September 2008
SLAC-R-905

Search for $B \to (\rho/\omega)\gamma$ decays at *BABAR*

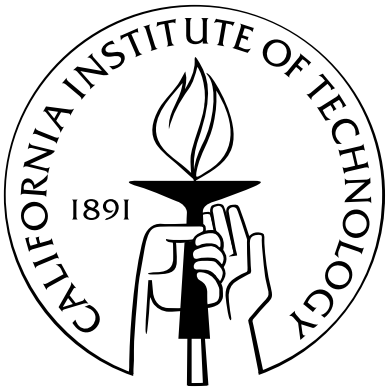
Thesis by

Timofei Piatenko

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2009

(Defended July 28, 2008)

Work supported in part by US Department of Energy contract DE-AC02-76SF00515

To my wife, Rimma, who helped me get through six years of graduate school, and my son, Andrei, whose arrival finally pushed me out of the door.

Acknowledgements

I would like to thank David Cassel and Lawrence Gibbons who got me interested in particle physics while I was an undergraduate at Cornell, Alan Weinstein for introducing me to the Caltech physics department, my adviser David Hitlin for his guidance through the years, Mark Wise and Frank Porter for teaching me particle physics, Ilya Narsky for sharing his deep knowledge of Statistics that made this work possible, Alex Samuel and Alexei Dvoretskii for their senior student wisdom, and all the folks who contributed to this analysis in the past couple of years.

Abstract

We present the results of the search for the decays $B^{0/\pm} \to \rho^{0/\pm}\gamma$ (previously observed) and $B^0 \to \omega\gamma$ (for which currently only an upper limit exists). Together with $B \to K^*\gamma$ decays, $B \to (\rho/\omega)\gamma$ allow us to measure the ratio of CKM-matrix elements $|V_{td}/V_{ts}|$. The analysis is based on the full *BABAR* dataset of 424.35 fb^{-1} corresponding to 465 million $B\overline{B}$ pairs, and makes heavy use of multivariate classification techniques based on decision trees. We find $\mathcal{B}(B^\pm \to \rho^\pm\gamma) = (1.20^{+0.42}_{-0.38} \pm 0.20) \times 10^{-6}$, $\mathcal{B}(B^0 \to \rho^0\gamma) = (0.95^{+0.23}_{-0.21} \pm 0.06) \times 10^{-6}$, and $\mathcal{B}(B^0 \to \omega\gamma) = (0.51^{+0.27}_{-0.24} \pm 0.10) \times 10^{-6}$, where the first error is statistical and the second is systematic. We do not observe a statistically significant signal in the latter channel and set an upper limit at $\mathcal{B}(B^0 \to \omega\gamma) < 0.9 \times 10^{-6}$ (90% C.L.). We also measure the isospin and $\mathrm{SU(3)_F}$ violating quantities $\Gamma(B^+ \to \rho^+\gamma)/2\Gamma(B^0 \to \rho^0\gamma) - 1 = -0.43^{+0.25}_{-0.22} \pm 0.10$ and $\Gamma(B^0 \to \omega\gamma)/\Gamma(B^0 \to \rho^0\gamma) - 1 = -0.49^{+0.30}_{-0.27} \pm 0.10$.

Contents

List of Figures

List of Tables

1 Introduction

At *BABAR*, which is a particle detector located in the PEP-II storage rings at the Stanford Linear Accelerator Center (SLAC), we study weak decays of the B meson by colliding electrons and positrons at the energy equivalent to the mass of the $\Upsilon(4S)$ resonance. Because the latter decays almost entirely into $B\overline{B}$ pairs, and because they are produced nearly at rest in the rest frame of the $\Upsilon(4S)$, we end up with a distinctively isotropic pattern of decay products, in sharp contrast to the jets produced by the hadronization of lighter quarks produced in the continuum. In addition, having two identical B mesons in each decay allows us to tag one of them, using a meson or a lepton decay signature to determine the flavor, and to infer information about the other, signal B.

Furthermore, using different energies for the electron and the positron beams, we introduce a boost between the laboratory frame and the e^+e^- center of mass (CM) frame, which allows us to easily find the separation between the two B mesons in the event. We also rely on the precise knowledge of the beam parameters to reconstruct two kinematic variables

$$\begin{aligned} \Delta E &\equiv E_B^* - E_{\rm beam}^* \\ m_{\rm ES} &\equiv \sqrt{E_{\rm beam}^{*2} - \mathbf{p}_B^{*2}} \end{aligned}$$

where E_B^* and $\mathbf{p}_B^*$ are the energy and momentum of the B candidate measured in the e^+e^- CM frame, while E_{beam}^* is the CM energy of each beam. We use $m_{\rm ES}$ in place of the meson mass because the energy of the beam is more precisely known. For a correctly reconstructed candidate, ΔE should be close to zero, while $m_{\rm ES}$ should be

close the nominal B mass.

While the original purpose of the *BABAR* experiment was to study the phenomenon known as CP violation, possibly responsible for the matter-antimatter asymmetry we observe in nature, as the experiment comes to a close, attention is shifting towards rare decays of the B meson. In this particular analysis, we attempt to reconstruct exclusively some of the rarest decay channels, $B \to (\rho/\omega)\gamma$, which proceed via second-order quantum loop processes known as Radiative Penguins. Apart from providing direct constraints on parameters of the Standard Model (SM) of elementary particle physics, penguin processes are also sensitive to new particles that may appear in the loop and alter the expected decay rates.

As in all typical *BABAR* (and other high-energy physics) analyses, we proceed in several well-defined steps. First, data is collected by the detector, maintained and operated by various university and research laboratory groups that are part of this large collaboration. This raw data is processed to select only the "interesting" events that are most likely to come from B decays. Physical quantities, such as the momenta and energies of the particles, are reconstructed from primary information collected by the various parts of the detector, which comes in the form of voltage and current readings. Then, reconstructed quantities are combined to build a complete picture of an event, where tracks and energy deposits are matched to form particle candidates, which are in turn combined with certain kinematic assumptions to form composite particles that decayed inside the detector.

Once lists of all particles in an event are formed, an individual analyst can use this information to select the events that fall within the topic of the analysis. For example, radiative penguin decays are identified by the presence of a highly energetic photon candidate that originates, along with a light meson, from the primary B candidate. The rest of the analysis relies heavily on statistical techniques to separate as cleanly as possible signal events from background and measure total yields or asymmetries to the highest precision possible, which implies maximizing statistical significance while minimizing the statistical and systematic uncertainties. Mathematically, an analysis consists of event classification, regression, and hypothesis testing.

Our main challenge is separating signal events from a very large amount of background from light quark decays, known as continuum. This includes $e^+e^- \to udsc$ and is often combined with leptonic $e^+e^- \to \tau^+\tau^-$ processes. The former are the main sources of our background, as they produce copious amounts of π^0s and ηs, which decay to $\gamma\gamma$ pairs where one of the photons can be energetic enough to fake our signal signature. Along with continuum events, other B decays, in particular $B \to K^*\gamma$ where a kaon is misidentified as a pion, or $B \to \rho\pi^0$ and $B \to \rho\eta$ decays where one of the photons from π^0 or η decay is missed, also contribute.

Therefore, we rely on the large number of variables that describe our events to carry out multivariate classification and remove much of the background. We then proceed to model the remaining distributions with signal and background components and perform maximum likelihood estimation of model parameters. Finally, we apply hypothesis testing methods, usually based on likelihood ratios, to determine the statistical significance of our measurements. In this analysis, we measure branching fractions, meaning that we are interested in the fraction of the total decay rate represented by our particular channels. Therefore, the measure of significance for us is simply the amount of deviation from zero in the total signal event yield.

Finally, it's important to note that we adopt the blind analysis methodology often used in physics experiments. We rely on precise modeling of the data in Monte Carlo (MC) simulations, which include a detailed physical description of all the detector components as well as all known information on elementary particle kinematics and their interactions with bulk matter. The analysis is carried out entirely on the MC simulated data, until all selection criteria and fit models are finalized.[1] Only then does one "open the box" and perform that final measurement on the actual on-peak data collected by the detector. The intention is to minimize any potential bias in the choice of analysis procedure originating from the desire of the analyst to maximize the significance of the results.

[1] A small amount of off-resonance data collected by the detector at a slightly lower energy is sometimes used for validation studies.

2 Theoretical background and previous results

2.1 Theoretical background

The Standard Model of elementary particle physics attempts to describe the most fundamental constituents of matter and their interaction via the known forces, excluding gravity. Matter is postulated to consist of fermionic particles with fractional spin quantum numbers that interact via exchange of bosons with integer spins. Fermions further divide into two sectors: leptons (electrons, muons, taus, and corresponding neutrinos) and quarks, which we are most interested in at *BABAR*. Quarks are grouped into three generations with an up (positive charge) and a down (negative charge) type quark in each one.

These particles interact via the strong or the electroweak force by exchanging a spin-1 gluon, photon, W, or Z boson. Exchange of the latter two between two quarks allows them to transform between up and down types and among the three generations. This mechanism is summarized concisely in the mathematical formalism of the SM in the form of the Cabbibo-Kobayashi-Maskawa (CKM) matrix [2], which in turn may be represented graphically as a collection of triangles due to the unitarity constraint on its rows and

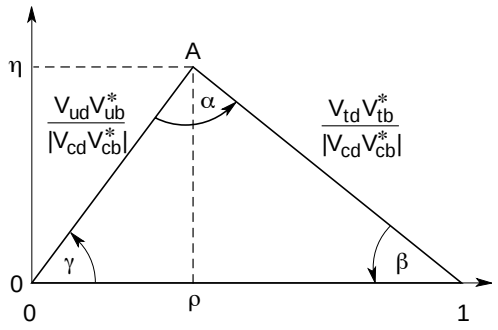


Figure 2.1: CKM triangle relevant to *BABAR* physics

columns. Figure 2.1 shows the CKM triangle most relevant to the physical processes studied at *BABAR*[1].

Within the SM, flavor-changing-neutral-current (FCNC) transitions among same type quarks of different generations are forbidden at tree level (lowest order processes described by Feynman diagrams without loops) because each CKM matrix element matches an up type quark to a down type quark. Thus, the leading order FCNC processes proceed via one-loop electroweak, so-called penguin diagrams where a weak boson is emitted and reabsorbed by a quark. The particular Feynman diagram that describes $B \to (\rho/\omega)\gamma$ decays studied in this analysis is shown in Figure 2.2.

Because the virtual top quark dominates the loop due to its high mass, $b \to d\gamma$ decays are sensitive to the value of the CKM matrix element V_{td}. In the context of elementary particle theories beyond the SM, such as supersymmetry, new virtual particles may appear in the loop, leading to measurable effects on experimental observables such as branching fractions and CP (charge-parity) asymmetries [3].

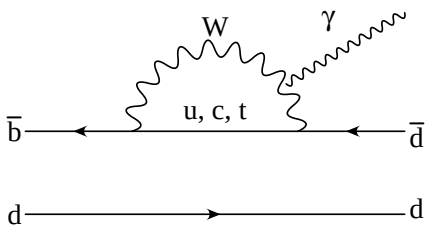


Figure 2.2: Feynman diagram for a $b \to d\gamma$ transition

Branching fraction results from recent next-to-leading order (NLO) SM theoretical calculations for the ρ isospin triplet and the ω singlet are listed in Table 2.1. The uncertainties of these predictions are rather large, mainly due to long-distance non-perturbative QCD effects, unaccounted for in the simplified approach to calculating the decay rates for these processes in the framework of the Heavy Quark Effective Theory (HQET). Normalizing to the branching fraction for the corresponding $b \to s\gamma$ decay, $B \to K^*\gamma$, eliminates some of these uncertainties. This branching fraction ratio is related to the ratio of CKM matrix elements $|V_{td}/V_{ts}|$ via [7]:

$$\frac{\mathcal{B}(B \to (\rho/\omega)\gamma)}{\mathcal{B}(B \to K^*\gamma)} = S_{\rho/\omega}\left|\frac{V_{td}}{V_{ts}}\right|^2 \left(\frac{1 - m_\rho^2/M_B^2}{1 - m_{K^*}^2/M_B^2}\right)^3 \zeta^2[1 + \Delta R]. \qquad (2.1)$$

Here, the isospin factor $S_{\rho/\omega}$ equals 1 and 0.5 for charged and neutral modes, respectively. ζ is the ratio of the ρ and K^* form factors computed using HQET approach, where the top quark and the W boson are integrated out of the effective interaction Hamiltonian. ΔR accounts for differences in the decay dynamics: the weak annihilation contribution for the decay $B^{\pm} \to \rho^{\pm}\gamma$ and the W boson exchange contribution for the neutral channels [4, 6, 8, 9]. It is also worth noting that ΔR is itself dependent on V_{td}/V_{ts} and can be expressed as follows:

$$\begin{aligned}\Delta R &= \left[2\epsilon_A F_1 + \epsilon_A^2(F_1^2 + F_2^2)\right]\left(1 - \frac{2A^{(1)K^*}}{C_7^{(0)\text{eff}}}\right) - \frac{2A^{(1)K^*}}{C_7^{(0)\text{eff}}} \\ &+ \frac{2}{C_7^{(0)\text{eff}}}\text{Re}\left[A_{sp}^{(1)\rho} - A_{sp}^{(1)K^*} + F_1(A^u + \epsilon_A A^{(1)t}) + \epsilon_A(F_1^2 + F_2^2)A^u\right] \quad (2.2)\end{aligned}$$

where $F_1 = -|\lambda_u|\cos\alpha$, $F_2 = -|\lambda_u|\sin\alpha$, $|\lambda_u| \simeq (1 - \lambda^2/2)|V_{ub}/V_{td}|$, $\lambda = V_{us} = 0.2265$, and α is one of the angles of the CKM triangle [4]. C_i are known as Wilson Coefficients and represent strong interaction effects above the scale $\mu \sim m_b$. They arise in HQET as the coefficients for the corresponding four-quark operators of various orders in the effective Hamiltonian that takes a B meson to the final $\rho\gamma$ state. The most general form for the amplitude of the decay can be written in terms of the Fermi Constant, Wilson Coefficients, and CKM factors λ_i as [10]:

$$A(B \to \rho\gamma) = \frac{G_F}{\sqrt{2}}\sum_i \lambda_i C_i(\mu)\langle\rho\gamma|O_i(\mu)|B\rangle. \quad (2.3)$$

From Equation 2.1, we see that radiative B decays provide the same CKM constraint as the ratio of B_d^0 and B_s^0 mixing frequencies, for which a precise measurement became recently available [11]. Comparing $|V_{td}/V_{ts}|$ results obtained with these independent methods provides an important cross check of the SM, as significant inconsistencies would be a sign of new physics processes influencing decays of B mesons and B^0 oscillations in different ways.

The other measurement that comes out of studying the branching ratios for the $B \to (\rho/\omega)\gamma$ transitions (independent of the corresponding $B \to K^*\gamma$ decay modes)

is the isospin and $SU(3)_F$ breaking[2] between the charged and the neutral channels, defined as:

$$\Delta_\rho = \frac{\mathcal{B}(B^\pm \to \rho^\pm\gamma)\tau_{B^0}}{2\mathcal{B}(B^0 \to \rho^0\gamma)\tau_{B^+}} - 1 \quad (2.4)$$

$$\Delta_\omega = \frac{\mathcal{B}(B^0 \to \omega\gamma)}{\mathcal{B}(B^0 \to \rho^0\gamma)} - 1. \quad (2.5)$$

The amount by which these quantities differ from zero may signal evidence for new physics. In particular, for the ratio of the ρ modes, this deviation is expected to come from the weak annihilation contribution in the charged and W-exchange in the neutral channels. In the case of isospin violation between the ρ^0 and the ω modes, the symmetry is broken by the different values of the form factors.

However, these effects are all expected to be rather small, and observing an unexpectedly large deviation implies contribution from Feynman diagrams unaccounted for in the framework of the SM. The caveat to this is that theoretical uncertainties on second-order contributions are rather large, especially where the form factors are involved. Different theorists seem to agree on predictions for Δ_ω and give values around -0.20 ± 0.09 [6]. There is a bit more variation when it comes to Δ_ρ, which depends strongly on the value of the CKM angle γ. Ball, Jones, and Zwicky, for example, provide a range of predictions from -0.05 ± 0.07 at $\gamma = 40^\circ$ to 0.11 ± 0.03 at $\gamma = 70^\circ$ [6].

2.2 Previous experimental results

After earlier searches by CLEO (using 10 million $B\overline{B}$ events) [12] and *BABAR* (using 211 million $B\overline{B}$ events) [13], which had not yielded significant signals, $b \to d\gamma$ transitions were first observed by the Belle collaboration in the decay $B^0 \to \rho^0\gamma$ [14] using a sample of 386 million $B\overline{B}$ pairs.

The latest *BABAR* measurement [15], which was based on 347 million $B\overline{B}$ events, confirmed Belle's $B^0 \to \rho^0\gamma$ observation and found the first evidence of the decay

[2]For the lack of a better name, we choose the convention used by Ali and Parkhomenko[4].

$B^{\pm} \to \rho^{\pm}\gamma$; no significant signal in the $B^0 \to \omega\gamma$ channel was seen, though. These results, which are detailed in Table 2.2, are in good agreement with the latest preliminary Belle measurement (based on a sample of 657 million $B\overline{B}$ pairs) [16] and theoretical predictions (see Figure 2.3).

Using the most recent *BABAR* measurement of the $B \to K^*\gamma$ branching fraction [17] and Equation 2.1 with theory input from [6], the published *BABAR* $B \to (\rho/\omega)\gamma$ branching fraction translates into $|V_{td}/V_{ts}| = 0.200^{+0.021}_{-0.020} \pm 0.015$, where the first error is experimental and the second comes from theoretical uncertainties. As shown in Figure 2.4, this agrees well within errors with the recent B_s mixing results [11] and a global, independent CKM fit [18]. *BABAR* also measured the isospin and $\mathrm{SU(3)_F}$ breaking ratios $\Delta_\rho = -0.35 \pm 0.27$ and $\Delta_\omega = -0.49 \pm 0.63$.

Although experimental measurements seem to indicate some evidence for significant isospin breaking between the charged and the neutral channels, for the purpose of event selection criteria optimizations and estimates of signal yields from simulated data (described in later chapters), we assume $\mathcal{B}(B^0 \to \rho^0\gamma) = \mathcal{B}(B^0 \to \omega\gamma) = 0.5 \times 10^{-6}$ and $\mathcal{B}(B^{\pm} \to \rho^{\pm}\gamma) = 1.0 \times 10^{-6}$ throughout this analysis. This is loosely based on the theoretical predictions by Ball, Jones, and Zwicky [6] as well as Ali and Parkhomenko [4].

	$\mathcal{B}(B^{\pm} \to \rho^{\pm}\gamma)/10^{-6}$	$\mathcal{B}(B^0 \to \rho^0\gamma)/10^{-6}$	$\mathcal{B}(B^0 \to \omega\gamma)/10^{-6}$
Ali *et al* [4]	$1.37 \pm 0.26 \pm 0.09$	$0.65 \pm 0.12 \pm 0.03$	$0.53 \pm 0.12 \pm 0.02$
Bosch *et al* [5]	$1.58^{+0.53}_{-0.46}$	$0.76^{+0.26}_{-0.23}$	–
Ball *et al* [6]	1.16 ± 0.26	0.55 ± 0.13	0.44 ± 0.10

Table 2.1: NLO SM branching fraction predictions. Where two sets of errors are present, the first is theoretical, the second experimental.

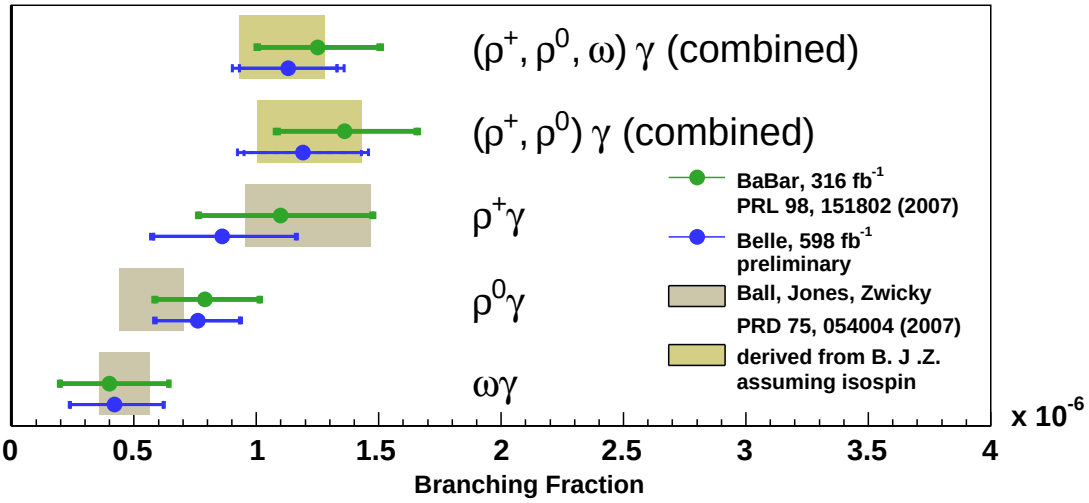


Figure 2.3: Comparison of previous branching fraction measurements with theory predictions

Mode	$n_{\rm sig}$	Σ	$\epsilon(\%)$	$\mathcal{B}(10^{-6})$	Belle $\mathcal{B}(10^{-6})$	Belle Σ
$B^{\pm} \to \rho^{\pm}\gamma$	$42.0^{+14.0}_{-12.7}$	3.8σ	11.0	$1.10^{+0.37}_{-0.33} \pm 0.09$	$0.87^{+0.29+0.09}_{-0.27-0.11}$	3.3σ
$B^0 \to \rho^0\gamma$	$38.7^{+10.6}_{-9.8}$	4.9σ	14.1	$0.79^{+0.22}_{-0.20} \pm 0.06$	$0.78^{+0.17+0.09}_{-0.16-0.10}$	5.0σ
$B^0 \to \omega\gamma$	$11.0^{+6.7}_{-5.6}$	2.2σ	7.9	$0.40^{+0.24}_{-0.20} \pm 0.05$	$0.40^{+0.19}_{-0.17} \pm 0.13$	2.6σ
$B \to (\rho/\omega)\gamma$		6.4σ		$1.25^{+0.25}_{-0.24} \pm 0.09$	$1.14 \pm 0.20^{+0.10}_{-0.12}$	6.2σ
$B \to \rho\gamma$		6.0σ		$1.36^{+0.29}_{-0.27} \pm 0.10$	$1.21^{+0.24}_{-0.22} \pm 0.12$	5.8σ

Table 2.2: The signal yield ($n_{\rm sig}$), significance (Σ) including systematic errors, efficiency (ϵ), and branching fraction ($\mathcal{B}$) for each mode. The second set of errors for ($\mathcal{B}$) are systematic. Branching fraction results from Belle [16] are included for comparison.

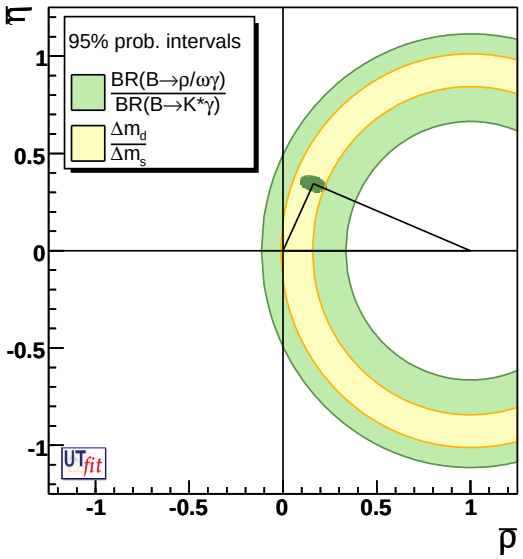


Figure 2.4: Contraints on $|V_{td}/V_{ts}|$ from the *BABAR* run 1–5 measurement of $B \to (\rho/\omega)\gamma$ (dark band), CDF B_s mixing results (light band), and a global CKM fit (ellipse).

3 PEP-II and the *BABAR* detector

3.1 PEP-II

PEP-II [23] is an asymmetric energy e^+e^- collider that uses separate storage rings for electrons and positrons before bringing them together at the energy of $\sqrt{s} = 10.58$ GeV, corresponding to the peak production of the $b\bar{b}$ resonance $\Upsilon(4s)$ with the cross-section of $\sigma(e^+e^- \to \Upsilon(4s)) = 1.05$ nb. The high-energy ring (HER) stores electrons with the energy of 9.0 GeV, while the low-energy ring (LER) holds 3.1 GeV positrons. Both rings are located in the PEP tunnel with circumference of 2.2 km. Figure 3.5 shows a schematic representation of PEP-II.

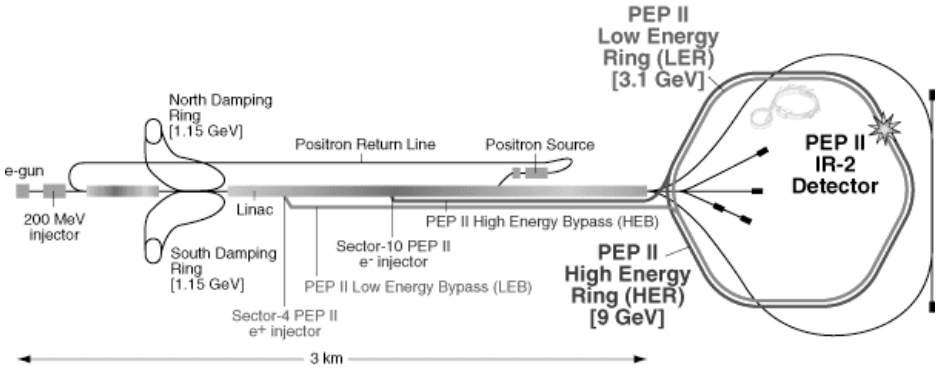


Figure 3.5: Schematic representation of the linear accelerator and the PEP-II storage rings

Apart from the desired $e^+e^- \to \Upsilon(4s)$ reaction, processes that take place in the interaction region include elastic Bhabha scattering with $\sigma(e^+e^- \to e^+e^-) \approx 40$ nb, muon production with $\sigma(e^+e^- \to \mu^+\mu^-) = 1.16$ nb, tau production with $\sigma(e^+e^- \to \tau^+\tau^-) = 0.94$ nb, and light quark production with $\sigma(e^+e^- \to u\bar{u}) = 1.39$

nb, $\sigma(e^+e^- \to d\bar{d}) = 0.35$ nb, $\sigma(e^+e^- \to s\bar{s}) = 0.35$ nb, and $\sigma(e^+e^- \to c\bar{c}) = 1.30$ nb. To study these processes with real data, PEP-II is run 40 MeV below the $\Upsilon(4s)$ resonance for a fraction of the operating time. This off-peak data can be used, for example, to validate Monte Carlo simulations used for event selection optimization.

The asymmetric nature of e^+e^- collisions at PEP-II results in a boost of $\beta\gamma = 0.56$ between the center-of-mass (CM) and the laboratory frames. This translates to a typical separation between the two B meson vertices of about 250 μm, within the resolution of the *BABAR* silicon vertex tracker. This separation distance is a good discriminator between signal $B\overline{B}$ events and background light quark (or continuum) decays.

The design luminosity for the PEP-II collider was 3.0×10^{33} cm^{-2}s^{-1}. However, the actual performance far exceeded this target, as the record peak luminosity reached 12.1×10^{33} cm^{-2}s^{-1} on August 16, 2006. Over the lifetime of the experiment, PEP-II delivered 557 fb^{-1} of data, of which 531 fb^{-1} were successfully recorded by the *BABAR* detector.

3.2 The *BABAR* detector

The *BABAR* detector is a multi-system particle detector operating in the PEP-II collider interaction region. In order of increasing distance from the interaction point where electron and positron beams are brought head-on, it consists of the silicon vertex tracker (SVT), the drift chamber (DCH), the detector of internally reflected Cherenkov light (DIRC), the CsI(Tl) electromagnetic calorimeter (EMC), a 1.5 T superconducting magnet, and the instrumented flux return (IFR). Detector control and monitoring, event triggering, and data readout are performed by the data acquisition system (DAQ).

The origin of the *BABAR* coordinate system coincides with the nominal interaction point. The z axis is pointing in the direction of the electron beam along the beam line. The y axis points upward, while the x axis points horizontally outward from the center of the storage rings.

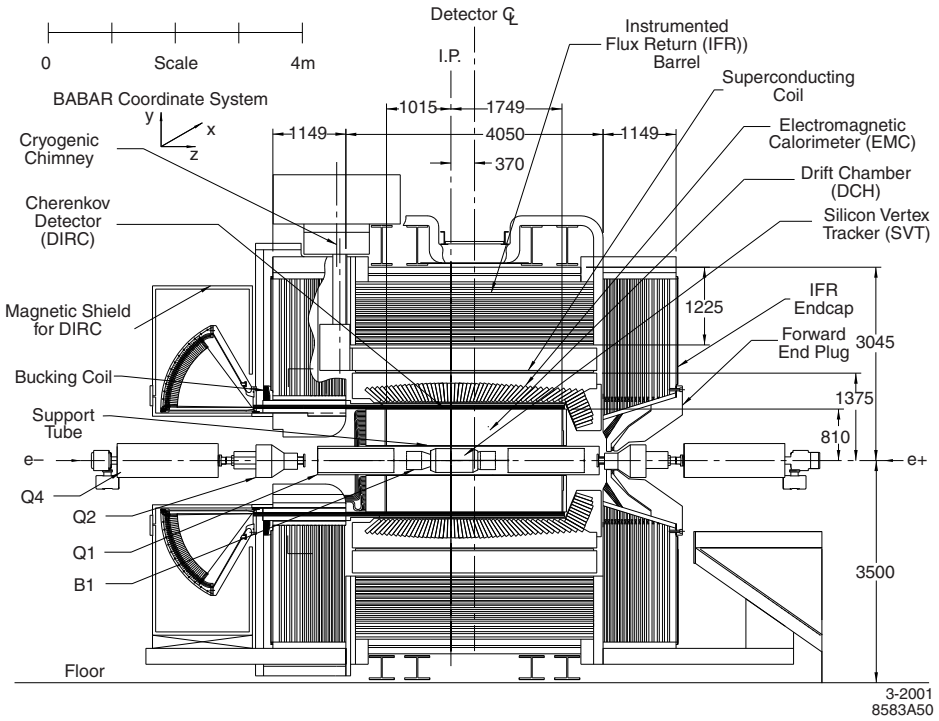


Figure 3.6: Schematic drawing of the *BABAR* detector: side view

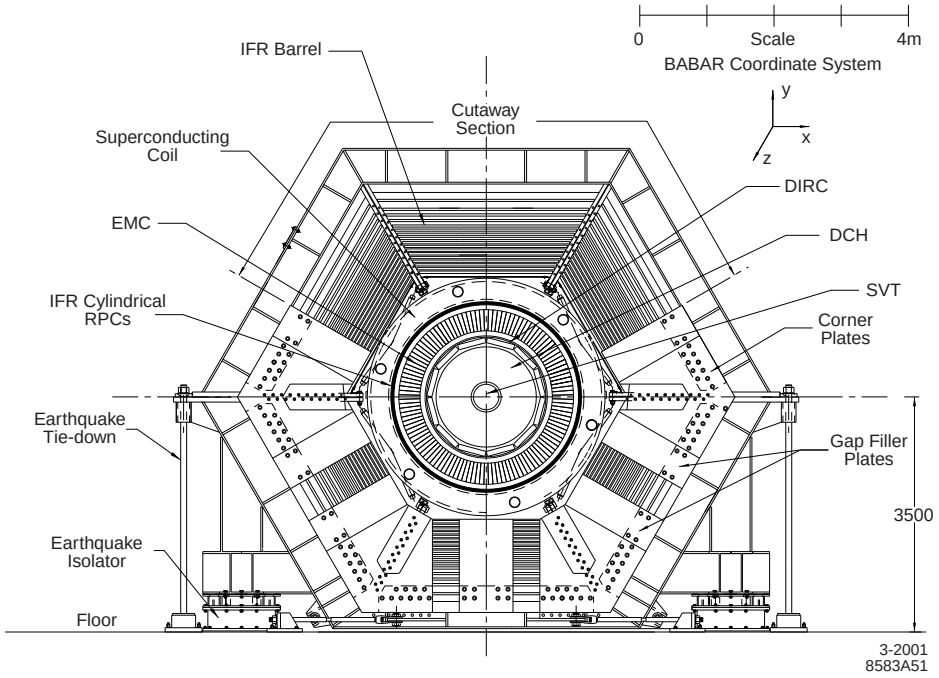


Figure 3.7: Schematic drawing of the *BABAR* detector: front view

3.3 The silicon vertex tracker

The SVT is the innermost component of the *BABAR* detector responsible for precise vertexing and, in combination with the drift chamber, charged particle tracking and identification through dE/dx measurements. Many of the decay products of a B meson have low transverse momentum, and the SVT provides standalone tracking for particles with $p_t < 120$ MeV/c, which is the cutoff for a reliable measurement using the DCH.

The tracker consists of five unequally-spaced layers of double-sided silicon microstrip sensors, where strips on one side are oriented parallel to the beam to provide angular information, while strips on the other side are oriented perpendicular to the beam to measure z. Each sensor is 300 μm thick. The innermost layer is only 33 mm from the interaction point, while the outermost layer is at the distance of 146 mm. Figure 3.9 demonstrates the layout and dimensions of the SVT as well as its angular coverage, which is constrained by the final beam focusing magnets located just outside the tracker. Figure 3.8 shows the layout of the five SVT layers from a cross-section perspective. The segments overlap slightly to provide better coverage at the edges.

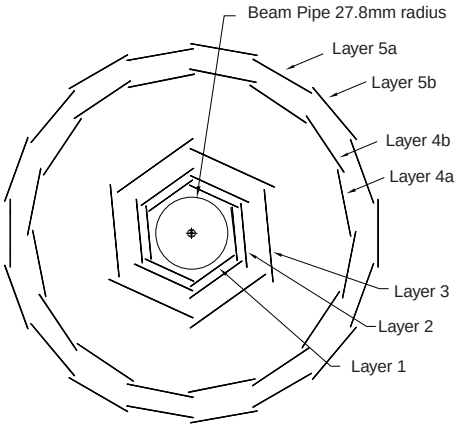


Figure 3.8: The configuration of the SVT layers

The helical trajectory of a charged track in a magnetic field is characterized by the following five parameters: the distance of closest approach to the z axis, d_0; the corresponding azimuthal angle, ϕ_0; the corresponding distance along the z axis, z_0; the dip angle, λ, of the helix; and the curvature of the track, $\kappa \sim 1/p_t$. The precision of the first four of these is usually dominated by the uncertainties on the SVT measurements: $\sigma_{d_0} = 23$ μm, $\sigma_{\phi_0} = 0.43$ mrad, $\sigma_{z_0} = 29$ μm, and $\sigma_{\tan\lambda} = 0.53 \times 10^{-3}$.

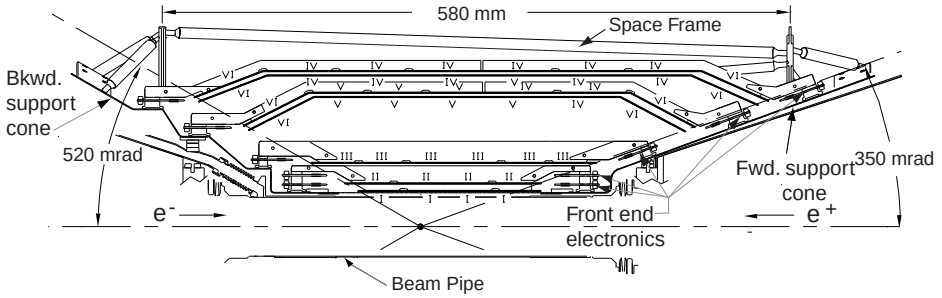


Figure 3.9: Schematic drawing of the silicon vertex tracker: side view

3.4 The drift chamber

The DCH consists of 40 layers (10 superlayers) and 7104 hexagonal drift cells. It is designed to provide precise charged particle tracking, such that a particle with $p_t > 180$ MeV/c will be recorded in all the 40 layers. For particles with momenta of less than 700 MeV/c, the DCH provides dE/dx measurements necessary for particle identification.

As shown in Figure 3.10, the subsystem is offset from the interaction point by 370 mm in the direction of the electron beam to improve forward coverage based on the PEP-II boost. Superlayers 1, 4, 7, and 10 are axial, with wires parallel to the z axis. The wires in the other superlayers are positioned at angles of 45–76 mrad, alternating in sign, to enable z coordinate measurements.

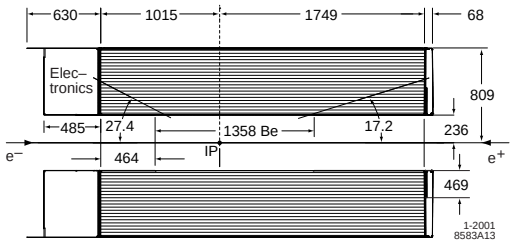


Figure 3.10: Schematic drawing of the DCH: side view

The DCH is filled with a 4:1 mixture of helium and isobutane gases. An incident particle ionizes the gas, and the resulting free electrons are accelerated toward the high-voltage sensor wires. Colliding with other molecules along the way, these elec-

trons produce further ionization electrons, resulting in an avalanche effect. The time of arrival of this avalanche determines the distance of closest approach of the particle to the wire, while the integrated charge provides dE/dx information with resolution of 7.5%. The DCH dominates the precision of momentum measurement for most tracks, parametrized by

$$\sigma_{p_t}/p_t = (0.13 \pm 0.01)\% \cdot p_t + (0.45 \pm 0.03)\%.$$

3.5 The detector of Cherenkov light

The DRC is a particle identification system built on the principles that a charged particle traveling faster than the speed of light in a refractive medium ($v > c/n$) emits a cone of radiation, and that the magnitudes of angles are preserved upon reflection from a flat surface. This radiation is called Cherenkov light, governed by $\cos\theta_c = 1/n\beta$, where θ_c is the opening angle.

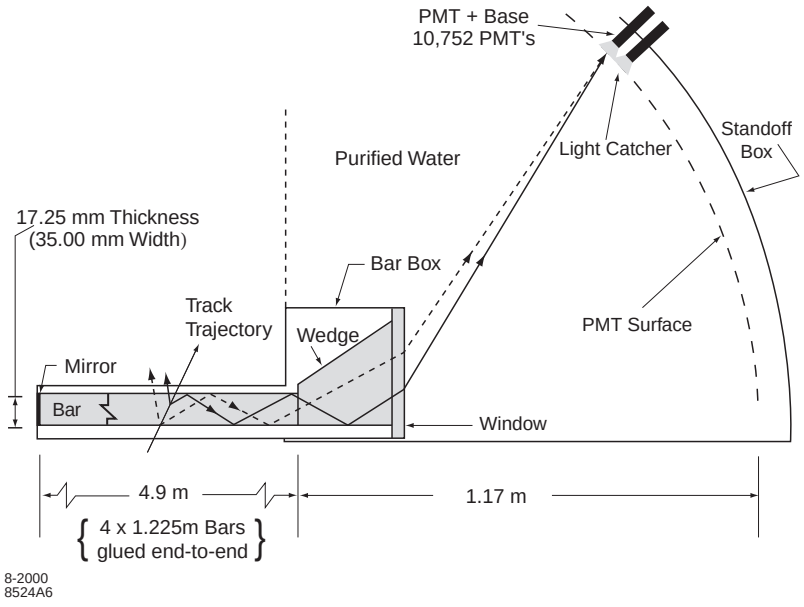


Figure 3.11: Schematic drawing of the Cherenkov light detector: side view

The detector (shown in Figure 3.11) consists of 144 synthetic quartz bars that trap Cherenkov light via total internal reflection. Photons that make it all the way

down to the front of a bar are reflected back by a mirror. On the back end, the photons are allowed to propagate into a standoff box filled with 6000 liters of purified water. Photons travel through the water, where the refractive index matches that of the quartz, preserving the conical pattern and the Cherenkov angle.

The final image is detected by an array of 10,752 photomultiplier tubes (PMTs), and the arrival time is used to associate hits with charged tracks reconstructed in the DCH. The timing resolution measured with $\mu^+\mu^-$ events is 1.7 ns, only slightly worse than the intrinsic PMT spread of 1.5 ns. The corresponding angular resolution is 2.5 mrad. The DRC allows for a 4.2σ separation between kaons and pions at 3 GeV/c. This is particularly important in this analysis, where the signal modes produce only pions in the final state, while backgrounds from kinematically similar $B \to K^*\gamma$ decays produce pairs of kaons and pions.

3.6 The electromagnetic calorimeter

The EMC is the most important component of *BABAR* hardware for this analysis, since it is responsible for detecting photons, which include the primary energetic photon from $B \to (\rho/\omega)\gamma$ decay, as well as the softer photon pairs coming from π^0 and η decays. Thus, the EMC is essential both for detecting signal events and for distinguishing them from backgrounds.

