A Monte Carlo Study of $B^{\pm} \to K^{\pm}\gamma\gamma$ Decays at *BABAR*

by

Ada Rubin

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Master of Science (Physics)

Advisor: Sridhara Dasu

University of Wisconsin – Madison

2004

Contents

List of Figures

List of Tables

1 Introduction

1.1 Theoretical Background

The Standard Model (SM) is the name given to the set of accepted theories describing fundamental particles, of which all matter is composed, and the forces governing their interactions. To date, it has proved successful in predicting measurable quantities such as particle lifetimes, branching fractions, and asymmetries. Within the SM, the basic building bocks of matter are fermions, spin 1/2 particles which include quarks and leptons. Their properties are summarized in (Fig. 1 [1]).

FERMIONS matter constituents spin = 1/2, 3/2, 5/2, ...

Leptons spin = 1/2

Flavor	Mass GeV/c^2	Electric charge
ν_e electron neutrino	$<1\times10^{-8}$	0
e electron	0.000511	−1
ν_μ muon neutrino	<0.0002	0
μ muon	0.106	−1
ν_τ tau neutrino	<0.02	0
τ tau	1.7771	−1

Quarks spin = 1/2

Flavor	Approx. Mass GeV/c^2	Electric charge
u up	0.003	2/3
d down	0.006	−1/3
c charm	1.3	2/3
s strange	0.1	−1/3
t top	175	2/3
b bottom	4.3	−1/3

Figure 1: Particle properties.

There are four forces that govern the interactions of quarks and leptons: the gravitational, electromagnetic, weak and strong forces. Both the theory of strong interactions (quantum chromodynamics or QCD) and the unified theory of weak and electromagnetic interactions (electroweak) are incorporated into the SM. The SM forces are mediated by particles with integer spin called bosons; their properties are summarized in (Fig. 2 [1])

Whether a certain force will interact with a certain particle is dependent on the particle's physical properties. For instance, only charged particles can be affected by the electromagnetic force. Therefore, a neutrino will be affected by neither by an electric nor magnetic field. Likewise, only quarks, or "colored" particles, can interact with gluons, the mediators of the strong force. All known particles, though, are susceptible to the weak force.

The weak force is the only means by which a quark can couple to a quark of a different flavor. In the SM, flavor changing neutral currents (FCNC) are forbidden. Therefore, two quarks of the same charge and different flavor can never couple. For example, there can be no direct coupling between a b quark and an s or d quark. Flavor changing can only occur via charged weak currents, ($W^{\pm}$), so a b quark can decay into an s only as shown in (Fig. 3 [7]). This sort of decay process is referred to as a penguin decay.

BOSONS force carriers spin = 0, 1, 2, ...

Unified Electroweak spin = 1		
Name	Mass GeV/c^2	Electric charge
γ photon	0	0
W^-	80.4	−1
W^+	80.4	+1
Z^0	91.187	0

Strong (color) spin = 1		
Name	Mass GeV/c^2	Electric charge
g gluon	0	0

Figure 2: Boson properties.

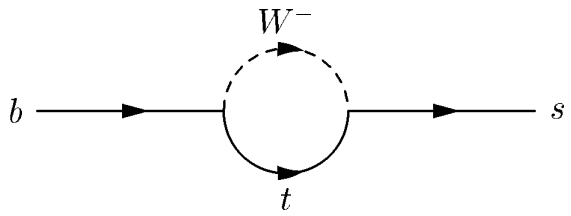


Figure 3: $b \to s$ penguin decay.

The decay for which we are searching, $B^{\pm} \rightarrow K^{\pm}\gamma\gamma$, is an example of a penguin decay, or more specifically, a radiative penguin decay since at least one photon is emitted. A variety of Feynman diagrams can be built to describe $B^{\pm} \rightarrow K^{\pm}\gamma\gamma$ as evinced in Figure 4 [6]. Investigation of penguin decays of B mesons such as $B \rightarrow K^{*}\gamma$ and $B \rightarrow Kl^{+}l^{-}$, have confirmed SM predictions regarding their behavior [3]. As will be described in the succeeding paragraphs, $b \rightarrow s\gamma\gamma$ decays, as is the case with other penguin decays, are perfectly suited, from both theoretical and experimental standpoints, to explore the limits of our current knowledge of effective FCNC decays.

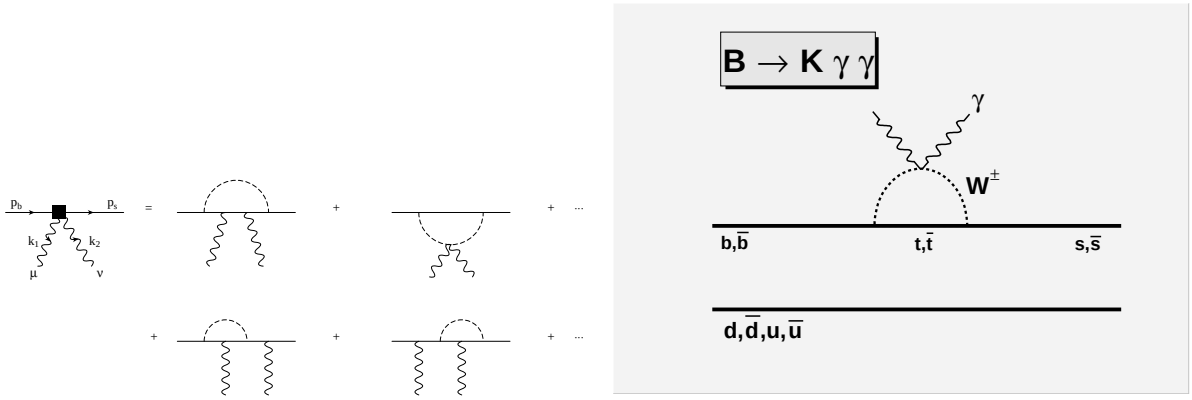


Figure 4: $B^{\pm} \rightarrow K^{\pm}\gamma\gamma$ decay diagrams. Left: Diagrams can be broken into two categories: reducible (bottom) decays are similar to $B \rightarrow K\gamma$ decays with an additional photon radiating from one of the quark legs and irreducible (top) decays where both photons are emitted by the loop particles. Right: Irreducible example of $B^{\pm} \rightarrow K^{\pm}\gamma\gamma$.

The interconversion of quarks via FCNC's is described by the Cabbibo-Kobayashi-Maskawa (CKM) matrix. The elements of this matrix are the proportionality constants used in determining the branching fractions or amplitudes of decay modes. Cross-generational, as well as inter-generational, couplings are possible because of quark mixing described by CKM. Therefore, the matrix will only enter into terms of the Lagrangian that describe quarks interacting via the charged weak current. The hadronic part of the Lagrangian density predicted by the SM for charged current coupling is of the form

$$-\tfrac{g}{\sqrt{2}} J^{\mu}_{cc} W^{\dagger}_{\mu} + \text{ Hermitian conjugate}$$

Coupled to the W-boson field, $W^{\dagger}_{\mu}$ is the quark current

$$J^{\mu}_{cc} = \begin{pmatrix} \bar{u_L} & \bar{c_L} & \bar{t_L} \end{pmatrix} \gamma^{/mu} V_{CKM} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}$$

$\bar{u_L}, \bar{c_L}, \bar{t_L}, d_L, s_L, b_L$ are the left-handed quark fields, the only quark fields that can couple to the charged W-boson field [17].

The 90% confidence limits on the magnitudes of the elements of the CKM matrix are

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.9741 - 0.9756 & 0.219 - 0.226 & 0.0025 - 0.0048 \\ 0.219 - 0.226 & 0.9732 - 0.9748 & 0.038 - 0.044 \\ 0.004 - 0.014 & 0.037 - 0.044 & 0.9990 - 0.9993 \end{pmatrix}$$

The CKM matrix is commonly parametrized with respect to the Cabibbo angle in order to emphasize its features.

$$V \approx \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}$$

The Cabibbo angle is $\lambda = 0.02229 \pm .0022$ [16]. In light of this, it is apparent that elements of the matrix that are of the order λ^3 are very tiny thus suppressing such decays. The phase angle, η, is needed to explain CP violation.

CP, or charge-parity, violation is a surprising and not very well understood consequence on the SM. C and P are operators that can act on a particle's function of state (or wave function or Lagrangian density function, etc.). The charge conjugation or C operator, when acting on a particle's function of state, will transform the particle into into its own antiparticle.

$$C|B\rangle = |\bar{B}\rangle$$

The parity operator, P, reverses the sign of the three spatial elements of the particle's motional four vectors,

$$P\Psi(t, \tilde{\mathbf{x}}) = \Psi(t, -\tilde{\mathbf{x}})$$

The combination of the two operators, CP, will reverse a particle's momentum and helicity (which describes the relative directions of a particle's spin and momentum) and turns it into its own anti-particle. If CP is violated for a decay mode, the anti-particle version of the decay will not be equivalent.

The large sample of B mesons produced to study CP violation at *BABAR* and other B-factories, also allow us to look for rare decays of the B meson such as radiative penguin decays. Study of rare B decays can provide stringent tests of the Standard Model and yield information about any new physics that may enhance the rates and CP asymmetry predictions. For example, $B^{\pm} \to K^{\pm}\gamma\gamma$ decays can potentially provide evidence of physics beyond the SM since penguin decay loops are very sensitive to non-SM effects. The presence of the heavier particles (W, t) in the loops can interact or be replaced by proposed heavy charged particles such as charged Higgs or SUSY particles. It should be further noted that in radiative penguin decays such as $b \to s\gamma$ and $b \to s\gamma\gamma$ (the inclusive mode that includes $B^{\pm} \to K^{\pm}\gamma\gamma$), any one of the charged particles (the b, s, t, or W) can emit photons. These photons are more energetic making these decays more accessible experimentally [7].

1.2 Predictions and Experimental Status

The most current SM prediction for the branching fraction of $B^{\pm} \rightarrow K^{\pm}\gamma\gamma$is

$$2.7 \times 10^{-7} \leq BR(B^{\pm} \rightarrow K^{\pm}\gamma\gamma) \leq 5.542 \times 10^{-7} \text{ [2]}$$

where the invariant mass of the photons is greater that 3 GeV. This prediction has incorporated QCD corrections into the effective Hamiltonian and taken into account all significant resonance contributions ($B- > K\eta$, $B- > K\eta'$, $B- > K\eta_c$).

The most current non-SM predictions are

	$tan\beta$	$BR(b \rightarrow s\gamma\gamma)(\times 10^{-7})$
SM		1.60
2HDM	0.5	0.23-14.67
Model I	1	0.23-1.26
	10	1.57-1.59
2HDM	0.5	2.19-16.67
Model II	1	2.07-9.65
	10	2.03-7.76

where $tan\beta = \xi$ for the first Two Higgs Model Doublet Model and $tan\beta = 1/\xi$ for the second model [6]. ξ vacuum expectation value of the Higgs field.

