Simulation and Analysis of Novel Leptoquark Pair Production Mechanism

Casey Hampson

August 3, 2024

Contents

1 Introduction

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 [5]. 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.

Once such example is the decay of the B-meson [6], 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 [7]. Clearly, LQ's are very attractive sources of new physics.

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 $SU(3)_c \times SU(2)_L \times U(1)_Y$, as shown in Table 1. Notably, all five are triplets under $SU(3)_c$, which is expected, as they couple to a quark and must therefore carry color. This also means that it is possible for there to be quark-quark interactions with a LQ, however we will not consider this for a similar reason as to why we aren't considering vector LQ's. Each multiplet has a number of charged eigenstates; the R_2 leptoquark, for instance, has two charged eigenstates: $R_2^{+5/3}$ and $R_2^{+2/3}$. This comes from the Gell-Mann-Nishijima formula, in our normalization, given by:

$$Q = I_3 + Y, \tag{1}$$

where Y is the hypercharge of the LQ and I_3 is the third component of the weak isospin of the LQ.

As a scalar particle, the LQ's couple to the quark and lepton via a Yukawa interaction, with the strength of the interaction determined by the magnitude of the coupling constant y. However, here, the Yukawa coupling is actually a 3×3 matrix, whose indices correspond to the fermion generation. As an example, the $LQ - q - \ell$ interaction part of the S_1 Lagrangian is given here [1]:

$$\mathcal{L}_{\text{int}} = Y_{1,ij}^{RR} \bar{u}_i^c \ell_j S_1^\dagger + Y_{1,ij}^{LL} \left(\bar{Q}_i^{c\intercal} i\sigma_2 L_j \right) S_1^\dagger \tag{2}$$

Here, Q_i and L_i are left-handed quark and lepton $SU(2)$ doublets, and u_i and ℓ_i are right-handed $SU(2)$ singlets. The superscripts c and $\intercal$ mean charge conjugation and transposition of the $SU(2)$ doublets, respectively. The parentheses in the second term denote contraction in $SU(2)$ space. Note that there are two terms, corresponding to the two different coupling types for the S_1 LQ, one for coupling of left-handed quarks to left-handed leptons, and one for coupling of right-handed quarks to right-handed leptons. If we were to, for instance, select $Y_{11}^{RR} \neq 1.0$ and all others to zero, then the LQ would couple only right-handed, first-generation quarks to right-handed, first generation leptons (which is the electron, since there are no

$SU(3)_c \times SU(2)_L \times U(1)_Y$	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
$(\bar{\mathbf{3}}, \mathbf{1}, 4/3)$	$\tilde{S}_1$	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.

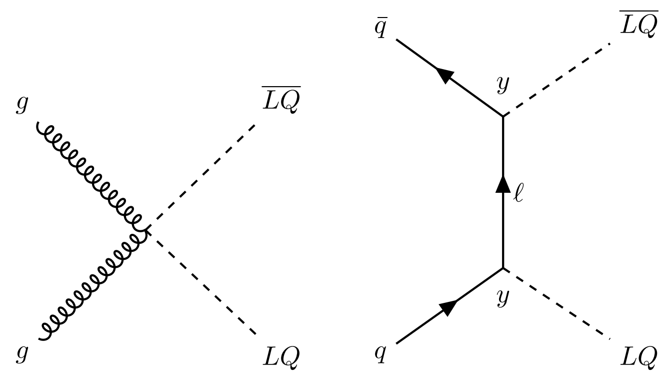

Figure 1: Diagrams contributing to conventional LQ pair production at the LHC at leading order.

right-handed neutrinos). In other words, the branching ratio $\beta(S_1 \to ue) = 1.0$ (note that due to charge conservation, the S_1 LQ cannot decay into an electron and a down-type quark).

1.1 Conventional LQ Production

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.

Based on the contributions to the single production cross section, we have that the amplitudes are proportional to only a single Yukawa coupling. So, the form of the total cross section is given by

$$\sigma_{\text{single}} = f(m_{\text{LQ}})|y_i|^2, \tag{3}$$

where the function f is dependent on the mass of the LQ. On the other hand, the pair production cross section takes a more complicated form:

$$\sigma_{\text{pair}} = f_{\text{QCD}}(m_{\text{LQ}}) + f_{\text{int}}(m_{\text{LQ}})|y_i|^2 + f_{\text{t-chan}}(m_{\text{LQ}})|y_i|^4. \tag{4}$$

In this case, we have the first term arising due to the QCD component, in which there are no fermions and thus no Yukawa coupling dependence. The third term is due to the t-channel diagram as shown before, and therefore that term depends quartic-ly on the Yukawa coupling, and the middle term is the interference between the two. Evidently, this cross section becomes dominated by the QCD contribution for small magnitudes of the Yukawa coupling, but for larger couplings, the quartic term in the pair production cross-section can begin to dominate. However, the mass dependence also plays a larger part in the pair production terms, meaning that the cross section drops off faster for higher LQ masses when compared to single production, hence, in the higher mass regime, pair production no longer dominates.

