ICHEP94 Ref. GLS0392 **CLEO CONF 94-1**
Submitted to Pa 21,2 Pl 9,19

First Measurement of the Inclusive Rate for the Radiative Penguin Decay $b \to s\gamma$

B. Barish,[1] M. Chadha,[1] S. Chan,[1] D.F. Cowen,[1] G. Eigen,[1] J.S. Miller,[1] C. O'Grady,[1] J. Urheim,[1] A.J. Weinstein,[1] M. Athanas,[2] W. Brower,[2] G. Masek,[2] H.P. Paar,[2] J. Gronberg,[3] R. Kutschke,[3] S. Menary,[3] R.J. Morrison,[3] S. Nakanishi,[3] H.N. Nelson,[3] T.K. Nelson,[3] C. Qiao,[3] J.D. Richman,[3] A. Ryd,[3] D. Sperka,[3] H. Tajima,[3] M.S. Witherell,[3] R. Balest,[4] K. Cho,[4] W.T. Ford,[4] D.R. Johnson,[4] K. Lingel,[4] M. Lohner,[4] P. Rankin,[4] J.G. Smith,[4] J.P. Alexander,[5] C. Bebek,[5] K. Berkelman,[5] K. Bloom,[5] T.E. Browder,[5*] D.G. Cassel,[5] H.A. Cho,[5] D.M. Coffman,[5] D.S. Crowcroft,[5] P.S. Drell,[5] D. Dumas,[5] R. Ehrlich,[5] P. Gaidarev,[5] M. Garcia-Sciveres,[5] B. Geiser,[5] B. Gittelman,[5] S.W. Gray,[5] D.L. Hartill,[5] B.K. Heltsley,[5] S. Henderson,[5] C.D. Jones,[5] S.L. Jones,[5] J. Kandaswamy,[5] N. Katayama,[5] P.C. Kim,[5] D.L. Kreinick,[5] G.S. Ludwig,[5] J. Masui,[5] J. Mevissen,[5] N.B. Mistry,[5] C.R. Ng,[5] E. Nordberg,[5] J.R. Patterson,[5] D. Peterson,[5] D. Riley,[5] S. Salman,[5] M. Sapper,[5] F. Würthwein,[5] P. Avery,[6] A. Freyberger,[6] J. Rodriguez,[6] S. Yang,[6] J. Yelton,[6] D. Cinabro,[7] T. Liu,[7] M. Saulnier,[7] R. Wilson,[7] H. Yamamoto,[7] T. Bergfeld,[8] B.I. Eisenstein,[8] G. Gollin,[8] B. Ong,[8] M. Palmer,[8] M. Selen,[8] J. J. Thaler,[8] K.W. Edwards,[9] M. Ogg,[9] A. Bellerive,[10] D.I. Britton,[10] E.R.F. Hyatt,[10] D.B. MacFarlane,[10] P.M. Patel,[10] B. Spaan,[10] A.J. Sadoff,[11] R. Ammar,[12] P. Baringer,[12] A. Bean,[12] D. Besson,[12] D. Coppage,[12] N. Copty,[12] R. Davis,[12] N. Hancock,[12] M. Kelly,[12] S. Kotov,[12] I. Kravchenko,[12] N. Kwak,[12] H. Lam,[12] Y. Kubota,[13] M. Lattery,[13] M. Momayezi,[13] J.K. Nelson,[13] S. Patton,[13] R. Poling,[13] V. Savinov,[13] S. Schrenk,[13] R. Wang,[13] M.S. Alam,[14] I.J. Kim,[14] Z. Ling,[14] A.H. Mahmood,[14] J.J. O'Neill,[14] H. Severini,[14] C.R. Sun,[14] F. Wappler,[14] G. Crawford,[15] C. M. Daubenmier,[15] R. Fulton,[15] D. Fujino,[15] K.K. Gan,[15] K. Honscheid,[15] H. Kagan,[15] R. Kass,[15] J. Lee,[15] R. Malchow,[15] M. Sung,[15] C. White,[15] M.M. Zoeller,[15] F. Butler,[16] X. Fu,[16] B. Nemati,[16] W.R. Ross,[16] P. Skubic,[16] M. Wood,[16] M. Bishai,[17] J. Fast,[17] E. Gerndt,[17] R.L. McIlwain,[17] T. Miao,[17] D.H. Miller,[17] M. Modesitt,[17] D. Payne,[17] E.I. Shibata,[17] I.P.J. Shipsey,[17] P.N. Wang,[17] M. Battle,[18] J. Ernst,[18] L. Gibbons,[18] Y. Kwon,[18] S. Roberts,[18] E.H. Thorndike,[18] C.H. Wang,[18] J. Dominick,[19] M. Lambrecht,[19] S. Sanghera,[19] V. Shelkov,[19] T. Skwarnicki,[19] R. Stroynowski,[19] I. Volobouev,[19] G. Wei,[19] P. Zadorozhny,[19] M. Artuso,[20] M. Gao,[20] M. Goldberg,[20] D. He,[20] N. Horwitz,[20] G.C. Moneti,[20] R. Mountain,[20] F. Muheim,[20] Y. Mukhin,[20] S. Playfer,[20] Y. Rozen,[20] S. Stone,[20] X. Xing,[20] G. Zhu,[20] J. Bartelt,[21] S.E. Csorna,[21] Z. Egyed,[21] V. Jain,[21] D. Gibaut,[22] K. Kinoshita,[22] and P. Pomianowski[22]

(CLEO Collaboration)

[1] *California Institute of Technology, Pasadena, California 91125*
[2] *University of California, San Diego, La Jolla, California 92093*
[3] *University of California, Santa Barbara, California 93106*
[4] *University of Colorado, Boulder, Colorado 80309-0390*
[5] *Cornell University, Ithaca, New York 14853*
[6] *University of Florida, Gainesville, Florida 32611*
[7] *Harvard University, Cambridge, Massachusetts 02138*
[8] *University of Illinois, Champaign-Urbana, Illinois, 61801*
[9] *Carleton University, Ottawa, Ontario K1S 5B6 and the Institute of Particle Physics, Canada*
[10] *McGill University, Montréal, Québec H3A 2T8 and the Institute of Particle Physics, Canada*
[11] *Ithaca College, Ithaca, New York 14850*
[12] *University of Kansas, Lawrence, Kansas 66045*
[13] *University of Minnesota, Minneapolis, Minnesota 55455*
[14] *State University of New York at Albany, Albany, New York 12222*
[15] *Ohio State University, Columbus, Ohio, 43210*
[16] *University of Oklahoma, Norman, Oklahoma 73019*
[17] *Purdue University, West Lafayette, Indiana 47907*
[18] *University of Rochester, Rochester, New York 14627*
[19] *Southern Methodist University, Dallas, Texas 75275*
[20] *Syracuse University, Syracuse, New York 13244*
[21] *Vanderbilt University, Nashville, Tennessee 37235*
[22] *Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 24061*

Abstract

We have measured the branching ratio for the inclusive radiative penguin process $b \to s\gamma$. We find

$$BR(b \to s\gamma) = (2.32 \pm 0.51 \pm 0.29 \pm 0.32) \times 10^{-4} \quad ,$$

where the first error is statistical, the second error is the additive systematic error from uncertainty in yield, and the third error is the multiplicative systematic error from uncertainty in efficiency, which includes model dependence.

---

*Permanent address: University of Hawaii at Manoa

I. INTRODUCTION

Last year CLEO observed [1] the decay $B \to K^*(892)\gamma$, thereby establishing the existence of the radiative penguin process $b \to s\gamma$. However, the fraction of the inclusive $b \to s\gamma$ rate which hadronizes into $B \to K^*(892)\gamma$ exhibits large model variation. The inclusive $b \to s\gamma$ branching ratio is more reliably predicted [2], at $(2.75 \pm 0.80) \times 10^{-4}$ for a top quark mass of 175 GeV. The branching ratio is sensitive to the existence of a charged Higgs [3], to anomalous $WW\gamma$ couplings [4], and other non-Standard-Model phenomena [5]. Here we present the first measurement of the inclusive branching ratio for $b \to s\gamma$.

Our signature for $b \to s\gamma$ is a photon from B-meson decay with an energy between 2.2 and 2.7 GeV. Model calculations indicate that 75-90% of the photons from $b \to s\gamma$ decays lie in this range. The background of photons in this range from other B decay processes is small and can be reliably determined. The background from continuum processes under the Υ(4S) resonance is large. We suppress it with various techniques, and then eliminate what is left by an On resonance $-$ Off resonance subtraction.

The outline of this paper is as follows. In Section II we briefly describe the detector and the data sample and outline the general procedures for event selection. In Section III we describe two methods for suppressing continuum processes. Section IV deals with the determination of backgrounds from $B\bar{B}$ events. Yields are presented in Section V and efficiency calculations discussed in Section VI. Branching ratios, with their statistical and systematic errors, are given in Section VII, and the results are compared with theory in Section VIII.

II. DETECTOR, DETECTOR SAMPLE, EVENT SELECTION

The data used in this analysis were taken with the CLEO-II detector operating at the Cornell Electron Storage Ring (CESR), and consist of 2.01 fb^{-1} taken on the Υ(4S) resonance and 0.96 fb^{-1} taken at a center-of-mass energy 60 MeV below the resonance. (The on resonance sample consists of 2.15 million $B\bar{B}$ events and 6.6 million continuum events.)

CLEO II is a general purpose solenoidal detector. It detects and measures charged particles over 90% of 4π steradians with a system of cylindrical drift chambers. Its barrel and endcap CsI electromagnetic calorimeters cover 98% of 4π. In the "good barrel region" used for detecting the high energy photons for this analysis, 70% of 4π, energy resolution is σ_E/E=2% near 2.5 GeV. A more complete description of CLEO II is given in Ref. [6].

We select events that pass general hadronic event selection criteria. We further require that the event contain a high energy cluster in the good barrel region ($|\cos\theta_\gamma| < 0.7$) of the electromagnetic calorimeter, not matched to a charged particle track from the tracking chambers. As the vast majority of photons, both from continuum and $B\bar{B}$ events, are decay products of π^0 and η, we discard those high energy clusters which, when paired with another γ in the event, have a $\gamma\gamma$ mass consistent with a π^0 or η. Finally, we require that the lateral energy distribution of the cluster be consistent with that of a single isolated photon, thus suppressing random overlaps, merged π^0's, and non-photon clusters.