No experimental limits have been placed, so far, on the limit of $B \rightarrow K\gamma\gamma$, but limits have been placed on $B_s \rightarrow \gamma\gamma$, a member of the same inclusive decay family, $b- > s\gamma\gamma$. As such they are expected to behave similarly. The upper limit placed by L3 on the branching fraction is

$$Br(B_s \rightarrow \gamma\gamma) < 14.8^{-}5 \text{ at the 90\% confidence level. [9]}$$

while the SM predicts the actual value to be 5×10^{-7} [18]. We hope to place a significantly lower upper limit, on the order of 10^{-7}, for $B^{\pm} \rightarrow K^{\pm}\gamma\gamma$.

1.3 Analysis Overview

The purpose of this study is to isolate $B^{\pm} \rightarrow K^{\pm}\gamma\gamma$ events for the purpose of putting an upper limit on the branching fraction at the 90% confidence level. This is simply an expectation of the branching fraction based on simulated Monte Carlo data and assuming the branching fraction is 2×10^{-7}. We will begin with a discussion of the details of the experiment, starting with a description of the *BABAR* detector. We will then describe the means by which we reduced the data in order to isolate the decay. Finally, we describe how an upper limit is placed on the branching fraction with the collected data.

2 PEP-II and the *BABAR* Detector

The *BABAR* detector has been designed to detect the results of e^+e^- collisions within it. The e^+e^- collisions have been optimized to produces $\Upsilon(4S)$ particles which are the $4S$ resonances of $b\bar{b}$ quarks. The $\Upsilon(4S)$ decays into $B\bar{B}$ meson pairs nearly 100% of the time. Hence, this facility has been dubbed a "B Factory". [1].

2.1 PEP-II

PEP-II refers to the two storage rings, the high energy ring (HER) and low energy ring (LER) that supply the colliding e^+e^- beams to the *BABAR* detector. The two beams are of differing energies, 3.1 and 9 GeV. These energies provide a boost, $\beta\gamma = .57$, to the resultant $\Upsilon(4S)$ and B's in the lab frame. The boost carries the B particles downstream in the direction of the higher energy beam. This forward motion enables the decay products to separate to an average distance of 300 μm, allowing scientists to observe the distances between their points of decay. The luminosities delivered since *BABAR* began operations is shown in Fig. 5

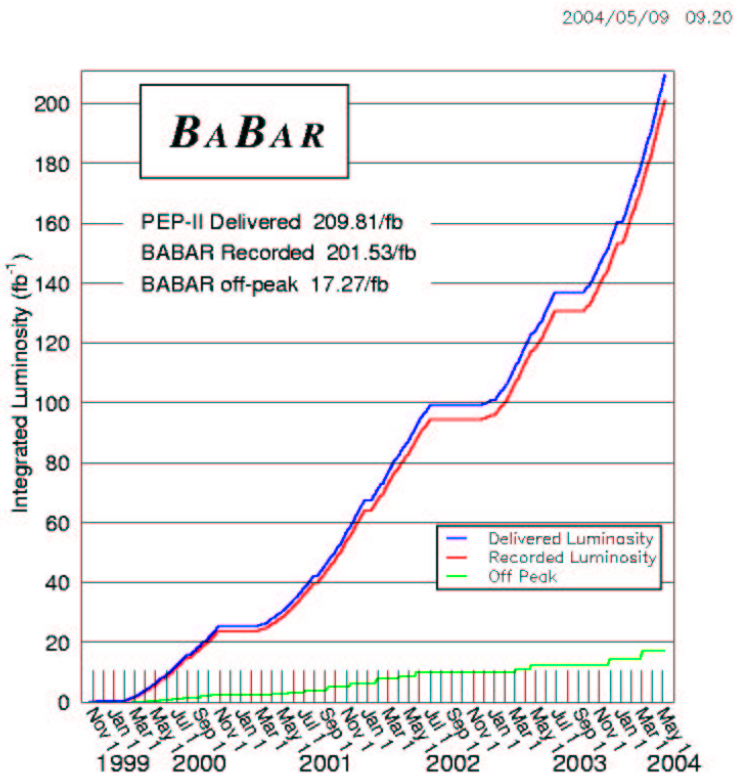


Figure 5: Integrated PEP-II-delivered and *BABAR*-recorded luminosities since 15, October 1999.

[1] Most of the diagrams and information cited in this section came from [19]

2.2 The *BABAR* Detector

The *BABAR* detector is composed of five detecting layers and a magnet layer. Design of the detector has been optimized for the detection of particles commonly involved in B decays such as electrons, muons, pions, and kaons.

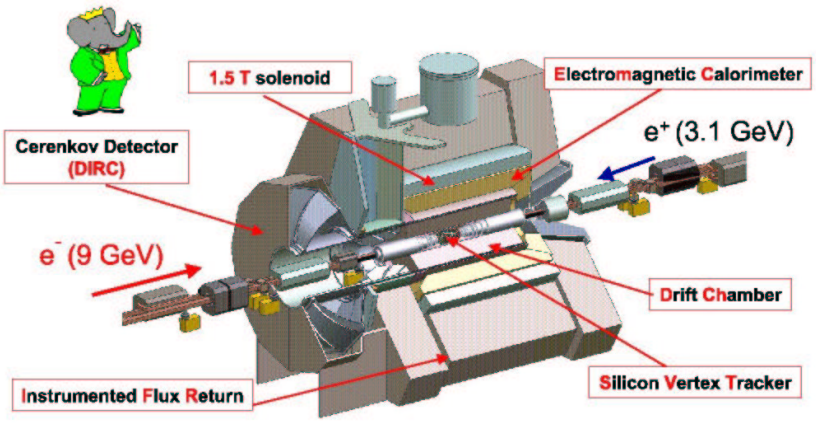


Figure 6: The *BABAR* detector assembly.

2.2.1 Silicon Vertex Detector (SVT)

The SVT is the innermost layer of the detector and is, therefore, essential in the reconstruction of B decay vertices. As such, its design requires very accurate position information for charged particles. As such,it should be able to resolve the vertices of the B decays which are, on average, 300 μm apart. The resolutions are given for each layer are given as a function of angle with respect to the beam axis in Fig. 7 along with a schematic of the layers (Fig. 8). The inner three layers are closest to the beam-pipe. Like all silicon microstrip detectors, the SVT modules function by measuring a current induced by the ionization of a silicon substrate.

2.2.2 Drift Chamber (DCH)

The DCH provides further tracking information, momenta and angles, about charged particles that pass through it. Tracks are reconstructed from the signals on wires (Fig. 9) suspended in the helium/isobutane filled chamber. Similar to the SVT, these signals are stimulated by ionization of the gas by the passing of charged particles. Resolution is a key design requirement of the DCH, as well, but for more than the distance between tracks. In order to fit a reconstructed track to an incoming track reconstructed by the SVT (and also to the DIRC, the next layer), the DCH must be able to spatially resolve hits that are at least 140 μm apart. de/dx resolution of low momentum particles ($< 700MeV$) that haven't enough energy to reach the outer layers must be good for identification of such particles;

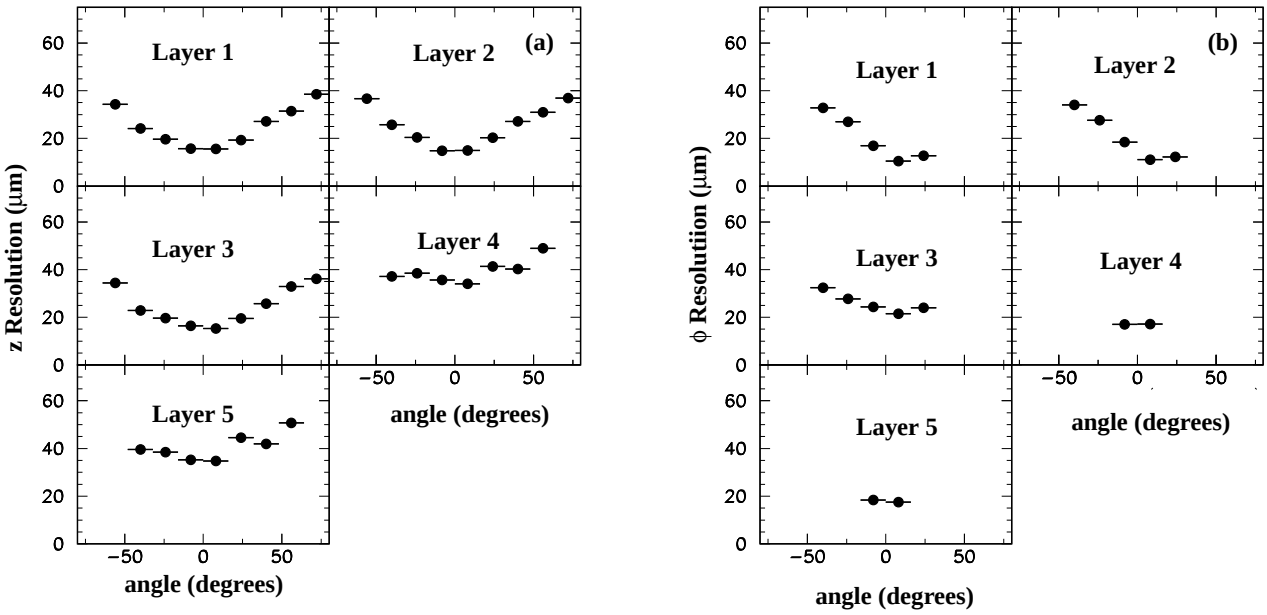


Figure 7: SVT hit resolutions in the z (left) and ϕ (right) with respect to the angle.

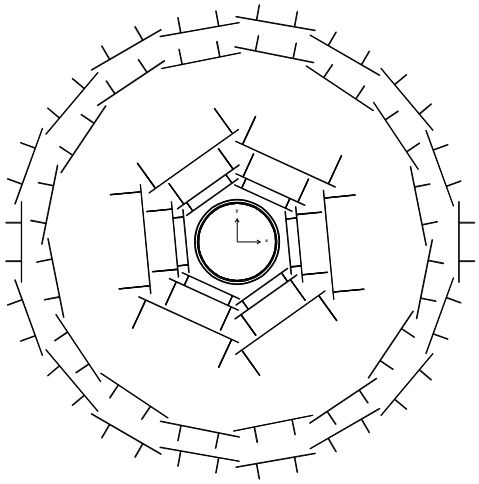


Figure 8: Cross-sectional view of the SVT.

it needs a de/dx be at least 7% to differentiate K's and π's. Momentum resolution of high energy particles ($> 1GeV$), such as those coming from B an D decays, must be high as well to reliably reconstruct events kinematically; the resolution on the transverse momentum is found to be $\sigma_{p_T} \approx 0.03 \times p_T$ for tracks with $p_T > 1GeV$.

