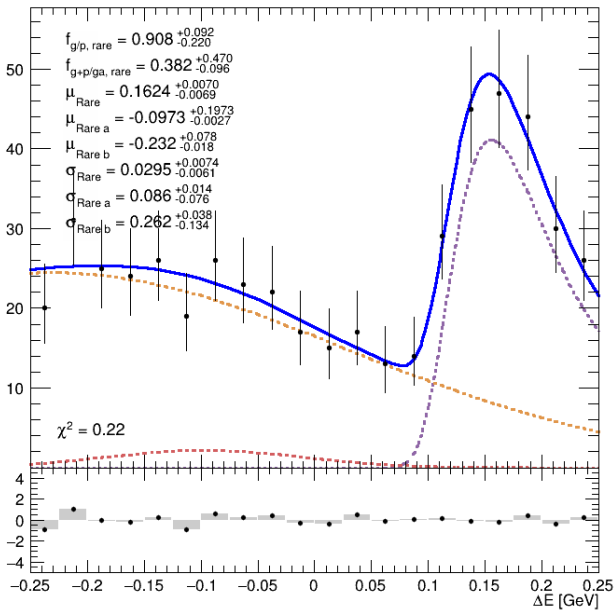


(h) Channel 2 (K_S)

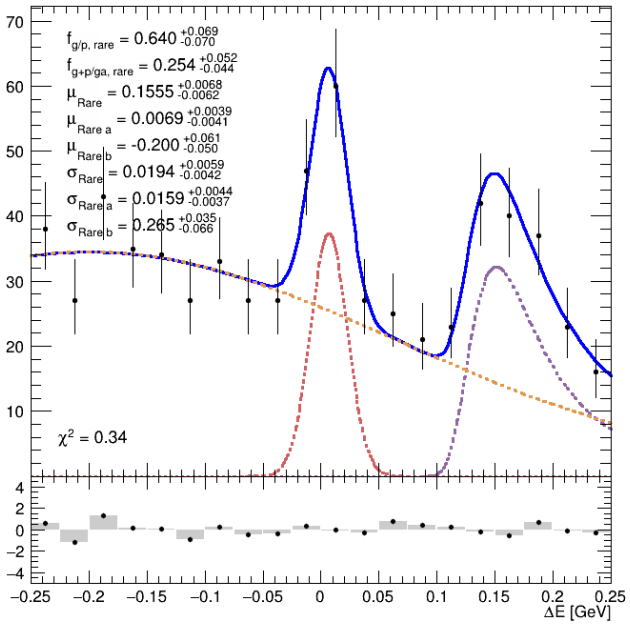
Figure 2.11: Models and their fits to MC data used in analysis: (f)-(h) signal+crossfeed models



(i) Channel 0 (ϕ)

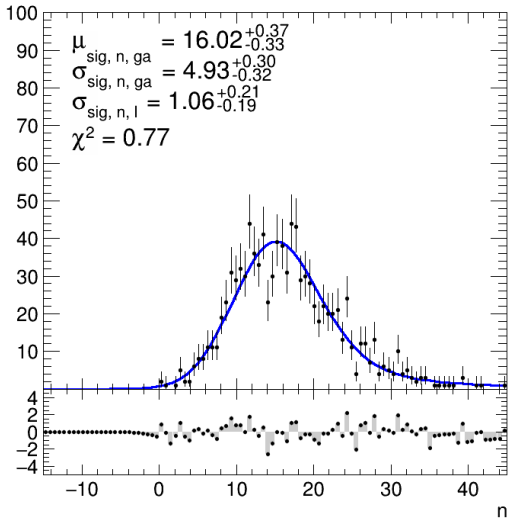


(j) Channel 1 (K^*)

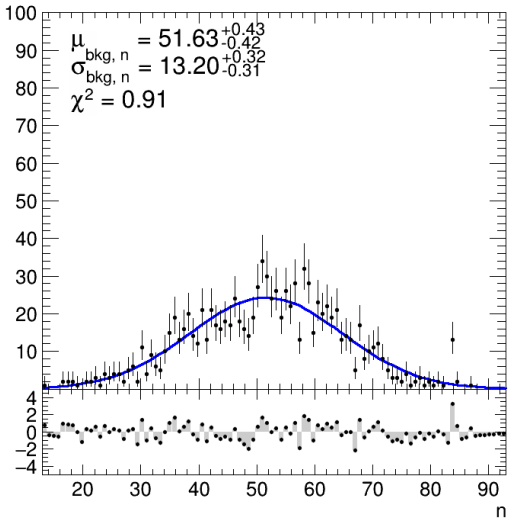


(k) Channel 2 (K_S)

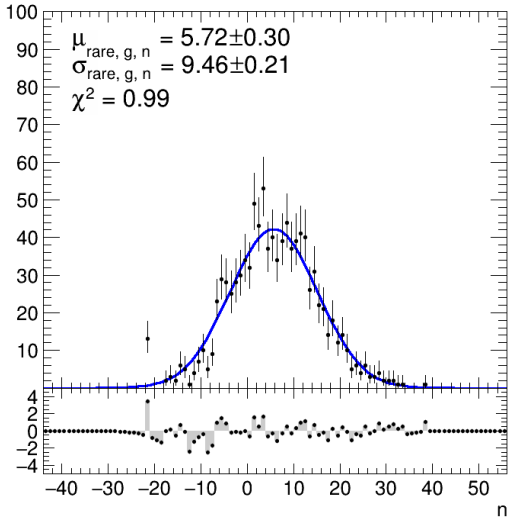
Figure 2.11: Models and their fits to MC data used in analysis: (i)-(k) mixedrare models



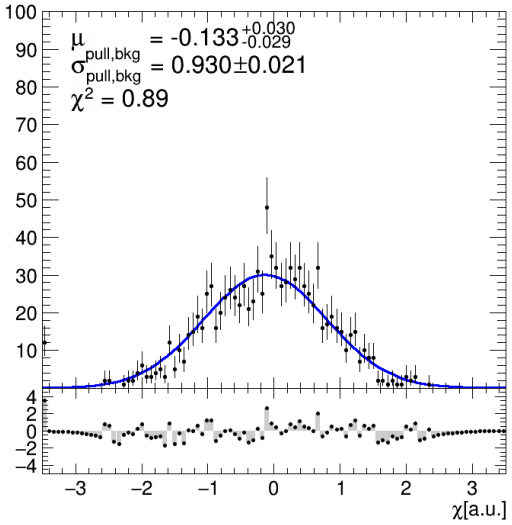
(a) Signal number of events



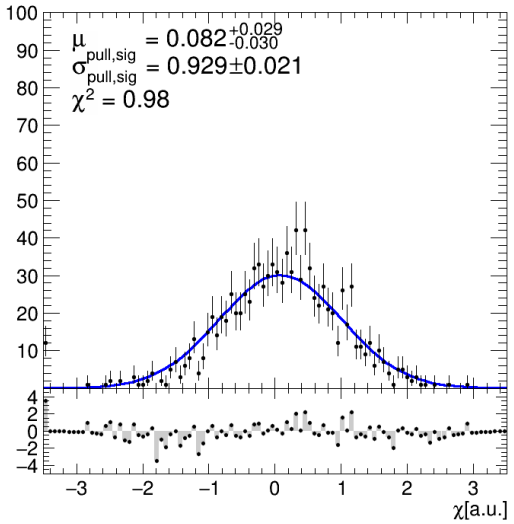
(b) Background number of events



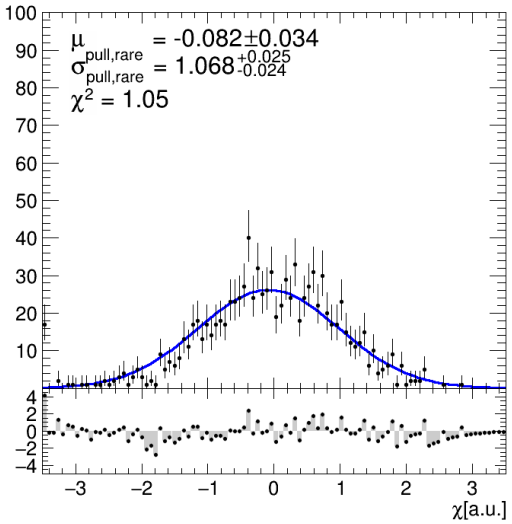
(c) MixedRare number of events



(d) Background pull

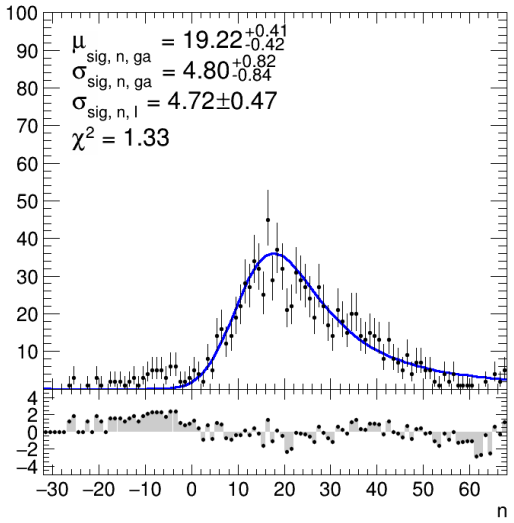


(e) Signal pull

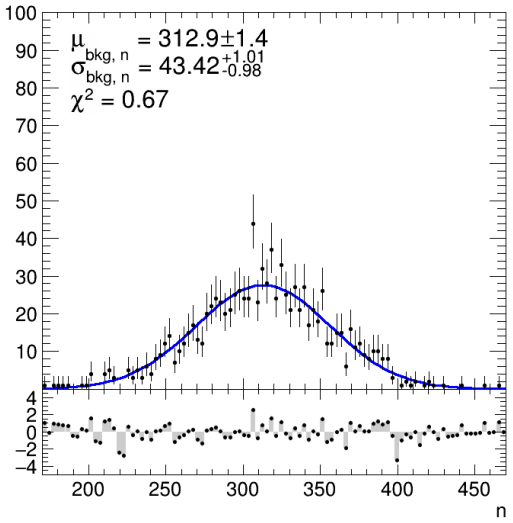


(f) MixedRare pull

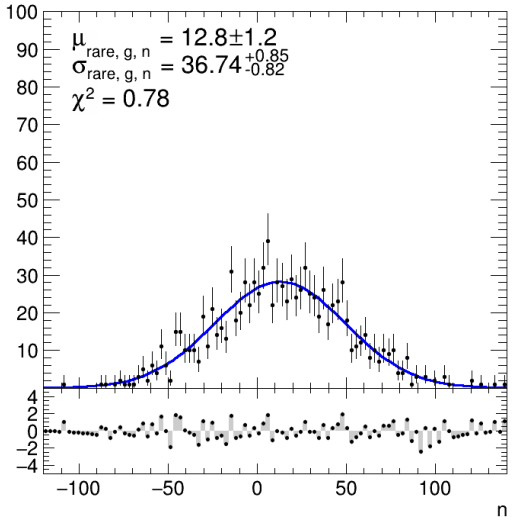
Figure 2.12: ToyMC validation of fitter program for channel 0 (ϕ). (a)-(c) numbers of events fitted (d)-(e) pull plots



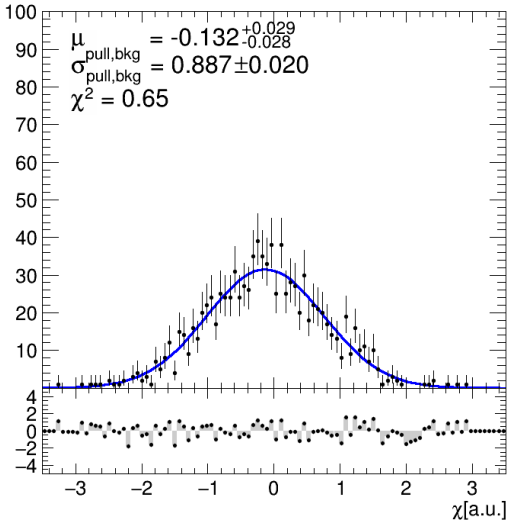
(a) Signal number of events



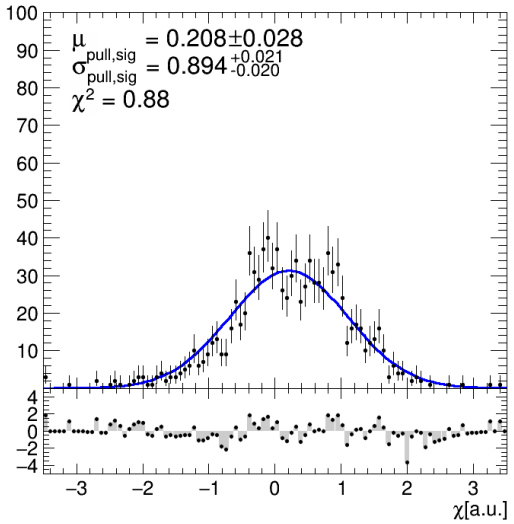
(b) Background number of events



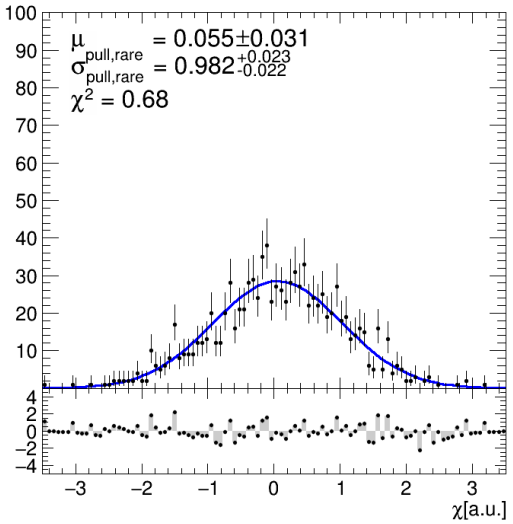
(c) MixedRare number of events



(d) Background pull

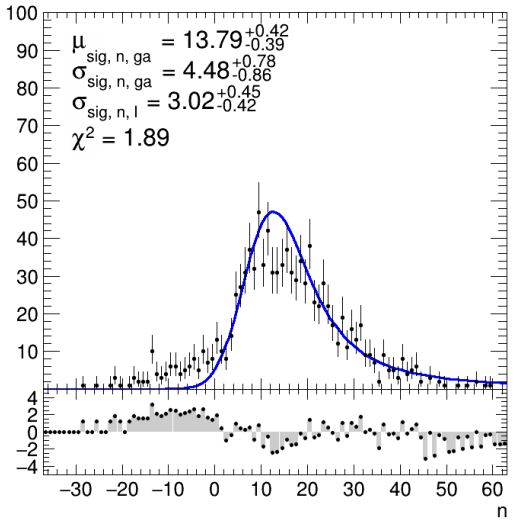


(e) Signal pull

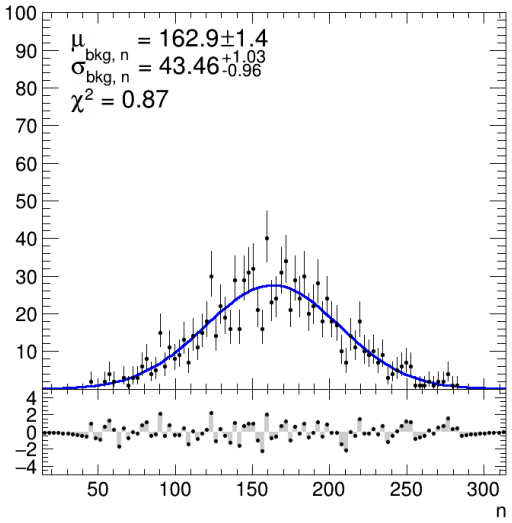


(f) MixedRare pull

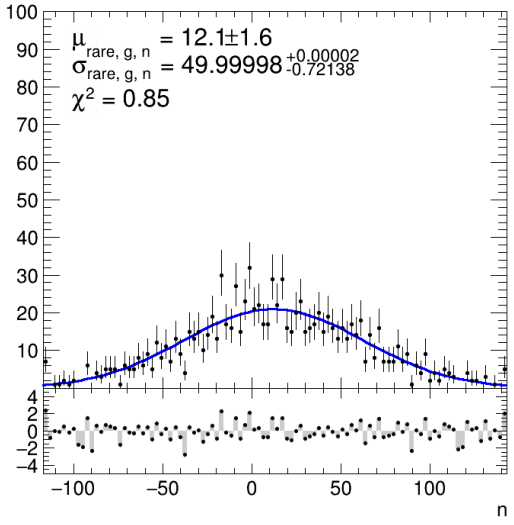
Figure 2.13: ToyMC validation of fitter program for channel 1 (K^*). (a)-(c) numbers of events fitted (d)-(e) pull plots



