SLAC-R-1027

A Search for the Rare Decay B^0 \to gamma gamma

By

Andrew Michael Ruland

2010

SLAC National Accelerator Laboratory, Stanford University, Stanford CA 94309

Work funded in part by DOE Contract DE-AC02-76SF00515

The Dissertation Committee for Andrew Michael Ruland
certifies that this is the approved version of the following dissertation:

A Search for the Rare Decay $B^0 \to \gamma\gamma$

Committee:

Jack L. Ritchie, Supervisor

Karol Lang

Roy F. Schwitters

Gerald Hoffmann

David Lambert

A Search for the Rare Decay $B^0 \to \gamma\gamma$

by

Andrew Michael Ruland, B.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2010

For my family

Acknowledgments

Like any large undertaking, the work presented here would not have been possible without the support of many through encouragement and discussion. First and foremost, I want to thank my advisor, Jack Ritchie, for all his efforts in making this analysis come to fruition. Not only through the many discussions we had about this analysis and physics in general but also for the opportunities allowed to me being on site at the SLAC National Accelerator Laboratory during the last four years.

Thanks go to all my fellow graduate students and post-docs that provided great friendship and fantastic lunch conversations and most of all tech support when I had no idea what was going on: Joe Tuggle, Bryan Fulsom, Jake Anderson, Jim Hirschauer, Tim Piatenko, Karsten Koeneke, Tae Min Hong, Steve Sekula, Sudan Paramesvaran, and Mark Tibbets. A special thanks is reserved for my fellow Texans Chris Schilling and Brad Wray for putting up with me day in and day out in our ever changing office conditions.

And although not directly a part of this document, the time I spent as the EMC operations manager was probably the most interesting and enjoyable part of my time on *BABAR*. Many thanks need to go to the team of experts who helped cover my short comings when it was needed most: Martin Kocian, Jong Yi, Tim West, Chris West, Jen Watson, and Roberto Sacco (for excellent preparation).

And finally to my family - Mike, Anne and Beth. Without their constant encouragement, love and support I am positive I would not have had the

drive to complete this part of my life.

A Search for the Rare Decay $B^0 \to \gamma\gamma$

Publication No. _____

Andrew Michael Ruland, Ph.D.
The University of Texas at Austin, 2010

Supervisor: Jack L. Ritchie

This thesis describes a search for the rare radiative decay $B^0 \to \gamma\gamma$, where the charged congugate mode is implied throughout. These decays are highly suppressed in the Standard Model where the branching fraction is expected to be of order 10^{-8}. In some new physics scenarios this could be enhanced by up to an order of magnitude to 10^{-7}. Therefore an observation of a significant signal above the Standard Model prediction could be a sign of new physics.

The search for this rare decay was performed using the data collected with the *BABAR* detector at the SLAC National Accelerator Laboratory PEP-II storage ring operating at the $\Upsilon(4S)$ resonance. The analysis uses a dataset with an integrated luminosity of 425.7 fb^{-1} corresponding to 467 million $B\overline{B}$ pairs. A signal yield of $21.3^{+12.8}_{-11.8}$ events with a significance of $1.88\,\sigma$ was measured using an unbinned extended maximum likelihood fit. An upper limit on the branching fraction is set at the 90% confidence level of $\mathcal{B}(B \to \gamma\gamma) < 3.2 \times 10^{-7}$. This is about two times more stringent than the best upper limit of less than 6.2×10^{-7} published by the Belle collaboration.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter introduces the Standard Model of particle interactions, the CKM matrix for quark mixing, and the dynamics of electroweak decays which are responsible for the decay $B \to \gamma\gamma$. It then describes the physics of e^+e^- collisions in relation to B physics and concludes with a description of the theoretical and experimental background of the decay $B \to \gamma\gamma$.

1.1 The Standard Model

To understand the structure of natural world three questions need answering:

1. What are the particles that comprise matter?
2. What are the forces that those particles interact by?
3. How can the interactions of the particles and forces be calculated?

The Standard Model (SM) is the present theoretical framework that answers these questions. It describes the particles that make up matter (quarks and leptons), the forces through which the particles interact called *strong*, *weak*, and *electromagnetic*, and how to calculate these interactions. Together with the general relativistic description of gravity (GR), the two theories give an almost complete description of the natural world.

The SM is a quantum field theory where the interactions between the particles arise due to the local gauge symmetry of the Lagrangian under the group $SU(3) \times SU(2) \times U(1)$. The $SU(3)$ subgroup governs the strong interactions while the $SU(2) \times U(1)$ subgroup describe the combined weak and electromagnetic interactions. Each subgroup is associated with a number of spin-1 gauge bosons that mediate the interactions between particles, where the number is equal to the number of generators of the group structure. The strong force is mediated by eight massless gluons, g, the weak force by the massive $W^{\pm}$ and Z^0 bosons, and the electromagnetic force by a single massless photon, γ.

The Lagrangian, $\mathcal{L}$, of a massless fermion field is given by

$$\mathcal{L}_0 = i\bar{\psi}_n \gamma^\mu \partial_\mu \psi_n, \tag{1.1}$$

where repeated Greek indices are summed over, γ^μ are the Dirac gamma matrices and ψ_n represents a quark or lepton field. By requiring that $\mathcal{L}$ obey various symmetries other terms are added that represent the various interactions. Finally, to account for the masses of the fermions and gauge bosons, two scalar Higgs fields are introduced.

The force responsible for the decay analyzed in this thesis, $B \to \gamma\gamma$, is the electroweak force where the interactions are based on the combined symmetry group $SU(2) \times U(1)$. The theory was developed in the 1960s by Glashow, Salaam, and Weinberg and is given the name GSW theory [1–3]. The three generators of the $SU(2)$ group correspond to a triplet of vector bosons that couple to weak isospin quantum numbers

$$\mathbf{W}^\mu = \begin{pmatrix} W_1^\mu \\ W_2^\mu \\ W_3^\mu \end{pmatrix}. \tag{1.2}$$

while the $U(1)$ generator contributes a single vector boson, B^μ, that couples to weak hypercharge. The corresponding electroweak interaction term in the Lagrangian is given by

$$\mathcal{L}_{EW} = -i[g_W \bar{\chi}\gamma^\mu(1-\gamma^5)\chi\boldsymbol{\tau}\cdot\mathbf{W}^\mu + \frac{g'}{2}\bar{f}\gamma_\mu f B^\mu] \tag{1.3}$$

where g_W and g' are the $SU(2)$ and $U(1)$ coupling constants, γ^μ are the Dirac gamma matrices, $\boldsymbol{\tau}$ are the three Pauli matrices, and f are the up and down components of the $SU(2)$ weak doublets, χ, of quarks and leptons. The three generations of the quark doublets and singlets include

$$Q = \begin{pmatrix} u \\ d \end{pmatrix}_L \begin{pmatrix} c \\ s \end{pmatrix}_L \begin{pmatrix} t \\ b \end{pmatrix}_L \quad \text{and } q = u_R, d_R, s_R, c_R, b_R, t_R$$

while the three lepton weak doublets consist of a massive lepton and its associated neutrino and the charged singlets

$$L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \quad \text{and } \ell = e_R, \mu_R, \tau_R.$$

where L means left-handed and R means right-handed.

The algebra of the $SU(2)$ weak isospin is identical to that of spin-$\frac{1}{2}$ fermions. In a similar fashion to the construction of the raising and lower operators in spin space, the first and second component of the W triplet can be combined to produce charge raising and lowering operators that correspond to the physical charged vector bosons $W^\pm$. All weak isospin fermion doublets are invariant under this transformation. The charged current interaction is a vector-minus-axial interaction, and the Lagrangian for this interaction becomes

$$\mathcal{L}_{CC} = g_W[\bar{f}\gamma^\mu(1-\gamma^5)W^+_\mu f + \bar{f}\gamma^\mu(1-\gamma^5)W^-_\mu f]. \tag{1.4}$$

The $1 - \gamma^5$ term operates on the isospin doublets, and projects out the left-handed component. This shows that the charged weak interaction operates exclusively on the weak isospin doublets.

In the GSW theory, the $SU(2) \times U(1)$ symmetry is broken resulting in the mixing of the W_μ^3 with the B_μ through the weak angle, θ_W. They form two linear combinations that correspond to the massive neutral boson mediating the neutral weak interaction, Z_μ, the massless photon that mediates the electromagnetic interaction, A_μ,

$$\begin{aligned} A_\mu &= B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \\ Z_\mu &= -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \end{aligned} \tag{1.5}$$

The origin of this symmetry breaking is the Higgs mechanism [4–6]. It is responsible for the masses of the fermions and the $W^\pm$ and Z^0 gauge bosons that mediate the weak force. The details of this process are beyond the scope of this thesis. The Higgs boson, the gauge boson associated with this symmetry breaking is yet to be observed in 2010 but its observation is the main goal for the Large Hadron Collider at CERN, currently beginning to take data.

1.2 The CKM Mechanism

The charged currents described in the previous section seem to operate only within generations of fermions (*i.e.* $u \leftrightarrow d$). Experimentally it is observed that generation changing charged currents do exist as evidenced by the decay $K^+ \to \mu^+\nu_\mu$, which is an example of a $s \to u$ transition. The description of these flavor changing processes in the quark sector is explained using the Cabbibo-Kobayashi-Maskawa (CKM) mechanism [7, 8].

The CKM mechanism states that the flavor eigenstates that participate in the weak interactions are not the same as the mass eigenstates of the Hamiltonian. Stately differently, the physical particles that are measured are actually mixtures of different flavor eigenstates. The relation between the mass and flavor eigenstates is given by the CKM matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_{mass} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.6}$$

When this new term is included into charged weak interaction term of the Lagrangian, eq. (1.4) becomes.

$$\mathcal{L}_{charged} = g_W[\bar{u}_i^L \gamma^\mu W_\mu^+ V_{ij} d_j^L + \bar{d}_i^L \gamma^\mu W_\mu^- V_{ij}^* u_j^L]. \tag{1.7}$$

The strengths of the flavor changing charged currents are then determined by the magnitudes and phases of the off-diagonal elements of the CKM matrix.

The CKM matrix is a 3×3 complex unitary matrix. A unitary 3×3 matrix can be parameterized in terms of three real parameters and six complex phases. The three real parameters are analogous to the angles in a three dimensional rotation. For the complex phases, five can be removed through redefinitions of the relative phases between the quark fields leaving a single complex phase. If this phase is non-zero this allows for an asymmetry under the combined operation of charge conjugation (C) and parity reversal (P). The CKM matrix can thus be parameterized using the following notation

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1.8}$$

where $c_{ij} = \cos\theta_{ij}$, $s_{ij} = \sin\theta_{ij}$, and δ is the complex phase.

The three angles, θ_{ij}, characterize the amount of mixing between the quark generations. They are experimentally found to be small: $s_{12} \approx 0.22$, $s_{23} \approx 0.05$, and $s_{13} \approx 0.01$ which allows for an expansion, due to Wolfenstein [9], of the CKM elements in (1.8) in the parameter λ that makes clear their relative magnitudes:

$$V_{CKM} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4), \tag{1.9}$$

The parameters A, ρ, and η are of order unity, with η parameterizing *CP* violation in the Standard Model.

There are six equations that result from the unitarity constraint on the CKM matrix:

$$\sum_i V_{ij} V_{ik}^* = 0. \tag{1.10}$$

Each equation represents a triangle in the complex plane whose area is related to the *CP* violating phase. The important equation for B physics is where $k = b$, representing the decay of a b type quark and given by the equation:

$$V_{ud} V_{ub}^* + V_{cd} V^* V_{cb} + V_{td} V_{tb}^* = 0 \tag{1.11}$$

Using the Wolfenstein parameterization of this equation, the triangle can be represented in the complex $\rho - \eta$ plane as shown in Figure 1.1. The experimental measurement of the angles and sides of this triangle is a stringent test of *CP* violation due to the CKM picture of the Standard Model.

1.3 Flavor Changing Neutral Currents

Flavor Changing Neutral Currents (FCNC) are processes that change the flavor of a fermion without changing the charge, e.g., $b \to d$. These

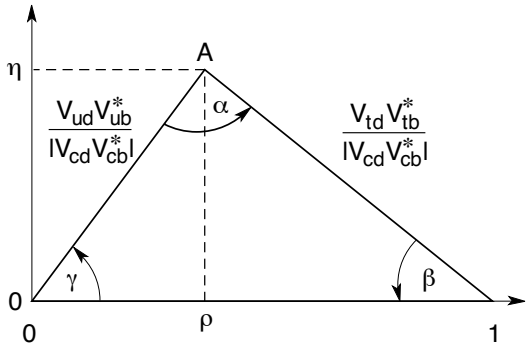


Figure 1.1: The unitarity triangle of the B meson in the complex $\rho - \eta$ plane.

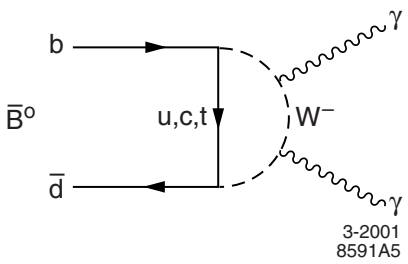


Figure 1.2: Standard Model Feynman diagram for the decay $\bar{B}^0 \to \gamma\gamma$, which is also a flavor changing neutral current transition.

transitions do not occur in the SM at tree level but are allowed in higher-order processes involving a charged W and quark loop. As such, they are highly suppressed compared to other weak interactions by means of the GIM mechanism [10]. The decay $B \to \gamma\gamma$ is one such example and is called an effective flavor-changing neutral current. A Feynman diagram showing this decay along with the $W - q$ loop is shown in Figure 1.2. These types of decays are suppressed by at least a factor of g_W compared to tree level weak decays due to the extra vertex factor associated with the loop. An additional suppression occurs due to the small size of the off-diagonal CKM factor in the

decay amplitude Although the combined suppressions make FCNC processes very rare, they are important because the rates of these processes are sensitive to new physics. The particles from new physics scenarios can take the place of the SM model bosons in the loop at the same level and have a measurable effect on the decay rates. Therefore a measurement of the rate at an enhanced level could provide hints at the type of physics beyond the SM. The work presented in this thesis is an attempt to measure the branching fraction of the flavor changing neutral current $B \to \gamma\gamma$.

1.4 e^+e^- Physics

The two mile long linear accelerator at SLAC National Accelerator Laboratory accelerates electrons and positrons to high energies for use in colliding beam experiments. When the electrons and positrons are at sufficient energies to produce the desired physics, they are each injected into storage ring with a circumference of 2.2 km. For *BABAR* these energies correspond to 9.0 and 3.1 GeV for the electron and positron beam, respectively. Each ring circulates the electrons and positrons in opposite directions and brings them into collision at the heart of the *BABAR* detector called the interaction point (IP). Chapter 2 goes into the details of the PEP-II accelerator and *BABAR* detector.

The electrons and positrons at *BABAR* are brought into collision with a center-of-mass energy of 10.58 GeV, which corresponds to the mass of the $\Upsilon(4S)$ meson. The label Υ refers to a meson composed of a b and $\bar{b}$ quark. The $(4S)$ label refers to the fourth excited state of this resonance which is similar to an excited state in a hydrogen atom. Figure 1.3 shows the energies of the different Υ resonances. The $4S$ resonance energy is important for B physics because the energy is large enough to produce a light $q\bar{q}$ pair, allowing the

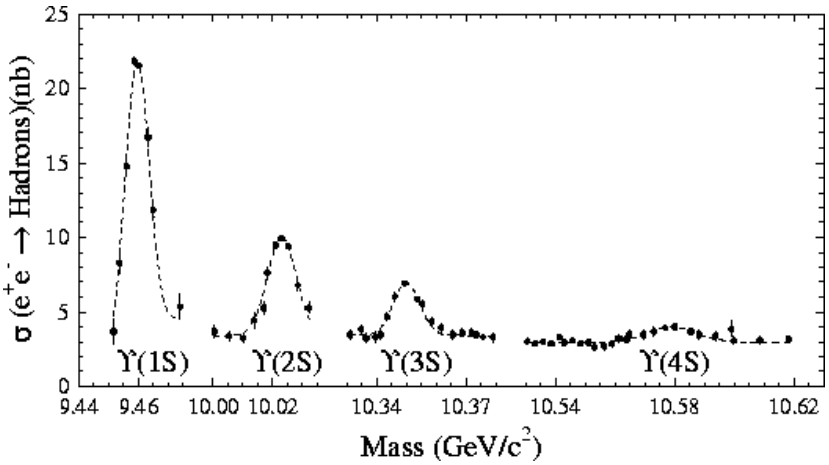


Figure 1.3: Upsilon resonances as a function of the center-of-mass-energy. The $\Upsilon(4S)$ resonance is important because there is enough energy to be able to produce a pair of B mesons.

Table 1.1: Phyical properties of B mesons and their quark content.

	B^+ $(u\bar{b})$	B^- $(\bar{u}b)$	B^0 $(d\bar{b})$	$\bar{B}^0$ $(\bar{d}b)$
mass (GeV/c^2)	5.2792 ± 0.0003		5.2795 ± 0.0003	
lifetime (τ) (s)	$(1.638 \pm 0.011) \times 10^{-12}$		$(1.525 \pm 0.009) \times 10^{-12}$	

$\Upsilon(4S)$ to decay to a pair of B mesons. Some physical properties of B mesons are given in Table 1.1

The use of an electron-positron collider provides many advantages over hadron colliders for precision measurements. The interaction of an electron and positron produces a virtual photon or Z^0 boson which then immediately decays into the particles that can be detected by *BABAR*. In contrast, in a hadron collider like in the LHC, the interactions are governed by QCD processes, where the colliding hadrons fragment creating a large number of secondary particles

that flow along jets. This makes the background for lepton colliders much lower than for hadron collisions, allowing for complete event reconstruction.

1.5 The Decay $B \to \gamma\gamma$

In the standard model, the decay $B \to \gamma\gamma$ proceeds through the second order radiative weak transition $b \to d\gamma\gamma$, followed by an annihilation. An example Feynman diagram for this process is shown in Figure 1.2. This mode is theoretically interesting because it allows for a study of the non-trivial QCD effects of the B decay with a kinematically simple hadron-less two-photon final state. Additionally, since the two-photon final state can be in a *CP*-even or *CP*-odd final state, this decay could be a non-traditional channel to study *CP* violation.

The physics of heavy quark transitions is often done in the framework of Operator Product Expansion (OPE) [11]. This is based on the assumption that the short-distance perturbative effects in QCD can be separated or factored out from the long-distance non-perturbative piece. The effective Hamiltonian of the transition can then be written as the sum over products of coefficients, $\mathcal{C}_i(\mu)$, called Wilson coefficients and local operators, $\mathcal{O}_i(\mu)$. The $\mathcal{C}_i(\mu)$ characterize the short-distance perturbative QCD effects above the energy scale μ. The operators, $\mathcal{O}_i(\mu)$, then characterize the long-distance non-perturbative physics below the μ energy scale. This parameterization is said to "integrate out" the particles of rest mass greater than μ. For B decays this energy scale is the mass of the b-quark, m_b, and it integrates out contributions from particles that are heavy compared to m_b such as the W, Z and t quark. The effective Hamiltonian for this scheme applied to a $B \to \gamma\gamma$ transition can be written

as [12]

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_j V_{jd}^* V_{jb} \left[\mathcal{C}_1^j \mathcal{O}_1^j + \mathcal{C}_2^j \mathcal{O}_2^j + \sum_{i=3}^{8} \mathcal{C}_i \mathcal{O}_i \right] + \text{h.c.} \tag{1.12}$$

where j is a quark flavor index representing either a u or c quark, G_F is the Fermi coupling constant and V is a CKM matrix element. The most important terms for and $b \to d\gamma\gamma$ transition is the seventh term, $\mathcal{C}_7\mathcal{O}_7$, called the magnetic penguin operator, and the four-quark operators, $Q_{1,2}^j$. Using this formalism a estimate of the branching fraction for $B \to \gamma\gamma$ is calculated to be

$$\mathcal{B}(B \to \gamma\gamma) = (3.1^{+6.4}_{-1.6}) \times 10^{-8}. \tag{1.13}$$

1.5.1 New Physics Enhancements

The small size of the *BABAR* dataset makes a measurement of the branching fraction of $B \to \gamma\gamma$ at the SM level unlikely. However, several new physics scenarios can provide enhancements to the rate by up to an order of magnitude. Since flavor changing neutral currents occur beginning at the one-loop level in the SM, new heavy particles such as a non-SM Higgs or a SUSY particle can enter the loop in place of the up-type quark. These new contributions would add extra terms to the amplitude in the and could alter the rate. Therefore any difference in the SM and observed rates could provide evidence for new physics.

One scenario showed that with the addition of a second weak Higgs doublet, the branching fraction could be enhanced up to the level of 10^{-7} [13]. Another calculation of the branching fraction uses a SM scenario but adds an additional space dimension. In this case the exchange of scalar Higgs particles within the loop gives an enhancement of the branching fraction of about 6% [14].

Table 1.2: Current published experimental results for $B \to \gamma\gamma$ from *BABAR* and Belle. All upper limits are given at 90% CL.

Experiment	$\mathcal{B}(B \to \gamma\gamma)$
L3 Collaboration [17]	$< 3.9 \times 10^{-5}$
BABAR [16]	$< 1.7 \times 10^{-6}$
Belle [15]	$< 6.2 \times 10^{-7}$

1.5.2 Previous Searches

The branching fraction for the decay $B \to \gamma\gamma$ has not been measured as of 2010. Several previous attempts to observe this decay mode have been attempted, but in each case an upper limit was set. The most recent result comes from Belle analyzed using a dataset corresponding to an integrated luminosity of 104 fb^{-1}. An upper limit on the branching fraction was set at less than 6.2×10^{-7} at the 90% CL [15]. A previous search for this decay by the *BABAR* Collaboration was done using a smaller dataset corresponding to an integrated luminosity of $20\,\text{fb}^{-1}$. Again, the results were consistent with zero signal events and an upper limit of less than 1.7×10^{-6} at 90% CL was set [16]. The previous results are shown in Table 1.2. This thesis will describe the search for the decay $B \to \gamma\gamma$ using the *BABAR* dataset whose integrated luminosity of 425.7 fb^{-1} is four times larger than was used for the current best upper limit.

Chapter 2

The B Factory and the *BABAR* Detector

To study the physics processes described in the Chapter 1, large numbers of B mesons must be produced. The B-Factory, consisting of the PEP-II accelerator and the *BABAR* detector, located at the SLAC National Accelerator Laboratory[1] in Menlo Park, CA was designed and constructed for this purpose. This chapter will introduce the experimental apparatus used to produce B mesons and detect their decay products.

2.1 The PEP-II Accelerator

The PEP-II (Positron-Electron Project II) accelerator is an asymmetric circular e^+e^- collider shown in Figure 2.1. It consists of a high energy ring (HER) of 9.0 GeV electrons and a low energy ring (LER) of 3.1 GeV positrons circulating in opposite directions. The beams are brought into collision near the center of the *BABAR* detector at a center-of-mass energy of 10.58 GeV which corresponds to the mass of the $\Upsilon(4S)$. The $\Upsilon(4S)$ decays nearly 100% of the time into a pair of $B\overline{B}$ mesons ($m_B = 5.279\,\mathrm{GeV}/c^2$). The rest frame of the $\Upsilon(4S)$ is Lorentz boosted from the lab frame by a factor of $\beta\gamma = 0.56$ because of the asymmetric energies of the beams. This boost causes the $B\overline{B}$ pair to travel far enough apart before they decay for the tracker to be able separate

[1] The name of the lab was changed in 2008 from the Stanford Linear Accelerator Center.

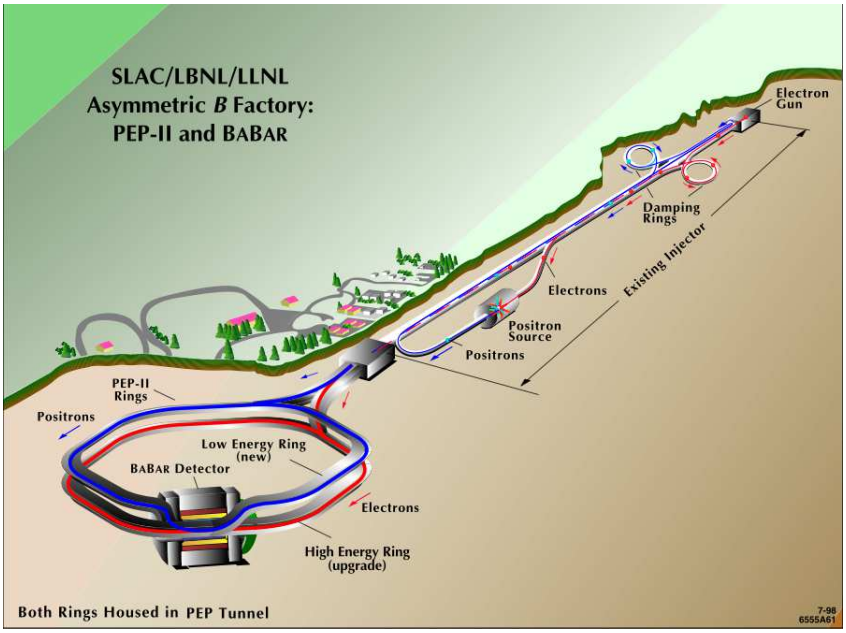


Figure 2.1: PEP-II is the circular blue/red ring.

the decay vertices. This ability to separate the vertices is necessary for the measurement of time-dependent *CP* violating decays.

2.2 The *BABAR* Detector

The *BABAR* detector was designed for the primary physics goal of measuring *CP* violating asymmetries in the decays of neutral B mesons. Additionally, with the high luminosity achieved by PEP-II, large samples of charm and τ events are available for precision studies as well as for rare decay searches, including the decay $B \to \gamma\gamma$. To achieve this large physics program the detector was designed to satisfy the following requirements

- A large geometrical acceptance including small polar angles relative to

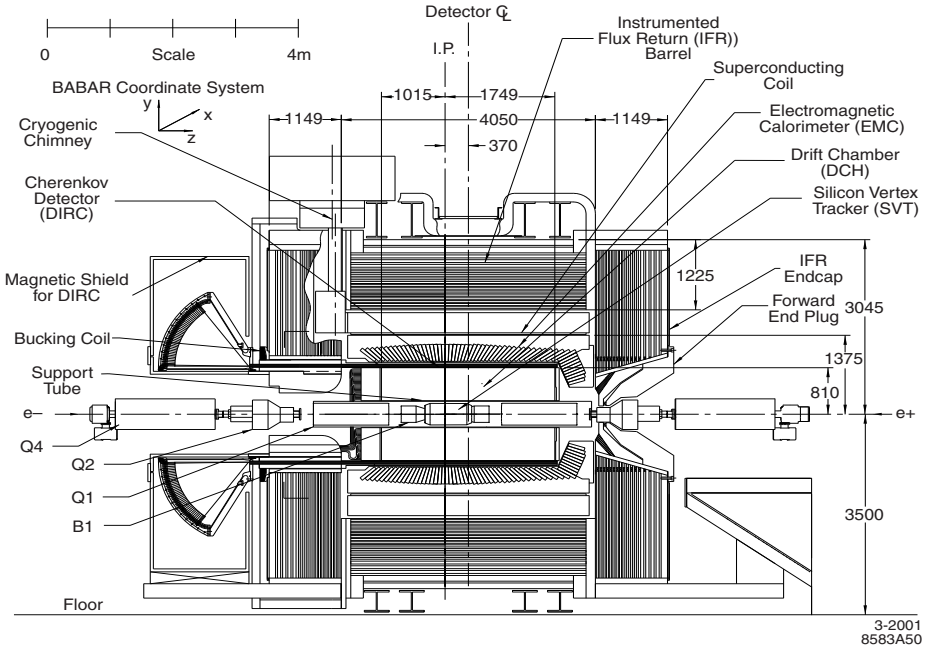
the boost direction;

- Excellent reconstruction efficiency for charged particles down to 60 MeV/c and for photons to 20 MeV;
- Very good momentum resolution to separate small signals from background;
- Good energy and angular resolution for photon detection from π^0, η, and radiative decays;
- Good vertex resolution from the tracker;
- Very good particle identification over a wide range of energies.

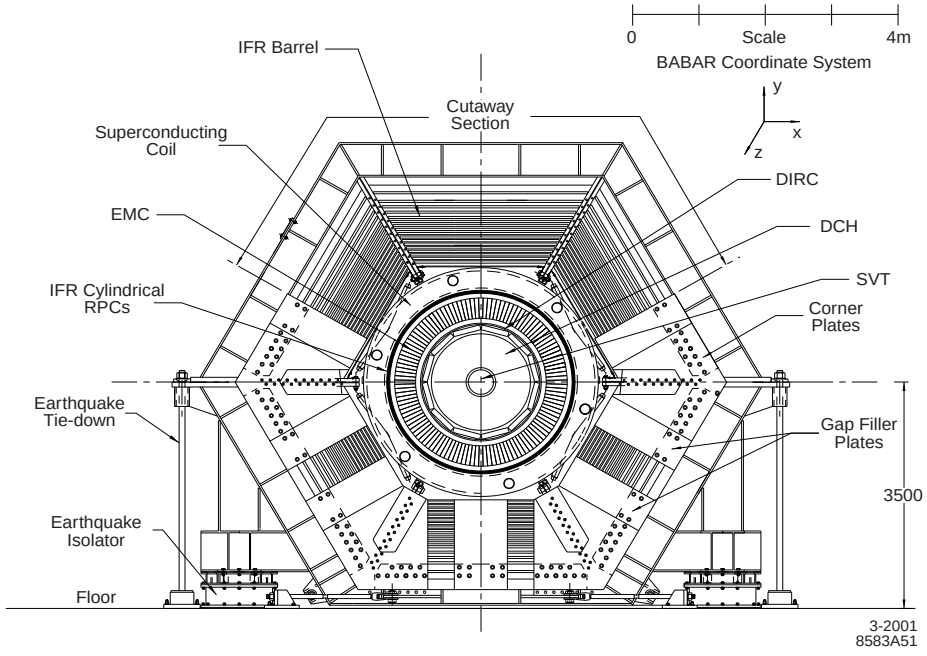
To accomplish all of the requirements, the *BABAR* detector is composed of several individual sub-detectors that work in tandem covering different tasks. The inner detector contains the tracking system, which is made up of a silicon vertex tracker (SVT), and a drift chamber (DCH), a particle ID detector using internally reflected Cherenkov light (DIRC), and an electromagnetic calorimeter (EMC). These components are surrounded by a super conducting magnet producing a field of 1.5T. The outermost sub-detector is steel flux return instrumented for muon and neutral hadron identification. The detector surrounds the interaction region (IR) and is offset in the direction of the LER by 0.37cm in order to maximize the acceptance of the $\Upsilon(4S)$ decay. Figure 2.2 shows the detector in cross section.

2.2.1 SVT: Silicon Vertex Tracker

The inner most sub-detector is the silicon vertex tracker which was designed to provide precise reconstruction of charged particle trajectories and



(a) *BABAR* longitudinal view



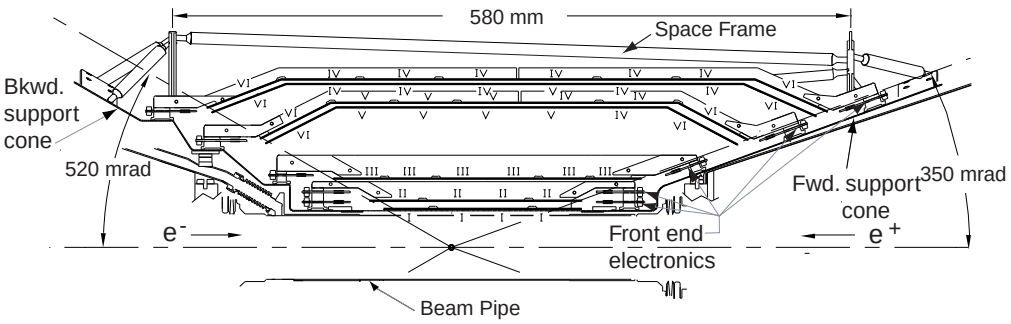
(b) *BABAR* transverse view

Figure 2.2: The *BABAR* Detector.

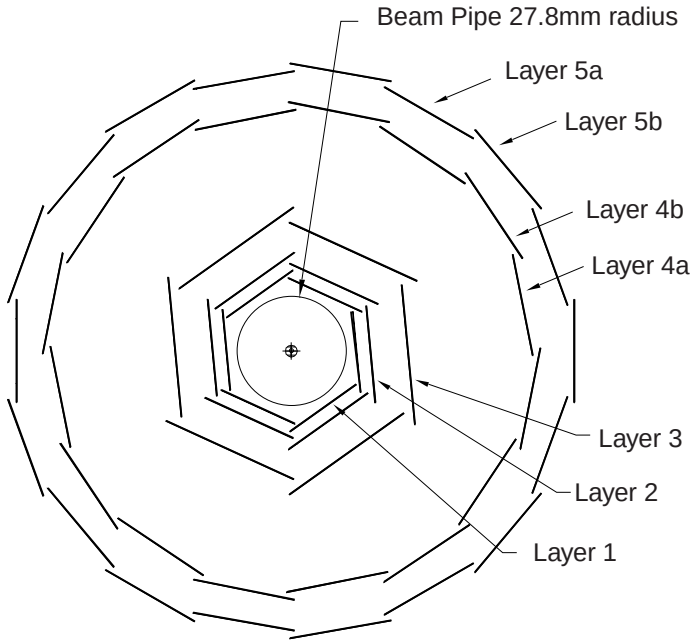
decay vertices near the IP. The SVT is constructed of five layers of two-sided silicon sensors. The five layers are arranged into modules containing 6, 6, 6, 16, and 18 units from inner to outer radius. The sensors on opposite sides of each layer are oriented perpendicularly to each other. For example on one side the strips run parallel to the beam axis and measure the azimuthal angle ϕ, while on the opposite side the strips are transverse to the beam axis and measure the z position. Figure 2.3 shows the design and orientation of the SVT in the longitudinal and transverse planes.

The three inner layers of the SVT are flat strips of silicon mounted as close to the beam pipe as possible to minimize the impact of multiple scattering. The outer layers 4 and 5 have an *arched* design chosen to minimize the amount of silicon necessary for full solid angle azimuthal coverage and to increase the crossing angle for particles near the edge of acceptance. The innermost layer is 3.4 cm in radius compared to the beam pipe radius of 2.78 cm, while the outermost layer is at a radius of 14.4 cm. Figure 2.3(a) shows an asymmetric orientation of the SVT in z because the Lorentz boost has the decay products preferentially moving in the forward direction. The polar angle coverage of the SVT is $20 < \theta < 150$ degrees which corresponds to 90% of the solid angle in the CM frame.

The SVT is made of 104 silicon wafer sensors that are 300 μm thick which constitutes a total area of silicon equal to 0.96 m^2 corresponding to about 4% of a radiation length. The wafers are built upon n-type substrate with p^+ and n^+ strips on opposite sides. When a charged particle passes through a wafer it creates many electron-hole pairs whose charge is read out through the front end electronics. If the total charge is greater than some threshold, typically about 0.4 fC, the event is recorded as a hit.



(a) SVT longitudinal cross section (y-z plane).



(b) SVT transverse cross section (x-y plane).

Figure 2.3: SVT cross section.

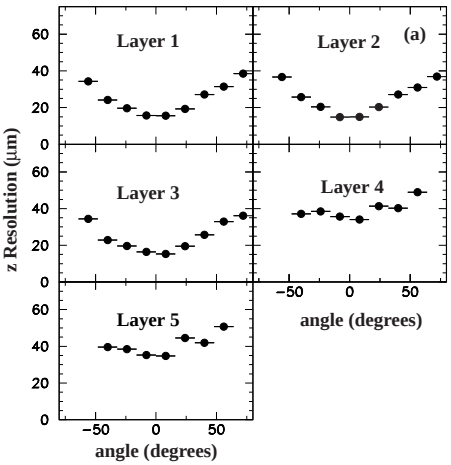
Due to its close proximity to the beam line the SVT is particularly vulnerable to radiation damage. In order to monitor the accumulated radiation dose a set of 12 silicon PIN monitoring diodes and two chemical vapor deposit (CVD) diamond sensors are installed between the beam pipe and innermost layer of the SVT. The radiation sensors have the ability to abort the PEP beams if the radiation in the environment becomes bad enough to have the ability to cause damage to the SVT.

The resolution of the SVT is dependent on the distance between readout strips which for the innermost(outermost) layer is 50(210) μm. The ϕ and z resolution of each layer is shown in Figure 2.4. The mean separation in z between the two B decay vertices is about 250 μm, which the SVT resolution of between 15–40 μm is sufficient to distinguish.

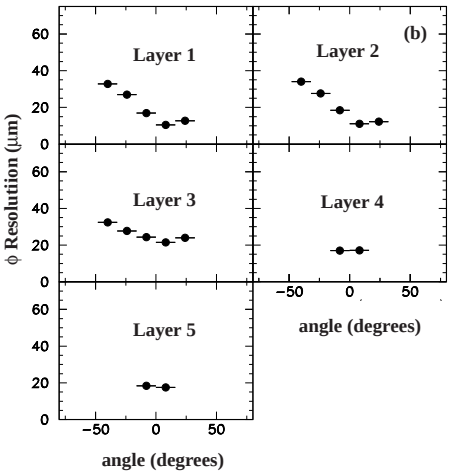
2.2.2 DCH: Drift Chamber

The Drift Chamber is used to measure, with high precision, the trajectories of charged particles and their energy loss due to ionization. It is solely responsible for the reconstruction of decay vertices for particles with lifetimes long enough to travel outside the SVT volume. Together the DCH and the SVT make up the tracking system for *BABAR*. Figure 2.5 shows the DCH in longitudinal cross section.

The DCH is a cylinder 276.4 cm long with an inner(outer) radius of 23.6(80.9) cm, containing 7104 drift cells. The drift cells are arranged into 10 superlayers which each contains 4 layers of cells. Figure 2.6 shows the layout of the innermost 4 superlayers of the drift cells. Each cell is hexagonal in shape with dimensions 11.9 $\times$ 19.0 mm along the radial and azimuthal directions, respectively. Each cell is composed of one tungsten-rhenium sense



(a) z hit resolution



(b) ϕ hit resolution

Figure 2.4: SVT hit resolution.

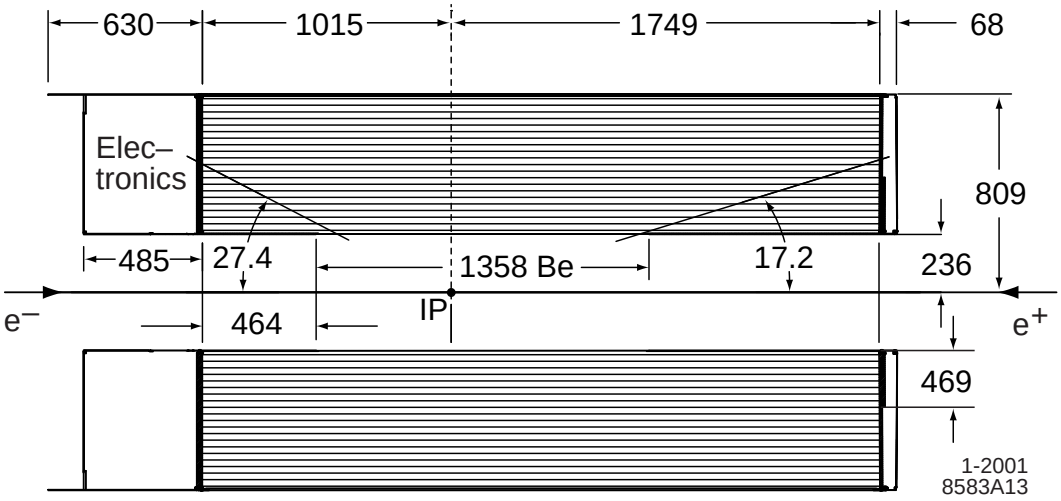


Figure 2.5: Longitudinal cross section of the Drift Chamber with dimensions given in mm and degrees. The center of the DCH is offset 370 cm from the IP.

wire, maintained at high voltage, surrounded by six aluminum wires.

The DCH is filled with an 80:20 mixture of helium and isobutane gas that the charged particles ionize as they traverse the chamber.

2.2.3 DRC: Detector of Internally Reflected Cherenkov Light

The *CP*-violating and rare B decays discussed in the previous chapter require knowledge of the flavor of the B meson in the decay, and so the ability to distinguish between charged pions and kaons is essential. The Detector of Internally Reflected Cherenkov light (DRC) was designed for particle identification above momentum of 700 MeV/c where the DCH is no longer able to distinguish between the two species using dE/dx measurements.

When a charged particle moves through a medium at a velocity greater than the speed of light in that medium it emits photons called Cherenkov light.

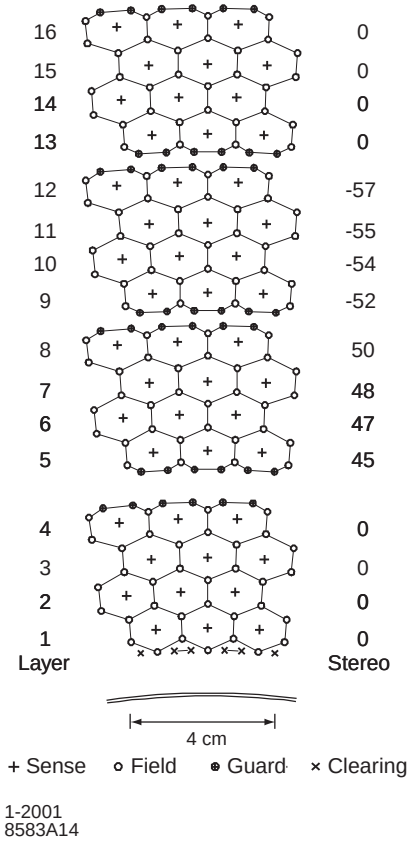


Figure 2.6: Schematic layout of the 4 innermost superlayers. The numbers on the right are the stereo angles (mrad) for the sense wires in each layer. The double solid line at the bottom is the beryllium inner wall of the DCH structure.

The angle between the velocity of the particle and the direction in which the photons are emitted, θ_c, is given by

$$\cos\theta_c = \frac{1}{n\beta} = \frac{c}{nv}, \tag{2.1}$$

where n is the index of refraction of the medium, v is the particle's velocity, and c the speed of light. Different species of particles can be identified by a measurement of their momentum and θ_c.

The DRC concept is based on the principle that the incident and reflected angles of light from a flat surface are equal in magnitude. As a particle passes through the DRC material, Cherenkov photons are emitted and transmitted by total internal reflection to the back of the detector where they are imaged in a standoff box filled with purified water by an array of photomultiplier tubes (PMTs). By using the timing of the electronic signals and the position and angles from the tracking system, the Cherenkov angles θ_c and ϕ_c can be calculated and can be used in the identification of the particle species. Here ϕ_c is the azimuthal angle of a Cherenkov photon around the track direction.

The DRC is constructed of 144 bars arranged into a 12 sided polygon barrel that is coaxial with the *BABAR* z-axis. Each bar is made of fused synthetic silica with index of refraction, $n = 1.473$. The bars are 17-mm-thick, 35-mm-wide, and 4.9-m-long and constructed of 4 smaller bars of length 1.225-m-long glued end to end. Figure 2.7 shows a schematic of the DRC longitudinal cross section and bar box.

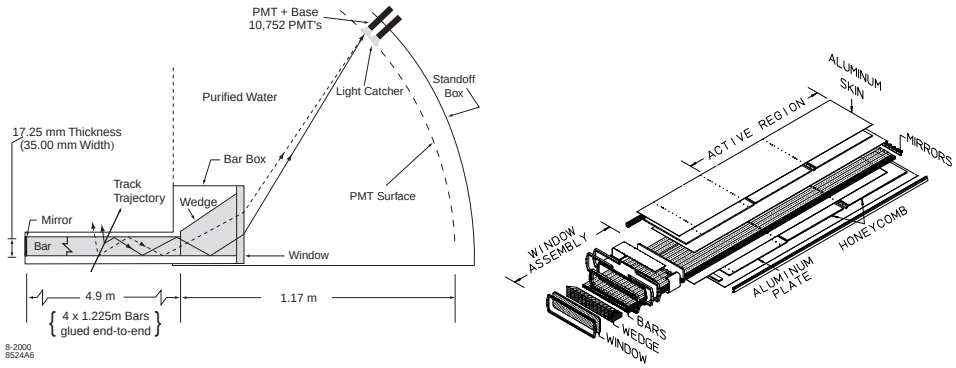


Figure 2.7: The left figure shows a schematic of the DRC longitudinal cross section. The right figure shows a schematic of one of the 12 bar boxes.

2.2.4 EMC: Electromagnetic Calorimeter

2.2.4.1 Design and Layout

The most important system for the measurement of the branching fraction of $B \to \gamma\gamma$ is the electromagnetic calorimeter (EMC). Positioned just outside the DRC radius, it is a total absorption calorimeter responsible for the measurement of photon energies as well as the identification of electrons from other charged tracks by using energy deposition and shower shape along with momentum measurements from the tracking system.

The design goals were excellent EM shower detection efficiency, energy and angular resolution from 20 MeV up to 9 GeV. This allows for the reconstruction of π^0 and η decays into two photons. The electron identification is used for flavor tagging of B mesons decaying into semi-leptonic states, as well as for the study of rare decays of B and D mesons.

The EMC is constructed from 6580 cesium iodide crystals doped with 0.1% thallium, CsI(Tl). These are total-absorption scintillating crystals meaning that the incident particle is destroyed and its energy is converted into

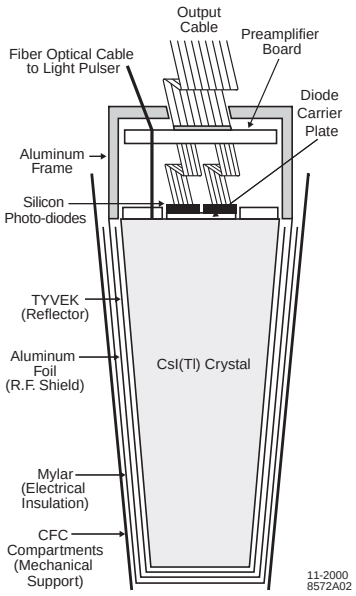


Figure 2.8: A drawing of an EMC crystal and the photodiode readout attached to the back.

scintillation photons. The scintillation photons can be detected using photodiodes and related to the incident particle energy. This material was chosen because its properties allow for very good energy and angular resolution at *BABAR* energies. Additionally its short radiation length allows a small volume of calorimeter to fully contain the EM showers thus reducing the cost and size of the subsystem and detector. The crystals, shown in Figure 2.8, have a trapezoidal cross section to minimize dead space in azimuthal coverage. The crystals range in length from 16.0 radiation lengths, X_0, in the rear of the calorimeter up to 17.5 X_0 in the forward section. This corresponds to crystal lengths of 29.6 to 32.4 cm, respectively. The average dimensions of a crystal front(back) face are 4.7×4.7 cm^2(6.0×6.1 cm^2).

The crystals are arranged in two mechanically independent structures

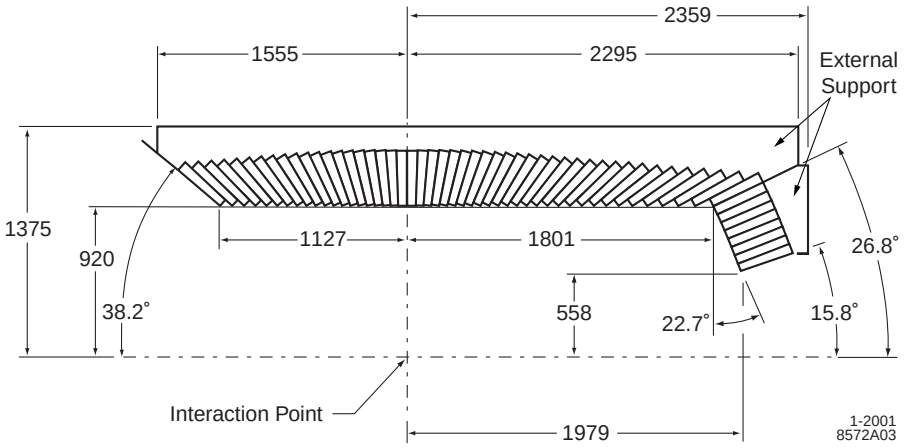


Figure 2.9: A longitudinal view of the EMC showing its' top half with dimensions given in mm. The calorimeter is symmetric around the z-axis which is the horizontal dotted line going through the interaction point.

called the barrel and endcap. A schematic drawing of a longitudinal cross-section is shown in Figure 2.9. The barrel is a cylinder of inner radius 92 cm composed of 48 rings along the beam axis each with 120 crystals. The endcap is a conic section with eight rings along the z-axis each composed of 80, 100, or 120 crystals depending on the radial distance from the beam line. All crystal axes are pointed towards the interaction point with a slight non-projectivity, ranging from 15–45 mrad, to minimize the chance for particles to traverse the gaps between the crystals. In the azimuthal direction the crystals are completely projective. This geometry provides full azimuthal coverage and from 15.8° to 141.8° polar angle coverage which corresponds to 90% solid angle coverage in the center-of-mass frame.

When a photon or electron enters the calorimeter it loses energy mainly through pair production and brehmsstrahlung processes. The energy loss continues as the particle moves through the crystal creating a shower of photons

and e^+e^- pairs until the incident particle energy reaches a critical value which is about 10 MeV in CsI(Tl). The scintillation photons are collected by two silicon photodiodes glued to the rear face of the crystal. These diodes are run at 50 V and have a quantum efficiency of 85% at the peak wavelength of CsI(Tl) scintillation, about 565 nm. Additionally each crystal is wrapped in a reflector material to keep the scintillation light inside the crystal and a aluminum foil layer to shield them from electronic interference.

2.2.4.2 EMC Calibration

The calibration of the EMC is done in two stages. First, the measured photodiode signal is related to the incident energy of individual crystals. This is done at a low and high energy point due to the non-uniformity of the crystal light yields. The crystal response in between is then calculated using a logarithmic interpolation. The low energy single crystal calibration uses a 6.13 MeV photon produced in the decay of a neutron irradiated fluid called Fluorinert through the reaction

$$^{19}\mathrm{F} + \mathrm{n} \rightarrow \ ^{16}\mathrm{N} + \alpha, \quad ^{16}\mathrm{N} \rightarrow \ ^{16}\mathrm{O}^* + \beta, \quad ^{16}\mathrm{O}^* \rightarrow \ ^{16}\mathrm{O} + \gamma.$$

The activated fluid is circulated through the detector passing the front face of the crystals and the response is measured. The high energy single crystal calibration (3-8 GeV) is done using Bhabha events ($e^+e^- \rightarrow e^+e^-$). The calibration constants constrain the deposited energy of the $e^\pm$ to that predicted by a GEANT based MC simulation as a function of polar angle.

Second, a correction for energy loss in dead material, shower leakage, and energy not associated with the cluster is calculated relating the deposited cluster energy to that of the incident electron or photon. In the low energy

region from 70 MeV to 2 GeV, the calibration is done using a sample of symmetric π^0s. Here the invariant mass of the two photons is constrained to be PDG π^0 mass. In the high energy region from 400 MeV to 6 GeV, the calibration is done using a sample of $e^+e^- \to \mu\mu\gamma$ events. The measured photon energy is related to the incident energy by kinematically fitting the $\mu\mu\gamma$ particles.

