

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

James Brau

Spencer Chang

Dev Sinha

Chair

Advisor

Core Member

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

© 2020 Jason Tyler Barkeloo
This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs (United States) License.

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This dissertation includes previously published and unpublished co-authored material.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 Science, Physics, 2012, Miami University
Bachelor of Science, Physics, 2010, Wittenberg University

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

Chapter	Page
I. SEARCH STRATEGY	1
1.1. Major Backgrounds	1
1.1.1. Overlap Removal	2
1.1.2. Duplicate Event Removal	3
1.2. Event Selection	3
1.3. Event Classification: Neural Network Optimization	4
1.3.1. Input Variables	5
1.3.2. Architecture	11
Training and Validation of Neural Networks	14
1.3.3. Hidden Layer Studies	16
1.3.4. B-Tagging Working Point Studies	20
1.4. Initial Event Selection	23
1.4.1. Triggers	24
1.5. Data and MC Pre-Selection Cuts	25
1.6. Background Evaluation: Control and Validation Regions	25
1.6.1. Backgrounds Without Photons	25
1.6.2. Background With Photons	26
1.7. Fake Rates	26
1.7.1. Jet \rightarrow Lepton Fakes	26
1.7.2. Electron \rightarrow Photon Fakes	26

Chapter	Page
1.7.3. Jet \rightarrow Photon Fakes	26
1.8. Signal Region	26
II. ANALYSIS AND RESULTS	27
2.1. Uncertainties	27
2.2. Statistical Treatment of Results	27
2.3. Limit on Branching Ratio $t \rightarrow q\gamma$	27
III. COMPLEMENTARY SEARCHES AND OUTLOOK	28
3.1. Comparison with Complementary Searches	28
3.2. Future Directions	28
3.3. Conclusion	28
REFERENCES CITED	29

LIST OF FIGURES

Figure		Page
1.1.	Normalized variables showing the shapes of neural network input variables for the μ +jets channel: γ_{iso} , $\text{topo}E_{T\text{cone}40}$, γ_{pT} , $m_{q\gamma}$, $m_{l\gamma}$, m_{bW} , and $\Delta R_{j\gamma}$	8
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 $p_T(q)$	9
1.3.	Normalized variables showing the shapes of neural network input variables for the μ +jets channel: $\Delta R_{l\gamma}$, E (lepton), \cancel{E}_T , and $p_T(b)$	10
1.4.	Pictorial representation of neural network architecture with 3 input variables, 2 hidden layers with 4 nodes each, and 1 output layer.	13
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.	16
1.6.	Accuracy plots for both channels for the 2 hidden layer neural network	17
1.7.	Loss plots for both channels for the 2 hidden layer neural network	18
1.8.	Normalized neural network output signal and background distribution plots are shown for both search channels for a varying number of hidden layers.	19
1.9.	Significance plots for both channels for the 2 hidden layer neural network.	19
1.10.	Accuaracy and loss plots for the e +jets channel at 70%, 77%, and 85% b-tagging working points.	21
1.11.	Accuaracy and loss plots for the μ +jets channel at 70%, 77%, and 85% b-tagging working points.	21
1.12.	Neural network output and significance plots for the e +jets channel at 70%, 77%, and 85% b-tagging working points.	22

Figure	Page
1.13. Neural network output and significance plots for the μ +jets channel at 70%, 77%, and 85% b-tagging working points.	22

LIST OF TABLES

Table		Page
1.1.	Separation of normalized variables between signal and background in the e+jets and μ +jets channels for the variables used as input to the final neural network.	6
1.2.	Branching ratio values with a significance of 2 after neural network optimization	20
1.3.	The electron trigger requirements in the event selections	24
1.4.	The muon trigger requirements in the event selections	25

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, $t\bar{t} \rightarrow bl\nu q\gamma$. 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 proceses, $t\bar{t}$ 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:

$t\bar{t} + \gamma$, $W + \text{jets} + \gamma$, $Z + \text{jets} + \gamma$ 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 1.7.1.

Fake Photons: The number of events with fake photons are estimated using a $Z \rightarrow e^+e^-$ tag-and-probe method discussed in Section 1.7.2. 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

neutrino, and a large transverse W mass. The transverse W mass requirement selects events that have a lepton and \cancel{E}_T which are consistent with a leptonically decaying W boson, defined as:

$$m_T^W = \sqrt{2 \times p_{T_l} \times \cancel{E}_T \times (1 - \cos(\phi_l - \phi_{\text{MET}}))}$$

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

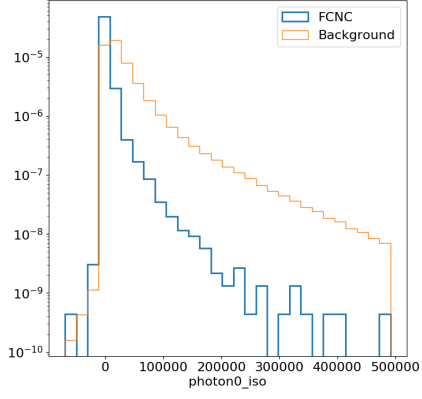
$$\text{Separation} = \sum_i^{\text{bins}} \frac{n_{si} - n_{bi}}{n_{si} + n_{bi}}$$

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
\cancel{E}_T	0.47	0.70
$p_T(b)$	0.51	0.53

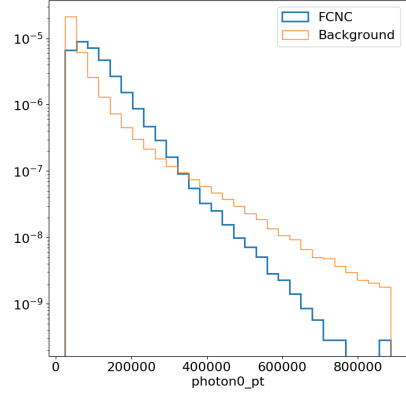
TABLE 1.1. Separation of normalized variables between signal and background in the e+jets and μ +jets channels for the variables used as input to the final neural network.

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 perform 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 $m_{l\gamma}$.

