ABSTRACT

Title of Dissertation: A SIMULTANEOUS MEASUREMENT OF THE BRANCHING FRACTIONS OF TEN *B* TO DOUBLE CHARM DECAYS

Chung Khim Lae, Doctor of Philosophy, 2006

Directed By: Associate Professor Douglas Roberts
Department of Physics

This dissertation presents a simultaneous measurement of the branching fractions of ten $B \to D^{(*)}\bar{D}^{(*)}$ decays. The measurements are derived from a sample of 2.32×10^8 $B\bar{B}$ pairs collected by the BABAR detector at the PEP-II *B* Factory located at Stanford Linear Accelerator Center. The branching fractions ($\times 10^{-4}$) are:

- $-0.10 \pm 0.44 \pm 0.15$ (< 0.59) for $B^0 \to D^0\bar{D}^0$
- $1.01 \pm 1.07 \pm 0.35$ (< 2.92) for $B^0 \to D^{*0}\bar{D}^0$
- $-1.31 \pm 1.05 \pm 0.41$ (< 0.92) for $B^0 \to D^{*0}\bar{D}^{*0}$
- $2.81 \pm 0.43 \pm 0.45$ for $B^0 \to D^+D^-$
- $5.72 \pm 0.64 \pm 0.71$ for $B^0 \to D^{*+}D^-$
- $8.11 \pm 0.57 \pm 0.97$ for $B^0 \to D^{*+}D^{*-}$
- $3.76 \pm 0.57 \pm 0.45$ for $B^- \to D^-D^0$
- $3.56 \pm 0.52 \pm 0.39$ for $B^- \to D^{*-}D^0$
- $6.30 \pm 1.32 \pm 0.93$ for $B^- \to D^-D^{*0}$
- $8.14 \pm 1.17 \pm 1.11$ for $B^- \to D^{*-}D^{*0}$

The first uncertainty is statistical while the second is systematic. The number in parentheses is the 90% upper limit using the Feldman-Cousins method with systematic uncertainties taken into account. These measurements are consistent with the Standard Model predictions using the factorization assumption.

A SIMULTANEOUS MEASUREMENT OF THE BRANCHING FRACTIONS OF TEN *B* TO DOUBLE CHARM DECAYS

by

Chung Khim Lae

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Associate Professor Douglas Roberts, Chairman/Advisor
Professor Abolhassan Jawahery
Professor Rabindra Mohapatra
Professor Andrew Baden
Professor Robert Gammon, Dean's Representative

Table of Contents

1 Introduction

Since the dawn of civilization, man has wondered about the building blocks of nature: “Is the world made of some elementary particles that cannot be divided further?” “What are these particles and how do we build matter from them?” “Are there rules that govern the behavior of these particles?” For more than 2000 years, we have been pursuing the answers to these questions. Today, physicists have a picture of the fundamental world. This picture is called the Standard Model.

In the following section, I will give a brief overview of the Standard Model, hopefully, in a manner a non-physicist will find it easy to understand. No mathematics is involved. This is mainly to introduce someone unfamiliar with particle physics the names and terms that we will use later. The reader can jump to Section 1.3 if desired.

1.1 The Standard Model

The Standard Model is a theory describing the fundamental particles and how they interact with one another. According to the Standard Model, all known matter are composites of particles called quarks (q) and leptons (l), whose interactions are governed by three types of forces called weak, electromagnetic and strong[1].

To give an example, consider the structure of an atom. The atom consists of electrons, which are leptons, orbiting a dense nucleus due to the electromagnetic force

[1] The force gravity is not described by the Standard Model. Its strength is too weak compare to the other three forces and so it can be neglected in particle interactions.

between the electrons and the nucleus. The nucleus is made of protons and neutrons held together by the strong force, which overcomes the electromagnetic repulsion between the positively charged protons. Each proton or neutron is in turn made of three quarks.

Altogether, there are six types of leptons: electron (e^-), muon (μ^-), tau (τ^-) and their neutrino partners ν_e, ν_μ, ν_τ. The quarks also come in six flavors or types known as up (u), down (d), strange (s), charm (c), bottom (b) and top (t). These particles and their charges are summarized in Table 1.1. The lightest quarks and leptons, u, d, e^- and ν_e, made up more than 90% of the ordinary matter that we can see in the universe.

Quarks	Charge (e)	Leptons	Charge (e)
d, s, b	$-\frac{1}{3}$	e^-, μ^-, τ^-	-1
u, c, t	$+\frac{2}{3}$	ν_e, ν_μ, ν_τ	0

Table 1.1: The elementary particles of the Standard Model with their charges given in units of the proton's charge.

For every particle, the Standard Model says there is an anti-particle of the same mass but opposite charge. For example, the anti-particle of an electron is the positron, denoted as e^+ [1]. Similarly, anti-$\mu^- \equiv \mu^+$ and anti-$\tau^- \equiv \tau^+$. For a quark or neutrino, its anti-partner is denoted by an over-bar like $\bar{u}$. When a particle collides with its anti-particle, they annihilate to produce other particles. Although we do not see anti-matter in nature, its existence has been confirmed in experiments [2], [3].

Unlike leptons, quarks do not exist alone in nature. They always come in pairs or triplets. A quark will bind with an anti-quark to give a meson, or with two other quarks to

give a baryon. The opposite is true for an anti-quark. Collectively, mesons and baryons are called hadrons. A list of some hadrons that we will encounter is given in Table 1.2.

Hadrons	Quark Composition	Charge (e)
π^+, π^-	$u\bar{d}$, $d\bar{u}$	+1, −1
π^0	$u\bar{u} - d\bar{d}$	0
K^+, K^-	$u\bar{s}$, $s\bar{u}$	+1, −1
K_S, K_L	$d\bar{s} \pm s\bar{d}$	0
D^+, D^-	$c\bar{d}$, $d\bar{c}$	+1, −1
D^0, $\bar{D}^0$	$c\bar{u}$, $u\bar{c}$	0
B^+, B^-	$u\bar{b}$, $b\bar{u}$	+1, −1
B^0, $\bar{B}^0$	$d\bar{b}$, $b\bar{d}$	0
$Y(4S)$	$b\bar{b}$	0
proton (p^+)	uud	+1
neutron (n^0)	udd	0

Table 1.2: A list of common hadrons with their charges given in units of the proton's charge. Each hadron's charge is given by summing the charges of the quarks that made up the hadron, and using the fact that an anti-quark has an opposite charge to its quark partner. Note that π^0, K_S, K_L and $Y(4S)$ are their own anti-particles.

In the Standard Model, the quarks and leptons interact by exchanging particles. The exchange particle responsible for the electromagnetic force is the photon (γ), more

commonly known as light[1]. For the weak force, there are three exchange particles: W^+, W^- and Z^0, also known as the weak bosons. W^+ and W^- are anti-particles of each other while Z^0 and γ are anti-particles of themselves. For the strong force, the exchange particles are called gluons (g) which are exchanged by the quarks but not the leptons.

Although the Standard Model is successful in predicting all the particles that we have observed in the laboratories so far, there are still some open questions that it cannot address. For one, it is unable to explain the great difference between matter and anti-matter in our universe. From the Big Bang Theory, the universe should have started out with equal amounts of matter and anti-matter, but experimentally, we have determined that less than 0.01% of the present universe consists of anti-matter [4]. If the Big Bang Theory is correct[2], why is there so much matter now?

It turns out that a process has to occur during the early universe to create the matter and anti-matter imbalance. This process is called *CP* violation [5].

1.2 *CP* Violation

To understand *CP* violation, we need to know two operations: charge conjugation (C) and parity inversion (P). C is an operation that turns a particle to its anti-particle or vice versa, and P is an operation that reverses the space coordinates $\mathbf{x} \rightarrow -\mathbf{x}$. *CP* is then

---

[1] In the world of elementary particles, light or electromagnetic radiation behaves like a particle. We call this particle a photon.

[2] We have strong reasons to believe that the Big Bang Theory is correct. Observations that support the theory are: 1) the Hubble expansion of the universe; 2) the cosmic microwave background; 3) the relative abundance of light elements; and 4) the formation and evolution of galaxies.

the combined operation of C and P. If the equations of motion (or physics) remain the same after an operation, the operation is said to be conserved and we call it a symmetry. If not, we say the operation is violated.

It is known that the strong and electromagnetic forces both conserve C and P, but not the weak force. This came as a surprise to many physicists at the time parity violation was discovered in 1957 [6], [7], as it was expected that the weak force would conserve C and P just like the other forces[1]. It was then believed that the combined operation CP would instead be conserved by the weak force and be a symmetry of nature. However, this was proven wrong in 1964 [8]. To restore the symmetry, a third operation, conserved by both the strong and electromagnetic forces, was introduced. This operation changed the time coordinate $t \rightarrow -t$ and was called time-reversal (T). Today, it is believed that the combined operation CPT is a fundamental symmetry of nature.

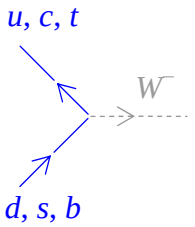


Figure 1.1: Weak interaction of the quarks.

In the Standard Model, CP violation is manifested when a quark changes flavor via the weak bosons $W^{\pm}$ as shown in Figure 1.1. From the diagram, there are nine possible

[1] In a sense, physicists prefer a theory to have symmetries. A theory with symmetries is more elegant as the calculations are much easier to perform.

ways the quarks can change flavor while conserving the charge at the interaction point. For each way, the strength of interaction is given by a number, and we can write these numbers in a 3 × 3 matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [9]:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1.1}$$

In general, V is complex to also account for flavor-changing of the anti-quarks (to give a total of 18 real numbers), but for physical results, V has to be unitary. That is,

$$V^{\dagger}V = 1, \tag{1.2}$$

where $V^{\dagger}$ is the transpose of the complex conjugate of V. From the unitarity of V, we can obtain six relations. The one favored by physicists is given by the first and third columns:

$$V_{ud}V_{ub}^{*} + V_{cd}V_{cb}^{*} + V_{td}V_{tb}^{*} = 0. \tag{1.3}$$

Equation (1.3) can be represented by a triangle in the complex plane as shown below, where one of the sides is normalized to one. This is called the Unitarity Triangle.

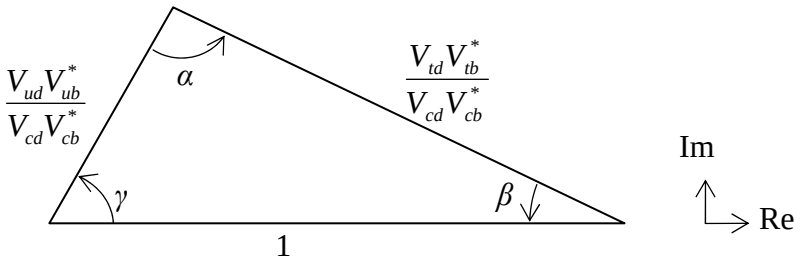


Figure 1.2: Unitarity Triangle.

One can actually construct two more triangles from Equation (1.2), but for these triangles, one side is much shorter than the other two, and so *CP* violating effects will be hard to observe. For the Unitarity Triangle, its sides are comparable in length as indicated

by experiments, and so physics related to the Unitarity Triangle (*B* decays) will have measurable *CP* quantities. This is why Equation (1.3) is favored by physicists.

It can be shown (using a result from [10]) that the area of the Unitarity Triangle is proportional to the size of *CP* violation. Hence, if the area is zero, *CP* is conserved. Using the current measurements of the CKM elements, we find that the size of *CP* violation is several orders of magnitude smaller than what is required to account for the matter-dominance that we observe today [11]. Therefore, there must be additional sources of *CP* violation from new physics beyond the Standard Model. It is by studying *CP* violation that we hope to uncover the new physics and increase our understanding of the universe.

By studying the weak decays of *B* mesons, each of the sides and angles of the Unitarity Triangle can be independently measured. Hence, we can over-constrain and test the accuracy of the Standard Model by making these measurements from many different decays. Any inconsistency would indicate the existence of new physics.

1.3 Double Charm Decays

The double charm decays $B \rightarrow D_{(s)}^{(*)}\bar{D}_{(s)}^{(*)}$, where $B = B^+$ or B^0 and $D_{(s)}^{(*)} = D^+$, D^0, D_s^+, D^{*+}, D^{*0} or D_s^{*+}, provide us with a rich field to study *CP* violation. (Here, "D^*" denotes an excited state of *D* and charged conjugated decays are included implicitly.) But before we can make any *CP* measurement, we need to understand the theory behind these decays so that we can compare reliably our results with those from other decays.

Central to the theory of double charm decays is the factorization assumption. To see how factorization works, let us consider the decay $B^0 \rightarrow D^+D^-$. We can represent this decay on a diagram shown in Figure 1.3. On such a diagram, quarks and leptons are

denoted by arrows and the exchange particles by dashed lines. The time axis is pointing to the right and a particle moving backward in time is treated as an anti-particle.

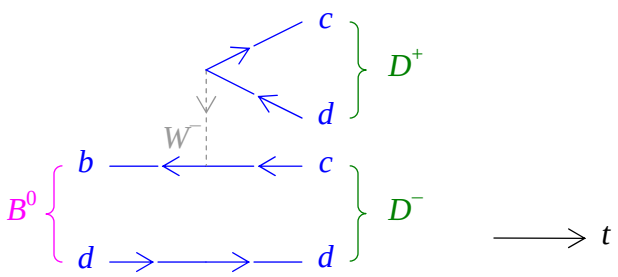


Figure 1.3: Tree diagram showing $B^0 \to D^+D^-$ decay.

Besides the tree diagram in Figure 1.3, the decay can also proceed via other paths like the one-loop diagram shown in Figure 1.4, also known as a penguin diagram. Usually, a path with more loops is less likely to occur and so to first approximation, one can just work with the tree diagram. In general, there are many diagrams and one must add up all their contributions to calculate the branching fraction of $B^0 \to D^+D^-$, which is the probability of a B^0 meson decaying into a pair of D^+D^- mesons. This is a quantity that we can measure in experiments.

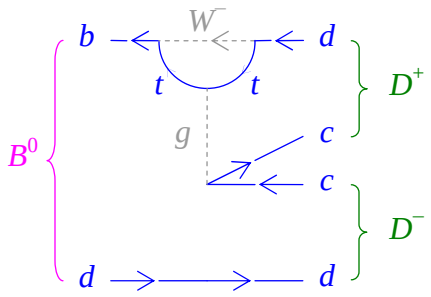


Figure 1.4: A one-loop (penguin) diagram for $B^0 \to D^+D^-$ decay.

To simplify the calculation, we ignore the strong interactions between the two daughters. Then, we can decompose the decay diagram into two parts as shown in Figure 1.5. This is the factorization assumption.

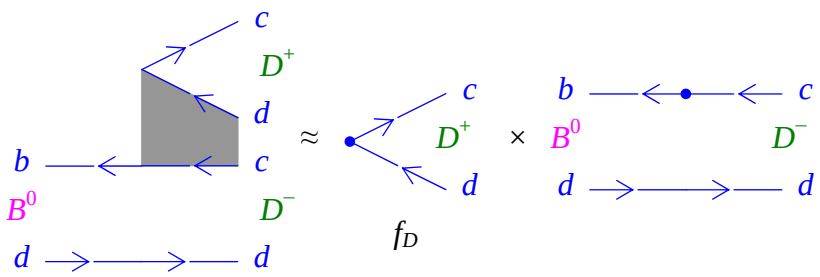


Figure 1.5: An illustration of the factorization assumption. The shaded region indicates the strong interactions that are neglected between the two daughters.

For simplicity, let us call the first term (*cd* vertex) as the decay constant f_D and the second term as the form factor[1]. To give a physical interpretation of these quantities, we can think of the decay constant as measuring the probability of a quark and anti-quark to bind together to form a meson while the form factor measures the probability that the flavor-changed quark ($\bar{b}$) and the spectator quark (d) will form a meson in the final state.

Now, if we look at the leptonic decay $D^+ \to l^+ \nu_l$, where $l = e, \mu, \tau$ (see Figure 1.6), we find that it also contains the *cd* vertex. Since the leptons do not interact strongly, the leptonic decay is not complicated by the strong interactions between its decay products. Hence, we can obtain f_D directly from the branching fraction of $D^+ \to l^+ \nu_l$. Similarly, the

[1] The decay constant and form factor are actually defined such that the first and second terms are proportional to them.

form factor of $B^0 \rightarrow D^+D^-$ can be obtained from the semi-leptonic decay $B^0 \rightarrow D^-l^+\nu_l$. Therefore, the basic idea of factorization is to write the branching fraction in terms of common quantities that we can measure from other decays.

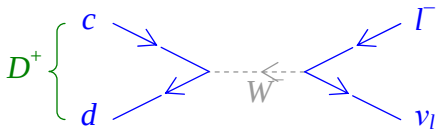


Figure 1.6: Diagram showing the leptonic decay $D^+ \rightarrow l^+\nu_l$.

Whether factorization is a valid assumption depends on the amount of strong interactions between the daughters. If the daughters are light, they will fly away quickly before they have a chance to interact and factorization will be a good approximation. On the other hand, if the daughters are heavy as in the case of the charmed mesons $D_{(s)}^{(*)}$, the assumption may fail. Hence, it is important to test the factorization assumption for the theory of double charm decays.

One way to test the factorization assumption is to measure the branching fractions and compare them to the predicted values. Listed in Table 1.3 are the branching fractions of $B \rightarrow D_{(s)}^{(*)}\bar{D}^{(*)}$ using three different models to calculate the decay constants and form factors. These models are the constituent quark model (CQM) [12], the light-front quantum chromodynamics (LFQCD) [13] and heavy quark symmetry with corrections (HQSC) [14]. We will not delve into these models here, but instead ask the interested reader to look at the references for details. Since the values are very close, a single measurement of a branching fraction is not enough and a simultaneous measurement of several branching fractions is required to pick out the correct model.

Decay[6]	CQM	LFQCD	HQSC
$B^0 \to D^{(*)0}\overline{D}^{(*)0}$	0.00	0.00	0.00
$B^0 \to D^+ D^-$	0.38	0.40	0.31
$B^0 \to D^{*+} D^-$	0.73	0.78	0.71
$B^0 \to D^{*+} D^{*-}$	1.06	0.99	0.91
$B^- \to D^- D^0$	0.42	0.44	0.33
$B^- \to D^{*-} D^0$	0.54	0.49	0.45
$B^- \to D^- D^{*0}$	0.39	0.36	0.31
$B^- \to D^{*-} D^{*0}$	1.16	1.08	0.98
$B^0 \to D_s^+ D^-$	9.70	10.33	8.25
$B^0 \to D_s^{*+} D^-$	12.49	11.42	10.80
$B^0 \to D_s^+ D^{*-}$	9.19	8.50	7.67
$B^0 \to D_s^{*+} D^{*-}$	28.78	27.09	25.51
$B^- \to D_s^- D^0$	10.58	11.26	8.94
$B^- \to D_s^{*-} D^0$	13.65	12.47	11.73
$B^- \to D_s^- D^{*0}$	10.02	9.27	8.34
$B^- \to D_s^{*-} D^{*0}$	31.37	29.52	27.69

Table 1.3: Predictions of branching fractions ($\times 10^{-3}$) from three different models using the factorization assumption (from Table IV of [14]). These values have uncertainties around 15–20%, mainly from the calculations of the decay constants and form factors.

[6] We use $B^0 \to D^{*+} D^-$ to denote both $B^0 \to D^{*+} D^-$ and $B^0 \to D^{*-} D^+$ decays.

Another way to test the factorization assumption is through the ratios of branching fractions. In Table 1.4, we list a few of these ratios for double charm decays predicted from the factorization approach. A different method to determine the decay constants and form factors is used in this case [15]. Since some of these values are related to the ratio of the decay constants of D^+ to D_s^+, it will be interesting to compare the ratio of branching fractions of D^+ to D_s^+ leptonic decays with those obtained from double charm decays. A difference will indicate a presence of non-factorizable effects in double charm decays.

Decays	Ratio	Decays	Ratio
$\frac{D_s^+ D^{*-}}{D_s^+ D^-}$	0.70	$\frac{D^+ D^-}{D_s^+ D^-}$	$0.05\left(\frac{f_D}{f_{D_s}}\right)^2$
$\frac{D_s^{*+} D^-}{D_s^+ D^-}$	0.70	$\frac{D^{*+} D^-}{D_s^+ D^-}$	$0.072\left(\frac{f_D}{f_{D_s}}\right)^2$
$\frac{D_s^{*+} D^{*-}}{D_s^+ D^-}$	1.81	$\frac{D^{*+} D^{*-}}{D_s^+ D^-}$	$0.092\left(\frac{f_D}{f_{D_s}}\right)^2$

Table 1.4: Ratios of branching fractions for $B^0 \rightarrow D_{(s)}^{(*)+} D^{(*)-}$ decays (from Table III of [15]).

Besides testing the factorization assumption, the branching fractions can also be used to determine theoretical uncertainties in the factorization approach. One of these uncertainties is the size of final-state interactions which can be significant in decays when the daughters are heavy. Such effects are assumed zero when calculating the penguin corrections to the angle β of the Unitarity Triangle for $B^0 \rightarrow D^{*+} D^{*-}$ decay. (For details,

please refer to [16].) Therefore, it is important to check the size of these effects before we compare the value of β with those from other decays. The decays $B^0 \to D^{(*)0}\bar{D}^{(*)0}$ are most suitable for this purpose. In the absence of final-state interactions, their branching fractions are predicted to be negligible. Hence, by measuring their branching fractions, we can determine the size of final-state interactions in double charm decays.

Another source of theoretical uncertainty comes from the non-factorizable effects in the exchange diagram of $B^0 \to D^{*+}D^{*-}$ (see Figure 1.7) when measuring the angle γ from $B^0 \to D_{(s)}^{*+}D^{*-}$ decays [17]. In this case, since the diagram similar to Figure 1.3 does not exist for $B^0 \to D_s^{*+}D_s^{*-}$, this decay proceeds mainly through the exchange diagram and so the ratio of the branching fractions of $B^0 \to D_s^{*+}D_s^{*-}$ to $B^0 \to D^{*+}D^{*-}$ can give an estimate of the size of the exchange contribution.

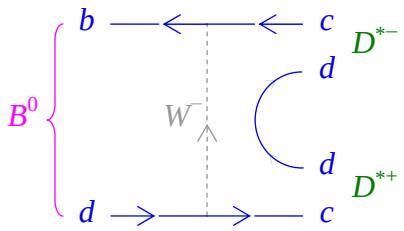


Figure 1.7: Exchange diagram for $B^0 \to D^{*+}D^{*-}$ decay.

1.4 Summary

As we have seen, measuring the branching fractions is an important step toward determining the amount of *CP* violation in double charm decays. We also see that one branching fraction is not enough, and that a simultaneous measurement of several branching fractions is necessary to verify the factorization assumption.

In this analysis, we present a simultaneous branching fraction measurement of ten double charm decays[7]: $B^0 \to D^0\bar{D}^0$, $B^0 \to D^{*0}\bar{D}^0$, $B^0 \to D^{*0}\bar{D}^{*0}$, $B^0 \to D^+D^-$, $B^0 \to D^{*+}D^-$, $B^0 \to D^{*+}D^{*-}$, $B^- \to D^-D^0$, $B^- \to D^{*-}D^0$, $B^- \to D^-D^{*0}$ and $B^- \to D^{*-}D^{*0}$. Hopefully, this will provide a stringent test on the factorization assumption through the ratios of branching fractions.

[7] We use $B^0 \to D^{*0}\bar{D}^0$ to denote both $B^0 \to D^{*0}\bar{D}^0$ and $B^0 \to \bar{D}^0D^{*0}$ decays.

2 Detector

At BABAR, B mesons are produced by colliding e^- and e^+ beams at a center-of-mass energy[8] of 10.58 GeV, corresponding to the mass of $\Upsilon(4S)$ (a $b\bar{b}$ meson), which in turn decays almost exclusively to a pair of $B^0\bar{B}^0$ or B^+B^- mesons. The rate of production at peak performance is roughly ten $B\bar{B}$ pairs per second.

To detect the decay products efficiently at such a rate, a detector has to be constructed from five sub-detectors: silicon vertex tracker (SVT), drift chamber (DCH), detector of internally reflected Cherenkov light (DIRC), electromagnetic calorimeter (EMC) and instrumented flux return (IFR). This is the BABAR detector. Its layout is given in Figure 2.1, which shows the SVT and DCH inside a uniform magnetic field of 1.5 T provided by a superconducting solenoid. Together, they form the tracking system responsible for measuring the momentum and position of a charged particle.

---

[8] We will assume that the reader is familiar with a few physical concepts like momentum, energy and the center-of-mass frame. The electron volt or eV is frequently used as a unit of energy in particle physics. 1 eV = 1.602×10^{-19} J, which is the energy to move an electron through a potential difference of one volt, and 1 GeV = 10^3 MeV = 10^9 eV. Because mass, momentum and energy are related through the equation $E^2 = (pc)^2 + (mc^2)^2$, where c is the speed of light in vacuum, we also express mass in units of eV/c^2 and momentum in eV/c. For example, the mass of an electron is 9.11×10^{-31} kg = 0.511 MeV/c^2.

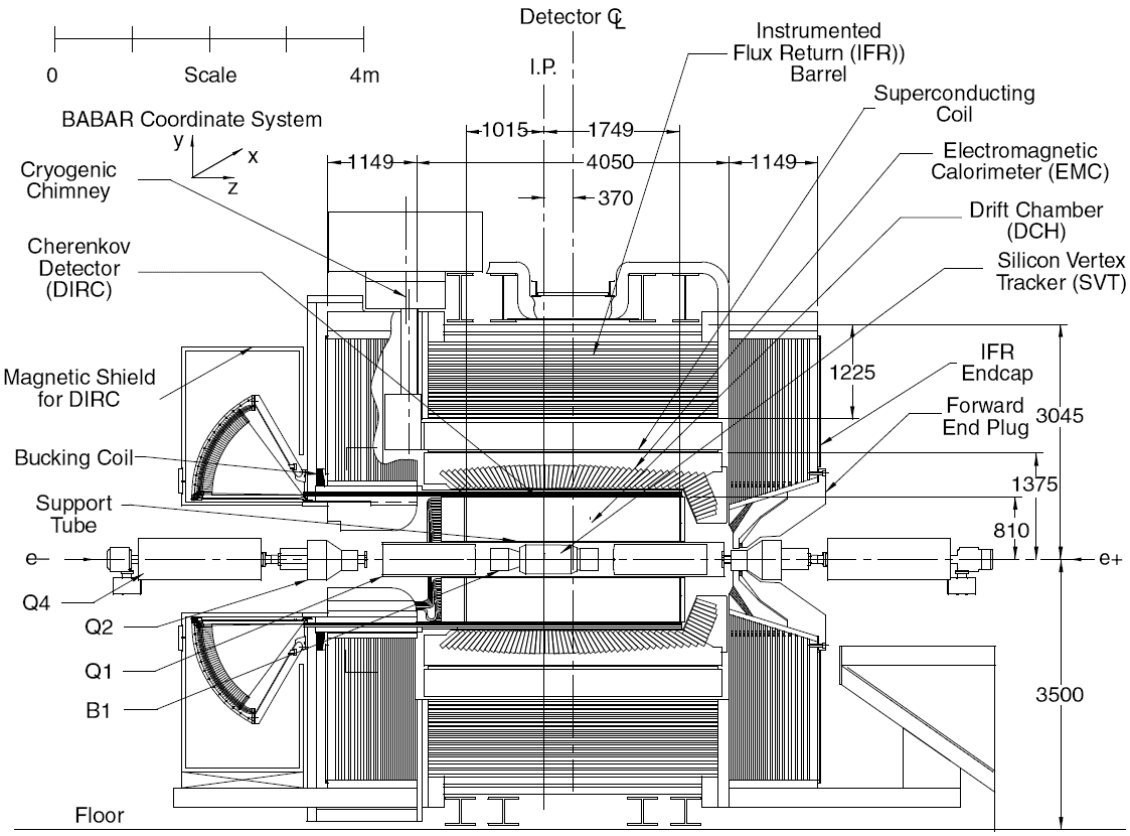


Figure 2.1: Longitudinal section of the BABAR detector.

