SEARCH FOR THE FLAVOR-CHANGING NEUTRAL CURRENT IN TOP

PAIR EVENTS WITH AN ASSOCIATED PHOTON USING 13 TEV

PROTON-PROTON COLLISION DATA COLLECTED WITH THE ATLAS

DETECTOR

by

JASON TYLER BARKELOO

A DISSERTATION

Presented to the Department of Physics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2020

DISSERTATION APPROVAL PAGE

Student: Jason Tyler Barkeloo

Title: Search for the Flavor-Changing Neutral Current in Top Pair Events With an Associated Photon Using 13 TeV Proton-Proton Collision Data Collected With the ATLAS Detector

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics by:

David Strom	Chair
James Brau	Advisor
Spencer Chang	Core Member
Dev Sinha	Institutional Representative

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2020

DISSERTATION ABSTRACT

Jason Tyler Barkeloo

Doctor of Philosophy

Department of Physics

March 2020

Title: Search for the Flavor-Changing Neutral Current in Top Pair Events With an Associated Photon Using 13 TeV Proton-Proton Collision Data Collected With the ATLAS Detector

Abstract for FCNC here.

This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Jason Tyler Barkeloo

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:
  University of Oregon, Eugene, Oregon
  Miami University, Oxford, Ohio
  Wittenberg University, Springfield, Ohio

DEGREES AWARDED:
  Doctor of Philosophy, Physics, 2020, University of Oregon
  Master of Sceience, Physics, 2012, Miami University
  Bachelor of Science, Physics, 2010, Wittenberg Univeristy

PROFESSIONAL EXPERIENCE:

  University of Oregon, Graduate Research Assistant, ATLAS Experiment, June 2014-Present

  University of Oregon, Graduate Teaching Assistant, September 2012-June 2014

  Miami University, Teaching and Research Assistant, August 2010 - May 2012

  Wittenberg University, Undergraduate Researcher, August 2008 - May 2010

GRANTS, AWARDS AND HONORS:

  Weiser Senior Teaching Assistant Award – University of Oregon – 2014

  American Association of Physics Teachers Outstanding Teaching Assistant – Miami University – 2012

PUBLICATIONS:

  J. Barkeloo, J. Brau, M. Breidenbach, R. Frey, D. Freytag, C. Gallagher, R. Herbst, M. Oriunno, B. Reese, A. Steinhebel, D. Strom. "A silicon-tungsten electromagnetic calorimeter with integrated electronics for the International Linear Collider" J. Phys.: Conf. Ser. **1162** 012016 (2019).

J. Kangara, A. Hachtel, M. C. Gillette, J. Barkeloo, E. Clements, S. Bali. "Design and construction of cost-effective fail-safe tapered amplifier systems for laser cooling and trapping experiments", Am. J. Phys. **82**(8), 805 - 817 (2014).

A. Hachtel, J. Kleykamp, D. Kane, M. Marshall, B. Worth, J. Barkeloo, J. Kangara, J. Camenisch, M. Gillette, S. Bali. "An undergraduate lab on measurement of radiative broadening in atomic vapor", Am. J. Phys. **81**(6), 471 (2013).

Additional ATLAS Collaboration publications can be found:
http://inspirehep.net/search?p=exactauthor%3AJason.Barkeloo

TABLE OF CONTENTS

LIST OF FIGURES

ix

x

LIST OF TABLES

CHAPTER I

SEARCH STRATEGY

This chapter will describe the major backgrounds and outline a search strategy used in the analysis. The kinematic regions for the signal will be defined as well as the introduction of neural networks to assist with the separation of signal and background like events.

1.1. Major Backgrounds

There are a large number of Standard Model processes that can end up in the signal region and share a similar final state topology as the studied signal process, $t\bar{t} \to bl\nu q\gamma$. All of these processes are modeled with Monte Carlo(MC) simulation, with the exception of particles that fake other particles (leptons and photons) which are done using data-driven techniques because they are poorly modeled with MC. A full list of the Monte Carlo samples used for each background can be found in Appendix **??**.

The dominant backgrounds in this search are Standard Model $t\bar{t}$, Standard Model $t\bar{t}+\gamma$, as well as W+jets and W+jets with an associated photon (W+jets+γ). These along with minor backgrounds (Standard Model single top events, Z+jets (Z+jets+γ), $t\bar{t}$+Vector Boson, diboson) are all modeled with MC and some data-driven estimates and corrections are applied. The various backgrounds modeled in this search as summarized here:

SM Processes: SM $t\bar{t}$, W+jets, Z+jets, single top, diboson, $t\bar{t}$+V are modeled with MC simulations. Control and validation regions are designed to test preformance of the largest of these background proces, $t\bar{t}$ and W+jets. A

discussion of the modeling of the major SM processes without photons and a derivation of additional scale factors for these processes is shown in Section 1.6.1.

SM Processes with an associated photon*:* SM + associated photon processes: $t\bar{t} + \gamma$, W+jets+γ, Z+jets+γ are modeled with MC simulation and overlap removal is applied to the SM processes to remove events with similar phase spaces since these processes are a subset of the SM process. MC simulations are created with additional statistics for these processes. Additional details are presented in Section 1.6.2.

Fake Leptons*:* Non-prompt or fake electrons and muons can arise from semi-leptonic decay of b and c quarks. For electrons additional contributions from photon conversions and jets in the electromagnetic calorimeter can lead to lepton fakes. Muons can be faked via energetic showers in the hadronic calorimeter or from hadrons that punch-through the hadronic calorimeter. The matrix method is used to estimate the number of events with fake leptons as described in Section 1.7.2.

Fake Photons*:* The number of events with fake photons are estimated using a $Z \rightarrow e^+e^-$ tag-and-probe method discussed in Section 1.7.1. Further, the ABCD method is used in Section 1.7.3 to estimate the number of fake photon events that arise from a jet faking a photon.

1.1.1. Overlap Removal

Tracks and energy deposits within the detector can, in some cases, be used to reconstruct multiple objects. To prevent using these tracks and deposits multiple

times a standard overlap removal procedure is applied to objects. First, electrons that share tracks with any other electrons are removed. Any electron sharing a track with a muon is then also removed. Any jet that is found within $\Delta R < 0.2$ of an electron is removed. Then any jet with less than 3 tracks associated with it within $\Delta R < 0.2$ of a muon object is removed. After that any muon found withing $\Delta R < 0.4$ of a jet is removed and any photon within $\Delta R < 0.4$ of an electron or muon object is removed.

1.1.2. Duplicate Event Removal

As specialized higher statistic samples are used for processes with prompt photons a double counting of events could occur with the nominal MC samples. For example, in addition to the $t\bar{t}$ sample a sample of $t\bar{t} + \gamma$ events are used. This is true for the W+jets/Z+jets and special samples of W+jets+γ and Z+jets+γ. Therefore a truth based matching scheme is used to remove events in the nominal samples that match with the photon types produced in the specialized +γ samples i.e., they contain a truth photon that does not originate from a hadron or lepton.

1.2. Event Selection

This analysis is searching for $t\bar{t}$ candidate events where one of the top quarks decays through the most common decay path (a W boson that decays leptonically to an electron or muon and a bottom quark) and the other through the FCNC diagram to an up type light quark and a photon. The selected decay path is then $t\bar{t} \rightarrow Wbq\gamma \rightarrow l\nu bq\gamma$ such that the final state will contain at least two jets exactly one of which is b-tagged using the MV2c10 algorithm, exactly one massive lepton (an electron or muon), exactly one highly energetic photon, missing transverse energy from the

3

neutrino, and a large transverse W mass. The transverse W mass requirement selects events that have a lepton and $\not{E}_T$ which are consistent with a leptonically decaying W boson, defined as:

$$m_T^W = \sqrt{2 \times p_{T_l} \times \not{E}_T \times (1 - \cos(\phi_l - \phi_{\mathrm{MET}}))}$$

which is largest when the lepton and missing transverse energy are close in ϕ as would be expected from the decay products of a boosted W boson. Events are selected loosely to include all possible kinematic regions of interest and then skimmed down to the individual regions for various studies. This includes events that will not enter the signal region such as events with 0 photons or events with 0 b jets that are used to derive and test scale factors on the largest background samples and account for mismodeling of MC simulations.

For the events that have a chance of entering the final Signal Region i.e., events with exactly one b-tagged jet and exactly one photon, a neural network analysis is preformed to help separate the signal from the background using a variety of high dimensional cuts. The neural network training and testing are described in Section 1.3. The neural network is then applied to all events with 1 b-tagged jet and 1 photon and greatly increases signal purity while separating out the most dominant backgrounds, in particular $t\bar{t}$ and $t\bar{t} + \gamma$.

1.3. Event Classification: Neural Network Optimization

To help distinguish signal events from the majority of background events neural networks were employed for event classification. Neural networks are multivariate methods that take a variety of inputs and output a number between 0 and 1. The

output value is a discriminating variable that will be used to classify events and determine which events make it into the final Signal Region selection. Signal-like events accumulate towards 1 while background-like events cluster around 0. Two neural networks are trained, one for the electron+jets final state and one for the muon+jets final state. This section will discuss the neural network studies completed and their uses in the search for FCNC events.

1.3.1. Input Variables

A wide variety of input variables to the neural network were studied in detail. Studies were done using only low level variables such as the kinematic variables (p_T, η, ϕ, E) of the physics objects in the signal region. This was done as a complex enough neural network should be able to figure out useful high level/event level variables (i.e. invariant masses, geometric separations) but in practice a combination of some of these low level variables and high level variables used as inputs to the neural network proved to give the best separation and projected limits. Using physical intuition to guide the neural network proved to be a valuable tool.

Combinations of 29 input variables were tested to start with however variables such as η and ϕ tend to not have significant weights in the neural network and are left out in favor the the high level variables that include them (e.g., ΔR values). A measure of how different the variables are between signal and background is the Separation. Table 1.1 shows the separation values for the variables that are inputs to the final neural network. Comparisons between the shapes of the input variables for the μ+jets channel are shown in Figures 1.1, 1.2, and 1.3

$$\text{Separation} = \sum_i^{bins} \frac{n_{si} - n_{bi}}{n_{si} + n_{bi}}$$

Variable	Separation e+jets	Separation μ+jets
$p_T(\gamma)$	22.97	24.01
$m_{q\gamma}$	22.65	28.31
γ_{iso}	18.62	41.32
m_{bW}	11.10	11.70
$m_{l\gamma}$	9.00	7.51
$\Delta R_{j\gamma}$	4.59	5.66
ΔR_{bl}	4.99	4.47
m_T^W	3.16	3.37
S_T	3.78	3.32
n_{jets}	1.70	2.03
χ_W^2	1.37	1.91
$p_T(q)$	2.46	2.82
$\Delta R_{l\gamma}$	1.40	1.19
E (lepton)	0.86	0.89
$\not{E}_T$	0.47	0.70
$p_T(b)$	0.51	0.53

TABLE 1.1. Separation of normalized variables between signal and bacground in the e+jets and μ+jets channels for the variables used as input to the final neural network.

Typically the kinematic variables with photon information have the biggest separation values. This is expected because the signal photon comes directly from the decay of a top quark and is much more energetic than background photons. Shape comparison plots for the e+jets channel and additional plots for other investigated variables are shown in Appendix **??**. The largest difference in separation between the e+jets and μ+jets channels is the photon isolation value. This is due to the fact that all backgrounds are included and fake photon contamination from a large Z+jets background are expected. Both networks preform similarly in their separation of signal and background events. The network is able to learn and compensate for this behavior with the help of other variables that include the lepton and photon: $\Delta R_{l\gamma}$ and $m_{l\gamma}$.

