

ECE 539 Project  
December 20, 2003

An Artificial Neural Network Approach for  
Separating Small Signal from  
Large Background  
for the Decay Process  $B \rightarrow K^* \gamma$   
in High-energy Particle Physics

Mousumi Datta  
*The University of Wisconsin-Madison*  
*ECE 539 Final Project*  
*Dec 19, 2003*

**Abstract**

High-energy particle physics deals basically with the study of the ultimate constituent of matter and the nature of interactions between them in order to understand the history and evolution of the universe. Study of various particle decay processes are of particular interest in this field. Some of the decay

processes of interest have quite small rate (probability). The decay process  $B \rightarrow K^*\gamma$  is one such rare process which has theoretical predictions on the rate of  $\sim 7 \times 10^{-5}$ . The goal is to identify this rare signal from various other decay processes which mimics this process (background). Both high signal selection efficiency and good background rejection are crucial for identifying the signal. This report describes the application of an artificial neural network for  $B \rightarrow K^*\gamma$  signal selection and background suppression.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Data Samples</b>	<b>7</b>
2.1	Training and Validation Samples . . . . .	7
2.2	Monte Carlo (MC) Simulation for training and validation . . . . .	7
2.3	validation Sample . . . . .	7
<b>3</b>	<b>Background Suppression</b>	<b>8</b>
3.1	Background Suppression Variables . . . . .	8
3.1.1	Shape Variables . . . . .	8
3.1.2	Extra variables for ANN . . . . .	9
3.2	Cut and Count Analysis : Existing Analysis . . . . .	9
3.3	ANN Based Analysis . . . . .	9
<b>4</b>	<b>Artificial Neurel Network Software and Tools</b>	<b>10</b>
<b>5</b>	<b>Multi-Layer Perceptron (MLP)</b>	<b>11</b>
5.1	MLP Setup . . . . .	11
5.1.1	Input Nodes . . . . .	11
5.1.2	Hidden Layer Setup Selection . . . . .	11
<b>6</b>	<b>MLP Training and Testing Results</b>	<b>12</b>
<b>7</b>	<b>ANN Validation</b>	<b>15</b>
<b>8</b>	<b>Summary</b>	<b>17</b>
<b>9</b>	<b>References</b>	<b>18</b>
<b>A</b>	<b>Background Suppression Variables</b>	<b>19</b>
A.1	Background Suppression Variables . . . . .	19
A.1.1	Shape Variables . . . . .	19
A.1.2	18 Energy Cones . . . . .	19
A.1.3	$R'_2$ . . . . .	21
A.1.4	Tagside Flavor . . . . .	21

# 1 Introduction

High-energy particle physics deals basically with the study of the ultimate constituent of matter and the nature of interactions between them in order to understand the history and evolution of the universe. Experimental research in this field is carried out with giant particle accelerators and their associated detection equipments. Without going into much details about particle physics, I am going to briefly mention the basic concepts which are relevant to this project.

A major type of experiment in high-energy particle physics is the collider experiment, where beams of particles and anti-particles are accelerated at very high energy and are collided. The collision of particles and anti-particles produces high energy. In that high energy environment various particles, which used to exist in the early stage of the universe, are produced. During the early stage the universe was at very high temperatures and many particles which are not seen at present used to exist. In the present day the universe is much cooler and these particles can only be created in the particle accelerators. By producing these particle in the accelerators and studying the decay of these particles physicists try to understand the history and evolution of the universe. This report is related to one of such decay process  $B \rightarrow K^*\gamma$ .

The dataset used for this report is from *BABAR* experiment in Stanford Linear Accelerator Center (SLAC). In this experiment electron beam and its anti-particle beam are accelerated (at PEP-II accelerator) and then are collided at a point, known as the interaction point. The *BABAR* detector [[1]] is located around the interaction point and detects particles produced from the collision of the two beams.

The beam energies are controlled in such a way that from the collision a particle known as “ $\Upsilon(4s)$ ” is created almost at rest in the center of mass frame of the two beams. This  $\Upsilon(4s)$  particle then decays into two particles: “ $B$ ” and its anti-particle “ $\bar{B}$ ”, which also decay into various other particles. One of the decay process of “ $B$ ” or “ $\bar{B}$ ” is  $K^*\gamma$ . The probability of  $B \rightarrow K^*\gamma$  decay process is in the order of  $\sim 7 \times 10^{-5}$ .

The  $K^*$  produced from  $B$  or  $\bar{B}$ , decays immediately into a two particles :  $K_s^0$  and  $\pi^\pm$ . The particle  $\pi^\pm$  is a charged particle and travels through the volume of the detector before it decays.  $K_s^0$  also decays to a  $\pi^+$  and a  $\pi^-$  which then travels through the detector. The  $\gamma$  is neutral and goes through any further decay.

The charged particles,  $\pi^+$  and  $\pi^-$ , interact with the detector through ionization and trajectory of these particles are traced. The neutral particles are detected by their energy deposit in the calorimeter part of the detector. The detector records the energy and the momentum (both magnitude and direction) of all the particles coming from the interaction point so that the decay process can be reconstructed. For example, for  $B \rightarrow K^*\gamma$  decay process a  $\pi^+$  and a  $\pi^-$  particles are selected to form as  $K_s^0$  particle. Then the  $K_s^0$  is combined with another charged  $\pi$  to form a  $K^*$ . Finally the  $K^*$  is combined with a  $\gamma$  particle and the whole decay process is reconstructed. The reconstruction technique can select real signal process as well as background processes which fake the signature of the signal decay.

The signal decay process has the following properties in the center of mass frame of the two beams:

- The initial particle  $\Upsilon(4s)$  is produced almost at rest.

- The  $\Upsilon(4s)$  decays to a “B” and a “ $\bar{B}$ ”, both of which are also produced at rest because of their mass.
- Since the B and  $\bar{B}$  are at rest their decay products, are uniformly distributed. The decay topology of these processes are referred as “spherical”, shown in Figure 1.