III. CONTINUUM SUPPRESSION AND SUBTRACTION

Continuum suppression is complicated by the presence of two different continuum processes, $e^+e^- \to q\bar{q}$, with the high energy photon coming from the $q\bar{q}$ hadronization (unvetoed π^0 and η decays, $\omega \to \pi^0\gamma$, etc.), and initial state radiation (ISR) $e^+e^- \to q\bar{q}\gamma$. We have developed two complementary continuum suppression techniques, each of which suppresses both $q\bar{q}$ and ISR backgrounds. In the first technique, we combine eight event-shape variables into a single variable using a neural network [7] – [8]. In the second technique, we attempt B reconstruction with several $b \to s\gamma$ decay modes, accepting the event if it is reconstructible in at least one mode. These two techniques are described in greater detail below. We carry analyses through with the two techniques in Sections IV-VI, and combine the results in Section VII.

A. Continuum Suppression with a Neural Network

The following eight variables are effective at discriminating between $b \to s\gamma$ and continuum events.

R_2. This is the normalized second Fox-Wolfram moment [9]. R_2 is large for $q\bar{q}$ events and small for $B\bar{B}$ events (but $b \to s\gamma$ has higher R_2 than generic $B\bar{B}$ does). For ISR events the distribution in R_2 is similar to that for $b \to s\gamma$ events. The R_2 distributions for $b \to s\gamma$, $q\bar{q}$, and ISR Monte Carlo events are shown in Fig. 1a.

$S_\perp$. This variable takes advantage of the fact that the high energy photon gives the event an axis. It is defined as the sum of the magnitudes of the momenta transverse to the photon direction, summed over all particles more than 45° away from the photon axis, divided by the sum of the magnitudes of the momenta of all particles except the photon. The $S_\perp$ distributions for $b \to s\gamma$, $q\bar{q}$, and ISR Monte Carlo events are shown in Fig. 1b. Note that $q\bar{q}$ events have small $S_\perp$, while some ISR events have large $S_\perp$.

To obtain better suppression of ISR, we use two variables evaluated in the rest frame of the e^+e^- following the radiation of the high energy photon (the primed frame). R_2' is the normalized Fox-Wolfram second moment evaluated in this frame, excluding the photon from the computation. Fig. 1c shows the distributions of R_2' for $b \to s\gamma$, $q\bar{q}$, and ISR Monte Carlo events.

$\cos\theta'$. This is the cosine of the angle between the photon and the thrust axis of the rest of the event, in the primed frame. It peaks towards 1.0 for $b \to s\gamma$, and is isotropic for ISR. The $\cos\theta'$ distributions for $b \to s\gamma$, $q\bar{q}$, and ISR Monte Carlo events are shown in Fig. 1d.

The cone surrounding the high energy photon ("forward cone") will contain little additional energy in $b \to s\gamma$ and ISR events, while in $q\bar{q}$ events it will contain $\sim (E_{beam} - E_\gamma)$. The cone away from the high energy photon ("backward cone") will contain little energy for ISR events, $\sim (E_{beam} - E_\gamma)$ for $b \to s\gamma$ events, and $\sim E_{beam}$ for $q\bar{q}$ events. Energy distributions for forward 20° and 30° cones and backward 20° and 30° cones, the cone energy variables used, are shown in Fig. 2.

In evaluating the eight variables just described, we include all well-measured charged particle tracks, and all calorimeter showers above 50 MeV not matched to charged particle tracks.

The neural network [7] – [8] combines the 8 input variables into a single output variable, r, which ranges from -1 to $+1$, and tends toward $+1$ for $b \to s\gamma$ events and -1 for continuum events. To obtain an output variable which discriminates well between signal and background, the net must be *trained*. This is accomplished by optimizing the free parameters of the neural net function with large Monte Carlo samples of $b \to s\gamma$ (400K events) and continuum (200M events). By using a neural network, we exploit correlations among the eight input variables. [1] The distributions in the net variable r, for Monte Carlo $b \to s\gamma$ and continuum events and for off-resonance data, are shown in Fig. 3. There is substantial discrimination between $b \to s\gamma$ and continuum, and good agreement between continuum Monte Carlo and off-resonance data.

We perform a weighted sum of candidate events with an r-dependent weighting that is large for r near $+1$ and small for r near -1. The weighting is chosen to maximize separation between signal and background. This procedure is equivalent to making a one-parameter fit to the background-subtracted r distribution with the r distribution expected for $b \to s\gamma$.

B. Continuum Suppression via B Reconstruction

In this technique the continuum is suppressed by searching for at least one combination of particles in the event which fits the $B \to X_s\gamma$ decay hypothesis. A fully inclusive analysis is achieved by using a large number of exclusive decay modes with γ, $K^\pm$ or K_s^0 (reconstructed in $\pi^+\pi^-$ decays), and up to four pions (which may include zero or one π^0). The attempt to reconstruct any of these exclusive modes in an event provides continuum suppression, and does not necessarily give the correct interpretation of the decay channel. Cuts are optimized on signal and background Monte Carlo for best sensitivity to the inclusive rate. No attempt is made to measure the $b \to s\gamma$ rate mode-by-mode, which would require higher statistics, removal of reconstruction ambiguities and subtraction of cross-feed, and most likely would need different cut optimization.

$K^\pm$ and $\pi^\pm$ candidates are required to have measured dE/dx values within $\pm 3\sigma$ of the mean for the appropriate mass hypothesis. The $K^\pm$ cuts are tightened to 2.5 σ after the best track combination is selected. Any charged particle pair consistent with a displaced vertex and with mass within 3σ of the K_s^0 mass is accepted as a K_s^0 candidate. Photons for a π^0 candidate must have energies above 60 MeV. Photon pairs within -3.5σ and $+3.0$ σ of the π^0 mass are accepted as π^0 candidates. They are constrained to the π^0 mass before being combined with other particles to form a candidate B.

For each set of candidate particles, we form the total energy E and momentum P from the sums of the energies and vector momenta, respectively. The energy difference ΔE, beam-constrained mass M, and mass difference ΔM are then calculated as $\Delta E = E - E_{beam}$, $M = \sqrt{E_{beam}^2 - P^2}$, and $\Delta M = M -$ 5.279 GeV. For the correct set of candidate particles, ΔE and ΔM will be zero, within measurement error. Consequently we construct a variable $\chi_B^2 = (\Delta M/\sigma_M)^2 + (\Delta E/\sigma_E)^2$, where σ_E and σ_M are the expected measurement errors. χ_B^2

[1]We use a two layer, fully-connected neural network with 10 nodes in the first layer, and 7 in the second.

reflects how well the kinematic properties of the set of candidate particles represent a B meson.

For each event with a high energy photon, there are many combinations of candidate particles to consider. Within the same decay mode we pick the combination that minimizes an overall χ^2, which includes χ^2_B together with contributions from dE/dX, and K^0_s and π^0 mass deviations if any of these particles are involved. We then choose the best combinations in the $K^{\pm}$ and K^0_s classes, separately by minimizing χ^2_B plus the contribution from the kaon (dE/dX or $M_{\pi^+\pi^-}$). Finally, we choose between $K^{\pm}$ and K^0_s modes on the basis of the lower χ^2_B. Thus, only one combination per event is allowed.

For the chosen candidate we calculate $\cos\theta_{tt}$, the cosine of the angle between the thrust axis of the $B \to X_s\gamma$ candidate tracks and photons, and the thrust axis for all other particles in the event. The distribution in $\cos\theta_{tt}$, for Monte Carlo $b \to s\gamma$ and continuum events is shown in Fig. 4. The background (continuum) events peak strongly at $\pm$ 1, while the distribution for signal ($b \to s\gamma$) events is flat. The distribution in χ^2_B for Monte Carlo background and signal events is shown in Fig. 5. While the continuum has a peak at small χ^2_B, it also has a substantial component at large χ^2_B. The final cuts, $|\cos\theta_{tt}| < 0.6$ and $\chi^2_B <$ 6.0, were determined with an optimization procedure using signal and background Monte Carlo samples.

C. Continuum Subtraction

Remaining continuum background is eliminated with a "continuum subtraction" using off-resonance data. To obtain a result free from continuum background, the scaled off-resonance yield is subtracted from the on-resonance yield; scaling corrects for the very different integrated luminosities and slightly different center-of-mass energy of the two samples.

The standard assumptions for scaling are that the number of continuum events is proportional to luminosity and inversely proportional to E^2_{cm}, and that momenta are proportional to E_{cm}. By comparing continuum Monte Carlo events at nearby center-of-mass energies, we find that the standard assumptions overestimate the continuum subtraction by 1.5% for the neural net analysis and by 3.0% for the B-reconstruction analysis. This is due to a change in the character of continuum events with E_{cm}. At a lower energy, continuum events are less jet-like, and therefore less effectively suppressed. The B-reconstruction technique suppresses the continuum harder than the neural net technique, and therefore has a larger oversubtraction. Having determined the size of the oversubtraction, we apply an appropriate correction.

IV. BACKGROUND FROM THE Υ(4S)

We determine the background from Υ(4S) decays other than $b \to s\gamma$ via a two-step procedure. In the first step, the yield predicted by analyzing $b \to c$ and $b \to u$ Monte Carlo samples is found. This yield is dominated by γ's that are decay products of π^0 or η and have escaped the π^0 or η veto.

In the second step, the high momentum portions of the momentum spectra for π^0 and η as predicted by the Monte Carlo samples are compared with those observed in On-Off-subtracted data. In this step, we apply all the cuts and procedures used for $b \to s\gamma$, treating the high momentum π^0 or η as if it were the high energy γ. We then obtain a correction to the yield obtained in the first step, based on the difference between the Monte Carlo and measured π^0 and η momentum spectra.

With this two step procedure, we depend on the correctness of the Monte Carlo only for predicting the very small component of background not from π^0 or η, and for predicting the π^0 and η veto efficiencies. The π^0 and η momentum spectra for that region of phase space passing our cuts, the key ingredient to determining the background, are obtained from data.

Assuming the Υ(4S) has a partial width for decay to non-$B\bar{B}$ final states comparable to those of the Υ(1S), Υ(2S), and Υ(3S), the Υ(4S) will decay to non-$B\bar{B}$ final states with a branching ratio near 10^{-3}. We determine the background from this source using Υ(3S) data and find it to be very small.

Backgrounds from the Υ(4S) are listed in Table 1.

V. YIELDS

Figure 6a shows the photon energy spectrum obtained with the neural net procedure. The dotted histogram represents the scaled off-resonance data. The points with error bars represent the sum of scaled off-resonance data and calculated Υ(4S) background. The solid histogram represents the on-resonance data. For energies below 2.2 GeV, there is a substantial Υ(4S) background contribution, as indicated by the difference between the dotted histogram and the points. This background is well estimated, as indicated by the agreement between the point and the solid histogram for the lowest bin, where we expect no contribution from signal. Above 2.7 GeV, the expected contributions from signal and from Υ(4S) background are negligible. In this control region, 3 of the 4 points agree with the solid histogram. In the signal region, 2.2-2.7 GeV, the solid histogram is significantly above the data points. This difference is the signal for $b \to s\gamma$.