In the following sections, we give a brief overview of each sub-detector and its performance pertinent to this analysis. We will, however, omit the IFR, since its main purpose is to detect muons which are not used in our analysis. For a full description of the BABAR detector, one can refer to [18].

2.1 SVT

The SVT is made out of five layers of silicon wafers cylindrically centered about the beam pipe. Each wafer has readouts on both sides to record the φ and z coordinates, thus allowing a 3D measurement of a charged particle's position. (The r coordinate being given by the layer number.)

The SVT is designed to measure the angles and positions of charged particles just outside the beam pipe. The mean vertex resolution of a B meson is typically around 100 μm. This precision allows us to reconstruct $B^0 \rightarrow D^+D^-$ decays, where the distance between the two D mesons in the x-y plane is ~275 μm. The SVT also provides tracking for a particle with transverse momentum (momentum in the x-y plane) $p_t < 120$ MeV/c. This is important for detecting slow pions from $D^{*+} \rightarrow D^0\pi^+$ decays.

To study the efficiency of detecting slow pions, $D^{*+} \rightarrow D^0\pi^+$ decays are selected by reconstructing events of the type $B \rightarrow D^{*+}X$ followed by $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$. Figure 2.2 shows the mass difference, $\delta M = m(D^{*+}) - m(D^0)$, for the total sample and the sub-sample of events in which the slow pion has been reconstructed in both the SVT and the DCH. The difference between these two distributions shows the contribution from SVT standalone tracking, both in terms of the gain of signal events and of the resolution. The gain in efficiency is mostly at very low momenta, and the resolution is affected by multiple scattering and limited track length of the slow pions. From simulation, we determine the slow pion efficiency to be above 80% for $p_t > 70$ MeV/c.

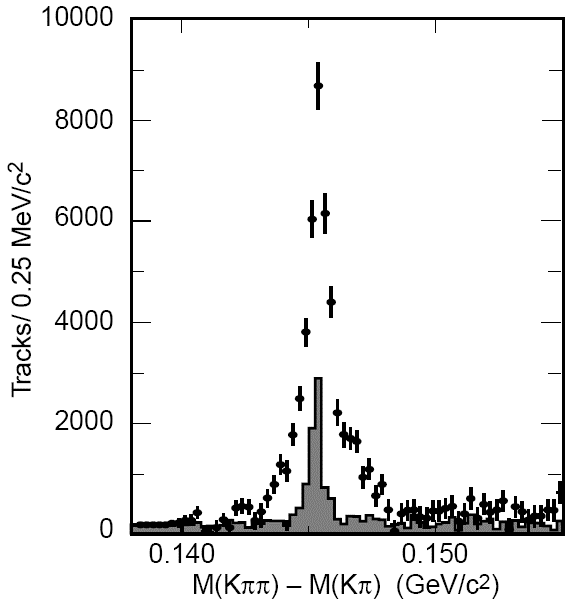


Figure 2.2: δM distributions (background-subtracted) of slow pions for all detected events (data points) and for events in which the slow pion is reconstructed both in the SVT and DCH (shaded histogram).

Besides measuring the momentum of a charged particle, the SVT can also measure the energy deposited by the particle from the pulse height recorded, and so perform particle identification through the rate of energy loss over distance, dE/dx [19]. The dE/dx resolution from the SVT is around 14%, which can separate $K^{\pm}$ and $\pi^{\pm}$ up to a momentum of 500 MeV/c.

2.2 DCH

The DCH consists of 40 cylindrical layers of drift cells centered around the beam pipe. Each cell is made out of six field wires arranged in a hexagon and one sense wire in the center. To give a reading of the z coordinate, some layers have their wires aligned at a

small angle to the *z*-axis. Thus, the DCH is capable of both tracking a charged particle and measuring its momentum.

Not only does the DCH complement the measurements of the SVT, it also increases the tracking efficiency of long-lived particles that decay outside the SVT, especially $K_S \rightarrow \pi^+\pi^-$ which is used in this analysis. Tracking efficiency in the DCH is around 98% for $p_t > 200$ MeV/c and polar angle $\theta > 500$ mrad. The dE/dx resolution is 7.5% and allows π-K separation up to p_t of 700 MeV/c (see Figure 2.3).

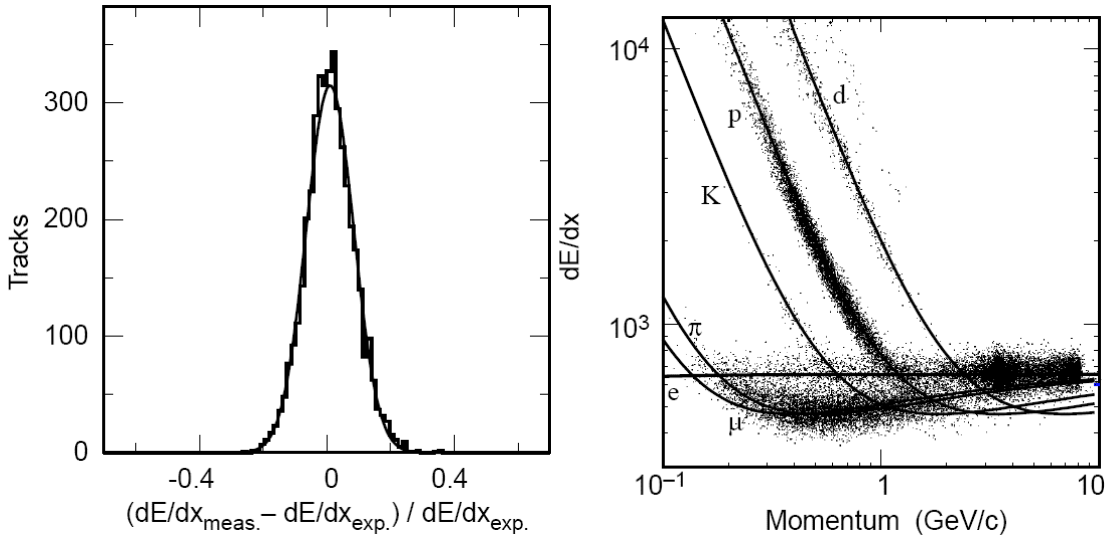


Figure 2.3: Left: Difference between the measured and expected dE/dx in the DCH for $e^{\pm}$. The curve is a Gaussian fit to data with a width of 7.5%. Right: Measurement of dE/dx in the DCH as a function of track momentum. The curves show the Bethe-Bloch predictions [19] derived from selected control samples of particles of different masses.

2.3 DIRC

The DIRC is a novel device based on the concept of Cherenkov radiation. A charged particle traveling faster than light in an optically dense medium will emit a cone of light whose apex angle depends on its velocity in the medium. In the DIRC, the emitted Cherenkov light travels along tubes of fused silica via total internal reflection (thus preserving its apex angle) to the backward end of the BABAR detector, where it is detected by photomultiplier tubes.

The main purpose of the DIRC is to provide particle identification for $K^{\pm}$ and $\pi^{\pm}$ at high momentum up to 4.5 GeV/c. (Particle identification below 700 MeV/c relies on dE/dx in the SVT and DCH.) This is important for reconstructing a D decay containing a kaon daughter whose momentum peaks around 1 GeV/c. Such decays are common in this analysis.

The figures below show the performance of the DIRC in kaon identification. From Figure 2.4, we see that the background is greatly reduced when DIRC information is used for reconstructing D mesons. From Figure 2.5, we find that the kaon efficiency is usually greater than 90% and the pion fake rate less than 10%.

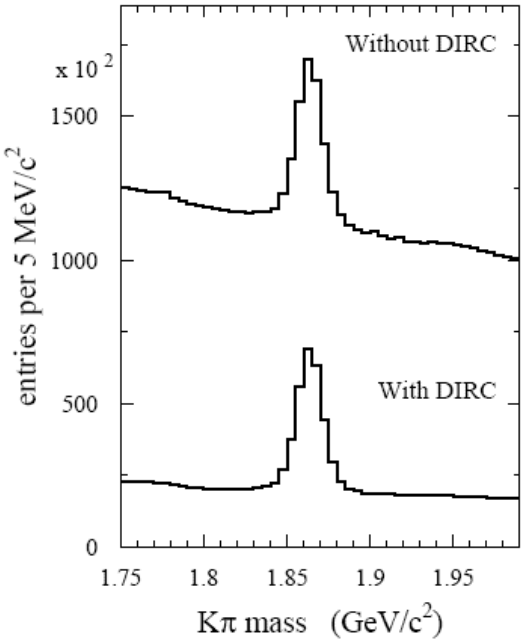


Figure 2.4: Mass of $D^0(K\pi)$ with and without the use of the DIRC for kaon identification.

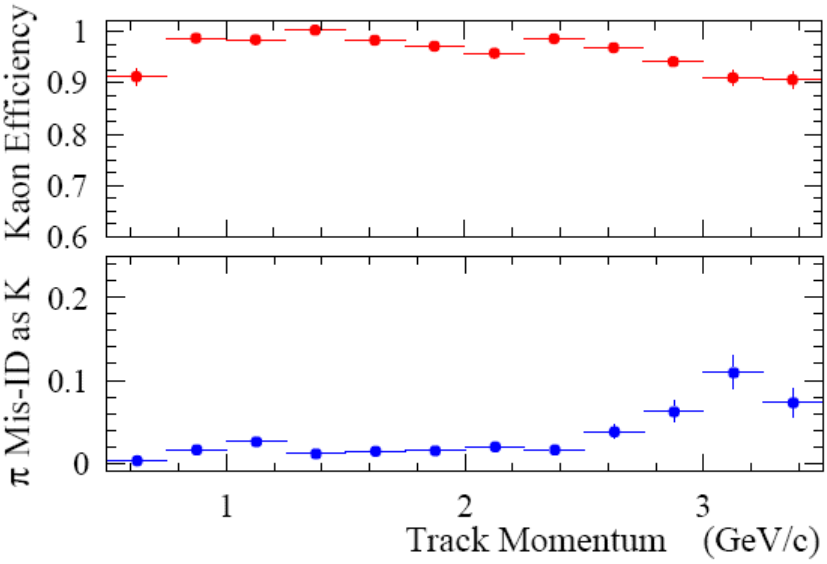


Figure 2.5: Efficiency (top) and fake rate (bottom) for the selection of $K^{\pm}$ as a function of track momentum, determined using $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$.

2.4 EMC

The EMC is an array of finely segmented crystals made of thallium-doped cesium iodide. These crystals have high light yields and short radiation lengths, thus allowing for high detector efficiency and shower containment.

The EMC measures electromagnetic showers from $e^{\pm}$ and γ for energies between 0.02 GeV and 9 GeV. The energy resolution σ_E/E ranges from 2% to 5%, while the angular resolution goes from 3 mrad to 12 mrad, resulting in a width of 6.9 MeV/c^2 for the $\pi^0(\gamma\gamma)$ mass (see figure below).

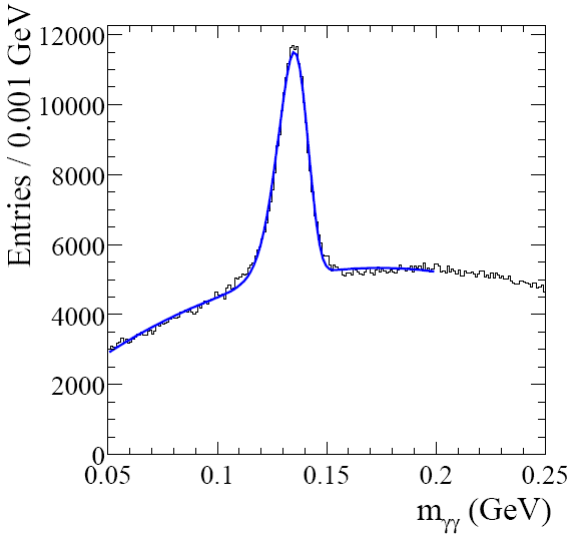


Figure 2.6: Mass of $\pi^0(\gamma\gamma)$. The curve is a fit to data with a peak at 135.1 MeV/c^2.

3 Reconstruction of Double Charm Decays

3.1 Reconstruction Procedure

The reconstruction of $B \to D^{(*)}\bar{D}^{(*)}$ starts from the bottom of the decay trees. To illustrate, consider a decay tree of $B^0 \to D^{*+}D^{*-}$ shown in Figure 3.1. We begin by selecting suitable $\pi^{\pm}$, $K^{\pm}$ and γ candidates from the stable particles detected by our detector. Then, we combine the relevant candidates to give the composites higher up the decay tree. For example, a K^- candidate is combined with a π^+ candidate to give a D^0 candidate, which in turn is combined with another π^+ candidate to form a D^{*+} candidate. At each level, we apply a set of selections to reduce the background while keeping about 90% of the signal. We move up the decay tree this way until we reach the B^0 meson.

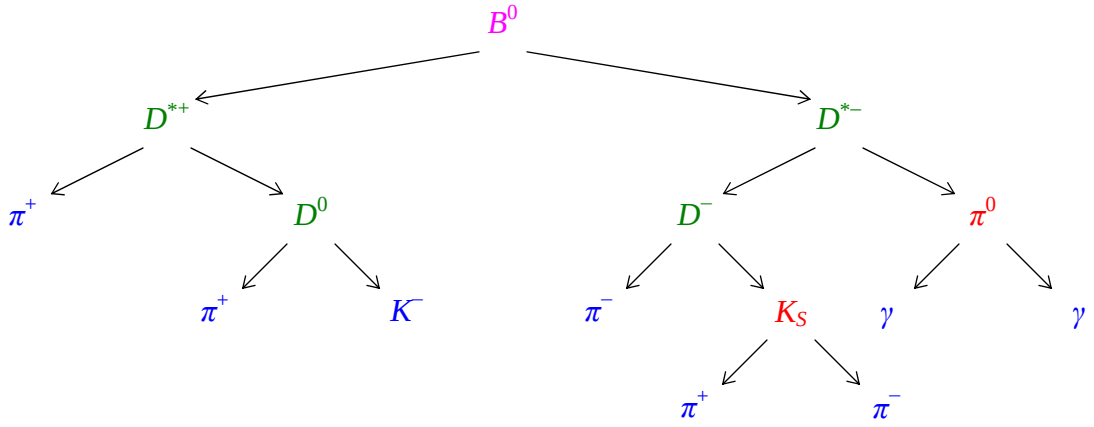


Figure 3.1: A possible decay tree for $B^0 \to D^{*+}D^{*-}$.

Before we go on to describe our selection criteria in detail, let us, for the ease of explanation, define a few terms and notations:

- E: particle's energy.
- $\mathbf{p} \equiv (p_x, p_y, p_z)$: particle's 3-momentum.
- $p \equiv |\mathbf{p}|$: particle's momentum (magnitude).
- $p_t \equiv \sqrt{p_x^2 + p_y^2}$: particle's transverse momentum.
- CM frame: center-of-mass frame of the colliding beams
- p_{CM}: particle's momentum in CM frame.
- m_{PDG}: particle's mass given by the Particle Data Group (PDG) [20].
- Kinematic fit: creation of a parent by adding its daughters' momenta and energies, assuming conservation of momentum and energy.
- Vertex-constrained fit: a kinematic fit where the daughters are constrained to come from a common point.
- Mass-constrained fit: a kinematic fit where the parent's mass is constrained to its PDG value.
- m: parent's fitted mass before a mass constraint.
- $\Delta m \equiv m - m_{PDG}$.
- P(χ^2): χ^2-probability (a measure of the goodness-of-fit) from a constrained fit; a poor fit has a P(χ^2) close to zero.
- Beam spot: region where the e^+ and e^- beams collide.

Note: In all the figures below, the vertical dashed lines denote the selections that are used in our analysis. Their efficiencies (fraction of selected signal) are given at the top right-hand corner. All figures are derived from simulated Monte Carlo events.

3.2 Event Level Selection

At event level, (that is, before making any composite), there is a powerful variable that distinguishes between $e^+e^- \rightarrow q\bar{q}$ ($q = u$, d, s or c) and $B\bar{B}$ events. This variable is given by the momentum distribution of all the stable particles measured by the detector and is called the Fox-Wolfram ratio [21]. It is defined as $R_2 = H_2 / H_0$, where

$$H_l = \frac{1}{s} \sum_{i,j} p_i p_j P_l(\cos\theta_{ij}) , \tag{3.1}$$

s is the square of sum of beam energies in the CM frame;

P_l is the Legendre Polynomial of order l;

i and j are indices running over all the stable particles; and

θ_{ij} is the angle between particles i and j.

The ratio characterizes the shape of an event's topology. For a jet-like event like $e^+e^- \rightarrow \tau^+\tau^-$, since the CM energy (10.58 GeV) is much higher than the taus' masses ($m_\tau = 1.78$ GeV/c^2), the taus move away quickly and most of the taus' daughters are produced in two cones directed opposite to each other. Hence, R_2 is close to one. For an isotropic event like $e^+e^- \rightarrow B\bar{B}$, since $m_B = 5.28$ GeV/c^2, the mass-energy difference between e^+e^- and $B\bar{B}$ is almost zero and the B mesons are essentially at rest. The decay products of B will be produced evenly in all directions, resulting in R_2 being close to zero. Figure 3.2 shows the distributions of R_2 for different types of simulated events. For our event selection, we take $R_2 \leq 0.6$. This has an efficiency of almost 100% for signal events while rejecting about 5% $c\bar{c}$, 11% uds and 73% $\tau^+\tau^-$ background.

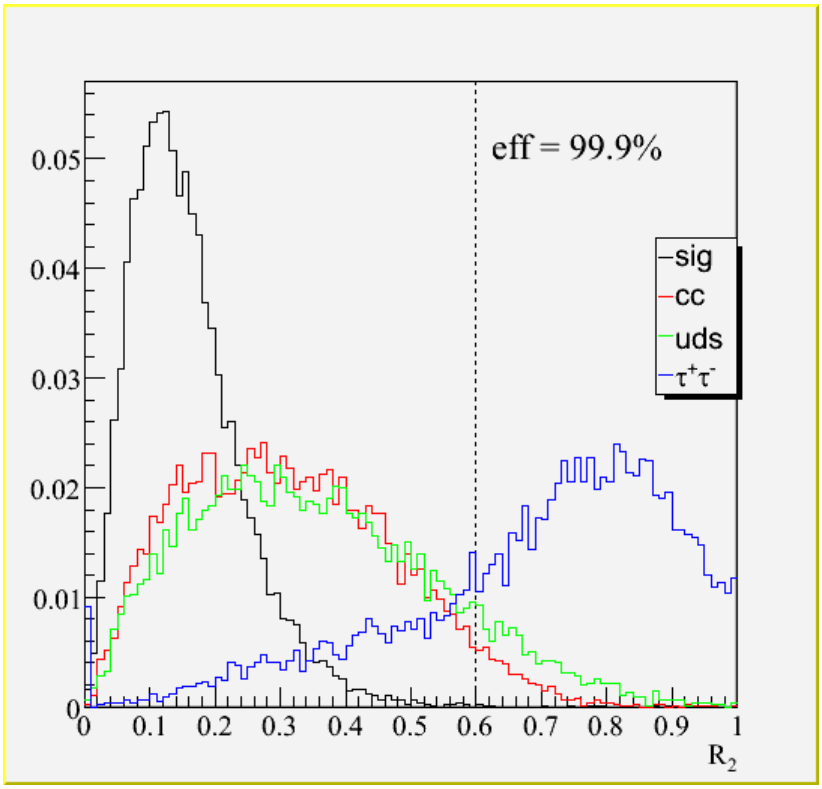


Figure 3.2: Normalized distributions of R_2 for $e^+e^- \rightarrow c\bar{c}$, $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s$), $e^+e^- \rightarrow \tau^+\tau^-$ and signal events. R_2 is zero if there is no detected particle in the event.

3.3 Selection of $\pi^{\pm}$

In the presence of a magnetic field, a charged particle bends to form a helical track whose radius depends on its transverse momentum. From the hits in the SVT and DCH, we reconstruct the charged tracks by fitting the hits to helices via the Kalman filter algorithm [22], which has a high efficiency of 98% in the DCH. By default, the tracks' masses are constrained to the PDG value of $\pi^{\pm}$ (139.57 MeV/c^2). Candidates for $\pi^{\pm}$ are then selected from the list of charged tracks using the following criteria:

- Distance of closest approach to the beam spot in x-y plane, $d_0 \leq 1.5$ cm
- Distance of closest approach to the beam spot along the z-axis, $z_0 \leq 10$ cm
- Number of SVT + DCH hits ≥ 5
- $p \leq 10$ GeV/c

This set of cuts is very loose and mainly serves to reduce the cosmic and machine backgrounds that do not originate from the beam spot. From Figure 3.3, we see that the d_0 and z_0 cuts help to lower the number of $e^{\pm}$ arising from pair production ($\gamma \rightarrow e^+e^-$) or from secondary decays like $\pi^0 \rightarrow e^+e^-\gamma$ and $K^+ \rightarrow \pi^0 e^+ \nu_e$. The cuts also lower the number of protons or anti-protons coming from particles interacting with the detector hardware.

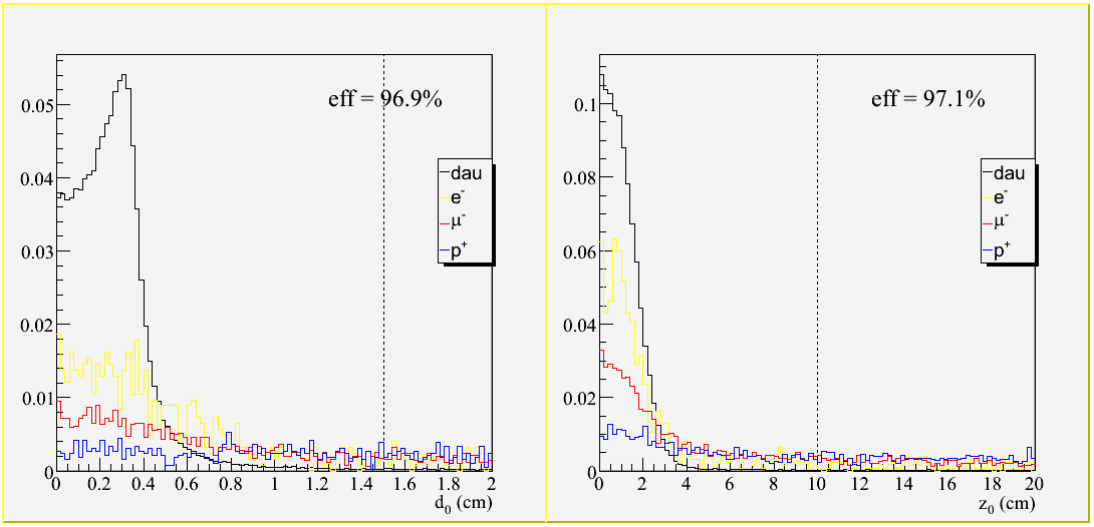


Figure 3.3: Normalized distributions of d_0 (left) and z_0 (right) for e^-, μ^-, p^+ and daughters (K^+ or π^+) of $D^{(*)}$. The anti-particles are included in the histograms.

From this list of good tracks, a sub-list called PionSoft is created for the slow pion of $D^{*+} \rightarrow D^0\pi^+$. Because the mass difference between D^{*+} and D^0 is small (PDG value of

145.42 MeV/c^2), the daughter π^+ will have a low momentum and so we can impose the additional cut for PionSoft:

- $p_{CM} \leq 0.45$ GeV/c

From the list of good tracks, another sub-list is created for the daughters of D mesons. It is called GoodTracksLoose and has the additional cuts:

- $p_t \geq 0.1$ GeV/c (Figure 3.4, left plot)
- Number of DCH hits ≥ 12 (Figure 3.4, right plot)

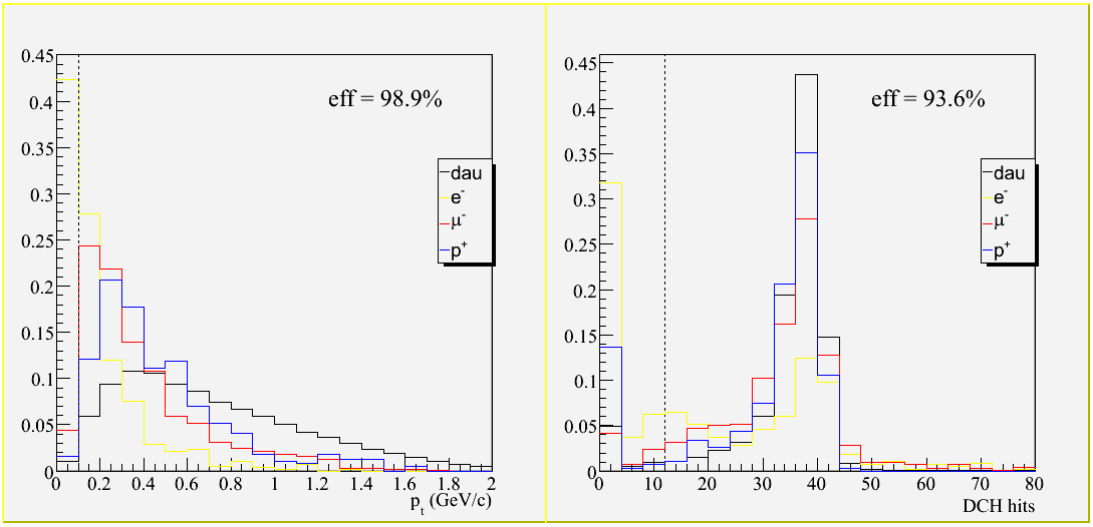


Figure 3.4: Normalized distributions of p_t (left) and number of DCH hits (right) for e^-, μ^-, p^+ and daughters (K^+ or π^+) of $D^{(*)}$ (with anti-particles included).

3.4 Selection of γ

A photon creates a shower that stretches across several crystals when it deposits its energy in the EMC. From the shower shape, it is possible to distinguish a photon from

other types of particles. Hence, similar to the event shape variable R_2, a quantity f_{LAT} is defined for the shower shape based on the lateral distribution of deposited energy [23]:

$$f_{\mathrm{LAT}} = \frac{\sigma_r^2}{\sigma_r^2 + E_1 + E_2}, \tag{3.2}$$

where

$$\sigma_r^2 = \frac{1}{r_a^2}\sum_{i=3}^{n} |\,\mathbf{r}_i - \mathbf{r}\,|^2\,, \tag{3.3}$$

$$\mathbf{r} = \sum_{i=1}^{n} E_i \mathbf{r}_i \Big/ \sum_{i=1}^{n} E_i\,, \tag{3.4}$$

$\mathbf{r}_i$ is the radius vector of crystal i;

E_i is the energy deposited in crystal i with $E_1 \geq E_2 \geq \ldots \geq E_n$;

$n \geq 3$ is the shower size (number of crystals occupied by the shower); and

r_a is the average distance between the centers of two adjacent counters.