2.2.4.3 EMC Performance

The performance of the EMC is measured by the resolution. At low energies, the resolution is measured directly from the 6.13 MeV photons used in the single crystal calibration. For high energies the resolution is derived from Bhabha events. An overall fit to the resolution measurements of the EMC results in:

$$\frac{\sigma_E}{E} = \frac{(2.32 \pm 0.30)\%}{\sqrt[4]{E(\,\mathrm{GeV})}} \oplus (1.85 \pm 0.12)\%, \tag{2.2}$$

The angular resolution uses symmetric π^0 and η decays to two photons where the angular distribution can be calculated. The resolution function is found to be equal in polar and azimuthal angle and is given by the relation:

$$\sigma_\theta = \sigma_\phi = \left(\frac{3.87 \pm 0.07}{\sqrt{E(\,\mathrm{GeV})}} + 0.00 \pm 0.004 \right) \mathrm{mrad}. \tag{2.3}$$

2.2.5 IFR: Instrumented Flux Return

The outermost subsystem is the Instrumented Flux Return (IFR). It is designed to detect muons with a high efficiency and neutral hadrons over a wide range of momenta and angles. It also serves to contain the flux of the 1.5 T magnetic field produced by the superconducting solenoid. Good muon

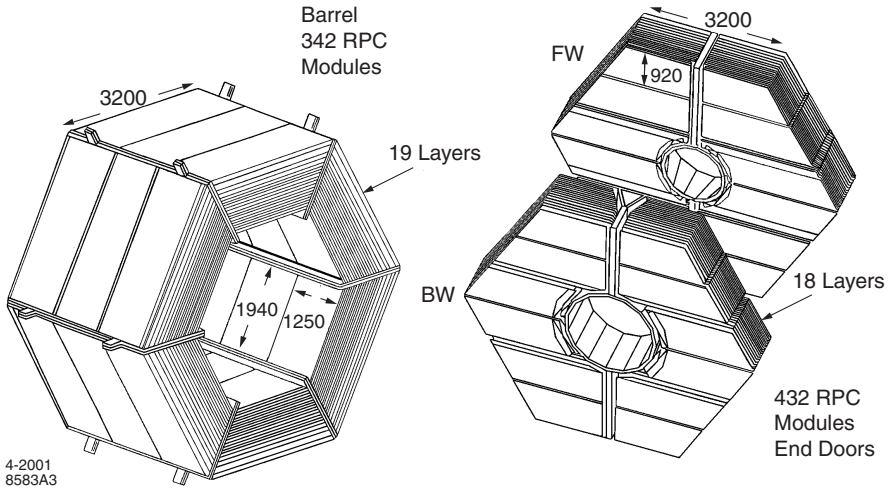


Figure 2.10: Schematic drawings for the IFR barrel and endcap forward (FW) and backward (BW) doors. All dimensions are in mm.

detection is important for tagging the flavor of the B mesons that undergo semi-leptonic decays, as well as rare decays of B and D mesons. The ability to identify K_L^0 also is essential for study of CP eigenstates in exclusive B decays.

The IFR is a hexagonal barrel section with three regions in z and two endcap modules segmented into west and east sides with six vertical regions in each as shown in Figure 2.10. The barrel and endcap modules have a total of 19 and 18 layers of steel interleaved with detectors respectively. The width of the steel layers vary with the radial distance from the IP. The innermost layer of steel is 2 cm in width while the outermost layer has a thickness of 10 cm. This was done as Monte Carlo studies showed this improved muon and K_L^0 identification.

The detectors in between steel layers in both the barrel and endcap were originally resistive plate chambers (RPCs) but in 2004 and 2006 the bar-

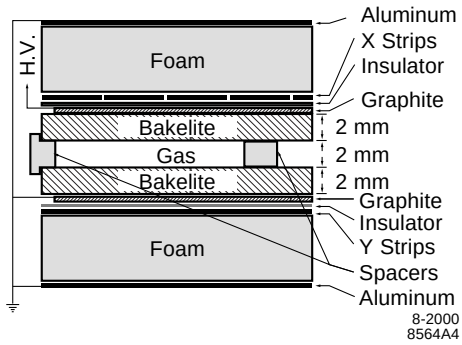


Figure 2.11: Cross section of an RPC.

rel RPCs were replaced with limited streamer tubes (LSTs) due to a degradation in RPC performance. Both types of detectors work by detecting a discharge (streamer) created by a charged particle passing through the gas in each RPC/LST.

The RPCs are made of Bakelite and graphite surrounding a 2-mm gap that contains a mixture of 56.7% argon, 38.8% freon, and 4.5% isobutane. A cross section of a typical RPC is shown in Figure 2.11. The RPCs are capable of readout in two dimensions by having orthogonal aluminum readout strips on either side of the module.

The LSTs are tubes of polyvinyl chloride with a cross-section of 15×17 mm^2, and length 38 cm. The tubes are connected end-to-end in groups of 4 to give a total length of $\sim$ 3.5 m. A flat module is made by placing 8 tubes next to each other into a honeycomb structure. The inner surface of each tube is coated with graphite to act as a cathode, while a 100 μm silver plated anode wire runs down the center and is kept at high voltage, 5500 V. The tube is filled with a gas mixture of 89% carbon dioxide, 8% isobutane, and 3% argon. Again two-dimensional readout is available by running aluminum strips along

the bottom of the tubes, perpendicular to the anode wire directions.

2.2.6 Trigger and Data Flow

The trigger is designed to recognize the signature of interesting physics events among the thousands of interactions that take place each second and save those events to disk for later analysis. It is a two level system that is made of a hardware based selector called the Level 1 trigger (L1), and a software based selector called the Level 3 trigger (L3).

2.2.6.1 Level 1

The Level 1 trigger is a hardware based trigger that uses information from charged tracks in the DCH, clusters in the EMC, and tracks in the IFR to make a decision on whether the event should be considered for further analysis. Each of the three subsystem has their own trigger called the Drift Chamber Trigger (DCT), Electromagnetic Calorimeter Trigger (EMT), and the Instrumented Flux Return Trigger (IFT), respectively. These triggers each generate a summary of the energy and position of tracks or clusters and send the information, called a *primitive*, to the Global Level Trigger (GLT).

The GLT synchronizes the information from the primitives and provides basic matching between the positions of tracks and clusters in the different systems. It then produces specific triggers dependent on the content of the event and passes this information to the Fast Control and Timing System (FCTS). The FCTS can mask triggers or pre-scale them depending of the physics processes that are of interest during running. If a trigger remains after the FCTS, the event is given an L1 Accept signal and is read out for processing by the L3 trigger.

The L1 trigger is designed with an output rate of about 1 kHz at design luminosity, but near the end of the data taking period in 2007/2008 with PEP-II delivering luminosity around 10^{34} cm^{-2}s^{-1} it was operating at about 3.5 kHz. The L1 trigger is greater than 99.9% efficient for $B\overline{B}$ events, greater than 98% efficient for $e^+e^- \to q\overline{q}$ ($q = u, d, s, c$), and greater than 94.5% for $\tau^+\tau^-$ events.

2.2.6.2 Level 3

The Level 3 trigger (L3) is implemented in software and uses the output of the Level 1 trigger as input and has access to all of the event information to determine it's output. L3 performs initial events reconstruction and classification, selection using filters, and monitoring. The L3 trigger is comprised of three phases.

1. Events are classified into L3 input lines by combining the output lines of the FCTS using a logical OR.

2. For each of the L3 input line, a pass/fail flag is assigned to the line.

3. L3 output lines are formed by a logical OR of flags from the input lines that were run in the previous step. If the event passes the L3 output it is written to disk for analysis.

After passing through both L1 and L3 triggers, $B\overline{B}$ events have an efficiency of greater than 99.9%, $e^+e^- \to q\overline{q}$ ($q = u, d, s, c$) have and efficiency greater than 95%, and $\tau^+\tau^-$ events are 92% efficient. At design luminosity the L3 output has about 13% of events useful for physics analysis and 40% are useful for calibration and diagnostics.

2.2.6.3 Data Flow

Events that pass the L3 selection are passed to a program that writes them to an *extended tagged container* (XTC) file. Each file contains about one hours worth of colliding beam data called a *run*. All XTC files are then sent to a computing farm where they undergo further processing and reconstruction.

Initially, some events are taken from the run and used to calibrate the subsystems in a process known as *prompt calibration* (PC). This is done to calibrate each subsystem as the detector conditions changed with time. PC is done on a rolling basis, meaning that the calibrations are occurring whenever *BABAR* is recording data. The second step in this process is called *event reconstruction* (ER), which takes the raw data from the XTC files and turns is into the lists of particles used in physics analysis. A detailed description of this process is given in the next chapter.

Chapter 3

Data Processing

The data recorded by the detector are signals from phototubes, sense wires, photodiodes, resistive plate chambers, and silicon strip sensors. To turn these into objects that can be used in a physics analysis, particles and their associated 4-momentum, requires a large set of software tools. This chapter describes the event reconstruction techniques used in *BABAR* to turn those electronic signals into charged tracks and neutral clusters. Additionally, the algorithm used for B counting is discussed. Lastly, a brief description of the simulation of *BABAR* data used to study specific signal and background processes is discussed.

3.1 Event Reconstruction

All events that pass the L3 trigger are written out to disk for reconstruction. Initially the data passes through a calibration phase called *prompt calibration* (PC) that implements many of the calibrations discussed in the previous chapter. Afterwards, usually within a 48 hours, the data is sent to a computing center in Padova, Italy for event reconstruction. The next sections describe the steps used to turn the raw measurements made by the detector into lists of reconstructed particles for physics analysis. The output of the reconstruction are stored in ROOT [18] collections.

3.1.1 Photon Reconstruction

Photons are detected in the EMC from their scintillation photons after undergoing an electromagnetic shower. The energy is usually contained within a number of adjacent crystals called a *cluster*. Any crystal inside a cluster where the adjacent crystal energies are smaller is referred to as a *bump*. Consequently a cluster can contain more than one bump. Cluster reconstruction starts by finding crystals whose energy is greater than 10 MeV. The cluster is then built by adding neighboring crystals whose energy is at least 1.0 MeV or if they are contiguous with crystals of energy greater than 3.0 MeV. If the cluster contains a single bump then the full energy of the cluster is associated with the bump. For clusters with multiple bumps the detected energy is split between them using an iterative algorithm. This method assigns a weight, w_i, to each crystal and the energy of the bump is given as $E_b = \sum_i w_i E_i$, where i is the index of all crystals in the cluster. The weights are given by the equation

$$w_i = E_i \frac{\exp(-2.5 r_i / r_M)}{\sum_j E_j \exp(-2.5 r_j / r_M)} \tag{3.1}$$

where j is an index referring to all crystals in the cluster, r_j is the radius from the bump center to the center of the j^{th} crystal, and r_M is the Molière radius of CsI(Tl) (3.8 cm). Initially all weights are set to one, iterated through, and the bump centroid is recalculated. The iteration stops when the bump centroid is constant within 1 mm. A bump is associated with a charged track by projecting the track helix onto the front face of the EMC. If the distance between the bump centroid and track impact are consistent with the momentum and angle of the track, the cluster is associated with the charged particle. If no tracks in the event are consistent then the cluster is assumed to be a photon.

Additionally, the shape of the cluster is also relevant and is characterized quantities called the lateral moment (LAT), and the second moment. The

LAT [19] is defined as the ratio of all but the two highest energy crystals in the cluster, weighted by the square of the distance to the cluster center to the sum of all crystal energies in the cluster weighted by the square of the distance between the center of adjacent crystals which is 5.0 cm. This is given by the equation

$$LAT = \frac{\sum_{i=3}^{N} E_i r_i^2}{\sum_{i=3}^{N} E_i r_i^2 + E_1 r_0^2 + E_2 r_0^2}, \quad \text{where } E_1 \geq E_2 \geq \ldots \geq E_N, \tag{3.2}$$

Here E_i is the energy of a crystal in the cluster, r_i is the distance from the i^{th} crystal center to the bump crystal center, and r_0 is the distance between crystal centers.

The second moment is the ratio of the sum of the crystal energies weighted by the square of the angular distance from the center of the cluster to the cluster energy and is given by the equation

$$2^{\text{nd}} \text{ Moment} = \frac{\sum_i E_i \alpha_i^2}{\sum_i E_i}. \tag{3.3}$$

During reconstruction lists of photons are created by selecting clusters based on their energy and lateral moment. All clusters in an event not matched to a track are placed on a list called `CalorClusterNeutral` (`CCN`). This list can contain clusters with any number of bumps. All clusters in the `CCN` list that contain only a single bump are placed in a list called `CalorNeutral` (`CN`). This list is further refined to a list called `GoodPhotonLoose` (`GPL`) by selecting those single bump clusters with energy greater than 30 MeV and with LAT less than 0.8. The photons used for this analysis are selected from the `GPL` list.

3.1.2 Track Reconstruction

A charged particle moving in the presence of a magnetic field will follow the trajectory of a helix, which can be parameterized by five quantities: (1) p_T, the transverse momentum, (2) *dip* angle, λ, is the angle between the transverse and longitudinal momentum, (3) d_0, the distance of closest approach (DOCA) to the origin in the $x-y$ plane, (4) ϕ_0 is the azimuthal angle at d_0, (5) z_0 is the distance of closest approach to the origin in the z direction. These parameters are used to parameterize the trajectory of the particle by

$$\begin{aligned} x &= r\sin\phi - (r+d_0)\sin\phi_0 \\ y &= -r\cos\phi - (r+d_0)\cos\phi_0 \\ z &= z_0 + l\ \tan\lambda \\ p_x &= p_T\cos\phi \\ p_y &= p_T\sin\phi \\ p_z &= p_T\tan\lambda \end{aligned} \tag{3.4}$$

where r is the radius of the track, given by $r = p_T/qB_z$, and $\phi = \phi_0 + l/r$ with l the parameter along the track.

Charged tracks are reconstructed using the output of the Level 3 DCH trigger. A helical fit is done on the track segments using a Kalman filter method [20] that takes into account measurement uncertainties as well as uncertainties due to multiple scattering and energy loss. The tracks are then extrapolated into the SVT, where they are added to the tracks from the standalone SVT tracking. Finally the full track is refit using the information from both the SVT and DCH. Tracks in the SVT not matched to any DCH track are also refit and are kept as SVT-only tracks. All reconstructed tracks in

the event are put into a list called `ChargedTracks` (CT) where by default all particles are assigned a mass equal to a pion. Other lists are constructed by requiring the tracks to pass some predefined selection criteria. This analysis uses a subset of this list called `GoodTracksLoose` (GTL). The tracks in this list are required to have $|d_0| < 1.5$ cm, $|z_0| < 2.5$ cm, $p_T < 0.05$ GeV/c, and $|p| < 10$ GeV/c.

3.2 B Counting

To measure the branching fraction for some process

$$X \to Y + Z + \ldots \tag{3.5}$$

one needs to know the number of parent particles, X, produced. For a measurement of $\mathcal{B}(B \to \gamma\gamma)$ this means the number of B^0 and $\overline{B}^0$ mesons in the data sample needs to be known. In *BABAR* only the total number of $B\overline{B}$ pairs is counted so to get the number of neutral B mesons a measurement of the $\Upsilon(4S) \to B^0\overline{B}^0$ branching fraction is also necessary.

B counting at *BABAR* is done by comparing data taken at the $\Upsilon(4S)$ resonance energy, called on-resonance, with that taken 40 MeV below, called off-resonance. In on-resonance data, the number of $B\overline{B}$ events is equal to the total number of hadronic events (defined below) minus the number of non-$B\overline{B}$ hadronic events. The number of non-$B\overline{B}$ hadronic events is found by luminosity scaling the off-resonance data up to the on-resonance energy.

The number of $\Upsilon(4S)$ produced, N_Υ, is given by the equation

$$N_\Upsilon = N_{on} - M_{on} \cdot R_{off} \cdot \kappa \ , \tag{3.6}$$

where N_{on} is the number of on-resonance hadronic events, M_{on} is the number of on-resonance μ-pairs, R_{off} is the ratio of hadronic events to μ-pairs in

off-resonance data, and κ is a constant to account for the differences in cross-section and efficiencies at the different CM energies.

The hadronic events above are defined as follows. The event is required to have at least three tracks in the tracking volume. The event shape variable called R_2, which is the ratio of the second to zeroth Fox-Wolfram moment, is required to be less than 0.5. R_2 is a measure of the sphericity of an event [21]. Also the primary event vertex, found using charged tracks, is required to be within 0.5 cm of the IP in the $x--y$-plane and six cm in the z-plane and the total energy measured in the detector must be greater than 4.5 GeV. About 95% of simulated $\Upsilon(4S) \to B\overline{B}$ events pass the above selections.

Muon pairs are selected using the following criteria. The two highest momentum tracks in an event must have $|\vec{p}|$ greater than 4 and 2 GeV/c, respectively. The total amount of energy deposited in the EMC by the tracks must add to less than 2 GeV, and the invariant mass of the two tracks must be greater than 7.5 GeV/c^2. Additionally, the two tracks must be back to back in the center-of-mass frame to within 10°.

3.3 Data Simulation

The simulation of *BABAR* data proceeds through two steps.

1. Physics processes such as particle decay and B mixing are generated using EvtGen [22].

2. Particles produced in (1) are propagated through a detailed model of the *BABAR* detector using a GEANT4 simulation [23].

The EvtGen package generates physics processes using decay amplitudes, instead of probabilities, allowing for interference in the decay. This requirement is a must in order to correctly simulate the mechanics of the B-mixing or a time dependent decay. The GEANT4 toolkit models the passage of particles through matter, simulating processes like multiple scattering, brehmsstrahlung, and photon conversion.

Simulated data are produced in amounts with respect to the overall recorded luminosity of the detector. For decays of the $\Upsilon(4S)$ to charged and neutral B pairs the equivalent luminosity of the simulated samples is about three times the data. Continuum data, $e^+e^- \to q\overline{q}$ ($q = u, d, s$) is produced with the same luminosity as the recorded data, while heavier $e^+e^- \to c\bar{c}$ is produced at a level of twice the data. $e^+e^- \to \tau^+\tau^-$ events are produced at about half the amount of data. To track the changes in the detector conditions throughout the data taking period from 1999–2008 from radiation damage and mechanical problems the simulated data samples are broken down by time period. This means that the running conditions of the detector are present in the simulated data for a certain set of the events.

3.4 Datasets

Table 3.1 shows the size, in integrated luminosity and number of $B\overline{B}$ pairs, of the on and off-resonance datasets used in this analysis. The number of $B\overline{B}$ pairs is calculated using the prescription described in Section 3.2 and the luminosity is calculated using Bhabha events. The amount of simulated data produced to study background distributions of continuum and exclusive decay processes is shown in Tables 3.2 and 3.3.

Table 3.1: Total integrated luminosity of each run period in on and off-resonance data. For on-resonance data the number of $B\overline{B}$ pairs is calculated using the technique described in Section 3.2. The off-resonance data is taken with a center-of-mass energy approximately 40 MeV below the on-resonance data.

Run	On-resonance		Off-resonance
	$\mathcal{L}$ [fb^{-1}]	$N_{B\bar{B}}$	$\mathcal{L}$ [fb^{-1}]
1	20.4	22,396,342	2.6
2	61.1	67,394,307	6.9
3	32.3	35,569,248	2.5
4	100.3	110,449,802	10.1
5	133.2	147,190,396	14.5
6	78.4	84,358,838	7.8
Total	425.7	467,358,933	44.4

Table 3.2: The number of continuum ($e^+e^- \to q\bar{q}$) and $B\overline{B}$ MC events produced at the on-resonance center of mass energy for each data taking period of *BABAR*. The equivalent luminosities of each sample are calculated using the given cross-sections.

Run	$B^0\overline{B}^0$	B^+ B^-	uds	$c\bar{c}$	$\tau^+\tau^-$
1	37,058,000	35,862,000	47,180,000	58,900,000	20,378,000
2	100,100,000	101,758,000	130,858,000	168,844,000	55,546,000
3	48,296,000	48,394,000	66,892,000	83,974,000	27,988,000
4	165,722,000	167,568,000	213,380,000	252,830,000	90,032,000
5	240,768,000	243,210,000	317,846,000	366,758,000	132,232,000
6	132,270,000	128,586,000	162,156,000	201,162,000	69,030,000
Total	724,214,000	725,378,000	938,312,000	1,132,468,000	395,206,000
σ [nb]	0.525	0.525	2.09	1.30	0.90
$\mathcal{L}$ [fb^{-1}]	1379.5	1381.7	448.9	871.1	439.1

Table 3.3: The number of exclusive MC events generated in the full *BABAR* simulation and the luminosity of a on-resonance dataset required to contain that number of events. A branching fraction of 1×10^{-7} is defined for signal MC in this table. All other branching fractions are taken from the PDG except where noted. A branching fraction with a less than sign is an upper limit at the 90% confidence level. The mode number is an internal *BABAR* bookkeeping system for the production of simulated data.

Mode Number	Decay	N_{evt}	$\mathcal{B}$ ($\times 10^{-7}$)	Equivalent $\mathcal{L}$ [fb^{-1}]
1764	$B^0 \to \gamma\gamma$	1,962,000	1.0	18,685,700
1043	$B^0 \to \pi^0\ \pi^0$	3,889,000	16.2 ± 3.1	2,286,300
2623	$B^0 \to \eta\eta$	434,000	5 ± 3 [24]	229,630
2624	$B^0 \to \eta\pi^0$	203,000	9 ± 4 [25]	128,888
1983	$B^0 \to \rho^0\gamma$	587,000	8.6 ± 1.5	650,055
995	$B^\pm \to \rho^\pm\gamma$	587,000	9.8 ± 2.5	601,127
1940	$B^\pm \to \rho^\pm\pi^0$	3,889,000	109.0 ± 14.0	339,799
1984	$B^0 \to \omega\gamma$	587,000	$4.4^{+1.8}_{-1.6}$	1,270,560
1587	$B^\pm \to K^\pm\ \pi^0$	3,889,000	129.0 ± 6.0	287,117
3134	$B^0 \to K^{*0}\ \gamma$	5,828,000	401.0 ± 20.0	138,416
3135	$B^\pm \to K^{*\pm}\ \gamma$	5,828,000	403.0 ± 26.0	137,729
1442	$B^0 \to K^0_S\pi^0$	3,931,000	33.9 ± 2.1	764,042
10174	$B^0 \to K^0_S\eta$	650,000	< 1.1	5,627,705

Chapter 4

Event Selection

This chapter describes the selection of events used to measure $\mathcal{B}(B \to \gamma\gamma)$. It begins with a description of the reconstruction of B candidates using from the charged tracks and neutral cluster described in Chapter 3. Then descriptions of all event cuts are given. Additionally, a description of the backgrounds from exclusive MC modes is discussed. Finally, the effect of all selections on signal MC is shown.

4.1 B Candidate Reconstruction

The B mesons used in this analysis are reconstructed by kinematically combining photons, i.e. adding their 4-momentum. The charged tracks and neutral clusters from reconstruction are used to create intermediate short lived particles. These intermediate particles are then combined again, and again to create a decay chain all the way back to the B meson parent particle.

An algorithm called `Add4` is used to reconstruct the decay of $B \to \gamma\gamma$. This algorithm works by simply adding the 4-momentum of two candidates together to create a potential parent candidate. In *BABAR*, all photon candidates from the `GPL` list have their momentum calculated as if they were produced at the origin of the *BABAR* coordinate system. In reality, the origin is not where the photons would be produced in a real B decay because the IP is offset from

the origin. To correct for this difference, the photon momentum are corrected by using the event vertex as the production point and recalculating. These 'corrected' photon candidates are then combined by the `Add4` algorithm.

For each pairing of the photons in an event, a parent candidate 4-momentum is created. Most of these pairings are combinations of photons that do not correspond to any true B meson in the event. What is needed is a way to separate the random unphysical pairings from events that correspond to signal decays. The primary means for doing this is to find variable distributions that are different for correctly reconstructed B decays from incorrectly reconstructed states. In *BABAR* there are two widely used variables that have this feature (1) $m_{\rm ES}$, the beam energy substituted mass, and (2) ΔE, the difference in energy of the reconstructed B meson and half the center of mass energy.

The beam energy substituted mass is defined as

$$m_{\rm ES} \equiv \sqrt{E^{*2}_{beam} - \vec{p^*}_B} \tag{4.1}$$

where E^*_{beam} is the center-of-mass beam energy, and $\vec{p^*}_B$ is the center-of-mass 3-momentum of the B candidate. For a correctly reconstructed B meson decay this variable will peak at the true B mass, $m_B = 5.279$ GeV/c^2.

The other quantity, ΔE, is defined as

$$\Delta E \equiv E^*_B - E^*_{beam} \tag{4.2}$$

where the asterisks denote the quantity is measured in the CM frame. In the center of mass frame for on-resonance data the beam energy is equal to one-half the $\Upsilon(4S)$ mass which should be the B mass. This quantity measures how close the CM energy of the reconstructed B candidate is to the expected

energy as determined by the measurement of the beam energy. For a correctly reconstructed B candidate this variable should peak at zero.

4.2 Event Preselection

To reduce the number of events that are analyzed a process called a *skim* is run over all reconstructed events. The skim looks at event related quantities to determine if it could be a candidate signal event. If the event carries a signature of a $B \to \gamma\gamma$ decay, two high energy neutral clusters, the photons are combined to form B meson candidates. If the B passes preliminary cuts on $m_{\rm ES}$ and ΔE the event is saved for further analysis.

Events that are consistent with $\Upsilon(4S) \to B\overline{B}$ are selected through the use of two filters called `BGFMultiHadron` and `BGFNeutralHadron`. The filter is a binary tag associated with the event, meaning an event either passes the filter or is rejected. The `BGFMultiHadron` tag selects events that contain several charged hadrons. Events must contain more than two charged tracks, mainly to reject Bhabha scattering. The filter requires events to have an $R_2/R_0 < 0.98$, where $R_l = \sum_{i,j} \frac{|p_i^*||p_j^*|}{s} P_l(\cos\theta_{ij}^*)$. P_l is the l^{th} order Legendre polynomial, p_i^* and p_j^* are the center-of-mass momentum of two particles, θ_{ij}^* is the angle between the momentum, and s is the center of mass energy. The sum runs over all *charged* particles in the event.

The `BGFNeutralHadron` filter is used select events with neutral hadrons and is orthogonal to `BGFMultiHadron`. This filter requires events to have no more than two charged tracks and looks at neutral clusters with energy greater than 100 MeV. Neutral clusters with energy greater than 500 MeV are labeled as photons. Events that have two tracks must also contain at least two photons or three neutral clusters. Events with one track must contain at least two

photons or four clusters, and no track events must have at least three photons or six clusters. Finally the ratio R_2/R_0 is required to be less than 0.95 where now the sum runs over both charged and neutral clusters. Events selected for analysis are required to pass one of these filters.

If an event passes one of these filters then a B meson is reconstructed using photons from two lists named `GoodPhotonLoose` and `gammaConversionDefault` described in Section 3.1. Photons on the combined list with center-of-mass energy in the range $1.15 \leq E^* \leq 3.5$ GeV, are pairwise combined by adding their 4-momentum to create a B meson candidate. If there are more than two photons in this energy range then more than one B candidate can be formed. All B candidates are saved for further analysis that pass

- $5.1 < m_{\rm ES} < 5.3$ GeV/c^2
- $|\Delta E| < 0.5$ GeV.

The effect of the skim selection on signal and background MC and on/off-resonance data samples is shown in Table 4.1.

At this stage an event can contain more than one B candidate although this is rare. In signal MC, 0.06% of events have more than one B candidate and in *uds* MC the percentage is 1.03%. Figure 4.1 shows the distribution of B candidates per event for signal, *uds*, $c\bar{c}$, and $\tau^+\tau^-$ MC samples after the skim selections. To avoid a bias when selecting a B from multiple candidate events only events where exactly one B candidate is reconstructed are used in this analysis. In signal MC this selection is greater than 99.9% efficient.

Table 4.1: Efficiency of the skim selection on signal MC, background MC, and on-resonance and off-resonance data.

Dataset	**Pre-skim**	**Post-skim**	$\epsilon[\%]$
$B \to \gamma\gamma$	1,962,000	1,371,411	69.9
$B^0\overline{B}^0$ MC	724,332,000	1965	0.00027
B^+B^- MC	725,378,000	2537	0.00035
$c\overline{c}$ MC	1,132,468,000	335,466	0.030
uds MC	938,312,000	1,028,134	0.110
τ^+ τ^- MC	395,206,000	216,475	0.055
OffPeak Data	607,063,825	162,673	0.027
OnPeak Data	6,363,687,525	1,692,603	0.021

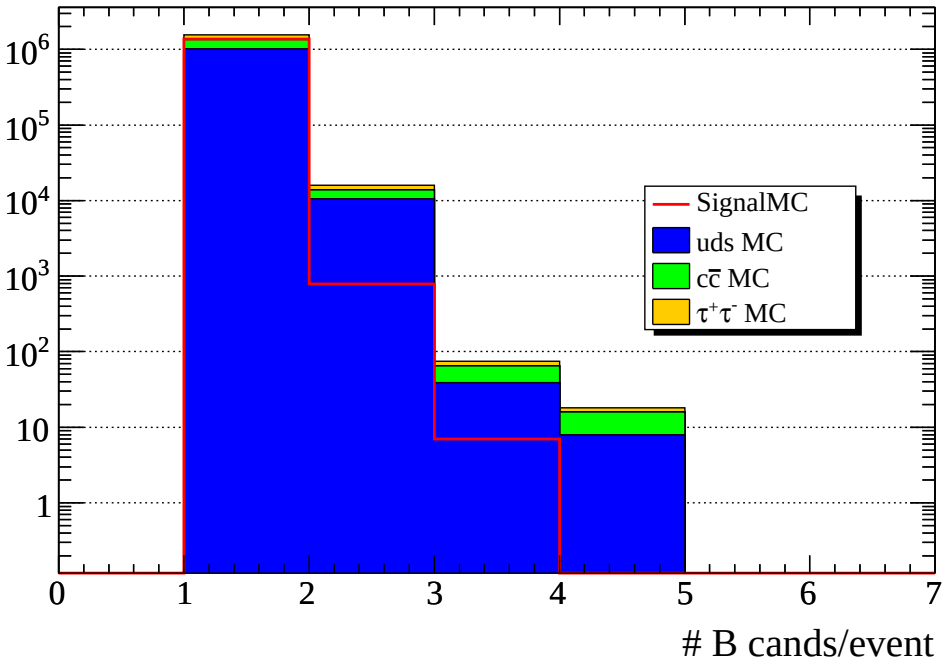


Figure 4.1: The number of B candidates per event that pass all of the skim cuts.

4.3 Event Categories

The photons used to reconstruct the B candidates are detected using two orthogonal methods and as such require different techniques to analyze. The first kind of photon is one that travels through the detector volume and undergoes an electromagnetic shower in the EMC volume which is reconstructed as a high energy cluster. These photons are present as GPL photons. The second type is a photon which undergoes a conversion to an e^+e^- pair through interactions with the material in the detector. Figure 4.2 shows the amount of material, in radiation lengths, in front of each detector subsystem as a function of polar angle. A photon has a higher probability of converting the more material that it traverses. The tracks from the daughter leptons are detected by the tracking system and are combined into the parent photon using geometric constraints on the track and an invariant mass constraint. These photons are present in the `GCD` list.

Four event categories are defined based on whether one of the photons used to reconstruct the B candidate undergoes a pair-conversion.

1. Both photon candidates are directly detected photons in the EMC.
2. The lower lab energy photon candidate undergoes pair-conversion.
3. The higher lab energy photon candidate undergoes pair-conversion
4. Both photon candidates undergo a conversion to an e^+e^- pair.

Table 4.2 shows the fraction of events in each category for signal MC. Only events from category 1 where the B candidate is reconstructed from `GPL` photons are used in the measurement. An attempt was made to incorporate events

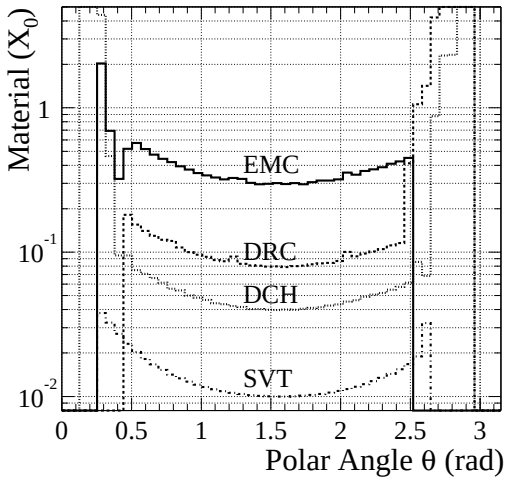


Figure 4.2: Amount of material, in radiation lengths, in front of each subsystem in the detector as a function of polar angle.

with photon conversions, but they were not used as the improvement to the upper limit was negligible. That work is presented in Appendix A.

4.4 Initial Selection

Loose cuts on photon quantities are made to clean up the sample of photons in the dataset. They are made by comparing distributions in signal

Table 4.2: The fraction of events, in signal MC, in each of the four categories.

Class	Fraction of Signal Events
1	0.922
2	0.030
3	0.046
4	0.002

and background MC with the aim of keeping virtually all signal events while removing obvious backgrounds. No event shape variables are used at this stage. Two types of selections are defined: (1) event level selections and (2) photon selections.

As a bookkeeping tool, each B candidates daughter photons are labeled according to their energy in the lab frame. The photon with the higher (lower) lab energy is called *gamHigh* (*gamLow*).

4.4.1 Event Level Selection

These cuts are applied at the event level before any cuts on the photon clusters are applied. They are designed to remove specific kinds of background events.

- **Number of** `GoodTracksLoose` ≥ 3
  This cut helps in removing $\tau^+\tau^-$ events.

- **Total Event Energy** < 15.0 GeV
  The total event energy should add up to the sum of the beam energy, but in practice it is lower due to the detector acceptance and neutrinos going undetected. This cut helps to remove events where the long scintillation time of the CsI(Tl) crystals can fake an energy deposit in the calorimeter from an out of time cluster. Section 4.5 describes out of time clusters.

4.4.2 `GoodPhotonLoose` Selection

These cuts remove events where the cluster of a high energy photon candidate is poorly reconstructed and to suppress backgrounds from merged π^0s and neutral hadrons.

- **# of Crystals the Cluster** > 10

  A photon interacting with the crystals in the EMC will undergo and electromagnetic shower that is usually contained within several adjacent crystals, mean of around 20, as is shown in Figure 4.4. This cut is designed to remove high energy clusters where all the energy is deposited within a small number of crystals which is inconsistent with the cluster shape produced from a photon.

- **Cluster Lateral Moment,** $0.1 \leq LAT \leq 0.6$

  A high energy photon in the EMC will have a definite distribution of the Lateral moment (3.2), that for *BABAR* should peak around 0.4. The cut on the cluster lateral moment is used to reject clusters where the shape is inconsistent with that of a high energy photon.

- **Cluster time,** $6100 \leq t \leq 6450$

  The long scintillation time of the CsI(Tl) EMC crystals and the high luminosity running of the PEP-II accelerator allows for the potential of events to overlap in time. This mainly occurs with Bhabha and two-photon events ($e^+e^- \rightarrow \gamma\gamma$) which can fake a signal. This cut rejects events where the photon cluster time is inconsistent with being from the current event. The cluster time is described in Section 4.5.

- **Cluster Isolation:** > 25 **cm**

  This selects events where the distance between the signal photon cluster and the nearest charged or neutral cluster in the event is larger than 25 cm. The purpose is to suppress backgrounds from merged π^0s and neutral hadrons.

- **EMC Fiducial Acceptance:** $0.4 < \theta < 2.4$ **rad**

  This cut selects only those events where the cluster is fully contained within the EMC volume. This removes the extreme forward and backward sections of the calorimeter. In the forward section the inner three rings of the endcap are poorly calibrated due to the material in front of the crystals. Additionally, Figure 4.6 shows that applying the polar angle cut greater than 0.4 radians removes pile-up events that have cluster energies comparable to a signal photon that could potentially produce a fake signal. The extreme backward edge of the calorimeter is also removed, to ensure that the entire cluster is contained within the crystals for an accurate measure of the cluster energy.

Figures 4.3 and 4.4 show the distributions of these variables for signal and background MC. The background MC is a combination of uds, $c\bar{c}$, $\tau^+\tau^-$, and $B\overline{B}$ MC that is weighted to a luminosity equivalent to the on resonance data of $425.7\,\mathrm{fb}^{-1}$.

4.5 Event Pileup

The high luminosity running of the PEP-II accelerator and the long scintillation time of the CsI(Tl) EMC crystals presents a problem for events that occur very close in time. If the shower from a near by in time event is still present when a consecutive event is recorded, a fake signal can occur. For example, if a Bhabha scattering event, $e^+e^- \to e^+e^-$, were nearby in time the tracks of the leptons would no longer be detected but the crystals may still be scintillating producing an energy deposit in the EMC. With no tracks matched to the crystals this would be reconstructed as a photon when there

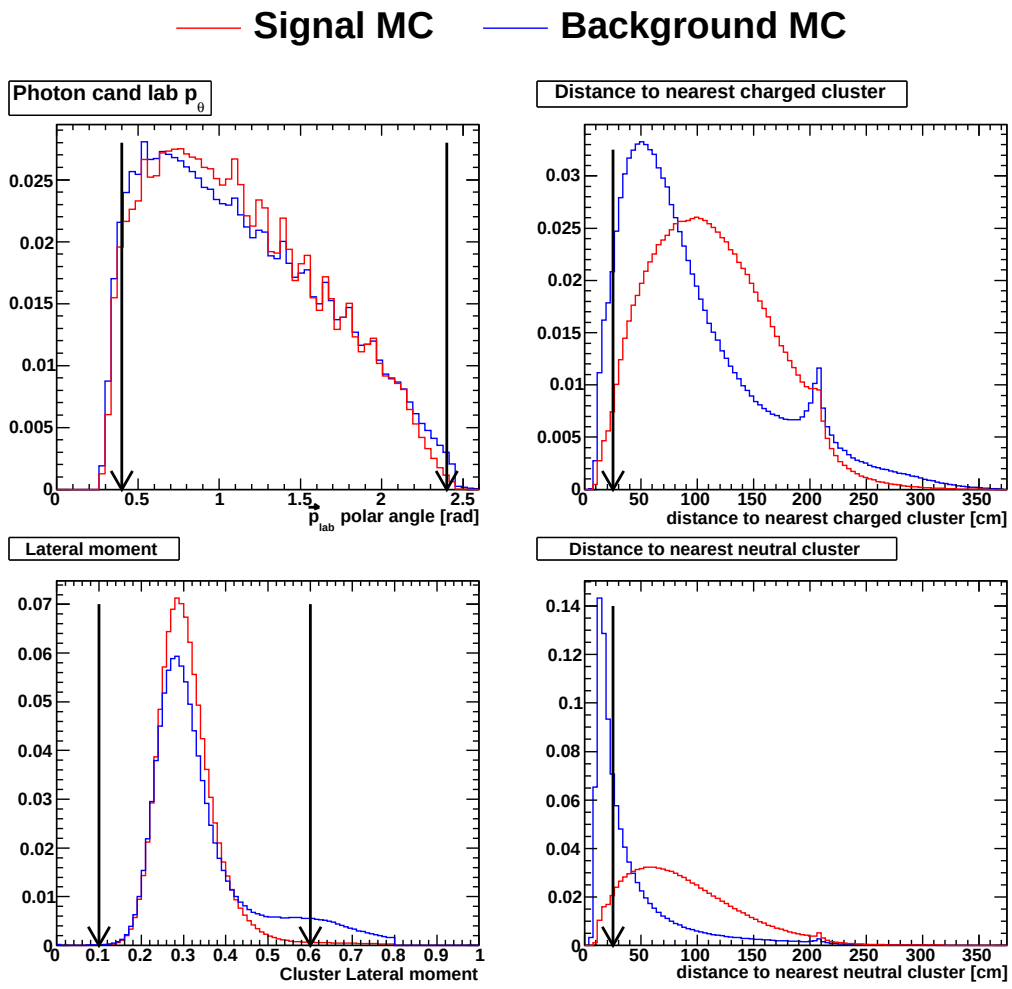


Figure 4.3: Distribution of variables used for initial selection in signal and background MC. The arrow indicates the value of the cut. If two arrows are present the accepted region is between them. The bump in the Distance to nearest charged cluster plot corresponds to the nearest charged cluster directly across the calorimeter barrel.

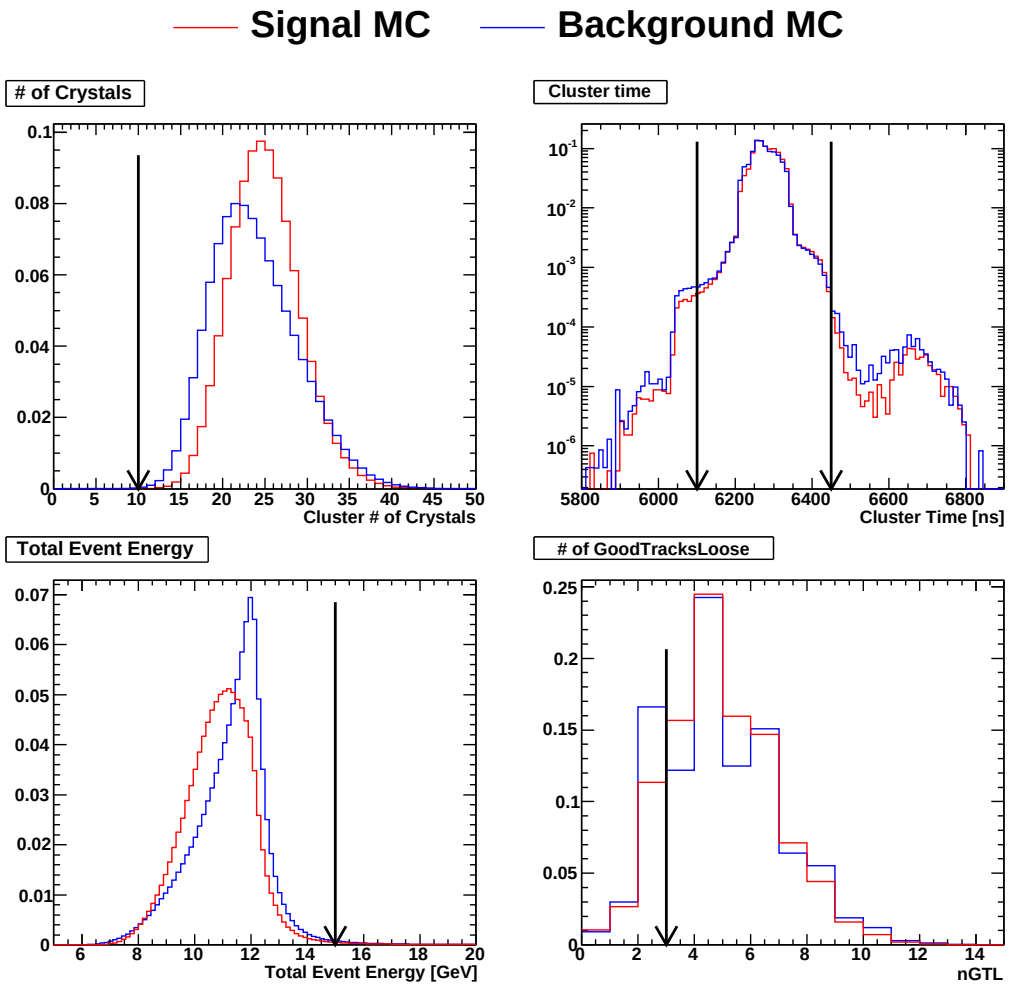


Figure 4.4: Distribution of variables used for initial selection in signal and background MC. The arrow indicates the value of the cut. If two arrows are present the accepted region is between them.

was not one initially. If the initial cluster has a high enough energy, the left over scintillation may produce a cluster that could fake a signal photon. At the $\Upsilon(4S)$ resonance the main processes that contribute to this type of behavior are Bhabha scattering, $e^+e^- \to e^+e^-$, and $e^+e^- \to \gamma\gamma$ due to their large cross sections. Additionally these are two body processes and their kinematics are very similar to the signal process $B \to \gamma\gamma$. The out-of-time clusters are rejected using cuts on the energy of the event and a 'time' quantity associated with the EMC cluster.

The total energy for an event is equal to the sum of the beam energies, although it is usually smaller due to the detector acceptance and the presence of neutrinos. Since out-of-time clusters are contributing extra energy, the total measured energy of the subsequent event where a fake signal could be measured will be larger. Therefore selecting events below an energy threshold can aid in rejection.

The EMC reconstruction software calculates a 'time' quantity associated with each reconstructed cluster. This is a derived quantity that is not directly a measure of the overall event time but is useful. Instead, this quantity is calculated in the feature extraction of the front end electronics on the detector. It is an energy weighted average of samplings of the crystal signal waveforms with respect to the time of the L1 trigger. For out-of-time clusters, the time value will be different than for in-time clusters due to their peaks occurring at different time.

Figure 4.5 shows a scatter plot of the total energy in an event versus the time of EMC clusters. Events that are outside of the red box are considered out-of-time. This shows that a cut on both the energy of less than 15 GeV and time of $6100 < t < 6450$ ns are effective at rejecting pileup events.

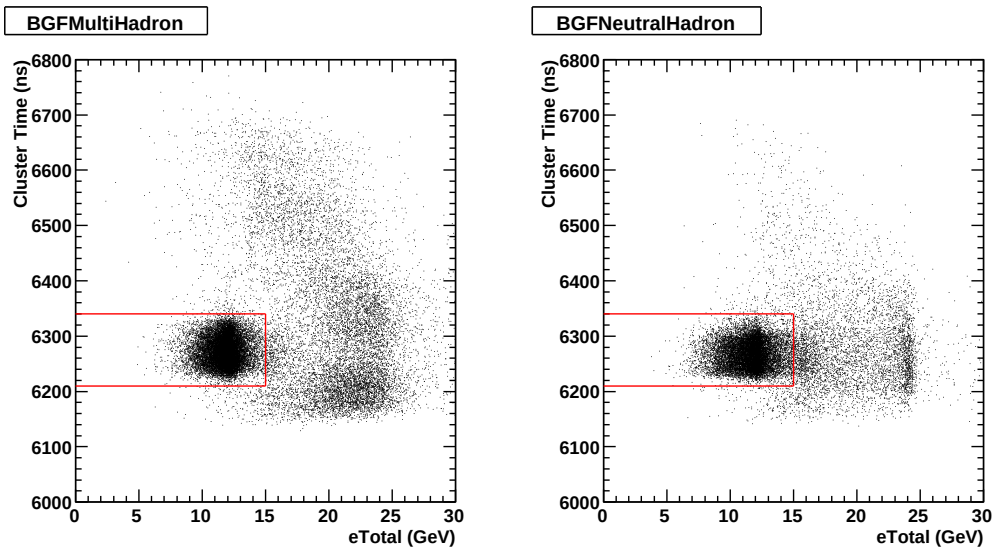


Figure 4.5: Time of the cluster versus the total energy in the event (sum of charged and neutral particle energies). Events outside the red box are considered out-of-time events.

Figure 4.6 shows the energy of out-of-time clusters as a function of the cluster polar angle, θ. This plot indicates the most of the out-of-time cluster energy is large, above 5 GeV, making it less likely to fake a signal event. It also shows the presence of out-of-time clusters with energies that could more easily fake signal clusters at polar angles below 0.4 rad. This is the motivation for selecting only those events where the photon clusters polar angle is greater than 0.4 rad.

4.6 π^0 and η Rejection

The major source of background photons for this analysis come from the decays of high energy π^0s and ηs as shown in Figure 4.7. Background from

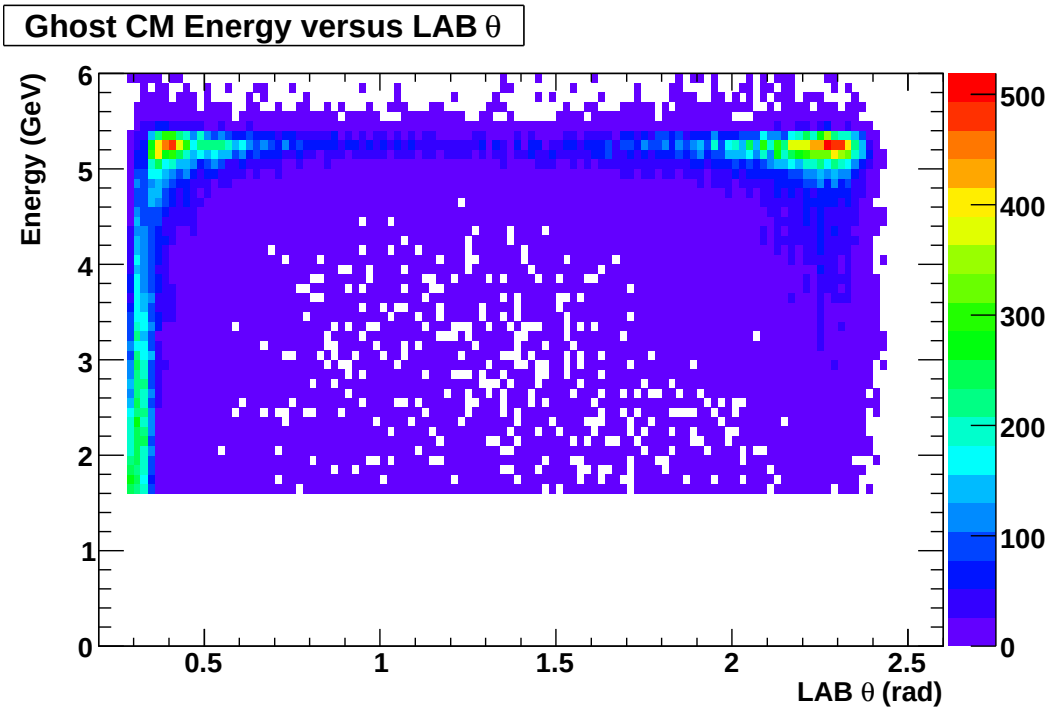


Figure 4.6: Lab energy of the out-of-time clusters from Bhabha and $\gamma\gamma$ events versus the lab polar angle θ of the cluster. The cutoff in energy below 1.6 GeV is a consequence of event selection for the pileup cluster study.

these decays are normally rejected by pairing the signal photon candidate with the other photons in the event and calculating the invariant mass. If the mass is within a window of the π^0 or η mass, the event is removed. The efficiency of rejecting π^0s and ηs can be improved by using a likelihood ratio rejection technique that uses energy of the non-signal photon in addition to the the invariant mass. This technique was developed for a branching fraction measurement of the mode $B \to \rho(\omega)\gamma$ [26, 27] at *BABAR*.