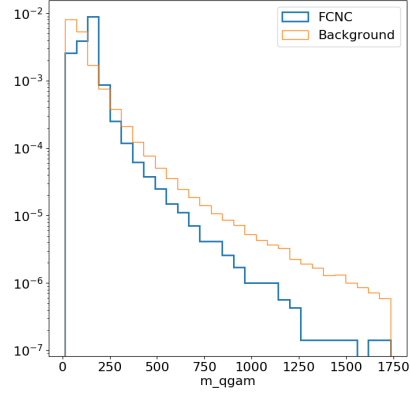
The neural networks are trained on MC events that have a chance of being in the signal region after basic event level cuts and optimized for signal significance. Only events with 1 photon (> 15 GeV) and 1 bjet (MV2c10 77% working point) are classified by the neural network. The 77% working point was 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



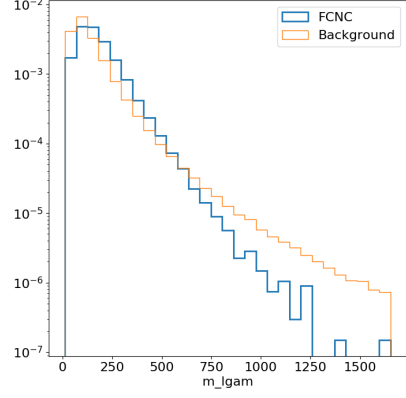
(a) $\gamma_{iso} \text{ topo}E_{T\text{cone}40}$



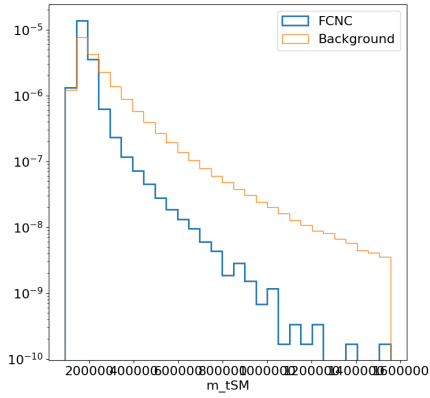
(b) γ_{p_T}



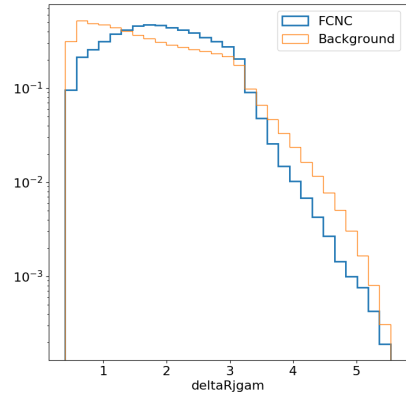
(c) $m_{q\gamma}$



(d) $m_{l\gamma}$

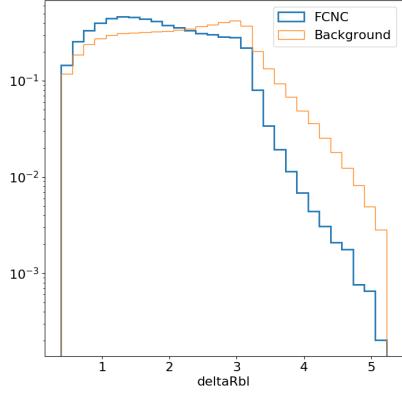


(e) m_{bW}

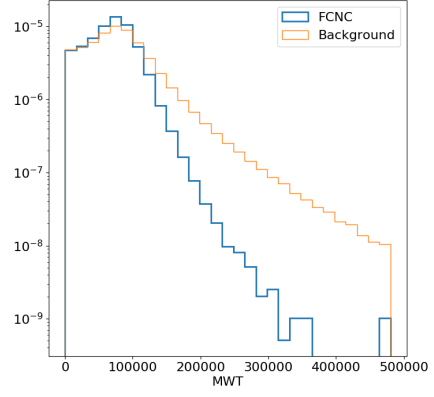


(f) $\Delta R_{j\gamma}$

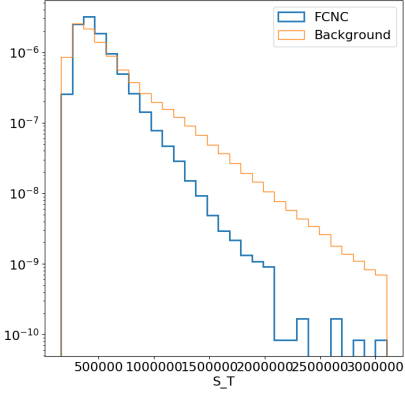
FIGURE 1.1. Normalized variables showing the shapes of neural network input variables for the μ +jets channel: $\gamma_{iso} \text{ topo}E_{T\text{cone}40}$, γ_{p_T} , $m_{q\gamma}$, $m_{l\gamma}$, m_{bW} , and $\Delta R_{j\gamma}$