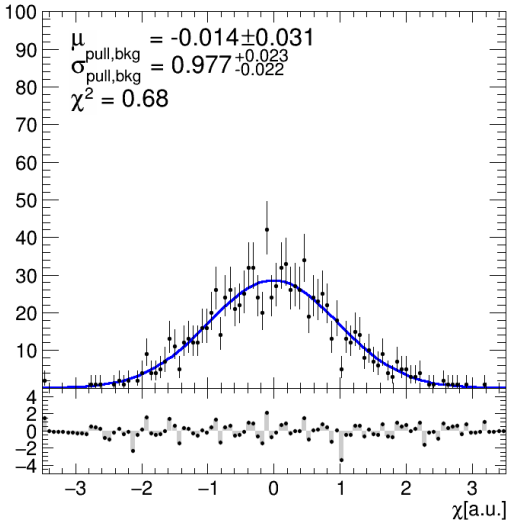
(a) Signal number of events



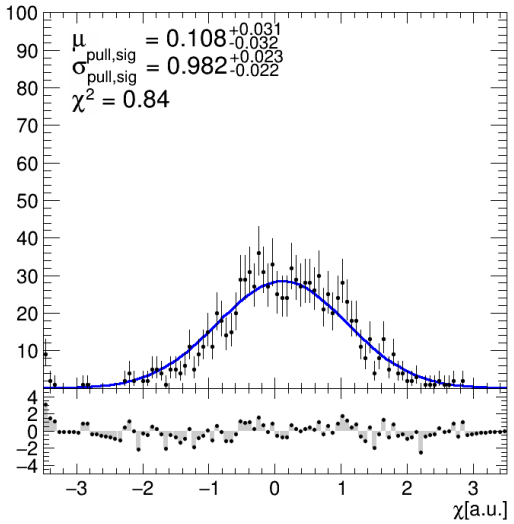
(b) Background number of events



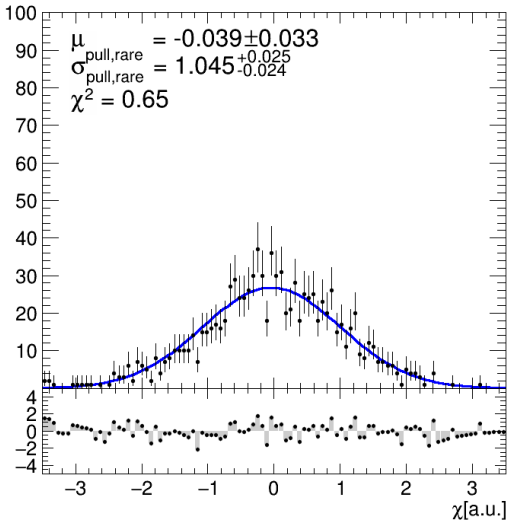
(c) MixedRare number of events



(d) Background pull

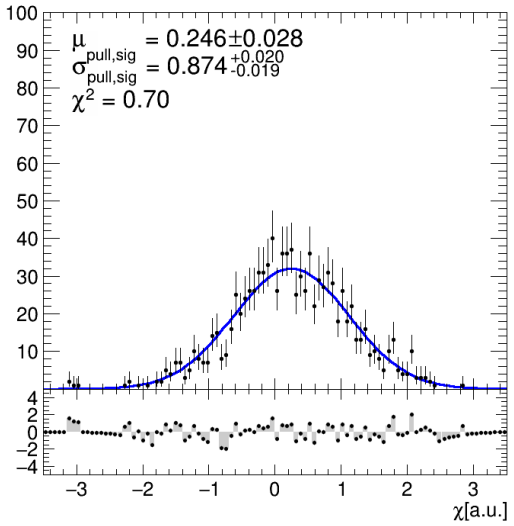


(e) Signal pull

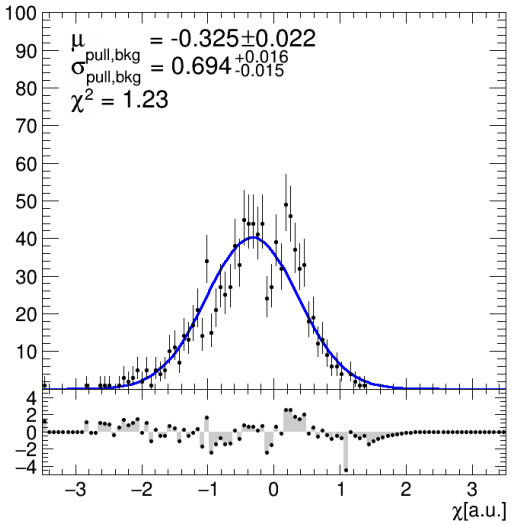


(f) MixedRare pull

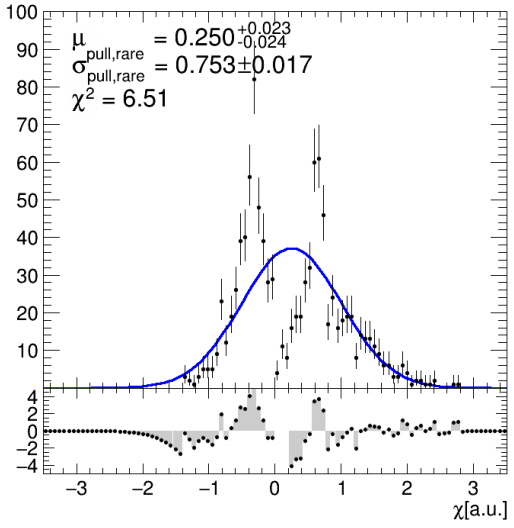
Figure 2.14: ToyMC validation of fitter program for channel 2 (K_S). (a)-(c) numbers of events fitted (d)-(e) pull plots



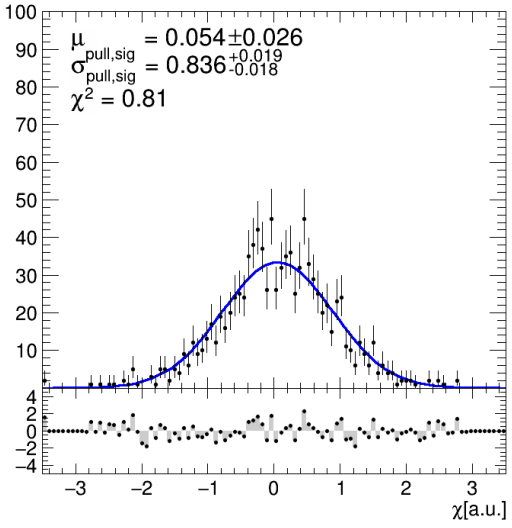
(a) Signal pull



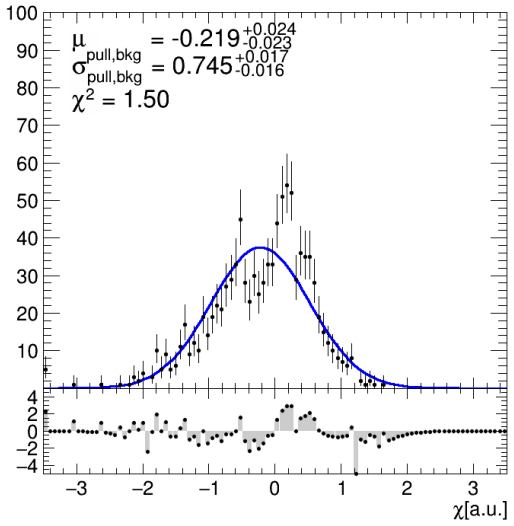
(b) Background pull



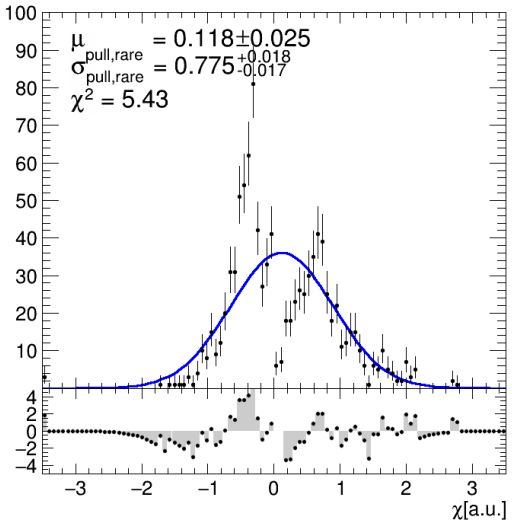
(c) MixedRare pull



(d) Signal pull

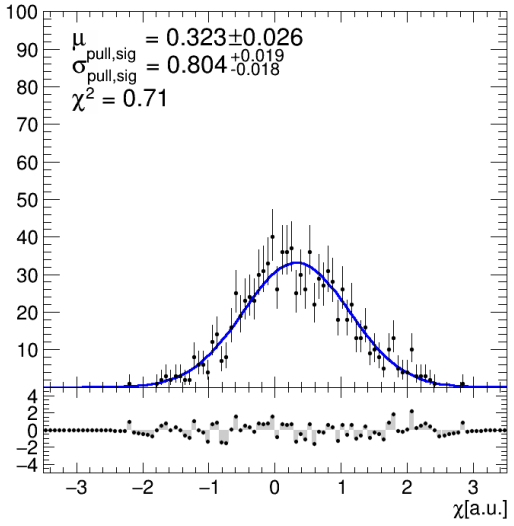


(e) Background pull

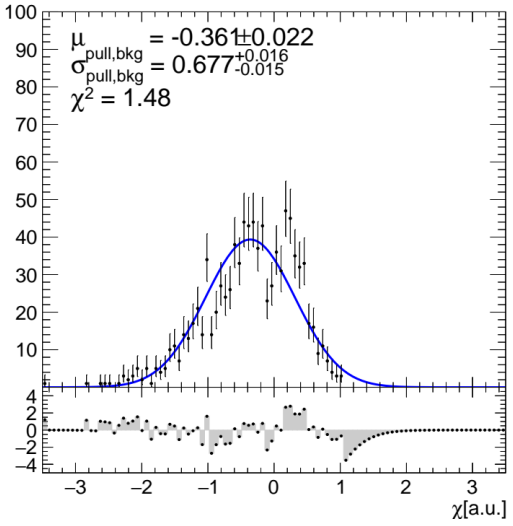


(f) MixedRare pull

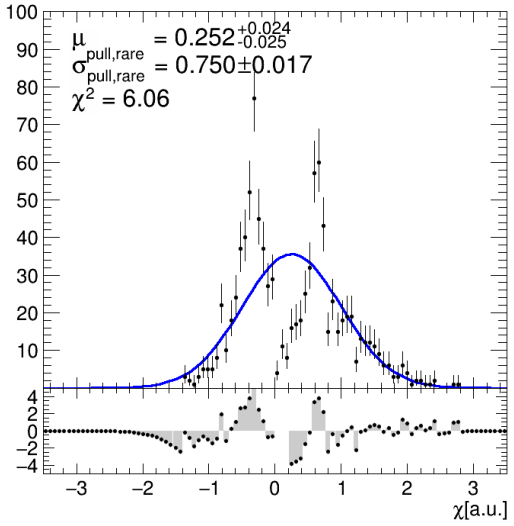
Figure 2.15: ToyMC pull plots of (a)-(c) channel 1 (K^*) and (d)-(f) channel 2 (K_S) in which n_{sig} and n_{rare} were constrained to positive numbers