The likelihood ratio is defined to be a ratio of probabilities for the signal photon candidate to be produced from the decay of a π^0 or η, and from the

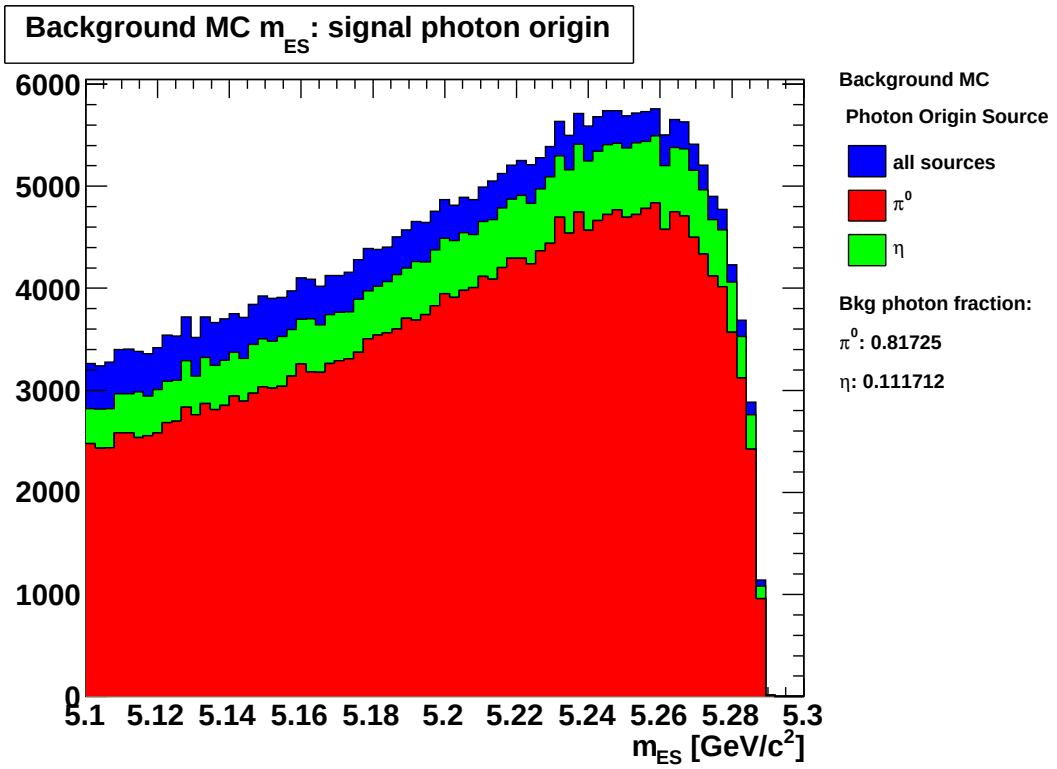


Figure 4.7: Origin of the signal photon candidate from background MC. More than 90% of the photons in the background originate from the high energy decays of π^0 and ηs.

signal decay, $B \to \gamma\gamma$. It is defined as

$$LR(\theta) \equiv \frac{\mathcal{P}_b(M(\gamma_s\gamma_2), E_{\gamma_2}|\theta)}{\mathcal{P}_s(M(\gamma_s\gamma_2), E_{\gamma_2}|B \to \gamma\gamma) + \mathcal{P}_b(M(\gamma_s\gamma_2), E_{\gamma_2}|\theta)} \tag{4.3}$$

where θ is a label for either π^0 or η, $M(\gamma_s\gamma_2)$ is the invariant mass of the photon pair, and E_{γ_2} is the lab energy of the non-signal photon. $\mathcal{P}_s$ and $\mathcal{P}_b$ are probabilities density functions constructed from two-dimensional histograms normalized to one. The signal probability, $\mathcal{P}_s$, gives the probability for the signal photon candidate to have been produced from the decay $B \to \gamma\gamma$. The

background probability, $\mathcal{P}_b$, gives the probability for the photon to have been produced from a π^0 or η decay depending on the value of the label θ. A likelihood ratio value close to zero indicates the photon is more likely to have been produced in a signal decay, while a value close to one means it was more likely to be produced from a π^0/η decay.

The probability density function for signal, $\mathcal{P}_s$, is constructed using signal MC. For each event, the signal photon candidate is paired with all non-signal photon candidates in the event and the reconstructed invariant mass and lab energy of the soft photon is recorded. The distributions of these quantities from signal MC are shown in red in Figures 4.8 and 4.9

The probability density function for π^0 (η), $\mathcal{P}_b$, is constructed from uds, $c\bar{c}$, and $\tau^+\tau^-$ MC. The signal photon candidate is required to come from a π^0 (η) decay and the second photon from the decay is also required to be within the calorimeter volume. If both of these conditions are met then the reconstructed invariant mass and lab energy of the non-signal photon candidate are recorded. The distribution of inputs to the π^0 (η) PDFs are shown in blue in Figures 4.8 and 4.9.

Each signal photon candidate is paired with all non-signal photon candidates in the event and both a π^0 and η likelihood ratio is calculated for each pairing. The largest value of each likelihood ratio is associated with the signal photon candidate corresponding to the most π^0 and η-like pairings in the event. The π^0 and η likelihood ratio distributions are shown in Figure 4.10. The π^0 (η) distribution in blue(green) is for photons from continuum MC that are required to be produced in a π^0 (η) decay. The signal distributions in red are for photons from signal MC that are required to come from the signal decay, $B \to \gamma\gamma$. Figure 4.11 shows the advantage of using the likelihood

ratio rejection technique over just an invariant mass cut. At all signal efficiencies, the likelihood ratio provides superior rejection of photons from π^0 and η decays.

4.7 Merged π^0s

The primary source of background photons come from the decay of high energy π^0s. When the opening angle of the daughter photons is small enough, the EMC is unable to resolve them and the resulting cluster can be mistaken for a high energy photon and fake a signal event. This type of event is referred to as a *merged* π^0 and can be rejected by cutting on a quantity called the merged π^0 consistency [28]. The consistency is calculated from the energy and second moment, eq. (3.3), of the cluster. The consistency is defined between zero, merged π^0-like, and one, photon-like. Figure 4.12 shows the distribution of the merged π^0 consistency for signal and background MC. The consistency is calculated during the reconstruction of data the described in Section 3.1 and is available to all analyses involving high energy photons. The distribution of the merged π^0 consistency is shown in Figure 4.12. The large zero bin is due to a technical problem that occurs when this quantity is persisted in the data. If the calculated consistency is less than 0.01, the value is not stored. These candidates are assigned a consistency value of 0.

4.8 Rejection of Continuum Background

The B mesons produced in $\Upsilon(4S)$ decays are almost at rest in the center-of-mass frame and decay isotropically. In contrast, particles produced in continuum events are collimated in jets along the axis of the initial $q\overline{q}$

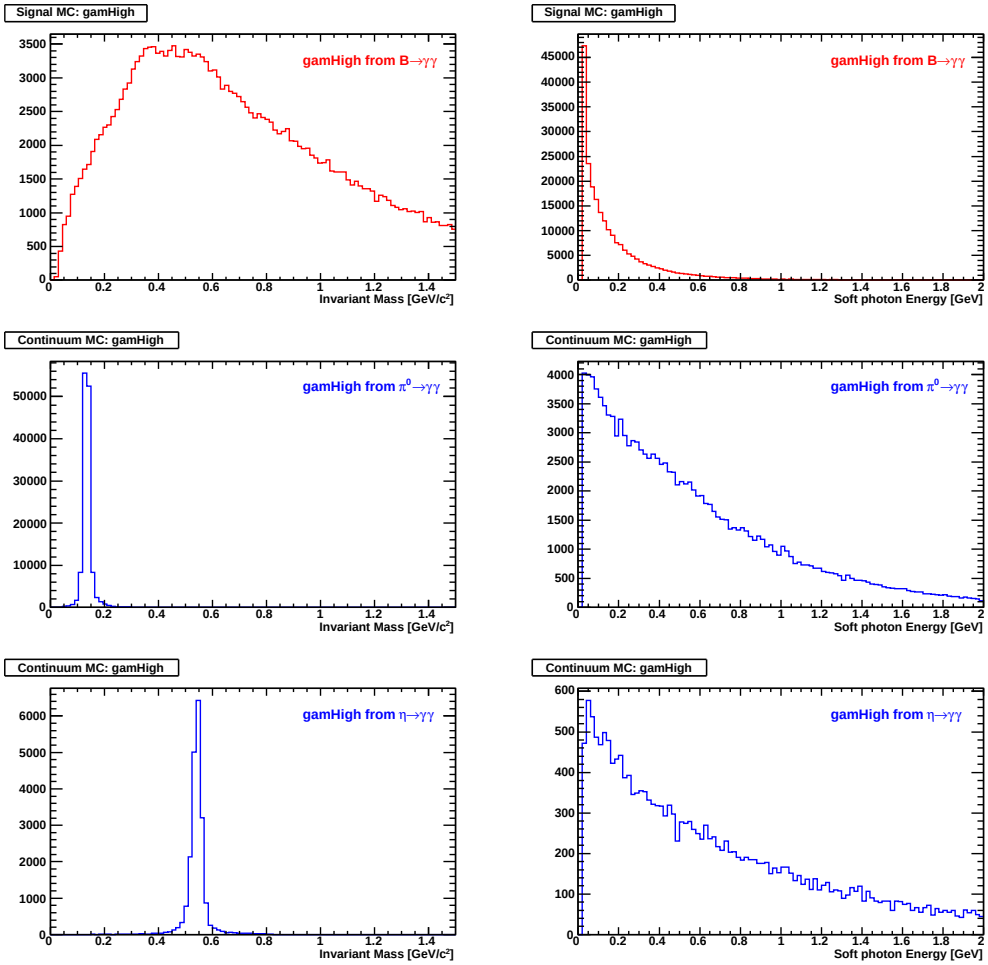


Figure 4.8: Distributions of inputs to the likelihood ratio PDFs for the larger lab energy signal photon candidate. (Top) signal, (middle) π^0, and (bottom) η. The left column shows the invariant mass distribution of the signal photon candidate with another photon in the event. The right column shows the distribution of the non-signal photon candidates reconstructed lab energy.

pairs. These different topologies make rejecting continuum events possible using quantities related to the event shape.

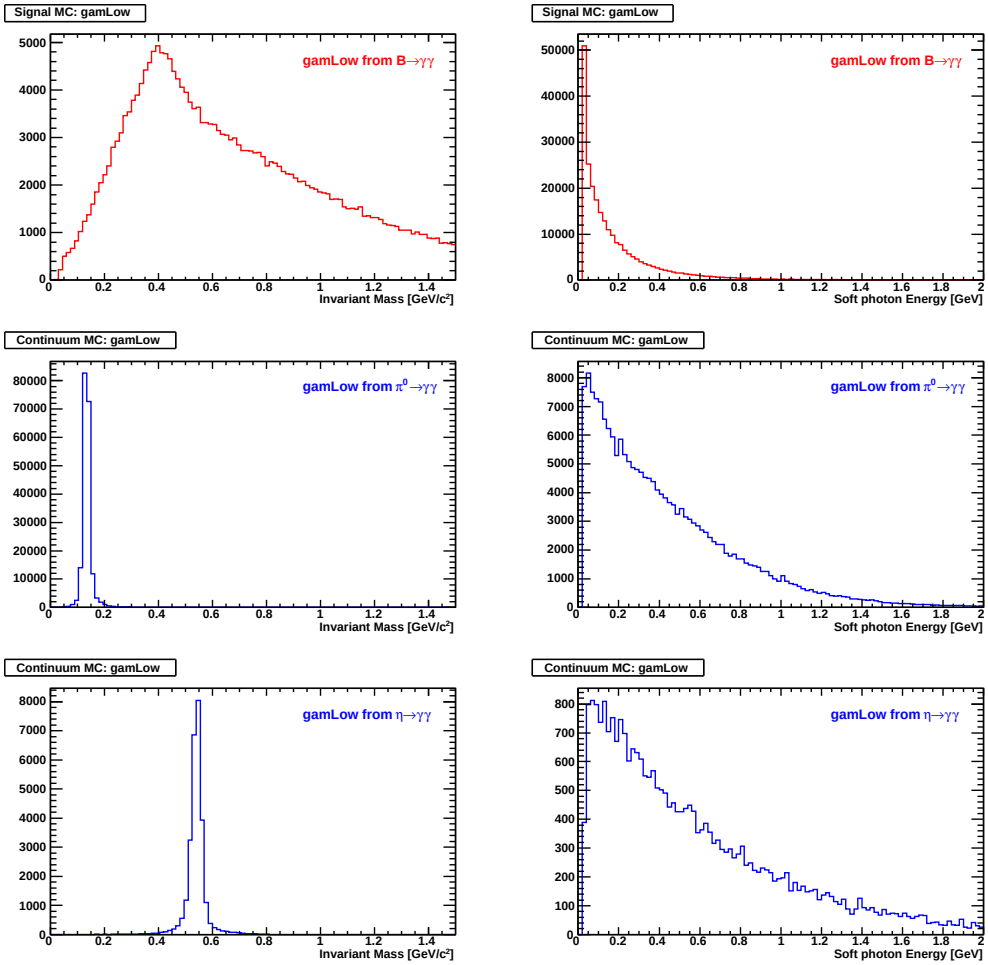


Figure 4.9: Distributions of inputs to the likelihood ratio PDFs for the smaller lab energy signal photon candidate. (Top) signal, (middle) π^0, and (bottom) η. The left column shows the invariant mass distribution of the signal photon candidate with another photon in the event. The right column shows the distribution of the non-signal photon candidate reconstructed lab energy.

A multivariate classifier called a Neural Network (NN)[1] is used to reject

[1]Boosted and Bagged Decision Trees were also investigated as to whether they would

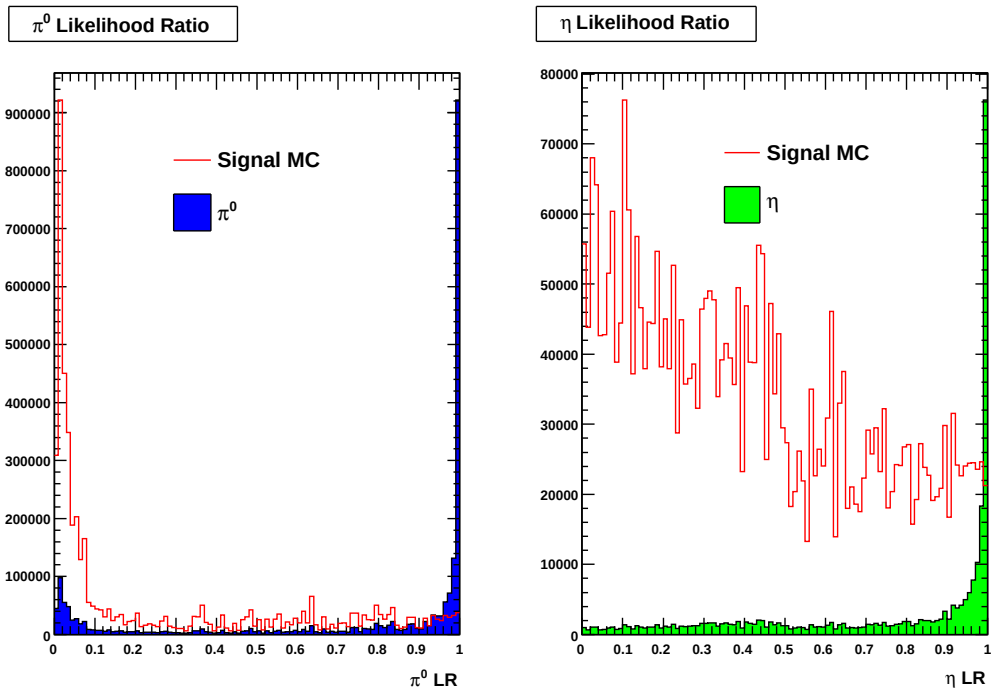


Figure 4.10: Likelihood ratio distributions for signal photon candidates from signal and background MC. (Left) Output π^0 likelihood ratio for signal and continuum + $B\overline{B}$ MC. The signal photon candidates in the blue distribution are required to originate from a π^0 decay. (Right) η likelihood ratio distribution for signal and continuum + $B\overline{B}$ MC. The photon in the green distribution is required to come from an η decay.

continuum ($e^+e^- \to q\overline{q}$) and $e^+e^- \to \tau^+\tau^-$ events. A multivariate classifier acts like a function in the sense that it takes a vector of quantities as input and outputs a single number that describes the input in terms of signal and background. For this analysis the input is a vector of event shape quantities and the output will be a number between zero and one that describes whether

provide better signal to background separation. The NN offered larger background rejection for all signal efficiencies.

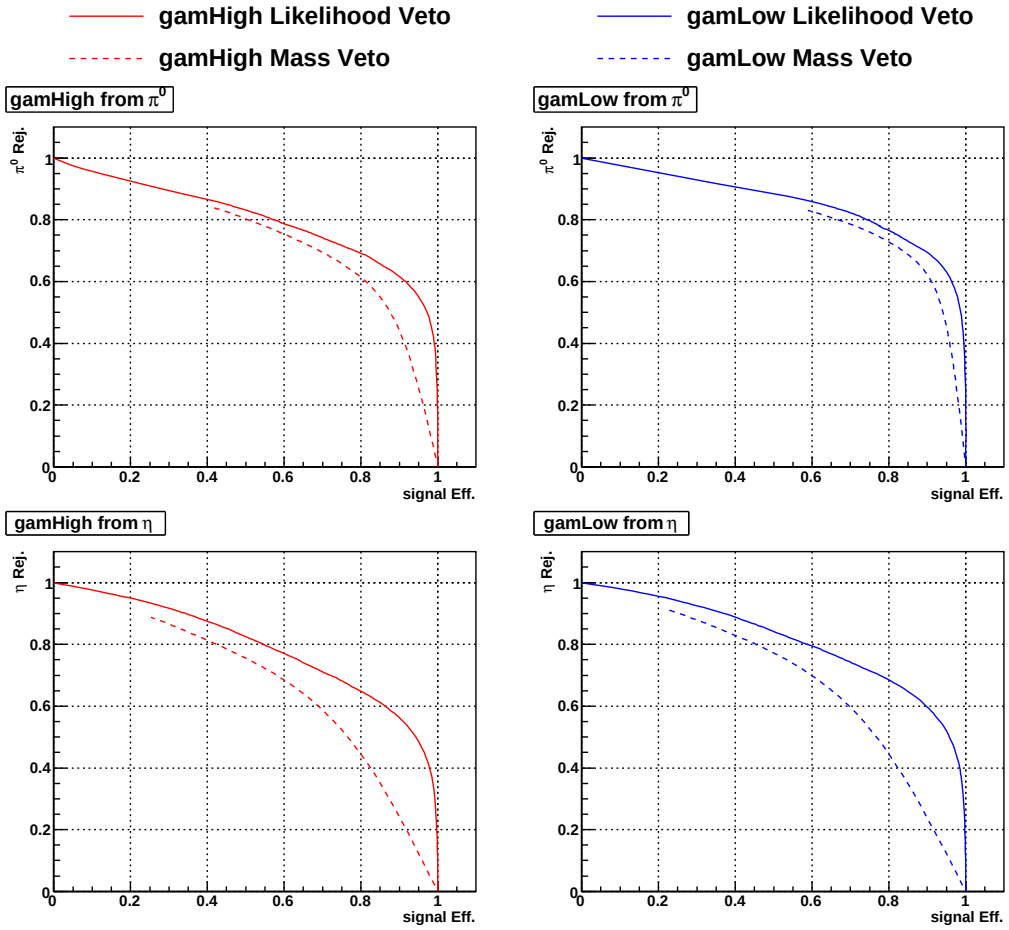


Figure 4.11: Signal efficiency versus π^0 and η rejection for the likelihood ratio and a cut on the invariant mass of the photon pairings. At all signal efficiencies that likelihood ratio provides superior rejection over the invariant mass cut.

the event resembles signal (near one) or background (near zero). The computer package StatPatterRecognition [29] was used to construct, train, and test the Neural Network. This package is a computer program developed for high energy physics experiments and written in the C++ language that provides advanced statistical tools to aid in the separation of signal and background.

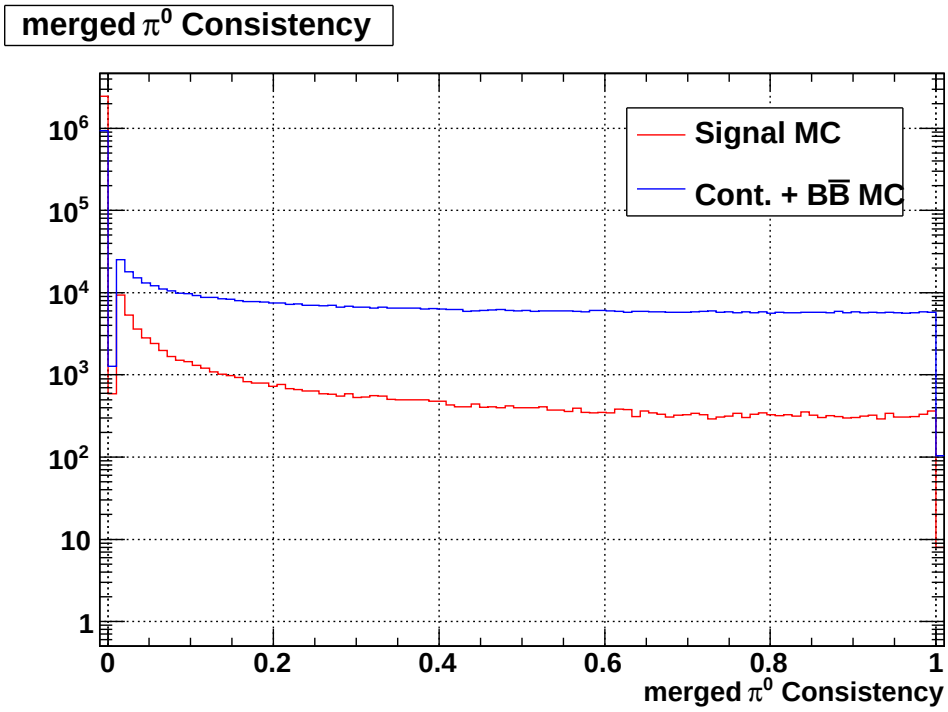


Figure 4.12: Distribution of the merged π^0 consistency in signal and background MC. The background MC sample is weighted to the on-resonance luminosity of $425.7\,\mathrm{fb}^{-1}$. The signal distribution is scaled to the same height as the background distribution. The large zero bin is due to the fact that consistency values less than 0.01 are not stored and are assigned a value of zero.

The signal events used in training and testing the the NN are selected from signal MC while the background is a combination of uds, $c\bar{c}$, and $\tau^+\tau^-$ MC events. They are required to pass the skim and initial selections described in Sections 4.2 and 4.4. Additionally a cut on the merged π^0 consistency for both photons of less than 0.1 is applied. The remaining events are randomly divided into two disjoint sets called a training and testing set. The random division is to ensure that for each set all parts of the MC datasets are sampled.

Table 4.3: Breakdown of the number of events used for the neural network training and testing samples for each MC mode.

MC sample	N_{events}
signal	22,600
uds	19,200
$c\bar{c}$	2,000
$\tau^+\tau^-$	1,400

The training set is used to tune the parameters of the NN for optimal signal and background separation while the testing dataset is used as an independent sample to verify the performance. The relative proportions of uds, $c\bar{c}$, and $\tau^+\tau^-$ events in the background training and testing set are calculated to be the same as expected in on-resonance data with the same selections applied. Table 4.3 shows the make up of the training and testing sets.

The event shape quantities used as inputs to the NN were chosen by comparing the distributions in signal and background MC. Those variables that showed an ability to distinguish between signal and continuum MC were used and are described below. Figures 4.13–4.17 show a comparison of their distributions in signal and continuum MC with off-resonance data overlaid. The off-resonance data contains only continuum events and matches the continuum MC.

- **Cluster isolation from nearest neutral cluster**

  This quantity measures the linear distance between the centroid of a signal photon candidate cluster and the nearest photon cluster. The other photon is from the `CalorNeutral` list. This distance is calculated for each of the signal photon candidates. The `CalorNeutral` list is de-

scribed in Section 3.1.1.

- **Cluster isolation from nearest charged cluster**

  This quantity measures the linear distance between the centroid of a signal photon candidate cluster and the nearest cluster that is associated with a charged track. This distance is calculated for each of the signal photons.

- **B momentum polar angle**

  The quantity is the polar angle of the momentum vector of the reconstructed B momentum in the laboratory frame.

- **R_2 in the CM frame**

  This quantity is the ratio of the 2^{nd} to the 0^{th} Fox-Wolfram moment of the event calculated in the center of mass frame. The Fox-Wolfram moments of order ℓ ($\ell = 0, 1, 2...$) are given by:

$$H_\ell = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{s} P_\ell(\cos\theta_{ij}) \tag{4.4}$$

  where the summation is over all final state particles in the event, $\vec{p}_i$ and $\vec{p}_j$ are the momenta of particles i and j, θ_{ij} is the angle between them, s is the center-of-mass energy, and P_ℓ are the Legendre polynomials of order ℓ.

- **Event Sphericity in the CM frame**

  The sphericity of an event is a measure of the total $p_\perp^2$ with respect to the event axis. For events with an isotropic distribution, i.e. signal events, S should be closer to 1, while for jet-like continuum events the value of S will be closer to zero. The sphericity tensor is given by the formula:

$$S^{\alpha\beta} = \frac{\sum_i^N p_i^\alpha p_i^\beta}{\sum_i^N |\vec{p}_i|^2}, \tag{4.5}$$

where the indices $\alpha, \beta = 1, 2, 3$ denote the x, y, and z components of the 3-momentum of the i^{th} particle in the event. This is a 3×3 tensor and can be diagonalized resulting in three eigenvalues: λ_1, λ_2, and λ_3 with the following relations:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \tag{4.6}$$

$$\lambda_1 + \lambda_2 + \lambda_3 \;\; = \;\; 1 \tag{4.7}$$

The event sphericity is defined to be

$$S = \frac{3}{2}(\lambda_1 + \lambda_2). \tag{4.8}$$

- **Rest of Event Sphericity in CM frame**

  This is the event sphericity, eq. (4.5), in the CM frame where the sum runs over all charged and neutral particles in the rest of the event (ROE), where the ROE is defined to be all charged tracks and neutral clusters excluding the two signal candidates.

- **Thrust axis polar angle with gamHigh removed**

  The thrust, T, of a set of particles with momentum, $\vec{p}_i$, is given by

$$T = \frac{\sum_i |\hat{n} \cdot \vec{p}_i|}{\sum_j |\vec{p}_j|} \tag{4.9}$$

  where $\hat{n}$ is a unit-vector in the direction that maximizes the value of T, called the thrust axis. Events with many particles lying along a certain axis have a larger thrust than events with a more isotropic distribution. $B\overline{B}$ events decay isotropically and will have smaller thrust values than jet-like continuum events.

This quantity is the polar angle of the thrust axis for all particles in the event except for the higher lab energy signal photon.

- **Thrust axis polar angle with gamLow removed**

  This quantity is analogous to the quantity above except that the lower lab energy photon rather than the higher lab energy photon is removed from the summation in eq. (4.9).

- **2^{nd} Angular Moment of Rest of Event Thrust axis**

  Useful topological variables that measure the angular distribution of energy and momentum flow in an event are called angular moments and are given by

$$L_\ell = \frac{\sum_i^N |\vec{p}_i| \cos^\ell \theta_i}{\sum_i^N |\vec{p}_i|}, \tag{4.10}$$

  where the summation is over all particles, $\vec{p}_i$ is the momentum of the i^{th} particle, and θ_i is the angle of the particle momentum with respect to the thrust axis. The NN used in this analysis uses the second angular moment with $\ell = 2$.

- **Rest of Event 1^{st} Angular Moment with gamHigh removed**

  This is the angular moment, (4.10), with $\ell = 1$ where the summation is over all particles in the event except for the two signal photons. The thrust axis is defined by all particles in the event except for the higher lab energy signal photon.

- **Rest of Event 2^{nd} Angular Moment with gamHigh removed**

  This is the same as above except with $\ell = 2$ in (4.10).

- **Rest of Event 3^{rd} Angular Moment with gamHigh removed**
  This is the same as above except with $\ell = 3$ in (4.10).

- **R_2' of event in recoil frame of gamLow**
  This quantity is the ratio of the second to zeroth order Fox-Wolfram moment, eq. (4.4), with the momentum and angles measured in the center-of-mass frame. The summation runs over all particles in the event except for the lower lab energy signal photon.

- **Event $\vec{\mathbf{p}}_T$**
  This is the total momentum of the event that is perpendicular to the z-axis. This quantity is calculated using all charged and neutral particles in the event.

- **Total Missing Energy of event**
  This quantity is a measure of the total energy in an event minus the beam energy.

- **# of Neutrals**
  This is the number of objects in the `CalorNeutral` list, which is described in Section 3.1.1.

- **# of `GoodTracksLoose`**
  This is the number of charged tracks in the event on the `GoodTracksLoose` list, which is described in Section 3.1.2.

The training is done by systematically varying the configuration of the NN's internal parameters such as the number of hidden layers, the number of nodes in a layer, and the number of training cycles and comparing the

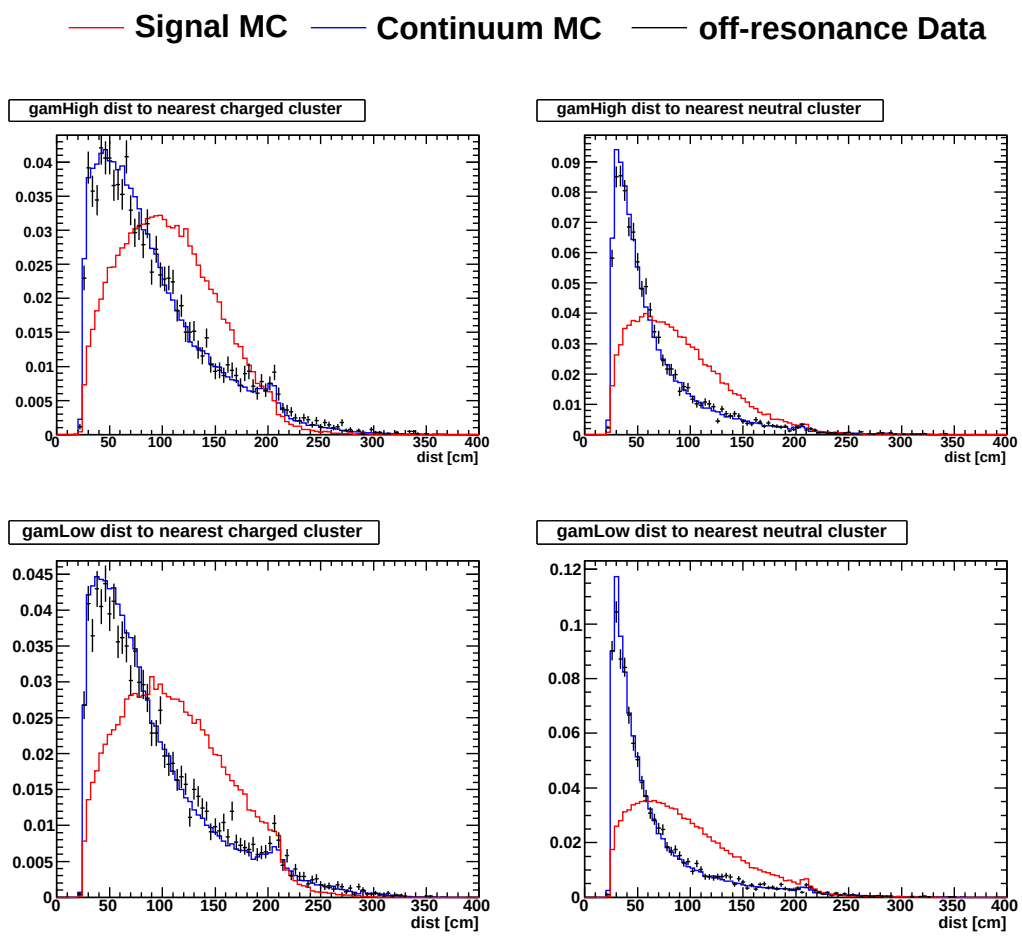


Figure 4.13: Input variables distributions to the NN classifier for signal MC, continuum MC, and offPeak data. The Continuum MC is a weighted sample of uds, $c\bar{c}$, $\tau^+\tau^-$ MC. All distributions are normalized to unity.

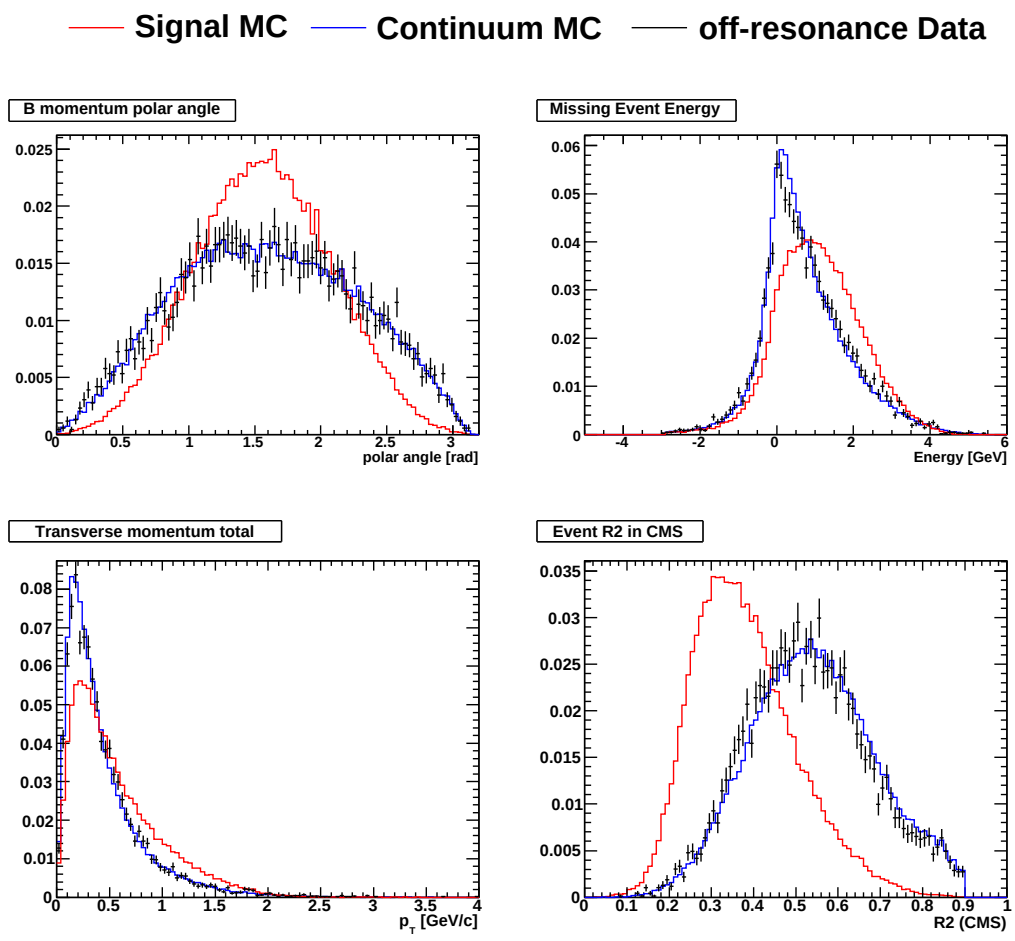


Figure 4.14: Input variables distributions to the NN classifier for signal MC, continuum MC, and offPeak data. The Continuum MC is a weighted sample of uds, $c\bar{c}$, $\tau^+\tau^-$ MC. All distributions are normalized to unity.

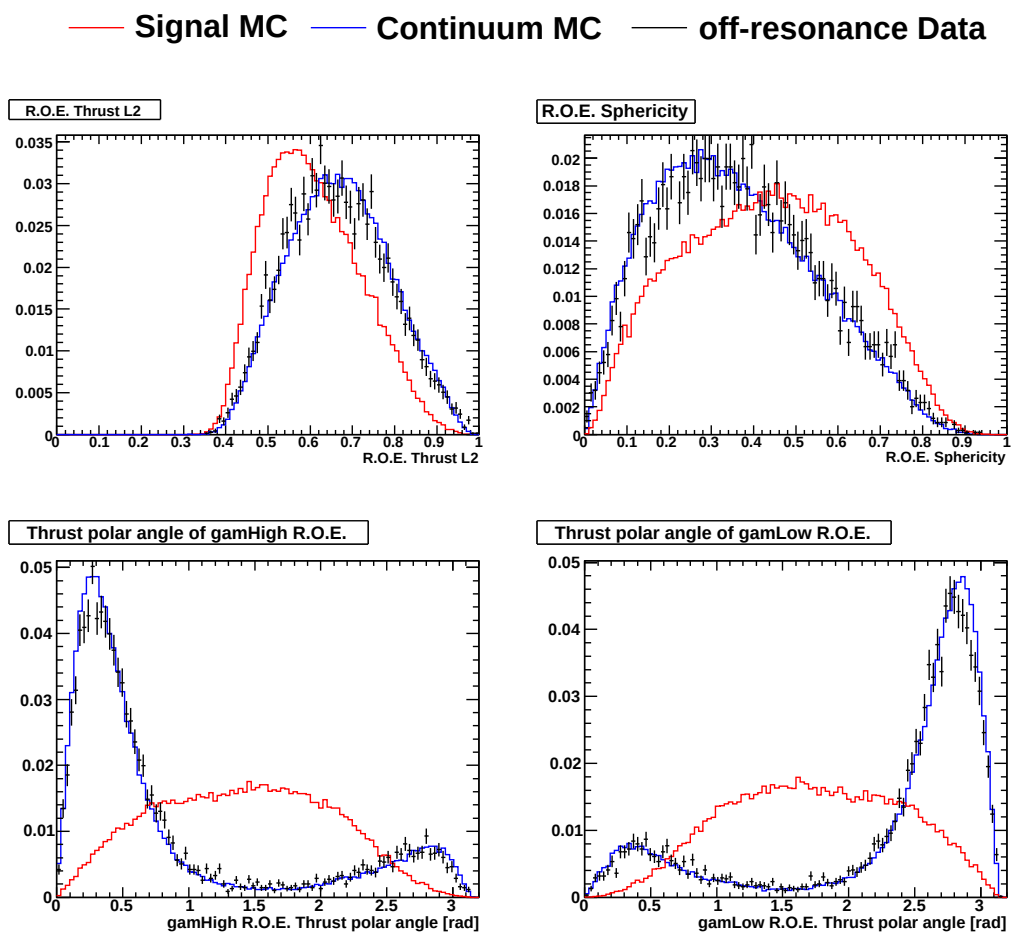


Figure 4.15: Input variables distributions to the NN classifier for signal MC, continuum MC, and offPeak data. The Continuum MC is a weighted sample of uds, $c\bar{c}$, $\tau^+\tau^-$ MC. All distributions are normalized to unity.

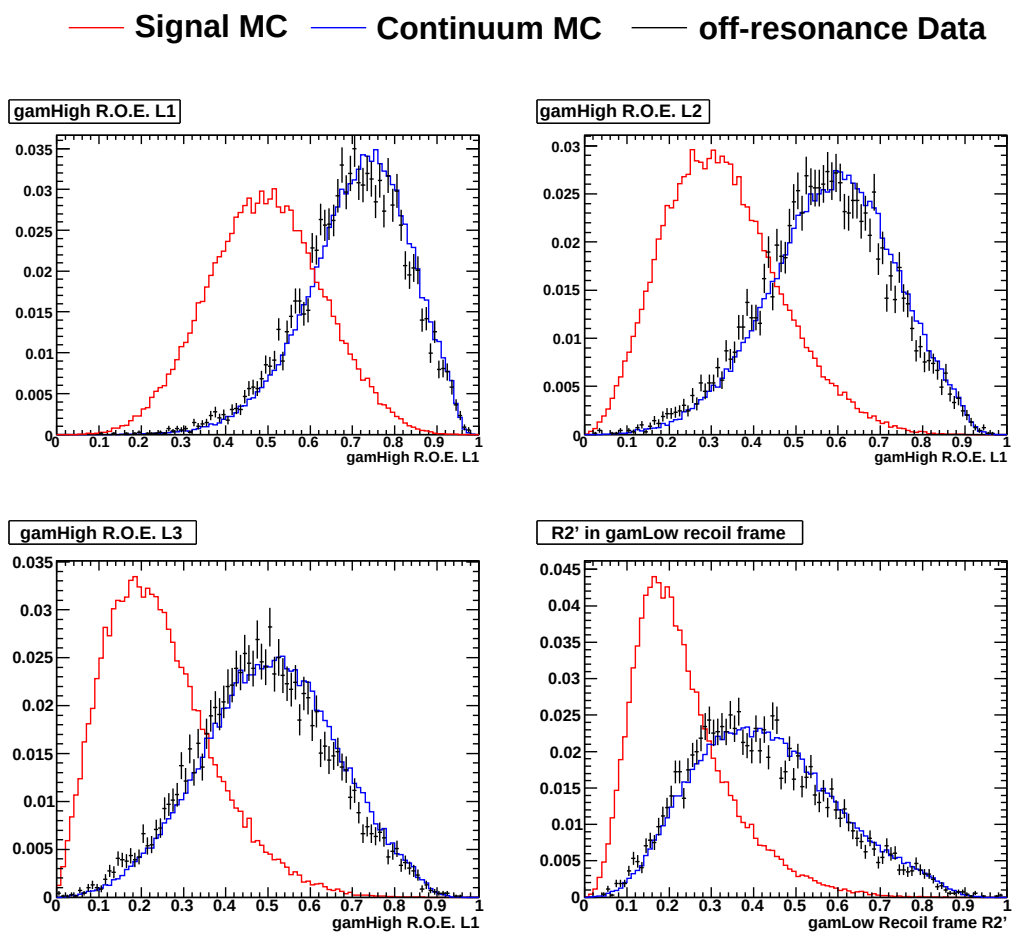


Figure 4.16: Input variables distributions to the NN classifier for signal MC, continuum MC, and offPeak data. The Continuum MC is a weighted sample of uds, $c\bar{c}$, $\tau^+\tau^-$ MC. All distributions are normalized to unity.

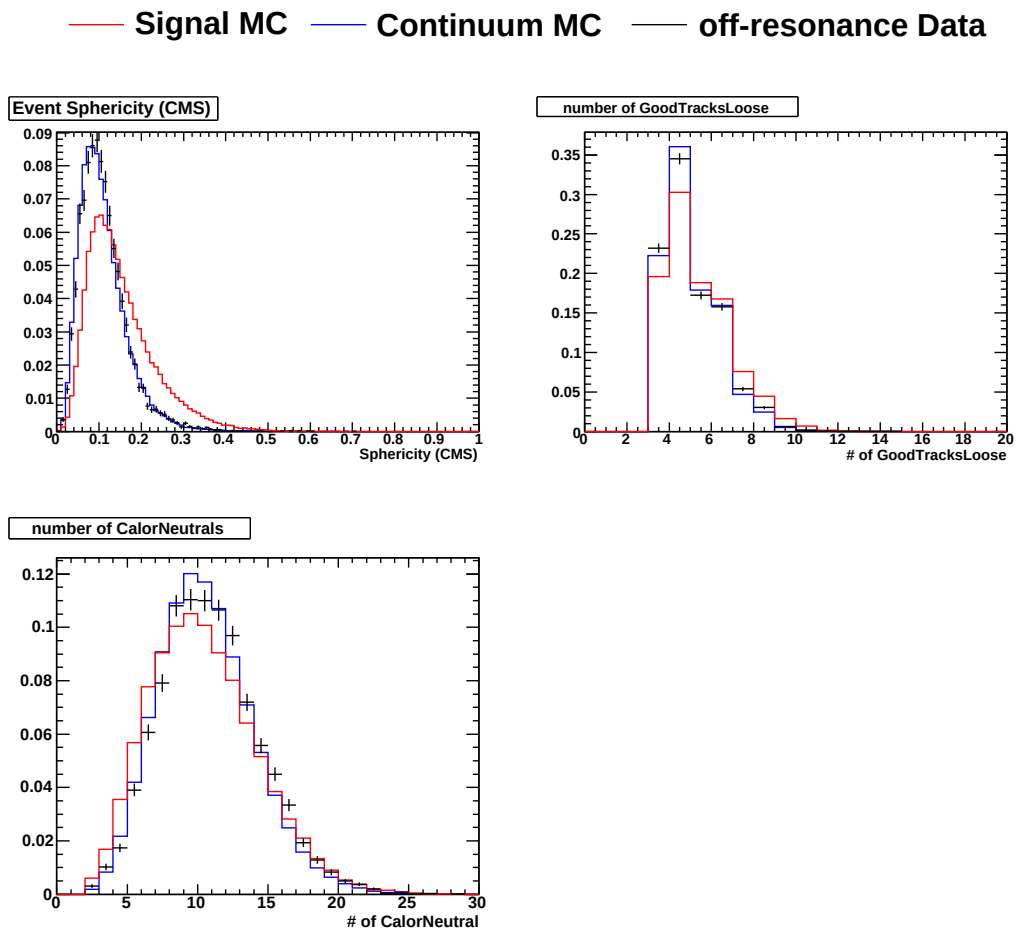


Figure 4.17: Input variables distributions to the NN classifier for signal MC, continuum MC, and offPeak data. The Continuum MC is a weighted sample of uds, $c\bar{c}$, $\tau^+\tau^-$ MC. All distributions are normalized to unity.

quadratic loss per event as a function of the number of training cycles. The quadratic loss per event is defined as

$$L = (y - f(x))^2 \tag{4.11}$$

where y is the the true class of the event, one for signal and zero for background, and $f(x)$ is the output of the NN. For a properly trained NN, the loss per event as a function of the number of training cycles should decrease and asymptotically converge to a stable value. The NN configuration that has the smallest loss per event is chosen as optimal. Figure 4.18 shows the training loss per event for the NN used in this analysis. This NN configuration has 400 cycles, and a $19 : 35 : 1$ configuration meaning one input layer with 19 nodes, one hidden layer with 35 nodes, and one output layer with one node. Figure 4.19 shows the output distribution of the optimal NN on the testing sample and Figure 4.20 shows the signal efficiency vs. the background rejection performance for the NN.

4.9 Optimization of Final Cuts

The values of the cuts on the π^0 and η likelihood veto, NN output, and the number of `GTL` are optimized using a figure of merit (FOM) called the *Punzi* FOM [30] defined as

$$Punzi \equiv \frac{\varepsilon_{sig}}{a/2 + \sqrt{B}}, \tag{4.12}$$

where ε_{sig} is the efficiency of the cuts on signal MC, B is the number of background events passing the cuts in the signal region. The value a is the number of sigmas corresponding to one-sided Gaussian tests at significance α. The optimization described here uses a value of $a = 3$. This FOM is useful for

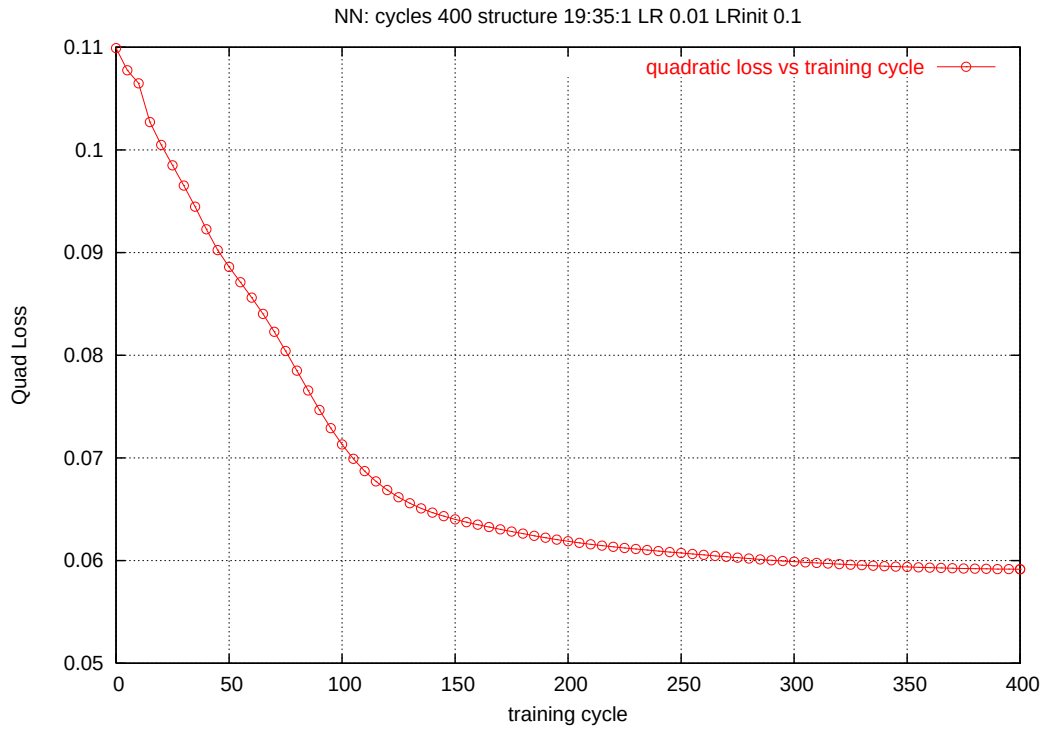


Figure 4.18: The quadratic loss per event as a function of the number of training cycles for the optimal neural network.

any search, and for the mode $B \to \gamma\gamma$ in particular, since no assumption on the branching fraction is required.