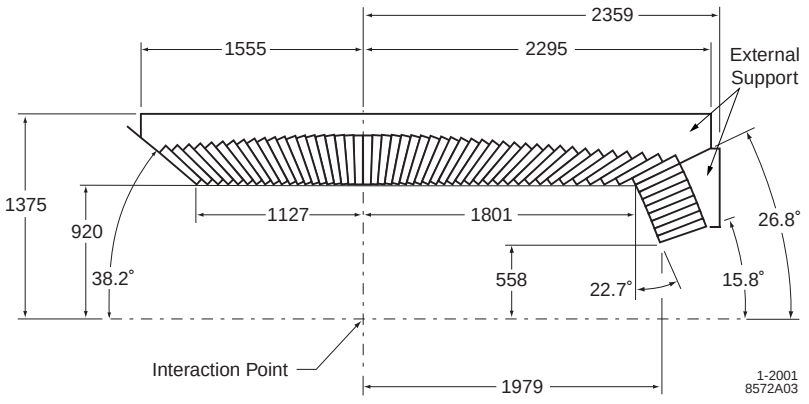


Figure 3.12: Schematic drawing of the electromagnetic calorimeter: side view

The EMC is a total absorption calorimeter composed of 6580 thallium-doped CsI crystals, arranged in a central barrel and a forward endcap (illustrated in Figure 3.12), all pointing toward the interaction point. The crystals are 4.7×4.7 cm in cross-section and vary in length from 29.6 cm (16 radiation lengths) to 32.4 cm (17.5 radiation lengths). Each crystal has a thin reflective coating to contain the scintillation light produced by an incident particle. Two silicon photodiodes with quantum efficiency of 85% read out the signal from the back of each crystal (see Figure 3.13).

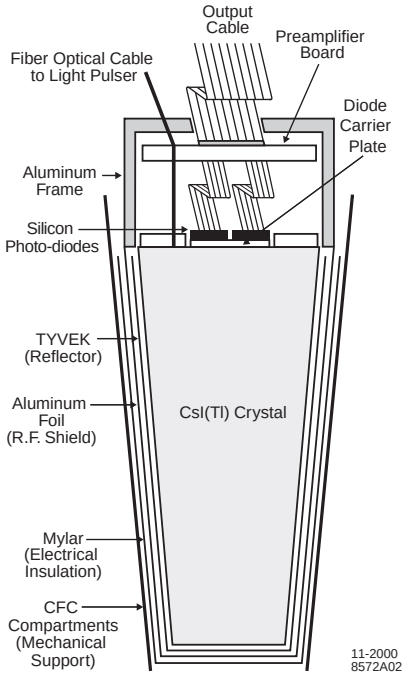


Figure 3.13: Schematic drawing of an EMC crystal

The calorimeter is calibrated using 6.13 MeV gamma rays from activated flourinet fluid, as well as with $e^+e^- \rightarrow e^+e^-$ Bhabha scattering events that produce 3–9 GeV clusters for high-energy calibration. The energy and angular resolutions are parametrized as follows

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

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

3.7 The instrumented flux return

The IFR serves as a flux return for the 1.5 T magnetic field and as a support structure for the entire *BABAR* detector. It's also instrumented to detect muons, K_L, and neutrons, none of which appear in this analysis. It is a hexagonal structure consisting of a barrel and two endcaps, as shown in Figure 3.14.

The original IFR design made use of resistive plate chambers (RPCs), which

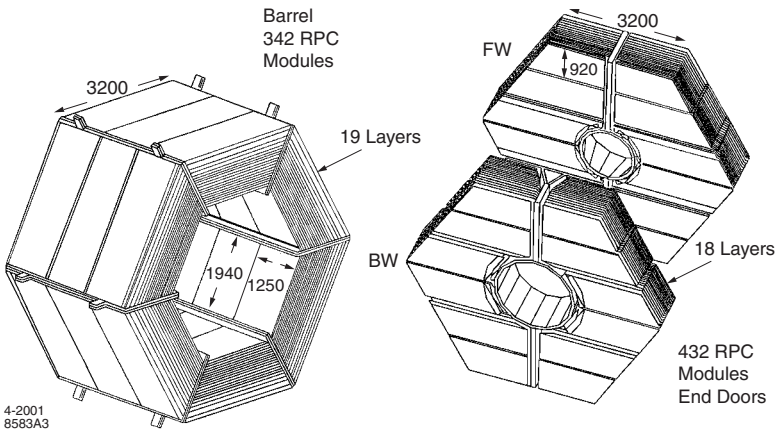


Figure 3.14: Schematic drawing of the instrumented flux return

consist of an argon-freon-isobutane gas mixture sandwiched between two layers of graphite-coated bakelite, foam, and aluminum strips. An ionizing particle induces charge and changes the capacitance of the system, which can be read out by electronics. However, the performance of the RPCs slowly degraded over the years, prompting a replacement technology based on limited streamer tubes (LSTs) to be installed in 2004 and 2006.

The LSTs are long and narrow rectangular cells ($1.5 \times 1.7 \times 400$ cm) of graphite-coated PVC and a silver-plated wire. Each cell is filled with a CO_2-isobutane-argon gas mixture. These detectors work on the principle of an electromagnetic avalanche produced by an ionizing particle traversing the gas-filled gap between the anode and the cathode. A "streamer" is the result of a cascade of secondary avalanches that occur after the primary avalanche grows enough to cancel out the external electric field and saturates.

3.8 Trigger and data acquisition

The purpose of the *BABAR* trigger system is to pick out the events of interest and reduce the logging rate to a reasonable value that can be handled by modern electronic systems. The trigger consists of a hardware part (Level 1) and a software component

(Level 3, for historical reasons). The Level 1 trigger combines three sub-triggers from the DCH, EMC, and IFR and feeds the information to a global trigger (GLT). The drift chamber trigger uses both the timing and z coordinate information to reject background events. The GLT matches calorimeter clusters to drift chamber tracks for physics triggering, while also using IFR information for diagnostic triggering on $\mu^+\mu^-$ pairs and cosmic rays. The overall rate for Level 1 trigger is about 2.5 kHz at luminosity of 8×10^{33} cm^{-2}s^{-1}.

The Level 3 trigger operates on an online Linux computer farm comprised of 28 Dell 1650 (Dual Pentium-III 1.4 GHz) capable of processing an individual event in about 4 ms. Here, quick reconstruction of DCH helices and EMC clusters is performed, and events are selected for logging. These events are then further grouped into runs and written to tape for long-term storage. Event rates are further reduced by about a factor of 10 by the Level 3 trigger, which also vetoes Bhabha scattering events.

Apart from hosting the Level 3 trigger, the *BABAR* online event processing system is also responsible for general monitoring tasks related to data acquisition, real-time data quality monitoring, detector calibrations performed regularly during normal detector operation, and running an event display that provides visual event reconstruction information.

Figure 3.15 shows a flow chart that illustrates Level 1 trigger operations and provides the design numbers for transmission rates among the various components. Figure 3.16 shows a very general overview of the *BABAR* online system.

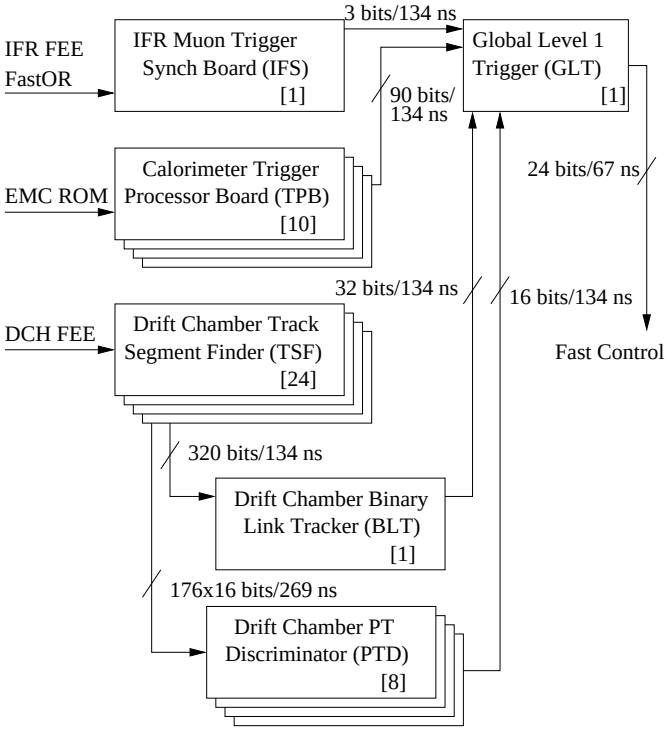


Figure 3.15: Flow chart representing Level 1 trigger operation

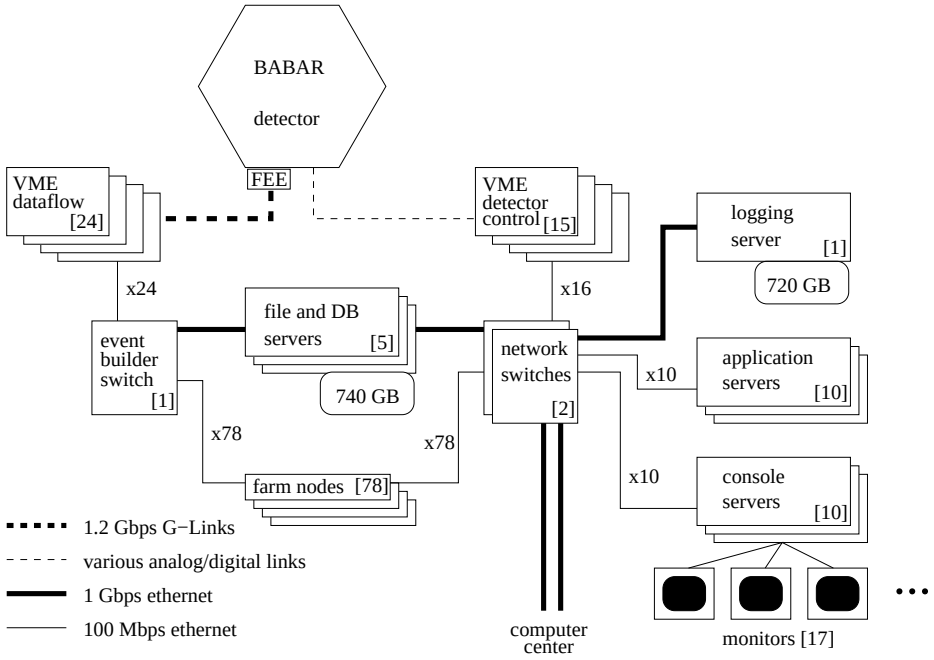


Figure 3.16: Schematic drawing of the *BABAR* online system

4 Initial event selection

4.1 Event reconstruction overview

We aim to fully (exclusively) reconstruct three decay modes:

$$B^{\pm} \to \rho^{\pm}\gamma \quad , \quad \rho^{\pm} \to \pi^{\pm}\pi^0 \qquad \text{with} \quad \mathcal{B}(\rho^{\pm} \to \pi^{\pm}\pi^0) \approx 100\%;$$
$$B^0 \to \rho^0\gamma \quad , \quad \rho^0 \to \pi^+\pi^- \qquad \text{with} \quad \mathcal{B}(\rho^0 \to \pi^+\pi^-) \approx 100\%;$$
$$B^0 \to \omega\gamma \quad , \quad \omega \to \pi^+\pi^-\pi^0 \quad \text{with} \quad \mathcal{B}(\omega \to \pi^+\pi^-\pi^0) \approx 89\%.$$

Event reconstruction proceeds in several distinct stages. First, a "skim" is used to run over all the decays reconstructed in the *BABAR* detector and select only the ones that could potentially contain the signal we are looking for. This is extremely important, as the entire dataset of several TB is far too large to work with in practice, and there are many possible categories of B decays, whose signatures are very different from each other.

Next, standard *BABAR* lists of particles reconstructed from primary event information collected by the detector using centralized software routines are used to combine individual candidates into composites, until one has a full picture of the decay. In our case, we combine low-energy photons into π^0s, then combine pions (and kaons for $B \to K^*\gamma$ decays) into ρ/ω (and K^*) mesons, and finally combine these with a high-energy photon to form the signal B meson candidate. At all stages, we require that the daughter particles originate from the same source (based on kinematic fits and geometric constraints) and that the mass of the resulting parent particle is consistent with the hypothesis assigned to it.

Finally, various event shape quantities are computed and saved for the reconstructed event, including spatial distributions of the decay products and angular

relationships between individual (or groups of) particles. It is important to note that multiple reconstructions are possible for the same event, in which case we save all the possibilities and decide which particular one to choose at a later stage of the analysis.

The main challenge, as is generally the case in all high-energy physics analyses, is to separate the events of interest from numerous background events that mimic our signal. Thus, it is possible to mis-reconstruct a kaon as a pion, allowing $B \to K^*\gamma$ background to leak in, or to use one of the photons originating from a π^0 or η decay as the primary photon candidate, after having lost the other photon during reconstruction. Since π^0 and η are light mesons, they are produced in large numbers in the continuum (i.e., decays that happen slightly off-resonance and do not proceed via the $\Upsilon(4S)$).

Because the $\omega(782)$ has a width of about 8 MeV, which is comparable to the reconstruction resolution in the final state $\pi^+\pi^-\pi^0$, a tight cut on the invariant mass of the resonance is a powerful handle to reject combinatorial background. The $\rho(770)$, however, has a width of about 150 MeV, which provides a far less stringent background rejection criterion.

Due to the two-body kinematics and the low mass of the particles in the final state, the photon and the meson are produced with very high momentum, $\mathbf{p}^*$, in the B meson reference frame. The high momentum of this signal photon allows reduction of the combinatorial background due to π^0 and η decays. Nevertheless, the background from the continuum is overwhelming, and developing a dedicated classification tool for continuum suppression is crucial to achieve the sensitivity needed for the measurement. To leverage a larger set of event variables, we moved from an artificial neural network, which was used in previous searches for $B \to (\rho/\omega)\gamma$ decays, to a decision-tree-based selection, described in detail in the next chapter.

In addition to high levels of continuum backgrounds, signal events also need to be separated from $B\overline{B}$ decays, such as $B \to X_s\gamma$ (mostly $B \to K^*\gamma$) and $B \to \rho(\pi^0/\eta)$. By using a mixture of continuum and $B\overline{B}$ Monte Carlo for multivariate classifier training (see Section 4.6), we can reduce the problem of separating signal from background to event selection based on a single variable — classifier output.

4.2 Datasets used

The analysis was performed using *BABAR*'s full runs 1–6 on-peak data sample of 424.35 fb^{-1}. Additionally, 41.37 fb^{-1}of off-peak data were used for validation studies. Monte Carlo samples used for optimization of selection criteria and for analysis technique validation are summarized in Table 4.3.

Mode	**Signal**	**Run**	$B^0\bar{B}^0$	B^+B^-	$c\bar{c}$	uds	τ
$B^\pm \to \rho^\pm\gamma$	587K	1	37200K	26072K	58900K	47180K	20378K
$B^0 \to \rho^0\gamma$	587K	2	103352K	103124K	168844K	130858K	55546K
$B^0 \to \omega\gamma$	587K	3	48466K	49766K	83974K	66892K	27988K
		4	167332K	167994K	252830K	206764K	90032K
		5	241224K	244192K	366758K	317846K	132234K
		6	105210K	100818K	156912K	127926K	56436K
Total			702784K	691966K	1088218K	897466K	382614K
Luminosity in fb^{-1}:			1338.6	1318.0	837.1	429.4	407.0

Table 4.3: Summary of $B \to (\rho/\omega)\gamma$ MC datasets for Runs 1–6

For the $B \to K^*\gamma$ ($K^+\pi^-$ and $K^\pm\pi^0$ modes) control sample, we used Runs 1–5 on-peak sample of 341.8 fb^{-1}, while run 6 was added for $K^{*+} \to K^+\pi^0$ mode, increasing the data to 376.24 fb^{-1}. The MC samples are summarized in Table 4.4.

Mode	**Signal**	**Run**	$B^0\bar{B}^0$	B^+B^-	$c\bar{c}$	uds	τ
$K^{*0} \to K^+\pi^-$	1164K	1	37200K	36968K	54214K	21164K	20378K
$K^{*+} \to K^+\pi^0$	195K	2	103498K	103124K	168844K	130858K	55606K
		3	50556K	49766K	83974K	49182K	27988K
		4	156446K	167994K	252830K	213380K	90032K
		5	244812K	244322K	360578K	301758K	132234K
Total			555312K	602174K	920440K	716342K	326238K
Luminosity in fb^{-1}:			1128.6	1147.0	708.0	342.7	347.1

Table 4.4: Summary of $B \to K^*\gamma$ MC datasets for Runs 1–5

4.3 Skim selection

We use the `BtoRhoGamma` skim in *BABAR*'s `FilterTools` package, which contains the skims for all possible *BABAR* analyses, to apply the following basic event selection criteria:

1. The event passes either the `BGFMutliHadron` (simply meaning we have identified at least three charged tracks in the event) or the `BGFNeutralHadron` (which selects photons with $\mathbf{p}^* > 500\,\mathrm{MeV}/c$ and $-0.75 < \cos(\theta_\gamma) < 0.96$) tag filter.

2. There are at least two tracks in the `GoodTracksLoose` list, which requires tracks to have at least $100\,\mathrm{MeV}/c$ of transverse momentum, with maximum momentum of $10\,\mathrm{GeV}/c$, at least 12 hits recorded in the drift chamber, and imposes basic restrictions on the point of closest approach to the interaction point.

3. The event shape: the ratio of the 2^{nd} and O^{th} Fox-Wolfram moments [21] is calculated using all tracks in the event, $R_2^{All} < 0.9$.

4. The primary photon energy is in the range [1.5, 3.5] GeV.

5. The meson candidate mass is in the range [0.5, 1.2] GeV/c^2 and [0.5, 1.3] GeV/c^2 for the ρ^0 and $\rho^\pm$ modes, respectively, and within 0.05 GeV/c^2 around the nominal ω mass.

6. Loose kinematic cuts are $5.1 < m_{\mathrm{ES}} < 5.5$ GeV/c^2 and $-0.6 < \Delta E < 0.6$ GeV.

7. The kinematics of the ρ^0 daughters are compatible with a common vertex.

Skim efficiencies for Monte-Carlo-generated events are listed in Table 4.5. Note that the skim selects events across all three signal modes, allowing cross-feed candidates to be counted. Further processing of the events by the `RhoGammaTools` analysis package (which performs the full candidate reconstruction described earlier, calculates and saves various event and candidate-level quantities, and applies further selection criteria, described in detail below) constrains events to be consistent with only one specific signal mode.

4.4 PID selection

A very important part of this analysis is correctly identifying the pions coming from ρ/ω decay. Misidentifying kaons as pions allows $B \to K^*\gamma$ background to leak though into our signal region. The optimum selection, based on *BABAR*'s standard particle identification algorithms, was determined via a detailed study in the previous runs 1–4 analysis of $B \to (\rho/\omega)\gamma$ decays. *BABAR* deploys various classification techniques, such as likelihood ratios and artificial neural networks, to correctly identify particles based on their attributes and place them into corresponding lists. There is a progression of selector algorithms with increasingly stringent selection criteria, which always introduce the mutually exclusive choice of better background rejection vs higher signal efficiency. These are labeled from `VeryLoose` to `VeryTight` in *BABAR* software.

The best pion selector for the $B \to \rho\gamma$ modes was found to be the likelihood-based `piLHVeryTight`, combined with a minimal significance criteria of 0.001 on the Poisson probability of the number of the photons seen in the DIRC for a particular candidate being different from the expected number for the assigned pion hypothesis. Because the $B \to K^*\gamma$ background is not as important in the $B^0 \to \omega\gamma$ ($\omega \to \pi\pi\pi^0$) decay mode, the `piLHTight` selector is the better choice there. Also, as enforcing the DIRC consistency criteria actually decreased signal significance, this cut was not applied to the $B^0 \to \omega\gamma$ channel.

4.5 Precuts

Table 4.6 summarizes fixed selection criteria applied to all modes before further optimization is carried out. Included are six precuts on photon quality, which are simultaneously optimized using the Bump Hunter (or PRIM) algorithm [24], although four of these are also included in bagged decision trees later, as described in Section 5.8.1. The photon quality precuts are meant to clean up the input to the decision tree classifier, reducing the size of the training sample and thus making it's job much easier, as well as to simplify the estimation of systematic errors later on by cutting away

regions of poor agreement between data and simulated events. See Figure 4.17 for the comparison plots of normalized signal vs background distributions. The variables are as follows:

- `GammaECal` : laboratory frame calibrated energy of the EMC cluster
- `Gammas9s25` : ratio of the sums of the energies of the central 9 EMC crystals to the central 25 crystals surrounding the centroid
- `GammaLat` : lateral moment of the EMC cluster
- `GammasecMom` : second moment of the EMC cluster
- `GammaZ20` : absolute value of the complex Zernike(2,0) moment [25] of the EMC cluster
- `GammaA42` : absolute value of the complex Zernike(4,2) moment of the EMC cluster.

4.6 Combined candidate selection

On top of the selection criteria described so far, many variables are combined in a sophisticated classification algorithm based on decision trees and described in the next chapter. Final selection also includes choosing the best B meson candidate in events where multiple candidates were reconstructed. Such candidates are chosen so that the mass of the daughter meson (ρ or ω) is closest to the nominal particle mass, as found in [26].

The final signal efficiencies for the entire candidate selection process (i.e., the candidates that pass all cuts and end up in the final fit) are: 6.1% for $B^0 \to \omega\gamma$, 8.5% for $B^0 \to \rho^0\gamma$, and 4.9% for $B^\pm \to \rho^\pm\gamma$. For comparison, the corresponding efficiencies for light quark (uds) continuum events are all near 10^{-7}.

data set	**pre-skim**	**post-skim**	ϵ[%]
$B^0\bar{B}^0$	702784000	333791	0.0477
B^+B^-	702558000	453767	0.0647
$c\bar{c}$	1088218000	16888023	1.5539
uds	897466000	36352462	4.0515
$\tau^+\tau^-$	382614000	4821779	1.2595
$\rho^0\gamma$	587000	423508	72.1479
$\rho^\pm\gamma$	587000	366158	62.3779
$\omega\gamma$	587000	356081	60.6612

Table 4.5: `BtoRhoGamma` skim efficiencies (runs 1–6 `R22d`)

Description	$B^0 \to \rho^0\gamma$	$B^\pm \to \rho^\pm\gamma$	$B^0 \to \omega\gamma$
High-energy γ	$-0.74 < \cos(\theta_\gamma) < 0.93$ Number of EMC crystals > 4 No problem crystal $> 25\,\text{cm}$ isolation		
Tracking	`GoodTracksLoose` requirements for all charged tracks		
PID	`VeryTight` π ID + DIRC photon consistency		`tight` π ID
π^0 selection	n/a	$0.115 < m_{\pi^0} < 0.150$ (GeV/c^2)	
R_2^{All}	< 0.7		
Converted γ veto	$0.10 \leq m_{\pi^0}^{conv} \leq 0.16$ and $0.50 \leq m_\eta^{conv} \leq 0.59$ (GeV/c^2)		
`GammaZ20`	> 0.82		
`Gammas9s25`	$[0.93, 0.99]$		
`GammaLat`	$[0.12, 0.51]$		
`GammaA42`	< 0.08		
`GammasecMom`	< 0.002		
`GammaECal`	$[1.5, 4.4]$ (relaxed from $[1.89, 4.14]$) GeV/c^2		

Table 4.6: Fixed cuts applied before optimization is carried out for the remaining criteria

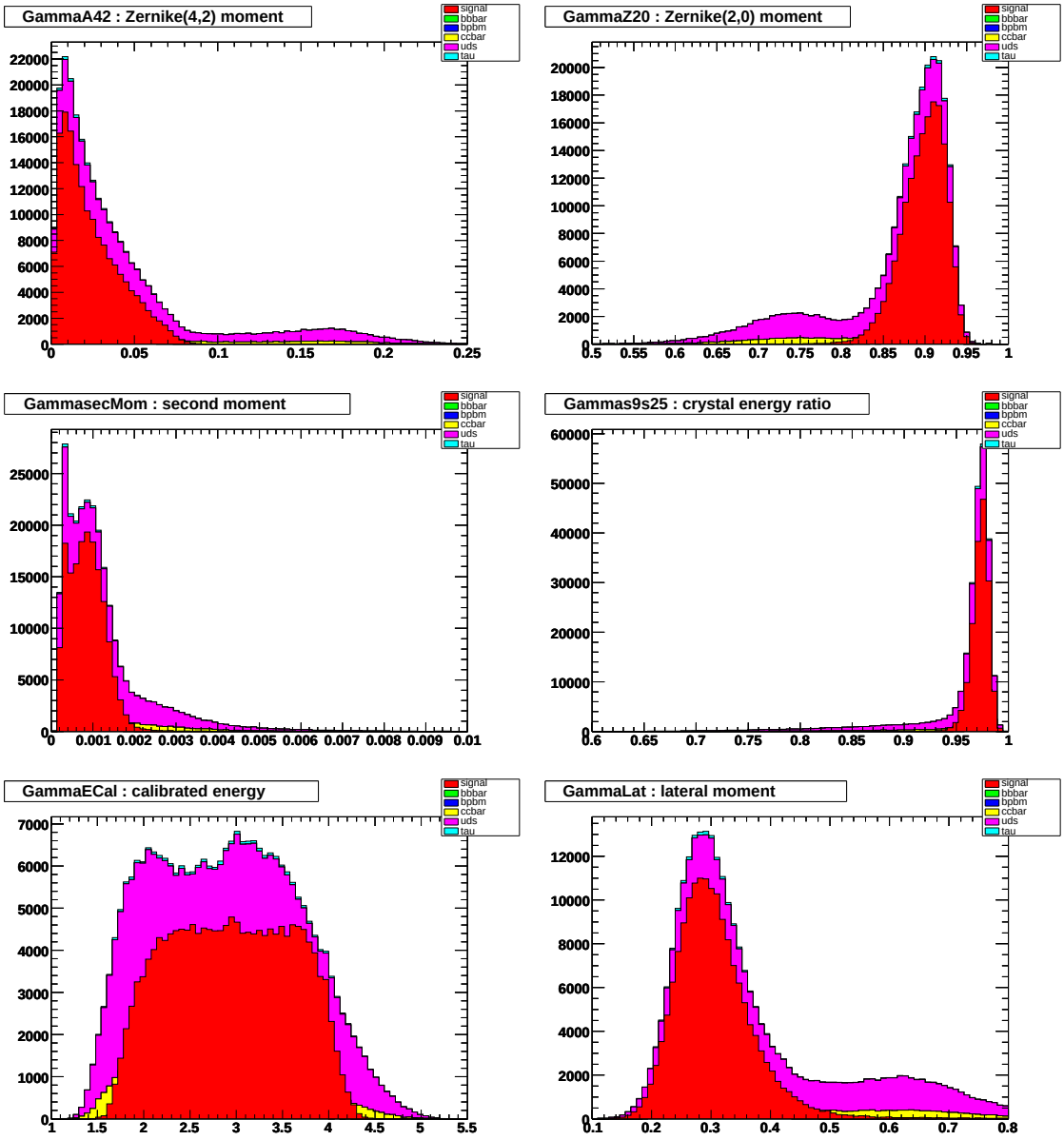


Figure 4.17: Plots of signal (arbitrarily scaled) vs stacked background MC (scaled to 316 $fb^{-1}$of data) for photon quality precuts optimized with Bump Hunter

5 Classification of events

5.1 Overview

In the previous *BABAR* analysis of $B \to (\rho/\omega)\gamma$ decays, the statistical significance of signal yield was slightly below 5σ in the ρ^0 mode and 4σ in the $\rho^\pm$ mode. We (and the subsequent update from the Belle experiment[14]) also failed to find any significant evidence for signal in $B^0 \to \omega\gamma$. Therefore, we decided to try to apply a different classification algorithm to suppress our overwhelming background. The (artificial) neural network (NN) that was previously used was already at its performance limit due to poor performance scaling with the number of input dimensions, as well as poor ability to deal with correlated and/or redundant inputs. We decided to look into the possibility of using boosted or bagged decision trees, which perform much better in higher dimensions and deal gracefully with varied, correlated, irrelevant, and redundant inputs. Because the $B^0 \to \omega\gamma$ mode produced the lowest and least significant yield in all previous analyses, we chose to concentrate on it for all of our initial tests of decision tree performance. A new `C++`-based package written by former *BABAR* collaborator Ilya Narsky was adopted as our primary tool. Now open-source software, `StatPatternRecognition` [27] is available for download from `http://sourceforge.net/projects/statpatrec` (documentation is also available in the corresponding `README` file).

5.2 Decision trees, boosting, and bagging

Decision trees are a well-known tool in machine learning, widely used for classification problems (in our case, separating signal from background). In HEP, a decision tree can be built by choosing the most useful event variables and making decision splits on them. This is similar to making a cut on a kinematic variable, except that this is done over and over for each new tree node. In `StatPatternRecognition`, a decision tree is built by trying all possible splits on all possible variables and choosing the one that optimizes a given figure of merit (FOM).

Although one can specify various FOMs at runtime, the usual criteria to optimize is either the Gini index [28] (used throughout this analysis) or cross-entropy [29]. In this case, one is concerned mostly with the fraction of correctly classified events and does not have to worry about applying correct luminosity weights to the different MC samples used in training and validation, as is required, for example, in the case of statistical significance optimization. Optimizing the Gini index is equivalent to maximizing the purity of child nodes in a tree, so that one node will contain almost entirely signal, while the other almost entirely background events.

`StatPatternRecognition` also does not apply any pruning algorithms, which are often used in other implementations to remove nodes, which produce splits that are no better than a random choice, or to merge nodes to further improve the overall FOM.

Because trees are simple to build and easy to visualize, they are a powerful algorithmic tool. However, a single tree cannot rival the power of the complexity built into a neural network. Thus, we extend the capabilities of tree-based algorithms by applying two modern statistical tools: *Boosting* (*adaptive boosting* or *AdaBoost*) and *bootstrap aggregation* (*bagging*).

Boosting was introduced by Freund and Shapire in the form of the AdaBoost algorithm in 1999 [30]. It consists of combining weak classifiers (individual trees) into a weighted sum that becomes a new, strong classifier. In the process of building new trees to classify our events, we increase the weights of those that have been

misclassified so far, so as to force the algorithm to pay them more attention in the future. The final decision of whether to classify a certain event as signal or background is made by a collective vote of all the weak classifiers. In the framework of `StatPatternRecognition`, one can control the minimum size of a terminal tree node (one that is not split further) as well as the number of weak classifiers to combine. Generally, one wants to build $O(100)$ trees with large terminal nodes for the algorithm to perform near optimally.

Bagging (or bootstrap aggregating) is the brain child of Leo Breiman, dating back to 1996 [31]. Here, the idea is to use bootstrap replicas of data, i.e., resampling the data a certain number of times, each time selecting with replacement a subset of training events. Thus, one "averages" over the bootstrap samples to gain better accuracy than a single-shot approach. Bootstrap was originally introduced as a way to estimate certain parameters of a statistical distribution, such as correlations among variables, for which there is no close-form algebraic expression. In the case of decision trees, it's a different way of producing a more accurate decision based on some collective vote. While the parameters of Bagger in `StatPatternRecognition` are the same as for AdaBoost, in this case, the user should make the terminal nodes small to pick up on the variation among the bootstrap replicas of training data.

5.3 Multivariate classifier training

The actual training of a decision-tree-based classifier is done in the following way. Samples of signal and background Monte Carlo are produced by applying all preselection criteria. The samples are then split randomly in half, the first subset becoming the training and the second the testing/validation samples. In earlier studies, our training sample consisted of about 30,000 signal and 45,000 background events, while for the final classifier training, we roughly doubled these numbers. The training algorithm is fed a vector of values corresponding to each input variable for each candidate in the sample. Once the classification is complete, it is applied to the testing sample, and the validation curve (exponential loss for AdaBoost or quadratic loss for

Bagger[3]) is plotted (e.g., see Figure 5.18). From this graph, we can see whether the training has converged (i.e., the loss function has leveled off, and we've fixed all the misclassifications we could), or if we need to add more cycles (i.e., build more trees). If the loss function starts shooting back up, we have overtrained and need to use fewer cycles. From these validation curves, we can see that most of the work is done in under 100 cycles, the next 100 providing a small performance increase, and after about 200 cycles, each additional one provides only a small incremental contribution.

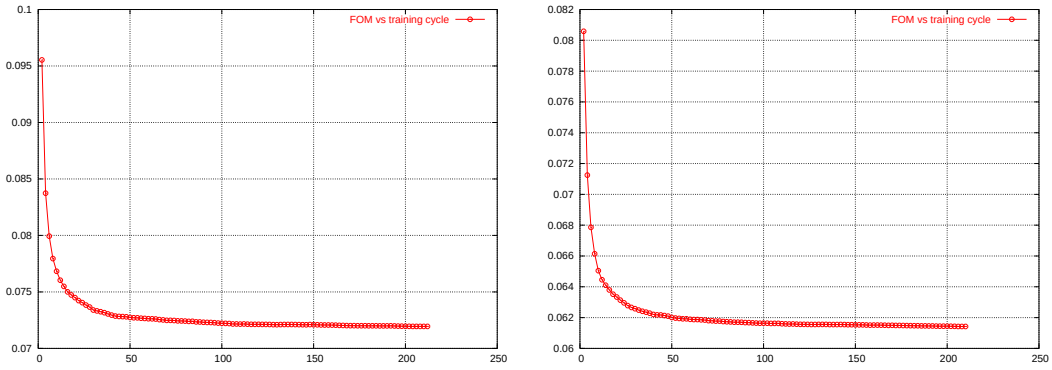


Figure 5.18: $B^0 \to \omega\gamma$ (left) and $B^0 \to \rho^0\gamma$ (right) Bagger validation curves (quadratic loss) as a function of training cycle (number of trees built)

5.4 Study of NN vs AdaBoost performance

Because this analysis is an update of earlier searches for $B \to (\rho/\omega)\gamma$ decays at *BABAR*, much of the initial work went into developing an improved analysis framework that would approach the goal of extracting the desired signal events from data in a substantially different manner from the previous iterations. Therefore, several studies were performed that compared fairly the background suppression performance of NNs used for the most recent *BABAR* measurement [15].

The first thing to investigate was how the performance of the existing NN compares to that of a boosted decision tree with the same set of input variables. Because

[3]A loss function measures the difference between the target and the result of classification. Thus, for quadratic loss used in bagging, a signal event with classifier output of 0.5 corresponds to a loss of $(1 - 0.5)^2 = 0.25$.

the previous *BABAR* analysis also relied on Bump Hunter [24] to optimize the cuts for remaining variables that were not included in the NN, as well as the cut on the NN output itself, we included these variables in the decision tree inputs to produce a fair comparison. In order to assess the performance, we compared the signal vs background efficiency curves produced by scanning the output of each algorithm, with the aim of achieving the lowest possible background efficiency for a given signal efficiency (or, alternatively, the highest signal efficiency for a given background efficiency).

Figures 5.19 through 5.21 show the normalized distributions of signal and background for AdaBoost trained with `StatPatternRecognition` next to the NN, as well as the efficiency curves for both. In these plots, the NN output is shown with the rest of the cuts already applied, which accounts for the difference in the behavior of the efficiency curves. Thus, we do not see the NN efficiencies reach 1, as is the case with AdaBoost. The optimal cut value, based on maximizing signal significance, is shown on the plot of AdaBoost output. A vertical line is also drawn on the NN plot to show the nominal cut applied in the previous analysis. Horizontal lines on the efficiency curve plots show the background level for each of these cuts. By following these lines, one can directly compare the performance (in terms of gain or loss in signal efficiency for the given background efficiency) of both the old and the new background rejection methods. Efficiency curves for AdaBoost are shown for both the training and the testing samples.

From this quick study, we determined that the $B^0 \to \omega\gamma$ mode can easily be improved, while the $B \to \rho\gamma$ modes require more effort. This study also demonstrated that while the old NN shows a better ability to concentrate the signal events near 1, AdaBost does better at moving most of the background to 0 ($[0, 1]$ being the standard output range for these multivariate classification algorithms). This makes AdaBoost the more desirable method, as the overall goal is to reduce the background, while keeping enough signal to extract a statistically significant measurement. However, both classifiers exhibit an unwanted feature — the tendency for the signal to peak near 0 and for background to peak near 1.

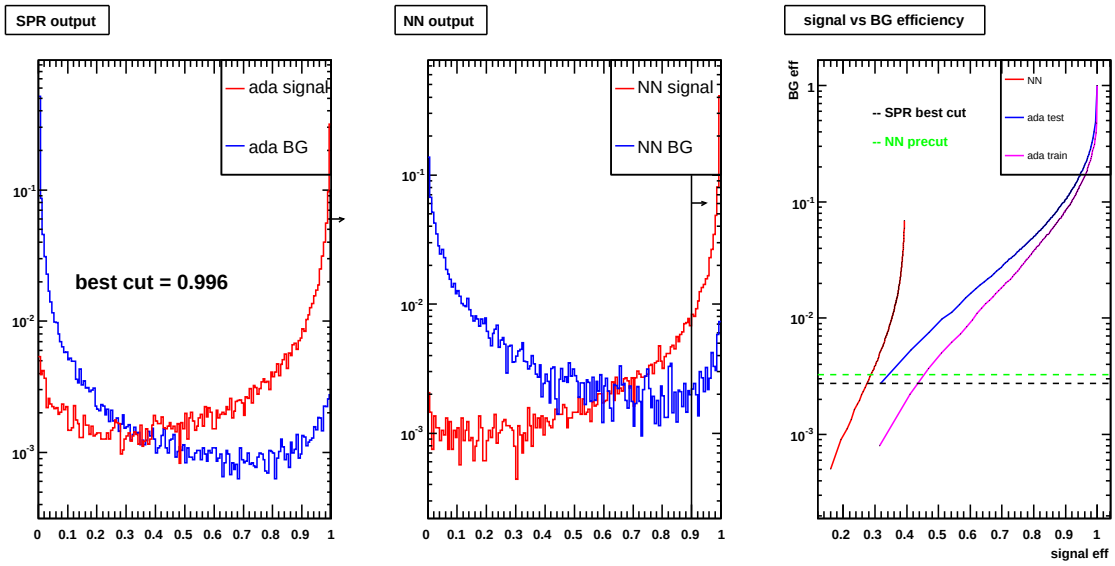


Figure 5.19: $B^0 \to \omega\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with NN variables from previous *BABAR* search for $B \to (\rho/\omega)\gamma$

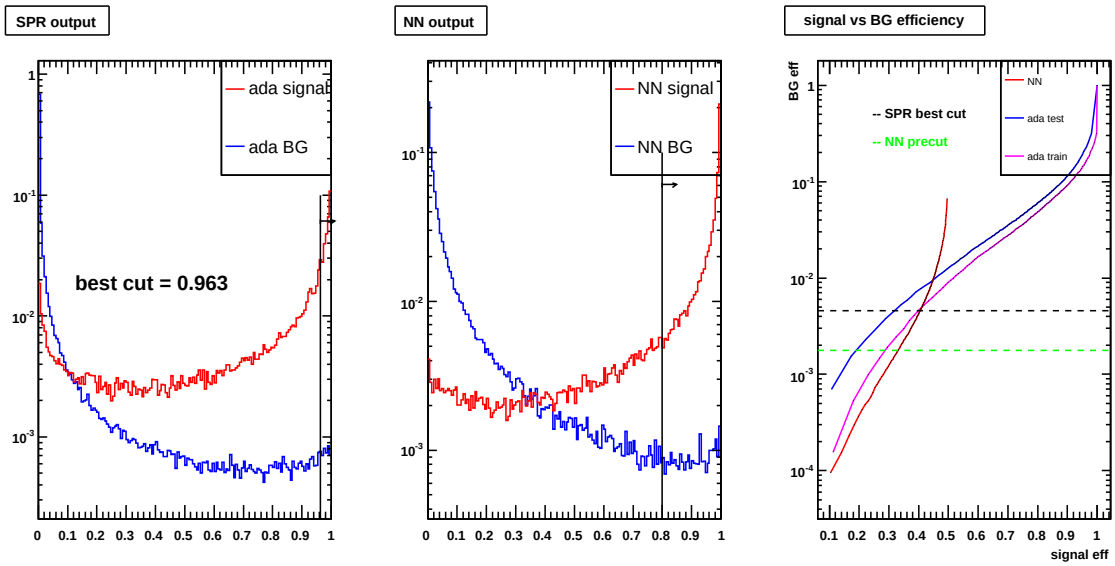


Figure 5.20: $B^{\pm} \to \rho^{\pm}\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with NN variables from previous *BABAR* search for $B \to (\rho/\omega)\gamma$

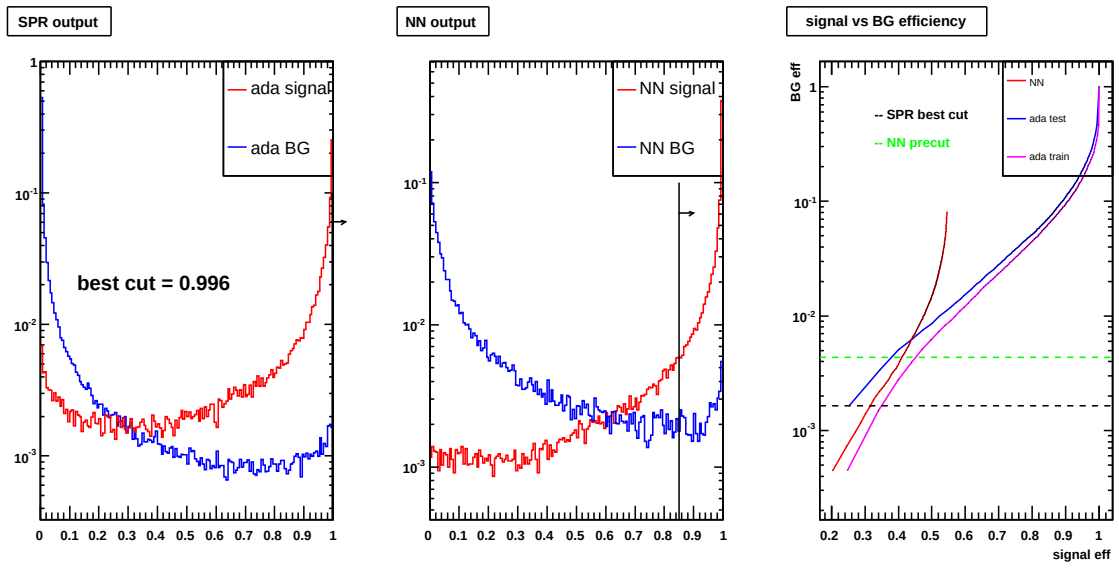


Figure 5.21: $B^0 \to \rho^0\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with NN variables from previous *BABAR* search for $B \to (\rho/\omega)\gamma$

5.5 Study of variable categories

The next study performed was to determine the contribution to overall classifier performance coming from different variable categories: B meson variables, daughter meson (ρ/ω) variables, γ variables, and $\pi^{(0)}$ variables. Here, the $B^0 \to \omega\gamma$ mode is used for illustration, being the channel we concentrate mostly on. Figures 5.22 through 5.25 show the progression of AdaBoost performance, compared to the same NN shown in previous section, as these categories of variables are added to all rest-of-event (ROE) variables available in the n-tuples. This demonstrates several points: a) ROE variables by themselves do not provide enough separating power, b) even with signal B meson variables added, we still can't beat the NN c) all categories of variables add something to the overall performance.

