

PAPER • OPEN ACCESS

Application of pivoting adversarial networks in search for four top quark production in CMS

To cite this article: V Wachirapusitanand *et al* 2019 *J. Phys.: Conf. Ser.* **1380** 012069

View the [article online](#) for updates and enhancements.

You may also like

- [Probing FCNC couplings in single top-quark production associated with a neutral gauge boson in future lepton colliders](#)
Sara Khatibi and Mehrnoosh Moallemi
- [Constraining \$ggtt\$ operators from four-top production: a case for enhanced EFT sensitivity](#)
Cen Zhang and
- [The top quark \(20 years after its discovery\)](#)
E E Boos, O E Brandt, D S Denisov et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Presenting more than 2,400
technical abstracts in 50 symposia



**ECS Plenary Lecture
featuring
M. Stanley Whittingham,**
Binghamton University
Nobel Laureate –
2019 Nobel Prize in Chemistry



Register now!



Application of pivoting adversarial networks in search for four top quark production in CMS

V Wachirapusitanand¹, N Srimanobhas^{1,2} and F Blekman³

¹Department of Physics, Faculty of Science, Chulalongkorn University, 254 Phayathai Rd
Wangmai Pathumwan Bangkok 10330, Thailand

²CERN, European Organization for Nuclear Research, Geneva, Switzerland

³Physics Department, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

E-mail: vichayanun.wachirapusitanand@cern.ch

Abstract. One burden of high energy physics data analysis is uncertainty within the measurement, both systematically and statistically. Even with sophisticated neural network techniques that are used to assist in high energy physics measurements, the resulting measurement may suffer from both types of uncertainties. Fortunately, most types of systematic uncertainties are based on knowledge from information such as theoretical assumptions, for which the range and behaviour are known. It has been proposed to mitigate such systematic uncertainties by using a new type of neural network called adversarial neural network (ANN) that would make the discriminator less sensitive to these uncertainties, but this has not yet been demonstrated in a real-life LHC analysis. This work investigates ANNs using as a benchmark the search for the production of four top quarks, an extremely rare physics process at the LHC and one of the important processes that can prove or disprove the Standard Model. The search for four top quarks in some cases is sensitive to large systematic uncertainties. The expected cross section upper limit for four top quark production is calculated using traditional neural networks and adversarial neural networks based on simulated proton-proton collisions within the Compact Muon Solenoid detector within Large Hadron Collider, and are compared to existing results. The improvement and further considerations to the search for rare processes at the LHC will be discussed.

1. Introduction

The production of four top quarks ($t\bar{t}t\bar{t}$) is one of the important particle production processes to determine whether or not the standard model (SM) can provide precise predictions. However, the measurement of the process suffers from its dominant background, $t\bar{t}$ production. $t\bar{t}t\bar{t}$ production has its cross section, as predicted by SM at NLO and $\sqrt{s} = 13$ TeV, is 9.2 fb. On the other hand, the background process has its cross section, as predicted by SM at NLO and $\sqrt{s} = 13$ TeV, to be 831 pb, which is 90 000 times larger than the cross section of the $t\bar{t}t\bar{t}$ production. To make matters worse, the search also suffers from large uncertainties, both systematically and statistically. Fortunately, most systematic uncertainties are modelled from well-known theoretical uncertainties and estimates, and can be modelled with particle collision simulations, while statistical uncertainties can be diminished by acquiring larger data samples over time. The most recent search conducted by the Compact Muon Solenoid (CMS) collaboration [1] uses a boosted decision tree (BDT), which can be trained to discriminate between $t\bar{t}t\bar{t}$ production and $t\bar{t}$ production, but the BDT itself is not designed to be resilient against systematic uncertainties.



To handle issues with systematic uncertainties, Louppe et al. [2] suggested that a new type of neural network called a pivoted Adversarial Neural Network (ANN) can deliver a discriminator that is resilient to certain systematic uncertainties. This, however, has only been tested with toy examples and has never been used in real-world cases such as a complete analysis at the LHC. In this work, we are applying the adversarial neural network approach to investigate possible gains of pivoted ANNs in the search for $t\bar{t}t\bar{t}$ production, particularly in single lepton channel.

2. Four top quark analysis with traditional neural network

The adversarial neural network approach contains two parts: a discriminator network and an adversarial network. Before training with the adversarial network, it is important to get the structure of the discriminator network right first. In this section, we will go through the design of the discriminator neural network (NN) and use the network to calculate the expected cross section limit and expected significance of the search, and will be compared to previous results.