The neural networks are trained on MC events that have a chance of being in the signal region after basic event level cuts and optimized for signal significance. Only events with 1 photon (> 15 GeV) and 1 bjet (MV2c10 77% working point) are classified by the neural network. The 77% working point was chosen by training the neural network on events with only 1 bjet at each working point: 70%, 77%, and 85% and picking the network and working point with the best estimated significance. The b-tagging neural network study is shown in Section 1.3.4
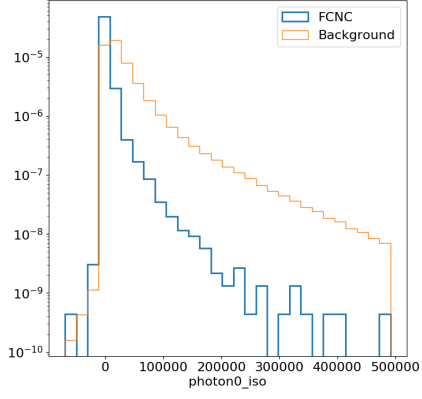
(a) γ_{iso} topo$E_T$cone40
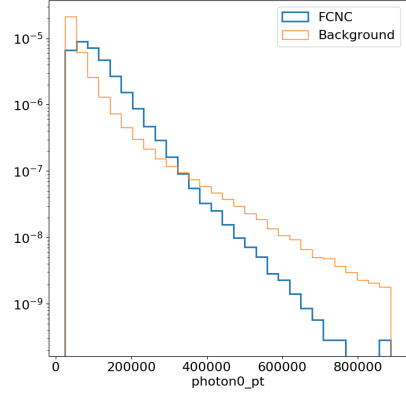
(b) γ_{p_T}
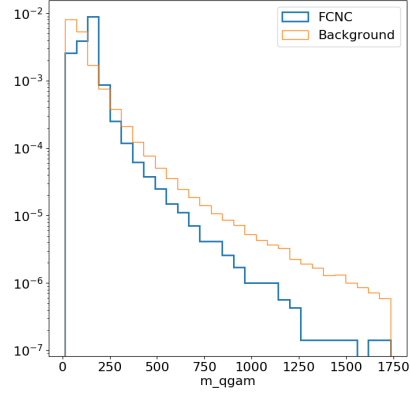
(c) $m_{q\gamma}$
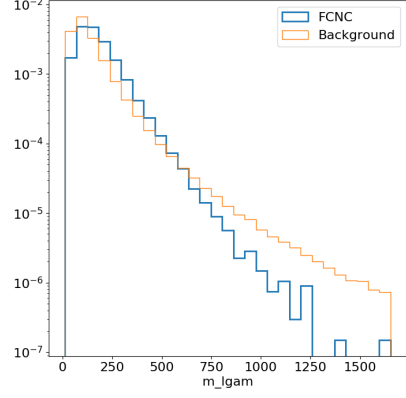
(d) $m_{l\gamma}$
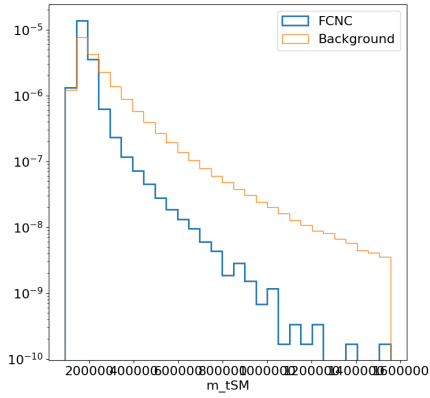
(e) m_{bW}
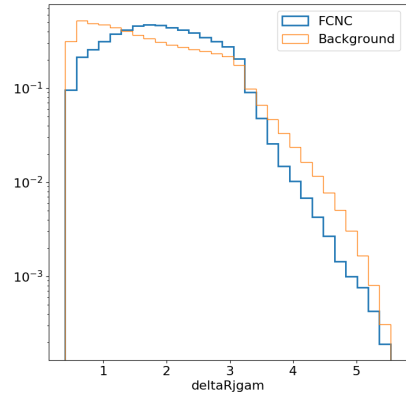
(f) $\Delta R_{j\gamma}$

FIGURE 1.1. Normalized variables showing the shapes of neural network input variables for the μ+jets channel: γ_{iso} topo$E_T$cone40, γ_{p_T}, $m_{q\gamma}$, $m_{l\gamma}$, m_{bW}, and $\Delta R_{j\gamma}$
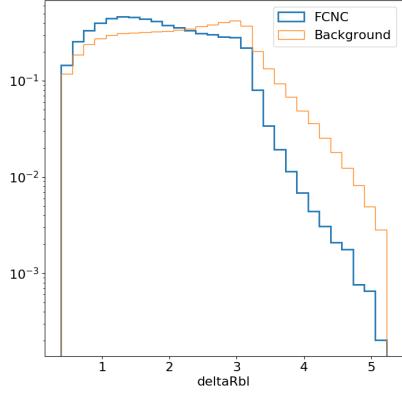
(a) ΔR_{bl}

(b) m_T^W

(c) S_T
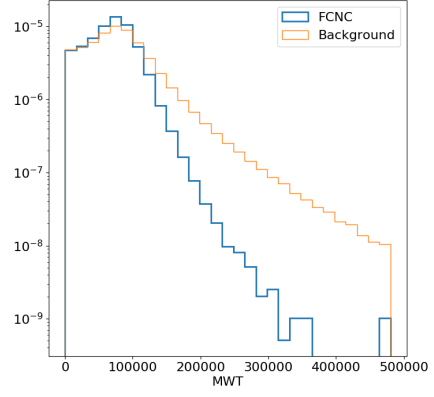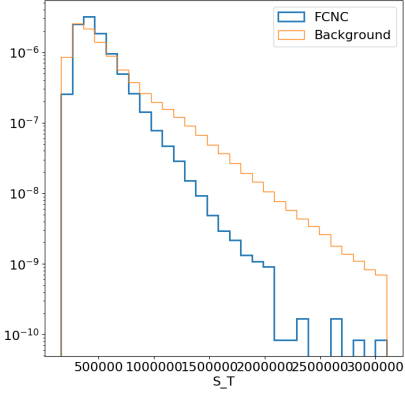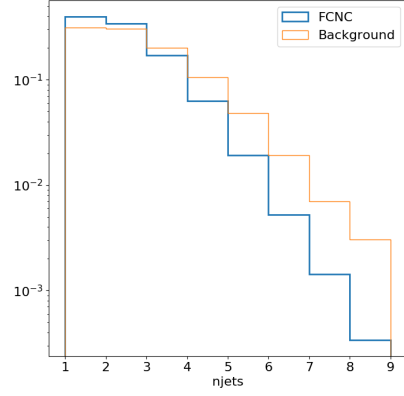
(d) n_{jets}
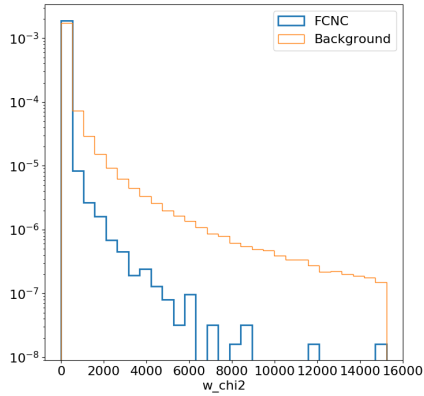
(e) χ_W^2
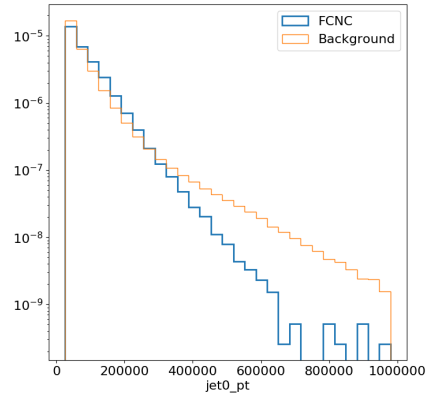
(f) $p_T(q)$

FIGURE 1.2. Normalized variables showing the shapes of neural network input variables for the μ+jets channel: ΔR_{bl}, m_T^W , S_T, n_{jets}, χ_W^2, and $p_T(q)$
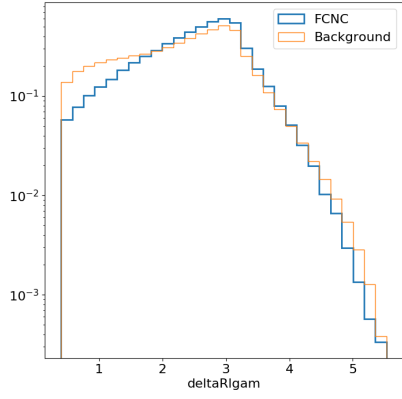
(a) $\Delta R_{l\gamma}$
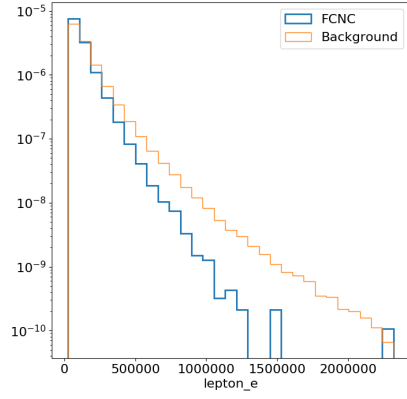
(b) E (lepton)
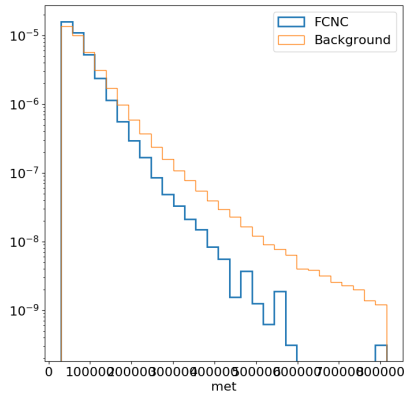
(c) $\not{E}_T$
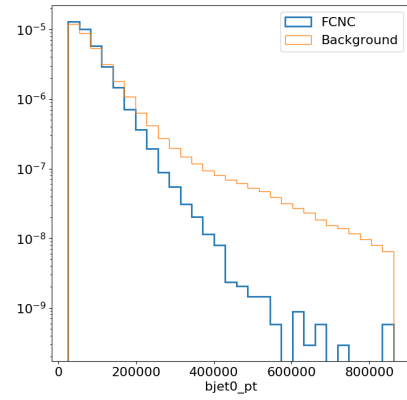
(d) $p_T(b)$

FIGURE 1.3. Normalized variables showing the shapes of neural network input variables for the μ+jets channel: $\Delta R_{l\gamma}$, E (lepton), $\not{E}_T$, and $p_T(b)$

1.3.2. Architecture

A variety of architecures of dense neural networks are studied using KERAS[?
] on top of the TENSORFLOW backend [?]. Each network has a number of input
nodes equal to the number of input variables. Networks with one, two, and three
hidden layers are investigated each with 20 nodes. The output layer contains only a
single node. Every node in one layer is connected to every node in the next layer and
the previous layer. Every connection is assigned a weight that is optimized during
the training of the network. For every node in the network a value is computed using
the weights and input values of the previous nodes using an activation function.
Nodes with the highest output of this function are more important to the fit. The
activation function used on the internal nodes in this search is the Rectified Linear
Unit activation function.

$$ReLU(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

The output layer uses the sigmoid function, $\sigma(x)$, as an activation function. The
sigmoid function maps the output smoothly to the range (0,1).

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Every training step the weights of each node are updated following an optimization
algorithm, in this case the ADAM optimizer[?]. This optimizer follows the steepest
gradient to reach the minimum of the parameter of interest called the loss function.
The loss function used for these classification neural networks is the binary cross

entropy:

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(p(y_i)) + (1 - y_i)\log(1 - p(y_i))$$

where y is a binary indicator (0 or 1) if class label is the correct classification for observation and p is the predicted probability observation is the class label (0 or 1). The logarithmic nature of this loss function means it applys small values to correctly assigned event but more harshly punishes mismatching of events. Therefore having a similar number of signal and background events that get weighted similarly can improve the behavior of the network. In rare decay searches typically the amount of signal events is significantly smaller than the amount of background events in the training sample. Using the weight functionality in keras the total number of signal events can be scaled to be similar to the number of background events.

Weighting the signal events this way allows the network to separate the signal and background events in a way that is significantly less harsh than without the weights by taking advantage of the loss function being used. This improves the estimated significance of the neural network cut after the signal events are rescaled to their proper normalization values.

