Search for Beyond Standard Model Physics at *BABAR*

Thesis by

Yunxuan Li

In Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2022

Defended May 20, 2022

Yunxuan Li
ORCID: 0000-0001-9004-2471

ACKNOWLEDGEMENTS

I would like to express my deepest and sincerest gratitude to my advisors Prof. David G. Hitlin and Prof. Frank C. Porter. They introduced me to the field of experimental high energy physics, created a great research environment to allow me to try new things, and taught me how to be a great physicist as role models. I want to thank David for all the insightful discussions. During my PhD study, I was always impressed by his physics intuition generating new ideas, understanding complicated results, and getting insights from the numbers. It always enlightened me to think from first principles when I struggled with complex projects. I also want to thank him for sharing lots of past experiences in the high energy physics area. These cherished experiences give me a picture of how this field was developed and a sense of what I was really doing in a historical perspective. I also want to thank Frank for his continuous and detailed support on the thesis. His expertise and academic rigor in physics and statistics encouraged me to question authenticity and validate each hypothesis before taking them for granted. I also want to thank him for allowing me to test my immature and even crazy ideas.

I wish to thank Prof. Bertrand Echenard for initiating me into the area of dark matter and spending a great deal of time and effort discussing the project, diagnosing the analysis strategy, debugging the programs, and writing papers. I acknowledge his patience and hand-on support for my growth toward becoming a good researcher, and I really enjoyed the time we worked together.

I thank my group members Daniel Chao, Jae Hong Kim, Dexu Lin, James Y. Oyang, Léo Borrel, Sophie Middleton, Jason M. Trevor, and Viktor A. Shcherbatyuk for the help and encouragement during my Ph.D. journey.

I thank all the members in the *BABAR* Collaboration, especially Fabio Anulli, Gerard Bonneaud, and Janis McKenna. Thank you for the great opportunities to let me represent the collaboration and present my research in conferences.

I want to thank my good friend Prof. Yiqiu Ma, for not only his immeasurable support on my career development as a senior, but also his open-mindedness to share in-depth thoughts as a friend. I want to thank my good friends Baoyi Chen and Hao Xie. Thank you for your time spent with me on my random and sometimes naive ideas, and your companionship along the journey. I also want to thank Tao Liang, Ruofei Shen, Chenshuo Yue, Zilong Chen, Xingsheng Luan, Xiaolin Mao, Junlong Kou, Xiang Li, Boqiang Shen, Yu-An Chen, and Yichuan Song ... for all the fun times together.

I am extremely grateful to my parents and sister. Thank you for your love and unconditional support. I really thank my wife Yicheng Wang. Thank your for being in my life and always being there for me.

ABSTRACT

This thesis reports searches for beyond Standard Model physics in e^+e^- collisions in two directions.

We report the first search for a dark matter bound state. The existence of dark matter bound states could arise in a simple dark sector model in which a dark photon (A') is light enough to generate an attractive force between dark fermions. We report herein a search for a $J^{PC} = 1^{--}$ darkonium state, the Υ_D, produced in the reaction $e^+e^- \to \gamma\Upsilon_D, \Upsilon_D \to A'A'A'$, where the dark photons subsequently decay into pairs of leptons or pions, using data collected with the BABAR detector. No significant signal is observed, and we derive limits on the $\gamma - A'$ kinetic mixing (ϵ) as a function of the dark sector coupling constant for $0.001 < m_{A'} < 3.16$ GeV and $0.05 < m_{\Upsilon_D} < 9.5$ GeV. Bounds on the mixing strength ϵ down to $5 \times 10^{-5} - 10^{-3}$ are set for a large fraction of the parameter space.

We also report a measurement of $\mathcal{R}(D) = \mathcal{B}(B \to \bar{D}\tau\bar{\nu}_\tau)/\mathcal{B}(B \to \bar{D}\ell\bar{\nu}_\ell)$ and $\mathcal{R}(D^*) = \mathcal{B}(B \to \bar{D}^*\tau\bar{\nu}_\tau)/\mathcal{B}(B \to \bar{D}^*\ell\bar{\nu}_\ell)$, where ℓ refers to either an electron or muon. We select samples by reconstructing tag-side B mesons in semileptonic decays and signal-side τ in a purely leptonic decay. Using data collected with the BABAR detector, we measure $R(D) = 0.316 \pm 0.062(\text{stat}) \pm 0.019(\text{syst})$ and $R(D^*) = 0.226 \pm 0.022(\text{stat}) \pm 0.012(\text{syst})$, which agree with the Standard Model expectations by 0.26σ and 1.10σ, respectively. Taken together, the results are in agreement with the Standard Model within 1.51σ level.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Part I

Introduction

Chapter 1

THE STANDARD MODEL OF PARTICLE PHYSICS

The ultimate goal of particle physics is to understand the fundamental law of matter in the universe, which is made of elementary particles. In other words, elementary particles are the building blocks of the universe and are thought to have no internal structure. According to the idea of reductionism, if we understand the characteristics of elementary particles and how they interact with each other, we are able to fully explain and link together all physical phenomena in the universe. The corresponding theory to model elementary particles and their interactions is often referred as the Theory of Everything (TOE).

Developed during the twentieth century, the standard model (SM) is a quantum field theory to describe three of four fundamental forces: electromagnetism, strong, and weak interactions. Although it does not include gravity, the standard model is currently our best attempt to unify the fundamental interactions together, and it has been extensively tested by a series of experiments. Some famous experiments include discovery of Z boson at CERN, W boson at CERN and Fermilab, J/ψ meson at SLAC and BNL, and Higgs boson at CERN.

The Lagrangian of the standard model is

$$\begin{aligned} L_{SM} = &- \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}\mathrm{Tr}[\boldsymbol{W}_{\mu\nu}\boldsymbol{W}^{\mu\nu}] - \frac{1}{4}\mathrm{Tr}[\boldsymbol{G}_{\mu\nu}\boldsymbol{G}^{\mu\nu}] \\ &+ i\bar{\psi}\not{D}\psi + h.c. \\ &+ \sum_{i,j=e,\mu,\tau} \psi_i y_{ij}\psi_j\phi + h.c. \\ &+ |D_\mu\phi|^2 - V(\phi). \end{aligned} \tag{1.1}$$

The first term is the gauge term, where $B_{\mu\nu}, \boldsymbol{W}_{\mu\nu}$ and $\boldsymbol{G}_{\mu\nu}$ are the field tensors for the U(1), SU(2), and SU(3) groups. Using the weak mixing angle θ_w, the $A_\mu, Z_\mu, W^\pm_\mu$ bosons are mixtures of $\boldsymbol{W}_{\mu\nu}$ and $\boldsymbol{G}_{\mu\nu}$ components:

$$\begin{aligned} &A_\mu = W_{11\mu}\sin\theta_w + B_\mu\cos\theta_w,\ B_\mu = A_\mu\cos\theta_w - Z_\mu\sin\theta_w \\ &Z_\mu = W_{11\mu}\cos\theta_w - B_\mu\sin\theta_w,\ W_{11\mu} = -W_{22\mu} = A_\mu\sin\theta_w + Z_\mu\cos\theta_w \\ &W^+_\mu = W^{-*}_\mu = W_{12\mu}/\sqrt{2},\ W_{12\mu} = W^*_{21\mu} = \sqrt{2}W^+_\mu. \end{aligned} \tag{1.2}$$

The second term involves kinetic terms of fermions and how fermions interact with gauge fields. ψ is a vector including all fermion fields. Explicitly, the Lagrangian can be re-written as

$$\begin{aligned} L_{F,F-G} =& i\bar{\psi}\not{D}\psi + h.c. \\ =& (\bar{\nu}_L, \bar{l}_L)\tilde{\sigma}^{\mu} iD_{\mu} \begin{pmatrix} \nu_L \\ l_L \end{pmatrix} + \bar{l}_R\sigma^{\mu} iD_{\mu} l_R + \bar{\nu}_R\sigma^{\mu} iD_{\mu}\nu_R + h.c. \\ &+ (\bar{u}_L, \bar{d}_L)\tilde{\sigma}^{\mu} iD_{\mu} \begin{pmatrix} u_L \\ d_L \end{pmatrix} + \bar{u}_R\sigma^{\mu} iD_{\mu} u_R + \bar{d}_R\sigma^{\mu} iD_{\mu} d_R + h.c., \end{aligned} \tag{1.3}$$

where $l = (e, \mu, \tau), \nu = (\nu_e, \nu_{\mu}, \nu_{\tau}), u = (u, c, t)$, and $d = (d, s, b)$ are all three-component generation indices. The σ are the Pauli matrices satisfying

$$\begin{aligned} \sigma^{\mu} &= \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right] \\ \tilde{\sigma}^{\mu} &= [\sigma^0, -\sigma^1, -\sigma^2, -\sigma^3]. \end{aligned} \tag{1.4}$$

D_{μ} is covariant derivative combining electromagnetic, weak, and strong interactions in the form of

$$D_{\mu} = \partial_{\mu} - ig'YB_{\mu} - igW_{\mu}^a T^a - ig_s G_{\mu}^a t^a$$

where g', g, g_s are the coupling strengths of the hypercharge, weak and strong interaction. Y, T^a, t^a are the corresponding hypercharge operator, SU(2) generator, and SU(3) generator.

The interaction between fermions and gauge fields can be further decomposed to charged and neutral currents. Taking leptons as an example, its charged current part of the Lagrangian is

$$\begin{aligned} L_L^{CC} &= -\frac{g}{2\sqrt{2}}(j_{W,L}^{\mu} W_{\mu} + h.c.) \\ j_{W,L}^{\mu} &= \bar{\nu}_l \gamma^{\mu}(1 - \gamma^5) l. \end{aligned} \tag{1.5}$$

Its neutral current includes the QED Lagrangian as well as a weak neutral current:

$$\begin{aligned} L_L^{NC} &= L_L^{\gamma} + L_L^{Z} \\ L_L^{\gamma} &= -ej_{\gamma,L}^{\mu} A_{\mu},\ j_{\gamma,L}^{\mu} = -\bar{e}\gamma^{\mu} e \\ L_L^{Z} &= -\frac{g}{2\cos\theta_w} j_{Z,L}^{\mu} Z_{\mu},\ j_{Z,L}^{\mu} = \bar{\nu}_l \gamma^{\mu}(c_V^{\nu} - c_A^{\nu}\gamma^5)\nu_l + \bar{l}\gamma^{\mu}(c_V^{l} - c_A^{l}\gamma^5) l. \end{aligned} \tag{1.6}$$

The last two terms are related to the Higgs mechanism. They describe the Higgs field and how the Higgs field gives mass to gauge and fermion fields. For the

Higgs-gauge interaction, since the hypercharge $Y = +1$, we have

$$D_\mu = \partial_\mu + ig\boldsymbol{W}_\mu \cdot \frac{\boldsymbol{\sigma}}{2} + ig'\boldsymbol{B}_\mu \cdot \frac{1}{2}.$$

So the Higgs-gauge term can be written as

$$\begin{aligned} L_{H,G} =& \frac{1}{2}(\partial_\mu H)(\partial^\mu H) - \lambda v^2 H^2 - \lambda v H^3 - \frac{1}{4}\lambda H^4 \\ &+ \frac{v^2 g^2}{4} W_\mu^+ W^{-\mu} + \frac{v^2 g^2}{8\cos\theta_w{}^2} Z_\mu Z^\mu \\ &+ \frac{v g^2}{2} W_\mu^+ W^{-\mu} H + \frac{v^2 g^2}{4\cos\theta_w{}^2} Z_\mu Z^\mu H + \frac{v^2 g^2}{4} W_\mu^+ W^{-\mu} H^2 + \frac{v^2 g^2}{8\cos\theta_w{}^2} Z_\mu Z^\mu H^2, \end{aligned} \tag{1.7}$$

from which we can get the mass of the Higgs as well as the mass of W and Z bosons:

$$\begin{aligned} m_H &= \sqrt{2\lambda v^2} \\ m_W &= \frac{vg}{2} \\ m_Z &= \frac{vg}{2\cos\theta_w}. \end{aligned} \tag{1.8}$$

Since the mass term of fermion fields $m\bar{\psi}\psi = m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$ is not gauge invariant, the way to give mass to fermions is via Yukawa coupling. Once diagonalized, it can be written as

$$\begin{aligned} L_{Yukawa} &= \sum_{i,j=e,\mu,\tau} \psi_i y_{ij} \psi_j \phi \\ &= \left(\frac{y_l v}{\sqrt{2}}\right) \bar{l}l - \frac{y_l}{\sqrt{2}} h\bar{l}l \\ &+ \left(\frac{y_u v}{\sqrt{2}}\right) \bar{u}u - \frac{y_u}{\sqrt{2}} h\bar{u}u \\ &+ \left(\frac{y_d v}{\sqrt{2}}\right) \bar{d}d - \frac{y_d}{\sqrt{2}} h\bar{d}d, \end{aligned} \tag{1.9}$$

so the mass of leptons and quarks are

$$\begin{aligned} m_i &= \frac{y_l^i v}{\sqrt{2}}, i = e, \mu, \tau \\ m_i &= \frac{y_u^i v}{\sqrt{2}}, i = u, c, t \\ m_i &= \frac{y_d^i v}{\sqrt{2}}, i = d, s, b. \end{aligned} \tag{1.10}$$

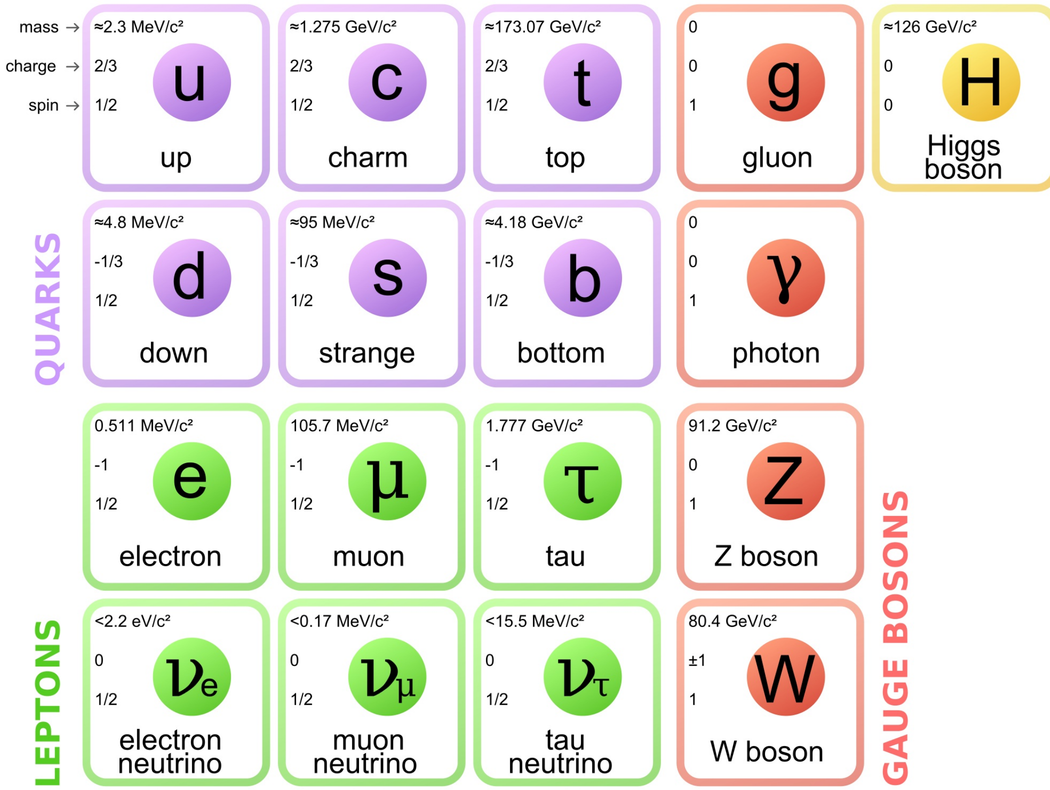

Figure 1.1: The Standard Model of particle physics. The picture is taken from [1].

Fig. 1.1 shows all the elementary particles in the standard model. In conclusion, the standard model has 17 types of elementary particles. 12 of them are fermions, which are the building blocks of the matter in our universe. The two type of fermions, *quark* and *lepton*, are distinguished by whether they carry color charge. In addition, we have 4 gauge bosons which act as force carriers: photon for electromagnetism, gluon for strong interaction, and W and Z boson for weak interaction. We also have the Higgs boson to explain the mass of the other particles.

Chapter 2

BEYOND STANDARD MODEL PHYSICS

Although the standard model has been tested by many experiments, it is still an incomplete theory due to the lack of explanation of the following phenomena and theoretical puzzles:

- Gravity: the standard model does not explain gravity, which is currently best described with general relativity and explains the gravitational force due to curved space-time from massive objects.
- Dark matter: The existence of dark matter has been supported by many cosmological and astrophysical observations, and it is measured to be 4-5 times that of ordinary matter in our universe. However, particles and interactions in the SM only explain the physics of ordinary matter, leaving dark matter unexplained.
- Dark energy: Dark energy is around 70% of the entire energy in our universe, and it is also not explained in the SM.
- Neutrino masses: the neutrinos in the SM are massless, however, neutrino oscillation experiments show that they do have mass.
- Hierarchy problem: The SM introduces a Higgs field to explain particle masses, however, the mass parameter in the Higgs field needs to be fine-tuned to cancel quantum correlations. However, relying on a numerical cancellation to explain the large discrepancy between weak force and gravity is uncomfortable for many physicists.
- Strong CP problem: The CP symmetry in the strong interaction sector is allowed to be broken in theory, while no experiments observe such violation.

Therefore, one of the main goals of particle physics is to search for and understand the physics beyond the SM. To search for BSM physics, particle physics has three categories of approaches, referred to as the energy frontier, the cosmic frontier, and the intensity frontier. While the energy and cosmic frontiers often provide a direct search for the production of new particles, the intensity frontier provides indirect probes of new physics with intense beam sources and super-sensitive detectors. With much better sensitivity, intensity frontier experiments can probe for new physics at mass scales of hundreds to thousands of TeV.

Part II

PEP-II and *BABAR*

The primary goal of the experiment is to study CP violation in the decays of B mesons. With its high luminosity, a sensitive measurement of the CKM matrix element V_{ub} can be made, which is a fundamental parameter in the Standard Model. This high-luminosity experiment is designed such that electrons and positrons collide at the $\Upsilon(4S)$ resonance, which then decays to $B\bar{B}$ pairs more than 95% of the time. The B mesons subsequently decay to other particles, which can be detected or reconstructed from specific sub-systems of the *BABAR* detector. At the hardware level, charged or neutral final-state particles generate hits or clusters in the detector. The optimized particle identification algorithms (PID) or cluster-finding algorithms are applied to identify final-state particles and reconstruct the entire decay process.

Besides the $\Upsilon(4S)$ resonance, the *BABAR* experiment also operated at the $\Upsilon(3S)$ and $\Upsilon(2S)$ resonances, as well as off-resonance to better understand the continuum backgrounds. The *BABAR* experiment can be also applied to study a range of other physics including, but not limited to, other B meson decays, the physics of charm and τ leptons, and two-photon physics.

Chapter 3

THE PEP-II ACCELERATOR

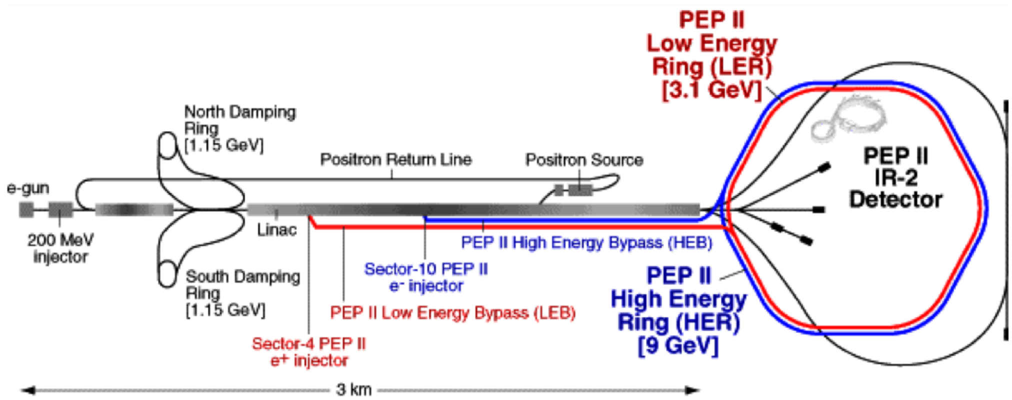

Figure 3.1: The design of the PEP-II accelerator.

The PEP-II accelerator is designed to deliver the B mesons to the experiment. It is an e^+e^- collider operating at the $\Upsilon(4S)$ resonance of 10.58 GeV. The two storage rings are asymmetric with 9 GeV for the electron beam and 3.1 GeV for the positron beam, which is a key design feature of PEP-II, as shown in Fig. 3.1. This design results in B mesons with significant momenta in the lab frame, and its decay length can be measured.

The PEP-II system includes the following four major subsystems: Injector, High-Energy Ring (HER), Low-Energy Ring (LER), and Interaction Region (IR) [2]. The Injector includes the extraction and transport lines from the SLAC two-mile linac. The HER is responsible for the 9 GeV electron beam. It is originated from PEP ring but with new vacuum, RF, feedback, and diagnostics systems. The LER is a new ring mounted above the 9-GeV HER, responsible for the 3.1 GeV positron ring. The IR is designed to focus the beam spots, bring two beams into head-on collision, and separate them cleanly.

Chapter 4

THE *BABAR* DETECTOR

In order to achieve the sensitivity required, the *BABAR* detector needs to have maximum possible acceptance in the center-of-mass (CM) system, excellent vertex resolution to measure B meson decay time, and great discrimination between charged particles to tag B meson decays. The detector mainly includes six components: Silicon Vertex Tracker (SVT), Drift Chamber (DCH), Detector of Internally Reflected Cherenkov light (DIRC), Caesium Iodide Electromagnetic Calorimeter (EMC), a 1.5 T superconducting coil, and Instrumented Flux Return (IFR). The details of design [3] are summarized below.

4.1 Silicon Vertex Tracker (SVT)

The main goal of the SVT is to reconstruct the decay vertices of the two B mesons to determine the time between two decays, and measure time-dependent CP asymmetries in B meson decays. The SVT is also responsible for the detection of low-momentum charged particles which do not reach the drift chamber.

As shown in Fig. 4.1, the SVT includes five concentric cylindrical layers of double-sided silicon detectors, divided in azimuth into modules. The inner three layers are barrel-style structures with six detector modules. The outer two layers have 16 and 18 detector modules, respectively, and are employed with a new arch structure so that detectors are electrically connected across an angle.

The silicon vertex detectors use double-sided silicon strip detectors coupled with polysilicon bias resistors. The p^+ and n^+ strips are fabricated on around 300-μm-thick high-resistivity silicon. The front-end signal processing is performed by ICs mounted on hybrid circuits, fabricated in a radiation-hard CMOS technology.

4.2 Drift Chamber (DCH)

The drift chamber is the main tracking system of the *BABAR* detector used to reconstruct tracks with transverse momentum larger than 100 MeV/c. The performance goal of the DCH includes a spatial resolution better than 140 μm, a PID for low momentum tracks by dE/dx with resolution of 7%, and a resolution of $\sigma_{p_t} \approx 0.3\% \times p_t$ for momentum above 1 GeV/c.

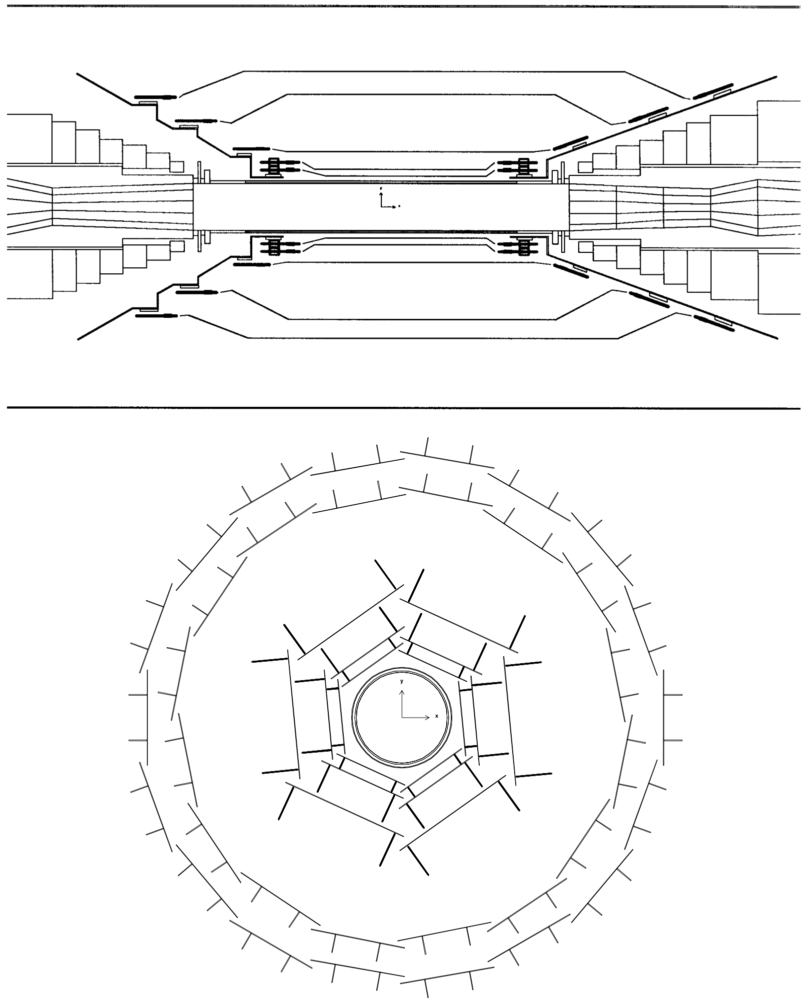

Figure 4.1: The cross-sectional view of the silicon vertex tracker design (upper) along the beam axis; (bottom) orthogonal to the beam axis.

The side view and layout of drift chamber are shown in Fig. 4.2. The drift chamber is a 280-cm-long cylinder. Its flat endplates are made of aluminum. The forward endplate is made thinner in the acceptance region of the detector compared with the rear endplate. The drift cells are arranged in 10 superlayers of 4 layers each, scheduled using the pattern AUVAUVAUVA for axial (A) and stereo (U, V) superlayers.

4.3 Detector of Internally Reflected Cherenkov Light (DIRC)

As shown in Fig. 4.3, the DIRC is designed to provide excellent kaon identification to distinguish between the two-body decay modes $B^0 \to \pi^+\pi^-$ and $B^0 \to K^+\pi^-$. It is also designed to identify muons in the low-momentum range. The DIRC system includes 144 long, straight bars of synthetic quartz arranged in a 12-sided polygonal

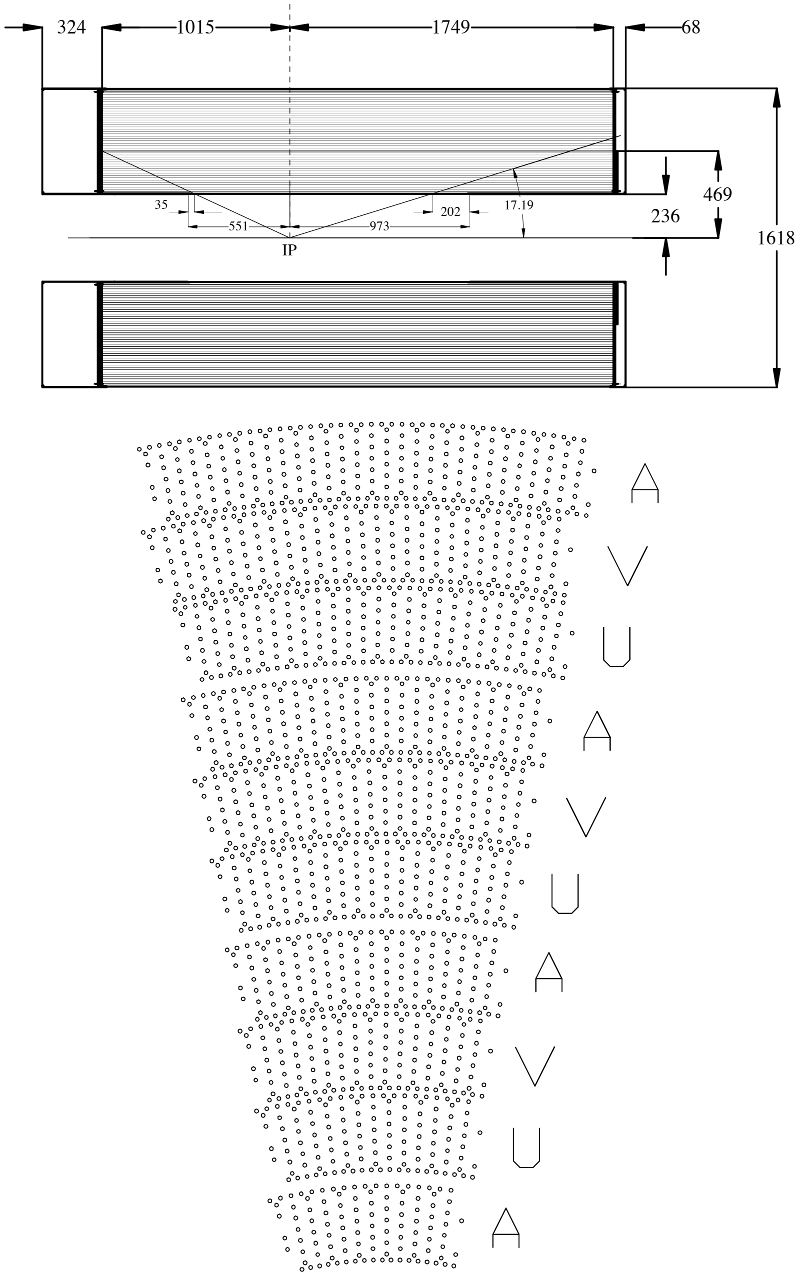

Figure 4.2: (upper) The side view of drift chamber; (bottom) The cell layout of the drift chamber.

barrel.

4.4 Electromagnetic Calorimeter (EMC)

The primary goal of EMC is to measure the energy of photon and electrons with precise energy and angular resolution. When a photon or an electron travels through the crystal, it interacts with the thallium-doped cesium iodide (CsI(Tl)) and triggers

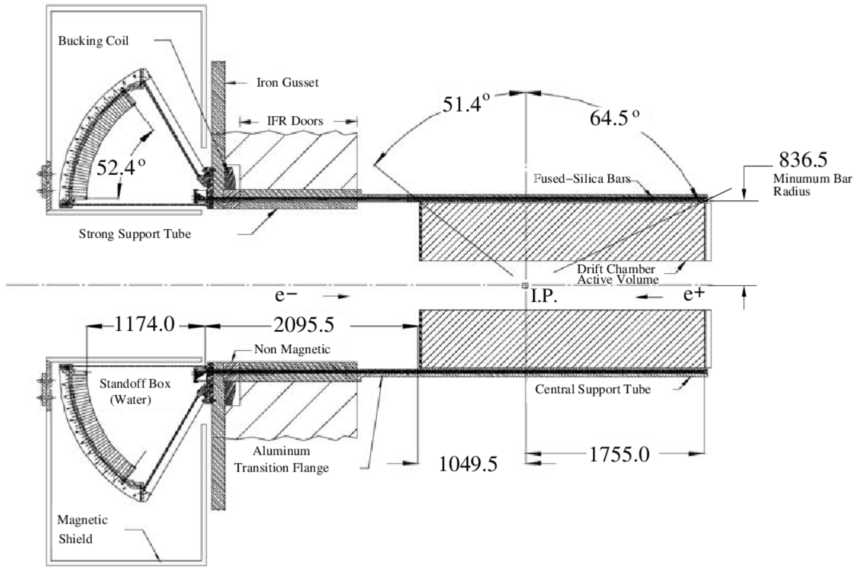

Figure 4.3: The geometry of DIRC.

an electromagnetic cascade and deposits its energy to create scintillation light. The photons are then reflected inside the crystal and collected with the PIN diodes.

As shown in Fig. 4.4, the calorimeter includes a cylindrical barrel section and a forward conic endcap. It is located asymmetrically about the interaction point with the inner radius extending 112.7 cm in the backward direction and 180.9 cm in the forward direction. The barrel crystals covers angle between $-0.775 \leq \cos\theta \leq 0.892$ in the laboratory frame, corresponding to $-0.916 \leq \cos\theta \leq 0.715$ in the c.m.s. frame.

4.5 Instrumented Flux Return (IFR)

The IFR shown in Fig. 4.5 is designed to return the flux of the 1.5T magnet and to provide muon identification and neutral hadron detection. The system consists of a central part (Barrel) and two plugs (End Caps) to cover the angle from 300 mrad in the forward direction to 400 mrad in the backward direction. The resistive plate chambers (RPC) are constructed of bakelite sheets separated by thick polycarbonate spacers with cylindrical symmetry and fiberglass frame.

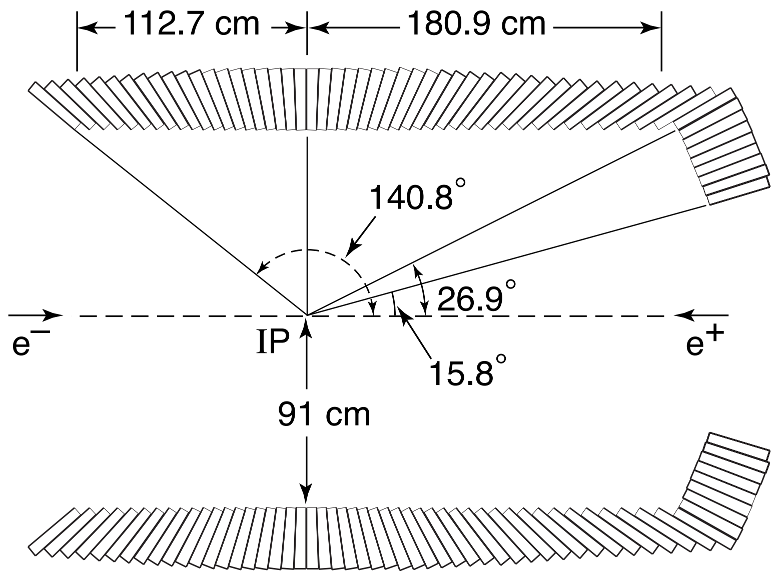

Figure 4.4: The side view of the electromagnetic calorimeter barrel and forward endcap.

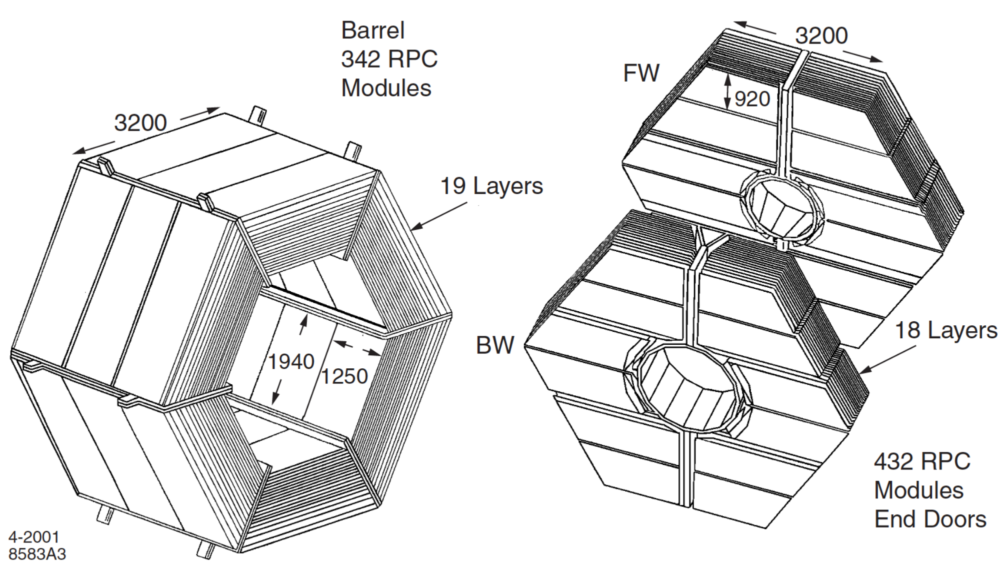

Figure 4.5: The overview of the IFR system.

Chapter 5

DATA RECORD AND LUMINOSITY

The *BABAR* experiment operated from 1999 to 2008. An integrated luminosity of 531 fb^{-1} was collected mostly at the $\Upsilon(4S)$ resonance, but also at the $\Upsilon(3S)$, $\Upsilon(2S)$, as well as off-resonance. Fig. 5.1 shows the integrated luminosity over time. The luminosity for each run is shown in Table 5.1.

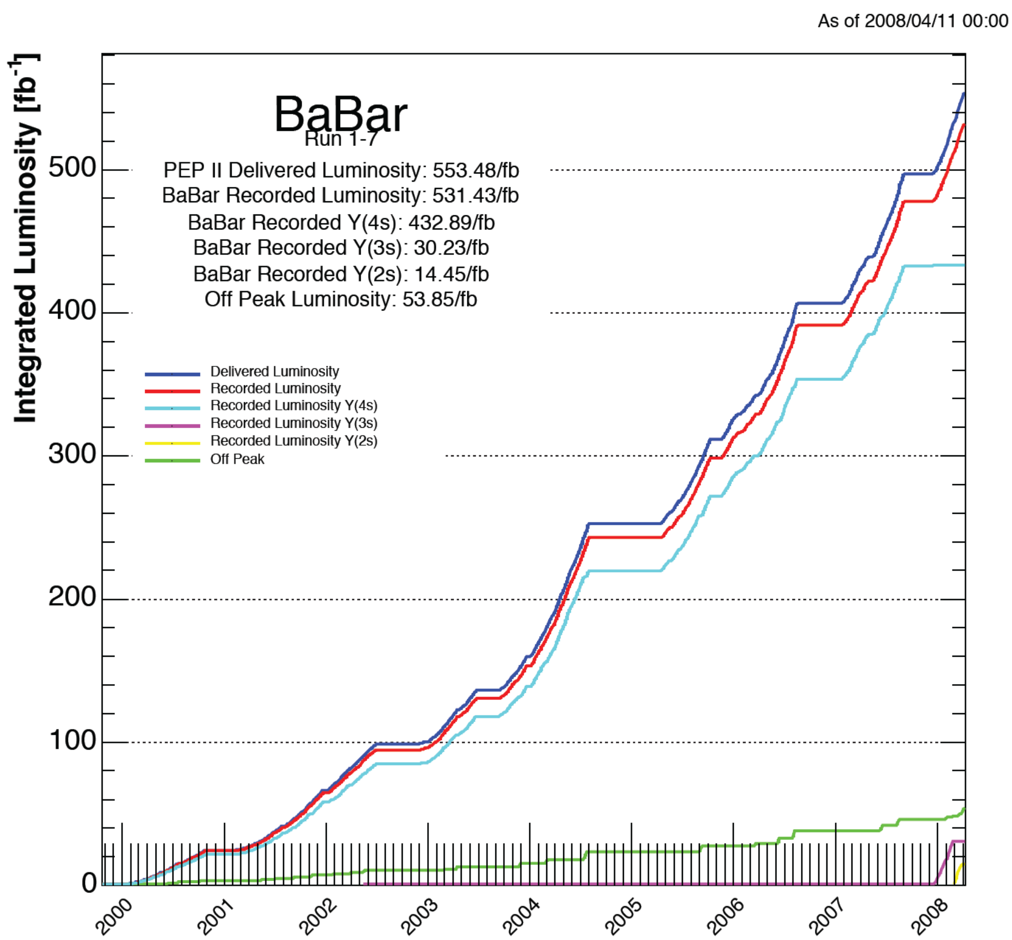

Figure 5.1: The integrated luminosity over time.