A photon, being a neutral particle, does not leave a track in the SVT or DCH. An electron, on the other hand, leaves a track and creates a shower in the EMC. Therefore, a list of photon candidates called PhotonLoose, to be used for reconstructing $\pi^0 \rightarrow \gamma\gamma$, is made from the EMC showers as follows:

- Showers that are not matched to any tracks
- $E \geq 30$ MeV
- $f_{\mathrm{LAT}} \leq 0.8$ (Figure 3.5)

For the slow photons in $D^{*0} \rightarrow D^0\gamma$, a tighter list, PhotonSoft, is used:

- $E \geq 0.1$ GeV
- $p_{\mathrm{CM}} \leq 0.45$ GeV/c

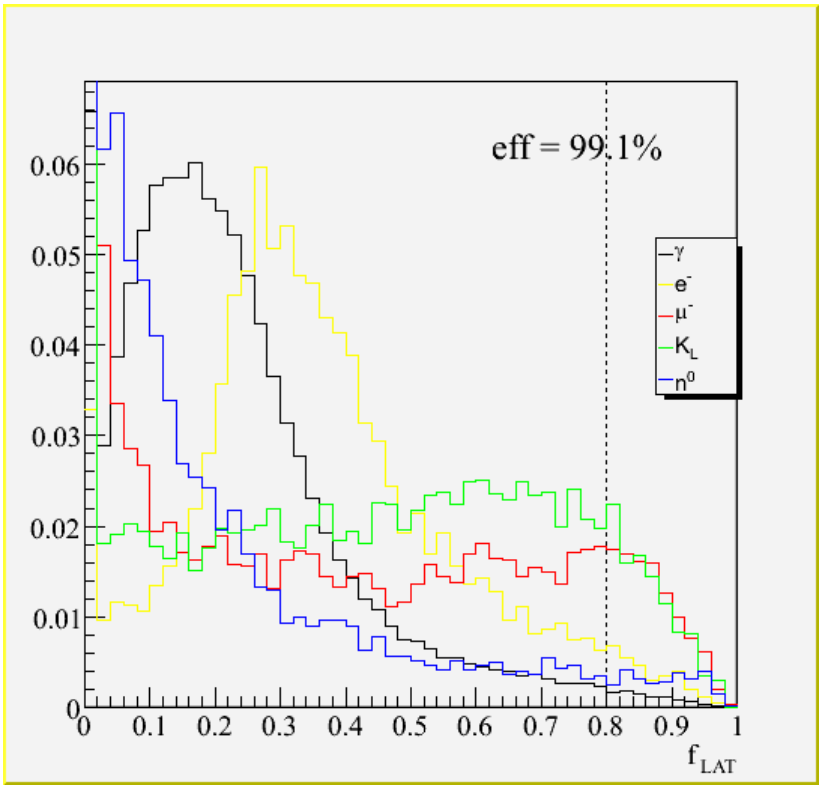


Figure 3.5: Normalized distributions of f_{LAT} for γ, e^{-}, μ^{-}, K_L and n^0 with anti-particles included. f_{LAT} is 0 if a shower size is less than 3.

3.5 Selection of π^0

π^0 decays to two photons almost 100% of the time [20]. We reconstruct this decay by combining two photons from PhotonLoose under the following conditions:

- Kinematic fit, where the photons are assumed to come from the beam spot
- $115 \leq m \leq 150$ MeV/c^2 (Figure 3.6)
- Mass constraint

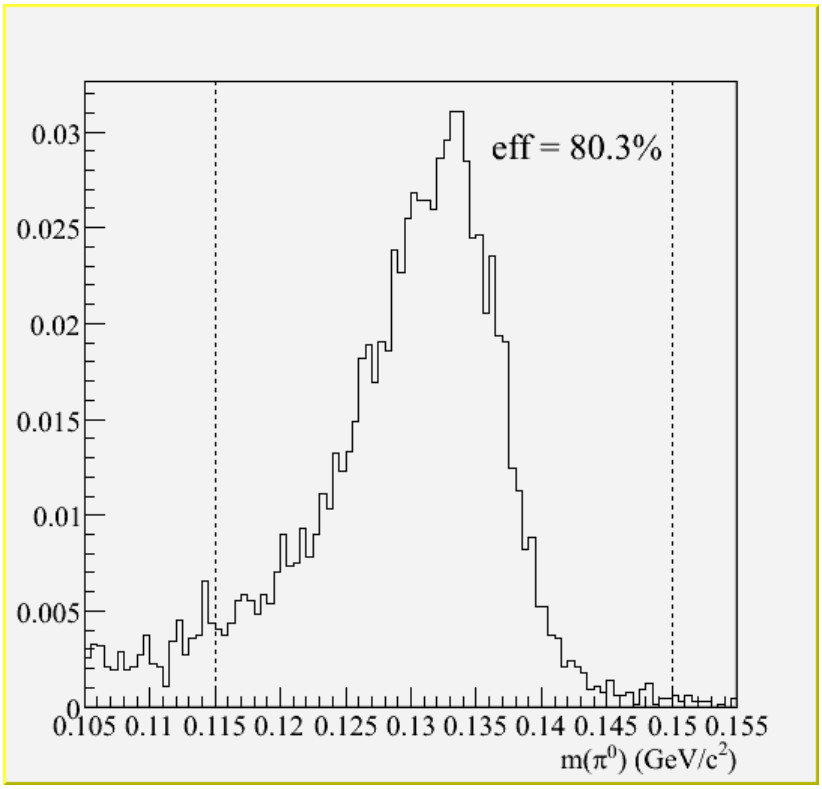


Figure 3.6: Normalized distribution of m for π^0 daughters of $D^{(*)}$.

From this list, two types of π^0 are created. The first type, known as Pi0Soft, for the slow pions of $D^{*+} \rightarrow D^+\pi^0$ and $D^{*0} \rightarrow D^0\pi^0$ decays, has the cut:

- $p_{CM} \leq 0.45$ GeV/c

The second type, for the daughter of $D^0 \rightarrow K^-\pi^+\pi^0$ decay, is called Pi0Default and has the cut:

- $E \geq 0.2$ GeV

3.6 Selection of $K^{\pm}$

Particle identification (PID) is used to select $K^{\pm}$ from GoodTracksLoose. Charged tracks are identified as $K^{\pm}$ based on the dE/dx measurements in both the SVT and DCH,

and also on the Cherenkov measurements in the DIRC. Two types of PID criteria are considered. A tight PID is used for kaons (except $D^0 \to K^-\pi^+$) in $B^0 \to D^{(*)0}\bar{D}^{(*)0}$ decays, where we expect to see high background-to-signal ratios. For all other decays, a loose PID, optimized to distinguish between kaons and pions, is used. For kaons that are daughters of D mesons, the loose criterion has an efficiency of more than 95% and a pion fake rate of less than 8% (see Figure 3.7). Although the proton fake rate is high, this is not a concern as the number of protons in our samples is low.

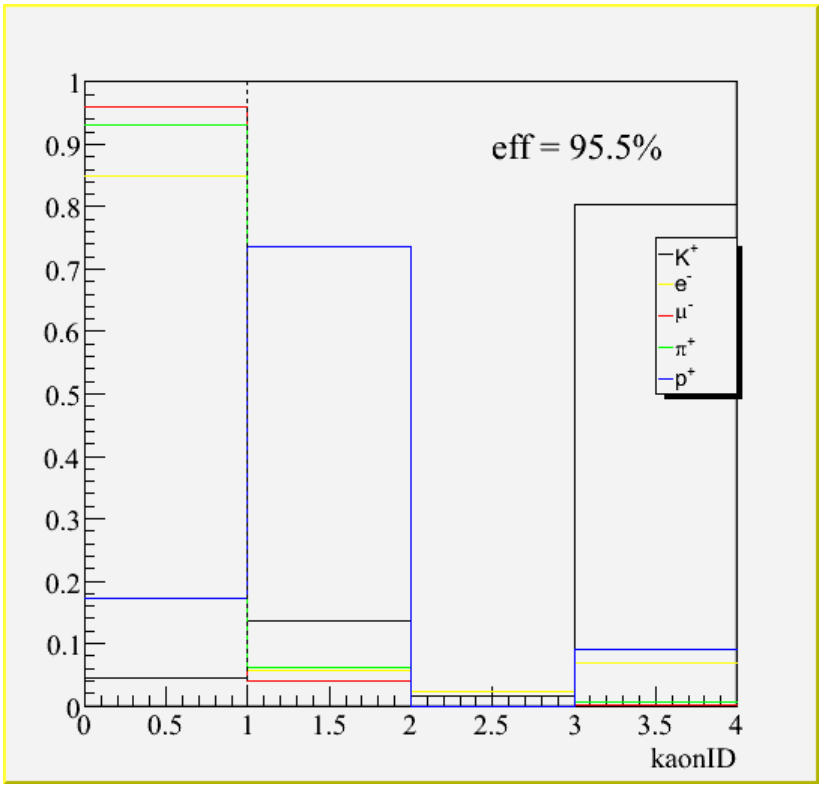


Figure 3.7: Normalized distributions of kaon PID quality (kaonID) for K^+, e^-, μ^-, π^+ and p^+ (anti-particles included). A charged track has kaonID ≥ 1 (kaonID ≥ 2) if it passes the loose (tight) criterion. Otherwise, kaonID = 0.

3.7 Selection of K_S

$K_S \rightarrow \pi^+\pi^-$ decay is reconstructed by fitting two charged tracks as follows:

- Vertex constraint
- $|\Delta m| \leq 15$ MeV/c^2
- $p \geq 0.2$ GeV/c
- P(χ^2) ≥ 0.001

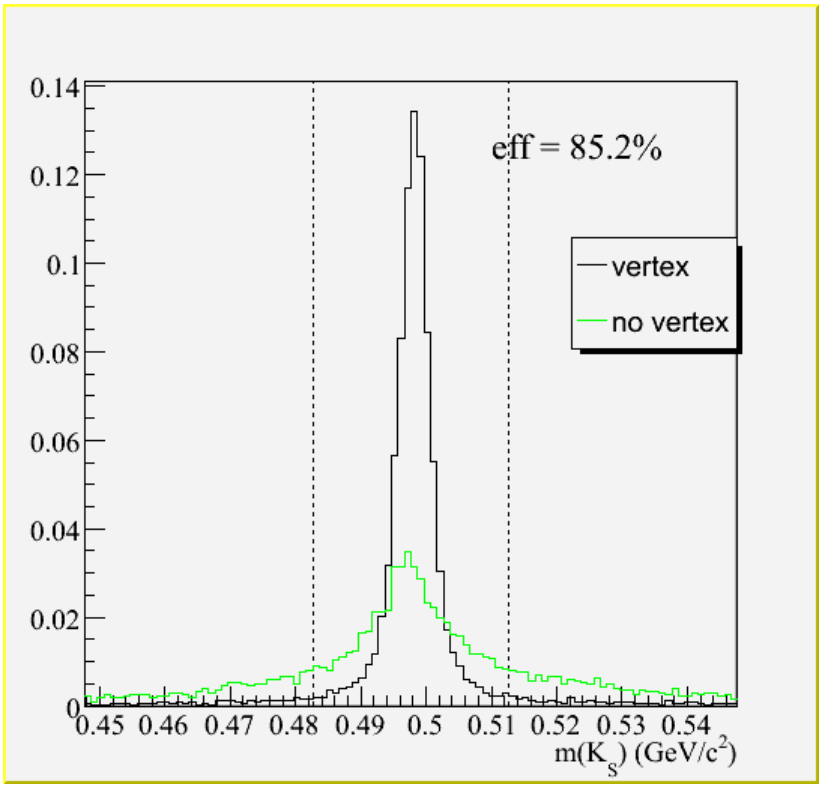


Figure 3.8: Normalized distributions of $m(K_S)$ before (green) and after (black) a vertex constraint.

Figure 3.8 shows the improvement in the mass resolution when a vertex constraint is used. This improvement allows us to cut away most of the background. To further

reduce the background due to bad fits from random combinations of charged tracks, we cut on $P(\chi^2) \geq 0.001$ which removes about 50% of the bad fits with a 14% drop in signal.

3.8 Selection of *D*

We use $D^0 \to K^-\pi^+$, $D^0 \to K^-\pi^+\pi^0$, $D^0 \to K^-\pi^+\pi^+\pi^-$, $D^0 \to K_S\pi^+\pi^-$, $D^+ \to K^-\pi^+\pi^-$ and $D^+ \to K_S\pi^+$ decays to make *D* candidates. These decays are chosen because of their high branching fractions and low backgrounds compared to the other *D* decays. Selected $\pi^\pm$ (from GoodTracksLoose), π^0 (from Pi0Default), $K^\pm$ and K_S candidates are combined in the following order:

- Vertex constraint
- $|\Delta m| \leq 20$ MeV/c^2, except for $D^0 \to K^-\pi^+\pi^0$ where $|\Delta m| \leq 40$ MeV/c^2
- Mass constraint
- $1.2 \leq p_{CM} \leq 2.2$ GeV/c
- $P(\chi^2) \geq 0.001$

Figure 3.9 shows the mass resolution and the efficiency of the mass cut for each decay. The wider resolution for $D^0 \to K^-\pi^+\pi^0$ is due to the poorer energy resolution of π^0. As we will see later, the mass constraint improves the energy resolution of the *B* candidates. For *B* decaying to two charm mesons, the CM momentum of the daughters is around 1.8 GeV/c (see Figure 3.10). Hence, the CM momentum cut reduces the number of charm mesons from $c\bar{c}$ events, whose CM momenta are much higher on average. The $P(\chi^2)$ cut further lowers the background due to random combinations.

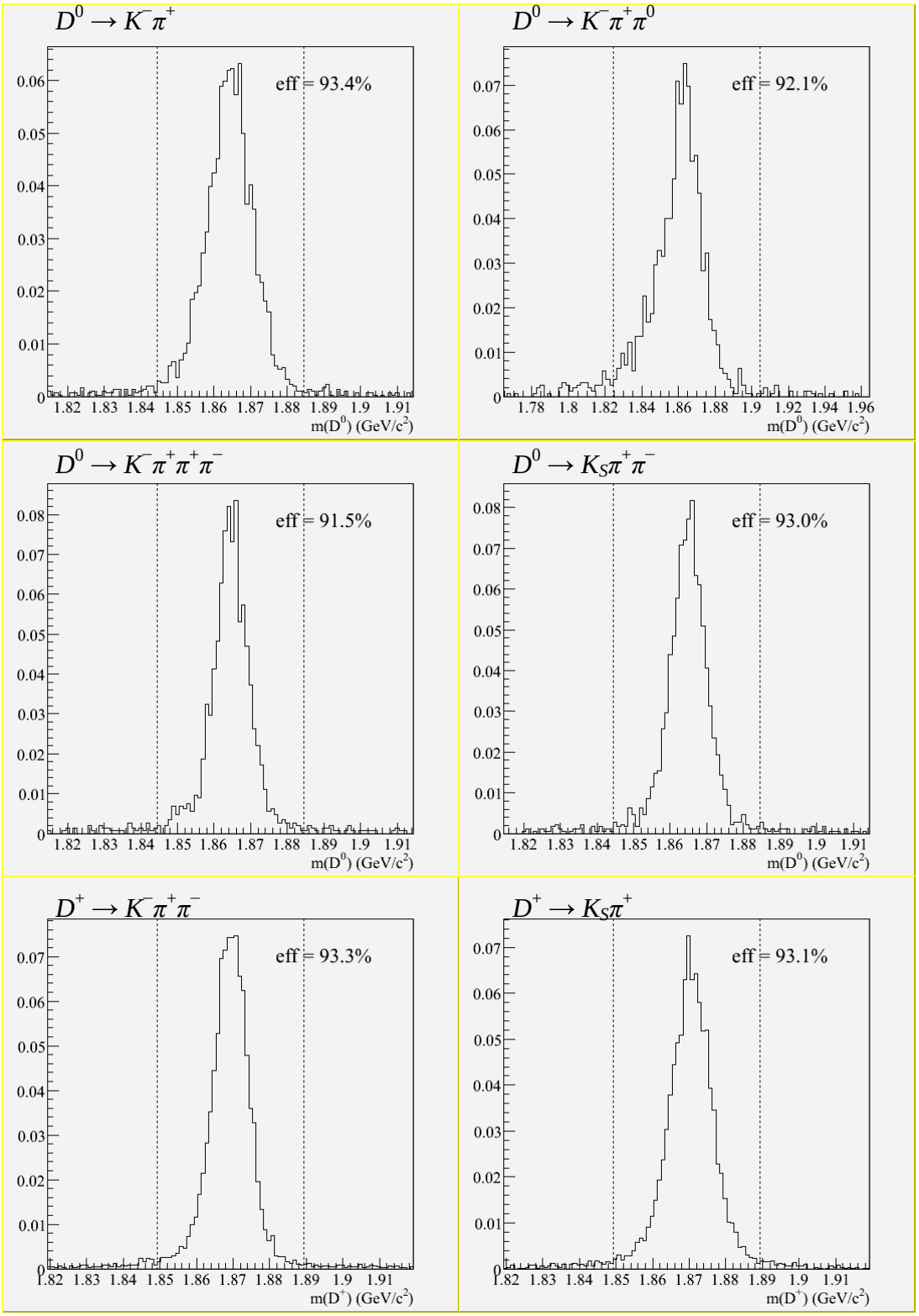


Figure 3.9: Normalized distribution of $m(D)$ for each D decay.

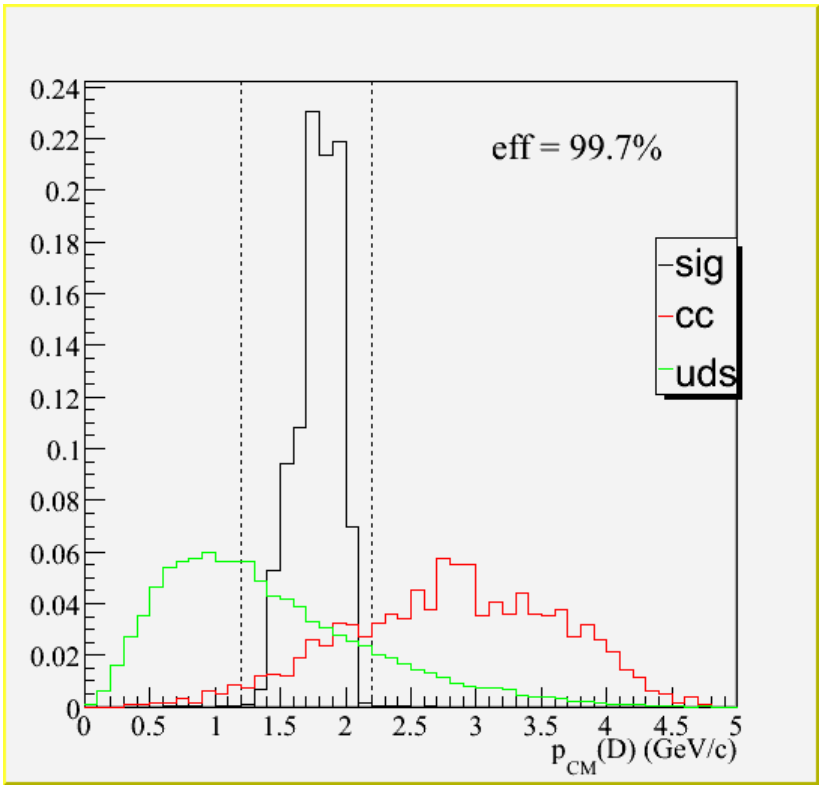


Figure 3.10: Normalized distributions of p_{CM} for real D decays from signal and $c\bar{c}$ events, and fake D decays from *uds* events.

For decays involving K_S, we use the long lifetime of the K_S meson to reduce the background due to fake K_S candidates:

- K_S flight length divided by its error $(l/\sigma_l) \geq 3$ (Figure 3.11, left plot)

For $D^0 \rightarrow K^-\pi^+\pi^0$, the decay proceeds mainly through the resonances $\rho^+(\pi^+\pi^0)$, $K^{*-}(K^-\pi^0)$ and $K^{*0}(K^+\pi^-)$ [20]. Hence, the decay probability for a resonant $D^0 \rightarrow K^-\pi^+\pi^0$ decay is relatively much higher than its non-resonant decay. Thus, we define a variable, known as the Dalitz weight, based on the decay probability by normalizing its maximum to one. The decay model is taken from [24]. To select mainly resonant decays and reduce the mostly non-resonant combinatorial background, we apply the cut:

- Dalitz weight ≥ 0.06 (Figure 3.11, right plot)

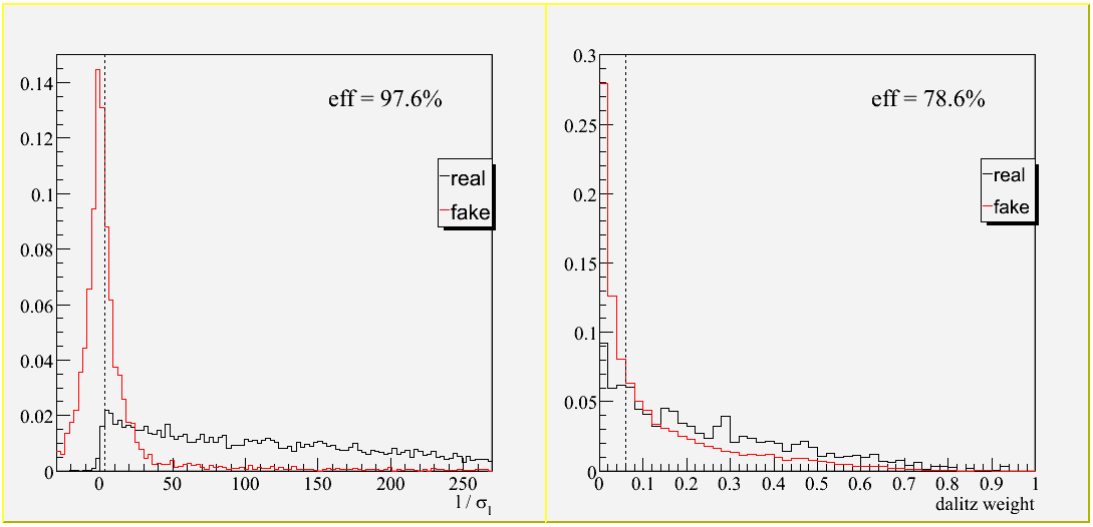


Figure 3.11: Left: Normalized distributions of l/σ_l for real (black) and fake (red) K_S candidates in $D^0 \to K_S\pi^+\pi^-$ and $D^+ \to K_S\pi^+$ reconstruction. Right: Normalized distribution of Dalitz weight for real (black) and fake (red) $D^0 \to K^-\pi^+\pi^0$ candidates.

3.9 Selection of D^*

D^* candidates are constructed from $D^{*+} \to D^0\pi^+$, $D^{*+} \to D^+\pi^0$, $D^{*0} \to D^0\pi^0$ and $D^{*0} \to D^0\gamma$ decays. The decay $D^{*+} \to D^+\gamma$ is not considered due to its low branching fraction of 1.6% [20]. Slow $\pi^\pm$ and π^0 candidates are taken from PionSoft and Pi0Soft respectively while the photon is taken from PhotonSoft. Note that the slow pions and photons have CM momenta less than 450 MeV/c each (see Figure 3.12). All the decays are reconstructed using kinematic fits. For $D^{*+} \to D^0\pi^+$, a vertex constraint is also used. Define $\delta M \equiv m(D^*) - m(D)$. Then, the following mass cuts are applied:

- $139.6 \leq \delta M \leq 151.3$ MeV/c^2 for $D^{*+} \to D^0\pi^+$
- $135.0 \leq \delta M \leq 146.3$ MeV/c^2 for $D^{*+} \to D^+\pi^0$

- $135.0 \leq \delta M \leq 149.3$ MeV/c^2 for $D^{*0} \rightarrow D^0\pi^0$
- $100.0 \leq \delta M \leq 170.0$ MeV/c^2 for $D^{*0} \rightarrow D^0\gamma$

The lower limit of δM for each of the first three decays is due to the mass of the slow pion. Since D is mass-constrained, $\delta M - \delta M_{\mathrm{PDG}} = \Delta m(D^*)$, and the above cuts correspond to $|\Delta m| \leq 5.85$ MeV/c^2 for $D^{*+} \rightarrow D^0\pi^+$, $|\Delta m| \leq 5.66$ MeV/c^2 for $D^{*+} \rightarrow D^+\pi^0$ and $|\Delta m| \leq 7.14$ MeV/c^2 for $D^{*0} \rightarrow D^0\pi^0$. From Figure 3.13, the efficiency for $D^{*0} \rightarrow D^0\gamma$ is 92% and around 98% for the other decays.

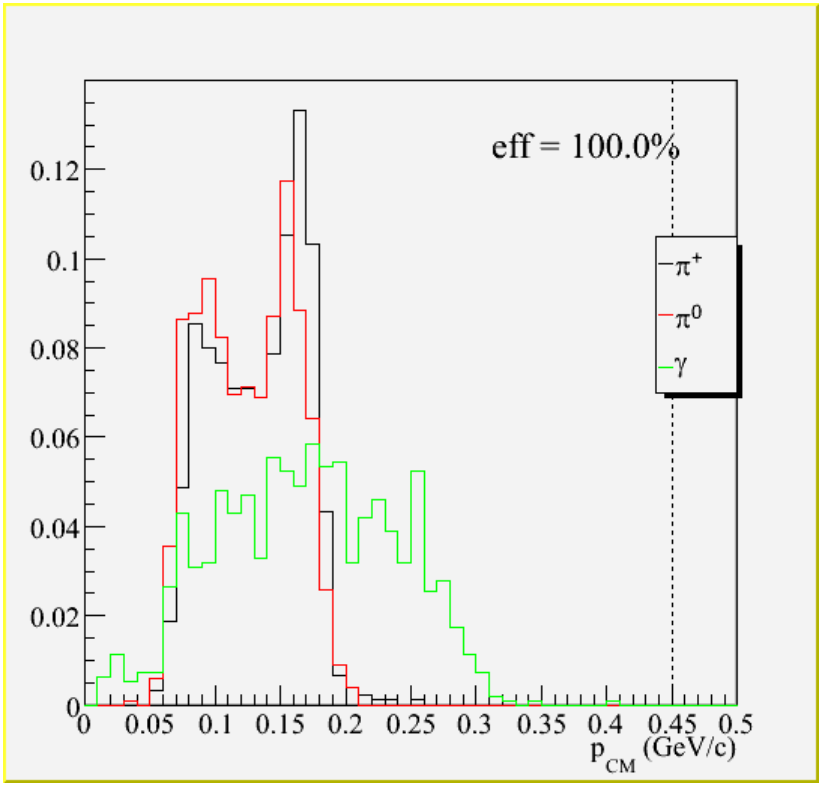


Figure 3.12: Normalized distributions of p_{CM} for π and γ daughters of D^*.

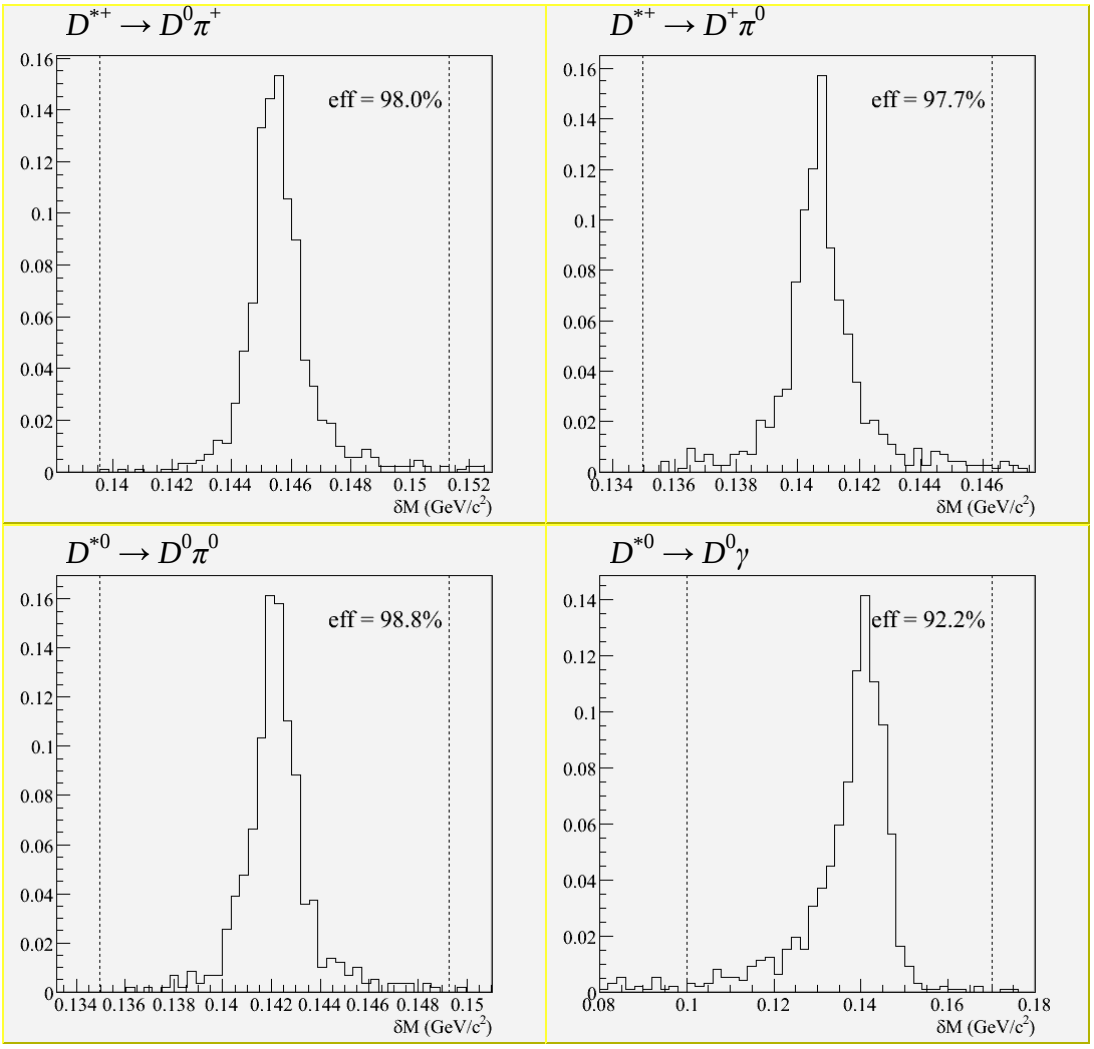


Figure 3.13: Normalized distribution of δM for each D^* decay.