Various hyperparameters are used as inputs into the neural network as well as the optimizer used. The ADAM optimizer has a default learning rate of 0.001 which was not changed throughout these studies. The learning rate corresponds to the amount that weights are updated during training. A learning rate that is too large can mean the network never settles into a local minima as it is always missing the minima or at the very least it can take much longer to converge into a minima. As the neural network training for this search always converged quickly and to a similar value after being tested multiple different times the learning rate was not adapted.
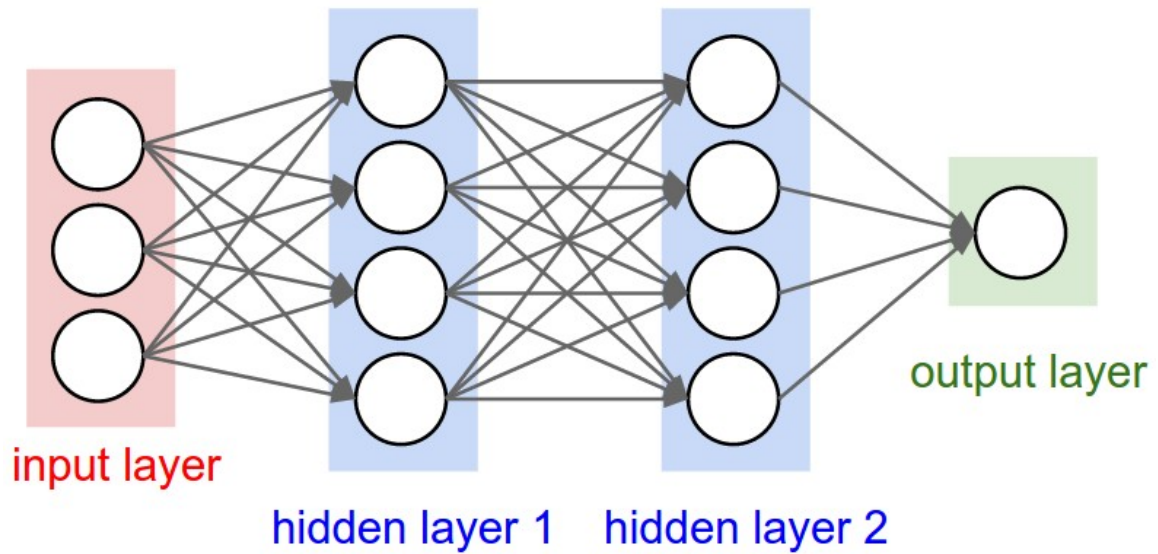
FIGURE 1.4. Pictoral representation of neural network architecture with 3 input variables, 2 hidden layers with 4 nodes each, and 1 output layer[**?**].

Another hyperparameter of note is the batch size which defines the number of samples that are propagated through the network at once. The batch size is of crucial importance in how long the training of the network takes. A set of 1000 training samples with a batch size of 100 will propagate each set of 100 samples through the neural network every epoch, so 10 separate batches. A larger batch size means that each epoch of the training takes a shorter amount of time. However, as the weights are updated after each batch the network can take many more epochs to converge as the weights are being updated less frequently. A batch size of 100 was used while training the networks presented in this chapter. Larger batch sizes were tested with the only difference being the time each epoch took and the total time the network took to converge.

Epochs are the total number of times the network has been trained over the entire training set. All of the networks were allowed up to 200 epochs to converge with a KERAS patience value set to 50. The loss function minimization would be

done every batch and after each epoch the best possible value of the loss function is found. If this value is better than any previous epoch the network is allowed to train for 50 more epochs until 50 epochs have passed without finding a new minimum loss function value which then terminates the training. All models converge early and are terminated typically between epoch 80 and 120 meaning the loss function was minimized between epoch 30 and 70.

One method employed to avoid overtraining the network dropout regularization was used on each of the hidden layers. Dropout has the effect of simulating a large number of networks with very different network structures by removing nodes randomly throughout the training. A dropout rate of 20% was used meaning that for every batch 20% of the weights of the hidden layer nodes were set to 0. This forces the network to not become overly dependent on any given node and learning the data 'by heart' as opposed to recognizing the trends in the sample.

1.3.2.1. Training and Validation of Neural Networks

The input variables into the neural network are preprocessed using the ROBUSTSCALAR method implemented in **scikit-learn**[?]. The preprocessing is done so that the input variables exist on a similar scale. As the network is tasked with learning how to combine these inputs through a series of linear combinations and nonlinear activation function values a disparity in the scales of the input values can lead to awkward loss function topology that will focus on certain parameter gradients instead of treating them all similarly. Normalizing the values to a standard scale allows the network to learn the optimal parameters for each input node more quickly and efficiently. This means that less focus can be used on the optimization of the

hyperparameters for the network as the scales of the inputs do not need to be learned by the network itself.

Each input variable in the neural network, x, is scaled by the following equation:

$$z = \frac{x - m}{q_3 - q_1}$$

where m is the median of the distribution, q_1 and q_3 are the first and third quartile. This changes the distribution of the input variable distributions to be centered around zero.

A second method to avoid overtraining the neural network is to make use of a train-test split to split the signal and background samples into 3 independent randomized sets before training the neural network. The samples are split into a training set of 64% of the samples, a test set containing 20% of the samples, and the remaing 16% are a validation set. The training and test sets are used during the training of the network while the validation set is used to compute performance of the trained neural network.

One measure of the performance of the network is the accuracy. The KERAS default accuracy measure is defined:

$$\text{accuracy} = \frac{N(\text{event}_{NN} \geq 0.5|\text{signal}) + N(\text{event}_{NN} < 0.5|\text{background})}{N(\text{signal}) + N(\text{background})}$$

where $N(\text{event}_{NN} \geq 0.5|\text{signal})$ $(N(\text{event}_{NN} \geq 0.5|\text{signal}))$ is the number of signal (background) events with $P_{\text{signal}} \geq 0.5$ $(P_{\text{signal}} < 0.5)$. Essentially, the accuracy is a measure of the mean of how often correct prediction values occur assuming a cut on the output of ≥ 0.5.

1.3.3. Hidden Layer Studies

The general performance of the neural network was studied with a varying number of hidden layers (1, 2, and 3) in both the e+jets and μ+jets channels. All of the networks are trained on the same set of variables and with the same train-test split input data. For each of the channels the *Receiver Operating Charactersitic* (ROC) curves are shown in Figure 1.5. The ROC curves show the value of $1 - \epsilon_{bkg}$ as a function of the true positive rate, ϵ_{signal}. A figure of merit is the Area Under the Curve (AUC) which is a measure of how close the resulting values are to the optimal value of unity.
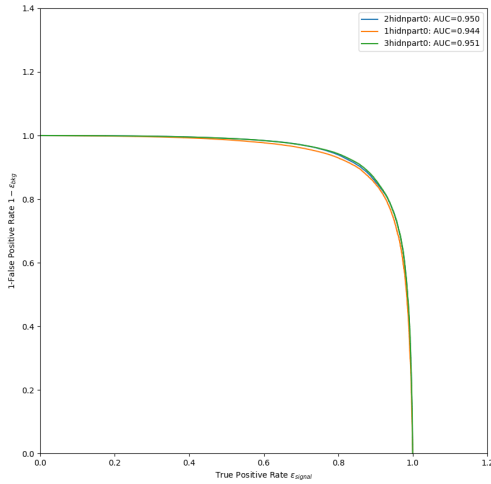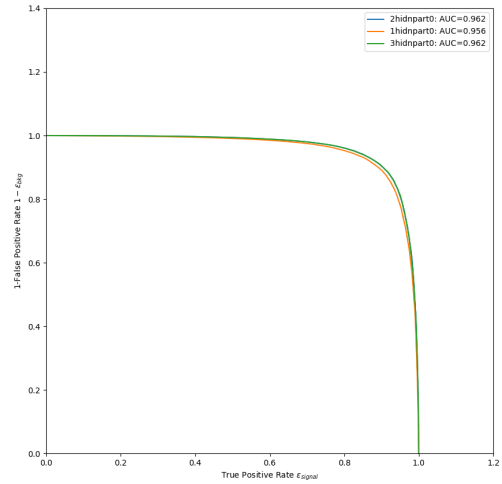
(a) e+jets ROC Curves   (b) μ+jets ROC Curves

FIGURE 1.5. ROC Curves are shown for both search channels for a varying number of hidden layers. Orange lines correspond to one hidden layer, blue to 2 hidden layers and green to 3 hidden layers. The blue and green curves have near identical AUC values: 0.950 and 0.951 for the e+jets case and 0.962 for the μ+jets cases.

The AUC for 2 hidden layers and 3 hidden layers are identical, to rounding errors, for both channels. As such the network with 2 hidden layers has been chosen as it

is computationally simpler. The normalized neural network output values are shown in Figure 1.8. Adding a second hidden layer significantly improves the performance of the network but a third layer does not. The output shapes change slightly adding the third hidden layer due to the network learning differently about the same data. However, as the AUC shows, the performance of 2 and 3 hidden layers is identical. Figures 1.6 and 1.7 show the accuracy metric and the loss function as a function of the training epoch for the networks trained with 2 hidden layers. The accuracy plot behavior is expected as the validation data sets do not have dropout regularization applied to them. These networks are also trained without further reduction of Z+jets background meaning the e+jets sample has a larger background contamination that makes the validation testing more volatile. This is due to the increased number of similar events in that sample that can be more heavily dependent on specific weights across the network for identification.
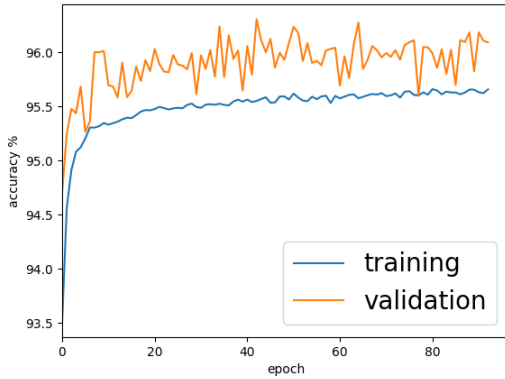
(a) e+jets Accuracy Curves (b) μ+jets Accuracy Curves
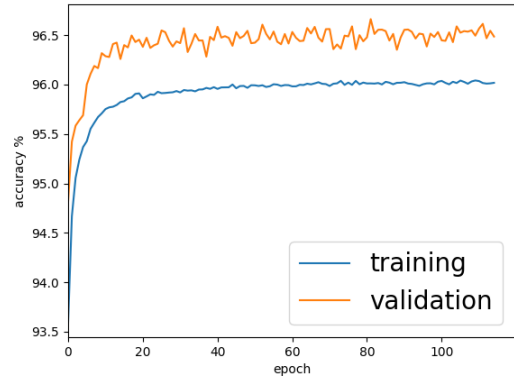
FIGURE 1.6. Accuracy plots for both channels for the 2 hidden layer neural network
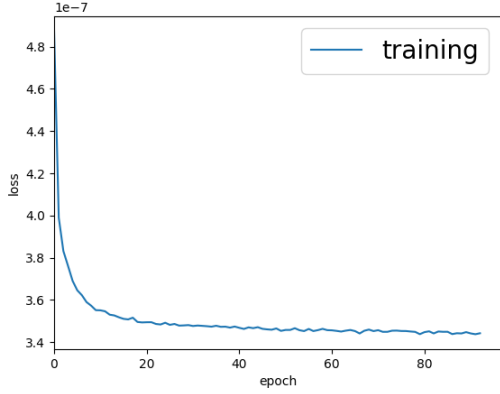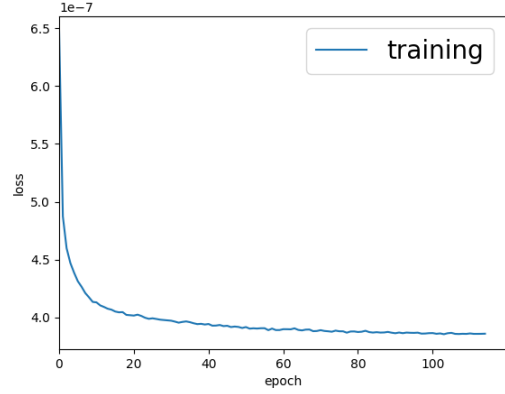
(a) e+jets Loss Curve　　　　　　　(b) μ+jets Loss Curve

FIGURE 1.7. Loss plots for both channels for the 2 hidden layer neural network

The main metric used in choosing which network has the best physics reach is the significance:

$$\text{significance} = \frac{N_s}{\sqrt{N_s + N_b}}$$

where N_s is the number of signal events that pass the cut and N_b is the number of background events that pass the neural network cut. After the model has been fully trained it is tested on all of the Monte Carlo for signal and background. The signal samples are normalized to various branching ratios (in the range $10^{-5} \rightarrow 3 \times 10^{-3}$) and full LHC Run-2 Luminosity and the significance is calculated as a function of the cut on the output of the neural network $P(\text{signal})$. The network with the output cut for the smallest branching ratio with a maximum significance of 2 is chosen, a rough estimate of where the expected limit could be set. The significance as a function of the neural network output cut is shown in Figure 1.9.

