Dark Photon Search at *BABAR*

Ross N. Greenwood

Office of Science, Science Undergraduate Laboratory Internship (SULI)

Massachusetts Institute of Technology

Stanford National Accelerator Laboratory

Stanford, CA

August 12, 2011

Prepared in partial fulfillment of the requirements of the Office of Science, Department of Energy's Science Undergraduate Laboratory Internship under the direction of Arafat Gabareen Mokhtar at the Stanford National Accelerator Laboratory.

Participant:

_____
Signature

Research Advisor:

_____
Signature

TABLE OF CONTENTS

ABSTRACT

Dark Photon Search at *BABAR*. ROSS N. GREENWOOD (Massachusetts Institute of Technology, Cambridge, MA 02139) ARAFAT GABAREEN MOKHTAR (Stanford National Accelerator Laboratory, Stanford, CA 94025)

Presented is the current progress of a search for the signature of a dark photon or new particle using the BaBar data set. We search for the processes $e^+e^- \rightarrow \gamma_{ISR}A', A' \rightarrow e^+e^-$ and $e^+e^- \rightarrow \gamma_{ISR}\gamma, \gamma \rightarrow A', A' \rightarrow e^+e^-$, where γ_{ISR} is an initial state radiated photon of energy $E_\gamma >= 1$ GeV. Twenty-five sets of Monte Carlo, simulating e^+e^- collisions at an energy of 10.58 GeV, were produced with different values of the A' mass ranging from 100 MeV to 9.5 GeV. The mass resolution is calculated based on Monte Carlo simulations. We implement ROOT's Toolkit for Multivariate Analysis (TMVA), a machine learning tool that allows us to evaluate the signal character of events based on many of discriminating variables. TMVA training is conducted with samples of Monte Carlo as signal and a small portion of Run 6 as background. The multivariate analysis produces additional cuts to separate signal and background. The signal efficiency and sensitivity are calculated. The analysis will move forward to fit the background and scan the residuals for the narrow resonance peak of a new particle.

INTRODUCTION

Based on observational evidence from astronomical surveys of the internal motions of distant galaxy clusters, it is believed that electromagnetically interacting matter contributes to only about 4% of the mass-energy of the universe. Current theory suggests that the remainder is made up of two additional components in a ratio of approximately three to one, respectively: dark energy, which is responsible for accelerating the metric expansion of space, and dark matter - nonbaryonic matter particles with nonzero mass (inferred from its gravitational effects on nearby visible matter) that may experience the weak or strong nuclear forces, or yet undiscovered interactions.

One candidate constituent of dark matter is a theoretical dark photon (symbol: A'). The dark photon is a massive, weakly interacting boson whose existence is postulated in models of supersymmetry [2]. It may serve as the force-carrying gauge boson for particle interactions unique to the dark sector, or even provide a means for interaction between dark matter and Standard Model particles. Theory suggests the possibility of dark photon production in electron-positron annihilation events. By the process $e^+e^- \rightarrow A' \rightarrow f^+f^-$, the electron-positron pair may annihilate directly to a dark photon, which would decay to a fermion pair. Alternatively, the e^+e^- pair annihilates to a virtual Standard Model photon, which produces a dark photon via an intermediate virtual fermion loop [2].

The BaBar experiment (1999 - 2008) recorded asymmetric collisions of electrons and positrons supplied by the linear accelerator at the SLAC National Accelerator Laboratory. BaBar focused on the study of CP violation by production of B mesons, hadrons containing a quark and an anti-quark of which one has bottom quark flavor (b or $\bar{b}$). B mesons can be produced during the decay of the $\Upsilon(4S)$ ($b\bar{b}$) meson. During Runs 1-6, collisions took place at center-of-mass energies corresponding to the masses of $\Upsilon(4S)$. The collision energy was tuned to the masses of the $\Upsilon(2S)$ and $\Upsilon(3S)$.

We searched the *BABAR* data set for evidence of a dark photon or a new particle, probing a mass range from 100 MeV to 9.5 GeV. Events featuring e^+e^- annihilation preceded by emission of an energetic photon ($E_\gamma >= 1\,\text{GeV}$) from either particle were selected. By requiring initial state radiation (ISR), we access a range of values for the relativistic mass m_{ee} of the e^+e^- system at the collision event. The main background process (those with the same final state as the sample) are radiative Bhabha scattering ($e^+e^- \rightarrow \gamma_{ISR}e^+e^-$) and two-photon production ($e^+e^- \rightarrow \gamma_{ISR}e^+e^-\gamma\gamma, \gamma\gamma \rightarrow e^+e^-$) where one of the e^+e^- pairs goes undetected.