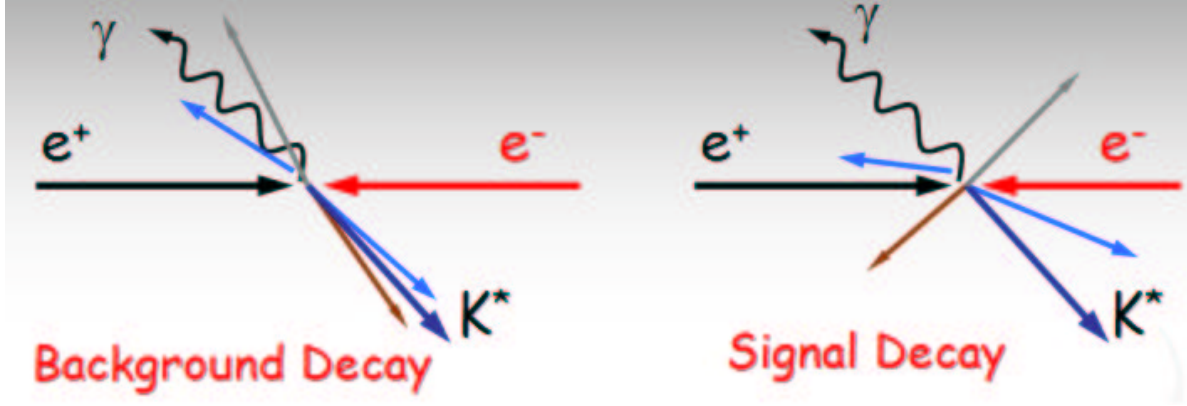


Figure 1: Event topology of signal event and background event.

From the collision of the two beams only about 10% of the time the  $\Upsilon(4s)$  particle is created. For the  $K^*\gamma$  decay a major source of background comes from the processes where the collision of the electron and positron beams don't create a  $\Upsilon(4s)$ . The background decay process has the following properties in the center of mass frame of the two beams:

- For non- $\Upsilon(4s)$  processes, instead of  $B\bar{B}$ , a different “particle anti-particle” pairs are produced.
- Unlike  $B\bar{B}$  pair, which are produced at rest, the particle anti-particle pairs have been produced with almost equal and opposite high momentum. Therefore, the decay produces two back to back jet of particles. This decay topology is referred as “jet-like”, shown in Figure 1.

The difference in the decay topology between the real signal decays and the background is exploited to reject these background decays. Some features or variables are defined to characterize the difference between the topology of the signal and background decays.

In the existing background rejection procedure, features which has good separation between signal and background are identified and used. Decays which falls in the “jet-like” topology dominated region in the feature space are rejected. This above procedure is referred as “cut and count”. The term “cut value” for a feature indicates the the decision boundary which is chosen in that feature distribution to separate “jet-like” and “spherical” decay topology. For this analysis, the cut value is chosen by maximizing a quantity called significance(S),

$$S = \frac{n_s}{\sqrt{n_s+n_b}},$$

where “ $n_s$ ” is the expected number of signal events and “ $n_b$ ” is the expected number of background events.

Several features are used for cut and count analysis. The optimization technique uses a linear regression method, which requires that the features are uncorrelated. With the current selection procedure for “cut and count” analysis the background efficiency is 13% and the signal efficiency is 66%. The background to signal ratio is high (about 1.1:1.0). It is necessary to decrease the background to signal ratio to obtain a better experimental result.

The goal of this project is to improve background rejection for the same signal efficiency obtained from “cut and count” analysis by using an artificial neural network (ANN). The reasons for replacing “cut and count” analysis with a ANN based analysis are the following.

- With the ANN features with have correlation can also be used.
- For some of the features the separation between signal and background regions may not be large enough to use them in the cut and count method. These features can still be used as inputs to ANN, which will increase both signal selection efficiency and background rejection for this process.

## 2 Data Samples

### 2.1 Training and Validation Samples

A common technique in high-energy particle physics is to use Monte Carlo (MC) simulation to study the signal selection and background rejection criteria. The selection procedure is optimized on MC simulation. Various data control samples, which do not contain any signal are used to validate the selection procedure in the MC events. At the final stage the selection are applied on the data sample of interest. The different types of data samples used for this project are outlined in this section. Total number of selected events used for ANN training, testing and validation are listed in table 1.

### 2.2 Monte Carlo (MC) Simulation for training and validation

Following samples of signal MC and background MC events are generated.

- Signal MC consists of following decay process.

$$\begin{aligned} & \Upsilon(4s) \rightarrow B\bar{B} \\ & B \rightarrow K^*\gamma \text{ and } \bar{B} \rightarrow X \\ & \text{or} \\ & \bar{B} \rightarrow K^*\gamma \text{ and } B \rightarrow X \end{aligned}$$

- Background MC consists of process where the collision of the beams do not produce  $\Upsilon(4s) \rightarrow B\bar{B}$ .

The simulated events are in reasonably good agreement with real data.

### 2.3 validation Sample

Validation sample consists of off-resonance data and a signal contron sample.

- Off-resonance Data : For off-resonance data sample the beam energy is tuned in such a way that there is not enough energy to create  $\Upsilon(4s)$ . Therefore this sample does not contain any signal and only contain the non- $\Upsilon(4s)$  processes, which are the major sources of background. Off-resonance data is used to validate the background rejection by ANN.
- Control Sample for Signal : To vadiate the signal selection efficiency by ANN, another decay process with similar decay topology as signal is used. For this purpose the following MC sample is used.

$$\begin{aligned} & B \rightarrow D\pi \text{ and } \bar{B} \rightarrow X \\ & \text{or} \\ & \bar{B} \rightarrow D\pi \text{ and } B \rightarrow X \end{aligned}$$

Table 1: List of data and MC samples used for training, testing and validation of the ANN.

Sample	Number of events
Training	
Signal MC	1167
Background MC	1167
Testing	
Signal MC	1167
Background MC	1167
Validation	
Off-resonance data	1262
Signal Control Sample	1498

The features for the decay topology are obtained by treating “D” as “ $K^*$ ” and the “ $\pi$ ” as “ $\gamma$ ”. This sample is used to validate the signal selection efficiency obtained from ANN.

### 3 Background Suppression

Background suppression technique uses the difference between the topologies of the signal and background. As mentioned earlier for signal the decay products are roughly spherically distributed, whereas the distribution of the decay products from background have a jet-like distribution.

In this section the “features”, also referred to as “variables”, which can distinguish between the two types of decay topology are listed. Appendix A contains the detailed description of all these variables and plots of the variable distributions for signal and background events. From the plots and descriptions in appendix A show that all these variable shapes have distinction between signal and background events.

#### 3.1 Background Suppression Variables

In this section the background suppression variables are listed. The details description and the distributions of background suppression variables can be found in appendix section A.1.1.

