Search for flavour-changing neutral currents in top pair events with an associated photon using $\sqrt{s} = 13$ TeV with the ATLAS detector

Jason Barkeloo[a] and Jim Brau[a]

[a]*University of Oregon*

This note presents the search for flavour-changing neutral currents in processes with top pairs where one top decays through the flavour-changing neutral current decay mode (to an up type quark and a photon) and the other decays through the most common standard model mode to a b-quark and a W boson. The W boson then decays leptonically defining the channels searched (electron+jets and muon+jets). Using the entire Run-2 data set of $\sqrt{s}$ = 13 TeV data collected using the ATLAS experiment between 2015 and 2018 corresponding to a total integrated luminosity of 139 fb^{-1}. A neural network was developed to separate signal and background events using both low level kinematic variables and high level variable combinations as inputs. The signal events contain an isolated very high p_T photon, a lepton, a b-tagged jet, another light jet, and missing transverse energy. Various data driven techniques were used to estimate contributions to the background from events with a hard scatter photon or a photon faked by either a jet or an electron. As there is no sensitivity for observation an upper limit on the branching ratio of this process is set for each channel and a combination of the channels.

Contents

1 Change Log

Version 0.1

- First Draft

2 Introduction

The Standard Model of particle physics has proven itself an exceptional and resilient theory since the combination of the electromagnetic and weak interactions in 1961[?]. Further theoretical work combined the Higgs Mechanism[?, ?] with the electroweak theory[?, ?]. The resiliency of this theoretical model has been tested to further degrees of accuracy over the decades with one of my most recent portions being the experimental confirmation of the Higgs Boson in 2012[?, ?] using the Large Hadron Collider (LHC). Further precision measurements are ongoing at various experiments at the LHC, including the ATLAS experiment.

However, the Standard Model is known to have flaws and disagreements with nature. For example, the Standard Model predicts massless neutrinos which is in conflict with experimental observation of neutrino flavor oscillation and doesn't provide an explanation for dark matter particles or their interactions with currently known particles. While these large gaps in the Standard Model are well known every precision measurement made has yet to give any significant new hints toward physics beyond the Standard Model. One new pathway to look for these hints at the LHC is through top quark decays.

The top quark was first observed at Fermilab's Tevatron in 1995[?] but the increase in energy and amount of data at the LHC has produced orders of magnitude more top quarks than previously seen opening up a pathway to precision measurements of the properties of the top quark. The top quark is the heaviest fundamental particle with a mass of 172.51 ± 0.27 (stat) ± 0.42 (syst)[?]. This large mass also means that top top quark lifetime is very short ($5 * 10^{-25}$ s) and decays before it can hadronize. This allows us to probe its' branching ratios and decay modes directly. The Standard Model predicts that the top quark decays through the charged current mode nearly 100% of the time; t→ qW (q = b,s,d)[?]. The Standard Model also predicts a rare branching ratio of the top quark through a flavor changing neutral current (FCNC) process, to a neutral boson (photon, Z boson, Higgs Boson, or gluon) and up type quark with a branching ratio on the order of 10^{-14}.

Precision measurements are an important litmus test for the Standard Model. Predicted branching ratios for FCNC processes in top quark decays are far beyond the experimental reach of the LHC and any observation of these decay modes would be a sure sign of new physics. Branching ratios are an important measurement due to a litany of theories for new physics beyond the Standard Model (BSM). These BSM theories predict enhancements in the top sector by many orders of magnitude such as Minimal Supersymmetric models[?], R-parity-violating Supersymmetic models[?] and two Higgs doublet models[?] can all increase this branching ratio many orders of magnitude. Even a null result to a search will set an upper limit on the branching ratio that can assist in ruling out future physical models based on their amount of large top sector enrichment.

This note presents the ATLAS search for top FCNCs with an associated photon in the decay mode of $t\bar{t}$ events using the entire Run 2 dataset at the LHC, containing combined 2015-2018 datasets taken by the ATLAS experiment totaling 139 fb^{-1} of integrated luminosity taken at $\sqrt{s}$= 13 TeV. The analysis presented looks for an excess of events coming from top quark pair produced events where one top quark decays to the most likely decay mode (a bottom quark and W boson) and the other to an up type quark (up or charm) and a photon. The W boson decay determines the final state channels looked for, two separate channels are considered: the electron+jets and muon+jets channels.

Section ?? a theoretical background will be presented for both the Standard Model with a closer view on the usual extensions to include the FCNC vertices. Following this Section ?? will discuss the LHC and the ATLAS experiment used in the creation of the dataset used in the analysis. In Section ?? the special

signal simulation requirements will be presented as well as the common background event simulation methodology. The search strategy including the creation of signal, control and validation regions and the training of a neural network will be examined in Section **??**. Section **??** will discuss the results and the conclusions drawn from these results will be presented in Chapter **??**.

3 Object Reconstruction

After the events are simulated, or collected in case of real data, the collections of energy deposits within the detector systems must be transformed into meaningful physics objects through reconstruction. Reconstruction is typically done in two major parts using the specialized detectors covered in Chapter **??**. The Inner Detector and Muon System turn patterns of hits within the tracking detectors into tracks that have direction and momentum information. The calorimeter system transforms the energy deposits within the calorimeters into calibrated energy deposits with a particular position. These tracks and calorimeter deposits are used to create physics objects (electrons, muons, etc.) by using particle identification techniques to reconstruct the underlying physics event. For the analysis presented in this dissertation, the final state signal particles that need to be reconstructed are one lepton (an electron or a muon), one photon, two quarks (one light flavor and one b quark), and one neutrino (missing transverse energy as it is the only particle that does not interact with the detector). Each of these particles has a particular signature in the subdetectors of the ATLAS detector, shown in Figure **??**.
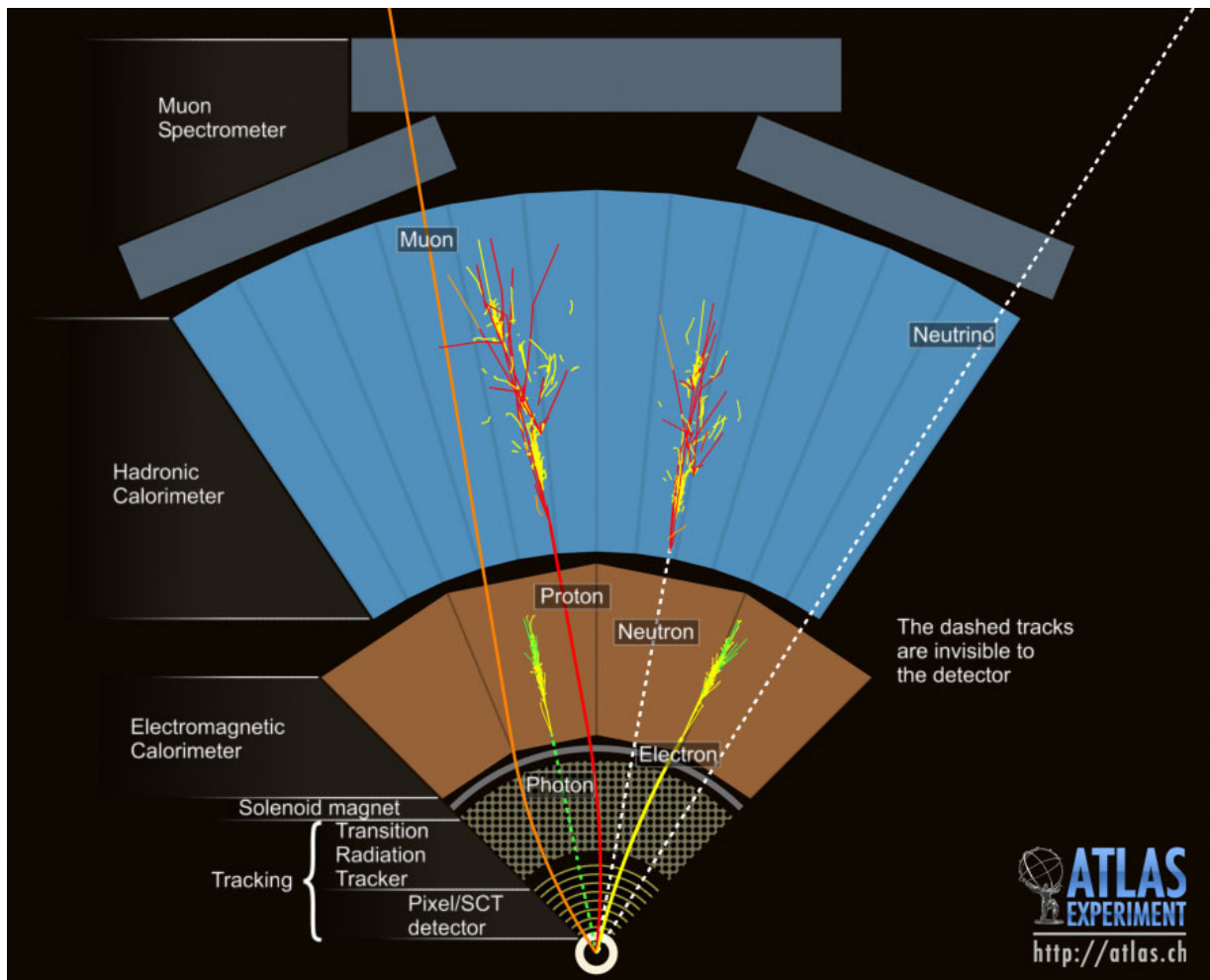

Figure 1: Cross section of a simulated ATLAS detector showing how various particles interact with ATLAS subsystems. Solid lines indicate interactions while dashed lines indicate that no interactions typically occur in that section of the detector. [**?**]

3.1 Electrons

Electrons interacting with within the ATLAS detector leave a track in the Inner Detector as well as a cluster of energy in the electromagnetic calorimeter. The track and cluster are required to be matched together to be identified as an electron candidate[?]. As electrons move through the detector they create electromagnetic showers through bremsstrahlung which can produce electron-positron pairs. The process continues as the particles continue to give energy to the detector. This collection of electrons, positrons, and photons creates a signature energy cluster in the calorimeter.

Electron identification algorithms are applied to the electron candidates to separate prompt and isolated electron candidates from electrons that come from backgrounds such as converted photons and misidentified jets. The electron idenification algorithms use a sliding window (3×5 in $\eta \times \phi$) in the high granularity section of the LAr electromagnetic calorimeter to search for electron cluster "seeds" greater than 2.5 GeV. Clusters are created around these seeds to form the electromagnetic shower and remove possible duplicate electron signals by containing them within the cluster. Further pattern recognition for the track fitting allows even larger amounts of energy into the shower to account for bremsstrahlung in the shower shape. Tracks and clusters are then matched to give electron candidates.

Electrons coming from background jets or photon conversion are called non-prompt as they do not originate from a signal object/the primary vertex. In order to reject these electrons, other discriminating variables are used in addition to the track-cluster matching. These variables inclue the amount of energy leakage into the hadronic calorimeter, thee shower development throughout the electromagnetic calorimeter, and the amount of radiation measured in the TRT. Three electron identification working points are used: Loose, Medium, and Tight. Each of these operating points have their own level of background rejection and signal efficiency. Working points with higher background rejection are a subset of those with lower background rejection.

Isolation variables are another useful tool in the identification of signal electrons from converted photons produced in hadron decays and light hadron misidentification. These variables are defined by a cone size around the electron candidate and are the sum of the transverse variable (momentum or energy) of all of the tracks within the cone, $p_T^{\mathrm{cone0.2}}$ with a cone of $\Delta R = 0.2$ (or 10 GeV/E_T, for high energy electrons) and $E_{T,Topo}^{\mathrm{varcone0.4}}$ with a cone defined in a similar manner.

Because the LAr calorimeter is a sampling calorimeter, the energy deposits must be calibrated and scaled such that the true electron energy is read out and not just the small amount of energy deposited into the active layers as discussed in Section ??. The energy scale is calibrated to be uniform throughout the detector. Any residual differences between data and simulation are corrected. The calibration strategy was developed for optimal preformance in LHC Run 1[?] and updated for the conditions of LHC Run 2[?].

3.2 Muons

Muons behave differently from other particles as they traverse the detector. They act as minimum-ionizing-particles (MIPs) throughout the calorimeter. The Muon Spectrometer (MS), Section ??, specializes in precision measurements of muons. The Inner Detector (ID) plays a pivotal role in the identification of muons as it offers an independent measure of the muon characteristics. The muron reconstruction process uses a specific set of variables as well[?]. These variables include: *q/p significance*: the difference in the ratio of track charge and momentum measured with the ID and MS, ρ': the difference between the

transverse momenta measured with the ID and MS, and χ^2 of the combined track fit using tracks from both the ID and MS.

Muons are separated out into four separate types depending on their interactions with the various subdetectors. The best muon candidates are combined muons that use hits in the MS to trace back to a track in the ID in order to reconstruct the entire muon track. Segment-tagged muons are muon candidates that leave a track in the ID but only a segment in the MS instead of a full track. Segment-tagged muons can occur because of the muon having low p_T or crossing through a region of the MS with reduced acceptance. Extrapolated muons require only tracks in the MS and are used in regions of η, ϕ phase space that the ID does not cover. Calorimeter-tagged muons are muons identified by MIPs in the calorimeters and are used to find muons that cross the ID and MS in regions where cabling might prevent particle detection.

Muons also have their own set of isolation criteria which is track-based $p_T^{\text{varcone0.3}}$, with a cone of $\Delta R = \min(0.3, 10 \text{ GeV}/p_T)$. Similar to electrons various working points are available at the analysis level for muons. These working points are named similarly: Loose, Medium, Tight, and High-p_T in order of background rejection.

High p_T jets that punch through the hadronic calorimeter can leave tracks in the MS which could be identified as muons. These would be identified as a bad or a fake muon because of the high-hit multiplicities they leave in the MS as opposed to a single track left by a muon as it is a MIP. Another source of bad muons is a mismeasured ID track that gets incorrectly matched to segments in the MS. Fake muons are a source of fake missing transverse energy, $\slashed{E}_T$

3.3 Photons

Photons behave very similarly to electrons in the calorimeter in that they also produce an electromagnetic shower in the calorimeter. However, they are neutrally charged particles meaning that they should not leave a track in the ID as they do not bend and produce bremsstrahlung photons traveling through the magnetic field. Prompt photons pair-produce electrons in the tracker, but this process can be identified as the associated cluster in the electromagnetic calorimeter is matched to two tracks with opposite charge. This process produces what is called a converted photon. Unconverted photons have no matching tracks associated with an electromagnetic cluster.

Prompt photons produce narrower energy deposits in the electromagnetic calorimeter and have smaller leakage into the hadronic calorimeter compared to background photons. The energy contained within narrow structure in $\eta \times \phi$ strips compared to the energy containd in a larger section can help identify prompt from non-prompt photons [?]. Cuts on this and the other variables listed in Table ?? are tuned to reduce dependency of identification efficiency on the pileup conditions of Run 2.

3.4 Jets

Contrasting with electromagnetic showers produced by electrons and photons, hadronic showers form through QCD processs. Quarks very quickly undergo showering by emitting gluons which further produce quark-antiquark pairs, analogous to the photons and pair-produced electron-positron pairs of

Category	Description	Name	*loose*	*tight*
Acceptance	$|\eta| < 2.37$, with $1.37 \leq |\eta| < 1.52$ excluded	-	✓	✓
Hadronic Leakage	Ratio of E_T in the first sampling layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < |\eta|$ or $|\eta| > 1.52$)	R_{had_1}	✓	✓
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < |\eta| < 1.37$)	R_{had}	✓	✓
EM Middle Layer	Ratio of the energy in 3×7 $\eta \times \phi$ cells over the energy in 7×7 cells centered around the photon cluster position	R_η	✓	✓
	Lateral shower width, $\sqrt{(\Sigma E_i \eta_i^2)/(\Sigma E_i) - ((\Sigma E_i \eta_i)/(\Sigma E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	ω_{η_2}	✓	✓
	Ratio of the energy in 3×2 $\eta \times \phi$ strips, over the energy of 3×6 cells centered around the photon cluster position	R_ϕ		✓
EM Strip Layer	Lateral shower width, $\sqrt{(\Sigma E_i (i - i_{max})^2)/(\Sigma E_i)}$, where i runs over all strips in a window of 3×2 $\eta \times \phi$ strips, and i_{max} is the index of the highest-energy strip calculated from three strips aroudn the strip with maximum energy deposit	$\omega_{s\,3}$		✓
	Total lateral shower width $\sqrt{(\Sigma E_i (i - i_{max})^2)/(\Sigma E_i)}$, where i runs over all strips in a window of 20×2 $\eta \times \phi$ strips, and i_{max} is the index of the highest-energy strip measured in the strip layer	$\omega_{s\,tot}$		✓
	Energy outside the core of the three central strips but within seven strips divided by energy within the three central strips	f_{side}		✓
	Difference between the energy associated with the second maximum in the strip layer and the energy reconstructed in the strip with the minimum value found between the first and second maxima	ΔE_s		✓
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in the secondary maximum in the cluster to the sum of these energies	E_{ratio}		✓
	Ratio of the energy in the first layer to the total energy of the EM cluster	f_1		✓

Table 1: Photon identification variables used for *loose* and *tight* photon identification, taken from [?]