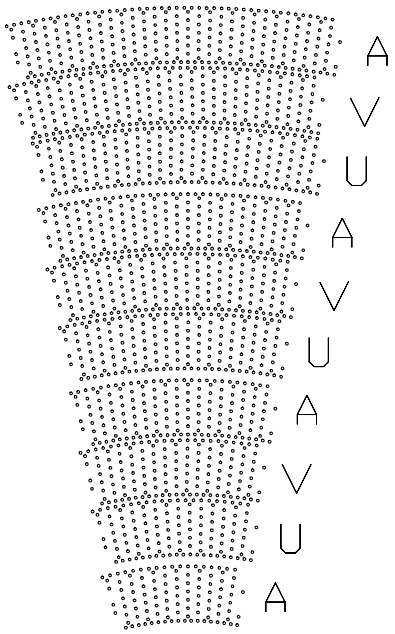


Figure 9: Wire map for one quadrant of the Drift Chamber endplate. A = Axial wires running parallel to z. U,V = Stereo wires are tilted at an angle w.r.t. z for Δz resolution.

2.2.3 Detector of Internally Reflected Cherenkov Light (DIRC)

The main purpose of the DIRC is to determine the mass of charged particles passing through it, thus identifying the particle, essential in the study of rare decays. The DIRC is composed of a ring of synthetic quartz bars. As a result of the Cherenkov effect, when a charged particle hits one of these bars photons are emitted at an angle $\theta_c = cos^{-1}(1/\beta n)$ (Fig. 10). These photons are then transmitted via internal reflection to the end of the bar where they can be detected by a series if PMT's. Combining this information, specifically the value of β, with the momentum information provided by the DCH the mass of the particle can be extracted. The DIRC has a can resolve Cherenkov angle $9.6mrad.$ so it can separate K's and pi's in the momentum range $0.8 - 4GeV$.

2.2.4 Electromagnetic Calorimeter (EMC)

A calorimeter, as the name implies, measures energy. A particle will deposit its energy in the EMC cesium-iodide crystal in the form of light which is detectable with a series of photodiodes that are attached to the crystals. It is very important that the energy resolution

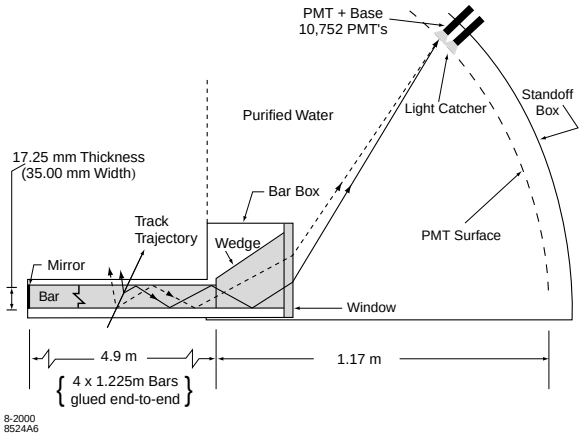


Figure 10: Schematic of a single DIRC quartz bar.

of the EMC be quite high in order to measure with high precision the energy of photons, especially important in identifying π^0's and η's. Energy resolution, a function of the energy, has been determined to be

$$\frac{\sigma}{E} = \frac{(2.32 \pm 0.30)\%}{\sqrt{4}E(GeV)} \bigoplus (1.85 \pm 0.12)\%$$

which has been calculated from the following fit (Fig. 11).

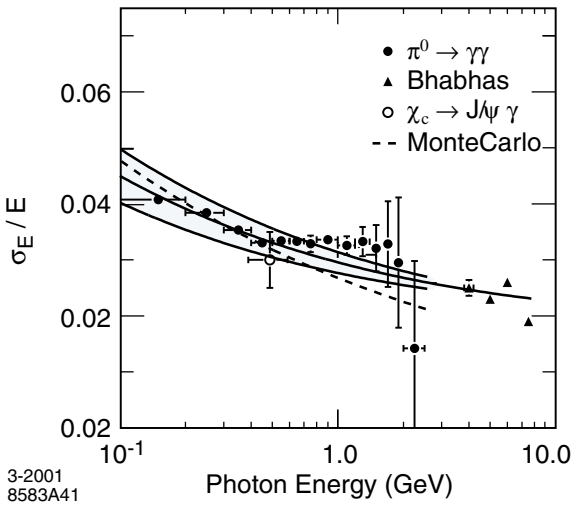


Figure 11: The energy resolution for the EMC measured for photons and electrons from various processes. The solid curve is a fit to Equation 6 and the shaded area denotes the rms error of the fit.

It should be further noted that the EMC can detect EM showers with energy as low as $20MeV$. Additionally PID information is provided by the shower shape which is different for muons, electrons and hadrons.

2.2.5 Magnet

Between the EMC and IFR lies a 1.5T solenoidal magnet that makes measurement of the transverse momentum and particle charge possible.

2.2.6 Instrumented Flux Return (IFR)

The sole purpose of the IFR is μ and neutral hadron (esp. K_L) identification. To this end, it is composed of 18-19 layers of resistive plate chambers (RPC). Layers of steel are interspersed between RPC layers. Hadrons are expected to interact with the steel more toward the inner part of the IFR while muons are expected to make it further to the outer layers and can ionize the gas in the center layer of the RPC's (Fig. 12).

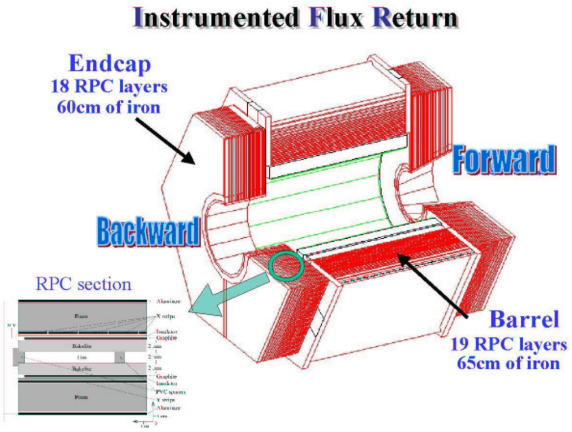


Figure 12: .

Sadly, the IFR has not performed well under experimental stress and efficiencies have steadily decreased since *BABAR* began operation. Fortunately, this does not affect my analysis since I'm not reconstructing decay channels involving μ's or K_L's.

3 Data Set

The analysis is based on results extracted from Monte Carlo (MC) simulated events. The MC data is especially tuned by *BABAR*'s MC Simulation and Production group to simulate detector conditions. We use the fifth generation *BABAR*Monte Carlo samples as well as off-peak data. Off-peak data is produced when one or more of the beams do not reach the energies (9 and 3.1 GeV) required to produce $BbarB$ events. It is used as a check of the MC. $11.95fb^{-1}$ of off-peak data was collected for this paper.

3.1 Monte Carlo Data set

The $e+e-$ collisions provided to the detector by PEP-II will produce a $B\bar{B}$ particle only about 2% of the time. Background filters are quite effective in eliminating $e+e \rightarrow l+l$ events, but are quite ineffectual in eliminating $e+e \rightarrow u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}$ events called continuum background. Therefore, to isolate the $e+e--> b\bar{b}$ events, all other e+e- decays, especially continuum background, should be removed from the data set. We do this by comparing continuum and signal MC and exploiting the differences by placing cuts. Once these cuts are placed, we then count the remaining events in order to approximate efficiencies.

	Events	Luminosity
$B^{\pm} \rightarrow K^{\pm}\gamma\gamma$	6,666	$63.5ab^{-1}$ $(BF = 2 \times 10^{-7})$
uds	444,600,000	$212.7fb^{-1}$
$c\bar{c}$	289,650,000	$222.8fb^{-1}$
$B^0\bar{B}^0$	348,812,000	$664.4fb^{-1}$
B^+B^-	335,916,000	$639.8fb^{-1}$

Table 1: Number of events generated with corresponding luminosities.

4 Selection

The first step in any analysis at *BABAR* is to separate signal events from the data. We start by placing cuts on variables that are known to be distributed differently for continuum background ($e^+e^- \to c\bar{c}$, $e^+e^- \to u\bar{u}$, $e^+e^- \to d\bar{d}$, $e^+e^- \to s\bar{s}$) and B events ($e^+e^- \to b\bar{b}$). To that end, the selection criteria that we apply in the analysis can be summarized in several steps. Each step is described in detail in the next paragraphs.

4.1 Two Photon Event Filter

We use a fast filter that selects only events with at least two high energy neutral photons to reduce the amount of data read and speed up the analysis. To pass the filter, an event must meet the following requirements:

- The energy of the highest energy photon in the event, E^*_γ, must lie in the interval $1.5\,\mathrm{GeV} < E^*_\gamma < 3.5\,\mathrm{GeV}$ in the CMS frame,
- The energy second highest energy photon in the event, E^*_γ, must lie in the interval $0.5\,\mathrm{GeV} < E^*_\gamma < 3.5\,\mathrm{GeV}$ in the CMS frame,

This filter selects only $\sim .023\%$ of the generic B decays, $\sim 1.1\%$ of the *uds* and $\sim .51\%$ of the $c\bar{c}$ samples.

4.2 $B^\pm \to K^\pm\gamma\gamma$ Event Reconstruction

We then find all combinations of one kaon and two photons per event that could possibly be our signal. B candidates are then reconstructed from these signal candidates. Loose cuts are then applied to the reconstructed B candidates to reduce the multiplicity of signal candidates:

- $|\Delta E^*| < 0.5$ GeV,
  where $\Delta E = E^*_{\gamma_{high}} + E^*_{\gamma_{low}} + E^*_K - E^*_{beam}$

  and $E^*_{beam} = 5.29GeV$ is the energy of the beam in the CMS of the e^+e^- collision (or half the mass of the $\Upsilon(4S)$).

- $m_{\mathrm{ES}} > 5$ GeV
  $m_{ES} = \sqrt{E^{*2}_{beam} - (\vec{p^*_{\gamma_{high}}} + \vec{p^*_{\gamma_{low}}} + \vec{p^*_K})}$

After Beta level cuts are applied and the final states are reconstructed, the efficiencies are shown in Table 2 and average multiplicities are shown in Table 3.

Sample	Efficiency (%)
$B^{\pm} \to K^{\pm}\gamma\gamma$	20.8
$c\overline{c}$	5.51×10^{-2}
uds	3.65×10^{-3}
$B^{+}\bar{B^{-}}$	4.68×10^{-4}
$B^{0}\bar{B^{0}}$	2.23×10^{-4}

Table 2: Efficiencies obtained from MC signal and background events after Beta level.

Sample	Multiplicity
$B^{\pm} \to K^{\pm}\gamma\gamma$	1.2
$c\overline{c}$	1.1
uds	1.1
$B^{+}\bar{B^{-}}$	1.2
$B^{0}\bar{B^{0}}$	1.2

Table 3: Multiplicities obtained from MC signal and background events after Beta level.

Variable	Cut
Minimum Transverse momentum	0.1 GeV
Maximum momentum	10 GeV
Minimum number of DCH hits	12
Minimum fit χ^2 probability	0
Maximum DOCA in XY plane	1.5 cm
Minimum Z DOCA	-10 cm
Maximum Z DOCA	10 cm

Table 4: Cuts applied by the *GoodTracksLoose* selection.