##### 3.1.1 Shape Variables

By studying the kinematics of signal and background the following variables are identified. These variables are described in detail in A.1.1. The variables distributions for signal and background are shown in figure 7.

- Magnitude of the cosine of the thrust angle  $\Theta_T$  in beam center-of-mass frame ( $|\cos\Theta_T|$ ).

- $|\cos\Theta_B|$ , magnitude of the cosine of the angle between reconstructed B and beam direction in the center-of-mass frame.
- Magnitude of the cosine of the helicity angle  $|\cos\Theta_H|$ .

These variables are uncorrelated with each other and are used in cut and count analysis.

### 3.1.2 Extra variables for ANN

- 18 Energy Cones: See A.1.2 for details.
- $R'_2$  : See A.1.3 for details.
- Tag-side Flavor: See A.1.4 for details.

Some of these above variables have some correlation with the shape variables. These variables doesn't have good separation between signal and background. Therefore these variables cannot be used in cut and count procedure.

## 3.2 Cut and Count Analysis : Existing Analysis

The “cut and count” analysis is briefly mentioned in the introduction section. Due to the limitation of this technique, only shape variables mentioned in section 3.1.1 are used. The optimized signal selection efficiency is 66% and background efficiency is 13%. Background to signal ratio using this method is about 1.1:2.

## 3.3 ANN Based Analysis

For ANN based analysis a “Feed Forward Multi-Layer Perceptron (MPL) Structure” is used. All 23 variables listed in sections 3.1.1 and 3.1.2 are used as input to the MLP. After training the ANN, it is applied on a separate testing sample, which never used for training. The ANN output value of the testing sample above which about 66% of the testing data are located, is chosen as the “decision boundary” for that ANN. The ANN is then applied on validation sample or actual data where the signal events are searched. If the feature vectors for a decay has ANN output greater than or equal to the ANN output value at the chosen “decision boundary”, that decay is categorized as signal; otherwise the decay is categorized as background.

The goal for this project is to get a lower background level compared to the cut and count analysis for the same signal efficiency.

## 4 Artificial Neural Network Software and Tools

For ANN training, validation and testing the Matlab code written by Professor Yu Hen Hu have been used. Matlab files are obtained from the course website <http://www.cae.wisc.edu/ece539/matlab/index.html> are bp.m, bpconfig.m, cvgtest.m, bpdisplay.m, bptest.m, rsample.m, randomize.m, actfun.m, actfunp.m, partunef.m, scale.m, randomize.m. The ANN code is based a “Feed Forward Model” of “Multi-Layer Perceptron (MPL) Structure”.

Additional matlab code written for this project and the modified code are submitted with the project report. The list of matlab files are the following.

1. bp.m : The matlab file bp.m is modified to store the final weights and MLP output of the training and testing sample to txt files.
2. dumpWbest.m : This code dumps the final weights of the MLP in a text file. This code is called in modified version of bp.m.
3. reconWbest.m : This code reconstructs the weight matrices from the text file created by dumpWbest.m containing all the final weights of the MLP.
4. getDecisionBoundary.m : Scan the MLP output of the testing sample. For each scanned MLP output value obtain signal and background efficiency and make plot for signal/background efficiency as a function of the location of the decision boundary.
5. plotMCTrainData.m : Plots MLP outputs for training and testing data
6. applyNNoutOfpeak.m : Apply MLP to off-peak data to validate the background rejection efficiency obtained from MLP.
7. applyNNoutDpiMC.m : Apply MLP to signal control sample to validate the Signal selection efficiency obtained from MLP.

The ASCII files containing data samples used for training, testing and validation are also submitted with the project. Following are the list of these files.

1. MCtest : Testing sample containing both signal and background MC events.
2. MCtrain : Training sample containing both signal and background MC events.
3. snns-12405-9-99-offpeak-va.pat : Off-peak data sample used for validation of background efficiency.
4. snns-12405-9-1455-DpiMC-va.pat :  $D\pi$  signal control sample used for validation of signal efficiency.

## 5 Multi-Layer Perceptron (MLP)

The goal for this project is to distinguish between two different decay topology: “spherical” vs “Jet-like”. Therefore the this problem is a “pattern classification problem”. To find the optimal setting for the MLP, various setting were experimented. This section describes the selection procedure for the MLP setting.

### 5.1 MLP Setup

Training and testing samples used for this project are listed in table 1. The training sample and testing samples were kept completely independent. The testing sample has never been used for training purposes.

All the input feature vectors in the data file are in the range between 0 and 1. No additional scaling to the input feature vectors have been applied. For hidden layers hyperbolic tangent activation function and for output layer sigmoid activation function is used. The MLP output is scaled between 0.2 and 0.8.

Learning rate is set to 0.1 and the momentum term is set to 0.8. Other settings contain “max epoch to run = 300”, “epoch size = 64” and “epoch between conv = 40”. If testing on tuning set meets no improvement for 30 training cycles, the training is terminated.

#### 5.1.1 Input Nodes

The features vectors or variables described in section 3 as used as the input to the MLP. The total number of input nodes 23.

#### 5.1.2 Hidden Layer Setup Selection

In order to find a reasonable hidden layer setup for the MLP, the performance of MLP is looked at for different hidden layer setups. For each hidden layer setup, the training and testing procedure described in section 3.3 is performed for 10 times.

As described before for each MLP training the decision boundary is chosen to obtained 66% signal efficiency on the testing sample. The MLP setting which provides the lowest average background efficiency is chosen. If two MLP setting has similar average background efficiency, then the one with relatively simpler structure is selected.

Table 2 lists the background efficiency on the testing sample for MLP setting of 23-8-4-1, 23-10-5-1 and 23-12-6-1, for 66% signal efficiency. “23-8-4-1” setting indicates that the MLP has 23 input nodes, two hidden layers with 8 and 4 nodes and 1 output layer. Same convention is used throughout this report.

Since the mean and the variance of background efficiencies for these settings are comparable, the relatively simple setting “23-8-4-1” is selected.

Table 2: lists the background efficiency on the testing sample for MLP setting of 23-8-4-1, 23-10-5-1 and 23-12-6-1, with the signal efficiency is fixed to 66%. For each setting we 10 trainings have been performed.