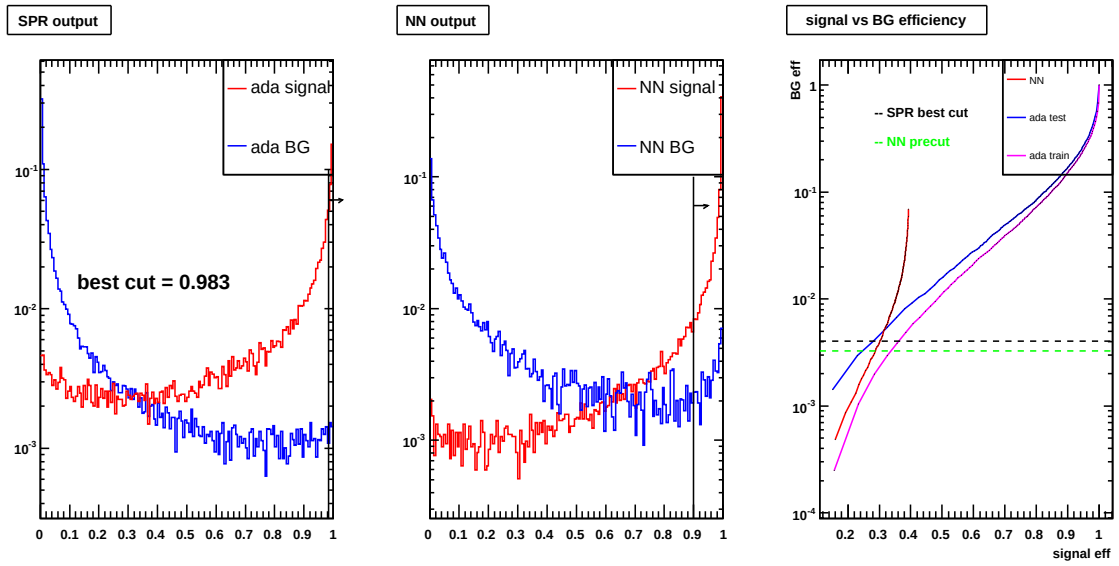


Figure 5.22: $B^0 \to \omega\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with ROE and B meson variables only

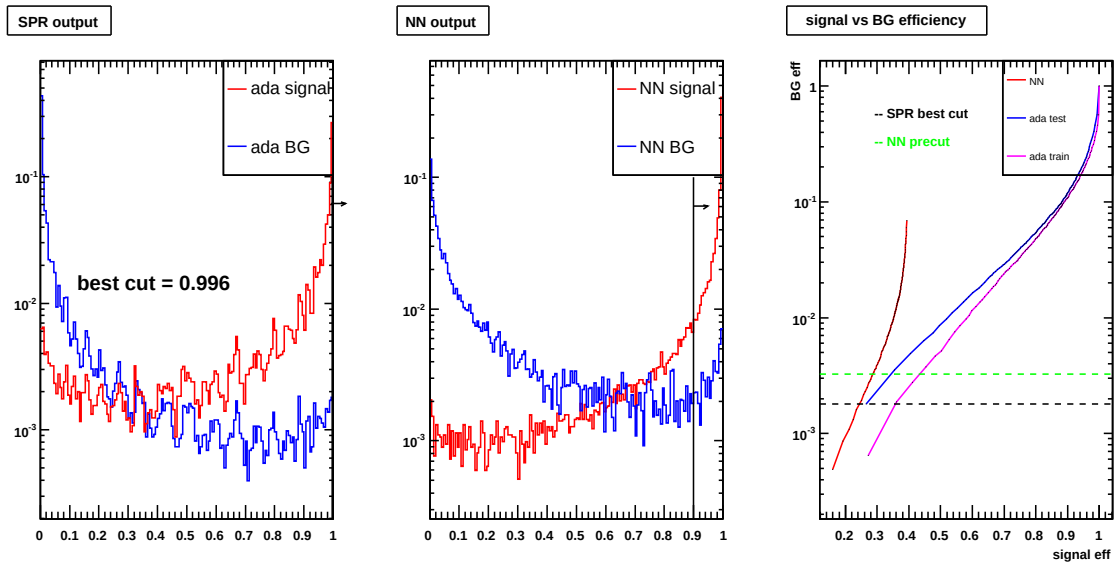


Figure 5.23: $B^0 \to \omega\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with ROE, B meson, and ω variables

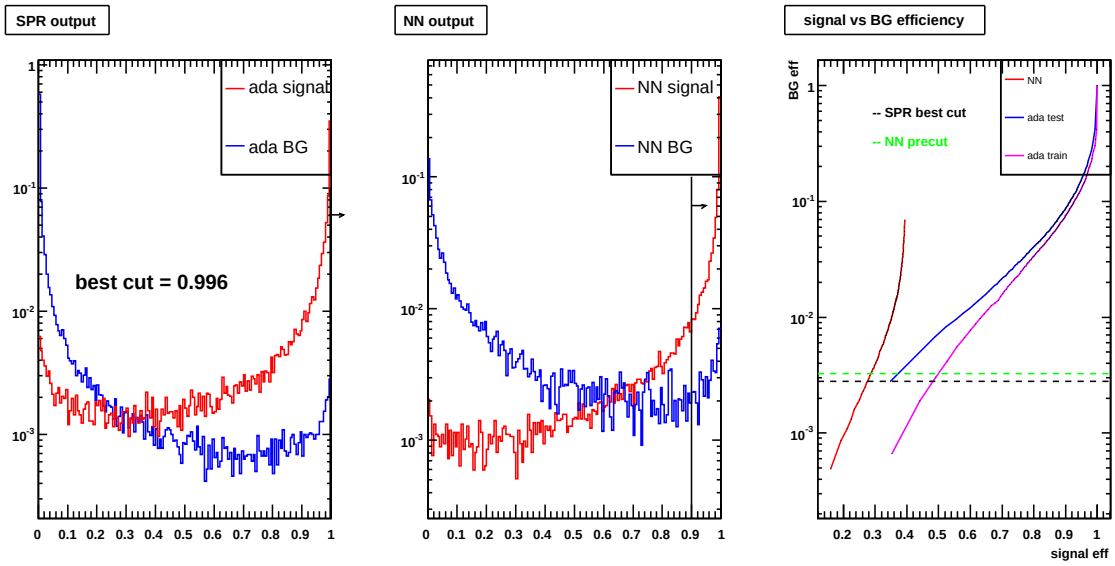


Figure 5.24: $B^0 \to \omega\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with ROE, B meson, ω, and γ variables

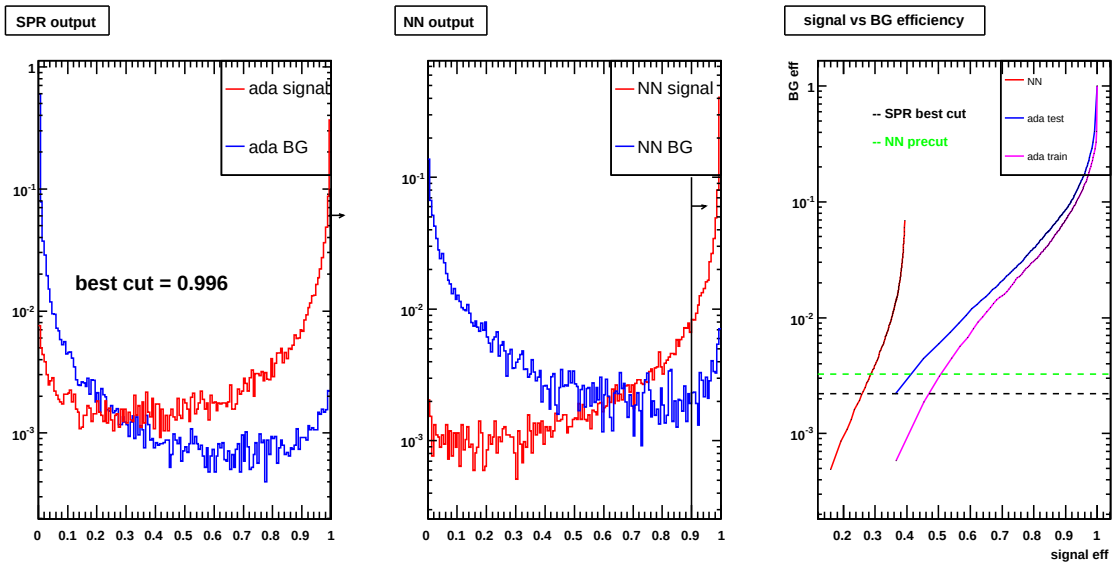


Figure 5.25: $B^0 \to \omega\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with ROE, B meson, ω, γ, and π^0 variables

5.6 Study of potential gain from extending the input variable list

Once we determined that adding variables from different categories helps significantly boost classifier performance, it seemed clear that there was a lot of potential in utilizing a much larger variable set than the one deployed for NN inputs, where signal-side variables were avoided, except for the cosine of the polar angle of the B candidate in the center of mass frame. Therefore, we performed a study of AdaBoost performance with the full set of n-tuple variables (~ 300) vs that of the original NN (~ 30). The results, shown in Figures 5.26 through 5.28, demonstrate a very high potential for performance gain.

One has to remember that in this order-of-magnitude study we include variables that are not well modeled in Monte Carlo simulations, as well as variables highly correlated with our primary kinematic variables $m_{\rm ES}$ and ΔE. However, even once the input list is pruned, one expects a performance gain over the old approach. We tested this hypothesis by removing obviously problematic variables from the list and re-training AdaBoost. Bagger was also trained for both cases. Figures 5.29 and 5.30 show that the differences in output were found to be very small, thus proving that there exist enough extra kinematic variables to warrant the expansion of the variable input list by about a factor of 2.

However, it is important to note here that momenta of particles were retained in the pruned input variable list. Therefore, these plots demonstrate that a very high level of signal and background separation can be obtained with decision-tree-based methods, but the results would only be appropriate for a simple cut-and-count analysis due to correlations of particle momenta with our primary fit variables, $m_{\rm ES}$ and ΔE. Because even with this level of separation, there still remains a large number of background events near the signal peak at AdaBoost or Bagger output value of 1, such a treatment is unrealistic. Therefore, we remove particle momenta from the list of classifier inputs in all of the following studies.

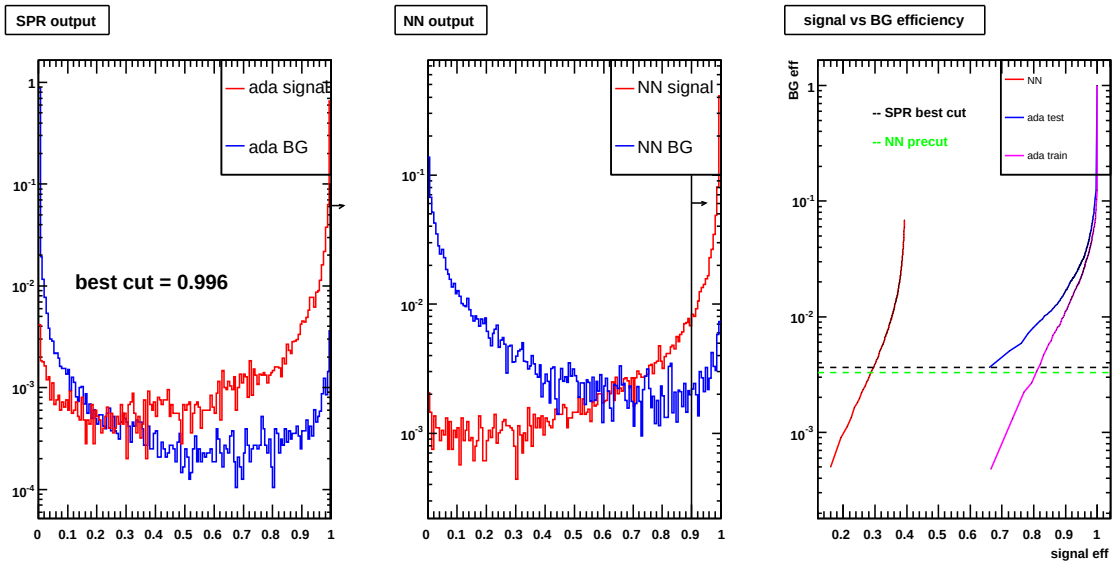


Figure 5.26: $B^0 \to \omega\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with 303 variables available in the full ROOT n-tuples

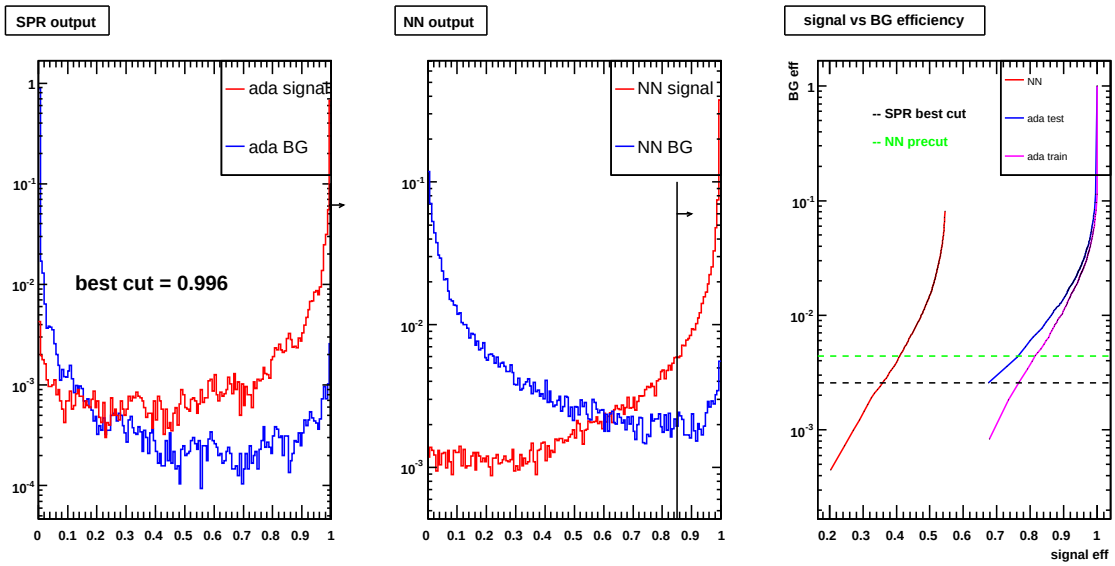


Figure 5.27: $B^0 \to \rho^0\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with 296 variables available in the full ROOT n-tuples

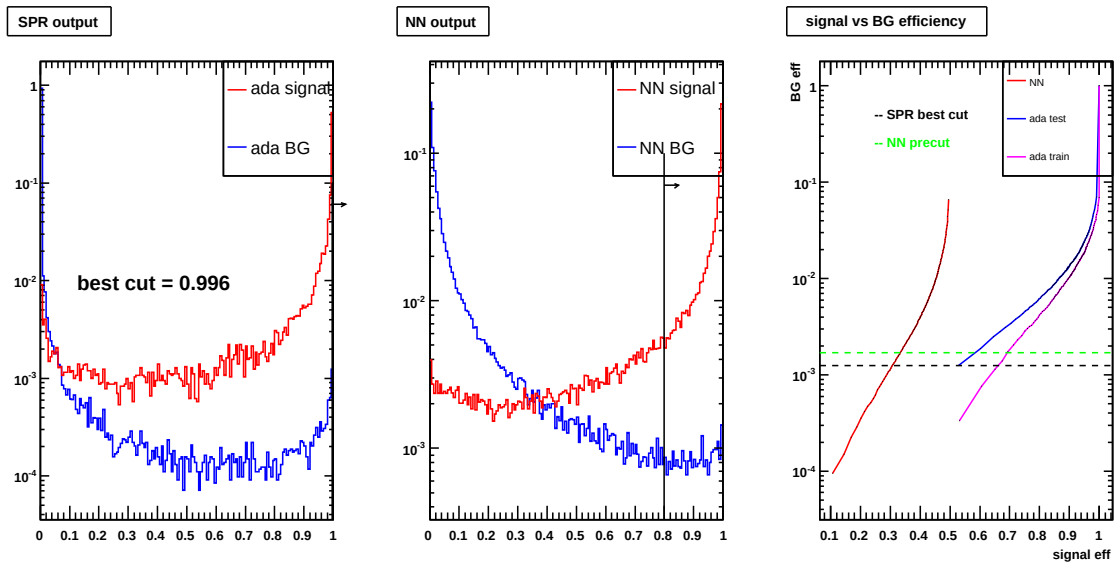


Figure 5.28: $B^{\pm} \to \rho^{\pm}\gamma$ AdaBoost (signal in red, combined backgrounds in blue) with 266 variables available in the full ROOT n-tuples

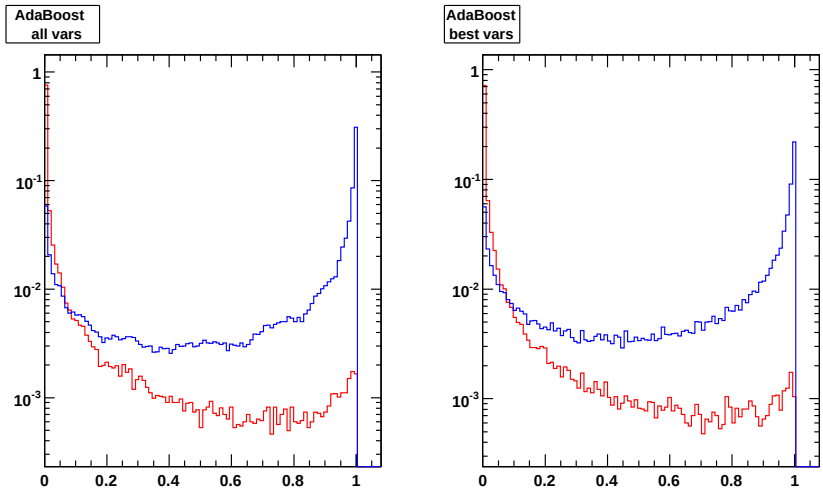


Figure 5.29: $B^0 \to \omega\gamma$ AdaBoost normalized output (signal in blue, background in red) compared for the cases where all 303 available variables were used as inputs (left) vs a pruned list of only 70 inputs (right)

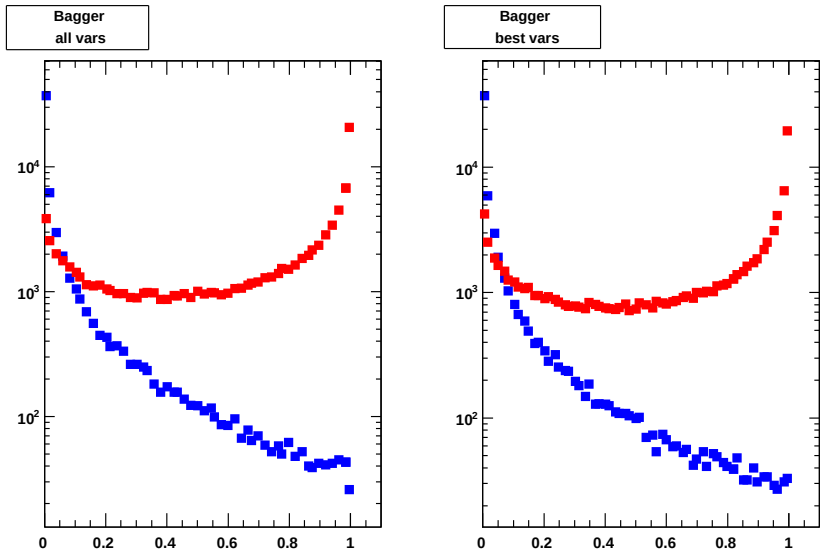


Figure 5.30: $B^0 \to \omega\gamma$ Bagger normalized output (signal in red, background in blue) compared for the cases where all 303 available variables were used as inputs (left) vs a pruned list of only 70 inputs (right)

The studies described above demonstrated the potential for improvement over the previous *BABAR* analysis of $B \to (\rho/\omega)\gamma$ decays by switching to a decision-tree-based

mode	**classifier**	N_{signal}	$N_{background}$	$S/\sqrt{S+B}$
$B^0 \to \rho^0\gamma$	AdaBoost	11.0	17.4	2.06
$B^0 \to \rho^0\gamma$	Bagger	14.3	27.7	2.20

Table 5.7: Summary of AdaBoost vs Bagger performance comparison using cut-and-count analysis in the signal region for $B^0 \to \rho^0\gamma$

classification scheme leveraging the power of boosting or bagging techniques and expanding the input variable set considerably. This improvement is made possible by the ability of decision trees to deal with correlated and irrelevant inputs and to perform well in higher dimensions, both of which are the downfall of neural networks.

5.7 AdaBoost vs Bagger

Based on the preliminary results described so far, we decided to pursue this new strategy and develop in parallel two classifiers based on boosted and bagged decision trees. As seen in Figures 5.29 and 5.30, while AdaBoost and Bagger perform similarly for signal MC events, Bagger produces better-behaved background distributions that fall off steadily as they approach 1, rather than peaking there (thus mimicking signal), as in the case of AdaBoost. However, to really be able to compare them, two studies had to be carried out. First, we looked at the signal vs background efficiency curves, similar to the ones that compared AdaBoost to the NN earlier. We also looked at the corresponding FOM plots to determine which classifier provides better signal sensitivity (defined as $S/\sqrt{S+B}$). Then, we performed a simple cut-and-count analysis in the signal region based on a cut on AdaBoost or Bagger output and compared the plots of $m_{\rm ES}$ and ΔE in the fit region (see Table 5.7).

Figures 5.32 and 5.33 compare efficiency and signal significance curves for AdaBoost and Bagger for $B^0 \to \omega\gamma$ and $B^0 \to \rho^0\gamma$ modes, respectively. Figure 5.34 shows the distributions of signal and background components in $m_{\rm ES}$ and ΔE for the $B^0 \to \omega\gamma$ mode using a cut at 0.96 on AdaBoost output, while Figure 5.35 shows

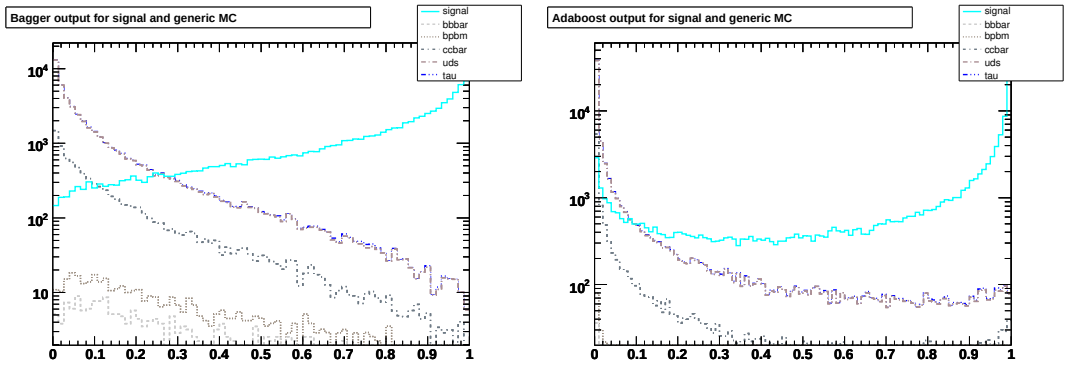


Figure 5.31: Raw output of Bagger (left) and AdaBoost (right) in $B^0 \to \omega\gamma$ mode. Signal is plotted against stacked backgrounds on a log scale.

the same distributions using the same cut on Bagger output. These plots clearly demonstrate the superiority of Bagger over AdaBoost.

What's more, AdaBoost seems to have a really unwanted feature in that it lacks an extremum in the signal significance curve, thus making it hard to justify a choice of cut on the output. We believe this is due to the previously mentioned problem with AdaBoost output for background events — namely the tendency of the background distribution to peak near 1, similar to the signal. In the case of Bagger, the background distribution drops off steadily near 1, allowing for a much cleaner cut on the output that rejects most of the background while retaining enough signal to produce peaks in $m_{\rm ES}$ and ΔE. See Figure 5.31 for a comparison of raw outputs for the two classifiers. It is clear that although AdaBoost produces more sharply peaking distributions for both signal and background events, Bagger has the better-behaved output of the two in the more-interesting-to-us region near 1.

These studies of classifier performance led us to ultimately select bagged in favor of boosted decision trees.

5.8 Variable selection

Several studies were carried out in order to prune the list of input variables to AdaBoost or Bagger to a more manageable set of well-modeled quantities that are not

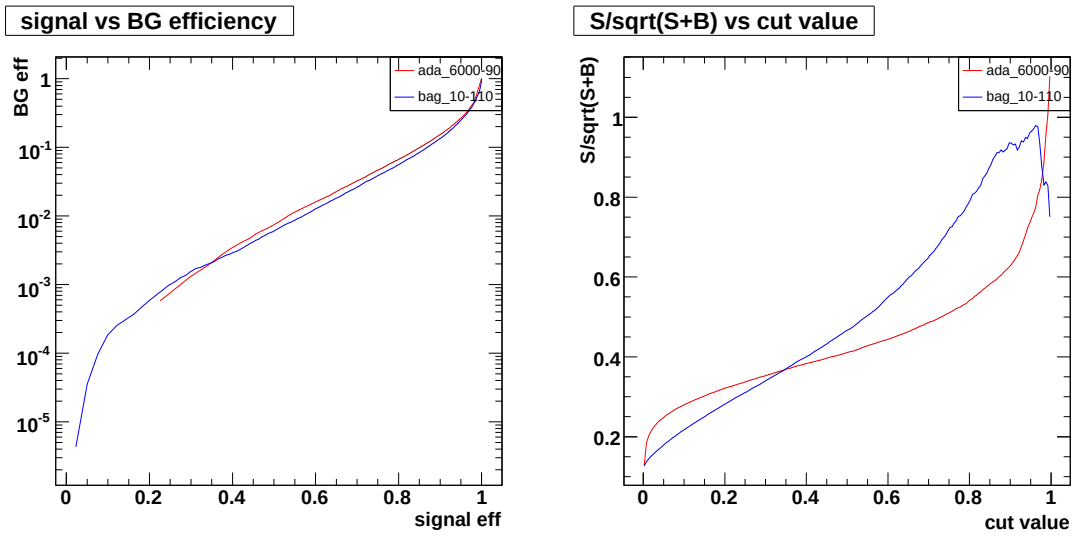


Figure 5.32: Signal vs background efficiency curves (left) and signal significance curves (right) for AdaBoost (red) and Bagger (blue) trained for $B^0 \to \omega\gamma$ mode

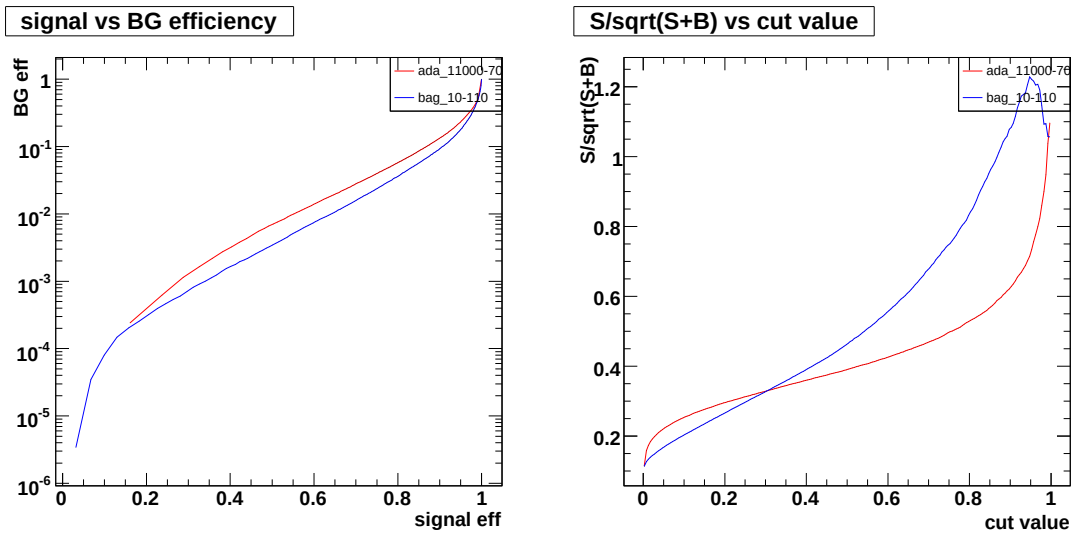


Figure 5.33: Signal vs background efficiency curves (left) and signal significance curves (right) for AdaBoost (red) and Bagger (blue) trained for $B^0 \to \rho^0\gamma$ mode

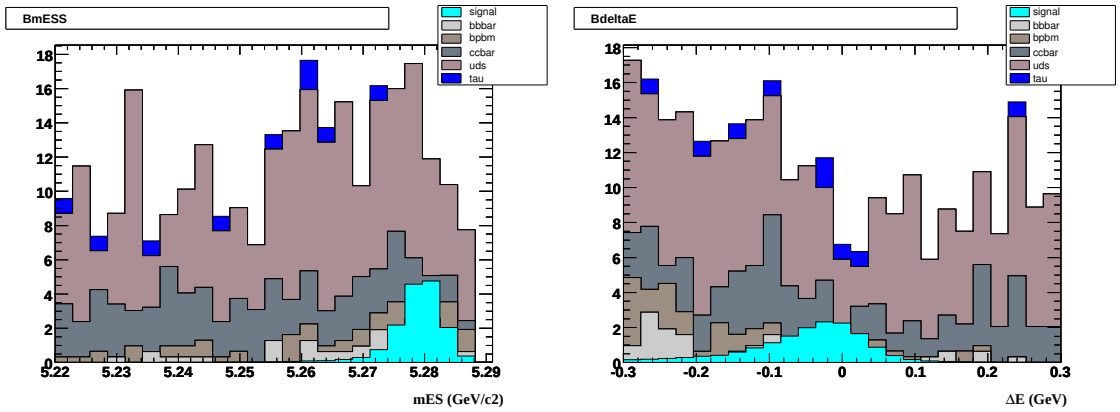


Figure 5.34: Distributions of MC signal overlaid on top of stacked background components for m_{ES} (left) and ΔE (right) for $B^0 \to \omega\gamma$ mode after a cut on AdaBoost output

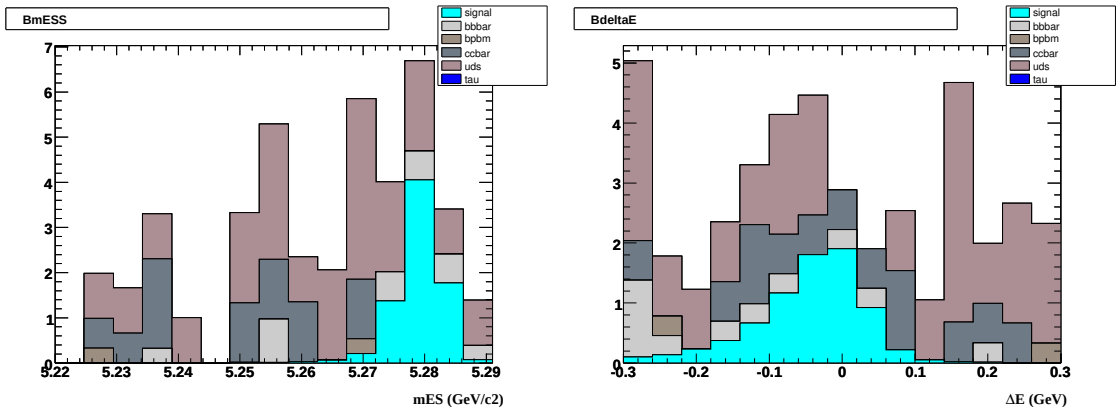


Figure 5.35: Distributions of MC signal overlaid on top of stacked background components for m_{ES} (left) and ΔE (right) for $B^0 \to \omega\gamma$ mode after a cut on Bagger output

strongly correlated with our primary fitting variables $m_{\rm ES}$ and ΔE. Before anything else, input variables were checked for compatibility between off-peak data and combined, luminosity-scaled generic Monte Carlo. If distributions exhibited visually-detectable differences, the corresponding variables were removed from training. We'll return to this study in Section 6.7.

5.8.1 Photon quality precuts

As described in Section 4.5, photon quality cuts are applied to our samples before any classifier training is produced. While the original idea was to remove these six variables from the list of Bagger inputs, studies showed that it is beneficial to keep them even after rectangular cuts are applied. This is likely due to correlations among the variables, which are impossible to see from the plots, yet can be utilized by the flexible multidimensional classifier. However, the cut on `GammaECal` was already relaxed to avoid biases. The variable is also rather strongly correlated with ΔE (see Figure 5.36). Also, there was some concern regarding data-MC agreement for `GammasecMom` at lower values. Although the difference is not very dramatic, when we considered the fact that the optimal cut value for this variable never changed in any of our optimizations (using the Bump Hunter in both the old and the new analyses), we did not see the benefit in retaining it the list of inputs. Therefore, we decided that it would be best to remove the above two variables from the list of Bagger inputs.

5.8.2 Correlations with $m_{\rm ES}$ and ΔE

Two different studies of variable correlations were carried out. First, we determined which of the potential Bagger inputs were correlated with ΔE at around 10% level or higher. One of the variables removed after this study (`GammaECal`) has already been mentioned. The other variables that were eliminated through this process are `BcosHelB(Gamma/Meson)`, all correlated with ΔE at the 30% level in *uds* and with $m_{\rm ES}$ at the 20% level in signal Monte Carlo. These are the two calculations of the helicity angle of the B meson in the CM frame using either the meson or the photon

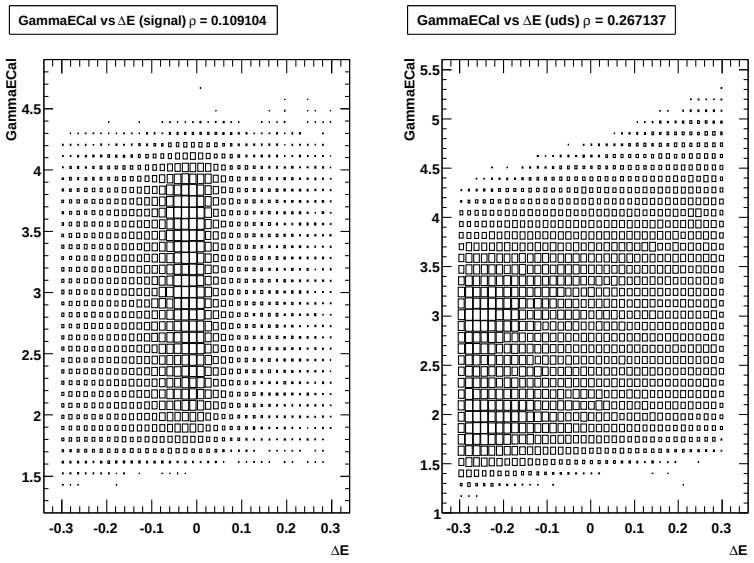


Figure 5.36: Correlations of `GammaECal` vs ΔE in signal and *uds* continuum MC

daughter.

5.8.3 Signal photon variables

Next, we studied the correlations between the signal photon variables and the rest of the Bagger inputs, demonstrating that the former can be treated as an independent sample, thus justifying the $e^+e^- \to \mu\mu\gamma$ study described in Section 9.2 to evaluate the systematic error associated with the precuts. Due to the very large number of associated correlation plots, we decided not to include them here.

5.8.4 Dalitz and helicity angles and $\pi^+\pi^-$ displacement for $B^0 \to \omega\gamma$

The Dalitz angle (defined as the angle between the π^+ and π^0 momenta in the $\pi^+\pi^-$ rest frame), the helicity angle (defined in Table 5.8), and the $\pi^+\pi^-$ displacement (distance between the B and the ω vertices) were considered as inputs to the Bagger, as well as components of the final fit. However, the Dalitz angle and the $\pi^+\pi^-$ displacement have no corresponding quantities in the $B \to K^*\gamma$ control sample, used

to validate the performance of the Bagger. On top of that, Bagger trained with the Dalitz angle as one of the inputs performed almost indistinguishably compared to the Dalitz-free configuration. The performance of the Bagger trained without $\pi^+\pi^-$ displacement also did not suffer much (based on split counts, this could be explained by the substantial increase in contribution from the χ^2 probability of the ω vertex). Therefore, it was decided to simplify the classifier architecture further by removing these two variables. After attempting to use the Dalitz angle as one of the components of the final fit and failing to find any improvement, the variable was completely left out of the analysis.

While removing the helicity angles from Bagger configuration did not result in a significant performance drop for $B^0 \to \omega\gamma$ and $B^\pm \to \rho^\pm\gamma$ modes, we decided to keep this variable for all three signal modes to maintain uniformity. The reason for its lack of contribution in the case of these two modes, based again on the number of decision splits, is its strong correlation with the cosine of the angle between the two photons originating from the π^0 decay. As the latter variable is not present in the $B^0 \to \rho^0\gamma$ ($\rho^0 \to \pi^+\pi^-$) mode, we chose to keep the helicity angle in all the signal modes, and split the load between it and $\theta_{\gamma\gamma}$ for the modes containing a π^0.

5.8.5 Meson masses

Due to the extremely broad nature of the ρ resonance ($\sim 150\,\mathrm{MeV}$), it is impractical to include the meson mass as one of the components of the fit. However, including this variable in the Bagger configuration is problematic from the point of view of using the $B \to K^*\gamma$ control sample for validation for exactly the same reason. We therefore attempted to remove the meson mass from Bagger training and instead used simple cuts (optimized using the Bump Hunter algorithm in the last round of this analysis). Thus, the ranges of $633 < m_{\pi^+\pi^-} < 957\,\mathrm{MeV}/c^2$, $636 < m_{\pi\pi^0} < 932\,\mathrm{MeV}/c^2$, and $764 < m_{\pi^+\pi^-\pi^0} < 795\,\mathrm{MeV}/c^2$ for ρ^0, $\rho^\pm$, and ω, respectively, were used. These cuts in combination with a retrained Bagger recovered the performance of the original configuration.

5.8.6 Likelihood ratio vetoes for $\pi^0/\eta \to \gamma\ \gamma$ decays

As noted previously, $\pi^0/\eta \to \gamma\ \gamma$ decays produce a very large amount of background where one of the daughter photons has a high enough momentum to mimic the signature of our signal events. These events are vetoed in our analysis using the Likelihood Ratio approach.

The π^0 and η likelihood ratio functions are defined as

$$\mathcal{LR}(\pi^0/\eta) = \frac{P(m_{\gamma_1\gamma_2}, E_{\gamma_2}|\pi^0/\eta)}{P(m_{\gamma_1\gamma_2}, E_{\gamma_2}|\rho/\omega) + P(m_{\gamma_1\gamma_2}, E_{\gamma_2}|\pi^0/\eta)},$$

where P is the probability density function (PDF) given in terms of the invariant mass of the photon pair, $M(\gamma_1\gamma_2)$, and the energy of of the lower-energy photon in the lab frame, E_{γ_2}. We determine the PDFs by performing 2-dimensional kernel density estimation[32] using signal and continuum MC samples. For the signal PDF, γ_1 is matched using MC truth information to have originated from the B meson, while for the π^0/η PDFs, both γ_1 and γ_2 are truth-matched to have originated from the same mother particle.