If a dark photon exists and we implement the appropriate selection criteria, the particle will manifest in the data as a narrow resonance peak in the candidate event count plotted against the e^+e^- invariant mass. In this analysis, the significance is optimized as

$$\frac{S}{\sqrt{B} + \frac{3}{2}}$$

, where S is the number of signal events and B is the number of background events. We begin with a blind analysis using a small portion of the available data, composed of events with a dielectron mass corresponding to the $\Upsilon(4S)$ resonance ($10.58\,\text{MeV}/c^2$). After perfecting our optimization techniques, we will unblind the analysis to include all compatible data from Runs 1-7.

MATERIALS AND METHODS

Twenty-five Monte Carlo simulation production (SP) modes with discrete dark photon generated mass values ranging from 100 MeV to 9.5 GeV were produced. Bhabha scattering and two-photon production were also simulated. Two terms are used to describe the mass with respect to Monte Carlo simulations: the *generated mass* is a parameter submitted to the event generator as the true mass of the simulated dark photon with a resonance of zero

width; the *reconstructed mass* is the invariant mass of the e^+e^- system calculated from the kinematics of the generated event and simulated to account for detector resolution.

We began with a data set containing approximately 2.3% of Run 6, which contains events with pre-ISR m_{ee} tuned to the $\Upsilon(4S)$ mass of 10.58 GeV. We require at least two e tracks and one photon with energy greater than 1 GeV in the final state. Up to ten additional photons are allowed with energies less than 100 MeV. If multiple candidate pairs of e tracks are available, the pair with the best vertex fit (best χ^2) is selected, such that there is only one candidate e^+e^- pair per event. Our analysis is focused on dark photon events with an e^+e^- final state, though $\mu^+\mu^-$, $\pi^+\pi^-$, and $e^\pm\mu^\mp$ final states are also considered for the purpose of understanding the background.

The analysis thus far can be summarized in the following steps:

1. Resolution $\sigma(m)$

   Quantify the spread of the reconstructed dark photon mass produced by a Monte Carlo simulation.

2. Multivariate Analysis (MVA)

   ROOT's Toolkit for Multivariate Analysis (TMVA) package is a machine learning tool that allows us to evaluate the most likely character of events (signal vs. background) based on many discriminating variables [1]. Train a TMVA program to recognize signal events by comparing sample signal events generated in Monte Carlo simulations to the sample background, a small portion of Run 6.

3. Signal efficiency $\epsilon(m)$

   Determine the rate of successful reconstruction of events generated in Monte Carlo.

4. Sensitivity

   Predict the potential discovery.

3

Resolution

The reconstructed mass histogram for each Monte Carlo mode was fit to the sum of two Crystal Ball functions. The widths and means of the two functions are set equal, but the exponential tails are on opposite sides and otherwise free to vary. The histograms and fits are found in Figure 1. The width is considered as the resolution at the mean value of the mass. The twenty-five resolution values were plotted with the generated mass of each Monte Carlo mode and fit to a third order polynomial.

Multivariate Analysis

We used TMVA to characterize events based on nine variables:

[Unless otherwise stated, kinematic calculations were performed in the center-of-momentum frame of the e^+e^- system after initial state radiation.]

- $EXUnc$: Unconstrained total energy of $\gamma_{ISR}e^+e^-$ calculated in the laboratory frame

- PTX : Transverse momentum of $\gamma_{ISR}e^+e^-$

- $XvtxProb$: Vertex probability

- $e2cms$: Energy of the 2nd highest energy photon

- apl : Positive cosine angle between γ_{ISR} and the plane defined by the lepton tracks

- $mmiss2$: Squared missing mass

- $costhmiss$: Cosine polar angle of the missing momentum vector

- $thglfCM$: Angle between γ_{ISR} and the fastest e track

- $thfCM$: Polar angle of the e^- track

4

The TMVA classification methods employed in this analysis are Boosted Decision Tree (BDT), Likelihood, and Fisher. An assessment of each method's performance based on signal efficiency and background reduction, sampled in Figure 3, identified BDT as the most successful classifier. We ran a separate TMVA analysis for each Monte Carlo mode, and interpolated the results to produce a continous function of the BDT output parameter mva_{BDT} for the full range of m_{ee}.

To determine the most appropriate cut on mva_{BDT}, we used the well understood J/ψ resonance peak as a reference. We varied the mimimum mva_{BDT} value in a cut of the Run 6 background and fit the region of the J/ψ to the stum of a Gaussian and a linear function using the RooFit package in ROOT. We then calculated the signal-to-background ratio of the J/ψ peak. The data and fits are shown in Figure 4 and the results summarized in Figure 5. We found the cut $mva_{BDT} > 0.2$ to produce the best signal-to-background ratio while maintaining an acceptable level of signal retention.

