Radiative Leptonic B Decays

Edwad Tann Chen

SLAC-R-900

Prepared for the Department of Energy
under contract number DE-AC02-76SF00515

Printed in the United States of America. Available from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161.

Radiative Leptonic B Decays

Thesis by

Edward Tann Chen

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended September 25, 2006)

Acknowledgments

First off, I would like to acknowledge my thesis advisor, Professor David Hitlin. This thesis would not have been possible without you. I have been lucky to have someone with such a depth and breadth of knowledge about high energy physics as my advisor. I learned a lot from you. Also, I have really appreciated your positive attitude, availability, and willingness to do whatever it took to help take my project to completion. And, it was good to have you on my side as an advocate within the *BABAR* collaboration. Not all graduate students are this fortunate. Thank you.

In addition, I would like to thank Gregory Dubois-Felsmann, a post-doc who, along with David and me, formed the radiative leptonic analysis group at Caltech. In so many ways, this thesis is a product of your vision. Working with you on a regular basis and getting constant, insightful feedback was what took me from a green, clueless graduate student to a fairly competent and knowledgeable analyst. Along with Dave, you are really one of the top physicists on *BABAR* in terms of knowledge and experience. Plus, you are one of the sharpest persons I have had the opportunity to meet in my life. I always had confidence that when I got stuck, you would have some insight that would get me unstuck. I really appreciate that you took time to teach me and help me develop as a physicist. Thank you.

I also owe many thanks to the other members of the Caltech *BABAR* group. Professor Frank Porter has been an integral part of my graduate school experience, both as a teacher and a researcher. His knowledge of statistics (not to mention physics in general), especially applied to high energy physics analyses, is tremendous, and I have been grateful to have had him as a resource. Also, post-docs Yury Kolomensky, Anders Ryd, Ilya Narsky, Scott Metzler, and Justin Albert have both helped me and taught me much. I have also been fortunate to have had as my fellow graduate students Alexei Dvoretskii, Alex Samuel, Tim Piatenko, and Dave Doll.

Because this analysis was conducted within the *BABAR* experiment, I am grateful to the members of the *BABAR* collaboration and the PEP-II staff for the tremendous amount of work done to get this experiment up and running, so that I could have all this data to analyze. In addition, this analysis went through the entire *BABAR* review process, and I am grateful to the collaboration for their useful feedback. In particular, our review committee of David Strom, Hojeong Kim, and Jack Ritchie did a great amount of work in reviewing the analysis. Finally, the Leptonic B and C analysis working group, in particular conveners Steve Robertson, Steve Sekula, and Paul Jackson, were really instrumental in taking this analysis from its start to finish.

This analysis required some input from the theoretical community. In particular, Dan Pirjol has been really responsive to our requests for help. Also, C. S. Kim and Y. Yang provided useful input for the η modes. I thank you all.

I would like to thank both my oral candidacy examination committee (David Hitlin, Barry Barish, Frank Porter, and Mark Wise), and thesis defense committee

(David Hitlin, Sunil Golwala, Frank Porter, and Mark Wise) for taking time out of their busy schedules to participate. My thesis defense committee was great in terms of taking time to read drafts of this dissertation and provide feedback.

Behind the scenes, the support staff at Caltech was instrumental in making my graduate school experience so much smoother. In particular, many thanks to Donna Driscoll, Betty Smith, Lupe Llamas, and Juan Barayoga. In addition, I typeset this thesis using a template written by Ling Li, Daniel Zimmerman, and Mike Kelsey.

Graduate school would not have been nearly a tenth of the enjoyable experience it was without the company of friends. Some of you went through similar experiences and that made a world of difference. While many have had an impact in my life in one way or another, I'd like to mention some names, at the risk of omitting some equally important ones: Vlatko Balic, Emily Balic, Alexei Dvoretskii, Tanja Bosak, Ben Weiss, Keith Matthews, Michele Vallisneri, Marc Hill, Laura Vernon, Jason T. T. McKeever, Ali Husain, Arkadas Ozakin, Paul Skerritt, Greg Vaitsas, Sylvie Brunet, David Cote, Wahid Bhimji, Jamie Boyd, Nick Barlow, Amir Farbin, Kun-Chin Lin, Christina Von Mayrhauser, Attila Kovacs, Ben Schupp, Dave Doll, Juancarlos Chan, Jean Giffin, and Bea Bartlett. It was heartwarming to be cheered at the finish line by you.

Finally, I like to thank my family (father Joseph, mother Fong-Jiao, brothers Jason and Michael, and sister-in-law Vivian) for their love and support throughout graduate school.

Abstract

We present the results of a search for B^+ meson decays into $\gamma\ell^+\nu_\ell$, where $\ell = e, \mu$. We use a sample of 232 million $B\bar{B}$ meson pairs recorded at the $\Upsilon(4S)$ resonance with the *BABAR* detector at the PEP-II B factory. We measure a partial branching fraction $\Delta\mathcal{B}$ in a restricted region of phase space that reduces the effect of theoretical uncertainties, requiring the lepton energy to be in the range 1.875 and 2.850 GeV, the photon energy to be in the range 0.45 and 2.35 GeV, and the cosine of the angle between the lepton and photon momenta to be less than -0.36, with all quantities computed in the $\Upsilon(4S)$ center-of-mass frame. We find $\Delta\mathcal{B}(B^+ \to \gamma\ell^+\nu_\ell) = (-0.3^{+1.3}_{-1.5}(\text{statistical})^{+0.6}_{-0.6}(\text{systematic}) \pm 0.1(\text{theoretical})) \times 10^{-6}$, under the assumption of lepton universality. Interpreted as a 90% confidence-level Bayesian upper limit, the result corresponds to 1.7×10^{-6} for a prior flat in amplitude, and 2.3×10^{-6} for a prior flat in branching fraction.

Contents

Tables

Figures

Chapter 1

Introduction

This dissertation presents the experimental search for the radiative leptonic B decay $B^+ \to \ell^+ \nu_\ell \gamma$, where $\ell = e, \mu$.[1] While a measurement of the branching fraction could, in principle, be of importance in flavor physics, the current primary importance is in quantum chromodynamics (QCD), where theoretical and experimental uncertainties relating to hadronic dynamics are still relatively large. In particular, within the context of the current state of affairs in our phenomenological knowledge of B-meson physics, a branching fraction measurement of the radiative leptonic decay mode is most useful for constraining λ_B, the first inverse moment of the B-meson light-cone distribution amplitude.

We will begin with an introduction of the standard model, and progress within its framework to a discussion of radiative leptonic B decays.

[1] Throughout this thesis, charge-conjugate modes are implied.

1.1 Standard Model Overview

Our most successful theoretical model of fundamental particles and their interactions is the standard model.[2] It describes the strong and electroweak interactions, but not the gravitational interaction. If we assume that neutrinos are massless, then the standard model requires 19 input parameters: the three charged-lepton masses, six quark masses, three gauge coupling constants, three quark-mixing angles and one complex phase, a Higgs mass and quartic coupling constant, and the QCD vacuum angle.

The primary theoretical tool of the standard model is relativistic quantum field theory. In particular, one can write down the Lagrangian for a theory, split it into free and interaction terms, and then use the perturbative methods of Feynman diagrams to calculate the amplitude for a particular scattering process.

If one starts with a Lagrangian that only includes free fermionic fields that are representations of a symmetry group and imposes a local internal gauge symmetry upon it, one is forced to introduce gauge fields and interaction terms. The standard model Lagrangian has an $SU(3) \otimes SU(2) \otimes U(1)$ gauge structure, leading to three kinds of charge: color, weak isospin, and weak hypercharge. The fundamental fields are the spin-1/2 leptons, spin-1/2 quarks, and spin-1 gauge bosons. In addition, there is the scalar Higgs field, which has not yet been observed.

[2]The standard model discussion is largely based on References [1]–[4].

1.2 Symmetries

From a purely physical standpoint, any symmetry of a physical system implies that a measurement should be invariant under the symmetry operation. For example, a branching-fraction measurement should not be dependent upon a rotation or boost of the system. Symmetries of a physical system should correspond to the invariance of the Lagrangian under symmetry operations.

We have already mentioned the implications of requiring local gauge symmetry. In addition to gauge symmetry, the effects of the discrete operations of parity, time-reversal, and charged conjugation on the Lagrangian are of particular interest. Parity (P) reverses handedness: $\vec{x} \rightarrow -\vec{x}$. Time reversal ($T$) reverses the direction of time: $t \rightarrow -t$. Parity and time reversal are spacetime operations. Charge conjugation (C) takes a fermion with a given spin orientation into an antifermion with the same spin orientation.

If we require that our theory be Lorentz invariant, then the CPT theorem states that our theory should also be symmetric under the combined operations of CPT. However, the separate operations of C and P are not good symmetries in chiral gauge theories, like electroweak theory. In addition, CP was originally believed to be a good symmetry, but in fact is violated in weak decays, as will be discussed in more detail in section 1.4.2.

1.3 Quantum Chromodynamics

1.3.1 Theory Overview

QCD is an SU(3) gauge theory of strong (color-charge-based) interactions. QCD is a theory of unbound quarks, yet at the same time, one result predicted by QCD is that unbound quarks do not exist in nature because of the nature of the variation of its coupling constant with distance. In quantum electrodynamics, the physical coupling constant, α, decreases (increases) with larger (smaller) distance scales due to the screening effect of charge:

$$\alpha(q^2) = \frac{\alpha(\mu^2)}{1 - \frac{\alpha(\mu^2)}{3\pi}\ln(\frac{q^2}{\mu^2})}, \tag{1.1}$$

where q is the momentum transfer (or inverse distance scale) involved in the interaction and μ is an energy scale or renormalization mass. In QCD, because gluons carry color charge, the opposite effect happens to the coupling constant, α_S, resulting in asymptotic freedom:

$$\alpha_S(q^2) = \frac{12\pi}{(11 \cdot 3 - 2 \cdot 6)\ln(q^2/\mu^2)}, \tag{1.2}$$

where we have assumed the number of colors is three and the number of flavors is six. In QCD, μ is typically taken to be Λ_{QCD}, the scale at which hadronization occurs.

In addition, we observe in nature the phenomenon of quark confinement, and while this does make sense within the context of strong coupling, it is not a necessary byproduct—and has yet to be analytically proved. It is quark confinement, or more

specifically, hadronization, which makes theoretical modeling of processes involving hadrons in the initial and final states difficult.

1.3.2 *B* Physics

Because our decay mode involves a photon radiating off of an internal quark in the B meson, an understanding of the B meson's internal dynamics is needed. In particular, the theoretical uncertainties are QCD related.

There are typically at least two scales involved in B physics: $\Lambda_{\rm QCD}$ and m_b (the b-quark mass). Given that m_b is relatively large, it is very useful as an expansion parameter.

The operator product expansion (OPE) allows us to separate the physics of a decay amplitude into two regimes, with a separation scale μ: the Wilson coefficients summarize all contributions from scales higher than μ, while the operators (vertices) encapsulate the physics at scales lower than μ. Typically one chooses μ to separate the short-distance contributions from the long-distance contributions.

One can use the OPE to write a general effective weak Hamiltonian, $\mathcal{H}_{\rm eff}$:

$$\mathcal{H}_{\rm eff} = \frac{G_F}{\sqrt{2}} \sum_n V_n C_n(\mu) Q_n \tag{1.3}$$

where V_n is a CKM factor, C_n is the Wilson coefficient at the renormalization scale μ, and Q_n is a local operator of the decay.

The amplitude $\mathcal{M}$ from initial state i to final state f is $\mathcal{M}(i \to f)$ is then be

written as:

$$\mathcal{M}(i \to f) = \langle f|\mathcal{H}_{\text{eff}}|i\rangle = \frac{G_F}{\sqrt{2}} \sum_n V_n C_n(\mu) \langle f|Q_n|i\rangle(\mu). \tag{1.4}$$

Note that the operator matrix elements have a renormalization scale and scheme dependence, but these dependencies must cancel out the dependencies in the Wilson coefficients because the amplitude is physical.

In semileptonic decays, one uses the vacuum-insertion approximation and assumes factorization. One then parameterizes the resulting matrix elements of quark bilinears with form factors.

1.4 Electroweak Interactions

1.4.1 Theory Overview

In electroweak theory, the weak currents, along with the electromagnetic current, form a symmetry group, $SU(2)_L \otimes U(1)_Y$. The weak isospin operators T^i $(i = 1, 2, 3)$ generate the symmetry group $SU(2)_L$, and the weak isospin current only couples to left-handed fermions. The weak hypercharge operator Y generates the symmetry group $U(1)_Y$. It is defined as $Y/2 = Q - T^3$,[3] where Q is the electromagnetic charge operator and T^3 is the third component of weak isospin.

In order for the theory to remain renormalizable, one cannot simply introduce mass terms for the gauge fields and fermions into the Lagrangian. It postulated that the nonzero vacuum expectation value of the Higgs field gives rise to electroweak

[3]Some authors use a convention in which the weak hypercharge includes the factor of one-half.

gauge symmetry breaking, resulting in particle masses. The masses are not predicted, but are inputs to the standard model in the form of Yukawa couplings.

At low order and low energies, one can perform calculations involving the weak interaction without needing the full power of a renormalizable theory. In particular, the momentum of the heavy gauge boson can be integrated out at tree level.

1.4.2 CKM Matrix and *CP* Violation

One of the interesting aspects of weak interactions is that there are flavor-changing charged currents. This is modeled by introducing the Cabibbo-Kobayashi-Maskawa (CKM) matrix, V, into the charged-current term ($\mathcal{L}_{\mathrm{CC}}$) of the electroweak Lagrangian:

$$\mathcal{L}_{\mathrm{CC}} = \frac{g}{\sqrt{2}} \left(\bar{u_{\mathrm{L}}}\ \bar{c_{\mathrm{L}}}\ \bar{t_{\mathrm{L}}}\right) V \gamma^{\mu} \begin{pmatrix} d_{\mathrm{L}} \\ s_{\mathrm{L}} \\ b_{\mathrm{L}} \end{pmatrix} W_{\mu} + \mathrm{h.c.}, \tag{1.5}$$

where q_{L} ($q = u, c, t, d, s, b$) are the left-handed quark mass-eigenstate fields, $g/\sqrt{2}$ is the dimensionless weak coupling, and V is:

$$V \equiv \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1.6}$$

In the standard model, the CKM matrix is unitary, implying nine parameters. However, five phases can be absorbed by the quark fields. This leaves four parameters:

three rotation angles and one physical, complex phase. The complex phase generates CP violation in weak decays.

There are six orthogonality conditions for the CKM matrix. The orthogonality condition between the first and third column of V is:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{1.7}$$

This orthogonality condition (as well as the other five) can be represented by a "unitarity triangle." Figure 1.1 depicts the unitarity triangle in the complex plane, while figure 1.2 shows the unitarity triangle after dividing the sides by $|V_{cd}V_{cb}|$.

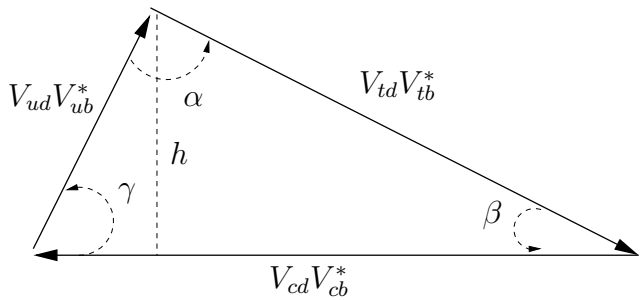


Figure 1.1. The standard unitarity triangle.

The angles of the triangle are defined by:

$$\alpha \equiv \left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \tag{1.8}$$

$$\beta \equiv \left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \tag{1.9}$$

$$\gamma \equiv \left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{1.10}$$

The height h is given by $|V_{ud}V_{ub}\sin\gamma|$, and the area is equal to $|V_{cd}V_{cb}|h/2$. These are

all physical quantities that can be measured by CP asymmetries in B decays.

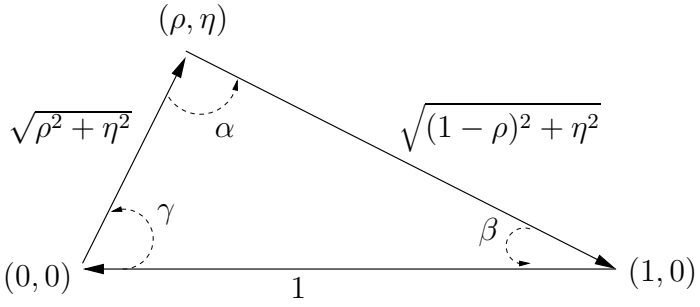


Figure 1.2. The standard unitarity triangle, after rescaling, in terms of the Wolfenstein parameterization.

In order to incorporate unitarity constraints and make clearer the relative measured magnitudes of its elements, one typically parameterizes the CKM matrix. One typical parameterization is the Wolfenstein parameterization. Using the Wolfenstein parameterization, the CKM matrix can be rewritten in terms of four parameters: λ, A, ρ, and η:

$$V = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4). \tag{1.11}$$

The parameters λ and A are relatively well measured, while ρ and η, corresponding to the coordinates of the unitarity triangle given in figure 1.2, are less so. One constraint on ρ and η comes from semileptonic B decays:

$$\rho^2 + \eta^2 = |V_{ud}V_{ub}|. \tag{1.12}$$

The other constraint comes from neutral B-meson mixing:

$$(1-\rho)^2+\eta^2 = |V_{td}V_{tb}| \propto \Delta m_B \cdot f_B^{-2}, \tag{1.13}$$

where V_{xy} are elements of the CKM matrix, Δm_B is the mass difference between the B_d mixing eigenstates, and f_B is the B-meson decay constant. Its theoretical value ranges from 175 to 370 MeV, depending on the method of calculation [5]–[8].

In the CLEO paper from 1997 detailing the first search for the radiative leptonic B decay [9], the state of phenomenological and experimental affairs was such that they believed that a measurement of the branching fraction of the radiative leptonic mode would be most useful for constraining f_B, assuming an external constraint on $|V_{ub}|$. However, f_B is now better known, and thus in the context of radiative leptonic B decays is useful primarily as an external input for constraints on our knowledge of the B meson's internal dynamics.

1.5 Fully Leptonic B Decays

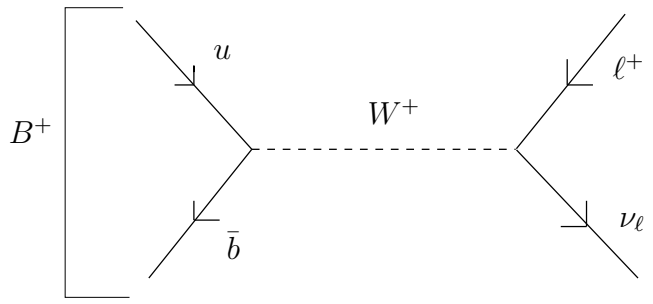


Figure 1.3. Tree-level Feynman diagram for $B^+ \to \ell^+\nu_\ell$.

Before we discuss the radiative leptonic B decay mode, it is instructive to first dis-

Table 1.1. Standard model estimates of the branching fractions of $B^+ \to \ell^+ \nu_\ell$.

Leptonic Channel	Branching Fraction
e	6.9×10^{-12}
μ	2.9×10^{-7}
τ	5.5×10^{-5}

cuss a closely-related decay mode, the purely leptonic mode, depicted in the Feynman diagram in figure 1.3. The tree-level branching fraction is:

$$\mathcal{B}(B^+ \to \ell^+ \nu) = \frac{G_F^2 |V_{ub}|^2}{8\pi} f_B^2 \tau_B m_B m_\ell^2 \left[1 - \frac{m_\ell^2}{m_B^2}\right]^2. \tag{1.14}$$

Unfortunately, the purely leptonic modes are helicity suppressed by a factor of m_ℓ^2, resulting in the estimated branching fractions shown in table 1.1.

Thus, the electron and muon channels are not observable at current luminosities. The tau channel analysis has the added difficulty of one or more extra missing neutrinos. However, recent analyses at *BABAR* and Belle are approaching the standard model value. For example, the Belle experiment has a recent measurement of the branching fraction of $B^+ \to \tau^+ \nu_\tau$:

$$\mathcal{B}(B^+ \to \tau^+ \nu_\tau) = (1.79^{+0.56}_{-0.49}(\text{stat})^{+0.46}_{-0.51}(\text{syst})) \times 10^{-4}, \tag{1.15}$$

which corresponds to [10]:

$$f_B = 0.229^{+0.036}_{-0.031}(\text{stat})^{+0.034}_{-0.037}(\text{syst})\,\text{GeV}. \tag{1.16}$$

1.6 Radiative Leptonic B Decays

1.6.1 Overview

In the radiative leptonic decay mode (depicted in the Feynman diagram in figure 1.4), the presence of the photon can remove the helicity suppression present in the purely leptonic decay mode if the photon is emitted from the B meson, rather than from the lepton.[4] Standard-model-based theoretical estimates of the radiative leptonic B branching fraction put it between 2.0×10^{-6} and 5.0×10^{-6} [11], which for the electron channel is several orders of magnitude higher than its corresponding purely leptonic mode branching fraction.

One can still calculate f_B, which would no longer be directly observed. But as it turns out, the radiated photon serves as a probe of the B-meson's internal dynamics, resulting in the introduction of moments of the B-meson distribution amplitude. These moments show up in other processes such as B to light-hadron decays, and therefore measuring the radiative leptonic B decay is more useful for these studies than for constraining f_B.

We will now discuss the standard model calculation of the branching fraction of the radiative leptonic B decay, in order to show how the branching fraction can depend on these moments. Assuming no higher-order diagrams involving loops between the

[4] We will ignore the radiation off the lepton ("internal bremsstrahlung") in the discussion below because it is also helicity suppressed.

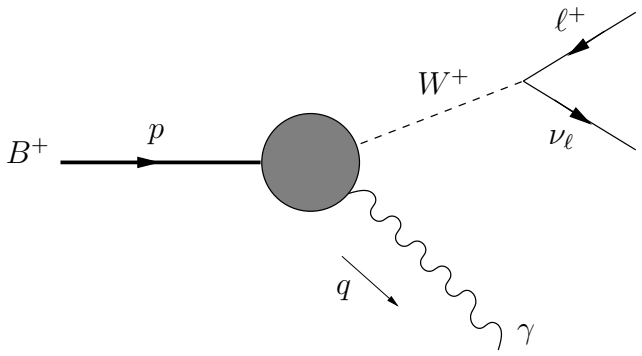


Figure 1.4. Tree-level Feynman diagram for $B^+ \to \ell^+ \nu_\ell \gamma$.

quarks and leptons, we can write the effective Hamiltonian, $\mathcal{H}_{\text{eff}}$, as:

$$\mathcal{H}_{\text{eff}} = \left(\frac{g}{2\sqrt{2}} |V_{ub}| H_\sigma \right) \frac{1}{m_W^2 - q^2} \left(\frac{g}{2\sqrt{2}} L^\sigma \right), \tag{1.17}$$

where $g/\sqrt{2}$ is the dimensionless weak coupling, H is the hadronic current, L is the leptonic current, m_W is the mass of the charged W boson, and q is the 4-momentum of the virtual W. We have also factored out $|V_{ub}|$ from the hadronic matrix element.

At energies well below the charged W mass ($q^2 \ll m_W^2$), we can define an effective 4-fermion vertex that is not dependent on q.[5] Setting $G_F/\sqrt{2} = g^2/(8m_W^2)$, we rewrite the effective Hamiltonian as:

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} |V_{ub}| H_\sigma L^\sigma. \tag{1.18}$$

Comparing with equation (1.3), one can see that the single local operator is the product of the two currents, with a Wilson coefficient of one. The amplitude is then obtained by calculating the matrix elements between initial and final states.

[5]By uncertainty principle arguments, the distance traveled by the W is too small to resolve.

The tree-level leptonic current is simply:

$$L^{\sigma} = \bar{u_{\nu}}\gamma^{\sigma}(1-\gamma^{5})u_{\ell}, \tag{1.19}$$

where u_{ℓ} and u_{ν} are the spinors for the lepton and lepton neutrino, respectively. The leptonic matrix element is also straightforward to determine.

The hadronic matrix element is a challenge to model theoretically because it involves nonperturbative strong-interaction physics. One approach to determining the hadronic current is to use factorization to separate the perturbative and nonperturbative physics. Quantities based on nonperturbative physics are universal in the sense that they show up in similar processes, so that one can extract the nonperturbative physics values from the measurement of one process to make predictions for another process, and vice versa. The factorization approach is explored in References [11] and [12] and we will summarize their calculations in the rest of this section.

First, one writes the hadronic matrix element in terms of two form factors, $f_V(E_\gamma)$ and $f_A(E_\gamma)$ (vector and axial-vector, respectively):

$$\begin{aligned} &\frac{1}{\sqrt{4\pi\alpha}}\langle\gamma(\epsilon^*,q)|\bar{b}\gamma_{\mu}(1-\gamma_5)u|B(v)\rangle \\ &= \varepsilon(\mu,\epsilon^*,v,q)f_V + i[\epsilon^*(v\cdot q) - q_{\mu}(\epsilon^*\cdot v)]f_A), \end{aligned} \tag{1.20}$$

where $\varepsilon(\mu,a,b,c) = \varepsilon_{\mu\alpha\beta\sigma}a^{\alpha}b^{\beta}c^{\sigma}$, q is the signal photon 4-momentum, ϵ is the signal photon polarization 4-vector , v is the signal B 4-velocity. We work in the B-meson rest frame, where its 4-momentum p is equal to $(m_B, \vec{0})$.

One then introduces light-cone coordinates $l = (l_+, l_-, \vec{l}_\perp)$, where $l_\pm = (l_0 \pm l_3)/\sqrt{2}$ and $\vec{l}_\perp = (l_1, l_2)$. We choose the signal photon momentum to lie along the l_- direction.

Now, in order to calculate the form factors, we start with the general factorized form of the hadronic current, which is:

$$H^\mu = \int d^4z \tilde{\Phi}(z)\tilde{T}(z) = \int \frac{d^4k}{(2\pi)^4} T(k) \int d^4z e^{ik\cdot z}\tilde{\Phi}(z), \tag{1.21}$$

where $T(k)$ is the hard-scattering amplitude, $\tilde{T}(z)$ is its Fourier transform, and $\tilde{\Phi}(z)$ is the B light-cone distribution function.

In Reference [12], they find that $T(k)$ depends only on k_+, so that the remaining components can be integrated out of $\Phi(k)$ (the Fourier transform of $\tilde{\Phi}(z)$). Using these results, one can then rewrite the hadronic current as:

$$H^\mu = \int \frac{dk_+}{2\pi} Tr\{\Phi(k_+)T(k_+)\}. \tag{1.22}$$

After some calculation, the form factors, including the leading Λ/m_b correction, are found to be:[6]

$$f_V(E_\gamma) = f_A(E_\gamma) = \frac{f_B m_B}{2E_\gamma}\left(\frac{Q_u}{\lambda_B} - \frac{Q_b}{m_b}\right), \tag{1.23}$$

where Q_i is the quark charge, m_b is the b quark mass, and λ_B is the first inverse moment of the B-meson light-cone distribution amplitude:

$$\lambda_B^{-1} = \frac{1}{\sqrt{2}} \int_0^\infty \frac{dk_+}{k_+}\Phi(k_+). \tag{1.24}$$

[6] Λ is the factorization scale.

The value of λ_B is typically taken to be on the order of $\lambda_{\rm QCD}$. For example, in Reference [13], they use $350 \pm 150\,\mathrm{MeV}$.

Now, we incorporate the above results in order to calculate the branching fraction. The differential branching fraction with respect to the lepton and photon energies (E_ℓ and E_γ, respectively) is given by:

$$\begin{aligned}\frac{d^2\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma)}{dE_\ell dE_\gamma} &= \frac{\alpha G_F^2 |V_{ub}|^2 \tau_B m_B^3}{16\pi^2}\{x[x^2 + 2y(y-x-1)+1][F_V^2 + F_A^2] \\ &\quad -2x(1-x)(1+x-2y)F_V F_A\},\end{aligned}$$

where $x \equiv 1 - 2E_\gamma/m_B$ and $y \equiv 2E_\ell/m_B$.

If we integrate over the lepton energy, we get the differential branching fraction with respect to the photon energy:

$$\frac{d\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma)}{dE_\gamma} = \frac{\alpha G_F^2 |V_{ub}|^2 \tau_B m_B^4}{48\pi^2} x(1-x)^3[F_V^2 + F_A^2]. \tag{1.25}$$

If we substitute in the form factors into the above expression and integrate over the photon energy, we obtain:

$$\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma) = \alpha \frac{G_F^2 |V_{ub}|^2}{288\pi^2} f_B^2 \tau_B m_B^5 \left(\frac{2}{3\lambda_B} + \frac{1}{3m_b}\right)^2. \tag{1.26}$$

Thus, if we take the value of f_B from lattice calculations, then a measurement of the branching fraction of the radiative leptonic decay allows us to constrain λ_B. However, there are higher-order moments in the B light-cone distribution amplitude

whose presence can affect the branching fraction calculations by as much as 50%.

Putting aside the issue of higher-order moments, one area in which a constraint on λ_B could be of value is in the study of light hadronic two-body B decays, as mentioned earlier. In these decays (written in general form as $B \to M_1 M_2$, where $M_{1,2}$ are light mesons such as pions), one of the factorized diagrams is shown in figure 1.5, and the contribution of this diagram to the hadronic matrix element can be written as:

$$\int_0^1 d\xi du dv T_i^{II}(\xi, u, v)\Phi_B(\xi)\Phi_{M_1}(v)\Phi_{M_2}(u), \tag{1.27}$$

where Φ_X is the light-cone distribution amplitude for meson X, and T_i^{II} is the hard scattering function obtained from perturbative calculations, and the index i runs over the local operators of the effective weak Hamiltonian [13].

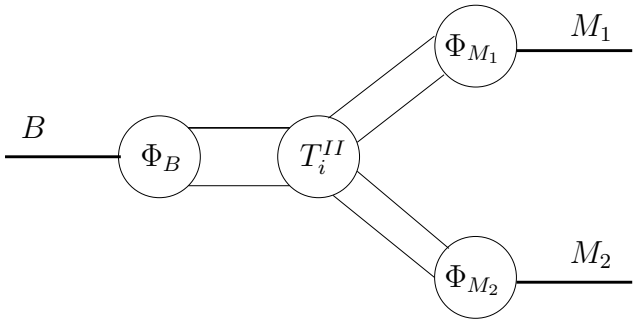


Figure 1.5. One contributing factorized diagram for the process $B \to M_1 M_2$.

Note that if factorization holds for both light hadronic two-body B decays as well as radiative leptonic B decays, then the light-cone distribution amplitude Φ_B should be the same for the two processes. Thus, any constraints that we can set on the moments of Φ_B (for example, λ_B) through constraints or measurements of the radiative leptonic branching fraction will allow theorists to make predictions about

light hadronic two-body decays.

1.6.2 Previous Measurements

In 1997, CLEO published results based on $2.5\,\mathrm{fb}^{-1}$ of data and set 90% confidence-level upper limits of 2.0×10^{-4} and 5.2×10^{-5} on the branching fractions of the electron and muon channels, respectively [9]. They performed a cut-based analysis that looked for a signal lepton and photon, assigned all other visible particles to the recoil B, and then reconstructed the energy and momentum of the signal neutrino. Some examples of their cut variables were: Fox-Wolfram 2nd moment (R2), total visible charge, tag lepton energy, recoil B mass and energy, and lepton-photon opening angle.

There is also a 2004 Belle conference paper which presents preliminary results based on $140\,\mathrm{fb}^{-1}$ [14]. They set 90% confidence-level upper limits of 2.2×10^{-5} and 2.3×10^{-5} on the branching fractions of the electron and muon channels, respectively. These results involve fitting to the lepton energy spectrum.

Chapter 2

Experimental Apparatus

2.1 Overview

The *BABAR* experimental apparatus consists of an e^+e^- storage ring-collider ("PEP-II") and a detector located at an interaction point. The experiment was primarily designed towards the goal of making precision measurements of *CP*-violating asymmetries in neutral B-meson decays. However, the high luminosity of PEP-II and the reconstruction capabilities of the *BABAR* detector allow us to make other precision measurements and perform searches for rare processes.

A brief description of the apparatus is as follows: In the laboratory frame, an electron beam of 9.0 GeV is made to collide head-on with a positron beam of 3.1 GeV at the interaction region, which is enclosed by a water-cooled beryllium beam pipe.[1] The detector apparatus, with approximate cylindrical symmetry, surrounds the beam pipe. This apparatus consists of an inner detector surrounded by a superconducting solenoid with a field strength of 1.5 tesla, and an outer detector that is an instrumented flux return (IFR). The inner detector, going from the interaction region out-

[1] Beryllium (low Z value of 4) is used to minimize multiple Coulomb scattering.

wards, is made up of a silicon vertex tracker (SVT), drift chamber (DCH), detector of internally reflected Cherenkov light (DIRC), and an electromagnetic calorimeter (EMC). A general layout of the detector can be seen in figure 2.1, with a more detailed longitudinal cross section provided in figure 2.2.

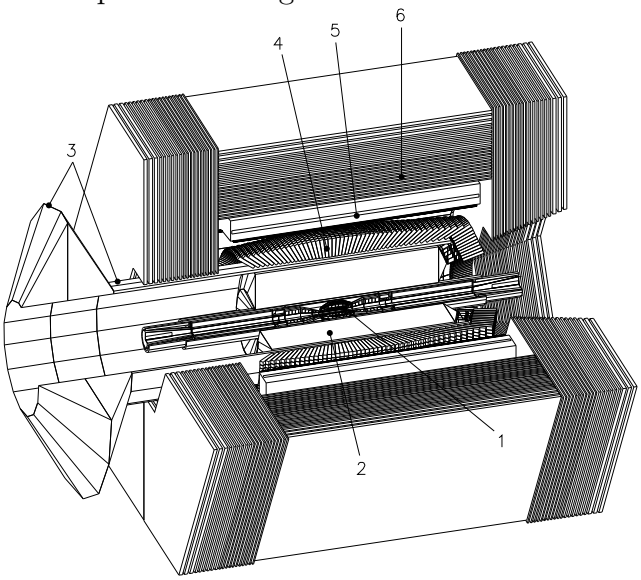


Figure 2.1. General layout of the *BABAR* detector, with the following key: 1 SVT, 2 DCH, 3 DIRC, 4 EMC, 5 Magnet, and 6 IFR.

The experimental apparatus has been described in detail elsewhere [15], so we will focus primarily on elements of the apparatus pertaining directly to the reconstruction of $B^+ \to \ell^+ \nu_\ell \gamma$.

2.2 PEP-II

PEP-II is a high luminosity asymmetric e^+e^- collider, with a center-of-mass (CM) energy generally around 10.58 GeV,[2] the mass of the spin-1 $\Upsilon(4S)$ resonance.[3] The

[2] "CM" will refer to the center-of-mass frame of the $\Upsilon(4S)$, unless stated otherwise.

[3] We label this energy, the "on-peak" or "on-resonance" energy. "Off-resonance" or "off-peak" samples have also been produced, by tuning the beam energies such that the CM energy is below the mass of the $\Upsilon(4S)$. These samples are devoid of $B\overline{B}$ pairs.

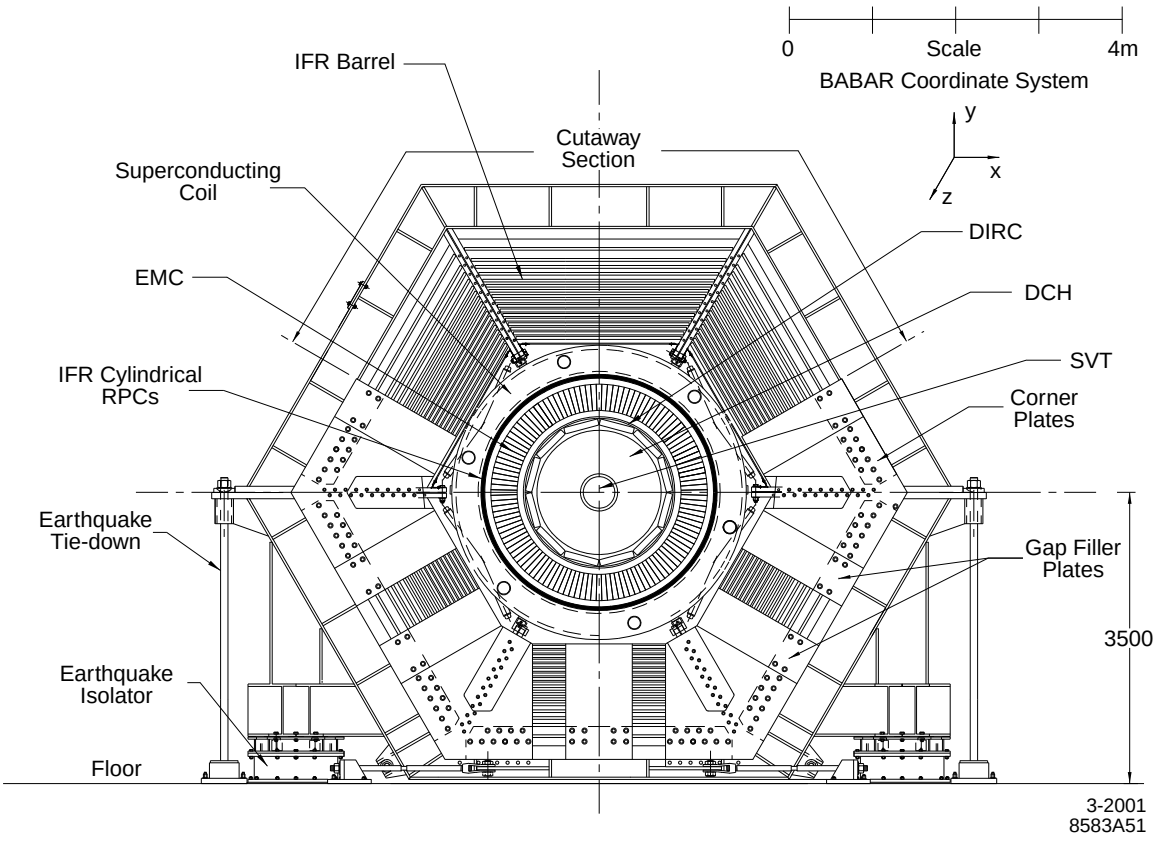


Figure 2.2. Longitudinal cross section of the *BABAR* detector.

$\Upsilon(4S)$ decays more than 96% of the time into spin-0 B pairs, of which approximately 50% are neutral and 50% are charged. Thus at this energy one obtains a large and relatively clean sample of B mesons. table 2.1 gives the effective cross section (including detector acceptance) of e^+e^- to various channels at the PEP-II on-peak CM energy.

Table 2.1. Effective cross sections of e^+e^- at 10.58 GeV. All values are approximate values obtained from the *BABAR* Physics Book [16].

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

B mesons are produced almost at rest in the $\Upsilon(4S)$ frame, with momenta close to 320 MeV.[4] In a conventional, symmetric collider, this implies a flight distance for the B^0 of about 30 micrometers in one lifetime (1.5 ps). Given the resolution limits of feasible experimental configurations, it would thus be rather difficult to reconstruct B decay vertices at the level of precision needed for time-dependent *CP* analyses.

The *BABAR* experiment circumvents these constraints by using asymmetric beam energies to produce the $\Upsilon(4S)$ in flight: a 9 GeV electron collides with a 3.1 GeV positron, resulting in a Lorentz boost to the $\Upsilon(4S)$ of $\beta\gamma \approx 0.56$ ($\gamma \approx 1.146$). The B^0 now travels an average distance of 250 μm in the laboratory, which is also the average

[4] We will use natural units throughout this thesis, setting $c = 1$.

separation between the B pairs.

A negative consequence of these asymmetric lab energies is that, relative to a symmetric energy collider, the forward boost "pushes" more particles outside of the detector acceptance region, resulting in a loss of charged track and neutral reconstruction efficiency. For example, the forward detector acceptance is around >0.3 rad (>17°) in the laboratory frame,[5] and this corresponds to an acceptance of >0.5 rad (>30°) in the CM frame for a massless particle.

The design luminosity of PEP-II is $3 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, with a design beam current of 0.75 A (2.15 A) for the high (low) energy ring. This design luminosity has been greatly exceeded. For example, during the data-taking period pertaining to this analysis dataset (October 1999–August 2004), a peak luminosity of $9.2 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ was achieved, more than three times the design luminosity.[6] In addition, the integrated luminosity has exceeded design goals.

2.3 Silicon Vertex Tracker

The silicon vertex tracker is a precision tracking device designed to reconstruct charged particle trajectories near the interaction region. The precision reconstruction allows one to measure the position of B decay vertices. The SVT also contributes particle ID information through energy loss (dE/dx).

The SVT has five layers of double-sided silicon strip sensors. The layout of the

[5] There is no exact value for the acceptance, as this depends on not just the physical dimensions of the detector, but also the ability of the apparatus and our software to correctly reconstruct particles whose trajectories may lie near the edges of the acceptance region.

[6] One measures the absolute luminosity from e^+e^- and $\mu^+\mu^-$ production rates.

SVT layers is shown in figure 2.3.

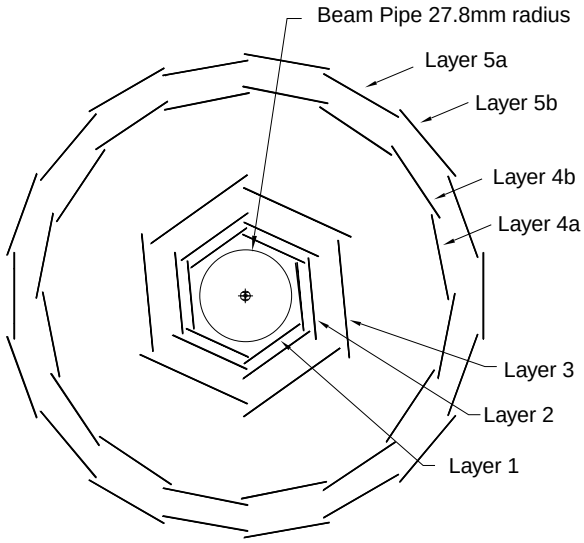


Figure 2.3. Transverse cross section of the SVT and the beam pipe. The five layers are shown, as well as the layout of the modules within each layer.

Each layer is organized into modules, which in turn are made up of wafers that contain the strip sensors. The strips collect and measure the induced charge from charged particles that ionize the sensors. There are a total of 150,000 channels.

The strips on one side of the sensor run parallel to the beam axis and measure the azimuthal angle ϕ with a resolution of 10–30 μm, while the strips on the other side of the sensor are oriented transversely to the beam axis and measure z, with a resolution of 15–50 μm. The cosine of the SVT polar angle acceptance in the laboratory frame is between -0.87 and 0.96.

A charged particle with sufficient transverse momentum will record a hit in each of the five layers. For a particle (a slow pion, for example) with transverse momentum less than 120 MeV, the SVT serves as a stand-alone tracking device because the particle will either never enter the DCH, or will not record enough hits in the DCH

to be reconstructed.

One can convert the calibrated pulse height into a measurement of dE/dx. Given that there are only 10 "hits" (five layers times two) maximum, with a minimum requirement of 4 hits, the SVT dE/dx is less powerful than the DCH dE/dx (up to 40 hits) for tracks that make it through both. The SVT dE/dx information provides a 2σ separation between pions and kaons at momenta up to 500 MeV, and between kaons and protons at momenta beyond 1 GeV.

2.4 Drift Chamber

The drift chamber is the primary charged particle detection/tracking device, with secondary functions in triggering and particle identification. Like the SVT, it works on the principle of ionization loss: charged particles ionize the gas inside the DCH, and this ionization is measured by wires running along the length of the DCH.

The DCH is cylindrical in shape, with a length of approximately 2.8 m, inner radius of 0.24 m, and outer radius of 0.81 m. The gas is a mixture of He (80%) and isobutane (20%).

The internal measurement unit of the DCH is the drift cell—a single sense wire surrounded by six field wires, thus resulting in a hexagonal cell shape, which provides a good balance between packing and approximate circular symmetry. The approximate dimensions of the drift cell are 11.9 mm (radial) by 19.0 mm (azimuthal). The field wires are at ground potential, while the sense wires are at a high positive voltage

(1960 V).[7] This results in an avalanche gain of about 5×10^4.

The DCH consists of 7104 drift cells, arranged in four groups of ten superlayers each (40 layers total). Within each superlayer, each layer has the same number of cells, and wires have the same stereo angle, as illustrated in figure 2.4. There is staggering between layers so as to provide left-right ambiguity resolution within a superlayer.

Each drift cell through which the charged track passes outputs a signal pulse that is converted into a drift time and integrated charge. Using calibrated parameters, one then converts these values into drift distance and dE/dx, respectively.

The DCH has a tracking efficiency of about $98 \pm 1\%$ per track above 200 MeV. The average spatial resolution of the DCH is 125 μm. DCH dE/dx provides better than 2σ separation between pions and kaons at momenta up to 700 MeV.

2.5 DIRC

The detector of internally reflected Cherenkov light (DIRC) is a ring-imaging Cherenkov detector, providing particle ID by determining a particle's velocity. It consists of a radiating medium of synthetic fused silica, which has a mean index of refraction of 1.473. The silica take the form of long, thin, rectangular bars. At one end of the bars is a water-filled expansion region, instrumented with photomultiplier tubes. The other end of each bar has a mirror, to direct light toward the opposing end.

The idea behind the DIRC is as follows: A charged particle of velocity β traversing

[7] Some of the earliest data recorded by *BABAR* used a DCH operating voltage of 1900 V.

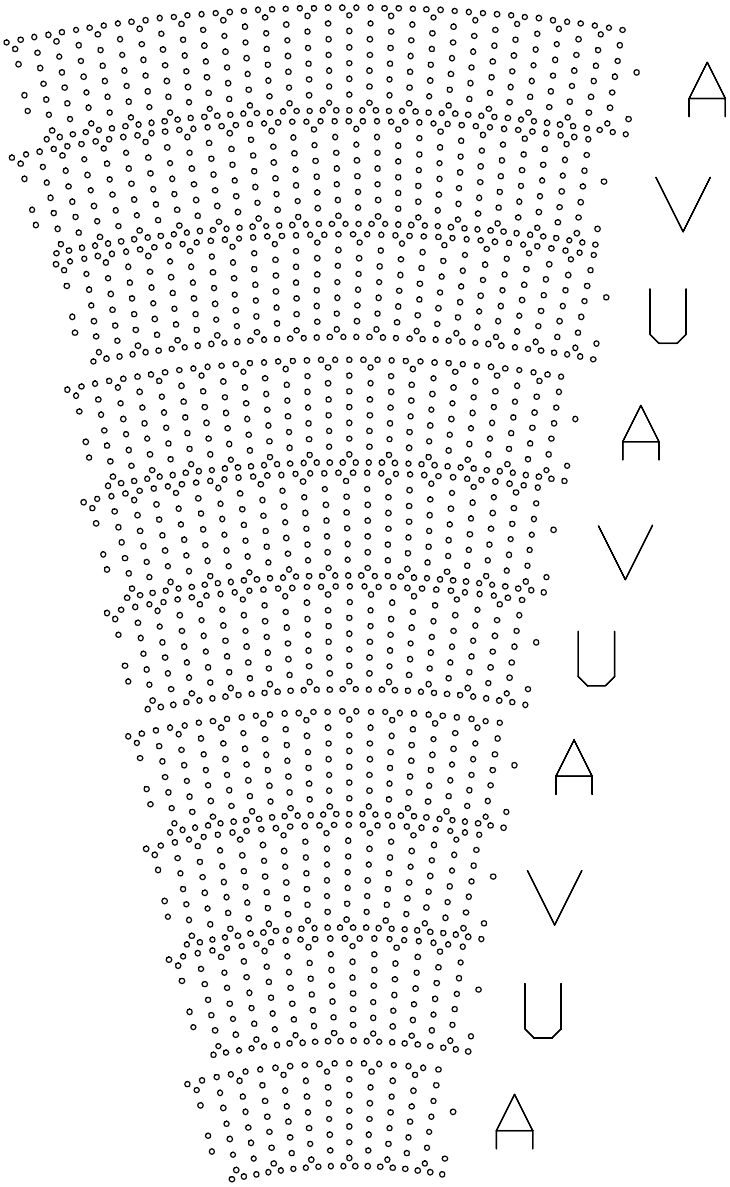


Figure 2.4. DCH cell layout. There are ten superlayers, each with a specific wire orientation: (A) axial, and (U,V) stereo.

a medium with index of refraction $n(\lambda)$ will coherently emit light if $\beta > 1/n$. The angle of Cherenkov photons, θ_C, is given by the expression: $\cos\theta_C = 1/(\beta n)$. For particles of a high enough velocity, some of these photons will be internally reflected to the instrumented end of the bar, with their angle preserved and measured by the photomultiplier tubes. A schematic of the DIRC, illustrating the path of Cherenkov photons, can be seen in figure 2.5.

The value of θ_C, along with the momentum measurement from the tracking system, allows one to calculate a likelihood of a particular particle ID hypothesis. In addition, the number of photons emitted is proportional to $\sin^2\theta_C$, and is used as an input in the likelihood. At 4 GeV, the DIRC provides about a 2.2σ separation between kaons and pions.

2.6 Electromagnetic Calorimeter

The EMC is designed to measure energy and angular resolution, and is the only subsystem that detects neutral particles such as photons. It also plays an important role in electron identification.

The detection unit of the EMC is the crystal. The *BABAR* EMC uses thallium-doped cesium iodide crystals. Particles interact with the calorimeter material (strong interactions, bremsstrahlung, pair production, etc.), losing energy. An electromagnetic shower, which is a cascade of alternating bremsstrahlung and pair production processes, forms, and the photons in this shower are detected by silicon photodiodes.

The EMC consists of a barrel and an endcap. The barrel has 48 rings of 120

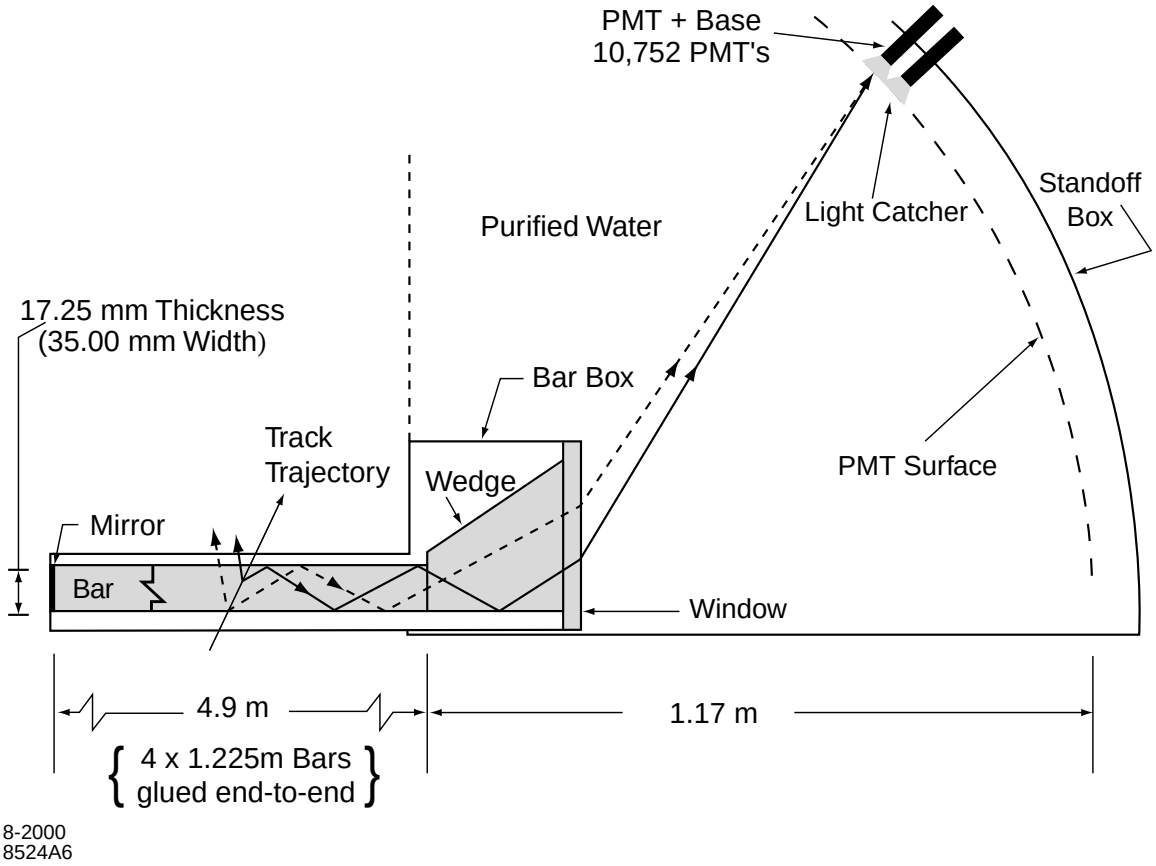


Figure 2.5. Schematic of the DIRC, illustrating the path of Cherenkov photons.

crystals each, and the endcap has 820 crystals in 8 rings, giving a total of 5760 crystals. Figure 2.6 illustrates the layout of the EMC crystals via a longitudinal cross section.

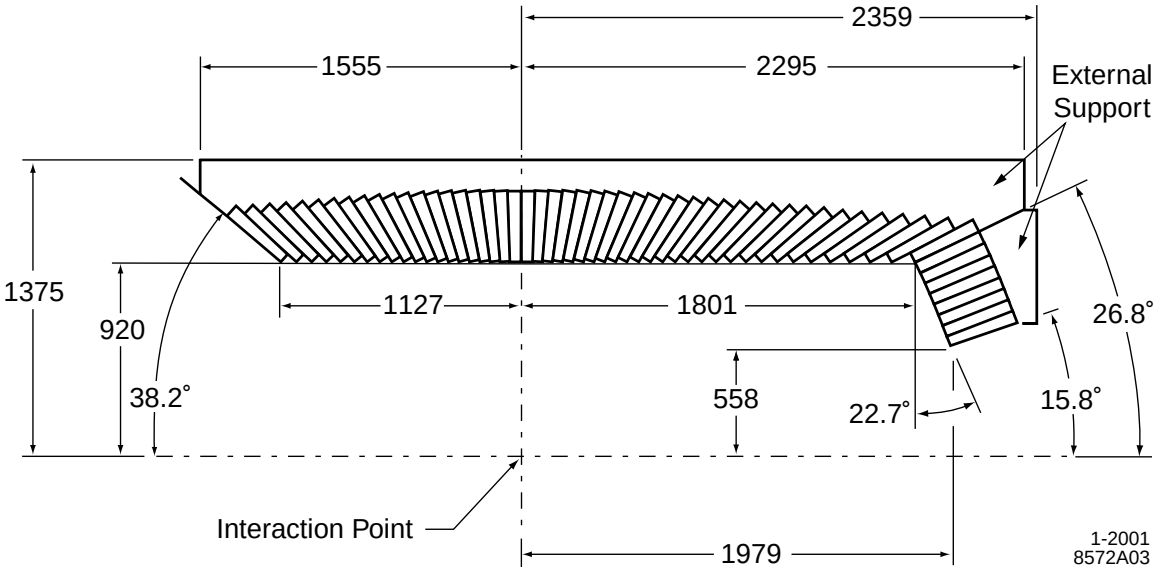


Figure 2.6. A longitudinal cross section of the top-half of the EMC, with all dimensions in mm.

The energy resolution of the EMC is:

$$\sigma_E/E = \frac{(2.32 \pm 0.30)}{\sqrt[4]{E(\,\mathrm{GeV})}}\% \oplus (1.85 \pm 0.12)\%. \tag{2.1}$$

The first term on the left arises primarily from photon statistics, as well as from electronic noise and background photons. The second term on the left is due to nonuniformity in the light collection, calibration errors, and the effects of the material between the crystals.

The angular resolution of the EMC is:

$$\sigma_\theta = \sigma_\phi = \left(\frac{3.87 \pm 0.07}{\sqrt{E(\,\mathrm{GeV})}} + 0.00 \pm 0.04 \right) \mathrm{mrad}. \tag{2.2}$$

This resolution is due to transverse crystal size and distance from the interaction point. All things held equal, a smaller transverse crystal size provides better photon separation and resolution. However,the choice of transverse crystal size itself (typically between 5 to 6 cm) was determined by the Moliére radius (approximately 3.8 cm) of CsI(Tl) as well as the readout complexity (more crystals means more channels).

EMC reconstruction uses pattern recognition to find and separate energy deposits. The reconstruction starts with "digis", which are crystals with at least 1 MeV of measured energy. A "cluster" is a 10 MeV digi with some surrounding digis. A "bump" is a local maxima in a set of clusters. Finally, one can either associate bumps with charged tracks, or identify the bump as a neutral candidate.

Particle ID is based upon shower shapes, and E/p. In particular E/p is close to one for electrons, but not other charged particles.

2.7 Instrumented Flux Return

The IFR, the outermost detector, has the dual purpose of being the flux return for the magnetic field as well as a detector of particles that penetrate through the EMC. The IFR plays an important role in muon identification, allowing us to separate pions and muons.[8]

[8]Pions and muons have a similar mass, so the DIRC is not very useful.

It consists of a barrel, and backward and forward endcaps. A general illustration of the IFR components can be seen in figure 2.7. The barrel consists of 19 resistive plate chamber (RPC) detector layers sandwiched between a total of 65 cm of iron. The endcaps each consist of 18 RPC layers sandwiched between a total of 60 cm of iron. The iron tends to absorb hadrons, while allowing muons to penetrate. The RPC strips in the barrel provide ϕ (19.77 to 33.5 mm pitch) and z (38.5 mm pitch) information, while the RPC strips in the endcaps provide x (28.4 mm pitch) and y (38 mm pitch) information.

An RPC consists of a 2 mm gap filled with an argon-freon-isobutane mixture, which is surrounded by bakelite plates, graphite, a PVC insulator, then capacitive readout strips. There is a voltage of 8 kV between the graphite surfaces, resulting in a quenched spark when ionizing particles pass through. The bakelite surfaces facing each other were coated with linseed oil in order to smooth out the electric field.[9]

The muon detection efficiency is close to 90% for muons in the lab momentum range of 1.5 to 3 GeV. The pion misidentification rate is about 6–8%.[10]

2.8 Data Acquisition and Computing

The high luminosity environment provided by PEP-II requires, in turn, a computing system with the capability to handle high rates and large volumes of data. Thus, the primary functions of the data acquisition and computing system are to acquire,

[9]Unfortunately, the coating was not applied uniformly enough to work effectively. This resulted in a degradation in the performance of the IFR over time.

[10]There is about a 5% practical limit due to pions that "punch-through" (penetrate all layers) and pions that decay into muons before entering the IFR.

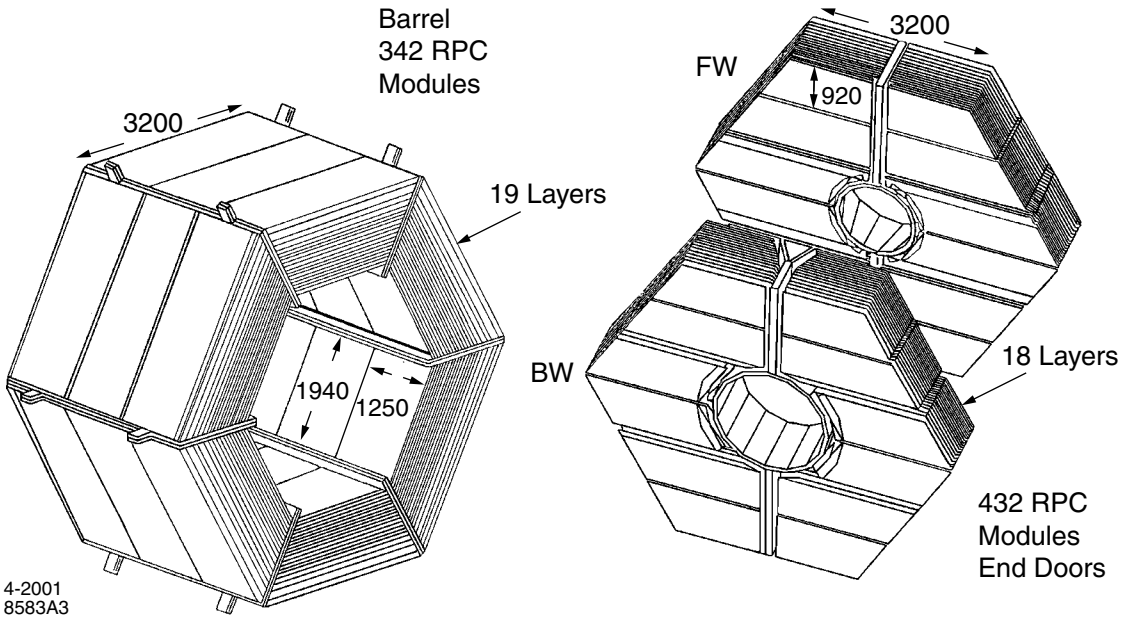


Figure 2.7. Views of the IFR barrel (left) and forward and backward endcaps (right).

process, store, and manage the data. A secondary function is to provide monitoring of the detector systems.

The *BABAR* data acquisition and computing system can be divided into three areas: online, prompt reconstruction, and offline. The online system can be roughly broken up into major subsystems: online dataflow (ODF), online event processing (OEP), logging manager (LM), online detector control (ODC), and online run control (ORC). There is also the trigger system, composed of a hardware (L1) and software (L3) trigger.

ODF handles the acquisition of raw digitized data starting from the detector subsystem electronics all the way to the assembly of an event.[11] OEP handles the processing of an event and passes the event to the logging manager. ODC handles

[11] An event can be thought of as the total collection of readout detector data in a particular time window.

the control and monitoring of the environmental conditions of the detector systems. ORC manages the various systems by allowing one to sequence a set of detector/data-taking states whose transitions can be manually controlled by a GUI or automatically triggered by certain running conditions.

We will now discuss in more detail some aspects of the online computing system.

2.8.1 Trigger

The primary purpose of the trigger system is to select with extremely high efficiency events of interest, in particular B events, while rejecting background events. For example, it is important to reject beam-induced background by identifying events with tracks that do not appear to originate from the interaction point.

The final design accepted event rate is around 120 Hz.

2.8.1.1 Level 1 Trigger

The hardware-based L1 trigger assembles DCH, EMC, and IFR information in order to make a triggering decision with low latency. The L1 trigger works at continuous sampling mode rate of 3.7 MHz.

The DCH trigger (DCT) is the primary charged-track trigger. It starts by looking for track segments, which are then passed in parallel to both a track assembler and a transverse track momentum discriminator. The output from the DCT are "trigger primitives," which are summary data indicating the presence of long tracks (tracks that reach all 10 DCH superlayers), short tracks (tracks that reach 5 DCH superlayers), and high transverse-momentum tracks.

The EMC trigger (EMT) is the energy trigger. It looks for various patterns of energy deposits, and distinguishes between minimum-ionizing particles (MIPs), intermediate energy particles, and high energy particles.

The IFR trigger (IFT) triggers on $\mu^+\mu^-$ and cosmic ray events. Seven specific trigger primitives are passed to the global trigger. These trigger patterns involve the detection of a signal in various sectors of the IFR and correspond to various muon event topologies.

The global trigger (GLT) receives trigger primitives from the DCT, the EMT, and the IFT every 134 ns (7.46 MHz), processes them to form triggers,[12] and then delivers them to the fast control and timing system (FCTS), where the final trigger decision is made.

The typical output rate of L1 is 1 kHz.

2.8.1.2 Level 3 Trigger

The L3 trigger is technically a part of OEP, but is an important enough component of the data acquisition process that it merits a separate discussion.

The L3 trigger is a software-based trigger that runs on the OEP computing farm (typically 32 nodes). Each event passing L1 is routed to one of the nodes for processing. The fact that L3 is software-based means that there is a great deal of flexibility in configuring it.

One important role of L3 is to reduce beam-induced background events produced

[12] A trigger in this context is some combination of trigger primitives that is an indication of a particular physics process or a set of physics processes.

near the IP, because L1 does not have enough vertexing resolution to reject such events. Another is to reject Bhabha events, which dominate, by using EMC information to identify electrons.

The DCH L3 algorithm, L3DCH, refines and augments the L1 selection method, starting with the track segments from L1, and making use of DCH information. The track-finding involves using a look-up table containing hit patterns of Monte Carlo generated tracks. If a track is recognized, then the information is passed to a track-fitting algorithm, which can perform vertexing.

The EMC L3 algorithm, L3EMC, finds crystals with significant energy deposits, and then identifies clusters. Clusters that pass a minimum energy criterion (100 MeV) are kept, along with their shape information.

At the typical luminosities used to collect our data, the L3 output rate of 120 Hz can be roughly partitioned into 70-90 Hz for physics events and 30-50 Hz for special (calibration/diagnostic) events. Desired physics events (hadrons, $\tau\tau$, and $\mu\mu$) make up about 13% of the total event output.

The final selection efficiencies for various physics processes are given in table 2.2.

Table 2.2. The final (L1 and L3 combined) selection efficiencies for various physics processes. The values are computed from Monte Carlo simulations.

Process	Efficiency(%)
$b\overline{b}$	>99.9
$c\overline{c}$	98.9
uds	95.8
$\tau^+\tau^-$	92.0

2.8.2 Online Dataflow

First, at the detector level, signals are measured and digitized by each subsystem's front-end electronics (FEE). There are 234K channels at this level. These data are stored in buffers, and in the case of the DCH and IFR, passed to the L1 trigger.

In all of the subsystems except the EMC, if the L1 trigger accepts the event, readout modules (ROMs) then read the data from the buffers of the FEEs, in time window around the trigger signal. The EMC ROMs, on the other hand, continuously read out from EMC FEE into a buffer pipeline and the EMT. This is the "segment" level, with 133 ROMs total.

A master ROM, called the "Slot 0 ROM," assembles the segments from the 133 ROMs into "fragments." At the fragment level, there are 23 Slot 0 ROMs.

Finally, the fragments are passed to OEP, where they are assembled into a single event.

2.8.3 Online Event Processing

OEP assembles event fragments into events, processes them, and then passes accepted events to the logging manager, where their data are written to disk. As mentioned in the L3 discussion, OEP runs simultaneously on a computing farm of 32 nodes. This multinode system allows us to minimize dead time.[13]

In addition to the processing performed by the L3 trigger, OEP also handles the monitoring of event data ("fast monitoring"). We will discuss this now in more detail.

[13]L3 dead time can lead to lost physics events if the detector hardware buffers are overrun and are unable to pass event information to an L3 farm node. In addition, a multinode system is more robust to node failures than a single-node system.

2.8.3.1 Distributed Histogramming

In principle, while one could monitor the output from a single node,[14] one gets better statistics (and thus potentially faster detection of problems) when one combines the output of many nodes.[15]

Multinode, or *distributed*, histogramming is implemented in the *BABAR* online computing environment via a set of C++ software packages, all with the common prefix DHP (Distributed Histogramming Package). Also, DHP solves the problem of interprocess communication (IPC), allowing us to separate the various aspects of histogram interfacing and management. DHP software is based on CORBA.[16] Any object (a histogram, for example) which will be passed around needs to be implemented within the DHP framework.

We can break up DHP processes into four types: accumulator, reader, requestor, and consumer. The DHP accumulator process is the process that books and fills histograms (or other objects). An accumulator runs on each of the farm nodes. Paired with the accumulator is a DHP reader process, which reads out histogram information. The DHP requestor process runs on a single node and is responsible for coordinating the reader processes. Its job within the context of OEP fast monitoring is to combine the histograms from the different farm nodes. Finally, a DHP consumer (or multiple DHP consumers) can query a requestor to obtain histograms.

[14] This was indeed the case for the first few years of data taking.

[15] One could also in principle page between the outputs of different nodes, but this would be tedious.

[16] CORBA (common object request broker architecture) is a distributed object computing infrastructure that allows one to pass high-level objects across processes, nodes, and networks.

2.8.3.2 Fast Monitoring

During data-taking periods, data are, in general, continuously acquired, with a team of typically two on-site "shift-takers" responsible for very general detector operations for a period of eight hours. Part of their responsibility is detector monitoring, using the fast-monitoring system. Using a Java-based browser written to interface with DHP objects, the shift-taker views HTML pages displaying DHP histograms and scalers,[17] along with descriptions and alerting procedures.

The fast-monitoring system also has the important role of allowing a shift-taker to perform data-quality monitoring. For instance, there may be problems that do not require immediate attention or the need to stop data taking, but might nonetheless warrant attention. In these cases, a shift-taker would make an entry in the logbook so that the run quality group would be alerted.

One of the components of the fast-monitoring system is the "automated" monitoring system. The purpose of the automated system is to supplement the manual monitoring process by alerting the shift-taker when potential problems occur in the detector subsystems.

The primary unit of automated monitoring is the comparison record. The comparison record defines a single automated test and its alerting response or responses to certain outcomes of the test. A typical type of test involves comparing two distributions: one live histogram vs. one reference histogram from a file, or a histogram vs. a function. The comparison can either perform a statistical comparison of entire

[17] Scalers are essentially time series. In their simplest form, they are values measured at a particular time. However, they can also be averaged or integrated values up to a particular time.

spectra, or look for individual bins with suspicious behavior (such a comparison is useful for occupancy histograms). Another type of test involves either calculating and checking the standard moments of a histogram, or fitting a histogram and then checking the results of the fit parameters.

If a test indicates that there is a problem, then the comparison record handles the alerting response(s), which can have varying levels of severity. There are three main types of responses: email, log file, and error logger. The email response is configured with an email address and an email message template (the level of severity might be filled in later by the automated monitoring process). A log file response involves writing error messages to a text file.[18] The error logger response utilizes an error logger that writes to a database. A GUI-based error browser (displaying a list of error messages) can perform specific queries on this database, but typically runs in continuous-update mode, where new error messages show up on top. This error browser is viewed by the shift-taker during periods of data taking. Levels of severity are indicated by colors (nonflashing or flashing).