The optimization uses signal MC for the calculation of the signal efficiency, and on-resonance sideband data to calculate B. The on-resonance data sideband regions are defined through fits to the $m_{\rm ES}$ and ΔE distributions in signal MC. The signal $m_{\rm ES}$ distribution is fit with a Crystal Ball shape [31–33] and a $\pm 3\sigma$ window around the peak of the distribution is taken as the signal region. This corresponds to $m_{\rm ES}$ greater than 5.27 GeV/c^2. The Crystal Ball

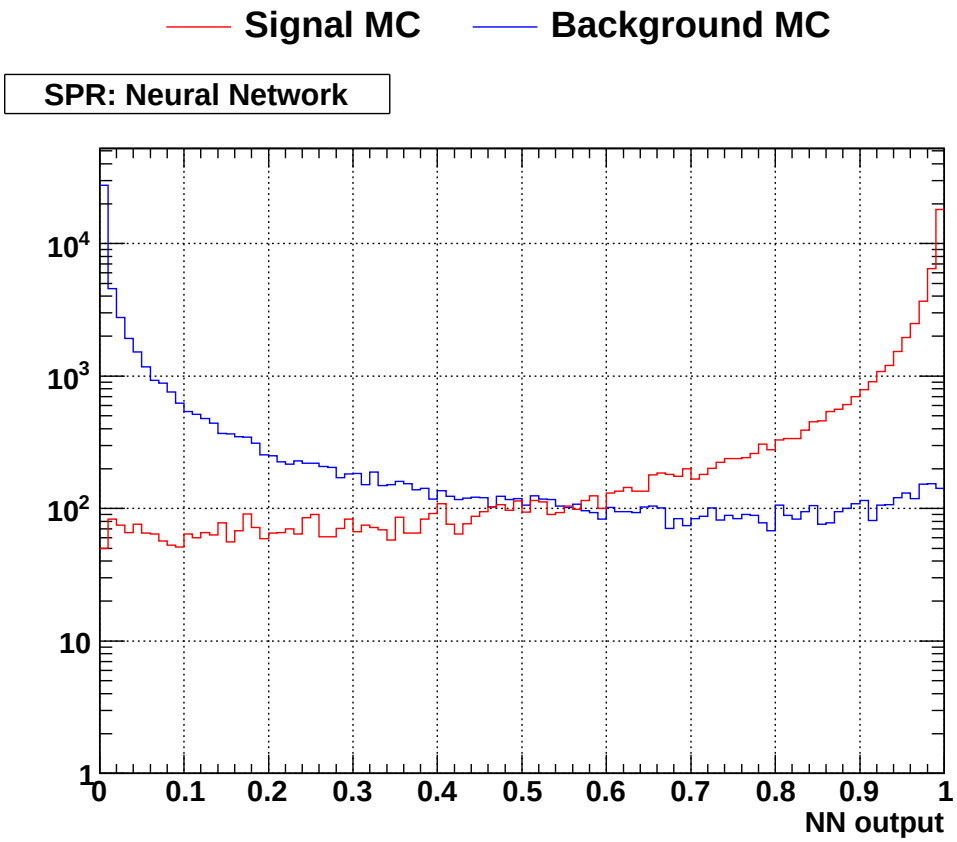


Figure 4.19: The Neural Network output distributions for signal and continuum MC. The distributions are scaled to unit area. Signal is red and Background is blue.

shape is defined as

$$f(x;\alpha,n,\bar{x},\sigma) = N \cdot \begin{cases} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-\bar{x}}{\sigma})^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leq -\alpha \end{cases} \tag{4.13}$$

$$\text{where} \quad A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) \text{ and } B = \frac{n}{|\alpha|} - |\alpha|.$$

Similarly, the signal ΔE distribution is fit with a doubled sided Gaus-

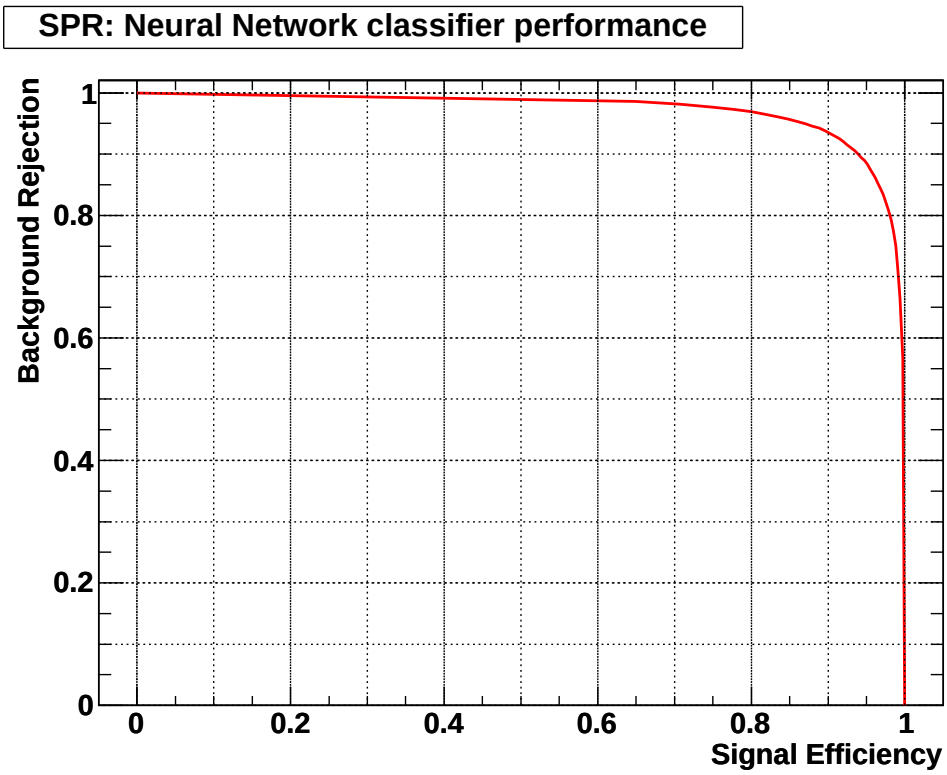


Figure 4.20: The signal efficiency versus background rejection of the neural network for the testing sample.

sian function with a low side tail parameter that is known in *BABAR* as the Cruijff shape and is given by

$$\text{Cruijff}(x;\mu,\sigma_{L,R},\alpha_{L,R}) = \exp\left[\frac{-(x-\mu)^2}{2\sigma_{L,R}^2+\alpha_{L,R}(x-\mu)^2}\right] \qquad (4.14)$$

where μ is the mean, σ is the gaussian width, and α is the tail parameter. The subscripts L and R denote the left and right side of the distribution peak, respectively. Again a $\pm 3\sigma$ window around the peak is defined as the signal

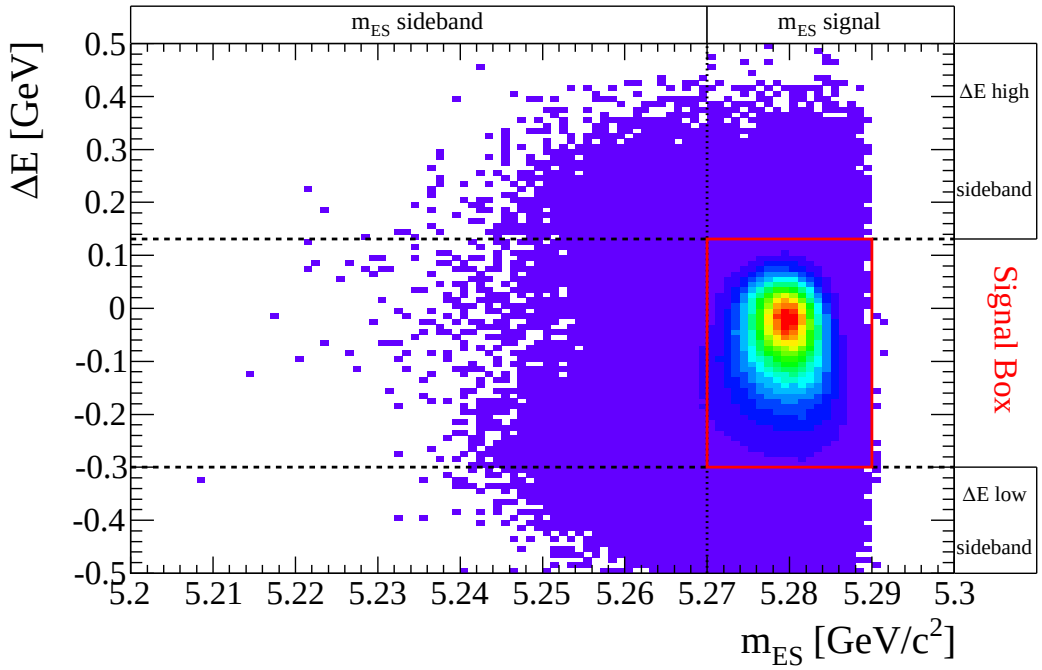


Figure 4.21: Definition of signal region and sidebands shown with signal MC. The signal region is defined as a $\pm 3\sigma$ window from fits the $m_{\rm ES}$ and ΔE distributions. The $m_{\rm ES}$ sideband is less than 5.27 GeV/c^2. The ΔE sidebands are the regions where $\Delta E < -0.3$ and $\Delta E > 0.1312$ GeV.

box which corresponds to $-0.3 \leq \Delta E \leq 0.1312$ GeV. Figure 4.21 shows the region topology for the $m_{\rm ES}$–ΔE space for signal MC.

The values of the cuts are optimized through an iterative process. One cut is varied over a range with all other cuts held constant. The value of the cut in the range that maximizes the FOM is chosen as optimum and assigned to the variable. The process is repeated to find the optimum value for the next cut again holding all other cuts constant. This process is repeated until all values are stable through consecutive iterations.

The events used to optimize the cut values are required to pass the skim and initial selections described in Sections 4.2 and 4.4 as well as a cut on the merged π^0 consistency less than 0.01. A fit region is also defined such that $m_{\rm ES}$ is

- $m_{\rm ES} \geq 5.2$ GeV/c^2
- $-0.5 \leq \Delta E \leq 0.5$ GeV

The signal efficiency, ε_{sig}, is calculated from signal MC as the ratio of events in the signal region to the total number generated during simulation production. To get the number of events in the signal region a 1 dimensional unbinned extended maximum likelihood fit is performed on the signal $m_{\rm ES}$ distribution using a Crystal Ball shape. The ratio of the area of the PDF in the signal region to the area in the fit region is multiplied by the fit yield to find the number of signal events that pass the selections. This number is used to calculate the efficiency. Figure 4.22 shows an example of a fit to the $m_{\rm ES}$ distribution in signal MC using a Crystal Ball shape for a cut on the π^0 likelihood veto less than 0.8.

This analysis is performed 'blind' and so the data inside the signal region is not accessible during event selection. The number of background events, B, is estimated using the data in the ΔE sideband in a two step process. First, the $m_{\rm ES}$ distribution in on-resonance ΔE sideband data is fit with an ARGUS shape [34] using a one dimensional unbinned extended maximum likelihood fit. The ARGUS shape is given by

$$f(x;m,p,c) = N \cdot x \cdot \left(1 - \left(\frac{x}{m}\right)^2\right)^p \cdot \exp\left(c\left(1 - \left(\frac{x}{m}\right)^2\right)\right). \qquad (4.15)$$

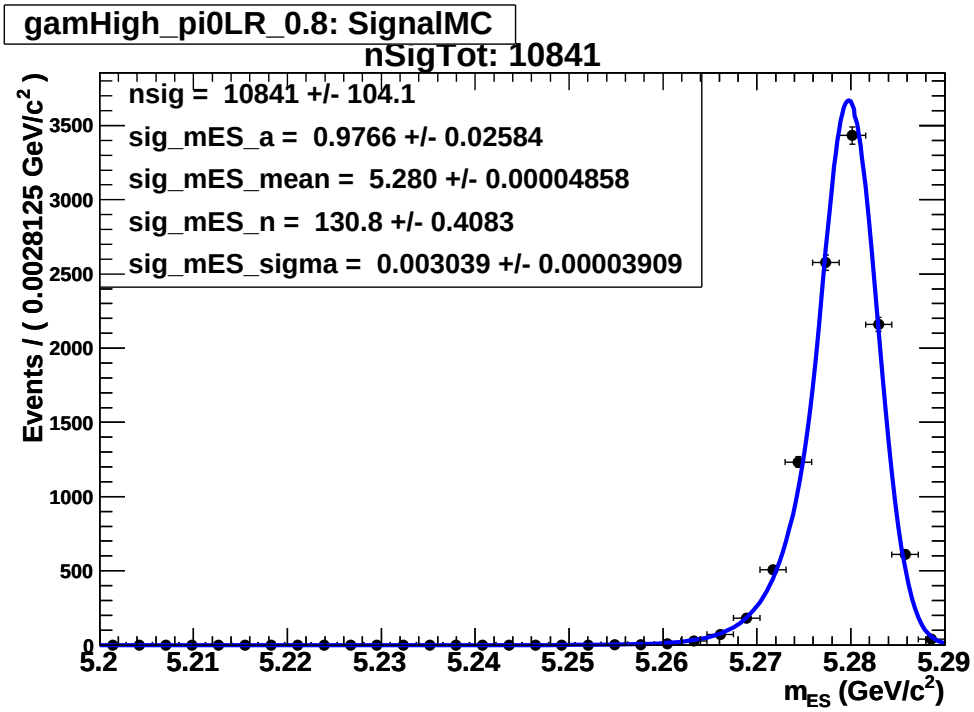
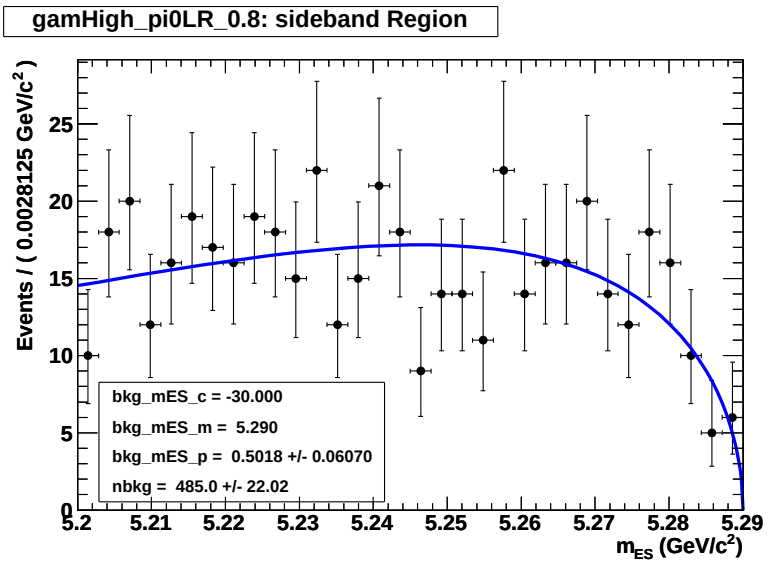
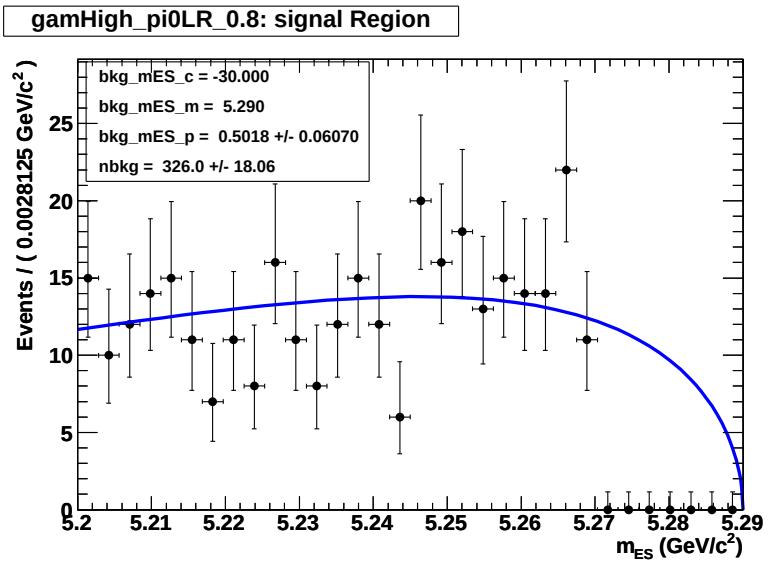


Figure 4.22: A fit to the $m_{\rm ES}$ distribution in signal MC with a Crystal Ball shape. This figure shows the fit after a cut on the $\pi^0 LR < 0.8$ has been applied.

The parameters of the ARGUS are fixed to the sideband fit values and the PDF is refit to the blinded ΔE signal region allowing the normalization to float. The shape of the $m_{\rm ES}$ distribution is assumed to be the same in the signal and sideband regions. A knowledge of the number of events above 5.27 GeV/c^2 in the sideband region and the ratio of the area under the PDFs in the signal and sideband region allow the number of background events, B, to be estimated. Figure 4.23 shows an example of the fits to the $m_{\rm ES}$ distribution for both the ΔE sideband and blinded signal regions. again for a cut on the π^0 LR less than 0.8.



(a) On-resonance ΔE sideband fit



(b) On-resonance blind signal fit

Figure 4.23: (a) A fit to on-resonance ΔE sideband data using an ARGUS shape. (b) The ARGUS shape fit to the blinded signal $m_{\rm ES}$ distribution with the parameters fixed from the sideband fit.

Table 4.4: Selection variables and their final cut values. The values of the π^0 and η LR are chosen as the average of the individual values obtained for the gamHigh and gamLow values.

Selection Variable	Optimized Value
# of B cands	1
`nGTL`	≥ 3
Event Energy	< 15.0 GeV
Dist. to Charged	> 25 cm
Dist. to Neutral	> 25 cm
$\vec{p}_\theta$	$0.4 < \theta < 2.4$
Cluster LAT	$0.15 \leq \ LAT \ \leq 0.5$
Cluster time	$6200 \leq$ time ≤ 6350
# of Crystals	> 10
merged π^0 Consistency	< 0.01
$\pi^0 LR$	≤ 0.84
η LR	≤ 0.84
SPR NN	≥ 0.54

Figures 4.24–4.26 show the FOM as a function of the cut on the optimized variable. The final cuts are shown with an arrow and are listed in Table 4.4 along with the non-optimized cuts. Additionally, as a check, the optimization was run over the other cuts that were not optimized. No indication was found that the initial cut values were not appropriate. Cluster time, as discussed previously was not optimized so there are no plots for this variable. For the number of crystals, the data–MC agreement is not very good and the cut on this variable is not optimized either. The value of 10 was chosen looking at a sample of 'clean' photons in a data sample from $e^+e^- \to \mu\mu\gamma$. This value keeps all "good" photons in both data and MC. The FOM plots for the non-optimized variables are in Appendix B.

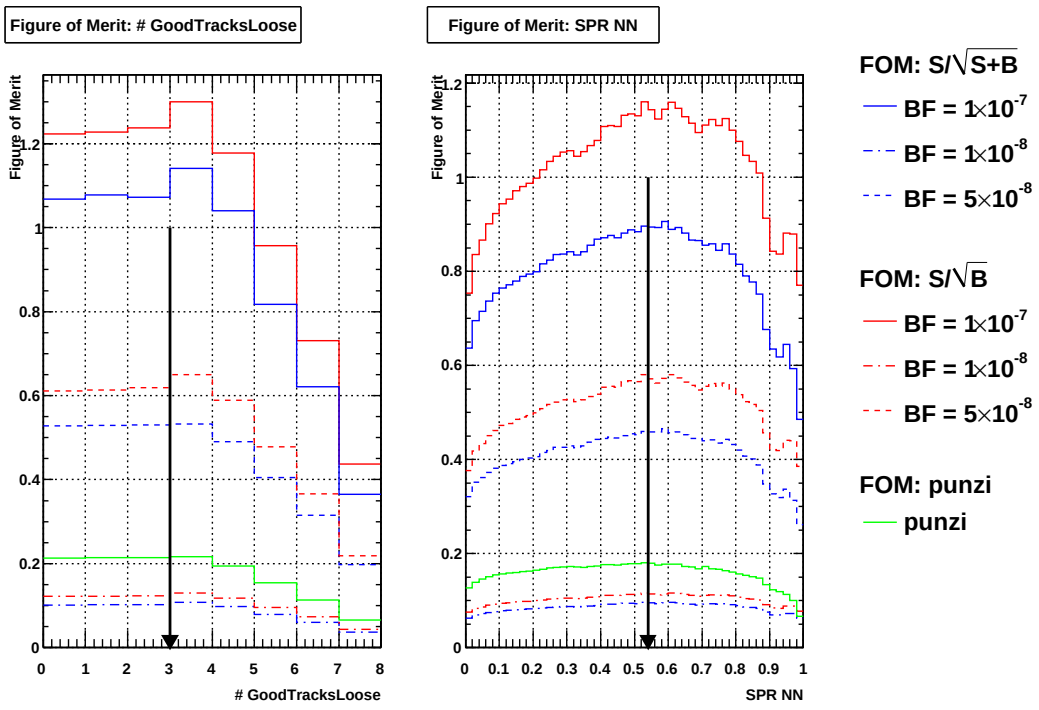


Figure 4.24: Figure of merit as a function of the cut on the optimized variables. Different assumed signal BF are shown to verify that the optimized values don't depend on the branching fraction. The final optimized values are taken where the histogram peaks and shown by the arrow.

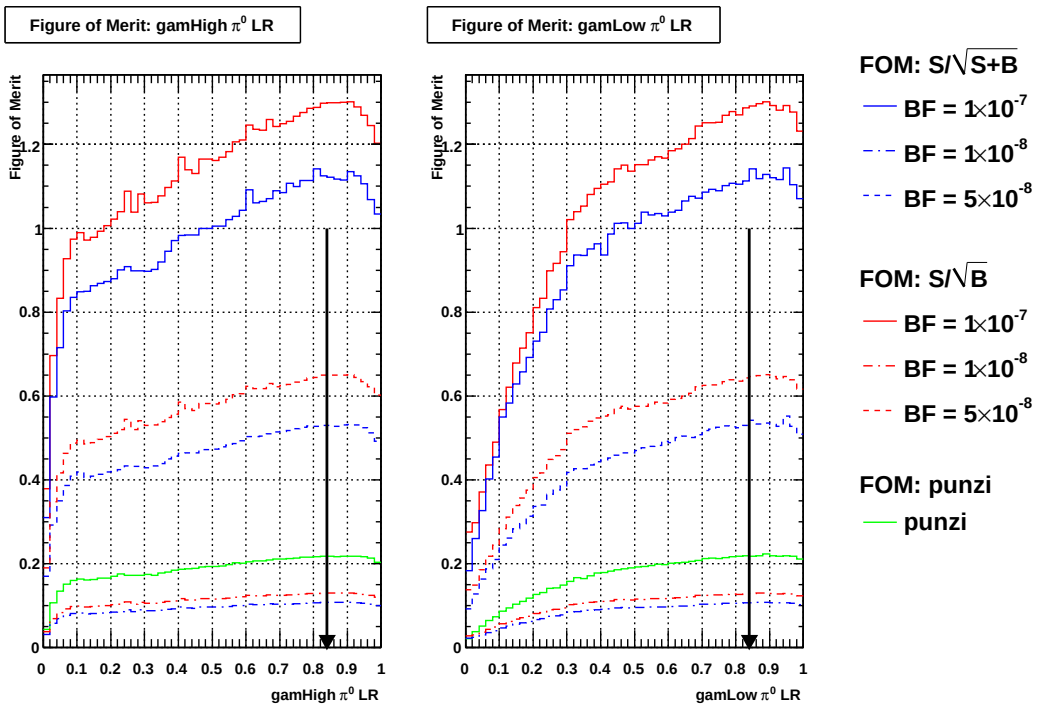


Figure 4.25: Figure of merit as a function of the cut on the optimized variables. Different assumed signal BF are shown to verify that the optimized values don't depend on the branching fraction. The final optimized values are taken where the histogram peaks and shown by the arrow.

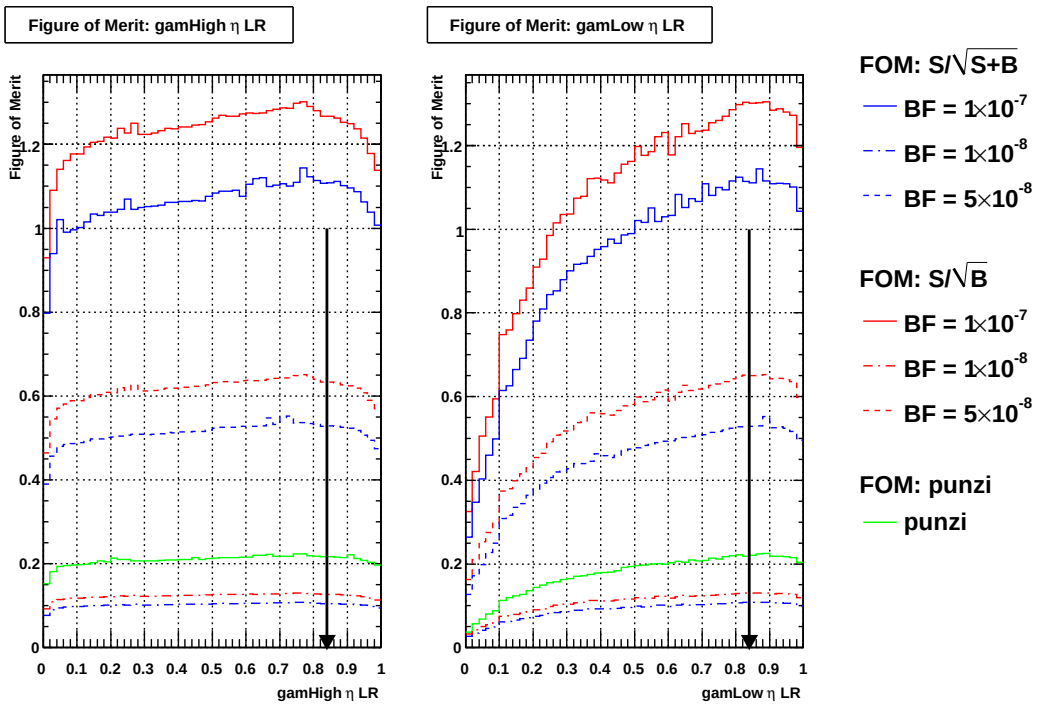


Figure 4.26: Figure of merit as a function of the cut on the optimized variables. Different assumed signal BF are shown to verify that the optimized values don't depend on the branching fraction. The final optimized values are taken where the histogram peaks and shown by the arrow.

4.10 Signal Efficiency

The equation to calculate the branching fraction for $B \to \gamma\gamma$ is given by

$$\mathcal{B}(B \to \gamma\gamma) = \frac{N_{sig}}{\varepsilon_{sig} \cdot 2 \cdot N_{B^0\overline{B}^0}} \tag{4.16}$$

where N_{sig} is the number of signal events, $2 \cdot N_{B^0\overline{B}^0}$ is the number of neutral B mesons in the data sample, and ε_{sig} is the signal efficiency derived from signal MC. The total number of signal events that pass the final selections is 551152 out of a total of 1962000 events generated. This gives a total signal efficiency of $26.7 \pm 0.03\%$. Table 4.5 shows the efficiencies of the final cuts on signal MC.

4.11 Peaking Backgrounds

Decay processes that contain high energy photons could potentially pass all final cuts and mimic a signal event. To see which decays contribute the final cuts were applied to exclusive B MC collections to determine the number of events that are expected in the on resonance data. When the mode has an upper limit published, that number is used as the assumed branching fraction for a worst case scenario. The expected number of peaking events in the fit region is 3.13 ± 0.53. This is comparable to the number of signal events expected, and so a peaking component is included in the fit. Because the $m_{\rm ES}$ and ΔE peaking background distributions, shown in Figure 4.27, can not be easily fit with a function, a histogram PDF is used to model the shapes. Table 4.6 shows a summary of the expected numbers of events for each mode in on-resonance data. The individual mode $m_{\rm ES}$ and ΔE distributions and the event scatter in the $m_{\rm ES} \times \Delta E$ plane are shown in Appendix D.

Table 4.5: Performance of each cut on signal MC. The efficiencies are calculated with respect to the total number of signal events generated in the full *BABAR* MC simulation. All cuts include the cuts listed above them.

# of events Generated:	1,962,000	
# of events passing Skim:	1,371,411	($\varepsilon = 0.699$)
nB=1:	1,370,618	($\varepsilon = 0.698$)

Cut	**Value**	Cumulative	
		# pass	ε
nGTL	≥ 3	1,063,339	0.542
E_{tot}	< 15.0 GeV	1,054,073	0.537
Dist to charged	> 25 cm	1,010,012	0.515
Dist to neutral	> 25 cm	899,871	0.459
$\vec{p}$ polar angle	$0.4 < \theta < 2.4$	834,201	0.425
Lateral Moment	$0.15 \leq$LAT≤ 0.5	798,601	0.407
cluster time	$6200 \leq t \leq 6350$ ns	775,818	0.395
# of Crystals	> 10	775,817	0.395
merged π^0 Cons.	< 0.01	767,376	0.391
π^0 LR	≤ 0.84	690,796	0.352
η LR	≤ 0.84	559,712	0.285
Neural Net	≥ 0.54	523,286	0.267
	Total Signal Efficiency		0.267

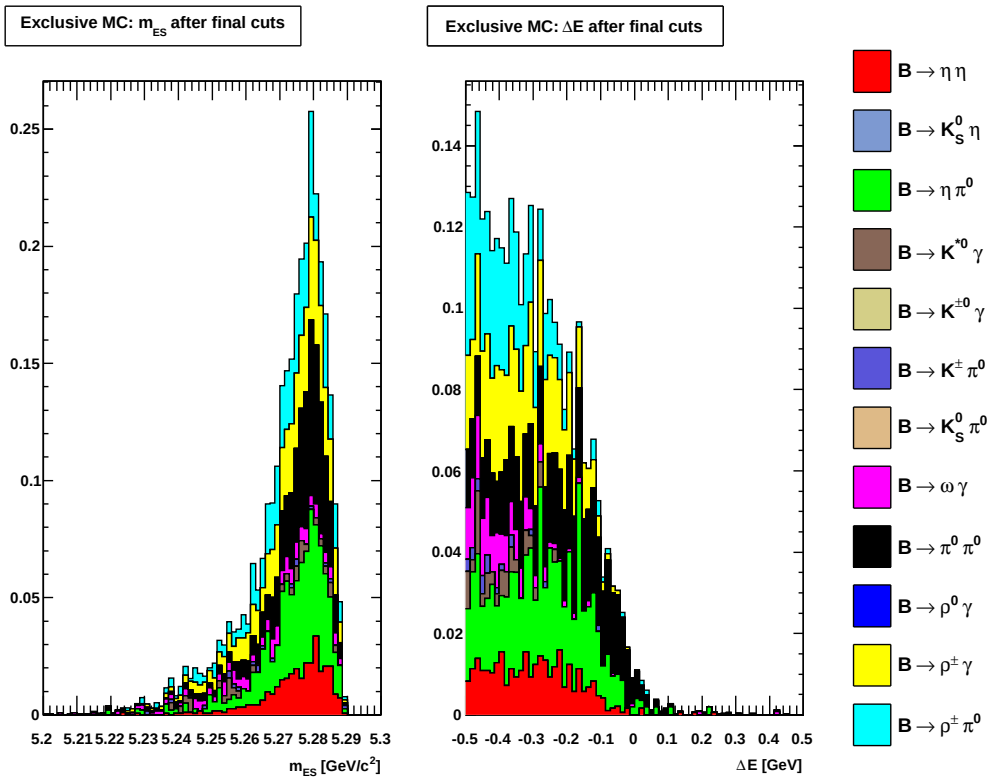


Figure 4.27: $m_{\rm ES}$ and ΔE distributions of a luminosity weighted combination of exclusive modes listed in Table 4.6.

Table 4.6: Number of events passing the optimized cuts for each exclusive MC dataset and the number expected in our fit and blinded signal region for a dataset of 425.7 fb^{-1}. The errors are from the branching fractions. The value of $\mathcal{L}$ factor is derived by dividing the on-resonance luminosity by the equivalent luminosity of the exclusive mode dataset. If no events pass the selections, we assume one event. The equivalent luminosity of each dataset is given in Table 3.3. Plots of $m_{\rm ES}$, ΔE, and the 2D $m_{\rm ES}$-ΔE plane for the events surviving the cuts for each mode are shown in Appendix D.

Num	Mode	$\mathcal{B}_{assumed}$ 10^{-7}	N_{gen}	$\mathcal{L}$ factor	In Fit Region		In Signal Region	
					N	N_{exp}	N	N_{exp}
1043	$B^0 \to \pi^0\ \pi^0$	16.2 ± 3.1	3,889,000	0.000186	3574	0.665 ± 0.128	2139	0.398 ± 0.077
2623	$B^0 \to \eta\eta$	5 ± 3[24]	434,000	0.0005	601	0.310 ± 0.186	246	0.127 ± 0.076
2624	$B^0 \to \eta\pi^0$	9 ± 4[25]	203,000	0.002	353	0.699 ± 0.313	168	0.333 ± 0.150
1940	$B^\pm \to \rho^\pm\pi^0$	109.0 ± 14.0	3,889,000	0.0013	418	0.524 ± 0.072	71	0.089 ± 0.016
995	$B^\pm \to \rho^\pm\gamma$	9.8 ± 2.5	587,000	0.00071	233	0.174 ± 0.032	18	0.013 ± 0.004
1983	$B^0 \to \rho^0\gamma$	8.6 ± 1.5	587,000	0.00065	0	< 0.00065	0	< 0.00065
1984	$B^0 \to \omega\gamma$	$4.4^{+1.8}_{-1.6}$	587,000	0.00033	1821	0.610 ± 0.347	622	0.208 ± 0.119
1587	$B^\pm \to K^\pm\ \pi^0$	129.0 ± 6.0	3,889,000	0.0015	0	< 0.0015	0	< 0.0015
3135	$B^\pm \to K^{*\pm}\ \gamma$	403.0 ± 26.0	5,828,000	0.0031	28	0.087 ± 0.017	3	0.009 ± 0.005
3134	$B^0 \to K^{*0}\ \gamma$	401.0 ± 20.0	5,828,000	0.0031	19	0.059 ± 0.014	1	0.003 ± 0.003
1442	$B^0 \to K^0_S\pi^0$	33.9 ± 2.1	3,931,000	0.00039	0	< 0.0004	0	< 0.0004
10174	$B^0 \to K^0_S\eta$	< 1.1	650,000	0.00007	79	< 0.0059	22	< 0.0016
				Total Expected		3.13 ± 0.53		1.18 ± 0.22

Chapter 5

Fitting the Data

This section describes the fitting strategy employed to extract a signal yield from the data. It also discusses the technique used to validate and check for biases in the fit. Finally, a the procedure used to set an upper limit on the branching fraction for $B \to \gamma\gamma$ is discussed.

5.1 Maximum Likelihood Fit

To extract a signal yield from the on-resonance data, an unbinned extended maximum likelihood fit [35] is performed. For this technique, loose selections are imposed on the discriminating variables in the data so a larger selection of events are present where the shapes can be fit. For each variable x in the fit, a probability density function (PDF), $\mathcal{P}_j(x, \vec{\alpha}_j)$, is constructed that describes the distribution for each class of events j, where j can be signal, background, or peaking background. The parameters of the PDF, $\vec{\alpha}_j$, are chosen for the best fits the distribution in data, and are either fixed beforehand from a source other than the data or are determined during the fit.

A likelihood function for each event in the fitted sample is computed at the values of the fit variables and has the form

$$\mathcal{L}_i = N_{sig}\mathcal{P}_{sig}(x_i, \vec{\alpha}_{sig}) + N_{bkg}\mathcal{P}_{bkg}(x_i, \vec{\alpha}_{bkg}) + N_{peak}\mathcal{P}_{peak}(x_i), \tag{5.1}$$

where N_{sig}, N_{bkg}, and N_{peak}, are the number of events for signal, background,

Table 5.1: Components of the total PDF used in the maximum likelihood fit.

Observable	Signal	Background	Peaking
$m_{\rm ES}$	Crystall Ball	ARGUS	Histogram
ΔE	Cruijff	Polynomial $O(1)$	

and peaking background components. The values of the N's are called yields. The total likelihood for the sample is then the product of the likelihoods for the individual events. A factor is multiplied to the likelihood product that describes the Poissonian nature of the yields with respect to the overall number of events in the data sample. The total likelihood is:

$$\mathcal{L} = \frac{e^{-N}}{N!} \prod_{i=0}^{N} (N_{sig}\mathcal{P}_{sig}(x_i, \vec{\alpha}_{sig}) + N_{bkg}\mathcal{P}_{bkg}(x_i, \vec{\alpha}_{bkg}) + N_{peak}\mathcal{P}_{peak}(x_i)) \quad (5.2)$$

where $N = N_{sig} + N_{bkg} + N_{peak}$.

The maximum likelihood fit works by finding the maximum of this function with respect to the PDF parameters, $\vec{\alpha}_j$ and the yields. This fit is done with the `Minuit` package [36] which uses an iterative instead of analytical method for finding extremum and so what is actually occurring is a minimization of the negative logarithm of this function.

5.2 PDFs

This section describes the PDFs used to fit the $m_{\rm ES}$ and ΔE distributions. A summary of this information is listed in Table 5.1.

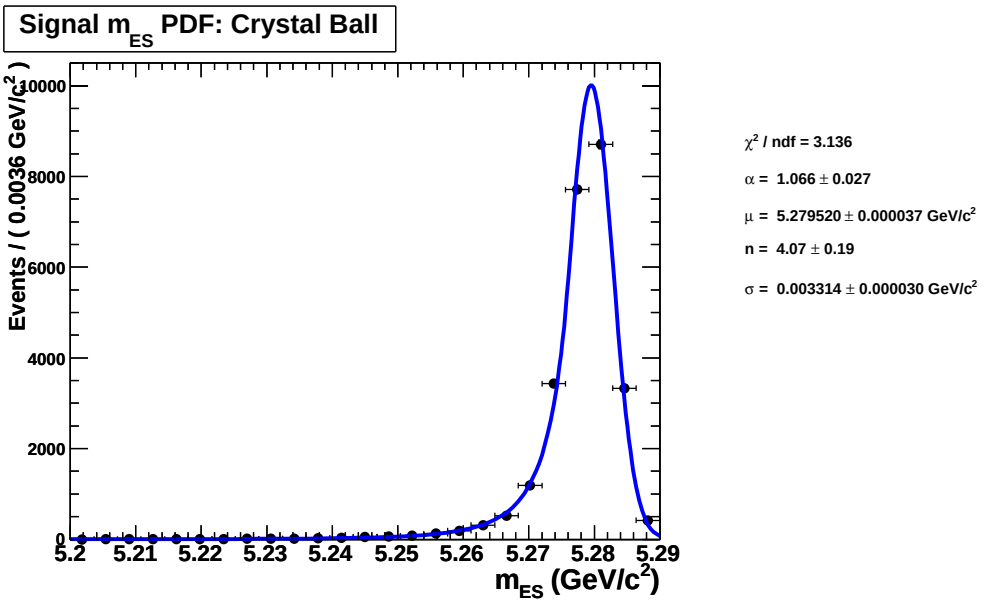
Table 5.2: Values of the fixed signal PDF parameters for the $m_{\rm ES}$ Crystal Ball and ΔE Cruijff shapes. The values are obtained from fits to signal MC.

Parameter	Value
Signal $m_{\rm ES}$: Crystal Ball	
α	1.066 ± 0.027
μ	5.279 ± 0.00004
n	4.07 ± 0.19
σ	0.0033 ± 0.00003
Signal ΔE: Cruijff	
α_L	0.218 ± 0.006
α_R	0.129 ± 0.003
μ	-0.008 ± 0.002
σ_L	0.101 ± 0.002
σ_R	0.038 ± 0.001

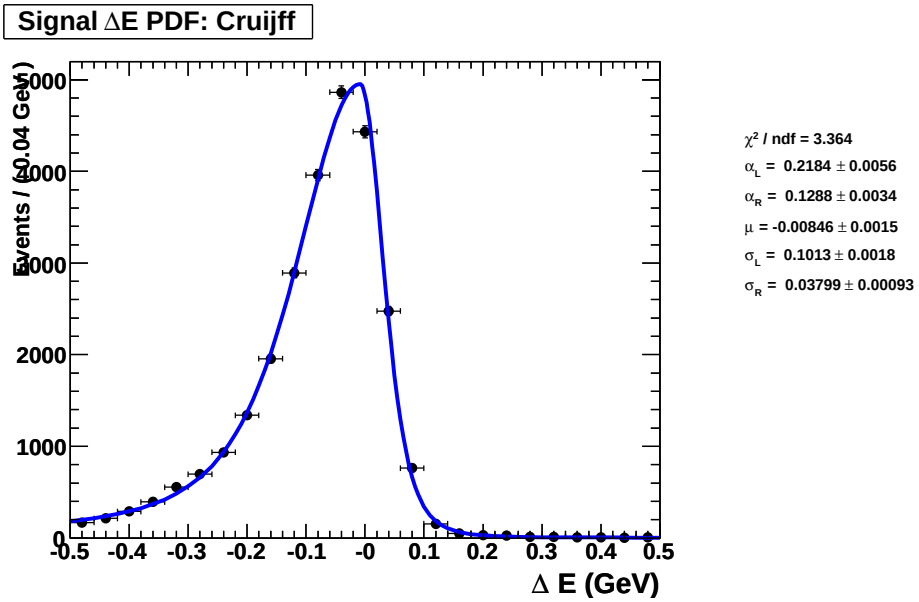
5.2.1 Signal PDFs

The signal $m_{\rm ES}$ distribution is described by a Crystal Ball shape, (4.13). The parameters are fixed in the final fit to the values determined from a fit to signal MC in the range 5.2–5.3 GeV/c^2. Figure 5.1(a) shows the fit to signal MC and values of the parameters used in the fit for the signal yield.

The signal ΔE distribution is described best by a double sided Gaussian function with a low side tail parameter that is known in *BABAR* as a Cruijff shape, eq. (4.14). These parameters are fixed to values determined from a fit over the entire ΔE range in signal MC shown in Figure 5.1(b). Although the shape of the Cruijff on the low ΔE side of the peak is not in perfect agreement, the discrepancy will be dealt with as a systematic by varying the PDF, but the systematic error is expected to be small. All signal parameters are fixed in the final fit and are listed in Table 5.2.



(a) Signal $m_{\rm ES}$ Crystal Ball shape



(b) Signal ΔE Cruijff shape

Figure 5.1: Signal $m_{\rm ES}$ and ΔE PDF shapes from a fit to signal MC after all final cuts are applied.

5.2.2 Background PDFs

The background $m_{\rm ES}$ distribution is fit with an ARGUS shape, eq. (4.15), with the endpoint fixed to 5.29 GeV/c^2 and the slope allowed to float in the final fit. Figure 5.2(a) shows the fit to the combined ΔE sideband in on-resonance data. The parameters listed in the figures are used to generate the background distributions for the toy MC studies.

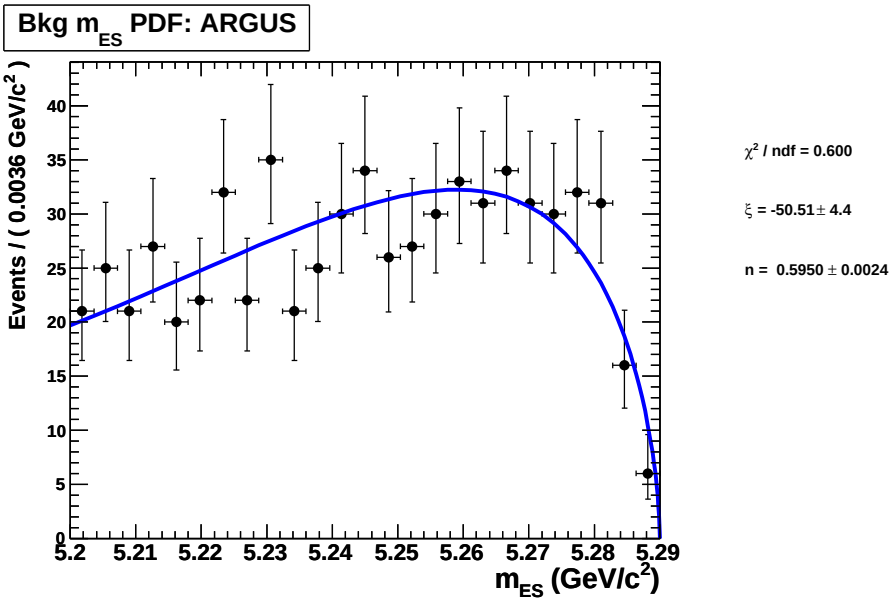
The background ΔE distribution is fit with a first order polynomial. The slope is determined from the $m_{\rm ES}$ sideband in on-resonance data, $m_{\rm ES} < 5.27\,{\rm GeV}/c^2$, and is floated in the final fit. Figure 5.2(b) shows the fit to the $m_{\rm ES}$ sideband.

5.2.3 Peaking Background PDFs

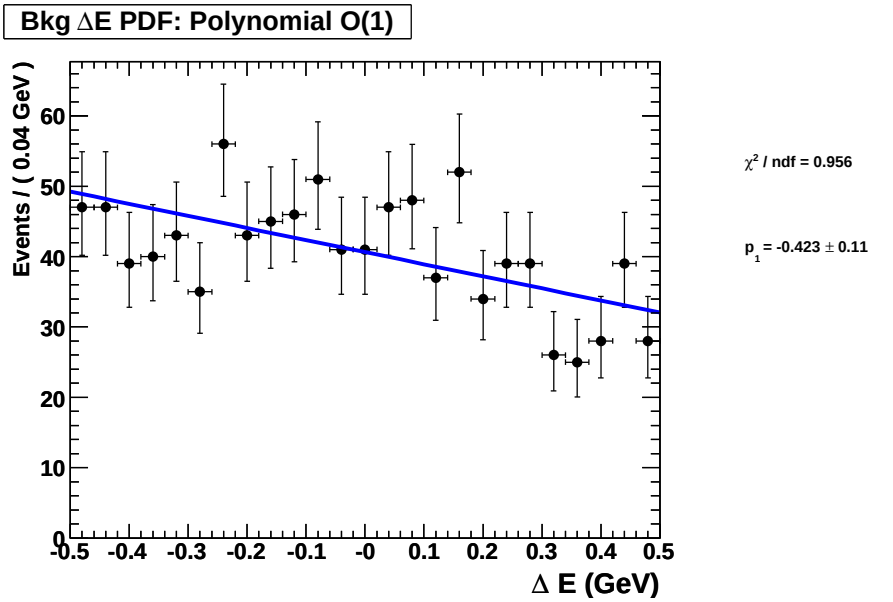
The number of peaking background events in the signal region, 1.18 ± 0.22, is comparable to the number of signal expected, 4.59, using the SM branching fraction. These events are dealt with by taking advantage of the difference in shapes of the signal and peaking background distributions. A 2D histogram PDF in $m_{\rm ES}$ and ΔE is used with the normalization fixed to the expected number of events from MC. Figure 5.3 show the 1D projections of the histogram PDF using a 25 bins which is also used for the final fit.

5.3 Fit Validation

To validate and check for biases in the maximum likelihood fit a series of "toy" MC experiments were performed. This technique involved generating mock datasets by randomly sampling the PDFs of the fit model and then fitting the datasets with the full extended maximum likelihood fit. The PDFs used

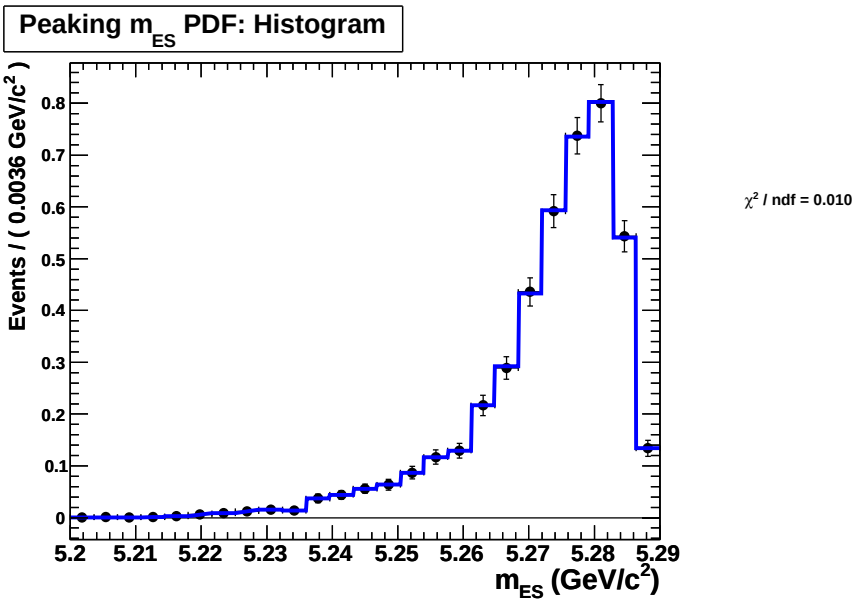


(a) Background $m_{\rm ES}$ ARGUS shape

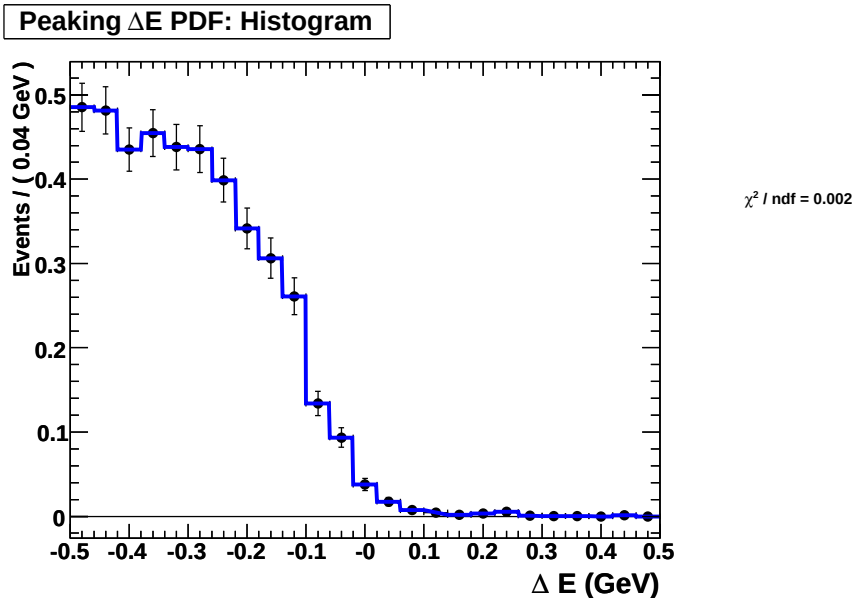


(b) Background ΔE Polynomial shape

Figure 5.2: Background $m_{\rm ES}$ and ΔE PDF shapes from a fit to on-resonance sideband data after all final selections are applied. The $m_{\rm ES}$ projection is over the ΔE combined sideband data, while the ΔE projection is over the $m_{\rm ES}$ sideband data.



(a) Peaking Background $m_{\rm ES}$ Distribution



(b) Peaking Background ΔE Distribution

Figure 5.3: 1D projections of the histogram PDF used in the fit as the peaking background component. The number of bins in the PDF is 25.

to generate the toy datasets are set to those values expected from data. The known values of the generated signal and background yields can be compared to those returned from the fit. By running many such toy experiments, the distributions of a quantity called the *pull* give a measure of the performance of the fit. For some parameter of interest, x, the pull is defined as

$$\text{Pull} \equiv \frac{x_{fit} - x_{gen}}{\sigma_{x_{fit}}} \tag{5.3}$$

and should follow a Gaussian distribution with mean zero (no bias) and width equal to one (fit error is consistent with resolution of distribution).

Two types of toy experiments are performed: (1) *pure* toy experiments were the signal and background events are sampled from the PDFs and (2) *signal-embedded* toy experiments where the background events are sampled from the background PDF while the signal events are randomly selected from the MC. This section describes the results for both the pure and signal-embedded toy MC studies. Example fits to the toy datasets are shown in Appendix E.

5.3.1 Pure Toy MC

To check that the fit model is performing accurately 2000 pure toy MC experiments are performed. For each experiment a dataset containing the total number of events expected was generated from the PDF shapes listed in Section 5.2. The number of signal events generated is dependent on the assumed branching fraction for $B \to \gamma\gamma$. Three different branching fractions were tested. The yields and pulls for all three assumed signal branching fraction are shown Figures 5.4–5.6 with the results summarized in Table 5.3.

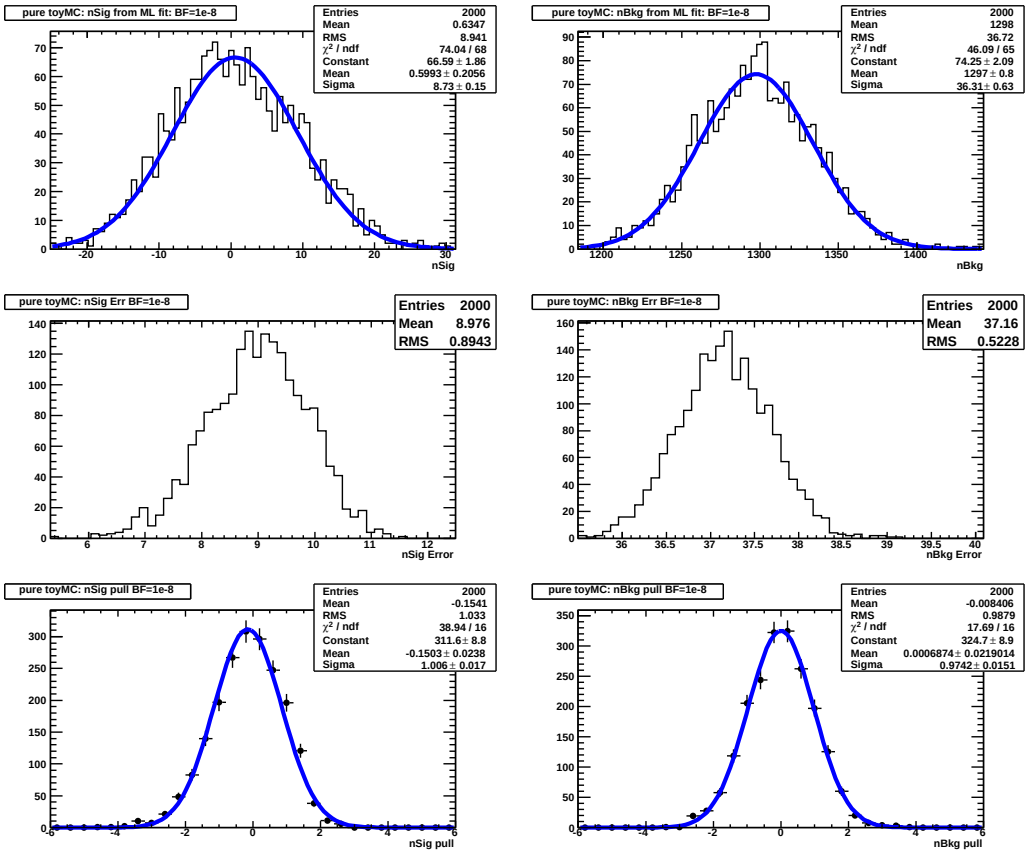


Figure 5.4: Yields, yield errors, and pull plots with a Gaussian fit overlaid for pure toy MC experiments assuming a signal branching fraction of 1×10^{-8}. The left column of plots is for the signal yield and the right column is for background. The number of signal events was generated according to a Poisson distribution with a mean value of 1.21.