As with every supervised ML model training, data is required to train the neural network. The data used to train the discriminator network are simulated datasets, the same as used in [1], containing simulated proton-proton collision events at 13 TeV centre-of-mass energy. Events in the dataset are required to contain only a single muon with 7 jets or a single electron with 8 jets and follow the same corrections and basic selections as in [1]. These simulated collision events are created using Monte Carlo techniques and include the simulation of the CMS detector with 2016 condition using Geant4 (v.9.4) [3]. Events simulated from $t\bar{t}t\bar{t}$ production are classified as signal events, while events simulated from $t\bar{t}$ production are classified as background events. There are also smaller background processes which account to approximately 13% of all backgrounds and are not included in the training. Only collision events containing 9 or more reconstructed particle flow (PF) jets [4] in the training dataset are used to train the discriminator network, since they have more signal to total events ratio than events containing fewer jets.

The input variables for the discriminator neural network are derived based on the difference, both kinematic and topological, between $t\bar{t}t\bar{t}$ production and $t\bar{t}$ production. The final list of input variables to be inputted into the discriminator network contains 48 variables, such as HT (scalar sum of the transverse momentum of each reconstructed jet) and 5th and 6th highest jet transverse momentum. Fifteen variables in the list are also used in the previous search mentioned in [1].

Regarding the network structure, the discriminator network contains one batch normalization layer after an input layer, which automatically normalizes the input, allowing us to input the data directly without normalizing the variables first. Several hidden layers and neurons follow the batch normalization layer and are permutated with hyperparameter adjustment procedure, whereas the final hidden layer is set to use tanh activation function. The hidden layers are followed by an output layer with sigmoid activation.

Hyperparameter adjustments are carried out to find the optimal configuration of the hidden layers within the discriminator network. The search considered the number of hidden layers and the number of neurons in each layer. Studies have shown that a configuration with three hidden layers with 200, 200, and 100 neurons, respectively, has the best area under Receiver Operating Characteristic (ROC) curve. This can be considered as a measure of the performance of the network structure.

The final discriminator network, with its structure obtained via hyperparameter adjustments, is used to calculate the NN output for each collision event from $t\bar{t}t\bar{t}$ production and all background processes within the same final states. Histograms of output distributions are then created, categorised by single electron channel and single muon channel, number of particle jets in an event and number of particle jets tagged as originating from bottom quarks. An example of the histograms is shown in figure 1. Each histogram is binned in such a way that each bin contains roughly equal amounts of background events, allowing stable statistical behaviour of

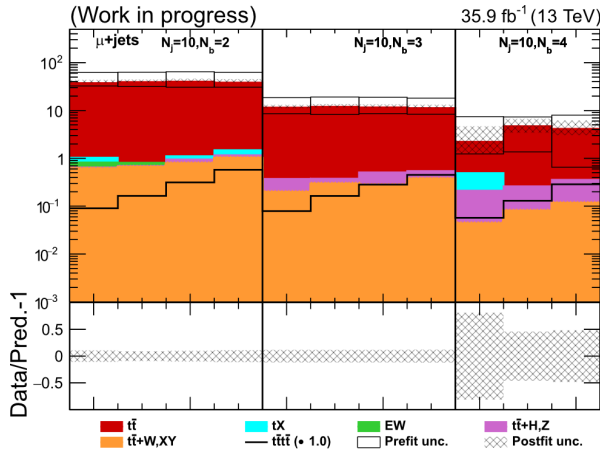


Figure 1. Histogram of NN output distribution in a category with 10 reconstructed jets within single muon channel. The histogram is further divided based on a number of jets tagged as originating from bottom quarks.

Table 1. Expected limits and significance above SM background prediction from a traditional NN and an adversarial neural network (ANN) for single lepton channel, compared to the sensitivity of boosted decision tree (BDT) shown in [1].

	For 35.8 fb ⁻¹ data			For 200 fb ⁻¹ data	
	BDT	NN	ANN	NN	ANN
Expected limit (fb)	90 ⁺⁴² ₋₂₇	78 ⁺³⁷ ₋₂₄	83 ⁺³⁹ ₋₂₅	36 ⁺¹⁷ ₋₁₁	41 ⁺¹⁹ ₋₁₂
Expected significance	0.21	0.25	0.23	0.52	0.46

the simultaneously binned maximum-likelihood fit used to constrain the data.

Numbers of background and signal events from each bin in each populated histogram, as well as systematic uncertainties in each bin, are then used to calculate the expected limits on the cross section and significance above SM background using CL_s method [5]. Table 1 shows the expected limits and significance for the traditional NN discriminator. BDT output values derived in [1] are also shown to compare to results from traditional NN. As table 1 shows, the traditional NN discriminator slightly improves uncertainty on the expected limit and increases the expected significance.