Figure 7a shows the photon energy spectrum obtained with the B-reconstruction procedure. The symbols have the same significance as in Fig. 6, and the features are similar. In particular, there is a substantial Υ(4S) background below 2.2 GeV, and the calculated background agrees well with data. The excess of solid histogram over points between 2.2 and 2.7 GeV, signifying $b \to s\gamma$, is apparent. By 2.8 GeV the efficiency has fallen to zero, and so the absence of signal above this point has no significance.

Figures 6b and 7b show the photon energy spectra for on-resonance minus off-resonance minus Υ(4S) background, the $b \to s\gamma$ signal, for the two analysis procedures. The curve shown on the figures is the spectrum calculated with a spectator model [11]. Agreement between calculation and data is satisfactory.

With the neural net analysis, there is an excess between 3.0 and 3.5 GeV, visible in Fig. 6b, with a statistical significance of 1.8 σ. We have searched for explanations other than "statistical fluctuation" and have found none. The four points in the control region above 2.7 GeV have a summed χ^2 of 4.9, for four degrees of freedom. A χ^2 this large or larger is expected in 30% of samples. We have modified the B-reconstruction procedure to test for an excess above 2.7 GeV with that general approach, and find none.

The B-reconstruction technique selects a "best" X_s candidate for $B \to X_s\gamma$ decay. Thus, in addition to providing a photon energy spectrum, it also provides an "apparent X_s mass spectrum." The background-subtracted apparent X_s mass spectrum is shown in Fig. 8. This spectrum must not be taken literally. Reconstruction ambiguities and cross-feed must be taken into account before quantitative use can be made of it. Yet, two qualitative features can be seen: (1) There is clear evidence for $B \to K^*(892)\gamma$; and (2) There is clear evidence for X_s systems in the 1-2 GeV range. The rate for $K^*(892)\gamma$ extracted by a fit to the distribution in Fig. 8 is consistent with our previous measurement [1].

The yields between 2.2 and 2.7 GeV, for the two analyses, for on-resonance data, scaled off-resonance data, and calculated $\Upsilon(4S)$ background, are given in Table 1.

VI. EFFICIENCY

To determine the efficiency for detecting a $b \to s\gamma$ decay with either procedure, one needs a model for the $b \to s\gamma$ decay process. While there are several calculations [10] of the rates for $b \to s\gamma$ decays to individual kaon resonances, the only fully-inclusive model is that of Ali and Greub [11], a spectator model that includes gluon bremsstrahlung and higher order radiative effects. We use that model. It gives the X_s mass distribution for the dressed process $B \to X_s\gamma$, but does not describe the decay of X_s. For that we use conventional models of quark hadronization.

The Ali-Greub spectator model, as normally used, describes two aspects of the B meson, namely the b-quark mass and the b-quark momentum distribution, with a single parameter, P_F. This tight coupling of b-quark mass and momentum was invented independently by Ali and Pietarinen [12], Poling [13], and Altarelli *et al.* [14], as a means of meeting the constraints of energy and momentum conservation, and was applied to semileptonic B decay. It may not be appropriate for $b \to s\gamma$, where the average b-quark mass is an important parameter. As an alternative, Ali and Greub [15] suggested making the spectator quark mass an adjustable parameter. We have done so, fitting CLEO's lepton spectrum from semileptonic B decay [16] with a spectator model [12] [14], varying P_F, charm quark mass, and spectator quark mass. The data do not constrain the momentum-averaged b-quark mass tightly. Consequently we take $< m_b > = 4.87 \pm 0.10$ GeV, an average of two recent calculations [17]. For $< m_b >=$ 4.87 GeV, the fit to the lepton spectrum gives $P_F = 270 \pm 20$ MeV/c. We use that value for modeling $b \to s\gamma$, but with an error of ± 40 MeV/c to allow for variations of P_F with m_b.

We split the efficiency into two components; (i) The fraction of $b \to s\gamma$ decays giving a photon with energy between 2.2 and 2.7 GeV, and (ii) the efficiency for detecting a $b \to s\gamma$ decay, given that it has a photon in the 2.2 - 2.7 GeV interval. The first component depends on the model for $b \to s\gamma$ decays, and is independent of experimental procedures. The second component depends on experimental procedures, and less strongly on the $b \to s\gamma$ model.

The fraction of $b \to s\gamma$ decays between 2.2 - 2.7 GeV is 0.85, the efficiency for detecting such decays is 0.38 with the neural net analysis and 0.104 with the B-reconstruction analysis. Systematic errors in these numbers will be discussed in the next section.

VII. BRANCHING RATIOS, SYSTEMATIC ERRORS

A. Individual Branching Ratio

The background-subtracted yields given in Table 1 and the efficiencies given in Section VI, along with a sample size of 2.15×10^6 $B\bar{B}$ events, combine to give the following $b \to s\gamma$ branching ratios.

$$BR(b \to s\gamma, 2.2 < E_\gamma < 2.7) = (1.60 \pm 0.57) \times 10^{-4}, \text{and } (2.35 \pm 0.53) \times 10^{-4} \quad ,$$

for neural net and B-reconstruction analyses, respectively. The conditional branching ratio $BR(b \to s\gamma, 2.2 < E_\gamma < 2.7)$ is for *laboratory* photon energies between 2.2 and 2.7 GeV, *not* B-meson rest frame energies. Extrapolating to the full energy range,

$$BR(b \to s\gamma) = (1.88 \pm 0.67) \times 10^{-4}, \text{and } (2.76 \pm 0.63) \times 10^{-4} \quad ,$$

for neural net and B-reconstruction analyses, respectively. The errors given are statistical only.

B. Consistency Checks

As checks, we have performed several variations on both analysis procedures. For the neural net analysis, we have replaced the weighted sum with a cut, $r > 0.6$. We have also dropped two of the cone energy variables and combined the remaining 6 variables not with a neural network but instead with a brute force technique that selects the hypervolume in the 6-dimensional variable space containing volume elements with signal/background above some optimum cut value. Finally, we have performed conventional "square cut" analyses with subsets of the variables. As one moves down, from neural net weighted sum analysis to square cut analysis, the statistical power progressively decreases. Yet the results of all analyses are in good agreement.

For the B-reconstruction analysis, we have varied the method used to find the best combination of candidate B decay products by removing the ΔM or dE/dX terms from the overall χ^2 and by distorting the measured mass or photon energy by amounts larger than the detector resolution. We have also replaced the cuts on χ^2_B and $\cos\theta_{tt}$ by fits to the background-subtracted data. In every variation, the data closely follow the behavior of the $b \to s\gamma$ signal predicted by the simulation. We have also compared the observed signal and the signal simulation for a number of distributions characterizing $B \to X_s\gamma$ decays: multiplicity, $K^\pm/K^0_s$ ratio, fraction of events with π^0, and charge of the B candidate. Agreement is excellent in all variables.

C. Combining Results from the Two Methods

The results obtained with neural net and B-reconstruction techniques can be combined to give a result more accurate than either. We do so by taking a weighted average, allowing for the correlations between the two analyses. To avoid bias, the relative weighting of the two results is determined from a Monte Carlo study, which indicates that the two methods have approximately equal statistical power, thus suggesting a 50:50 weighting.

We determine the degree of statistical correlation between the two analyses by performing ensembles of toy Monte Carlo experiments whose input includes the observed correlations from standard Monte Carlo samples of signal and background. These studies indicate that the statistical error on the combined result is 1.12 times larger than the error would be if the analyses were uncorrelated. The statistical error on the difference between the two results is 0.89 times smaller than the error would be if the analyses were uncorrelated. Thus, the difference between the two fully-inclusive measurements is $(0.88 \pm 0.82) \times 10^{-4}$. The two analyses agree within 1.1 σ. Their average has a statistical error of 0.51×10^{-4}.

D. Systematic Errors

Systematic errors are of three varieties, those that affect the computation of yield (additive systematic errors), those that affect the computation of efficiency given a model for $b \to s\gamma$ (multiplicative systematic errors), and the variation in computed efficiency due to the choice of model for $b \to s\gamma$ (model dependence).

1. Additive Systematic Errors

The systematic error in yield due to the On$-$Off subtraction has a component from the ΔE_{cm} correction, taken as half the correction, and a component from the On-Off luminosity ratio uncertainty of $\pm$ 0.25% [18]. Since both analyses obtain the ΔE_{cm} correction by the same procedure, we assume the systematic errors are fully correlated. The errors from the luminosity ratio uncertainty are also fully correlated. The systematic errors obtained for the two analyses and the combined result are given in Table 2.

There is also a systematic error in yield from the $B\bar{B}$ subtraction. It has components from $b \to c$ Monte Carlo statistics, $b \to u$ Monte Carlo statistics, π^0 corrections, η corrections, and π^0 veto modeling. These are listed in Table 2. In combining the errors from the two analyses, we assume all but the $b \to u$ Monte Carlo statistics are fully correlated (there we used independent MC samples).

2. Multiplicative Systematic Errors

Systematic errors in the efficiency aside from modeling $b \to s\gamma$ include uncertainties in modeling the other B and in simulating the detector response. We have investigated charged particle track-finding efficiency, photon-finding efficiency, rates for losing good photons because of random π^0 or η vetoes and random matches to charged tracks, and π^0-finding efficiency. To study modeling the other B, we have compared Monte Carlo and data distributions in the neural net variable r for inclusive semimuonic B decay, treating the muon as the high energy photon. Finally, we have varied many cuts in the B-reconstruction analysis, and seen how well the Monte Carlo follows the data. From these studies we conclude that the multiplicative systematic error on the combined result is 7%. This number is listed in Table 3.

3. Model Dependence

A full description of the process $b \to s\gamma$ must give the mass distribution of X_s, and the decays of X_s as a function of its mass, for the dressed process $B \to X_s\gamma$. We take the mass distribution from the Ali-Greub spectator model, with $< m_b >= 4.87 \pm 0.10$ GeV, and $P_F = 270 \pm 40$ MeV/c.

A second area of uncertainty in modeling $b \to s\gamma$ is the hadronization process. The critical variable is the mean multiplicity of the decay. We take $\pm$ 1/4 unit of multiplicity as the uncertainty here. Other aspects of X_s decay, in addition to multiplicity, are the fraction of final state particles that are π^0's (expected to be 1/3 of the pions if the initial state is an equal mix of charged and neutral B's, $f_{+-}/f_{00} = 1$),and the division between $K^\pm$ and $K^0/\bar{K}^0$ (expected to be 1:1 under the same assumption).