The comparison records are set up by detector subsystem experts using configuration files, and can be adjusted at any time.[19] One can also specify in the configuration file the frequency of testing. The automated monitoring application reads the configuration files, sets up the comparison records, and then runs comparisons at the specified or default (2 minutes) comparison frequency.

[18] This is more useful for less severe problems.

[19] Any changes to comparison records involves restarting the automated monitoring process, and this can be done without restarting any other online component. One can also update reference histograms at any time.

Chapter 3

Analysis Overview

In a large sample of e^+e^- events reconstructed in the *BABAR* detector, corresponding to an integrated luminosity of $210.49\,\mathrm{fb}^{-1}$, we search for the decay $B^+ \to \ell^+\nu_\ell\gamma$, where $\ell = e, \mu$. The primary challenges are the following: (1) reconstructing the missing signal neutrino,[1] (2) understanding the diverse sources of background, removing them when possible, and (3) extracting the signal branching fraction (BF).

In order to give the reader a general idea of what follows, we will sketch out a road map of the rest of this dissertation. The rest of this chapter will discuss some of the broader analysis issues. In chapter 4, we will discuss our Monte Carlo (MC) model for the signal decay mode, followed in chapter 5 by a discussion of the various data and MC samples used.

From there, we move into the finer details of our analysis, starting in chapter 6 with a description of our selection criteria. In chapter 7 we discuss event reconstruction and preliminary event selection, followed in chapter 8 by a description of the optimization of our selection criteria.

[1] We use the term "signal" to refer to the B meson which decays into our signal mode, as well as its products, and "recoil" to refer to the other B meson.

We will then discuss the signal extraction fit procedure in chapter 9. Before presenting the results of the fits, we perform validation studies using control samples (chapter 10), and a sideband region fit (chapter 11).

In chapter 12, we present the fit results. We evaluate our systematic errors in chapter 13. We will present the full results, conclusions, and outlook in chapter 14.

3.1 Analysis Procedure

3.1.1 Procedure Overview

Our analysis is based on an inclusive reconstruction of the signal neutrino. We assign the highest center-of-mass (CM) energy lepton and highest energy CM photon to be our signal lepton and signal photon,[2] form a recoil B candidate from the remaining tracks and neutral clusters,[3] and then reconstruct the missing momentum and energy as a candidate signal neutrino. Using a number of selection variables that include event shape variables, signal-side (kinematics), and loose cuts on the recoil B, we maximize signal sensitivity via cut optimization.[4] The signal is then extracted using a fit to the event yields in a set of signal and sideband regions. Since both the electron and muon decay modes are rare, and our signal extraction procedure is potentially sensitive to fluctuations in our selection variable distributions, we blind ourselves to the signal extraction regions in the data, so as not to bias ourselves.

[2] As stated previously, "CM" refers to the center-of-mass frame of the $\Upsilon(4S)$, unless it is explicitly stated otherwise.

[3] "Recoil" will be used throughout this thesis to refer to the nonsignal B and its decay products.

[4] The term "cuts" is used interchangeably with the term "selection criteria." In addition, the term "loose" means that some or most of the selection criteria have been relaxed.

Unlike in a tagged analysis where one has control over the recoil (but with the cost of lower signal efficiencies), our inclusive analysis, with its loose recoil B reconstruction chosen to improve signal efficiency, potentially results in much higher backgrounds from a variety of sources, including continuum events. Ideally, one would be able to correctly model all of these different sources with Monte Carlo (MC) so that one could take advantage of potentially higher MC statistics in a signal extraction fit.

However, in the electron channel, it appears that the continuum MC background cannot be used to accurately estimate the shape of the continuum background in the data. In an earlier attempt of this analysis based on a smaller dataset with earlier reconstruction software, we observed a large excess of events in the off-peak data sample compared to the continuum (non-$B\overline{B}$) MC prediction. Based on a number of tests, we suspect that this discrepancy is due to two-photon background, which will be described in more detail in section 3.2. One thing to emphasize here is that the *BABAR* experiment does not have a comprehensive generic two-photon MC generator, so that we cannot adequately model two-photon physics.

We also observed a smaller, but still sizable excess of off-peak events in the muon channel. We were unable to understand these events in terms of unmodeled properties of the muon identification system, and therefore suspect an additional source, such as two-photon events. In the absence of a generic two-photon MC generator, it is difficult to confirm our interpretations of these excess events. Regardless of interpretation, we have chosen to use off-peak data in the muon-channel signal extraction as well.

Our analysis was designed to take these observations into account. In the ab-

sence of an adequate full continuum MC model, the off-peak data sample is used to determine the continuum background. However, because the ratio of off-peak to on-peak data is approximately 1:10, using the off-peak data introduces sizable statistical uncertainties into the background determination and signal extraction.

In the optimization of the selection criteria in these analyses we rely primarily on MC samples. We avoid bias in determination of signal and background efficiencies by splitting our samples into two—one half is used for cut optimization ("training sample") and the other for measuring the signal efficiency or background contamination in our signal extraction fit ("validation sample"). Due to the very small quantity of off-peak data available and the need to blind it, we are forced to use continuum MC in optimizing the analysis even though it will not be used in the final signal extraction. Because our cut variable distributions may differ in shape and size between continuum MC and the blinded off-peak data, this means that our cut optimization may not be optimal—but it is still unbiased.

Before the final unblinding, in order to convince ourselves of the validity of our analysis procedure, we study data and MC in control samples and with fits to an unblinded sideband region.

3.1.2 Neutrino Reconstruction

We now provide an overview of our neutral reconstruction technique, and introduce the idea of "scaling".

We label the B that decays via the signal mode the "signal B", and the other B

the "recoil B." One way to reconstruct the recoil B is to sum up the 4-momenta of all track and neutral candidates in the event (excluding those of the signal lepton and photon). The procedure is to sum up the component 4-vectors in the lab frame and boost the summed 4-vector into the CM frame. We call this the "unscaled" recoil B candidate.

Using the CM momentum 3-vectors of the unscaled recoil B, signal lepton, and signal photon, we can reconstruct the missing CM momentum 3-vector $\vec{p}_{\rm miss}$:

$$\vec{p}_{\rm miss} = -(\vec{p}_{{\rm rec}B,{\rm unscaled}} + \vec{p}_{\ell\gamma}), \tag{3.1}$$

where $\vec{p}_{{\rm rec}B,{\rm unscaled}}$ is the 3-momentum of the unscaled recoil B and $\vec{p}_{\ell\gamma} \equiv \vec{p}_\ell + \vec{p}_\gamma$ is the sum of the 3-momenta of the signal lepton and photon. We take this as the 3-momentum $\vec{p}_{\nu,{\rm unscaled}}$ of the "unscaled" neutrino candidate.[5]

The energy of the unscaled neutrino candidate is:

$$E_{\rm miss} = 2E_{\rm beam} - E_{\ell\gamma} - E_{{\rm rec}B,{\rm unscaled}}, \tag{3.2}$$

where $E_{\rm beam}$ is the beam energy,[6] $E_{\ell\gamma} \equiv E_\ell + E_\gamma$ is the sum of the signal lepton and signal photon energies, and $E_{{\rm rec}B,{\rm unscaled}}$ is the energy of the unscaled recoil B candidate.

We improve our reconstruction of the recoil B by exploiting the beam energy constraint. When the beam energies are on-peak, we expect the magnitude of the

[5]The expression $\vec{p}_{\nu,{\rm unscaled}}$ will be used interchangeably with $\vec{p}_{\rm miss}$.

[6]We define the "beam energy" as the energy of a single beam in the CM frame, unless stated otherwise.

B momentum ($\equiv |\vec{p}_B|$) to be calculable from the charged B mass (which we take to be 5.2791 GeV) and the beam energy, using the standard formula for relating energy, 3-momentum, and invariant mass ($E^2 + |\vec{p}|^2 = m^2$), referred to henceforth as the "invariant mass formula." We define a "scaled" recoil B candidate by retaining the measured direction of the unscaled candidate, but taking advantage of the beam constraint by rescaling its 3-momentum to the expected B scalar momentum, approximately 320 MeV. In addition, we constrain its energy to be the beam energy.

The "scaled" neutrino 3-momentum $\vec{p}_{\nu,\text{scaled}}$ is then defined by using the scaled recoil B in the missing-momentum calculation:

$$\vec{p}_{\nu,\text{scaled}} = -(\vec{p}_{\text{rec}B,\text{scaled}} + \vec{p}_{\ell\gamma}), \tag{3.3}$$

where $\vec{p}_{\text{rec}B,\text{unscaled}}$ is the 3-momentum of the scaled recoil B. The energy of the scaled neutrino, E_ν, is obtained by taking the difference between the beam energy and the sum of the lepton and photon energies:

$$E_\nu = E_{\text{beam}} - E_{\ell\gamma}. \tag{3.4}$$

Figure 3.1 shows a simplified geometric illustration of the difference between the unscaled and scaled neutrino 3-momentum reconstruction. As one can see, the two reconstructions will result in different directions for the reconstructed neutrino 3-momentum if the unscaled and scaled recoil B 3-momenta have different magnitudes.

For off-peak events, the beam energy is less than the charged B mass, so one cannot

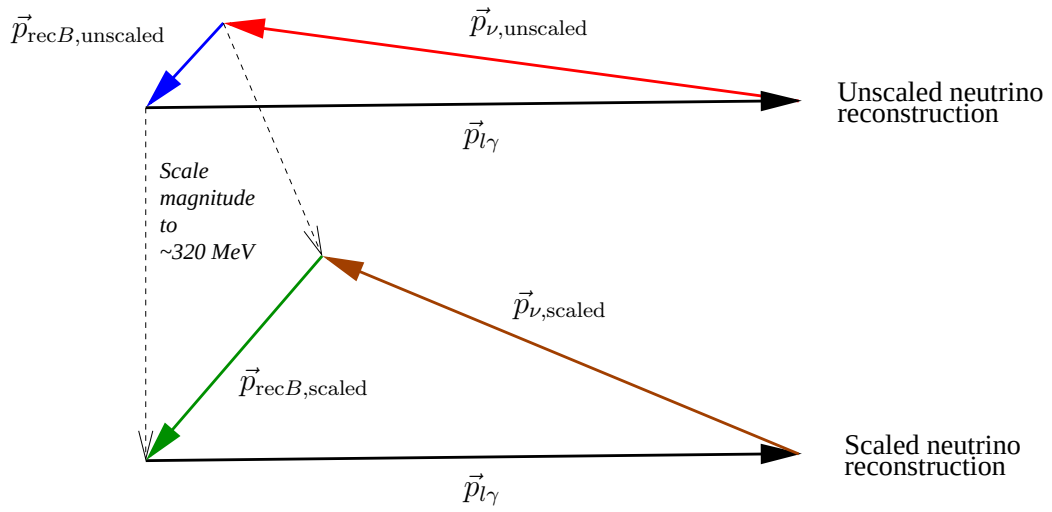


Figure 3.1. A geometric illustration of the neutrino reconstruction in both the unscaled (upper) and scaled (lower) cases.

calculate $|\vec{p}_B|$ using the invariant mass formula, since no B mesons can be created at these energies. Because we are still interested in reconstructing a scaled recoil B candidate and a scaled signal neutrino candidate in the off-peak and have their 4-momentum spectra potentially closely match those in on-peak signal and continuum MC, we need an equivalent (or nearly equivalent) formulation. Our choice was to scale the charged B mass down by a factor of beam energy/nominal on-peak beam energy, where the nominal MC on-peak beam energy is $E^*_{\mathrm{beam,MC}} = 5.2891\,\mathrm{GeV}$. We then substitute the scaled B mass into the invariant mass formula to calculate $|\vec{p}_B|$.

3.2 Two-Photon Background

An important limiting factor in this analysis is our inability to model all of our continuum background. In earlier studies, based on a smaller data sample and an earlier version of our reconstruction software, we found discrepancies between off-

peak data and continuum MC in both the sizes (around a factor of two for the electron channel) and shapes of the distributions of some of our analysis selection variables, even after all analysis selection criteria were applied. We hypothesized that this background discrepancy is mostly due to unmodeled two-photon events ($e^+e^- \rightarrow e^+e^-(\gamma\gamma)^* \rightarrow e^+e^-X$, where X can be any set of allowed particles). We also believe that unmodeled higher-order QED events may be another source of the discrepancy.

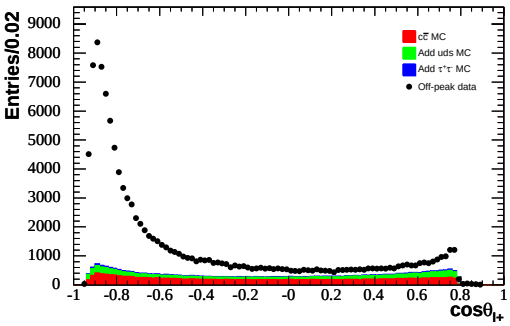
In the electron channel, we suspect that our background is largely composed of single-tagged two-photon events in which the tagged beam electron (or positron),[7] which tends to have relatively high energy, is assigned to be our signal lepton. In addition, these single-tagged events will have some missing energy due to the other beam positron (or electron) being undetected. If our hypothesis is true, then we should expect a larger discrepancy in the electron channel vs. the muon channel, and this is indeed the case.[8]

In the electron channel, we studied the low track multiplicity background discrepancy by examining radiative Bhabha and 4-electron final state MC samples, and found that these samples do account for much of the low-multiplicity background discrepancy, at least at loose cut levels. Unfortunately, these samples have limited statistics, so one cannot accurately judge the contributions of these modes because no events

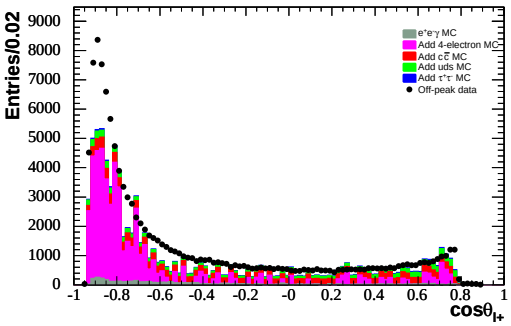
[7] We can classify two-photon events by their reconstruction in the detector. In an "untagged" event, both the beam electron and beam positron are not reconstructed (typically because their paths lie outside of the detector acceptance). In a "single-tagged" event, either the beam electron or the beam positron, but not both, are reconstructed. In a "double-tagged" event, both are reconstructed.

[8] It is possible that two-photon events could also be a source of background in the muon channel if the two-photon final state has high energy muons or other high energy particles that could be misidentified as muons.

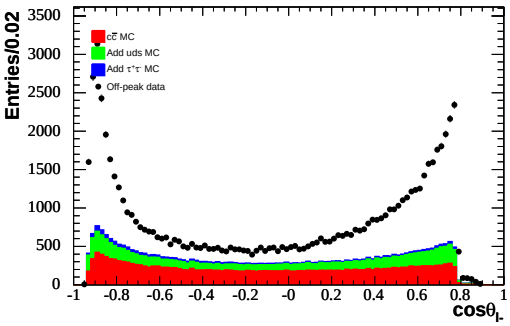
survive moderate or tight selection criteria. We could apply a final low multiplicity cut to be safe, but we found that the primary source of discrepancy after final analysis cuts are applied is in the higher track multiplicity region (greater than 4 charged tracks). In figures 3.2-3.4, we compare the electron-channel MC and off-peak distributions of some variables, with only a basic, loose event selection applied. Clearly there is a continuum data/MC discrepancy, even with the extra limited-statistics continuum MC samples included. Similarly, in figure 3.5, we show a data/MC continuum comparison for the muon channel, and there is also an evident discrepancy.



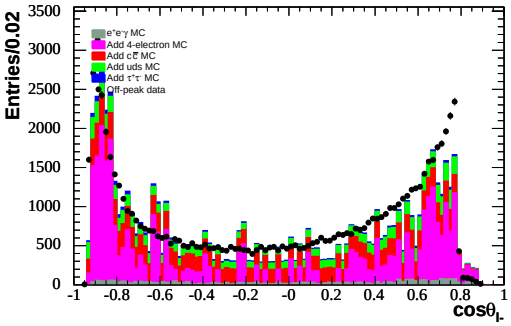
(a) Signal electron (positive charge) lab θ distributions.



(b) Signal electron (positive charge) lab θ distributions, including 4-electron and radiative Bhabha MC samples.

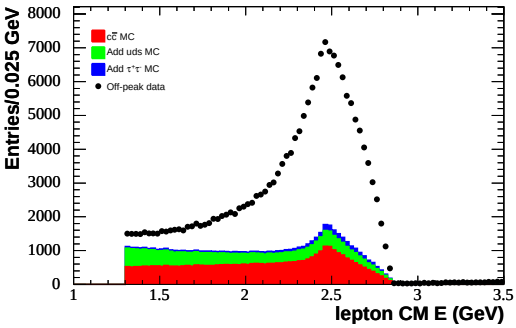


(c) Signal electron (negative charge) lab θ distributions.

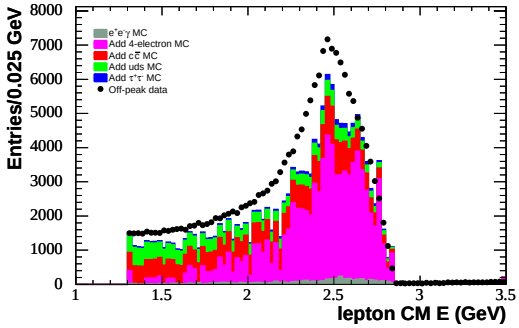


(d) Signal electron (negative charge) lab θ distributions, including 4-electron and radiative Bhabha MC samples.

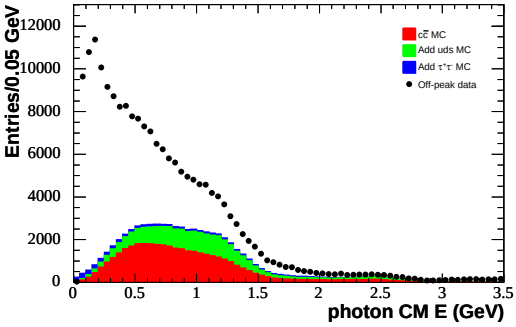
Figure 3.2. Signal electron lab θ distributions (split by charge) in the signal region for the electron channel, MC vs. off-peak data. Scaling is to off-peak luminosity. Only a very basic, loose event selection has been applied.



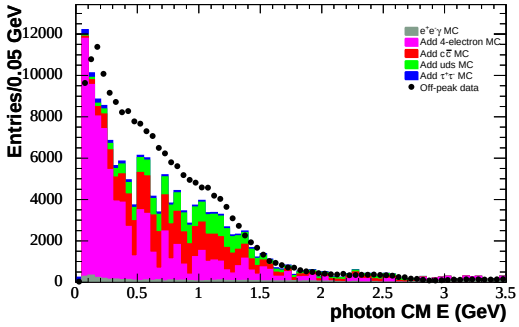
(a) Signal lepton CM energy distributions.



(b) Signal lepton CM energy distributions, including 4-electron and radiative Bhabha MC samples.

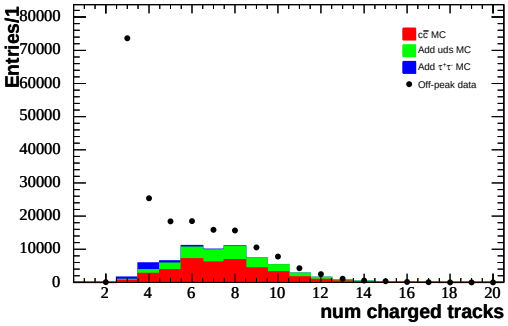


(c) Signal photon CM energy distributions.

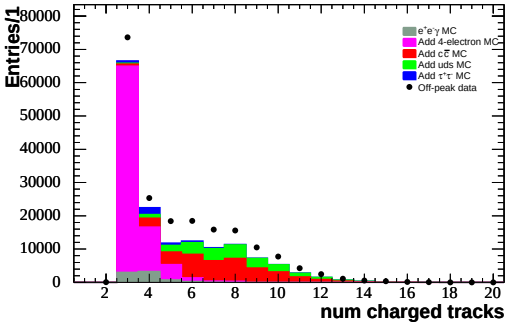


(d) Signal photon CM energy distributions, including 4-electron and radiative Bhabha MC samples.

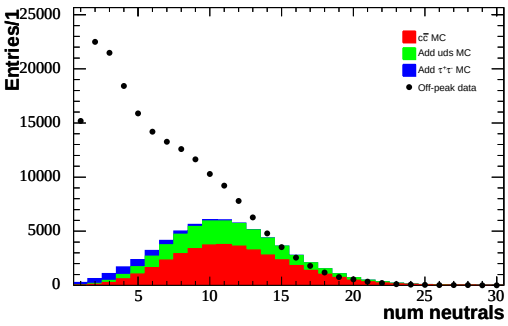
Figure 3.3. Signal lepton and signal photon CM energy distributions in the signal region for the electron channel, MC vs. off-peak data. Scaling is to off-peak luminosity. Only a very basic, loose event selection has been applied.



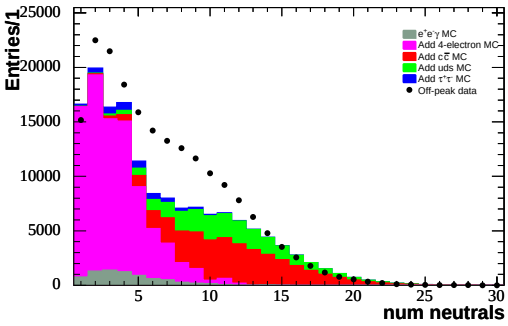
(a) Number of charged tracks distributions.



(b) Number of charged tracks distributions, including 4-electron and radiative Bhabha MC samples.

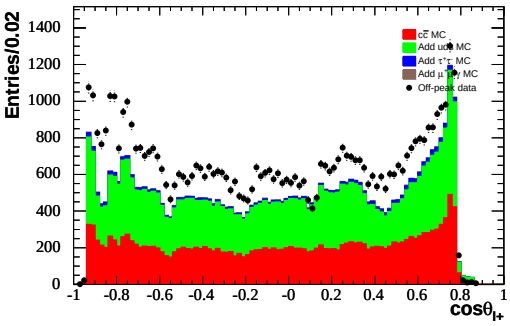


(c) Number of neutrals distributions.

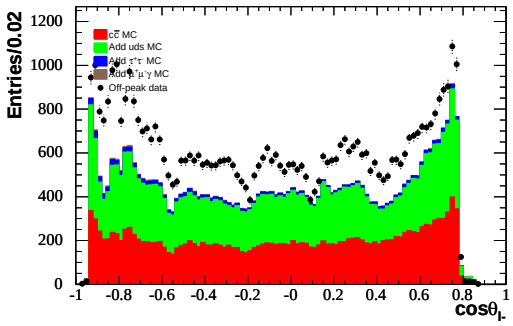


(d) Number of neutrals distributions, including 4-electron and radiative Bhabha MC samples.

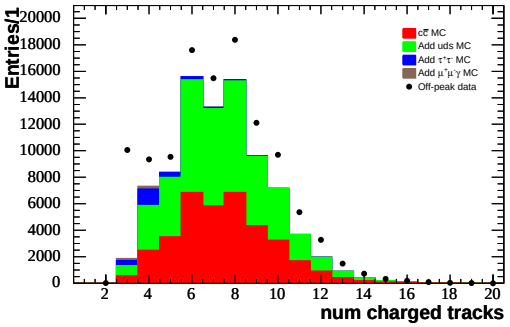
Figure 3.4. Number of charged tracks and number of neutrals distributions in the signal region for the electron channel, MC vs. off-peak data. Scaling is to off-peak luminosity. Only a very basic, loose event selection has been applied.



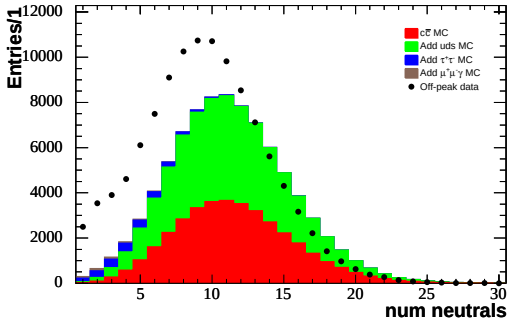
(a) Signal muon (positive charge) lab θ distributions.



(b) Signal muon (negative charge) lab θ distributions.



(c) Number of charged tracks distributions.



(d) Number of neutrals distributions.

Figure 3.5. Signal muon (separated by charge) lab angle, number of charged tracks, and number of neutrals distributions in the signal region for the muon channel, MC vs. off-peak data. Scaling is to off-peak luminosity. Only a very basic, loose event selection has been applied.

3.3 Blinding Strategy

As mentioned briefly in section 3.1, we blinded ourselves to both on-peak and off-peak data in order to avoid biasing our event selection. In particular, this meant that we did not use the on-peak or off-peak data in our selection criteria optimizations, or look at the marginal distributions (all selection criteria have been applied) of the selection variables until after we unblinded. However, we were still free to examine on-peak and off-peak data distributions before signal region unblinding in two cases.

First, we looked at these distributions at much looser cut levels. Our rule-of-thumb is that we need at least ten to twenty times the statistics of our final event selection in the signal region to feel comfortable about not biasing ourselves. The drawback is that the shapes of the distributions change as cuts are applied and/or tightened due to correlations among cuts.

Second, we unblinded in the on-peak and off-peak data in sideband regions. This can be done even before the cut optimization is completed, as long as we choose sideband regions far enough from the potential signal region. These regions should, by their very nature, have little to no signal events relative to the signal region. Ideally, one chooses sideband variables with high background statistics where the shapes of the other cut variable distributions would be as uncorrelated as possible in the background. However, even in the sideband region of a single variable, one would expect a different composition of background events, which can be useful for studying certain types of background. In chapter 11, we discuss the results of a preunblinding sideband region fit.

Chapter 4

Signal MC Generator Models

The first step in this analysis is to generate signal events to better understand the $B^+ \to \ell^+ \nu_\ell \gamma$ mode. In this section we discuss the models used to generate our signal MC samples. In particular, we use one model (Korchemsky, Pirjol, and Yan) to generate the signal decay, simultaneously overlaying internal bremsstrahlung processes modeled by a software package called `PHOTOS`.

4.1 Korchemsky, Pirjol, and Yan Model

For our signal decay model, we use the tree-level hadronic matrix element of Korchemsky, Pirjol, and Yan (KPY) [11], as given earlier in section 1.6 in equation (1.20):

$$\varepsilon(\mu, \epsilon^*, v, q) f_V + i[\epsilon^*(v \cdot q) - q_\mu(\epsilon^* \cdot v)] f_A), \tag{4.1}$$

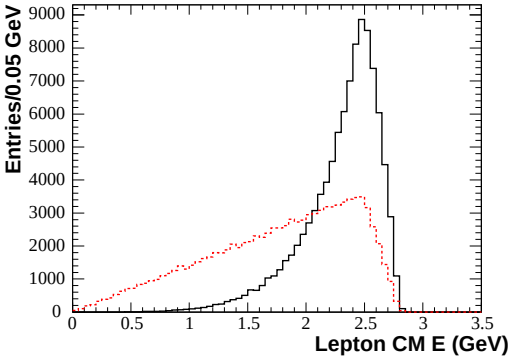
where $\varepsilon(\mu, a, b, c) = \varepsilon_{\mu\alpha\beta\sigma} a^\alpha b^\beta c^\sigma$, q is the signal photon 4-momentum, ϵ is the signal photon polarization 4-vector , v is the signal B 4-velocity, and the form factors f_V, f_A are equal in magnitude and proportional to E_γ^{-1}.

We use this model to generate our signal events, applying a minimum cutoff to the photon energy of 350 MeV, since the KPY model uses a $\Lambda_{\rm QCD}/E_\gamma$ expansion.[1] Because the analysis selects the highest CM energy photon to be the signal photon we will effectively end up with low efficiency for events in which the signal photon has low energy.

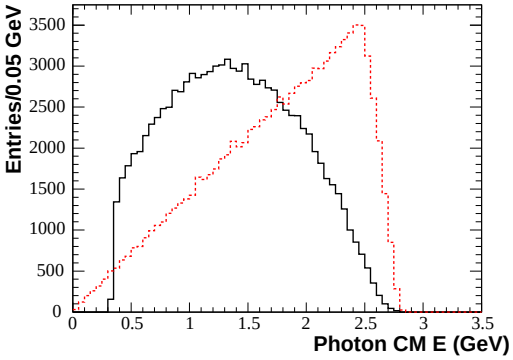
The resulting lepton, photon, and neutrino CM energy distributions, along with the corresponding Dalitz plot are shown in figure 4.1. The lepton, photon, and neutrino CM energy distributions for a phase-space (PS) model uniform on the Dalitz plot are also plotted for comparison. (The rising distributions of the PS model are simply due to integration over the available triangular phase space.) One can see that in moving from a PS model to the KPY model, the photon energy distribution no longer rises, but peaks at around 1.3 GeV, and the lepton energy peaks more sharply towards higher energies. The lepton and photon tend to be emitted back-to-back, so their angle in the CM, $\cos\theta_{\ell\gamma}$, is a potentially useful cut variable. A comparison of the signal particle opening angles in the KPY model vs. the PS model can be seen in figure 4.2.

We also generate events in which the axial vector form factor is set to 0. (We refer to this as the "fa0 model," while referring to the previous model as the "standard signal model.") The lepton and photon CM energy distributions for the KPY model with and without the axial vector form factor can be compared in figure 4.3. A comparison of the signal particle opening angles can be seen in figure 4.4. As will

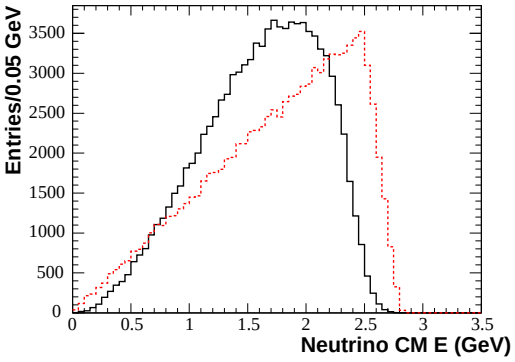
[1] All references to the "full BF," with respect to the signal mode, from herein will refer to the partial BF with a 350 MeV minimum energy restriction on the signal photon energy in the signal B rest frame. In section 14.2, the full BF results will be corrected to account for this cutoff.



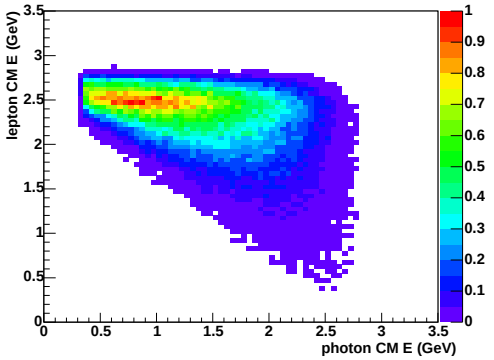
(a) Signal lepton CM energy distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is the phase space decay model.



(b) Signal photon CM energy distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is the phase space decay model.

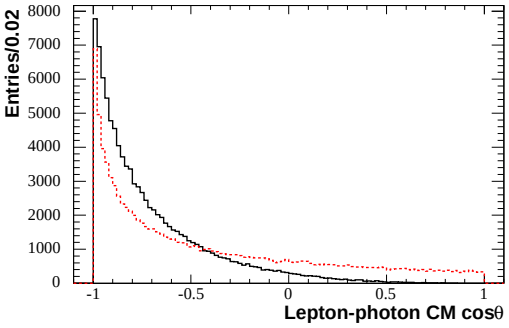


(c) Signal neutrino CM energy distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is the phase space decay model.

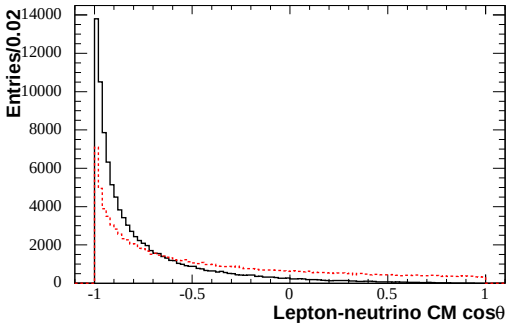


(d) Dalitz plot for KPY model (no PHOTOS).

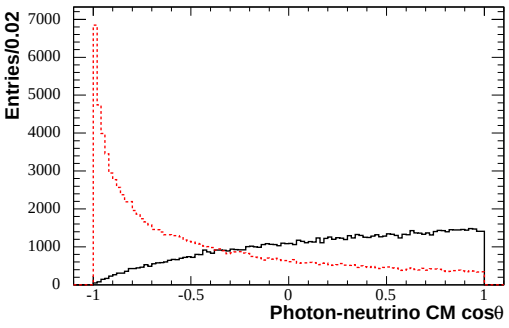
Figure 4.1. Generator-level comparisons of $B^+ \to \ell^+ \nu_\ell \gamma$ signal decay particle energies in the KPY and phase space models for the electron channel.



(a) Signal lepton-photon CM opening-angle distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is the phase space decay model.



(b) Signal lepton-neutrino CM opening-angle distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is the phase space decay model.



(c) Signal photon-neutrino CM opening-angle distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is the phase space decay model.

Figure 4.2. Generator-level comparisons of $B^+ \to \ell^+ \nu_\ell \gamma$ signal decay particle opening angles in the KPY and phase space models for the electron channel.

be discussed in section 13.6, the fa0 model is unrealistic, but will still be useful for evaluating systematic uncertainties.

When the axial vector form factor is set to 0, the lepton and neutrino energy spectrums are almost exactly the same. The lepton and neutrino still have a tendency to come out back-to-back, but the lepton and neutrino now have an equal tendency to come out back-to-back with the photon, as opposed to the non-fa0 case where the neutrino has a higher tendency to come out in the direction of the photon.

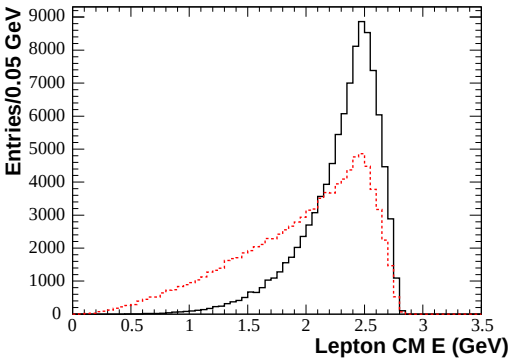
It is important to note that, since the lepton and photon energy distributions are model-dependent, we must avoid relying too heavily on their shape.

4.2 PHOTOS

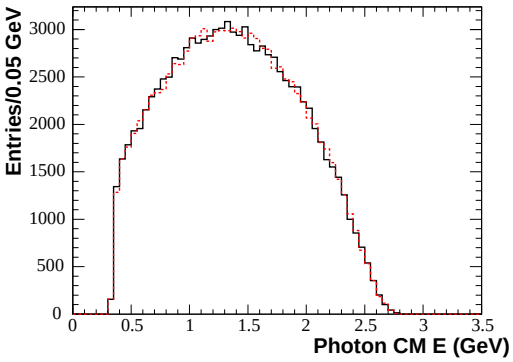
Bremsstrahlung is modeled using a software package called PHOTOS [17].[2] After including internal bremsstrahlung processes, approximately 20% of our generated electron-channel signal events have at least one additional signal-side photon with a minimum energy of 350 MeV. The photon multiplicity shown in figure 4.5. One could potentially misreconstruct the recoil B energy by a couple of GeV by not assigning these extra radiated photons to the signal side, but since we have minimum CM energy criteria applied to the lepton and photon, this problem is mitigated. Also, the signal electron loses a mean CM energy of only about 50 MeV due to PHOTOS.

The small effect of PHOTOS on the electron energy distribution can be seen in figure 4.6(a). Also, from figure 4.6(b) it is clear that the difference between the

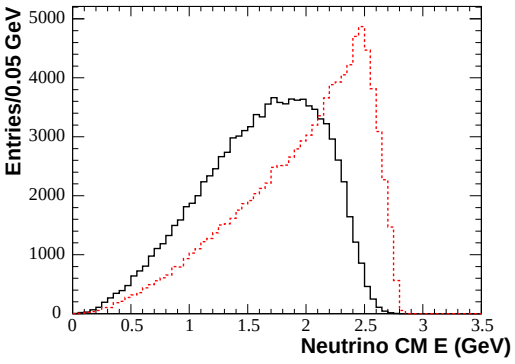
[2] PHOTOS is activated when generating any $B\overline{B}$ MC events.



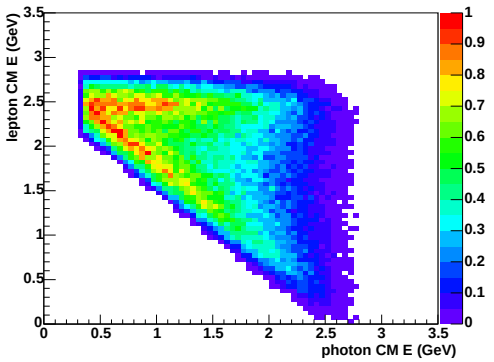
(a) Signal lepton CM energy distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is KPY model (no PHOTOS), axial vector form factor set to 0.



(b) Signal photon CM energy distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is KPY model (no PHOTOS), axial vector form factor set to 0.

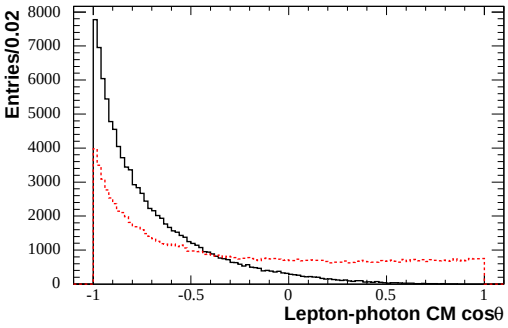


(c) Signal neutrino CM energy distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is KPY model (no PHOTOS), axial vector form factor set to 0.

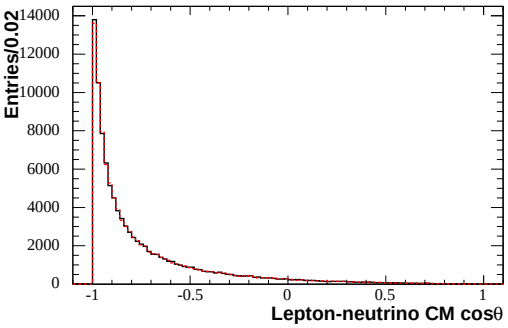


(d) Dalitz plot for KPY model (no PHOTOS), axial vector form factor set to 0.

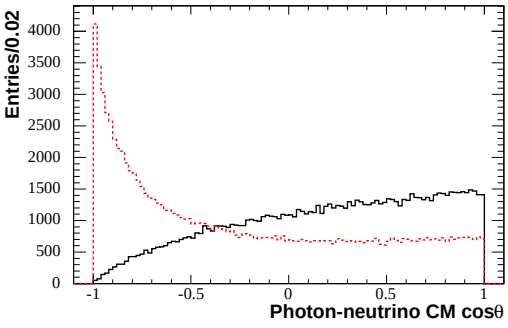
Figure 4.3. Generator-level comparisons of $B^+ \to \ell^+ \nu_\ell \gamma$ signal decay particle energies in the KPY model (no PHOTOS), with and without the axial vector form factor set to 0 for the electron channel.



(a) Signal lepton-photon CM opening-angle distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is KPY model (no PHOTOS), axial vector form factor set to 0.



(b) Signal lepton-neutrino CM opening-angle distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is KPY model (no PHOTOS), axial vector form factor set to 0.



(c) Signal photon-neutrino CM opening-angle distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is KPY model (no PHOTOS), axial vector form factor set to 0.

Figure 4.4. Generator-level comparisons of $B^+ \to \ell^+ \nu_\ell \gamma$ signal decay particle opening angles in the KPY model (no PHOTOS), with and without the axial vector form factor set to 0 for the electron channel.

electron and muon energy spectrum in our model is almost entirely due to the extra radiation in the electron channel modeled by PHOTOS.

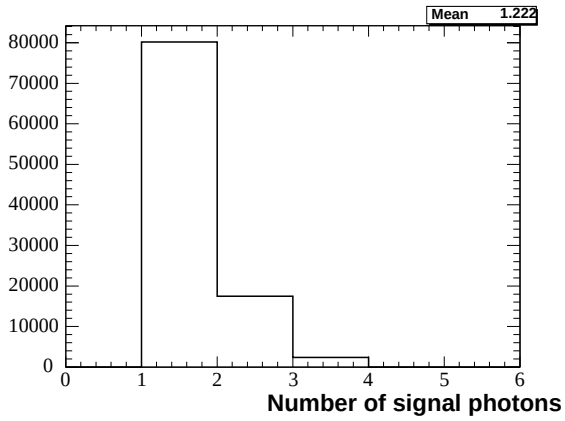
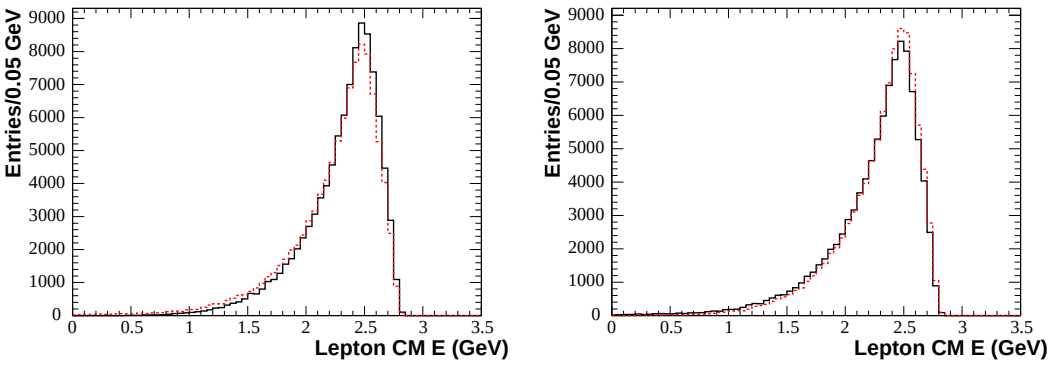


Figure 4.5. Generator-level signal photon multiplicity distributions for the electron channel, with PHOTOS turned on.



(a) Signal lepton CM energy distribution, electron channel. Black (solid) is KPY model (no PHOTOS). Red (dashed) is KPY model with PHOTOS.

(b) Signal lepton CM energy distribution for KPY model with PHOTOS. Black (solid) is electron channel. Red (dashed) is muon channel.

Figure 4.6. More generator-level comparisons of the lepton energy distributions at generator level.

Chapter 5

Analysis Samples

Our analysis samples can be broken up along different lines: data vs. MC, signal vs. background, or $B^+ \to \ell^+ \nu_\ell \gamma$ vs. control sample. On-peak data are fitted to extract the signal, while off-peak data and MC samples provide the fit templates for the signal extraction. MC samples are also used to determine signal efficiency, as well as to determine the optimal separation of signal from background. Control samples provide a validation of our analysis, and play a role in determining some systematic uncertainties. A summary of our $B^+ \to \ell^+ \nu_\ell \gamma$ analysis samples can be found in table 5.1.

5.1 Data

The data consist of on-peak and off-peak data, broken up into four distinct "run periods": Run 1 (October 1999–October 2000), Run 2 (February 2001–June 2002), Run 3 (December 2002–June 2003), and Run 4 (September 2003–July 2004). These run periods include different detector conditions and performance, as well as changes in the data acquisition system. However, the reconstruction software is consistent

Table 5.1. Samples used, totaled over the four run periods. The luminosity values for generic B events are based on B-counting values as given in table 5.2. The luminosity values for generic continuum events are based on the cross sections given in table 5.6.

	Total	
MC Sample	# Events (K)	Lumi(fb^{-1})
$B^+ \to e^+\nu_e\gamma$	1,172	354902.72
$B^+ \to e^+\nu_e\gamma$ (fa0)	1,166	353090.98
$B^+ \to \mu^+\nu_\mu\gamma$	1,171	354600.37
$B^+ \to \mu^+\nu_\mu\gamma$ (fa0)	1,166	353087.32
On-peak	3,088,954	210.49
Off-peak	286,942	21.59
B^+B^-	465,012	844.25
$B^0\overline{B}^0$	466,530	847.04
$c\overline{c}$	355,335	273.33
uds	554,294	265.21
$\tau^+\tau^-$	360,410	383.41
$e^+e^-\gamma$	7,054	6.49
$e^+e^- \to e^+e^-e^+e^-$	5,255	0.12
$\mu^+\mu^-\gamma$	224,250	200.37
$B^+ \to \pi^0\ell^+\nu_\ell$	512	1687.45
$B^+ \to \rho^0\ell^+\nu_\ell$	756	1607.44
$B^+ \to \eta\ell^+\nu_\ell$	348	1977.55
$B^+ \to \eta'\ell^+\nu_\ell$	336	3927.82
$B^+ \to \omega\ell^+\nu_\ell$	460	2113.67
$B^0 \to \pi^-\ell^+\nu_\ell$	352	2507.07
$B^0 \to \rho^-\ell^+\nu_\ell$	466	2141.19

across the run periods. The breakdown of our samples by run period is given in table 5.3.

Table 5.2. B-counting values. The errors on the total are obtained by adding the respective run errors in quadrature.

		Error		
Run Period	# B Events	Statistical	Systematic	Total
Run 1	21167252.1	24177.3	232839.8	234091.7
Run 2	66395899.7	43195.1	730354.9	731631.1
Run 3	34067156.0	30960.7	374738.7	376015.5
Run 4	110107681.0	55519.8	1211184.5	1212456.3
Total	231,737,988.8	80569.0	1481563.2	1483752.3

Table 5.3. Samples used, broken up by run period.

	Run 1		Run 2		Run 3		Run 4	
MC Sample	# Events (K)	Lumi (fb^{-1})	# Events (K)	Lumi (fb^{-1})	# Events (K)	Lumi (fb^{-1})	# Events (K)	Lumi (fb^{-1})
$B^+ \to e^+\nu_e\gamma$	129	39503.17	316	95544.10	159	48310.75	568	171544.70
$B^+ \to e^+\nu_e\gamma$ (fa0)	130	39809.39	310	93729.97	158	48006.91	568	171544.70
$B^+ \to \mu^+\nu_\mu\gamma$	129	39503.17	315	95241.75	159	48310.75	568	171544.70
$B^+ \to \mu^+\nu_\mu\gamma$ (fa0)	130	39809.39	310	93729.97	156	47399.23	570	172148.73
On-peak	267,668	19.45	864,182	60.23	444,392	31.05	1,512,712	99.76
Off-peak	28,755	2.31	91,966	6.93	31,525	2.40	134,696	9.94
B^+B^-	24,246	44.55	119,356	216.53	67,974	123.92	253,436	459.25
$B^0\overline{B}^0$	28,762	52.85	120,300	218.24	61,310	111.77	256,159	464.18
$c\overline{c}$	21,346	16.42	100,076	76.98	57,700	44.38	176,213	135.55
uds	30,168	14.43	177,601	84.98	104,110	49.81	242,415	115.99
$\tau^+\tau^-$	22,098	23.51	99,792	106.16	69,852	74.31	168,668	179.43
$e^+e^-\gamma$	1,300	1.20	1,546	1.42	790	0.73	3,418	3.15
4-electron	585	0.01	1,391	0.03	714	0.02	2,565	0.06
$\mu^+\mu^-\gamma$	22,814	20.38	70,190	62.72	16,578	14.81	114,668	102.46
$B^+ \to \pi^0\ell^+\nu_\ell$	52	374.21	124	881.06	62	442.70	114	809.10
$B^+ \to \rho^0\ell^+\nu_\ell$	76	352.69	184	843.08	94	432.82	112	512.60
$B^+ \to \eta\ell^+\nu_\ell$	50	287.09	120	680.30	64	364.61	114	645.56
$B^+ \to \eta'\ell^+\nu_\ell$	50	590.58	114	1329.50	58	679.74	114	1328.00
$B^+ \to \omega\ell^+\nu_\ell$	78	361.97	180	824.75	88	405.19	114	521.76
$B^0 \to \pi^-\ell^+\nu_\ell$	83	276.27	209	686.87	106	350.08	114	374.24
$B^0 \to \rho^-\ell^+\nu_\ell$	137	294.06	333	705.73	172	366.31	1,138	2409.07

The on-peak sample consists of electron-positron collision data at a CM energy mostly within 1 MeV of the $\Upsilon(4S)$ peak. At this energy, the cross section for producing charged and neutral B pairs via the $\Upsilon(4S)$ resonance is approximately 1.1 nb [16]. However, we do not use this cross section to determine the number of B mesons, but instead use the numbers provided by the *BABAR* B-counting group, which are calculated using dimuon events. The B-counting numbers are summarized in table 5.2. The total luminosity of the on-peak sample is $210.49\,\text{fb}^{-1}$.

The off-peak sample was obtained by setting the total CM energy of PEP-II to approximately 39 MeV below the $\Upsilon(4S)$ mass (61 MeV below the B-pair production threshold), resulting in a pure sample of continuum events. The total luminosity of the off-peak sample is $21.59\,\text{fb}^{-1}$, about a tenth of the on-peak luminosity.

5.2 Monte Carlo

In order to best reproduce the data, Monte Carlo events, after being generated, are processed by a GEANT4-based [18] detector simulation using conditions obtained from data, and are thus also broken up the same way: Run 1, Run 2, Run 3, and Run 4. However, their generated luminosities do not necessarily correspond to that of on-peak data, so we reweight the MC sample event counts accordingly, as will be discussed in more detail in section 5.4. The reconstruction software is the same as that used to process data.

5.2.1 Signal Monte Carlo Sample

For both the electron and muon signal channels, we generate two sets of MC samples, one in which the vector and axial vector form factors are assumed to be equal, and one in which the axial vector form factor is set to 0, as described in section 4.1. The other B is decayed generically, as will be described in section 5.2.2.

5.2.2 Generic Monte Carlo

Preliminary MC studies confirmed the presence of both $B\overline{B}$ and non-$B\overline{B}$ backgrounds. Therefore, for the electron and muon channels, we look at $e^+e^- \rightarrow \Upsilon(4S) \rightarrow b\overline{b}$ ("$B\overline{B}$," or broken up by charge:"B^+B^-," "$B^0\overline{B}^0$"), $e^+e^- \rightarrow c\overline{c}$ ("$c\overline{c}$"), $e^+e^- \rightarrow q\overline{q}, q = u, d, s$ ("uds"), and $e^+e^- \rightarrow \tau^+\tau^-$ ("$\tau^+\tau^-$") MC. In addition, for the muon channel, we look at $\mu^+\mu^-\gamma$ ("radiative dimuon") MC.

$B\overline{B}$ MC events were generated using `EvtGen` [19], supplemented with `JETSET` [20]. For the continuum, $c\overline{c}$ and uds events were generated by `JETSET`, $\tau^+\tau^-$ events by `KORALB` [21], and radiative dimuon samples by `KK2F` [22].

As discussed earlier in section 3.2, while we do not have any generic two-photon MC available, we did perform some studies at loose cuts levels with exclusive QED and two-photon samples, in particular, radiative Bhabha and 4-electron final-state samples. The radiative Bhabha events were generated with `Bhwide` [23] and the 4-electron events were generated with `Diag36` [24].

5.2.3 Semileptonic Monte Carlo

Preliminary studies showed that semileptonic events, especially $b \to u\ell\nu_\ell$, are a significant source of background, as one would expect, given that such events have similar event topologies to our signal decay. We understand $b \to c\ell\nu_\ell$ decays better than $b \to u\ell\nu_\ell$ decays, and thus do not study $b \to c\ell\nu_\ell$ events separately from other generic B decays. However, $b \to u\ell\nu_\ell$ decays warrant a separate, more detailed study.

5.2.3.1 Seven $b \to u\ell\nu_\ell$ Decay Modes

In particular, we study seven $b \to u\ell\nu_\ell$ decay modes with separate MC samples: $B^+ \to h^0\ell^+\nu_\ell$ ($h^0 = \pi^0, \rho^0, \eta, \eta', \omega$) and $B^0 \to h^-\ell^+\nu_\ell$ ($h = \pi^-, \rho^-$), respectively referred to below as the "h mode."[1] These samples are generated with a flat q^2 distribution, where q^2 is the invariant mass of the virtual $W^\pm$ boson. The events are then reweighted using q^2-dependent weights based on the chosen theoretical model. The separation of these modes from the generic B sample also allows us to adjust the values of their BFs.

The primary models used, along with the assumed BFs, are listed in table 5.4. (Alternative models will be used in evaluating systematic uncertainties.)

The charged π (1.38×10^{-4}) and ρ (2.14×10^{-4}) branching fractions are taken from the most recent *BABAR* measurement [28]. We obtain the neutral π and ρ branching fractions by multiplying the corresponding charged branching fraction by two (isospin symmetry) times a factor of 1.081 (lifetime ratio between charged and

[1] We also use the terms "π mode" and "ρ mode" to refer to either the charged or neutral decay modes.

Table 5.4. Primary model weights and branching-fraction comparison for $b \to u\ell\nu$ MC. The corresponding muonic channels have the same values and use the same models. The FNAL04 [25] weights are based on lattice QCD results, while the BALL05 [26] and BALL04 [27] weights are based on light-cone sum rules.

MC Sample	Primary Weight	BF
$B^0 \to \pi^- e^+ \nu_e$	FNAL04	1.38×10^{-4}
$B^0 \to \rho^- e^+ \nu_e$	BALL05	2.14×10^{-4}
$B^+ \to \eta e^+ \nu_e$	BALL04	8.00×10^{-5}
$B^+ \to \eta' e^+ \nu_e$	BALL04	3.89×10^{-5}
$B^+ \to \omega e^+ \nu_e$	BALL05	1.16×10^{-4}
$B^+ \to \pi^0 e^+ \nu_e$	FNAL04	7.46×10^{-5}
$B^+ \to \rho^0 e^+ \nu_e$	BALL05	1.16×10^{-4}

neutral B mesons), where we use the value cited in the same *BABAR* measurement [28] for consistency. The ω branching fraction, using SU(3) flavor symmetry, is set equal to the ρ^0 branching fraction. The η value is from the PDG [29] and is based on a 2003 measurement by the CLEO collaboration [30]. The η' value is calculated from the η result using an SU(3) flavor symmetry factor of 4.32/2.10 (approximately 2.057) obtained from the ratio of predicted branching fractions for the η and η' modes [31].

5.2.3.2 Other $b \to u\ell\nu_\ell$ Decay Modes

In addition to the seven semileptonic $b \to u\ell\nu_\ell$ modes that we treat separately, there are other $b \to u\ell\nu_\ell$ decay modes ("$B \to X_u\ell\nu_\ell$"), both resonant and nonresonant, that can contribute to our background. The dynamics of these other decay modes are not well determined theoretically, especially in the limited region of phase space of our analysis. In our $B\overline{B}$ MC generator, $B \to X_u\ell\nu_\ell$ decays are handled by either the Isgur-Wise [32] model (resonant decays) or the De Fazio-Neubert [33] model (nonresonant decays). We do not treat $B \to X_u\ell\nu_\ell$ events separately, but instead,

leave them in the sample of generic $B\overline{B}$ MC. Our signal extraction fits show an excess of generic B events in data relative to the number predicted by MC; we suspect that mismodeling of $B \to X_u \ell \nu_\ell$ is largely the reason. The resulting large uncertainty in $B \to X_u \ell \nu_\ell$ will ultimately be included as a systematic (section 13.5.1).

5.3 Control Sample: $B^+ \to \pi^+ \overline{D}^0 (\to K^+ \pi^-)$

Our primary control sample mode is $B^+ \to \pi^+ \overline{D}^0 (\to K^+ \pi^-)$ (also abbreviated as $\overline{D}^0 \pi^+$). We choose this sample because there are three decay products, with one decay product (the hard pion from the B decay) of high energy, and the sample is clean with a relatively large branching fraction. The hard pion is treated as the signal electron, the kaon as the signal photon, and the soft pion (from the D decay) as the signal neutrino. We use this sample to study our event shape and neutrino reconstruction variables.

There are clearly differences in the kinematics and dynamics of the $B^+ \to \ell^+ \nu_\ell \gamma$ signal mode and this control sample mode. Because the hard pion has an energy of approximately 2.31 GeV, we use it as a proxy for the high energy signal lepton in the $B^+ \to \ell^+ \nu_\ell \gamma$ decay. However, the hard pion energy spectrum is largely monochromatic (the boost from the B rest frame to the CM frame does smear it out a bit), so the shape of the spectrum does not resemble that of the corresponding lepton spectrum.

Since the "photon" (kaon) is the decay product of the monochromatic D^0, its energy spectrum is smeared out, and ranges from the kaon mass to more than 2.8 GeV,

somewhat resembling the $B^+ \to \ell^+ \nu_\ell \gamma$ signal photon energy spectrum in range.

What is extremely useful about the $\overline{D}^0 \pi^+$ control sample is that we can reconstruct the "neutrino" (soft pion) candidate. This allows us to study the neutrino reconstruction because we can "throw" out the soft pion, while retaining its kinematic information.

In addition, the recoil reconstruction will potentially be different between the $B^+ \to \ell^+ \nu_\ell \gamma$ signal mode and this control sample mode due to phenomena such as hadronic splitoffs and bremsstrahlung. In particular, if any energy deposits from the $\overline{D}^0 \pi^+$ decay are associated with recoil candidates, this will result in a ΔE spectrum shifted upwards, relative to that of the $B^+ \to \ell^+ \nu_\ell \gamma$ decay. On the other hand, excess photons produced by bremsstrahlung from the signal lepton in the $B^+ \to \ell^+ \nu_\ell \gamma$ decay will be associated with the recoil, resulting in an upwards ΔE shift relative to the $\overline{D}^0 \pi^+$ control sample ΔE.

We study this control sample with signal MC, generic B MC, off-peak data, and on-peak data. A clean sample of events based on a full reconstruction of the signal-side B was available.

5.4 Luminosity Factors

As mentioned in section 5.1, we break each data and MC sample up by run period. The effective generated luminosities for each of the periods for our various samples are, in general, not identical with those of on-peak data, so we reweight the events by run period luminosity factors in order to obtain the correct predicted event counts.

For off-peak data, the luminosity factor is the ratio of the on-peak run period luminosity to the off-peak run period luminosity, where we use the luminosities given in table 5.3.

For a continuum MC sample i, the luminosity factor, α_{ik}, for run period k is:

$$\alpha_{ik} = \frac{\sigma_i \mathcal{L}_{k,\mathrm{On}}}{N_{ik}}, \tag{5.1}$$

where σ_i is the total cross section for sample i (given in table 5.6), $\mathcal{L}_{k,\mathrm{On}}$ is the on-peak luminosity for run period k (given in table 5.3), and N_{ik} is the number of sample i events used from run period k (also given in table 5.3).

For a generic B MC sample i, the luminosity factor, α_{ik}, for run period k is:

$$\alpha_{ik} = \frac{N_{Bk}}{2N_{ik}}, \tag{5.2}$$

where N_{Bk} is the number of B events obtained from B-counting for run period k (given in table 5.2), and N_{ik} is the number of sample i events used for run period k. In the MC, we weight charged and neutral B mesons under the assumption that they have an equal production ratio, so therefore there is an extra factor of one-half.

For the rest of the MC samples, we use branching fractions instead of cross sections to calculate the luminosity factors, and the luminosity factor, α_{ik}, for sample i and run period k is:

$$\alpha_{ik} = \frac{N_{Bk}\mathcal{B}_i}{N_{ik}}, \tag{5.3}$$

where N_{Bk} is the number of B events obtained from B-counting for run period k,

N_{ik} is the number of sample i events used for run period k, and $\mathcal{B}_i$ is the assumed branching fraction for the sample mode (given in table 5.5). (The factor of one-half is canceled out by a factor of two due to two B mesons per event.)

Table 5.5. Input branching fractions used for MC samples. The corresponding muonic channels use the same values. In the relevant cases we provide the uncertainties.

Mode	Branching Fraction	Uncertainty	
		Statistical	Systematic
$B^+ \to e^+\nu_e\gamma$	3×10^{-6}	-	-
$B^0 \to \pi^- e^+\nu_e$	1.38×10^{-4}	0.10	0.16
$B^0 \to \rho^- e^+\nu_e$	2.14×10^{-4}	0.21	0.48
$B^+ \to \eta e^+\nu_e$	0.80×10^{-4}	0.31	0.16
$B^+ \to \eta' e^+\nu_e$	3.89×10^{-5}	-	-
$B^+ \to \omega e^+\nu_e$	1.16×10^{-4}	-	-
$B^+ \to \pi^0 e^+\nu_e$	7.46×10^{-5}	-	-
$B^+ \to \rho^0 e^+\nu_e$	1.16×10^{-4}	-	-
$B^+ \to \overline{D}^0\pi^+$	$4.91 \pm 0.21 \times 10^{-3}$	-	-
$D^0 \to K^-\pi^+$	$3.81 \pm 0.09 \times 10^{-2}$	-	-
$B^+ \to \overline{D}^{*0}\pi^+$	$4.6 \pm 0.4 \times 10^{-3}$	-	-
$D^{*0} \to \gamma D^0$	$3.81 \pm 0.29 \times 10^{-1}$	-	-

Table 5.6. Assumed cross sections for the generic MC samples. All samples except the $\mu^+\mu^-\gamma$, $e^+e^-\gamma$, and $e^+e^- \to e^+e^-e^+e^-$ samples use the *BABAR* Physics Book [16] value. The other values are obtained from generator-level studies.

MC Sample	Cross Section (nb)
$c\overline{c}$	1.30
uds	2.09
$\tau^+\tau^-$	0.94
$\mu^+\mu^-\gamma$	1.12
$e^+e^-\gamma$	1.09
$e^+e^- \to e^+e^-e^+e^-$	43.20

Chapter 6

Selection Variables

Before extracting our signal BF, it is necessary to reduce the presence of background events in the final data sample as much as possible. We accomplish this through multivariate classification, with our selection criteria targeted towards signal events. We divide our selection variables into five categories: signal side, recoil B reconstruction, neutrino reconstruction, event shape, and anti-two-photon. The optimization of the variables is described in detail in chapter 8, but we occasionally refer to it in this chapter.

6.1 Signal Side

We select the most energetic lepton and the most energetic photon in the CM frame for each event and label them the signal lepton and signal photon. Earlier, in section 4.1, we discussed the dynamics of the signal side at the MC generator level.

6.1.1 Signal Lepton and Photon CM Energies

Due to the dynamics of our signal mode decay, we expect a lepton energy spectrum that is peaked toward higher energies. This provides discriminating power in removing charmed semileptonic events, in which the presence of a D meson softens the lepton energy spectrum. The lepton CM energy spectrum in our signal MC is model-dependent, so despite the characteristic peak shape in the nominal model, it would be unwise to extract the signal from data by fitting this spectrum.

The photon CM energy spectrum is approximately quadratic, peaking around a quarter of the B mass. Relative to the lepton energy spectrum, the photon energy spectrum is softer and broader, thus providing less discriminating power against background. However, a low-side cut is still useful for eliminating background due to the presence of low-energy neutral clusters.

6.1.2 Cosine of the Lepton-Photon Angle, $\cos\theta_{\ell\gamma}$

Another signal side cut variable that depends on the dynamics of our signal decay is the cosine of the angle in the CM frame between the lepton and photon, $\cos\theta_{\ell\gamma}$, also mentioned in section 4.1. Due to the helicity structure of the decay, the lepton and photon tend to come out back-to-back, resulting in a peak at -1 in the $\cos\theta_{\ell\gamma}$ distribution. Background events from $e^+e^- \to q\overline{q}$, where $q = u, d, s, c$ also peak at -1 (lepton and photon from opposite jets), but not as sharply.

6.1.3 Cosine of the Implied B-$\ell\gamma$ Angle, $\cos\theta_{BY}$

This variable is the angle between the implied signal B and the lepton-photon combination. In events with a B decaying to $Y\nu$ (Y, the combination of the rest of the daughters of the B, is $l\gamma$ for our analysis), it is useful to use a combination of the knowledge of the B mass, the beam energy, and the reconstructed 3-vectors of the constituents of Y to calculate an estimate of the angle between the hypothetical B and the composite Y.

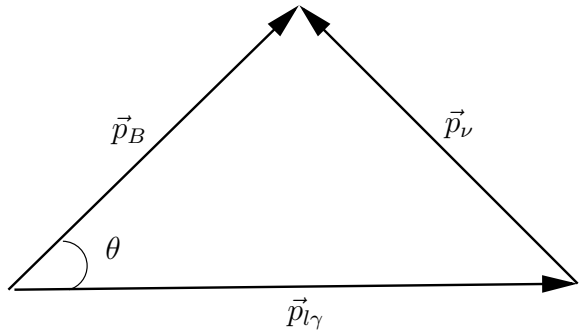


Figure 6.1. The angle θ used in the selection variable $\cos\theta_{BY}$.

More specifically, starting with the vectors as shown in figure 6.1, we calculate $\cos\theta_{BY}$ using the law of cosines:

$$\cos\theta_{BY} = \frac{(E_{\rm beam}/2 - E_\ell - E_\gamma)^2 - |\vec{p}_{\ell\gamma}|^2 - |\vec{p}_B|^2}{-2|\vec{p}_B||\vec{p}_{\ell\gamma}|}. \tag{6.1}$$

Due to resolution effects or an incorrect decay hypothesis, this variable can take on values far outside of the -1 and 1 range for the cosine of a physical angle.

In the B rest frame, the distribution of $\cos\theta_{BY}$ for correctly reconstructed decays is flat, because decays of the B (spin-0) are uniformly distributed in Ω. However, because we do not have accurate knowledge of the B rest frame, we must calculate

$\cos\theta_{BY}$ in the CM frame. The effect of boosting to this frame is that the angle between the lepton-photon combination and the B gets compressed, leading to larger values of $\cos\theta_{BY}$. This effect is shown in a generator-level study in figure 6.2.

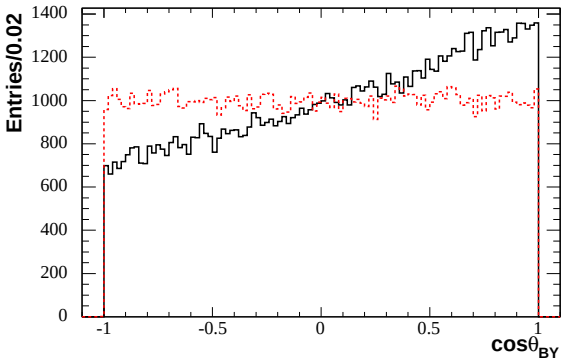


Figure 6.2. Generator-level distributions of $\cos\theta_{BY}$ for the electron channel. Black (solid) is $\cos\theta_{BY}$ computed in the CM frame. Red (dashed) is $\cos\theta_{BY}$ computed after boosting into the B rest frame.

We calculate $|\vec{p}_B|$ using the beam energy and a charged B mass of 5.2791 GeV, as described earlier in section 3.1.2. This includes any correction for off-peak energies, also described earlier.

In the $B^+ \to \pi^+ \overline{D}^0 (\to K^+\pi^-)$ control sample, we calculate $\cos\theta_{BY}$ differently because we need to account for the mass of the "neutrino" (charged pion):

$$\cos\theta_{BY} = \frac{(E_{\text{beam}}/2 - E_\ell - E_\gamma)^2 - m_{\pi^\pm}^2 - |\vec{p}_{\ell\gamma}|^2 - |\vec{p}_B|^2}{-2|\vec{p}_B||\vec{p}_{\ell\gamma}|}, \tag{6.2}$$

where $m_{\pi^\pm}$ is the charged pion mass.

6.1.4 Rejection of π^0s

6.1.4.1 Variable Description

Modes such as $B^+ \to \pi^0 \ell^+ \nu_\ell$ where the signal photon candidate is a decay product of a high energy π^0 comprise a non-trivial portion of our background, so a π^0 veto is important. Our veto consists of finding the invariant masses of all 4-momentum combinations of the signal photon with other photons in the event, and then requiring that they all lie outside of a specified range around the π^0 mass. For an event-based veto, we choose the invariant mass combination closest to the π^0 mass and use that as our π^0 antiselection variable, calling it "clpi0."

There are two ways to get a peak in the clpi0 variable distribution: first, if the signal photon candidate is really the decay product of a π^0 and it is combined with another photon candidate which actually is the other daughter of the π^0, and second, if there are enough photons in the event, a peak can be formed by combinatorics due to the bias of the closest π^0 mass selection. The second case does indeed result in a combinatoric peak in the signal mode, but the overall distribution is broader than that of background.

Because we are vetoing π^0s, we define the clpi0 "signal" region as the region away from the π^0 peak. The clpi0 "sideband" region is defined as the region covering the π^0 peak.

For the $B^+ \to \pi^+ \overline{D}^0 (\to K^+ \pi^-)$ control sample, we need to modify the clpi0 selection variable since it is calculated assuming an actual signal photon candidate, which has a zero mass. But the control sample signal "photon" has the kaon mass,

so we circumvent this by creating a massless candidate whose 4-vector is combined with each of the photon candidate 4-vectors in the event to find the closest π^0 mass combination. Starting with the original control sample signal "photon" (kaon) 4-vector $p_K \equiv (E_K, \vec{p}_K)$ in the lab frame, we define a new, massless 4-vector, p'_K:

$$p'_K \equiv (E_K, \alpha\vec{p}_K), \tag{6.3}$$

where α is the scale factor $E_K/|\vec{p}_K|$.[1] This adjusted massless 4-vector is used *only* in the calculation of the clpi0 variable.

6.1.4.2 Sideband Region of clpi0

We define the clpi0 sideband region to be between 124 and 146 MeV.[2] In this region, there is the expectation that the composition of events differs from that in the clpi0 signal region (the anti-π^0 region). For example, we expect an excess of $B^+ \to \pi^0\ell^+\nu_\ell$ events in this sideband. We use this region to study anti-two-photon variables (section 6.5.2), as well as to perform a preunblinding validation of our fit procedure (chapter 11).

6.1.5 Photon Shower Shape

Some potential signal photon background is due to merged π^0s, hadronic splitoffs, neutrons, anti-neutrons, and K^0_Ls since all of these can leave showers in the calorimeter

[1] The α used here is just for illustrative purposes and has no relation to any other αs used anywhere else in this thesis.

