A sensitivity study for asymmetric leptoquark production in the $2\ell SS + 1\tau$ channel
– Summer Student Report

Casey Hampson

CERN Summer Student 2024

Supervisor: André Sopczak

IEAP CTU in Prague

Abstract

This project aims to determine the sensitivity of a novel mechanism to pair produce asymmetric leptoquarks (LQs), i.e. those that are not charge conjugates of each other, within the $2\ell SS + 1\tau$ channel. We briefly outline the phenomenology behind this novel mechanism, then detail the methodology by which LQ events were produced and subsequently tested for sensitivity against many backgrounds.

1 Introduction

1.1 The Leptoquark

The Leptoquark (LQ) is a proposed Beyond the Standard Model (BSM) particle that can couple to both leptons and quarks, thus allowing one to transform into another. Such a particle has been theorized for decades, where it was first found in many Grand Unification Theories (GUT) such as that by Georgi and Glashow in 1974 [1]. Experimental interest in LQ's has been generally unremarkable; however, in recent years, due to the increasing energies of particle collisions at the LHC giving rise to newer discoveries, interest in the LQ search has increased, as they can provide explanations for a number of discrepencies between Standard Model (SM) predictions and experimental results.[1]

Once such example is the decay of the B-meson [2], in which a flavor-changing neutral current process such as $b \to s\ell\ell$ deviates from Standard Model (SM) predictions, signalling a violation of Lepton Flavor Universality (LFU). Another example is that LQ's can radiatively generate Majorana neutrino masses [3]. Both of these phenomena can be explained by considering the leptoquark – clearly, LQ's are very attractive sources of new physics.

A more in-depth discussion of this project with further details can be found in Ref. [4].

LQ's are found as either vector or scalar particles; we will focus solely on the scalar LQ's, as the novel pair production method we will introduce in this report has not been studied for the vector LQ's. There are five multiplets with varying representations under the standard model gauge group $\mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, as shown in Table 1.

1.2 Conventional LQ Production

Current/conventional LQ production at the LHC consists of a number of different mechanisms, largely categorized as either pair production or single/resonant production. There are two main contributions to normal pair production, as shown in Figure 1. This involves a QCD-driven component, as well as a Yukawa-driven component in which a lepton is exchanged in the t-channel. Note that in both cases, the final state LQ's are charge conjugates of each other; in other words, an LQ and its anti-particle are produced.

[1] Of course, these discrepancies are still well within the standard 5σ limit before anything is considered a "discovery;" the standard model is (so far) bulletproof. Nevertheless, small discrepencies *are* present in certain areas more than others, and the LQ is one particle that can resolve such discrepencies.

$\mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$	Symbol	Q-L Chirality	F
$(\bar{\mathbf{3}}, \mathbf{3}, 1/3)$	S_3	LL	-2
$(\mathbf{3}, \mathbf{2}, 7/6)$	R_2	$RL,\ LR$	0
$(\mathbf{3}, \mathbf{2}, 1/6)$	$\tilde{R}_2$	RL	0
$(\bar{\mathbf{3}}, \mathbf{1}, 4/3)$	$\tilde{S}_2$	RR	-2
$(\bar{\mathbf{3}}, \mathbf{1}, 1/3)$	S_1	$LL,\ RR$	-2

Table 1: The representations of the scalar LQ multiplets under the standard model gauge group, the accepted symbols in the literature, the chirality types of the quark and lepton that the LQ couples to, and the fermion number of the LQ.

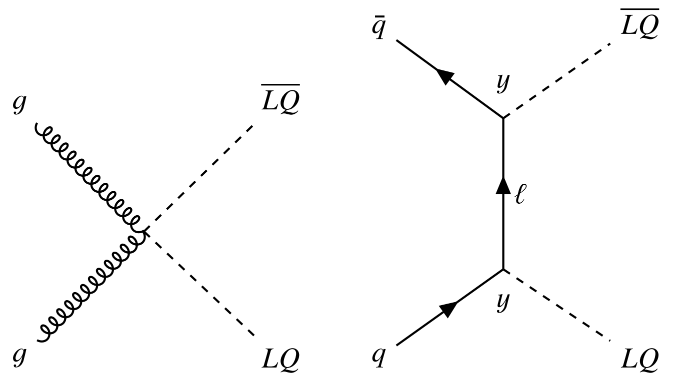

Figure 1: Diagrams contributing to conventional LQ pair production at the LHC at leading order.