In previous searches for $B \to (\rho/\omega)\gamma$ decays at *BABAR*, the likelihood ratios were not included in a multidimensional classifier, and their performance was assessed separately. Here, we choose to include these functions as inputs to the Bagger. Thus, their contribution to the overall performance can only be estimated.

5.8.7 Final list of classifier inputs

The final list of Bagger inputs is given in Table 5.8 for $B^0 \to \omega\gamma$. The variables are sorted by their contribution to overall classifier performance, based on the number of decision splits. The inputs are basically identical in the other two modes, but the order of their importance is a bit different. You can see the top 15 variables for the $B \to \rho\gamma$ modes in Table 5.9. The final Monte Carlo output curves for the $B^0 \to \rho^0\gamma$ Bagger are shown in Figure 5.37. The signal efficiencies for the Bagger cuts alone are: 34.4% for $B^0 \to \omega\gamma$, 31.1% for $B^0 \to \rho^0\gamma$, and 19.0% for $B^\pm \to \rho^\pm\gamma$.

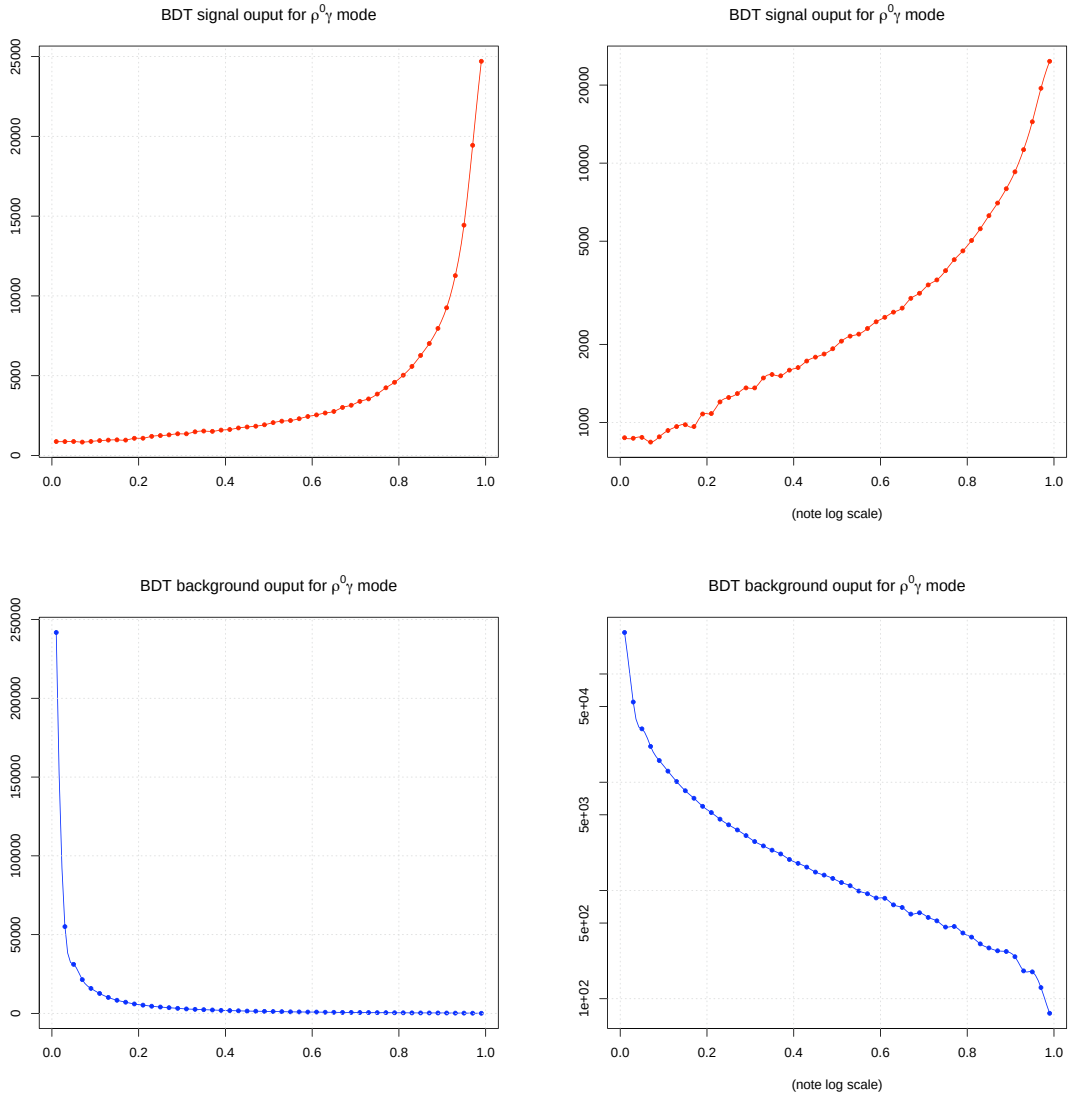


Figure 5.37: Final MC output signal (top) and background (bottom) curves for $B^0 \to \rho^0\gamma$ Bagger. Linear scale plots are shown on the left, logarithmic scale plots on the right

Variable name	**Number of decision splits**	**Variable definition**
`BdeltaZFitOverErr`	48980	Δz, the displacement between the two B vertices in the z direction, divided by the error
`Pi0cosGammaAngle`	36450	Cosine of the angle between the two γ daughters of π^0 in the lab frame
`R2All`	32969	The ratio of the second to the zeroth Fox-Wolfram moments for the event, using `ChargedTracksAcc` and `GoodNeutralLooseAcc`
`BcosThetaCM`	30392	Cosine of the polar angle of the CM momentum of the signal B
`BroeGamL3`	30265	The third normalized Legendre moment of ROE boosted into the CM frame with respect to the ROE thrust axis
`BzCP`	24146	The z component of the decay vertex of the signal B
`GammabestEtaLR`	23109	Likelihood Ratio based veto for $\eta \to \gamma\gamma$ events
`BzTag`	22016	The z component of the decay vertex of the tag B
`BRoeMissMass`	21892	Missing mass in the rest of the event

Variable name	**Number of decision splits**	**Variable definition**
`OmegahelicityAngle`	21876	The angle between the normal of the $\pi^+\pi^-$ plane and the B flight direction, both in the ω rest frame. (For the $(\rho/K^*)\gamma$ modes, it is the flight direction of a ρ/K^* daughter wrt the direction of the reconstructed B meson in the ρ/K^* rest frame)
`GammabestPi0LR`	21198	Likelihood Ratio based veto for $\pi \to \gamma\gamma$ events
`OmegacosthCM`	21131	Cosine of the polar angle of the CM momentum of the ω
`BRoePerpPBeam`	20988	Rest of event momentum perpendicular to the beam direction
`BrecoilR2prime50`	16867	R2 of EMC bumps of energy greater than 50 MeV in the γ recoil system
`BRoeAplanarity`	16408	Aplanarity of the rest of the event
`BKaonValueMor`	16024	Kaon-based B-tag
`Pi0invMass`	15723	Invariant mass of the $\pi^0 \to \gamma\ \gamma$ system
`BroeGamL2`	15178	The second normalized Legendre moment of ROE boosted into the CM frame with respect to the ROE thrust axis
`BRoeDeltaE`	15092	The ΔE of the tag B

Variable name	**Number of decision splits**	**Variable definition**
GammadistCh50	14566	The centroid distance to the nearest charged EMC bump of energy greater than 50 MeV
GammadistNe	14508	The centroid distance to the nearest neutral EMC bump
BRoePerpPThrust	14417	Rest of event momentum perpendicular to the thrust axis
BrecoilR2primeCh	13397	R2 of charged tracks in the γ recoil system
BKaonSlowPionValue04	13336	K-π B-tag
GammaZ20	13063	The absolute value of the complex Zernike(2,0) moment
Pi0cosHelicity	12610	The cosine of the π^0 helicity angle
BcosAngGamThrust	12250	The cosine of the angle of the γ wrt the ROE thrust axis
BRoeKinLepTracks2P3CM	12245	
BRoeKinLepTracks1CosThetaMissCM	12154	
GammaA42	12018	The absolute value of the complex Zernike(4,2) moment
Gammas9s25	11891	The ratio of the sums of the energies of the central 9 EMC crystals to the central 25 crystals surrounding the centroid
BRoeKinLepTracks2E90W	11040	
BRoePlanarity	10473	The planarity of the rest of the event
BRoeSphericity	10420	The spericity of the rest of the event

Variable name	**Number of decision splits**	**Variable definition**
BroeGamL1	10166	The first normalized Legendre moment of ROE boosted into the CM frame with respect to the ROE thrust axis
GammaLat	10132	The lateral moment of the photon shower shape
BRoeKinLepTracks1E90W	9677	
BKinematicLeptonValue04	9643	A lepton-based B-tag
BMaxPstarValueMor	8100	Maximum CM momentum B-tag
BRoeKinLepTracks1P3CM	7972	
BroeThrL2	7286	The second normalized Legendre moment of the thrust of the rest of the event
BroeThrL1	6477	The first normalized Legendre moment of the thrust of the rest of the event
BrecoilR2prime30	5460	R2 of EMC bumps of energy greater than 30 MeV in the γ recoil system
BRoeMissECM	5315	CM missing energy of the rest of the event
BRoeElectron1Pid	5283	
BroeThrL3	5175	The third normalized Legendre moment of the thrust of the rest of the event
BRoeElectron1P3CM	5129	

Variable name	**Number of decision splits**	**Variable definition**
BRoeKShort1P3CM	4631	
BRoeKCharged1P3CM	4557	
BRoeElectron1E90W	4386	
BRoeMuon1E90W	2732	
BRoeElectron1CosThetaMissCM	2571	
BRoeMuon1P3CM	2541	
BRoeKCharged2P3CM	1585	
BRoeMuon1CosThetaMissCM	1464	

Table 5.8: Definitions of classifier (Bagger) training input variables for $B^0 \to \omega\gamma$ (some rest-of-event variables are self-explanatory). Variables are analogous for the $B \to \rho\gamma$ modes.

$\rho^0\gamma$		$\rho^\pm\gamma$	
Number of decision splits	**Variable name**	**Number of decision splits**	**Variable name**
79532	BdeltaZFitOverErr	98566	BdeltaZFitOverErr
54972	R2All	69787	R2All
51145	BcosThetaCM	56497	BroeGamL3
45449	BroeGamL3	51676	BcosThetaCM
40920	Rho0helicityAngle	49988	Pi0cosGammaAngle
40622	BzCP	48680	BRoePerpPBeam
35767	BRoeMissMass	46173	BRoeMissMass
34995	GammabestEtaLR	43944	RhohelicityAngle
34703	BRoePerpPBeam	43646	BKaonValueMor
32832	BzTag	38990	GammabestEtaLR
32789	GammabestPi0LR	37764	GammabestPi0LR
31739	Rho0chi2Prob	34803	BrecoilR2prime50
29608	BrecoilR2prime50	34322	BRoeAplanarity
29588	BKaonValueMor	32624	RhocosthCM
27603	Rho0costhCM	31459	BRoePerpPThrust

Table 5.9: Top 15 Bagger inputs for $\rho^0\gamma$ and $\rho^\pm\gamma$ modes

6 The fit strategy

6.1 General fitter configuration

The fits for all modes, including the control sample $B \to K^*\gamma$, were carried out using the `RooFit`[33] framework though the `RooRarFit`[34] configuration package. All fits employ the extended unbinned maximum likelihood technique, where the combined probability density function (PDF) for all the events, built from signal and background components, is augmented by a Poisson variation factor for the signal yield. Thus the yield becomes one of the parameters of the fit and can be extracted directly from the result.

Various components were included in the fit procedure at different times (as will be described in subsequent sections) for different modes. In the end, we settled on a rather simple 2-dimensional fit using $m_{\rm ES}$ and ΔE only. In this configuration, the two dimensions are taken as statistically independent, resulting in a combined likelihood function that is a product of two uncorrelated PDFs for each fit model component.

While the (transformed) output of the neural network was used as one of the fit components in the last round of this analysis[15], the shape of the Bagger output for signal is changing far too gradually in the fit region, and the overall number of signal events is too small. We therefore decided to impose a harder cut on the raw Bagger output, leaving the $m_{\rm ES}$-ΔE fit region rather clean. This cut was varied for combined, luminosity-weighted Monte Carlo fit for each signal mode, until the signal significance ($S/\sqrt{S+B}$) was maximized. The values obtained using this procedure were $bag > 0.94$ and $bag > 0.93$ for the ρ and ω modes, respectively. In the previous analysis, the cut on the NN output was similarly optimized, but then subsequently

relaxed to allow more events in the fit region, since the output itself was one of the fit components.

Meson helicity angles were also used in the previous analysis, but dropped this time around in favor of keeping this information as one of the inputs to the Bagger. While Bagger performance did not suffer significantly when helicity angles were removed, as mentioned earlier, including these angles as another fit dimension did not help the performance either. And, due to the similarity between the signal and background helicity shapes after the tight Bagger cut is imposed, it seems highly unlikely that this variable will be of any use in the fit.

We also attempted to use the Dalitz angle as another fit dimension in the $B^0 \to \omega\gamma$ mode, but the shape turned out to be too similar to that of the continuum background component, leading to zero gain in performance. Therefore, unlike the previous round of analysis, this variable was in the end dropped completely.

In general, the setup is fundamentally different from the previous round of this analysis, where the approach was to cut loosely on the NN, leaving a (relatively) large number of events in the fit region, and to attempt to decompose the resulting distribution into signal and background components using a sophisticated multidimensional fit. We now go in the opposite direction by allowing the multidimensional classifier to do most of the work separating signal from background, resulting in a sparsely populated fit region and allowing for a simpler 2-dimensional fit.

6.2 $B^0 \to \omega\gamma$ fitter

As can be seen in Figure 6.38, the PDF shapes for the three components of the likelihood function (signal, $B\overline{B}$, and continuum) were determined from luminosity-weighted MC samples. The functions used were: Crystal Ball [35] (signal and $B\overline{B}$) and ARGUS [36] (continuum) in $m_{\rm ES}$, Cruiff (signal), exponential ($B\overline{B}$), and a line (continuum) in ΔE.[4]

[4]In all signal-mode ML fits, the only floating parameters are: signal yield, continuum yield, and parameters of the continuum PDFs in both $m_{\rm ES}$ and ΔE. In $B \to K^*\gamma$ fits with Bagger cut applied, we also float signal mean and width(s).

The Crystal Ball function is parametrized as:

$$f(x|\mu,\sigma,\alpha,n) = C \cdot \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & , \frac{x-\mu}{\sigma} > \alpha \\ \left(\frac{n}{\alpha}\right)^n \exp\left(-\frac{\alpha^2}{2}\right)\left(\frac{n}{\alpha} - \alpha + \frac{x-\mu}{\sigma}\right)^{-n} & , \frac{x-\mu}{\sigma} \leq \alpha \end{cases} \tag{6.6}$$

where α determines where the usual Gaussian turns into a power function with the tail parameter n, and C is the overall normalization.

The ARGUS function is parametrized as:

$$f(x|b, E_{beam}) = C \cdot \left(1 - \frac{x^2}{E_{beam}^2}\right)^{\frac{1}{2}} \exp\left[-b\left(1 - \frac{x^2}{E_{beam}^2}\right)\right] \tag{6.7}$$

where E_{beam}, the beam energy, is the endpoint of the spectrum, and b is the shape parameter.

The Cruiff function is used internally in the *BABAR* collaboration and is parametrized as:

$$f(x|\mu,\sigma_{L,R},\alpha_{L,R}) = C \cdot \exp\left(\frac{-(x-\mu)^2}{2\sigma_{L,R}^2 + \alpha_{L,R}(x-\mu)^2}\right), \tag{6.8}$$

where $\sigma_{L,R}$ are the left and right-side width of this modified Gaussian, and $\alpha_{L,R}$ are the measures of the corresponding tails.

Figure 6.39 shows the result of a luminosity-sampled combined MC fit (i.e., each type of MC was randomly sampled to select the number of events expected in our on-peak datset). Signal and background components are also shown with dashed lines. The numbers are summarized in Table 6.10 for all three signal modes. Note that the number of $B\overline{B}$ events is fixed from PDF fit to $B\overline{B}$ MC. This is in fact done for all three modes by integrating the $m_{\rm ES}$ distribution of $B\overline{B}$ MC shown in Figures 6.38, 6.41, and 6.44.

The fits were validated with both pure and signal-embedded toy MC (1000 experiments each). The results for these in $B^0 \to \omega\gamma$ mode are shown in Figure 6.40 and display no significant biases. Toy MC experiments produce a distribution of signal event yields centered on the expected number, eliminating the need for any corrections.

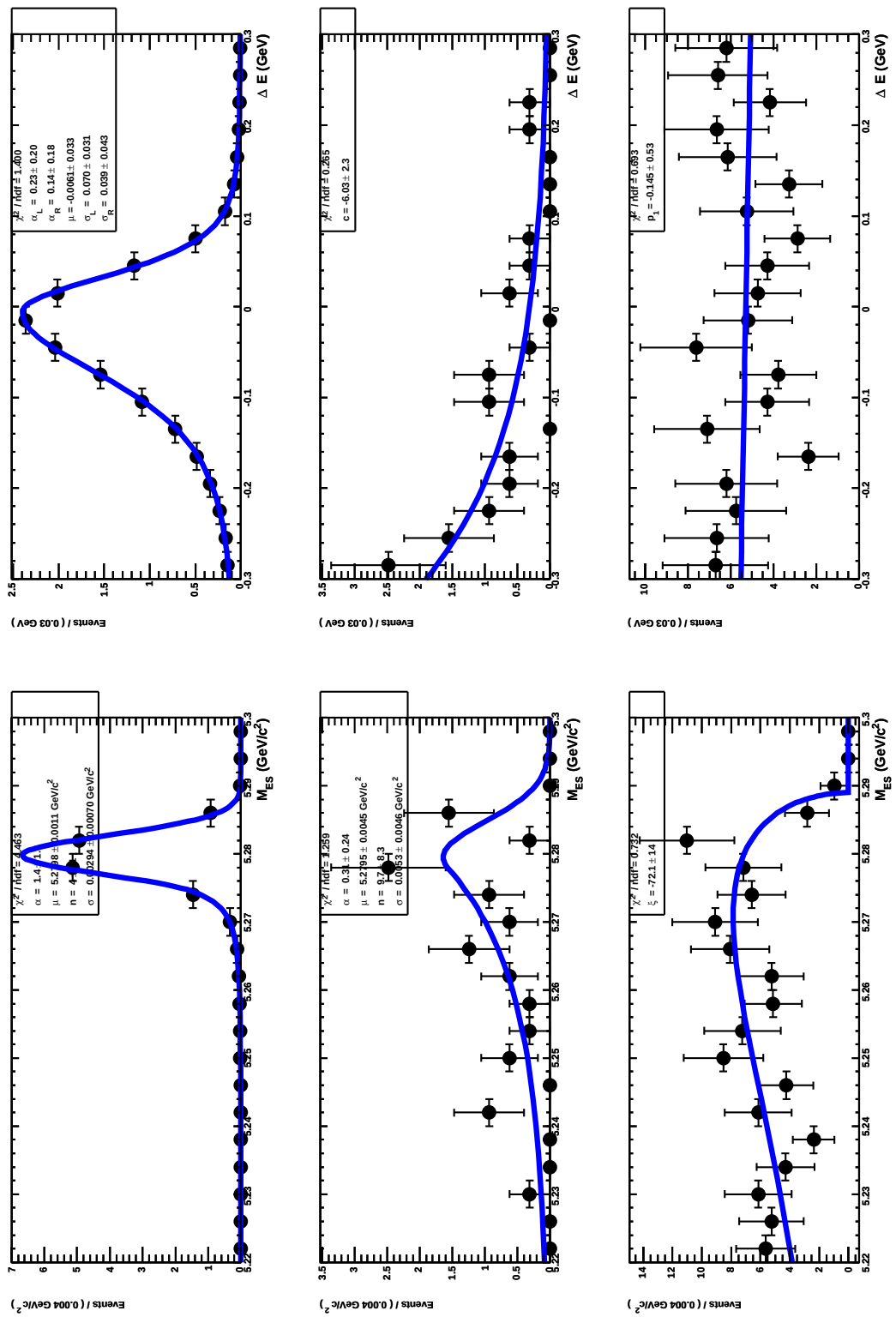


Figure 6.38: PDF shapes for $m_{\rm ES}$ (left) and ΔE (right) determined from Monte Carlo for the $B^0 \to \omega\gamma$ mode. Signal on top, $B\overline{B}$ in the middle, continuum on the bottom

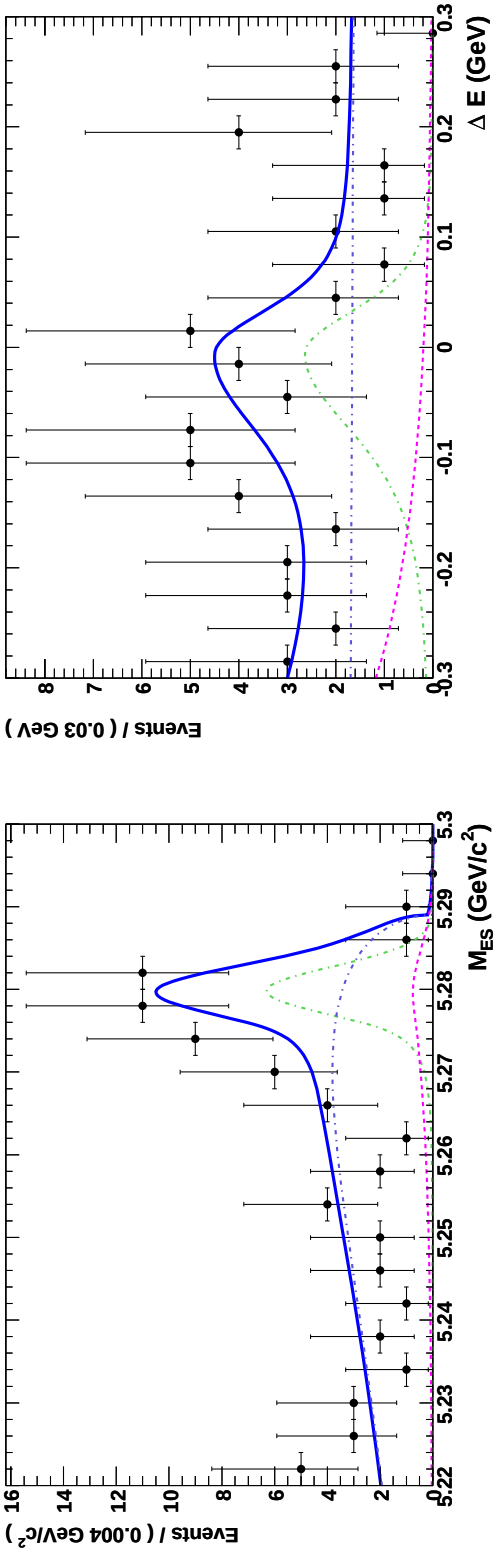


Figure 6.39: Luminosity-sampled, combined Monte Carlo fit for the $B^0 \to \omega\gamma$ mode

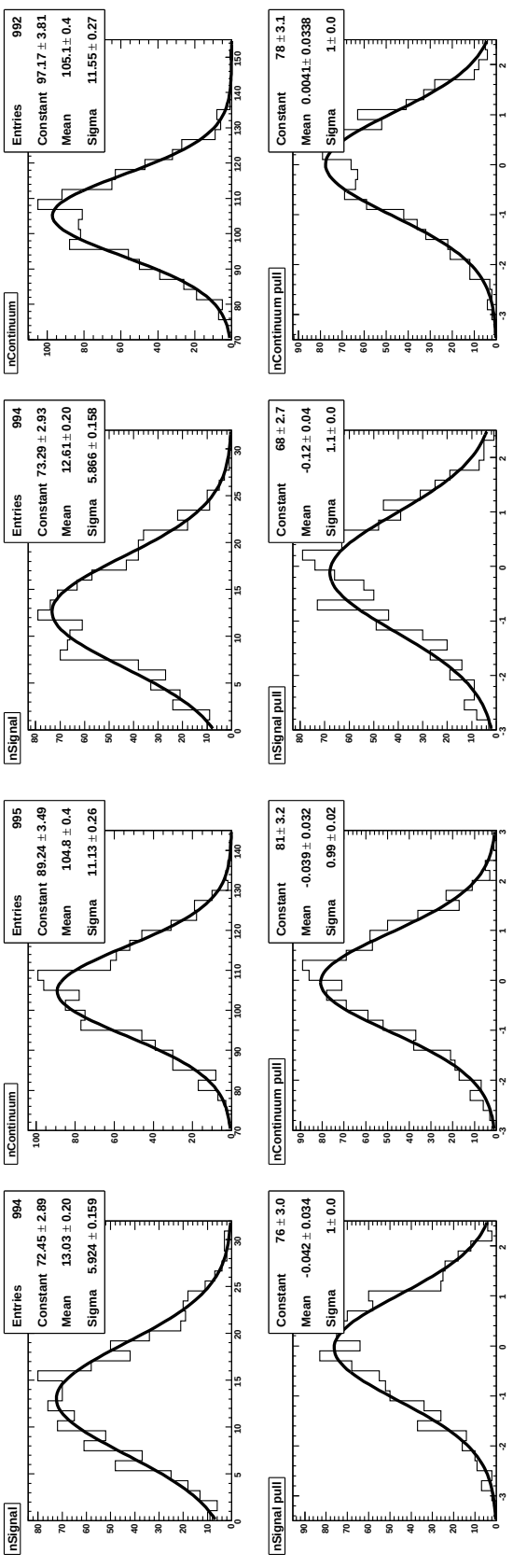


Figure 6.40: Plots of signal and continuum yields (top) and their pulls (bottom) in pure (left four) and signal-embedded (right four) toy MC for the $B^0 \to \omega\gamma$ mode

6.3 $B^0 \to \rho^0\gamma$ fitter

The fitter for this mode is almost analogous to the $B^0 \to \omega\gamma$ one, except that the PDF for the $B\overline{B}$ component in $m_{\rm ES}$ is fit only in the range of $[5.25, 5.3]$. The very few events that appear at the lower range of $m_{\rm ES}$ can be absorbed by the continuum. This also affects the fixed $B\overline{B}$ yield, as the events in the tail are not counted. In addition, the PDF for the $B\overline{B}$ component in ΔE contains a peaking Crystal Ball contribution, which is taken from dedicated $B \to K^*\gamma$ MC (see Section 6.6). The PDF plots are shown in Figure 6.41 and the luminosity-sampled MC fit in Figure 6.42. The plots summarizing the results of toy MC studies are given in Figure 6.43. Again, no significant biases are observed, and no corrections are necessary.

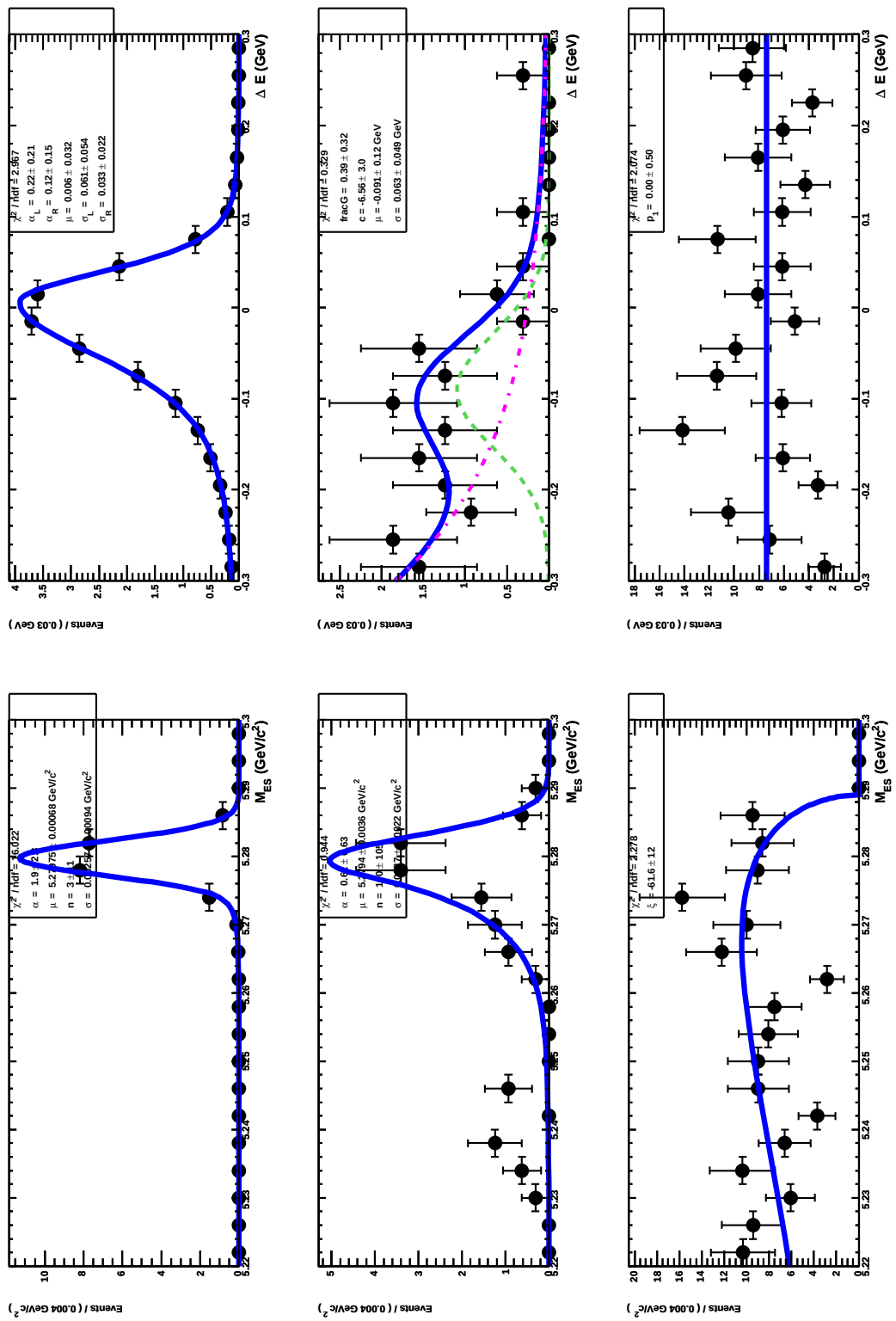


Figure 6.41: PDF shapes for $m_{\rm ES}$ (left) and ΔE (right) determined from Monte Carlo for the $B^0 \to \rho^0\gamma$ mode. Signal on top, $B\overline{B}$ in the middle, continuum on the bottom

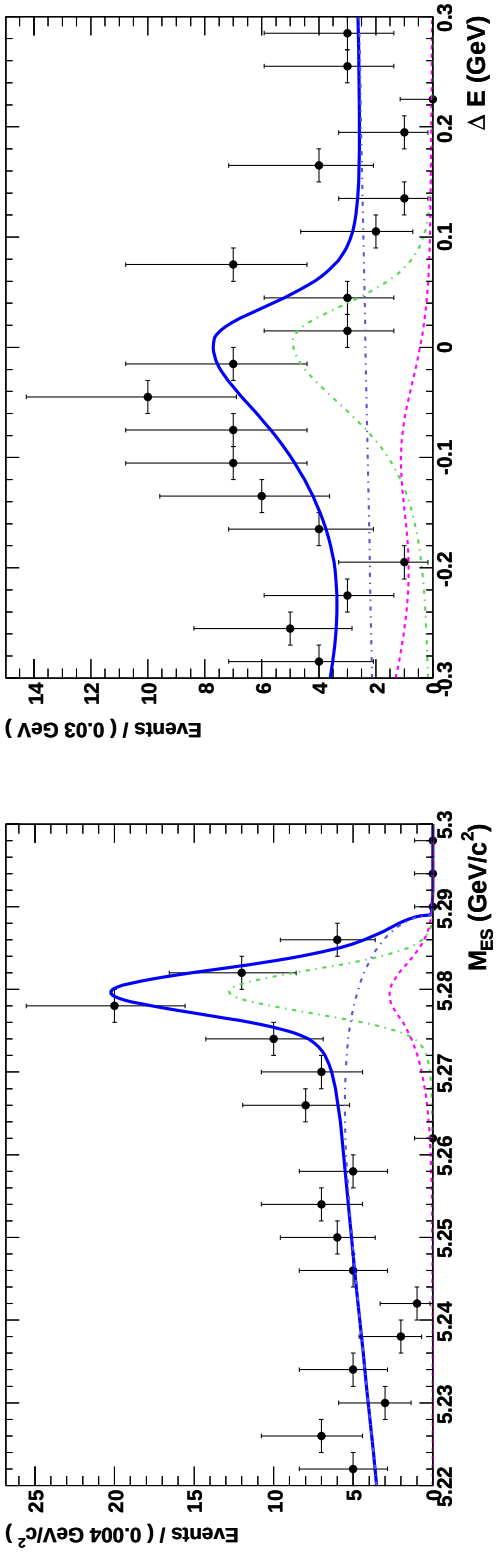


Figure 6.42: Luminosity-sampled, combined Monte Carlo fit for the $B^0 \to \rho^0\gamma$ mode

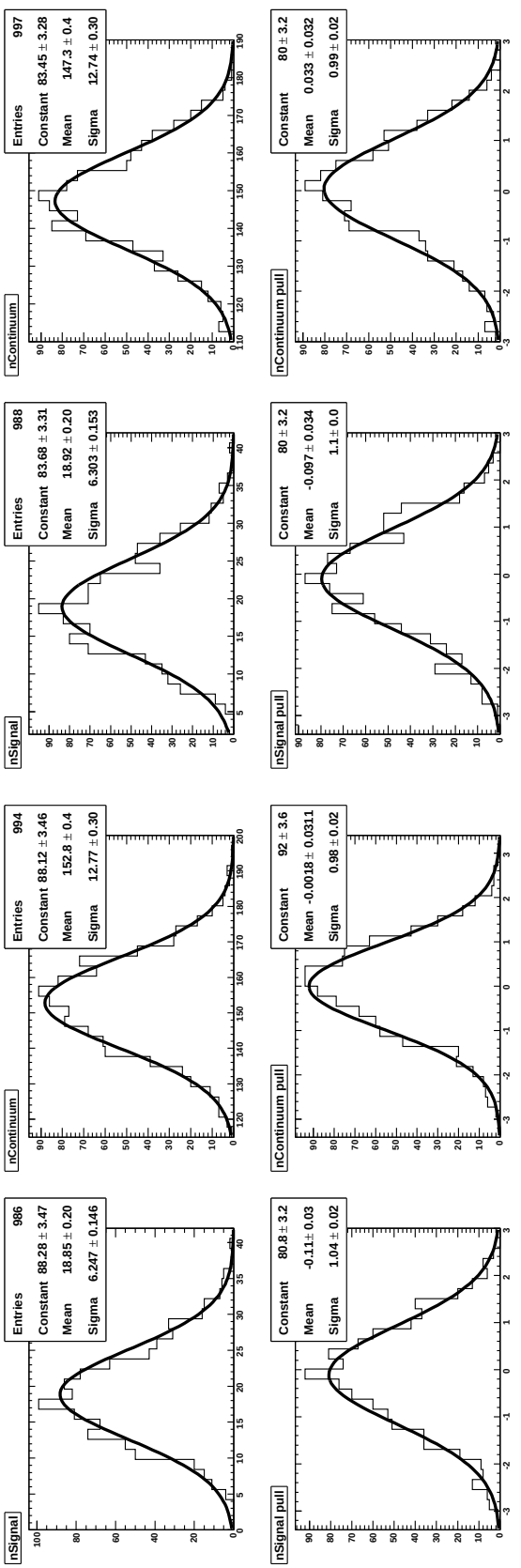


Figure 6.43: Plots of signal and continuum yields (top) and their pulls (bottom) in pure (left four) and signal-embedded (right four) toy MC for the $B^0 \to \rho^0\gamma$ mode

6.4 $B^{\pm} \to \rho^{\pm}\gamma$ fitter

The fitter for this mode is a mix of the fitters for the neutral modes: the PDF for the $B\overline{B}$ component in $m_{\rm ES}$ is given by a Crystal Ball fit over the whole $m_{\rm ES}$ range, as in $B^0 \to \omega\gamma$, while the same component in ΔE is given by an exponential with a Crystal Ball contribution taken from dedicated $B \to K^*\gamma$ and $B^{\pm} \to \rho^{\pm}\pi^0$ MC (see Section 6.6), similar to $B^0 \to \rho^0\gamma$. The PDF projection plots are shown in Figure 6.44 and the luminosity-sampled MC fit in Figure 6.45. The plots summarizing the results of toy MC studies are given in Figure 6.46. As in the other two modes, no significant biases are present, and no corrections are necessary.