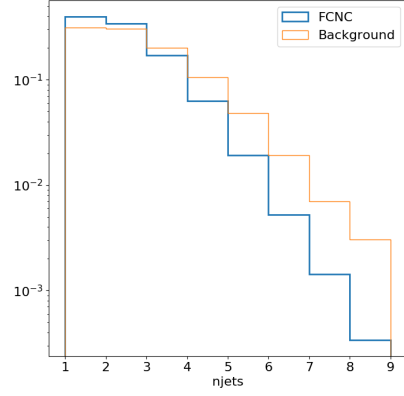
(a) ΔR_{bl}



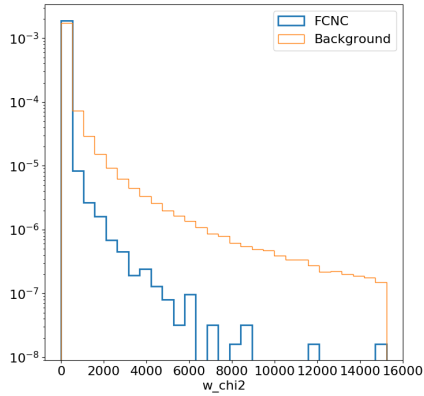
(b) m_T^W



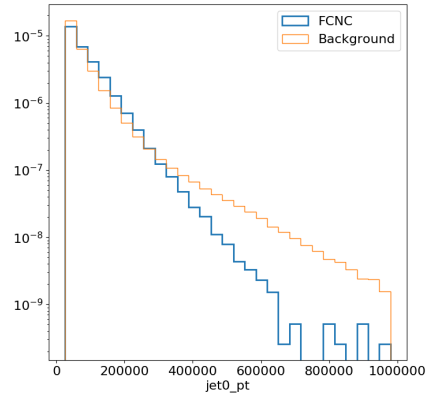
(c) S_T



(d) n_{jets}

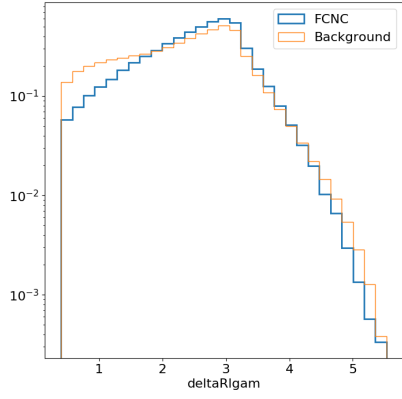


(e) χ_W^2

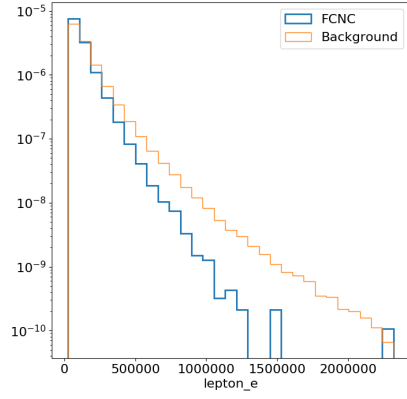


(f) $p_T(q)$

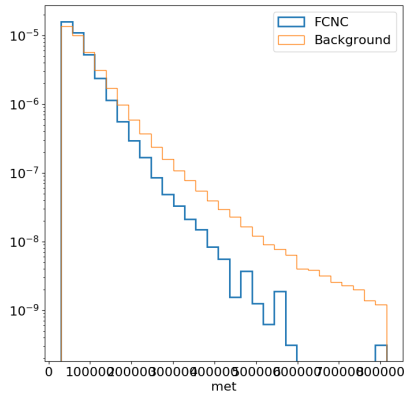
FIGURE 1.2. Normalized variables showing the shapes of neural network input variables for the μ +jets channel: ΔR_{bl} , m_T^W , S_T , n_{jets} , χ_W^2 , and $p_T(q)$



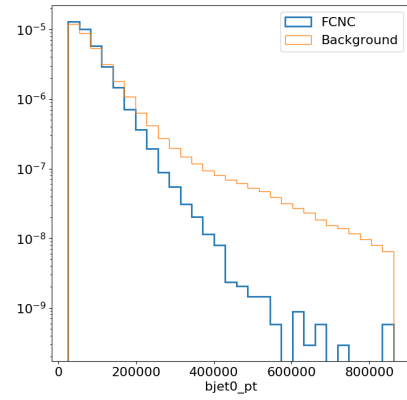
(a) $\Delta R_{l\gamma}$



(b) E (lepton)



(c) \cancel{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), \cancel{E}_T , and $p_T(b)$

1.3.2. Architecture

A variety of architectures of dense neural networks are studied using KERAS[79] on top of the TENSORFLOW backend [80]. 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[81]. 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 applies 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.

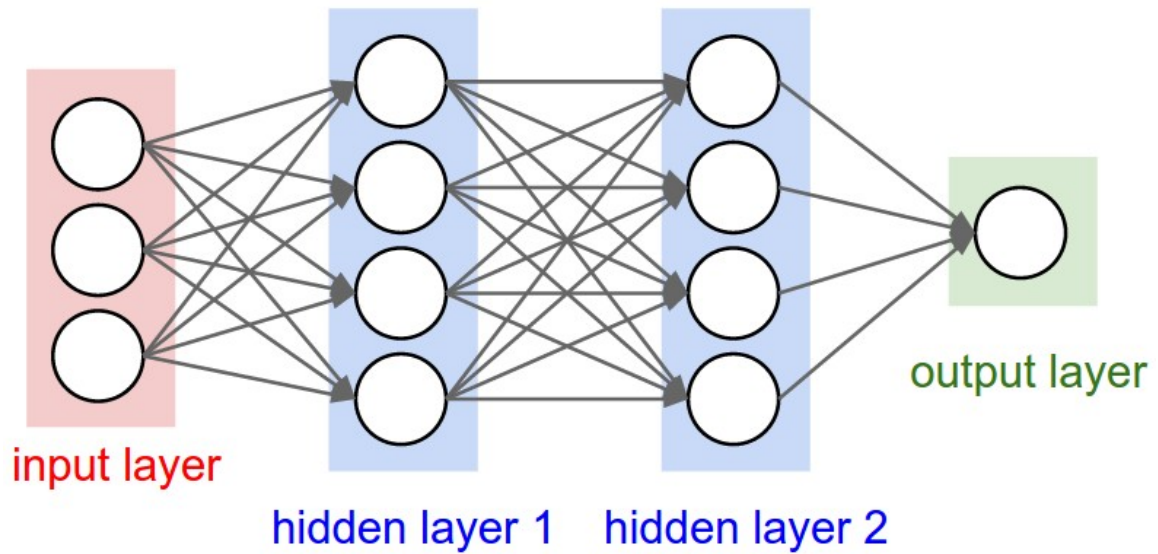


FIGURE 1.4. Pictorial representation of neural network architecture with 3 input variables, 2 hidden layers with 4 nodes each, and 1 output layer[82].

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**[83]. 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 remaining 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.