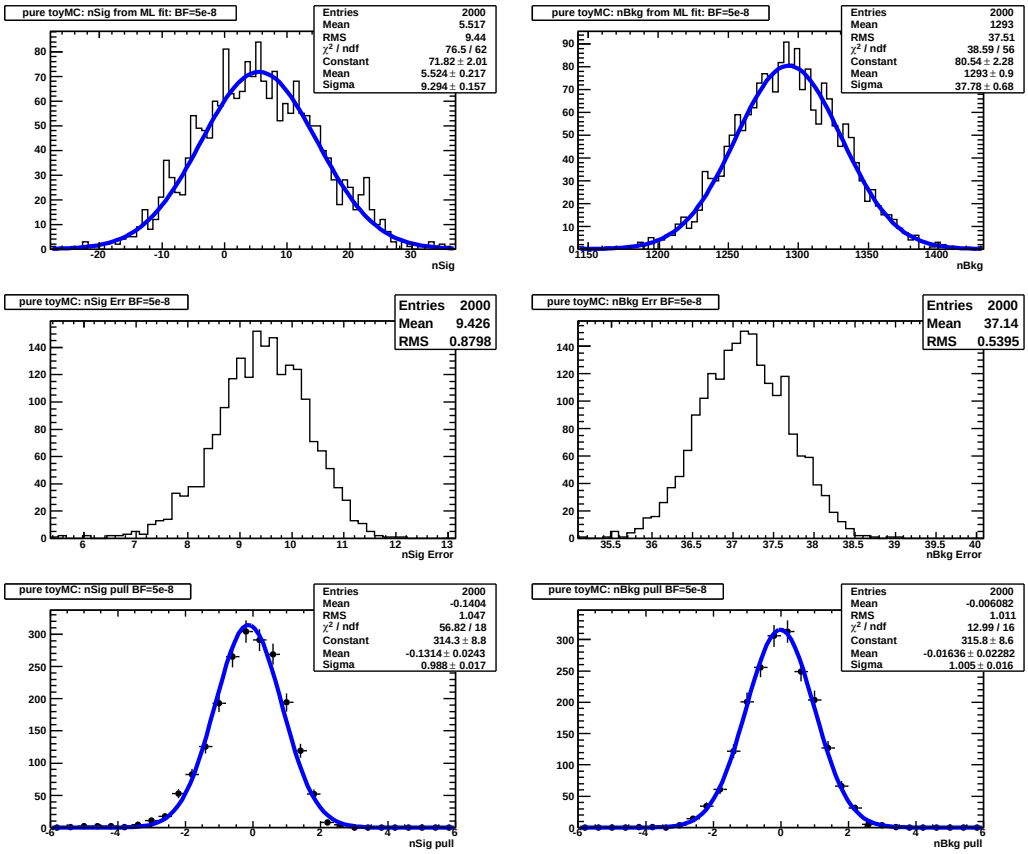


Figure 5.5: Yields, yield errors, and pull plots with a Gaussian fit overlaid for pure toy MC experiments assuming a signal branching fraction of 5×10^{-8}. The left column of plots is for the signal yield and the right column is for background. The number of signal events was generated according to a Poisson distribution with a mean value of 6.03.

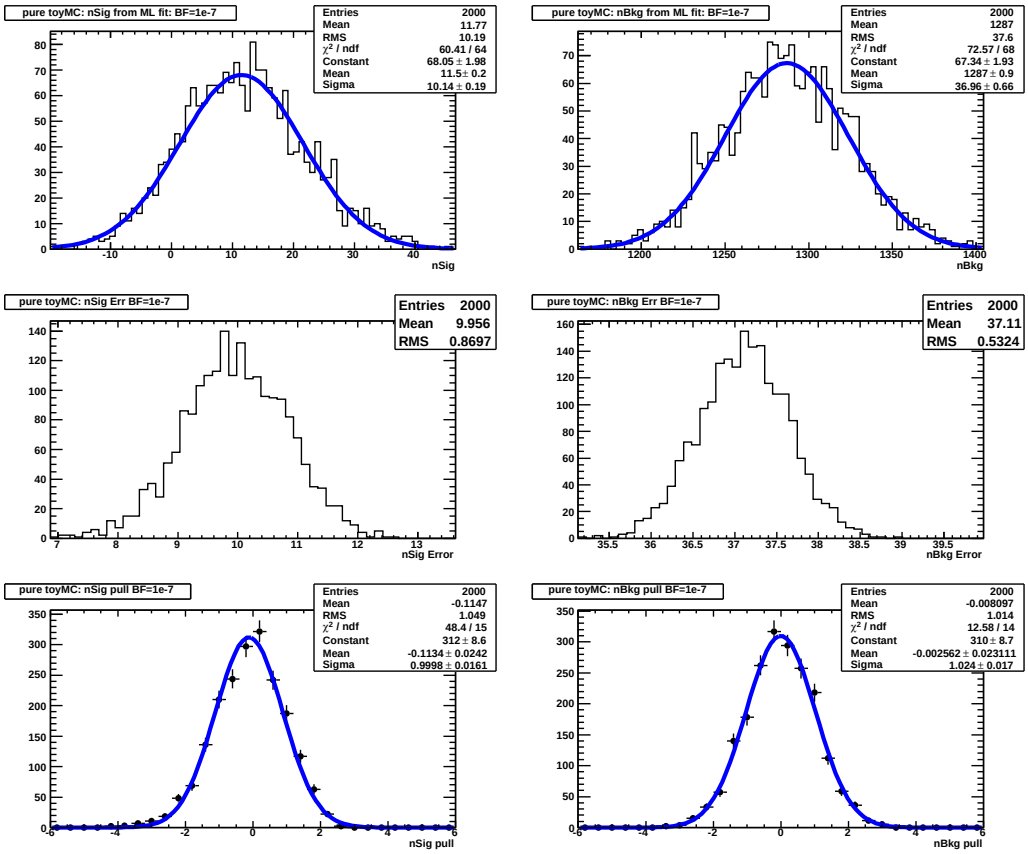


Figure 5.6: Yields, yield errors, and pull plots with a Gaussian fit overlaid for pure toy MC experiments assuming a signal branching fraction of 1×10^{-7}. The left column of plots is for the signal yield and the right column is for background. The number of signal events was generated according to a Poisson distribution with a mean value of 12.1.

Table 5.3: Summary of results for pure Toy MC experiments using the optimized cuts. All parameters come from a fit to the distributions of the pure toy studies.

$\mathcal{B}_{assumed}$		Generated	Mean (fit)	Width (fit)	Pull Mean	Pull width
1×10^{-8}	N_S	1.21	0.599 ± 0.206	8.73 ± 0.15	-0.150 ± 0.024	1.001 ± 0.017
	N_B	1297	1297 ± 0.8	36.31 ± 0.63	0.001 ± 0.022	0.974 ± 0.015
5×10^{-8}	N_S	6.03	5.524 ± 0.217	9.294 ± 0.157	-0.131 ± 0.024	0.958 ± 0.017
	N_B	1293	1293 ± 0.9	37.78 ± 0.68	-0.016 ± 0.023	1.005 ± 0.016
1×10^{-7}	N_S	12.1	11.5 ± 0.2	10.14 ± 0.19	-0.115 ± 0.024	0.999 ± 0.016
	N_B	1287	1287 ± 0.9	36.96 ± 0.66	-0.003 ± 0.023	1.024 ± 0.017

5.3.2 Signal-Embedded Toy MC

To check that the fit model does not introduce a bias on the signal yield and account for correlations among the fit variables or other improperly modeled effects, 2000 signal-embedded toy MC experiments are performed. For each experiment a dataset containing the total number of events expected in on-resonance data was generated from the PDF shapes listed in Section 5.2. Instead of generating the signal events from the PDFs, they are embedded into the background dataset directly from the signal MC. The number of signal events generated is dependent on the assumed branching fraction for $B \to \gamma\gamma$. Three different branching fractions were tested. The yields and pulls for all three assumed signal branching fraction are shown Figures 5.7–5.9 with the results summarized in Table 5.4.

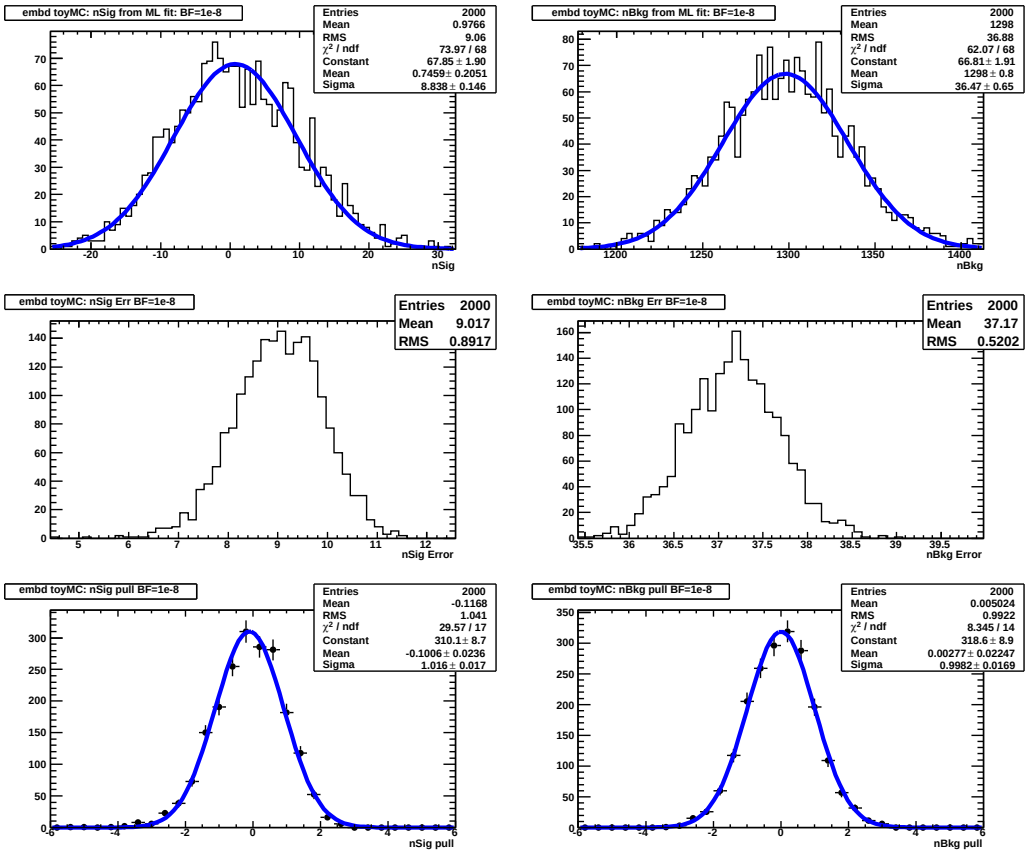


Figure 5.7: Yields, yield errors, and pull plots with a Gaussian fit overlaid for signal-embedded toy MC experiments assuming a signal branching fraction of 1×10^{-8}. The left column of plots is for the signal yield and the right column is for background. The number of signal events was generated according to a Poisson distribution with a mean value of 1.21.

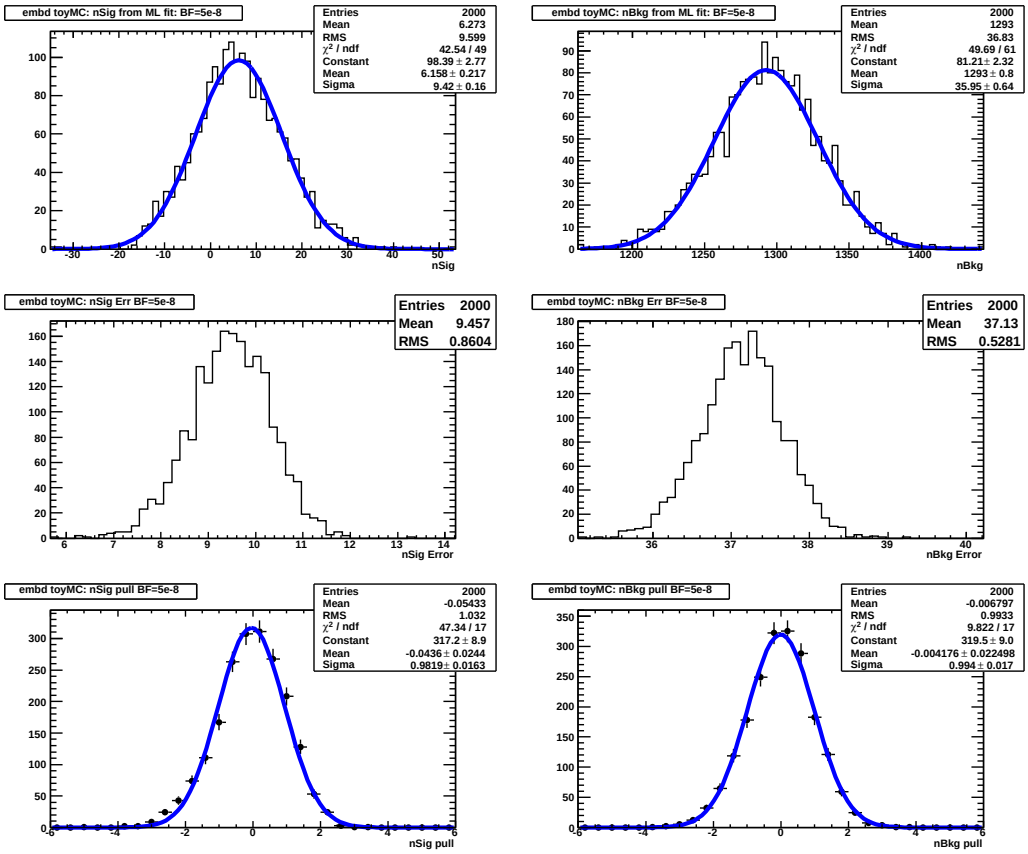


Figure 5.8: Yields, yield errors, and pull plots with a Gaussian fit overlaid for signal-embedded toy MC experiments assuming a signal branching fraction of 5×10^{-8}. The left column of plots is for the signal yield and the right column is for background. The number of signal events was generated according to a Poisson distribution with a mean value of 6.03.

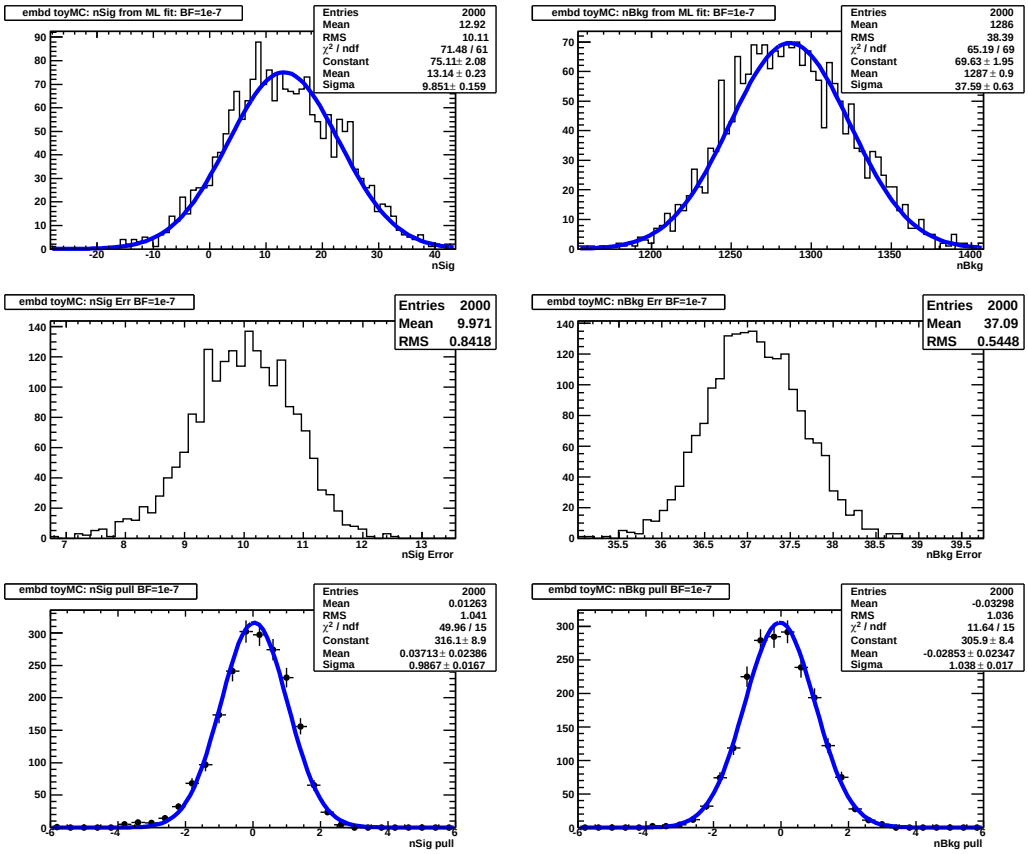


Figure 5.9: Yields, yield errors, and pull plots with a Gaussian fit overlaid for signal-embedded toy MC experiments assuming a signal branching fraction of 1×10^{-7}. The left column of plots is for the signal yield and the right column is for background. The number of signal events was generated according to a Poisson distribution with a mean value of 12.1.

Table 5.4: Summary of results for signal embedded Toy MC experiments using the optimized cuts. All parameters come from a fit to the distributions of the embedded toy MC experiments.

$\mathcal{B}_{assumed}$		Generated	Mean (fit)	Width (fit)	Pull Mean	Pull sigma
1×10^{-8}	N_S	1.21	0.746 ± 0.205	8.838 ± 0.146	-0.101 ± 0.024	1.016 ± 0.017
	N_B	1297	1298 ± 0.8	36.47 ± 0.65	0.003 ± 0.022	0.998 ± 0.017
5×10^{-8}	N_S	6.03	6.158 ± 0.217	9.42 ± 0.16	-0.044 ± 0.024	0.982 ± 0.016
	N_B	1293	1293 ± 0.8	35.95 ± 0.64	-0.004 ± 0.022	0.994 ± 0.017
1×10^{-7}	N_S	12.1	13.14 ± 0.23	9.851 ± 0.159	0.037 ± 0.024	0.987 ± 0.017
	N_B	1287	1287 ± 0.9	37.59 ± 0.63	-0.029 ± 0.023	1.038 ± 0.017

5.4 Upper Limit

The SM expectation for the branching fraction of $B \to \gamma\gamma$ is about 3×10^{-8}, and in the absence of a substantial enhancement only a few signal events are expected. The equation for the branching fraction is given by

$$\mathcal{B}(B \to \gamma\gamma) = \frac{N_{sig}}{2 \cdot N_{B^0\bar{B}^0} \cdot \epsilon_{sig}} \tag{5.4}$$

where N_{sig} is the signal yield of the maximum likelihood fit, ϵ_{sig} is the signal efficiency determined from signal MC, and $N_{B^0\bar{B}^0}$ is the number of neutral B meson pairs in the data sample. For the SM branching fraction only a few signal events are expected and so an upper limit is set. To calculate the upper limit, the likelihood function from the EML fit is integrated from zero up to 90% of its total area. This value corresponds to the 90% confidence level upper limit. An example of the likelihood curve from a signal-embedded toy MC experiment is shown in Figure 5.10. The blue curve is the likelihood where the peak corresponds to the signal yield of the fit. The red line is drawn at the value of N_{sig} where 90% of the area under the curve lies to the left. This value of H_{sig} is used in eq. (5.4) to calculate the upper limit at the 90% confidence Figure 5.11 shows the distribution of the upper limits for each assumed signal branching fractions for pure and signal embedded studies calculated this way. In these plots, the value of $N_{B^0\bar{B}^0}$ is 226.2×10^6, and ϵ_{sig} is 0.267. The mean values of the upper limit distributions are shown in Table 5.5.

Table 5.5: Mean of the upper limit distributions in Figure 5.11 using the stated $B \to \gamma\gamma$ branching fractions.

Assumed $\mathcal{BF}$	Expected UL @ 90% CL ($\times 10^{-7}$)	
	pure Toy MC	signal-embedded Toy MC
1×10^{-8}	1.56	1.55
5×10^{-8}	1.85	1.93
1×10^{-7}	2.30	2.38

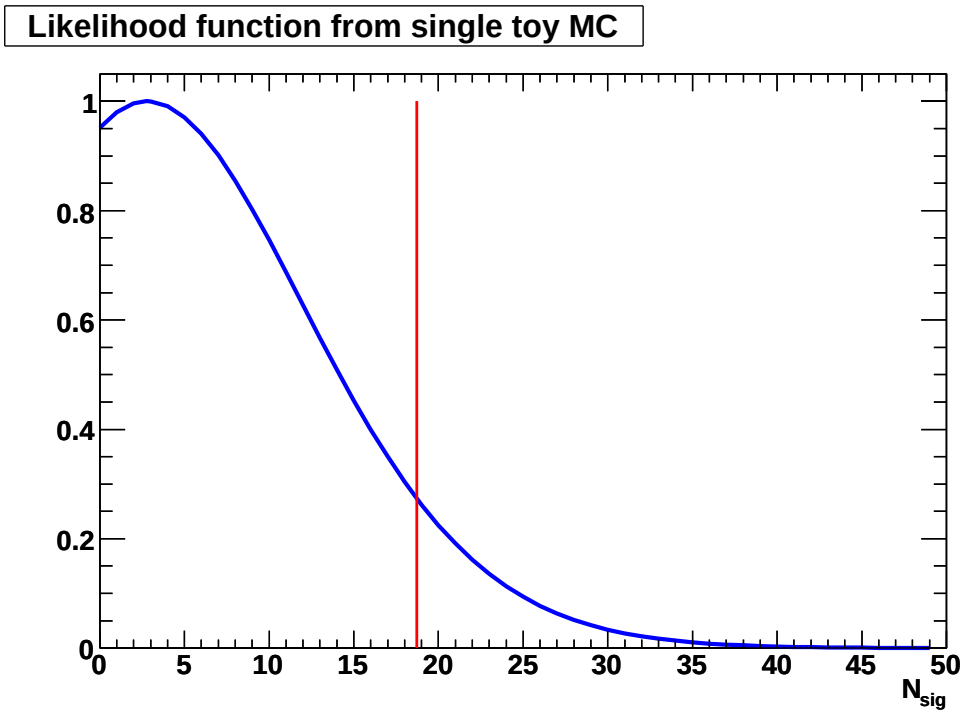


Figure 5.10: Example likelihood function from a signal-embedded toy MC experiment. The maximum of the blue curve is the signal yield, and the red line indicated the value where 90% of the area under the curve is to the left corresponding to the 90% confidence level upper limit on the signal yield.

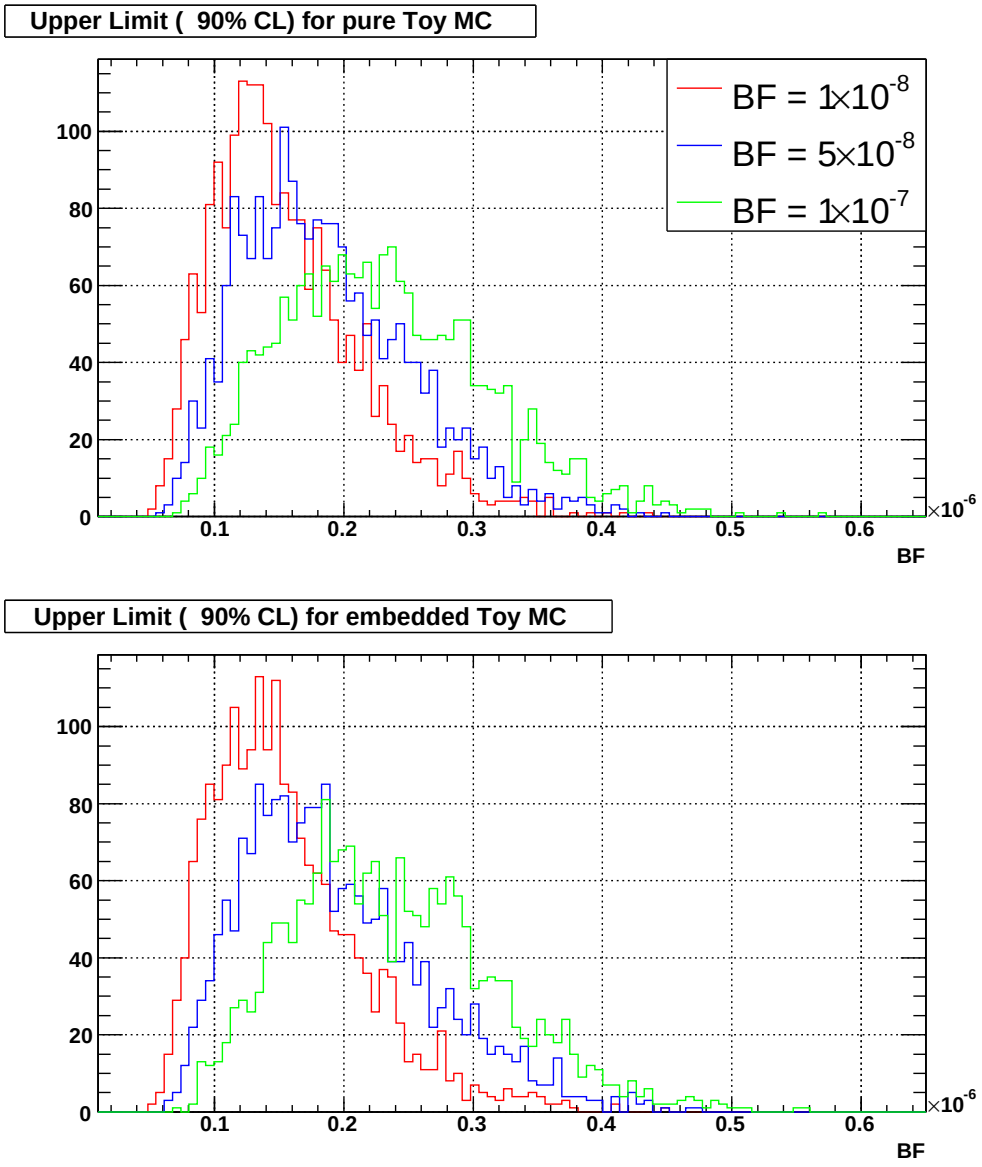


Figure 5.11: Distribution of the upper limits calculated for the toy MC experiments by integrating the likelihood function up to 90%. The top plot is for pure toy MC and the bottom is for signal-embedded toy MC.

Chapter 6

Results

This chapter describes the result of the maximum likelihood fit to the unblinded dataset. The process of using the signal yield from the fit to set an upper limit on the branching fraction of $B \to \gamma\gamma$ is then described. Finally, the identified sources of systematic errors are discussed and the procedure for incorporating them into the final limit is discussed.

6.1 Fit to Unblinded Data Sample

In on-resonance data, 1679 events remain in the fit region after the final cuts listed in Table 4.4 are applied. The data is fit with the extended maximum likelihood fit described in Chapter 5, and the signal yield is found to be $21.3^{+12.8}_{-11.8}$ events with a significance of $1.88\,\sigma$. The projection of the PDF onto the fit variables $m_{\rm ES}$ and ΔE is shown in Figure 6.1. A more instructive way to see the signal events is to apply a cut to the non-plotted variable in the signal region and redraw the projection, i.e., apply a cut on ΔE to be in the signal region, $-0.3 < \Delta E < 0.1312$ GeV, and draw the $m_{\rm ES}$ projection. This is shown in Figure 6.2. The likelihood curve from the maximum likelihood fit is shown in Figure 6.3. Unfortunately, the low significance value of the result does not allow for us to claim an observation of this branching fraction so an upper limit is set using the procedure described in Section 5.4.

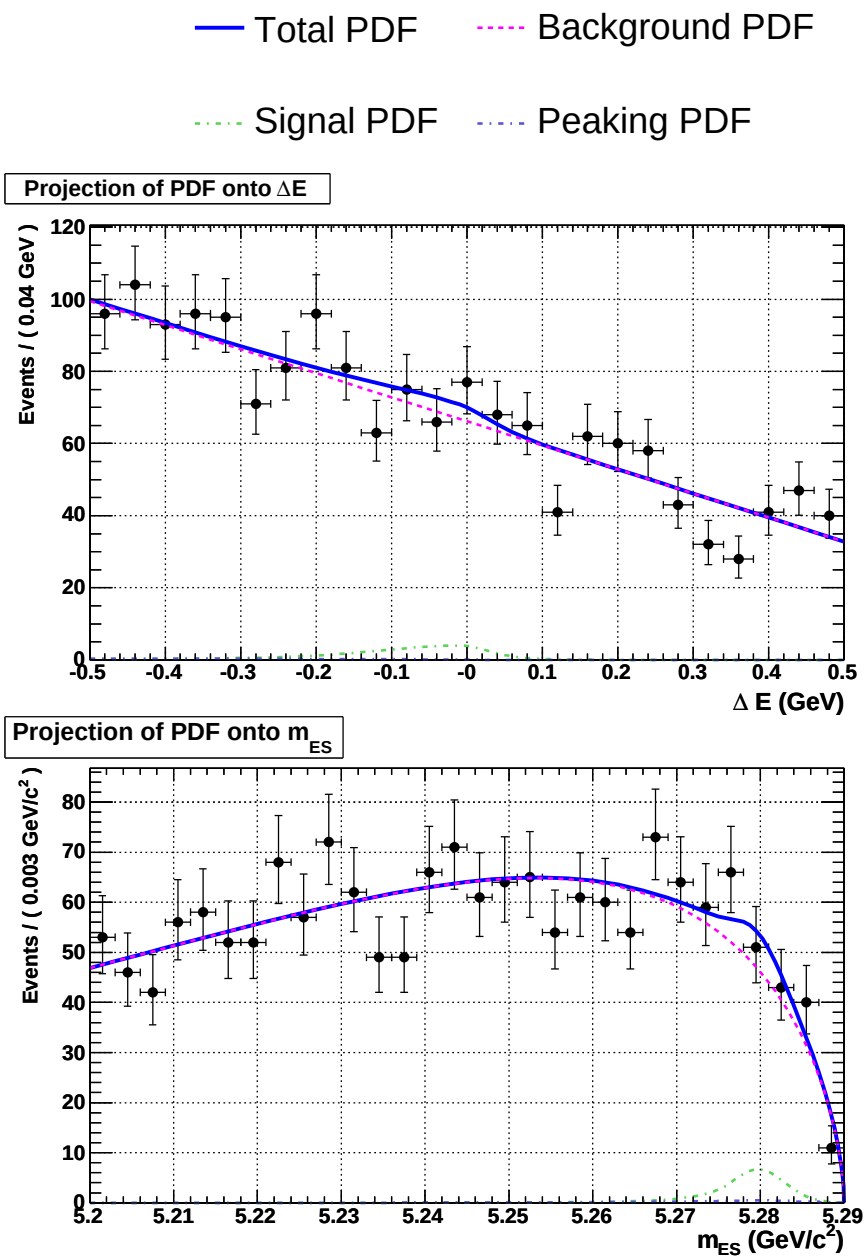


Figure 6.1: The projection of the PDF onto $m_{\rm ES}$ and ΔE. This plot shows all three components of the total PDF (signal, background, and peaking background) as dashed lines and the sum total as the solid blue line.

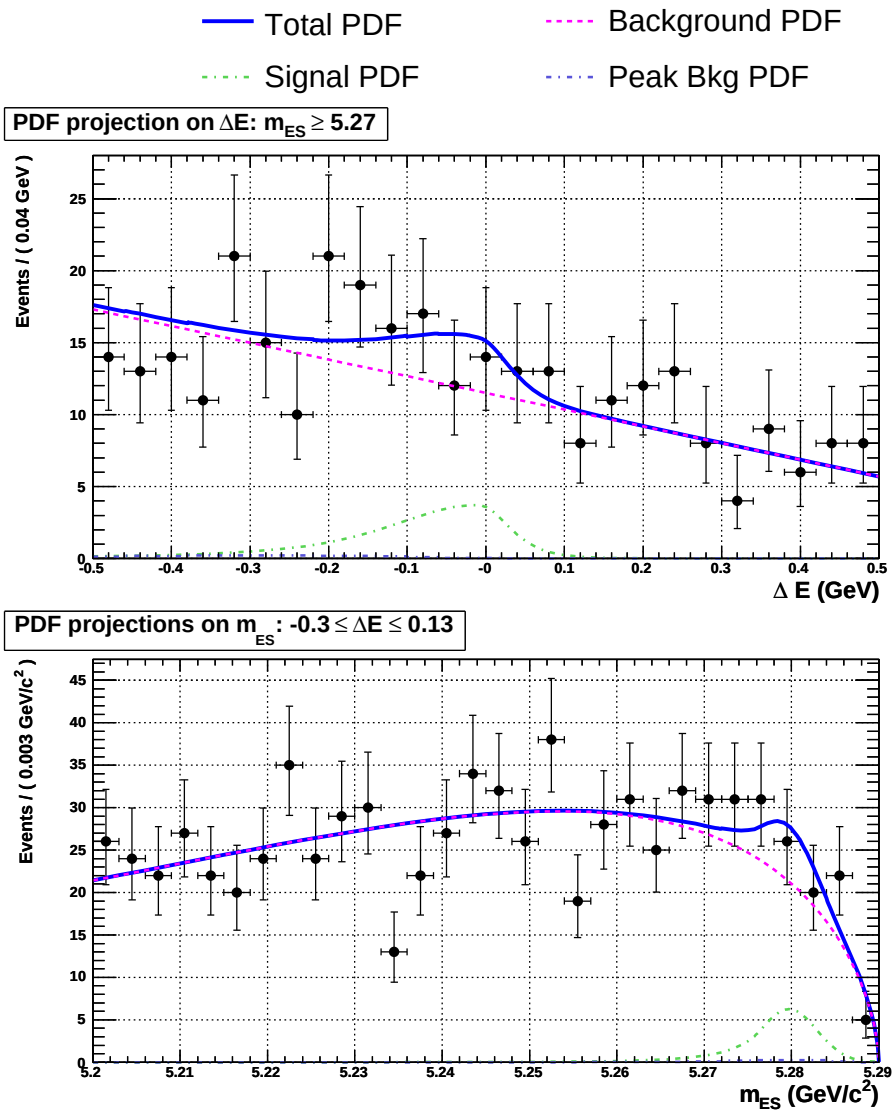


Figure 6.2: Projections of the PDFs onto $m_{\rm ES}$ and ΔE when a cut is applied to the other variable. For the $m_{\rm ES}$ projection the range of ΔE is $-0.3 < \Delta E < 0.1312$ GeV. For the ΔE projection there is a cut on $m_{\rm ES}$ of greater than 5.27 GeV/c^2.

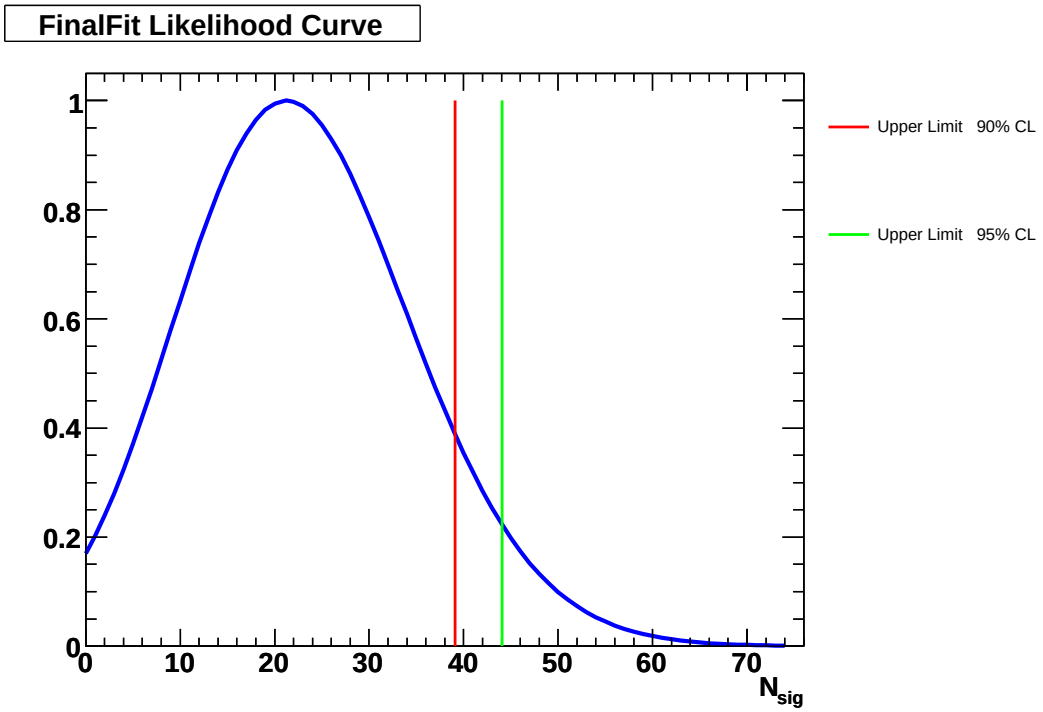


Figure 6.3: Likelihood curve from the extended maximum likelihood fit to the unblinded on-resonance data sample. The peak of the curve represents the signal yield. The red (green) line shows the point below where 90 (95)% of the area is.

6.2 Systematic Errors

In addition to the statistical errors due to the finite size of the data sample, there is an additional uncertainty due to systematic effects that must be applied. Differences in the data and MC distributions of the variables lead to an error in the determination of the efficiency, signal yield and ultimately to a determination of the upper limit for this mode. The following were considered as sources of systematic errors. A summary of the systematic errors is given in Table 6.1.

Table 6.1: This table shows the systematic errors calculated for this analysis. Multiplicative errors are effects that modify the efficiency used in the branching fraction calculation. Additive errors directly affect the signal yield, N_{sig}. The total systematic error is the sum in quadrature of the multiplicative and additive errors in units of events.

Source	**Systematic Error**
Multiplicative Errors (%)	
B-counting	1.6
Track eff.	0.14
Single γ eff.	3.4
Cluster Time	1.0
π^0/η LR	2.0
Neural Network	3.0
Additive Errors (nEvents)	
Fit Sys.	0.55

6.2.1 B Counting

The determination of the number of $B\overline{B}$ events in *BABAR* data is equal to $(467.4 \pm 5.14) \times 10^6$ with a relative uncertainty of 1.1%. To calculate the number of neutral Bs from the luminosity the latest $\Upsilon(4S)$ branching fraction from the PDG is used: $\mathcal{B}(\Upsilon(4S) \to B^0\overline{B}^0) = (48.4 \pm 0.6)\%$. The number of $B^0\overline{B}^0$ pairs is calculated to be $(226.2 \pm 3.7) \times 10^6$ with a systematic error of 1.6%.

6.2.2 Tracking Efficiency

The tracking efficiency can affect this analysis as there is a requirement on the minimum number of GoodTracksLoose in the event greater than two. The charged particle tracking group in *BABAR* has studied the track finding efficiency using τ decays, and found that no efficiency correction should be

applied. However, a systematic error for the GoodTracksLoose list of 0.236% per track, should be applied. This will only affect events with 3 tracks. In signal MC this corresponds to 19.7% of events after the initial selections are applied. A systematic error of $3 \cdot 0.236\% \cdot 0.197 = 0.14\%$.

6.2.3 Single Photon Efficiency

Systematic uncertainties for the selections on the photon are determined using $e^+e^- \to \mu\mu\gamma$ control samples made from the *BABAR* Neutral particles working group. The efficiency is calculated for data and MC when all final selections are applied and the ratio is 0.987. We apply this difference as a systematic error. Additionally the selection on the distance to the nearest charged and neutral cluster of 25 cm is made. This selection was studied in a previous *BABAR* analysis and a 2% discrepancy between data and MC was found. These errors are added in quadrature for the total photon selection systematic error. The systematic error assigned is $\sqrt{2 \cdot (1.3\%^2 + 2\%^2)} = 3.4\%$.

6.2.4 Cluster Time

Since cutting on cluster time is unusual within *BABAR*, this cut was investigated as a possible source of systematic error. The efficiencies of the cluster time cut (applied to both photons) is compared between signal MC and the on-resonance data by using a sample of events in which both photon candidates are consistent with begin from the decay of a π^0, although they do not have the same parent. Such candidates are unlikely to be out of time. The π^0 likelihood ratio is reversed to to select such events. The MC efficiency of this cut is 0.969 ± 0.004 and for data it is 0.972 ± 0.005. The ratio is then 1.006 ± 0.007 and a systematic error of 0.7% is assigned for each photon. The

total systematic error for both photons is the sum in quadrature and gives 1.0%.

6.2.5 π^0/η LR

The π^0/η LR veto systematic effect was studied in a *BABAR* analysis of the decay $B \to \rho\gamma$ [26, 27]. The effect was estimated by embedding high energy photons from $B \to \rho/\omega\gamma$ signal MC into $B \to D\pi$ MC and data samples. The systematic error for this selection is taken from this source. A 1% systematic error was quoted for both the π^0 and η LR vetoes. These errors are added in quadrature for each photon and an overall systematic error of 2.0% is assigned.

6.2.6 Neural Network

MC simulated events are used to train and test the SPR Neural Network classifier and so differences between the data and MC need to be accounted for. To determine the systematic uncertainty for the NN response, the relative efficiency between data and a luminosity weighted MC sample in the fit region, with the signal region excluded, is computed. With all final cuts applied, the efficiency of the NN cut at 0.54 is found to be 0.209 ± 0.004 in on-resonance data and 0.214 ± 0.004 in background MC. The relative efficiency is calculated to be 0.977 ± 0.026. No efficiency correction is applied and a 3% systematic error is assigned.

6.2.7 Fit systematics

A source of systematic error are the differences between the PDFs used in the maximum likelihood fit and the true distributions in the data. These differences may cause a bias in the signal yield and affect the branching frac-

tion calculation. All fixed parameters are possible sources of these systematic errors. Also, an error on the peaking background yield or for the pdf shapes would result in a systematic error. We have estimated the systematic errors in the fit using embedded toys.

To study the effect of fixing the PDF shapes from the signal MC distributions, all parameters listed in Table 5.2 are varied by $\pm 1\sigma$ and the on-resonance data sample is refit and the difference in the signal yield is calculated. After this procedure is run for all fixed parameters, the systematic error is given by adding in quadrature all the differences. To estimate this error, the parameter variations are run using signal-embedded toy MC datasets and the results were found to be 0.16 and 0.2 events when correlations are and are not taken into account between the PDF parameters, respectively. The distribution of the systematic errors for the toy MC experiments is shown in Figure 6.4. When this procedure was run on the unblinded on-resonance dataset, the systematic error on N_{sig} was found to be equal to 0.27 and 0.56 events when correlations are and are not included, respectively.

In addition to changing the signal shape parameters, different shapes were fit to the signal ΔE distribution. Figure 6.5 shows the signal ΔE distribution overlaid with a Cruijff, Crystal Ball, and Novosibirsk shape. The Cruijff shape was replaced with a Crystal Ball shape, and the difference in the signal yield of about -0.15 events was observed in signal-embedded toy MC experiments. This shift may indicate a small bias that may be included as an additional systematic error. Figure 6.6 shows the difference in signal yield between the Cruijff and Crystal Ball shape.

Since the peaking background shape and yield are fixed from MC this becomes a possible source of systematic error. The error for this will be based

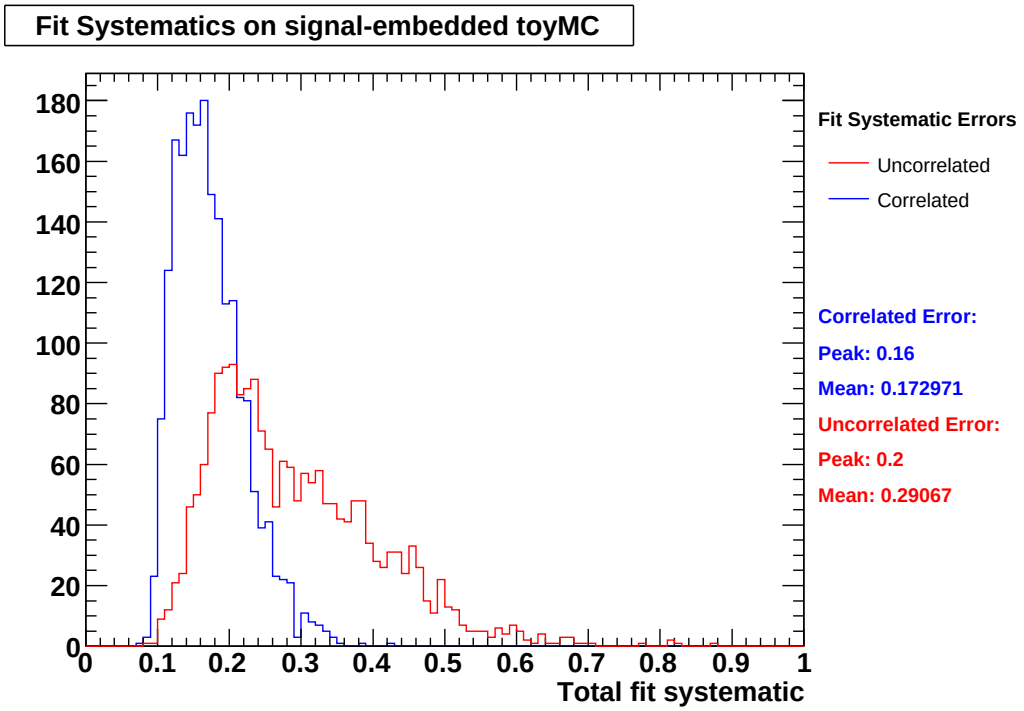


Figure 6.4: The systematic error on the signal yield is found by varying the fixed signal parameters by $\pm 1\,\sigma$ from their central values for 2000 signal-embedded toy MC studies. The blue histogram takes into account the correlations between the parameters while the red histogram does not.

on varying the peaking yield and shape in the fit. As shown in Table 4.6, 3.13 ± 0.53 events are expected in our fit region. As a worst case scenario the errors are added linearly to give 3.13 ± 1.11 events. The systematic error will be estimated using this larger variation. Figure 6.7 shows the difference in the signal yield is about 0.15 events when the peaking component is varied by $\pm 1\sigma$.

Additionally the peaking shape is varied by replacing the shape in Figure 4.27 by the exclusive mode with the most peaking exclusive mode which is

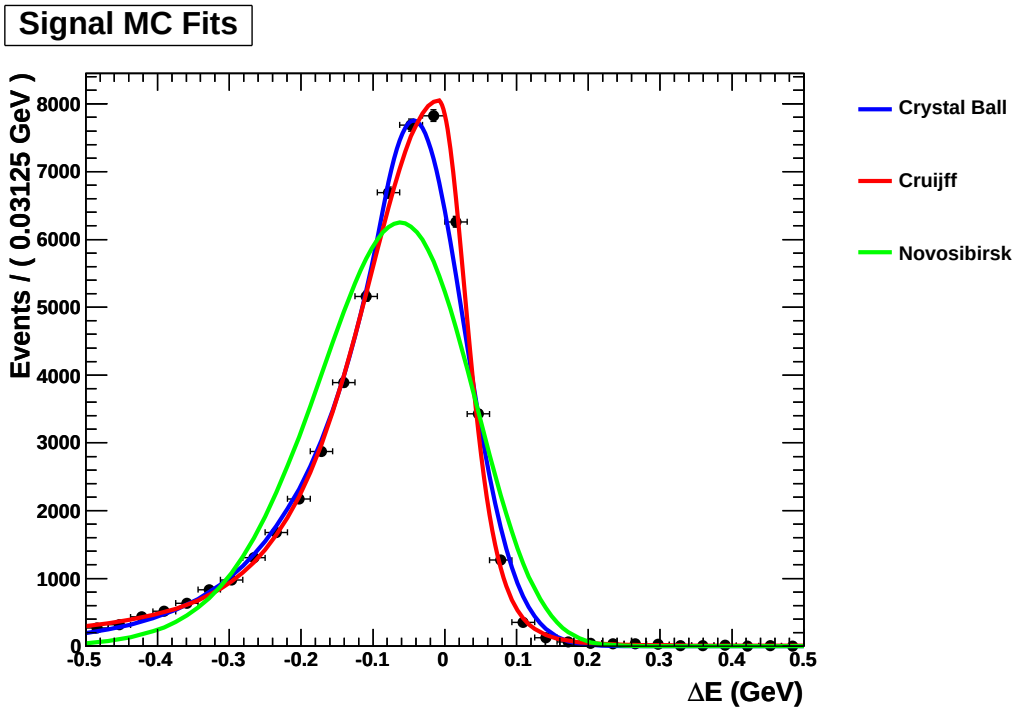


Figure 6.5: The ΔE distribution in signal MC fit with a Crystal Ball, Cruijff and Novosibirsk shape. The Cruijff is found to give the best fit.

the $B \to \pi^0\ \pi^0$ shape shown in Figure 6.8. The difference in the signal yields for a peaking background estimate of 3.13 events and $3.13 \pm 1\sigma$ is shown in Figure 6.9. The mean value of 0.46 events events is taken as the basis for the systematic error estimate.

The total fit systematic error is based on three components. From varying the fixed parameters in the signal pdf an error of 0.27 events is found, 0.15 events from the choice of a Cruijff shape for signal ΔE, and 0.46 events from the peaking background yield and pdf shape. Theses errors are added in quadrature to obtain the total fit systematic error of 0.55 events.

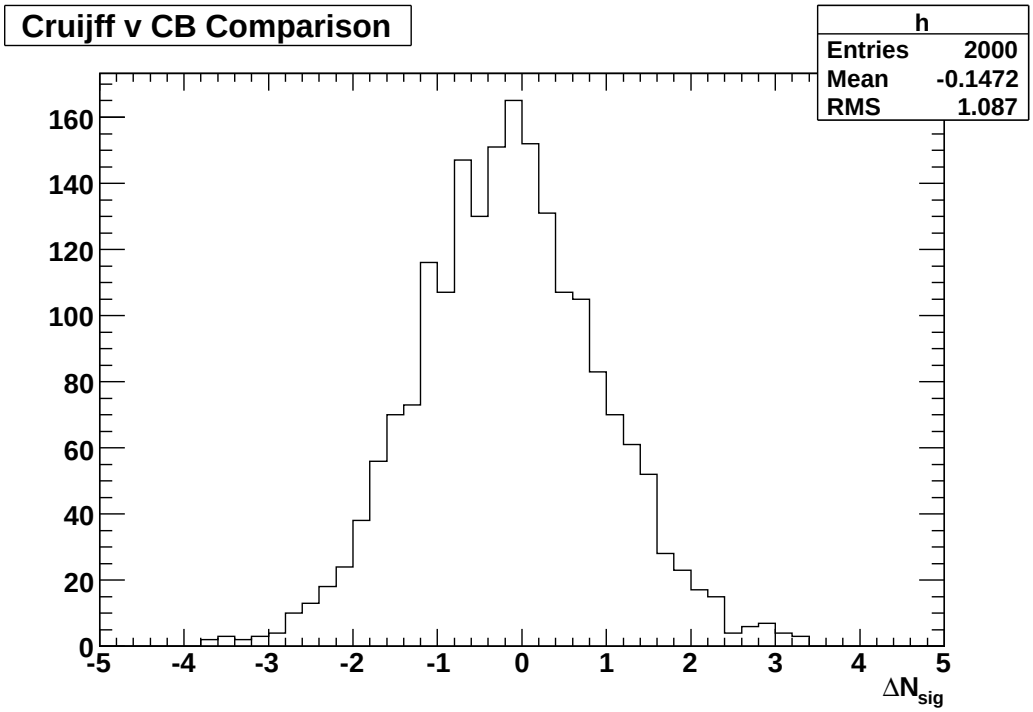


Figure 6.6: Difference in signal yield when the signal ΔE distribution is fit with a Cruijff or Crystal Ball shape to the same dataset.