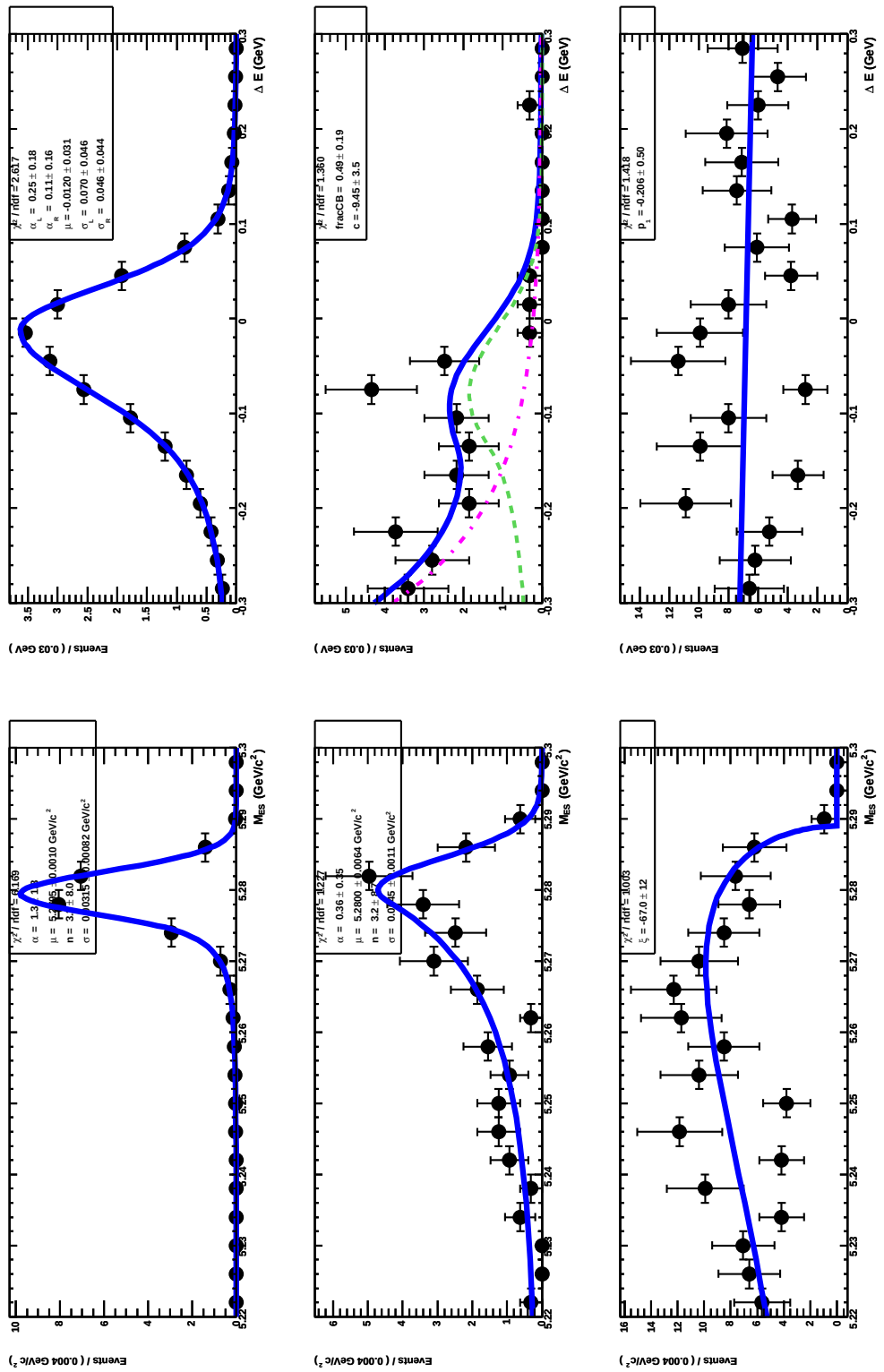


Figure 6.44: PDF shapes for $m_{\rm ES}$ (left) and ΔE (right) determined from Monte Carlo for the $B^{\pm} \to \rho^{\pm}\gamma$ mode. Signal on top, $B\overline{B}$ in the middle, continuum on the bottom

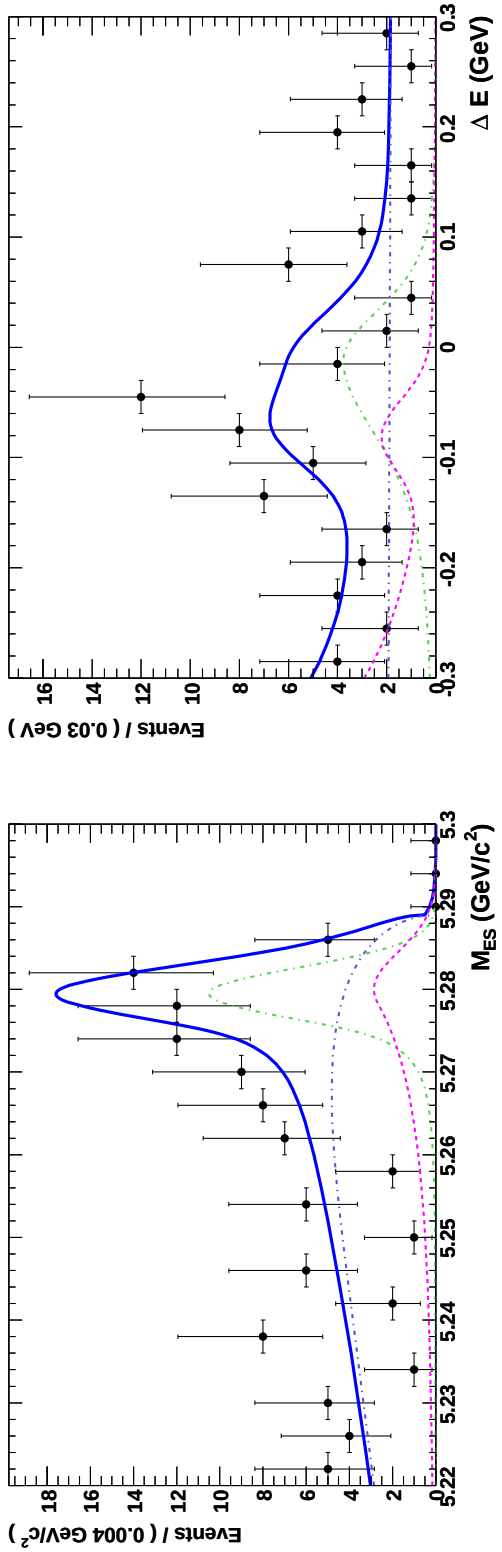


Figure 6.45: Luminosity-sampled, combined Monte Carlo fit for the $B^{\pm} \to \rho^{\pm}\gamma$ mode

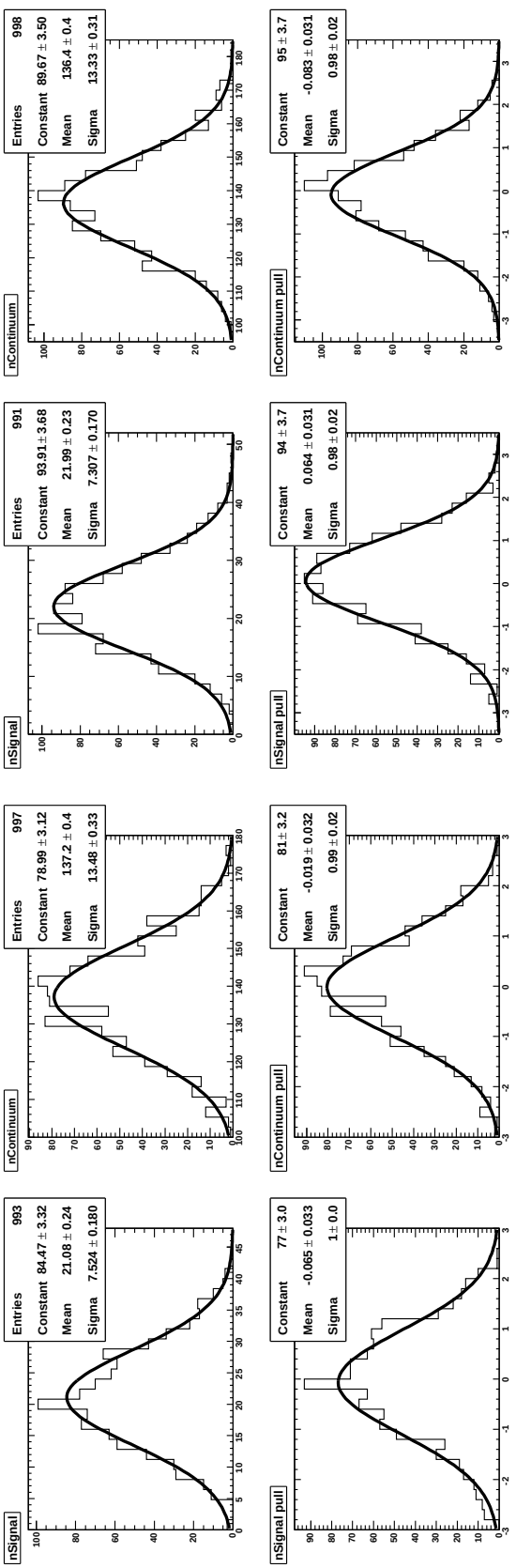


Figure 6.46: Plots of signal and continuum yields (top) and their pulls (bottom) in pure (left four) and signal-embedded (right four) toy MC for the $B^{\pm} \to \rho^{\pm}\gamma$ mode

	N_{signal}	$N_{B\bar{B}}$ **(fixed)**	$N_{udsc\tau}$	**Significance**
Sampled MC				
$B^0 \to \omega\gamma$	$12.3^{+6.2}_{-5.4}$ (13 expected)	10.2	$104.9^{+12.1}_{-11.3}$ (105 expected)	2.6σ
$B^0 \to \rho^0\gamma$	$19.1^{+6.8}_{-6.1}$ (19 expected)	10.5	$150.6^{+13.7}_{-12.9}$ (148 expected)	3.9σ
$B^\pm \to \rho^\pm\gamma$	$21.2^{+8.2}_{-7.4}$ (21 expected)	26.0	$135.6^{+14.2}_{-13.4}$ (137 expected)	3.2σ
Weighted MC				
$B^0 \to \omega\gamma$	$15.4^{+0.0}_{-5.6}$	10.2	$102.7^{+11.9}_{-11.2}$	3.2σ
$B^0 \to \rho^0\gamma$	$21.0^{+6.8}_{-6.0}$	10.5	$148.6^{+13.5}_{-12.8}$	4.4σ
$B^\pm \to \rho^\pm\gamma$	$19.0^{+7.8}_{-7.0}$	26.0	$138.2^{+14.2}_{-13.4}$	3.1σ

Table 6.10: Summary of luminosity-sampled and luminosity-weighted combined Monte Carlo fits for all three signal modes

6.5 $B \to K^*\gamma$ fitter for cross check

The purpose of performing these cross-check fits is twofold. First, we want to make sure we extract the correct branching fractions for these previously measured decays ($B^0 \to K^{*0}\gamma$ ($K^{*0} \to K^+\pi^-$) and $B^\pm \to K^{*\pm}\gamma$ ($K^{*\pm} \to K^\pm\pi^0$)) and make sure that on-peak data agrees well with our combined MC sample. Second, we want to extract the fit parameters (in particular the means and the widths of the signal distributions) from both MC and on-peak fits and use the differences as correction factors for the final on-peak data fits in our signal modes.

The kinematics of $B \to K^*\gamma$ decays are almost identical to our signal modes. This allows us to use the $B \to K^*\gamma$ sample for several validation studies, including Bagger and fitter cross checks. Figures 6.47 and 6.48 show overlay plots of normalized signal distributions in $m_{\rm ES}$ and ΔE for $B^0 \to \rho^0\gamma$ vs $B^0 \to K^{*0}\gamma$ and $B^\pm \to \rho^\pm\gamma$ vs $B^\pm \to K^{*\pm}\gamma$. Plots in Figure 6.49 show the distributions of Bagger output in signal and generic MC for the same four modes. These demonstrate that the $B \to K^*\gamma$ and

$B \to \rho\gamma$ signal shapes are very similar, as are the multivariate classifier outputs used for final event selection.

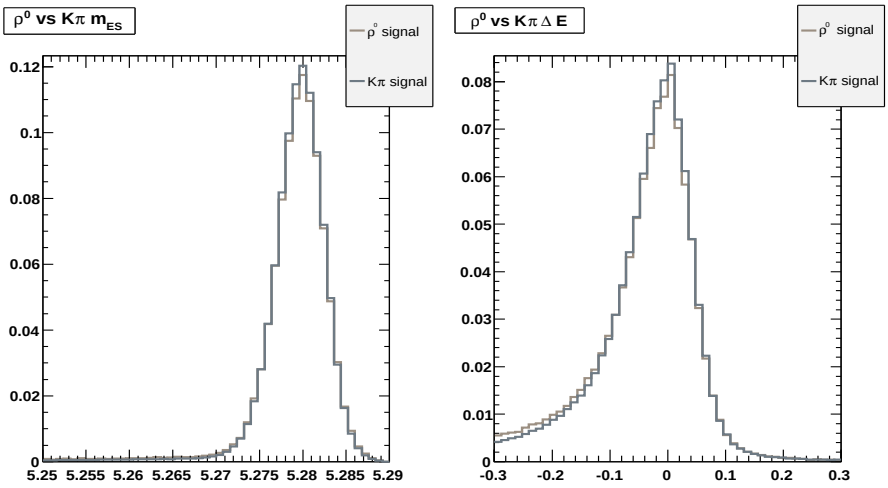


Figure 6.47: Signal MC $m_{\rm ES}$ and ΔE for $B^0 \to \rho^0\gamma$ vs $B^0 \to K^{*0}\gamma$

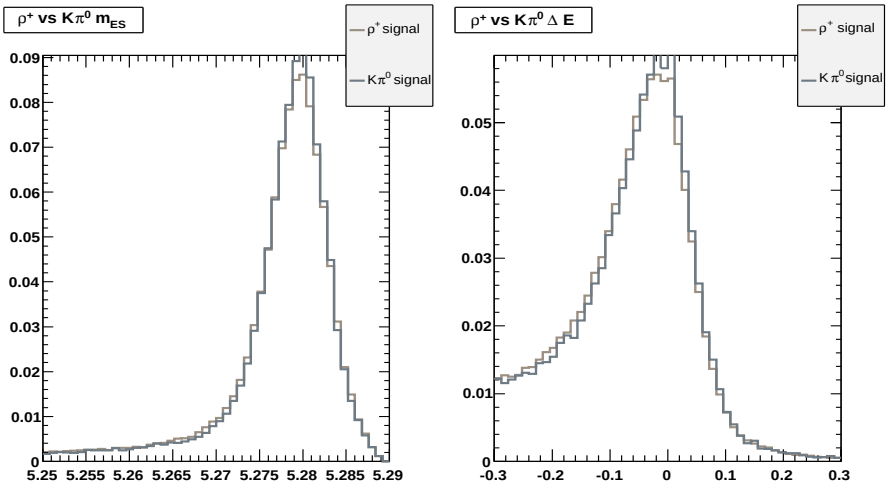


Figure 6.48: Signal MC $m_{\rm ES}$ and ΔE for $B^\pm \to \rho^\pm\gamma$ vs $B^\pm \to K^{*\pm}\gamma$

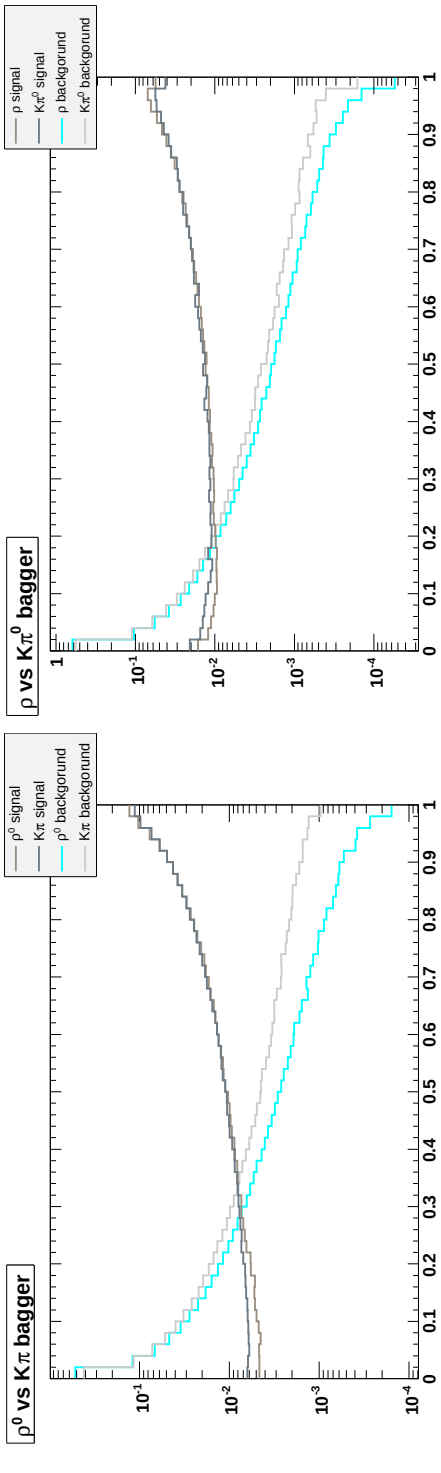


Figure 6.49: Signal and generic MC Bagger output for $B^0 \to \rho^0\gamma$ vs $B^0 \to K^{*0}\gamma$ and $B^\pm \to \rho^\pm\gamma$ vs $B^\pm \to K^{*\pm}\gamma$

Since the statistics for $B \to K^*\gamma$ are much higher than for our signal modes, we did not use the full on-peak dataset for the cleaner $K^+\pi^-$ mode, where we only used *BABAR* runs 1–5 data. We did, however, include run 6 data for $K^\pm\pi^0$ mode, which has a much lower signal-to-noise ratio. This is simply a result of trying to avoid potential technical problems associated with upgrading to a new *BABAR* dataset and the associated release version.

The fitter configurations for the two $B \to K^*\gamma$ modes are again very similar to those described above. In $K^+\pi^-$ mode, we have a small peaking Gaussian component for the $B\overline{B}$ PDF in ΔE, similar to $B^0 \to \rho^0\gamma$. For $K^+\pi^0$ mode, we perform two different sets of fits. The first uses the same PDFs as the $B^0 \to \omega\gamma$ fit. The results of this are used to correct the signal shape parameters in $B^0 \to \omega\gamma$ and $B^\pm \to \rho^\pm\gamma$ modes, based on data-MC differences. The errors on the signal shape parameters obtained from this fit are used for systematic variations of fixed signal shape parameters in the above two signal modes. For the other fit, we split the signal into truth-matched and self-crossfeed parts. This is done because the two differ quite significantly if no cut on the Bagger output is made. The motivation for this split comes from the fact that $m_{\rm ES}$-ΔE correlation is much stronger for the self-crossfeed part, thus leading to a large bias in the fit if it is not handled separately. The separation cuts this bias in half. This becomes important when we extract signal yield from $B \to K^*\gamma$ to calculate Bagger efficiency systematic error (discussed in Section 9.4), where we need to perform the fit both with and without the Bagger cut applied.

		N_{signal}	$N_{udsc\tau}$	$m_{\rm ES}$ μ	$m_{\rm ES}$ σ	ΔE μ	ΔE σ_L	ΔE σ_R
$K^+\pi^-$	**MC**	987 ± 33	124 ± 19	5.2797 ± 0.00009	0.0026 ± 0.00007	0.0064 ± 0.0040	0.0590 ± 0.0037	0.0324 ± 0.0026
	Data	1036 ± 33	122 ± 17	5.2794 ± 0.00009	0.0026 ± 0.00007	-0.0060 ± 0.0038	0.0469 ± 0.0033	0.0368 ± 0.0023
$K^+\pi^0$	**MC**	195 ± 17	34 ± 10	5.2798 ± 0.0003	0.0029 ± 0.0002	-0.0034 ± 0.0088	0.0676 ± 0.0086	0.0382 ± 0.0061
	Data	239 ± 15	25 ± 7	5.2794 ± 0.0002	0.0028 ± 0.0002	-0.0133 ± 0.0094	0.0534 ± 0.0081	0.0377 ± 0.0056

Table 6.11: Summary of luminosity-weighted combined Monte Carlo and on-peak data fits for $B \to K^*\gamma$ ($K^{*0} \to K^+\pi^-$and $K^{*+} \to K^+\pi^0$) cross check.

The PDF projection plots for $B \to K^*\gamma$ are shown in Figures 6.50 and 6.52, while the luminosity-weighted MC and on-peak data fits are shown in Figures 6.51 and 6.53. Table 6.11 summarizes the results (yields and PDF parameters). Note that the on-peak luminosity used for $K^{\pm}\pi^0$ mode was 376.2 fb^{-1} (run 6 data included), while MC was matched to only 341.8 fb^{-1} (runs 1–5) — the on-peak luminosity used for $K^+\pi^-$ mode. With signal efficiencies of 10.6% and 4.2% and on-peak luminosities of 341.8 fb^{-1} and 376.2 fb^{-1} for the $K^+\pi^-$ and $K^+\pi^0$ modes respectively, we obtain:

$$\begin{aligned}
\mathcal{B}(B^0 \to K^{*0}\gamma\ (K^{*0} \to K^+\pi^-)) &= (4.11 \pm 0.13) \times 10^{-5} \quad (Data) \\
\mathcal{B}(B^0 \to K^{*0}\gamma\ (K^{*0} \to K^+\pi^-)) &= (3.89 \pm 0.13) \times 10^{-5} \quad (MC) \\
\mathcal{B}(B^{\pm} \to K^{*\pm}\gamma\ (K^{*+} \to K^+\pi^0)) &= (4.34 \pm 0.28) \times 10^{-5} \quad (Data) \\
\mathcal{B}(B^{\pm} \to K^{*\pm}\gamma\ (K^{*+} \to K^+\pi^0)) &= (3.88 \pm 0.34) \times 10^{-5} \quad (MC)
\end{aligned}$$

which compare favorably with the generated MC, based on world averages of $4.01 \pm 0.20 \times 10^{-5}$ and $4.03 \pm 0.26 \times 10^{-5}$ for $B^0 \to K^{*0}\gamma$ and $B^{\pm} \to K^{*\pm}\gamma$, respectively [26].

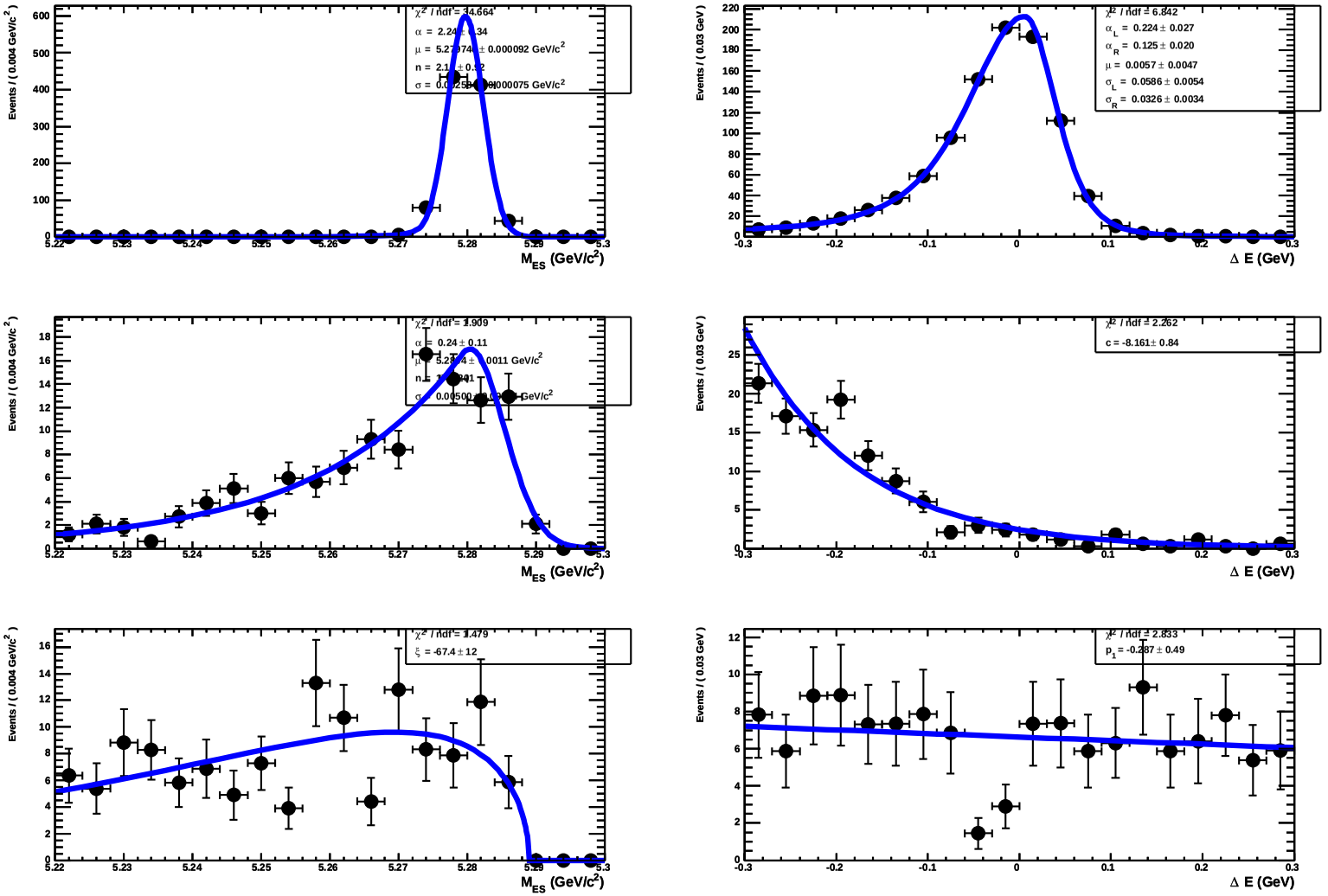


Figure 6.50: PDF shapes for $m_{\rm ES}$ (left) and ΔE (right) determined from Monte Carlo for the $B^0 \to K^{*0}\gamma$ mode. Signal on top, $B\overline{B}$ in the middle, continuum on the bottom

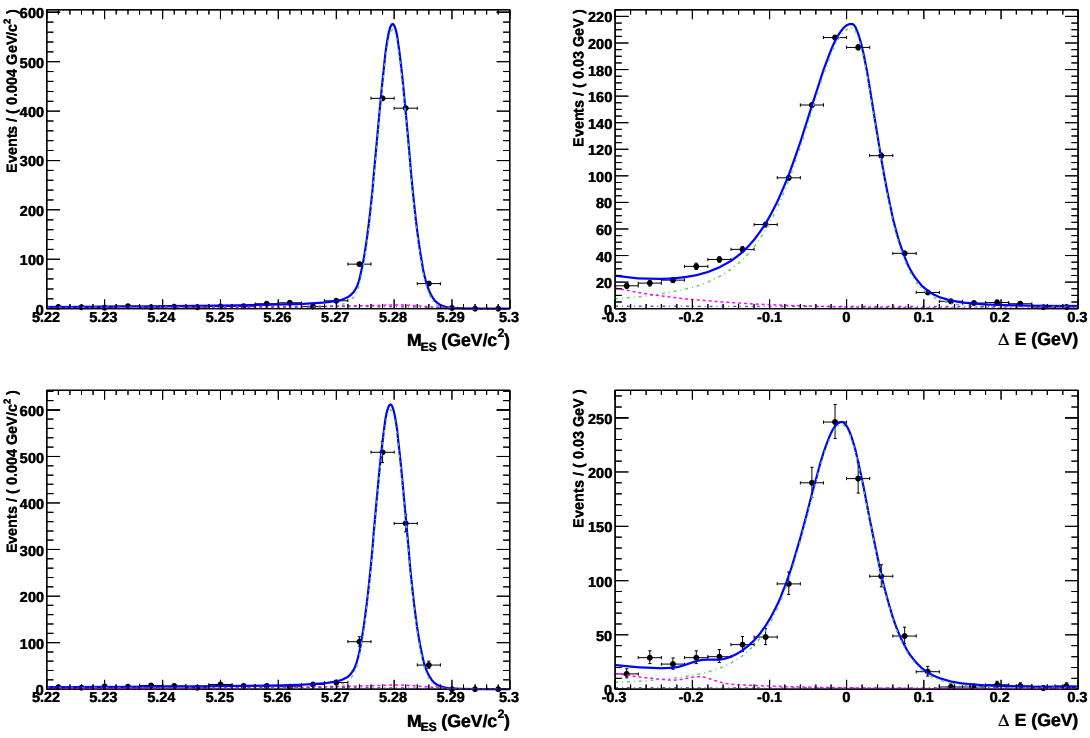


Figure 6.51: Luminosity-weighted, combined Monte Carlo (top) and on-peak data (bottom) fits for the $B^0 \to K^{*0}\gamma$ mode

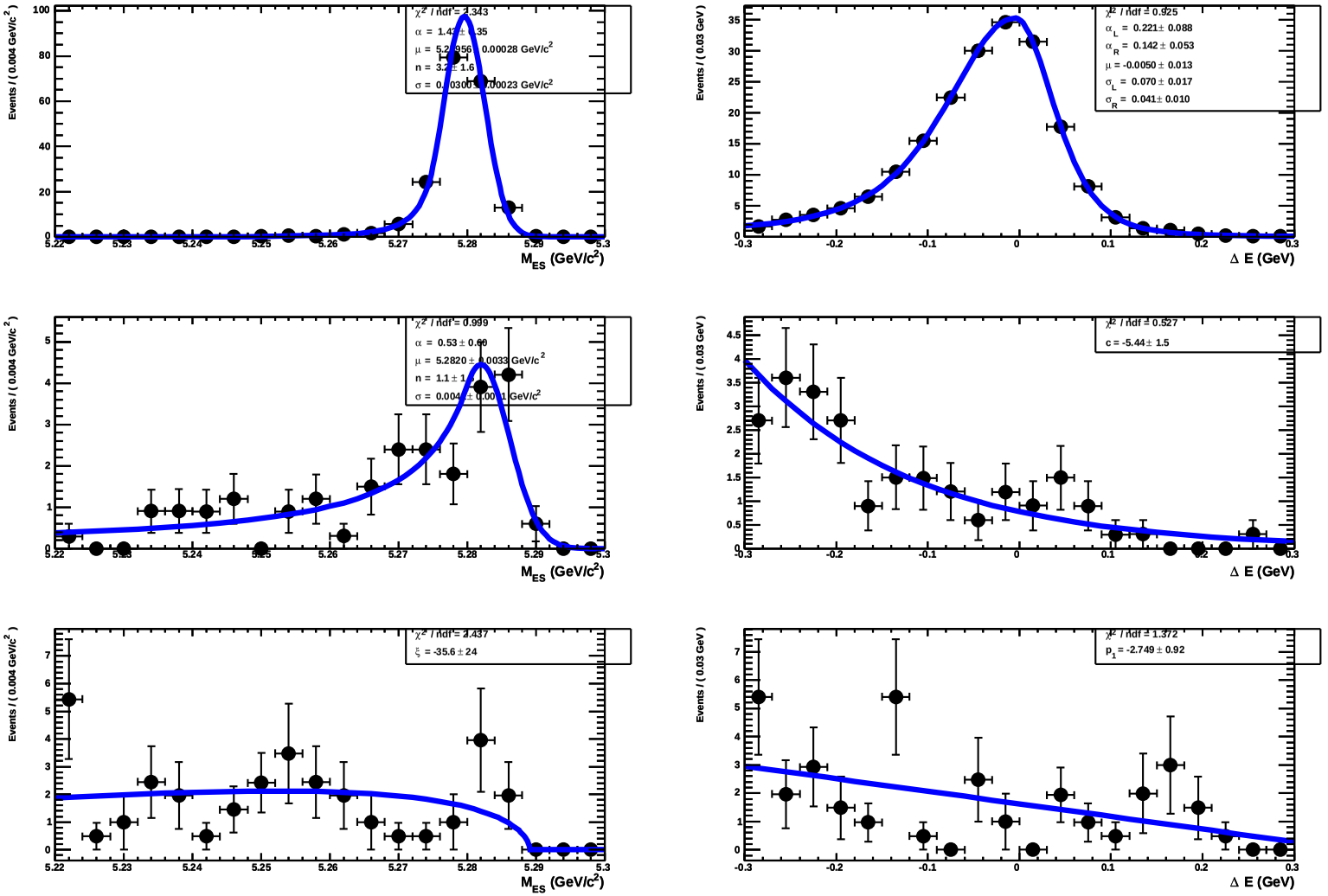


Figure 6.52: PDF shapes for $m_{\rm ES}$ (left) and ΔE (right) determined from Monte Carlo for the $B^{\pm} \to K^{*\pm}\gamma$ mode. True signal on top, self-crossfeed signal followed by $B\overline{B}$ in the middle, continuum on the bottom

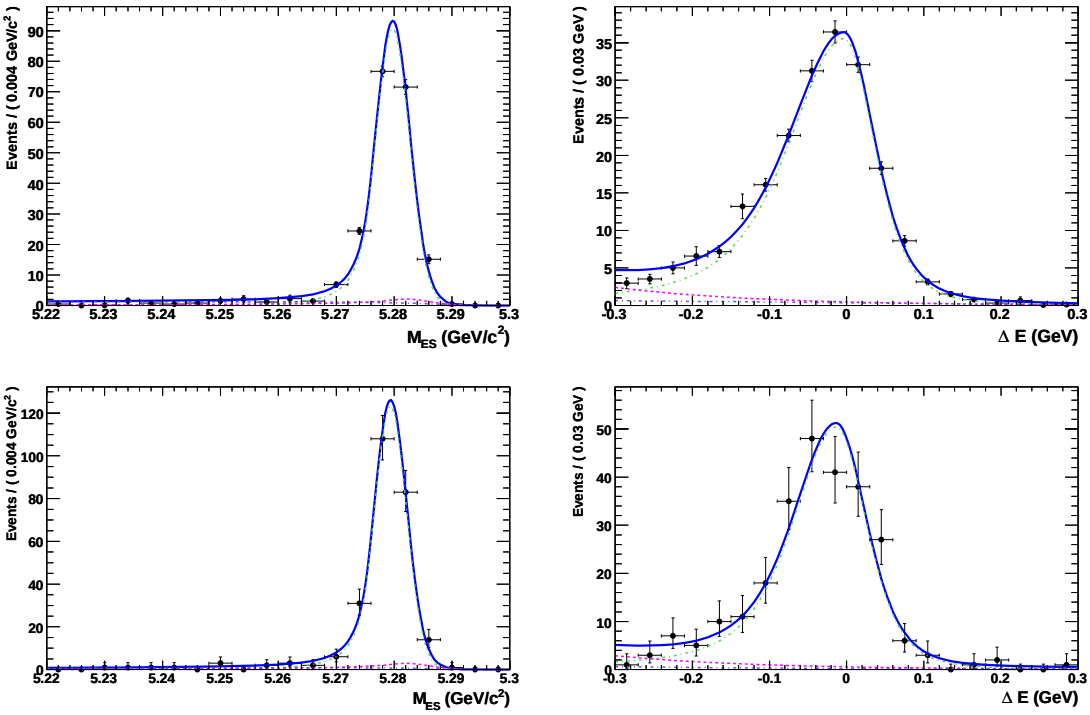


Figure 6.53: Luminosity-weighted, combined Monte Carlo (top) and on-peak data (bottom) fits for the $B^{\pm} \to K^{*\pm}\gamma$ mode

6.6 Peaking $B\overline{B}$ component in $B \to \rho\gamma$

We performed a detailed study of the composition of $B\overline{B}$ backgrounds in $B \to \rho\gamma$ modes using dedicated $B \to K^*\gamma$, $B \to X_s\gamma$, $B^\pm \to \rho^\pm\eta$, and $B^\pm \to \rho^\pm\pi^0$ MC samples. Figures 6.54 and 6.58 show distributions of all these various backgrounds after all selection criteria have been applied. We found that $B \to K^*\gamma$ describes completely the peak we observe in the generic $B\overline{B}$ distribution in $B^0 \to \rho^0\gamma$. In $B^\pm \to \rho^\pm\gamma$, however, the story turned out to be a bit more complicated. Here, $B \to K^*\gamma$ contributes only about 2 events, whereas the overall number of $B\overline{B}$ candidates in generic MC is 26. Thus, we had to look at other potential sources of peaking backgrounds, including $B^\pm \to \rho^\pm\pi^0$ and $B^\pm \to \rho^\pm\eta$. The latter has a sizable contribution of about 7 events, but is not peaking as much around $\Delta E = -0.1$. In fact, combined with $B \to X_s\gamma$, it is described well by the exponential component of the $B\overline{B}$ PDF. $B^\pm \to \rho^\pm\pi^0$, on the other hand, peaks in a similar fashion to $B \to K^*\gamma$, and together they account for about 5 or 6 events. We therefore use the combined shape from these two backgrounds to describe the peak in generic $B\overline{B}$ distribution in Section 6.4.

Second row of Figure 6.54 and bottom row of Figure 6.58 show the combined $B\overline{B}$ background distributions for $B^0 \to \rho^0\gamma$ ($B \to K^*\gamma$ and $B \to X_s\gamma$) and $B^\pm \to \rho^\pm\gamma$ ($B \to K^*\gamma$, $B^\pm \to \rho^\pm\eta$, $B^\pm \to \rho^\pm\pi^0$, and $B \to X_s\gamma$). These agree well with the plots in Figures 6.41 and 6.44 in Sections 6.3 and 6.4, respectively. The only noticeable difference is the bin around $\Delta E = -0.09$ in $B^\pm \to \rho^\pm\gamma$ mode, which comes out low in these studies but high in generic $B\overline{B}$ MC. However, it is likely to be a fluctuation in the latter, where we have lower statistics. Otherwise, the fit is forced to produce a peak that seems too narrow, especially compared to $B^0 \to \rho^0\gamma$ mode.

We performed an additional study using the $B^\pm \to \rho^\pm\gamma$ mode to assess the stability of the $B\overline{B}$-component contribution to the fit. The shift of the ΔE $B\overline{B}$ peaks in $B \to \rho\gamma$ modes with respect to the signal peak helps the fit to separate these components. However, there's a possibility that the $m_{\rm ES}$-ΔE correlations are important enough that the nominal PDF, which is a simple product of two 1-dimensional PDFs, does not provide the correct model. This could potentially introduce a bias, because

the ΔE $B\overline{B}$ peak overlaps the signal, particularly in the $B^{\pm} \to \rho^{\pm}\gamma$ mode.

We studied the 2-dimensional $m_{\rm ES}$-ΔE distribution of all $B\overline{B}$ events in $B^{\pm} \to \rho^{\pm}\gamma$ mode that pass our combined candidate selection process. This is shown in Figure 6.59, illustrating that, at least with the current statistics, the correlation is negligible. However, we also performed 500 toy MC experiments embedding random $B\overline{B}$ events, and redoing the fit. The results, shown in Figure 6.60, demonstrate that the fit remains stable, from which we conclude that the current fit model adequately describes the peaking $B\overline{B}$ component.

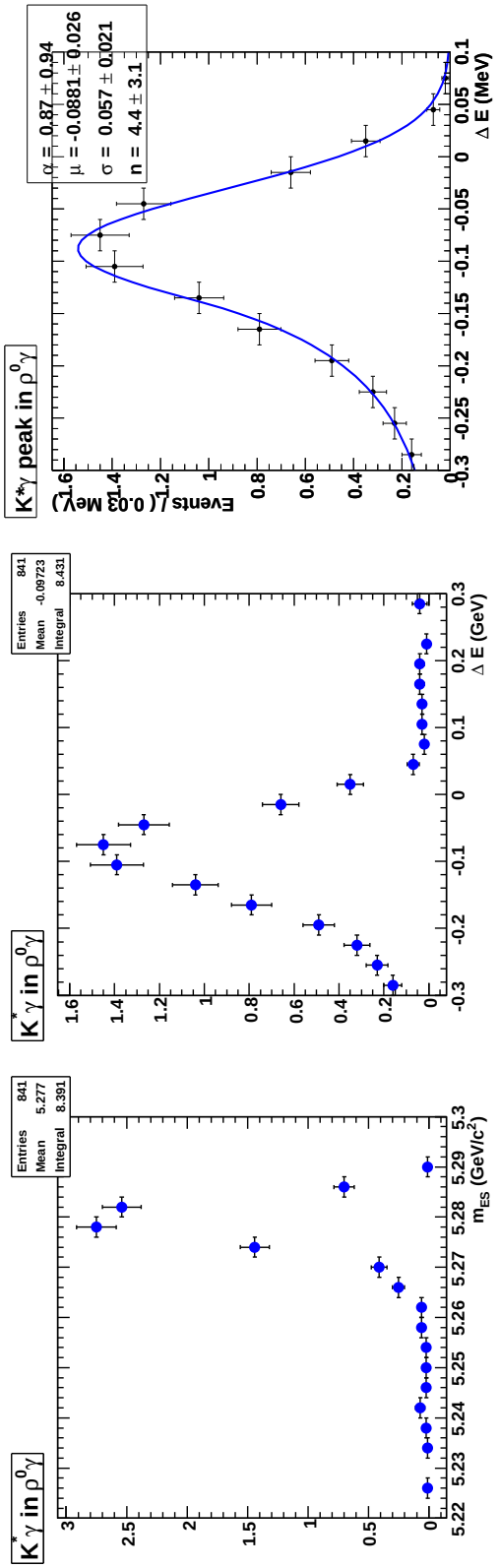


Figure 6.54: Plots for peaking B-background studies ($B \to K^*\gamma$ and $B \to X_s\gamma$ only) in $B \to \rho\gamma$

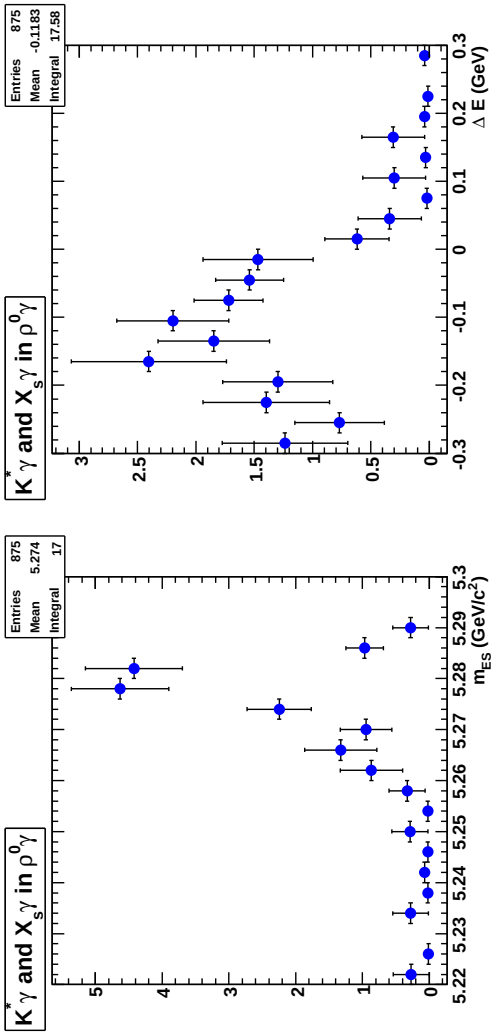


Figure 6.54: Plots for peaking B-background studies continued

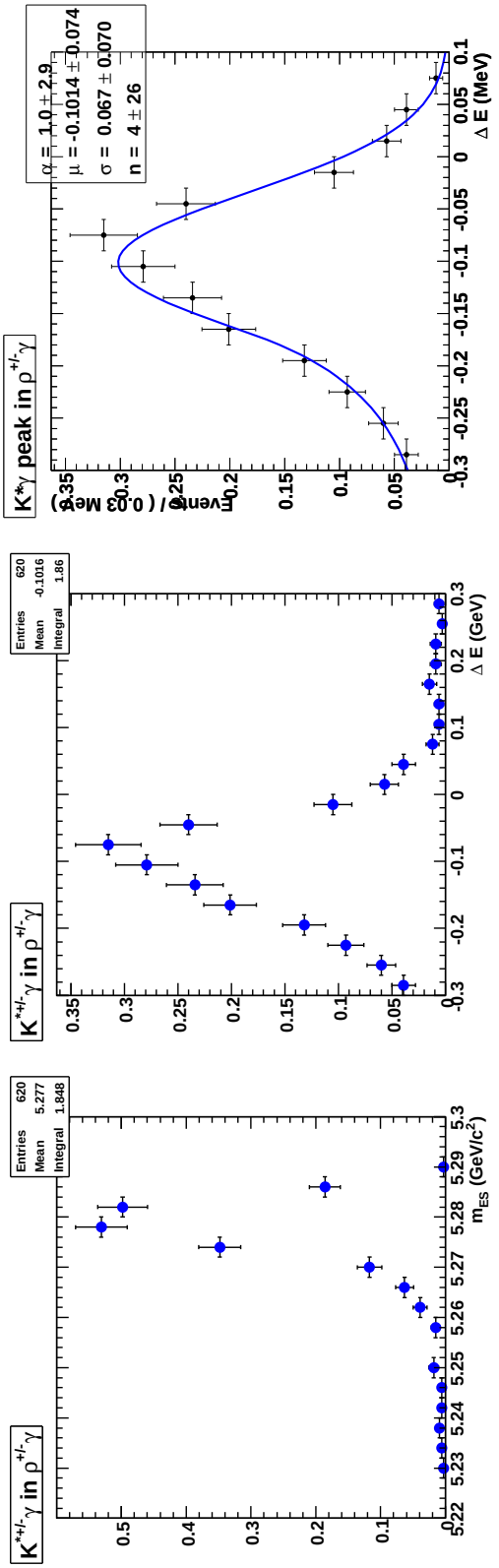


Figure 6.54: Plots for peaking B-background studies continued

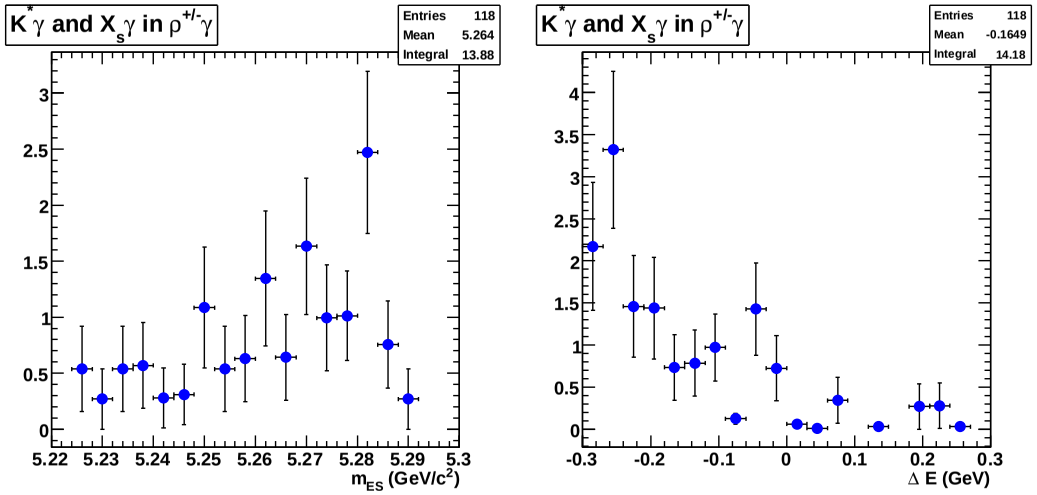


Figure 6.54: Plots for peaking B-background studies continued

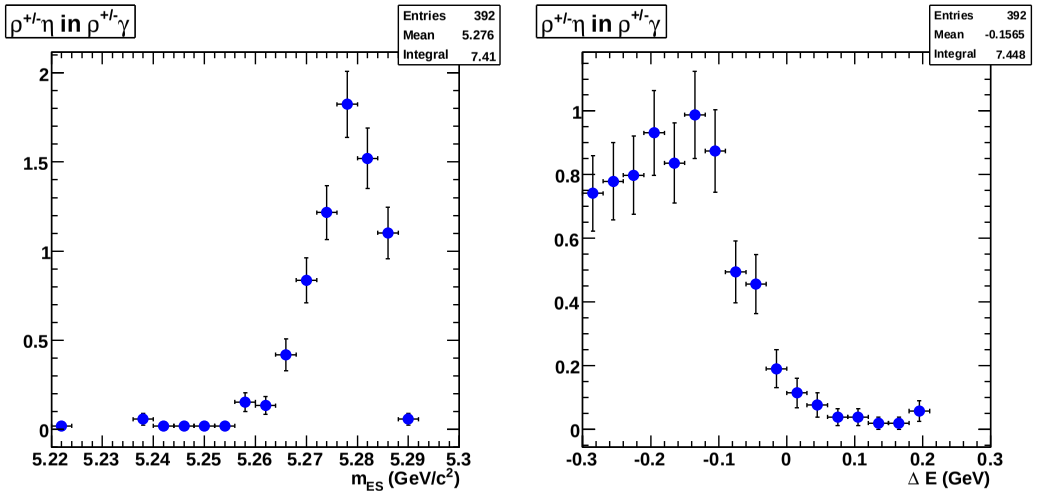


Figure 6.55: Plots for peaking B-background studies (adding $B^{\pm} \to \rho^{\pm}\eta$ and $B^{\pm} \to \rho^{\pm}\pi^{0}$) in $B^{\pm} \to \rho^{\pm}\gamma$

