SEARCH FOR BARYON-NUMBER-VIOLATING PROCESSES IN B^- DECAY TO FINAL STATE $\overline{\Xi}{}^0_C\overline{\Lambda}{}^-_C$ AT BELLE AND DESIGN OF BELLE II TOP TRIGGER SYSTEM

by

Tianping Gu

Bachelor of Science, University of Science and Technology of China, 2018

Submitted to the Graduate Faculty of
the Dietrich School of Arts and Sciences in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Tianping Gu

It was defended on

December 5th, 2023

and approved by

Vladimir Savinov, University of Pittsburgh

John Alison, Carnegie Mellon University

Brian Batell, University of Pittsburgh

Hanna Salman, University of Pittsburgh

James Mueller, University of Pittsburgh

Dissertation Director: Vladimir Savinov, University of Pittsburgh

SEARCH FOR BARYON-NUMBER-VIOLATING PROCESSES IN B^- DECAY TO FINAL STATE $\overline{\Xi}^0_C\overline{\Lambda}^-_C$ AT BELLE AND DESIGN OF BELLE II TOP TRIGGER SYSTEM

Tianping Gu, PhD

University of Pittsburgh, 2023

This dissertation has two parts which describe my work on the Belle and Belle II experiments. The first part provides a detailed account of my search for baryon-number violating processes using Belle data. The second part outlines my contributions to the Belle II trigger based on information from Time-of-Propagation (TOP) counters.

In the first part I report the results of the first search for B^- decay to final state $\overline{\Xi}^0_c\overline{\Lambda}^-_c$ using 711 fb^{-1} of data collected at the $\Upsilon(4S)$ resonance with the Belle detector at the KEKB asymmetric-energy e^+e^- collider. The results are interpreted in terms of a direct baryon-number-violating B^- decay and, alternatively, in terms of $\Xi^0_c - \overline{\Xi}^0_c$ oscillations which follow the Standard Model decay $B^- \to \Xi^0_c\overline{\Lambda}^-_c$. No evidence for baryon number violation is observed and 95% confidence-level upper limits are set on the ratio of baryon-number-violating and Standard Model branching fractions $\mathcal{B}(B^- \to \overline{\Xi}^0_c\overline{\Lambda}^-_c)/\mathcal{B}(B^- \to \Xi^0_c\overline{\Lambda}^-_c) < 2.7\%$ and on $\Xi^0_c - \overline{\Xi}^0_c$ oscillation angular frequency $\omega < 0.76$ ps^{-1}.

In the second part I report on my studies in the area of TOP-based trigger system (TOP TRG) which provides precise collision timing information for the Belle II Trigger system. The development of TOP TRG system, including the hardware, firmware design, and performance analysis are presented. The standalone TOP TRG achieves the efficiency of 92.5% for barrel cosmic muons. By incorporating CDC-TOP matching, an efficiency of 75% and a timing resolution of 12 ns are achieved for hadronic events. I discuss further possible TOP TRG improvements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First of all, I want to express my sincere gratitude to my advisor, Prof. Vladimir Savinov, for his guidance, motivation, and mentorship during my PhD studies. I will always remember the days we spent together discussing physics, developing, and analyzing code and firmware. Vladimir has always been supportive in all aspects of my life. He makes every effort to create a comfortable and productive work environment for everyone. I feel fortunate to have such a wonderful PhD advisor.

I'd also express my appreciation to the members of my PhD committee, Prof. John Alison, Prof. Brian Batell, Prof. Hanna Salman, and Prof. James Mueller, for their valuable feedback and support.

The physical analysis work was accomplished under the guidance and efforts of many Belle members. I appreciate the guidance and suggestions from Jim Libby, Karim Trabelsi, Yun-Tsung Lai, Lena Solovieva, and Thomas Kuhr.

I would like to thank all the people I met in Belle TRG and TOP group: Taichiro Koga, Hideyuki Nakazawa, Yuuji Unno, Yoshihito Iwasaki, Vasily Shebalin, Martin Bessner, Kurtis Nishimura, and Gary Varner. I learned many aspects of the experiment from them.

I also want to thank all my friends at Pitt: Erfei Wang, Tong Pang, Zehua Wang, Qian Song, Yuze Wang, Tianyi Wang and others. I feel lucky to be friends and have worked with them.

Finally, I want to express my deepest appreciation to my family. Their support and encouragement have been a constant source of inspiration, helping me overcome challenges throughout these years.

It has been a long and rewarding journey, and I feel incredibly fortunate to have crossed paths with so many remarkable people along the way. Thank you!

1.0 INTRODUCTION

1.1 BARYON ASYMMETRY AND BARYOGENESIS

The Big Bang should have created equal amounts of matter and antimatter in the early universe. However, the universe today appears to consist predominantly of matter rather than antimatter. The baryon asymmetry of the universe (BAU) can be defined as the difference between the number of baryons N_B and antibaryon $N_{\bar{B}}$, divided by the number of photons N_γ. It is measured using two different indirect probes, Big Bang Nucleosynthesis (BBN) and the Cosmic Microwave Background (CMB), to be $(N_B - N_{\bar{B}})/N_\gamma \approx N_B/N_\gamma = (6.10 \pm 0.4) \times 10^{-10}$ [1], which is a clear asymmetry between baryons and antibaryons.

Baryogenesis is a physical process that is hypothesized to have taken place during the early universe to produce baryon asymmetry. Sakharov proposed three necessary conditions for baryogenesis [2]:

1. Baryon number violation (BNV). Without baryon number violation, no system can evolve from a state with baryon number $B = 0$ to a state with $B \neq 0$.

2. C and CP violation. C and CP violation is required because otherwise for each process that generates a baryon asymmetry, there would be a C or CP conjugate process that generates the same asymmetry with the opposite sign and the same probability. Therefore, even in the presence of baryon number violation, there would be no net effect without C and CP violation.

3. Departure from thermal/chemical equilibrium. CPT theorem assures that masses of particles and antiparticles are equal and hence their thermal equilibrium densities remain equal. Thus, the baryogenesis process must happen during the thermal non-equilibrium phase.

The first condition, BNV, is satisfied in the Standard Model (SM) since the quantum number B is anomalous, and non-perturbative sphaleron transitions [3] violate baryon number. Sphaleron processes are efficient at temperatures above the electroweak scale but are exponentially suppressed at zero temperatures. However, there is no experimental evidence of baryon number violation decays so far. The second condition, CP violation, comes from the irreducible complex phase in the Cabibbo-Kobayashi-Maskawa matrix [4, 5] and it has been proven to be insufficient to account for the BAU [6, 7]. Concerning the third condition, the electroweak phase transition could have provided the departure from thermal equilibrium if it were strongly first order. However, it has been understood that the electroweak phase transition in the Standard Model is a smooth crossover [8].

In the following sections, I focus on the first Sakharov condition, baryon number violation.

1.2 BARYON NUMBER VIOLATION

Baryon number (B) is defined as $B = \frac{1}{3}(n_q - n_{\bar{q}})$, where n_q is the number of quarks and $n_{\bar{q}}$ is the number of antiquarks. Baryons (e.g. proton and neutron) have $B = 1$, antibaryons have $B = -1$ and mesons have $B = 0$.

Baryon number is observed to be a good symmetry and the stability of ordinary matter is attributed to the baryon number conservation. For example, the proton, the lightest particle which $B \neq 0$, would be absolutely stable if the baryon number is exactly conserved.

However, the absolute conservation of baryon number has been questioned both theoretically and experimentally for many years. Experimentally, the baryon asymmetry of the universe could not be explained without the baryon number violation process. Theoretically, the baryon number conservation is an accidental symmetry, i.e., not protected by any gauge symmetry, of the SM Lagrangian, which is true only at the perturbative level. Baryon number could be violated non-perturbatively via sphaleron transition. Therefore, various theories beyond the Standard Model with BNV processes have been proposed and various experimental efforts have been performed to search for baryon number violation.

Although the baryon number B and lepton number L symmetries are broken in the

SM, $B - L$ remains to be a conserved quantum number. This, however, is not so in various SM extensions. In the language of effective field theories based on Wilson's Operator Product Expansion, the lowest dimension operators which allow BNV are dimension-six and dimension-seven and require $|\Delta(B - L)| = 0$ and 2, respectively [9, 10, 7]. This framework inspires to search for BNV in a variety of processes such as

1. proton decay in which $\Delta B = -\Delta L = 1$.
2. τ decays in which $\Delta B = \Delta L = \pm 1$ while keeping $|\Delta(B - L)| = 0$.
3. neutron-antineutron/baryon-antibaryon oscillations in which $\Delta B = 2$.
4. $B/D/\Lambda$ decay in which $\Delta B = \Delta L = \pm 1$
5. neutrinoless double-beta decay in which $\Delta L = 2$.

1.3 EXPERIMENTAL EFFORTS

1.3.1 Proton Decay

Grand Unified Theory (GUT) predicts protons will decay via the exchange of a massive gauge boson between two quarks. The favored gauge-mediated decay mode is $p \to e^+\pi^0$. Super-Kamiokande is currently the largest detector for observation of proton decay and has an excellent capability to observe this decay mode [11, 12].

Super-Kamiokande (SK) is located about 1,000 m under Mount Ikeno in Gifu Prefecture, Japan. The detector was designed to detect high-energy neutrinos and search for proton decay. It consists of a cylindrical stainless steel tank that is 41.4 m tall and 39.3 m in diameter, holding 50220 metric tons of ultrapure water. By using water as the source of protons, a proton from either hydrogen or oxygen would decay into a positron and a π^0. The π^0 will decay to two photons immediately. The positron and two photons produce three electromagnetic showers respectively and are then observed by super-Kamiokande.

Searching for proton decay via $p \to e^+\pi^0$ has been performed with the data of Super-Kamiokande from April 1996 to May 2018, but no candidates have been found [13]. The lower limit on the lifetime for this mode is set to: $\tau(p \to e^+\pi^0) > 2.4 \times 10^{34}$ yrs at the 90%

confidence level.

Proton decay mode $p \to \mu^+\pi^0$ is also studied and the limit on the lifetime is set to: $\tau(p \to \mu^+\pi^0) > 1.6 \times 10^{34}$ yrs at the 90% confidence level.

1.3.2 τ Decays

Baryon number violation in charged lepton decays often implies lepton number violation (LNV), which is necessarily associated also with lepton *flavor* violation (LFV), though possible existence of the latter phenomenon does not imply the former BSM scenario which is more relevant for BNV via leptogenesis. The observation of neutrino oscillations was the first evidence for LFV (though only via mixing, not at perturbative tree level of any SM extension). It implies that LFV and/or LNV may also exist in the charged sector but with much smaller branching fractions. Many BSM theories predict enhanced LNV in τ^- decays concerning μ^- decays [14]. Therefore, several experiments performed searches for baryon number and lepton number violation in τ decays.

LHCb collaboration searched for two LNV and BNV decay modes $\tau^- \to \bar{p}\mu^+\mu^-$ and $\tau^- \to p\mu^-\mu^-$ with LHCb data sample from 2011, corresponding to an integrated luminosity of 1.0 fb^{-1} collected at $\sqrt{s} = 7$ TeV [15]. The observed limits at 90% (95%) confidence level for the branching fractions are

$$B(\tau^- \to \bar{p}\mu^+\mu^-) < 3.3\ (4.3) \times 10^{-7}$$

and

$$B(\tau^- \to p\mu^-\mu^-) < 4.4\ (5.7) \times 10^{-7}.$$

Belle collaboration searched for six LNV and BNV decay modes $\tau^- \to \bar{p}e^+e^-$, pe^-e^-, $\bar{p}e^+\mu^-$, $\bar{p}e^-\mu^+$, $p\mu^-\mu^-$ and $\bar{p}\mu^+\mu^-$ with 711 fb^{-1} of data recorded at the $\Upsilon(4S)$ resonance and 121 fb^{-1} of data recorded near the $\Upsilon(5S)$ peak with the Belle detector at the KEKB asymmetric-energy e^+e^- collider [16]. The 90% confidence level upper limits for the branching fractions have been measured to be

$$B(\tau^- \to \bar{p}e^+e^-) < 3.0 \times 10^{-8},$$

$$B(\tau^- \to pe^-e^-) < 3.0 \times 10^{-8},$$

$$B(\tau^- \to \bar{p}e^+\mu^-) < 2.0 \times 10^{-8},$$

$$B(\tau^- \to \bar{p}e^-\mu^+) < 1.8 \times 10^{-8},$$

$$B(\tau^- \to p\mu^-\mu^-) < 4.0 \times 10^{-8},$$

and

$$B(\tau^- \to \bar{p}\mu^+\mu^-) < 1.8 \times 10^{-8}.$$

1.3.3 Neutron-antineutron Oscillations

Neutron-antineutron ($n - \bar{n}$) oscillations are another promising way of searching for the baryon number violation. It is inspired by $K^0 - \bar{K}^0$ mixing which violates strangeness by two units. $n - \bar{n}$ oscillations violate the baryon number by two units and $B - L$ by two units as well. $n - \bar{n}$ oscillations would test a different sector of BNV compared to proton/τ decay, where $|\Delta B| = 1$ and $\Delta(B - L) = 0$. Moreover, the effective operators responsible for $n - \bar{n}$ oscillations have dimension-nine and are suppressed by five inverse powers of a mass scale, so they probe new physics at mass scales of TeV and above, which is much lower than the GUT energy scale [17] of $\sim 10^{15} - 10^{16}$ GeV. In principle, there are two ways to search for $n - \bar{n}$ oscillations: free neutrons and neutrons bound inside nuclei.

The basic idea for experiments that use free neutrons is to prepare a beam of slow neutrons that propagate freely to a distant annihilation target. $n - \bar{n}$ oscillations could happen during the time of propagation and pure n state could oscillate to $n - \bar{n}$ mix state. The generated $\bar{n}$ annihilates with the target and generates several secondary pions which are then detected by the tracking detector. The current best limit for free $n - \bar{n}$ oscillations was performed at the Institut Laue-Langevin (ILL) in Grenoble in 1994 [18]. The ILL experiment used a cold neutron beam from a 58 MW research reactor with a neutron current of 1.25×10^{11} neutrons/s incident on the annihilation target. It achieved a limit of $\tau_{n\bar{n}} > 0.86 \times 10^8$ s.

Experiments that search for $n - \bar{n}$ oscillations with bounded neutrons are performed in large underground detectors. The current best limit for bounded $n - \bar{n}$ oscillations comes

from the Super-Kamiokande experiment [19]. By using the water as the source of neutrons, antineutrons generated from $n-\bar{n}$ oscillations are expected to annihilate quickly with one of the surrounding nucleons and produce multiple secondary hadrons that could be detected. With the data recorded in the course of 1489 days between May 31, 1996 and July 15, 2001, super-Kamiokande has set the lower limit on the lifetime for bounded neutrons to be 1.9×10^{32} yrs at the 90% confidence level. The corresponding lower limit for the oscillation time of free neutrons was calculated to be 2.7×10^{8} s using a theoretical value of the nuclear suppression factor of $0.517 \times 10^{23}\ s^{-1}$.

1.3.4 Light Baryon-antibaryon Oscillations

$\Lambda - \bar{\Lambda}$ oscillations are proposed to be a new way to probe BNV [20]. By coherent $\Lambda\bar{\Lambda}$ production in $J/\Psi \to \Lambda\bar{\Lambda}$ decay process, one can set an upper limit on $\Lambda - \bar{\Lambda}$ oscillation time to about 10^{-6} s. No experimental results in this are have been reported so far.

1.3.5 Charmed or Beautiful Baryon-antibaryon Oscillations

Charmed and bottom baryon-antibaryon oscillations have been proposed as a possible mechanism for baryogenesis [21]. The proposed model introduces CP-violating oscillations of neutral, heavy-flavor baryons into antibaryons at rates which are within a few orders of magnitude of their lifetimes. The model introduces four new particles: three light Majorana fermions and a colored scalar. The lightest of these fermions is typically long lived (on collider time scales) and may be produced in decays of bottom and possibly charmed baryons. Alternatively, such baryons could be created in the early universe via out-of-equilibrium decays of this Majorana fermion after hadronization but before nucleosynthesis. Therefore, this novel approach to baryogenesis also fulfills the out-of-equilibrium Sakharov condition. The discussed model could be easily embedded in an R-parity-violating supersymmetric theory, providing important connections to solving the puzzle of dark matter and the unification of fundamental forces.

LHCb recently searched for $\Xi_b^0 - \overline{\Xi}_b^0$ oscillations with pp collision data sample corresponding to an integrated luminosity of 3 fb^{-1} collected at $\sqrt{s} = 7$ TeV and 8 TeV [22]. Two

decay modes $\Xi_b^{'-} \to \Xi_b^0 \pi^-$ and $\Xi_b^{*-} \to \Xi_b^0 \pi^-$ were used, where $\Xi_b^{'-}$ and Ξ_b^{*-} are the narrow resonances $\Xi_b^{'}(5935)^-$ and $\Xi_b^{*}(5955)^-$. The upper limit on the $\Xi_b^0 - \bar{\Xi}_b^0$ oscillation rate was measured to be $\omega < 0.08$ ps^{-1} at the 95% confidence level.

1.3.6 B Meson Decays

BABAR collaboration searched for six LNV and BNV B meson decay modes $B^0 \to \Lambda_c^+ l^-$, $B^- \to \Lambda l^-$ and $B^- \to \bar{\Lambda} l^-$, where the lepton l is a muon or an electron [23]. The data sample consists of 429 fb^{-1} recorded at the $\Upsilon(4S)$ resonance with BABAR detector at the PEP-II e^+e^- collider [24]. No significant signal is observed for any decay modes and the 90% confidence level upper limits for branching fractions are measured to be

$$B(B^0 \to \Lambda_c^+ \mu^-) < 180 \times 10^{-8},$$

$$B(B^0 \to \Lambda_c^+ e^-) < 520 \times 10^{-8},$$

$$B(B^0 \to \Lambda \mu^-) < 6.2 \times 10^{-8},$$

$$B(B^0 \to \Lambda e^-) < 8.1 \times 10^{-8},$$

$$B(B^0 \to \bar{\Lambda} \mu^-) < 6.1 \times 10^{-8},$$

and

$$B(B^0 \to \bar{\Lambda} e^-) < 3.2 \times 10^{-8}.$$

1.3.7 D Meson Decays

BES III collaboration searched for four LNV and BNV D meson decay modes $D^+ \to \bar{\Lambda}(\bar{\Sigma}^0) e^+$ and $D^+ \to \Lambda(\Sigma^0) e^+$ [25]. The data sample consists of 2.93 fb^{-1} at $\sqrt{s} = 3.773$ GeV with BES III detector at the BEPCII [26]. No significant signal was observed for any decay modes and the 90% confidence level upper limits for branching fractions were measured to be

$$B(D^+ \to \Lambda e^+) < 1.1 \times 10^{-6},$$

$$B(D^+ \to \bar{\Lambda} e^+) < 6.5 \times 10^{-7},$$

$$B(D^+ \to \Sigma^0 e^+) < 1.7 \times 10^{-6},$$

and

$$B(D^+ \to \bar{\Sigma}^0 e^+) < 1.3 \times 10^{-6}.$$

1.3.8 Λ Hyperon Decays

CLAS collaboration searches for eight BNV decay modes in which Λ decays to a charged meson and a charged lepton [27]. The meson is either a $\pi^\pm$ or $K^\pm$ and the lepton is either a $e^\mp$ or $\mu^\mp$. Λ was produced by photon beam incident on a liquid hydrogen target through the exclusive reaction $\gamma p \to K^+\Lambda$ [28].

The data sample consists of roughly 1.8×10^6 reconstructed Standard Model $\Lambda \to p\pi^-$ decays. No significant signal is observed for any decay modes and the 90% confidence level upper limits for branching fractions are set in the range $(4 - 200) \times 10^{-7}$.

2.0 PARTICLES INTERACTIONS IN MATTER

In this section, I review the interactions of charged and neutral particles (photons and K_L^0 mesons) in matter [29]. The charged particles of interest are divided into two categories: (1) electrons and (2) heavy particles (i.e., charged particles heavier than electrons, such as $\pi^{\pm}$, $K^{\pm}$, proton and muon) according to their different behaviors when interacting with matter. I focus on processes of interest for particle detectors such as ionization, bremsstrahlung, pair production, and Cherenkov radiation.

2.1 ELECTRONIC ENERGY LOSS BY HEAVY PARTICLES

Charged particles traversing the detector and interacting with matter lose energy in collisions with electrons, leading to ionization, atomic, or collective excitation. The mean rate of energy loss by moderately relativistic heavy charged particles is well described by the Bethe-Bloch equation:

$$\left\langle -\frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right],$$

where the coefficient $K = 4\pi N_A r_e m_e c^2$, N_A is Avogadro s number, r_e is the classical electron radius, m_e is electron mass, c is the speed of light in vacuum, z is the charge number of incident particles, Z is the atomic number of the absorber and A is the atomic mass of the absorber. W_{max} is the maximum possible energy transfer to an electron in a single collision

$$W_{max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e/M + (m_e/M)^2},$$

and I is mean excitation energy. The theory of energy loss by ionization and excitation as given by Bethe is based on the first-order Born approximation. This free-electron approximation is not valid if W is not large compared to electron binding energies. Correction due to electronic binding is accounted for by introducing this effective ionization energy I, which is a geometric average of the excitation energies of the medium weighed by the corresponding oscillator strength. $\delta(\beta\gamma)$ is density effect correction to ionization energy loss due to the polarization of the medium at high energies.

Bethe equation is valid in the region $0.1 \lesssim \beta\gamma \lesssim 1000$ with an accuracy of a few percent. For momenta outside this region, the predictions of the Bethe equation need to be corrected: for smaller momenta, the mean excitation energy correction needs to be taken into account, and for larger momenta density effect correction begins to be important. Mass stopping power vs. $\beta\gamma$ predicted by Bethe equation for muons in copper is shown as the "Bethe" region in Fig. 1.

2.2 MULTIPLE SCATTERING THROUGH SMALL ANGLES

In addition to inelastic collisions with the atomic electrons, particles passing through matter experience elastic Coulomb scattering from nuclei. Since nuclei are usually heavier than incoming particles, the energy transfer is negligible, but each scattering adds a small deviation to the incoming particle's trajectory. The statistical outcome from a combination of many small-angle scatters is a multiple scattering angle whose distribution is approximately Gaussian (due to the Central Limit Theorem). The angular dispersion can be calculated as

$$\theta_0 = \frac{13.6}{\beta cp}\, z \sqrt{\frac{x}{X_0}} \left[1 + 0.038 \ln\left(\frac{x}{X_0}\right)\right],$$

where p, βc, and z are the momentum, speed, and charge number of the incident particle, respectively. x/X_0 is the thickness of the medium in units of radiation length (discussed in the next section).

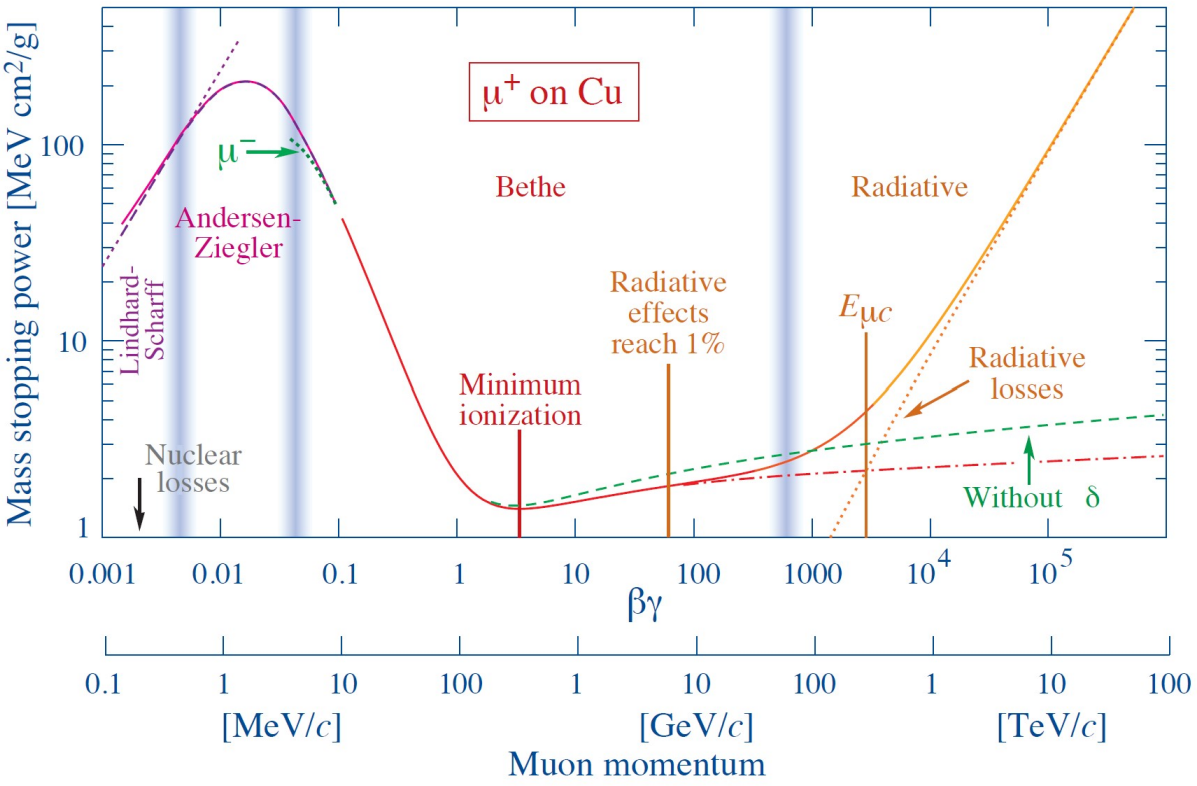


Figure 1: Mass stopping power (dE/dx) vs. $\beta\gamma$ for muons in copper. Mass stopping power in the "Bethe" region is calculated using the Bethe equation.

2.3 ELECTRON INTERACTIONS IN MATTER

Stopping power for electrons differs from that for heavy particles because of the small electron mass and spin configuration in the initial and final states. At low energies, electrons primarily lose energy by ionization, other processes (Møller scattering, Bhabha scattering, annihilation) also contribute. At high energies, electrons predominantly lose energy in matter by bremsstrahlung which is electromagnetic radiation produced by the deceleration of a charged particle.

Fig. 2 shows the energy losses by electrons through different processes in lead. While ionization losses rise logarithmically with energy, bremsstrahlung losses rise nearly linearly. The critical energy E_c (not shown in figure) is defined as the energy where these two energy

loss rates are equal to each other.

Radiation length for high-energy electrons High-energy electrons predominantly lose energy in matter via bremsstrahlung. The radiation length (X_0) is the characteristic length that describes the energy loss of electrons and could be approximated by the following analytical formula:

$$X_0 = \frac{716.4\,\mathrm{g\,cm^{-2}}A}{Z(Z+1)\ln\left(287/\sqrt{Z}\right)}.$$

Note that X_0 in this formula is measured in units of $\mathrm{g/cm^{-2}}$, therefore facilitating the calculation of the radiation length for complex compounds. To express the radiation length in units of length, the above formula needs to be divided by density. The radiation length is the mean distance over which the electron energy is reduced by a factor of $1/e$ due to radiation losses only. The energy loss by bremsstrahlung can be expressed using radiation length (after dividing by density):

$$-\frac{dE}{dx} = \frac{E}{X_0}$$

Radiation lengths are different for different materials, and higher-Z materials have shorter radiation lengths. Electromagnetic calorimeters are usually built with high-Z materials to fully absorb the electron energy. Since the discussed formula is only an approximation, it should be used only when data are not available.

2.4 PHOTON INTERACTIONS IN MATTER

Processes contributing to the scattering and absorption of photons in matter include Rayleigh scattering, photoelectric effect, Compton scattering, and pair production. Fig. 3 shows the contributions of different processes to the photon interaction cross section vs. energy in carbon and lead. The different processes involving photons are described below.

1. Rayleigh scattering. A photon (elastically) scatters from bound electrons without ionizing or exciting the atom. The recoil energy of the atom is generally negligible and the photon essentially retains its energy.

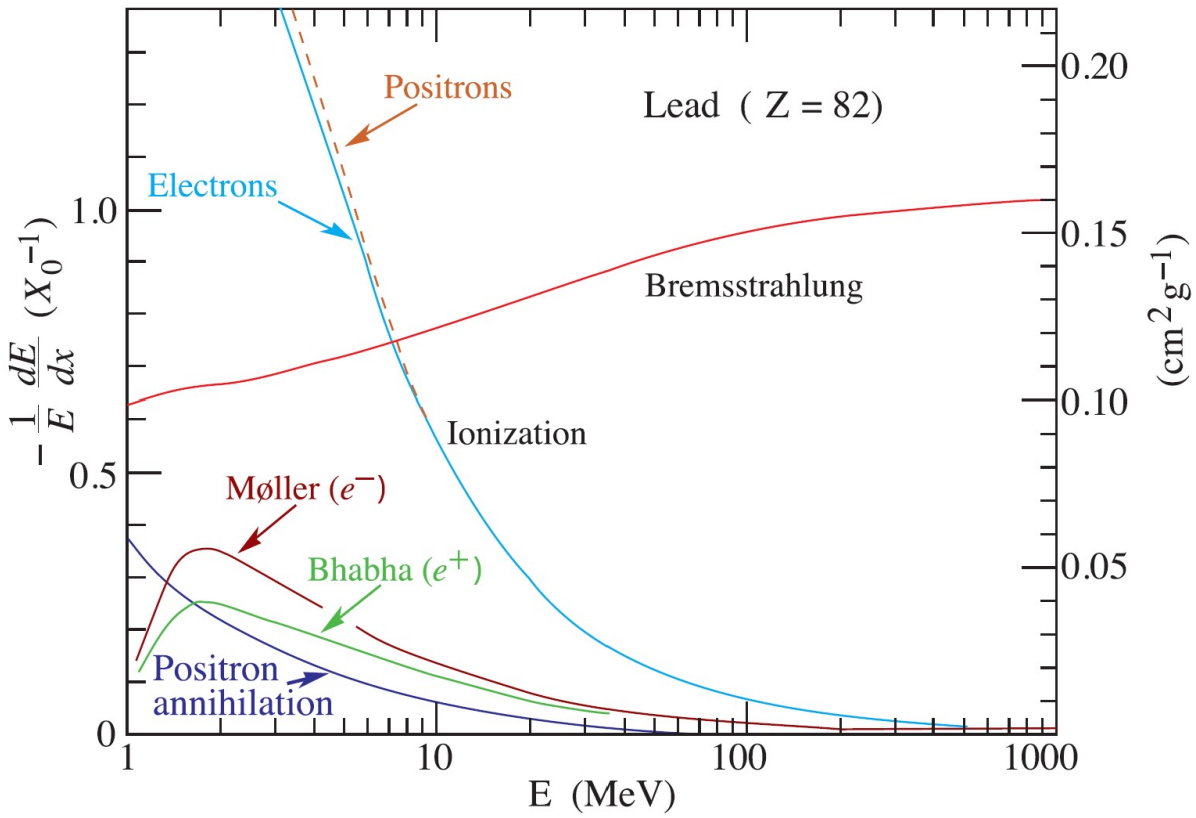


Figure 2: Fractional energy loss per radiation length in lead vs. electron or positron energy.

2. Photoelectric effect. A photon is absorbed and an electron is released from its atomic binding state. The photoelectric effect is the dominant photon process at low energies. The photoelectric cross section exhibits discontinuities (absorption edges) as photoelectric thresholds of various atomic levels are reached.
3. Compton scattering. A photon collides with an electron and transfers part of its energy to the electron. Compton scattering is inelastic scattering. Compton effect is the dominant photon process at MeV energies.
4. Pair production. An electron-positron pair can be created when (and only when) a photon passes through the Coulomb field of a nucleus or atomic electron (needed for conservation of momentum). Pair production is the dominant photon interaction process at high energies (at and above a hundred MeV).

Mean free path for high-energy photons High-energy photons predominantly lose en-

ergy in matter by e^+e^- pair production. Similar to high-energy electrons, the radiation length is also a characteristic length for the penetration of high-energy photons in matter. The mean free path for pair production by a high-energy photon is:

$$\lambda = \frac{9}{7} X_0,$$

where X_0 is the radiation length. Therefore the photon intensity I at depth x can be expressed using the radiation length:

$$-\frac{dI(x)}{dx} = \frac{7}{9}\frac{I}{X0}.$$

Electromagnetic cascades As discussed in previous sections, high-energy electrons and positrons mainly lose energy through bremsstrahlung, while high-energy photon interactions are dominated by pair production. If the energy of an incident electron, positron, or photon is high enough (GeV), it initiates an electromagnetic cascade as pair production and bremsstrahlung generate more and more electrons and photons with lower energy. The cascade process stops when electron energy drops below the critical energy E_c and no more particles are generated in the shower. The rest of the energy is lost via ionization and excitation.

As discussed in the previous section, the radiation length X_0 is a characteristic length for the energy loss for both high-energy electrons and photons. It is also a central quantity for describing the longitudinal development of electromagnetic showers. It is therefore convenient to introduce the scale variable t:

$$t = x/X_0$$

Assuming that bremsstrahlung and pair production always take place after a radiation length X_0 and the energy is equally split between the two outgoing particles, the number of particles after t would be

$$N = 2^t,$$

and the energy per particle would be

$$E(t) = E_0/2^t,$$

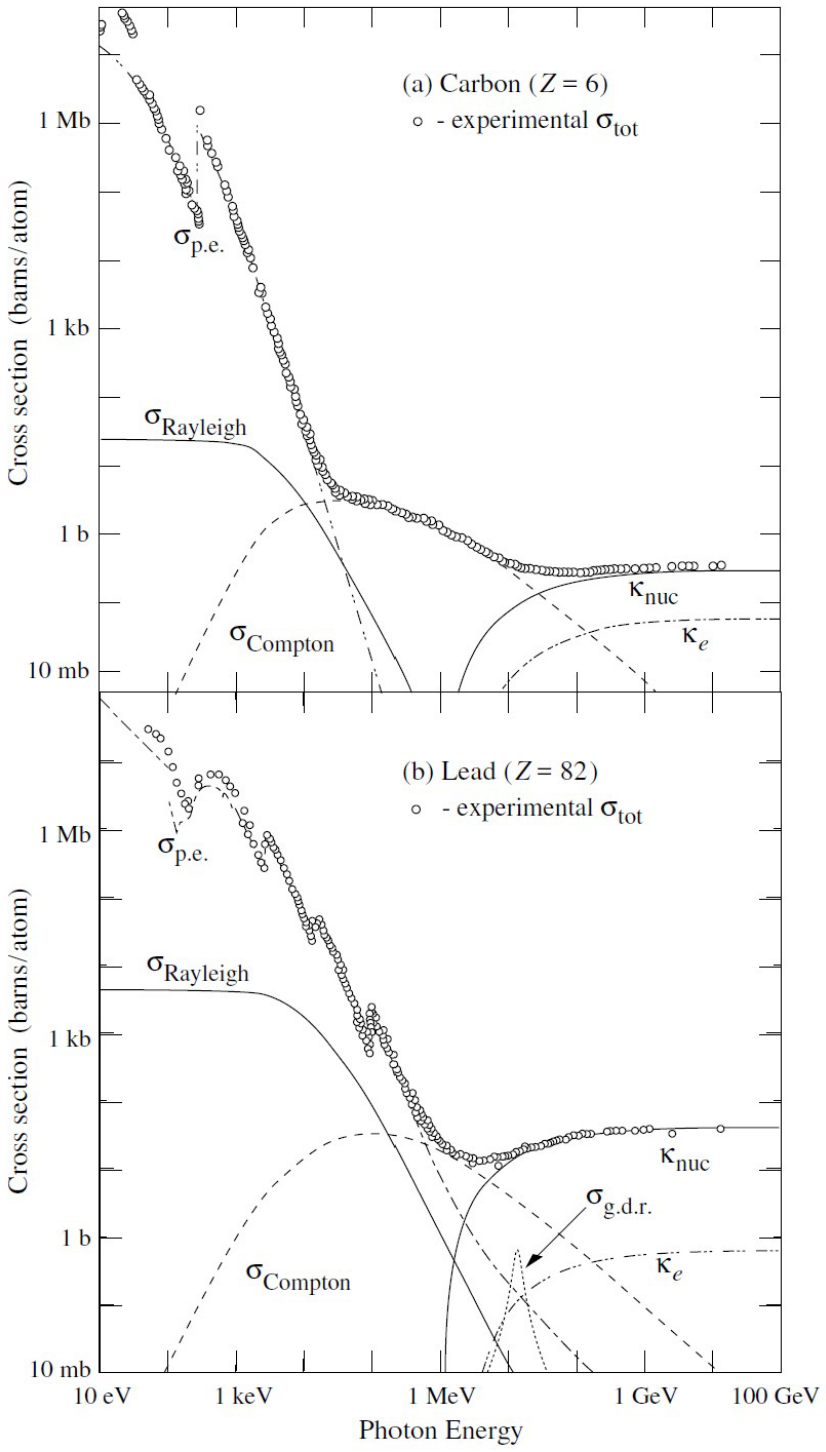


Figure 3: Contributions of different processes to total cross section of photon interaction with matter vs. photon energy in carbon and lead. $\sigma_{\text{p.e.}}$: atomic photoelectric effect, σ_{Rayleigh}: Rayleigh scattering, σ_{Compton}: Compton scattering, κ_{nuc}: pair production, nuclear field, κ_e: pair production, electron field, $\sigma_{\text{g.d.r.}}$: photonuclear interactions

where E_0 is the energy of the incident electron. If we assume the cascade process stops when the energy per particle drops to the critical energy E_c, then the depth of the maximum t_{max} would depend logarithmically on the initial energy:

$$t_{max} \propto \ln\left(\frac{E_0}{E_c}\right).$$

The transverse development of electromagnetic showers can be described with the Molière radius R_M:

$$R_M \approx \frac{21\,\text{MeV}}{E_c} X_0,$$

where E_c is in units of MeV.

2.5 HADRON INTERACTIONS IN MATTER

Hadrons traveling through matter experience elastic and inelastic interactions with nuclei through the strong force. The total hadron-nucleon interaction cross section is the sum of the elastic and inelastic contributions:

$$\sigma_{total} = \sigma_{elastic} + \sigma_{inelastic}.$$

Fig. 4 shows the total and elastic proton-proton cross sections. At high energy, σ_{total} is dominated by the inelastic part $\sigma_{inelastic}$. The inelastic cross section can be parameterized in terms of the atomic mass A and the hadron-proton cross section σ_0:

$$\sigma_{inelastic} = \sigma_0 A^{2/3},$$

therefore the mean free path λ is:

$$\lambda \propto \frac{A^{1/3}}{\rho},$$

where ρ is density of material and A is its atomic mass. The mean free path for inelastic processes is also known as interaction length. The energy of the incident hadron can be described as

$$-\frac{dE}{dx} = \frac{E}{\lambda}.$$

Hadronic showers Just like electrons and photons create electromagnetic showers via electromagnetic interactions, hadrons create hadronic showers via strong interactions. Similar to the radiation length X_0 for electromagnetic showers, the interaction length λ is a characteristic length for describing the development of hadronic showers. Interaction length is generally longer than the radiation length, and hadronic showers are therefore spatially longer than electromagnetic showers. This is why hadronic calorimeters are usually larger than electromagnetic ones.

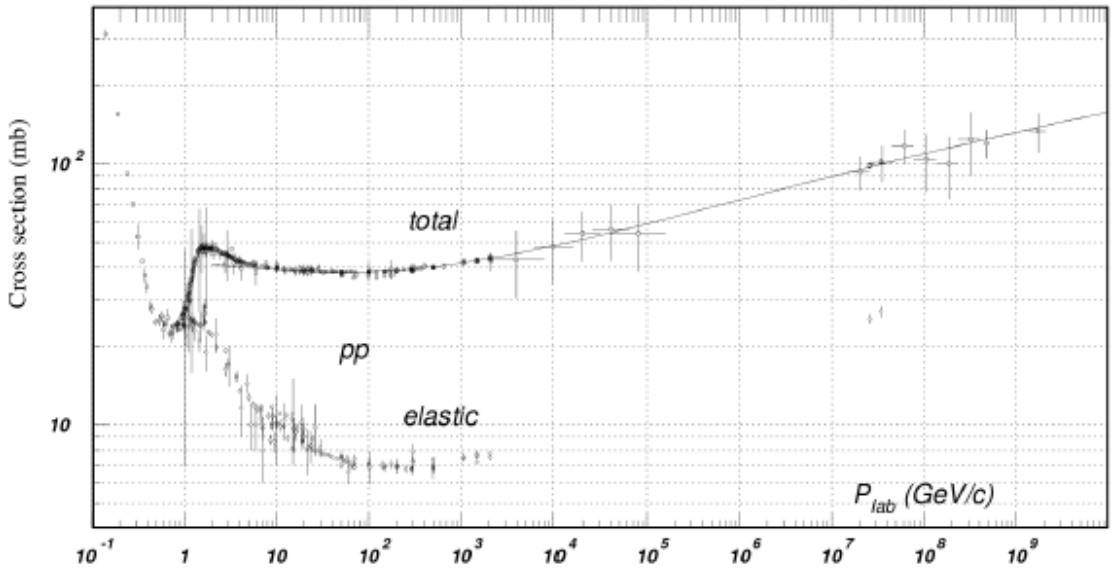


Figure 4: Total and elastic proton-proton cross sections.

2.6 CHERENKOV RADIATION AND TRANSITION RADIATION

When a charged particle is traversing a medium where its speed exceeds the speed of light in that medium, similar to shock waves created in the air by an object traveling faster than the speed of sound in the air, such particle radiates Cherenkov radiation. When a charged particle crosses the interface between materials with different optical properties, it emits transition radiation. Energy losses associated with these types of radiation are small, but they are both widely used in high-energy particle physics detectors. To further differentiate between the two forms of electromagnetic radiation briefly discussed in this

section, Cherenkov radiation is emitted when a relativistic charged particle passes through a homogeneous dielectric medium, transition radiation is emitted when a charged particle traverses an inhomogeneous medium. In practice, an interface between two homogeneous dielectric media is usually used to make charged particles emit transition radiation.

The particle's threshold speed for Cherenkov radiation is c/n, where n is the index of refraction in the traversed medium. Cherenkov radiation is emitted in a cone of angle θ_c, where

$$\cos\theta_c = c/n$$

w.r.t. the direction of the particle's velocity. The number of Cherenkov photons produced by a charged particle of charge ze per unit path length and wavelength is

$$\frac{d^2N}{d\lambda\, dx} = \frac{2\pi\alpha z^2}{\lambda^2}\left(1 - \frac{1}{\beta^2 n^2(\lambda)}\right) = \frac{2\pi\alpha z^2}{\lambda^2}\sin^2\theta_c,$$

where α is fine structure constant and λ is the wavelength.

Transition radiation is another form of electromagnetic radiation, and it is produced when a charged particle crosses the boundary between two media with different indices of refraction. This effect can be explained by boundary conditions for the electric field at the interface between two dielectric media. The total energy emitted when a particle with charge ze crosses the boundary between vacuum and a medium with plasma frequency ω_p is

$$I = \alpha z^2 \gamma \hbar \omega_p / 3.$$

The number of emitted photons is usually small, e.g., for an electron with the relativistic factor $\gamma = 2 \times 10^3$ only 0.8 X-ray photons are produced, on average, in a typical microstrip gas chamber. This is why detectors of transition radiation usually have many layers.

3.0 THE BELLE DETECTOR

This chapter provides a summary of the principles of operation of various particle detection technologies as most relevant for the Belle detector [30] which operated at the interaction region of the KEKB e^+e^- collider [31] at the national laboratory KEK near Tsukuba in Japan between 2000 and 2010. The Belle detector and KEKB collider (one of the two original B-factories) were built to study CP-violating phenomena, however, over the past 20 years, Belle data have been also successfully used to study a large number of diverse particle physics phenomena [32].

KEKB is an asymmetric e^+e^- collider, with electrons having the energy of 8.0 GeV and positrons having the energy of 3.5 GeV, giving 10.58 GeV center-of-mass energy, which is equal to the mass of the $\Upsilon(4S)$ meson, the short-lived $J^{PC} = 1^{--}$ bound state of the $b\bar{b}$ quark-antiquark pair of lowest invariant mass right above the production threshold for B meson pairs. As the result of an e^+e^- collision, besides the elastic scattering, the electron and positron could annihilate into an $\Upsilon(4S)$ meson or pairs of photons, muons, τ leptons, lighter quarks, i.e., $q\bar{q}$ ($q \in u, d, s, c$). A variety of radiative processes are also possible, a prominent example being a radiative return to lower invariant mass $J^{PC} = 1^{--}$ states after initial state radiation (ISR). Many of the particles born in the process of electron-positron annihilation are short-lived, so they decay almost instantaneously into photons, lighter leptons, and hadrons. The Belle detector is used to detect and reconstruct these particles as they traverse the detector and interact with its passive and active components. The former components are often used in tandem with the latter components, as additional interactions the original particles undergo in passive materials produce additional particles, often particle showers. The response of the detector is recorded using the active components. The digitized response of the detector is stored in the form of "hits", which are later used to

reconstruct the original particles and to perform physics analyses. Generally, we can observe only the most stable end products of the collision – the final state particles. These are, primarily, electrons[1], muons, photons, pions, charged kaons, protons and K_L^0. The original decay process is reconstructed based on the measurements performed using these final state particles.

A schematic side view of the Belle detector is shown in Fig. 5.

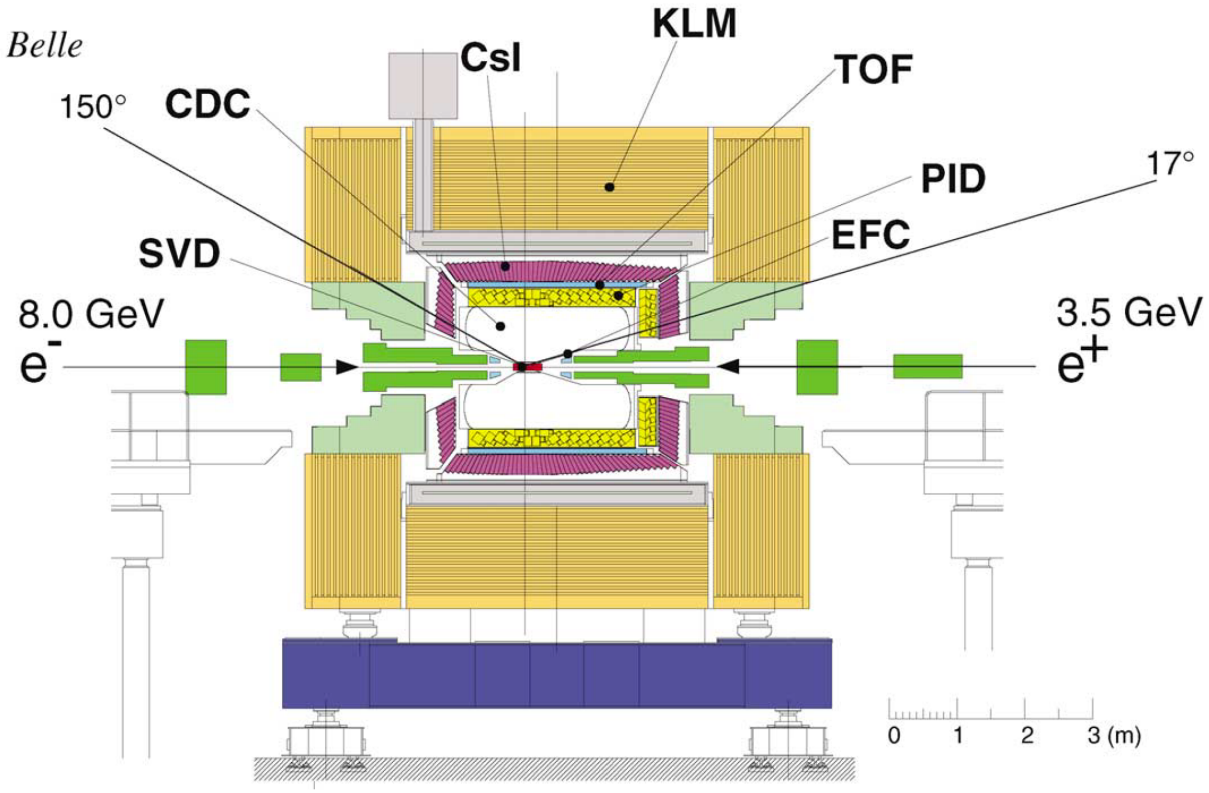


Figure 5: Side view of the Belle detector.

3.1 SILICON VERTEX DETECTOR (SVD)

Silicon vertex detector is designed to precisely measure decay vertices of short-lived B mesons (whose lifetimes are of the order of 1.5 ps) produced in $\Upsilon(4S)$ decay. The primary goal of the Belle experiment is to study time-dependent CP-violation in the decays of B

[1]Charge conjugate states, i.e., antiparticles such as the positrons, are implicitly included in all discussions throughout this dissertation, unless specifically mentioned.

mesons. The Lorentz boost of KEKB ($(\beta\gamma)_{\Upsilon(4S)} = 0.425$) results in an average B mesons z-vertex position difference of about 200 μm. SVD is designed to measure the z-vertex position difference with a precision of 100 μm therefore allowing for the necessary proper time resolution of the order of 1 ps between decay times of the two B mesons. Information from SVD is used to reconstruct the trajectories (tracks) of final state particles. By providing a stronger constraint on their parameters close to the interaction point, a significantly more precise vertex reconstruction becomes possible than without using SVD information. SVD is also useful for identifying and measuring the decay vertices of D mesons and τ leptons.

SVD is divided into layers of different radii around the beam pipe. Each layer is made of ladders consisting of double-sided silicon strip detectors (DSSD). The DSSDs provide depleted $p-n$ junctions. Fig. 6 shows the principles of operation of DSSD. Charged particles passing the junctions create electron and hole pairs along their trajectories. The created electrons and holes drift to the n^+ and p^+ strips on the surface of the DSSDs. The n^+ strips are aligned perpendicular to the beam axis and the p^+ strips are aligned along the beam axis. Therefore they can provide measurements of charged tracks in the $r-\phi$ and z direction.

There were two versions of SVD used on Belle during its data taking. SVD1 operated from 1999 to 2003 and SVD2 operated from 2003 to 2010, when the data taking stopped. SVD1 has 3 layers located at different radii around the beam pipe, while SVD2 has 4 layers.

SVD1 has three layers at radii 30, 45.5 and 60.5 mm and covers a solid angle $23° < \theta < 139°$. In total 102 DSSDs are used, each DSSD consists of 1280 sense strips and 640 readout pads on opposite sides. The z-strip pitch is 42 μm and the ϕ-strip pitch is 25 μm. The overall DSSD size is $57.5 \times 33.5\,\mathrm{mm}^2$. Fig. 7 shows the schematic view of SVD1.

SVD2 has four layers at radii 20, 43.5, 70, and 80 mm and covers a solid angle $17° < \theta < 150°$. In total 246 DSSDs are used. The DSSD size of the first three layers is $79.6 \times 28.4\,\mathrm{mm}^2$ while the size of the fourth layer is $76.4 \times 34.9\,\mathrm{mm}^2$.

3.2 CENTRAL DRIFT CHAMBER (CDC)

Central drift chamber (CDC) is designed to:

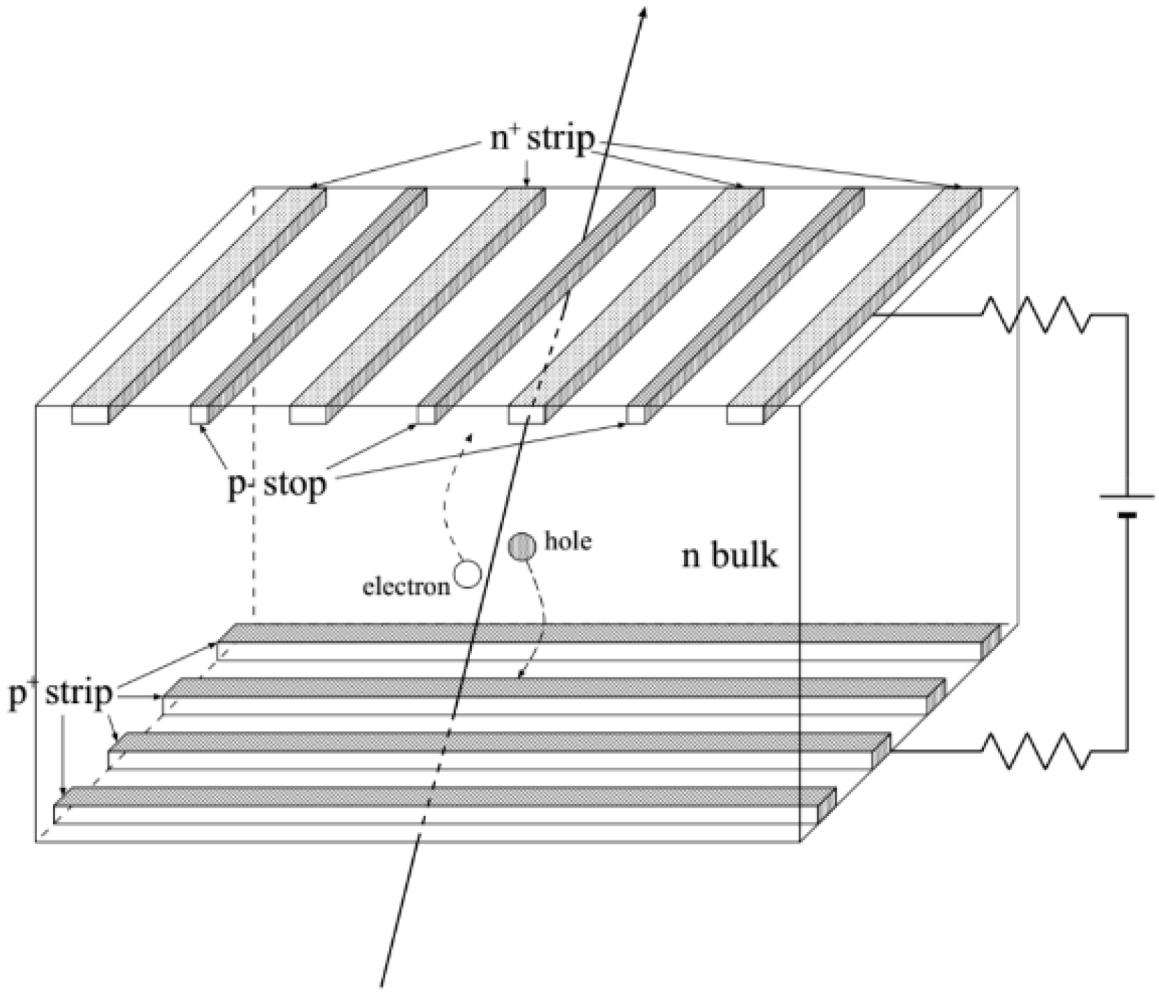


Figure 6: Principles of operation of DSSD.

1. reconstruct charged particles' tracks and precisely measure their momenta. The physics goals of the Belle experiment require a momentum resolution of $\sigma_{p_t}/p_t \sim 0.5\%\sqrt{1+p_t^2}$ (p_t in units of GeV/c) for all charged particles with $p_t \geq 100$ MeV/c in the range of polar angles of $17^\circ < \theta < 150^\circ$.

2. measure the ionizing energy losses dE/dx of the incident charged particles contributing to particle identification. In particular, low-momenta charged particles, which do not reach the PID, can only be identified by CDC.

3. provide efficient and reliable trigger decisions based on information about charged particles traversing the detector.

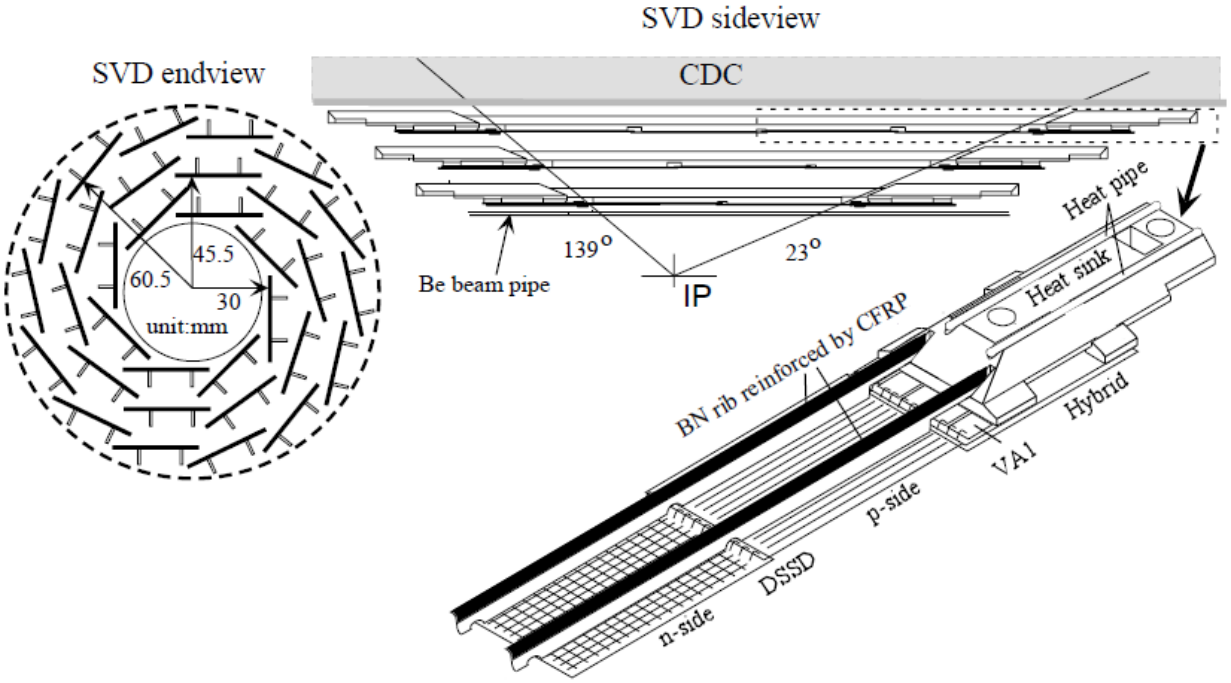


Figure 7: The schematic view of SVD1.

Charged particles passing the CDC ionize the gas along their trajectories. Electrons and ions created in ionization drift towards anode and cathode wires, respectively. Close to the wires, due to the strong gradient of the electric field, an avalanche develops therefore amplifying the electric pulse detected by the sense wires. The axial layers provide a position measurement in the $r - \phi$ plane for measurement of the transverse momentum, while the small-angle stereo layers provide additional z-position information.

The ionization energy losses dE/dx by the charged particles in the CDC are determined from the signal amplitudes. As discussed in the previous section, the dE/dx distributions follow the Bethe-Bloch equation and depend on the velocity and mass of the incident charged particles and, consequently, the dE/dx measurements allow to discriminate between different kinds of charged particles. The dE/dx measurements are included in the particle identification and are of particular importance for the $K^{\pm}/\pi^{\pm}$ separation which allows to distinguish between $b \to u$ and $b \to d$ transitions on a statistical basis.

Fig. 8 shows dE/dx distributions vs. momentum for different species of charged particles. As we can see, charged particles of the same momenta but different masses experience

different ionization energy losses dE/dx, making it possible to discriminate between different particle species.

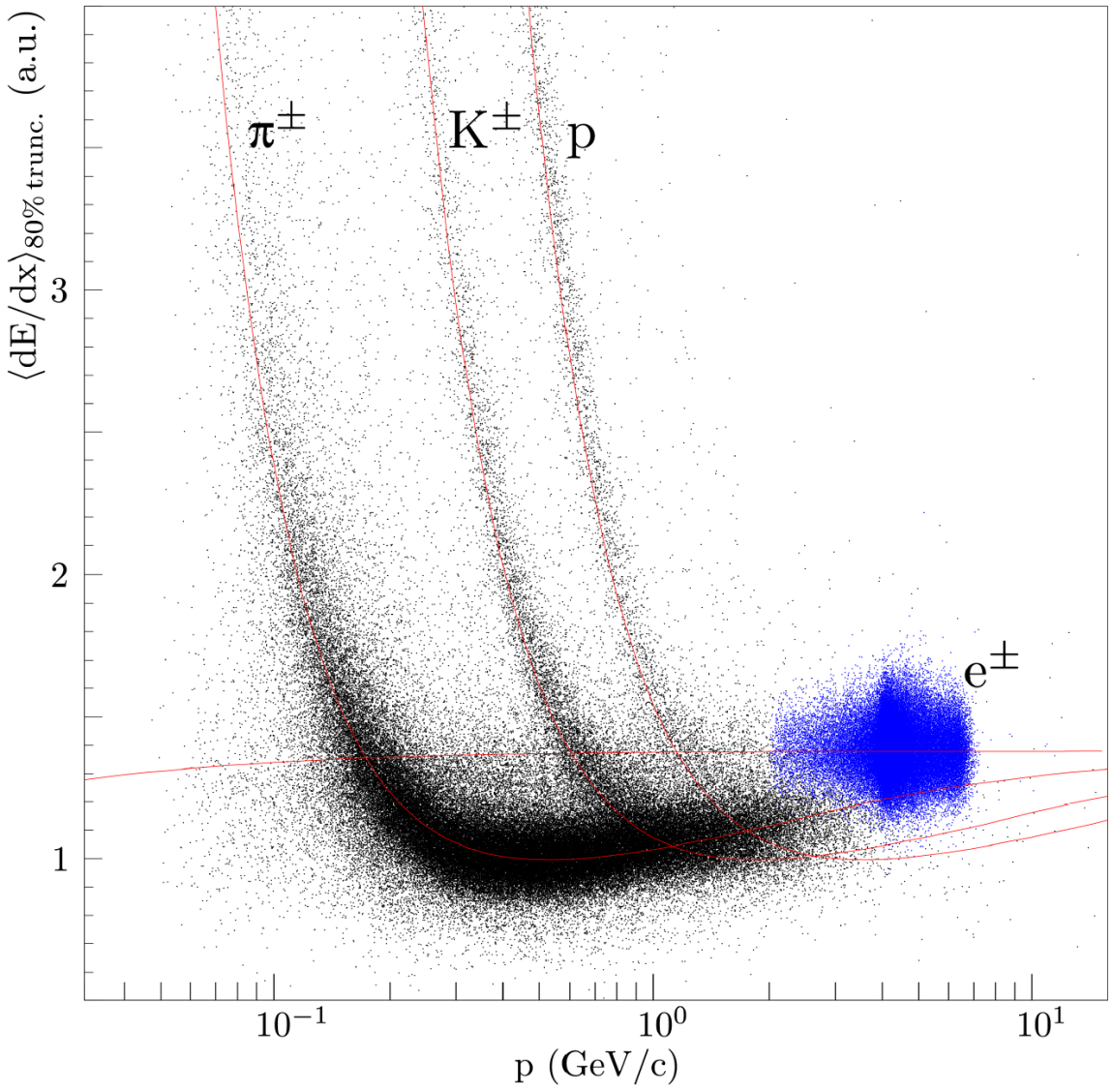


Figure 8: Truncated means of the dE/dx measurements performed using CDC vs. momentum for different species of charged particles. Distributions are from collision data recorded by Belle.

The structure of CDC is shown in Fig. 10. CDC is 2400 mm long and covers the polar angle of $17° < \theta < 150°$. The inner and outer radii are 83 mm and 888 mm. The chamber has 50 cylindrical layers, each containing between three and six either axial or small-angle stereo layers, and three cathode strip layers. CDC has a total of 8400 drift cells, each drift

cell is nearly square and is surrounded by field wires along the beam axis. Fig. 9 shows the cell structure and wire configuration.

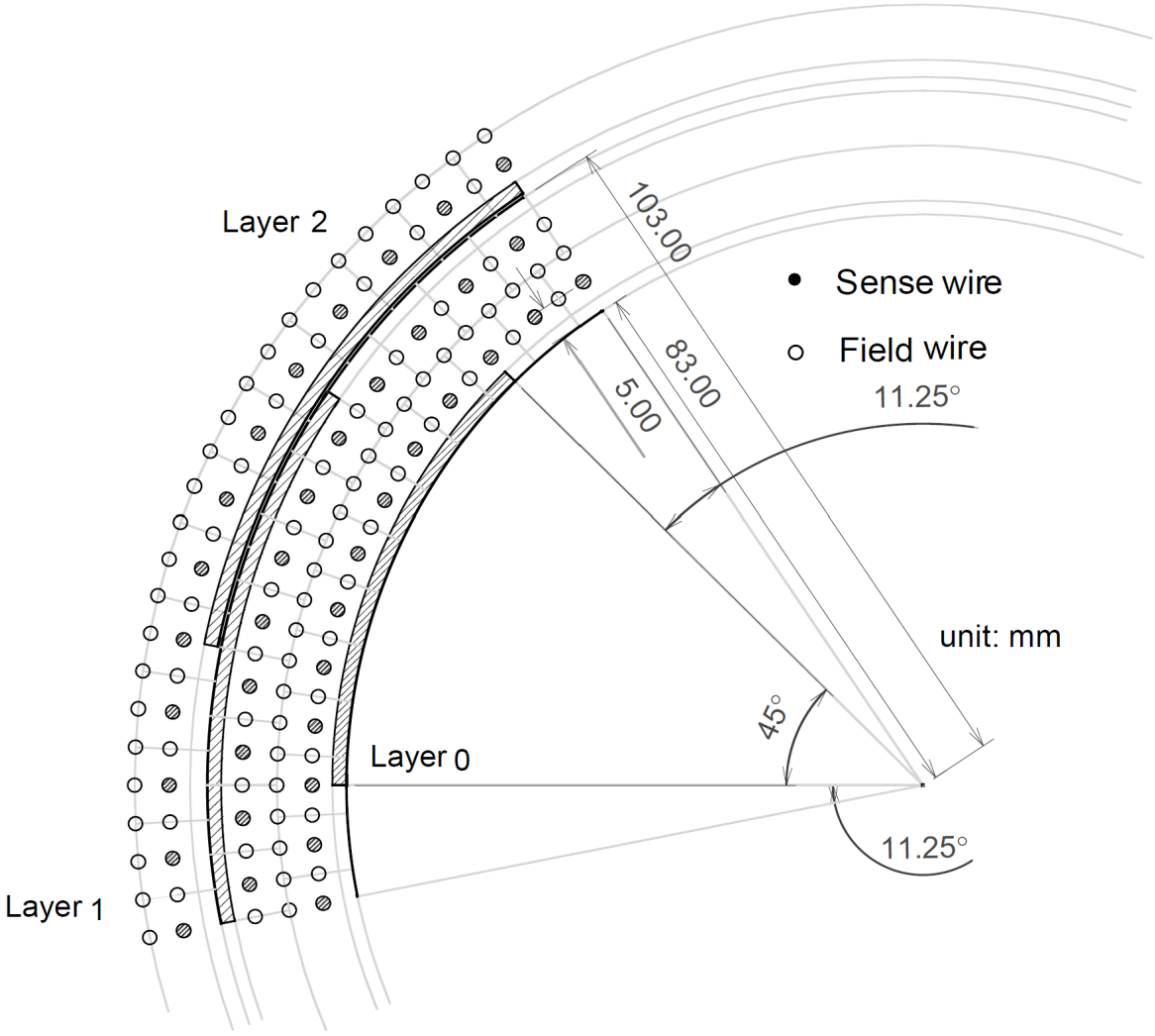


Figure 9: Cell structure and the cathode sector configuration.

The CDC is filled with 50% helium–50% ethane gas mixture. Pure helium has a large radiation length, therefore it reduces the effect of multiple Coulomb scattering, which is especially important for lower-momenta particles, as their trajectories are especially affected by multiple scattering. Adding ethane, a saturated hydrocarbon, decreases the radiation length, but increases the ionization, decreases diffusion, and gives a faster drift speed. This mixture has a long radiation length (640 m) and has a smaller photo-electric cross section than argon-based gas mixtures which reduces background from synchrotron radiation, therefore improving efficiency and resolution.

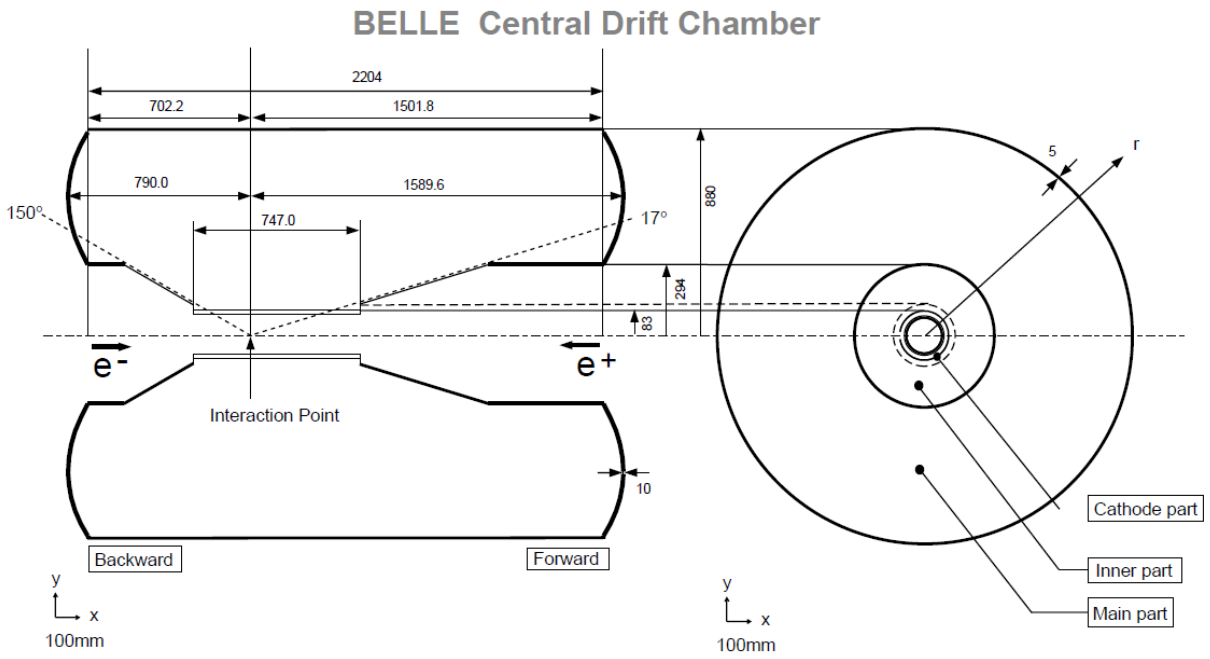


Figure 10: The schematic view of CDC.

3.3 AEROGEL CHERENKOV COUNTER (ACC)

The system of Aerogel Cherenkov counters (ACC) is designed to provide information for particle identification, in particular to separate kaons from pions. It provides particle identification information in a momentum range $1.2\,\mathrm{GeV} < p < 3.5\,\mathrm{GeV}$ and thus is complementary to the dE/dx measurement in the CDC and the time-of-flight measurement by the TOF (described later in this document).

Charged particles passing a medium with a velocity larger than the speed of light in the medium radiate Cherenkov radiation. For a particle with mass m, momentum p and velocity β, Cherenkov light is emitted if

$$n > 1/\beta = \sqrt{1 + (m/p)^2}.$$

The material of the medium is chosen such, that $\pi^{\pm}$ mesons with momenta larger than 1 GeV/c generate Cherenkov radiation, but the heavier K mesons with the same momenta are below the threshold velocity and do not generate Cherenkov radiation. The ACC is a

threshold counter and does not image the cone of the Cherenkov light, which would have provided further information as its opening angle depends on the particle's velocity.

As shown in Fig. 11, the ACC is divided into barrel and forward endcap parts and covers a polar angle of $17° < \theta < 127°$. The barrel ACC consists of 960 counter modules segmented into 60 cells in the ϕ-direction. The endcap ACC consists of 228 counter modules arranged in 5 concentric layers. All of the counter modules point to the interaction region.

The counter modules consist of layers of silica aerogels arranged in aluminum boxes of $12 \times 12 \times 12$ cm^3 in size. The refractive index of the silica aerogels ranges from 1.01 to 1.03 and depends on the polar angle. The Cherenkov light is detected by fine-mesh photomultipliers (PMTs) attached to the boxes.

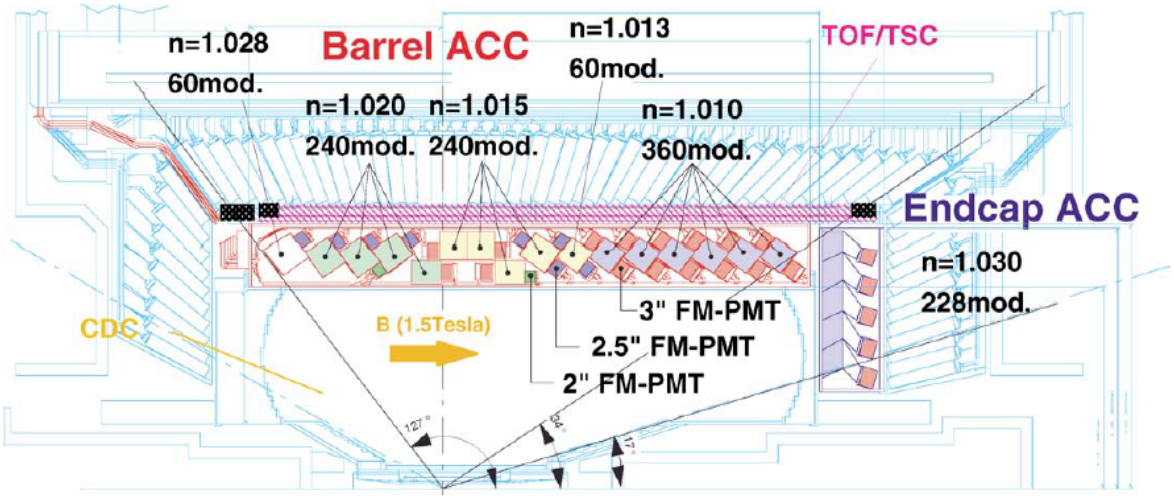


Figure 11: The schematic view of ACC.

3.4 TIME-OF-FLIGHT COUNTER (TOF)

The system of time-of-flight (TOF) counters (a.k.a. modules, when assembled with PMTs) is designed to measure the time that a particle needs to travel from the interaction point to the TOF module. TOF is sensitive to charged particles with momentum less than 1.2 GeV/c and is thus complementary to the ACC. TOF also provides fast timing signals for the trigger system due to its good time resolution of approximately 100 ps.

The mass m of a particle is related to the measured time-of-flight T by

$$m = \frac{p}{c}\sqrt{\left(\frac{cT}{L}\right)^2 - 1},$$

where p is momentum measured by the CDC and SVD and L denotes the helical distance traveled by the particle from the interaction point to the TOF module. Mass distribution obtained from TOF measurements contributes to particle identification. Fig. 12 shows the predictions of Monte Carlo simulation for the invariant mass distribution from TOF measurements for particles of momenta below 1.2 GeV/c. As we can see, TOF counters allow to discriminate among $\pi^{\pm}$, $K^{\pm}$, and proton candidates.

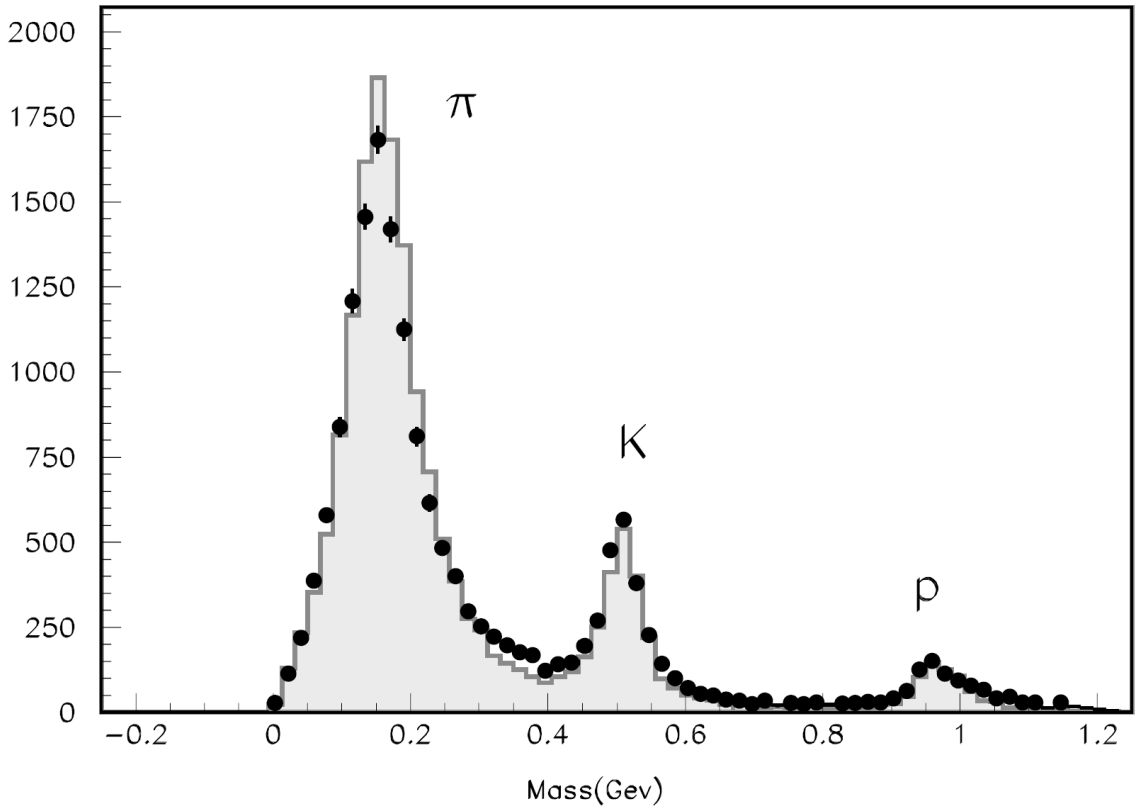


Figure 12: The invariant mass distribution from TOF measurements for particles of momenta below 1.2 GeV/c in MC simulation.

The TOF system consists of 64 modules placed in the barrel of the detector and covers a polar angle of $34° < \theta < 120°$. Each TOF module consists of a plastic scintillation counter attached to a PMT. Charged particles traversing the plastic scintillators produce photons which are then detected by the PMTs.

3.5 ELECTROMAGNETIC CALORIMETER (ECL)

Electromagnetic calorimeter (ECL) is designed to measure the energy and position of electromagnetic showers. ECL provides an excellent energy resolution in a wide range of energies between 20 MeV/c^2 and 8 GeV/c^2. This range includes the low-energy photons originating from the cascades and high-energy photons coming directly from B decays. Luminosity measurements and calibrations are performed using ECL measurements for $e^+e^- \to e^+e^-$ and $e^+e^- \to \gamma\gamma$ processes.

ECL is also used to contribute to particle identification. Electron identification in Belle relies primarily on a comparison of the charged particle momentum and the energy deposits in the ECL. The ratio of deposited energy to momentum, E/p, is close to 1.0 for electrons, while it is smaller for other charged particles. Also, the shapes of showers differ between different species of particles.

High-energy electrons and positrons mainly lose energy through bremsstrahlung, while high-energy photon interactions are dominated by pair production. If the energy of an incident electron, positron, or photon is high enough (GeV), an electromagnetic cascade is initiated via pair production or bremsstrahlung which then generate more and more electrons and photons of lower energy.

The radiation length X_0 is a characteristic length for the energy loss by both high-energy electrons and photons, it is also a central quantity for describing the longitudinal development of electromagnetic showers. The depth of maximum t_{max} of the shower in the unit of radiation length X_0 depends logarithmically on the initial energy:

$$t_{max} \propto \ln\left(\frac{E_0}{E_c}\right),$$

where E_c is the critical energy. ECL needs to have sufficient radiation length X_0 to absorb the particle energy. A detailed discussion can be found in previous sections.

Fig. 13 shows the schematic view of ECL. ECL consists of 8376 scintillating cesium-iodide crystals doped with thallium, CsI(Tl), and covers a polar angle of $12° < \theta < 155°$. Barrel ECL contains 6624 crystals, and forward and backward endcaps contain 1152 and 960 crystals, respectively. CsI(TI) crystals in the barrel ECL have forward and backward areas

of 5.5 cm×5.5 cm and 6.5 cm×6.5 cm, respectively. The 30 cm crystal length, which equals 16.2 radiation length X_0, is chosen to ensure almost 100% absorption of electromagnetic energy.

ECL has an energy resolution of

$$\frac{\sigma_E}{E} = (1.34) \oplus \frac{0.066}{E} \oplus \frac{0.81}{E^{1/4}}\%,$$

where the energy E is measured in units of GeV. The energy resolution is limited by electronic noise, parameterized by the first term, and by shower leakage fluctuations, parameterized by the second and third terms.

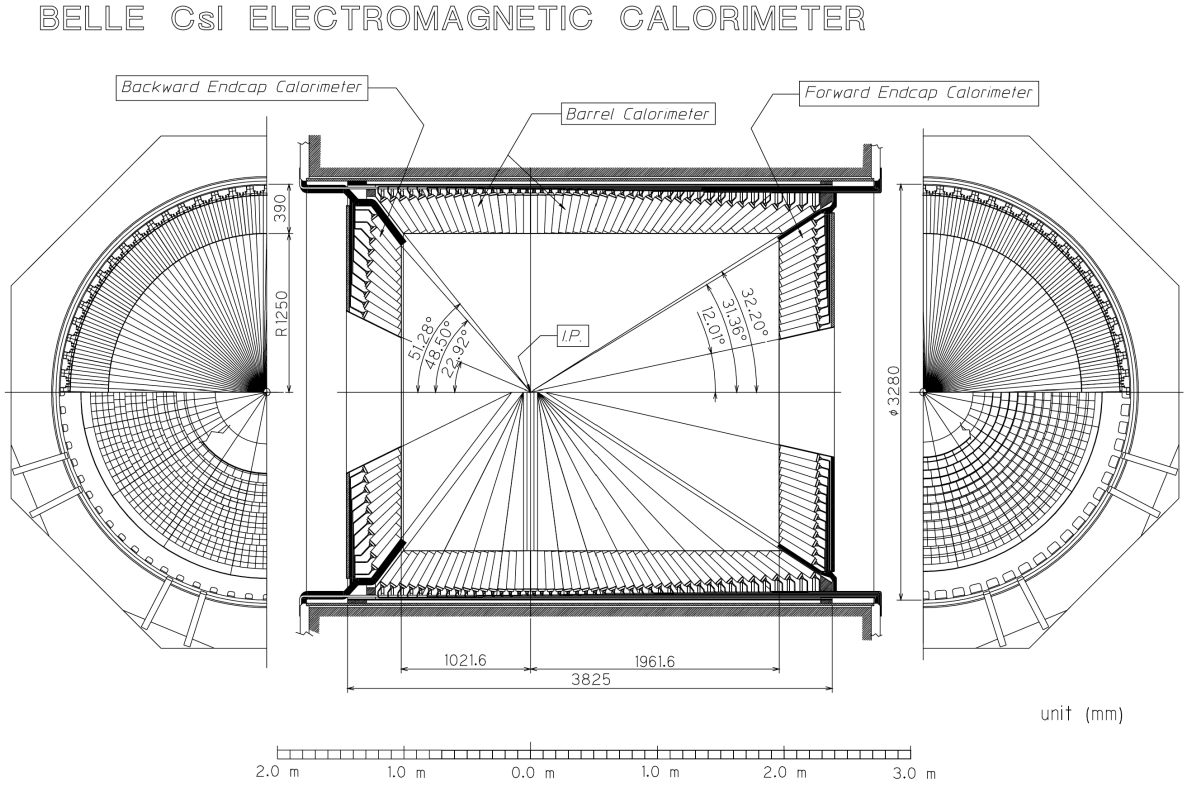


Figure 13: The schematic view of ECL.

3.6 K_L^0 AND MUON DETECTION SYSTEM (KLM)

K_L^0 and muon detection system (KLM) is designed to identify muons and long-lived neutral kaons.

The Belle detector has a homogeneous electromagnetic calorimeter (ECL) but no hadron calorimeter (HCAL) because the center of mass energy of the experiment is too low to reach a reasonable energy resolution for hadrons with an HCAL. KLM tries to compensate for the missing hadron calorimeter. It can not measure the energy but it can indicate the presence of a hadron by the shower it creates in the detector. In combination with the tracking detectors, it can also determine if the particle is charged or not by matching the shower with the track's projection.

K_L^0 creates a shower in KLM via strong interactions just like electrons and photons create electromagnetic showers via electromagnetic interactions. K_L^0 passing the iron plates of KLM produces showers of ionizing particles. KLM measures the location of the shower and thus the flight direction of K_L^0 can be determined. KLM can not determine the energy of K_L^0 due to the large fluctuations in the size of the shower. Since K_L^0 is neutral, it leaves no hit along its path in SVD, CDC, and TOP. Clusters of hits found in KLM or ECL without a matching CDC/SVD track are assumed to be due to a K_L^0 candidate, though, in principle, could be due to some other relatively long-lived neutral hadron, such as neutron.

KLM is also used for particle identification of weakly interacting muons. Muons leave long clean tracks in KLM that match with extrapolated CDC/SVD tracks. With sufficient momentum, muons pass all other detector components and penetrate several layers of the KLM. Reconstructed tracks from the CDC that can be matched to a series of hits in the KLM are assumed to be due to muons, which could, in principle, also come from in-flight decays of pions or charged kaons.

KLM is composed of alternating layers of iron plates (passive detector material, a.k.a. the radiator, where showers are initiated) and superlayers of resistive plate counters (RPC) that detect the charged particles (including particles produced in the showers). The barrel part of KLM contains 14 iron layers and 15 RPC superlayers and covers a polar angle of $20^\circ < \theta < 155^\circ$.

As shown in Fig. 14, the superlayers are composed of two glass-electrode RPC modules with high bulk resistivity separated by a gas-filled gap. A charged particle passing the gap initiates a streamer in the gas. The streamer in the gas causes a local discharge of the plates, that induces a signal on the external readout strips along the θ and ϕ directions.

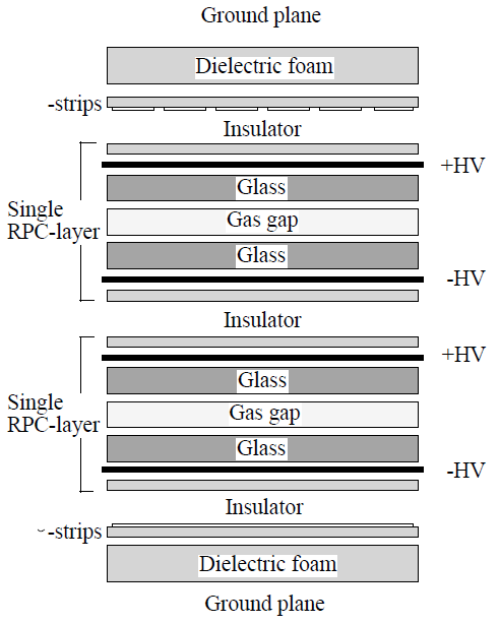


Figure 14: Cross section of a KLM superlayer.