Signal Efficiency

The efficiency was calculated for each Monte Carlo simulation as the ratio of the number of reconstructed events to the number of generated events. The results are plotted in Figure 7.

$$\epsilon(m_{ee}) = \frac{N_{rec}}{N_{gen}}. \tag{1}$$

Sensitivity

The sensitivity was calculated as the expected error on the branching ratio σ_{BR}, the fraction of the total e^+e^- annihilation cross section that includes dark photon production, corresponding to a 5σ discovery.

$$\sigma_{BR} = \frac{5\sqrt{\alpha \cdot N_{bkg}(m)}}{\epsilon(m) \cdot N_{ee}} \tag{2}$$

5

α is the ratio of the number of events in Runs 1-7 to the number of events in Run 6. N_{bkg} is the number of background events with reconstructed masses close to m_{ee}, within $1.5\sigma(m_{ee})$, where $\sigma(m_{ee})$ is the resolution. N_{ee} is the total number of electron-positron collision events in Runs 1-7.

DISCUSSION AND OUTLOOK

In summary, we have made significant progress in a search for a dark photon signature using the *BABAR* data set, calculating the simulated resolution, establishing cuts based on multivariate analysis, determining our signal efficiency, and predicting the sensitivity. However, the analysis is still underway, and has yet to produce conclusive results. The next step is to fit the reconstructed m_{ee} distribution. We will first fit the background, including recognized features (i.e. J/ψ) and assuming that no other signal is present. We will then iterate through the range of possible reconstructed A' masses, attempting to fit to a double sided Crystal Ball function at each mass value on top of the background. We aim to produce a measure of the likelihood that a discovery resonance peak sits atop the background at that mass value. Finally, we will unblind the analysis to include all available *BABAR* data from Runs 1-7.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Backes, et al. *Toolkit for Multivariate Analysis in ROOT, User's Guide.* European Organization for Nuclear Research (CERN). (2009).

[2] R. Essig, et al. "Probing dark forces and hidden light sectors at low-energy e^+e^- colliders." SLAC National Accelerator Laboratory. (2009).
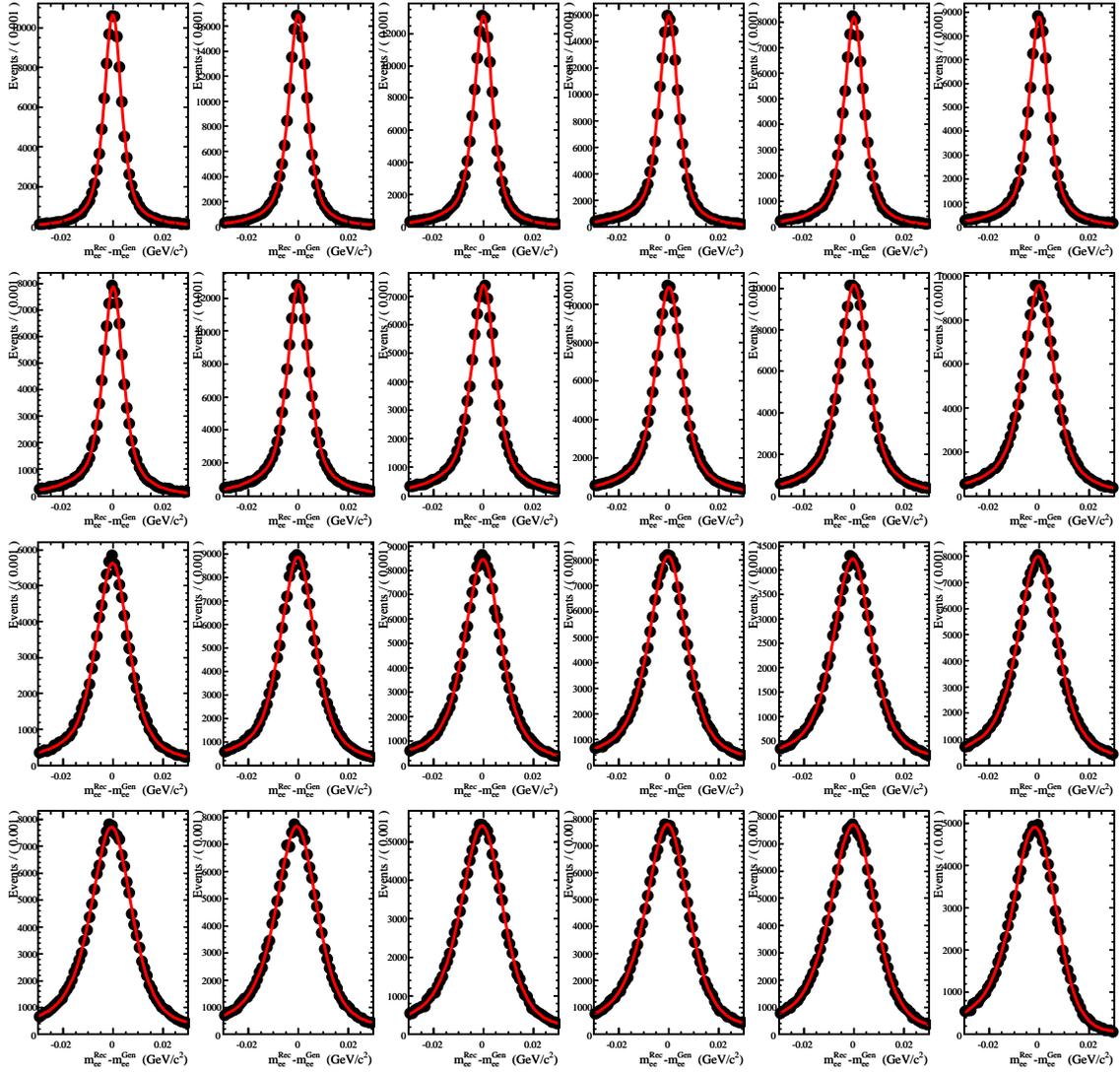
FIGURES