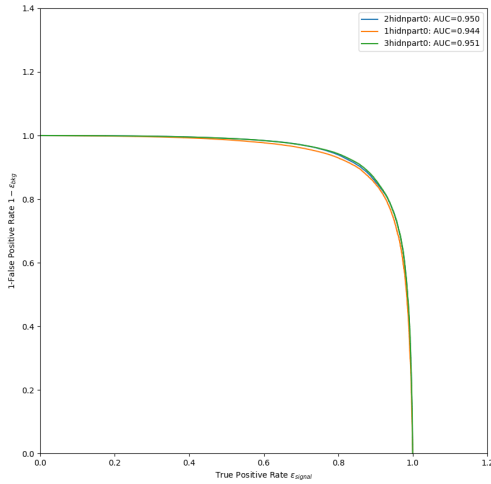
One measure of the performance of the network is the accuracy. The KERAS default accuracy measure is defined:

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

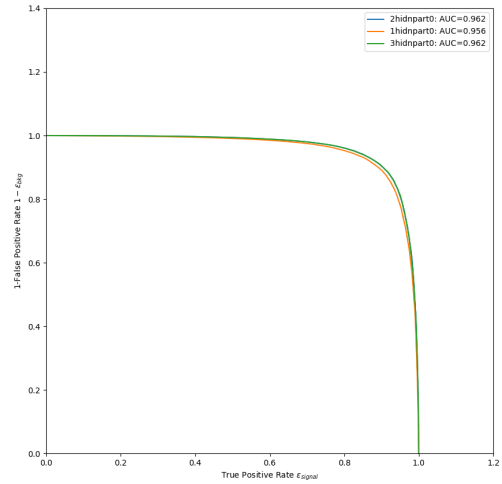
where $N(\text{event}_{NN} \geq 0.5|\text{signal})$ ($N(\text{event}_{NN} < 0.5|\text{background})$) is the number of signal (background) events with $P_{\text{signal}} \geq 0.5$ ($P_{\text{signal}} < 0.5$). Essentially, the accuracy is a measure of the mean of how often correct prediction values occur 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 Characteristic* (ROC) curves are shown in Figure 1.5. The ROC curves show the value of $1 - \epsilon_{\text{bkg}}$ 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

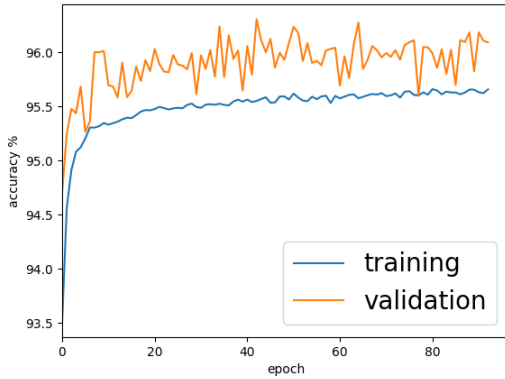


(b) μ +jets ROC Curves

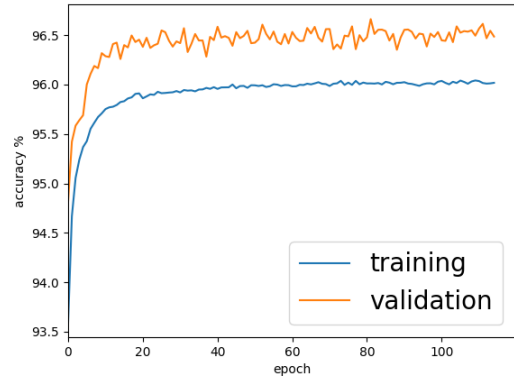
FIGURE 1.5. ROC Curves are shown for both search channels for a varying number of hidden layers. Orange lines correspond to one hidden layer, blue to 2 hidden layers and green to 3 hidden layers. The blue and green curves have near identical AUC values: 0.950 and 0.951 for the e +jets case and 0.962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

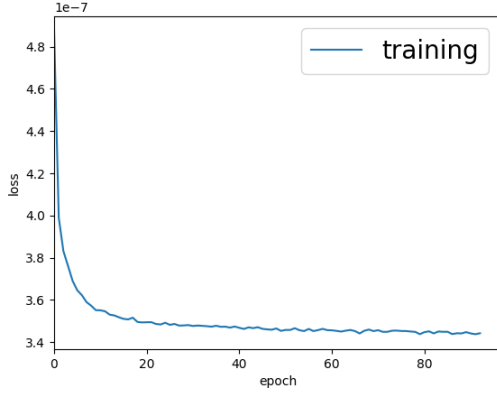


(a) e +jets Accuracy Curves

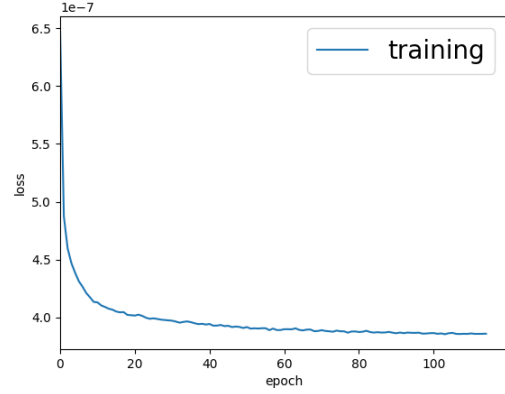


(b) μ +jets Accuracy Curves

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.