The two analyses depend on the features of the modeling of $b \to s\gamma$ decay in different ways. For the neural net analysis, the efficiency falls roughly linearly with increasing M_{X_s}, and at $M_{X_s} = 1.4$ GeV the fractional change in efficiency for a 100 MeV change in M_{X_s} is $\approx 10\%$. For the B-reconstruction analysis, there is practically no dependence on M_{X_s} except at high masses, near 2.0 GeV. For the neural net analysis, the efficiency *rises* with increasing multiplicity, by 0 to 20% per unit of multiplicity, depending on M_{X_s}. For the B-reconstruction analysis, the efficiency *falls* with increasing multiplicity, by 20% per unit of multiplicity. For the net analysis, the efficiency falls typically 6% of itself when a $\pi^\pm$ is replaced with a π^0. For the reconstruction analysis, this effect is much larger, typically 50%. For the net analysis, the efficiency for modes involving $K^\pm$ is typically a factor of 1.5 larger than that for equivalent modes involving $K^0/\bar{K}^0$, while for the reconstruction analysis, the efficiency ratio is 3.0.

Table 3 lists the systematic errors from model dependence: from $\delta m_b = \pm 100$ MeV, $\delta P_F = \pm 40$ MeV/c, from hadronization uncertainties, and from $\delta(f_{+-}/f_{00}) = \pm 0.2$. For the errors from m_b and P_F, we separately show the systematic error on the fully inclusive branching ratio and the error on the conditional branching ratio, $2.2 < E_\gamma < 2.7$ GeV. The total systematic error on the combined result is smaller than that on either individual result.

E. Final Results

Combining the two analyses, and including systematic errors, we obtain

$$BR(b \to s\gamma, 2.2 < E_\gamma < 2.7) = (1.98 \pm 0.43 \pm 0.25 \pm 0.23) \times 10^{-4} \quad ,$$

and

$$BR(b \to s\gamma) = (2.32 \pm 0.51 \pm 0.29 \pm 0.32) \times 10^{-4} \quad ,$$

where the first error is statistical, the second error is the additive systematic error from uncertainty in yield, and the third error is the multiplicative systematic error from uncertainty in efficiency. The conditional branching ratio $BR(b \to s\gamma, 2.2 < E_\gamma < 2.7)$ is for *laboratory* photon energies between 2.2 and 2.7 GeV, *not* B rest frame energies.

For the purpose of constraining extensions of the Standard Model, it is desirable to have upper and lower limits on the fully-inclusive branching ratio. To obtain these, we combine the statistical and additive systematic errors in quadrature, since the additive systematic error is dominated by the statistics of the π^0 and η corrections. We split the multiplicative systematic error into a component due to δm_b, and the rest, and show the dependence on m_b explicitly, writing:

$$BR(b \to s\gamma) = (2.32 \pm 0.59)(1 \pm 0.091)(1 - (m_b - 4.87)) \times 10^{-4},$$

where m_b is in GeV. The 95% confidence level upper and lower limits on the first factor are 3.29 and 1.35, respectively. To conservatively allow for the multiplicative systematic error of 9.1%, we increase (decrease) the upper (lower) limit by $1.64 \times 9.1\%$, obtaining:

$$BR(b \to s\gamma) < 3.78 \times (1.0 - (m_b - 4.87)) \times 10^{-4},$$

and

$$BR(b \to s\gamma) > 1.15 \times (1.0 - (m_b - 4.87)) \times 10^{-4}.$$

We obtain

$$1 \times 10^{-4} < BR(b \to s\gamma) < 4 \times 10^{-4}.$$

VIII. IMPLICATIONS

The rate for $b \to s\gamma$ has been calculated [2] within the framework of renormalization-group-improved perturbation theory. There is now general agreement among theorists on the leading logarithmic calculation. Short distance QCD effects due to hard gluon exchange enhance the the rate by a factor of 2-3. The largest theoretical uncertainty stems from the choice of the renormalization scale μ. There have been calculations of some, but not all, of the next-to-leading logarithmic QCD corrections [19]. For consistency, we do *not* include these NLL corrections. In Fig. 9, we show the prediction for the $b \to s\gamma$ branching ratio as a function of the top quark mass m_t. Following Ali *et al.* [20] and Buras *et al.* [21], we have taken $m_b/2 < \mu < 2m_b$ for the range in the renormalization scale. Also shown in Fig. 9 is our measurement of the branching ratio, and the range of m_t suggested by recent CDF evidence [22]. The measured value of the branching ratio is consistent with the theoretical prediction for any value of m_t in the 100-200 GeV range.

As an example of the sensitivity of $b \to s\gamma$ to new physics, we consider a Two-Higgs-Doublet model [3], with down and up quarks acquiring mass from separate vacuum expectation values v_1 and v_2, i.e., a Model II THDM. The model has two parameters: the charged Higgs mass M_H, and the ratio of vacuum expectation values $tan\beta = v_2/v_1$. In Fig. 10 we plot the contour in $M_H - tan\beta$ space corresponding to our upper limit (allowing for an additional theoretical uncertainty of 10%). We have used $m_t = 175$ GeV, and have used $\mu = 2m_b$ to make the theoretical prediction as small as possible. All points below the contour will have calculated branching ratios that are ruled out by our upper limit.

As a second example, we consider anomalous $W-W-\gamma$ couplings, described by the parameters $\Delta\kappa$ and λ [4]. In Fig. 11 we plot contours in $\Delta\kappa - \lambda$ space corresponding to our $b \to s\gamma$ upper and lower limits (including a 10% theoretical uncertainty aside from the scale μ). For each contour, we use either $\mu = 2m_b$ or $\mu = m_b/2$, whichever gives the more conservative limit. Unshaded regions in $\Delta\kappa - \lambda$ space are excluded by our limits because those regions would have too large, or too small a value for the $b \to s\gamma$ branching ratio. The allowed regions are shown as two cross-hatched strips. The constraint from $p\bar{p}$ collider results on $\bar{p}p \to W\gamma X$ [23] is shown as the ellipse. The two approaches are complementary.

IX. CONCLUSIONS

We have measured the branching ratio for $b \to s\gamma$. Comparisons of data with theory are now limited as much by theoretical uncertainties as by experimental errors. The completion of a full next-to-leading logarithmic calculation is highly desirable, as is further theoretical work on $b \to s\gamma$ models. Our result is fully consistent with the Standard Model. It rules out a charged Higgs lighter than 260 GeV in a Two-Higgs-Doublet, Model II, and rules out a portion of the parameter space of anomalous $WW\gamma$ couplings still permitted by $p\bar{p}$ collider measurements.

X. ACKNOWLEDGEMENTS

We thank Ahmed Ali and Christoph Greub for performing several spectator model calculations for us, and for many interesting conversations. Joanne Hewett and Tom Rizzo have helped us understand the implications of $b \to s\gamma$ for non-standard models. We gratefully acknowledge the effort of the CESR staff in providing us with excellent luminosity and running conditions. This work was supported by the National Science Foundation, the U.S. Dept. of Energy, the Heisenberg Foundation, the SSC Fellowship program of TNRLC, Natural Sciences and Engineering Research Council of Canada, and the A.P. Sloan Foundation.

REFERENCES

[1] R. Ammar *et al.* (CLEO Collaboration), Phys. Rev. Lett. **71**, 674 (1993).

[2] S. Bertolini *et al.*, Phys. Rev. Lett. **59**, 180 (1987); N. Deshpande *et al.*, Phys. Rev. Lett. **59**, 183 (1987); B. Grinstein *et al.*, Phys. Lett. **B202**,138 (1988); R. Grigjanis *et al.*, Phys. Lett. **B224**, 209 (1989); G. Cella *et al.*, Phys. Lett. **B248**, 181 (1990); A.J. Buras *et al.*, Max Planck Institute preprint MPI-Ph/93-77.

[3] T. G. Rizzo, Phys Rev. **D38**, 820 (1988); W.-S. Hou and R.S. Wiley, Phys Lett. **B202**, 591 (1988); C.Q.Geng and J.N. Ng, Phys Rev. **D38**, 2858 (1988); V. Barger, J.L. Hewett, and R.J.N. Phillips, Phys. Rev. **D41**, 3421 (1990); J.L. Hewett, Phys. Rev. Lett. **70**, 1045 (1993); V. Barger, M. Berger, and R.J.N. Phillips, Phys. Rev. Lett. **70**, 1368 (1993).

[4] S.-P. Chia, Phys. Lett. **B240**, 465 (1990); K.A. Peterson, Phys. Lett. **B282**, 207 (1992); T.G. Rizzo, Phys. Lett. **B315**, 471 (1993); U. Baur, in Proceedings of the *Summer Workshop on B Physics*, Snowmass, CO, 1993; X.-G. He and B. McKellar, Phys. Lett. **B320**, 165 (1994).

[5] For a recent review of implications of $b \to s\gamma$ for non-Standard models, see J.L. Hewett, SLAC-PUB-6521 (May 1994).

[6] Y. Kubota *et al.* (CLEO Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. **A320**, 66 (1992).

[7] Rumelhart D.E. and McClelland J.L., *Parallel Distributed Processing*, Vol. **1–2** (Bradford Books, MIT Press, Cambridge, MA) 1986.

[8] Minsky M.L. and Papert S., *Perceptrons: an Introduction to Computational Geometry* (exp. ed.) (MIT Press, Cambridge, MA) 1986.

[9] G. Fox and S. Wolfram, Phys. Rev. Lett. **41**, 1581 (1978).

[10] T. Altomari, Phys. Rev. **D37**, 677 (1988); A. Ali, T. Ohl, and T. Mannel, Phys. Lett. **B298**, 195 (1993); D. Atwood and A. Soni, SLAC-PUB-6425 (1994), submitted to Z. Phys. **C**.

[11] A. Ali and C. Greub, Phys. Lett. **B259**, 182 (1991).

[12] A. Ali and E. Pietarinen, Nucl. Phys. **B154**, 519 (1979).

[13] R.A. Poling, Ph.D. Thesis, University of Rochester (1981).

[14] G. Altarelli *et al.*, Nucl. Phys. **B208**, 365 (1982).

[15] A. Ali and C. Greub, private communication.

[16] CLEO preprint CLEO CONF 93-19.

[17] M. Voloshin and Y. Zaitsev, Sov. Phys. Usp. **30**, 553 (1987); C.T.H. Davies *et al.*, submitted to Phys. Rev. Lett.