We can also expect to use more data recorded from particle detectors to gain better results, both in terms of expected limit uncertainty and expected significance. Expected limits and significances are also calculated with an adjustment of integrated luminosity to 200 fb⁻¹, which is the approximate size of Run 2 CMS dataset recorded in 2015–2018. As shown in table 1, the expected limit uncertainty decreases and expected significance increases when 200 fb⁻¹ data is used. Again, traditional NN can still deliver a smaller uncertainty range and better expected significance than BDT used in [1].

3. Adversarial neural network outlook

The discriminator network obtained in section 2 only discriminates input events into signal-like and background-like events, and has not been trained with an adversary network. The adversary network, added by the ANN approach, will take the discriminator output as its input and will assess the input event that gives such output. Both networks are trained in turns, where the adversary is trained while the discriminator is locked, and vice versa.

In our application case, we can set both network's loss function, which is a function determining the accuracy of NN predictions, in such a way that the adversary can guess whether or not an event contains a specific systematic uncertainty, and the discriminator must give its

output value so that the adversary cannot do its job. With this way, we will get a distribution of the discriminator output to be the same for events with and without such systematic uncertainty. In our case of training, we aim to train against a normalization uncertainty on events containing additional heavy flavour jets in $t\bar{t}$ production, which is called internally as HeavyFlav.

We also introduced a hyperparameter, called λ , to control the importance of how well the adversarial network assesses the input event. The λ parameter will appear in loss functions to control such behaviour of the network. We have iterated the training with varying values of λ to find the optimal values which cause the adversarial network to be unable to assess whether an event contains the trained uncertainty.

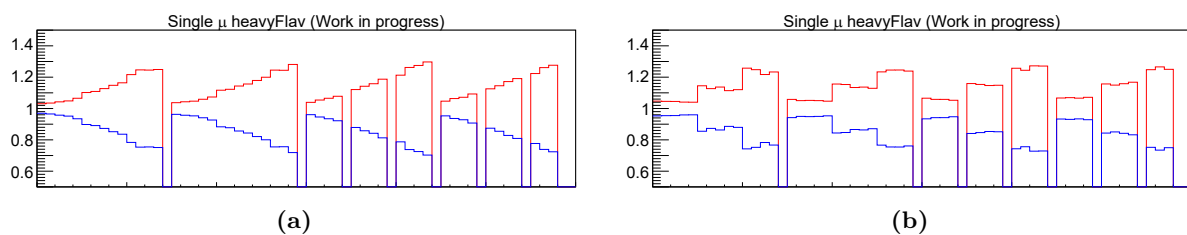


Figure 2. Uncertainty distribution due to heavy flavour modelling as a ratio, calculated with traditional NN (a) and adversary NN (b), as determined in the single muon channel.

After training a discriminator network with an adversary network, using the optimal value of λ , the discriminator can be extracted and used to calculate the output for signal and background events. We observed that the distribution of HeavyFlav uncertainty from the ANN approach becomes flattered (see figure 2), causing the data to be constrained more easily during the calculation of expected limits. Nevertheless, we have found that the expected limit uncertainty from ANN does not improve with our training over the results with traditional NN alone (see table 1). With a reduction in the dependence of the shape of the considered systematic uncertainty, observed in figure 2, future studies may focus on reducing the impact of systematic uncertainties coming from multiple sources.

4. Summary

A discriminator based on the traditional neural network can already give us a better expected limit uncertainty and significance for $t\bar{t}t\bar{t}$ production. The results are more pronounced for large data samples, which is expected to be achieved over time. We expect that, with the correct design of the adversarial neural network, we can use the network to pivot the discriminator network and finally achieve smaller uncertainty on expected limits and better expected significance for $t\bar{t}t\bar{t}$ production searches.

Acknowledgements

This work has been supported in part by the Thailand Research Fund under contract no. MRG5980195, the Chulalongkorn Academic into Its 2nd Century Project Advancement Project (Thailand), and het Fonds Wetenschappelijk Onderzoek ODYSSEUSII programme (Belgium).

References

- [1] CMS Collaboration 2019 arXiv:1906.02805
- [2] Louppe G, Kagan M and Cranmer K 2017 *Adv. Neural Inf. Process. Syst.* **30** ed Guyon I et al (New York: Curran Associates) pp 981–90
- [3] Agostinelli S *et al* 2003 *Nucl. Instrum. Methods Phys. Res. A* **506**(3) 250–303
- [4] CMS Collaboration 2017 *J. Instrum.* **12** 10003
- [5] Cowan C, Cranmer K, Gross E and Vitells O 2010 *Eur. Phys. J. C* **71** 1554