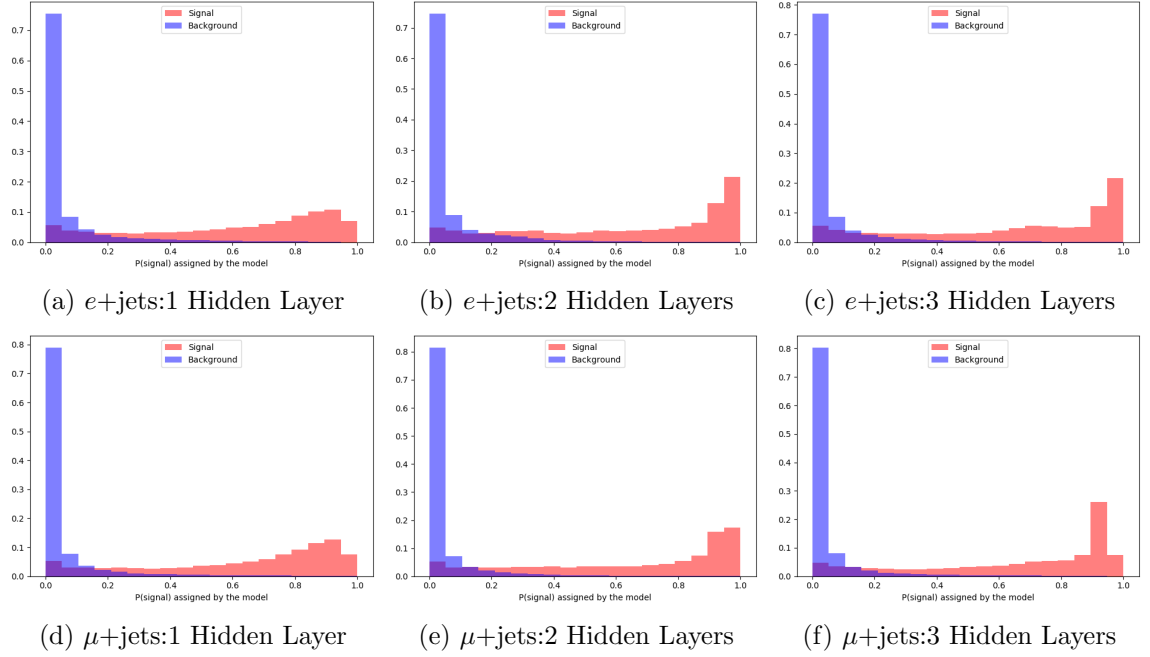


FIGURE 1.8. Normalized neural network output signal and background distribution plots are shown for both search channels for a varying number of hidden layers.

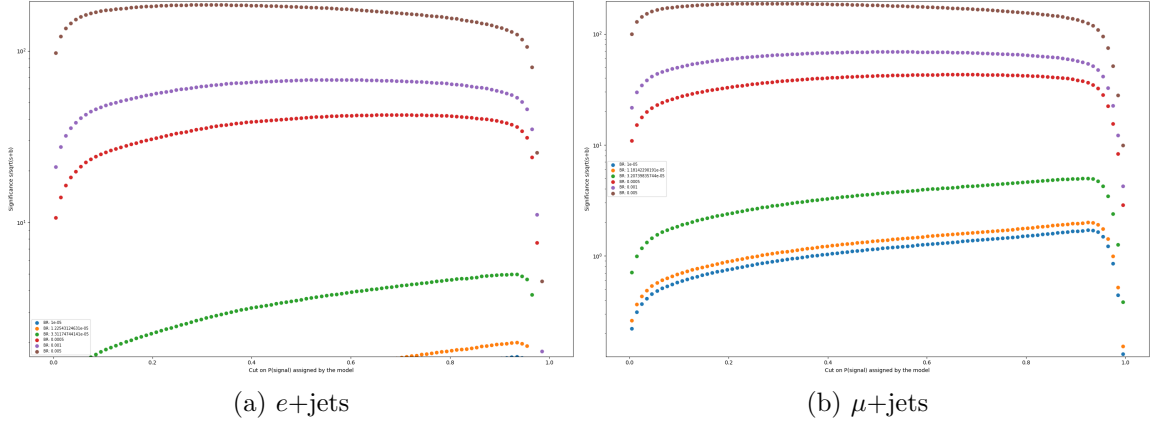


FIGURE 1.9. Significance plots for both channels for the 2 hidden layer neural network. The green points correspond to a branching ratio with a maximum significance of 5, the orange to a maximum significance of 2. The $e+jets$ ($\mu+jets$) branching ratio with max significance of 2 is 1.22×10^{-5} (1.18×10^{-5}). 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.

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 perform 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

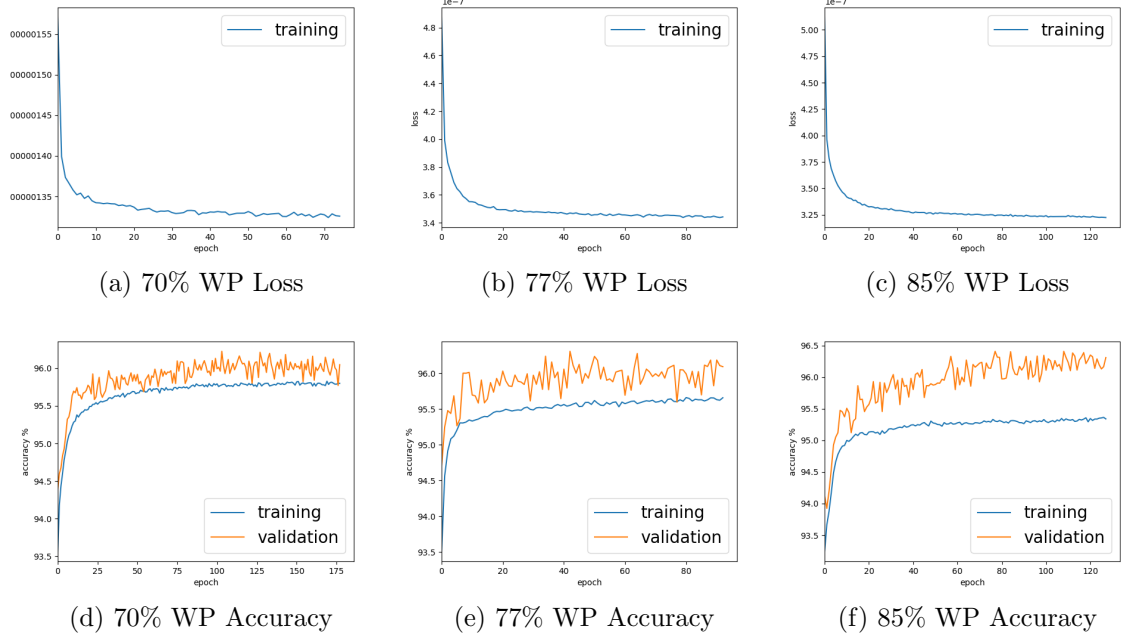


FIGURE 1.10. Accuracy and loss plots for the e +jets channel at 70%, 77%, and 85% b-tagging working points.