(a) e+jets:1 Hidden Layer   (b) e+jets:2 Hidden Layers   (c) e+jets:3 Hidden Layers

(d) μ+jets:1 Hidden Layer   (e) μ+jets:2 Hidden Layers   (f) μ+jets:3 Hidden Layers
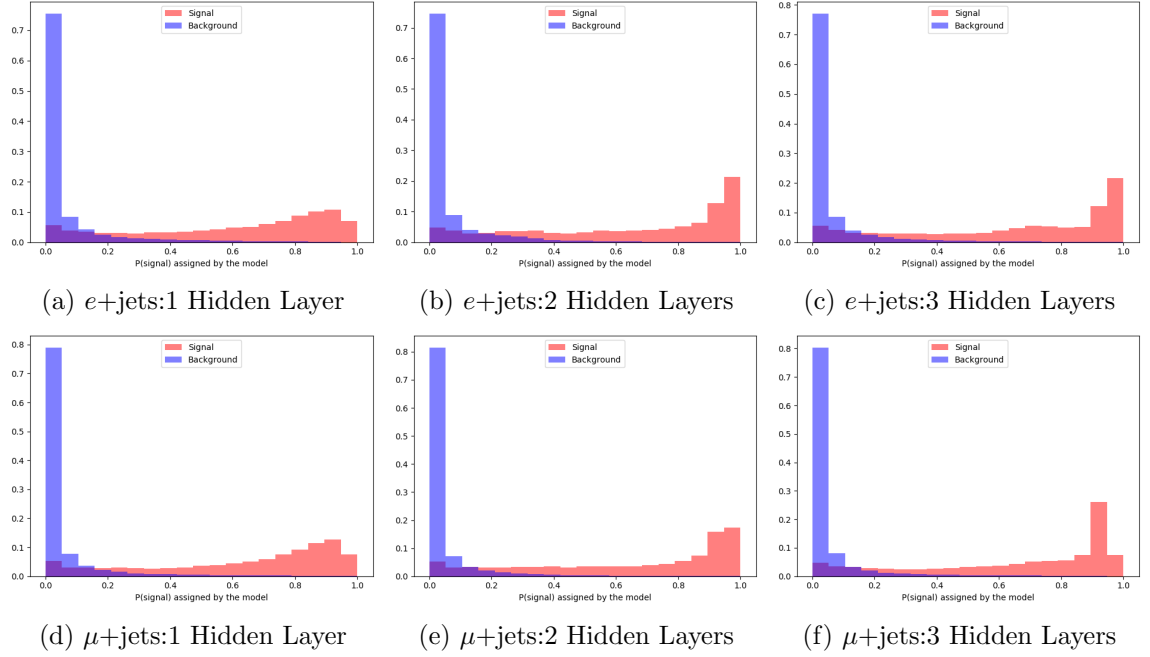
FIGURE 1.8. Normalized neural network output signal and background distribution plots are shown for both search channels for a varying number of hidden layers.

(a) e+jets   (b) μ+jets
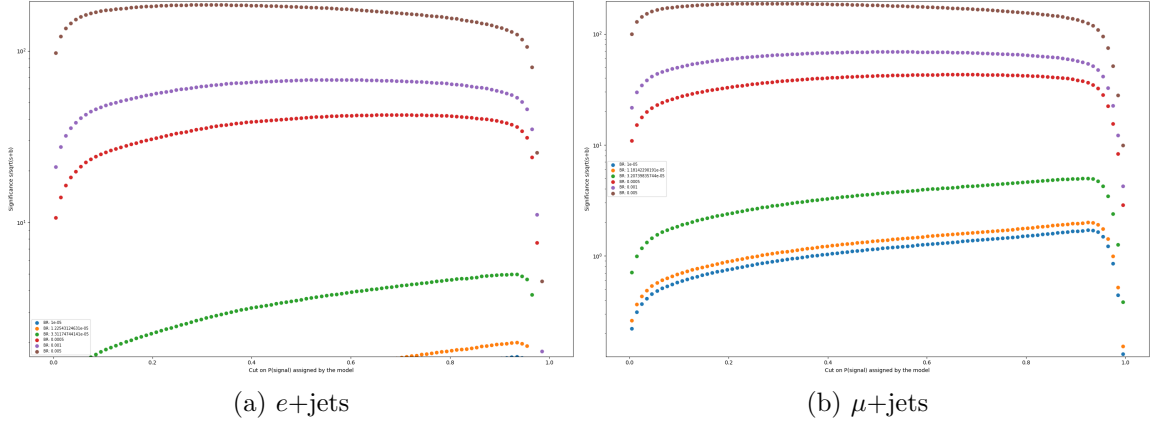
FIGURE 1.9. Significance plots for both channels for the 2 hidden layer neural network. The green points correspond to a branching ratio with a maximum significance of 5, the orange to a maximum significance of 2. The e+jets (μ+jets) branching ratio with max significance of 2 is $1.22 \times 10^{-5} (1.18 \times 10^{-5})$. The blue, red, purple, and brown points correspond to branching ratios of 1×10^{-5}, 5×10^{-4}, 1×10^{-3}, and 5×10^{-3}, respectively.

19

1.3.4. B-Tagging Working Point Studies

The b-tagging working point selection was also done with similar neural network studies. Three neural networks were trained with the datasets using the jet information and total scaled events for each of the major b-tagging working points: 70%, 77%, and 85%. Changing the working point changes a number of things about the signal and background data sets such as which jets are b tagged and therefore which jets are combined into the higher level variables (e.g., $m_{q\gamma}$ and m_{Wb}). The total number of events that pass the preselection to the neural network are also changed for all of the datasets since the neural network are only trained on events with 1 b-tagged jet. Similar sets of plots to Section 1.3.3 will be presented in this section.

This selection of neural networks were trained in parallel with one, two, and three hidden layers. The only results shown are the 2 hidden layer outputs as they preform equally or better to the others as previously discussed. The accuracy and loss plots for these networks are shown in Figures 1.10 and 1.11. Following that the neural network output and significance plots are shown in Figures 1.12 and 1.13.

The result of these studies is the choice of using the 77% working point for b-tagged jets. The branching ratio with significance of 2 is found for each network and reported in Table 1.2.

B-Tag Working Point	e+jets Branching Ratio	μ+jets Branching Ratio
70%	1.25×10^{-5}	1.31×10^{-5}
77%	1.23×10^{-5}	1.18×10^{-5}
85%	1.27×10^{-5}	1.19×10^{-5}

TABLE 1.2. Branching ratio values with a significance of 2 after neural network optimization
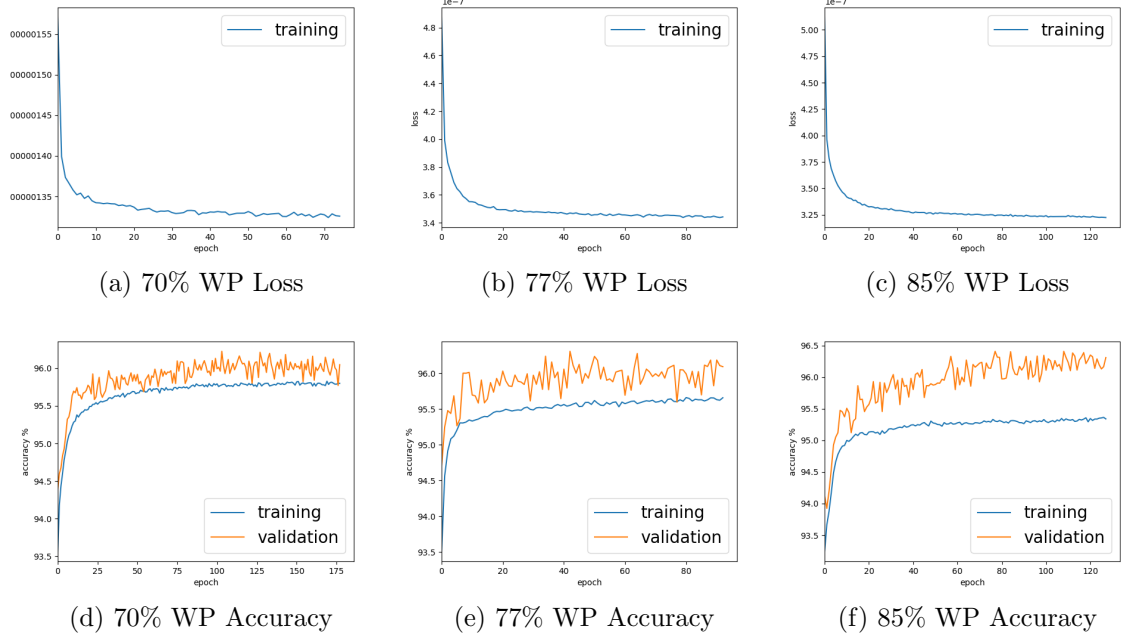
(a) 70% WP Loss   (b) 77% WP Loss   (c) 85% WP Loss

(d) 70% WP Accuracy   (e) 77% WP Accuracy   (f) 85% WP Accuracy

FIGURE 1.10. Accuaracy and loss plots for the e+jets channel at 70%, 77%, and 85% b-tagging working points.

(a) 70% WP Loss   (b) 77% WP Loss   (c) 85% WP Loss

(d) 70% WP Accuracy   (e) 77% WP Accuracy   (f) 85% WP Accuracy

FIGURE 1.11. Accuaracy and loss plots for the μ+jets channel at 70%, 77%, and 85% b-tagging working points.

(a) 70% WP NN output   (b) 77% WP NN output   (c) 85% WP NN output

(d) 70% WP Significance   (e) 77% WP Significance   (f) 85% WP Significance
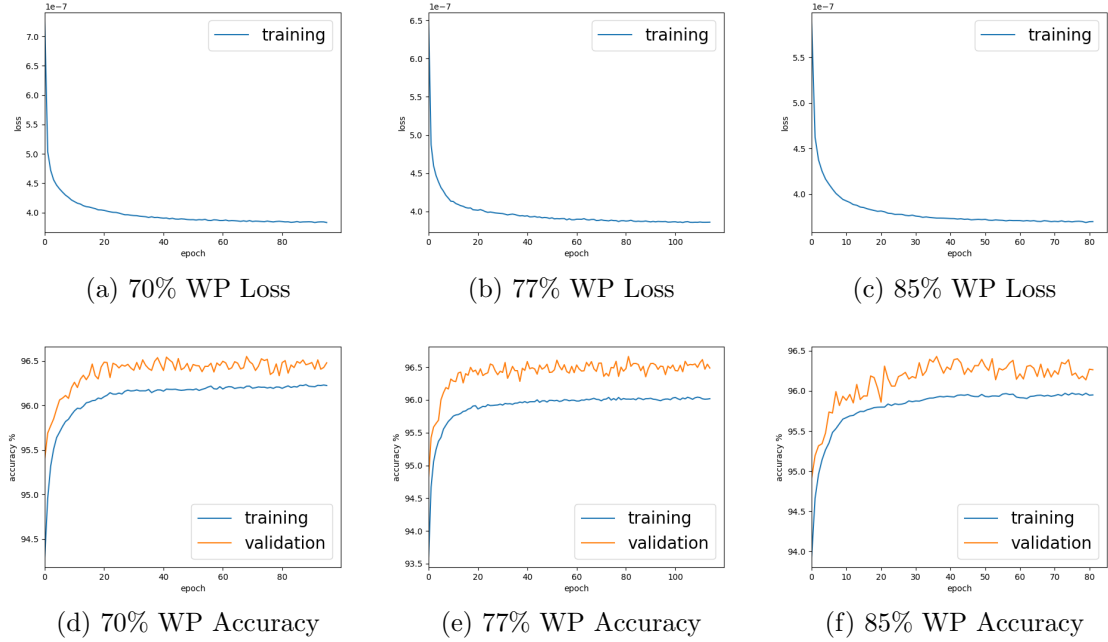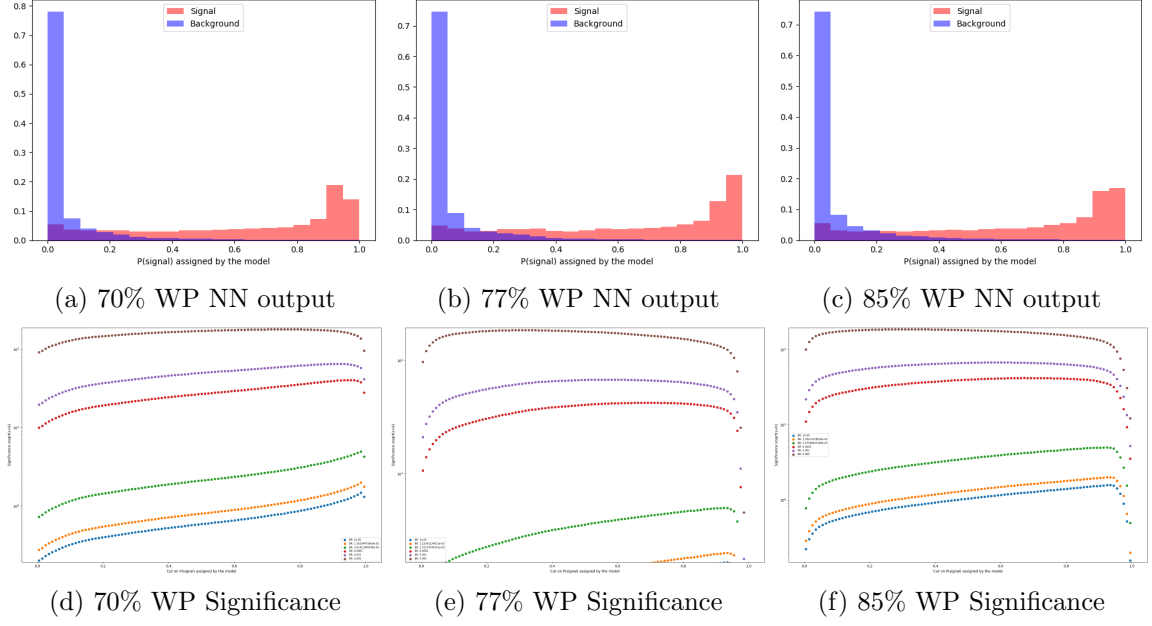
FIGURE 1.12. Neural network output and significance plots for the e+jets channel at 70%, 77%, and 85% b-tagging working points.