The toy results have the benefit of averaging over many toys, while the data is a single experiment that may fluctuate. The method used to calculate the fit systematic is given by the formula

$$\text{Fit Syst. Error} = \sqrt{0.27^2 + 0.15^2 + 0.46^2} = 0.55 \text{ events.} \tag{6.1}$$

6.3 Upper Limit with Systematic Errors

To incorporate the systematic error into the upper limit, the likelihood function from the EML fit, shown in Figure 6.3, is convolved with a Gaussian shape whose width is equal to the total systematic error. The systematic error

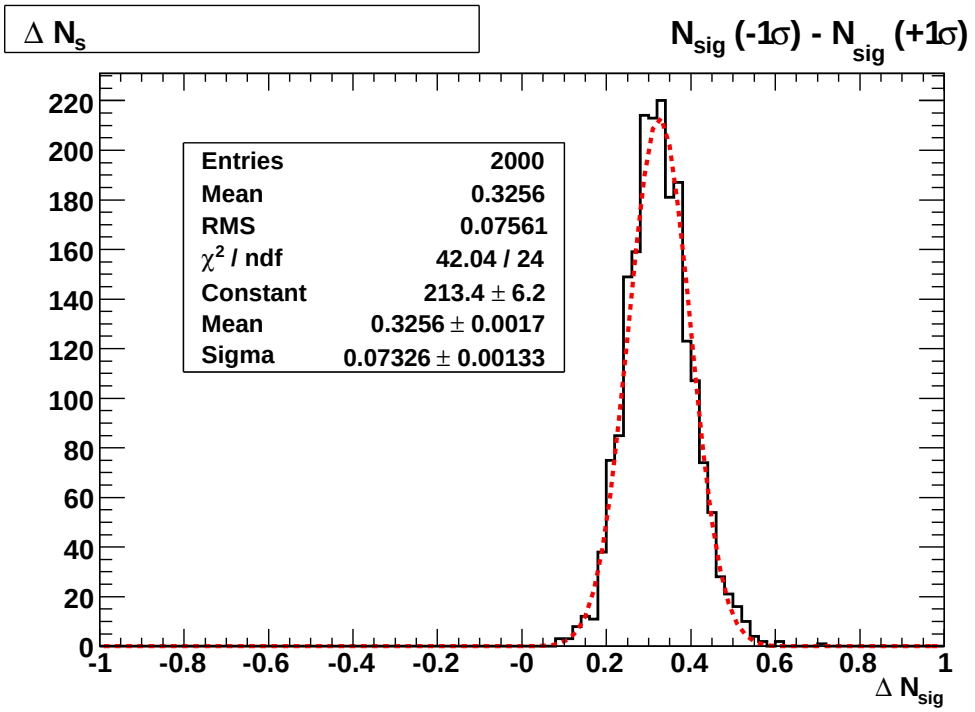
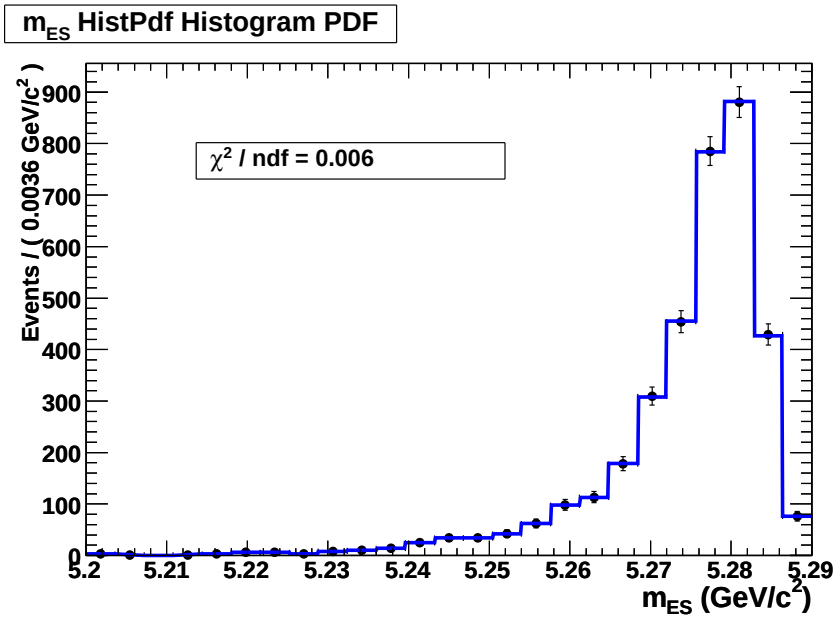
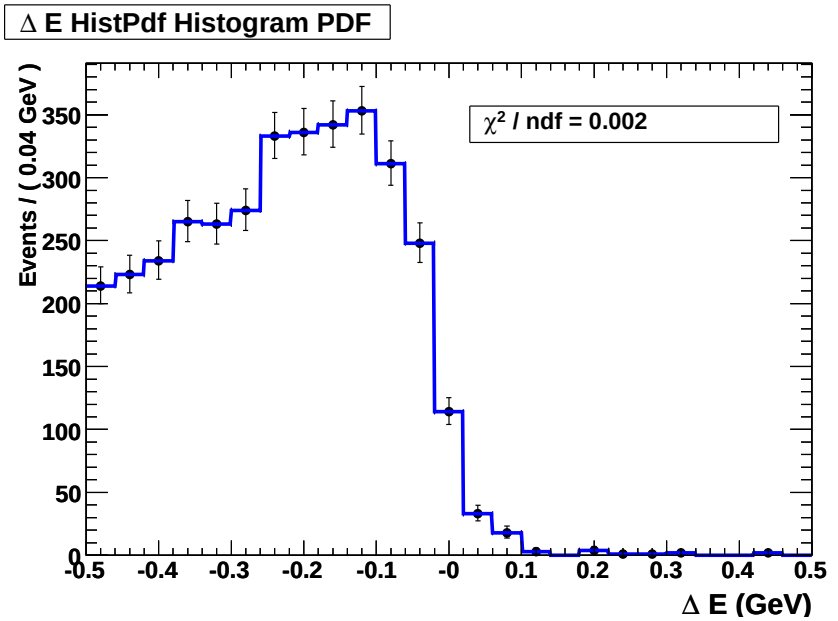


Figure 6.7: Difference in the signal yield when fitting the same toy dataset and the peaking yield is fixed to $\pm 1\,\sigma$ from its mean value of 3.13 events. The red dashed line is a Gaussian fit to the distribution with the fit parameters listed in the legend.

from all multiplicative sources is 5.3% which corresponds to 1.13 events. The additive systematic error, found by varying the fixed parameters of the fit, is 0.55 events. These are added in quadrature to give a total systematic error of 1.26 events. The likelihood curve from the fit to on-resonance data is smeared by convolving it with a Gaussian of width 1.26. Figure 6.10 shows the result of the likelihood convolution. The smeared likelihood curve essentially rests on top of the original curve showing the the systematic errors are small compared to the statistical error. The upper limit including systematic errors is obtained



(a) $B \to \pi^0\pi^0$ $m_{\rm ES}$ shape



(b) $B \to \pi^0\pi^0$ ΔE shape

Figure 6.8: ΔE and $m_{\rm ES}$ histogram PDFs used as the peaking background shape. The shapes are from the $B \to \pi^0\ \pi^0$ exclusive MC mode.

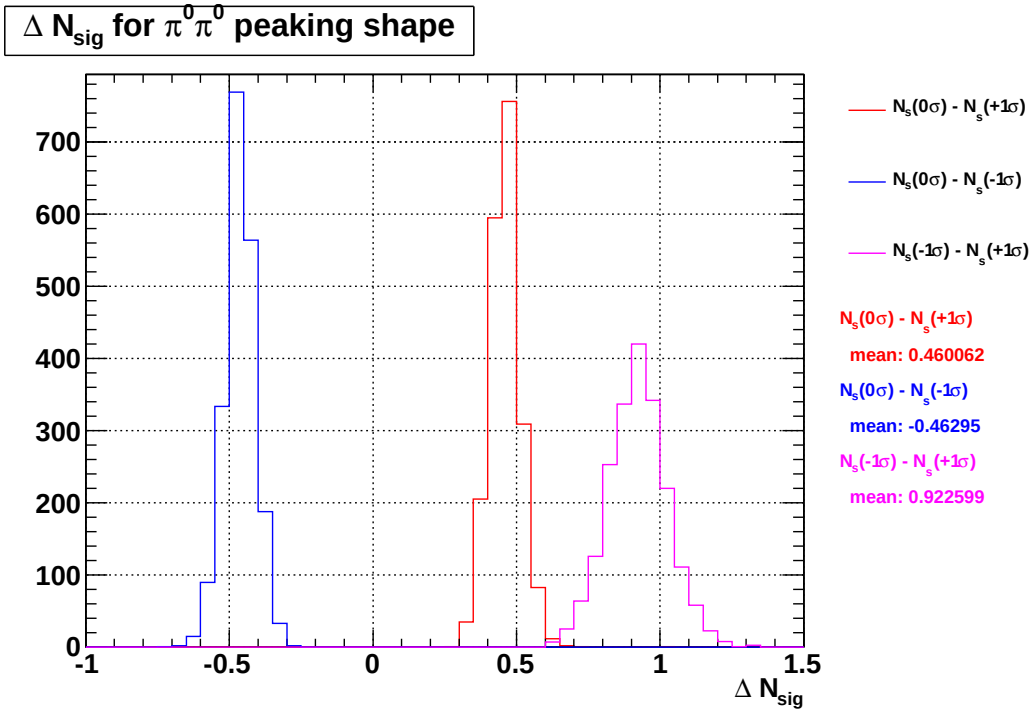


Figure 6.9: Difference in the signal yield when the expected peaking background yield is varied by $\pm 1\sigma$ from its mean value of 3.13 and the $m_{\rm ES}$ and ΔE PDF shapes are modeled by the $B \to \pi^0\ \pi^0$ shape of Figure 6.8.

by integrating the smeared likelihood curve up to 90%. The value of the signal yield where this occurs is 39.19 events and using equation (5.4) the branching fraction upper limit is calculated to be

$$\boxed{\mathcal{B}(B \to \gamma\gamma) < 3.2 \times 10^{-7} \quad (@\ 90\%\ \mathrm{CL}).} \tag{6.2}$$

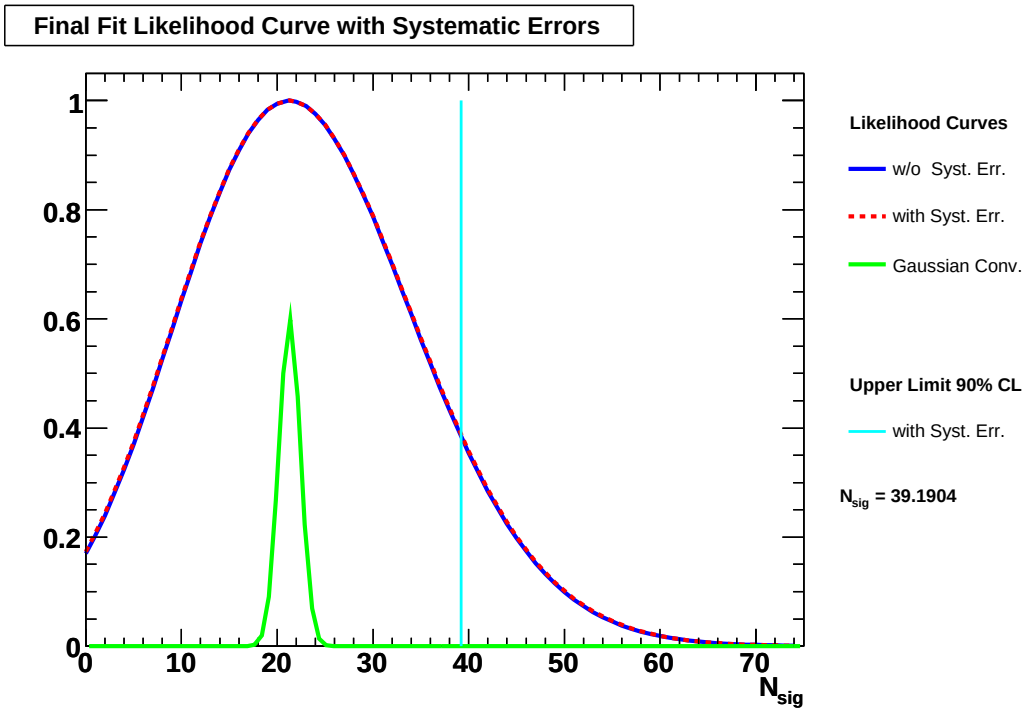


Figure 6.10: Likelihood curve before (blue) and after (red dashed) smearing by convolution with a Gaussian whose width is equal to the total systematic error (green), 1.25 events. The upper limit at the 90% CL is shown as the vertical cyan line and is equal to $N_{sig} = 39.19$ events.

6.4 Conclusions

This thesis discussed a search for the rare effective flavor changing neutral current decay $B \to \gamma\gamma$. With the dataset recorded by the *BABAR* detector, a signal yield of $21.3^{+12.8}_{-11.8}$ events with a significance of $1.88\,\sigma$ was extracted from the on-resonance data. An upper limit on the branching fraction is set at the 90% confidence level to be less than 3.2×10^{-7}. Even with the limited statistics dataset this limit is about twice as stringent as the previous best limit of less than 6.2×10^{-7} published by the Belle Collaboration. Unfortu-

nately this result is unable to definitively say anything about potential new physics beyond the Standard Model.

This analysis was performed on a data sample of integrated luminosity equal to 425.7 fb^{-1}, which corresponds to the entire $\Upsilon(4S)$ dataset recorded by *BABAR* during its lifetime, and is dominated by statistical uncertainties. As such, further improvements on the measurement of the branching fraction are possible by analyzing the Belle Collaboration data of $B\overline{B}$ events which is currently slightly less than twice as large as *BABAR* at about 750 fb^{-1}. In order to gain significant reduction on the statistical error a new generation of B Factories is required with the ability to record much large datasets. Currently there are two proposals with this goal in the planning stages. The first is an upgrade to the existing KEKB machine in Japan which would increase the instantaneous luminosity to around 10^{36} cm^{-2} s^{-1} [37]. The other is new B factory, called SuperB, which would reuse some of the current *BABAR* detector hardware while utilizing a new accelerator design to achieve a similar luminosity [38]. With these high luminosity machines a dataset with an integrated luminosity of tens of ab^{-1} is foreseen after some years of operation, which would lead to a reduction in experimental uncertainties by at least an order of magnitude from current results. Not only could this lead to a measurement of the branching fraction for $B \to \gamma\gamma$, but also other rare decays sensitive to new physics would be possible to observe.

Appendices

Appendix A

Photon Conversions

When a photon travels through material, there is a non-zero probability it will interact with a nucleus and undergo a conversion to an electron-positron pair through the reaction $\gamma + N \rightarrow e^+e^- + N$. The measure of this probability is called the radiation length, X_0, usually expressed in units of g cm^{-2}. For a high energy photon it is defined as 7/9 of the mean free path for pair production.

The *BABAR* detector is a massive object and consequently for any particles produced near the IP to reach the calorimeter it must pass through some amount of material. The amount of material the particle must traverse is dependent on the polar angle of its trajectory. Figure 4.2 shows the amount of material, in radiation lengths, in front of each sub-detector. Because this is a nonzero amount of material, some photons will undergo a conversion to an e^+e^- pair. For this analysis, this means that a signal event could go undetected affecting our efficiency and ability to make the best branching fraction measurement possible. These types of events can be recovered though, since the e^+e^- pair is detected by the tracking system. By placing basic kinematic constraints on the two tracks, such as invariant mass and distance of closest approach, a list of converted photon candidates can be built. For this reason the photon candidates used to reconstruct the B candidates in this analysis are drawn from two orthogonal lists. The `GoodPhotonLoose` candidates are measured as clusters in the EMC. The `gammaConversionDefault` photon can-

didates are made by pairwise combination of tracks. Since each B candidate is made from two photon candidates and there are two categories of photon candidates, the B candidates can be classified into four categories

Category 1 Both photon candidate daughters are GPL photon candidates.

Category 2 The higher lab energy photon candidate is a GPL photon and the lower lab energy photon candidate is a conversion.

Category 3 The lower lab energy photon candidate is a GPL photon and the higher lab energy photon candidate is a conversion.

Category 4 Both photon candidate daughters are converted photon candidates.

Only categories 1, 2 and 3 are relevant since the probability that both photons undergo a conversion is very small (about 0.2% from signal MC). The distribution of event categories in signal MC shows that more than 99.9% of events are in these three categories. Table 4.2 shows the content of event categories in signal MC and Figure A.1 shows the distribution of event categories for signal, $c\bar{c}$, $\tau^+\tau^-$, uds, B^+B^-, and $B^0\overline{B}^0$ MC samples.

The event category distributions between different MC samples in Figure A.1 are very different in shape. This is not intuitive since the process that produces the photon should not influence the way that it moves through the detector material. Therefore, the source of the photon should not effect whether it undergoes a conversion. In particular the distributions in $c\bar{c}$, and generic $B\overline{B}$ MC show a much larger fraction of events in categories two, three, and four. This reason for this is due to the way the converted photon list is constructed. The `GCD` list is made by a pairwise combination of tracks from

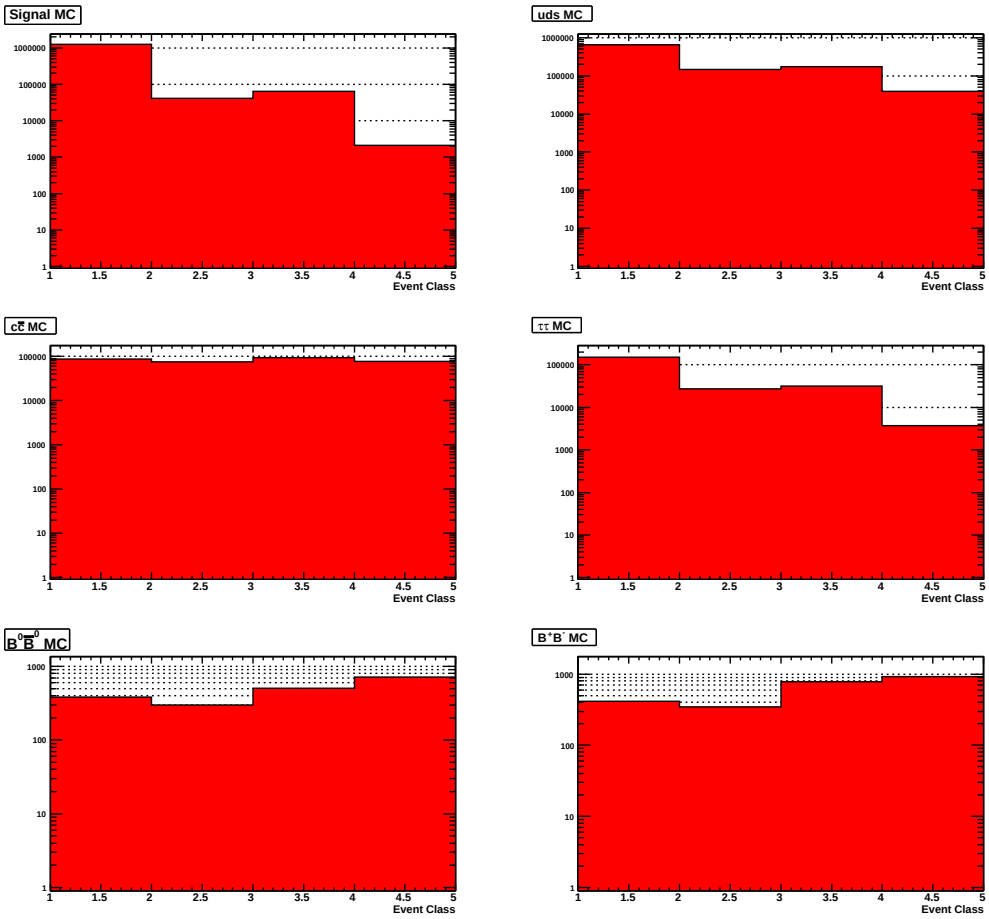


Figure A.1: Distribution of event categories in signal, $c\bar{c}$, $\tau^+\tau^-$, uds, B^+B^-, and $B^0\overline{B}^0$ MC. Bins 1, 2, 3, and 4 correspond to event categories 1, 2, 3 and 4, respectively.

Table A.1: Mean number of `ChargedTracks` per event in different MC samples.

MC sample	Track Mult.
signal	5.6
uds	5.4
$c\overline{c}$	7.2
$\tau^+\tau^-$	3.2
$B\overline{B}$	8.8

the `ChargedTracks` list. Each track is assigned the electron mass hypothesis, its 3-momentum recalculated, and the tracks are combined. If the invariant mass of the pair is less than 30 MeV/c^2, and the tracks satisfy some geometrical constraints on their distance of closest approach, the resulting candidate is placed on the `GCD` list. What this shows is that events with more tracks closer together have a larger chance of producing false conversion candidates. Table A.1 shows the multiplicities of tracks/event for the different MC samples. The $c\overline{c}$ and $B\overline{B}$ samples have a higher track multiplicity and therefore have a higher portion of events in the conversion categories.

We can get a better handle on exactly what types of tracks the photon conversions are being reconstructed from by looking at the MC Truth information. The production of the MC datasets proceeds in two stages, described in Section 3.3. In the first stage, when EvtGen produces the physics processes, a list of the particles produced by the simulator are persisted with the event in the datastore. The list of particles is accessible at later times to be able to compare what was generated with respect to what was reconstructed. What this means is that the identities of the particles used to create the photon conversion candidate can be recovered and tested to see whether they were truly electrons and positrons or just a random track combination that passed

the cuts. What is actually stored is a quantity called the Lund ID. It is a unique identifier label for each particle, e.g, electrons have a Lund ID of 11 and positrons have a Lund ID of -11. When the reconstructed charged tracks used to create the converted photon candidates are matched to their generated partners we see that in non-signal processes the majority of particles used are pions. Figure A.2 shows the Lund IDs of the particles that were used to create the conversions.

To clean up the list and select only true photon conversions, particle identification (PID) is applied to the daughter particles of the conversion. All charged tracks in *BABAR* are run through PID algorithms during reconstruction and a bit mask integer of which selectors the track passed is persisted in the datastore with the charged track. Since photons that undergo a conversion produce and $e^{\pm}$ pair, conversion candidates whose daughter constituents pass an electron selector are kept as true conversions. The algorithm chosen to identify real conversions is called *eLHSelector*. This algorithm uses information from the DCH, DRC, and EMC to decide whether the track and cluster have the characteristics associated with an electron. Both daughter tracks are required to pass this selector which gives a signal efficiency of 69.9%. The distribution of event categories after the selector is applied is shown in Figure A.3. The event categories for all MC samples is now very similar to signal. For the daughter charged tracks that pass the PID selector, the Lund ID of those remaining shows that they are almost all $e^{\pm}$, Figure A.4. The pollution from pions and Ks is consistent with the efficiency of this selector.

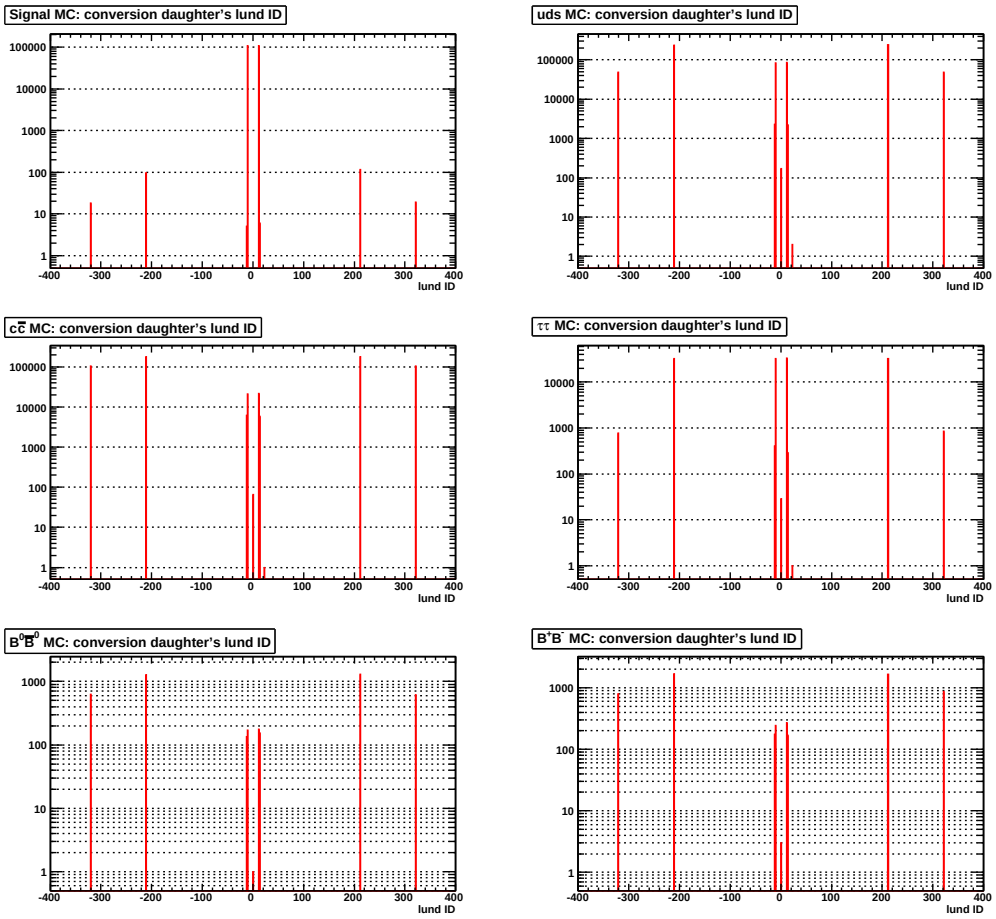


Figure A.2: Lund ID of the generated particle used to create a photon conversion candidate. Electrons and Positrons have a Lund ID of 11 and -11, respectively, $\pi^{\pm}$ Lund IDs are ± 221, and $K^{\pm}$ Lund ID are ± 321.

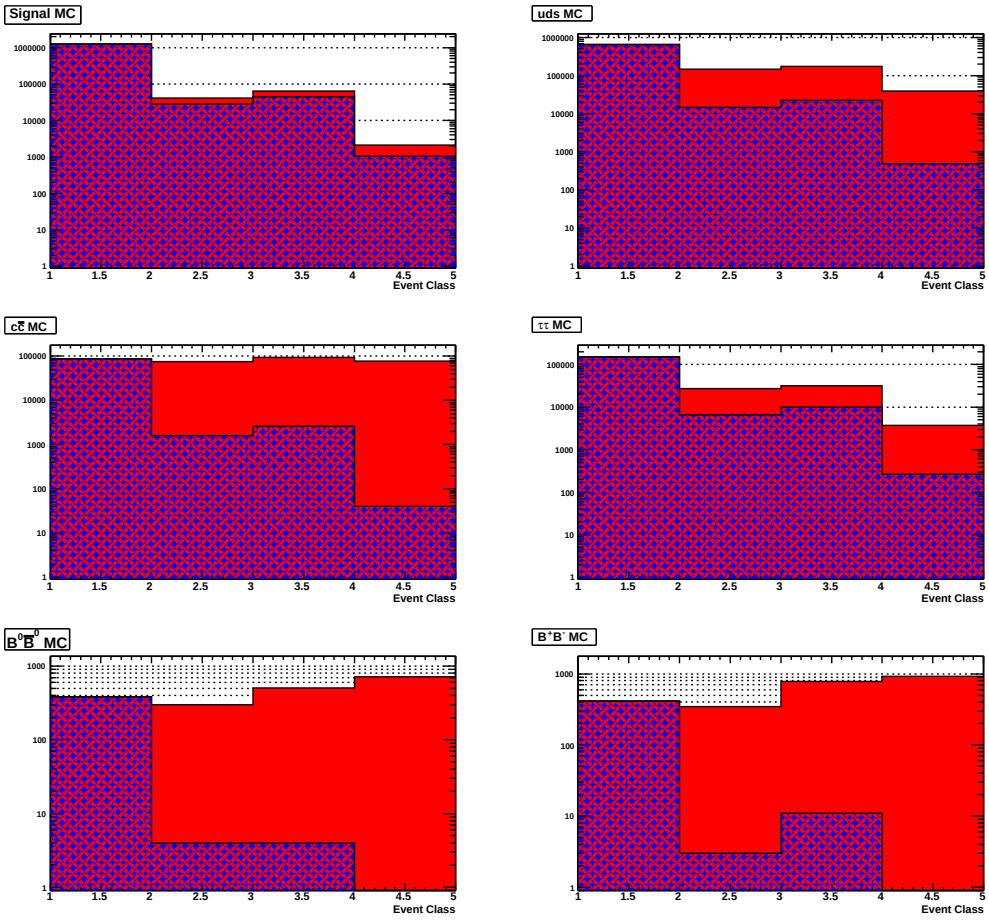


Figure A.3: Distribution of event categories in signal, $c\bar{c}$, $\tau^+\tau^-$, uds, B^+B^- and $B^0\overline{B}^0$ MC after the electron PID algorithm has been applied to the daughter tracks used to create the photon conversion candidate. Bins 1, 2, 3 and 4 correspond to event categories 1, 2, 3 and 4, respectively.

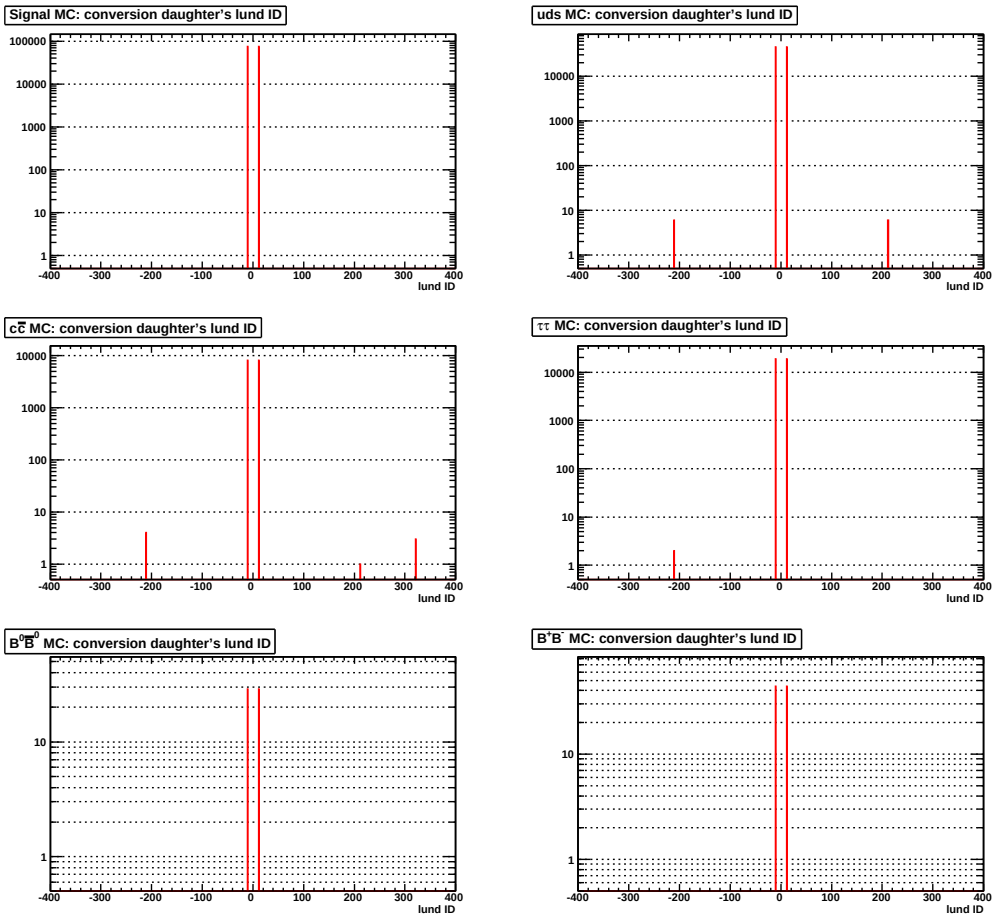


Figure A.4: Lund ID of the generated particles used to create a photon conversion candidate. The Lund ID of $e^{\pm} = \pm 11$, $\pi^{\pm} = 221$, and $K^{\pm} = \pm 321$.

A.1 Distance to Nearest Cluster

Since a photon conversion candidate is not reconstructed from an EMC cluster, the distance to the nearest charged and neutral cluster used to calculate a neural network response is not an available quantity. To be able to use the same neural network for categories two and three events as with category one events, a method to calculate this distance from a converted photon was developed. This method uses the direction of the photon momentum at its point of conversion to find where the centroid of the cluster would have occurred. This calculated cluster centroid is called the *expected cluster centroid*. The expected centroid is found by locating the point of the front face of the EMC crystal the photon would have entered had it not undergone a conversion. The x, y, and z coordinates of the centroid are then calculated by extrapolating 12.5 cm into the crystal along the line from the photon conversion to the point on the front face of the crystal. The distance is calculated from the linear distance between this point and the centroid of the nearest cluster, whether charged or neutral.

To test if this method works, it is applied to `GPL` photons with whose distance it can be compared. Figure A.5 shows a comparison between the distance calculated from the reconstructed cluster centroid of a `GPL` photon to the distance from the method outlined above to calculate the shower centroid which shows very good agreement.

Figure A.6 shows the distances calculated for the converted photon candidates overlaid on the distance histogram from Figure A.5. The shapes of the distance distributions are not the same because the events were sampled from different distributions. The probability to convert depends on the amount of material the photon traverses. Consequently, the photons that travel in the

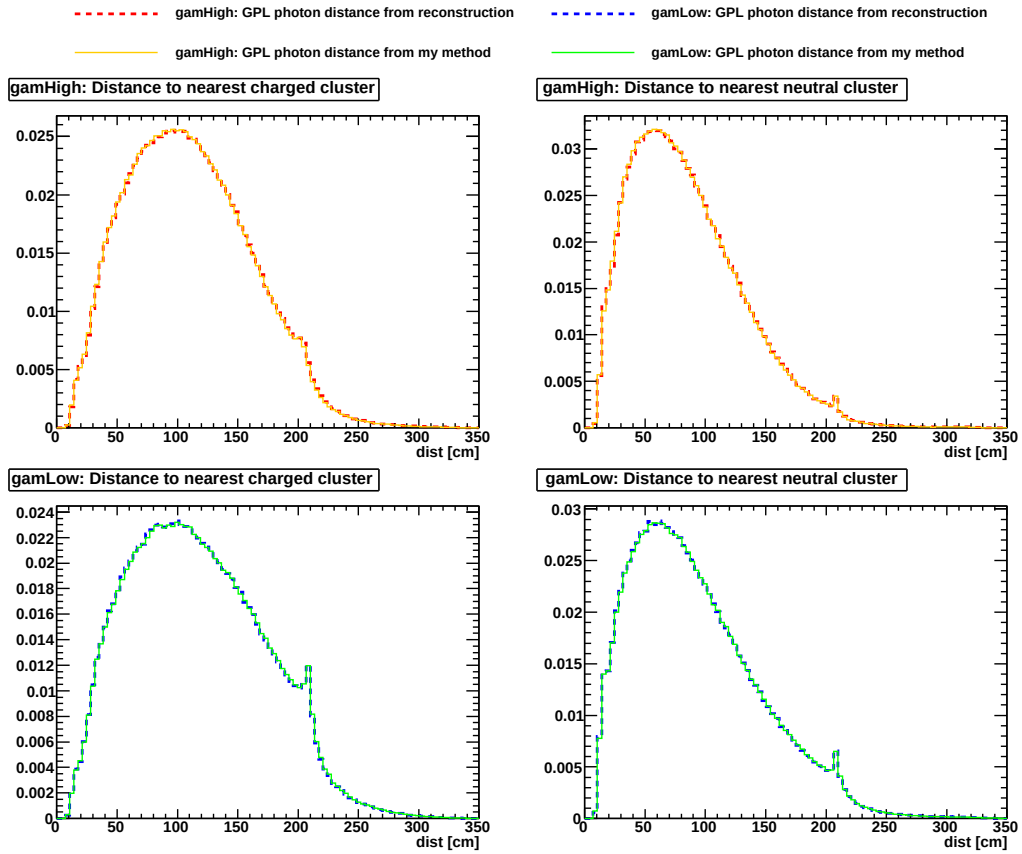


Figure A.5: Distance between the cluster centroid and the nearest charged (left) and neutral (right) clusters in the EMC for signal MC. The dashed lines show the distance using the cluster centroid calculated from the EMC cluster. The solid lines show the distance using the expected cluster centroid calculated from the direction of the photon momentum. These plots show photons from category one. The peak in the distributions at about 210 cm is consistent with the nearest EMC object on the opposite side of the calorimeter.

most forward direction, i.e., smaller polar angle, are more likely to undergo a conversion, as can be seen from Figure 4.2. When the distance distributions for converted photons were compared to the non-converted photon sample

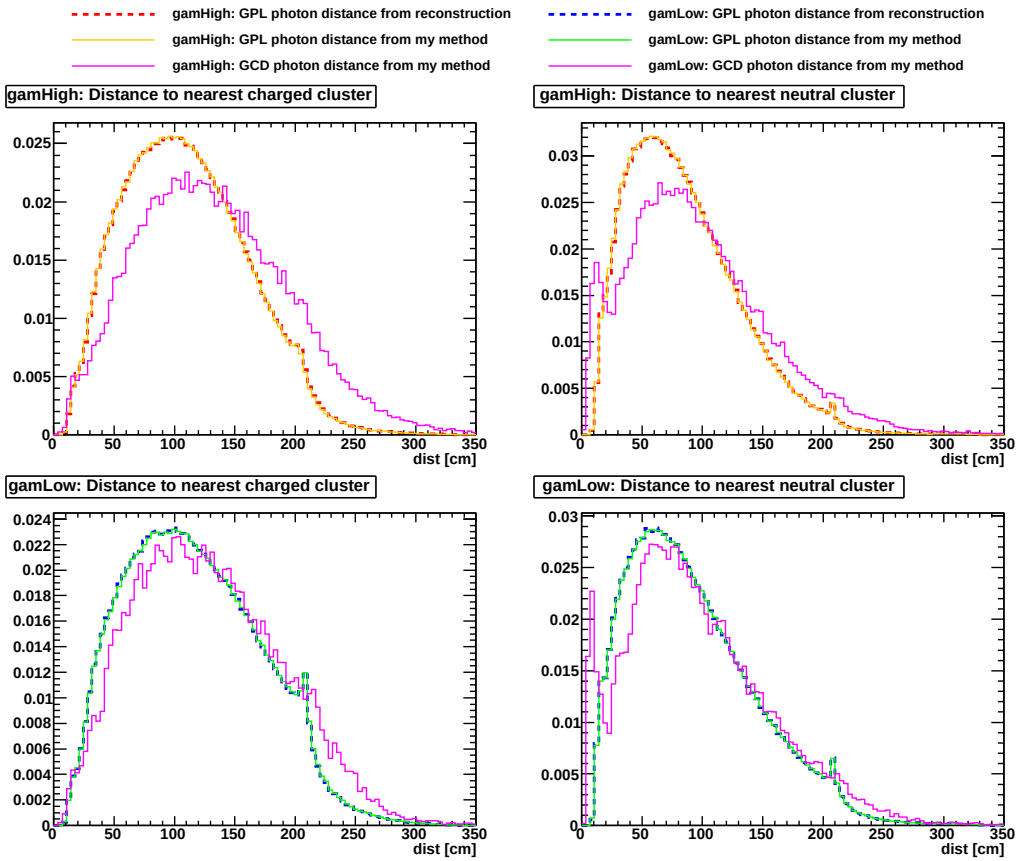


Figure A.6: Distance between cluster centroid and nearest charged (left) and neutral (right) cluster centroid in the EMC for signal MC. The magenta histogram is the distances for the converted photons using the expected cluster centroid. The peak near 10 cm in the neutral distance plot is due to the brehmsstrahlung from the daughter electrons.

in bins of polar angle, θ, and energy, the agreement was better, although not identical due to the coarse binning used. The distribution is better for the lower energy photon, gamLow, than for the higher energy photon, gamHigh, due to geometry considerations as well. The more forward photon is always the higher energy photon and will have a large probability of converting. At smaller polar

angles the sampling populations are dissimilar and the disagreement is more pronounced. With a method to calculate the distance for converted photons, a NN response can be calculated for category two and three events which allows them to be included in the likelihood fit.

A.2 Maximum Likelihood Fit with Conversions

The same PDF shapes used for category one events are also used to parameterize the category two and three $m_{\rm ES}$ and ΔE distributions. These distributions are narrower in events with a converted photon due to the better resolution of the tracking systems than the EMC. Because of the difference in the two types of event categories the $m_{\rm ES}$ and ΔE distributions have different PDF parameterizations and are shown in Figures A.7–A.10 for signal and background.

The fitting model is also changed because of this difference. The fit becomes a 2D simultaneous extended maximum likelihood fit where the data is split into two categories by whether or not the event has a conversion. When testing this scenario, a peaking component was not included in the overall PDF. A simultaneous fit means that the distributions of the different categories are fit separately, but the yields are connected. For example, the overall signal yield, N_{sig}, is related to the individual signal yields for categories one, N^1_{sig} and two plus three, N^{2+3}_{sig} by the relation $N_{sig} = N^1_{sig} + N^{2+3}_{sig}$. The relative signal yield, the fraction of events in categories two plus three to that of category one, is fixed to the value 5.1% found from signal MC. For the background yield, $N_{bkg} = N^1_{bkg} + N^{2+3}_{bkg}$, but the relative yield is not fixed. Because we have added new information to the likelihood function, eq. (5.2)

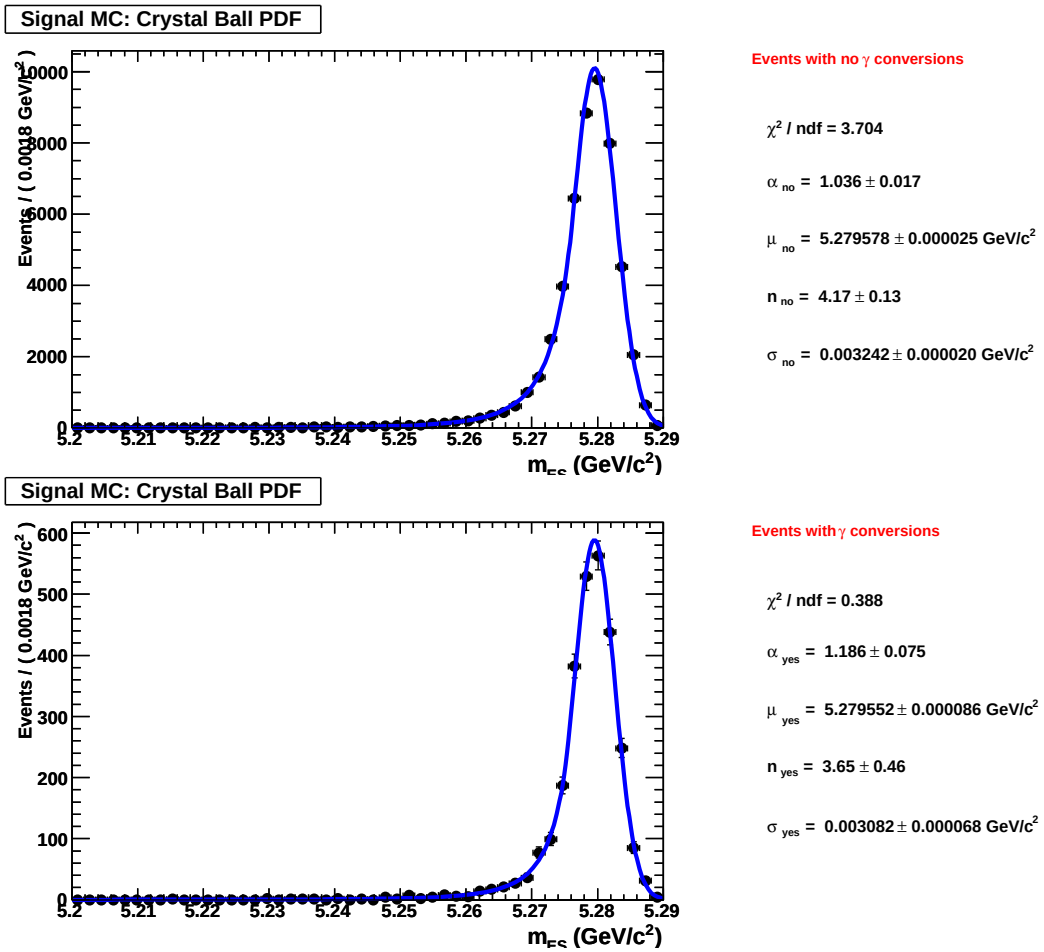


Figure A.7: Signal $m_{\rm ES}$ PDF shapes from signal MC using a crystal ball PDF. The final cuts are applied to the MC sample and 2D fit is performed to calculate the parameters.

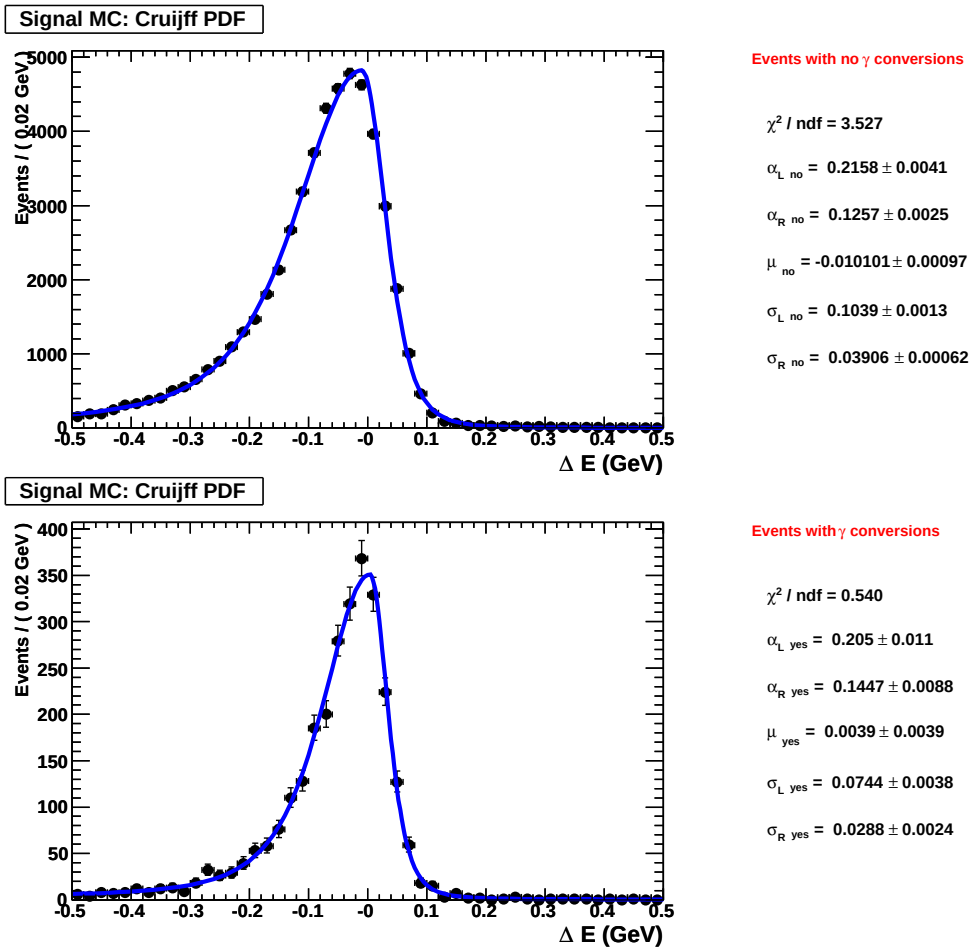


Figure A.8: Signal ΔE PDF shapes from signal MC using a Cruijff PDF. The final cuts are applied to the MC sample and 2D fit is performed to calculate the parameters.

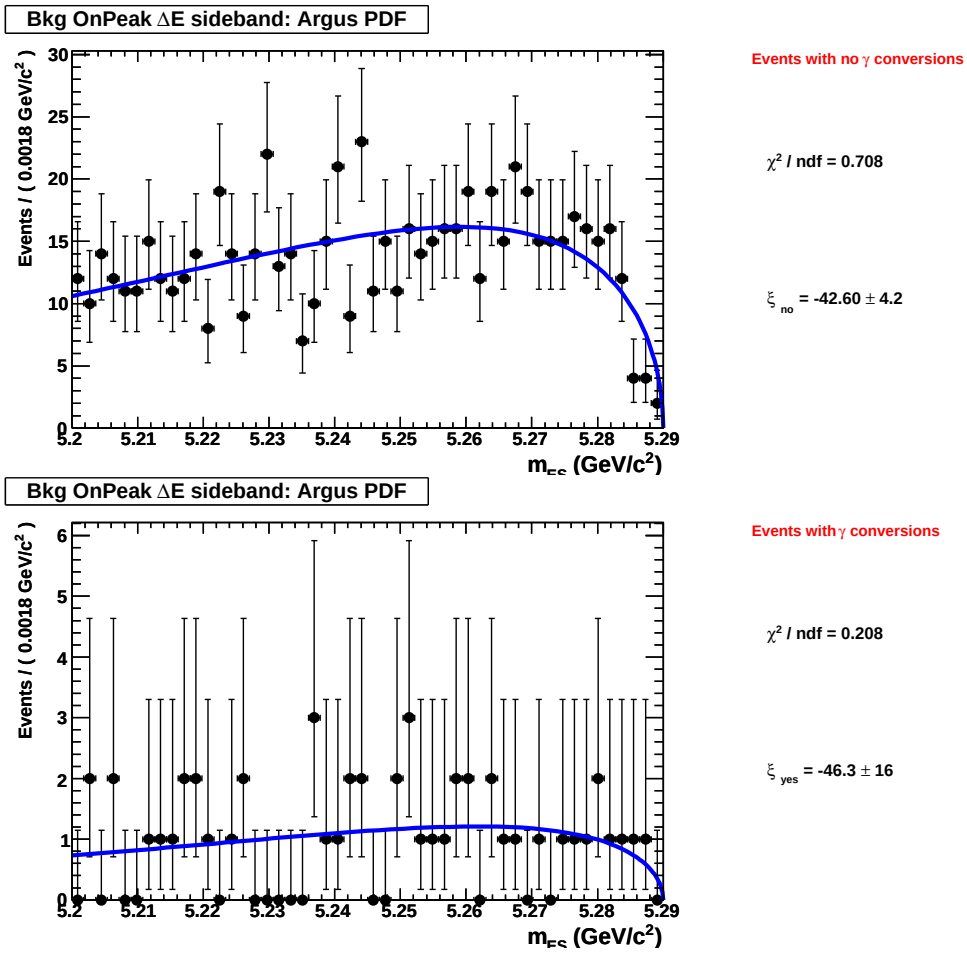


Figure A.9: Background $m_{\rm ES}$ PDFs from onPeak sideband data. The final cuts are applied and a 2D fit is performed to calculate the parameters. The endpoint is fixed to 5.279 GeV/c^2.

becomes

$$\mathcal{L} = \frac{e^{-N}}{N!} \sum_{c=1}^{2} \prod_{i=0}^{N} (N_{sig}^{c} \mathcal{P}^{c}(x_i, \vec{\alpha}_i) + N_{bkg}^{c} \mathcal{P}^{c}(x_i, \vec{\alpha}_i)) \tag{A.1}$$

where i is the index of each data point x, $\vec{\alpha}$ are the floating parameters in the fit, c is a label for the event categories which for this analysis c is equal to two. Again all signal shape parameters for the $m_{\rm ES}$ and ΔE distributions are fixed in the fit.