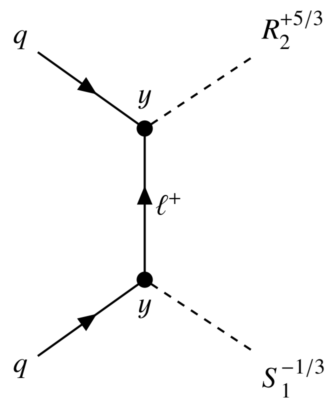

Figure 2: One possible contribution towards the asymmetric pair production of the S_1 and R_2 LQ's. Notably, their charges and fermion numbers are different.

2 Asymmetric Pair Production

In a recent study [5], a novel method for pair producing LQ's that are not charge conjugates of each other, called "asymmetric" production, has been put forward, with the possibility that its cross sections are of similar or higher order than those of the conventional single and pair production methods mentioned in the previous section, for suitable masses and Yukawa coupling magnitudes. This novel method also comes with a few added benefits, such as the possibility for quark-quark initial states as opposed to quark-antiquark initial states, which is an advantage of the LHC due to lessened PDF suppression.

The main requirements to initiate this pair production are that the two leptoquarks in the final state couple to a lepton of the same chirality and flavor, and that their fermion numbers differ. As an example, the R_2 LQ can couple to the left-handed SU(2) quark doublets and the right-handed SU(2) lepton singlets (and vice-versa), and the S_1 LQ, as mentioned before, can couple to both the right and left-handed SU(2) quark and lepton singlets. Because of the similar lepton coupling, we have the ability to asymmetrically produce an R_2 and an S_1 LQ in the final state from a quark-quark initial state, so long as the corresponding Yukawa coupling matrix elements are non-zero. The (only) feynman diagram for the leading order contribution for an S_1/R_2 asymmetric production is given in Figure 2.

2.1 The S_1 and R_2 Case and the $2\ell SS + 1\tau$ Channel

To further examine the S_1/R_2 scheme, we take into consideration the fact that we are interested in the $2\ell SS + 1\tau$ analysis channel, which corresponds to a final state consisting of two same-sign leptons and a hadronically-decaying τ. In order to try and isolate this channel as much as possible, we will try to wisely choose our non-zero Yukawa coupling matrix elements. For instance, we will only consider scenarios in which the LQ couples to a right-handed lepton which eliminates final states containing neutrinos (left-handed neutrinos, to the best of our current knowledge, do not exist). The relevant interaction terms in the Lagrangian are:

$$\mathcal{L}_{\text{int}} = Y^{RR}_{1,ij}\bar{u}^c_i \ell_j S^\dagger_1 + Y^{LR}_{2,ij}\left(\bar{Q}^\intercal_i \ell_j R_2\right) + \text{h.c.} \tag{1}$$

Now, if we expand out the second term into the charge eigenstates of the R_2 multiplet, we get

$$\mathcal{L}_{\text{int}} = Y^{RR}_{1,ij}\bar{u}^c_{R,i} e_{R,j} S^\dagger_1 + Y^{LR}_{2,ij}\bar{u}_{L,j} e_{R,i}(R_2^{+5/3})^* + (Y_2 V_\dagger)^{LR}_{ij}\bar{d}_{L,j} e_{R,i}(R_2^{+2/3})^*, \tag{2}$$

where we have made a couple of important assumptions: 1) the Cabbibo-Kobayashi-Maskawa (CKM) matrix is taken to be the identity, and 2) there is mass degeneracy not only within LQ multiplets but also across multiplets, i.e. all LQs will have the same mass.

Next, we want to isolate LQ decays into heavy particles, which would correspond to having Yukawa coupling matrix elements:

$$Y^{XR}_{i3} \neq 0, \text{ where } X = L, R \text{ and } i = 1, 2, 3. \tag{3}$$

This will allow the LQ to couple to only right-handed, third-generation leptons, which is only the tau. Since we also want a decay into a top quark, we need $Y^{XR}_{33} \neq 0$, but we will let the LQ couple to all quark generations to reduce PDF suppressions. Table 2 contains some cross sections for various masses and Yukawa couplings. As one would intuitively expect, higher mass LQs have a significantly lower cross section, and similarly, a lower coupling strength also decreases the cross section.

m_{LQ}	y	σ
1500 GeV	0.1	1.769×10^{-7}
	0.5	1.105×10^{-4}
	1.0	1.728×10^{-3}
2000 GeV	0.1	2.385×10^{-8}
	0.5	1.492×10^{-5}
	1.0	2.376×10^{-4}
2500 GeV	0.1	3.338×10^{-9}
	0.5	2.089×10^{-6}
	1.0	3.491×10^{-5}

Table 2: The cross sections in picobarns for various mass and Yukawa coupling magnitudes. This corresponds to the Yukawa matrix elements chosen in Equation (3).