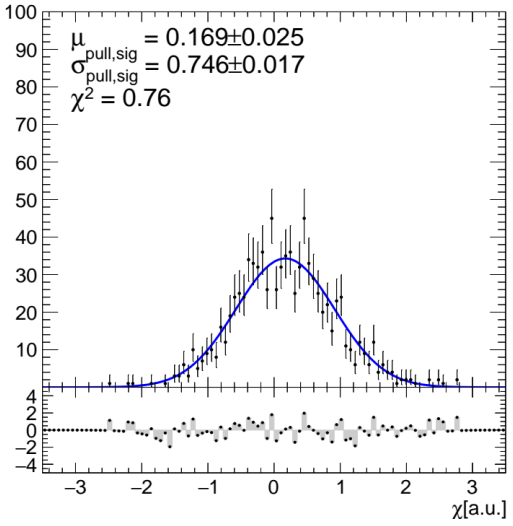
(a) Signal



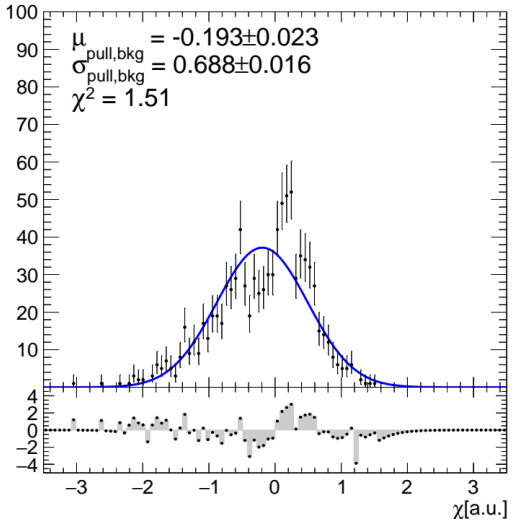
(b) Background



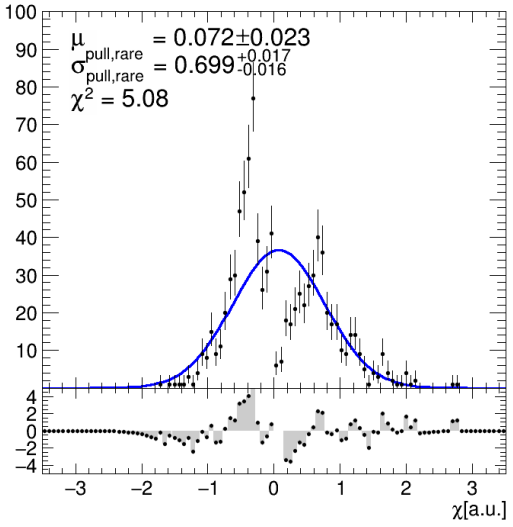
(c) MixedRare



(d) Signal

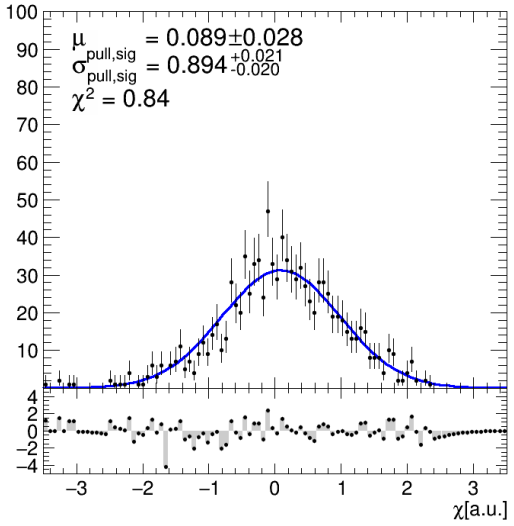


(e) Background

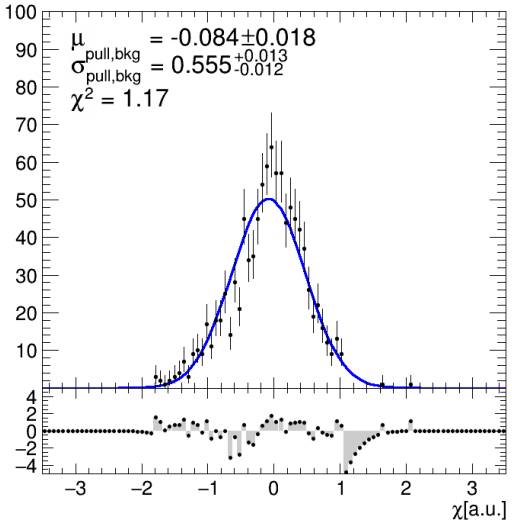


(f) MixedRare

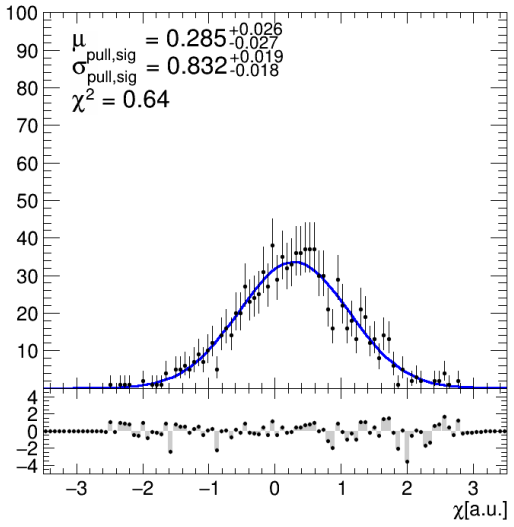
Figure 2.16: ToyMC pull plots of (a)-(c) channel 1 (K^*) and (d)-(f) channel 2 (K_S) in which n_{sig} and n_{rare} were constrained to positive numbers and fits with $n_{\text{sig}} < 1$ cut off



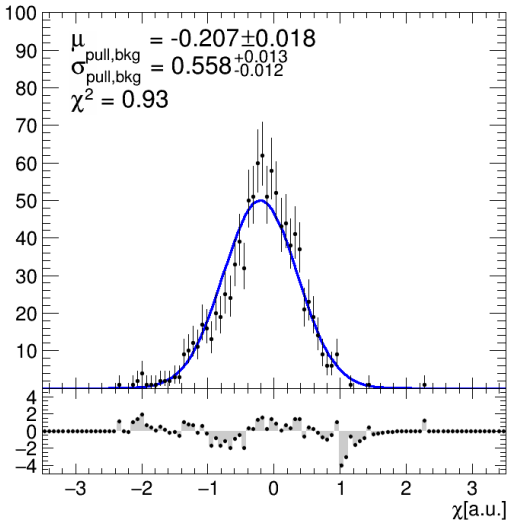
(a) Signal



(b) Background



(c) Signal



(d) Background

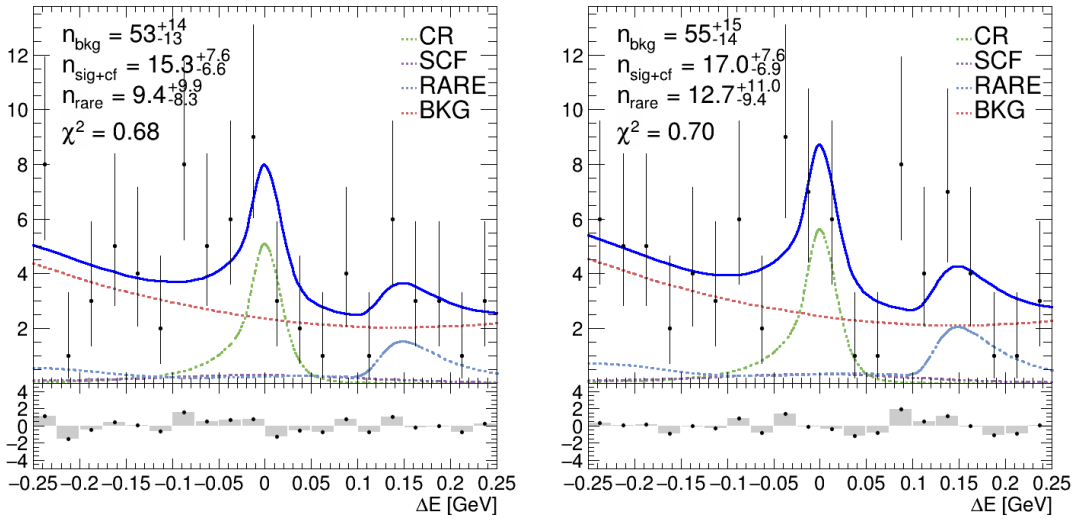
Figure 2.17: ToyMC pull plots of (a)-(b) channel 1 (K^*) and (c)-(d) channel 2 (K_S) without mixedrare component

CHANNEL 0 (ϕ mode)												
	Signal+CF				**Background**				**Mixedrare**			
	Real	Fitted	Error	Pull	Real	Fitted	Error	Pull	Real	Fitted	Error	Pull
	13	15.3	7.6	0.30	56	53	14	-0.21	10	9.4	9.9	-0.06
	22	17.0	7.6	-0.66	56	55	15	-0.07	5	13.7	-9.4	0.93
	12	12.4	-5.9	0.07	53	42	14	-0.79	6	14.6	-9.1	0.95
	10	6.0	5.8	-0.69	49	58	-13	0.69	7	0.6	9.1	-0.70
	9	10.6	-5.9	0.27	55	63	-13	0.62	5	-8.0	7.8	-1.67
	20	9.2	6.8	-1.59	46	65	-13	1.46	2	-4.4	7.3	-0.88
Avg:	14.33	11.75	2.67	-0.38	52.50	56.00	0.67	0.28	5.83	4.32	2.60	-0.24
			BIAS	-2.58			BIAS	3.50			BIAS	-1.52

CHANNEL 1 (K^* mode)												
	Signal+CF				**Background**				**Mixedrare**			
	Real	Fitted	Error	Pull	Real	Fitted	Error	Pull	Real	Fitted	Error	Pull
	16	12	13	-0.31	302	289	44	-0.30	13	24	-35	0.31
	15	14	12	-0.08	331	343	-44	0.27	11	-9	37	-0.54
	15	12	13	-0.23	311	341	-41	0.73	8	-29	33	-1.12
	16	-1	12	-1.42	321	379	-44	1.32	11	-40	36	-1.42
	23	28	-13	0.38	318	262	44	-1.27	6	49	-36	1.19
	16	12	13	-0.31	234	256	-47	0.47	13	83	-40	1.75
Avg:	16.83	12.83	8.33	-0.33	302.83	311.67	-14.67	0.20	10.33	13.00	-0.83	0.03
			BIAS	-4.0			BIAS	8.83			BIAS	2.67

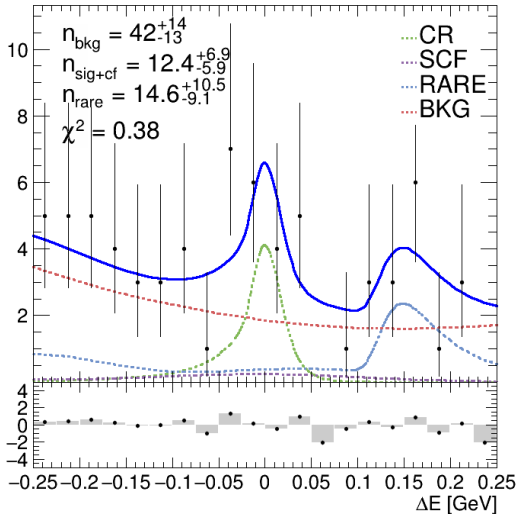
CHANNEL 2 (K_S mode)												
	Signal+CF				**Background**				**Mixedrare**			
	Real	Fitted	Error	Pull	Real	Fitted	Error	Pull	Real	Fitted	Error	Pull
	20	23	-11	0.27	151	202	-41	1.24	18	-41	44	-1.34
	9	14.9	-9.9	0.60	163	218	-38	1.45	12	-50	41	-1.51
	15	8	11	-0.64	150	173	-42	0.55	12	-6	46	-0.39
	16	6	12	-0.83	154	122	43	-0.74	18	56	-46	0.83
	16	25	-12	0.75	197	195	45	-0.04	11	0	51	-0.22
	16	-1.2	9.6	-1.79	169	214	-41	1.10	9	-23	44	-0.73
Avg:	15.33	12.62	-0.05	-0.27	164.00	187.33	-12.33	0.59	13.33	-10.67	30.00	-0.56
			BIAS	-2.72			BIAS	23.33			BIAS	-24.00

Table 2.12: Summary of unconstrained test fits as seen on figures 2.18, 2.19 and 2.20. Only errors relevant for pull calculation are listed here (if the real value is bigger than the fitted value, only positive errors are listed and if the real value is below the fitted value, only negative errors are listed).

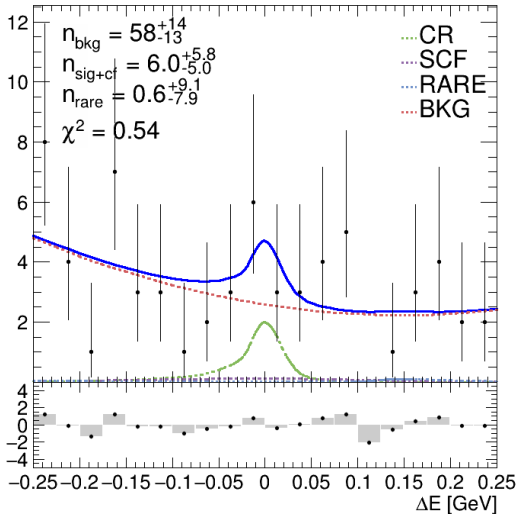


(a)

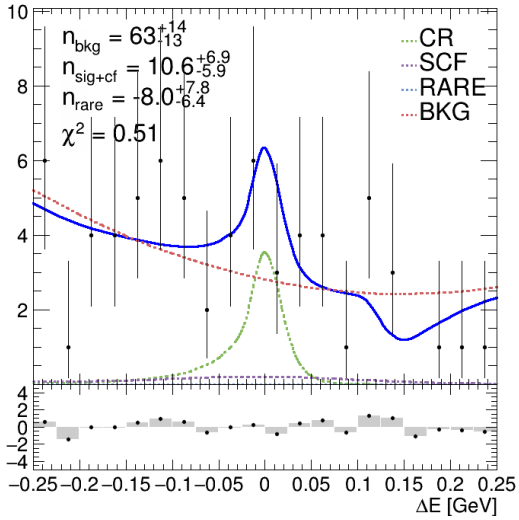
(b)



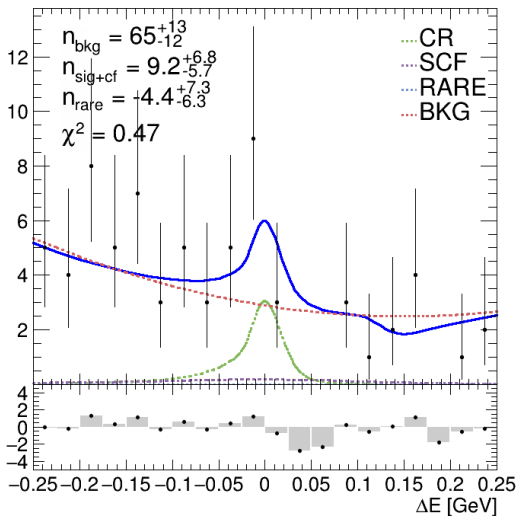
(c)



(d)

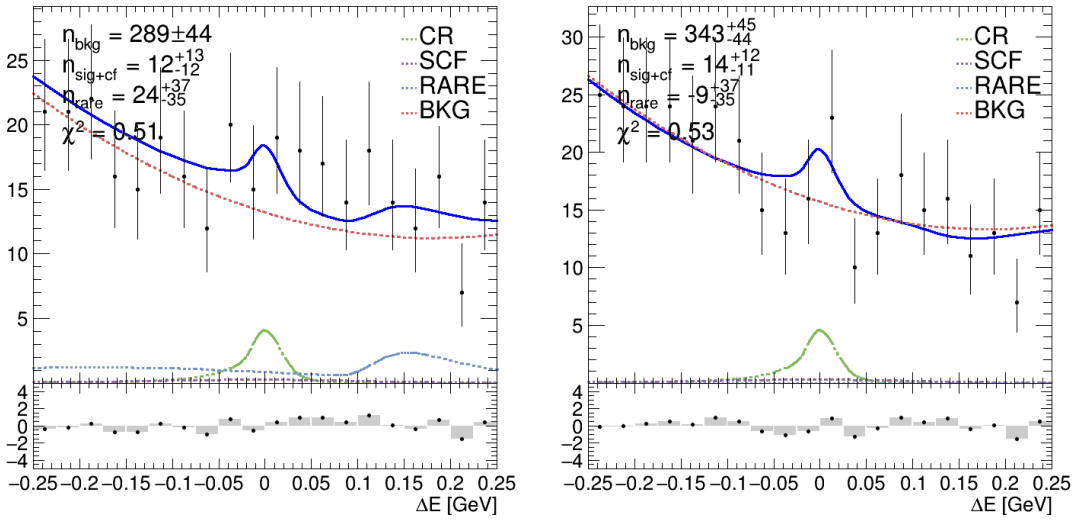


(e)