2 Asymmetric Pair Production

In a recent study [4], 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 particularly preferrable for the LHC due to lessened PDF suppression.

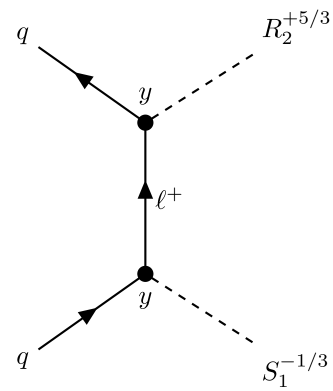

Figure 2: One possible contribution towards the asymmetric pair production of the S_1 and R_2 LQ's. Notably, their charges and fermion numbers are different.

The main requirements to initiate this pair production is 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 the both the right-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 feynman diagram for an S_1/R_2 asymmetric production is given in Figure 2.

2.1 The S_1 and R_2 Case

To examine further the S_1/R_2 scheme, we introduce the interaction terms of the Lagrangian for, say, coupling to right-handed leptons (which, again, excludes the neutrinos and leaves only the charged leptons):

$$\mathcal{L}_{\text{int}} = Y_{1,ij}^{RR} \bar{u}_i^c \ell_j S_1^\dagger + Y_{2,ij}^{LR} \left(\bar{Q}_i^\intercal \ell_j R_2 \right) + \text{h.c.} \tag{5}$$

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

$$\mathcal{L}_{\text{int}} = Y_{1,ij}^{RR} \bar{u}_{R,i}^c e_{R,j} S_1^\dagger + Y_{2,ij}^{LR} \bar{u}_{L,j} e_{R,i} (R_2^{+5/3})^* + (Y_2 V_\dagger)_{ij}^{LR} \bar{d}_{L,j} e_{R,i} (R_2^{+2/3})^*, \tag{6}$$

where now we have added chirality subscripts for the fermion states, LQ charges as superscripts which are in terms of the charge of the positron, and asterisks to represent charge conjugations of specific LQ eigenstates. Further, since the previous Lagrangian was in terms of the weak eigenstates of the fermions,[1] when we induce spontaneous symmetry breaking, we need to introduce the Cabbibo-Kobayashi-Maskawa (CKM) matrix to relate the weak eigenstates of the down-type quarks to their mass eigenstates. Fortunately for us, the contents of this matrix is not important for our study, so, for simplicity, we will take it to be the identity. We will make the additional assumptions for simplicity that not only is there mass degeneracy within LQ multiplets, but also among different multiplets; i.e. in every process, the LQ's will be assumed to have the same mass.[2]

2.2 The $2\ell\ell SS + 1\tau$ Channel

In the interest of analysis of this new pair production method, we need to pick a channel to study. Our group specializes in the $2\ell\ell SS + 1\tau$ channel, which says that there are two same-sign leptons and one hadronically decaying tau in the final state. One way we can analyze this channel is to choose our Yukawa couplings such

[1] Ordinarliy, the weak eigenstates are denoted with primes, but for the purposes of this study where this difference is not relevant since we set the CKM matrix to the identity, we dropped the primes.

[2] As will be mentioned later, will end indeed up considering asymmetric masses by requesting a few extra grid points with final state LQ's of different masses, but only for future supplementary study.

m_{LQ}	y	σ
1500 GeV	0.1	1.23×10^{-7}
	0.5	1.232×10^{-7}
	1.0	7.676×10^{-5}
2000 GeV	0.1	0.001 201
	0.5	1.403×10^{-8}
	1.0	8.829×10^{-6}
2500 GeV	0.1	0.000 141 8
	0.5	1.023×10^{-6}
	1.0	1.769×10^{-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 (2.2).