The only way to asymmetrically pair-produce the S_1 and R_2 LQ's and have them decay into top quarks and taus is shown in Figure 3. The S_1 LQ will subsequently decay into a tau and a top quark and the R_2 LQ will subsequently decay into an anti-tau and a top quark.

One example decay chain for the taus and top quarks are:

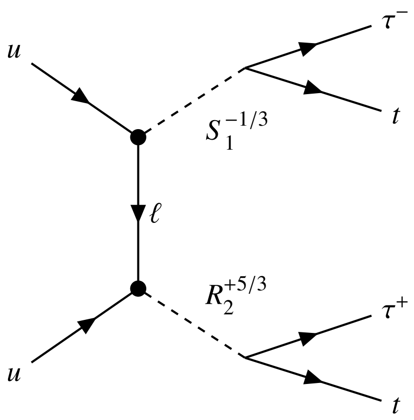

Figure 3: The contribution to $pp \to S_1 R_2 \to t\tau t\tau$ at leading order.

$$S_1 \to t\tau^- \to W^+ b\tau^- \to qqb\tau^- \tag{4}$$
$$R_2 \to t\tau^+ \to W^+ b\tau^+ \to \ell^+ \nu b\tau^+, \tag{5}$$

which gives us a satisfactory final state for our analysis channel.

3 Methodology

As this is a novel production method, there are a number of steps we must take in order to begin analyzing this process. We will make use of the ATLAS experiment's Athena software, which contains a number of helpful tools and wrappers around popular event generators, simulators, and reconstruction algorithms. There are three main steps for the production of the signal in the ATLAS experiment: testing and production of validation plots, approval from subgroup conveners, and submission of a ticket to generate events on the grid.

3.1 Testing and Production of Validation Plots

To start to generate events for our chosen process, we will use MadGraph5_aMC@NLO, a popular program that is used for event generation and cross-section calculation. It comes equipped with many tools for doing these calculations within the regime of the standard model, but for anything outside of this, it requires a supplementary "model" that describes all of the details of the BSM process we want to simulate. The typical pipeline for this is to use the FeynRules package within Mathematica, from which a Universal FeynRules Output (UFO) model is produced that can be imported into MadGraph5_aMC@NLO. Fortunately for us, models have already been made for the scalar LQ's; see Refs. [6, 7]. A scenario with multiple LQ's present requires a simple combination of the individual models, and it is one such combination model that we will use.

The Athena software in the ATLAS experiment provides the `Gen_tf.py` "transform" script, which wraps around MadGraph5_aMC@NLO and Pythia8 (among other generators) and takes in a "JobOptions" file, a pseudo-Python script. The JobOptions file contains all the information relevant for event generation and parton showering. For instance, the desired process is specified using MadGraph5_aMC@NLO syntax, Yukawa couplings, masses, and the number of events are specified, and filters can be added which apply generator-level kinemetic cuts. In the command line, the name of the output EVNT file is specified, which contains the raw event data after the showering is complete.

The EVNT file cannot be directly parsed without considerable effort, so we need to produce so-called "TRUTH derivations". This is done with another one of Athena's transform scripts called `Derivation_tf.py`, and requires only a few command-line inputs such as the input file and the name of the xAOD-formatted output file. There are a few different types of TRUTH formats which contain varying levels of the full truth record, from a direct copy of the EVNT data in xAOD format to a lite version that contains only the absolute necessary information for an analysis.

Once the TRUTH derivations have been made, we can then apply a simple analysis using Athena's EventLoop framework, which is a tool that lets one easily parse through xAOD files and extract relevant information (among many others). Inside this analysis program, we can grab all of the kinematics for all of the final state particles, including the LQ's and their intermediate decay products. All that remains is to use ROOT to place them into histograms, which is a trivial exercise. An example validation plot is shown in Fig. 4.