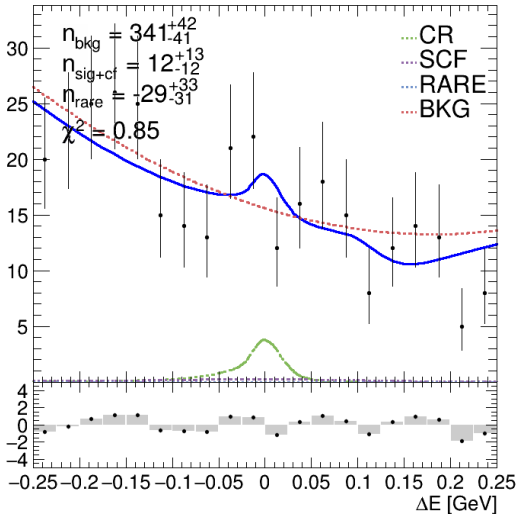
(f)

Figure 2.18: Fits of the whole model to six test MC datasets without constraints for fitted variables for channel 0 (ϕ)

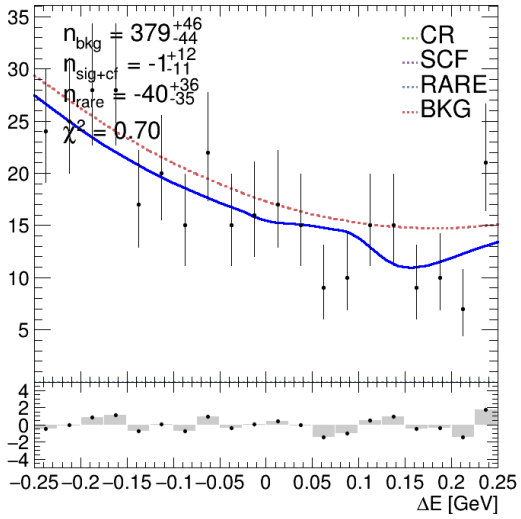


(a)

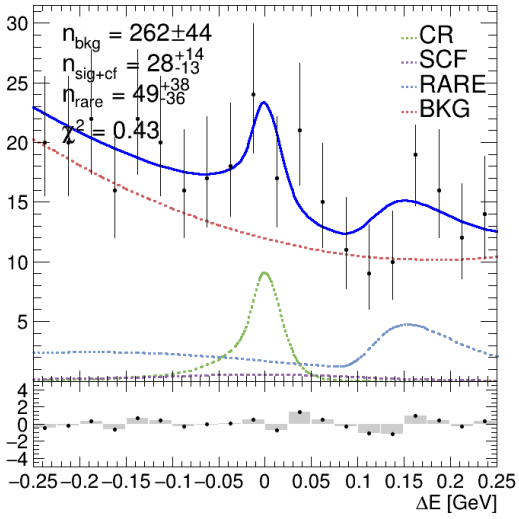
(b)



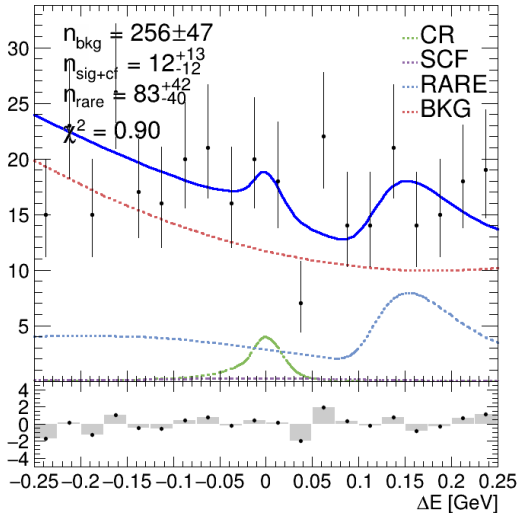
(c)



(d)

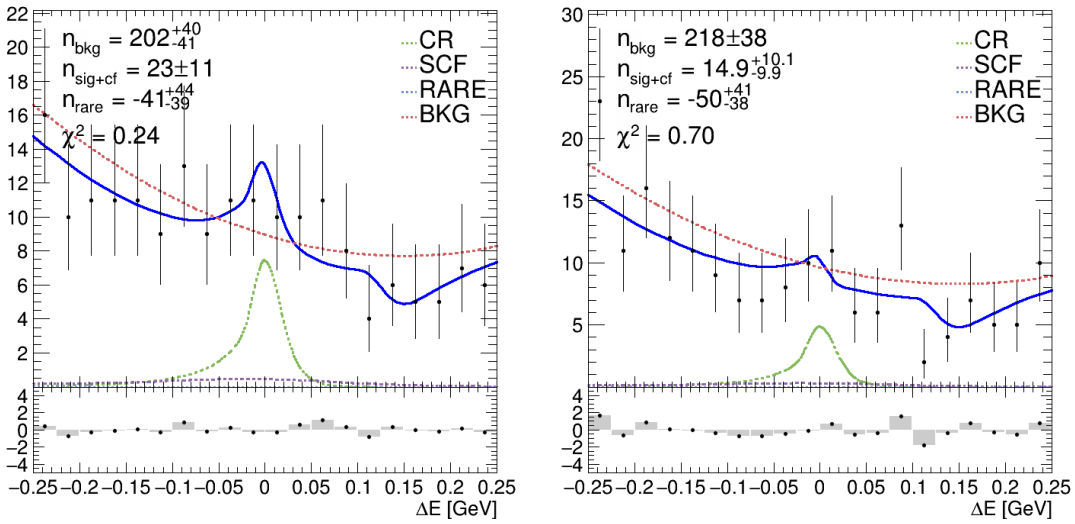


(e)



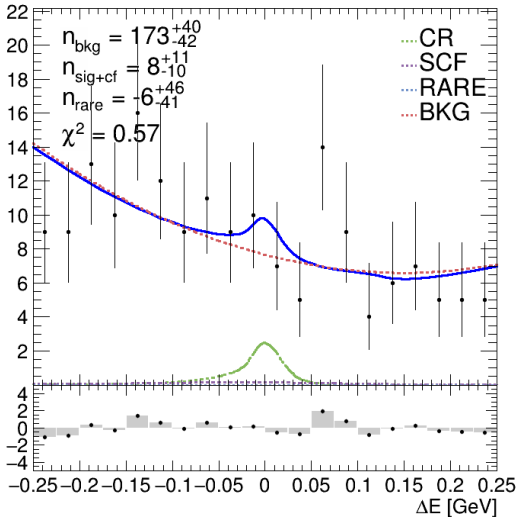
(f)

Figure 2.19: Fits of the whole model to six test MC datasets without constraints for fitted variables for channel 1 (K^*)

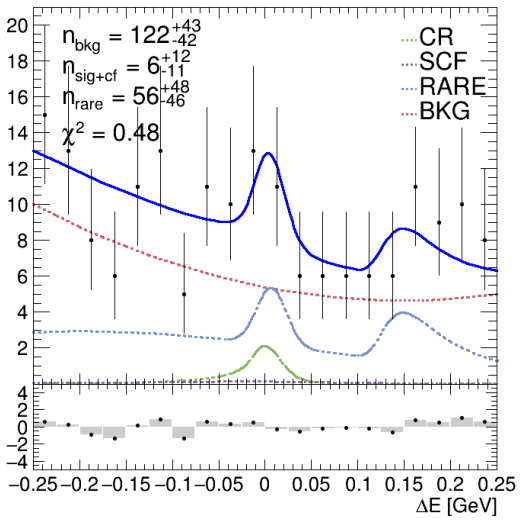


(a)

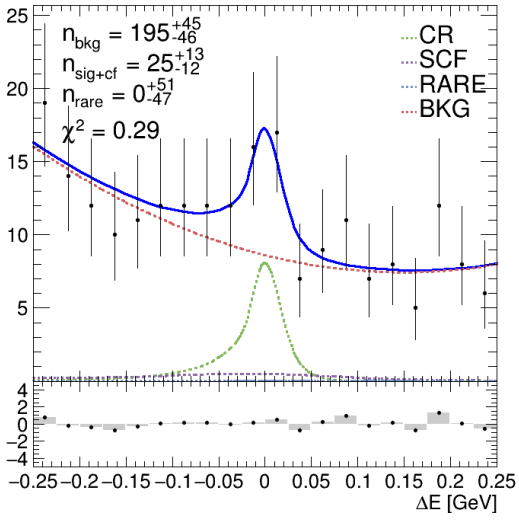
(b)



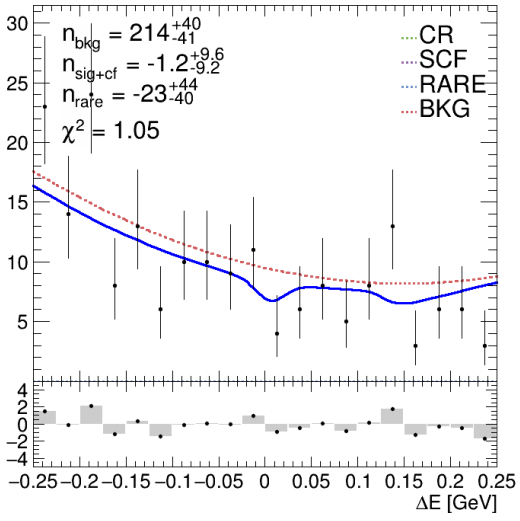
(c)



(d)

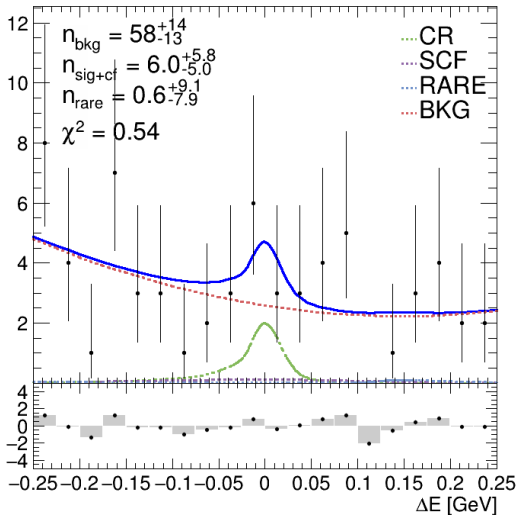


(e)

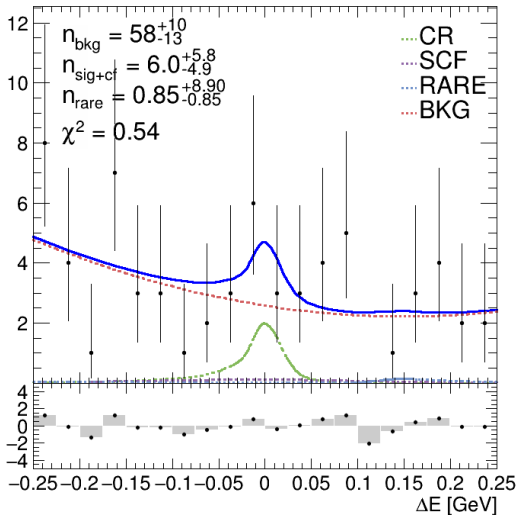


(f)

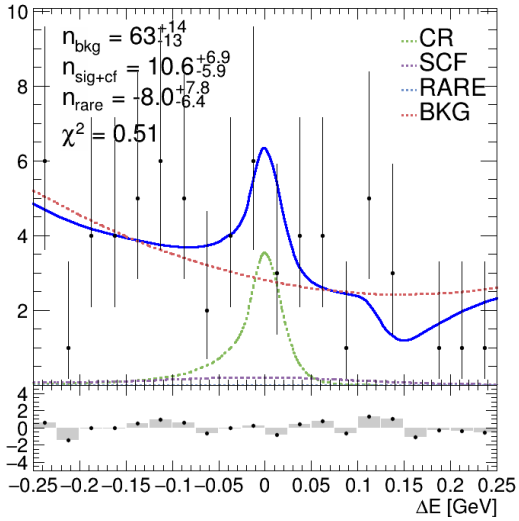
Figure 2.20: Fits of the whole model to six test MC datasets without constraints for fitted variables for channel 2 (K_S)