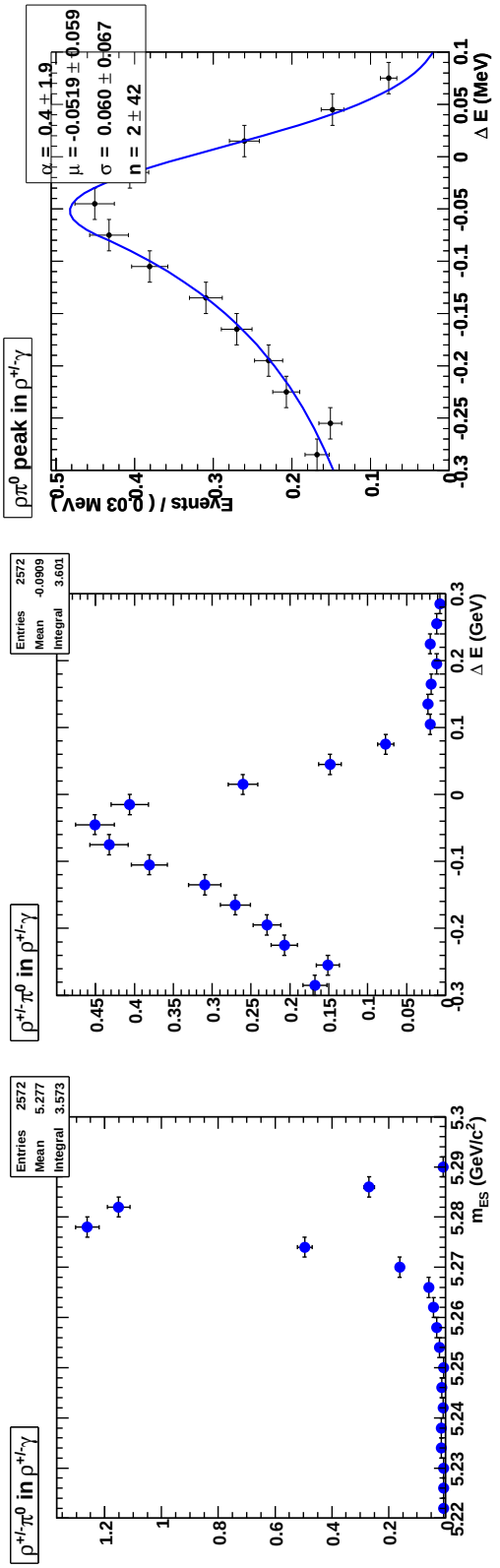


Figure 6.56: Plots for peaking B-background studies continued

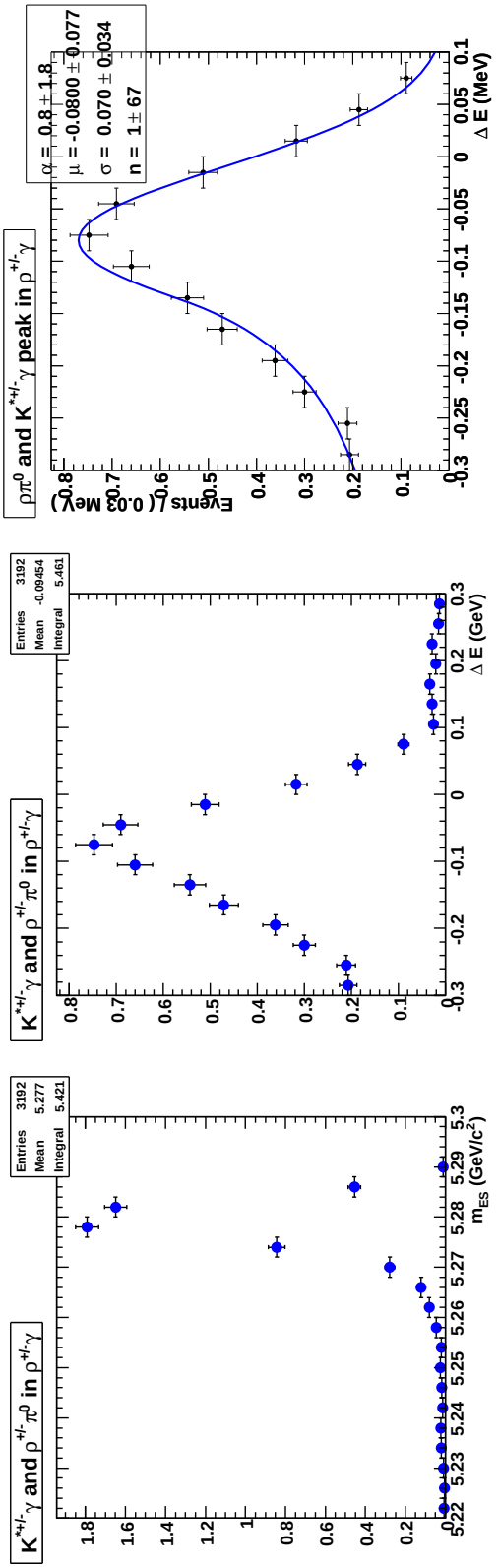


Figure 6.57: Plots for peaking B-background studies continued

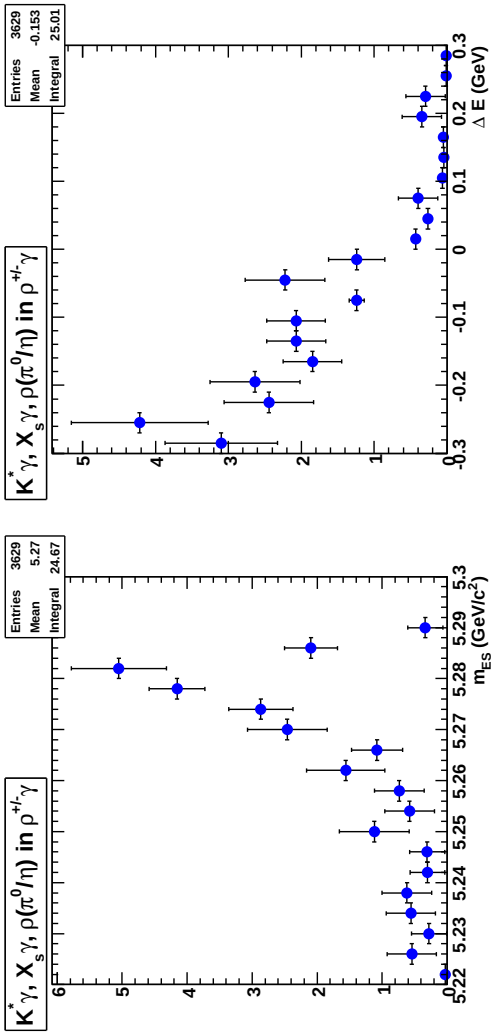


Figure 6.58: Plots for peaking B-background studies continued

6.7 Simultaneous $B \to (\rho/\omega)\gamma$ fitter

The individual decay modes can be combined together via the quark model prediction:

$$\Gamma(B^0 \to \rho^0\gamma) = \Gamma(B^0 \to \omega\gamma) = \frac{1}{2} \cdot \Gamma(B^\pm \to \rho^\pm\gamma). \tag{6.9}$$

Following [8], the combined branching fraction is defined as

$$\mathcal{B}(B \to (\rho,\omega)\gamma) = \frac{1}{2} \cdot \left(\mathcal{B}(B^\pm \to \rho^\pm\gamma) + \frac{\tau_{B^+}}{\tau_{B^0}} \cdot \left[\mathcal{B}(B^0 \to \rho^0\gamma) + \mathcal{B}(B^0 \to \omega\gamma)\right]\right) \tag{6.10}$$

with the B meson lifetimes being $\tau_{B^+} = 1.643 \pm 0.010$ and $\tau_{B^0} = 1.528 \pm 0.009$.

We define the combined "effective signal yield" as $N_{\rm eff}^{sim} = N_{B\overline{B}} \times \mathcal{B}(B \to (\rho,\omega)\gamma)$, so that it includes the efficiencies of all three individual modes. With the combined branching fraction defined above, and assuming the number of $B^0\overline{B}^0$ decays is the same as the number of B^+B^- decays, each being half the total $N_{B\overline{B}}$, this becomes:

$$N_{\rm eff}^{sim} = \frac{N_{sig}(B^\pm \to \rho^\pm\gamma)}{\epsilon\,(B^\pm \to \rho^\pm\gamma)} + \frac{\tau_{B^+}}{\tau_{B^0}} \cdot \left[\frac{N_{sig}(B^0 \to \rho^0\gamma)}{\epsilon\,(B^0 \to \rho^0\gamma)} + \frac{N_{sig}(B^0 \to \omega\gamma)}{\epsilon\,(B^0 \to \omega\gamma)}\right]. \tag{6.11}$$

We determine this combined "effective signal yield" directly from the simultaneous maximum likelihood fit. Each component of this fit is identical to the one of the corresponding individual fit. The only difference is that the signal yield for each mode is given as a function of the combined "effective signal yield" using the assumption of Equation 6.9 above:

$$\begin{aligned} N_{sig}\,(B^\pm \to \rho^\pm\gamma) &= \tfrac{1}{2} \cdot N_{\rm eff}^{sim} \cdot \epsilon\,(B^\pm \to \rho^\pm\gamma) \\ N_{sig}\,(B^0 \to \rho^0\gamma) &= \tfrac{1}{4} \cdot \tfrac{\tau_{B^0}}{\tau_{B^+}} \cdot N_{\rm eff}^{sim} \cdot \epsilon\,(B^0 \to \rho^0\gamma) \\ N_{sig}\,(B^0 \to \omega\gamma) &= \tfrac{1}{4} \cdot \tfrac{\tau_{B^0}}{\tau_{B^+}} \cdot N_{\rm eff}^{sim} \cdot \epsilon\,(B^0 \to \omega\gamma)\,. \end{aligned} \tag{6.12}$$

Thus, we vary only one yield parameter in the combined fit, while the individual three yields are derived from it using the corresponding mode-by-mode signal efficiencies obtained from MC. These efficiency numbers also include systematic corrections and errors described in the next section.

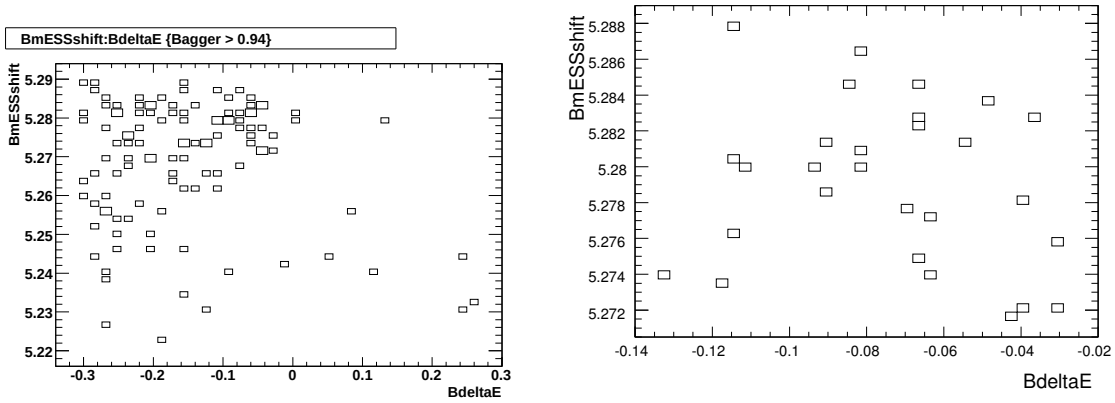


Figure 6.59: 2-dimensional distribution of $B\overline{B}$ MC in $B^{\pm} \to \rho^{\pm}\gamma$ mode

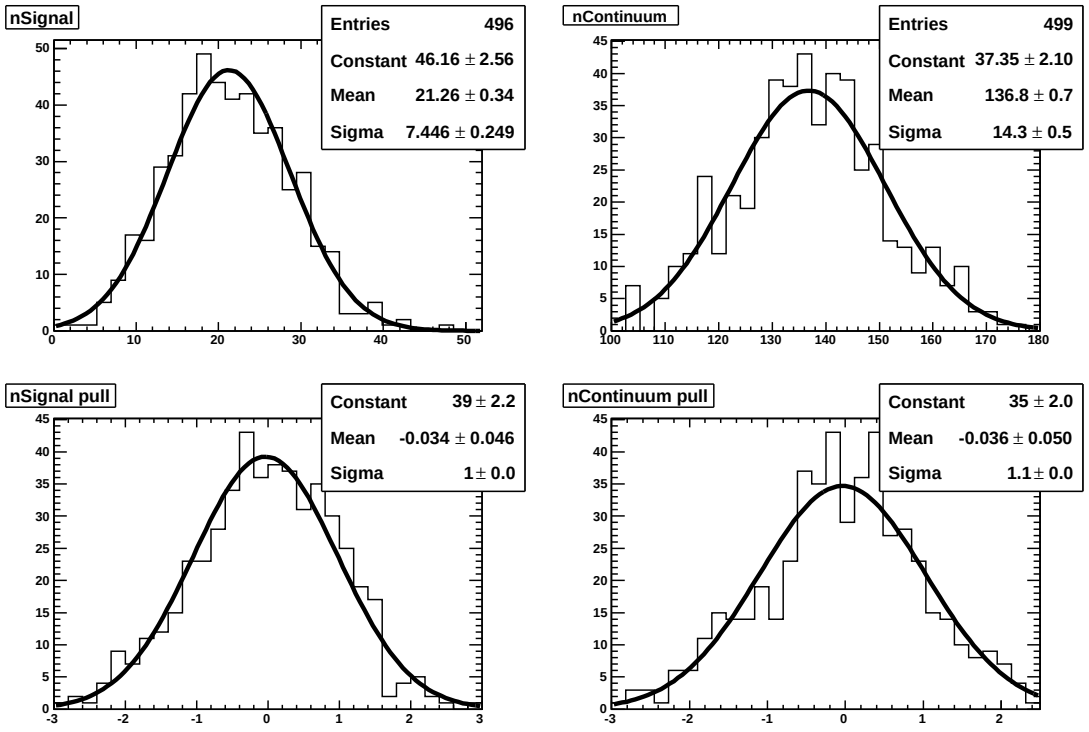


Figure 6.60: Plots of signal and continuum yields (top) and their pulls (bottom) in B-embedded toy MC for the $B^{\pm} \to \rho^{\pm}\gamma$ mode

Figure 6.61 shows the results of the luminosity-sampled MC fit combining all three signal modes. The total effective yield is 905^{+213}_{-198}, consistent with the expected number of 865. The plots also show each individual PDF contribution; for clarity, $B\overline{B}$ components are omitted from the $m_{\rm ES}$ plots, and continuum components are omitted from the ΔE projections. Figure 6.62 shows results of pure and signal-embedded toy MC studies for this fit. Neither show any significant biases (there's a slight positive pull in embedded toys).

Figures 6.63 and 6.64 give the corresponding plots for the combined $B \to (\rho^0/\rho^\pm)\gamma$ fit, which uses the exact same motivation as above, but ignores the $B^0 \to \omega\gamma$ data. The relative sizes of individual contributions are again kept at $2:1$ ratio, while the combined effective yield is equivalent to that obtained for the full fit above. Again, toy MC studies confirm that no significant biases are present. The one-shot luminosity-sampled MC fit gives the combined effective yield of 877^{+246}_{-225}, again consistent within errors with the expected yield of 865.

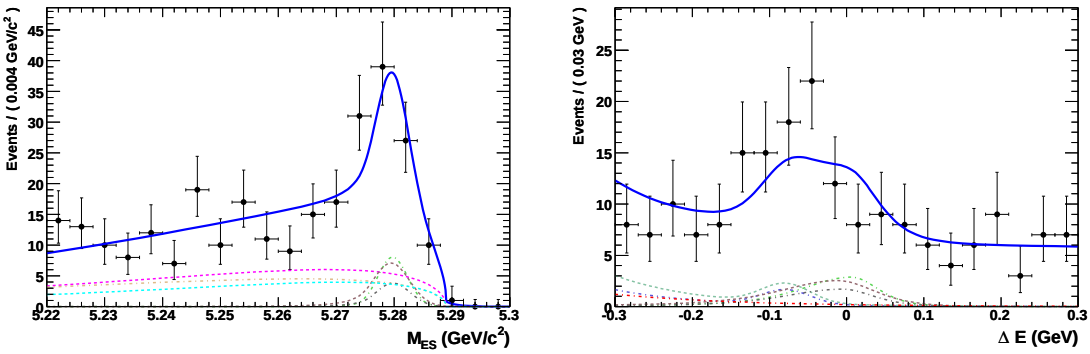


Figure 6.61: Luminosity-sampled Monte Carlo fit for combined $B \to (\rho/\omega)\gamma$

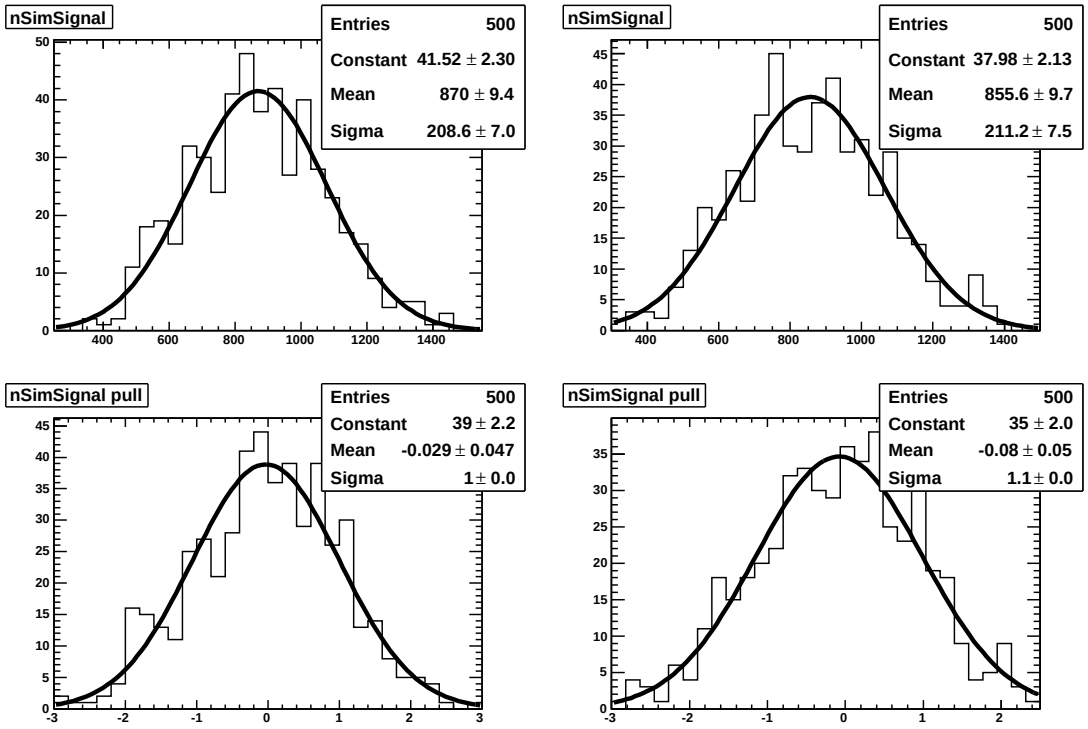


Figure 6.62: Pure (left) and signal-embedded (right) toy MC for combined $B \to (\rho/\omega)\gamma$ fit. Effective signal yields on top, pulls on the bottom

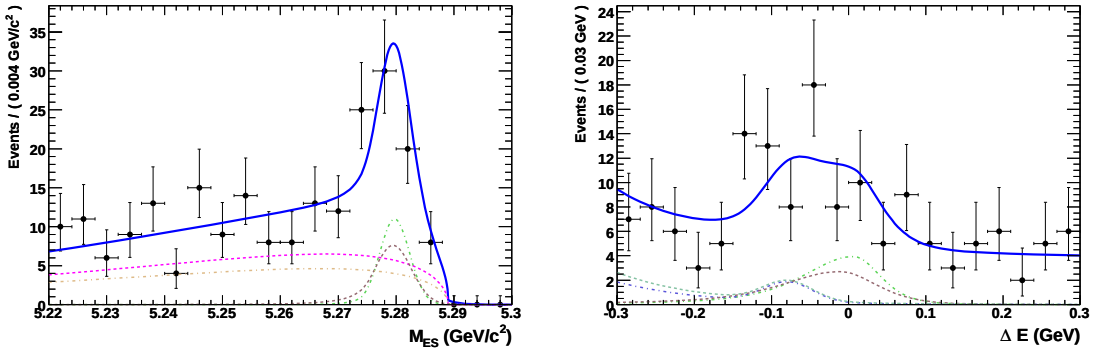


Figure 6.63: Luminosity-sampled Monte Carlo fit for combined $B \to (\rho^0/\rho^\pm)\gamma$

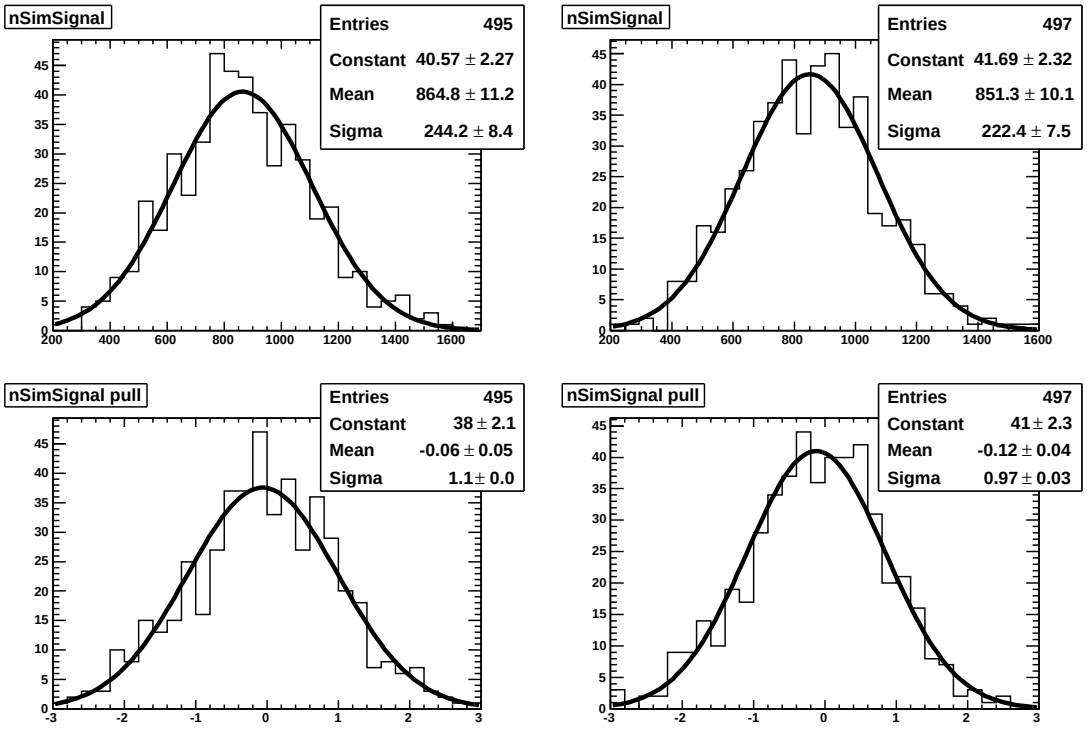


Figure 6.64: Pure (left) and signal-embedded (right) toy MC for combined $B \to (\rho^0/\rho^\pm)\gamma$ fit. Effective signal yields on top, pulls on the bottom

7 Validation studies

Several studies were performed to validate the analysis technique. These fall into two different categories: validating the analysis against previous iterations attempted at *BABAR*, and performing studies that validate the use of MC-simulated samples in event selection, classification, and modeling. We will not cover the former category, as it is internal to *BABAR* analysis framework and depends on such details as the software release version used for event reconstruction. It suffices to say that the current analysis was validated against earlier results in the sense that distributions of the variables used for event selection (such as the Bagger inputs) were not significantly different from those seen in the past.

Because we rely so heavily on MC simulated data samples, we need to find ways to cross-check that the simulation is indeed a good representation of our data. In order to achieve this goal, we combine the various simulated continuum MC datasets listed in Table 4.3 with appropriate luminosity weights that represent their expected proportional contribution in real data, and we compare the resulting distributions to those in the off-peak data sample. Only the variables for which the agreement is good are kept in the analysis. Because of the large number of variables used in this analysis, we do not include all the plots here, but a sample is shown in Figure 7.65.

Another step in validating our analysis technique is to make sure that the control samples we use are indeed appropriate representations of our signal modes. Some of these cross-checks have already been described in Section 6.5, where we demonstrated that the distributions of the fit variables, $m_{\rm ES}$ and ΔE, as well as the outputs of the Bagger, agree very well between the $B^0 \to \rho^0\gamma$ and $B^\pm \to \rho^\pm\gamma$ and the corresponding $B \to K^*\gamma$ modes. Another study that was carried out compared the individual

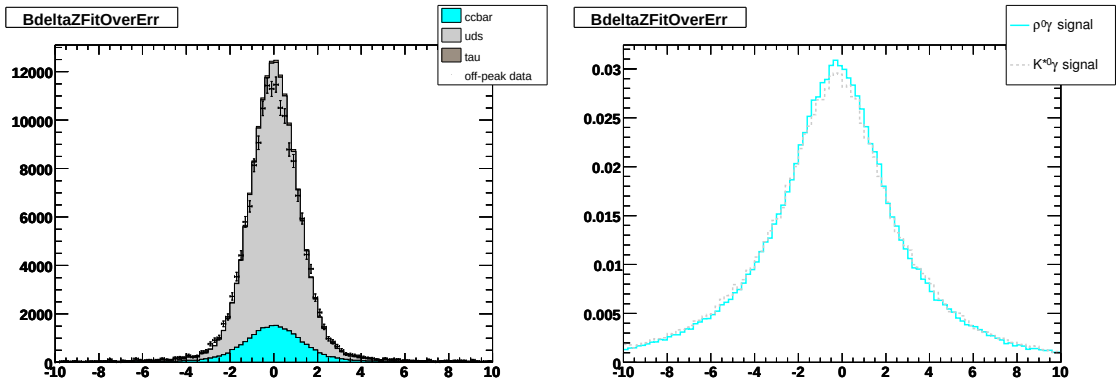


Figure 7.65: Distributions of combined continuum MC vs off-peak data in $B^0 \to \omega\gamma$ (left) and signal MC in $B^0 \to \rho^0\gamma$ vs $B^0 \to K^{*0}\gamma$ (right). The separation between the two B vertices is used as an example.

Bagger inputs between $B \to \rho\gamma$ and corresponding $B \to K^*\gamma$ modes using signal MC samples. Again, we do not include plots for all the variables and provide a sample in Figure 7.65. Agreement was very good, indicating that $B \to K^*\gamma$ is indeed, as expected, an appropriate control sample to be used for validation and systematic effect evaluation for Bagger and for the fits.

Finally, to get a sense for what differences in background levels we may expect in real data as compared to the combined MC fits, on which we base our expectations, we look at $m_{\rm ES}$, ΔE, and Bagger distributions in off-peak data and in continuum MC. Figure 7.66 shows the results, which are mostly limited by the low off-peak statistics. We see on average higher background yields in data, which is not unexpected. In the last round of analysis, we observed about 25% more continuum background in the on-peak data fits than we expected from MC. The Bagger distributions here look reasonable within available off-peak statistics, leaving no strong reason to suspect a large systematic difference in classifier performance between data and MC. In fact, the distributions in on-peak data seem to follow the shape of the combined MC rather well (although there is again the tendency to have more background events there).

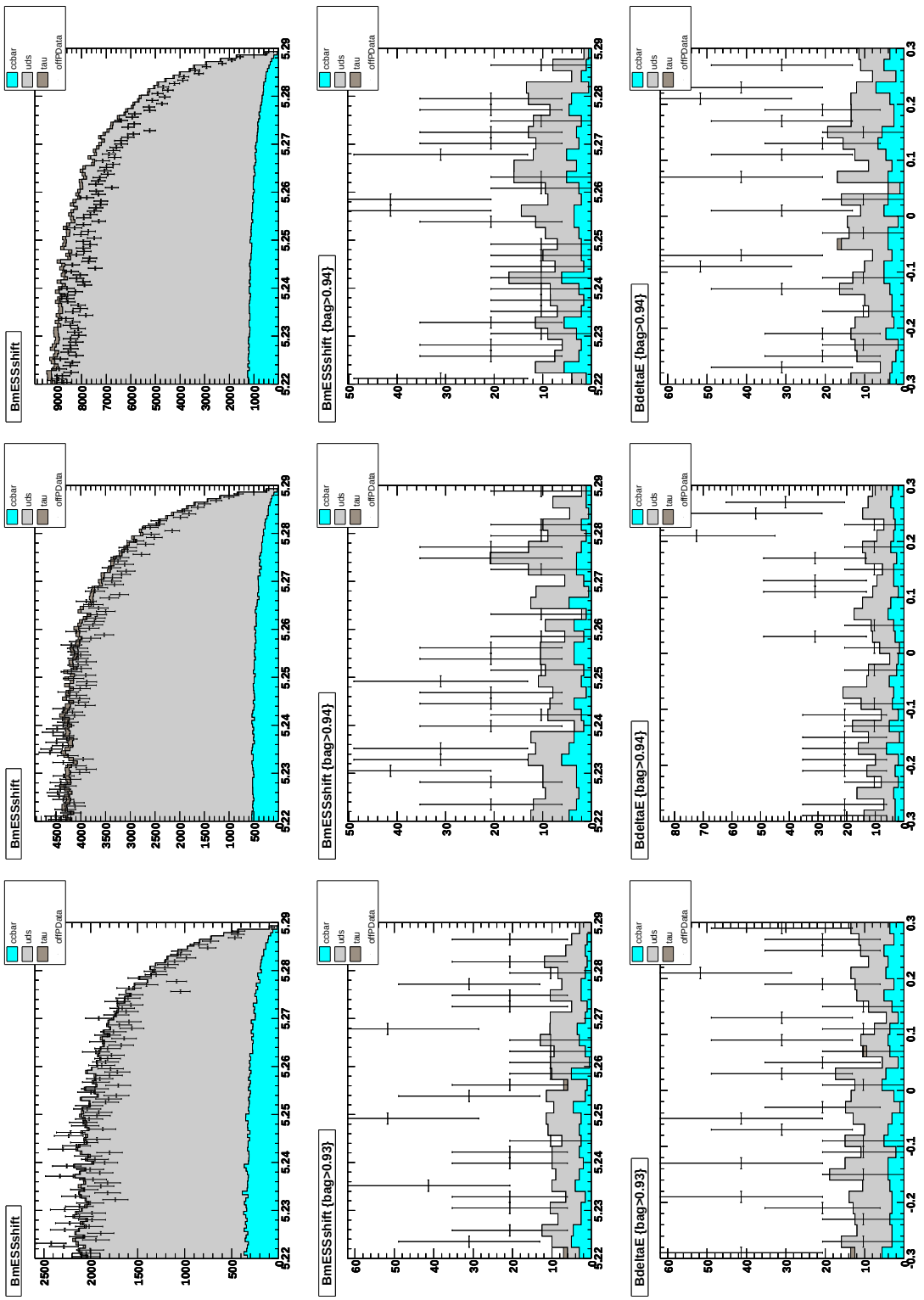


Figure 7.66: Plots for pre-unblinding data-MC studies: columns correspond to the individual modes ($B^0 \to \omega\gamma$, $B^0 \to \rho^0\gamma$, and $B^\pm \to \rho^\pm\gamma$); rows correspond to pre-Bagger cut $m_{\rm ES}$, post-Bagger cut $m_{\rm ES}$ and ΔE, and the high end of the Bagger output range (next page) in continuum MC vs off-peak and combined MC vs on-peak data.

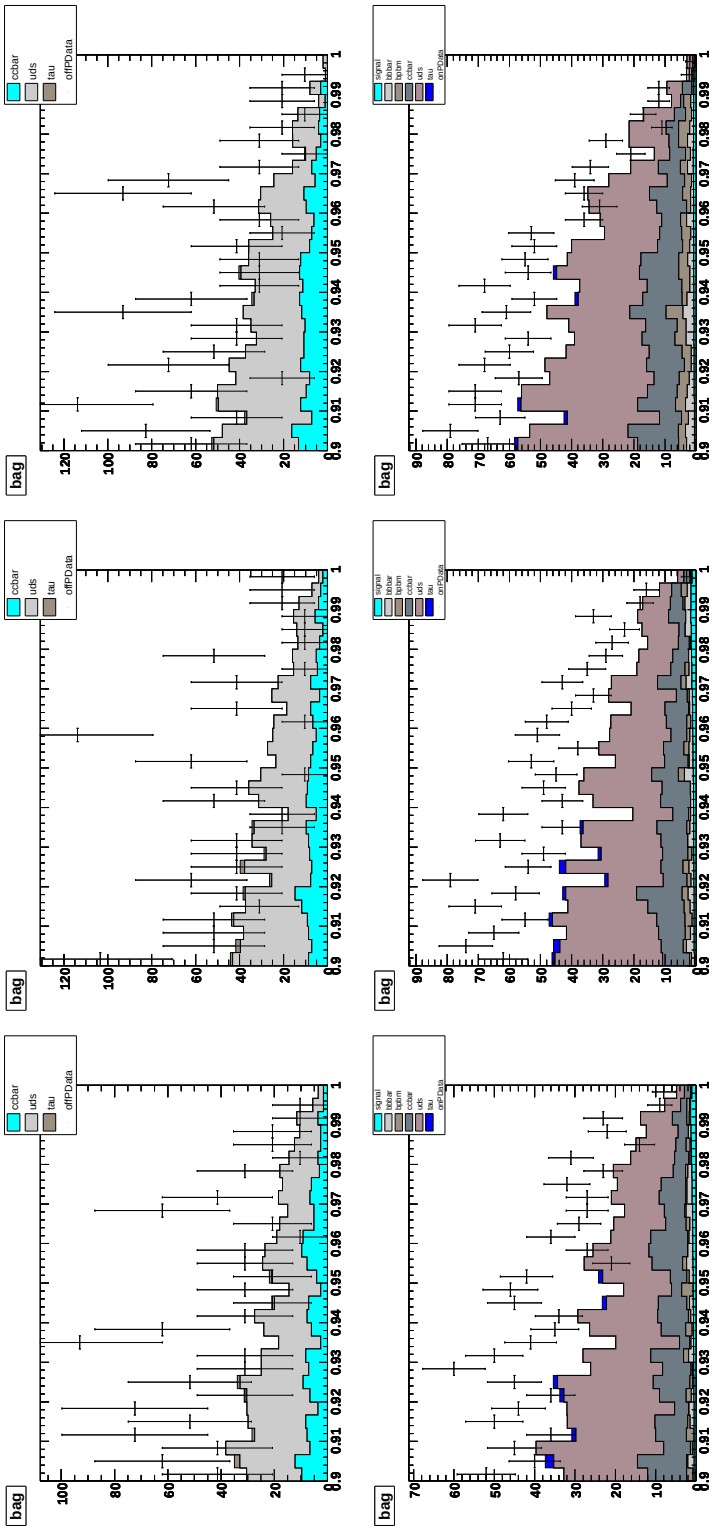


Figure 7.66: Pre-unblinding studies continued

8 Results

8.1 Branching fractions

Table 8.12 lists the results of unblinded on-peak data fits to all the signal modes, as well as the $\rho^0/\rho^\pm$ and ρ/ω combinations. Unfortunately, we do not observe a significant signal yield in the $B^0 \to \omega\gamma$ mode, and we do not improve on the previous measurement of the $B^\pm \to \rho^\pm\gamma$ channel either.

Table 8.13 summarizes on-peak data fits to the three signal modes using only *BABAR* runs 1–5 data, corresponding to 383 million $B\overline{B}$ pairs. Compared to the previously published *BABAR* measurement using runs 1–5 data[15], our statistical significance is only slightly better for the neutral and slighly worse for the charged modes. The results are in good agreement. It appears that the increase in total integrated luminosity provided by run 6 (82 million $B\overline{B}$ pairs) benefits only the $B^0 \to \rho^0\gamma$ channel, while contributing mostly background events to the other two. More details are provided in Section 8.1.1 below. Our results directly supersede those previously published by the *BABAR* collaboration.

	N_{signal}	ϵ	$N_{udsc\tau}$	**Significance**	$\mathcal{B}$ $(\times 10^6)$
$B^0 \to \omega\gamma$	$12.4^{+6.6}_{-5.7} \pm 1.6$	6.1%	$164.9^{+14.5}_{-13.8}$	2.2σ	$0.50^{+0.27}_{-0.23} \pm 0.09$
$B^0 \to \rho^0\gamma$	$34.9^{+8.6}_{-7.9} \pm 1.2$	8.5%	$271.4^{+18.0}_{-17.3}$	5.4σ	$0.97^{+0.24}_{-0.22} \pm 0.06$
$B^\pm \to \rho^\pm\gamma$	$23.3^{+8.1}_{-7.3} \pm 4.2$	4.9%	$178.1^{+15.5}_{-14.8}$	3.2σ	$1.20^{+0.42}_{-0.37} \pm 0.20$
$B \to (\rho/\omega)\gamma$	$1509.8^{+276}_{-262} \pm 104.4$	—	—	6.5σ	$1.63^{+0.30}_{-0.28} \pm 0.15$
$B \to \rho\gamma$	$1604.6^{+317}_{-299} \pm 131.5$	—	—	6.0σ	$1.73^{+0.34}_{-0.32} \pm 0.17$

Table 8.12: Summary of on-peak data fits for all signal modes

	N_{signal}	**Significance**
$B^0 \to \omega\gamma$	$11.4^{+6.0}_{-5.1}$	2.5σ
$B^0 \to \rho^0\gamma$	$25.9^{+7.7}_{-7.0}$	4.7σ
$B^\pm \to \rho^\pm\gamma$	$23.6^{+8.2}_{-7.4}$	3.9σ

Table 8.13: Summary of on-peak data fit results for all signal modes, using only *BABAR* runs 1–5 data

Figures 8.67 through 8.69 show the $m_{\rm ES}$ and the ΔE projection plots for the on-peak data fits for all three signal modes. Here, signal region cuts are applied to the variable that is not being plotted to visually enhance the signal-to-noise ratio. Figures 8.70 and 8.71 do the same for the simultaneous, isospin-constrained fits to the $\rho^\pm$ and ρ/ω combinations.