163 electromagnetic showers. When quarks have enough energy they hadronize by producing bound states of
164 particles. These particles are typically pions or mesons that are measured by the ATLAS detector. The
165 top quark is the only quark that decays before hadronization because it decays so fast (5×10^{-25} s). The
166 spray of hadrons coming from a quark from the initial interaction is called a jet and is a collection of
167 detector objects that are traced back and assigned to the quark(s) in the final state of the interaction. These
168 algorithms are called jet-finding algorithms. Pictoral representations of the same event reconstructed with
169 four various algorithms is shown in Figure **??**.
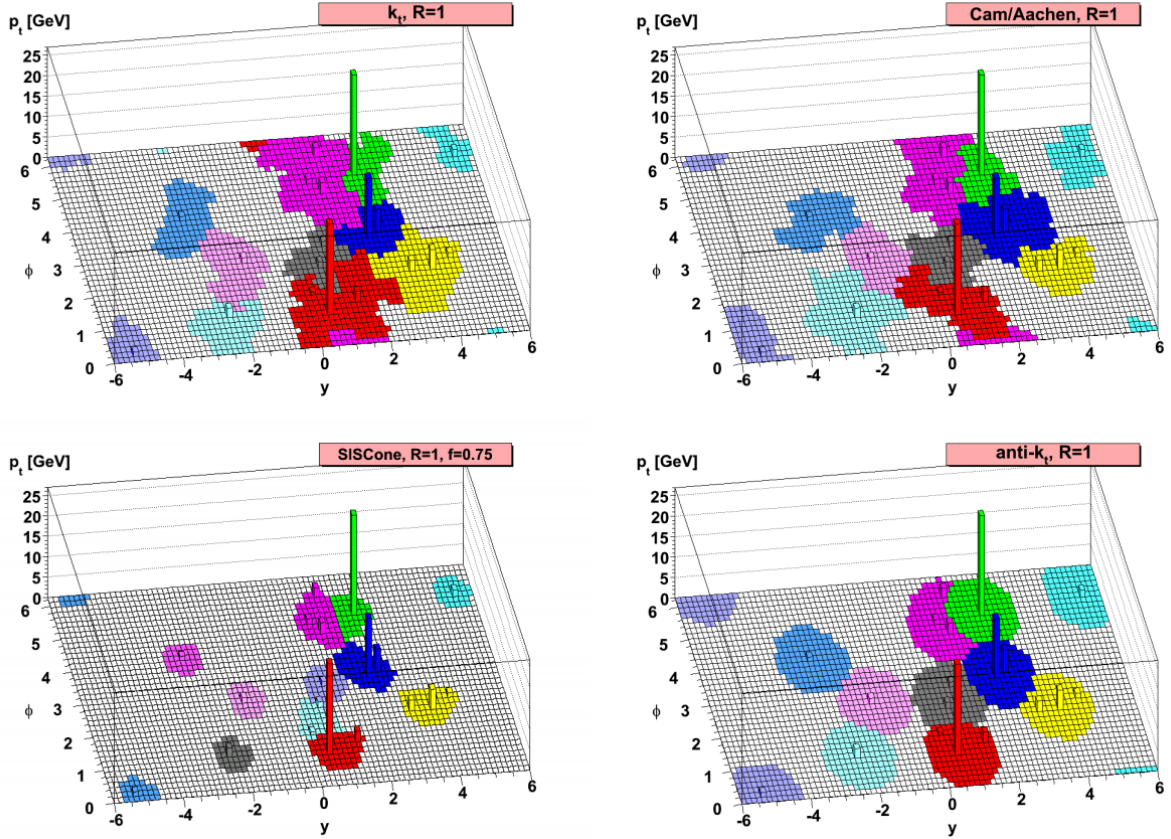
Figure 2: A sample parton-level event with many random soft jet objects, clustered with four different jets algorithms, illustrating the areas of the resulting hard jets. For kt and Cambridge/Aachen the detailed shapes are in part determined by the specific set of ghosts used, and change when the ghosts are modified [**?**]

170 The jets in this analysis use the anti-k_T algorithm [**?**] with a radius parameter $R = 0.4$. Jets are not physical
171 objects but collections of clustered particles so how they are defined can change the physics objects that are
172 eventually analyzed. The anti-k_T algorithm is preferred because it is infrared and collinear safe. Infrared
173 safe jet algorithms do not merge two jets with a soft emission between them. Adding or removing a
174 soft term between two jets should not change which objects are called jets. Collinear safe jet algorithms
175 do not change the jet collection if the high transverse momentum particles is split or merged . Another
176 added benefit of the anti-k_T jet finding algorithm is that it produces roughly circular jet objects, thereby
177 simplifying the calculation of the energy density and simplifying the calibration of the jet object.

178 The anti=k_T algorithm calculates the distance between an object i and all possible jet objects j (d_{ij}) and

179 the beam (d_{iB})

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R}, \qquad d_{iB} = k_{ti}^{2p}$$

180 where k_{ti} is the transverse momentum, Δ is the distance between the objects, and $p = -1$. This is a general
181 form for the type of algorithm where the inclusive k_T algorithm has a p value of 1 and the inclusive
182 Cambridge/Aachen algorithm has a p value of 0 [?]. The algorithm then follows that if d_{ij} is smaller than
183 d_{iB} then objects i and j are merged, otherwise i is labeled as a jet and removed from the list of entries of
184 possible jet objects. This is repeated for all entries in the list of possible jet objects.

185 Jet cleaning is also applied to remove events with jets built from known noisy parts of the calorimeter due
186 to particular calorimeter cells or non-collision background in those areas [?]. To reduce selecting jets that
187 originate from pileup interactions, another requirement on the jet object is made on the jet vertex tagger
188 [?, ?] as follows:

189 1. For jets with 20GeV $< p_T <$ 60GeV and $|\eta| < 2.4$: if any jet is bad AND that jet is not marked as
190   pileup by JVT, then reject the event

191 2. For jets with 20GeV $< p_T <$ 60GeV and $|\eta| \geq 2.4$: if any jet is bad, then reject the event

192 3. For jets with $p_T \geq$ 60GeV: if any jet is bad, then reject the event

3.4.1 B-Jets

194 While jets originate from any quark, jets coming from b quarks can be identified due to their decay products.
195 B quarks hadronize into b-hadrons which have a relatively long lifetime compared to many other hadrons
196 produced from light quarks. The longer lifetime and the relativistic speeds at which the hadrons travel
197 mean the particle travels a measureable distance before it decays ($400 - 500\mu m$)[?]. Thus, the vertex
198 reconstructed from the energy coming from a b hadron decay can be traced back to a point that does not
199 correspond to the primary vertex of the event. A pictoral representation of a b quark decay is shown in
200 Figure ??. The b-jet vertex is called the secondary vertex.

201 In addition to the secondary vertex, other variables are helpful in identifiying jets coming from b quarks.
202 By back tracing the tracks within the displaced vertex the minimum distance between the track and the
203 interaction point can be measured, known as the impact parameter. Reconstructing the decay chain of the
204 jet is also used in determining the providence of the jet. This information is used in a multivariate analysis
205 (MVA) to identify jets coming from b quarks and reject jets coming from light quarks.

206 The MVA used in this analysis is the MV2c10, the discriminant used for b-jet identification [?]. The output
207 distributions for various flavors of jets as well as background rejection and signal efficiency plots are
208 shown in Figure ??. The c10 in the algorithm name refers to the background training sample of the MVA
209 consisting of a 10% fraction of c-jets. The 77% efficiency fixed-cut working point for b-jet identification
210 was chosen for this analysis, discussed in Section ??. Differences in efficiency of b-tagging between data
211 and simulation is taken into account with working point specific scale factors provided by the ATLAS
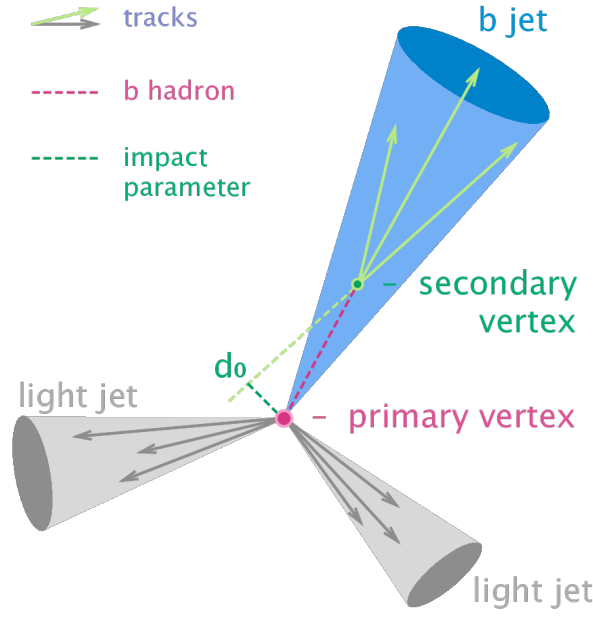212 Flavour Tagging Combined Performance group.

Figure 3: Pictoral representation of an event with a b-jet showing the secondary vertex and impact parameter [**?**]
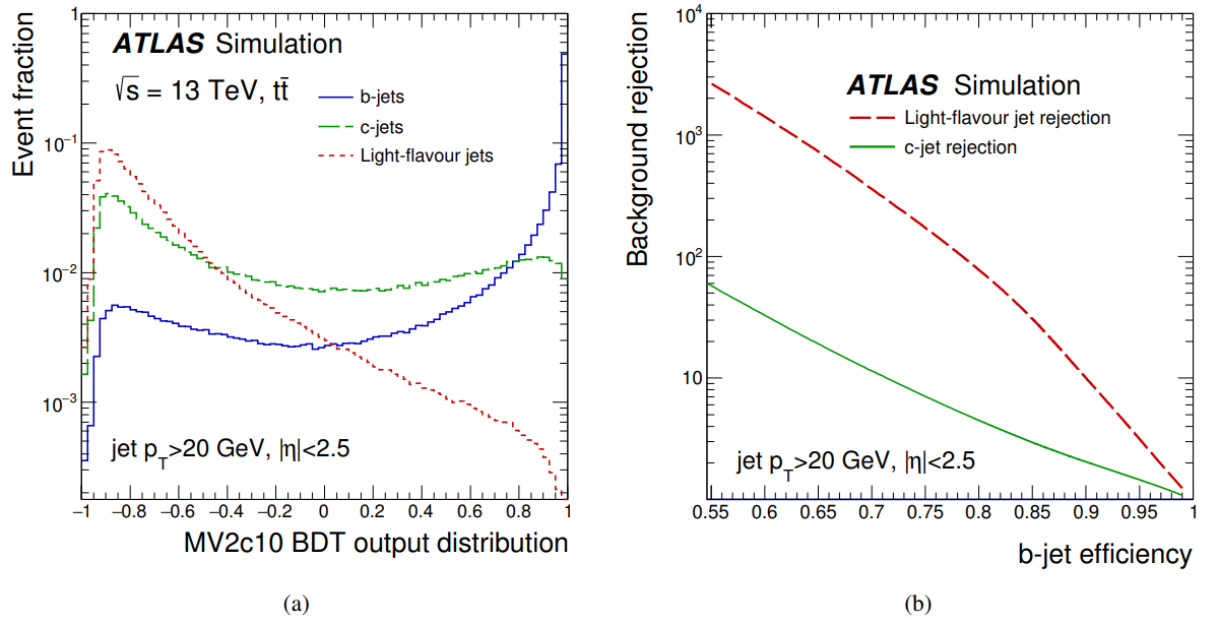

Figure 4: The MV2c10 output for b, c, and light flavored jets in simulated $t\bar{t}$ and the background rejection as a function of the b-jet efficiency [**?**]

3.5 Missing Transverse Energy

The remaining signal object that has yet to be discussed is the neutrino coming from the W boson decay. Neutrinos do not interact with the detectors as they pass through the ATLAS detector. The only way to measure any properties of the neutrino in ATLAS events is to use conservation of momentum. As previously mentioned the collision energy is unknown as partons do not carry a consistent fraction of the

beam proton energy. However, in the transverse plane to the beamline the total momentum is known to be very small. Before the collision there is on the order of 1 GeV of momentum in the transverse plane. Therefore, the total transverse momentum of the collision products should be approximately zero.

Any imbalance in the momentum is referred to as Missing Transverse Momentum ($\not{E}_T$). The negative vector sum of all reconstructed objects plus an additional soft term are used to calculate the missing energy in the x-plane and the y-plane[?]. A magnitude and an aziumuthal angle are calculated to give the $\not{E}_T$ vector in the transverse plane but this does not directly correspond to a neutrino which also has a momentum in the z direction.

3.5.1 Neutrino Reconstruction

In this analysis the signal contains only one source of missing energy, therefore all of the missing energy can be used to reconstruct a neturino object. There is an ambiguity in the choice of the neutrino z-momentum. To find the z-momentum a χ^2 minimization is done:

$$\chi^2 = \chi^2_{\text{SMTop}} + \chi^2_{\text{W}}$$

$$\chi^2 = \frac{(m_{\text{bjet},l,\nu} - m_t)^2}{\sigma^2_{\text{SMTop}}} + \frac{(m_{l,\nu} - m_W)^2}{\sigma^2_W}$$

The widths $sigma_{\text{SMTop}}$ and σ^2_W are determined from signal Monte Carlo. The event objects are combined to calculate the invariant mass of the top quark (the combination of the b-jet, lepton, and neutrino) and the W boson (combination of the lepton and neutrino). The χ^2 minimization is is done while varying the z-momentum of the neturino. The neutrino momentum that corresponds to the smallest χ^2 value is given assigned to the neutrino object for further use in the analysis. The χ^2 values are also used as a discriminating variable and fed into a neural network (Section **??**).

3.6 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.

3.7 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.

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 **??**

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

- Overlap Removal as discussed in Section **??**

- 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 **??** and in the discussion of kinematic region creation throughout the rest of the analysis e.g., Section **??**.

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.

4.2 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:

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

Table 2: 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 3: The muon trigger requirements in the event selections

- 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 ??.

Other variables of interest, i.e. those being used as inputs into the neural network are also showed in this section. Figure ?? shows the S_T and m_T^W distributions. Figure ?? 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 ?? and for the muon channel in Figure ??. The neural network output of these events are shown in Figure ??

(a) electron channel
(b) muon channel
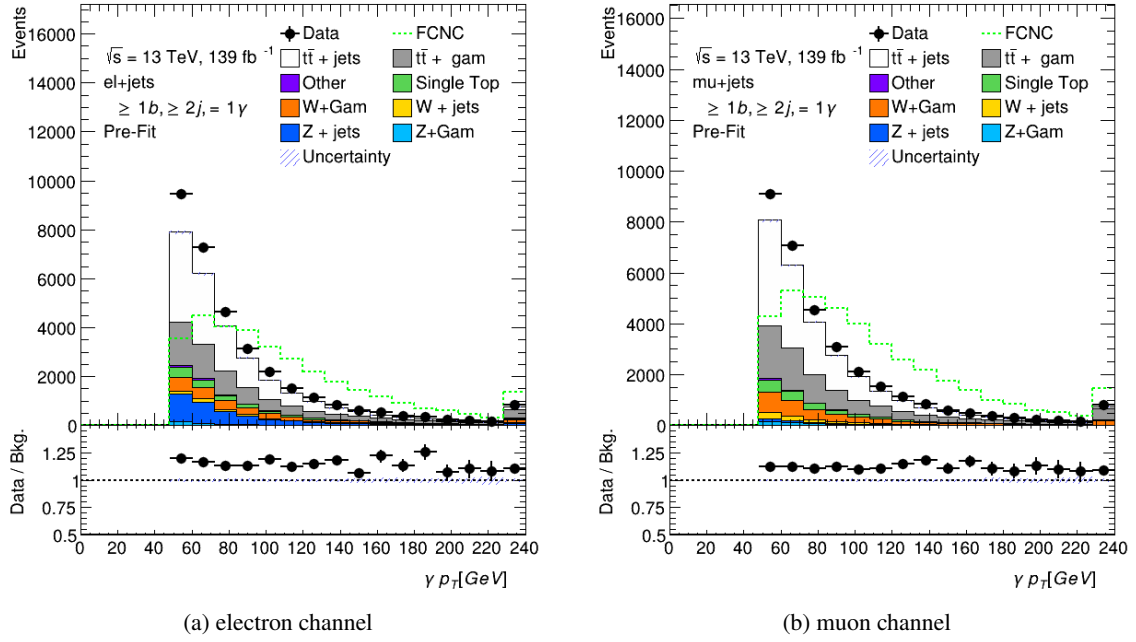
Figure 5: Photon p_T in the signal region pre-preselection region. FCNC signal branching ratio is scaled to 1%.

4.3 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 ?? and regions with a real photon are described in Section ??.

4.3.1 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:

(a) electron channel

(b) muon channel

(a) electron channel

(b) muon channel
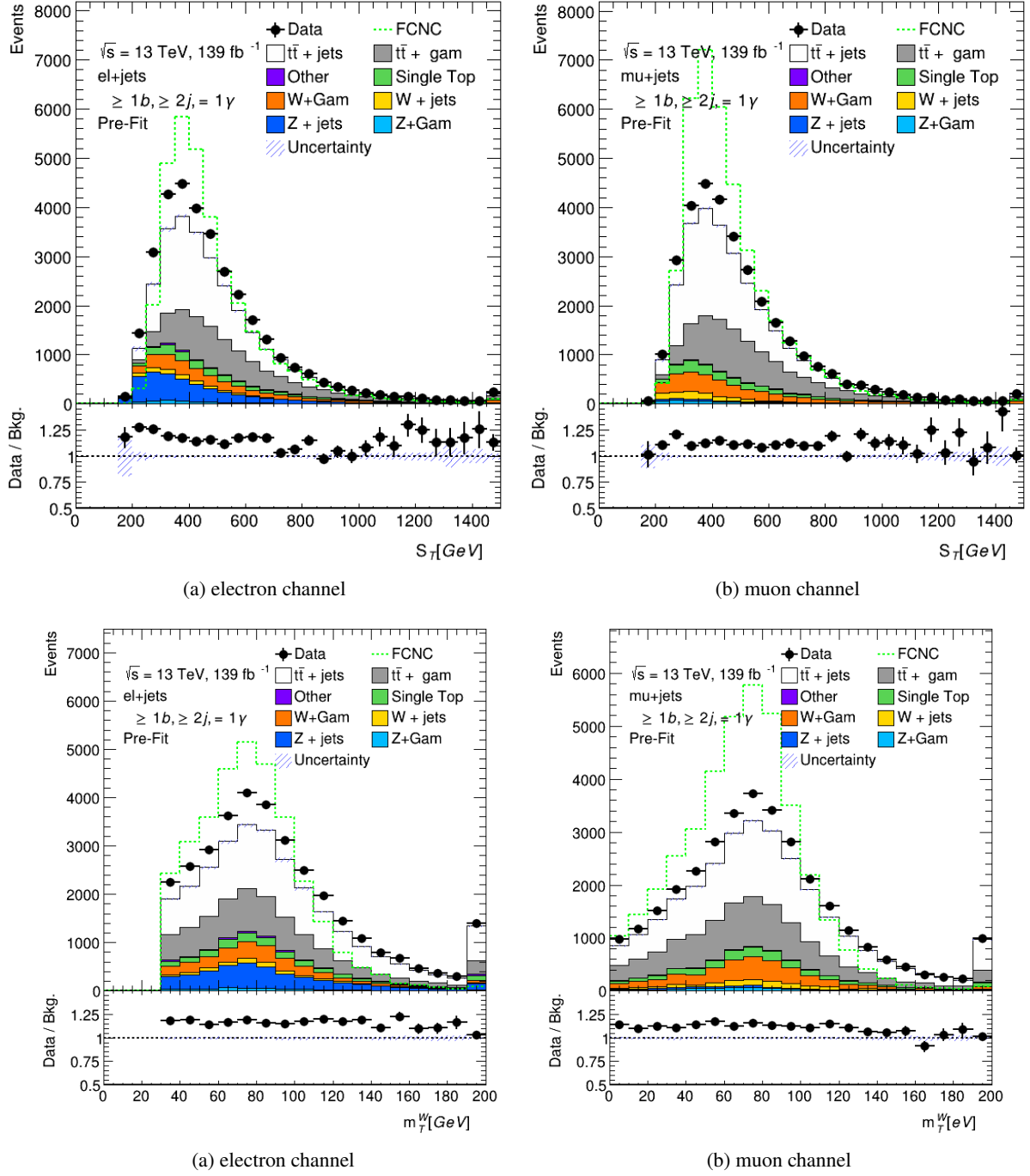
Figure 6: S_T and m_T^W in the signal region pre-selection region. FCNC signal branching ratio is scaled to 1%.