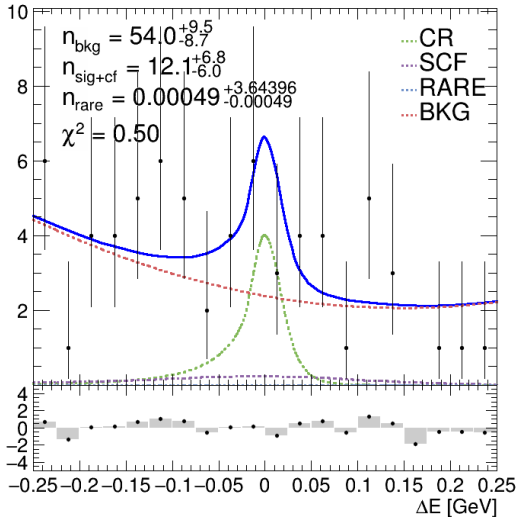
(a) Without constraints



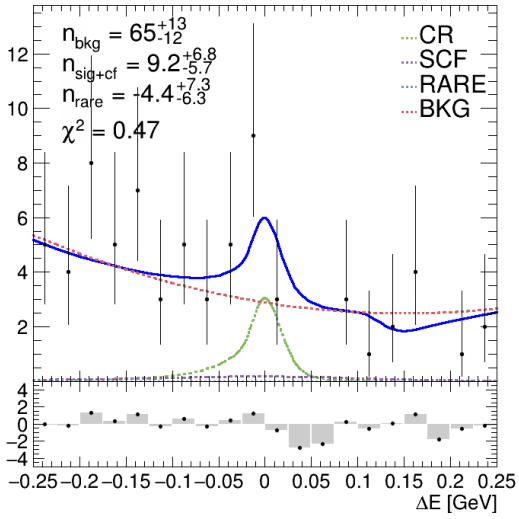
(b) With constraints



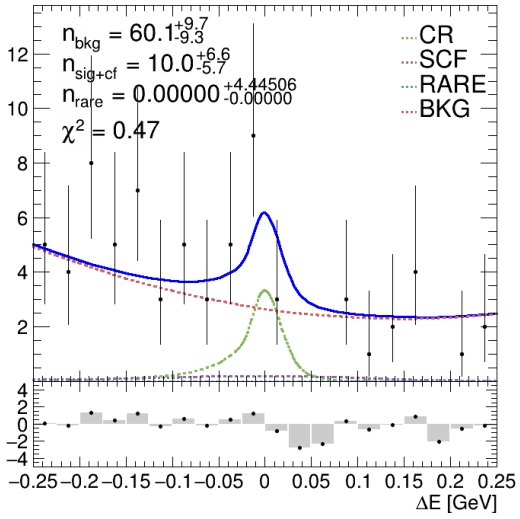
(c) Without constraints



(d) With constraints

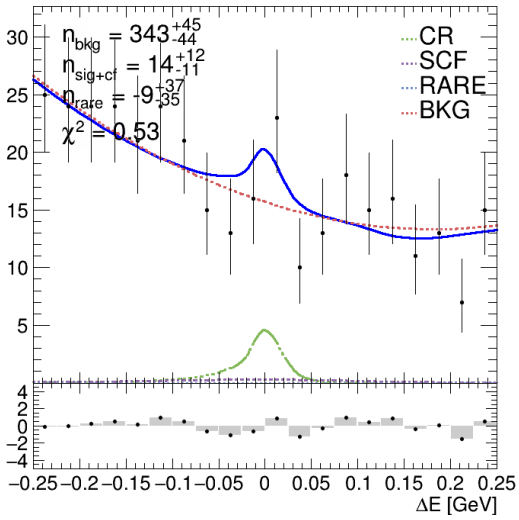


(e) Without constraints

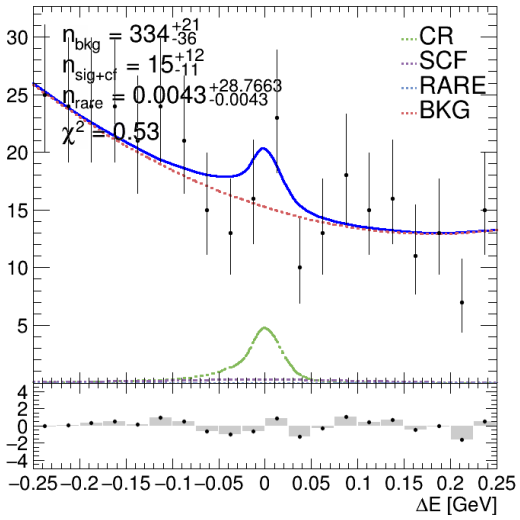


(f) With constraints

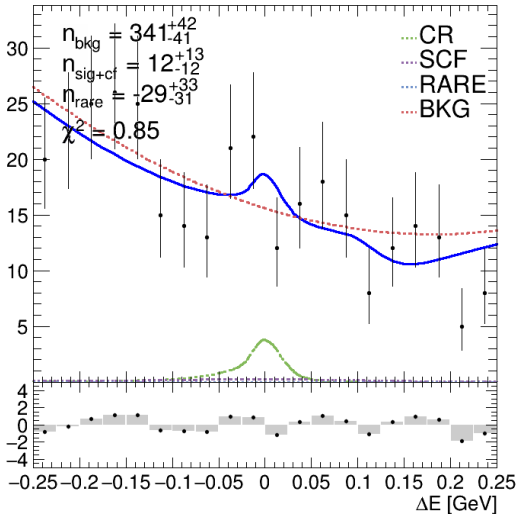
Figure 2.21: Comparisons of fits of the whole model to six test MC datasets between variables with and without constraints for channel 0 (ϕ)



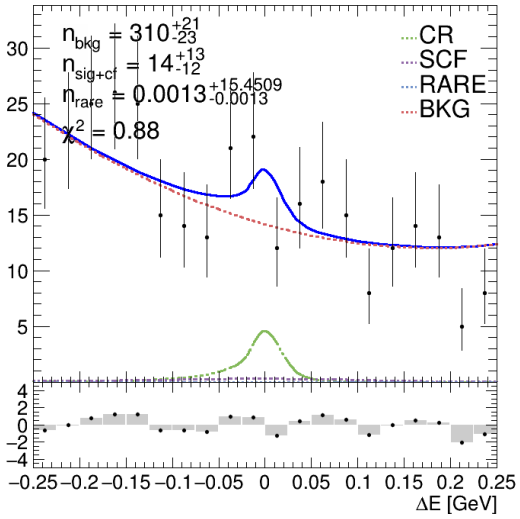
(a) Without constraints



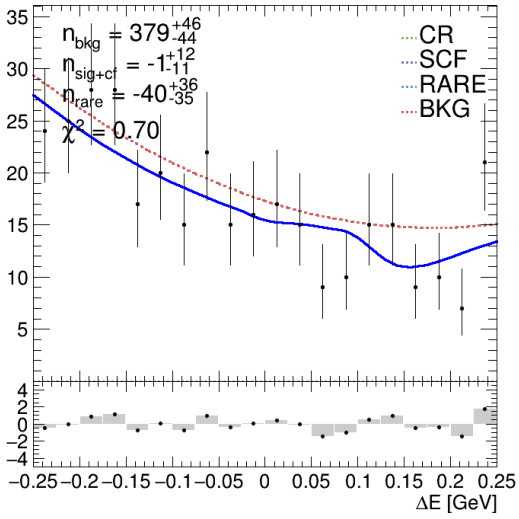
(b) With constraints



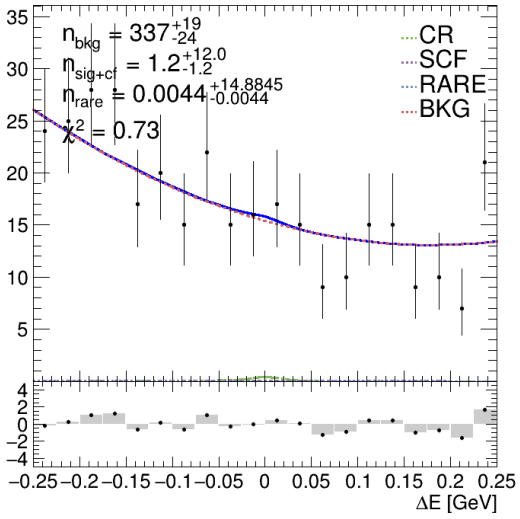
(c) Without constraints



(d) With constraints

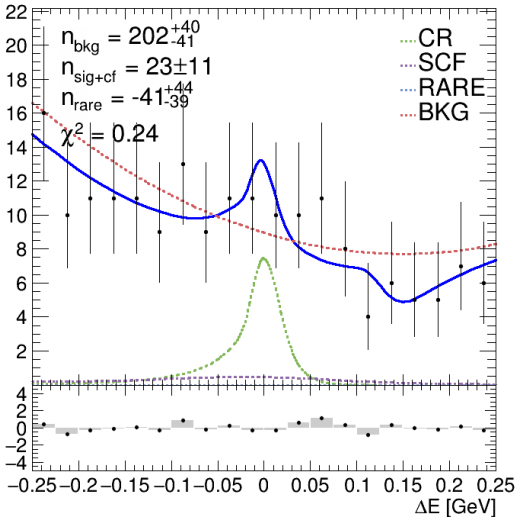


(e) Without constraints

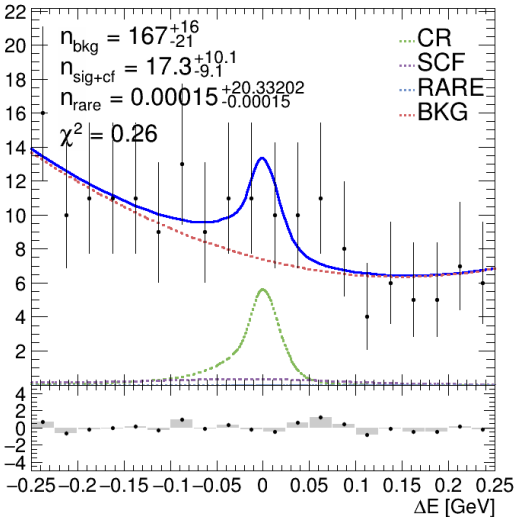


(f) With constraints

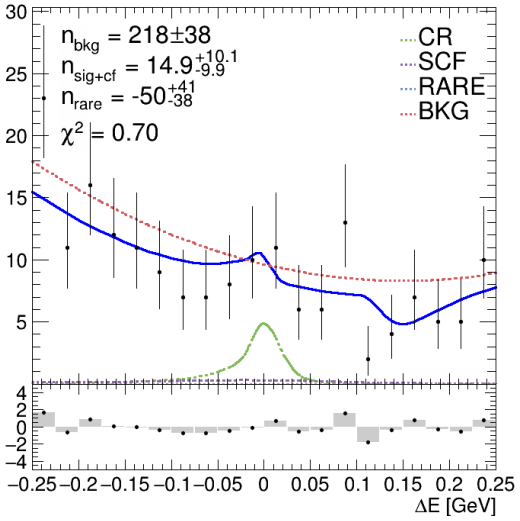
Figure 2.22: Comparisons of fits of the whole model to six test MC datasets between variables with and without constraints for channel 1 (K^*)



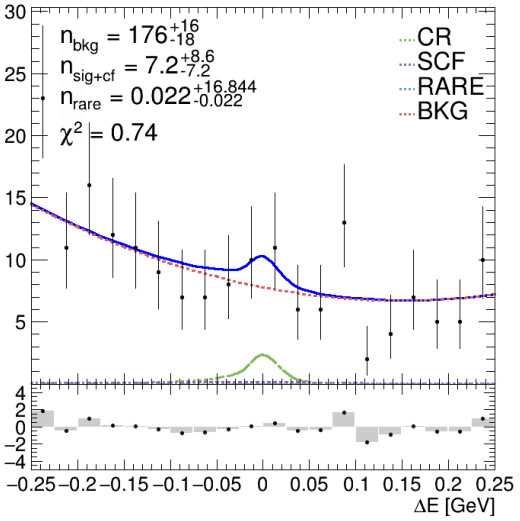
(a) Without constraints



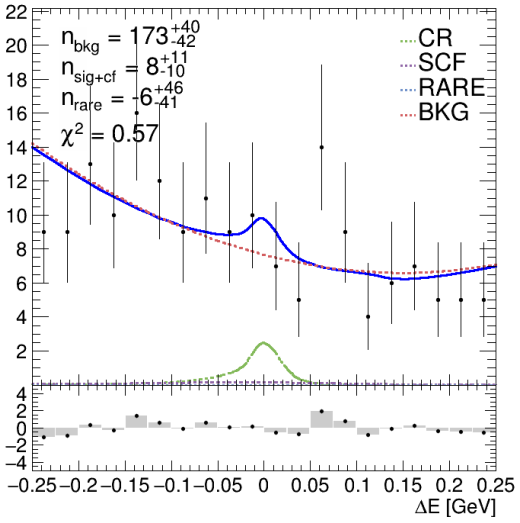
(b) With constraints



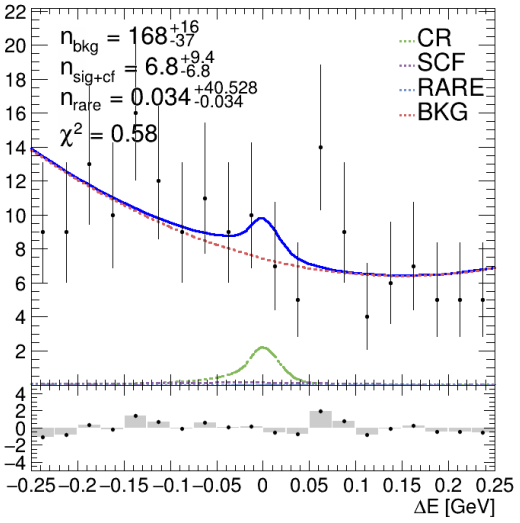
(c) Without constraints



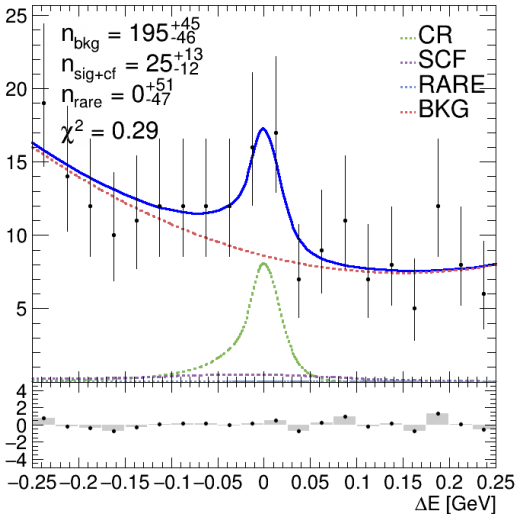
(d) With constraints



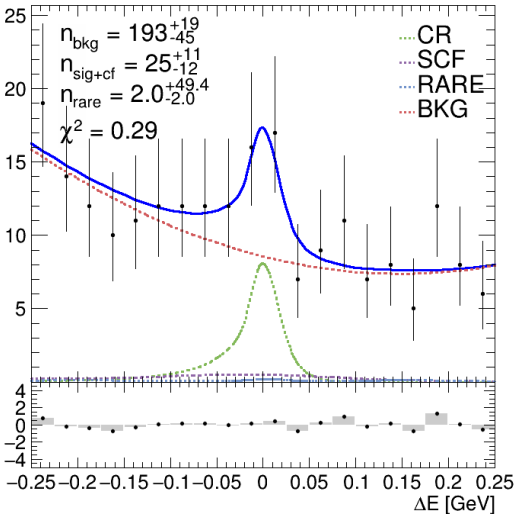
(e) Without constraints



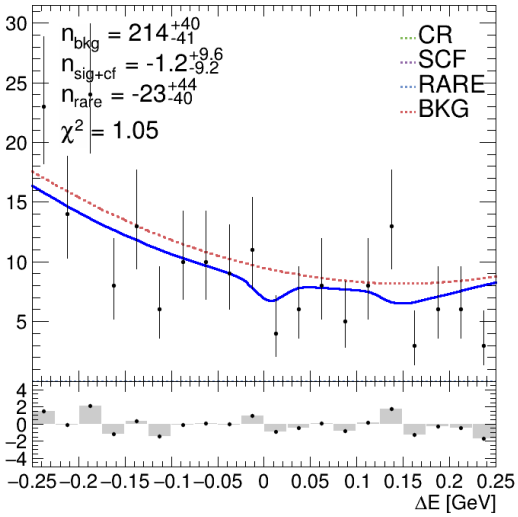
(f) Without constraints



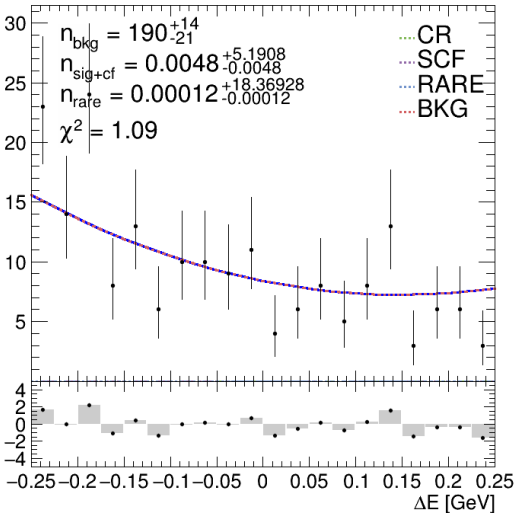
(g) Without constraints



(h) With constraints



(i) Without constraints



(j) With constraints

Figure 2.23: Comparisons of fits of the whole model to six test MC datasets between variables with and without constraints for channel 2 (K_S)

Variable	Cut low [GeV]	Cut high [GeV]	Difference
$E_\gamma(D_s^*)$	Ch0: 0.149 Ch1: 0.182 Ch2: 0.177	-	Used helicity angle between γ and D_s cut
$\Delta M_{D_s^*-D_s}$	0.130	0.158	Minor
M_{bc}	0.27	-	None
M_ϕ	1.0074	1.0314	None
M_{K*}	0.818	0.965	None
M_{K_S}	0.488	0.508	None
M_{D_s}	1.954	1.982	Minor
BDTG	Ch0: 0.6 Ch1: 0.78 Ch2: 0.48	-	Used Fisher discriminant

Table 2.13: Cuts used for control channel analysis and explanation of their difference to [25, 36]. Minor difference means difference of a few MeV, in the range of statistical error. Ch0 is channel with ϕ decaying into FSP, Ch1 channel with K^* and Ch2 with K_S.