3.10 Selection of *B*

A *B* candidate is constructed from two $D^{(*)}$ candidates that have passed the selections described in the above sections. Instead of cutting on its mass and energy directly, we make use of the CM beam energy to improve on the mass and energy resolutions. Hence, we define the following two variables in the CM frame:

$$m_{\mathrm{ES}} = \sqrt{E_{\mathrm{beam}}^2 - p_B^2}\ ,\ \Delta E = E_B - E_{\mathrm{beam}}\,. \tag{3.5}$$

Figure 3.14 shows the improvement in ΔE resolution when we mass-constrain the D candidates. Although for our signal, m_{ES} ranges from 5.27 to 5.29 GeV/c^2 and ΔE from −50 to 50 MeV, we cut loosely on m_{ES} and ΔE for the possibility of fitting the m_{ES} distribution and studying the background. Thus, the B candidates are selected as such:

- Vertex constraint
- $5.2 \leq m_{\mathrm{ES}} \leq 5.3$ GeV/c^2
- $\Delta E \leq 0.2$ GeV

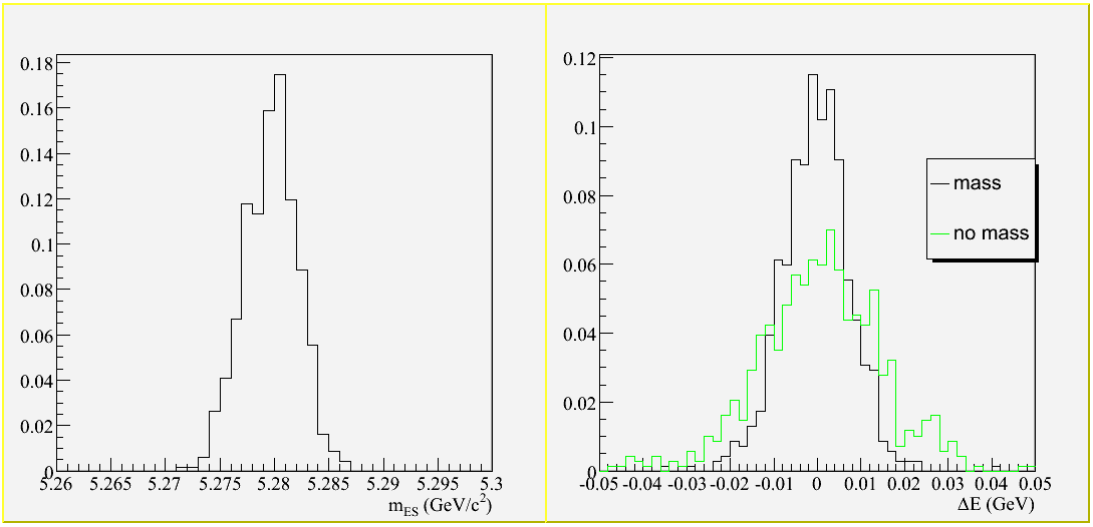


Figure 3.14: Normalized distributions of m_{ES} (left) and ΔE (right) for real $B^0 \rightarrow D^+D^-$ candidates. The green histogram shows the ΔE distribution when the daughters $D^{\pm}$ are not mass-constrained.

As you can see, there is still a lot of background left in our list of B candidates. In the next chapter, we will describe our procedure to optimize the signal significance and improve the signal-to-background ratio.

4 Optimization of Selection Variables

4.1 Definitions

We begin by introducing two terms that will be used throughout this dissertation: "*B* mode" and "sub-mode". A *B* mode refers to one of the *B* decays like $B^0 \rightarrow D^{*+}D^-$. A sub-mode refers to a specific decay tree like $B^0 \rightarrow D^{*+}D^- \rightarrow (D^0\pi^+)\,(K_S\pi^-) \rightarrow ((K^-\pi^+\pi^0)\pi^+)\,(K_S\pi^-)$, where one daughter D^{*+} decays to $D^0\pi^+$ with D^0 decaying to $K^-\pi^+\pi^0$, and the other daughter D^- decays to $K_S\pi^-$. Hence, a *B* mode can have several sub-modes[9].

Moreover, when we mention the branching fraction of a *B* mode, we are referring to the parent *B* decay (with the $D^{(*)}$ daughters decaying to all possible modes), while the branching fraction of a sub-mode is the product of the branching fractions of the two $D^{(*)}$ decays taken from Table 4.1 (excluding the *B* decay's branching fraction). For example, the branching fraction of the sub-mode given in the first paragraph is

$$0.677 \times 0.131 \times 0.0095 = 8.4 \times 10^{-4}.$$

[9] The word "mode" alone can refer to a *B* mode or a sub-mode, and will be used interchangeably with the word "decay".

D^* Decays	BF (%)	D Decays	BF (%)
$D^{*+} \to D^0\pi^+$	67.7 ± 0.5	$D^0 \to K^-\pi^+$	3.8 ± 0.09
$D^{*+} \to D^+\pi^0$	30.7 ± 0.5	$D^0 \to K^-\pi^+\pi^0$	13.1 ± 0.9
$D^{*0} \to D^0\pi^0$	61.9 ± 2.9	$D^0 \to K^-\pi^+\pi^+\pi^-$	7.46 ± 0.31
$D^{*0} \to D^0\gamma$	38.1 ± 2.9	$D^0 \to K_S\pi^+\pi^-$	2.03 ± 0.12
		$D^+ \to K^-\pi^+\pi^-$	9.1 ± 0.6
		$D^+ \to K_S\pi^+$	0.95 ± 0.06

Table 4.1: Branching fractions (BF) of $D^{(*)}$ decays that we reconstruct [20]. The branching fraction of $K_S \to \pi^+\pi^-$ (0.6895 ± 0.0014) and a factor of half from $K^0 \to K_S$ are included for the decays containing K_S.

4.2 Event Samples

The events that we use to optimize our selection variables come from simulation or Monte Carlo (MC). Two types of simulated events are generated. The first, known as signal MC, represents the signal decays that we reconstruct. The second, called generic MC, reproduces the background from all other decays. On average, about 1.2×10^5 signal MC events are produced per sub-mode. For generic MC, the breakdown is shown below.

Event Type	Events Generated ($\times 10^6$)	Relative to Real Data
$B^0\overline{B}^0$	547	4.7
B^+B^-	539	4.7
$c\overline{c}$	418	1.5
$u\overline{u}$ + $d\overline{d}$ + $s\overline{s}$ (uds)	716	1.6

Table 4.2: Breakdown of generic MC events.

4.3 Selection Variables

Besides ΔE that is used in our pre-selection, there are three other variables that we look at. The first variable, known as the mass likelihood, is defined based on the masses of the $D^{(*)}$ daughters:

$$\begin{aligned} L_{\text{mass}} &= G\left(z_m(D);0,1\right)\times G\left(z_m(\overline{D});0,1\right) \\ &\times\left[fG\left(\Delta m(D^*);\mu_1,\sigma_1\right)+(1-f)G\left(\Delta m(D^*);\mu_2,\sigma_2\right)\right], \\ &\times\left[\bar{f}G\left(\Delta m(\overline{D}^*);\overline{\mu}_1,\overline{\sigma}_1\right)+(1-\bar{f})G\left(\Delta m(\overline{D}^*);\overline{\mu}_2,\overline{\sigma}_2\right)\right] \end{aligned} \tag{4.1}$$

where $z_m = \Delta m / s_m$, s_m is the mass uncertainty from the vertex-constrained fit;

$\Delta m = m - m_{\text{PDG}}$;

$0 < f, \overline{f} < 1$; and

$G(x; \mu, \sigma)$ is a Gaussian probability density function with mean μ and width σ:

$$G(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{4.2}$$

Instead of cutting on the $D^{(*)}$ masses directly, we combine the information into a likelihood variable which is more efficient in rejecting background. We divide $\Delta m(D)$ by

its uncertainty so that the same model (standard Gaussian with $\mu = 0$ and $\sigma = 1$) can be used for all D decays. For the D^* mass, it is modeled by a sum of two Gaussians with one having a wider width to accommodate the tails in the distribution (see Figure 3.13). In the case where a D^* meson is absent from a B mode, the corresponding G term is set to one. We assume that $z_m(D)$ and $\Delta m(D^*)$ are independent. This is supported by Figure 4.1. The values of f, μ_1, μ_2, σ_1 and σ_2 depend on the D^* decay. They are given in Table 4.3. For selection purposes, we use the variable $-\log(L_{\text{mass}})$, where a smaller value indicates a "better" B candidate.

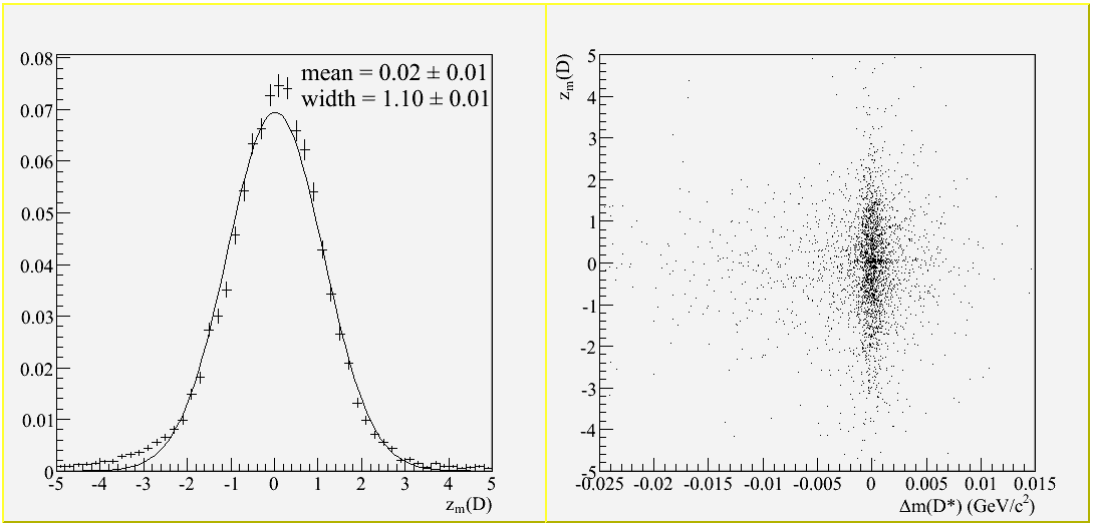


Figure 4.1: Left: Normalized distribution of $z_m(D)$ for all D decays in signal MC. The curve is a fit to a Gaussian with mean and width given in the plot. Right: A scatter plot of $z_m(D)$ vs. $\Delta m(D^*)$.

D^* Decay	f	μ_1 (MeV/c^2)	μ_2 (MeV/c^2)	σ_1 (MeV/c^2)	σ_2 (MeV/c^2)
$D^{*+} \to D^0\pi^+$	0.6623	0.000	0.260	0.3224	1.513
$D^{*+} \to D^+\pi^0$	0.5289	0.000	0.161	0.5775	1.509
$D^{*0} \to D^0\pi^0$	0.5489	−0.105	0.126	0.6354	1.733
$D^{*0} \to D^0\gamma$	0.6517	0.000	−6.83	4.527	12.51

Table 4.3: Parameters used to model $\Delta m(D^*)$ distributions as determined from signal MC.

The second variable exploits our ability to distinguish the long D lifetime. Denoted as l_D, it is the sum of the D candidates' flight lengths divided by the sum's uncertainty. The flight lengths and uncertainty are obtained from fitting the decay tree of B via the method described in [25], so that the correlation between the flight lengths is accounted for properly in the uncertainty. Note that the flight length is allowed to be negative in the fit. In this way, fake $B \to D^{(*)}\bar{D}^{(*)}$ decays will have an average l_D of zero while real decays will have a positive mean (see Figure 4.2).

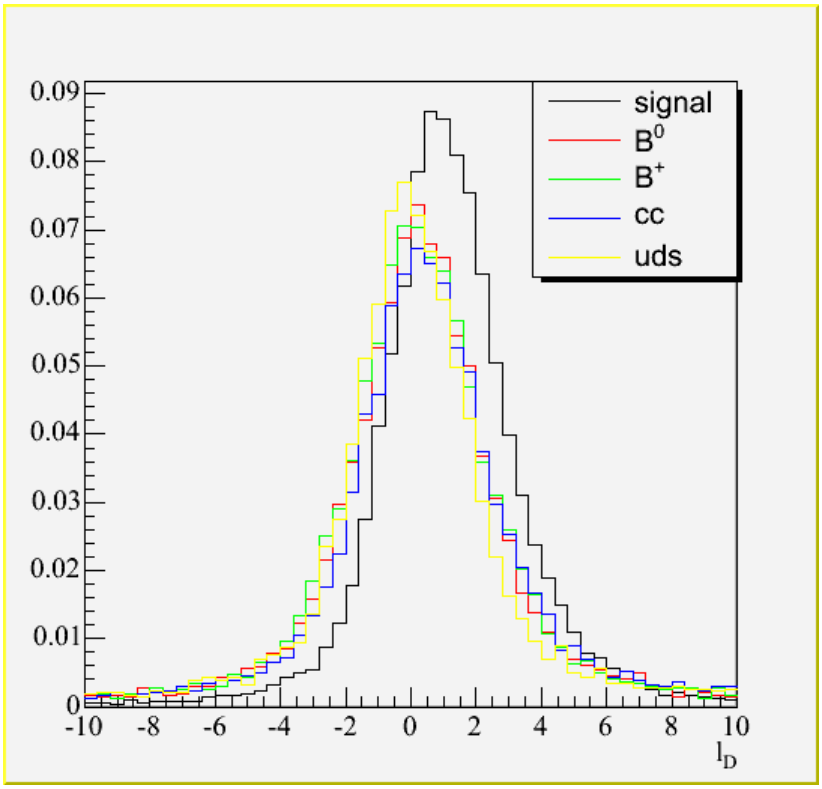


Figure 4.2: Normalized distributions of l_D for reconstructed $B^0 \rightarrow D^0\overline{D}^0$ events from signal and generic MC.

The third variable is a combination of the event's shape variables, namely, the *B*'s momentum, its thrust vector and the momentum distribution of the rest-of-event (tracks and showers that do not belong to *B*). The thrust vector **T** of *B* is defined such that

$$T = \max_{\{\mathbf{A};|\mathbf{A}|=1\}} \left\{ \sum_i |\, \mathbf{p}_i \cdot \mathbf{A} \,| \right\} \Big/ \sum_i |\, \mathbf{p}_i \,|, \qquad \mathbf{T} = T\mathbf{A}_{\max}, \tag{4.3}$$

where $\mathbf{p}_i$ is the momentum of a *B*'s daughter in CM frame; and

$\mathbf{A}_{\max}$ is the unit vector that maximizes the above sum, called the thrust axis.

To describe the momentum distribution of the rest-of-event, we consider the momentum flows in nine equal polar angular intervals of 10° each around $\mathbf{A}_{\max}$ (see

Figure 4.3). The momentum flow in each interval is the scalar sum of the CM momenta of all tracks and showers (not from the *B* candidate) pointing in that interval.

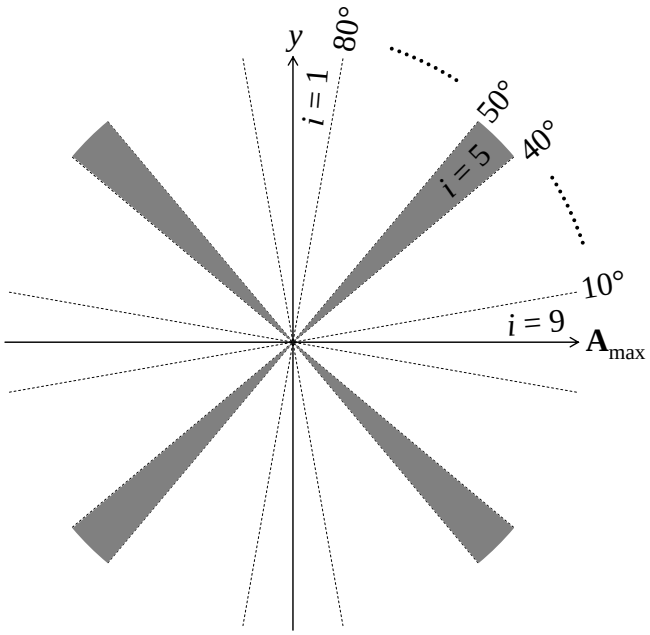


Figure 4.3: Illustration of the polar angular intervals. Imagine the cones formed by rotating the dashed lines of 10°, 20°, …, 80° about the thrust axis $\mathbf{A}_{\max}$ of *B*. The shaded region between the cones (on both sides of the *y*-axis) given by the 40° and 50° lines is the fifth polar angular interval.

For a better separation between signal and $c\bar{c}$ background, we combine the shape variables into a Fisher's linear discriminant [26]:

$$F = c_0 + \sum_{i=1}^{11} c_i x_i \ , \tag{4.4}$$

where x_i, i = 1, …, 9, is the momentum flow in the interval i;

$x_{10} = |\cos(\theta_{\mathbf{T}})|$, $\theta_{\mathbf{T}}$ is the angle between $\mathbf{A}_{\max}$ and the beam axis; and

$x_{11} = |\cos(\theta_B)|$, θ_B is the angle between the *B*'s CM momentum and the beam axis.

The coefficients c_i are determined from a separate study and shown in Table 4.4. The separation between signal and background can be seen in Figure 4.4.

i	0	1	2	3	4	5
c_i	2.00000	−0.26033	−0.17384	−0.16185	0.00003	0.06730
i	6	7	8	9	10	11
c_i	0.37850	0.69457	0.78548	0.48972	0.71547	0.86215

Table 4.4: Coefficients of F. The polar angle decreases from i = 1 to i = 9.

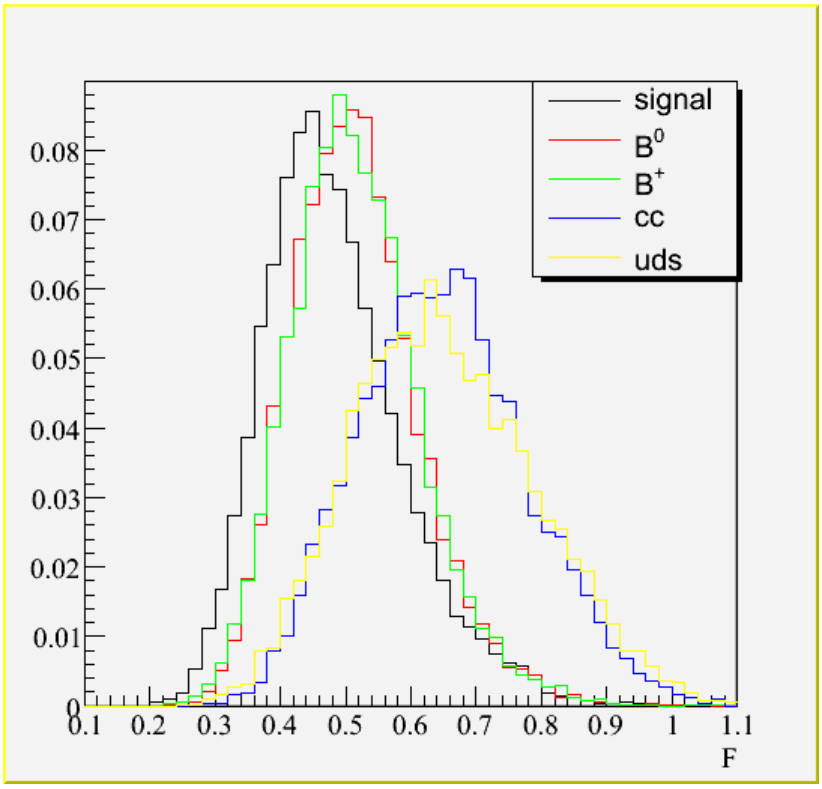


Figure 4.4: Normalized distributions of F for reconstructed $B^0 \rightarrow D^0\bar{D}^0$ events from signal and generic MC.

4.4 Procedure

The optimization of the selection variables is performed by minimizing a function f (usually taken to be the fractional uncertainty of the branching fraction) with respect to the cut variables $|\Delta E|$, $-\log(L_{\text{mass}})$, $-l_D$ and F. For $B^0 \to D^{(*)0}\overline{D}^{(*)0}$, where we expect to see no signal, we use the formula given in [27]. For the other B modes, we use the square of the expected fractional uncertainty in the branching fraction. Therefore,

$$f = \begin{cases} \dfrac{1+\sqrt{N_{\text{bkg}}}}{N_{\text{sig}}}, & \text{for } B^0 \to D^{(*)0}\overline{D}^{(*)0} \\ \dfrac{N_{\text{sig}}+N_{\text{bkg}}}{N_{\text{sig}}^2}, & \text{for other } B \text{ modes} \end{cases}, \tag{4.5}$$

where N_{sig} (N_{bkg}) is the expected number of signal (background) events in data.

From signal MC, we can obtain N_{sig} for each B mode via the formula

$$N_{\text{sig}} = N_B \mathcal{B} \sum_{\text{sub-mode } i} n_i \mathcal{b}_i / N_i \,, \tag{4.6}$$

where $N_B = 2.32 \times 10^8$ is the number of $B\overline{B}$ events in data;

$\mathcal{B}$ is the branching fraction of the B mode;

i denotes a sub-mode of B;

N_i is the number of generated events for sub-mode i;

$\mathcal{b}_i$ is the branching fraction of sub-mode i; and

n_i is the number of MC events in the signal region for sub-mode i.

Because each signal MC sample is generated with the $D^{(*)}$ mesons decaying solely to a given sub-mode, we multiply n_i by $\mathcal{b}_i$ to match the actual number in data. For N_{bkg}, a similar formula is applied where we scale the numbers of events in the signal region in generic MC to the expected numbers in data. The signal region is defined as the interval

$m_{ES} \geq 5.27$ GeV/c^2 after all the cuts are applied. A B candidate passes a cut if its cut variable is less than the cut value. In an event when more than one B candidate per B mode passes all the cuts, the one with the lowest $-\log(L_{mass})$ is selected.

In the optimization, since the $D^{(*)}$ mass and ΔE resolutions vary across the sub-modes, different $|\Delta E|$ and $-\log(L_{mass})$ cut values are used for each sub-mode. To reduce possible differences in resolutions between data and MC, a lower bound is set for each cut value such that its efficiency is at least 95%. We also set an upper bound that is two times higher than the lower bound to avoid keeping too much background. For $-l_D$, only one cut value per B mode is used. The same is true for F. Therefore, taking $B^0 \rightarrow D^+D^-$ as an example, where there are three sub-modes (two decays for a $D^{\pm}$ meson), a total of eight cut values (three $|\Delta E|$ cuts, three $-\log(L_{mass})$ cuts, one $-l_D$ cut and one F cut) are used to minimize f.

At the end of each minimization, we check if a sub-mode is useful by comparing f before and after removing the mode. We keep a sub-mode if f is higher after removal. We also check if it is beneficial to keep a $-l_D$ or F cut, since each cut will have an associated systematic uncertainty. We keep such a cut if it decreases the expected fractional uncertainty $\sqrt{N_{sig} + N_{bkg}} \big/ N_{sig}$ by more than 3%.

Hence, the optimization for each B mode consists of a series of minimization and cuts-checking until all the cuts are stable.

Note that, in Equation (4.6), we have to assume a branching fraction for the B mode which is either taken from a previous experiment or from theory. To test the reliability of the cuts obtained, we repeat the optimization assuming different values for the branching fractions. We also increase N_{bkg} by 10% or 20% in case of a possible

difference in the background estimation between data and MC. We find that the cuts remain stable after all these tests.