Figure 1: The reconstructed mass histograms and double Crystal Ball fits for 24 Monte Carlo modes.
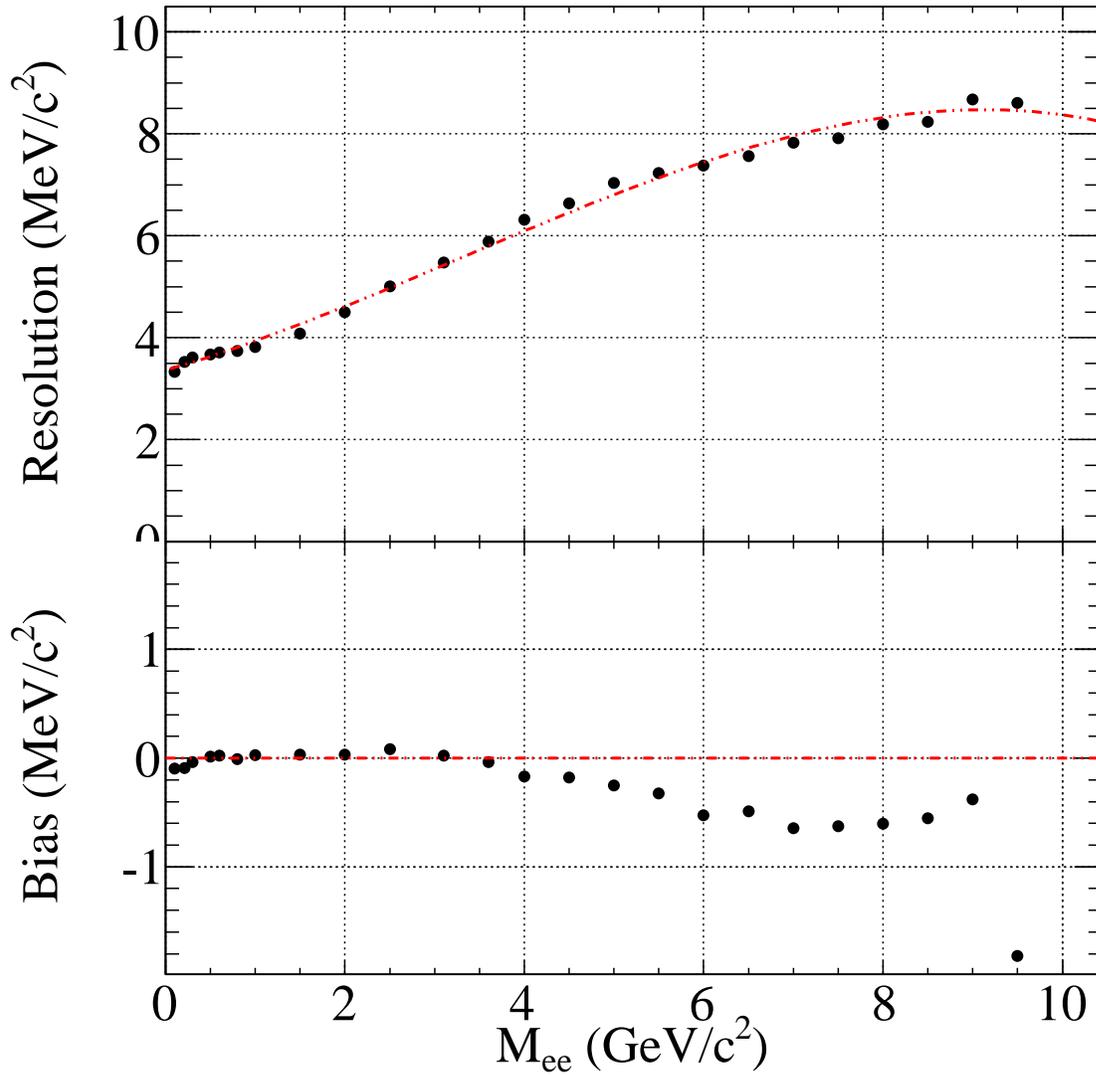
8