2.6 Validation

2.6.1 Control Channel Validation

To facilitate the correctness of the analysis software, a control channel method was adopted. In this method, we looked at already analyzed channel and compared our results with the previous, published results. For our purpose, a decay channel $B^0 \to D_s^{*\mp} + \pi^\pm$ was chosen thanks to its similarity with our decay channel. This channel is not only similar in its decay products, but also because of its rarity. Its branching fraction reaching $(2.1 \pm 0.4) \times 10^{-5}$ [45]. We analyzed 3 sub-channels with identical D_s^* decay products as those of our channel.

The previous analysis of this channel can be found in [25, 36]. In our analysis we tried to replicate this previous one as well as possible. The cuts chosen and differences between our analysis and analysis in [25, 36] can be seen in table 2.13. One major difference between our analysis and analysis in [25, 36] is, that we used asymmetrical PID cuts with $\mathrm{PID}(\pi/K) = 1 - \mathrm{PID}(K/\pi) = 0.6$ and they used symmetrical PID cuts with $\mathrm{PID}(K/\pi) = \mathrm{PID}(\pi/K)$ for every $K^\pm$ not originating from ϕ and not being a sibling of K^*. For those originating from ϕ, they used $\mathrm{PID}(K/\pi) > 0.1$. For siblings of K^*, they used the cut $\mathrm{PID}(K/\pi) > 0.8$.

The results of fitting (models are the same as in table 2.9) PDF fit can be seen

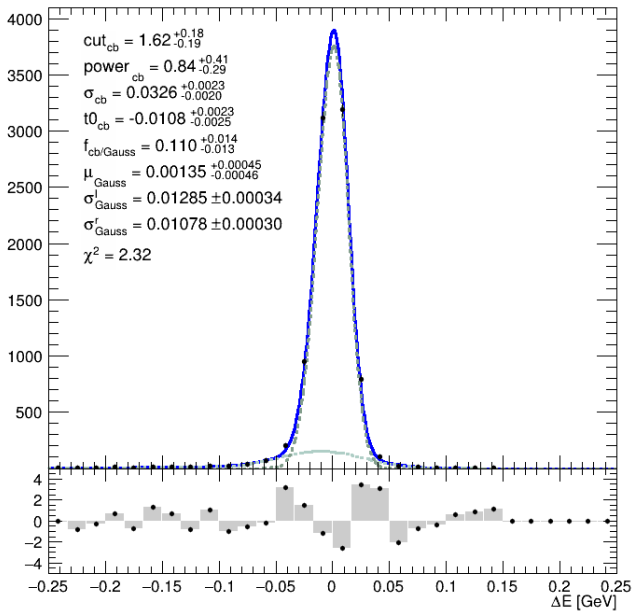
	Ch0 (ϕ)	**Ch1** (K^*)	**Ch2** (K_S)
Previous analysis	32.3±8.1	29.2±9.6	13.1±6.8
Our analysis	$32.0^{+7.0}_{-6.4}$	$23.0^{+8.0}_{-7.1}$	$17.2^{+6.7}_{-5.7}$

Table 2.14: Signal yields comparison between our analysis and analysis in [25, 36]

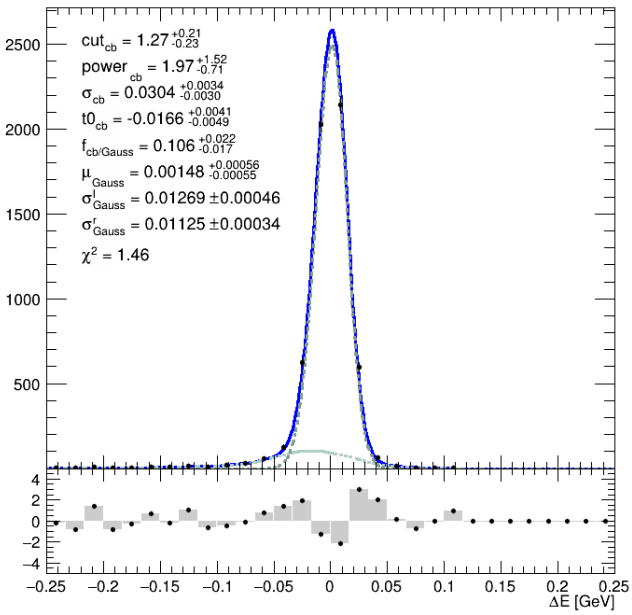
.

in figure 2.24. The background model for this analysis was set the same for every channel also. We also used smaller dataset for signal fit, instead of one million events per channel, we generated only one hundred thousand events per channel.

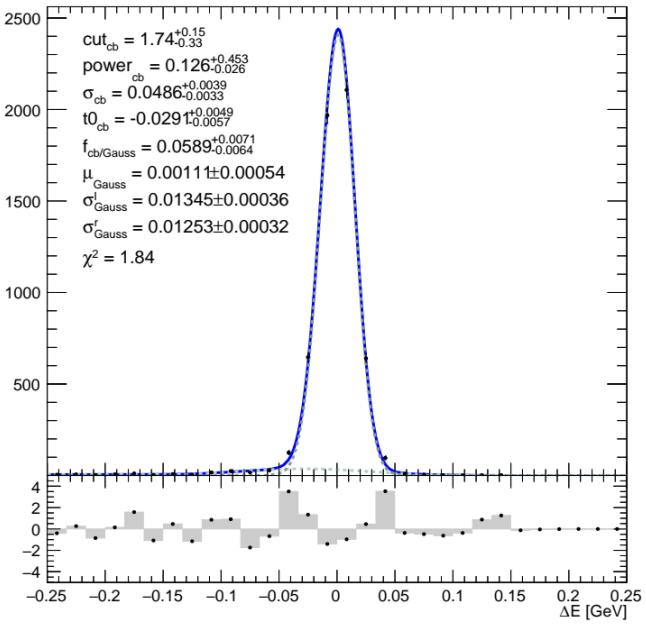
To account for the difference between simulation and experimental data, a fudge factor widening the signal distribution was used. The value of this fudge factor was taken from [25, 36], explicitly $F = 1.08$. The resulting yields can be seen in figure 2.25 and signal yields in table 2.14, with a comparison between our analysis and [25, 36] as well.



(a) Channel 0 (ϕ)

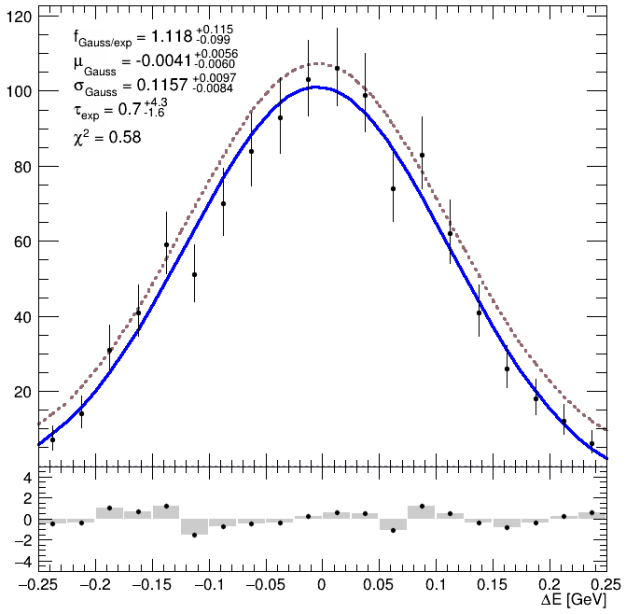


(b) Channel 1 (K^*)

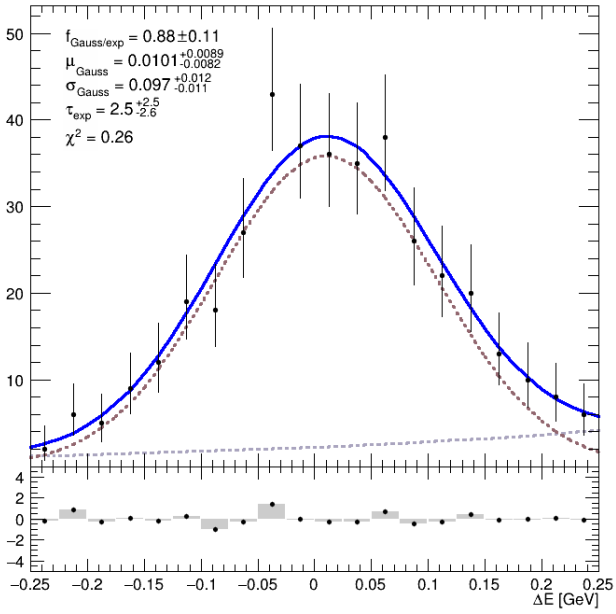


(c) Channel 2 (K_S)

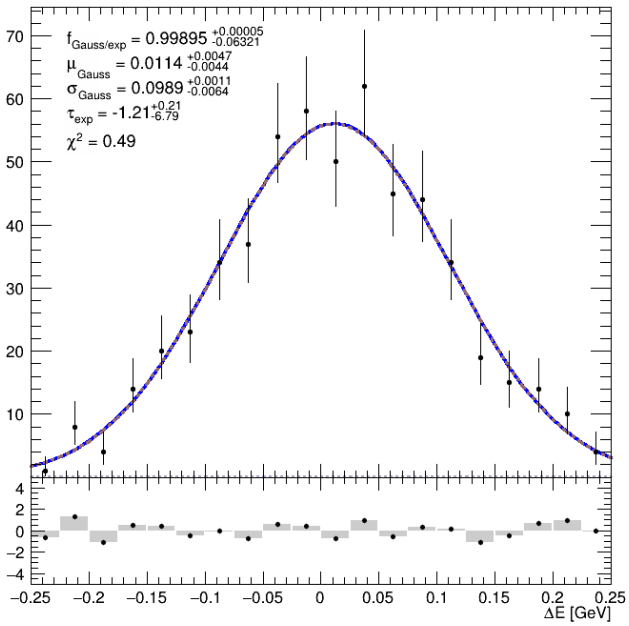
Figure 2.24: Models and their fits to MC data used in control channel analysis: (a)-(c) signal models



(d) Channel 0 (ϕ)

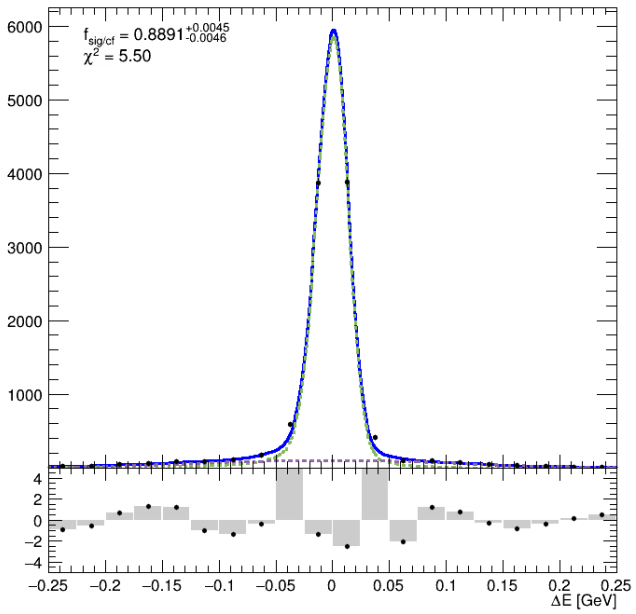


(e) Channel 1 (K^*)

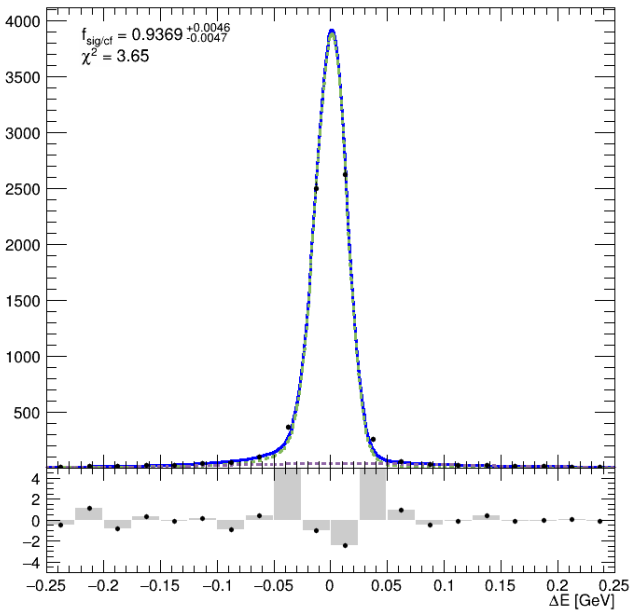


(f) Channel 2 (K_S)

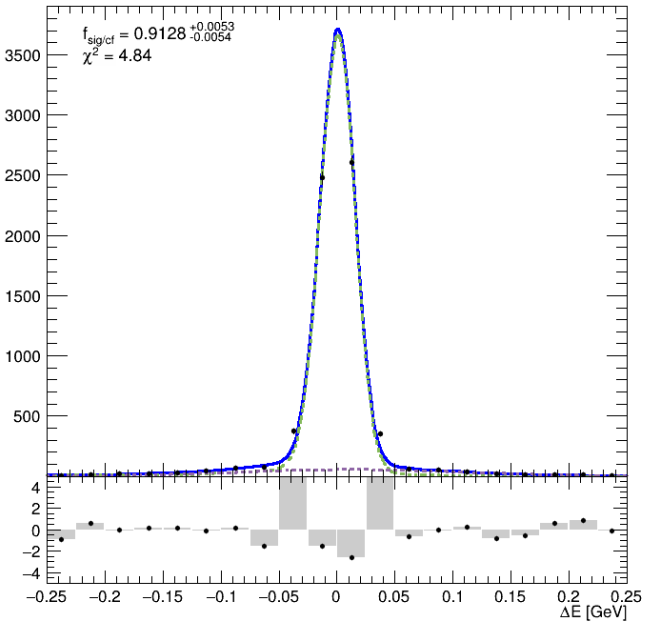
Figure 2.24: Models and their fits to MC data used in control channel analysis:(d)-(f) crossfeed models



(g) Channel 0 (ϕ)

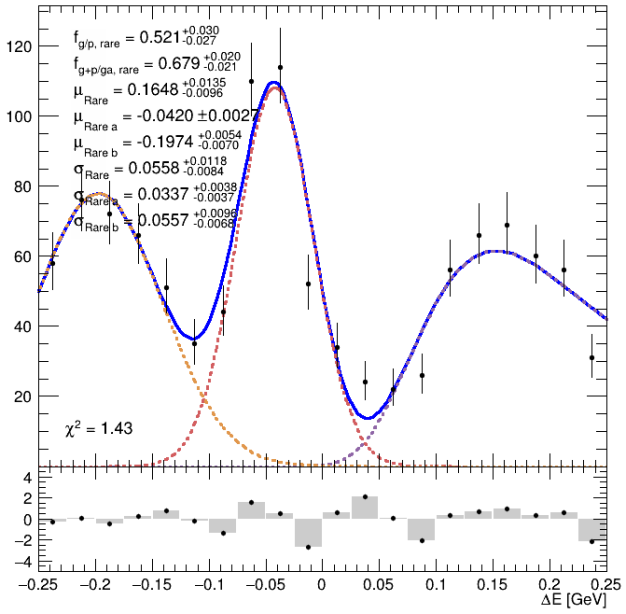


(h) Channel 1 (K^*)