[18] G. Crawford *et al.* (CLEO Collaboration), Nucl. Instrum. Methods Phys. Res. **A345**, 429 (1994).

[19] M. Misiak, Phys. Lett **B269**, 161 (1991); Nucl. Phys. **B393**, 23 (1993); A. J. Buras *et al.*, Nucl. Phys. **B370**, 69 (1992), **B375**, 501 (1992), **B400**, 37 (1993), and **B400**, 75 (1993); M. Ciuchini *et al.*, Phys. Lett. **B301**, 263 (1993); Nucl. Phys. **B415**, 403 (1994); G. Cella *et al.*, Phys. Lett. **B325**, 227 (1994).

[20] A. Ali, C. Greub, and T. Mannel, in Proc. ECFA Workshop on B-Meson Factory, eds. R. Aleksan and A. Ali, DESY, March 1993.

[21] A.J. Buras *et al.*, Max Planck Institute preprint MPI-Ph/93-77.

[22] F. Abe *et al.* (CDF Collaboration), Phys. Rev. Lett. **73**, 225 (1994).

[23] "Electroweak Boson Pair Production at CDF", The CDF Collaboration, Contribution to this conference GLS0357; D0 Collaboration measurement of $W\gamma X$ rates, proceedings of 1993 Moriond conference.

Table 1: Yields for the two analysis procedures.

	Neural Net	B Reconstruction
On	3013 ± 59	281 ± 17
Off(scaled)	2618 ± 73	155 ± 18
$\Upsilon(4S)$ background		
$b \to c$	50.7 ± 5.1	12 ± 2
$b \to u$	11.9 ± 4.0	2 ± 1
π^0 correction	50.2 ± 27.7	-0.7 ± 2.3
η correction	16.5 ± 33.7	2.0 ± 8.5
Non-$B\bar{B}$	2.3	
$\Upsilon(4S)$ total	132 ± 44	15 ± 9
On$-$Off$-\Upsilon(4S)$	263 ± 94 ± 49	110 ± 25 ± 10

Table 2: Additive Systematic Errors

	Neural Net		B reconstruction		Combined
	Events	B.R. $[10^{-4}]$	Events	B.R. $[10^{-4}]$	B.R. $[10^{-4}]$
On−Off subtraction					
ΔE_{cm} correction	± 19.7	± 0.14	± 2.5	± 0.06	$\pm$ 0.10
Lum ratio	± 6.6	$\pm$ 0.05	$\pm$ 0.4	$\pm$ 0.01	$\pm$ 0.03
SubTotal		± 0.15		± 0.06	± 0.10
$B\bar{B}$ background					
$b \to c$ Generic	$\pm$ 5.1	$\pm$ 0.036	$\pm$ 2	± 0.050	± 0.043
(MC statistics)					
$b \to u$ Generic	± 4.0	± 0.029	± 1	± 0.025	± 0.019
(MC statistics)					
Corrections:					
to π^0 generic:	± 27.7	± 0.198	± 2.3	± 0.058	± 0.128
to η generic:	± 33.7	0.240	± 8.5	± 0.214	± 0.227
π^0 Veto modeling	$\pm$ 14.1	$\pm$ 0.101	$\pm$ 1.9	$\pm$ 0.048	± 0.075
SubTotal		$\pm$ 0.330		$\pm$ 0.233	± 0.275
Total		$\pm$ 0.36		$\pm$ 0.24	± 0.29

Table 3: Model Dependence and Other Multiplicative Systematic Errors. The systematic errors for the conditional branching ratio, where they differ from those for the fully inclusive branching ratio, are shown in parentheses.

	Neural Net	B reconstruction	Combined
Model Dependence			
$\delta m_b = \pm$ 100 MeV	$\pm 14.5\%$	$\pm 5.7\%$	$\pm 10.1\%$
(for $2.2 < E_\gamma < 2.7$ GeV)	$(\pm 11.9\%)$	$(\pm 2.9\%)$	$(\pm 7.4\%)$
$\delta P_F = \pm$ 40 MeV/c	$\pm 1.2\%$	$\pm 1.4\%$	$\pm 0.1\%$
(for $2.2 < E_\gamma < 2.7$ GeV)	$(\pm 2.9\%)$	$(\pm 0.3\%)$	$(\pm 1.6\%)$
hadronization uncertainty	$\pm 1.0\%$	$\pm 9.0\%$	$\pm 5.0\%$
$\delta(f_{+-}/f_{00}) = \pm 0.2$	$\pm 1.0\%$	$\pm 4.9\%$	$\pm 3\%$
Subtotal	$\pm 14.6\%$	$\pm 11.8\%$	$\pm 11.7\%$
(for $2.2 < E_\gamma < 2.7$ GeV)	$(\pm 12.3\%)$	$(\pm 10.7\%)$	$(\pm 9.6\%)$
Other Multiplicative	$\pm 7.7\%$	$\pm 9.5\%$	$\pm 7.0\%$
Total	$\pm 16.5\%$	$\pm 15.1\%$	$\pm 13.6\%$
(for $2.2 < E_\gamma < 2.7$ GeV)	$(\pm 14.5\%)$	$(\pm 14.3\%)$	$(\pm 11.8\%)$

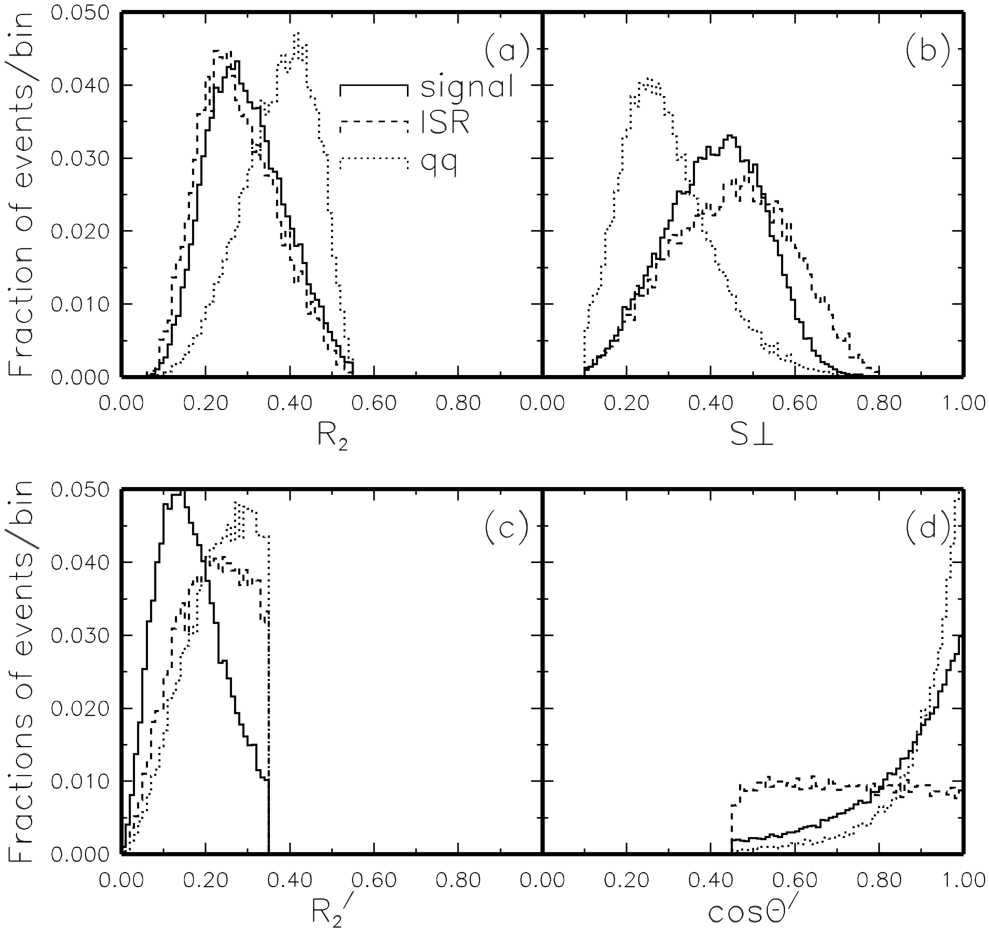

FIG. 1. $\mathrm{R}_2, \mathrm{S}_\perp, \mathrm{R}_2', \mathrm{Cos}(\theta')$ distributions for simulations of signal (solid), ISR (dashed) and $q\overline{q}$ (dotted). Some loose cuts, to eliminate obvious continuum events, have already been applied.

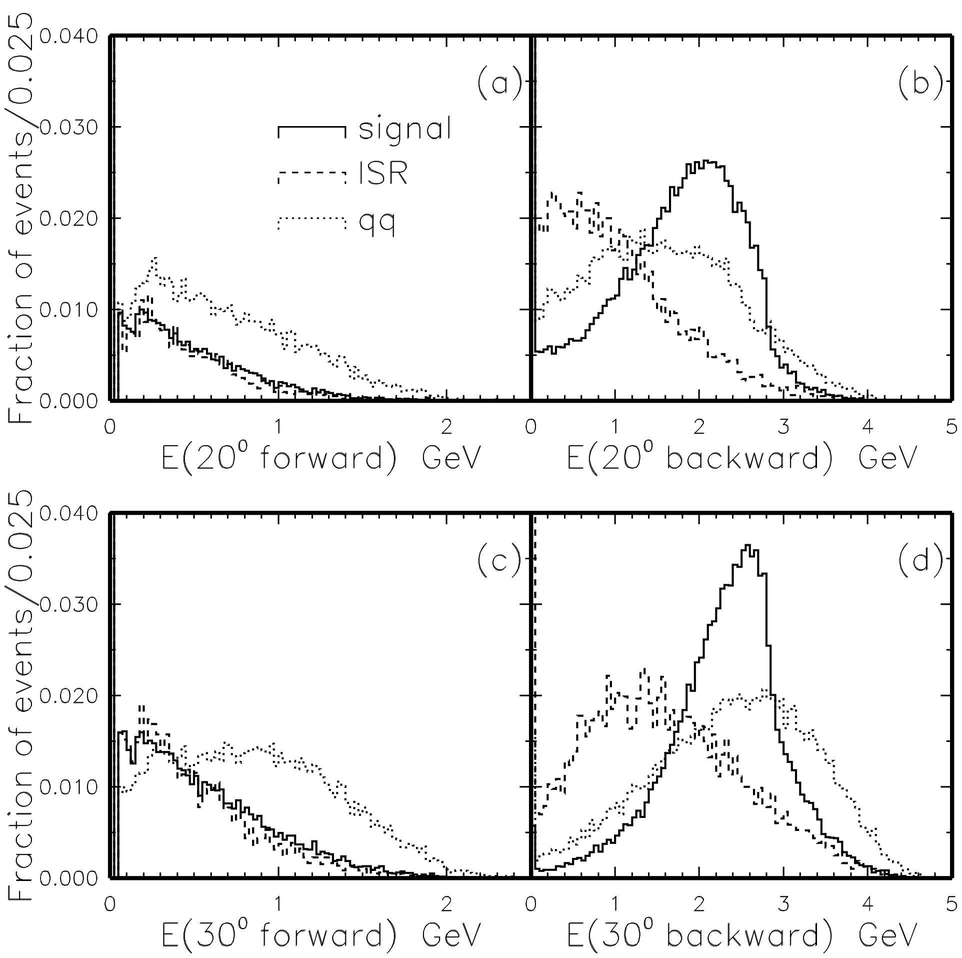

FIG. 2. Distributions of energy cone variables for simulations of signal (solid), ISR (dashed) and $q\overline{q}$ (dotted). Some loose cuts, to eliminate obvious continuum events, have already been applied.