3.7 EXTREME FORWARD CALORIMETER (EFC)

Extreme forward calorimeter (EFC) is designed to extend the coverage of the polar angle concerning ECL. EFC detects electrons and photons in the extreme forward region $6.4° < \theta < 11.5°$ and in the backward region $163.3° < \theta < 171.2°$. EFC can also mask the beam-related background for CDC and be used as a beam luminosity monitor by KEKB.

Since EFC is exposed to very high levels of radiation, it is made from radiation-hard bismuth germanium oxide crystals (BGO), and the scintillation light is collected by photo-diodes.

3.8 TRIGGER AND DATA ACQUISITION SYSTEM

The trigger system is used to decide whether an event is of interest for physics analysis and has to be recorded by the data acquisition system (DAQ) for analyses at a later stage, or should be discarded. Events of interest are hadronic $\Upsilon(4S)$ decays, $q\bar{q}$ ($q \in u, d, s, c$) continuum events, two-photon processes ($e^+e^- \to e^+e^-X$, where X is produced, to lower order

in QED, by two space-like photons emitted by the electron and positron), $e^+e^- \to \tau^+\tau^-$, Bhabha scattering, diphoton ($e^+e^- \to \gamma\gamma$), and other processes. Discarded are background events from synchrotron radiation, interactions of the beam with residual gas in the beam pipe or events due to cosmic rays ("cosmics"). Generally, events of low final state particle multiplicities used primarily for calibration purposes are prescaled (or suppressed), so only a small fraction of such events are recorded.

The trigger system includes 3 stages: the online hardware Level-1 trigger, the online software Level-3 trigger, and the offline software Level-4 trigger. This nomenclature is such for historical reasons, which explains why there is no Level-2 trigger on Belle. The Level-1 trigger system is shown in Fig. 15. It consists of the sub-detector trigger systems and the central trigger system called the Global Decision Logic (GDL). CDC provides $r-\phi$ and $r-z$ track trigger signals. TOF provides timing information and measurements of the topology and multiplicity of hits by charged tracks. ECL provides trigger-based energy deposit and energy cluster counting. KLM provides information about muon-based triggering. ECL and EFC together trigger on Bhabha and diphoton events. GDL processes trigger signals from sub-detectors within 1.85 μs after the collision and provides a final trigger decision 2.2 μs after the collision.

Once DAQ receives Level-1 trigger decisions from GDL, it reads out the digitized information from all sub-detectors.

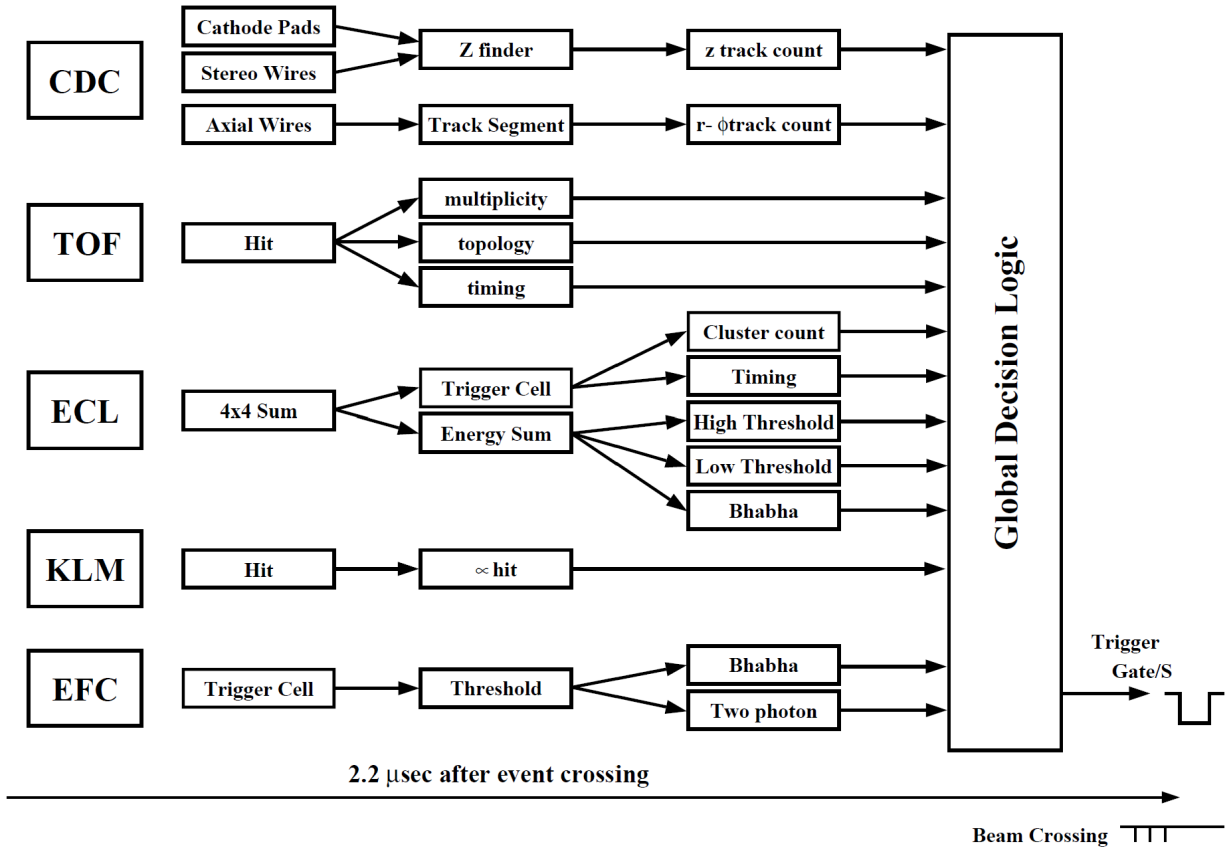


Figure 15: The Level-1 trigger system of the Belle detector.

4.0 SEARCH FOR BARYON-NUMBER-VIOLATING PROCESSES IN B^- DECAY TO FINAL STATE $\overline{\Xi}{}^0_C\overline{\Lambda}{}^-_C$ AT BELLE

4.1 INTRODUCTION

The main goal of this analysis is to search for baryon-number-violating processes that involve Ξ^0_c. We search for the decay $B^- \to \overline{\Xi}{}^0_c\overline{\Lambda}{}^-_c$ forbidden in the Standard Model (SM) and interpret the results also in terms of the previously-discovered [33] SM decay $B^- \to \Xi^0_c\overline{\Lambda}{}^-_c$ followed by the beyond-the-SM (BSM) $\Xi^0_c - \overline{\Xi}{}^0_c$ oscillations in the light of baryogenesis model [21] discussed in Chapter 1. We also present our results as the ratio of the branching fractions for the SM-forbidden and SM-allowed channels which allows us to cancel many systematic uncertainties. We provide a detailed description of the analysis after presenting a brief recap of the Introduction to this dissertation.

4.1.1 Motivation

One of the greatest challenges in particle physics is to figure out why the universe is composed primarily of matter instead of equal amounts of matter and antimatter. This matter-antimatter asymmetry should have an explanation, generally dubbed baryogenesis, which is the motivation for a wide range of theories predicting BSM physics. Pioneer of the field, Andrei Sakharov, proposed three conditions for baryogenesis in 1966, which are baryon number violation (BNV), C and CP violation, and departure from thermal/chemical equilibrium [2]. BNV is a necessary condition to produce the predominance of baryons over antibaryons but, until now, there has been no direct experimental evidence for this phenomenon.

Many different processes can be used to search for BNV, *e.g.* decays of the τ lepton and B mesons [34, 35], proton decay [36], and baryon-antibaryon oscillations [37]. In this project, we focus on the search for BNV which, if observed, would most likely be due to charmed baryon-antibaryon oscillations [21]. An observation of baryon-antibaryon oscillations would further our understanding of baryogenesis and would be a pathway to understanding the matter-antimatter asymmetry of the universe.

LHCb recently searched for $\Xi_b^0 - \overline{\Xi}_b^0$ oscillations and set an upper limit [22] on the oscillation rate of $\omega < 0.08$ ps^{-1} at the 95% confidence level (CL). In our analysis we search for $\Xi_c^0 - \overline{\Xi}_c^0$ oscillations using the SM decay $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ (which has a relatively large branching fraction 9.5×10^{-4} [38]) followed by the oscillation of Ξ_c^0 into $\overline{\Xi}_c^0$. As charmed baryons have a relatively short lifetime (of the order of 0.3 ps), we are not able to resolve their decay vertices well with Belle. As a result, from the analysis perspective, the SM decay $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ followed by the oscillation of Ξ_c^0 into $\overline{\Xi}_c^0$ would look like the direct BNV decay $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$. In our analysis, we do not distinguish between the two BNV hypotheses and generally describe the analysis as the search for BNV using the *channel* $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$. We interpret our results using both hypotheses and, also, in terms of the ratio of branching fractions for the BNV and SM decays, where most systematic uncertainties cancel in the ratio.

4.1.2 Experimental Approach

In our analysis, we search for BNV in the channel $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$ using the entire Belle data sample collected at $\Upsilon(4S)$. We focus on the exclusive channel $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$. We may consider including other channels such as $B^- \to \overline{\Xi}_c^0 \overline{\Sigma}_c^- / \overline{\Xi}_c^{\prime 0} \overline{\Lambda}_c^- / \overline{\Xi}_c^{\prime 0} \overline{\Sigma}_c^-$ in the updated analysis in the future.

Fig. 16 shows the Feynman diagram for the SM-allowed decay $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ and other B decays where we could detect final state $\Xi_c^0 \overline{\Lambda}_c^-$ inclusively. Charge of $\overline{\Lambda}_c^-$ tags the flavor of B meson, so we know the flavor of the other charmed baryon born in the SM decay of B^-. Ξ_c^0 is expected, but it could oscillate into its antiparticle. Of course, the direct BNV decay could, in principle, happen also, however, if such final state is observed, it would more likely

be due to the SM decay followed by charmed baryon-antibaryon oscillations.

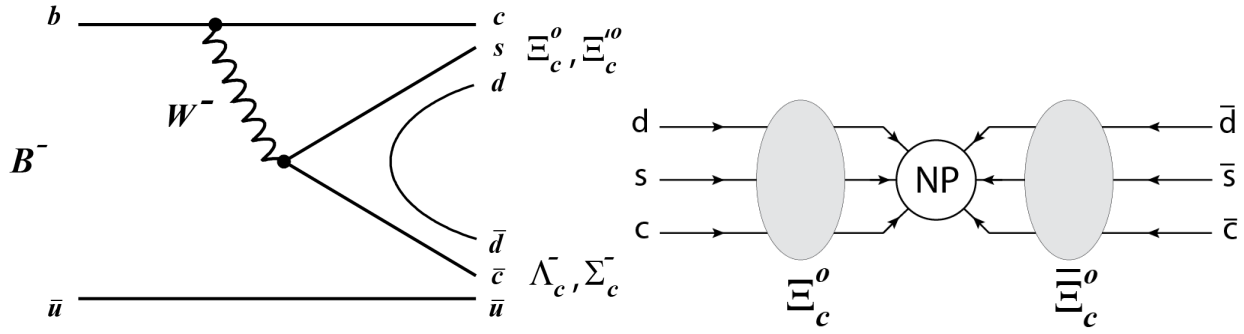


Figure 16: Left: feynman diagram for the previously-measured SM-allowed decay $B^- \to \Xi_c^0\overline{\Lambda}_c^-$ [33] and other B decays where we could detect final state $\Xi_c^0\overline{\Lambda}_c^-$ inclusively. Right: diagram for the $\Xi_c^0 - \overline{\Xi}_c^0$ oscillation.

To detect the SM decays we use three Ξ_c^0 decay channels ($\Xi_c^0 \to \Xi^-\pi^+/\Lambda K^-\pi^+/pK^-K^-\pi^+$) whose branching fractions were measured by Belle recently [33]. Two $\overline{\Lambda}_c^-$ decay channels ($\overline{\Lambda}_c^- \to \overline{p}K_S^0/\overline{p}K^+\pi^-$) are used to perform self-flavor-tagging. We show a summary of the resulting 6 decay channels of $B^- \to \Xi_c^0\overline{\Lambda}_c^-$ and their branching fractions [38] in Table 1. These channels allow us to calibrate our BSM analysis using previous Belle measurements.

	channel	Product of all relevant branching fractions(10^{-7})
1	$\Xi_c^0 \to \Xi^-\pi^+, \overline{\Lambda}_c^- \to \overline{p}K_S^0$	1.20
2	$\Xi_c^0 \to \Xi^-\pi^+, \overline{\Lambda}_c^- \to \overline{p}K^+\pi^-$	6.86
3	$\Xi_c^0 \to \Lambda K^-\pi^+, \overline{\Lambda}_c^- \to \overline{p}K_S^0$	0.78
4	$\Xi_c^0 \to \Lambda K^-\pi^+, \overline{\Lambda}_c^- \to \overline{p}K^+\pi^-$	4.46
5	$\Xi_c^0 \to pK^-K^-\pi^+, \overline{\Lambda}_c^- \to \overline{p}K_S^0$	0.61
6	$\Xi_c^0 \to pK^-K^-\pi^+, \overline{\Lambda}_c^- \to \overline{p}K^+\pi^-$	3.46

Table 1: Branching fractions for 6 decay channels of $B^- \to \Xi_c^0\overline{\Lambda}_c^-$ used in our analysis. For Ξ_c^0 decays, Belle measured branching fractions are used, see Table 23.

4.2 MC SIMULATION AND DATA SAMPLE

We use the entire 711 fb^{-1} of $\Upsilon(4S)$ on-resonance data. This dataset contains 772 million $B\bar{B}$ pairs collected with the Belle detector [30] at the asymmetric-energy e^+e^- collider KEKB [31] at $\sqrt{s} = 10.58$ GeV.

We generated 1,000,000 signal MC events (500,000 events for each of the SM and BNV modes) for each decay channel using MC generator EvtGen [39]. Signal MC sample was generated using toolkit mcproduzh [40], therefore the events are distributed over experiments proportionally to the statistics of actual Belle data. The $\Upsilon(4S) \to B^+B^-$ process is simulated with VSS model, and all other decays in the signal decay chain are simulated by PHSP model. FSR was simulated using the "afterburner" MC generator PHOTOS.

We use 6 streams of generic MC for $\Upsilon(4S) \to B\bar{B}$, $e^+e^- \to q\bar{q}$ to study background, where each "stream" is represented by a simulated sample of statistics approximately equivalent to the full Belle dataset.

4.3 SELECTION CRITERIA

4.3.1 Final-State Charged Particle Candidates Selection

1. Charged pion candidates are required to satisfy the binary PID requirements $R_{K,\pi} < 0.4$ and $R_{e,hadron} < 0.95$. Here $R_{K,\pi}$ is the K/π likelihood ratio $\frac{L_K}{L_K+L_\pi}$, where L_i is the likelihood for each particle hypothesis determined using information from different detector subsystems. $R_{e,hadron}$ is *electron/hadron* likelihood ratio.

2. The kaon candidates are required to satisfy $R_{K,\pi} > 0.4$ and $R_{e,hadron} < 0.95$.

3. The proton and antiproton candidates are required to satisfy $R_{p/\bar{p},K} > 0.6$ and $R_{p/\bar{p},\pi} > 0.6$, where $R_{p/\bar{p},K}$ and $R_{p/\bar{p},\pi}$ are p/K and p/π likelihood ratios.

4.3.2 Transverse Momentum Requirement for Final-State Charged Particle Candidates

Track reconstruction efficiency depends on the transverse momentum of the reconstructed particle. Low transverse momenta tracks are not necessarily reliably simulated, so their reconstruction could be associated with a large systematic uncertainty. Ideally, we would like to select final state particle candidates of transverse momenta in the region where systematic uncertainties are small and the efficiency is large. Doing so presents a challenge in our analysis.

Fig. 17 shows the transverse momentum ($p_\perp$) distributions and reconstruction efficiencies for final-state charged particles (pions and protons) for signal MC events in channel 1 (as explained in Table 1). The plots in red color show distributions of transverse momenta in MC truth. The plots in green color show transverse momentum distributions for reconstructed final-state charged particles. The plots in blue color show reconstruction efficiencies for final-state charged particles as functions of their transverse momenta. Fig. 18 shows $p_\perp$ distributions and reconstruction efficiencies with polar angle acceptance cut $17° - 150°$.

Since K^0_S and Λ candidates are selected from standard BASF banks, we apply the $p_\perp > 50$ MeV/c requirement to all kaon/pion/proton candidates in all channels except for daughter particles of Λ and K^0_S candidates. We remove candidates which $p_\perp < 50$ MeV/c.

The distributions of transverse momenta and reconstruction efficiencies for all channels are shown in the Appendix.

In our analysis, we use the SM decay as the control channel. This allows us, among other studies, to investigate the effects of $p_\perp$ cuts on the branching fraction measurement. We compare the branching fractions measured with 60% efficiency, 50 MeV/c and without $p_\perp$ cuts to decide if $p_\perp$ cuts are necessary. Anyway, the systematic effects should cancel in the ratios of our future BNV and SM results.

4.3.3 Selection of K^0_S and Λ Candidates

K^0_S candidates that decay into π^+ and π^- are selected from VEE2 bank with NISKSFINDER method [41, 42].

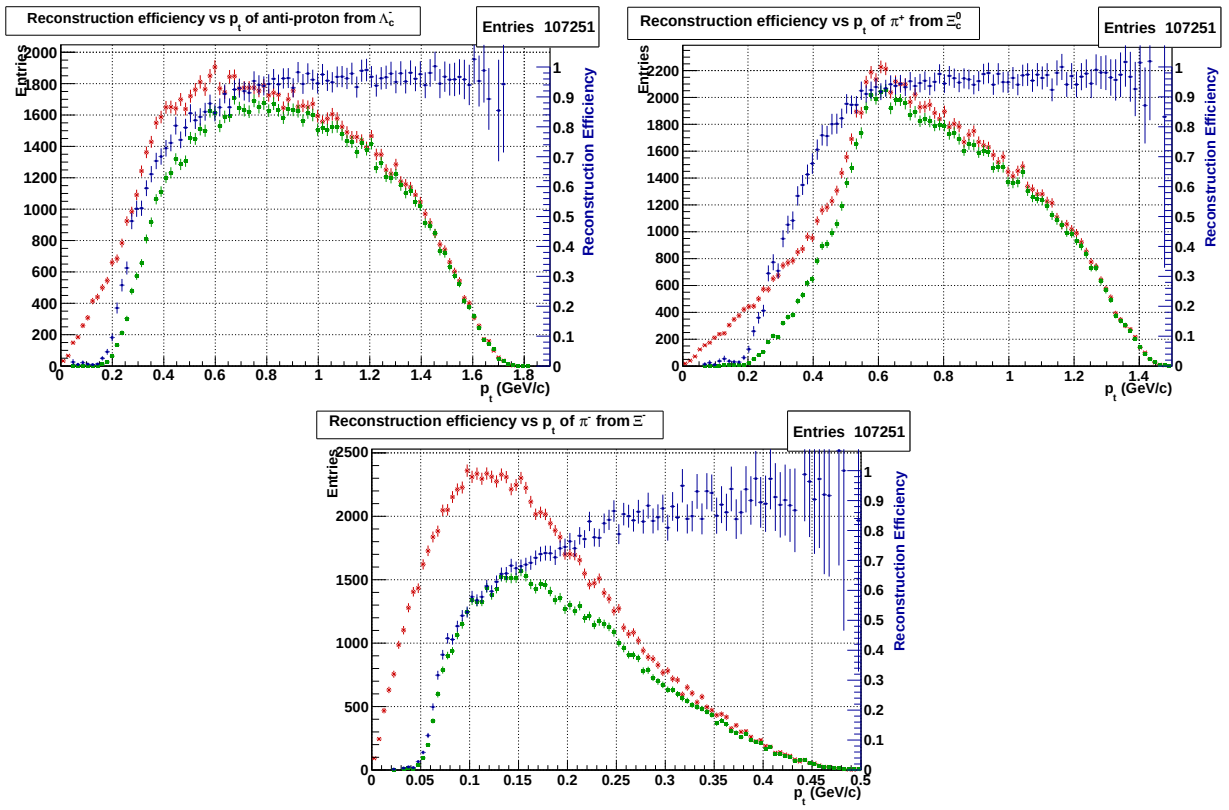


Figure 17: Reconstruction efficiency vs transverse momentum of final-state charged particles in signal MC for ch1. Generated transverse momentum distribution (*i.e.* MC truth, in red color). Reconstructed transverse momentum distributions (in green color). Reconstruction efficiency vs $p_\perp$ (in blue color).

Λ candidates that decay into p and π^- are selected from VEE2 bank with FINDLAMBDA package. Λ candidates of any quality from VEE2 bank are allowed in the analysis.

4.3.4 Reconstruction of Ξ^- Candidates

Ξ^- candidates are reconstructed in the decay channel

$$\Xi^- \to \Lambda\pi^-.$$

In order to improve the ΔE and invariant mass resolutions, the following reconstruction procedure is followed:

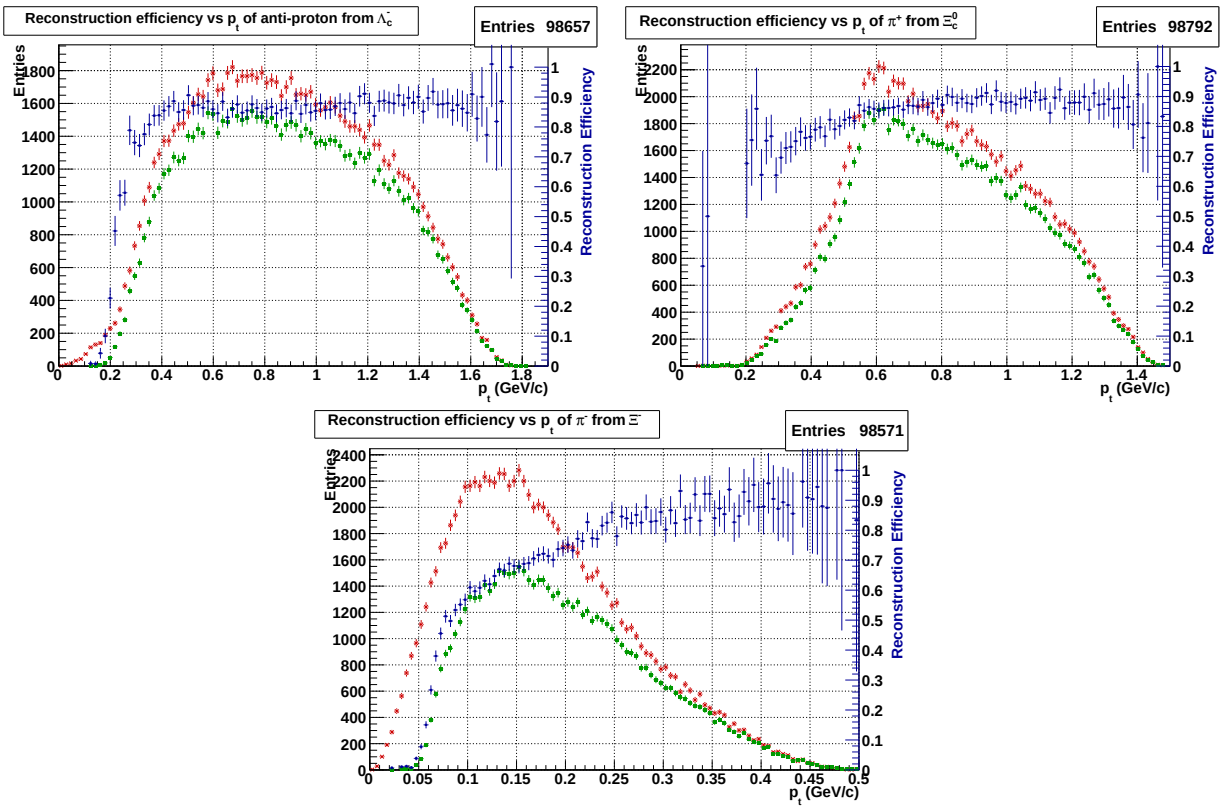


Figure 18: Reconstruction efficiency vs transverse momentum with polar angle acceptance cut $17° - 150°$ of final-state charged particles in signal MC for ch1.

1. Apply vertex fit to Λ candidates taken from VEE2 bank (of any quality), using the vertex from bank as the seed.

2. Apply mass-vertex fit to Λ.

3. Reconstruct Ξ^- using the four momentum from Λ mass-vertex fitting.

4. Apply mass-vertex fit to Ξ^- and require $\chi^2/n.d.f < 100$, where $n.d.f$ is the number of degrees of freedom for the mass-vertex fit.

4.3.5 Reconstruction of Ξ_c^0 Candidates

Ξ_c^0 candidates are reconstructed in the decay channels

$$\Xi_c^0 \to \Xi^- \pi^+,$$

$$\Xi_c^0 \to \Lambda K^- \pi^+,$$

and

$$\Xi_c^0 \to pK^-K^-\pi^+.$$

In decay channels $\Xi_c^0 \to \Xi^-\pi^+/pK^-K^-\pi^+$, mass-vertex fit is applied.

In decay channel $\Xi_c^0 \to \Lambda K^-\pi^+$, in order to improve the ΔE and invariant mass resolutions, the following reconstruction procedure is followed:

1. Apply vertex fit to Λ candidates taken from VEE2 bank (of any quality), using the vertex from bank as the seed.

2. Apply mass-vertex fit to Λ.

3. Reconstruct Ξ_c^0 using the four momentum from Λ mass-vertex fitting.

4. Apply mass-vertex fit to Ξ_c^0 and require $\chi^2/n.d.f < 100$, where $n.d.f$ is the number of degrees of freedom for the mass-vertex fit.

4.3.6 Reconstruction of $\overline{\Lambda}_c^-$ Candidates

$\overline{\Lambda}_c^-$ candidates are reconstructed in the decay channels

$$\overline{\Lambda}_c^- \to \overline{p}K_S^0$$

and

$$\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-.$$

In decay channel $\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-$, mass-vertex fit is applied.

In decay channel $\overline{\Lambda}_c^- \to \overline{p}K_S^0$, in order to improve the ΔE and invariant mass resolutions, the following reconstruction procedure is followed:

1. Apply vertex fit to K_S^0 candidates taken from VEE2 bank, using the vertex from bank as the seed.

2. Apply mass-vertex fit for K_S^0.

3. Reconstruct $\overline{\Lambda}_c^-$ using the four momentum from K_S^0 mass-vertex fitting.

4. Apply mass-vertex fit to $\overline{\Lambda}_c^-$ and require $\chi^2/n.d.f < 100$, where $n.d.f$ is the number of degrees of freedom for the mass-vertex fit.

4.3.7 Reconstruction of Ξ_c^0, $\overline{\Lambda}_c^-$, Ξ^-, and Λ Candidates

Fig. 19 shows the invariant mass distributions of Ξ_c^0, $\overline{\Lambda}_c^-$, Ξ^-, and Λ candidates in signal MC sample for ch1. The reconstructed invariant mass of each particle candidate is calculated after a vertex fit (without mass constraint). Vertex fits with mass constraints are applied to daughter particles in the decay chain. These distributions are fit with the double-Gaussian function and first-order Chebychev polynomial function.

We choose the invariant mass cuts as $|M_{\Xi_c^0} - m_{\Xi_c^0}| < 20$ MeV/c^2, $|M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-}| < 10$ MeV/c^2, $|M_{\Xi^-} - m_{\Xi^-}| < 10$ MeV/c^2, $|M_{\Lambda} - m_{\Lambda}| < 10$ MeV/c^2. Ξ_c^0 and $\overline{\Lambda}_c^-$ candidates are selected in about 3σ interval.

We tried to optimize the invariant mass cuts based on the figure of merit $\frac{S}{uncertainty\ of\ S}$, where S is the signal yield extracted from the maximum likelihood fit. Not surprisingly, the optimized values of our invariant mass cuts are channel-dependent. We decided not to use the optimized invariant mass cuts because, first of all, this would mean that the systematic uncertainties would also be different between different channels. Second of all, as BNV modes generally have less background than SM modes, the optimized values of such cuts would also depend on BNV or SM mode. Finally, we eventually decided to measure the ratio (of BNV to SM), so prior to this decision we were using the same cuts as in the original SM analysis. These cuts (now used for both SM and BNV modes) are highly efficient, and the systematic uncertainties associated with their values are small. Therefore, we decided to use the same mass cuts (20/10/10MeV) as in the previous SM mode studies [33] (this is justified also by extensive validation of the analysis we performed using the SM channels).

The invariant mass distributions for reconstructed candidates in signal MC for all six channels are shown in the Appendix.

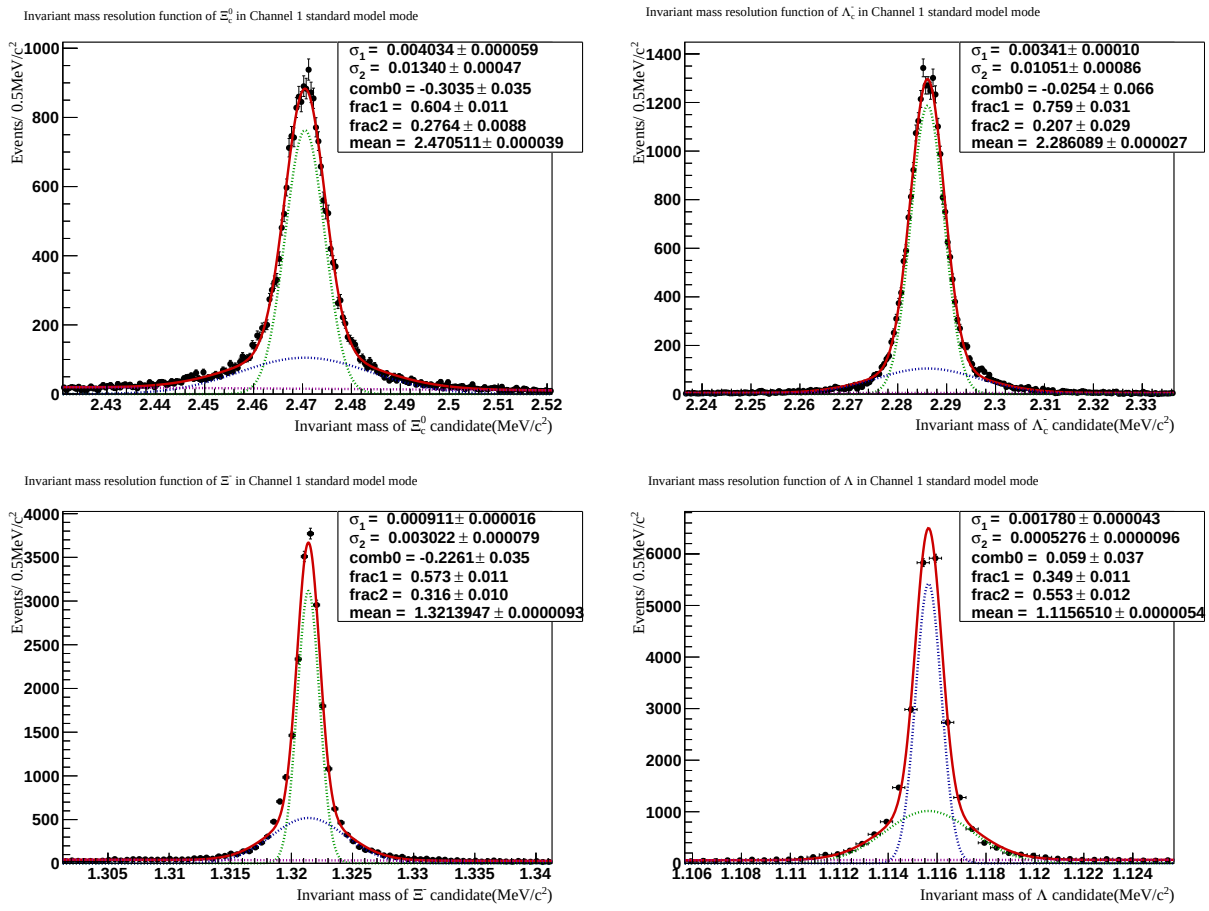


Figure 19: Invariant mass distributions for $\Xi_c^0, \overline{\Lambda}_c^-, \Xi^-, \Lambda$ candidates in signal MC sample for ch1 ($\Xi_c^0 \to \Xi^-\pi^+$, $\overline{\Lambda}_c^- \to \overline{p}K_S^0$). The invariant mass distribution of each particle candidate is after a vertex fit while vertex fits with mass constraints are applied to daughter particles in the decay chain. The red solid lines show the fit results with the double-Gaussian function and first-order Chebychev polynomial function.

4.3.8 Reconstruction of B^- Candidates

B^- signal candidates are reconstructed in the decay channel

$$B^- \to \Xi_c^0 \overline{\Lambda}_c^-.$$

B^- are selected using the beam-energy-constrained mass $M_{\rm bc}$ and the energy difference (ΔE). Here $M_{\rm bc} = \sqrt{(E_{beam})^2 - (\overrightarrow{p}_B)^2}$ and $\Delta E = E_B - E_{beam}$. E_{beam} is the beam energy,

$\overrightarrow{p}_B$ and E_B are the momentum and energy of B meson candidates in the center-of-mass (CM) frame. The candidates must satisfy the selection criteria

$$M_{\rm bc} > 5.2 \text{ GeV/c}^2$$

$$|\Delta E| < 0.25 \text{ GeV}$$

In addition, we define the signal region as

$$M_{\rm bc} > 5.27 \text{ GeV/c}^2$$

$$|\Delta E| < 0.02 \text{ GeV}$$

Particle candidates	Selection criteria
$\pi^{\pm}$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$, $p_{\perp} > 50$ MeV/c
$K^{\pm}$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$, $p_{\perp} > 50$ MeV/c
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$, $p_{\perp} > 50$ MeV/c
Λ	FindLambda selection from vee2, $\lvert M_{\Lambda} - m_{\Lambda}\rvert < 10$ MeV/c^2
K^0_S	nisKsFinder selection from vee2
Ξ^-	$\lvert M_{\Xi^-} - m_{\Xi^-}\rvert < 10$ MeV/c^2, $\chi^2 < 100$
Ξ^0_c	$\lvert M_{\Xi^0_c} - m_{\Xi^0_c}\rvert < 20$ MeV/c^2, $\chi^2 < 100$
$\overline{\Lambda}^-_c$	$\lvert M_{\overline{\Lambda}^-_c} - m_{\overline{\Lambda}^-_c}\rvert < 10$ MeV/c^2, $\chi^2 < 100$
B^-	$M_{\rm bc} > 5.2$ GeV/c^2, $\lvert\Delta E\rvert < 0.25$ GeV

Table 2: Selection criteria. The reconstructed invariant mass of each particle candidate is calculated after a vertex fit (without mass constraint). Vertex fits with mass constraints are applied to daughter particles in the decay chain.

4.3.9 Best Candidate Selection

After applying all selection criteria, in the 6 decay channels, the percentages of reconstructed signal MC events that contain more than one candidate are 5.9%, 8.3%, 10.7%, 13.4%, 15.4%, and 17.3%, respectively. Fig. 20 shows the number of candidates per event before the best candidate selection.

We use combined χ^2 obtained from mass-vertex fits to Ξ_c^0, $\overline{\Lambda}_c^-$ and Ξ^- (if involved) to choose the best candidate. The candidate with the smallest combined χ^2 value is chosen as the best candidate. As shown in Fig. 20 by histograms with the error bars, events where the best candidate is fully MC-tagged account for 96.6%, 93.5%, 92.6%, 89.6%, 87.4% and 84.7%, respectively, for the 6 decay channels used in our analysis.

Events with more than one candidate where the best candidate is fully MC-tagged account for 93.5%, 82.9%, 84.4%, 76.3%, 76.0%, and 71.6%, respectively.

We also tried to include χ^2 of Λ into the combined χ^2 for best candidate selection. Take channel 1 SM mode as an example (we consider only B^- candidates here), events with more than one candidate make only 564 events out of 9837 events = 5.7% of the sample. Among these 564 events, 345 events have exactly the same Λ candidate. In the remaining 219 events, χ^2 of the Λ from the best candidate is smaller than χ^2 of the Λ from the next to the best candidate in 127 events, which is more than 50%, but not statistically significant.

However, when Λ χ^2 is added to the combined χ^2 we use to identify the best candidate, 178 out of 219 events where the Λ candidates are different still have smaller combined χ^2 associated with our best candidate selected without using the Λ. In other words, adding this additional χ^2 to our best candidate selection would change the outcome for 41 out of 219 events, i.e., 19% of 219 events. Finally, the outcome changes just for 41 out of 9837 events, which is 0.4% of selected signal MC events.

Note that selection criteria $M_{\rm bc} > 5.2$ GeV/c^2, $|\Delta E| < 0.25$ GeV are used. In section 3.10, the fraction of MC-tagged best candidates in signal region, *i.e.* $M_{\rm bc} > 5.27$ GeV/c^2 and $|\Delta E| < 0.02$ GeV, will be further discussed.

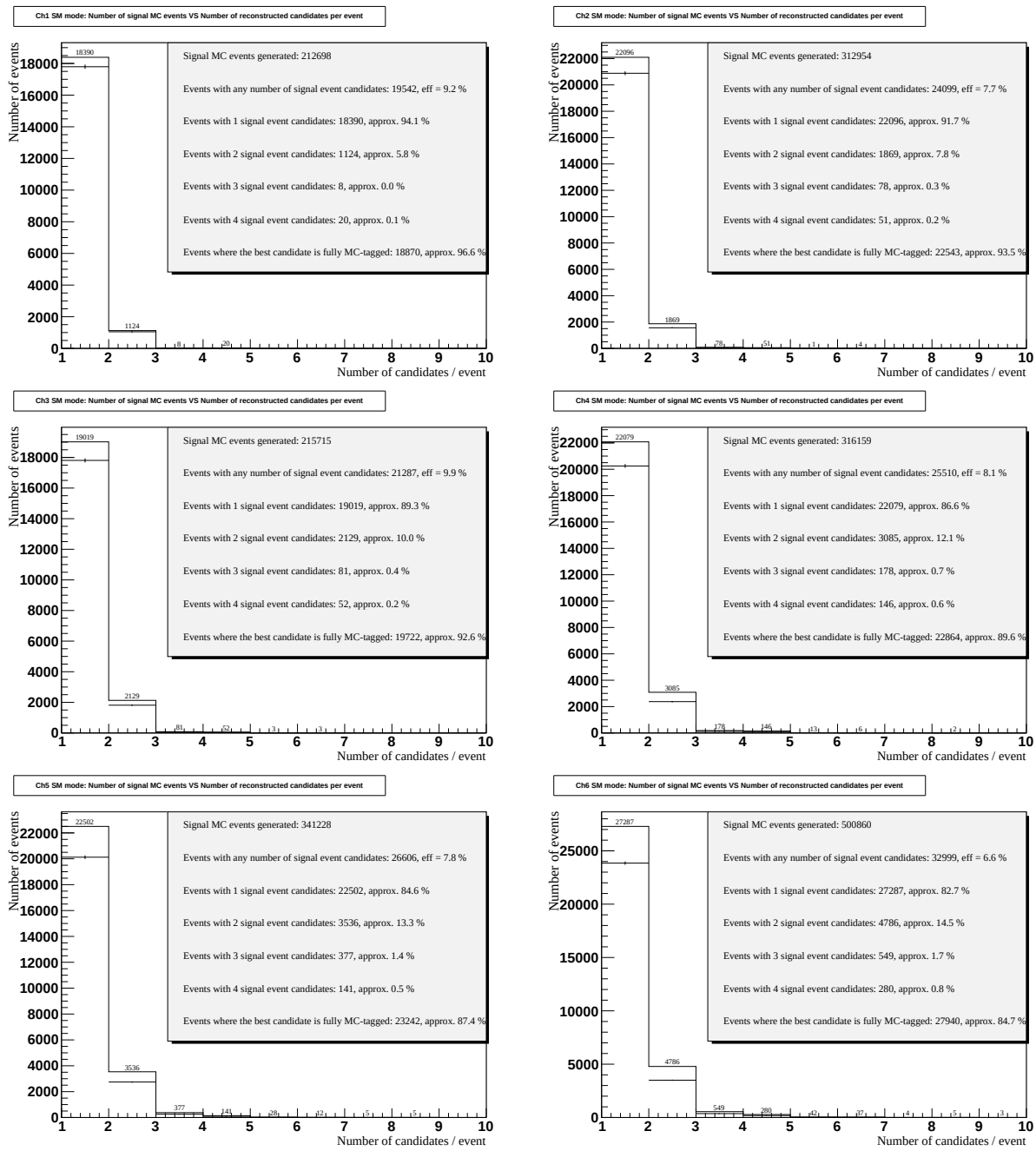


Figure 20: Distributions of signal candidate multiplicity (solid histograms) and distributions of best candidates which are fully MC tagged (histograms with the error bars), *i.e.* correctly reconstructed for signal MC events in each of the 6 channels used in our analysis.

4.3.10 Cut Flow Analysis for Signal Events

In this section, cut flow analysis is performed to investigate the effect of each selection criterion on the signal MC sample. Selection criteria are applied successively. The single cut efficiency, overall efficiency, fraction of events in the signal region, fraction of single candidate events, and fraction of events where the best candidate is fully MC-tagged in the signal region are all shown in the cut flow tables.

Table 3 shows the cut flow table for ch1. The efficiency of baseline selection is 17.2%. Each of the cuts is approximately 90%-efficient and the fraction of MC-tagged events grows after each cut. The rest of the tables exhibit similar dynamics for the other 5 channels.

Parameter	Requirement	Efficiency(this cut)	Efficiency(overall)	SignalRegion	SingleCand	MCtagged(SR)
Baseline	Best candidate	N/A	17.20 %	77.90 %	62.50 %	91.50 %
$p_\perp$	$p_\perp > 50$ MeV/c	100.00 %	17.20 %	77.90 %	62.50 %	91.60 %
$\pi^\pm$	$R_{\pi,K} > 0.6,\ R_{e,hadron} < 0.95$	90.12 %	15.50 %	79.90 %	67.20 %	91.80 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	89.68 %	13.90 %	82.80 %	73.40 %	92.00 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	82.01 %	11.40 %	93.20 %	86.60 %	95.00 %
Λ	$\lvert M_\Lambda - m_\Lambda \rvert < 10$ MeV/c^2	98.25 %	11.20 %	93.50 %	87.70 %	95.60 %
Ξ^-	$\lvert M_{\Xi^-} - m_{\Xi^-} \rvert < 10$ MeV/c^2	95.54 %	10.70 %	95.00 %	89.30 %	96.60 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0} \rvert < 20$ MeV/c^2	95.33 %	10.20 %	96.90 %	92.40 %	97.10 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-} \rvert < 10$ MeV/c^2	90.88 %	9.27 %	97.70 %	93.10 %	97.50 %
$M_{\rm bc}$	$M_{\rm bc} > 5.2$ GeV/c^2	99.46 %	9.22 %	98.50 %	94.20 %	97.50 %
ΔE	$\lvert \Delta E \rvert < 0.25$ GeV	100.00 %	9.22 %	98.50 %	94.30 %	97.50 %

Table 3: Cut flow table for ch1 signal MC.

Parameter	Requirement	Efficiency(this cut)	Efficiency(overall)	SignalRegion	SingleCand	MCtagged(SR)
Baseline	Best candidate	N/A	28.00 %	32.40 %	12.20 %	89.30 %
$p_\perp$	$p_\perp > 50$ MeV/c	99.64 %	27.90 %	32.40 %	12.20 %	89.30 %
$\pi^\pm$	$R_{\pi,K} > 0.6,\ R_{e,hadron} < 0.95$	83.51 %	23.30 %	37.20 %	18.60 %	90.40 %
$K^\pm$	$R_{\pi,K} < 0.6,\ R_{e,hadron} < 0.95$	79.40 %	18.50 %	48.40 %	30.10 %	91.00 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	75.14 %	13.90 %	66.50 %	51.90 %	91.20 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	75.54 %	10.50 %	79.90 %	66.00 %	94.40 %
Λ	$\lvert M_\Lambda - m_\Lambda \rvert < 10$ MeV/c^2	98.10 %	10.30 %	80.20 %	66.90 %	95.00 %
Ξ^-	$\lvert M_{\Xi^-} - m_{\Xi^-} \rvert < 10$ MeV/c^2	96.21 %	9.91 %	81.50 %	68.30 %	96.00 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0} \rvert < 20$ MeV/c^2	95.06 %	9.42 %	83.20 %	70.70 %	96.50 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-} \rvert < 10$ MeV/c^2	83.55 %	7.87 %	92.70 %	88.70 %	97.40 %
$M_{\rm bc}$	$M_{\rm bc} > 5.2$ GeV/c^2	97.97 %	7.71 %	95.30 %	91.50 %	97.40 %
ΔE	$\lvert \Delta E \rvert < 0.25$ GeV	99.74 %	7.69 %	95.60 %	91.70 %	97.40 %

Table 4: Cut flow table for ch2 signal MC.

Parameter	Requirement	Efficiency(this cut)	Efficiency(overall)	SignalRegion	SingleCand	MCtagged(SR)
Baseline	Best candidate	N/A	22.30 %	54.20 %	34.70 %	93.50 %
$p_\perp$	$p_\perp > 50$ MeV/c	100.00 %	22.30 %	54.20 %	34.70 %	93.50 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	92.83 %	20.70 %	58.00 %	41.00 %	93.60 %
$K^\pm$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$	77.78 %	16.10 %	73.50 %	60.20 %	94.60 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	86.96 %	14.00 %	78.20 %	67.30 %	94.80 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	87.86 %	12.30 %	85.20 %	74.70 %	95.90 %
Λ	$\lvert M_\Lambda - m_\Lambda \rvert < 10$ MeV/c^2	97.56 %	12.00 %	86.20 %	75.20 %	96.20 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0} \rvert < 20$ MeV/c^2	91.67 %	11.00 %	91.30 %	86.60 %	96.80 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-} \rvert < 10$ MeV/c^2	90.00 %	9.90 %	92.40 %	87.30 %	97.30 %
$M_{\rm bc}$	$M_{\rm bc} > 5.2$ GeV/c^2	98.69 %	9.77 %	94.00 %	89.20 %	97.30 %
ΔE	$\lvert \Delta E \rvert < 0.25$ GeV	99.80 %	9.75 %	94.30 %	89.50 %	97.30 %

Table 5: Cut flow table for ch3 signal MC.

Parameter	Requirement	Efficiency(this cut)	Efficiency(overall)	SignalRegion	SingleCand	MCtagged(SR)
Baseline	Best candidate	N/A	40.20 %	19.20 %	8.53 %	91.40 %
$p_\perp$	$p_\perp > 50$ MeV/c	100.00 %	40.20 %	19.20 %	8.53 %	91.40 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	85.07 %	34.20 %	23.90 %	14.40 %	92.10 %
$K^\pm$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$	62.28 %	21.30 %	39.30 %	25.40 %	93.60 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	66.20 %	14.10 %	61.00 %	46.40 %	94.10 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	80.14 %	11.30 %	71.40 %	56.00 %	95.30 %
Λ	$\lvert M_\Lambda - m_\Lambda \rvert < 10$ MeV/c^2	98.23 %	11.10 %	72.30 %	56.50 %	95.50 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0} \rvert < 20$ MeV/c^2	90.09 %	10.00 %	77.50 %	65.80 %	96.30 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-} \rvert < 10$ MeV/c^2	82.00 %	8.20 %	87.70 %	82.40 %	97.40 %
$M_{\rm bc}$	$M_{\rm bc} > 5.2$ GeV/c^2	96.95 %	7.95 %	91.20 %	85.90 %	97.40 %
ΔE	$\lvert \Delta E \rvert < 0.25$ GeV	99.37 %	7.90 %	91.90 %	86.70 %	97.40 %

Table 6: Cut flow table for ch4 signal MC.

Parameter	Requirement	Efficiency(this cut)	Efficiency(overall)	SignalRegion	SingleCand	MCtagged(SR)
Baseline	Best candidate	N/A	39.20 %	18.00 %	12.40 %	87.20 %
$p_\perp$	$p_\perp > 50$ MeV/c	99.74 %	39.10 %	18.00 %	12.40 %	87.30 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	96.42 %	37.70 %	19.10 %	14.30 %	88.30 %
$K^\pm$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$	49.07 %	18.50 %	43.90 %	27.70 %	92.90 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	64.86 %	12.00 %	69.00 %	60.60 %	93.90 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	85.00 %	10.20 %	76.70 %	68.50 %	95.00 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0} \rvert < 20$ MeV/c^2	85.98 %	8.77 %	85.70 %	82.40 %	96.40 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-} \rvert < 10$ MeV/c^2	90.08 %	7.90 %	86.60 %	83.20 %	96.80 %
$M_{\rm bc}$	$M_{\rm bc} > 5.2$ GeV/c^2	98.73 %	7.80 %	88.00 %	84.10 %	96.80 %
ΔE	$\lvert \Delta E \rvert < 0.25$ GeV	99.49 %	7.76 %	88.40 %	84.50 %	96.80 %

Table 7: Cut flow table for ch5 signal MC.

Parameter	Requirement	Efficiency(this cut)	Efficiency(overall)	SignalRegion	SingleCand	MCtagged(SR)
Baseline	Best candidate	N/A	70.50 %	6.41 %	4.55 %	86.50 %
$p_\perp$	$p_\perp > 50$ MeV/c	99.86 %	70.40 %	6.40 %	4.56 %	86.50 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	89.20 %	62.80 %	8.07 %	7.64 %	87.40 %
$K^\pm$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$	44.75 %	28.10 %	20.90 %	13.00 %	92.20 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	43.06 %	12.10 %	55.60 %	43.00 %	93.10 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	80.08 %	9.69 %	64.90 %	52.20 %	94.60 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0}\rvert < 20$ MeV/c^2	86.17 %	8.35 %	72.70 %	63.50 %	96.20 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-}\rvert < 10$ MeV/c^2	82.04 %	6.85 %	81.90 %	79.70 %	97.00 %
$M_{\rm bc}$	$M_{\rm bc} > 5.2$ GeV/c^2	97.08 %	6.65 %	85.10 %	82.30 %	96.90 %
ΔE	$\lvert\Delta E\rvert < 0.25$ GeV	99.25 %	6.60 %	85.80 %	83.00 %	96.90 %

Table 8: Cut flow table for ch6 signal MC.

4.3.11 Cut Flow Analysis for Background Events

In this section, cut flow analysis is performed to investigate the effect of each selection criterion on the generic MC sample. The number of charged/mixed/continuum background events surviving the selection criteria in full fit range ($M_{\rm bc} > 5.2$ GeV/c^2, $|\Delta E| < 0.05$ GeV) and signal region ($M_{\rm bc} > 5.27$ GeV/c^2, $|\Delta E| < 0.02$ GeV) are shown in the cut flow tables. 6 streams of generic MC are used in this cut flow analysis.

Table 9 to 14 show the cut flow table for generic MC. We can see that background contributions are small, so we concluded that no further background suppression is needed in our analysis.

Parameter	Requirement	Charged	Mixed	Continuum	Elimination rate	Charged(SR)	Mixed(SR)	Continuum(SR)	Elimination rate(SR)
Baseline	Best candidate	490	745	2077	N/A	38	54	120	N/A
$p_\perp$	$p_\perp > 50$ MeV	490	742	2075	0.15 %	38	54	119	0.47 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	475	719	1988	3.78 %	37	52	116	2.84 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	429	609	1769	11.79 %	26	47	102	14.63 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	100	127	301	81.19 %	9	13	19	76.57 %
Λ	$\lvert M_\Lambda - m_\Lambda\rvert < 10$ MeV/c^2	82	110	254	15.53 %	8	12	13	19.51 %
Ξ^-	$\lvert M_{\Xi^-} - m_{\Xi^-}\rvert < 10$ MeV/c^2	50	60	147	42.38 %	6	3	8	48.48 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0}\rvert < 20$ MeV/c^2	32	34	67	48.25 %	6	0	2	52.94 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-}\rvert < 10$ MeV/c^2	10	9	22	69.17 %	4	0	1	37.50 %

Table 9: Cut flow table for ch1 generic MC.

Parameter	Requirement	Charged	Mixed	Continuum	Elimination rate	Charged(SR)	Mixed(SR)	Continuum(SR)	Elimination rate(SR)
Baseline	Best candidate	38678	45780	75585	N/A	1878	2401	3881	N/A
$p_\perp$	$p_\perp > 50$ MeV	38578	45675	75435	0.22 %	1871	2395	3873	0.26 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	26317	31781	54091	29.74 %	1300	1770	2755	28.43 %
$K^\pm$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$	18119	21637	37368	31.26 %	920	1218	1949	29.84 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	11838	14027	25482	33.42 %	641	873	1334	30.32 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	1554	1864	3200	87.11 %	76	155	167	86.03 %
Λ	$\lvert M_\Lambda - m_\Lambda \rvert < 10$ MeV/c^2	1226	1448	2577	20.66 %	61	126	124	21.86 %
Ξ^-	$\lvert M_{\Xi^-} - m_{\Xi^-} \rvert < 10$ MeV/c^2	636	749	1361	47.71 %	27	60	72	48.87 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0} \rvert < 20$ MeV/c^2	265	296	586	58.23 %	14	29	33	52.20 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-} \rvert < 10$ MeV/c^2	47	88	102	79.34 %	3	11	4	76.32 %

Table 10: Cut flow table for ch2 generic MC.

Parameter	Requirement	Charged	Mixed	Continuum	Elimination rate	Charged(SR)	Mixed(SR)	Continuum(SR)	Elimination rate(SR)
Baseline	Best candidate	4464	5941	13101	N/A	279	315	742	N/A
$p_\perp$	$p_\perp > 50$ MeV	4462	5941	13099	0.02 %	279	315	742	0.00 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	3593	4932	11470	14.92 %	223	245	655	15.94 %
$K^\pm$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$	2782	3745	8991	22.39 %	173	188	520	21.55 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	2149	2890	7153	21.43 %	140	145	405	21.68 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	736	935	2282	67.58 %	53	54	137	64.64 %
Λ	$\lvert M_\Lambda - m_\Lambda \rvert < 10$ MeV/c^2	583	736	1842	20.04 %	46	42	103	21.72 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0} \rvert < 20$ MeV/c^2	248	315	723	59.32 %	15	17	45	59.69 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-} \rvert < 10$ MeV/c^2	89	118	171	70.61 %	6	7	9	71.43 %

Table 11: Cut flow table for ch3 generic MC.

Parameter	Requirement	Charged	Mixed	Continuum	Elimination rate	Charged(SR)	Mixed(SR)	Continuum(SR)	Elimination rate(SR)
Baseline	Best candidate	595086	612237	661972	N/A	25714	26237	31261	N/A
$p_\perp$	$p_\perp > 50$ MeV	594833	611956	661642	0.05 %	25688	26224	31234	0.08 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	350539	366634	428644	38.67 %	15932	16326	20747	36.25 %
$K^\pm$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$	187859	189584	227146	47.24 %	9345	9074	11748	43.09 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	93577	92642	127660	48.08 %	4906	4597	6790	45.99 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	25706	23848	32159	73.97 %	1449	1199	1703	73.30 %
Λ	$\lvert M_\Lambda - m_\Lambda \rvert < 10$ MeV/c^2	19931	18461	25184	22.20 %	1157	941	1337	21.05 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0} \rvert < 20$ MeV/c^2	8286	7780	10388	58.39 %	442	409	538	59.56 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-} \rvert < 10$ MeV/c^2	1839	1809	2002	78.64 %	115	105	112	76.10 %

Table 12: Cut flow table for ch4 generic MC.

Parameter	Requirement	Charged	Mixed	Continuum	Elimination rate	Charged(SR)	Mixed(SR)	Continuum(SR)	Elimination rate(SR)
Baseline	Best candidate	4440	4499	5521	N/A	198	229	262	N/A
$p_\perp$	$p_\perp > 50$ MeV	4436	4494	5514	0.11 %	198	228	260	0.44 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	3739	3735	4371	17.99 %	168	184	207	18.51 %
$K^\pm$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$	1705	1721	2060	53.69 %	76	99	93	52.06 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	914	800	1034	49.91 %	48	52	51	43.66 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	369	293	292	65.28 %	24	14	18	62.91 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0} \rvert < 20$ MeV/c^2	154	135	109	58.28 %	10	12	9	44.64 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-} \rvert < 10$ MeV/c^2	66	55	20	64.57 %	4	3	0	77.42 %

Table 13: Cut flow table for ch5 generic MC.

Parameter	Requirement	Charged	Mixed	Continuum	Elimination rate	Charged(SR)	Mixed(SR)	Continuum(SR)	Elimination rate(SR)
Baseline	Best candidate	329288	262680	187687	N/A	12114	9491	7721	N/A
$p_\perp$	$p_\perp > 50$ MeV	329101	262522	187535	0.06 %	12104	9495	7712	0.05 %
$\pi^\pm$	$R_{\pi,K} > 0.6$, $R_{e,hadron} < 0.95$	232513	186373	128677	29.72 %	8511	6715	5300	29.97 %
$K^\pm$	$R_{\pi,K} < 0.6$, $R_{e,hadron} < 0.95$	93070	75131	51056	59.96 %	3991	3147	2415	53.46 %
$p/\overline{p}$	$R_{p/\overline{p},\pi} > 0.6, R_{p/\overline{p},K} > 0.6$	27404	20879	17100	70.18 %	1393	1051	952	64.45 %
χ^2	$\chi^2 < 100$ for $\Xi_c^0, \overline{\Lambda}_c^-$	8590	5823	4373	71.27 %	478	311	246	69.52 %
Ξ_c^0	$\lvert M_{\Xi_c^0} - m_{\Xi_c^0}\rvert < 20$ MeV/c^2	3509	2390	1738	59.35 %	201	125	87	60.10 %
$\overline{\Lambda}_c^-$	$\lvert M_{\overline{\Lambda}_c^-} - m_{\overline{\Lambda}_c^-}\rvert < 10$ MeV/c^2	914	579	333	76.09 %	54	32	19	74.58 %

Table 14: Cut flow table for ch6 generic MC.

We observe no peaking backgrounds in generic MC nor in sideband data. We have at least 6 particles in our final states. Usually, background for such final states is dominated by combinatorics. We rely on data sideband and generic MC to estimate background shape.

4.3.12 $M_{\rm bc}$ and ΔE Distributions for Signal and Generic MC

Figure 21 shows the $M_{\rm bc}$ and ΔE distributions for ch1 signal MC after applying all the selection criteria. $M_{\rm bc}$ and ΔE distributions of signal have good resolutions. The correlation factor between $M_{\rm bc}$ and ΔE in the range $M_{\rm bc} > 5.2$ GeV/c^2, $|\Delta E| < 0.25$ GeV is 0.0325.

Figure 22 shows the $M_{\rm bc}$ and ΔE distributions for ch1 and ch2 generic MC after applying all the selection criteria. Here we combine the channels with different tags (ch1 and ch2) due to low statistics of ch1 generic MC, as shown in Table 9. Six streams of generic MC are used in Figure 22.

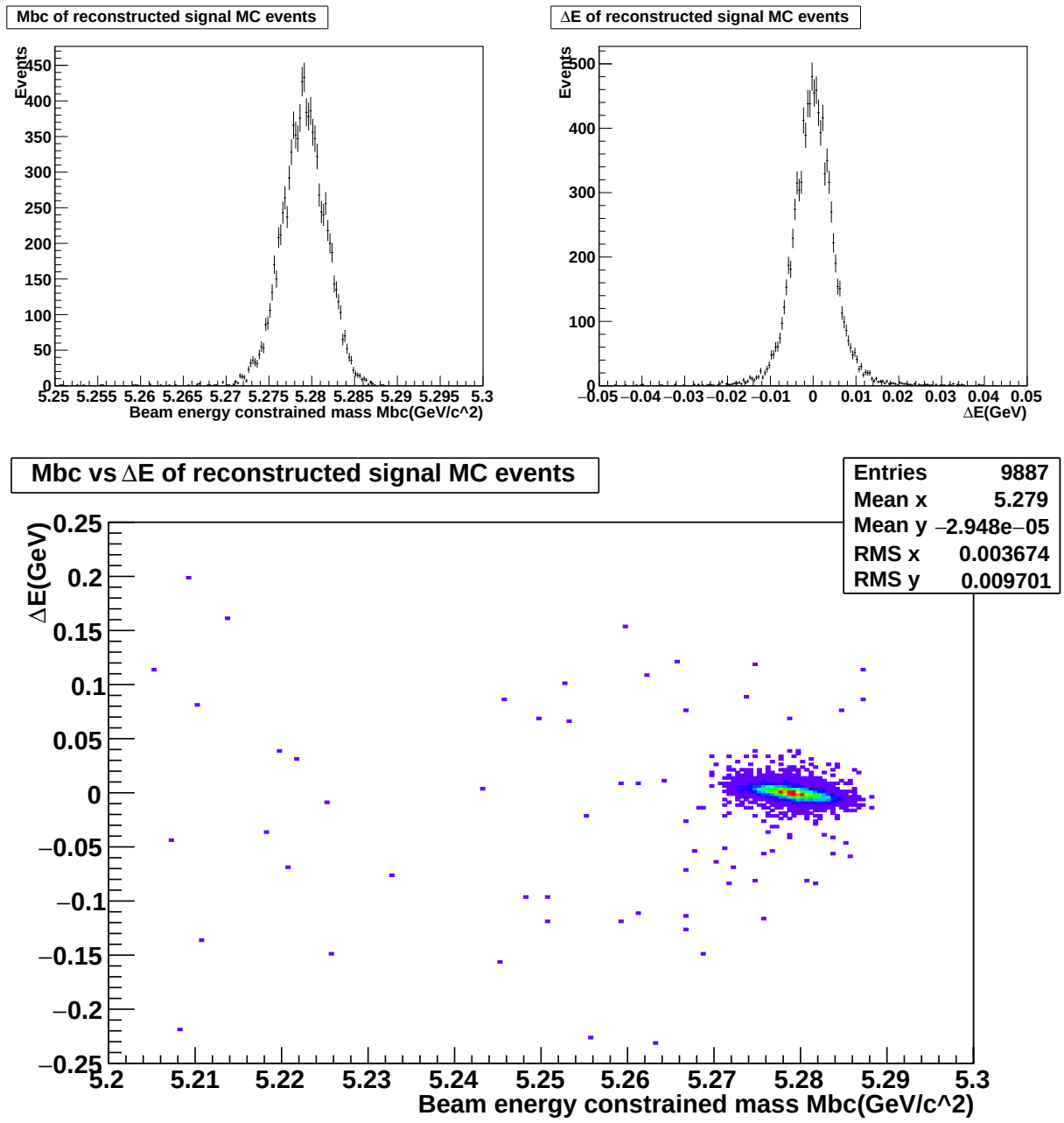


Figure 21: $M_{\rm bc}$ and ΔE distributions for ch1 signal MC.

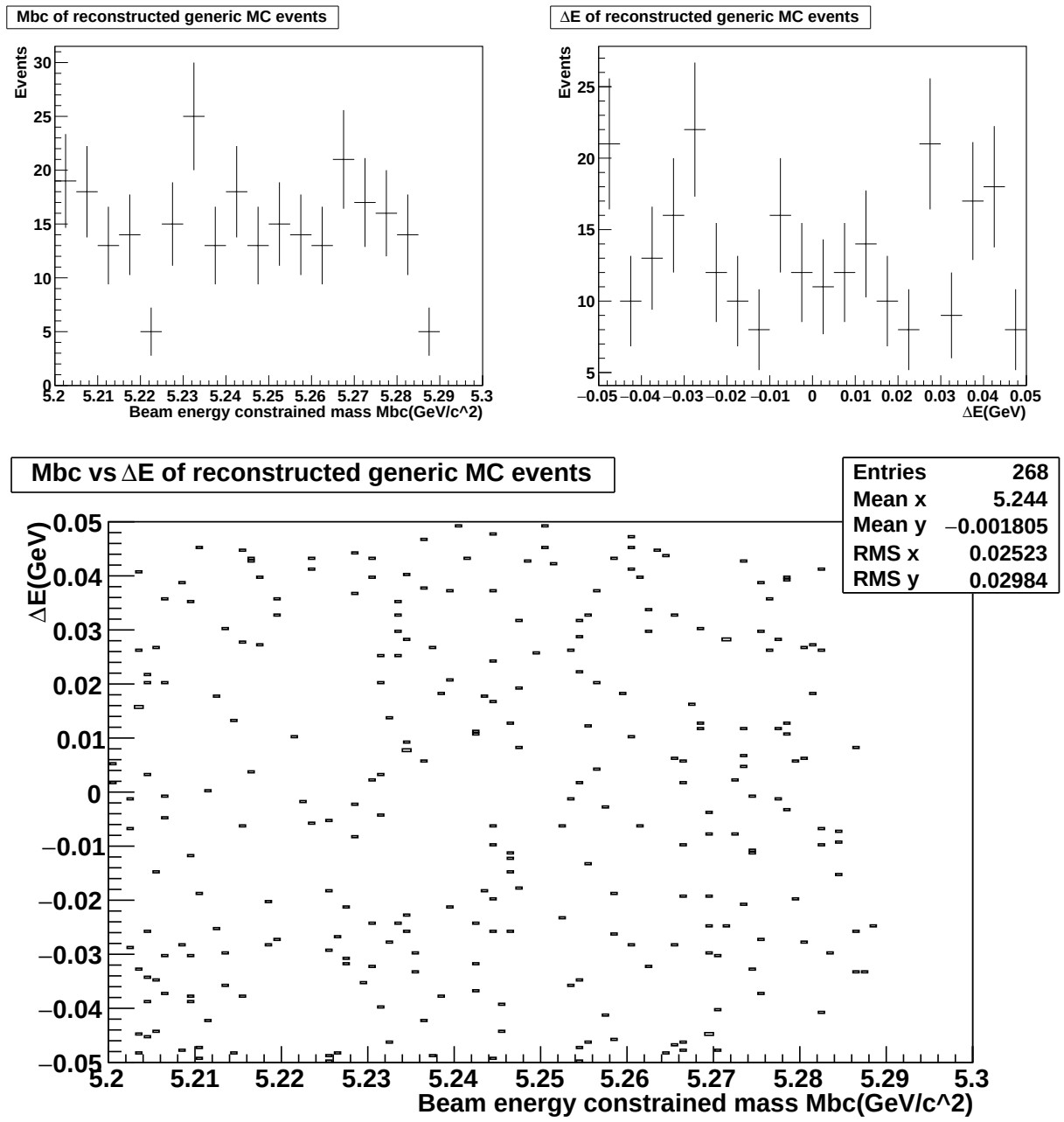


Figure 22: $M_{\rm bc}$ and ΔE distributions for ch1&2 generic MC. Six streams of generic MC samples are used.

4.4 FITTING

To statistically separate signal from background we perform an unbinned 2-dimensional extended maximum likelihood fit to $M_{\rm bc}$ and ΔE distributions. The likelihood function is defined as

$$\mathcal{L} = \frac{e^{-(N_{sig}+N_{bkg})}}{N!} \prod_{i=1}^{N} (N_{sig}\mathcal{P}_{sig} + N_{bkg}\mathcal{P}_{bkg}) \tag{4.1}$$

where N is the total number of candidates, N_{sig} and N_{bkg} are fit parameters representing the numbers of signal and background events, and $\mathcal{P}_{sig}$ and $\mathcal{P}_{bkg}$ are probability density functions (PDF) for signal and background distributions.

To handle signal correlations between $M_{\rm bc}$ and ΔE shown in Fig. 21, a 2-dimensional smoothed histogram of the second order of smoothness is used to model signal PDF. Both signal PDFs and the 2D shape of the smoothed histograms are shown in the next section. To avoid being affected by small non-Gaussian tails, only events in the signal region $M_{\rm bc} > 5.27$ GeV/c^2 and $|\Delta E| < 0.02$ GeV are used to obtain signal PDFs.

The background $M_{\rm bc}$ distribution is modeled with an ARGUS function and background ΔE is modeled with a first-order Chebychev polynomial using 6 streams of generic MC sample. The 2D PDFs for the background are assumed to be factorizable, i.e., $\mathcal{P}_{bkg}(M_{\rm bc}, \Delta E) = \mathcal{P}_{bkg}(M_{\rm bc}) \times \mathcal{P}_{bkg}(\Delta E)$. As shown in Fig. 22, the correlations between $M_{\rm bc}$ and ΔE for the background are negligible. Background PDFs are obtained by fitting events in the entire $M_{\rm bc} > 5.2$ GeV/c^2 and $|\Delta E| < 0.25$ GeV range.

In the following sections, to follow the previous analysis, pairs of $\overline{\Lambda}_c^-$ channels are analyzed together. We show signal and background PDFs for ch1&2, ch3&4, and ch5&6. The results of individual fits to each of the 6 decay channels are shown in the Appendix.

Both SM and BNV modes are investigated.

4.4.1 SM Mode Fits

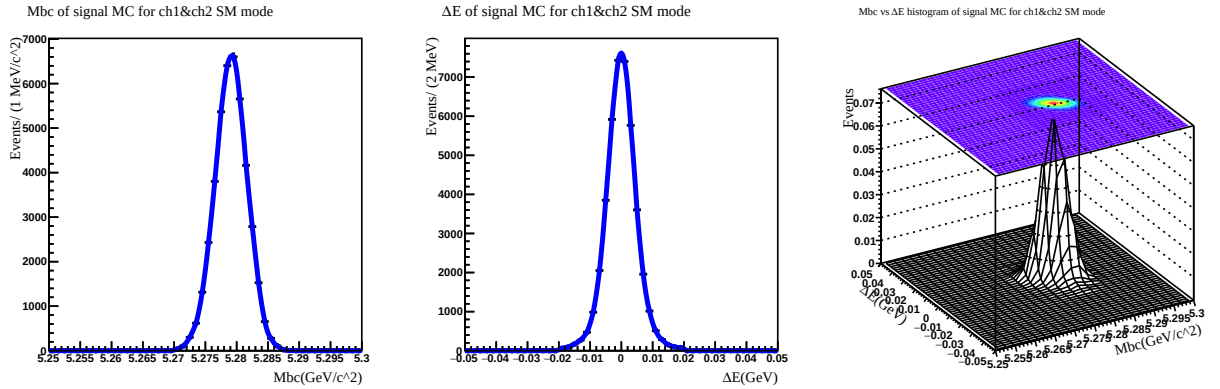


Figure 23: 2D smoothed histogram and projections of ch1&2 signal MC, SM mode.

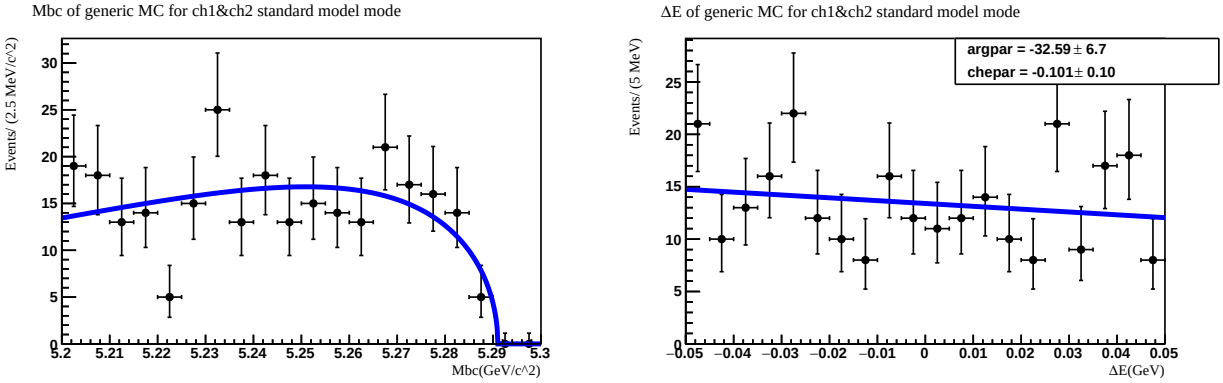


Figure 24: Projections of the 2D fit for ch1&2 generic MC, SM mode.

4.4.1.1 Ch1&2: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Xi^- \pi^+$

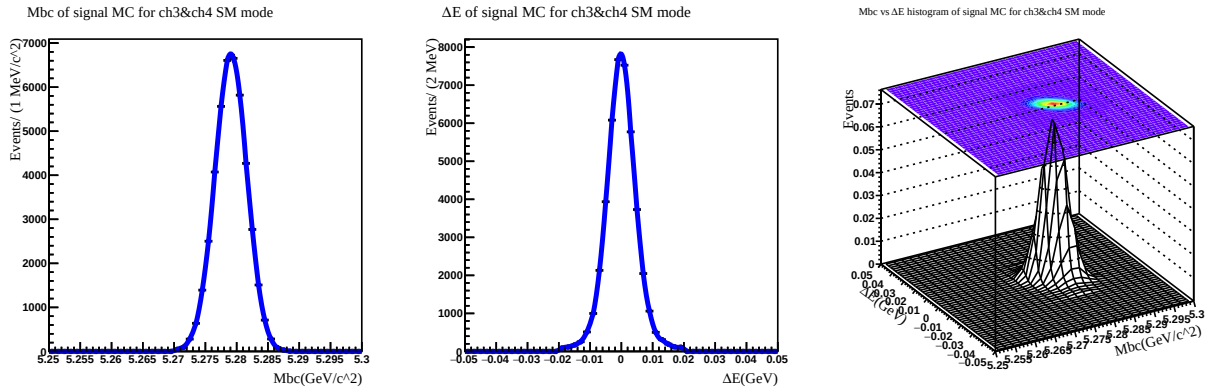


Figure 25: 2D smoothed histogram and projections of ch3&4 signal MC, SM mode.

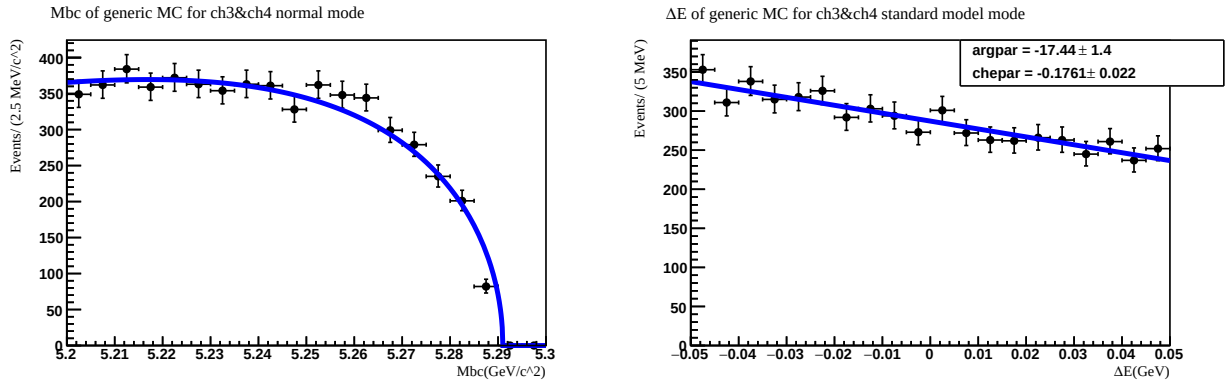


Figure 26: Projections of the 2D fit for ch3&4 generic MC, SM mode.

4.4.1.2 Ch3&4: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Lambda K^- \pi^+$

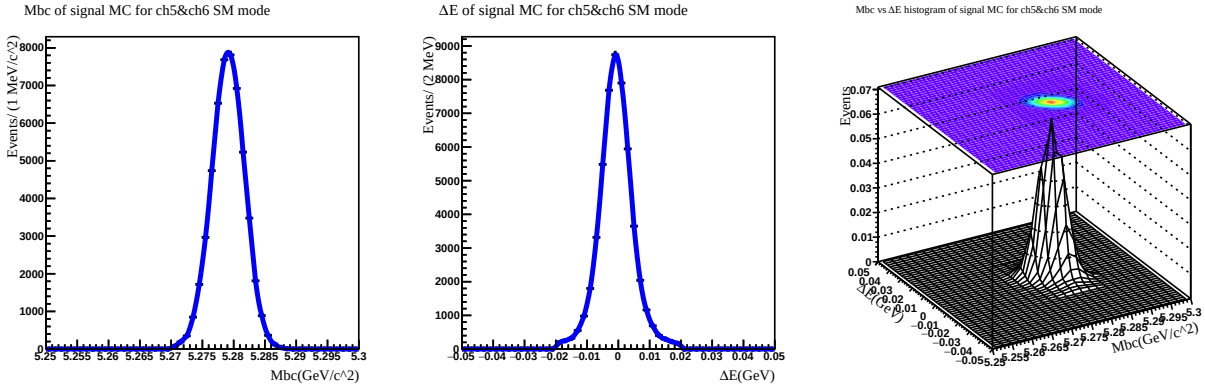


Figure 27: 2D smoothed histogram and projections of ch5&6 signal MC, SM mode.

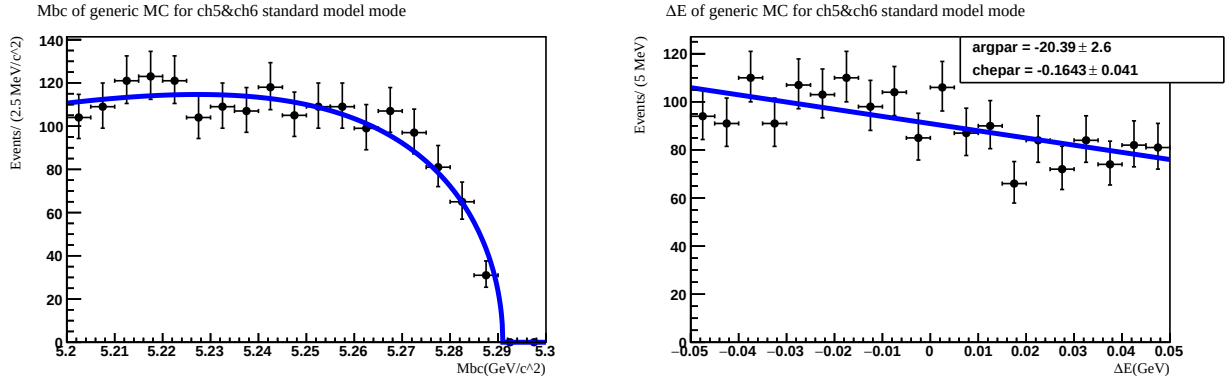


Figure 28: Projections of the 2D fit for ch5&6 generic MC, SM mode.

4.4.1.3 Ch5&6: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to pK^- K^- \pi^+$

4.4.2 BNV Mode Fits

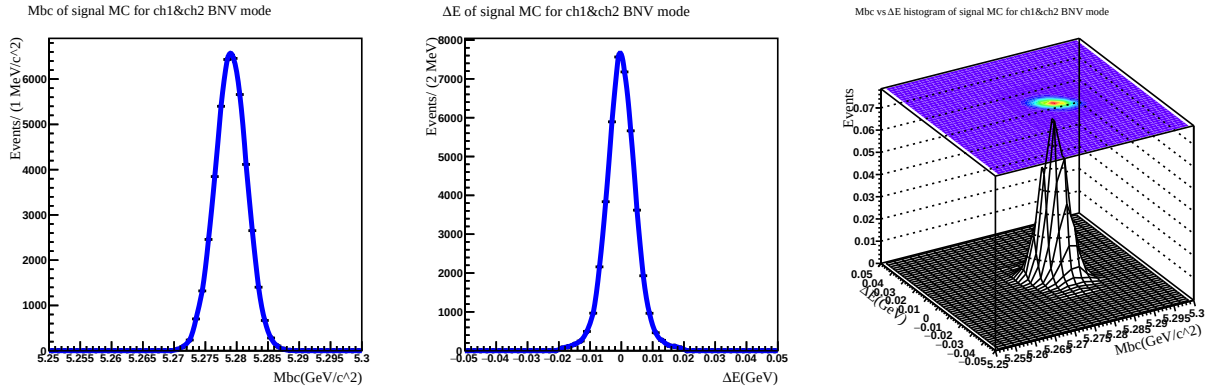


Figure 29: 2D smoothed histogram and projections of ch1&2 signal MC, BNV mode.

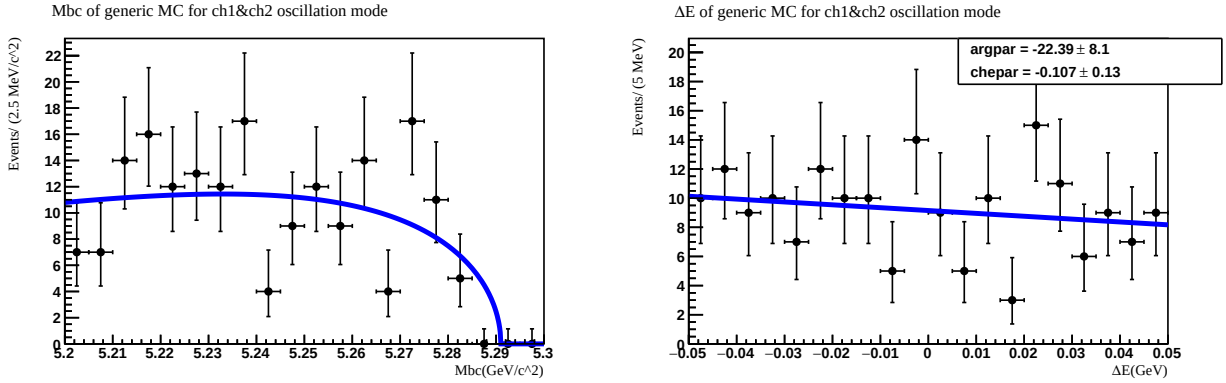


Figure 30: Projections of the 2D fit for ch1&2 generic MC, BNV mode.

4.4.2.1 Ch1&2: $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Xi^- \pi^+$

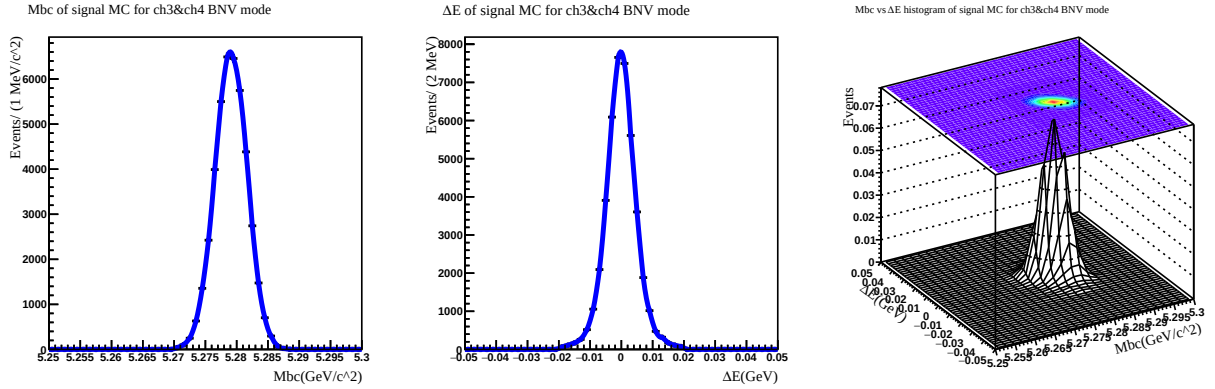


Figure 31: 2D smoothed histogram and projections of ch3&4 signal MC, BNV mode.

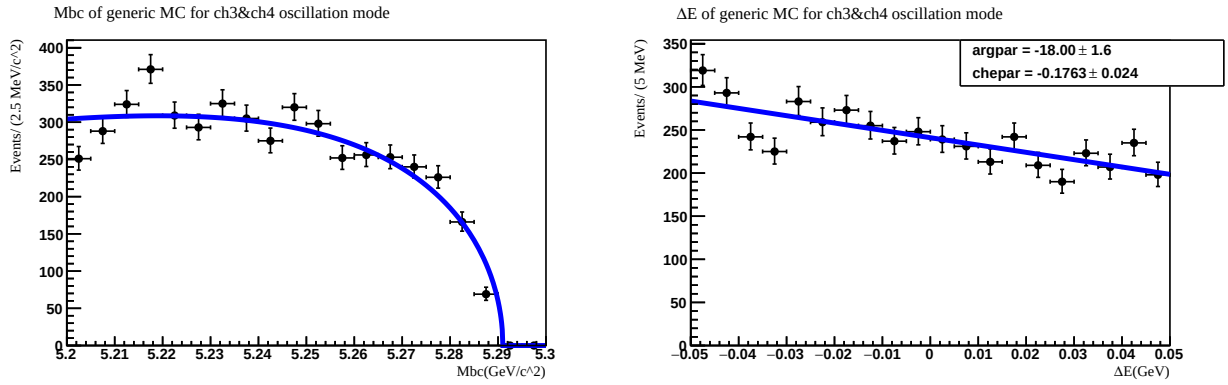


Figure 32: Projections of the 2D fit for ch3&4 generic MC, BNV mode.

4.4.2.2 Ch3&4: $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Lambda K^- \pi^+$

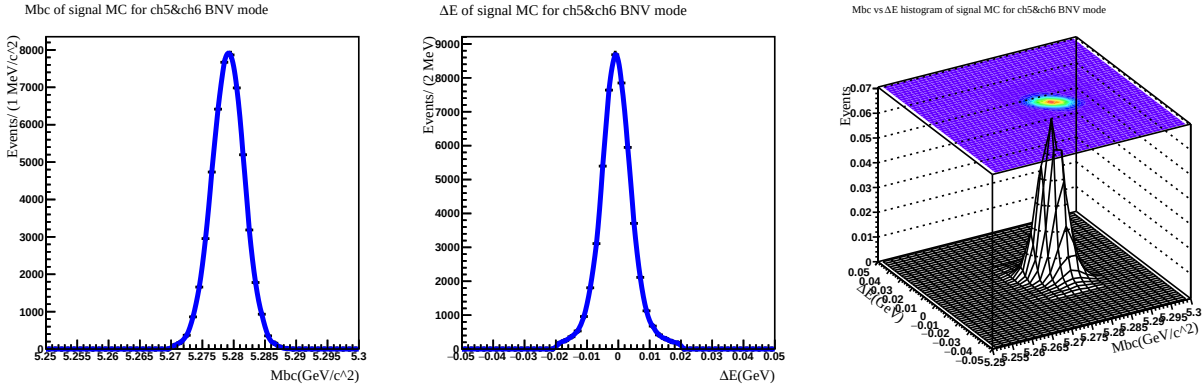


Figure 33: 2D smoothed histogram and projections of ch5&6 signal MC, BNV mode.