MLP	background	background	background
	efficiency for	efficiency for	efficiency for
structure	23-8-4-1	23-10-5-1	23-12-6-1
trial 1	0.096055	0.102058	0.099485
trial 2	0.111492	0.100343	0.098628
trial 3	0.100343	0.097770	0.098628
trial 4	0.102058	0.096913	0.096055
trial 5	0.092624	0.104631	0.091767
trial 6	0.090051	0.102916	0.094340
trial 7	0.102058	0.094340	0.096913
trial 8	0.094340	0.102916	0.100343
trial 9	0.099485	0.098628	0.104631
trial 10	0.099485	0.102916	0.102058
Average	0.0988	0.1003	0.0983
Variance	3.6417e-05	1.0950e-05	1.3925e-05

## 6 MLP Training and Testing Results

In this section the training and testing results for the final MLP is reported. Figure 2 shows the training error, which indicates that the MLP was not over trained. Figure 3 shows the MLP output for signal and background events in the training and testing sample. The MLP output distribution for signal(background) in training and testing data has good agreement.

The signal and background efficiency in the testing sample is looked at by using about 500 differnt MLP output values between 0.2 to 0.8 (MLP output range) as the decision bounday. In plot 4 shows the signal efficencies VS decision boundary point and the background efficiency VS the decision bounday point. From above scan one can choose a decision boundary to get the desired value for the signal or background efficeincy.

For this project choice of 66% signal efficieny is made. The corresponding decision boudary is the MLP output value is 0.622745 and the background efficiency in training sample is 10.4%.

On the training sample the background efficiency goes down from 13% to 10.4%. Comparing the existing cut and count based background efficiency ( $\varepsilon_{cut}$ ) and the MLP based background efficiency ( $\varepsilon_{MLP}$ ) one gets

$$\frac{\varepsilon_{MLP}}{\varepsilon_{cut}} = 0.8$$

Therefor by using MLP, the background rejection in the testing sample has been improved by around 20%. The background to signal ratio improves to 0.88:1.0 compared to the the ratio of 1.1:1.0 obtained from cut and count method.

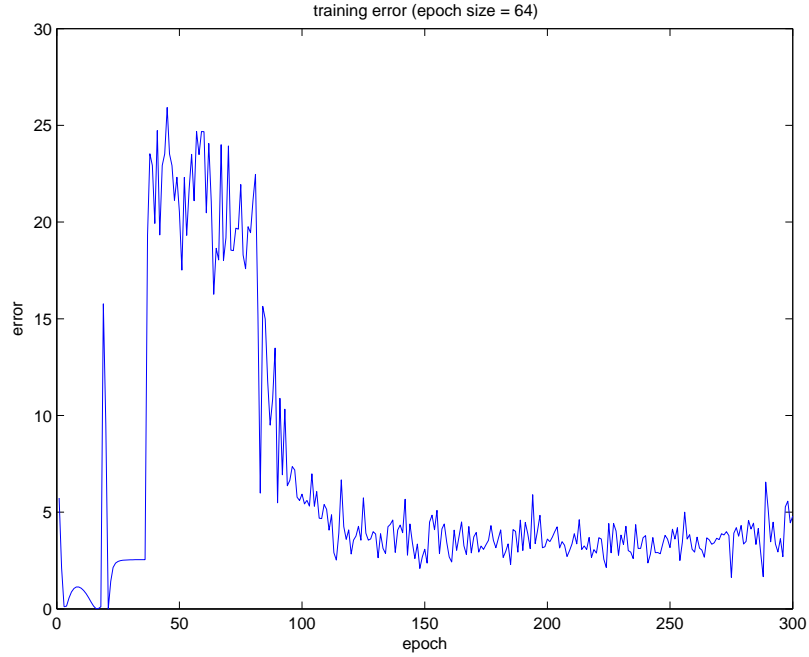


Figure 2: Training error for the final MLP.

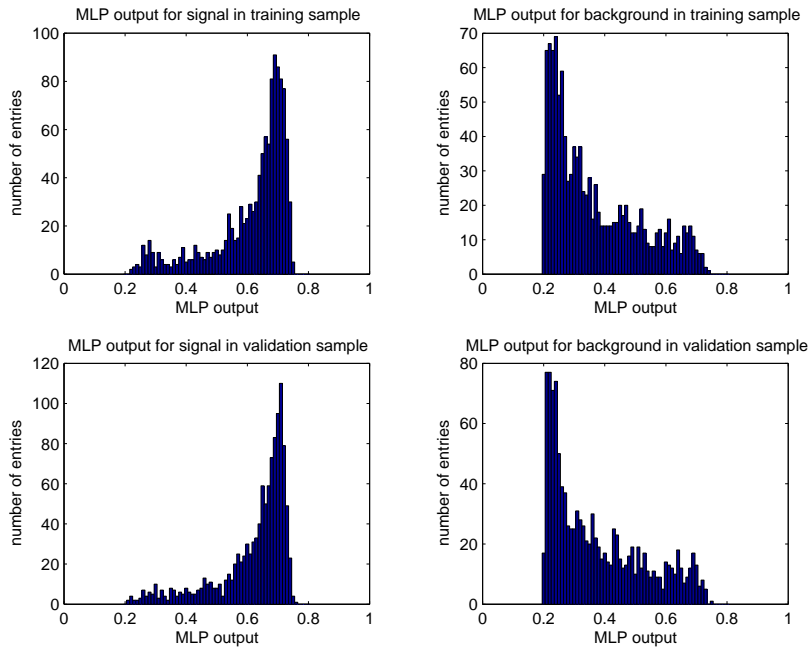


Figure 3: MLP outout for training and testing sample.