[2] This region was chosen after the cut optimization was completed, and does not overlap with the clpi0 signal regions in either the electron- or the muon-channel analyses.

that can be mistaken for high-energy photons. One of the methods for improving photon identification is to look at the shower shape in the calorimeter. In particular, electromagnetic showers tend to have energy deposits concentrated in two or three calorimeter crystals. So by calculating the shower shape's lateral moment, one can better discriminate between electromagnetic and hadronic showers [34]:

$$LAT \equiv \frac{\sum_{i=3}^{N} E_i r_i^2}{\sum_{i=3}^{N} E_i r_i^2 + E_1 r_0^2 + E_2 r_0^2}, \tag{6.4}$$

where i runs over the index set of shower crystals ordered from highest to lowest energy (E_i) deposited, r_i refers to the ith crystal's radius (in polar coordinates) in the plane perpendicular to the line between the interaction point and the shower center, and r_0 is the average distance between two crystals (5cm). For the signal photon, we call this variable "gLAT."

Because we do not have complete confidence that this variable is correctly modeled in the MC, we apply a conservative, fixed cut on gLAT at <0.55, rather than including this variable in the multivariate cut optimization described below in chapter 8.

6.1.6 Photon Fiducial

We reject events in which the photon lab momentum direction does not pass fiducial cuts. In particular, the lab polar angle is required to be between 0.326 and 2.443 radians.

6.1.7 Other Signal Lepton Cuts

One potential source of signal lepton background is the decay mode $J/\psi \rightarrow \ell^+\ell^-$, where the J/ψ is produced in B decays or in the continuum. We veto this background by requiring that the signal electron (muon) does not pair up with an oppositely charged electron (muon) to form a candidate with an invariant mass between 2.5 and 3.3 GeV (2.8 and 3.3 GeV).

Another source of lepton background, applicable only to the electron channel, is the photon-conversion process, a process in which a photon interacts with the detector and forms an e^+e^- pair. We veto this background by requiring that the signal electron does not come from a photon conversion. Specifically, we require that the signal electron (positron) does not pair up with any other positron (electron) in the event to form a candidate with an invariant mass of less than 30 MeV.

Both the J/ψ and the photon-conversion vetoes provide very little discriminating power against modeled background, but are applied nonetheless, because they are extremely efficient with respect to our signal modes.

6.2 Recoil B Reconstruction

After the signal lepton and photon candidates are chosen, the rest of the tracks and neutral clusters are combined to form the recoil B candidate, as described in some detail in section 3.1.2. We can impose kinematic cuts on this loosely reconstructed recoil B, in particular on the variables ΔE and $m_{\rm ES}$, which are standard variables in exclusive B reconstruction analyses. However, it is important to stress that unlike a

standard exclusive analysis that relies on fitting the ΔE and $m_{\rm ES}$ of the signal side B to extract signal, our loose, inclusive recoil B reconstruction reduces the power of these variables. While $m_{\rm ES}$ is still useful and important for our final signal extraction, ΔE does not have good enough resolution to be usable in a fit.

6.2.1 Recoil Missing Energy, ΔE

The ΔE variable is calculated as the difference between the total CM energies of the non-signal charged-track (pion mass hypothesis) and photon candidates and the beam energy.[3] For a perfectly reconstructed B, ΔE should be zero.

Because of detector-related and particle misidentification effects, ΔE is limited in discrimination power. If the detector has perfect coverage, resolution, PID, and ability to detect all neutral particles, ΔE is zero for a signal event, while for a background event, ΔE is nonzero since there will be too many or too few track or neutral candidates assigned to the recoil B.

However, if we relax our perfect detector conditions, throwing out neutrinos, and ignoring charged and neutral particles outside of the fiducial volume, ΔE becomes much less useful. The ΔE distribution for signal is further smeared out when one includes other factors like resolution, particle misidentification, merged π^0s, hadronic splitoffs, and unreconstructed or misreconstructed neutral particles such as neutrons and K_L^0s.

[3] As stated earlier, we define the beam energy as the energy of a single beam in the CM frame, unless stated otherwise.

6.2.2 Recoil Invariant Mass, $m_{\mathbf{ES}}$

The variable $m_{\rm ES}$ is the invariant mass of the recoil B calculated from its unscaled CM 3-momentum and the known beam energy (in order to remove the effect of the resolution of the measured energy):

$$m_{\rm ES} = \sqrt{E^2_{\rm beam} - |\vec{p}_{{\rm rec}B,{\rm unscaled}}|^2}. \tag{6.5}$$

For a perfectly reconstructed B, $m_{\rm ES}$ should be the B mass.

Because the endpoint of the $m_{\rm ES}$ spectrum is dependent on the beam energy (an important issue in the off-peak data, where there is no B), we compensate to first order by adding to $m_{\rm ES}$ a value equal to the difference between the nominal B energy and the actual B energy (which we take to be the beam energy). As in the adjustment of $|\vec{p}_B|$ for off-peak energies, described in section 3.1.2, we use a nominal on-peak beam energy value defined by $E^*_{\rm beam,MC}$.

Because $m_{\rm ES}$ is not nearly as sensitive to detector-related effects as ΔE and has a different shape for signal, B background, and continuum background, we use its shape as an input to our signal extraction fit (chapter 9).

6.3 Neutrino Reconstruction

6.3.1 Neutrino Quality Variable, nuEP

6.3.1.1 Variable Description

We described the details of the neutrino reconstruction in section 3.1.2. To check the quality of our neutrino reconstruction, we define a variable, nuEP, which is the difference between the energy and the 3-momentum magnitude of the scaled neutrino candidate (nuEP $\equiv E_\nu - |\vec{p}_{\nu,\mathrm{scaled}}|$). When the neutrino is reconstructed perfectly for a signal event, the value of nuEP should be 0. Detector effects and imperfect reconstruction smear out the nuEP distribution for signal (and background).

In background events, the nuEP distribution does not peak at zero, but instead tends to rise with nuEP. Because the nuEP distribution has a different shape for signal, B background, and continuum background, in particular in combination with m_{ES}, we use their 2D distributions as inputs to our signal extraction fit, to be described in chapter 9.

The reason we use the scaled neutrino momentum for the nuEP calculation is that it has a better resolution for signal than a variable calculated using the unscaled neutrino momentum. Figure 6.3 shows the electron signal distributions of nuEP and two alternatives based on the unscaled neutrino momentum: $E_\nu - |\vec{p}_{\mathrm{miss}}|$ and $E_{\mathrm{miss}} - |\vec{p}_{\mathrm{miss}}|$. It is clear from this comparison that the scaled energy minus the scaled momentum magnitude gives the best signal resolution.

For the $B^+ \to \pi^+ \overline{D}^0 (\to K^+\pi^-)$ control sample studies, we have to make some

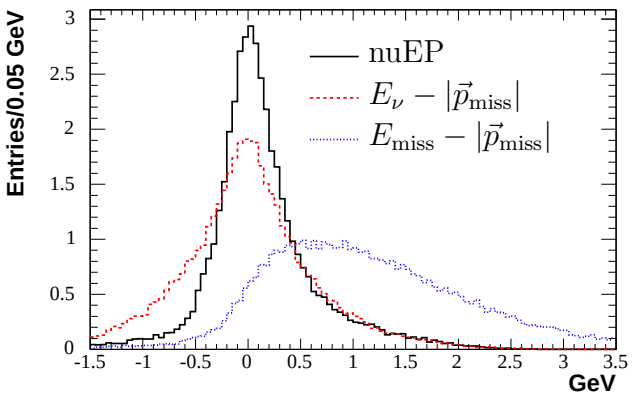


Figure 6.3. Comparison of nuEP with two other computations of the difference of the reconstructed neutrino energy minus scalar momentum, using different pairs of reconstructed values.

changes to the nuEP calculation. For our signal mode, nuEP is calculated assuming a zero mass for the missing particle. In the $B^+ \to \pi^+ \overline{D}^0 (\to K^+ \pi^-)$ control sample, the "neutrino" has the charged pion mass, so nuEP is not a useful variable. So, we define a new variable called nuEPM, which includes the reconstructed candidate mass:

$$\text{nuEPM} \equiv E_\nu - ((\vec{p}_{\nu,\text{scaled}})^2 + m^2)^{1/2}, \tag{6.6}$$

where m is 139.57 GeV for the control sample, and 0 for our $B^+ \to \ell^+ \nu_\ell \gamma$ signal mode. For our $B^+ \to \ell^+ \nu_\ell \gamma$ signal mode, nuEP and nuEPM are equivalent.

6.3.1.2 Correlation with $\cos\theta_{BY}$

It is also worth noting that there is a strong correlation between the nuEP and $\cos\theta_{BY}$ variables, due to similarities in their calculations, as can be seen in 2D distribution in figure 6.4 for signal MC events.

The calculation of nuEP uses the lepton-photon 3-momentum, $\vec{p}_{\ell\gamma}$, and the scaled

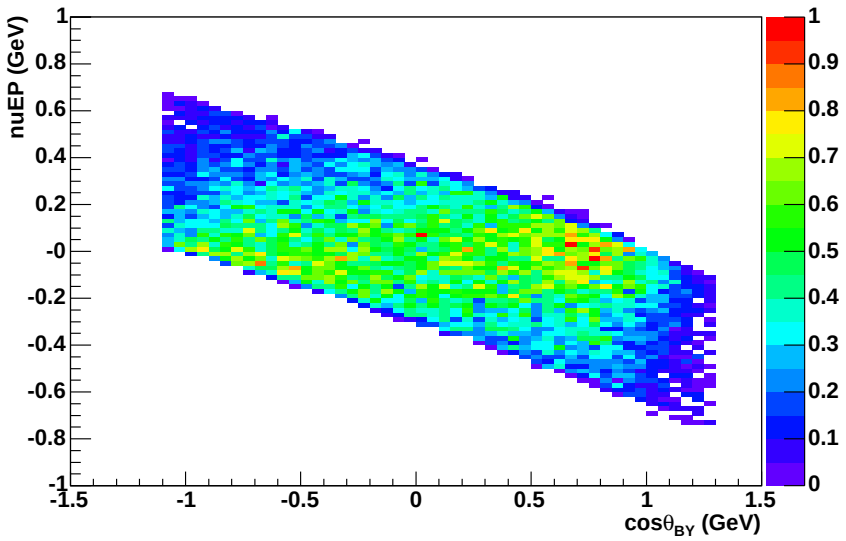


Figure 6.4. Reconstructed electron signal MC distribution of nuEP vs. $\cos\theta_{BY}$, with a basic event selection applied, as well as a restriction that $\cos\theta_{BY}$ be between -1.1 and $1.3\,\mathrm{GeV}$.

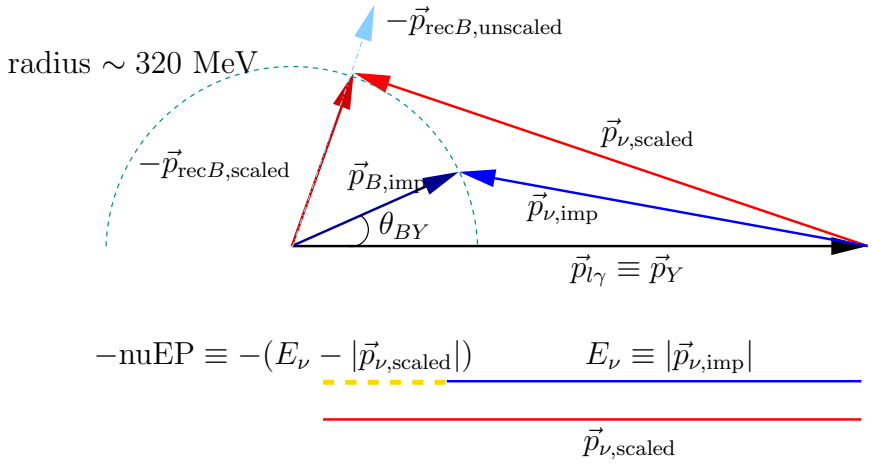


Figure 6.5. A geometric illustration of the neutrino reconstruction in both the nuEP and $\cos\theta_{BY}$ calculations. The "imp" refers to a implied reconstructed vector in the $\cos\theta_{BY}$ calculation when assuming a massless neutrino and a fixed B momentum of about $320\,\mathrm{MeV}$.

neutrino 3-momentum obtained by using the scaled recoil B candidate, with its direction obtained from the sum of visible charged tracks and neutral candidates (excluding the signal candidates) and its magnitude set equal to the expected scalar B momentum (as described earlier in section 3.1.2).

As detailed in section 6.1.3, the variable $\cos\theta_{BY}$ is calculated assuming that the signal decay occurs, i.e. there is a B with a momentum of about 320 MeV in the CM frame which decays to the lepton and photon candidates and a massless neutrino. This construction uses the same lepton-photon 3-momentum, $\vec{p}_{\ell\gamma}$, as used for nuEP. The calculation of the angle between $\vec{p}_{\ell\gamma}$ and the signal B momentum 3-vector leaves an ambiguity in the azimuthal angle about $\vec{p}_{\ell\gamma}$ for the direction of the signal B. Up to this ambiguity, if the lepton and photon are reconstructed perfectly, then the signal B, recoil B, and neutrino 3-momenta will also be reconstructed perfectly.

The two constructions are illustrated in figure 6.5, in which the azimuth of the 3-momentum of the recoil B in the $\cos\theta_{BY}$ reconstruction has been rotated about $\vec{p}_{\ell\gamma}$ to lie in the same half plane as the scaled recoil B 3-momentum in the nuEP reconstruction.

The two reconstructed B directions will most likely not be the same; the measured recoil B reconstruction is unlikely to be perfect, while the $\cos\theta_{BY}$ reconstruction, independent of the recoil B, relies on different assumptions. Because of these differences in the recoil B reconstruction, the resulting neutrino 3-vectors in turn will also be different; nuEP can be viewed as the difference between the length of the reconstructed 3-momentum vectors in the $\cos\theta_{BY}$ and nuEP reconstructions.

The correlation between $\cos\theta_{BY}$ and nuEP arises because for a given value of $\cos\theta_{BY}$, the minimum and maximum of values of nuEP are constrained, as can be seen in figure 6.5. For example, if $\cos\theta_{BY}$ is one, then θ_{BY} is zero so that the reconstructed B direction and the $\ell\gamma$ direction are the same. For a perfect reconstruction in the nuEP calculation, we would get the direction of the scaled reconstructed recoil B correct and nuEP would be zero. If we get the direction of the reconstructed scaled recoil B completely wrong, i.e. 180 degrees off, then nuEP would be twice the radius of the circle, or about 640 MeV (with an overall minus sign). As can be seen figure 6.4, the "band" in nuEP for a given value of $\cos\theta_{BY}$ is indeed about 640 MeV wide.

6.3.2 Neutrino Polar Angle

To help ensure that our missing particle is indeed a neutrino and not merely some other unobserved particle, we place a fiducial cut on the neutrino lab momentum direction, requiring it to point into the detector. This cut is fixed and not optimized. We found it useful to place fiducial cuts on both the scaled and unscaled neutrino polar angles. In our tables, these are labeled the scaled and unscaled ν lab θs.

6.4 Event Shape

Our event shape variables are largely useful for discriminating between signal and continuum background, where events are topologically jettier.

6.4.1 Thrust Variable

The $|\cos\theta_T|$ variable is the absolute value of the cosine of the angle between the thrust axes of the reconstructed B and lepton-photon system in the CM frame. The thrust axis is defined as the axis which minimizes the total transverse momentum for a set of particles.

In our signal decay, we expect the orientation of the thrust axis for the lepton-photon system to be uncorrelated with that of the recoil B for two reasons. First, in the rest frame of the B, we expect the orientation of its thrust axis to be uniformly distributed over the solid angle. While we expect the boost of the B to the CM frame to introduce correlations, the magnitude of the boost is not large enough to result in a significant correlation of the resultant CM-based thrusts. Second, removing the missing neutrino from the signal B thrust calculation guarantees that the thrust calculated using only the lepton-photon system cannot be correlated with that of the recoil B.

However, in a jet-like event where the lepton and photon come from opposite jets, we expect a correlation, so the angle between the thrust vectors of the lepton-photon system and the recoil B candidate should be a discriminant against continuum background. This correlation can also occur in a nonsignal $B\overline{B}$ event when the reconstructed signal lepton and photon come from opposite B mesons. One would expect a bias towards large thrust angles, that is higher values of $|\cos\theta_T|$.

After determining the thrust axis, we arbitrarily choose one of its two directions in order to define a thrust 3-vector. We then calculate the cosine of the angle between

the thrust 3-vectors using the dot product of the recoil thrust vector and the $\ell\gamma$ thrust vector. Since the directions of the vectors are arbitrarily chosen, the angle can be either θ or $\pi - \theta$, and so the cosine is ambiguous up to a sign. Therefore we take the absolute value of the cosine of the angle.

6.4.2 Fisher Discriminant, $\mathcal{F}$

The purpose of a Fisher discriminant is to perform a parametric multivariate classification by first training on samples that have been preclassified. We use the Fisher discriminant to remove continuum background, and therefore include event shape variables in the discriminant. By making the assumption that the joint probability distribution functions of signal and background are multivariate gaussians,[4] one can use the prior probabilities of signal and background (their expected fraction) to calculate the log-odds of an event being signal or background. A set of decision boundaries (separating hyperplanes) are parameterized by the log-odds criteria. For example, one could choose a simple decision boundary such that the log-odds ratio is zero (equal probabilities). However, we use a figure-of-merit to determine our decision boundary.

Linear Fisher discriminant analysis involves the assumption that the signal and background both share the same covariance matrix, but have different first moments. Quadratic discriminant analysis does not involve this assumption. The only things that are needed for the calculation of the Fisher discriminant are the first and second moments of the signal and background samples, as well as the prior probabilities, and

[4] Our event shape selection criteria are not distributed as multivariate gaussians, but one can relax this assumption as long as there is a decent signal and background separation. We are not concerned with the statistical properties of the discriminant, but rather only its classification performance.

these values are readily calculable.

Based on a pure MC-based analysis (using continuum MC for continuum background), we found that the most effective discriminant between signal and background was a linear Fisher discriminant calculated with the momentum-weighted 0th and 2nd Legendre moments (L_0, L_2) of the recoil B about the lepton-photon thrust axis in the CM frame. These moments measure the energy flow about the lepton-photon vector.

One calculates the weighted Legendre moments as follows: Start with the thrust vector, $\vec{t}$, of the lepton-photon system in the CM frame. Using the second Legendre polynomial, $P_2(x) = (1/2)(3x^2 - 1)$, calculate L_2 as:

$$L_2 = \frac{1}{2} \sum_i |\vec{p}_i|(3\cos^2\theta_i - 1), \tag{6.7}$$

where $\vec{p}_i$ are the CM 3-momenta of the tracks and neutrals that comprise the recoil B candidate, and θ_i are the angles between $\vec{p}_i$ and $\vec{t}$.

Similarly, using the zeroth Legendre polynomial, $P_0(x) = 1$, L_0 is defined as:

$$L_0 = \sum_i |\vec{p}_i|. \tag{6.8}$$

We added the 1st and 3rd Legendre moments to the Fisher discriminant, as well as the thrust variable, but did not find that any of these additions improved the Fisher performance. Neither did using a quadratic Fisher discriminant. The coefficients of the Fisher discriminant were determined using MC events with some basic selection criteria applied.

The equation of the electron-channel linear Fisher discriminant is:

$$\mathcal{F}_{\mathrm{E}} = -4.1761 + 0.427743 L_0 + -1.85883 L_2. \tag{6.9}$$

The equation of the muon-channel linear Fisher discriminant is:

$$\mathcal{F}_{\mathrm{M}} = -3.11153 + 0.00788145 L_0 + -1.59365 L_2. \tag{6.10}$$

6.4.3 $R2_{\mathrm{All}}$

This variable is the ratio of the 2nd to the 0th Fox-Wolfram moments [35] using all charged and neutral particles in the event, in the CM frame.

The definition of the lth Fox-Wolfram moment is:

$$H_\ell \equiv \frac{1}{E_{tot}^2} \sum_{i,j} |\vec{p}_i||\vec{p}_j| P_\ell(\cos\theta_{ij}), \tag{6.11}$$

where i and j are index sets over the list of candidates, E_{tot} is the total energy of the candidates, θ_{ij} is the opening angle between the ith and jth candidates, and P_ℓ is the ℓth-order Legendre polynomial.

$R2_{\mathrm{All}}$ lies between 0 and 1. Spherical, $B\overline{B}$ events tend to have lower values of $R2_{\mathrm{All}}$, while jet-like continuum events tend to have their $R2_{\mathrm{All}}$ values shifted higher. Thus, $R2_{\mathrm{All}}$ is useful for removing continuum background. However, we fix the value of $R2_{\mathrm{All}}$ in our analysis because the Fisher discriminant is more useful for removing continuum background, and allowing the cut value of $R2_{\mathrm{All}}$ to float in the optimization

as well led to overtraining.

6.5 Anti-Two-Photon

Due to our inability to model generic two-photon processes in MC, we use off-peak data in our final signal extraction to model the background from continuum events. Because of limited off-peak statistics, this procedure contributes significantly to the statistical uncertainty on the result. It is therefore useful to reduce the size of the off-peak contribution, even in the absence of a two-photon simulation.

6.5.1 Cosine of the Lepton Angle, $\cos\theta_\ell$

Because there is no preferred z-direction for the signal lepton in the CM frame, we expect the distribution of the cosine of its CM polar angle θ to be flat. (The same holds for any background events originating from B decays.) On the other hand, for single-tagged two-photon events we would expect an excess of initial state electrons with small-angle scattering, resulting in a charge-angle correlation for these events in the electron-channel analysis, with electrons (positrons) heading in the positive (negative) z-direction.

For both the electron and muon channels, other higher-order QED processes may also be a source of background. The processes can produce an excess of leptons in the forward and backward directions.

The details of the optimization of this variable will be described in section 8.3.

6.5.2 Two-Photon Parameters (Electron and Muon)

We consider other anti-two-photon variables based on physical grounds, study their behavior in the off-peak data in the clpi0 sideband region (defined earlier in section 6.1.4.2 to be between 124 and 146 MeV), and then apply cuts on the potentially useful ones in our final event selection. There is no guarantee that the behavior of these variables will be the same between the clpi0 signal and sideband regions. Choosing cut boundaries so as to maintain high signal efficiency minimizes this concern.

6.5.2.1 Electron Two-Photon Parameter

In the electron channel, we looked at a number of potential anti-two-photon selection variables (besides the lepton angle). The two most useful ones were TagLepW and actNuPZ.

TagLepW is defined as the invariant mass, W, of the two-photon system, calculated using all charged and neutral candidates except the signal lepton. Thus, it assumes the hypothesis that the event is a single-tagged two-photon event, with the signal lepton being the tag lepton, and the other lepton disappearing down the beampipe. We would expect the distribution of TagLepW to be distinctive for single-tagged two-photon events.

The other variable is actNuPZ, which is the missing CM energy in the z-direction. We would expect that in a single-tagged event, there would be a longitudinal momentum imbalance arising from an initial-state electron or positron going down the beampipe unobserved.

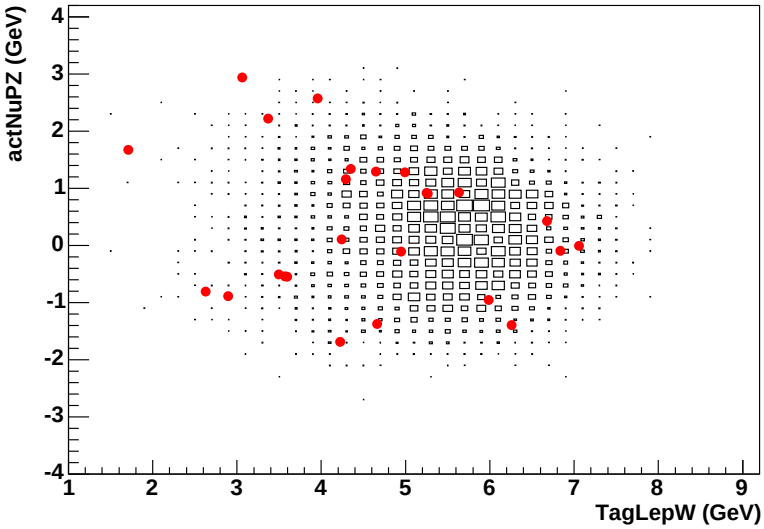


Figure 6.6. Distribution of actNuPZ vs. TagLepW for the electron channel, validation sample, in the clpi0 sideband region (defined to be between 124 and 146 MeV), after the rest of the final analysis cuts (table 8.2) have been applied. In addition, we apply cuts on nuEP ($[-1, 0.1]$ GeV) and m_{ES} ($[4.9, 5.3]$ GeV), corresponding to a nuEP signal region and expanded m_{ES} signal and sideband region. The box plot is the signal distribution. The solid red circles are the off-peak data.

By plotting TagLepW vs. actNuPZ in the clpi0 sideband, we found a distinct elliptical shape for the distribution in signal MC, peaking strongly in the center, while the off-peak data exhibited a much flatter, scattered distribution, with numerous events far from the core of the signal peak, as can be seen in figure 6.6.

Our strategy was thus to fit the signal distribution shape in the clpi0 signal region with an ad hoc PDF and place a cut on the value of the PDF—effectively defining a contour in the TagLepW–actNuPZ plane.[5] By parameterizing a set of contours in signal MC using one parameter, one has a potentially effective tool for removing two-photon background, with little loss in signal efficiency. We refer to this parameter in the electron channel as the "electron two-photon parameter."

We determine the electron two photon parameter by first parameterizing both the TagLepW and actNuPZ distributions with double gaussians. The electron two-photon parameter is then a function of the log of the product of these double gaussians.

Figure 6.7 shows the separate fits to the TagLepW and actNuPZ distributions in the electron signal MC training sample,[6] with a basic set of selection criteria applied. We combined these double gaussians by multiplying them. The maximum of this product is approximately 16.9, so we divided the product by 16.9. Then, we took the negative of the log of this normalized product so that the parameter is distributed with a maximum near 0, tailing off to 0 in the positive direction. In summary, the

[5] The TagLepW vs. actNuPZ distribution peaks also in the clpi0 signal region for signal events.

[6] As a reminder, we split the MC samples 50/50 into a training and validation sample for the purposes of determining unbiased selection criteria.

electron two-photon parameter is:

$$-\ln\left[\frac{f(\text{TagLepW})g(\text{actNuPZ})}{16.9}\right] \tag{6.12}$$

with the TagLepW double gaussian:

$$f(x) = 2.5e^{\left[\frac{1}{2}\left(\frac{x-5.1}{0.76}\right)^2\right]} + 3.1e^{\left[\frac{1}{2}\left(\frac{x-6.2}{0.59}\right)^2\right]} \tag{6.13}$$

and the actNuPZ double gaussian:

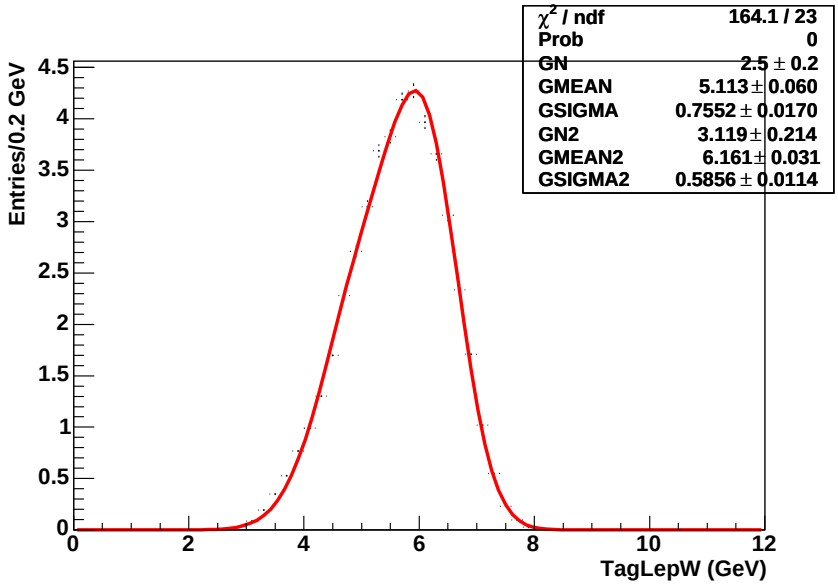
$$g(y) = 3.1e^{\left[\frac{1}{2}\left(\frac{y-0.7}{0.62}\right)^2\right]} + 2.9e^{\left[\frac{1}{2}\left(\frac{y+0.5}{0.62}\right)^2\right]}. \tag{6.14}$$

Figure 6.8 shows the distribution of the electron two-photon parameter and some contours of the distribution, as applied to the signal MC validation sample.

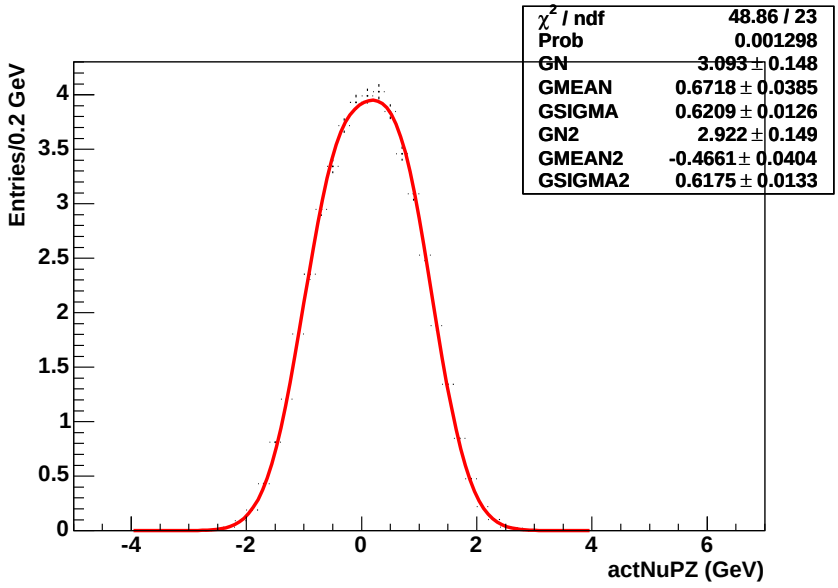
The next step was to "optimize" the electron two-photon parameter against background. We found that an electron two-photon parameter cut of <2.34 was reasonable, in that it excludes a good portion of the tail-end off-peak background seen in the clpi0 sideband, and had a reasonably high marginal signal efficiency in the validation sample of 96%.

6.5.2.2 Muon Two-Photon Parameter

In the muon channel, we expect less two-photon background because single-tagged two-photon events, in which a beam electron or positron is reconstructed as the signal muon, do not contribute. However, signal muons could still originate from

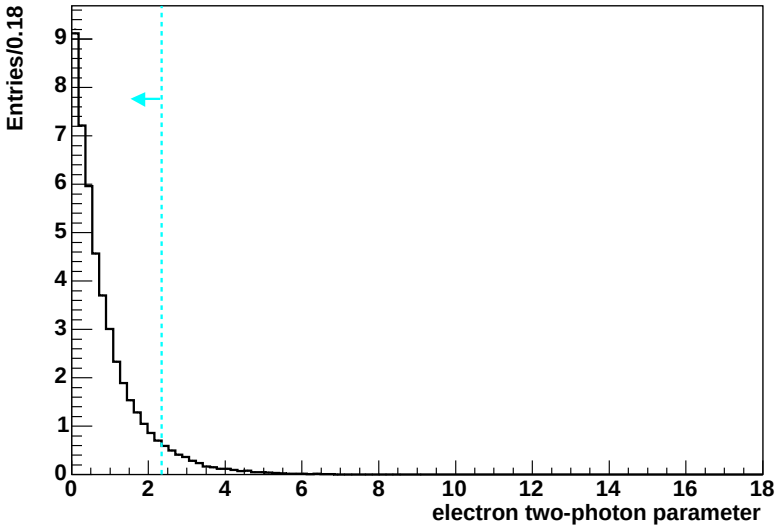


(a) Fit to TagLepW.

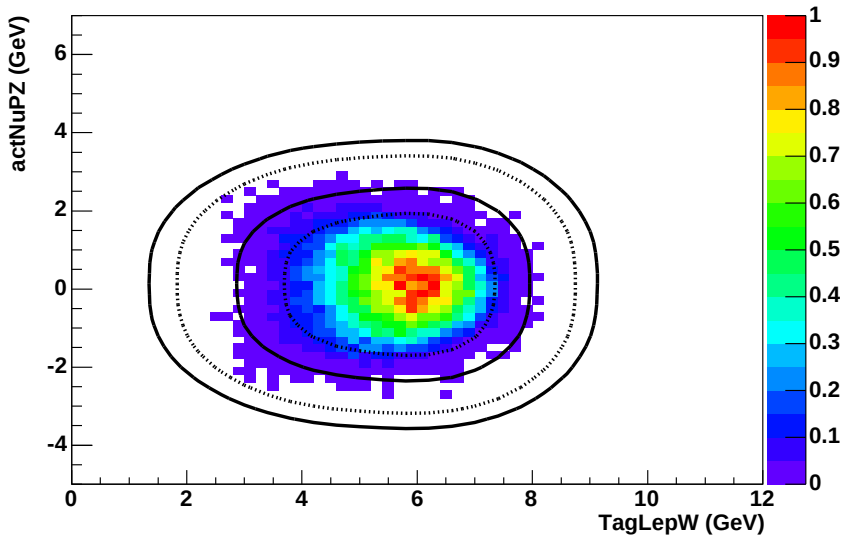


(b) Fit to actNuPZ.

Figure 6.7. Double-gaussian fits to the electron signal MC training sample distributions for determining the electron two-photon parameter.



(a) Distribution of the electron two-photon parameter.



(b) Some contours (2.34, 5, 10, 13) of the electron two-photon parameter, plotted over the distribution of actNuPZ vs. TagLepW.

Figure 6.8. Distribution and contours of the electron two-photon parameter distribution for the electron signal MC validation sample. The innermost contour at 2.34 is the location of the cut.

a two-photon system—either true muons, e.g. from $c \rightarrow s\mu\nu_\mu$ decays, or hadrons misidentified as muons.

Because the signal muon is not expected to arise from the initial state, the use of TagLepW is inappropriate. Instead, we looked at a different variable, AllW, defined as the invariant mass, W, of the hypothetical two-photon system, now calculated using all charged and neutral candidates. Thus, it assumes the hypothesis that this event is an untagged two-photon event, with both initial-state leptons disappearing down the beampipe. We would expect the distribution of W to be distinct for untagged two-photon events.

We also use the actNuPZ variable that is used in the electron-channel analysis. However, given that both beam leptons disappear down the beampipe in an untagged two-photon event, we would expect less of a longitudinal CM momentum imbalance vs. an single-tagged two-photon event, though we do not expect the deflected beam leptons to have completely counterbalanced momenta.

By plotting AllW vs. actNuPZ, we found a distinct elliptical shape to the contours in the signal MC, peaking strongly in the center, while the off-peak sideband data exhibited a somewhat more scattered distribution, as can be seen in figure 6.9. As with the electron-channel analysis, we parameterized the set of contours in signal MC using one parameter, the "muon two-photon parameter." It is clear, however, the muon two-photon parameter is less powerful than the electron two-photon parameter. Nonetheless, we place a fixed cut with little loss in signal efficiency, in the hope that it does remove two-photon background.

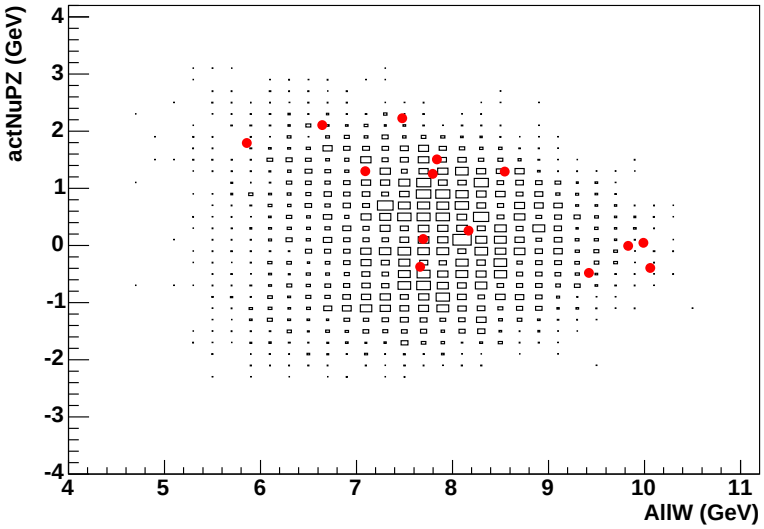


Figure 6.9. Distribution of actNuPZ vs. AllW for the muon channel, validation sample, in the clpi0 sideband region (defined to be between 124 and 146 MeV), after the rest of the final analysis cuts (table 8.3) have been applied. In addition, we apply cuts on nuEP ($[-1, 0.1]$ GeV) and m_{ES} ($[4.9, 5.3]$ GeV), corresponding to a nuEP signal region and expanded m_{ES} signal and sideband region. The box plot is the signal distribution. The solid red circles are the off-peak data.

We determine the muon two-photon parameter by fitting AllW to a double gaussian and actNuPZ to a sum of three gaussians (triple gaussian). The muon two-photon parameter is then a function of the log of the product of the double and triple gaussians.

Figure 6.10 shows the separate fits to the AllW and actNuPZ distributions in the muon signal MC training sample, with some basic selection criteria applied. We combine the double and triple gaussians by multiplying them. The maximum of this product is approximately 5.19, so we divided the product by 5.2. Then, we took the negative of the log of this normalized product, so that the parameter is distributed with a maximum near 0, tailing off to 0 in the positive direction. In summary, the muon two-photon parameter is:

$$-\ln\left[\frac{f(\text{AllW})g(\text{actNuPZ})}{5.2}\right] \tag{6.15}$$

with the AllW double gaussian:

$$f(x) = 1.2e^{\left[\frac{1}{2}\left(\frac{x-7.6}{0.84}\right)^2\right]} + 1.6e^{\left[\frac{1}{2}\left(\frac{x-8.7}{0.63}\right)^2\right]} \tag{6.16}$$

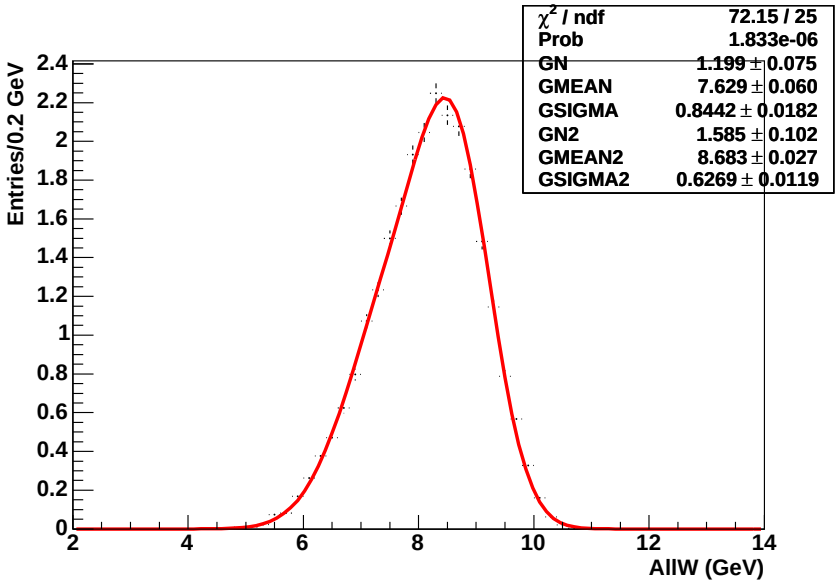
and the actNuPZ triple gaussian:

$$g(y) = 1.5e^{\left[\frac{1}{2}\left(\frac{y-0.7}{0.58}\right)^2\right]} + 5.9e^{\left[\frac{1}{2}\left(\frac{y+1.1}{0.33}\right)^2\right]} + 1.9e^{\left[\frac{1}{2}\left(\frac{y+0.3}{0.51}\right)^2\right]}. \tag{6.17}$$

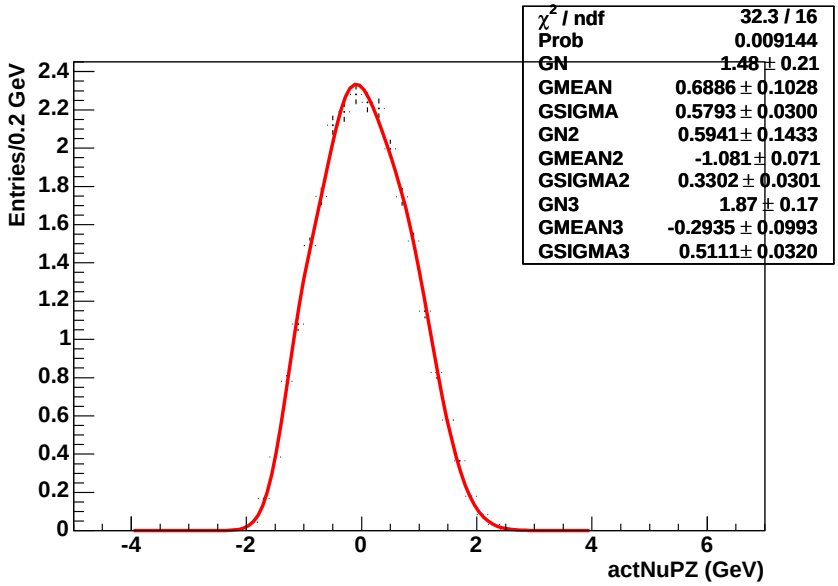
Figure 6.11 shows the distribution of the muon two-photon parameter and some

contours of the distribution, as applied to the signal MC validation sample.

The next step was to "optimize" the muon two-photon parameter against background. We found that a muon two-photon parameter cut of <2.88 was reasonable, in that it excludes a good portion of the tail-end off-peak background seen in the clpi0 sideband, and had a reasonably high marginal signal efficiency in the validation sample of 97%.

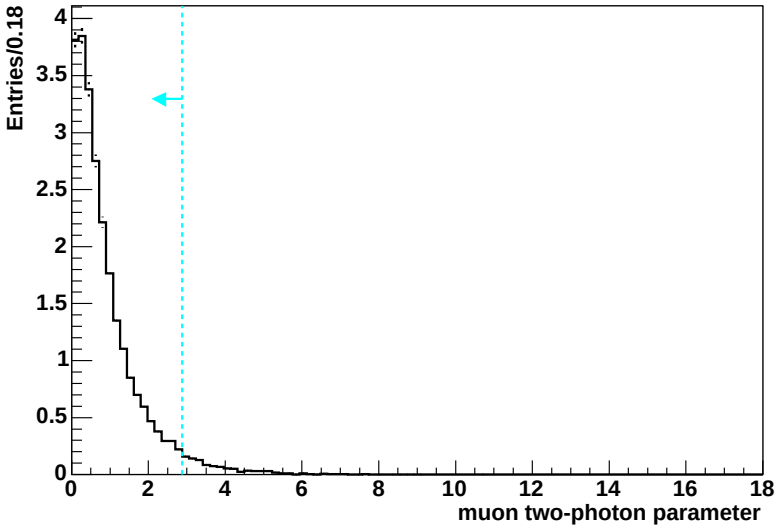


(a) Fit to AllW.

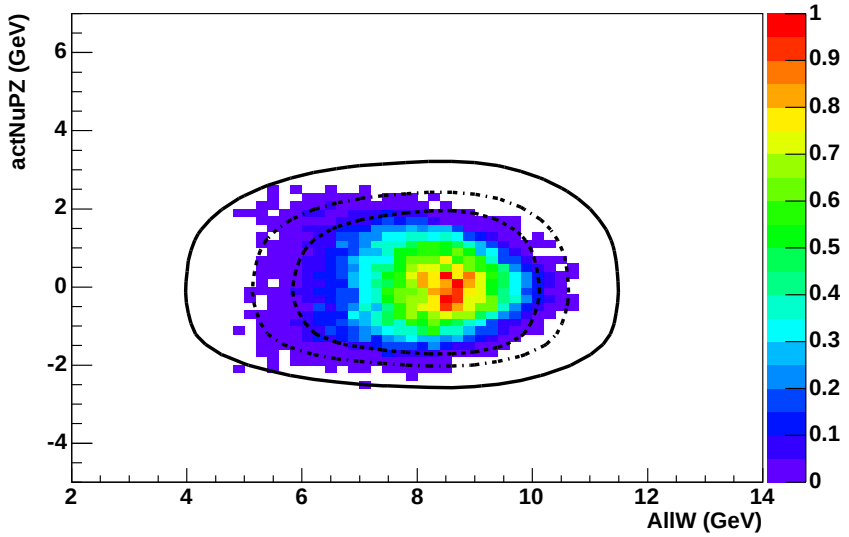


(b) Fit to actNuPZ.

Figure 6.10. Double-gaussian fits to the muon signal MC training sample distributions for determining the muon two-photon parameter.



(a) Distribution of the muon two-photon parameter.



(b) Some contours (2.88, 5, 10) of the muon two-photon parameter, plotted over the distribution of actNuPZ vs. TagLepW.

Figure 6.11. Distribution and contours of the muon two-photon parameter distribution for the muon signal MC validation sample. The innermost contour at 2.88 is the location of the cut.

Chapter 7

Preliminary Selection

In this chapter we discuss event reconstruction and preliminary event selection. Preliminary event selection is a necessary step before we optimize our selection criteria, due to limitations in data storage and processing speeds.

7.1 Event Reconstruction

Starting with real event data that has been stored by the logging manager after passing the L3 software trigger (section 2.8), or simulated MC samples that must pass the same criteria, we reconstruct our signal event by first finding a signal lepton and a signal photon, and then combining the remaining visible charged tracks and photons into a recoil B.

The signal-side selection is based on finding the highest CM energy lepton and highest CM energy photon in the event. By looking at generator-level MC, we can determine the accuracy of the highest energy selection, with and without energy cuts on the lepton and photon. The results are summarized in table 7.1. One can see that the highest-energy lepton selection, as expected due to the signal lepton's higher and

sharper peak, is more accurate than the highest-energy photon selection.

Table 7.1. Accuracy of the highest-energy lepton, photon selections for the electron channel, KPY model. "Cuts" refer to $1.9 < E_l < 3.0$ GeV and $0.6 < E_\gamma < 2.8$ GeV.

Type	Percent Accuracy
Correct electron	99.2
Correct photon	91.1
Correct electron and photon	90.4
With cuts, correct electron	99.6
With cuts, correct photon	91.8
With cuts, correct electron and photon	91.5

Charged track candidates are reconstructed by their hits in the SVT and DCH. All charged tracks (with the exception of the signal lepton) are assigned the pion mass hypothesis. Lepton identification will be described in more detail in sections 7.1.1 and 7.1.2. Photon candidates (both signal and nonsignal) are reconstructed by looking for "bumps" in the EMC not matched to a charged track. Bumps are required to have a lab energy of less than 10 GeV.

We refine the signal lepton track selection with stricter criteria. We require the signal lepton lab momentum be less than 10 GeV. In addition, we require that the absolute value of its distance-of-closest-approach with respect to the interaction point be less than 1.5 cm in the $x-y$ plane, and less than 10 cm along the z-axis. The signal lepton track is also required to have a minimum transverse momentum of 100 MeV, with a minimum of 12 hits in the DCH.

7.1.1 Electron ID

Electron particle identification is based largely on information from the EMC. One particularly useful discriminating variable is E/p, the ratio of deposited energy to charged track momentum. Electrons tend to deposit their entire energy in the EMC, so their measured E/p is close to one. Muons and hadrons, on the other hand, deposit less energy in the calorimeter, so their E/p is typically much lower. For example, at 1.2 GeV, the mean E/p for an electron is about 0.9, while it is about 0.4 for a charged pion, and 0.15 for a muon. (all widths are less than 0.05). However, there are tails due to shower leakage, insensitive calorimeter material, and other resolution effects.

Another useful electron ID variable is the lateral moment of the shower shape, described earlier in section 6.1.5. Electrons tend to have more localized showers, so their lateral moments are smaller than those of hadrons.

The efficiencies of the electron identification for electrons, pions, kaons, and protons are approximately 94%, 0.05%, 0.4%, and 0.2%, respectively.[1]

7.1.2 Muon ID

The IFR provides the most useful information for discriminating muons from pions. In particular, IFR penetration and cluster shape are distinguishing criteria. In addition, the DIRC can provide supplementary identification information.

The efficiencies of the muon identification for muons, pions, kaons, and protons are approximately 50%, 2%, 1%, and 1%, respectively.

[1] These efficiencies vary by lab angle, lab momentum, charged, run period, and between data and MC. These are average numbers.

7.1.3 Data/MC Corrections

In order to bring our data and MC samples into better agreement, particle ID (PID) efficiencies are adjusted in the MC using an algorithm called "PID tweaking." This involves using PID efficiency tables to accept or reject particles from PID lists in such a way so as to bring the MC PID efficiencies and misidentification rates more in line with data.[2] More specifically, for each particle in a PID list, PID tweaking will:

- reject an accepted track with probability $1 - \epsilon_{\text{data}}/\epsilon_{\text{MC}}$ if $\epsilon_{\text{data}} < \epsilon_{\text{MC}}$, or
- accept a rejected track with probability $1 - (1 - \epsilon_{\text{data}})/(1 - \epsilon_{\text{MC}})$ if $\epsilon_{\text{data}} > \epsilon_{\text{MC}}$,

where ϵ is the efficiency.

7.2 Preliminary Event Selection

In order to reduce disk storage and processing time, we apply some preliminary selection criteria to our samples. In general, the preliminary selection criteria were chosen to be looser than the anticipated final criteria, so that their efficiency effects are transparent in the final analysis. In addition, some of the preliminary selection criteria were chosen because their values are not allowed to float in the optimization. The criteria are given in table 7.2.

[2] PID efficiency tables are determined using specific PID control samples.

Table 7.2. Preliminary selection criteria for both the electron- and muon-channel analyses. "Lepton," "photon," and "neutrino" (ν) refer to the signal candidates.

Selection Variable	Criteria
photon lab θ	[0.326, 2.443]
scaled ν lab θ	[0.3, 2.443]
unscaled ν lab θ	[0.3, 2.443]
$R2_{\mathrm{All}}$	<0.5
$\cos\theta_{\ell\gamma}$	<0
$\cos\theta_{BY}$	$(-1.1, 1.3)$
electron photon-conv veto (e channel only)	Passed
J/ψ veto	Passed
lepton CM E (GeV)	(1.3, 3.2)
photon CM E (GeV)	(0, 3.2)
gLAT	<0.55
ΔE (GeV)	<2
m_{ES} (GeV)	>4.6

Chapter 8

Cut Optimization

Starting with the event sample obtained as described in chapter 7, we refine our event selection through cut optimization. The following is an outline of the cut optimization procedure:

1. Split all MC samples in half into separate training and validation samples. (section 8.1)

2. Optimize the lepton angle cut(s) using a combination of signal MC, nonsignal $B\overline{B}$ MC, and off-peak data. (section 8.3)

3. Optimize the rest of the floated cuts, fixing the lepton angle cut(s), on pure MC. (section 8.4)

4. Optimize nuEP and m_{ES} signal and sideband boundaries using toy MC simulation of signal extraction fit. (section 9.3)

8.1 Training/Validation Split

In order to avoid overtraining, we split our MC samples into 2 separate samples, labeled *training* and *validation*. All optimizations were done with the training sample, and their respective performances evaluated using the validation sample.

The only exception was off-peak data, which was not split, due to limited statistics. Because of this, our only use of off-peak data in determining cuts was in loose-cut, high-statistics regimes. In particular, we use off-peak data to optimize the lepton angle cuts, as will be described below in section 8.3.

8.2 Figure-of-Merit (FOM)

Before defining the FOM used in our optimization, let us start with the signal significance, ξ, defined as the ratio of signal to error on the measurement, S/E, where S is the number of signal events, and E is the error, or uncertainty, on the measured number of events. If we assume Poisson statistics with no separate uncertainty on the background, then the significance is equal to $S/\sqrt{S+B}$, where B is the number of background events scaled to the on-peak luminosity. In order to arrive at the FOM, we set $S = N_B \epsilon_S \mathcal{B}$, assign the significance ξ to be 3, and solve the resulting quadratic equation for the branching fraction, $\mathcal{B}$, of the signal:[1]

$$\mathcal{B} = \frac{9 + \sqrt{81 + 36B}}{2N_B\epsilon_S}, \tag{8.1}$$

[1]There are two solutions to the quadratic equation, but only one yields a positive branching fraction.

where N_B is the total number of charged and neutral B mesons, and ϵ_S is the signal efficiency. This branching fraction is our FOM, and our optimization procedure chooses the set of cuts which minimizes the FOM in the training sample.

However, if there is an uncertainty, σ_B, associated with the background estimate, then the significance is equal to $S/\sqrt{S+B+\sigma_B^2}$. This variance arises from the Poisson statistics of finite-sized samples.

We can extend our definition of significance to include multiple background sources:

$$\xi = \frac{S}{\sqrt{S+\sum_i (B_i+\sigma_{Bi}^2)}} = \frac{S}{\sqrt{S+\sigma_{BTot}^2}}, \tag{8.2}$$

where i is an index set of the background sources. The equation for the FOM is now written as:

$$\text{FOM} \equiv \text{BF} = \frac{9+\sqrt{81+36\sigma_{BTot}^2}}{2N_B\epsilon_S}. \tag{8.3}$$

Including the variance on the separate background sources is important, because we want to account for the larger variances associated with continuum background due to the off-peak fit, relative to the variances associated with B background. Since we are optimizing on continuum MC and not off-peak data, and the continuum MC background itself has relatively small uncertainties, we scale up these uncertainties by the off-peak luminosity factors (on the order of 8-10 depending on the run period) to capture the magnitude of the off-peak uncertainties. One would expect that by doing so, the optimization procedure should preferentially remove more continuum background relative to B background.

8.3 Preliminary Optimization: Lepton Lab Angle

We optimize the cut on the lepton lab angle using signal MC, nonsignal $B\overline{B}$ MC, and off-peak data. The inclusion of off-peak data is to provide a two-photon component to the spectrum. We optimize the angle separately by charge for the electron channel, but do not do so for the muon channel, since a charge asymmetry should only arise from initial-state electrons or positrons from the beam. The results of the optimization can be seen in table 8.1. The corresponding FOM and efficiency plots can be seen in figures 8.1 and 8.2.

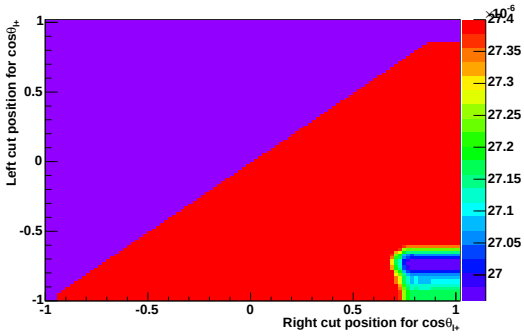
Table 8.1. Optimized cuts on the signal lepton candidate lab angle. Cuts are only applied where applicable.

• Signal e^+ lab angle: $-0.74 < \cos(\theta_{e^+}) < 0.78$
• Signal e^- lab angle: $-0.94 < \cos(\theta_{e^-}) < 0.7$
• Signal muon lab angle: $-1.0 < \cos(\theta_{\mu}) < 0.78$

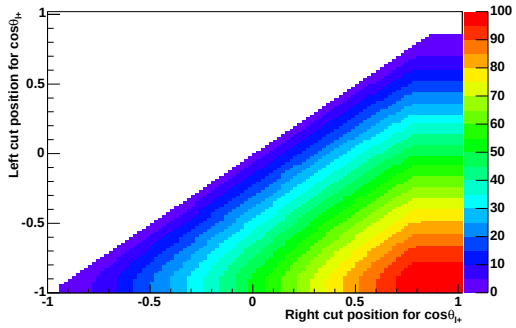
8.4 Iterative Cut Optimization Procedure

We use an iterative cut optimization procedure to select an optimized hyperrectangular region in our multivariate space. The procedure for the iterative optimization is:

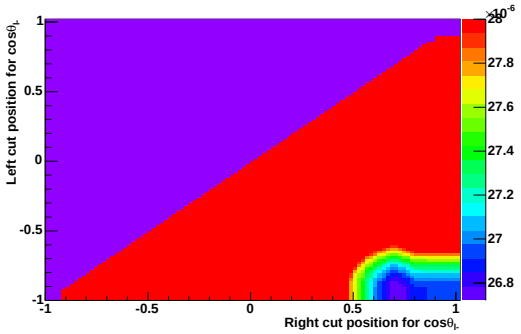
1. Using the training sample, fill an FOM histogram for each adjustable cut variable, where each bin corresponds to a particular cut position (one-sided) or a pair of cut positions (two-sided), with all other cuts applied.



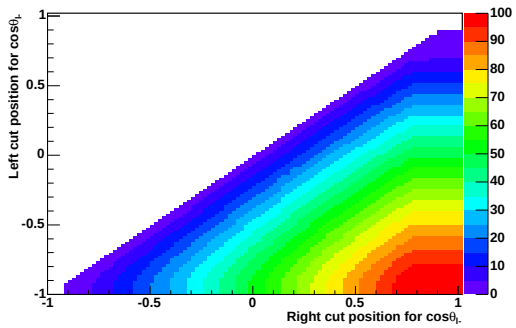
(a) FOM distribution for positrons.



(b) Efficiency distribution for positrons.

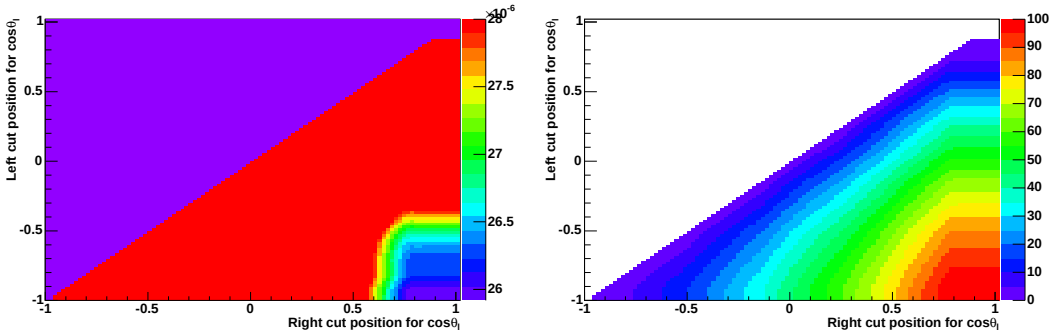


(c) FOM distribution for electrons.



(d) Efficiency distribution electrons.

Figure 8.1. Electron-channel results of the optimization of the signal electron cosine lab angle cuts, performed separately for positive leptons (upper) and negative leptons (lower). Sample is a combination of training sample MC (signal and $B\overline{B}$) and all off-peak data.



(a) FOM distribution for muons (both charges).

(b) Efficiency distribution for muons (both charges).

Figure 8.2. Muon-channel results of the optimization of the signal muon cosine lab angle cuts. Sample is a combination of training sample MC (signal and $B\overline{B}$) and all off-peak data.

2. For each floating cut variable, find the cut or pair of cuts that produces the optimal FOM.

3. Out of these variables, find the cut variable for which its optimization produces the optimal FOM.

4. Apply the cut or pair of cuts corresponding to the optimal FOM.

5. Repeat steps 1-4.

   - If the new optimal FOM is better, repeat steps 1-4.
   - If the new optimal FOM is worse and less than 10 iterations have been attempted, repeat steps 1-4.
   - If the new optimal FOM is worse and at least 10 iterations have been attempted, unapply only the last cut change or pair of cuts changes, but keep any other changes. This set of changes corresponds to the optimal

set.

The final selection criteria obtained from the optimization are given in tables 8.2 and 8.3 for the electron- and muon-channel analyses, respectively. These selection criteria are applied to the validation sample, with the selected events used as input to our BF extraction fit.

Table 8.2. Final selection criteria for the electron channel. "Electron" (e), "photon," and "neutrino" (ν) refer to the signal candidates.

Selection Variable	Criteria
photon lab θ	$[0.326, 2.443]$
scaled ν lab θ	$[0.3, 2.443]$
unscaled ν lab θ	$[0.3, 2.443]$
e^+: $\cos\theta_{\rm Lab}$	$(-0.74, 0.78)$
e^-: $\cos\theta_{\rm Lab}$	$(-0.94, 0.7)$
$\lvert\cos\theta_T\rvert$	<0.98
$R2_{\rm All}$	<0.5
$\mathcal{F}_{\rm E}$	> -2.7
$\cos\theta_{\ell\gamma}$	< -0.42
$\cos\theta_{BY}$	$(-1.1, 1.1)$
electron photon-conv veto	Passed
electron J/ψ veto	Passed
lepton CM E (GeV)	$(2, 2.85)$
photon CM E (GeV)	$(0.65, 2.35)$
gLAT	<0.55
e two-photon parameter	<2.34
ΔE (GeV)	<0.9
clpi0 (MeV)	$(\leq 123)\lvert\rvert(\geq 147)$
$m_{\rm ES}$ (GeV)	Fit
nuEP (GeV)	Fit

Table 8.3. Final selection criteria for the muon channel. "Muon" (μ), "photon," and "neutrino" (ν) refer to the signal candidates.

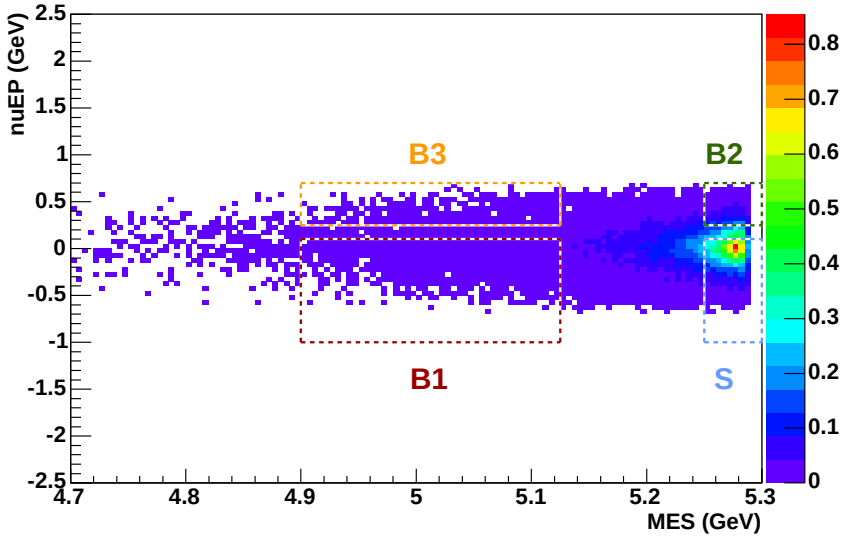
Selection Variable	Criteria
photon lab θ	$[0.326, 2.443]$
scaled ν lab θ	$[0.3, 2.443]$
unscaled ν lab θ	$[0.3, 2.443]$
μ: $\cos\theta_{\rm Lab}$	$(-1.0, 0.78)$
$\lvert\cos\theta_T\rvert$	<0.86
$R2_{\rm All}$	<0.5
$\mathcal{F}_{\rm M}$	> -2.8
$\cos\theta_{\ell\gamma}$	< -0.36
$\cos\theta_{BY}$	$(-1.05, 1.0)$
muon J/ψ veto	Passed
lepton CM E (GeV)	$(1.875, 2.775)$
photon CM E (GeV))	$(0.45, 2.35)$
gLAT	<0.55
μ two-photon parameter	<2.88
ΔE (GeV)	$(-2.5, 0.7)$
clpi0 (MeV)	$(\leq 116)\|\|(\geq 148)$
$m_{\rm ES}$ (GeV)	Fit
nuEP (GeV)	Fit

Chapter 9

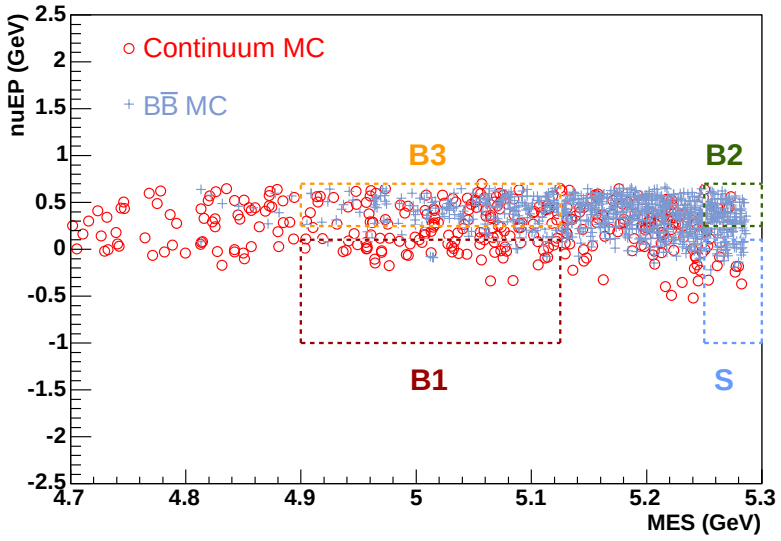
Branching-Fraction Extraction

Our signal extraction procedure is a compromise between a simple cut-and-count analysis and a full-blown maximum likelihood fit with many degrees of freedom. We start with two-dimensional distributions in signal and background in the nuEP and $m_{\rm ES}$ variables. In order to avoid dependence on the fine details of these distributions, we select four relatively large regions in this distribution where the proportions of signal to $B\overline{B}$ to continuum background vary. The region with the largest proportion of signal is labeled the "S" region and the other three (sideband) regions are labeled "B1," "B2," and "B3." In particular, B1 is relatively high in continuum background, B2 is relatively high in $B\overline{B}$ background, and B3 is high in both backgrounds. The relative sizes of the event counts in the four regions have distinct patterns for each of the three sample groups—signal, $B\overline{B}$, and continuum—which should be distinguishable in a fit.

Figure 9.1 shows the distributions of nuEP vs. $m_{\rm ES}$ for the electron-channel validation sample for signal and background. The regions and the sample event counts in these regions are given in tables 9.1 and 9.2, respectively. The details of the optimization of these regions will be described in section 9.3.

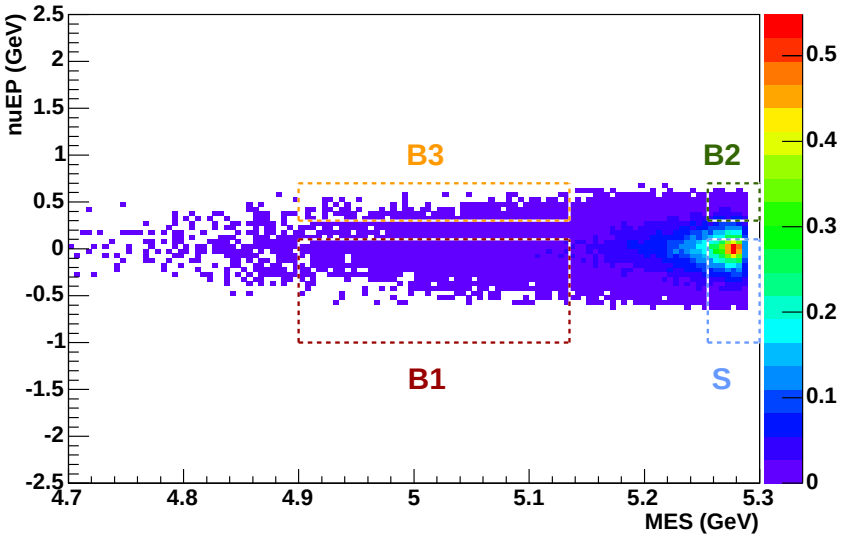


(a) Signal distribution.

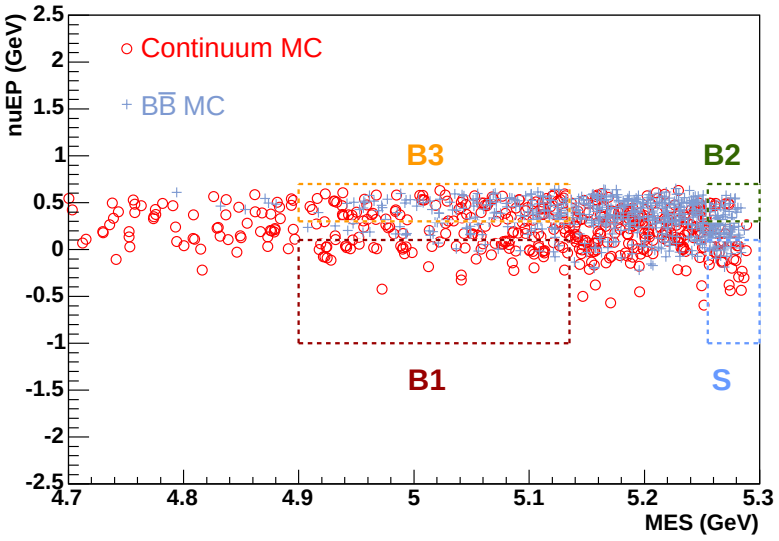


(b) Background distributions.

Figure 9.1. Electron-channel nuEP vs. m_{ES} distributions after all cuts, validation sample.



(a) Signal distribution.



(b) Background distributions.

Figure 9.2. Muon-channel nuEP vs. m_{ES} distributions after all cuts, validation sample.

Table 9.1. Region definitions for the fit, electron channel.

Region	Boundaries	
	nuEP (GeV)	$m_{\rm ES}$ (GeV)
S	$[-1, 0.1]$	$[5.25, 5.3]$
B1	$[-1, 0.1]$	$[4.9, 5.125]$
B2	$[0.25, 0.7]$	$[5.25, 5.3]$
B3	$[0.25, 0.7]$	$[4.9, 5.125]$

Likewise, for the muon channel, figure 9.2 shows the distributions of nuEP vs. $m_{\rm ES}$ for the validation sample for signal and background. The regions and the sample event counts in these regions are given in tables 9.3 and 9.4, respectively.