8.1.1 Study of $B^0 \to \omega\gamma$ yields by *BABAR* runs

Failing to observe a significant signal yield in the $B^0 \to \omega\gamma$ mode, and noting that excluding run 6 from the dataset does not significantly change the yield, we decided to perform additional fits to attempt to discover how the yield varies depending on the data-taking period. Figures 8.72, 8.73, and 8.74 show the $m_{\rm ES}$ projections for on-peak data fits in the $B^0 \to \omega\gamma$ channel using *BABAR* runs 1–4 (corresponding to 236 million $B\overline{B}$ pairs), runs 1–5 (383 million $B\overline{B}$ pairs), and run 5 by itself (147 million $B\overline{B}$ pairs). We see that the first fit appears to contain mostly background events, the second fit looks very similar to our nominal result that uses the whole available *BABAR* dataset, and the latter has a very clean signal signature — virtually background free. Thus, it appears that the majority of our signal events come from run 5 alone, while the other 5 runs dilute these with a large amount of background. One can view this as an upward fluctuation in one third of our data, or as an unlucky downward fluctuation in the other two thirds. Based on the numbers in Table 8.13, it seems plausible that the $B^\pm \to \rho^\pm\gamma$ mode suffers from a similarly uneven distribution of events in our data sample.

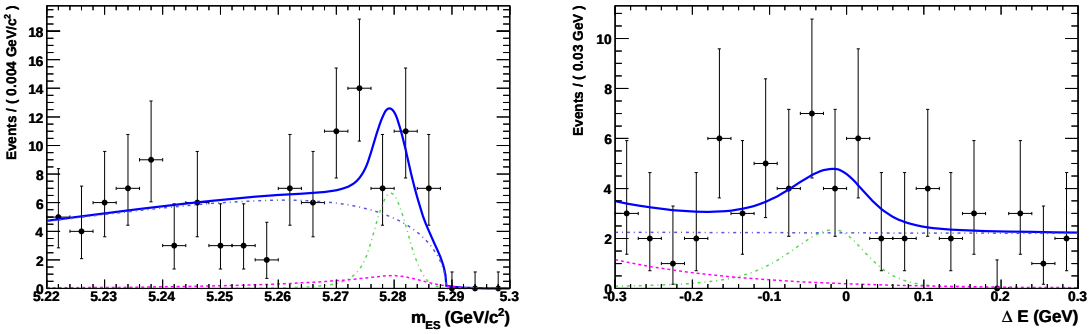


Figure 8.67: $m_{\rm ES}$ and ΔE projections for the on-peak data fit in the $B^0 \to \omega\gamma$ mode

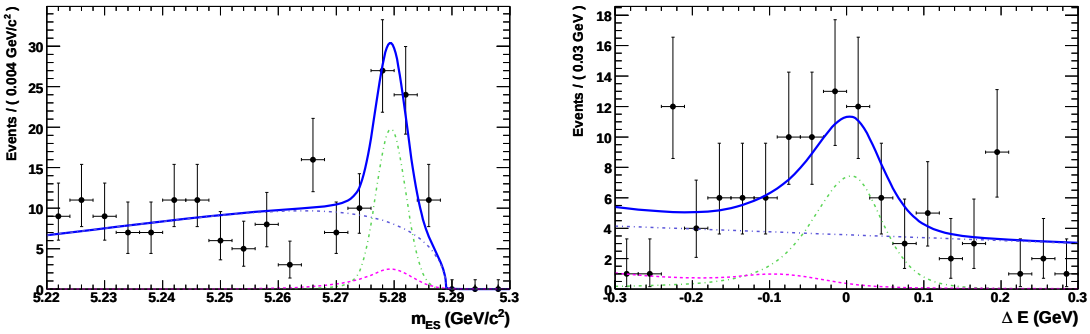


Figure 8.68: $m_{\rm ES}$ and ΔE projections for the on-peak data fit in the $B^0 \to \rho^0\gamma$ mode

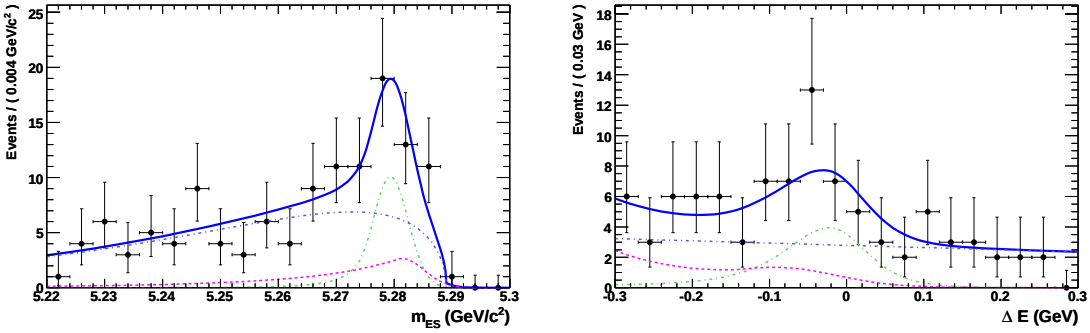


Figure 8.69: $m_{\rm ES}$ and ΔE projections for the on-peak data fit in the $B^\pm \to \rho^\pm\gamma$ mode

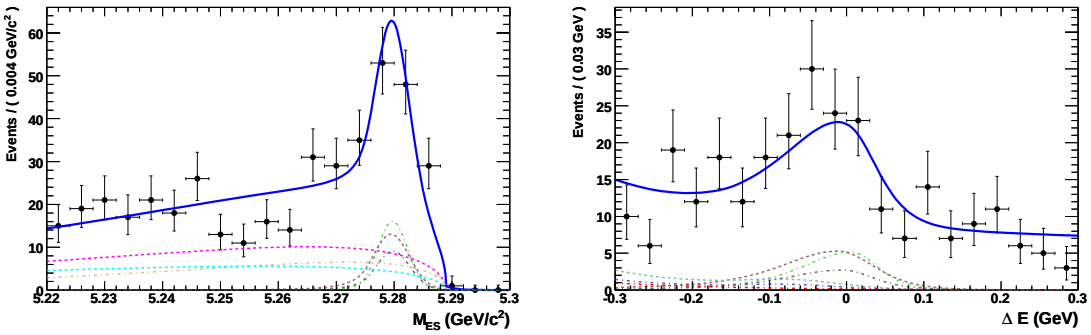


Figure 8.70: $m_{\rm ES}$ and ΔE projections for the on-peak data fit for the combined $B \to (\rho/\omega)\gamma$ mode

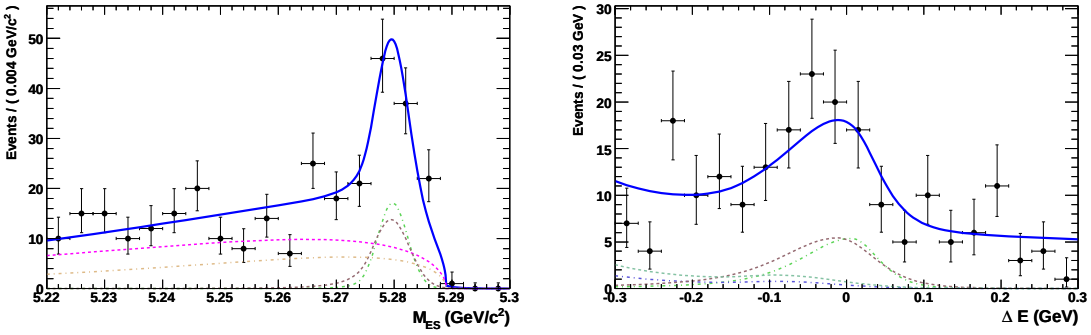


Figure 8.71: $m_{\rm ES}$ and ΔE projections for the on-peak data fit for the combined $B \to \rho\gamma$ mode

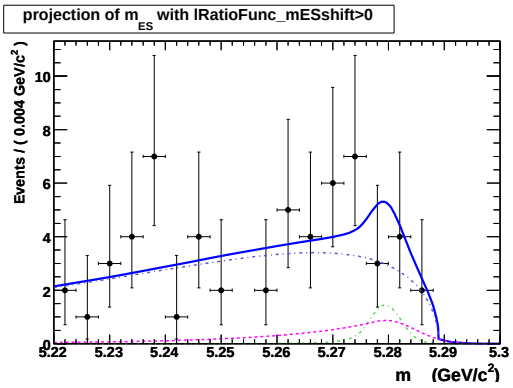


Figure 8.72: On-peak data fit in the $B^0 \to \omega\gamma$ mode using *BABAR* runs 1–4 only

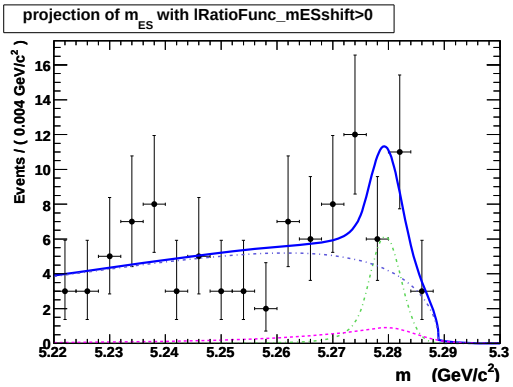


Figure 8.73: On-peak data fit in the $B^0 \to \omega\gamma$ mode using *BABAR* runs 1–5 only

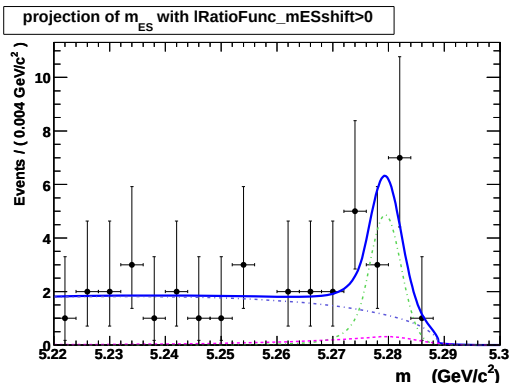


Figure 8.74: On-peak data fit in the $B^0 \to \omega\gamma$ mode using *BABAR* run 5 only

8.2 Upper limit for $B^0 \to \omega\gamma$ mode

Because we do not observe the $B^0 \to \omega\gamma$ signal above the 3σ level of statistical significance, we also set an upper limit on this branching fraction using a simple Bayesian technique. Bayes' theorem applied to the extraction of signal yield via fitting a parametric model to a set of data points implies:

$$\mathcal{L}(\text{parameters}|\text{data}) = P(\text{data}|\text{parameters}) = \frac{P(\text{parameters}|\text{data})P(\text{data})}{P(\text{parameters})}$$

where the left side of the equation is our likelihood function. Assuming flat priors $P(\text{data})$ and $P(\text{parameters})$, we readily obtain the probability distribution for the model parameters. Because we use an extended likelihood approach, where the event yield is incorporated into the fit model, setting an upper limit becomes as straightforward as integrating the likelihood curve with all other parameters held fixed.[5] Thus, we find $\mathcal{B}(B^0 \to \omega\gamma) < 0.9 \times 10^{-6}$ at the 90% confidence level.

8.3 Isospin and $\mathrm{SU}(3)_\mathrm{F}$

To test the hypothesis of isospin symmetry between the charged and neutral ρ channels and to find the degree of $\mathrm{SU}(3)_\mathrm{F}$ symmetry breaking between ρ^0 and ω channels, we measure

$$\Delta_\rho = \frac{\mathcal{B}(B^+ \to \rho^+\gamma)\tau_{B^0}}{2\mathcal{B}(B^0 \to \rho^0\gamma)\tau_{B^+}} - 1 = -0.43^{+0.25}_{-0.22}\ (stat) \pm 0.10\ (syst), \tag{8.13}$$

$$\Delta_\omega = \frac{\Gamma(B^0 \to \omega\gamma)}{\Gamma(B^0 \to \rho^0\gamma)} - 1 = -0.49^{+0.30}_{-0.27}\ (stat) \pm 0.10\ (syst). \tag{8.14}$$

Though less consistent than previous results, this number is still in agreement with the theoretical expectation [8].

[5] Note that this procedure is not possible in the frequentist interpretation of statistics.

8.4 Determination of $|V_{td}/V_{ts}|$

Using the current world average experimentally measured value of $\mathcal{B}(B \to K^*\gamma)$ [26], we calculate

$$R_\rho = \frac{\overline{\mathcal{B}}[B \to (\rho/\omega)\gamma]}{\mathcal{B}(B \to K^*\gamma)} = 0.039 \pm 0.008 \tag{8.15}$$

$$R_{\rho/\omega} = \frac{\mathcal{B}(B \to \rho\gamma)}{\mathcal{B}(B \to K^*\gamma)} = 0.042 \pm 0.009. \tag{8.16}$$

This result is used to obtain the ratio of CKM elements $|V_{td}/V_{ts}|$ by means of the following equation

$$\frac{\mathcal{B}(B \to (\rho/\omega)\gamma)}{\mathcal{B}(B \to K^*\gamma)} = S_{\rho,\omega}\left|\frac{V_{td}}{V_{ts}}\right|^2 \left(\frac{1 - m_\rho^2/M_B^2}{1 - m_{K^*}^2/M_B^2}\right)^3 \zeta^2[1 + \Delta R]. \tag{8.17}$$

Following [6], we choose the values $\zeta = 0.85 \pm 0.07$, and $\Delta R = 0.1 \pm 0.1$. We find

$$|V_{td}/V_{ts}|_\rho = 0.235^{+0.026}_{-0.025} \pm 0.018, \tag{8.18}$$

$$|V_{td}/V_{ts}|_{\rho/\omega} = 0.229^{+0.024}_{-0.023} \pm 0.017, \tag{8.19}$$

where the first error is experimental and the second is theoretical.

We would like to stress that we consider the individual branching fractions and the asymmetries to be the primary results of this analysis. While an important constraint on the Standard Model, the calculation of $|V_{td}/V_{ts}|$ requires theoretical input, where assumptions differ from one theorist to another.

9 Systematic errors

Various systematic effects have to be taken into account, including the modeling of reconstruction efficiency of charged and neutral particles, photon precut efficiencies, particle ID modeling, and the data-MC differences in signal efficiency of the cut on the multivariate classifier (Bagger). There are also systematic effects due to the choice of the fitting model. The errors are summarized in Table 9.17 and are described in detail below.

9.1 π^0 efficiency from Neutral AWG

The efficiency for π^0 reconstruction is studied by the Neutral Reconstruction Analysis Working Group (AWG) in the *BABAR* collaboration. The standard approach is to use the double ratio of $\tau \to \rho\nu$ to $\tau \to \pi\nu$ decays in data and MC. Based on these studies, we apply an efficiency correction of **0.971** to the $B^{\pm} \to \rho^{\pm}\gamma$ mode and **0.968** to the $B^0 \to \omega\gamma$ mode. The associated π^0 systematic error is **3%**.

9.2 Photon quality cuts using $e^+e^- \to \mu\mu\gamma$ control sample

Systematic uncertainties for the photon precuts on photon energy, number of crystals, `GammaZ20`, `GammaA42`, `Gammas9s25`, and lateral and second moments of the EMC cluster are determined using the $e^+e^- \to \mu\mu\gamma$ samples published by Neutral Reconstruction AWG. Figure 9.75 shows comparison between data and MC for the relevant photon variables.

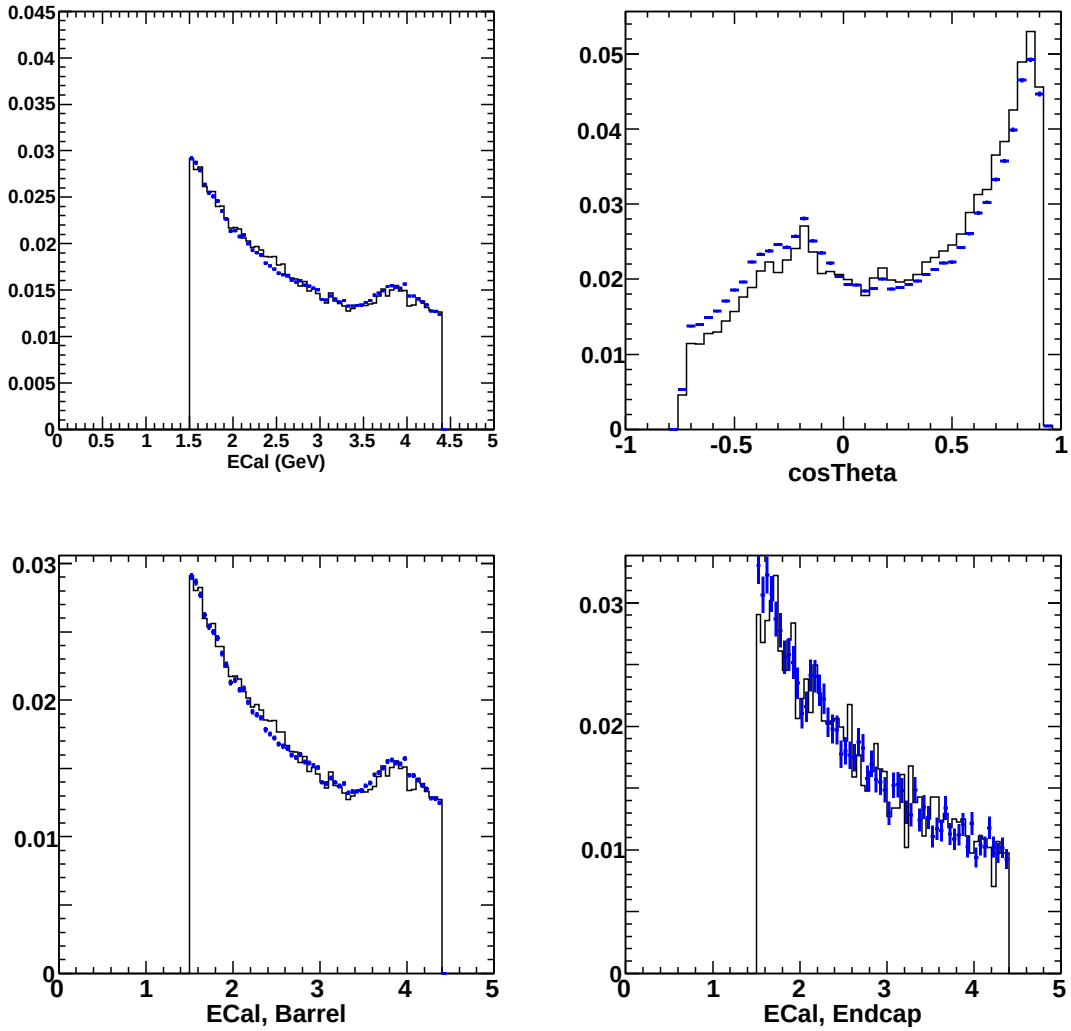


Figure 9.75: Distributions of photon variables for $\mu\mu\gamma$ data (points with error bars) and MC (histograms)

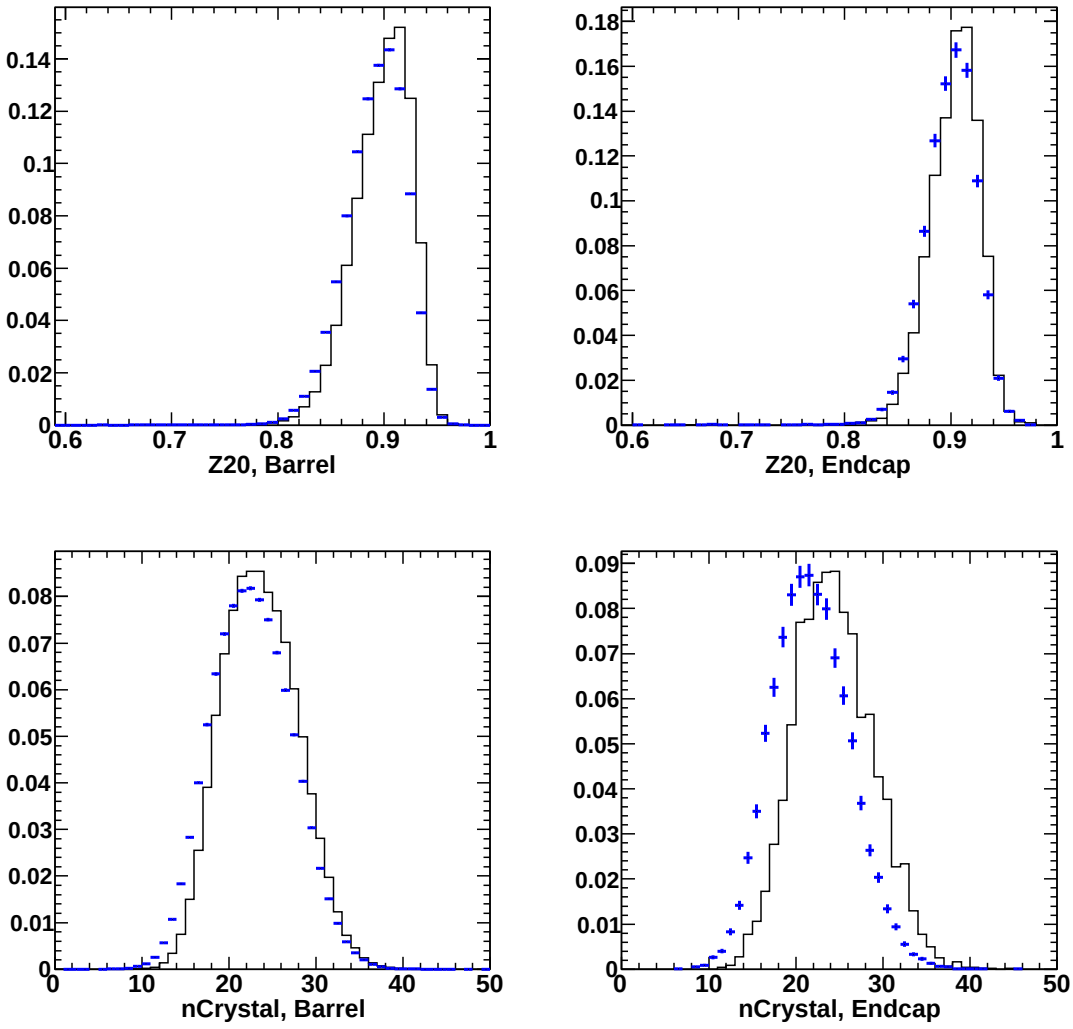


Figure 9.75: Distributions of photon variables for $\mu\mu\gamma$ continued

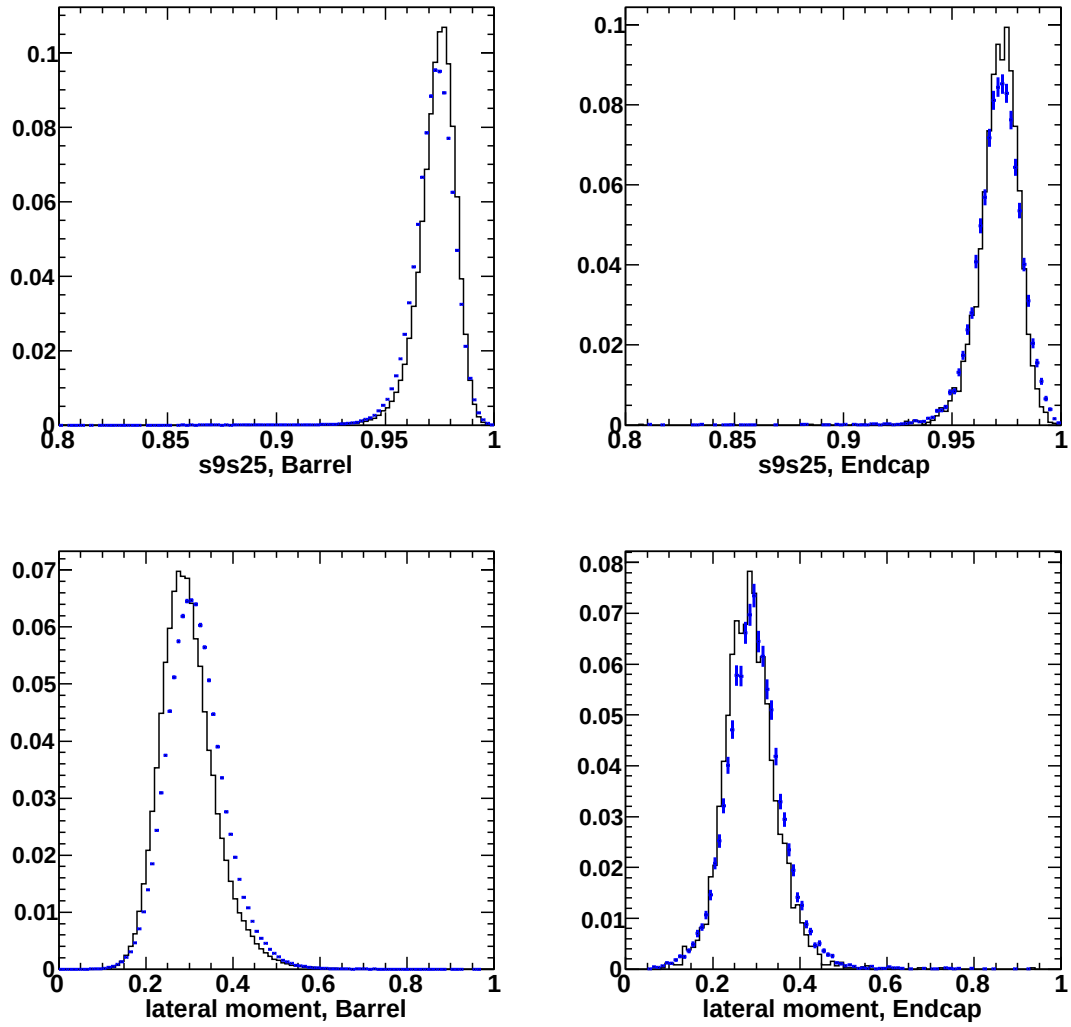


Figure 9.75: Distributions of photon variables for $\mu\mu\gamma$ continued

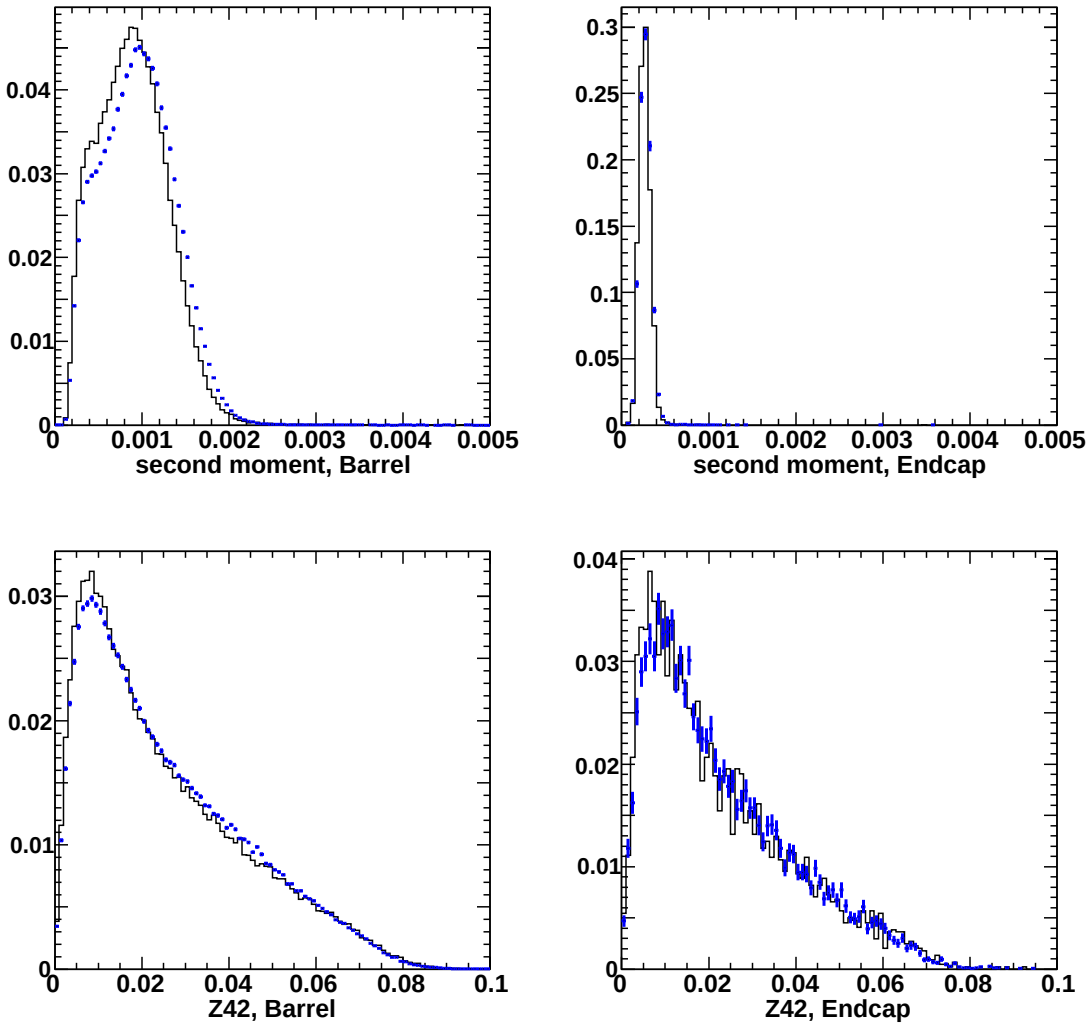


Figure 9.75: Distributions of photon variables for $\mu\mu\gamma$ continued

Figure 9.76 shows the dependency of the photon variables on photon energy (ECal) and acceptance angle (θ) for $B \rightarrow(\rho/\omega)\gamma$ signal MC. The $\mu\mu\gamma$ events are weighted so that the ECal and $cos(\theta)$ distributions match those of $B \rightarrow(\rho/\omega)\gamma$ signal MC. Figure 9.77 shows the weights in bins of ECal and $cos(\theta)$. The data-to-MC efficiency ratio for the photon precuts is calculated to be 0.996 ± 0.020. We don't apply signal efficiency corrections and quote a **2%** systematic error. EMC bump isolation cut of 25 cm was also studied in detail by the Neutral Reconstruction AWG, and a **2%** discrepancy between data and MC was found. These errors are added in quadrature to become the photon selection systematic error.

9.3 Particle ID systematic

Since we use a particle identification criteria that is different from those defined by the *BABAR* PID group for $B \rightarrow \rho\gamma$, we need to validate its performance ourselves. We do this using the standard PID sample of D^* decays, which, after background subtraction, provide clean kinematically selected samples of pions and kaons. There are two aspects of the PID that need to be checked: the $K^{\pm}$ mis-identification rate, and the $\pi^{\pm}$ efficiency. The first is the more critical because $K^{\pm}$ could cause a bias in the fit yield (an additive systematic), whereas $\pi^{\pm}$ efficiency is simply a multiplicative systematic. Figure 9.78 shows the performance of our PID criteria for the full generic $B\overline{B}$ and $c\overline{c}$ MC and D^* decay data (883033 events).

Kaon mis-identification rates are low ($\approx 1\%$), and there is generally a good agreement between data and MC. However, at high momentum (above 3 GeV/c), the agreement between D^* data and $B\overline{B}$ MC is poor. Since the $K^{\pm}$ momenta in $B\overline{B}$ decays are generally low (below 3 GeV/c), the high momentum $K^{\pm}$s are produced and boosted, almost exclusively, in the forward direction. These high momentum kaons enjoy better DIRC resolution, leading to lower mis-identification rates. The agreement between D^* data and $c\overline{c}$ MC at high momentum is much better; high momentum $K^{\pm}$s are present throughout the entire detector in both samples. Therefore, the $c\overline{c}$ MC PID rates were used to weight the $\pi^{\pm}$ ($B \rightarrow (\rho/\omega)\gamma$ sample) and $K^{\pm}$

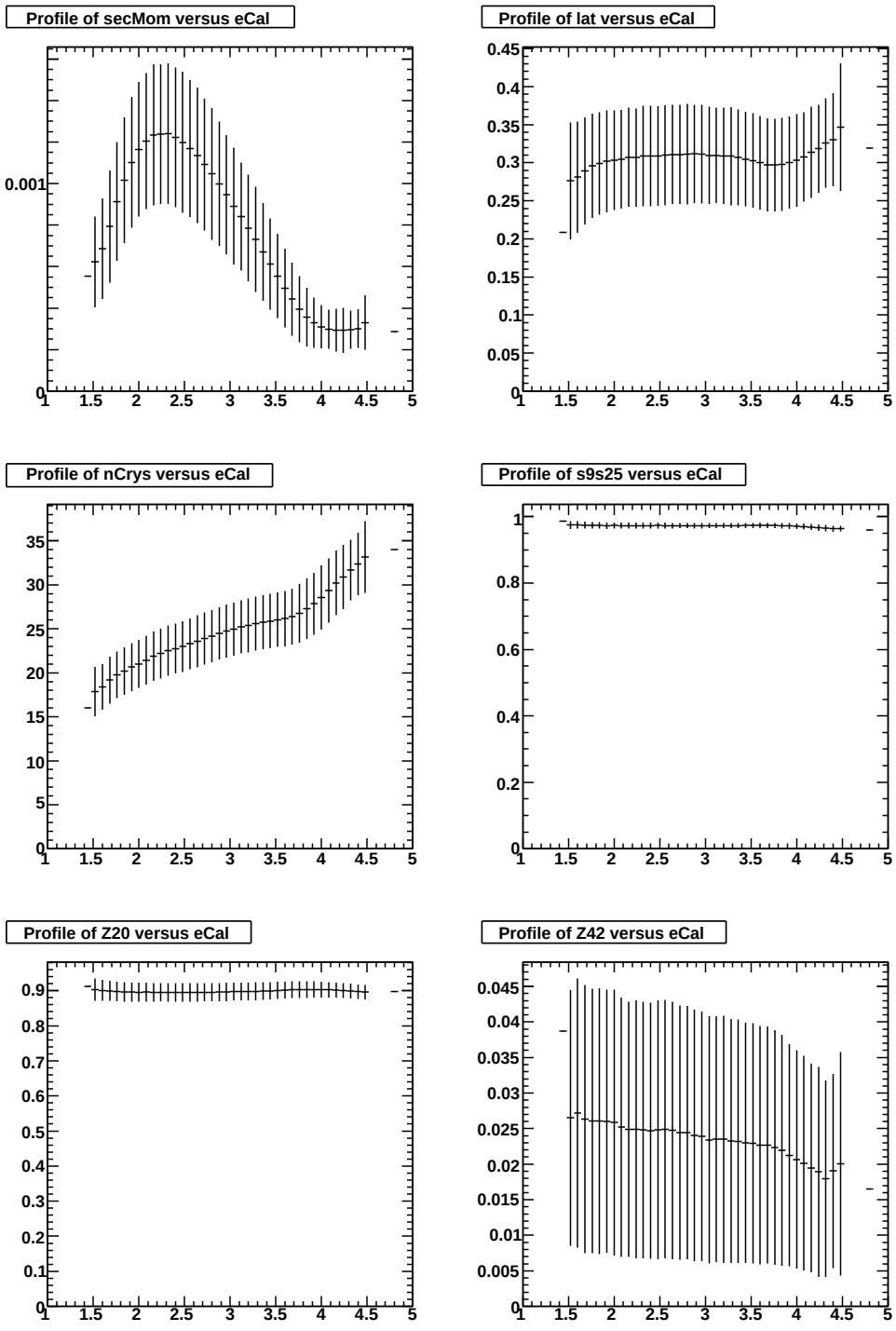


Figure 9.76: Profile histograms of the photon precut variables for $B \rightarrow (\rho/\omega)\gamma$ signal MC

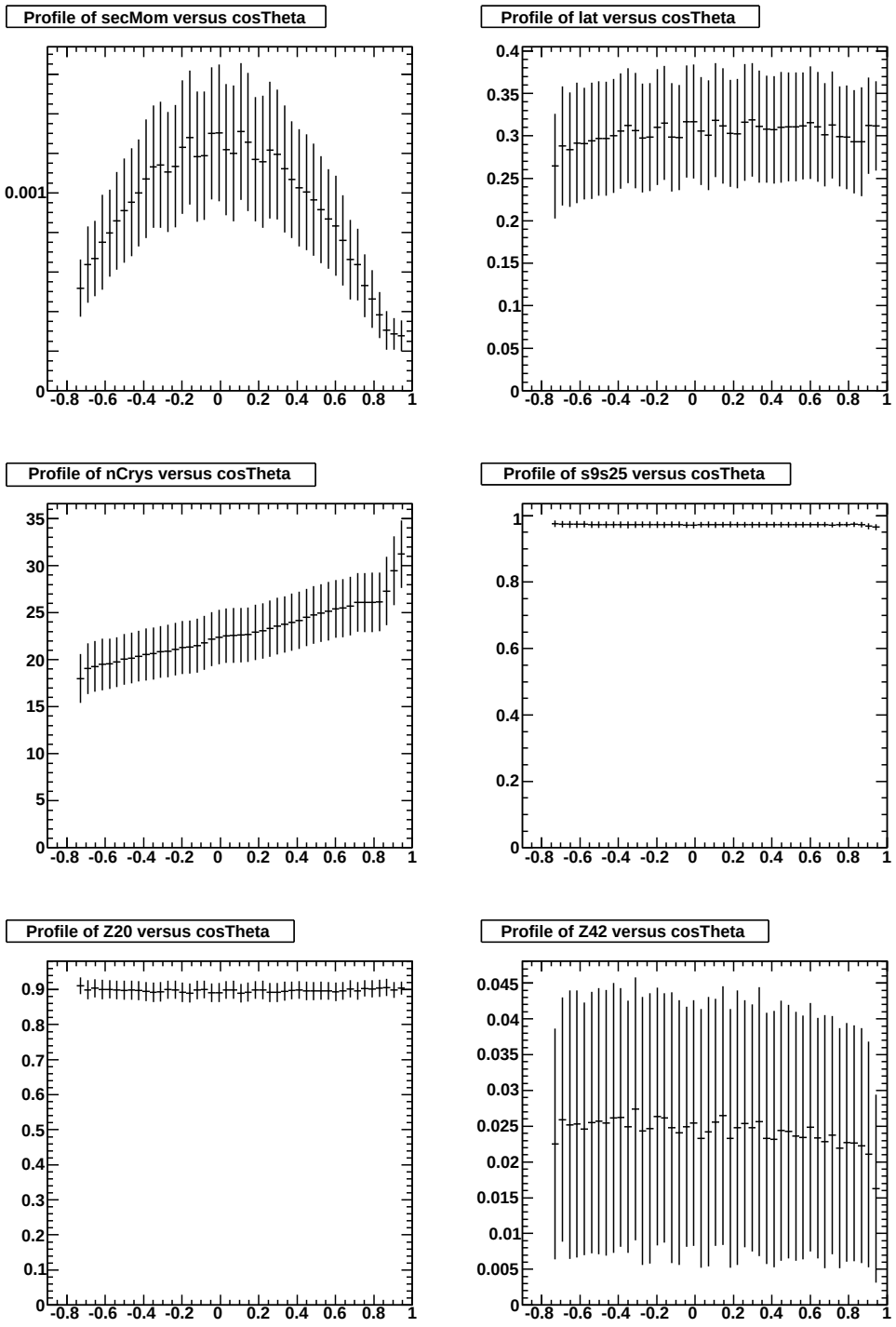


Figure 9.76: Profile histograms of the photon precut variables continued

($B^0 \to K^{*0}\gamma$ sample) momentum spectra.

Weighting the momentum spectrum of $\pi^{\pm}$s from $B \to (\rho/\omega)\gamma$ decays (kaons from $B^0 \to K^{*0}\gamma$ decays) with the D^* data $\pi^{\pm}$ efficiencies (D^* data $K^{\pm}$ mis-identification rates), integrating, then dividing by the same momenum spectrum weighted by $c\bar{c}$ MC $\pi^{\pm}$ efficiencies ($c\bar{c}$ MC $K^{\pm}$ mis-identification rates) gives the relative difference in the mis-identification rates predicted by the MC to those projected using the D^* decays data.