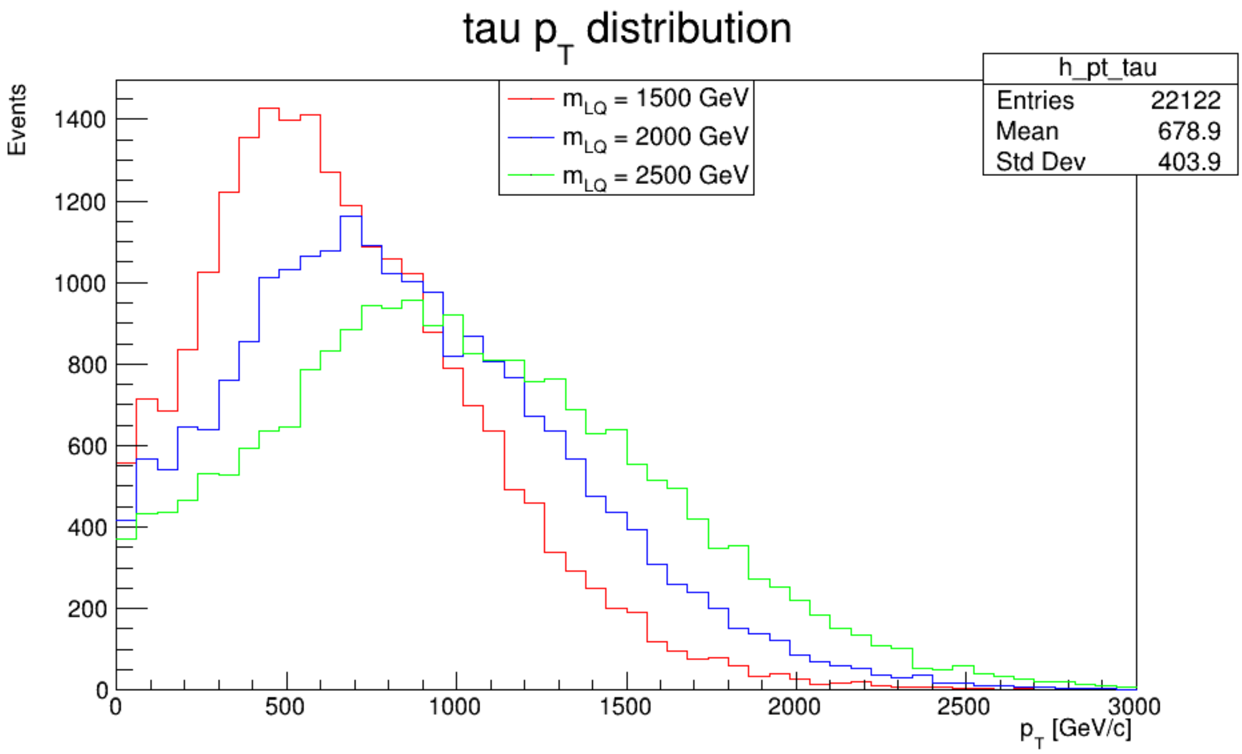

Figure 4: Validation plot of the transverse momentum of the τ generated from test runs for the LQ model.

3.2 Approval and Submission to the Grid

To make statistically valid predictions, we need to produce a very large number of events, more than could be reasonably produced even on a decently powerful single machine. Additionally, if the events are not produced within ATLAS (i.e. on a personal computer), their validity may be called into question. So, our events will need to be produced on ATLAS's central grid, which splits up the jobs and lets them all run in parallel on high-performance machines to improve efficiency. Before this can be done, approval must be granted by the relevent convening groups. In our case, this is the Lepton+X group, of which Tau+X is a subgroup. After approval is granted, we must submit a request with details about this approval, the relevant JobOptions files for each mass (or some other parameter) point, and a few additional details. Then, the ATLAS MC coordinators will run the generation, simulation, and reconstruction.

4 Analysis

There are a number of SM processes that contribute also to the $2\ell SS + 1\tau$ channel, including Higgs production, heavy boson ($W^{\pm}$, Z) production, top quark production, and several others. Because of this, we need to find a way to separate these ordinary SM "backgrounds" from our LQ pair production "signal", and the best way to achieve this is with machine learning. This is because there are dozens of relevent features that are present in the ntuples created from the event generation, and many machine learning models have an edge over conventional fitting algorithms when it comes to many-dimensional scenarios such as this.