(a) 70% WP NN output   (b) 77% WP NN output   (c) 85% WP NN output

(d) 70% WP Significance   (e) 77% WP Significance   (f) 85% WP Significance
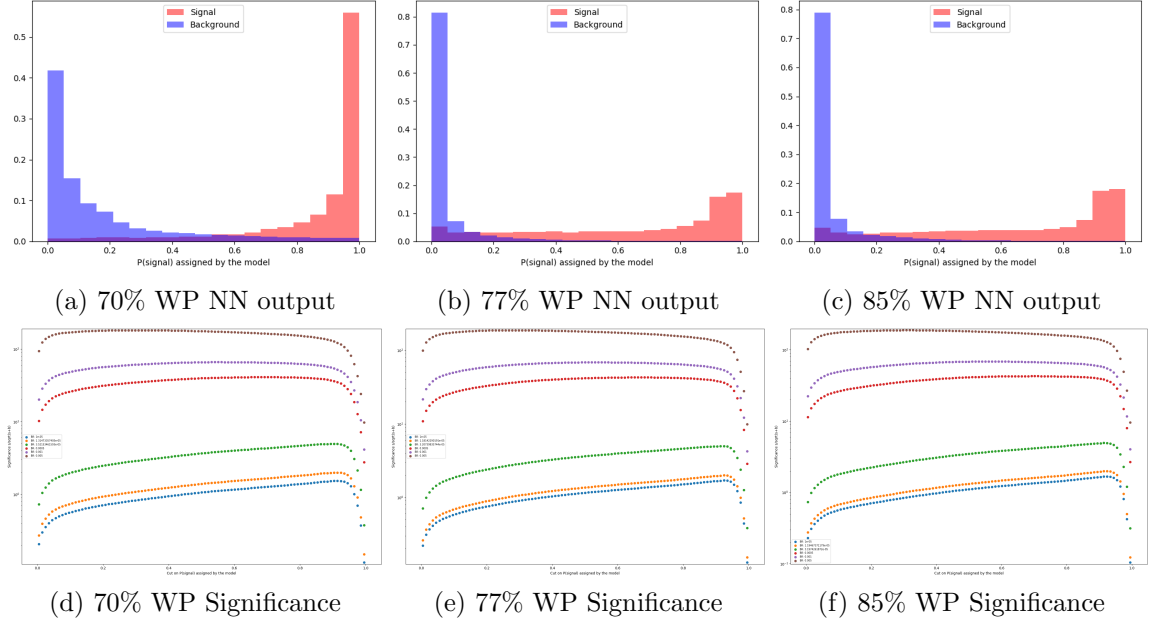
FIGURE 1.13. Neural network output and significance plots for the μ+jets channel at 70%, 77%, and 85% b-tagging working points.

1.4. Initial Event Selection

Initial event selection is done to ensure that events that are accepted into the analysis are not contaminated by extremely noisey detector environments and happened during times when the ATLAS detector was accepting events properly. All of the events have the same initial set of criteria for determining whether or not the event is looked at any further for this analysis, applying to both MC and Data. These initial checks are as follows:

- Only events occuring during runs good for physics

- Good Calorimeter status: Ensures that the LAr and Tile calorimeters are not experiencing a noise burst at the time of the event

- Requires a primay vertex to be reconstructed for the event which ensures timing of further reconstructed objects are placed with the correct vertex

- Global Trigger Decision: Selects events based on wheter they passed one of the triggers including the trigger thresholds, further discussion in Section 1.4.1

- Trigger Match: Select events where an electron or muon matches the trigger

- Overlap Removal as discussed in Section 1.1.1

- Ignore events that have a bad muon, occurs mostly in the transition region and the cathode strip chamber regions.

- Jet Cleaning: Removes events with jets formed from calorimeter information from sources that have nothing to do with the energy flow from the initial hard scatter interaction

These basic event selection values are applied to every event, in both MC and Data. On top of these various kinematic cuts are added to form the additional analysis level objects and regions used in the analysis. These additional kinematic cuts are examined more closely in Section 1.5 and in the discussion of kinematic region creation throughout the rest of the analysis e.g., Section 1.6.

1.4.1. Triggers

Different HLT triggers are used for data taking periods for each year of Run 2. This analysis takes advantage of single lepton triggers for electrons and muons to dramatically reduce backgrounds due to QCD events without leptons.

Year	p_T threshold [GeV]	Identification Menu	Isolation menu	L1 Seed
2015	≥ 24	Medium	None	L1EM20VH
	≥ 60	Medium	None	-
	≥ 120	Loose	None	-
	≥ 26	Tight	Gradient (Loose)	-
2016-2018	≥ 60	Medium	None	-
	≥ 140	Loose	None	-

TABLE 1.3. The electron trigger requirements in the event selections

Year	p_T threshold [GeV]	Identification Menu	Isolation menu	L1 Seed
2015	≥ 20	None	Gradient (Loose)	L1MU15
	≥ 50	None	None	-
2016-2018	≥ 26	None	Gradient (Medium)	-
	≥ 50	None	None	-

TABLE 1.4. The muon trigger requirements in the event selections

1.5. Data and MC Pre-Selection Cuts

The Signal Region pre-selection is defined to select events that have an opportunity to enter the final search selection. This pre-selection selects events with exactly one massive lepton, at least two jets (at least one of which is b-tagged at the 77% working point), transverse momentum and exactly one photon such that it resembles the expected final-state toplogy for the signal. All of the events have the same initial set of criteria for determining whether or not the event is looked at any further for this analysis, applying to both MC and Data. These initial checks are as follows:

- Exactly 1 lepton (electron or muon) $p_T > 25$ GeV

- At least two good jets ($p_T > 25$ GeV)

- At least one b-tag (MV2c10, 77% Working point)

- $\not{E}_T > 30$ GeV and $m_T^W > 30$ GeV (for events with electrons)

- $\not{E}_T > 20$ GeV and $\not{E}_T + m_T^W > 60$ GeV (for events with muons)

- Exactly 1 photon, $p_T > 50$ GeV

These plots are also produced before additional scale factors to account for mismodelling of various processes are taken into account. These include the fake

rate scale factors for processes where a truth electron or hadron are reconstructed to a photon and scaling to account for further mismodeling based on the order of the MC events produced (leading order, next-to-leading order, etc.). Only statistical uncertainties are shown

Signal photons, which originate from a top quark decay, are very high p_T whereas background photons typically result from soft processes. A cut on the photon candidate p_T removes much of the backgrounds while keeping a majority of the signal. The photon p_T in the preselection region is shown in Figure 1.14.

(a) electron channel   (b) muon channel
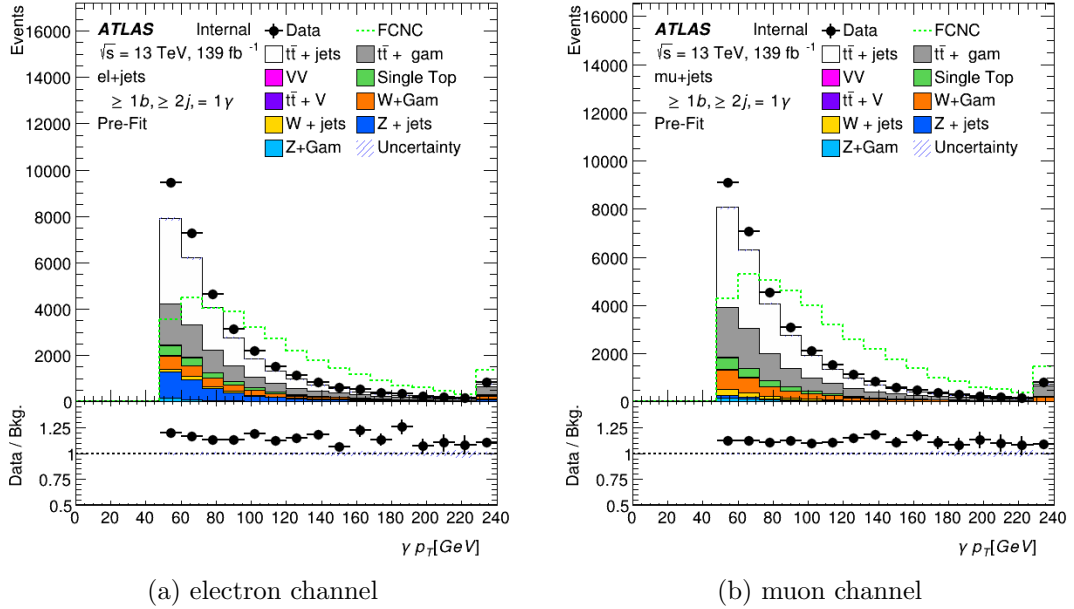
FIGURE 1.14. Photon p_T in the signal region pre-preselection region. FCNC signal branching ratio is scaled to 1%

Other variables of interest, i.e. those being used as inputs into the neural network are also showed in this section. Figure 1.15 shows the S_T and m_T^W distributions. Figure 1.16 shows the invariant mass distributions for both top quark candidates, m_{Wb} and $m_{q\gamma}$. The kinematic variables for the electron channel are shown in Figure

1.17 and for the muon channel in Figure 1.18. The neural network output of these events are shown in Figure 1.19