Figure 2: The resolution (top) and bias (displacement of the mean of the fit function) (bottom), plotted against m_{ee}.
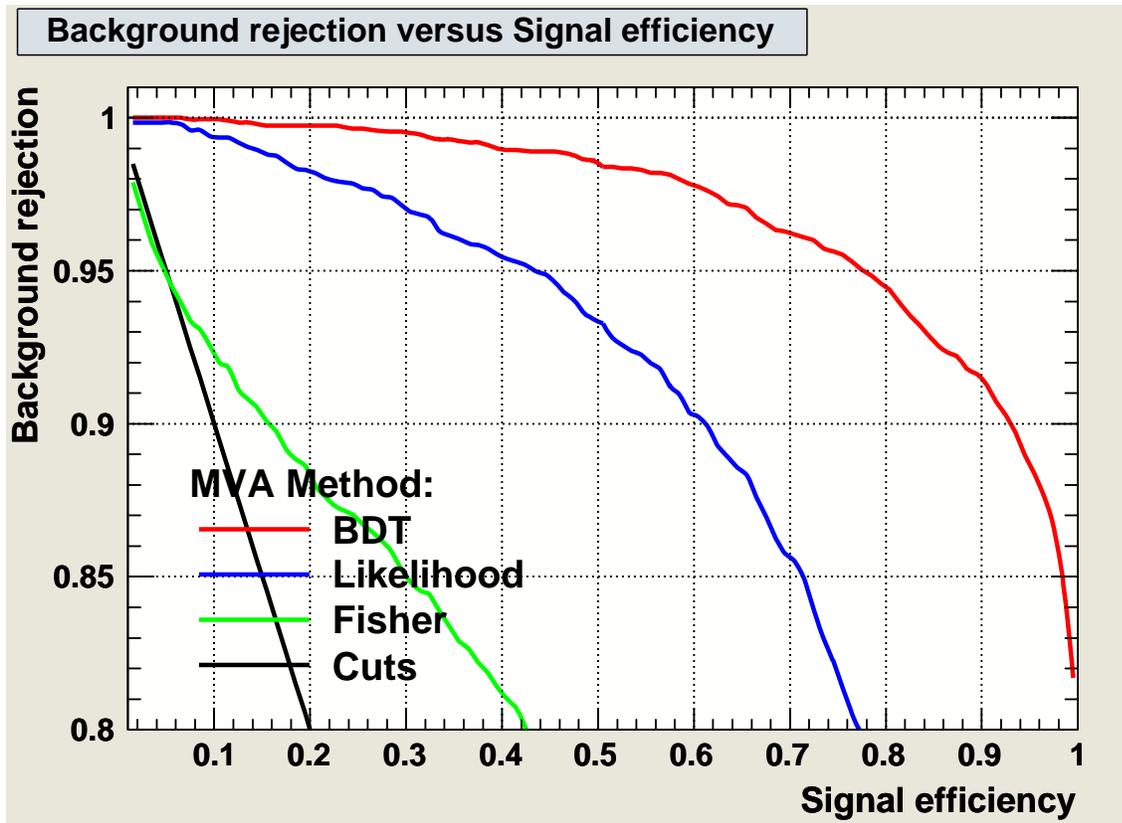
Figure 3: Background rejection plotted against