Because $b \to u\ell\nu_\ell$ decays are expected to contribute a significant fraction of the $B\overline{B}$ background, and because the branching fractions for these decays have different associated uncertainties, we treat seven exclusive $b \to u\ell\nu_\ell$ decay modes separately from the remaining "generic B" decays.[1] This gives us a total of ten component sample groups.

Conceptually, the signal extraction is based on fitting the total event counts in each of the four nuEP–$m_{\rm ES}$ regions with parameters describing the contributions of these ten groups. The branching fraction for signal and the magnitude of the generic B background are always allowed to vary in the fit.[2] The contribution from continuum background is fixed in magnitude because it is determined from off-peak data. The branching fractions for the seven semileptonic modes are determined from three independent measurements, of $\mathcal{B}(B^0 \to \pi^-\ell^+\nu_\ell)$, $\mathcal{B}(B^0 \to \rho^-\ell^+\nu_\ell)$, and $\mathcal{B}(B^+ \to \eta\ell^+\nu_\ell)$;

[1] In the rest of this thesis, "generic B" will refer to the generic B sample with the seven semileptonic samples removed.

[2] We allow the generic B background magnitude to float in the fit in order to absorb potential scale differences between data and MC.

Table 9.2. Region event counts, electron channel, validation sample, scaled to on-peak luminosities. For any specific modes, the assumed branching fractions are given in table 5.5.

	Counts			
Sample	S	$B1$	$B2$	$B3$
Signal MC (BF $= 3 \times 10^{-6}$)	21.19	3.74	3.15	0.94
Total continuum MC	14.12	67.76	10.51	117.41
Total $B\overline{B}$ MC	67.92	56.13	136.98	354.43
Generic B MC (incl. X_u)	22.42	40.22	85.68	317.35
Generic B MC (w/o X_u)	9.38	29.84	49.76	258.55
$B \to X_u \ell \nu_\ell$ MC	13.04	10.37	35.92	58.79
$B^0 \to \pi^- \ell^+ \nu_\ell$ MC	1.47	0.64	0.31	2.14
$B^0 \to \rho^- \ell^+ \nu_\ell$ MC	6.53	7.10	13.21	19.37
$B^+ \to \eta \ell^+ \nu_\ell$ MC	12.55	4.63	19.28	5.94
$B^+ \to \eta' \ell^+ \nu_\ell$ MC	0.36	0.05	0.47	1.61
$B^+ \to \omega \ell^+ \nu_\ell$ MC	0.73	0.16	3.45	2.56
$B^+ \to \pi^0 \ell^+ \nu_\ell$ MC	23.16	2.78	12.58	3.69
$B^+ \to \rho^0 \ell^+ \nu_\ell$ MC	0.70	0.56	2.00	1.76

the other four branching fractions are determined using isospin and SU(3) flavor symmetry, as described in section 5.2.3. The uncertainties on the measurements are taken into account in the fit by allowing $\mathcal{B}(B^0 \to \pi^- \ell^+ \nu_\ell)$, $\mathcal{B}(B^0 \to \rho^- \ell^+ \nu_\ell)$, and $\mathcal{B}(B^+ \to \eta \ell^+ \nu_\ell)$ to vary within their uncertainties, which we take to be their combined statistical and experimental systematic uncertainties. The theoretical uncertainties on these measurements are treated as systematic uncertainties in this analysis.

This fit requires knowledge of the relative contributions to each of the four nuEP–$m_{\rm ES}$ regions from the different component sample groups; we refer to these as the "templates" for the sample groups. The templates for signal, generic B, and the seven exclusive $b \to u \ell \nu_\ell$ modes are obtained from MC simulation. The continuum template is obtained from off-peak data.

Table 9.3. Region definitions for the fit, muon channel.

Region	Boundaries	
	nuEP (GeV)	$m_{\rm ES}$ (GeV)
S	$[-1, 0.1]$	$[5.255, 5.3]$
B1	$[-1, 0.1]$	$[4.9, 5.135]$
B2	$[0.3, 0.7]$	$[5.255, 5.3]$
B3	$[0.3, 0.7]$	$[4.9, 5.135]$

Because of the limited sample sizes used to determine the background templates, especially the continuum template derived from off-peak data, it is important to account for the effect of the statistical uncertainties on the templates in the fit. In particular, the large uncertainty on the off-peak template makes a major contribution to the final extracted signal branching-fraction uncertainty. We do this by allowing the templates to vary in the fit within their statistical uncertainties. This introduces an additional $10 \times 4 = 40$ measured parameters, bringing us to a total of 47 measured parameters.

The resulting fit minimizes a negative log-likelihood function. The full likelihood function is built up from the product of 3 gaussian probability density functions (PDFs) representing the distributions of the three measured SL BFs and 44 Poisson PDFs representing the event counts in the four $m_{\rm ES}$–nuEP regions in the ten templates and in the data.

With the template uncertainties included, the fit extracts the branching fraction of signal, the three semileptonic (SL) branching fractions, the B background normalization factor, and 40 scale factors reflecting the statistical uncertainties of the ten templates, giving us a total of 45 fitted parameters with two degrees of freedom in

Table 9.4. Region event counts, muon channel, validation sample, scaled to on-peak luminosities. For any specific modes, the assumed branching fractions are given in table 5.5.

	Counts			
Sample	S	$B1$	$B2$	$B3$
Signal MC (BF $= 3 \times 10^{-6}$)	13.12	3.27	0.96	0.48
Total continuum MC	36.00	82.51	10.36	131.08
Total $B\overline{B}$ MC	56.97	54.79	54.69	244.33
Generic B MC (incl. X_u)	21.69	40.20	32.59	221.81
Generic B MC (w/o X_u)	13.49	28.40	18.72	191.78
$B \to X_u \ell \nu_\ell$ MC	8.21	11.80	13.88	30.03
$B^0 \to \pi^- \ell^+ \nu_\ell$ MC	1.48	1.84	0.71	2.09
$B^0 \to \rho^- \ell^+ \nu_\ell$ MC	6.38	3.30	7.13	12.23
$B^+ \to \eta \ell^+ \nu_\ell$ MC	12.87	4.30	8.81	4.35
$B^+ \to \eta' \ell^+ \nu_\ell$ MC	0.10	0.08	0.31	0.94
$B^+ \to \omega \ell^+ \nu_\ell$ MC	1.18	0.25	0.41	0.60
$B^+ \to \pi^0 \ell^+ \nu_\ell$ MC	12.92	3.83	4.24	1.84
$B^+ \to \rho^0 \ell^+ \nu_\ell$ MC	0.34	0.99	0.50	0.46

our fit.[3]

9.1 Fit Details

We now describe the details of the likelihood fit. For each of the four nuEP–$m_{\rm ES}$ regions, "S," "B1," "B2," and "B3," we must determine the predicted event counts of each of the component sample groups. We have more than ten different samples, so each one of these samples, which we will call a "subsample," belongs to a sample group, as summarized in table 9.5. We combine the charged and neutral generic B MC into a single sample group, and in our toy MC studies, we combine the continuum MC subsamples into a continuum sample group.

[3]If we were to choose to let the continuum normalization factor float as well (which we do not), we would have 46 fitted parameters with one degree of freedom.

Table 9.5. Component sample groupings used for the BF extraction fit. For the continuum group, the MC samples are used for toy MC studies, and the off-peak sample is used for the final fit. Also, the radiative dimuon MC sample is only used in the muon analysis. Also, recall that the generic B samples have the events from the seven specific semileptonic decay modes removed.

Sample Group	Subsample
Signal	$B^+ \to \ell^+ \nu_\ell \gamma$ MC
Generic B	Generic B^+B^- MC
	Generic $B^0\overline{B}^0$ MC
Continuum (for studies)	$c\bar{c}$ MC
	uds MC
	$\tau^+\tau^-$ MC
	$\mu^+\mu^-\gamma$ MC
Continuum (for fit)	Off-peak data
SL1	$B^0 \to \pi^-\ell^+\nu_\ell$ MC
SL2	$B^0 \to \rho^-\ell^+\nu_\ell$ MC
SL3	$B^+ \to \eta\ell^+\nu_\ell$ MC
SL4	$B^+ \to \eta'\ell^+\nu_\ell$ MC
SL5	$B^+ \to \omega\ell^+\nu_\ell$ MC
SL6	$B^+ \to \pi^0\ell^+\nu_\ell$ MC
SL7	$B^+ \to \rho^0\ell^+\nu_\ell$ MC

For a given subsample i (generic B^+B^- MC, for instance), we define the scaled event count t_{ij}, essentially a predicted event count, for a particular subsample and region:

$$t_{ij} = \sum_k \alpha_{ik}\gamma_{ijk}m_{ijk}, \tag{9.1}$$

where j indexes the four regions, k indexes the four run periods, α_{ik} is the run period luminosity factor (described in section 5.4), γ_{ijk} is a q^2-weighting factor for each SL sample for a given run period and region (and takes a value of one for all other samples), and m_{ijk} is the unscaled event count and is based on a flat-q^2 model for the SL samples.

We then sum the scaled templates within each sample group to form the sample group templates: t_{Sj}, t_{Bj}, t_{Cj}, t_{1j}, t_{2j}, t_{3j}, t_{4j}, t_{5j}, t_{6j}, and t_{7j}, for signal, generic B background, continuum background, and the seven SL modes respectively.

We define $\hat{n}_j$ to be the predicted, or fitted, number of events in region j, to be compared to the observed event counts in the data, n_j. Setting aside for a moment the issue of allowing the templates themselves to vary in the fit within their statistical uncertainties, we compute the predicted event counts as follows:

$$\begin{aligned}\hat{n}_j &= \beta_S t_{Sj} + \beta_B t_{Bj} + \beta_C t_{Cj} + \\ &\quad \beta_1 t_{1j} + \beta_2 t_{2j} + \beta_3 t_{3j} + \beta_4 t_{4j} + \beta_5 t_{5j} + \beta_6 t_{6j} + \beta_7 t_{7j},\end{aligned}$$

where we have introduced a fitted normalization factor β_X for each sample group X. A normalization factor of one means that the scale prediction for a sample group was correct within fit uncertainties. Because we determine the continuum contribution from off-peak data, we fix $\beta_C = 1$ in the fit.

For the signal and the seven SL decay modes, it is more convenient to express the fit in terms of a branching fraction instead of a normalization factor. In addition, as explained above, we derive the seven SL branching fractions from only three independent measurements. We therefore rewrite the event count prediction as follows:

$$\begin{aligned}\hat{n}_j &= \kappa_S \mathcal{B}_S t_{Sj} + \beta_B t_{Bj} + \beta_C t_{Cj} + \\ &\quad \kappa_1 \mathcal{B}_\pi t_{1j} + \kappa_2 \mathcal{B}_\rho t_{2j} + \kappa_3 \mathcal{B}_\eta t_{3j} + \\ &\quad \kappa_4 \mathcal{B}_\eta t_{4j} + \kappa_5 \mathcal{B}_\rho t_{5j} + \kappa_6 \mathcal{B}_\pi t_{6j} + \kappa_7 \mathcal{B}_\rho t_{7j},\end{aligned} \tag{9.2}$$

where $\mathcal{B}_S$ is the fitted signal branching fraction, and $\mathcal{B}_\pi$, $\mathcal{B}_\rho$, and $\mathcal{B}_\eta$ are the fitted values of $\mathcal{B}(B^0 \to \pi^-\ell^+\nu_\ell)$, $\mathcal{B}(B^0 \to \rho^-\ell^+\nu_\ell)$, and $\mathcal{B}(B^+ \to \eta\ell^+\nu_\ell)$, respectively. The constant factors κ_X, defined in table 9.6, take into account the normalizations of the sample group templates, as well as the isospin and SU(3) symmetry factors and the B^+–B^0 lifetime ratio, which relate the branching fractions of the SL modes. The κ_X factors are computed taking into account the assumed branching fractions when generating the final templates in the case of MC, which as mentioned earlier reflect the number of expected events at the on-peak luminosity.

Table 9.6. The κ factors used in the fit. These are based on the branching fractions given in table 5.5. Because the SL samples include both the electron and muon channels, we double the BFs given in table 5.5. The value of 0.4861 for $B^+ \to \eta'\ell^+\nu_\ell$ is a factor obtained from theory which relates the η and η' BFs, as described in section 5.2.3. The value of 1.081 for the ω, π^0, and ρ^0 modes is a lifetime ratio between between charged and neutral B mesons.

Sample Group	Factor
$B^+ \to \ell^+\nu_\ell\gamma$ MC	$\kappa_S = 1/(3 \times 10^{-6})$
$B^0 \to \pi^-\ell^+\nu_\ell$ MC	$\kappa_1 = 1/(2.76 \times 10^{-4})$
$B^0 \to \rho^-\ell^+\nu_\ell$ MC	$\kappa_2 = 1/(4.28 \times 10^{-4})$
$B^+ \to \eta\ell^+\nu_\ell$ MC	$\kappa_3 = 1/(1.6 \times 10^{-4})$
$B^+ \to \eta'\ell^+\nu_\ell$ MC	$\kappa_4 = 1/(7.78 \times 10^{-5} * 0.4861)$
$B^+ \to \omega\ell^+\nu_\ell$ MC	$\kappa_5 = 1/(1.98 \times 10^{-4} * 2 * 1.081)$
$B^+ \to \pi^0\ell^+\nu_\ell$ MC	$\kappa_6 = 1/(1.28 \times 10^{-4} * 2 * 1.081)$
$B^+ \to \rho^0\ell^+\nu_\ell$ MC	$\kappa_7 = 1/(1.98 \times 10^{-4} * 2 * 1.081)$

We can now define a fit with seven measured parameters and five (six) fitted parameters. The seven measured parameters are the four measured event counts and three SL branching fractions. The five (six) fitted parameters are the signal branching fraction, the three SL branching fractions, and the generic B (and continuum)

normalization factor(s). We write the likelihood for this fit as follows:

$$\mathcal{L} = \mathcal{G}(\mathcal{B}'_\pi, \mathcal{B}_\pi, \sigma'_\pi)\mathcal{G}(\mathcal{B}'_\rho, \mathcal{B}_\rho, \sigma'_\rho)\mathcal{G}(\mathcal{B}'_\eta, \mathcal{B}_\eta, \sigma'_\eta) \prod_j P(n_j, \hat{n}_j). \tag{9.3}$$

where j indexes the region, $\mathcal{G}(x, \mu, \sigma)$ is the value of the gaussian distribution with mean μ and sigma σ at x, $\mathcal{B}'_X$ (σ'_X) is the measured value (combined statistical and systematic uncertainty) of the BF of mode X as given in table 5.5, and $P(x, \mu)$ is the value of the Poisson distribution with mean μ at x.

As mentioned earlier, we want to account for the statistical uncertainties in our templates, particularly when the continuum template is derived from off-peak data. We therefore allow the template values for each sample group and region to vary, introducing 40 scale factors ζ_{Xj} that multiply the template values in the fit. Thus, equation (9.2) must be modified as follows:

$$\begin{aligned} \hat{n}_j \;=\; & \kappa_S \mathcal{B}_S t_{Sj} \zeta_{Sj} + \beta_B t_{Bj} \zeta_{Bj} + \beta_C t_{Cj} \zeta_{Cj} + \\ & \kappa_1 \mathcal{B}_\pi t_{1j} \zeta_{1j} + \kappa_2 \mathcal{B}_\rho t_{2j} \zeta_{2j} + \kappa_3 \mathcal{B}_\eta t_{3j} \zeta_{3j} + \\ & \kappa_4 \mathcal{B}_\eta t_{4j} \zeta_{4j} + \kappa_5 \mathcal{B}_\rho t_{5j} \zeta_{5j} + \kappa_6 \mathcal{B}_\pi t_{6j} \zeta_{6j} + \kappa_7 \mathcal{B}_\rho t_{7j} \zeta_{7j}. \end{aligned} \tag{9.4}$$

In order to account for the variation of these scale factors, we introduce additional terms into the likelihood function—an additional Poisson PDF to constrain each template scale factor. The Poisson distribution applies to the actual *unscaled* event counts; however, in the fit model, as given by equation (9.4), the scale factors ζ multiply the *scaled* total event counts, which are scaled sums across run periods and

sample groups.

In principle, this can only be accounted for by having a separate PDF for each run and sample group, but this would be combinatorially intractable. We make the simplifying assumption that we may treat as a unit the statistical fluctuations across run periods and across subsamples within a sample group. (In the final fit, the only nontrivial sample group is generic B, which combines charged and neutral generic B MC. In our toy MC tests, where we use the continuum MC sample group, it is composed of a combination of different continuum MC subsamples (table 9.5); in the final signal extraction off-peak data will be used instead).

This assumption is implemented by summing up the *unscaled* event counts within a sample group and across run periods:

$$m_{Xj} = \sum_{ik} m_{ijk}, \tag{9.5}$$

where j indexes the region, k indexes the four run periods, X refers to the sample group, i is an index set of subsamples which correspond to sample group X, and m_{ijk} is the unscaled event count introduced earlier in equation (9.1),

We can now introduce new, template-variation terms into the fit of the form $P(m_{Xj}, \hat{m}_{Xj})$, where $\hat{m}_{Xj} = m_{Xj}\zeta_{Xj}$ and j indexes the region. The full likelihood function, $\mathcal{L}$, is then:

$$\begin{aligned}\mathcal{L} &= \mathcal{G}(\mathcal{B}'_\pi, \mathcal{B}_\pi, \sigma'_\pi)\mathcal{G}(\mathcal{B}'_\rho, \mathcal{B}_\rho, \sigma'_\rho)\mathcal{G}(\mathcal{B}'_\eta, \mathcal{B}_\eta, \sigma'_\eta) \\ &\quad \cdot \prod_j P(n_j, \hat{n}_j)P(m_{Sj}, \hat{m}_{Sj})P(m_{Bj}, \hat{m}_{Bj})P(m_{Cj}, \hat{m}_{Cj})\end{aligned}$$

$$\cdot P(m_{1j}, \hat{m}_{1j})P(m_{2j}, \hat{m}_{2j})P(m_{3j}, \hat{m}_{3j})P(m_{4j}, \hat{m}_{4j})$$

$$\cdot P(m_{5j}, \hat{m}_{5j})P(m_{6j}, \hat{m}_{6j})P(m_{7j}, \hat{m}_{7j}). \quad (9.6)$$

To summarize, we potentially have 46 fitted parameters: the branching fraction of signal ($\mathcal{B}_S$), the three semileptonic branching fractions ($\mathcal{B}_\pi$,$\mathcal{B}_\rho$,$\mathcal{B}_\eta$), the normalization factors for generic B and continuum background (β_B, β_C), and the 40 scale factors (ζ_{Xj}) reflecting the statistical uncertainties of the ten templates. However, in our final fit to data in both the electron and muon channels, we will fix the normalization of the continuum background (β_C is fixed to one in the fit), and therefore have 45 fitted parameters. The 47 measured parameters are the four values of n_j, the three SL branching fractions ($\mathcal{B}'_\pi$,$\mathcal{B}'_\rho$,$\mathcal{B}'_\eta$), and the 40 values of m_{Xj}.

9.2 Toy MC Studies

In order to study and validate the fit procedure, we use toy MC studies. We generate sets of a large number of simulated experiments, varying the signal branching fraction between each set, perform the fit for each experiment, and then fit the output distributions of each set. The different signal branching fractions we tried are given in table 9.7.

To generate our events, we start with the unscaled 2D nuEP vs. $m_{\rm ES}$ template distributions for each subsample, broken down by run period. Because the fit accounts for fluctuations in the templates, we also need to vary the generation templates between different toy MC experiments to be consistent. By *generation templates*, we

Table 9.7. The different branching fractions used in the toy MC studies of the signal extraction fit.

Branching Fraction
0
1×10^{-6}
3×10^{-6}
5×10^{-6}
1×10^{-5}
5×10^{-5}
1×10^{-4}

mean the templates that are used in the process of generating the final event counts in each of the four regions. The *fit* templates, however, as given by m_{Xj} in section 9.1, remain constant from experiment to experiment.

Even though the fit only accounts for fluctuations in the average sample *group* templates with the four different run periods combined, and not the *subsample* templates separated by run period, we choose to fluctuate the unscaled subsample templates separated by run, using Poisson statistics, and then combine them using the appropriate run period and training/validation split factors to form scaled generation templates, in the cases of the non-SL templates. (We can evaluate this assumption in our final toy MC results by examining the pull distributions of the fitted parameters).

In the case of the continuum fit component, we use the continuum MC because the off-peak data are blinded. However, we are interested in the behavior of our fit when the continuum template has large, off-peak-sized statistical uncertainties, so we use a modified procedure to give the continuum MC template off-peak-sized uncertainties. First, for each run period, we have an off-peak luminosity factor, as given by the ratio of on-peak luminosity to off-peak luminosity. Now, for each

subsample of the continuum MC group, we have a scaled template for each run period, which is the result of taking the unscaled template and scaling it by its luminosity and training/validation split factors. We then create new unscaled templates for each subsample and run period by dividing the scaled template by the off-peak luminosity factor. We use these new unscaled templates as the basis for fitting and generation templates for continuum MC.

For the SL templates, we randomize the three branching fractions by dividing the measured BF by the measured combined statistical and systematic errors to get a percentage uncertainty. We then generate three Gaussian random numbers with mean one and standard deviations given by the percentage uncertainties. These three random numbers are branching-fraction scale factors, and are applied to their appropriate seven SL modes. (The $B^0 \to \pi^- \ell^+ \nu_\ell$ BF scale factor is also applied to the $B^+ \to \pi^0 \ell^+ \nu_\ell$ generation template, the $B^0 \to \rho^- \ell^+ \nu_\ell$ factor is also applied to the $B^+ \to \rho^0 \ell^+ \nu_\ell$ and $B^+ \to \omega \ell^+ \nu_\ell$ templates, and the $B^+ \to \eta \ell^+ \nu_\ell$ factor is also applied to the $B^+ \to \eta' \ell^+ \nu_\ell$ template.) After the unscaled generation templates are fluctuated, as in the non-SL cases described in the previous paragraph, we apply the appropriate run period, training/validation split, q^2-weighting, and BF scale factors to obtain the scaled generation templates for the SL modes.

Once we have the scaled generation templates, we combine them into one total scaled generation template. We then allow the event counts in each of the four regions of this template to vary according to Poisson statistics. This gives us the final event counts in each of the four regions for a single experiment. We then fit these event

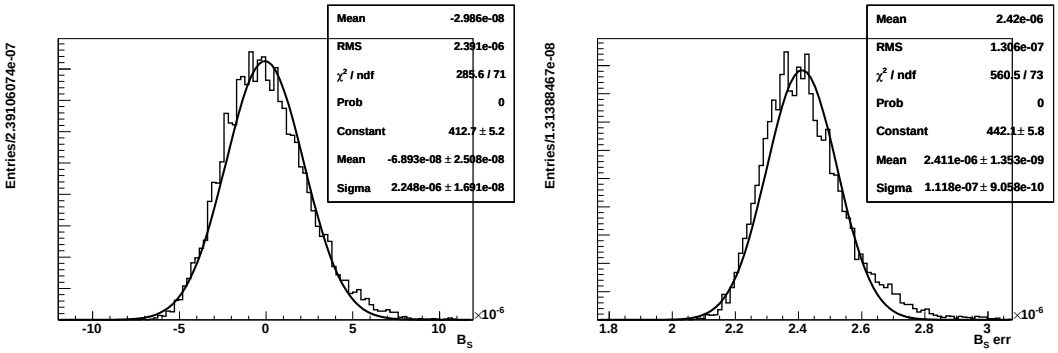
Table 9.8. Electron-channel results of 10,000 toy MC experiments, using different signal BFs. The mean and RMS in the table are the mean and RMS of the distribution of the fit-obtained signal BF.

Branching Fraction	Mean	RMS
0	-3.00×10^{-8}	2.39×10^{-6}
1×10^{-6}	1.00×10^{-6}	2.43×10^{-6}
3×10^{-6}	2.95×10^{-6}	2.50×10^{-6}
5×10^{-6}	4.93×10^{-6}	2.55×10^{-6}
1×10^{-5}	9.93×10^{-6}	2.70×10^{-6}
5×10^{-5}	4.99×10^{-5}	3.61×10^{-6}
1×10^{-4}	9.98×10^{-5}	4.50×10^{-6}

counts as we would after unblinding (except that we use continuum MC instead of off-peak data for our continuum template).

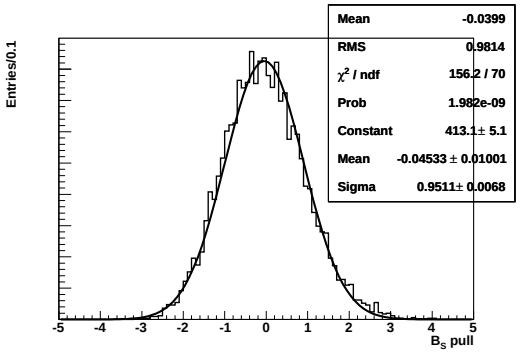
Using the above procedure to generate experiments, we had 10,000 experiments per set. We used the training sample MC template distributions after applying all of the cuts obtained at the end of the optimization detailed in chapter 8, with the exception of the nuEP and $m_{\rm ES}$ cuts.

In figure 9.3, we provide the toy MC results for the electron channel, assuming a signal BF of 0. In these figures we show the fitted value, the fitted error, and the pull of the fitted parameter, for the signal BF. We summarize the results for different BFs in table 9.8.



(a) Fitted value of $\mathcal{B}_S$.

(b) Fitted error of $\mathcal{B}_S$.



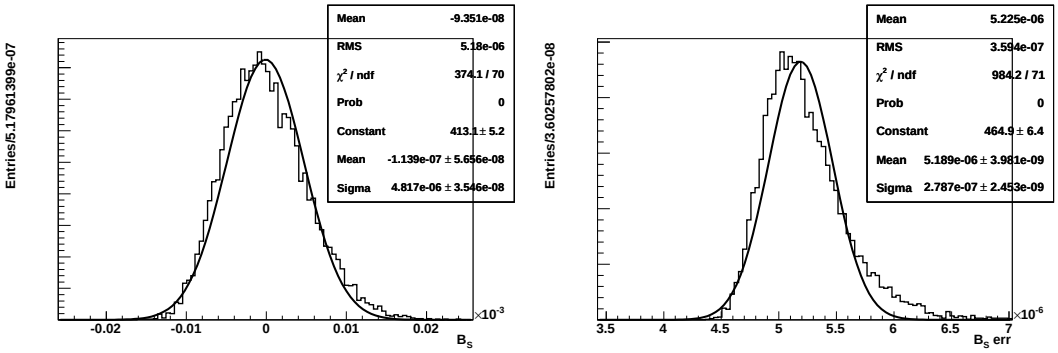
(c) Pull of fitted $\mathcal{B}_S$.

Figure 9.3. The $\mathcal{B}_S$ results from 10,000 toy MC experiments for the electron channel, assuming a signal BF of 0. Provided are the fitted value, the fitted error, and the pull of the fitted parameter. All distributions are also fitted to gaussians, with the gaussian fit output given in the last 5 lines of the statistics box.

Table 9.9. Muon-channel results of 10,000 toy MC experiments, using different signal BFs. The mean and RMS in the table are the mean and RMS of the distribution of the fit-obtained signal BF.

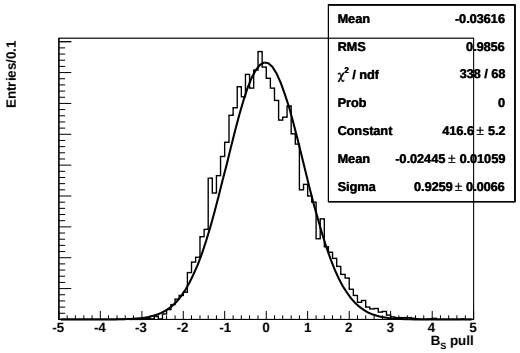
Branching Fraction	Mean	RMS
0	-9.35×10^{-8}	5.18×10^{-6}
1×10^{-6}	8.10×10^{-7}	5.21×10^{-6}
3×10^{-6}	2.76×10^{-6}	5.22×10^{-6}
5×10^{-6}	4.80×10^{-6}	5.24×10^{-6}
1×10^{-5}	9.78×10^{-6}	5.35×10^{-6}
5×10^{-5}	4.97×10^{-5}	6.17×10^{-6}
1×10^{-4}	9.97×10^{-5}	6.95×10^{-6}

In figure 9.4, we provide the results for the signal BF from 10,000 toy MC experiments for the electron channel, assuming a signal BF of 0. We also summarize the results for different BFs in table 9.9.



(a) Fitted value of $\mathcal{B}_S$.

(b) Fitted error of $\mathcal{B}_S$.



(c) Pull of fitted $\mathcal{B}_S$.

Figure 9.4. The $\mathcal{B}_S$ results from 10,000 toy MC experiments for the muon channel, assuming a signal BF of 0. Provided are the fitted value, the fitted error, and the pull of the fitted parameter. All distributions are also fitted to gaussians, with the gaussian fit output given in the last 5 lines of the statistics box.

9.3 Toy MC Region Optimization

We also use toy MC simulations to optimize the signal and sideband region boundaries of $m_{\rm ES}$ and nuEP. This involves varying the boundaries and running a set of toy MC experiments for each boundary. The boundary corresponding to the set with the lowest fitted mean of the signal branching-fraction variance distribution is taken as our optimized boundary.

We optimized our boundaries using the training sample MC as the source for our generating and fit template distributions. For each boundary, we generated a set of 10,000 experiments. In generating each experiment, we assumed a branching fraction for signal of 3×10^{-6}.[4] Continuum MC was used to model the continuum background, but was given off-peak statistics.

For both $m_{\rm ES}$ and nuEP, there are two signal region boundaries and two sideband region boundaries. For both variables, we fix the "outer" signal and sideband boundaries to reasonable values and vary only the "inner" boundaries.

A brute force optimization method would be to try every possible combination of the inner $m_{\rm ES}$ and nuEP boundaries. Given the very large number of possible combinations, we decided instead to perform an iterative optimization.

In the first step, we fixed the $m_{\rm ES}$ boundaries and optimized the nuEP boundaries, and vice versa, with the fixed boundaries chosen to be reasonable. The second step was a repetition of the first step, except the fixed values of $m_{\rm ES}$ and nuEP were obtained from the first step. The third step was a repetition of the second step, except

[4] We found no difference in the optimized boundaries when assuming a signal BF of either 0 or 3×10^{-6}.

the fixed values of $m_{\rm ES}$ and nuEP were obtained from the second step. After the third step, we examined the two FOM values resulting from the new $m_{\rm ES}$ and nuEP boundaries and chose to apply the new boundaries (either $m_{\rm ES}$ or nuEP) corresponding to the better variance.

The final results of this optimization are summarized in tables 9.1 and 9.3 for the electron and muon channels, respectively.

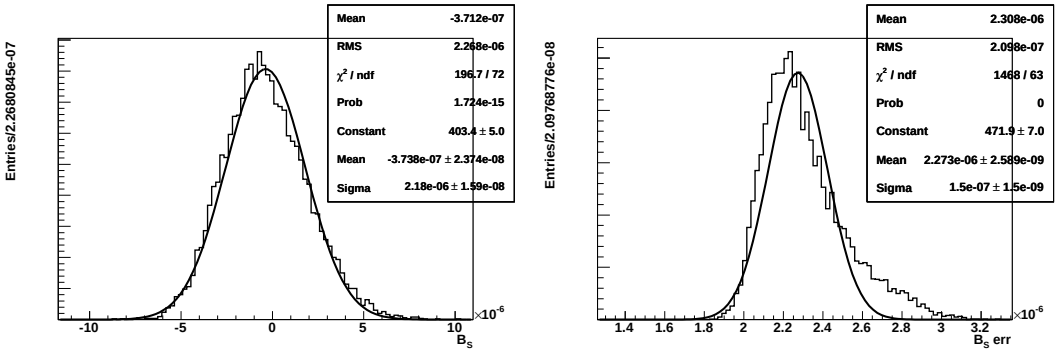
The marginal cut variable distributions and signal efficiency tables in the optimized signal region of $m_{\rm ES}$ and nuEP, with all other cuts optimized as described in chapter 8, can be seen in appendices A and B.

9.4 Joint Fit Study

We also extract a BF result based on a joint fit to the electron and muon channels. This is done by simultaneously performing the electron- and muon-channel fits, constraining the fitted signal BF and the three fitted SL BFs to be the same between channels. The normalization factor for generic B background is allowed to float separately for the electron and muon channels. In particular, this allows us to account for unmodeled B background with different efficiencies in the electron and muon channels.

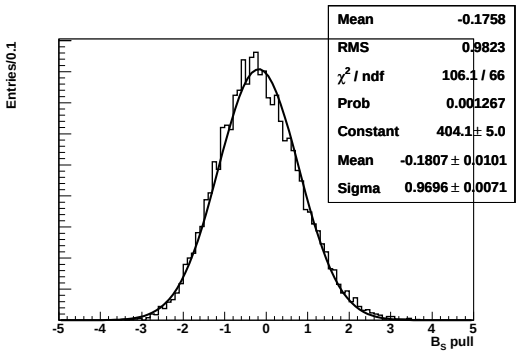
Before unblinding both channels, we first studied this fit with toy MC studies using only MC for the templates. Examination of the events comprising the templates showed that the channels have no events in common, so we can generate events independently between the two channels.

We evaluated the results of a joint fit, using 10,000 toy MC experiments, with a signal BF of 0. The RMS of the resulting signal BF distribution for the joint fit is 2.27×10^{-6}, as can be seen in figure 9.5(a).



(a) Fitted value of $\mathcal{B}_S$.

(b) Fitted error of $\mathcal{B}_S$.



(c) Pull of fitted $\mathcal{B}_S$.

Figure 9.5. The $\mathcal{B}_S$ results from 10,000 toy MC experiments for the joint electron- and muon-channel fit, assuming a signal BF of 0. Provided are the fitted value, the fitted error, and the pull of the fitted parameter. All distributions are also fitted to gaussians, with the gaussian fit output given in the last 5 lines of the statistics box.

This may be compared with the RMS values of the signal BF distributions for the electron- and muon-channel fits: 2.39×10^{-6} and 5.18×10^{-6}, respectively (see tables 9.8 and 9.9), and with their weighted average of 2.17×10^{-6}.[5] We speculate that the discrepancy between the naive average and the result of the joint fit may be due to correlated contributions to the two fits, such as from the exclusive $b \to u\ell\nu_\ell$ measurements.

[5]We use 0 as the central value, assigning a weight of $1/\sigma_i^2$, where σ_i is the RMS and i is an index referring to either the electron or muon channel.

Chapter 10

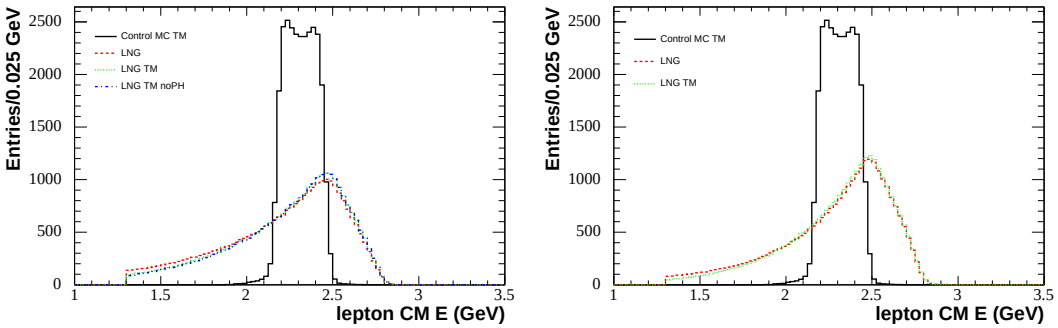
Control Sample Studies

In order to check the validity of the most important aspects of our analysis technique and perform systematic studies, we study the $B^+ \to \pi^+ \overline{D}^0 (\to K^+\pi^-)$ control sample. The $B^+ \to \pi^+ \overline{D}^0 (\to K^+\pi^-)$ control sample (abbreviated as $\overline{D}^0\pi^+$) is useful for studying our neutrino reconstruction and recoil-based selection variables. As mentioned in section 5.3, we choose this sample because there are three decay products, with one decay product (the hard pion) of high energy, and the sample is clean, with a relatively large branching fraction. The hard pion is treated as the signal lepton, the kaon as the signal photon, and the soft pion as the signal neutrino.

In this chapter, we will begin by comparing MC cut variable distributions for $\overline{D}^0\pi^+$ vs. $B^+ \to \ell^+\nu_\ell\gamma$. Then, we will use this control sample to compare cut variable distributions in data and MC. We will need to perform a background subtraction on the data, replicated in the MC, because the data sample is not clean enough due to the presence of background in the signal region (tight cuts around $\Delta E_{\rm con}$, $m_{\rm ES,con}$, and m_{D^0}).

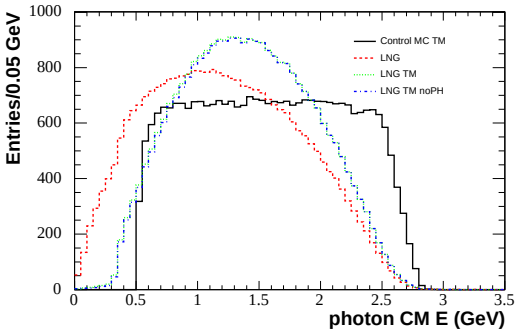
10.1 MC Comparison of $\overline{D}^0\pi^+$ vs. $B^+ \to \ell^+\nu_\ell\gamma$

First, we compare the distributions of $\overline{D}^0\pi^+$ MC vs. $B^+ \to \ell^+\nu_\ell\gamma$ MC. We expect differences in some of the distributions due to different physics. Figures 10.1–10.4 show some comparisons of different distributions for the control sample MC vs. the electron and muon $B^+ \to \ell^+\nu_\ell\gamma$ channels.

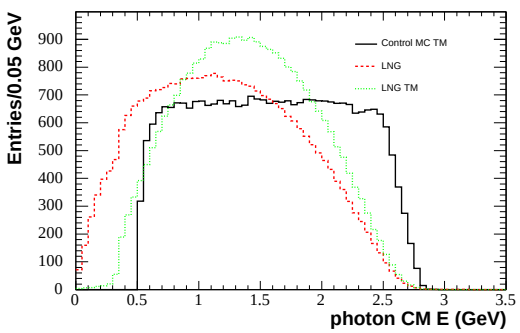


(a) Signal lepton CM energy, electron channel.

(b) Signal lepton CM energy, muon channel.

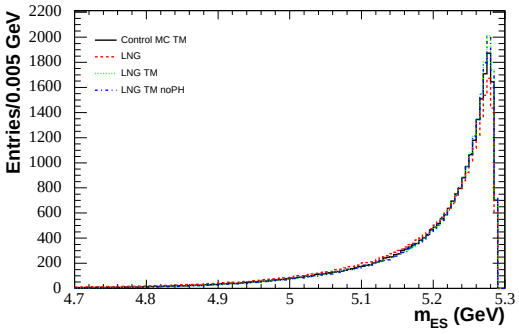


(c) Signal photon CM energy, electron channel.

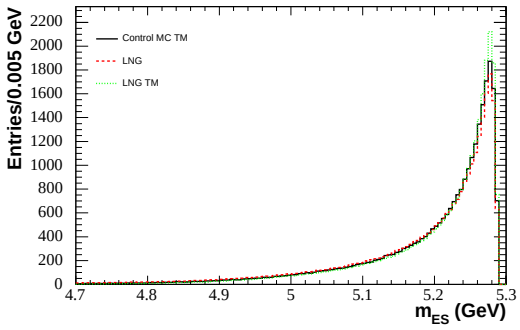


(d) Signal photon CM energy, muon channel.

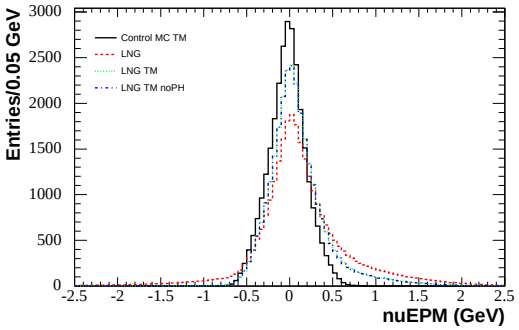
Figure 10.1. Comparison of signal lepton CM energy and signal photon CM energy in $\overline{D}^0\pi^+$ MC ("Control MC") vs. $B^+ \to \ell^+\nu_\ell\gamma$ ("LNG") MC for both the electron and muon channels. "TM" stands for truth-matched. "No PH" stands for "no PHOTOS." Scaling is by total integral, except in the "No PH" histogram, where the scale factor is the same as for the "TM" histogram.



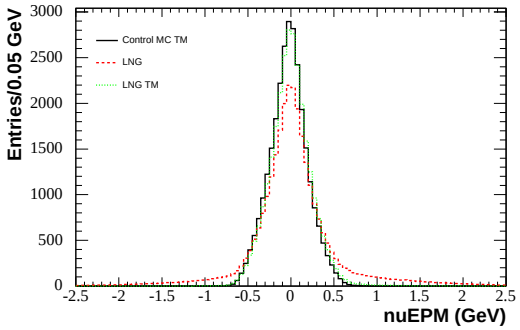
(a) $m_{\rm ES}$ (Recoil), electron channel.



(b) $m_{\rm ES}$ (Recoil), muon channel.

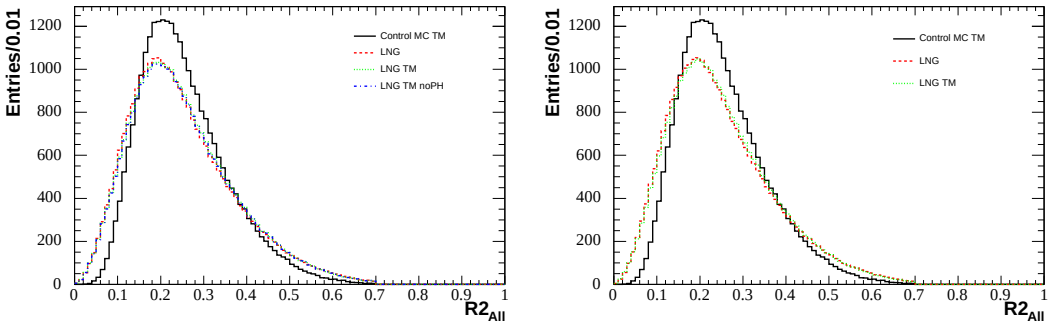


(c) nuEPM, electron channel.



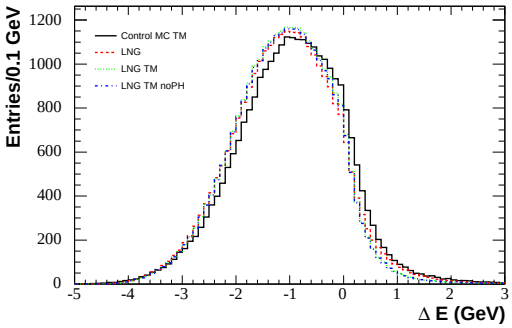
(d) nuEPM, muon channel.

Figure 10.2. Comparison of $m_{\rm ES}$ (Recoil) and nuEPM in $\overline{D}^0\pi^+$ MC ("Control MC") vs. $B^+ \to \ell^+\nu_\ell\gamma$ ("LNG") MC for both the electron and muon channels. "TM" stands for truth-matched. "No PH" stands for "no PHOTOS." Scaling is by total integral, except in the "No PH" histogram, where the scale factor is the same as for the "TM" histogram.

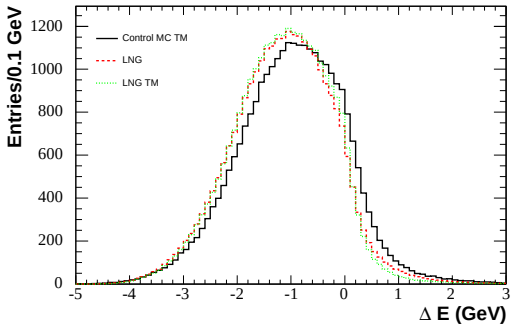


(a) $R2_{\mathrm{All}}$, electron channel.

(b) $R2_{\mathrm{All}}$, muon channel.

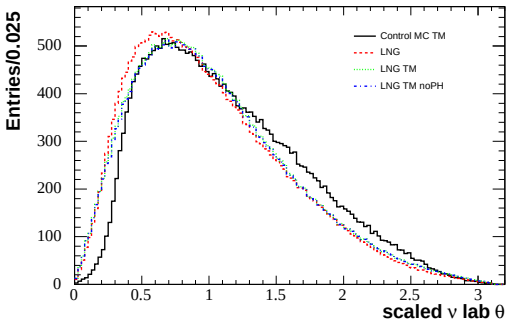


(c) ΔE, electron channel.

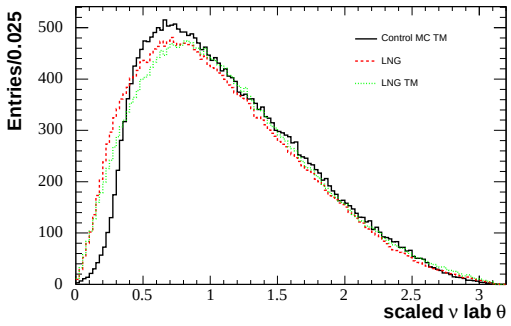


(d) ΔE, muon channel.

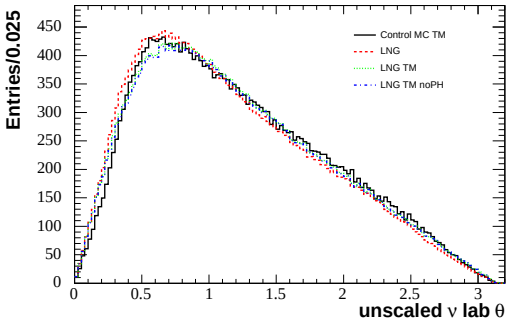
Figure 10.3. Comparison of $R2_{\mathrm{All}}$ and recoil ΔE in $\overline{D}^0\pi^+$ MC ("Control MC") vs. $B^+ \to \ell^+\nu_\ell\gamma$ ("LNG") MC for both the electron and muon channels. "TM" stands for truth-matched. "No PH" stands for "no PHOTOS." Scaling is by total integral, except in the "No PH" histogram, where the scale factor is the same as for the "TM" histogram.



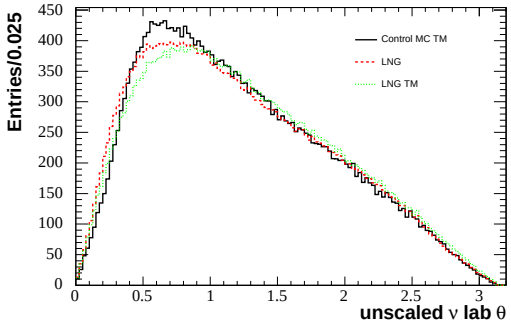
(a) Scaled neutrino lab θ, electron channel.



(b) Scaled neutrino lab θ, muon channel.



(c) Unscaled neutrino lab θ, electron channel.



(d) Unscaled neutrino lab θ, muon channel.

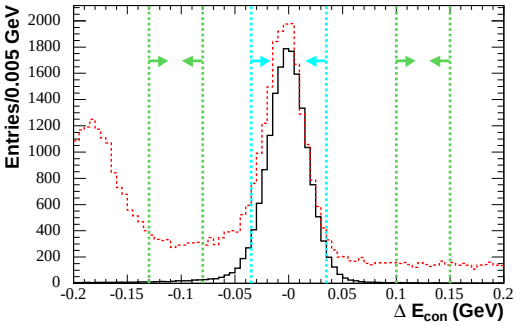
Figure 10.4. Comparison of scaled and unscaled neutrino lab θ in $\overline{D}^0\pi^+$ MC ("Control MC") vs. $B^+ \to \ell^+\nu_\ell\gamma$ ("LNG") MC for both the electron and muon channels. "TM" stands for truth-matched. "No PH" stands for "no PHOTOS." Scaling is by total integral, except in the "No PH" histogram, where the scale factor is the same as for the "TM" histogram.

10.2 Background Subtraction and Control Sample Data/MC Comparison

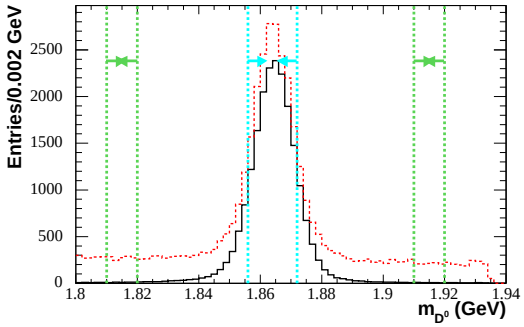
The overall goal here is to compare a clean sample of $\overline{D}^0\pi^+$ events in on-peak data and MC, with the MC distributions scaled to on-peak luminosities. The MC sample is clean, by composition, and only combinatoric background due to misreconstruction is possible. The data sample, however, requires a background subtraction in order to be cleaned up. For consistency, the same background subtraction procedure needs to be performed on MC.

We choose $\Delta E_{\rm con}$ (as opposed to $m_{\rm ES,con}$ or m_{D^0}) as our background subtraction variable, because it should have the least peaking background in the signal region, thus allowing us to perform more reliable fits to signal and background. We define the $\Delta E_{\rm con}$ signal region to be between -35 and 35 MeV, and the two $\Delta E_{\rm con}$ sideband regions to be between -150 and -100 MeV, and between 80 and 130 MeV. We also use an $m_{\rm ES,con}$ signal region of greater than 5.274 GeV and a m_{D^0} signal region of between 1.856 and 1.872 GeV.

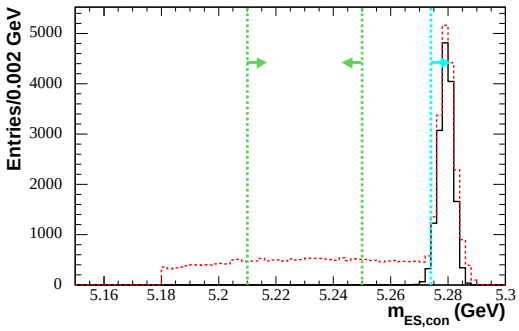
We accumulate histograms in both the triple ($\Delta E_{\rm con}$, $m_{\rm ES,con}$, m_{D^0}) signal region and the $\Delta E_{\rm con}$ sideband region (m_{D^0} and $m_{\rm ES,con}$ signal region) for our variables of interest. The marginal distributions of $\Delta E_{\rm con}$, m_{D^0}, and $m_{\rm ES,con}$ in the triple signal region can be seen in figure 10.5.



(a) $\Delta E_{\rm con}$ distributions.



(b) m_{D^0} distributions.



(c) $m_{\rm ES,con}$ distributions.

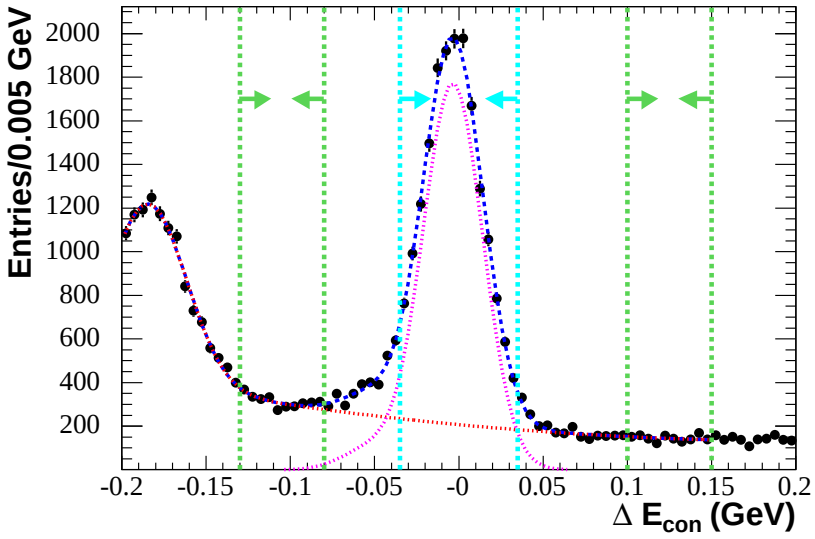
Figure 10.5. Marginal distributions of $\Delta E_{\rm con}$, m_{D^0}, and $m_{\rm ES,con}$ in the signal region of all three variables for the $\overline{D}^0\pi^+$ control sample. Black (solid) is MC. Red (dashed) is data. The aqua dashed lines define the signal region and the green dashed lines define the sideband regions. Event counts are based on on-peak luminosities.

For variable X, let us label its $\Delta E_{\rm con}$ signal and sideband region histograms $H_{\rm Sig}$ and $H_{\rm SB}$, respectively. Also, we assume that each of these histograms has two unknown component histograms: H^S and H^B, with the superscript S or B referring to the signal or background component, respectively. The goal of the background subtraction is to obtain the background-subtracted histogram $H^S_{\rm Sig} = H_{\rm Sig} - H^B_{\rm Sig}$, where the subtraction is performed bin-by-bin between the two histograms and the equality is a bin-by-bin equality.

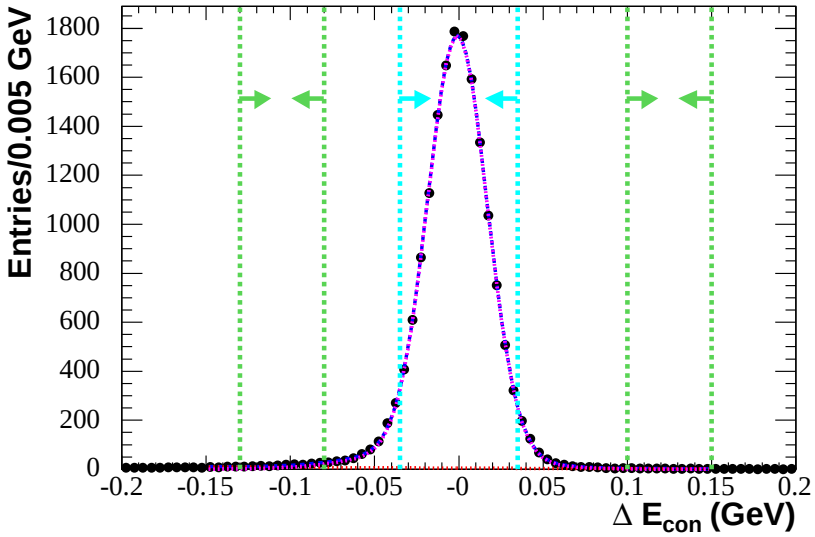
We next fit the $\Delta E_{\rm con}$ distribution for signal and sideband. In the on-peak data, we fit the signal to a double gaussian, and the background to the sum of a second-order polynomial and gaussian. In the MC, we fit the signal to a double gaussian and the background to a second-order polynomial. The results of the fits to on-peak data and MC are given in tables 10.1 and 10.2, and figure 10.6.

By integrating the fit equations over the signal and sideband boundaries we can calculate the amount of signal and background in the signal and sideband regions, labeling these quantities $S_{\rm Sig}$, $B_{\rm Sig}$, $S_{\rm SB}$, and $B_{\rm SB}$, with the S,B referring to the number of signal or background events, and the subscripts Sig, SB referring to the signal and sideband region, respectively. In order to perform the background subtraction on the variable X distribution we need two factors, α and β, where α is the ratio of $B_{\rm Sig}$ to $B_{\rm SB}$, and β is the ratio of $S_{\rm Sig}$ to $S_{\rm SB}$.

We now make two assumptions: First, we assume that the shape of the background in variable X is approximately the same in the $\Delta E_{\rm con}$ signal and sideband regions. Second, we assume that the shape of the signal in variable X is approxi-



(a) Fit to on-peak data.



(b) Fit to MC.

Figure 10.6. Fits to $\Delta E_{\rm con}$ distributions in on-peak data and MC, $\overline{D}^0\pi^+$ control sample. The black points are the data. The blue dashed line is the complete fit to the data. The magenta and red dashed lines are the signal and background components of this fit, respectively. The aqua dashed lines define the signal region and the green dashed lines define the sideband regions.

Table 10.1. The fit results to $\Delta E_{\rm con}$ for on-peak data, $\overline{D}^0\pi^+$ signal control sample. The results correspond to the fit in figure 10.6(a).

Parameter	Value
χ^2/ndf	80.51/58
Probability	0.03
Signal gaussian 1 normalization	$1.77 \pm 0.02 \times 10^3$
Signal gaussian 1 mean	$-3.55 \pm 0.35 \times 10^{-3}$
Signal gaussian 1 sigma	$1.78 \pm 0.03 \times 10^{-2}$
Signal gaussian 2 normalization	$1.00 \pm 0.13 \times 10^2$
Signal gaussian 2 mean	$-5.14 \pm 0.48 \times 10^{-2}$
Signal gaussian 2 sigma	$1.73 \pm 0.40 \times 10^{-2}$
Background $p0$	$2.07 \pm 0.06 \times 10^2$
Background $p1$	$-7.24 \pm 0.41 \times 10^2$
Background $p2$	$1.75 \pm 0.51 \times 10^3$
Background gaussian normalization	$8.19 \pm 0.22 \times 10^2$
Background gaussian mean	$-1.84 \pm 0.01 \times 10^{-1}$
Background gaussian sigma	$2.24 \pm 0.10 \times 10^{-2}$

mately the same in the $\Delta E_{\rm con}$ signal and sideband regions. These assumptions lead to the following two histogram "equations":

$$\begin{aligned} H^B_{\rm Sig} &\approx \alpha H^B_{\rm SB} \\ H^S_{\rm Sig} &\approx \beta H^S_{\rm SB}, \end{aligned}$$

where the approximate equality is a bin-by-bin approximate equality, and the product of a scalar and a histogram is a bin-by-bin scaling of the histogram by the scalar factor. These equations imply that we can use the ratio of event counts in $\Delta E_{\rm con}$ signal vs. sideband to scale the signal and background shapes in variable X so that we can perform the background subtraction.

With these two assumptions, one can show that the background-subtracted his-

Table 10.2. The fit results to $\Delta E_{\rm con}$ for the $\overline{D}^0\pi^+$ signal control MC sample. The results correspond to the fit in figure 10.6(b).

Parameter	Value
χ^2/ndf	117.4/51
Probability	6.74×10^{-8}
Signal gaussian 1 normalization	$1.70 \pm 0.01 \times 10^{3}$
Signal gaussian 1 mean	$-7.95 \pm 5.80 \times 10^{-4}$
Signal gaussian 1 sigma	$1.76 \pm 0.01 \times 10^{-2}$
Signal gaussian 2 normalization	$7.45 \pm 5.07 \times 10^{1}$
Signal gaussian 2 mean	$-1.54 \pm 0.09 \times 10^{-2}$
Signal gaussian 2 sigma	$4.05 \pm 0.13 \times 10^{-2}$
Background $p0$	4.63 ± 0.94
Background $p1$	$-2.70 \pm 0.21 \times 10^{1}$
Background $p2$	$5.33 \pm 4.54 \times 10^{1}$

togram $H^S_{\rm Sig}$ for variable X can be obtained from its unsubtracted region histograms and the fitted $\Delta E_{\rm con}$ signal and sideband event counts as follows:

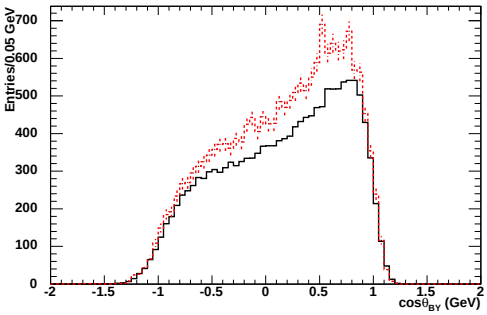
$$\begin{aligned} H^S_{\rm Sig} &\approx \frac{1}{1-\frac{\alpha}{\beta}}(H_{\rm Sig} - \alpha H_{\rm SB}) \\ &\approx \gamma(H_{\rm Sig} - \alpha H_{\rm SB}), \end{aligned}$$

where γ is used to label the overall correction factor due to the presence of signal events in the sideband region.

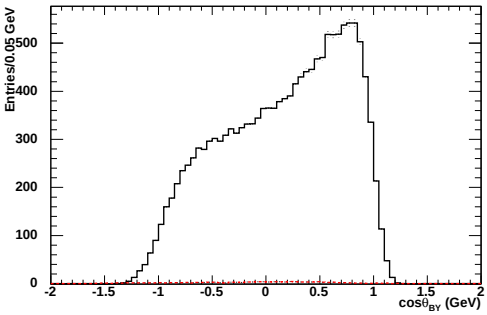
Table 10.3. Factors for $B^+ \to \pi^+\overline{D}^0(\to K^+\pi^-)$ control sample $\Delta E_{\rm con}$ background subtraction.

Factor	Value
α (Data)	0.64012
γ (Data)	1.00182
α (MC)	0.64071
γ (MC)	1.00355

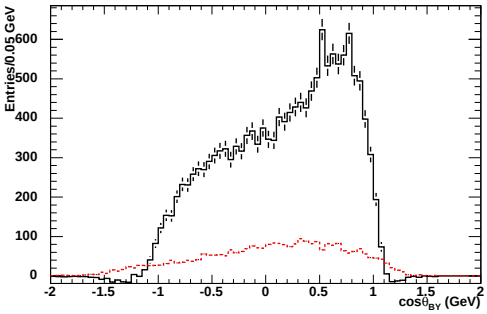
The calculated α and γ factors for data and MC are given in table 10.3. With these factors and the $\Delta E_{\rm con}$ signal and sideband region histograms for our variables of interest, we perform our background subtractions. An explicit set of plots before and after the background subtraction is shown in figure 10.7.



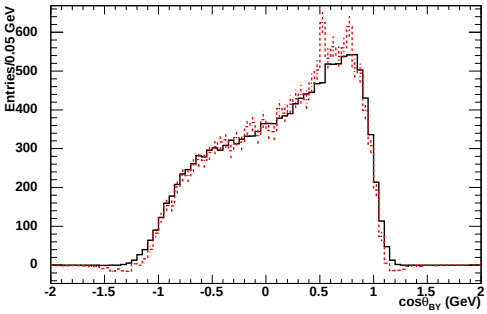
(a) MC (black, solid) and on-peak data (red, dashed) distributions in the triple $\Delta E_{\rm con}$, m_{D^0}, and $m_{\rm ES,con}$ signal region before background subtraction.



(b) Distribution (red, dashed) for MC in the $\Delta E_{\rm con}$ sideband, with the appropriate signal to sideband scaling factor applied, and distribution (black, solid) for MC after this sideband distribution was subtracted off.



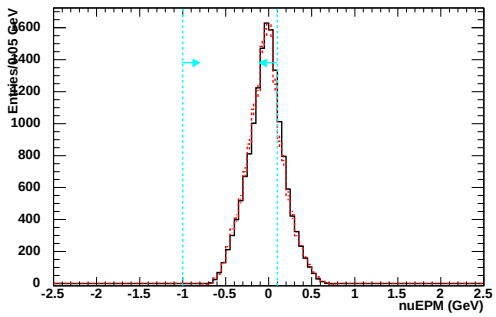
(c) Distribution (red, dashed) for on-peak data in the $\Delta E_{\rm con}$ sideband, with the appropriate signal to sideband scaling factor applied, and distribution (black, solid) for on-peak data after this sideband distribution was subtracted off.



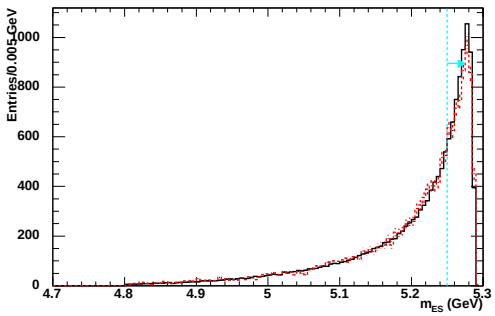
(d) MC (black, solid) and on-peak data (red, dashed) distributions after background subtraction.

Figure 10.7. Plots illustrating the $\overline{D}^0\pi^+$ control sample $\Delta E_{\rm con}$ background subtraction for the $\cos\theta_{BY}$ selection variable. Event counts are based on on-peak luminosities.

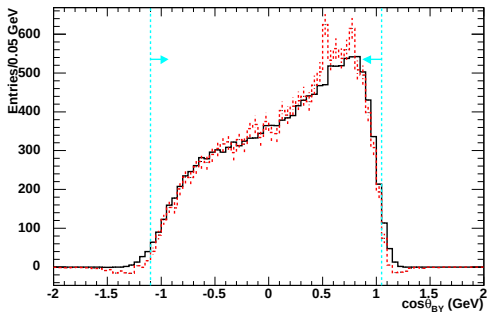
The results after background subtraction for our selection variables are shown in figures 10.8–10.11. The results are, in general, quite good, and do give us confidence in the validity of our analysis procedure.



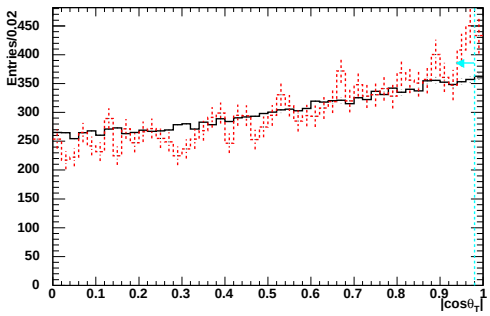
(a) nuEPM distributions.



(b) $m_{\rm ES}$ distributions.

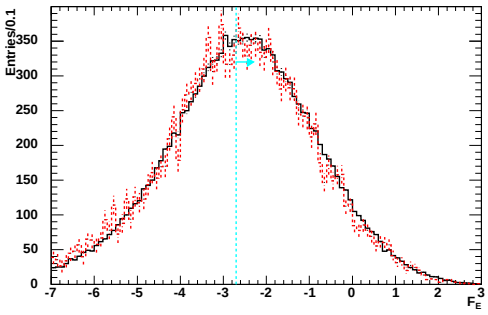


(c) $\cos\theta_{BY}$ distributions.

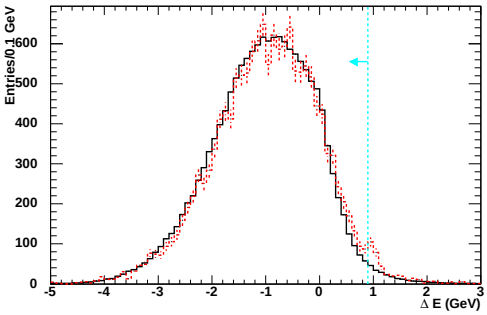


(d) $|\cos\theta_T|$ distributions.

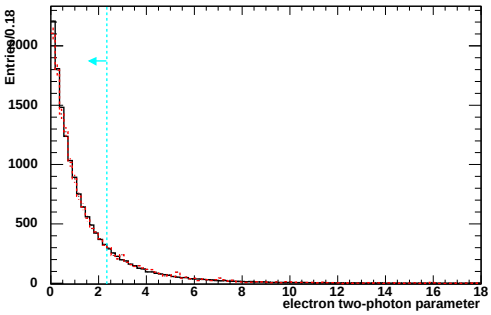
Figure 10.8. The nuEPM, $m_{\rm ES}$, $\cos\theta_{BY}$, and $|\cos\theta_T|$ $\overline{D}^0\pi^+$ control sample distributions after the $\Delta E_{\rm con}$ background subtraction, as described in section 10.2. Black (solid) is MC. Red (dashed) is on-peak data. The vertical aqua dashed lines are the cut locations in the final signal region of the electron analysis. Event counts are based on on-peak luminosities. More distributions follow in figures 10.9, 10.10, and 10.11.



(a) Electron channel $\mathcal{F}_{\rm E}$ distributions.

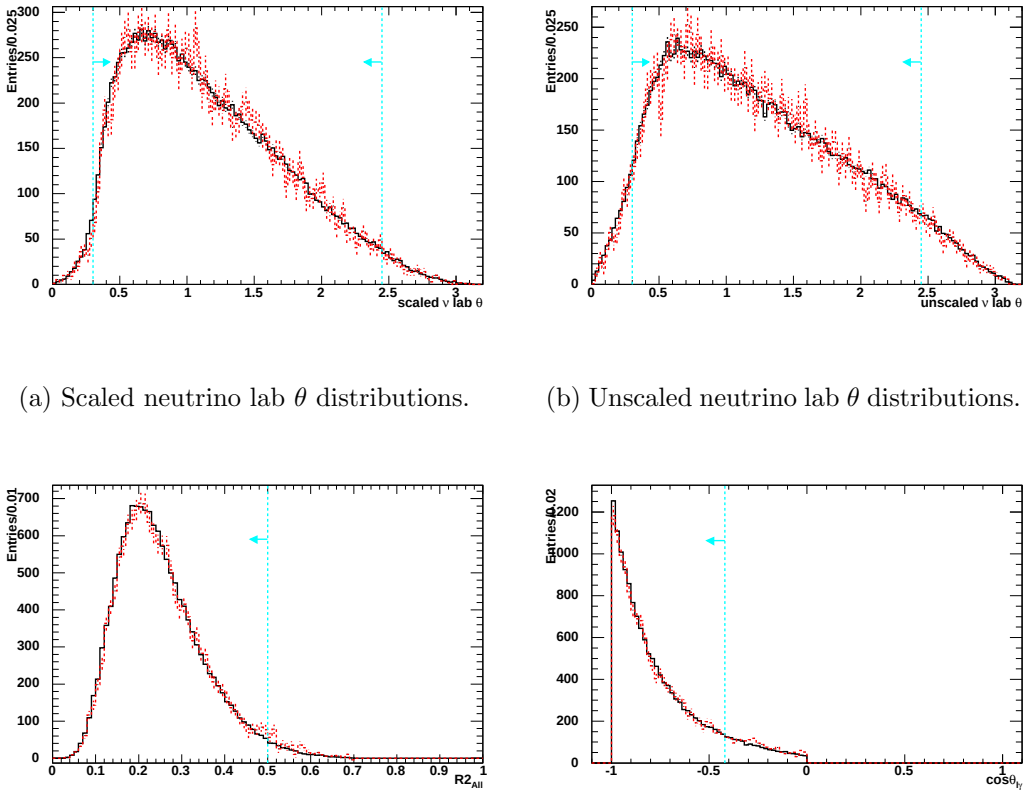


(b) ΔE distributions.