Run period	Luminosity (fb^{-1})		Data sample on-peak
	on-peak	off-peak	
Run1	20.4	2.6	$\Upsilon(4S)$
Run2	61.3	6.7	$\Upsilon(4S)$
Run3	32.3	2.4	$\Upsilon(4S)$
Run4	99.6	10.0	$\Upsilon(4S)$
Run5	133.2	14.3	$\Upsilon(4S)$
Run6	78.3	7.8	$\Upsilon(4S)$
Run7	27.9	2.6	$\Upsilon(3S)$
Run8	13.6	1.4	$\Upsilon(2S)$
Total	465.7	47.9	

Table 5.1: Integrated luminosity of on-peak and off-peak data collected for each run.

Part III

Search for Dark Matter at *BABAR*

Chapter 6

SEARCH FOR DARKONIUM IN ELECTRON-POSITRON COLLISIONS

6.1 Introduction

The existence of dark matter is overwhelmingly supported by astrophysical and cosmological observations, and it is one of the most important tasks of contemporary physics to understand its nature and properties. The concept of dark matter can be traced back to 1880s, when Lord Kelvin estimated the number of dark bodies in the Milky Way from the observed velocity dispersion of the stars orbiting the center of the galaxy. From his results, he concluded that many of our stars may be dark bodies. In 1933, Fritz Zwicky reached a similar conclusion by applying the virial theorem to the Coma Cluster in an attempt to estimate its mass, finding a result 400 times larger than the visible mass. During the last decades, much evidence for dark matter has been accumulated from measurements of galaxy rotation curves, velocity dispersions, galaxy clusters, and gravitational lensing.

Many dark matter candidates have been proposed during the last decades. A long-leading paradigm is in the form of Weakly Interacting Massive Particles (WIMPs), as the dark matter self-annihilation cross section is roughly consistent with a weak-scale massive particle interacting via the electroweak force. Many beyond Standard Model theories, such as Supersymmetry [4] or models containing extra dimensions, include WIMP-like candidates. However, the absence of observation of WIMPs so far has motivated the exploration of alternative models, such as light dark matter and the existence of a dark sector.

The dark sector (also known as hidden sectors) is a collection of particles that do not interact directly with SM particles, i.e. the dark sector fields are singlets under the SM gauge groups. They are motivated by many BSM theories, for example, in string theory constructions and type-II compactifications. The dark sector has its own symmetries, which could be arbitrarily complex. If the dark sector contains an extra U(1) symmetry, the dark and visible sectors are indirectly connected via the so-called vector portal, namely they can interact via kinetic mixing between the dark gauge field (A'_μ) and the SM hypercharge field $F^{\mu\nu}$:

$$\epsilon F^{\mu\nu,Y} F_{\mu\nu,D} \qquad F_{\mu\nu,D} = \partial_\mu A'_\nu - \partial_\nu A'_\mu,$$

with ϵ denoting the mixing strength, as shown in Fig. 6.1. Values of mixing strength in the $10^{-12} - 10^{-3}$ range have been predicted in the literature [5, 6, 7]. The kinetic mixing term can be removed by redefining the vector field

$$A'_\mu \longrightarrow A'_\mu, \;\; A_\mu \longrightarrow A_\mu + \epsilon A'_\mu,$$

so that SM fermions pick up a small dark charge $\sim \epsilon e$ leading to "milli-charged" interaction between the two sectors. The dark photon can be massive, and its mass can arise via the Higgs or the Stueckelberg mechanisms. Its mass range is predicted to be in the MeV-GeV range [8, 9, 10, 11], though much smaller (sub-eV) masses are also possible [12].

Besides the vector portal, there are a few other indirect interactions that can connect the dark sector to the SM, including the scalar and Fermion portals. In this analysis we focus on the vector portal.

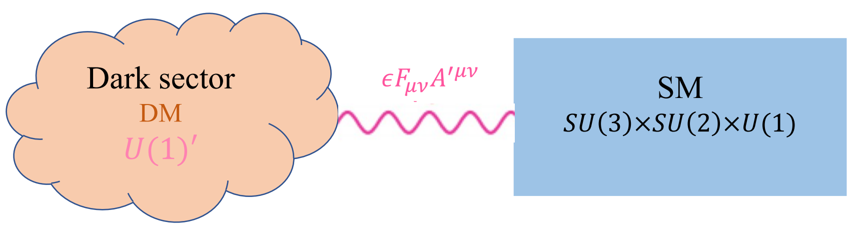

Figure 6.1: The dark sector scenario under study: The dark sector has an additional $U(1)'$ group and the corresponding dark photon A' (γ' in the picture) plays a role as intermediate messenger between SM and dark sector (hidden sector).

6.2 Dark Matter Bound States Model

The possibility of dark sector bound states was first proposed by An et al.[13]. A specific model contains a Dirac dark matter field (χ) charged under an additional U(1) gauge group in the dark sector, with the corresponding vector boson acting as a mediator between the dark sector and SM via kinetic mixing. The mass of dark matter particles is O(few GeV). The Lagrangian for this model is

$$\begin{aligned}\mathcal{L} =& \mathcal{L}_{SM} + \bar{\chi} i \gamma^{\mu}(\partial_{\mu} - i g_D A'_{\mu})\chi - m_{\chi}\bar{\chi}\chi \\ & - \frac{1}{4}A'_{\mu\nu}A'^{\mu\nu} - \frac{\epsilon}{2}F_{\mu\nu}A'^{\mu\nu} + \frac{1}{2}m^2_{A'}A'_{\mu}A'^{\mu},\end{aligned} \tag{6.1}$$

where A'_{μ} is the vector boson mediator with field strength $A'_{\mu\nu}$, $F_{\mu\nu}$ is the SM hypercharge field, ϵ is the kinetic mixing strength, and $\alpha_D = \frac{g_D^2}{4\pi}$ is the strength of the dark electromagnetic interaction. When g_D is sufficiently large, the force between the dark fermions mediated becomes attractive, resulting in the formation of dark matter bound states ($\chi\bar{\chi}$).

The two lowest (1S) bound states, $^1S_0$ and $^3S_1$, are respectively denoted η_D and Υ_D in analogy with the SM. The critical conditions for the existence of stable bound states has been determined numerically [14],

$$1.68 m_{A'} \leq \alpha_D m_{\chi}.$$

If we assume $\alpha_D = 0.5$ and $m_{\chi} = 3.5$ GeV, the dark photon must be lighter than 1 GeV, as shown in Fig. 6.2. The value of α_D has to be large enough to produce a bound state; m_{χ} higher than 3.5 GeV has been excluded by other measurements [15, 16].

The quantum numbers of the two lowest (1S) bound states suggest the following production mechanisms at electron-positron colliders:

$$e^+e^- \to \eta_D + A'$$
$$e^+e^- \to \Upsilon_D + \gamma.$$

These production processes are mediated by a mixed $\gamma - A'$ propagator, as shown in Fig. 6.3. The η_D (Υ_D) decays into 2 (3) pairs of dark photons, leading to multi-lepton and/or multi-quark final states.

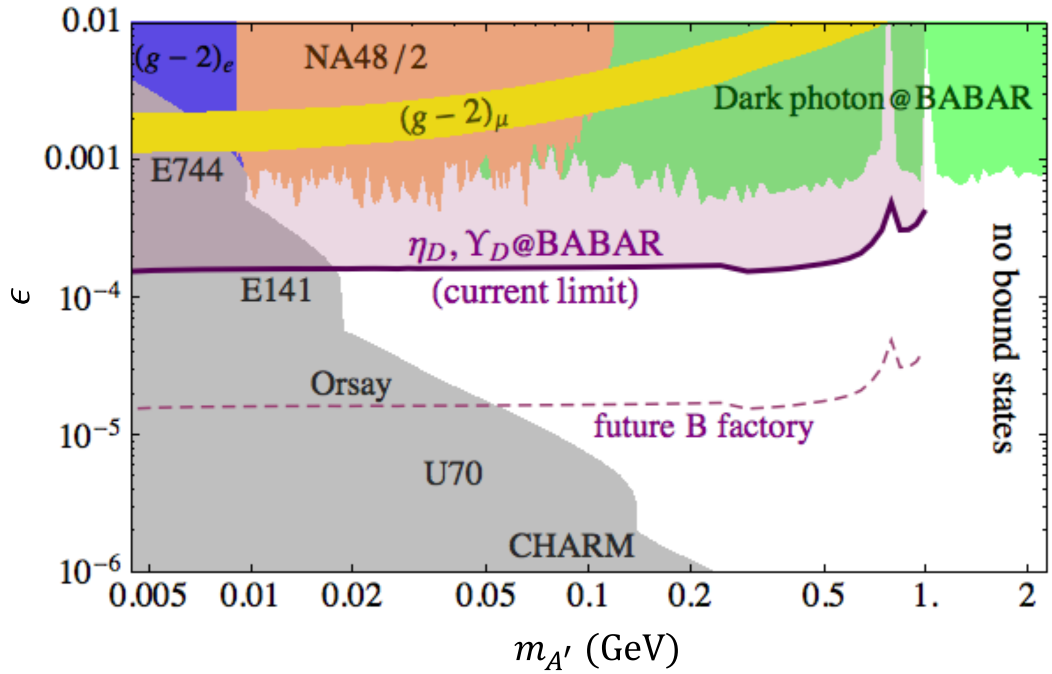

Figure 6.2: Constraints on the dark photon parameter space from a re-interpretation of the *BABAR* dark Higgsstrahlung searches. The solid purple curve corresponds to the current *BABAR* limit for the parameters $\alpha_D = 0.5$, $m_\chi = 3.5$ GeV. The figure is taken from [13].

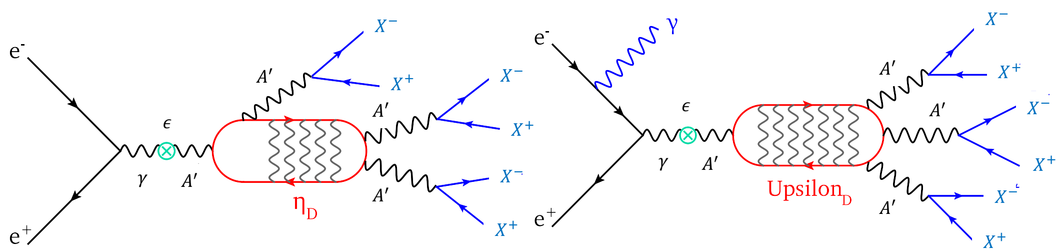

Figure 6.3: Diagram for η_D and Υ_D production and decay at *BABAR* . The figure is taken from [13].

This analysis is searching for Υ_D production. The e^+e^- emit an initial state radiation (ISR) photon to produce a Υ_D state. The Υ_D then decays into three dark photons, and these dark photons subsequently decay to lepton or quark-antiquark pairs. In this analysis, we will only search for $A' \to X^+X^-(X = e, \mu, \pi)$.

The Υ_D production cross-section via initial state radiation is calculated in [13] by applying a non-relativistic expansion to the dark matter field in the Υ_D rest frame, and using the relation between the matrix element and the wave function [17] to derive the effective kinetic mixing term between Υ_D and the photon

$$\mathcal{L}_{eff} = -\frac{1}{2}\epsilon\epsilon_D F_{\mu\nu}\Upsilon_D^{\mu\nu}, \;\; \epsilon_D = \sqrt{\frac{\alpha_D}{2m_\chi^3}}R_{\Upsilon_D}(0). \tag{6.2}$$

When $m_{A'} \ll \alpha_D m_\chi$, the following differential cross section is obtained

$$\begin{aligned}\frac{d\sigma_{e^+e^-\to\gamma\Upsilon_D}}{d\cos\theta} \approx & \frac{2\pi\alpha^2\epsilon^2\epsilon_D^2}{s}\left(1-\frac{4m_\chi^2}{s}\right) \\ & \times\left[\frac{8s^2(s^2+16m_\chi^4)\sin^2\theta}{(s-4m_\chi)^2(s+4m_e^2-(s-4m_e^2)\cos 2\theta)^2}-1\right],\end{aligned} \tag{6.3}$$

where θ is the angle between γ and the initial e^- in the center-of-mass frame.

The Υ_D particle will subsequently decay to three dark photons; the differential decay rate is calculated by generalizing the approach in [18] to massive dark photon case,

$$\begin{aligned}&\frac{d\Gamma(\Upsilon_D\to A'A'A')}{dx_1dx_2} = \frac{2\alpha_D^3[R_{\Upsilon_D}(0)]^2}{3\pi m_\chi^2} \\ &\times\frac{39x^8+4x^6F_6-16x^4F_4+32x^2F_2+256F_0}{(x^2-2x_1)^2(x^2-2x_2)^2(x^2+2(x_1+x_2-2))^2},\end{aligned} \tag{6.4}$$

where $x_{1,2} = E_{1,2}/m_\chi$, $x = m_{A'}/m_\chi$, and

$$\begin{aligned}F_6 &= x_1^2+(x_1+x_2)(x_2-2)-30 \\ F_4 &= (x_1^2+x_1x_2-2x_1)(3x_2-10)-10x_2(x_2-2)-21 \\ F_2 &= x_1^4+2x_1^3(x_2-2)+x_1^2(x_2(3x_2-22)+28)+ \\ &\quad 2x_1(x_2-2)(x_2(x_2-9)+12)+x_2(x_2-2)(x_2(x_2-2)+24)+24 \\ F_0 &= x_1^4+2x_1^3(x_2-2)+x_1^2(3x_2(x_2-3)+7)+ \\ &\quad x_1(x_2-1)(x_2-2)(2x_2-3)+(x_2-1)^2(x_2(x_2-2)+2).\end{aligned} \tag{6.5}$$

We assume that the coupling of dark photons to SM fermions is universal and proportional to their charge. Based on the following constraints,

$$\begin{aligned}&\bullet\ \mathcal{B}(A'\to hadrons)/\mathcal{B}(A'\to l^+l^-) = \sigma(e^+e^-\to hadrons)/\sigma(e^+e^-\to l^+l^-) = R, \\ &\bullet\ \sum_{l=e,\mu,\tau}\mathcal{B}(A'\to l^+l^-)+\mathcal{B}(A'\to hadrons) = 1,\end{aligned} \tag{6.6}$$

we obtain the dark photon branching ratios into specific final states, as shown in Fig. 6.4.

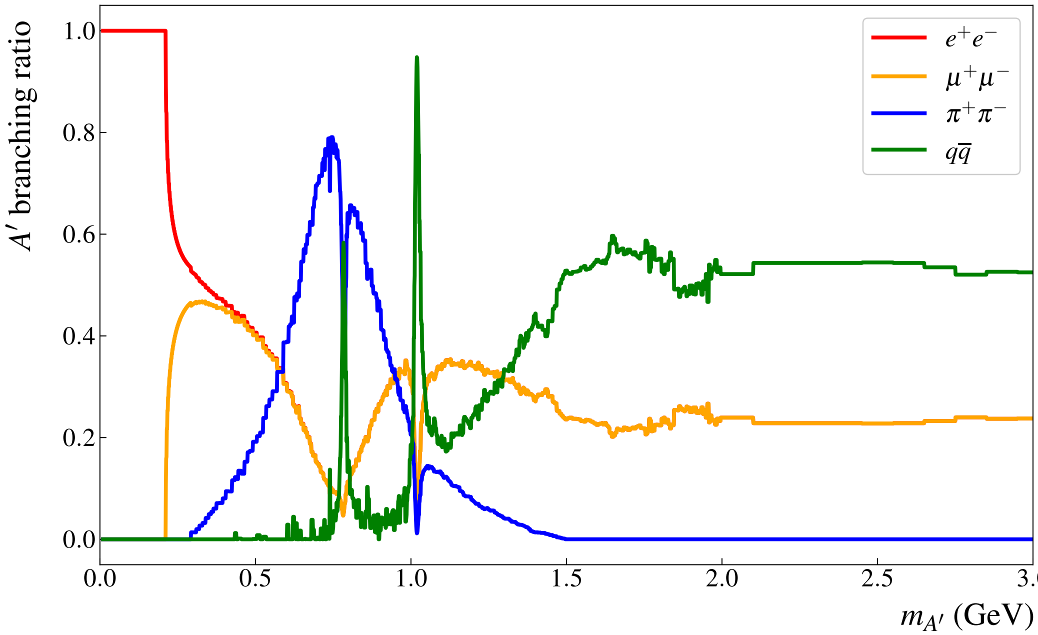

Figure 6.4: Dark photon branching ratios to specific final states as a function of the dark photon mass. The two peaks around 0.78 GeV and 1.0 GeV for $q\bar{q}$ correspond to the ω and ϕ resonances.

Integrating (6.3) over θ, and combining with (6.4) and the dark photon branching ratios, we obtain the total cross section of the signal channel $e^+e^- \to \gamma\Upsilon_D, \Upsilon_D \to A'A'A', A' \to X^+X^-(X = e, \mu, \pi)$, which is a function of ϵ and $m_{A'}$ only.

One complication arises from the dark photon lifetime, which can become sufficiently large to produce displaced decay vertices and affect the mass resolution and signal efficiency. The partial decay width of the dark photon in the case of $m_{A'} > 2m_e$ is given by $\Gamma \approx \frac{1}{3} m_{A'} \alpha \epsilon^2$, where $m_{A'}$ is the dark photon mass, α is the strength of SM electromagnetic interaction, ϵ is the kinetic mixing strength. Therefore, dark photon decay length l in the laboratory frame is

$$l = \gamma v c \tau = \frac{p}{m^2} \cdot \frac{3\hbar c}{\alpha \epsilon^2}, \tag{6.7}$$

which becomes significant when the kinetic mixing and the dark photon mass are small. For instance, if $m_{A'} = 10$ MeV and $\epsilon = 10^{-4}$, the flight length of dark photon is around 0.2 m. The effect of the dark photon flight length becomes non-negligible above $O(100\ \mu\text{m})$, a few times the resolution of the beam spot location.

6.3 Dataset and Signal Monte Carlo Simulations

Signal MC events are generated using MadGraph, which calculates matrix elements from first principles. Our final states consist of three pairs of leptons or pions, leading to 10 different possibilities. We do not consider the $3\pi^+\pi^-$ combination, as this channel has much more background and much lower signal-to-noise ratio than the other possibilities. These final states can be divided into three categories based on the number of pion pairs: $0\pi^+\pi^-$ pair (C_0), $1\pi^+\pi^-$ pair (C_1) and $2\pi^+\pi^-$ pairs (C_2).

We simulate zero-lifetime signal MC events in the range $0 < m_{\Upsilon_D} \leq 10$ GeV and $0 < m_{A'} \leq 3$ GeV. The number of events simulated for each final state is listed in Table 6.1. The dark photon and dark Upsilon mass hypotheses we generated are shown in Fig. 6.5. The range of dark Upsilon mass is [0.05, 0.15, 0.3, 0.5, 1, 1.5, 2, 2.25, 3, 4, 4.5, 5, 6, 7, 7.5, 8, 9, 10] GeV. The range of dark photon mass is [0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.47, 0.5, 0.71, 0.75, 1, 1.42, 1.5, 2, 2.38, 2.5, 2.85, 3] GeV. The generated events are passed through the detector response simulation based on GEANT4 [19] and reconstructed using the same software chain as the experimental data. The dark photons are reconstructed using the prompt reco sequence.

Signal MC events with non-zero lifetimes are also generated using MadGraph and reconstructed using the same software chain as the experimental data using prompt reco sequence. The dark photon flight lengths we simulated are [0.1, 1, 10] mm. The range of dark Upsilon mass is [0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10] GeV. The range of dark photon mass is [0.02, 0.05, 0.1 , 0.2, 0.5] GeV. This signal MC will be used to evaluate the impact of dark photon lifetime on the signal efficiencies.

The background is complex due to multiple lepton and pion pairs in the final states and cannot be simulated accurately by current generators. Instead, we use 5% of the data (Run3 data) as a background sample to optimize our selection criteria and validate the fitting procedure. We assume that the signal component is negligible in the validation data.

6.4 Candidate Reconstruction

We search for the reaction $e^+e^- \rightarrow \gamma\Upsilon_D, \Upsilon_D \rightarrow A'A'A'$, where the dark photon subsequently decays into a lepton or pion pair. We conduct our analysis by first identifying events containing exactly six charged tracks, and we form all possible combinations of three dark photon candidates. The electrons, muons and pion pairs are selected from loose PID algorithms (a track may satisfy several PID

Category	Final state	Number of events generated
C_0	$3e^+e^-$ $2e^+e^-1\mu^+\mu^-$ $1e^+e^-2\mu^+\mu^-$ $3\mu^+\mu^-$	2040083
C_1	$2e^+e^-1\pi^+\pi^-$ $2\mu^+\mu^-1\pi^+\pi^-$ $1e^+e^-1\mu^+\mu^-1\pi^+\pi^-$	1103956
C_2	$1e^+e^-2\pi^+\pi^-$ $1\mu^+\mu^-2\pi^+\pi^-$	1103956

Table 6.1: Number of signal Monte Carlo events generated for each category of final states.

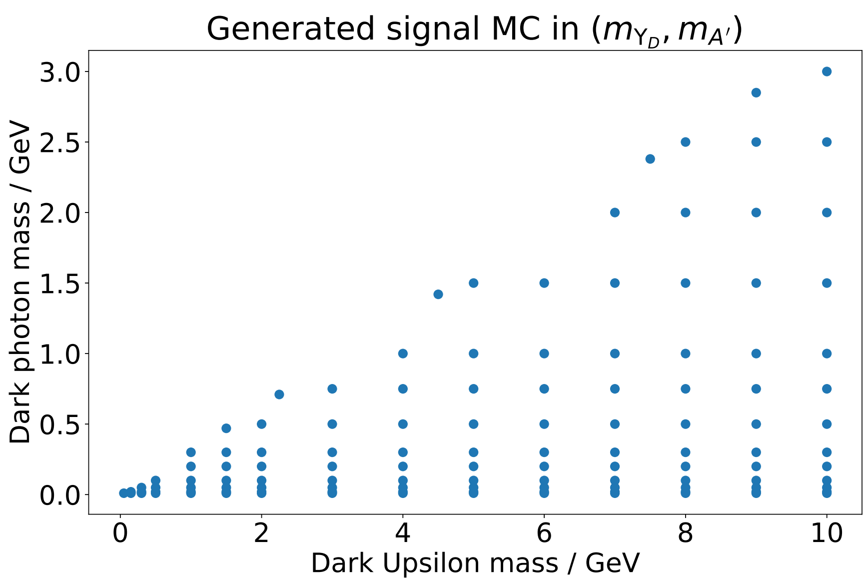

Figure 6.5: Monte Carlo-generated Υ_D and dark photon masses.

algorithms and be used multiple times with different mass hypotheses). To limit the combinatoric background, we require the presence of at least two tracks loosely identified as either electrons or muons, and the maximum mass difference between the three dark photon candidates must be less than 0.5 GeV. Tighter PID and criteria are applied at a later stage.

We then perform a kinematic fit with a beam constraint to each reconstructed Υ_D candidate. No energy constraint is applied, to allow for the possibility of initial state radiation or Bremsstrahlung from electrons. The ISR photon kinematics can then be inferred, based on the dark photon and Υ_D kinematics. For signal events, the recoil mass should be compatible with the photon hypothesis ($\sim$ 0 GeV), while background events produce a wide distribution. The ISR photon can be observed or not, and this analysis treats both cases simultaneously. If the extrapolated ISR photon is emitted in the EMC acceptance, a corresponding photon must be found. Otherwise, only a small amount of extra neutral energy should be seen.

Same-sign A' candidates are also reconstructed. For our analysis with 6 charged tracks grouped into 3 neutral dark photon pairs, the same-sign A' candidates are obtained by swapping two particles of the same charge with opposite charge in different pairs. The mass difference between the same-sign candidates tends to be smaller for background events than signal events. The background is mostly combinatorics and the same-sign mass difference is distributed more broadly than signal events, i.e, the probability of having a small same-sign mass difference is larger for background than signal.

6.5 Event Selection

To further improve the signal purity, we use a multivariate classifier based on the following variables.

- χ^2: The probability of the constrained fit on the Υ_D candidate, imposing a beam constraint to the six tracks. Signal events should have small value, as they should be compatible with the beam spot and their tracks can form a common vertex. Background events are spread continuously.

- *PIDpass*: The particle type (e, μ, π) assigned to each track must be compatible with the PID requirements described in Table 6.2. We explored several PID requirements by changing the selectors and the required number of charged tracks, and choose the ones optimizing the signal significance.

- *massdiff*: Maximum mass difference between the three dark photon candidates. Given the reconstructed masses of the three dark photons $m^1_{A'}, m^2_{A'}, m^3_{A'}$, we compute

  $$massdiff = max(|m^1_{A'} - m^2_{A'}|, |m^2_{A'} - m^3_{A'}|, |m^1_{A'} - m^3_{A'}|).$$

  Signal events should be peaked around 0, while background events are spread continuously.

- *ISR_costh*: The cosine of the polar angle of the reconstructed ISR photon in the laboratory frame.

- *bestISRPhoton*: A categorical feature indicating whether the emitted ISR photon is found in the calorimeter. Its value can be $\{-2, -1, 0\}$. Its value is -2 when the reconstructed ISR photon polar angle is outside the detector acceptance. Its value is -1 when reconstructed ISR photon polar angle is within the detector acceptance, but not found in the calorimeter, i.e, there is a missing energy in this event. Its value is 0 when the reconstructed ISR photon polar angle is within the detector acceptance and detected by calorimeter. The ISR photon is considered to be detected if a neutral cluster is found in the calorimeter with an energy within 10% of that reconstructed from the Υ_D measurement, and the cosine of the angle between the neutral cluster and reconstructed ISR photon directions smaller than 0.1.

- *ISR_mass2*: Invariant mass of reconstructed ISR photon $m^2_{ISR} = (p_{e^+e^-} - p_{\Upsilon_D})^2$. Signal candidates should be compatible with the photon hypothesis, peaking around 0, while background candidates span a large range of values.

- *E_extra*: Sum of neutral clusters in the ECAL, excluding the ISR photon candidate and Bremsstrahlung photons. Signal events are expected to have small extra neutral energy, while the background events can have larger values.

- *A_helicity*: The average helicity angle of three dark photons. The helicity angle is defined as the angle between the Υ_D flight in CM frame and A' flight in Υ_D rest frame, as illustrated in Fig. 6.6.

- *A_angle*: The average of angles between the dark photons in the Υ_D rest frame, as illustrated in Fig. 6.6.

- *A_dihedral*: The average dihedral angles between the dark photons. For two dark photons $A'_1 \to X_1^+ X_1^-$ and $A'_2 \to X_2^+ X_2^-$, the dihedral angle between A'_1 and A'_2 is the angle between the planes defined by $X_1^+ X_1^-$ and $X_2^+ X_2^-$, as illustrated in Fig. 6.6.

- *A_poslepton_helicity*: The average helicity angle of the positive lepton for each dark photon decay, as illustrated in Fig. 6.6.

- *A_FlightLen_avg*: The average flight length of the "zero-lifetime" dark photons.

- *same_sign_massdiff*: For each candidate with three pairs of opposite charged particles, we construct its same sign candidates by swapping two same particles with opposite charge in different pairs. If there is only one type of particle in final states, i.e, in the form of $X_1^+X_1^-, X_2^+X_2^-, X_3^+X_3^-$ where X indicates symbol of particles (e, μ, π) and integer indicates which dark photon it decays from, we have $3 \times C_3^2 = 9$ possible same sign candidates:

$$\begin{aligned}
&X_1^+X_2^+, X_1^-X_2^-, X_3^+X_3^- \\
&X_1^+X_2^+, X_1^-X_3^-, X_3^+X_2^- \\
&X_1^+X_2^+, X_3^-X_2^-, X_3^+X_1^- \\
&X_1^+X_3^+, X_2^+X_1^-, X_2^-X_3^- \\
&X_1^+X_3^+, X_2^+X_2^-, X_1^-X_3^- \\
&X_1^+X_3^+, X_2^+X_3^-, X_1^-X_2^- \\
&X_2^+X_3^+, X_1^+X_1^-, X_2^-X_3^- \\
&X_2^+X_3^+, X_1^+X_2^-, X_1^-X_3^- \\
&X_2^+X_3^+, X_1^+X_3^-, X_1^-X_2^- .
\end{aligned} \tag{6.8}$$

If there are two types of particles in final states, i.e, $X_1^+X_1^-, X_2^+X_2^-, Y_3^+Y_3^-$ where X and Y are different particles from e, μ, π, we have only one possible candidate:

$$X_1^+X_2^+, X_1^-X_2^-, Y_3^+Y_3^- .$$

If there are three types of particles in final states, namely $e^+e^-, \mu^+\mu^-, \pi^+\pi^-$, we will consider pions as muons and use the previous rule.
The same sign mass candidate for tracks X_1 and X_2 is calculated as

$$\begin{aligned}
p_{A'} &= p_{X_1} + p_{X_2} \\
m_{A'} &= E_{A'}^2 - \mathbf{p}_{A'}^2 .
\end{aligned} \tag{6.9}$$

We choose the minimum same-sign mass difference among all possible combinations for a candidate. The mass difference between the same-sign candidates tends to be smaller for background events than signal events. The background is mostly combinatorics and the same-sign mass difference is distributed more broadly than signal events, i.e, the probability of having a small same-sign mass difference is larger for background than signal.

The distributions of these features for all three categories are shown in Fig. 6.7 and 6.8, in which each entry comes from one candidate. Correlations among these features are shown in Fig. 6.9. The background candidates are taken from Run3 data,

Final state	PID Selection Criteria
$3e^+e^-$	at least 4 tracks pass tight electron PID and 5 track passes loose electron PID
$2e^+e^-1\mu^+\mu^-$	at least 3 tracks pass loose electron PID and 1 track passes loose muon PID
$1e^+e^-2\mu^+\mu^-$	at least 1 track passes loose electron PID and 3 tracks pass loose muon PID
$3\mu^+\mu^-$	at least 3 tracks pass tight muon PID and 5 tracks pass loose muon PID
$2e^+e^-1\pi^+\pi^-$	at least 3 tracks pass loose electron PID
$1e^+e^-1\mu^+\mu^-1\pi^+\pi^-$	at least 1 track passes tight electron PID and 1 track passes loose muon PID
$2\mu^+\mu^-1\pi^+\pi^-$	at least 3 tracks pass loose muon PID and 2 tracks pass tight muon PID
$1e^+e^-2\pi^+\pi^-$	at least 1 track passes tight electron PID
$1\mu^+\mu^-2\pi^+\pi^-$	at least 1 track passes loose muon PID

Table 6.2: PID selection rules of different final states.

while the signal candidates are selected from a mixture of MC samples generated at all different masses.

Machine learning classifier

Using the features described above, our next step is to choose and train machine learning models for signal/background classification. The dataset we use is a mixture of Run3 data as negative (background) sample and signal MC as positive (signal) sample. We separate the dataset into two disjoint subsets acting as a training dataset and a validation dataset respectively. Models are first trained using the training dataset, while hyper-parameters are selected based on the models' performance on validation dataset to reduce over-training. We tried both Random Forest (RF) and Gradient Boosting Decision Tree (GBDT) models, and we found that RF has better performance, so we use RF in this analysis. The classification probability of RF models is discrete by construction. For example, a RF classifier containing 100 trees will have score values separated by increments of 0.01, which is insufficient for our purpose. In order to obtain a continuous classification probability and determine the hypothesis testing rejection region, we perform a logistic regression on the RF output, as shown in Fig. 6.10. Specifically, we train the RF with our training dataset first. The outcome of each decision tree in the RF indicates which leaf node our event ends. This information is represented by a binary vector $\mathbf{d} = [0, 0, 0, ..., 1, ..., 0]$,

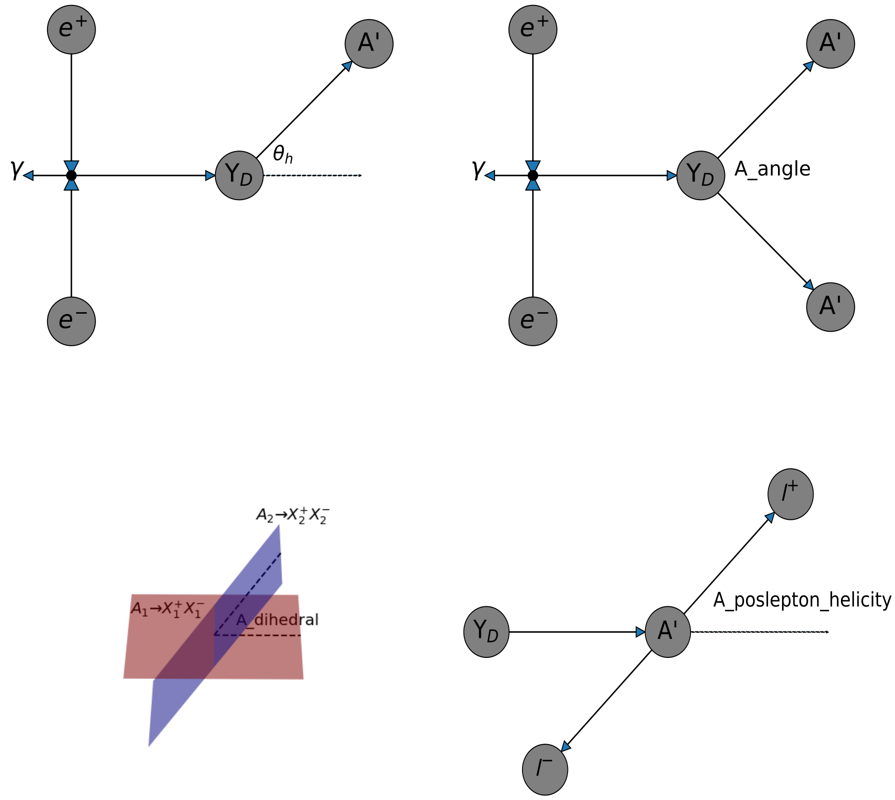

Figure 6.6: Illustration of angle-related physical variables used in the MVA. *A_helicity* (upper left), *A_angle* (upper right), *A_dihedral* (bottom left), and *A_poslepton_helicity* (bottom right).

indicating that the event ended in the i^{th} leaf node. Given a RF with n decision trees, the binary representation of the full RF is a stack of decision trees' representations $[\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_n]$. To obtain a continuous score function, we fit a Logistic Regression function on top of this binary representation of training samples as our classification model. We denote this model as *stacked RF*. All of models are trained and validated using the scikit-learn package.

We trained stacked RF under a different number of assembled decision trees (*n_estimators*). Other hyper-parameters such as the *criterion* (the function to measure the quality of a split) and the *min_samples_split* (minimum number of samples required to split an internal node), are kept as default values, as they are believed to have been well-tuned. Under the figure of merit defined below, we find that RF with 60 trees offers stable and close to optimal performance, and we select this

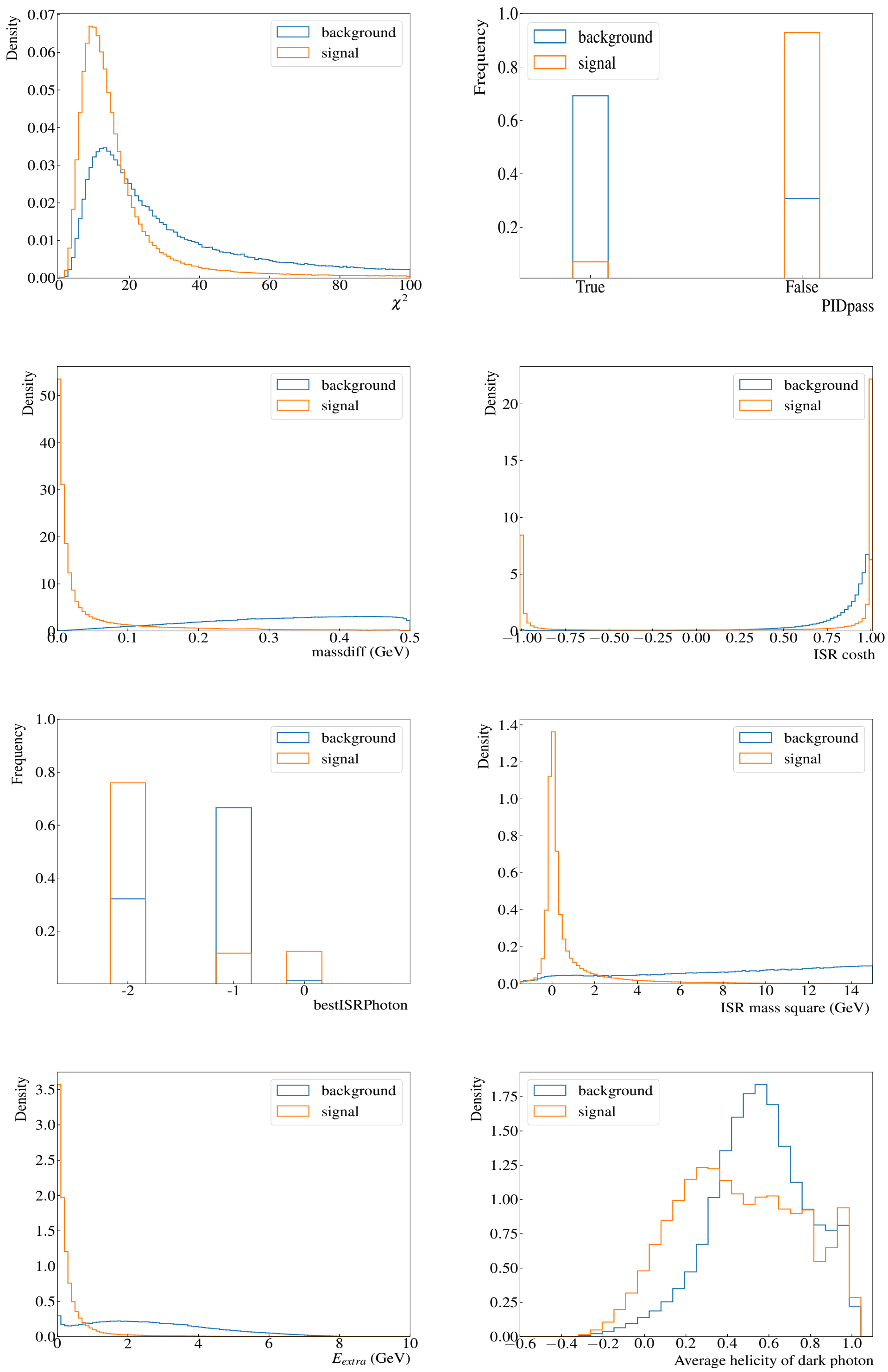

Figure 6.7: Signal and background density distributions of input variables.