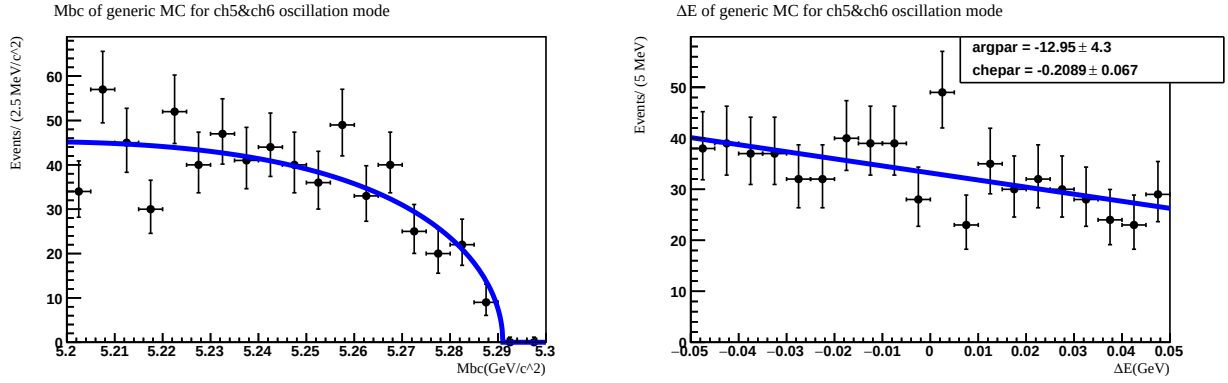


Figure 34: Projections of the 2D fit for ch5&6 generic MC, BNV mode.

4.4.2.3 Ch5&6: $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to pK^- K^- \pi^+$

4.5 ENSEMBLE TESTS

To test our fitting model for possible biases, we perform ensemble tests with toy MC experiments. In each toy MC experiment we generate $N^{gen} = N^{gen}_{sig} + N^{gen}_{bkg}$ events, where N^{gen}_{sig} is the number of signal events and N^{gen}_{bkg} is the number of background events, according to the signal and background shapes obtained from 2D fits described in Chapter 4.4. Poisson fluctuations of N^{gen} are allowed. The numbers of signal events N^{gen}_{sig} are varied, and for each hypothesis, the toy MC experiment is repeated 5000 times. The expected number of background events N^{gen}_{bkg} is estimated using 6 streams of generic MC and is scaled using sideband data.[1]

For ch1&2, we predict 50 signal events, 73 background events in SM mode, and 0 signal events, 27 background events in BNV mode.

For ch3&4, we predict 34 signal events, 1354 background events for SM mode and 0 signal events, 1058 background events for BNV mode.

For ch5&6, we predict 19 signal events, 425 background events for SM mode and 0 signal events, 143 background events for BNV mode.

Unbinned 2-dimensional extended maximum likelihood fit is performed for each sample generated in our toy MC experiments to extract signals N^{fit}_{sig}. Ideally, we expect N^{fit}_{sig} to be unbiased with respect to N^{gen}_{sig}

Fig. 35 shows the extracted numbers of signal events N^{fit}_{sig}, the error on the extracted number of signal events σ^{fit}_{sig}, the distribution of pull P and negative log-likelihood, $-\log\mathcal{L}$, for ch1&2 in BNV mode with 2D fit to $M_{\rm bc}$ and ΔE. The pull P is defined as

$$P = \frac{N^{fit}_{sig} - N^{gen}_{sig}}{\sigma^{fit}_{sig}}$$

For unbiased σ^{fit}_{sig} and large statistics, the pull distribution is expected to be a Gaussian with zero mean and unit width.

Fig. 35 through 40 show these distributions for different N^{gen}_{sig} hypotheses for ch1&2 BNV mode. These distributions show unbiased results for large numbers of signal events, however,

[1] This is a blind analysis, i.e., the signal region in the data sample is blinded.

bias is observed for the zero signal events hypothesis. This is due to the small total number of events in the signal region and is expected for such small statistics. As the number of events in the signal region becomes smaller, the 2D fits start failing more and more frequently. The failure rate can be quantified by the fraction of toy MC experiments in which the signals N_{sig}^{fit} extracted are below -8 (we set the lower bound of N_{sig} to -10). Take ch1&2 BNV mode as an example, in Fig. 35, a total of 5000 toy experiments are performed and N_{sig}^{fit} in 2727 of them are below -8, so the failure rate is 54.5%.

This problem can be partially alleviated by performing a fit to a single variable instead of multiple variables, which increases the expected number of events in the signal region. So in addition to 2D fits, we perform 1D fits to $M_{\rm bc}$. Fig. 36 shows the results of the 1D ensemble test for ch1&2 BNV modes. N_{sig}^{fit} in 657 out of 5000 toy MC experiments are below -8, so the failure rate is 13.1%, compared with 54.5% for 2D fits.

In the Appendix, we discuss the reasons for multiple bumps observed in the log-likelihood distributions in Fig. 35 and Fig. 36.

We then use a set of confidence intervals obtained from the frequentist method to determine the upper limit on the branching fraction.

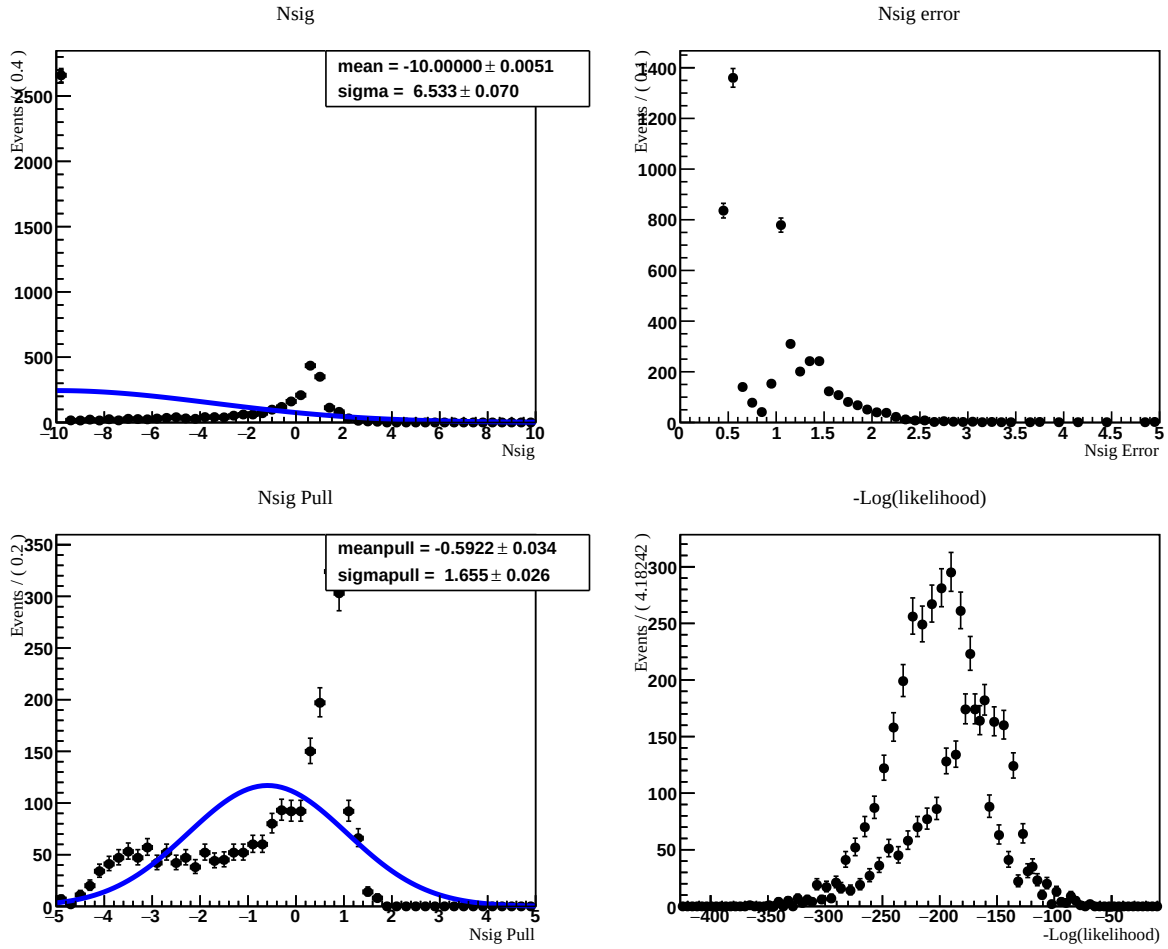


Figure 35: Zero signal events hypothesis ensemble test with 27 background events for ch1&2, BNV mode. 2D fit to $M_{\rm bc}$ and ΔE. 5000 toy MC experiments are performed and N_{sig}^{fit} in 2727 of them are below -8.

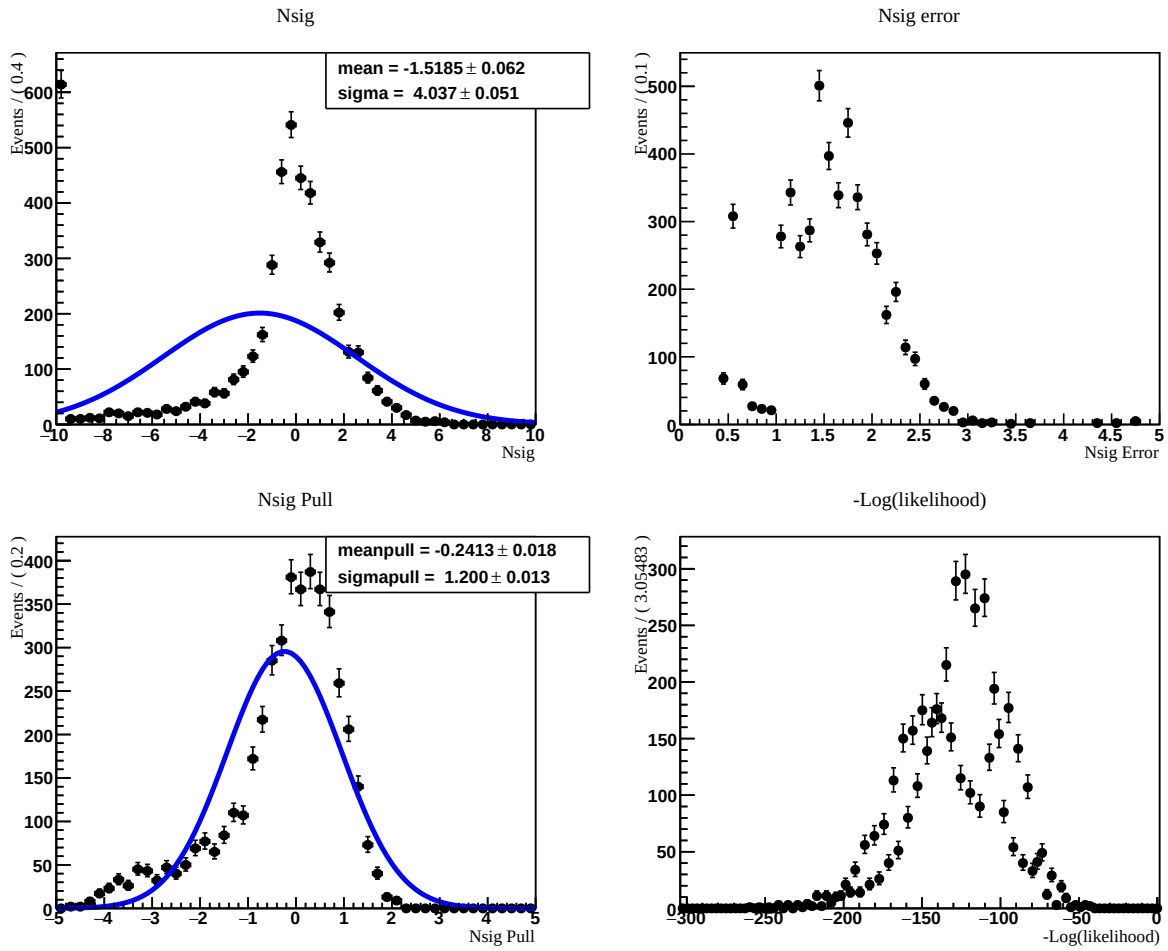


Figure 36: Zero signal events hypothesis ensemble test with 27 background events for ch1&2, BNV mode. 1D fit to $M_{\rm bc}$. 5000 toy MC experiments are performed and N_{sig}^{fit} in 657 of them are below -8.

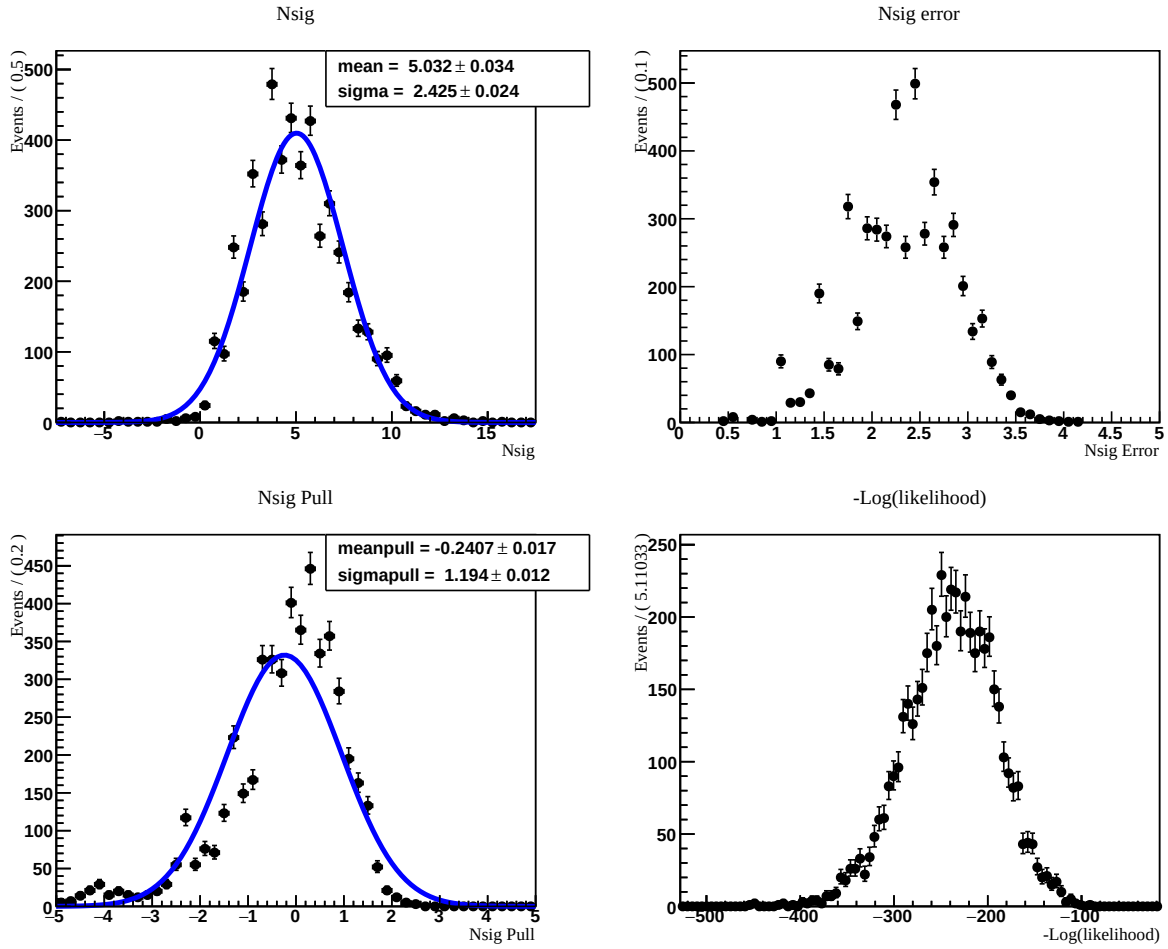


Figure 37: Five signal events hypothesis ensemble test with 27 background events for ch1&2, BNV mode. 2D fit to $M_{\rm bc}$ and ΔE. 5000 toy MC experiments are performed.

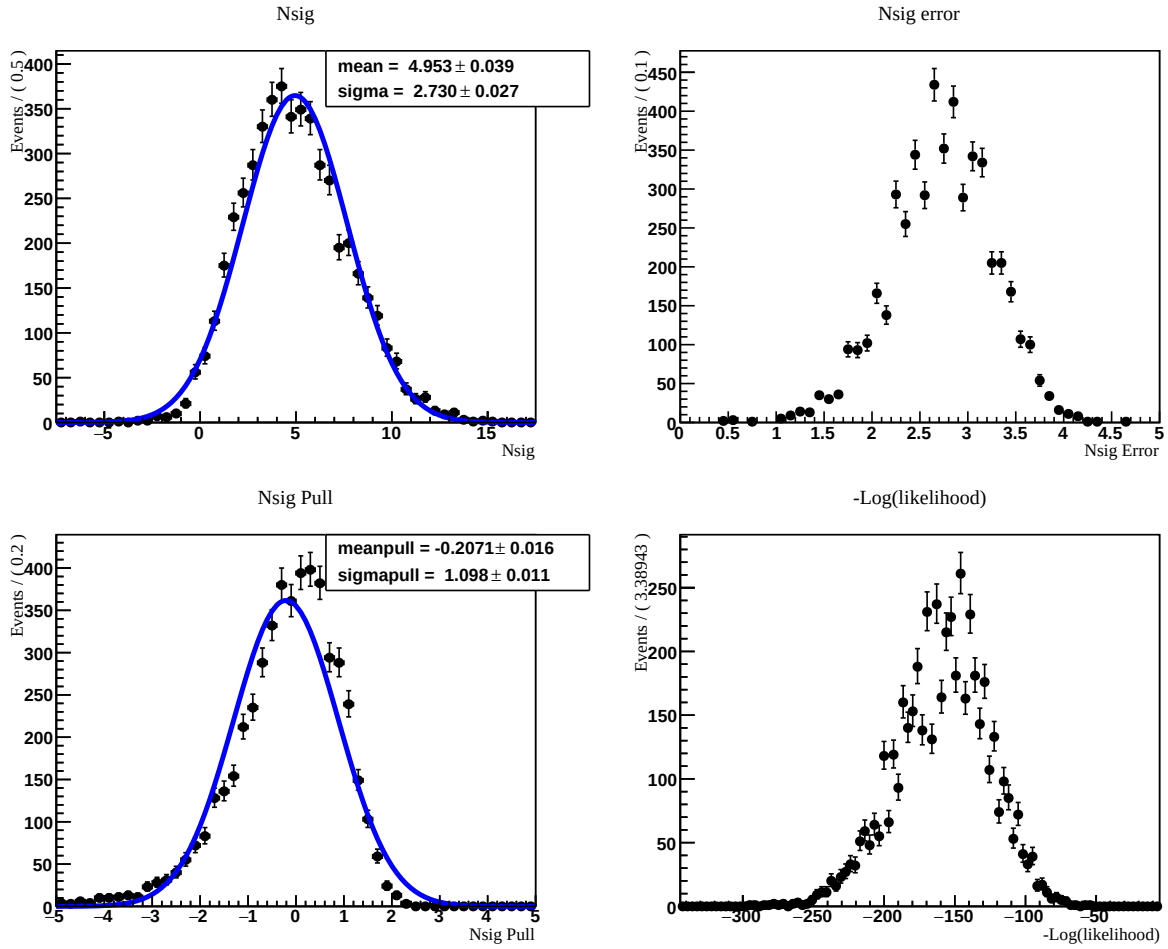


Figure 38: Five signal events hypothesis ensemble test with 27 background events for ch1&2, BNV mode. 1D fit to $M_{\rm bc}$. 5000 toy MC experiments are performed.

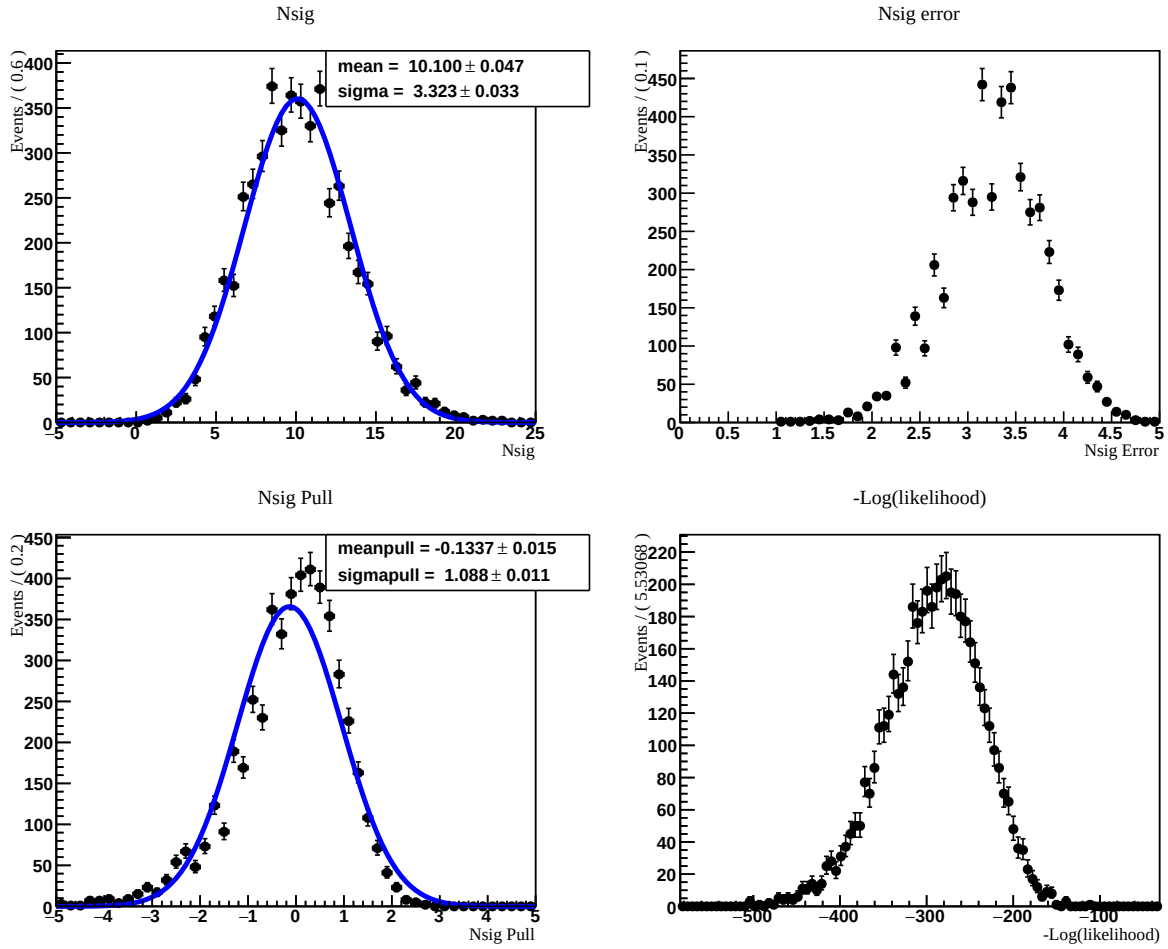


Figure 39: Ten signal events hypothesis ensemble test with 27 background events for ch1&2, BNV mode. 2D fit to $M_{\rm bc}$ and ΔE. 5000 toy MC experiments are performed.

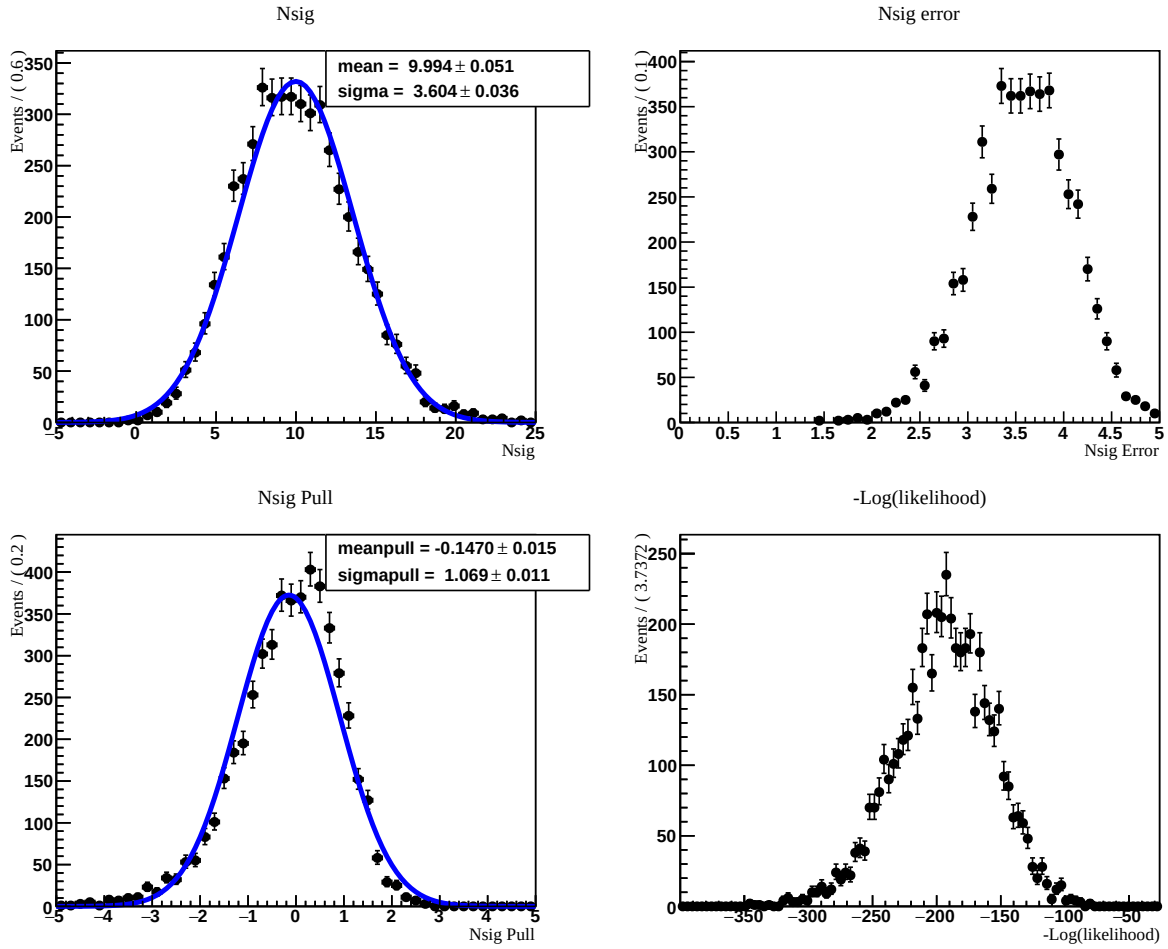


Figure 40: Ten signal events hypothesis ensemble test with 27 background events for ch1&2, BNV mode. 1D fit to $M_{\rm bc}$. 5000 toy MC experiments are performed.

4.5.1 Confidence Belt (2D fits)

The 90% confidence level belt is prepared using the frequentist method. The lower bounds for 90% CL belt are given by N_{sig}^{fit} with 5% of the results below this value. The upper bounds are given by N_{sig}^{fit} with 5% of the results above this value.

The belts shown in red color in Figures 41 through 43 show the confidence belts obtained from 2D fitting for SM and BNV modes.

The linearity of the fit model is also tested. For each fixed N_{sig}^{gen}, the N_{sig}^{fit} distribution is fit with a Gaussian to obtain the mean of the distribution, which is expected (for unbiased fits) to be close to N_{sig}^{gen}. The green line in the confidence belt shows the relationship between

the Gaussian mean for the number of signal events obtained from fitting and the number of signal events generated in toy MC experiments (N_{sig}^{gen}).

4.5.2 Confidence Belt (1D fits)

As we make the number of signal events in our toy MC experiments smaller, the 2D fits start failing more and more frequently. This happens because of very low statistics in the signal region. This problem can be partially alleviated by performing a fit to a single variable instead of multiple variables. Figures 36 to 40 show the ensemble test result for zero, five and ten signal events hypothesis using 1D $M_{\rm bc}$ fit. The problem of deviation from Gaussian is somewhat alleviated.

The blue belts in Figures 41 through 43 show the confidence belts with 1D fitting. While the red lines in the confidence belts show the result of the linearity test for 1D fitting.

Our strategy for this analysis is to start with a 1D $M_{\rm bc}$ fit. If a 1D fit shows a sign of possible signal (e.g., 2.5σ significance), we would also perform a 2D fit to improve significance for possible discovery.

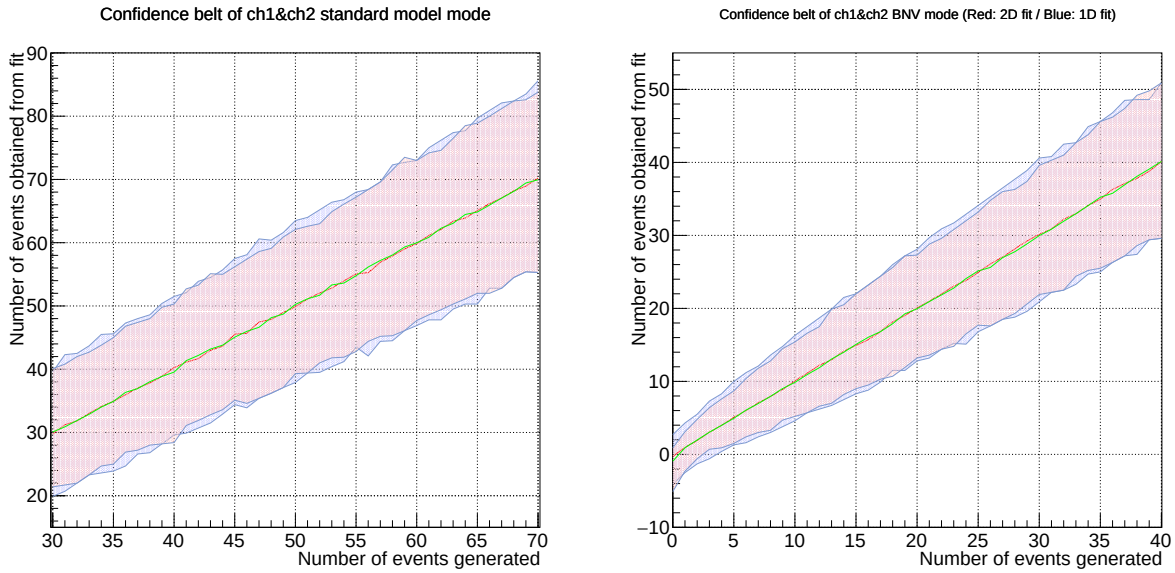


Figure 41: Confidence belts for ch1&2, SM and BNV modes.

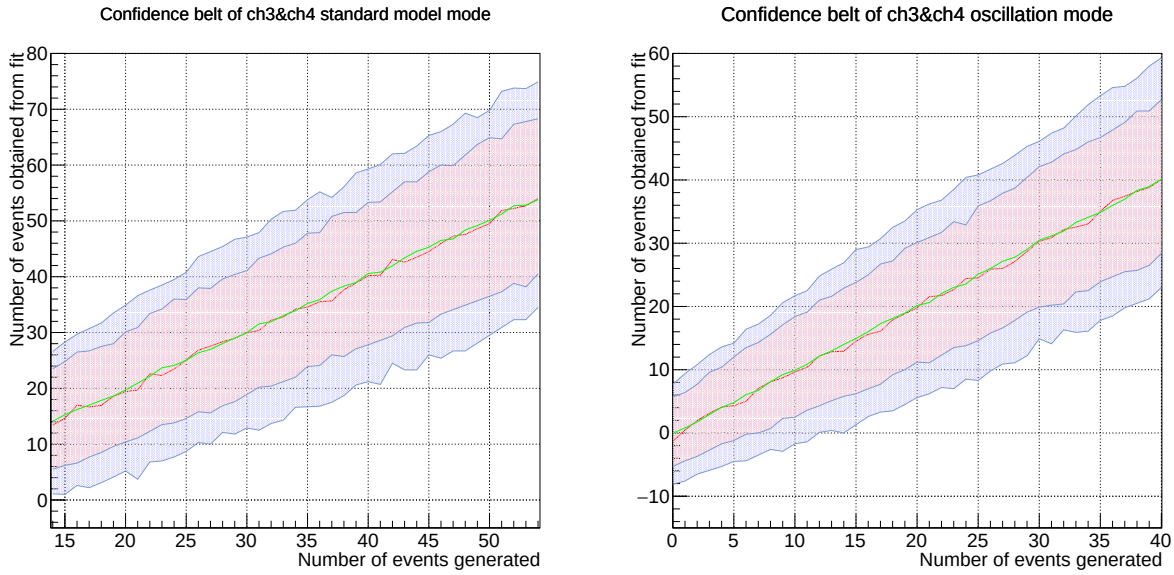


Figure 42: Confidence belts for ch3&4, SM and BNV modes.

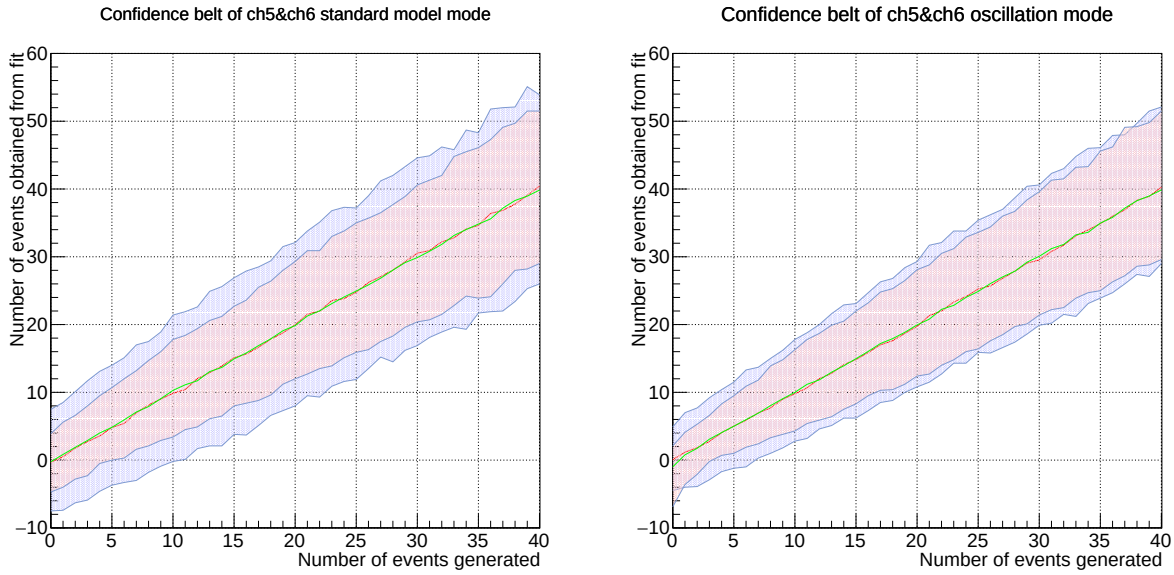


Figure 43: Confidence belts for ch5&6, SM and BNV modes.

4.6 BRANCHING FRACTION SYSTEMATICS IN CONTROL CHANNELS

In this section, we use all six SM modes as control channels to investigate the effects of different cuts on the branching fraction measurement.

In section 6.1, we compare the branching fractions measured with 60% efficiency, 50 MeV/c and without $p_\perp$ cuts to decide if $p_\perp$ cuts introduce a bias due to possible differences between data and MC.

In section 6.2, we compare the branching fraction systematics due to the $\chi^2/n.d.f < 100$ cuts.

4.6.1 Systematics of $p_\perp$ Cuts

Low transverse momenta tracks are associated with large systematic uncertainties. In this analysis, 60%-efficiency and 50 MeV/c $p_\perp$ cuts are applied to reduce the systematics.

4.6.1.1 Fitting Models With and Without $p_\perp$ Cuts The signal $M_{\rm bc}$ distribution is modeled with a Gaussian function and signal ΔE distribution is modeled with a double-Gaussian function using signal MC samples. Signal region $M_{\rm bc} > 5.27$ GeV/c$^2, |\Delta E| < 0.02$ GeV is taken as the fit range for signal PDF.

The background $M_{\rm bc}$ distribution is modeled with an ARGUS function and background ΔE is modeled with a first-order Chebychev polynomial using 6 streams of generic MC sample. $M_{\rm bc} > 5.2$ GeV/c$^2, |\Delta E| < 0.05$ GeV is taken as the fit range for background PDF.

Tables 15 through 17 show the fitting models with and without $p_\perp$ cuts. Different $\overline{\Lambda}_c^-$ tags are mixed.

Parameter	No $p_\perp$ cut	60% efficiency $p_\perp$ cut	50 MeV/c $p_\perp$ cut
$\mu_{M_{\rm bc}}$	5.279084 ± 0.000012	5.279087 ± 0.000014	5.279084 ± 0.000012
$\sigma_{M_{\rm bc}}$	$(2.5556 \pm 0.0089) \times 10^{-3}$	$(2.5472 \pm 0.01002) \times 10^{-3}$	$(2.5560 \pm 0.0089) \times 10^{-3}$
$\mu_{\Delta E}$	$(-5.4440 \pm 2.2946) \times 10^{-5}$	$(-4.0081 \pm 2.5273) \times 10^{-5}$	$(-5.4187 \pm 2.2946) \times 10^{-5}$
$Fraction$	0.2033 ± 0.01624	0.1940 ± 0.01856	0.2037 ± 0.01632
$\sigma_{1_{\Delta E}}$	$(8.3546 \pm 0.2663) \times 10^{-3}$	$(8.0284 \pm 0.2846) \times 10^{-3}$	$(8.3335 \pm 0.2650) \times 10^{-3}$
$\sigma_{2_{\Delta E}}$	$(3.9263 \pm 0.0465) \times 10^{-3}$	$(3.8632 \pm 0.0484) \times 10^{-3}$	$(3.9240 \pm 0.0448) \times 10^{-3}$
$ArgusParameter$	-31.97 ± 6.629	-39.32 ± 8.528	-32.59 ± 6.666
$ChebyParameter$	-0.1082 ± 0.1016	-0.3464 ± 0.1273	-0.1009 ± 0.1019

Table 15: Fitting models of ch1&2 with and without $p_\perp$ cuts.

Parameter	No $p_\perp$ cut	60% efficiency $p_\perp$ cut	50 MeV/c $p_\perp$ cut
$\mu_{M_{\rm bc}}$	5.279059 ± 0.000012	5.279071 ± 0.000014	5.279059 ± 0.000012
$\sigma_{M_{\rm bc}}$	$(2.5516 \pm 0.0088) \times 10^{-3}$	$(2.5370 \pm 0.0101) \times 10^{-3}$	$(2.5518 \pm 0.0088) \times 10^{-3}$
$\mu_{\Delta E}$	$(-5.5330 \pm 2.2890) \times 10^{-5}$	$(-8.6430 \pm 2.5810) \times 10^{-5}$	$(-5.4960 \pm 2.2900) \times 10^{-5}$
$Fraction$	0.1673 ± 0.01234	0.1559 ± 0.01381	0.1667 ± 0.01231
$\sigma_{1_{\Delta E}}$	$(9.6800 \pm 0.3994) \times 10^{-3}$	$(9.3870 \pm 0.4375) \times 10^{-3}$	$(9.6960 \pm 0.4022) \times 10^{-3}$
$\sigma_{2_{\Delta E}}$	$(4.0290 \pm 0.0396) \times 10^{-3}$	$(3.9550 \pm 0.0432) \times 10^{-3}$	$(4.0300 \pm 0.0396) \times 10^{-3}$
$ArgusParameter$	-17.52 ± 1.448	-17.86 ± 1.922	-17.44 ± 1.448
$ChebyParameter$	-0.1759 ± 0.02247	-0.2005 ± 0.02972	-0.1761 ± 0.02248

Table 16: Fitting models of ch3&4 with and without $p_\perp$ cuts.

Parameter	No $p_\perp$ cut	60% efficiency $p_\perp$ cut	50 MeV/c $p_\perp$ cut
$\mu_{M_{\rm bc}}$	5.279074 ± 0.000012	5.279069 ± 0.000015	5.279074 ± 0.000012
$\sigma_{M_{\rm bc}}$	$(2.6140 \pm 0.0082) \times 10^{-3}$	$(2.6010 \pm 0.01093) \times 10^{-3}$	$(2.6134 \pm 0.0082) \times 10^{-3}$
$\mu_{\Delta E}$	$(-7.8550 \pm 0.2288) \times 10^{-4}$	$(-7.1150 \pm 0.2888) \times 10^{-4}$	$(-7.8510 \pm 0.2288) \times 10^{-4}$
$Fraction$	0.221 ± 0.01191	0.1847 ± 0.01454	0.2201 ± 0.01185
$\sigma_{1_{\Delta E}}$	$(1.0520 \pm 0.0389) \times 10^{-2}$	$(1.0260 \pm 0.0526) \times 10^{-2}$	$(1.0550 \pm 0.0392) \times 10^{-2}$
$\sigma_{2_{\Delta E}}$	$(4.1660 \pm 0.0424) \times 10^{-3}$	$(4.0630 \pm 0.0050) \times 10^{-3}$	$(4.1680 \pm 0.0423) \times 10^{-3}$
$ArgusParameter$	-20.35 ± 2.566	-16.17 ± 4.371	-20.39 ± 2.569
$ChebyParameter$	-0.1649 ± 0.04052	-0.14 ± 0.06962	-0.1643 ± 0.04061

Table 17: Fitting models of ch5&6 with and without $p_\perp$ cuts.

4.6.1.2 SM Signal Extraction. Signal yields are extracted using the fitting models described in Section 6.1 for SM mode using full Belle data. Figures 44 through 46 show the results of signal extraction. Note that the fits shown in these figures are performed to a narrower range of $M_{\rm bc}$ and ΔE that will be used in our analysis to obtain the final results. Such ranges are used to perform the most meaningful comparison with the previous SM results. This is why the numbers of background events obtained from the fits shown here may appear to be inconsistent with the numbers of background events mentioned in Section 5 (Ensemble Tests). When the fits are performed in our full analysis region, the numbers are consistent with each other.

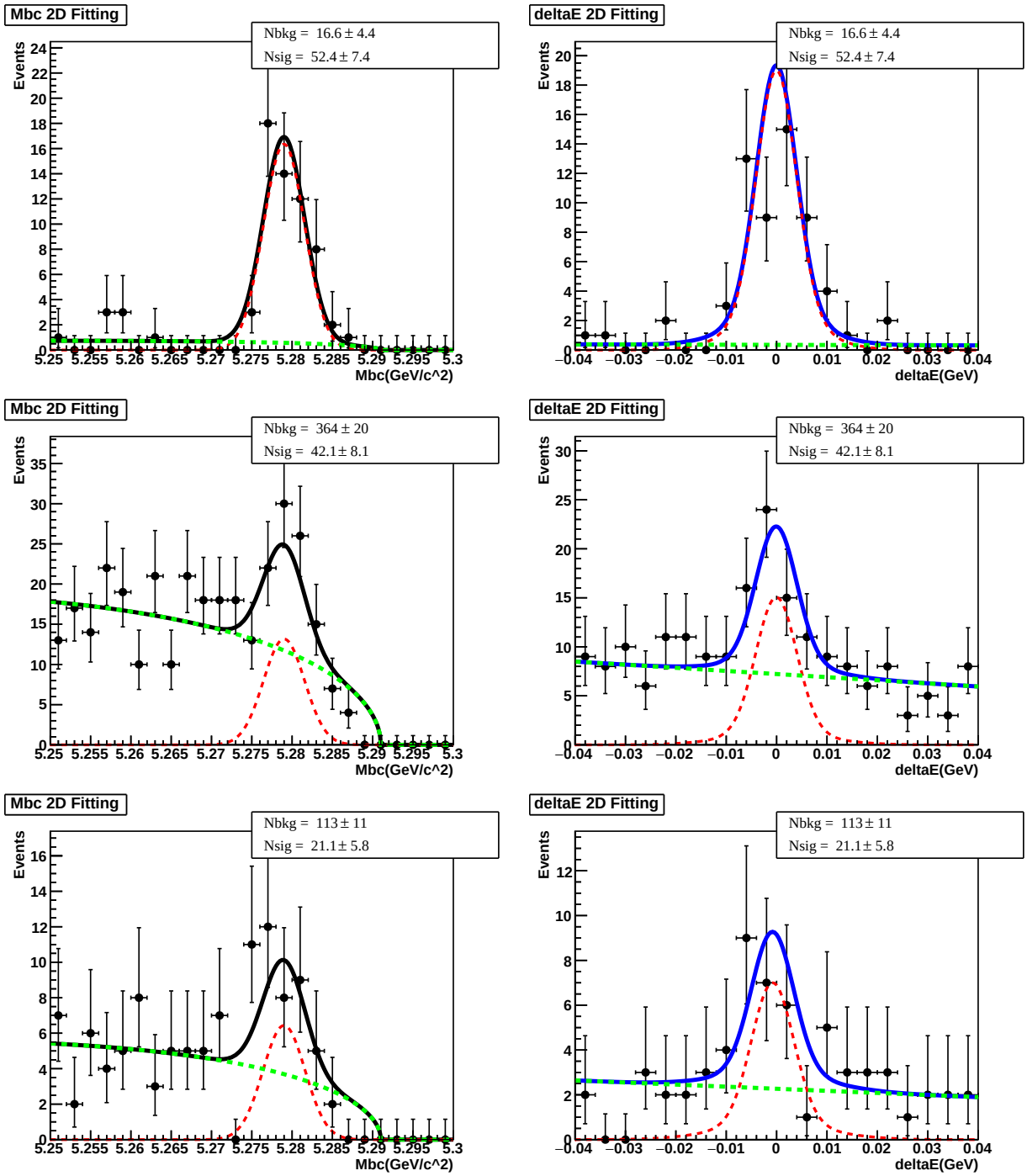


Figure 44: SM signal extraction for ch1&2, ch3&4, ch5&6 with $p_{\perp}$ >50 MeV/c cuts.

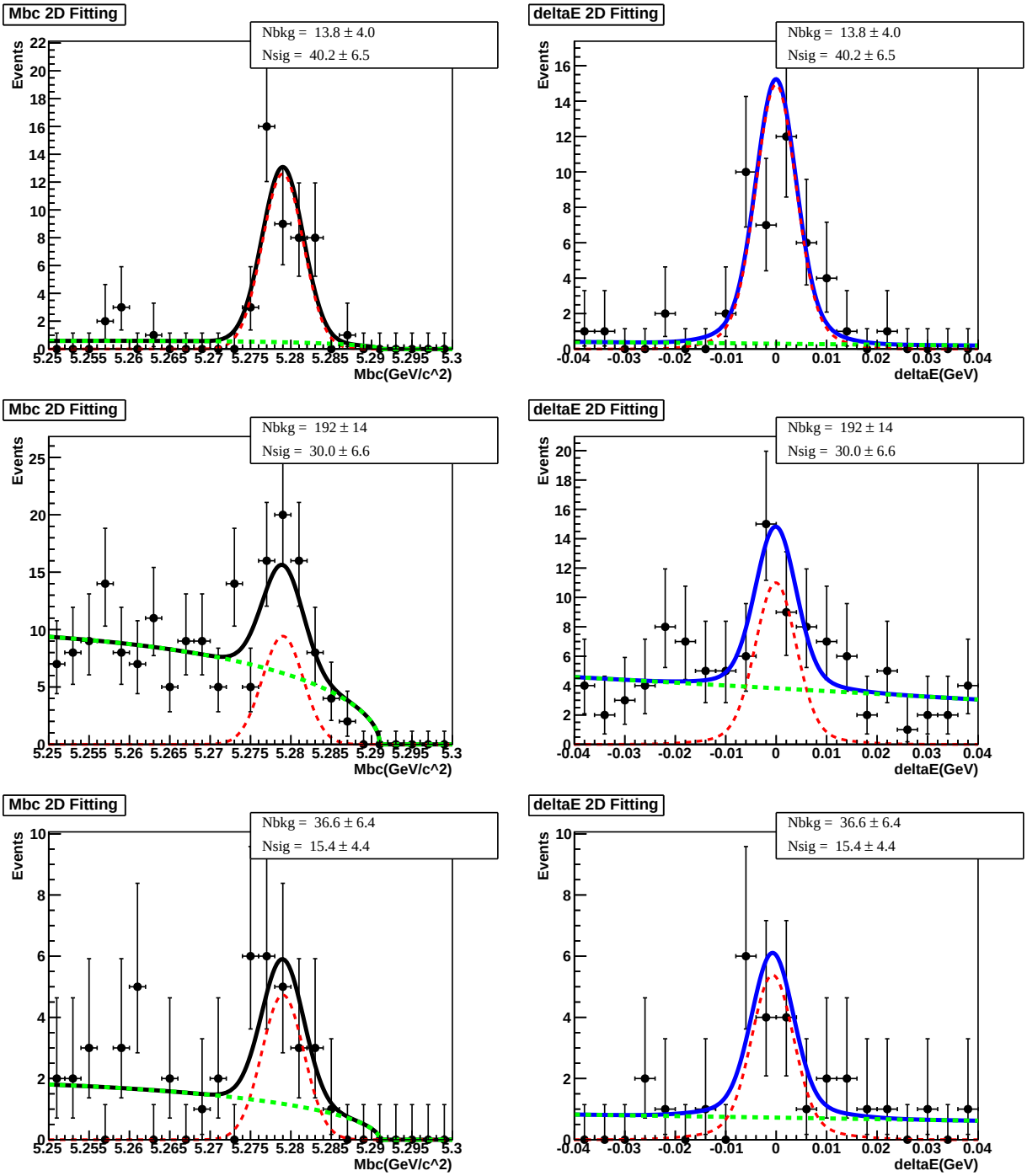


Figure 45: SM signal extraction for ch1&2, ch3&4, ch5&6 with 60% efficiency $p_\perp$ cuts.

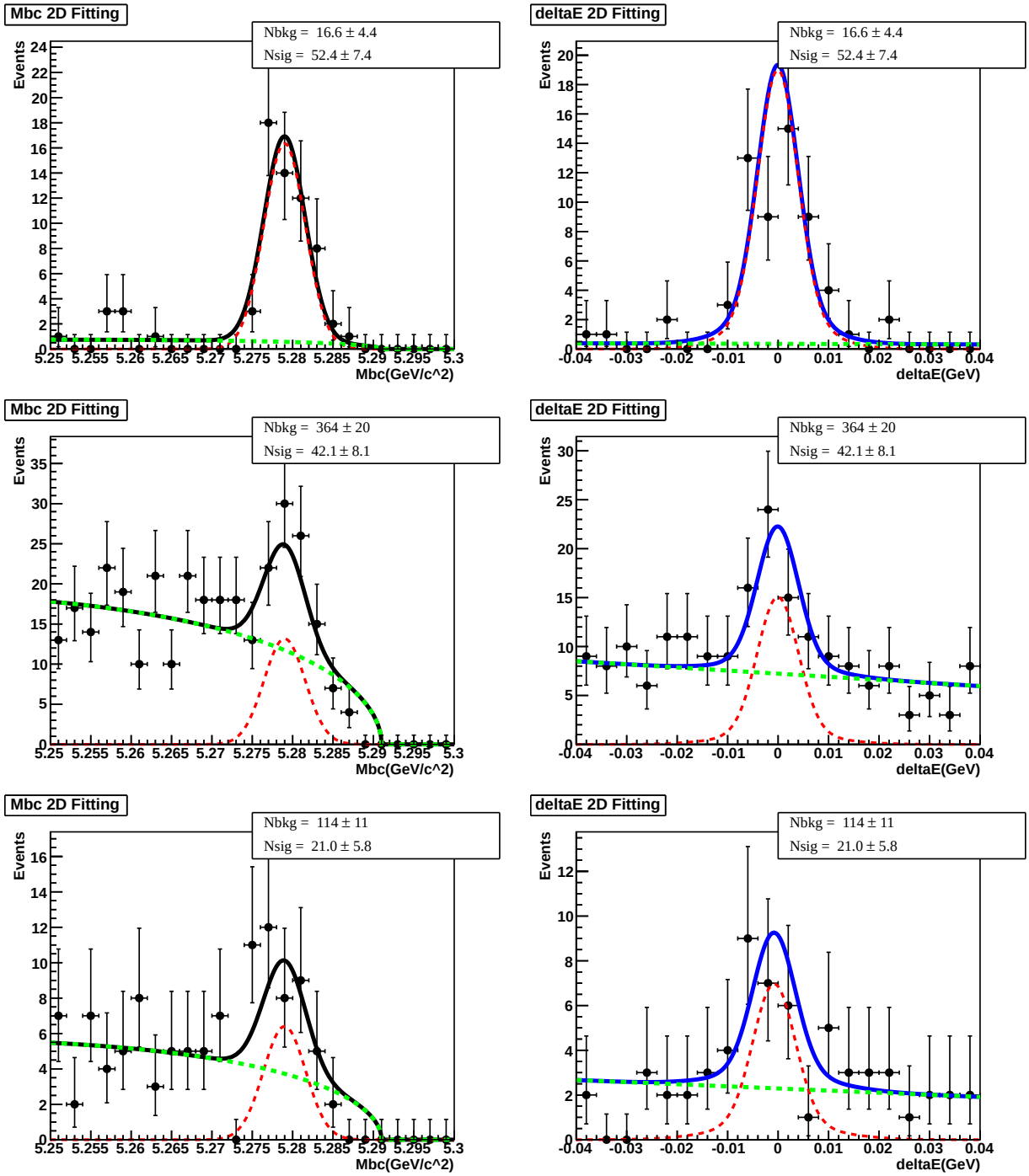


Figure 46: SM signal extraction for ch1&2, ch3&4, ch5&6 without $p_\perp$ cuts.

4.6.1.3 Measurement of Branching Fraction

The SM branching fraction for ch1&2 can be calculated as

$$\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)\mathcal{B}(\Xi_c^0 \to \Xi^- \pi^+) = \frac{N_{sig}}{N_{B^-/B^+}(\epsilon_1 \mathcal{B}_1 + \epsilon_2 \mathcal{B}_2)}$$

where $\mathcal{B}_1 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)\mathcal{B}(\Xi^- \to \Lambda\pi^-)\mathcal{B}(\Lambda \to p\pi^-)$, $\mathcal{B}_2 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K^+\pi^-)\mathcal{B}(\Xi^- \to \Lambda\pi^-)\mathcal{B}(\Lambda \to p\pi^-)$. ϵ_1 and ϵ_2 are reconstruction efficiencies for ch1 and ch2, respectively. N_{sig} is the number of signal events obtained from fitting. The number 771×10^6 B^+/B^- is used for this estimate.

The SM branching fraction for ch3&4 can be calculated as

$$\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)\mathcal{B}(\Xi_c^0 \to \Lambda K^- \pi^+) = \frac{N_{sig}}{N_{B^+/B^-}(\epsilon_3 \mathcal{B}_3 + \epsilon_4 \mathcal{B}_4)}$$

where $\mathcal{B}_3 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)\mathcal{B}(\Lambda \to p\pi^-)$, $\mathcal{B}_4 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K^+\pi^-)\mathcal{B}(\Lambda \to p\pi^-)$, ϵ_3 and ϵ_4 are reconstruction efficiencies for ch3 and ch4, respectively.

The SM branching fraction for ch5&6 can be calculated as

$$\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)\mathcal{B}(\Xi_c^0 \to p K^- K^- \pi^+) = \frac{N_{sig}}{N_{B^+/B^-}(\epsilon_5 \mathcal{B}_5 + \epsilon_6 \mathcal{B}_6)}$$

where $\mathcal{B}_5 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)$, $\mathcal{B}_6 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K^+\pi^-)$, ϵ_5 and ϵ_6 are reconstruction efficiencies for ch5 and ch6, respectively.

Tables 18 and 19 show the quantities used in the calculation.

	Parameter	Factor
1	$\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Xi}^+\pi^-)$	1.43%
2	$\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Lambda} K^+\pi^-)$	1.45%
3	$\mathcal{B}(\overline{\Xi}_c^0 \to \overline{p} K^+ K^+\pi^-)$	0.48%
4	$\mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K_S^0)$	1.59%
5	$\mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K^+\pi^-)$	6.28%
6	$\mathcal{B}(K_S^0 \to \pi^+\pi^-)$	69.2%
7	$\mathcal{B}(\bar{\Lambda} \to \overline{p}\pi^+)$	63.9%
8	$\mathcal{B}(\overline{\Xi}^+ \to \overline{\Lambda}\pi^+)$	99.9%

Table 18: Quantities used to measure the branching fractions.

Table 20 shows the branching fractions measured with, and without $p_\perp$ cuts and published result [33].

Reconstruction efficiency	No $p_\perp$ cut	60% efficiency $p_\perp$ cut	50 MeV/c $p_\perp$ cut
Channel 1	9.20%	8.07%	9.19%
Channel 2	7.71%	5.38%	7.70%
Channel 3	9.87%	8.23%	9.87%
Channel 4	8.05%	5.32%	8.05%
Channel 5	7.81%	4.87%	7.80%
Channel 6	6.59%	3.30%	6.58%

Table 19: Reconstruction efficiencies for SM mode with and without $p_\perp$ cuts.

	Ch1&2	Ch3&4	Ch5&6
Physical review letters, 122, 082001(2019)			
Signal Yield	44.8 ± 7.3	24.1 ± 5.5	16.6 ± 5.4
BR	$(1.71 \pm 0.28) \times 10^{-5}$	$(1.11 \pm 0.25) \times 10^{-5}$	$(5.47 \pm 1.78) \times 10^{-6}$
No $p_\perp$ cut			
Signal Yield	52.4 ± 7.4	42.1 ± 8.1	21.0 ± 5.8
BR	$(1.81 \pm 0.26) \times 10^{-5}$	$(1.39 \pm 0.27) \times 10^{-5}$	$(5.44 \pm 1.50) \times 10^{-6}$
60% efficiency $p_\perp$ cut			
Signal Yield	40.2 ± 6.5	30.0 ± 6.6	15.4 ± 4.4
BR	$(1.91 \pm 0.31) \times 10^{-5}$	$(1.43 \pm 0.32) \times 10^{-5}$	$(7.65 \pm 2.19) \times 10^{-6}$
50 MeV/c $p_\perp$ cut			
Signal Yield	52.4 ± 7.4	42.1 ± 8.1	21.1 ± 5.8
BR	$(1.82 \pm 0.26) \times 10^{-5}$	$(1.39 \pm 0.27) \times 10^{-5}$	$(5.48 \pm 1.51) \times 10^{-6}$

Table 20: SM branching fractions measured with and without $p_\perp$ cuts. $\chi^2/n.d.f < 100$ cuts are applied. Invariant mass cuts and $\chi^2/n.d.f < 15$ cuts are applied in published result [33].

4.6.2 Systematics of χ^2 Cuts

To suppress combinatorial background, cuts are applied to χ^2 from mass-vertex fit. Previous SM measurement applied cuts $\chi^2/n.d.f < 15$ to each reconstructed particle, where $n.d.f$ is the number of degrees of freedom for the mass-vertex fit. $\chi^2/n.d.f < 15$ cuts lead to around 20% efficiency loss (after all other cuts are applied).

In this analysis, we decided to use 95% efficient cuts $\chi^2/n.d.f < 100$ for Ξ_c^0, $\overline{\Lambda}_c^-$ and Ξ^- candidates. This is the cut value used in many Belle analyses.

Branching fractions measured with $\chi^2/n.d.f < 50$ are compared with the ones measured with $\chi^2/n.d.f < 100$ to study the systematics of χ^2 cuts.

The fitting models are established following the same conditions as described in the previous section, along with the SM signal extraction and branching fraction measurement. Table 21 shows the branching fractions measured with different χ^2 cuts. As we can see, χ^2 cuts lead to rather small changes in signal yields and measurement of branching fractions.

	Ch1&2	Ch3&4	Ch5&6
$\chi^2(\Xi_c^0, \overline{\Lambda}_c^-, \Xi^-) < 100$		$\chi^2(\Xi_c^0, \overline{\Lambda}_c^-) < 100$	
Signal Yield	40.2 ± 6.5	30 ± 6.6	15.4 ± 4.4
BR	$(1.91 \pm 0.31) \times 10^{-5}$	$(1.43 \pm 0.32) \times 10^{-5}$	$(7.65 \pm 2.19) \times 10^{-6}$
$\chi^2(\Xi_c^0, \Xi^-) < 100, \chi^2(\overline{\Lambda}_c^-) < 50$		$\chi^2(\Xi_c^0) < 100, \chi^2(\overline{\Lambda}_c^-) < 50$	
Signal Yield	40.1 ± 6.5	29.9 ± 6.5	16.6 ± 5.4
BR	$(1.92 \pm 0.31) \times 10^{-5}$	$(1.44 \pm 0.31) \times 10^{-5}$	$(7.53 \pm 2.16) \times 10^{-6}$
$\chi^2(\overline{\Lambda}_c^-, \Xi^-) < 100, \chi^2(\Xi_c^0) < 50$		$\chi^2(\overline{\Lambda}_c^-) < 100, \chi^2(\Xi_c^0) < 50$	
Signal Yield	40.3 ± 6.5	29.9 ± 6.4	15.2 ± 4.3
BR	$(1.94 \pm 0.31) \times 10^{-5}$	$(1.46 \pm 0.31) \times 10^{-5}$	$(7.81 \pm 2.21) \times 10^{-6}$
$\chi^2(\Xi_c^0, \overline{\Lambda}_c^-) < 100, \chi^2(\Xi^-) < 50$			
Signal Yield	38.4 ± 6.4		
BR	$(1.85 \pm 0.31) \times 10^{-5}$		
$\chi^2(\Xi_c^0, \overline{\Lambda}_c^-, \Xi^-) < 50$		$\chi^2(\Xi_c^0, \overline{\Lambda}_c^-) < 50$	
Signal Yield	38.4 ± 6.4	29.9 ± 6.4	14.7 ± 4.3
BR	$(1.89 \pm 0.31) \times 10^{-5}$	$(1.47 \pm 0.31) \times 10^{-5}$	$(7.64 \pm 2.21) \times 10^{-6}$

Table 21: SM branching fractions measured with different χ^2 cuts. Invariant mass cuts and $p_\perp$ cuts are applied.

4.7 UPPER LIMIT ESTIMATE FOR BNV BRANCHING FRACTION

4.7.1 Estimate Using ch1 and ch2

In this section, we estimate the upper limit for the BNV branching fraction using the confidence belt obtained in Chapter 4.5. First, we consider the ch1&2 in BNV mode. The branching fraction of the BNV mode can be calculated as

$$\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) = \frac{N_{sig}}{N_{B^-/B^+}(\epsilon_1 \mathcal{B}_1 + \epsilon_2 \mathcal{B}_2)}$$

where $\mathcal{B}_1 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Xi}^+\pi^-)\mathcal{B}(\overline{\Xi}^+ \to \overline{\Lambda}\pi^+)\mathcal{B}(\overline{\Lambda} \to \overline{p}\pi^+)$, $\mathcal{B}_2 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Xi}^+\pi^-)\mathcal{B}(\overline{\Xi}^+ \to \overline{\Lambda}\pi^+)\mathcal{B}(\overline{\Lambda} \to \overline{p}\pi^+)$. ϵ_1 and ϵ_2 are reconstruction efficiencies for ch1 and ch2, respectively, shown in Table 22. N_{sig} is the number of signal events obtained from fitting. The number 771×10^6 B^+/B^- is used for this estimate.

Assuming that we observe no events in the data we can now use the confidence belt shown in Fig. 41. The upper limit estimated from the confidence belt for this value is the height of the upper blue boundary at the value of 2 on the X-axis (*i.e.* roughly 2 events). We then calculate the 95% CL upper limit on the branching fraction using the information as

$$\begin{aligned}\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) &= \frac{2}{(771.6 \times 10^6) \times (0.0143) \times (0.999) \times (0.639) \times \{9.22\% \times (0.0159) \times (0.692) + 7.56\% \times (0.0628)\}} \\ &= 4.93 \times 10^{-5}\end{aligned}$$

4.7.2 Estimate Using ch3 and ch4

The branching fraction of ch3&4 in BNV mode can be calculated as

$$\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) = \frac{N_{sig}}{N_{B^+/B^-}(\epsilon_3 \mathcal{B}_3 + \epsilon_4 \mathcal{B}_4)}$$

where $\mathcal{B}_3 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Lambda} K^+\pi^-)\mathcal{B}(\overline{\Lambda} \to \overline{p}\pi^+)$, $\mathcal{B}_4 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Lambda} K^+\pi^-)\mathcal{B}(\overline{\Lambda} \to \overline{p}\pi^+)$, ϵ_3 and ϵ_4 are reconstruction efficiencies for ch3 and ch4, respectively, shown in Table 22.

Assuming that we observe no events in the data we can now use the confidence belt shown in Fig. 42. The upper limit estimated from the confidence belt for this value is the height of the upper blue boundary at the value of 7 on the X-axis (*i.e.* roughly 7 events). We then calculate the 95% CL upper limit on the branching fraction using the information as

$$\begin{aligned}\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) &= \frac{7}{(771.6 \times 10^6) \times (0.0145) \times (0.639) \times \{9.83\% \times (0.0159) \times (0.692) + 8.12\% \times (0.0628)\}} \\ &= 1.58 \times 10^{-4}\end{aligned}$$

4.7.3 Estimate Using ch5 and ch6

The branching fraction of ch5&6 in BNV mode can be calculated as

$$\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) = \frac{N_{sig}}{N_{B^+/B^-}(\epsilon_5 \mathcal{B}_5 + \epsilon_6 \mathcal{B}_6)}$$

where $\mathcal{B}_5 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p} K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{p}K^+K^+\pi^-)$, $\mathcal{B}_6 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{p}K^+K^+\pi^-)$, ϵ_5 and ϵ_6 are reconstruction efficiencies for ch5 and ch6, respectively, shown in Table 22.

Assuming that we observe no events in the data we can now use the confidence belt shown in Fig. 43. The upper limit estimated from the confidence belt for this value is the height of the upper blue boundary at the value of 4 on the X-axis (*i.e.* roughly 4 events). We then calculate the 95% CL upper limit on the branching fraction using the information as

$$\begin{aligned}\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) &= \frac{4}{(771.6 \times 10^6) \times (0.0048) \times \{8.15\% \times (0.0159) \times (0.692) + 6.90\% \times (0.0628)\}} \\ &= 2.07 \times 10^{-4}\end{aligned}$$

	Reconstruction efficiency
Channel 1	9.22%
Channel 2	7.56%
Channel 3	9.83%
Channel 4	8.12%
Channel 5	8.15%
Channel 6	6.90%

Table 22: Reconstruction efficiencies used to estimate the upper limit in data.

4.8 SIMULTANEOUS FIT TO ALL CHANNELS

Unbinned 2-dimensional extended maximum likelihood fits to $M_{\rm bc}$ and ΔE distributions are performed separately for ch1&2, ch3&4, and ch5&6. The likelihood functions are defined

as

$$\mathcal{L}_{12} = e^{-(N_{sig\_12}+N_{bkg\_12})} \prod_{i=1}^{N_{12}} (N_{sig\_12}\mathcal{P}_{sig\_12} + N_{bkg\_12}\mathcal{P}_{bkg\_12})$$

$$\mathcal{L}_{34} = e^{-(N_{sig\_34}+N_{bkg\_34})} \prod_{i=1}^{N_{34}} (N_{sig\_34}\mathcal{P}_{sig\_34} + N_{bkg\_34}\mathcal{P}_{bkg\_34})$$

$$\mathcal{L}_{56} = e^{-(N_{sig\_56}+N_{bkg\_56})} \prod_{i=1}^{N_{56}} (N_{sig\_56}\mathcal{P}_{sig\_56} + N_{bkg\_56}\mathcal{P}_{bkg\_56})$$

To perform a simultaneous fit to all six channels, the likelihood function is defined as

$$\mathcal{L}_{total} = \mathcal{L}_{12} \times \mathcal{L}_{34} \times \mathcal{L}_{56}$$

The common fitting variable is $\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)$, assuming branching fractions of sub-decays are known

$$N_{sig\_12} = \mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) \times (\epsilon_1 \mathcal{B}_1 + \epsilon_2 \mathcal{B}_2) \times N_{B^+/B^-}$$
$$N_{sig\_34} = \mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) \times (\epsilon_3 \mathcal{B}_3 + \epsilon_4 \mathcal{B}_4) \times N_{B^+/B^-}$$
$$N_{sig\_56} = \mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) \times (\epsilon_5 \mathcal{B}_5 + \epsilon_6 \mathcal{B}_6) \times N_{B^+/B^-}$$

where ϵ_i is reconstruction efficiency for channel i,

$$\mathcal{B}_1 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p}K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Xi}^+\pi^-)\mathcal{B}(\overline{\Xi}^+ \to \overline{\Lambda}\pi^+)\mathcal{B}(\overline{\Lambda} \to \overline{p}\pi^+)$$
$$\mathcal{B}_2 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Xi}^+\pi^-)\mathcal{B}(\overline{\Xi}^+ \to \overline{\Lambda}\pi^+)\mathcal{B}(\overline{\Lambda} \to \overline{p}\pi^+)$$
$$\mathcal{B}_3 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p}K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Lambda}K^+\pi^-)\mathcal{B}(\overline{\Lambda} \to \overline{p}\pi^+)$$
$$\mathcal{B}_4 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Lambda}K^+\pi^-)\mathcal{B}(\overline{\Lambda} \to \overline{p}\pi^+)$$
$$\mathcal{B}_5 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p}K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{p}K^+K^+\pi^-)$$
$$\mathcal{B}_6 = \mathcal{B}(\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-)\mathcal{B}(\overline{\Xi}_c^0 \to \overline{p}K^+K^+\pi^-)$$

4.8.1 Simultaneous Fit to Six SM Channels Using $\mathcal{B}(\Xi_c^0)$ Measured by Belle

To validate our fitting model, we perform a simultaneous fit to six SM channels and extract $\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)$ which we compare with the previous Belle measurement. For purposes of this comparison we use the Ξ_c^0 branching fractions measured by Belle (instead of PDG averages).

Fig. 47 shows the simultaneous fit result to six SM channels:

$$\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-) = (1.03 \pm 0.11) \times 10^{-3}, \tag{4.2}$$

while the current PDG value is

$$\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-) = (0.95 \pm 0.23) \times 10^{-3}. \tag{4.3}$$

Therefore, the result of our SM analysis agrees reasonably well with the previous measurement.

Note that the shown PDG value is actually from Belle, measured using the missing mass technique and the same data sample. The same previous analysis also measured Ξ_c^0 branching fractions. Therefore, our result is, indeed, supposed to be strongly correlated with PDG value, except for some differences in the analysis procedures.

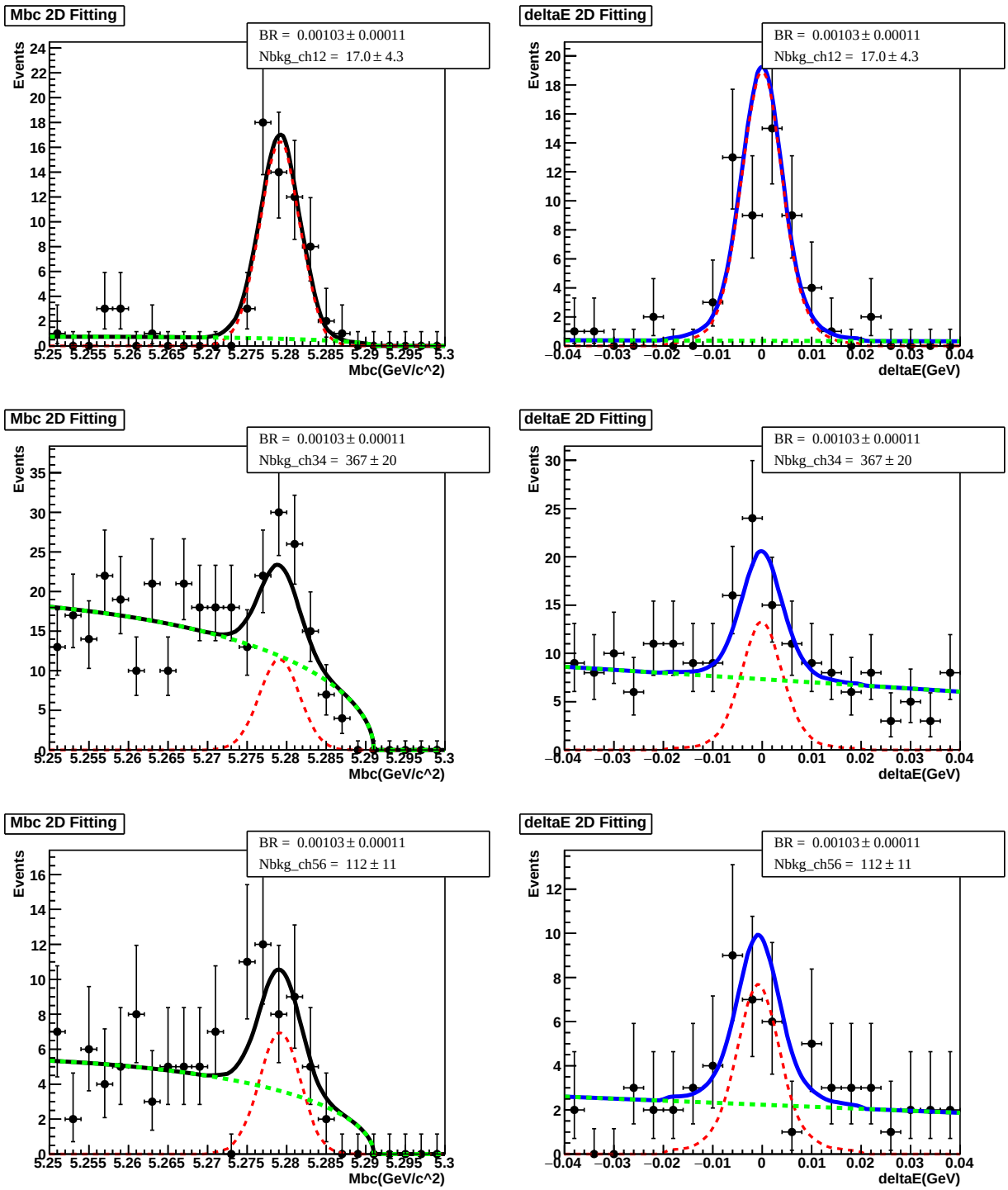


Figure 47: Simultaneous fit result for SM decay $\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)$

4.8.2 Pseudo-experiments: Simultaneous Fit to Six BNV Channels

Fig. 48 and 49 show the results of pseudo-experiments assuming 10% and 0% SM mode branching fraction, respectively.

Note that in a simultaneous fit for BNV channels, we use PDG values for Ξ_c^0 branching fractions instead of their values measured by Belle. Table. 23 shows the Ξ_c^0 branching fractions from Belle and PDG.

Branching fractions	PDG	Belle
$\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Xi}^+\pi^-)$	$(1.43 \pm 0.32)\%$	$(1.80 \pm 0.52)\%$
$\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Lambda}K^+\pi^-)$	$(1.45 \pm 0.33)\%$	$(1.17 \pm 0.38)\%$
$\mathcal{B}(\overline{\Xi}_c^0 \to \overline{p}K^+K^+\pi^-)$	$(0.48 \pm 0.12)\%$	$(0.58 \pm 0.24)\%$

Table 23: Ξ_c^0 branching fractions from Belle and PDG.

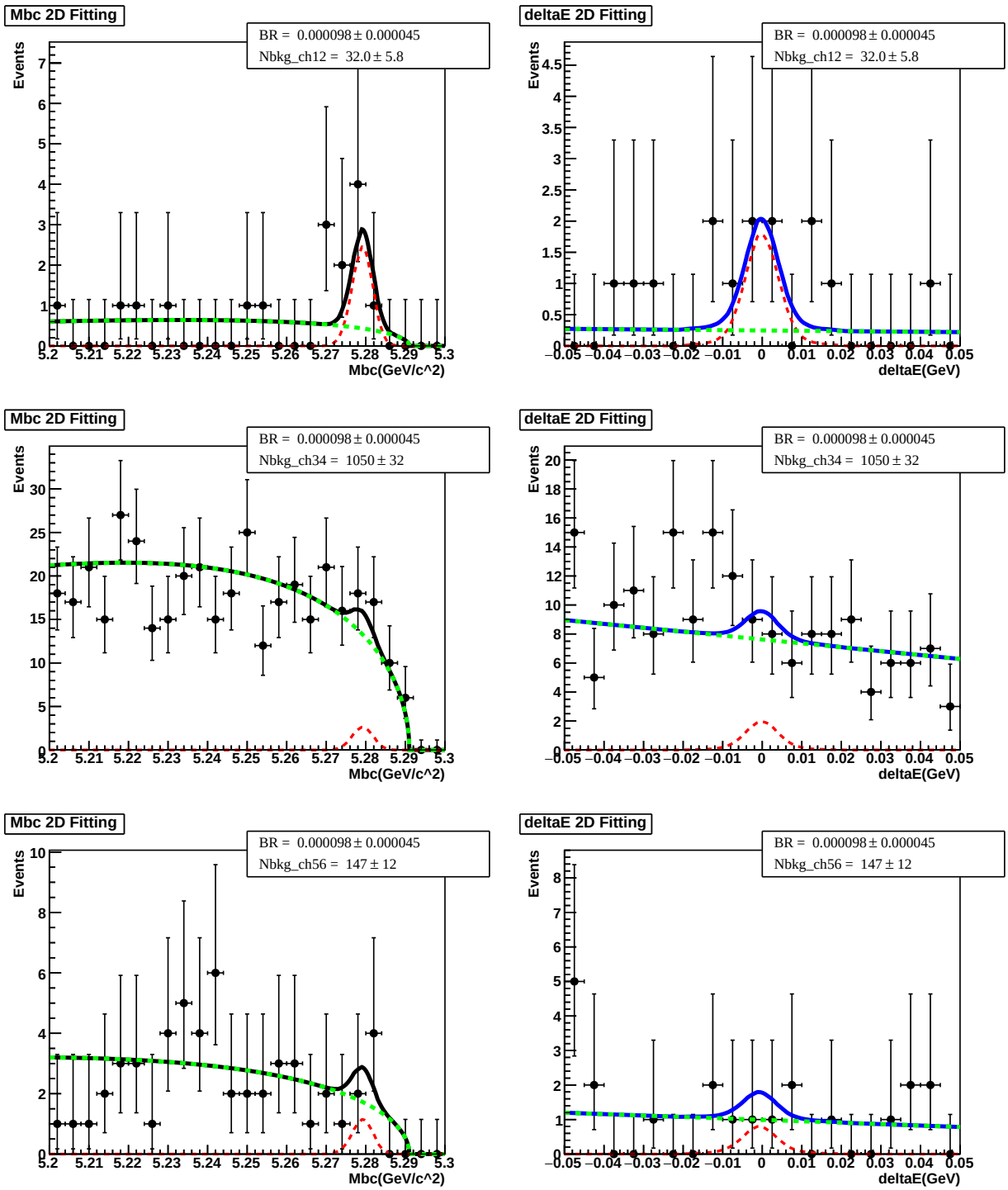


Figure 48: Pseudo-experiment assuming 10% branching fraction of SM mode: simultaneous fit result for BNV decay $\mathcal{B}(B^- \to \overline{\Xi}_c^{0}\overline{\Lambda}_c^{-})$.

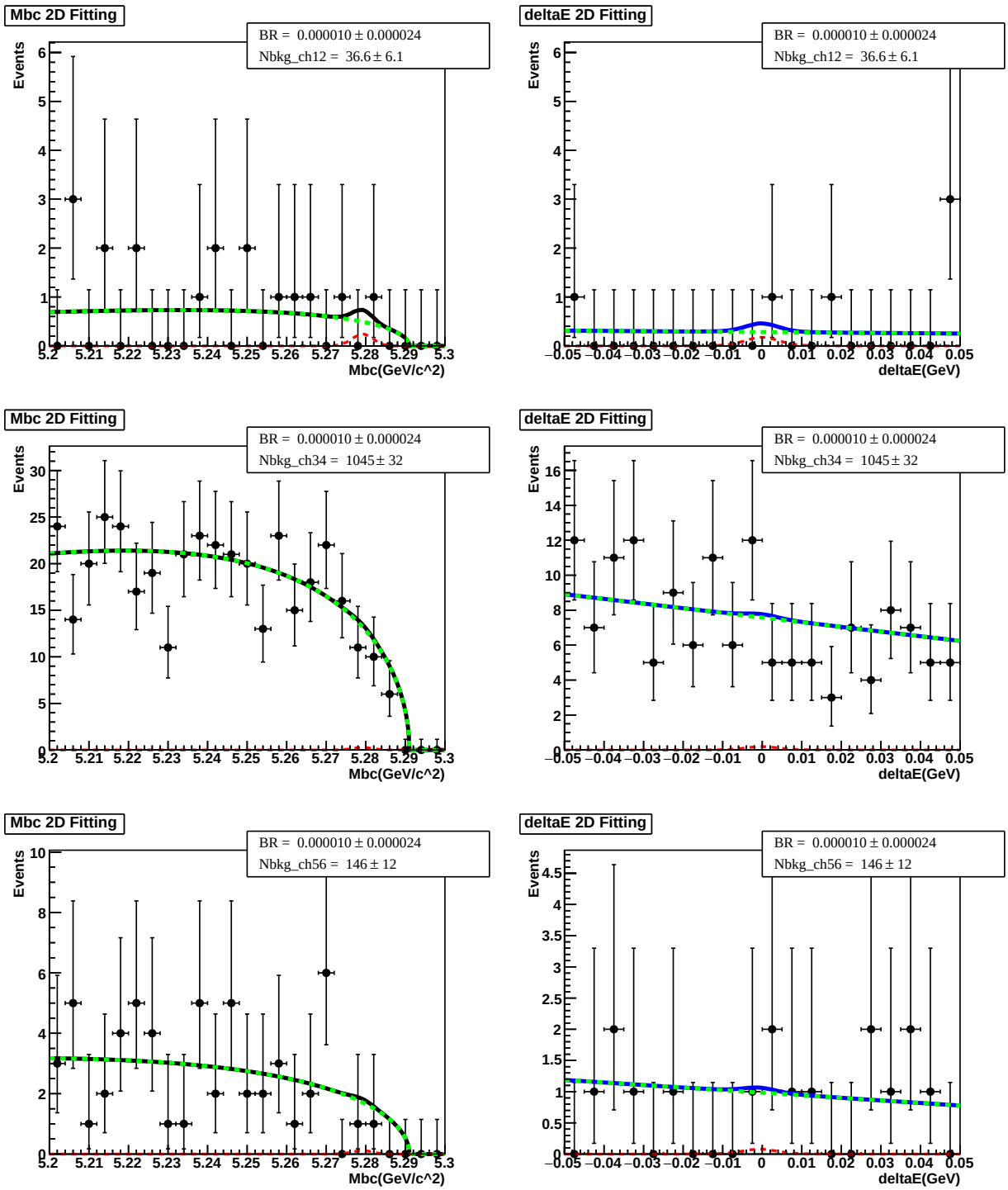


Figure 49: Pseudo-experiment assuming 0% branching fraction: simultaneous fit result for BNV decay $\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)$.

4.8.3 Ensemble test: Simultaneous Fit to Six BNV Channels

To test our fitting model for possible biases, we perform ensemble tests with toy MC experiments. Figures 50 through 55 show ensemble test results for 0/5/10% SM mode branching fraction hypotheses. 5000 toy MC experiments are performed for each hypothesis. 27/1058/143 background events are generated for ch1&2, ch3&4 and ch5&6, respectively. 2D fit to $M_{\rm bc}$ and ΔE, and 1D fit to $M_{\rm bc}$ are performed.

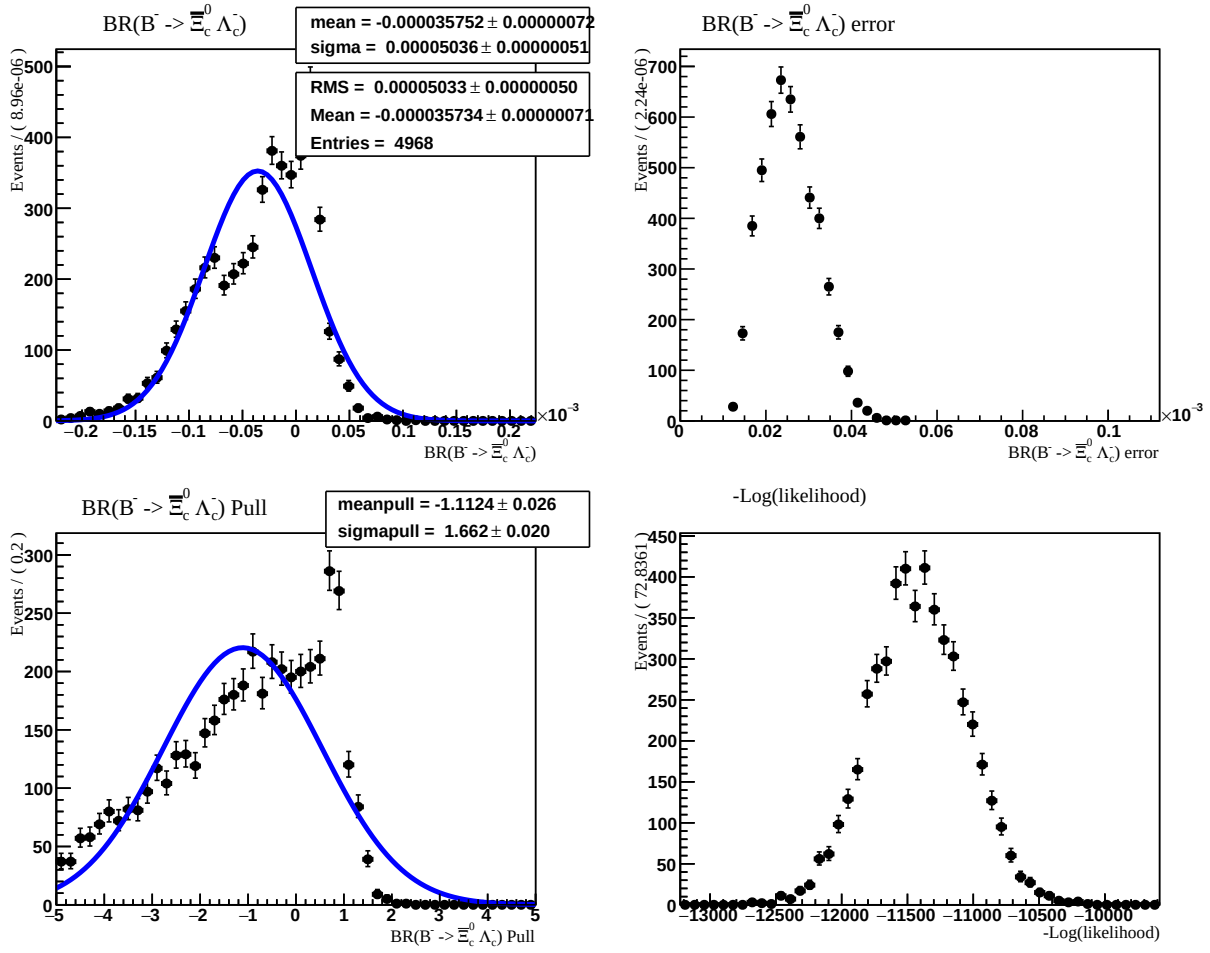


Figure 50: 0% branching fraction of SM mode hypothesis ensemble test for $\mathcal{B}(B^- \to \overline{\Xi}{}_c^0 \overline{\Lambda}{}_c^-)$. 2D fit to $M_{\rm bc}$ and ΔE. 5000 toy MC experiments are performed.