(c) Electron two-photon parameter distributions.

Figure 10.9. The electron channel $\mathcal{F}_{\rm E}$, ΔE, and electron two-photon parameter $\overline{D}^0\pi^+$ control sample distributions after the $\Delta E_{\rm con}$ background subtraction, as described in section 10.2. Black (solid) is MC. Red (dashed) is on-peak data. The vertical aqua dashed lines are the cut locations in the final signal region of the electron analysis. Event counts are based on on-peak luminosities. More distributions follow in figures 10.10 and 10.11.



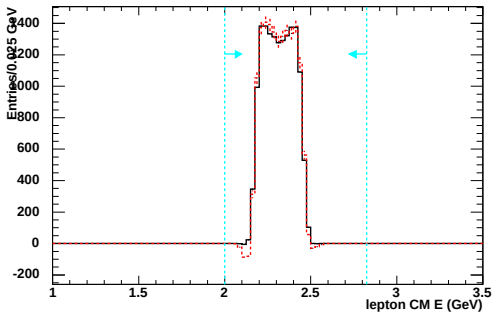
(a) Scaled neutrino lab θ distributions.

(b) Unscaled neutrino lab θ distributions.

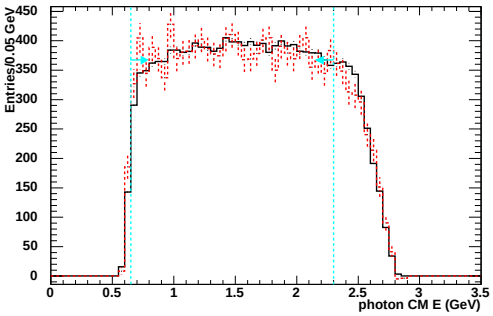
(c) $R2_{\mathrm{All}}$ distributions.

(d) $\cos\theta_{\ell\gamma}$ distributions.

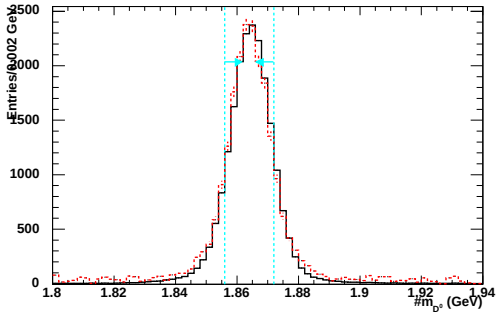
Figure 10.10. The scaled neutrino lab θ, unscaled neutrino lab θ, $R2_{\mathrm{All}}$, and $\cos\theta_{\ell\gamma}$ $\overline{D}^0\pi^+$ control sample distributions after the ΔE_{con} background subtraction, as described in section 10.2. Black (solid) is MC. Red (dashed) is on-peak data. The vertical aqua dashed lines are the cut locations in the final signal region of the electron analysis. Event counts are based on on-peak luminosities. More distributions follow in figure 10.11.



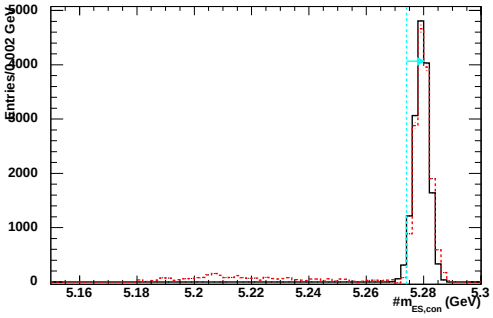
(a) Signal lepton CM energy distributions.



(b) Signal photon CM energy distributions.



(c) m_{D^0} distributions.



(d) $m_{\mathrm{ES,con}}$ distributions.

Figure 10.11. The signal lepton CM energy, signal photon CM energy, m_{D^0}, and $m_{\mathrm{ES,con}}$ $\overline{D}^0\pi^+$ control sample distributions after the ΔE_{con} background subtraction, as described in section 10.2. Black (solid) is MC. Red (dashed) is on-peak data. The vertical aqua dashed lines are the cut locations in the final signal region of the electron analysis. Event counts are based on on-peak luminosities.

Chapter 11

Fit to π^0-Veto Sideband

As a validation of our fit procedure, described earlier in chapter 9, we perform a fit to the on-peak data in the clpi0 sideband (defined earlier in section 6.1.4.2 to be between 124 and 146 MeV). The signal BF fit procedure described in chapter 9 takes advantage of isospin relations and fits for the charged and neutral π SL modes using a single parameter. For the fit in the clpi0 sideband, we modify this procedure and separately fit for the charged and neutral π modes by using a separate BF parameter for each.

To include separate BF parameters for the charged and neutral π modes, we modify equation (9.2):

$$\begin{aligned}\hat{n}_j &= \kappa_S \mathcal{B}_S t_{Sj} + \beta_B t_{Bj} + \beta_C t_{Cj} + \\ &\quad \kappa_1 \mathcal{B}_{\pi^+} t_{1j} + \kappa_2 \mathcal{B}_\rho t_{2j} + \kappa_3 \mathcal{B}_\eta t_{3j} + \\ &\quad \kappa_4 \mathcal{B}_\eta t_{4j} + \kappa_5 \mathcal{B}_\rho t_{5j} + \kappa'_6 \mathcal{B}_{\pi^0} t_{6j} + \kappa_7 \mathcal{B}_\rho t_{7j},\end{aligned}$$

where we now have $\mathcal{B}_{\pi^+}$ and $\mathcal{B}_{\pi^0}$ as the BF parameters for the π^+ and π^0 modes respectively, and κ'_6 is equal to $1/(1.49 \times 10^{-4})$ (since isospin factors are no longer

needed).

In addition, equation (9.6) can be modified in one of two ways. First, one can constrain the BF of the π^0 mode in the fit, with the resulting likelihood:

$$\begin{aligned}\mathcal{L} &= \mathcal{G}(\mathcal{B}'_{\pi^+}, \mathcal{B}_{\pi^+}, \sigma'_{\pi^+})\mathcal{G}(\mathcal{B}'_{\pi^0}, \mathcal{B}_{\pi^0}, \sigma'_{\pi^0})\mathcal{G}(\mathcal{B}'_{\rho}, \mathcal{B}_{\rho}, \sigma'_{\rho})\mathcal{G}(\mathcal{B}'_{\eta}, \mathcal{B}_{\eta}, \sigma'_{\eta}) \\ &\cdot \prod_j P(n_j, \hat{n}_j)P(m_{Sj}, \hat{m}_{Sj})P(m_{Bj}, \hat{m}_{Bj})P(m_{Cj}, \hat{m}_{Cj}) \\ &\cdot P(m_{1j}, \hat{m}_{1j})P(m_{2j}, \hat{m}_{2j})P(m_{3j}, \hat{m}_{3j})P(m_{4j}, \hat{m}_{4j}) \\ &\cdot P(m_{5j}, \hat{m}_{5j})P(m_{6j}, \hat{m}_{6j})P(m_{7j}, \hat{m}_{7j}),\end{aligned}$$

where once again $\mathcal{B}'_X$ (σ'_X) is the measured value (combined statistical and systematic uncertainty) of the BF of mode X as given in table 5.5, except now split up for the charged and neutral π modes.

However, we can also allow the π^0 mode BF to float completely in the fit, so that the resulting likelihood is:

$$\begin{aligned}\mathcal{L} &= \mathcal{G}(\mathcal{B}'_{\pi^+}, \mathcal{B}_{\pi^+}, \sigma'_{\pi^+})\mathcal{G}(\mathcal{B}'_{\rho}, \mathcal{B}_{\rho}, \sigma'_{\rho})\mathcal{G}(\mathcal{B}'_{\eta}, \mathcal{B}_{\eta}, \sigma'_{\eta}) \\ &\cdot \prod_j P(n_j, \hat{n}_j)P(m_{Sj}, \hat{m}_{Sj})P(m_{Bj}, \hat{m}_{Bj})P(m_{Cj}, \hat{m}_{Cj}) \\ &\cdot P(m_{1j}, \hat{m}_{1j})P(m_{2j}, \hat{m}_{2j})P(m_{3j}, \hat{m}_{3j})P(m_{4j}, \hat{m}_{4j}) \\ &\cdot P(m_{5j}, \hat{m}_{5j})P(m_{6j}, \hat{m}_{6j})P(m_{7j}, \hat{m}_{7j}).\end{aligned}$$

When allowing the π^0 mode BF to float, one can also constrain our signal mode BF to be 0 in the fit in order to improve the measurement of the π^0 mode BF.

As with the standard fit, we fit to the unblinded on-peak event counts in the four regions of nuEP and $m_{\rm ES}$ in the clpi0 sideband region using signal and $B\overline{B}$ sideband templates from MC, and continuum sideband templates from off-peak data.

11.1 Electron Channel

First, we perform the fit and constrain the π^0 mode branching fraction. The unblinded region event counts in the clpi0 sideband region are provided in table 11.1. The fit results can be seen in table 11.2. The fitted value of $\mathcal{B}_S$ comes out to be, within statistical errors, 0, as one might expect. The fitted SL BF values are consistent with the input means in their gaussian PDFs. Also, the scale factor for $B\overline{B}$ is near 1, indicating that the MC estimate of the efficiency of generic B events is accurate.[1] We also provide the results of fits to 10,000 toy MC experiments (with a constrained π^0 mode branching fraction) using off-peak data as the continuum template in the table.

Next, we allow the π^0 mode branching fraction to float. The results can be seen in table 11.3. The fitted value of $\mathcal{B}_S$ comes out to be, within statistical errors, 0, as one might expect. The fitted value of $\mathcal{B}_{\pi^0}$ also comes out to be, within statistical errors, 0, so that the fit is not an improvement in the measurement of the π^0 mode BF. The other fitted SL BF values are consistent with the input means in their gaussian PDFs. Once again, the scale factor for generic B is near 1. We also provide the results of fits to 10,000 toy MC experiments (with the π^0 mode branching fraction allowed to

[1] We allow for the possibility that the shape of generic B events is correct, but not the size, hence the reason for allowing the scale factor for generic B to float. However, a large deviation from one would indicate some serious errors in our MC modeling of generic B events.

Table 11.1. Region event counts, electron channel, validation sample (for MC), scaled to on-peak luminosities, in the π^0 sideband region. For any specific modes, the branching fractions are given in table 5.5. The continuum MC is not used in the fit, but is just provided for comparison vs. off-peak data.

	Counts			
Sample	S	$B1$	$B2$	$B3$
On-peak	115.00	208.00	229.00	851.00
Off-peak	20.08	134.85	22.97	280.21
Signal MC	4.33	0.62	0.72	0.16
Total continuum MC	14.82	45.38	12.13	112.55
Total $B\overline{B}$ MC	87.30	76.89	218.65	528.21
Generic B MC	31.79	49.83	129.08	463.27
$B^0 \to \pi^-\ell^+\nu_\ell$ MC	3.38	3.14	0.41	2.93
$B^0 \to \rho^-\ell^+\nu_\ell$ MC	16.06	10.85	43.93	39.91
$B^+ \to \eta\ell^+\nu_\ell$ MC	3.75	2.44	7.44	5.39
$B^+ \to \eta'\ell^+\nu_\ell$ MC	0.06	0.07	0.58	0.44
$B^+ \to \omega\ell^+\nu_\ell$ MC	1.25	0.25	1.31	3.03
$B^+ \to \pi^0\ell^+\nu_\ell$ MC	29.81	9.20	34.50	9.35
$B^+ \to \rho^0\ell^+\nu_\ell$ MC	1.20	1.11	1.39	3.89

float) using off-peak data as the continuum template in the table.

Finally, we allow the π^0 mode branching fraction to float, and fix our signal mode branching fraction to be 0. The results can be seen in table 11.4. The fitted value of $\mathcal{B}_{\pi^0}$, while an improvement over the previous fit, still comes out to be, within statistical errors, 0, so that the fit is not an overall improvement in the measurement of the π^0 mode BF. The other fitted SL BF values are consistent with the input means in their gaussian PDFs. Once again, the scale factor for generic B is near 1. We also provide the results of fits to 10,000 toy MC experiments (with the π^0 mode branching fraction allowed to float and the signal branching fraction set to 0 in experiment generation and fits) using off-peak data as the continuum template in the table.

Table 11.2. Results of fitting to the π^0 sideband in the electron channel, with the π^0 SL mode branching-fraction distribution modeled in the fit by a gaussian. The fit results for the 40 template scale factors are not included in this table.

Fit Parameter	Fitted Value (Parabolic Error)	Fit Constraint (Uncertainty)
$\mathcal{B}_S \times 10^6$	4.47(1402)	-
β_B	1.03(12)	-
$\mathcal{B}_{\pi^0} \times 10^4$	1.47(20)	1.28(9)
$\mathcal{B}_{\pi^+} \times 10^4$	2.76(38)	2.76(19)
$\mathcal{B}_\rho \times 10^4$	4.05(101)	4.28(52)
$\mathcal{B}_\eta \times 10^4$	1.55(70)	1.60(35)

Table 11.3. Results of fitting to the π^0 sideband in the electron channel, with the π^0 SL mode branching fraction allowed to float completely. In the table we provide the fit constraint values for the π^0 SL branching fraction just for reference, even though they are not inputs to the fit. The fit results for the 40 template scale factors are not included in this table.

Fit Parameter	Fitted Value (Parabolic Error)	Fit Constraint (Uncertainty)
$\mathcal{B}_S \times 10^6$	21.91(2504)	-
β_B	1.11(15)	-
$\mathcal{B}_{\pi^0} \times 10^4$	0.12(164)	1.28(9)
$\mathcal{B}_{\pi^+} \times 10^4$	2.76(38)	2.76(19)
$\mathcal{B}_\rho \times 10^4$	4.28(105)	4.28(52)
$\mathcal{B}_\eta \times 10^4$	1.60(70)	1.60(35)

11.2 Muon Channel

First, we perform the fit and constrain the π^0 mode branching fraction. The unblinded region event counts in the clpi0 sideband region are provided in table 11.5. The fit results can be seen in table 11.6. The fitted value of $\mathcal{B}_S$ comes out to be, within statistical errors, 0, as one might expect. The fitted SL BF values are consistent with the input means in their gaussian PDFs. We also provide the results of fits to 10,000

Table 11.4. Results of fitting to the π^0 sideband in the electron channel, with the π^0 SL mode branching fraction allowed to float completely. In the table we provide the fit constraint values for the π^0 SL branching fraction just for reference, even though they are not inputs to the fit. The signal branching fraction $\mathcal{B}_S$ is constrained to be 0. The fit results for the 40 template scale factors are not included in this table.

Fit Parameter	Fitted Value (Parabolic Error)	Fit Constraint (Uncertainty)
β_B	1.06(14)	-
$\mathcal{B}_{\pi^0} \times 10^4$	1.25(97)	1.28(9)
$\mathcal{B}_{\pi^+} \times 10^4$	2.76(38)	2.76(19)
$\mathcal{B}_{\rho} \times 10^4$	4.17(104)	4.28(52)
$\mathcal{B}_{\eta} \times 10^4$	1.58(70)	1.60(35)

toy MC experiments (with a constrained π^0 mode branching fraction) using off-peak data as the continuum template in the table.

However, the scale factor for $B\overline{B}$ indicates that there is excess of approximately 20% of generic B events in the data. After some considerable investigation along different paths, we concluded that the excess could be due to nonresonant $B \rightarrow X_u \ell \nu_\ell$ background that is probably not modeled completely correctly, especially in the limited region of phase space that applies to our analysis. Therefore, we will treat any unblinded excess of $B\overline{B}$ background extracted in our fit as a systematic, as will be discussed in section 13.5.

Next, we allow the π^0 mode branching fraction to float. The results can be seen in table 11.7. The fitted value of $\mathcal{B}_S$ comes out to be, within statistical errors, 0, as one might expect. The fitted value of $\mathcal{B}_{\pi^0}$ also comes out to be, within statistical errors, 0, so that the fit is not an improvement in the measurement of the π^0 mode BF. The other fitted SL BF values are consistent with the input means in their gaussian

Table 11.5. Region event counts, muon channel, validation sample (for MC), scaled to on-peak luminosities, in the π^0 sideband region. For any specific modes, the branching fractions are given in table 5.5. The continuum MC is not used in the fit, but is just provided for comparison vs. off-peak data.

	Counts			
Sample	S	$B1$	$B2$	$B3$
On-peak	100.00	162.00	169.00	617.00
Off-peak	28.76	48.84	10.04	150.77
Signal MC	2.79	0.60	0.23	0.11
Total continuum MC	15.91	70.70	10.54	131.27
Total $B\overline{B}$ MC	65.66	70.44	137.10	408.82
Generic B MC	22.55	41.83	89.66	360.75
$B^0 \to \pi^-\ell^+\nu_\ell$ MC	4.38	5.70	0.82	4.66
$B^0 \to \rho^-\ell^+\nu_\ell$ MC	11.24	9.43	23.95	28.68
$B^+ \to \eta\ell^+\nu_\ell$ MC	3.10	1.27	2.61	4.09
$B^+ \to \eta'\ell^+\nu_\ell$ MC	0.01	0.00	0.12	0.56
$B^+ \to \omega\ell^+\nu_\ell$ MC	0.44	0.43	0.49	1.24
$B^+ \to \pi^0\ell^+\nu_\ell$ MC	22.18	10.89	19.04	5.27
$B^+ \to \rho^0\ell^+\nu_\ell$ MC	1.76	0.90	0.40	3.58

PDFs. Once again, the scale factor for generic B is greater than 1. We also provide the results of fits to 10,000 toy MC experiments (with the π^0 mode branching fraction allowed to float) using off-peak data as the continuum template in the table.

Finally, we allow the π^0 mode branching fraction to float, and fix our signal mode branching fraction to be 0. The results can be seen in table 11.8. The fitted value of $\mathcal{B}_{\pi^0}$, while an improvement over the previous fit, still comes out to be, within statistical errors, 0, so that the fit is not an overall improvement in the measurement of the π^0 mode BF. The other fitted SL BF values are consistent with the input means in their gaussian PDFs. Once again, the scale factor for generic B is greater than 1. We also provide the results of fits to 10,000 toy MC experiments (with the π^0 mode branching fraction allowed to float and the signal branching fraction set to

Table 11.6. Results of fitting to the π^0 sideband in the muon channel, with the π^0 SL mode branching-fraction distribution modeled in the fit by a gaussian. The fit results for the 40 template scale factors are not included in this table.

Fit Parameter	Fitted Value (Parabolic Error)	Fit Constraint (Uncertainty)
$\mathcal{B}_S \times 10^6$	4.58(2100)	-
β_B	1.20(12)	-
$\mathcal{B}_{\pi^0} \times 10^4$	1.50(20)	1.28(9)
$\mathcal{B}_{\pi^+} \times 10^4$	2.77(38)	2.76(19)
$\mathcal{B}_\rho \times 10^4$	4.36(103)	4.28(52)
$\mathcal{B}_\eta \times 10^4$	1.61(70)	1.60(35)

Table 11.7. Results of fitting to the π^0 sideband in the muon channel, with the π^0 SL mode branching fraction allowed to float completely. In the table we provide the fit constraint values for the π^0 SL branching fraction just for reference, even though they are not inputs to the fit. The fit results for the 40 template scale factors are not included in this table.

Fit Parameter	Fitted Value (Parabolic Error)	Fit Constraint (Uncertainty)
$\mathcal{B}_S \times 10^6$	$-12.02(3740)$	-
β_B	1.15(14)	-
$\mathcal{B}_{\pi^0} \times 10^4$	2.58(201)	1.28(9)
$\mathcal{B}_{\pi^+} \times 10^4$	2.77(38)	2.76(19)
$\mathcal{B}_\rho \times 10^4$	4.27(105)	4.28(52)
$\mathcal{B}_\eta \times 10^4$	1.60(70)	1.60(35)

0 in experiment generation and fits) using off-peak data as the continuum template

in the table.

Table 11.8. Results of fitting to the π^0 sideband in the muon channel, with the π^0 SL mode branching fraction allowed to float completely. In the table we provide the fit constraint values for the π^0 SL branching fraction just for reference, even though they are not inputs to the fit. The signal branching fraction $\mathcal{B}_S$ is constrained to be 0. The fit results for the 40 template scale factors are not included in this table.

Fit Parameter	Fitted Value (Parabolic Error)	Fit Constraint (Uncertainty)
β_B	1.17(13)	-
$\mathcal{B}_{\pi^0} \times 10^4$	2.03(110)	1.28(9)
$\mathcal{B}_{\pi^+} \times 10^4$	2.77(38)	2.76(19)
$\mathcal{B}_{\rho} \times 10^4$	4.29(105)	4.28(52)
$\mathcal{B}_{\eta} \times 10^4$	1.59(70)	1.60(35)

Chapter 12

Fit Results

We discuss the extraction of the full branching fractions, the computation of the 90% Bayesian upper limits, and the conversion of these results to partial branching fractions in a restricted region of phase space. All computations in this chapter will include only statistical uncertainties associated with the fit. The incorporation of systematic uncertainties will follow in chapter 13.

12.1 Branching Fractions

For the muon and electron channels, we perform the fit on unblinded on-peak data, using the unblinded off-peak data as the continuum template. We also perform a joint fit to the data, where the electron- and muon-channel branching fractions are constrained to be equal. We first discuss and present the muon-channel separate fit results, though there will be some discussion of the electron and joint fits included.

For the muon-channel separate fit, the unblinded event counts are given in table 12.1. These results are graphically displayed in figure 12.1(f). (Other data and MC comparison distributions are given in the other subfigures of figure 12.1.) The fit

values are given in tables 12.2, 12.3, and 12.4.

In order to obtain MINOS-based [36] asymmetric errors on the signal BF, we trace out a negative log-likelihood curve,[1] as seen in figure 12.2. This is done by fixing the signal BF to a series of values (sampled in increments of 1×10^{-7}), and refitting, thus obtaining a minimum negative log-likelihood corresponding to each value. The minimum point of the curve corresponds to our fitted signal BF and negative log-likelihood ($-\ln \mathcal{L}_{\min}$). Because we are using twice the negative log-likelihood, the BFs corresponding to the intersection of the curve with a horizontal line at $-2\ln \mathcal{L}_{\min} + 1$ give us the minimum and maximum of a frequentist two-sided 68% confidence-level interval, if conditions such as a large sample limit hold. We therefore also need to check the coverage of this MINOS confidence interval using 10,000 toy MC events, and do find that it gives a relatively correct coverage (67.69% for the separate electron fit, 68.5% for the separate muon fit, and 63.77% for the joint fit).

The signal BF fit results for the muon channel (as well as the electron channel, averaged results, and joint fit) are given in table 12.5.

For the electron-channel separate fit, the unblinded event counts are given in table 12.6. These results are graphically displayed in figure 12.3(f). (Other data and MC comparison distributions are given in the other subfigures of figure 12.3.) The fit values are given in tables 12.7, 12.8, and 12.9. The scanned electron-channel likelihood can be seen in figure 12.4.

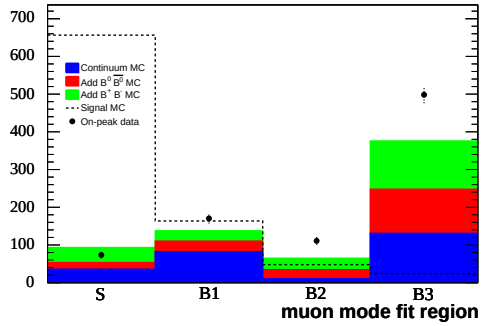
For the joint fit, the unblinded event counts are given in tables 12.10 and 12.11.

[1] It is not necessary to trace out the curve to obtain the asymmetric errors since the errors are outputted by MINOS, but this curve will be needed for both the upper-limit analysis and the incorporation of systematic uncertainties.

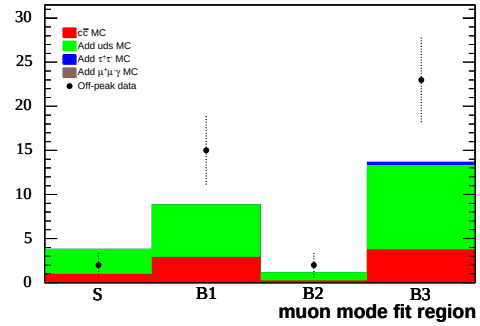
These results are graphically displayed in figure 12.5. The obtained values of the fit parameters are given in tables 12.12, 12.13, 12.14, 12.15, and 12.16. The scanned joint likelihood can be seen in figure 12.6.

Table 12.1. Region event counts, muon channel separate fit, scaled to on-peak luminosities. In addition, the fit results are provided. For any specific modes, the assumed branching fractions are given in table 5.5. The continuum MC is not used in the fit, but is provided just for a comparison with off-peak data. "Scaled" means scaled to on-peak luminosities. "Unscaled" means the raw event count.

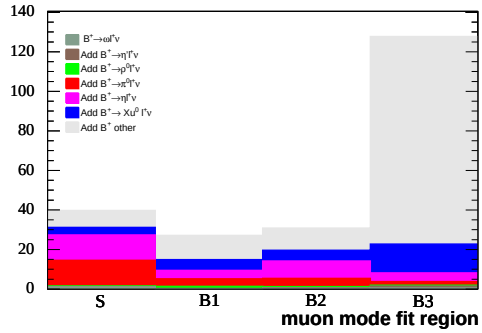
	Counts			
Sample	S	$B1$	$B2$	$B3$
Fitted total	72.73	175.88	104.12	500.28
On-peak	73.00	170.00	111.00	498.00
Fitted signal	−14.19	−3.54	−1.04	−0.52
Signal MC (BF $= 3 \times 10^{-6}$)	13.12	3.27	0.96	0.48
Fitted total B	62.96	62.55	64.42	290.22
Total $B\overline{B}$ MC	56.97	54.79	54.69	244.33
Fitted generic B	26.25	47.54	40.98	266.94
Generic B MC (incl. X_u)	21.69	40.20	32.59	221.81
Generic B MC (w/o X_u)	13.49	28.40	18.72	191.78
$B \to X_u \ell \nu_\ell$ MC	8.21	11.80	13.88	30.03
Fitted $B^0 \to \pi^- \ell^+ \nu_\ell$	1.48	1.82	0.73	2.10
$B^0 \to \pi^- \ell^+ \nu_\ell$ MC	1.48	1.84	0.71	2.09
Fitted $B^0 \to \rho^- \ell^+ \nu_\ell$	6.52	3.35	7.38	12.48
$B^0 \to \rho^- \ell^+ \nu_\ell$ MC	6.38	3.30	7.13	12.23
Fitted $B^+ \to \eta \ell^+ \nu_\ell$	14.07	4.67	9.74	4.75
$B^+ \to \eta \ell^+ \nu_\ell$ MC	12.87	4.30	8.81	4.35
Fitted $B^+ \to \eta' \ell^+ \nu_\ell$	0.11	0.09	0.34	1.02
$B^+ \to \eta' \ell^+ \nu_\ell$ MC	0.10	0.08	0.31	0.94
Fitted $B^+ \to \omega \ell^+ \nu_\ell$	1.21	0.26	0.42	0.61
$B^+ \to \omega \ell^+ \nu_\ell$ MC	1.18	0.25	0.41	0.60
Fitted $B^+ \to \pi^0 \ell^+ \nu_\ell$	12.97	3.82	4.31	1.85
$B^+ \to \pi^0 \ell^+ \nu_\ell$ MC	12.92	3.83	4.24	1.84
Fitted $B^+ \to \rho^0 \ell^+ \nu_\ell$	0.35	1.01	0.52	0.47
$B^+ \to \rho^0 \ell^+ \nu_\ell$ MC	0.34	0.99	0.50	0.46
Fitted continuum	23.96	116.86	40.73	210.58
Off-peak scaled	22.97	158.08	17.37	219.69
Off-peak unscaled	2.00	15.00	2.00	23.00
Total continuum MC	36.00	82.51	10.36	131.08



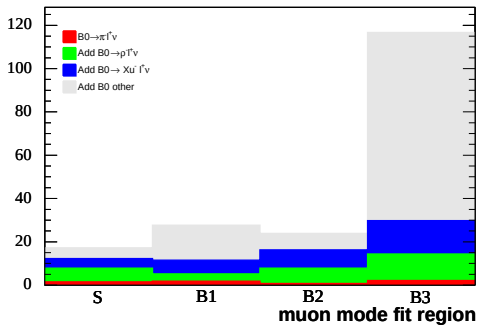
(a) Summary of data and MC distributions in the four fit regions. Signal BF = 1.5×10^{-4} (scaled up for visibility).



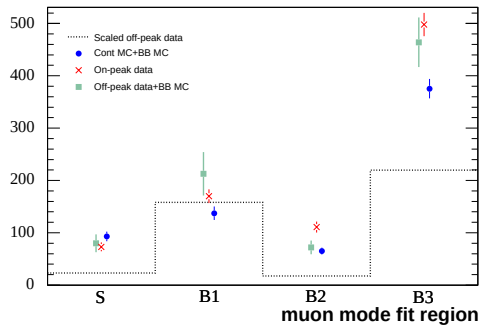
(b) Continuum data and MC distributions in the four fit regions.



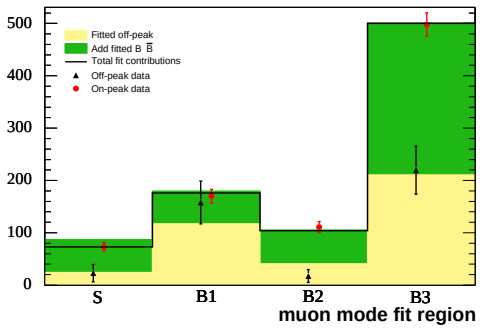
(c) Generic B^+B^- MC distributions in the four fit regions.



(d) Generic $B^0\overline{B}^0$ MC distributions in the four fit regions.



(e) Data and MC distributions in the four fit regions.



(f) Histogram of the counts in the four fit regions for the data and the fit results.

Figure 12.1. Distributions of data and MC in the four regions for the muon-channel separate fit.

Table 12.2. The main parameter results (fit value and parabolic error) for the muon-channel separate fit. See tables 12.3 and 12.4 for the template fit results.

Fit Parameter	Fitted Value (Parabolic Error)	Fit Constraint (Uncertainty)
$\mathcal{B}_S \times 10^6$	-3.24(480)	-
β_B	1.21(22)	-
$\mathcal{B}_\pi \times 10^4$	2.77(38)	2.76(19)
$\mathcal{B}_\rho \times 10^4$	4.37(10)	4.28(52)
$\mathcal{B}_\eta \times 10^4$	1.75(69)	1.60(35)

Table 12.3. The template parameter results (fit value and parabolic error) for the muon-channel separate fit for regions S and B1.

Sample	S Fitted Value (Parabolic Error)	B1 Fitted Value (Parabolic Error)
Signal	1.000(10)	1.000(19)
Generic B	1.002(158)	0.980(110)
Continuum	1.040(730)	0.739(94)
$B^0 \to \pi^- \ell^+ \nu_\ell$	1.001(378)	0.988(442)
$B^0 \to \rho^- \ell^+ \nu_\ell$	1.001(183)	0.994(228)
$B^+ \to \eta \ell^+ \nu_\ell$	1.001(107)	0.995(173)
$B^+ \to \eta' \ell^+ \nu_\ell$	1.000(707)	0.999(577)
$B^+ \to \omega \ell^+ \nu_\ell$	1.000(333)	0.998(499)
$B^+ \to \pi^0 \ell^+ \nu_\ell$	1.000(109)	0.994(212)
$B^+ \to \rho^0 \ell^+ \nu_\ell$	1.001(1000)	0.993(444)

Table 12.4. The template parameter results (fit value and parabolic error) for the muon-channel separate fit for regions B2 and B3.

Sample	B2 Fitted Value (Parabolic Error)	B3 Fitted Value (Parabolic Error)
Signal	1.000(35)	1.000(49)
Generic B	1.042(128)	0.997(48)
Continuum	2.340(730)	0.959(192)
$B^0 \to \pi^- \ell^+ \nu_\ell$	1.024(723)	0.998(408)
$B^0 \to \rho^- \ell^+ \nu_\ell$	1.014(169)	0.999(147)
$B^+ \to \eta \ell^+ \nu_\ell$	1.013(145)	0.999(208)
$B^+ \to \eta' \ell^+ \nu_\ell$	1.004(410)	1.000(267)
$B^+ \to \omega \ell^+ \nu_\ell$	1.003(334)	1.000(353)
$B^+ \to \pi^0 \ell^+ \nu_\ell$	1.014(227)	0.999(316)
$B^+ \to \rho^0 \ell^+ \nu_\ell$	1.017(719)	0.998(996)

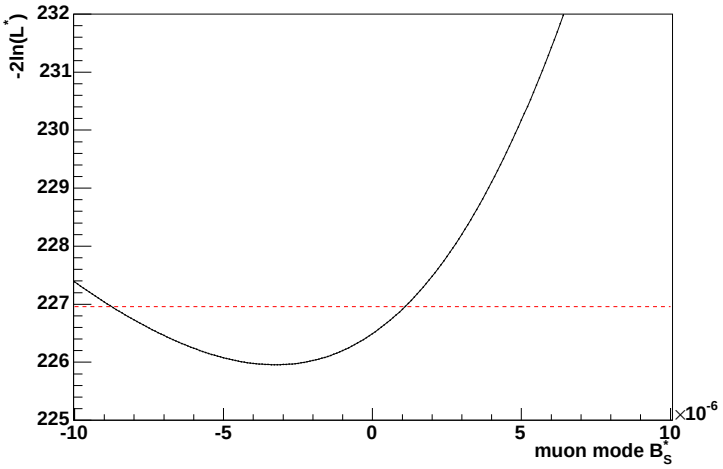


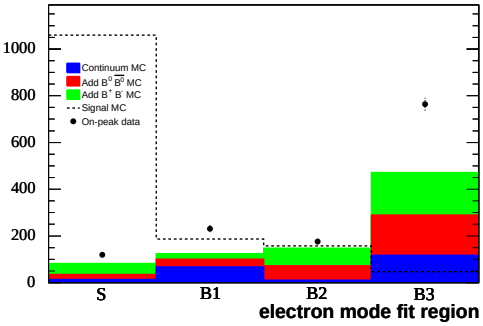
Figure 12.2. Muon-channel plot of $-2\ln\mathcal{L}^*$ vs. $\mathcal{B}^*_S$, where $\mathcal{L}^*$ is the likelihood obtained from fitting to unblinded on-peak data, fixing the signal BF to $\mathcal{B}^*_S$. The dashed red line marks $-2\ln\mathcal{L}_{\rm min} + 1$, where $\mathcal{L}_{\rm min}$ is the likelihood obtained from floating the signal BF in the fit.

Table 12.5. Comparison of signal BF results for all three fits, as well as the average of the electron- and muon-channel results. The numbers in parentheses result from correcting for bias (as obtained from toy MC).

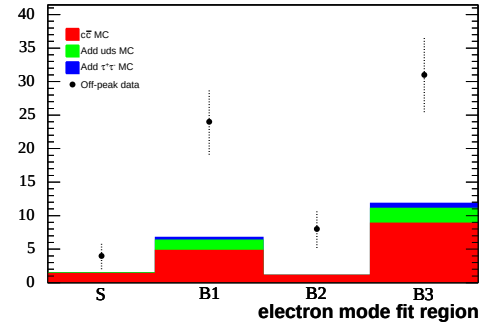
	Central Value	Parabolic Error	MINOS Error(-)	MINOS Error(+)
Muon	-3.24×10^{-6} (-3.32×10^{-6})	4.80×10^{-6}	-5.50×10^{-6}	4.33×10^{-6}
Electron	0.18×10^{-6} (0.22×10^{-6})	3.96×10^{-6}	-4.44×10^{-6}	3.61×10^{-6}
Average	-1.21×10^{-6} (-1.22×10^{-6})	3.05×10^{-6}	-3.46×10^{-6}	2.77×10^{-6}
Joint	-1.19×10^{-6} (-0.56×10^{-6})	3.16×10^{-6}	-3.43×10^{-6}	2.97×10^{-6}

Table 12.6. Region event counts, electron-channel separate fit, scaled to on-peak luminosities. In addition, the fit results are provided. For any specific modes, the assumed branching fractions are given in table 5.5. The continuum MC is not used in the fit, but is provided just for a comparison with off-peak data. "Scaled" means scaled to on-peak luminosities. "Unscaled" means the raw event count.

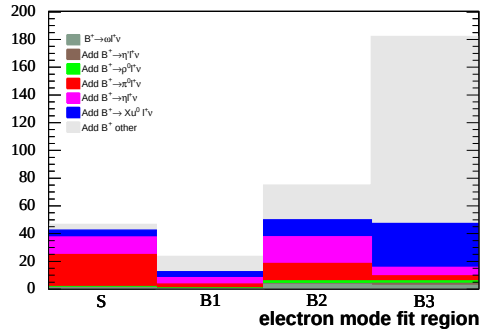
	Counts			
Sample	S	$B1$	$B2$	$B3$
Fitted total	117.34	238.90	187.43	749.89
On-peak	119.00	231.00	176.00	764.00
Fitted signal	1.26	0.22	0.19	0.06
Signal MC (BF $= 3 \times 10^{-6}$)	21.19	3.74	3.15	0.94
Fitted total B	67.63	58.44	137.94	391.35
Total $B\overline{B}$ MC	67.92	56.13	136.98	354.43
Fitted generic B	25.15	43.93	92.00	356.83
Generic B MC (incl. X_u)	22.42	40.22	85.68	317.35
Generic B MC (w/o X_u)	9.38	29.84	49.76	258.55
$B \to X_u \ell \nu_\ell$ MC	13.04	10.37	35.92	58.79
Fitted $B^0 \to \pi^- \ell^+ \nu_\ell$	1.46	0.63	0.30	2.14
$B^0 \to \pi^- \ell^+ \nu_\ell$ MC	1.47	0.64	0.31	2.14
Fitted $B^0 \to \rho^- \ell^+ \nu_\ell$	6.22	6.71	12.42	18.49
$B^0 \to \rho^- \ell^+ \nu_\ell$ MC	6.53	7.10	13.21	19.37
Fitted $B^+ \to \eta \ell^+ \nu_\ell$	10.12	3.71	15.39	4.79
$B^+ \to \eta \ell^+ \nu_\ell$ MC	12.55	4.63	19.28	5.94
Fitted $B^+ \to \eta' \ell^+ \nu_\ell$	0.29	0.04	0.38	1.30
$B^+ \to \eta' \ell^+ \nu_\ell$ MC	0.36	0.05	0.47	1.61
Fitted $B^+ \to \omega \ell^+ \nu_\ell$	0.70	0.15	3.25	2.44
$B^+ \to \omega \ell^+ \nu_\ell$ MC	0.73	0.16	3.45	2.56
Fitted $B^+ \to \pi^0 \ell^+ \nu_\ell$	23.03	2.74	12.34	3.67
$B^+ \to \pi^0 \ell^+ \nu_\ell$ MC	23.16	2.78	12.58	3.69
Fitted $B^+ \to \rho^0 \ell^+ \nu_\ell$	0.67	0.53	1.87	1.69
$B^+ \to \rho^0 \ell^+ \nu_\ell$ MC	0.70	0.56	2.00	1.76
Fitted continuum	48.45	180.23	49.30	358.49
Off-peak scaled	41.41	239.67	78.96	294.49
Off-peak unscaled	4.00	24.00	8.00	31.00
Total continuum MC	14.12	67.76	10.51	117.41



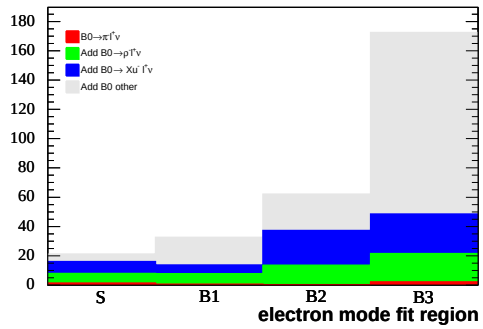
(a) Summary of data and MC distributions in the four fit regions. Signal BF = 1.5×10^{-4} (scaled up for visibility).



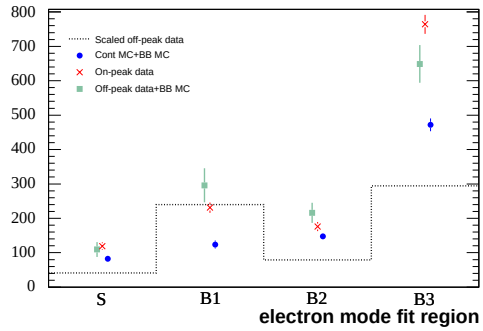
(b) Continuum data and MC distributions in the four fit regions.



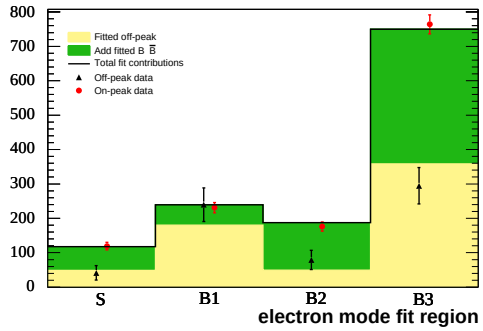
(c) Generic B^+B^- MC distributions in the four fit regions.



(d) Generic $B^0\overline{B}^0$ MC distributions in the four fit regions.



(e) Data and MC distributions in the four fit regions.



(f) Histogram of the counts in the four fit regions for the data and the fit results.

Figure 12.3. Distributions of data and MC in the four regions for the electron-channel separate fit.

Table 12.7. The main parameter results (fit value and parabolic error) for the electron-channel separate fit. See tables 12.8 and 12.9 for the template fit results.

Fit Parameter	Fitted Value (Parabolic Error)	Fit Constraint (Uncertainty)
$\mathcal{B}_S \times 10^6$	0.18(396)	-
β_B	1.11(19)	-
$\mathcal{B}_\pi \times 10^4$	2.74(38)	2.76(19)
$\mathcal{B}_\rho \times 10^4$	4.07(10)	4.28(52)
$\mathcal{B}_\eta \times 10^4$	1.29(68)	1.60(35)

Table 12.8. The template parameter results (fit value and parabolic error) for the electron-channel separate fit for regions S and B1.

Sample	S Fitted Value (Parabolic Error)	B1 Fitted Value (Parabolic Error)
Signal	1.000(8)	1.000(18)
Generic B	1.008(152)	0.982(109)
Continuum	1.170(576)	0.752(68)
$B^0 \to \pi^- \ell^+ \nu_\ell$	1.002(302)	0.997(407)
$B^0 \to \rho^- \ell^+ \nu_\ell$	1.002(163)	0.994(168)
$B^+ \to \eta \ell^+ \nu_\ell$	1.002(110)	0.996(179)
$B^+ \to \eta' \ell^+ \nu_\ell$	1.001(500)	0.999(706)
$B^+ \to \omega \ell^+ \nu_\ell$	1.001(302)	0.999(408)
$B^+ \to \pi^0 \ell^+ \nu_\ell$	1.002(86)	0.996(217)
$B^+ \to \rho^0 \ell^+ \nu_\ell$	1.003(579)	0.991(700)

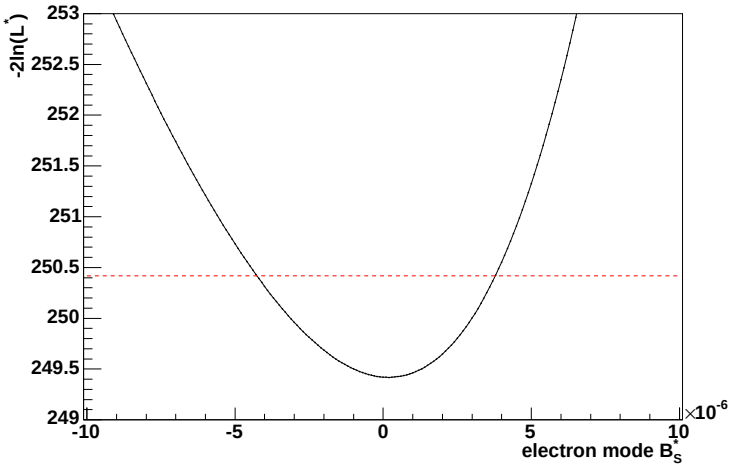
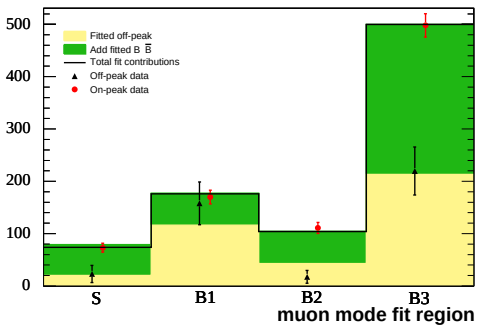
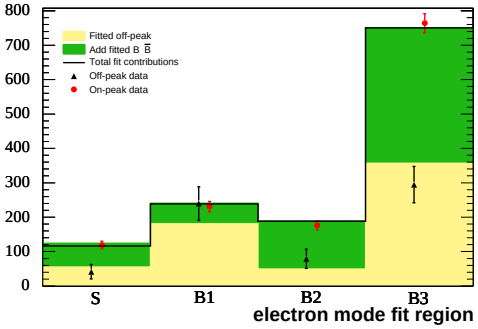


Figure 12.4. Electron-channel plot of $-2\ln\mathcal{L}^*$ vs. $\mathcal{B}^*_S$, where $\mathcal{L}^*$ is the likelihood obtained from fitting to unblinded on-peak data, fixing the signal BF to $\mathcal{B}^*_S$. The dashed red line marks $-2\ln\mathcal{L}_{\rm min}+1$, where $\mathcal{L}_{\rm min}$ is the likelihood obtained from floating the signal BF in the fit.



(a) Muon-channel histogram of the counts in the four fit regions for the data and the fit results.



(b) Electron-channel histogram of the counts in the four fit regions for the data and the fit results.

Figure 12.5. Muon- and electron-channel histograms of the counts in the four fit regions for the unblinded data and the fit results, for the joint fit.

Table 12.9. The template parameter results (fit value and parabolic error) for the electron-channel separate fit for regions B2 and B3.

Sample	B2 Fitted Value (Parabolic Error)	B3 Fitted Value (Parabolic Error)
Signal	1.000(19)	1.000(35)
Generic B	0.965(75)	1.011(40)
Continuum	0.620(180)	1.217(188)
$B^0 \to \pi^- \ell^+ \nu_\ell$	0.991(700)	1.006(380)
$B^0 \to \rho^- \ell^+ \nu_\ell$	0.989(120)	1.004(103)
$B^+ \to \eta \ell^+ \nu_\ell$	0.992(93)	1.002(163)
$B^+ \to \eta' \ell^+ \nu_\ell$	0.997(332)	1.001(196)
$B^+ \to \omega \ell^+ \nu_\ell$	0.991(211)	1.002(219)
$B^+ \to \pi^0 \ell^+ \nu_\ell$	0.989(122)	1.003(214)
$B^+ \to \rho^0 \ell^+ \nu_\ell$	0.977(437)	1.006(450)

Table 12.10. Muon-channel fit results and region event counts, from the joint fit, scaled to on-peak luminosities. In addition, the fit results are provided. For any specific modes, the assumed branching fractions are given in table 5.5. The continuum MC is not used in the fit, but is provided just for a comparison with off-peak data. "Scaled" means scaled to on-peak luminosities. "Unscaled" means the raw event count.

	Counts			
Sample	S	$B1$	$B2$	$B3$
Fitted total	73.95	176.00	103.92	499.59
On-peak	73.00	170.00	111.00	498.00
Fitted signal	−5.19	−1.29	−0.38	−0.19
Signal MC (BF $= 3 \times 10^{-6}$)	13.12	3.27	0.96	0.48
Fitted total B	59.13	60.97	61.71	286.53
Total $B\overline{B}$ MC	56.97	54.79	54.69	244.33
Fitted generic B	25.76	47.15	40.69	265.07
Generic B MC (incl. X_u)	21.69	40.20	32.59	221.81
Generic B MC (w/o X_u)	13.49	28.40	18.72	191.78
$B \to X_u \ell \nu_\ell$ MC	8.21	11.80	13.88	30.03
Fitted $B^0 \to \pi^- \ell^+ \nu_\ell$	1.47	1.81	0.72	2.08
$B^0 \to \pi^- \ell^+ \nu_\ell$ MC	1.48	1.84	0.71	2.09
Fitted $B^0 \to \rho^- \ell^+ \nu_\ell$	6.17	3.18	7.00	11.85
$B^0 \to \rho^- \ell^+ \nu_\ell$ MC	6.38	3.30	7.13	12.23
Fitted $B^+ \to \eta \ell^+ \nu_\ell$	11.33	3.77	7.85	3.83
$B^+ \to \eta \ell^+ \nu_\ell$ MC	12.87	4.30	8.81	4.35
Fitted $B^+ \to \eta' \ell^+ \nu_\ell$	0.09	0.07	0.28	0.82
$B^+ \to \eta' \ell^+ \nu_\ell$ MC	0.10	0.08	0.31	0.94
Fitted $B^+ \to \omega \ell^+ \nu_\ell$	1.14	0.25	0.40	0.58
$B^+ \to \omega \ell^+ \nu_\ell$ MC	1.18	0.25	0.41	0.60
Fitted $B^+ \to \pi^0 \ell^+ \nu_\ell$	12.84	3.79	4.28	1.83
$B^+ \to \pi^0 \ell^+ \nu_\ell$ MC	12.92	3.83	4.24	1.84
Fitted $B^+ \to \rho^0 \ell^+ \nu_\ell$	0.33	0.96	0.49	0.45
$B^+ \to \rho^0 \ell^+ \nu_\ell$ MC	0.34	0.99	0.50	0.46
Fitted continuum	20.01	116.33	42.59	213.25
Off-peak scaled	22.97	158.08	17.37	219.69
Off-peak unscaled	2.00	15.00	2.00	23.00
Total continuum MC	36.00	82.51	10.36	131.08

Table 12.11. Electron-channel fit results and region event counts, from the joint fit, scaled to on-peak luminosities. In addition, the fit results are provided. For any specific modes, the assumed branching fractions are given in table 5.5. The continuum MC is not used in the fit, but is provided just for a comparison with off-peak data. "Scaled" means scaled to on-peak luminosities. "Unscaled" means the raw event count.

	Counts			
Sample	S	$B1$	$B2$	$B3$
Fitted total	116.18	238.75	187.72	750.21
On-peak	119.00	231.00	176.00	764.00
Fitted signal	−8.37	−1.48	−1.24	−0.37
Signal MC (BF $= 3 \times 10^{-6}$)	21.19	3.74	3.15	0.94
Fitted total B	69.19	59.16	140.10	393.84
Total $B\overline{B}$ MC	67.92	56.13	136.98	354.43
Fitted generic B	25.41	44.13	92.30	358.28
Generic B MC (incl. X_u)	22.42	40.22	85.68	317.35
Generic B MC (w/o X_u)	9.38	29.84	49.76	258.55
$B \to X_u \ell \nu_\ell$ MC	13.04	10.37	35.92	58.79
Fitted $B^0 \to \pi^- \ell^+ \nu_\ell$	1.47	0.63	0.30	2.14
$B^0 \to \pi^- \ell^+ \nu_\ell$ MC	1.47	0.64	0.31	2.14
Fitted $B^0 \to \rho^- \ell^+ \nu_\ell$	6.35	6.84	12.65	18.85
$B^0 \to \rho^- \ell^+ \nu_\ell$ MC	6.53	7.10	13.21	19.37
Fitted $B^+ \to \eta \ell^+ \nu_\ell$	11.10	4.06	16.84	5.25
$B^+ \to \eta \ell^+ \nu_\ell$ MC	12.55	4.63	19.28	5.94
Fitted $B^+ \to \eta' \ell^+ \nu_\ell$	0.31	0.05	0.42	1.43
$B^+ \to \eta' \ell^+ \nu_\ell$ MC	0.36	0.05	0.47	1.61
Fitted $B^+ \to \omega \ell^+ \nu_\ell$	0.71	0.16	3.31	2.49
$B^+ \to \omega \ell^+ \nu_\ell$ MC	0.73	0.16	3.45	2.56
Fitted $B^+ \to \pi^0 \ell^+ \nu_\ell$	23.15	2.75	12.38	3.68
$B^+ \to \pi^0 \ell^+ \nu_\ell$ MC	23.16	2.78	12.58	3.69
Fitted $B^+ \to \rho^0 \ell^+ \nu_\ell$	0.68	0.54	1.90	1.72
$B^+ \to \rho^0 \ell^+ \nu_\ell$ MC	0.70	0.56	2.00	1.76
Fitted continuum	55.36	181.07	48.86	356.75
Off-peak scaled	41.41	239.67	78.96	294.49
Off-peak unscaled	4.00	24.00	8.00	31.00
Total continuum MC	14.12	67.76	10.51	117.41

Table 12.12. The main parameter results (fit value and parabolic error) for the joint fit. See tables 12.13, 12.14, 12.15, and 12.16 for the template fit results.

Fit Parameter	Fitted Value (Parabolic Error)	Fit Constraint (Uncertainty)
$\mathcal{B}_S \times 10^6$	-1.19(316)	-
$\beta_B(e)$	1.12(18)	-
$\beta_B(\mu)$	1.20(22)	-
$\mathcal{B}_\pi \times 10^4$	2.75(38)	2.76(19)
$\mathcal{B}_\rho \times 10^4$	4.15(10)	4.28(52)
$\mathcal{B}_\eta \times 10^4$	1.41(68)	1.60(35)

Table 12.13. The template parameter results (fit value and parabolic error) for the muon channel from the joint fit for regions S and B1.

Sample	S Fitted Value (Parabolic Error)	B1 Fitted Value (Parabolic Error)
Signal	1.000(10)	1.000(19)
Generic B	0.992(155)	0.979(110)
Continuum	0.870(510)	0.736(93)
$B^0 \to \pi^- \ell^+ \nu_\ell$	0.997(377)	0.988(441)
$B^0 \to \rho^- \ell^+ \nu_\ell$	0.997(182)	0.994(228)
$B^+ \to \eta \ell^+ \nu_\ell$	0.998(107)	0.996(173)
$B^+ \to \eta' \ell^+ \nu_\ell$	0.999(706)	0.999(577)
$B^+ \to \omega \ell^+ \nu_\ell$	0.998(333)	0.998(499)
$B^+ \to \pi^0 \ell^+ \nu_\ell$	0.998(108)	0.994(212)
$B^+ \to \rho^0 \ell^+ \nu_\ell$	0.996(994)	0.994(444)

Table 12.14. The template parameter results (fit value and parabolic error) for the muon channel from the joint fit for regions B2 and B3.

Sample	B2 Fitted Value (Parabolic Error)	B3 Fitted Value (Parabolic Error)
Signal	1.000(35)	1.000(49)
Generic B	1.043(128)	0.998(48)
Continuum	2.450(730)	0.971(192)
$B^0 \to \pi^- \ell^+ \nu_\ell$	1.025(724)	0.999(408)
$B^0 \to \rho^- \ell^+ \nu_\ell$	1.013(169)	0.999(147)
$B^+ \to \eta \ell^+ \nu_\ell$	1.011(144)	0.999(208)
$B^+ \to \eta' \ell^+ \nu_\ell$	1.003(409)	1.000(267)
$B^+ \to \omega \ell^+ \nu_\ell$	1.003(334)	1.000(353)
$B^+ \to \pi^0 \ell^+ \nu_\ell$	1.015(227)	0.999(316)
$B^+ \to \rho^0 \ell^+ \nu_\ell$	1.017(718)	0.999(997)

12.2 Upper Limits

12.2.1 Overview

In addition to the two-sided confidence intervals, we will also quote 90% Bayesian upper limits. This requires the choice of a prior probability distribution function (PDF) in the signal branching fraction. We calculate the upper limit using two different priors, both of which take values of 0 for negative values of the signal branching fraction: a prior flat in the branching fraction ("flat BF prior"), and a prior flat in the square root of the branching fraction ("flat amplitude prior"). The second choice is based on the belief that the prior PDF for $|V_{ub}|$ is flat.

As will be shown, the choice of prior does make a large enough difference in the upper limit that we will quote upper limits for each choice of prior.

Table 12.15. The template parameter results (fit value and parabolic error) for the electron channel from the joint fit for regions S and B1.

Sample	S Fitted Value (Parabolic Error)	B1 Fitted Value (Parabolic Error)
Signal	1.000(7)	1.000(18)
Generic B	1.014(152)	0.982(109)
Continuum	1.337(496)	0.755(68)
$B^0 \to \pi^- \ell^+ \nu_\ell$	1.003(302)	0.997(407)
$B^0 \to \rho^- \ell^+ \nu_\ell$	1.004(163)	0.994(168)
$B^+ \to \eta \ell^+ \nu_\ell$	1.003(110)	0.996(179)
$B^+ \to \eta' \ell^+ \nu_\ell$	1.002(501)	0.999(706)
$B^+ \to \omega \ell^+ \nu_\ell$	1.002(302)	0.999(408)
$B^+ \to \pi^0 \ell^+ \nu_\ell$	1.004(86)	0.996(217)
$B^+ \to \rho^0 \ell^+ \nu_\ell$	1.006(580)	0.991(700)

12.2.2 Implementation

In section 12.1, we traced out the negative log-likelihood curve (times a factor of two) for each of the three fits (electron, muon, and joint) by fixing the signal BF to a series of values, and refitting. We convert this log-likelihood curve to a likelihood curve by dividing each point by a factor of -2 and exponentiating. A comparison of these curves can be seen in figure 12.7. One thing to note about these likelihood curves is that their values are negligible beyond BF $= 1.5 \times 10^{-5}$.

Using Bayes' rule, we calculate the posterior conditional PDF of the signal BF:

$$p(\mathcal{B}_S|\mathbf{x}) = \frac{\mathcal{L}(\mathbf{x}|\mathcal{B}_S)p(\mathcal{B}_S)}{\int_{-\infty}^{\infty} \mathcal{L}(\mathbf{x}|\mathcal{B}_S)p(\mathcal{B}_S)d\mathcal{B}_S}, \tag{12.1}$$

where $\mathbf{x}$ is shorthand for the data, $\mathcal{B}_S$ is the signal branching fraction, $p(\mathcal{B}_S)$ is the

Table 12.16. The template parameter results (fit value and parabolic error) for the electron channel from the joint fit for regions B2 and B3.

Sample	B2 Fitted Value (Parabolic Error)	B3 Fitted Value (Parabolic Error)
Signal	1.000(19)	1.000(35)
Generic B	0.964(75)	1.010(40)
Continuum	0.620(180)	1.211(185)
$B^0 \to \pi^- \ell^+ \nu_\ell$	0.991(700)	1.006(380)
$B^0 \to \rho^- \ell^+ \nu_\ell$	0.988(120)	1.004(103)
$B^+ \to \eta \ell^+ \nu_\ell$	0.991(93)	1.003(163)
$B^+ \to \eta' \ell^+ \nu_\ell$	0.997(332)	1.001(196)
$B^+ \to \omega \ell^+ \nu_\ell$	0.991(211)	1.002(219)
$B^+ \to \pi^0 \ell^+ \nu_\ell$	0.988(122)	1.003(214)
$B^+ \to \rho^0 \ell^+ \nu_\ell$	0.976(436)	1.006(450)

prior PDF for the signal branching fraction, and $\mathcal{L}(\mathbf{x}|\mathcal{B}_S)$ is the likelihood of $\mathcal{B}_S$ given data $\mathbf{x}$.

For the upper-limit extraction, we choose a probability of $1-\alpha$ and integrate to find the branching-fraction upper limit, $\mathcal{B}_S^{\rm UL}$:

$$1-\alpha = \int_{-\infty}^{\mathcal{B}_S^{\rm UL}} p(\mathcal{B}_S|\mathbf{x})d\mathcal{B}_S = \frac{\int_{-\infty}^{\mathcal{B}_S^{\rm UL}} \mathcal{L}(\mathbf{x}|\mathcal{B}_S)p(\mathcal{B}_S)d\mathcal{B}_S}{\int_{-\infty}^{\infty} \mathcal{L}(\mathbf{x}|\mathcal{B}_S)p(\mathcal{B}_S)d\mathcal{B}_S}. \tag{12.2}$$

Because we are determining the 90% upper limit, are using nonnegative priors, and have an effective signal BF cutoff is at $\mathcal{B}_S = 1.5\times10^{-5}$,[2] equation 12.2 becomes:

$$0.90 = \int_{0}^{\mathcal{B}_S^{\rm UL}} p(\mathcal{B}_S|\mathbf{x})d\mathcal{B}_S = \frac{\int_{0}^{\mathcal{B}_S^{\rm UL}} \mathcal{L}(\mathbf{x}|\mathcal{B}_S)p(\mathcal{B}_S)d\mathcal{B}_S}{\int_{0}^{1.5\times10^{-5}} \mathcal{L}(\mathbf{x}|\mathcal{B}_S)p(\mathcal{B}_S)d\mathcal{B}_S}. \tag{12.3}$$

[2] As mentioned earlier, this is the last sampled point for the likelihood and beyond this point, the value of the likelihood is nearly zero.

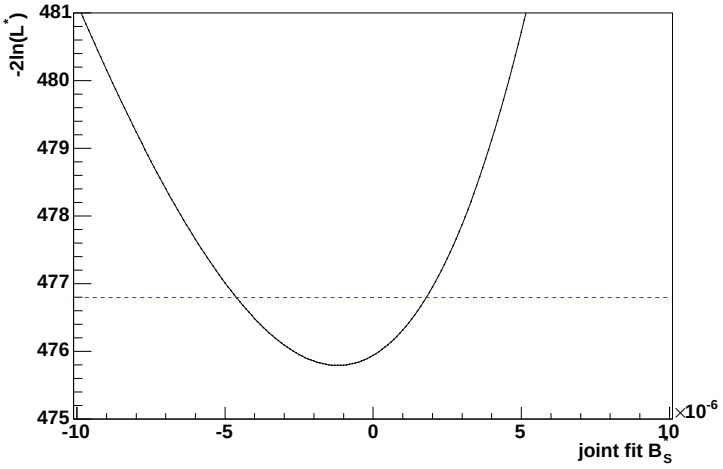
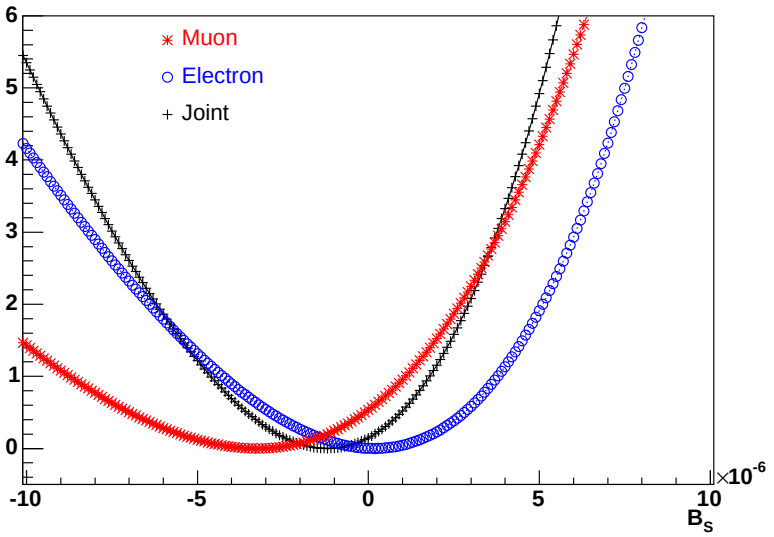
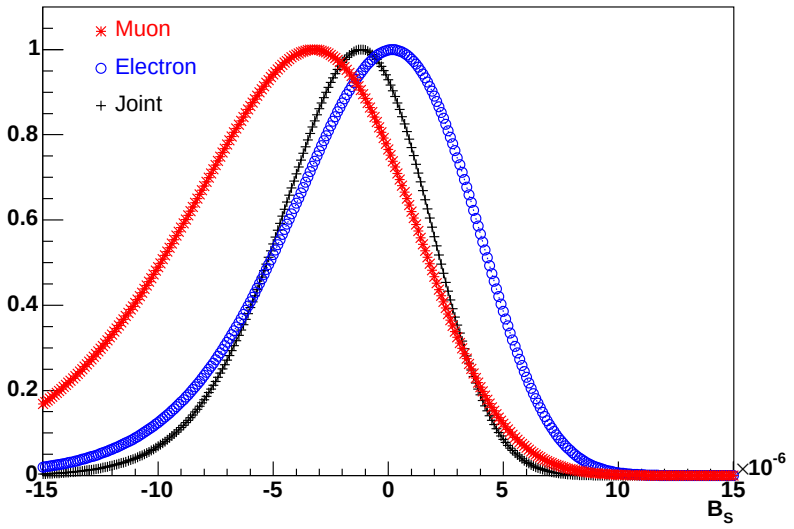


Figure 12.6. Joint fit plot of $-2\ln\mathcal{L}^*$ vs. $\mathcal{B}_S^*$, where $\mathcal{L}^*$ is the likelihood obtained from fitting to unblinded on-peak data, fixing the signal BF to $\mathcal{B}_S^*$. The dashed red line marks $-2\ln\mathcal{L}_{\rm min}+1$, where $\mathcal{L}_{\rm min}$ is the likelihood obtained from floating the signal BF in the fit.

For the assumption of the flat BF prior, we implement this upper-limit calculation as follows: We first start with the a set of sampled likelihood points with the signal BF ranging from 0 to 1.5×10^{-5}. We calculate the integral for the entire BF range, $I_{\rm ent}$, using a simple trapezoidal algorithm: for each $\mathcal{B}_S$ interval, we calculate the area by multiplying the length of the interval by the average of the likelihood values at the start and end of the interval boundaries. Then, we repeat the integration to find the value of $\mathcal{B}_S$ at which the value of the integral is equal to $0.90I_{\rm ent}$. This involves finding the first $\mathcal{B}_S$ interval with an incremental area that results in the integral exceeding $0.90I_{\rm ent}$ and then performing a linear interpolation (consistent with our trapezoidal integration method) within this interval to find the value of $\mathcal{B}_S$ that gives a total integral of $0.90I_{\rm ent}$. This gives us a 90% Bayesian upper limit, which also needs to be corrected for bias, using values obtained from toy MC studies.



(a) Distributions of refitted $-2\ln\mathcal{L}$ for different fixed values of the signal BF. The minimum likelihoods are adjusted to be zero.



(b) Distributions of refitted $\mathcal{L}$ for different fixed values of the signal BF. The maximums are adjusted to be one.

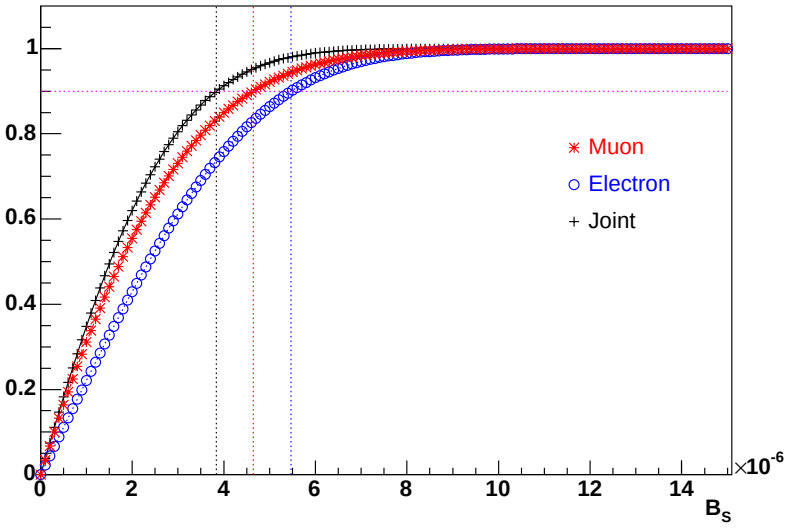
Figure 12.7. Distributions of refitted $-2\ln\mathcal{L}$ and $\mathcal{L}$ for different fixed values of the signal BF, for all three fits.

For the assumption of the flat amplitude prior, we have two equivalent options: integrate using this prior ($\propto 1/\sqrt{\mathcal{B}_S}$) over $d\mathcal{B}_S$, or integrate using a flat prior over $d\sqrt{\mathcal{B}_S}$. The first option is more difficult to accurately implement numerically due to the the prior going to infinity at small values of $\mathcal{B}_S$, so therefore we pick the second option. This involves converting the sampled likelihood points to be in terms of $\sqrt{\mathcal{B}_S}$ instead of $\mathcal{B}_S$, resulting in nonuniform intervals. Then we perform the same procedure as for the case with the flat BF prior, but solve for the value of $\sqrt{\mathcal{B}_S}$ that gives an integral which is 90% of the total integral. Finally, we square this value to obtain the upper limit. This gives us a 90% Bayesian upper limit, which also needs to be corrected for bias, using values obtained from toy MC studies.

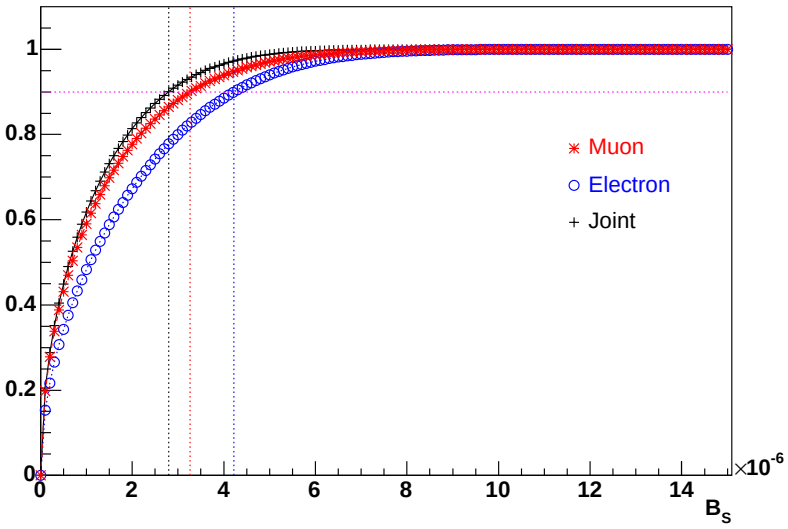
Table 12.17. Comparison of signal BF results, including 90% Bayesian upper limits, for all three fits, including only statistical uncertainties. All values have been multiplied by 10^6. Numbers in parentheses are after correcting for toy MC bias. "UL BF prior" means the upper limit was obtained using the flat BF prior. "UL amp prior" means the upper limit was obtained using the flat amplitude prior.