4.5 Results

We summarize our optimized cut values in the tables below. A B candidate is selected if all its selection variables ($|\Delta E|$, $-\log(L_{\text{mass}})$, $-l_D$ and F) are less than the given cut values. The lower bounds that are used in the optimization for the cut values of $|\Delta E|$ and $-\log(L_{\text{mass}})$ are also shown in Table 4.5 to Table 4.14. In these tables, we employ a simplified notation to represent the decay tree of a sub-mode. For example, the decay tree $B^0 \to D^{*+}D^- \to (D^0\pi^+)\,(K_S\pi^-) \to ((K^-\pi^+\pi^0)\pi^+)\,(K_S\pi^-)$ is simply written as $(K\pi\pi^0)\pi\,(K_S\pi)$. Finally, we show the expected signal and background events in the last table. From now on, events from data and MC that have passed all the cuts will be used for our analysis.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi)\,(K\pi)$	15.1	15.1	8.8	4.4
$(K\pi)\,(K\pi\pi^0)$	18.7	18.7	5.2	5.2
$(K\pi)\,(K\pi\pi\pi)$	16.1	15.5	6.8	5.6
$(K\pi)\,(K_S\pi\pi)$	15.6	15.6	5.6	5.6
$(K\pi\pi^0)\,(K\pi\pi\pi)$	19.0	19.0	6.2	6.2
$(K\pi\pi\pi)\,(K\pi\pi\pi)$	15.9	15.8	6.9	6.8

Table 4.5: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^0 \to D^0\bar{D}^0$.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi)\pi^0$ $(K\pi)$	17.4	16.3	7.6	6.2
$(K\pi)\pi^0$ $(K\pi\pi^0)$	19.6	19.6	6.7	6.6
$(K\pi)\pi^0$ $(K\pi\pi\pi)$	17.4	16.7	7.2	7.2
$(K\pi\pi^0)\pi^0$ $(K\pi)$	19.5	19.4	6.8	6.8
$(K\pi\pi\pi)\pi^0$ $(K\pi)$	16.8	16.8	7.4	7.4
$(K\pi)\gamma$ $(K\pi)$	24.5	24.4	5.8	5.8

Table 4.6: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^0 \to D^{*0}\bar{D}^0$.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi)\pi^0$ $(K\pi)\pi^0$	35.1	17.7	12.2	7.8
$(K\pi)\pi^0$ $(K\pi\pi^0)\pi^0$	23.0	20.3	8.4	8.4
$(K\pi)\pi^0$ $(K\pi\pi\pi)\pi^0$	27.3	17.9	9.6	9.4
$(K\pi)\gamma$ $(K\pi)\pi^0$	25.5	25.4	7.6	7.6

Table 4.7: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^0 \to D^{*0}\bar{D}^{*0}$.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi\pi)$ $(K_S\pi)$	15.5	15.0	6.3	5.4
$(K\pi\pi)$ $(K\pi\pi)$	15.1	15.0	6.0	5.6

Table 4.8: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^0 \to D^+D^-$.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi)\pi$ $(K_S\pi)$	17.2	16.4	8.0	7.2
$(K\pi)\pi$ $(K\pi\pi)$	18.0	16.2	8.2	7.4
$(K\pi\pi^0)\pi$ $(K\pi\pi)$	19.4	19.4	8.4	8.4
$(K\pi\pi\pi)\pi$ $(K\pi\pi)$	18.6	16.7	8.8	8.8
$(K_S\pi\pi)\pi$ $(K\pi\pi)$	21.9	16.4	9.6	8.6

Table 4.9: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^0 \to D^{*+}D^-$.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi)\pi$ $(K\pi)\pi$	35.5	18.3	13.0	9.0
$(K\pi)\pi$ $(K\pi\pi^0)\pi$	33.8	21.3	12.0	9.8
$(K\pi)\pi$ $(K\pi\pi\pi)\pi$	30.4	19.1	17.3	10.4
$(K\pi)\pi$ $(K_S\pi\pi)\pi$	35.2	18.6	19.8	10.0
$(K\pi)\pi$ $(K\pi\pi)\pi^0$	25.5	17.6	10.5	8.6
$(K\pi\pi^0)\pi$ $(K\pi\pi^0)\pi$	34.5	23.9	10.6	10.4
$(K\pi\pi^0)\pi$ $(K\pi\pi\pi)\pi$	29.6	21.5	11.0	11.0
$(K\pi\pi^0)\pi$ $(K_S\pi\pi)\pi$	23.5	21.2	18.3	11.0
$(K\pi\pi^0)\pi$ $(K\pi\pi)\pi^0$	27.4	20.4	9.5	9.4
$(K\pi\pi\pi)\pi$ $(K\pi\pi\pi)\pi$	23.5	19.3	11.7	11.4
$(K\pi\pi\pi)\pi$ $(K\pi\pi)\pi^0$	18.2	18.2	9.8	9.8
$(K_S\pi\pi)\pi$ $(K\pi\pi\pi)\pi$	23.7	19.4	11.0	11.0

Table 4.10: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^0 \to D^{*+}D^{*-}$.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi)$ $(K_S\pi)$	16.4	15.0	9.6	4.8
$(K\pi)$ $(K\pi\pi)$	15.5	15.2	7.3	5.0
$(K\pi\pi^0)$ $(K\pi\pi)$	19.2	19.2	5.8	5.8
$(K\pi\pi\pi)$ $(K\pi\pi)$	15.4	15.4	6.5	6.4
$(K_S\pi\pi)$ $(K\pi\pi)$	15.5	15.4	6.2	6.2

Table 4.11: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^- \to D^- D^0$.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi)\pi$ $(K\pi)$	18.1	16.3	8.6	7.0
$(K\pi)\pi$ $(K\pi\pi^0)$	20.2	19.9	8.5	7.6
$(K\pi)\pi$ $(K\pi\pi\pi)$	17.1	16.6	8.2	8.2
$(K\pi)\pi$ $(K_S\pi\pi)$	19.0	16.7	8.0	8.0
$(K\pi\pi^0)\pi$ $(K\pi)$	25.9	19.5	7.8	7.8
$(K\pi\pi^0)\pi$ $(K\pi\pi^0)$	22.5	22.5	7.8	7.8
$(K\pi\pi^0)\pi$ $(K\pi\pi\pi)$	19.5	19.5	8.8	8.8
$(K\pi\pi\pi)\pi$ $(K\pi)$	19.0	16.8	9.3	8.4
$(K\pi\pi\pi)\pi$ $(K\pi\pi^0)$	20.4	20.3	9.4	9.2
$(K\pi\pi\pi)\pi$ $(K\pi\pi\pi)$	17.1	17.1	9.0	9.0
$(K_S\pi\pi)\pi$ $(K\pi)$	16.9	16.7	15.1	8.0
$(K_S\pi\pi)\pi$ $(K\pi\pi^0)$	19.7	19.7	9.2	9.0
$(K\pi\pi)\pi^0$ $(K\pi)$	16.4	16.3	6.6	6.6

Table 4.12: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^- \to D^{*-}D^0$.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi)\pi^0$ $(K_S\pi)$	23.5	16.4	10.3	6.6
$(K\pi)\pi^0$ $(K\pi\pi)$	25.9	16.4	9.9	6.8
$(K\pi\pi^0)\pi^0$ $(K_S\pi)$	19.5	19.4	7.0	7.0
$(K\pi\pi^0)\pi^0$ $(K\pi\pi)$	20.0	19.9	7.5	7.4
$(K\pi\pi\pi)\pi^0$ $(K\pi\pi)$	17.1	17.1	8.0	8.0
$(K_S\pi\pi)\pi^0$ $(K\pi\pi)$	16.6	16.6	9.2	8.2
$(K\pi)\gamma$ $(K\pi\pi)$	24.6	24.6	6.4	6.4

Table 4.13: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^- \to D^- D^{*0}$.

Sub-Mode	$\|\Delta E\|$ (MeV)		$-\log(L_{\text{mass}})$	
	Cut Value	Lower Bound	Cut Value	Lower Bound
$(K\pi)\pi^0$ $(K\pi)\pi$	35.7	17.9	14.6	8.2
$(K\pi)\pi^0$ $(K\pi\pi^0)\pi$	40.9	20.5	11.5	9.4
$(K\pi)\pi^0$ $(K\pi\pi\pi)\pi$	34.0	18.0	11.7	9.6
$(K\pi)\pi^0$ $(K\pi\pi)\pi^0$	19.1	17.3	8.2	8.2
$(K\pi\pi^0)\pi^0$ $(K\pi)\pi$	21.0	20.8	17.5	8.8
$(K\pi\pi^0)\pi^0$ $(K\pi\pi^0)\pi$	23.9	23.2	9.8	9.8
$(K\pi\pi^0)\pi^0$ $(K\pi\pi\pi)\pi$	30.6	20.5	9.6	9.6
$(K\pi\pi\pi)\pi^0$ $(K\pi)\pi$	26.0	18.0	9.2	9.2
$(K\pi\pi\pi)\pi^0$ $(K\pi\pi^0)\pi$	21.4	19.4	10.7	9.6
$(K\pi\pi\pi)\pi^0$ $(K\pi\pi\pi)\pi$	20.6	18.5	10.4	10.4
$(K\pi)\gamma$ $(K\pi)\pi$	43.6	26.8	8.9	8.0
$(K\pi)\gamma$ $(K\pi\pi^0)\pi$	29.3	28.2	8.7	8.6
$(K\pi)\gamma$ $(K\pi\pi\pi)\pi$	27.3	27.0	9.0	9.0

Table 4.14: Optimized $|\Delta E|$ and $-\log(L_{\text{mass}})$ cut values for the sub-modes of $B^- \to D^{*-}D^{*0}$.

B Mode	$-l_D$		F	
	Cut Value	Efficiency (%)	Cut Value	Efficiency (%)
$B^0 \to D^0\bar{D}^0$	0.4	89.4	0.51	71.0
$B^0 \to D^{*0}\bar{D}^0$	0.4	88.0	0.53	75.6
$B^0 \to D^{*0}\bar{D}^{*0}$	1.6	97.3	0.60	89.6
$B^0 \to D^+D^-$	−1.3	82.5	0.62	90.8
$B^0 \to D^{*+}D^-$	–	100	–	100
$B^0 \to D^{*+}D^{*-}$	–	100	–	100
$B^- \to D^-D^0$	−0.5	83.2	0.53	75.0
$B^- \to D^{*-}D^0$	1.3	96.6	0.53	75.7
$B^- \to D^-D^{*0}$	0.0	88.2	0.53	75.3
$B^- \to D^{*-}D^{*0}$	–	100	0.60	88.6

Table 4.15: Optimized $-l_D$ and F cut values and their efficiencies for each B mode. A dash indicates a cut is not applied.

B Mode	BF ($\times 10^{-4}$)	$N_{\rm sig}$	$N_{\rm bkg}$	$\frac{\sqrt{N_{\rm sig}+N_{\rm bkg}}}{N_{\rm sig}}$ (%)	$\frac{N_{\rm sig}}{N_{\rm sig}+N_{\rm bkg}}$ (%)
$B^0 \to D^0 \bar{D}^0$	0.1	3.9	246.4	406	1.6
$B^0 \to D^{*0} \bar{D}^0$	0.1	1.3	122.1	855	1.1
$B^0 \to D^{*0} \bar{D}^{*0}$	0.1	0.7	51.9	1040	1.4
$B^0 \to D^+ D^-$	1.6	44.0	53.3	22.4	45.2
$B^0 \to D^{*+} D^-$	8.0	236.3	159.5	8.42	59.7
$B^0 \to D^{*+} D^{*-}$	8.0	288.1	86.7	6.72	76.9
$B^- \to D^- D^0$	3.2	127.7	364.0	17.4	26.0
$B^- \to D^{*-} D^0$	3.2	131.0	246.2	14.8	34.7
$B^- \to D^- D^{*0}$	3.2	57.0	247.2	30.6	18.7
$B^- \to D^{*-} D^{*0}$	8.0	203.6	364.9	12.2	35.8

Table 4.16: Assumed branching fraction, and expected signal, background, fractional uncertainty and purity for each B mode.

5 Determination of Branching Fractions

In this chapter, we will show the steps involved in measuring the branching fractions of $B \to D^{(*)}\bar{D}^{(*)}$ decays. We will start by deriving an equation that will allow us to find the branching fractions of several decays simultaneously. Using this equation, we can systematically deal with two major problems that we encounter in this analysis: 1) the false reconstruction of events from one of our *B* modes to another; and 2) the background from decays other than our *B* modes. Then, we will describe our method of extracting the signal from our samples of events that have passed all the selections mentioned in the previous chapter in order to obtain the quantities appearing in the equation. In the last section, we will solve the equation to determine our branching fractions.

5.1 Theory

Let e_{ij} be the efficiency of reconstructing in mode i for an event generated in mode j. Then, the number of events reconstructed in mode i is

$$D_i = \sum_j e_{ij} N_j \,, \tag{5.1}$$

where N_j is the number of events generated in mode j.

We can split the generated events into two sets: one set containing the modes whose branching fractions we are measuring (signal modes); the other set containing all the other modes (generic modes). Let us denote the signal modes by S. Then, we can write Equation (5.1) as

$$\sum_{J \in S} e_{IJ} \mathcal{B}_J N + \sum_{J' \notin S} e_{IJ'} N_{J'} = D_I \,, \tag{5.2}$$

where $\mathcal{B}_J$ is the branching fraction of mode J; and

N is the total number of events generated.

Defining the vector $\mathbf{P} = [P_I]$ as

$$P_I \equiv \sum_{J' \notin S} e_{IJ'} N_{J'}, \qquad (I \in S), \tag{5.3}$$

and the matrix $\mathbf{C} = [C_{IJ}]$ as

$$C_{IJ} \equiv e_{IJ}, \qquad (I, J \in S), \tag{5.4}$$

Equation (5.2) becomes

$$N \sum_{J \in S} C_{IJ} \mathcal{B}_J = D_I - P_I, (I \in S). \tag{5.5}$$

We shall call P_I as the peaking background (the reason for this name will be clear later), and C_{IJ} as cross feed. P_I is the amount of background from generic decays while C_{IJ} is the amount of false reconstruction of events from one signal mode to another. To solve for the branching fractions, we simply invert the matrix **C** to get

$$\boldsymbol{\mathcal{B}} = \mathbf{C}^{-1}(\mathbf{D} - \mathbf{P}) / N . \tag{5.6}$$

If our MC is perfect, we can easily count the number of events in the signal region $m_{\mathrm{ES}} \geq 5.27$ GeV/c^2 to obtain **P** from generic MC, **C** from signal MC and **D** from real data. However, we choose not to count since there is a difference in the efficiencies between MC and data. Instead, we fit the m_{ES} distribution to extract the signal. This method is described below.

5.2 m_{ES} Fit

Our method of signal extraction is via a fit to the m_{ES} distribution known as the m_{ES} fit. The m_{ES} distribution can be thought of having two components. One is a signal

component from correctly reconstructed events, which is Gaussian in shape and peaks at the B mass. The other is a background component from badly reconstructed $c\bar{c}$, *uds* and generic $B\bar{B}$ events, whose shape can be described by the Argus function [28],

$$A(m_{\mathrm{ES}};\kappa,E_0)=\begin{cases}\sqrt{u}e^{-\kappa u}, u=1-(m_{\mathrm{ES}}/E_0)^2, & \text{for } m_{\mathrm{ES}}<E_0\\ 0, & \text{for } m_{\mathrm{ES}}\geq E_0\end{cases}. \tag{5.7}$$

Therefore, the function to fit the m_{ES} distribution is given by a sum of a Gaussian function and an Argus function:

$$f(m_{\mathrm{ES}})=\Delta m_{\mathrm{ES}}\left(N_{\mathrm{sig}}G(m_{\mathrm{ES}};\mu,\sigma)+N_{\mathrm{bkg}}A(m_{\mathrm{ES}};\kappa,E_0)\right). \tag{5.8}$$

This is known as the m_{ES} function and is defined on the interval $5.2\leq m_{\mathrm{ES}}\leq 5.3$ GeV/c^2.

In the fit to MC, the mean μ is fixed to 5.2795 GeV/c^2 for B^0 modes or 5.2792 GeV/c^2 for B^- modes. For data, it is fixed to 5.2798 GeV/c^2 for B^0 modes or 5.2796 GeV/c^2 for B^- modes. These values are obtained from a fit to common B modes in signal MC and in data.

The width σ is fixed in the fit for both MC and data. Each B mode has its own value which is determined from signal MC. For data, since there are not enough events to determine σ for each B mode, the same value in MC is used.

κ is called the shape parameter. It is fixed as per B mode in MC but free to vary in data. The fixed value is obtained from a fit to generic MC after fixing μ and σ.

E_0 is the end-point of the m_{ES} distribution; a limit determined by the beam energy. For purpose of fitting, we fix E_0 to 5.2891 GeV/c^2 in MC or 5.2895 GeV/c^2 in data. These values are obtained from a fit to all modes in generic MC or in data respectively.

Δm_{ES} is the bin width used in the fit and is fixed to 2 MeV/c^2, corresponding to fifty bins in the interval $5.2\leq m_{\mathrm{ES}}\leq 5.3$ GeV/c^2.

N_{sig} is called the signal yield. It can be interpreted as the number of signal events in the sample. N_{sig} is allowed to be negative in the fit.

N_{bkg} is the normalization constant for the Argus function. It is a free parameter and scales with the amount of background in the m_{ES} distribution. It is constrained to be positive in the fit.

Finally, the fit is performed using the least χ^2 method. In the following sections, we will describe how we use the m_{ES} fit to obtain **P**, **C** and **D** in Equation (5.6).

5.3 Peaking Background

We can rewrite Equation (5.3) as

$$P_I = \sum_{J' \notin S} n_{IJ'} , \tag{5.9}$$

where $n_{IJ'}$ is the number of events reconstructed in B mode I from mode J', and determine P_I as a signal yield from generic MC[10].

Each event in generic MC is weighted by a factor to correct for the efficiency difference between data and MC. Each type of generic MC is also scaled appropriately to match the amount of collected data before fitting the m_{ES} distribution of events reconstructed in B mode I to obtain P_I. In the m_{ES} fits, μ, σ and κ are fixed as mentioned in the earlier section.

---

[10] We will now explain the name "peaking background" for P_I. The Argus function is adequate to model the $c\bar{c}$ and *uds* backgrounds that are not expected to peak in the signal region, hence the signal yield will not include such events. However, it is possible for events from a few generic $B\bar{B}$ decays to peak at the B mass. In this case, P_I will contain these peaking generic $B\bar{B}$ events, thus the name "peaking background".

Table 5.1 shows the signal yields from each type of generic MC. As expected, the peaking backgrounds from $c\bar{c}$ and *uds* events are not significant (less than three standard deviations from zero). Hence, to reduce the uncertainties on P_I, we assume no peaking background from $c\bar{c}$ and *uds* events, and use only $B^0\bar{B}^0$ and B^+B^- MC to determine P_I.

Furthermore, the table indicates significant peaking backgrounds for $B^0 \to D^0\bar{D}^0$, $B^0 \to D^{*0}\bar{D}^0$, $B^0 \to D^+D^-$ and $B^- \to D^-D^{*0}$. For $B^0 \to D^+D^-$, this is mainly due to B^0 decaying to D^-K^+X or $D^-\pi^+X$, where D^- decays to $K^+\pi^-\pi^-$ or $K_S\pi^-$, and X is K_S, ρ, a_1 or ω. The light mesons (K^+X) or (π^+X) fake a $D^+ \to K^-\pi^+\pi^-$ or $D^+ \to K_S\pi^+$ decay.

B Mode	$B^0\bar{B}^0$	B^+B^-	$c\bar{c}$	*uds*
$B^0 \to D^0\bar{D}^0$	16.11 ± 3.16	−10.27 ± 3.50	−0.29 ± 4.07	−1.27 ± 1.48
$B^0 \to D^{*0}\bar{D}^0$	8.28 ± 2.02	−4.71 ± 2.51	−4.69 ± 1.71	0.00 ± 0.00
$B^0 \to D^{*0}\bar{D}^{*0}$	1.04 ± 1.04	−1.70 ± 1.71	−0.81 ± 1.59	0.00 ± 0.00
$B^0 \to D^+D^-$	9.79 ± 2.18	0.12 ± 0.82	−0.18 ± 1.84	0.00 ± 0.00
$B^0 \to D^{*+}D^-$	6.84 ± 2.92	0.04 ± 1.47	7.87 ± 4.70	0.69 ± 1.59
$B^0 \to D^{*+}D^{*-}$	0.19 ± 2.16	−0.66 ± 0.70	−0.48 ± 2.60	0.00 ± 0.00
$B^- \to D^-D^0$	−2.12 ± 3.71	3.82 ± 3.82	7.44 ± 6.03	0.16 ± 2.87
$B^- \to D^{*-}D^0$	1.05 ± 3.38	7.95 ± 2.84	−1.37 ± 3.72	0.00 ± 0.00
$B^- \to D^-D^{*0}$	−1.18 ± 2.46	28.44 ± 3.93	−7.57 ± 2.93	−1.62 ± 1.34
$B^- \to D^{*-}D^{*0}$	−2.81 ± 3.48	−2.12 ± 2.52	−1.64 ± 3.41	0.00 ± 0.00

Table 5.1: Peaking background from each type of generic MC.

Sub-Mode	Efficiency ($\times 10^{-4}$)	Peaking Background
$(K\pi\pi)$ $(K_S\pi)$	2.11 ± 0.02	8.46 ± 1.51
$(K\pi\pi)$ $(K\pi\pi)$	9.50 ± 0.08	3.57 ± 1.50

Table 5.2: Branching fraction-weighted efficiencies from signal MC and peaking background from $B^0\overline{B}^0$ MC for the sub-modes of $B^0 \rightarrow D^+D^-$.

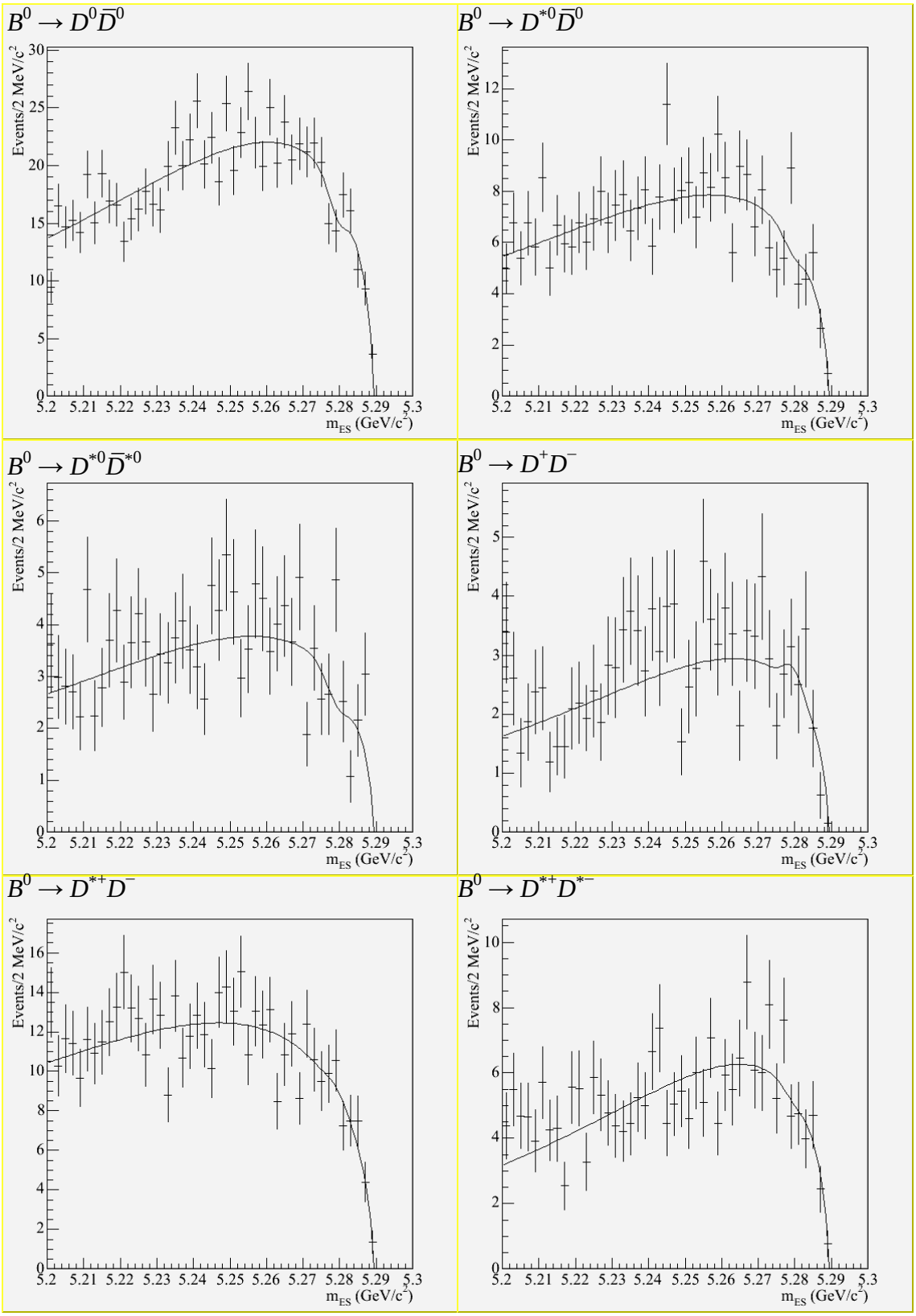
To see if we can reduce the peaking background of $B^0 \rightarrow D^+D^-$, we look at each sub-mode's contribution and its efficiency (weighted by the sub-mode's branching fraction) from signal MC. From Table 5.2, one can see it would be beneficial to remove the sub-mode $(K\pi\pi)$ $(K_S\pi)$. The same study is done for the other *B* modes which show possible peaking background. We remove sub-modes with high peaking background and low efficiencies so that the overall peaking background for the *B* mode is consistent with zero to two standard deviations. We list the sub-modes being removed below.

- $B^0 \rightarrow D^0\overline{D}^0$: $(K\pi)$ $(K_S\pi\pi)$, $(K\pi)$ $(K\pi)$.
- $B^0 \rightarrow D^{*0}\overline{D}^0$: $(K\pi\pi^0)\pi^0$ $(K\pi)$, $(K\pi\pi\pi)\pi^0$ $(K\pi)$.
- $B^0 \rightarrow D^+D^-$: $(K\pi\pi)$ $(K_S\pi)$.
- $B^0 \rightarrow D^{*+}D^-$: $(K\pi)\pi$ $(K_S\pi)$.
- $B^- \rightarrow D^-D^0$: $(K\pi)$ $(K_S\pi)$.
- $B^- \rightarrow D^{*-}D^0$: $(K\pi\pi^0)\pi$ $(K\pi\pi^0)$.
- $B^- \rightarrow D^-D^{*0}$: $(K\pi)\gamma$ $(K\pi\pi)$, $(K\pi\pi\pi)\pi^0$ $(K\pi\pi)$, $(K\pi\pi^0)\pi^0$ $(K_S\pi)$, $(K\pi)\pi^0$ $(K_S\pi)$.

The remaining sub-modes after removing the above dirty modes will be used for our analysis from now on. Using the clean modes, we determine P_I from $B^0\overline{B}^0$ and B^+B^- MC which is shown in Table 5.3. The m_{ES} fits are shown below the table.

B Mode	I	μ (GeV/c^2)	σ (MeV/c^2)	κ	P_I
$B^0 \to D^0\bar{D}^0$	1	5.2795	2.48	45.98	-7.99 ± 4.10
$B^0 \to D^{*0}\bar{D}^0$	2	5.2795	2.68	41.00	-2.09 ± 2.57
$B^0 \to D^{*0}\bar{D}^{*0}$	3	5.2795	3.05	40.26	-1.98 ± 1.92
$B^0 \to D^+D^-$	4	5.2795	2.44	51.08	1.04 ± 1.78
$B^0 \to D^{*+}D^-$	5	5.2795	2.50	31.36	0.94 ± 3.08
$B^0 \to D^{*+}D^{*-}$	6	5.2795	2.61	54.74	-0.87 ± 2.33
$B^- \to D^-D^0$	7	5.2792	2.47	45.15	-1.63 ± 5.19
$B^- \to D^{*-}D^0$	8	5.2792	2.51	29.94	0.96 ± 3.73
$B^- \to D^-D^{*0}$	9	5.2792	2.64	30.58	2.74 ± 2.65
$B^- \to D^{*-}D^{*0}$	10	5.2792	3.09	47.26	-5.20 ± 4.35

Table 5.3: Peaking background (P_I) from $B^0\bar{B}^0$ and B^+B^- MC for each B mode. Also shown are the fixed values of μ, σ and κ used in the $m_{\rm ES}$ fits.