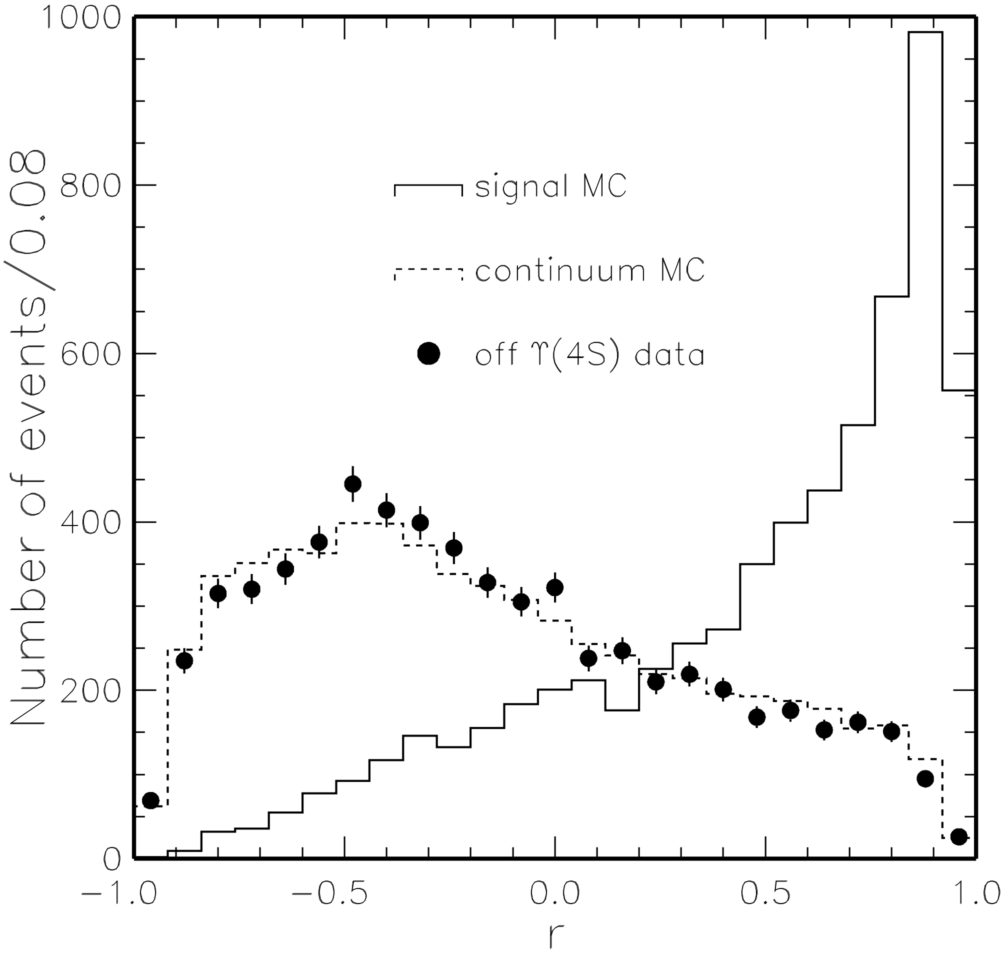

FIG. 3. r distribution for simulations of signal (solid) and background (dashed). r distribution for OFF resonance data (points).

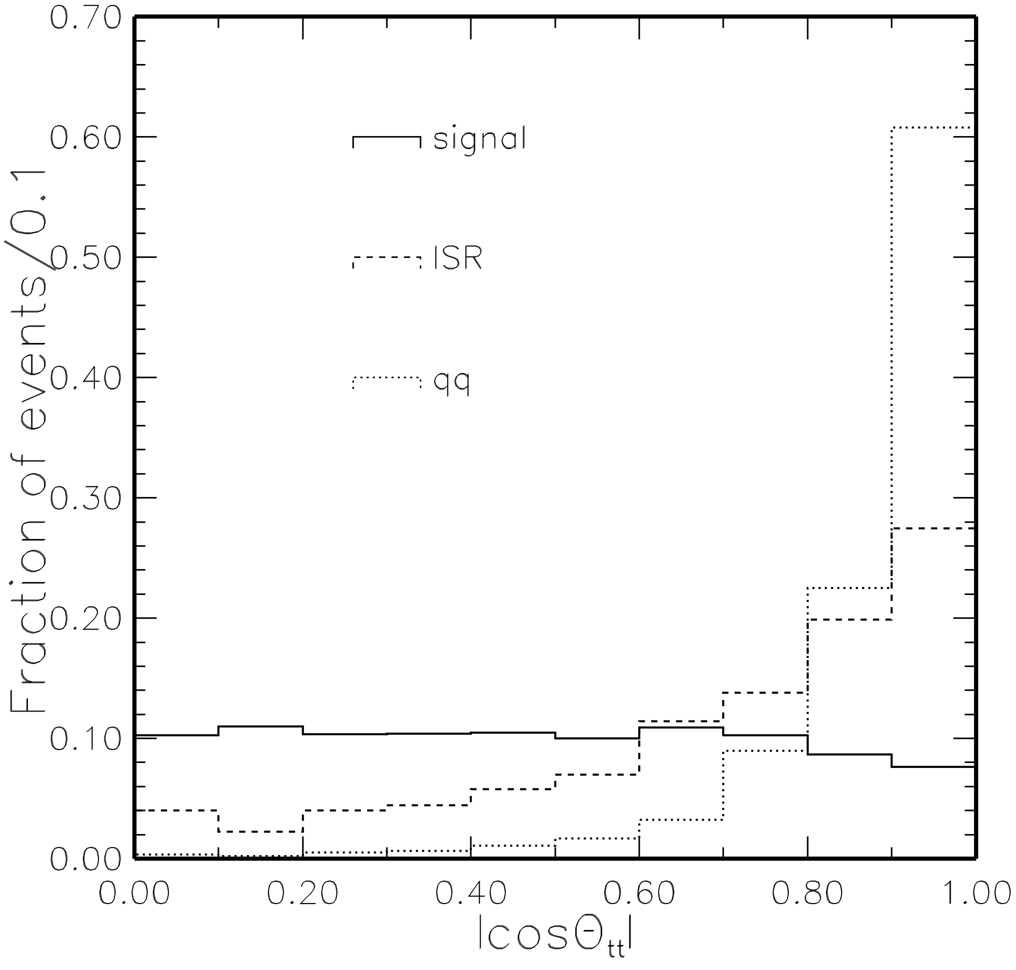

FIG. 4. $|\cos(\theta_{tt})|$ distribution for simulations of signal (solid), q$\overline{\text{q}}$ (dotted) and initial state radiation (dashed).

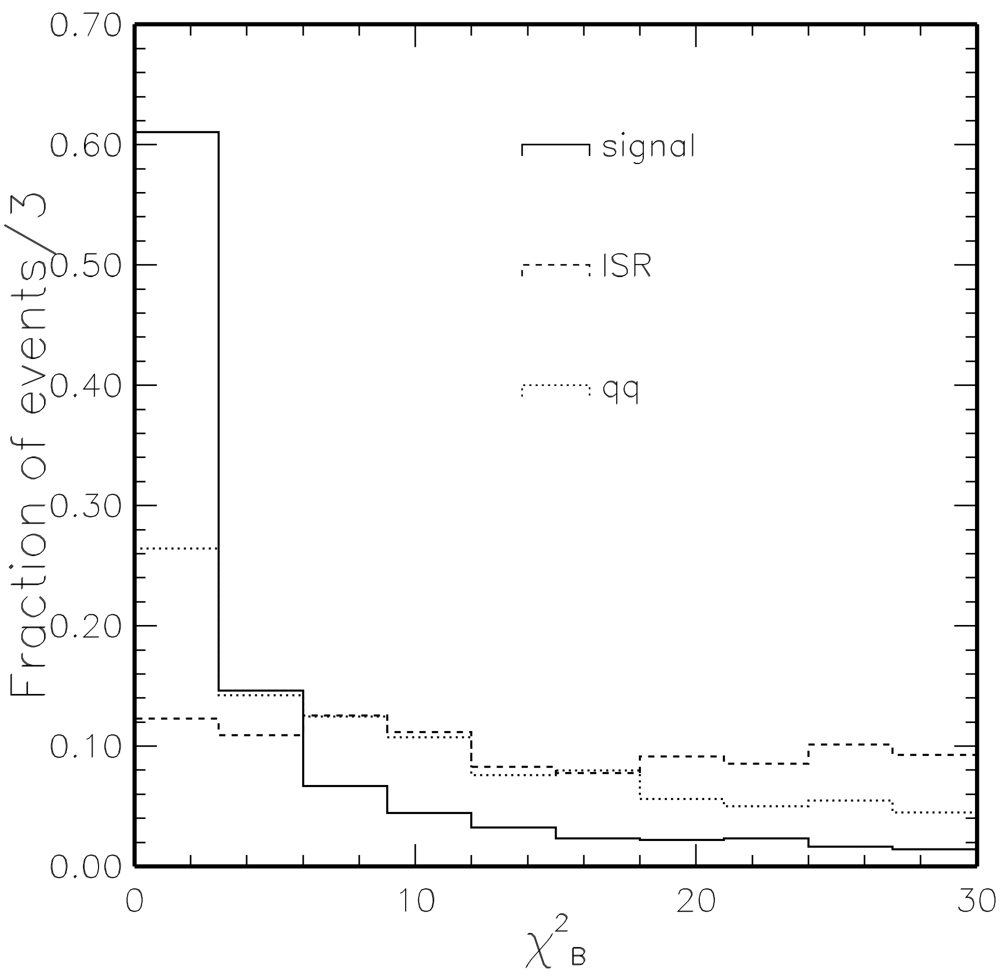

FIG. 5. $\chi^2{}_B$ distribution for simulations of signal (solid), $q\overline{q}$ (dotted) and initial state radiation (dashed).