All Values $\times 10^6$	Fitted Value	Parabolic Error	MINOS Error(-)	MINOS Error(+)	UL BF Prior	UL Amp Prior
Muon	−3.24 (−3.32)	4.80	−5.50	4.33	<4.64 (<4.57)	<3.27 (<3.20)
Electron	0.18 (0.22)	3.96	−4.44	3.61	<5.47 (<5.51)	<4.22 (<4.26)
Joint	−1.19 (−0.56)	3.16	−3.43	2.97	<3.84 (<4.46)	<2.80 (<3.42)

The cumulative distribution functions of the signal BF posterior PDFs can be seen in figure 12.8, for both choices of prior. A summary of the results is given in table 12.17. There are a couple of things to note: First, as mentioned previously,



(a) Cumulative distribution functions assuming the flat BF prior.



(b) Cumulative distribution functions assuming the flat amplitude prior.

Figure 12.8. Cumulative distribution functions in signal BF assuming either the flat BF prior or the flat amplitude prior, for all three fits. The vertical lines indicate the 90% upper limit. The horizontal lines are at 0.90.

the choice of prior makes a significant difference. The upper limits are lower when using the flat amplitude priors, and this is in line with our intuition because the flat amplitude prior goes as $1/\sqrt{\mathcal{B}_S}$, thus biasing upper limits towards lower values of $\mathcal{B}_S$.

The second interesting thing to note in the upper-limit results is that the muon-channel upper limits are lower than the electron-channel upper limits, even though the electron channel has better signal BF fit uncertainties. To examine this effect, we used a gaussian likelihood function to calculate an upper limit with a flat BF prior, and found that the much lower central value of the muon-channel fit was the reason for lower upper limit.

12.3 Partial Branching Fractions

12.3.1 Overview

As will be discussed in more detail in section 13.6, there is a potential systematic uncertainty due to the theoretical uncertainty on the signal MC decay model. In particular, we select events in a restricted region of phase space on the signal side by applying cuts on the signal lepton CM energy, the signal photon CM energy, and their CM opening angle. Thus, our signal efficiency, and consequently, our extracted total BF, are sensitive to the dynamics of the signal model. Unfortunately, at the present, there is little theoretical guidance in determining the systematic uncertainty of our signal MC model over the entire signal kinematic region.

One alternative model is the "fa0" model,[3] described earlier in section 4.1, in which

[3]Yet another study that could have be performed involves making small changes rather than

the axial vector form factor is set to 0 (as opposed to being set equal to the vector form factor as is the case in our standard signal model). As was shown previously in that section, we obtain a significantly different signal phase space distribution when using the fa0 model instead of the standard signal model.

In addition, based on the opinion of theorists, the fa0 model is not a valid model. Therefore, we have adopted a convention often used by semileptonic analyses: instead of quoting our measurements in terms of the full BF, $\mathcal{B}$, we quote in terms of the partial BF, $\Delta\mathcal{B}$, in the restricted region of phase space. The idea is that the efficiency is relatively flat in the restricted region, so theoretical uncertainty due to lack of knowledge of the decay dynamics is greatly reduced.

Quoting in terms of $\Delta\mathcal{B}$ has two main implications: First, we shift the responsibility of calculating the theoretical error on the full extracted BF to the theorist. Second, the remaining theoretical uncertainty is due to the uncertainty in the conversion from the full BF to the partial BF, and this uncertainty will be calculated in section 13.6.

12.3.2 Calculation of $\Delta\mathcal{B}$

In order to obtain a partial BF measurement from a full BF measurement, we calculate a conversion factor, ϵ.[4] We determine ϵ as follows: We reconstruct an event without applying any selection criteria. We then look at the MC generator-level kinematic information for the reconstructed signal lepton and photon. By applying

extreme changes in the relative sizes of the two form factors.

[4]This factor is different from any ϵ factors used elsewhere.

the appropriate kinematic selection, we can determine the efficiency, ϵ, or conversion factor of this selection.

Table 12.18. Kinematic regions used in quoting partial BFs.

	Original		
	Electron	Muon	Envelope
CM lepton energy (GeV)	[2, 2.85]	[1.875, 2.775]	[1.875, 2.85]
CM photon energy (GeV)	[0.65, 2.35]	[0.45, 2.35]	[0.45, 2.35]
$\cos\theta_{\ell\gamma}$	< -0.42	< -0.36	< -0.36

For the joint fit, the procedure for extracting $\Delta\mathcal{B}$ is not as simple as for the separate mode fits because this fit involves two nonoverlapping kinematic regions,[5] as can be seen in table 12.18. The question is what kinematic region should be used to determine the partial BF of the joint fit. The decision was to quote $\Delta\mathcal{B}$ for the *envelope* of the electron and muon kinematic regions (also given in table 12.18). In addition, we quote $\Delta\mathcal{B}$ in this envelope region for the separate electron- and muon-channel fits so that a direct comparison of $\Delta\mathcal{B}$ can be made.

The calculation of $\Delta\mathcal{B}$ for the envelope region involves calculating new conversion factors for the envelope region for the separate electron- and muon-channel fits. We obtain the joint fit conversion factor by averaging the electron- and muon-channel conversion factors, with weights obtained from the uncertainties on the BFs of the respective fits. The conversion factors are given in table 12.19.

In order to obtain the results in terms of $\Delta\mathcal{B}$, we simply multiply all values by the conversion factor ϵ for a chosen kinematic region. The converted central values and 90% Bayesian upper limits are given in table 12.20 (statistical uncertainties only).

[5]The analysis was not originally designed with the joint fit in mind.

Table 12.19. Conversion factors, ϵ, used to obtain partial BFs. The definitions of the kinematic regions are given in table 12.18.

Kinematic Region	Muon ϵ	Electron ϵ	Joint ϵ
Original	0.40	0.41	–
Envelope	0.40	0.48	0.45

Table 12.20. Comparison of signal $\Delta\mathcal{B}$ results, including 90% Bayesian upper limits, for all three fits, including only statistical uncertainties. All values have been multiplied by 10^6. "UL BF prior" means the upper limit was obtained using the flat BF prior. "UL amp prior" means the upper limit was obtained using the flat amplitude prior. The definitions of the kinematic regions are given in table 12.18. "Original" refers to the original kinematic regions. "Envelope" refers to the envelope of the electron and muon kinematic regions.

All Values $\times 10^6$	Central Value	Parabolic Error	MINOS Error(-)	MINOS Error(+)	UL BF Prior	UL Amp Prior
Original						
Muon	−1.32	1.91	−2.19	1.73	<2.09	<1.46
Electron	0.09	1.62	−1.82	1.48	<2.42	<1.86
Envelope						
Muon	−1.33	1.92	−2.20	1.74	<2.10	<1.47
Electron	0.11	1.90	−2.13	1.73	<2.84	<2.17
Joint	−0.25	1.41	−1.53	1.33	<2.24	<1.70

Chapter 13

Systematics

Systematic uncertainties are considered in four categories: experimental (essentially data/MC discrepancies) uncertainties, uncertainties related to the number of B mesons, $B\overline{B}$-background uncertainties, and theoretical uncertainties. Although it will be clear when the full results are summarized in section 14.1, it is worth noting here that our measurements are dominated by statistical uncertainties, and that systematic uncertainties will ultimately be a relatively small fraction of the overall uncertainty when added in quadrature to the statistical uncertainty.

13.1 Treatment

Our measurements include both multiplicative and additive systematic uncertainties. Multiplicative systematics are quoted in percentage (or fractional) terms, relative to the signal BF. Additive systematics are quoted in absolute terms, and are added in quadrature to the statistical uncertainties.

Multiplicative systematics cannot be evaluated by simply multiplying the central BF value by the multiplicative fractional effect because the resulting value of the

systematic would be highly sensitive to small fluctuations in the central BF value and would ignore the probabilistic distribution of the extracted signal BF. Because our evaluation of multiplicative systematics should take into account the likelihood distribution of the signal BF, we convolute the likelihood function with a gaussian whose width incorporates multiplicative effects. Using notation consistent with section 12.2.2 (except that we drop the xs representing the data in writing the likelihoods), we write the expression for the convoluted signal branching-fraction likelihood $\mathcal{L}_{\rm new}$ that includes the effect of a single multiplicative systematic m (quoted as a fraction):

$$\mathcal{L}_{\rm new}(\mathcal{B}_S) = \int_{-\infty}^{\infty} \mathcal{L}_{\rm orig}(\mathcal{B}'_S)\mathcal{G}(\mathcal{B}_S, \mathcal{B}'_S, \sigma(\mathcal{B}'_S))d\mathcal{B}'_S, \tag{13.1}$$

where $\mathcal{L}_{\rm orig}$ is the original likelihood function, $\mathcal{G}(x, \mu, \sigma)$ is the value of a gaussian distribution with mean μ and sigma σ at x, and $\sigma(\mathcal{B}_S) \equiv |m \cdot \mathcal{B}_S|$.

We then determine the total asymmetric two-sided 68% confidence-level interval by finding the BF values at which we obtain $-2\ln\mathcal{L}_{\rm min}+1$, where $\mathcal{L}_{\rm min}$ is the minimum value of $\mathcal{L}_{\rm new}$. The systematic uncertainties are extracted from the full interval by subtracting off the statistical uncertainties in quadrature. It is simple to extend this analysis to multiple systematics by including all of their effects in the calculation of $\sigma(\mathcal{B}_S)$.

We classify additive systematics in two ways: those whose evaluation does not include implicit multiplicative effects, and those whose evaluation does include implicit multiplicative effects. Implicit multiplicative effects are due to a simultaneous scaling of the signal and background templates.

We evaluate additive systematics that do not have associated implicit multiplicative effects by adjusting the appropriate fit templates, and then refitting. We then compare the change in the extracted signal BF, relative to the result from the original fit, in order to obtain the additive effect.

An additive systematic that has an associated implicit multiplicative effect needs to be evaluated by first treating it as if it had no associated implicit multiplicative effect, i.e. we refit after adjusting the appropriate templates. The additive effect we obtain from such an analysis is incorrect in that the implicit multiplicative effect is incorrectly accounted for because it does not allow for the distribution of the statistical likelihood function. Therefore, we subtract off the incorrect multiplicative effect (equal to the multiplicative fraction m times the central value of the BF) to obtain the correct additive effect.[1] Then, the multiplicative effect is evaluated correctly by incorporating it into the convolution, as described above.

In order to determine a Bayesian 90% upper limit, we integrate the full convoluted likelihood function with a prior for the signal BF, using the same procedure described in section 12.2 in determining upper limits (where only statistical uncertainties were taken into account).

[1] In all of our additive systematics with associated implicit multiplicative effects, the additive and multiplicative effects have 100% correlation, so we perform a straight subtraction, rather than a subtraction in quadrature.

13.2 Experimental Systematics

13.2.1 Particle Identification

Particle identification (PID) systematics arise from discrepancies in data/MC PID efficiencies. As mentioned in section 7.1.3, we use PID tweaking to bring data/MC efficiencies into better agreement. The PID-tweaking correction is applied in bins of lepton lab momentum and lab angle. Each one of these bins has an associated statistical error, due to the statistics of the control samples. We average these statistical errors to determine the systematic error on the lepton PID. The size of the systematic effect is 2.2% (3.5%) for the electron (muon) channel.

Data/MC PID discrepancies affect all MC samples in a correlated fashion, so we must take these correlations into account. We adjust the signal MC template up and down by the systematic effect. We do not need to adjust the generic B template because its magnitude is allowed to float in the fit, nor do we adjust the continuum template because it is obtained from off-peak data.

In the case of the seven semileptonic background modes, there is the complicating fact that we constrain their BFs in the fit to externally measured central values within their statistical and experimental systematic errors, with the experimental systematic error already including PID systematics. For the η modes,[2] this is not a problem because CLEO and *BABAR* experimental systematics are uncorrelated. Therefore, we adjust the η templates up and down by the systematic effect (simultaneously

[2]The η' mode is included when we refer to the η modes.

with the adjustment of the signal template). For the π and ρ modes,[3] the externally measured values and uncertainties are also obtained from *BABAR* [28]. In fact, the PID systematics for these modes should completely cancel out if the lepton spectra match completely between our analysis and the *BABAR* π and ρ analysis.[4] The lepton spectra, in fact, do not completely match, therefore we do not have a complete cancellation. We decide to be conservative and leave in the systematic effects in the fit constraints on the π and ρ BFs.[5]

Because there is an implicit multiplicative systematic, we extract the additive systematic as described in section 13.1 by removing the incorrectly calculated multiplicative effect. We do this for both scenarios (adjusting the templates upwards and downwards) and take the average effect as the additive systematic, and include the multiplicative effect in the convolution of the likelihood function. The evaluated additive and multiplicative effects are given in table 13.11, presented in section 13.7.

13.2.2 Tracking Efficiency

Our signal lepton must pass track selection criteria (section 7.1), and the standard *BABAR* prescription for evaluating tracking systematics is to make a bias correction of -0.8% and in addition, include a 1.3% multiplicative effect.

Using reasoning similar to that given in section 13.2.1, we apply the 1.3% multi-

[3]The ω mode is included when we refer to the ρ modes.

[4]For simplicity, assume a single SL mode whose experimental values also come from *BABAR*. Now, say there is a systematic effect that results in the *BABAR* MC PID efficiency being lower than the data PID efficiency. Then the extracted SL BF would be too high, and would need to be corrected downwards. On the other hand, for a given BF, the SL template would need to be adjusted upwards. The combined effects cancel out in the SL template evaluated at the adjusted BF.

[5]It is worth noting that the overall effect of PID systematics is tiny, so there is little cost to taking a conservative approach.

plicative effect only to the signal and η templates (upwards and downwards). However, we must apply the bias correction to the signal and all 7 semileptonic templates. Therefore, we have three evaluation scenarios:

1. Adjust the templates for the signal and η modes by $-(1.3+0.8)\%$, and templates for the π and ρ modes by -0.8%.

2. Adjust the templates for the signal, η, π, and ρ modes by -0.8%.

3. Adjust the templates for the signal and η modes by $(1.3-0.8)\%$, and templates for the π and ρ modes by -0.8%.

After removing the implicit and incorrect multiplicative effect, we take the worst of the additive effects as the additive systematic, and include the multiplicative effect in the convolution of the likelihood function. The evaluated additive and multiplicative effects are given in table 13.11, presented in section 13.7.

13.3 Neutral Reconstruction

The *BABAR* prescription (determined using control samples) for evaluating the systematic on neutral reconstruction is to use a systematic error of $(1.9-0.2E)\%$, where E is the signal photon lab energy(GeV). We use the signal photon lab energy spectrum in the S region of the fit and calculate the average error over this spectrum. This effect needs to be separately determined for the signal and seven semileptonic modes. Table 13.1 shows the averaged effects.

Table 13.1. Averaged multiplicative corrections due to neutral reconstruction efficiency systematics. The joint corrections are obtained by weighing the electron- and muon-channel corrections by their respective luminosity-weighted event counts in the S region.

Sample	Averaged Systematic (%)		
	Muon	Electron	Joint
Signal	1.58	1.57	1.58
$B^0 \to \pi^- \ell^+ \nu_\ell$	1.75	1.67	1.71
$B^0 \to \rho^- \ell^+ \nu_\ell$	1.67	1.61	1.63
$B^+ \to \eta \ell^+ \nu_\ell$	1.59	1.59	1.59
$B^+ \to \eta' \ell^+ \nu_\ell$	1.73	1.50	1.55
$B^+ \to \omega \ell^+ \nu_\ell$	1.56	1.59	1.57
$B^+ \to \pi^0 \ell^+ \nu_\ell$	1.62	1.58	1.59
$B^+ \to \rho^0 \ell^+ \nu_\ell$	1.71	1.71	1.71

In order to calculate the systematic effect, we simultaneously adjust the templates both up and down. After removing the implicit and incorrect multiplicative effect, we take the average of the additive effects as the additive systematic, and include the multiplicative effect in the convolution of the likelihood function. The evaluated additive and multiplicative effects are given in table 13.11, presented in section 13.7.

13.3.1 Shape

Our signal BF extraction fit is sensitive to the shapes of the different component templates in the four regions of nuEP and $m_{\rm ES}$ (S, B1, B2, and B3), as described in detail in chapter 9. There are ten different component samples, and nine of these component template shapes are obtained from MC. Therefore, we need to estimate the effect of potential data/MC discrepancies in nuEP and $m_{\rm ES}$, in signal and background.

Our approach is study the $B^+ \to \pi^+ \overline{D}^0 (\to K^+ \pi^-)$ control sample.[6] In order to

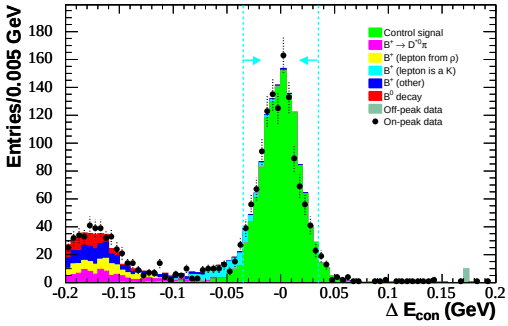
[6]Recall that we use nuEPM instead of nuEP in order to account for the "neutrino" mass in the

have a better handle on the background contributions in the $\Delta E_{\rm con}$ signal region (-35 to 35 MeV), we do not perform a background subtraction as described in section 10.2. Instead, we compare unsubtracted on-peak data versus a "hybrid" sample made up of $B\overline{B}$ MC and off-peak data.[7] We also use MC truth information to categorize the various contributions of $B\overline{B}$ MC. In addition, unlike in section 10.2, where our analysis selection criteria are not applied, we do apply any applicable analysis selection criteria (and these are different for the electron and muon channels, even though the control sample itself is the same).

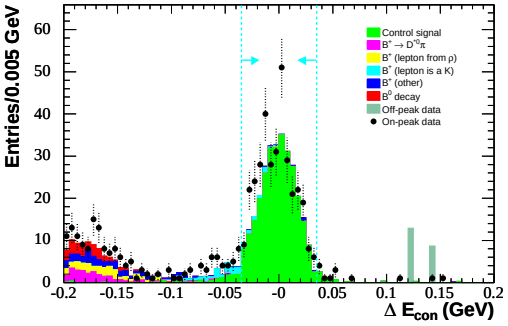
Figures 13.1 and 13.2 show a comparison of the on-peak and hybrid distributions of $\Delta E_{\rm con}$, with the applicable electron- and muon-channel analysis cuts applied, respectively. It is evident from these comparisons that even in the $\Delta E_{\rm con}$ signal region, which is clearly dominated by the control sample signal decay modeled in MC, there is an on-peak/MC discrepancy that cannot be explained by a mismodeling of background.

control sample.

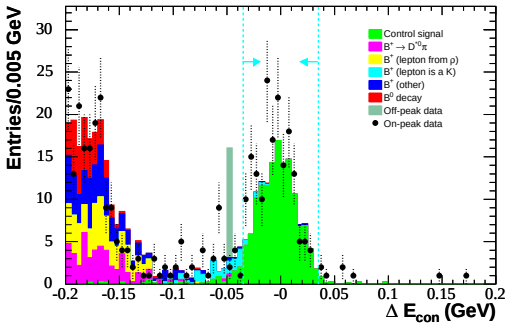
[7] Recall that the off-peak data has limited statistics (approximately one tenth of the on-peak), so there are large off-peak component fluctuations in the comparison plots.



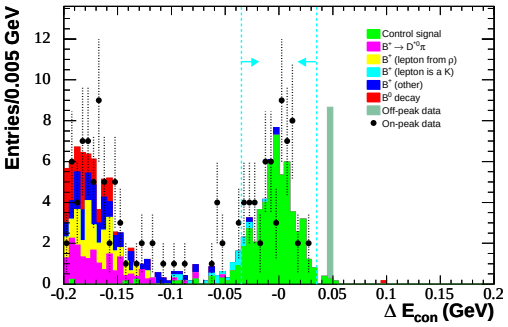
(a) S region (nuEPM signal, m_{ES} signal).



(b) B1 region (nuEPM signal, m_{ES} sideband).

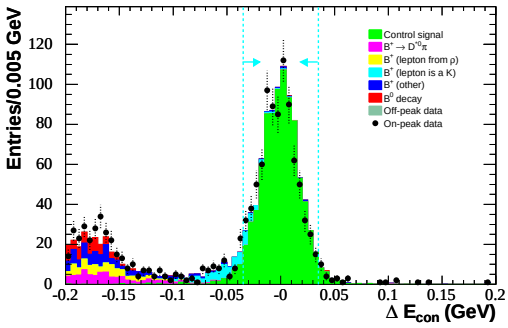


(c) B2 region (nuEPM sideband, m_{ES} signal).

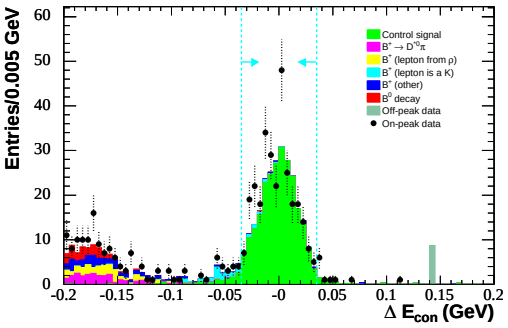


(d) B3 region (nuEPM sideband, m_{ES} sideband).

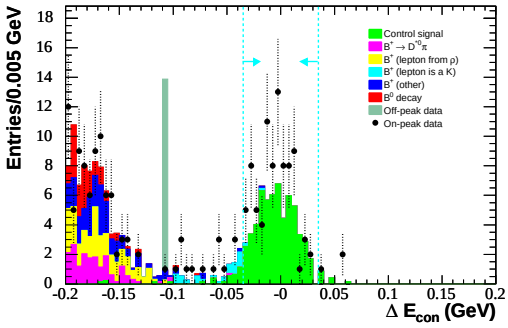
Figure 13.1. Control sample comparison of unsubtracted on-peak data and the hybrid sample, after all applicable electron-channel analysis cuts have been applied. The turquoise dotted lines indicate the ΔE_{con} signal region of -35 to 35 MeV. The regions of nuEPM and m_{ES} are defined in table 9.1.



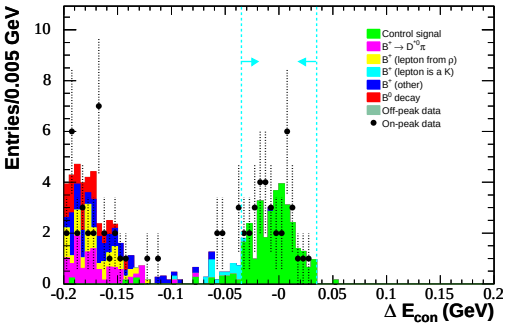
(a) S region (nuEPM signal, m_{ES} signal).



(b) B1 region (nuEPM signal, m_{ES} sideband).



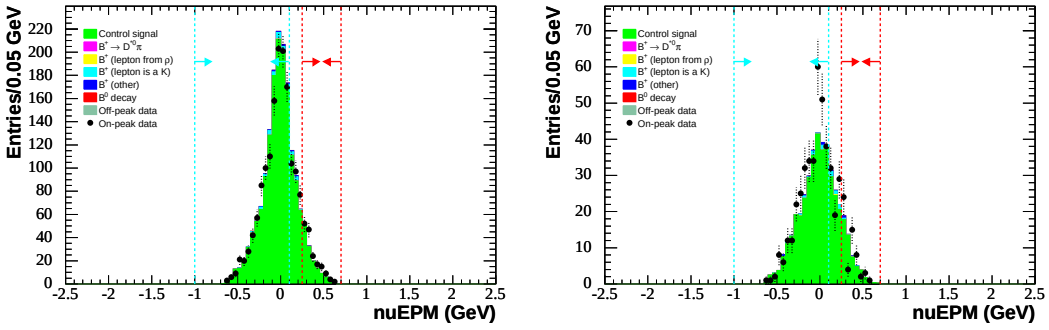
(c) B2 region (nuEPM sideband, m_{ES} signal).



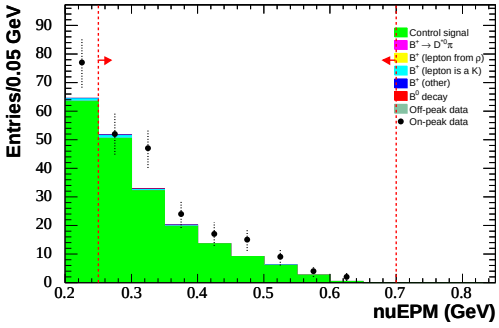
(d) B3 region (nuEPM sideband, m_{ES} sideband).

Figure 13.2. Control sample comparison of unsubtracted on-peak data and the hybrid sample, after all applicable muon-channel analysis cuts have been applied. The turquoise dotted lines indicate the ΔE_{con} signal region of -35 to 35 MeV. The regions of nuEPM and m_{ES} are defined in table 9.3.

Figures 13.3 and 13.4 show a comparison of the on-peak and hybrid distributions of nuEPM, in the $\Delta E_{\rm con}$ signal region, with the applicable electron- and muon-channel analysis cuts applied, respectively. Table 13.2 shows the event count discrepancies in the $\Delta E_{\rm con}$ region. The discrepancy in the B2 region appears rather large on a percentage basis. In figures 13.3(c) and 13.4(c) we have visually expanded the B2 region. It appears that the discrepancies arise from rather small changes in the underlying data, as the B2 region lies in the tail of the nuEPM distribution.

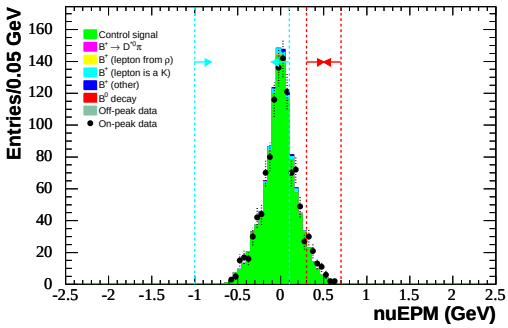


(a) S and B2 regions ($m_{\rm ES}$ signal). (b) B1 and B3 regions ($m_{\rm ES}$ sideband).

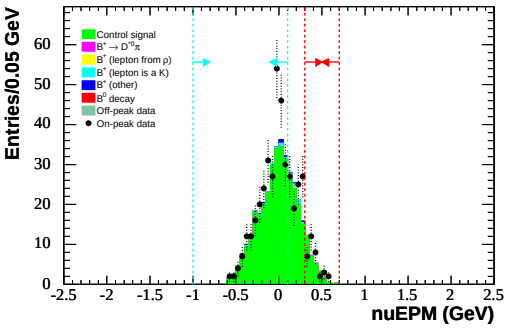


(c) B2 region ($m_{\rm ES}$ signal).

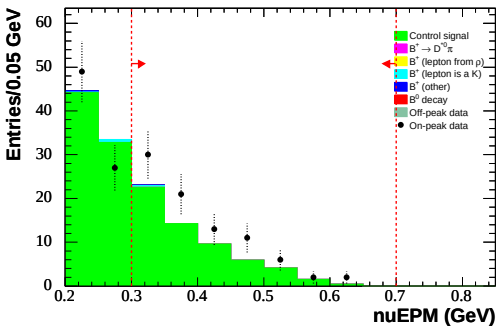
Figure 13.3. Control sample comparison of unsubtracted on-peak data and the hybrid sample, after all applicable electron-channel analysis cuts have been applied, along with a $\Delta E_{\rm con}$ cut of -35 to 35 MeV. The turquoise dotted lines bound the nuEPM signal region, and the red dotted lines bound the nuEPM sideband region. The B2 region is visually expanded in the lower histogram for illustrative purposes.



(a) S and B2 regions (m_{ES} signal).



(b) B1 and B3 regions (m_{ES} sideband).



(c) B2 region (m_{ES} signal).

Figure 13.4. Control sample comparison of unsubtracted on-peak data and the hybrid sample, after all applicable muon-channel analysis cuts have been applied, along with a ΔE_{con} cut of -35 to 35 MeV. The turquoise dotted lines bound the nuEPM signal region, and the red dotted lines bound the nuEPM sideband region. The B2 region is visually expanded in the lower histogram for illustrative purposes.

Table 13.2. Control sample comparison of on-peak data and hybrid sample (scaled to on-peak luminosities) event counts in the $\Delta E_{\rm con}$ signal region (-35 to 35 MeV).

	S	B1	B2	B3
Muon-Channel Analysis Cuts				
Control on-peak	837	287	85	34
Control hybrid	861	252	59	31
Data/hybrid ratio	97%	114%	144%	111%
Electron-Channel Analysis Cuts				
Control on-peak	1213	338	170	57
Control hybrid	1263	292	137	52
Data/hybrid ratio	96%	116%	124%	110%

Given that there appear to be data/MC discrepancies, there are two issues: (1) the applicability of this control sample study to our analysis, and (2) the evaluation of the shape systematic.

We believe that the discrepancies seen in the control sample study can be directly applied to the signal MC template, because the reconstruction of the signal and recoil sides is largely the same. For the other component MC samples, i.e. all other $B\overline{B}$ decays, in the fit, the applicability is not as clear-cut, because the recoil reconstruction (or misreconstruction, to be more specific) is different, and not in an unambiguous way because background events can be reconstructed as signal in a number of ways, especially since the recoil B decays generically. However, we feel that some degree of data/MC discrepancy in the $B\overline{B}$ background is already conservatively accounted for in our evaluation of the $B \to X_u \ell \nu_\ell$ systematic (section 13.5), and that applying the control sample discrepancies to the $B\overline{B}$ background should provide a fair (though not exact) modeling of background shape discrepancies.

Using the results of the control sample study, we evaluate the shape systematic in two different ways, taking the worst effect as the systematic. One method is to apply the within-region discrepancies from the control sample study to all of the MC templates. We do this by fluctuating the templates in the four fit regions on the scales given by the control sample studies ("random fluctuations study"). We perform 100 toy MC experiments for each type of fit (electron, muon, joint) in which we randomly fluctuate the template regions by choosing four (representing S, B1, B2, and B3) uniformly distributed random factors on the scale of the control sample discrepancies, simultaneously adjust all of the MC template regions by these factors, and then refit. We then take the 90th worst effect (in order to remove the effects of tails, which can be sensitive to the number of experiments performed).

The second method is to shift the nuEPM distribution ("shift study"). As can be seen in figures 13.3(c) and 13.4(c), a shift in the nuEPM distribution could explain the data/MC discrepancy. In the $m_{\rm ES}$ signal region, we shift the MC templates by 22 MeV, a value which was obtained by minimizing a pseudo-χ^2 statistic that simultaneously took into account the S and B2 region discrepancies. (A similar analysis of the B1 and B3 regions did not yield an unambiguous shift, and therefore we apply no shift there.) We apply the shift and obtain new MC templates, refitting with these templates to obtain the systematic effect.

The evaluated effects using the random fluctuations and shift studies are given in table 13.3. As can be seen, the effects are larger when using the random fluctuations study, and so we take these effects as the measure of our shape systematic.

Table 13.3. Results of the two different shape systematic studies, given in terms of a shift in the extracted signal BF, relative to the standard fit.

Mode	Shift Study	Random Fluctuations Study
Muon	0.80×10^{-6}	0.82×10^{-6}
Electron	0.35×10^{-6}	0.48×10^{-6}
Joint	0.47×10^{-6}	0.63×10^{-6}

13.3.2 Cut Efficiency

Because our shape systematic is evaluated after most cuts[8] are applied, we only need to evaluate cut efficiency systematics for the remaining cuts. In particular, we have the signal lepton CM energy and angle, the signal photon CM energy, the signal photon lab angle, and the signal photon shower shape lateral moment (gLAT). All photon- and muon-related systematics were evaluated using a radiative dimuon sample. All electron-related systematics were evaluated using a radiative Bhabha control sample. The evaluated multiplicative effects are given in table 13.4. As one can see, the dominant cut efficiency systematic is due to the lepton CM energy.

Table 13.4. Cut efficiency multiplicative systematics.

Cut	Muon	Electron
Signal photon lab angle	0.4%	0.4%
Signal photon CM energy	0.5%	0.5%
Signal photon LAT	0.0%	0.0%
Signal lepton CM energy	6.0%	5.0%
Signal lepton CM angle	0.0%	0.0%
Total	6.0%	5.0%

In order to calculate the systematic effect, we simultaneously adjust the templates

[8] "Most cuts" means that we apply any cuts which are applicable to the $B^+ \to \pi^+ \overline{D}^0 (\to K^+ \pi^-)$ control sample.

(all MC templates except generic B) both up and down. (In the joint fit, we apply the different electron- and muon-channel effects separately to their respective templates). After removing the implicit and incorrect multiplicative effect, we take the average of the additive effects as the additive systematic, and include the multiplicative effect in the convolution of the likelihood function. The evaluated additive and multiplicative effects are given in table 13.11, presented in section 13.7.

13.4 Uncertainties in the Number of B Mesons

13.4.1 Charged/Neutral B Ratio

We have assumed that charged and neutral B mesons are produced in equal proportions at the $\Upsilon(4S)$ resonance energy. We determine the systematic uncertainty by varying the proportion between the extremes of the interval specified by $f_{+-}/f_{00} = 1.020 \pm 0.034$ [29], the current published value for the charged-to-neutral B production ratio at the $\Upsilon(4S)$. Our assumed central value of one differs from the published central value, and so our variation is asymmetric.

Using reasoning similar to that of section 13.2.1, we apply the effect only to the signal and η templates (upwards and downwards). After removing the implicit and incorrect multiplicative effect, we take the worst of the additive effects as the additive systematic, and include the multiplicative effect in the convolution of the likelihood function. The evaluated additive and multiplicative effects are given in table 13.11, presented in section 13.7.

13.4.2 B counting

As mentioned earlier in section 5.1, there is a 1.1% uncertainty on the number of B pairs. The scale of each of our MC templates is dependent on the total number of B pairs, and therefore we must take this uncertainty into account. Using reasoning similar to that of section 13.2.1, we apply the 1.1% multiplicative effect only to the signal and η templates (upwards and downwards). After removing the implicit and incorrect multiplicative effect, we take the average of the additive effects as the additive systematic, and include the multiplicative effect in the convolution of the likelihood function. The evaluated additive and multiplicative effects are given in table 13.11, presented in section 13.7.

13.5 B-Background Uncertainties

13.5.1 $B \to X_u \ell \nu_\ell$

Because we do not have a good handle on the uncertainties in our modeling of $B \to X_u \ell \nu_\ell$ background in MC, we take a conservative approach to evaluating the systematic effect. We try the following three scenarios, taking the worst effect as the systematic:

1. Evaluate the effect of fixing the generic B factor, β_B to 1.

2. Evaluate the effect of setting the $B \to X_u \ell \nu_\ell$ contribution to the generic B template to 0.

3. Evaluate the effect of doubling the $B \to X_u \ell \nu_\ell$ contribution to the generic B template.

In the muon, electron, and joint fits, the worst effect, by far, is due to fixing the generic B factor to 1. What this means is that we assume that all of the excess generic B background that is not predicted by the MC (scalewise) is due to $B \to X_u \ell \nu_\ell$. The evaluated additive effects are given in table 13.11, presented in section 13.7.

13.5.2 Semileptonic π and ρ Modes

There are systematic uncertainties due to theoretical uncertainties on the branching fractions of the π and ρ modes. In addition, there is uncertainty in their form factor (q^2-dependent) models.

As in the signal extraction fit, we group the charged and neutral π modes together, and the charged and neutral ρ modes and the ω mode together. Within each group, we vary the branching fractions and/or the theoretical form factor models simultaneously.

Our determination of the theoretical systematic errors for these modes is largely based on the *BABAR* treatment from which we obtained the π and ρ mode measurements [28]. Besides the relation of the form factor models and the extracted BFs, an equally important issue is the crossfeed between the π and ρ modes.[9] In fact, the largest variation in the π mode BF is due to variation in the ρ mode form factor model. A naive scan of form factor models and BFs would very likely result in a larger systematic error than an intelligent scan that simultaneously took into account

[9] The *BABAR* treatment involves a simultaneous fit to the π and ρ modes so changes in the modeling of one mode will affect the extracted results of the other mode.

the relation between BFs and form factor models in both the π and ρ modes.

For the π modes, our default form factor model is FNAL04, based on lattice QCD results [25]. More specifically, it is a fit to the FNAL04 result, using a parameterization of Becirevic and Kaidalov (BK) [37]. This is necessary for extending lattice calculations to the low-q^2 (large hadron momentum) regime. For the π mode form factor of interest, the BK parameterization extracts a normalization factor and a factor, α. The value of α for the FNAL04 result is 0.63 ± 0.05. The associated branching fraction for $B^0 \to \pi^- e^+ \nu_e$ obtained by the *BABAR* analysts using this parameterization is 1.38×10^{-4}, the same as their actual measurement.

The actual *BABAR* measurement of the π mode branching fraction involved a fit based on the BK parameterization.[10] The fitted value of α for the *BABAR* measurement was 0.61 ± 0.09, with virtually the same central value as for FNAL04. We vary the π form factor model to use this central value (probably unnecessary) of α along with a model using a value one sigma higher (equal to 0.70). One could, in principal, modify the branching-fraction value downwards when adjusting the value of α upwards by one sigma, but we do not expect the scale of the change to be large, and therefore forgo this.

Another lattice result is HPQCD04 [38]. One can also perform a a BK-based fit to this result, obtaining a value of $\alpha = 0.41 \pm 0.07$.

We vary the π form factor model to use this central value of α along with a model using a value one sigma lower (equal to 0.34). The associated branching fraction for $B^0 \to \pi^- e^+ \nu_e$ obtained by the *BABAR* analysts using this parameterization is

[10] We will refer to the *BABAR* BK-based fit as the "BK fit."

1.40×10^{-4}. Once again, one could modify the branching-fraction value upwards when adjusting the value of α further downwards by one sigma, but we do not expect the scale of the change to be large (given that the branching fraction changes slightly even with a relatively large charge in α), and therefore forgo this. There is also a very small change in the ρ mode branching fraction due to crossfeed when using the LPQCD04 model for the π form factor, with the central value of the BF of $B^0 \to \rho^- e^+ \nu_e$ going from 2.14×10^{-4} to 2.13×10^{-4}.

For the ρ modes, we have two form factor models: BALL05 and ISGW2. The BALL05 model is based on light-cone sum rule calculations [26]. It is our default model. The ISGW2 model is a constituent quark model [32]. Switching from BALL05 to ISWG2 results in a relatively significant change in BFs, increasing the π BF from 1.38×10^{-4} to 1.46×10^{-4} (a 5.8% increase) and decreasing the ρ BF from 2.14×10^{-4} to 1.85×10^{-4} (a 13.6% decrease).

Ideally, a ρ form factor model would be available which would have the opposite effect of the ISGW2, relative to the BALL05 model, decreasing the π mode BF and increasing the ρ mode BF. Instead, we symmetrize the BF changes due to the switch to ISGW2, while using the BALL05 model. This results in a π mode BF of 1.30×10^{-4} and a ρ mode BF of 2.43×10^{-4}.

In our scan, we fix the π model and BF at the default values while varying the ρ model and BF, and vice versa. We switch to the ISGW2 model and resultant BF in the ρ mode, with the π mode BF also changing as described previously, while varying the π mode form-factor model. We also keep the BALL05 model and use

the symmetrized (ISGW2-determined magnitudes) ρ and π BFs, while changing the π mode form-factor model. We do not, however, simultaneously decrease or increase both the ρ and π BFs by the ISGW2-determined magnitudes, which would be done in a naive scan.

The form-factor models and BF values that we scan over are listed in table 13.5. In the muon, electron, and joint fits, the worst effect arises from the scenario in which the π model is varied such that the parameter α is one sigma higher than the central value obtained from the BK fit, as can be seen in table 13.6 (the bottommost row). This effect is additive, and we take it as our theoretical systematic for the π and ρ modes.

Table 13.5. BF and form-factor model scan variations for determining π and ρ mode theoretical systematics. Errors are provided in parentheses. The term "default" refers to the model used in the fit (chapter 12). The term "1 sigma" refers to variations in the value of α. All BF values have been multiplied by 10^4.

π Model	ρ Model	π BF$\times 10^4$	ρ BF$\times 10^4$
FNAL04 (default)	BALL05 (default)	1.38(10)	2.14(21)
BK fit	BALL05	1.38(10)	2.14(21)
HPQCD04	BALL05	1.40(10)	2.13(22)
FNAL04	ISGW2	1.46(10)	1.85(15)
HPQCD04 - 1 sigma	BALL05	1.40(10)	2.13(22)
BK fit + 1 sigma	BALL05	1.40(10)	2.14(21)
FNAL04	BALL05	1.38(10)	2.43(24)
FNAL04	BALL05	1.30(10)	2.43(24)
FNAL04	BALL05	1.30(10)	2.14(21)
BK fit	BALL05	1.30(10)	2.14(21)
BK fit + 1 sigma	BALL05	1.30(10)	2.14(21)

Table 13.6. Results of the π and ρ mode BF and form-factor model scan, given in terms of a shift in the extracted signal BF, relative to the standard fit. See table 13.5 for more explanation of the terms used in the table. All shift values have been multiplied by 10^6.

		BF Shift $\times 10^6$		
π Model	ρ Model	Muon	Electron	Joint
FNAL04 (default)	BALL05 (default)	0	0	0
BK fit	BALL05	−0.15	−0.16	−0.15
HPQCD04	BALL05	0.10	0.05	0.08
FNAL04	ISGW2	−0.28	−0.69	−0.38
HPQCD04 - 1 sigma	BALL05	−0.20	−0.28	−0.23
BK fit + 1 sigma	BALL05	0.43	0.48	0.45
FNAL04	BALL05	−0.44	−0.23	−0.31
FNAL04	BALL05	−0.08	0.15	0.05
FNAL04	BALL05	0.36	0.38	0.36
BK fit	BALL05	0.05	0.24	0.23
BK fit + 1 sigma	BALL05	0.82	0.88	0.83

13.5.3 Semileptonic η Modes

The strategy for determining the systematic error due to theoretical uncertainty in the BFs of the η and η' modes is to vary the η and η' mode BFs simultaneously by $\pm 10\%$ for each theoretical model. This prescription is based on the strategy used in the CLEO η mode measurement [30], where they determined the 10% error based on their π mode measurements. The resulting effect is additive, and we average the effect when varying the η BF up and down. The only model we have available for the η mode is BALL04, a light-cone sum rule model [27]. The evaluated additive effects are given in table 13.11, presented in section 13.7.

In addition, as mentioned in section 5.2.3, we use a theoretical factor of 2.057 [31] to relate the η and η' BFs. After consultation with the authors of the paper from which this factor was obtained, we use an uncertainty of ± 0.02 and scan over the

range of this factor. The resulting shift in the signal BF is negligible ($\mathcal{O}(10^{-10})$), and therefore we ignore this systematic.

13.6 Theoretical Uncertainties

As mentioned earlier in section 12.3, we convert our full BF into a partial BF because there is little theoretical guidance in determining the accuracy of the signal decay model over the entire signal kinematic region. Given that the partial BF is based on a greatly restricted kinematic region, we expect that its central value should be relatively insensitive to the choice of the signal decay model.

One possibility is to reweight standard[11] signal model MC events over the kinematic region so that the resulting distribution is indeed flat, and perform an analysis with the new signal template. Instead, we use the fa0 model.

The first step in this evaluation is to extract the full signal BF using the fa0 signal template in place of the standard signal template. Table 13.7 shows a comparison of the templates for the electron and muon channels. Table 13.8 shows the extracted signal BF results based on using the fa0 signal templates. As one would expect, the results have less statistical power given the lower signal efficiency.

The next step in evaluating the systematic is to calculate the conversion factor. The method is exactly the same as in the standard case, except that we use the fa0 sample. The fa0-based conversion factors (as well as those from the standard fit) for

[11] As mentioned previously in section 4.1, we use the term "standard" to refer to the standard signal model used in our analysis (as opposed to the fa0 model). We also use "standard" to refer to refer to the standard fit (as opposed to fits used in systematic studies in which templates are modified).

Table 13.7. Comparison of the standard and fa0 signal model region event counts scaled to on-peak luminosities, for the electron and muon channels.

	Counts			
Sample	S	$B1$	$B2$	$B3$
Electron standard signal MC	21.19	3.74	3.15	0.94
Electron fa0 signal MC	11.22	2.00	1.73	0.55
fa0/standard ratio	0.53	0.53	0.55	0.59
Muon standard signal MC	13.12	3.27	0.96	0.48
Muon fa0 signal MC	7.28	1.83	0.50	0.26
fa0/standard ratio	0.55	0.56	0.52	0.54

Table 13.8. The extracted signal BF central value and parabolic error when using the fa0 signal model, relative to the standard fits. "Relative" is defined such that the standard fit result is subtracted off.

Mode	Fitted Value	Parabolic Error
Muon standard fit value	-3.24×10^{-6}	4.80×10^{-6}
(fa0 relative value)	(-2.67×10^{-6})	(3.87×10^{-6})
Electron standard fit value	0.18×10^{-6}	3.96×10^{-6}
(fa0 relative value)	(0.11×10^{-6})	(3.54×10^{-6})
Joint standard fit value	-1.19×10^{-6}	3.16×10^{-6}
(fa0 relative value)	(-1.13×10^{-6})	(2.69×10^{-6})

the original and envelope regions are given in table 13.9.

We then calculate the fa0-based value of $\Delta\mathcal{B}$ and compare it to the value of $\Delta\mathcal{B}$ obtained using the standard fit. In order to determine the systematic uncertainty, one cannot simply take the difference of their values because there are multiplicative effects that need to be accounted for. Instead, we calculate an adjusted conversion factor, which is defined as the conversion factor that would take the BF extracted from the standard fit to the fa0-based $\Delta\mathcal{B}$ value. The difference between the adjusted conversion factor and the conversion factor obtained from the standard analysis is then treated as a multiplicative systematic, applied to a likelihood-weighted value of

Table 13.9. Conversion factors, ϵ, used to obtain partial BFs, for the standard and fa0 fits. The definitions of the kinematic regions are given in table 12.18.

Signal Template	Kinematic Region	Muon ϵ	Electron ϵ	Joint ϵ
Standard	Original	0.40	0.41	–
Standard	Envelope	0.40	0.48	0.45
fa0	Original	0.22	0.23	–
fa0	Envelope	0.23	0.27	0.26

$\Delta\mathcal{B}$. These results are summarized in table 13.10. As one can see, the systematic is extremely small.

Table 13.10. Theoretical systematics on $\Delta\mathcal{B}$, in absolute terms.

Kinematic Region	Muon	Electron	Joint
Original	0.03×10^{-6}	0.08×10^{-6}	-
Envelope	0.03×10^{-6}	0.08×10^{-6}	0.07×10^{-6}

13.7 Systematics Results

The summary of the multiplicative and additive systematics are given in table 13.11, in terms of the total BF. We do not list the theoretical systematic in this table because this systematic is meaningful only in terms of $\Delta\mathcal{B}$.

Table 13.11. Additive and multiplicative systematic uncertainties for the muon, electron, and joint fits, in terms of the total BF. All additive systematic values have been multiplied by 10^6. The values in parentheses are multiplicative systematics, with their effects evaluated separately (signs are given to indicate the direction of correlation with the associated additive systematic).

Systematic	Muon	Electron	Joint
Tracking efficiency	0.082(-1.3%)	0.061(-1.3%)	0.060(-1.3%)
Particle ID	0.111(3.5%)	0.022(2.2%)	0.004(2.1%)
Neutral reconstruction	0.027(1.58%)	0.080(1.58%)	0.019(1.58%)
B counting	0.034(-1.1%)	0.011(-1.1%)	0.008(-1.1%)
Charged to neutral B ratio	0.322(-9.4%)	0.040(-9.4%)	0.126(-9.4%)
Shape	0.745	0.451	0.606
Selection efficiency	0.109(6.0%)	0.275(5.0%)	0.236(6.0%)
η mode BF	0.689	0.306	0.434
π,ρ mode theory	0.816	0.876	0.833
$B \to X_u \ell^+ \nu_\ell$ BF	1.120	0.478	0.650
Total (+)	+1.74	+1.28	+1.34
Total (-)	−2.16	−1.23	−1.42

Chapter 14

Conclusions

14.1 Full Results

The two-sided results in terms of $\mathcal{B}$, the signal BF, are summarized in table 14.1. The Bayesian 90% upper limits (not including the uncertainty on the signal MC decay model) are summarized in table 14.2. Compared with the CLEO measurements given in section 1.6.2 (2.0×10^{-4} and 5.2×10^{-5} for the electron and muon channels, respectively), we have improved upon the upper limits by better than an order of magnitude. We are also in the realm of the standard model prediction of 2.0×10^{-6} to 5.0×10^{-6}.

The two-sided results in terms of $\Delta\mathcal{B}$ for the original and envelope kinematic regions are summarized in table 14.3. The Bayesian 90% upper limits (including the uncertainty on the signal MC decay model) are summarized in table 14.4.

Table 14.1. Comparison of signal $\mathcal{B}$ two-sided results for all three fits. All values have been multiplied by 10^6. The theoretical uncertainty on the signal MC decay model is *not* included.

	Central Value	Statistical Uncertainty	Systematic Uncertainty	Combined Uncertainty
Muon	-3.32	$+4.33$ -5.50	$+1.74$ -2.16	$+4.67$ -5.91
Electron	0.22	$+3.61$ -4.44	$+1.28$ -1.23	$+3.83$ -4.61
Joint	-0.56	$+2.97$ -3.43	$+1.34$ -1.42	$+3.26$ -3.71

Table 14.2. Comparison of the 90% Bayesian upper-limit results for all three fits, for the two different choices of prior, in terms of the full BF, $\mathcal{B}$. The theoretical uncertainty on the signal MC decay model is *not* included in the calculation.

	Prior	
	Flat in BF	Flat in Amplitude
Muon	$<5.24 \times 10^{-6}$	$<3.66 \times 10^{-6}$
Electron	$<5.92 \times 10^{-6}$	$<4.53 \times 10^{-6}$
Joint	$<5.01 \times 10^{-6}$	$<3.81 \times 10^{-6}$

14.2 Adjusted Results

As mentioned in section 4.1, all full BF results were determined without correcting for the generator-level restriction of 350 MeV on the minimum signal photon energy. We must adjust the full BF results to correct for this restriction.

In the tree-level model of KPY [11], we take the signal photon energy spectrum in the signal B rest frame to be symmetric about 1.32 GeV, with a minimum of 0 GeV and a maximum of 2.64 GeV. With these assumptions, the correction factor on the full BF is found to be 1.0515. The two-sided results in terms of the adjusted full signal BF are summarized in table 14.5. The adjusted Bayesian 90% upper limits

Table 14.3. Comparison of signal $\Delta\mathcal{B}$ two-sided results for all three fits. All values have been multiplied by 10^6. The definitions of the kinematic regions are given in table 12.18. "Original" refers to the original kinematic regions. "Envelope" refers to the envelope of the electron and muon kinematic regions.

	Central Value	Statistical Uncertainty	Systematic Uncertainty	Theoretical Uncertainty
Original				
Muon	−1.32	$^{+1.73}_{-2.19}$	$^{+0.69}_{-0.86}$	0.03
Electron	0.09	$^{+1.48}_{-1.82}$	$^{+0.52}_{-0.50}$	0.08
Envelope				
Muon	−1.33	$^{+1.74}_{-2.20}$	$^{+0.80}_{-0.87}$	0.03
Electron	0.11	$^{+1.73}_{-2.13}$	$^{+0.61}_{-0.59}$	0.08
Joint	−0.25	$^{+1.33}_{-1.53}$	$^{+0.60}_{-0.64}$	0.07

(not including the uncertainty on the signal MC decay model) are summarized in table 14.6.

14.3 Theoretical Implications

Using the 90% Bayesian upper limit values on the adjusted full BF (thus ignoring any theoretical uncertainty), we can set lower limits on λ_B using equation (1.26). We take m_B to be 5.279 GeV and τ_B to be 1.638 ps. In addition, we use the input values of $f_B = 216\,\text{MeV}$ [39], $|V_{ub}| = 4.31 \times 10^{-3}$ [29], and $m_b = 4.70\,\text{GeV}$ [29]. The results are summarized in table 14.7. As one can see, the limits on λ_B are still within range of λ_{QCD}.

Table 14.4. Comparison of the 90% Bayesian upper-limit results for all three fits, for the two different choices of prior, in terms of the partial BF, $\Delta\mathcal{B}$. The definitions of the kinematic regions are given in table 12.18. "Original" refers to the original kinematic regions. "Envelope" refers to the envelope of the electron and muon kinematic regions.

	Prior	
	Flat in BF	Flat in Amplitude
Original		
Muon	$<2.09 \times 10^{-6}$	$<1.46 \times 10^{-6}$
Electron	$<2.43 \times 10^{-6}$	$<1.86 \times 10^{-6}$
Envelope		
Muon	$<2.10 \times 10^{-6}$	$<1.47 \times 10^{-6}$
Electron	$<2.84 \times 10^{-6}$	$<2.18 \times 10^{-6}$
Joint	$<2.25 \times 10^{-6}$	$<1.71 \times 10^{-6}$

Table 14.5. Comparison of the adjusted full BF two-sided results for all three fits. All values have been multiplied by 10^6. The theoretical uncertainty on the signal MC decay model is *not* included.

	Central Value	Statistical Uncertainty	Systematic Uncertainty	Combined Uncertainty
Muon	-3.49	$+4.56$ -5.79	$+1.83$ -2.27	$+4.91$ -6.22
Electron	0.23	$+3.79$ -4.67	$+1.35$ -1.29	$+4.03$ -4.85
Joint	-0.59	$+3.13$ -3.60	$+1.41$ -1.49	$+3.43$ -3.90

14.4 Outlook

From the standpoint of *BABAR* data, one can perform this analysis again using the anticipated final dataset of $1\,\mathrm{ab}^{-1}$, if not earlier. One could design an analysis around fitting for not just the radiative leptonic channel, but also the semileptonic channels (as opposed to constraining their BFs as we do). This would require a redesigned multivariate (probably nonlinear) analysis over a larger region of phase space, given

Table 14.6. Comparison of the 90% Bayesian upper-limit results for all three fits, for the two different choices of prior, in terms of the adjusted full BF. The theoretical uncertainty on the signal MC decay model is *not* included in the calculation.

	Prior	
	Flat in BF	Flat in Amplitude
Muon	$<5.51 \times 10^{-6}$	$<3.85 \times 10^{-6}$
Electron	$<6.22 \times 10^{-6}$	$<4.77 \times 10^{-6}$
Joint	$<5.27 \times 10^{-6}$	$<4.00 \times 10^{-6}$

Table 14.7. Various calculated lower-limit constraints on λ_B, corresponding to the various measured upper-limit values in table 14.6.

UL Type	Limit on λ_B
Muon (flat BF)	>527
Muon (flat amp.)	>638
Electron (flat BF)	>494
Electron (flat amp.)	>569
Joint (flat BF)	>540
Joint (flat amp.)	>625

that our present analysis is designed to remove as much semileptonic background as possible through optimized selection criteria. For example, our analysis procedure unconditionally vetoes π^0s, and an analysis involving a simultaneous fit would need to relax this.

However, the preference might be to concentrate efforts on a tagged analysis, given that generic continuum backgrounds are much more manageable.

Ultimately, with enough signal events, one would like to measure not just the BF, but the full signal lepton and photon energy spectra (as well as the lepton-photon opening angle), as these would provide guidance in understanding the dynamics of these decays.

Bibliography

[1] F. Halzen and A. Martin, *Quarks & Leptons*, (John Wiley & Sons, New York 1984).

[2] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory*, (Westview Press, Boulder 1995).

[3] J. Donoghue, E. Golowich, and B. Holstein, *Dynamics of the Standard Model*, (Cambridge University Press, Cambridge 1992).

[4] G. Branco, L. Lavoura, and J. Silva, *CP Violation*, (Oxford University Press, New York 1999).

[5] J. M. Flynn, in *Proceedings of the 28th international conference on high-energy physics, Warsaw, Poland, 25-31 July 1996,* (World Scientific, Singapore 1997).

[6] M. Neubert, Phys. Rev. D **45**, 2451 (1992).

[7] C. R. Allton, C. T. Sachrajda, V. Lubicz, L. Maiani, and G. Martinelli, Nucl. Phys. B **349**, 598 (1991).

[8] C. Alexandrou, F. Jegerlehner, S. Gusken, K. Schilling, and R. Sommer, Phys. Lett. B **256**, 60 (1991).

[9] T. E. Browder et al. [CLEO Collaboration], Phys. Rev. D **56**, 11 (1997).

[10] K. Ikado et al. [Belle Collaboration], Phys. Rev. Lett. **97**, 251802 (2006) [arXiv:hep-ex/0604018].

[11] G. P. Korchemsky, D. Pirjol, and T. M. Yan, Phys. Rev. D **61**, 114510 (2000).

[12] S. Descotes-Genon and C. T. Sachrajda, Nucl. Phys. B **650**, 356 (2003).

[13] M. Beneke, G. Buchalla, M. Neubert, and C. T. Sachrajda, Phys. Rev. Lett. **83**, 1914 (1999).

[14] K. Abe et al. [Belle Collaboration], arXiv:hep-ex/0408132.

[15] B. Aubert et al. [BaBar Collaboration], Nucl. Instrum. Methods A **479**, 1 (2002).

[16] The BaBaR Collaboration, The BaBar Physics Book: Physics At An Asymmetric B Factory, edited by P. F. Harrison and H. R. Quinn, Report No. SLAC-R-0504, 1998.

[17] P. Golonka, B. Kersevan, T. Pierzchala, E. Richter-Was, Z. Was, and M. Worek, Comput. Phys. Commun. **174**, 818 (2006) [arXiv:hep-ph/0312240].

[18] S. Agostinelli et al. [GEANT4 Collaboration], Nucl. Instrum. Meth. A **506**, 250 (2003).

[19] D. Lange, Nucl. Instrum. Methods A **462**, 152 (2001).

[20] T. Sjostrand, S. Mrenna, and P. Skands, JHEP **0605**, 026 (2006) [arXiv:hep-ph/0603175].

[21] S. Jadach and Z. Was, Comput. Phys. Commun. **85**, 453 (1995).

[22] S. Jadach, B. F. L. Ward, and Z. Was, Comput. Phys. Commun. **130**, 260 (2000) [arXiv:hep-ph/9912214].

[23] S. Jadach, W. Placzek, and B. Ward, Phys. Lett. B **390**, 298 (1997).

[24] F. Berends, P. Daverfeldt, and R. Kleiss, Nucl. Phys. B **253**, 441 (1985).

[25] M. Okamoto et al., Nucl. Phys. Proc. Suppl. **140**, 461 (2005) [arXiv:hep-lat/0409116].

[26] P. Ball and R. Zwicky, Phys. Rev. D **71**, 014029 (2005).

[27] P. Ball and R. Zwicky, Phys. Rev. D **71**, 014015 (2005).

[28] B. Aubert et al. [BaBaR Collaboration], Phys. Rev. D **72**, 051102 (2005).

[29] W. M. Yao et al. [Particle Data Group], J. Phys. G **33**, 1 (2006).

[30] S. B. Athar et al. [CLEO Collaboration], Phys. Rev. D **68**, 072003 (2003).

[31] C. S. Kim and Y. D. Yang, Phys. Rev. D **65**, 017501 (2002) [arXiv:hep-ph/0107226].

[32] N. Isgur and M. B. Wise, Phys. Rev. D **43**, 819 (1991); D. Scora and N. Isgur, Phys. Rev. D **52**, 2783 (1995) [arXiv:hep-ph/9503486].

[33] F. De Fazio and M. Neubert, JHEP **9906**, 017 (1999) [arXiv:hep-ph/9905351].

[34] A. Drescher it et al., Nucl. Instrum. Meth. A **237**, 464 (1985).

[35] G. C. Fox and S. Wolfram, Phys. Rev. Lett. **41**, 1581 (1978).

[36] F. James, Comput. Phys. Commun. **20**, 29 (1980).

[37] D. Becirevic and A. B. Kaidalov, Phys. Lett. B **478**, 417 (2000).

[38] J. Shigemitsu et al., Nucl. Phys. Proc. Suppl. **140**, 464 (2005) [arXiv:hep-lat/0408019].

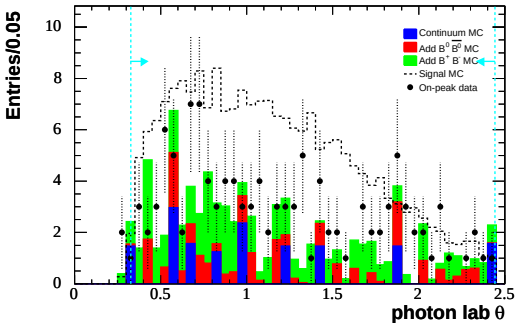
[39] A. Gray et al. [HPQCD Collaboration], Phys. Rev. Lett. **95**, 212001 (2005).

Appendix A

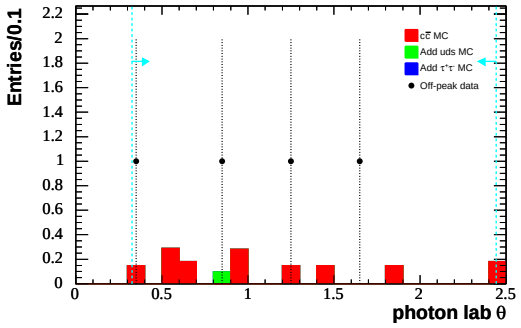
Marginal Cut Distributions

A.1 Electron Channel

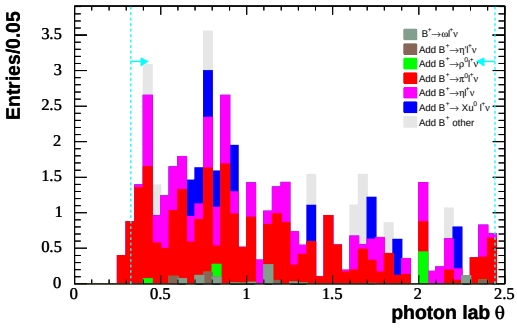
In this appendix we provide the marginal cut variable distributions for the electron channel, validation sample, as well as the distributions of other variables of interest. The m_{ES} and nuEP cuts are the optimized signal region cuts given in table 9.1. The rest of the cut values are given in table 8.2.



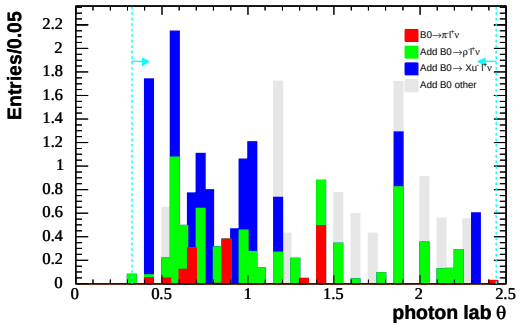
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



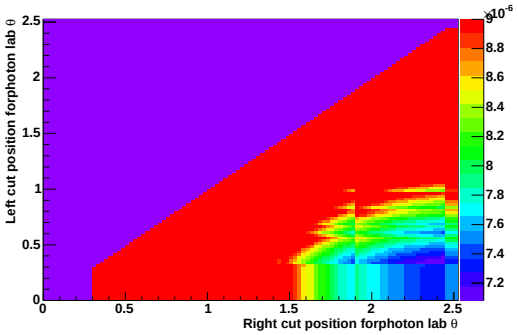
(b) Continuum marginal distributions.



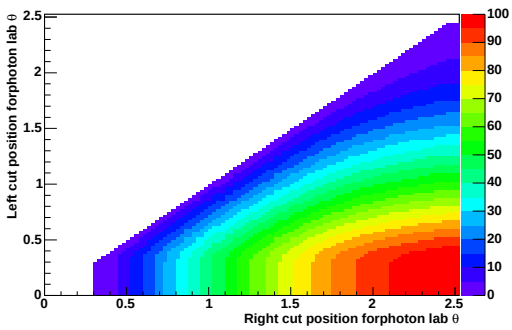
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

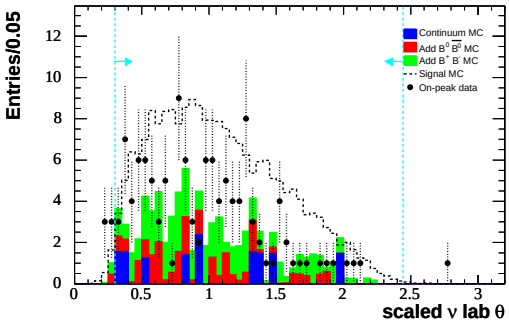


(e) FOM distribution.

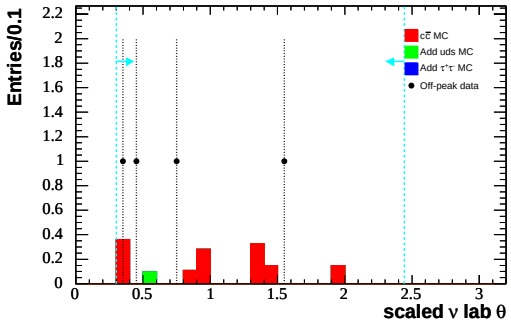


(f) Efficiency distribution.

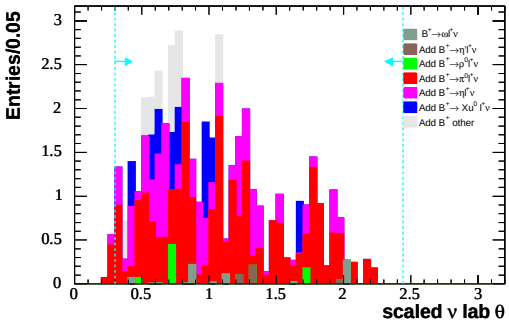
Figure A.1. Distributions of the photon lab θ in the signal region for the electron channel, validation sample.



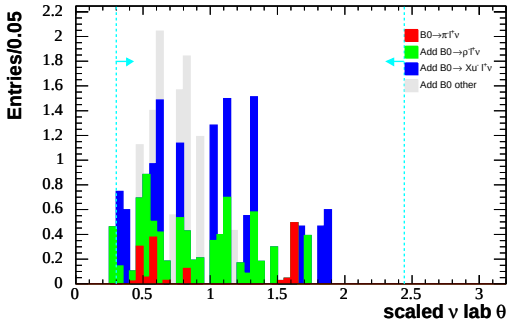
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



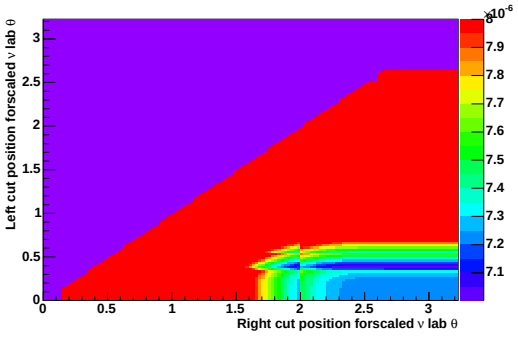
(b) Continuum marginal distributions.



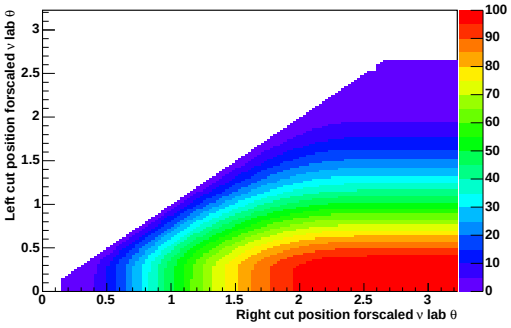
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

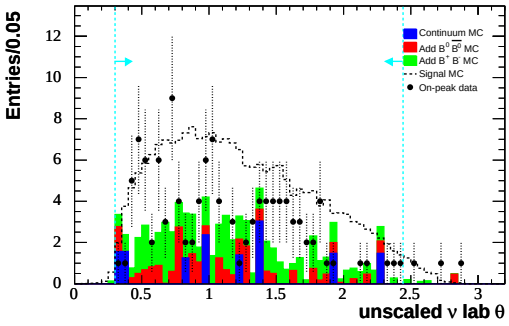


(e) FOM distribution.

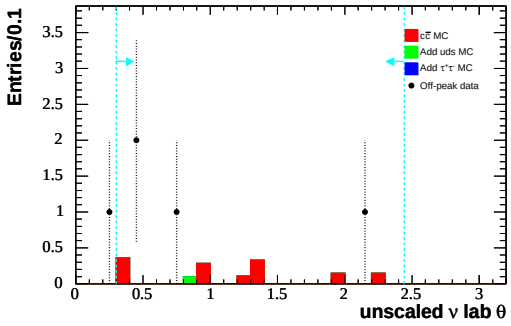


(f) Efficiency distribution.

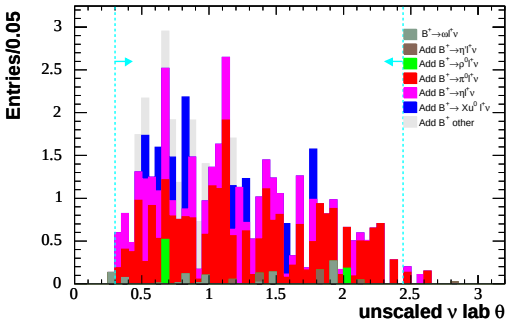
Figure A.2. Distributions of the scaled neutrino lab θ in the signal region for the electron channel, validation sample.