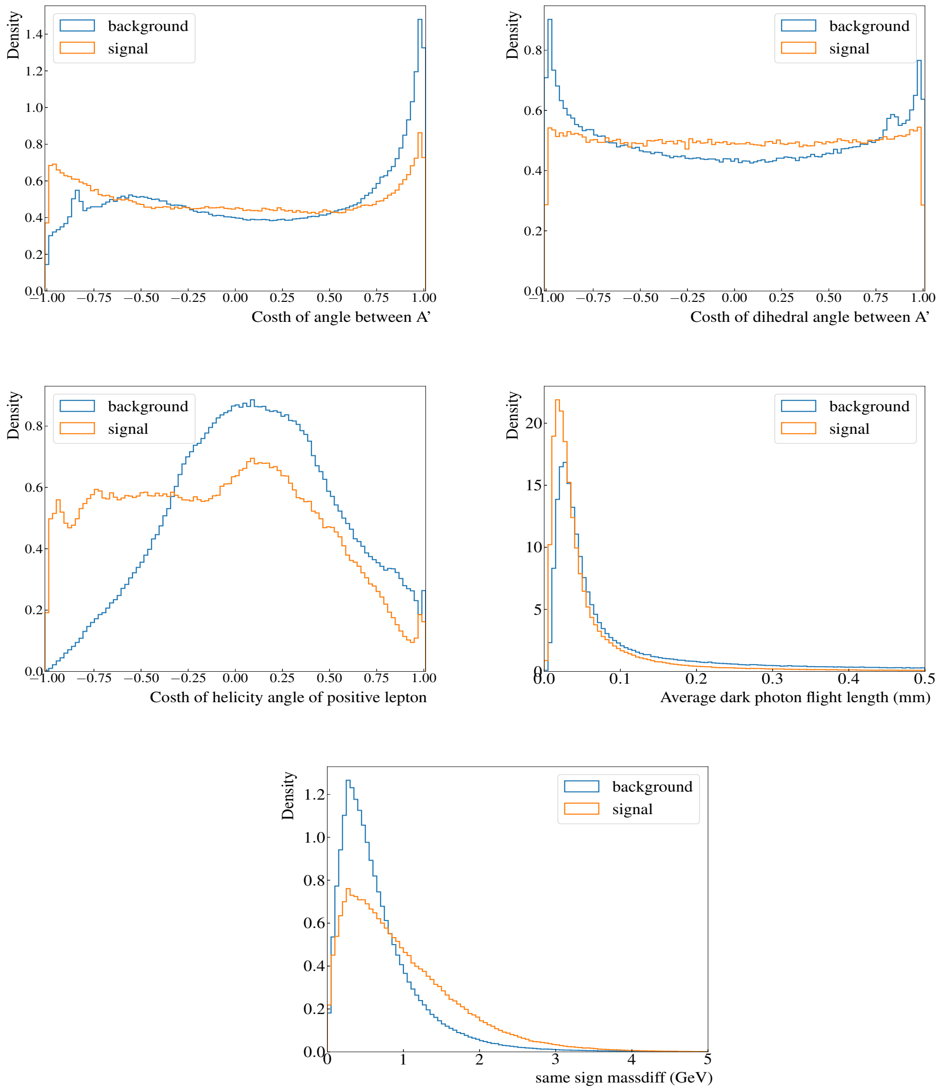

Figure 6.8: Signal and background density distributions of input variables.

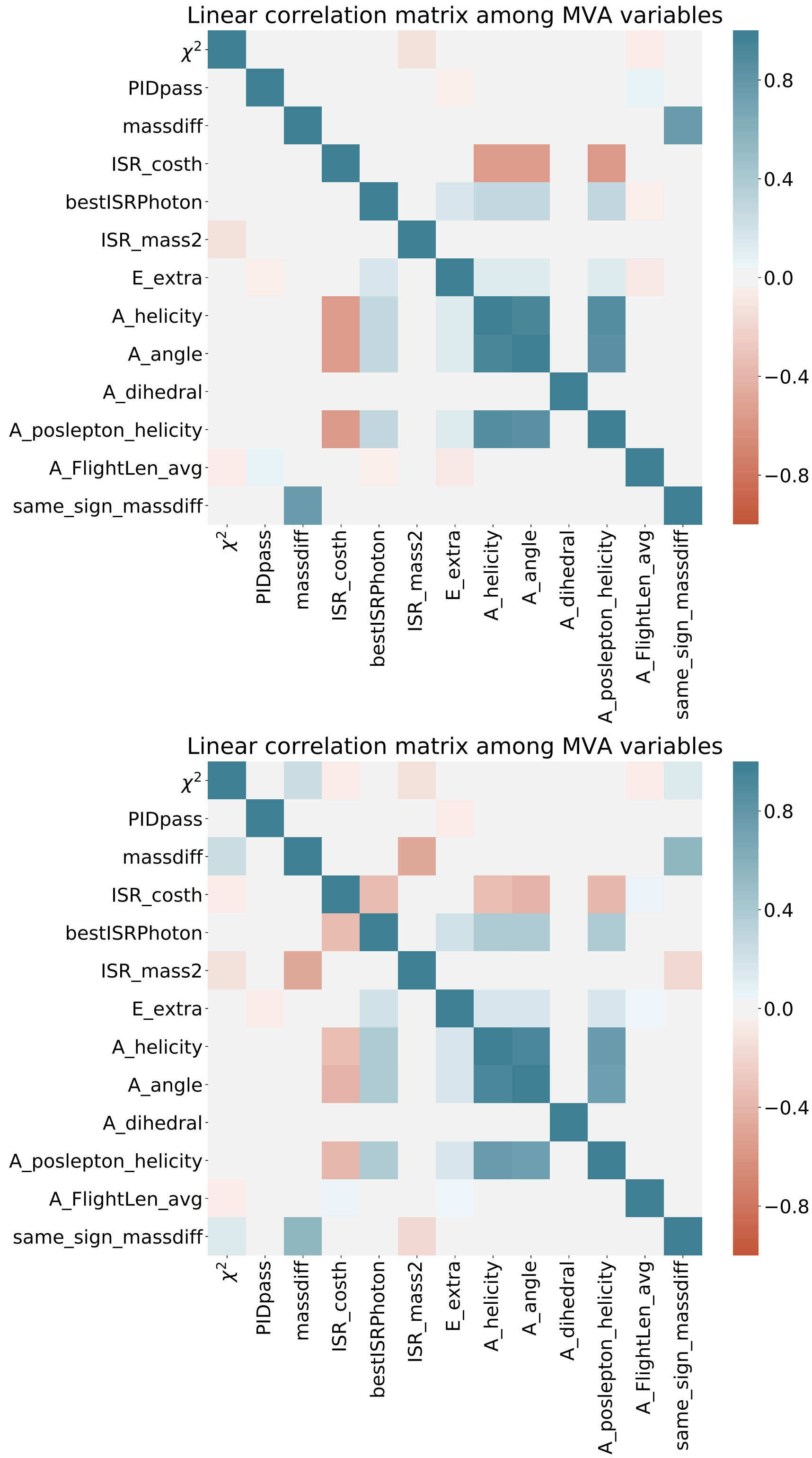

Figure 6.9: Linear correlation matrix of variables used for MVA training for signal MC (upper) and Run3 data (bottom).

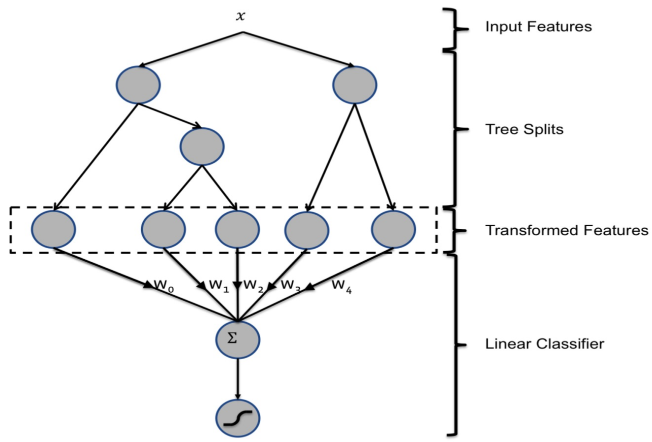

Figure 6.10: Illustration of the stacked Random Forest model. Figure taken from [20].

configuration.

Selection criteria optimization

The distribution of the classifier outputs is shown in Fig. 6.11 and Fig. 6.12 for both prompt and displaced dark photon decays. The optimal criteria on the MVA output (*ml_score*) is determined by optimizing the average cross-section upper limit over the whole mass range, assuming every observed event is a signal candidate. This approach is clearly conservative, but as it can be seen in Fig. 6.11 and 6.12, the background rises rapidly in the vicinity of the optimal cut, while the signal efficiency varies smoothly. Minimizing the background level (ideally completely suppressing it) therefore has a small effect on the signal sensitivity, but greatly reduces the associated systematic uncertainties and facilitates the determination of the signal significance as well.

More specifically, we optimize the classifier by minimizing the figure of merit $\frac{\hat{\lambda}(n)}{\epsilon}$, where $\hat{\lambda}(n)$ is the 90% CL upper limit on the number of signal events given n observed events in the signal region (assumed to be signal), and ϵ is the signal efficiency averaged over the phase space.

We determined n and ϵ as a function of the classifier cut by fitting the classifier output distributions, and integrated the density over the cut value. We apply a binned fit to the *ml_score* histogram of the data events in the range of $0 \leq ml\_score \leq 8$, by fitting a function of the form $f(x) = e^{a \cdot x^2 + b \cdot x + c}$. The fit is performed for each category of C_0, C_1 and C_2 separately. The fitting results are shown in Fig. 6.13. The dashed vertical lines indicates the optimal cuts. The range of $0 \leq ml\_score \leq 8$ ensures that the fit is completely dominated by background events and still offers a good estimation of the data. Then, for a given cut c, we estimate the expected number of observed events by integrating the fitting density over the cuts: $n = N \cdot \int_c^{+\infty} f(x)dx$, where N is the total number of data events in the corresponding category.

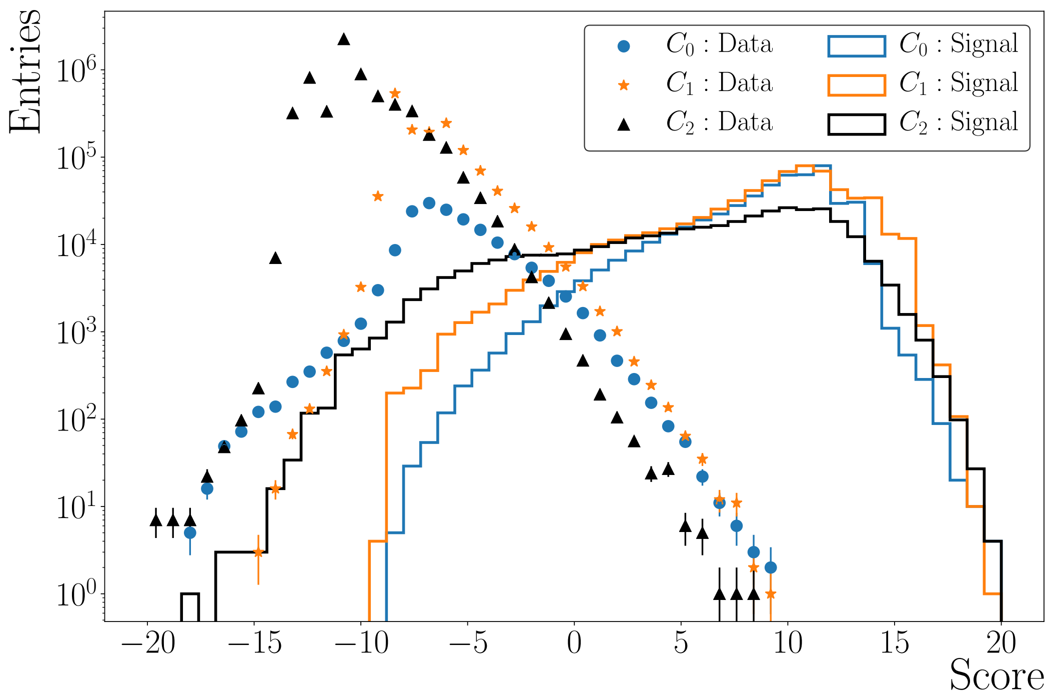

Figure 6.11: The distribution of the classifier scores for each event category for the data (markers) and signal Monte Carlo (solid lines) samples, for the prompt dark photon decays. The MC simulations are arbitrarily normalized.

Best candidate selection

The best candidate selection aims to assign each event with one single best matching candidate. If multiple candidates pass the MVA selection for one event, the best candidate is selected based on its final state, according to the following sequence of hypotheses:

$$6e, 4e2\mu, 2e4\mu, 6\mu, 4e2\pi, 2e2\mu2\pi, 4\mu2\pi, 2e4\pi, 2\mu4\pi.$$

The sequence of hypotheses is ordered according to the purity of the final states to minimize mis-identification among channels.

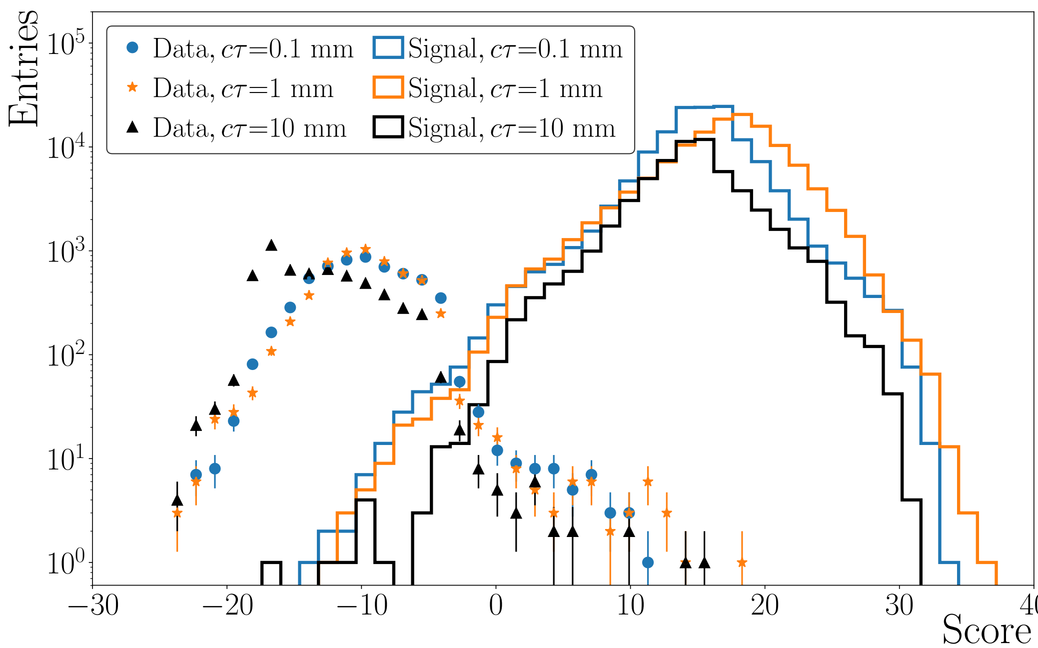

Figure 6.12: The distribution of the classifier scores for data (markers) and signal MC simulations (solid lines) for dark photon lifetimes corresponding to (top) $c\tau_{A'}$ = 0.1mm, (middle) $c\tau_{A'}$ = 1mm, and (bottom) $c\tau_{A'}$ = 10mm. The MC simulations are arbitrarily normalized.

Event selection results

For signal with prompt decays, a total of 69 events pass all the selection criteria. The corresponding $(m_{\Upsilon_D}, m_{A'})$ distribution is shown in Fig. 6.14. Most events correspond to $e^+e^- \to q\bar{q}$ process. The events near $m_{\Upsilon_D} \sim 0.1$ GeV and $m_{A'} \sim 0.05$ GeV arise from $e^+e^- \to \gamma\gamma\gamma$ events in which all three photons convert to e^+e^- pairs.

For signal with displaced decays, a total of 56, 33, and 31 events are selected for the $c\tau_{A'} = 0.1$, 1, and 10 mm data sample, respectively. The resulting mass distributions are shown in Fig. 6.15.

6.6 Signal Modeling and Efficiency

Mass resolution

The A' and Υ_D mass resolutions are used for both determining the size of a given scan box, as well as calculating the scan step size. The mass resolutions are estimated by fitting the A' (Υ_D) mass distributions with Crystal Ball functions (CBF) for each MC sample. One example of the fits is shown in Fig. 6.16, in which we fit the Υ_D and dark photon mass spectrum at m_{Υ_D} = 8 GeV, $m_{A'}$ = 0.5 GeV. The CBF captures the energy loss in low tails and gives a better estimation of the $(\Delta_{m_\Upsilon}, \Delta_{m_{A'}})$ resolution

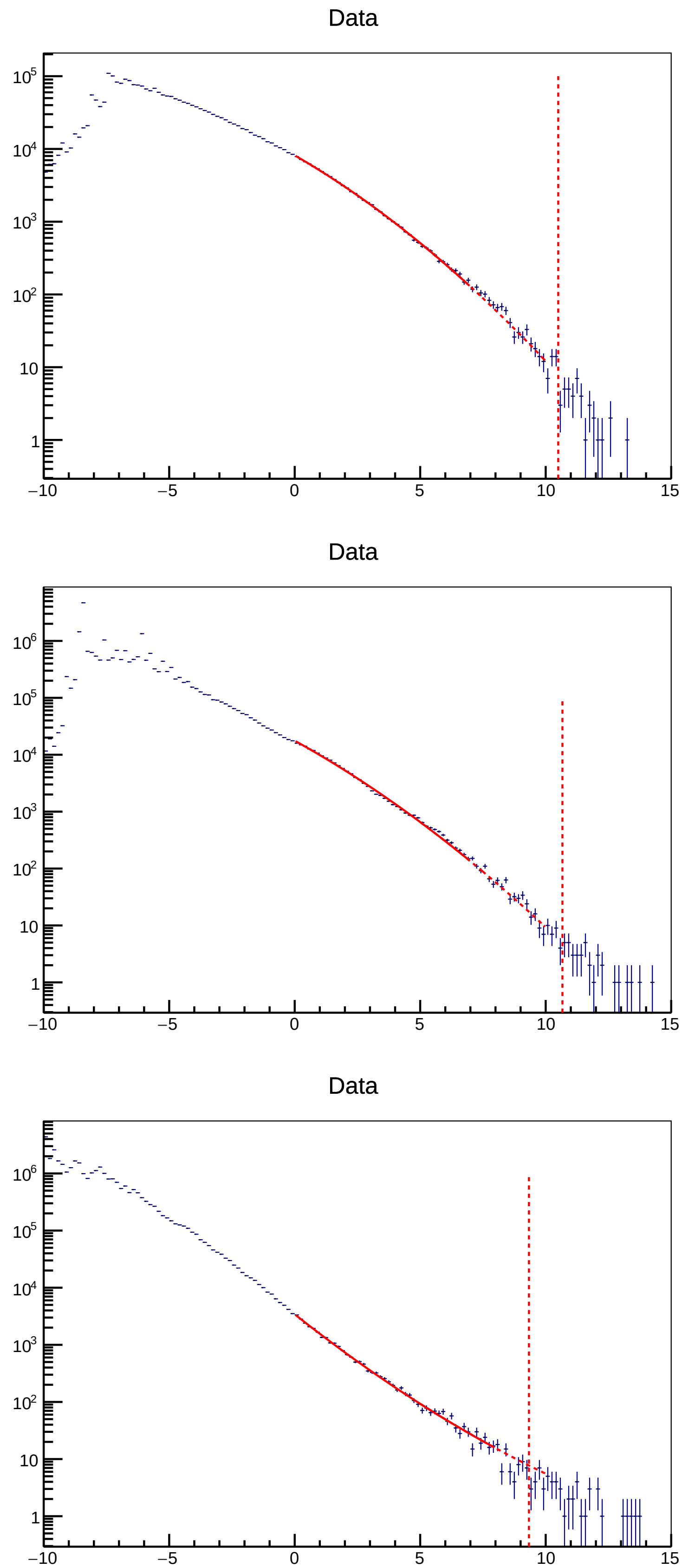

Figure 6.13: Fits to the classification score distributions of data with a function of the form $p(x) = e^{a \cdot x^2 + b \cdot x + c}$, for C_0 (upper), C_1 (middle), and C_2 (bottom) category of events.

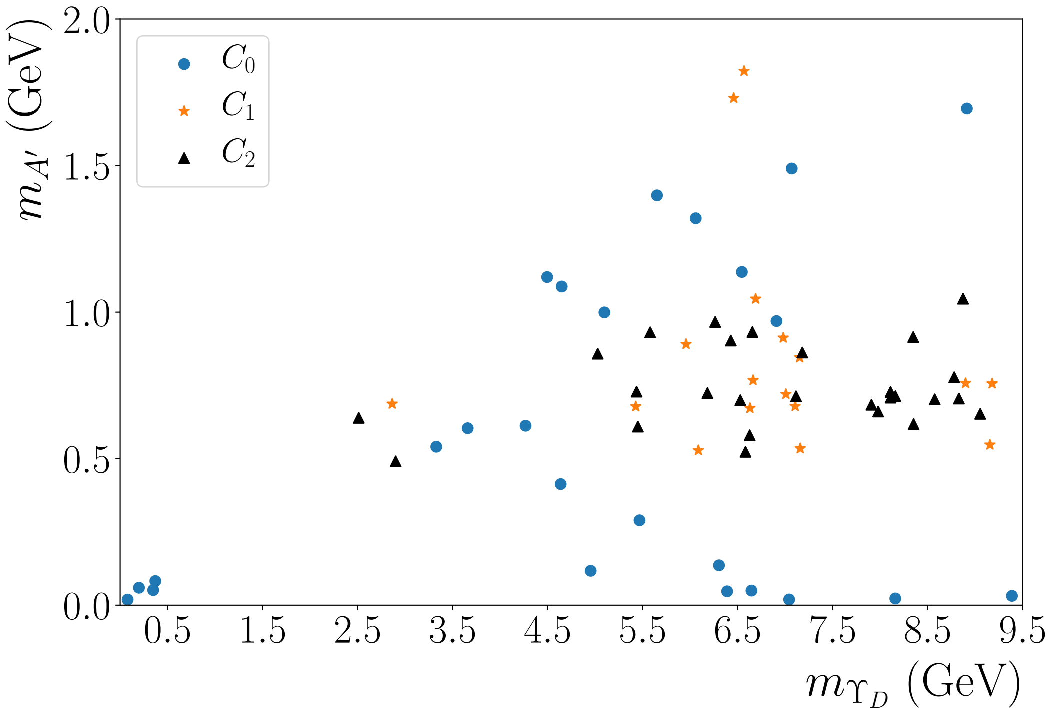

Figure 6.14: The $(m_{\Upsilon_D}, m_{A'})$ distribution for events passing all selection criteria for prompt dark photon decays.

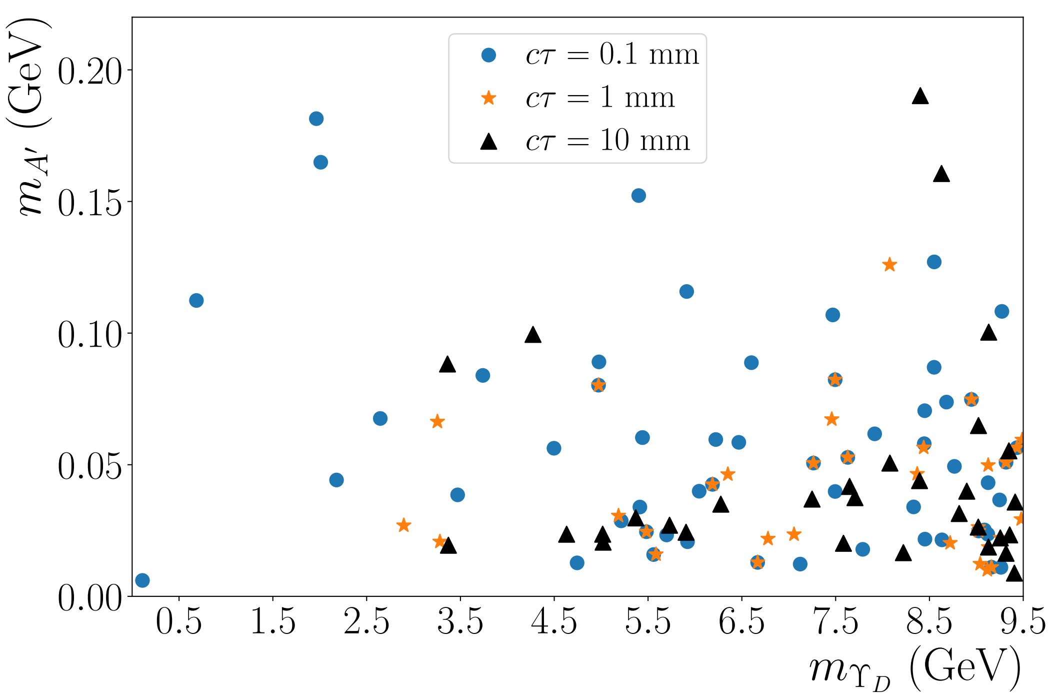

Figure 6.15: The $(m_{\Upsilon_D}, m_{A'})$ mass distribution of event candidates passing all selection criteria for the datasets optimized for each dark photon lifetime.

than a simple Gaussian fit. While the fits are not perfect, they are sufficient for the purpose of estimating the mass resolution.

To evaluate the mass resolution as a function of $(m_{\Upsilon_D}, m_{A'})$, we interpolate the results at known points using a 2-dimensional smooth spline. Smooth spline function aims to fit a polynomial function on data and guarantees the smoothness of the function at the same time. To minimize the fitting uncertainties, we fit each category of events individually. The results are shown in Fig. 6.17. The Υ_D mass resolutions are at the level of 20 MeV. The dark photon mass resolutions are at the level of 3 MeV.

Signal efficiency

The signal efficiency is evaluated by counting the fraction of simulated events passing the selection criteria in the signal window at a given point, normalized to the total number of generated events. The size of the signal window is given by the mass resolutions at that point, namely $4\Delta_{m_\Upsilon}$ and $4\Delta_{m_{A'}}$.

To evaluate the signal efficiency as a function of $(m_{\Upsilon_D}, m_{A'})$, we also use a 2-dimensional smooth spline technique for interpolation, fitting each category separately. The degrees of the polynomial functions used for interpolation are increased until the model is stable. The results of the signal acceptance, selection efficiency, and signal efficiency are shown in Fig. 6.18.

The signal acceptance is low when m_{Υ_D} and $m_{A'}$ are low, which stops us from improving the signal efficiency in the low mass region. The MVA prefers to select events in the range of 6 GeV $\leq m_{\Upsilon_D} \leq$ 9 GeV, thus the selection efficiencies are higher in this region.

For each signal window, the signal efficiencies from C_0, C_1, and C_2 are weighted by branching fractions to obtain the total signal efficiency of the signal window. The total signal efficiency is then used for signal extraction. The result is shown in Fig. 6.19. The two horizontal bands around $m_{A'} = 0.78$ GeV and 1.0 GeV correspond to the ω and ϕ resonances. The drop in the branching fraction $A' \to X^+X^-, X = e, \mu, \pi$ at $m_{A'} = 0.78$ GeV and 1.0 GeV is due to the A' decaying predominantly into $\pi^+\pi^-\pi^0$ and K^+K^-, respectively.

Lifetime dependency

The signal efficiencies depend on the dark photon lifetime. When the dark photons have non-zero lifetime, the χ^2 fit of assuming prompt decay becomes larger and the mass resolution worse. The larger the dark photon lifetime, the smaller the signal

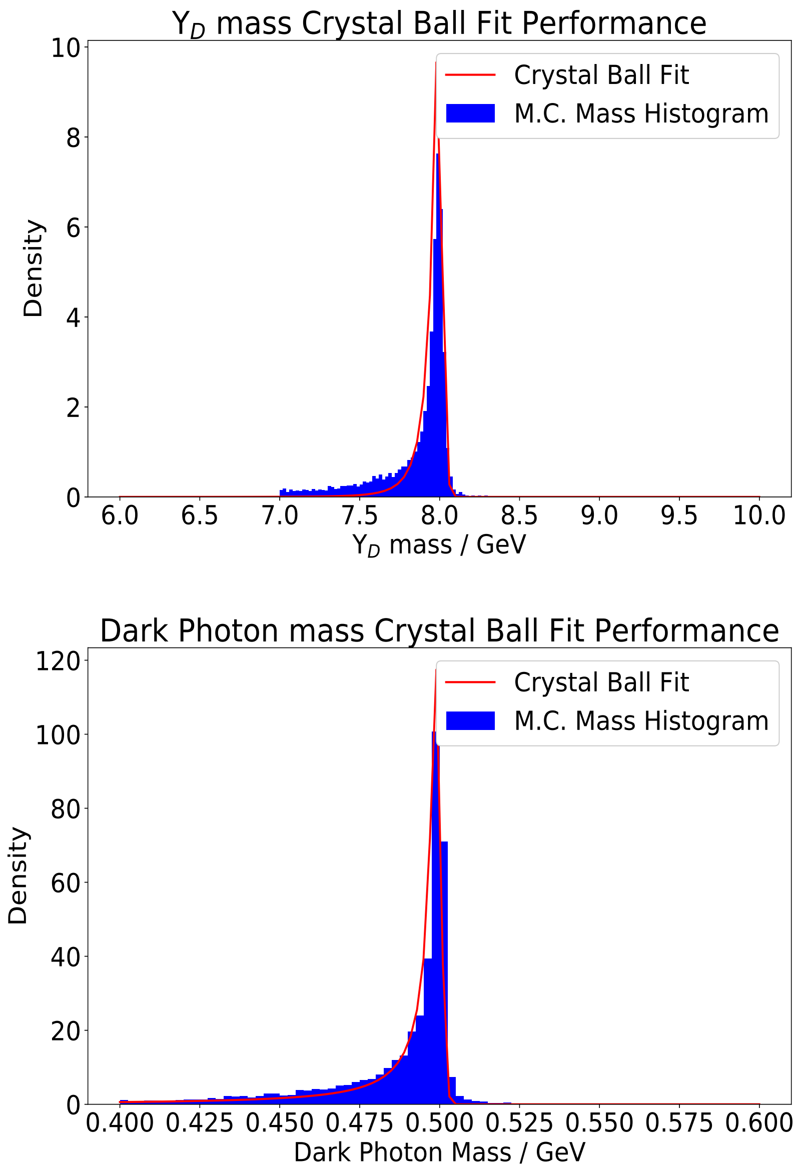

Figure 6.16: Example of CBF fits to the Υ_D and A' mass spectrum (m_{Υ_D} = 8.0 GeV and $m_{A'}$ = 0.5 GeV).

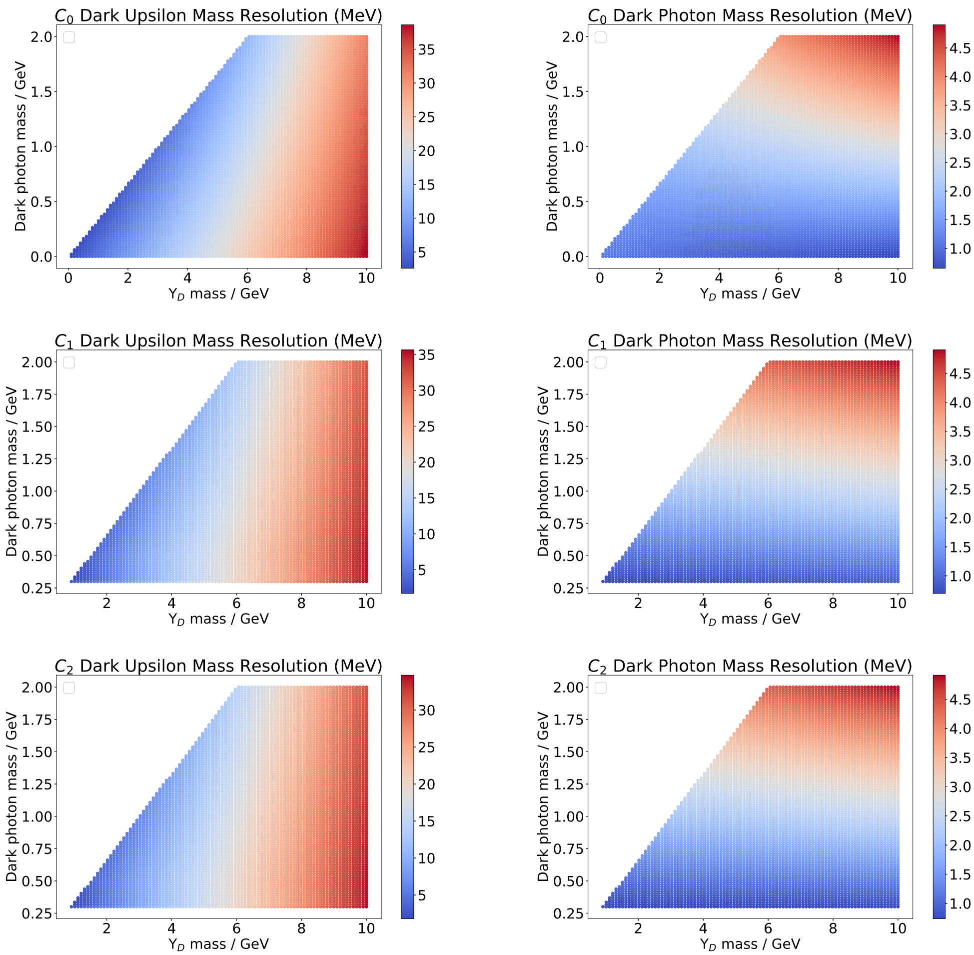

Figure 6.17: (Left) The Υ_D and (right) the dark photon mass resolution as a function of m_{Υ_D} and $m_{A'}$ for each category of events.

efficiency. We generate and reconstruct the signal MC with non-zero lifetime to estimate and interpolate the impact of lifetime on the signal efficiencies. We use the functional form $\log(\epsilon) = a \cdot \log(l) + b$ to fit the relation between the average signal efficiencies (ϵ) and the dark photon lifetime (l). We checked that the results obtained for each mass hypothesis are compatible with those obtained by considering the average efficiency, but the later are more robust and were chosen to describe the global dependence on the lifetime. For a given mass hypothesis $(m_{\Upsilon_D}, m_{A'})$ with prompt signal efficiency ϵ_0, the relation between the signal efficiency and the dark photon lifetime under the mass hypothesis is reasonably described by:

$$\log(\epsilon) = a \cdot (\log(l) - \log(l_0)) + \log(\epsilon_0), \tag{6.10}$$

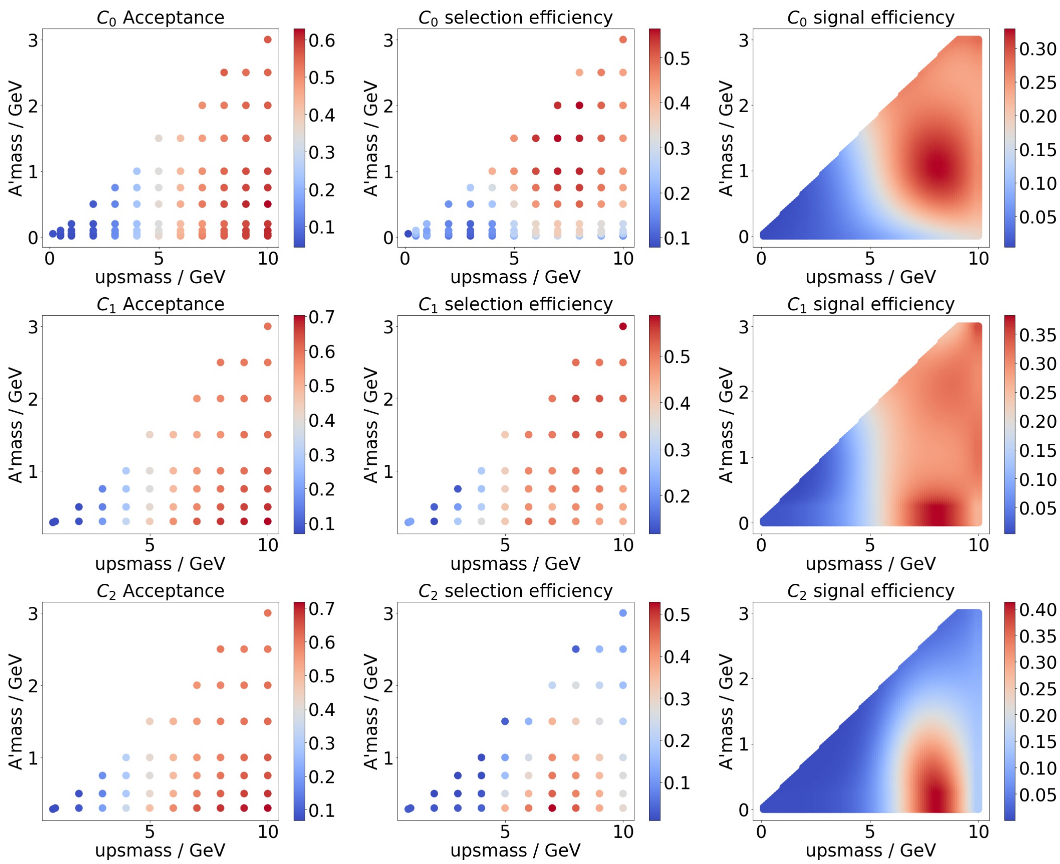

Figure 6.18: The acceptance, selection efficiency, and signal efficiency as a function of m_{Υ_D} and $m_{A'}$ for each category of events.

with $l_0 = 0.1$ mm, and ϵ_0 the efficiency for prompt decays. For a dark photon flight lengths above 100 μm, this interpolated relation will be used to correct the signal efficiency and establish the kinetic mixing strength upper limit. When the dark photon flight length is below 100 μm, we use the efficiency determined for prompt decays, as the displaced decay vertices of the dark photon have negligible effect on the signal efficiencies.