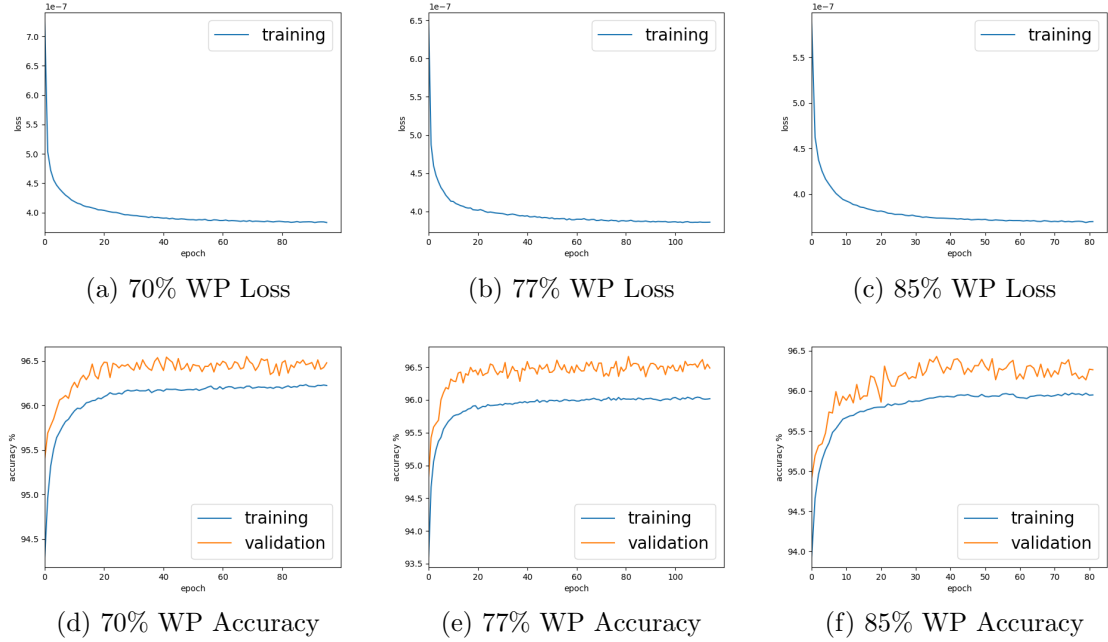


FIGURE 1.11. Accuracy and loss plots for the μ +jets channel at 70%, 77%, and 85% b-tagging working points.

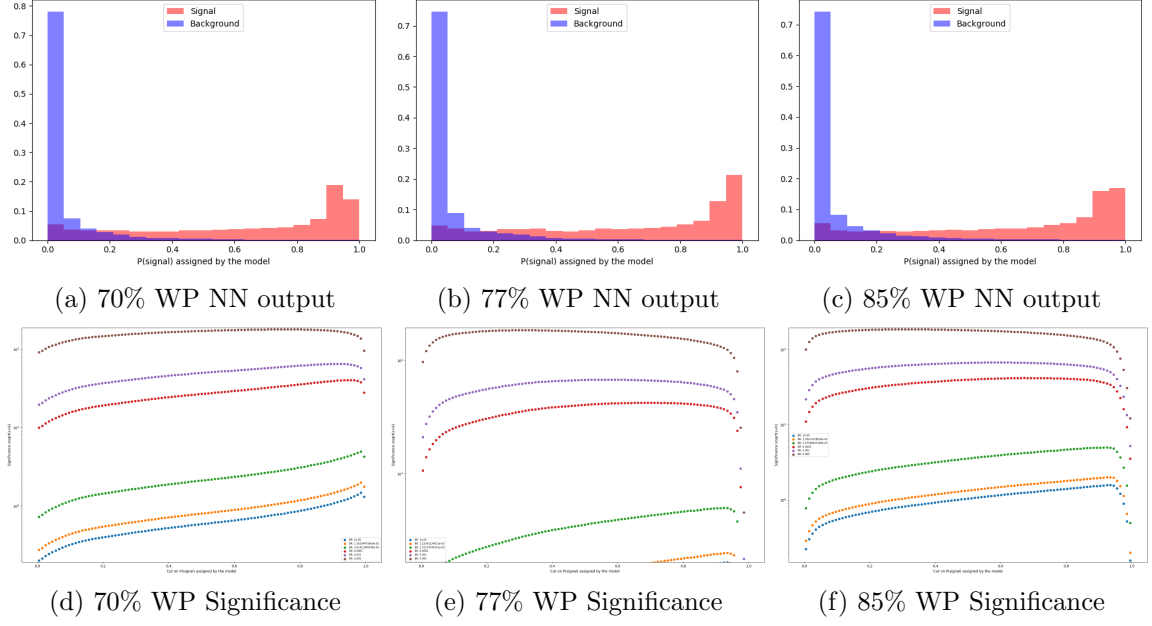


FIGURE 1.12. Neural network output and significance plots for the e +jets channel at 70%, 77%, and 85% b-tagging working points.