4.3 Global Event Selection

We then make cuts based on variable that describe the entire event. For instance, we require that there are at least 2 good tracks per event (good according to the criteria quoted in Table 4).

Additionally, to suppress bhabbas, radiative bhabbas and τ events we require that the second Fox-Wolfram moment be less than 0.9. The Fox-Wolfram moments R_l, $l = 0, 1, 2, \ldots$, are defined by:

$$R_l = \sum_{i,j} \frac{|\mathbf{p}_i|\,|\mathbf{p}_j|}{E_{\rm vis}^2} P_l(\cos\theta_{ij}) \ ,$$

where θ_{ij} is the opening angle between particles i and j, $E_{\rm vis}$ is the total visible energy of the event and $P_l(x)$ are the Legendre polynomials.

4.4 Particle Quality Criteria

After the Beta level reconstruction of the B candidates, we apply quality cuts on the daughters of the B.

4.4.1 Photon selection

Photon quality cuts Photon selection is based on the quality of its signal, determination of which is done by the following cuts:

- photon angle w.r.t the beam-line in the LAB frame, θ_γ, is in the interval: $-0.74 < \cos\theta_\gamma < 0.93$,
- the EMC cluster contains no noisy or dead crystals
- the EMC cluster second moment < 0.002 (Fig. 14)
  (See below for explanation of second moment.)
- 15 to 35 crystals hit in the calorimeter by the higher energy photon, 10 to 30 crystals hit in the calorimeter by the lower energy photon , (Fig. 13)

The second moment is a measure of how circular the shower in the calorimeter is, and a cut on this reduces background from merged π^0's or η's. If the two photons from a π^0 or η decay are merged (nearly parallel so that they can not be resolved by the EMC) and form a cluster, this cluster is more likely to be oval shaped. Clusters from one photon should be spherical which gives a smaller second moment.

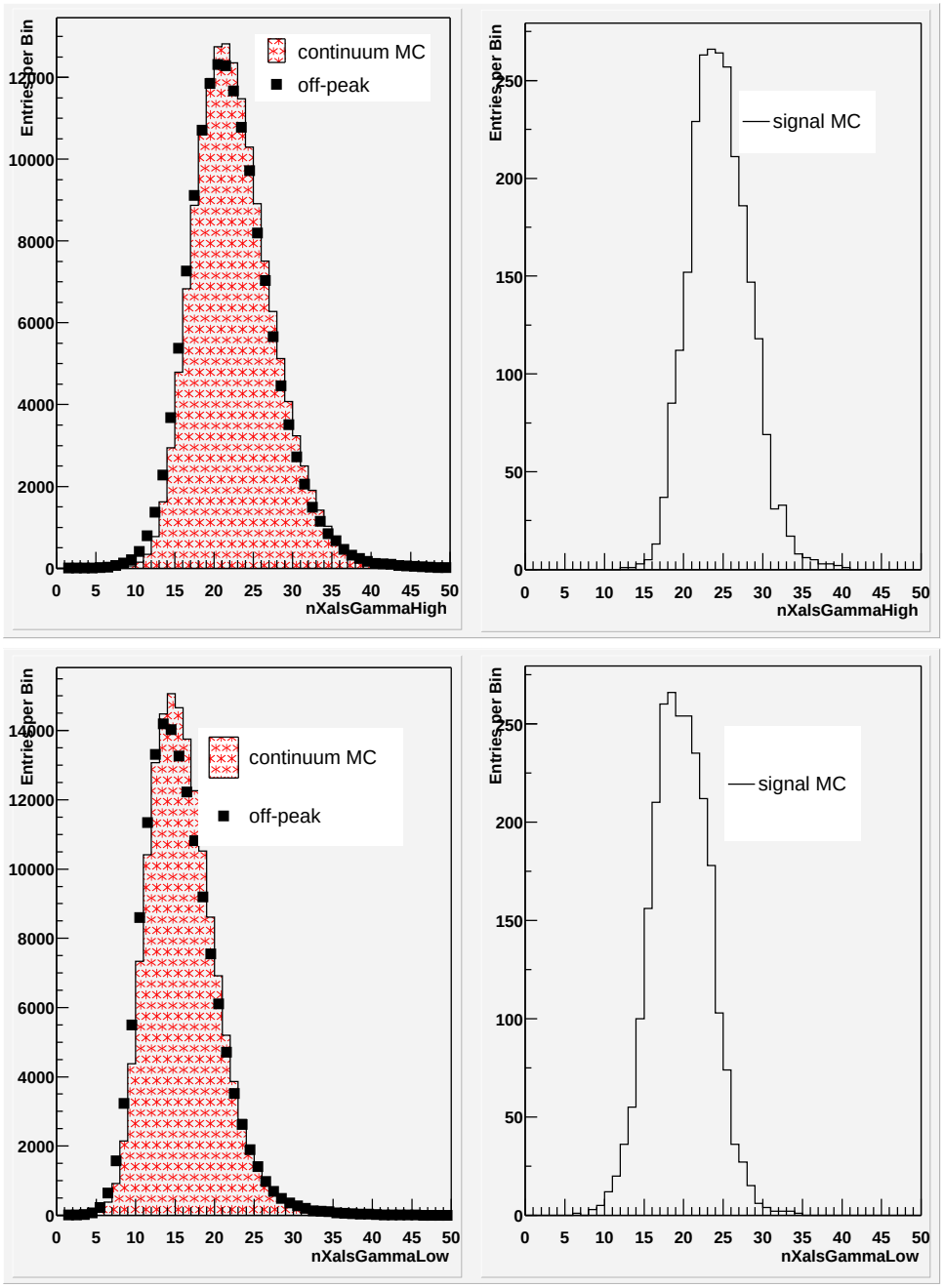


Figure 13: Number of Crystals hit for both photons.

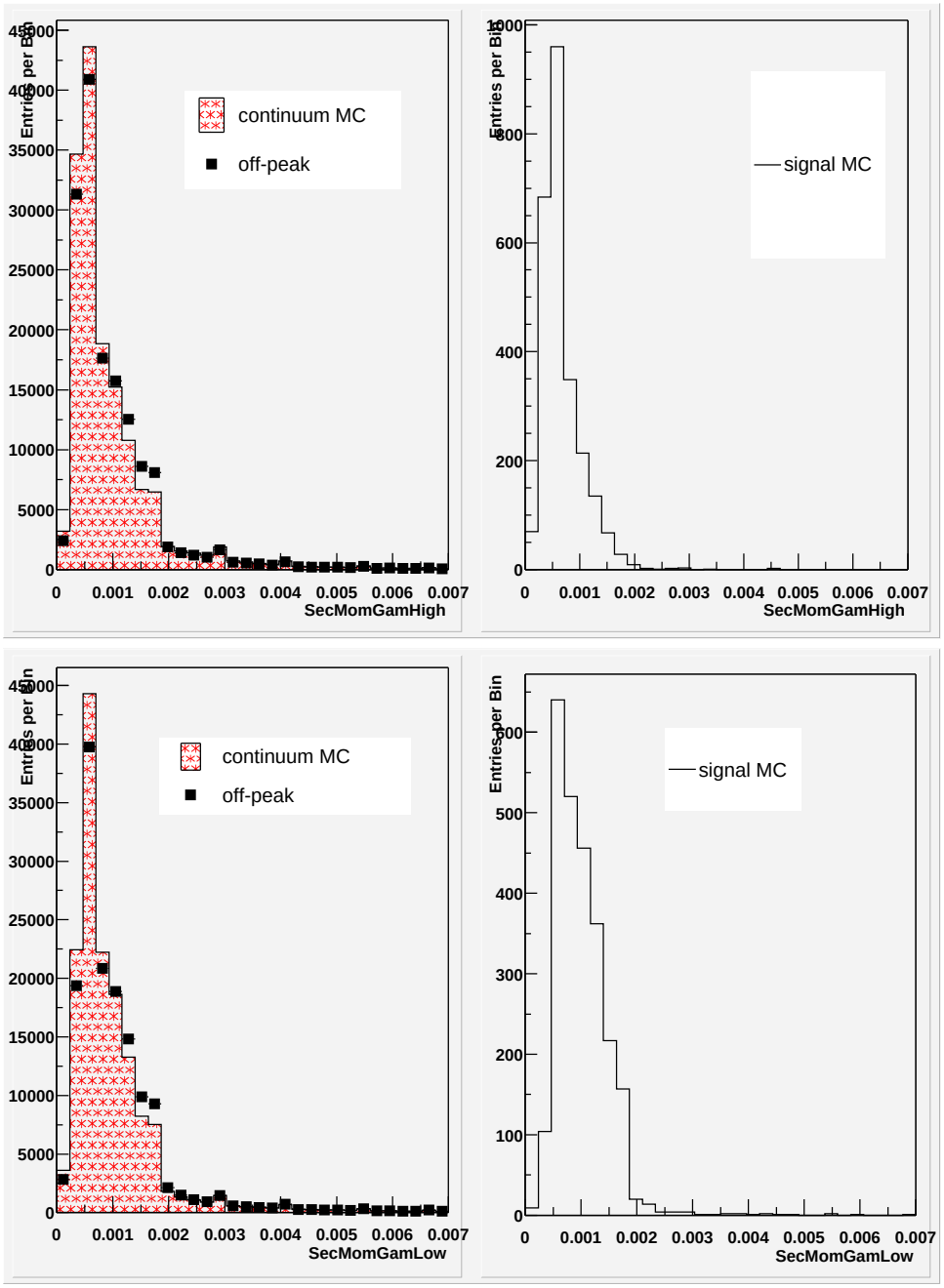


Figure 14: Second moment hit for both photons.

Bump distance cut Furthermore, in order to improve the suppression of background coming from π^0's or η's, we implement an isolation cut on the photon bump: the distance between the photon bump and another bump or track in the event is computed. The corresponding distribution can be seen in Fig. 15. The distance between the photon and any other bump or track on the event is asked to be greater than 25 cm. This cut is done early in the analysis to reduce multiplicity.

π^0 and η vetoes The likelihood that any one photon in an event is the the product of a π^0 or η decay is quite high, so we want to ensure that this is not this case for our photon candidates. First, we must make sure that our two photon candidates did not come from the decay of the same π^0 or η. We can do this by checking that the invariant mass of the mass of the photons combined, $p_{\gamma_1} \cdot p_{\gamma_2}$, does not equal the mass of neither a π^0 nor η. It turns out that this cut is not needed since we are requiring $p_{\gamma_1} \cdot p_{\gamma_2} > 3GeV$.