326   • All of the Initial Event Selection as outlined in Section **??**

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

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

(a) electron channel

(b) muon channel
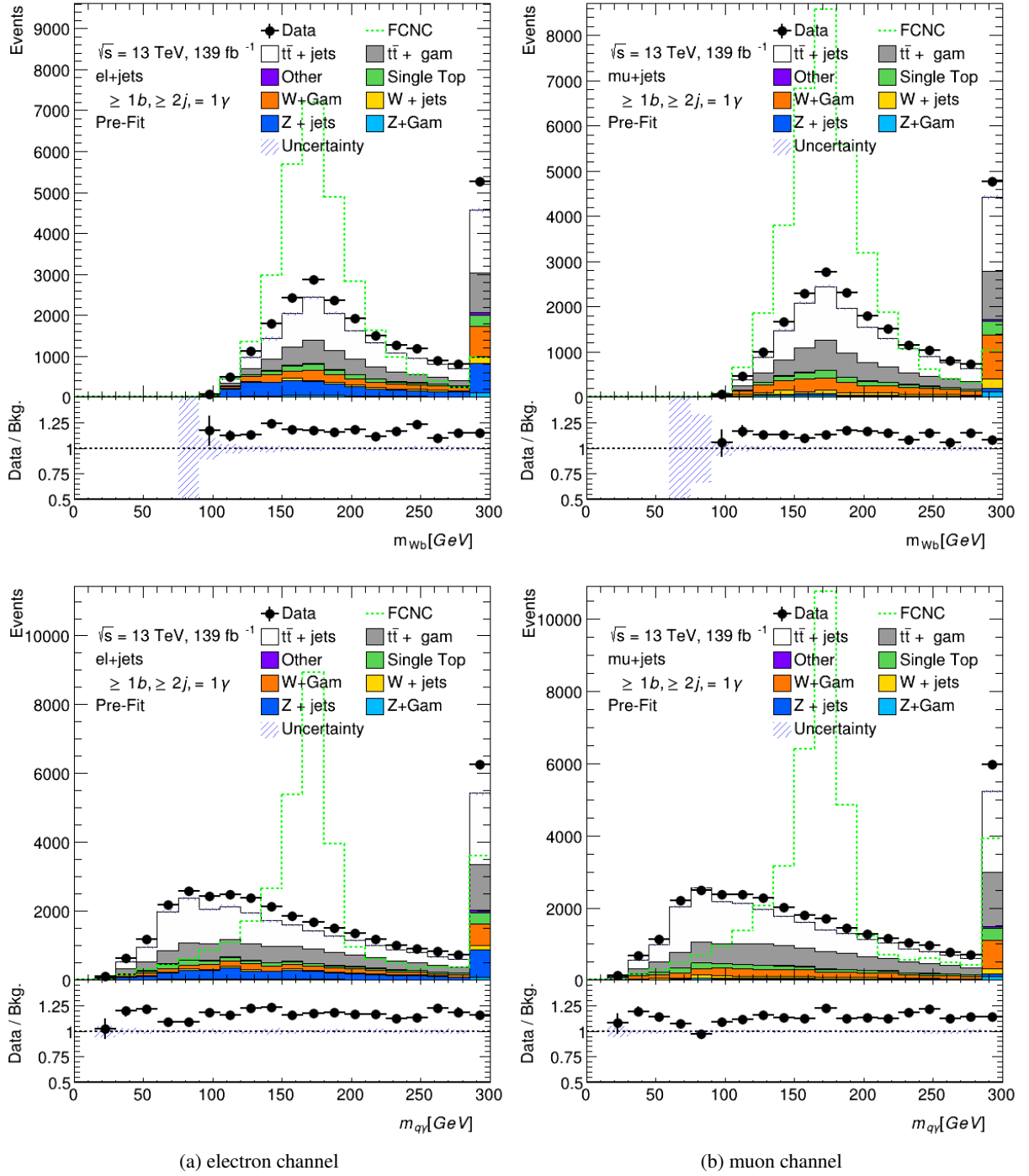
Figure 7: Top mass candidates in the signal region pre-selection: m_{Wb} and $m_{q\gamma}$. FCNC signal branching ratio is scaled to 1%.

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

330   – Validation Region: $n_{\text{jets}} = 4$
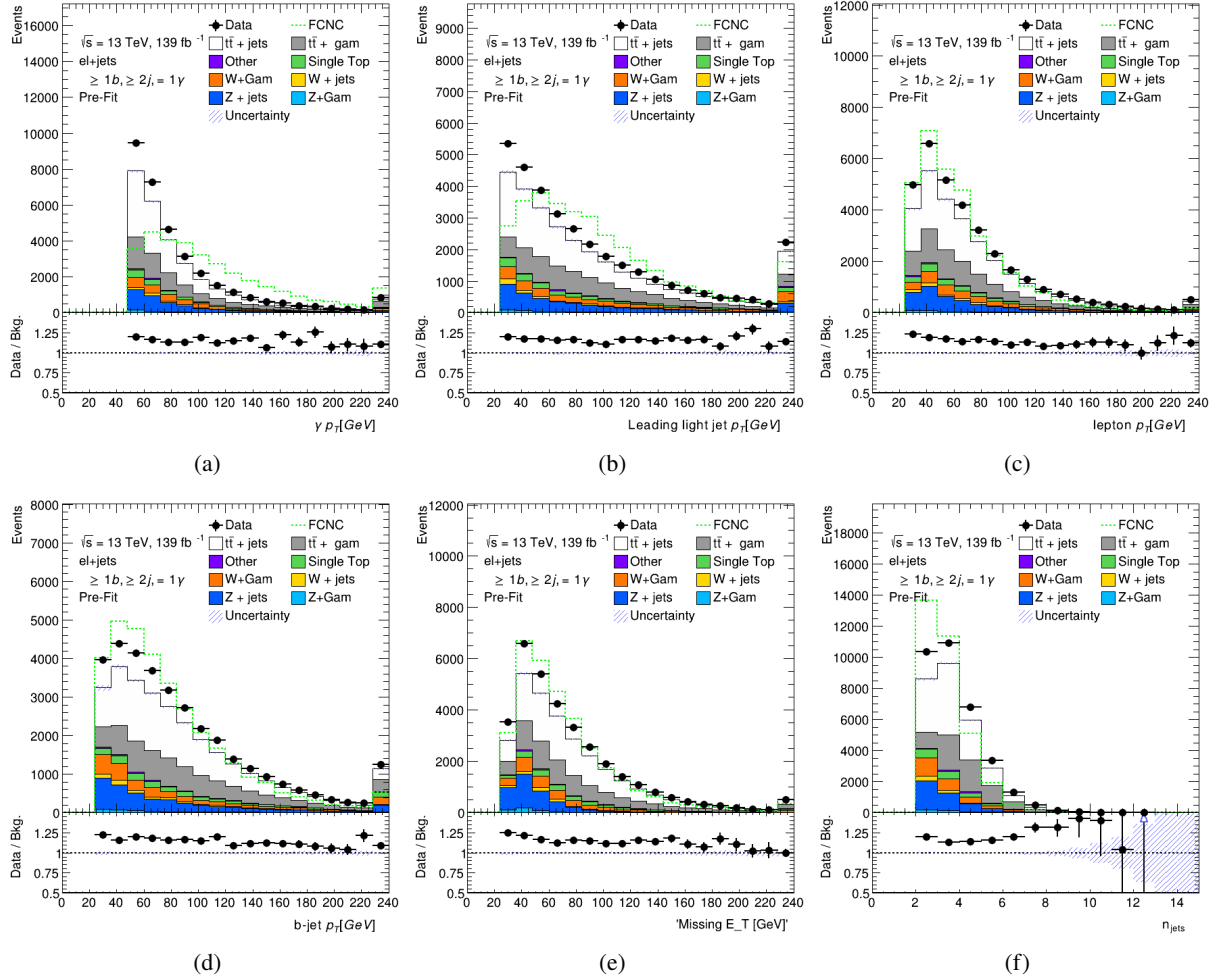
Figure 8: Photon p_T (a), leading light jet p_T (b), lepton p_T (c), b-jet p_T (d), $\not{E}_T$ (e), and n_{jets} (f) plots in the signal region pre-selection for the electron+jets channel. FCNC signal branching ratio is scaled to 1%.

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

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

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

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

335   • 0 photons, $p_T > 15$ GeV

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

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

338 Figure ?? shows the S_T distribution in both electron and muon channels before scale factors are calculated
339 for all three kinematically separate regions. The large mismodelling occurs at low S_T values as expected
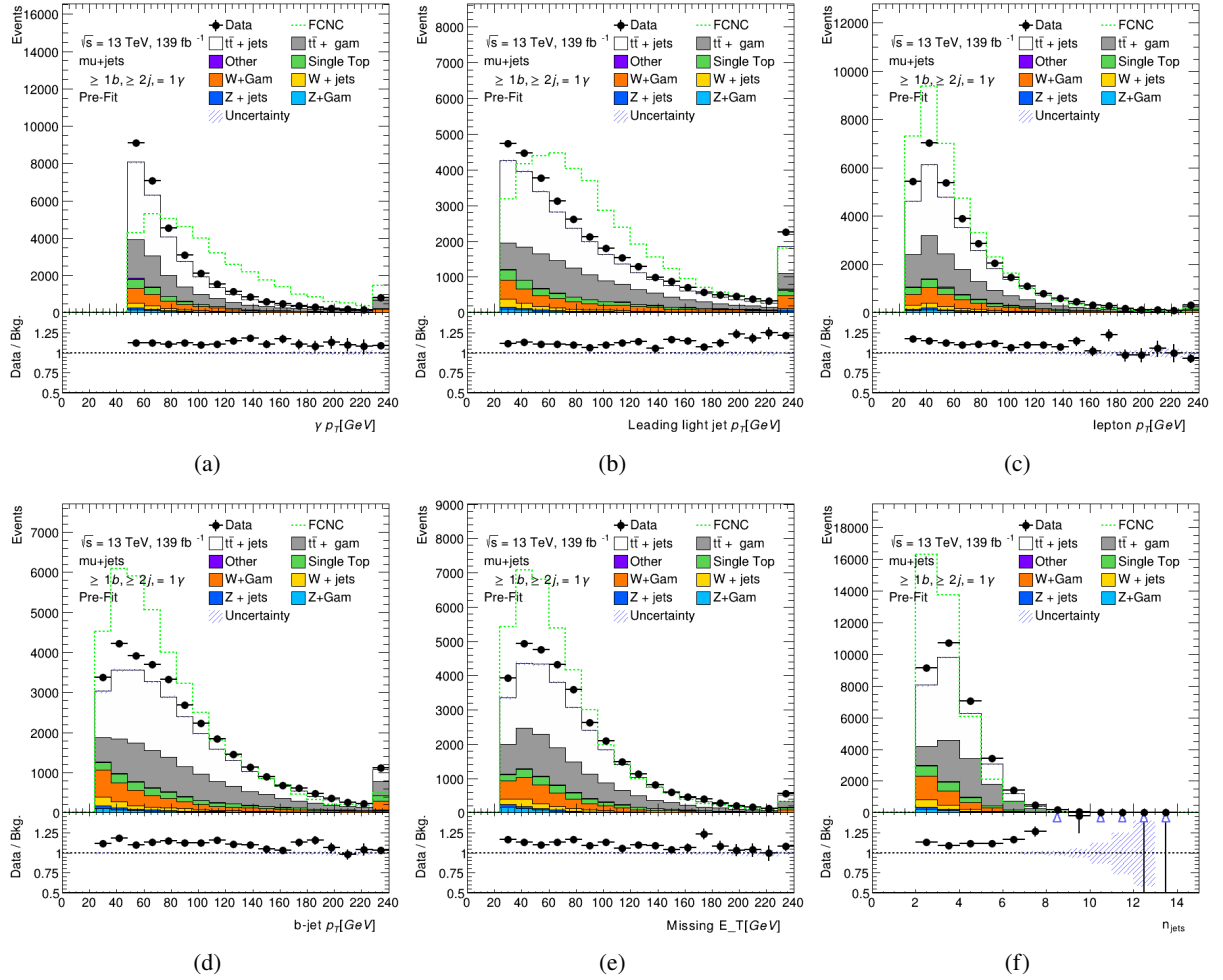
Figure 9: Photon p_T (a), leading light jet p_T (b), lepton p_T (c), b-jet p_T (d), $\not{E}_T$ (e), and n_{jets} (f) plots in the signal region pre-selection for the muon+jets channel. FCNC signal branching ratio is scaled to 1%.

340 as QCD processes will typically add low energy jets to the events. The Figures **??**(electron channel) and
341 **??**(muon channel) show various event-level variable plots for the validation region after the scale factors
342 have been applied.

The derived scale factors using these regions are shown in Table **??**

Sample	e+jets SF	μ+jets SF
W+jets	1.22	1.25
$t\bar{t}$	1.06	1.01

Table 4: Derived $t\bar{t}$ and W+jets scale factors for QCD multijet backgrounds.

343

(a) electron channel

(b) muon channel
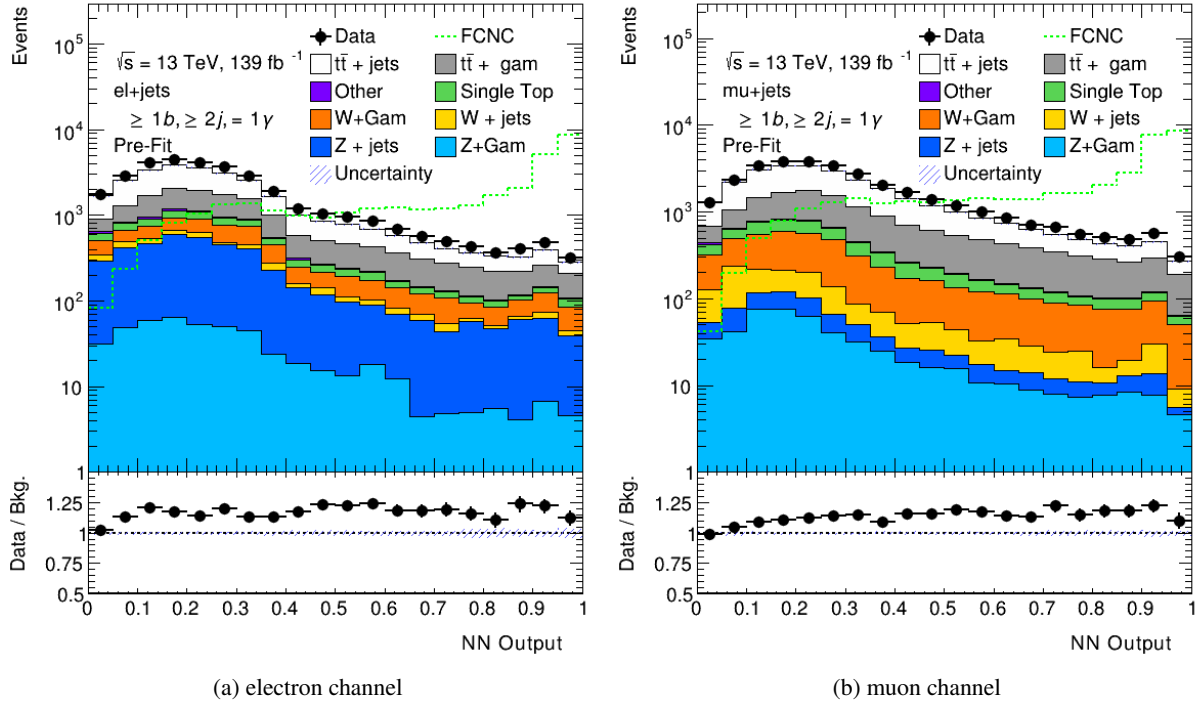
Figure 10: Output of the Neural Network in the signal region pre-selection region. FCNC signal branching ratio is scaled to 1%.

4.3.2 Background With Photons

Standard Model processes that are produced with an extra real photon are an 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 **??**. 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.