6.7 Systematic Uncertainties

The following source of systematic uncertainties are considered:

- Efficiency MC statistics ($\sigma_{\epsilon_{stat}}$): The uncertainties on the efficiency estimation due to the limited Monte Carlo simulation are taken as a systematic uncertainty. Our estimate of signal efficiency is $\hat{\epsilon} = \frac{\sum_{i=1}^{N} x_i}{N}$, where $x_i \sim Bernoulli(\epsilon)$ indicates whether the i^{th} event is selected as signal, and it follows a Bernoulli distribution with a true signal efficiency ϵ probability of taking 1. Therefore,

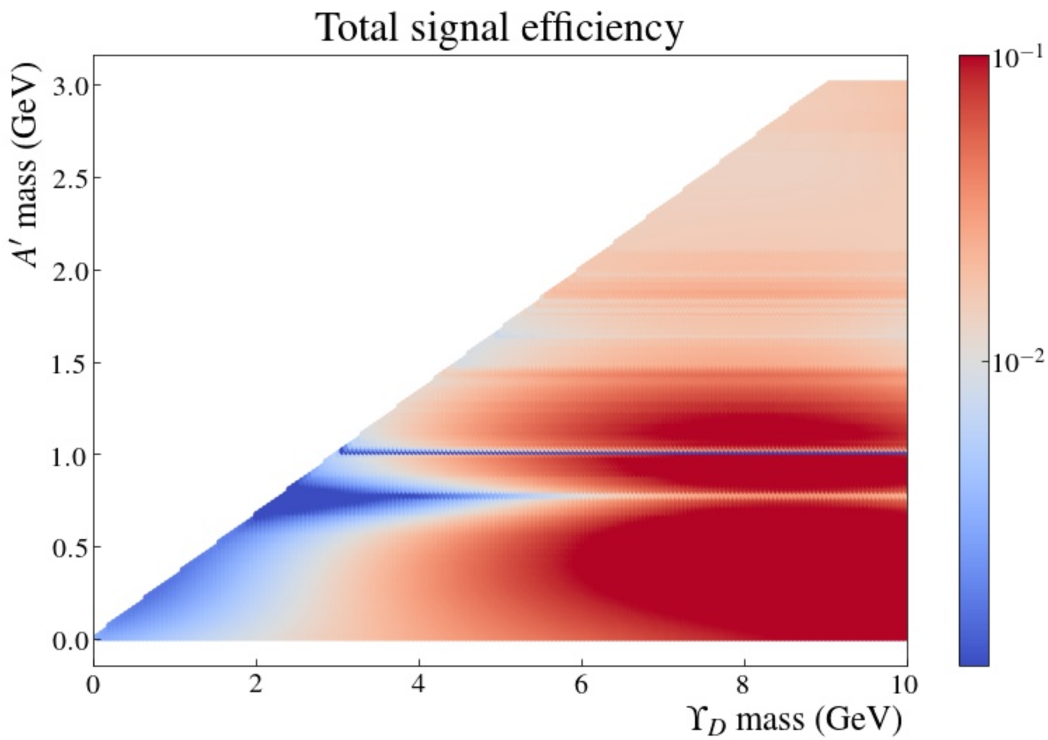

Figure 6.19: Signal efficiency (including the branching fractions) as a function of m_{Υ_D} and $m_{A'}$. The two horizontal bands around $m_{A'} = 0.78$ GeV and 1.0 GeV correspond to the ω and ϕ resonances.

the statistical uncertainty of the signal efficiency estimate is $\sigma_{\hat{\epsilon}} = \sqrt{\frac{\hat{\epsilon}(1-\hat{\epsilon})}{N}}$, where $\hat{\epsilon}$ is the estimated signal efficiency and N is the number of simulated events.

- Efficiency Interpolation ($\sigma_{\epsilon_{int}}$): This is the uncertainty due to the interpolation of the signal efficiency, as a function of m_{Υ_D} and $m_{A'}$. For each M.C. sample, we have both an estimated signal efficiency (ϵ_{est}) and a predicted signal efficiency (ϵ_{eval}). The former was obtained by counting the fraction of simulation events passing the classifiers and within our scanning windows, while the latter is the output of smooth spline functions obtained by removing this M.C. point and refitting the spline interpolation. The interpolation uncertainty at this point is defined as $|\epsilon_{est} - \epsilon_{eval}|/\epsilon_{est}$. To evaluate the interpolation uncertainty of any point, we fit a smooth spline function on the interpolation uncertainties for each category of events. The average interpolation uncertainty is within the scale of 8%, which indicates that the spline functions provide a good interpolation.

- PID: A systematic efficiency uncertainty of 2% per muon, 2% per pion, and 1% per electron is used. The PID uncertainty is added linearly for each final state, and we average the PID uncertainties within each category of events, assuming each final state within category occurs with equal probability. The final uncertainty is 9% for C_0, 10% for C_1 and 11% for C_2.

- Luminosity: The uncertainty on the luminosity for Run 1-6 is taken to be 0.5%, following the result from BAD 2186. A systematic of 1.2% is taken for the fraction of Run7 data missing B-counting information. The luminosity weighted uncertainty is 0.6%.

- Tracking efficiency: The systematic uncertainty is taken to be 0.2% per track, determined from various control sample, added linearly for the 6 tracks.

- Branching fraction: This fraction of uncertainties come from uncertainties of photon decay to $X^+X^-, X = e, \mu, \pi$. The uncertainty on the measurement of the ratio R [21] is propagated to the product of branching fraction. This uncertainty is mostly at the level of 1% for all category of events.

All systematic uncertainties listed above are summed in quadrature to obtain the total systematic uncertainty for each category. We also combine the systematic uncertainties from each of the categories together. The correlations arising from the uncertainties on the PID and branching fractions are taken into consideration in the combination process. The combined total systematic uncertainties are shown in Fig. 6.20. The average scale of the systematic uncertainties is $\sim$ 11%; the main contributor is the PID uncertainty.

There is a horizontal band around 0.7 GeV $\leq m_{A'} \leq$ 0.85 GeV that corresponds to the region where the $A' \to \pi^+\pi^-$ branching fraction is higher. This band results from the higher C_2 category of branching fractions in the region. Without considering the correlation among the three categories, the relation between combined relative uncertainty (σ^{rel}_{comb}) and each category's relative systematic uncertainty ($\sigma^{rel}_i, i = C_0, C_1, C_2$) is:

$$(\sigma^{relative}_{comb})^2 = \sum (\sigma^{relative}_i)^2 \cdot w_i$$

$$w_i = \frac{(BF_i \cdot \epsilon_i)^2}{(\sum\limits_i BF_i \cdot \epsilon_i)^2}.$$

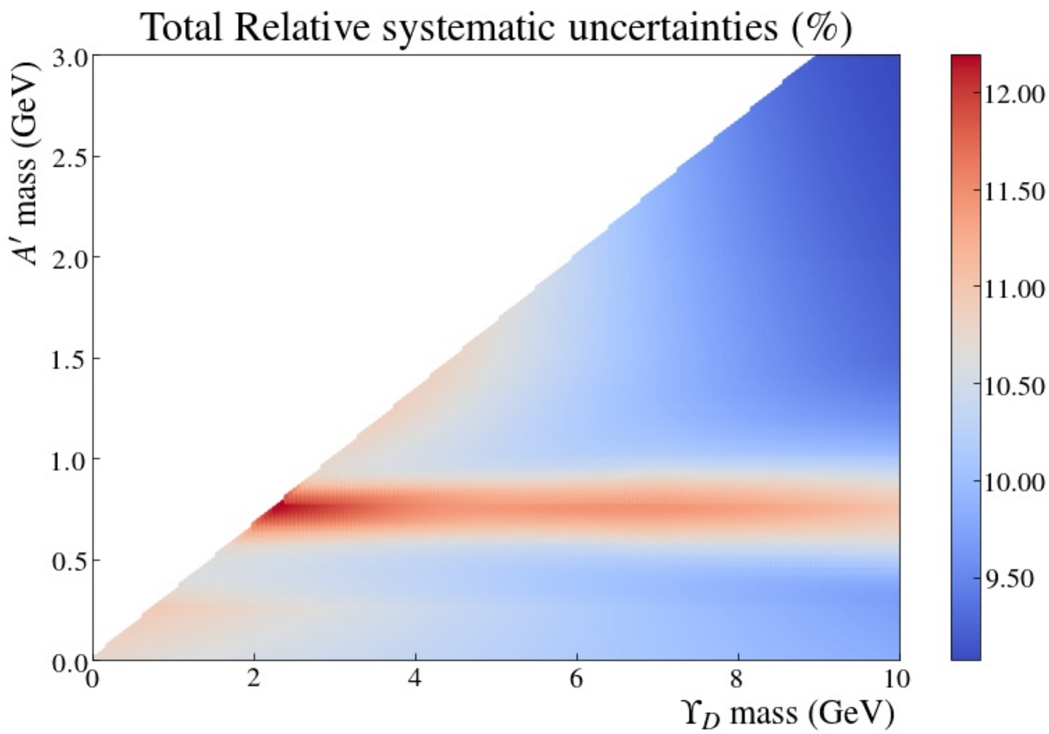

Figure 6.20: The total systematic uncertainties as a function of m_{Υ_D} and $m_{A'}$ for combined systematic uncertainties.

6.8 Signal Significance Estimation

This section discusses the procedure to estimate the significance of a potential signal. There are three steps to accomplish this:

1. Sideband Dataset: Construct a sideband dataset to estimate the background characteristics, since no events were selected in the optimization sample.

2. Distribution Structure: Use the sideband dataset to study the background distribution as a function of $(m_{\Upsilon_D}, m_{A'})$ and establish the background estimation procedure.

3. Signal significance: Determine the distribution of the number of observed signal events under the null hypothesis.

Sideband dataset

To obtain the distribution of background events in a signal region, we would need to study background events selected by our classifiers, which is infeasible for two reasons. First, there are no generators to simulate the relevant backgrounds accurately. Second, no events are selected in the optimization sample.

Instead, we construct a sideband dataset to solve this challenge. Given our classifier with an optimal selection criteria *optimal_cut* on the classifier score, our sideband region is defined as $[optimal\_cut - \delta, optimal\_cut]$, where δ is a parameter we set. If δ is too small, we won't have enough statistics, while the sideband dataset may be too biased if δ is too large. Our method is to select δ such that each sideband contains at least 500 events for each category of events, while making sure that δ is no smaller than the prediction uncertainty of each classifier, which is the standard deviation of *optimal_cut* estimated via bootstrapping. Run3 data events falling into the sideband regions constitute the *sideband datasets*. We claim that sideband datasets are a good approximation of the background distribution in the signal region. Fig. 6.21 illustrates the sideband region for the C_2 category of events.

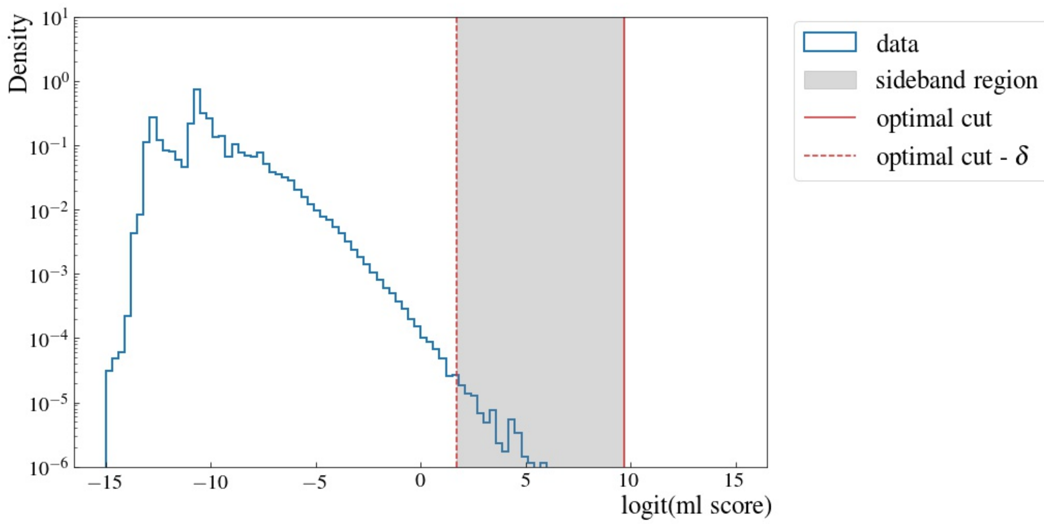

Figure 6.21: Distribution of the ml_score for the C_2 category for data. The sideband region is shown as a red rectangle. The optimal cut on the classifier score is shown as a solid line. The width of signal region is the δ we set to contain at least 500 events.

Background structure

To understand the distribution of background events in the signal region, we plot the distributions of sideband events on $m_{\Upsilon_D} - m_{A'}$ space, shown in Fig. 6.22. The dimuon and dipion thresholds are indicated by horizontal lines. For the C_0 and C_1 category of events, the distribution of sideband events are uniform. There is also a small band for photon conversions in the C_0 sample. For the C_2 category of events, the distribution of sideband events is also uniform, but there is a one-dimensional band when the dark photon mass is in the range of 0.7-0.8 GeV, corresponding to the ρ meson.

The distributions of sideband events provide a prescription on how to estimate the number of background events in the scanning windows. Under the assumption that the background distributions in the signal region are smooth, a nearest-neighbor method can be applied to estimate the number of background events. Specifically, the estimated number of background events B within a scanning window is obtained by calculating the average number of observed events of its nearest left and right scanning windows, as illustrated in Fig. 6.23. We do not take up and down windows due to the horizontal band structure in the background distribution. If the scanning window is at the left (right) boundary of $m_{\Upsilon_D} - m_{A'}$ space, we only use the nearest right (left) scanning window for estimation. When the scanning window is above dimuon threshold, the uniformity of the background guarantees the unbiasedness of the estimated number of background events. Using the left and right windows also works when we scan over the one-dimensional band produced by ρ or ω meson decays. When the scanning window is under the dimuon threshold, it is still a good estimation as long as background distribution is smooth along the Υ_D axis.

Signal significance

Generally, three factors are needed to estimate the signal significance: (1) Hypotheses: our null hypothesis (H_0) is that there is no darkonium signal. (2) Statistic: a test statistic that has discrimination between the null hypothesis (no signal) and alternative hypothesis (have signal). The statistic we use is the maximum number of observed signal events N_{sig}^{max} among all signal windows. For each signal window, the number of observed signal events equals the number of observed events minus the estimated background: $N_{sig} = N_{obs} - B$. (3) Distribution of statistic: we use a bootstrapping technique to estimate the distribution of N_{sig}^{max}. The distribution of N_{sig}^{max} will be used to calculate the signal significance (p-value) of the observed signal events in the final data.

There are two steps to obtain the distribution of N_{sig}^{max}. First, we estimate the expected number of background events based on the optimal expected number of observed events on data. The second step estimates the N_{sig}^{max} distribution with a bootstrap procedure (i.e, Toy MC). For each category C_i, we sample a number of events N_i from a Poisson distribution. We then bootstrap N_i events from the sideband dataset,

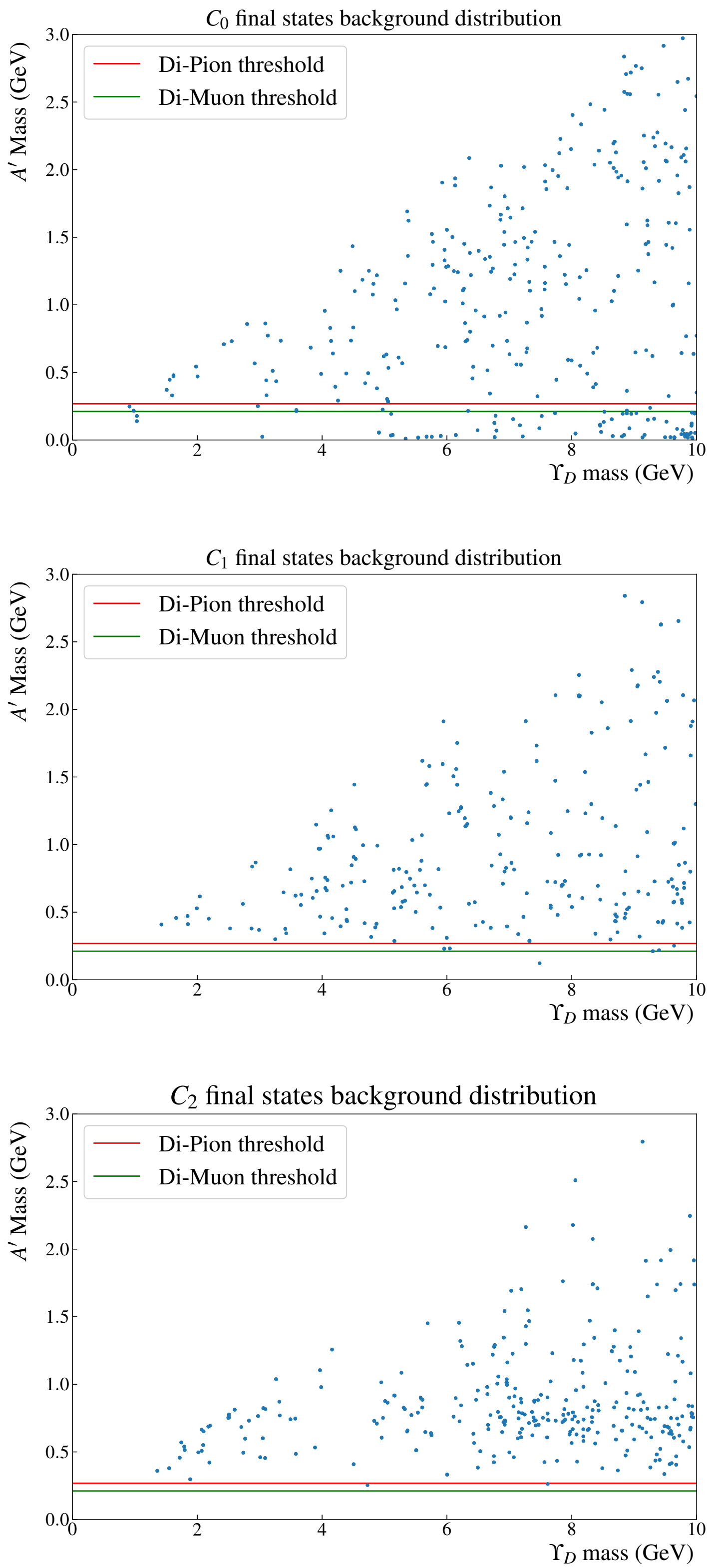

Figure 6.22: The distributions of sideband events for the (upper) C_0, (middle) C_1, and (bottom) C_2 categories of events.

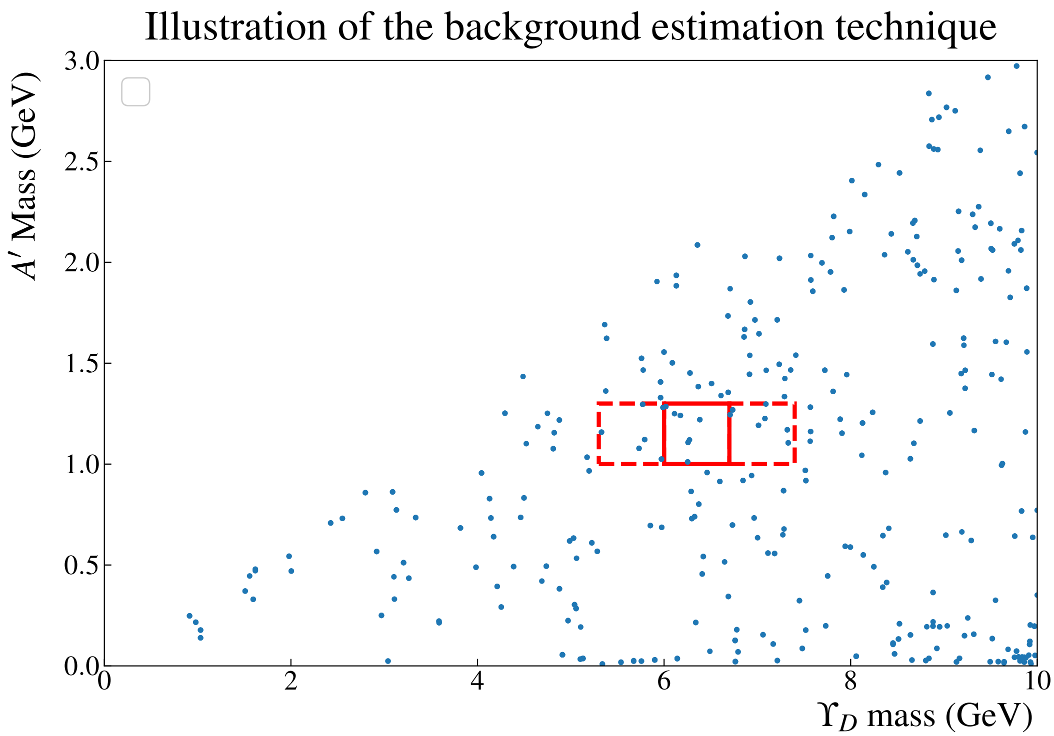

Figure 6.23: Illustration of the background estimate method for a given mass hypothesis in the $m_{\Upsilon_D} - m_{A'}$ space. The dataset displayed here is sideband data for C_0 category of events.

and run our scanning procedure on the $m_{\Upsilon_D} - m_{A'}$ space. Since all categories are summed together when establishing scanning windows, a single scan is necessary here. The maximum number of events among all signal windows gives one entry of the N_{sig}^{max} distribution. We run 10,000 bootstraps to obtain the distribution of N_{sig}^{max} shown in Fig. 6.24.

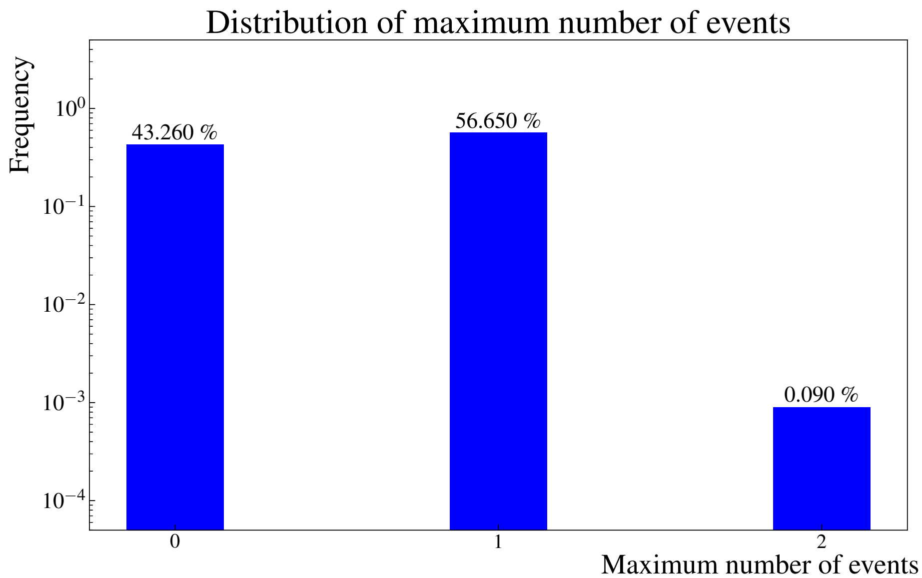

Figure 6.24: The distribution of the maximum number of signal events after scanning the parameter space $0 < m_{\Upsilon_D} < 10$ GeV, $0 < m_{A'} < 3$ GeV.

6.9 Signal Extraction and Upper Limit of Coupling

In this section, we first describe the procedure to extract the potential signal, then the results of upper limit of the coupling constant we obtain from data.

Signal extraction

We first combine the three categories (C_0, C_1, C_2) together before scanning for the signals: (1) The combined Υ_D and dark photon mass resolutions are taken as the maximum mass resolutions of three categories:

$$\sigma_{m_{\Upsilon_D}} = \max\{\sigma_{m_{\Upsilon_D}}(C_0), \sigma_{m_{\Upsilon_D}}(C_1), \sigma_{m_{\Upsilon_D}}(C_2)\}$$
$$\sigma_{m_{A'}} = \max\{\sigma_{m_{A'}}(C_0), \sigma_{m_{A'}}(C_1), \sigma_{m_{A'}}(C_2)\}.$$

(2) The signal efficiencies are combined as the average of a categories' signal efficiencies weighted by their branching fractions $\bar{\epsilon} = BF(C_0) \cdot \epsilon_{C_0} + BF(C_1) \cdot \epsilon_{C_1} + BF(C_2) \cdot \epsilon_{C_2}$. (3) The systematic uncertainties from three categories are combined.

The signal is extracted by scanning the Υ_D mass versus the dark photon mass plane. As mentioned above, the signal region is defined as a rectangular window $m_{\Upsilon_D} - 4\sigma_{m_{\Upsilon_D}} < m_{\Upsilon_D} < m_{\Upsilon_D} + 4\sigma_{m_{\Upsilon_D}}$ and $m_{A'} - 4\sigma_{m_{A'}} < m_{A'} < m_{A'} + 4\sigma_{m_{A'}}$ for a given Υ_D and photon mass ($m_{\Upsilon_D}, m_{A'}$). An estimate of the background is obtained from the nearest left and right signal windows. The signal significance is obtained from the distribution of the maximum number of signal events derived from the bootstrap procedure (Fig. 6.24). We do not see a clear signal; therefore upper limits on the signal cross section are extracted, as described below.

Suppose for a certain scanning window, we observe n_i events passing the selection criteria for each category C_i ($i = 0, 1, 2$). The combined signal efficiency is $\bar{\epsilon}$. The estimated number of background events is $\hat{B}$. The combined systematic uncertainty is σ_{tot}. We assume the total number of observed events $n = \sum_{i=0,1,2} n_i$ follows a Poisson distribution with parameter $\lambda = \mathcal{L}\sigma\epsilon + B$, where $\mathcal{L}$ is luminosity, σ is cross section of dark Upsilon decaying to three dark photons, ϵ is unknown true signal efficiency, and B is unknown true number of background. We also assume that the evaluated signal efficiency $\bar{\epsilon}$ follows a Gaussian distribution whose mean value is the true signal efficiency ϵ and the standard deviation is the estimated total systematic uncertainty σ^2_{tot}. The likelihood for this scanning window is:

$$L(n|\lambda = \mathcal{L}\sigma\epsilon + B) = Poisson(n|\lambda) \cdot \mathcal{N}(\bar{\epsilon}|\epsilon, \sigma^2_{tot}) \cdot Poisson(\hat{B}|B).$$

There are two nuisance parameters in the likelihood, ϵ and B. We apply a profile likelihood method to obtain a 90% confidence level limit on the cross section. For a likelihood L with parameters (θ, ψ), where θ are the nuisance parameters and ψ are the parameters of interest, the profile likelihood method takes two steps to optimize the L: For each ψ, we have a curve $L_\psi(\lambda)$ over λ. We first evaluate the maximal L over λ, then we choose the ψ that is maximum over all these curves.

To take the signal lifetime into consideration, we apply an iterative algorithm to find the converged signal efficiency and converged upper limits. We use Equation 6.7 to compute the theoretical dark photon lifetime and Equation 6.10 to correct the signal efficiencies. The detailed algorithm is described in Table 6.3.

Results

We do not observe a significant signal of dark matter bound state Υ_D. The 90% CL results on cross sections for both prompt and displaced signal decays are shown in Fig. 6.25 and 6.26. We derive separate limits for α_D values set to 0.1, 0.3, 0.5, 0.7, 0.9, and 1.1. Constraints for different values of α_D, $m_{A'}$, and m_{Υ_D} are shown in Fig. 6.27. The results compared with previous search are shown in Fig. 6.28 for different α_D. Bounds on the mixing strength ε down to $5 \times 10^{-5} - 10^{-3}$ are set for

Algorithm to derive the limit when the lifetime is non-zero
Step 0: Compute the zero-lifetime kinetic mixing strength upper limit (ϵ) using profile likelihood method.
Step 1: Calculate the corresponding dark photon lifetime (l) using Equation 6.7. If the lifetime is smaller than 10 um, go to **Step 5**; otherwise go to **Step 2**.
Step 2: Update the signal efficiency under the dark photon lifetime l using Equation 6.10.
Step 3: Update the corresponding mixing strength upper limit ϵ using the updated signal efficiency.
Step 4: Repeat **Step 1** until the procedure has converged.
Step 5: Return the converged signal efficiency, the cross-section upper limit and the mixing strength upper limit.

Table 6.3: Algorithm to find converged signal efficiency, cross-section upper limits, and kinetic mixing strength upper limits.

a wide range of dark photon mass from MeV to few GeV. Compared with previous search, our analysis shows that the constraints on $\gamma - A'$ coupling strength can be improved by a large fraction of parameter space and can be more than one order of magnitude for 40 MeV $\leq m_{A'} \leq$ 200 MeV with strong electromagnetic interaction in the dark sector, demonstrating the dark matter bound state as a sensitive probe of dark matter.

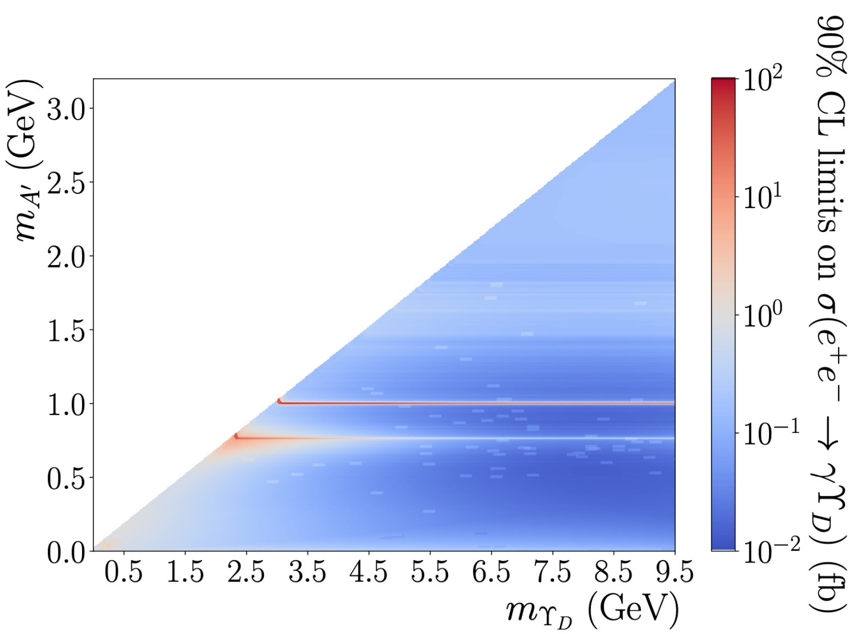

Figure 6.25: The 90% CL upper limits on the $e^+e^- \to \gamma\Upsilon_D$ cross section for prompt dark photon decays.

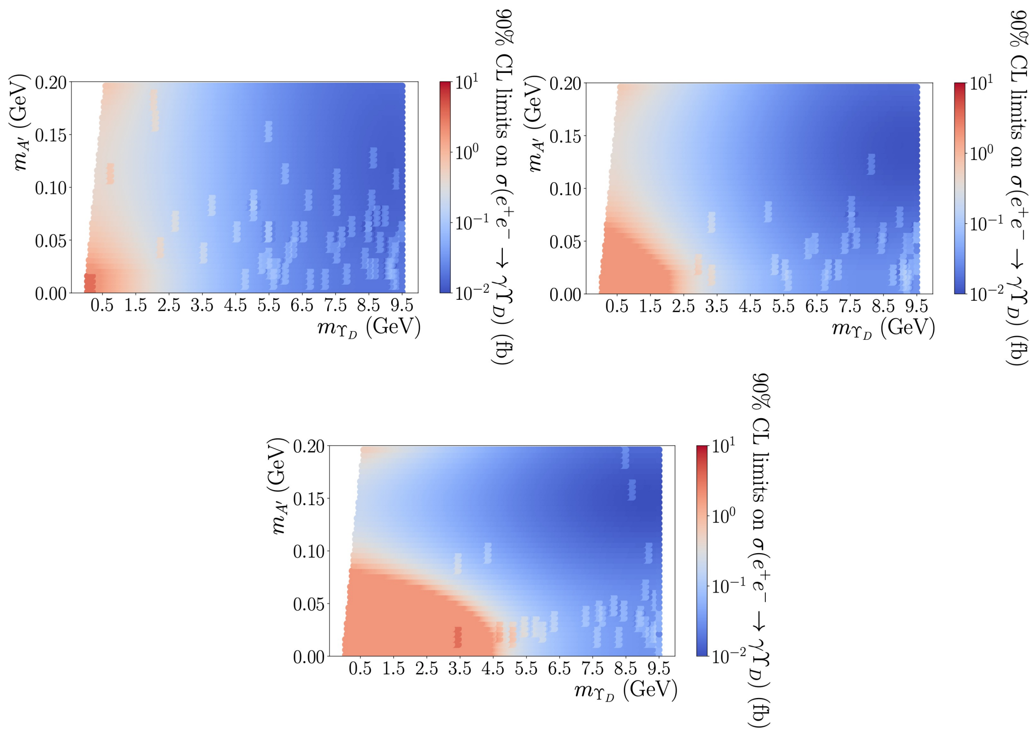

Figure 6.26: The 90% CL upper limits on the $e^+e^- \to \gamma\Upsilon_D$ cross section for dark photon lifetimes corresponding to (top left) $c\tau_{A'} = 0.1$ mm, (top right) $c\tau_{A'} = 1$ mm, and (bottom) $c\tau_{A'} = 10$ mm.

6.10 Summary and Outlook

In summary, we explore the sensitivity of collider probes on dark sector searches with the existence of dark sector structures. Dark matter bound states could exist under a simple dark photon model when the dark photon is light enough to generate an attractive force between dark fermions. We report the first search for a dark sector bound state decaying into three dark photons in the range 0.001 GeV $< m_{A'} <$ 3.16 GeV and 0.05 GeV $< m_{\Upsilon_D} <$ 9.5 GeV. We do not observe significant signals. The limits on the $\gamma - A'$ kinetic mixing ϵ are derived at the level of $5 \times 10^{-5} - 10^{-3}$, depending on the values of the model parameters. These measurements improve upon existing constraints over a significant fraction of dark photon masses below 1 GeV for large values of the dark sector coupling constant.

In the future, the search for η_D bound state can also be also included. With the current experiment data, the upper limits on the cross section (in the absence of a signal) could be improved by around a factor of 2, leading to an improvement on the constraints on the kinetic mixing strength by about a factor of $\sqrt{2}$. With 50 ab^{-1} luminosity collected by future B factory experiments (Belle II), the limits on the

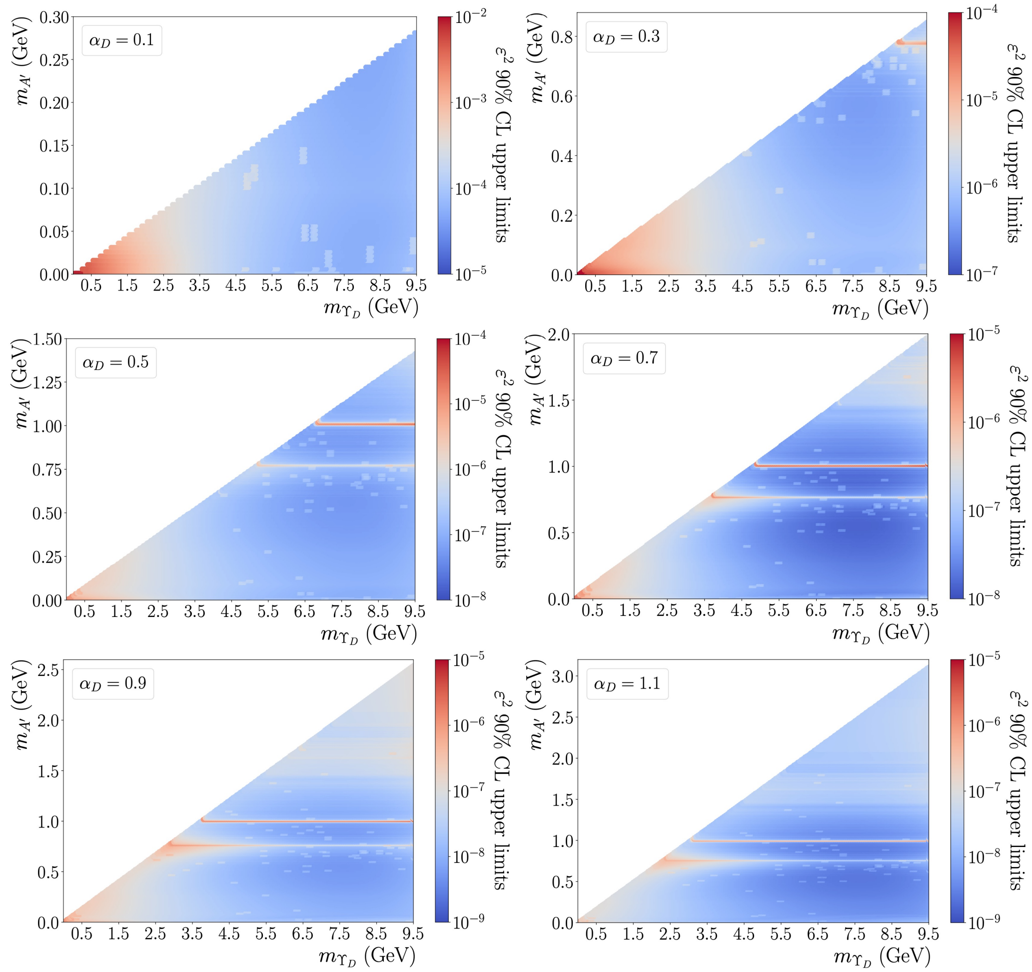

Figure 6.27: The 90% CL upper limits on the kinetic mixing ϵ^2 as a function of the Υ_D mass, m_{Υ_D}, and dark photon mass, $m_{A'}$, assuming (top left) $\alpha_D = 0.1$, (top right) $\alpha_D = 0.3$, (middle left) $\alpha_D = 0.5$, (middle right) $\alpha_D = 0.7$, (bottom left) $\alpha_D = 0.9$, and (bottom right) $\alpha_D = 1.1$.

kinematic mixing strength can potentially reduce by at least one order of magnitude.