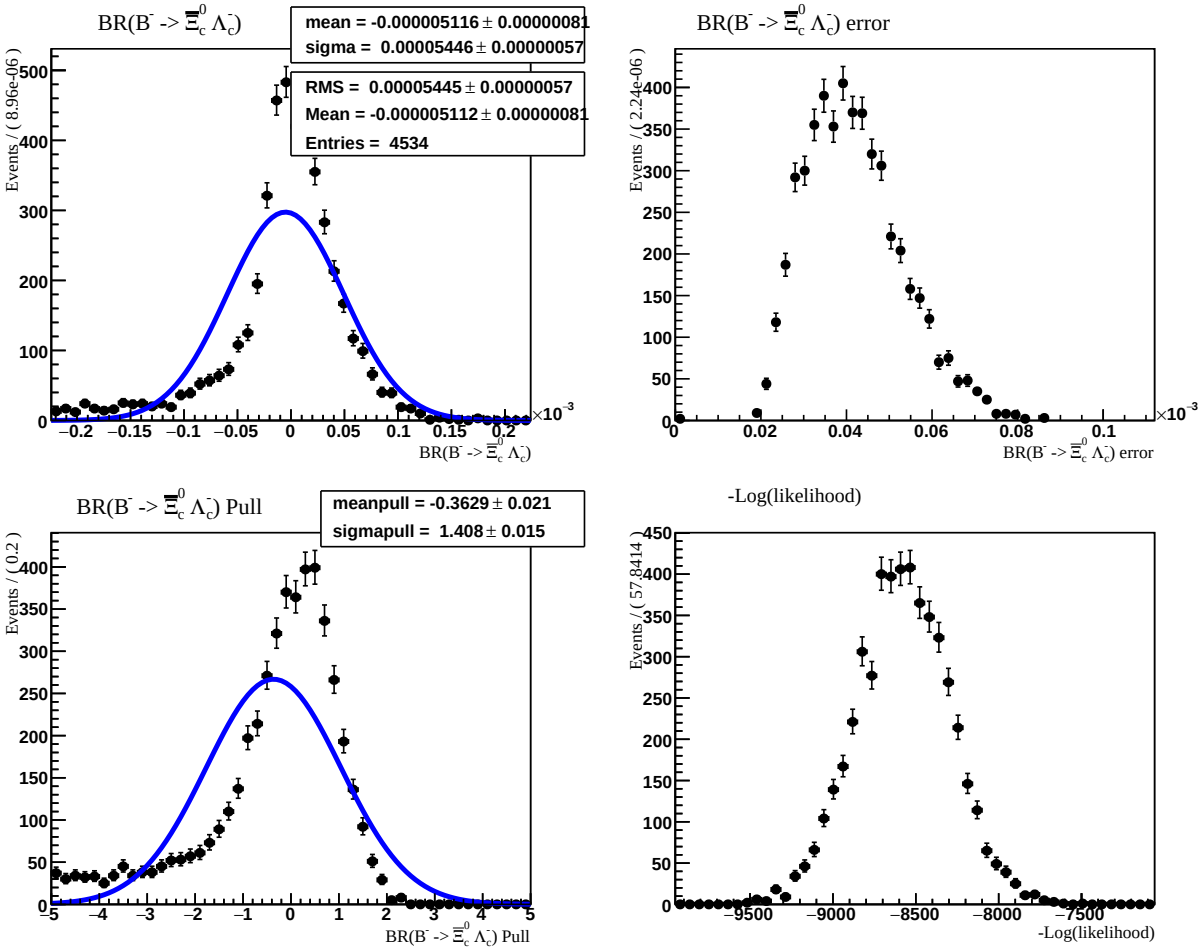


Figure 51: 0% branching fraction of SM mode hypothesis ensemble test for $\mathcal{B}(B^- \to \overline{\Xi}{}_c^0 \overline{\Lambda}{}_c^-)$. 1D fit to $M_{\rm bc}$. 5000 toy MC experiments are performed.

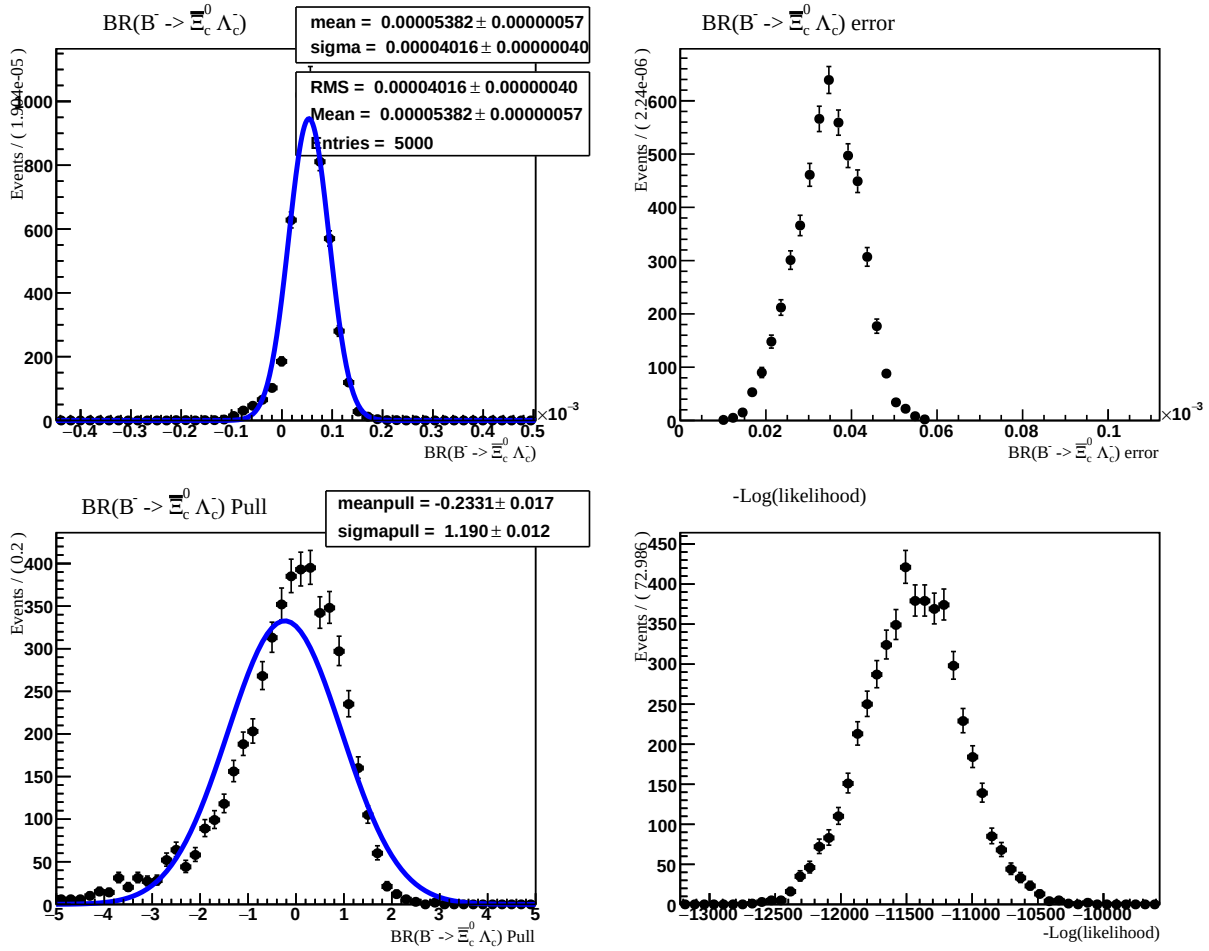


Figure 52: 5% branching fraction of SM mode hypothesis ensemble test for $\mathcal{B}(B^- \to \overline{\Xi}{}_c^0 \overline{\Lambda}{}_c^-)$. 2D fit to $M_{\rm bc}$ and ΔE. 5000 toy MC experiments are performed.

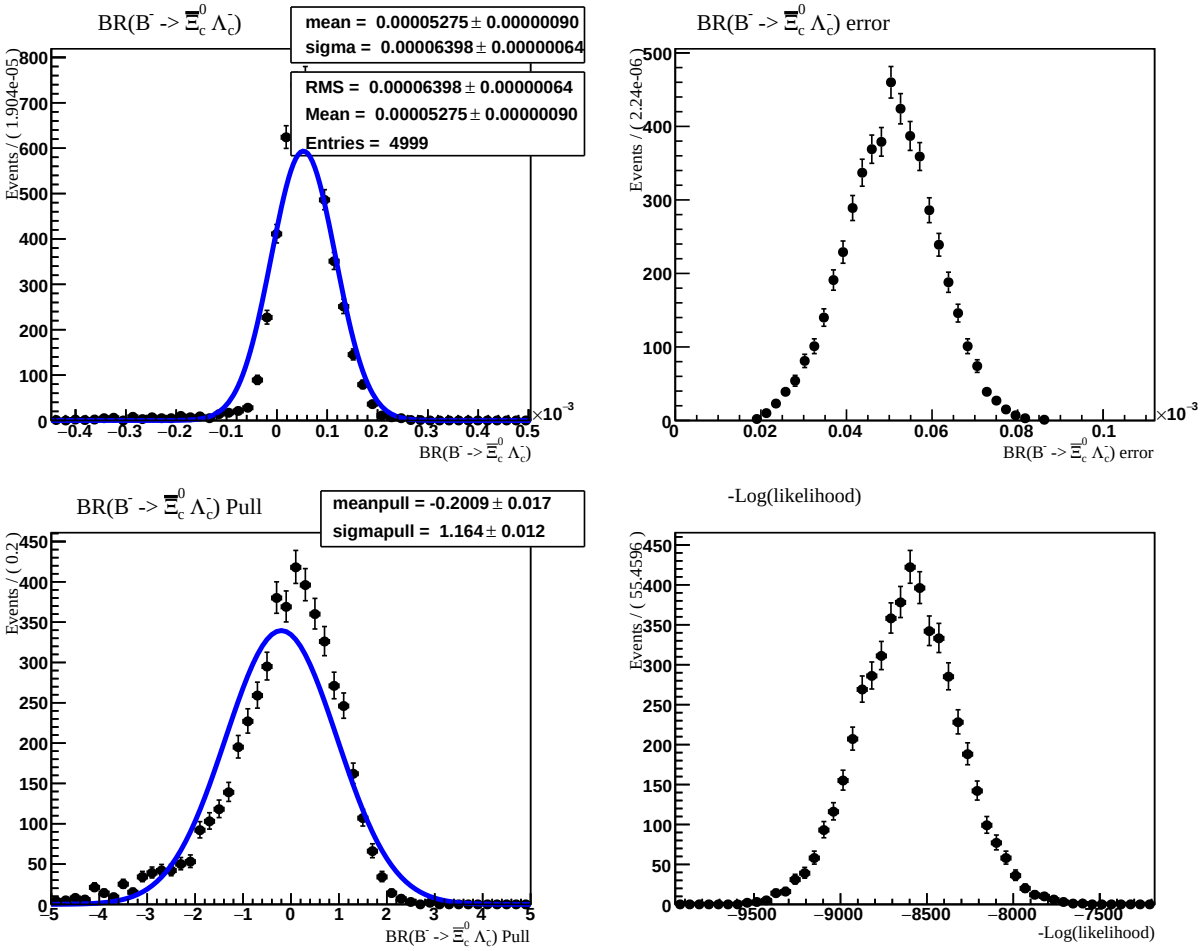


Figure 53: 5% branching fraction of SM mode hypothesis ensemble test for $\mathcal{B}(B^- \to \overline{\Xi}{}_c^0 \overline{\Lambda}{}_c^-)$. 1D fit to $M_{\rm bc}$. 5000 toy MC experiments are performed.

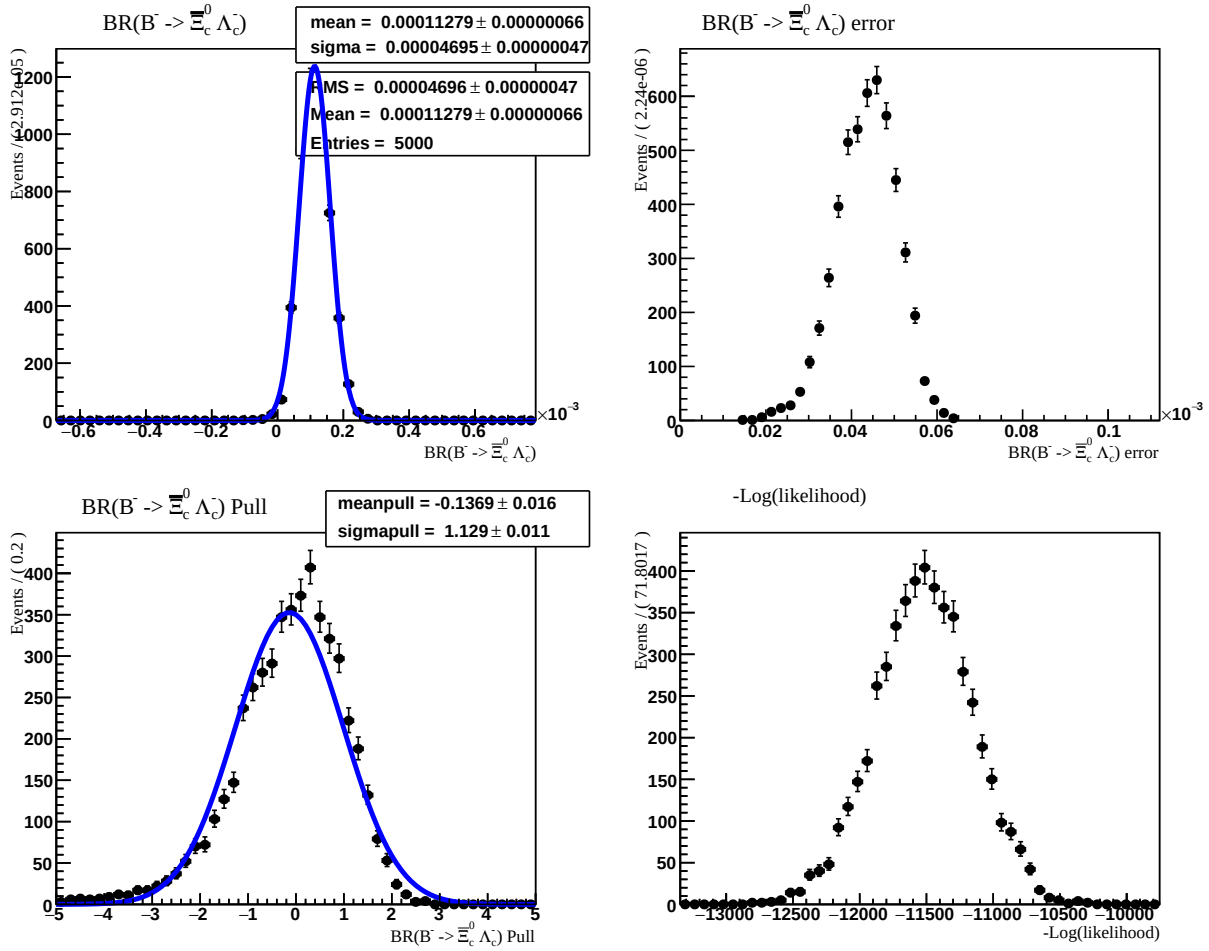


Figure 54: 10% branching fraction of SM mode hypothesis ensemble test for $\mathcal{B}(B^- \to \overline{\Xi}{}_c^0 \overline{\Lambda}{}_c^-)$. 2D fit to $M_{\rm bc}$ and ΔE. 5000 toy MC experiments are performed.

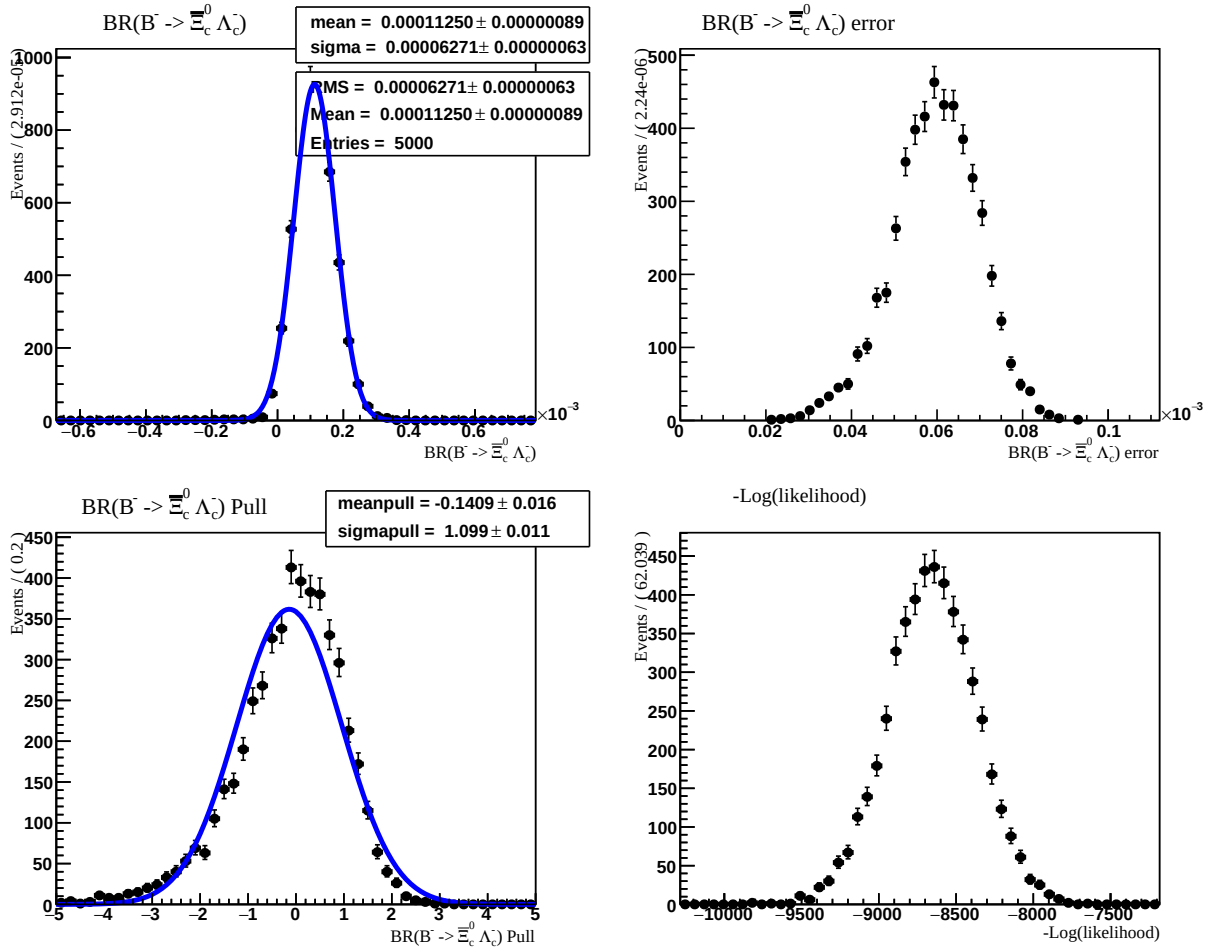


Figure 55: 10% branching fraction of SM mode hypothesis ensemble test for $\mathcal{B}(B^- \to \overline{\Xi}{}_c^0 \overline{\Lambda}{}_c^-)$. 1D fit to $M_{\rm bc}$. 5000 toy MC experiments are performed.

4.8.4 Confidence Belt: Simultaneous Fit to Six BNV Channels

Fig. 56 shows the 90% confidence level belt prepared using the frequentist method. The belt shown in red color in Figures 56 shows the confidence belts obtained from the 2D fit for BNV modes and the belt in blue color shows the confidence belts obtained from the 1D fit.

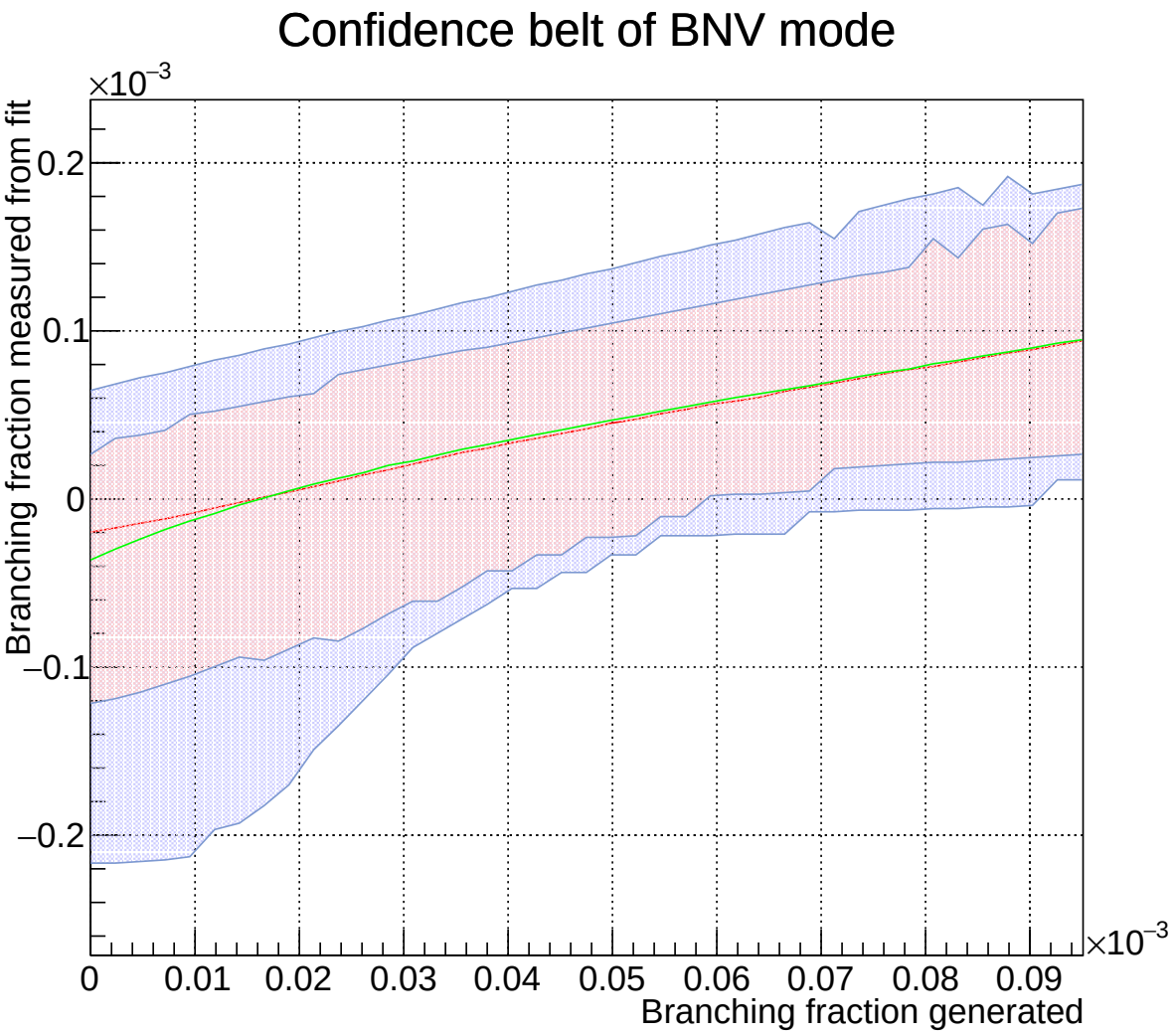


Figure 56: Confidence belt for BNV Decay $\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)$.

4.9 SYSTEMATIC UNCERTAINTIES

The following sources of systematic uncertainties will be considered in our final estimate of the BNV branching fraction upper limit.

4.9.1 Integrated Luminosity

1.4% systematic uncertainty is assigned due to the luminosity measurement.

4.9.2 Daughter Branching Fractions

Daughter branching fractions and the uncertainties associated.

Daughter decay	Branching fraction
$\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Xi}^+\pi^-)$	$(1.43 \pm 0.32)\%$
$\mathcal{B}(\overline{\Xi}_c^0 \to \overline{\Lambda}K^+\pi^-)$	$(1.45 \pm 0.33)\%$
$\mathcal{B}(\overline{\Xi}_c^0 \to \overline{p}K^+K^+\pi^-)$	$(0.48 \pm 0.12)\%$
$\mathcal{B}(\overline{\Lambda}_c^- \to \overline{p}K_S^0)$	$(1.59 \pm 0.08)\%$
$\mathcal{B}(\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-)$	$(6.28 \pm 0.32)\%$
$\mathcal{B}(K_S^0 \to \pi^+\pi^-)$	$(69.2 \pm 0.05)\%$
$\mathcal{B}(\bar{\Lambda} \to \overline{p}\pi^+)$	$(63.9 \pm 0.5)\%$
$\mathcal{B}(\overline{\Xi}^+ \to \bar{\Lambda}\pi^+)$	$(99.887 \pm 0.035)\%$

Table 24: Daughter branching fractions and uncertainties from PDG.

	Uncertainties due to daughter branching fractions
Ch1	22.9%
Ch2	23.0%
Ch3	23.3%
Ch4	23.3%
Ch5	25.5%
Ch6	25.5%

Table 25: Uncertainties associated with daughter branching fractions.

4.9.3 Particle Identification

In this section, we investigate the particle identification efficiencies and their uncertainties due to (binary) PID criteria applied to pion and proton candidates. We do not consider PID corrections for daughters of K_S^0 and Λ candidates since they are selected from standard BASF banks. Meanwhile, our electron identification requirement for kaon/pion candidates ($R_{e,hadron} < 0.95$) is very loose and of high efficiency (99%), so we do not consider efficiency correction and uncertainties due to electron identification.

The particle identification (PID) efficiencies and uncertainties are determined using the tools and tabulated data provided by the PID joint group [43]. Studies of systematics for kaon/pion and proton identification were performed using inclusive D^* and $\Lambda \to p\pi$ data samples, respectively.

The overall PID efficiency correction is the weighted sum of efficiency correction from the PID table over valid $p_{\rm lab}$ and $\cos\theta_{\rm lab}$ regions. Associated systematic uncertainty is estimated as

$$\sigma_{\rm PID} = \sqrt{\sigma^2_{\rm ratio} + \Delta^2_{\rm eff}}, \tag{4.4}$$

where $\sigma_{\rm ratio} = \sigma_{\rm syst} + \sigma_{\rm stat} + 0.003$, $\sigma_{\rm syst}$ and $\sigma_{\rm stat}$ are the weighted sum of systematic and statistical uncertainties for the data/MC efficiency ratio over all valid regions, and 0.003 is a constant term representing run-dependent effects. $\Delta_{\rm eff}$ is the difference between the weighted sum of efficiencies ratio for tracks in valid regions and the weighted sum of efficiencies ratio for all tracks (assuming MC is correct for invalid regions).

A weighted sum of PID corrections for SVD1 and SVD2 data samples (ρ_1 and ρ_2) is evaluated to estimate total PID correction ρ:

$$\rho = \frac{140\epsilon_1\rho_1 + 571\epsilon_2\rho_2}{140\epsilon_1 + 571\epsilon_2}, \tag{4.5}$$

where the numbers $140\epsilon_1$ and $571\epsilon_2$ are the products of reconstruction efficiencies and integrated luminosities for SVD1 and SVD2 data samples.

Table 26 shows the PID efficiency corrections and uncertainties for all six channels. The overall PID efficiency correction for each channel is the product of the correction factors for all corresponding charged tracks. The overall PID systematics is the sum of systematic uncertainties for all corresponding charged tracks. PID corrections and uncertainties for proton candidates are the weighted sums for positive and negative tracks.

In conclusion, we assign 2.6%, 5.1%, 2.9%, 5.4%, 6.0%, and 8.6% uncertainties for our channels 1 through 6, respectively.

Ch1	Correction	Uncertainty
π from Ξ	0.9654	0.0096
π from Ξ_c^0	0.9997	0.0126
p from Λ_c	0.9776	0.0032
Total	0.9434	0.0255

Ch2	Correction	Uncertainty
π from Ξ	0.9652	0.0096
π from Ξ_c^0	0.9997	0.0127
π from Λ_c	0.9862	0.0122
K from Λ_c	1.0070	0.0132
p from Λ_c	0.9704	0.0031
Total	0.9299	0.0507

Ch3	Correction	Uncertainty
π from Ξ_c^0	0.9889	0.0122
K from Ξ_c^0	1.0072	0.0132
p from Λ_c	0.9783	0.0032
Total	0.9744	0.0287

Ch4	Correction	Uncertainty
π from Ξ_c^0	0.9884	0.0122
K from Ξ_c^0	1.0073	0.0133
π from Λ_c	0.9863	0.0121
K from Λ_c	1.0069	0.0131
p from Λ_c	0.9713	0.0031
Total	0.9603	0.0539

Ch5	Correction	Uncertainty
p from Ξ_c^0	0.9681	0.0034
K from Ξ_c^0	1.0130	0.0206
K from Ξ_c^0	1.0130	0.0206
π from Ξ_c^0	0.9994	0.0125
p from Λ_c	0.9785	0.0033
Total	0.9715	0.0604

Ch6	Correction	Uncertainty
p from Ξ_c^0	0.9692	0.0033
K from Ξ_c^0	1.0132	0.0208
K from Ξ_c^0	1.0132	0.0208
π from Ξ_c^0	0.9993	0.0123
π from Λ_c	0.9858	0.0122
K from Λ_c	1.0070	0.0132
p from Λ_c	0.9716	0.0031
Total	0.9590	0.0857

Table 26: PID correction factors and relative systematic uncertainties for six analysis channels.

4.9.4 Charged Track Reconstruction

0.35% systematic uncertainty is assigned for each charged track of transverse momentum above 200 MeV/c [44]. 1.2% systematic uncertainty is assigned for each charged track of transverse momentum below 200 MeV/c [44]. The overall uncertainty for each track is the weighted average of the uncertainty for tracks which p_T below and above 200 MeV/c.

For example, ch1 has three charged tracks (we consider tracks from K_S^0 and Λ seperately), π from Ξ, π from Ξ_c^0 and p from Λ_c. 45% of π from Ξ tracks are below 200 MeV/c, so the overall uncertainty we assign is $0.45 * 1.2\% + 0.55 * 0.35\% = 0.73\%$. All π from Ξ_c^0 and p from Λ_c tracks are above 200 MeV/c, so we assign 0.35% uncertainty for both tracks. The total tracking uncertainty assigned to ch1 is 1.43%.

Table 27 shows the tracking reconstruction systematic uncertainties for all six channels.

	Tracking uncertainty
Channel 1	1.43%
Channel 2	2.29%
Channel 3	1.21%
Channel 4	2.07%
Channel 5	2.41%
Channel 6	3.27%

Table 27: Track reconstruction uncertainties for six channels.

4.9.5 K_S^0 Reconstruction

1.57% systematic uncertainty is assigned for K_S^0 reconstruction using NISKSFINDER package [45].

4.9.6 Λ Reconstruction

According to a previous study [46], uncertainty associated with Λ reconstruction would not be larger than 3%, so we assign 3% systematic uncertainty for Λ reconstruction.

4.9.7 PDF Parametrization

Systematic uncertainties due to the signal and background PDF parametrization are estimated by measuring the change in the signal yield while varying every PDF parameter by $\pm 1\sigma$.

4.9.8 MC Statistics

Reconstruction efficiency uncertainty is determined as $\sqrt{\epsilon(1-\epsilon)/N}$, where ϵ is the reconstruction efficiency, and N is the number of generated signal MC events.

4.9.9 Summary of Systematic Uncertainties

Systematics sources	ch1	ch2	ch3	ch4	ch5	ch6
Luminosity	1.4%					
MC statistics	0.68%	0.63%	0.65%	0.60%	0.58%	0.52%
Daughter branching fraction	22.9%	23.0%	23.3%	23.3%	25.5%	25.5%
Tracking	1.43%	2.29%	1.21%	2.07%	2.41%	3.27%
PID	2.6%	5.1%	2.9%	5.4%	6.0%	8.6%
K_S^0	1.57%	0%	1.57%	0%	1.57%	0%
Λ	3%	3%	3%	3%	0%	0%

Table 28: Summary table of uncertainties for six channels.

4.9.10 Systematic Uncertainties in the Ratio of the Branching Fractions for the BNV and SM Modes

We plan to present the results as the ratio of the branching fractions for the BNV and SM modes $\frac{\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)}$. In this approach, most of the systematic uncertainties discussed in Section 9.9 cancel. We discuss the possible remaining resources of systematic uncertainties in this section.

4.9.10.1 Uncertainty due to finite MC statistics The first remaining systematic uncertainty would be due to finite MC statistics. Possible subtle differences between the overall detection efficiencies for the SM and BNV modes are vanishingly small. The ratio of efficiencies is defined as $R = \frac{\epsilon_1}{\epsilon_2} = \frac{N_{s1}/N_1}{N_{s2}/N_2}$, where ϵ_1 and ϵ_2 are reconstruction efficiencies for BNV and SM modes, respectively.

We consider the effect of uncorrelated N_{s1} and N_{s2} only:

$$\begin{aligned} dR^2 &= \left(\frac{N_2}{N_1}\right)^2 * \left[\frac{1}{N_{s2}^2} * dN_{s1}^2 + \left(\frac{N_{s1}}{N_{s2}^2}\right)^2 * dN_{s2}^2\right] \\ &= \left(\frac{\epsilon_1}{\epsilon_2}\right)^2 * \left(\frac{1}{N_{s1}} + \frac{1}{N_{s2}}\right) \\ &= \left(\frac{\epsilon_1}{\epsilon_2}\right)^2 * \left(\frac{1}{\epsilon_1 * N1} + \frac{1}{\epsilon_2 * N2}\right) \\ \frac{dR}{R} &= \sqrt{\frac{1}{\epsilon_1 * N_1} + \frac{1}{\epsilon_2 * N_2}} \end{aligned}$$

There are 150069 events for SM mode and 152199 events for BNV mode for all 6 channels after applying all selection criteria, then the relative systematics $dR/R = \sqrt{1/150069 + 1/152199} = 0.36\%$. We quote 0.5% to take into account that N_1 and N_2 are also fluctuating (though are correlated with N_{s1} and N_{s2}).

4.9.10.2 Uncertainty due to daughter branching fractions In this section, we utilize the error propagation to investigate the effect of the uncertainties in daughter branching fractions on the ratio of BNV and SM branching fractions. Since we combine the two decay channels with $\bar{\Lambda}_c^-$ tags, the efficiency differences between different channels do not allow to cancel the systematics due to the uncertainties in daughter branching fractions completely. However, we demonstrate that the systematics due to this source of uncertainty is negligible.

Take ch1 and ch2 as example:

$$N_{SM} = \mathcal{B}(B^- \to \Xi_c^0 \bar{\Lambda}_c^-) \times (\epsilon_{ch1SM}\mathcal{B}_1 + \epsilon_{ch2SM}\mathcal{B}_2) \times N_{B^+/B^-}$$

$$N_{BNV} = \mathcal{B}(B^- \to \bar{\Xi}_c^0 \bar{\Lambda}_c^-) \times (\epsilon_{ch1BNV}\mathcal{B}_1 + \epsilon_{ch2BNV}\mathcal{B}_2) \times N_{B^+/B^-},$$

where

$$\mathcal{B}_1 = \mathcal{B}(\bar{\Lambda}_c^- \to \bar{p}K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)\mathcal{B}(\bar{\Xi}_c^0 \to \bar{\Xi}^+\pi^-)\mathcal{B}(\bar{\Xi}^+ \to \bar{\Lambda}\pi^+)\mathcal{B}(\bar{\Lambda} \to \bar{p}\pi^+)$$
$$\mathcal{B}_2 = \mathcal{B}(\bar{\Lambda}_c^- \to \bar{p}K^+\pi^-)\mathcal{B}(\bar{\Xi}_c^0 \to \bar{\Xi}^+\pi^-)\mathcal{B}(\bar{\Xi}^+ \to \bar{\Lambda}\pi^+)\mathcal{B}(\bar{\Lambda} \to \bar{p}\pi^+).$$

Then the ratio between BNV and SM branching fractions could be calculated as

$$R = \frac{\mathcal{B}(B^- \to \bar{\Xi}_c^0\bar{\Lambda}_c^-)}{\mathcal{B}(B^- \to \Xi_c^0\bar{\Lambda}_c^-)} = \frac{N_{BNV}}{N_{SM}}\frac{\epsilon_{ch1SM}\mathcal{B}_1 + \epsilon_{ch2SM}\mathcal{B}_2}{\epsilon_{ch1BNV}\mathcal{B}_1 + \epsilon_{ch2BNV}\mathcal{B}_2} = \frac{N_{BNV}}{N_{SM}}\frac{\epsilon_{ch1SM} + \epsilon_{ch2SM}\frac{\mathcal{B}_2}{\mathcal{B}_1}}{\epsilon_{ch1BNV} + \epsilon_{ch2BNV}\frac{\mathcal{B}_2}{\mathcal{B}_1}}.$$

For $r = \frac{\mathcal{B}_2}{\mathcal{B}_1}$, Ξ_c^0-related daughter branching fractions cancel and only $\bar{\Lambda}_c^-$ tag-related daughter branching fractions remain. As shown in Table. 24, the relative uncertainties in Ξ_c^0 branching fractions are around 20%, and the relative uncertainties in $\bar{\Lambda}_c^-$ branching fractions are around 5%. The uncertainty in $\frac{\mathcal{B}_2}{\mathcal{B}_1}$ is then calculated as

$$r = \frac{\mathcal{B}_2}{\mathcal{B}_1} = \frac{\mathcal{B}(\bar{\Lambda}_c^- \to \bar{p}K_S^0)\mathcal{B}(K_S^0 \to \pi^+\pi^-)}{\mathcal{B}(\bar{\Lambda}_c^- \to \bar{p}K^+\pi^-)} = \frac{(1.59 \pm 0.08)\% * (69.2 \pm 0.05)\%}{(6.28 \pm 0.32)\%} = 5.71 \pm 0.41.$$

The relative uncertainty in $\frac{\mathcal{B}_2}{\mathcal{B}_1}$ is around 7%. Note that r is only related to the $\bar{\Lambda}_c^-$ decays and is the same for ch3 and ch4, ch5 and ch6.

For the uncertainty in the ratio between BNV and SM, substitute into the formula the efficiencies for SM and BNV modes ($\epsilon_{ch1SM} = 9.19\%$, $\epsilon_{ch1BNV} = 9.22\%$, $\epsilon_{ch2SM} = 7.70\%$, $\epsilon_{ch2BNV} = 7.56\%$), then

$$R = \frac{\mathcal{B}(B^- \to \bar{\Xi}_c^0\bar{\Lambda}_c^-)}{\mathcal{B}(B^- \to \Xi_c^0\bar{\Lambda}_c^-)} = \frac{N_{BNV}}{N_{SM}}\frac{\epsilon_{ch1SM} + \epsilon_{ch2SM} * r}{\epsilon_{ch1BNV} + \epsilon_{ch2BNV} * r} = \frac{N_{BNV}}{N_{SM}}\frac{9.19\% + 7.70\% * r}{9.22\% + 7.56\% * r}$$

$$\frac{\Delta R}{R} = \frac{\partial R}{\partial r}\frac{\Delta r}{R} = \frac{(\epsilon_{ch2SM} * \epsilon_{ch1BNV} - \epsilon_{ch1SM} * \epsilon_{ch2BNV}) * \Delta r}{(\epsilon_{ch1SM} + \epsilon_{ch2SM} * r)(\epsilon_{ch1BNV} + \epsilon_{ch2BNV} * r)}$$
$$= \frac{(7.70\% * 9.22\% - 7.56\% * 9.19\%) * 0.41}{(9.19\% + 7.70\% * 5.71)(9.22\% + 7.56\% * 5.71)} = 0.022\%$$

The relative uncertainty is related to the second order of the small difference between SM and BNV mode efficiencies and is thus negligible.

In the calculation shown on the previous page, we assumed that the efficiencies are known with 3 significant digits. Assuming that the difference between the numbers 9.19% and

9.22% is actual (i.e., not statistical), we estimated the systematics on the ratio of branching fractions due to the uncertainties in daughter branching fractions of just 0.022%. Taking ch1 SM mode as an example, the number of fully reconstructed events is around 20000 and the statistical uncertainty on the (absolute) efficiency is about 0.05%. Therefore, the SM and BNV efficiencies could actually be even more similar to each other, which would further reduce the systematic uncertainty on the ratio of branching fractions due to the uncertainties in daughter branching fractions. Assuming that the central values of estimated efficiencies actually differ from each other by a larger amount, e.g., the efficiencies are 9.14% and 9.27%, instead of 9.19% and 9.22%, respectively, the estimate of 0.022% becomes 0.032%, which is still a very small marginally larger number. Assuming that the actual efficiencies are 9.24% and 9.17% (again, instead of central values of 9.19% and 9.22%) results in the systematic uncertainty of just 0.010%.

4.10 INTERPRETATION OF THE RESULTS IN TERMS OF THE OSCILLATION FREQUENCY

We already measured the branching fractions for both BNV and SM decays: $\mathcal{B}(B^- \to \overline{\Xi}_c^0\overline{\Lambda}_c^-)$ and $\mathcal{B}(B^- \to \Xi_c^0\overline{\Lambda}_c^-)$. Now we would like to convert the ratio of BNV and SM branching fractions to the oscillation frequency of Ξ_c^0.

4.10.1 Formalism of Ξ_c^0 and Anti-Ξ_c^0 Mixing

The time evolution of the $\Xi_c^0 - \overline{\Xi}_c^0$ mixing is described by the Schrödinger equation as

$$i\frac{\partial}{\partial t}\begin{pmatrix} \Xi_c^0(t) \\ \overline{\Xi}_c^0(t) \end{pmatrix} = (\mathbf{M} - \frac{i}{2}\mathbf{\Gamma})\begin{pmatrix} \Xi_c^0(t) \\ \overline{\Xi}_c^0(t) \end{pmatrix}, \tag{4.6}$$

where the $\mathbf{M}$ and $\mathbf{\Gamma}$ matrices are Hermitian and CPT invariance requires $M_{11} = M_{22} = M$ and $\Gamma_{11} = \Gamma_{22} = \Gamma$.

The Ξ_c^0 and $\overline{\Xi}_c^0$ are produced as flavor eigenstates and evolve as the eigenstates Ξ_{cH}^0 and Ξ_{cL}^0 of the Hamiltonian $\mathbf{M} - \frac{i}{2}\mathbf{\Gamma}$ with masses and widths M_H,Γ_H and M_L,Γ_L. The mixing

process is characterized by the mass and width differences, so mixing parameters x and y are defined as

$$x = \frac{\Delta M}{\Gamma} = \frac{M_H - M_L}{\Gamma}, y = \frac{\Delta\Gamma}{\Gamma} = \frac{\Gamma_H - \Gamma_L}{2\Gamma},$$

where $\Gamma = (\Gamma_H + \Gamma_L)/2$.

Let us consider the evolution of Ξ_c^0 produced at time $t = 0$ to BNV final state. This process can proceed through a direct BNV decay, or via $\Xi_c^0 - \overline{\Xi}_c^0$ oscillations followed by the SM decay. Assuming small mixing $|x|, |y| \ll 1$ and no CP violation, the time-dependent ratio of the BNV decay rate to the SM decay rate is described as [47]

$$r(t) = \frac{P(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{P(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = (R_D + \sqrt{R_D} y' \Gamma t + \frac{x'^2 + y'^2}{4} \Gamma^2 t^2) e^{-\Gamma t},$$

where R_D is the ratio of the direct BNV decay rate to the SM decay rate, $x' = x \cos\delta + y \sin\delta$, $y' = -x \sin\delta + y \cos\delta$, and δ is the strong phase difference between direct BNV decay and SM decay. The first term in this equation is due to direct BNV decay, the last term is due to mixing (oscillations), and the middle term is due to the interference between them.

The time-integrated ratio of the BNV decay rate to the SM decay rate is described as

$$R = R_D + \sqrt{R_D} y' + \frac{x'^2 + y'^2}{2}.$$

In this last formula, the last term does not depend on the strong phase difference anymore.

4.10.2 Formalism of Ξ_c^0 and Anti-Ξ_c^0 Oscillations

In this analysis, when we interpret our results, we assume the $\Xi_c^0 - \overline{\Xi}_c^0$ oscillation hypothesis only and no direct BNV decay, which leads to $R_D = 0$. Then the time-dependent ratio $r(t)$ becomes

$$r(t) = \frac{x^2 + y^2}{4} \Gamma^2 t^2 e^{-\Gamma t} = \left[\left(\frac{\Delta M}{2} \right)^2 + \left(\frac{\Delta\Gamma}{4} \right)^2 \right] t^2 e^{-\Gamma t} = \omega^2 t^2 e^{-\Gamma t},$$

where ω is the angular frequency of oscillations.

The time-integrated ratio R becomes

$$R = \frac{x^2 + y^2}{2} = \frac{2\omega^2}{\Gamma^2} = 2\omega^2\tau^2,$$

where τ is the lifetime of Ξ_c^0.

In this analysis, we measure the time-integrated ratio R of the BNV decay rate to the SM decay rate, i.e., the ratio of BNV and SM branching fractions, and report the angular frequency of oscillations.

4.10.3 Effect of the Magnetic Field

In the presence of the magnetic field, the non-zero magnetic moment of baryons results in an energy splitting $\Delta E = 2\mu B$ of the baryon and anti-baryon states. This energy splitting would suppress the oscillation rate. The criterion for neglecting this effect is $|\Delta E|t/2 \ll 1$, where t is the time of propagation of the baryon [37].

We assume the magnetic moment of Ξ_c^0 to be comparable to the nuclear magneton $\mu_N = 3.15 \times 10^{-14}$ MeV/Tesla. The magnetic field of Belle solenoid is 1.5 T and the Lorentz boost factor is negligible. The lifetime τ of Ξ_c^0 is 1.519×10^{-13} s. $|\Delta E|t/2 = \mu B\tau < 10^{-5}$ and therefore the effect of the magnetic field can be safely ignored.

4.11 10% DATA UNBLINDING

To validate our analysis approach, upon securing the permission from our paper committee, we unblind 10% of data, which corresponds, approximately, to 77.1×10^6 $B\overline{B}$ pairs, and estimate the upper limit on the ratio between BNV and SM branching fractions. To do so we use 10% of randomly selected events from both the sideband and the previously-blinded region.

4.11.1 Branching Fractions for SM and BNV Modes

To extract $\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)$ and $\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)$, we perform simultaneous fits for six SM/BNV channels. We use PDG averages for the Ξ_c^0 branching fractions.

Fig. 57 shows the results of a simultaneous fit for SM mode:

$$\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-) = (1.32 \pm 0.40) \times 10^{-3} \tag{4.7}$$

This result should be compared with our full data sample result shown in Eq. 4.2. Taken at face values, these two sets of numbers could be used to evaluate the improvement in the uncertainty associated with the central value: $(0.40/1.32)/(0.11/1.03) = \sqrt{8.1}$, which is close to $\sqrt{10}$, expected on average.

Fig. 58 shows the results of a simultaneous fit for BNV mode:

$$\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) = (-0.76 \pm 1.10) \times 10^{-3}. \tag{4.8}$$

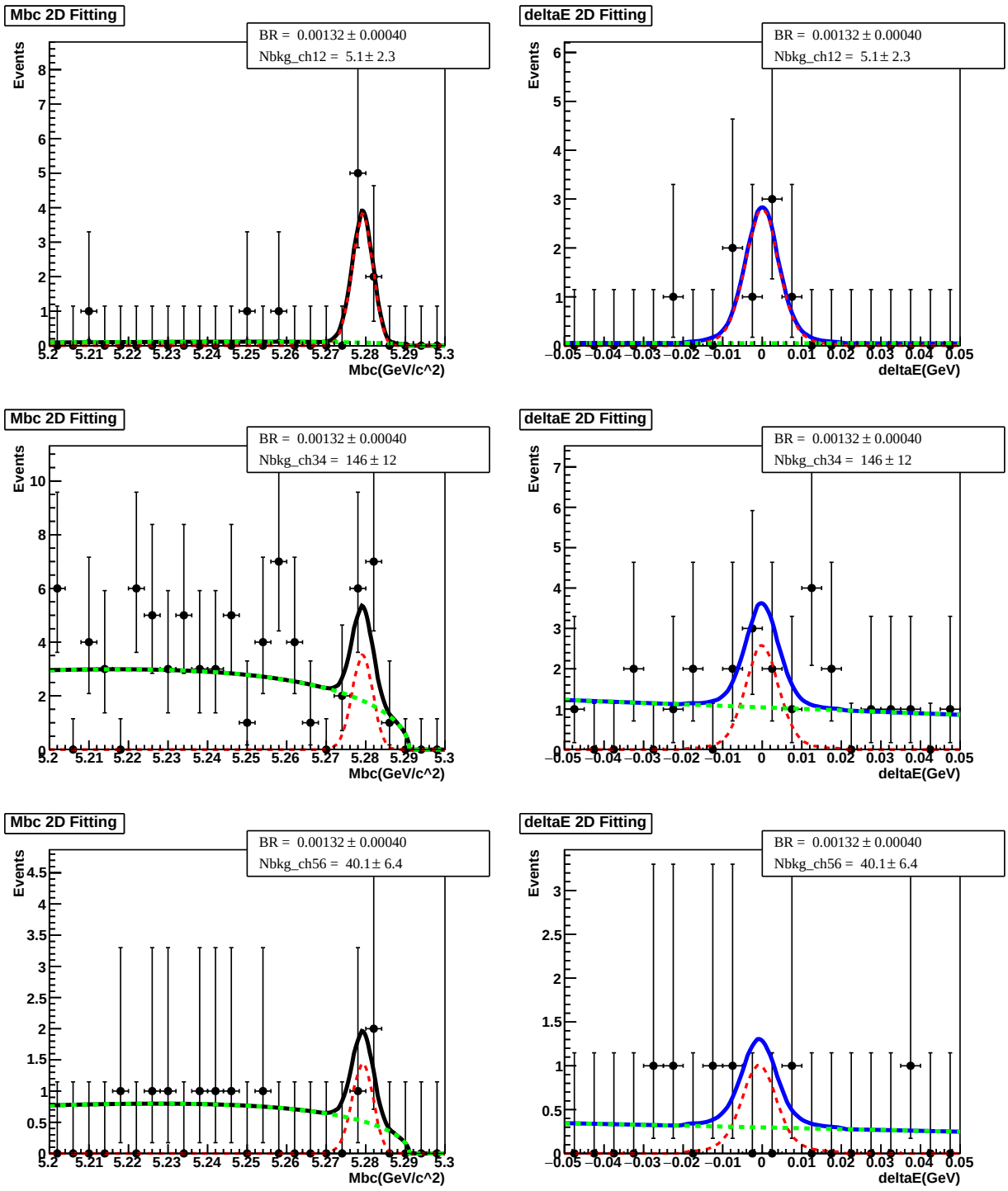


Figure 57: Simultaneous fit to SM mode using 10% of unblinded data.

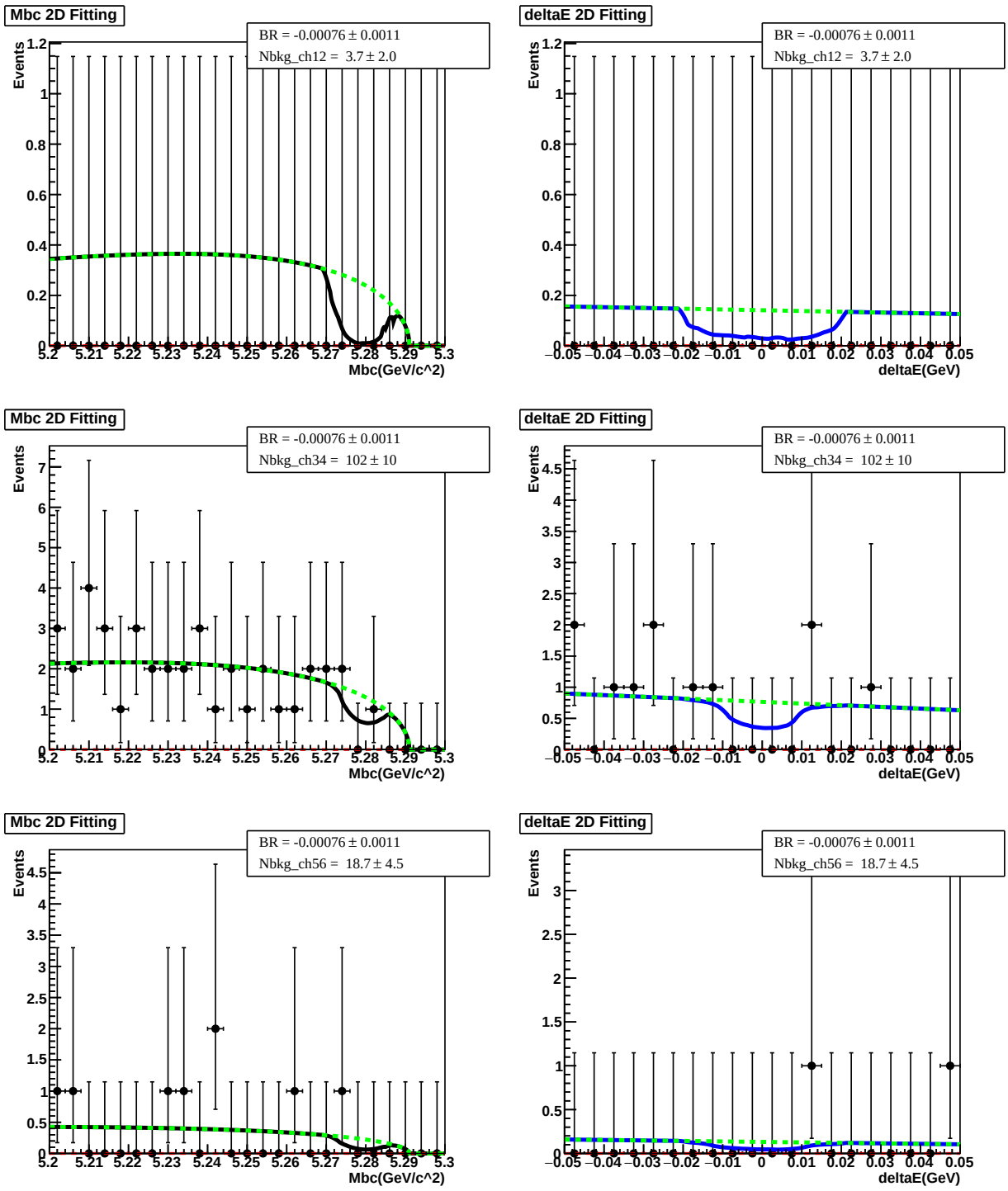


Figure 58: Simultaneous fit to BNV mode using 10% of unblinded data.

4.11.2 Confidence Belt for BNV Mode Prepared for 10% of Unblinded Data

Fig. 59 shows the 90% confidence level belt prepared for 10% unblinded data. 3/106/14 background events are generated for ch1&2, ch3&4 and ch5&6, respectively. The belt shown in red color in Figures 59 shows the confidence belts obtained from 2D fitting for BNV modes and the belt in blue color shows the confidence belts obtained from 1D fitting.

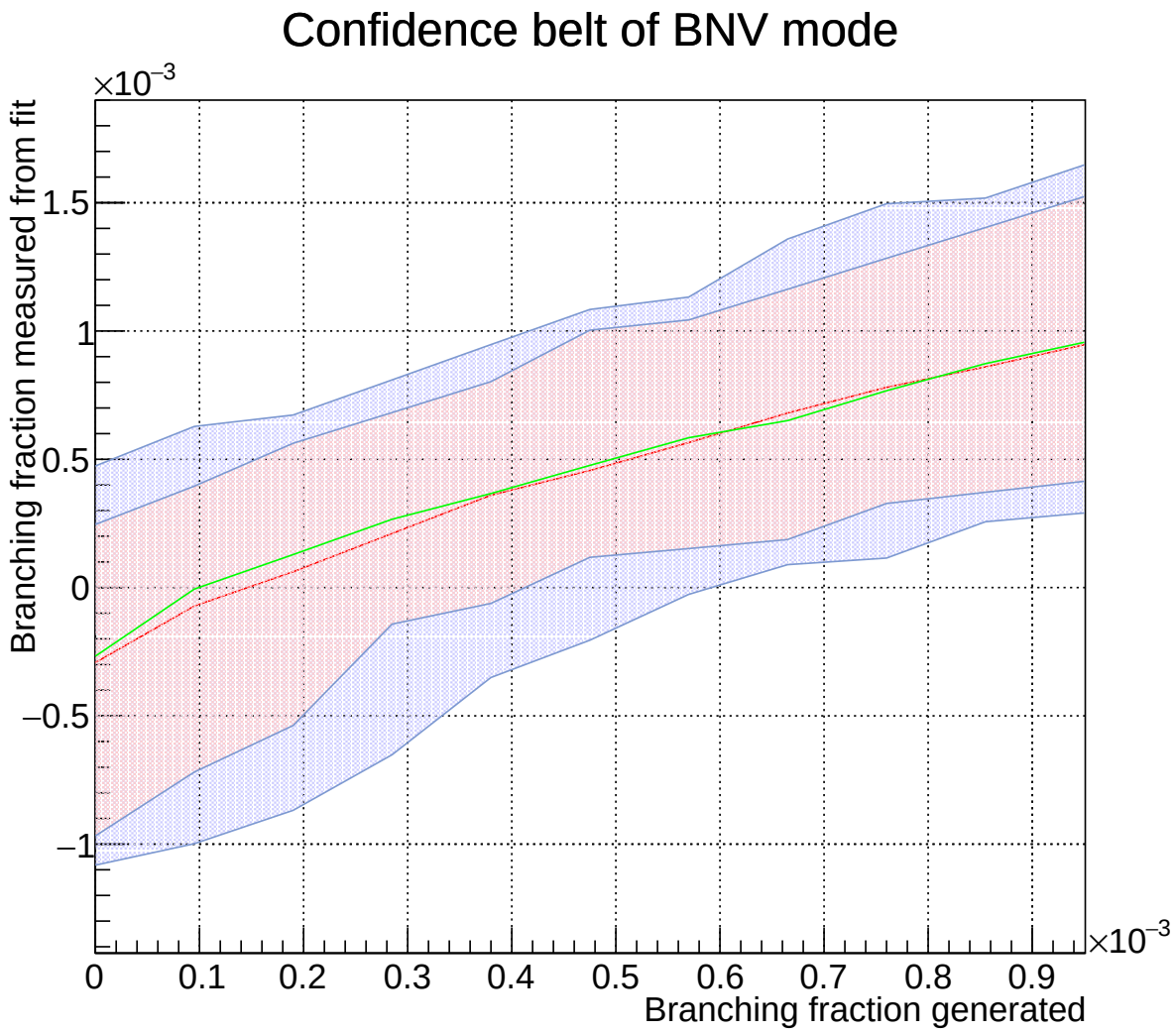


Figure 59: Confidence belt for BNV mode prepared for 10% of unblinded data.

Assuming that the measured central value of the BNV branching ratio in data is 0, the 95% confidence level upper limit obtained from this confidence belt would be 0.4 and 0.6 (in units of 10^{-3}) for 2D and 1D fits, respectively. The full-data prediction, based on Fig. 56, predicts 0.06 and 0.09, for the same respective scenarios. Therefore, for zero result in data,

the improvement in the upper limit is represented by the factor of, approximately, 7. This is consistent with the low-statistics BNV mode, where, if no signal and background events were present, we would expect the improvement by a factor of 10. On the other hand, if BNV mode analysis were dominated by background, the improvement of, approximately, $\sqrt{10}$ would be expected on average.

4.11.3 Upper Limit on the Ratio Between BNV and SM Branching Fractions

First, we estimate the upper limit on the ratio of branching fractions using the BNV fit result for 10% of data. Using the 2D confidence belt shown in Fig. 59, and the central value for the branching fraction obtained from the 2D fit to 10%-unblinded data (-0.76×10^{-3} from Eq. 4.8), the upper limit of the BNV branching fraction is estimated to be 7.9×10^{-5}. We then calculate the 95% CL upper limit on the time-integrated ratio between BNV and SM branching fraction as

$$R = \frac{P(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{P(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = \frac{7.9 \times 10^{-5}}{1.32 \times 10^{-3}} = 6.0\%, \tag{4.9}$$

where only central value is used for the SM result and no systematic uncertainties are taken into account.

Converted to the 95% CL upper limit on the oscillation angular frequency the result becomes

$$\omega = \frac{\sqrt{R/2}}{\tau} = \frac{0.173}{1.519 * 10^{-13}} = 1.14\ ps^{-1}.$$

Now we estimate the upper limit on the ratio of branching fractions assuming that the fit yields 0 result for 10% of data. Using the central value for the SM result (shown in Eq. 4.7) and confidence belt prepared for 10% of data (shown in Fig. 59), we expect

$$R = \frac{P(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{P(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = \frac{40.0 \times 10^{-5}}{1.32 \times 10^{-3}} = 30\%, \tag{4.10}$$

and

$$R = \frac{P(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{P(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = \frac{60.0 \times 10^{-5}}{1.32 \times 10^{-3}} = 45\%, \tag{4.11}$$

for 2D and 1D analyses, respectively.

Finally, we estimate the upper limit on the ratio of branching fractions assuming that the fit yields 0 result for the entire Belle data sample. Using the central value for the SM result (shown in Eq. 4.2) and confidence belt prepared for full data statistics (shown in Fig. 56), we expect

$$R = \frac{P(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{P(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = \frac{6.0 \times 10^{-5}}{1.03 \times 10^{-3}} = 6\%, \tag{4.12}$$

and

$$R = \frac{P(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{P(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = \frac{9.0 \times 10^{-5}}{1.03 \times 10^{-3}} = 9\%, \tag{4.13}$$

for 2D and 1D analyses, respectively, where the improvement with 10 times larger statistics is the factor of 5 according to Eq. 4.10 and 4.11. The improvement (the factor of 5) is less than expected (factor of 7 explained previously), because the result for the central value for SM branching fraction obtained using 10% of the data is 30% larger than the one obtained with full data.

4.12 FULL DATA UNBLINDING

We unblind the full data sample, which corresponds to 771×10^6 $B\overline{B}$ pairs, and set the upper limit on the ratio between BNV and SM branching fractions.

4.12.1 Branching Fractions for SM and BNV Modes

Simultaneous fits for six SM/BNV channels are performed to extract BNV and SM branching fractions. Fig. 60 shows the results for SM mode:

$$\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-) = (1.10 \pm 0.12) \times 10^{-3}, \tag{4.14}$$

Note the SM branching fraction measured here is different from the one measured in Section 8.1, which also uses full data sample. This is because, to be consistent in the comparison with the previous Belle measurement, we use the Ξ_c^0 branching fractions from Belle and fit range $M_{\rm bc} > 5.25$ GeV/c^2, $|\Delta E| < 0.04$ GeV in Section 8.1.

In this section, we use the Ξ_c^0 branching fractions from PDG averages and fit range $M_{\rm bc} > 5.2$ GeV/c^2, $|\Delta E| < 0.05$ GeV in both SM and BNV measurements.

Fig. 61 shows the results for BNV mode:

$$\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) = (-9.65 \pm 2.90) \times 10^{-5}. \tag{4.15}$$

We also performed a simultaneous fit to SM and BNV modes with SM branching fraction and the ratio of BNV and SM branching fractions as parameters, the results (central values and uncertainties) shown below are 100% consistent with the results obtained from the original fit to two branching fractions. Therefore (as is shown in the next subsection) we use our result for the BNV branching fraction, i.e., $\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)$, to set the upper limit on the ratio of BNV and SM branching fractions using the following results:

$$\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-) = (1.10 \pm 0.12) \times 10^{-3}, \tag{4.16}$$

$$\frac{\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = -0.0878 \pm 0.028 \tag{4.17}$$

Note that using the result of the fit for the ratio of branching fractions and the SM branching fraction yields the same result (to the third significant digit) because the two fitting parameters in either of the two fits are not correlated.

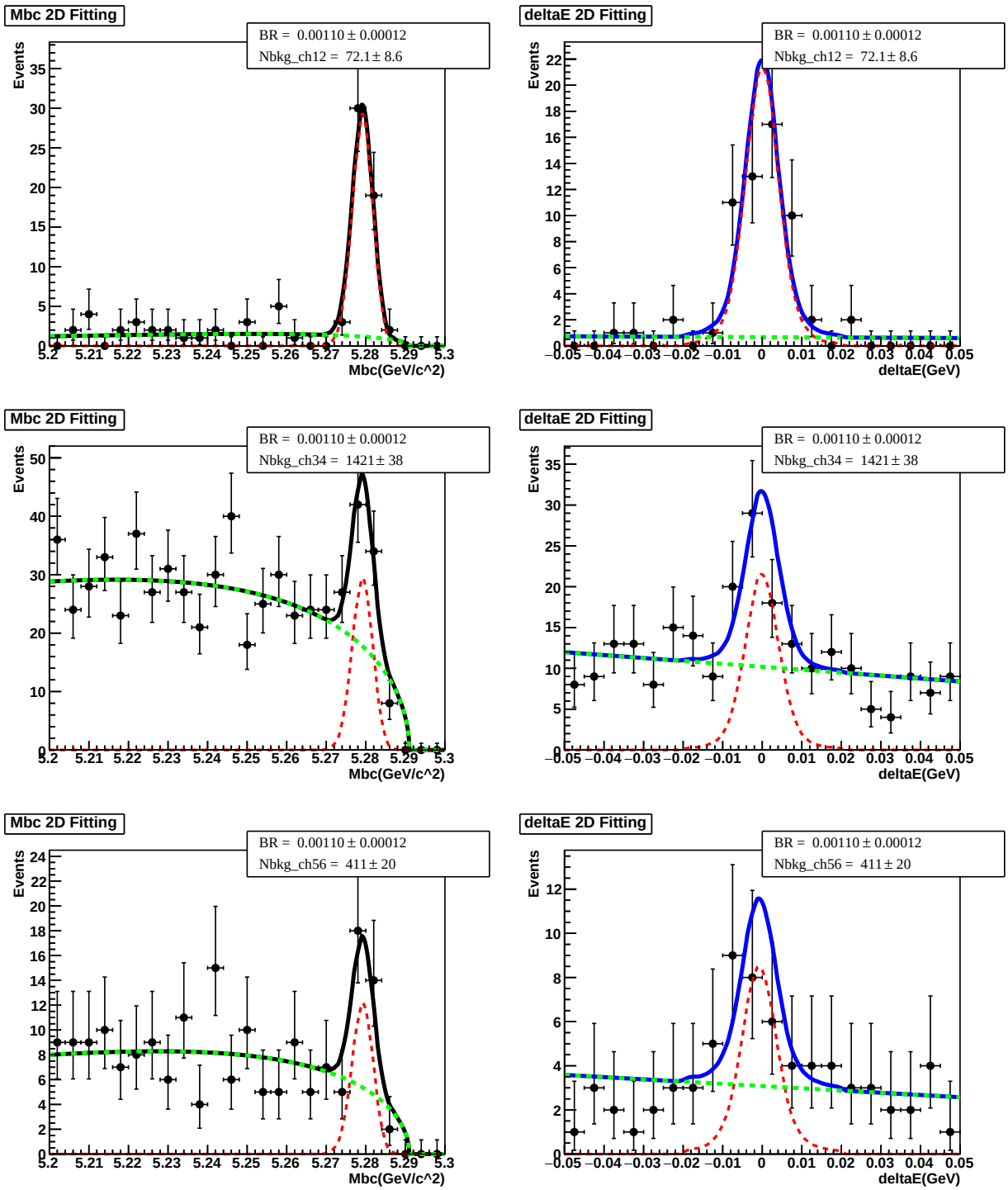


Figure 60: The results of a simultaneous fit for the SM channels using fully unblinded data.

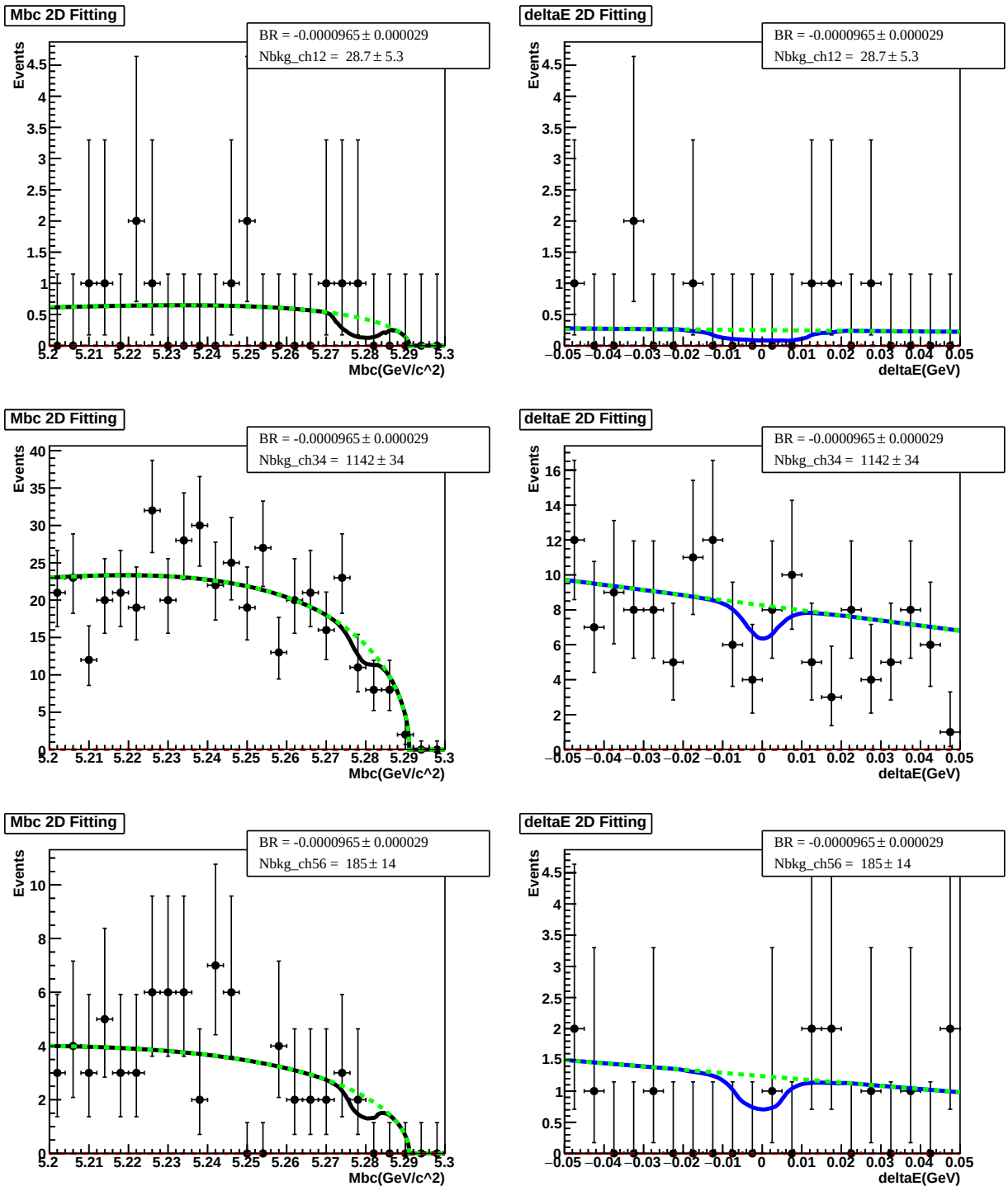


Figure 61: The results of a simultaneous fit for the BNV channels using fully unblinded data.

4.12.2 Upper Limit on the Ratio Between BNV and SM Branching Fractions Based on Central Value

Using the 2D confidence belt for full data sample shown in Fig. 56, and the central value for the branching fraction obtained from the 2D fit to fully unblinded data (-9.65×10^{-5} from Eq. 4.15), the upper limit of the BNV branching fraction is estimated to be 1.5×10^{-5} @ 95% CL. Note that (by virtue of the confidence belt approach) this result already includes the statistical uncertainties of BNV branching fractions but not SM branching fractions (discussed in Section 13.1). We then calculate the 95% CL upper limit on the time-integrated ratio between BNV and SM branching fraction as

$$R = \frac{P(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{P(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = \frac{1.5 \times 10^{-5}}{1.10 \times 10^{-3}} = 1.36\%, \tag{4.18}$$

at 95% CL, where the statistical uncertainties are already taken into account via the confidence belt technique. The remaining systematic uncertainties (besides those which cancel via the ratio) will be discussed in Section 13.

Converted to the 95% CL upper limit on the oscillation angular frequency the result becomes

$$\omega = \frac{\sqrt{R/2}}{\tau} = \frac{0.083}{1.519 * 10^{-13}} = 0.54\ ps^{-1}.$$

The sensitivity for the BNV branching fraction (the upper limit estimated assuming measured branching fraction of 0) is 6.0×10^{-5} using Fig. 56. We then estimate the sensitivity for the ratio between BNV and SM branching fractions to be

$$R = \frac{P(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{P(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = \frac{6.0 \times 10^{-5}}{1.10 \times 10^{-3}} = 5.45\%, \tag{4.19}$$

at 95% CL. Converted to sensitivity for the upper limit on the oscillation angular frequency the result becomes

$$\omega = \frac{\sqrt{R/2}}{\tau} = \frac{0.165}{1.519 * 10^{-13}} = 1.09\ ps^{-1}$$

at 95% CL.

4.13 UPPER LIMIT ON THE RATIO BETWEEN BNV AND SM BRANCHING FRACTIONS

In this section, we describe how we include statistical uncertainties in SM branching fractions and systematic uncertainties into the construction of confidence belt and thus estimate the upper limit for the ratio between BNV and SM branching fractions.

4.13.1 Including Statistical Uncertainty of SM Branching Fraction

In the estimate of the upper limit on the ratio between BNV and SM branching fractions (1.36% (1.4%)) in Section 12.2, we use the central value of measured SM branching fraction but do NOT yet include statistical uncertainty on its value.

To include the statistical uncertainty in the SM branching fraction into our estimate of the upper limit, we prepare a new 90% confidence belt (using the two signal variables) for the ratio between BNV and SM branching fractions (see Fig. 62).

We use 20 values of the BNV/SM ratio of branching fractions in the range between 0% and 10%, and, for each value of the BNV/SM ratio, we generate 5000 toy MC experiments. In each toy MC experiment for a specified BNV/SM ratio, we first randomly generate (i.e., sample) the SM branching fraction from a Gaussian distribution based on the measured SM branching fraction and its statistical uncertainty (see Eq. 4.14), and calculate the corresponding BNV branching fractions. We then translate the SM and BNV branching fractions to the number of signal events to be generated $N_{sig}^{SM/BNV}$ for SM and BNV modes, respectively.

$N^{SM/BNV} = N_{sig}^{SM/BNV} + N_{bkg}^{SM/BNV}$ events are generated according to the fit models described in Section 4.4, where $N_{bkg}^{SM/BNV}$ is the number of background events for SM and BNV modes, which are estimated using generic MC and scaled using sideband data (described in Section 4.5). Unbinned 2D extended maximum likelihood simultaneous fits for six SM/BNV channels are performed for each generated sample to extract the BNV and SM branching fractions and then the BNV/SM ratio is calculated.

When statistical uncertainty in the SM branching fraction is included as described here, our 95% CL upper limit on BNV/SM ratio (see Eq. 4.17) becomes 1.39% (1.4%).

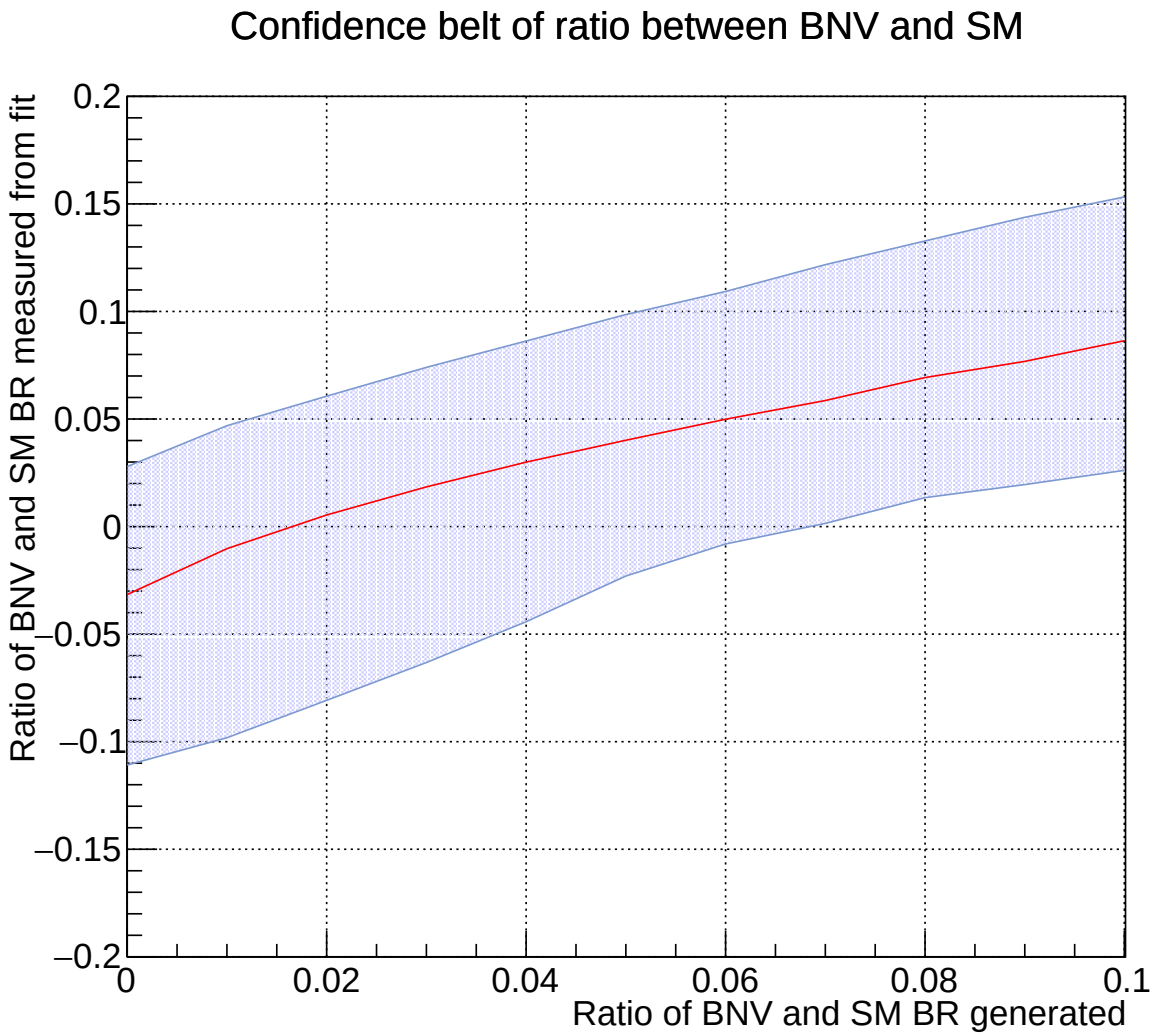


Figure 62: Confidence belt for the ratio between BNV and SM branching fractions constructed including statistical uncertainty in the SM result.

4.13.2 Including Systematic Uncertainty Due to PDF Shapes

We incorporate the systematic uncertainty due to fixed PDF shapes into our confidence belt using a similar method.

We again use 20 values of the BNV/SM ratio of branching fractions in the range between 0% and 10% and generate 5000 toy MC experiments for each value of the BNV/SM ratio. In each toy MC experiment, we generate our toy MC events based on the given SM and BNV branching fractions using our original PDFs (i.e., we do not vary PDF shapes in this

step because varying PDFs during event generation and also during fitting would result in including the same effect twice).

Then, before we perform a fit to toy MC samples, we randomly vary the SM and BNV PDF shapes. In our opinion, this is practically equivalent to varying the PDF shapes in the previous step. Each signal PDF is scaled with a scale factor randomly sampled from a Gaussian distribution with $\mu = 1$ and $\sigma = 0.1$, in order to increase or decrease the width of signal PDFs by 10%, on average. Our decision to use this number is based on various previous Belle studies that indicate that MC and data usually agree with each other better than that.

Background PDF parameters are varied by a few percent value, which is randomly selected from a Gaussian distribution, whose σ equals 20% of the relative uncertainty in the background PDF shapes. We decided to use the 20% number on the basis of our analysis for ch3&4 (channels with the largest statistics) when fitting full Belle data. We release the restriction on the background PDF parameters of ch3&4 in the simultaneous fit and extract the (relative) uncertainty, which equals 20%. We do not vary the background PDF parameters by $\pm 1\sigma$ as for signal PDF because varying the shape of ARGUS too much would immediately take us into an unphysical region.

To measure the ratio between BNV and SM branching fractions we fit our model with such randomly varied PDFs to toy MC events generated using the original PDFs.

Fig. 63 shows the 90% confidence belt (prepared using both signal variables) for the ratio of the BNV and SM branching fractions constructed including the PDF systematic uncertainties. Using the BNV/SM ratio measured from the data (Eq. 4.17), the 95% CL upper limit becomes 1.50% (1.5%).

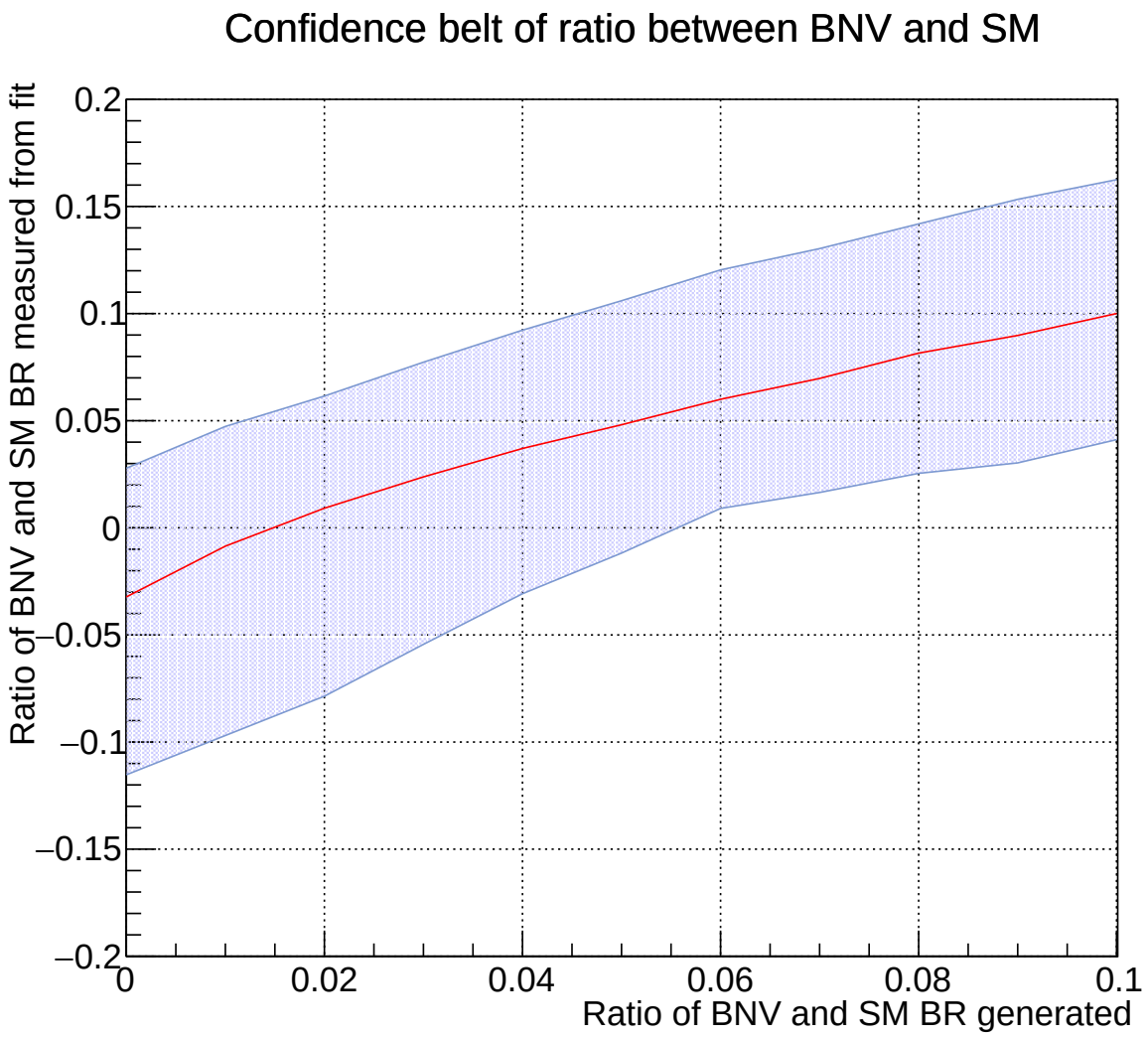


Figure 63: Confidence belt for the ratio between BNV and SM branching fractions constructed including the systematic uncertainties due to PDF shapes.

4.13.3 Including All Statistical and Systematic Uncertainty

In this section, we describe how we include both the statistical uncertainty in SM branching fractions and the systematic uncertainty due to fixed PDF shapes to construct our confidence belt used to obtain the final result.

Similarly to previous sections, we use 20 values of the BNV/SM ratio of the branching fractions in the range between 0% and 10% and generate 5000 toy MC experiments for each value of the BNV/SM ratio. In each toy MC experiment, first we randomly generate the

SM branching fraction based on its measured value (and the uncertainty) as described in the previous section. Then we generate our toy MC events based on the given SM and BNV branching fractions using our original PDFs. Finally, we randomly vary the SM and BNV PDF shapes (as described in the previous section) before we perform a fit to toy MC data. To measure the ratio between BNV and SM branching fractions we fit our model with such randomly varied PDFs to toy MC events generated using the original PDFs.