This process is done in a number of different ways based on the experiment – we will briefly list our steps here then fully describe them in the subsequent sections. First, we make some preliminary plots/tables using TRExFitter to show the raw number of events with and without weighting/selection criteria. Then, we take the full set of events and transform it to e.g. NumPy arrays so that it can be used to train a machine learning model. After the model is trained, the ntuples are fed through it to generate probabilities in what are called "friend" ntuples. Lastly, TRExFitter can generate full distributions and other statistical measures like cross section limits using the friend ntuples.

4.1 LQ Analysis

4.1.1 Pre-fit Yields

In Fig. 5 we give a couple of the pre-fit yields before the training has been done, specifically the leading τ p_T and the leading jet p_T. The τ p_T plot uses a logarithmic scale due to the large number of low p_T background events and larger number of high p_T signal (LQ) events..

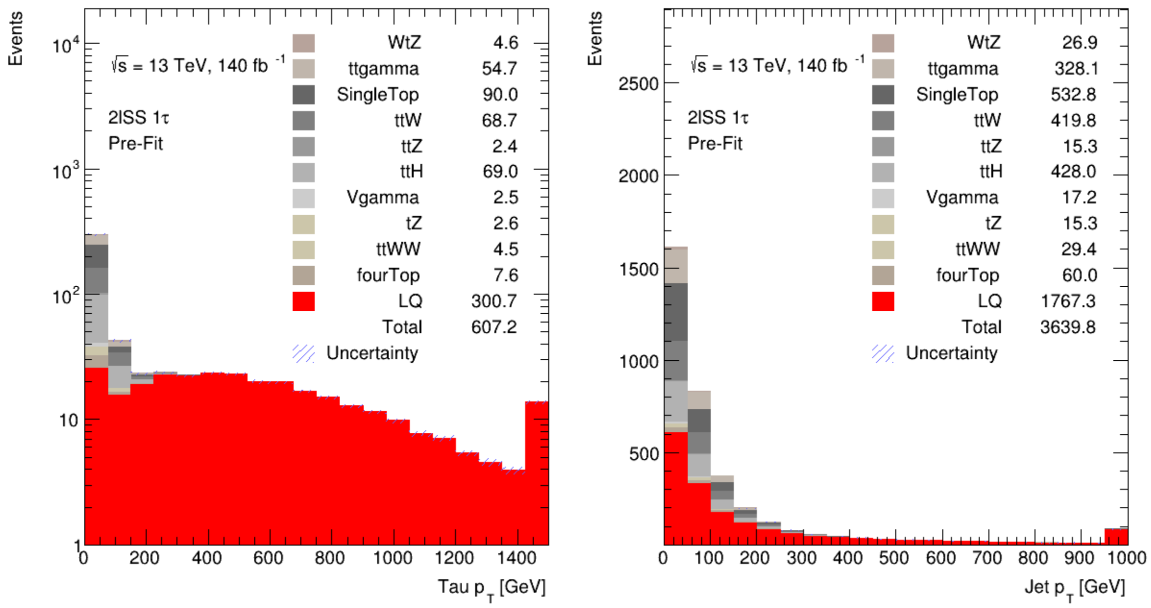

Figure 5: Pre-fit yields for the LQ signal and chosen backgrounds.

4.1.2 Machine Learning

As mentioned above, the next step in the analysis chain was to transform the events in to a machine learning-readable format, which in this case was NumPy arrays. The training was done with a simple

ResNet-6 model (for a description on this model, see [8]). Fortunately, the relatively small size of the ntuples coupled with the fact that the machine doing the training had an NVIDIA 4050 with CUDA installed allowed the entire pipeline, including the subsequent TRExFitter steps, to be done in only a couple of minutes. An example ROC curve for the LQ is given in Figure 6.

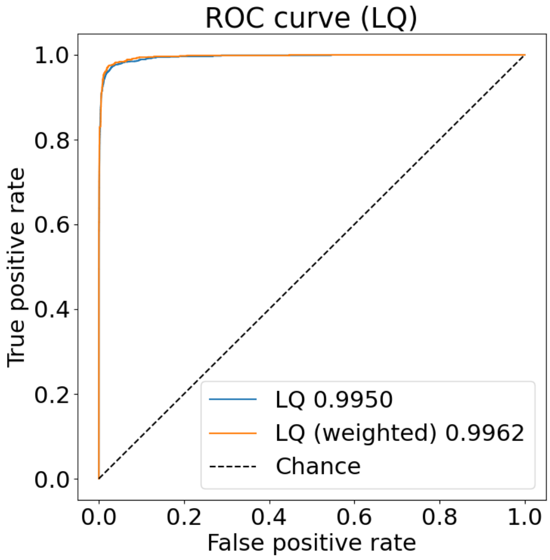

Figure 6: ROC curves for the LQ training for the $t\bar{t}H$ samples (left) and for the LQ samples (right).