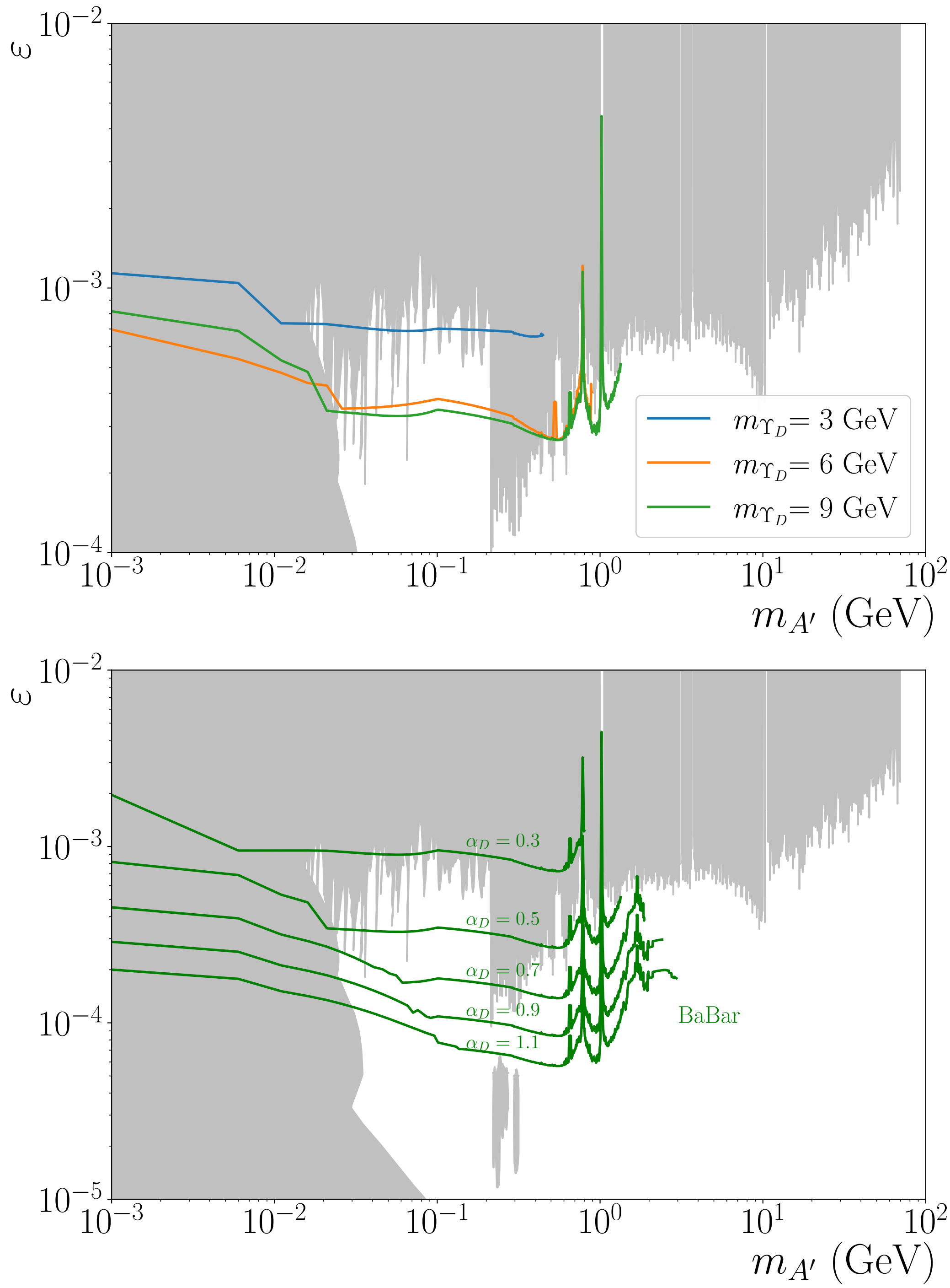

Figure 6.28: The 90% CL upper limits on the kinetic mixing ε for (top) various Υ_D masses assuming α_D = 0.5 and (bottom) various α_D values assuming m_{Υ_D} = 9 GeV together with current constraints (gray area).

Part IV

Precise Measurement of Semileptonic B Meson Decays

Chapter 7

MEASUREMENT OF $R(D)$ AND $R(D^*)$ USING SEMILEPTONIC TAGGING AND LEPTONIC τ DECAYS

7.1 Introduction

The excess of semitauonic decays $B \to D^{(*)}\tau\nu$ have been one of the most interesting puzzles in flavor physics in recent years [22]. The physical quantity to measure is the ratio between $B \to D^{(*)}\tau\nu$ and $B \to D^{(*)}l\nu$:

$$R(D^{(*)}) = \frac{\mathcal{B}(B \to D^{(*)}\tau^-\nu_\tau)}{\mathcal{B}(B \to D^{(*)}l^-\nu_l)}(l = e, \mu)$$

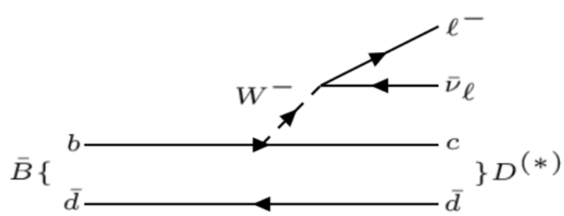

Figure 7.1: Feynman diagram of $B \to D^{(*)}\tau\nu$ decay.

In the Standard Model, $B \to D^{(*)}\tau\nu$ decay involves the semileptonic quark transition $b \to cl\bar{\nu}$ with $l = e, \mu, \tau$. The Feynman diagram is shown in Fig. 7.1. The b quark in the B meson decays to c quark via first-order electroweak interactions, and the mediator, a W boson, decays into pair of leptons $l\bar{\nu}$. The other quark $\bar{q}$ in the B meson binds with the c quark in the final state to form a $D^{(*)}$ meson. The Lagrangian of the system is

$$\mathcal{L} = \frac{G_F}{\sqrt{2}}|V_{cb}|J^{+\mu}_{\nu l}J_{cb,\mu} + h.c.,$$

which involves the quark currents $J^{\mu}_{cb} = \bar{\psi}_c\gamma^\mu(1 - \gamma^5)\psi_b$ and the leptonic currents $J^{\mu}_{\nu l} = \bar{\psi}_\nu\gamma^\mu(1 - \gamma^5)\psi_l$. The amplitudes of this scattering process are:

$$\mathcal{M}^{\lambda_l}_{\lambda_M}(q^2, x) = \frac{G_F}{\sqrt{2}}V_{cb}\sum_{\lambda_W}\eta_{\lambda_W}L^{\lambda_l}_{\lambda_W}H^{\lambda_M}_{\lambda_W},$$

where λ is the particle helicities, $M = D, D^*$ indicates the final state meson. The leptonic amplitudes $L^{\lambda_l}_{\lambda_W}$ can be analytically calculated with the standard framework of electroweak interactions [23]. Although the analytical calculation of hadronic amplitudes is intractable because of complex interactions between quarks and the self-interactions of the gluons, they can be expressed in terms of six functions, or the *form factors* depending only on q^2: $f_\pm(q^2)$ and $f_i(q^2), i = 1,2,3,4$, where $q = p_B - p_{D^{(*)}}$ is the four-momentum of the virtual W. One model to form factors is the Caprini-Lellouch-Neubert (CLN) model [24], which introduces dispersive constraints from heavy quark symmetry to provide relations between the form factors near zero recoil. The SM predictions for $R(D^{(*)})$ are:

$$R(D)_{\text{SM}} = 0.299 \pm 0.003,$$

$$R(D^*)_{\text{SM}} = 0.254 \pm 0.005.$$

The uncertainties on $R(D^{(*)})_{\text{SM}}$ come from the uncertainties of the form factor parameters.

If new particles couple proportionally with their masses, $B \to D^{(*)}\tau\nu$ are sensitive probes for non-SM contributions due to the heavy τ mass. The individual decay modes will suffer from large hadronic uncertainties related to the form factors and the Cabibbo-Kobayashi-Maskawa (CKM) elements. Normalizing the $B \to D^{(*)}\tau\nu$ to the corresponding decay with light leptons provides better sensitivity to new physics (NP).

Since the energy scale of NP should be far above the scale of B meson, we can integrate out higher degrees of freedom and form the following effective Hamiltonian:

$$\mathcal{H}_{\text{eff}} = C^{cb}_{SM} O^{cb}_{SM} + C^{cb}_{R} O^{cb}_{R} + C^{cb}_{L} O^{cb}_{L},$$

where

$$O^{cb}_{SM} = \bar{q}\gamma_\mu P_L b \bar{\tau}\gamma_\mu P_L \nu_\tau$$
$$O^{cb}_{R} = \bar{q} P_R b \bar{\tau} P_L \nu_\tau$$
$$O^{cb}_{L} = \bar{q} P_L b \bar{\tau} P_L \nu_\tau.$$

The first term is the SM charged current. The second and third terms are the four-fermion operators allowed by Lorentz invariance [25]. The SM Wilson coefficient $C^{cb}_{SM} = \frac{4G_F V_{cb}}{\sqrt{2}}$. The corresponding Wilson coefficients C^{cb}_R and C^{cb}_L affect the observables in the following way:

$$R(D) = R_{\text{SM}}(D)\left(1 + 1.5\Re\left[\frac{C^{cb}_R + C^{cb}_L}{C^{cb}_{SM}}\right] + 1.0\left|\frac{C^{cb}_R + C^{cb}_L}{C^{cb}_{SM}}\right|^2\right)$$

$$R(D^*) = R_{\text{SM}}(D^*)\left(1 + 0.12\Re\left[\frac{C_R^{cb} - C_L^{cb}}{C_{SM}^{cb}}\right] - 0.05\left|\frac{C_R^{cb} - C_L^{cb}}{C_{SM}^{cb}}\right|^2\right).$$

Therefore, any deviation between the SM prediction and observed value of $R(D^{(*)})$ may imply existence of NP, including

- Supersymmetry: In the Minimal Supersymmetric Standard Model (MSSM), a charged Higgs boson which couples proportionally to the masses of the fermions is introduced. The Higgs only couples significantly to τ, so that it contributes to the deviation between SM prediction and observed value of measured quantity.
- Lepton flavor universality (LU) violation: The $b \to cl\nu$ transition is a Flavor-Changing Charged Current (FCCC) test. If LU is violated, the Wilson coefficients would be different for the operators with the same structure but different lepton flavors, contributing additional term to $R(D^{(*)})$.

The first measurement of $R(D^{(*)})$ were at BABAR in 2008 [26], and the first observation of an excess was on 2012 [27]. The measurements use hadronic tags to fully reconstruct B mesons and leptonic τ decays. The measured value has a 3.4σ deviation from SM predictions. Since then, many other experiments have contributed their measurements to the quantity of interest [28, 29, 30]. As of 2019, the $R(D)$ and $R(D^*)$ exceeds the SM prediction by 1.4σ and 2.5σ respectively. Considering the $R(D) - R(D^*)$) correlation of -0.38, The difference with the SM predictions reported above, corresponds to about 3.08σ. Fig. 7.2 and Table 7.1 show a summary of previous measurements of this quantity.

The goal of this BABAR analysis is to provide another measurement of this quantity using, for the first time, a semileptonic tagging method.

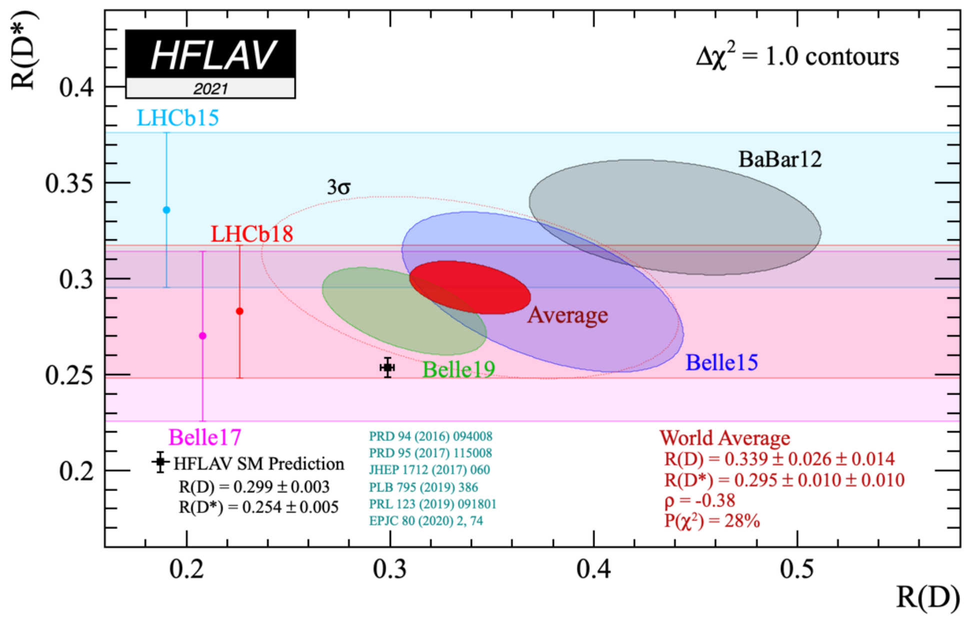

Figure 7.2: Visualization of previous measurements of $R(D)$ and $R(D^*)$. The plot is from HFLAV [22].

	$R(D)$	$R(D^*)$
BABAR 2013 [23]	$0.440 \pm 0.058 \pm 0.042$	$0.332 \pm 0.024 \pm 0.018$
Belle 2015 [28]	$0.375 \pm 0.064 \pm 0.026$	$0.293 \pm 0.038 \pm 0.015$
LHCb 2015 [29]	-	$0.336 \pm 0.027 \pm 0.03$
Belle 2016 [31]	-	$0.302 \pm 0.030 \pm 0.011$
Belle 2017 [32]	-	$0.270 \pm 0.035^{+0.028}_{-0.025}$
LHCb 2018 [33]	-	$0.291 \pm 0.019 \pm 0.029$
Belle 2019 [30]	$0.307 \pm 0.037 \pm 0.016$	$0.283 \pm 0.018 \pm 0.014$
Average (HFLAV Summer 2018)	$0.340 \pm 0.027 \pm 0.013$	$0.295 \pm 0.011 \pm 0.008$
Standard Model [22]	0.299 ± 0.003	0.258 ± 0.005

Table 7.1: Previous results.

7.2 Analysis Strategy Overview

Event types definition

For the $B\bar{B}$ system, the signal events are defined as $B_{tag} \longrightarrow D^{(*)}l\nu, B_{sig} \longrightarrow D^{(*)}\tau\nu$. On the tag B side, the B meson decays to $D^{(*)}l\nu$, where $l = e, \mu$. On the signal B side, the B meson decays to $D^{(*)}\tau\nu$, and the τ subsequently decays to leptons. We have two types of signal events, based on whether there is a D or D^* meson. The normalization events are defined as events with both the tag side B meson and signal side B meson decaying to $D^{(*)}l\nu$. One of the most important category of backgrounds arises from a $B \to D^{**}l\nu$ events where D^{**} denotes the excited charm meson states heavier than D^* because of the similar decay topology to signal events. Two other sources of background are combinatorial $B\bar{B}$ events and continuum events. The definition of these events is listed in Table 7.2.

Event type		Description
Signal event	signal D	One B decays to $D^{(*)}l\nu$, the other B decays to $D\tau\nu$, $\tau \to leptons$
	signal D^*	One B decays to $D^{(*)}l\nu$, the other B decays to $D^*\tau\nu$, $\tau \to leptons$
Normalization event	norm D	One B decays to $D^{(*)}l\nu$, the other B decays to $Dl\nu$
	norm D^*	Both B decay to $D^*l\nu$
D^{**} event		At least one B decays to $D^{**}(l/\tau)\nu$, where D^{**} denotes any excited charmed meson states that are not in the ground state $1S$ doublet. In this analysis, it includes $1P$ states D_0^*, D_1, D_1', D_2^*, and $2S$ states. We also allow non-resonant final states consisting of a $D^{(*)}$ and one pion.
Combinatorial $B\bar{B}$ event		Any $B\bar{B}$ events that are not signal and not normalization and not D^{**}.
Continuum event		Non-$B\bar{B}$ events produced in the detector

Table 7.2: Definition of event types in the $B\bar{B}$ system.

Reconstruction

The first step of this analysis is to identify primitive particles (e.g, $e, \mu, \pi^{\pm}, K, \gamma$), and reconstruct composite particles (e.g, π^0, K_S, $D^{(*)}$, B) by applying a series of selection criteria. When we reconstruct signal events, we first reconstruct a B_{tag}, then search for $D^{(*)}l$ in the remaining tracks and calorimeter clusters. The signal events can be categorized into four disjoint subsets based on the reconstructed type of D meson: $D^+l, D^0l, D^{*+}l, D^{*0}l$. We denote these as *channel_labels* for convenience. The details of the reconstruction and selection criteria are described in Section 7.4.

Estimate of $R(D^{(*)})$

To measure $R(D^{(*)})$, let us denote

$$\begin{aligned} P &:= \mathcal{B}(B \to D\tau\nu) \\ P^* &:= \mathcal{B}(B \to D^*\tau\nu) \\ Q &:= \mathcal{B}(B \to Dl\nu)(\text{average for } l = e \text{ or } \mu) \\ Q^* &:= \mathcal{B}(B \to D^*l\nu)(\text{average for } l = e \text{ or } \mu). \end{aligned} \tag{7.1}$$

Therefore, $R(D) = \frac{P}{Q}$ and $R(D^*) = \frac{P^*}{Q^*}$. If we denote the number of generated $B\bar{B}$ event as N, the relationship between the B decay branching fractions and the expected number of signal events generated in the detector would be

$$
\begin{aligned}
&N(\text{signal } D) = 2N\cdot(2Q+2Q^*)\cdot P\cdot\mathcal{B}(\tau\to leptons)\\
&N(\text{signal } D^*) = 2N\cdot(2Q+2Q^*)\cdot P^*\cdot\mathcal{B}(\tau\to leptons)\\
&N(\text{norm } D) = 4N\cdot(Q^2+2QQ^*)\\
&N(\text{norm } D^*) = 4N\cdot Q^{*2}.
\end{aligned}
\tag{7.2}
$$

Given the estimated number of generated signal events $\hat{N}(\text{signal } D)$ and $\hat{N}(\text{signal } D^*)$ and normalization events $\hat{N}(\text{norm } D)$ and $\hat{N}(\text{norm } D^*)$, the estimated $P^{(*)}$ can be derived from Equation (7.2):

$$
\begin{aligned}
\hat{P} &= \frac{\hat{N}(\text{signal } D)}{2\sqrt{N}\cdot\mathcal{B}(\tau\to leptons)\cdot\sqrt{\hat{N}(\text{norm } D)+\hat{N}(\text{norm } D^*)}}\\
\hat{P}^* &= \frac{\hat{N}(\text{signal } D^*)}{2\sqrt{N}\cdot\mathcal{B}(\tau\to leptons)\cdot\sqrt{\hat{N}(\text{norm } D)+\hat{N}(\text{norm } D^*)}}\\
\hat{Q} &= \frac{\sqrt{\hat{N}(\text{norm } D)+\hat{N}(\text{norm } D^*)}-\sqrt{\hat{N}(\text{norm } D^*)}}{\sqrt{4N}}\\
\hat{Q}^* &= \sqrt{\frac{\hat{N}(\text{norm } D^*)}{4N}}.
\end{aligned}
\tag{7.3}
$$

Therefore, the estimated $R(D^{(*)})$, as functions of the estimated number of generated events $\hat{N}(\text{signal } D)$ and $\hat{N}(\text{signal } D^*)$, would be

$$
\begin{aligned}
R(D) =&\frac{1}{2\mathcal{B}(\tau\to leptons)}\cdot\\
&\frac{\hat{N}(\text{signal } D)}{\sqrt{\hat{N}(\text{norm } D)+\hat{N}(\text{norm } D^*)}\cdot\left(\sqrt{\hat{N}(\text{norm } D)+\hat{N}(\text{norm } D^*)}-\sqrt{\hat{N}(\text{norm } D^*)}\right)}\\
R(D^*) =&\frac{1}{2\mathcal{B}(\tau\to leptons)}\cdot\\
&\frac{\hat{N}(\text{signal } D^*)}{\sqrt{\hat{N}(\text{norm } D)+\hat{N}(\text{norm } D^*)}\cdot\sqrt{\hat{N}(\text{norm } D^*)}}.
\end{aligned}
\tag{7.4}
$$

The branching fraction of τ decays to leptons is taken to be [34]

$$
\begin{aligned}
&\mathcal{B}(\tau\to leptons) = \mathcal{B}(\tau\to e\bar{\nu}_e\nu_\tau)+\mathcal{B}(\tau\to\mu\bar{\nu}_\mu\nu_\tau)\\
&= (35.24\pm 0.05)\%
\end{aligned}
\tag{7.5}
$$

$\hat{N}(\text{signal } D)$, $\hat{N}(\text{signal } D^*)$, $\hat{N}(\text{norm } D)$, and $\hat{N}(\text{norm } D^*)$ can be estimated as the extracted yields divided by the corresponding efficiencies.

7.3 Simulation Samples

The simulated data includes generic MC, signal MC, as well as normalization MC events.

The generic MC aims to represent the data collected in the detector by faithfully generating the possible results of the e^+e^- collisions. The generic MC includes two types of events: $B\bar{B}$ events and continuum events. $B\bar{B}$ events emulate both charged and neutral $B\bar{B}$ meson pairs produced from $\Upsilon(4S)$. Continuum sample simulates $e^+e^- \to q\bar{q}$. Other constributions (QED, two-photon processes) are negligible. The weighted collection of these types of events gives our best reproduction of what an on-peak collider run produces. To weight the generic MC, we need to consider the cross section of each type of event by the following formula:

$$w_i = \mathcal{L}\frac{\sigma_i}{N_i},$$

where $\mathcal{L}$ is the integrated detector data luminosity, σ_i is the corresponding cross section for event component i, and N_i is the number of event components i generated. The cross sections for each event type are listed in Table 7.3. The components of the generic MC dataset are listed in Table 7.4.

SP Mode	Mode type	Cross section (pb)
1235	B^+B^-	525.0
1237	$B^0\overline{B}^0$	525.0
1005	$c\overline{c}$	1300.0
998	uds	2090.0

Table 7.3: Cross sections used to convert the sizes generic simulated data to the equivalent on-peak dataset.

The signal (normalization) MC is generated by forcing every B meson decay in the signal (normalization) modes. The datasets we use are listed in Table 7.5 and Table 7.6, in which both charged and neutral $B\bar{B}$ meson pairs are simulated. The signal (normalization) MC datasets are used in the analysis described later.

Simulated Dataset Name	Mode Type	Collisions Generated	Multiplier
SP-1235-AllEventsSkim-Run1-R24a1	B^+B^-	34878000	0.306
SP-1235-AllEventsSkim-Run2-R24a1	B^+B^-	105561000	0.305
SP-1235-AllEventsSkim-Run3-R24a1	B^+B^-	56035000	0.303
SP-1235-AllEventsSkim-Run4-R24a1	B^+B^-	166784000	0.314
SP-1235-AllEventsSkim-Run5-R24a1	B^+B^-	215168000	0.323
SP-1235-AllEventsSkim-Run6-R24a1	B^+B^-	130336000	0.316
SP-1237-AllEventsSkim-Run1-R24a1	$B^0\overline{B}^0$	34941000	0.306
SP-1237-AllEventsSkim-Run2-R24a1	$B^0\overline{B}^0$	104188000	0.308
SP-1237-AllEventsSkim-Run3-R24a1	$B^0\overline{B}^0$	57888000	0.292
SP-1237-AllEventsSkim-Run4-R24a1	$B^0\overline{B}^0$	169801000	0.307
SP-1237-AllEventsSkim-Run5-R24a1	$B^0\overline{B}^0$	215953000	0.321
SP-1237-AllEventsSkim-Run6-R24a1	$B^0\overline{B}^0$	135224000	0.304
SP-1005-AllEventsSkim-Run1-R24a1	$c\overline{c}$	55254000	0.479
SP-1005-AllEventsSkim-Run2-R24a1	$c\overline{c}$	164722000	0.483
SP-1005-AllEventsSkim-Run3-R24a1	$c\overline{c}$	88321000	0.475
SP-1005-AllEventsSkim-Run4-R24a1	$c\overline{c}$	267308000	0.484
SP-1005-AllEventsSkim-Run5-R24a1	$c\overline{c}$	344275000	0.499
SP-1005-AllEventsSkim-Run6-R24a1	$c\overline{c}$	208664000	0.488
SP-998-AllEventsSkim-Run1-R24a1	uds	176404000	0.241
SP-998-AllEventsSkim-Run2-R24a1	uds	525504000	0.243
SP-998-AllEventsSkim-Run3-R24a1	uds	276381000	0.244
SP-998-AllEventsSkim-Run4-R24a1	uds	845899000	0.246
SP-998-AllEventsSkim-Run5-R24a1	uds	1110944000	0.249
SP-998-AllEventsSkim-Run6-R24a1	uds	655152000	0.250

Table 7.4: Generic simulated data. Multiplier is the factor by which the size of the corresponding on-peak dataset exceeds that of the given simulated dataset. More simulated datasets are usually generated to better study the decay characteristics.

7.4 Event Reconstruction

The event reconstruction procedure includes three steps, all of which are implemented using the *BABAR* offline analysis framework (version-52).

1. Pre-screening: We use a collection of loose criteria to remove obvious background events. This step helps reduce the amount of data to analyze.

2. Candidate reconstruction: We then attempt to reconstruct the B_{tag} and B_{sig} of signal events from a set of tracks of final state particles. The event reconstruction is a hierarchical process: we first identify primitive particles (e.g,

Simulated Dataset Name	Mode Type	N_{event} Collected
SP-11440-Run1-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D\tau(e,\mu)\nu$	2621000
SP-11440-Run2-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D\tau(e,\mu)\nu$	
SP-11440-Run3-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D\tau(e,\mu)\nu$	
SP-11440-Run4-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D\tau(e,\mu)\nu$	
SP-11440-Run5-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D\tau(e,\mu)\nu$	
SP-11440-Run6-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D\tau(e,\mu)\nu$	
SP-11441-Run1-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^*\tau(e,\mu)\nu$	2756000
SP-11441-Run2-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^*\tau(e,\mu)\nu$	
SP-11441-Run3-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^*\tau(e,\mu)\nu$	
SP-11441-Run4-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^*\tau(e,\mu)\nu$	
SP-11441-Run5-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^*\tau(e,\mu)\nu$	
SP-11441-Run6-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^*\tau(e,\mu)\nu$	
SP-11442-Run1-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D\tau(e,\mu)\nu$	2789000
SP-11442-Run2-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D\tau(e,\mu)\nu$	
SP-11442-Run3-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D\tau(e,\mu)\nu$	
SP-11442-Run4-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D\tau(e,\mu)\nu$	
SP-11442-Run5-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D\tau(e,\mu)\nu$	
SP-11442-Run6-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D\tau(e,\mu)\nu$	
SP-11443-Run1-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^*\tau(e,\mu)\nu$	2356000
SP-11443-Run2-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^*\tau(e,\mu)\nu$	
SP-11443-Run3-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^*\tau(e,\mu)\nu$	
SP-11443-Run4-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^*\tau(e,\mu)\nu$	
SP-11443-Run5-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^*\tau(e,\mu)\nu$	
SP-11443-Run6-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^*\tau(e,\mu)\nu$	

Table 7.5: Simulated signal data.

e, μ, γ), then reconstruct $D^{(*)}$ and B mesons, and then combine B and $\bar{B}$ pairs to reconstruct $\Upsilon(4S)$ candidates.

3. Candidate selection: When multiple candidates are involved in an event after reconstruction, we select a single best candidate, as defined below.

Pre-screening

The following set of broad selection criteria is first applied to loosely pre-select signal events. These selection criteria, studied from MC samples, are set to exclude obvious backgrounds but keep most signal and normalization events.

- Size of `ChargedTracks` ≤ 14,

Simulated Dataset Name	Mode Type	N_{event} Collected
SP-11438-Run1-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^{(*)}\ell\nu$	2443000
SP-11438-Run2-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^{(*)}\ell\nu$	
SP-11438-Run3-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^{(*)}\ell\nu$	
SP-11438-Run4-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^{(*)}\ell\nu$	
SP-11438-Run5-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^{(*)}\ell\nu$	
SP-11438-Run6-R24	$B^+ \to D^{(*)}\ell\nu$, $B^- \to D^{(*)}\ell\nu$	
SP-11439-Run1-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^{(*)}\ell\nu$	2869000
SP-11439-Run2-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^{(*)}\ell\nu$	
SP-11439-Run3-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^{(*)}\ell\nu$	
SP-11439-Run4-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^{(*)}\ell\nu$	
SP-11439-Run5-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^{(*)}\ell\nu$	
SP-11439-Run6-R24	$B^0 \to D^{(*)}\ell\nu$, $\bar{B}^0 \to D^{(*)}\ell\nu$	

Table 7.6: Simulated normalization data.

- Size of `GoodPhotonsLoose` ≤ 10,
- $-2 \leq Q_{total} \leq 2$,
- Apply tag filter `BGFMultiHadron`,
- Apply tag filter `TagL3`.

Candidate reconstruction

Events passing the pre-screening are then used to reconstruct B_{tag} and B_{sig}. Initially, each event is a collection of tracks of final-state particles. To reconstruct the signal candidates, we first identify final-state particles using built-in Particle Identification algorithms. The loosest selectors are applied to identify $e, \mu, \pi^{\pm}, K$, and γ.

The π^0 mesons are then reconstructed from $\pi^0 \to \gamma\gamma$ by combining a pair of photons, and K_S is reconstructed from the $K_S \to \pi^+\pi^-$ mode. The following are applied to select light mesons:

- π^0: `pi0AllDefault` and `pi0SoftDefaultMass`. Using the photon list above, reconstruct $\pi^0 \to \gamma\gamma$ with the mass of the pion candidate constrained to be between $[0.115, 0.15]$ GeV.
- K_S: `KsDefault`. Reconstruct $K_S \to \pi^+\pi^-$ using TreeFitter.

The D mesons are reconstructed based on the identified final state particles and light mesons. The D decay modes listed in Table 7.7 are considered in reconstruction, which sum up to 19.2% and 30.1% of D^+ and D^0 branching fractions. Other decay modes are not considered in this analysis due to higher backgrounds.

Decay Modes	Branching Fraction (%)	Number in Generic MC
$D^+ \to K^-\pi^+\pi^+$	9.46 ± 0.24	142202
$D^+ \to K_S\pi^+$	1.53 ± 0.06	22691
$D^+ \to K_S\pi^+\pi^0$	7.24 ± 0.17	430750
$D^+ \to K^-K^+\pi^+$	0.99 ± 0.026	62859
$D^0 \to K^-\pi^+$	3.88 ± 0.05	158718
$D^0 \to K^-\pi^+\pi^0$	14.3 ± 0.80	1235214
$D^0 \to K^-\pi^+\pi^-\pi^+$	8.06 ± 0.23	646768
$D^0 \to K_S\pi^+\pi^-$	2.85 ± 0.20	641954
$D^0 \to K_S\pi^0$	1.20 ± 0.04	94151
$D^0 \to K^-K^+$	0.40 ± 0.007	49909

Table 7.7: D meson decay modes used in the analysis.

The mass of D mesons ($m(D)$) is used to select correctly reconstructed candidates. Since the mass resolutions vary for different types of D meson decay modes, our criteria on $m(D)$ are different depending on the D decay modes. The distribution of $m(D)$ and the criteria is shown in Fig. 7.3.

– D decay without a π^0 in the final state: The mass resolution (σ_{m_D}) for this type of D decay modes is around 5 MeV. Therefore, we require $m(D)$ within [-15, 15] MeV of the nominal mass, which corresponds to $\pm 3\sigma_{m_D}$.

– D^+ decay with a π^0 in the final state: The mass resolutions for this type of D decay modes is around 12 MeV. We require $m(D^\pm)$ within [-36, 24] MeV of the nominal charged D mass, corresponding to [-3σ_{m_D}, +2σ_{m_D}].

– D^0 decay with a π^0 in the final state: The mass resolutions for this type of D decay modes is around 15 MeV. We require $m(D^\pm)$ within [-45, 30] MeV of the nominal neutral D mass, corresponding to [-3σ_{m_D}, +2σ_{m_D}].

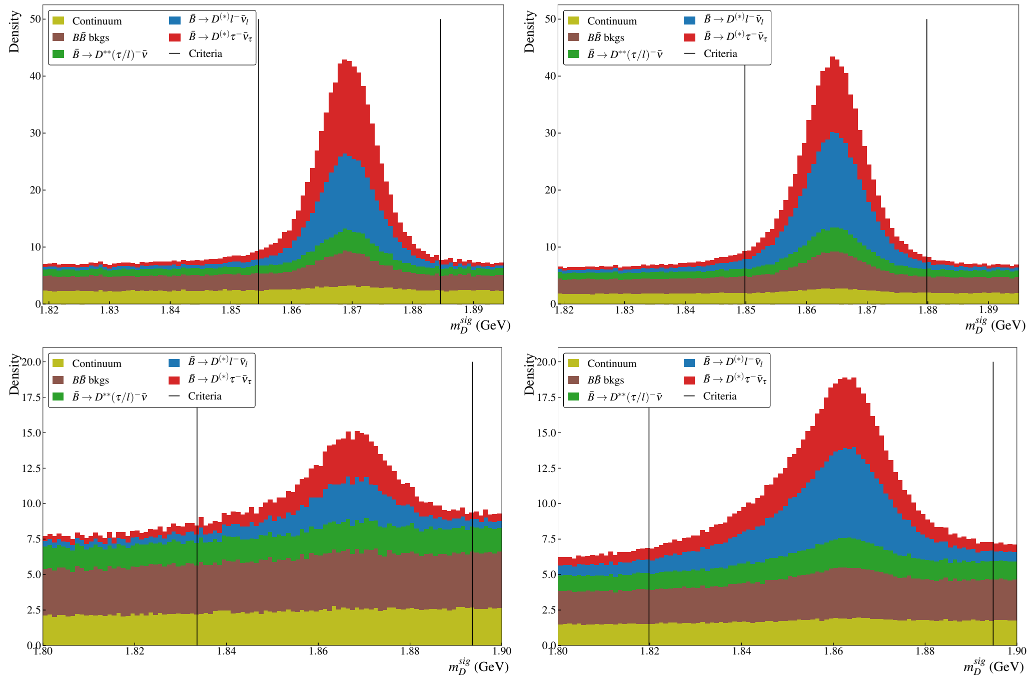

Figure 7.3: The distributions of reconstructed D meson mass for (upper left) D^+ decays without π^0, (upper right) D^0 decays without π^0, (bottom left) D^+ decays with π^0, and (bottom right) D^0 decays with π^0.

The D^* mesons are reconstructed in three decay modes. Other decay modes are not considered, due to the high backgrounds. The mass difference between reconstructed D^* meson and its daughter D meson (ΔM) is used to select well-reconstructed D^* mesons. The resolution on ΔM is around 2 MeV. Tighter criteria are applied when there is a π^0 in the final state, as large backgrounds arise from mis-reconstructed neutral pions. The distributions of mass differences for each type of decays are shown in Fig. 7.4.

- $D^{*+} \to D^0\pi^+$: We require ΔM to be within 2.5 MeV of the nominal mass difference.
- $D^{*+} \to D^+\pi^0$: We require ΔM to be within 2.0 MeV of the nominal mass difference.
- $D^{*0} \to D^0\pi^0$: We require ΔM to be within 2.0 MeV of the nominal mass difference.

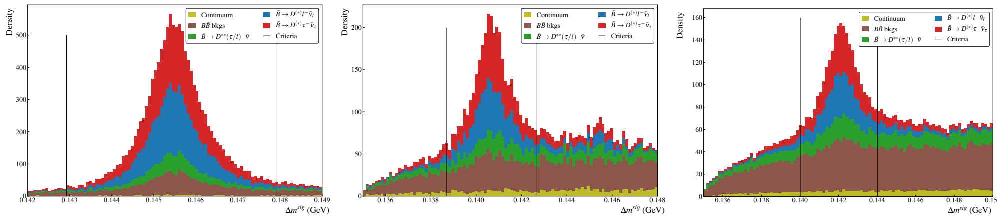

Figure 7.4: The distribution of mass difference between reconstructed D^* meson and its daughter D meson for (left) $D^{*+} \to D^0\pi^+$, (middle) $D^{*+} \to D^+\pi^0$, and (right) $D^{*0} \to D^0\pi^0$ decays.

The B_{tag} mesons are reconstructed by combining a $D^{(*)}$ and an electron or muon candidate. The $D^{(*)}l$ invariant mass is required to be at most 5.2791 GeV, and the p-value for the χ^2 test of the kinematic fit must be at least 0.001. The variable used to select a correctly reconstructed B_{tag} is the angle between the 3-momentum of the B_{tag} and the 3-momentum sum of its $D^{(*)}$ and lepton daughters ($\cos\theta^{tag}_{B-D^{(*)}l}$), defined as

$$\cos\theta^{tag}_{B-D^{(*)}l} = \frac{2E_{beam}E_{D^{(*)}l} - m_B^2 - m_{D^{(*)}l}^2}{2|\mathbf{p}_B| \cdot |\mathbf{p}_{D^{(*)}l}|}.$$

The value of $\cos\theta^{tag}_{B-D^{(*)}l}$ should be in the range of [-1, 1], as it has only one missing neutrino. Other events (signal, combinatorial, continuum events) tend to have more negative $\cos\theta^{tag}_{B-D^{(*)}l}$ values. We select events requiring $\cos\theta^{tag}_{B-D^{(*)}l} \in [-2, 1]$ to take the detector resolutions into consideration.

- $B^+ \to \bar{D}^0 e^+$.
- $B^+ \to \bar{D}^0 \mu^+$.
- $B^0 \to D^- e^+$.
- $B^0 \to D^- \mu^+$.
- $B^+ \to \bar{D}^{*0} e^+$.
- $B^+ \to \bar{D}^{*0} \mu^+$.
- $B^0 \to D^{*-} e^+$.
- $B^0 \to D^{*-} \mu^+$.

In each event with a selected B_{tag} candidate, we search for $D^{(*)}l$ among the remaining tracks and calorimeter clusters to reconstruct B_{sig}, using the B decay modes described above. The invariant mass of $D^{(*)}l$ candidates is required to be at most 5.2791 GeV, and the p-value of the χ^2 test must be at least 0.001.

The $\Upsilon(4S)$ candidates are reconstructed by combining B_{tag} and B_{sig}, requiring that the two B mesons must conserve charge. To further suppress backgrounds, we require:

- No extra charged tracks, K_S^0 or π^0 particles must be present.
- The extra neutral energy in the calorimeter $E_{extra} \leq 1.2$ GeV.
- The second Fox-Wolfram moment $R_2 \leq 0.4$.
- The 3-momentum magnitude of the B_{sig}'s lepton daughter in the CM frame $|p_l^{sig}| \leq 2$ GeV.

Candidate selection

We use the following algorithm to assign a truth-matched candidate for each signal MC event. The criteria for the best candidate is clear: if the reconstruction graph of a candidate matches exactly that of the truth, that is our best candidate. The truth matching algorithm is described in detail in [35].

About 91% of the reconstructed events have only a single candidate. If multiple candidates are observed in an event, we choose the candidate with the minimal E_{extra} as the best candidate to represent the event. Following this method, 95.8% of the reconstructed signal candidates are the truth-matched candidate.

Reconstruction results

Table 7.8 shows the composition of events selected by the reconstruction strategy evaluated on generic MC. 1.56% (3.18%) of our reconstructed generic MC sample are $B \to D\tau\nu$ ($B \to D^*\tau\nu$) events. The normalization events ($B \to D^{(*)}l\nu$) sum up to around 38%. Combinatorial $B\bar{B}$ background dominates and is roughly 28%, and we also have around 6.4% of continuum events.

Component	Proportion (%)
$B \to D\tau\nu$	1.56
$B \to D^*\tau\nu$	3.18
$B \to Dl\nu$	19.80
$B \to D^*l\nu$	18.11
$B \to D^{**}l\nu$	22.74
Combinatorial $B\bar{B}$	28.17
Continuum	6.44

Table 7.8: Proportions of each component after reconstruction, evaluated using the generic MC.

7.5 Signal Detection

We apply machine learning models to classify each type of event after reconstruction. Two binary classifiers C_1, C_2 are used, the corresponding scores denoted as z_1 and z_2. C_1 is used to identify signal events and normalization events from background D^{**}, $B\bar{B}$ combinatorial and continuum events. C_2 is used to separate signal and normalization events. The combination of z_1 and z_2 enables us to measure signal events based on their density shape difference from other event types. We do not apply any selection criteria on z_1 or z_2, however, the (z_1, z_2) spectrum of all event types is used to extract the signal yields.

C_1 classifier

The sample used to train and validate the C_1 classifier includes 350K events from the generic MC. The sample is divided into two parts using random sampling without replacement: training sample (240K) and validation sample (110K). We first use the training sample to train different classifiers and then use the validation sample to choose the best classifier.