Fig. 64 shows the 90% confidence belt (prepared using both signal variables) for the ratio of the BNV and SM branching fraction constructed including both statistical uncertainty in SM branching fraction and the PDF systematic uncertainties. Using the BNV/SM ratio measured from the data (Eq. 4.17), the 95% CL upper limit becomes 1.54% (1.5%).

To conclude, when we include the statistical and systematic uncertainties, our estimate of the upper limit increases from 1.36% to 1.39% and, finally, to 1.54%. Such trend is explained by the relatively steep rise of the lower boundary of our confidence belt when no BNV signal is expected.

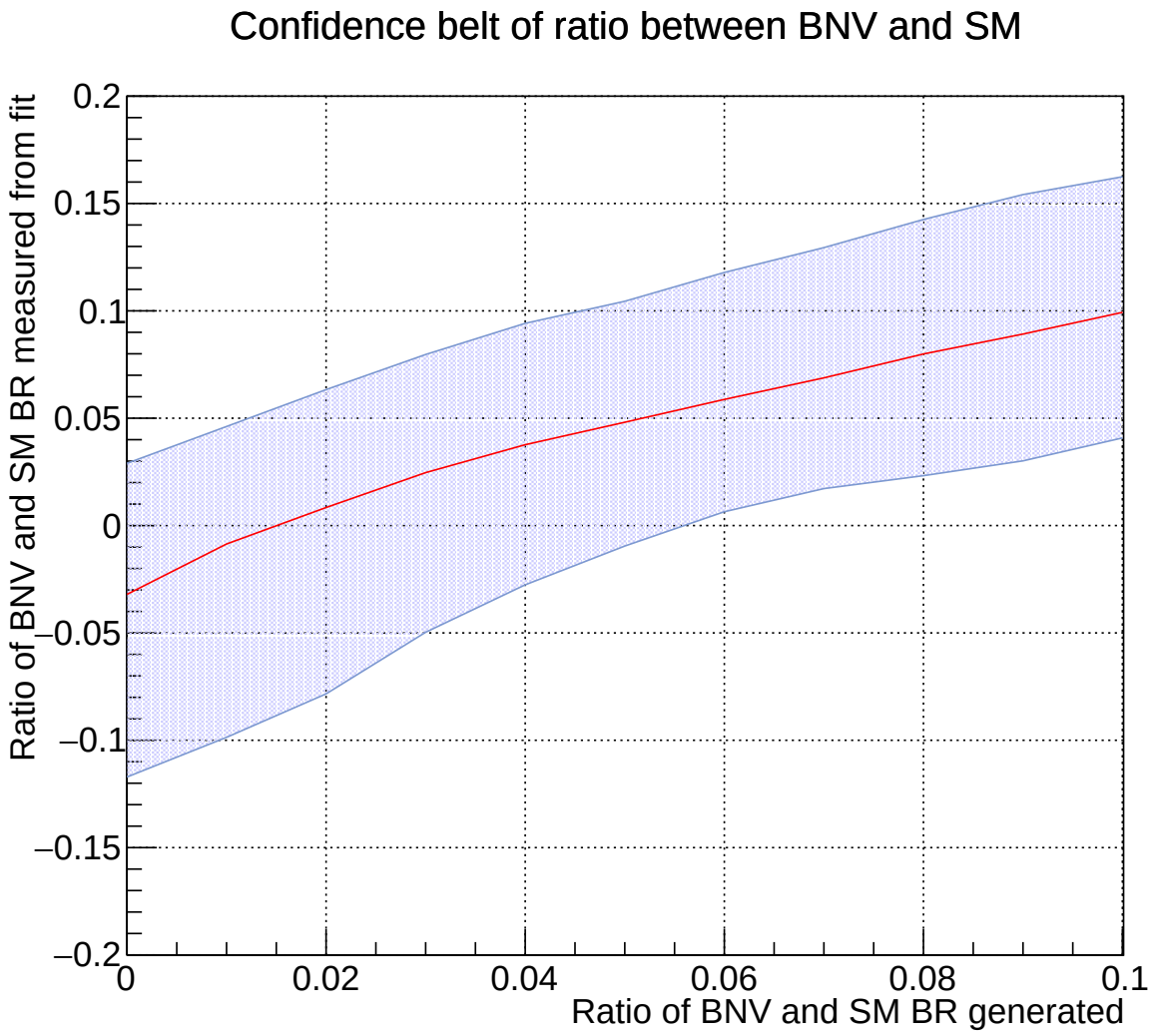


Figure 64: Confidence belt for the ratio between BNV and SM branching fractions constructed including the statistical uncertainty in the SM branching fraction and systematic uncertainties due to PDF shapes.

4.14 FITTING, CONFIDENCE BELT, AND UPPER LIMIT WITH RELEASED BACKGROUND PDFS

In this section, we present our results obtained with background shape PDF parameters extracted from signal data. This effort is motivated by insufficient statistics of generic MC for some of the channels and could be thought of as an alternative way to take background

PDF shape systematics into account. In one sentence, this approach results in a larger upper limit on the ratio of the BNV and SM branching fractions, increasing the final result (at 95% CL) from 1.5% to 2.7%. Further details of our investigations are described below.

4.14.1 Nominal Fit Results with Released Background PDFs

Simultaneous fits for six SM/BNV channels are performed with released background PDF shape parameters. The results of the fit for the SM channels are shown in Fig. 65. We obtain the SM branching fraction

$$\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-) = (1.13 \pm 0.12) \times 10^{-3}. \tag{4.20}$$

Fig. 66 shows the results for the BNV channels:

$$\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-) = (-7.78 \pm 2.70) \times 10^{-5}. \tag{4.21}$$

We also obtain the following central value for the ratio of BNV and SM branching fractions from the same fits:

$$\frac{\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} = -0.0688 \pm 0.025 \tag{4.22}$$

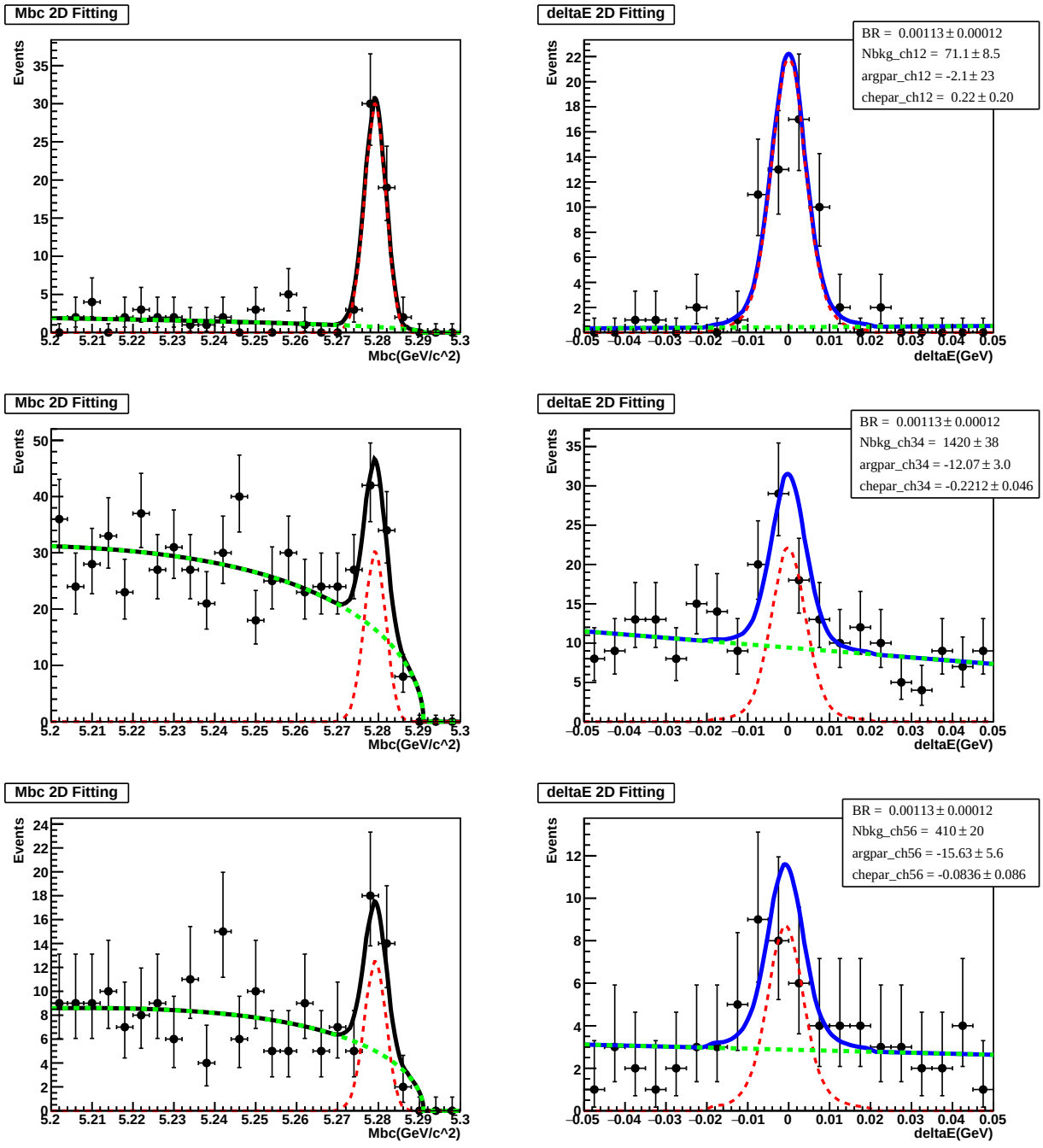


Figure 65: The results of a simultaneous fit for the SM channels using fully unblinded data with released background PDF shape parameters.

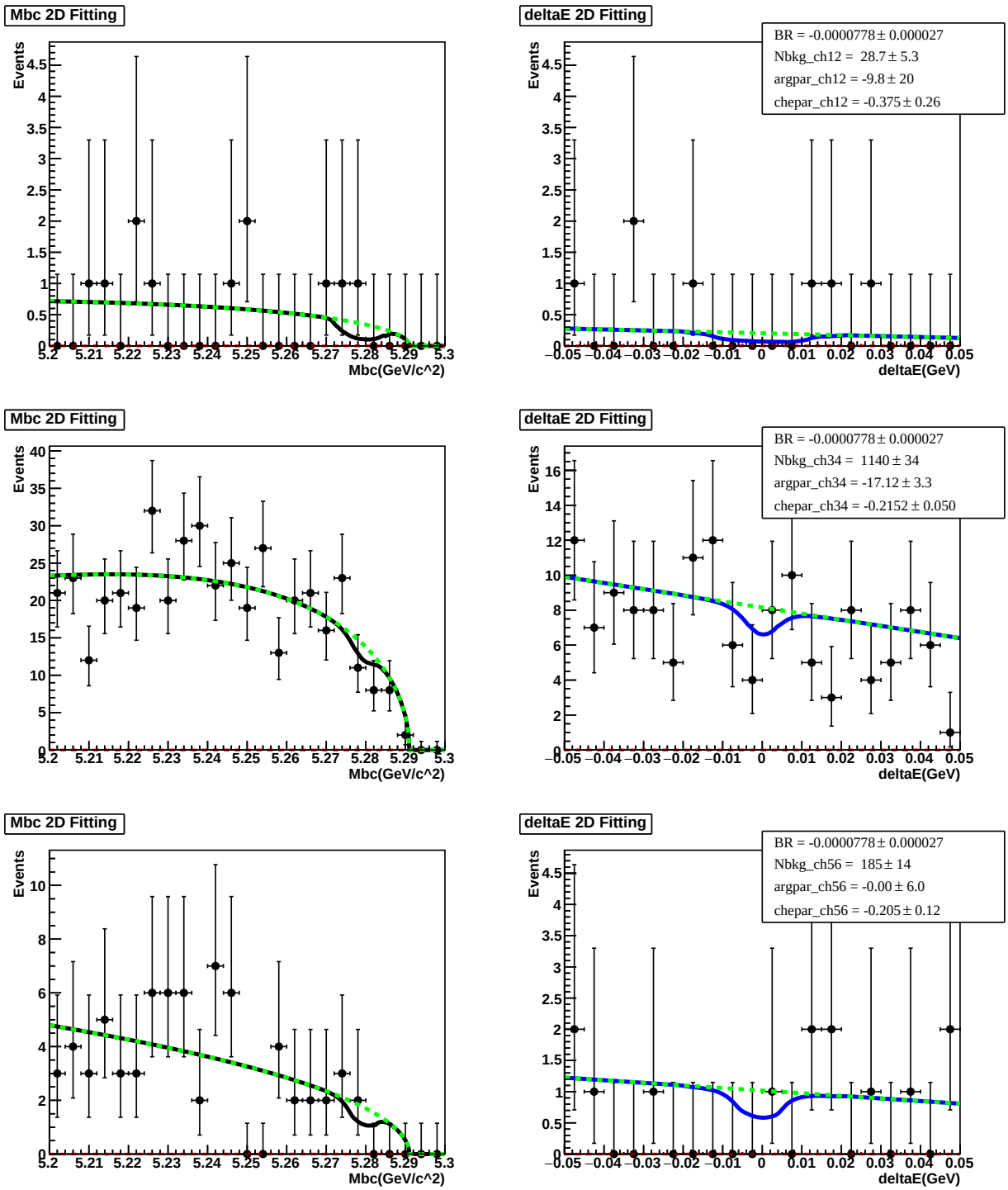


Figure 66: The results of a simultaneous fit for the BNV channels using fully unblinded data with released background PDF shape parameters.

4.14.2 Confidence Belts for Released Background PDFs

To obtain the result for released background PDF shape parameters, we prepare new confidence belts where toy MC experiments are generated using central values of background PDF shape parameters obtained from data and fitted with released background PDF shape parameters. Figures 67 through 70 show, respectively, the baseline confidence belt (without including any statistical and systematic uncertainties), the confidence belt including statistical uncertainties in SM branching fractions, the confidence belt including systematic uncertainties due to PDF shapes (to include systematic uncertainties due to background PDF shapes, the background $M_{\rm bc}$ is modeled with an ARGUS function with released threshold and background ΔE is modeled with a second-order Chebychev polynomial, this is being done in addition to releasing background PDF shape parameters in the nominal fit to data), and the confidence belt including both statistical and systematic uncertainties.

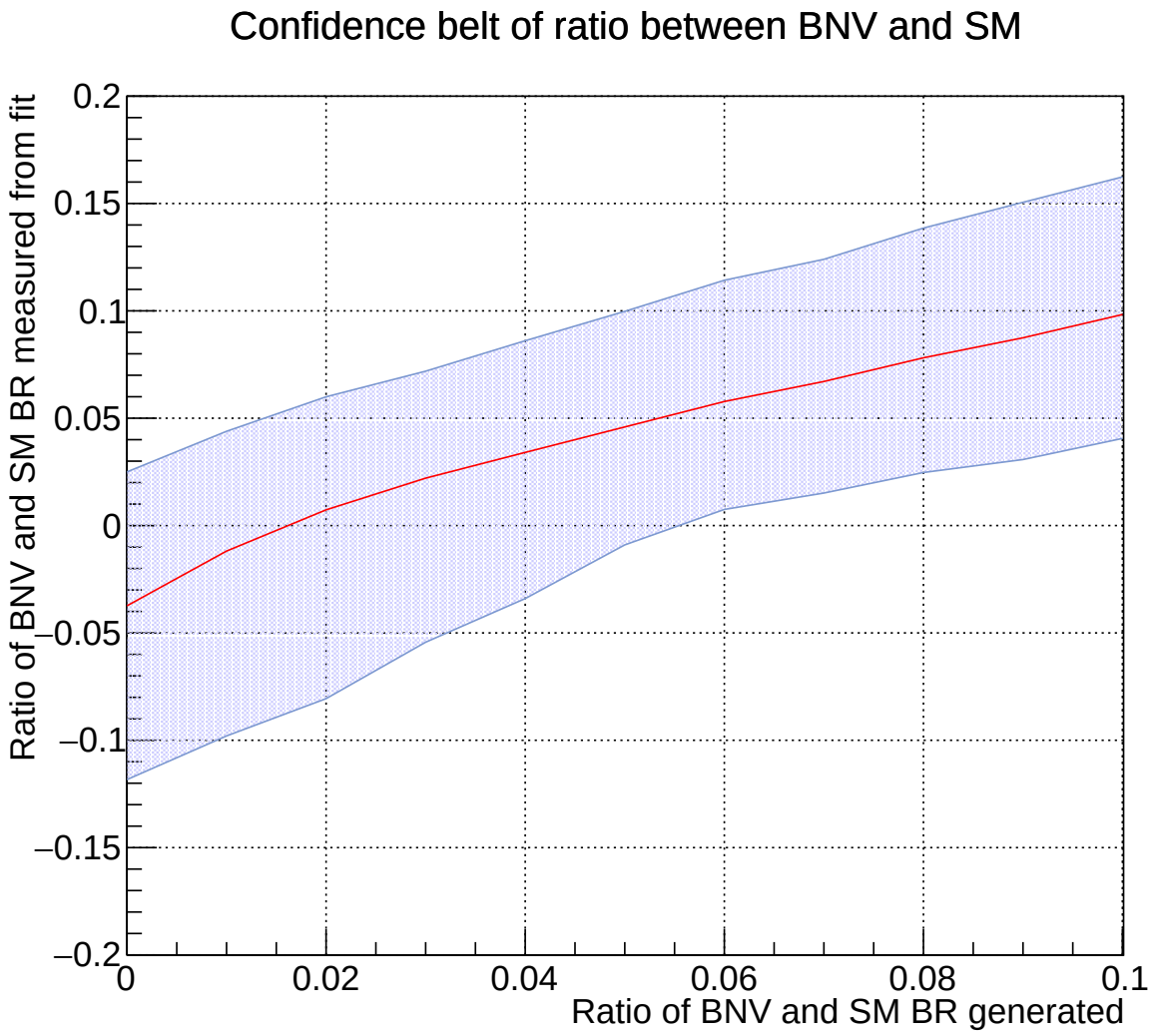


Figure 67: Confidence belt for the ratio between BNV and SM branching fractions constructed with released background PDF shape parameters.

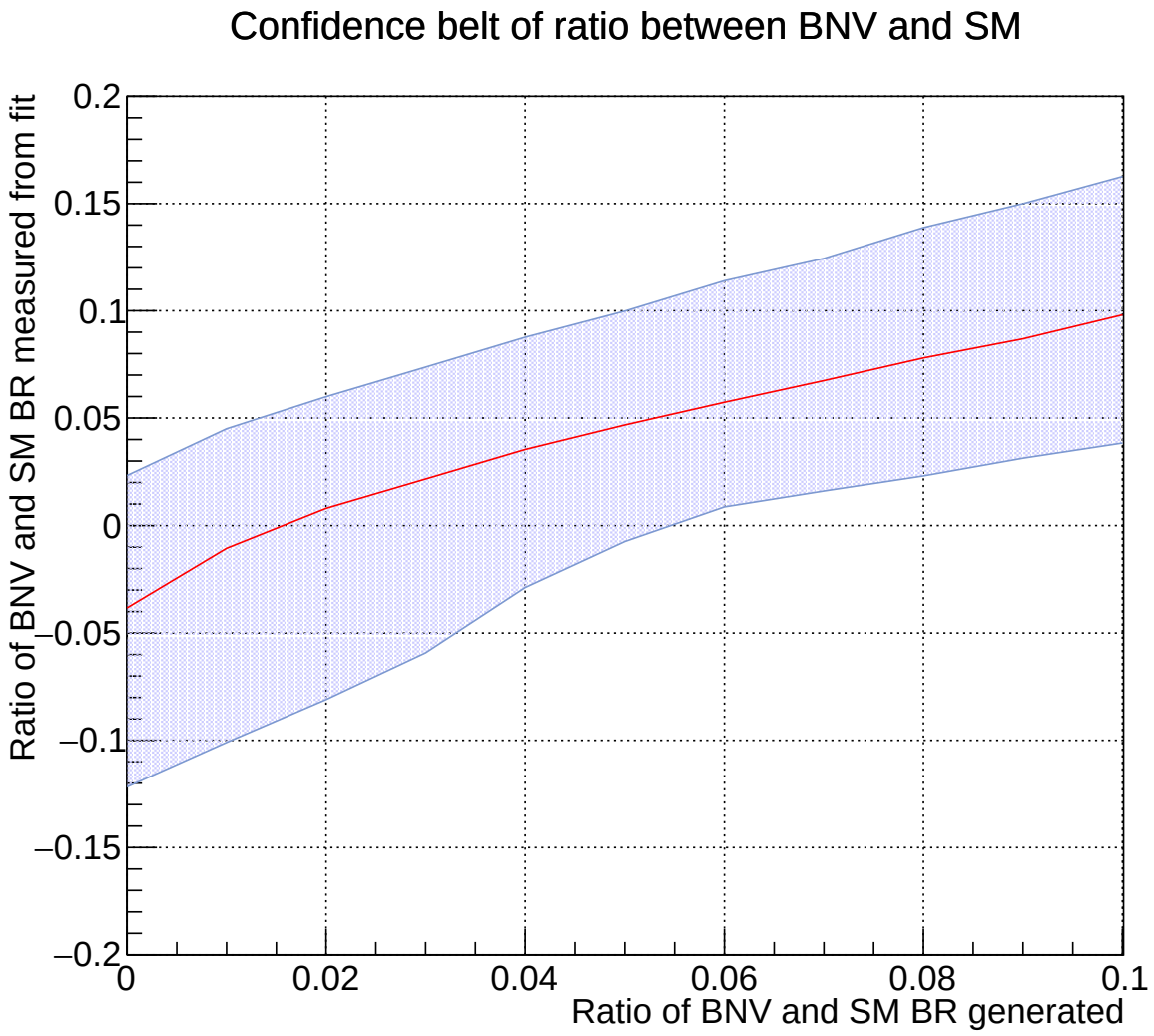


Figure 68: Confidence belt for the ratio between BNV and SM branching fractions constructed including the statistical uncertainties in SM branching fractions with released background PDF shape parameters.

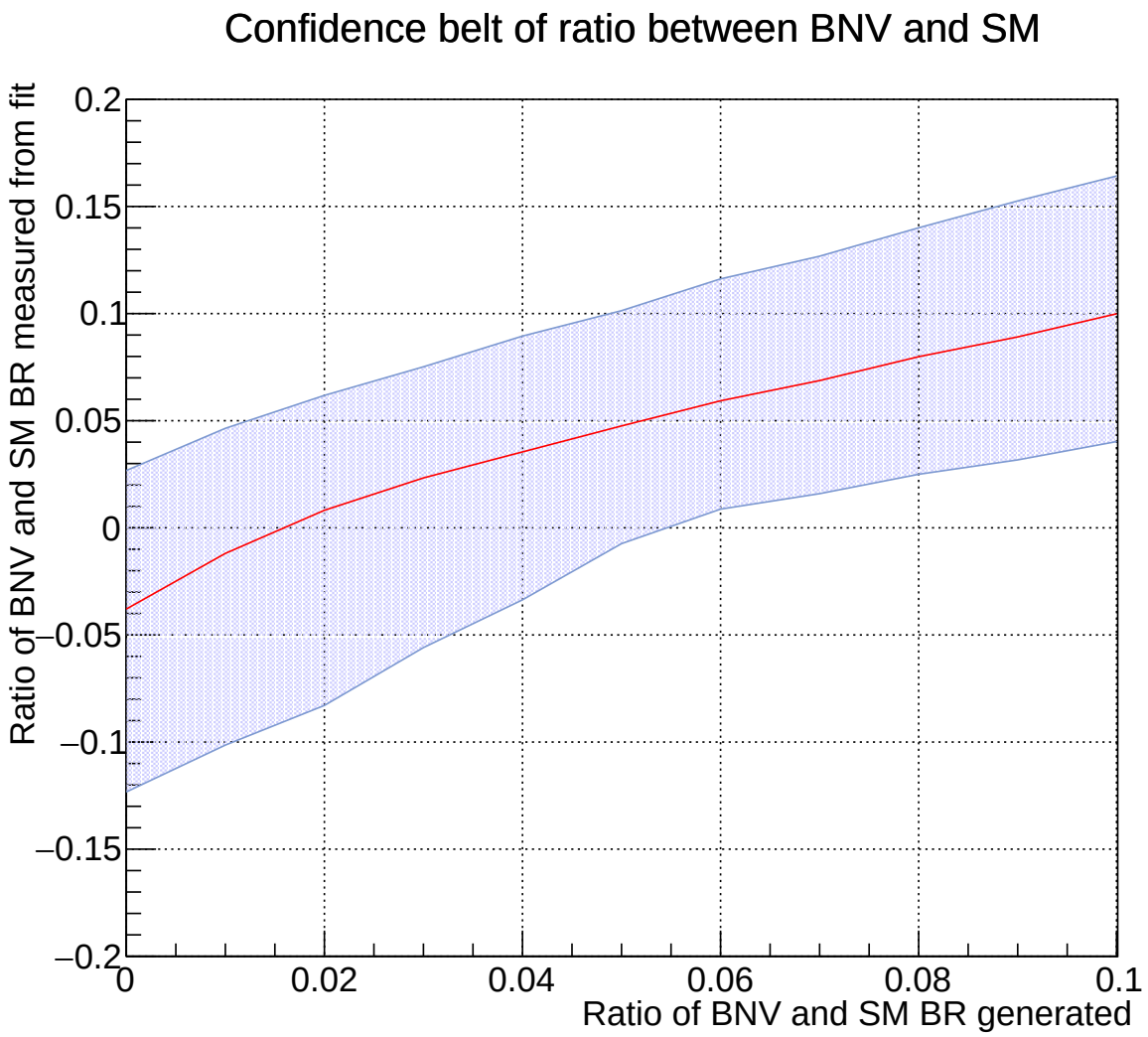


Figure 69: Confidence belt for the ratio between BNV and SM branching fractions constructed including the systematic uncertainties due to PDF shapes with released background PDF shape parameters.

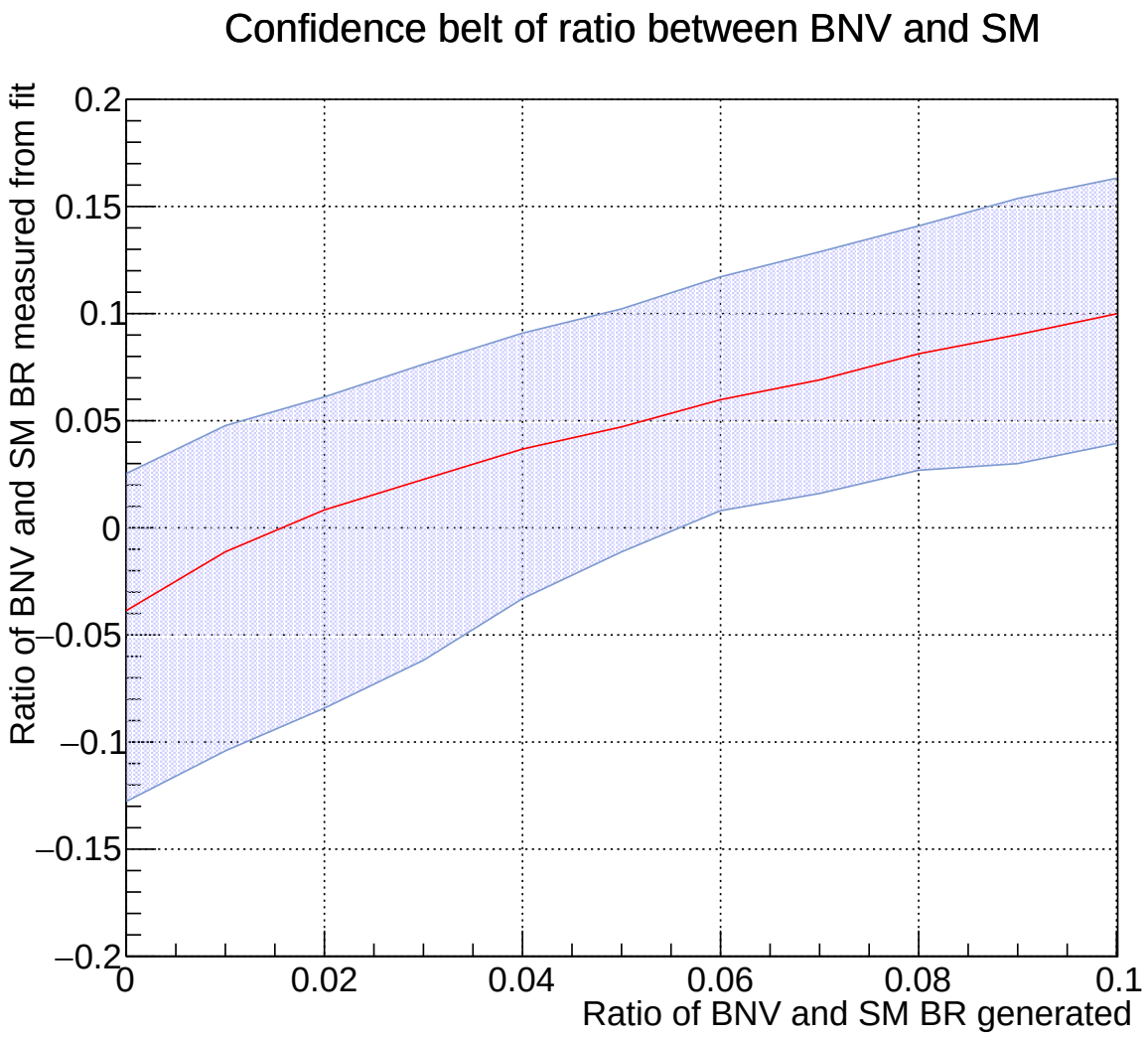


Figure 70: Confidence belt for the ratio between BNV and SM branching fractions constructed including both statistical and systematic uncertainties with released background PDF shape parameters.

4.14.3 Upper Limits with Released Background PDFs

Using the central value for the ratio between BNV and SM branching fractions of -0.0688 measured from data with released background PDFs and using confidence belts shown in Figures 67 through 70, the 95% CL upper limits are set as:

1. Baseline: 2.44%
2. Include statistical uncertainties in SM branching fractions: 2.55%

3. Include systematic uncertainties due to PDF shapes: 2.52%
4. Include both statistical and systematic uncertainties: 2.69%

The sensitivities for the BNV/SM ratio (the upper limit estimated assuming a measured ratio of 0) are:

1. Baseline: 5.55%
2. Include statistical uncertainties in SM branching fractions: 5.46%
3. Include systematic uncertainties due to PDF shapes: 5.46%
4. Include both statistical and systematic uncertainties: 5.58%

Converting 95% CL upper limit 2.7% and sensitivity 5.6% including all statistical and systematic uncertainties with released background PDFs to oscillation angular frequencies, the results become

$$\omega = \frac{\sqrt{R/2}}{\tau} = 0.76\ ps^{-1}$$

for the upper limit and

$$\omega = \frac{\sqrt{R/2}}{\tau} = 1.10\ ps^{-1}$$

for sensitivity.

4.15 CONCLUSIONS

In summary, using the full data sample collected by the Belle experiment at the $\Upsilon(4S)$ resonance, we performed the first search for the baryon-number-violating processes in B^- decay to final state $\overline{\Xi}_c^0\overline{\Lambda}_c^-$. We observe no evidence for baryon number violation and set the upper limit on the ratio between branching fractions for BNV and SM modes to be

$$R = \frac{\mathcal{B}(B^- \to \overline{\Xi}_c^0\overline{\Lambda}_c^-)}{\mathcal{B}(B^- \to \Xi_c^0\overline{\Lambda}_c^-)} < 2.7\% \text{ at } 95\% \text{ C.L.} \tag{4.23}$$

Assuming no direct BNV transitions in Ξ_c^0 decays, we set the upper limit on the $\Xi_c^0 - \overline{\Xi}_c^0$ oscillation angular frequency to be $\omega < 0.76$ ps^{-1} at 95% C.L.

Assuming a zero result for the B^- BNV branching fraction, the sensitivity for the ratio between BNV and SM branching fractions is estimated to be

$$R = \frac{\mathcal{B}(B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-)}{\mathcal{B}(B^- \to \Xi_c^0 \overline{\Lambda}_c^-)} < 5.6\% \text{ at } 95\% \text{ C.L.} \tag{4.24}$$

Under hypothesis of $\Xi_c^0 - \overline{\Xi}_c^0$ oscillations such zero result corresponds to a sensitivity $\omega = 1.10$ ps^{-1} at 95% C.L. for the oscillation angular frequency.

This is the first experimental result on the oscillations in the charmed baryon sector. The time-dependent charmed baryon-antibaryon oscillations will be further explored by the Belle II experiment at the SuperKEKB collider, where improved vertex resolution and expected integrated luminosity make such studies feasible.

5.0 TOP TRIGGER DESIGN REPORT

The development of firmware (FW) for the trigger (TRG) subsystem which uses the information from the time-of-propagation (TOP) sub-detector represents a major part of my PhD work. In this chapter, I first introduce the TOP sub-detector of the Belle II experiment, then I describe the TOP front-end electronics (FEE), and, finally, I discuss the design of TOP-based trigger (TOP TRG).

5.1 IMAGING TIME-OF-PROPAGATION (ITOP) COUNTER OVERVIEW

Particle identification (PID) systems play a key role in B-factory experiments. In particular, K/π discrimination ability is crucial for the reconstruction of B meson decays due to $b \to s$ and $b \to d$ transitions. The PID systems also play an important role in flavor tagging used to measure time-dependent CP asymmetries. As is discussed in Chapter 3, the Belle detector [48] PID system consists of barrel Time-of-Flight (TOF) counters, barrel and endcap Aerogel Cherenkov counters (ACC) and Central Drift Chamber (CDC). The Belle II experiment will eventually operate at the design luminosity of the SuperKEKB collider which is $L = 8 \times 10^{35}$ cm^{-2} s^{-1}, a factor of 40 times higher than that achieved by KEKB. The PID system of the detector was upgraded to cope with the higher beam-related backgrounds expected at SuperKEKB. The Time-of-Flight counters and Aerogel Cherenkov counters are replaced with the imaging Time-of-Propagation (iTOP) counters comprising the TOP sub-detector in the barrel and the redesigned Aerogel Ring Imaging Cherenkov (ARICH) sub-detector in the endcaps. Fig. 71 shows the schematic view of the Belle II detector and the sub-detector locations.

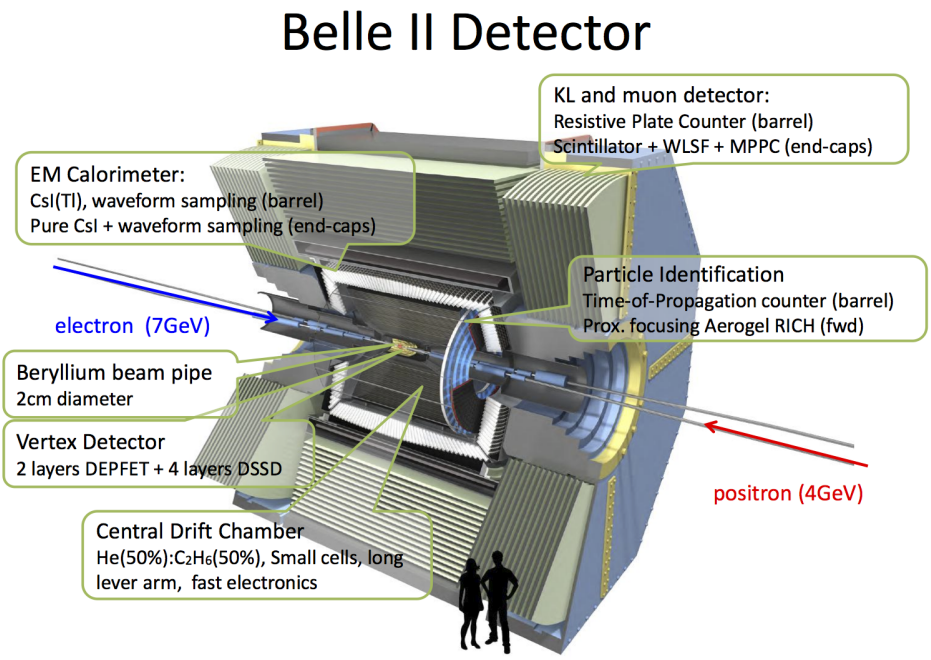
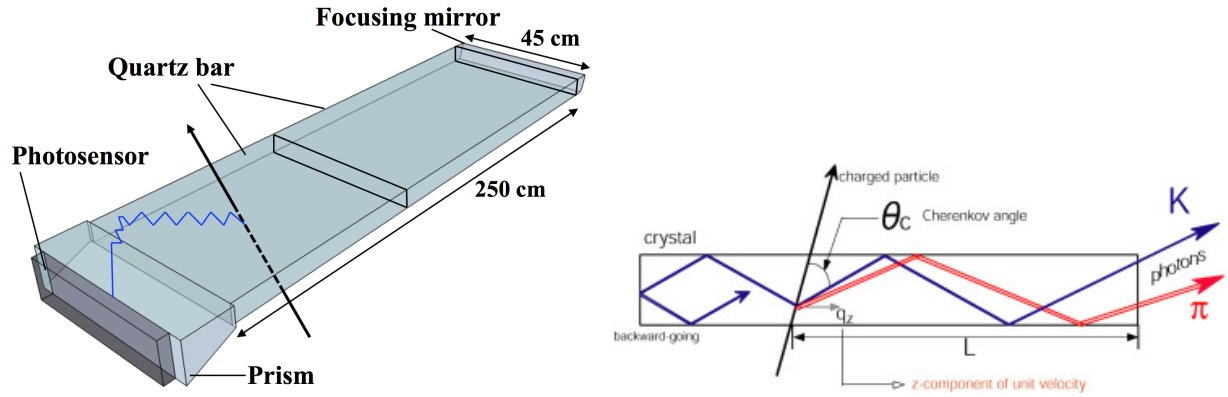


Figure 71: Schematic view of the Belle II detector.

The TOP system consists of 16 TOP modules, arranged cylindrically around the beam pipe between the ECL and CDC. Each TOP module consists of two fused silica bars of 125 cm length, 45 cm width, and 2.0 cm thickness glued together. A reflective mirror with a spherical surface is mounted at the forward end and an expansion prism is mounted at the back end. Attached to the expansion prism is an array of micro-channel-plate photomultiplier tubes (MCP-PMTs). Fig. 72 shows the conceptual view of an iTOP counter module.

Charged particles passing through a medium at a speed faster than the speed of light in the medium emit Cherenkov photons. The emission angle of the photons (Cherenkov angle) depends on the velocity of the charged particle and the refractive index of the radiator material. For a given momentum, each particle species (e.g., $p/K/\pi$) emits photons with a characteristic Cherenkov angle, which can be used to identify the incident particle species. Fig. 72 schematically demonstrates the difference between paths of propagation of Cherenkov photons from pions and kaons of the same momentum.

Due to the high refractive index of the silica bar, Cherenkov photons inside the bar experience total internal reflection, therefore preserving the angular information, until reaching



x

Figure 72: Left: conceptual view of an iTOP counter module. Right: schematic view of the photon propagation in an iTOP quartz bar. The red and blue lines show the path of propagation of Cherenkov photons from a pion and a kaon of the same momentum.

the expansion prism at the back end and then collected by the PMTs. Photons going forward are reflected by the spherical mirror at the forward end, eventually arriving at PMTs at the back end at a later time. The photon arrival time at the PMT arrays depends both on the time-of-flight of the charged particle from the collision point to the TOP detector (determined by the charged particle's velocity) and the time-of-propagation of Cherenkov photons inside the TOP detector (determined by the Cherenkov angle and thus by the charged particle's velocity). Fig. 73 shows a simulation of Cherenkov photon distribution in the horizontal position (i.e., along the longer edge of the expansion prism) and arrival time for 2 GeV/c kaons (blue) and pions (red) at normal incidence on the quartz bar [49].

For each event, the spatial and temporal distributions of the Cherenkov photons detected on the PMT arrays are compared to probability density functions (PDFs) for each particle hypothesis (e, μ, π, K, p) to compute likelihoods. The PDFs are based upon the momenta and incident points on the silica bars of the charged particles measured by CDC. TOP reconstruction algorithm allows a precise measurement of the speed of the incident charged particles and thus the particle species. In order to reach the required particle identification performance goal, the photon arrival times have to be measured with a resolution better than 100 ps [50].

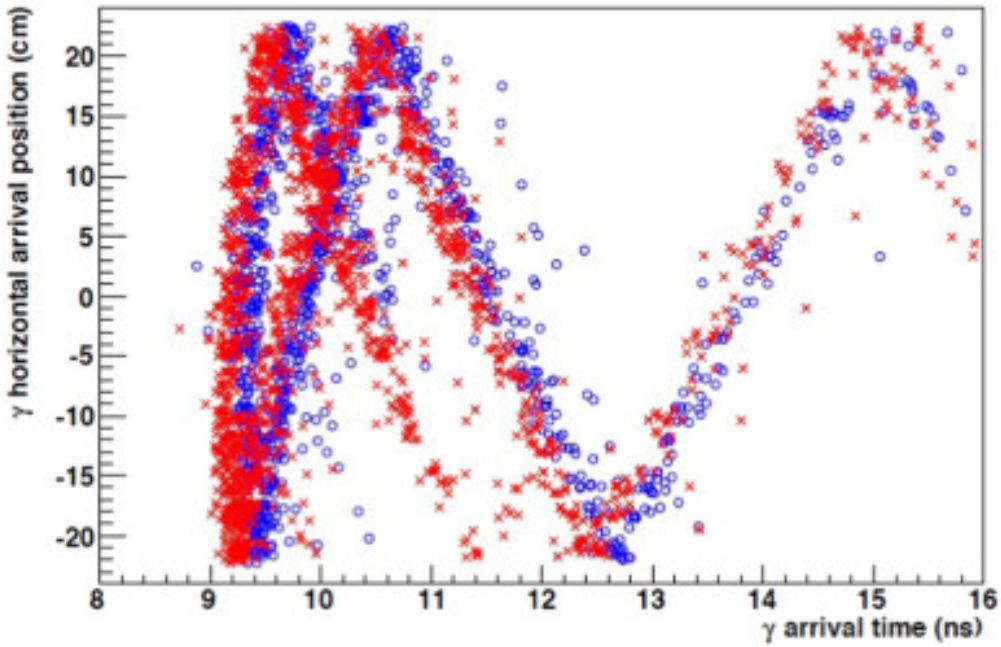


Figure 73: Simulation of Cherenkov photon distribution in horizontal position (i.e., along the longer edge of the expansion prism) and arrival time for 2 GeV/c kaons (blue) and pions (red) at normal incidence on the quartz bar.

5.2 TOP FRONT-END ELECTRONICS

TOP FEE [51] are responsible for reading out the signals of all 512 MCP-PMTs (8192 channels) with a single photon timing resolution of better than 100 ps at a L1 trigger rate of up to 30 kHz at the design luminosity of the SuperKEKB. Upon receiving the L1 decision, TOP FEE identify regions-of-interest (ROIs) in TOP FEE analog sampling memory followed by digitization, feature extraction, and data readout. However, TOP FEE could also provide precise collision time information, therefore contributing to the future L1 timing decision, by generating trigger timestamps and streaming them toward the TOP TRG system implemented in the FPGA fabric of backend electronics located off the detector. Timestamps represent 2ns-precise timing values obtained by the TOP FEE trigger discriminator circuit in response to PMT electric pulses generated in response to photons hitting their sensitive pads. Precise identification of collision time at the L1 trigger level is necessary for correctly identifying ROIs, especially in subsystems overwhelmed by beam-induced background, such

as the Silicon Vertex Detector (SVD). No spatial information (channel number) is currently used for timestamping. In this section, we briefly introduce the design of TOP front-end electronics.

Each of the 16 TOP modules is equipped with 4 board stacks. Each board stack is responsible for the readout of 128 MCP-PMTs. Fig. 74 shows a photograph and a schematic mechanical layout of a board stack.

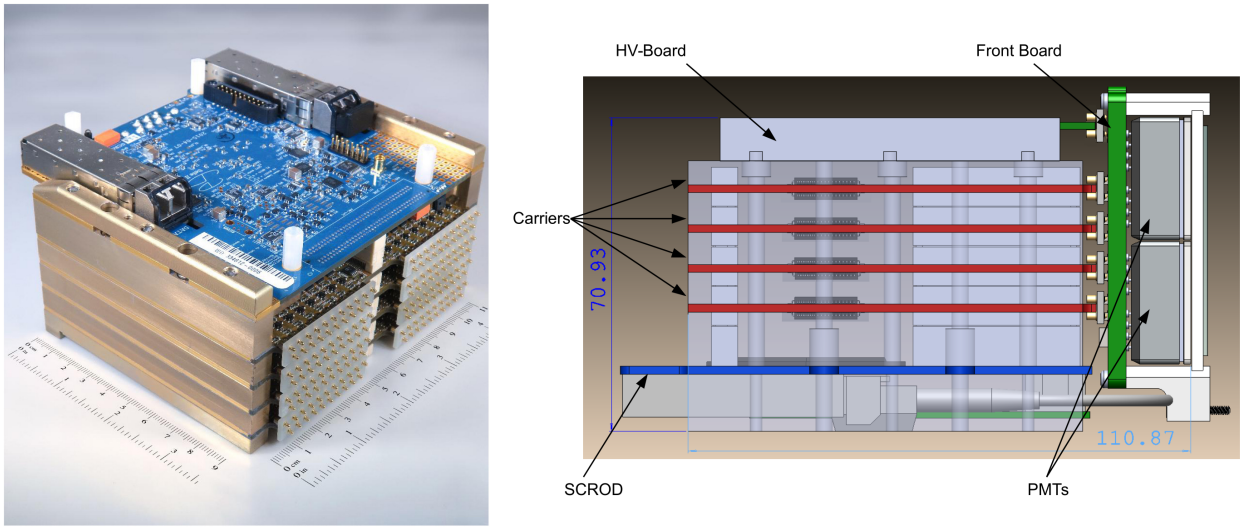


Figure 74: Photograph and a side view schematic mechanical layout of a board stack of the TOP front-end electronics.

The main components of a board stack include:

1. four ASIC Carrier boards, each of which is responsible for the sampling, digitization and transmission of data from 2 MCP-PMTs (32 channels). Each ASIC Carrier board is equipped with four 8-channel custom-designed waveform sampling Ice Ray Sampler version X (IRSX) application-specific integrated circuits (ASICs) [52].
2. one Standard Control Read-Out Data (SCROD) board which is responsible for receiving, processing digitized data from upstream carrier boards and transmitting it to the downstream data acquisition system (DAQ).
3. one High Voltage (HV) board, which supplies high voltages to the PMTs.
4. one Front board, which hosts the PMT array, provides HV connections and interconnects the PMTs and the carrier boards.

5.2.1 ASIC Carrier Board

Fig. 75 shows the top and bottom views of the ASIC Carrier board. Each ASIC Carrier board is equipped with four 8-channel IRSX ASICs. The readout and register control of the ASICs is handled by a 030-series Xilinx Zynq FPGA. The digitized data from the four ASIC Carrier boards are then transmitted to the single Standard Control Read-Out Data (SCROD) board.

IRSX ASIC uses switched-capacitor array memory to sample and store the analog waveforms. The sampling rate for each channel is 2.714 GSa/s. The analog storage is grouped into 512 banks of 64 storage cells, resulting in a total of 32,768 storage cells per channel. The buffer depth allows the storage of approximately 12 μs of the analog waveform. The latency of the Belle II trigger system is 5 μs, which is shorter than the buffer depth.

When a global L1 trigger decision arrives from the Belle II trigger system, regions of interest (ROI), which are 64 storage cells long, in the buffers are identified and digitized by a 12-bit Wilkinson-type analog-to-digital converter (ADC) in parallel. A ramp generator generates a ramp voltage while a 12-bit counter driven by an oscillator is started for each channel. The counter stops when the ramp voltage exceeds the stored voltage inside the capacitor. Therefore the value of the counter is proportional to the stored voltage. The digitized data are subsequently transmitted to SCRODs.

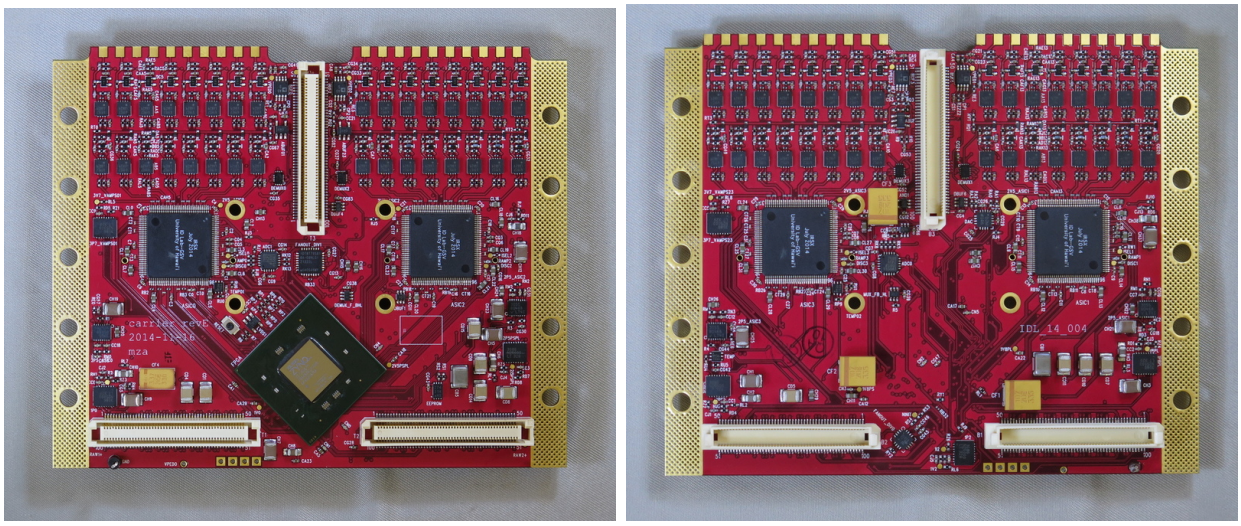


Figure 75: Top and bottom views of the ASIC Carrier board.

In addition to the TOP main data readout path described earlier, the IRSX ASIC also

features an independent data path called the trigger path. The trigger path on the ASIC Carrier board is responsible for generating and streaming trigger timestamps to the TOP TRG system that eventually contributes to the Belle II trigger decision-making. Whenever the analog signal from a PMT exceeds a programmable threshold, the trigger circuitry on the IRSX ASIC records the time and generates 16-bit trigger hits (timestamps). The trigger circuitry operates at 508 MHz and thus these timestamps carry 2 ns-precise timing information, which is much coarser than the main readout path. Four ASIC Carrier boards continuously stream out the timestamps to the SCROD Board. Timestamps are streamed from the Carrier boards to the SCROD at 32MHz (i.e., on average, at 16MHz/PMT) and at 128MHz from the SCROD to TOP trigger backend (i.e., TOP TRG).

It is important to highlight some of the differences between the TOP main readout path and the trigger path. The main readout path operates in burst mode, where it digitizes and transfers data packets (with precise waveforms) only when the global L1 trigger decision arrives. The trigger path operates in a streaming mode, continuously sending coarse timestamps to the backend. The main readout contains primarily valuable data digitized from narrow (in time) ROIs identified according to collision times associated with useful physics processes, while the trigger path is overwhelmed by the ever-present beam-induced photon background.

5.2.2 Standard Control, Read-Out, and Data (SCROD) Board

Fig. 76 shows the top and bottom views of the SCROD board. Its main component is a 045-series Xilinx Zynq FPGA. The SCROD receives digitized data from upstream carrier boards and performs pedestal subtraction and feature extraction[1] to calculate the hit time of the photon on the PMTs. The feature-extracted data are subsequently transmitted to the DAQ downstream. Furthermore, the SCROD board is responsible for distributing clock signals and timing of L1 triggers to the ASIC Carrier boards. Additionally, it merges the timestamp streams from the four ASIC Carrier boards into a single stream, which is then

[1] Alternatively, feature extraction could be performed in the DAQ off the detector. Such a scheme requires a wider bandwidth which is now available after the original DAQ of Belle II, COPPER [53], was upgraded to PCIe40 [54].

forwarded to the TOP TRG system downstream.

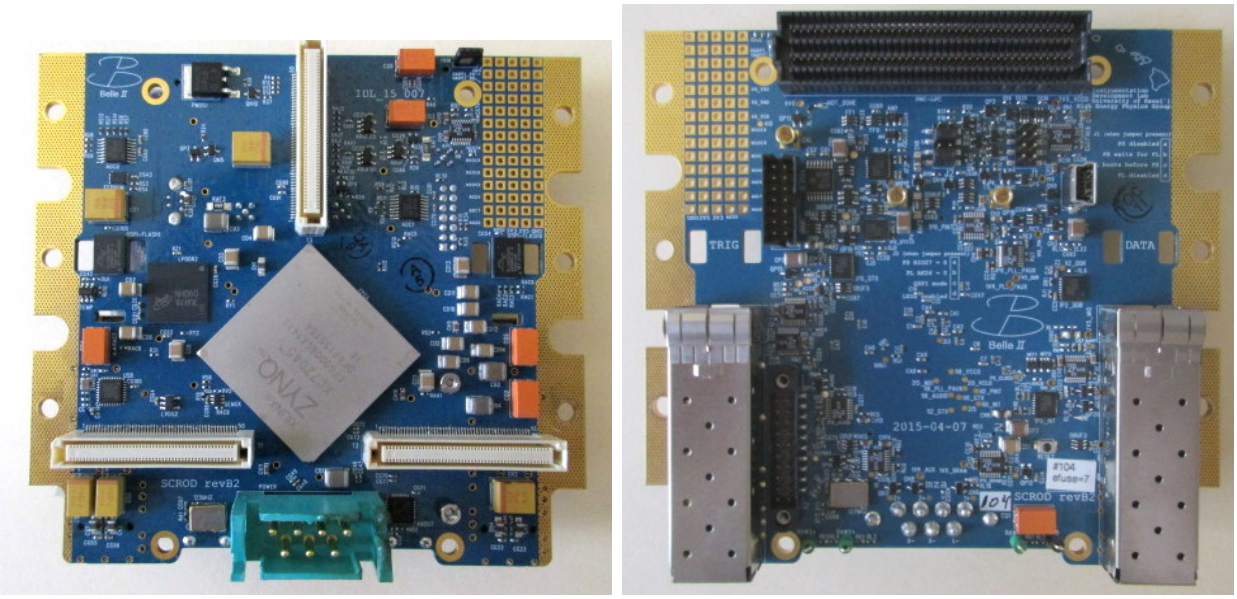


Figure 76: Top and bottom view of the SCROD board.

Fig. 77 shows the block diagrams of the SCROD board and the ASIC Carrier board.

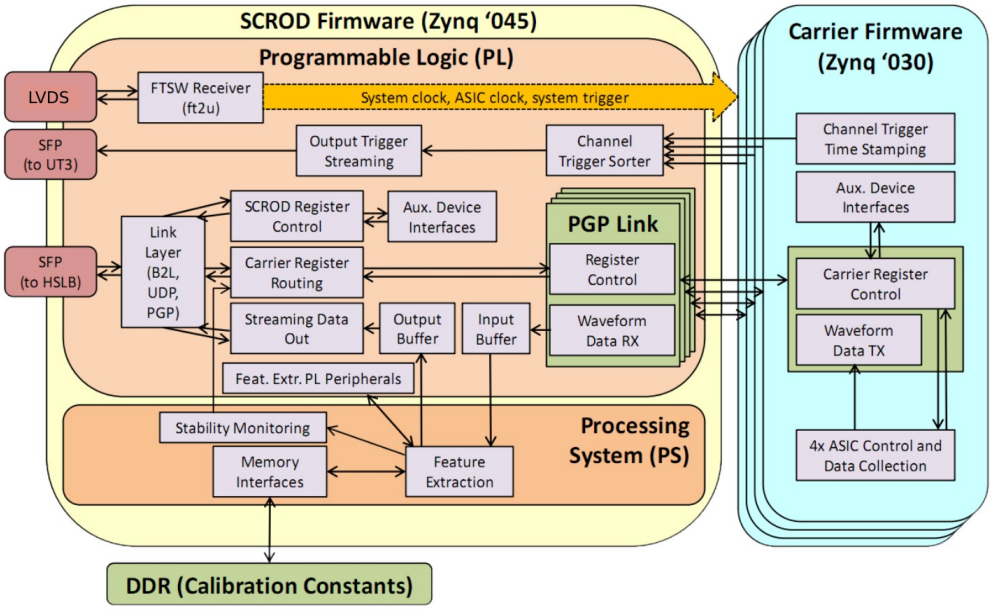


Figure 77: Block diagrams of the SCROD and the ASIC Carrier boards.

5.3 TOP TRG SYSTEM

The primary goal of the TOP TRG system is to provide precise collision timing (t_0) information to the Belle II trigger system. At the design luminosity of Super-KEKB, where the SVD sampling window will have to be significantly tightened to suppress the out-of-time hits, precise collision timing information would be necessary to reduce the data volume. To achieve full physics potential with Super-KEKB, the TOP TRG system is required to operate at almost 100% efficiency for hadronic events and collision timing identification precision of several ns. In addition, currently, TOP is the only subsystem capable of providing the required timing resolution for barrel dimuon events, which are critical for luminosity measurements and detector calibration.

The TOP TRG system consists of two 3rd generation Universal Trigger (UT3) boards, which will be discussed in detail later. Each UT3 board receives trigger timestamps from 8 TOP slots through 4x8 optical links. The data from four SCRODs of the same slot are merged and used to estimate the collision timing using a likelihood-based algorithm. Tentative timing decisions are made in parallel for 16 slots using 16 copies of the algorithm running on two UT3 boards. These slot-level timing decisions are collected on one of the boards, where a combined timing decision is made by correlating tentative timing decisions made for individual slots. The combined timing decision is then sent to the Global Decision Logic (GDL) system within a latency of 3.2 μs (w.r.t. presumed collision time). The slot-level timing decisions are also sent to the Global Reconstruction Logic (GRL) system for CDC-TOP matching.

In the following sections, I introduce the hardware, firmware design, and performance analysis of the TOP TRG system.

5.3.1 Hardware

The 3rd generation Universal Trigger (UT3) board is a custom-built FPGA board widely used in the Belle II trigger system. UT3 board is based on Xilinx Virtex-6 XC6VHX565T-FF1923, which contains 566,784 logic cells, 864 DPS48E1 slices, 32 Mb Block RAM, 40

GTX transceivers, and 24 GTH transceivers. GTX transceivers could operate at a line rate between 480 Mbps and 6.6 Gbps. GTH transceivers could operate at a line rate between 2.488 Gbps and 11.18 Gbps.

The UT3 board is composed of three single-slot 1U VME sub-boards, which are the main board, GTX board, and LVDS board. Fig. 78 to 81 show to top and bottom view of the UT3 board. Fig. 82 shows the block diagram of the UT3 board. On the UT3 main board, 24 GTH transceiver channels of the FPGA are connected to 6 quad (4-channel) small form-factor pluggable (QSFP+) optical transceivers. Each of these QSFP transceivers supports a data transfer rate of up to 4x10.3 Gbps. 40 GTX transceiver channels of the FPGA are connected to 10 4x6.25 Gbps QSFP+ optical transceivers on the UT3 GTX board.

Each of the main and the LVDS boards has a 64-channel LVDS connector, providing support for up to 128 pairs of LVDS signals. Furthermore, the main board has the Versa Module Eurocard (VME) interface and 256 Mb flash memory for configuration, while the LVDS board supports NIM, RJ45, and JTAG interfaces.

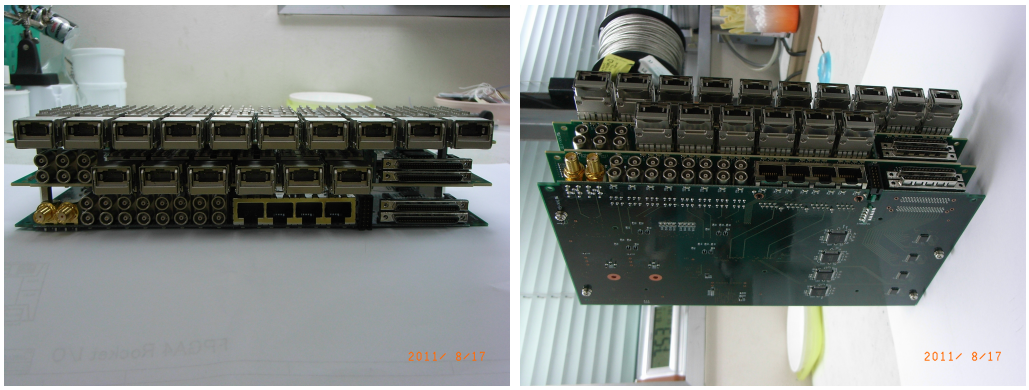


Figure 78: Top and bottom views of the UT3 board.

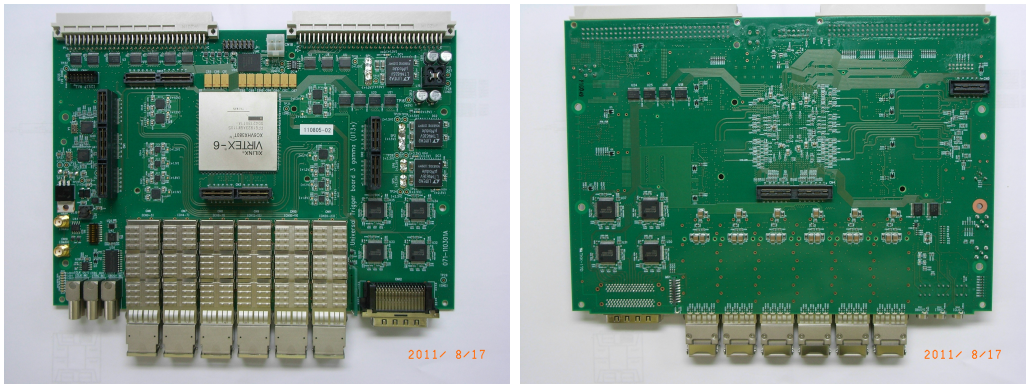


Figure 79: Top and bottom views of the UT3 main board.

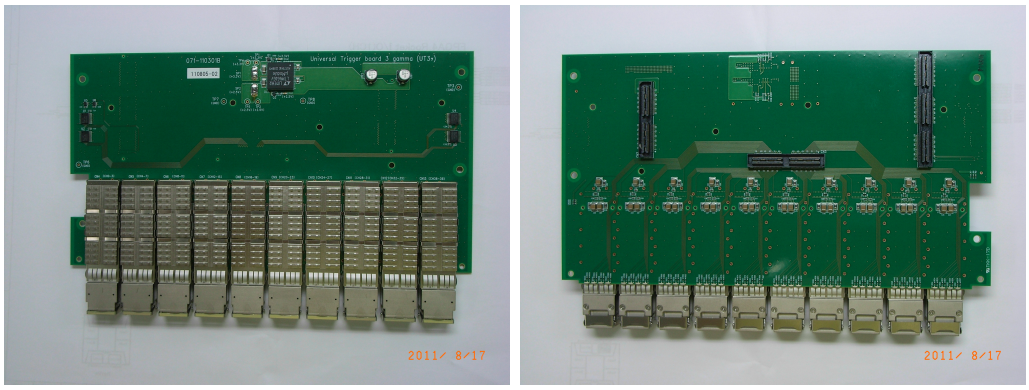


Figure 80: Top and bottom views of the UT3 GTX board.

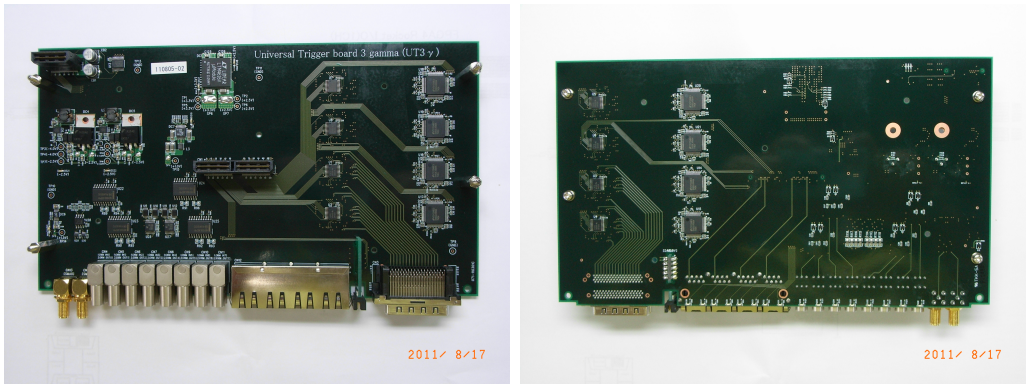


Figure 81: Top and bottom views of the UT3 LVDS board.

In 2022, the TOP TRG system was upgraded from UT3 to the 4th generation Universal Trigger (UT4) boards. The UT4 board is based on Xilinx UltraScale XCVU190-FLGB2104, which contains 2,349,900 system logic cells, 1800 DPS48E1 slices, 132.9 Mb Block RAM, 40 GTH transceivers and 36 GTY transceivers. GTH transceivers could operate at a rate between 0.5 Gbps and 16.3 Gbps. GTY transceivers could operate at a rate between 0.5 Gbps and 30.5 Gbps.

Similarly to UT3, the UT4 board is also composed of three boards, which are the main board, subboard, and LVDS board. Fig. 83 shows the block diagram of the UT4 board. 32 GTY transceiver channels are connected to eight 4x25.8 Gbps QSFP+ optical modules on the UT4 main board. 32 GTH transceiver channels are connected to eight 4x16.3 Gbps QSFP+ optical modules on the UT4 subboard. The Xilinx Artix-7 XC7A15T FPGA on the UT4 subboard is responsible for the VME interface and the configuration of the main FPGA from a 512 Mb flash memory. The UT4 LVDS board supports LVDS, NIM, RJ45, and JTAG interfaces.

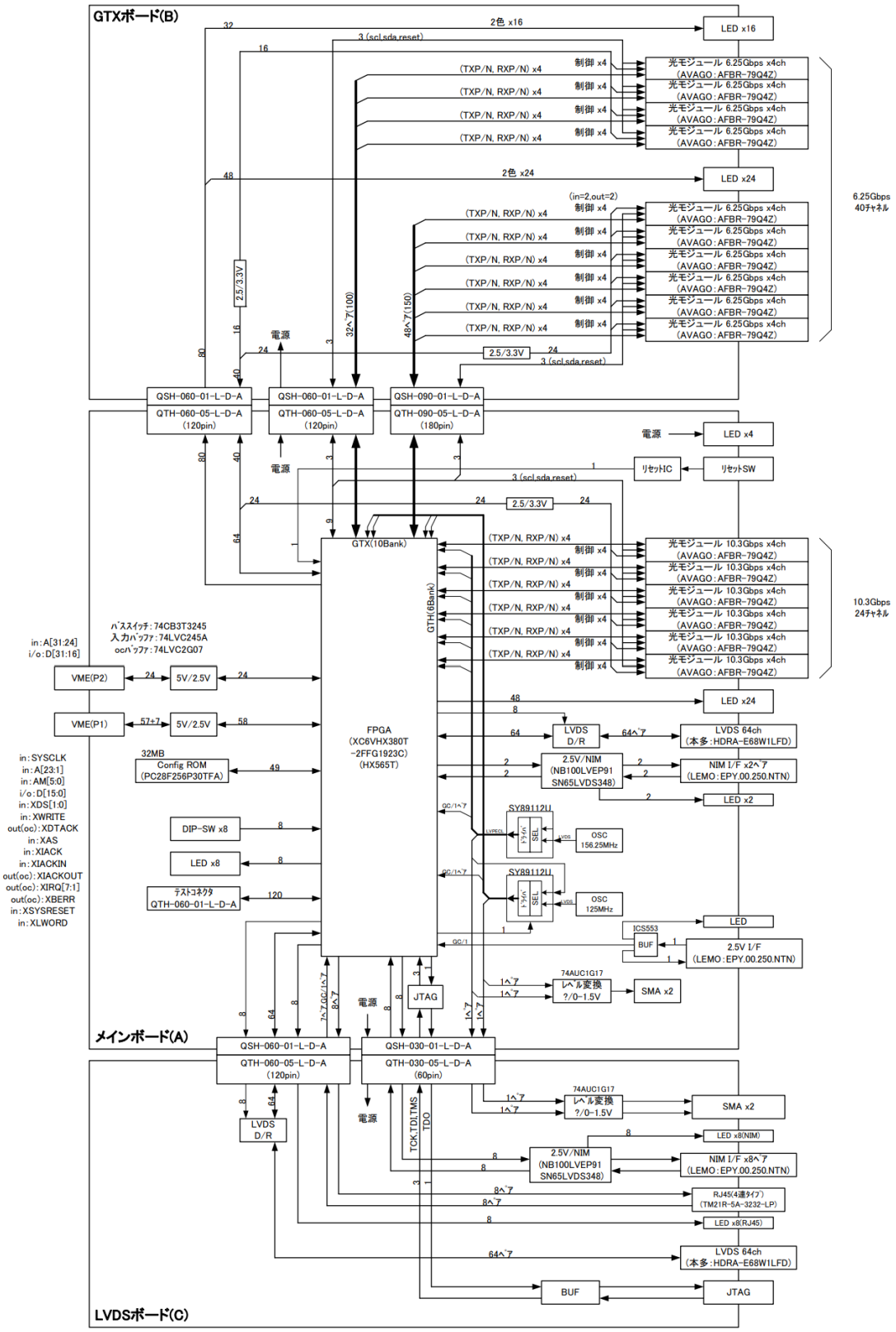


Figure 82: Block diagram of the UT3 board.

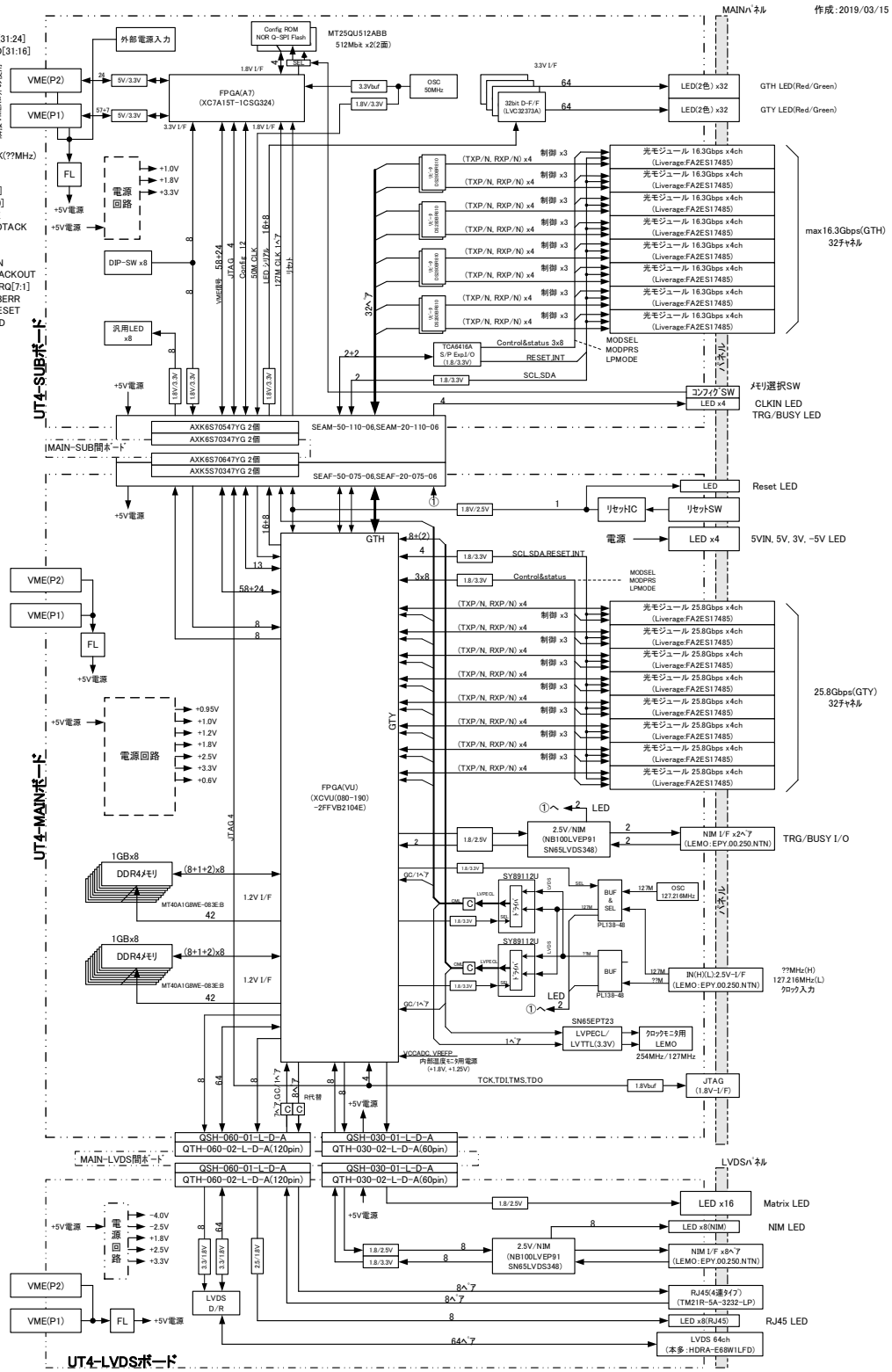


Figure 83: Block diagram of the UT4 board.

5.3.2 Design of TOP Trigger Firmware

Fig. 84 and 85 show the block diagrams of firmware running on TOP TRG TRANSMIT and RECEIVE boards.

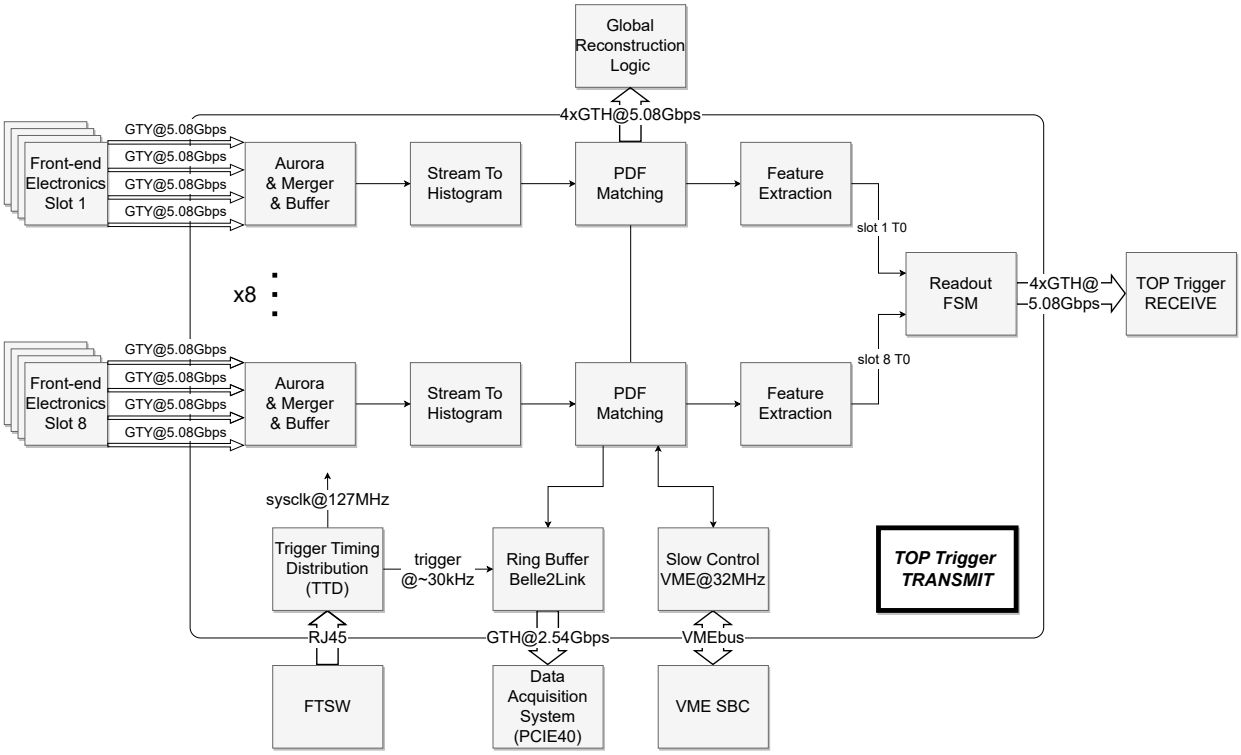


Figure 84: Block diagram of TOP TRG TRANSMIT board.

TOP TRG TRANSMIT board receives trigger timestamps from TOP slots 1 through 8 (upper half of the TOP detector), calculates the slot-level timing decisions based on a sliding window likelihood-based algorithm and transmits the results to TOP TRG RECEIVE board.

TOP TRG RECEIVE board receives trigger timestamps from TOP slots 9 through 16 (lower half of the TOP detector) and calculates slot-level timing decisions in the same way as the TRANSMIT board. In addition, it receives timing decisions for slots 1 through 8 from TOP TRG TRANSMIT board and makes the combined decisions by correlating the slot-level decisions of all 16 slots.

In the previous design of the TOP TRG system, a fixed time window algorithm was employed. This algorithm identified potential physics events by detecting multiple timestamps with similar values and the subsequent timestamps within a predefined fixed time range were matched with different PDFs. A long-standing challenge in such design was posed by the

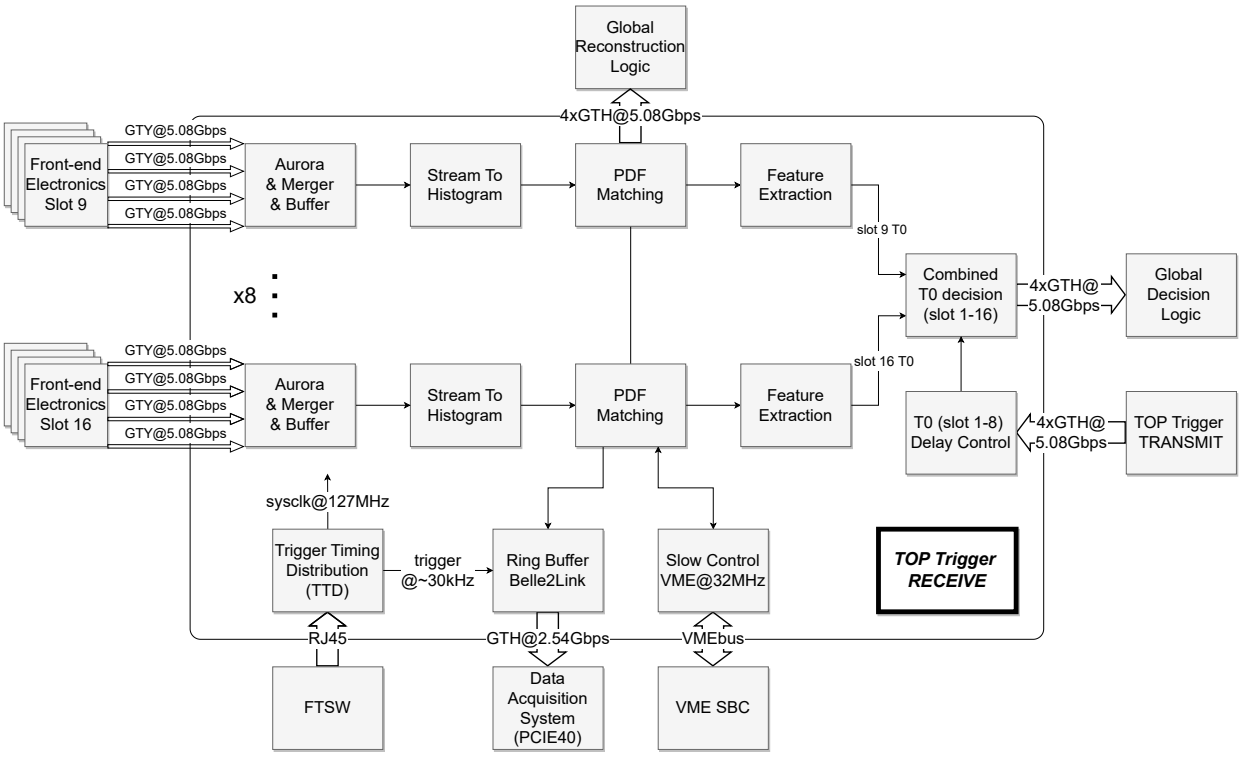


Figure 85: Block diagram of TOP TRG RECEIVE board.

"first signal hit" problem. Since occurrences of bunches of signal hits are rare, the probability that time windows are triggered by background-related hits gets higher as the background increases. Additionally, the fixed time range sometimes resulted in time windows ending in the middle of a cluster of signal hits. Consequently, the fixed time window algorithm proved to be efficient mainly for cosmic events but exhibited poor performance during collisions.

To improve trigger efficiency and timing resolution, a sliding window algorithm was developed and implemented. The sliding window algorithm consists of three major FW modules: stream to histogram, PDF matching, and feature extraction (Fig. 84). Histograms of trigger hits (timestamps) are being continuously updated in the stream to histogram module. The PDF matching module matches the timestamp histogram with several PDFs and evaluates tentative likelihoods and timing decisions on each clock cycle. The feature extraction module identifies the local maximum of the likelihood distribution, which represents the slot-level timing decision.

Fig. 86 shows the schematics of the fixed time window and the sliding window algorithms.

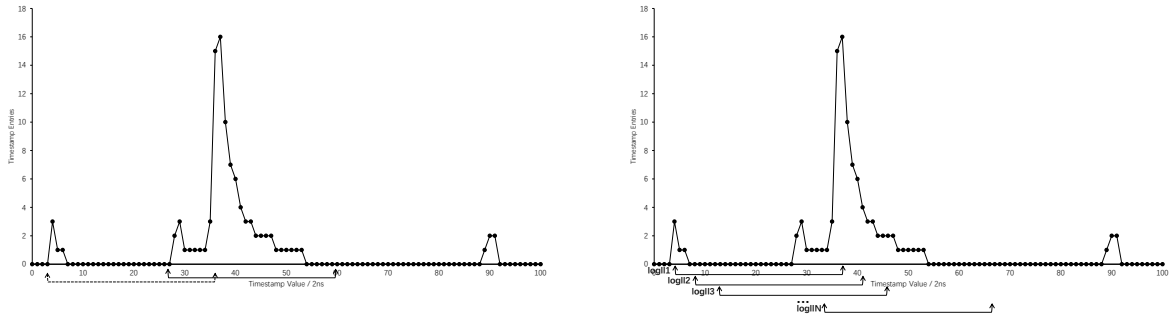


Figure 86: Schematics (on the left) of the fixed time window and (on the right) the sliding window algorithms.

In the following section, I discuss the features of each module shown in Fig. 84.

5.3.2.1 TOP FEE Data Link for Timestamp Stream The data transmission between TOP FEE and TOP TRG employs the Aurora 8B/10B protocol, which is a high-speed serial link layer communication protocol used for point-to-point communications. The Aurora 8B/10B IP core utilized on the TOP FEE side is based on Xilinx 7 Series GTX transceiver and the one used on the TOP TRG side is based on Xilinx UltraScale GTY transceiver. The Aurora core is configured as a duplex streaming interface. The line rate for each core is set to be 5.08 Gbps, supporting the 32 bit continuous timestamp stream at 127 MHz between each TOP SCROD and TOP TRG system. Since the TOP detector consists of 16 slots, with each slot comprising 4 SCRODs, there are 64 5.08 Gbps Aurora links in total. The input timestamp bandwidth for each UT4 board is 162.56 Gbps (32x5.08 Gbps).

Aurora 8B/10B IP core is also responsible for the lane initialization, the comma symbol code-word detection and alignment operation, clock compensation, error handling, and channel bonding.

5.3.2.2 Input Timestamp Buffer and Merger Before being used by the likelihood calculation module input timestamp streams from the four SCRODs of the same TOP slot are merged into one stream. To achieve this, the four timestamps streams from the Aurora core are individually written into four asynchronous dual-clock First-In-First-Out (FIFO) buffers. Then the round-robin merger module takes turns reading from these four FIFOs

and merges the data into a single stream.

The FIFOs have a depth of 256, which is longer than the number of timestamps generated in a typical hadronic event (approximately 100). This ensures that the buffers can handle the burst events without overflowing.

The asynchronous FIFOs also play an important role in clock domain crossing (CDC) between the Aurora core (operating at 127 MHz) and the likelihood calculation module (operating at 127/254 MHz). The Aurora core clock serves as the FIFO write clock and the likelihood calculation module clock serves as the FIFO read clock.

It is important to note that there is a disparity between the write and read bandwidth of the FIFOs. Even if the likelihood calculation module operates at 254 MHz, the FIFOs can overflow when the four SCROD timestamp streams exceed 63.5MHz (i.e. 8MHz/PMT). Further optimization and parallelization strategies would be necessary.

Finally, the single timestamp stream after the merger is further buffered and then streamed to the histogram module.

Fig. 87 shows the block diagrams of the buffer and merger module in TOP TRG FW, including the related clock domains.

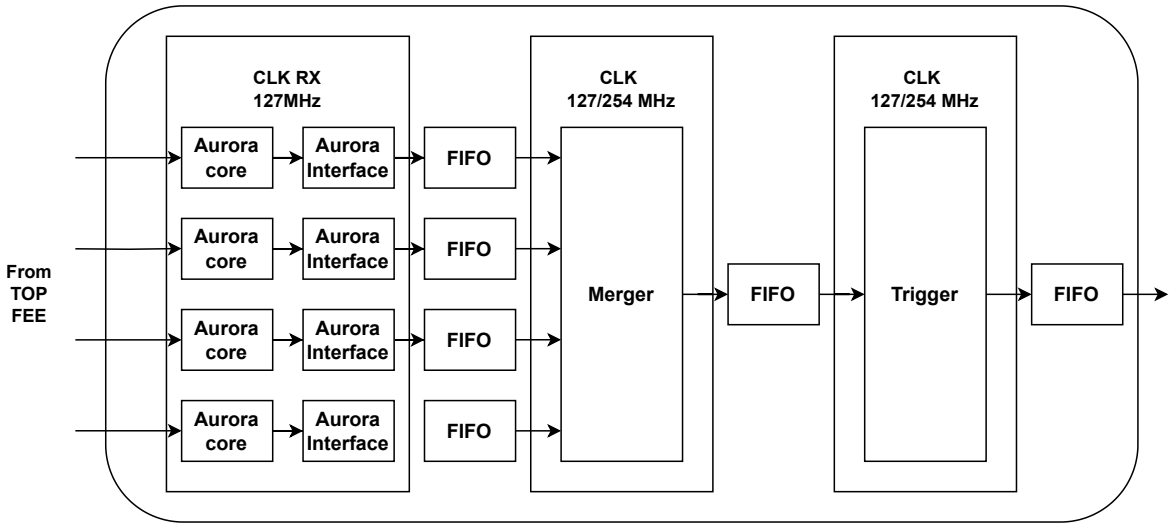


Figure 87: Block diagrams of TOP TRG input buffer, merger module, and related clock domains.