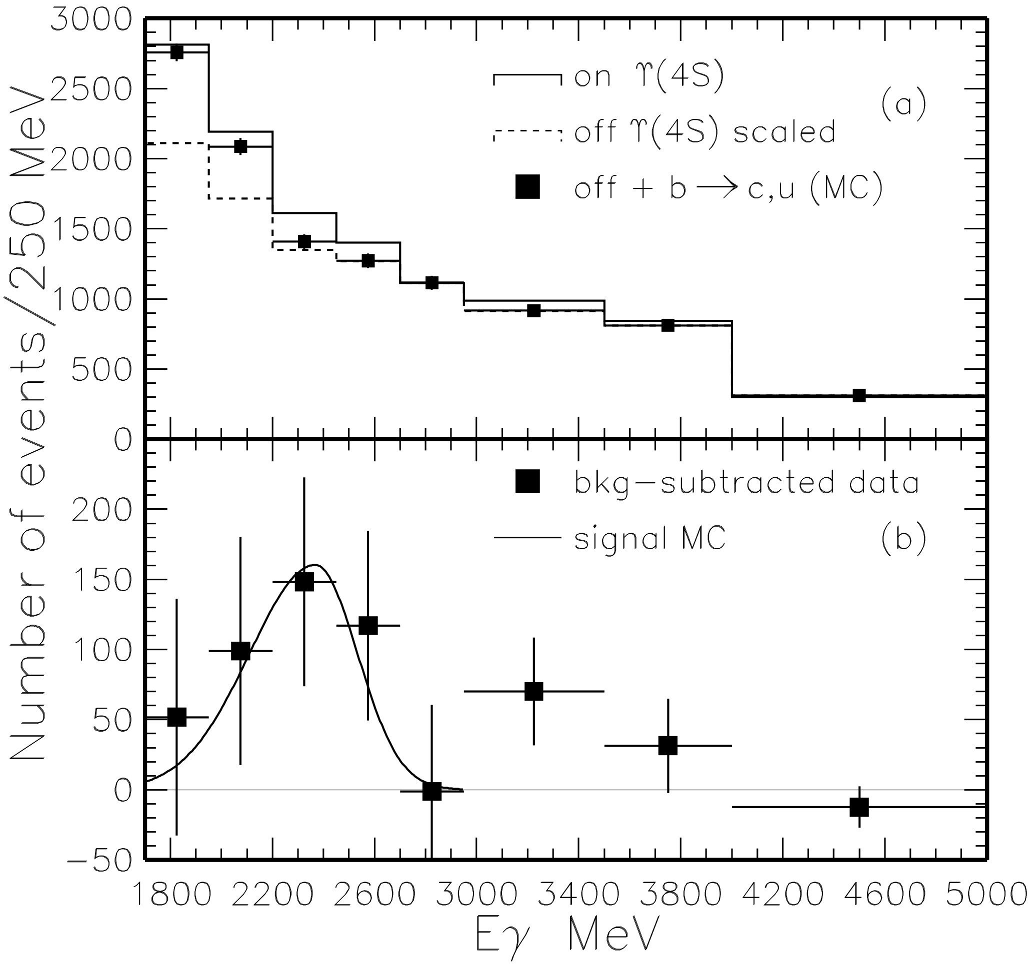

FIG. 6. Photon energy spectra from the neural net analysis.

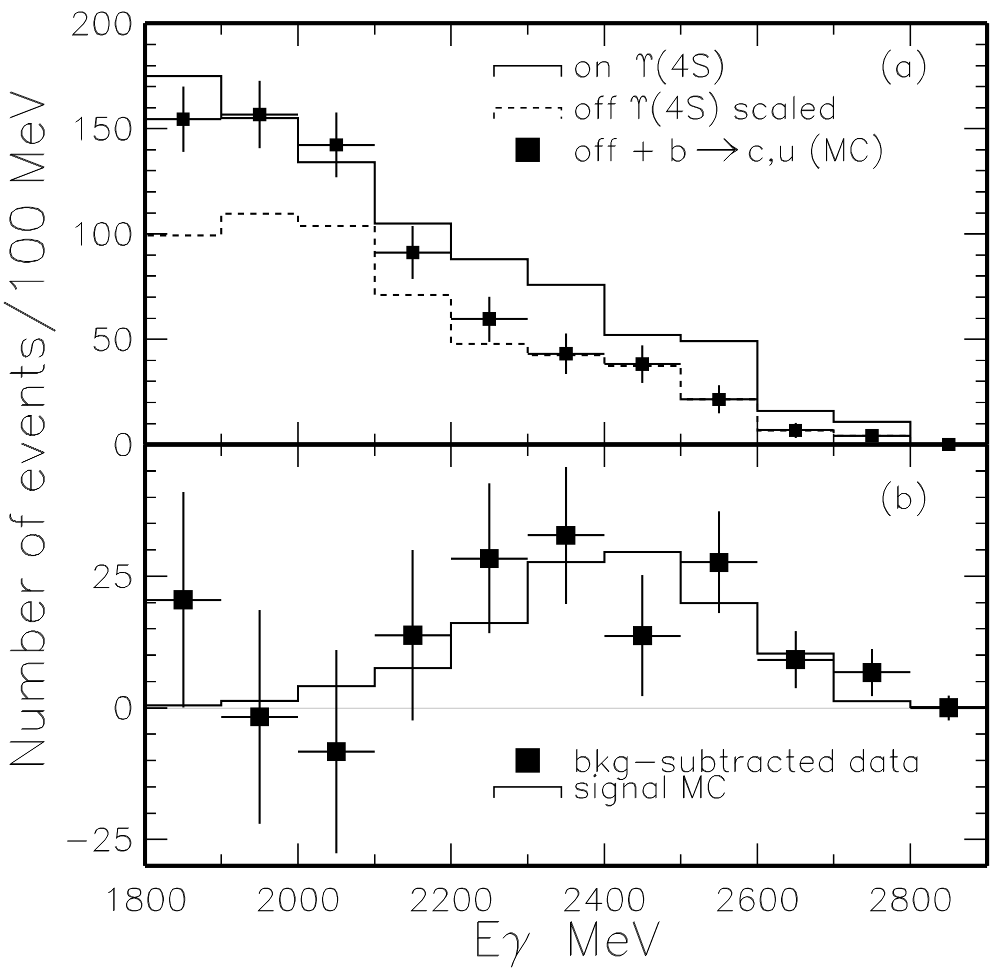

FIG. 7. Photon energy spectra from the B reconstruction analysis.

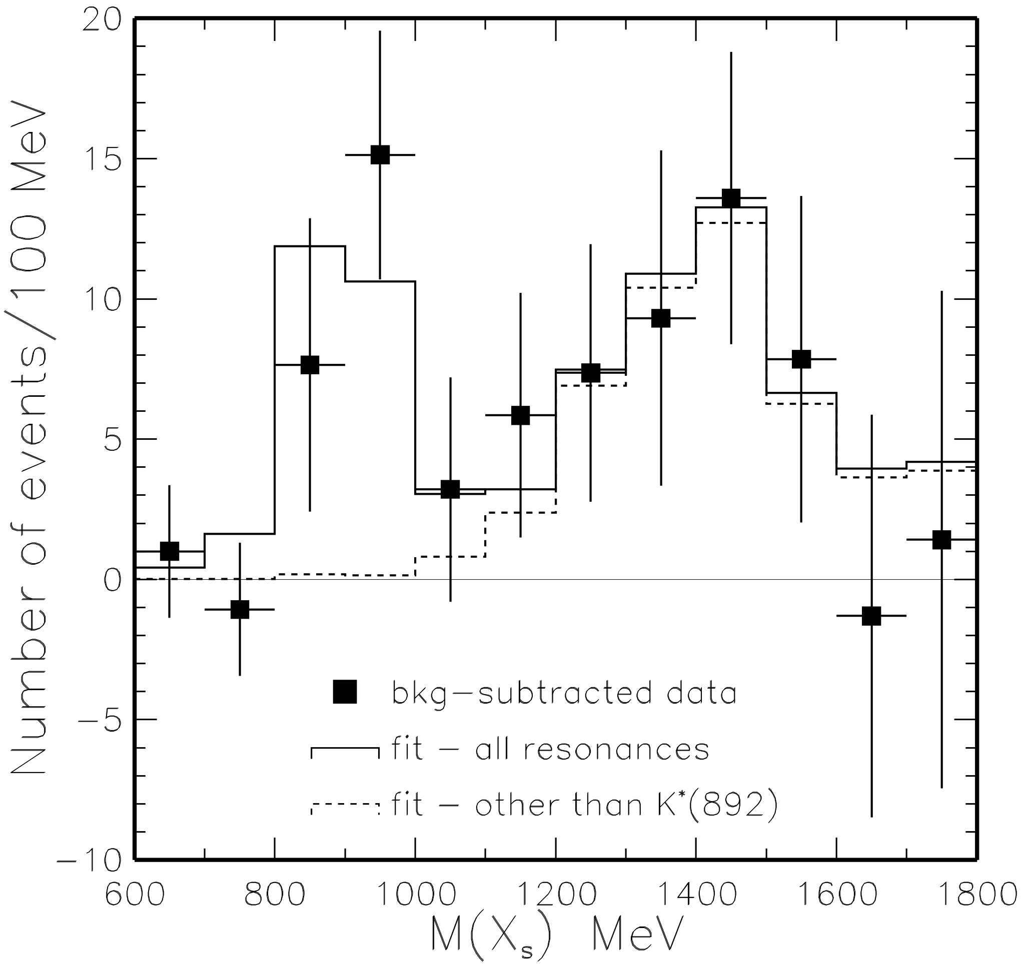

FIG. 8. Apparent X_s mass distribution from B reconstruction analysis.