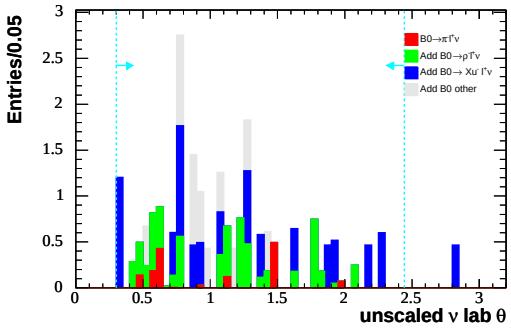
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



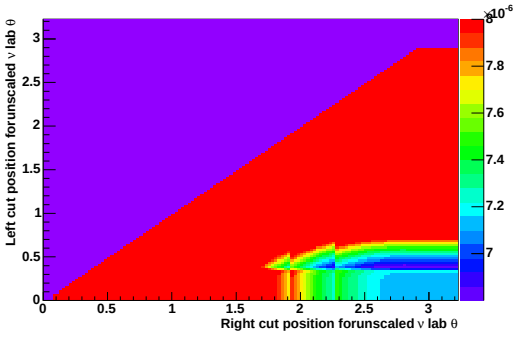
(b) Continuum marginal distributions.



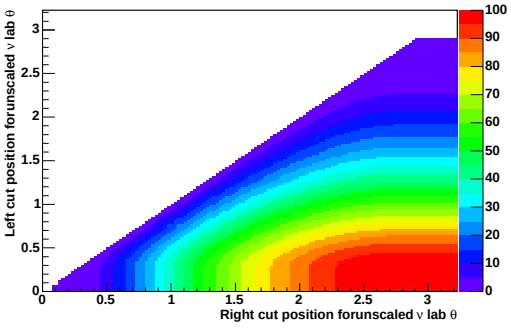
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

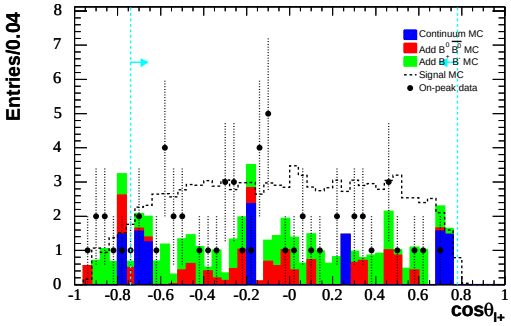


(e) FOM distribution.

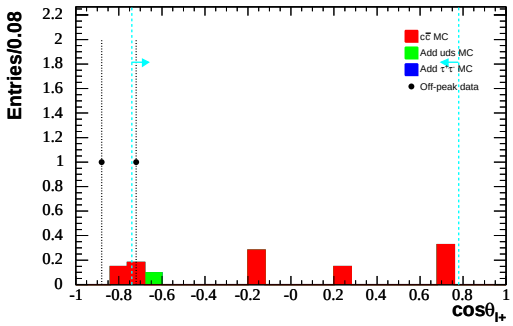


(f) Efficiency distribution.

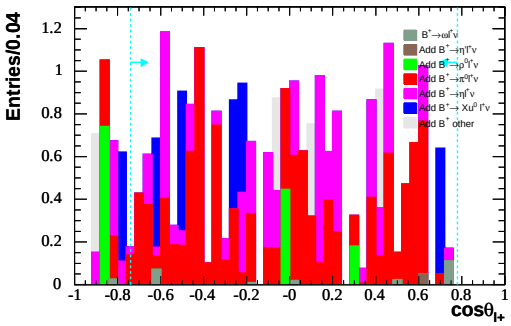
Figure A.3. Distributions of the unscaled neutrino lab θ in the signal region for the electron channel, validation sample.



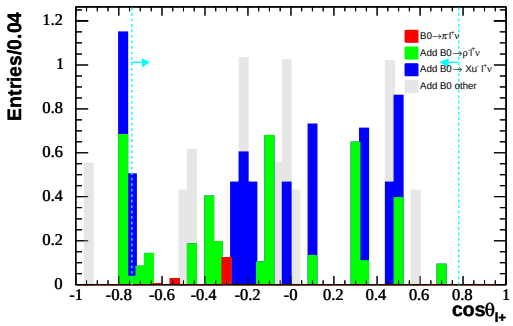
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



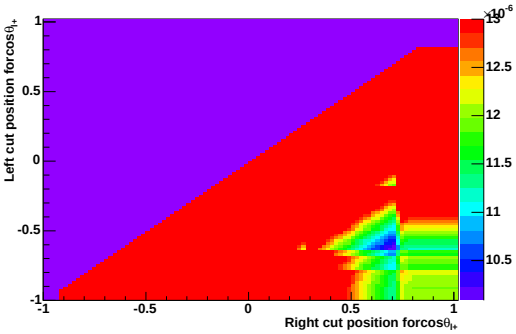
(b) Continuum marginal distributions.



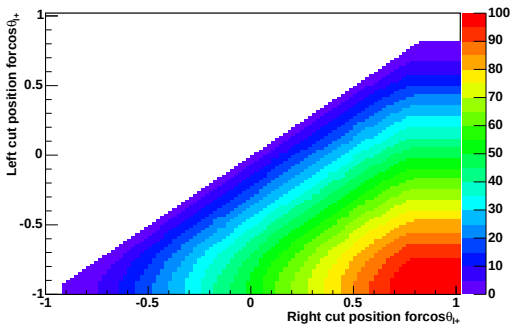
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

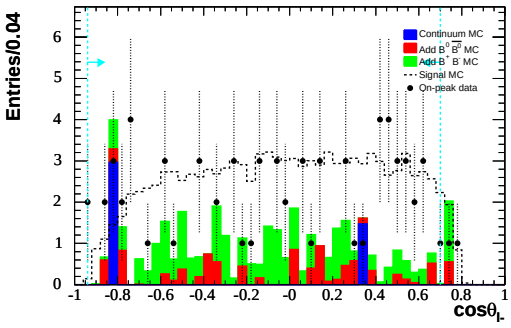


(e) FOM distribution.

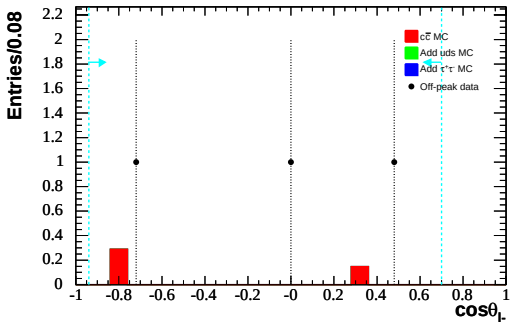


(f) Efficiency distribution.

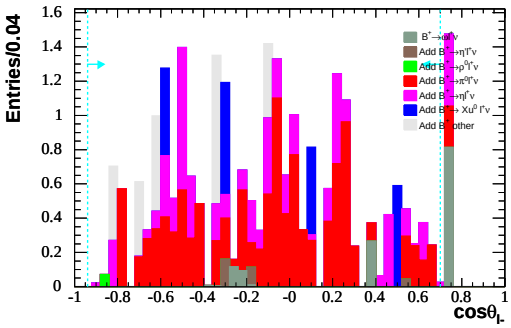
Figure A.4. Distributions of e^+: $\cos(\theta_{\mathrm{Lab}})$ in the signal region for the electron channel, validation sample.



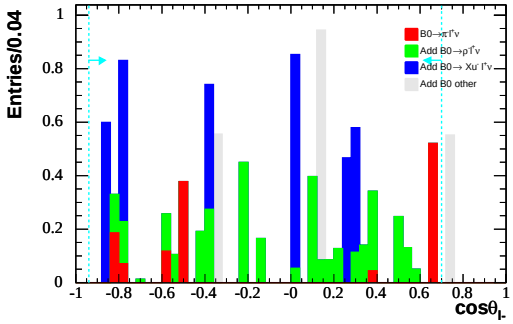
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



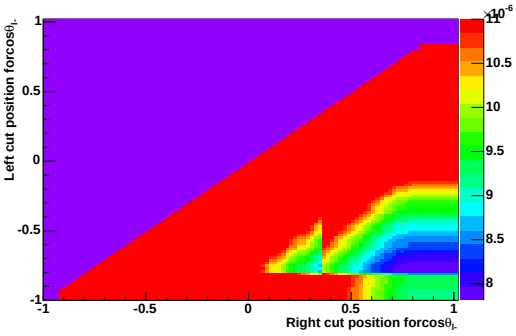
(b) Continuum marginal distributions.



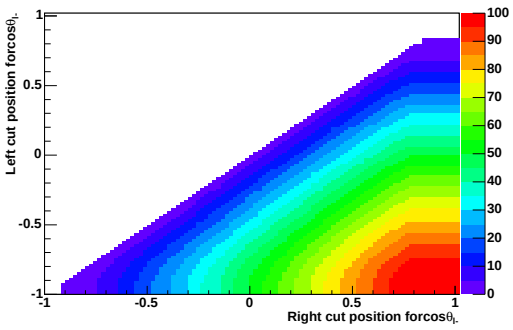
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

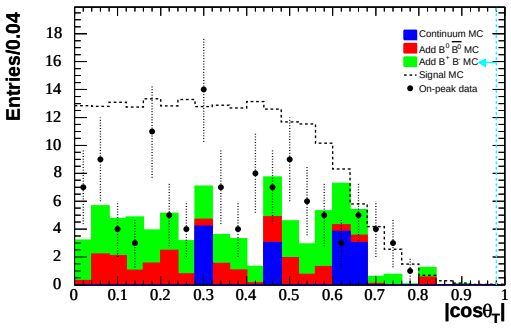


(e) FOM distribution.

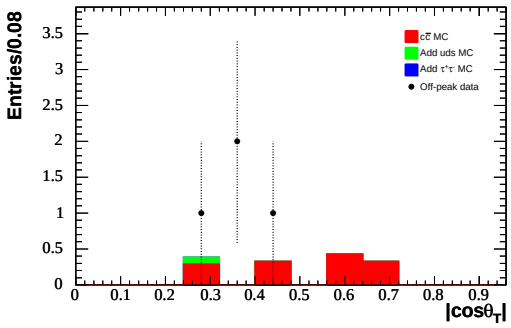


(f) Efficiency distribution.

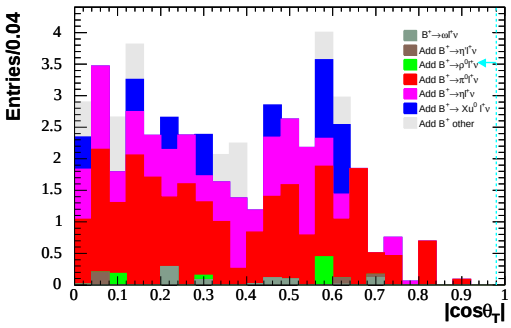
Figure A.5. Distributions of e^-: $\cos(\theta_{\mathrm{Lab}})$ in the signal region for the electron channel, validation sample.



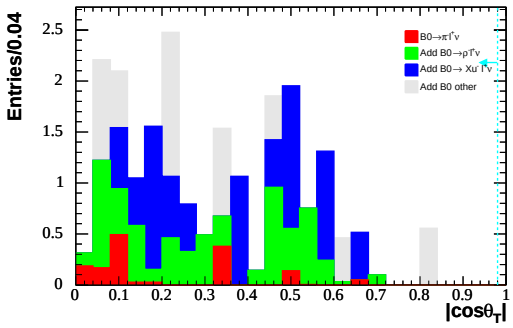
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



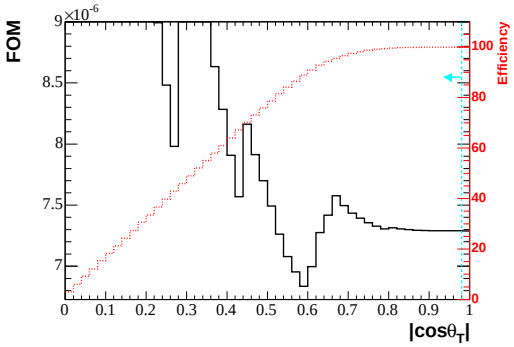
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

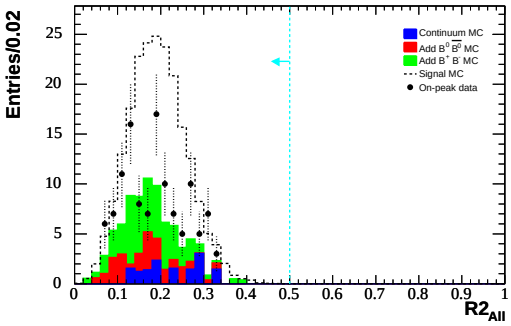


(d) Generic $B^0\overline{B}^0$ marginal distributions.

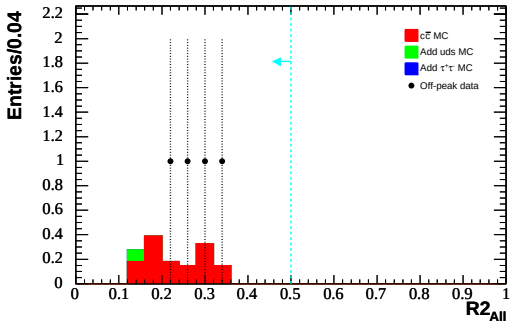


(e) FOM and efficiency distributions.

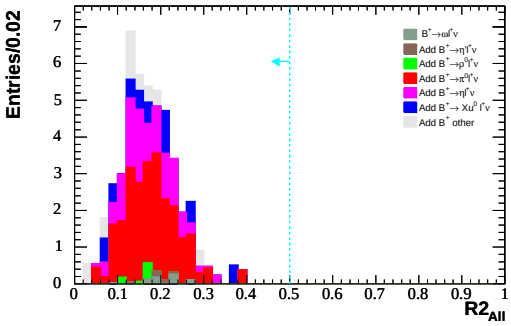
Figure A.6. Distributions of $|\cos\theta_T|$ in the signal region for the electron channel, validation sample.



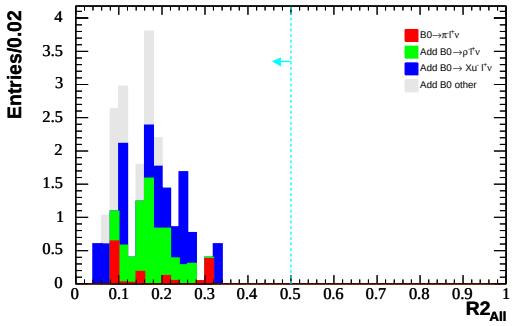
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



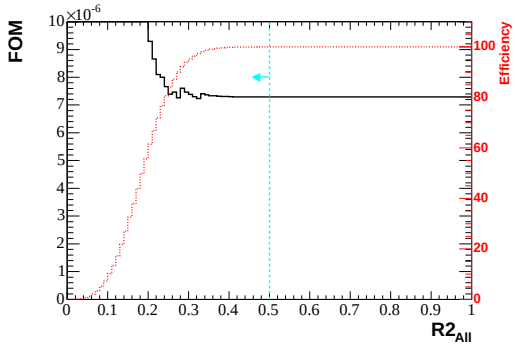
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

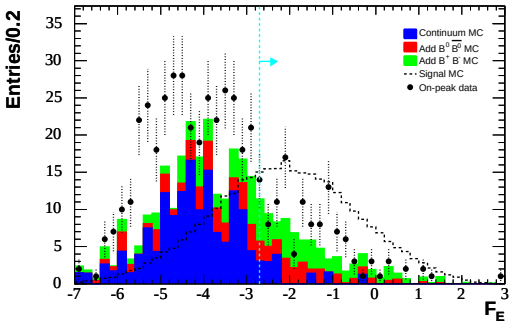


(d) Generic $B^0\overline{B}^0$ marginal distributions.

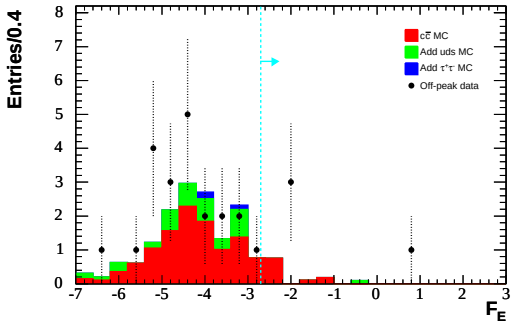


(e) FOM and efficiency distributions.

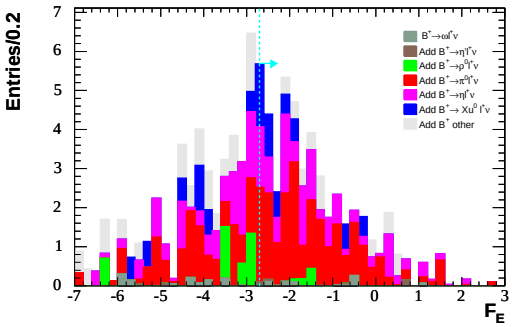
Figure A.7. Distributions of $R2_{\mathrm{All}}$ in the signal region for the electron channel, validation sample.



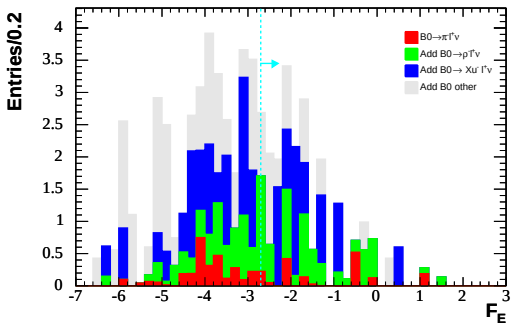
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



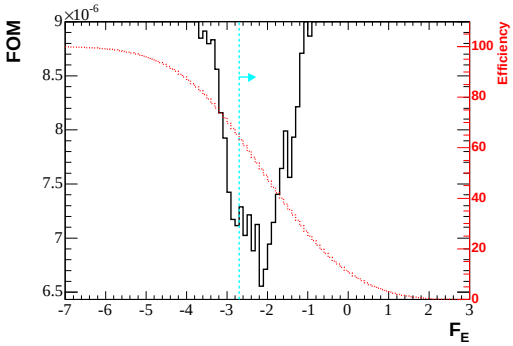
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

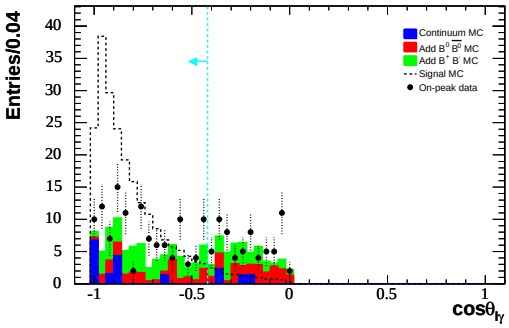


(d) Generic $B^0\overline{B}^0$ marginal distributions.

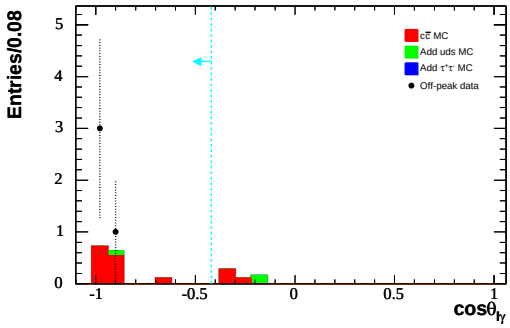


(e) FOM and efficiency distributions.

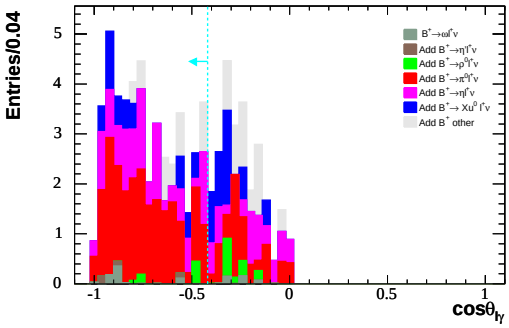
Figure A.8. Distributions of $\mathcal{F}_{\rm E}$ in the signal region for the electron channel, validation sample.



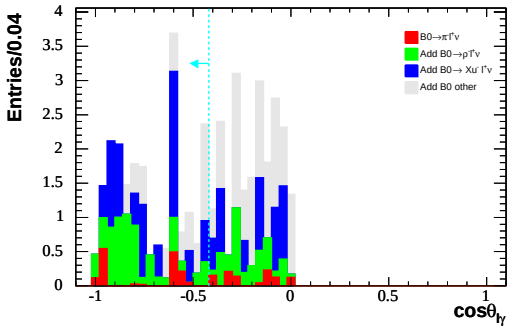
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



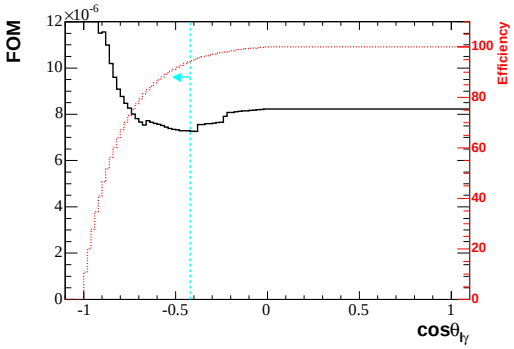
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

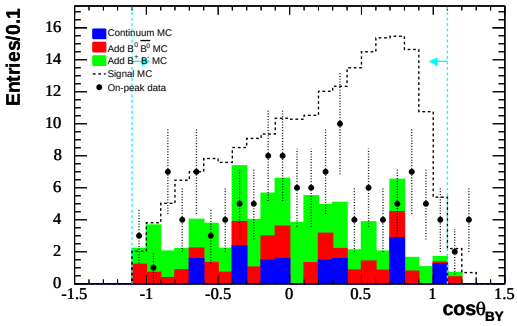


(d) Generic $B^0\overline{B}^0$ marginal distributions.

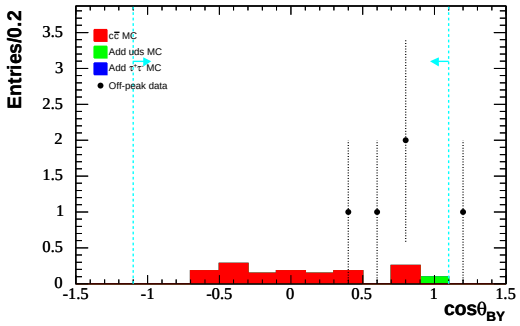


(e) FOM and efficiency distributions.

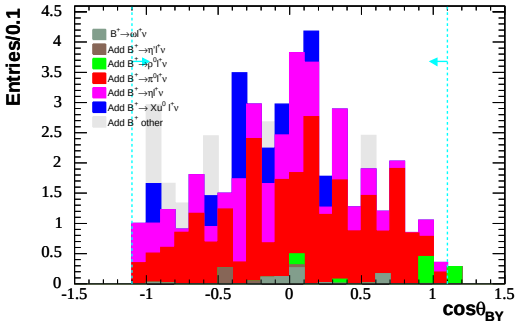
Figure A.9. Distributions of $\cos\theta_{\ell\gamma}$ in the signal region for the electron channel, validation sample.



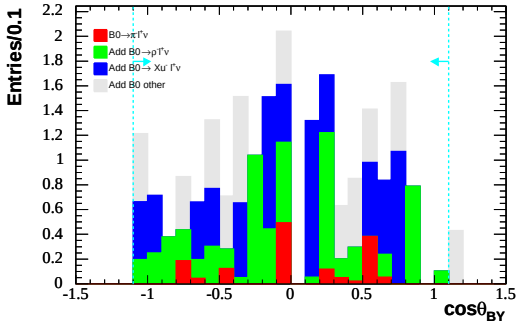
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



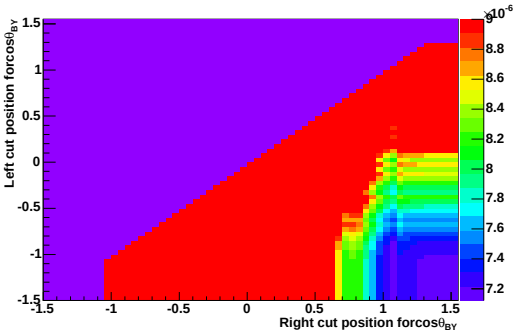
(b) Continuum marginal distributions.



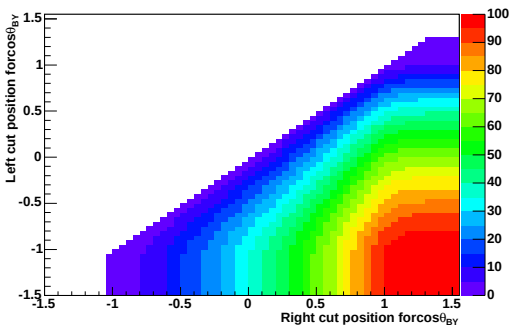
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

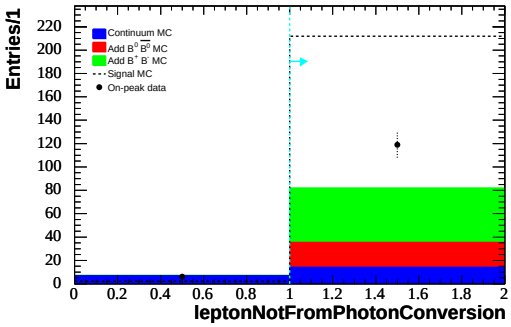


(e) FOM distribution.

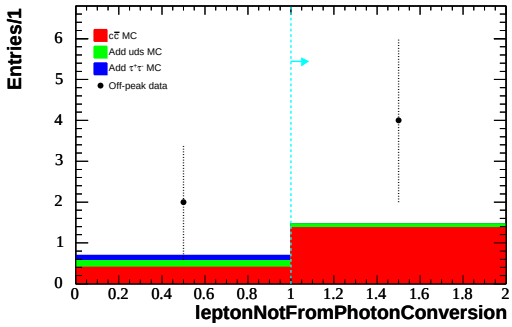


(f) Efficiency distribution.

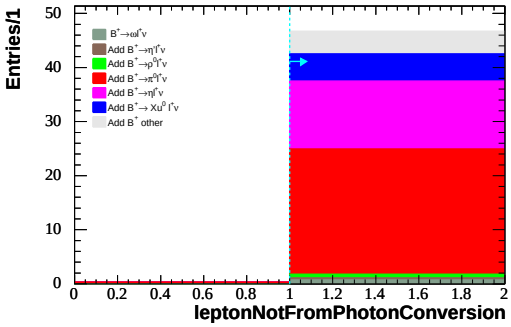
Figure A.10. Distributions of $\cos\theta_{BY}$ in the signal region for the electron channel, validation sample.



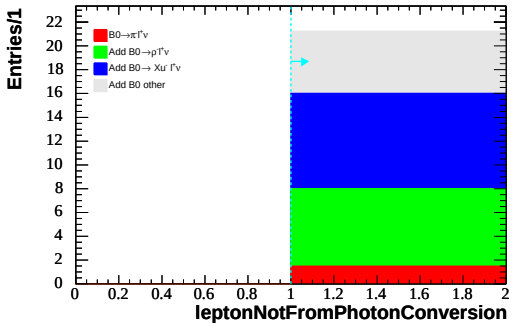
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



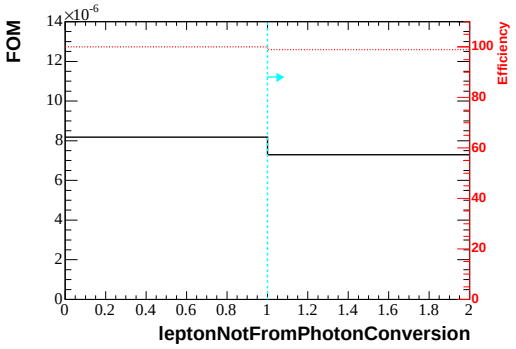
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

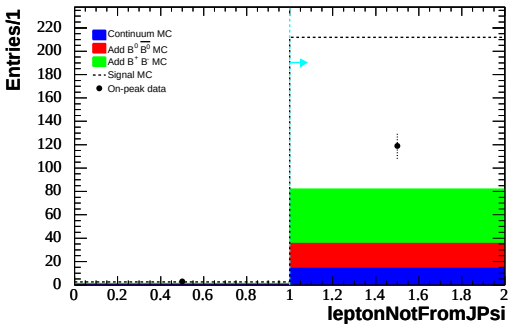


(d) Generic $B^0\overline{B}^0$ marginal distributions.

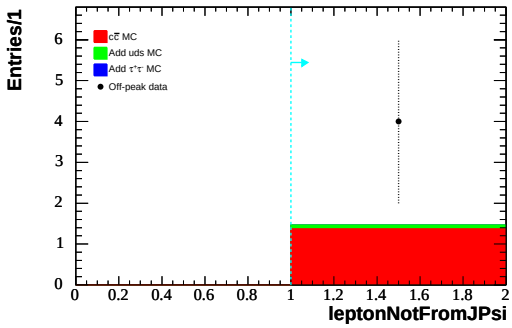


(e) FOM and efficiency distributions.

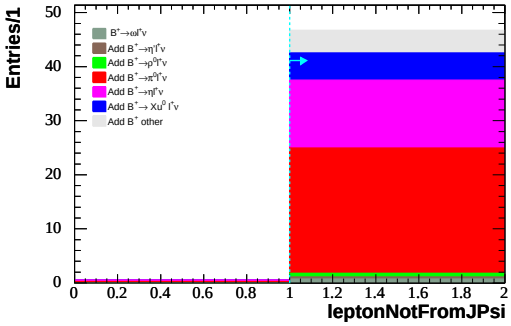
Figure A.11. Distributions of the electron photon-conversion veto in the signal region for the electron channel, validation sample.



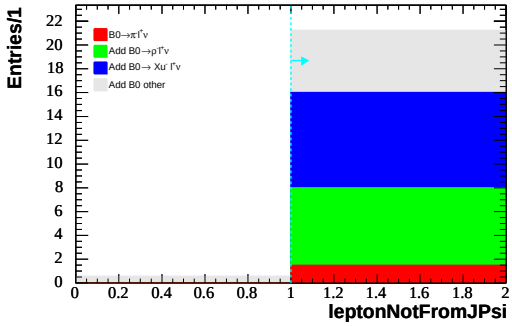
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



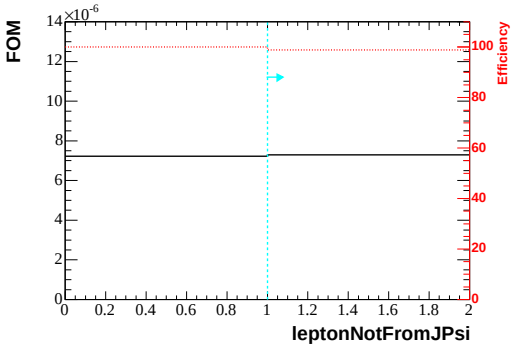
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

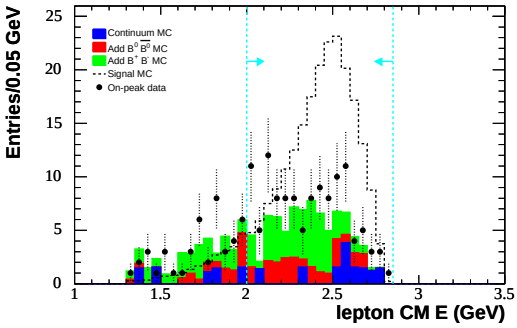


(d) Generic $B^0\overline{B}^0$ marginal distributions.

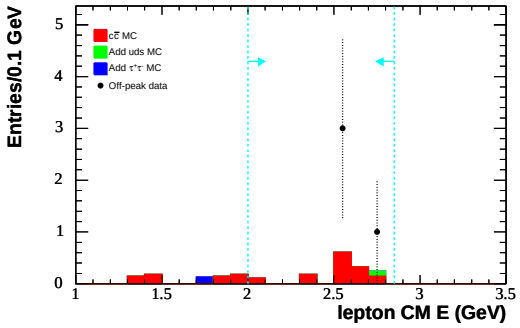


(e) FOM and efficiency distributions.

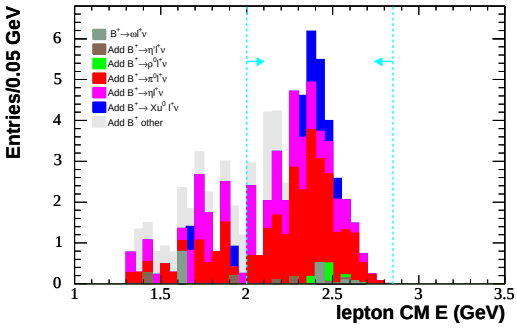
Figure A.12. Distributions of the electron J/ψ veto in the signal region for the electron channel, validation sample.



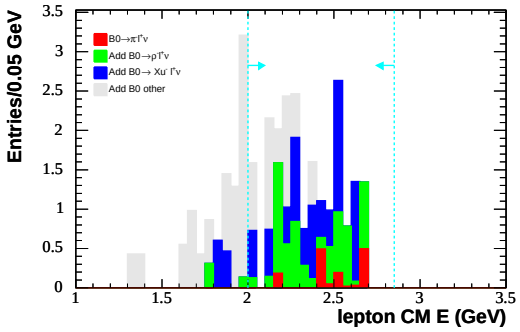
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



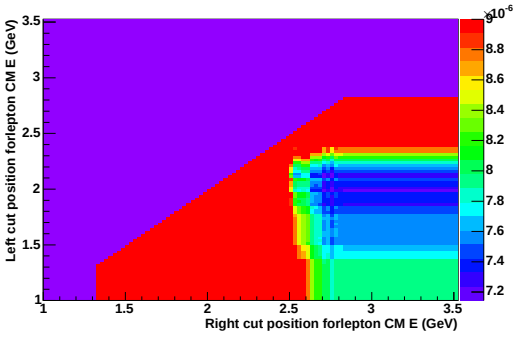
(b) Continuum marginal distributions.



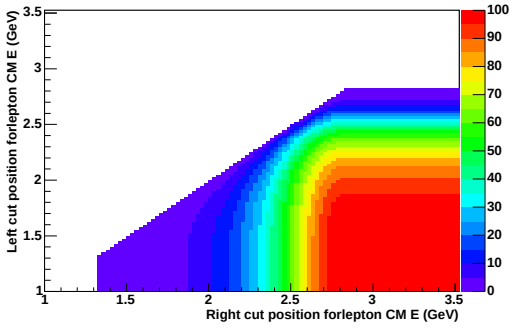
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

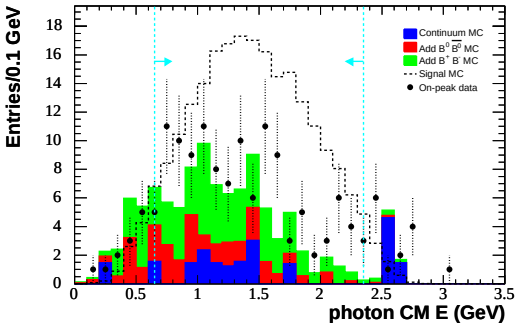


(e) FOM distribution.

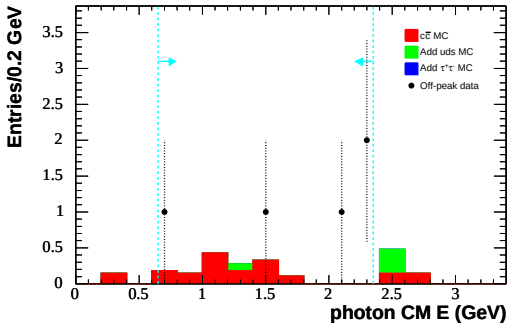


(f) Efficiency distribution.

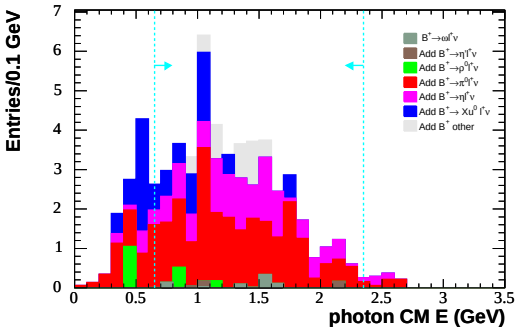
Figure A.13. Distributions of the signal lepton CM energy in the signal region for the electron channel, validation sample.



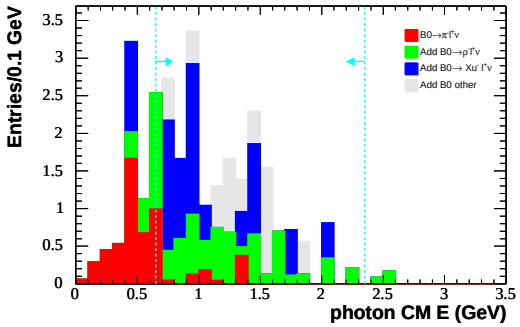
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



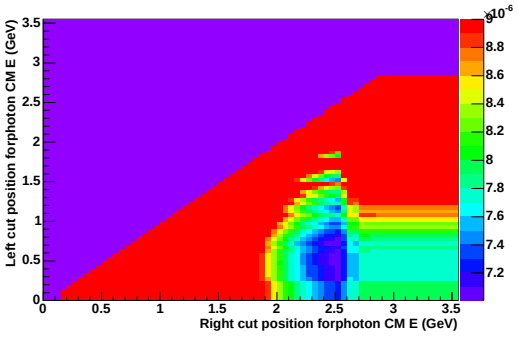
(b) Continuum marginal distributions.



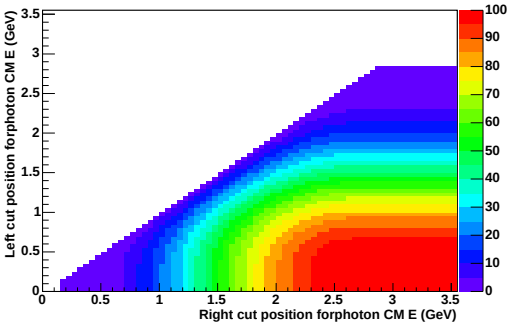
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

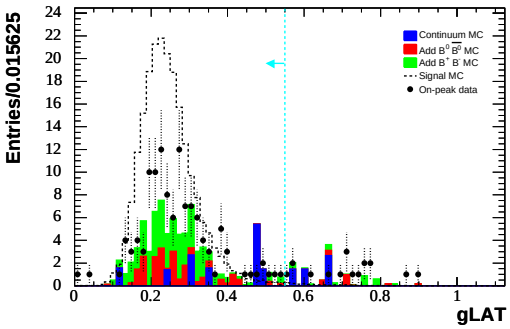


(e) FOM distribution.

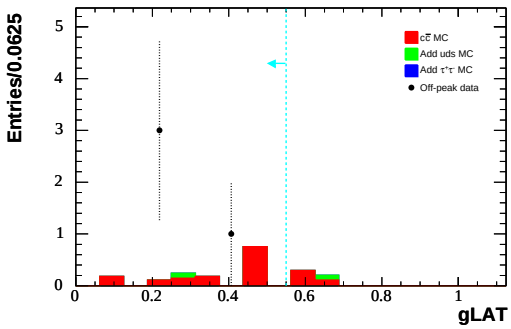


(f) Efficiency distribution.

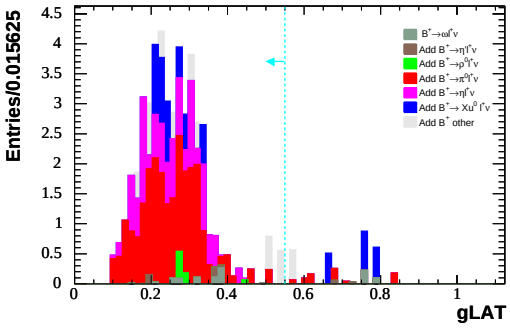
Figure A.14. Distributions of the signal photon CM energy in the signal region for the electron channel, validation sample.



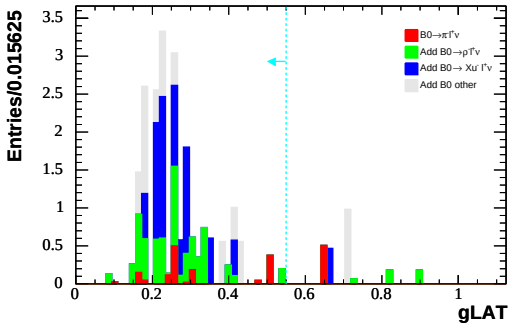
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



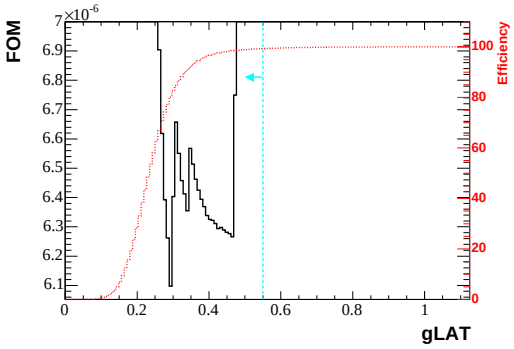
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

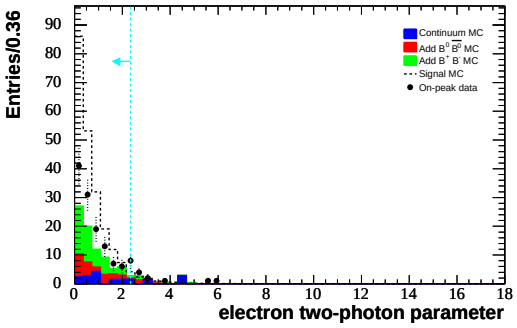


(d) Generic $B^0\overline{B}^0$ marginal distributions.

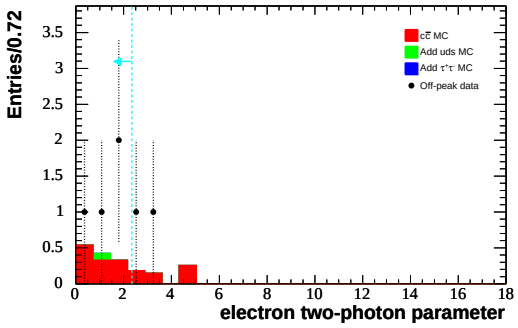


(e) FOM and efficiency distributions.

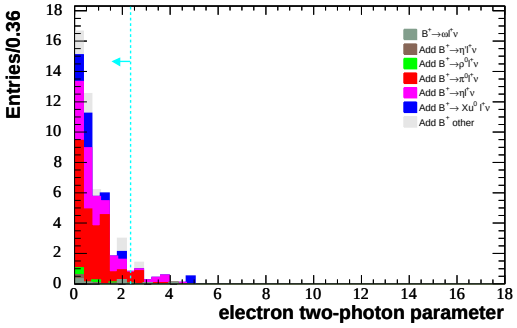
Figure A.15. Distributions of gLAT in the signal region for the electron channel, validation sample.



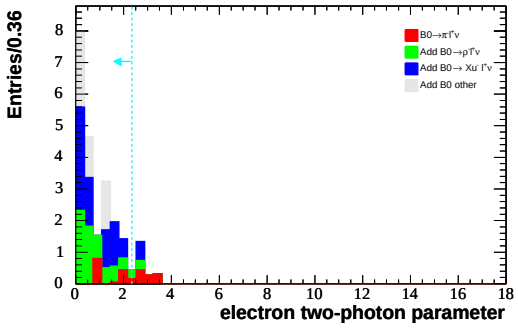
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



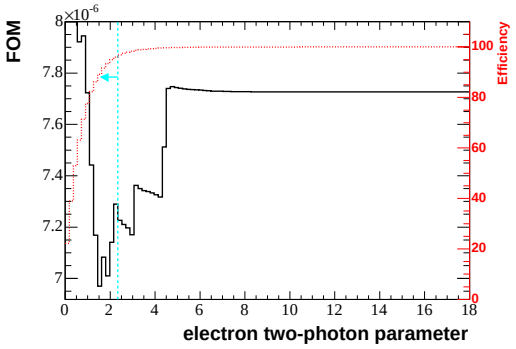
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

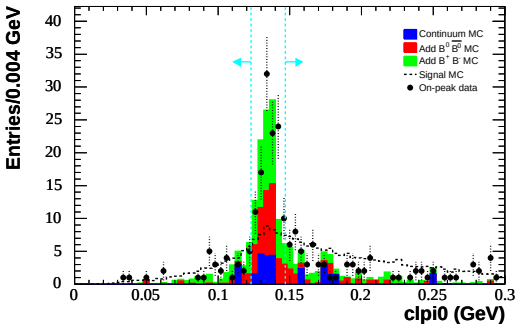


(d) Generic $B^0\overline{B}^0$ marginal distributions.

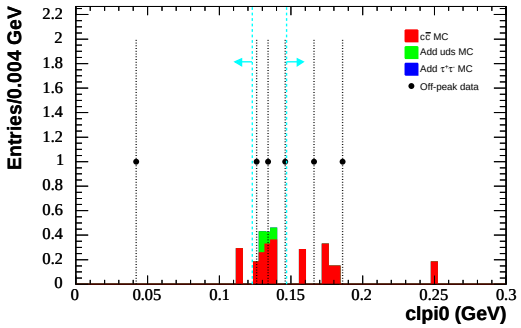


(e) FOM and efficiency distributions.

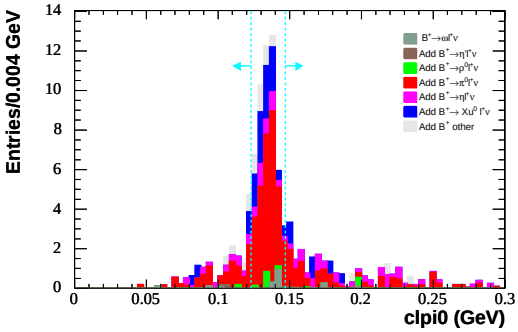
Figure A.16. Distributions of the electron two-photon parameter in the signal region for the electron channel, validation sample.



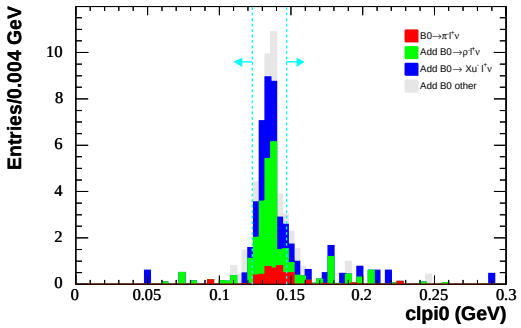
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



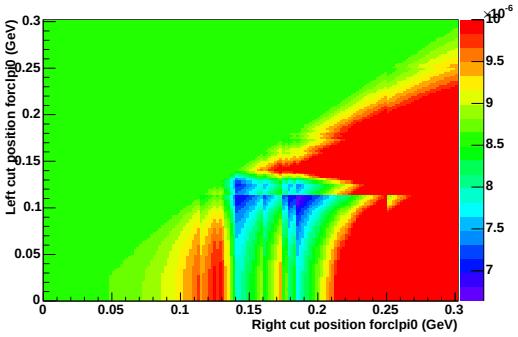
(b) Continuum marginal distributions.



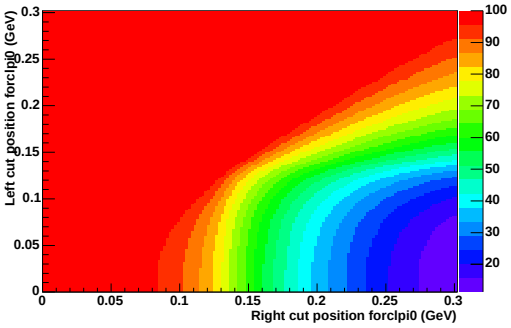
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

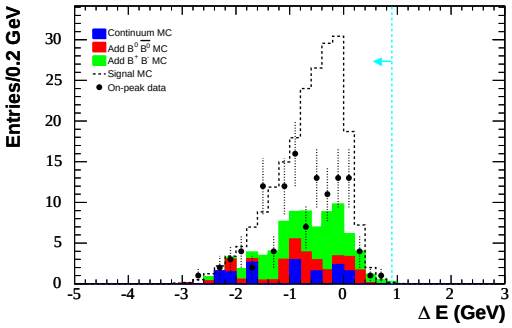


(e) FOM distribution.

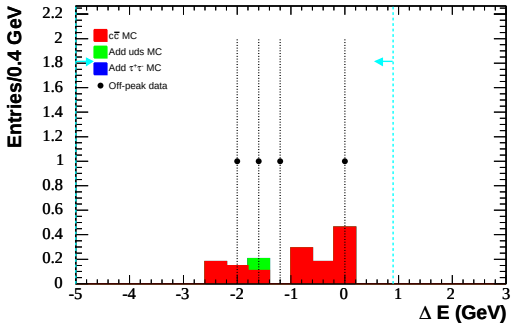


(f) Efficiency distribution.

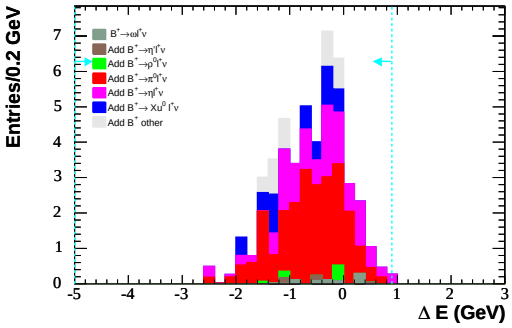
Figure A.17. Distributions of the closest π^0 mass in the signal region for the electron channel, validation sample.



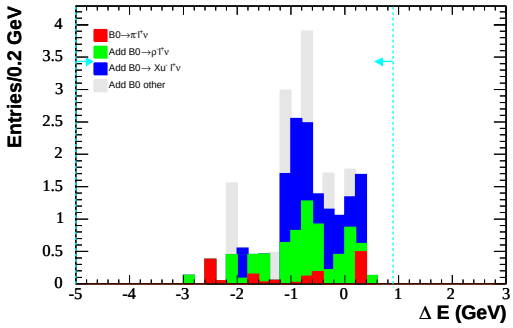
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



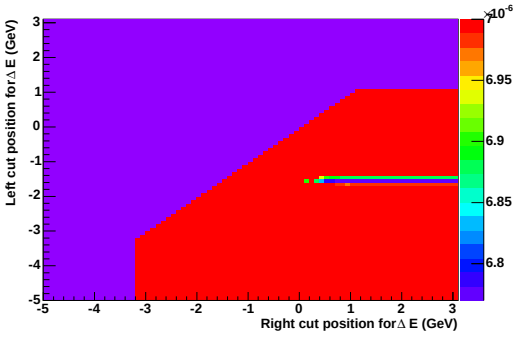
(b) Continuum marginal distributions.



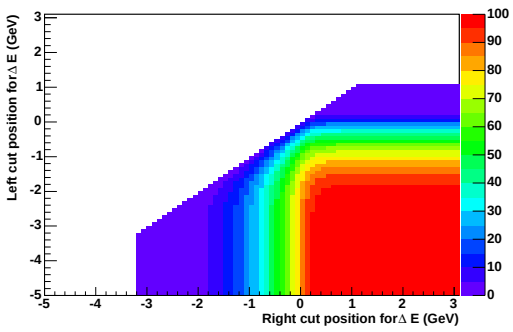
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

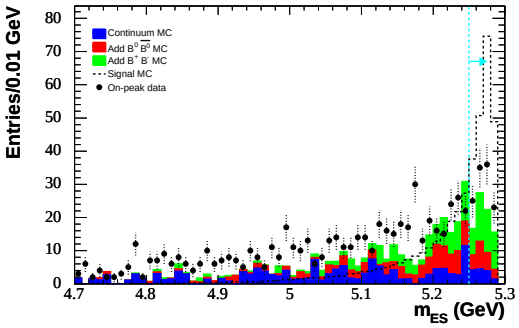


(e) FOM distribution.

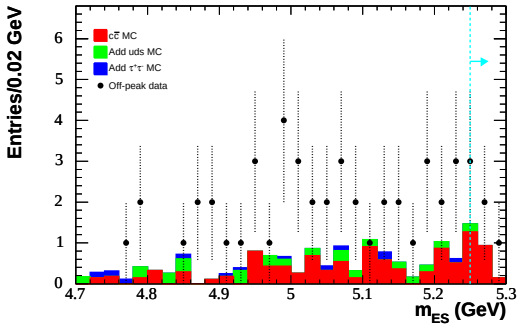


(f) Efficiency distribution.

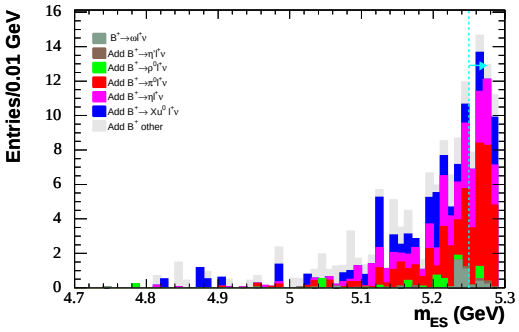
Figure A.18. Distributions of ΔE in the signal region for the electron channel, validation sample.



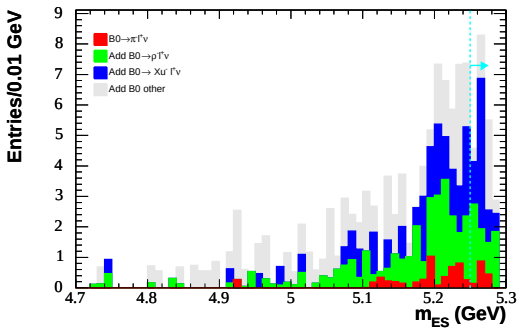
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



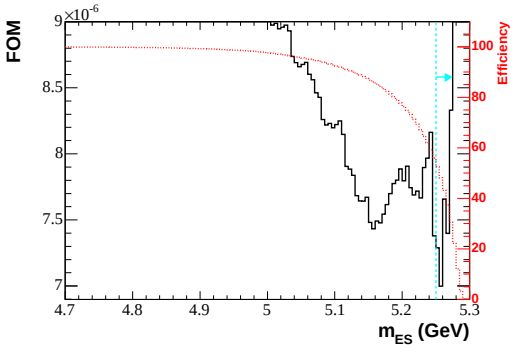
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

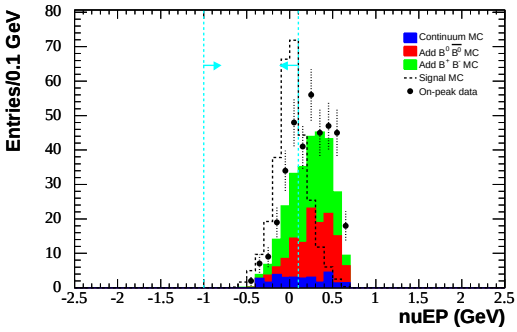


(d) Generic $B^0\overline{B}^0$ marginal distributions.

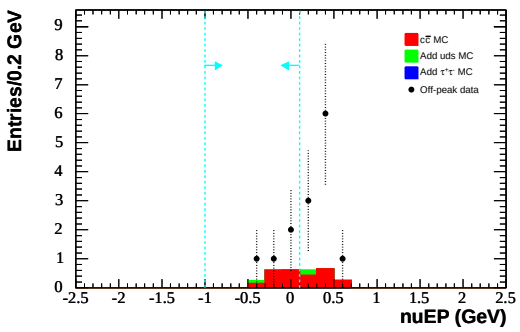


(e) FOM and efficiency distributions.

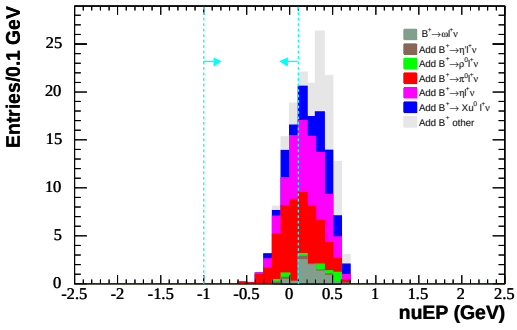
Figure A.19. Distributions of $m_{\rm ES}$ in the signal region for the electron channel, validation sample.



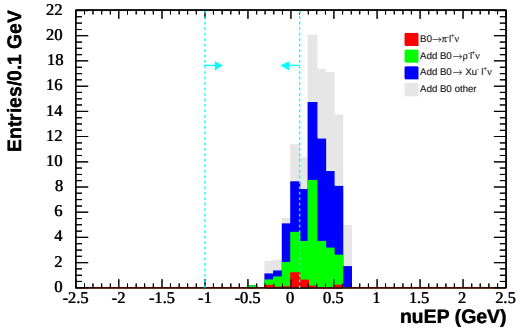
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



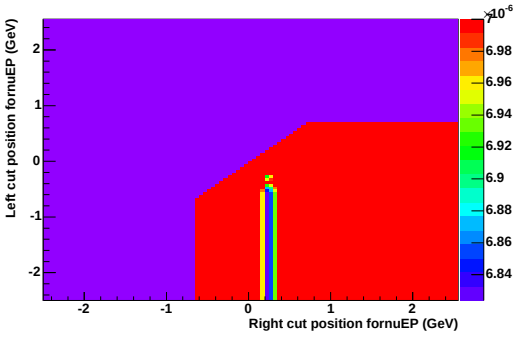
(b) Continuum marginal distributions.



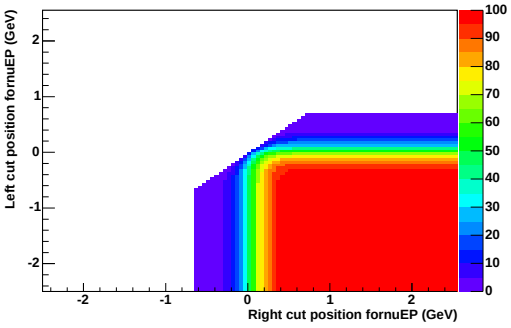
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

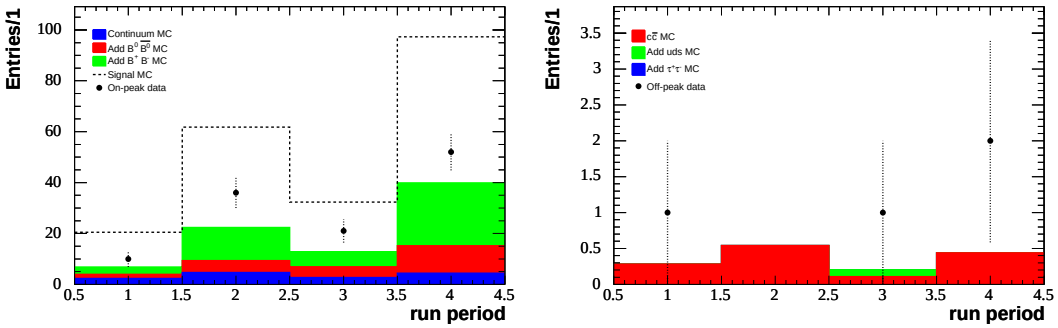


(e) FOM distribution.



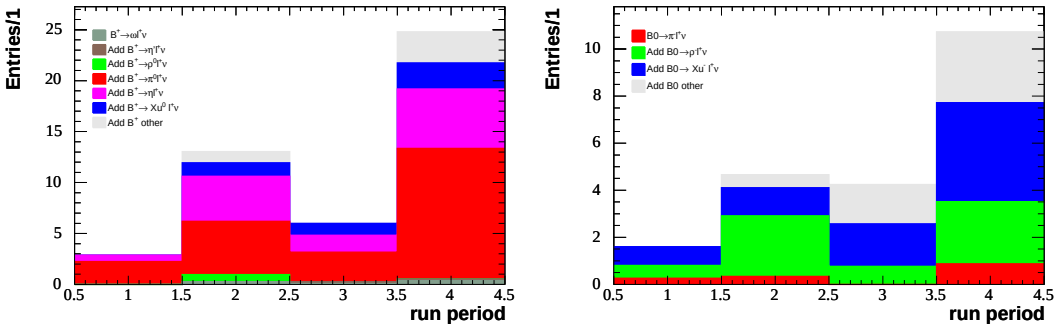
(f) Efficiency distribution.

Figure A.20. Distributions of nuEP in the signal region for the electron channel, validation sample.



(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.

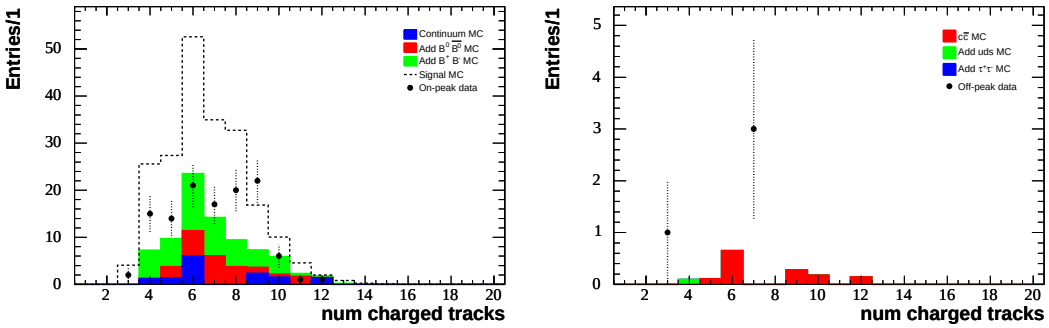
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

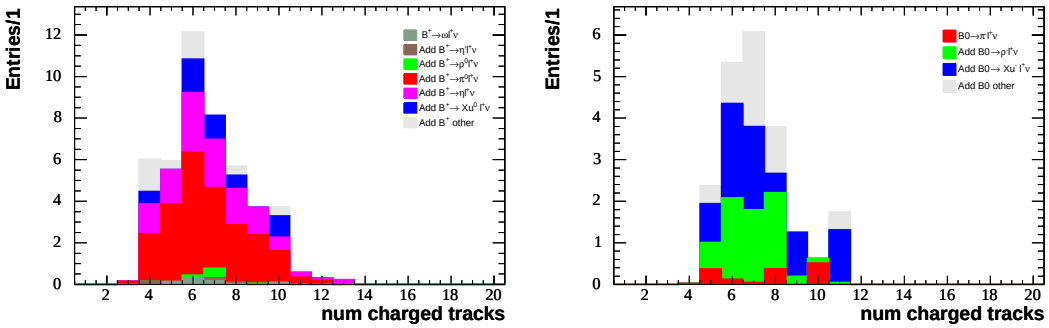
(d) Generic $B^0\overline{B}^0$ marginal distributions.

Figure A.21. Distributions of the run period event count in the signal region for the electron channel, validation sample.



(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.

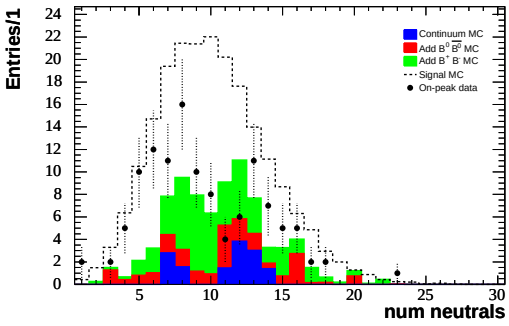
(b) Continuum marginal distributions.



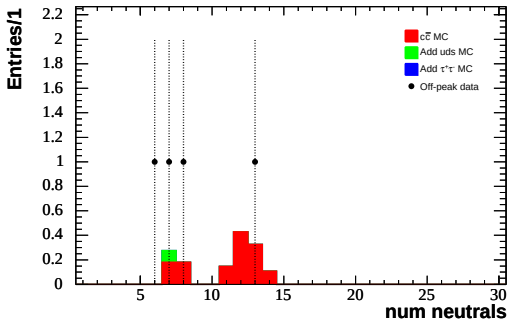
(c) Generic B^+B^- marginal distributions.

(d) Generic $B^0\overline{B}^0$ marginal distributions.

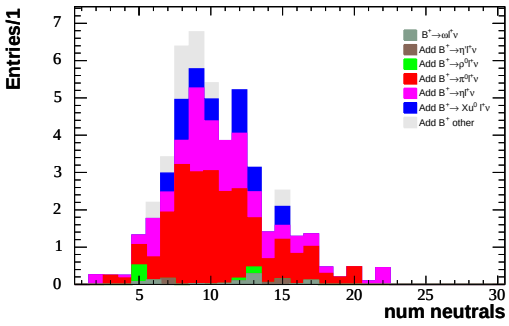
Figure A.22. Distributions of the number of charged tracks in the signal region for the electron channel, validation sample.



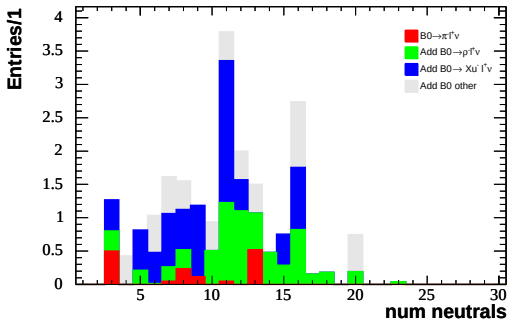
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

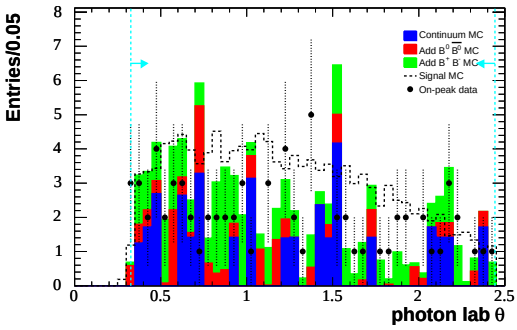


(d) Generic $B^0\overline{B}^0$ marginal distributions.

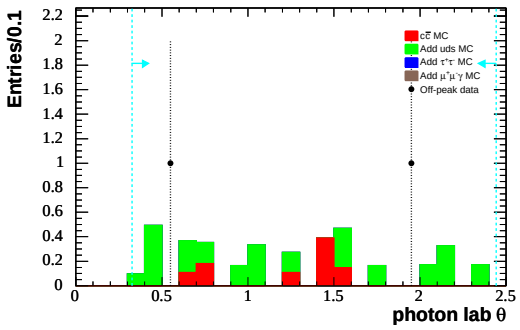
Figure A.23. Distributions of the number of neutrals in the signal region for the electron channel, validation sample.

A.2 Muon Channel

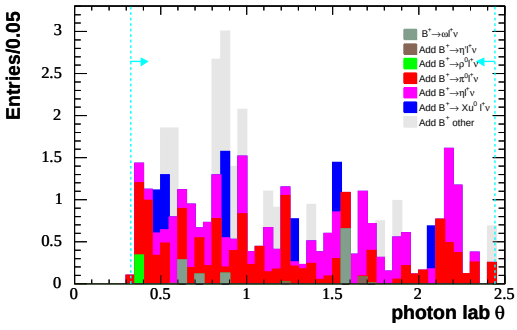
In this appendix we provide the marginal cut variable distributions for the muon channel, validation sample, as well as the distributions of other variables of interest. The m_{ES} and nuEP cuts are the optimized signal region cuts given in table 9.3. The rest of the cuts are given in table 8.3.



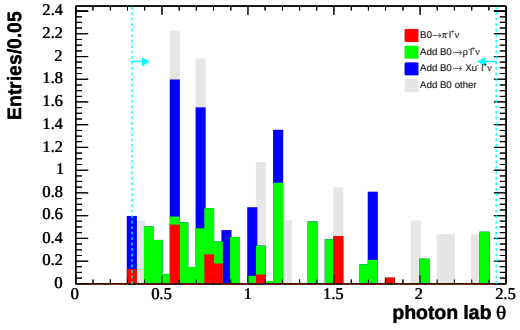
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



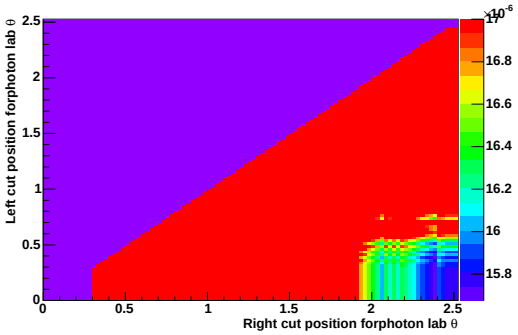
(b) Continuum marginal distributions.



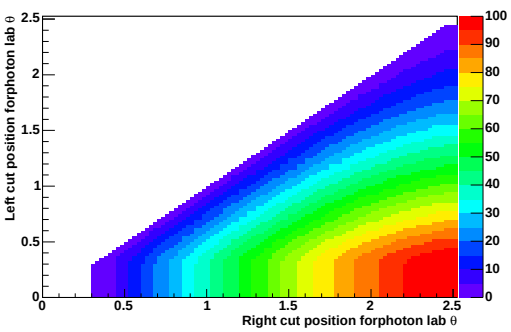
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

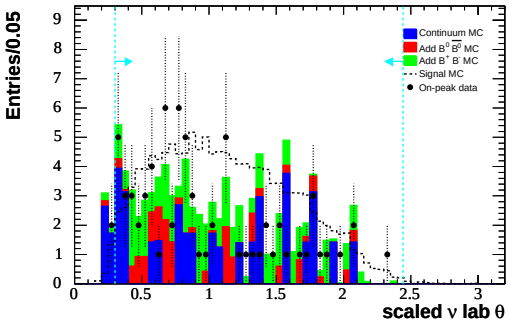


(e) FOM distribution.

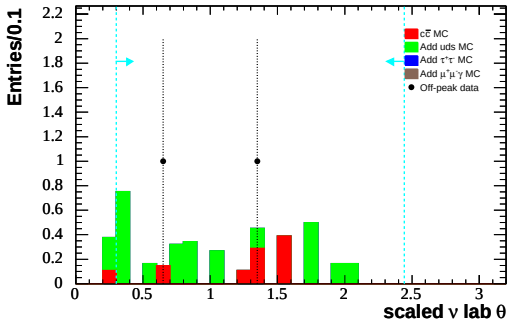


(f) Efficiency distribution.

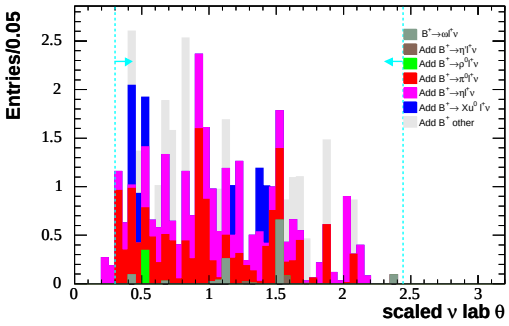
Figure A.24. Distributions of the photon lab θ in the signal region for the muon channel, validation sample.