signal efficiency for various MVA-based cuts.
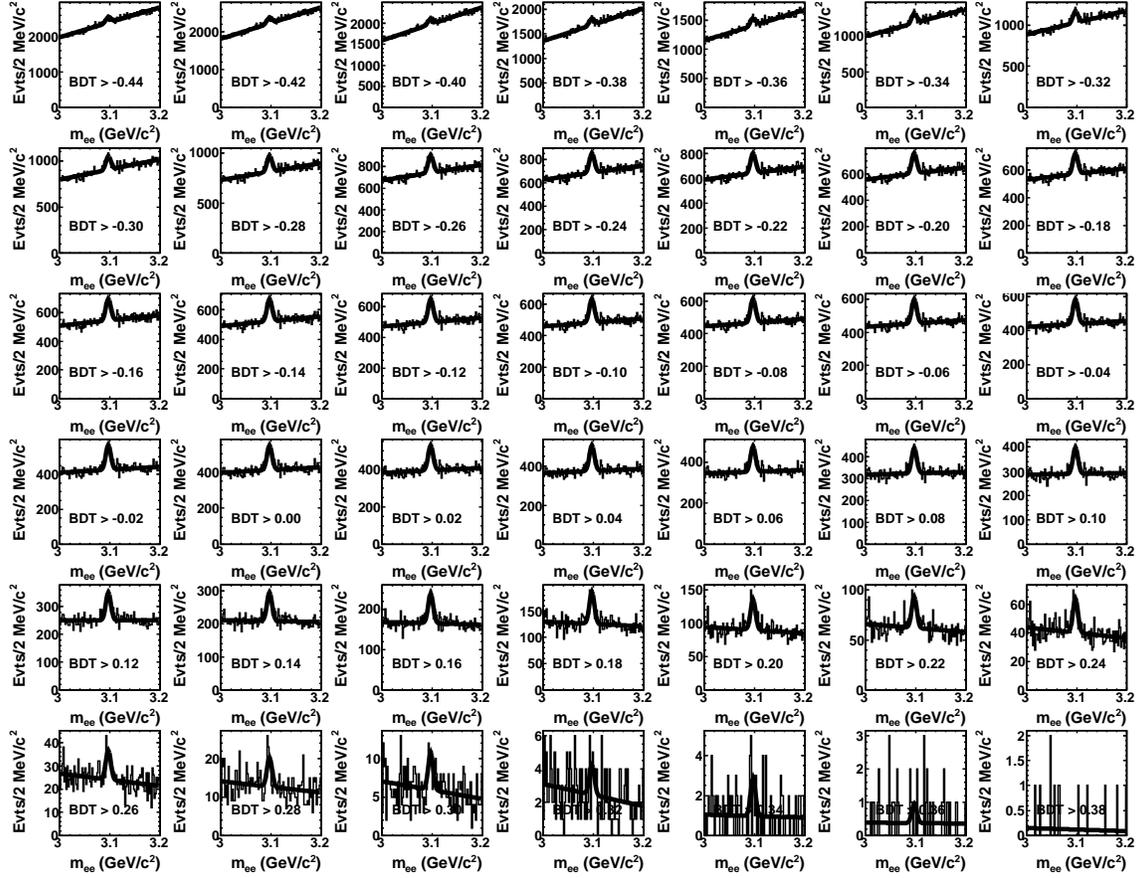
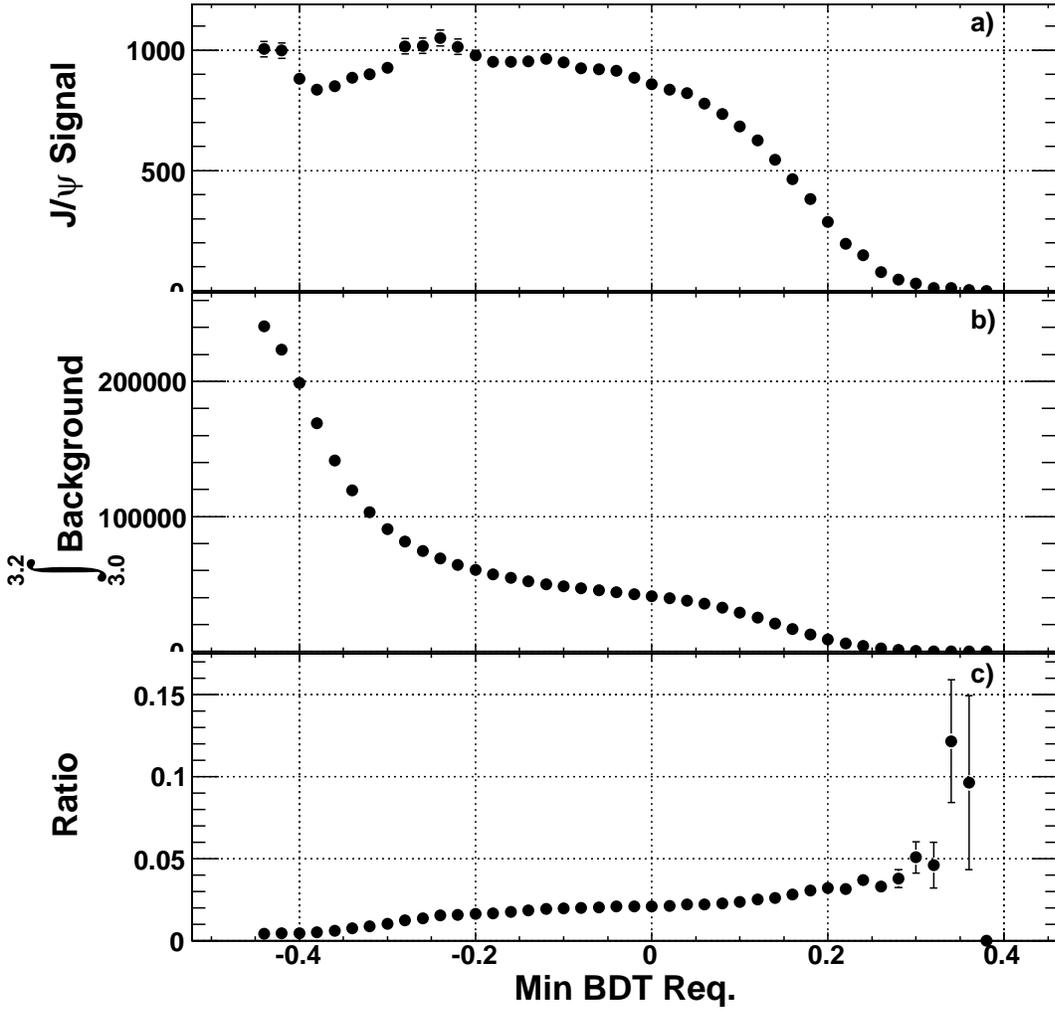Figure 4: Fits to the J/ψ resonance for different values of minimum mva_{BDT}