(a) electron channel

(b) muon channel

(a) electron channel

(b) muon channel
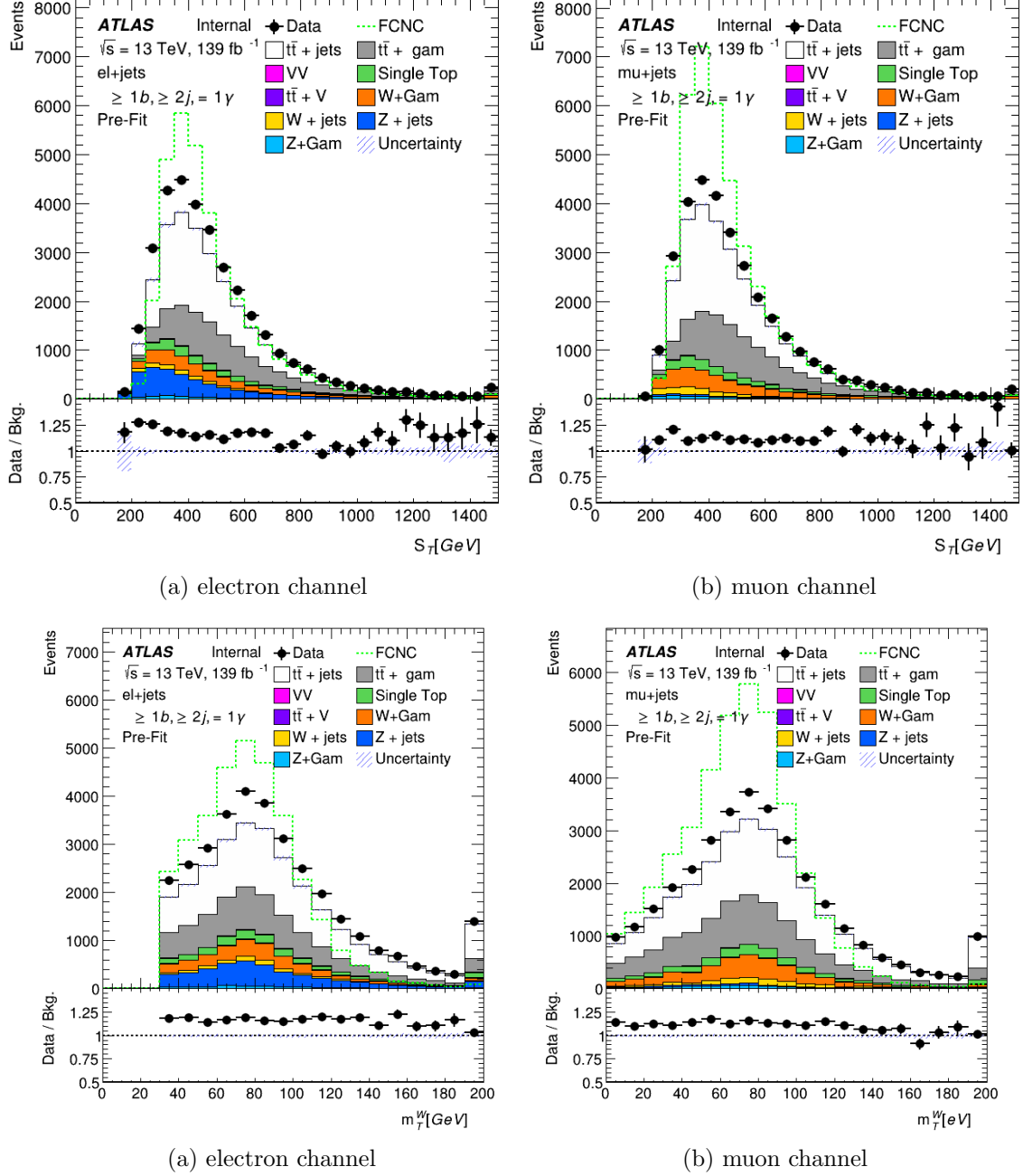
FIGURE 1.15. S_T and m_T^W in the signal region pre-selection region. FCNC signal branching ratio is scaled to 1%

(a) electron channel

(b) muon channel
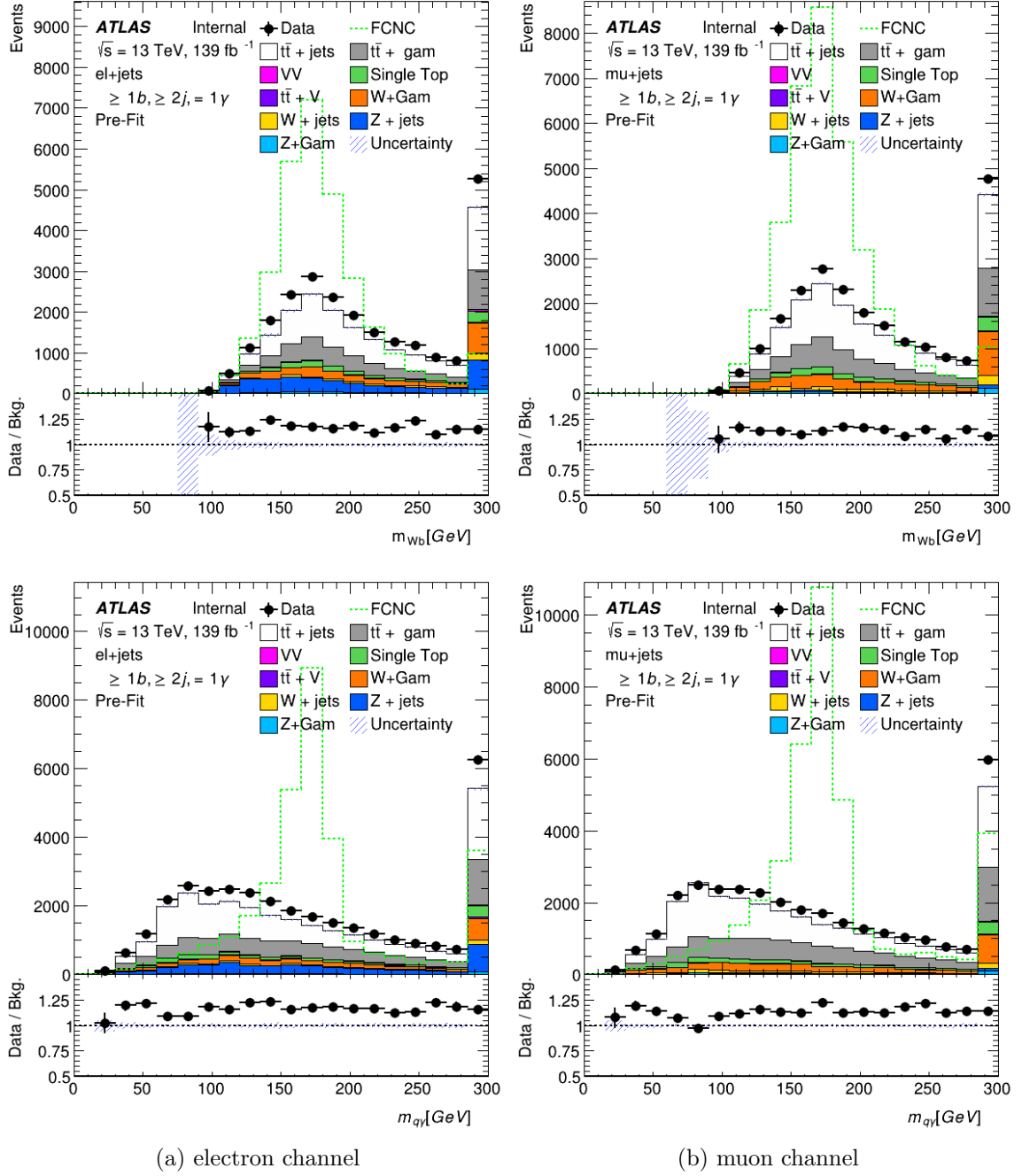
FIGURE 1.16. Top mass candidates in the signal region pre-selection: m_{Wb} and $m_{q\gamma}$. FCNC signal branching ratio is scaled to 1%
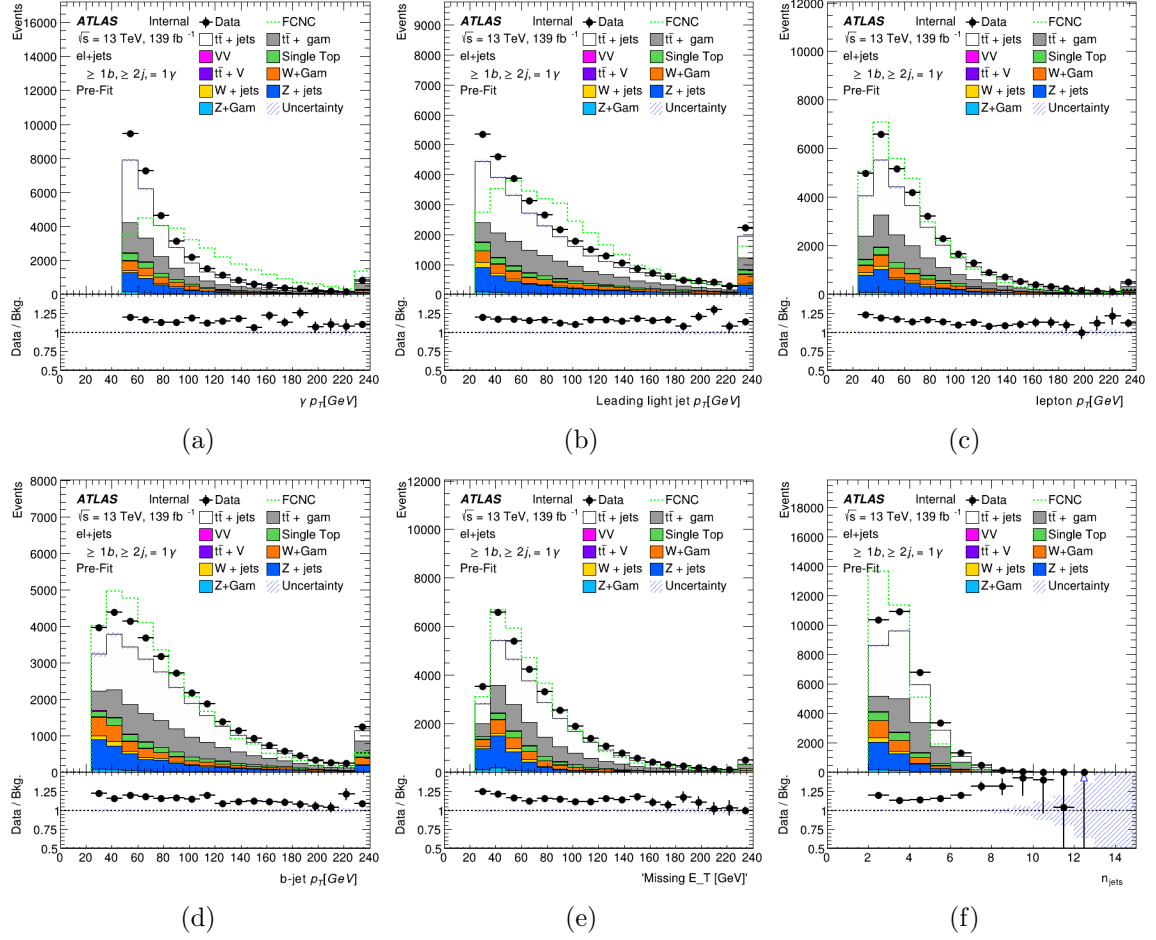
FIGURE 1.17. Photon p_T (a), leading light jet p_T (b), lepton p_T (c), b-jet p_T(d), $\not{E}_T$ (e), and n_{jets} (f) plots in the signal region pre-selection for the electron+jets channel. FCNC signal branching ratio is scaled to 1%

FIGURE 1.18. Photon p_T (a), leading light jet p_T (b), lepton p_T (c), b-jet p_T(d), $\not{E}_T$ (e), and n_{jets} (f) plots in the signal region pre-selection for the muon+jets channel. FCNC signal branching ratio is scaled to 1%

(a) electron channel　　　　　　　　(b) muon channel

FIGURE 1.19. Output of the Neural Network in the signal region pre-selection region. FCNC signal branching ratio is scaled to 1%