The other method we employ in determining whether a candidate photon has been produced by a π^0/η decay involves looping through the entire list of photons in an event and combining it with a candidate photon to compute the invariant mass. We cut events where the invariant mass of the candidate photon and some other photon in the event meet the following criteria

- for the π^0 veto (Fig. 16):
  - $|m_{\gamma_{sig}\gamma} - m_{\pi_0}| > 0.03GeV$
  - $E_\gamma > 0.05GeV$ (optimized)
- for the η veto (Fig. 17):
  - $|m_{\gamma_{sig}\gamma} - m_{\eta}| > 0.05GeV$
  - $E_\gamma > 0.3GeV$ (optimized)

4.4.2 $K^\pm$ selection

In addition to the *GoodTracksTight* requirements (Table 4), our kaon candidates must meet the requirements of the *KMicroTight* list which are listed in Table 5. We chose to use the Tight selection as it has a high rejection power and the loss in signal efficiency is acceptable. These criteria are optimized to keep the mis-id rate below 5% up to momenta of 4 GeV/c.

momentum range [GeV/c]	requirements
< 0.7	> 3 SVT hits
< 0.7	> 10 DCH hits
> 0.6	expected number of photons in DIRC > 0 $\theta_c > (0.31p + 0.48)rad$

Table 5: Cuts applied by the *KMicroTight* selection.

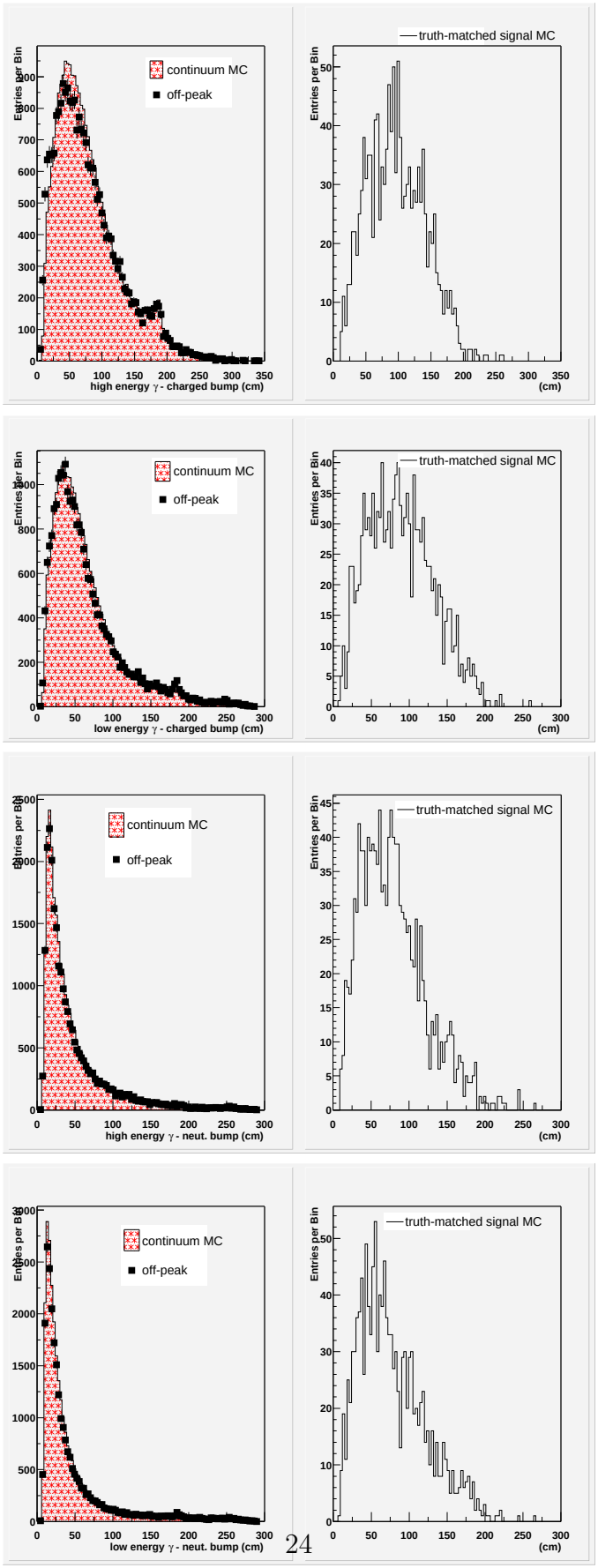


Figure 15: Charged(up two) and neutral(bottom two) bump distance distribution for signal(left) and background MC (right) events.

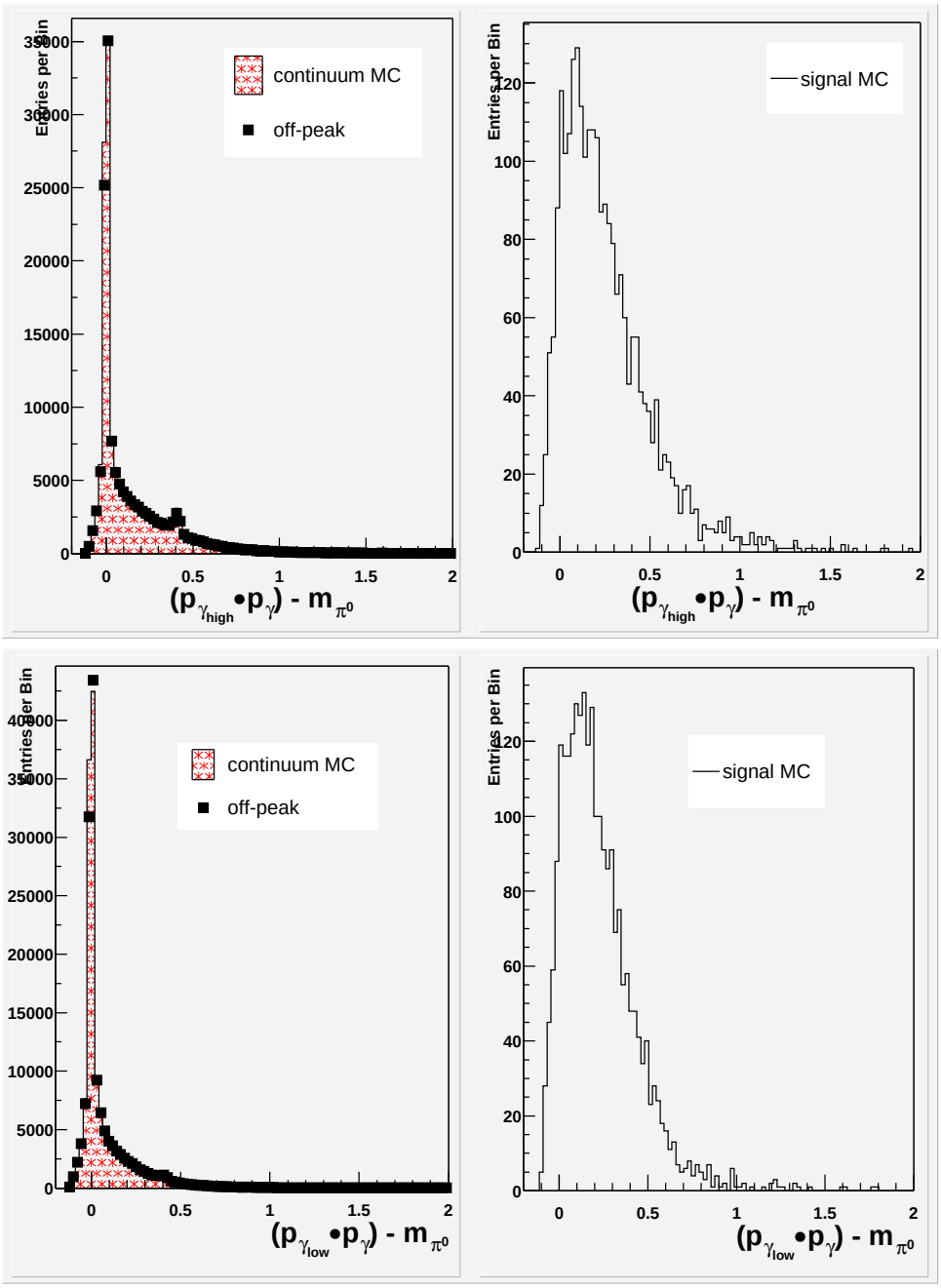


Figure 16: Distributions for $p_{\gamma_1} \cdot p_{\gamma_2} - m_{\pi^0}$ (in Gev) where the second photon minimizes the absolute value of this equation for a single event. The first photon is the high energy photon (top), low energy photon (bottom).

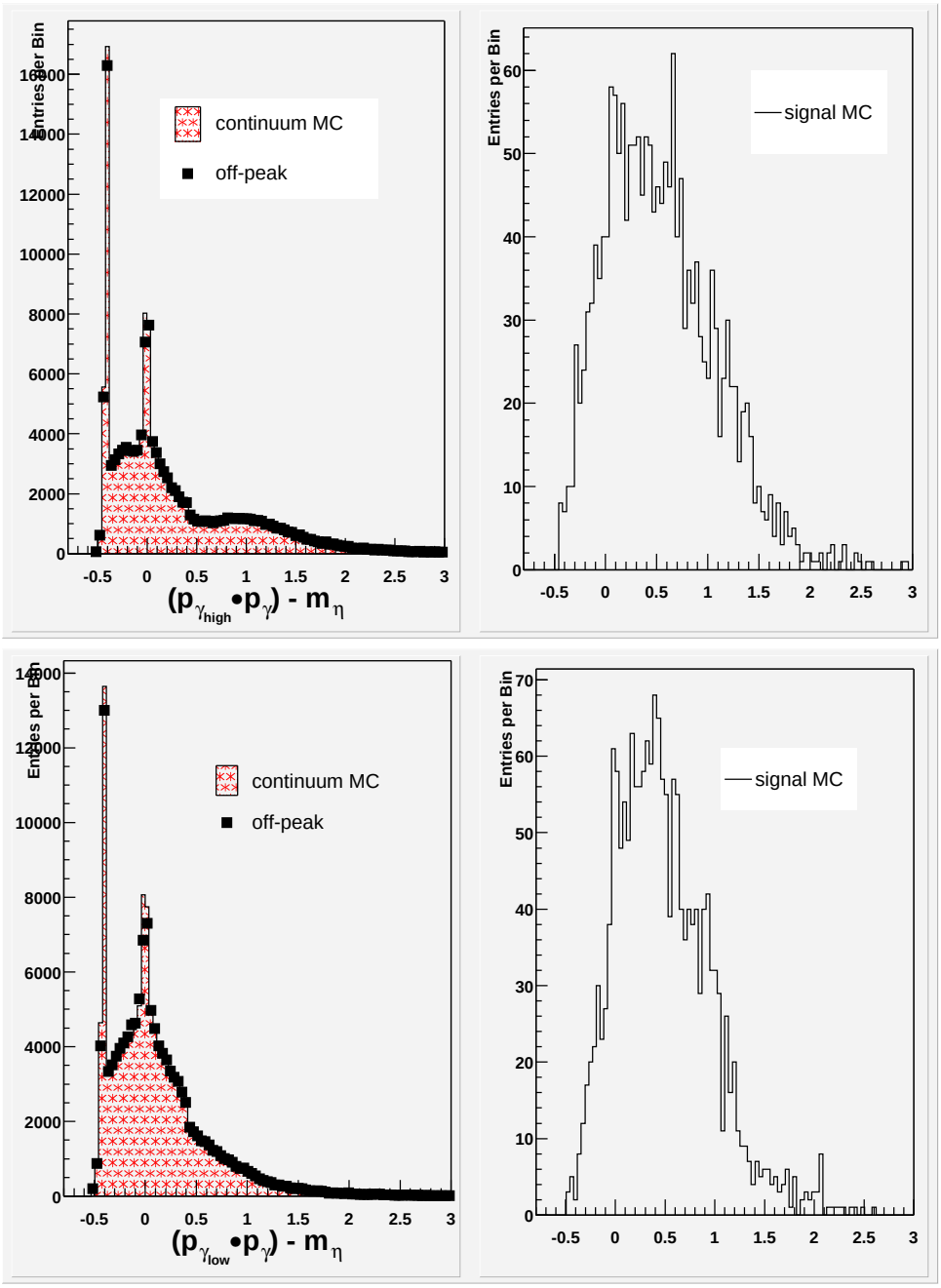


Figure 17: Distributions for $p_{\gamma_1} \cdot p_{\gamma_2} - m_\eta$ (in Gev) where the second photon minimizes the absolute value this equation for a single event. The first photon is the high energy photon (top), low energy photon (bottom).