4.3.3 W+γ Control Region

A validation region for W+jets+γ was created as it is one of the more dominat backgrounds other than $t\bar{t}$ and $t\bar{t} + \gamma$ events. The normalization for the W+jets+γ validation region enters as a free parameter into the final fit. The region selection for the W+jets+γ is as follows:

- All of the Initial Event Selection as outlined in Section **??**

- Exactly 1 lepton (electron or muon) $p_T > 25$ GeV
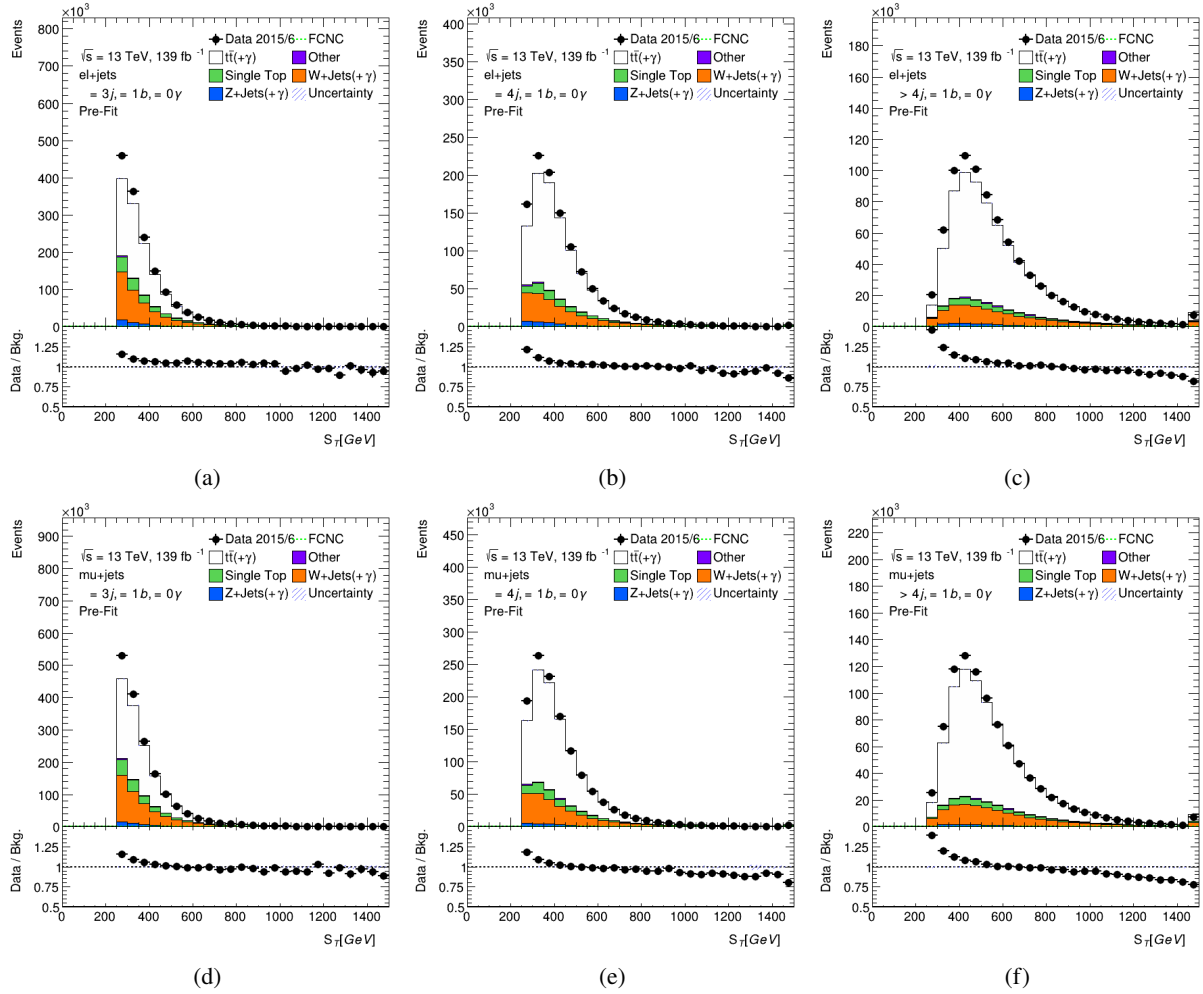
Figure 11: S_T distributions in the 3(a,d), 4(b,e), and 5+(c,f) jets control and validation regions. Electron channel is shown on the top and the muon channel on the bottom, before scale factors are determined.

360 • At least 2 Jets ($p_T > 25$ GeV)

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

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

363 • Exactly 0 b-tagged jet (MV2c10, 77% Working point)

364 • Exactly 1 photons, $p_T > 50$ GeV

365 • Photon isolation cuts: topo$E_T$cone40<4 GeV

366 • Z mass cut $|m_{l\gamma} - m_Z|V > 5$ GeV

367 Distributions of kinematic variables in the electron (muon) channels are shown in Figure ?? (??).

Figure 12: Event-level plots for the =4 jet validation region after scale factors have been applied in the electron channel. FCNC signal branching ratio is scaled to 1%.

Channel:	e+jets	μ+jets
W+jets+γ SF	PUT IN VALUE	PUT IN VALUE

Table 5: Fit result W+jets+γ normalization scale factors including both statistical and systematic uncertainties.

4.3.4 $t\bar{t} + \gamma$ Control Region

Another validation region was created for the other largest photon enriched samples, $t\bar{t} + \gamma$. The normalization for the $t\bar{t} + \gamma$ validation region enters as a free parameter into the final fit and the region selection is as follows:

- All of the Initial Event Selection as outlined in Section **??**

- Exactly 1 lepton (electron or muon) $p_T > 25$ GeV
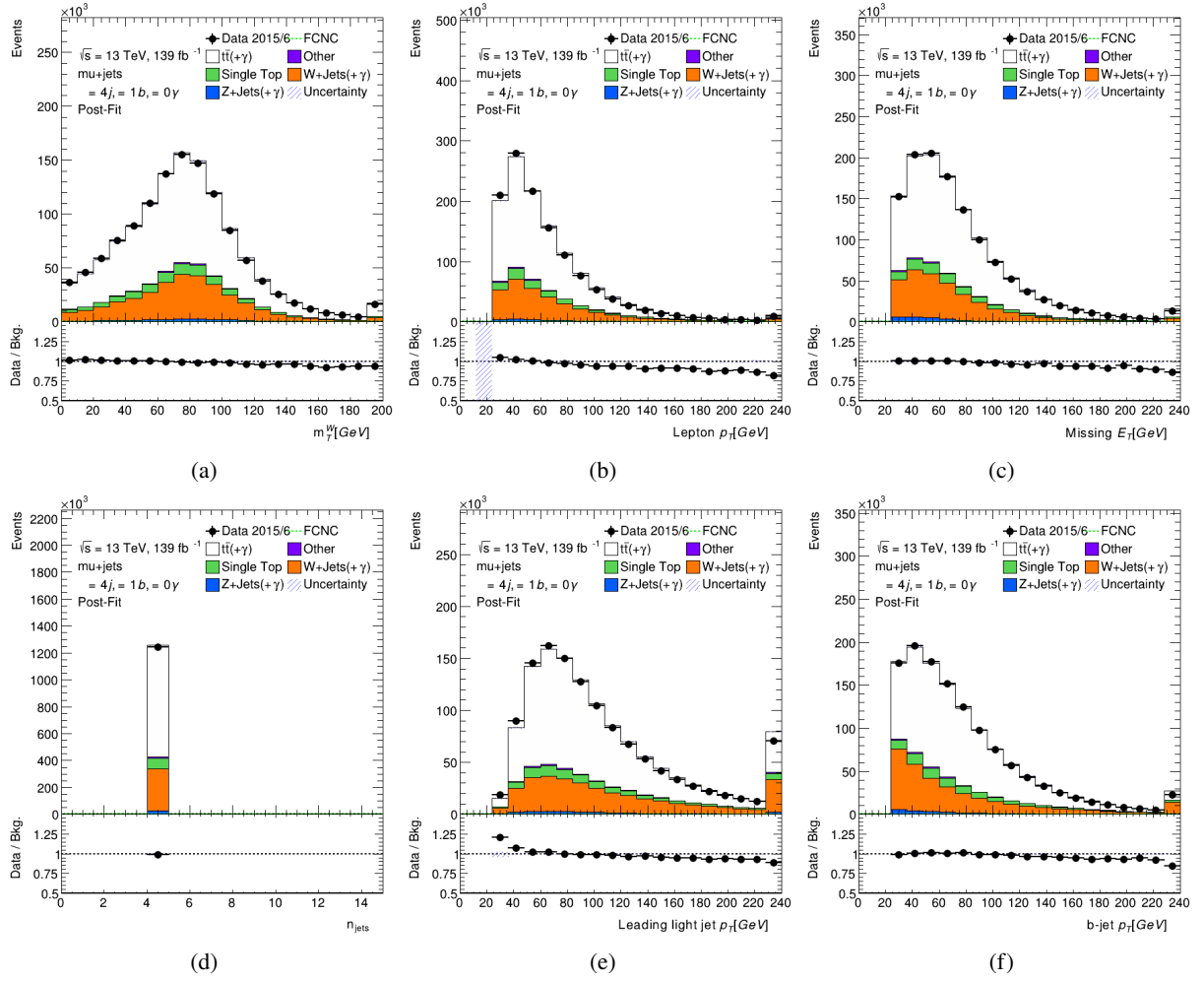
Figure 13: Event-level plots for the =4 jet validation region after scale factors have been applied in the muon channel. FCNC signal branching ratio is scaled to 1%.

- At least 4 Jets ($p_T > 25$ GeV)

- $\not{E}_T > 30$ 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)

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

- Exactly 1 photons, $p_T > 50$ GeV

- Photon isolation cuts: topo$E_T$cone40<4 GeV

- Reverse Neural Network Cut: NNOutput<0.93 electron channel, NNOutput<0.92 muon channel

Distributions of kinematic variables in the electron (muon) channels are shown in Figure ?? (??).
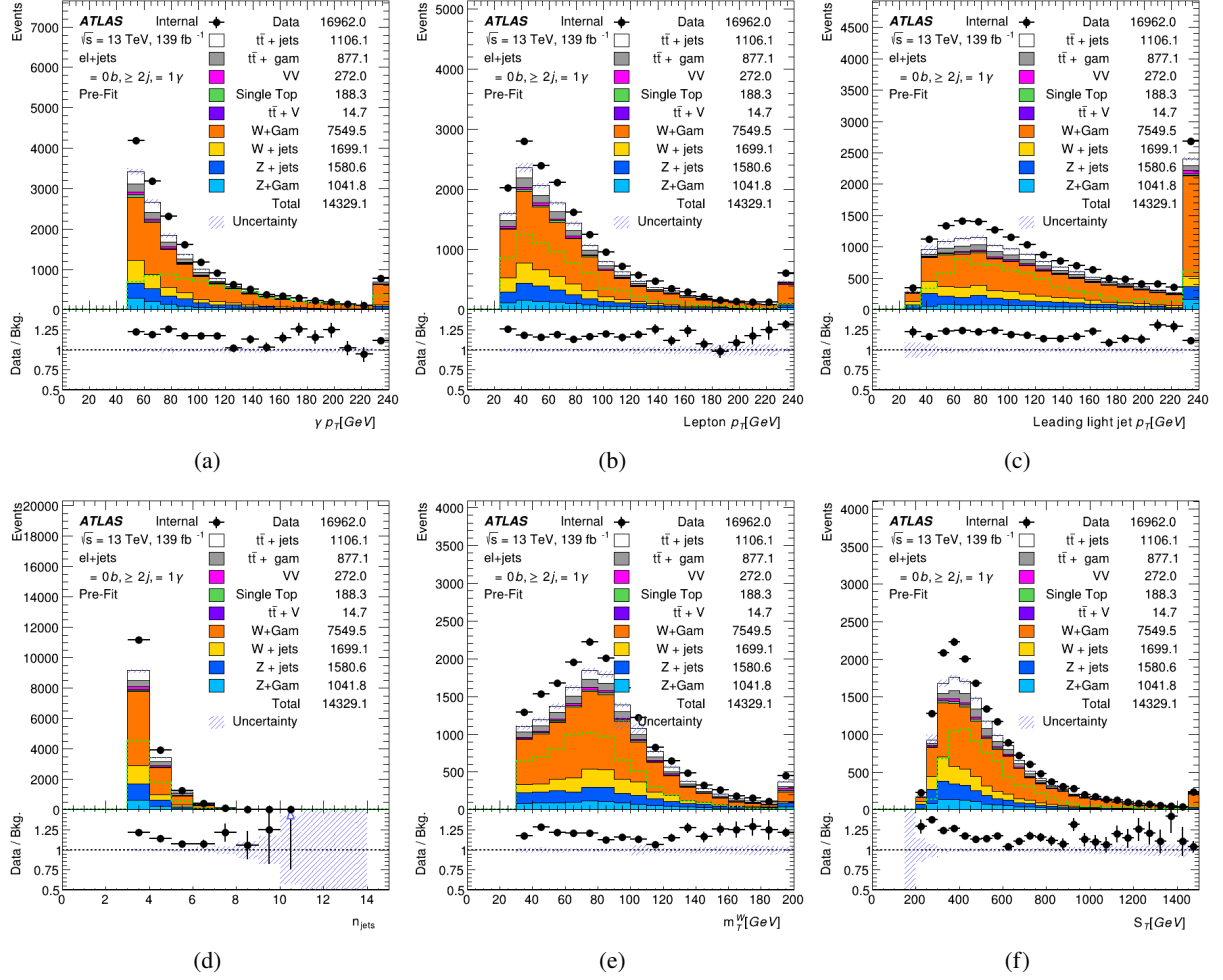
Figure 14: W+jets+γ validation region plots for the electron channel. The FCNC signal sample is scaled to 1%.

Channel:	e+jets	μ+jets
$t\bar{t} + \gamma$ SF	PUT IN VALUE	PUT IN VALUE

Table 6: Fit result $t\bar{t} + \gamma$ normalization scale factors including both statistical and systematic uncertainties.
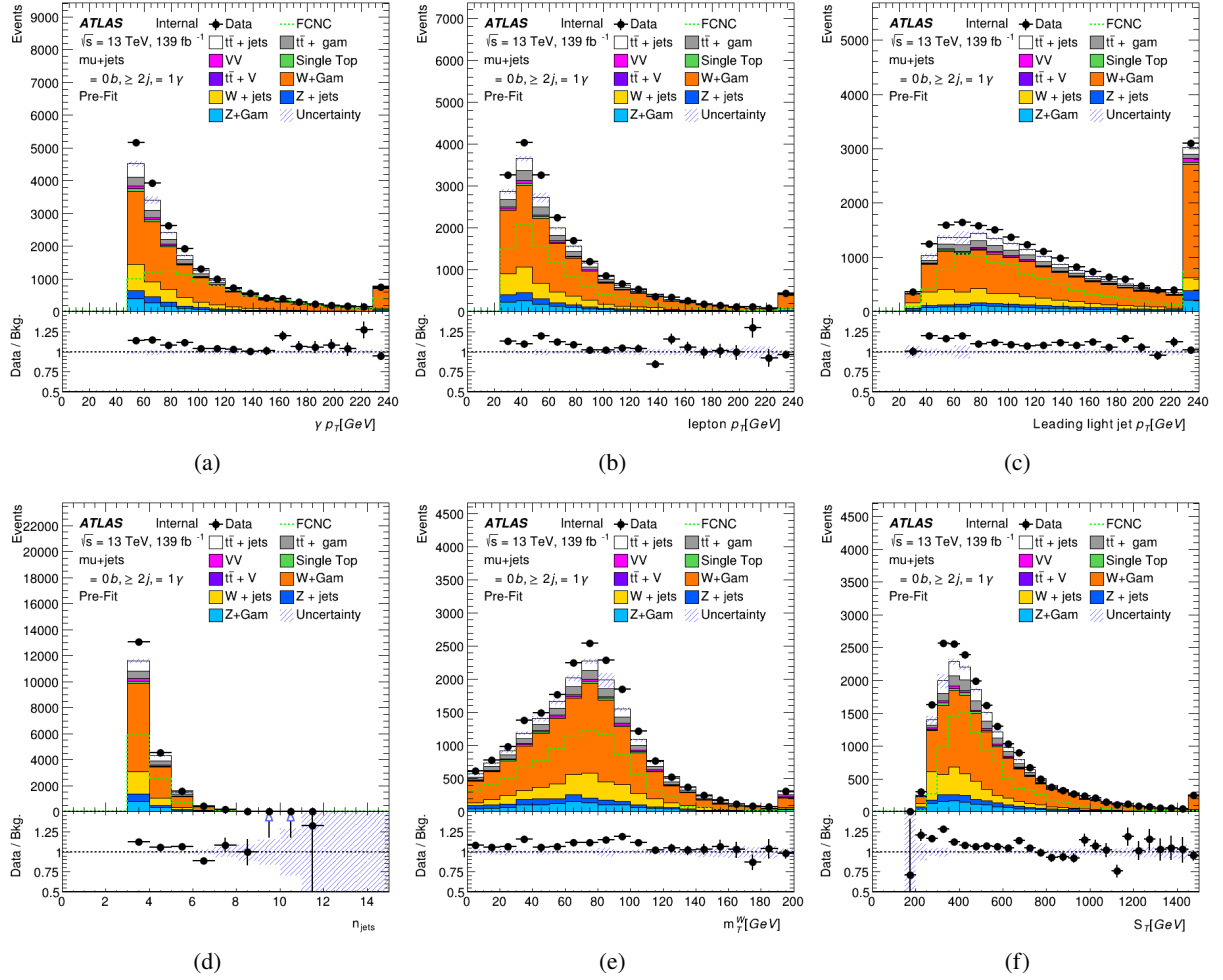
(a)

(b)

(c)

(d)

(e)

(f)