$B^0 \to D^0\bar{D}^0$
$B^0 \to D^{*0}\bar{D}^0$
$B^0 \to D^{*0}\bar{D}^{*0}$
$B^0 \to D^+D^-$
$B^0 \to D^{*+}D^-$
$B^0 \to D^{*+}D^{*-}$
Events/2 MeV/c^2
m$_{ES}$ (GeV/c^2)
5.2 5.21 5.22 5.23 5.24 5.25 5.26 5.27 5.28 5.29 5.3

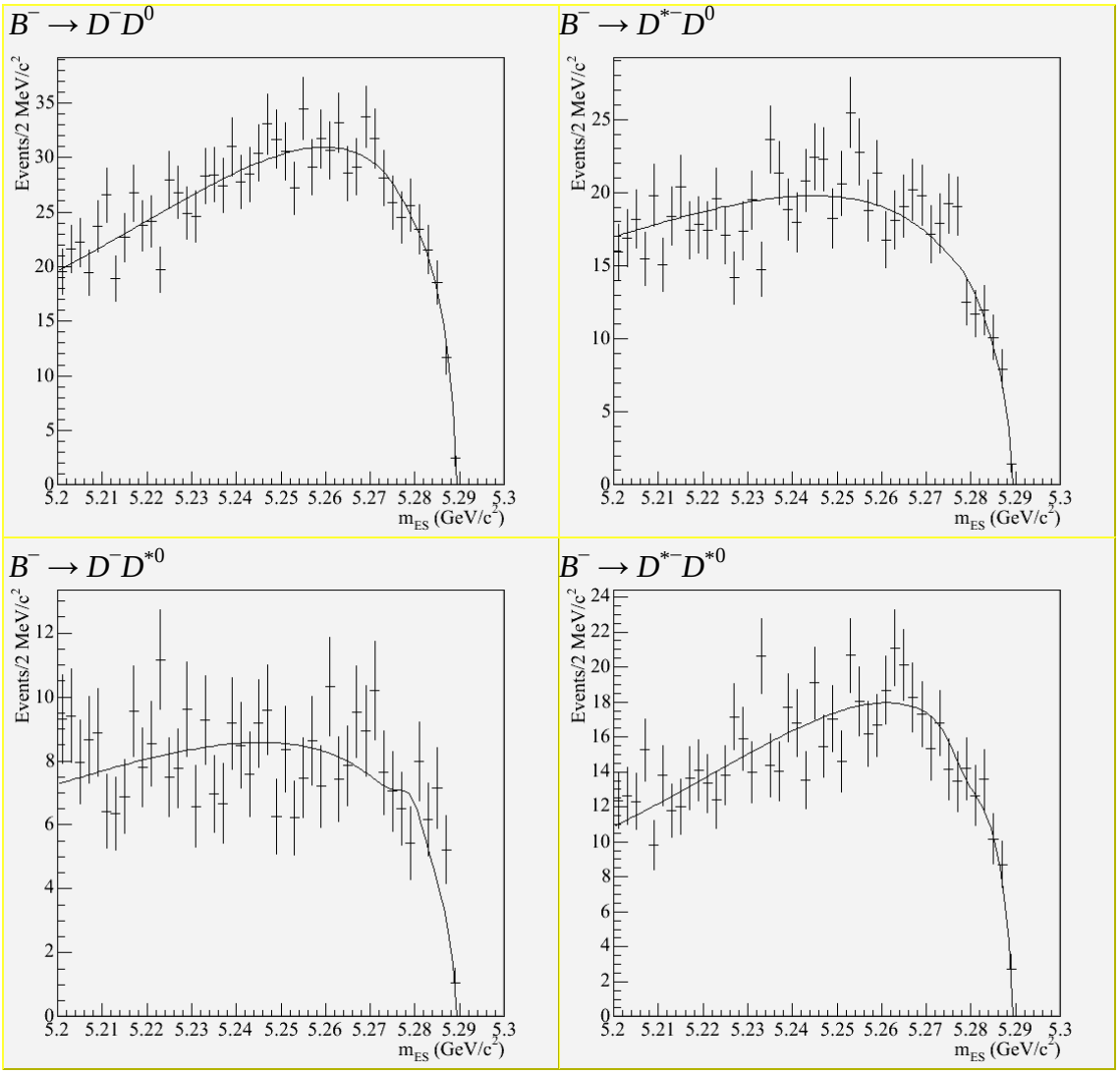


Figure 5.1: The m_{ES} fit to $B^0\overline{B}^0$ and B^+B^- MC for each B mode.

5.4 Cross Feed

Similar to P_I, we can determine C_{IJ} from signal MC by rewriting Equation (5.4),

$$C_{IJ} = \sum_{j\in\{J\text{'s sub-modes}\}} \mathcal{B}_j n_{Ij} / N_j, \qquad (I, J \in S), \tag{5.10}$$

where $\mathcal{B}_j$ is the branching fraction of sub-mode j;

n_{Ij} is the number of events reconstructed in B mode I from sub-mode j; and

N_j is the number of generated events in sub-mode j.

If we treat n_{Ij} as a signal yield from the m_{ES} distribution of sub-mode j, then C_{IJ} is a signal yield from a weighted sum of m_{ES} distributions. Hence, we determine C_{IJ} from signal MC as follows. After applying the efficiency correction factors to the events, the m_{ES} distributions for B mode I from the sub-modes j of J are weighted by b_j / N_j and summed together. If this weighted sum of distributions has more than fifty entries, we take C_{IJ} as the signal yield from an m_{ES} fit to the summed distribution, else C_{IJ} is assumed to be zero since a fit is unreliable when the number of entries is low. The maximum amount of cross feed neglected (by integrating over the signal regions of neglected C_{IJ}'s distributions of the same row) is 0.14% of C_{II}. In the m_{ES} fit for C_{IJ}, the parameters μ, σ and κ of row I are fixed according to Table 5.3.

Due to page constraint, we are unable to show the m_{ES} fits for the full 10×10 cross feed matrix. As an illustration, we show the m_{ES} fits for the modes with dominant non-diagonal elements in Figure 5.2. The cross feed matrix for the ten B modes is shown in Table 5.4.

In reality, the simple model (Gaussian for signal and Argus for background) is not enough to describe the shapes of the non-diagonal distributions[11]. However, since most of these elements are small compared to the diagonal, they would not affect the branching fractions by much. We perform a study based on a detailed simulation of the 2×2

[11] It is difficult to come up with a model that will fit for all the elements correctly.

distributions shown in Figure 5.2 for $B^0 \to D^{*+}D^{*-}$ and $B^- \to D^{*-}D^{*0}$ modes that have the biggest non-diagonal cross feed contributions. The study gives a maximum possible bias of 7% to the branching fractions when such a simple model is used in the m_{ES} fit. Since the bias is small compared to the total uncertainty, a correction for this bias is not applied to our results.

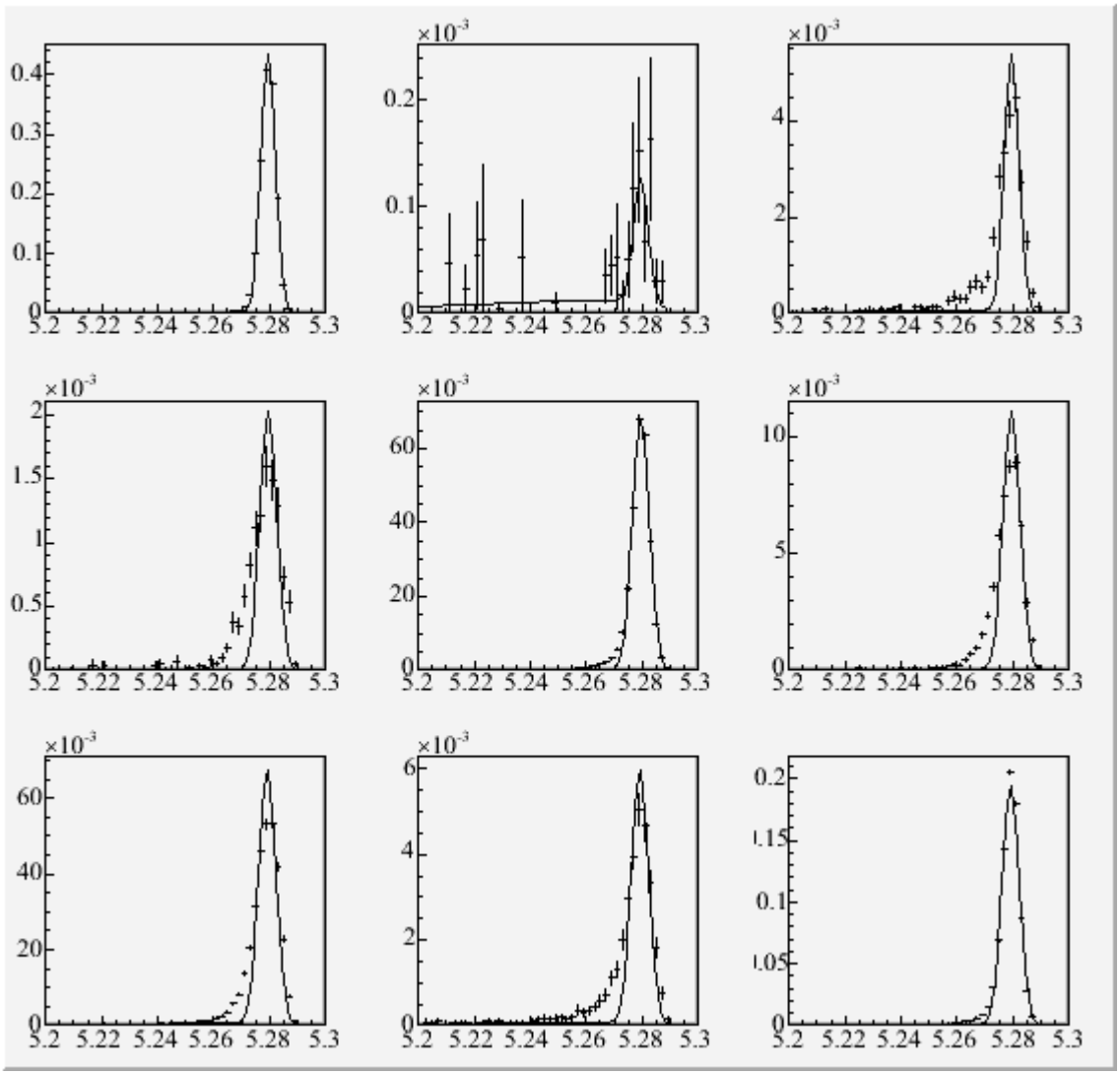


Figure 5.2: The m_{ES} fits for the 3 × 3 cross feed matrix for $B^0 \to D^{*+}D^{*-}$, $B^0 \to D^{*0}\bar{D}^{*0}$ and $B^- \to D^{*-}D^{*0}$ in the same order as Table 5.4.

$\times 10^{-3}$	$D^{*+}D^{*-}$	$D^{*0}\bar{D}^{*0}$	$D^{*-}D^{*0}$	$D^{*+}D^{-}$	$D^{-}D^{*0}$	$D^{*0}\bar{D}^{0}$	$D^{*-}D^{0}$	$D^{0}\bar{D}^{0}$	$D^{-}D^{0}$	$D^{+}D^{-}$
$B^0 \to D^{*+}D^{*-}$	1.424(6)	0.000(0)	0.018(1)	0.001(0)	0	0	0.000(0)	0	0	0
$B^0 \to D^{*0}\bar{D}^{*0}$	0.008(0)	0.260(2)	0.042(1)	0	0.000(0)	0.003(0)	0.001(0)	0	0	0
$B^- \to D^{*-}D^{*0}$	0.260(2)	0.023(1)	0.752(4)	0.000(0)	0	0.001(0)	0.007(1)	0	0	0
$B^0 \to D^{*+}D^{-}$	0.002(0)	0	0.001(0)	1.152(6)	0.008(1)	0	0.004(0)	0	0	0
$B^- \to D^{-}D^{*0}$	0.000(0)	0.001(0)	0.000(0)	0.041(1)	0.370(3)	0.001(0)	0.007(0)	0	0.000(0)	0.000(0)
$B^0 \to D^{*0}\bar{D}^{0}$	0.000(0)	0.002(0)	0.001(0)	0.000(0)	0.001(0)	0.340(2)	0.046(1)	0.000(0)	0	0
$B^- \to D^{*-}D^{0}$	0.004(0)	0.000(0)	0.003(0)	0.006(0)	0.004(0)	0.011(1)	1.351(6)	0	0	0
$B^0 \to D^{0}\bar{D}^{0}$	0.000(0)	0.000(0)	0.000(0)	0.000(0)	0.000(0)	0.001(0)	0.001(0)	1.202(10)	0.002(0)	0
$B^- \to D^{-}D^{0}$	0.000(0)	0.000(0)	0.000(0)	0.002(0)	0.002(0)	0.001(0)	0.001(0)	0.005(1)	1.493(9)	0.006(1)
$B^0 \to D^{+}D^{-}$	0	0	0	0	0.000(0)	0	0	0	0.001(0)	0.951(8)

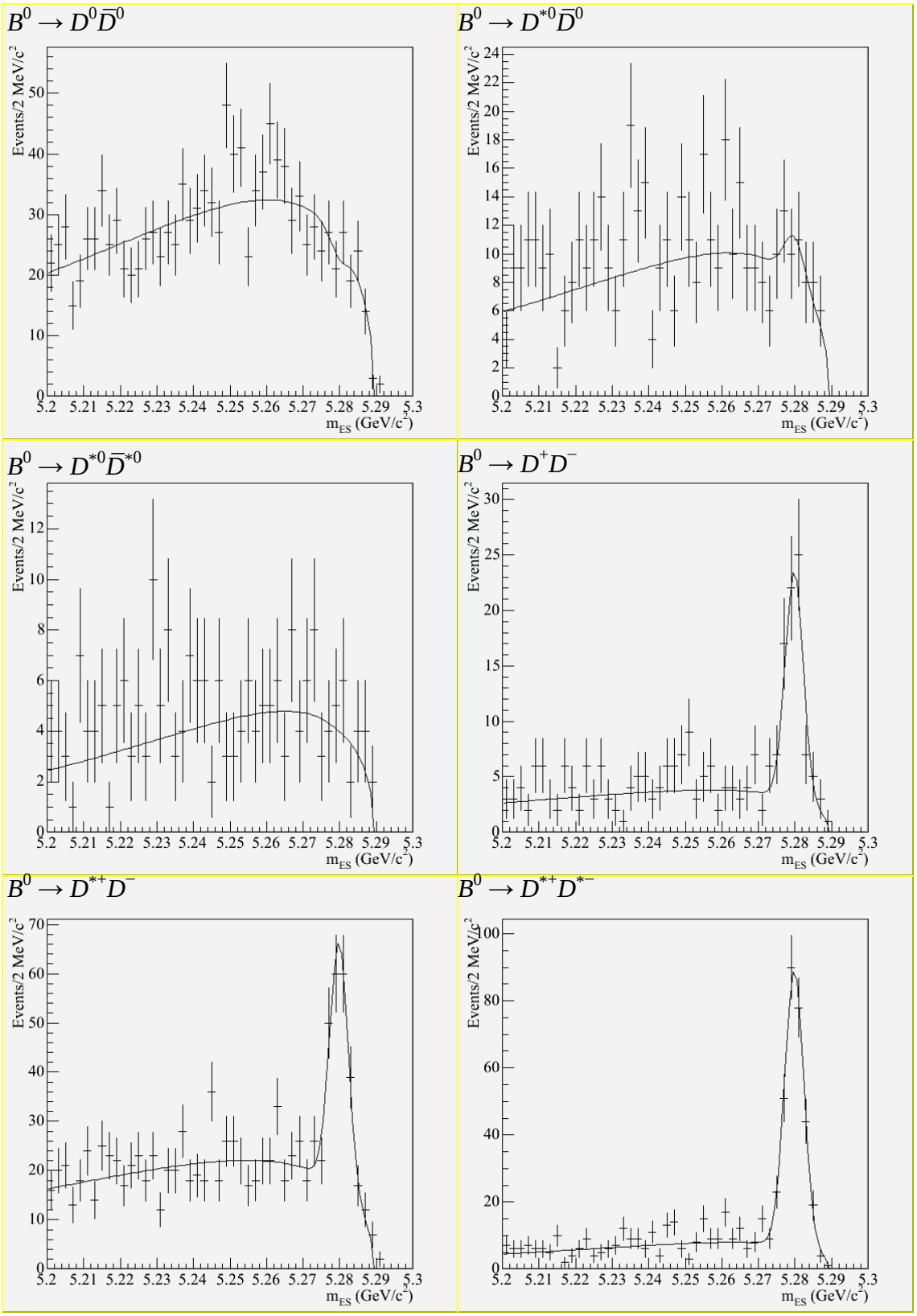
Table 5.4: Values and uncertainties (in parentheses) of the cross feed matrix for the ten *B* modes, grouped to show the sub-matrices (shaded) with significant cross feeds. Elements without uncertainties correspond to unfitted histograms with less than fifty entries. Their signal yields are assumed zero. The m_{ES} fits for the first 3×3 sub-matrix is shown in Figure 5.2.

5.5 Data Yields

D_I is determined as a signal yield from data (in short, known as a data yield). We perform an m_{ES} fit on data for events reconstructed in B mode I to get D_I. In the m_{ES} fit, μ and σ are fixed as shown in Table 5.5, while κ is free to vary. The data yields and m_{ES} fits are shown below.

Decay Mode	I	μ (GeV/c^2)	σ (MeV/c^2)	κ	D_I
$B^0 \to D^0\bar{D}^0$	1	5.2798	2.48	45.39	−10.60 ± 12.30
$B^0 \to D^{*0}\bar{D}^0$	2	5.2798	2.68	48.04	9.92 ± 8.36
$B^0 \to D^{*0}\bar{D}^{*0}$	3	5.2798	3.05	54.19	−0.32 ± 6.20
$B^0 \to D^+D^-$	4	5.2798	2.44	40.77	62.97 ± 9.45
$B^0 \to D^{*+}D^-$	5	5.2798	2.50	37.86	155.63 ± 17.04
$B^0 \to D^{*+}D^{*-}$	6	5.2798	2.61	50.15	269.93 ± 18.59
$B^- \to D^-D^0$	7	5.2796	2.47	41.15	129.44 ± 19.83
$B^- \to D^{*-}D^0$	8	5.2796	2.51	35.41	115.31 ± 16.36
$B^- \to D^-D^{*0}$	9	5.2796	2.64	36.40	62.94 ± 11.25
$B^- \to D^{*-}D^{*0}$	10	5.2796	3.09	48.21	185.28 ± 19.87

Table 5.5: Data yield (D_I) for each B mode. The parameters μ and σ are fixed in the m_{ES} fit while κ is obtained from the fit.


$B^0 \to D^0\bar{D}^0$
$B^0 \to D^{*0}\bar{D}^0$
$B^0 \to D^{*0}\bar{D}^{*0}$
$B^0 \to D^+D^-$
$B^0 \to D^{*+}D^-$
$B^0 \to D^{*+}D^{*-}$
Events/2 MeV/c^2
m$_{ES}$ (GeV/c^2)

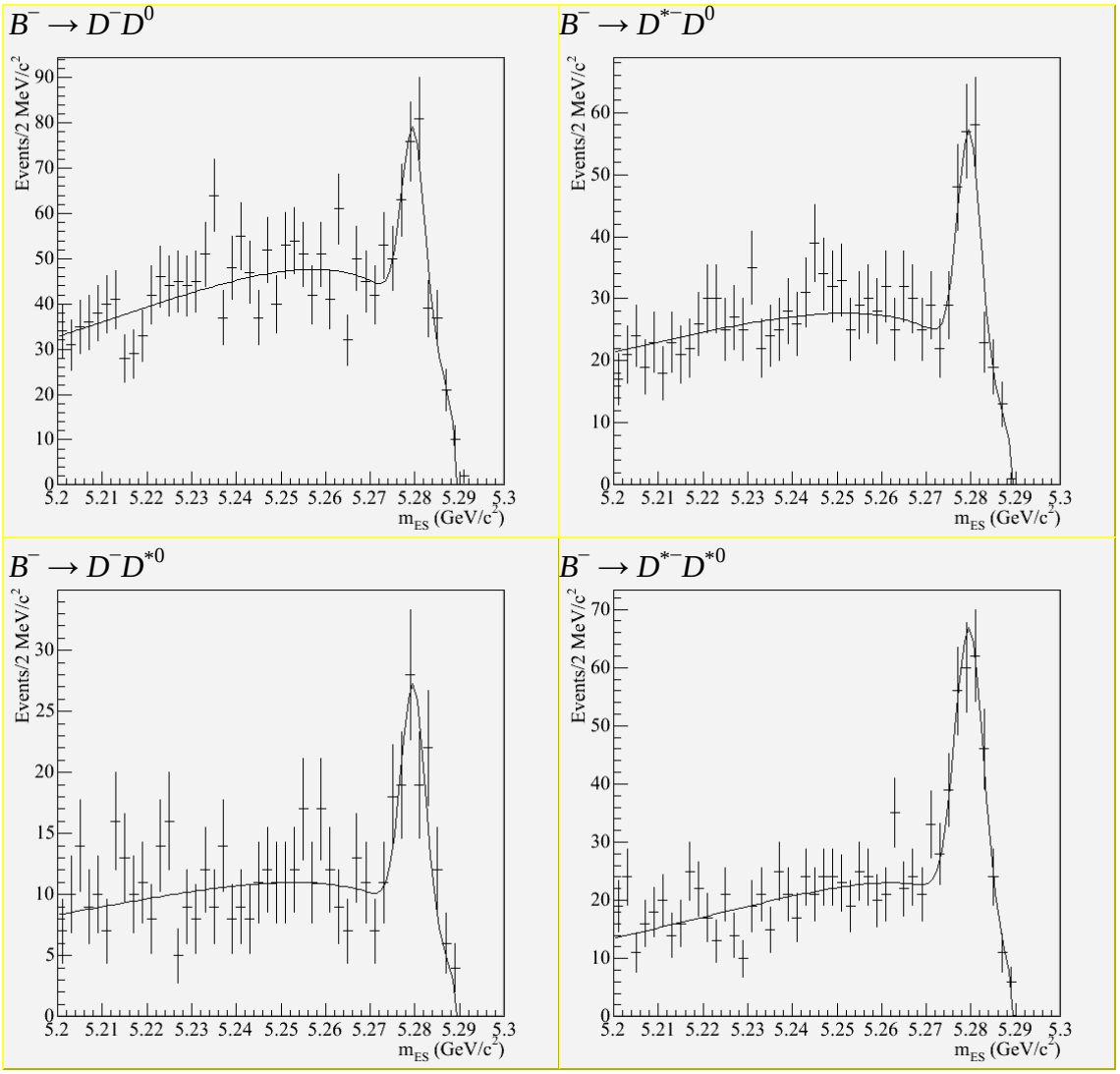


Figure 5.3: The m_{ES} fit to data for each B mode.

5.6 Branching Fraction Values

With all the ingredients **C**, **P** and **D** in place, we can now use Equation (5.6) to find the branching fractions of all the B modes. Since each $B\overline{B}$ event produces two B mesons, therefore for B^0 modes,

N = Number of B^0 mesons in data = 2 × Number of $B^0\overline{B}^0$ events.

Similarly for B^- modes. If we assume that the branching fractions of $Y(4S) \to B^0\overline{B}^0$ and $Y(4S) \to B^+B^-$ are equal, then for all B modes,

N = Number of $B\overline{B}$ events = $N_B = 2.32 \times 10^8$.

To propagate the errors of the cross feed matrix, we use the formula [29]:

$$\mathrm{cov}(\mathcal{B}_I, \mathcal{B}_J) = F_K F_L \,\mathrm{cov}(C^{-1}_{IK}, C^{-1}_{JL}) + C^{-1}_{IK} C^{-1}_{JL} \,\mathrm{cov}(F_K, F_L), \tag{5.11}$$

where $F_I = (D_I - P_I)/N_B$ and

$$\mathrm{cov}(C^{-1}_{IK}, C^{-1}_{JL}) = C^{-1}_{IM} C^{-1}_{NK} C^{-1}_{JP} C^{-1}_{QL} \,\mathrm{cov}(C_{MN}, C_{PQ}). \tag{5.12}$$

Summation over repeated indices is implied.

If we assume that the elements between **P** and **D** are uncorrelated, then Equation (5.11) becomes[12]

$$\mathrm{cov}(\mathcal{B}_I, \mathcal{B}_J) = F_K F_L \,\mathrm{cov}(C^{-1}_{IK}, C^{-1}_{JL}) + C^{-1}_{IK} C^{-1}_{JL} \,\mathrm{cov}(P_K, P_L) + C^{-1}_{IK} C^{-1}_{JL} \,\mathrm{cov}(D_K, D_L). \tag{5.13}$$

In this form, we can treat the uncertainties of branching fractions as having three parts. The first and second terms on the right-hand-side of the equation are, respectively, the systematic uncertainties from cross feed and peaking background, while the last term is the statistical uncertainty from data.

In using the formula, we assume that $\mathrm{cov}(C_{MN}, C_{PQ}) = 0$ for $M \neq P$ or $N \neq Q$, and $\mathrm{cov}(P_K, P_L) = \mathrm{cov}(D_K, D_L) = 0$ for $K \neq L$. The branching fractions and their uncertainties using these assumptions are shown in Table 5.6. The covariance matrix of $\boldsymbol{\mathcal{B}}$ (excluding the cross feed and peaking background systematic uncertainties) is shown in Table 5.7.

[12] The uncertainty of N_B is treated as a systematic uncertainty later.

B Mode	$\mathcal{B}_I$	$\sqrt{C_{IK}^{-1}C_{IL}^{-1}\,\mathrm{cov}(D_K, D_L)}$	$\sqrt{F_K F_L\,\mathrm{cov}(C_{IK}^{-1}, C_{IL}^{-1})}$	$\sqrt{C_{IK}^{-1}C_{IL}^{-1}\,\mathrm{cov}(P_K, P_L)}$
$B^0 \to D^0\bar{D}^0$	−0.10	0.44 (433.5%)	0.002 (1.6%)	0.15 (144.6%)
$B^0 \to D^{*0}\bar{D}^0$	1.01	1.07 (105.3%)	0.011 (1.1%)	0.33 (32.3%)
$B^0 \to D^{*0}\bar{D}^{*0}$	−1.31	1.05 (80.0%)	0.024 (1.9%)	0.32 (24.5%)
$B^0 \to D^+D^-$	2.81	0.43 (15.3%)	0.024 (0.8%)	0.08 (2.9%)
$B^0 \to D^{*+}D^-$	5.72	0.64 (11.2%)	0.032 (0.6%)	0.12 (2.0%)
$B^0 \to D^{*+}D^{*-}$	8.11	0.57 (7.0%)	0.032 (0.4%)	0.07 (0.9%)
$B^- \to D^-D^0$	3.76	0.57 (15.3%)	0.023 (0.6%)	0.15 (4.0%)
$B^- \to D^{*-}D^0$	3.56	0.52 (14.7%)	0.016 (0.4%)	0.12 (3.4%)
$B^- \to D^-D^{*0}$	6.30	1.32 (20.9%)	0.062 (1.0%)	0.31 (4.9%)
$B^- \to D^{*-}D^{*0}$	8.14	1.17 (14.4%)	0.048 (0.6%)	0.25 (3.1%)

Table 5.6: Branching fractions ($\mathcal{B}_I$) of all B modes. Statistical uncertainties from data are shown in the third column while systematic uncertainties due to cross feed and peaking background are shown in the fourth and fifth columns respectively. All values are in 10^{-4}.

$\times 10^{-8}$	$D^0\bar{D}^0$	$D^{*0}\bar{D}^0$	$D^{*0}\bar{D}^{*0}$	D^+D^-	$D^{*+}D^-$	$D^{*+}D^{*-}$	D^-D^0	$D^{*-}D^0$	D^-D^{*0}	$D^{*-}D^{*0}$
$D^0\bar{D}^0$	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$D^{*0}\bar{D}^0$	0.00	1.13	−0.02	0.00	0.00	0.00	0.00	−0.05	0.00	0.00
$D^{*0}\bar{D}^{*0}$	0.00	−0.02	1.10	0.00	0.00	0.01	0.00	0.00	0.00	−0.25
D^+D^-	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.00
$D^{*+}D^-$	0.00	0.00	0.00	0.00	0.41	0.00	0.00	0.00	−0.06	0.00
$D^{*+}D^{*-}$	0.00	0.00	0.01	0.00	0.00	0.32	0.00	0.00	0.00	−0.13
D^-D^0	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00
$D^{*-}D^0$	0.00	−0.05	0.00	0.00	0.00	0.00	0.00	0.27	−0.01	−0.01
D^-D^{*0}	0.00	0.00	0.00	0.00	−0.06	0.00	0.00	−0.01	1.73	0.00
$D^{*-}D^{*0}$	0.00	0.00	−0.25	0.00	0.00	−0.13	0.00	−0.01	0.00	1.37

Table 5.7: Covariance matrix for the branching fractions. Cross feed and peaking background systematic uncertainties are not included.

There are, of course, other sources of errors besides cross feed's and peaking background's that we have not considered. These errors are treated as systematic uncertainties (as opposed to statistical uncertainty which only depends on the amount of collected data) in the next chapter.