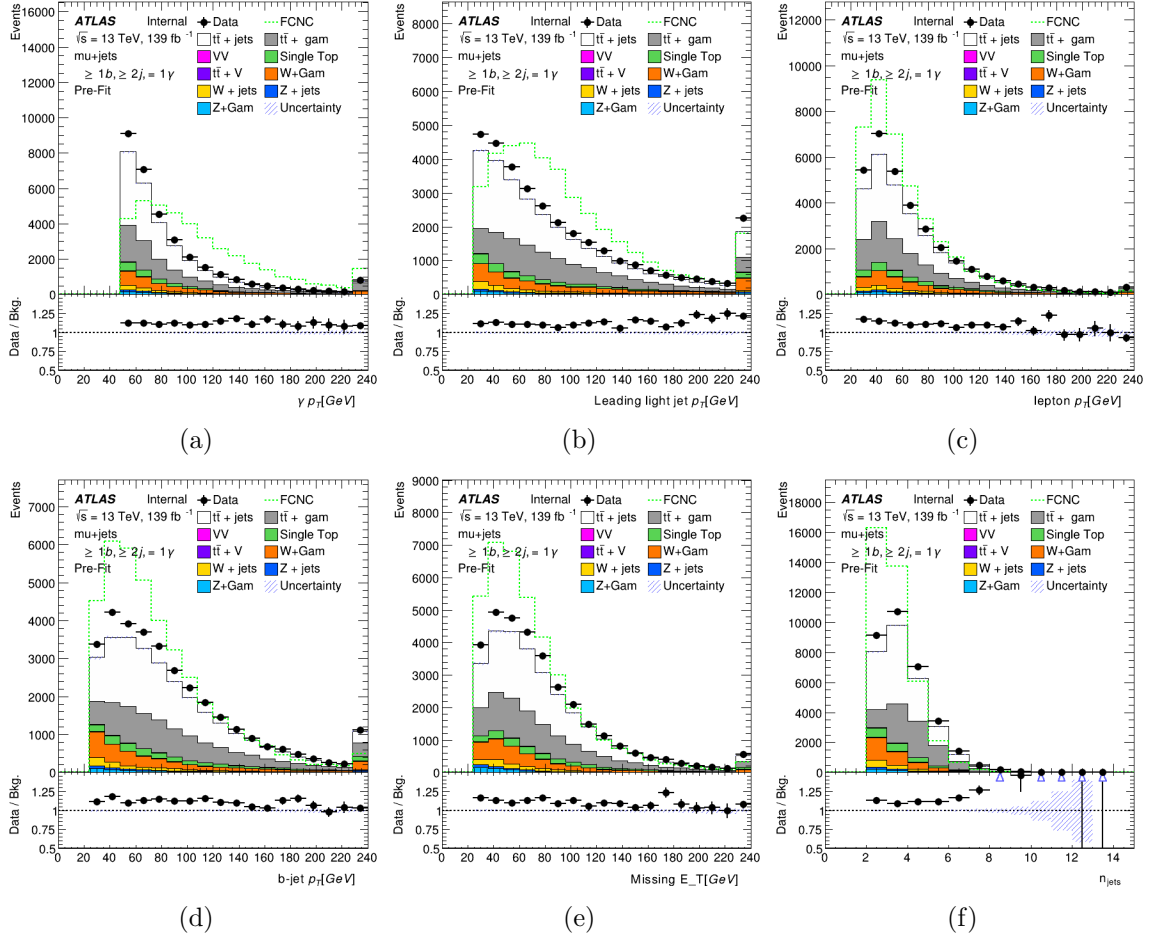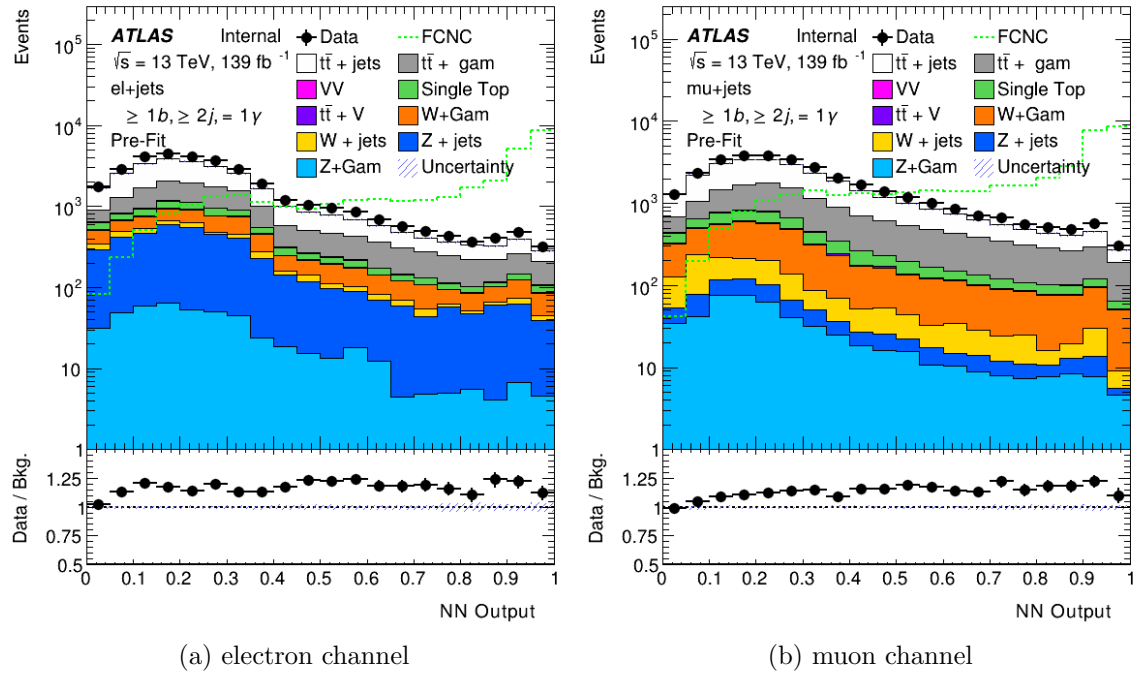## 1.6. Background Evaluation: Control and Validation Regions

Orthogonal regions to the signal region have been created to test the preformance of Monte Carlo samples. Control and validation regions are designed to isolate specific physics processes to determine and test the efficacy of scale factors that will be applied to the final signal region Monte Carlo evets. These control and validation regions need to be kinematically similar to the signal region such that derived scale factors can be translated directly into the signal region and orthogonal to make sure that there is little signal contamination in the regions. Regions have been made to test the major backgrounds expected in the signal region: $t\bar{t}$, W+jets, as well as events similar events produced with an associated photon: $t\bar{t} + \gamma$ and W+Jets+γ. Events without real photons are described in Section 1.6.1 and regions with a real photon are described in Section 1.6.2.

1.6.1. Backgrounds Without Photons

Various background processes that do not have a real photon produced in the events can still enter the signal region if an electron or jet is mis-reconstructed as a photon. Of these processes the largest contributors in the signal region are Standard Model $t\bar{t}$ and W+jets. As the LHC attains higher and higher energies the QCD multijet backgrounds become increasingly hard to model due to the non-perturbative nature of the interactions. A data-driven technique for accounting for these backgrounds is developed by scaling the major backgrounds to account for the QCD backgrounds that contribute extra jets to the major backgrounds. Designing a single control region satisfactorily close to the signal region is impossible. Thus, two control regions are designed, one which is W+jets rich and the other $t\bar{t}$ rich. Scale factors for these backgrounds are derived simultaneously and tested in a third

similar region for validation before being applied to other regions. These control and validation regions are defined as follows:

- All of the Initial Event Selection as outlined in Section 1.4

- Exactly 1 lepton (electron or muon) $p_T > 25$ GeV

- Number of Jets ($p_T > 25$ GeV) to define the regions

  * Control Region 1 (W+Jets enriched): $n_{\text{jets}} = 3$

  * Validation Region: $n_{\text{jets}} = 4$

  * Control Region 2 ($t\bar{t}$ enriched): $n_{\text{jets}} \geq 5$

- $\slashed{E}_T > 30$ GeV and $m_T^W > 30$ GeV (for events with electrons)

- $\slashed{E}_T > 20$ GeV and $\slashed{E}_T + m_T^W > 60$ GeV (for events with muons)

- Exactly 1 b-tagged jet (MV2c10, 77% Working point)

- 0 photons, $p_T > 15$ GeV

The efficiency of scale factors derived using control regions 1 (n_{jets}=3) and 2 ($n_{\text{jets}} \geq$5) are then tested in the validation region (n_{jets}=4). The scale factors for the $t\bar{t}$ and W+jets MC are derived using:

$$\begin{bmatrix} N(W)_{3j} & N(t\bar{t})_{3j} \\ N(W)_{5+j} & N(t\bar{t})_{5+j} \end{bmatrix} \begin{bmatrix} W_{SF} \\ t\bar{t}_{SF} \end{bmatrix} = \begin{bmatrix} N(\text{data-bkg})_{3j} \\ N(\text{data-bkg})_{5j} \end{bmatrix}$$

Figure 1.20 shows the S_T distribution in both electron and muon channels before scale factors are calculated for all three kinematically separate regions. The large mismodelling occurs at low S_T values as expected as QCD processes will typically
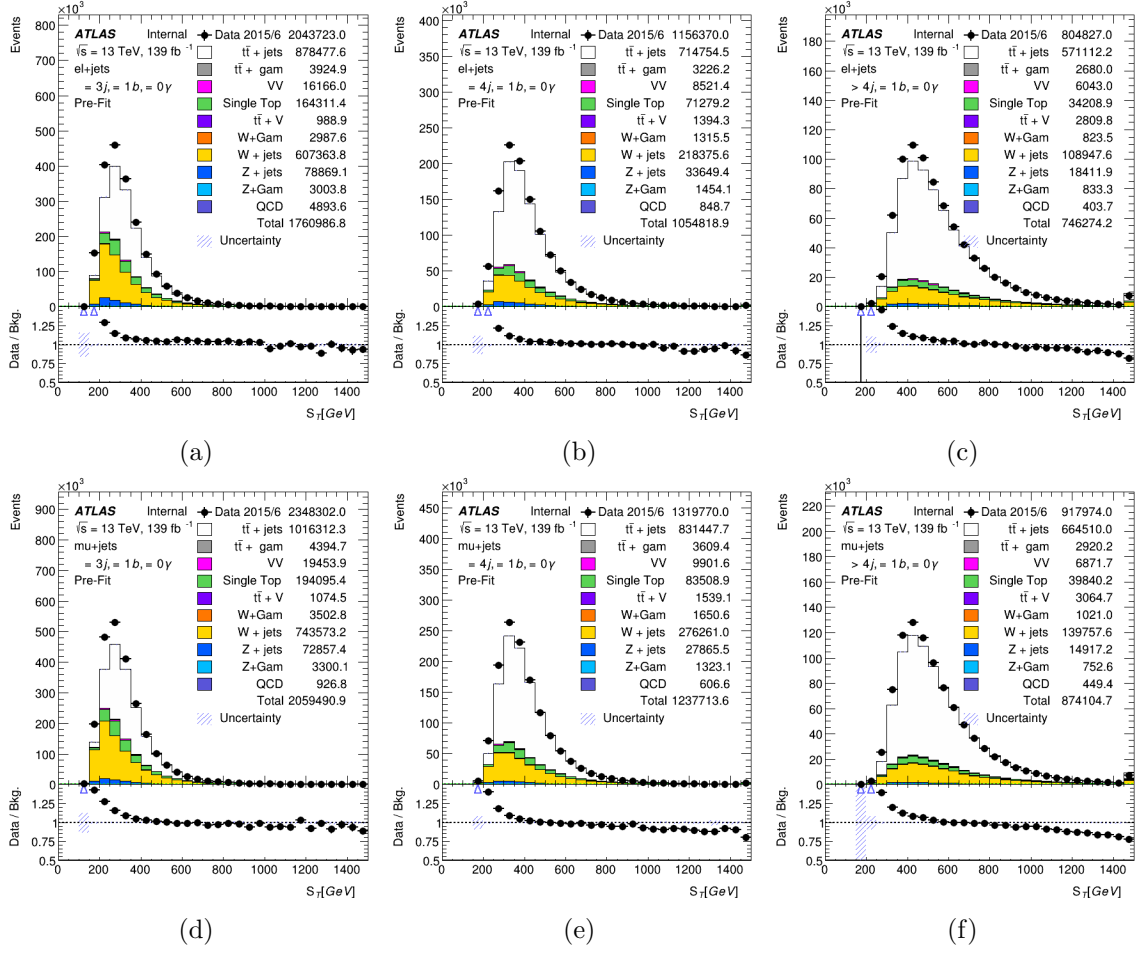
FIGURE 1.20. S_T distributions in the 3(a,d), 4(b,e), and 5+(c,f) jets control and validation regions. Electron channel is shown on the top and the muon channel on the bottom, before scale factors are determined.

add low energy jets to the events. The Figures 1.21(electron channel) and 1.22(muon channel) show various event-level variable plots for the validation region after the scale factors have been applied.

The derived scale factors using these regions are shown in Table 1.5

Sample	e+jets SF	μ+jets SF
W+jets	PUT IN VALUE	PUT IN VALUE
$t\bar{t}$	PUT IN VALUE	PUT IN VALUE

TABLE 1.5. Derived $t\bar{t}$ and W+jets scale factors for QCD multijet backgrounds.