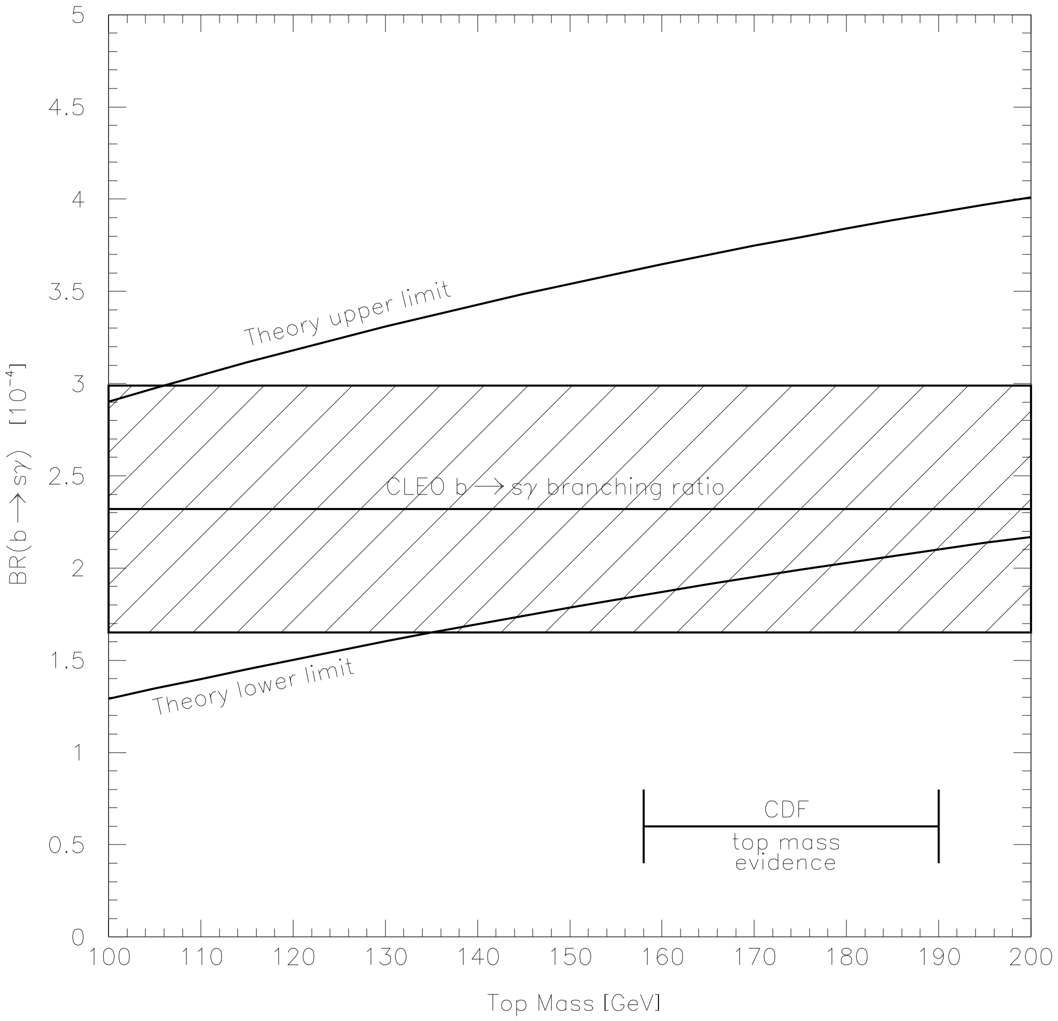

FIG. 9. Branching ratio for $b \to s\gamma$. The shaded region shows the one standard deviation range from this measurement. (The statistical and systematic errors have been combined in quadrature.) The range predicted by the Standard Model as a function of top quark mass is indicated by the curves. The range ($1\ \sigma$) of top mass according to the CDF evidence is also shown.

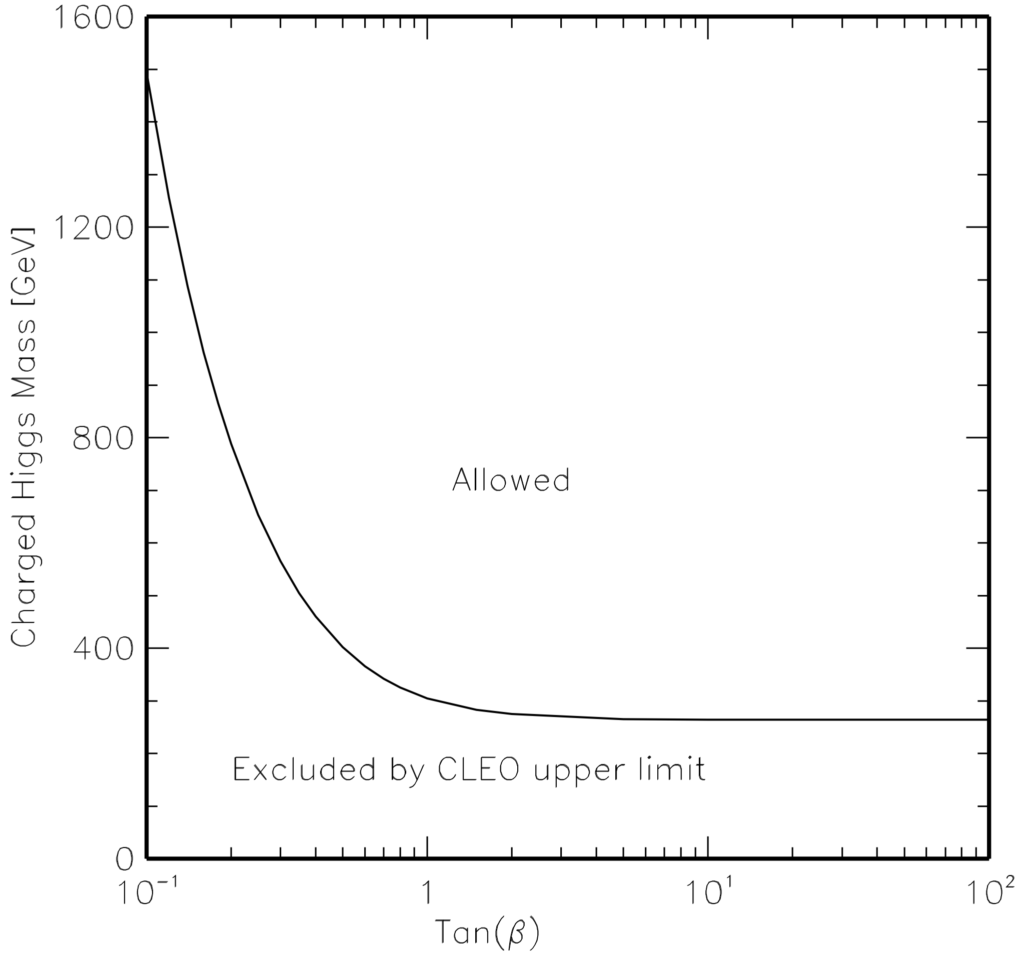

FIG. 10. Upper limit on Two-Higgs-Doublet-Model-II Higgs mass as a function of tan(β). The region below the curve is ruled out by the upper limit on the $b \to s\gamma$ branching ratio reported here.

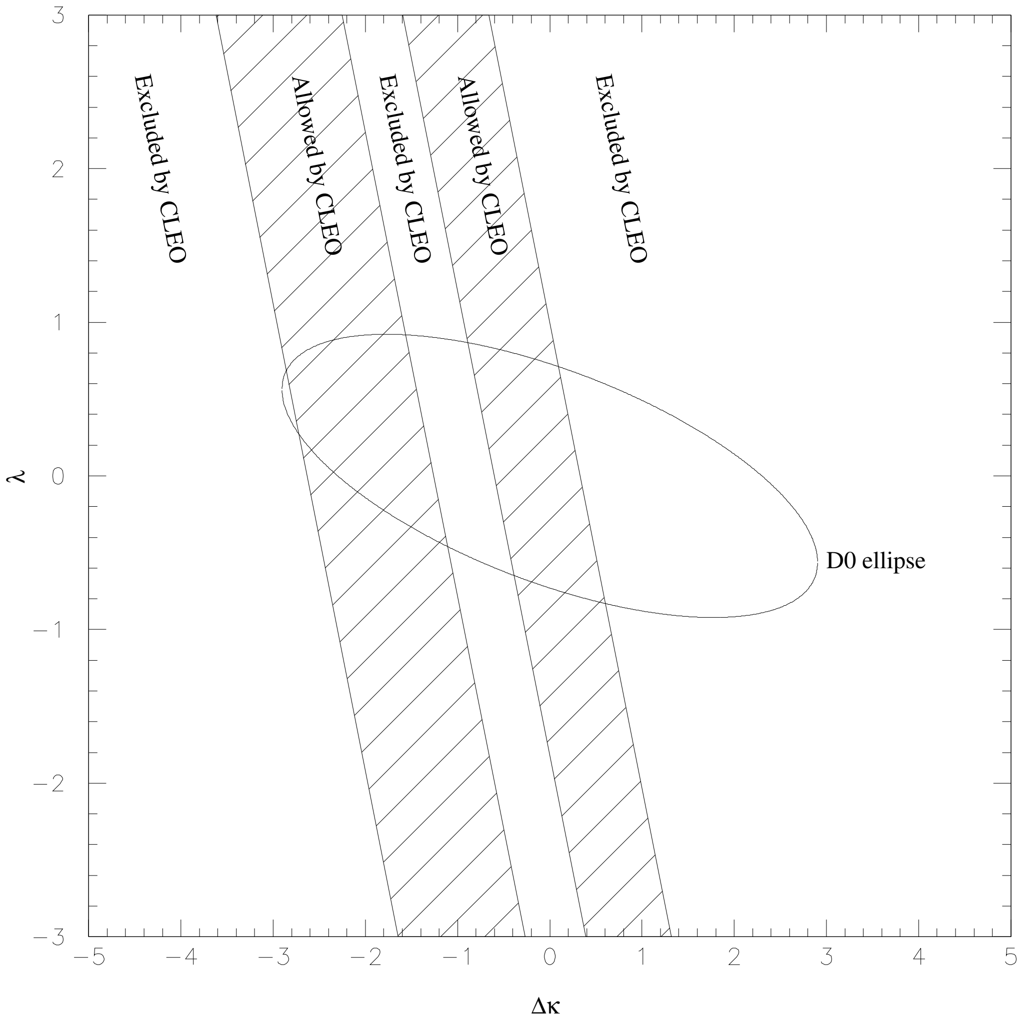

FIG. 11. Limits on anomalous $WW\gamma$ coupling parameters λ and $\Delta\kappa$. The shaded regions are consistent with the $b \to s\gamma$ branching ratio reported here. The space between the shaded regions is ruled out by CLEO's lower limit, while the spaces outside the shaded regions are ruled out by CLEO's upper limit. D0's yield of $p\overline{p} \to W\gamma X$ limits the allowed range to the interior of the ellipse (the CDF $p\overline{p} \to W\gamma X$ limit gives a similar ellipse).