6 Systematic Uncertainties

Other sources of uncertainties (besides cross feed and peaking background) that affect the branching fraction measurements are examined in this chapter. In evaluating their systematic uncertainties, one has to be careful in propagating the errors through the cross feed matrix. For convenience, we reproduce the error propagation formula below, where we have added the uncertainty of N_B:

$$\begin{aligned} \operatorname{cov}(\mathcal{B}_I, \mathcal{B}_J) &= F_K F_L \operatorname{cov}(C^{-1}_{IK}, C^{-1}_{JL}) + C^{-1}_{IK} C^{-1}_{JL} \operatorname{cov}(F_K, F_L) + \frac{\mathcal{B}_I \mathcal{B}_J}{N_B^2} \operatorname{var}(N_B) \\ \operatorname{cov}(C^{-1}_{IK}, C^{-1}_{JL}) &= C^{-1}_{IM} C^{-1}_{NK} C^{-1}_{JP} C^{-1}_{QL} \operatorname{cov}(C_{MN}, C_{PQ}) \end{aligned}, \quad (6.1)$$

and $\mathbf{F} = (\mathbf{D} - \mathbf{P}) / N_B$.

We can classify the errors into three types:

1. Error that does not need to be propagated through the cross feed matrix like the error of N_B.
2. Error that needs to be propagated through the first term like the cross feed errors.
3. Error that needs to be propagated through the second term like the data errors.

Therefore, depending on the nature of the error, it will be treated as one of the above types when propagating through Equation (6.1) to determine its systematic effect on the uncertainties of the branching fractions $\mathcal{B}_I$.

6.1 Number of Events in Data

The number of $B\bar{B}$ events ($N_B = 2.32 \times 10^8$) measured in data has an uncertainty of 2.6×10^6 events. This corresponds to a systematic uncertainty of 1.1% on $\mathcal{B}_I$.

6.2 Sub-modes' Branching Fractions

When calculating the cross feed matrix **C**, we use the $D^{(*)}$ branching fractions that are taken from the PDG to weight the MC events. These branching fractions can be split into four independent groups: D^{*+}, D^{*0}, D^0 and D^+. A systematic uncertainty is incurred from the uncertainties of each group.

To propagate the errors, one needs to expand C_{IJ} to the sub-mode level:

$$C_{IJ} = \sum_{j \in \{J\text{'s sub-modes}\}} e_{Ij} \mathcal{B}_j \,, \tag{6.2}$$

where $\mathcal{B}_j$ is the branching fraction of sub-mode j; and

e_{Ij} is the efficiency of reconstructing in mode I an event generated in mode j.

Then, we have

$$\operatorname{cov}(C_{MN}, C_{PQ}) = \sum_{n \in \{N\text{'s sub-modes}\}} \sum_{q \in \{Q\text{'s sub-modes}\}} \left(e_{Mn} e_{Pq} \operatorname{cov}(\mathcal{B}_n, \mathcal{B}_q) + \mathcal{B}_n \mathcal{B}_q \operatorname{cov}(e_{Mn}, e_{Pq}) \right). \tag{6.3}$$

The systematic uncertainty due to the sub-modes' branching fractions is therefore given by the first term in Equation (6.3). The efficiency e_{Ij} is obtained in the same way as **C**: via an m_{ES} fit to signal MC events generated in sub-mode j and reconstructed in B mode I, with the parameters μ, σ and κ fixed as before. The errors after propagating through Equation (6.1) give us the required systematic uncertainties of $\mathcal{B}$.

6.3 Efficiency Correction Factors

Each event in MC is weighted by a factor to correct for the efficiency difference between data and MC. Each factor is a product of efficiency corrections determined from four separate studies (which we will not go into details here):

- Tracking efficiency of charged particles ($\pi^{\pm}$, $K^{\pm}$)
- K_S reconstruction
- Reconstruction of neutral particles (π^0, γ)
- Kaon particle identification (PID)

From the results of these studies, an uncertainty (see Table 6.1) is assigned to each candidate based on its type. These uncertainties enter through the second term in Equation (6.3). The overall uncertainty of the correction factors on e_{Ij} depends on the sub-mode j. For example, the sub-mode $(K\pi\pi^0)\pi\,(K_S\pi\pi)\pi$ has four charged daughters from D^0 (one kaon and three $\pi^{\pm}$), two slow $\pi^{\pm}$, one K_S, one π^0, and one kaon with PID[13]. From Table 6.1, we have

$$\text{var}(e_{Ij}) = (e_{Ij})^2 \times [(4 \times 0.8 + 2 \times 2.2)^2 + (2.5)^2 + (3.0)^2 + (2.5)^2] \times 10^{-4}$$

$$= (7.93 \times 10^{-3})(e_{Ij})^2.$$

Since the first two uncertainties (0.8% and 2.2%) are derived from the same study, we treat them as 100% correlated and add them linearly. The same is true for π^0 and γ that appear together in a sub-mode. The systematic uncertainties of $\mathcal{B}$ due to the efficiency corrections are obtained after propagating the errors through Equation (6.1).

[13] The kaon gets two corrections: one from tracking, the other from PID.

Type	Uncertainty (%)
$\pi^{\pm}$ or $K^{\pm}$ daughters of D	0.8
Slow $\pi^{\pm}$ from $D^{*\pm}$	2.2
K_S	2.5
π^0	3.0
γ	1.8
Kaon identification	2.5

Table 6.1: Uncertainty of the efficiency correction for each type of candidate. Values of the same color are determined from the same study and hence correlated.

6.4 Selection Differences Between Data and MC

Each of the cuts used to select our events has slightly different efficiencies in data and in MC. As a result, **C** and **P** as estimated from MC will not reproduce exactly the values in data. In this section, we will consider the cuts $|\Delta E|$, $-\log(L_{\text{mass}})$, $-l_D$ and F for B selection (Section 4.3), and $\text{P}(\chi^2)$ and Dalitz weight for D selection (Section 3.8). The selection differences for $\pi^{\pm}$, $K^{\pm}$, K_S, π^0 and γ are treated in Section 6.3.

To study the selection differences, $B^0 \to D_s^{*+}D^{*-}$ events from data and generic MC are used. These events have similar decay shapes to $B \to D^{(*)}\bar{D}^{(*)}$ events. They are also reconstructed using the same $D^{(*)}$ decays with similar techniques as $B \to D^{(*)}\bar{D}^{(*)}$. The much larger branching fraction of $B^0 \to D_s^{*+}D^{*-}$ (about twenty times more than that

of $B^0 \to D^{*+}D^{*-}$) allows us to have a reasonably sized data sample from which to measure the systematic differences between data and MC.

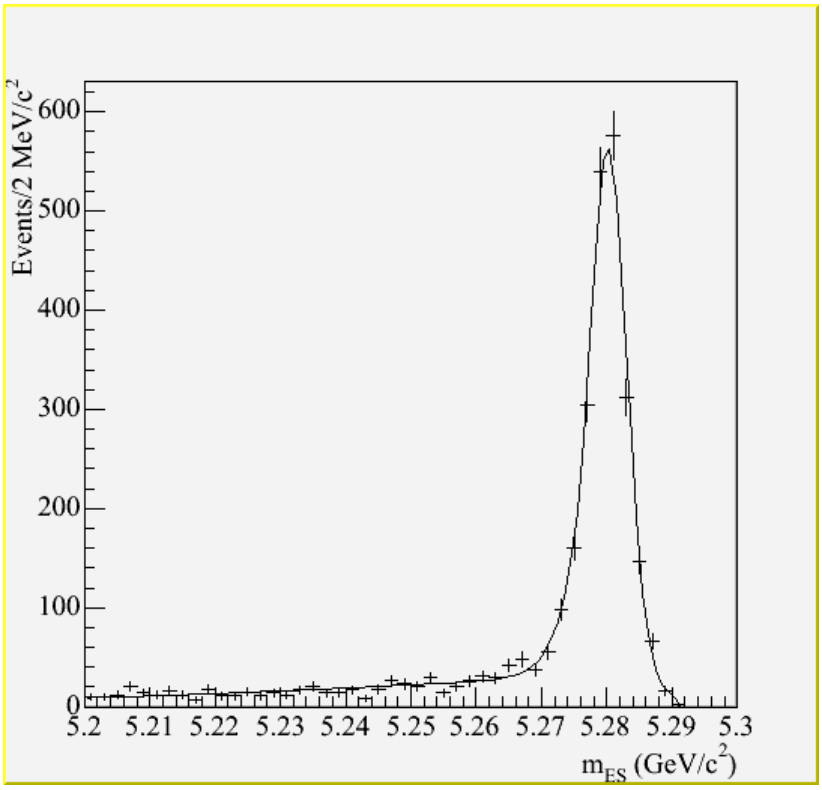


Figure 6.1: A fit to the m_{ES} distribution in data for $B^0 \to D_s^{*+}D^{*-}$ using a Crystal Ball + Argus function.

Figure 6.1 shows the m_{ES} distribution of $B^0 \to D_s^{*+}D^{*-}$ events after making cuts of $|\Delta E| \leq 35$ MeV, $-\log(L_{mass}) \leq 15$, $P(\chi^2) \geq 0.001$ (on both D_s and D candidates) and Dalitz weight ≥ 0.06. The fit is done using a Crystal Ball function for the signal and an Argus function for the background. The Crystal Ball function is defined as

$$CB(x;\mu,\sigma,\alpha) = \begin{cases} \exp(-z^2/2), & z \geq \alpha \\ \exp(\alpha^2/2 - \alpha z), & z < \alpha < 0 \end{cases} \qquad \left(z = \frac{x-\mu}{\sigma}\right). \tag{6.4}$$

Note that α is negative and denotes the point where the exponential tail begins.

We begin by finding the ratio $r \equiv N_{\text{data}}/N_{\text{MC}}$, where N_{data} (N_{MC}) is the number of $B^0 \to D_s^{*+}D^{*-}$ events seen in data (MC). N_{data} (N_{MC}) is determined from a fit to the m_{ES} distribution in data (MC) using a Crystal Ball + Argus function with μ and E_0 (end-point of the Argus function) fixed. To reduce the systematic differences between data and MC, we use the double ratio $r' \equiv r/r_0$, where r_0 is the ratio when the cuts $|\Delta E| \leq 35$ MeV, $-\log(L_{\text{mass}}) \leq 15$, $\text{P}(\chi^2) \geq 0.001$ and Dalitz weight ≥ 0.06 are applied.

For $\text{P}(\chi^2)$ and Dalitz weight, where the cuts are uniform across all B modes, we take the percentage change in r' before and after each cut (in addition to $|\Delta E| \leq 35$ MeV and $-\log(L_{\text{mass}}) \leq 15$) as the systematic uncertainty for that cut. The results are 3.80% for $\text{P}(\chi^2)$ and 1.56% for Dalitz weight. The $\text{P}(\chi^2)$ cut affects the efficiencies equally for all B modes, hence it is similar to N_B and we treat it as a Type 1 error. For the Dalitz weight cut, it is only applied to sub-modes containing the decay $D^0 \to K^-\pi^+\pi^0$, so it is treated as a Type 2 error.

For each of the cuts $|\Delta E|$, $-\log(L_{\text{mass}})$, $-l_D$ and F, where it varies across the sub-modes or B modes, we plot r' as a function of the cut value while fixing the other cuts to the values used to find r_0. The root-mean-square of r' over a range of cut values that gives the same efficiencies as $B \to D^{(*)}\bar{D}^{(*)}$ MC is taken as the systematic uncertainty for that cut. The results are shown in Figure 6.2. These errors are treated as Type 1.

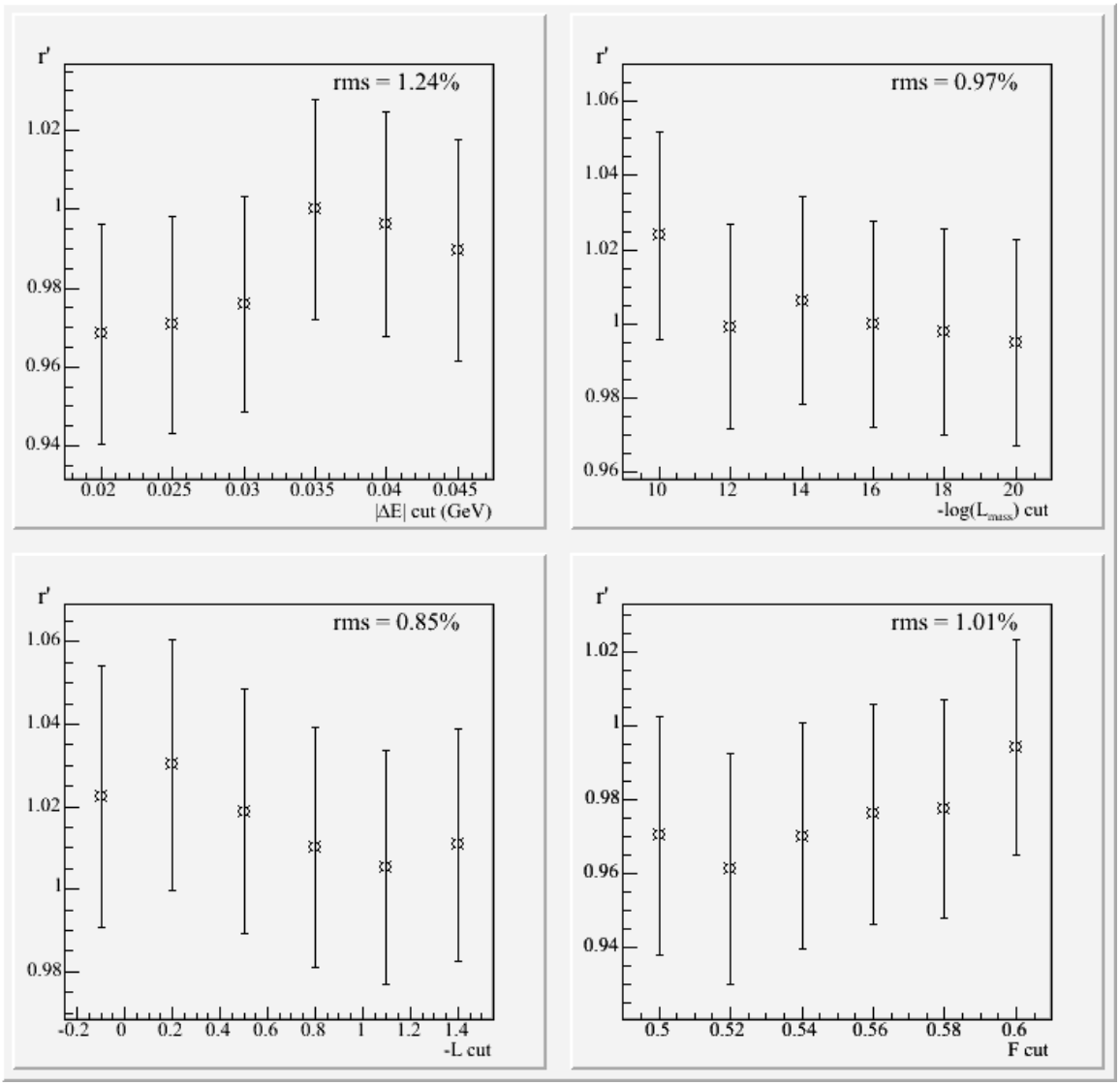


Figure 6.2: Graphs of r' vs. $|\Delta E|$, $-\log(L_{\text{mass}})$, $-l_D$ and F cuts. The root-mean-square (rms) of r' for each graph is also shown.

6.5 Fit Model

The data yield is obtained from an m_{ES} fit where the mean (μ), width (σ) and end-point (E_0) are fixed. These parameters are estimated and have uncertainties in them, giving rise to an uncertainty on the fit model.

For σ, the value is taken from signal MC for each B mode. To estimate its uncertainty due to possible differences between data's and signal MC's m_{ES} resolutions, we first look at this difference ($\Delta\sigma = \sigma_{data} - \sigma_{MC}$) for those modes of high purity, including $B^0 \to D_s^{*+} D^{*-}$. From Table 6.2, we see that these differences are consistent with zero, which justify our use of σ_{MC} in obtaining the data yields. We then find the weighted average of $\Delta\sigma$, which is 0.11 ± 0.08 MeV/c^2. As a conservative estimate, we repeat the data yield determination of each B mode by moving σ up and down by 0.2 MeV/c^2, and take the average change in data yield as the uncertainty of σ on the fit model.

B Mode	σ_{MC} (MeV/c^2)	σ_{data} (MeV/c^2)	$\Delta\sigma$ (MeV/c^2)
$B^0 \to D^{*+} D^-$	2.50 ± 0.01	2.47 ± 0.30	0.03 ± 0.30
$B^0 \to D^{*+} D^{*-}$	2.61 ± 0.01	2.49 ± 0.16	0.12 ± 0.16
$B^- \to D^{*-} D^{*0}$	3.09 ± 0.02	3.22 ± 0.36	0.13 ± 0.36
$B^0 \to D_s^{*+} D^{*-}$	2.64 ± 0.05	2.76 ± 0.10	0.12 ± 0.11

Table 6.2: m_{ES} resolutions of signal MC and data for high purity modes.

For μ and E_0, they are estimated from a fit to common modes in data. Hence, we move the parameters up and down by their uncertainties from the fit (0.2 MeV/c^2 for μ and 0.1 MeV/c^2 for E_0) to obtain their corresponding uncertainties on the fit model.

For each B mode I, the total uncertainty of the fit model is given by a quadratic sum of the three uncertainties from μ, σ and E_0. Since the fit model affects the data yield directly, the sum is propagated as a Type 3 error to give a systematic uncertainty on $\mathcal{B}_I$. The results are shown in Table 6.3.

B Mode	μ	σ	E_0	Sum	Systematic Uncertainty ($\times 10^{-4}$)
$B^0 \to D^0\bar{D}^0$	0.60	0.93	0.62	1.27	0.05
$B^0 \to D^{*0}\bar{D}^0$	0.02	0.52	0.09	0.52	0.07
$B^0 \to D^{*0}\bar{D}^{*0}$	0.00	0.16	0.08	0.18	0.07
$B^0 \to D^+D^-$	1.58	1.08	0.03	1.91	0.09
$B^0 \to D^{*+}D^-$	0.56	5.44	0.20	5.48	0.21
$B^0 \to D^{*+}D^{*-}$	1.24	4.65	0.19	4.82	0.15
$B^- \to D^-D^0$	0.71	4.64	0.04	4.70	0.14
$B^- \to D^{*-}D^0$	2.06	1.81	1.43	3.10	0.10
$B^- \to D^-D^{*0}$	0.10	3.75	0.12	3.75	0.44
$B^- \to D^{*-}D^{*0}$	0.71	6.83	0.07	6.87	0.40

Table 6.3: Uncertainties (changes in data yields) on the fit model due to fixing μ, σ and E_0 in the m_{ES} fits. The square-root of the quadratic sum of these uncertainties is given in the fifth column, which is used to find the systematic uncertainty due to the fit model.

6.6 Decay Models

For $B^0 \to D^{*0}\bar{D}^{*0}$, $B^0 \to D^{*+}D^{*-}$ and $B^- \to D^{*-}D^{*0}$ decays, the excited states of both daughters in each decay result in three possible configurations for the decay products. The distribution of the decay products is therefore a combination of the three configurations, with each having a certain probability that the B meson can decay to [30].

Using $B^0 \to D^{*+}D^{*-} \to (D^0\pi^+)\,(D^-\pi^0)$ as an example, let us define a frame, called the transversity frame, by the coordinates (see Figure 6.3):

- θ_1, the angle between the momentum of π^0 in the rest frame of D^{*-} and the momentum of D^{*-} in the rest frame of B^0;
- θ_{tr}, the angle between the normal to the decay plane of D^{*-} and the momentum of π^+ in the rest frame of D^{*+}; and
- the corresponding azimuthal angle φ_{tr}.

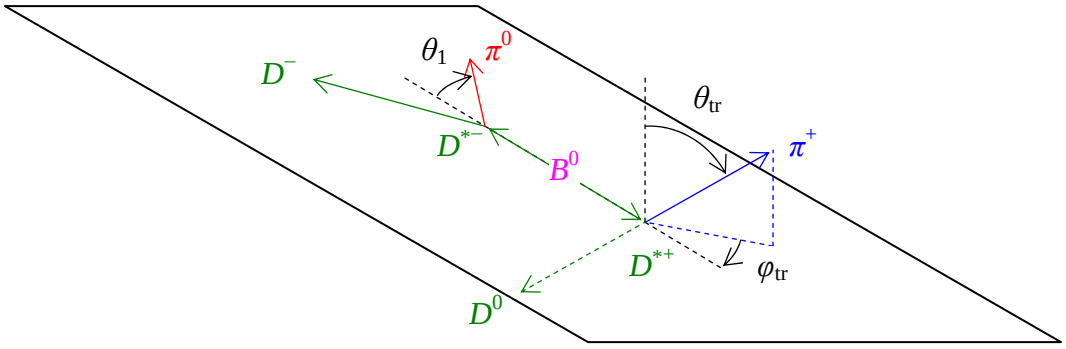


Figure 6.3: Transversity frame for $B^0 \to D^{*+}D^{*-} \to (D^0\pi^+)\,(D^-\pi^0)$. Each straight arrow represents a particle's momentum in its parent's rest frame.

In this frame, we can express the normalized angular distribution of the decay products as

$$\begin{aligned} f(\varphi_{tr}, \cos\theta_1, \cos\theta_{tr}) = \frac{9}{16\pi}\Bigg\{ & \frac{1-a}{2}(1-R)\sin^2\varphi_{tr}\sin^2\theta_1\sin^2\theta_{tr} \\ & + (1+a)(1-R)\cos^2\varphi_{tr}\cos^2\theta_1\sin^2\theta_{tr} \\ & + R\sin^2\theta_1\cos^2\theta_{tr} \\ & + \frac{1}{2}\sqrt{\frac{1-a^2}{2}}(1-R)\cos\delta\sin 2\varphi_{tr}\sin 2\theta_1\sin^2\theta_{tr}\Bigg\} \end{aligned} \quad , \qquad (6.5)$$

where[14] $-1 \le a \le 1$, $0 \le R \le 1$ and $-\pi \le \delta < \pi$.

Our signal MC is generated using a particular set of (a, R, δ), which may not give the true decay model in data. From Equation (6.5), it is clear that different values of a, R and δ will result in different p_t distributions of the slow pions. Since the efficiency of reconstructing a slow pion depends on its p_t (see Figure 6.4), an uncertainty from the lack of knowledge of the true decay model is incurred on **C**.

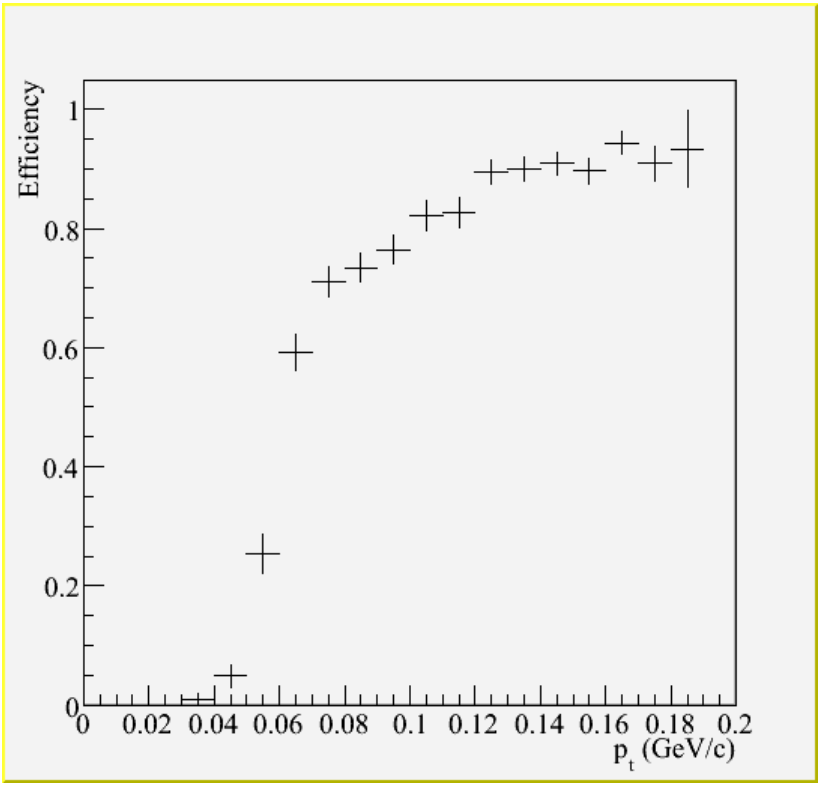


Figure 6.4: Efficiency of the slow $\pi^{\pm}$ as a function of its p_t from MC.

[14] We will not elaborate on the decay parameters a, R and δ, but just remark that R is one of the three decay probabilities, a is the ratio of the difference to the sum of the other two decay probabilities, and δ completes the set of three independent parameters that governs the decay model of a B meson decaying to two D^* mesons. A discussion of the decay parameters is given in [31].

To estimate this uncertainty, we produce a MC sample for each $B \to D^{*}\bar{D}^{*}$ mode using a uniform angular distribution. For each parameter a, R or δ, a random number is selected from a uniform distribution over the parameter's allowed range of values. For R of $B^0 \to D^{*+}D^{*-}$, since it has been measured in [31], a Gaussian distribution with mean and width fixed to the measured value and uncertainty is used instead. Each event in the MC sample is then weighted by the angular distribution given by the randomly selected a, R and δ, and the value of C_{II} (assuming $C_{IJ} = 0$ for $I \neq J$) is obtained as before. The process is repeated a thousand times and the root-mean-squares of C_{II} are determined, which are then propagated as Type 2 errors to give the systematic uncertainties of $\mathcal{B}$ due to the unknown decay models. The results are shown in Table 6.4, which also summarizes all the systematic uncertainties in the previous sections.

6.7 Summary

Systematic Uncertainty	$D^0\bar{D}^0$	$D^{*0}\bar{D}^0$	$D^{*0}\bar{D}^{*0}$	D^+D^-	$D^{*+}D^-$	$D^{*+}D^{*-}$	D^-D^0	$D^{*-}D^0$	D^-D^{*0}	$D^{*-}D^{*0}$
N_B	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
P(χ^2) cut	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8
$-\log(L_{\text{mass}})$ cut	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
$\lvert\Delta E\rvert$ cut	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
$-l_D$ cut	0.8	0.8	0.8	0.8	0.0	0.0	0.8	0.8	0.8	0.0
F cut	0.9	0.9	0.9	0.9	0.0	0.0	0.9	0.9	0.9	0.9
D^{*+} branching fractions	0.0	0.1	0.9	0.0	0.7	1.4	0.0	0.7	0.0	0.7
D^{*0} branching fractions	0.0	1.6	4.9	0.0	0.0	0.0	0.0	0.0	4.4	2.1
D^0 branching fractions	5.7	3.7	7.4	0.0	2.7	5.0	2.7	4.5	3.3	5.2
D^+ branching fractions	0.4	0.2	0.1	13.2	6.5	1.4	6.5	0.3	6.5	0.1
Tracking correction	4.7	3.0	7.9	4.8	6.5	7.9	4.4	6.0	3.8	6.0
Neutrals correction	1.9	2.9	8.4	0.0	1.0	2.5	1.0	1.6	4.3	4.6
K_S correction	0.0	0.1	0.0	0.0	0.2	0.3	0.2	0.2	0.3	0.1
Kaon PID correction	5.4	4.9	7.3	5.0	4.7	4.6	4.7	4.6	4.6	5.0
Dalitz weight cut	1.0	0.2	1.4	0.0	0.5	1.0	0.5	0.8	0.7	1.0
Decay model	0.1	0.0	6.1	0.0	0.0	1.0	0.0	0.0	0.0	4.1
Cross feed	44.6	6.7	5.4	3.1	3.6	1.8	3.6	2.8	7.0	4.9
Peaking background	144.6	32.3	24.5	2.9	2.0	0.9	4.0	3.4	4.9	3.1
Fit model	1.6	1.1	1.9	0.8	0.6	0.4	0.6	0.4	1.0	0.6
Total	151.7	34.2	31.0	16.1	12.3	12.0	11.9	11.0	14.8	13.6

Table 6.4: Systematic uncertainties (%) of the branching fractions of all B modes estimated from Equation (6.1). Type 1 errors are shaded in red, Type 2 in green and Type 3 in blue. The total is a square-root of the quadratic sum of the uncertainties in each column.