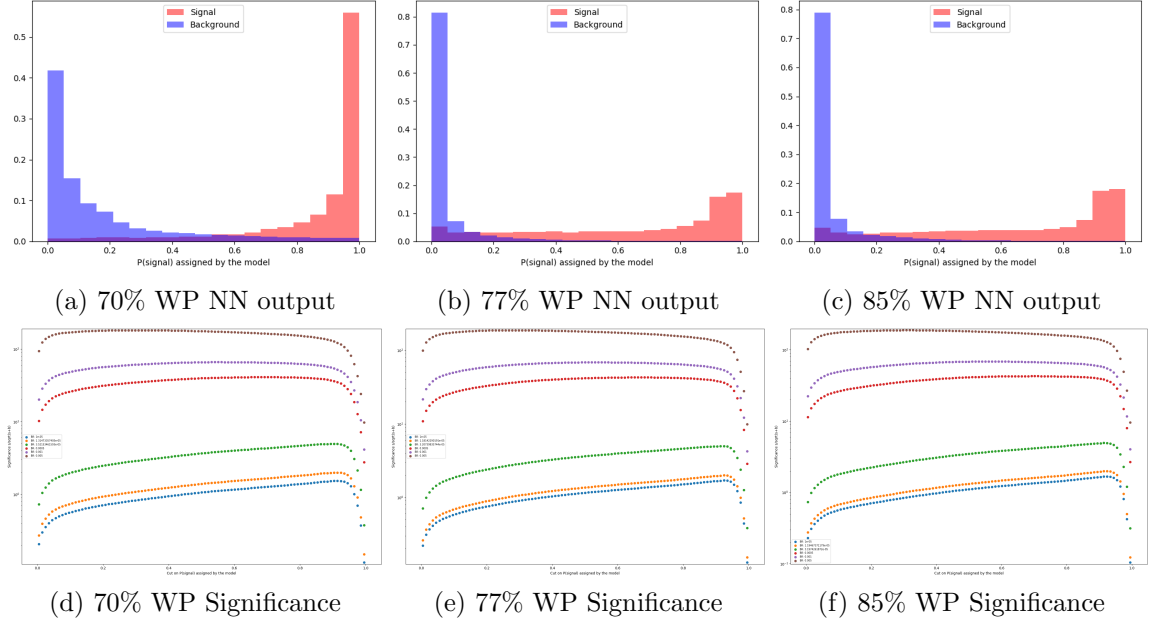


FIGURE 1.13. Neural network output and significance plots for the μ +jets channel at 70%, 77%, and 85% b-tagging working points.

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 noisy 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 occurring 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 primary 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 whether 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 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	-
2016-2018	≥ 26	Tight	Gradient (Loose)	-
	≥ 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 topology 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:

-

Pre Selection Plots

1.6. Background Evaluation: Control and Validation Regions

1.6.1. Backgrounds Without Photons

For scale factors

1.6.2. Background With Photons

1.7. Fake Rates

1.7.1. Jet \rightarrow Lepton Fakes

1.7.2. Electron \rightarrow Photon Fakes

1.7.3. Jet \rightarrow Photon Fakes

1.8. Signal Region

CHAPTER II

ANALYSIS AND RESULTS

2.1. Uncertainties

2.2. Statistical Treatment of Results

2.3. Limit on Branching Ratio $t \rightarrow q\gamma$

CHAPTER III

COMPLEMENTARY SEARCHES AND OUTLOOK

3.1. Comparison with Complementary Searches

3.2. Future Directions

HL-LHC and Beyond Future perspectives at Linear Colliders? - <https://www.sciencedirect.com/>

3.3. Conclusion

REFERENCES CITED

- [1] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.
- [2] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964. [,160(1964)].
- [3] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964. [,157(1964)].
- [4] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.
- [5] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc.*, C680519:367–377, 1968.
- [6] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [7] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.
- [8] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.
- [9] The ATLAS collaboration. Measurement of the top quark mass in the $t\bar{t} \rightarrow$ lepton+jets channel from $\sqrt{s}=8$ TeV ATLAS data. 2017.
- [10] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev.*, D98(3):030001, 2018.
- [11] J. A. Aguilar-Saavedra. Top flavor-changing neutral interactions: Theoretical expectations and experimental detection. *Acta Phys. Polon.*, B35:2695–2710, 2004.
- [12] J. J. Cao, G. Eilam, M. Frank, K. Hikasa, G. L. Liu, I. Turan, and J. M. Yang. SUSY-induced FCNC top-quark processes at the large hadron collider. *Phys. Rev.*, D75:075021, 2007.
- [13] Junjie Cao, Zhaoxia Heng, Lei Wu, and Jin Min Yang. R-parity violating effects in top quark FCNC productions at LHC. *Phys. Rev.*, D79:054003, 2009.
- [14] G. Eilam, J. L. Hewett, and A. Soni. Rare decays of the top quark in the standard and two Higgs doublet models. *Phys. Rev.*, D44:1473–1484, 1991. [Erratum: *Phys. Rev.* D59,039901(1999)].

- [15] Yuval Grossman. Introduction to flavor physics. In *Flavianet School on Flavour Physics Karlsruhe, Germany, September 7-18, 2009*, pages 111–144, 2014. [73(2014)].
- [16] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963. [648(1963)].
- [17] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.
- [18] Lincoln Wolfenstein. Parametrization of the Kobayashi-Maskawa Matrix. *Phys. Rev. Lett.*, 51:1945, 1983.
- [19] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev.*, D2:1285–1292, 1970.
- [20] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.
- [21] S. Abachi et al. Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. Lett.*, 74:2422–2426, 1995.
- [22] LHC Top Working Group. Lhctopwg summary plots, 2019.
- [23] V. M. Abazov et al. A precision measurement of the mass of the top quark. *Nature*, 429:638–642, 2004.
- [24] Markus Cristinziani and Martijn Mulders. Top-quark physics at the Large Hadron Collider. *J. Phys.*, G44(6):063001, 2017.
- [25] David Atwood, Laura Reina, and Amarjit Soni. Phenomenology of two Higgs doublet models with flavor changing neutral currents. *Phys. Rev.*, D55:3156–3176, 1997.
- [26] J. A. Aguilar-Saavedra and B. M. Nobre. Rare top decays $t \rightarrow c \gamma$, $t \rightarrow c g$ and CKM unitarity. *Phys. Lett.*, B553:251–260, 2003.
- [27] J. A. Aguilar-Saavedra. Effects of mixing with quark singlets. *Phys. Rev.*, D67:035003, 2003. [Erratum: *Phys. Rev.* D69,099901(2004)].
- [28] Jin Min Yang, Bing-Lin Young, and X. Zhang. Flavor changing top quark decays in r parity violating SUSY. *Phys. Rev.*, D58:055001, 1998.
- [29] Kaustubh Agashe, Gilad Perez, and Amarjit Soni. Collider Signals of Top Quark Flavor Violation from a Warped Extra Dimension. *Phys. Rev.*, D75:015002, 2007.
- [30] Mark Trodden. Electroweak baryogenesis. *Rev. Mod. Phys.*, 71:1463–1500, 1999.