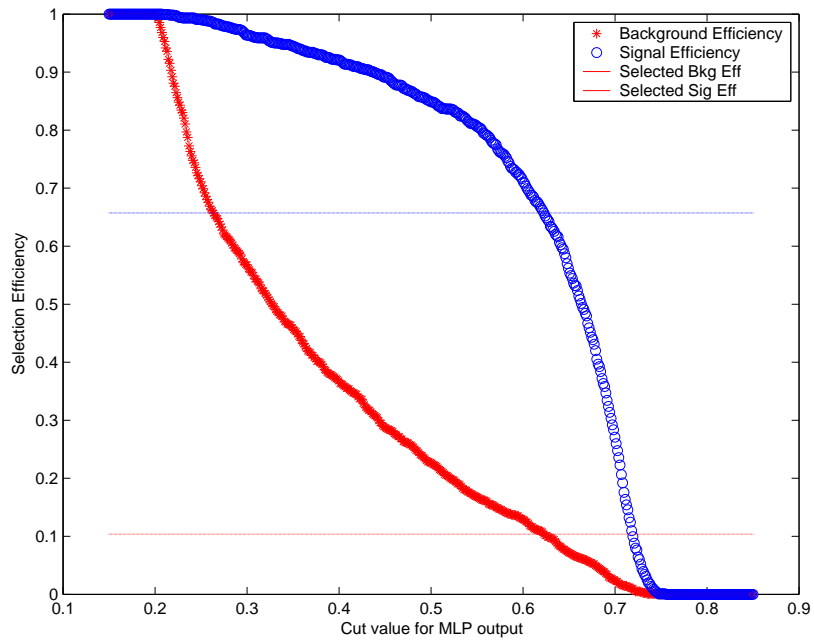


Figure 4: Signal and background efficiencies for different choices of decision boundary is plotted. The upper horizontal line indicates 66% the signal efficiency and the lower horizontal line shows the corresponding background efficiency.

## 7 ANN Validation

To validate the MLP two control samples described in section 2.3 are used.

- Off-peak data sample: This sample do not contain any signal and the MLP output is expected to be consistant with that of the background training and testing sample. Figure 5 shows the performance of the MLP on the off-peak data sample. The MLP output distribution is in

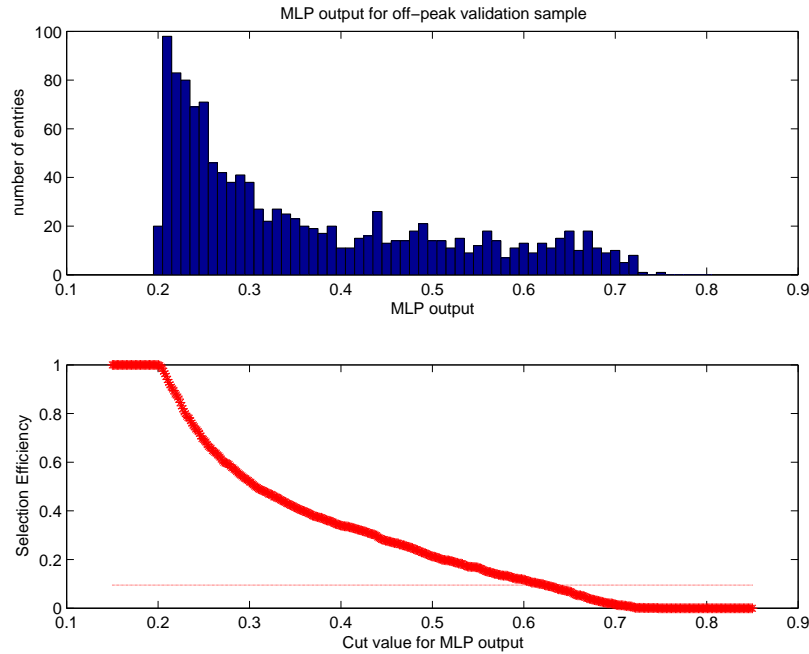


Figure 5: Results for applying the MLP on the signal control sample. The top plot shows the distribution of MLP output and the bottom plot shows the efficiencies for different decision boundary points.

agreement with that of background samples. For decision boundary of MLP output value at 0.622745, which is obtained from section 6, the background efficiency is 9.5%. This result is in agreement with the 10.4% background efficiency on the testing sample.

- $D\pi$  control sample: This sample do not contain any background and the decay topology is similar to the the signal decay. Therefore the MLP output is expected to be consistant with that of the signal training and testing sample. Figure 6 shows the performance of the MLP on the control sample. The MLP output distribution is in agreement with that of signal samples. For decision boundary of MLP output value at 0.622745, which is obtained from section 6, the signal efficiency is 66%. This result is in agreement with the 66% signal efficiency on the testing sample.

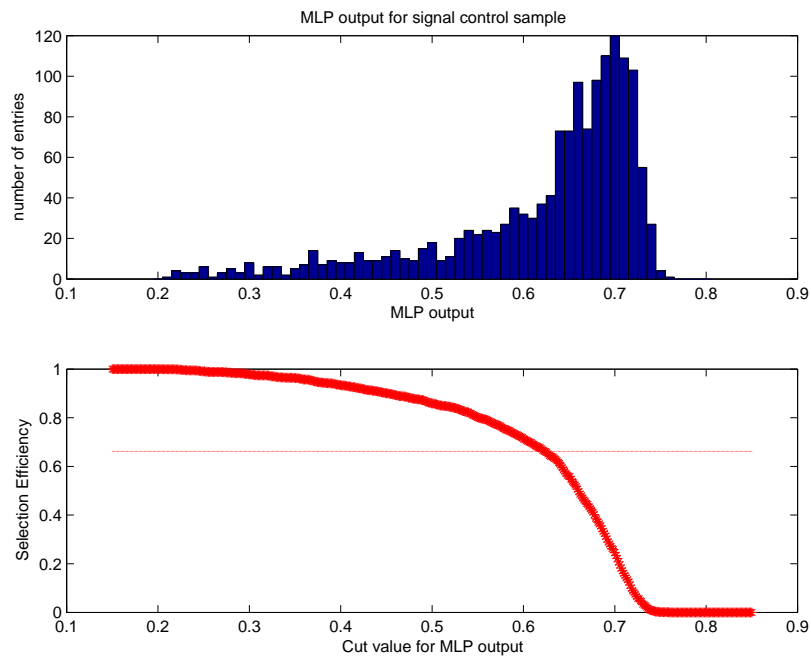


Figure 6: Results for applying the MLP on the signal control sample. The top plot shows the distribution of MLP output and the bottom plot shows the efficiencies for different decision boundary points.

## 8 Summary

Using ANN based selection the background rejection is improved by about 20% in training and validation samples compared to the existing analysis. This improved the background to signal ratio from 1.1:1.0 to 0.88:1.0, which is a significant improvement for selecting  $B \rightarrow K^* \gamma$  decay process.

The performance of the MLP is consistent between testing and validation samples, which gives us confidence that this MLP performance on the real data sample would be reliable.