5.3.2.3 Stream to Histogram Module Before matching with PDFs the input timestamp stream should be transformed into a timestamp histogram. The stream to histogram

module is responsible for maintaining a circular buffer with a width of 64 bins (128 ns), where each bin records the number of timestamps with specific values. The incoming timestamps are added to the circular buffer on every clock cycle. Within the 64-bin circular buffer, a 32-bin width timestamp histogram, which represents a time window of 64 ns, is used to match with PDFs. The boundary of the timestamp histogram advances based on the largest timestamps received so far.

Fig. 88 shows an example of how the timestamp histogram is updated in the stream to histogram module. In the top figure, the current 64-bin circular buffer stores the number of timestamps within the specific range [5974, 6037]. The incoming timestamp value is 6032 and thus the range of the 32-bin timestamp histogram is between 6001 and 6032.

The middle figure demonstrates the condition where a new incoming timestamp value of 6030 falls within the range of the existing histogram ([6001,6032]). In this case, the counter corresponding to the position of 6030 inside the histogram is incremented and the boundary of the histogram remains unchanged.

In the bottom figure, a new timestamp value of 6035 is outside the range of the current histogram. As a result, the histogram boundary aligns with the value of the new timestamp. The counters outside the range of the histogram are cleared accordingly.

Fig. 89 shows the finite state machine (FSM) implemented in the stream to histogram module. The module initiates in the IDLE state and transitions to the UPDATE state whenever a new timestamp arrives. During the UPDATE state, the timestamp histogram is updated following the discussed mechanisms. A corruption detection mechanism is implemented: a timestamp is identified as corrupted and skipped as long as its deviation from the two preceding and following timestamps exceeds a certain threshold. This mechanism is important for preventing the histogram from being broken by a single corrupted timestamp. The module enters the WAIT state when no timestamp arrives. If no timestamp is received for N clock cycles while in the WAIT state, the module transitions to the SLIDE state. During the SLIDE state, the boundary of the timestamp histogram shifts one bin per clock cycle to match with the fixed PDFs. This is equivalent to matching the shifted PDFs with fixed timestamp histograms. Upon the arrival of a new timestamp during the WAIT and SLIDE states, the FSM returns to the UPDATE state to update the timestamp histogram accord-

ingly. After the timestamp histogram has shifted for m clock cycles during the SLIDE state, the FSM progresses to the FINISH state and finally returns to the IDLE state. The values of n for the WAIT state and m for the SLIDE state are adjustable parameters, allowing for optimization based on specific event types and background conditions.

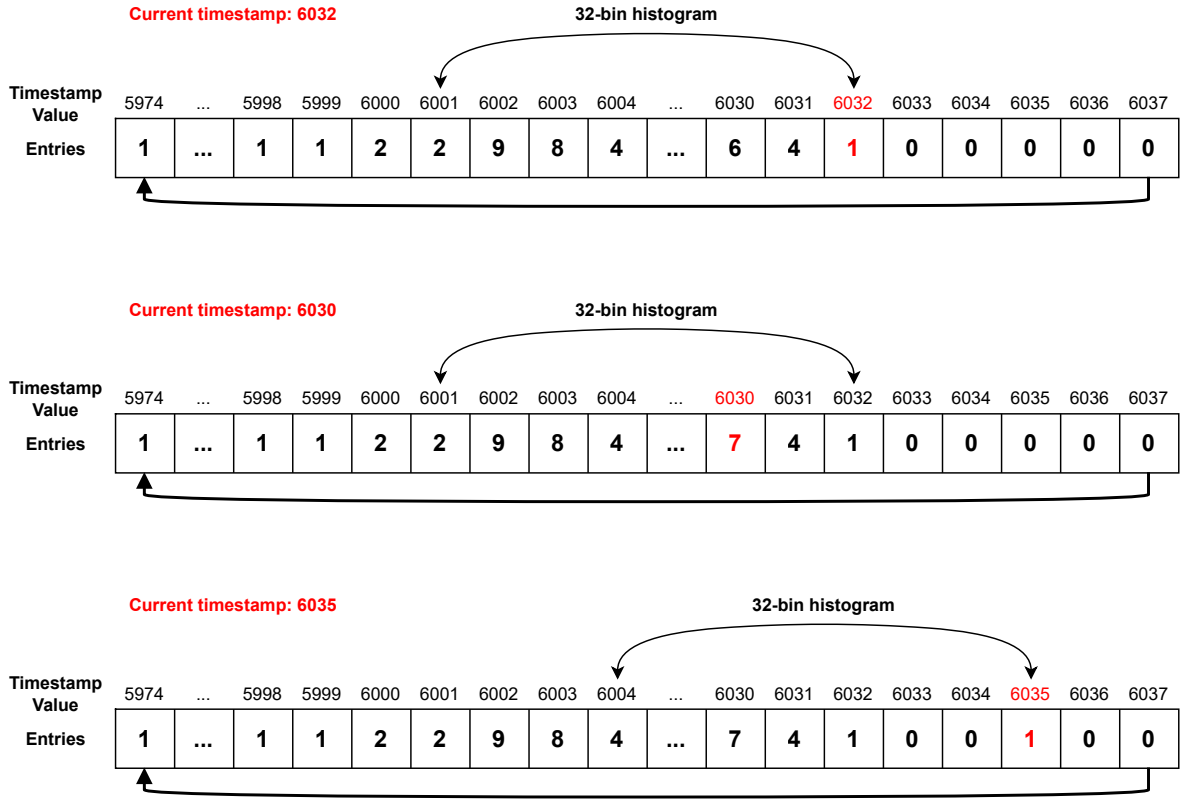


Figure 88: Schematic of the process of updating timestamp histogram.

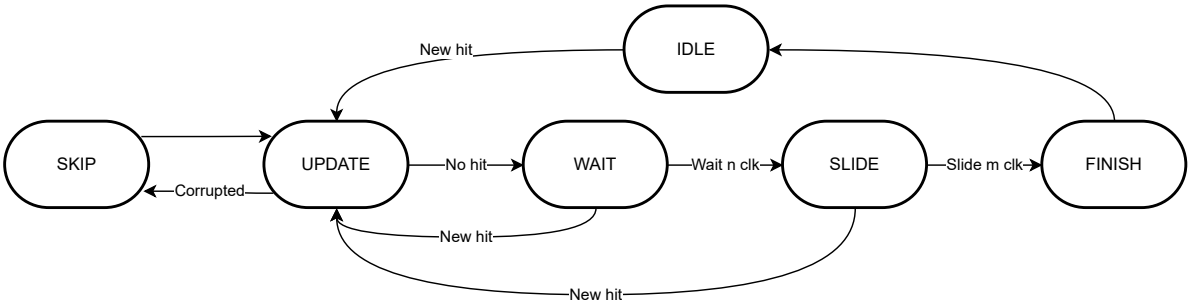


Figure 89: Schematic of the finite state machine implemented in the stream to histogram module.

5.3.2.4 PDF Matching The TOP TRG system utilizes a likelihood-based algorithm to determine the incident position of the charged particles on the TOP bars and the collision time (t_0).

The arrival time probability density functions (PDFs) of Cherenkov photons detected by the PMTs strongly depend on the incident position of charged particles on the TOP bars. Fig. 90 shows the three representative PDFs that illustrate different scenarios of a 3 GeV pion interacting with the TOP bar.

In the first PDF, the charged particle hit the back end of the TOP bar. Generated Cherenkov photons immediately reach PMTs and lead to a single peak in the PDF. A small fraction of photons travel forward to the end of the bar and are reflected back toward the PMTs. These photons result in a small tail in the PDF.

The second PDF represents the case where the pion hits the bar at normal incidence. In this case, the PDF exhibits two peaks of similar height. The first peak corresponds to those photons that move backward toward PMTs, while the second peak corresponds to photons that move forward and are reflected at the end of the TOP bar. Thus the second peak is delayed by approximately 20 ns compared to the first peak.

The third PDF corresponds to the case where the pion hits the forward end of the TOP bar. In such case, all generated photons must travel the entire length of the bar in a backward direction to reach the PMTs. Thus the single peak in the PDF is delayed by around 20 ns compared to collision time. Also, the peak is wider than in the first example.

Both TOP main readout and TOP TRG system use PDF-based algorithms, but it is important to note the differences:

1. TOP main readout is designed for particle identification, while TOP TRG aims to measure the incident position of the charged particles on the TOP bars and the collision time (t_0). The difference between the Cherenkov angles ($cos\theta_c = 1/n\beta$) of a 3 GeV kaon and pion is 0.65 degrees, which correspond to approximately 50 ps photon arrival time differences per meter of propagation in the normal incidence case. The time resolution of TOP main readout is approximately 100 ps, and it is thus capable of identifying the particle species aided by the reconstructed CDC tracks which inform the reconstruction about the origin of photons. In contrast, the time quantization of TOP TRG is 2 ns, and it does not have the resolution necessary for particle identification. The original design of TOP TRG does not use CDC TRG information therefore relying on PDF shapes to establish the collision time.

2. As discussed in the previous section, the photon arrival time on the PMT array depends on both the time-of-flight of the charged particle from the interaction point to the TOP detector and on the time-of-propagation of Cherenkov photons inside the TOP bar. The time-of-flight depends on the velocity (or momentum) of a charged particle. The reconstruction combines TOP main readout data with the track information from CDC to determine the time-of-flight (i.e., charged particle's point of incidence). In TOP TRG, and, generally, online, we do not have such CDC information promptly available. Fortunately, considering a 1 GeV kaon and 4 GeV kaon both at normal incidence to the TOP bars (located 1.2 m of radial distance away from the interaction region), the time-of-flight difference between them is approximately 0.4 ns, which could be ignored as compared to the current time resolution of TOP TRG (10 ns). TOP TRG can make a safe estimate of the time-of-flight based solely on the incident position. However, figuring out the point of incidence remains to be a formidable task.
3. The PDFs used for reconstruction using data from the TOP main readout incorporate both spatial and temporal information, but TOP TRG PDFs only use time information. Earlier simulations [55] demonstrated that utilizing time information alone can provide sufficient timing resolution. Including spatial information would require more trigger path bandwidth in TOP FEE and a larger FPGA fabric in TOP TRG.

For purposes of TOP TRG FW, more specifically, to identify the region of incidence with 2ns precision, the TOP bar is divided into 10 logical segments evenly. PDFs are generated for each segment either using GEANT-4 based Monte Carlo simulation or from real data after reconstruction. Fig. 91 shows the 10 PDFs generated for 10 segments evenly spaced along the bard. To facilitate the implementation in FPGA, these PDFs are transformed into log-likelihoods (i.e., χ^2 values) and scaled to integer numbers ranging from 0 and $2^{20} - 1$ (Fig. 92).

These 32-bin 20-bit PDFs are stored as Look-Up-Tables (LUT) on FPGA for quick access. PDFs are multiplied with the 32-bin timestamp histogram obtained from the upstream stream in the histogram module. The multiplications are performed in parallel for all 32 bins. Each multiplication utilizes a Xilinx multiplier IP core which is configured with 20-bit/7-bit inputs and 27-bit output. To meet timing constraints the latency is set to 2 clock

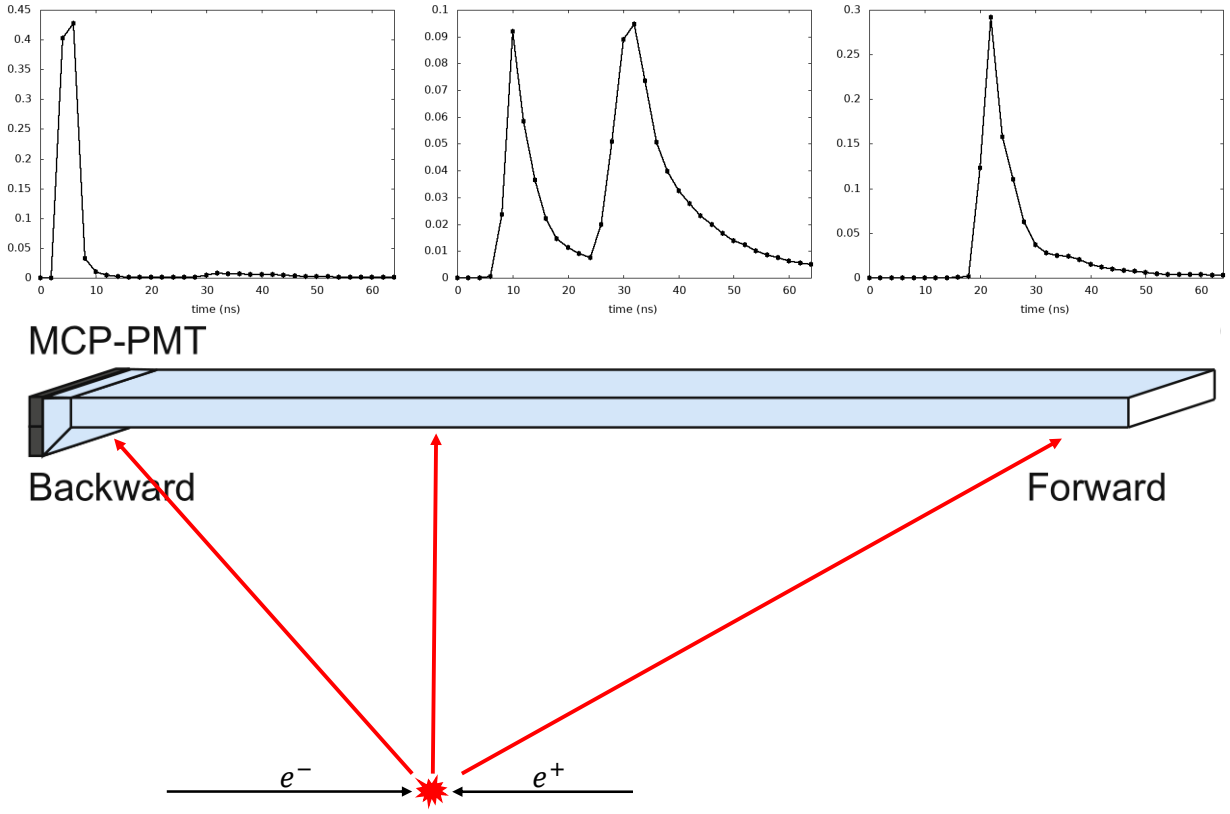


Figure 90: Three representative PDFs used in TOP TRG illustrating the scenarios where a 3 GeV pion is incident on the bar at three different locations.

cycles. The results of the 32 multiplications are then added together to form the overall log-likelihoods for each PDF hypothesis corresponding to 10 segments along the bar.

Next, the 10 log-likelihoods corresponding to the 10 PDFs are compared with each other. The PDF with the largest log-likelihood is selected as the best match, indicating the most probable incidence location and therefore the collision time. The estimated values for the collision time and hit position are then transferred to the next FW module for further analysis downstream.

5.3.2.5 Feature Extraction The feature extraction module is responsible for making a tentative slot-level t_0 decision by identifying the local maximum of the log-likelihood dis-

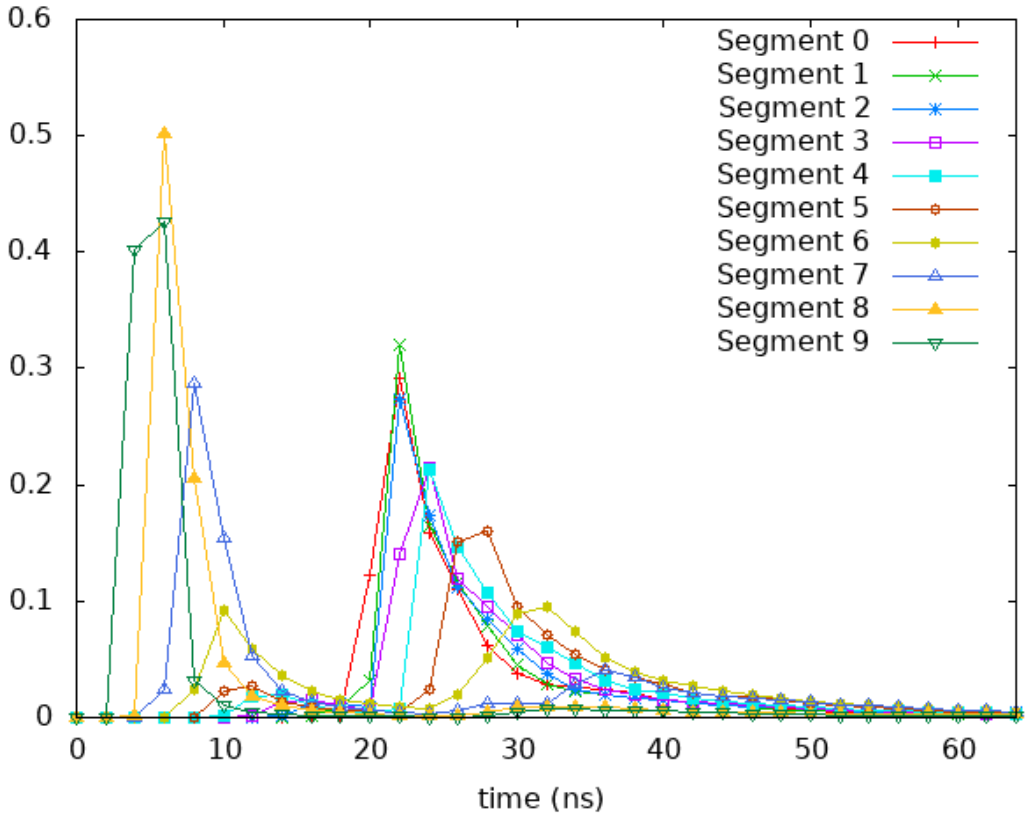


Figure 91: PDFs of TOP TRG for 10 segments.

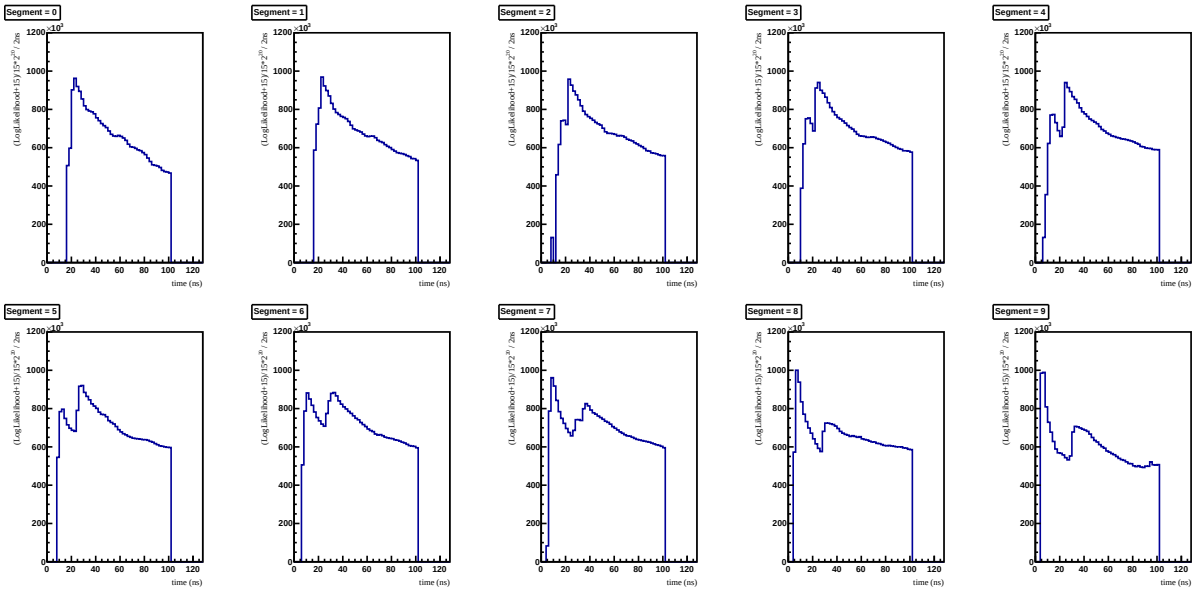


Figure 92: Log likelihood PDFs obtained from MC and used in TOP TRG timing algorithm.

tribution. The upstream PDF matching module calculates the log-likelihoods for the best-matched PDF and the corresponding t_0 decisions on every clock cycle. The slot-level timing decisions are represented by the clock cycles where log-likelihood achieves its maximum value. Fig. 93 illustrates the principle of the feature extraction algorithm implemented in TOP TRG. The feature extraction module identifies the local maximum by observing three consecutive rising likelihood values followed by one falling likelihood value. To filter out potential beam-related background, an adjustable threshold cut is applied to the likelihood values. Only likelihood values that exceed the threshold are considered in the feature extraction process.

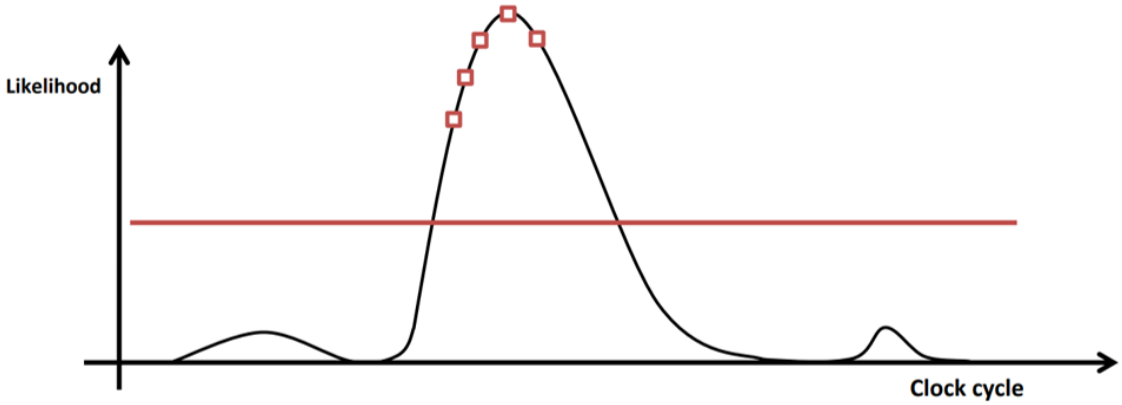


Figure 93: Principle of the feature extraction implemented in TOP TRG.

Fig. 94 shows the snapshots of the Xilinx ChipScope® traces for the real-time likelihood evaluation (likelihood versus clock cycle distribution) and subsequent feature extraction for slot 1 in a collision run. As Fig. 94 demonstrates, there is a trade-off between trigger efficiency and trigger rate. When using a low likelihood threshold cut, TOP TRG is overwhelmed by background-based decisions. Conversely, when the likelihood threshold cut is set too high, the physics events have a high probability of being cut off. As discussed later, to further improve the TOP TRG performance under conditions of high beam-induced background, CDC-TOP matching has been introduced into the TOP TRG system.

5.3.2.6 Combined Decision The combined t_0 decision module is responsible for combining slot-level decisions from all 16 slots and making combined decisions to be sent to the GDL. The combined t_0 decision module is implemented only on the TOP TRG RECEIVE

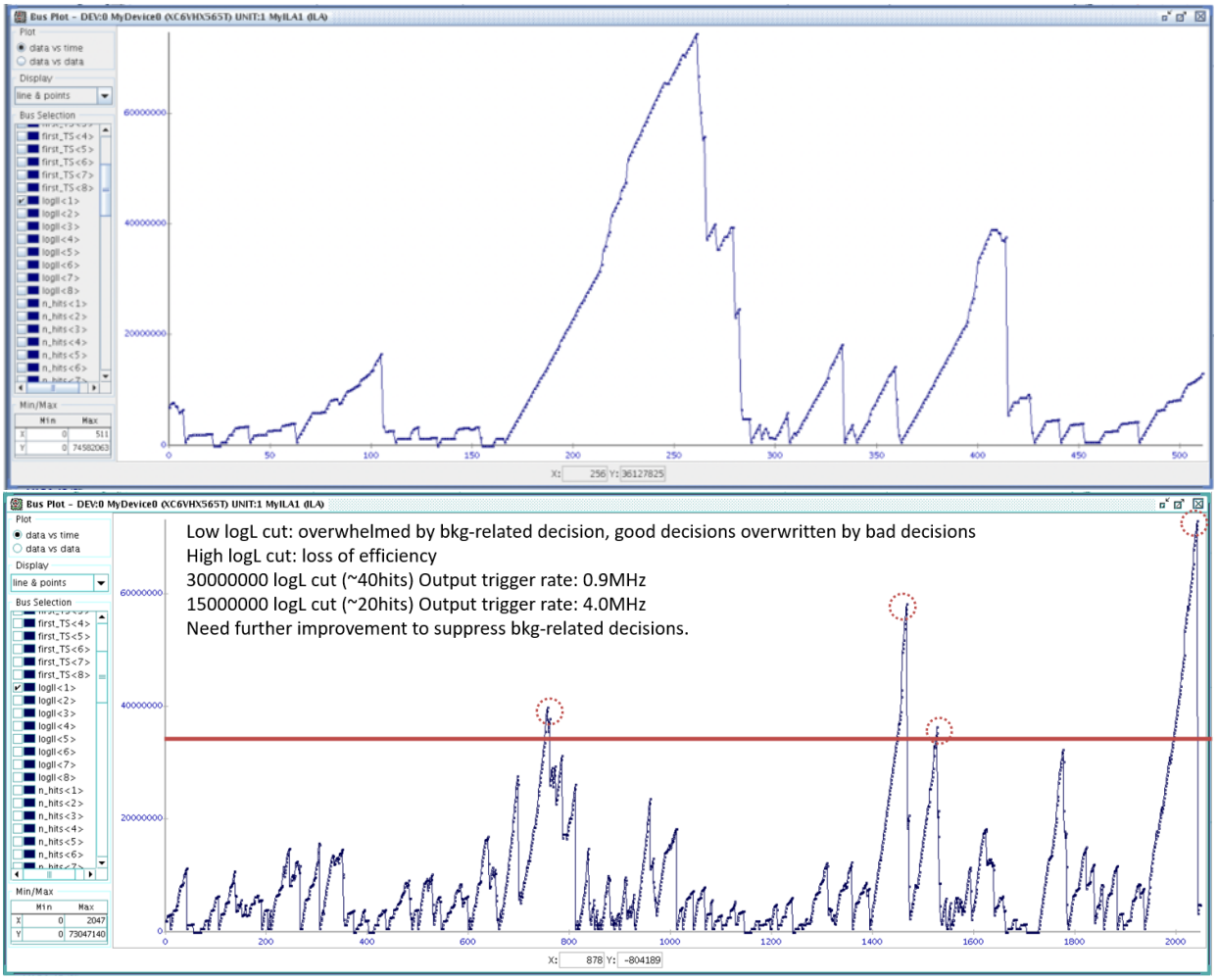


Figure 94: Snapshots of the ChipScope® traces representing likelihood evaluation and feature extraction in a collision run.

board. It collects the slot-level decisions for slots 1 through 8 from the TRANSMIT board and combines them with slot-level decisions for slots 9 through 16 prepared on the RECEIVE board. The slot-level decisions for slot 9 through 16 are delayed by 27 clock cycles (212.5 ns) to compensate for the data transmission latency between the two UT4 boards.

The latency of making a slot-level t_0 decision depends on the number of timestamps, i.e., the number of Cherenkov photons. Therefore, different slots make their t_0 decisions at different clock cycles. Once the combined t_0 decision module receives the first slot-level decision, it waits for other possible decisions from other slots. The waiting time is adjustable in order to adapt to different event types, luminosity, and background conditions. After

reaching the maximum waiting time, the module counts the number of slot-level decisions and adds all decisions. The results are fed into the division core to calculate the average of these slot-level decisions, which is the final combined t_0 decision.

Other than t_0 values associated with individual slots, several other pieces of slot-level information are also kept for further processing. Slot numbers and segment numbers of all voted slot-level decisions are used for back-to-back selection, which is useful for specific event types, *e.g.*, wide angle Bhabha scattering, and di-muon production. The number of voted slot-level decisions and the number of timestamps used in each slot-level decision provide valuable information about the multiplicity.

5.3.2.7 Aurora Link between UT4 Boards Aurora 8B/10B protocol is used for transmitting slot-level decisions for slots 1 through 8 from the TOP TRG TRANSMIT board to the TOP TRG RECEIVE board. Aurora 8B/10B IP cores on both sides utilize 4 bonded Xilinx UltraScale GTH transceivers. The line rate for the core is set to be 4x5.08 Gbps, supporting the 128-bit continuous stream of slot-level t_0 information at 127 MHz between two UT4 boards. The Aurora cores are configured as a duplex streaming interface. Aurora 8B/10B IP core is responsible for the lane initialization, comma alignment, clock compensation, error handling, and channel bonding.

5.3.2.8 Posting Combined Decisions to GDL The TOP TRG RECEIVE board posts combined t_0 decisions to GDL. For the data transmission between TOP TRG and GDL, a custom Aurora-like streaming protocol designed by NTUHEP group is used [56]. This protocol was originally designed to address the long latency issue in the CDC TRG, so its main advantage is the shorter latency compared to the Aurora protocol. Additionally, it supports line rates of up to 25 Gbps for GTY transceivers, while Aurora 8B/10B has a maximum line rate of 6.6 Gbps. The data link between TOP TRG and GDL is based on 4 GTH transceivers with 8B/10B encoding. The line rate is set at 4x5.08Gbps, supporting the 128-bit continuous stream at 127 MHz.

Due to the special mechanism of GDL selecting timing decisions from trigger subsystems (ECL TRG, CDC TRG, and TOP TRG), TOP TRG decision arrival time at the GDL is

expected to be 100% correlated with the TOP TRG timing value. As shown in Fig. 95, when the GDL logic signal is triggered, the GDL timing decision module only selects the TOP TRG t_0 decisions within ±28 clock cycles (of 127MHz clock) around the logic signal, which is equivalent to a time window of ±28*7.87 ns=220 ns. Decisions that arrive too early or too late are ignored. If more than one TOP TRG decision arrives within the ±28 clock cycles of the logic signal, only the earliest TOP TRG timing decision is used.

TOP TRG holds the decisions until a specific clock cycle that ensures a 100% correlation between the timing value and the time the decision is posted to the GDL. As there is no dedicated data pipe to store these decisions, if the trigger rate is high (more than 1 per 1280 clock cycles, i.e., above approx. 100kHz), the earlier TOP TRG decisions may be overwritten by the later decisions before being transmitted/posted to the GDL. Further consideration and optimization of the trigger rate are essential to avoid the loss of timing decision data due to such overwriting.

The information sent to GDL includes

1. combined t_0 decision (18 bits),
2. counter with respect to (FTSW) frame signal (11 bits),
3. flags for the slots voted for the combined t_0 (16 bits),
4. flag for back-to-back condition (1 bit),
5. segment number for each slot-level decision (4*16=64 bits).

5.3.2.9 Posting Slot-level Decisions to GRL Both TOP TRG TRANSMIT and RECEIVE boards post slot-level t_0 decisions and the numbers of timestamps used in each slot-level t_0 decision to GRL. As discussed more later, GRL performs CDC-TOP matching using the track information from CDC TRG and timing information from TOPTRG to filter out beam-related backgrounds.

Similarly to GDL, for the data transmission between TOP TRG and GRL, a custom protocol developed by the NTUHEP group is used. The data link is based on 4 GTH transceivers with 8B/10B encoding and the line rate is set at 4x5.08Gbps, supporting the 128-bit continuous stream at 127 MHz.

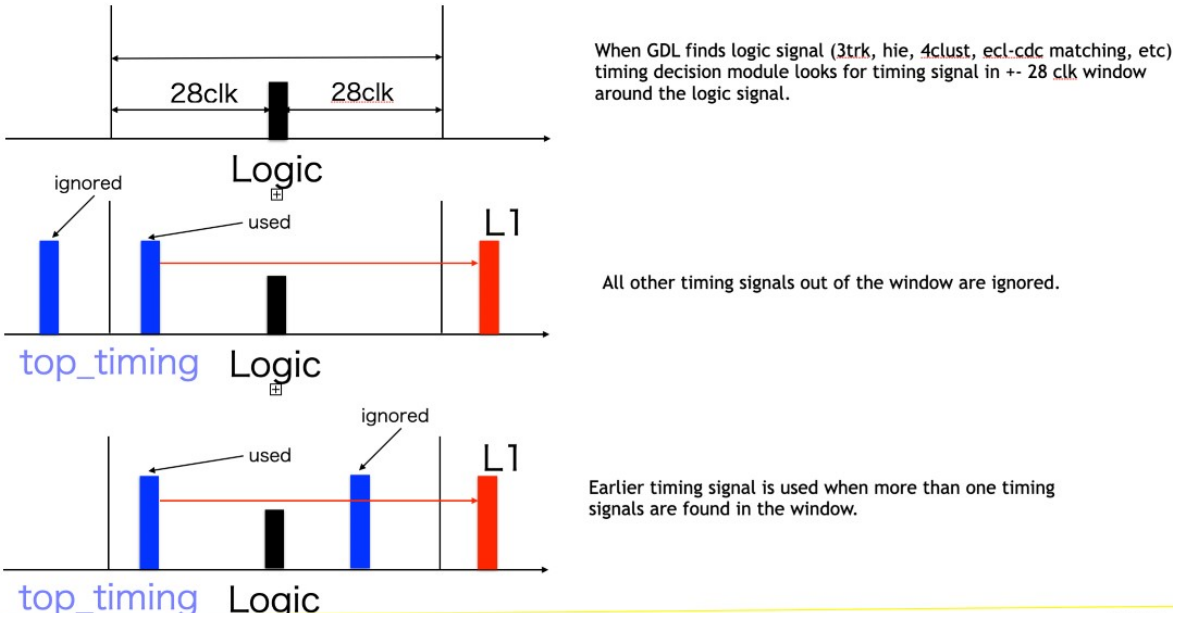


Figure 95: Mechanism of GDL selecting TOP TRG decisions.

Fig. 96 shows the data format used to send data from TOP TRG to GRL. These include a 1-bit valid flag, an 8-bit field used for the number of timestamps and 16-bit t_0 values for each slot-level decision. Due to limited bandwidth, TOP TRG sends the least significant 5 bits for each slot-level t_0 decision and only one field of the most significant 11 bits.

bit	0	1	2	3	4	5	6	7
[0:7]	Flag slot 1-8							
[8:15]	Hit numbers slot 1							
[16:23]	Hit numbers slot 2							
[24:31]	Hit numbers slot 3							
[32:39]	Hit numbers slot 4							
[40:47]	Hit numbers slot 5							
[48:55]	Hit numbers slot 6							
[56:63]	Hit numbers slot 7							

bit	0	1	2	3	4	5	6	7
[64:71]	Hit numbers slot 8							
[72:79]	MSB 11 bits of slot T0							
[80:87]				T0 slot 1-8				
[88:95]								
[96:103]								
[104:111]								
[112:119]								
[120:127]				Spare 5bits				

Figure 96: Data format used to send data from TOP TRG to GRL.

5.3.2.10 Belle II Trigger Timing Interface Belle II Trigger Timing (b2tt) is a Trigger and Timing distribution (TTD) and control protocol for Belle II. This protocol supports bidirectional communication based on a 254 Mbps serial data link [57]. The b2tt interface in TOP TRG FW is responsible for receiving clock and trigger signals from upstream Front-end Timing Switch (FTSW) VME module and returning status acknowledgment. FTSW and

TOP TRG boards are connected via a CAT-7 cable, where 4 shielded twisted copper pairs are used for the clock, distribution, collection, and modulated TDO signals, respectively.

One of the key b2tt functions is distributing the L1 trigger signal generated by GDL to all front-end electronics including the TOP TRG system. Once the TOP TRG receives the GDL L1 trigger, the corresponding event data are sent to DAQ via the belle2link protocol. Fig. 97 shows the schematic of the trigger and data paths of the DAQ system.

The upstream FTSW receives the clock signal from the clock master board. This clock signal is generated from a 127.216 MHz system clock, which itself is derived from a 509 MHz RF signal originating from the SuperKEKB accelerator. The FTSW then transmits this 127.216 MHz clock to the TOP TRG firmware, where it serves as the system clock for the core logic of the TOP TRG modules.

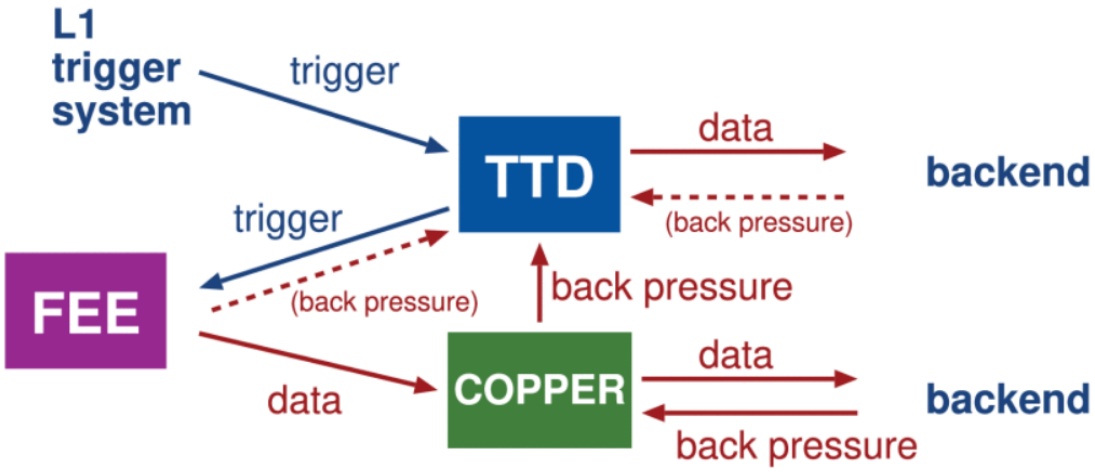


Figure 97: Trigger and data paths of the data acquisition system. COPPER-based DAQ has been recently replaced by the PCIe40-based DAQ system.

5.3.2.11 Belle2link Readout The Belle2link readout module on TOP TRG consists of three major parts:

1. b2tt module receives GDL L1 signals which trigger the readout process,
2. Belle2link protocol is used to transfer event data to the DAQ,
3. circular buffer is used to store the event data from the past few μs.

The basic principle of Belle2link readout is that the event data corresponding to the specific GDL L1 trigger can be retrieved within a fixed look-back time interval around the

L1 signal of 5 μs latency, therefore facilitating the readout of the data relevant for the time of the actual physics event. The circular buffer on the TOP TRG has a depth of 512 windows, and a new window of 1024-bit event data is stored in the buffer every 31.5 ns (32 MHz). This corresponds to, approximately, 16 μs time interval, which is longer than the GDL L1 latency. When the GDL L1 signal arrives, a continuous part of the buffer containing 96 windows is selected, which corresponds to 3 μs time interval centered around the event, according to the L1 signal. The data within this 3 μs time interval are sent to the Belle2link transmitter for readout and subsequent transfer to the DAQ system. The location of the 96 windows inside the 512-window circular buffer is selected by a controllable parameter.

Belle2link is a unified data transfer protocol used to transfer data from the FEE to the DAQ system using optical fibers. Belle2link is a bi-directional custom high-speed serial link protocol. In the TOP TRG system, the Belle2link protocol is implemented using a GTH transceiver with 8B/10B encoding, and the line rate is set at 2.54 Gbps. Belle2link attaches 6 words as the header and 3 words as the trailer to the raw data for verification and monitoring purposes. The data format of a Belle2link package is shown in Fig. 98, the length of the FEE data part is 32x32=1024 bits.

The Belle2link in the TOP TRG system is primarily used for trigger waveform readout. Since the belle2link data clock operates at 32 MHz, while the timestamp flow clock operates at 127 MHz, 4 successive timestamps are packed together at 127 MHz and sent to Belle2link at 32 MHz. Consequently, the Belle2link effectively reads out 384 successive timestamps within the 96 windows. The readout waveforms are utilized in offline FW simulations to compare the online and offline performance, facilitating further FW optimization. In addition to the waveform data, the Belle2link data package also includes slot-level t_0 decisions, the number of timestamps, segment numbers, and other firmware information which is used for Data Quality Monitoring (DQM) purposes.

5.3.2.12 VME Slow Control The VME module integrated into the TOP TRG system serves the purpose of slow control and monitoring. It communicates with VME single board computer (SBC) via the VMEbus interface. The C language programs developed for VME SBC (Linux OS) communicate with TOP TRG firmware using the driver controlling

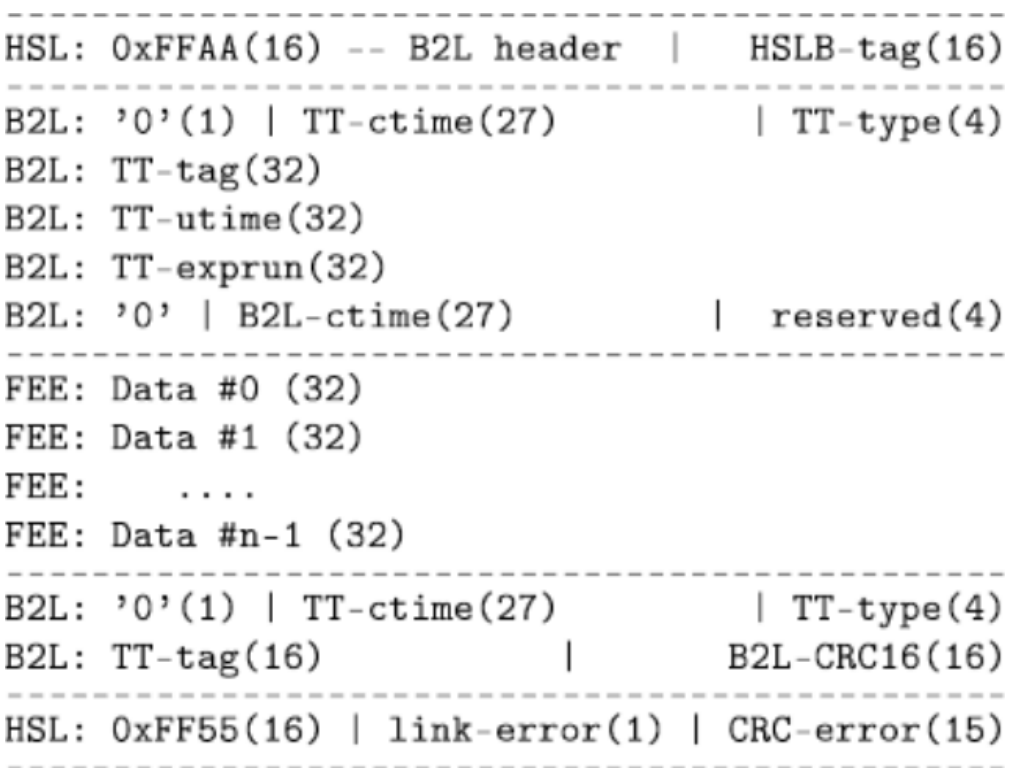


Figure 98: Data format of Belle2link package.

the VME interface chip, TUNDRA Universe-II. The VME SBC can read the data from the TOP TRG system, enabling real-time monitoring and analysis of the performance. Such data include counters for the numbers of timestamps, slot-level t_0 decisions, combined decisions, and link downtimes within a one-second interval. Additionally, the VME SBC is able to configure various parameters of the TOP TRG system during runtime, which includes adjusting parameters such as the likelihood threshold and maximum wait time, as well as downloading the firmware as needed. It allows for the fine-tuning of the TOP TRG system's operation for specific run conditions.

5.3.2.13 Resource Utilization Fig. 99 shows the FPGA resource utilization tables on the TOP TRG TRANSMIT and RECEIVE boards.

The PDF matching module in the TOP TRG system is the most resource-intensive part of the design. Matching with each PDF requires 32 20x7-bit multipliers, which consume approximately 3,000 Look-Up Tables (LUTs). Each TOP slot utilizes 30,000 LUTs to match

with a total of 10 PDFs. Considering that a single board is responsible for 8 TOP slots, each UT4 it utilizes around 240,000 LUTs.

There are still plenty of unused resources on the UT4 boards, which allows for further expansion and improvement. With the available resources, it is possible to match more PDFs in parallel, which would lead to a better timing resolution.

Resource	Utilization	Available	Utilization %
LUT	260389	1074240	24.24
LUTRAM	5285	231840	2.28
FF	398678	2148480	18.56
BRAM	236.50	3780	6.26
IO	178	702	25.36
GT	45	76	59.21
BUFG	34	1800	1.89
MMCM	3	30	10.00

Resource	Utilization	Available	Utilization %
LUT	264396	1074240	24.61
LUTRAM	6679	231840	2.88
FF	405049	2148480	18.85
BRAM	243	3780	6.43
IO	178	702	25.36
GT	45	76	59.21
BUFG	34	1800	1.89
MMCM	3	30	10.00

Figure 99: FPGA resource utilization on the TOP TRG TRANSMIT (top) and RECEIVE (bottom) boards.

5.3.3 TOP TRG Performance Studies

In this section, we analyze the TOP TRG performance for both cosmic and hadronic events. In order to perform TOP TRG performance studies, several tools have been developed and integrated into the Belle II Analysis Software Framework (basf2), which include (1) visualization and analysis tools for TOP TRG waveform readout and TOP main readout and (2) offline FW simulation based on Xilinx ISim simulator using TOP TRG waveform

readout and TOP main readout data.

5.3.3.1 Visualization of TOP TRG Timestamps and TOP Main Readout Hits

We developed the visualization tool which allows us to compare the data from the TOP TRG waveform readout with the data from the TOP main readout. This tool assists in optimizing the TOP TRG algorithm and helps detect any corruption or synchronization issues in the trigger path. The TOP hits from the TOP main readout are quantized to a precision of 2 ns for comparison with the trigger timestamp and for offline FW simulation. TOP TRG timestamps are obtained from the TOP TRG waveform readout data.

5.3.3.2 Offline Firmware Simulation Offline firmware simulation tools based on Xilinx ISE Simulator (ISim) are used to evaluate the algorithm performance under different run conditions (i.e., cosmic and collision data). Register Transfer Level (RTL) design of the sliding window slot-level algorithm, including Stream to Histogram, PDF matching and feature extraction modules (Fig. 84), is simulated using Xilinx ISim. Both the TOP TRG waveform and TOP main readout hits can be utilized in the simulation. The firmware simulation module is fully integrated into the basf2 framework. The basf2 unpacker module prepares the TOP TRG timestamps and TOP main readout hits from the raw data for each event. The event data, in a specific format, is then provided to the FW test bench, which is used to run the ISim simulation. The output results are subsequently retrieved by basf2 for detailed investigations. The simulated TOP TRG t_0 decisions are compared with the fully reconstructed event t_0 to evaluate the performance. This automated process enables the fast and comprehensive analysis of the TOP TRG system performance.

5.3.3.3 TOP TRG Performance for Cosmic Events Fig. 100 shows a cosmic event from exp 24 run 1513. The plot on the left shows an EventDisplay snapshot of a cosmic event including CDC TRG 2D tracks and TOP main readout hits. On the right side, the blue plot represents the timestamp values versus the timestamp arrival time at UT3 boards on slot 14. The red plot corresponds to the TOP main readout hits from slot 14, quantized to a precision of 2 ns and aligned with the TOP TRG timestamps relative to the GDL L1

signals. The bottom right plot provides a zoomed-in view, showing the agreement between the TOP TRG timestamps and the TOP main readout hits but with additional delays and large gaps.

In addition to the timestamps for the (presumed) muon from a cosmic event, there are more trigger timestamps than the TOP main readout hits. This is because TOP FEE feature extraction filters out electronic noise and abnormal pulses, preventing them from being interpreted as hits. However, such filtering is not possible for trigger timestamps, which represent all incoming signals, including electronic noise.

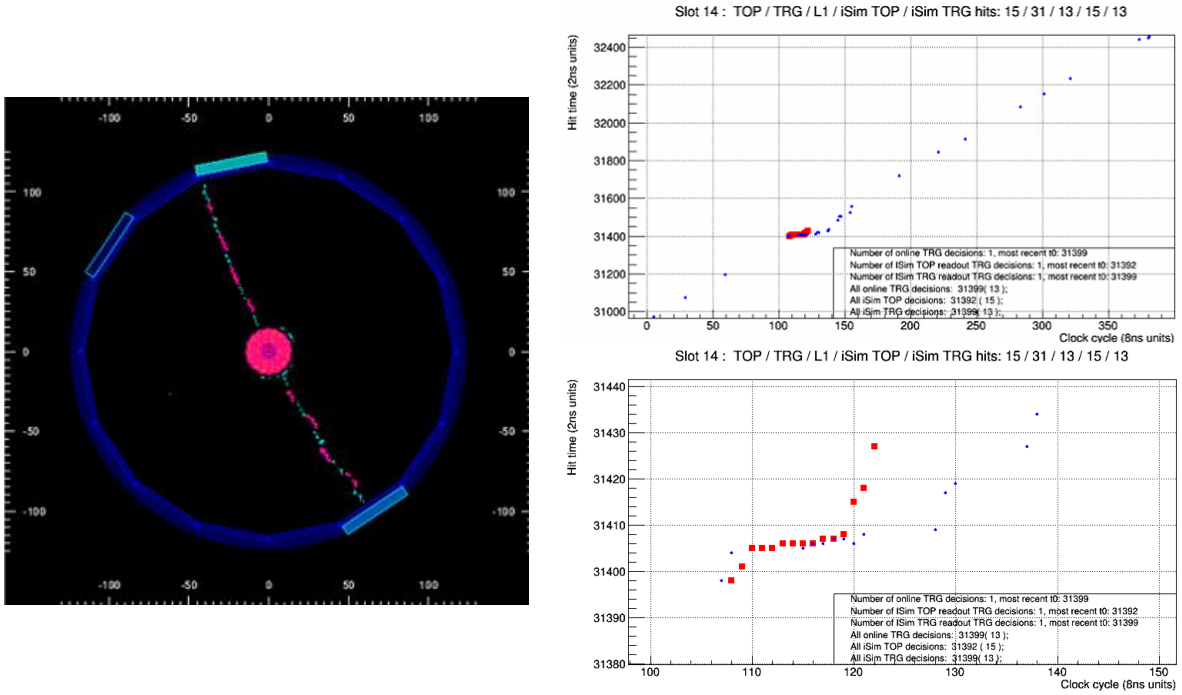


Figure 100: A cosmic event from exp 24 run 1513. The plot on the left shows the Event-Display snapshot of a cosmic event including CDC TRG 2D tracks and TOP main readout hits. Plot on the top right shows the comparison between TOP main readout hits quantized to 2 ns (in red) and TOP TRG timestamps (in blue) for slot 14 of this event. The plot on the bottom right shows the zoomed-in version.

For cosmic events, the background primarily consists of electronic noise, and the level of background noise is tolerable. Consequently, the TOP TRG t_0 decisions simulated using the TOP main readout hits and TOP TRG waveforms closely agree with each other, and are also close to online decisions. The "All online TRG decision" in Fig.100 corresponds to the online t_0 decision, while "All iSim TOP decisions" and "All iSim TRG decisions" represent the t_0

decisions simulated using TOP main readout hits and TOP TRG timestamps, respectively.

TOP TRG t_0 decisions based on electronic noise in cosmic events are rare. Although there is some efficiency loss due to electronic noise, the impact is minimal. Fig.101 shows the online TOP TRG t_0 decisions and TOP TRG t_0 simulated using TOP main readout hits, with respect to ECL t_0 for cosmic events of exp 24 run 1513. The online TOP TRG efficiency is approximately 92.5%, while the efficiency simulated using TOP main readout hits is around 96.3%. The small increase in efficiency can be attributed to the filtering of electronic noise in the TOP main readout, and, more importantly, to pairing of channels for TOP TRG purposes in TOP FEE. Such pairing, for high-quality electric pulses from actual photons, results in the number of timestamps always being smaller than the number of hits in the TOP main readout where such pairing is not implemented. As a result, the efficiency of the actual TOP TRG is lower than that estimated using hits from the TOP main readout. Efficiency is defined as the fraction of events in which the TOP TRG t_0 is within $\pm$100 ns with respect to the ECL t_0.

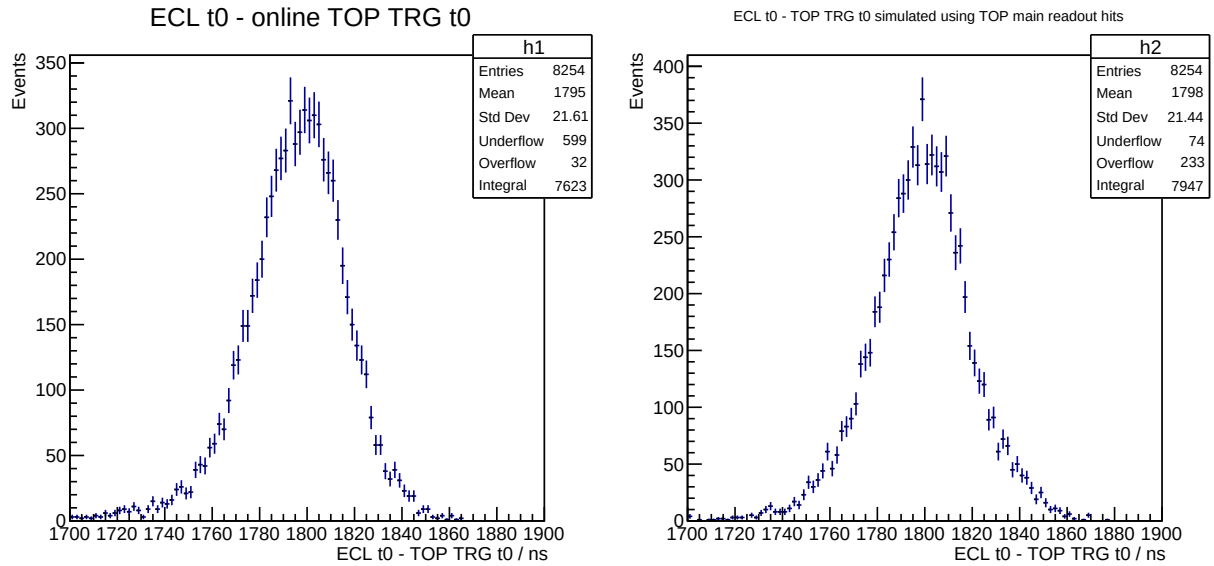


Figure 101: Cosmic events in exp 24 run 1513: online TOP TRG t_0 decisions and TOP TRG t_0 decisions simulated using TOP main readout hits, with respect to ECL t_0.

5.3.3.4 TOP TRG Performance for Hadronic Events Fig. 102 shows a hadronic event from exp 24 run 1551. The plot on the left shows the EventDisplay snapshot of an

average hadronic event. The plots on the right side show the timestamps/hits for slot 15 and 16. The trigger timestamps are shown with blue points and the TOP main readout hits are shown with red points.

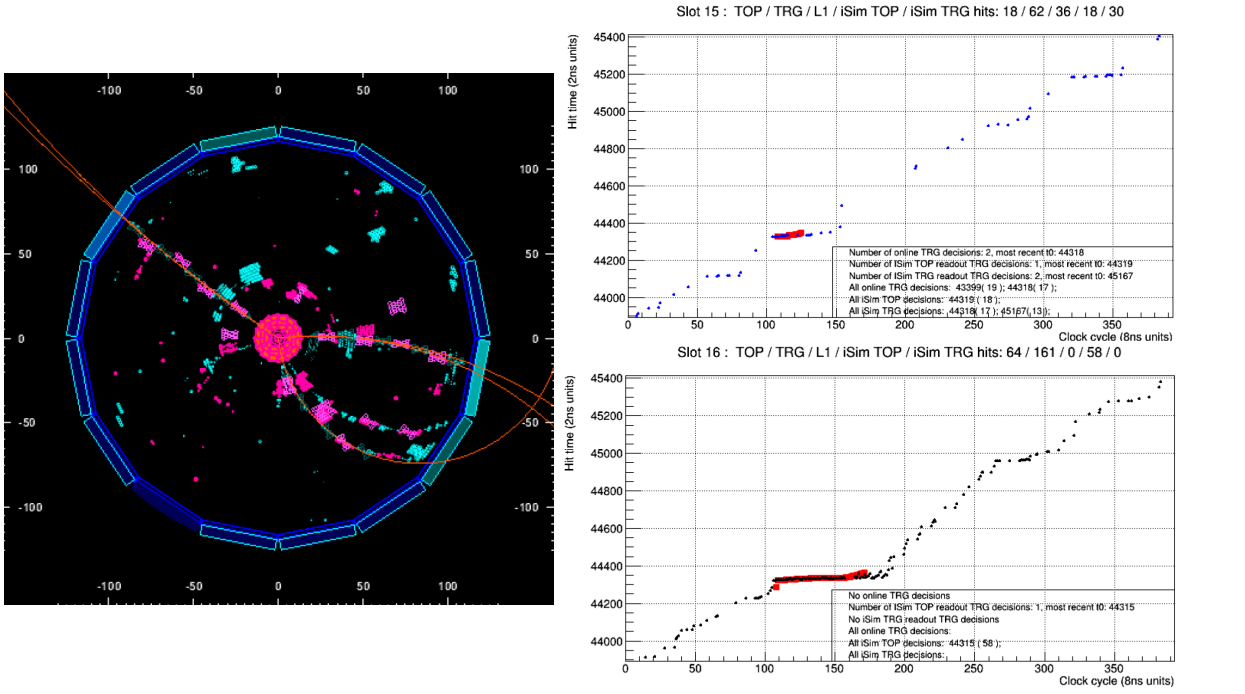


Figure 102: A hadronic event from exp 24 run 1551. The plot on the left shows the EventDisplay snapshot for this hadronic event. The plots on the right side show the timestamps/hits for slot 15 and 16, where trigger timestamps are shown with blue points and the TOP main readout hits are shown with red points.

In hadronic events, the presence of beam-related background photons poses a significant challenge for the TOP TRG system. Unlike electronic noise, which is not a concern for hadronic events, the beam-related background photons severely contaminate the trigger timestamp stream and affect the accuracy of the TOP TRG decisions. As shown in Fig. 102, a large number of trigger timestamps due to the beam-related background are observed, contaminating the trigger timestamps corresponding to the actual hadronic events.

The contamination from the beam-related background photons results in the overwriting or blocking of the decisions for actual hadronic events by decisions associated with the overwhelming background. While the TOP main readout has a very narrow digitization window of 200 ns with respect to GDL L1 signals, which helps reject the beam-related background, the TOP TRG is exceptionally sensitive to such background due to the absence

of a similar digitization window (which is impossible to implement, as TOP TRG needs to make its timing decision before the L1 strobe is generated).

Fig. 103 shows the TOP TRG t_0 decisions simulated using both TOP TRG waveform and TOP main readout hits, with respect to ECL t_0 for the hadron events of exp 24 run 2088. The simulation results show that the TOP TRG efficiency, when using the waveform, is approximately 34.0%, while the efficiency using the main readout hits is around 38.8%. The low TOP TRG efficiency when using the waveform, as we already discussed, is attributed to the overwhelming beam-related background due to the absence of a digitization window as in the TOP main readout. However, the low TOP TRG efficiency simulated using the main readout hits is due to the contamination of the combined TOP TRG decisions by background decisions from slots dominated by background hits. Unlike the TOP main readout data, which utilizes the CDC track information during reconstruction, the TOP TRG lacks this information from CDC TRG, making it more difficult to mitigate the impact of background decisions.

As the level of beam-related background increases, making accurate TOP TRG t_0 decisions in collision events becomes significantly more challenging. Therefore, it becomes necessary to correlate TOP TRG information with CDC 2D (or even 3D) TRG tracks or using ECL-based t_0 as a seed (so TOP TRG could be used for a more precise T0 decision). This approach will be discussed in the subsequent section.

5.3.4 CDC-TOP TRG-level Matching

As discussed in the previous section, TOP TRG efficiency reaches 92.5% for cosmic events, while only 34.0% for hadronic events. It has been realized that standalone TOP-based trigger is impossible due to excessively high beam-related background. Fig. 104 compares the TOP TRG timestamps (in blue) with TOP main readout hits (in red) for slot 1, 2, 5 and 15 of a hadronic event in exp 26 run 766. TOP TRG path is almost 100% saturated with timestamps, which are primarily due to the high beam-related background photons. These beam-related background timestamps affect TOP TRG decisions in two ways:

1. signal-based decisions are overwritten by background-based decisions made for the same

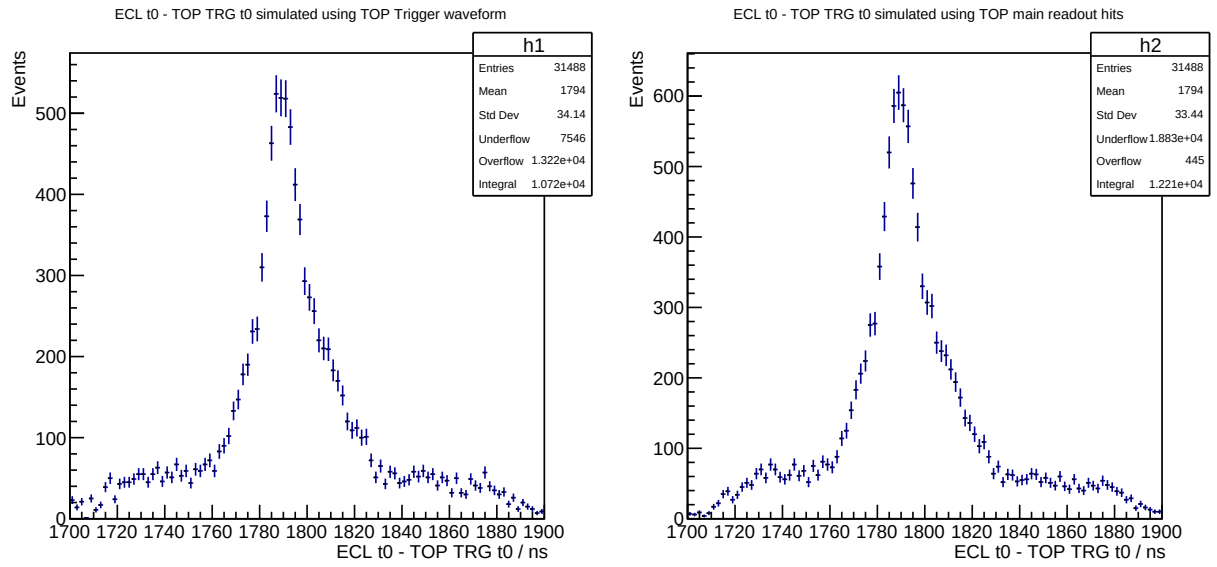


Figure 103: Hadronic events in exp 24 run 2088: TOP TRG t_0 decisions simulated using TOP TRG waveform and TOP main readout hits, with respect to ECL t_0.

slot,

2. combined decisions are contaminated by background-based decisions from the slots which are dominated by background photons.

TOP main readout solves the first problem using a very narrow digitization window of 200 ns with respect to GDL L1 decisions. The role of the TOP digitization window is demonstrated in Fig. 105. Almost every horizontal sequence of timestamps in this plot represents a burst of photons from beam-related background. These photons arrive at PMT in the same narrow interval of time. They differ from collision-related photons primarily by the burst's photon multiplicity. However, TOP TRG currently makes its timing decisions completely independently of ECL and CDC. Therefore, in contrast to the TOP main readout, TOP TRG is unable to reject background timestamps using a digitization/sampling window with respect to the actual L1 decision. The TOP main readout solves the second problem using CDC tracks. In the offline reconstruction, fully reconstructed tracks are extrapolated to TOP bars and only TOP hits near the points of incidence are used in the TOP reconstruction for PID purposes.

Inspired by the architecture of TOP main readout, we proposed two solutions for TOP TRG:

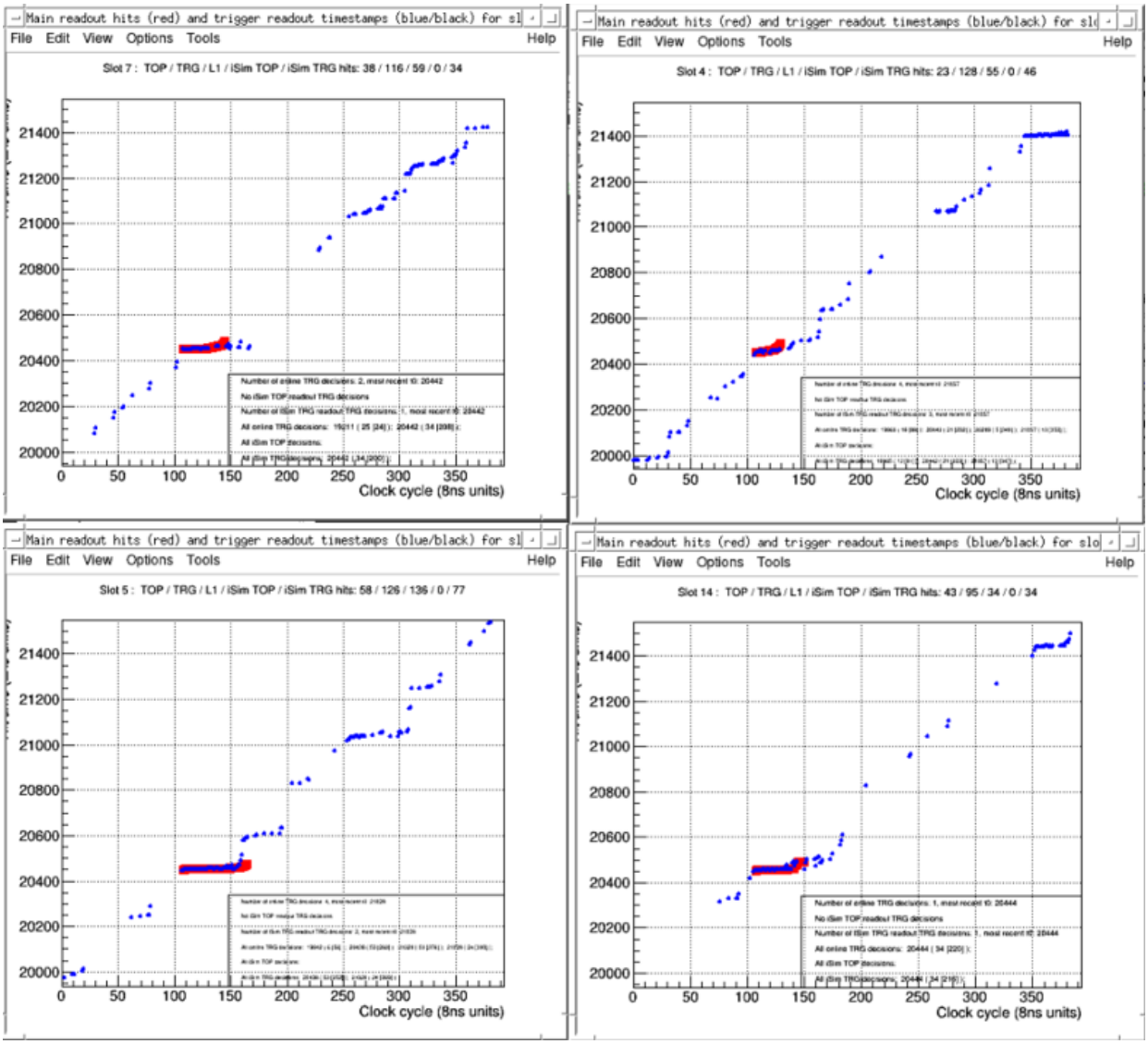


Figure 104: A hadronic event in exp 26 run 766: TOP TRG timestamps (in blue) and TOP main readout hits (in red) for slot 1, 2, 5, and 15.

1. consider only the slots matched to CDC TRG 2D tracks, and
2. use a narrow timing window (500-800 ns) around CDC TRG 2D to reject background-based decisions.

Fig. 106 demonstrates the concept of the CDC-TOP TRG-level match window. It corresponds to the GRL-CDC bit used for matching in GRL, which could be raised for a specific number of clock cycles when there is any CDC TRG 2D track pointing to the corresponding TOP slot.

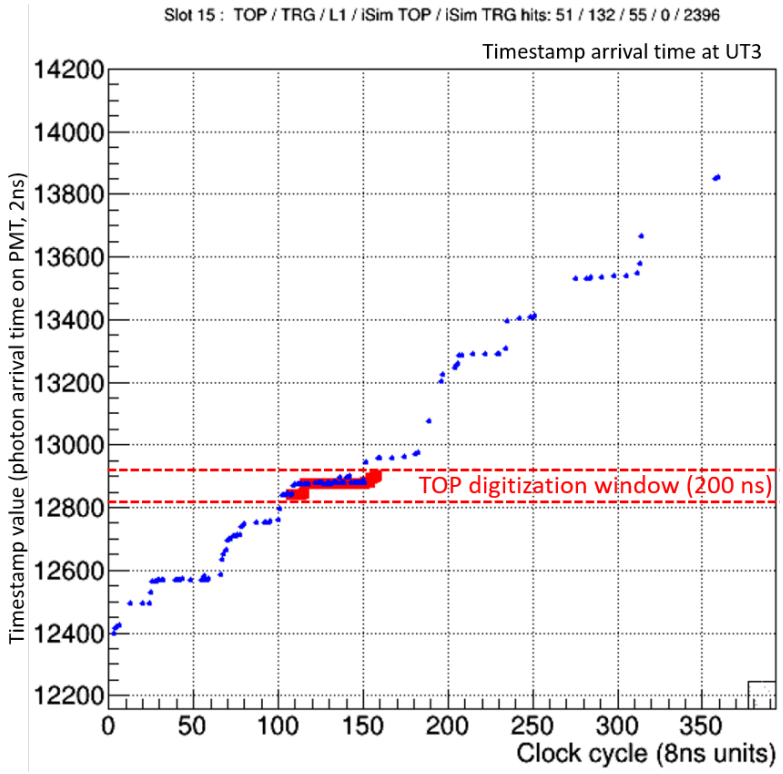


Figure 105: The explanation of the role of the digitization window in TOP main readout.

5.3.4.1 Performance Studies for Several CDC-TOP Matching Algorithms In order to determine the best-performing CDC-TOP matching algorithm, we performed offline simulations using several slot-level and combined algorithms. The following are the algorithms that we evaluated:

1. The slot-level decision is based on the maximum log-likelihood in the CDC-TOP matching window.

   The combined decision is the average of the slot-level decisions with CDC TRG 2D track matching.

   The combined decision generated by this algorithm is the one that is sent to GDL.
2. The slot-level decision is based on the maximum log-likelihood in the CDC-TOP matching window.

   The combined decision is the slot-level decision with the most hits, with CDC TRG 2D track matching.

   The combined decision generated by this algorithm is the one that is sent to GDL.

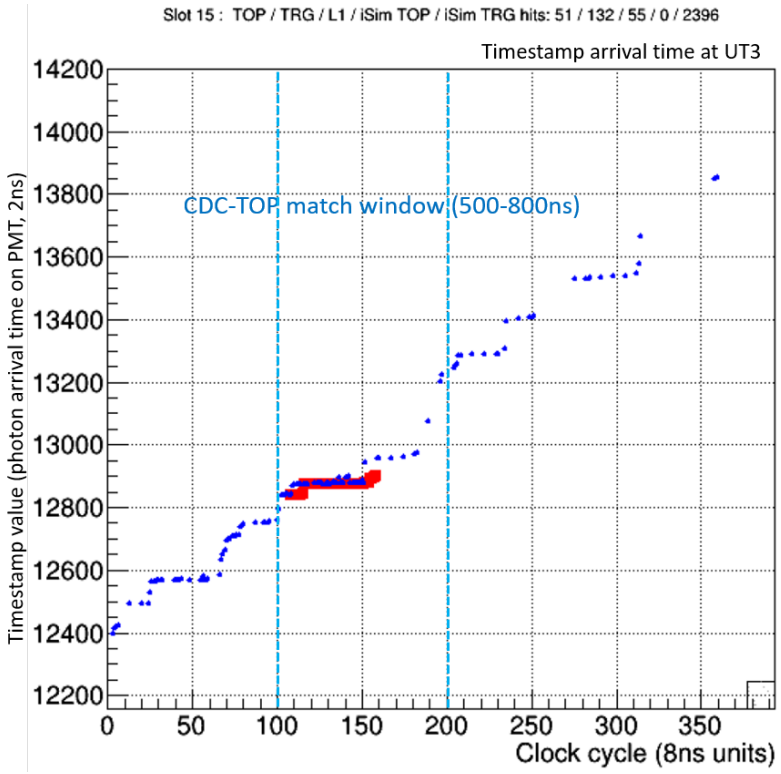


Figure 106: The explanation of the role of the proposed CDC-TOP matching window in TOP TRG.

3. The slot-level decision is generated on every clock cycle.

   The combined decision is the average of the slot-level decisions for the same clock cycle, with CDC TRG 2D track matching.

   The combined decision that has the largest number of slot-level decisions is the one that is sent to GDL.

4. The slot-level decision is generated on every clock cycle.

   The combined decision is the slot-level decision with the most hits for the same clock cycle, with CDC TRG 2D track matching.

   The combined decision that has the largest number of slot-level decisions is the one that is sent to GDL.

Fig. 107 shows the distributions of TOP TRG $t0$ w.r.t ECL t_0 for these four algorithms, using TOP TRG FW simulations using TOP TRG waveforms recorded for hadronic events in exp 26 run 766.

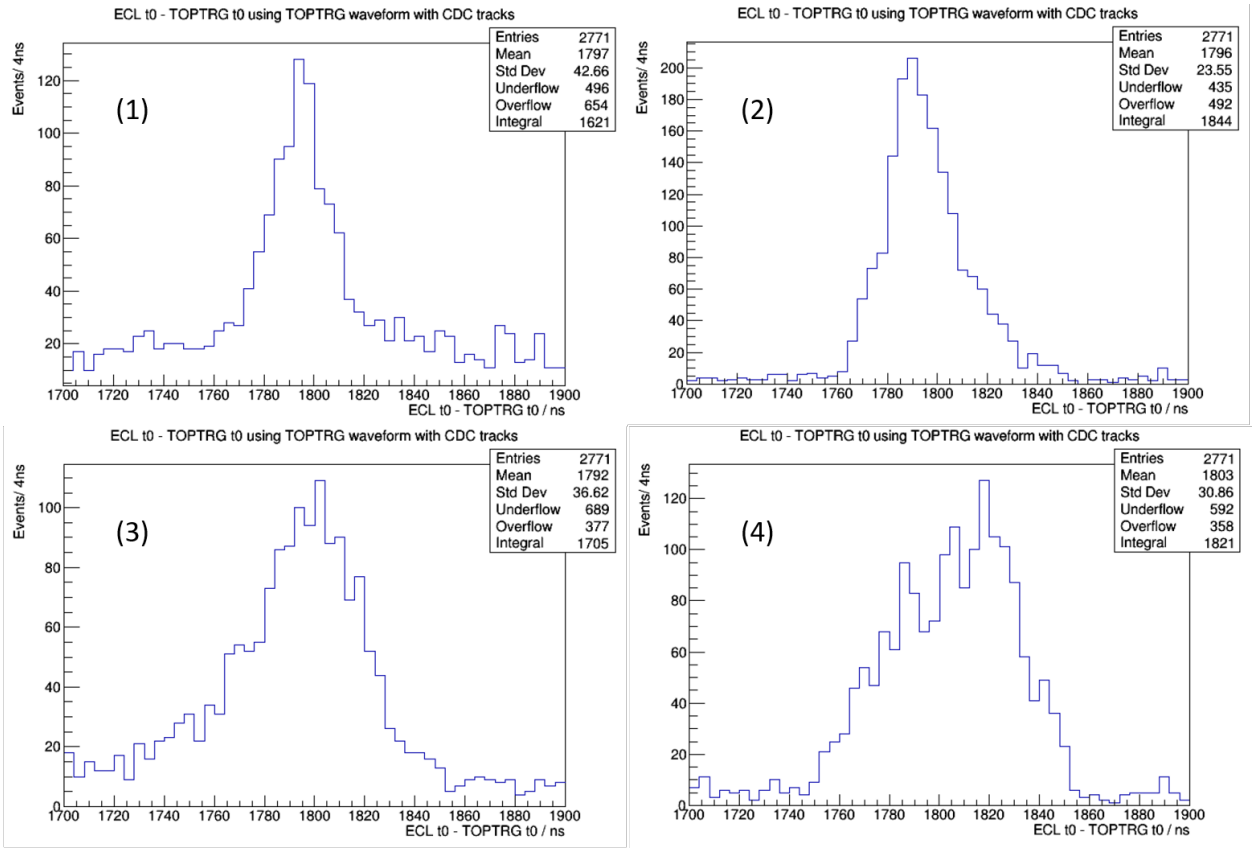


Figure 107: TOP TRG t_0 w.r.t ECL t_0 distributions for four algorithms, simulated using TOP TRG waveforms for hadronic events in exp 26 run 766.

The first algorithm, which utilizes the average of all slot decisions with CDC TRG 2D track matching as a combined decision, achieves an efficiency of 58.5%. The low efficiency and resolution are attributed to the contamination of background-based slot decisions.

The second algorithm, which selects the slot decision with the highest number of timestamps with CDC TRG 2D track matching as the combined decision, achieves an efficiency of 66.5% and an improved resolution of approximately 15 ns. The resolution and efficiency improve significantly compared to the first algorithm, which suggests that using the slot decision with the most timestamps instead of averaging over all slot-level decisions helps to mitigate contamination and suppresses the impact of beam-related background.

The third and fourth algorithms rely on the high particle multiplicity of hadronic events compared to the beam-related background. The third algorithm, which averages all slot decisions made on the same clock cycle, exhibits poor resolution and a low efficiency of

61.5%, similar to the first algorithm. The fourth algorithm, which selects the slot decision with the most hits among those made on the same clock cycle, achieves an efficiency of 65.7%. However, its timing resolution is worse than that of the second algorithm.

Overall, the second algorithm demonstrates the best performance, with an efficiency of approximately 67% for hadronic events and an ECL-TOP timing resolution of approximately 15 ns.

5.3.4.2 Performance for Hadronic Events of High Charged-particle Multiplicity

Fig. 108 shows the TOP TRG t_0 w.r.t ECL t_0 distributions for the second CDC-TOP matching algorithm for events of different charged-particle multiplicities, simulated using the TOP TRG waveform of hadronic events in exp 26 run 766.

By requiring at least one TOP slot with CDC TRG 2D tracks pointing to, the TOP TRG efficiency improves from 66.5% to 70.1%. This is because approximately 5% of hadronic events do not have CDC TRG 2D tracks. The TOP TRG efficiency reaches 73.0% when requiring at least two TOP slots with CDC TRG 2D tracks pointing to. For hadronic events with a higher charged-particle multiplicity, as estimated using CDC TRG information, *i.e.* with at least three TOP slots matching with projections of CDC TRG 2D tracks, the TOP efficiency surpasses 75% and the timing resolution becomes approximately 12 ns.

5.3.4.3 CDC-TOP TRG-level Match Summary

Offline simulation demonstrates that the second proposed CDC-TOP matching algorithm, which uses slot-level t_0 decision with the most hits and matching with CDC TRG 2D tracks, exhibits the best performance. The efficiency for hadronic events is 66.5% and the timing resolution is 15 ns.[2] For hadronic events of high charged-particle multiplicity, the efficiency exceeds 75% and (the combined ECL and TOP) timing resolution is 12 ns.

5.3.5 Summary, Challenges and Future work

In this section, we introduce the working principles and provide a detailed description of the FW design of the TOP TRG system. The standalone TOP TRG achieves the efficiency

[2] Note that estimating the t_0 resolution of TOP TRG requires removing the ECL TRG t_0 resolution.

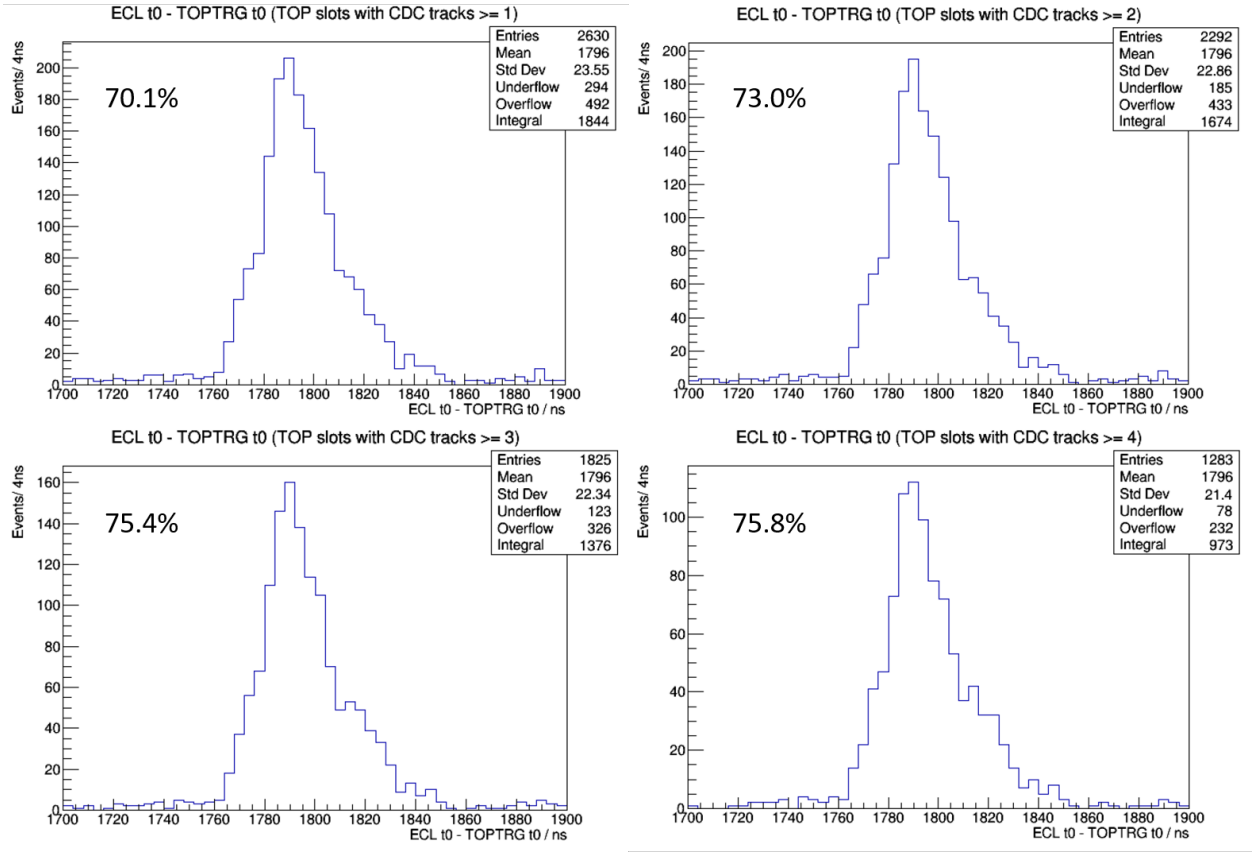


Figure 108: TOP TRG t_0 w.r.t ECL t_0 distributions for the second CDC-TOP matching algorithm for events of different charged-particle multiplicities, simulated using TOP TRG waveforms for hadronic events in exp 26 run 766.

of 92.5% for barrel cosmic muons but only 34.0% for hadronic events during collision. The low efficiency is attributed to the excessively high rate of beam-related background photons which illuminate the entire detector. To address this challenge, a study on CDC-TOP TRG-level matching has been performed to correlate the TOP TRG information with CDC TRG 2D track information. Offline simulation shows that by incorporating CDC-TOP matching, an efficiency of 75% and a timing resolution of 12 ns can be achieved for hadronic events.

Further developments will be carried out for the TOP TRG system to achieve better performance. Larger resources available on the UT4 board allow us to expand the width of the sliding window (current depth 64 ns) in the Stream to histogram module, which should mitigate yet another problem due to unsorted trigger hits arriving from different SCRODs of the same slot. Larger resources also allow for the matching of more PDFs with the timestamp

histogram in parallel, which should allow to improve timing resolution. Additionally, our current sliding window algorithm could not smoothly slide through time with 2 ns timestamps (which would require the firmware to operate at 508 MHz), therefore, the algorithm has to include the ability to quickly advance ("jump") in time when it is unable to keep up with more recent data. This shortcoming of the current implementation also affects timing resolution and must be improved. Streaming trigger data from the SCROD to TOP TRG has been successfully tested to operate correctly at the line rate of 10.16 Gbps. We are exploring to use Machine Learning algorithms to suppress beam-related background photons more efficiently.

APPENDIX A

DETECTOR TECHNOLOGIES

In this Appendix, at the conceptual level, I briefly discuss several detector technologies, some of which are used in the Belle and Belle II detectors.

A.1 PHOTON DETECTORS

Photon detectors are designed to detect the (usually very small) number of incident photons. Generally, photon detectors generate a detectable electrical signal proportional to the number of incident photons. This process involves three steps:

1. Generation of a primary photoelectron or electron-hole (e-h) pair by an incident photon by the photoelectric or photoconductive effect.

2. Multiplication of the photoelectron or electron-hole pair signal to detectable levels, usually by one or more multiplicative bombardment steps or an avalanche process.

3. Detection of charges induced by secondary electrons.

Photomultiplier tubes Vacuum photomultiplier tube (PMT) technologies can be broadly divided into two types: transmission-type and reflection-type. In the former technology, the photocathode material is deposited on the inside of a transparent window through which the photons enter, while in the latter technology, the photocathode material rests on a separate surface that the incident photons strike. The PMT cathode material has a low work function, thus when a photon hits the cathode, an electron can be liberated by the

photoelectric effect. The electron accelerated by the electric field hits a dynode which emits several($\sim$ 5) secondary electrons. The multiplication process is repeated typically about 10 times to generate a sufficient number of electrons, which are collected at the anode for delivery to the external circuit for sampling and eventual digitization.

Microchannel plate photomultiplier tubes Microchannel plate photomultiplier tubes (MCP-PMTs) incorporate a microchannel plate in place of conventional discrete dynodes. A typical MCP-PMT consists of two or more $\sim$ 1 mm-thick glass plates with densely packed $O(10\mu\text{m})$-diameter cylindrical holes (microchannels) between the photocathode and anode planes, separated by O(1 mm) gaps. Instead of discrete dynodes, the inner surface of each cylindrical hole with a length-to-diameter ratio of 40-100 serves as a continuous dynode. MCP-PMTs have the advantages of small thickness, good spatial resolution, excellent time resolution ($\sim$20 ps), and magnetic field tolerance up to 0.1 T.