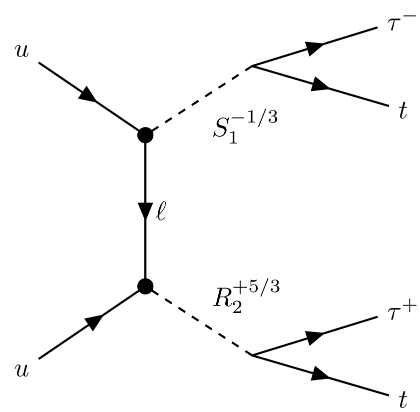

Figure 3: The contribution to $pp \to S_1 R_2 \to t\tau t\tau$ at leading order.

that the LQ's we pair produce both decay into heavy particles, like taus and top quarks. In order to achieve this, we would be looking at having

$$Y_{i3}^{XR} \neq 0, \text{ where } X = L, R \text{ and } i = 1, 2, 3. \tag{7}$$

This will allow the LQ to couple to only right-handed, third-generation leptons, so the tau. Since we also want a decay into a top quark, we need $Y_{33}^{XR} \neq 0$, but we will let the LQ couple to all quark generations to reduce PDF suppressions. Table 2 contains the cross sections for various masses and Yukawa couplings, where the Yukawa magnitude of each matrix element is identical.[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, where we have two up-type quarks (which, in the standard 4 or 5-flavor schemes, means only the up and charm quarks) in the initial state which exchange a lepton in the t-channel to pair produce the LQ's of interest. 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:

$$S_1 \to t\tau^- \to W^+ b\tau^- \to qqb\tau^- \tag{8}$$
$$R_2 \to t\tau^+ \to W^+ b\tau^+ \to \ell^+ \nu b\tau^+, \tag{9}$$

[3]In principle, it is entirely possible to specify that the Y_{33}^{XR} matrix element be greater than the others in order to isolate the decay $LQ \to t\tau$; this is something we delegate to a future study.

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 rely heavily/entirely on 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, a popular program that is used for event generation and cross-section calculation. It comes equipped with many tools for doing the aforementioned 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. Fortunately for us, models have already been made for the scalar LQ's; see Refs [2, 3]. 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 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 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 let's one easily parse through xAOD files and extract relevant information. 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. Four example validation plots are shown in 4.

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. For documentation purposes, this was done here: `https://indico.cern.ch/event/1439002/`. 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. This request was done here: `https://its.cern.ch/jira/browse/ATLMCPROD-11359`

4 Analysis

There are a number of SM processes that contribute also to the $2\ell\ell SS + 1\tau$ channel, including Higgs production, heavy boson ($W^{\pm}$, Z) production, top quark production, and several others; the full table of background (and signal) processes that are considered for the analysis is given in Table A. 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 described in Section 3, 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 list our steps here. First, we produce so-called "small" ntuples, which are created by skimming through the terabytes of raw background and signal ntuples and applying a selection criteria relevent for the $2\ell\ell SS + 1\tau$ channel to remove events that are not interesting. This is done to make subsequent steps significantly faster and more efficient due to a highly reduced event count and file size.[4] Then, the ROOT files are converted to numpy files which can then be read by PyTorch to train a machine learning model. After the machine learning model is trained, we can then plug the small ntuples into the model to create so-called "friend" ntuples, which are then loaded into TRExFitter to produce statistical analyses.

4.1 Testing

Before this analysis was run on the new LQ signal samples, we chose to run the analysis with some release 22 conventional LQ production samples as the signal in order to test that the pipeline was working. This involved a lot of trail and error, as the code that was provided to do the analysis had been made for release 21 samples, and there are considerable differences between the two. For instance, a non-neglible portion of

[4] Ordinarily, the small ntuple step is not done at the very beginning; the subsequent steps take a long time, but not a ridiculously long time, and statistics will be better with larger sample sizes. However, for the purposes of the summer student timeline, this step was expedited so that the analysis can be done in a reasonable time.