## 9 References

### References

- [1] *BABAR* Collaboration, B. Aubert *et al.*, Nucl. Instr. Methods **A479**, (2002) 1.
- [2] S. Agostinelli *et al.*, “Geant4 - A Simulation Toolkit” . SLAC-PUB-9350. Submitted to Nucl. Instr. Methods **A**.

## A Background Suppression Variables

Background suppression technique uses the difference between the topologies of the signal and background. As mentioned earlier for signal the decay products are roughly spherically distributed, whereas the distribution of the decay products from background have a jet-like distribution. In this section the “features”, also referred to as “variables”, which can distinguish between the two types of decay topology are described.

### A.1 Background Suppression Variables

#### A.1.1 Shape Variables

By studying the kinematics of signal and background the following variables are identified. The variables distributions for signal and background are shown in figure 7.

- Cosine of the thrust angle  $\Theta_T$  in beam center-of-mass frame ( $\cos\Theta_T$ ): The thrust axis ( $\vec{T}$ ) for an event is defined to be the direction which maximizes the sum of the longitudinal momenta of the particles along this direction. The angle between signal photon and the thrust axis in the center-of-mass frame of beams is defined as  $\Theta_T$ . It provides good separation between signal and background.

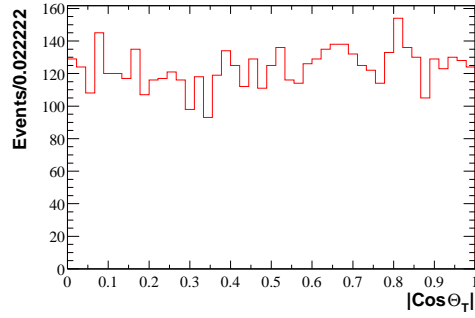
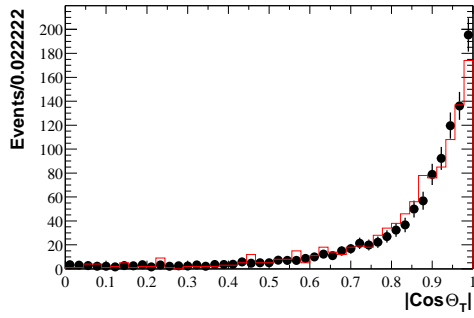
For background decays, which are jet-like, the thrust angle is close to 0 or 180 degrees. While for the signal decays, the signal photon direction is less correlated with the thrust axis and  $\cos\Theta_T$  is uniformly distributed, shown in a) of Figure 7. This variable provides most powerful distinction between signal and background decays.

- $\Theta_B$ , the angle between reconstructed B (from  $K^*\gamma$  decay) and beam direction in the center-of-mass frame. For signal decays this angle follows  $\cos\Theta_B$  follows a 2nd-Polynomial distribution. While for the background decays, the  $\cos\Theta_B$  distribution is roughly flat, shown in b) of Figure 7.
- Helicity angle  $\Theta_H$ , the angle between one daughter of  $K^*$  and the mother of  $K^*$  (the reconstructed B) in the rest frame of  $K^*$ .  $\cos\Theta_H$  is expected to follow a 2nd-Polynomial distribution for signal and be flat for background, shown in c) of Figure 7.

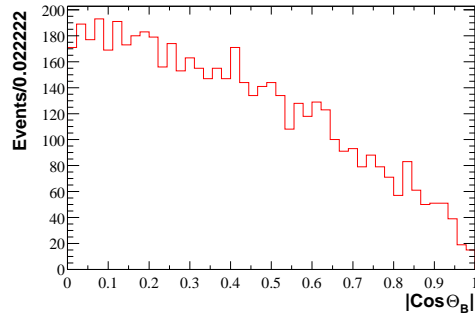
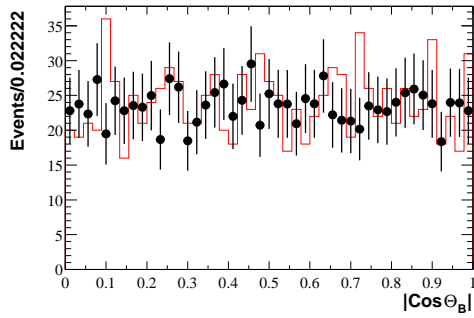
#### A.1.2 18 Energy Cones

In addition to the above shape variables, additional information on the decay topology can be obtained by considering the distribution of the decay products. This energy of final products is binned into cones of 10deg increment in the center of mass frame of the beams, around the signal photon direction, shown in Figure 8, where the labeling of the 18 energy cones are defined.

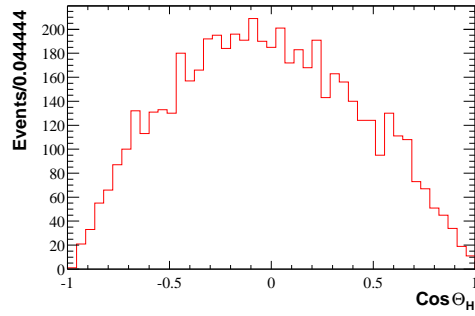
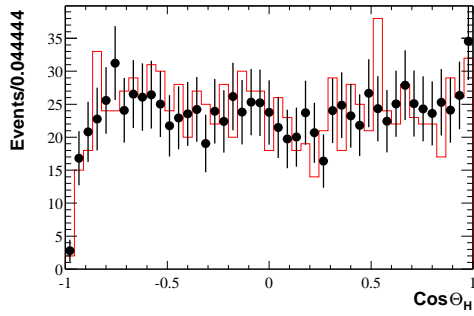
Figure 9, shows the distribution of 18 energy cones for background. The comparison of off-peak data and background Monte Carlo shows good agreement. The same distribution for Signal Monte Carlo is shown in Figure 10 also. We can see there is no significant difference between the signal and background events in any one energy cone. These variables can not be used in “cut



(a)



(b)



(c)

Figure 7: a)  $|\cos\Theta_T|$  distribution, b)  $|\cos\Theta_B|$  distribution and c)  $\cos\Theta_H$  distribution for background Monte Carlo(left) and signal Monte Carlo(right). The background Monte Carlo (red line) is compared with off-peak data (solid circle).