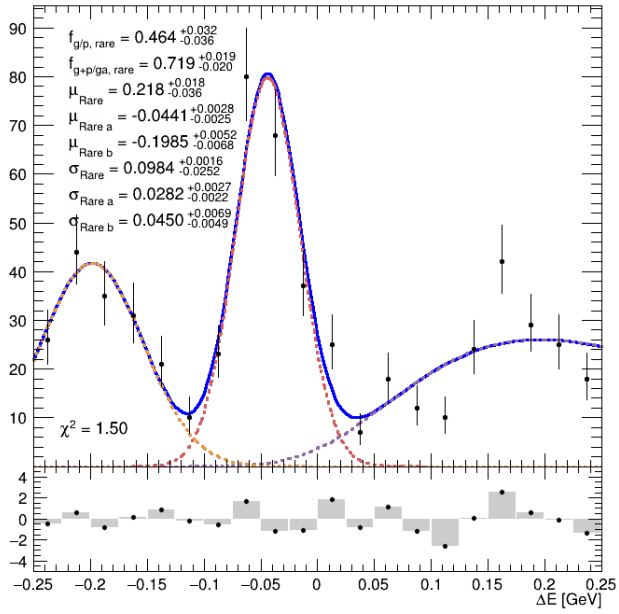


(i) Channel 2 (K_S)

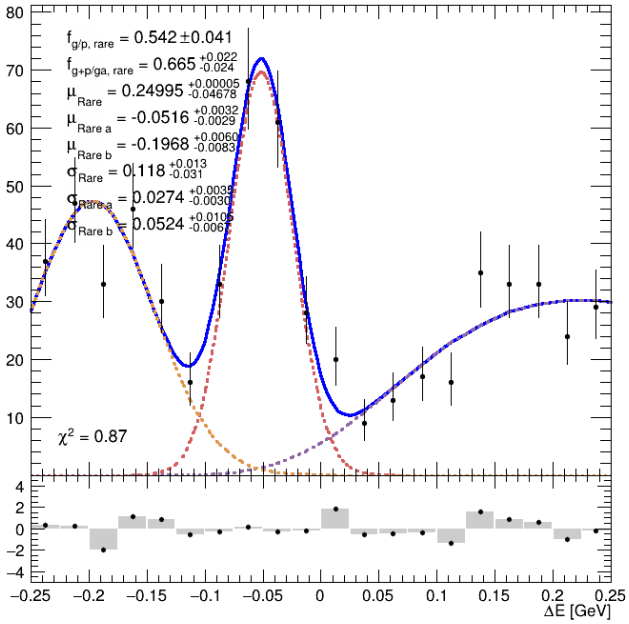
Figure 2.24: (g)-(i) signal+crossfeed models



(j) Channel 0 (ϕ)

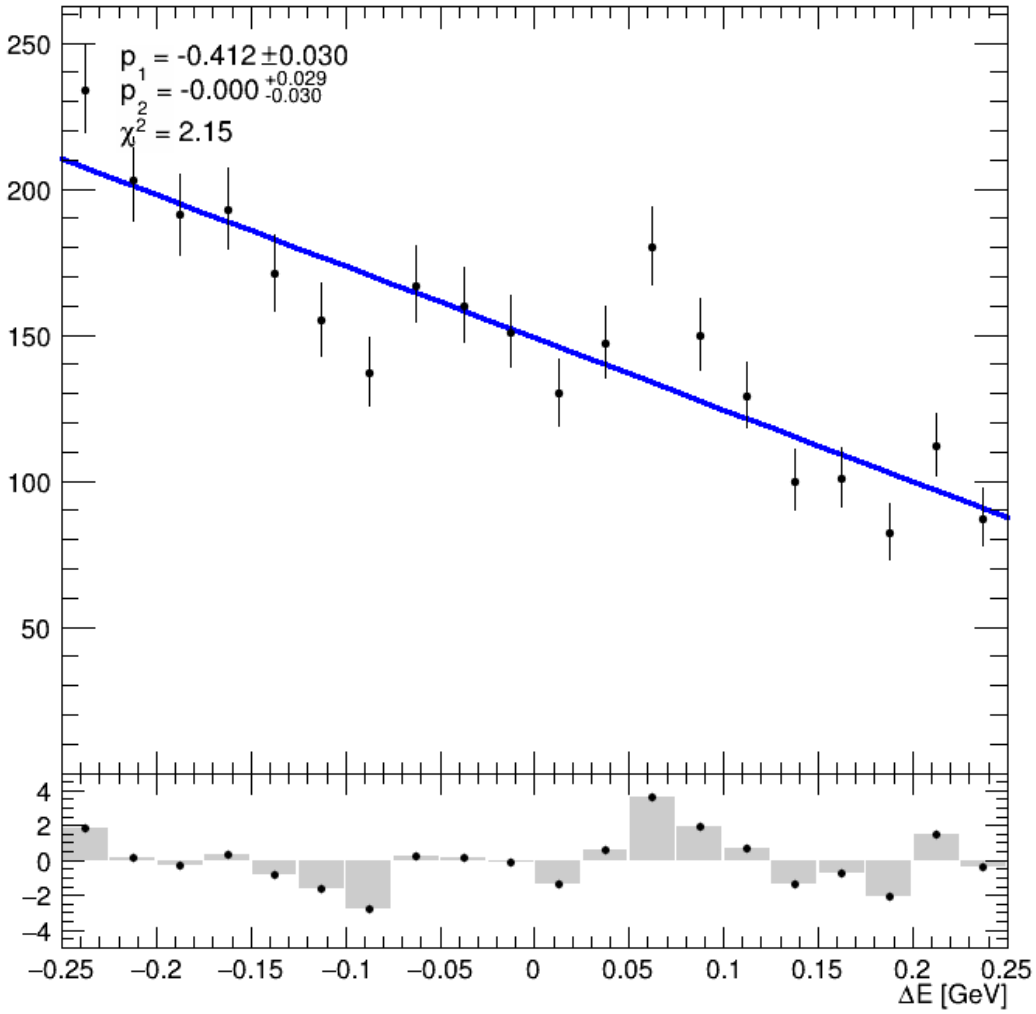


(k) Channel 1 (K^*)



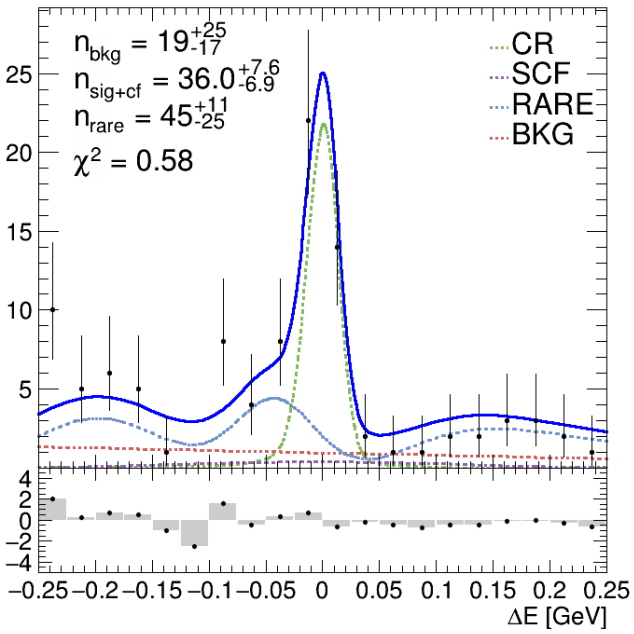
(l) Channel 2 (K_S)

Figure 2.24: Models and their fits to MC data used in control channel analysis: (j)-(l) mixedrare models

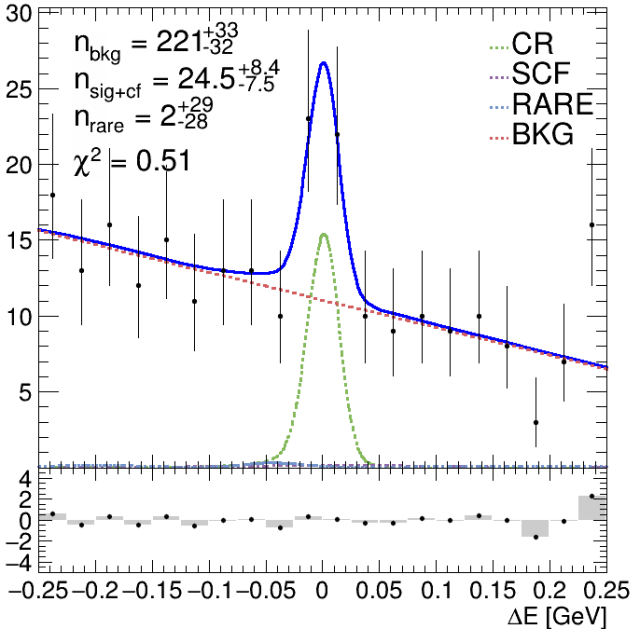


(m) Background model common for every channel

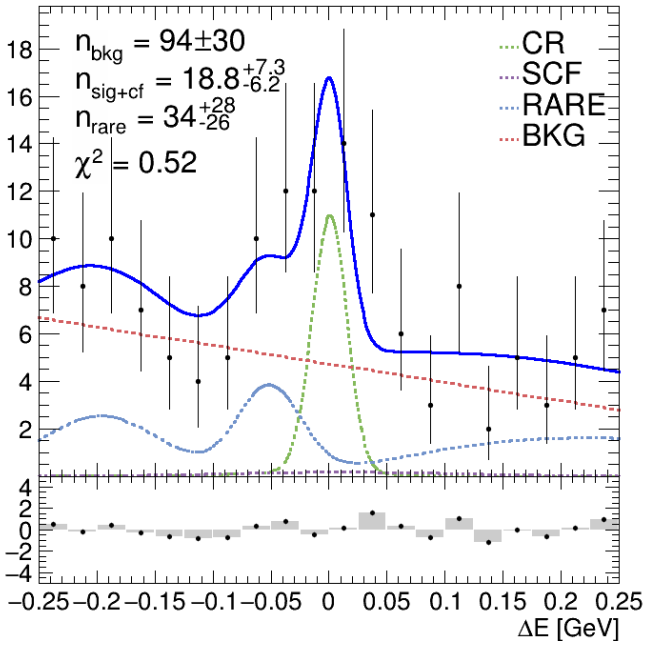
Figure 2.24: Models and their fits to MC data used in control channel analysis: (m) background model



(a) Channel 0 (ϕ)



(b) Channel 1 (K^*)



(c) Channel 2 (K_S)

Figure 2.25: Fit of the whole model to data

2.6.2 Sidebands Validation

Monte Carlo simulations that we compiled and tuned our analysis module onto, ´are just an approximation of reality. They base their decay parameters either on known data, or in case of mainly rareMC, they are an estimate based on our current knowledge and the Standard Model of particle physics. These do not necessarily correspond to the real data and real decay parameters.

In order to validate the monte carlo simulation and not to impede with our blind analysis, a sideband method was used. This method consists of setting the variable cuts to exclude our signal window. Two sideband regions were chosen for our analysis:

1. M_{bc} sideband cut $5.24 < M_{bc} < 5.276$
2. $\Delta M_{D*_s,D_s}$ sideband cut $0.096 < \Delta M_{D*_s,D_s} < 0.13$ and $0.158 < \Delta M_{D*_s,D_s} < 0.26$

the $\Delta M_{D*_s,D_s}$ cut is set to be three times the signal window of this variable. This cut was chosen to battle the possibility of low numbers of background and the fact, that M_{bc} distribution is inherently biased due to the method of the best candidate selection. We obtained results using 6 streams of genericMC merged with appropriately normalized 50 streams of rareMC. These were then normalized by the number of streams and plotted together with the 1 stream of experimental data. These results can be seen in figure 2.26 for M_{bc} sidebands, which are merged for all 3 channels due to the low number of events, and 2.27 for $\Delta M_{D_s^*,D_s}$ sidebands. These results show that no significant difference in background shape is present between the simulated and the real background, obtained from experimental data. There is, however, some difference in the number of events per bin. This signifies the Monte Carlo simulation is underestimated in our case in some channels (underestimation of MC events to real events can be clearly seen on figures 2.26 and 2.27-b. This justifies our decision of maintaining the physical cuts and not overly relying on cut optimization. However, it might be necessary to explore the reason behind it further and maybe account for this difference during the cut optimization process. Sidebands provided also show random character of the background and show no significant peaks.