Generally, a binary classifier aims to classify two categories of events (often labeled as positive/negative), based on a set of input variables. C_1 is a binary classifier to distinguish signal events and normalization events, from all types of background events. When training the C_1 classifier, we label signal and normalization events as positive, and label D^{**}, combinatorial $B\bar{B}$, and continuum events as negative. The following variables are used as inputs to train the C_1 classifier. The histograms of these variables for positive label (signal and normalization events) and each type of background event are shown in Fig. 7.5, 7.6, and 7.7.

- N_{tracks}: Number of charged tracks.

- R_2 All: Second Fox-Wolfram moment.
- M^2_{miss}: Square of the missing 4-vector of the event in the CM frame.
- E_{extra}: Extra neutral energy in the calorimeter.
- $\cos\theta_T$: Cosine of the angle between the thrust and the beam momentum.
- $|p_l^{tag}|$: 3-momentum magnitude of the B_{tag} lepton in the CM frame.
- $\cos\theta^{tag}_{B-D^{(*)}l}$: Cosine of the angle between the 3-momentum of the B_{tag} and the 3-momentum sum of its D and lepton daughters in the CM frame.
- $\cos\theta^{tag}_{D-l}$: Cosine of the angle between the 3-momentum of the D meson and the lepton daughter in the tag side.
- m_D^{tag}: Mass of the B_{tag} D meson daughter.
- Δm^{tag}: Mass difference between D^* and D meson in the tag side, if exists.
- $\cos\theta^{tag}_{D\,soft}$: Cosine of the angle between the D^* mesons daughters in the tag side in the CM frame.
- $|p^{tag}_{soft}|$: 3-momentum magnitude of the D^*'s soft daughter in the tag side in the CM frame.
- $|p_l^{sig}|$: 3-momentum magnitude of the B_{sig} lepton daughter in the CM frame.
- $\cos\theta^{sig}_{D-l}$: Cosine of the angle between the 3-momentum of the D meson and the lepton daughter in the sig side in the CM frame.
- χ^2: χ^2 of the B_{sig} vertex fit.
- m_D^{sig}: Mass of the B_{sig} D meson daughter.
- Δm^{sig}: Mass difference between D^* and D meson in the sig side, if it exists.
- $\cos\theta^{sig}_{D\,soft}$: Cosine of the angle between the D^* mesons' daughters in the sig side in the CM frame.
- $|p^{sig}_{soft}|$: 3-momentum magnitude of the D^*'s soft daughter in the sig side in the CM frame.
- $\cos\theta_{D^{(*)}l-D^{(*)}l}$: Cosine of the angle between the two Dl Dl systems in the CM frame.

- tag l electron PID: B_{tag} lepton daughter's electron PID level.
- tag l muon PID: B_{tag} lepton daughter's muon PID level.
- sig l electron PID: B_{sig} lepton daughter's electron PID level.
- sig l muon PID: B_{sig} lepton daughter's muon PID level.

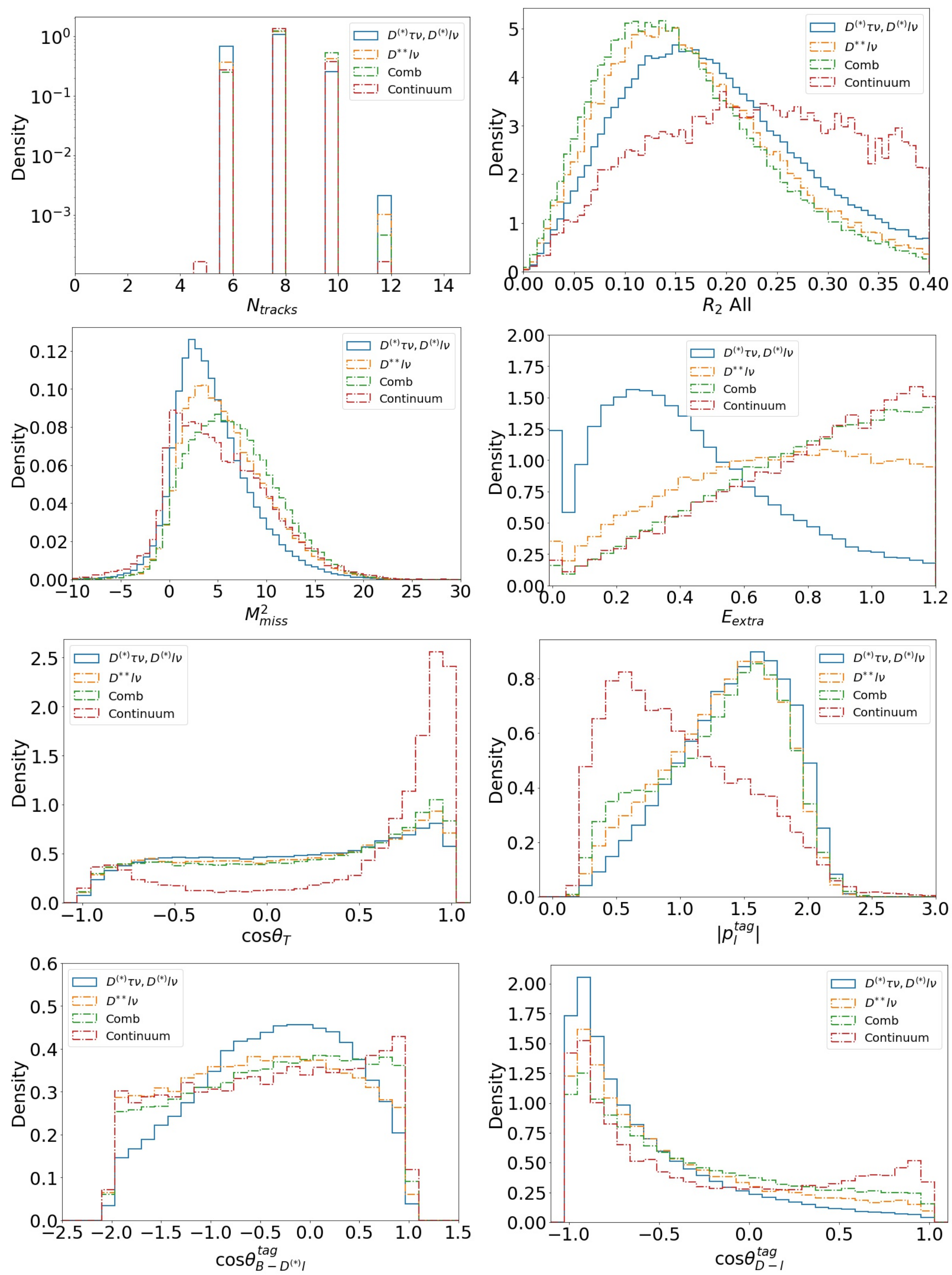

Figure 7.5: Histograms of variables used for the C_1 classifier.

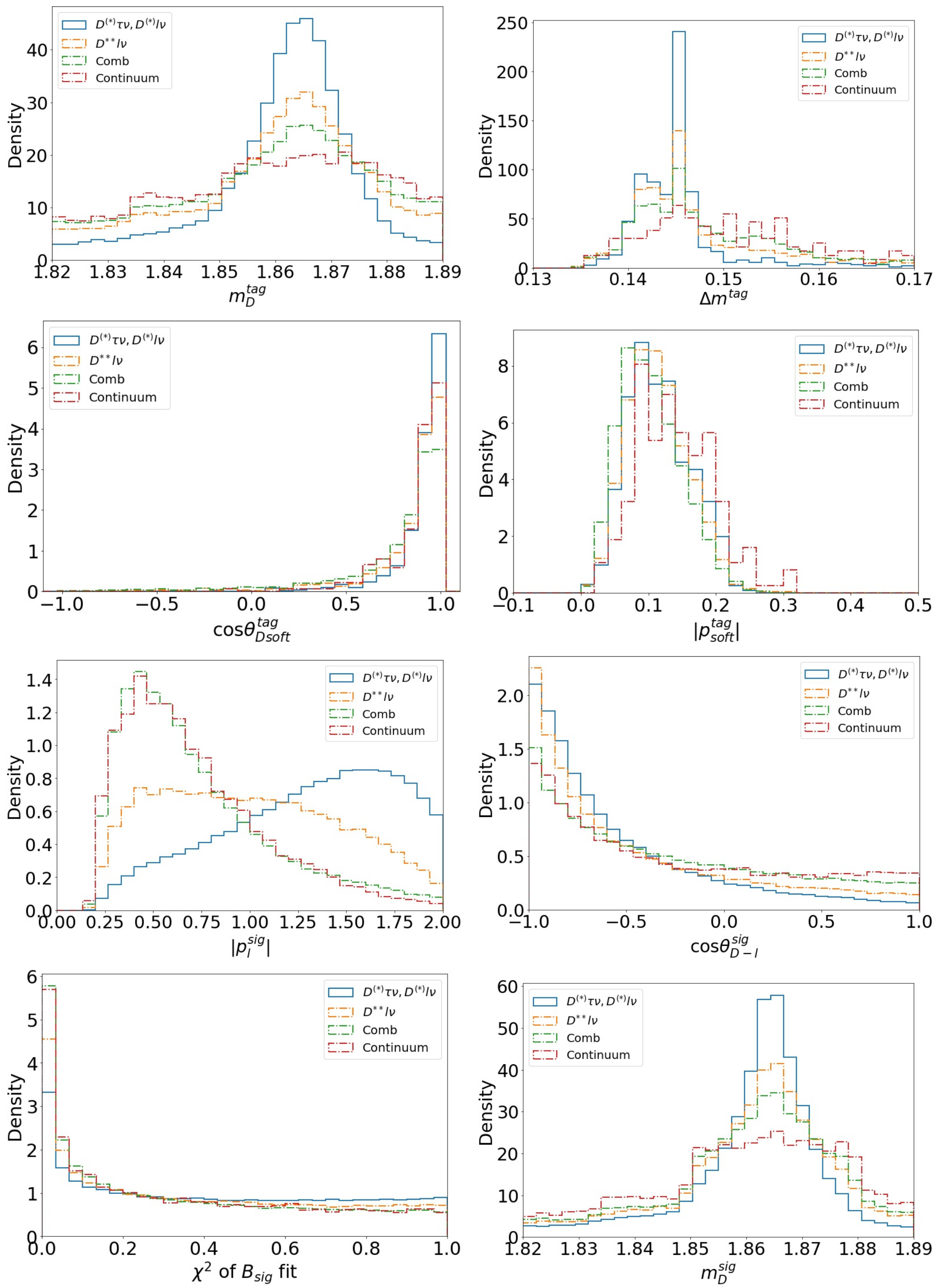

Figure 7.6: Histograms of variables used for the C_1 classifier.

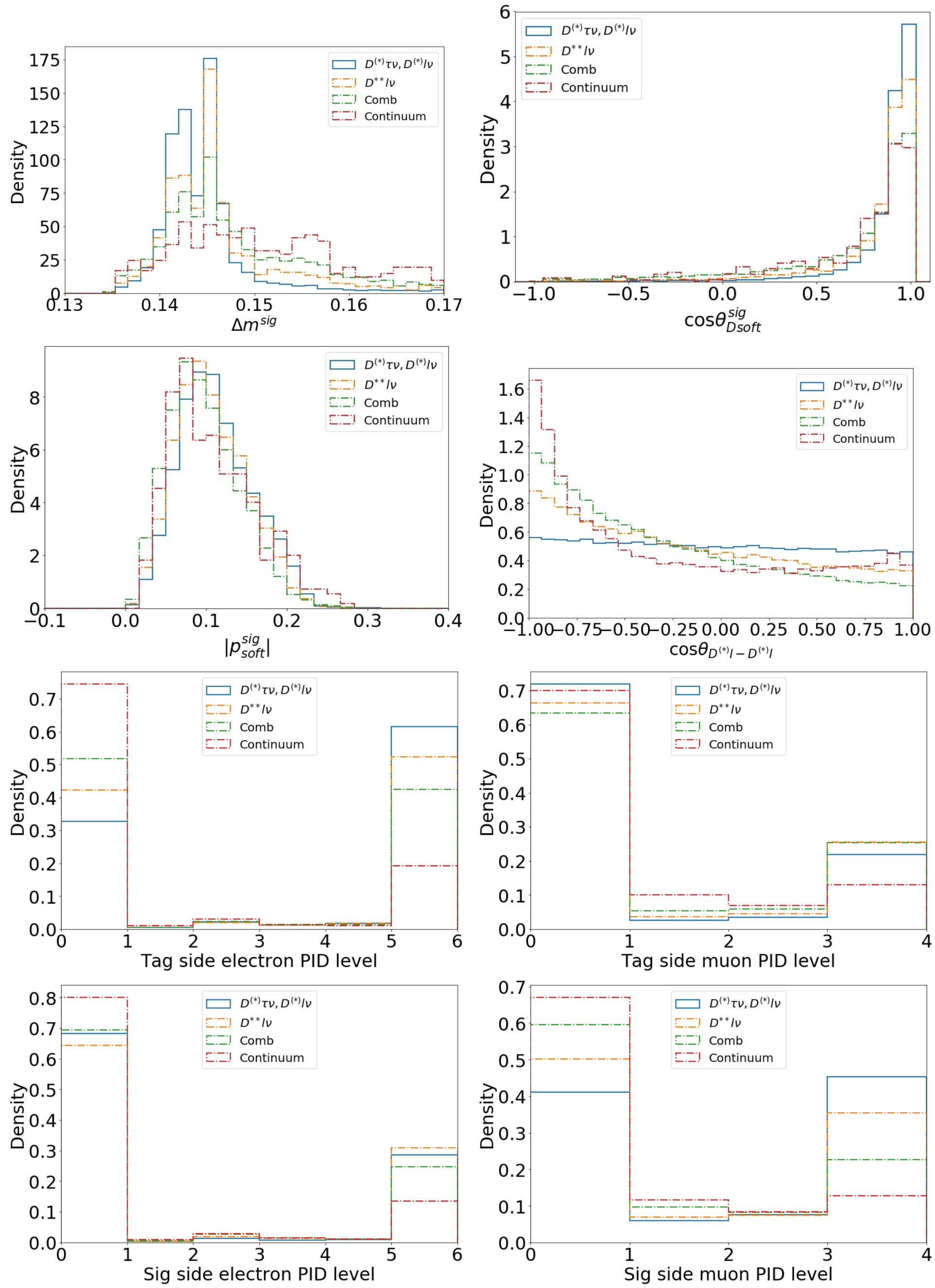

Figure 7.7: Histograms of variables used for the C_1 classifier.

We use a Gradient Boosting Decision Tree (BDT) [36] for the C_1 classifier. The number of decision trees ranges from 20 to 600. The model is implemented using the `scikit-learn` package. The metric used to evaluate the classification performance is the area under the ROC curve. The higher the score, the higher the classification power. Fig. 7.8 shows the relationship between the area under ROC curve and the number of trees. The classification performance becomes stable when the number of decision trees is above 500. We use a BDT with 600 trees as the final model. Fig. 7.9 shows the importance of variables for classification. E_{extra} and $|p_l^{sig}|$ are the most powerful variables to identify $B \to D^{(*)}(\tau/l)\nu$ from all types of backgrounds. For E_{extra}, the signal and normalization events usually have near-zero extra neutral energy, while background events have a wide distribution. The normalization decay has an energetic lepton produced by the D decay, leading to a higher $|p_l^{sig}|$ value than signal events, as well as all types of backgrounds. The output of the BDT score p is transformed using a logit function:

$$z_1 = \text{logit}(p) = \log \frac{p}{1-p}.$$

The z_1 distribution for all types of events is shown in Fig. 7.10. Signal and normalization events tend to have higher z_1 values than backgrounds.

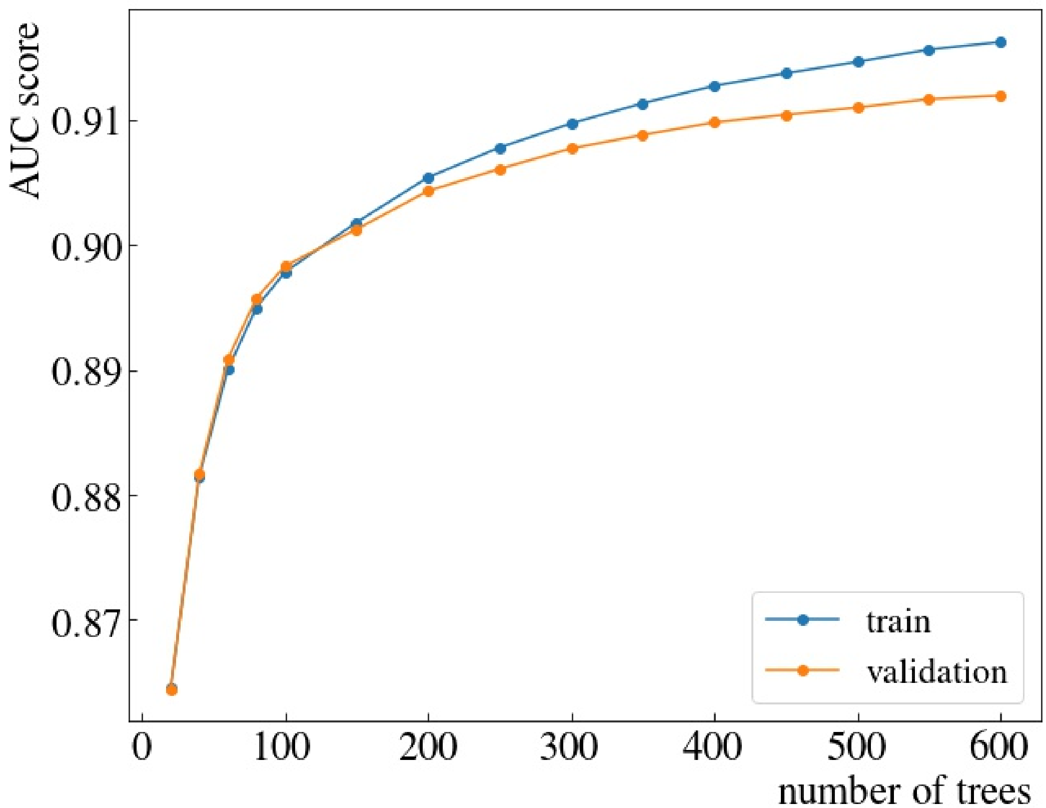

Figure 7.8: Area under the ROC curve for BDT classifiers with different numbers of trees.

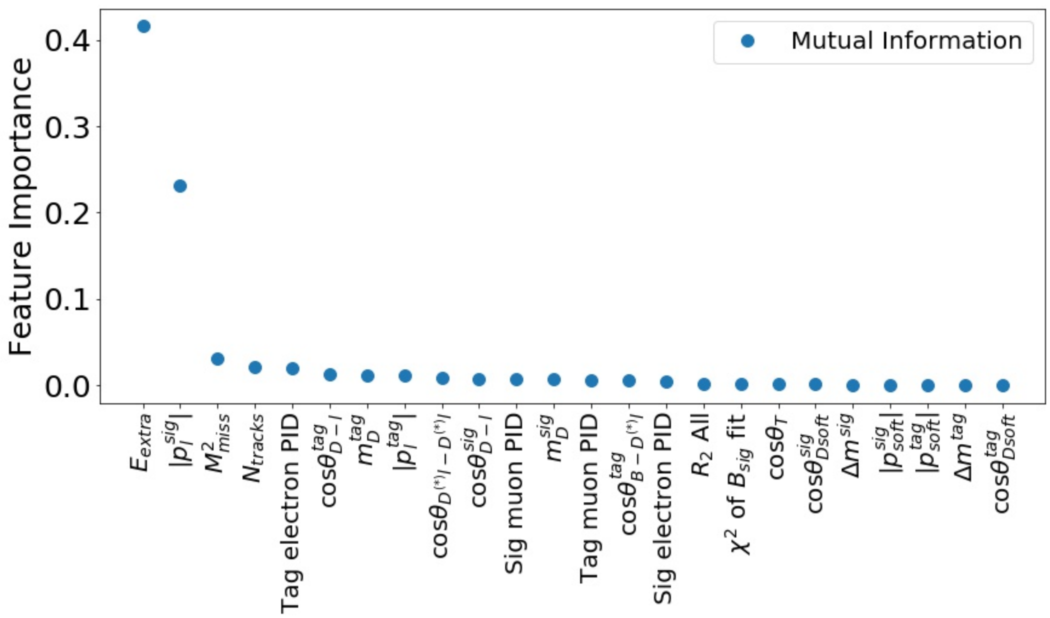

Figure 7.9: Importance of each variable for learning the C_1 classifier.

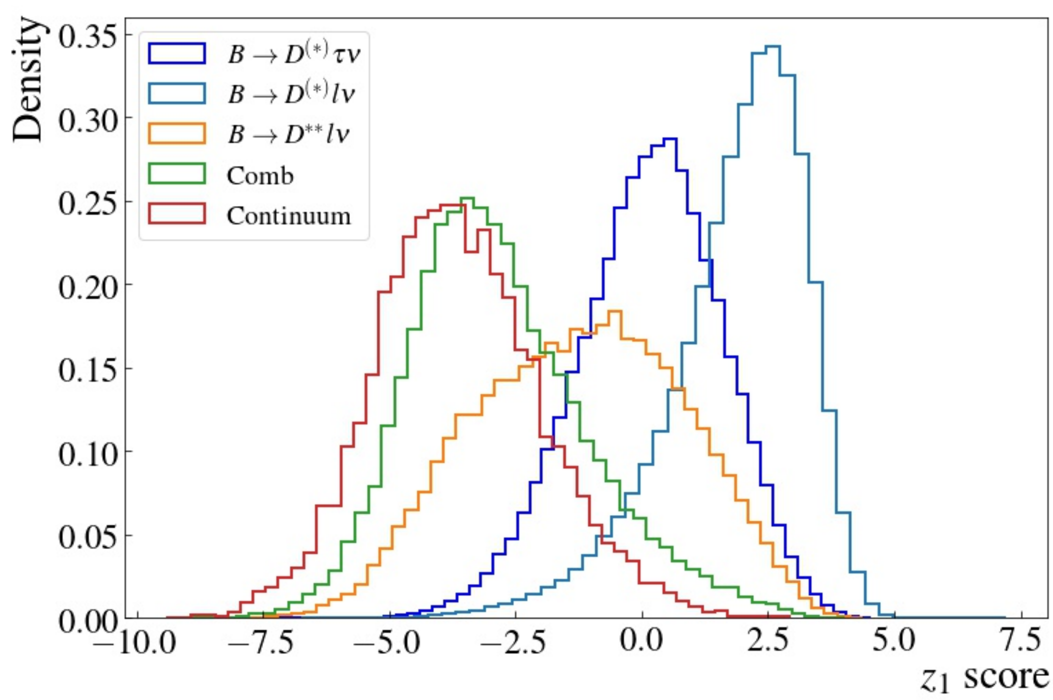

Figure 7.10: z_1 distribution for signal, normalization, $D^{**}l\nu$, $B\bar{B}$ combinatorial, and continuum events.

C_2 classifier

The C_2 Classifier aims to classify signal events and normalization events. Similar to the C_1 classifier, we divide the sample into training and validation samples. The training sample is first used to train the classifier, the validation sample is then applied to evaluate the performance of the classifiers. Signal events are labeled positive and normalization events are labeled negative before training. We use the same variables used for the C_1 classifier, with the addition of the following quantity:

- $\cos\theta^{sig}_{B-D^{(*)}l}$: Cosine of the angle between the 3-momentum of the B_{sig} and the 3-momentum sum of its D and lepton daughters.

The histogram of these variables for signal and normalization events are shown in Fig. 7.11.

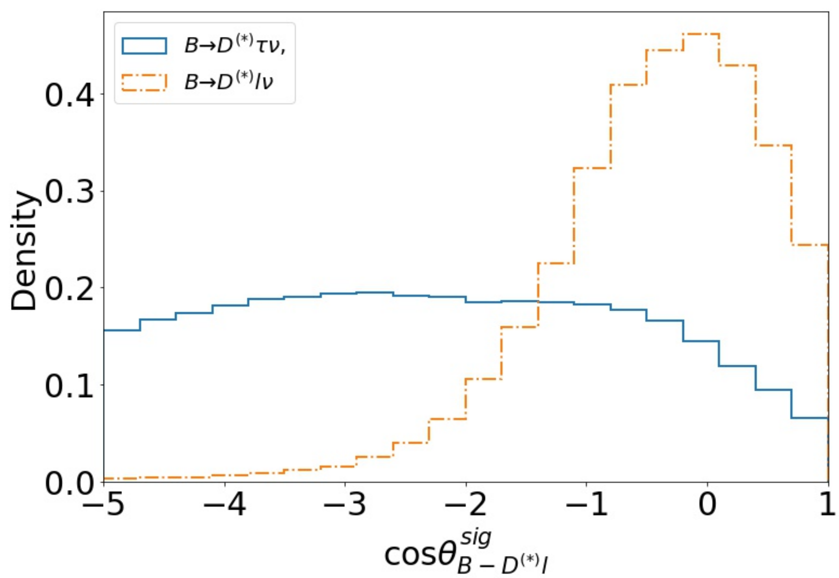

Figure 7.11: Histograms of variables used for the C_2 classifier.

Similar to the C_1 classifier, we use a BDT for the C_2 classifier. The number of decision trees ranges from 100 to 600. The classification performance is stable when the number of decision trees is above 100, as shown in Fig. 7.12. We use a BDT with 600 trees as the final model. Fig. 7.13 shows the importance of the variables for classification: $\cos\theta^{sig}_{B-D^{(*)}l}$ and $|p^{sig}_l|$ are the most powerful variables to distinguish signal from normalization events. For $\cos\theta^{sig}_{B-D^{(*)}l}$, normalization events have only a single neutrino, and the value should be from -1 to 1. However, signal events tend to have more negative values, due to the presence of three neutrinos in the final state. The e/μ produced from the τ decays in signal events have a softer $|p_l^{sig}|$ spectrum than the leptons produced from D decays for normalization events.

The output of the BDT score is transformed to z_2 using a logit function. The z_2 distribution for all types of events is shown in Fig. 7.14. Signal events tend to have a higher z_2 score, while normalization events tend to have a lower z_2 score. Backgrounds have z_2 scores in between.

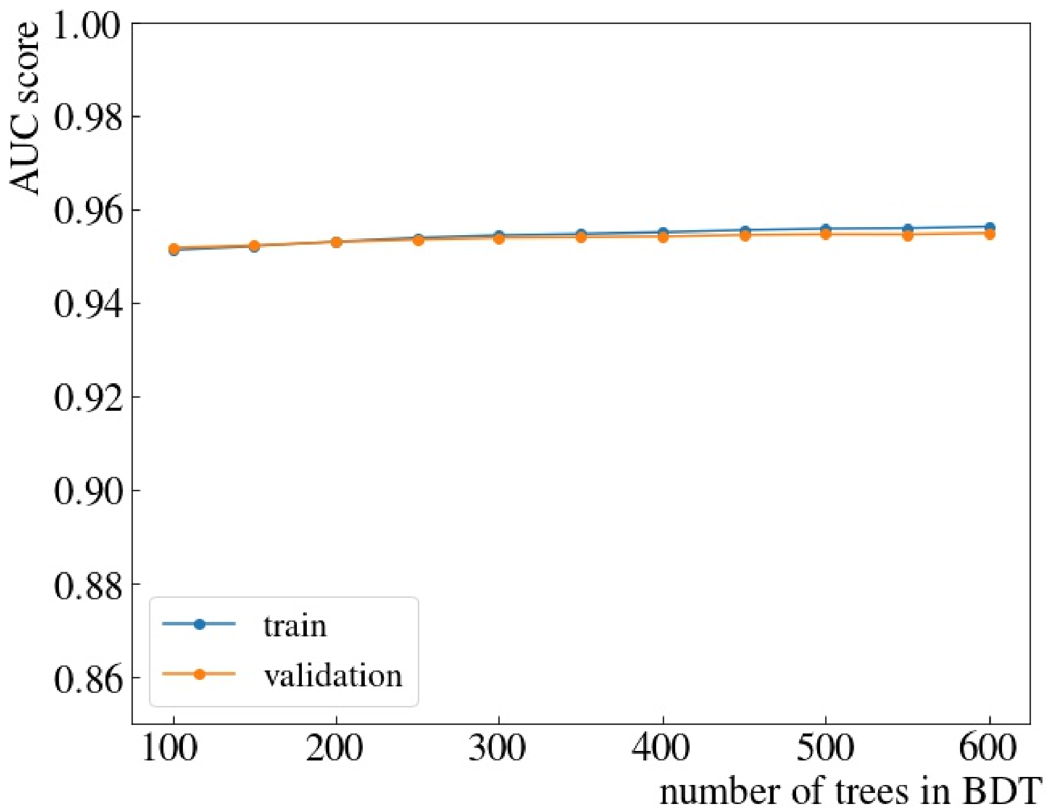

Figure 7.12: The area under the ROC curve for BDT classifiers with different number of trees.

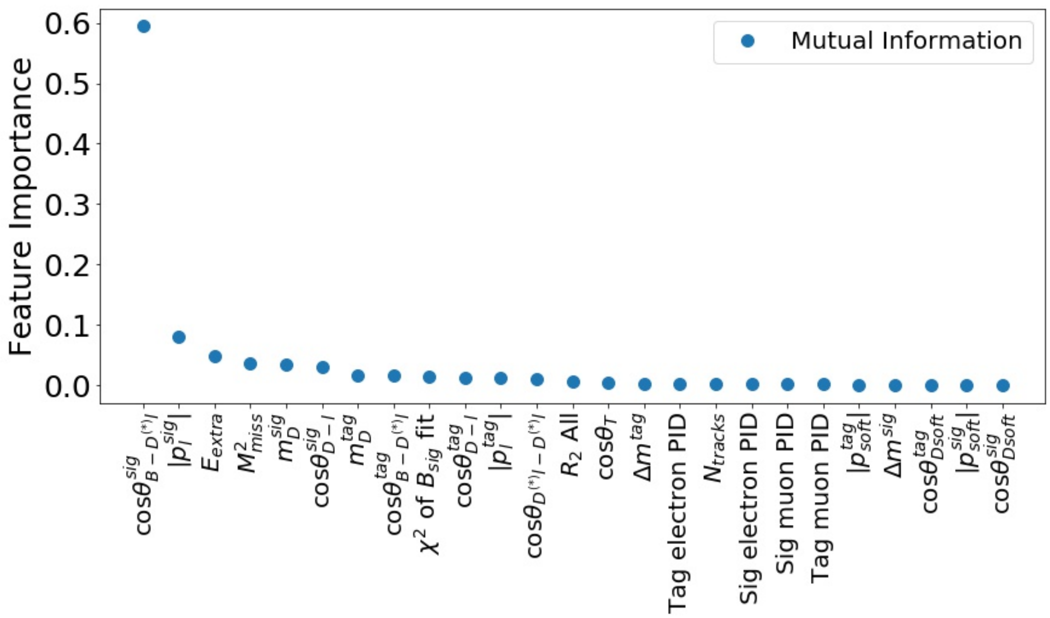

Figure 7.13: Importance of each variable for learning the C_2 classifier.

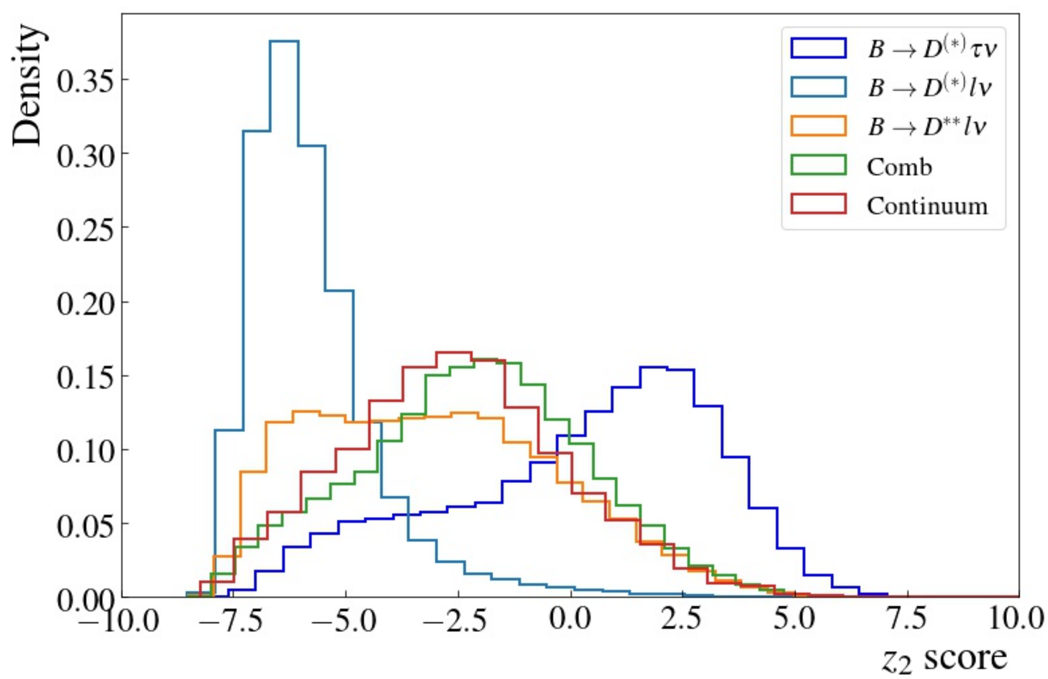

Figure 7.14: z_2 distribution for signal, normalization, $D^{**}l\nu$, $B\bar{B}$ combinatorial, and continuum events.

7.6 Signal Extraction

In this section we describe how the signal yields are estimated. Our procedure includes two steps, and is applied to each subset individually to extract the corresponding signal yields:

1. Density Estimation: estimate the 2-dimensional density $f(z_1, z_2)$ for each type of event.

2. Maximum Likelihood Estimation: estimate signal yields based on estimated densities, by solving a convex optimization problem.

Density estimation

Suppose a random variable X follows an unknown distribution f. If we have N observations of X, the objective of the density estimation procedure is to estimate f based on the N observations. Many approaches have been developed, both parametric and non-parametric, to solve this problem. One popular non-parametric approach is histogramming. Histogramming divides the range of X into m equally-sized bins, and assigns each observation to one of these bins. The normalized counts of each bin gives us the estimated density f.

Kernel density estimation (KDE) [37] is a non-parametric approach, which estimates f via data smoothing. Given N observations $\{x_i\}_{i=1}^N$ of X, the KDE estimate of f is:

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K(\frac{x - x_i}{h}),$$

where K is a kernel function, h is the bandwidth parameter to control the level of smoothing. General choices of kernel function can be Gaussian ($K(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}$) or Epanechnikov ($K(v) = \frac{3}{4}(1 - v^2)$), though the kernel function K can be any nonnegative bounded function satisfying

1. $\int_{-1}^{1} K(v)dv = 1$

2. $K(v) = K(-v)$

3. $\int_{-1}^{1} v^2 K(v)dv < \infty$.

The choice of bandwidth parameter h involves minimizing the mean squared error (MISE) between the estimated $\hat{f}$ and f:

$$MISE = \int (\hat{f} - f)^2 = \int \hat{f}^2 - 2 \int \hat{f} \cdot f + \int f^2,$$

which can be evaluated by summing over the observed data points of X. Practically, we use cross validation to choose the optimal bandwidth, which gives the same asymptotic accuracy:

$$CV(h) = \int \hat{f}^2 - 2N^{-1} \sum_i \hat{f}_{-i}(x_i),$$

where $\hat{f}_{-i}(x_i) = \frac{1}{Nh} \sum_{i \neq j} K(\frac{x_i - x_j}{h})$ is used to evaluate MISE by removing the contribution from point i.

In this analysis, we use an adaptive KDE to evaluate the 2-dimensional probability density function (PDF) of each event type. Adaptive KDE changes the bandwidth parameters depending on the densities. In lower data density regions, adaptive KDE chooses wider bandwidths to reduce the effect of outliers on the overall KDE, and provides better overall performance. The adaptive KDE of f is:

$$\begin{aligned}
\hat{f}(x) &= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h_i} K(\frac{x - x_i}{h_i}) \\
h_i &= h \times \lambda_i \\
\lambda_i &= (\frac{G}{\tilde{f}(x_i)})^{\alpha},
\end{aligned} \tag{7.6}$$

where $\tilde{f}(x_i)$ is a pilot estimate of $f(x_i)$, λ_i is local bandwidth factor to control the local smoothing, G is the geometric mean of $f(x_i)$, α is a sensitivity parameter between 0 and 1. We normally set its values to be 0.5.

We use an Epanechnikov function as the kernel function, and use grid search to find the optimal bandwidth h. The KDE using an Epanechnikov kernel is claimed to have better mean integrated squared error under the same amount of data, and is one of the popular kernels used in practice. We use the `bbrcit_kde` library to reduce the time complexity of performing KDE from $O(n^2)$ to $O(n)$. It applies kd-tree [38] algorithm and applies GPU computation to speedup the KDE computation. The learned densities for all event components for each of four subsets are shown in Fig. 7.15, 7.16, 7.17, and 7.18.

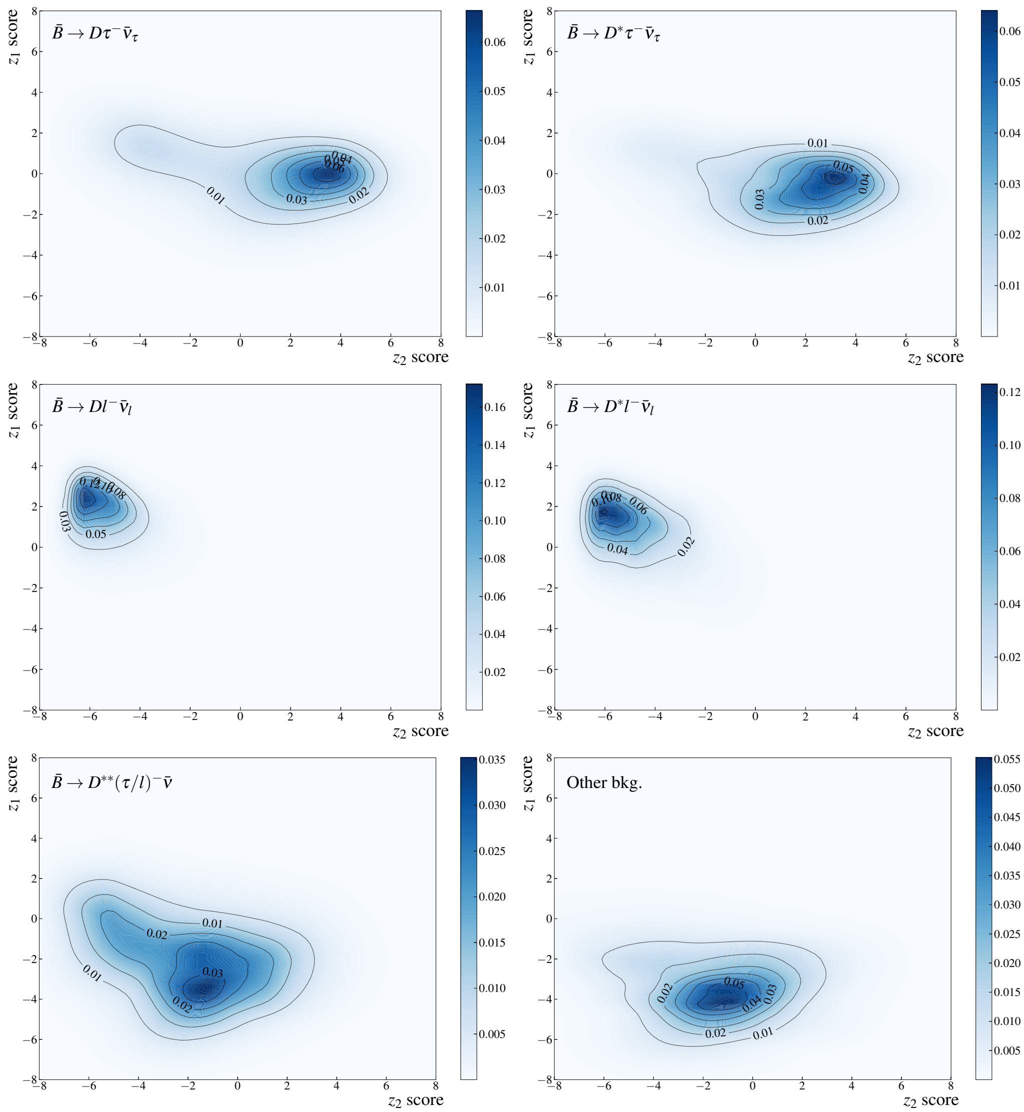

Figure 7.15: KDE learned densities for event components in the D^+l subset.