Figure 5: The normalizations of the J/ψ resonance (top) and background (middle), and the signal-to-background ratio (bootm) plotted against minimum required value of mva_{BDT}. At 0.2 on the x-axis, corresponding to a cut of $mva_{BDT} < 0.2$, there is a local maximum in the signal efficiency.
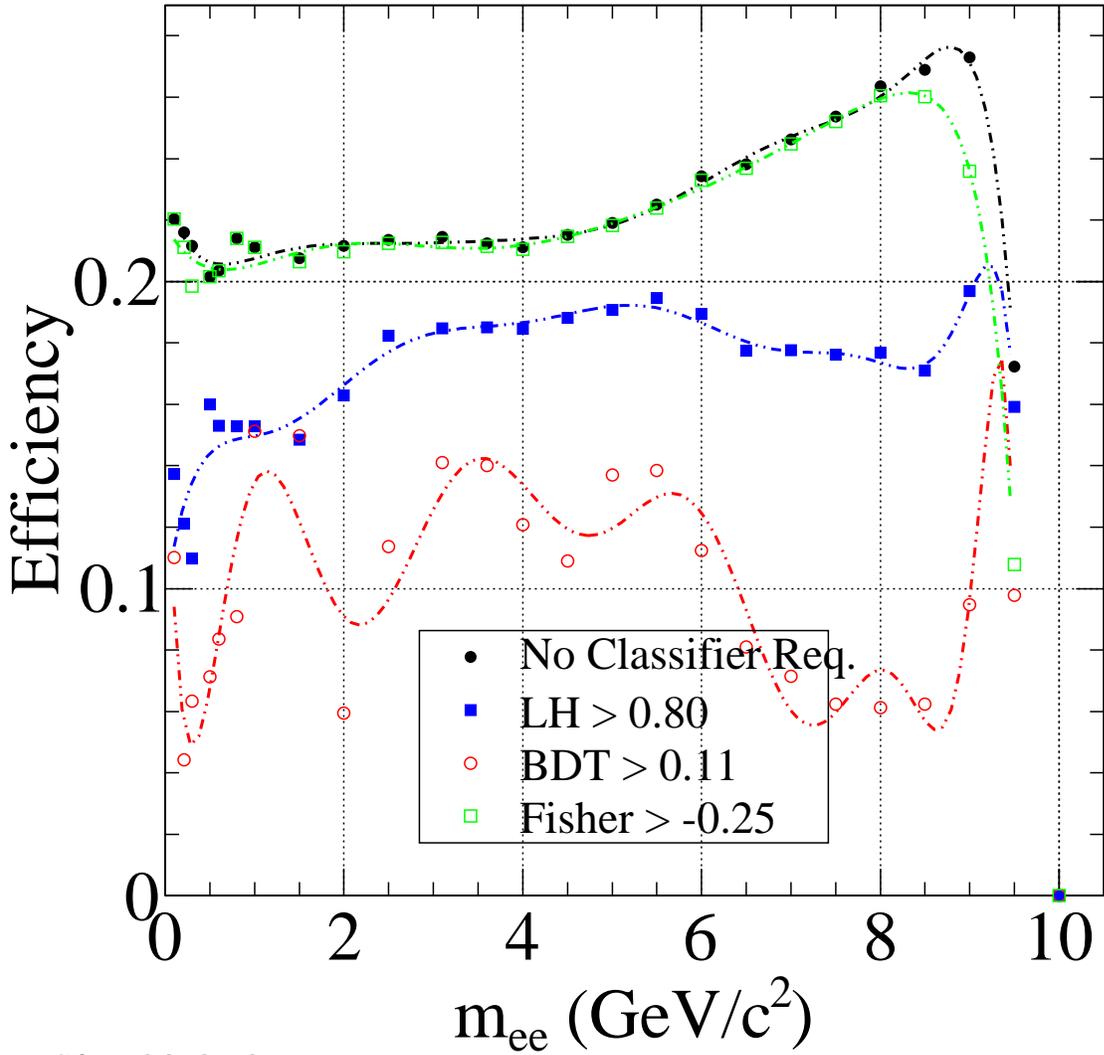
Figure 6: The signal efficiency $\epsilon(m_{ee})$ calculated from 25 Monte Carlo modes with fits to high order