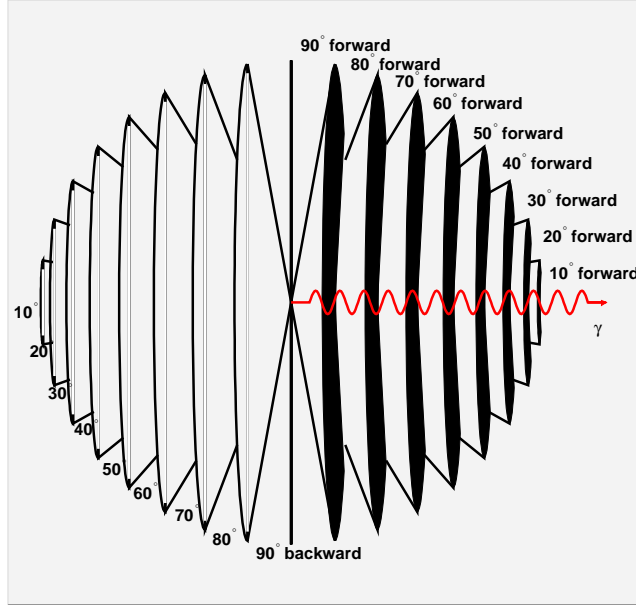


Figure 8: Energy Cones in the CMS respecting the  $\gamma$  momentum. The cone variable is obtained by summing contribution in a cone (exclusively) from charged and neutral particles with momentum pointing within the cone.

and count” analysis, but we can combine these variables into neural network to provide further discriminating power between data and Monte Carlo.

### A.1.3 $R'_2$

The Fox-Wolfram moments are defined as,

$$\sum_{i,j} [|p_i| \cdot |p_j| P_l(\cos\theta_{ij}) / E_{vis}^2],$$

where  $P_l$  are the Legendre polynomials,  $p_i$  are the particle momenta,  $\theta_{ij}$  is the opening angle between particle i and j and  $E_{vis}$  is the visible energy of all the particles in the event. The distribution for signal event and background event are shown in Figure 11. We can see the separation is not significant. This feature also has some correlations with  $\Theta_T$ . We can not use it in “cut and count analysis”.

### A.1.4 Tagside Flavor

The decays of  $\Upsilon(4S)$  events are performed through different physics interaction from non- $\Upsilon(4S)$  events. As a result, the content of the final products are also different from each other. (e.g. in  $\Upsilon(4S)$  events, 10% of decays contain electrons, while for non- $\Upsilon(4S)$  events almost there is no electron in the decay). To identify this feature, a “Tagside flavor” is defined and the distributions for signal and background are shown in Figure 11. We can see some difference are shown between

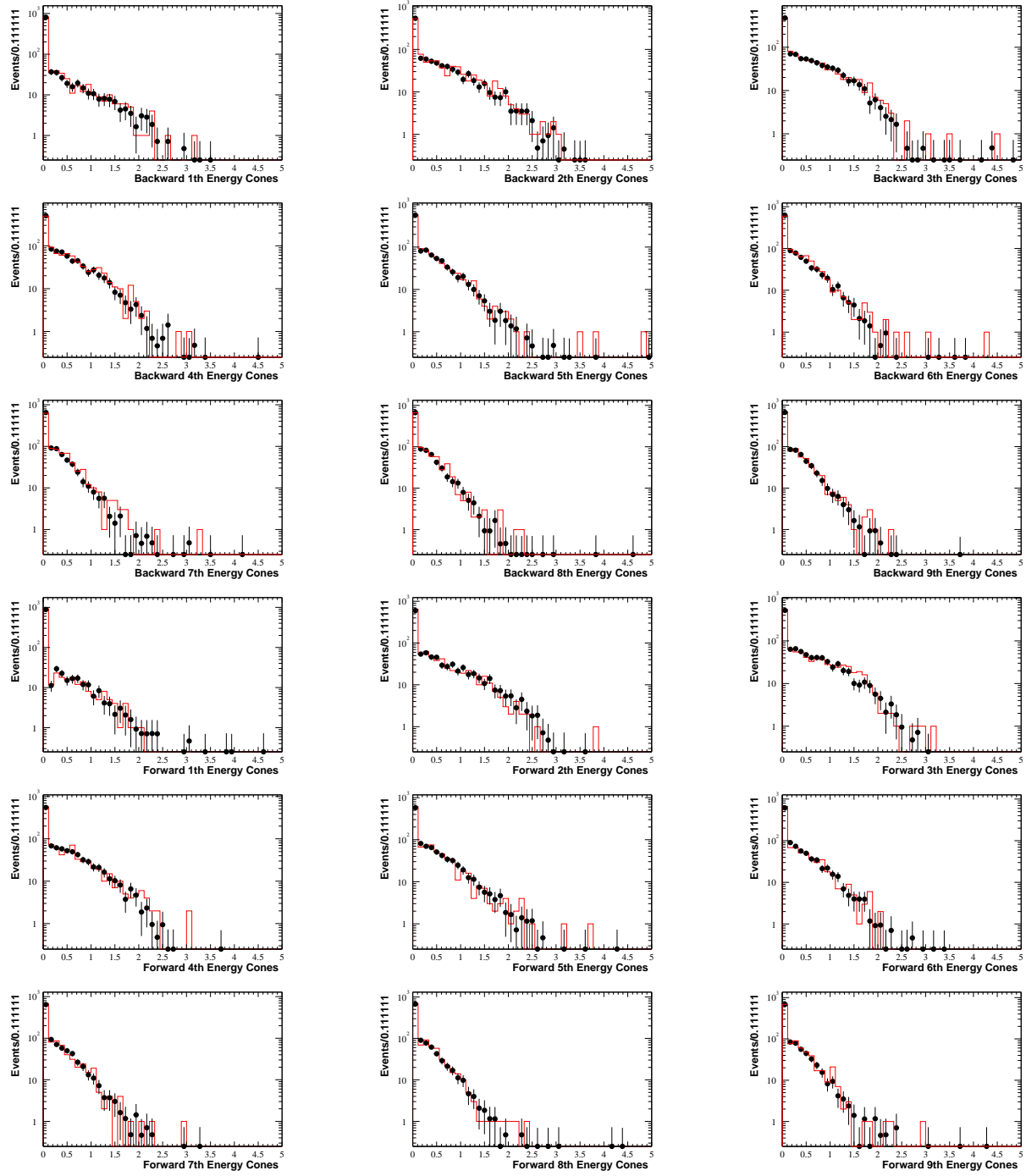


Figure 9: 18 energy cone distributions for background. The Monte Carlo(red line) is compared with off-peak data(solid circle).