4.1.3 Probability Distributions and Norm Factors

The last main step after the training was to pass all of the original ntuples through and create the friend ntuples, which are same as the original ones but with all the other features stripped, leaving only the network probability for each event. In our event generation request, we asked for three different mass points for the LQ, 1500 GeV, 2000 GeV, and 2500 GeV. What we did was train the model the events from all three mass points, then applied it individually to each mass point to generate the output distributions. These are given in Figure 7.

4.1.4 Cross Section Limits

The only other metric we are particularly interested in are the cross section/exclusion limits, which gives a measure of sensitivity for the LQ process. Essentially, this metric gives an upper limit on how large the cross section could be before it can no longer be exclused on the basis of it being statistical noise. In other words, if the true cross section is above the exclusion limit, we would expect to see enough events in the signal region that we couldn't rule out the process. The values at 68% and 95% confidence (1 and 2 σ respectively) and the median limits are given in Table 3.

Mass [GeV]	**1500**	**2000**	**2500**
Limit [pb]	0.00236	0.00265	0.00294
Limit $+1\sigma$ [pb]	0.00422	0.00478	0.00531
Limit $+2\sigma$ [pb]	0.00717	0.00820	0.00913
Limit -1σ [pb]	0.00136	0.00151	0.00167
Limit -2σ [pb]	0.00084	0.00094	0.00104

Table 3: Cross section limit for the three chosen LQ mass points.