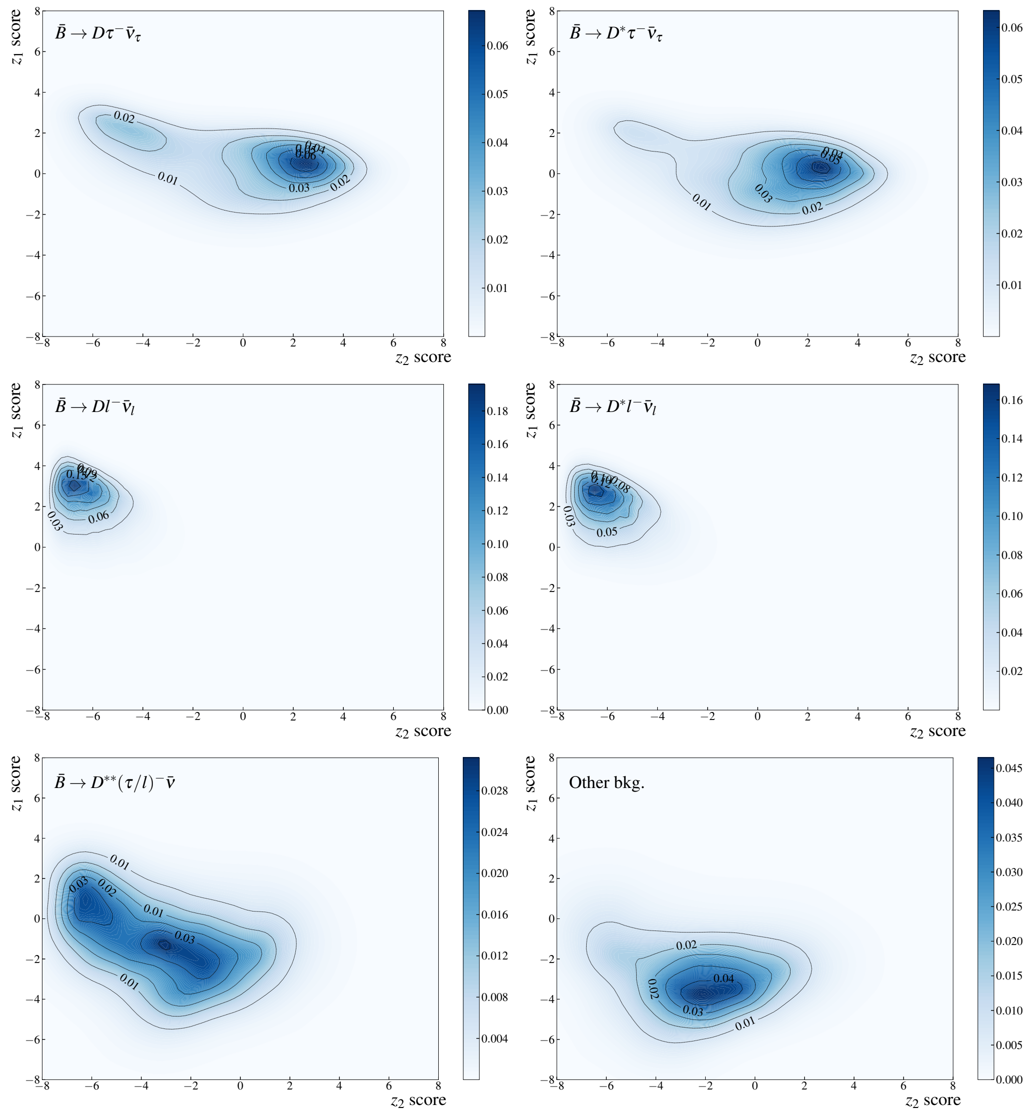

Figure 7.16: KDE learned densities for event components in the $D^0 l$ subset.

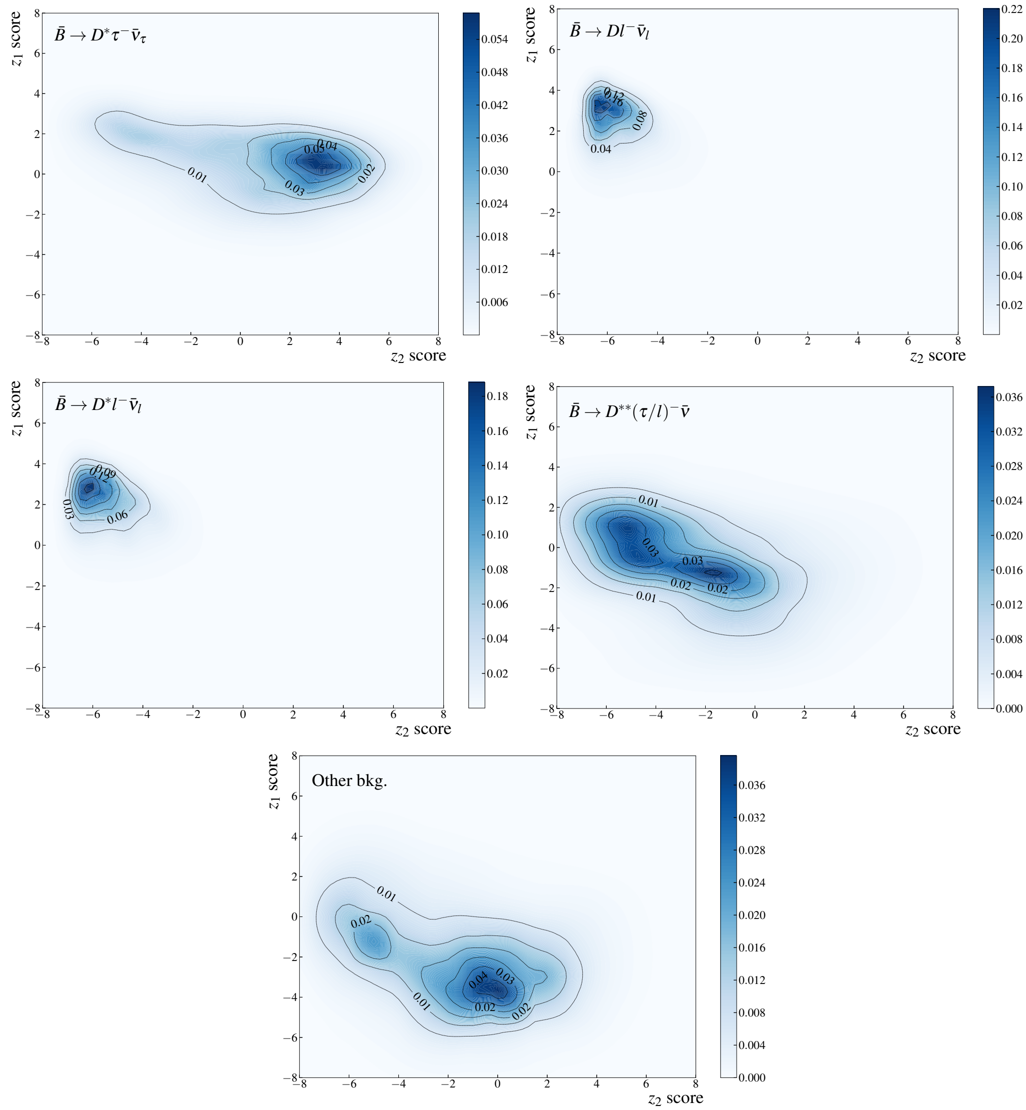

Figure 7.17: KDE learned densities for event components in the $D^{*+}l$ subset.

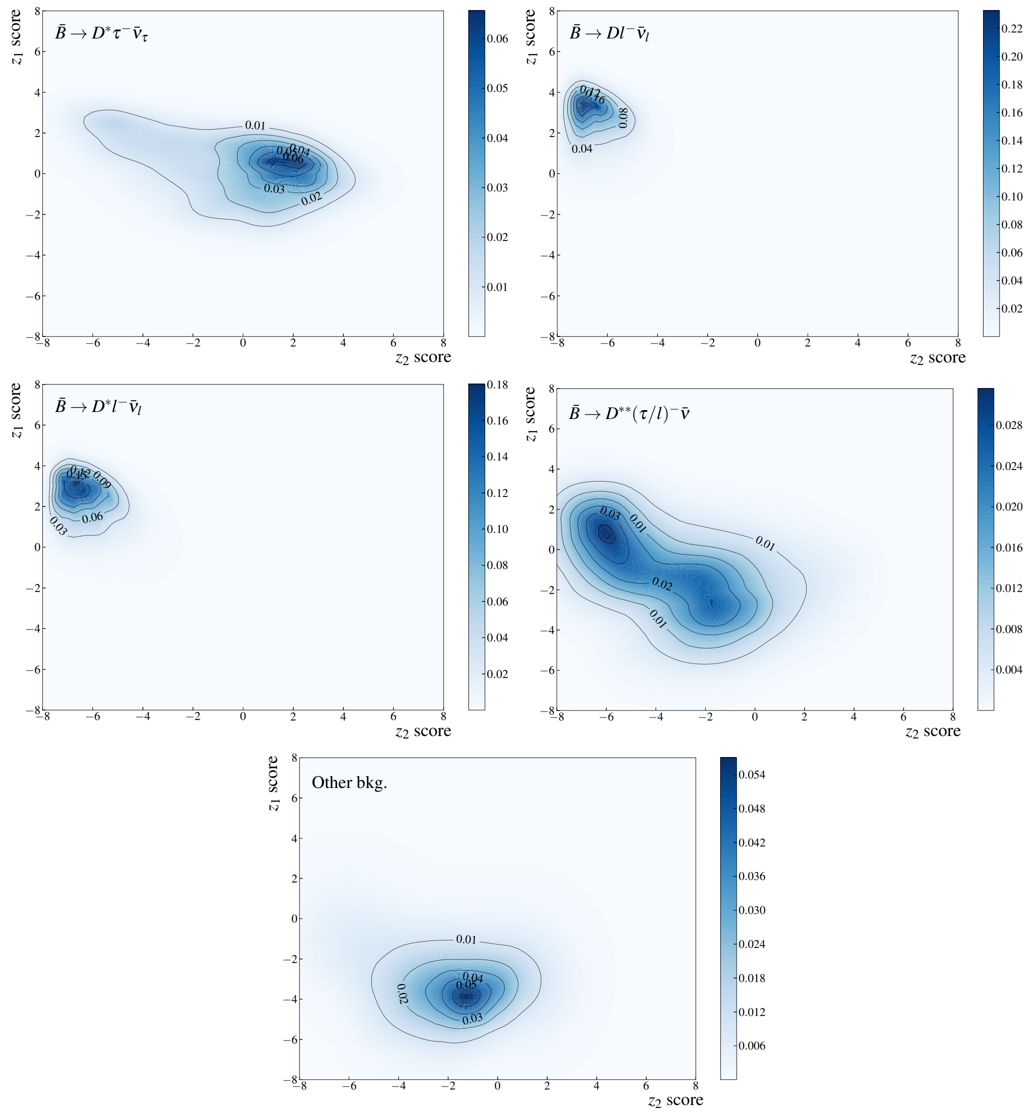

Figure 7.18: KDE learned densities for event components in the $D^{*0}l$ subset.

Maximum likelihood estimation

We extract the signal and normalization yields from each subset (*channel_label*) individually, and then combine them to get the overall yields. Given a subset with C components and the corresponding learned densities $f_j(z_1, z_2)$ $(j = 1, 2, ..., C)$, the $\mathbf{z} = (z_1, z_2)$ distribution of the subset can be written as

$$\text{Prob}(\mathbf{z}) = \sum_{j=1}^{C} p_j f_j(\mathbf{z}),$$

where p_j is the proportion of j's component. If we observe N events in the subset $\{\mathbf{z}\}_{i=1}^{N}$, the likelihood is

$$L = \prod_{i=1}^{N} \text{Prob}(\mathbf{z}_i) = \prod_{i=1}^{N} \left(\sum_{j=1}^{C} p_j \cdot f_j(\mathbf{z}_i) \right).$$

The MLE estimate of component proportions p_j $(j = 1, 2, ..., C)$ can be obtained by solving the following convex optimization problem:

$$\begin{aligned} &\max_{p_1, p_2, \ldots, p_C} \prod_{i=1}^{N} \left(\sum_{j=1}^{C} p_j \cdot f_j(\mathbf{z}_i) \right) \\ &\text{s.t.} \sum_{j} p_j = 1. \end{aligned} \tag{7.7}$$

We use the CVXOPT [39] python package to implement the above estimation. The procedure is applied to the generic MC to evaluate its performance. To estimate the statistical sensitivity, we bootstrap the generic MC 900 times and solve the MLE for each bootstrap sample. The standard deviations of the 900 estimated proportions are used as absolute statistical uncertainties for the MLE estimation. The extracted yields and the absolute statistical uncertainties for all the four subsets are listed in Table 7.9. The projected fitting results on z_1 and z_2 scores are shown in Fig. 7.19 and Fig. 7.20.

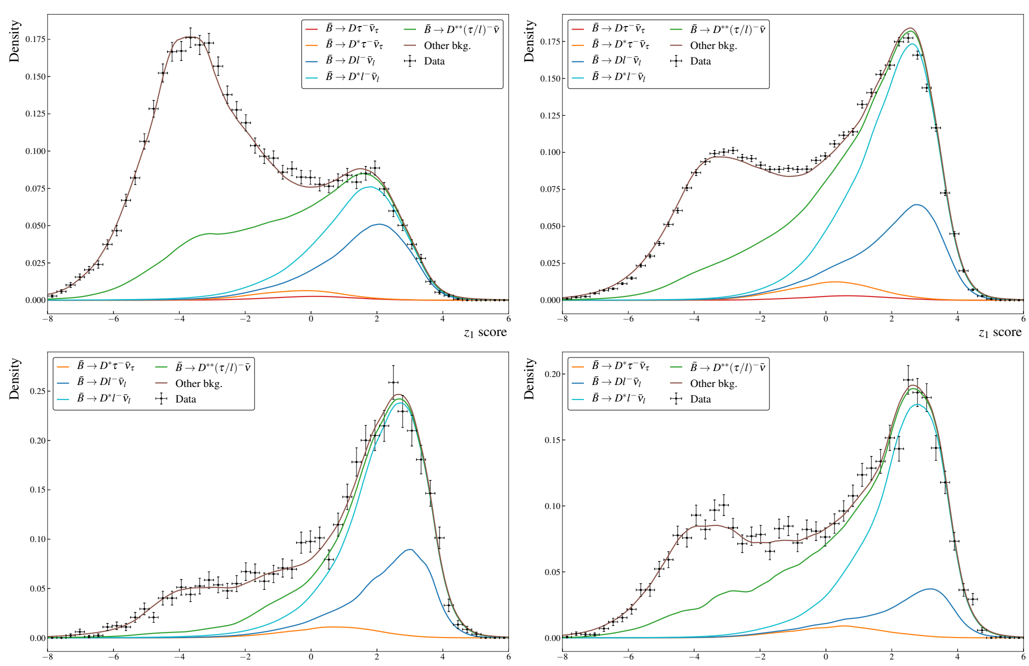

Figure 7.19: Comparison of z_1 score distributions of the data with the projections of fit results for (upper left) D^+l, (upper right) D^0l, (bottom left) $D^{*+}l$, and (bottom right) $D^{*0}l$ subsets.

Subset	Component	Extracted Yield
D^+l	$B \to D\tau\nu$	117 ± 44
	$B \to D^*\tau\nu$	191 ± 57
	$B \to Dl\nu$	1965 ± 110
	$B \to D^*l\nu$	1261 ± 146
	$B \to D^{**}l\nu$	3248 ± 233
	Other Bkgs	7258 ± 182
D^0l	$B \to D\tau\nu$	726 ± 273
	$B \to D^*\tau\nu$	2452 ± 285
	$B \to Dl\nu$	12012 ± 555
	$B \to D^*l\nu$	22808 ± 648
	$B \to D^{**}l\nu$	15594 ± 445
	Other Bkgs	22065 ± 307
$D^{*+}l$	$B \to D^*\tau\nu$	110 ± 17
	$B \to Dl\nu$	614 ± 127
	$B \to D^*l\nu$	1246 ± 139
	$B \to D^{**}l\nu$	292 ± 54
	Other Bkgs	665 ± 42
$D^{*0}l$	$B \to D^*\tau\nu$	175 ± 25
	$B \to Dl\nu$	479 ± 126
	$B \to D^*l\nu$	2154 ± 149
	$B \to D^{**}l\nu$	1387 ± 101
	Other Bkgs	1415 ± 70

Table 7.9: Fit results for the yields of all the components in four subsets.

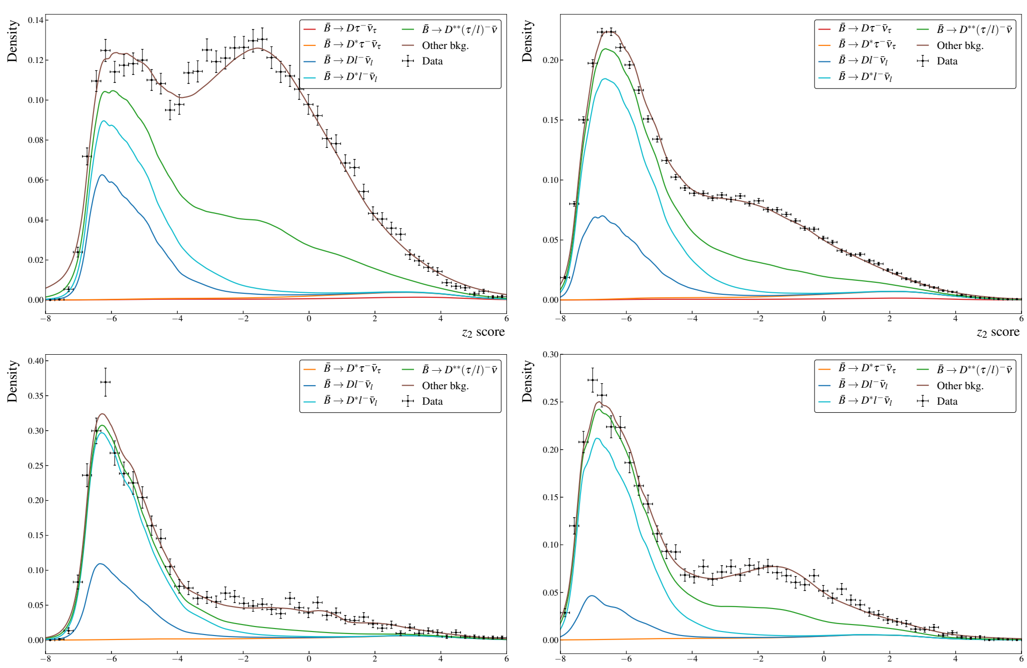

Figure 7.20: Comparison of z_2 score distributions of the data with the projections of fit results for (upper left) D^+l, (upper right) D^0l, (bottom left) $D^{*+}l$, and (bottom right) $D^{*0}l$ subsets.

Cross-check on PDFs modeling

Cross-checks are performed to study the modeling of PDF shapes. We first check the PDF modeling of normalization events by comparing the PDFs of MC with data on the normalization enriched region. The normalization enriched region is defined in (z_1, z_2) space as $z_1 > 2, z_2 < -4$, in which 94% events are normalization decays. The MC and data comparison shown in Fig. 7.21 indicates the normalization PDF shapes are well modeled.

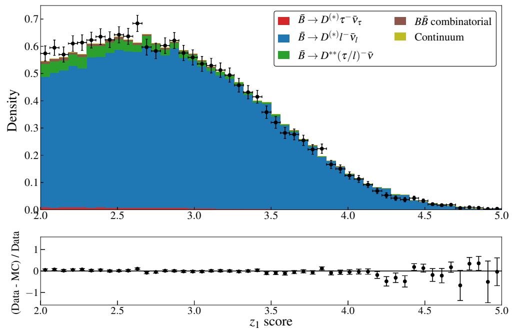

Figure 7.21: MC/data comparison of z_1 and z_2 score distributions in normalization enriched region.

We then check the PDF modeling of backgrounds (after on-peak background calibration) by comparing the PDFs of MC with data on background enriched region. The background enriched region is defined as $-4 \leq z_1 \leq -2$, in which 97% events are either $B \to D^{**}l\nu$, $B\bar{B}$ combinatorial, or continuum events. The MC and data comparison shown in Fig. 7.22 indicates the PDF shapes of backgrounds are also well-modeled.

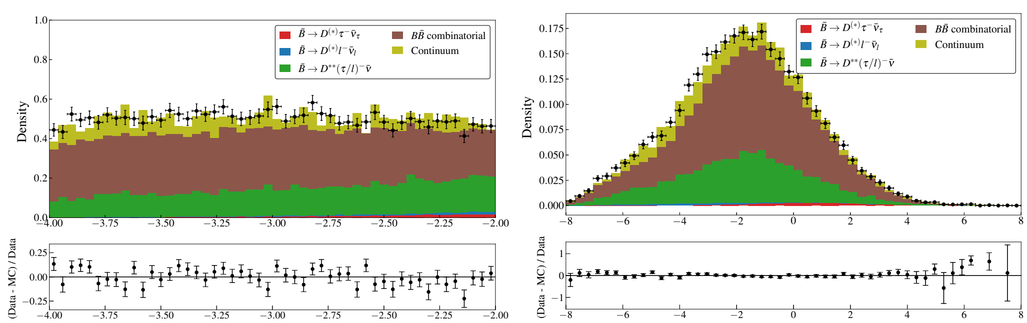

Figure 7.22: MC/data comparison of z_1 and z_2 score distributions in backgrounds enriched region.

7.7 Systematic Uncertainties

Procedure

Many quantities are required to precisely measure $\mathcal{R}(D^{(*)})$, for instance, the PDF shape of each event type, the physical parameter setup for the Monte Carlo simulations, and the detector efficiencies. We use our best knowledge of this information during the measurement, however, this information has associated uncertainties. Consequently, our measurement might be biased due to these sources of uncertainty. This section describes the systematic uncertainties that are important in this analysis, and how we evaluate them.

Different procedures are developed to evaluate each source of uncertainty. Uncertainties from form factors, branching fractions of $\bar{B} \to D^{**}(\tau/l)^{-}\bar{\nu}$, $b \to c\bar{c}$ decays, and D meson decays are evaluated based on the delta method [40]. We fluctuate every uncertain parameter up and down according to their 1σ uncertainties. Each time the parameter is changed, we re-evaluate the kernel densities and repeat the fit with the new PDFs. The changes in $\mathcal{R}(D^{(*)})$ are quoted as the corresponding systematic uncertainty. Uncertainties due to the limited size of the MC sample and background calibration are evaluated using a bootstrap algorithm [41].

$B \to D^{(*)}l\nu$ form factors

One source of uncertainty comes from the model behind the Monte Carlo simulation of B meson decays, in which form factors are required. However, the form factors used for the MC simulation have associated uncertainties, which contribute to the differential decay rate of $B \to D\tau\nu$ through angular distributions, and therefore the shapes of the PDFs. To transform the uncertainties on the form factors to $R(D^{(*)})$, we first transform the default model i, which is used to generate simulation events, to an updated parameterization model j, and use model j to evaluate the corresponding systematics.

The default form factor model used to simulate $B \to Dl\nu_l, l = e, \mu, \tau$ events is the

CLN [24] model. It uses dispersion relations to expand the form factors [42]

$$\begin{aligned}
\frac{V_1(\omega)}{V_1(1)} &= 1 - 8\rho^2 z + (51\rho^2 - 10)z^2 - (262\rho^2 - 84)z^3 + O(z^4) \\
\frac{S_1(\omega)}{V_1(\omega)} &= (1 + \Delta(\omega)) \\
\rho^2 &= 1.186 \pm 0.054 \\
V_1 &= 1.0816 \\
\Delta &= 1.0.
\end{aligned} \tag{7.8}$$

The expansions of the CLN model used to simulate $B \to D^* l\nu, l = e, \mu, \tau$ and baseline parameter settings are::

$$\begin{aligned}
\frac{h_{A_1}(\omega)}{F_1} &= 1 - 8\rho^2 z + (53\rho^2 - 15)z^2 - (231\rho^2 - 91)z^3 + O(z^4) \\
R_0(\omega) &= R_0(1) - 0.11(\omega - 1) + 0.01(\omega - 1)^2 + O(\omega^3) \\
R_1(\omega) &= R_1(1) - 0.12(\omega - 1) + 0.05(\omega - 1)^2 + O(\omega^3) \\
R_2(\omega) &= R_2(1) + 0.11(\omega - 1) - 0.06(\omega - 1)^2 + O(\omega^3) \\
F_1 &= 0.921 \\
\rho^2 &= 1.207 \pm 0.026 \\
R_0 &= 1.14 \\
R_1 &= 1.401 \pm 0.033 \\
R_2 &= 0.854 \pm 0.02.
\end{aligned} \tag{7.9}$$

To evaluate the systematic uncertainty due to the $B \to D^{(*)}(l/\tau)\nu_l$ form factors, we measure the central values of $R(D^{(*)})$ using the world average values of parameters in form factor model. Then we change the form factor parameters with the corresponding uncertainties by $\pm\sigma$. The difference between the values obtained before and after fluctuation are listed as systematic. The results are shown in Table 7.10.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
$B \to D(l/\tau)\nu$ form factor	0.42	0.13
$B \to D^*(l/\tau)\nu$ form factor	0.92	0.31

Table 7.10: Evaluated relative systematic uncertainties from the $B \to Dl\nu$ and $B \to D^* l\nu$ form factors.

$B \to D^{**}(l/\tau)\nu_l$ form factors

The systematic uncertainties from semileptonic B meson decays involving D^{**} are also evaluated, where D^{**} denotes the $L = 1$ excitations of the ground state D meson: D_0^*, D_1, D_1', and D_2^*. In this case, we use the LLSW B1 model [43] as baseline, fluctuate it to the LLSW B2 [43] model and take the difference between the two models as systematic uncertainties. The resulting systematic uncertainty is shown in Table 7.11.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
$B \to D^{**}(l/\tau)\nu$ form factor	0.48	0.18

Table 7.11: Evaluated relative systematic uncertainties from the $B \to D^{**}l\nu$ form factors.

$B \to D^{(*)}l\nu$ branching fractions

Branching fractions from semileptonic B meson decays are another source of systematic uncertainties, as they affect the relative abundance of the decays and thus the density distributions of event components, as well as the estimate of signal and normalization efficiencies. Some of the branching fractions used for the MC generation are different than the current world average. Therefore, before evaluating the systematic uncertainties associated with the branching fractions, we assign correction factors to re-weight the MC events:

$$\omega(x) = \frac{\omega_{W.A.}(x)}{\omega_{\texttt{DECAY.DEC}}(x)},$$

where $\omega(x)$ is the assigned weight for event x, $\omega_{W.A.}$ is the world average value of event x's branching fraction, and $\omega_{\texttt{DECAY.DEC}}(x)$ is the value used in the MC simulation.

The $B \to D^{(*)}l\nu$ decays are one of the dominant B meson decays in this analysis; we corrected and fluctuated their uncertainties to evaluate the systematic uncertainties according to the values listed in Table 7.12. The resulting systematic uncertainties are listed in Table 7.13.

$B \to D^{**}(l/\tau)\nu$ branching fractions

Semileptonic decays of $B \to D^{**}l\nu_l$ and $B \to D^{**}\tau\nu_\tau$ are an important background as they have topologies to the signal events. In general, D^{**} is defined as any excited

Decay Mode	MC Simulation Value	World Average Value	Uncertainty
$B^+ \to \overline{D}^{*0}\mu^+\nu_\mu$	0.0617	0.0566	0.0022
$B^+ \to \overline{D}^{*0}e^+\nu_e$	0.0617	0.0566	0.0022
$B^+ \to \overline{D}^0\mu^+\nu_\mu$	0.0224	0.0235	0.0009
$B^+ \to \overline{D}^0 e^+\nu_e$	0.0224	0.0235	0.0009
$B^0 \to D^{*-}\mu^+\nu_\mu$	0.057	0.0506	0.0012
$B^0 \to D^{*-}e^+\nu_e$	0.057	0.0506	0.0012
$B^0 \to D^-\mu^+\nu_\mu$	0.0207	0.0231	0.0010
$B^0 \to D^-e^+\nu_e$	0.0207	0.0231	0.0010

Table 7.12: Decay modes fluctuated to evaluate the $B \to D^{(*)}l\nu$ branching fractions on generic MC. The world average values are from [34].

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
$B \to D^{(*)}l\nu$ branching fraction	0.47	0.38

Table 7.13: Evaluated relative systematic uncertainties from $B \to D^{(*)}l\nu$ branching fractions on MC.

charmed meson states that is not in the ground-state $1S$ doublet [44]. In this analysis, we consider the following three D^{**} decay types:

- Resonant $D^{**}(1P)$ state, which includes the four lightest excited charmed meson states $D_0^*(0^+)$, $D_1(1^+)$, $D_1'(1^+)$ and $D_2^*(2^+)$.
- Resonant $D^{**}(2S)$ state, the radially-excited modes of $D^{(*)}$.
- Non-resonant D^{**} states. We allow unmeasured D^{**} states whose final states consist of a $D^{(*)}$ and one pion.

We estimate the systematic uncertainties from both $B \to D^{**}l\nu_l$ and $B \to D^{**}\tau\nu_\tau$ for these three types of D^{**} decays. We do not have MC samples for some $B \to D^{**}\tau\nu_\tau$ decays, so we use a 3-body phase space model for estimation. Table 7.14 shows all the decay modes taken into consideration. We do not consider other charmed states heavier than $D^{(*)}(2S)$ as their smaller phase space suppresses the branching fractions.

	Resonant $D^{**}(1P)$	Resonant $D^{**}(2S)$	Non-resonant $B \rightarrow D^{(*)}\pi l \nu_l$
$B \rightarrow D^{**} l \nu_l$	MC	MC	MC
$B \rightarrow D^{**} \tau \nu_\tau$	MC	Phase Space	Phase Space

Table 7.14: $B \rightarrow D^{**}(D^{(*)}\pi)(l/\tau)\nu$ decays and method used to estimate the systematic uncertainties. "MC" means the uncertainties can be directly estimated using MC sample. "Phase Space" means we do not have the corresponding MC samples, therefore a phase-space-based estimation is applied.

To evaluate the resonant $D^{**}(1P)$ for $B \rightarrow D^{**}(l/\tau)\nu$ decays, the following decays in the generic MC sample are fluctuated as listed in Table 7.15. The corresponding systematic uncertainties are obtained in Table 7.16.

Decay Mode	MC Simulation Value	Setup Value	Uncertainty
$B^+ \to \overline{D}_1^0 e^+ \nu_e$	0.0056	0.0096	0.001
$B^+ \to \overline{D}_0^{*0} e^+ \nu_e$	0.0049	0.0044	0.0008
$B^+ \to \overline{D}_2^{*0} e^+ \nu_e$	0.003	0.003	0.0004
$B^+ \to \overline{D}_1^{\prime 0} e^+ \nu_e$	0.009	0.002	0.0005
$B^+ \to \overline{D}_1^0 \mu^+ \nu_\mu$	0.0056	0.0096	0.001
$B^+ \to \overline{D}_0^{*0} \mu^+ \nu_\mu$	0.0049	0.0044	0.0008
$B^+ \to \overline{D}_2^{*0} \mu^+ \nu_\mu$	0.003	0.003	0.0004
$B^+ \to \overline{D}_1^{\prime 0} \mu^+ \nu_\mu$	0.009	0.002	0.0005
$B^+ \to \overline{D}_1^0 \tau^+ \nu_\tau$	0.0013	0.001	0.00014
$B^+ \to \overline{D}_0^{*0} \tau^+ \nu_\tau$	0.0013	0.0004	0.00015
$B^+ \to \overline{D}_1^{\prime 0} \tau^+ \nu_\tau$	0.002	0.00012	0.00005
$B^+ \to \overline{D}_2^{*0} \tau^+ \nu_\tau$	0.002	0.00021	0.00004
$B^0 \to D_2^{*-} e^+ \nu_e$	0.0023	0.0028	0.0004
$B^0 \to D_1^{\prime -} e^+ \nu_e$	0.0083	0.0019	0.00046
$B^0 \to D_0^{*-} e^+ \nu_e$	0.0045	0.00408	0.00074
$B^0 \to D_1^- e^+ \nu_e$	0.0052	0.0089	0.000911
$B^0 \to D_1^- \mu^+ \nu_\mu$	0.0052	0.0089	0.000911
$B^0 \to D_0^{*-} \mu^+ \nu_\mu$	0.0045	0.00408	0.00074
$B^0 \to D_1^{\prime -} \mu^+ \nu_\mu$	0.0083	0.0019	0.00046
$B^0 \to D_2^{*-} \mu^+ \nu_\mu$	0.0023	0.0028	0.0004
$B^0 \to D_1^- \tau^+ \nu_\tau$	0.0013	0.0009	0.00013
$B^0 \to D_0^{*-} \tau^+ \nu_\tau$	0.0013	0.0003	0.00014
$B^0 \to D_1^{\prime -} \tau^+ \nu_\tau$	0.002	0.00017	0.00005
$B^0 \to D_2^{*-} \tau^+ \nu_\tau$	0.002	0.00013	0.00004

Table 7.15: Decay modes fluctuated to evaluate the resonant $B \to D^{**}(1P)(l/\tau)\nu$ branching fractions on generic MC. The setup values are from [34, 45].

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
Resonant $B \to D^{**}(1P)(l/\tau)\nu$ branching fraction	1.87	0.35

Table 7.16: Evaluated relative systematic uncertainties from resonant $B \to D^{**}(1P)(l/\tau)\nu$ branching fractions.

The samples simulated $B \to D^{**}l\nu$ decays are used to evaluate the systematic uncertainties from resonant $B \to D^{**}(2S)l\nu$ decays, as listed in Table 7.17. The masses for $D^{**}(2S)$ and $D^{**}(2S)^*$ are around 2.47 GeV and 2.7 GeV. These events are then reconstructed, selected, and assigned (z_1, z_2) scores using the same procedure as for the generic MC. It is well known that there is a discrepancy of 1.5% between the inclusive branching fraction of semileptonic B decays and the sum of exclusive branching fractions. Conservatively, we use this sample to make up the 1.5% discrepancy. The densities of each event type are then re-evaluated. The difference between the measured $R(D^{(*)})$ on whether the $B \to D^{**}(2S)l\nu$ decays are included are taken as a systematic uncertainty. The resulting systematic uncertainty is shown in Table 7.18.

Decay Type	# event [10^6]	BABAR Dataset Name
$B^+ \to D^{**}(2S)(D\pi\pi)\ell\nu$	6.776	SP-11461-R24
$B^0 \to D^{**}(2S)(D\pi\pi)\ell\nu$	6.826	SP-11467-R24
$B^+ \to D^{**}(2S)(D^*\pi\pi)\ell\nu$	6.530	SP-11462-R24
$B^0 \to D^{**}(2S)(D^*\pi\pi)\ell\nu$	6.769	SP-11468-R24
$B^+ \to D^{**}(2S)^*(D\pi\pi)\ell\nu$	6.369	SP-11463-R24
$B^0 \to D^{**}(2S)^*(D\pi\pi)\ell\nu$	6.552	SP-11469-R24
$B^+ \to D^{**}(2S)^*(D^*\pi\pi)\ell\nu$	6.425	SP-11464-R24
$B^0 \to D^{**}(2S)^*(D^*\pi\pi)\ell\nu$	6.616	SP-11470-R24

Table 7.17: MC samples used for assessing the gap between inclusive and the sum of exclusive $B \to X_c\ell\nu$ branching fractions.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
$B \to D^{**}(2S)l\nu$ branching fraction	0.58	0.56

Table 7.18: Evaluated relative systematic uncertainties from the resonant $B \to D^{**}(2S)l\nu$ branching fractions.

For non-resonant $B \to D^{**}l\nu$ decays, we fluctuated both $B \to D^{(*)}\pi^+l\nu$ and $B \to D^{(*)}\pi^0l\nu$ decay modes simultaneously and the corresponding systematic uncertainties are listed in Table 7.19.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
Non-resonant $B \to D^{(*)}\pi l\nu$ branching fraction	2.05	1.46

Table 7.19: Evaluated relative systematic uncertainties from non-resonant $B \to D^{(*)}\pi l\nu$ branching fractions.

We do not have MC samples for resonant $B \to D^{**}(2S)\tau\nu$ and non-resonant $B \to D^{(*)}\pi\tau\nu$ decays. However, since $B \to D^{**}\tau\nu$ and $B \to D^{**}l\nu$ have similar topology in this analysis, we will use the branching ratio between $\mathcal{B}(B \to D^{**}\tau\nu)$ and $\mathcal{B}(B \to D^{**}l\nu)$ to estimate the systematic uncertainties.

We estimate the relative branching ratio $R(D^{**})$ with the available phase space Φ

$$R(D^{**}) = \frac{\mathcal{B}(B \to D^{**}\tau\nu)}{\mathcal{B}(B \to D^{**}l\nu)} \approx \frac{\Phi(B \to D^{**}\tau\nu)}{\Phi(B \to D^{**}l\nu)}.$$

The phase space of the three body decay $B \to Ml\nu$ is given by the integral

$$\begin{aligned}\Phi(B \to Ml\nu) &= \int \frac{d^3p_M}{2E_M}\frac{d^3p_l}{2E_l}\frac{d^3p_\nu}{2E_\nu}\delta^4(p_B - p_M - p_l - p_\nu) \\ &\propto \int_{m_l^2}^{(m_B-m_M)^2} dq^2\sqrt{1 - \frac{2m_l^2}{q^2} + \frac{m_l^4}{q^4}}\sqrt{(\frac{m_B^2 - m_M^2 + q^2}{2m_B})^2 - q^2}.\end{aligned} \tag{7.10}$$

We calculate $R(D^{**})$ as well as the $\Phi(B \to D^{**}l\nu)$ integrals numerically for a wide range of m_M, as shown in Fig. 7.23. The mass of the discovered excited D mesons are roughly in the range of $[2300, 3300]$ MeV, so we use $R(D^{**}) = 0.13$ at $m_{D^{**}} = 2300$ MeV to conservatively estimate the systematic uncertainty from resonant $B \to D^{**}(2S)\tau\nu$ decays. The results are shown in Table 7.20. For heavier excited D mesons, the phase space decreases and thus has a smaller effect on the measurement.