Figure 15: W+jets+γ validation region plots for the muon channel. The FCNC signal sample is scaled to 1%.
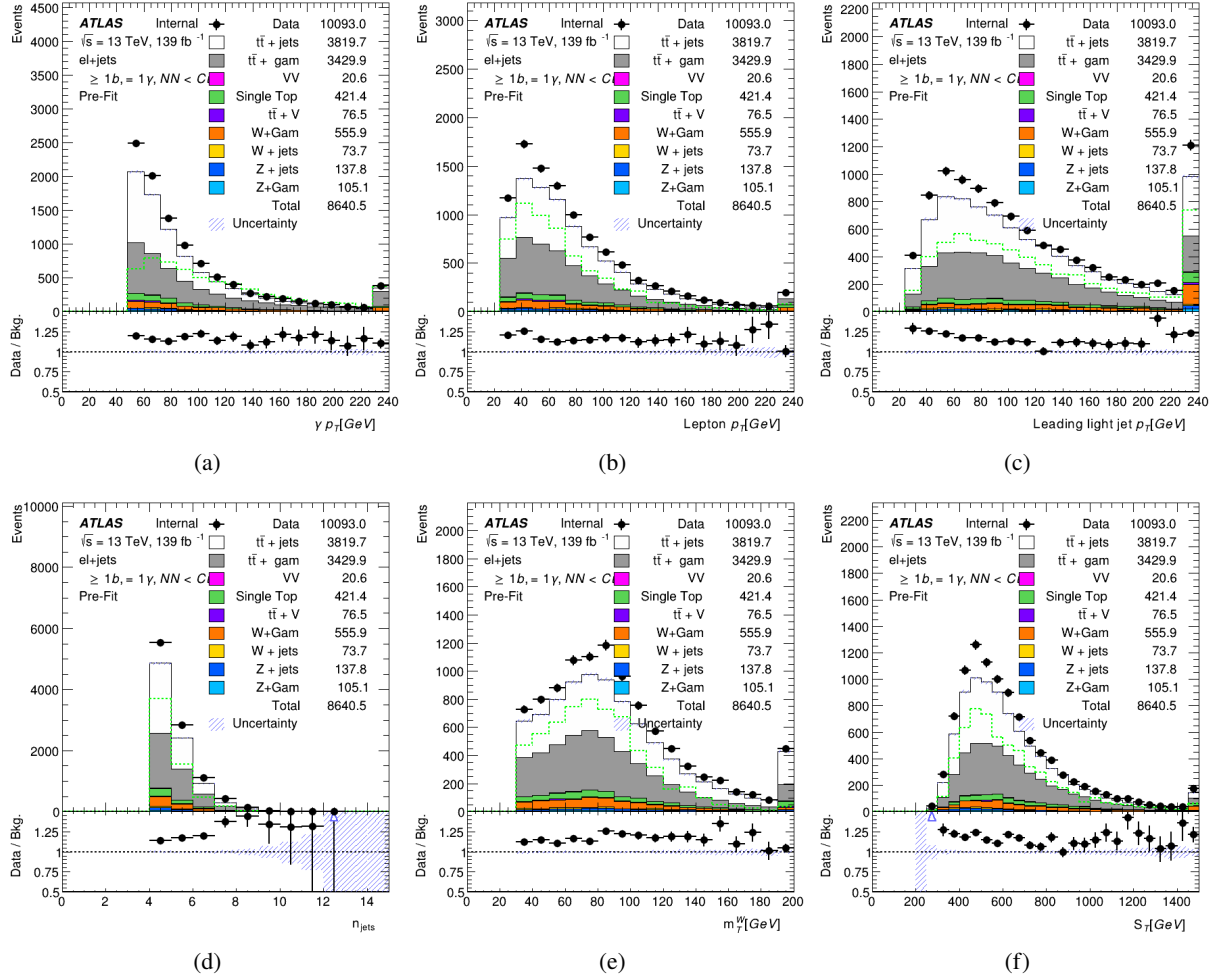
Figure 16: $t\bar{t}$+jets+γ validation region plots for the electron channel. The FCNC signal sample is scaled to 1%.
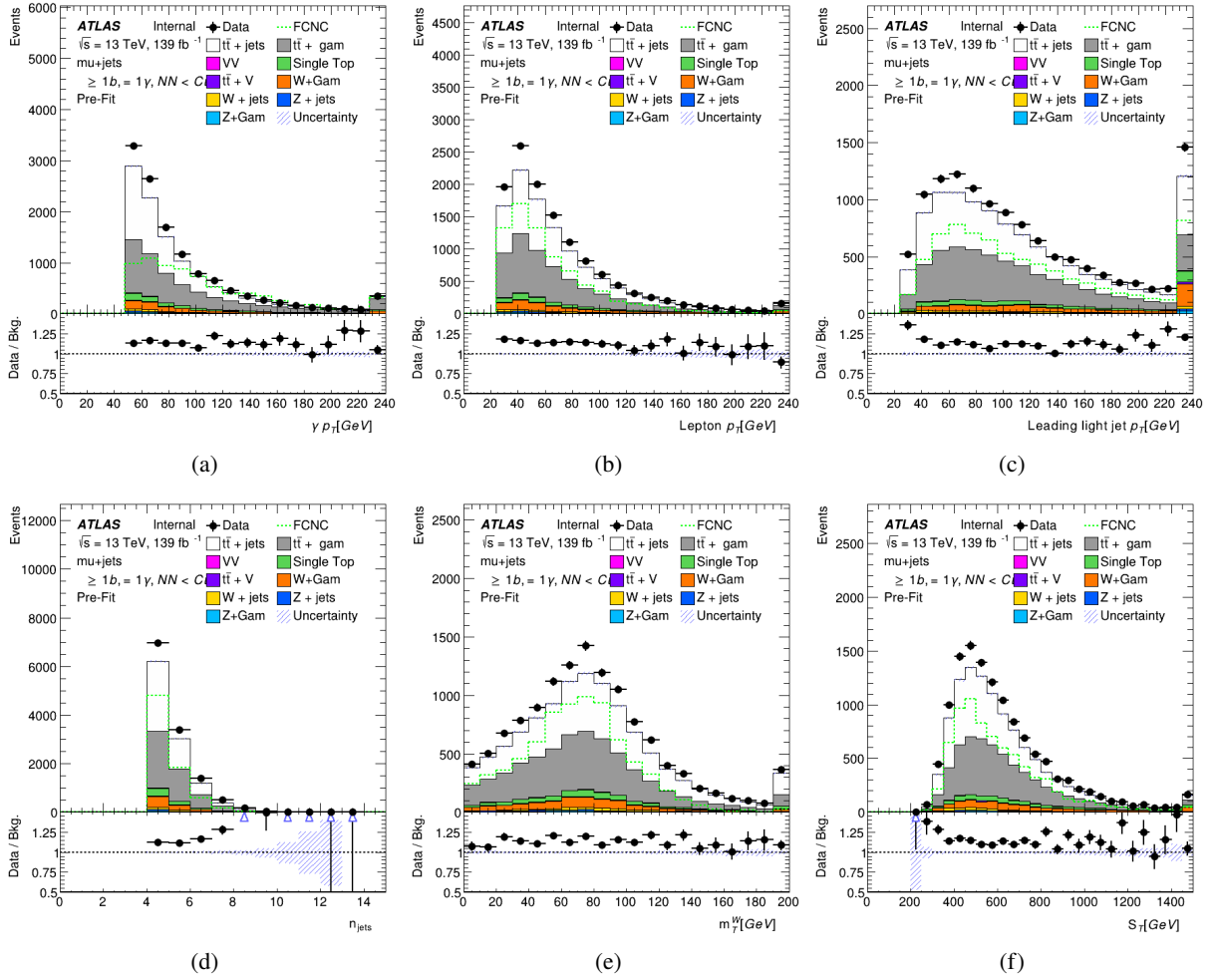
Figure 17: $t\bar{t}$+jets+γ validation region plots for the muon channel. The FCNC signal sample is scaled to 1%.

5 Background Estimations

6 Neural Network for Signal Background Separation

6.1 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.

6.1.1 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 **??** 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 **??**, **??**, and **??**

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

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

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 7: 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.

picking the network and working point with the best estimated significance. The b-tagging neural network study is shown in Section **??**

6.1.2 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.

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

473 not become overly dependent on any given node and learning the data 'by heart' as opposed to recognizing
474 the trends in the sample.

6.1.3 Training and Validation of Neural Networks

476 The input variables into the neural network are preprocessed using the RobustScalar method implemented
477 in **scikit-learn**[?]. The preprocessing is done so that the input variables exist on a similar scale. As the
478 network is tasked with learning how to combine these inputs through a series of linear combinations and
479 nonlinear activation function values a disparity in the scales of the input values can lead to awkward loss
480 function topology that will focus on certain parameter gradients instead of treating them all similarly.
481 Normalizing the values to a standard scale allows the network to learn the optimal parameters for each
482 input node more quickly and efficiently. This means that less focus can be used on the optimization of
483 the hyperparameters for the network as the scales of the inputs do not need to be learned by the network
484 itself.

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

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

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

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

493 One measure of the performance of the network is the accuracy. The Keras default accuracy measure is
494 defined:

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

495 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
496 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
497 prediction values occur assuming a cut on the output of ≥ 0.5.

6.1.4 Hidden Layer Studies

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

505 The AUC for 2 hidden layers and 3 hidden layers are identical, to rounding errors, for both channels. As
506 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 **??**. 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 **??** and **??** 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.

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 **??**.

6.1.5 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 **??** 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 **??** and **??**. Following that the neural network output and significance plots are shown in Figures **??** and **??**.

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 **??**.

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 8: Branching ratio values with a significance of 2 after neural network optimization

(a) γ_{iso} topo$E_T$cone40

(b) γ_{p_T}

(c) $m_{q\gamma}$
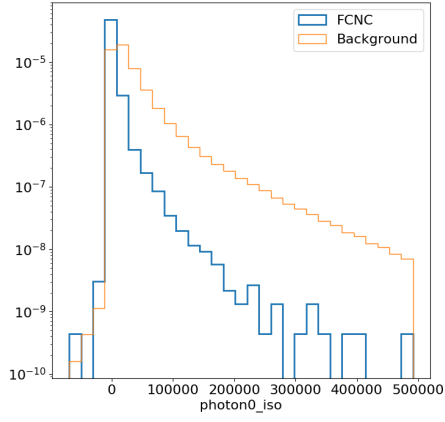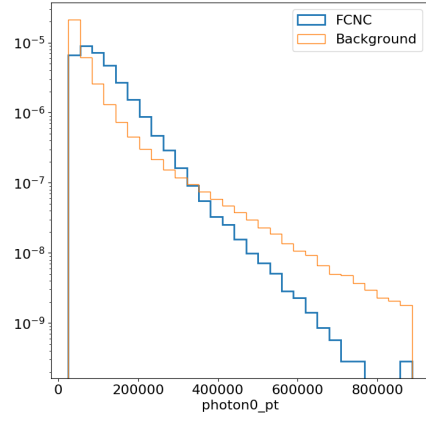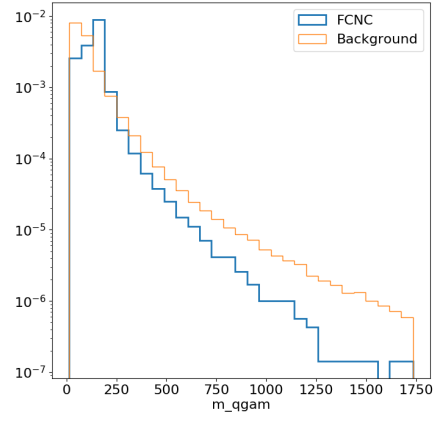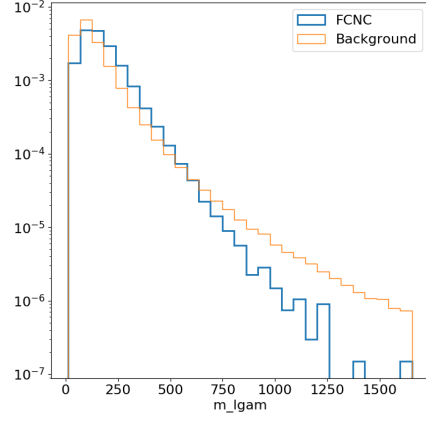
(d) $m_{l\gamma}$
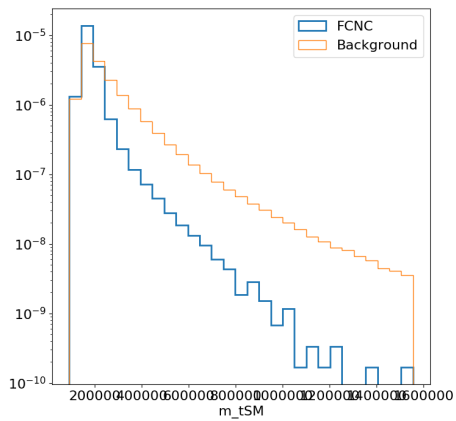
(e) m_{bW}
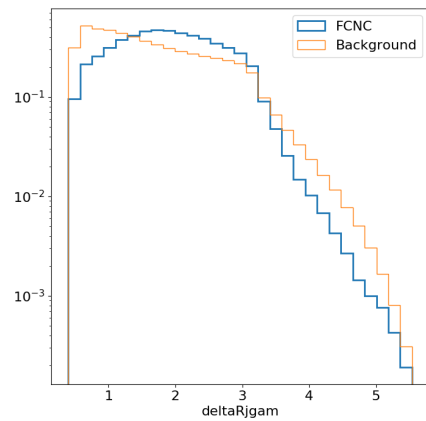
(f) $\Delta R_{j\gamma}$

Figure 18: 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) $p_T(q)$
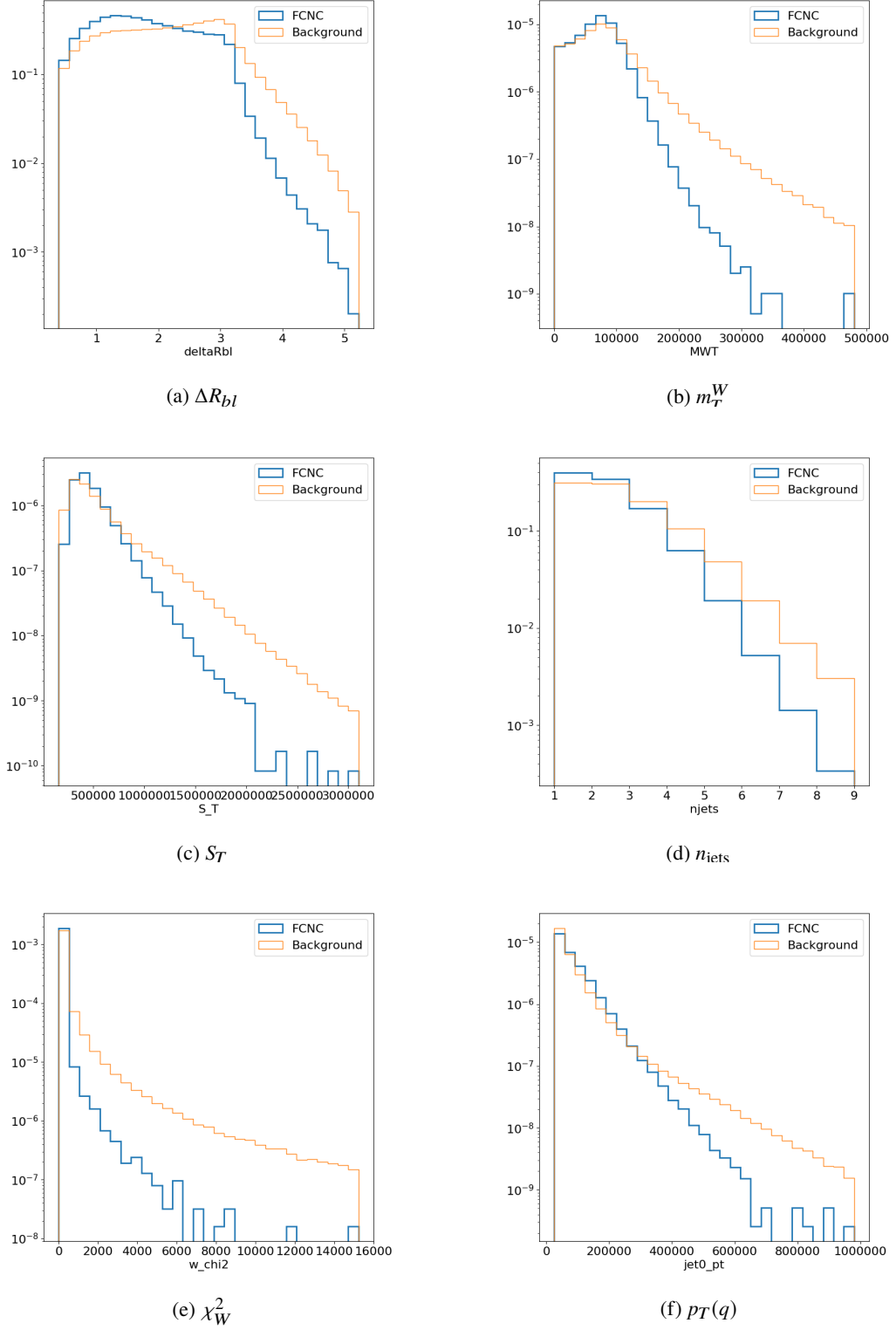
Figure 19: Normalized variables showing the shapes of neural network input variables for the μ+jets channel: ΔR_{bl}, m_T^W, S_T, n_{jets}, χ_W^2, and $p_T(q)$

(a) $\Delta R_{l\nu}$

(b) E (lepton)