7 Upper Limits

For $B^0 \rightarrow D^{(*)0}\bar{D}^{(*)0}$ decays whose data yields are consistent with zero, upper limits are determined for their branching fractions at the 90% confidence level. Such an upper limit can be estimated from the likelihood ratio method.

7.1 Likelihood Ratio Method

For an event α in data or generic MC reconstructed in B mode I, (I = 1, …, 10), the likelihood is given by

$$f_I^t(x_\alpha) = N_{\mathrm{sig},I}^t G(x_\alpha; \mu_I^t, \sigma_I^t) + N_{\mathrm{bkg},I}^t A(x_\alpha; \kappa_I^t, E_0^t) \quad (x = m_{\mathrm{ES}}), \tag{7.1}$$

where G is a Gaussian function and A is an Argus function. The event type (data or MC) is denoted by t. To account for cross feed and peaking background, we set

$$N_{\mathrm{sig},I}^{\mathrm{data}} = N_B \sum_{J=1}^{10} C_{IJ}\mathcal{B}_J + P_I, \qquad P_I = N_{\mathrm{sig},I}^{\mathrm{MC}}. \tag{7.2}$$

The likelihood for the sample of events of type t and reconstructed in mode I is

$$L_I^t = \frac{\lambda^n e^{-\lambda}}{n!} \prod_{\alpha=1}^{n} f_I^t(x_\alpha),\ \lambda = N_{\mathrm{sig}} + N_{\mathrm{bkg}}, \tag{7.3}$$

where n is the sample size which is treated to have a Poisson distribution; and

α denotes an event belonging to the sample.

For clarity, we have dropped the superscript "t" and subscript "I" on λ, n, N_{sig} and N_{bkg} in the above equation. The full likelihood for all our events in data and generic MC is

$$L(\boldsymbol{\mathcal{B}}, \mathbf{P}, \mathbf{C}, N_B, \mathbf{N}_{\mathrm{bkg}}^{\mathrm{data,\,MC}}, \boldsymbol{\mu}^{\mathrm{data,\,MC}}, \boldsymbol{\sigma}^{\mathrm{data,\,MC}}, \boldsymbol{\kappa}^{\mathrm{data,\,MC}}, E_0^{\mathrm{data,\,MC}}) = \prod_{I=1}^{10} \prod_{t=\mathrm{data,\,MC}} L_I^t. \tag{7.4}$$

The maximum likelihood estimates of $\boldsymbol{\mathcal{B}}$ are obtained by maximizing L with respect to $\boldsymbol{\kappa}^{\text{data}}$, $\mathbf{P}$ and $\boldsymbol{\mathcal{B}}$ while keeping the other parameters (with $N_B = 2.32 \times 10^8$ and $\boldsymbol{\sigma}^{\text{data}} = \boldsymbol{\sigma}^{\text{MC}}$) fixed to the values described in Section 5.2. In the maximum likelihood fit, the cross feed $\mathbf{C}$ is also fixed and taken from Table 5.4. The likelihood L_{max} from this fit is then used to find the likelihood ratio (for a given B mode I)

$$\text{LR}_I = 2\log\left(\frac{L_{\text{max}}}{L(\mathcal{B}_I)}\right), \tag{7.5}$$

where $L(\mathcal{B}_I)$ is the likelihood from a fit with $\mathcal{B}_I$ fixed.

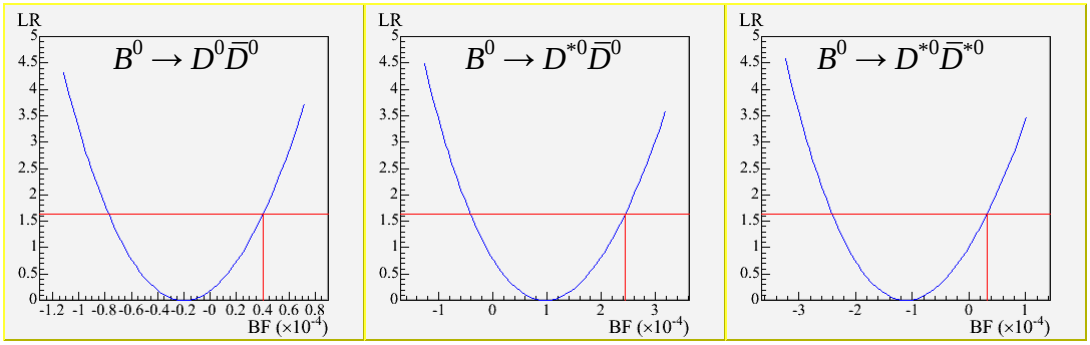


Figure 7.1: Graphs of LR_I vs. $\mathcal{B}_I$ for the modes $B^0 \to D^0\bar{D}^0$, $B^0 \to D^{*0}\bar{D}^0$ and $B^0 \to D^{*0}\bar{D}^{*0}$. The vertical line shows the value of $\mathcal{B}_I$ when LR = 1.6424 (horizontal line).

Figure 7.1 shows LR_I as a function of $\mathcal{B}_I$ for each of the $B^0 \to D^{(*)0}\bar{D}^{(*)0}$ decays. The horizontal red line denotes the value $\text{LR}_{80} = 1.6424$ such that $\text{P}(\chi_1^2 < \text{LR}_{80}) = 0.8$, where χ_1^2 is a χ^2 distribution with one degree of freedom. It can be shown that (under certain conditions) LR_I has approximately a χ_1^2 distribution [32]. (We verify this to be true

for our likelihood ratio via a simulation involving $B^0 \to D^{*0}\bar{D}^{*0}$ and $B^- \to D^{*-}D^{*0}$.) Since $\chi^2_1 < \mathrm{LR}_{80}$ corresponds to a two-sided 80% confidence interval, a 90% upper limit for the branching fraction is then given by the value of $\mathcal{B}_I$ when $\mathrm{LR}_I = \mathrm{LR}_{80}$.

The above method, however, does not guarantee that the upper limits will always be greater than zero. To ensure that the limits will always correspond to physical values, we make use of the Feldman-Cousins method [33].

7.2 Feldman-Cousins Method

Let $\mathcal{B}_{\mathrm{MLE}}$ be the maximum likelihood estimate of $\mathcal{B}$ for a given B mode. (We will drop the subscript "I" of denoting a B mode in this section.) From Section 7.1, we find that the distribution of $\mathcal{B}_{\mathrm{MLE}}$ is approximately Gaussian with width $s_{\mathcal{B}}$ given by

$$s_{\mathcal{B}} = \frac{\mathcal{B}_{90} - \mathcal{B}_{\mathrm{MLE}}}{\sqrt{\mathrm{LR}_{80}}}, \tag{7.6}$$

where $\mathcal{B}_{90}$ is the 90% upper limit from the likelihood ratio method.

Using the Gaussian approximation, we determine the measured mean

$$x_0 \equiv \frac{\mathcal{B}_{\mathrm{MLE}}}{s_{\mathcal{B}}}. \tag{7.7}$$

Then, applying the Feldman-Cousins method, a 90% upper limit[15] for $\mathcal{B}$ is obtained by multiplying the 90% upper limit of x_0 (from Figure 7.2) by the estimated width $s_{\mathcal{B}}$. The results are shown in Table 7.1.

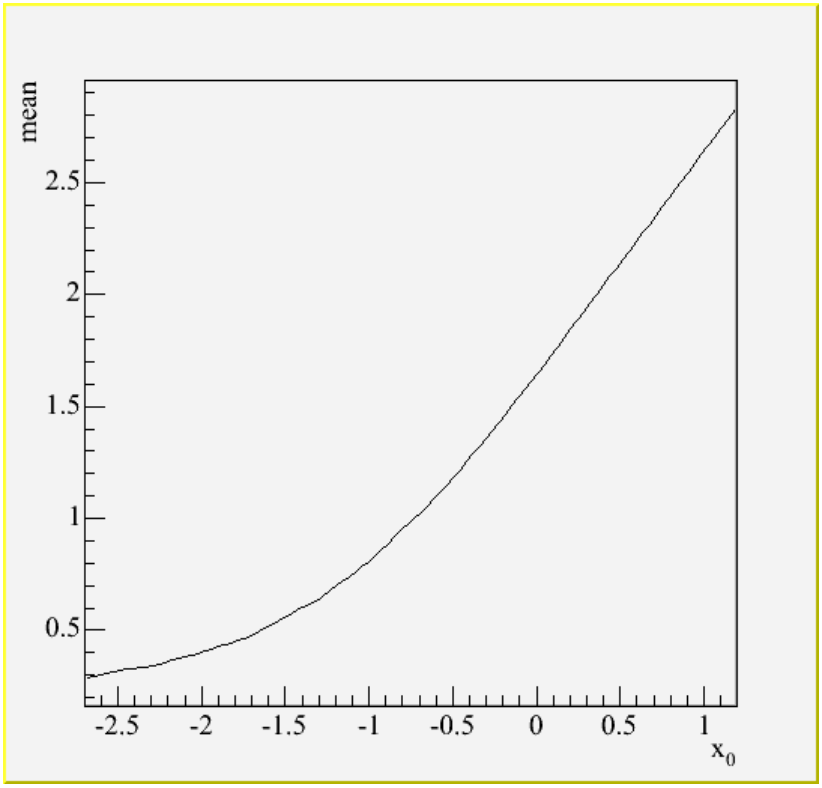


Figure 7.2: Feldman-Cousins 90% confidence intervals (constrained to be non-negative) for the mean of a Gaussian distribution with unit width. The curve shows the upper limits of the measured mean x_0 taken from Table X of [33]. The lower limits are all zero for the range of x_0 shown.

[15] The Feldman-Cousins method actually gives a two-sided confidence interval. For the $B^0 \to D^{(*)0}\overline{D}^{(*)0}$ decays, the lower limits are all zero at the 90% confidence level. Put in another way, we have not observed any signal for these decays at the 90% level.

B Mode	LR ($\times 10^{-4}$)	FC ($\times 10^{-4}$)
$B^0 \to D^0 \bar{D}^0$	0.40	0.59
$B^0 \to D^{*0} \bar{D}^0$	2.44	2.85
$B^0 \to D^{*0} \bar{D}^{*0}$	0.32	0.91

Table 7.1: 90% upper limits for the branching fractions of $B^0 \to D^{(*)0} \bar{D}^{(*)0}$ decays from the likelihood ratio (LR) and Feldman-Cousins (FC) methods.

7.3 Systematic uncertainties

Now, we add the systematic uncertainties to the upper limits. From Table 6.4, the dominant sources are the peaking background and the fit model. The systematic uncertainty from the peaking background is accounted for in constructing the likelihood L (see Equation (7.2)). For the fit model, its systematic uncertainty can be determined by adding to $-2\log L$ a χ^2 constraint

$$\sum_{I=1}^{10} \left(\frac{\sigma_I^{\text{data}} - \sigma_I^{\text{MC}}}{s_\sigma} \right)^2 , \tag{7.8}$$

and repeating the maximum likelihood fit with $\boldsymbol{\mu}^{\text{data}}$, $\boldsymbol{\sigma}^{\text{data}}$ and E_0^{data} free. s_σ is the weighted average of $\Delta\sigma$ obtained in Section 6.5 and is equal to 0.2 MeV/c^2.

We treat the rest of the systematic uncertainties as one and account for this single uncertainty by multiplying to each $\mathcal{B}_I$ in the likelihood L a scale factor A_I, which is free in the maximum likelihood fit, and adding to $-2\log L$ a χ^2 constraint

$$\sum_{I,J=1}^{10} (A_I - 1)(A_J - 1) \operatorname{cov}^{-1}(A_I, A_J) . \tag{7.9}$$

To simplify the fit, we assume that the scale factors are uncorrelated, so $\mathrm{cov}(A_I, A_J) = 0$ for $I \neq J$. For $I = J$, $\mathrm{cov}(A_I, A_J)$ is the quadratic sum of the systematic uncertainties of $\mathcal{B}_I$ in Table 6.4, with the contributions from peaking background and fit model excluded. Another fit is also done assuming the scale factors are 100% correlated. In this case, there is only one common scale factor A and we take $\mathrm{var}(A)$ to be the average of $\mathrm{cov}(A_I, A_I)$. The χ^2 constraint in (7.9) reduces to $A^2/\mathrm{var}(A)$.

A summary of the results from the Feldman-Cousins method is given below.

Systematic Uncertainty	$B^0 \to D^0\bar{D}^0$	$B^0 \to D^{*0}\bar{D}^0$	$B^0 \to D^{*0}\bar{D}^{*0}$
Fit model	0.59	2.86	0.91
Others (uncorrelated)	0.59	2.90	0.91
Others (100% correlated)	0.59	2.91	0.91
All (uncorrelated)	0.59	2.91	0.92
All (100% correlated)	0.59	2.92	0.92

Table 7.2: 90% upper limits ($\times 10^{-4}$) with systematic uncertainties included for the branching fractions of $B^0 \to D^{(*)0}\bar{D}^{(*)0}$ decays using the Feldman-Cousins method. By the fit's construction, the systematic uncertainty from the peaking background is naturally included in each value.

8 Conclusion

8.1 Summary of Results

We summarize our measurements of the $B \to D^{(*)}\bar{D}^{(*)}$ branching fractions in Table 8.1 and 90% upper limits for the $B^0 \to D^{(*)0}\bar{D}^{(*)0}$ branching fractions in Table 8.2. The covariance matrix for the branching fractions, including all systematic uncertainties, is given in Table 8.3.

Decay	BF ($\times 10^{-4}$)
$B^0 \to D^0\bar{D}^0$	$-0.10 \pm 0.44 \pm 0.15$
$B^0 \to D^{*0}\bar{D}^0$	$1.01 \pm 1.07 \pm 0.35$
$B^0 \to D^{*0}\bar{D}^{*0}$	$-1.31 \pm 1.05 \pm 0.41$
$B^0 \to D^+D^-$	$2.81 \pm 0.43 \pm 0.45$
$B^0 \to D^{*+}D^-$	$5.72 \pm 0.64 \pm 0.71$
$B^0 \to D^{*+}D^{*-}$	$8.11 \pm 0.57 \pm 0.97$
$B^- \to D^-D^0$	$3.76 \pm 0.57 \pm 0.45$
$B^- \to D^{*-}D^0$	$3.56 \pm 0.52 \pm 0.39$
$B^- \to D^-D^{*0}$	$6.30 \pm 1.32 \pm 0.93$
$B^- \to D^{*-}D^{*0}$	$8.14 \pm 1.17 \pm 1.11$

Table 8.1: Branching fractions (BF) of $B \to D^{(*)}\bar{D}^{(*)}$ decays. The first uncertainty is statistical while the second is systematic.

Decay	No Systematic ($\times 10^{-4}$)	With Systematic ($\times 10^{-4}$)
$B^0 \to D^0 \bar{D}^0$	0.59	0.59
$B^0 \to D^{*0} \bar{D}^0$	2.85	2.92
$B^0 \to D^{*0} \bar{D}^{*0}$	0.91	0.92

Table 8.2: 90% upper limits, with and without systematic uncertainties, for $B^0 \to D^{(*)0} \bar{D}^{(*)0}$ branching fractions from the Feldman-Cousins method.

$\times 10^{-8}$	$D^0\bar{D}^0$	$D^{*0}\bar{D}^0$	$D^{*0}\bar{D}^{*0}$	D^+D^-	$D^{*+}D^-$	$D^{*+}D^{*-}$	D^-D^0	$D^{*-}D^0$	D^-D^{*0}	$D^{*-}D^{*0}$
$D^0\bar{D}^0$	0.22	0.00	0.00	0.00	−0.01	−0.01	0.00	0.00	−0.01	−0.01
$D^{*0}\bar{D}^0$	0.00	1.25	−0.04	0.02	0.04	0.07	0.02	−0.02	0.05	0.07
$D^{*0}\bar{D}^{*0}$	0.00	−0.04	1.27	−0.03	−0.08	−0.15	−0.05	−0.06	−0.13	−0.53
D^+D^-	0.00	0.02	−0.03	0.39	0.26	0.22	0.16	0.08	0.26	0.16
$D^{*+}D^-$	−0.01	0.04	−0.08	0.26	0.91	0.55	0.26	0.19	0.37	0.46
$D^{*+}D^{*-}$	−0.01	0.07	−0.15	0.22	0.55	1.26	0.30	0.33	0.54	0.73
D^-D^0	0.00	0.02	−0.05	0.16	0.26	0.30	0.53	0.11	0.27	0.27
$D^{*-}D^0$	0.00	−0.02	−0.06	0.08	0.19	0.33	0.11	0.43	0.19	0.31
D^-D^{*0}	−0.01	0.05	−0.13	0.26	0.37	0.54	0.27	0.19	2.61	0.55
$D^{*-}D^{*0}$	−0.01	0.07	−0.53	0.16	0.46	0.73	0.27	0.31	0.55	2.60

Table 8.3: Covariance matrix for $B \to D^{(*)} \bar{D}^{(*)}$ branching fractions with all systematic uncertainties included.

8.2 Test of Factorization Assumption

Comparing Table 8.1 to Table 1.3, it is interesting to note that the theoretical predictions of the $B \rightarrow D^{(*)}\bar{D}^{(*)}$ branching fractions are generally higher than the experimental values. The differences in branching fractions between experiment and theory ($\Delta\mathcal{B}$) for the constituent quark model (CQM) [], the light-front quantum chromodynamics (LFQCD) [] and heavy quark symmetry with corrections (HQSC) [], are shown in Table 8.4. All the models rely on the factorization assumption and differ only in their calculations of the decay constants and form factors.

To see how each model fares in their predictions of the $B \rightarrow D^{(*)}\bar{D}^{(*)}$ branching fractions, we form the χ^2 test statistic

$$\chi^2 = \sum_{I,J=1}^{10} \Delta\mathcal{B}_I \Delta\mathcal{B}_J \, \mathrm{cov}^{-1}(\Delta\mathcal{B}_I, \Delta\mathcal{B}_J). \tag{8.1}$$

Since the theoretical uncertainties are not provided in the papers, they are excluded from the covariance between $\Delta\mathcal{B}_I$ and $\Delta\mathcal{B}_J$. Given that the estimates of the branching fractions are approximately Gaussian, this test statistic should have a χ^2 distribution with ten degrees of freedom (χ^2_{10}). The probability that χ^2_{10} is greater than χ^2 for each of the models, CQM, LFQCD and HQSC, is given in Table 8.5. Our results seem to favor the values predicted by HQSC, though the other two models cannot be ruled out since theoretical errors are ignored in the test. We conclude that our results are consistent with the factorization assumption. Furthermore, since the branching fractions of $B^0 \rightarrow D^{(*)0}\bar{D}^{(*)0}$ are consistent with zero, there is no evidence of final-state interactions in the double charm decays.

Decay	CQM	LFQCD	HQSC	Uncertainty ($\times 10^{-4}$)
$B^0 \to D^0 \bar{D}^0$	0.10	0.10	0.10	0.47
$B^0 \to D^{*0} \bar{D}^0$	1.01	1.01	1.01	1.12
$B^0 \to D^{*0} \bar{D}^{*0}$	1.31	1.31	1.31	1.13
$B^0 \to D^+ D^-$	0.99	1.19	0.29	0.62
$B^0 \to D^{*+} D^-$	1.58	2.08	1.38	0.95
$B^0 \to D^{*+} D^{*-}$	2.49	1.79	0.99	1.12
$B^- \to D^- D^0$	0.44	0.64	0.46	0.73
$B^- \to D^{*-} D^0$	1.84	1.34	0.94	0.65
$B^- \to D^- D^{*0}$	2.40	2.70	3.20	1.61
$B^- \to D^{*-} D^{*0}$	3.46	2.66	1.66	1.61

Table 8.4: Differences ($\times 10^{-4}$) in the branching fractions of $B \to D^{(*)} \bar{D}^{(*)}$ decays between experiment (our results) and theory for the models CQM, LFQCD and HQSC. The uncertainty of each difference is experimental only and hence the same for all the models.

Model	χ^2	$P(\chi^2_{10} > \chi^2)$
CQM	22.3	1.3%
LFQCD	19.6	3.4%
HQSC	14.6	14.7%

Table 8.5: The χ^2 test statistics and their probabilities for the branching fraction predictions from the models CQM, LFQCD and HQSC.

It is unfortunate that the theoretical uncertainties are not given by the authors[16] for us to do a proper χ^2 test and make a definitive statement on the factorization assumption. Hopefully, with the covariance matrix we have provided, one can perform a χ^2 test when the theoretical uncertainties become available.

Perhaps, a stronger test can be carried out by adding $B \to D_s^{(*)}\bar{D}^{(*)}$ decays to the measurement. In this case, one is looking at $D_s^+ \to \varphi\pi^+$, whose branching fraction has an uncertainty of 13% [34]. This is a major source of systematic uncertainty which needs to be reduced before measuring the branching fractions of $B \to D_s^{(*)}\bar{D}^{(*)}$ decays.

A better approach, as noted in the Introduction, is to measure ratios of branching fractions. In such a ratio, most of the experimental systematic uncertainties (such as those from the $D_{(s)}^{(*)}$ branching fractions and data-MC differences) will cancel out. The same is also true for the theoretical uncertainties. With twice the amount of data in the next year or so, the test of factorization using the double charm decays is certainly a possibility.

8.3 Final Words

This dissertation presents a simultaneous measurement of the branching fractions of ten $B \to D^{(*)}\bar{D}^{(*)}$ decays. In performing this analysis, we have encountered difficulties like cross feed and peaking background, and we show how we deal with these problems systematically via a matrix equation. Furthermore, the propagation of errors through the matrix equation, especially the systematic errors, is handled by applying a simple formula.

---

[16] Since in theory, the branching fractions are derived from common quantities, they will be correlated and a covariance matrix should preferably be given.

The determination of the upper limits is a challenge too, but we overcome it by constructing a likelihood ratio where the likelihood is computed from a simultaneous fit to both data and generic MC to account for the peaking background. Incorporating the systematic uncertainties to the upper limits is achieved by adding χ^2 constraints to the negative log-likelihood. These techniques are very general and in principle, can be applied to any number of decays. That is, adding $B \to D_s^{(*)}\bar{D}^{(*)}$ decays to the analysis will not be hard to do[17]. Finally, we conclude by stating that our measured branching fractions are consistent with the predictions from the Standard Model using the factorization approach, and a test of the factorization assumption in double charm decays through a simultaneous measurement of ratios of branching fractions is certainly feasible.

[17] It is only limited by the processing time of a computer.

References

[1] C. D. Anderson, The Positive Electron, Phys. Rev. 43, 491 (1933).

[2] O. Chamberlain, E. Segrè, C. Wiegand and T. Ypsilantis, Observation of Anti-Protons, Phys. Rev. 100, 947 (1955).

[3] B. Cork, G. R. Lambertson, O. Piccioni and W. A. Wenzel, Anti-Neutrons Produced from Anti-Protons in Charge-Exchange Collisions, Phys. Rev. 104, 1193 (1956).

[4] AMS Collaboration, Search for Anti-Helium in Cosmic Rays, Phys. Lett. B 461, 387 (1999)

[5] A. D. Sakharov, Violation of *CP* Invariance, *C* Asymmetry, and Baryon Asymmetry of the Universe, JETP Lett. 5, 24 (1967).

[6] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, Experimental Test of Parity Conservation in Beta Decay, Phys. Rev. 105, 1413 (1957).

[7] R. L. Garwin, L. M. Lederman and M. Weinrich, Observations of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: the Magnetic Moment of the Free Muon, Phys. Rev. 105, 1415 (1957).

[8] J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay, Evidence for the 2π Decay of the K_L Meson, Phys. Rev. Lett. 13, 138 (1964).

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

[10] C. Jarlskog, Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal *CP* Non-Conservation, Phys. Rev. Lett. 55, 1039 (1985).

[11] PDG, Review of Particle Physics: *CP* violation, Phys. Lett. B 592, 1 (2004).

[12] D. Melikhov and B. Stech, Weak Form Factors for Heavy Meson Decays: An Update, Phys. Rev. D 62, 014006 (2000).

[13] H. Y. Cheng, C. K. Chua, C. W. Hwang, Covariant Light-Front Approach for *s*-Wave and *p*-Wave Mesons: Its Application to Decay Constants and Form Factors, Phys. Rev. D 69, 074025 (2004).

[14] C. H. Chen, C. Q. Geng and Z. T. Wei, Factorization and Polarization in Two Charmed-Meson *B* Decays, hep-ph/0507295.

[15] J. L. Rosner, Determination of Pseudoscalar-Charmed-Meson Decay Constants from *B* Meson Decays, Phys. Rev. D 42, 3732 (1990).

[16] X. Y. Pham and Z. Z. Xing, *CP* Asymmetries in $B_d \to D^{*+}D^{*-}$ and $B_s \to D_s^{*+}D_s^{*-}$ Decays: *P*-wave Dilution, Penguin and Rescattering Effects, Phys. Lett. B 458, 375 (1999).

[17] A. Datta and D. London, Extracting γ from $B^0(t) \to D^{(*)+}D^{(*)-}$ and $B^0 \to D_s^{(*)+}D^{(*)-}$ Decays, Phys. Lett. B 584, 81 (2004).

[18] BABAR Collaboration, The BABAR Detector, Nucl. Instrum. Meth. A 479, 1 (2002).

[19] PDG, Review of Particle Physics: Passage of Particles Through Matter, Phys. Lett. B 592, 1 (2004).

[20] PDG, Review of Particle Physics: Mesons, Phys. Rev. D 66, 010001 (2002).

[21] G. C. Fox and S. Wolfram, Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes, Phys. Rev. Lett. 41, 1581 (1978).

[22] P. Billoir, Track Fitting with Multiple Scattering: A New Method, Nucl. Instrum. Meth. A 225, 352 (1984).

[23] DESY Collaboration, The ARGUS Electron-photon Calorimeter III. Electron-hadron Separation, Nucl. Instrum. Meth. A 237, 464 (1985).

[24] E687 Collaboration, Analysis of Three $D \to K\pi\pi$ Dalitz Plots, Phys. Lett. B 331, 217 (1994).

[25] W. D. Hulsbergen, Decay Chain Fitting with a Kalman Filter, Nucl. Instrum. Meth. A 552, 566 (2005).

[26] R. A. Fisher, The Use of Multiple Measures in Taxonomic Problems, Ann. Eugen. 7, 179 (1936).

[27] G. Punzi, Sensitivity of Searches for New Signals and its Optimization, physics/0308063 (2003).

[28] ARGUS Collaboration, Search for Hadronic $b \to u$ Decays, Phys. Lett. B 241, 278 (1990).

[29] M. Lefebvre, R. K. Keeler, R. Sobie and J. White, Propagation of Errors for Matrix Inversion, Nucl. Instrum. Meth. A 451, 520 (2000).

[30] G. Valencia, Angular Correlations in the Decay $B \to VV$ and *CP* Violation, Phys. Rev. D 39, 3339 (1989).

[31] BABAR Collaboration, Measurement of Time-Dependent *CP* Asymmetries and the *CP*-Odd Fraction in the Decay $B^0 \to D^{*+}D^{*-}$, Phys. Rev. Lett. 91, 131801 (2003).

[32] S. S. Wilks, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, Ann. Math. Stat. 9, 60 (1938).

[33] G. J. Feldman and R. D. Cousins, Unified Approach to the Classical Statistical Analysis of Small Signals, Phys.Rev. D 57, 3873 (1998).

[34] BABAR Collaboration, Measurement of the $B^0 \to D^{*-} D_s^{*+}$ and $D_s^+ \to \varphi\pi^+$ Branching Fractions, Phys. Rev. D 71, 091104 (2005).