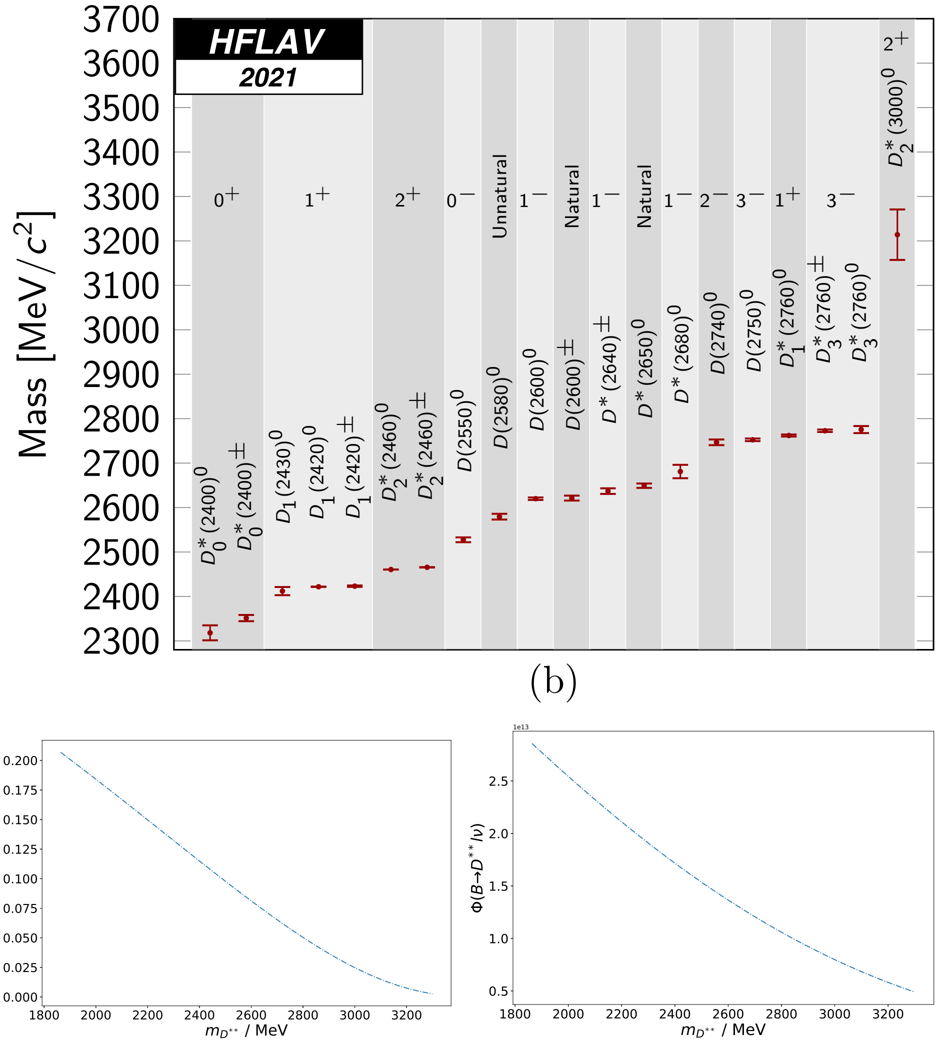

Figure 7.23: (upper) The current mass ranges for excited D meson states. (bottom left) The $R(D^{**})$ for different excited D meson masses. (bottom right) The $\Phi(B \to D^{**}l\nu)$ for different excited D meson masses.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
Resonant $B \to D^{**}(2S)\tau\nu$	0.08	0.07
Non-resonant $B \to D^{(*)}\pi\tau\nu$	0.27	0.19

Table 7.20: Evaluated relative systematic uncertainties from resonant $B \to D^{**}(2S)\tau\nu$ and non-resonant $B \to D^{(*)}\pi\tau\nu$ branching fractions.

D meson decay branching fractions

The uncertainties from D meson decay branching fractions are also considered. Since performing fluctuations on all D meson decay channels is impractical, we simply fluctuate the most common decay $D \to K\pi\pi$. Specifically, for each event involving a $D \to K\pi\pi$ decay, we assign a factor of $\omega = \omega_{W.A./\omega_{\mathtt{DECAY.DEC}}} = 0.992$. Table 7.21 shows the resulting systematic uncertainties.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
$D \to K\pi\pi$ branching fraction	0.84	0.70

Table 7.21: Evaluated relative systematic uncertainties from the $D \to K\pi\pi$ branching fractions.

$\Upsilon(4S)) \to B\bar{B}$ branching fractions

The branching fractions of $\mathcal{B}(\Upsilon(4S))$ decaying to charged or neutral B mesons are $\mathcal{B}(\Upsilon(4S) \to B^+B^-) = 51.4\% \pm 0.6\%$, and $\mathcal{B}(\Upsilon(4\mathrm{S}) \to B^0\bar{B}^0) = 48.6\% \pm 0.6\%$. To evaluate the effect of the uncertainty of the neutral/charged B meson ratio, we fluctuate the relative number of events between $B^0\bar{B}^0$ and B^+B^- in the amount of 1.1% and propagate it to the uncertainty of $R(D^{(*)})$. The results are in Table 7.22.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
$\mathcal{B}(\Upsilon(4S))$	0.48	0.43

Table 7.22: Evaluated relative systematic uncertainties from $\mathcal{B}(\Upsilon(4S))$ decays to charged or neutral B mesons.

Lepton efficiency

We assign a 1% relative uncertainty per electron or muon in an event, and assign a 1.1% relative uncertainty per kaon in an event. There are two light leptons and two kaons for a signal or normalization event, one from the B_{tag} and the other from B_{sig}. However, most sources of PID efficiency uncertainties will cancel out, since we measure the ratio between signal and normalization event. The only source left is the signal side lepton PID efficiency, because the signal side lepton momentum varies for signal and normalization events.

If we assume $\epsilon_{\mathrm{PID}}^{\mathrm{signal}} \sim \mathcal{N}(\mu_x, \sigma_x^2), \epsilon_{\mathrm{PID}}^{\mathrm{norm}} \sim \mathcal{N}(\mu_y, \sigma_y^2)$, then

$$\left(\frac{\sigma_{x/y}}{x/y}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2 - 2 \cdot \left(\frac{\sigma_x}{x}\right) \cdot \left(\frac{\sigma_y}{y}\right).$$

Based on [46, 47], the corresponding $R(D^{(*)})$ uncertainties are 0.29% and 0.40%.

Neutral pion efficiency

The soft π^0 efficiency uncertainty does not cancel in the ratio. Studies have been performed by the *BABAR* collaboration to measure the difference of the π^0 efficiency in data and simulation. Following BAD [48], an average π^0 correction $\omega_{\pi^0} = 0.958 \pm 0.009$ is applied, and the corresponding uncertainties are propagated to the results of $R(D^*)$. Although the soft pion efficiency is momentum dependent, the momentum for nearly all soft pions is in the range of $[0.1, 1]$ GeV in this analysis. Given the momentum-dependent correction factor as shown in Fig. 7.24, this is a second-order effect and can be neglected.

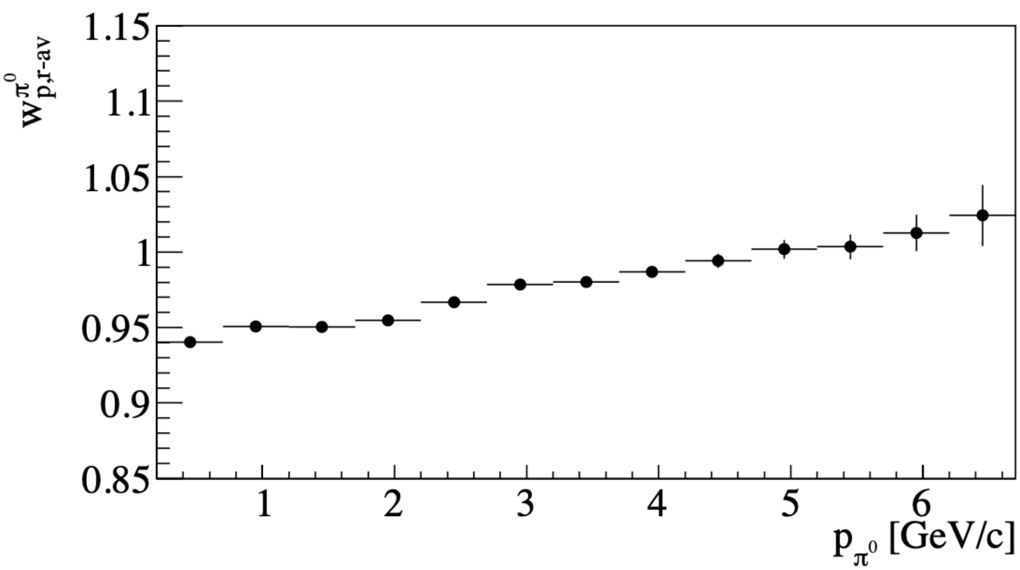

Figure 7.24: The π^0 efficiency correction values in dependence of the π^0 momentum. The figure is from [48].

PDF shape of on-peak backgrounds

Among the three types of backgrounds, the continuum background is characterized using off-peak data. However, the shape of $B \to D^{**} l \nu$ and combinatorial $B\bar{B}$ background are characterized using generic MC, which may be different from what happened in the detector. We denote the latter two backgrounds as $B\bar{B}$ backgrounds. To compare the shape difference of $B\bar{B}$ backgrounds between MC and data, we construct a sideband in which the data consists only of background events. The criteria to define the sideband is:

S1: $E_{extra} > 1.0$ GeV

S2: $|p_l^{sig}| < 0.4$ GeV

We assume that the signal and normalization events are negligible in the sideband region (estimated proportions are 0.4% for signal events and 3.8% for normalization events). A sample of 1% of on-peak data is used to demonstrate and calibrate the difference between MC and data, as shown in Fig. 7.25. Since the continuum background is well modeled using off-peak data, the discrepancy in the figure is a result of mis-modeling of $B\bar{B}$ background.

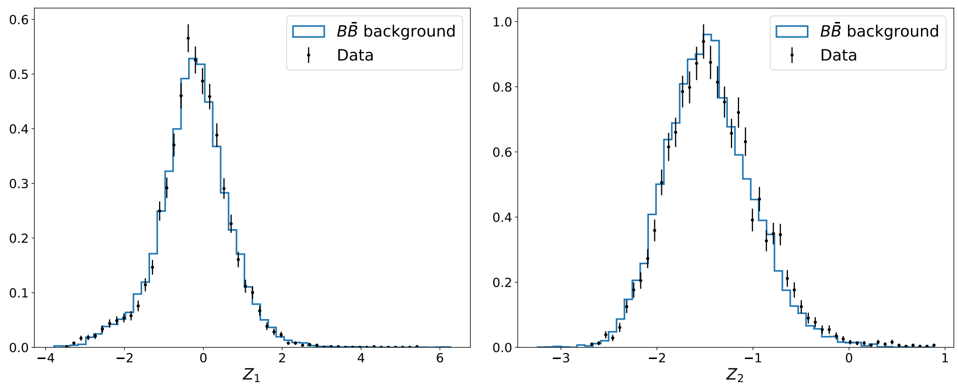

Figure 7.25: z_1 and z_2 distribution of sidebands for generic MC and sideband data. The difference between data and histogram indicates the discrepancy between MC and data.

The discrepancy of the (z_1, z_2) distribution between MC and data may introduce a bias on this measurement, therefore, we apply a correction factor based on the 2-dimensional shape difference as follows:

1. Let $g(z_1, z_2)$ be the (z_1, z_2) distribution of on-peak data in the sideband region. Since the signal and normalization events are negligible, its density can be decomposed as the sum of $B\bar{B}$ background and continuum components:

$$g(z_1, z_2) = p_{B\bar{B}} g_{B\bar{B}}(z_1, z_2) + p_{cont} g_{cont}(z_1, z_2)$$

$$p_{B\bar{B}} + p_{cont} = 1.$$

2. Let $f_{B\bar{B}}(z_1, z_2)$ be the density function of the sideband MC $B\bar{B}$ background.

3. The correction factors are:

$$\omega(z_1, z_2) = \frac{g_{B\bar{B}}(z_1, z_2)}{f_{B\bar{B}}(z_1, z_2)} = \frac{g(z_1, z_2) - p_{cont} g_{cont}(z_1, z_2)}{p_{B\bar{B}} f_{B\bar{B}}(z_1, z_2)}. \tag{7.11}$$

The systematic uncertainty from these correction factors is estimated using a bootstrap technique. Since the variance of $\omega(z_1, z_2)$ is due to the limited sample, to model the PDF shapes, we bootstrap the sideband sample of $B\bar{B}$ background, on-peak data, and off-peak data to capture its effect on the measurement.

1. We bootstrap the sideband sample of $B\bar{B}$ background, on-peak data, and off-peak data 100 times.
2. For each bootstrap sample, we calculate Equation 7.11 and apply the sideband calibration for each $B\bar{B}$ background events.
3. For each bootstrap sample, we use updated generic MC events to extract the signal and normalization yields and measure $R(D^{(*)})$.
4. We take the sample standard deviation of the measured $R(D^{(*)})$ as an estimate of the systematic uncertainty.

The results are shown in Table 7.23.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
$B\bar{B}$ background shape	0.73	0.37

Table 7.23: Evaluated relative systematic uncertainties from the calibration factor on the $B\bar{B}$ background shape difference between generic MC and data.

Sideband selection efficiency

Our definition of sideband does not guarantee the same selection efficiency for $B\bar{B}$ events and $D^{**}l\nu$ events. Therefore, the corresponding systematic is estimated. We apply the correction only for $B\bar{B}$ events and $D^{**}l\nu$ events, and take the half of the discrepancy as an estimate of the systematic uncertainty. The results are shown in Table 7.24.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)
Sideband Region	2.34	0.23

Table 7.24: Evaluated relative systematic uncertainties from the calibration factor on the $B\bar{B}$ background shape difference between generic MC and data.

Correlation between the uncertainties on $R(D)$ and $R(D^*)$

The correlations between $R(D)$ and $R(D^*)$ are used to combine all the uncertainties together and compare with the theoretical predictions. We use the MC samples to estimate the correlations for additive systematic uncertainties, while the correlations for multiplicative uncertainties can be clearly derived.

For the additive systematic uncertainties, a standard approach to estimate correlations is bootstrapping: we should bootstrap several samples and measure $R(D^{(*)})$ for each sample, and then estimate the correlations using the sample results. However, the above approach is impractical in this analysis, due to the lengthy computation required to evaluate the kernel densities for each bootstrapping sample. Therefore, we use importance sampling as an approximate method. For a given systematic uncertainty source X with standard deviation σ, we assume it follows a Gaussian distribution $X \sim \mathcal{N}(\mu, \sigma^2)$, and affects the $R(D^{(*)})$ measurements in the unknown form of $r^{(*)}(X)$. We fluctuate X up and down on an amount of $\pm\sigma/2, \pm\sigma$, and then we measure $R(D^{(*)})$ for each fluctuated sample. The correlation between $R(D)$ and $R(D^*)$ under the systematic uncertainty source X can be obtained by computing the weighted Pearson correlations between the following two samples:

$$[r(X-\sigma), r(X-\sigma/2), r(X), r(X+\sigma/2), r(X+\sigma)]$$
$$[r^*(X-\sigma), r^*(X-\sigma/2), r^*(X), r^*(X+\sigma/2), r^*(X+\sigma)],$$

with the weighting $[\text{Prob}[X = \mu - \sigma], \text{Prob}[X = \mu - \sigma/2], \text{Prob}[X = \mu], \text{Prob}[X = \mu + \sigma/2], and \text{Prob}[X = \mu + \sigma]]$, which are standard Gaussian densities at the corresponding $\pm\sigma/2, \pm\sigma$ away from mean value. This approach is applied to estimate correlations for all the additive systematic uncertainties as well as the $\mathcal{B}(\Upsilon(4S))$ uncertainty.

Among the multiplicative uncertainties, the PID efficiency, soft π^0 efficiency and tracking efficiency affect $R(D)$ and $R(D^*)$ equally, so it has a 100% correlation. The efficiency uncertainty due to limited statistics (MC Efficiency) is clearly uncorrelated. The uncertainty on $\mathcal{B}(\tau \to l^- \bar{\nu}_l \nu_\tau)$ affects all channels equally, so it also has 100% correlation. The correlation on the statistical uncertainty is evaluated by fitting the MLE multiple times, and is found to be negatively correlated.

Summary

Table 7.25 summarizes all the uncertainties taken into consideration, as well as the correlations between the uncertainties on $R(D)$ and $R(D^*)$. The effective correlation coefficient ρ_{tot} is calculated by adding the covariance matrices as follows:

$$\sum_i \begin{pmatrix} \sigma_i^2 & \rho_i\sigma_i\sigma_i^* \\ \rho_i\sigma_i\sigma_i^* & \sigma_i^{*2} \end{pmatrix} = \begin{pmatrix} \sigma_{\text{tot}}^2 & \rho_{\text{tot}}\sigma_{\text{tot}}\sigma_{\text{tot}}^* \\ \rho_{\text{tot}}\sigma_{\text{tot}}\sigma_{\text{tot}}^* & \sigma_{\text{tot}}^{*2} \end{pmatrix}, \tag{7.12}$$

where $\sigma_i^{(*)}$ refers to uncertainties on $R(D^{(*)})$, and i runs over each source of uncertainty.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)	Correlation
$B \to Dl\nu$ form factor	0.42	0.13	-0.18
$B \to D^*l\nu$ form factor	0.92	0.31	-0.19
$B \to D^{**}l\nu$ form factor	0.48	0.18	-0.90
$\mathcal{B}(B \to D^{(*)}l\nu)$	0.47	0.38	0.97
$\mathcal{B}(b \to c\bar{c})$	0.34	0.13	1
$\mathcal{B}(B \to D^{**}l\nu)$	2.83	1.60	-0.97
$\mathcal{B}(D)$	0.84	0.70	-0.40
PDF shapes MC statistics	4.12	4.37	-0.15
On-peak background calibration	2.45	0.44	-0.05
$\mathcal{B}(\Upsilon(4S))$	0.48	0.43	1
PID efficiency	0.29	0.40	1
Soft π^0 efficiency	0.84	1.25	1
$\mathcal{B}(\tau \to l^-\bar{\nu}_l\nu_\tau)$	0.16	0.16	1
Systematic Total	5.86	4.96	-0.20
Statistical Uncertainty	19.8	9.9	-0.92
Total	20.65	11.07	-0.82

Table 7.25: Summary of evaluated uncertainties (preliminary).

7.8 Results

Our preliminary results are

$$\begin{aligned} \mathcal{R}(D) &= 0.316 \pm 0.062 \pm 0.019 \\ \mathcal{R}(D^*) &= 0.226 \pm 0.022 \pm 0.012, \end{aligned} \tag{7.13}$$

where the first uncertainties are statistical and the second are systematic. The correlation between $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ is -0.82. The fitted yields for each event type in all of four subsets are listed in Table 7.9. The signal and normalization efficiencies are listed in Table 7.26. The comparison of our result with previous measurments is shown in Fig. 7.26

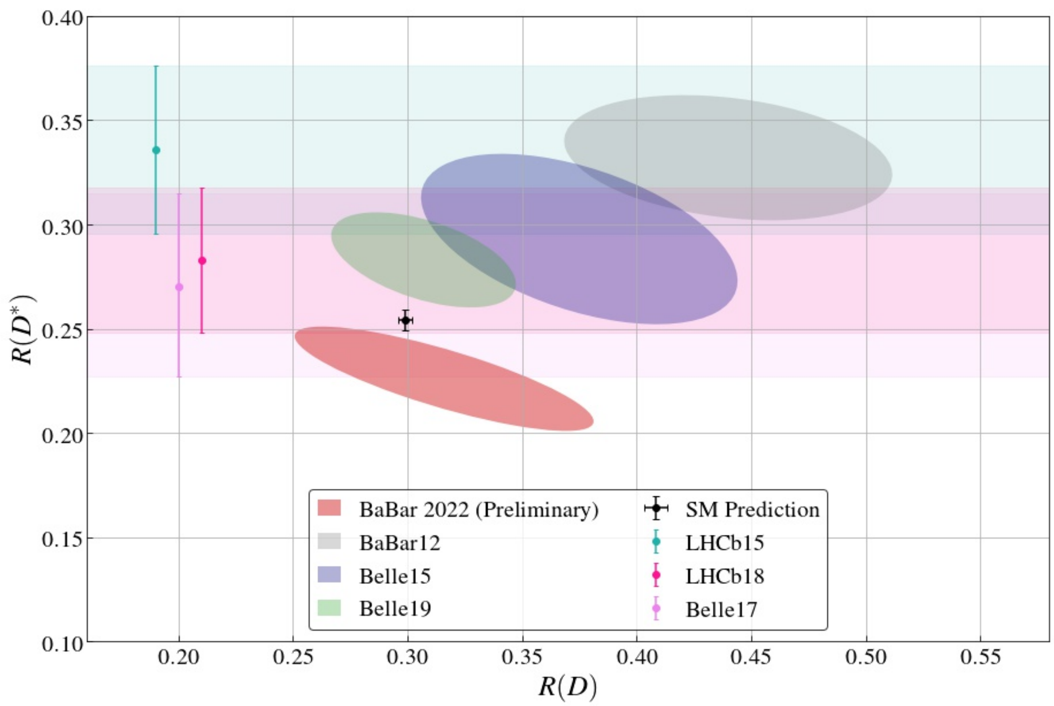

Figure 7.26: The comparison of preliminary results with previous measurments for $R(D^{(*)})$.

Event type	Efficiency (%)
$B \to D\tau\nu$	0.41
$B \to D^*\tau\nu$	0.47
$B \to Dl\nu$	0.46
$B \to D^*l\nu$	0.43

Table 7.26: Efficiencies for signal and normalization events.

7.9 Summary

In summary, we measure $R(D^{(*)})$ using a semileptonic tagging method and leptonic τ decays. Our preliminary results are $\mathcal{R}(D) = 0.316 \pm 0.062 \pm 0.019$ and $\mathcal{R}(D^*) = 0.226 \pm 0.022 \pm 0.012$, where the first uncertainties are statistical and the second are systematic. The measured $R(D)$, $R(D^*)$, and their combinations agree with the Standard Model predictions by 0.26σ, 1.10σ, and 1.51σ. This analysis is BABAR's first measurement of $R(D^{(*)})$ using a semileptonic tagging method.

The measurement strategy imposes as minimal assumptions from Monte Carlo samples as possible. For instance, instead of fixing background yields, they are also treated as free parameters when fitting the data. By doing this, our strategy can overcome potential simulation bias. It is also feasible because of the customer package developed to best model PDFs of decays with millions of MC samples. Due to this, our measurement strategy is unique and reliable.

Our combined results on $R(D^{(*)})$ agree with the Belle's semileptonic tagging measurement by 2.23σ. If we average the two semileptonic tagging measurements together and compare with the SM prediction, the combined difference is about 0.46σ.

Similar to previous measurements, this measurement is also statistically dominant. However, with future Belle II experiment target luminosity of 50 ab^{-1}, the statistical uncertainty on $R(D^{(*)})$ measurements can be reduced to only few percent, making it comparable with the systematic uncertainties and better to probe the new physics.

Part V

Conclusion

Chapter 8

CONCLUSION

This thesis focuses on collider searches of beyond Standard Model physics at the intensity frontier. With an integrated luminosity of 531 fb^{-1} collected at the *BABAR* experiment, both direct and indirect methods to probe for New Physics are performed with this unique experimental environment.

We first briefly review the Standard Model, which is the current best theory to describe the interactions between fundamental particles, and has been extensively tested by a series of experiments. However, the Standard Model is not a complete theory. Besides the fact that it does not include gravity, we also summarize its limitation in both experimental and theoretical aspects. The main goal of high energy physics is to search and understand the beyond Standard Model physics.

We then describe the details of the PEP-II accelerator and *BABAR* detector. It is an e^+e^- collider operating at the center-of-mass energy around 10.58 GeV. The asymmetric design enables the B meson decay length to be measurable in the lab frame. We also briefly summarize the main components of the *BABAR* detector.

We present our direct search for beyond Standard Model physics by looking for dark sectors. Dark sectors are new particle(s) interacting only feebly with ordinary matter mediated via portals, and have become an intriguing framework to explain the presence of dark matter in the Universe. While previous collider searches focus on identifying new mediators, we investigate the possibility of dark matter bound states as a probe for dark sector. This is also the first search for darkonium. In an absence of signal, we show that this search improves the existing constraints on the $\gamma - A'$ mixing strength over a significant fraction of dark photon masses below 1 GeV for large values of the dark sector coupling constant.

We then present our indirect search for beyond Standard Model physics by precision measurement of semileptonic B meson decays. We measure $R(D)$ and $R(D^*)$, which are sensitive probes for BSM physics, using a semileptonic tagging method. The results are dominated by statistical uncertainties, and agree with the Standard Model prediction within 1.51σ. This measurement applies a data-driven approach and is therefore robust to potential bias from simulation. The unique strategy is enabled by our custom-developed software for fast kernel density estimation powered by GPU technology, which enables us to best characterize signatures of decays with large samples.

In the appendix, we demonstrate the idea of applying deep transfer learning algorithms to reduce systematic uncertainties. With the adversarial neural network architecture, the multivariate classifiers can be trained to be insensitive to the small variance of event distributions. This framework can be applied to most high energy physics analysis to reduce the systematics from mis-modeling of signal or background samples.

BIBLIOGRAPHY

[1] "https://www.quantumdiaries.org/2014/03/14/the-standard-model-a-beautiful-but-flawed-theory".

[2] MS Zisman. "PEP-II: An Asymmetric B Factory: Conceptual Design Report". In: *SLAC Report* 418 (1993), pp. 2–93.

[3] D Boutigny et al. "The BABAR physics book: Physics at an asymmetric B factory". In: (1999).

[4] Gerard Jungman, Marc Kamionkowski, and Kim Griest. "Supersymmetric dark matter". In: *Physics Reports* 267.5-6 (1996), pp. 195–373.

[5] SA Abel et al. "Kinetic mixing of the photon with hidden U (1) s in string phenomenology". In: *Journal of High Energy Physics* 2008.07 (2008), p. 124.

[6] Mark Goodsell et al. "Naturally light hidden photons in LARGE volume string compactifications". In: *Journal of High Energy Physics* 2009.11 (2009), p. 027.

[7] Mark Goodsell, Saúl Ramos-Sánchez, and Andreas Ringwald. "Kinetic mixing of U (1) s in heterotic orbifolds". In: *Journal of High Energy Physics* 2012.1 (2012), p. 21.

[8] Nima Arkani-Hamed and Neal Weiner. "LHC signals for a superunified theory of dark matter". In: *Journal of High Energy Physics* 2008.12 (2008), p. 104.

[9] Pierre Fayet. "U-boson production in e+ e- annihilations, ψ and Y decays, and Light Dark Matter". In: *Physical Review D* 75.11 (2007), p. 115017.

[10] C Cheung. "C. Cheung, JT Ruderman, L.-T. Wang, and I. Yavin, Phys. Rev. D 80, 035008 (2009)." In: *Phys. Rev. D* 80 (2009), p. 035008.

[11] David E Morrissey, David Poland, and Kathryn M Zurek. "Abelian hidden sectors at a GeV". In: *Journal of High Energy Physics* 2009.07 (2009), p. 050.

[12] Haipeng An et al. "Direct detection constraints on dark photon dark matter". In: *Physics Letters B* 747 (2015), pp. 331–338.

[13] Haipeng An et al. "Probing the dark sector with dark matter bound states". In: *Physical review letters* 116.15 (2016), p. 151801.

[14] FJ Rogers, HC Graboske Jr, and DJ Harwood. "Bound eigenstates of the static screened Coulomb potential". In: *Physical Review A* 1.6 (1970), p. 1577.

[15] R Agnese et al. "Search for low-mass weakly interacting massive particles using voltage-assisted calorimetric ionization detection in the SuperCDMS experiment". In: *Physical review letters* 112.4 (2014), p. 041302.

[16] Daniel S Akerib et al. “First results from the LUX dark matter experiment at the Sanford Underground Research Facility”. In: *Physical review letters* 112.9 (2014), p. 091303.

[17] Geoffrey T Bodwin, Eric Braaten, and G Peter Lepage. “P hys. Rev., D51, 1125 (1995),[Erratum”. In: *Phys. Rev. D* 55 (1997), p. 5853.

[18] Andrea Petrelli et al. “NLO production and decay of quarkonium”. In: *Nuclear Physics B* 514.1-2 (1998), pp. 245–309.

[19] Sea Agostinelli et al. “GEANT4—a simulation toolkit”. In: *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303.

[20] Xinran He et al. “Practical lessons from predicting clicks on ads at facebook”. In: *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*. ACM. 2014, pp. 1–9.

[21] Kenzo Nakamura, Particle Data Group, et al. “Review of particle physics”. In: *Journal of Physics G: Nuclear and Particle Physics* 37.7A (2010), p. 075021.

[22] Yasmine Sara Amhis et al. “Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018”. In: (2019). updated results and plots available at `https://hflav.web.cern.ch/`. arXiv: `1909.12524 [hep-ex]`.

[23] BaBar Collaboration et al. “Measurement of an excess of $B \to D^{(*)}\tau\nu_\tau$ decays and implications for charged Higgs bosons”. In: *Phys. Rev. D* 88.072012 (2013), pp. 1303–0571.

[24] Irinel Caprini, M Neubert, and LP Lellouch. “Dispersive Bounds on the Shape of $B \to D^{(*)}l\nu$ Form Factors”. In: *Nucl. Phys. B* 530.hep-ph/9712417 (1998), pp. 153–181.

[25] Simone Bifani et al. “Review of Lepton Universality tests in B decays”. In: *Journal of Physics G: Nuclear and Particle Physics* 46.2 (2018), p. 023001.

[26] Bernard Aubert et al. “Observation of the Semileptonic Decays $B \to D^*\tau\nu_\tau$ and Evidence for $B \to D\tau\nu_\tau$”. In: *Physical review letters* 100.2 (2008), p. 021801.

[27] JP Lees et al. “Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays”. In: *Physical review letters* 109.10 (2012), p. 101802.

[28] Matthias Huschle et al. “Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}l^-\bar{\nu}_l$ decays with hadronic tagging at Belle”. In: *Physical Review D* 92.7 (2015), p. 072014.

[29] Roel Aaij et al. “Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$”. In: *Phys. Rev. Lett.* 115.LHCB-PAPER-2015-025 (2015), p. 111803.

[30] Giacomo Caria et al. “Measurement of $R(D)$ and $R(D^*)$ with a Semileptonic Tagging Method”. In: *Physical review letters* 124.16 (2020), p. 161803.

[31] A. Abdesselam *et al*. In: (2016). eprint: arXiv:1603.06711[hep-ex].

[32] S. Hirose *et al*. "Measurement of the τ Lepton Polarization and $R(D^*)$ in the Decay $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$". In: *Phys. Rev. Lett.* 118 (21 2017), p. 211801. DOI: 10.1103/PhysRevLett.118.211801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.118.211801.

[33] R. Aaij *et al*. "Measurement of the Ratio of the $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ Branching Fractions Using Three-Prong τ-Lepton Decays". In: *Phys. Rev. Lett.* 120 (17 2018), p. 171802. DOI: 10.1103/PhysRevLett.120.171802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.171802.

[34] PARTICLE DATA GROUP Collab. K. Olive et al. In: *Chin. Phys. C* 38 (2014), p. 090001.

[35] Daniel Shuteh Chao. "Measuring $R(D^{(*)})$ for $B \to D^{(*)}\tau\nu_\tau$ using Semileptonic Tags and Tau Decays to Hadrons". PhD thesis. California Institute of Technology, 2018.

[36] Jerome H Friedman. "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics* (2001), pp. 1189–1232.

[37] Bernard W Silverman. *Density estimation for statistics and data analysis*. Vol. 26. CRC press, 1986.

[38] Jon Louis Bentley. "Multidimensional binary search trees used for associative searching". In: *Communications of the ACM* 18.9 (1975), pp. 509–517.

[39] Martin S Andersen, Joachim Dahl, Lieven Vandenberghe, et al. "CVXOPT: A Python package for convex optimization". In: *abel. ee. ucla. edu/cvxopt* 88 (2013).

[40] Joseph L Doob. "The limiting distributions of certain statistics". In: *The Annals of Mathematical Statistics* 6.3 (1935), pp. 160–169.

[41] Bradley Efron. "Bootstrap methods: another look at the jackknife". In: *Breakthroughs in statistics*. Springer, 1992, pp. 569–593.

[42] Y Amhis et al. "Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2014". In: *arXiv preprint arXiv:1412.7515* (2014).

[43] Adam K Leibovich et al. "Semileptonic B decays to excited charmed mesons". In: *Physical Review D* 57.1 (1998), p. 308.

[44] Florian Bernlochner. "Impact of D^{**} on $B \to D^*l\nu$ and $R(D^{(*)})$". In: *Amplitude Analyses for Flavour Anomalies* (Bristol 2019). URL: https://indico.cern.ch/event/810429/contributions/3376226/attachments/1875340/3087759/Amplitudes.pdf.

[45] Florian U Bernlochner and Zoltan Ligeti. "Semileptonic $B_{(s)}$ decays to excited charmed mesons with e, μ, τ and searching for new physics with $R(D^{**})$". In: *Physical Review D* 95.1 (2017), p. 014022.

[46] Vuosalo C. O., Telnov A. V., and K. T. Flood1. “Muon Identification Using Decision Trees”. In: *BaBar Analysis Document* (2010). URL: https://babar.heprc.uvic.ca/BFROOT/www/Physics/BAD/vol18/01853.003.pdf.

[47] Alessandro Gaz et al. “Particle Identification Using Error Correcting Output Code Multiclass Classifier”. In: *BaBar Analysis Document* (2011). URL: https://www.slac.stanford.edu/babar-internal/BAD/doc/detail.html?docNum=2199.

[48] BaBar Collaboration et al. “Branching fractions measurement of $\tau^- \to K^- n\pi^0 \nu_\tau$ with $n = 0, 1, 2, 3$ and $\tau^- \to \pi^- n\pi^0 \nu_\tau$ with $n = 3, 4$”. In: *BaBar Analysis Document* 2318 (2012).

[49] Yaroslav Ganin and Victor Lempitsky. “Unsupervised domain adaptation by backpropagation”. In: *International conference on machine learning*. PMLR. 2015, pp. 1180–1189.

[50] Eric Tzeng et al. “Adversarial discriminative domain adaptation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 7167–7176.

[51] Gilles Louppe, Michael Kagan, and Kyle Cranmer. “Learning to pivot with adversarial networks”. In: *Advances in neural information processing systems* 30 (2017).

APPENDIX A: DEEP ADVERSARIAL NETWORK FOR SYSTEMATIC UNCERTAINTY REDUCTION

Monte Carlo simulation samples are almost always used in high energy physics to study the characteristics of particle decays and establish the measurement strategies. They are used to simulate the response of detectors in the event reconstruction stage, and to optimize criteria in the event selection stage. However, simulations are not always perfect, especially for background with high-multiplicity decay modes. The imperfection from simulations is usually considered as a source of systematic uncertainties in the measurements.

For a simple case in which experimental data is categorized into only two labels, signal and background, we usually generate and simulate signal MC sample S^{mc} and background MC sample B^{mc}. A typical and widely used strategy to establish event selection strategy is to optimize criteria using the combined samples of S^{mc} and B^{mc}. To estimate corresponding systematics, people compare B^{mc} distribution with experimental data on signal suppressed regions, and evaluate the difference as systematics. This strategy is reliable but not optimal, as it does not take into account the information of MC/data discrepancy when optimizing selection criteria.

The systematics due to MC/data discrepancy can be potentially reduced if the adversarial framework can be introduced. One architecture we played with is unsupervised domain adaptation [49], as shown in Fig. 8.1, while other adversarial architectures about unsupervised domain adaptation [50, 51] are also applicable. In this framework, the input variables $\mathbf{x}$ (normally physical variables of an event) are first transformed to $\mathbf{f}$ using $G_f(\mathbf{x};\theta_f)$, and then used to train two classifiers together. The first classifier $G_t(\mathbf{x};\theta_t)$, similar to traditional multivariate analysis algorithms in event selection, aims to distinguish between signal and backgrounds. The second classifier $G_d(\mathbf{x};\theta_d)$ aims to distinguish whether the input event is from MC simulation or experimental data. Each event has two labels, y_t to label signal/background, and y_d to label it is from MC simulation or experimental data. Therefore, a more robust classifier can be obtained by minimizing the combined loss function

$$\begin{aligned}
\mathcal{L} &= \mathcal{L}_t - \lambda\mathcal{L}_d \\
\mathcal{L}_t &= \mathcal{L}_t(G_t(G_f(x)), y_t) \\
\mathcal{L}_d &= \mathcal{L}_d(G_d(G_f(x)), y_d),
\end{aligned} \tag{8.1}$$

where $\mathcal{L}_t$ and $\mathcal{L}_d$ are the loss functions for signal/background classification and

MC/data classification. λ is a hyper-parameter to control the tradeoff between accuracy and robustness.

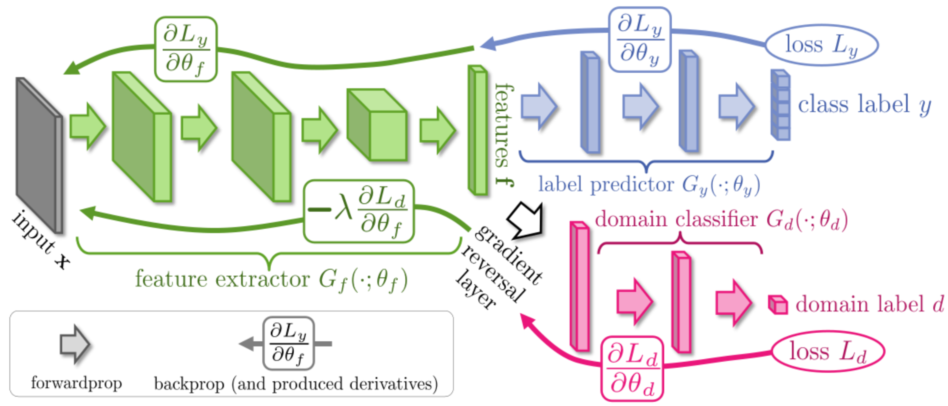

Figure 8.1: Architecture of unsupervised domain adaptation. The figure is taken from [49].

Although the algorithm does not give significant improvement in our analysis, as our analysis is dominated by statistical uncertainties, we believe this framework is sufficient general to reduce systematics when the mis-modeling of particle decays is not negligible.