4.5 Best candidate selection

After each signal candidate passes all the previous requirements there remain events that have more than one possible candidate. At this point, the best candidate per event is chosen to be the one that has the smallest ΔE^* which is defined in Section 4.2.

4.6 Neural Network

After the cuts so far, most of the background is comprised of continuum background in which the photons are coming from either initial state radiation (ISR) or through a π^0/η decay. These backgrounds can be suppressed using event shape variable cuts. The shape variables are described in the following. However, instead of applying a cut for each variable separately, they are combined in a Neural Net (NN) in order to improve the total rejection power.

The NN is a standard supervised learning, feed-forward, back-propagation, neural network [14]. We use the Stuttgart Neural Network Simulator [2], with a developed ROOT-based interface called RooSNNS. We input 24 variables into the NN with 2 hidden layers with 3 and 10 nodes respectively and with one output node to differentiate signal from continuum background. The details and validation are discussed in the next few sections.

4.6.1 Neural Network Basics

Our neural network is used to combine our twenty-four variables (see Section 4.6.3) into a single variable to separate the signal and background processes. The basic structure of a neural network is depicted in Figure 18. At the bottom of the diagram, event variables x_i enter the network as input nodes in red. Linear combinations of these variables are sent up to the hidden nodes (blue): the jth hidden node will receive receive the linear combination:

$$y_j = \sum a_{ij} x_j$$

. Each hidden node receives a different linear combination of the input variables. This input is then transformed by an "activation" function, in this case the tanh function (our x_j and $g(y_j)$ functions). The activation function determines how fast the output of the hidden layer varies as a function of the input. If the input range is small, a linear response is recovered; if it is large, a step response results. At each stage, there is "bias" node (purple) which provides a constant output. The bias node is used to set the zero-level of the output.
A linear combination of the output from the hidden nodes $g(y_j)$

$$z = \sum b_j g(y_j)$$

are then passed to the output node where it is again transformed by an activation function. The mathematical chain from input to output is shown on the right of Figure 18. A neural network can in principle have many hidden layers and output nodes. In general, neural

[2] http://www-ra.informatik.uni-tuebingen.de/SNNS/

network structures can be specified by the number of input nodes (network variables), the number of hidden layers with the number of nodes in each hidden layer, and number of output nodes. Hence the example in Figure 18 can be described as a neural network with four input nodes, one hidden layer with three nodes and a bias node, and one output node.

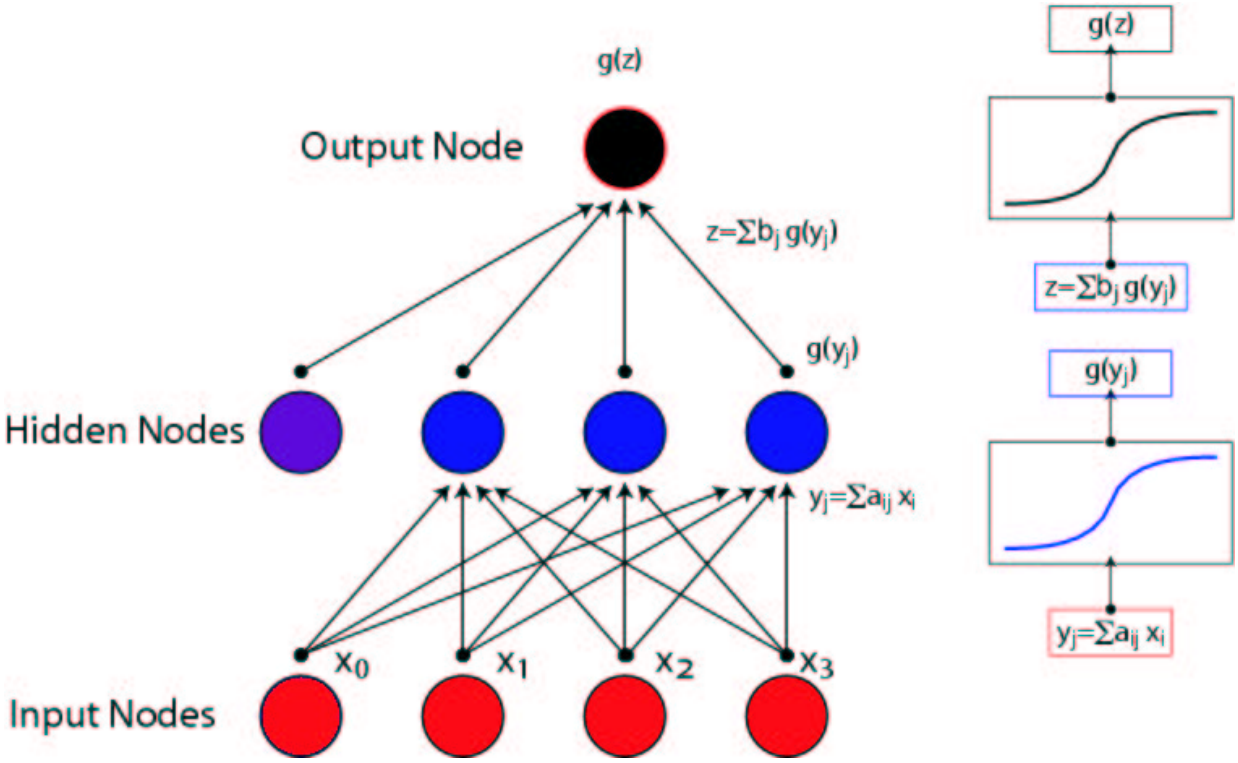


Figure 18: Basic structure of a single hidden layer neural network. Input nodes (red) corresponding to event variables are passed to "hidden" nodes (blue) as linear combinations. The line combination is then transformed by an "activation function," in this case the $\tanh(x)$. Linear combinations of the hidden node outputs are then passed to the output node (black), where it is transformed once again by the activation function to give the final neural network output.

4.6.2 Training the Neural Network

The free parameters of a neural network are the coefficients a_{ij} and b_j that are used to form the linear combination of input nodes to the hidden nodes, and the hidden node outputs to the output node respectively. The optimal set of coefficients are determined by a process known as "back-propagation." The performance of the neural network for any given set of coefficients is summarized by the sum-squared error (SSE):

$$\mathrm{SSE}(a_{ij}, b_j) = \sum_{a=1}^{N} [\mathrm{NN}(\vec{x_a}; a_{ij}, b_j) - F(\vec{x_a})]^2$$

Here, the x_a represents the vector of input variables for the ath event, $\mathrm{NN}(\vec{x_a}; a_{ij}, b_k)$ is the neural network output for this vector with the coefficients a_{ij} and b_j, and $F(\vec{x_a})$ is the desired output for this vector (e.g. 0 if it is a vector corresponding to a background event, 1 if it is a signal event). The SSE then represents a "χ^2" for the network configuration that can be minimized in a manner completely analogous to a fit via gradient descent. This is precisely the back-propagation algorithm: the derivatives of the SSE relative to changes in each of the coefficients are evaluated and the coefficients adjusted accordingly and iteratively to minimize the SSE.

For neural network training, usually two separate data sets are needed, one for the training purpose and another one for validation. The network performance is quantified by the Mean Squared Error (MSE) defined as:

$$MSE = \frac{SSE}{Number\ of\ Events}$$

The MSE is evaluated after each training cycle both for the training sample and the validation sample. Normally we want to stop the training when the MSE for validation sample reaches it's minimum. We use continuum Monte Carlo (including uds + $c\bar{c}$) and the truth-matched signal Monte Carlo data samples for the training and validation. The fractions of continuum Monte Carlo sets are scaled to correspond to the same luminosity.

The results of training can be seen in Fig. 19.

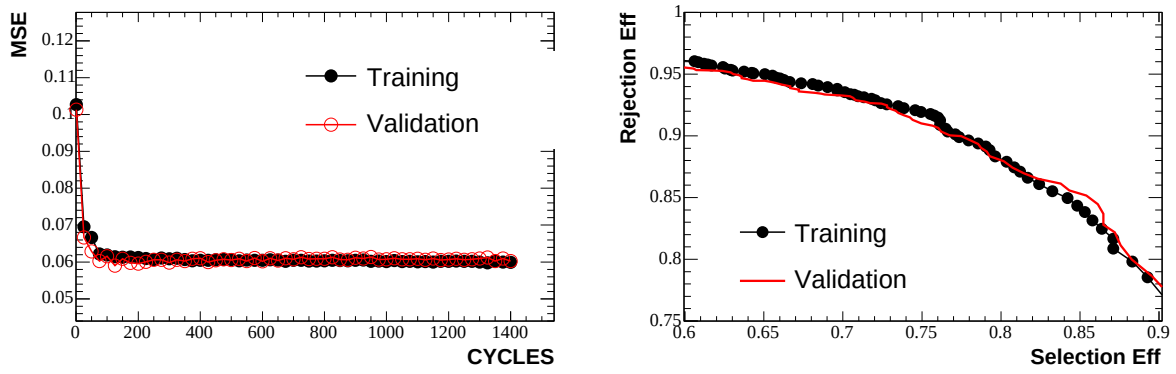


Figure 19: Left: Learning curve shows that MSE has reached minimum and converged for both training and validation samples. Right: Rejection efficiency can reach over 95% but at the price of lower selection efficiency, sample again converge nicely indicating proper training.

4.6.3 Input variables for the NN

The shape variables which are used as input to our NN are:

Thrust $\equiv 2max(\sum \vec{p_L} / \sum |\vec{p}|)$ where p_L is the longitudinal momentum along a unit vector $\vec{n}$, called the thrust axis, and it is summed over all particles.
$|\vec{p}|$ is summed over all particle s.t. $\vec{p} \cdot \vec{n} > 0$.

Sphericity $\equiv (3/2)min(\sum \vec{p_T^2}/\sum \vec{p}^2)$ where p_T is the transverse momentum perpendicular to a unit vector $\vec{n}$ that minimizes the equation and the sums are over all particles of the reaction.