(c) $\not{E}_T$
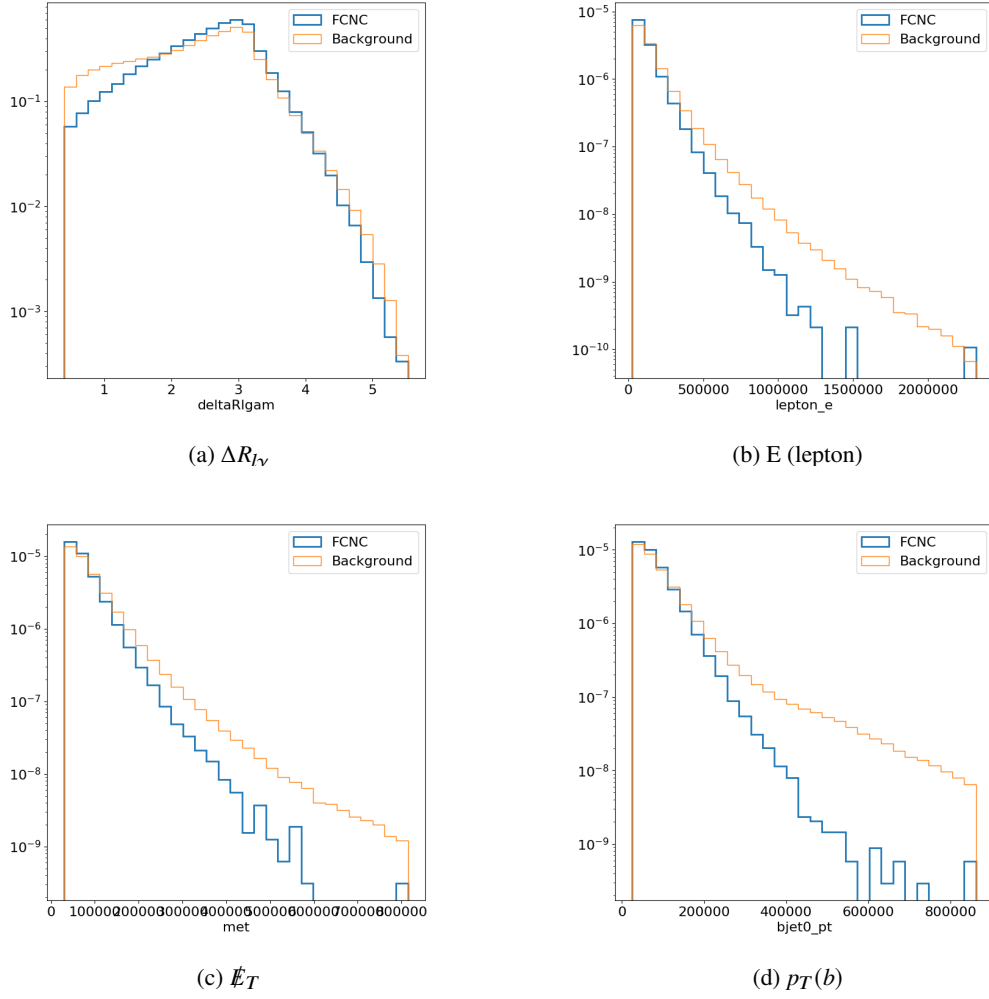
(d) $p_T(b)$

Figure 20: 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 $p_T(b)$
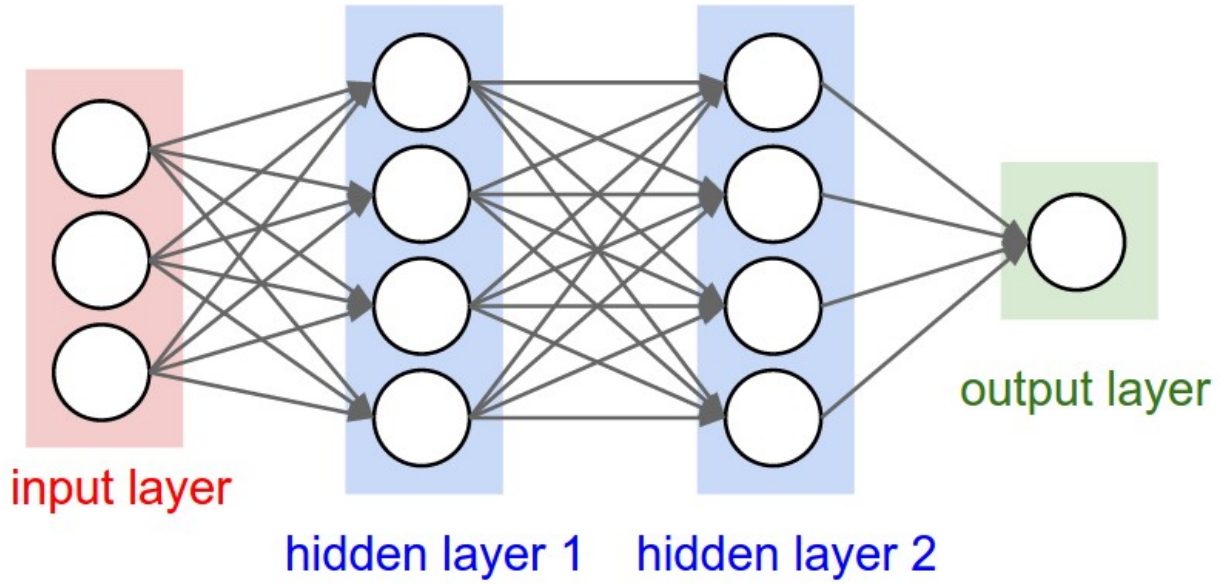
Figure 21: Pictoral representation of neural network architecture with 3 input variables, 2 hidden layers with 4 nodes each, and 1 output layer[?].
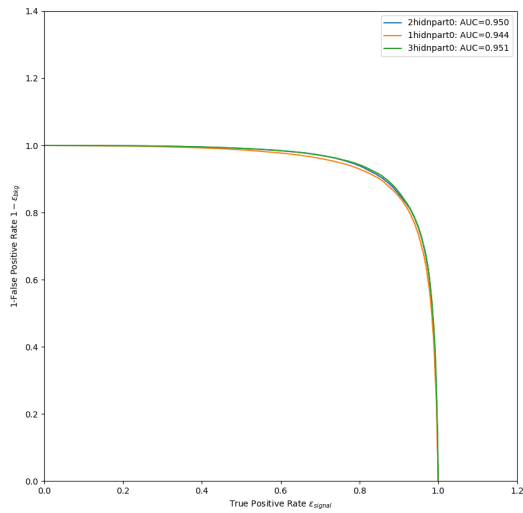
(a) e+jets ROC Curves
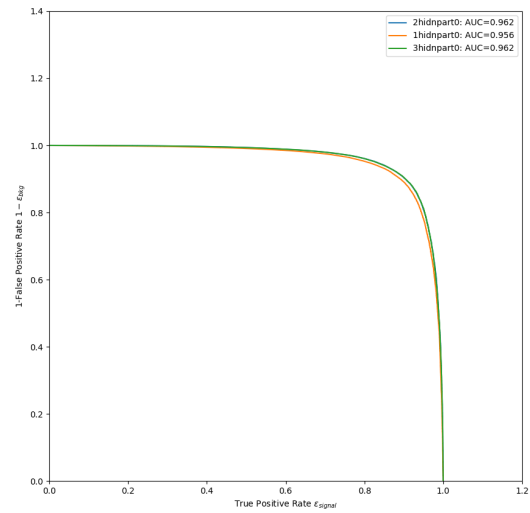
(b) μ+jets ROC Curves

Figure 22: 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.
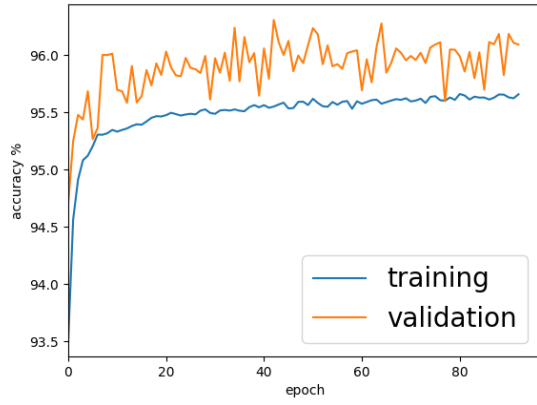
(a) *e*+jets Accuracy Curves
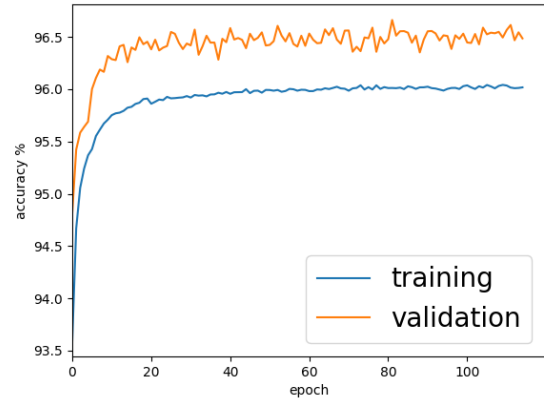
(b) *μ*+jets Accuracy Curves

Figure 23: Accuracy plots for both channels for the 2 hidden layer neural network
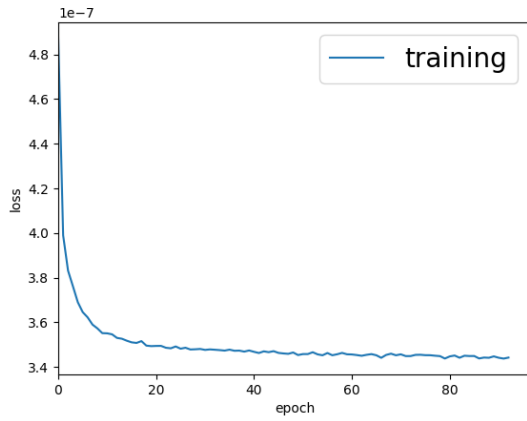
(a) *e*+jets Loss Curve
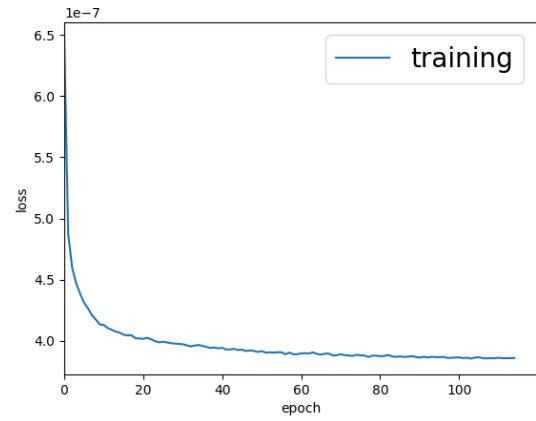
(b) *μ*+jets Loss Curve

Figure 24: Loss plots for both channels for the 2 hidden layer neural network

(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
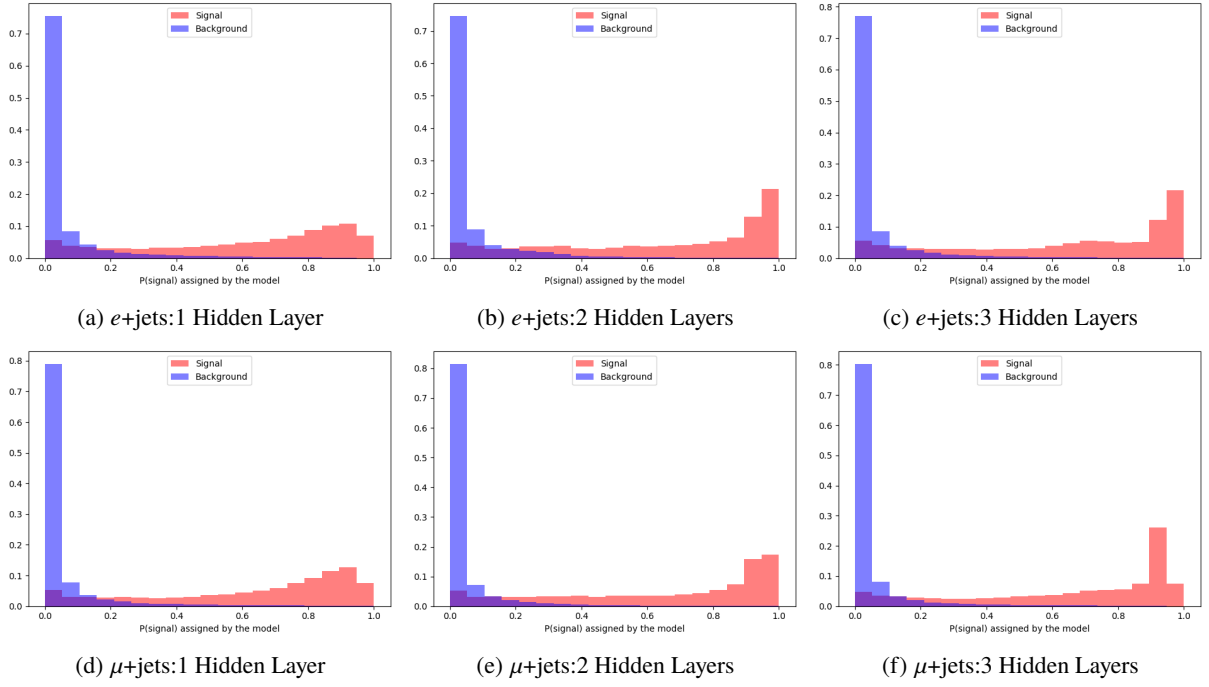
Figure 25: Normalized neural network output signal and background distribution plots are shown for both search channels for a varying number of hidden layers.

(a) *e*+jets   (b) *μ*+jets
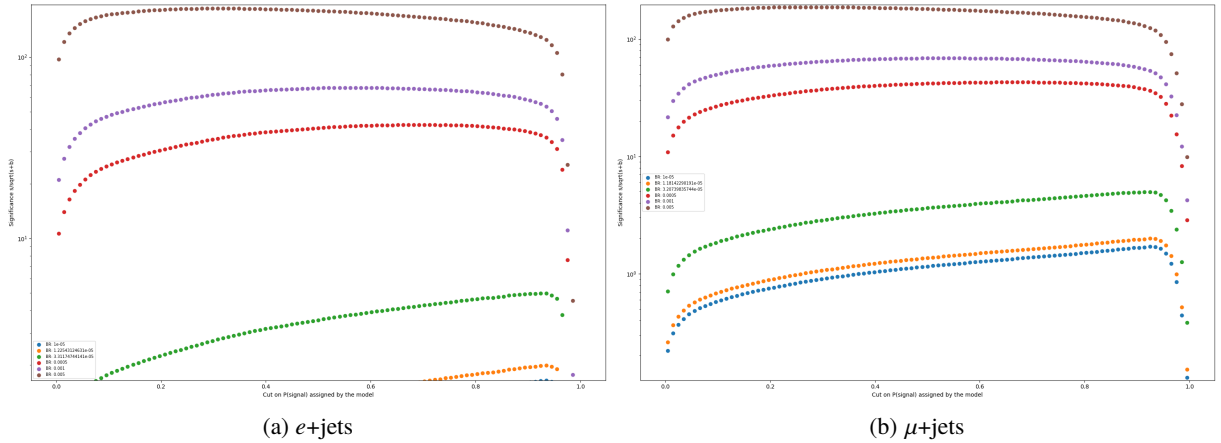
Figure 26: Significance plots for both channels for the 2 hidden layer neural network. The green points correspond to a branching ratio with a maximum significance of 5, the orange to a maximum significance of 2. The *e*+jets (*μ*+jets) branching ratio with max significance of 2 is $1.22 \times 10^{-5} (1.18 \times 10^{-5})$. 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.
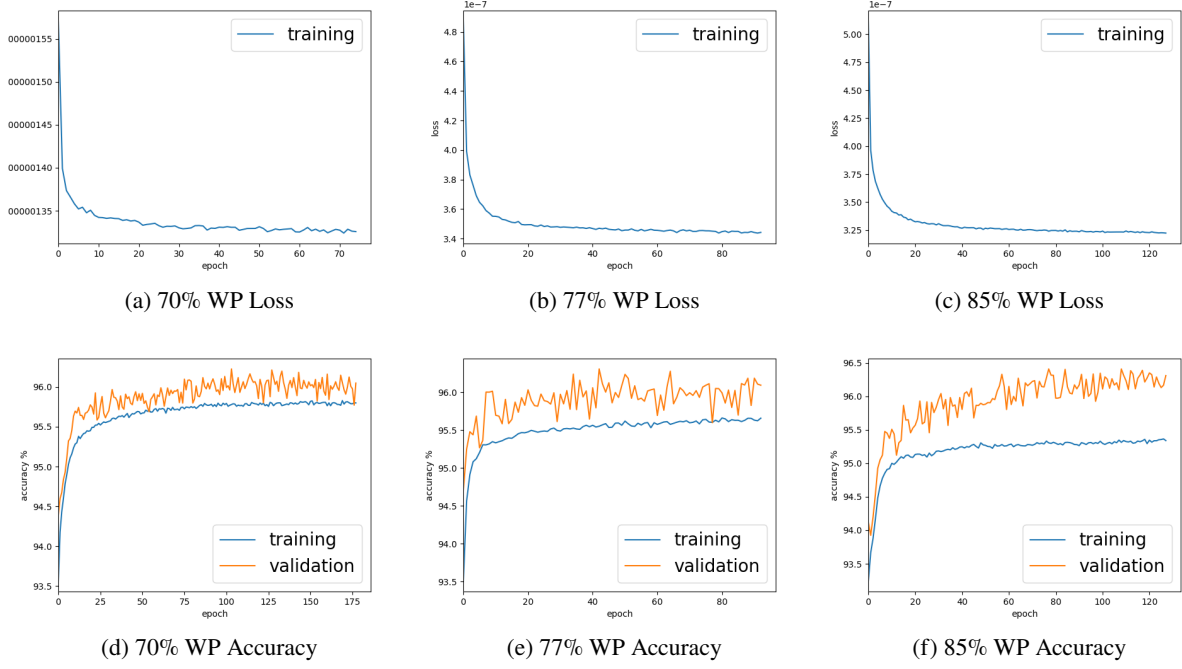
(a) 70% WP Loss