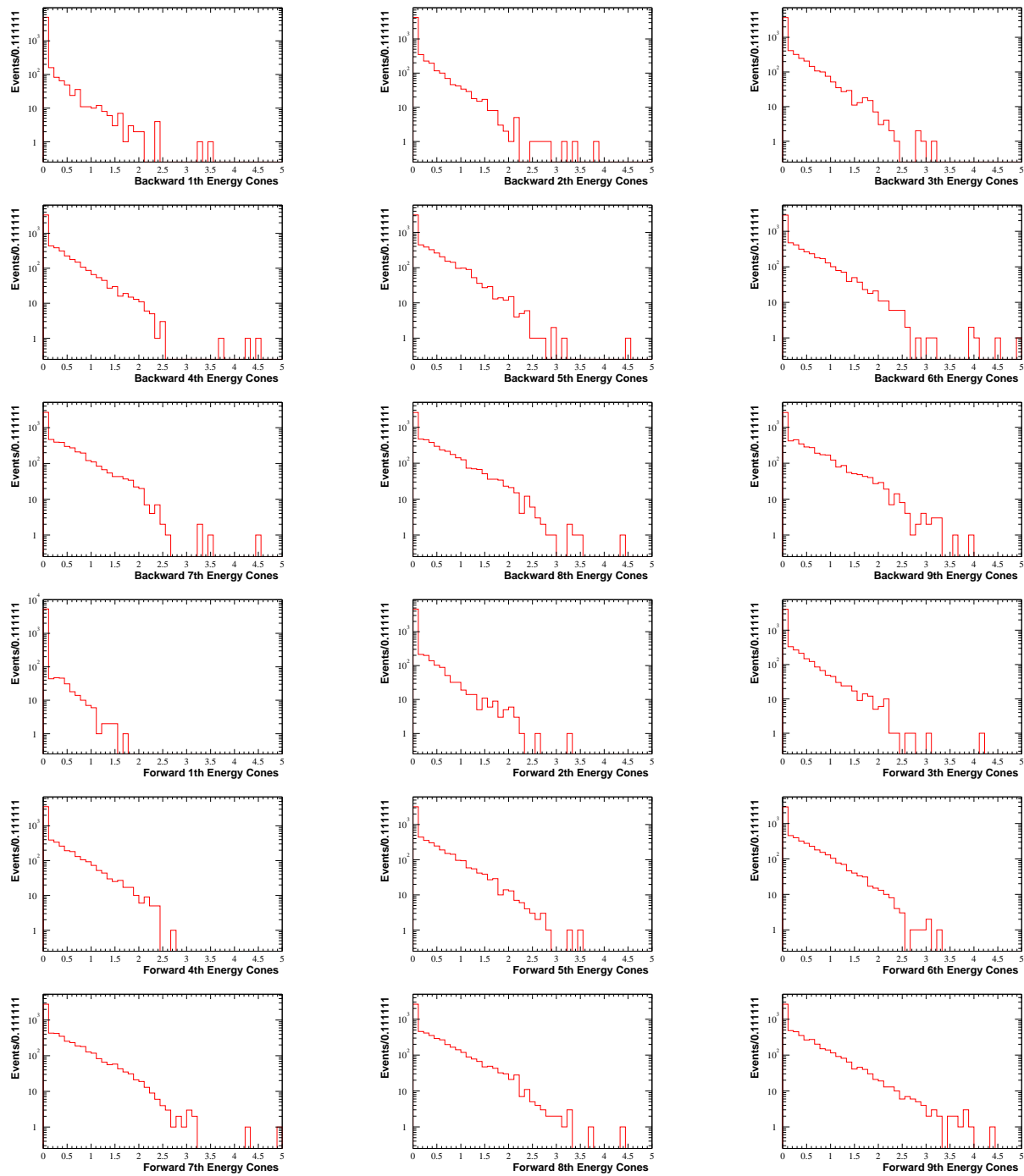
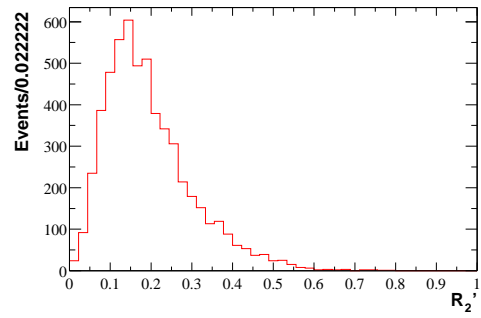
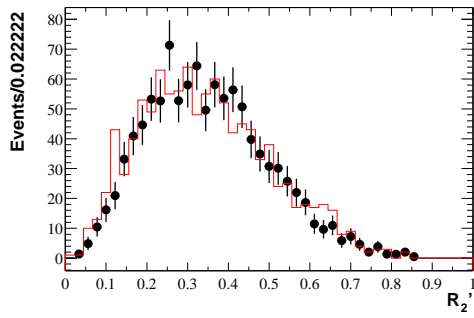
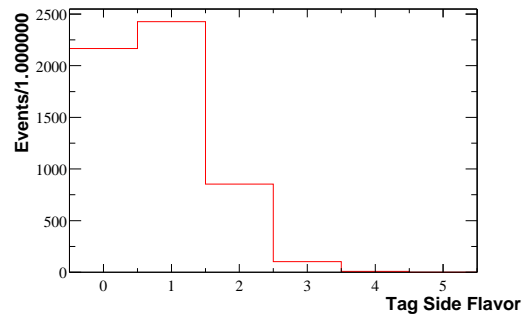
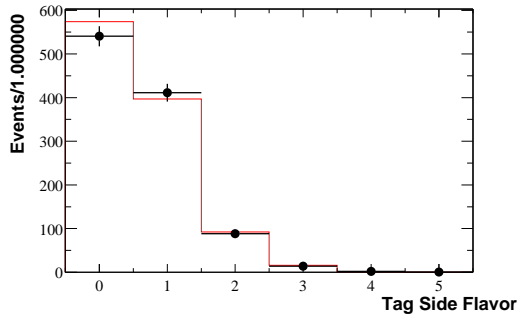


Figure 10: 18 energy cone distributions for signal Monte Carlo.



(a)



(b)

Figure 11: a)  $R_2'$  and b) Tagside Flavor distributions for background (left) and signal Monte Carlo(right). The background Monte Carlo (solid circle) is compared with off-peak data (red line).

the signal event and background event, but this difference is not good enough to apply any direct cut on it.