The results of the PID $K^{\pm}$ mis-identification and $\pi^{\pm}$ efficiency weighted momentum spectra study are summarized in Table 9.14. Uncertainties in $\pi^{\pm}$ efficiency affect the overall signal reconstruction efficiency, while $K^{\pm}$ mis-identification rates are directly reflected in the overall normalization of the $B \to K^*\gamma$ peaking background component included (and fixed) in the fits. Thus the former error is multiplicative, while the latter is additive. Based on this cross check, we found a relative uncertainty of **23%** in the overall $B \to K^*\gamma$ background normalization attributed to kaon mis-identification rates. We choose to vary the peaking $B\overline{B}$ contribution by a slightly more conservative **30%** in the fits. Agreement between data and MC is also quite good for $\pi^{\pm}$ efficiency. Based on this, we assign a **1%** per track multiplicative systematic in $B^0 \to \rho^0\gamma$ and $B^{\pm} \to \rho^{\pm}\gamma$ modes.

data set	$B\overline{B}$	$c\bar{c}$	**data**	$(B\overline{B} - data)/B\overline{B}$	$(c\bar{c} - data)/c\bar{c}$
$B^0 \to K^{*0}\gamma$	3147.88	4421.95	5434.32	0.726341	0.228941
$B^0 \to \rho^0\gamma$	328236	331301	330395	0.00657558	0.00273468
$B^{\pm} \to \rho^{\pm}\gamma$	159574	161431	161101	0.00957159	0.00204576
$B^0 \to \omega\gamma$	251909	262206	259542	0.030304	0.0101593

Table 9.14: Signal $\pi^{\pm}$ and $K^{\pm}$ momentum spectrum integrals after weighting with PID $\pi^{\pm}$ efficiency ($B \to (\rho/\omega)\gamma$ data samples) and $K^{\pm}$ mis-identification ($B^0 \to K^{*0}\gamma$ data sample) rates. The last two columns give the overall rates after averaging the PID weighted momentum bins using the $B\overline{B}$ and $c\bar{c}$ rates, respectively. The $c\bar{c}$ weighted overall rates were used in this analysis.

9.4 Bagger validation with $B \to K^*\gamma$

Together with the cross-check fits described in Section 6.5, the fits described here give us a handle on the systematic uncertainty associated with how well we understand the efficiency of the Bagger cut. We carry out the same 2-dimensional $m_{\rm ES}$-ΔE fits for the $K^+\pi^-$ and $K^\pm\pi^0$ modes as before, but remove the Bagger cut altogether. Thus we have the difference in the fraction of events passing this cut in Monte Carlo vs on-peak data, which we can then use to apply a correction and assign a systematic error to our $B \to (\rho/\omega)\gamma$ branching fraction measurements.

We begin by simply looking at signal MC numbers before and after the Bagger cut is applied, and we obtain the results shown in Table 9.15.

Mode	**Before cut**	**After cut**	ϵ
$K^+\pi^-$	423570	122813	29.0%
$K^\pm\pi^0$	38186	7159	18.7%

Table 9.15: Bagger cut signal MC efficiency in $B \to K^*\gamma$

To do the same exercise in on-peak data, we need to perform maximum likelihood fits to separate the signal from background. Results for these fits with the Bagger cut applied have already been presented in Table 6.11. The same fits performed with no Bagger applied (PDFs shown in Figures 9.79 and 9.82, MC and on-peak data fits in Figures 9.80 an 9.83) give us signal yields of 3986 ± 96 for $K^+\pi^-$ and 1321 ± 68 for $K^\pm\pi^0$ modes. However, after performing pure and signal-embedded toy MC studies (500 experiments each, shown in Figures 9.81 and 9.84), we find significant biases in the embedded toys that require correction. The mean of the signal yield distribution is shifted upward by 48 (from 3389, corresponding to 1.4%) in $K^+\pi^-$ and by 38 (from 916, corresponding to 4.1%) in $K^\pm\pi^0$. Applying these corrections, we end up with the numbers listed in Table 9.16.

Thus, in order to correct our data efficiency, we need to multiply it by **0.91** in $B^0 \to \rho^0\gamma$ and **0.88** in $B^\pm \to \rho^\pm\gamma$ mode. For $B^0 \to \omega\gamma$ channel, we use the average of the two, because, even though there's a π^0 in the final state similar to the $B^\pm \to \rho^\pm\gamma$,

Mode	**Before cut (corrected)**	**After cut**	ϵ
$K^+\pi^-$	3931 ± 96	1037 ± 33	26.4%
$K^\pm\pi^0$	1269 ± 68	210 ± 14	16.5%

Table 9.16: Bagger cut on-peak data efficiency in $B \to K^*\gamma$

the output of the Bagger is almost identical to $B^0 \to \rho^0\gamma$ and the efficiency of the cut is exactly the same for the same cut value.

There remains the question of statistical errors on the on-peak data yields. We believe that a correction is justified by the fact that Bagger inputs agree well in $B \to \rho\gamma$ and $B \to K^*\gamma$ (which is kinematically very similar to our signal). But the correction is obtained using central values for the data yields, while these also carry statistical errors themselves. Therefore, we apply the efficiency correction and take these errors as systematic errors. We obtain 3.2% error on the yield with the Bagger cut applied and 2.7% without it (taking this number from signal-embedded toy MC studies) for $K^+\pi^-$ mode. Similarly, the numbers are 6.7% and 6.4% for $K^\pm\pi^0$. Now we use the following facts:

$$f(x) = x \implies \frac{df}{f} = \frac{dx}{x} \quad \text{and} \quad f(x) = \frac{1}{x} \implies \frac{df}{f} = -\frac{dx}{x}.$$

Therefore, an error of 1% on either yield will translate to the same fractional error on the efficiency of the Bagger cut (since $\epsilon_{cut} = \frac{N \text{ after cut}}{N \text{ before cut}}$). And because the two sets of fits are independent, we can treat these errors as uncorrelated (i.e., an upward fluctuation in one yield does not need to correspond to a downward fluctuation in another). Then, adding the two sets of errors in quadrature, we end up with a **4.2%** error in $K^+\pi^-$ to be applied to $B^0 \to \rho^0\gamma$, and **9.3%** error in $K^+\pi^0$ to be applied to $B^\pm \to \rho^\pm\gamma$. Because we took the average of the two efficiency corrections for the ρ channels and applied that to the ω, we take half the quadrature sum of the corresponding errors, which comes out to be **5.1%**, as the error to be applied to the $B^0 \to \omega\gamma$ mode.

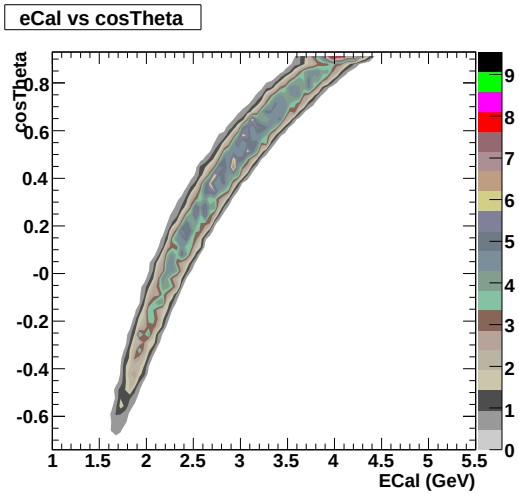


Figure 9.77: Distributions of weights applied to the $\mu\mu\gamma$ sample

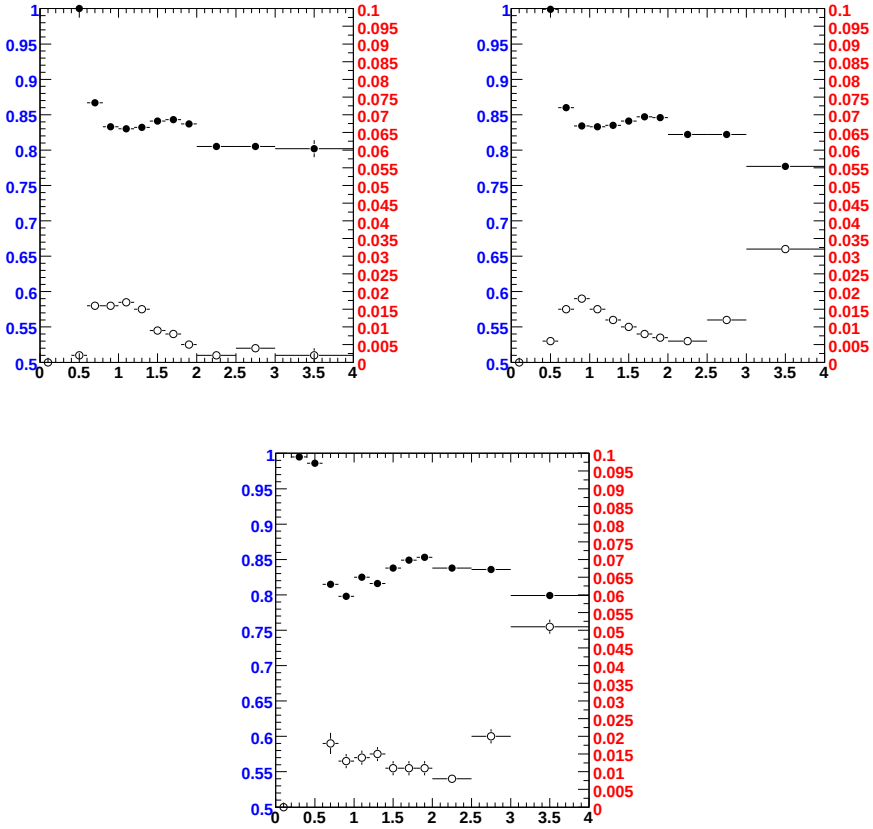


Figure 9.78: Particle ID performance in generic $B\overline{B}$ MC (left), generic $c\bar{c}$ MC (right), and Run 6 D^* data (bottom). Filled circles are for $\pi^{\pm}$ efficiency and use left-hand scale. Open circles are for $K^{\pm}$ mis-identification and use the right-hand scale.

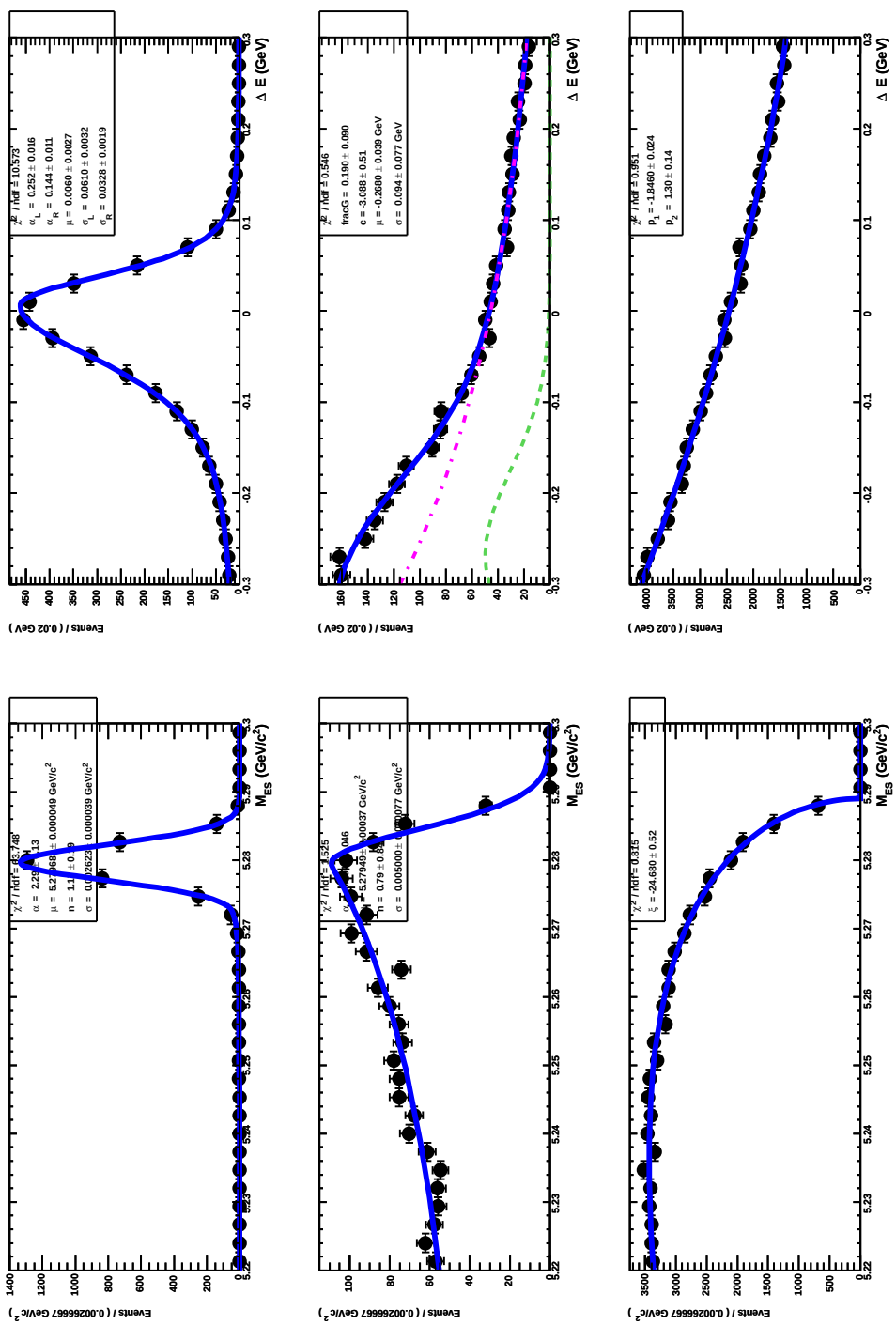


Figure 9.79: PDF shapes for $m_{\rm ES}$ (left) and ΔE (right) determined from Monte Carlo for the $B^0 \to K^{*0}\gamma$ ($K^{*0} \to K^+\pi^-$) mode. Signal on top, $B\overline{B}$ in the middle, continuum on the bottom. No Bagger cut is applied.

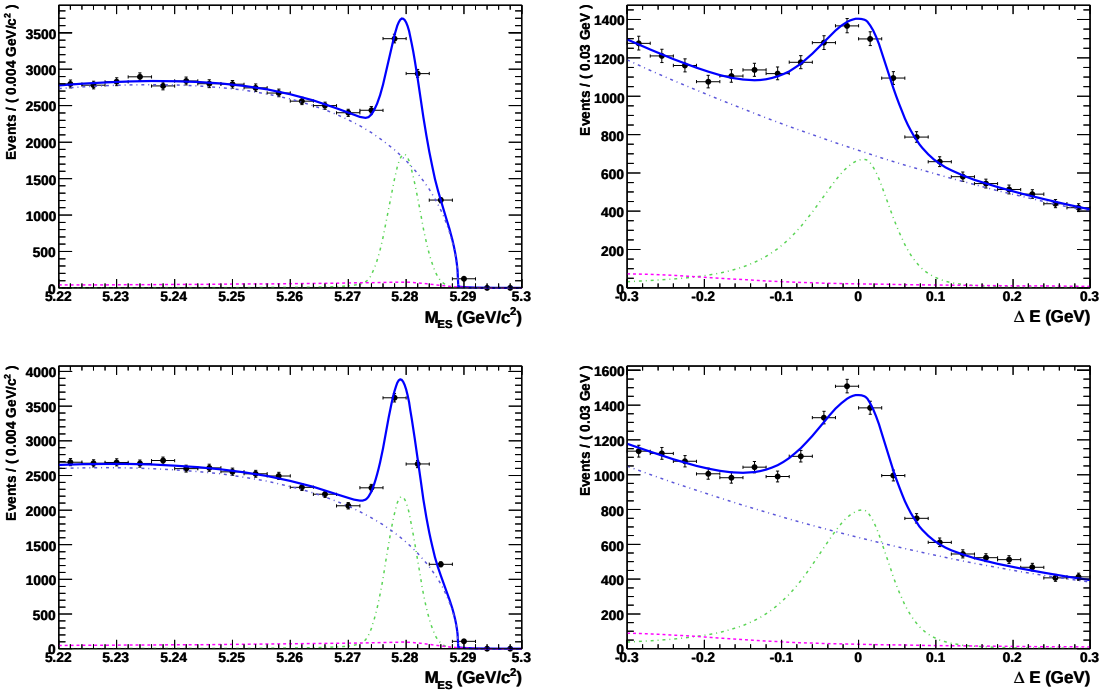


Figure 9.80: Luminosity-sampled, combined Monte Carlo (top) and on-peak data (bottom) fits for the $B^0 \to K^{*0}\gamma$ ($K^{*0} \to K^+\pi^-$) mode. No Bagger cut is applied.

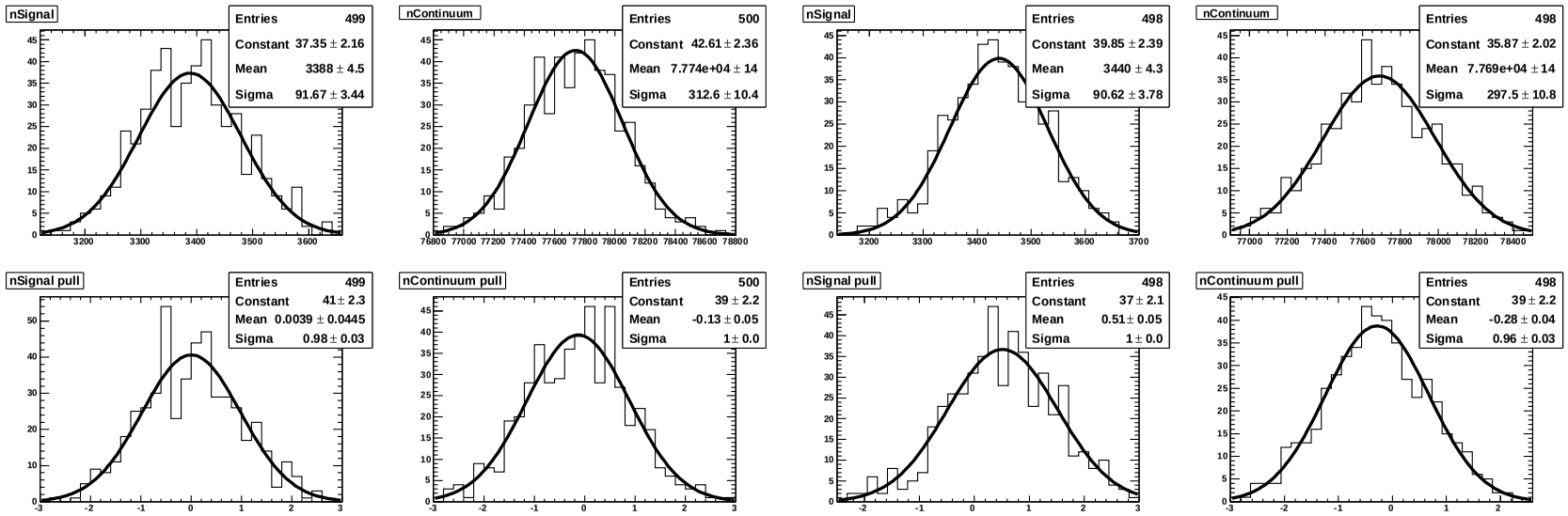


Figure 9.81: Plots of signal and continuum yields (top) and their pulls (bottom) in pure (left four) and signal-embedded (right four) toy MC for the $B^0 \to K^{*0}\gamma$ ($K^{*0} \to K^+\pi^-$) mode. No Bagger cut is applied. Expected signal and continuum yields are 3389 and 77762, respectively.

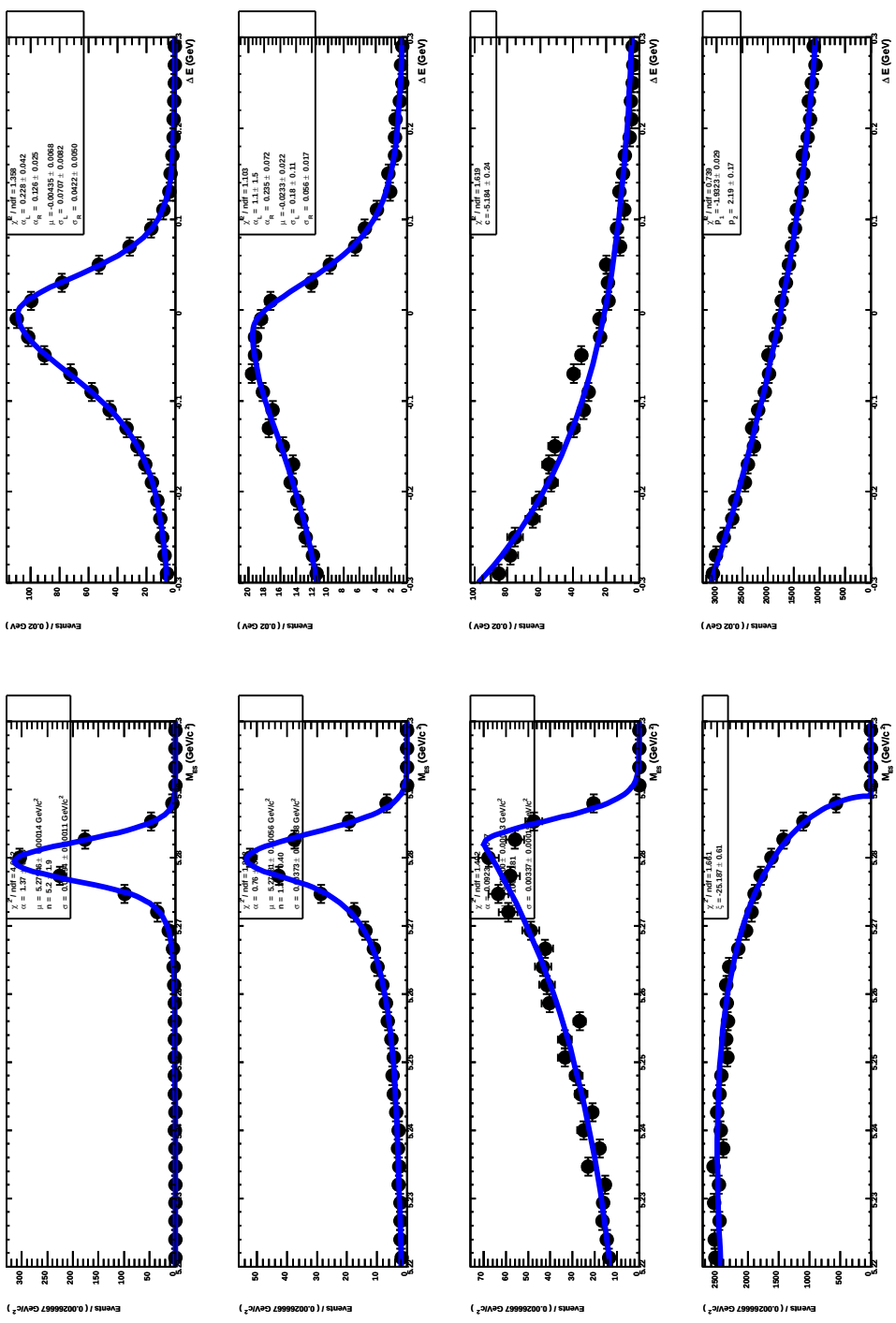


Figure 9.82: PDF shapes for $m_{\rm ES}$ (left) and ΔE (right) determined from Monte Carlo for the $B^{\pm} \to K^{*\pm}\gamma$ ($K^{*\pm} \to K^{\pm}\pi^0$) mode. True signal on top, self-crossfeed signal followed by $B\overline{B}$ in the middle, continuum on the bottom. No Bagger cut is applied.

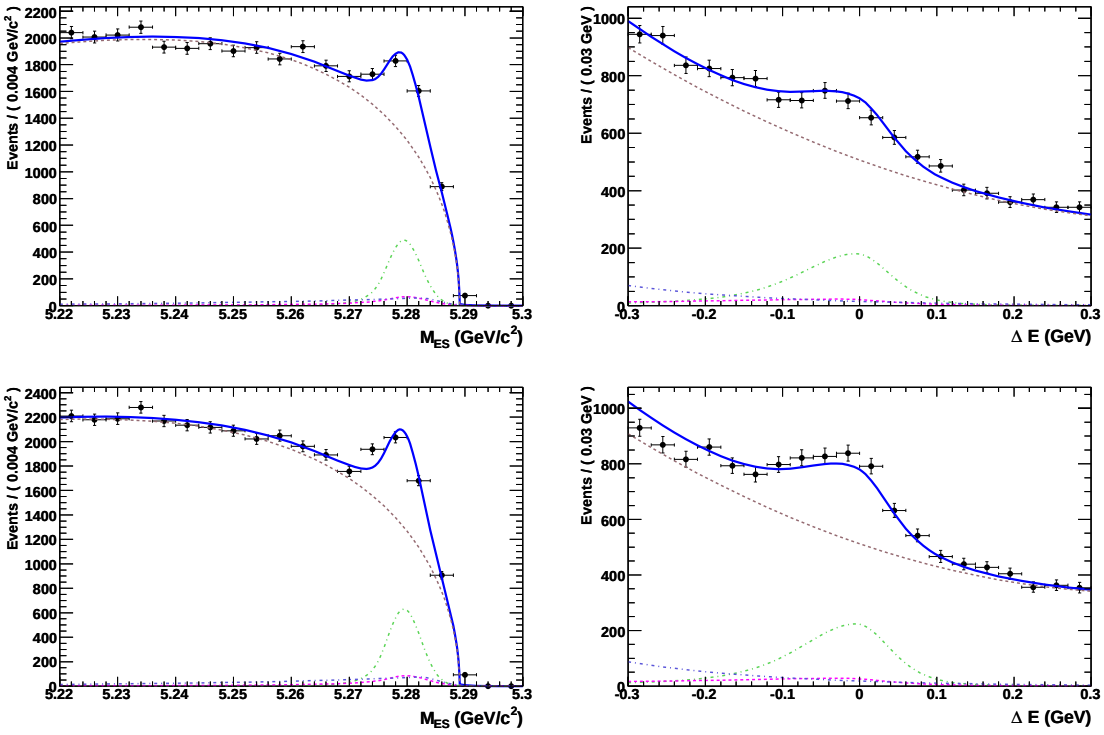


Figure 9.83: Luminosity-sampled, combined Monte Carlo (top) and on-peak data (bottom) fits for the $B^{\pm} \to K^{*\pm}\gamma$ ($K^{*\pm} \to K^{\pm}\pi^0$) mode. No Bagger cut is applied.

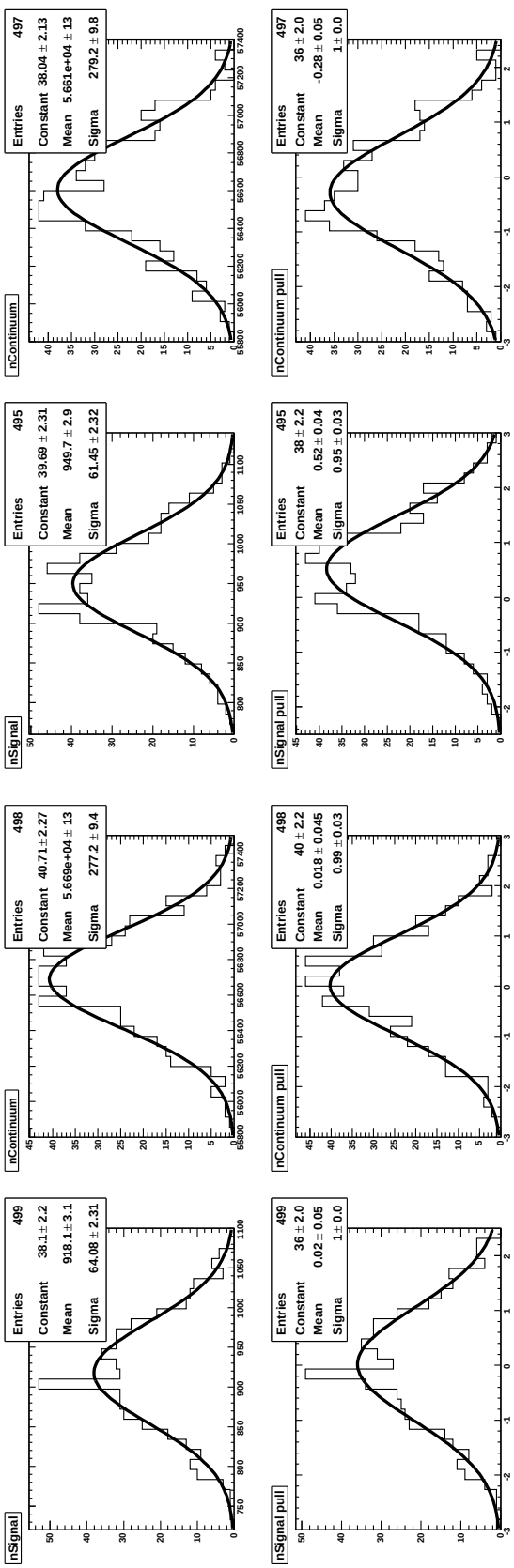


Figure 9.84: Plots of signal and continuum yields (top) and their pulls (bottom) in pure (left four) and signal-embedded (right four) toy MC for the $B^{\pm} \rightarrow K^{*\pm}\gamma$ ($K^{*\pm} \rightarrow K^{\pm}\pi^0$) mode. No Bagger cut is applied. Expected signal and continuum yields are 916 and 56682, respectively.

9.5 Fit model systematics

A maximum likelihood fit finds the best parameters for the chosen model, such that the model describes the data as closely as possible. The regression algorithm also returns the statistical errors for each parameter, based on the amount and distribution of the data sample. However, one can always ask the question of whether the choice of the model itself is the optimal one. Maybe a slightly different model could fit the data event better? This becomes especially important for those model parameters that are fixed in the fit and, therefore, have no associated statistical errors. In order to account for possible model variations in these parameters, we have to vary them explicitly by a specified amount and perform the fitting procedure again.

The typical thing to do is to vary each fixed parameter by the error one obtains from the individual component PDF fits, where all parameters are floating (before they are fixed for the final multidimensional maximum likelihood fit). However, we can do a bit better than that for the signal model parameters by taking those errors from the $B \to K^*\gamma$ control sample fits instead. In order to match the signal mode fits as closely as possible, we use the $B^\pm \to K^{*\pm}\gamma$ ($K^{*\pm} \to K^\pm\pi^0$) fit that does not split the signal into truth-matched and cross-feed components.

We also vary parameters of the $B\overline{B}$ background components in $m_{\rm ES}$ and ΔE. The fraction attributed to the peaking $B \to K^*\gamma$ backgrounds is varied by $\pm 30\%$ based on the $K^\pm$ mis-identification rate studied in Section 9.3 above. The overall contribution of the $B\overline{B}$ background to the likelihood function (the $B\overline{B}$ yield) is then varied by $\pm 15\%$ because the peaking component in ΔE accounts for about 30% of the entire PDF, and we choose to be a bit more conservative.

Table 9.17 summarizes all systematic errors and includes a breakdown of the fit model systematics into those originating from signal and those coming from the background components. The former is dominant for the $B^0 \to \omega\gamma$ channel, where we are limited by the low purity of the signal. In the cleanest $B^0 \to \rho^0\gamma$ channel, model errors from both sources are reasonably small. For the $B^\pm \to \rho^\pm\gamma$ mode, both the shape of the continuum background component, which rises as we go from

the low to the high end of the fit region, and the relatively larger $B\overline{B}$ contribution combine to produce a rather high background model error. The simultaneous fits are driven mostly by the high yield in the $B^0 \to \rho^0\gamma$ channel, and are therefore much less sensitive to the variations in backgrounds in the other two modes.

9.6 Other systematics

The other systematic effects we need to account for include charged particle reconstruction and the uncertainty on the overall number of $B\overline{B}$ pairs used in the analysis. The *BABAR* Tracking Efficiency Task Force provides standard recipes for all *BABAR* analyses dealing with charged tracks. From the tables they provide internally, we get a **0.43%** error for $B^0 \to \omega\gamma$ and $B^0 \to \rho^0\gamma$ modes, and **0.36%** for $B^\pm \to \rho^\pm\gamma$, since all our tracks are taken from the `GoodTracksLoose` list. As for the uncertainty on the size of the $B\overline{B}$ sample, all analyses in the *BABAR* collaboration use a centrally computed uncertainty on the recorded number of $\Upsilon(4s)$ decays. For the exact sample used in this analysis, we obtain a **1.1%** systematic error.

Source of error	$B^\pm \to \rho^\pm\gamma$	$B^0 \to \rho^0\gamma$	$B^0 \to \omega\gamma$	$B \to (\rho,\omega)\gamma$	$B \to \rho\gamma$
Tracking efficiency	0.4%	0.4%	0.4%	0.4%	0.4%
PID	1.0%	2.0%	1.0%	1.2%	1.4%
Photon selection	2.8%	2.8%	2.8%	2.8%	2.8%
π^0 reconstruction	3.0%	-	3.0%	2.0%	1.7%
Bagger cut efficiency	9.3%	4.2%	5.1%	7.5%	7.0%
Signal model	7.1	2.1	16.3	3.0	3.0
Background model	10.9	2.8	2.7	3.6	4.3
$B\overline{B}$ sample	1.1%	1.1%	1.1%	1.1%	1.1%
$\mathcal{B}(\omega \to \pi^+\pi^-\pi^0)$	-	-	0.8%	0.1%	-
Combined	16.7%	6.6%	17.9%	9.5%	9.5%

Table 9.17: Fractional systematic errors in % of the measured branching fractions

10 Summary

In conclusion, we have performed an analysis of $B \to (\rho/\omega)\gamma$ decays using the full dataset of 465 million $B\overline{B}$ pairs collected by the *BABAR* experiment, which ran from 1998 to 2008. We confirmed earlier observations of the $B^0 \to \rho^0\gamma$ channel, as well as evidence for $B^\pm \to \rho^\pm\gamma$. We do not observe a statistically significant signal in $B^0 \to \omega\gamma$ channel, and a study of signal yields in this mode performed for different periods of data-taking suggests that there's a great deal of fluctuation in the data. We measure the branching fractions to be $\mathcal{B}(B^\pm \to \rho^\pm\gamma) = (1.20^{+0.42}_{-0.37} \pm 0.20) \times 10^{-6}$, $\mathcal{B}(B^0 \to \rho^0\gamma) = (0.97^{+0.24}_{-0.22} \pm 0.06) \times 10^{-6}$, and $\mathcal{B}(B^0 \to \omega\gamma) = (0.50^{+0.27}_{-0.23} \pm 0.09) \times 10^{-6}$, and we set an upper limit for the $B^0 \to \omega\gamma$ channel at $\mathcal{B}(B^0 \to \omega\gamma) < 0.9 \times 10^{-6}$ at 90% confidence level using a simple Bayesian approach.

Decay rates for these channels probe isospin and $\mathrm{SU(3)_F}$ violation between the charged and neutral modes. Combined with branching fractions for $B \to K^*\gamma$ decays, they also provide a handle on the ratio of CKM matrix elements V_{td}/V_{ts}, giving us yet another constraint on the Standard Model of elementary particle physics:

$$\begin{aligned}
\Delta_\rho = \frac{\mathcal{B}(B^+ \to \rho^+\gamma)\tau_{B^0}}{2\mathcal{B}(B^0 \to \rho^0\gamma)\tau_{B^+}} - 1 &= -0.43^{+0.25}_{-0.22}\ (stat) \pm 0.10\ (syst), \\
\Delta_\omega = \frac{\Gamma(B^0 \to \omega\gamma)}{\Gamma(B^0 \to \rho^0\gamma)} - 1 &= -0.49^{+0.30}_{-0.27}\ (stat) \pm 0.10\ (syst), \\
|V_{td}/V_{ts}|_\rho &= 0.235^{+0.026}_{-0.025} \pm 0.018, \\
|V_{td}/V_{ts}|_{\rho/\omega} &= 0.229^{+0.024}_{-0.023} \pm 0.017.
\end{aligned}$$

Bibliography

[1] P. F. Harrison and H. R. Quinn [*BABAR* Collaboration], "The *BABAR* physics book: Physics at an asymmetric B factory." Papers from Workshop on Physics at an Asymmetric B Factory (*BABAR* Collaboration Meeting), Rome, Italy, 11-14 Nov 1996, Princeton, NJ, 17-20 Mar 1997, Orsay, France, 16-19 Jun 1997 and Pasadena, CA, 22-24 Sep 1997.

[2] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[3] See, for example, S. Bertolini, F. Borzumati, and A. Masiero, Nucl. Phys. B **294**, 321 (1987); H. Baer and M. Brhlik, Phys. Rev. D **55**, 3201 (1997) [arXiv:hep-ph/9610224]; J. L. Hewett and J. D. Wells, Phys. Rev. D **55**, 5549 (1997) [arXiv:hep-ph/9610323]; M. S. Carena, D. Garcia, U. Nierste, and C. E. M. Wagner, Phys. Lett. B **499**, 141 (2001) [arXiv:hep-ph/0010003].

[4] A. Ali and A. Parkhomenko, arXiv:hep-ph/0610149.

[5] S. W. Bosch and G. Buchalla, Nucl. Phys. B **621**, 459 (2002) [arXiv:hep-ph/0106081].

[6] P. Ball, G. W. Jones, and R. Zwicky, Phys. Rev. D **75**, 054004 (2007) [arXiv:hep-ph/0612081]; P. Ball and R. Zwicky, JHEP **0604**, 046 (2006) [arXiv:hep-ph/0603232].

[7] A. Ali and A. Y. Parkhomenko, Eur. Phys. J. C **23**, 89 (2002) [arXiv:hep-ph/0105302].

[8] A. Ali, E. Lunghi, and A. Y. Parkhomenko, Phys. Lett. B **595**, 323 (2004) [arXiv:hep-ph/0405075].

[9] S. W. Bosch and G. Buchalla, JHEP **0501**, 035 (2005) [arXiv:hep-ph/0408231].

[10] M. Neubert, arXiv:hep-ph/0012204.

[11] A. Abulencia et al. [CDF Collaboration], Phys. Rev. Lett. **97**, 242003 (2006) [arXiv:hep-ex/0609040].

[12] T. E. Coan et al. [CLEO Collaboration], Phys. Rev. Lett. **84**, 5283 (2000) [arXiv:hep-ex/9912057].

[13] B. Aubert et al. [*BABAR* Collaboration], Phys. Rev. Lett. **94**, 011801 (2005) [arXiv:hep-ex/0408034].

[14] K. Abe et al., Phys. Rev. Lett. **96**, 221601 (2006) [arXiv:hep-ex/0506079].

[15] B. Aubert et al. [*BABAR* Collaboration], Phys. Rev. Lett. **98**, 151802 (2007) [arXiv:hep-ex/0612017].

[16] N. Taniguchi, M. Nakao, and S. Nishida for the Belle Collaboration, arXiv:0804.4770 [hep-ex].

[17] B. Aubert et al. [*BABAR* Collaboration], Phys. Rev. Lett. **88**, 101805 (2002) [arXiv:hep-ex/0110065].

[18] http://utfit.roma1.infn.it/ckm-rare/ckm-rare.html

[19] B. Aubert et al. [*BABAR* Collaboration], Nucl. Instrum. Meth. A **479**, 1 (2002) [arXiv:hep-ex/0105044].

[20] S. Agostinelli et al. [GEANT4 Collaboration], Nucl. Instrum. Meth. A **506**, 250 (2003).

[21] G. C. Fox and S. Wolfram, Nucl. Phys. B **149**, 413 (1979) [Erratum—ibid. B **157**, 543 (1979)].

[22] B. Aubert et al. [*BABAR* Collaboration], Phys. Rev. Lett. **89**, 201802 (2002) [arXiv:hep-ex/0207042].

[23] E. D. Bloom et al., In the Proceedings of 4th European Particle Accelerator Conference (EPAC 94), London, England, 27 Jun–1 Jul 1994, 464–466.

[24] J. Friedman and N. Fisher, "Bump hunting in high dimensional data, Statistics and Computing," **9**, 123 (1999).

[25] M. R. Teague, "Image analysis via the general theory of moments," J. Opt. Soc. Am. **70**, 920 (1980).

[26] W. M. Yao et al. [Particle Data Group], J. Phys. G **33**, 1 (2006).

[27] I. Narsky, "StatPatternRecognition: A C++ Package for Statistical Analysis of High Energy Physics Data," physics/0507143, (2005).

[28] C. Gini, "Variabilit e mutabilit" (1912). Reprinted in *Memorie di metodologica statistica* (E. Pizetti and T. Salvemini). Rome: Libreria Eredi Virgilio Veschi (1955).

[29] C. E. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal **27**, 379–423, 623–656, (1948).

[30] Y. Freund and R. E. Shapire, "A short introduction to boosting," Journal of Japanese Society for Artificial Intelligence, **14(5)**, 771–780 (1999).

[31] L. Breiman, "Bagging Predictors," Machine Learning **24(2)**, 123140 (1996).

[32] E. Parzen, "On Estimation of a Probability Density Function and Mode," Ann. Math. Stat. **33(3)**, 1065 (1962).

[33] W. Verkerke and D. Kirkby, "The RooFit Toolkit for Data Modeling," http://roofit.sourceforge.net/.

[34] L. Zhang, "RooRarFit," http://rarfit.sourceforge.net/.

[35] J. Gaiser et al., Phys. Rev. D **34**, 711 (1986).

[36] H. Albrecht et al. [ARGUS Collaboration], Z. Phys. C **48**, 543 (1990).