(b) 77% WP Loss

(c) 85% WP Loss

(d) 70% WP Accuracy

(e) 77% WP Accuracy

(f) 85% WP Accuracy

Figure 27: 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 28: 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
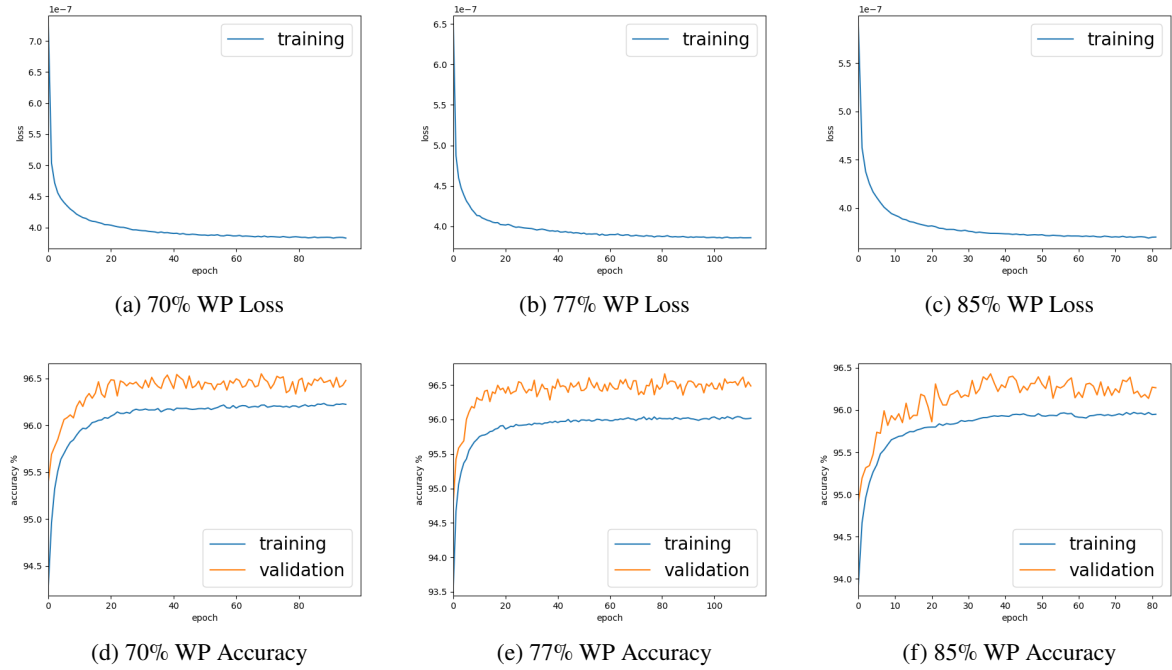
(d) 70% WP Significance

(e) 77% WP Significance

(f) 85% WP Significance
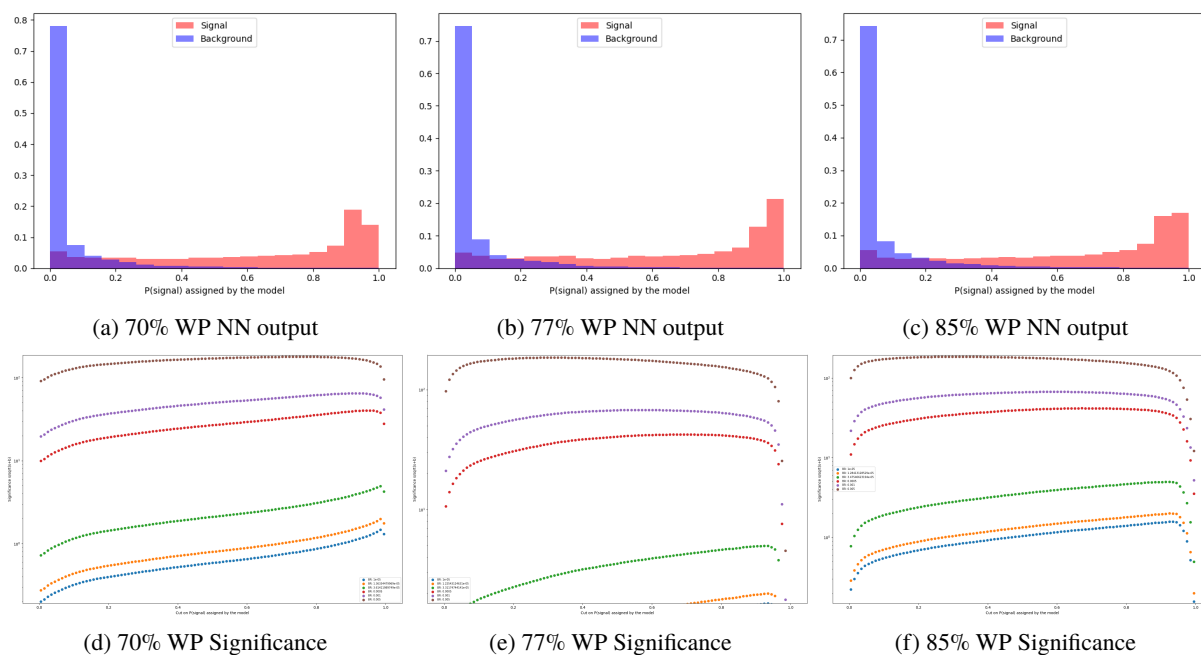
Figure 29: Neural network output and significance plots for the *e*+jets channel at 70%, 77%, and 85% b-tagging working points.

(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
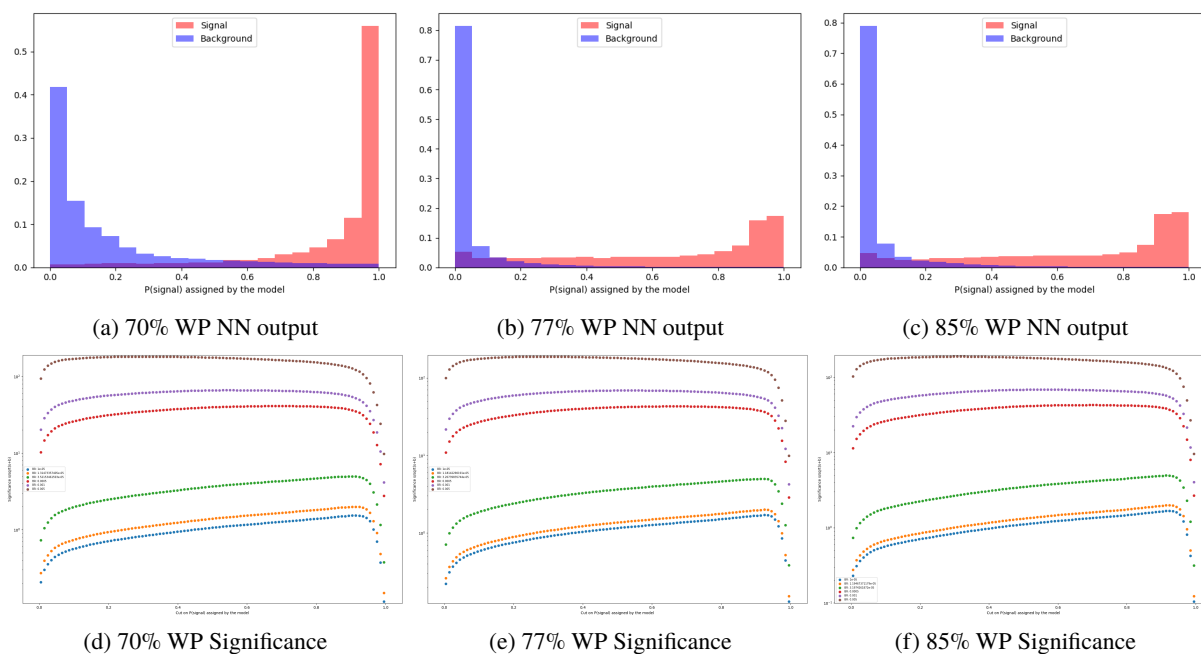
Figure 30: Neural network output and significance plots for the μ+jets channel at 70%, 77%, and 85% b-tagging working points.

540 ## 7 Analysis

540 Place your results here.

8 Conclusion

Place your conclusion here.

List of contributions

Jason Barkeloo	All Aspects of Analysis and Note
Jim Brau	Supervisor to Jason Barkeloo

Appendices

A Derivation Information (TOPQ1)

Preselection is applied to both data and MC samples using the derivation framework in order to reduce the xAOD sample size. TOPQ1 derivations were used and further skimmed for the specific n-tuples used in this analysis. TOPQ1 derivations are used for standard top analyses with at least one lepton and have the following definitions:

Particle	Definition
Electron	$|\eta| < 2.5$ and Electrons.DFCommonElectronsLHLoose
Muon	$|\eta| < 2.5$ and Muons.muonType=0 and Muons.DFCommonGoodMuon
Small-R Jet	$|\eta| < 2.5$ and AntiKt4EMTopoJets.DFCommonJets_Calib
Large-R Jet	$|\eta| < 2.5$ and AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets.DFCommonJets_Calib

Table 9: TOPQ1 Object Definitions

In addition to these object definitions TOPQ1 Derivations require trigger skimming of ≥ 1 lepton with $p_{\mathrm{T}} > 20\mathrm{GeV}$.

554 # B Complete List of Monte Carlo Samples Used

555 # C Additional Shape Comparison Plots: μ+jets channel

556 Various additional plots are shown in this appendix from the neural network creation and studies. Figure **??**
557 and **??** show additional shape comparisons in variables which are not included in the final neural network
558 model as they do not significantly change the fit values. In the cases of p_T or E variables with the higher
559 separation value were used as there is a large correlation between the two values and the other is shown in
560 this appendix. ΔR_{jb} was not included as the other 3 ΔR values had higher separation values and they are
561 all related to each other as they are the geometrically related.

562 The neutrino reconstruction is done using a minimization of

$$\chi_\nu^2 = \chi_{bW}^2 + \chi_W^2$$

563 . All three were investigated for their separation values and the χ_W^2 value had the largest separation.

564 # D Shape Comparison Plots: e+jets channel

(a) E (bjet)

(b) η_b

(c) ΔR_{jb}

(d) E (light jet)

(e) light jet η

(f) χ_ν^2
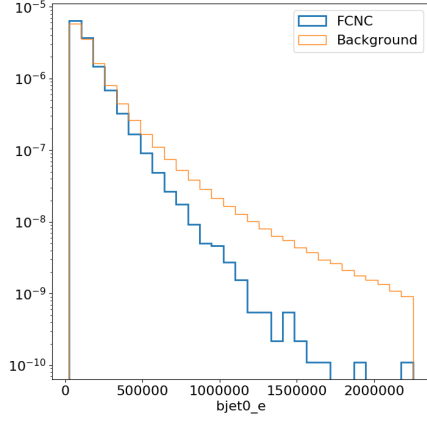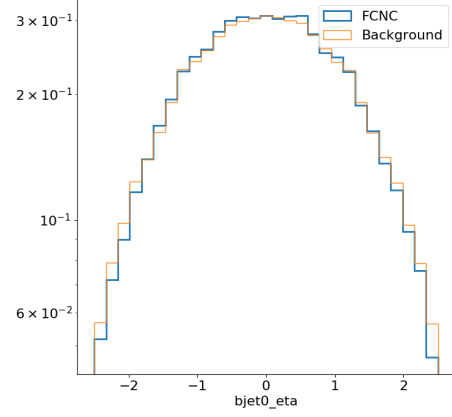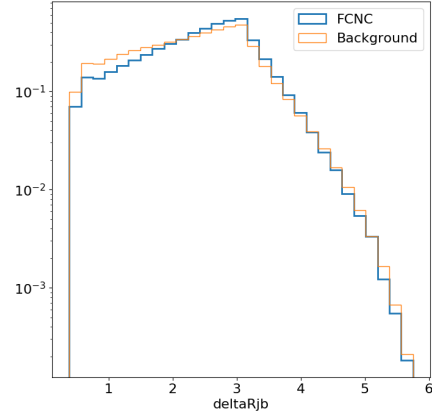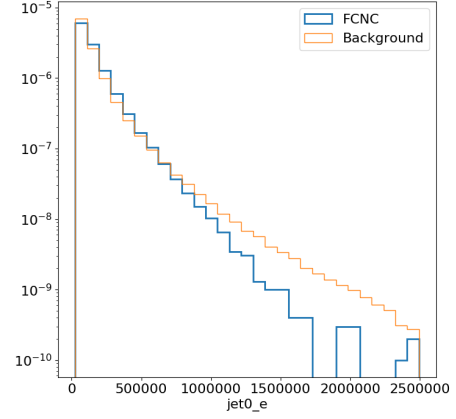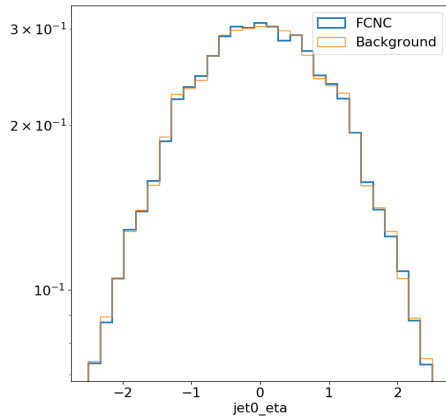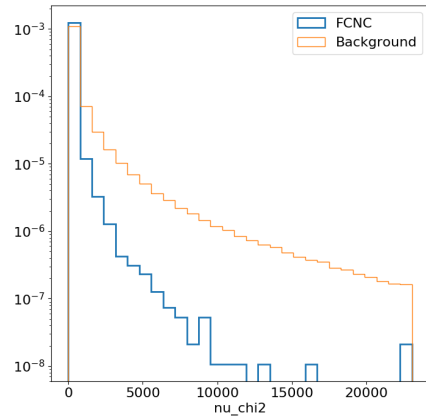
Figure 31: Normalized variables showing the shapes of neural network input variables for the μ+jets channel: [E (bjet), η_b, ΔR_{jb}, E (light jet), light jet η, and χ_ν^2 the total χ^2 fit value

(a) lepton p_T

(b) lepton η

(c) lepton isolation

(d) χ^2_{bW}
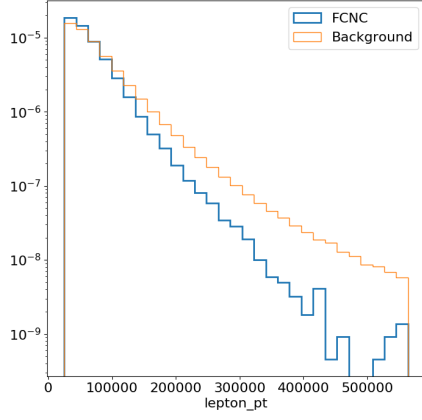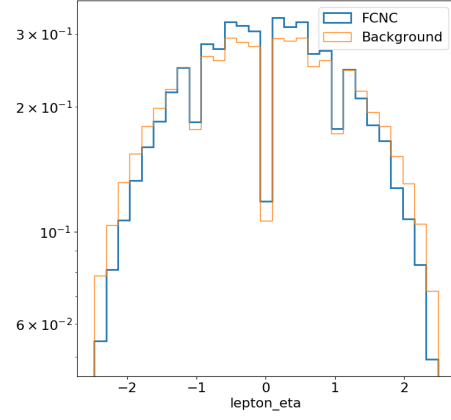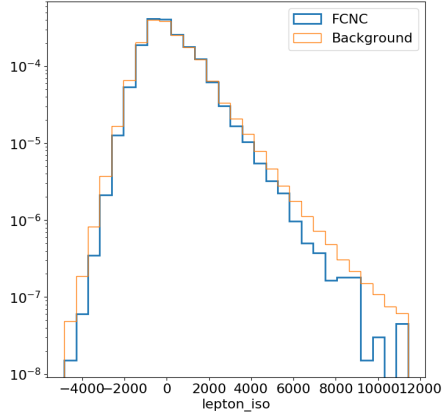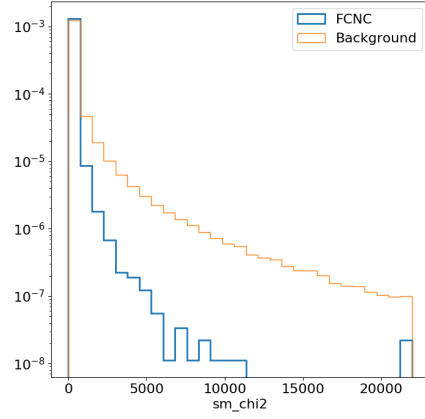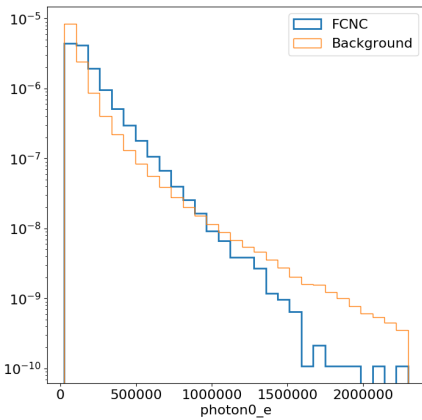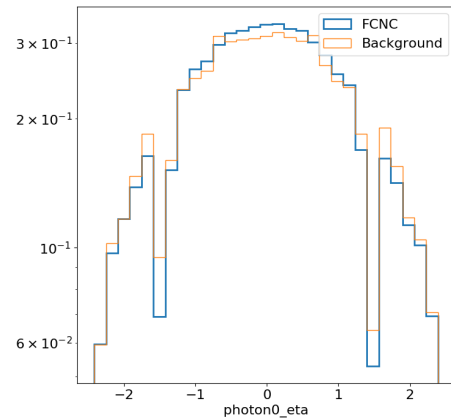
(e) photon E

(f) photon η

Figure 32: Normalized variables showing the shapes of neural network input variables for the μ+jets channel: [lepton p_T, lepton η, lepton isolation , chi^2_{SM} the bWχ^2 value from neutrino reconstrucion ,photon E, and photon η.