Hybrid photon detectors Hybrid photon detectors (HPD) combine the sensitivity of a vacuum PMT with the excellent spatial and energy resolutions of a silicon sensor. Photoelectrons emitted from the photocathode are accelerated through a large potential difference of the order of 10-20 kV before hitting the silicon sensor. The absorbed kinetic energy of the photoelectron gives rise to the creation of electron-hole pairs, which in the depleted silicon sensor results in a detectable current. The number of created electron-hole pairs per photoelectron, i.e., the gain of the device, is given by eV/w, where e is the electronic charge, V is the applied potential difference, and $w \approx 3.7$ eV is the mean energy required to create an e-h pair in Si at room temperature. Since the gain is achieved in a single step, one can expect to have the excellent resolution of a simple Poisson statistic with a large mean.

Gaseous photon detectors In a gaseous photon detector (GPD) a photoelectron in a suitable gas mixture initiates an avalanche in a strong-field region, producing a large number of secondary impact-ionization electrons.

Solid-state photon detectors Compared to traditional vacuum and gaseous photodetectors, solid-state devices are more compact, lightweight, rugged, tolerant to magnetic fields, and often cheaper. They also allow fine pixelization, are easy to integrate into large systems, and can operate at low electric potentials, while matching or exceeding most performance criteria.

A.2 CHERENKOV DETECTORS

Charged particles passing a medium with a velocity faster than the speed of light in the medium radiate Cherenkov radiation. The particle velocity v can be determined by the Cherenkov angle: $cos\theta_c = \frac{c}{nv}$, where n is the refractive index of the medium and c is the speed of light in a vacuum.

Cherenkov detectors aim at recording the Cherenkov radiation produced by charged particles. Cherenkov counters contain two main elements: (1) a radiator through which the charged particle passes and (2) a photodetector.

Cherenkov counters can be classified into two types: imaging or threshold type, based on whether they make use of Cherenkov angle information.

Threshold Cherenkov detectors, which do not use Cherenkov angle information, make decisions simply based on whether the velocity of the particle is above the Cherenkov velocity threshold $v = c/n$.

Imaging Cherekov detectors, which use Cherenkov angle information, can be used to track particles as well as to identify them. Cherenkov photons are imaged onto a detector, so that their Cherenkov angles can be measured. Typically, the Cherenkov cone is mapped as a distorted circle at the photodetector. Imaging methods include focusing with a lens or a mirror, proximity focusing, focusing through an aperture, time imaging and correlated tracking. Many different imaging counters have been developed, including Differential Cherenkov detector, Ring-Imaging Cherenkov detectors (RICH), Detection of Internally Reflected Cherenkov detector (DIRC), and Time of Propagation (TOP) counter.

A.3 CALORIMETERS

Calorimeters measure the energy and direction of particles by absorption in the material and the energy deposited via the electromagnetic or hadronic shower.

Calorimeters can measure not only the energy of charged particles (except muons), but also of photons and neutral hadrons such as long-lived neutral kaons, or neutrons. Calorime-

ters can also measure particle jets and aid in full event reconstruction. Through the measurement of missing energy (or missing transverse energy at hadron colliders), one can deduce the presence of invisible particles, such as neutrinos, and hypothetical particles, such as dark-matter candidates. Calorimeters can also be used for the identification of particles. Using information on the longitudinal and transverse shapes of the energy deposition, calorimeters can separate electrons, photons, hadrons, and muons.

The capability to measure particle and jet energies, missing energy and particle identification, makes calorimeter central components of modern high-energy physics experiments.

APPENDIX B

SUPPLEMENTARY MATERIAL FOR THE SEARCH FOR BARYON-NUMBER-VIOLATING PROCESSES IN B^- DECAY TO FINAL STATE $\overline{\Xi}_C^0 \overline{\Lambda}_C^-$ AT BELLE

B.1 DECAY CHAINS FOR CHANNELS USED IN THE ANALYSIS

The 6 decay chains used in our analysis are depicted pictorially in Fig. 109.

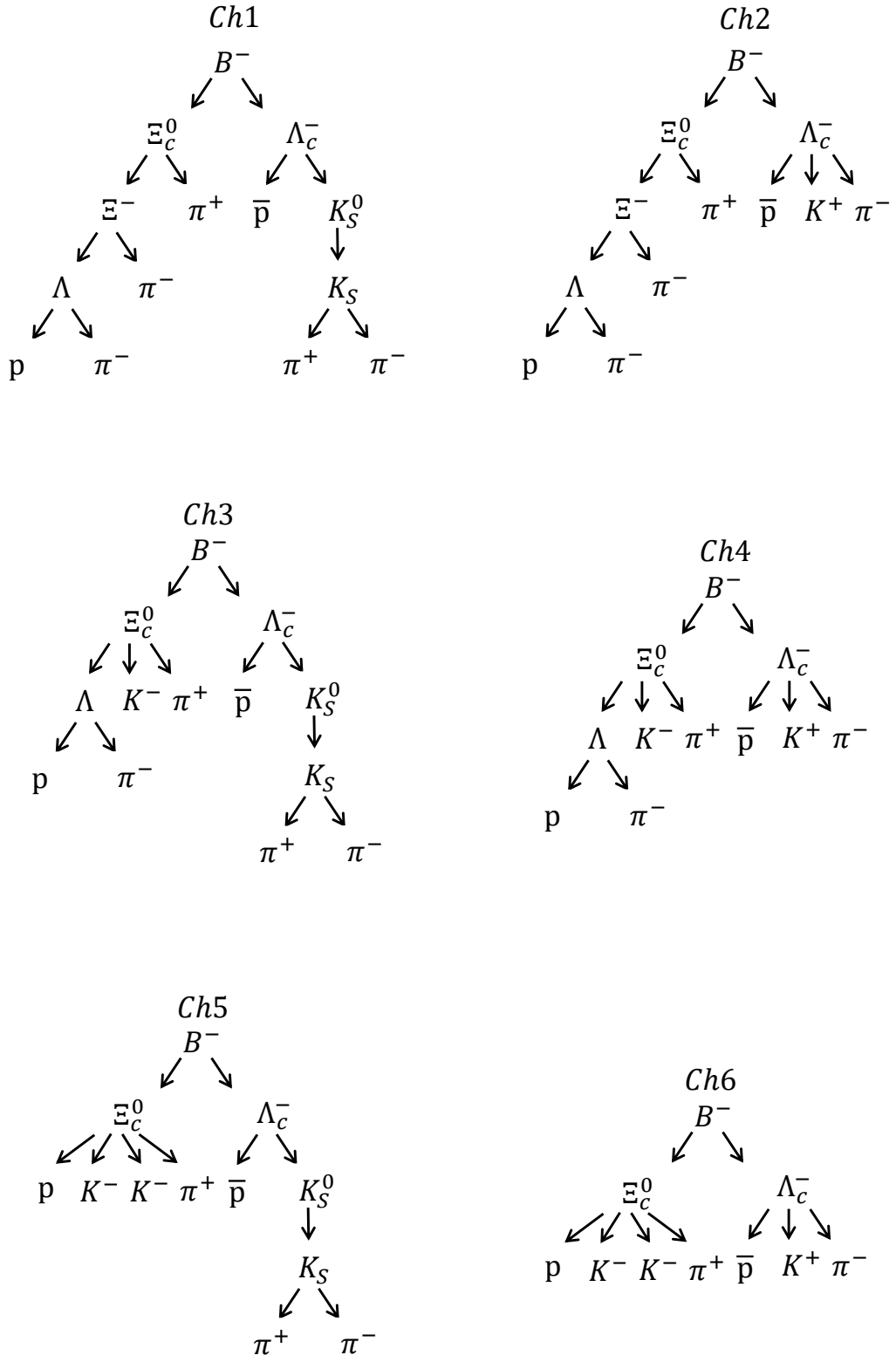


Figure 109: Decay chains for 6 (SM) channels used in our analysis.

B.2 EVTGEN DECAY MODELS USED IN THE ANALYSIS

EvtGen decay models used to generate the signal MC sample for each decay channels are shown below:

B.2.1 Ch1: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Xi^- \pi^+$ and $\overline{\Lambda}_c^- \to \overline{p} K_S^0$

```
Alias My_K0       K0
Alias My_anti-K0  anti-K0
ChargeConj My_K0 My_anti-K0

Alias My_B+      B+
Alias My_B-      B-
ChargeConj My_B+ My_B-

Alias My_Lambda_c+           Lambda_c+
Alias My_anti-Lambda_c-      anti-Lambda_c-
ChargeConj My_Lambda_c+ My_anti-Lambda_c-

Alias My_Xi_c0      Xi_c0
Alias My_anti-Xi_c0 anti-Xi_c0
ChargeConj My_Xi_c0 My_anti-Xi_c0

Decay Upsilon(4S)
0.5   My_B+   B-       VSS;
0.5   B+      My_B-    VSS;
Enddecay

Decay My_B+
0.5       My_anti-Xi_c0  My_Lambda_c+          PHOTOS PHSP;
0.5       My_Xi_c0       My_Lambda_c+          PHOTOS PHSP;
Enddecay
CDecay My_B-

Decay My_Xi_c0
1.000   Xi-  pi+   PHOTOS  PHSP;
Enddecay
CDecay My_anti-Xi_c0

Decay My_Lambda_c+
1.000         p+      My_anti-K0         PHOTOS PHSP;
Enddecay
CDecay My_anti-Lambda_c-

Decay My_K0
```

```
1.000     K_S0                                PHSP;
Enddecay
CDecay My_anti-K0
```

B.2.2 Ch2: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Xi^- \pi^+$ and $\overline{\Lambda}_c^- \to \overline{p} K^+ \pi^-$

```
Alias My_Lambda(1520)0       Lambda(1520)0
Alias My_anti-Lambda(1520)0  anti-Lambda(1520)0
ChargeConj My_Lambda(1520)0 My_anti-Lambda(1520)0

Alias My_K*0       K*0
Alias My_anti-K*0  anti-K*0
ChargeConj My_K*0 My_anti-K*0

Alias My_K0       K0
Alias My_anti-K0  anti-K0
ChargeConj My_K0 My_anti-K0

Alias My_B+      B+
Alias My_B-      B-
ChargeConj My_B+ My_B-

Alias My_Lambda_c+           Lambda_c+
Alias My_anti-Lambda_c-      anti-Lambda_c-
ChargeConj My_Lambda_c+ My_anti-Lambda_c-

Alias My_Xi_c0      Xi_c0
Alias My_anti-Xi_c0 anti-Xi_c0
ChargeConj My_Xi_c0 My_anti-Xi_c0

Decay Upsilon(4S)
0.5   My_B+   B-       VSS;
0.5   B+      My_B-    VSS;
Enddecay

Decay My_B+
0.5       My_anti-Xi_c0  My_Lambda_c+          PHOTOS PHSP;
0.5       My_Xi_c0       My_Lambda_c+          PHOTOS PHSP;
Enddecay
CDecay My_B-

Decay My_Xi_c0
1.000   Xi-  pi+   PHOTOS  PHSP;
Enddecay
CDecay My_anti-Xi_c0

Decay My_Lambda_c+
```

```
# according to PDG 2020
0.01960         p+    My_anti-K*0       PHOTOS PHSP;
0.01080         Delta++  K-             PHOTOS PHSP;
0.02200         My_Lambda(1520)0 pi+    PHOTOS PHSP;
0.03500         p+    K-  pi+           PHOTOS PHSP;
Enddecay
CDecay My_anti-Lambda_c-

Decay My_Lambda(1520)0
1.00   p+     K-    PHOTOS PHSP;
Enddecay
CDecay My_anti-Lambda(1520)0

Decay My_K*0
1.0000      K+  pi-                         PHOTOS VSS;
Enddecay
CDecay My_anti-K*0

Decay My_K0
1.000     K_S0                               PHSP;
Enddecay
CDecay My_anti-K0
```

B.2.3 Ch3: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Lambda K^- \pi^+$ and $\overline{\Lambda}_c^- \to \overline{p} K_S^0$

```
Alias My_K0       K0
Alias My_anti-K0  anti-K0
ChargeConj My_K0 My_anti-K0

Alias My_B+      B+
Alias My_B-      B-
ChargeConj My_B+ My_B-

Alias My_Lambda_c+           Lambda_c+
Alias My_anti-Lambda_c-      anti-Lambda_c-
ChargeConj My_Lambda_c+ My_anti-Lambda_c-

Alias My_Xi_c0      Xi_c0
Alias My_anti-Xi_c0 anti-Xi_c0
ChargeConj My_Xi_c0 My_anti-Xi_c0

Decay Upsilon(4S)
0.5   My_B+   B-       VSS;
0.5   B+      My_B-    VSS;
Enddecay

Decay My_B+
```

```
0.5       My_anti-Xi_c0  My_Lambda_c+          PHOTOS PHSP;
0.5       My_Xi_c0       My_Lambda_c+          PHOTOS PHSP;
Enddecay
CDecay My_B-

Decay My_Xi_c0
1.000   Lambda0  K-  pi+   PHOTOS  PHSP;
Enddecay
CDecay My_anti-Xi_c0

Decay My_Lambda_c+
1.000         p+      My_anti-K0         PHOTOS PHSP;
Enddecay
CDecay My_anti-Lambda_c-

Decay My_K0
1.000     K_S0                               PHSP;
Enddecay
CDecay My_anti-K0
```

B.2.4 Ch4: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Lambda K^- \pi^+$ and $\overline{\Lambda}_c^- \to \overline{p} K^+ \pi^-$

```
Alias My_Lambda(1520)0       Lambda(1520)0
Alias My_anti-Lambda(1520)0  anti-Lambda(1520)0
ChargeConj My_Lambda(1520)0 My_anti-Lambda(1520)0

Alias My_K*0       K*0
Alias My_anti-K*0  anti-K*0
ChargeConj My_K*0 My_anti-K*0

Alias My_K0       K0
Alias My_anti-K0  anti-K0
ChargeConj My_K0 My_anti-K0

Alias My_B+      B+
Alias My_B-      B-
ChargeConj My_B+ My_B-

Alias My_Lambda_c+           Lambda_c+
Alias My_anti-Lambda_c-      anti-Lambda_c-
ChargeConj My_Lambda_c+ My_anti-Lambda_c-

Alias My_Xi_c0      Xi_c0
Alias My_anti-Xi_c0 anti-Xi_c0
ChargeConj My_Xi_c0 My_anti-Xi_c0

Decay Upsilon(4S)
```

```
0.5   My_B+   B-       VSS;
0.5   B+      My_B-    VSS;
Enddecay

Decay My_B+
0.5       My_anti-Xi_c0  My_Lambda_c+          PHOTOS PHSP;
0.5       My_Xi_c0       My_Lambda_c+          PHOTOS PHSP;
Enddecay
CDecay My_B-

Decay My_Xi_c0
1.000   Lambda0  K-  pi+   PHOTOS  PHSP;
Enddecay
CDecay My_anti-Xi_c0

Decay My_Lambda_c+
# according to PDG 2020
0.01960         p+    My_anti-K*0       PHOTOS PHSP;
0.01080         Delta++  K-             PHOTOS PHSP;
0.02200         My_Lambda(1520)0 pi+    PHOTOS PHSP;
0.03500         p+    K-  pi+           PHOTOS PHSP;
Enddecay
CDecay My_anti-Lambda_c-

Decay My_Lambda(1520)0
1.00   p+     K-    PHOTOS PHSP;
Enddecay
CDecay My_anti-Lambda(1520)0

Decay My_K*0
1.0000      K+  pi-                         PHOTOS VSS;
Enddecay
CDecay My_anti-K*0

Decay My_K0
1.000     K_S0                               PHSP;
Enddecay
CDecay My_anti-K0
```

B.2.5 Ch5: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to pK^-K^-\pi^+$ and $\overline{\Lambda}_c^- \to \overline{p}K_S^0$

```
Alias My_K0       K0
Alias My_anti-K0  anti-K0
ChargeConj My_K0 My_anti-K0

Alias My_B+      B+
Alias My_B-      B-
```

```
ChargeConj My_B+ My_B-

Alias My_Lambda_c+           Lambda_c+
Alias My_anti-Lambda_c-      anti-Lambda_c-
ChargeConj My_Lambda_c+ My_anti-Lambda_c-

Alias My_Xi_c0      Xi_c0
Alias My_anti-Xi_c0 anti-Xi_c0
ChargeConj My_Xi_c0 My_anti-Xi_c0

Decay Upsilon(4S)
0.5   My_B+   B-       VSS;
0.5   B+      My_B-    VSS;
Enddecay

Decay My_B+
0.5       My_anti-Xi_c0  My_Lambda_c+          PHOTOS PHSP;
0.5       My_Xi_c0       My_Lambda_c+          PHOTOS PHSP;
Enddecay
CDecay My_B-

Decay My_Xi_c0
1.000   p+  K-  K-  pi+   PHOTOS  PHSP;
Enddecay
CDecay My_anti-Xi_c0

Decay My_Lambda_c+
1.000         p+      My_anti-K0         PHOTOS PHSP;
Enddecay
CDecay My_anti-Lambda_c-

Decay My_K0
1.000     K_S0                              PHSP;
Enddecay
CDecay My_anti-K0
```

B.2.6 Ch6: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to pK^-K^-\pi^+$ and $\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-$

```
Alias My_Lambda(1520)0       Lambda(1520)0
Alias My_anti-Lambda(1520)0  anti-Lambda(1520)0
ChargeConj My_Lambda(1520)0 My_anti-Lambda(1520)0

Alias My_K*0       K*0
Alias My_anti-K*0  anti-K*0
ChargeConj My_K*0 My_anti-K*0

Alias My_K0       K0
```

```
Alias My_anti-K0  anti-K0
ChargeConj My_K0 My_anti-K0

Alias My_B+      B+
Alias My_B-      B-
ChargeConj My_B+ My_B-

Alias My_Lambda_c+           Lambda_c+
Alias My_anti-Lambda_c-      anti-Lambda_c-
ChargeConj My_Lambda_c+ My_anti-Lambda_c-

Alias My_Xi_c0      Xi_c0
Alias My_anti-Xi_c0 anti-Xi_c0
ChargeConj My_Xi_c0 My_anti-Xi_c0

Decay Upsilon(4S)
0.5   My_B+   B-       VSS;
0.5   B+      My_B-    VSS;
Enddecay

Decay My_B+
0.5       My_anti-Xi_c0  My_Lambda_c+          PHOTOS PHSP;
0.5       My_Xi_c0       My_Lambda_c+          PHOTOS PHSP;
Enddecay
CDecay My_B-

Decay My_Xi_c0
1.000   p+  K-  K-  pi+   PHOTOS  PHSP;
Enddecay
CDecay My_anti-Xi_c0

Decay My_Lambda_c+
# according to PDG 2020
0.01960         p+    My_anti-K*0       PHOTOS PHSP;
0.01080         Delta++  K-             PHOTOS PHSP;
0.02200         My_Lambda(1520)0 pi+    PHOTOS PHSP;
0.03500         p+    K-  pi+           PHOTOS PHSP;
Enddecay
CDecay My_anti-Lambda_c-

Decay My_Lambda(1520)0
1.00   p+     K-    PHOTOS PHSP;
Enddecay
CDecay My_anti-Lambda(1520)0

Decay My_K*0
1.0000      K+  pi-                        PHOTOS VSS;
```

```
Enddecay
CDecay My_anti-K*0

Decay My_K0
1.000     K_S0                                PHSP;
Enddecay
CDecay My_anti-K0
```

B.3 TRANSVERSE MOMENTUM DISTRIBUTIONS AND RECONSTRUCTION EFFICIENCIES FOR FINAL-STATE CHARGED PARTICLES

Plots in Figures 110 through 115 show the transverse momentum distributions and reconstruction efficiencies for final-state charged particles (pion, kaon, and proton) in signal MC. The red plots show transverse momentum distributions in MC truth. The green plots show transverse momentum distributions of reconstructed final-state charged particles. The blue plots show the reconstruction efficiency of final-state charged particles.

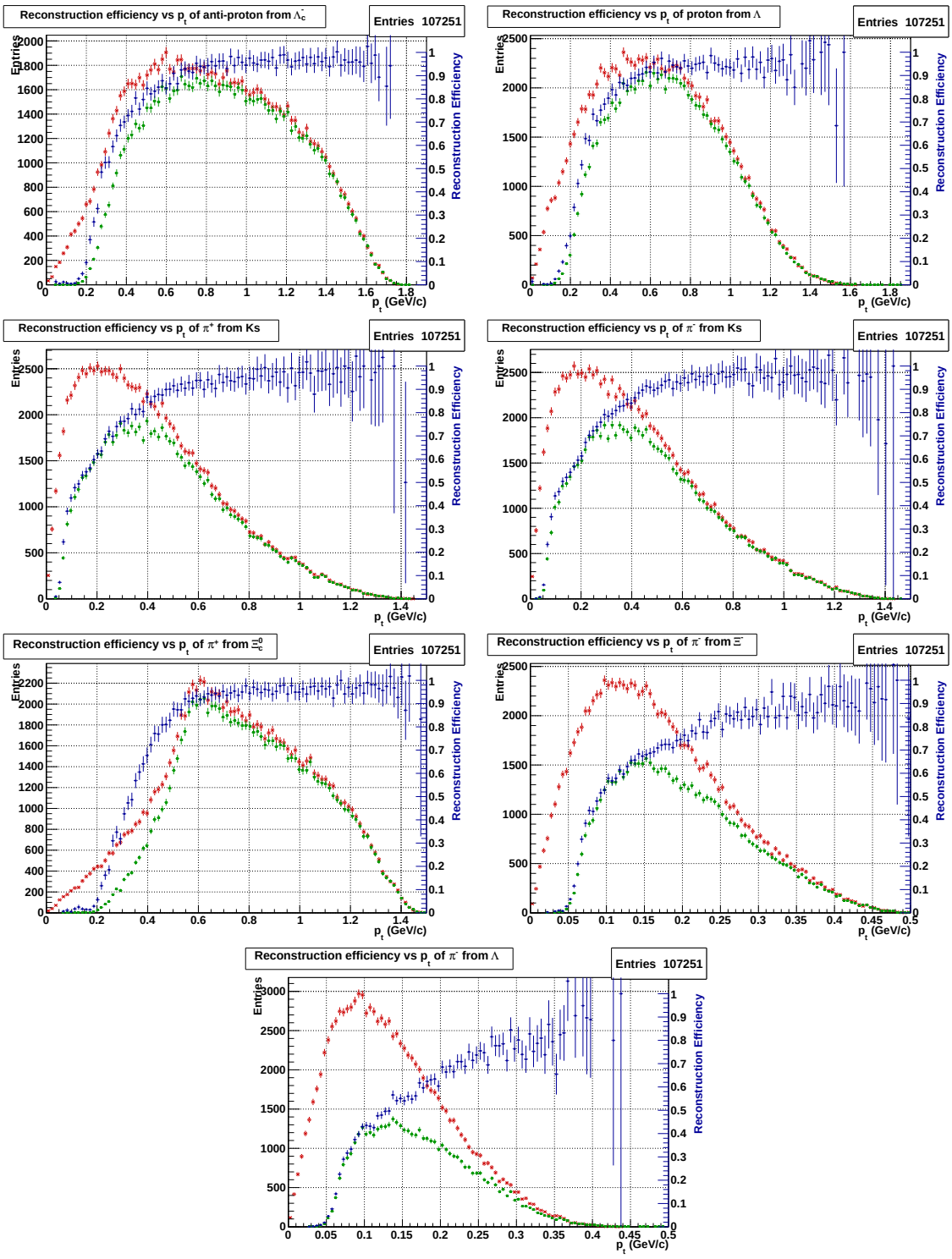


Figure 110: Reconstruction efficiency vs transverse momentum for final-state charged particles in signal MC of ch1. Generated transverse momentum distribution (*i.e.* MC truth, in red color). Reconstructed transverse momentum distributions (in green color). Reconstruction efficiency vs $p_\perp$ (in blue color).

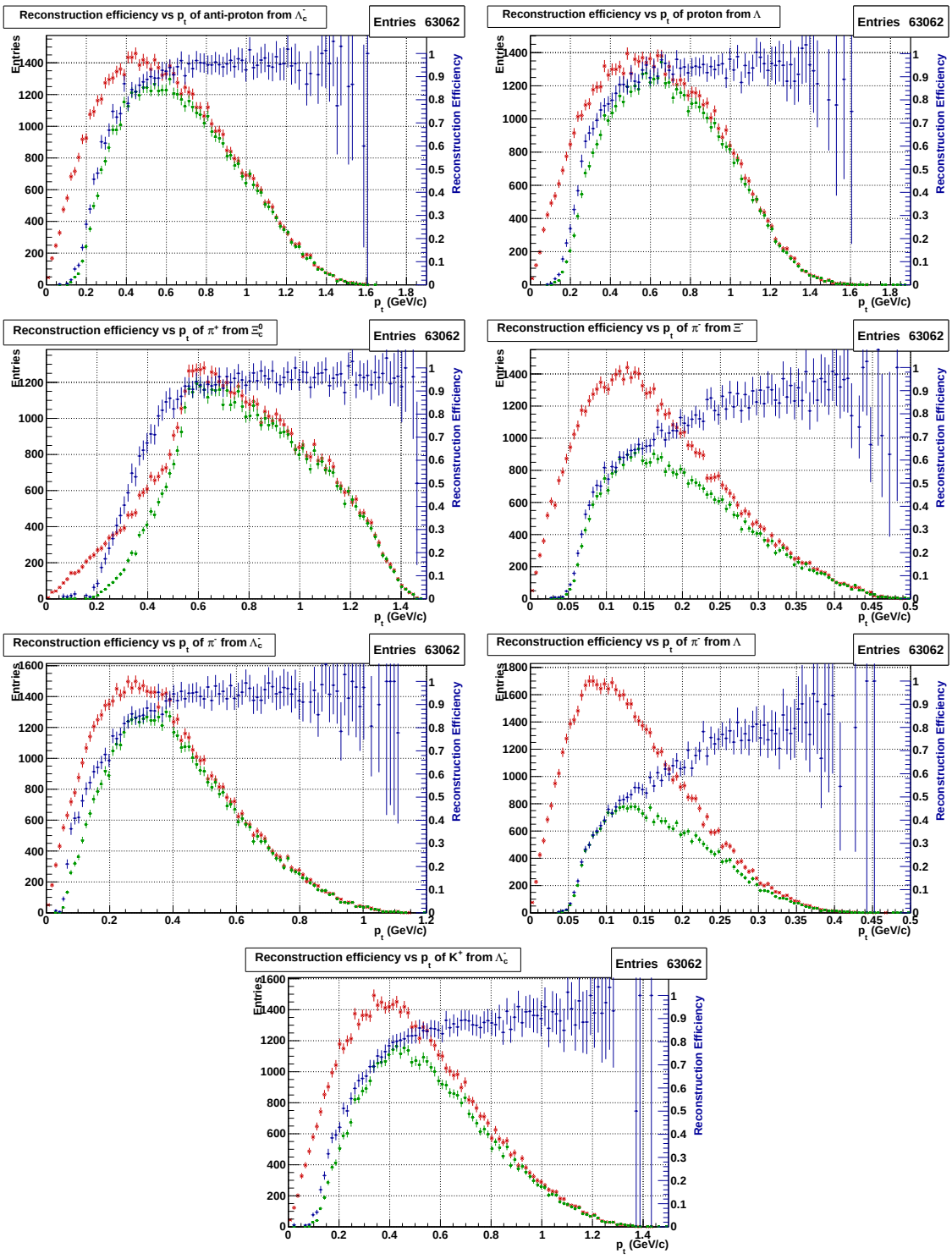


Figure 111: Reconstruction efficiency vs transverse momentum for final-state charged particles in signal MC of ch2. Generated transverse momentum distribution (*i.e.* MC truth, in red color). Reconstructed transverse momentum distributions (in green color). Reconstruction efficiency vs $p_{\perp}$ (in blue color).

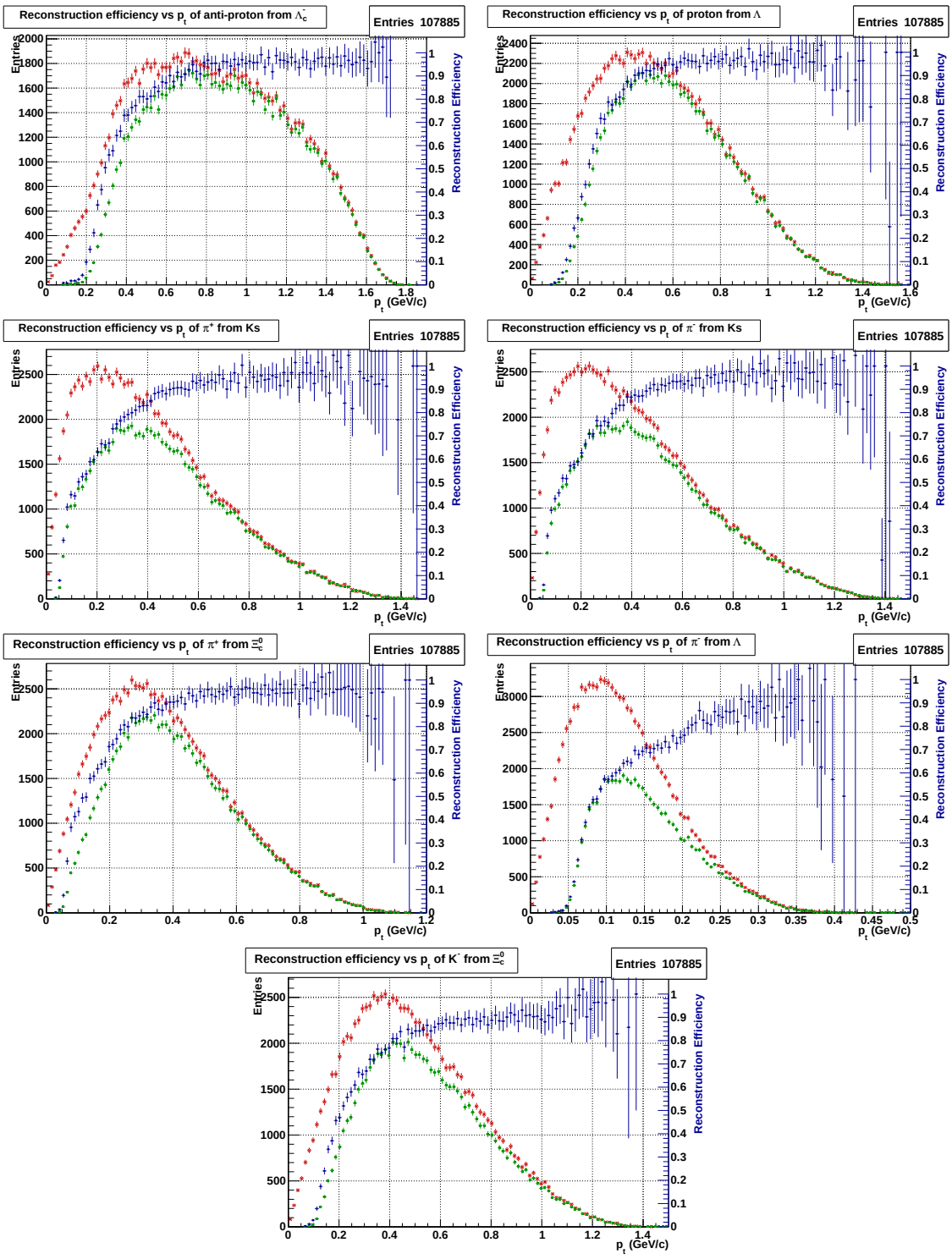


Figure 112: Reconstruction efficiency vs transverse momentum for final-state charged particles in signal MC of ch3. Generated transverse momentum distribution (*i.e.* MC truth, in red color). Reconstructed transverse momentum distributions (in green color). Reconstruction efficiency vs $p_{\perp}$ (in blue color).

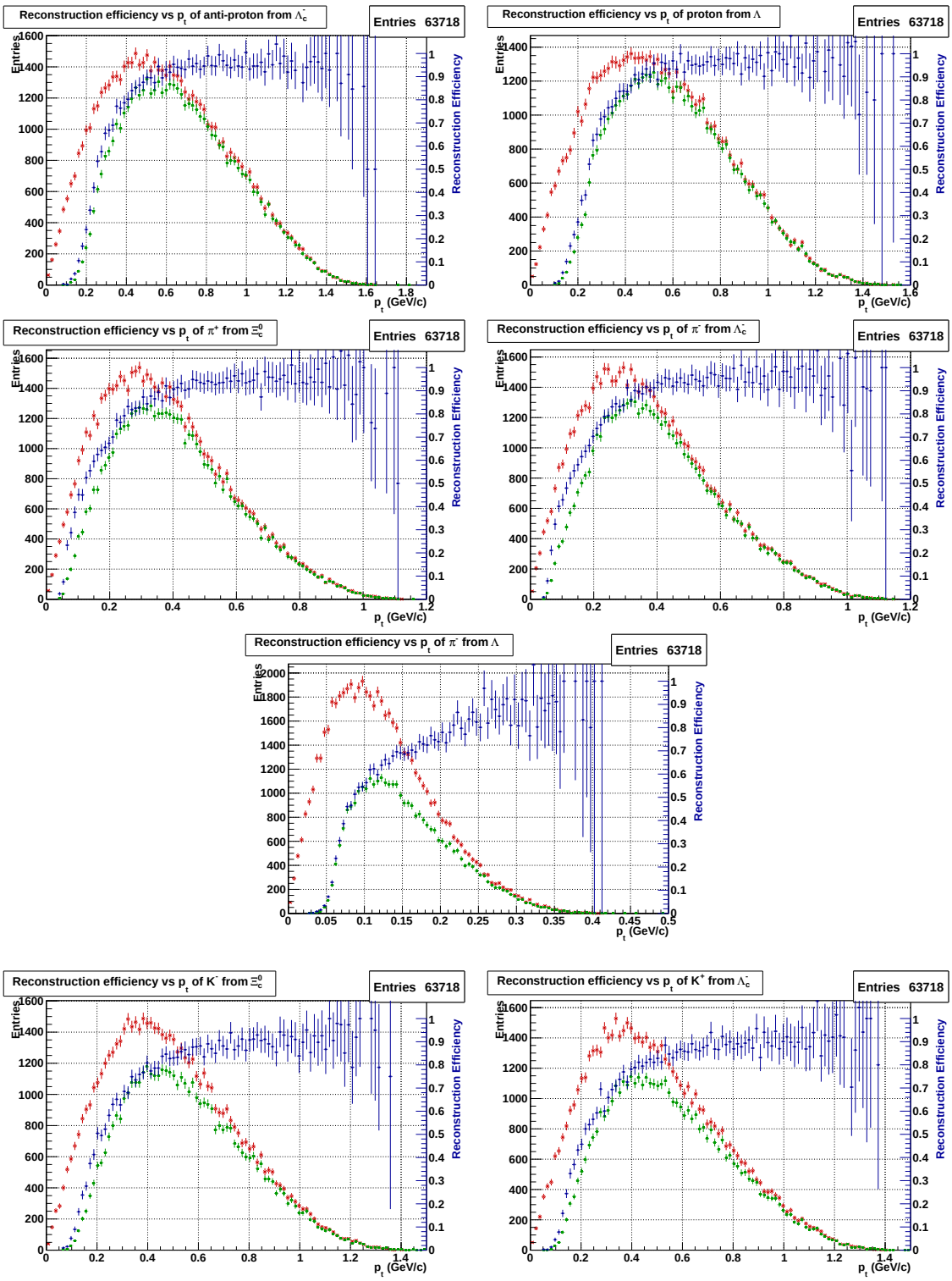


Figure 113: Reconstruction efficiency vs transverse momentum for final-state charged particles in signal MC of ch4. Generated transverse momentum distribution (*i.e.* MC truth, in red color). Reconstructed transverse momentum distributions (in green color). Reconstruction efficiency vs $p_\perp$ (in blue color).

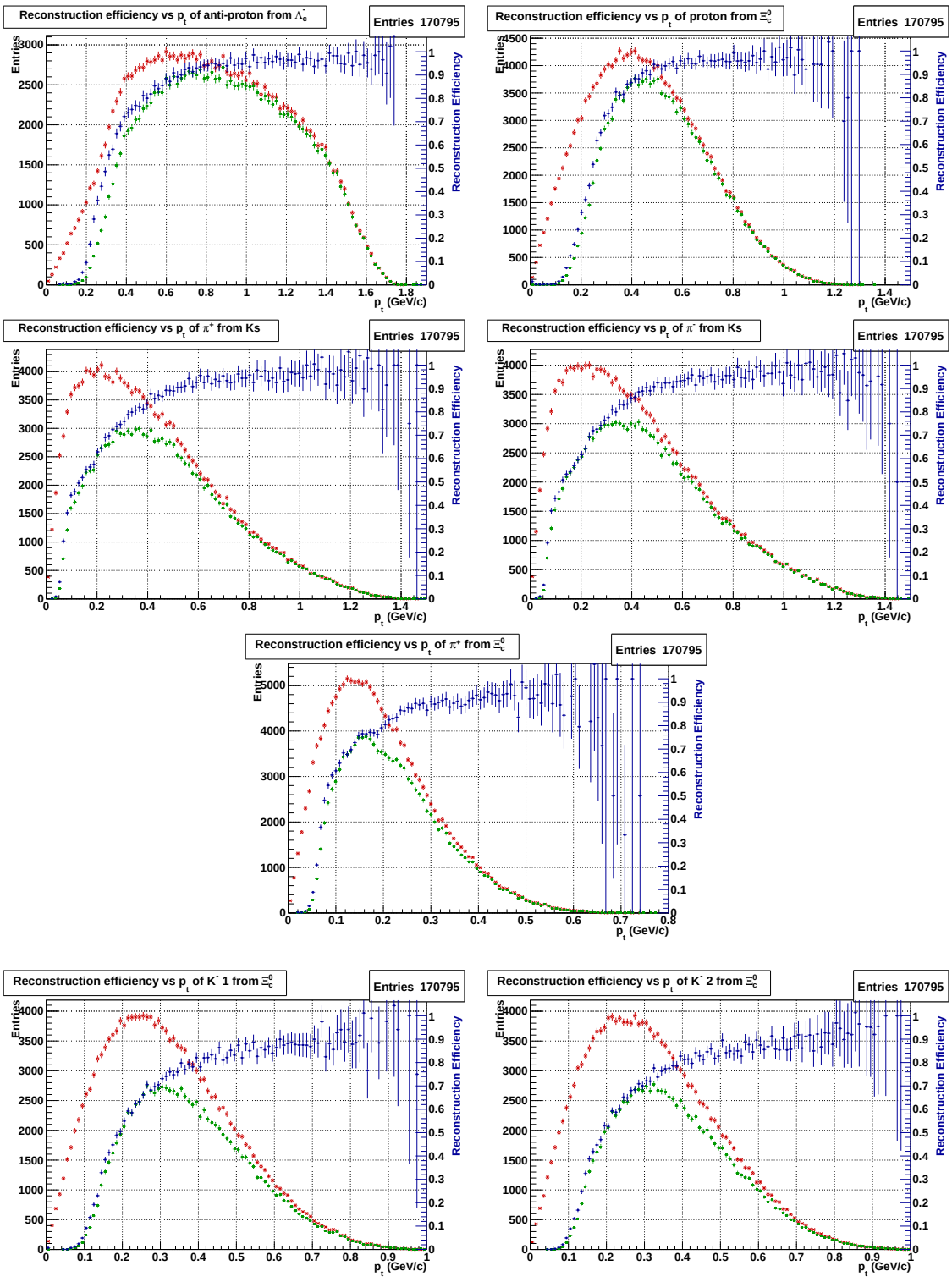


Figure 114: Reconstruction efficiency vs transverse momentum for final-state charged particles in signal MC of ch5. Generated transverse momentum distribution (*i.e.* MC truth, in red color). Reconstructed transverse momentum distributions (in green color). Reconstruction efficiency vs $p_{\perp}$ (in blue color).

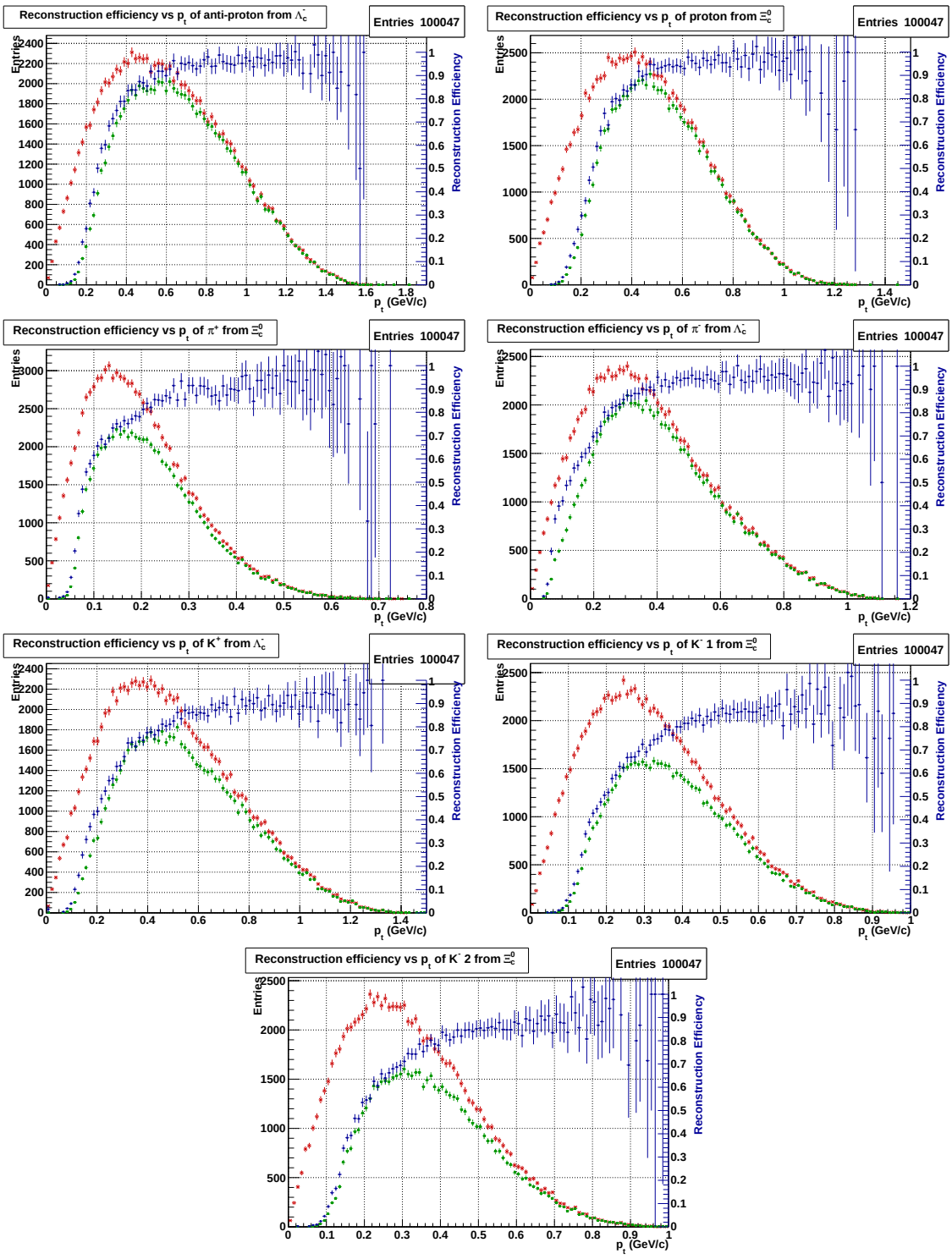


Figure 115: Reconstruction efficiency vs transverse momentum for final-state charged particles in signal MC of ch6. Generated transverse momentum distribution (*i.e.* MC truth, in red color). Reconstructed transverse momentum distributions (in green color). Reconstruction efficiency vs $p_{\perp}$ (in blue color).

B.4 INVARIANT MASS DISTRIBUTIONS FOR RECONSTRUCTED SIGNAL CANDIDATES

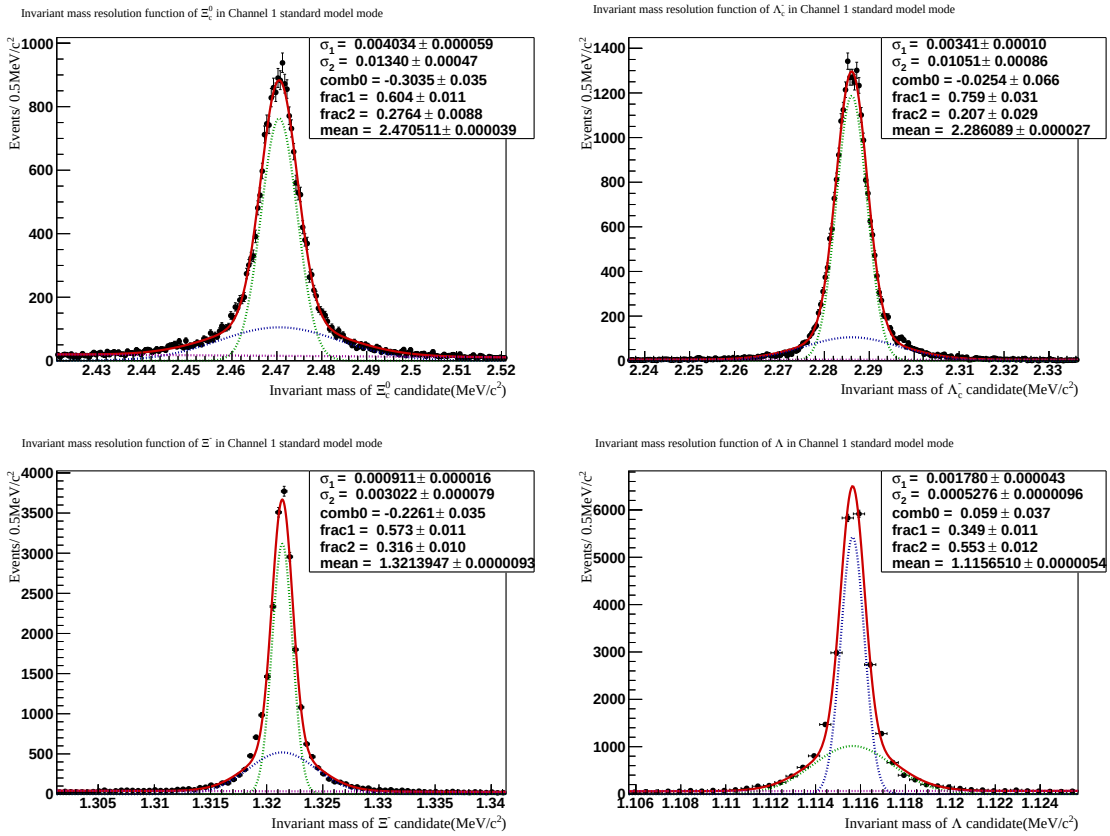


Figure 116: Invariant mass distributions for $\Xi_c^0, \overline{\Lambda}_c^-, \Xi^-, \Lambda$ candidates in signal MC sample for ch1 ($\Xi_c^0 \to \Xi^-\pi^+$, $\overline{\Lambda}_c^- \to \overline{p}K_S^0$). The red solid lines show the fit results with a double-Gaussian function and a first-order Chebychev polynomial.

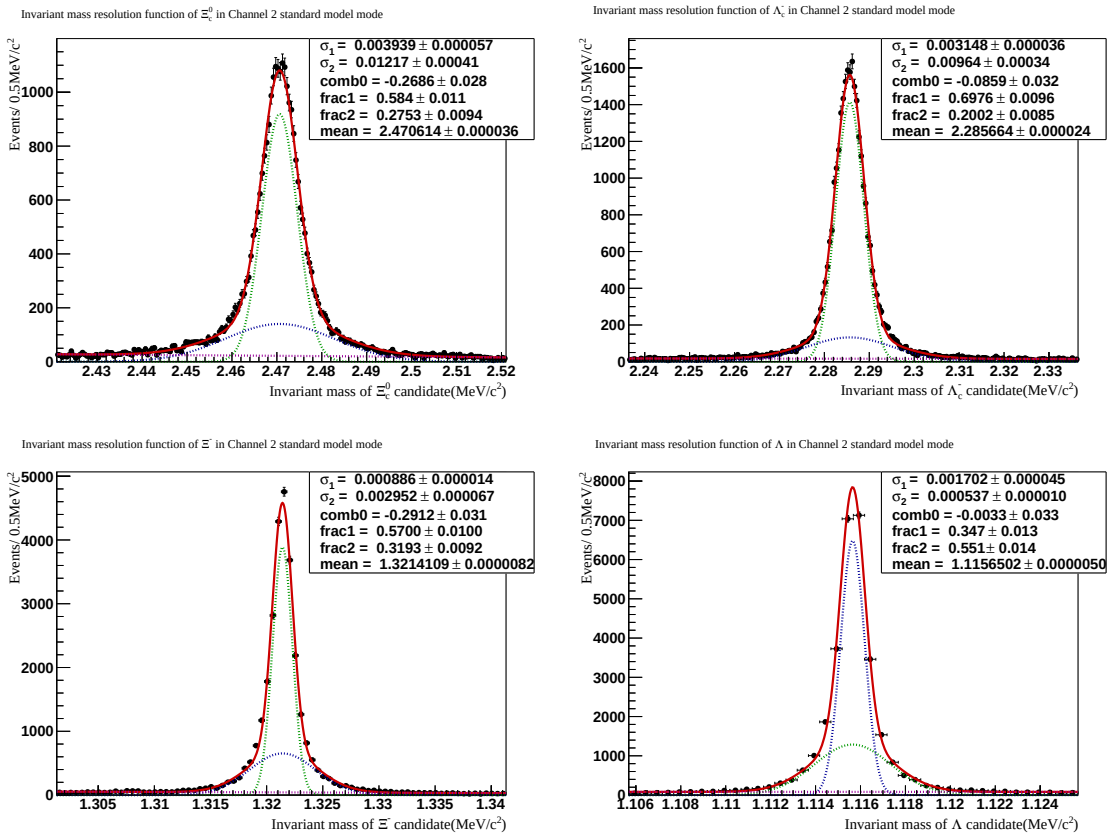


Figure 117: Invariant mass distributions for $\Xi_c^0, \overline{\Lambda}_c^-, \Xi^-, \Lambda$ candidates in signal MC sample for ch2 ($\Xi_c^0 \to \Xi^-\pi^+$, $\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-$). The red solid lines show the fit results with a double-Gaussian function and a first-order Chebychev polynomial.

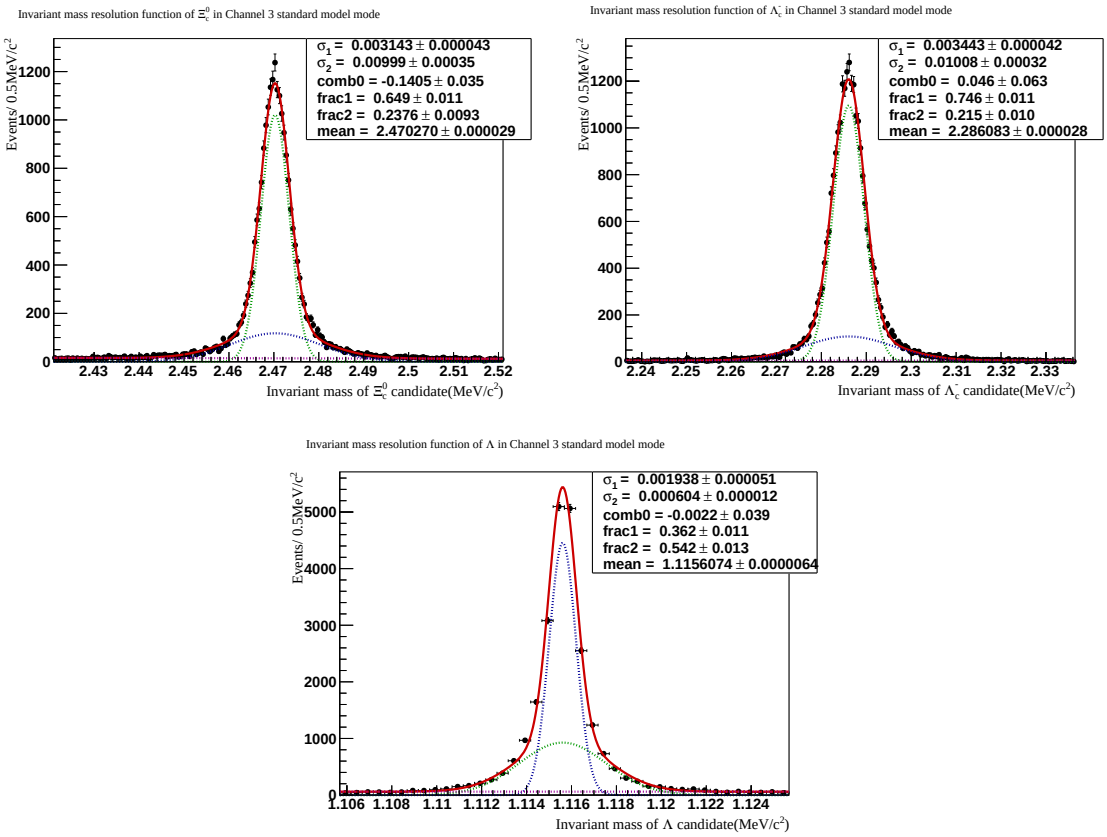


Figure 118: Invariant mass distributions for $\Xi_c^0, \overline{\Lambda}_c^-, \Lambda$ candidates in signal MC sample for ch3 ($\Xi_c^0 \to \Lambda K^- \pi^+$, $\overline{\Lambda}_c^- \to \overline{p} K_S^0$). The red solid lines show the fit results with a double-Gaussian function and a first-order Chebychev polynomial.

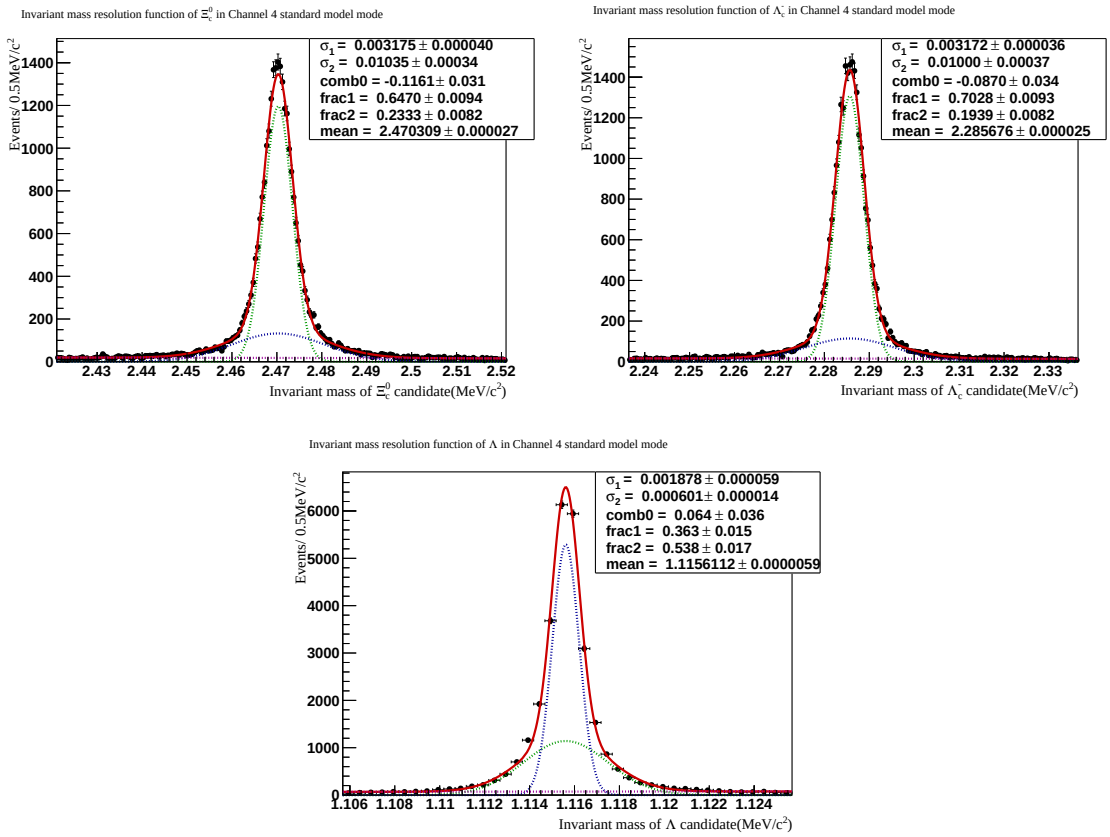


Figure 119: Invariant mass distributions for $\Xi_c^0, \overline{\Lambda}_c^-, \Lambda$ candidates in signal MC sample for ch4 ($\Xi_c^0 \to \Lambda K^- \pi^+$, $\overline{\Lambda}_c^- \to \bar{p} K^+ \pi^-$). The red solid lines show the fit results with a double-Gaussian function and a first-order Chebychev polynomial.

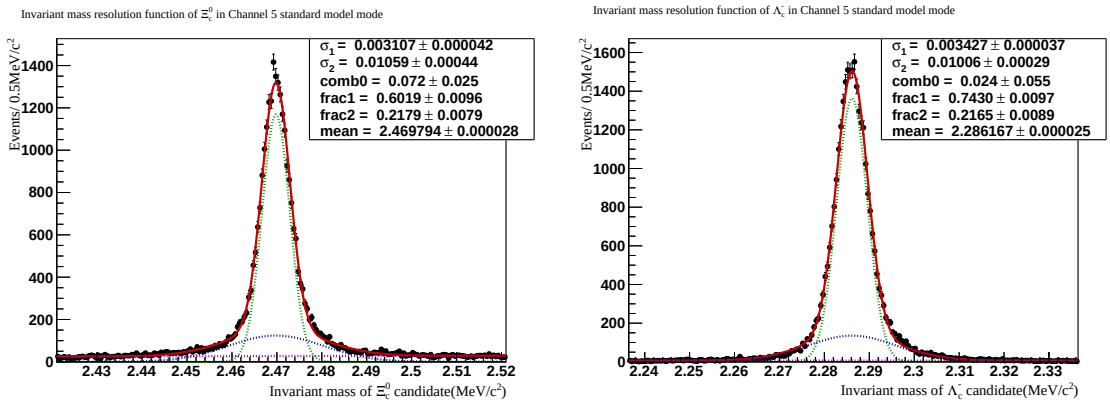


Figure 120: Invariant mass distributions for $\Xi_c^0, \overline{\Lambda}_c^-$ candidates in signal MC sample for ch5 ($\Xi_c^0 \to p K^- K^- \pi^+$, $\overline{\Lambda}_c^- \to \bar{p} K_S^0$). The red solid lines show the fit results with a double-Gaussian function and a first-order Chebychev polynomial.

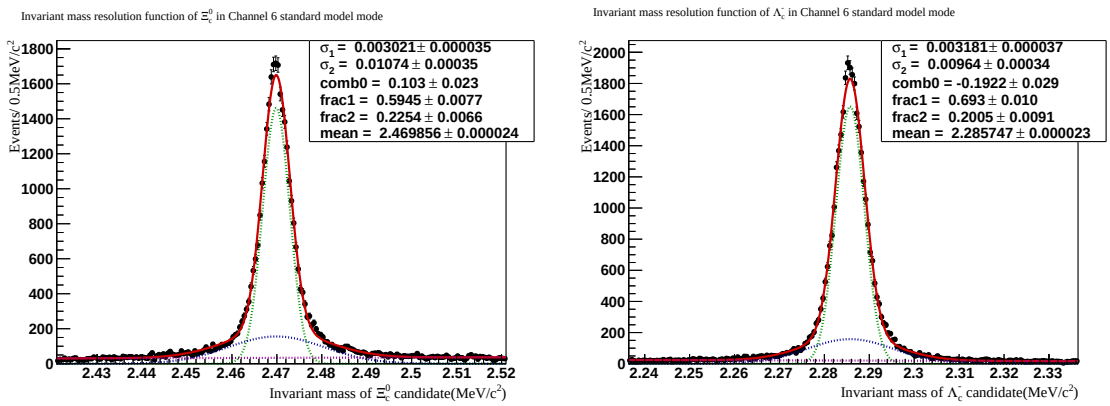


Figure 121: Invariant mass distributions for $\Xi_c^0, \overline{\Lambda}_c^-$ candidates in signal MC sample for ch6 ($\Xi_c^0 \to pK^-K^-\pi^+$, $\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-$). The red solid lines show the fit results with a double-Gaussian function and a first-order Chebychev polynomial.

B.5 SEPARATE FIT RESULTS FOR SIX CHANNELS WITH $P_\perp > 50$ MEV/C CUTS

B.5.1 SM: $B^- \to \Xi_c^0\overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Xi^-\pi^+$ and $\overline{\Lambda}_c^- \to \overline{p}K_S^0$

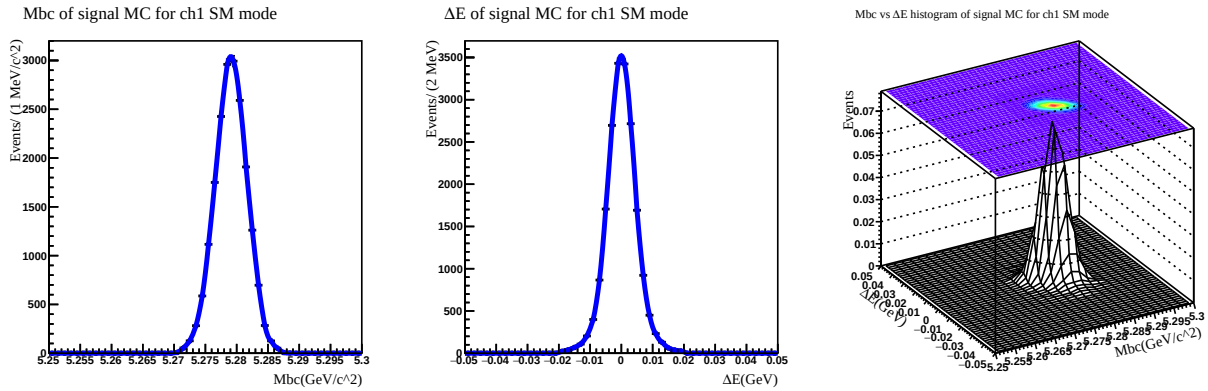


Figure 122: 2D smoothed histogram and projections of ch1 SM signal MC samples.

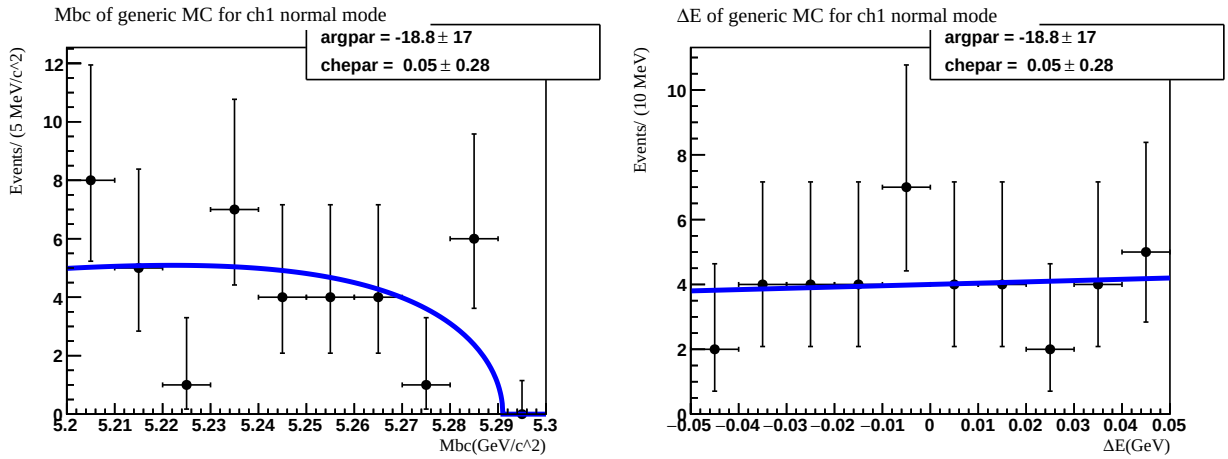


Figure 123: Projections of the 2D fit for ch1 SM generic background MC samples.

B.5.2 SM: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Xi^- \pi^+$ and $\overline{\Lambda}_c^- \to \overline{p} K^+ \pi^-$

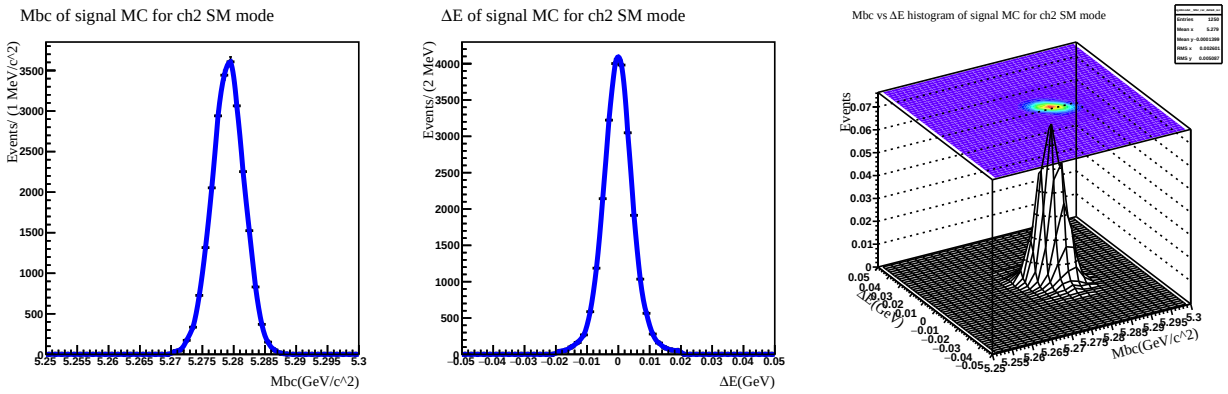


Figure 124: 2D smoothed histogram and projections of ch2 SM signal MC samples.

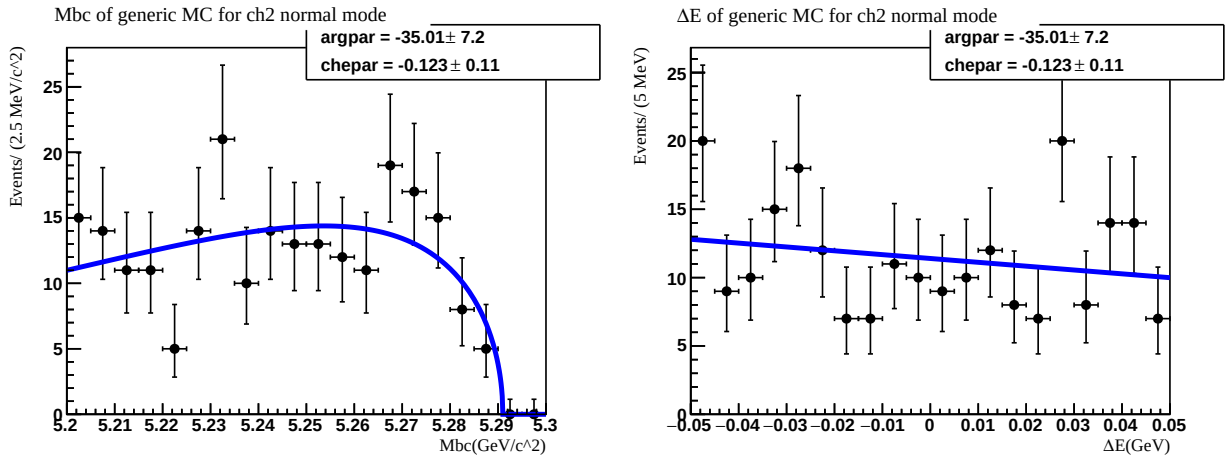


Figure 125: Projections of the 2D fit for ch2 SM generic background MC samples.

B.5.3 SM: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Lambda K^- \pi^+$ and $\overline{\Lambda}_c^- \to \overline{p} K_S^0$

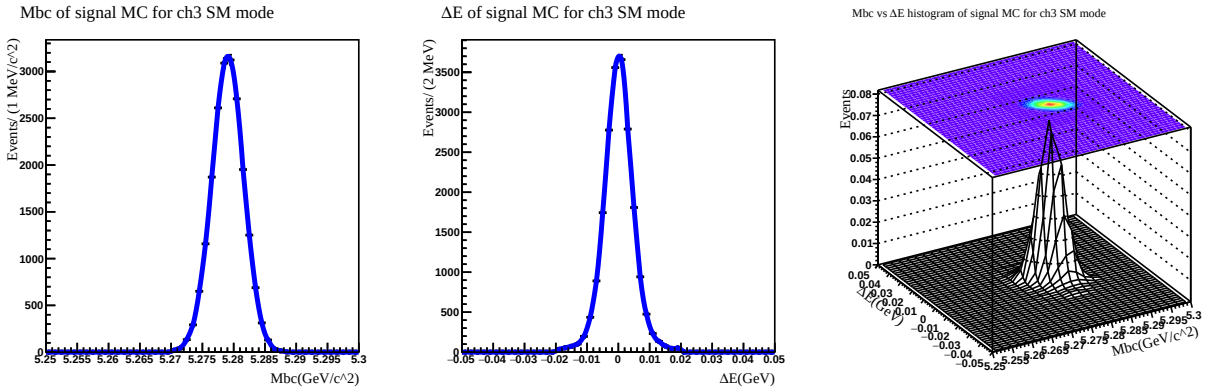


Figure 126: 2D smoothed histogram and projections of ch3 SM signal MC samples.

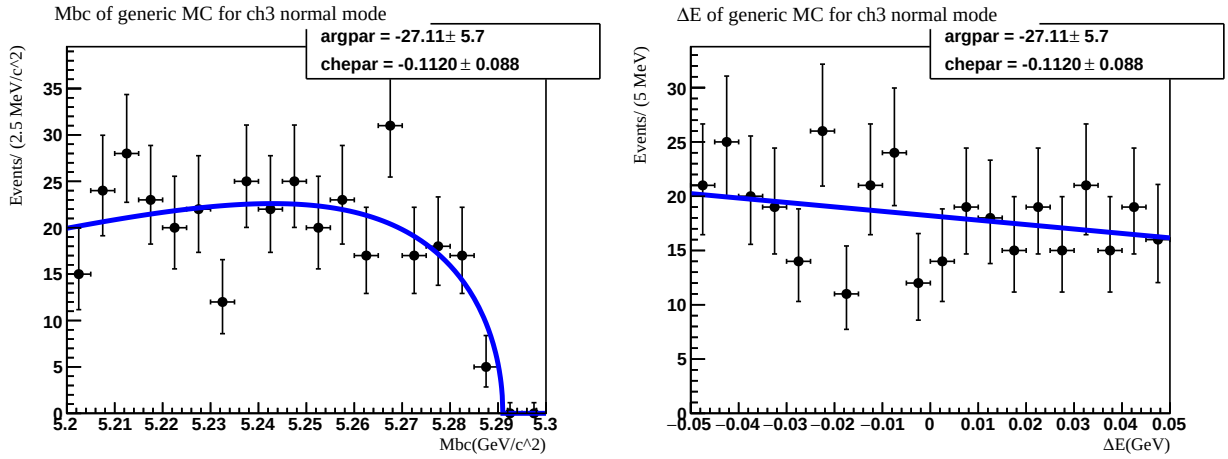


Figure 127: Projections of the 2D fit for ch3 SM generic background MC samples.

B.5.4 SM: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Lambda K^- \pi^+$ and $\overline{\Lambda}_c^- \to \overline{p} K^+ \pi^-$

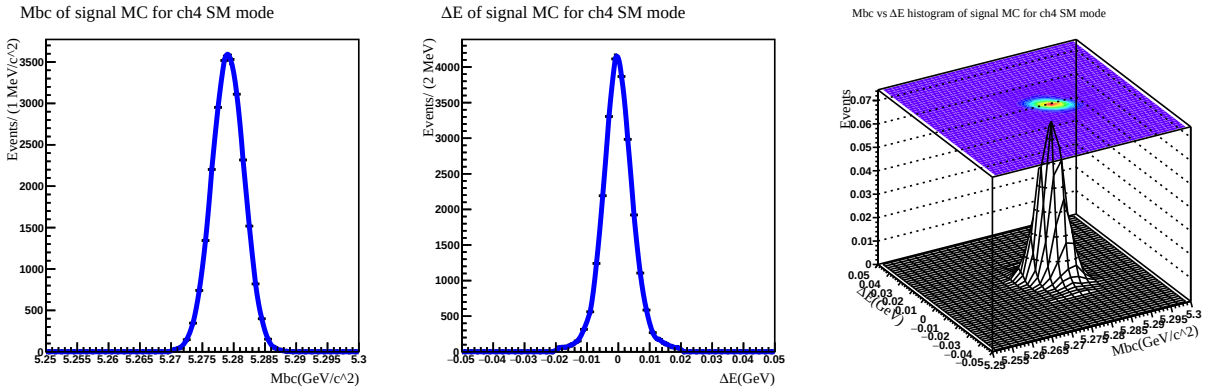


Figure 128: 2D smoothed histogram and projections of ch4 SM signal MC samples.

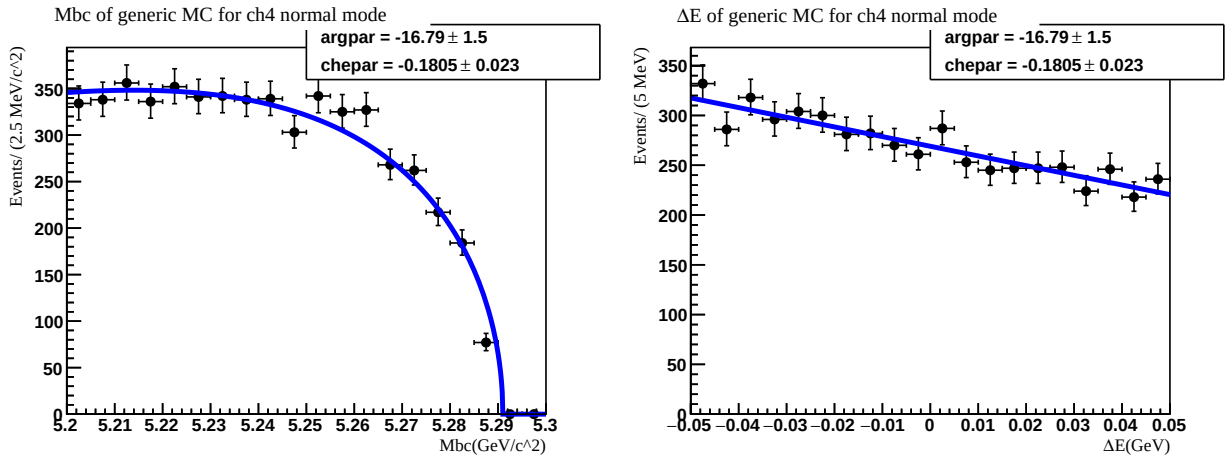


Figure 129: Projections of the 2D fit for ch4 SM generic background MC samples.

B.5.5 SM: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to pK^-K^-\pi^+$ and $\overline{\Lambda}_c^- \to \overline{p}K_S^0$

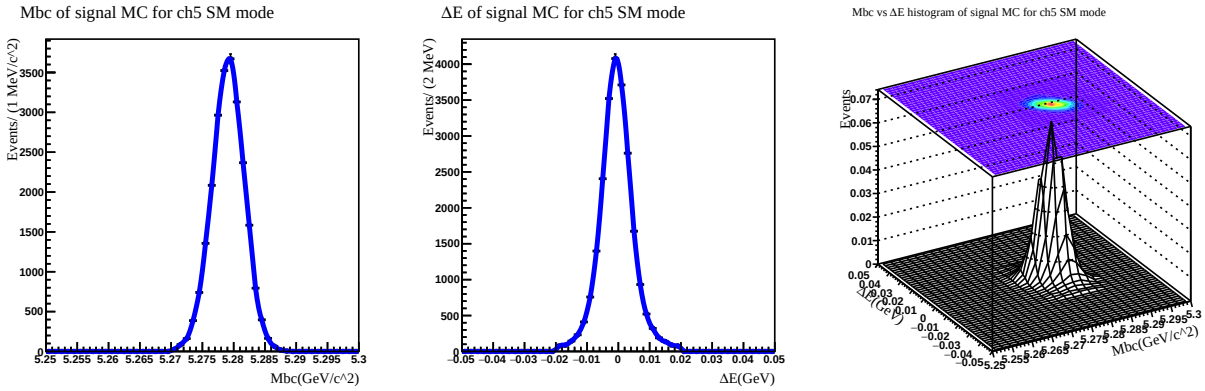


Figure 130: 2D smoothed histogram and projections of ch5 SM signal MC samples.

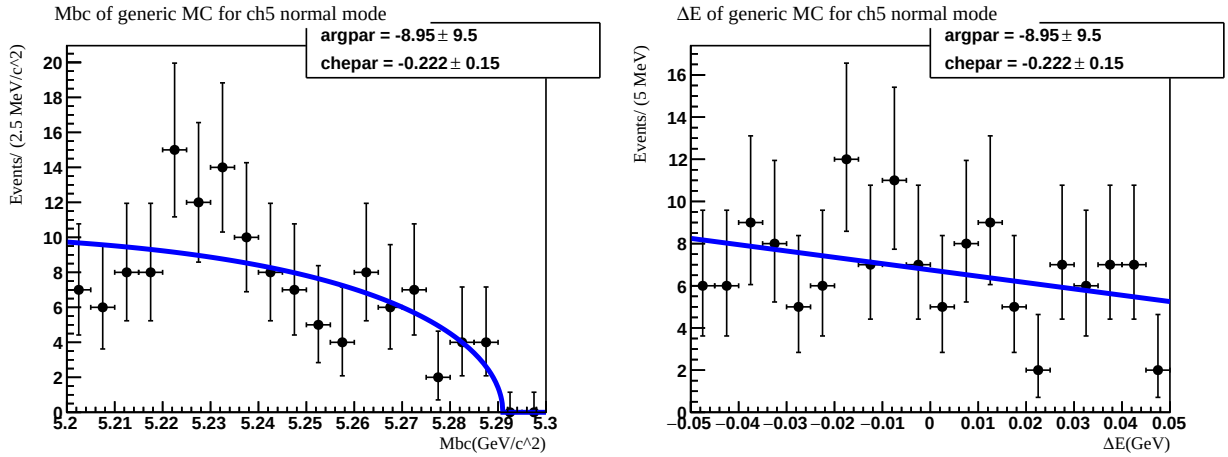


Figure 131: Projections of the 2D fit for ch5 SM generic background MC samples.

B.5.6 SM: $B^- \to \Xi_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to pK^-K^-\pi^+$ and $\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-$

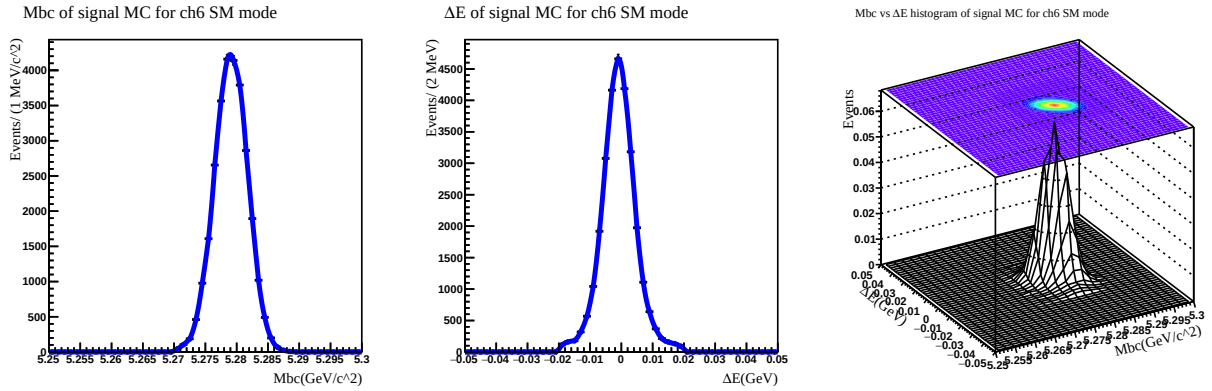


Figure 132: 2D smoothed histogram and projections of ch6 SM signal MC samples.

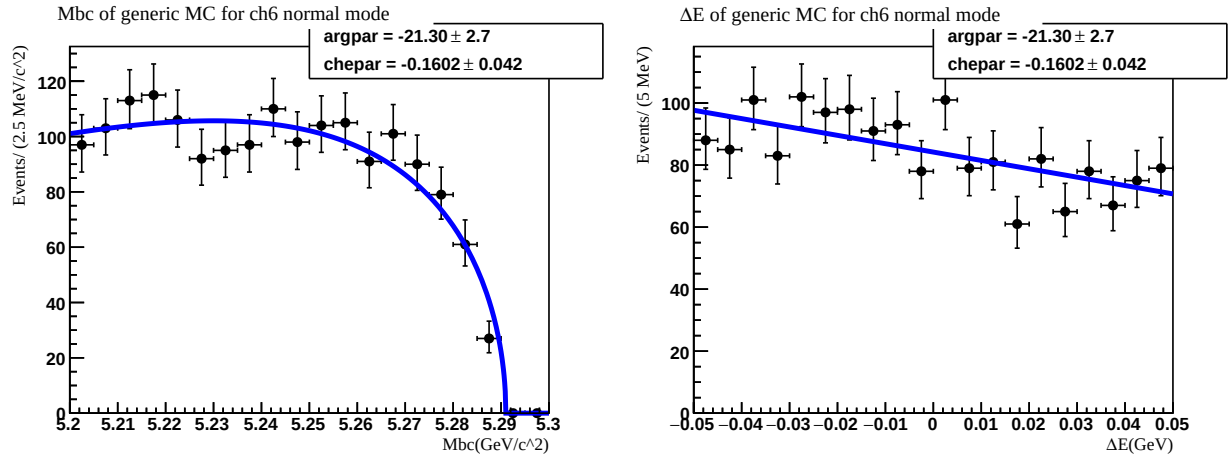


Figure 133: Projections of the 2D fit for ch6 SM generic background MC samples.

B.5.7 BNV: $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Xi^-\pi^+$, $\overline{\Lambda}_c^- \to \overline{p}K_S^0$

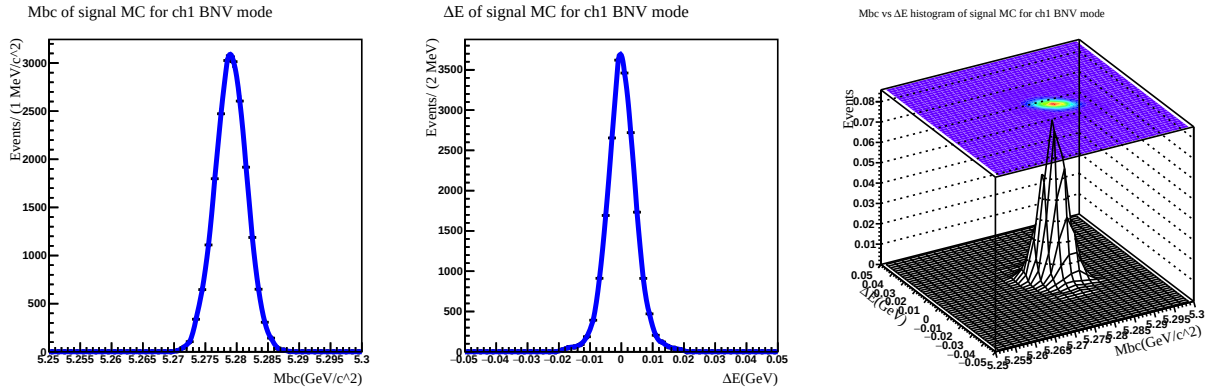


Figure 134: 2D smoothed histogram and projections of ch1 BNV signal MC samples.

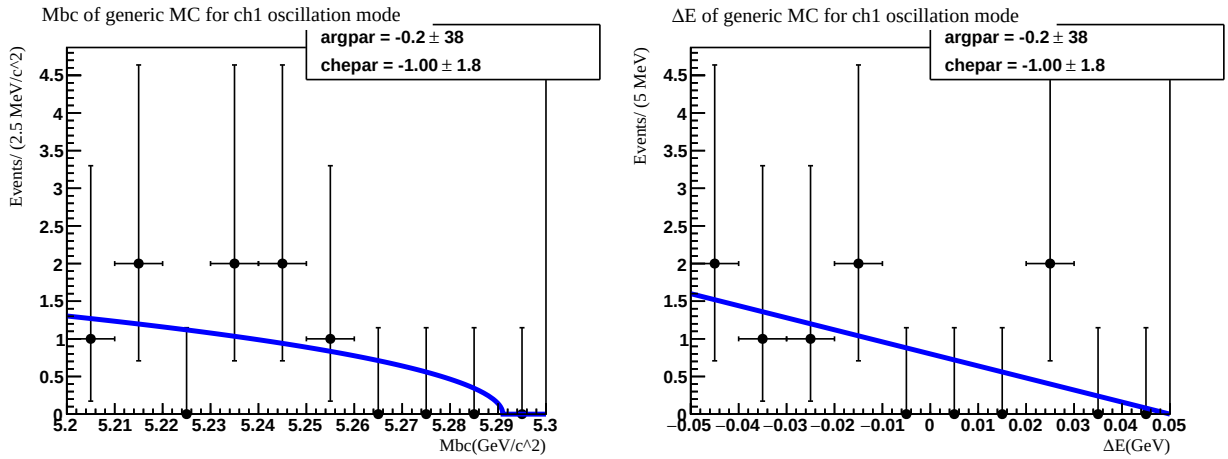


Figure 135: Projections of the 2D fit for ch1 BNV generic background MC samples.

B.5.8 BNV: $B^- \to \overline{\Xi}{}^0_c \overline{\Lambda}{}^-_c$ with $\Xi^0_c \to \Xi^- \pi^+$ and $\overline{\Lambda}{}^-_c \to \overline{p} K^+ \pi^-$

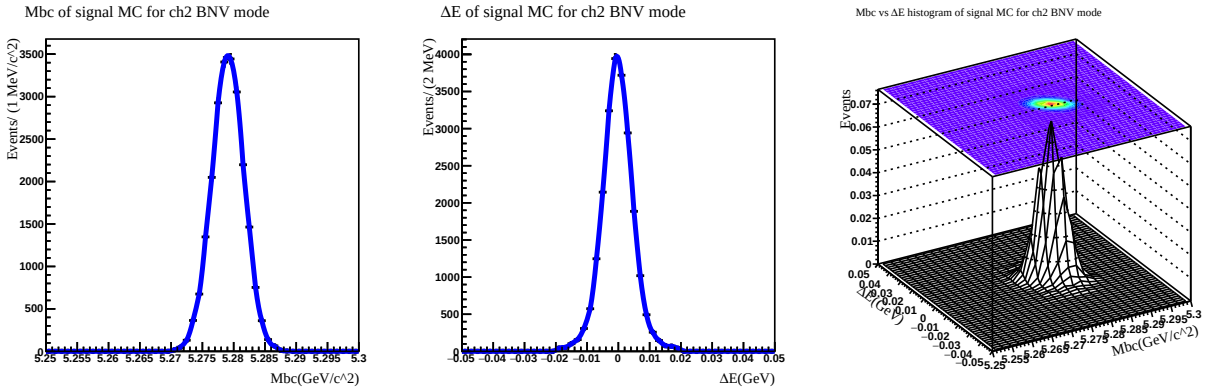


Figure 136: 2D smoothed histogram and projections of ch2 BNV signal MC samples.