$\cos(\theta_T)$: the angle between the thrust axis formed by the signal particles and the thrust axis formed by all other particles in the event. The thrust axis is the direction which maximizes the sum of the longitudinal momenta of the particles. In the case of a background event, the thrust axis corresponds to the axis along the primary quark-antiquark pair produced from e+e- annihilation. For a true B event, the quark-antiquark pair is typically at rest in the Υ(4S) rest frame so the thrust axis tends to be distributed isotropically. For continuum this is peaked toward one, while for signal it is lower.

$\cos(\theta_B)$: the angle of the reconstructed B candidate with respect to the beam direction computed in the CMS frame. For signal events, $\cos(\theta_B)$ should have a quadratic shape, peaking at zero; for background, it should be flat given the direction of the B.

$R2$: the ratio of second to zeroth Fox-Wolfram moment. A lower R2 is characteristic of the isotropic decay of signal events, as opposed to Bhabha and tau events and continuum background.
(For more thorough description see Sec.4.3.)

$E_\theta(1-18)$: energy flow cones in the CMS frame about the reconstructed B direction, in 10^o increments (18 total). By looking at the energy flow cones we can separate out jetty continuum from the more isotropic/spherical signal events which can be characterized by their energy distribution for the rest of event. We said "rest of the event" because the energy of the reconstructed or signal side of the event is removed from the cones.

Plots of the six shape variables can be found in Figure 20, they are normalized to the same area since relative shape is the important factor.

4.6.4 Training/Optimization of the NN

The training samples used in the NN were created from a reduced data set using all previously discussed cuts in Section 4.4. We use truth–matched signal events (as described in section 3) for signal and uds and $c\overline{c}$ MC for background [3]. If there are more B candidates per event, the one with the smallest ΔE^* is chosen. Total data sets of 1385 signal and 10615 background events are used for the final training sets. Two hidden layers with 3 and 10 hidden nodes is the best configuration for our NN. An increase or decrease in the number of hidden nodes actually shows a loss of signal/background differentiation.

The NN output is shown in Figure 22 for signal and continuum. Each candidate has a NN weight computed for it; 1 is signal like while 0 is continuum background like.

[3]Generic $B\ \overline{B}$ MC was not used as a background training sample. The generic $B\ \overline{B}$ sample introduces biases toward acceptances of lower multiplicity modes if used.

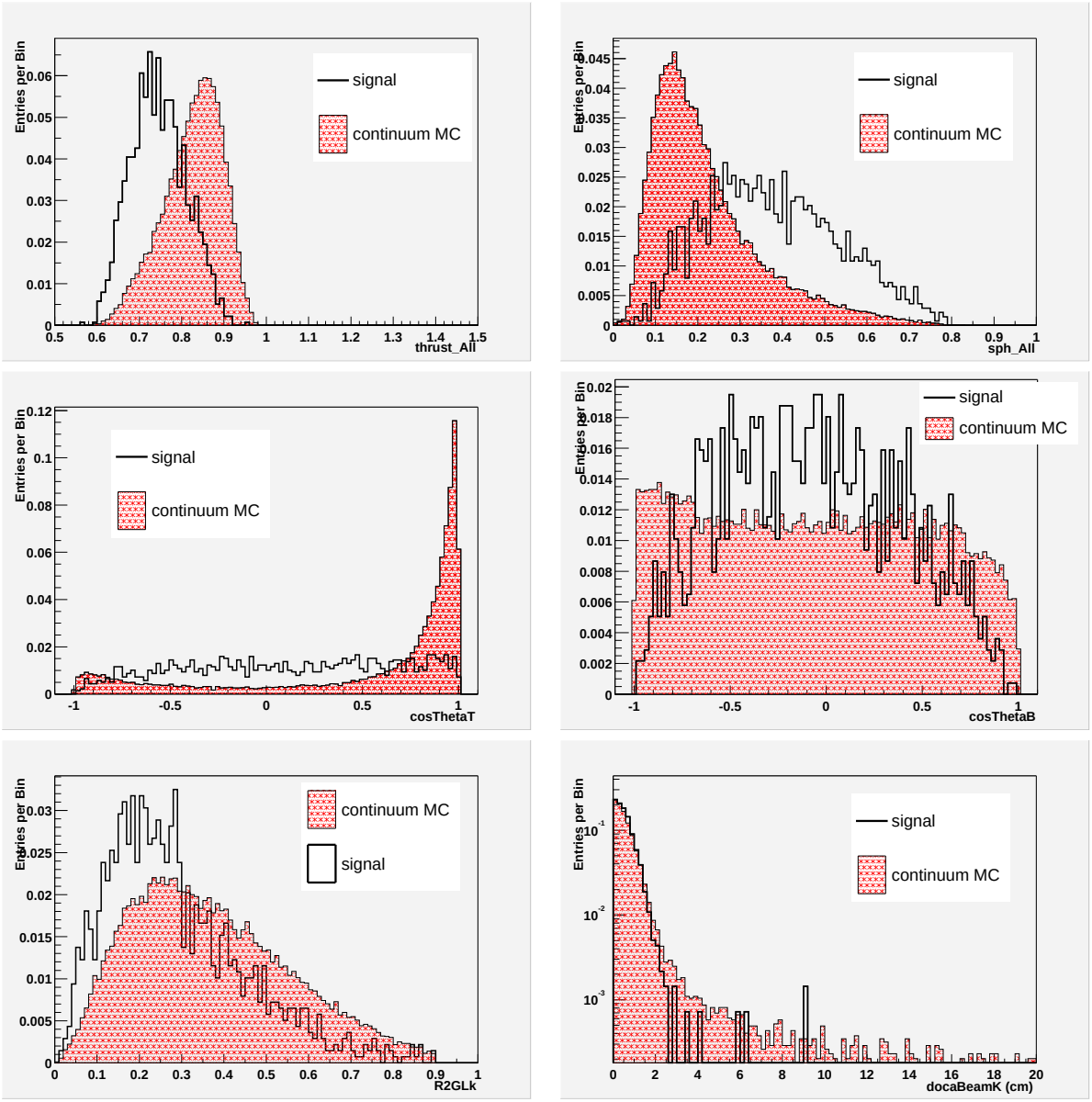


Figure 20: Shape variables going into the NN: $|\cos(thrust)$, $|\cos\theta_{B^*}|$, $|\cos\theta'|$, $W2$, $R2$ and $R2'$. The histograms are all normalized to an area of 1.

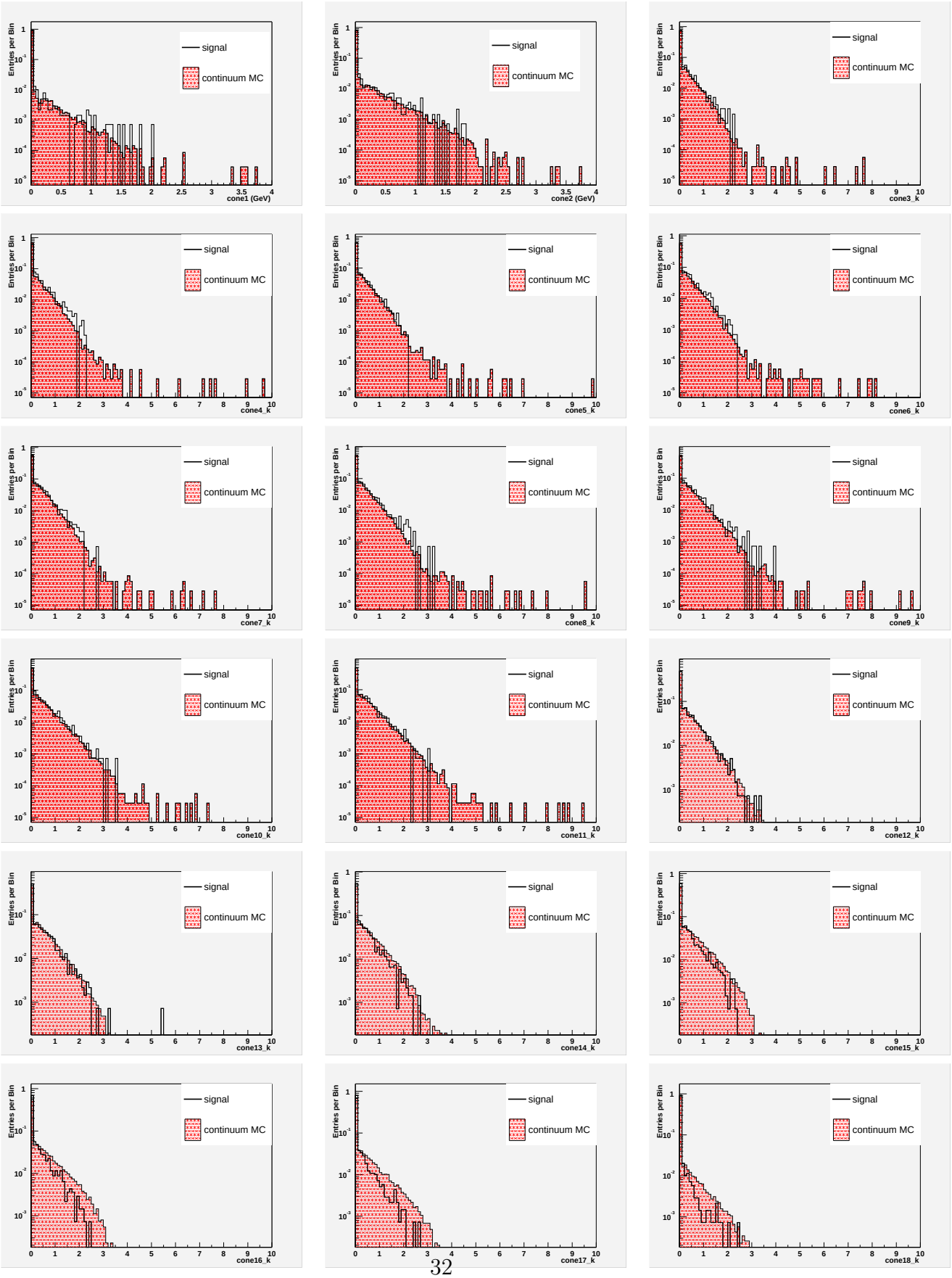


Figure 21: Energy cones going into the NN. The histograms are all normalized to an area of 1.