polynomials. Also plotted are the efficiencies calculated after applying the cuts from several different MVA methods.
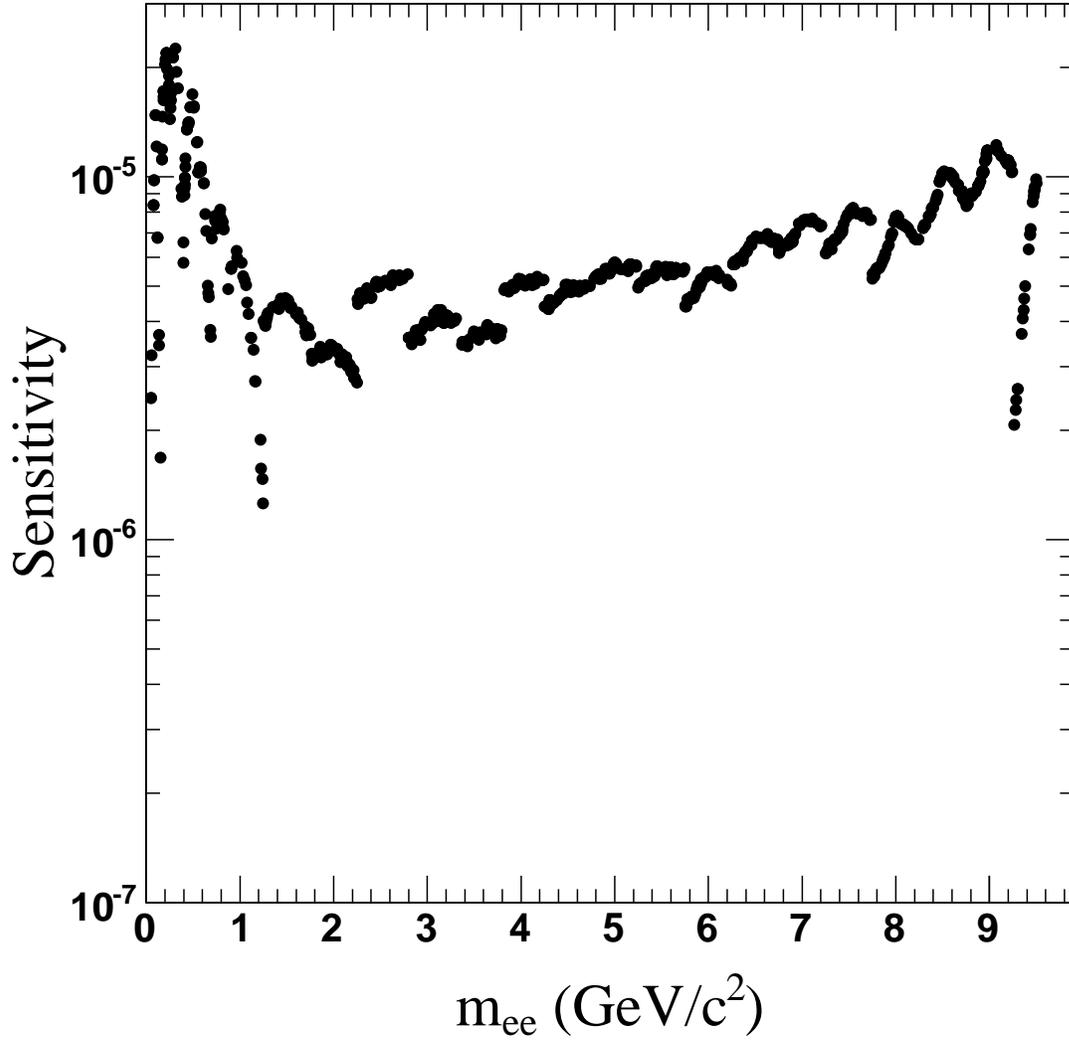
Figure 7: Sensitivity plotted against m_{ee}. Discontinuities are caused by the evaluation of the sensitivity based on parameters specific to each separate Monte Carlo mode. The sensitivity calculation at m_{ee} uses parameters from the Monte Carlo with generated mass closest to m_{ee}.

14