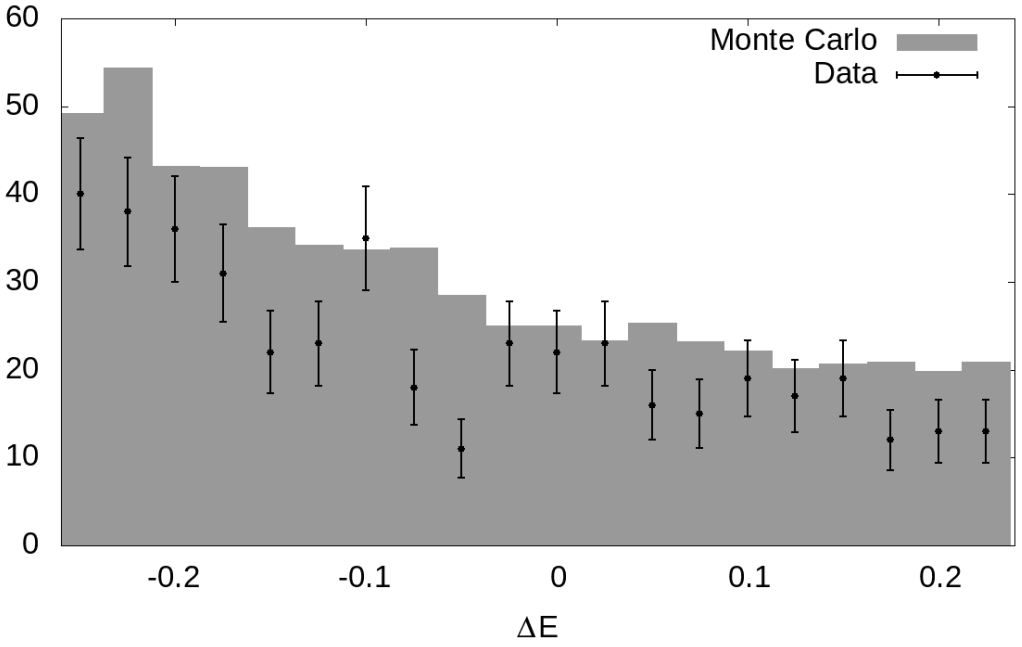
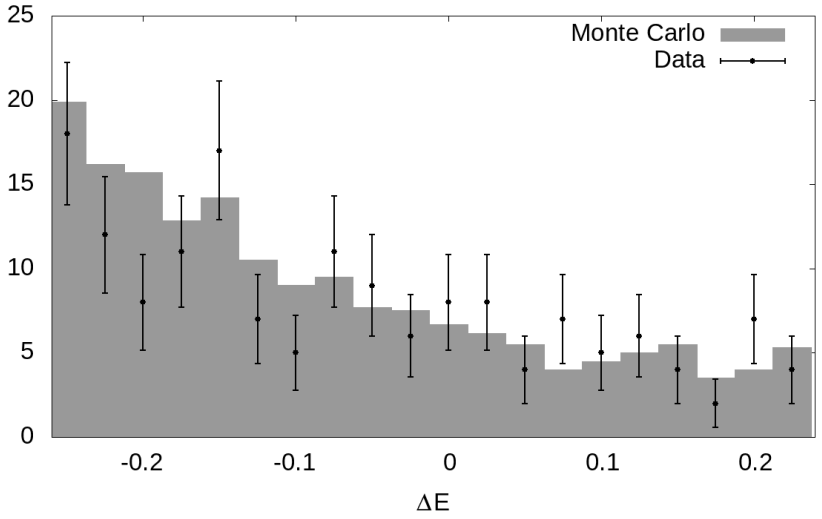
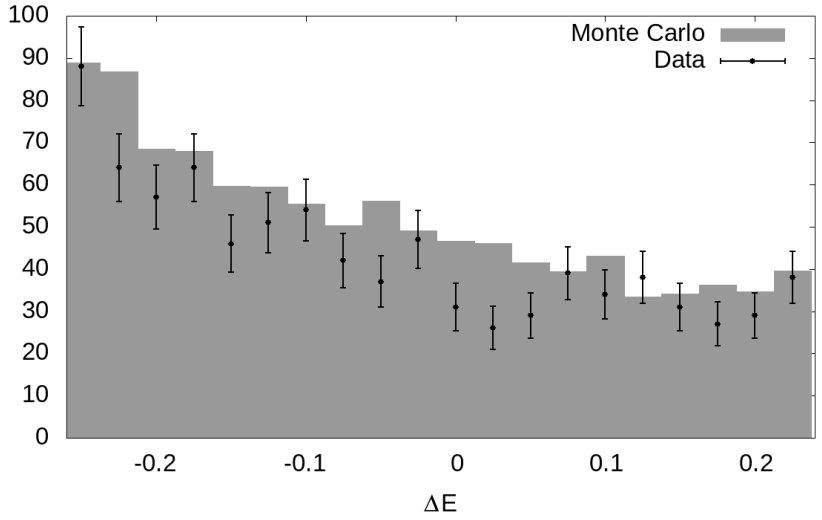


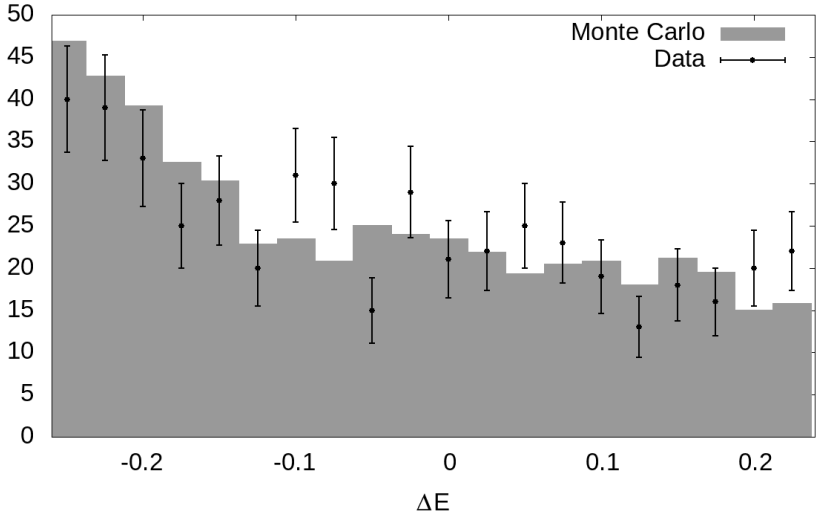
Figure 2.26: M_{bc} sidebands of our channel, comparison between experimental data and normalized Monte Carlo dataset



(a) Channel 0 (ϕ)



(b) Channel 1 (K^*)



(c) Channel 2 (K_S)

Figure 2.27: Sideband study of $\Delta M_{D_s^*,D_s}$ sidebands

Conclusion

The aim of our work was to create an analysis module and fine-tune it to analyze the decay channel of $B^0 \to D_s^{*\mp} + \rho^{\pm}$. The motivation behind this was to be able to extract a branching ratio of this decay.

To achieve our goal, we employed cut-based methodology and did an extensive study of Monte Carlo simulations. We developed a new method of cut optimization that has the potential to improve another cut-based analyses. The table 2.5 proves viability of this method.

We tested our module and fitter in three distinctive ways, each confirming viability of one of the aspects of our analysis - the toy Monte Carlo study, results of which can be seen in figures 2.12, 2.13 and 2.14, the control channel validation, which validates the module itself and sidebands study, which validates the simulations performed by Belle collaboration.

The results of the control channel study in table 2.14 and observe higher yields in 2 out of 3 channels studied with smaller errors than that of the previous study published in [25] proving the viability and hinting improvement in methodology.

The sidebands study, results of which can be seen in figures 2.26 and 2.27, has shown no significant difference between the real and simulated background shapes, however, provided some clues of underestimation of the Monte Carlo simulation. This might need to be explored further.

However, even though we managed to obtain slightly better results for control channel than the previous analysis, the rarity of our channel $B^0 \to D_s^{*\mp} + \rho^{\pm}$ combined with the fact, that ρ is a very wide resonance reconstructed from "invisible" neutral pion which has to be itself reconstructed from another particles visible for the detector, decreases the module's efficiency and the final signal yield. For this reason, although our channel has higher branching fraction than that of the control channel, the signal to noise ratio is considerably worse. This fact can be illustrated by test data fits in figures 2.18, 2.19 and 2.20. I personally find this fact interesting since it provides a challenge to fit.

With our thesis we prove the viability of our methodology and propose to use it in analysis of the real experimental data to obtain valuable physical information and proceed with yield fit of experimental data, statistical error estimation.

Bibliography

[1] Website of Belle Collaboration [Online]. `http://belle.kek.jp/`, 2018. Accessed 5.12.2018.

[2] Website of Belle-II Collaboration [Online]. `http://belle2.org/`, 2018. Accessed 5.12.2018.

[3] Aad, G., et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B716* (2012), 1–29.

[4] Aaij, R., et al. Evidence for *cp* violation in time-integrated $D^0 \to h^- h^+$ decay rates. *Phys. Rev. Lett. 108* (Mar 2012), 111602.

[5] Aaij, R., et al. Observation of *cp* violation in charm decays. *Phys. Rev. Lett. 122* (May 2019), 211803.

[6] Abashian, A., et al. The Belle Detector. *Nucl. Instrum. Meth. A479* (2002), 117–232.

[7] Abe, K., et al. Observation of large CP violation in the neutral B meson system. *Phys. Rev. Lett. 87* (2001), 091802.

[8] Abe, R., et al. BELLE/SVD2 status and performance. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 535*, 1 (2004), 379 – 383. Proceedings of the 10th International Vienna Conference on Instrumentation.

[9] Aegerter, M. A., Leventis, N., and Koebel, M. M. *Aerogels handbook*, 1 ed. Springer, 2011.

[10] Akopov, N., et al. The HERMES dual-radiator ring imaging Cherenkov detector. *Nuclear Instruments and Methods in Physics Research A 479* (Mar. 2002), 511–530.

[11] Albrecht, E., et al. The radiator gas and the gas system of compass rich-1. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 502*, 1 (2003), 266 – 269. Experimental Techniques of Cherenkov Light Imaging. Proceedings of the Fourth International Workshop on Ring Imaging Cherenkov Detectors.

[12] Arnold, R., et al. A Ring Imaging Cherenkov Detector: The DELPHI Barrel Rich Prototype, Part A: Experimental Studies of the Detection Efficiency and the Spatial Resolution. *Nucl. Instrum. Meth. A270* (1988), 255–288.

[13] Blum, W., Riegler, W., and Rolandi, L. *Particle Detection with Drift Chambers*, 2 ed. Accelerator Physics. Springer-Verlag, Heidelberg, 2008.

[14] Bosteels, M., Hahn, F., and Haider, S. LHCB RICH gas system proposal. Tech. Rep. LHCb-2000-079, CERN, Geneva, Apr 2001.

[15] Bryant, P. J. A brief history and review of accelerators.

[16] Büchner, M. A Study of Multivariate Techniques for Continuum Suppression. Bachelor thesis, Karlsruhe Institute of Technology.

[17] Cervenkov, D. $B^0 \to D^{*\mp} + \rho^{\pm}$. Belle internal note #1401.

[18] Chatrchyan, S., et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B716* (2012), 30–61.

[19] Christenson, J. H., Cronin, J. W., Fitch, V. L., and Turlay, R. Evidence for the 2π decay of the K_2^0 meson. *Phys. Rev. Lett. 13* (Jul 1964), 138–140.

[20] Close, F. *Particle physics: a very short introduction*. Oxford University Press, 2012.

[21] dalla torre, S., Levorato, S., Menon, G., Polak, J., Steiger, L., Šulc, M., and Tessarotto, F. A study of the rich gas refractive index. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 639* (05 2011), 271–273.

[22] Fang, F. Study of $K_S \to \pi^+\pi^-$ Selection. Belle internal note #323.

[23] Hoecker, A., et al. TMVA - Toolkit for Multivariate Data Analysis. arXiv:physics/0703039, 2007.

[24] Iijima, T., et al. Aerogel Cherenkov counter for the BELLE detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 453*, 1 (2000), 321 – 325. Proc. 7th Int. Conf on Instrumentation for colliding Beam Physics.

[25] Joshi, N. J., et al. Measurement of the branching fractions for B-0 –> Ds*+pi- and B-0 –> Ds*-K+ decays. *Phys. Rev. D81* (2010), 031101.

[26] Kobayashi, M., and Maskawa, T. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys. 49* (1973), 652–657.

[27] Kostelecky, V. A. The Status of CPT. In *Physics beyond the standard model. Proceedings, 5th International WEIN Symposium, Santa Fe, USA, June 14-19, 1998* (1998), pp. 588–600.

[28] Krištof, M. Gas electron multiplier based photon detection for Cherenkov imaging applications. Bachelor thesis, Charles University, Praha, 2016.

[29] Kurokawa, S.-I. Kekb status and plans.

[30] Landau, L. On the conservation laws for weak interactions. *Nuclear Physics 3*, 1 (1957), 127 – 131.

[31] Lee, T. D., and Yang, C. N. Question of parity conservation in weak interactions. *Phys. Rev. 104* (Oct 1956), 254–258.

[32] Moloney, G. R., et al. Design and Structure of the Upgraded Silicon Vertex Detector at Belle.

[33] Mordvinstev, A., Olah, C., and Tyka, M. Google AI Blog: Inceptionism: Going Deeper into Neural Networks [Online]. `https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html`, 2015. Accessed 10.12.2018.

[34] Moser, H.-G. The Belle II DEPFET pixel detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 831* (2016), 85 – 87. Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors.

[35] Mrvar, M., et al. First experience with belle ii aerogel rich detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (2018).

[36] N. J. Joshi, T. Aziz, K. T. Measurement of $\mathcal{BR}(B_{0s} * h)$. Belle internal note #995.

[37] Nishida, S. Study of Kaon and Pion Identification Using Inclusive D Sample. Belle internal note #779.

[38] Noether, E. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse 1918* (1918), 235–257.

[39] Quinlan, J. R. Induction of decision trees. *Mach. Learn. 1*, 1 (Mar. 1986), 81–106.

[40] Roodman, A. Blind analysis in particle physics.

[41] Roskot, M. Particle Identification using Ring Cherenkov Detector Technology at COMPASS Experiment. Master's thesis, Charles University of Prague, 2015.

[42] Ryd, A., et al. EvtGen A Monte Carlo Generator for B-Physics. *BAD 522* (Feb 2004).

[43] Schwinger, J. The theory of quantized fields. i. *Phys. Rev. 82* (Jun 1951), 914–927.

[44] Tajima, H., et al. Proper-time resolution function for measurement of time evolution of b mesons at the kek b-factory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 533*, 3 (2004), 370 – 386.

[45] Tanabashi, M., et al. Review of particle physics. *Phys. Rev. D 98* (Aug 2018), 030001.

[46] Wu, C. S., et al. Experimental test of parity conservation in beta decay. *Phys. Rev. 105* (Feb 1957), 1413–1415.

List of Figures

List of Tables

List of Abbreviations

ACC - Aerogel Cherenkov Counter

ANN - Artificial Neural Network

BASF - Belle AnalysiS Framework

BDT - Boosted Decision Tree

BDTB - BDT with Bagging

BDTG - BDT with Gradient boosting

BF - Branching Fraction

BFGS - Broyden-Fletcher-Goldfarb-Shannon method

BTT - "Bottom-to-top" ordering

CC - Control Channel

CDC - Central Drift Chamber

CERN - Conseil Européen pour la Recherche Nucléaire, European Organization for Nuclear Research

CF - (Self-)crossfeed

CFMlpANN - Clermont-Ferrand Multi-layer perceptron Artificial Neural Network

CMS - Center of Mass System

CP - Charge-Parity

CPT - Charge-Parity-Time

Ch - Channel

DEPFET - DEPleted Field Effect Transistor

DSSD - Double Sided Silicon Detector

EFC - Extreme Forward Calorimeter

EM - ElectroMagnetic

FDA - Function Discrimination Analysis

FM-PMT - Fine-Mesh Photo-Multiplier Tube

FOM - Figure-Of-Merit

FSP - Final State Particle(s)

FSR - Final State Radiation

FTD - Final Trigger Decision

GDL - Global Decision Logic

HER - High Energy Ring

IP - Interaction Point

ITD - Input Trigger Delay

KEK - Kō Enerugī Kasokuki Kenkyū Kikō, The High Energy Accelerator Research Organization in Tsukuba, Japan

KEKB - KEK B-factory, particle accelerator

KLM - K_L and Muon detector

KSFW - Kakuno-Super-Fox-Wolfram moments

LER - Low Energy Ring

MC - Monte Carlo simulation

MLP - Multi-Layer Perceptron

MLPBFGS - Multi-Layer Preceptron using BFGS method

MVA - MultiVariate Analysis

PDE - Probability Density Estimator

PDF - Probability Density Function

PDG - Particle Data Group

PID - Particle IDentification

PSNM - PreScale aNd Mask

ROC - Receiver Operating Characteristic

RPC - Resistive Plate Counter

SCF - Self-CrossFeed

SLAC - Standford Linear Accelerator Complex

SVD - Silicon Vertex Detector

TDC - Time-to-Digital Converter

TMD - TiMing Decision

TMVA - Toolkit for MultiVariate data Analysis

TOF - Time Of Flight detector