A.3 Toy Studies and Upper Limits

In analog to Section 5.3, 2000 pure and signal-embedded toy MC studies were run to validate the fitter and check for biases in the yields. The results of the validation studies showed no bias in the yields and that the fit was working as expected. Figure A.11 shows an example of a simultaneous fit to a signal-embedded toy MC dataset with an assumed signal branching fraction of 5×10^{-8}. For each toy MC experiment an upper limit on the branching fraction was calculated by integrating the likelihood function up to 90% of its area. It was then compared to upper limits obtained for toy studies that excluded events with conversions. The goal was to see if by including these events a better upper limit could be obtained. Figure A.12 shows a comparison between the upper limits with and without converted photon events. There is no difference in the limit obtained and for this reason these events were excluded from the main analysis.

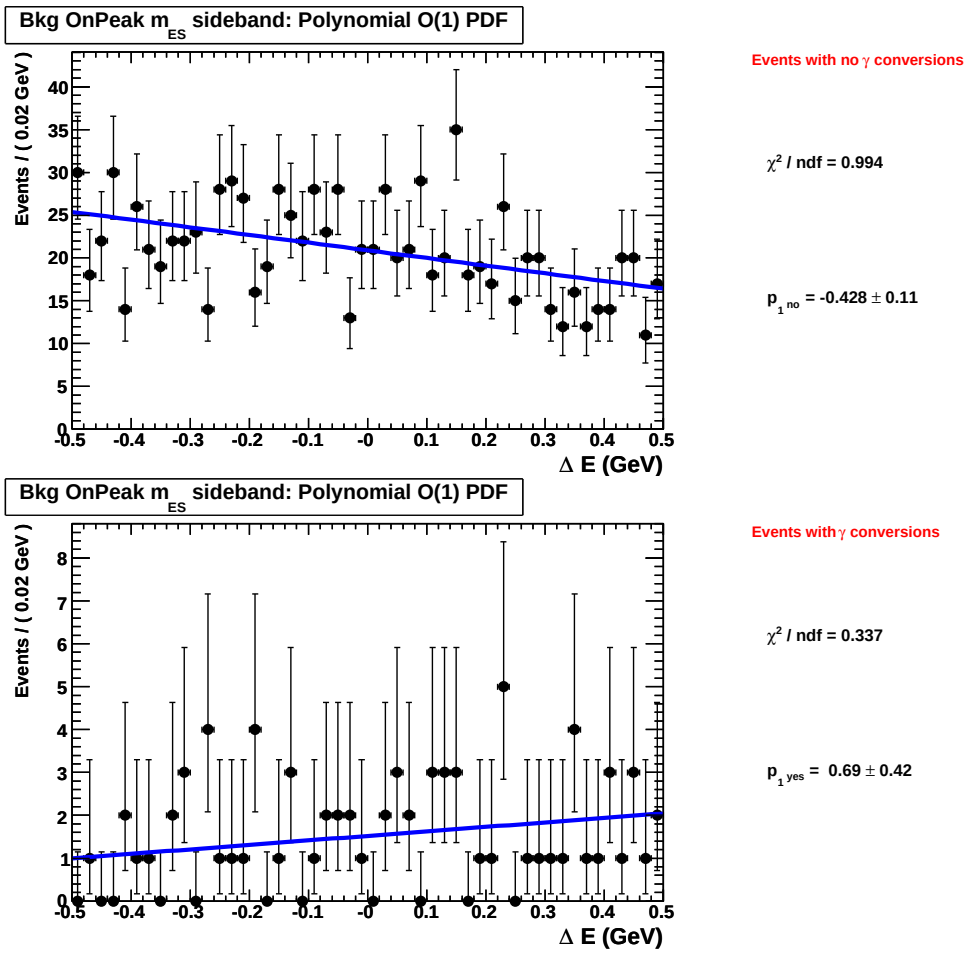


Figure A.10: Background ΔE PDFs from onPeak sideband data. The final cuts are applied and a 2D fit is performed to calculate the parameters.

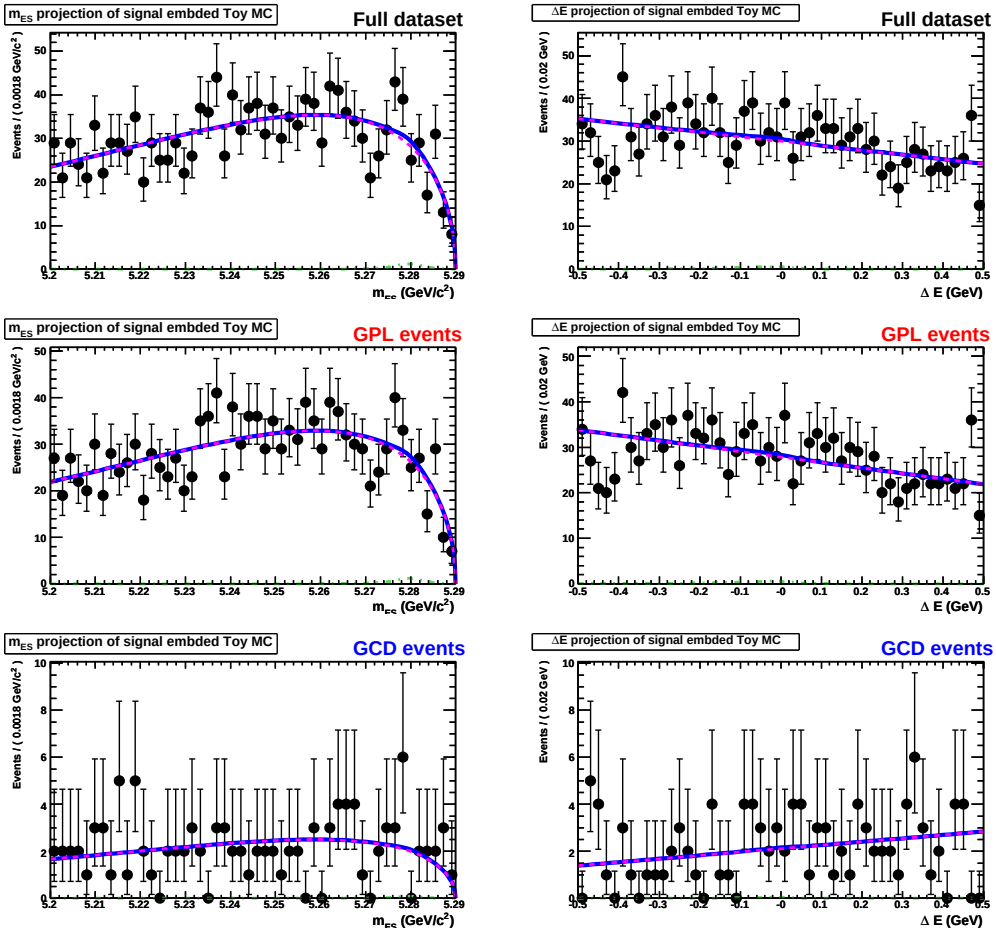


Figure A.11: Example of one of the 2000 signal embedded toy MC fit that were run assuming a signal branching fraction of 5×10^{-8}.

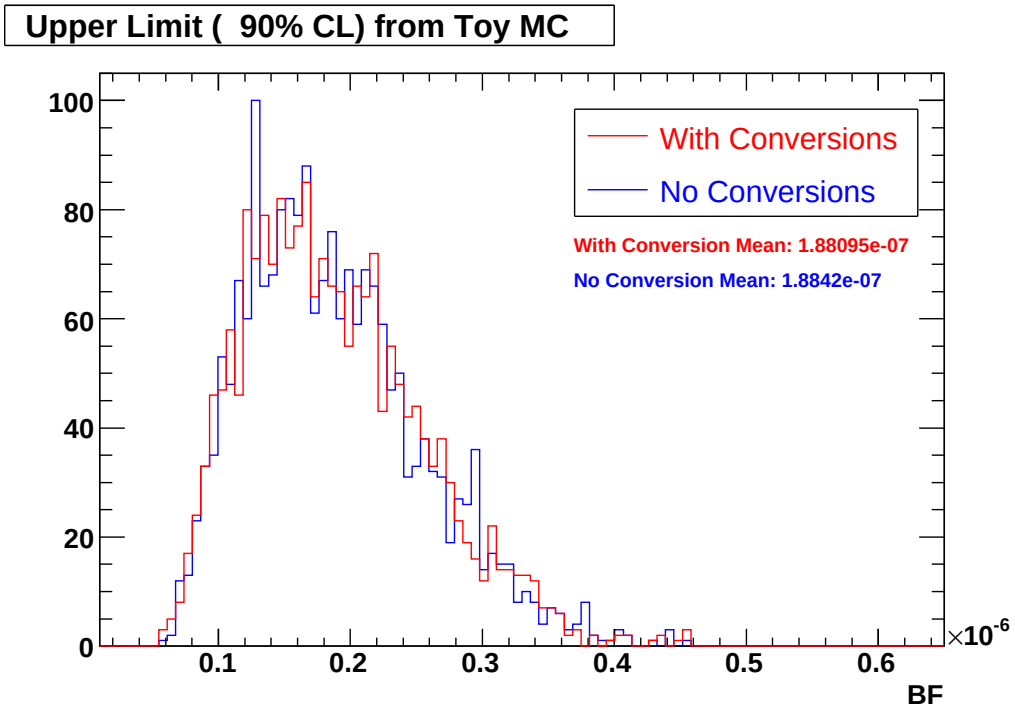


Figure A.12: Comparison of upper limits at 90% CL with and without photon conversions included in the event sample.

Appendix B

Non-optimized Variable Figure of Merit Plots

In addition to the variables whose cuts were determined by the optimization procedure described in Section 4.9, the same technique was applied to the other variables that were not optimized, e.g., photon momentum polar angle. This was done as a check that the chosen cut values were correct, and no evidence was found that those values should be changed. The following plots show the figure of merit as a function of the cut value for the non-optimized variables.

In the figures of this appendix, there are large spikes in the FOM values for certain cuts. These features correspond to failed fits from the optimization and should be ignored. If, during the optimization, a fit failed and produced an errant figure of merit value it was ignored when finding the maximum of the FOM. For these plots no time was spent to fixing the issue of failed fits, unlike the optimized variables, and so the failed fit spikes remain.

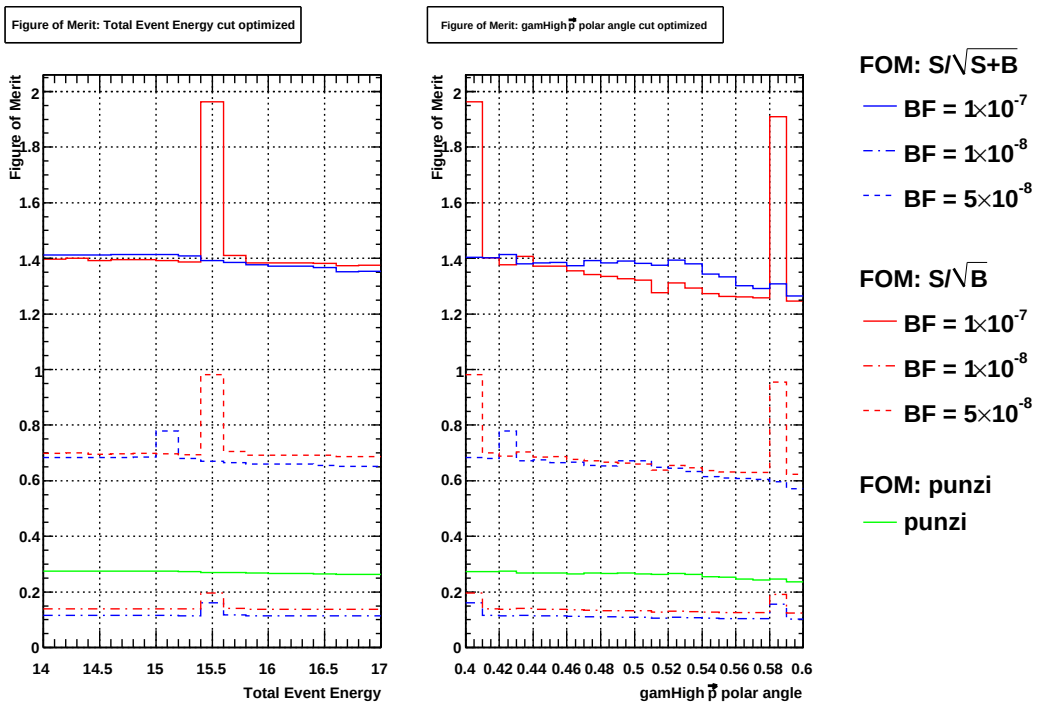


Figure B.1: Figure of Merit as a function of variable cut value. The plots are flat in this region meaning that the initial cut chosen is also the final cut. The spikes in the plots correspond to failed fits of the $m_{\rm ES}$ distribution during the optimization.

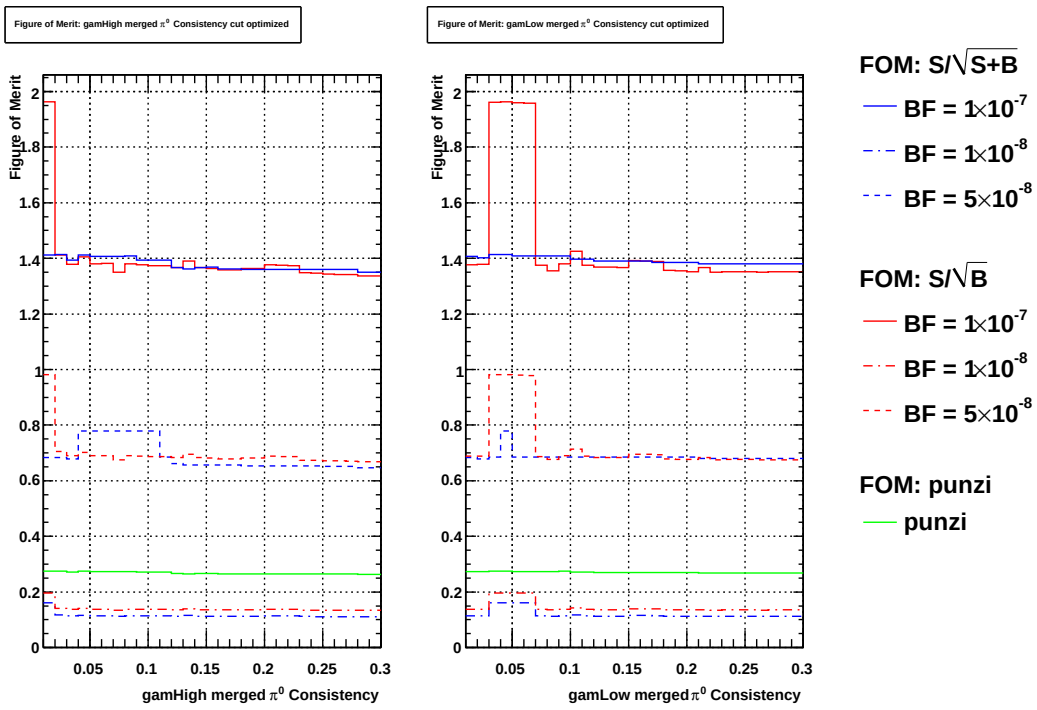


Figure B.2: Figure of Merit as a function of variable cut value. The plots are flat in this region meaning that the initial cut chosen is also the final cut. The spikes in the plots correspond to failed fits of the $m_{\rm ES}$ distribution during the optimization.

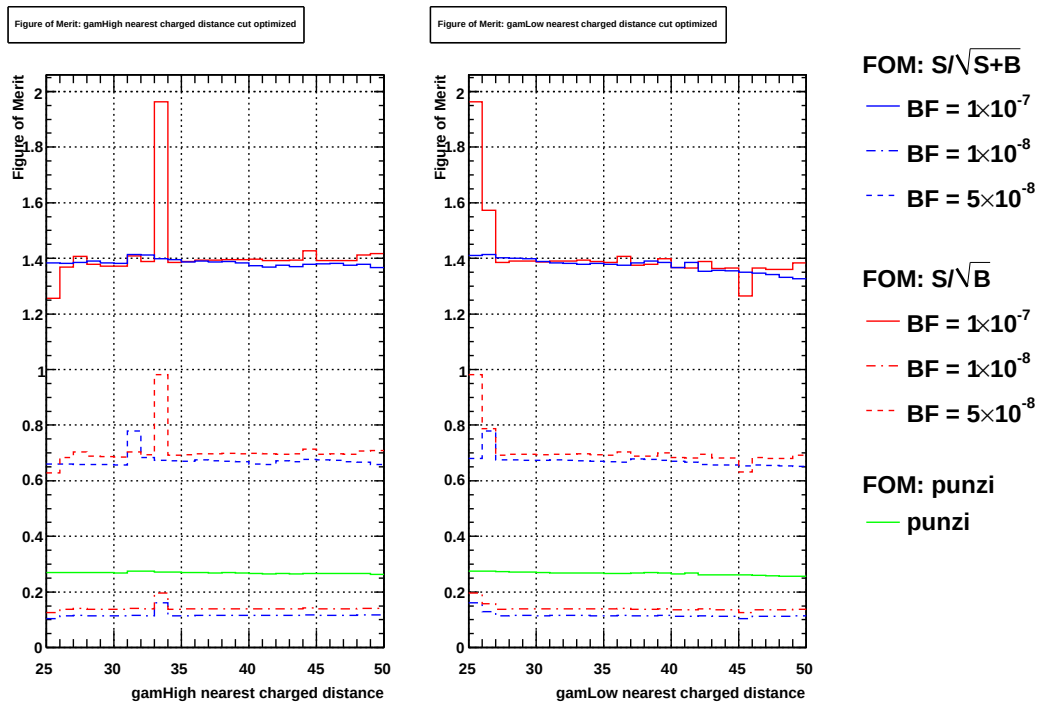


Figure B.3: Figure of Merit as a function of variable cut value. The plots are flat in this region meaning that the initial cut chosen is also the final cut. The spikes in the plots correspond to failed fits of the $m_{\rm ES}$ distribution during the optimization.

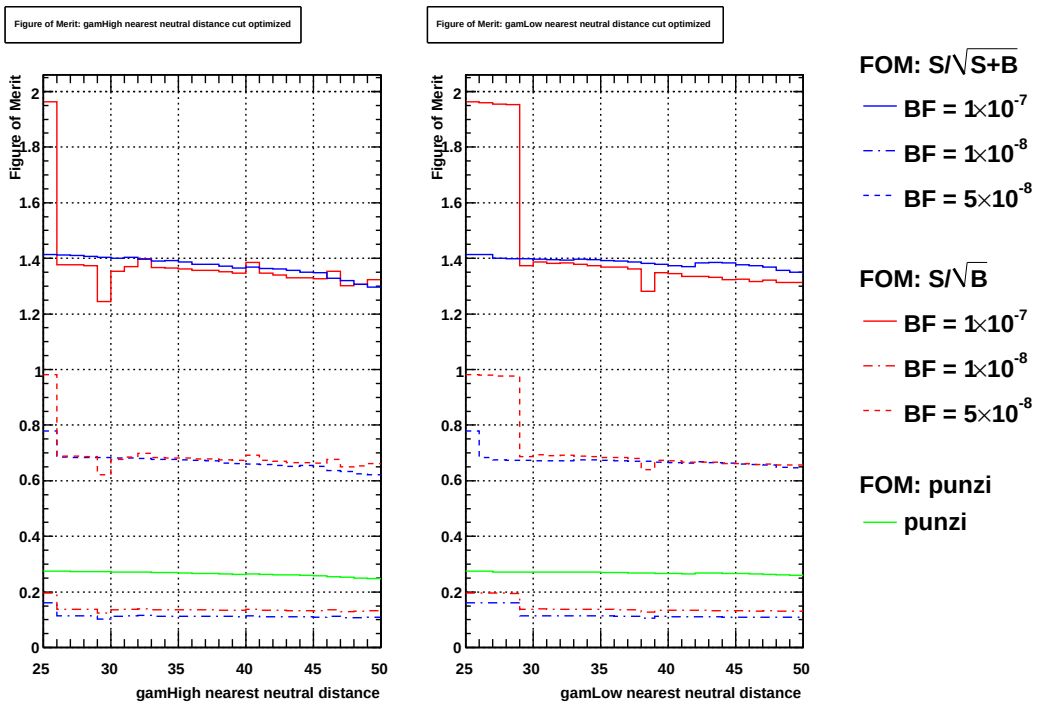


Figure B.4: Figure of Merit as a function of variable cut value. The plots are flat in this region meaning that the initial cut chosen is also the final cut. The spikes in the plots correspond to failed fits of the $m_{\rm ES}$ distribution during the optimization.

Appendix C

Final Cut Efficiencies

This appendix shows the efficiencies of the final cuts on uds, $c\bar{c}$, $\tau^+\tau^-$, and $B^0\overline{B}^0$ and B^+B^- MC both individually and cumulatively.

Table C.1: Cumulative efficiencies for uds, $c\bar{c}$, $\tau^+\tau^-$, and $B\overline{B}$ MC samples. Each row has all cuts listed above applied when calculating the efficiency. N_{gen} is the number of events generated of this type of MC process. All efficiencies are normalized to the number of events that pass the skim selections, N_{skim}.

		uds		$c\bar{c}$		$\tau^+\tau^-$		$B\overline{B}$	
	N_{gen}	938,312,000		1,132,468,000		395,206,000		1,449,710,000	
	N_{skim}	1,028,134		335,466		216,475		4,502	
Cut	**Value**	N_{pass}	ε	N_{pass}	ε	N_{pass}	ε	N_{pass}	ε
Class 1 Cuts									
# Bs / event	nB=1	1017506	0.990	332115	0.990	214464	0.991	4369	0.970
Both γs GPL	TRUE	658646	0.641	86192	0.257	151316	0.699	796	0.177
Event Level Cuts									
nGTL	≥ 3	511210	0.497	70574	0.210	40312	0.186	731	0.162
E_{tot}	< 15.0 GeV	506086	0.492	69677	0.208	40092	0.185	467	0.104
Neural Net	≥ 0.54	14396	0.014	3229	0.010	576	0.003	278	0.062
Cuts applied to both photons									
Dist to charged	> 25 cm	12503	0.012	2888	0.009	438	0.002	211	0.047
Dist to neutral	> 25 cm	9059	0.009	2268	0.007	363	0.002	153	0.034
$\vec{p}$ polar angle	$0.4 < \theta < 2.4$	8512	0.008	2108	0.006	328	0.002	140	0.031
Lateral Moment	$0.15 \leq$Lat≤ 0.5	4566	0.004	1343	0.004	224	0.001	59	0.013
cluster time	$6200 \leq t \leq 6350$ ns	4449	0.004	1290	0.004	214	0.001	57	0.013
# of Crystals	> 10	4446	0.004	1290	0.004	214	0.001	57	0.013
merged π^0 Cons.	< 0.01	3638	0.004	1129	0.003	193	0.001	51	0.011
π^0 LR	≤ 0.84	2115	0.002	664	0.002	149	0.001	12	0.003
η LR	≤ 0.84	1141	0.001	394	0.001	139	0.001	9	0.002

Table C.2: Efficiencies of the final cuts for uds, $c\bar{c}$, $\tau^+\tau^-$, and generic $B\overline{B}$ MC samples applied individually. N_{gen} is the number of events generated of this type of MC process. All efficiencies are normalized to the number of events that pass the skim selection, N_{skim}.

		uds		$c\bar{c}$		$\tau^+\tau^-$		$B\overline{B}$	
	N_{gen}	938,312,000		1,132,468,000		395,206,000		1,449,710,000	
	N_{skim}	1,028,134		335,466		216,475		4,502	
Cut	**Value**	N_{pass}	ε	N_{pass}	ε	N_{pass}	ε	N_{pass}	ε
Class 1 Cuts									
# Bs / event	nB=1	1017506	0.990	332115	0.990	214464	0.991	4369	0.970
Both γs GPL	TRUE	658646	0.641	86192	0.257	151316	0.699	796	0.177
Event Level Cuts									
nGTL	≥ 3	511210	0.776	70574	0.819	40312	0.266	731	0.918
E_{tot}	< 15.0 GeV	652494	0.991	85148	0.988	150772	0.996	525	0.660
Neural Net	≥ 0.54	11946	0.030	4361	0.051	2392	0.016	349	0.438
Cuts applied to both photons									
Dist to charged	> 25 cm	566475	0.860	76519	0.888	133122	0.880	559	0.702
Dist to neutral	> 25 cm	179281	0.272	30984	0.359	33744	0.223	322	0.405
$\vec{p}$ polar angle	$0.4 < \theta < 2.4$	592651	0.900	78413	0.910	137202	0.907	449	0.564
Lateral Moment	$0.15 \leq$Lat≤ 0.5	495610	0.752	61777	0.717	119778	0.792	269	0.338
cluster time	$6200 \leq t \leq 6350$ ns	638621	0.970	83318	0.967	145919	0.964	414	0.520
# of Crystals	> 10	657188	0.998	85738	0.995	151014	0.998	436	0.548
merged π^0 Cons.	< 0.01	224418	0.341	37989	0.441	42655	0.282	482	0.606
π^0 LR	≤ 0.84	94469	0.143	16541	0.192	14281	0.094	206	0.259
η LR	≤ 0.84	383738	0.583	51386	0.596	109283	0.722	467	0.587

Appendix D

Peaking Backgrounds

This appendix contains the $m_{\rm ES}$ and ΔE distributions for all of the exclusive MC modes that were investigated in the study of peaking backgrounds done in Section 4.11. After the final selections are applied to each MC sample the $m_{\rm ES}$ and ΔE distributions are plotted along with the distribution of event in the fit plane. The 1D plots are scaled to the expected number of events expected in the on resonance data. The black histograms represent the events in the whole fit plane while the solid red histograms correspond to the events inside the signal box which is defined as

$$\begin{aligned} 5.27 \leq \quad & m_{\rm ES} \quad \leq 5.29 \ \ {\rm GeV}/c^2 \\ -0.3 \leq \quad & \Delta E \quad \leq 0.1312 \ \ {\rm GeV} \end{aligned} \tag{D.1}$$

Figure D.1 shows the sum of all exclusive MC after being scaled to an equivalent luminosity equal to the on resonance data sample of $425.7\,{\rm fb}^{-1}$.

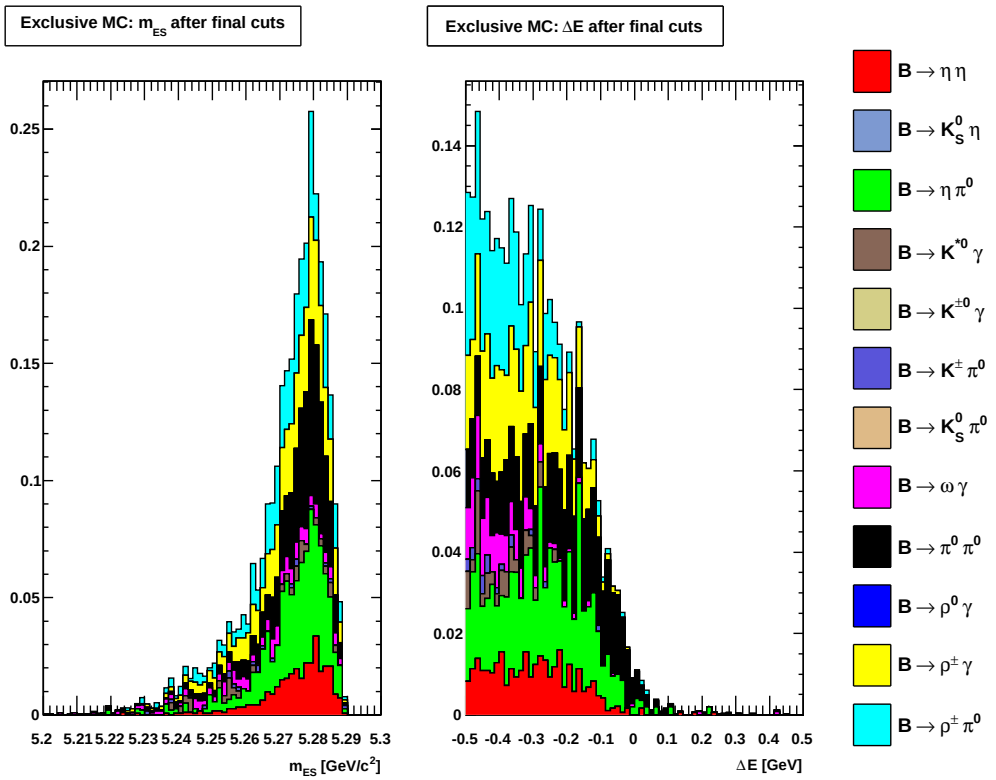


Figure D.1: $m_{\rm ES}$ and ΔE distributions of a luminosity weighted combination of modes listed in Table 4.6.

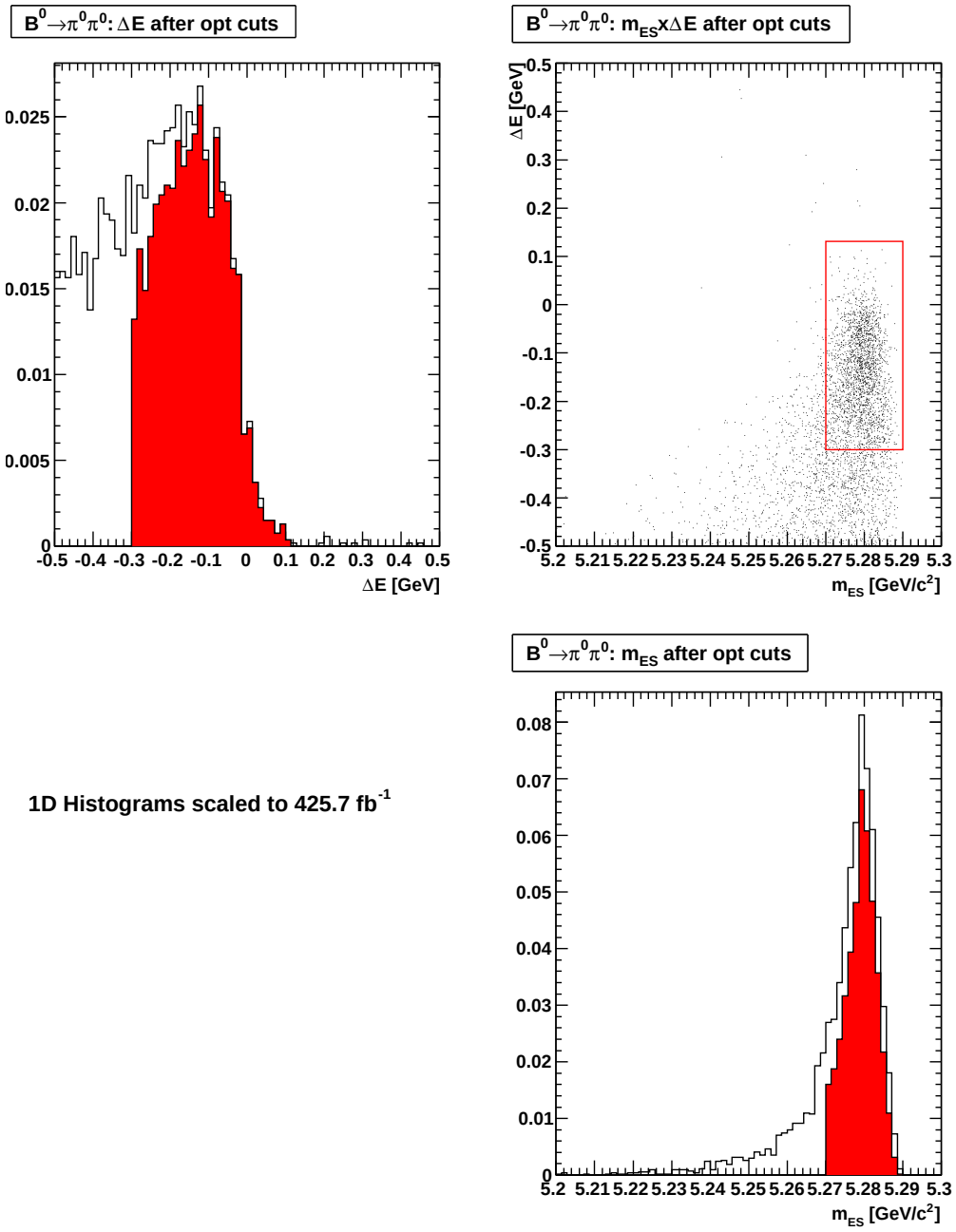


Figure D.2: Distribution of events in $B^0 \to \pi^0\ \pi^0$ MC that survive the optimized cut variables. The 1D histogram axis shows the numbers of event after being scaled to luminosity to find the expected numbers in data.

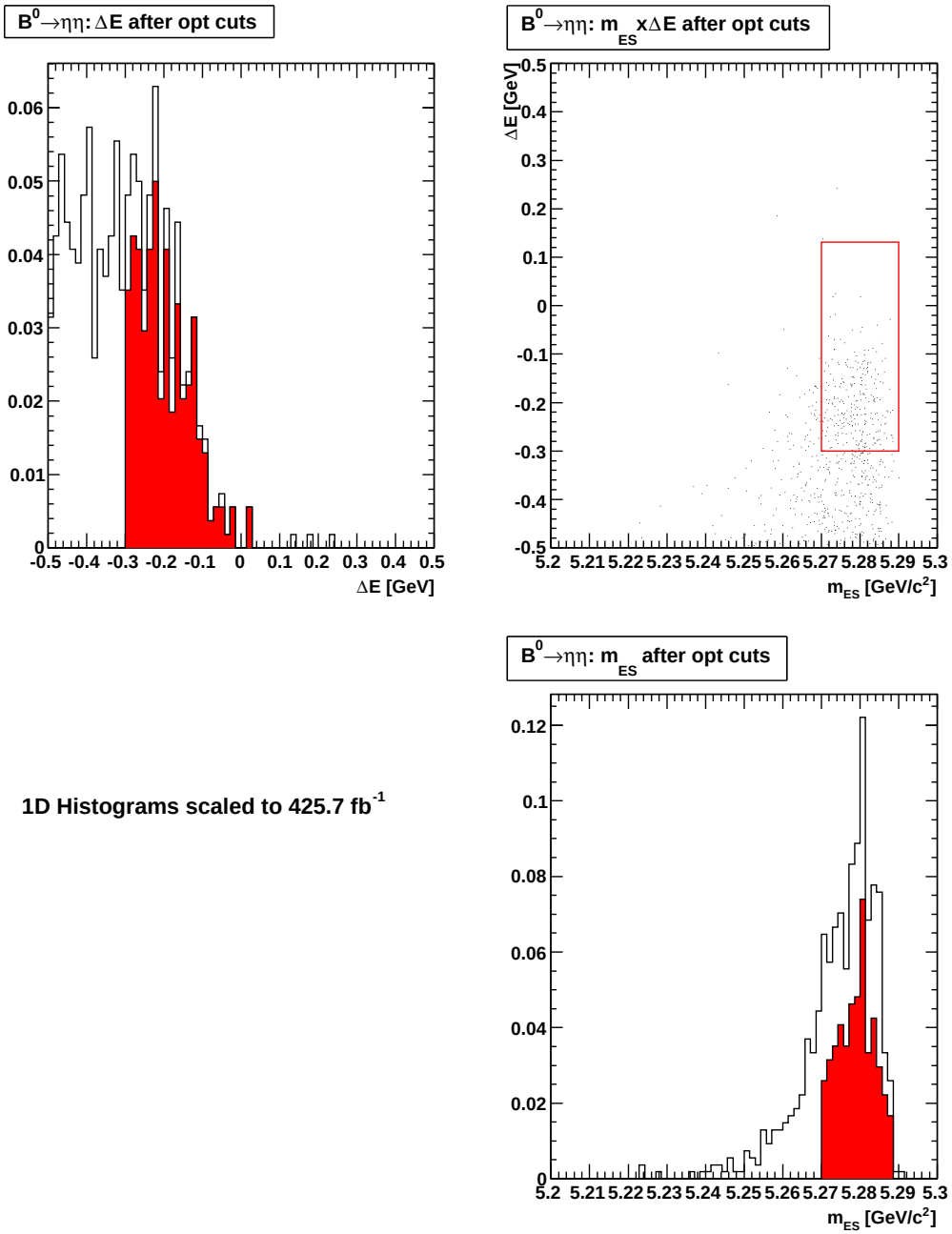


Figure D.3: Distribution of events in $B^0 \to \eta\eta$ MC that survive the optimized cut variables. The 1D histogram axis shows the numbers of event after being scaled to luminosity to find the expected numbers in data.

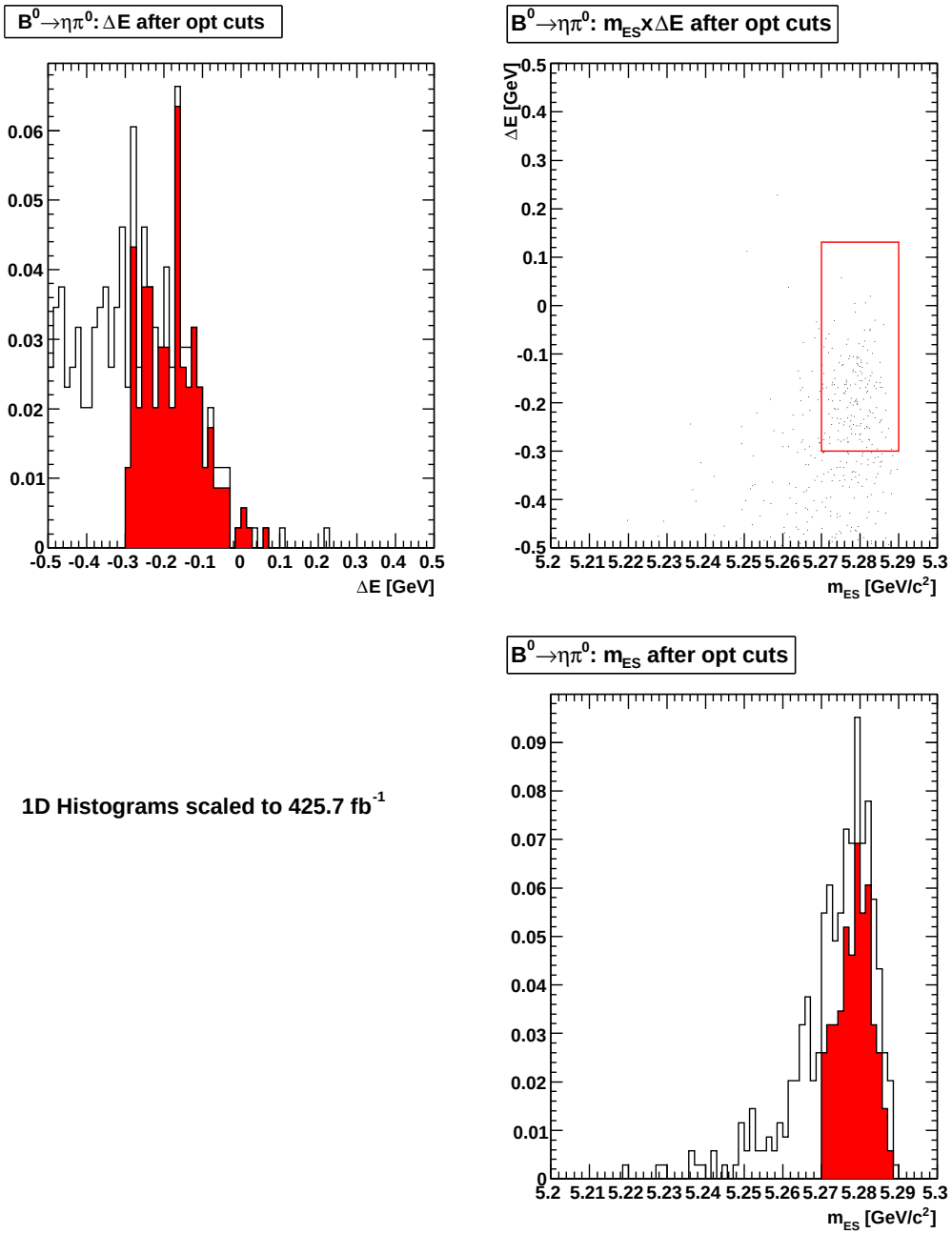


Figure D.4: Distribution of events in $B^0 \to \eta\pi^0$ MC that survive the optimized cut variables. The 1D histogram axis shows the numbers of event after being scaled to luminosity to find the expected numbers in data.

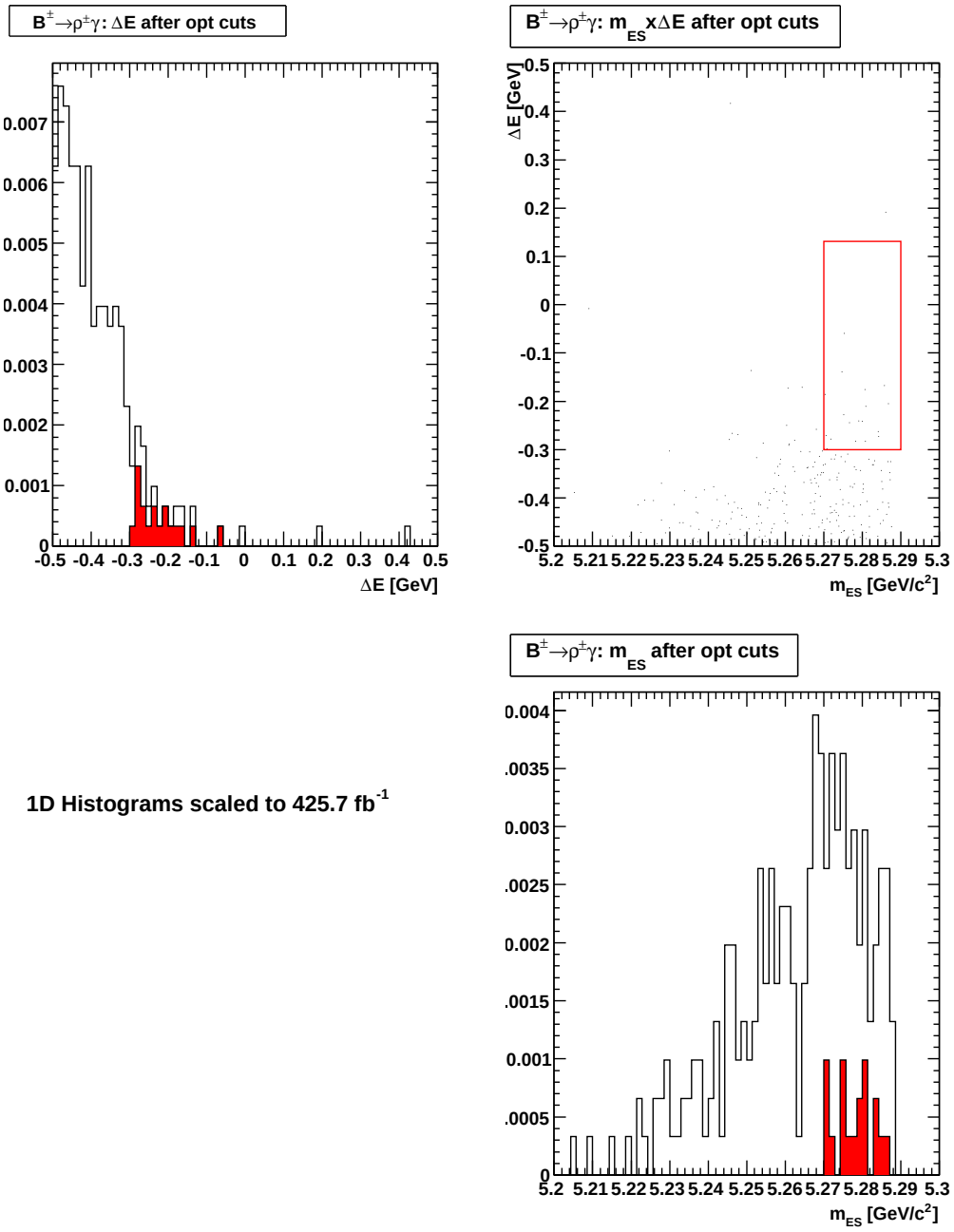


Figure D.5: Distribution of events in $B^{\pm} \to \rho^{\pm}\gamma$ MC that survive the optimized cut variables. The 1D histogram axis shows the number of events after being scaled to luminosity to find the expected numbers in data.

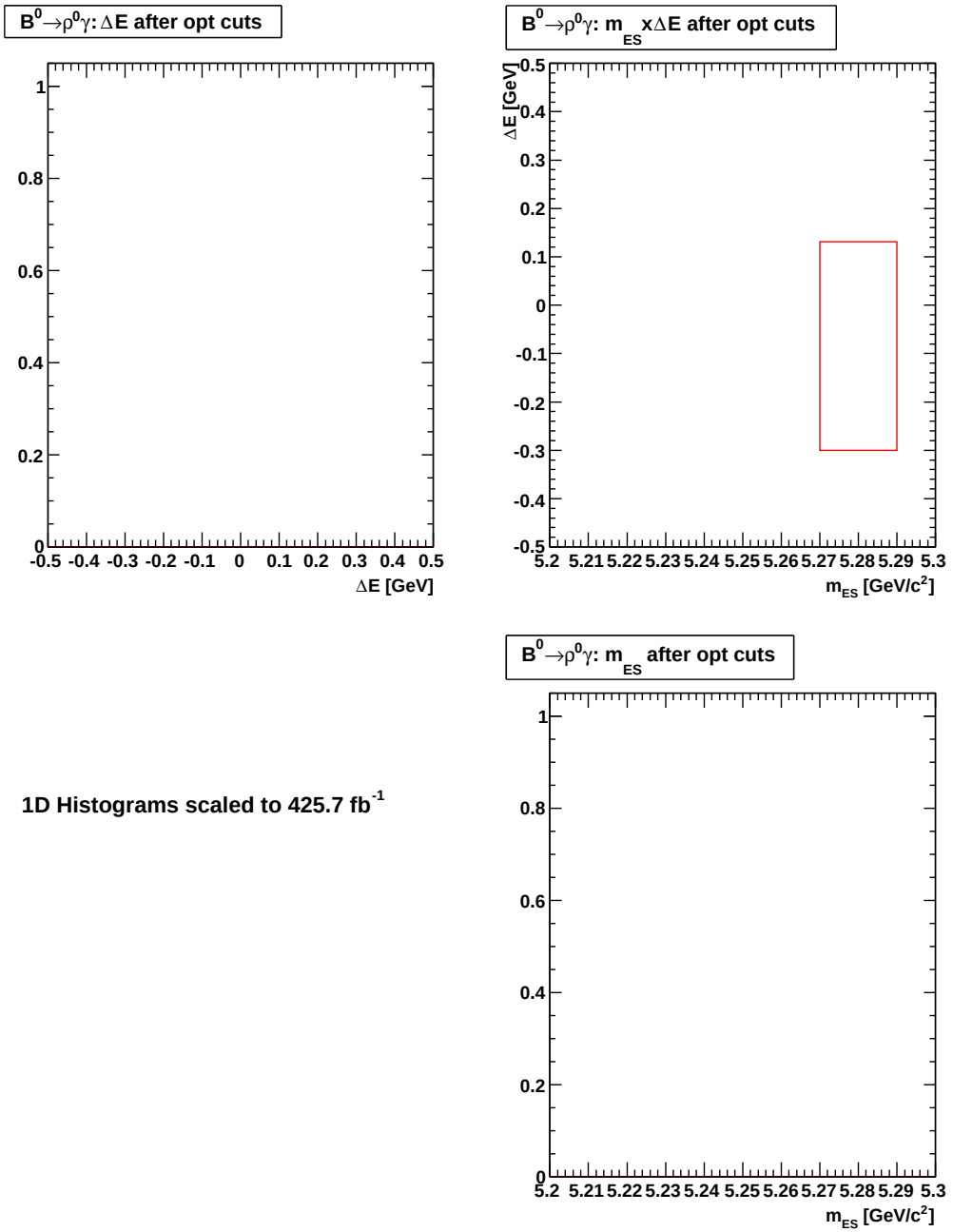


Figure D.6: Distribution of events in $B^0 \to \rho^0\gamma$ MC that survive the optimized cut variables. The 1D histogram axis shows the number of events after being scaled to luminosity to find the expected numbers in data. No events from this sample pass the optimized cuts.

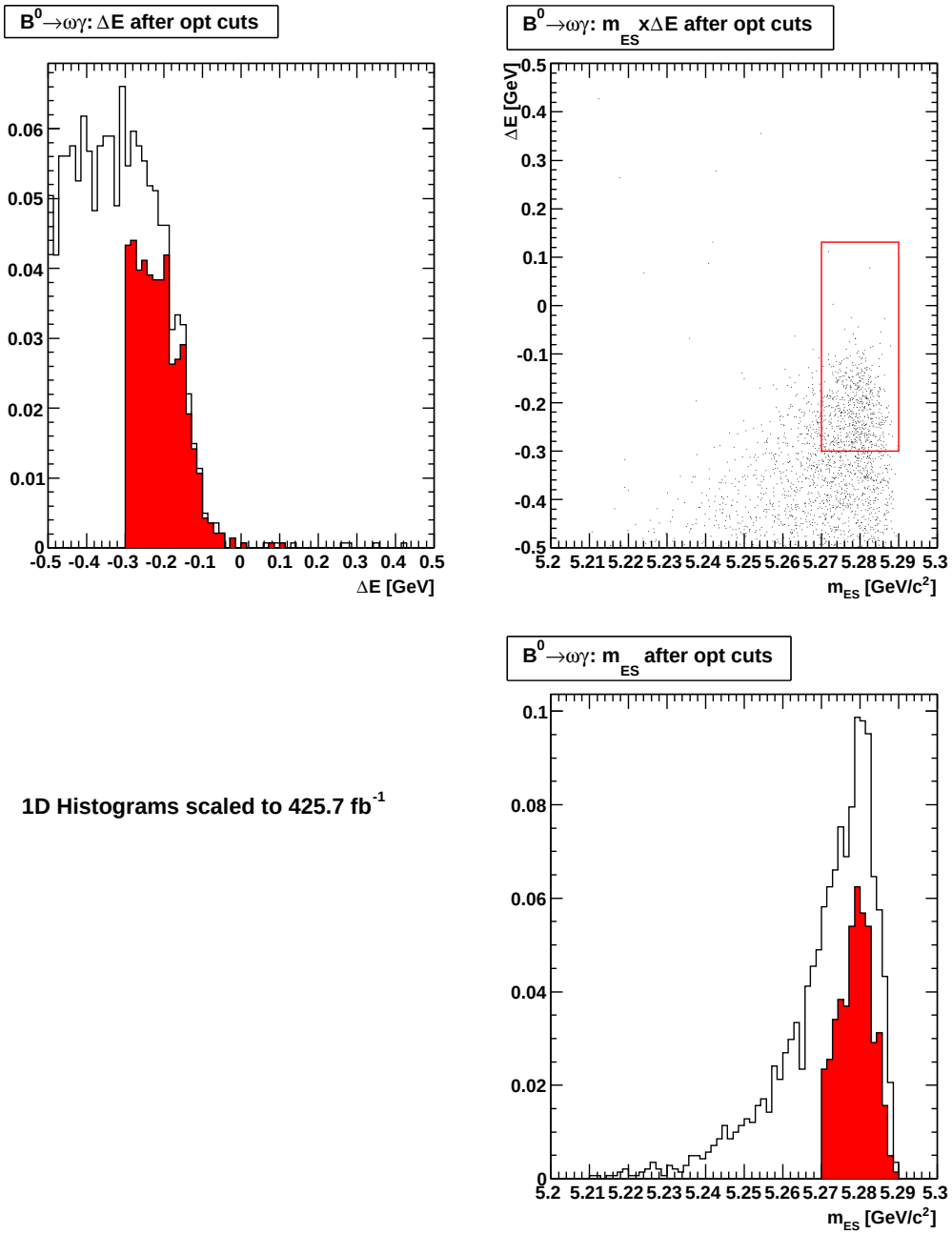


Figure D.7: Distribution of events in $B^0 \to \omega\gamma$ MC that survive the optimized cut variables. The 1D histogram axis shows the number of events after being scaled to luminosity to find the expected numbers in data.