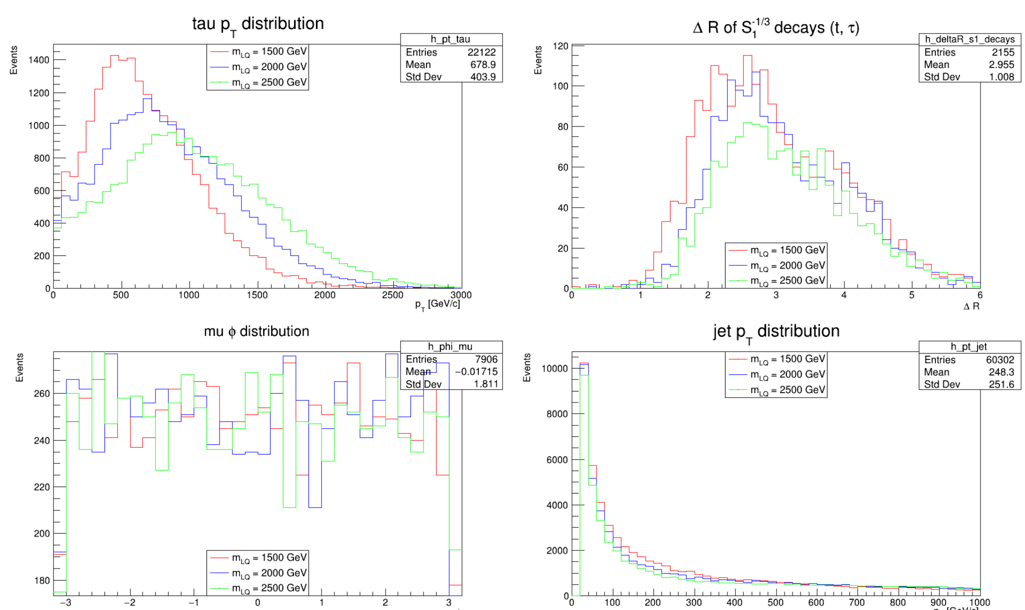

Figure 4: Validation plots generated from test runs for the LQ model. Top left shows the p_T distribution for the τ which came from the decay of the LQ's; top right shows the ΔR for the decay products (t and τ) of the S_1; bottom left shows the *phi* distribution for the final state μ's; and the bottom right shows the p_T distribution for the produced jets.

Figure 5: The raw number of events for each background and signal sample.

the branches in the ntuples had been completely removed or renamed (a list of some of the more important features from Rel.21 that are not present in Rel.22 is given in Appendix B).

Despite these challenges, we were able to get some preliminary results. Please note that the output from these analysis is to be taken with a grain of salt; many of the aforementioned missing branches were completely deleted from all selection criteria, and there were some samples from Rel.21 that were not reproduced in Rel.22, for reasons such as that certain generators were found to be better than others for certain processes, leading to the corresponding samples being discontinued.

Figure 5 contains the samples/yields for the background and signal processes with and without a $2\ell\ell SS + 1\tau$ selection criteria.

A List of DSIDs Considered for Analysis

Process	DSIDs
LQ	545824, 545825, 545826
$t\bar{t}H$	346343, 346344, 346345
$t\bar{t}W$	700168
$t\bar{t}W\_EW$	700205
$t\bar{t}Z$	504330, 504334, 504342
$t\bar{t}$	410470
VV	364250, 364253, 364254, 364255, 364283, 364284, 364285, 364286, 364287, 363355, 363356, 363357, 363358, 363360, 363489
Others	410560, 410408, 410646, 410470, 304014, 345705, 345706, 34572, 364242, 364243, 364244, 364245, 364246, 364247, 364248, 364249, 410081, 364156, 364157, 364158, 364159, 364160, 364161, 364162, 364163, 364164, 364165, 364166, 364167, 364168, 364169, 364170, 364171, 364172, 364173, 364174, 364175, 364176, 364177, 364178, 364179, 364180, 364181, 364182, 364183, 364184, 364185, 364186, 364187, 364188, 364189, 364190, 364191, 364192, 364193, 364194, 364195, 364196, 364197

Table 3: Table of all the DSID for the background and signal processes that are considered for the analysis. The "Others" category contains tZ, WtZ, tW, and ttt samples.

B Most Important Missing Rel.22 Branches

References

[1] Andreas Crivellin and Luc Schnell. Complete lagrangian and set of feynman rules for scalar leptoquarks. *Computer Physics Communications*, 271:108188, February 2022.

[2] Ilja Doršner, Svjetlana Fajfer, and Ajla Lejlić. Novel leptoquark pair production at lhc. *Journal of High Energy Physics*, 2021(5), May 2021.

[3] Ilja Doršner and Admir Greljo. Leptoquark toolbox for precision collider studies. *Journal of High Energy Physics*, 2018(5), May 2018.

[4] Ilja Doršner, Ajla Lejlić, and Shaikh Saad. Asymmetric leptoquark pair production at lhc. *Journal of High Energy Physics*, 2023(3), March 2023.

[5] H. Georgi and S. L. Glashow. Unity of All Elementary Particle Forces. *Phys. Rev. Lett.*, 32:438–441, 1974.

[6] Gudrun Hiller and Martin Schmaltz. R_K and future $b \to s\ell\ell$ physics beyond the standard model opportunities. *Phys. Rev. D*, 90:054014, 2014.

[7] Uma Mahanta. Neutrino masses and mixing angles from leptoquark interactions. *Physical Review D*, 62(7), September 2000.