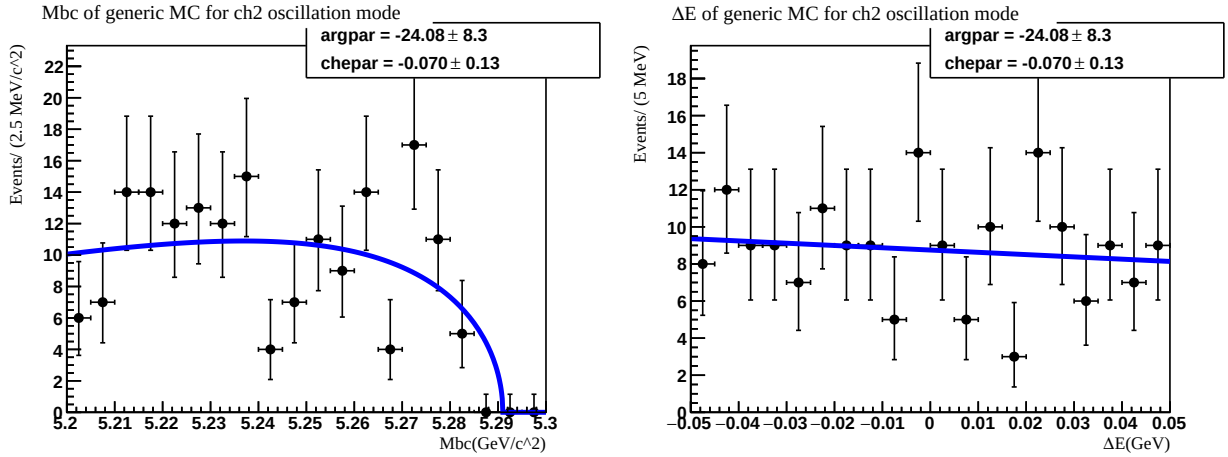


Figure 137: Projections of the 2D fit for ch2 BNV generic background MC samples.

B.5.9 BNV: $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Lambda K^- \pi^+$ and $\overline{\Lambda}_c^- \to \overline{p} K_S^0$

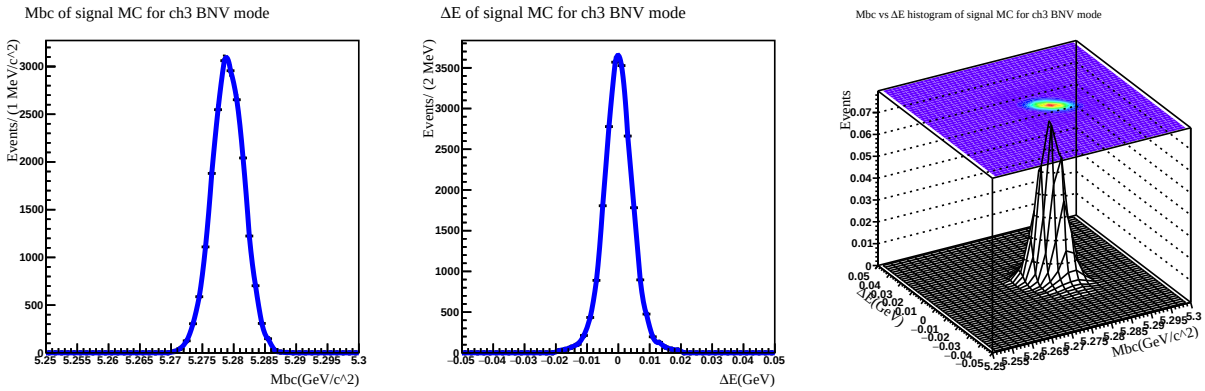


Figure 138: 2D smoothed histogram and projections of ch3 BNV signal MC samples.

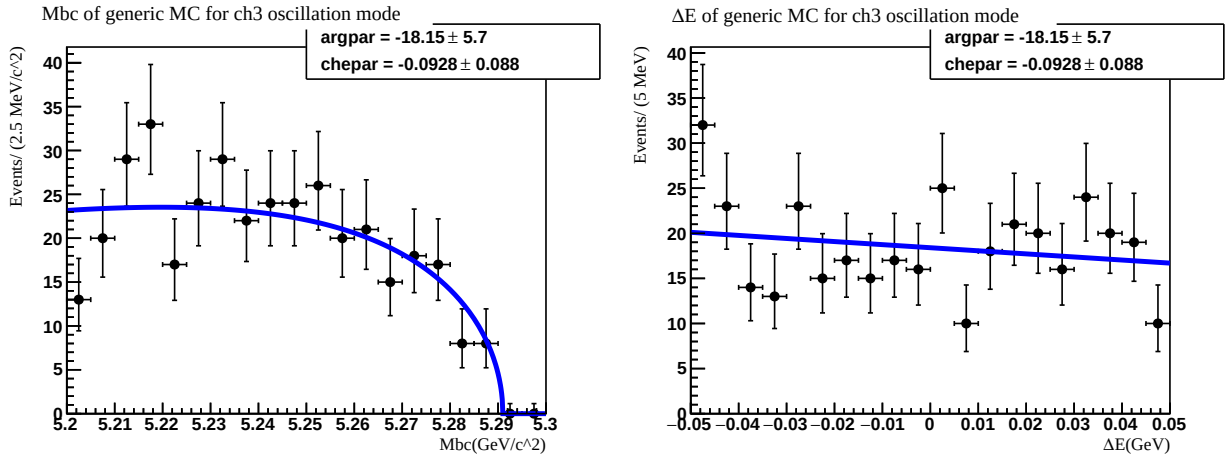


Figure 139: Projections of the 2D fit for ch3 BNV generic background MC samples.

B.5.10 BNV: $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to \Lambda K^- \pi^+$ and $\overline{\Lambda}_c^- \to \overline{p} K^+ \pi^-$

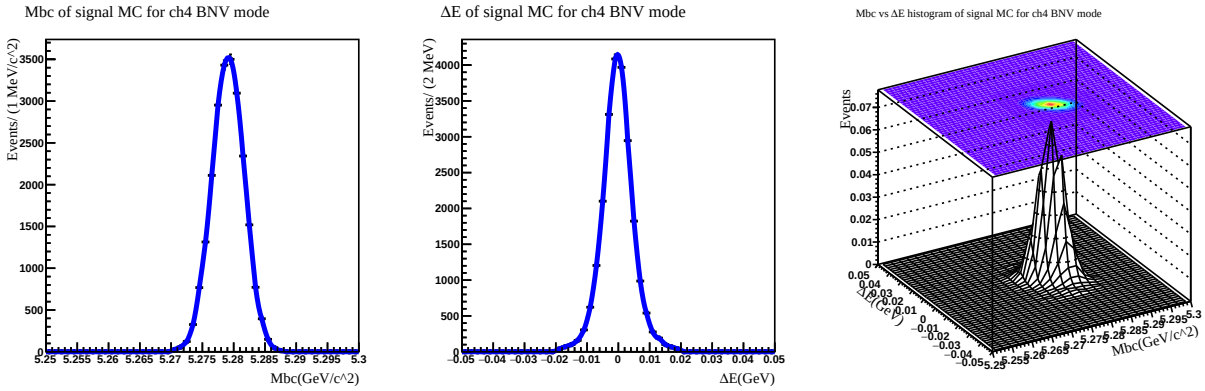


Figure 140: 2D smoothed histogram and projections of ch4 BNV signal MC samples.

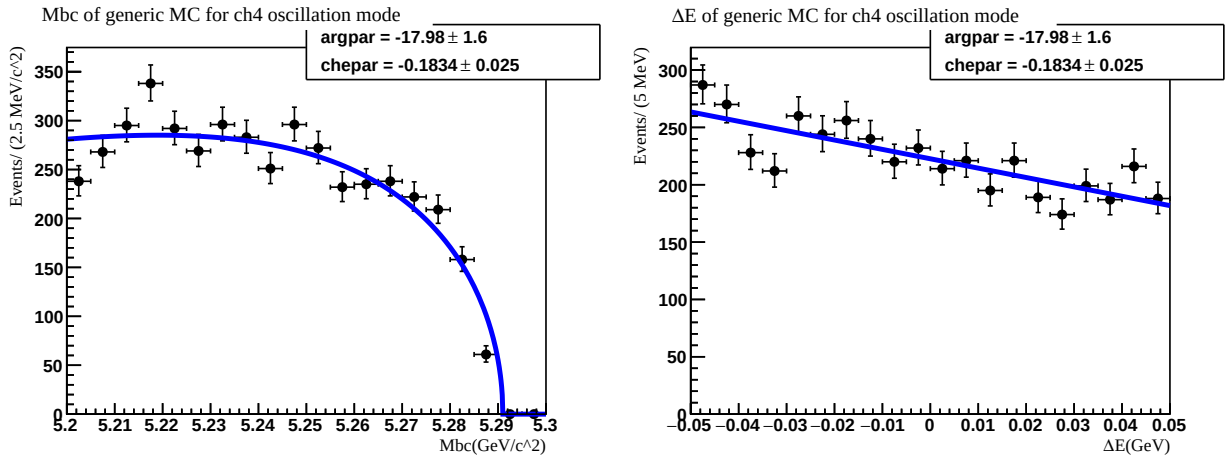


Figure 141: Projections of the 2D fit for ch4 BNV generic background MC samples.

B.5.11 BNV: $B^- \to \overline{\Xi}{}_c^0 \overline{\Lambda}{}_c^-$ with $\Xi_c^0 \to pK^-K^-\pi^+$ and $\overline{\Lambda}{}_c^- \to \overline{p}K_S^0$

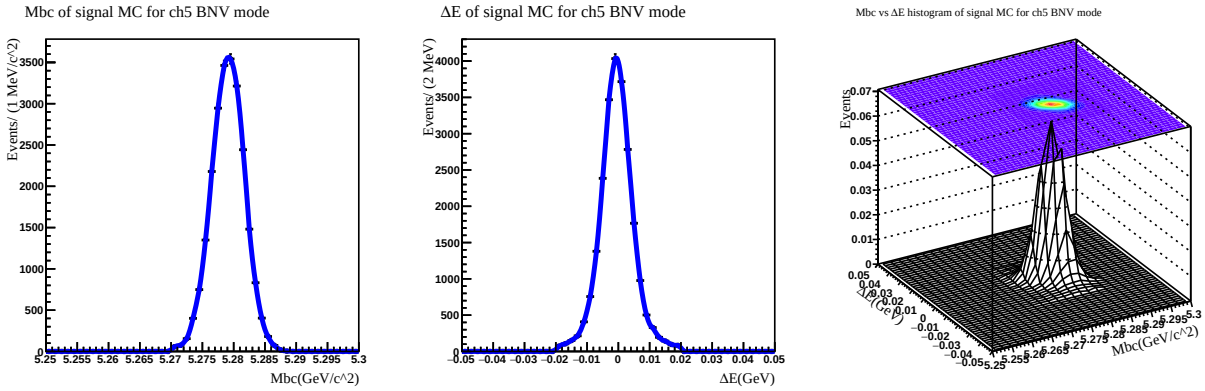


Figure 142: 2D smoothed histogram and projections of ch5 BNV signal MC samples.

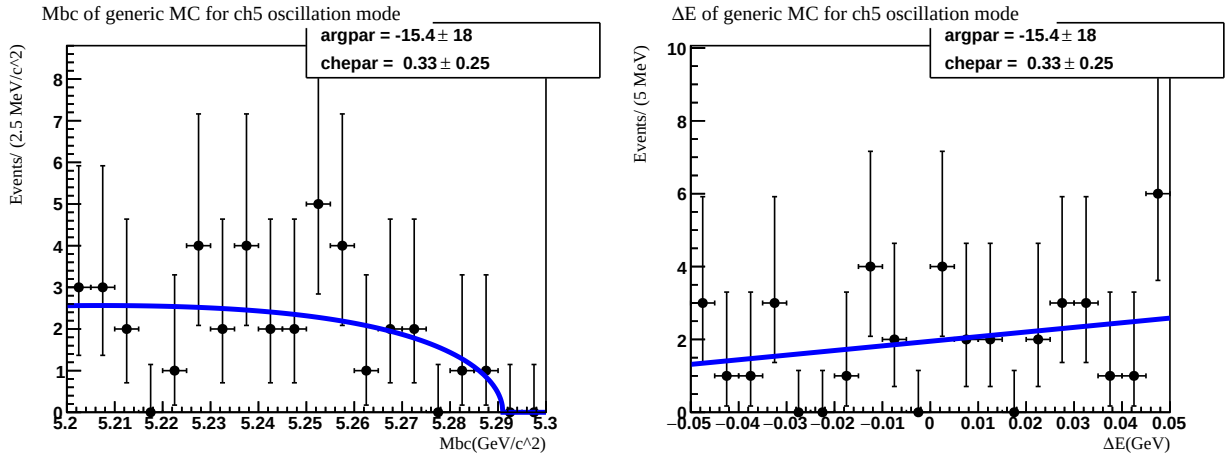


Figure 143: Projections of the 2D fit for ch5 BNV generic background MC samples.

B.5.12 BNV: $B^- \to \overline{\Xi}_c^0 \overline{\Lambda}_c^-$ with $\Xi_c^0 \to pK^-K^-\pi^+$ and $\overline{\Lambda}_c^- \to \overline{p}K^+\pi^-$

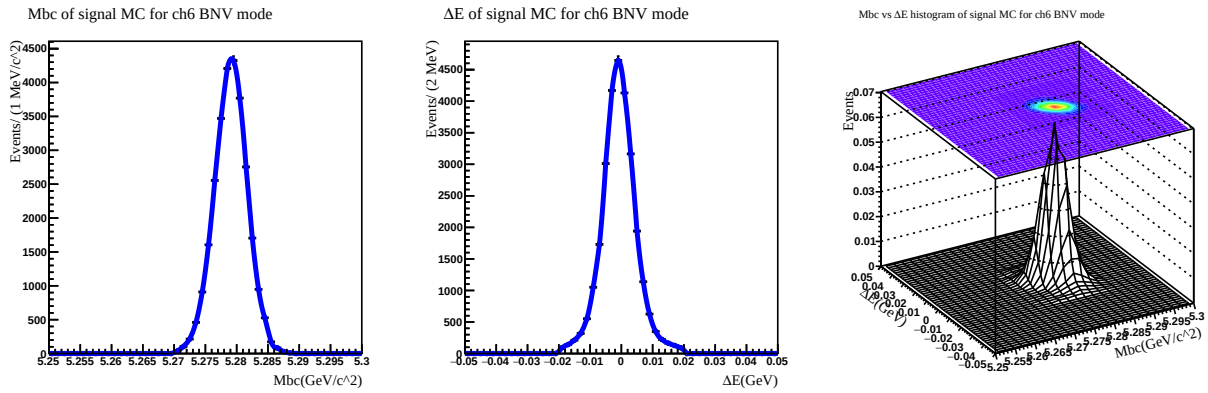


Figure 144: 2D smoothed histogram and projections of ch6 BNV signal MC samples.

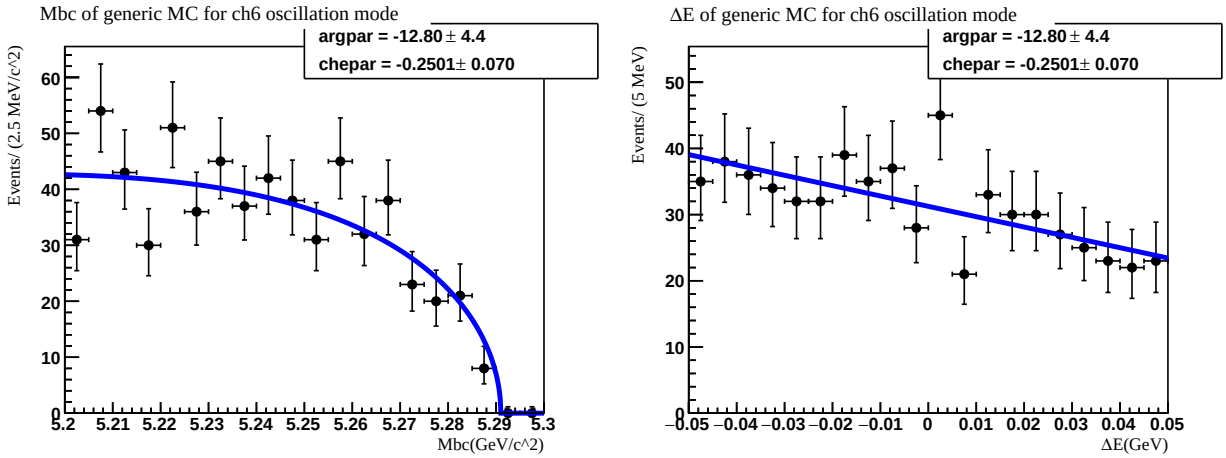


Figure 145: Projections of the 2D fit for ch6 BNV generic background MC samples.

B.6 DISCUSSION ABOUT THE SM BRANCHING FRACTIONS DIFFERENCE BETWEEN PREVIOUS MEASUREMENT AND OUR RESULT

As shown in Table. 20, there is a difference between SM branching fractions measured by us and previous Belle analysis. This is because our reconstruction workflow is different from the original analysis. The original analysis performed a vertex fit to the signal $B^{\pm}$ candidate, but we decided not to perform this vertex fit, because we found out that there is no additional background suppression gained but some efficiency loss needs to be included. The original analysis used this fit for best candidate selection, but we achieve the same goal

differently. Our vertex χ^2 cuts are different from those used in the original SM analysis: the previous analysis applied a very tight cut of $\chi^2 < 15$. It is interesting, however, that, as shown in Table. 29, applying such tight cut in our analysis does not make us more compatible with previous results. Our ΔE resolution is better (i.e., narrower) than in the previous analysis, which results in different signal PDFs. We think that there could be some other small differences (such as the already-mentioned best candidate selection). When the same data are analyzed correctly in such two different analyses, we think that $\pm 1\sigma$ statistical fluctuations could be tolerated (though are marginally acceptable).

Table. 29 shows a complete list of the results for SM branching fractions measured with different sets of cuts in our analysis and a comparison with the previous analysis.

	Ch1&2	Ch3&4	Ch5&6
No $p_\perp$ cut / $\chi^2 < 15$ / *Physical review letters*, 122, 082001(2019)			
Signal Yield	44.8 ± 7.3	24.1 ± 5.5	16.6 ± 5.4
BR	$(1.71 \pm 0.28) \times 10^{-5}$	$(1.11 \pm 0.25) \times 10^{-5}$	$(5.47 \pm 1.78) \times 10^{-6}$
No $p_\perp$ cut / $\chi^2 < 15$			
Signal Yield	41.5 ± 6.6	37.4 ± 6.8	16.4 ± 4.6
BR	$(1.67 \pm 0.26) \times 10^{-5}$	$(1.41 \pm 0.26) \times 10^{-5}$	$(5.04 \pm 1.41) \times 10^{-6}$
No $p_\perp$ cut / No χ^2 cut			
Signal Yield	53.5 ± 7.7	43.1 ± 9.4	19.7 ± 6.5
BR	$(1.76 \pm 0.25) \times 10^{-5}$	$(1.34 \pm 0.29) \times 10^{-5}$	$(4.68 \pm 1.55) \times 10^{-6}$
No $p_\perp$ cut / $\chi^2 < 100$			
Signal Yield	52.4 ± 7.4	42.1 ± 8.1	21.0 ± 5.8
BR	$(1.81 \pm 0.26) \times 10^{-5}$	$(1.39 \pm 0.27) \times 10^{-5}$	$(5.44 \pm 1.50) \times 10^{-6}$
60% efficiency $p_\perp$ cut / No χ^2 cut			
Signal Yield	41.1 ± 6.7	26.9 ± 7.2	15.5 ± 4.8
BR	$(1.86 \pm 0.30) \times 10^{-5}$	$(1.21 \pm 0.32) \times 10^{-5}$	$(7.04 \pm 2.18) \times 10^{-6}$
60% efficiency $p_\perp$ cut / $\chi^2 < 100$			
Signal Yield	40.2 ± 6.5	30.0 ± 6.6	15.4 ± 4.4
BR	$(1.91 \pm 0.31) \times 10^{-5}$	$(1.43 \pm 0.32) \times 10^{-5}$	$(7.65 \pm 2.19) \times 10^{-6}$
50 MeV/c $p_\perp$ cut / $\chi^2 < 100$			
Signal Yield	52.4 ± 7.4	42.1 ± 8.1	21.1 ± 5.8
BR	$(1.82 \pm 0.26) \times 10^{-5}$	$(1.39 \pm 0.27) \times 10^{-5}$	$(5.48 \pm 1.51) \times 10^{-6}$

Table 29: SM branching fractions measured with different sets of cuts.

B.7 DISCUSSION ABOUT LOG-LIKELIHOOD DISTRIBUTIONS FOR ENSEMBLE TESTS FOR LOW STATISTICS EXPERIMENTS

Fig. 35 and Fig. 36 show the log-likelihood distributions of the results of ensemble tests. These distributions indicate that the results are separated into several categories, each (as will soon become clear) represented by its own log-likelihood distribution. In this section we explain the origin of this effect in the results of ensemble tests and provide evidence that supports our explanation of the two bumps in any log-likelihood distribution for any fixed and small number of generated events in our fits. Then we use this information to explain multiple peaks in the log-likelihood distribution when the number of generated events is small and is allowed to fluctuate (according to Poisson statistics).

In our ensemble tests, according to the choice we make while using RooFit, the number of events generated for each toy MC ensemble follows Poisson distribution. If we fix the number of events generated (therefore not allowing Poisson fluctuations of this quantity), the log-likelihood distribution clearly demonstrates that the results separate into two distinct categories. It is easy to validate the following hypothesis: one distribution represents successful fits and the other distribution represents failed fits (which in this case, is loosely defined as $N_{sig}^{fit} < -8$, however, "failed" fits actually correspond to the parameter N_{sig}^{fit} being at the limit (-10 in our fits)). Fig. 146 shows the log-likelihood vs N_{sig}^{fit} distribution for 2D fits to ch1&ch2 BNV mode with 0 signal events and 27 background events. Projections of this distribution onto the log-likelihood axis are also shown. The distributions shown in this figure demonstrate that the log-likelihood distribution for failed fits is different from the log-likelihood distribution for successful fits. To remind you, in this context, fits are defined as failed when the parameter N_{sig}^{fit} is at the limit. Extending the range of the parameter N_{sig}^{fit} further left toward even more negative values extends the tail of the log-likelihood distribution for successful fits also increasing their fractional contribution to the ensemble. However, allowing N_{sig}^{fit} to be large and negative is unphysical and makes using the method of extended unbinned maximum likelihood invalid, as, strictly speaking, the extended likelihood function is no longer correctly defined in this region (i.e., for negative numbers of events). Below we show two more examples of the same bimodal behavior in the results of ensemble tests, which

is followed by a brief concluding discussion.

Fig. 147 and Fig. 148 show the log-likelihood distribution for 2D fit to ch1&ch2 BNV mode with 0 signal events and 22/32 background events, respectively. Again, no Poisson fluctuation is allowed. Both figures show a strong correlation between log-likelihood values and failed/successful fits.

Finally, when we allow Poisson fluctuations in the total number of generated events, we start observing Poisson-weighted ensembles of bimodal distributions (one for each total number of generated events), which, especially for (nominal) small numbers of generated events, exhibit a series of bumps in the log-likelihood distribution. Only when the total number of generated events becomes relatively large, these bumps merge together and form a relatively smooth log-likelihood distribution (also, as the number of generated events becomes large, the fraction of tests that fail becomes smaller, as we gradually enter the regime where the formalism of the extended maximum likelihood method works as expected).

We conclude that several categories of the results in log-likelihood distributions in Fig. 35 and Fig. 36 come from pairs of failed/successful groups of fits for individual numbers of events generated while following Poisson distribution. We also conclude that (obviously) extended maximum likelihood fitting is not necessarily the best analysis tool to apply to low-statistics samples.

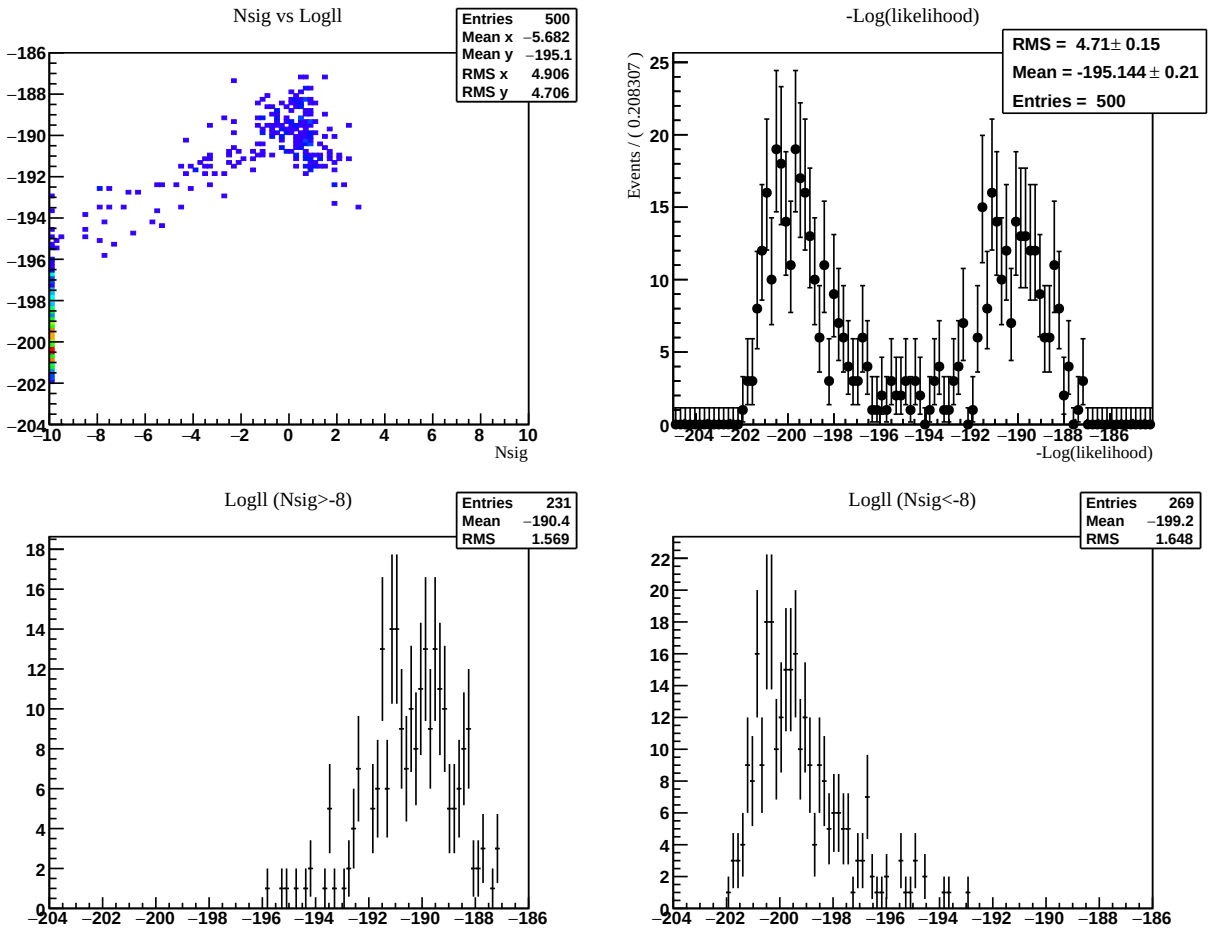


Figure 146: Ensemble test with 0 signal events and 27 background events (the total number of generated events is fixed at 27, i.e., no Poisson fluctuations are allowed) for ch1&ch2 BNV mode.

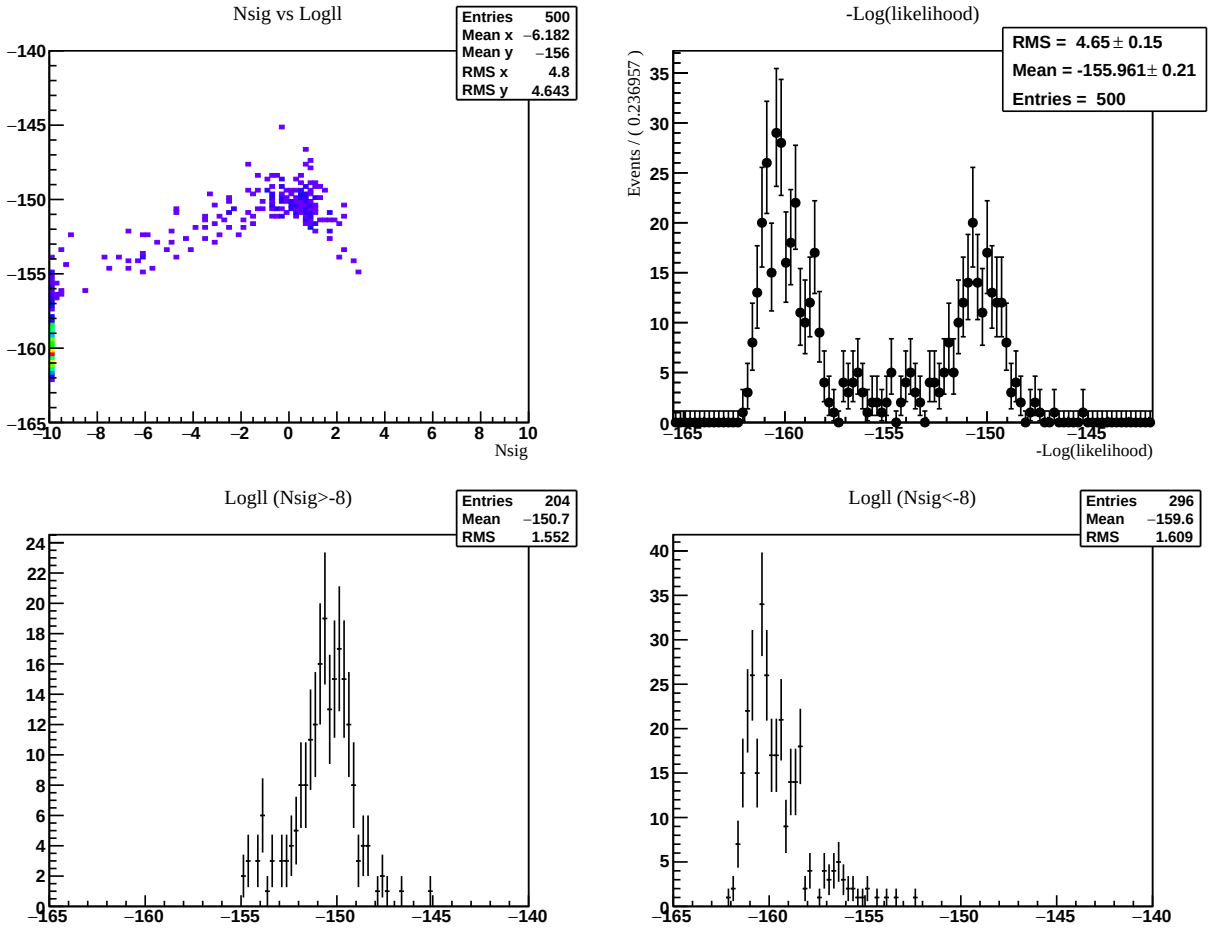


Figure 147: Ensemble test with 0 signal events and 22 background events (the total number of generated events is fixed at 22, i.e., no Poisson fluctuations are allowed) for ch1&ch2 BNV mode.

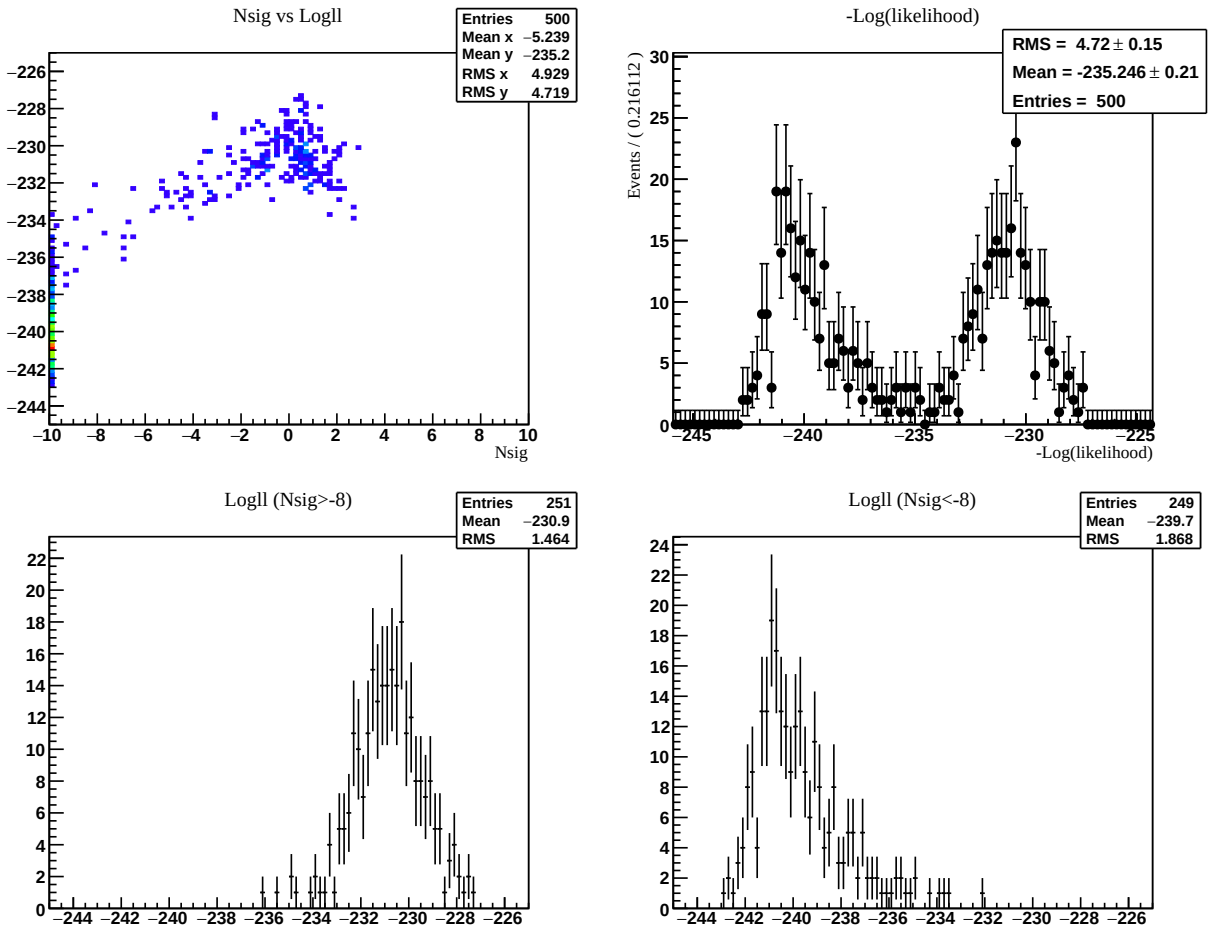


Figure 148: Ensemble test with 0 signal events and 32 background events (the total number of generated events is fixed at 32, i.e., no Poisson fluctuations are allowed) for ch1&ch2 BNV mode.

B.8 COMPARISON BETWEEN BLINDED DATA AND GENERIC MC

We unblind data in the region $M_{bc} < 5.26$ GeV/c^2, $|\Delta E| < 0.25$ GeV for both BNV and SM modes. In this section, we compare the numbers of events in this sideband between data and generic MC and show this comparison in Table 30. We also compare the actual distributions of M_{bc} and ΔE in data and generic MC in Figures 149, 150, 151, and 152. Comparison of p_T distributions for several final state particle candidates are shown in Figures 153 and 154. To present this comparison, we normalize the number of events in generic MC to the

number of events in data. All six streams of generic MC are used for this comparison.

Note that to compare generic MC and data we use a wider range of ΔE than we use in our fits to data ($5.2 < M_{bc} < 5.3$ GeV/c^2, $|\Delta E| < 0.05$ GeV). The effects observed on the left side in some of the $M_{\rm bc}$ and ΔE plots are not present for this narrower range of ΔE. We inspected these distributions before baseline event selection to verify that these effects are not due to statistical fluctuations but correctly represent the shape of combinatorial background in generic MC events and sideband data.

	SM mode data	SM mode generic MC	BNV mode data	BNV mode generic MC
Ch1	12	24	6	9
Ch2	291	396	128	126
Ch3	405	253	331	244
Ch4	5433	3877	4155	3096
Ch5	126	93	52	33
Ch6	1672	1189	656	496

Table 30: Numbers of events in data and generic MC sidebands ($M_{\rm bc} < 5.26$ GeV/c^2). Generic MC numbers are normalized to full data luminosity (by dividing by 6, as we use all 6 streams of generic MC).

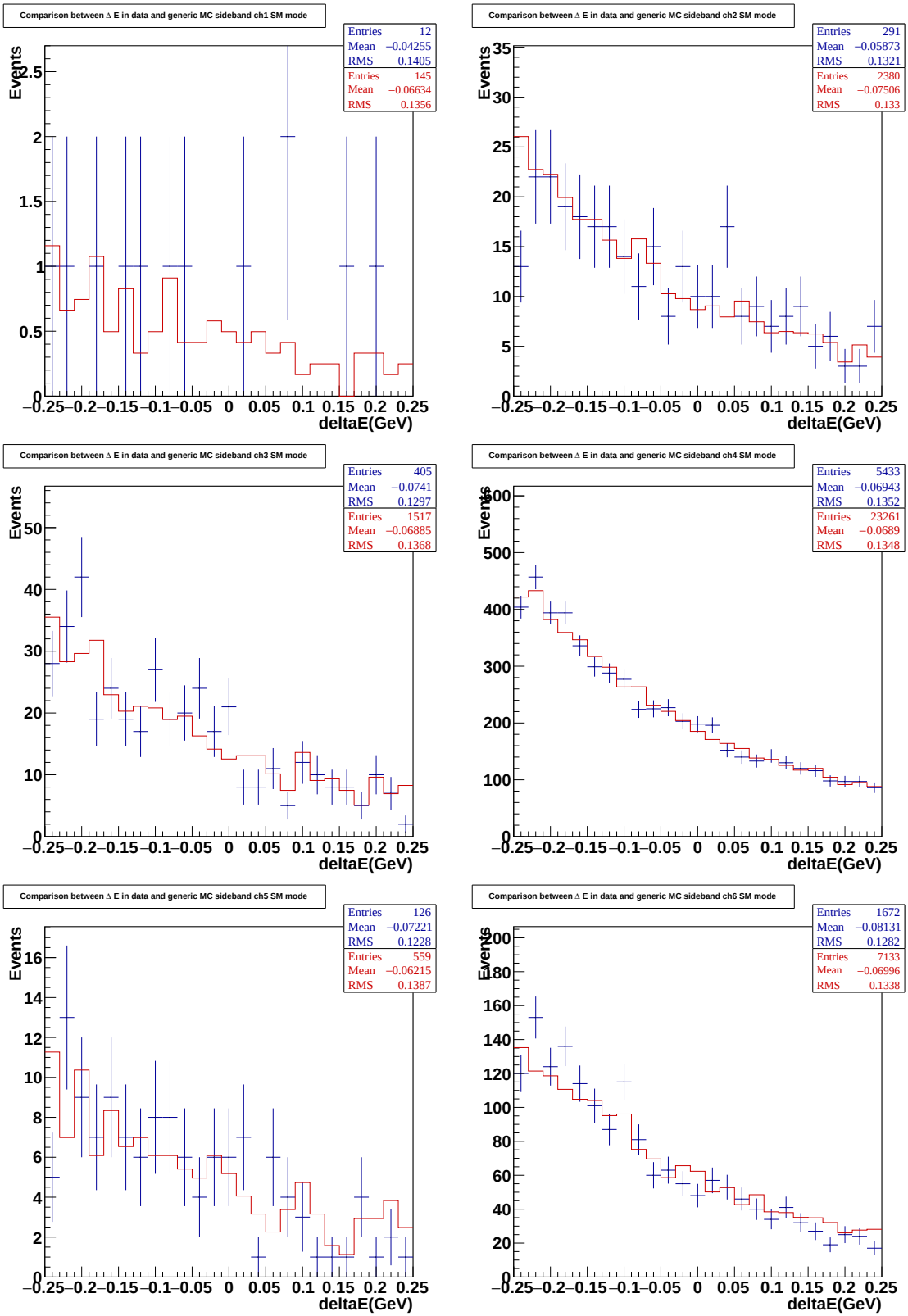


Figure 149: Comparison of ΔE distributions between data and generic MC in sideband for SM mode. Number of events in generic MC are normalized to the number of events in data.

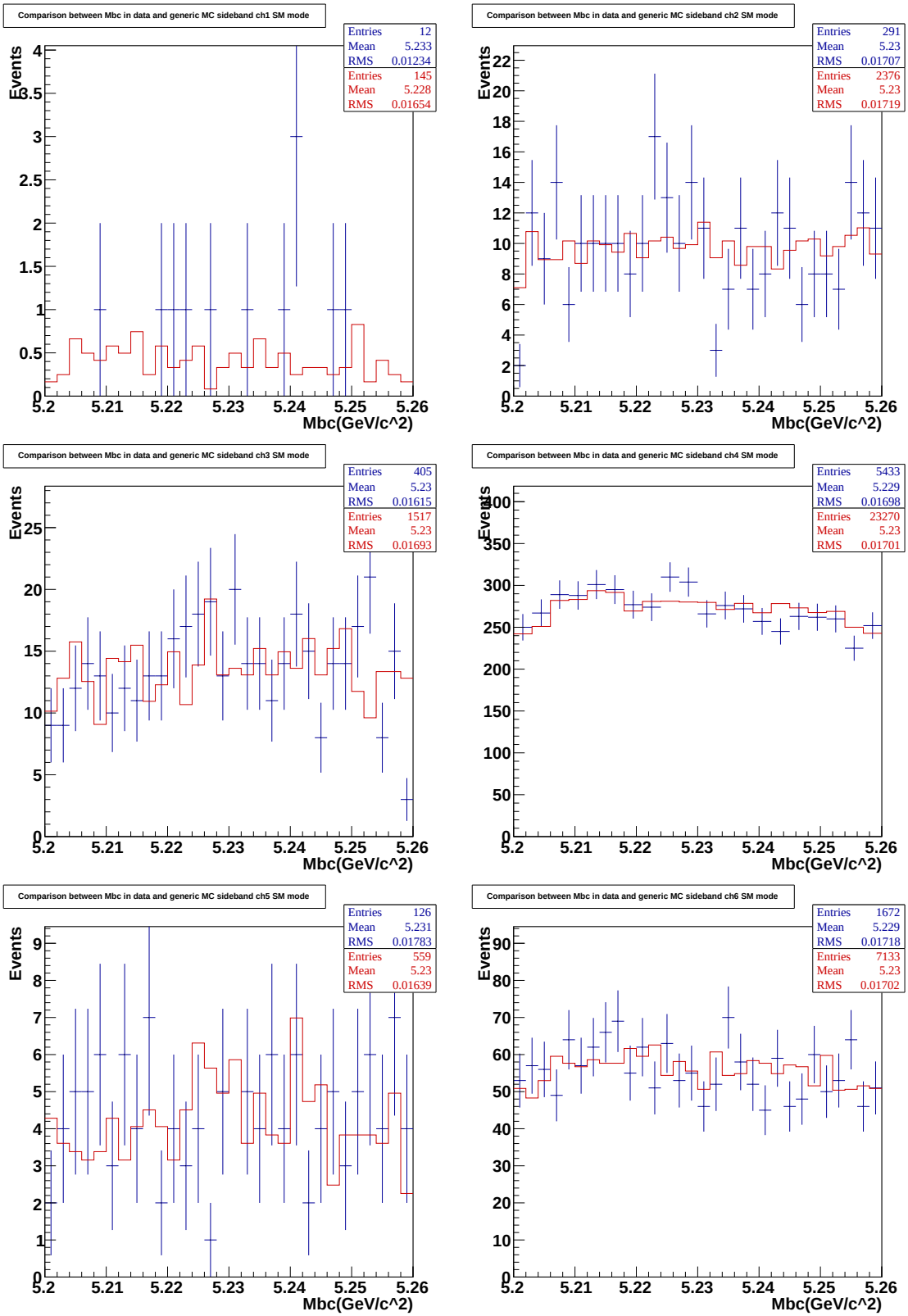


Figure 150: Comparison of M_{bc} distributions between data and generic MC in sideband for SM mode. Number of events in generic MC are normalized to the number of events in data.

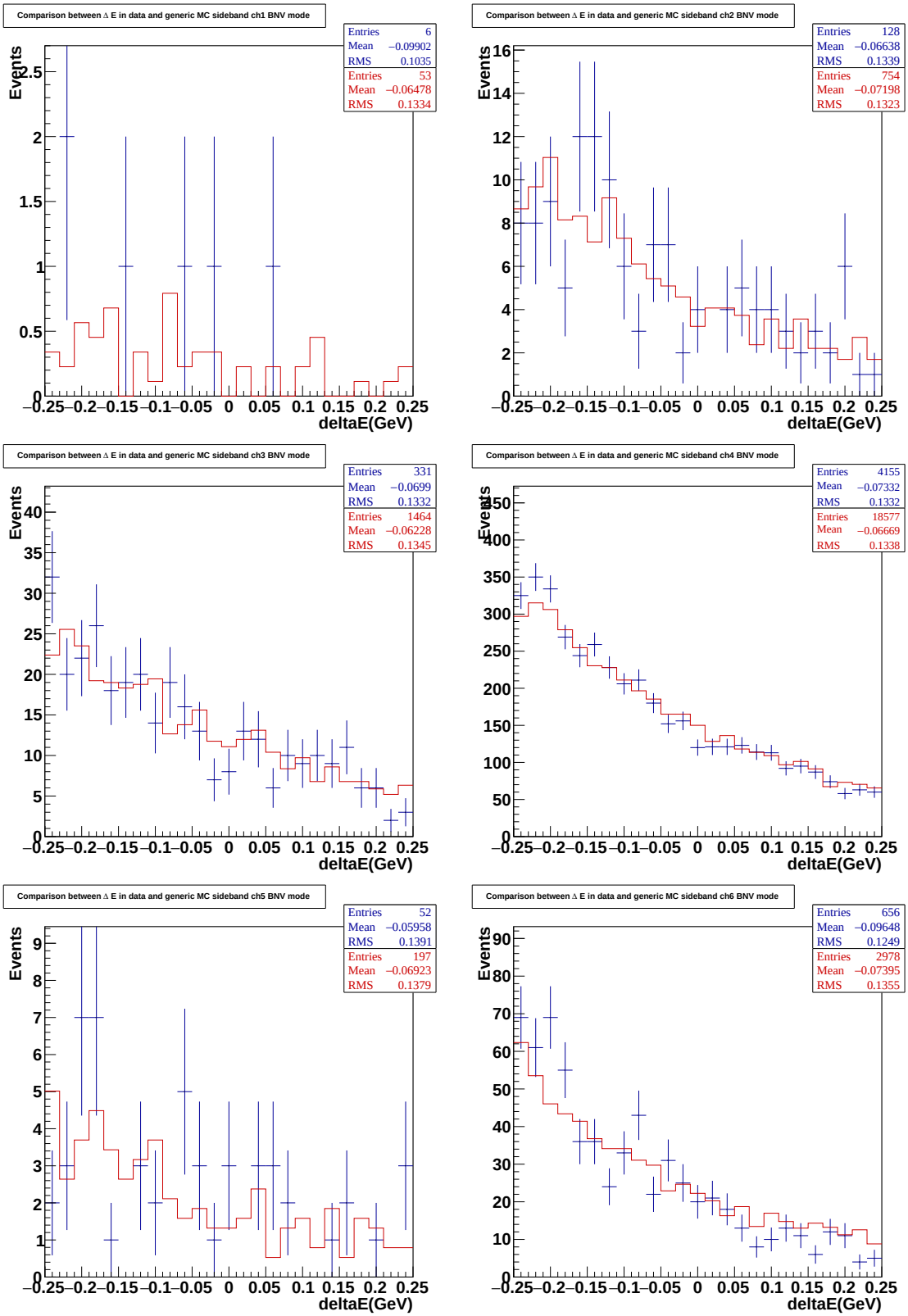


Figure 151: Comparison of ΔE distributions between data and generic MC in sideband for BNV mode. Number of events in generic MC are normalized to the number of events in data.

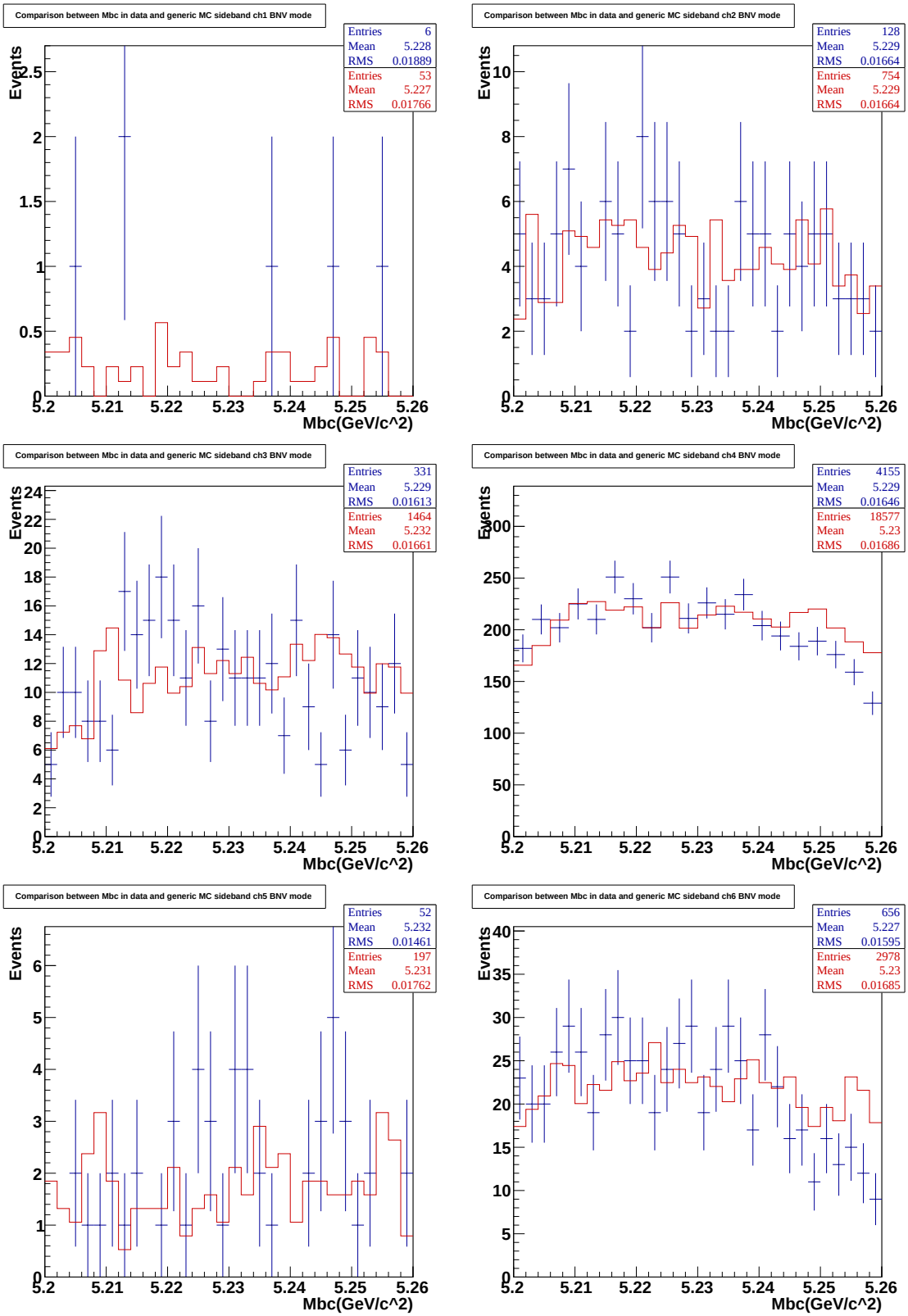


Figure 152: Comparison of M_{bc} distributions between data and generic MC in sideband for BNV mode. Number of events in generic MC are normalized to the number of events in data.

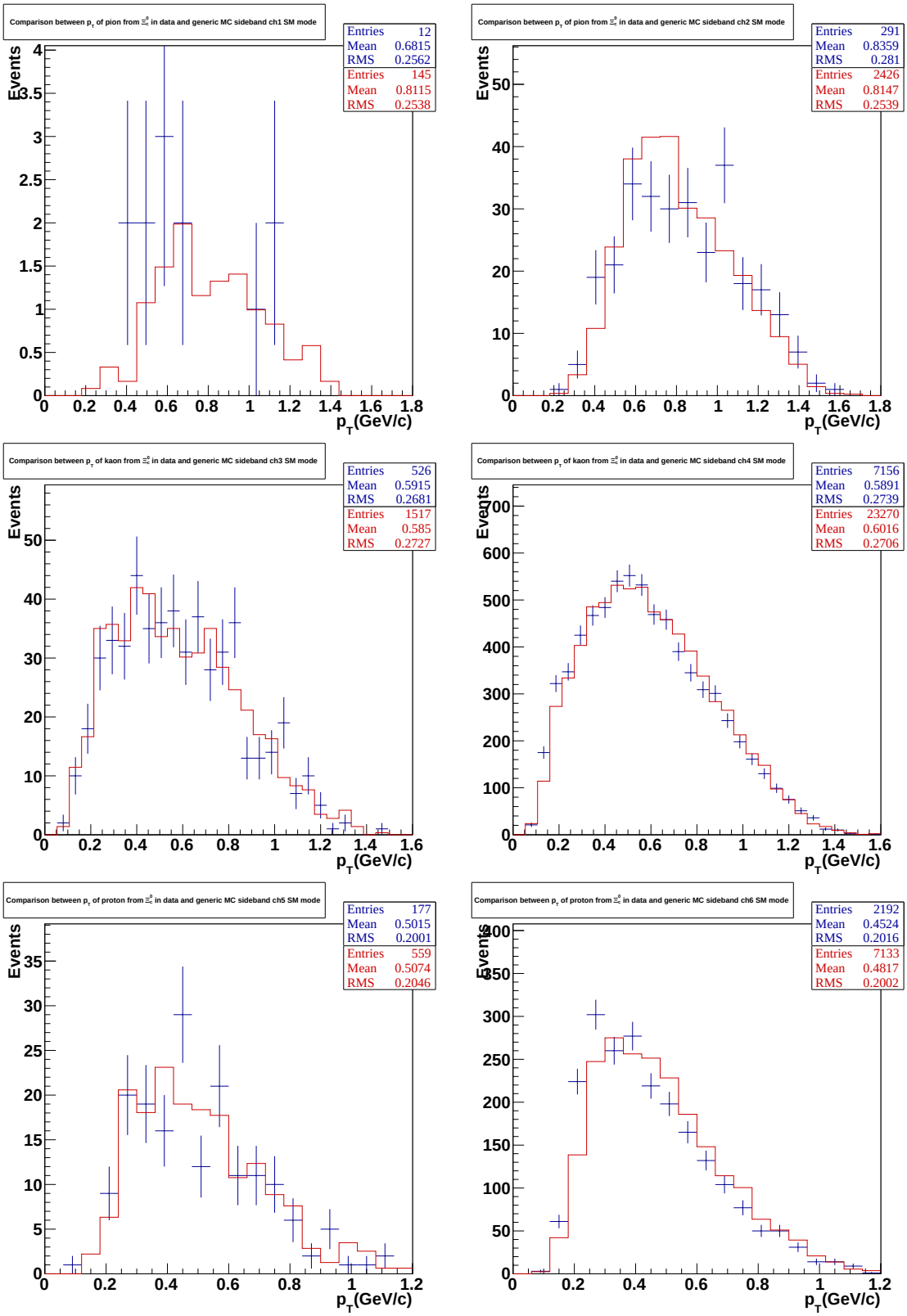


Figure 153: Comparison of p_T distributions between data and generic MC in sideband for SM mode. Number of events in generic MC are normalized to the number of events in data.

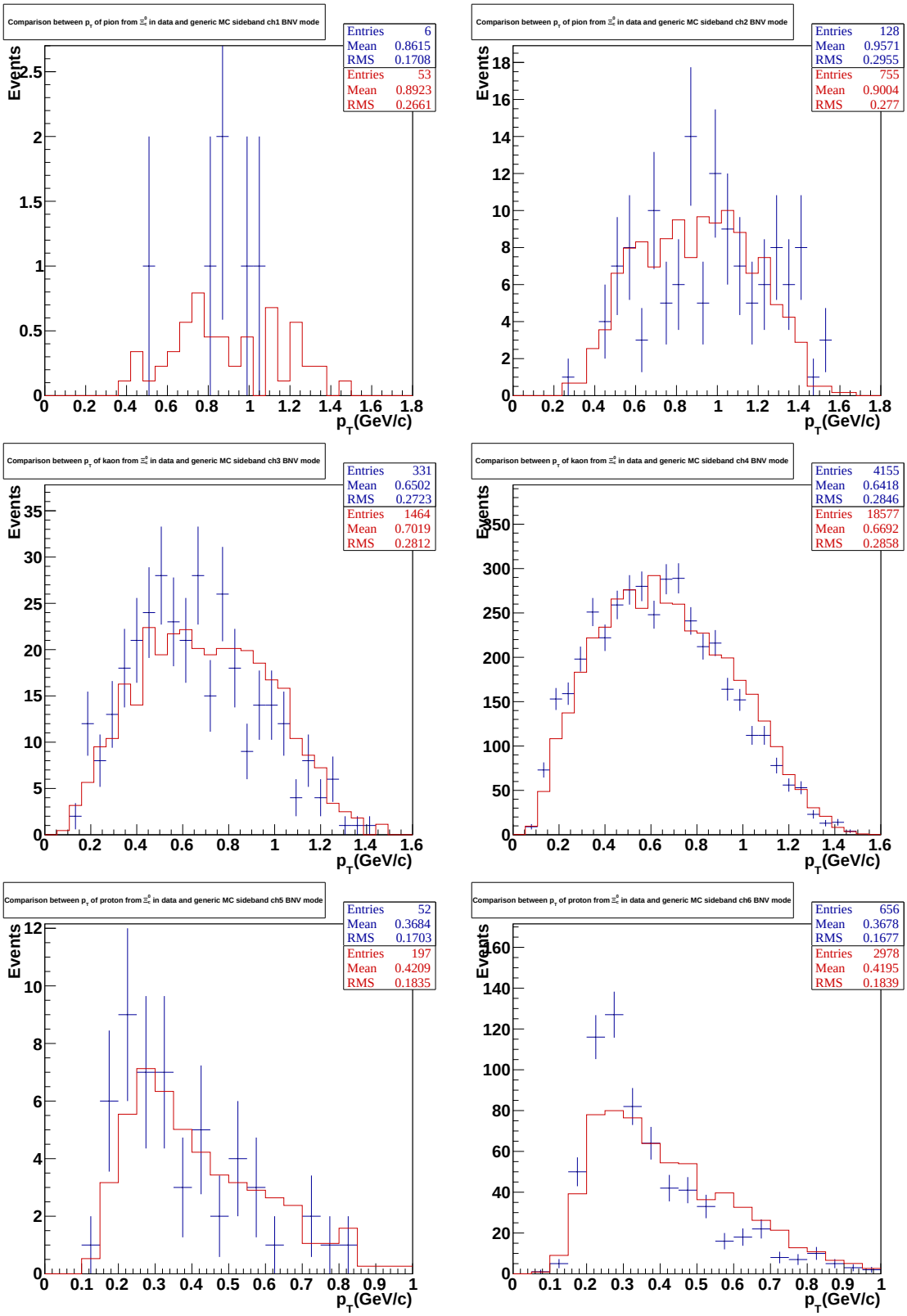


Figure 154: Comparison of p_T distributions between data and generic MC in sideband for BNV mode. Number of events in generic MC are normalized to the number of events in data.

B.9 CHANNELS EXCLUSION STUDIES

We considered excluding some channels (e.g., ch4/6) may improve sensitivity to BNV mode due to the large background. In this section, we demonstrate that the best sensitivity is achieved when all channels are included in the analysis.

Fig. 155 shows the comparison between 1D and 2D confidence belts with ch4 or ch6 excluded. As we can see, the best upper limit comes from all channels included.

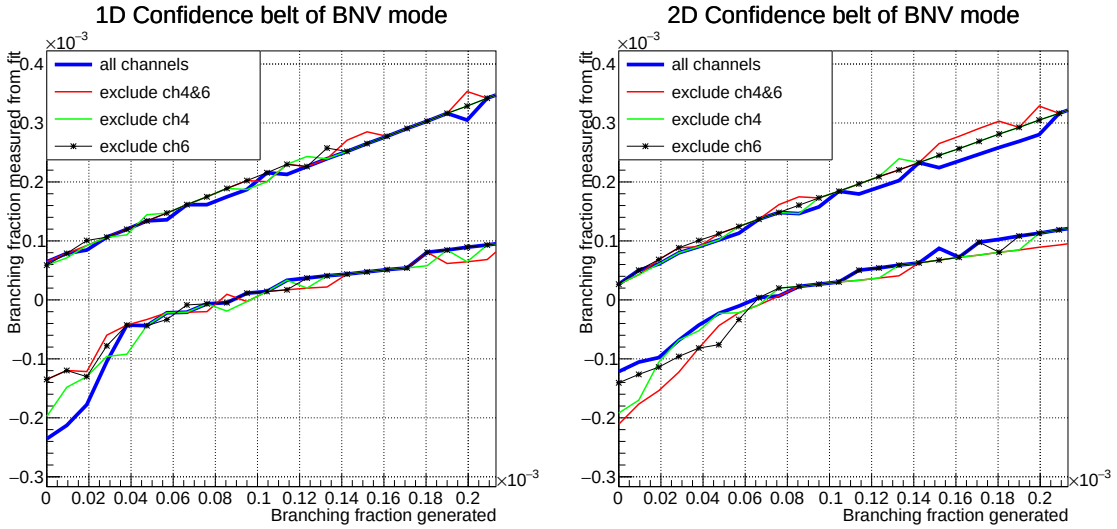


Figure 155: Comparison between 1D and 2D confidence belts for different scenarios.

Fig. 156 to 159 shows the 1D and 2D zero signal events hypothesis pseudo-experiments with ch4 or ch6 excluded. Table 31 summarizes the fit failure rate for the zero signal events hypothesis with ch4 or ch6 excluded. The fit failure rate is defined as the fraction of toy MC experiments in which the measured branching fraction is below -0.3×10^{-3}. The probability of fit failure increases when ch4 or ch6 is excluded, which is due to low background statistics in the signal region.

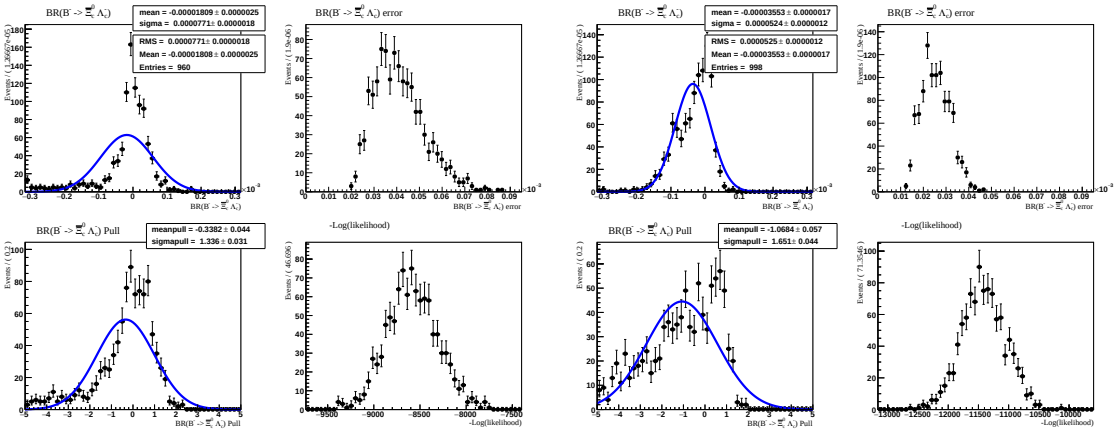


Figure 156: Zero signal events hypothesis 1D/2D pseudo-experiments with all channels included.

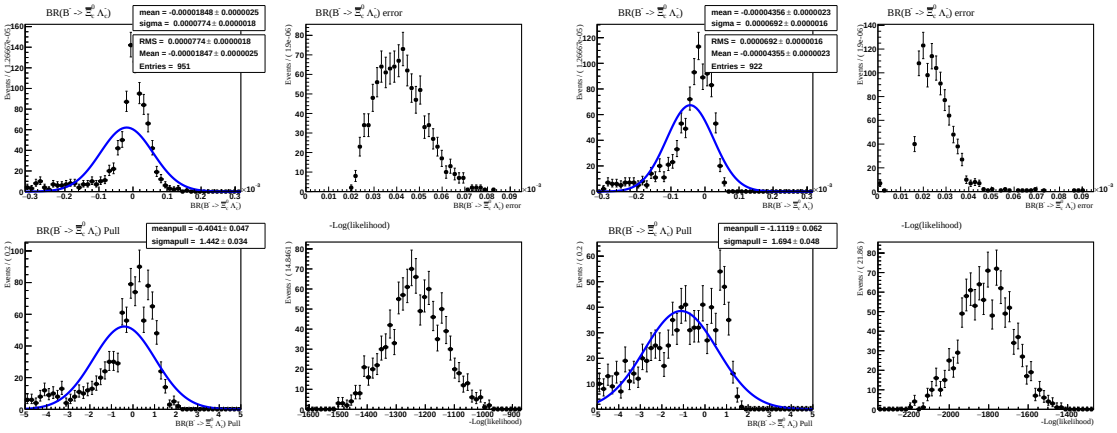


Figure 157: Zero signal events hypothesis 1D/2D pseudo-experiments with ch4 excluded.

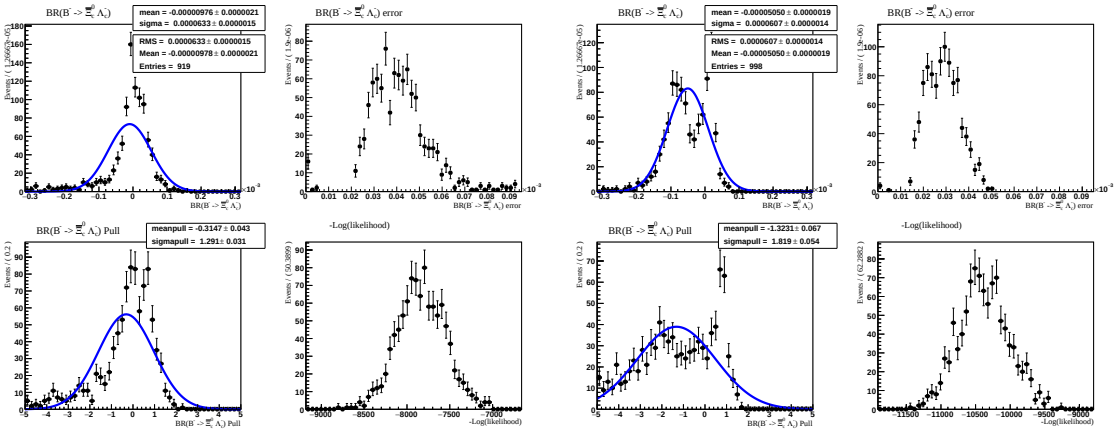


Figure 158: Zero signal events hypothesis 1D/2D pseudo-experiments with ch6 excluded.

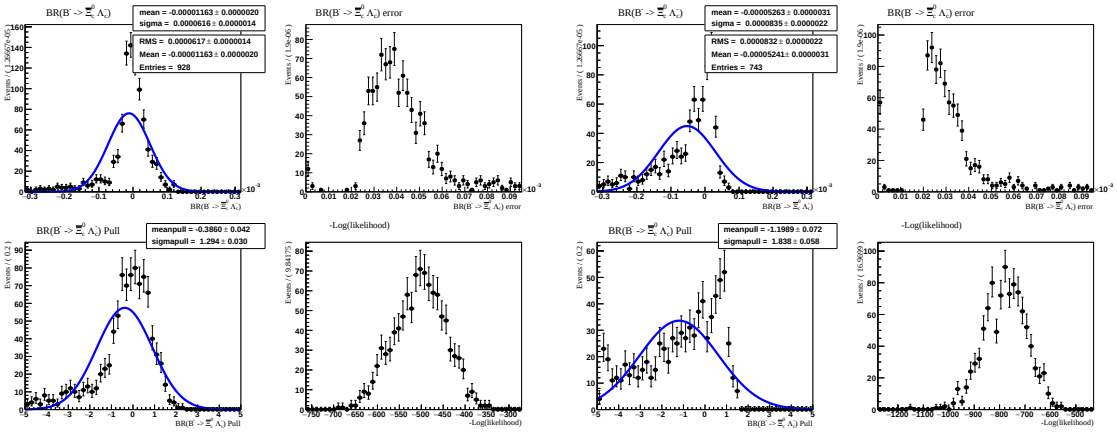


Figure 159: Zero signal events hypothesis 1D/2D pseudo-experiments with ch4 and ch6 excluded.

Zero signal events hypothesis	1D fit failure rate	2D fit failure rate
All channels	4.0%	0.2%
Exclude ch4	4.9%	7.8%
Exclude ch6	8.1%	0.2%
Exclude ch4&6	7.9%	25.7%

Table 31: 1D and 2D fit failure rate for zero signal events hypothesis with ch4 or ch6 excluded.

B.10 UPPER LIMIT ESTIMATE USING FELDMAN-COUSINS APPROACH

Table 32 shows the number of background events of SM/BNV mode in the full fit region ($5.2 < M_{\rm bc} < 5.3$ GeV/c^2, $|\Delta E| < 0.05$ GeV) and signal region ($5.27 < M_{\rm bc} < 5.29$ GeV/c^2, $|\Delta E| < 0.02$ GeV).

bkg	SM full	SM sr	BNV full	BNV sr
ch1	3	0.4	1	0.0
ch2	71	5.6	28	2.8
ch3	97	5.9	81	4.6
ch4	1257	73.6	979	58.3
ch5	30	1.6	10	0.5
ch6	395	23.0	135	6.0

Table 32: Number of background events of SM/BNV mode in the full fit region and signal region, estimated by generic MC and scaled by sideband data.

If we exclude ch4, the number of background events in the signal region is 14. If we exclude ch4&6, the number of background events in the signal region is 8. If we exclude ch3&4&6, the number of background events in the signal region is 3.3.

It is possible to use the Feldman-Cousins confidence belt for a Poisson process with background [58] which is used to obtain 90% C.L. intervals for the Poisson signal mean μ, for total events observed n_0, for known mean background b ranging from 0 to 15 events.

Table 33 shows the BNV branching fraction upper limit estimate using the Feldman-Cousins confidence belt.

The best result comes from ch4 excluded. If we measured 14 events in the signal region in data while the expected background is 14, the upper limit on the BNV branching fraction is 1.11×10^{-4}.

To compare with the conventional approach, if we assume the measured branching fraction of 0, the upper limit of 6.5×10^{-5} would be set at 95% CL (see Fig. 56), which is slightly better than if we used the Feldman-Cousins method.

Exclude ch4 (bkg = 14)			
Entries in signal region	10	14	18
90% C.L. intervals for Poisson signal mean	[0,3.27]	[0,7.5]	[0,12.16]
Upper limit for branching fraction	4.83×10^{-5}	1.11×10^{-4}	1.80×10^{-4}
Exclude ch4&6 (bkg = 8)			
Entries in signal region	4	8	12
90% C.L. intervals for Poisson signal mean	[0,1.98]	[0,5.99]	[0,11]
Upper limit for branching fraction	3.84×10^{-5}	1.16×10^{-4}	2.13×10^{-4}
Exclude ch3&4&6 (bkg = 3.3)			
Entries in signal region	0	4	8
90% C.L. intervals for Poisson signal mean	[0,1.06]	[0,5.10]	[0,10.49]
Upper limit for branching fraction	2.42×10^{-5}	1.16×10^{-4}	2.39×10^{-4}

Table 33: BNV branching fraction upper limit estimate using Feldman-Cousins confidence belt [58].

B.11 MEASURED BNV MODE BRANCHING FRACTION: A DISCUSSION

The measured BNV mode branching fraction of $(-9.65 \pm 2.90) \times 10^{-5}$ is more than three standard deviations below zero, however, according to our toy MC studies, the probability of such a negative (or even more negative) result is 18% (this is to be compared with 1.5% which

would be for the one-sided probability interval assuming Gaussian statistics). Fig. 160 shows the results of our toy MC studies (i.e., ensemble test) with zero signal events hypothesis for simultaneous fit to BNV mode for all channels. The distribution of measured branching fraction for BNV mode has a long negative tail, and the mean value is about -3.7×10^{-5}.

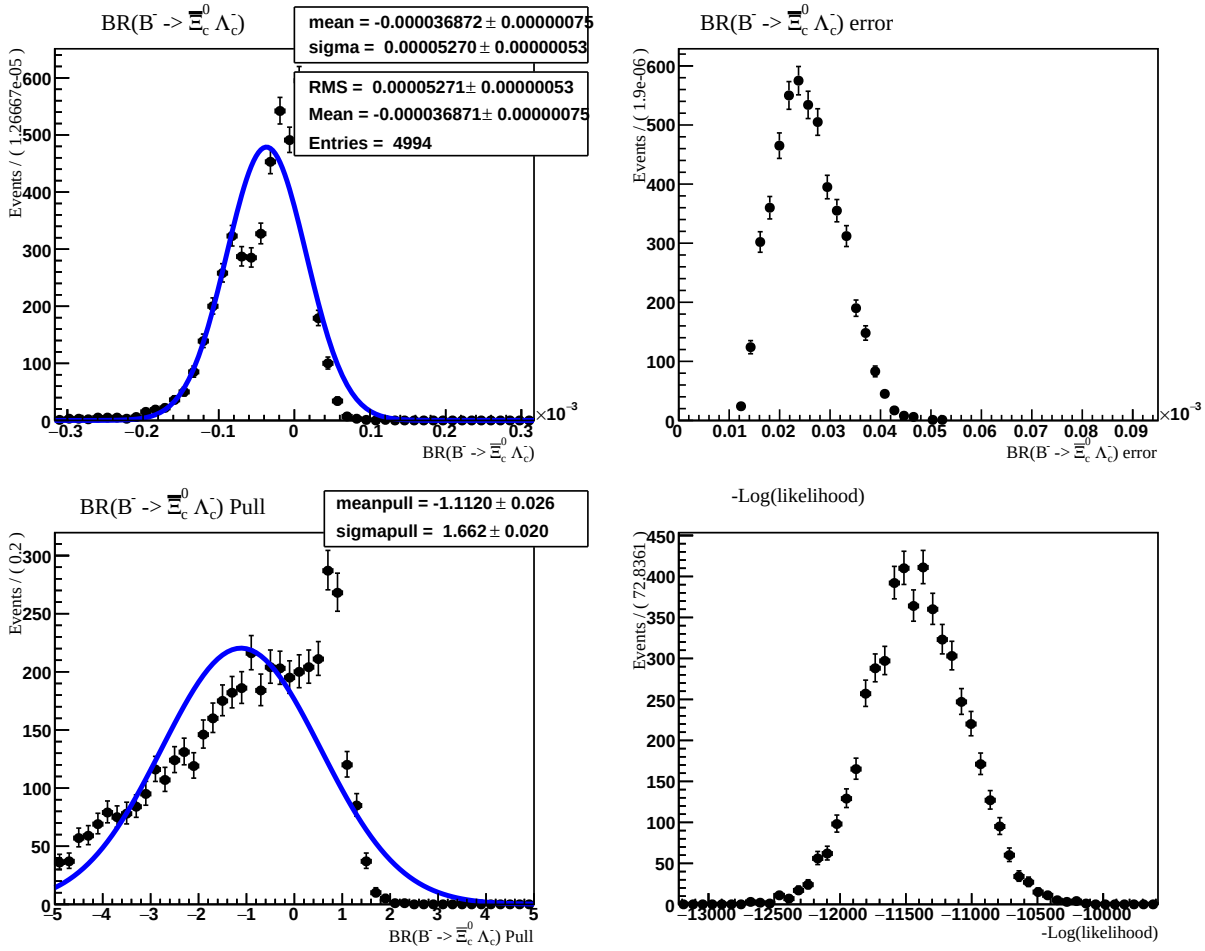


Figure 160: Zero signal events hypothesis ensemble test for simultaneous fit to BNV mode of all channels.

BIBLIOGRAPHY

[1] N. Aghanim *et al.* (Planck Collaboration). Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[2] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Sov. Phys. Usp.*, 34(5):392–393, 1991.

[3] P.B. Arnold and L.D. McLerran. Sphalerons, Small Fluctuations and Baryon Number Violation in Electroweak Theory. *Phys. Rev. D*, 36:581, 1987.

[4] N. Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963.

[5] S. Schrenk *et al.* (BaBar Collaboration). The unitarity triangle angle gamma (phi(3)) at BaBar and Belle. *Nucl. Phys. B Proc. Suppl.*, 142:279–286, 2005.

[6] P. Huet and E. Sather. Electroweak baryogenesis and standard model CP violation. *Phys. Rev. D*, 51:379–394, 1995.

[7] M.B. Gavela *et al.* . Standard model CP violation and baryon asymmetry. Part 2: Finite temperature. *Nucl. Phys. B*, 430:382–426, 1994.

[8] K. Kajantie, M. Laine, K. Rummukainen, and Mikhail E. Shaposhnikov. Is there a hot electroweak phase transition at $m_H \gtrsim m_W$? *Phys. Rev. Lett.*, 77:2887–2890, 1996.

[9] F. Wilczek and A. Zee. Operator Analysis of Nucleon Decay. *Phys. Rev. Lett.*, 43:1571–1573, 1979.

[10] S. Weinberg. Baryon and Lepton Nonconserving Processes. *Phys. Rev. Lett.*, 43:1566–1570, 1979.

[11] S. Fukuda *et al.* (Super-Kamiokande Collaboration). The Super-Kamiokande detector. *Nucl. Instrum. Meth. A*, 501:418–462, 2003.

[12] K. Abe *et al.* . Calibration of the Super-Kamiokande Detector. *Nucl. Instrum. Meth. A*, 737:253–272, 2014.

[13] A. Takenaka *et al.* (Super-Kamiokande Collaboration). Search for proton decay via $p \to e^+\pi^0$ and $p \to \mu^+\pi^0$ with an enlarged fiducial volume in Super-Kamiokande I-IV. *Phys. Rev. D*, 102(11):112011, 2020.

[14] T. Mori W. J. Marciano and J. M. Roney. Charged Lepton Flavor Violation Experiments. *Ann. Rev. Nucl. Part. Sci.*, 58:315–341, 2008.

[15] R. Aaij *et al.* (LHCb collaboration). Searches for violation of lepton flavour and baryon number in tau lepton decays at LHCb. *Phys. Lett. B*, 724:36–45, 2013.

[16] D. Sahoo *et al.* (Belle collaboration). Search for lepton-number- and baryon-number-violating tau decays at Belle. *Phys. Rev. D*, 102:111101, 2020.

[17] II *et al.* D. G. Phillips. Neutron-antineutron oscillations: theoretical status and experimental prospects. *Phys. Rept.*, 612:1–45, 2016.

[18] M. Baldo-Ceolin *et al.* . A New experimental limit on neutron - anti-neutron oscillations. *Z. Phys. C*, 63:409–416, 1994.

[19] K. Abe *et al.* (Super-Kamiokande collaboration). The Search for $n - \bar{n}$ oscillation in Super-Kamiokande I. *Phys. Rev. D*, 91:072006, 2015.

[20] Y. Liu *et al.* . Test baryon antibaryon oscillation in collider experiments. *J. Phys. Conf. Ser.*, 738(1):012043, 2016.

[21] K. Aitken, D. McKeen, T. Neder, and A. E. Nelson. Baryogenesis from oscillations of charmed or beautiful baryons. *Phys. Rev. D*, 96(7):075009, 2017.

[22] R. Aaij *et al.* (LHCb Collaboration). Search for baryon-number violating Ξ_b^0 oscillations. *Phys. Rev. Lett.*, 119(18):181807, 2017.

[23] P. del Amo Sanchez *et al.* (BaBar collaboration). Searches for the baryon- and lepton-number violating decays $B^0 \to \Lambda_c^+\ell^-$, $B^- \to \Lambda\ell^-$, and $B^- \to \bar{\Lambda}\ell^-$. *Phys. Rev. D*, 83:091101, 2011.

[24] B. Aubert *et al.* (BaBar collaboration). The BaBar detector. *Nucl. Instrum. Meth. A*, 479:1–116, 2002.

[25] M. Ablikim *et al.* (BESIII collaboration). Search for baryon- and lepton-number violating decays $D^0 \to \overline{p}e^+$ and $D^0 \to pe^-$. *Phys. Rev. D*, 105(3):032006, 2022.

[26] M. Ablikim *et al.* (BESIII collaboration). Design and Construction of the BESIII Detector. *Nucl. Instrum. Meth. A*, 614:345–399, 2010.

[27] M. E. McCracken *et al.* (CLAS Collaboration) . Search for baryon-number and lepton-number violating decays of Λ hyperons using the CLAS detector at Jefferson Laboratory. *Phys. Rev. D*, 92(7):072002, 2015.

[28] B. A. Mecking *et al.* . The CEBAF Large Acceptance Spectrometer (CLAS). *Nucl. Instrum. Meth. A*, 503:513–553, 2003.

[29] A. Lechner. Particle interactions with matter. *CERN Yellow Rep. School Proc.*, 5:47, 2018.

[30] A. Abashian *et al.* (Belle Collaboration). The Belle detector. *Nucl. Instrum. Meth. A*, 479:117–232, 2002.

[31] S. Kurokawa and E. Kikutani. Overview of the KEKB accelerators. *Nucl. Instrum. Meth. A*, 499:1–7, 2003.

[32] J. Brodzicka *et al.* (Belle Collaboration). Physics Achievements from the Belle Experiment. *PTEP*, 2012:04D001, 2012.

[33] Y. B. Li *et al.* (Belle Collaboration). First measurements of absolute branching fractions of the Ξ_c^0 baryon at Belle. *Phys. Rev. Lett.*, 122(8):082001, 2019.

[34] R. Godang *et al.* (CLEO Collaboration). Search for baryon and lepton number violating decays of the tau lepton. *Phys. Rev. D*, 59:091303, 1999.

[35] D. Sahoo *et al.* (Belle Collaboration). Search for lepton-number- and baryon-number-violating tau decays at Belle. *Phys. Rev. D*, 102:111101, 2020.

[36] R. M. Bionta *et al.* . The search for proton decay. *AIP Conf. Proc.*, 123:321–336, 1984.

[37] II *et al.* D. G. Phillips. Neutron-antineutron oscillations: theoretical status and experimental prospects. *Phys. Rept.*, 612:1–45, 2016.

[38] R. L. Workman *et al.* (Particle Data Group). Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[39] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152–155, 2001.

[40] S. Agostinelli *et al.* . GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[41] M. Feindt and U. Kerzel. The NeuroBayes neural network package. *Nucl. Instrum. Meth. A*, 559:190–194, 2006.

[42] H. Nakano. *Search for new physics by a time-dependent CP violation analysis of the decay $B \to K_S\eta\gamma$ using the Belle detector*. PhD thesis, Tohoku U., 2015.

[43] S. Nishida (Belle Collaboration). Study of kaon and pion identification using inclusive D^* sample. *Belle Note 779*.

[44] B. Bhuyan (Belle Collaboration). High P_T tracking efficiency using partially reconstructed D^* decays. *Belle Note 1165*.

[45] N. Dash, S. Bahinipati, V. Bhardwaj and K. Trabelsi (Belle Collaboration). Systematic uncertainty for the K^0_S reconstruction efficiency. *Belle Note 1472.*

[46] Y. J. Lee (Belle Collaboration). Lambda selection at Belle. *Belle Note 684.*

[47] L. Li, Z. Zhang and W. Yan (Belle Collaboration). $D^0 - \overline{D}^0$ mixing with time-dependent Dalitz analysis in wrong-sign decay $D^0 \to K^+\pi^-\pi^0$. *Belle Note 1381.*

[48] K. Inami. Development of a TOP counter for the super B factory. *Nucl. Instrum. Meth. A*, 595:96–99, 2008.

[49] J. Fast (Belle II Barrel PID group). The Belle II imaging Time-of-Propagation (iTOP) detector. *Nucl. Instrum. Meth. A*, 876:145–148, 2017.

[50] M. Bessner (Belle II iTOP group). Performance of the Belle II imaging Time-Of-Propagation (iTOP) detector in first collisions. *Nucl. Instrum. Meth. A*, 958:162318, 2020.

[51] D. Kotchetkov *et al.* . Front-end electronic readout system for the Belle II imaging Time-Of-Propagation detector. *Nucl. Instrum. Meth. A*, 941:162342, 2019.

[52] G. Varner *et al.* . The large analog bandwidth recorder and digitizer with ordered readout (LABRADOR) ASIC. *Nucl. Instrum. Meth. A*, 583:447–460, 2007.

[53] S. Yamada *et al.* . Data acquisition system for the Belle II experiment. In *19th Real Time Conference*, 2014.

[54] Q. D. Zhou *et al.* . PCI-Express based high-speed readout for the Belle II DAQ upgrade. *IEEE Trans. Nucl. Sci.*, 68(8):1818–1825, 2021.

[55] L. Macchiarulo *et al.* . A probability-optimized fast timing trigger for the Belle II time of propagation detector. In *2010 IEEE Nuclear Science Symposium, Medical Imaging Conference, and 17th Room Temperature Semiconductor Detectors Workshop*, pages 630–635, 2010.

[56] H. C. Lin S. M. Liu and C. H. Wang. The implementation of high-speed data transmission for 8B/10B protocol on FPGAs. In *2017 International Conference on Applied System Innovation (ICASI)*, page 810. IEEE, 2017.

[57] M. Nakao *et al.* . Performance of the unified readout system of Belle II. *IEEE Trans. Nucl. Sci.*, 68(8):1826–1832, 2021.

[58] G. J. Feldman and R. D. Cousins. A unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57:3873–3889, 1998.