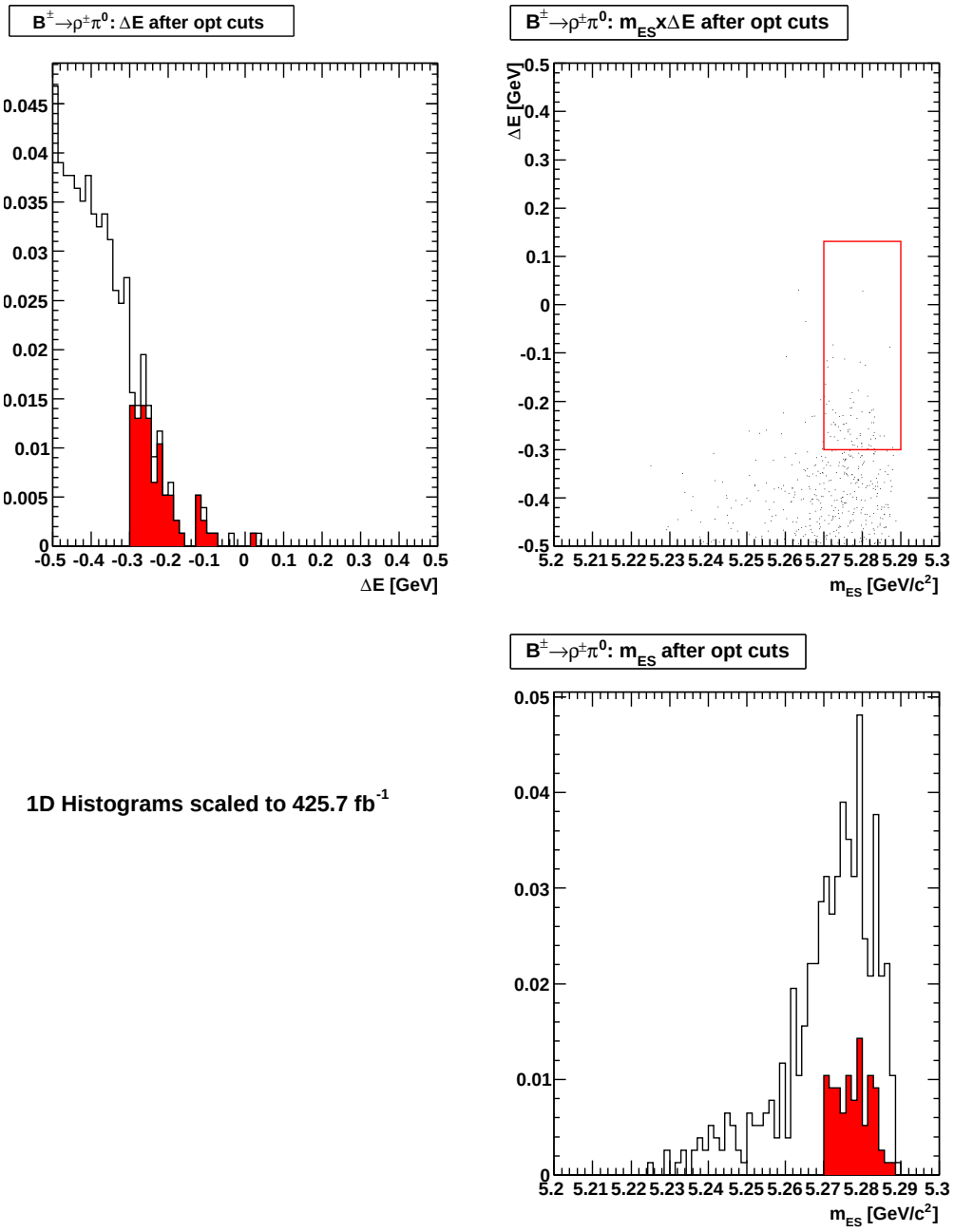


Figure D.8: Distribution of events in $B^{\pm} \to \rho^{\pm}\pi^0$ MC that survive the optimized cut variables. The 1D histogram axis shows the number of events after being scaled to luminosity to find the expected numbers in data.

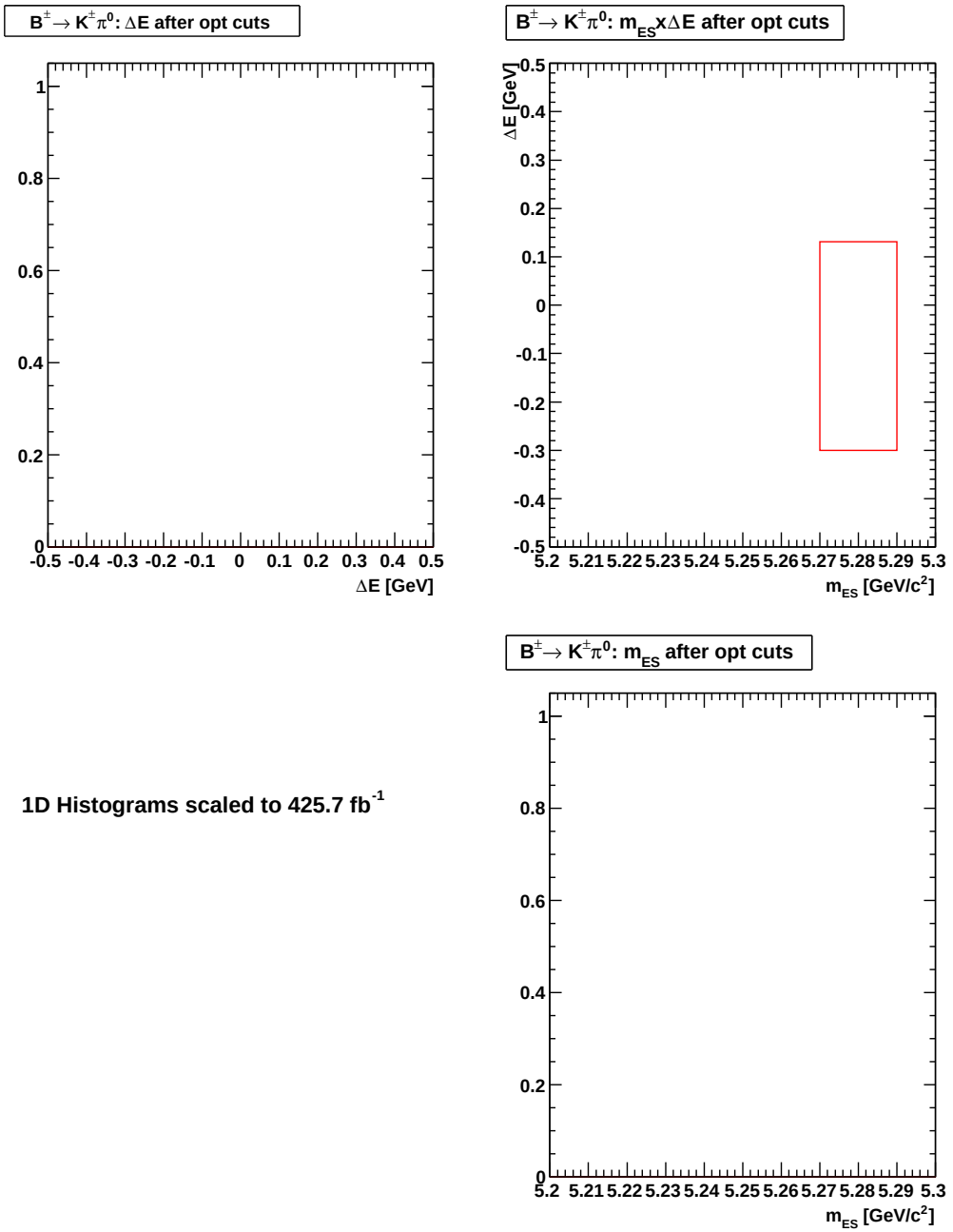


Figure D.9: Distribution of events in $B^{\pm} \to K^{\pm}\pi^0$ MC that survive the optimized cut variables. The 1D histogram axis shows the number of events after being scaled to luminosity to find the expected numbers in data. No events from this sample pass the optimized cuts.

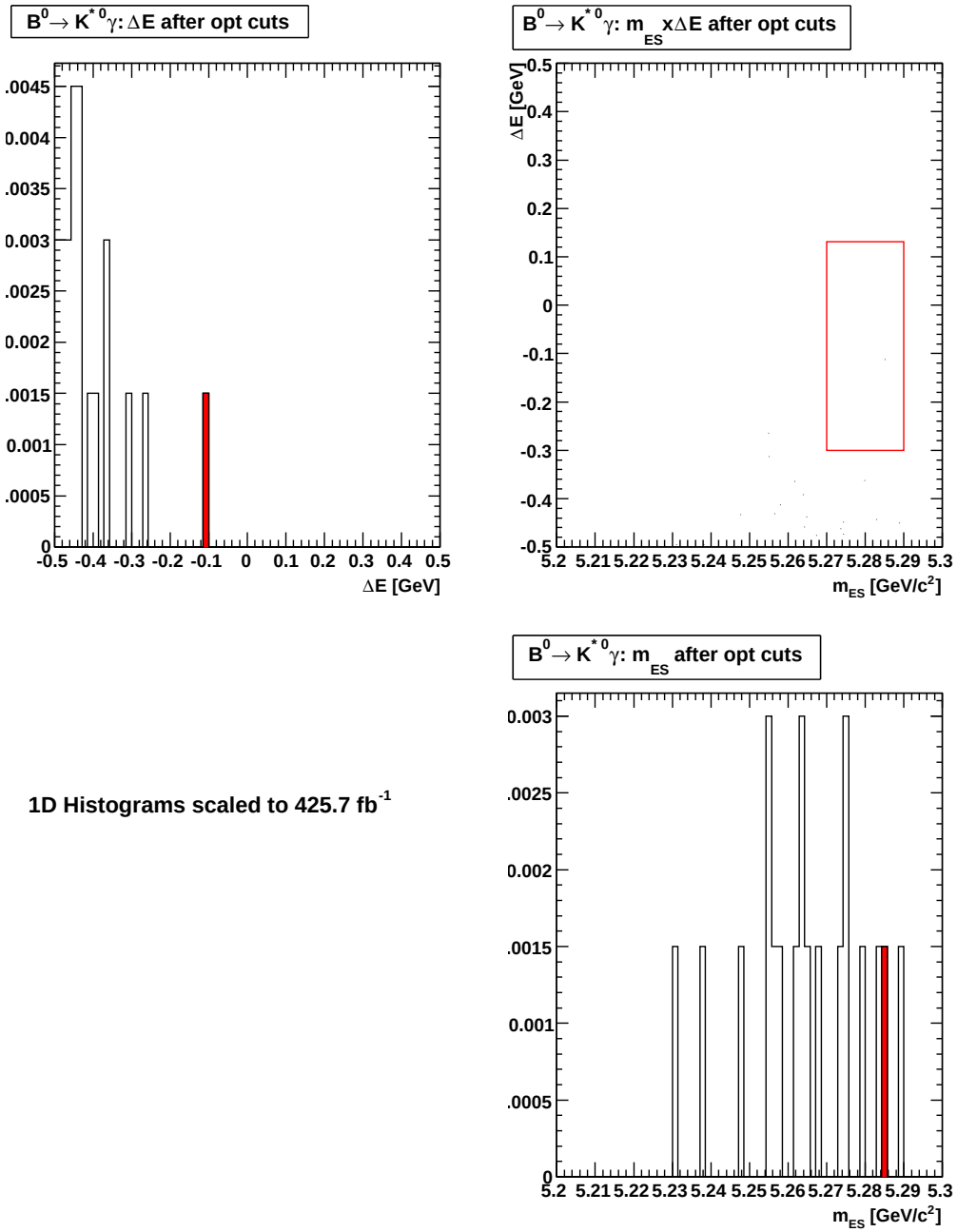


Figure D.10: Distribution of events in $B^0 \to K^{*0}\gamma$ MC that survive the optimized cut variables. The 1D histogram axis shows the number of events after being scaled to luminosity to find the expected numbers in data.

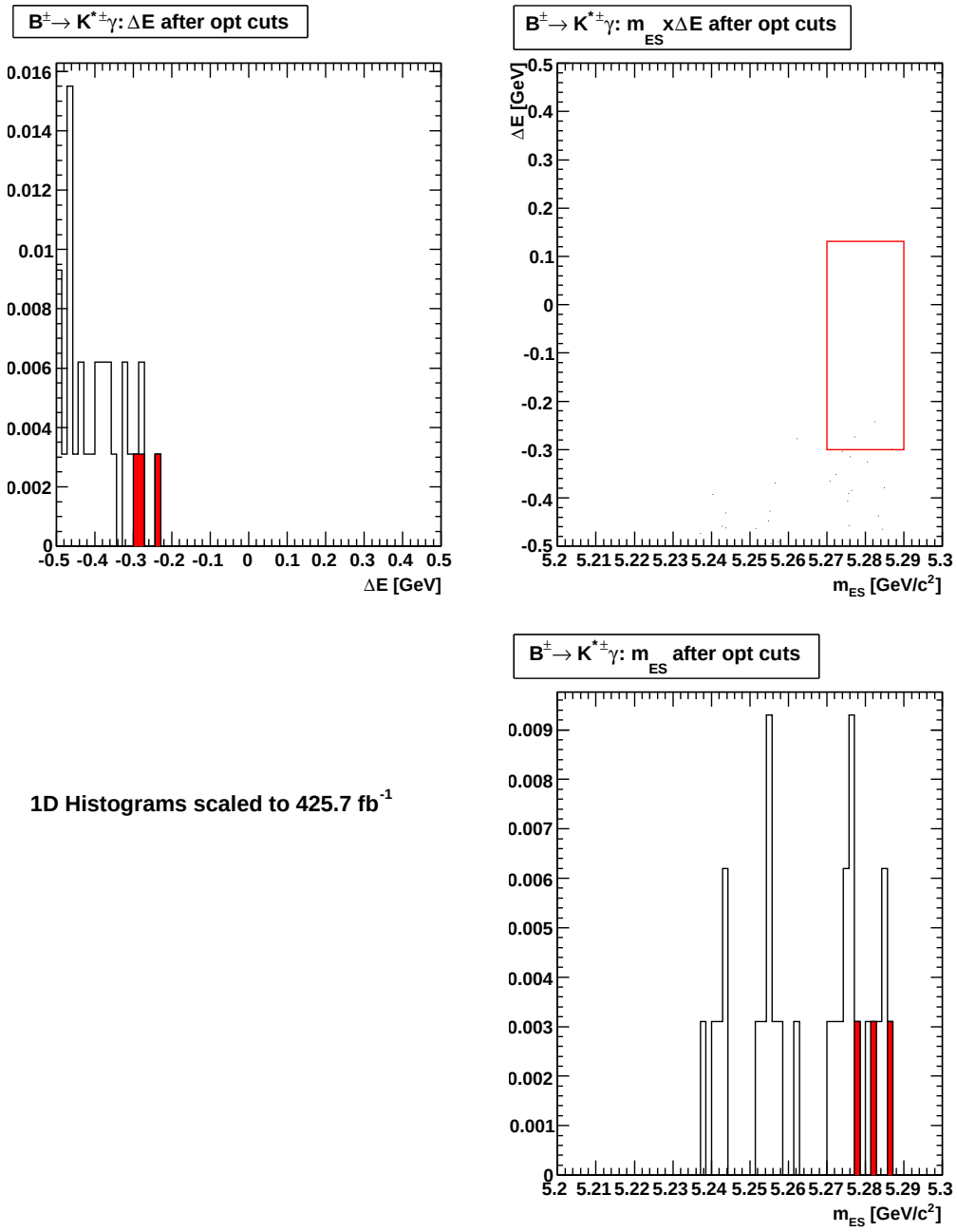


Figure D.11: Distribution of events in $B^{\pm} \to K^{*\pm}\gamma$ MC that survive the optimized cut variables. The 1D histogram axis shows the number of events after being scaled to luminosity to find the expected numbers in data.

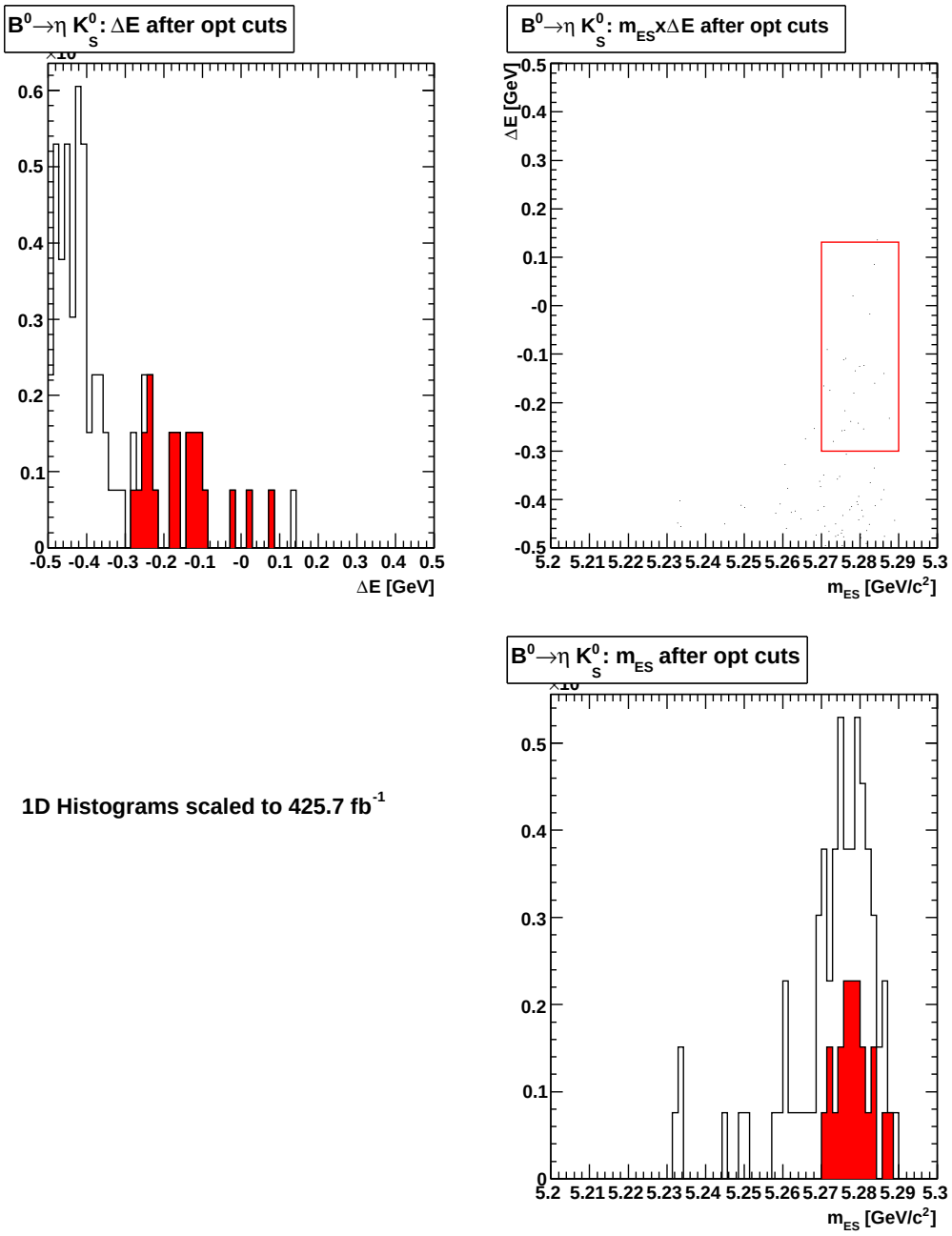


Figure D.12: Distribution of events in $B^0 \to K^0_s\eta$ MC that survive the optimized cut variables. The 1D histogram axis shows the number of events after being scaled to luminosity to find the expected numbers in data.

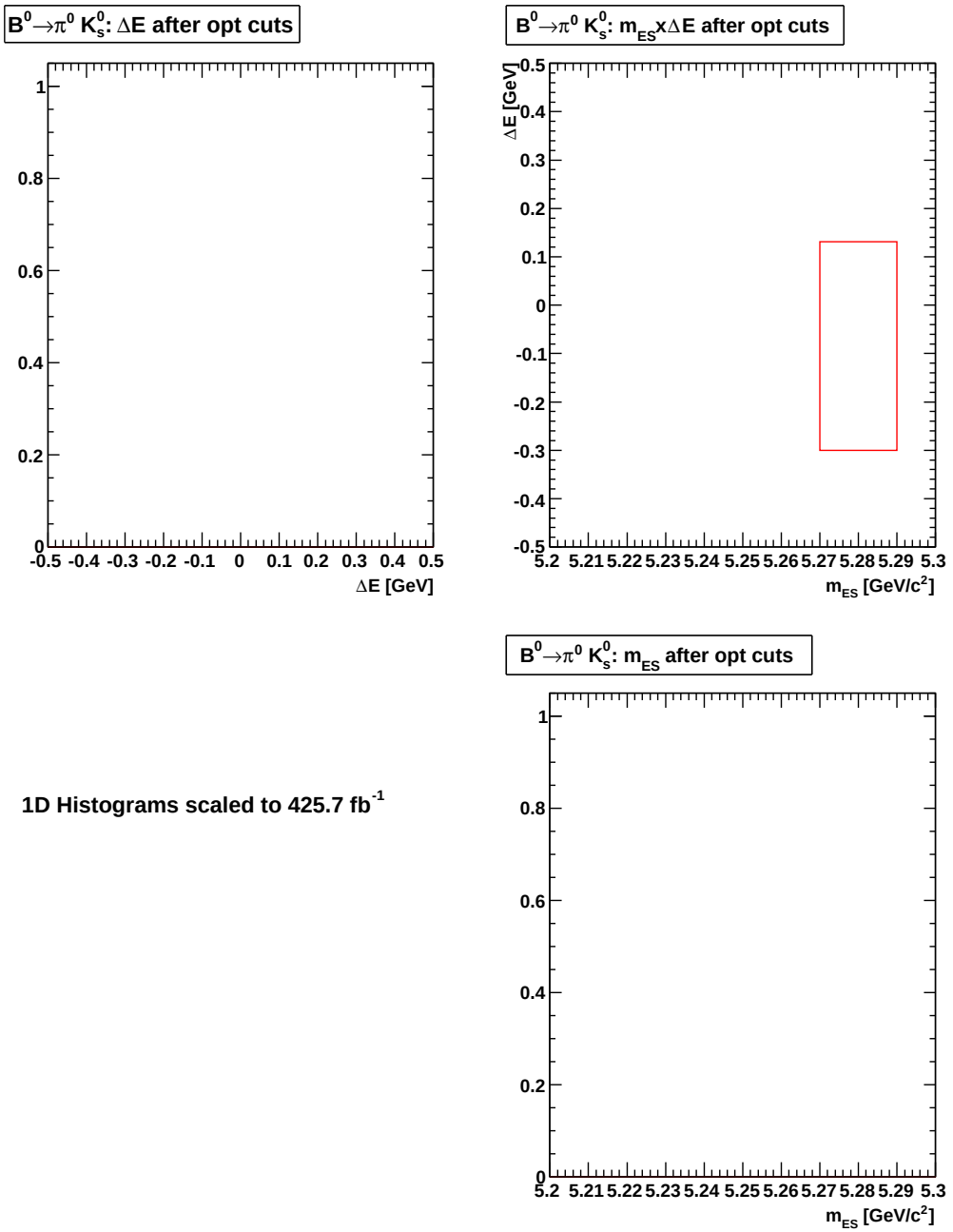


Figure D.13: Distribution of events in $B^0 \to K^0_S \pi^0$ MC that survive the optimized cut variables. The 1D histogram axis shows the number of events after being scaled to luminosity to find the expected numbers in data. No events from this sample pass the optimized selections.

Appendix E

Sample Toy MC fits

Each toy MC experiment proceeds by generating a dataset from the PDFs, or for signal-embedded toy MC from the background PDF and embedding into that dataset a number of signal events selected from the full *BABAR* simulation. Each dataset is fit with the full 2D extended maximum likelihood fit comprising a signal, background and peaking component. This appendix shows some examples of the fit for the pure and signal-embedded toy MC studies run in Section 5.3 for all three assumed signal branching fractions.

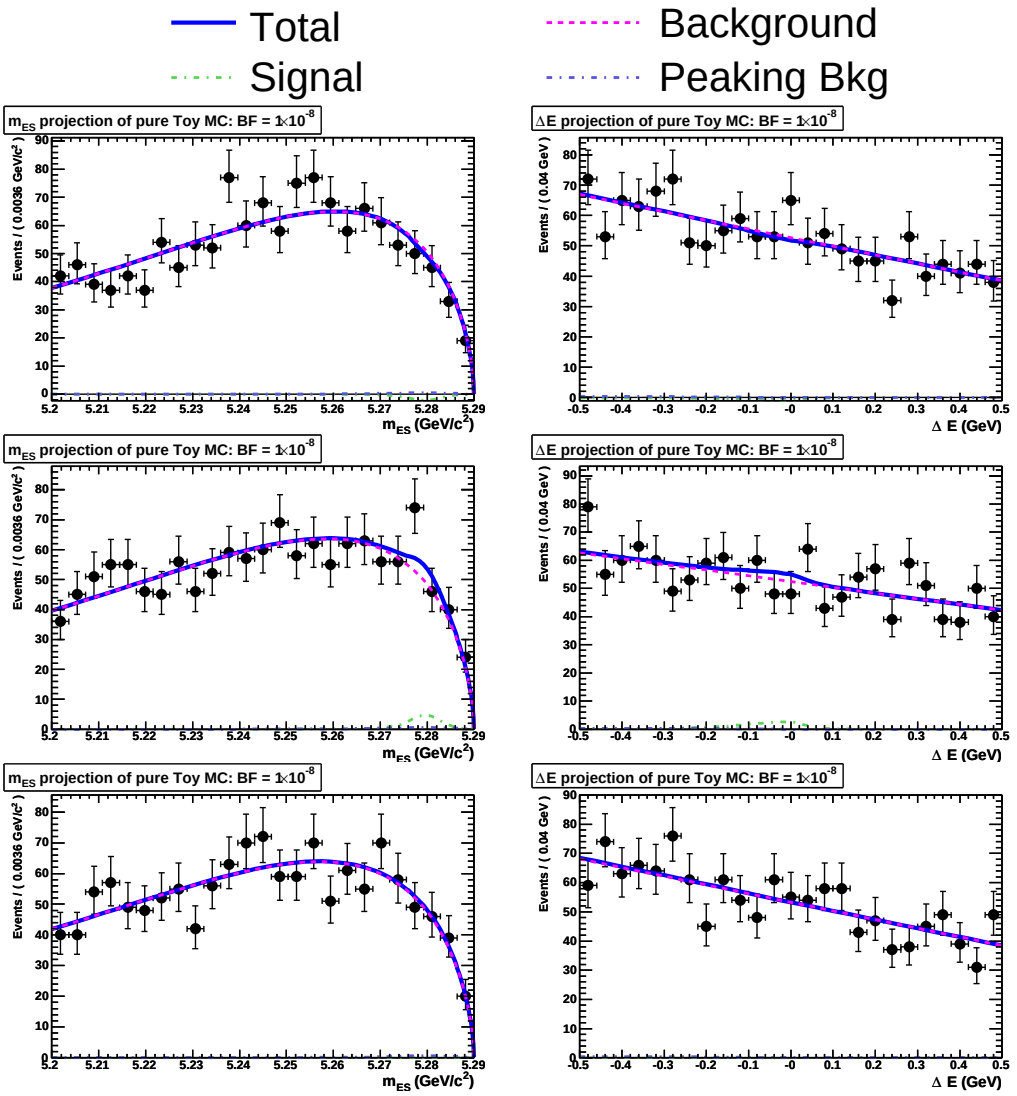


Figure E.1: Some example fits from the pure toy MC when assuming a signal branching fraction of 1×10^{-8}.

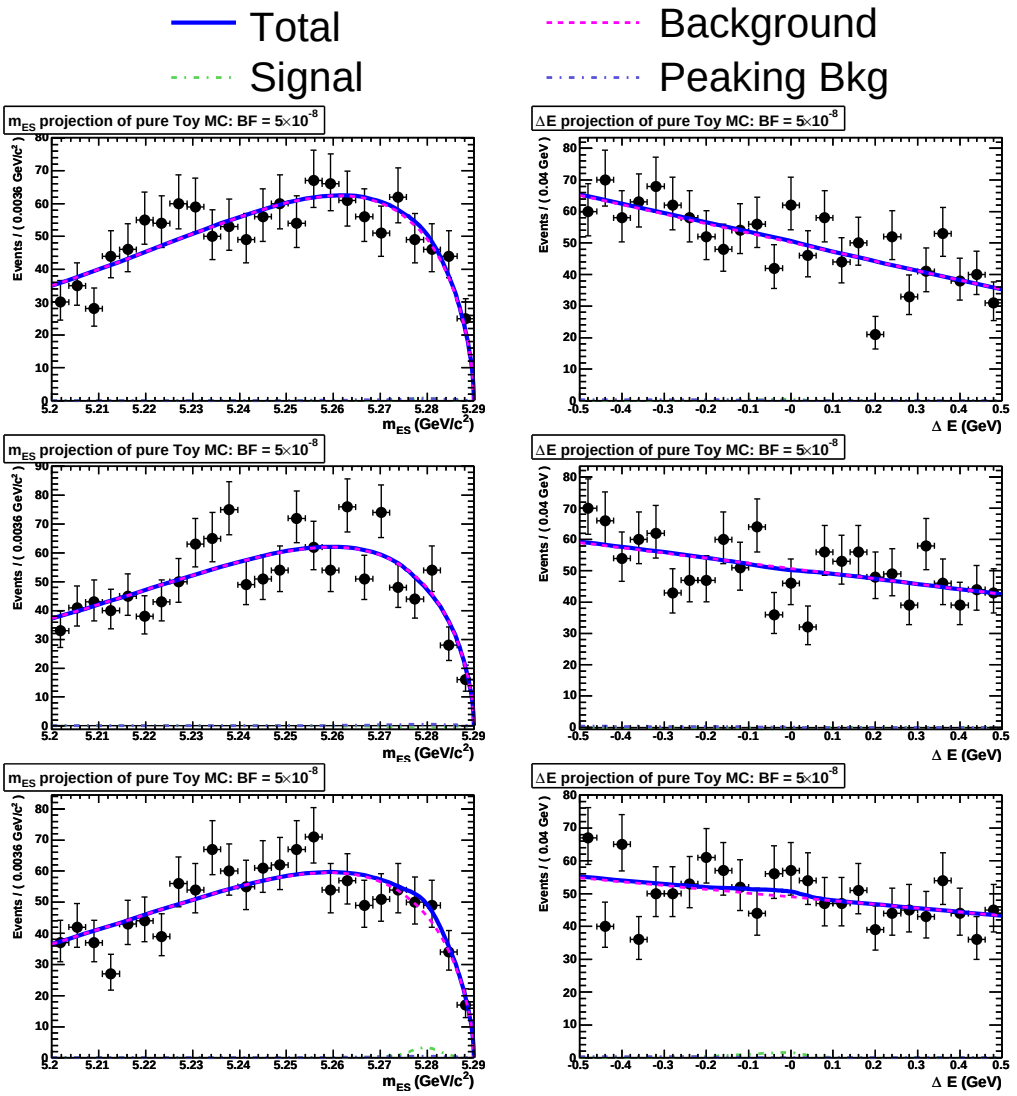


Figure E.2: Example of pure toy MC fits generated with an assumed signal branching fraction of 5×10^{-8}.

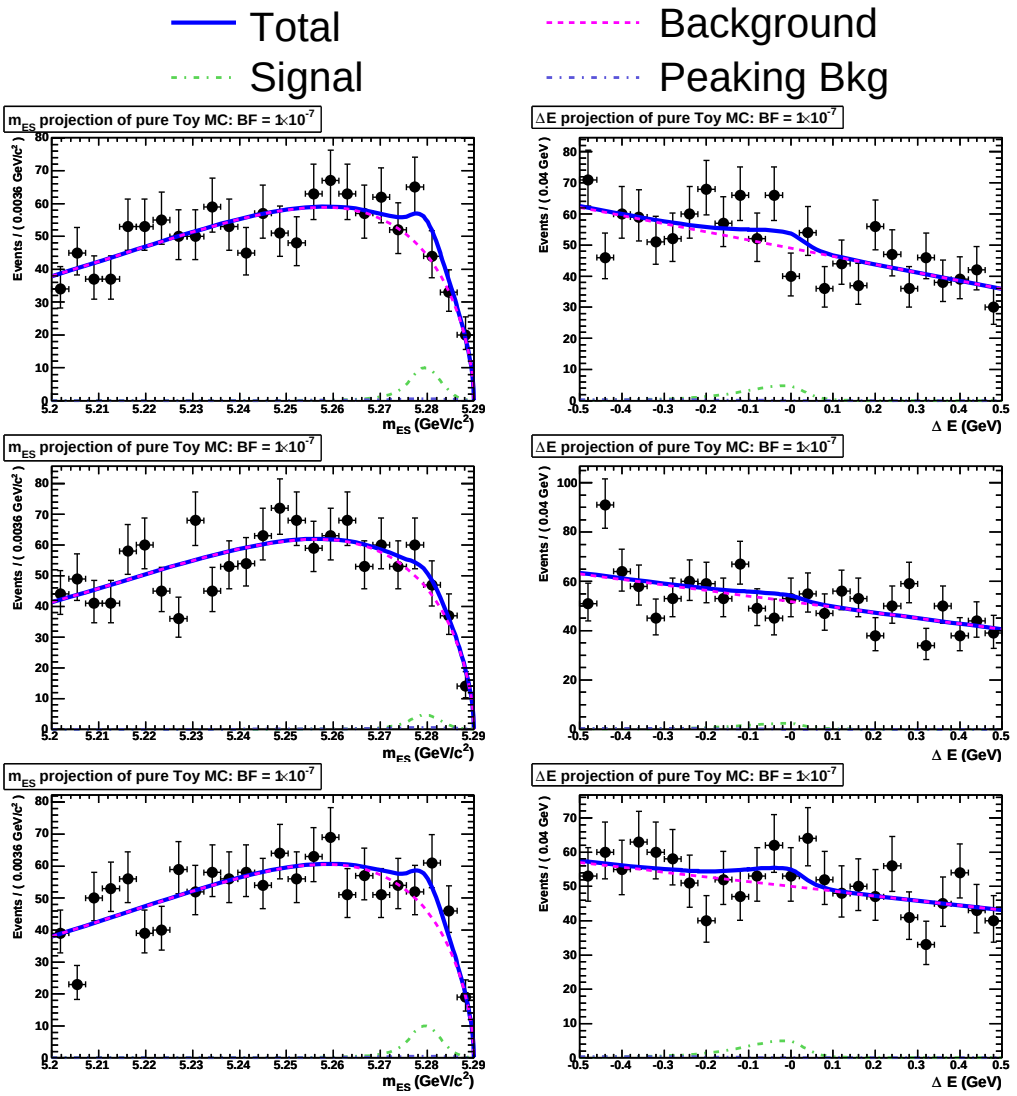


Figure E.3: Example pure toy MC fits generated with an assumed signal branching fraction of 1×10^{-7}.

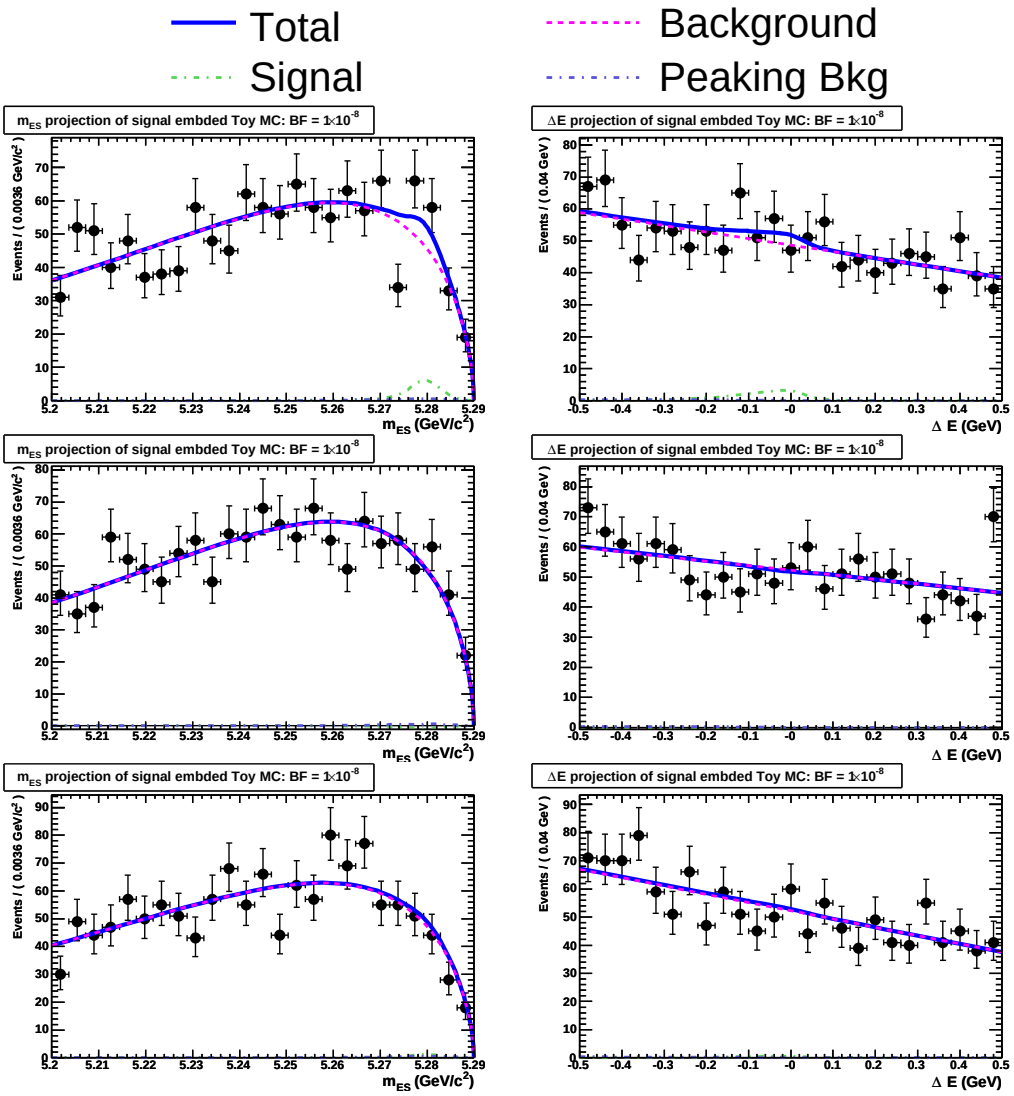


Figure E.4: Example of signal embedded toy MC fits generated with an assumed signal branching fraction of 1×10^{-8}.

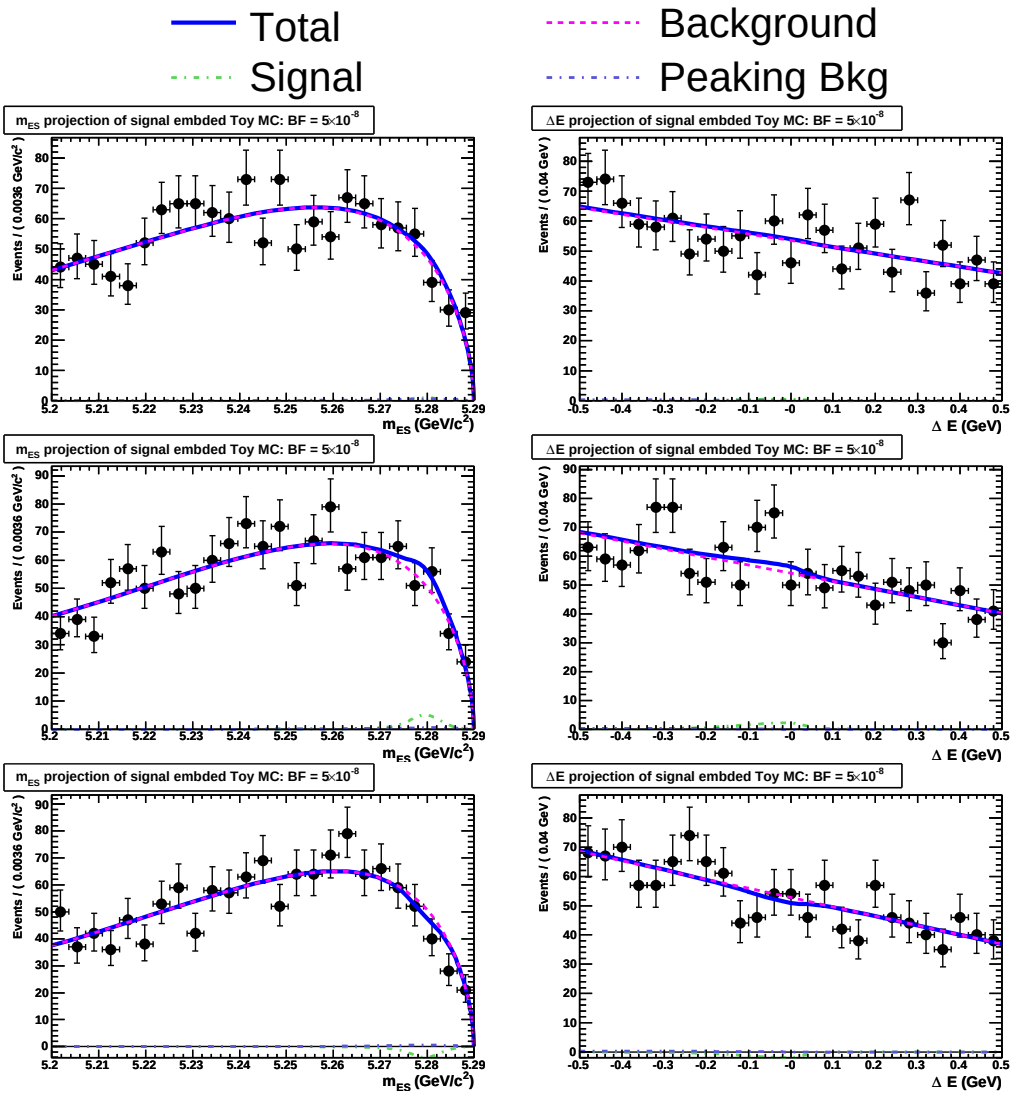


Figure E.5: Example of signal embedded toy MC fits generated with an assumed signal branching fraction of 5×10^{-8}.

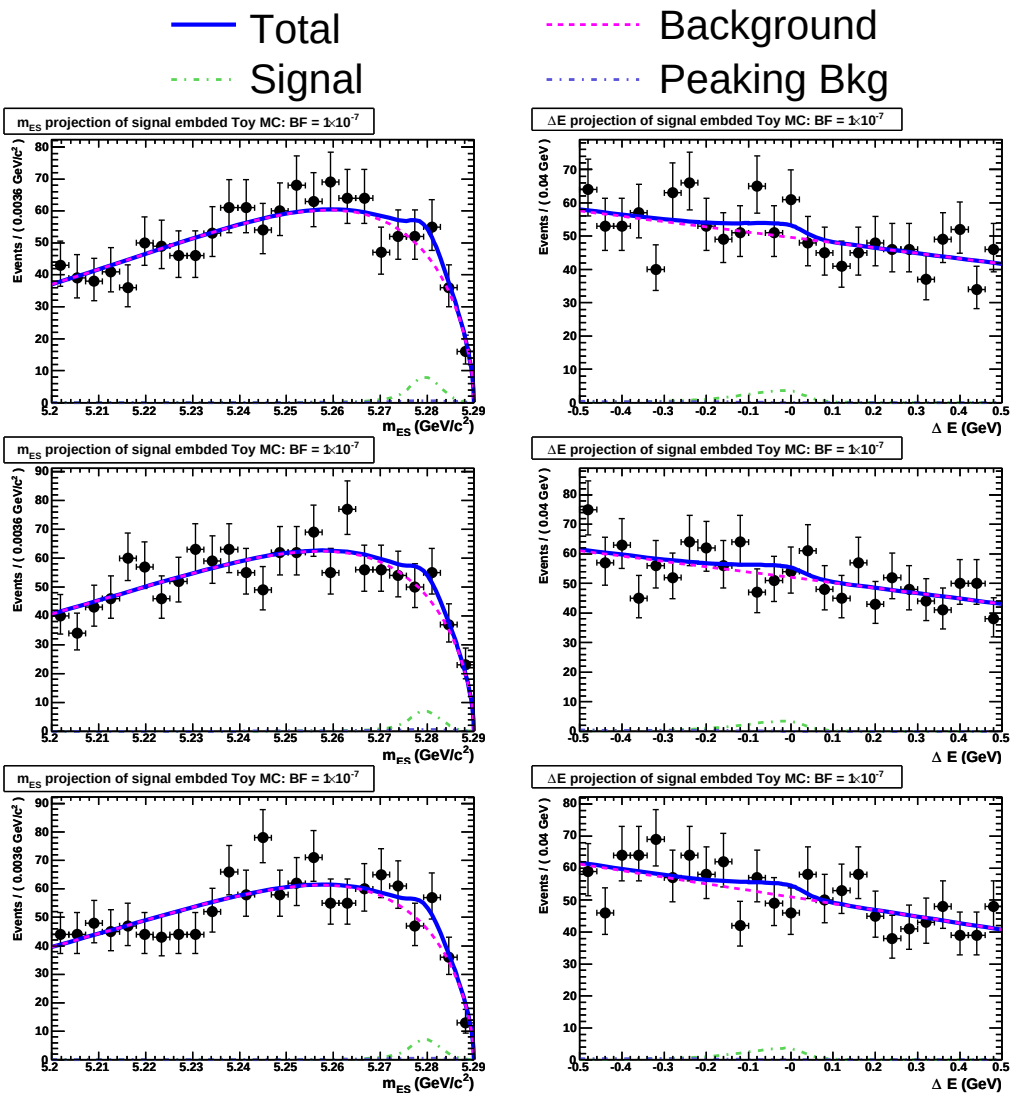


Figure E.6: Example of signal embedded toy MC fits generated with an assumed signal branching fraction of 1×10^{-7}.

Bibliography

[1] S.L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22: 579, 1961.

[2] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19(21):1264–1266, Nov 1967.

[3] N. Svartholm, editor. *Elementay Particle Physics*, 1968.

[4] P.W. Higgs. *Phys. Lett.*, 12:132, 1964.

[5] P.W. Higgs. *Phys. Rev. Lett.*, 23(508), 1964.

[6] P.W. Higgs. *Phys. Rev.*, 245:1156, 1966.

[7] N. Cabibbo. *Phys. Rev. Lett.*, 10:531, 1963.

[8] M. Kobayashi and T. Maskawa. *Prog. Theor. Phys.*, 49:652, 1973.

[9] L. Wolfenstein. Parameterization of the Kobayahi-Maskawa Matrix. *Phys. Rev. Lett.*, 51(21):1945, 1983.

[10] Glashow, S. L. and Iliopolous, J. and Maiani, L. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev. D*, 2(7):1285–1292, Oct 1970.

[11] K.G. Wilson and W. Zimmermann. Operator product expansion and composite field operators in the general framework of quantum field theory. *Commun. Math Phys.*, 24(2):87–106, 1972.

[12] Stefan W. Bosch and Gerhard Buchalla. The Double Radiative Decays $B \to \gamma\gamma$ in the Heavy Quark Limit. *Journal of High Energy Physics*, 2002 (08):054–054, 2002.

[13] T. M. Aliev and G. Turan. $B_s \to \gamma\gamma$ Decay in the Two-Higgs-Doublet Model. *Phys. Rev.*, D48:1176–1184, 1993.

[14] G. Devidze, A. Liparteliani, Ulf-G. Meißner. $B_{s,d} \to \gamma\gamma$ decay in the model with one universal extra dimension. *Physics Letters B*, 634:59–62, 2006.

[15] S. Villa et al. Search for the decay $B \to \gamma\gamma$. *Phys. Rev. D*, 73(5):051107, Mar 2006.

[16] B. Aubert et al. Search for the decay $B^0 \to \gamma\gamma$. *Phys. Rev. Lett.*, 87: 241803, 2001.

[17] L3 Collaboration. Search for the decays $B_{s,d} \to \gamma\gamma$. *Physics Letters B*, 363(1-2):137–144, 1995.

[18] Rene Brun and Fons Rademaker. ROOT – An Object Oriented Data Analysis Framework. *Nucl. Inst. and Methods A*, 389(1-2):81–86, 1997.

[19] A. Drescher et al. The Argus electron-photon calorimeter III. Electron-hadron separation. *Nucl. Inst. and Methods A*, 237(3):464 – 474, 1985.

[20] P. Bilori. *Nucl. Instrum. Meth. A*, 225(352), 1984.

[21] Geoffrey C. Fox and Stephen Wolfram. Event Shapes in e^+e^- Annihilation. *Nucl. Phys.*, B149:413, 1979.

[22] David J. Lange. The EvtGen Particle Decay Simulation Package. *Nucl. Instrum. Meth. A*, 462(1-2):152–155, 2001.

[23] S. Agostinelli et al. G4–A Simulation Toolkit. *Nucl. Instrum. Meth. A*, 506(3):250–303, 2003.

[24] B. Aubert et al. B meson decays to charmless meson pairs containing η or η' mesons. *Phys. Rev. D*, 80(11):112002, Dec. 2009.

[25] B. Aubert et al. Observation of $B^+ \to \eta\rho^+$ and search for B^0 decays to $\eta'\eta$, $\eta\pi^0$, $\eta'\pi^0$, and $\omega\pi^0$. *Phys. Rev. D*, 78(1):011107, Jul 2008. doi: 10.1103/PhysRevD.78.011107.

[26] F. Fang et al. Search for $B \to \rho\gamma$ and $B \to \omega\gamma$ in Run 1-5 data. *BaBar Analysis Document 1301, v. 12*, Dec. 2006.

[27] B. Aubert et al. Branching Fraction Measurements of $B^\pm \to \rho^\pm\gamma$, $B \to \rho\gamma$, and $B \to \omega\gamma$. *Phys. Rev. Lett.*, 98(15):151802, Apr 2007.

[28] D. Dujmic. Merged π^0 Reconstruction: From Cluster Shape to Mass. *BaBar Analysis Document #514, v3*, October 2002.

[29] I. Narsky. StatPatternRecognition: A C++ Package for Statistical Analysis of High Energy Physics Data. arXiv:physics/0507143v1, 2005.

[30] Giovanni Punzi. Sensitivity of Searches for New Signals and its Optimization. arxiv.org/abs/physics/0308063v2, 2003.

[31] M. J. Oreglia. *A Study of the Reactions* $\Psi' \to \gamma\gamma\Psi$. PhD thesis, SLAC, 1980. SLAC-236.

[32] J. Gaiser. *Charmonium Spectroscopy from the Radiative Decays of the J/ψ and ψ'*. PhD thesis, SLAC, 1982. SLAC-255.

[33] Tomasz Skwarnicki. *A Study of the Radiative Cascade Transitions between the Υ' and Υ Resonances*. PhD thesis, SLAC. DESY-F31-86-02.

[34] H. Albrecht et al. Search for hadronic $b \to u$ decays. *Physics Letters B*, 241(2):278 – 282, 1990. ISSN 0370-2693. doi: DOI:10.1016/0370-2693(90)91293-K.

[35] Roger Barlow. Extended Maximum Likelihood. *Nucl. Instrum. Meth. A*, 297(3):496–506, 1990.

[36] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975.

[37] Dolezal, Zdenek. Super KEKB and Belle II: Status of the KEK Super B Factory. (arXiv:0910.0388), 2009.

[38] Bona, M. and others. SuperB: A High-Luminosity Asymmetric e^+e^- Super Flavor Factory. Conceptual Design Report. (arXiv:0709.0451v2), 2007.