- [31] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1–102, 2012.
- [32] Andreas Crivellin and Ulrich Nierste. Chirally enhanced corrections to FCNC processes in the generic MSSM. *Phys. Rev.*, D81:095007, 2010.
- [33] Georges Aad et al. Search for flavour-changing neutral currents in processes with one top quark and a photon using 81 fb⁻¹ of *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment. 2019.
- [34] Fabian Kuger. Signal formation processes in micromegas detectors and quality control for large size detector construction for the atlas new small wheel. 08 2017.
- [35] Esma Mobs. The CERN accelerator complex. Complexe des accélérateurs du CERN. Jul 2016. General Photo.
- [36] The ATLAS Collaboration. Luminosity public results run 2, 2019.
- [37] The ATLAS Experiment at the CERN Large Hadron Collider. *J. Instrum.*, 3:S08003. 437 p, 2008.
- [38] M Capeans, G Darbo, K Einsweiler, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne, and R Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010.
- [39] *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997.
- [40] Yosuke Takubo. ATLAS IBL operational experience. *PoS*, Vertex2016:004, 2017.
- [41] Karolos Potamianos. The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider. *PoS*, EPS-HEP2015:261, 2015.
- [42] Bartosz Mindur. ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider. *Nucl. Instrum. Meth.*, A845:257–261, 2017.
- [43] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996.
- [44] *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996.

- [45] ATLAS Outreach. ATLAS Fact Sheet : To raise awareness of the ATLAS detector and collaboration on the LHC. 2010.
- [46] Moritz Backes. The ATLAS Trigger System: Ready for Run-2. *PoS, LeptonPhoton2015:045*, 2016.
- [47] Chaowaroj Wanotayaroj and Jim Brau. Search for a Scalar Partner of the Top Quark in the Jets+MET Final State with the ATLAS detector, Nov 2016. Presented 25 Oct 2016.
- [48] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.
- [49] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.
- [50] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.
- [51] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.
- [52] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [53] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.
- [54] Enrico Bothmann et al. Event Generation with Sherpa 2.2. *SciPost Phys.*, 7:034, 2019.
- [55] Johan Alwall et al. A Standard format for Les Houches event files. *Comput. Phys. Commun.*, 176:300–304, 2007.
- [56] Celine Degrande, Claude Duhr, Benjamin Fuks, David Grellscheid, Olivier Mattelaer, and Thomas Reiter. UFO - The Universal FeynRules Output. *Comput. Phys. Commun.*, 183:1201–1214, 2012.
- [57] R. A. Coimbra, P. M. Ferreira, R. B. Guedes, O. Oliveira, A. Onofre, R. Santos, and Miguel Won. Dimension six FCNC operators and top production at the LHC. *Phys. Rev.*, D79:014006, 2009.

- [58] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek. Dimension-Six Terms in the Standard Model Lagrangian. *JHEP*, 10:085, 2010.
- [59] J. A. Aguilar-Saavedra. A Minimal set of top anomalous couplings. *Nucl. Phys.*, B812:181–204, 2009.
- [60] Zenro Hioki, Kazumasa Ohkuma, and Akira Uejima. Effective Lagrangian Approach to Top Decay via Flavor Changing Neutral Current. In *29th International Conference on Lepton and Photon Interactions (LP2019) Toronto, Ontario, Canada, August 5-10, 2019*, 2019.
- [61] Artur Amorim, Juan Pedro Araque, Nuno Filipe Castro, Jose Santiago, and Rui Santos. Production of $t\gamma$, tZ and tH via Flavour Changing Neutral Currents. *PoS*, TOP2015:067, 2016.
- [62] Celine Degrande, Fabio Maltoni, Jian Wang, and Cen Zhang. Automatic computations at next-to-leading order in QCD for top-quark flavor-changing neutral processes. *Phys. Rev.*, D91:034024, 2015.
- [63] Joao Pequena. Event Cross Section in a computer generated image of the ATLAS detector. Mar 2008.
- [64] Morad Aaboud et al. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C79(8):639, 2019.
- [65] Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. *Eur. Phys. J.*, C74(10):3071, 2014.
- [66] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV, Aug 2016.
- [67] Sébastien Rettie. Muon identification and performance in the ATLAS experiment. *PoS*, DIS2018:097, 2018.
- [68] Morad Aaboud et al. Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016. *Eur. Phys. J.*, C79(3):205, 2019.
- [69] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008.
- [70] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. In *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pages 270–279, 1998.

- [71] Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector. Technical Report ATLAS-CONF-2015-029, CERN, Geneva, Jul 2015.
- [72] The ATLAS collaboration. Tagging and suppression of pileup jets. 2014.
- [73] S. Schramm. Howtocleanjets2016 : Jet-based event cleaning, 2016.
- [74] Albert M Sirunyan et al. Measurement of b hadron lifetimes in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C78(6):457, 2018. [Erratum: *Eur. Phys. J.*C78,no.7,561(2018)].
- [75] N. Bartosik. File:b-tagging diagram.png, 2016.
- [76] Francesco Armando Di Bello. Optimisation of the ATLAS b -tagging algorithms for the 2017-2018 LHC data-taking. *PoS*, EPS-HEP2017:733, 2017.
- [77] Morad Aaboud et al. Measurements of b -jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. *JHEP*, 08:089, 2018.
- [78] Morad Aaboud et al. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C78(11):903, 2018.
- [79] F. Chollet et al. Keras: The python deep learning library, 2019.
- [80] Martín Abadi et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016.
- [81] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. 2014.
- [82] A. Karpathy. Cs231n convolutional neural networks for visual recognition, 2015.
- [83] Fabian Pedregosa et al. Scikit-learn: Machine Learning in Python. *J. Machine Learning Res.*, 12:2825–2830, 2011.