(a) $\Delta R_{l\nu}$
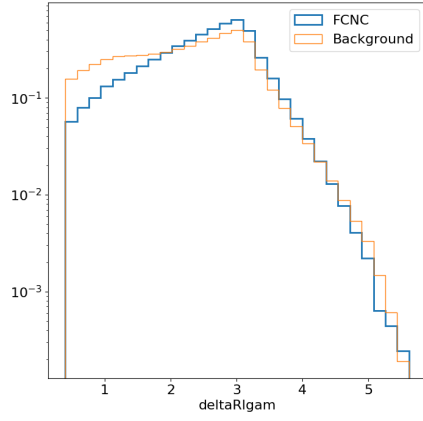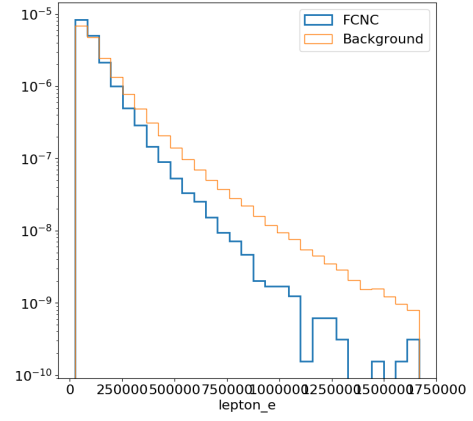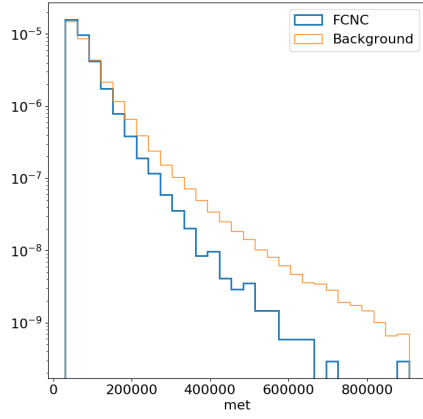
(b) E (lepton)

(c) $\not{E}_T$
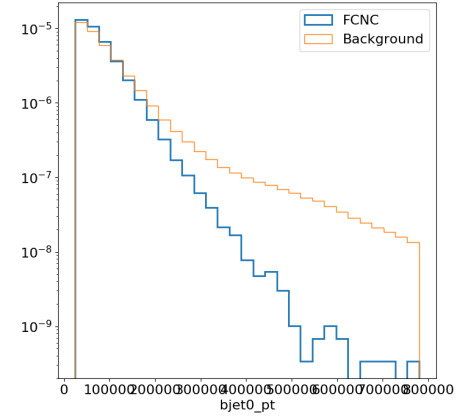
(d) $p_T(b)$

Figure 33: Normalized variables showing the shapes of neural network input variables for the e+jets channel: $\Delta R_{l\gamma}$, E (lepton), $\not{E}_T$, and $p_T(b)$

(a) γ_{iso} topo$E_T$cone40

(b) γ_{p_T}

(c) $m_{q\gamma}$
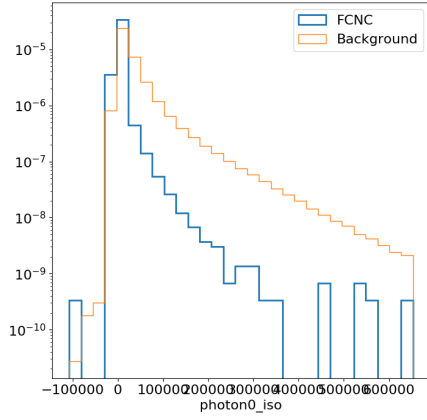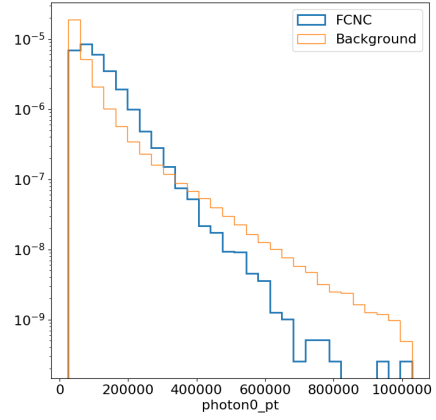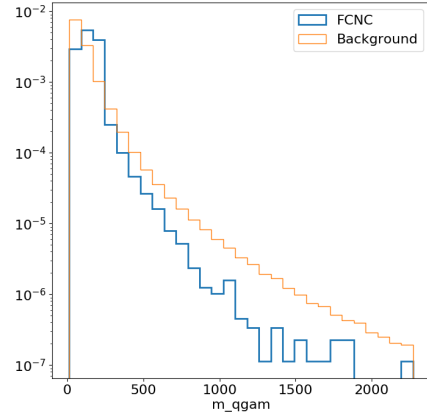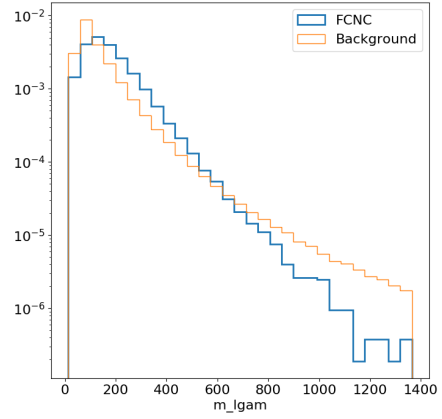
(d) $m_{l\gamma}$
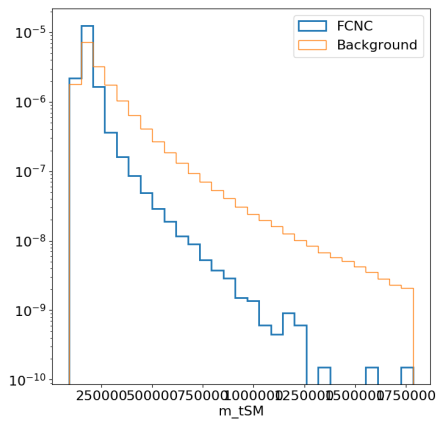
(e) m_{bW}
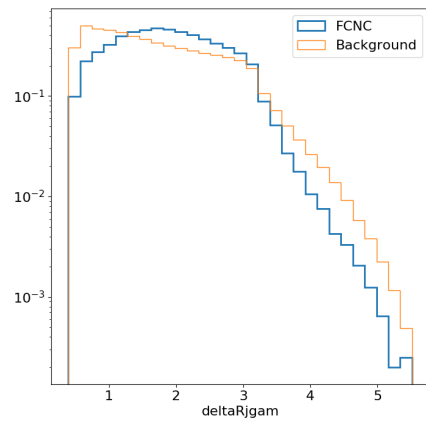
(f) $\Delta R_{j\gamma}$

Figure 34: Normalized variables showing the shapes of neural network input variables for the e+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) $p_T(q)$
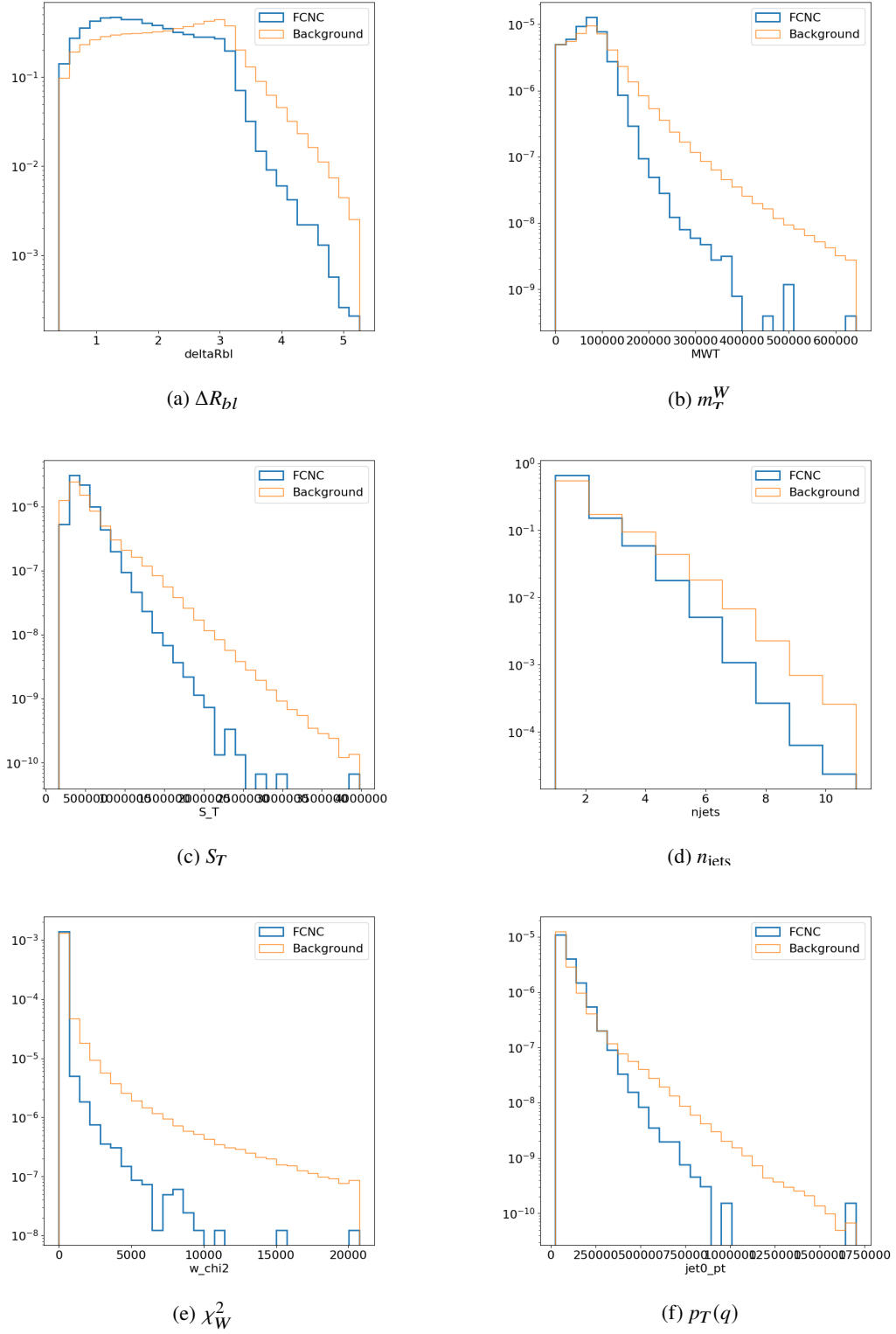
Figure 35: Normalized variables showing the shapes of neural network input variables for the e+jets channel: ΔR_{bl}, m_T^W, S_T, n_{jets}, χ_W^2, and $p_T(q)$
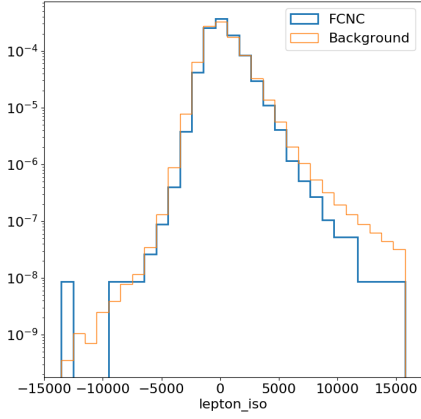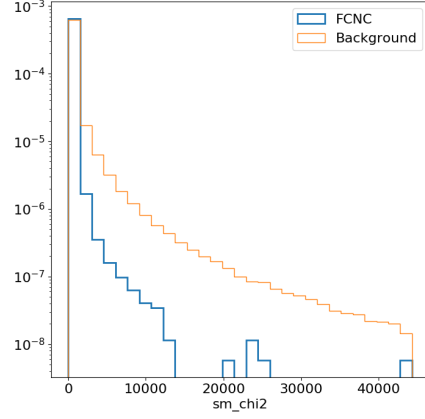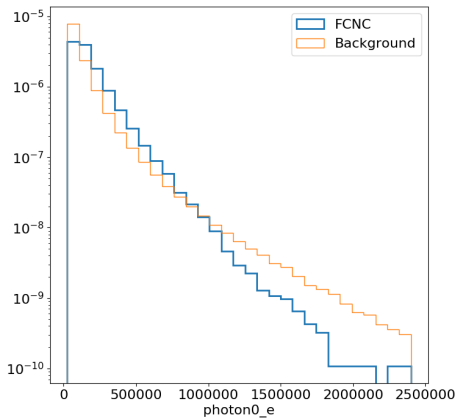
(a) E (bjet)

(b) η_b

(c) ΔR_{jb}

(d) E (light jet)

(e) light jet η

(f) χ^2_ν
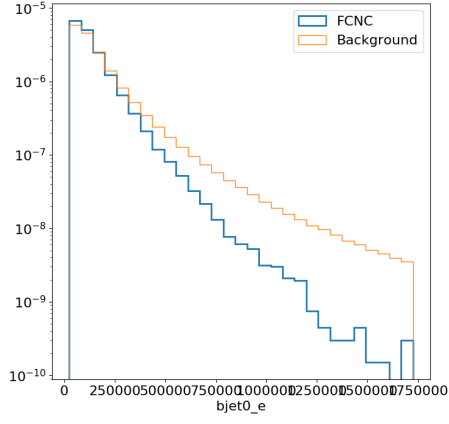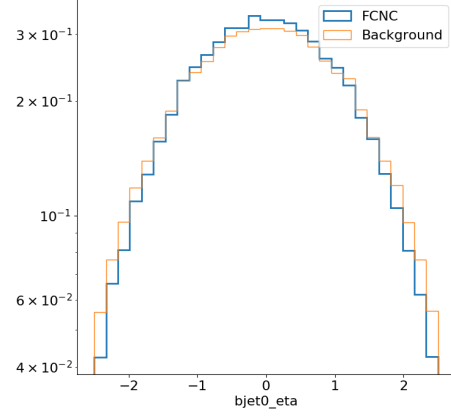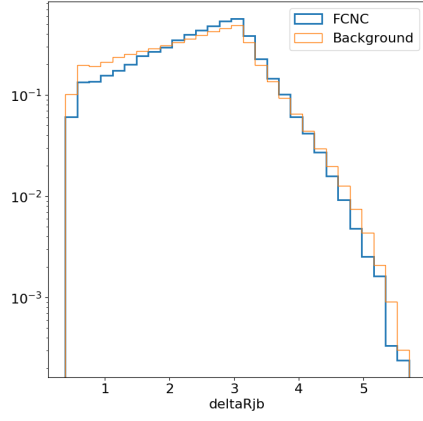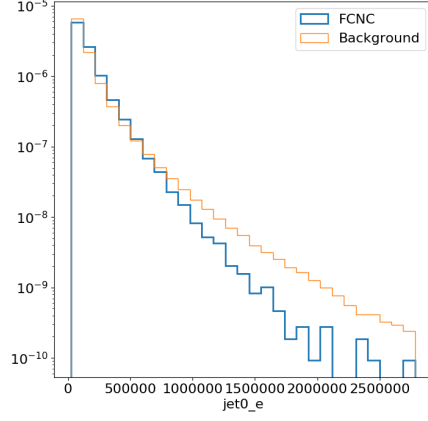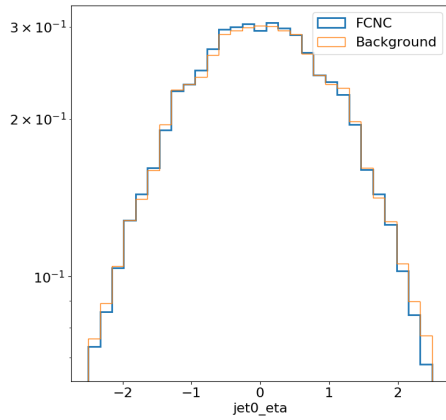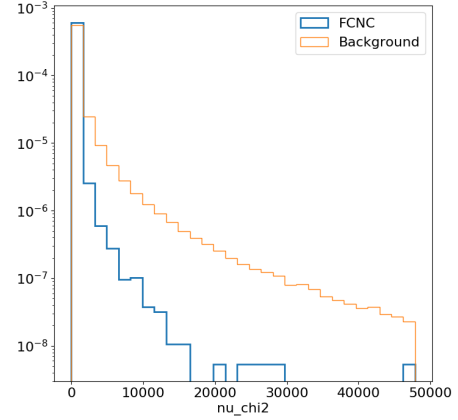
Figure 36: Normalized variables showing the shapes of neural network input variables for the e+jets channel: [E (bjet), η_b, ΔR_{jb}, E (light jet), light jet η, and χ^2_ν the total χ^2 fit value

(a) lepton p_T

(b) lepton η
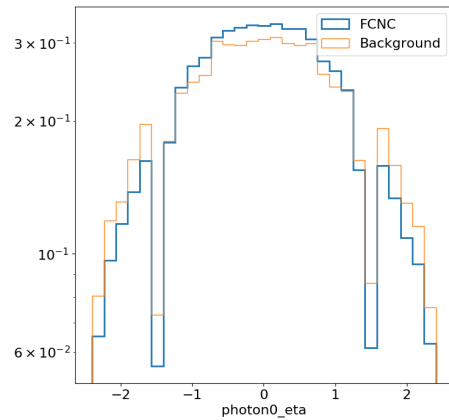
(c) lepton isolation

(d) χ^2_{bW}
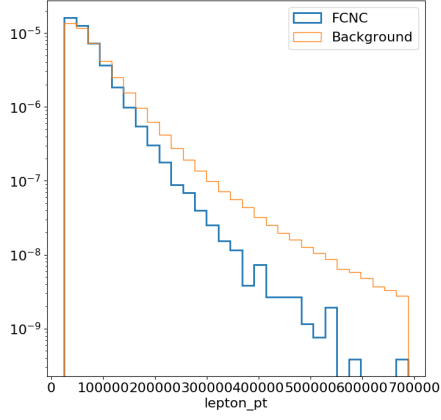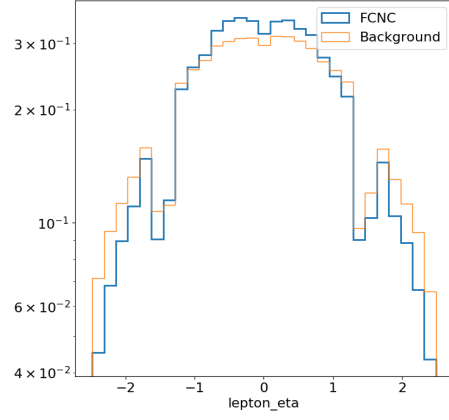
(e) photon E

(f) photon η

Figure 37: Normalized variables showing the shapes of neural network input variables for the e+jets channel: [lepton p_T, lepton η, lepton isolation , chi^2_{SM} the bWχ^2 value from neutrino reconstrucion, photon E, and photon η.