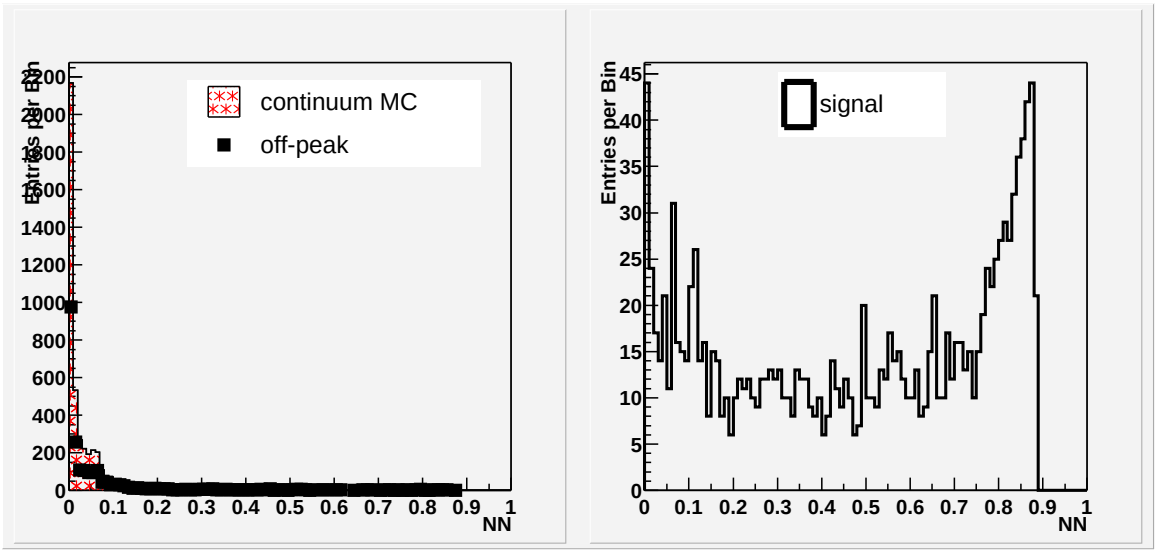


Figure 22: NN output for various MC samples: 1 is signal like, while 0 is continuum background like. The histograms are all normalized to an area of 1.

4.7 Cut Optimization for Neural Network Output

Cuts are performed on the network output to suppress the continuum background. To get the best performance from the NN, the cut value is optimized by acquiring maximum significance, which is defined as $S^2/(S+B)$ where S is the signal yield and B, the background yield. In Figure 23 , it shows significance vs. network cut value. Based on this we have placed a cut at 0.64.

4.8 The Signal Region

After the NN cut is applied we are still left with quite a few events. The last set of cuts placed on the data will be kinematic. Both $DeltaE$ and m_{ES} cuts are first optimized and then placed on the data. $DeltaE$ and m_{ES} were defined in Section 4.2. In Figure 24 you can see the significant amount of background remaining in the signal region.

4.9 Final Efficiencies

After the NN cut is applied, expected number of events for signal and background are tabulated in Table 6. As noted before, our signal is dominated by generic B^+B^- background. Further cuts can possibly eliminate some of this.

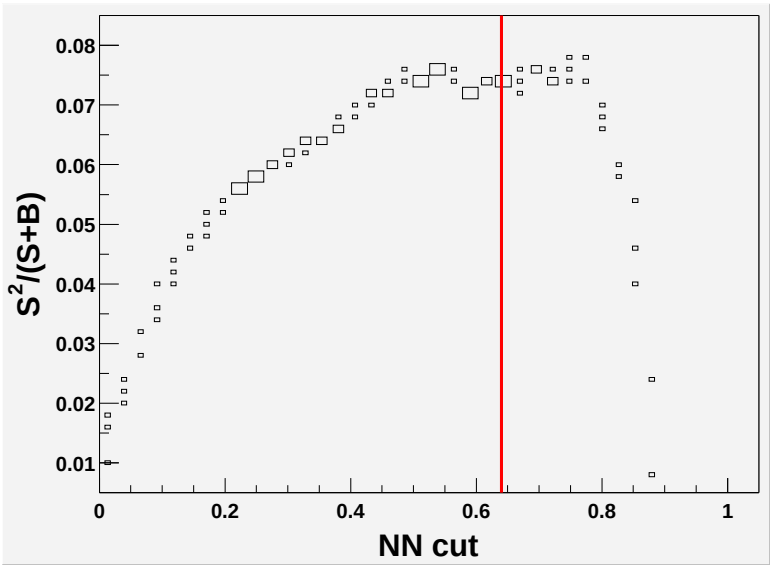
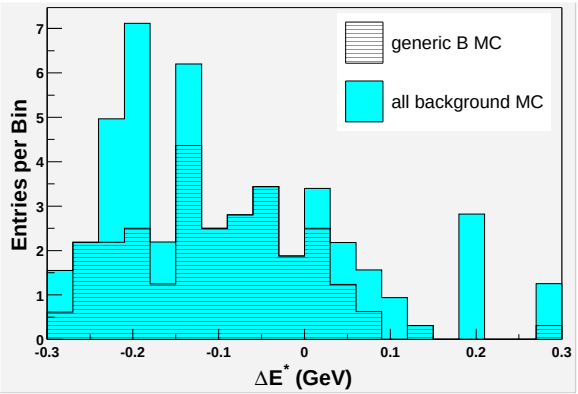


Figure 23: Significance vs cut on NN output.



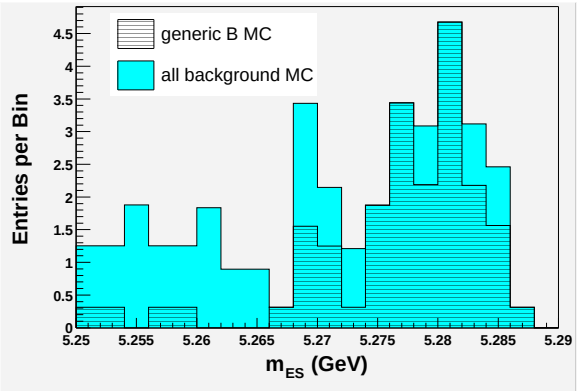


Figure 24:

		in signal region $5.273 < m_{ES} < 5.285$ $-0.12 < \Delta E < 0.08$	
	Raw Yields After NN cut	**Raw Yield**	**Exp. Yield $200 fb^{-1}$**
$B^{\pm} \to K^{\pm}\gamma\gamma$ $(BF = 2 \times 10^{-7})$	556	395	2.48±0.12
$B^0 B^0$	78	2	0.60±0.4
$B^+ B^-$	232	45	14.1±2.1
uds	400	1	0.94±0.9
$c\bar{c}$	196	1	0.90±0.9
Total expected background:			**16.5 ± 2.5**

Table 6: Expected number of events for signal and background.

5 Branching Fraction Upper Limit Extraction

Now that we have expected yields for our expected luminosity, we can use them to put an upper limit on the branching fraction (BF) at the 90% confidence level. In estimating a parameter its is useful to get an idea of its accuracy. By providing a limit, we know that our estimate has a 90% probabaility of being correct. To do this we have used a calculator program [4] that has been developed by the *BABAR* Statistics Working Group. It has been specifically developed to give confidence intervals on BF's for rare decay modes calculated from the number of observed events, the acceptance factor, the background estimate and the associated errors. It follows the approach of Highland and Cousins [5] who have devised a way of incorporating experimental error in the calculation of an upper limit. They have defined the probability of observing n total events to be

$$p_n = ke^{-RS}\sum_{j=0}^{n}(B+RS)^j/j!$$

where the normalization of K depends on B, the number of background events and n through

$$K^{-1} = \sum_{j=0}^{n} B^j/j!$$

and R is the estimated value of the branching fraction and S is the true value of the sensitivity. The probability of $B + RS$ is obviously assumed to have a Poisson distribution.

The derivative of the probability, p_n, is then taken with respect R and averaged over S. The average probability is then set equal to 0.1 to set the confidence level at 90%. The resultant equation is quite complex and can only be solved via numerical methods. The calculator does this complex calculation by generating values of "S" according to a normal distribution with the mean being the Sensitivity and the variance being the error on the Sensitivity, both submitted by the user.

The sensitivity, in our case, can defined as follows,

$$S = 2(\epsilon_s \pm \sigma_{\epsilon_s})L\sigma_{B^+B^-} \rightarrow (1.24 \pm .06) \times 10^7$$

where the error and its statistical fluctuation is $\epsilon_s \pm \sigma_{\epsilon_s} = .059 \pm .003$, the expected luminosity is $L = 200fb^{-1}$ and the $e + e- \rightarrow B^+B^-$ cross section is $\sigma_{B^+B^-} = \frac{1}{2} \times 1.05nb$.

The upper limit at the 90% confidence level is of the form:

$$BF < \frac{N_{UL}^{90}}{2\epsilon_s L\sigma_{B^+B^-}}$$

In accordance with the results of the last section, the number of events seen (signal + background) was set to 19 and the number of background events and corresponding errors set to 16.5 ± 2.5.

The 90% confidence level limits obtained are

$$BF < 8.1 \times 10^{-7}$$

6 Summary

The upper limit set on the branching fraction, at the 90% confidence level, was found to be 8.1×10^{-7}. It should be noted that this study was performed purely on MC data with event selection cuts based on the assumption that the branching fraction of $B^{\pm} \to K^{\pm}\gamma\gamma$ is 2×10^{-7}, a low estimate compared to the predictions of Section 1.2. We are currently investigating techniques to suppress the background coming from $B \to X_{su}\gamma$ decays. Since the current indirect experimental limit on this process is $\mathcal{O}(10^{-4})$, a new *BABAR* search with $200 fb^{-1}$ is warranted. It will be of interest if we measure an excess rate to this process as it will indicate significant new physics contributions or point to need for refinement of SM based calculations.

References

[1] Courtesy of $http://www.cpepweb.org/cpep_s m_l arge.html$

[2] S. R. Choudhury, G. C. Joshi, N. Mahajan & H. J. McKellar, Phys. Rev. D **67**, 074016 (2003)

[3] P. Singer & D. Zhang, Phys. Rev. D **56**, 4274 (1997)

[4] R.Barlow, Comput. Phys. Commun. **149**, 97 (2002)

[5] R. D. Cousins & V.L. Highland Nucl. Instrum. Methods **A320**, 331 (1992)

[6] L.Reina, G, Ricciardi & A. Soni Phys. Lett. B **396**, 231 (1997)

[7] K. Lingel, T. Skwarnicki & J. Smith, hep-ex/9804015

[8] A.Ali Nucl. Phys. B. proc. suppl **59**, 86 ((1997))

[9] L3 Collaboration Phys. Lett. B **363**, 137 (1995)

[10] M. Convery *et al.*, BAD665

[11] D. Bard, T. Beck, F. Di Lodovico, A. Eichenbaum, S. Playfer, Babar Analysis Document 611 "Direct CP Violation in $b \to s\gamma$ Decays".

[12] P. Harrison and H. Quinn ed., "The *BABAR* physics book" physics at an asymmetry B factory", SLAC report 504.

[13] G. C. Fox and S. Wolfram, Nucl. Phys. B **149**, 413 (1970)

[14] Reed & Marks II, Neural Smithing, MIT Press, 1999

[15] M. Mancinelli, S. Spanier, Babar Analysis Document 116 "Kaon selection at the BaBar experiment"

[16] K. Hagiwara *et al.*(Particle Data Group) Phys. Rev. D **66**, 2002 (010001)

[17] F. Halzen & A. Martin, Quarks & Leptons, Wiley 1984

[18] C. V. Chang, G. Lin & Y. Yao Phys. Lett. B **415**, 395 (1997)

[19] *BABAR*Collaboration Nucl. Instrum. Methods A.**479**, 1 (2002)