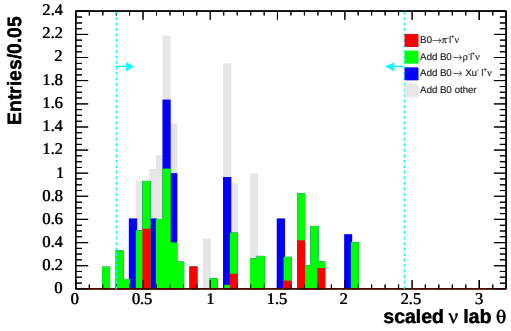
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



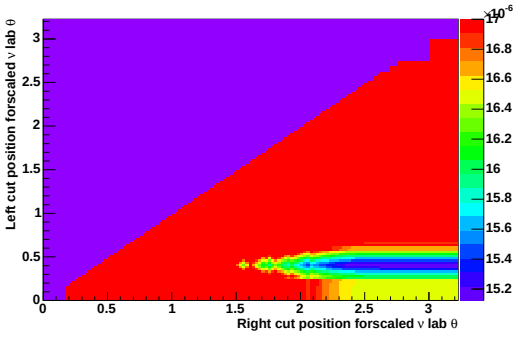
(b) Continuum marginal distributions.



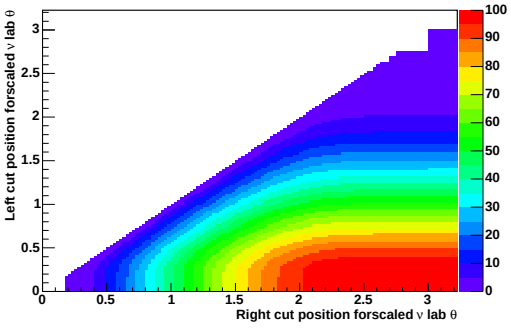
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

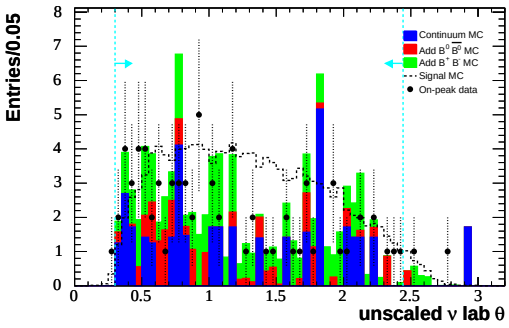


(e) FOM distribution.

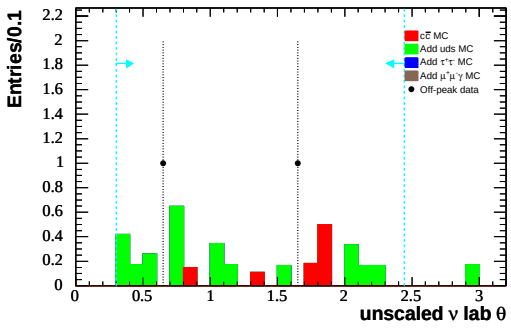


(f) Efficiency distribution.

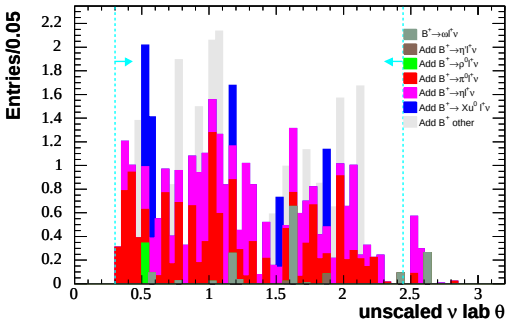
Figure A.25. Distributions of the scaled neutrino lab θ in the signal region for the muon channel, validation sample.



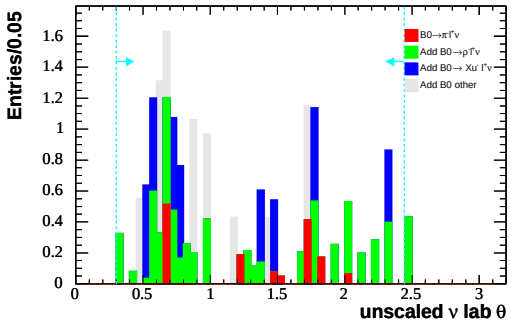
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



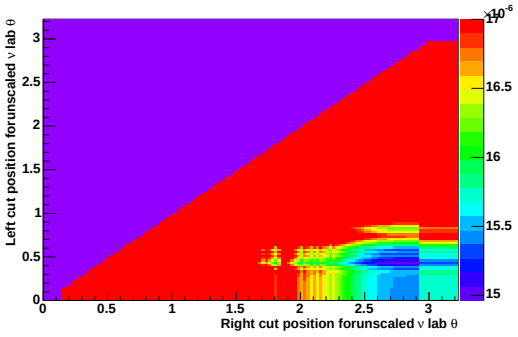
(b) Continuum marginal distributions.



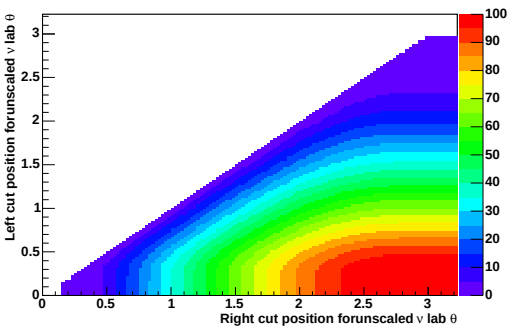
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

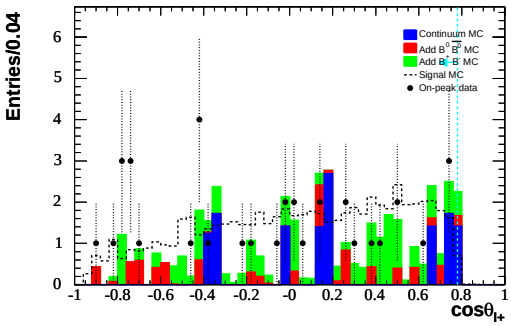


(e) FOM distribution.

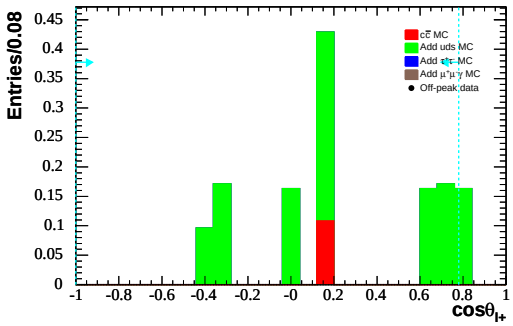


(f) Efficiency distribution.

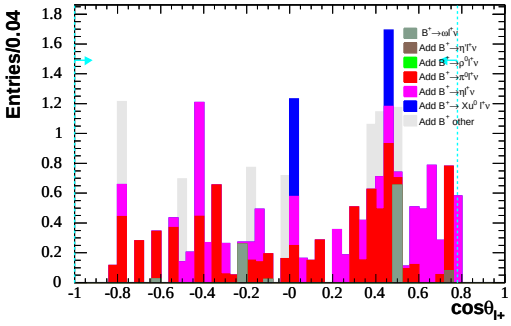
Figure A.26. Distributions of the unscaled neutrino lab θ in the signal region for the muon channel, validation sample.



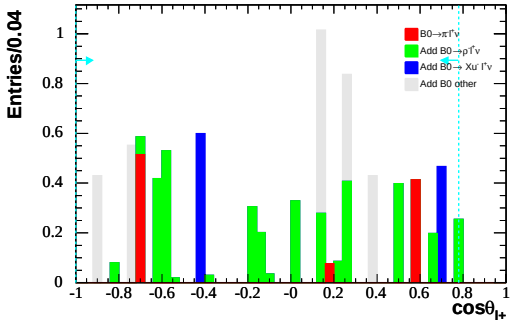
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



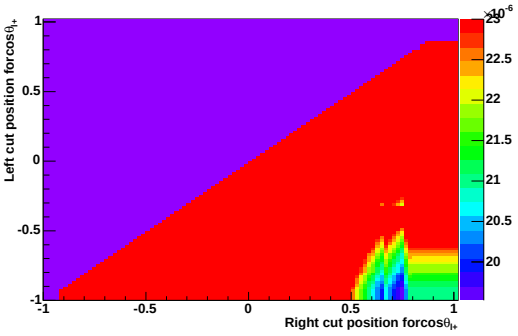
(b) Continuum marginal distributions.



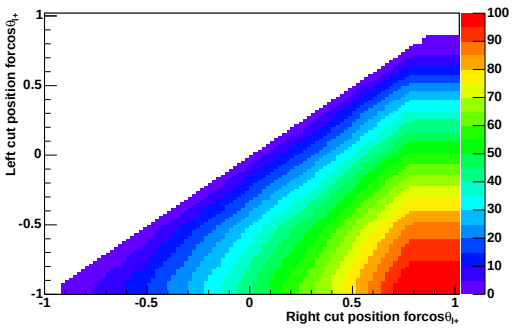
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

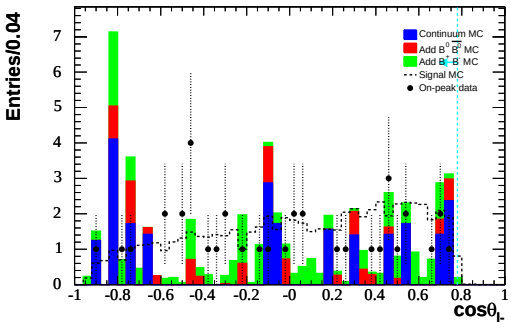


(e) FOM distribution.



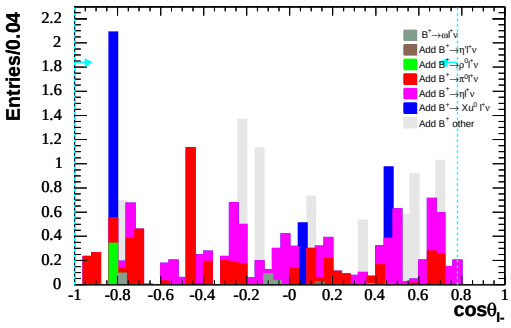
(f) Efficiency distribution.

Figure A.27. Distributions of μ^+: $\cos(\theta_{\mathrm{Lab}})$ in the signal region for the muon channel, validation sample.



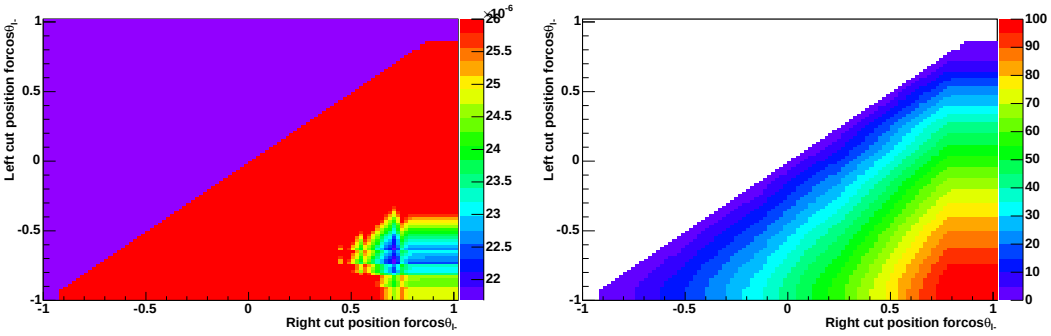
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.

(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

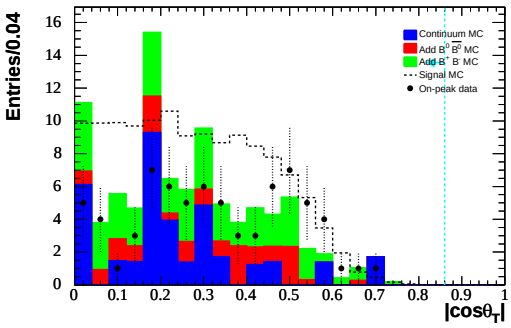
(d) Generic $B^0\overline{B}^0$ marginal distributions.



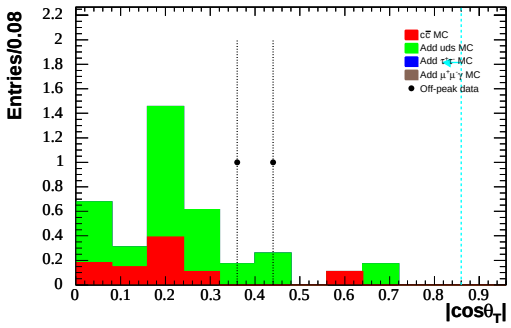
(e) FOM distribution.

(f) Efficiency distribution.

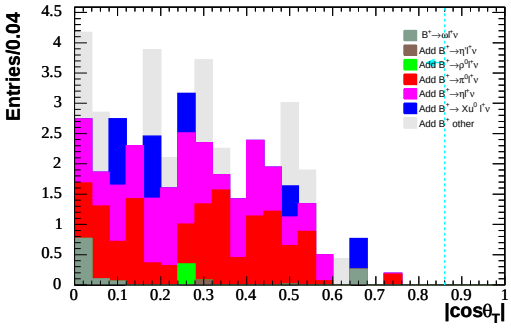
Figure A.28. Distributions of μ^-: $\cos(\theta_{\rm Lab})$ in the signal region for the muon channel, validation sample.



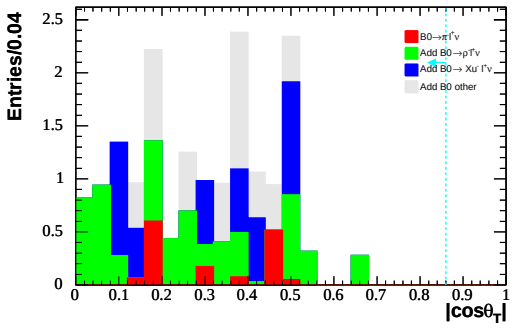
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



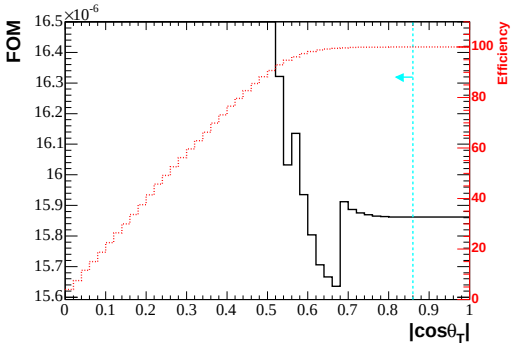
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

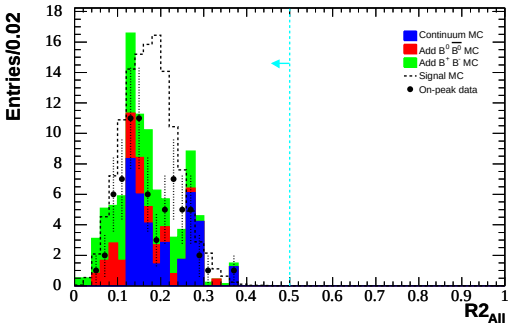


(d) Generic $B^0\overline{B}^0$ marginal distributions.

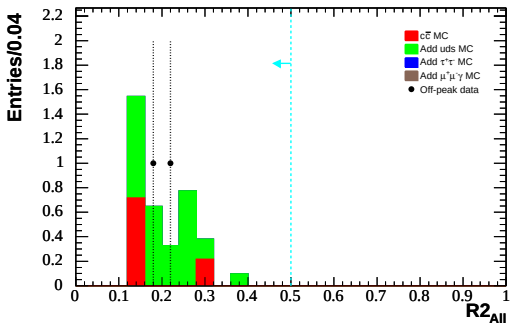


(e) FOM and efficiency distributions.

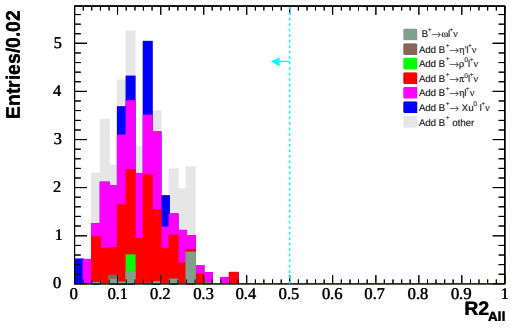
Figure A.29. Distributions of $|\cos\theta_T|$ in the signal region for the muon channel, validation sample.



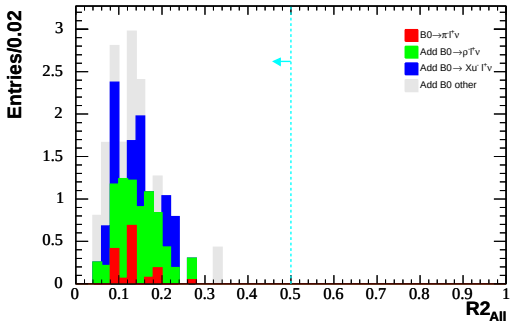
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



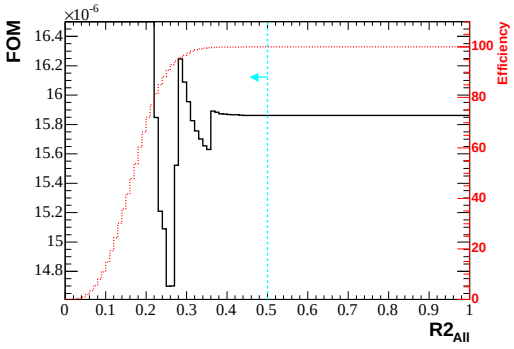
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

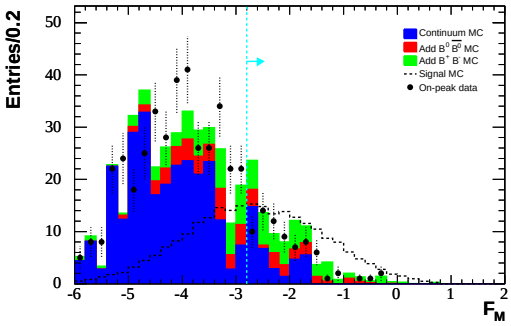


(d) Generic $B^0\overline{B}^0$ marginal distributions.

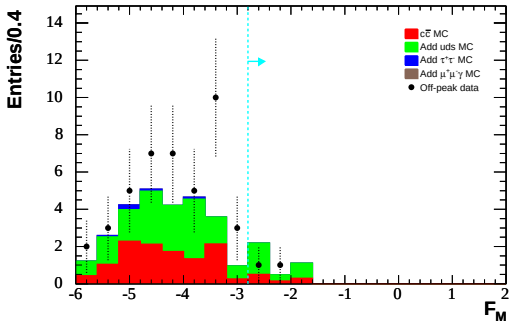


(e) FOM and efficiency distributions.

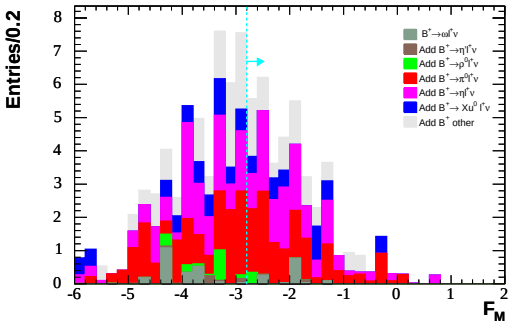
Figure A.30. Distributions of $R2_{\mathrm{All}}$ in the signal region for the muon channel, validation sample.



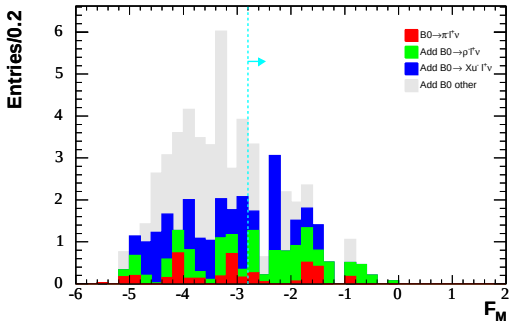
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



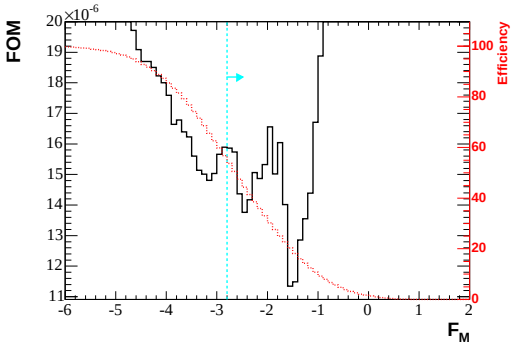
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

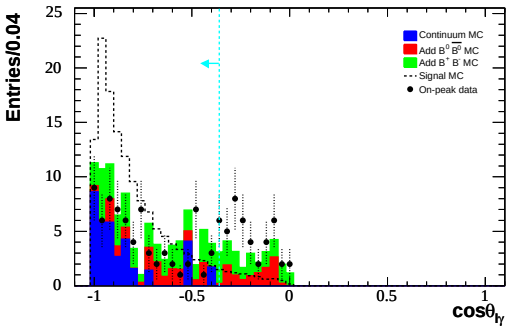


(d) Generic $B^0\overline{B}^0$ marginal distributions.

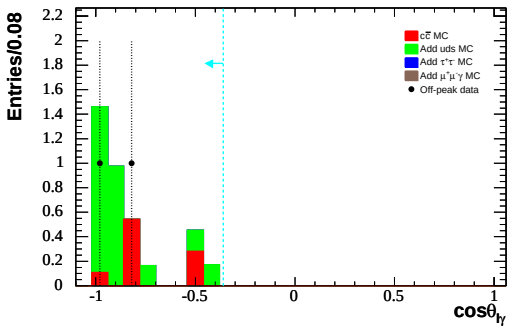


(e) FOM and efficiency distributions.

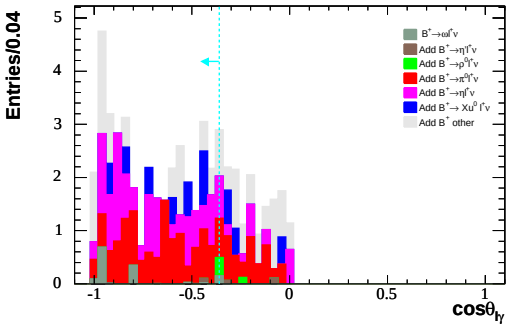
Figure A.31. Distributions of $\mathcal{F}_{\rm M}$ in the signal region for the muon channel, validation sample.



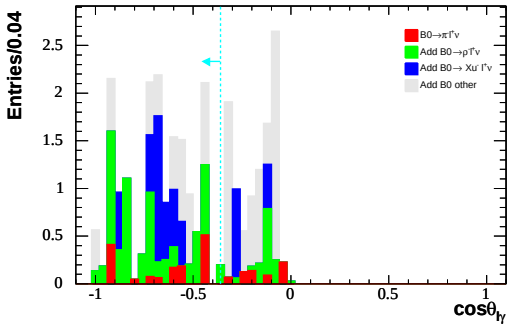
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



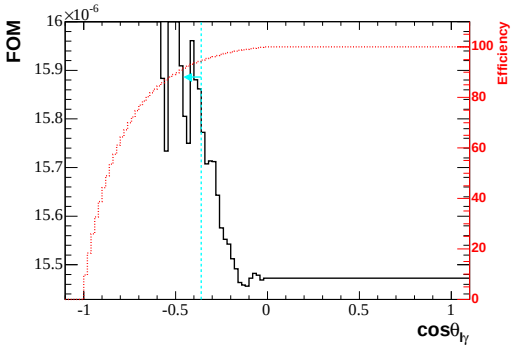
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

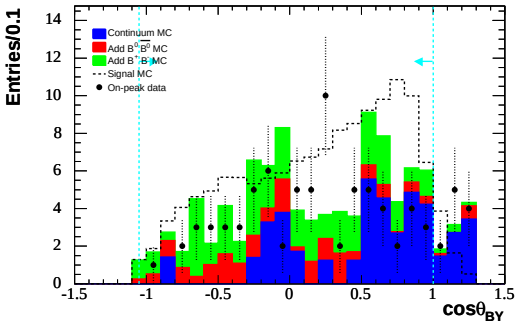


(d) Generic $B^0\overline{B}^0$ marginal distributions.

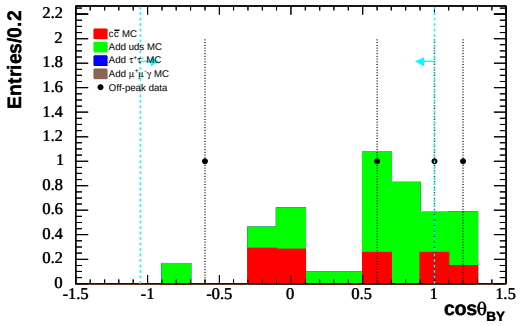


(e) FOM and efficiency distributions.

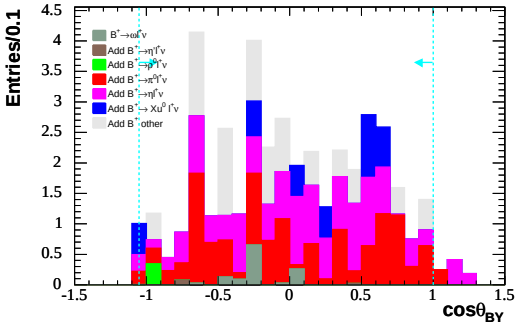
Figure A.32. Distributions of $\cos\theta_{\ell\gamma}$ in the signal region for the muon channel, validation sample.



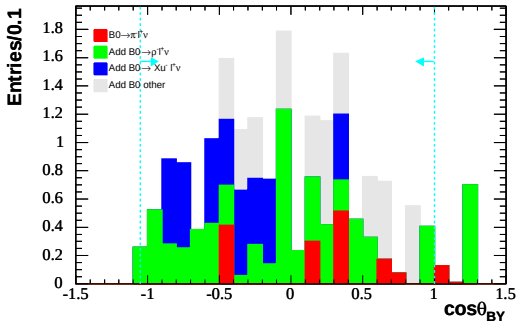
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



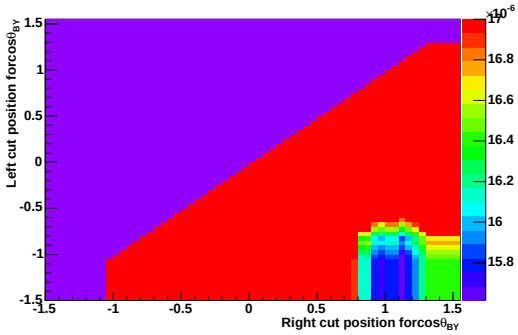
(b) Continuum marginal distributions.



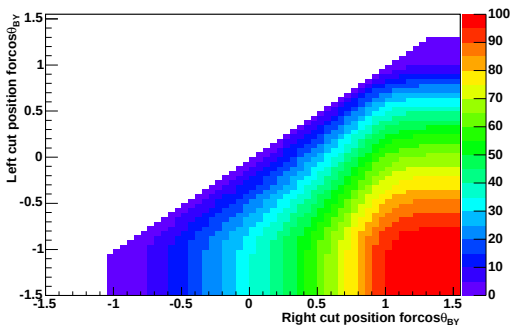
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

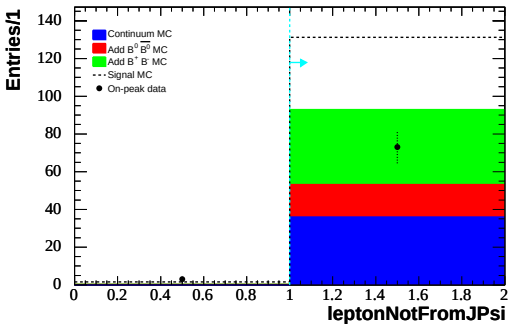


(e) FOM distribution.

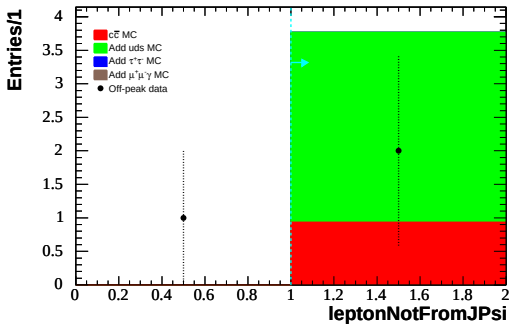


(f) Efficiency distribution.

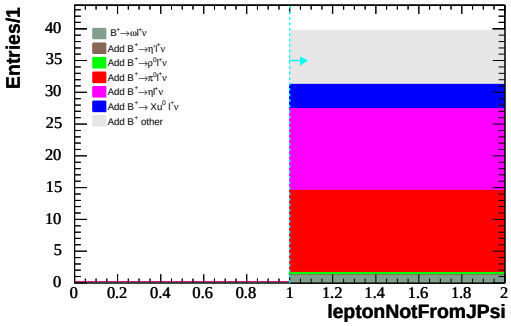
Figure A.33. Distributions of $\cos\theta_{BY}$ in the signal region for the muon channel, validation sample.



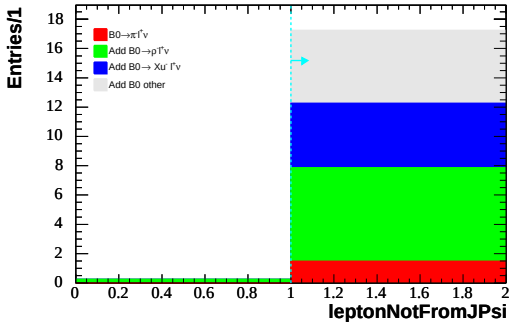
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



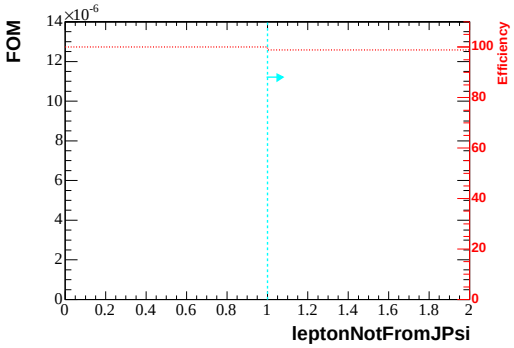
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

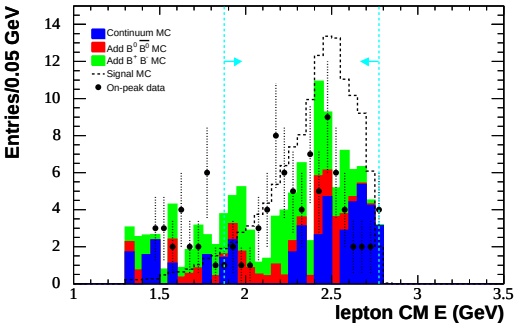


(d) Generic $B^0\overline{B}^0$ marginal distributions.

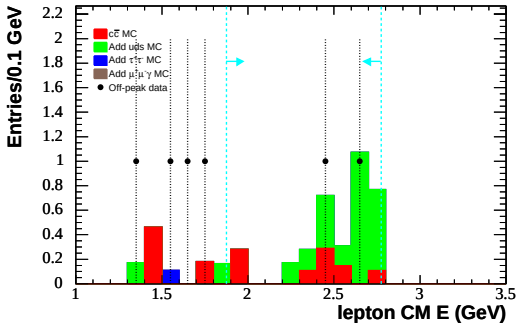


(e) FOM and efficiency distributions.

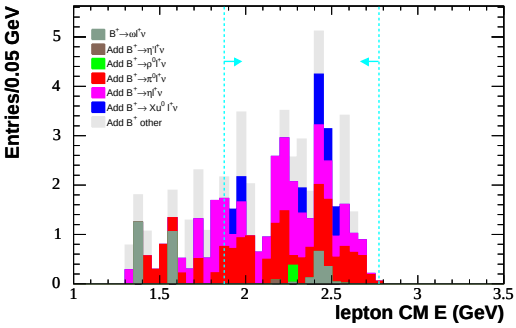
Figure A.34. Distributions of the muon J/ψ veto in the signal region for the muon channel, validation sample.



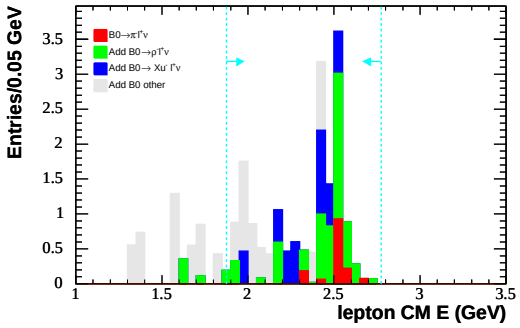
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



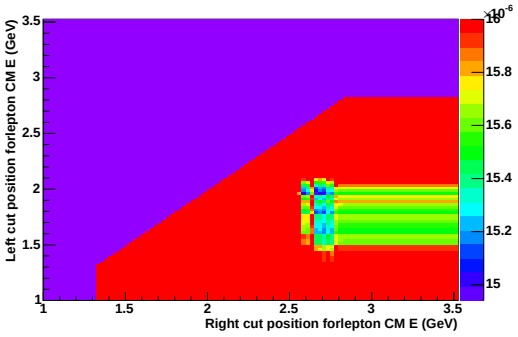
(b) Continuum marginal distributions.



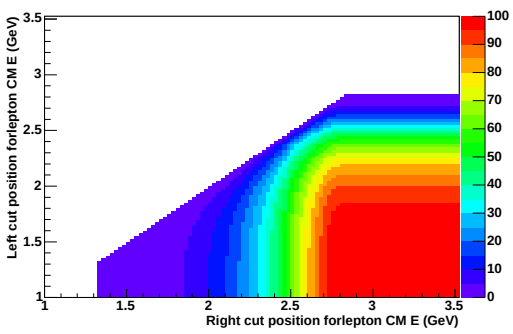
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

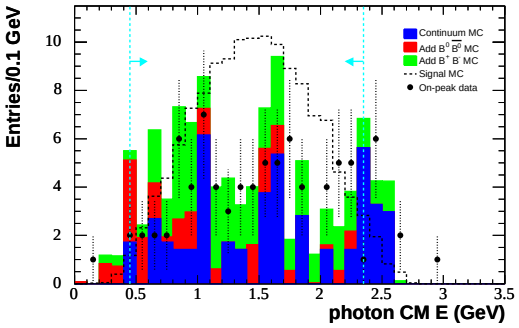


(e) FOM distribution.

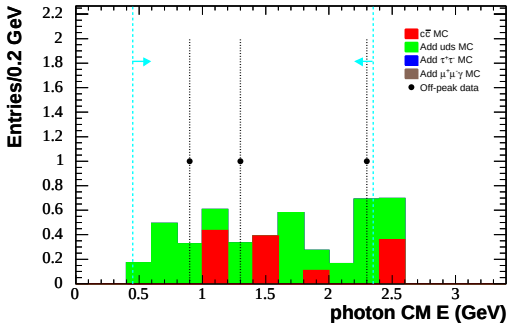


(f) Efficiency distribution.

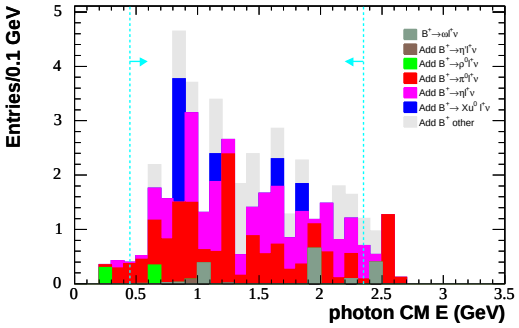
Figure A.35. Distributions of the signal lepton CM energy in the signal region for the muon channel, validation sample.



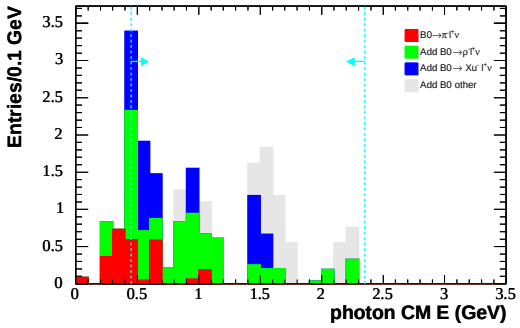
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



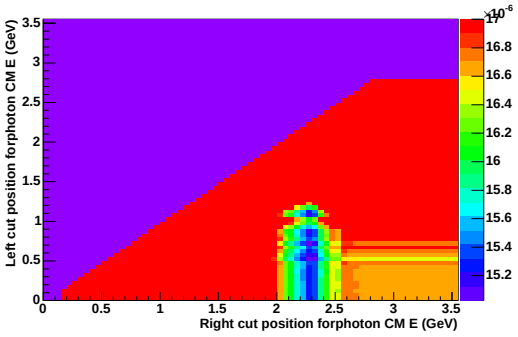
(b) Continuum marginal distributions.



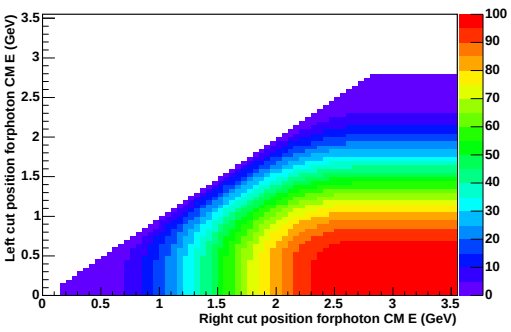
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

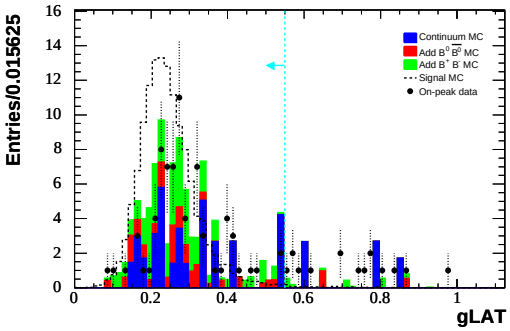


(e) FOM distribution.

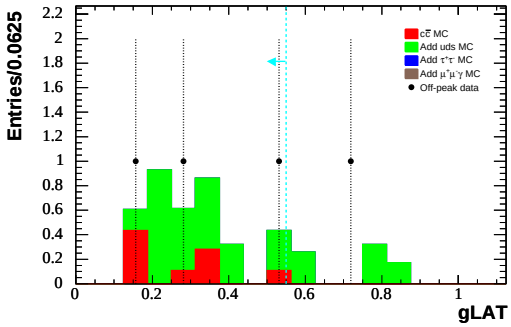


(f) Efficiency distribution.

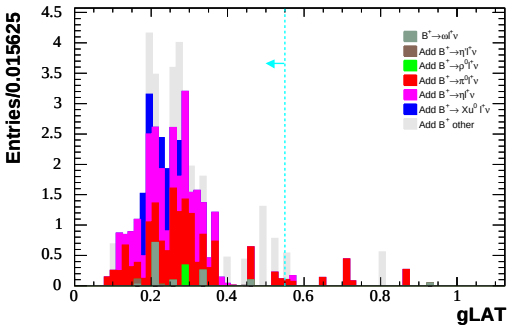
Figure A.36. Distributions of the signal photon CM energy in the signal region for the muon channel, validation sample.



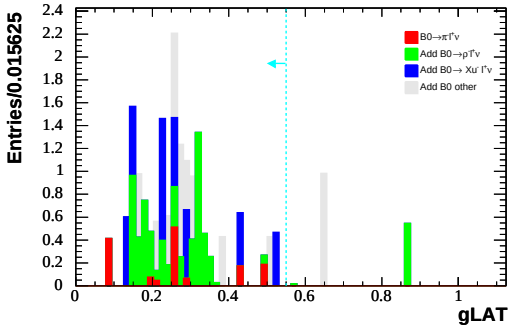
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



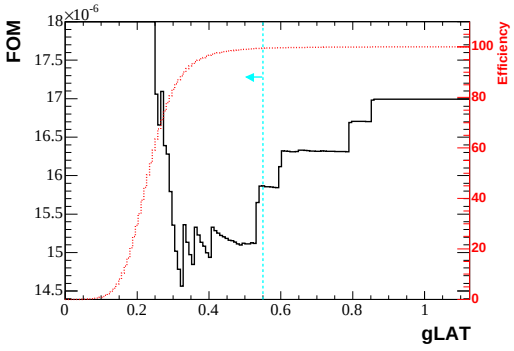
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

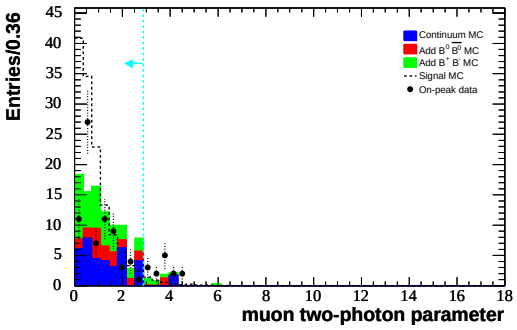


(d) Generic $B^0\overline{B}^0$ marginal distributions.

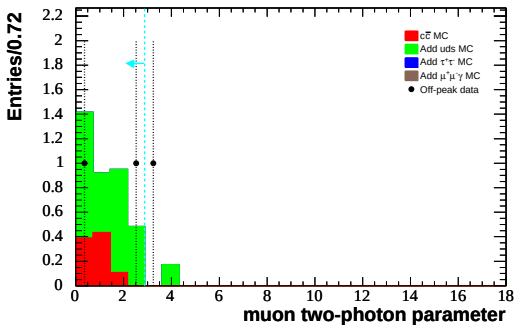


(e) FOM and efficiency distributions.

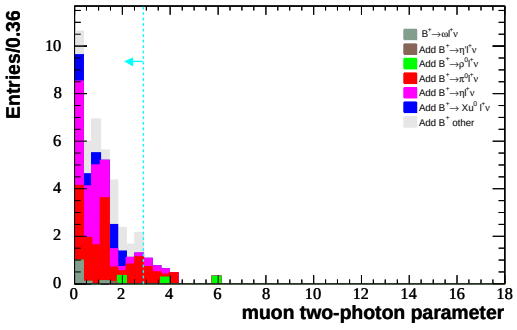
Figure A.37. Distributions of gLAT in the signal region for the muon channel, validation sample.



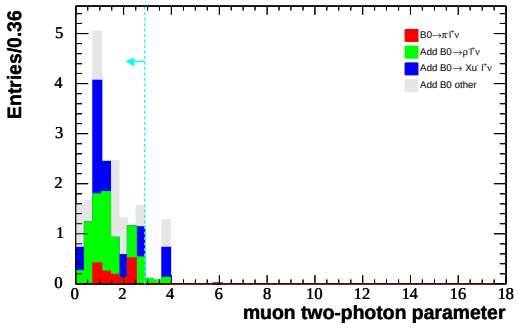
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



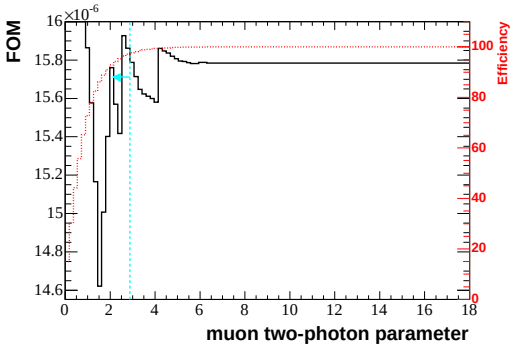
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

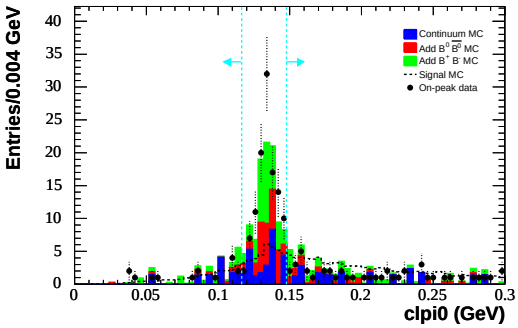


(d) Generic $B^0\overline{B}^0$ marginal distributions.

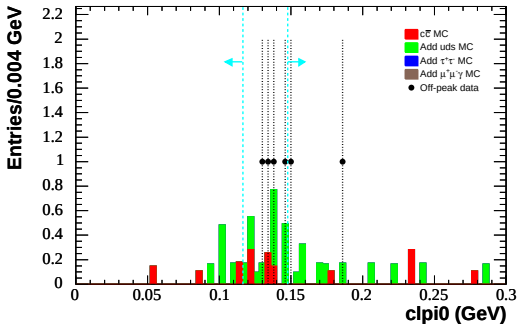


(e) FOM and efficiency distributions.

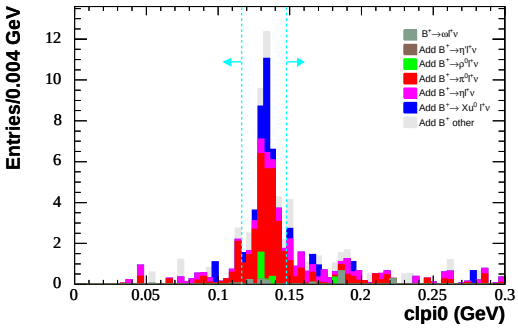
Figure A.38. Distributions of the muon two-photon parameter in the signal region for the muon channel, validation sample.



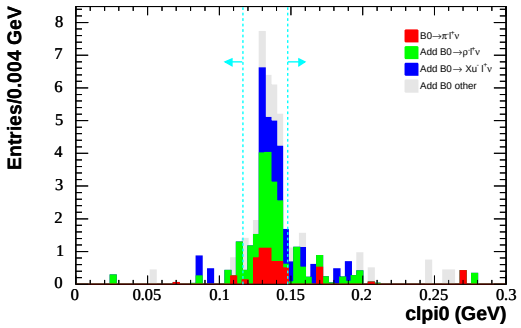
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



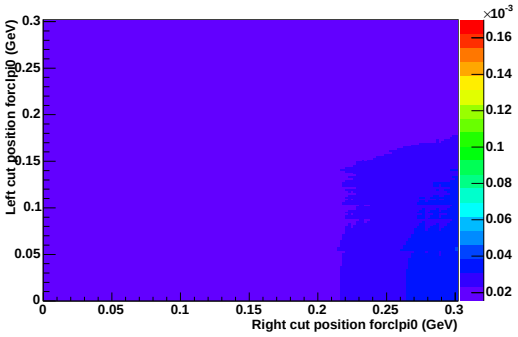
(b) Continuum marginal distributions.



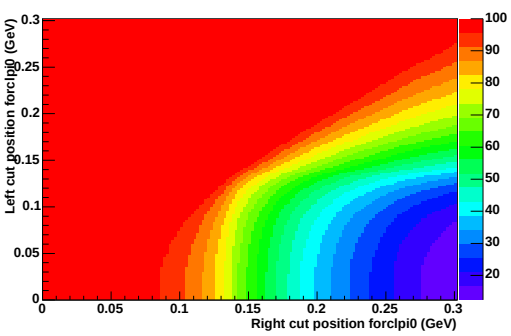
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

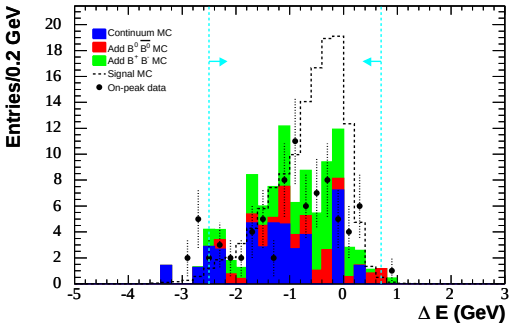


(e) FOM distribution.

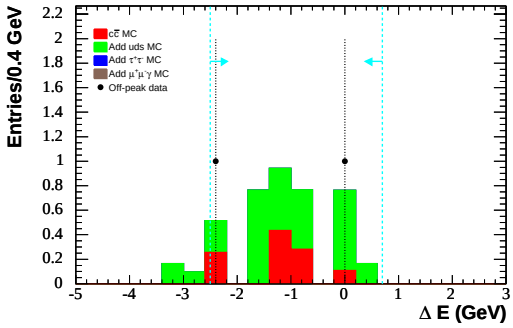


(f) Efficiency distribution.

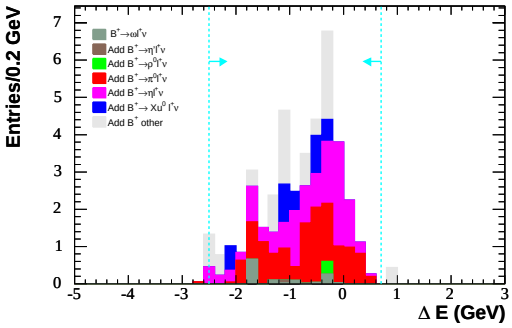
Figure A.39. Distributions of the closest π^0 mass in the signal region for the muon channel, validation sample.



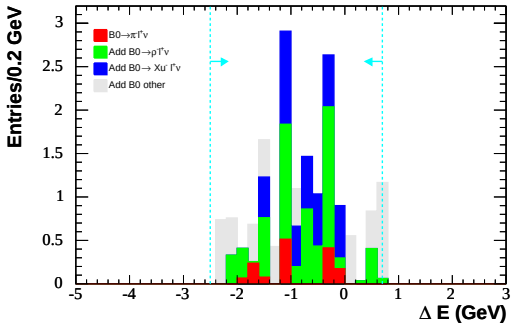
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



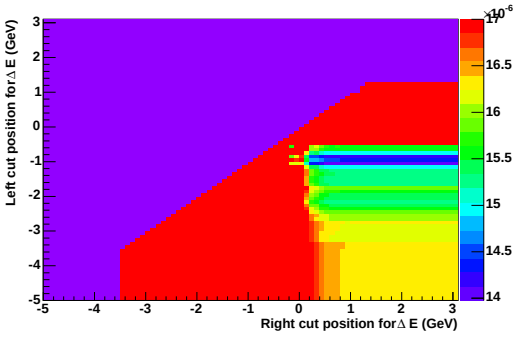
(b) Continuum marginal distributions.



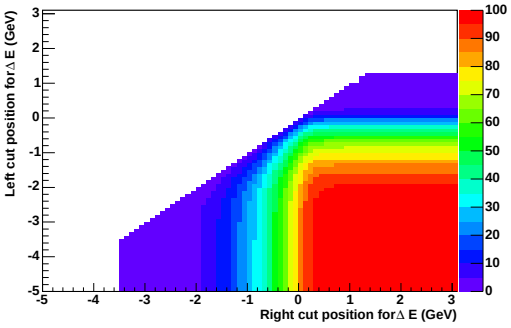
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

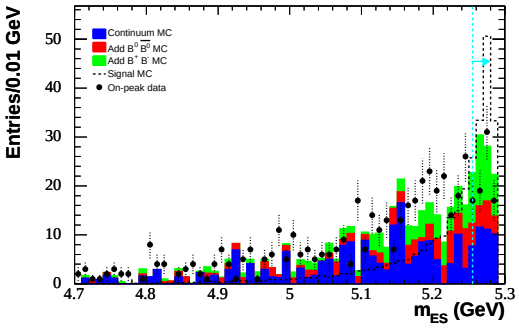


(e) FOM distribution.

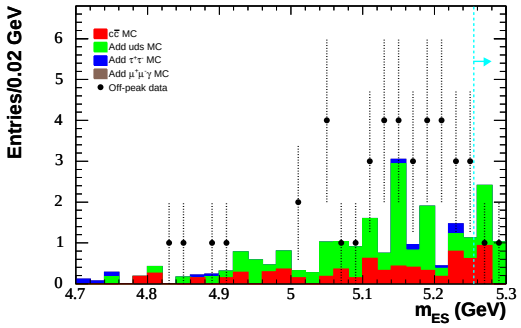


(f) Efficiency distribution.

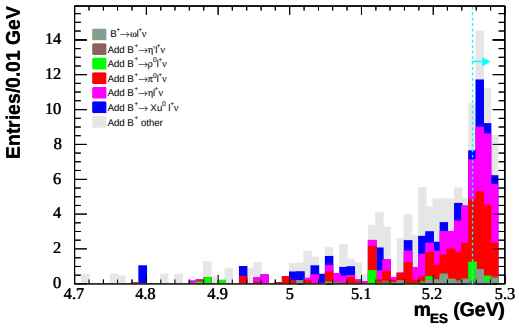
Figure A.40. Distributions of ΔE in the signal region for the muon channel, validation sample.



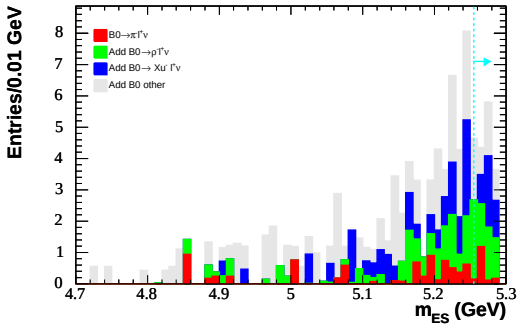
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



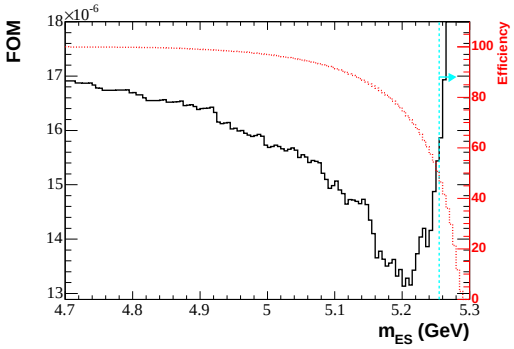
(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.

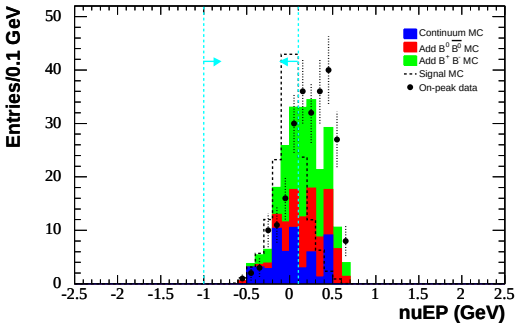


(d) Generic $B^0\overline{B}^0$ marginal distributions.

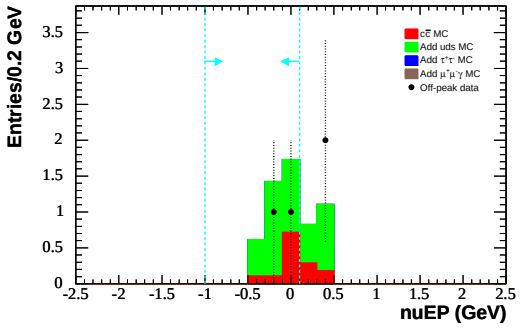


(e) FOM and efficiency distributions.

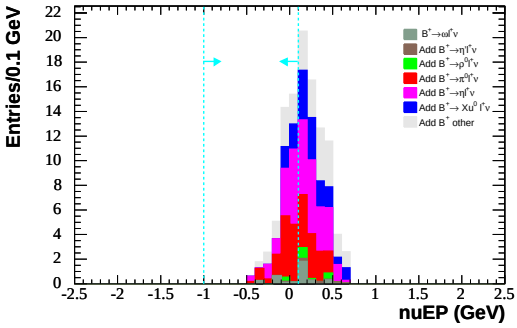
Figure A.41. Distributions of $m_{\rm ES}$ in the signal region for the muon channel, validation sample.



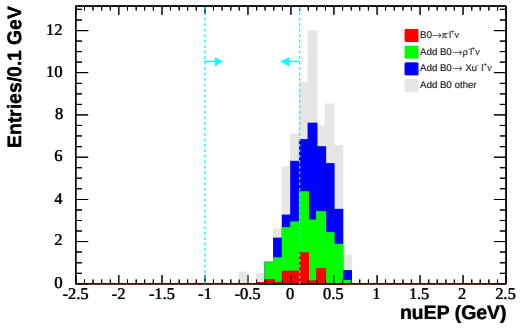
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



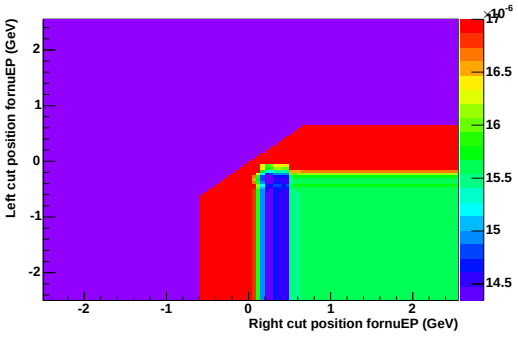
(b) Continuum marginal distributions.



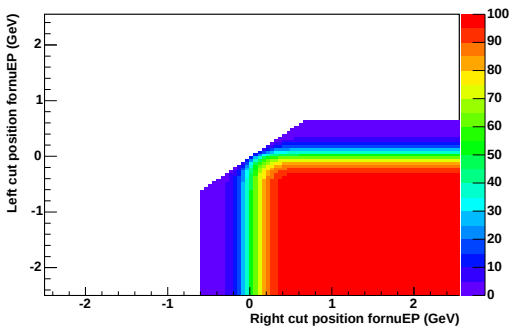
(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

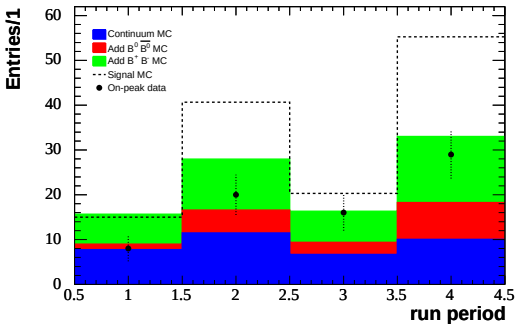


(e) FOM distribution.

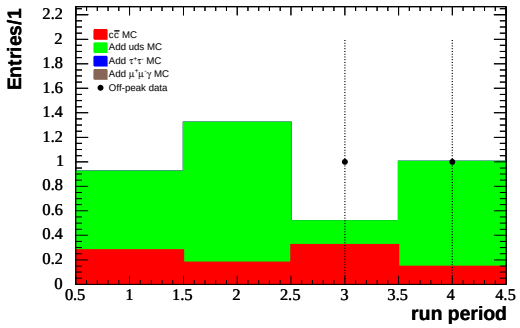


(f) Efficiency distribution.

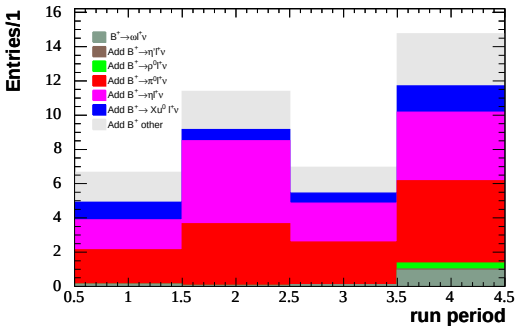
Figure A.42. Distributions of nuEP in the signal region for the muon channel, validation sample.



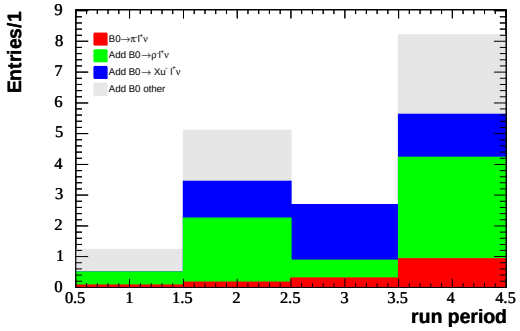
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



(b) Continuum marginal distributions.

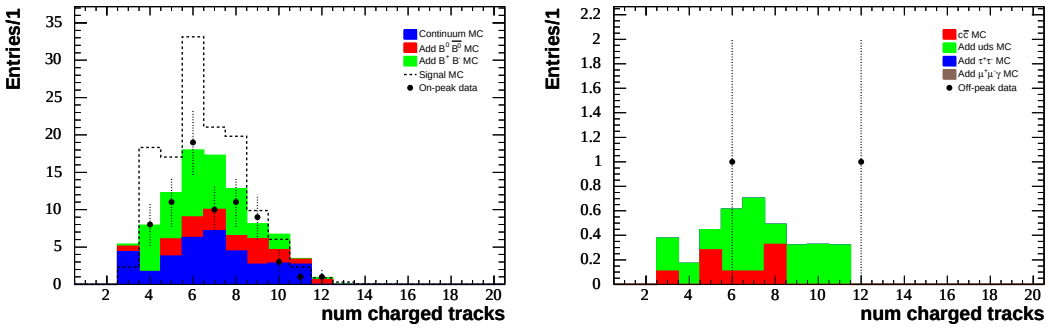


(c) Generic B^+B^- marginal distributions.



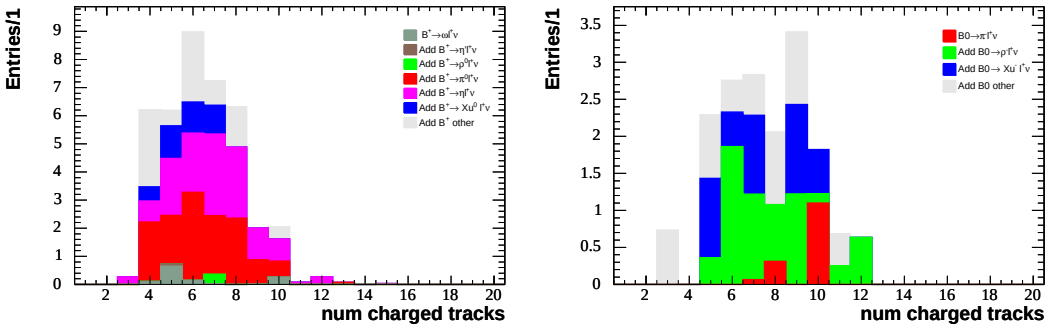
(d) Generic $B^0\overline{B}^0$ marginal distributions.

Figure A.43. Distributions of the run period event count in the signal region for the muon channel, validation sample.



(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.

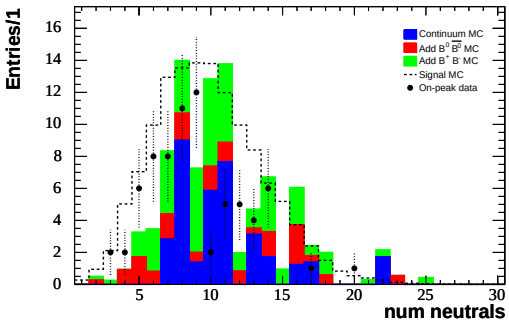
(b) Continuum marginal distributions.



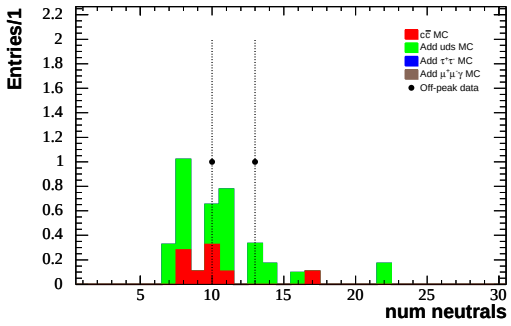
(c) Generic B^+B^- marginal distributions.

(d) Generic $B^0\overline{B}^0$ marginal distributions.

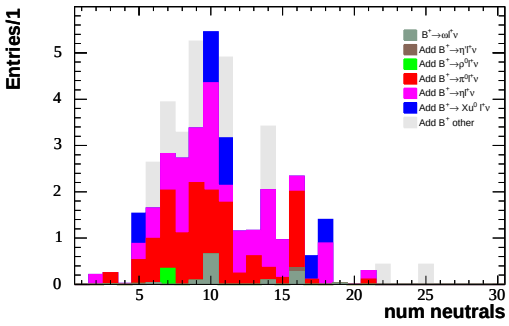
Figure A.44. Distributions of the number of charged tracks in the signal region for the muon channel, validation sample.



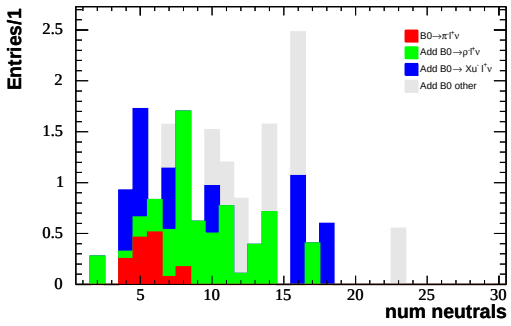
(a) Summary of marginal distributions. Signal BF $= 3 \times 10^{-5}$.



(b) Continuum marginal distributions.



(c) Generic B^+B^- marginal distributions.



(d) Generic $B^0\overline{B}^0$ marginal distributions.

Figure A.45. Distribution of the number of neutrals in the signal region for the muon channel, validation sample.

Appendix B

Efficiency Tables

In this appendix we provide the efficiency tables for the electron- and muon-channel analyses. The tables are broken up by double lines into two sections. The first section gives the original event counts (chapter 5), the effect of the training/validation split (50%/50%), and the effect of the preliminary event selection. The second section gives the efficiencies of the final optimized cuts.

B.1 Electron Channel

In this appendix we provide the efficiency tables for the electron channel, validation sample. The m_{ES} and nuEP cuts are the optimized signal region cuts given in table 9.1. The rest of the cut values are given in table 8.2. The final event counts in the Sequential column of the nuEP row correspond to the template values in the "S" region of the signal extraction fit, up to a factor of two for the training/validation split.

Table B.1. Electron-channel signal-mode selection efficiencies in the nuEP and $m_{\rm ES}$ signal region (described in section 9.3). The preliminary selection is described in chapter 7.

Cut	Overall		Marginal		Sequential	
Description	#Events	Eff(%)	#Events	Eff(%)	#Events	Eff(%)
Original number	695.21	–	695.21	–	695.21	–
Testing sample split	347.61	–	347.61	–	347.61	–
After prelim. selection	156.66	45.07	156.66	–	156.66	45.07
photon lab θ	155.31	44.68	10.64	99.53	155.31	44.68
scaled ν lab θ	143.70	41.34	10.72	98.81	142.44	40.98
unscaled ν lab θ	139.94	40.26	10.86	97.52	134.02	38.56
$e^{\pm}$ $\cos\theta_{\rm Lab}$	144.97	41.70	11.10	95.46	126.35	36.35
$\lvert\cos\theta_T\rvert$	153.91	44.28	10.59	99.99	124.04	35.68
$R2_{\rm All}$	156.66	45.07	10.59	100.00	124.04	35.68
$\mathcal{F}_{\rm E}$	88.45	25.44	16.76	63.22	70.46	20.27
$\cos\theta_{\ell\gamma}$	120.61	34.70	11.38	93.13	55.52	15.97
$\cos\theta_{BY}$	153.59	44.18	10.74	98.65	54.90	15.79
photon conv. veto	154.82	44.54	10.70	98.99	54.44	15.66
electron J/ψ veto	153.01	44.02	10.72	98.83	53.25	15.32
lepton CM E	130.20	37.46	11.73	90.29	47.30	13.61
photon CM E	126.40	36.36	11.52	91.99	41.44	11.92
gLAT	154.07	44.32	10.67	99.28	41.16	11.84
e two-photon parameter	118.40	34.06	11.04	95.97	36.22	10.42
ΔE	154.42	44.42	10.60	99.97	36.16	10.40
clpi0	117.15	33.70	12.90	82.15	29.69	8.54
$m_{\rm ES}$	63.86	18.37	20.20	52.45	15.11	4.35
nuEP	104.32	30.01	15.11	70.10	10.59	3.05

B.2 Muon Channel

In this appendix we provide the efficiency tables for the muon channel, validation sample. The $m_{\rm ES}$ and nuEP cuts are the optimized signal region cuts given in table 9.3. The rest of the cuts are given in table 8.3. The final event counts in the Sequential column of the nuEP row correspond to the template values in the "S" region of the signal extraction fit, up to a factor of two for the training/validation split.

Table B.2. Muon-channel signal-mode selection efficiencies in the nuEP and $m_{\rm ES}$ signal region (described in section 9.3). The preliminary selection is described in chapter 7.

Cut	Overall		Marginal		Sequential	
Description	#Events	Eff(%)	#Events	Eff(%)	#Events	Eff(%)
Original number	695.21	–	695.21	–	695.21	–
Testing sample split	347.61	–	347.61	–	347.61	–
After prelim. selection	117.19	33.71	117.19	–	117.19	33.71
photon lab θ	116.23	33.44	6.59	99.51	116.23	33.44
scaled ν lab θ	107.69	30.98	6.66	98.45	106.83	30.73
unscaled ν lab θ	104.26	29.99	6.75	97.16	100.02	28.77
$\mu^{\pm}$ $\cos\theta_{\rm Lab}$	116.91	33.63	6.57	99.84	99.81	28.71
$\lvert\cos\theta_T\rvert$	102.41	29.46	6.56	100.00	87.09	25.05
$R2_{\rm All}$	117.19	33.71	6.56	100.00	87.09	25.05
$\mathcal{F}_{\rm M}$	53.29	15.33	12.21	53.73	45.09	12.97
$\cos\theta_{\ell\gamma}$	95.50	27.47	7.07	92.79	37.15	10.69
$\cos\theta_{BY}$	110.58	31.81	6.89	95.24	35.49	10.21
muon J/ψ veto	115.08	33.11	6.64	98.86	34.74	9.99
lepton CM E	105.61	30.38	7.01	93.60	32.40	9.32
photon CM E	105.12	30.24	6.84	95.86	30.49	8.77
gLAT	115.40	33.20	6.60	99.45	30.27	8.71
μ two-photon param.	91.20	26.24	6.76	97.11	26.49	7.62
ΔE	107.38	30.89	6.65	98.58	25.60	7.36
clpi0	83.25	23.95	8.38	78.29	19.99	5.75
$m_{\rm ES}$	43.43	12.49	14.12	46.47	8.82	2.54
nuEP	83.09	23.90	8.82	74.35	6.56	1.89