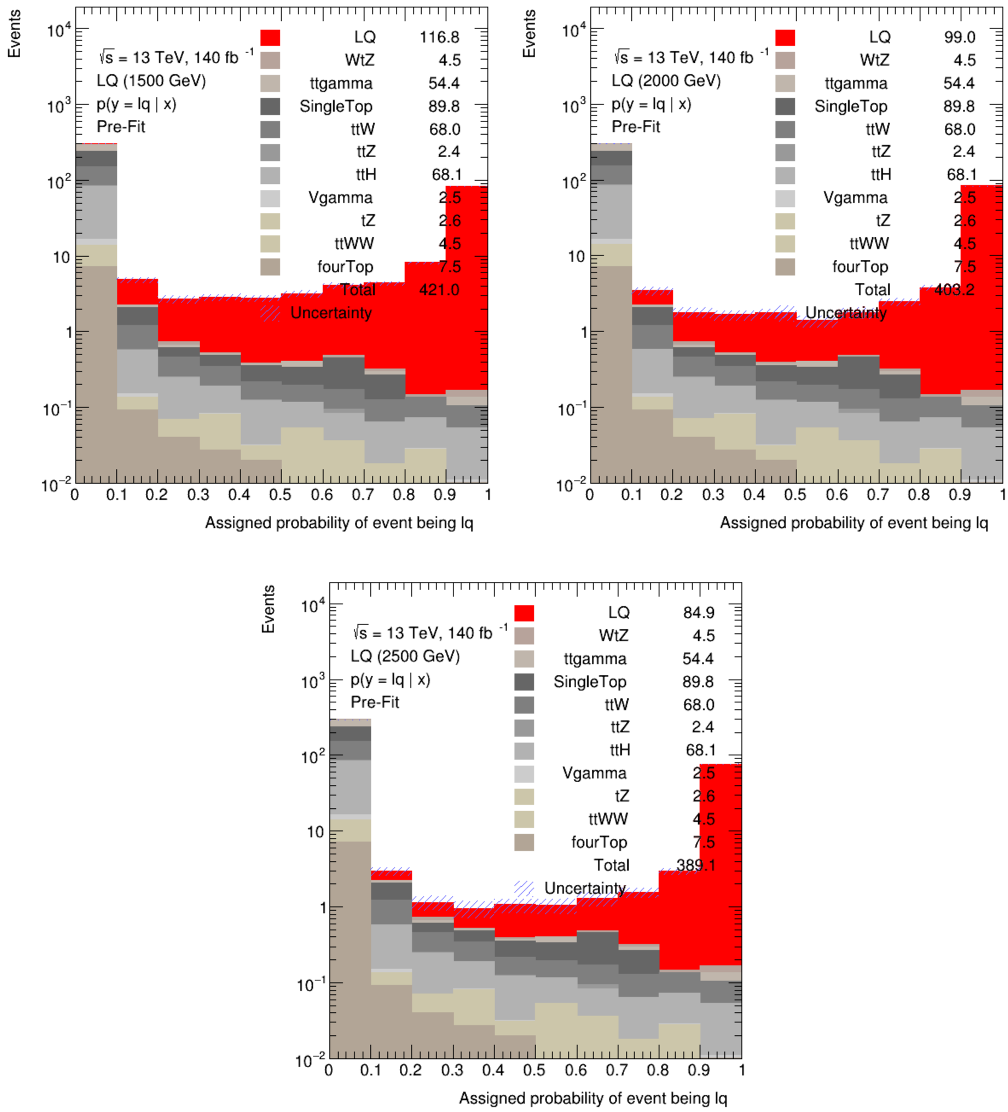

Figure 7: Probability distributions for m = 1500 GeV (left), m = 2000 GeV (middle), and m = 2500 GeV (right).

5 Discussion

For this summer student project, we investigated a novel method of leptoquark pair production which involves two leptoquarks in the final state that are not charge conjugates of each other. Additionally, there is the added benefit of a quark-quark initial state, which is desirable at the LHC. We outlined the steps taken to test the model and submit an official MC request on the grid for larger samples.

We then processed the new LQ events through the FastFrames framework, on which we were able to apply a machine learning setup to analyze the LQ events. We noted that there was very strong separation between the signal and the background, at the cost of the appearance of some over-fitting in a few areas. In future work, a good set of training features could be investigated as well as further work on the FastFrames side of things to provide some more variables with better/more efficient definitions. This could potentially help with some of the over-fitting. Also, a more advanced ResNet model with some fine-tuned hyperparameters can be used which could also alleviate this issue and possible give even better sensitivity.

6 Acknowledgements

I would like to express my deepest gratitude to my supervisor André Sopczak for his invaluable guidance, encouragement, and continuous support throughout this project. I would also like to thank the Conseil Européen pour la Recherche Nucléaire, the CERN & Society Foundation, and the University of Michigan for their support in making this summer student project possible.

References

1. Georgi, H. & Glashow, S. L. Unity of All Elementary Particle Forces. *Phys. Rev. Lett.* **32,** 438–441 (1974).
2. Hiller, G. & Schmaltz, M. R_K and future $b \to s\ell\ell$ physics beyond the standard model opportunities. *Phys. Rev. D* **90,** 054014. arXiv: 1408.1627 [hep-ph] (2014).
3. Mahanta, U. Neutrino masses and mixing angles from leptoquark interactions. *Physical Review D* **62.** ISSN: 1089-4918 (Sept. 2000).
4. Hampson, C. R. & Sopczak, A. *Sensitivity Study for Asymmetric Leptoquark Production in the 2lSS1tau Channel* ATL-COM-PHYS-2025-813. Geneva, 2025. https://cds.cern.ch/record/2943247.
5. Doršner, I., Lejlić, A. & Saad, S. Asymmetric leptoquark pair production at LHC. *Journal of High Energy Physics* **2023.** ISSN: 1029-8479 (Mar. 2023).
6. Doršner, I. & Greljo, A. Leptoquark toolbox for precision collider studies. *Journal of High Energy Physics* **2018.** ISSN: 1029-8479 (May 2018).
7. Doršner, I., Fajfer, S. & Lejlić, A. Novel leptoquark pair production at LHC. *Journal of High Energy Physics* **2021.** ISSN: 1029-8479 (May 2021).
8. He, K., Zhang, X., Ren, S. & Sun, J. *Deep Residual Learning for Image Recognition* 2015. arXiv: 1512.03385 [cs.CV]. https://arxiv.org/abs/1512.03385.
9. Crivellin, A. & Schnell, L. Complete Lagrangian and set of Feynman rules for scalar leptoquarks. *Computer Physics Communications* **271,** 108188. ISSN: 0010-4655 (Feb. 2022).
10. Yazykov, V. *Optimization of the ttH Selection Including Systematics Using Machine Learning with ATLAS Data* Presented 05 Sep 2023 (Prague, Tech. U., 2023). https://cds.cern.ch/record/2871484.