(a) (b) (c)

(d) (e) (f)

FIGURE 1.21. Event-level plots for the =4 jet validation region after scale factors have been applied in the electron channel.
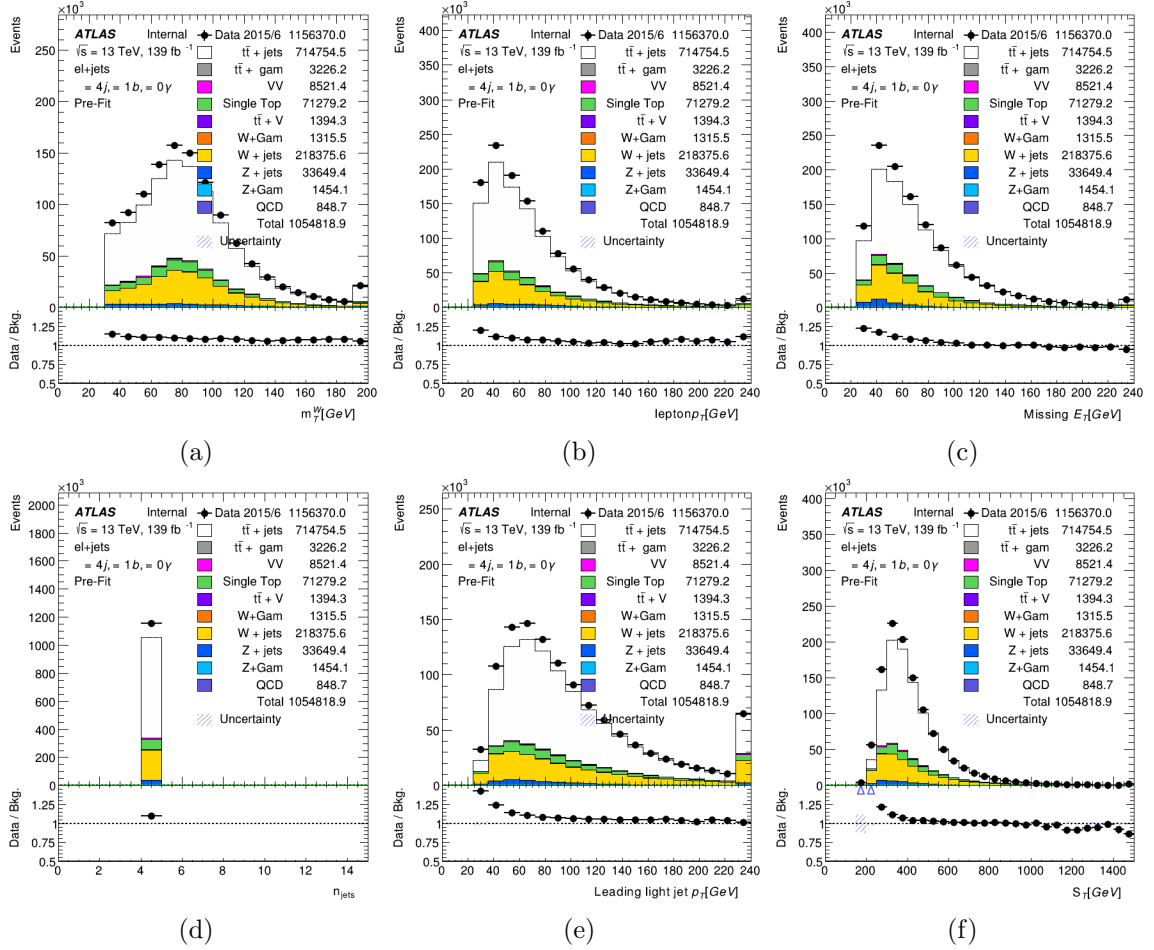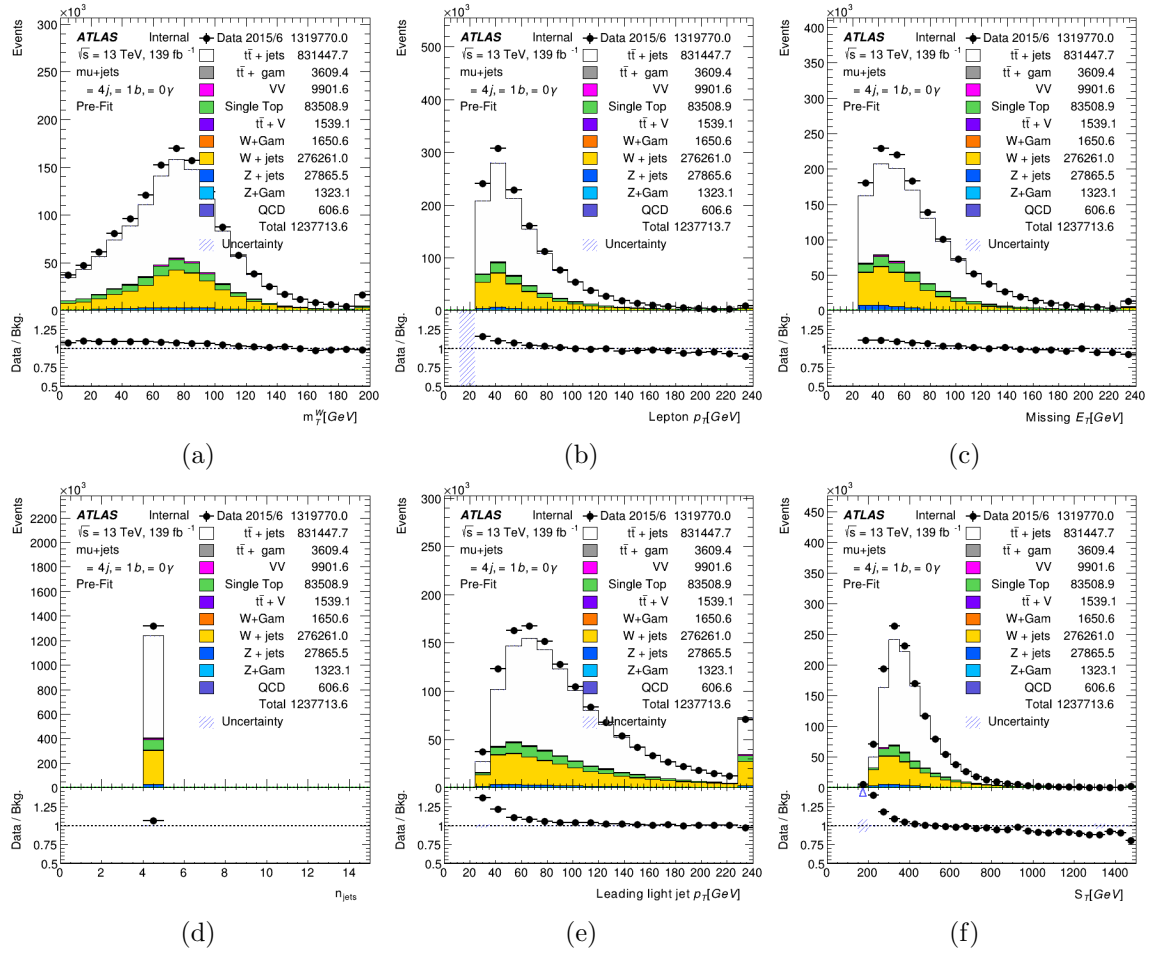
FIGURE 1.22. Event-level plots for the =4 jet validation region after scale factors have been applied in the muon channel.

1.6.2. Background With Photons

Standard Model processes that are produced with an extra real photon are irreducible background for this search as they can share the same final state as the signal events. The largest contributors of these irreducible backgrounds are the major background samples discussed in the previous section with an associated photon ($t\bar{t}+\gamma$ and W+jets+γ). Special Monte Carlo samples are produced for these samples (along with Z+jets+γ) that have higher statistics of these photon enriched events than the nominal samples. However, as these samples (X+jets+γ) are subsets of the nominal sample(X+jets) duplicate events must be removed from the nominal sample. This is done using the **MCTruthClassifier** tool which is detailed further in Section 1.7. All events with a photon from the hard scattering are removed from the X+jets samples as they are contained within the X+jets+γ samples.

1.6.2.1. W+γ Control Region

1.6.2.2. $t\bar{t} + \gamma$ Control Region

1.7. Fake Rates

Photons and leptons can be faked by various other particles in the detector depending on how they interact with the detector. Section 1.7.2 discusses how jets faking leptons are accounted for where Sections 1.7.1 and 1.7.3 discuss how electrons and jets can appear as photons and enter the signal region. Photon fake rates and scale factors are determined using data-driven techniques and truth information in the MC samples. Information in the truth record of the reconstruced photons is found using the **MCTruthClassifier** tool which preforms a *Truth to Cluster* matching

algorithm based on geometric separations, ΔR, between various truth-level physics objects.

A photon is considered as an electron fake if the truth particle ID is equal to the PDG ID of an electron or if the truth particle ID is equal to the PDG ID of a photon but a truth electron is within a distance of $\Delta R < 0.05$ as the photon is then assumed to have been originating from the electron.

A photon is considered a hadronic fake if the truth photon originates from any hadron or when the truth particle is a hadron. These hadrons can be any meson or baryon within the initial hard interaction.

Discussion of MCTruthClassifier stuffs

1.7.1. Electron $\rightarrow$ Photon Fakes

1.7.2. Jet $\rightarrow$ Lepton Fakes

The final state is determined by the decay of the W boson throughout this search, resulting in a charged lepton and a neutrino. However, this can be faked by the QCD multi-jet background if a jet is misidentified as a lepton. Non-prompt and fake leptons can pass the event selection and come from predominantly QCD multi-jet processes with an associated photon. These leptons can arise from semi-leptonic decay of bottom and charm quarks. The electron final state has additional contributions from photon conversions in the electromagnetic calorimeter while muons can be faked from highly energetic hadrons that punch-through to the hadronic calorimeter before beginning their shower leaving only tracks similar to muons throughout the electormagnetic calorimeter.

The estimation for the lepton fake background in the signal region channels follows the fully data-driven Matrix Method[**?**] approach which compares the number

of leptons with looser lepton idenification and isolation requirements than those used in the search. Electrons in the loose sample are required to have *Medium* identification criteria but no isolation requirement is applied. Similarly for muons the idenification criteria is *Medium* and the requirement on the isolation is dropped. The Matrix Method works under the assumption that the tight sample will contain mostly real leptons whereas the loose sample will be enriched with fake lepton events. Therefore it follows that the number of leptons in the loose (tight) sample is simply a combination of the number of real and fake leptons in those samples:

$$N^{\text{loose}} = N^{\text{loose}}_{\text{real}} + N^{\text{loose}}_{\text{fake}}$$

$$N^{\text{tight}} = N^{\text{tight}}_{\text{real}} + N^{\text{tight}}_{\text{fake}}$$

Then ϵ_{real} is defined as the probabilty of a real lepton in the loose sample to pass the tight selection and ϵ_{fake} as the probability of a fake lepton in the loose sample passing the tight selection. Then the number of events with fake leptons in the tight sample can be estimated applying the following equation:

$$N^{\text{tight}}_{\text{fake}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}}(\epsilon_{\text{real}} N^{\text{loose}} - N^{\text{tight}})$$

The real efficiencies, ϵ_{real}, are estimated using the tag-and-probe method in $Z \rightarrow e^+ e^-$ and $Z \rightarrow \mu^+ \mu^-$ regions and the fake efficiencies, ϵ_{fake}, are estimated in data samples which are dominated by non-prompt and fake leptons. The efficiency measurements are provided centrally in ATLAS and are explained in much more detail in [?]. Individual weights are applied to the event identification caluclated as

a function of these efficiencies which follow:

$$w_i = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}}(\epsilon_{\text{real}} - \delta_i)$$

where δ_i is 1 if the event passes the tight selection and 0 otherwise. These event weights are applied to the loose data sample to estimate the fake contributions from jets faking leptons.

1.7.3. Jet $\rightarrow$ Photon Fakes

1.8. Signal Region

REFERENCES CITED

[1] F. Chollet et al. Keras: The python deep learning library, 2019.

[2] Martín Abadi et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016.

[3] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. 2014.

[4] A. Karpathy. Cs231n convolutional neural networks for visual recognition, 2015.

[5] Fabian Pedregosa et al. Scikit-learn: Machine Learning in Python. *J. Machine Learning Res.*, 12:2825–2830, 2011.

[6] The ATLAS collaboration. Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. 2014.