STUDY OF t-CHANNEL PRODUCTION OF SCALAR LEPTOQUARKS AT LHCb AND CENTRAL ACCEPTANCE DETECTOR

Carrie R. Cox

UM-CERN REU 2023

August 11, 2023

Author Note

This project was made possible by the UM-CERN REU program, and by the Summer Student Programme operated by CERN.

Abstract

Scalar leptoquarks are a hypothetical, beyond standard model (BSM) particle that couples simultaneously to both quarks and leptons, carrying both a lepton number and baryon number. These particles provide a possible explanation for the flavor anomalies observed by the LHCb collaboration at CERN as well as by other B-factory experiments. Scalar leptoquarks appear as a propagator in the Drell-Yan process, and scalar leptoquark models with large couplings to heavy quark flavors link flavor anomalies with modifications in Drell-Yan dilepton kinematic distributions. This phenomenological study uses simulated events to examine how the inclusion of scalar leptoquark t-channel production of leptons affects the expected kinematic distributions of the leptons produced and whether a detectable signal could be expected at LHCb or a model Central Acceptance detector.

Keywords: Scalar Leptoquarks, Beyond Standard Model, LHCb, Machine Learning,

STUDY OF t-CHANNEL PRODUCTION OF SCALAR LEPTOQUARKS AT LHCb AND CENTRAL ACCEPTANCE DETECTOR

Scalar leptoquarks (SLQ) first appeared in some Grand Unification theories and have recently experienced a renaissance as an explanation for the hints of lepton flavor universality (LFU) violations in b-meson decay ratios[1]. These LFU anomalies have been observed in $b \to s\ell\ell$ and $b \to c\ell\nu$ decays at LHCb[2][3], and the inclusion of leptoquarks could accommodate both anomalies. Measurements of Drell-Yan (DY) production have historically been a pillar of research at the Large Hadron Collider (LHC), and specifically in searches for physics beyond the Standard Model (SM), and leptoquark models with large couplings to heavy quark flavors, such as top and bottom quarks, are able to link the b meson decay anomalies with modifications in DY dilepton distributions[4]. The t-channel leptoquark exchange contributes to $pp \to \ell\ell$ production, and at high energies, the effect of such new physics may be enhanced to a degree where we could expect a detectable deviation from SM dilepton distribution predictions, and thus place limits on the possible leptoquark masses or couplings. The goal of this research is to compare SM and SM+SLQ dilepton distribution predictions using the results from Monte Carlo (MC) simulations and study whether the LHCb or a Central Acceptance detector would have sensitivity to detecting t-channel SLQ production.

THEORETICAL FRAMEWORK

Leptoquarks are hypothetical colored bosons that carry both a lepton number and baryon number, and hence appear in many BSM models that unify matter. Because at low energies, any theory of unification must reduce to the SM, SLQs can only appear in five different representations, the $SU(2)_L$ singlets Φ_1 and $\widetilde{\Phi}_1$, the $SU(2)_L$ doublets Φ_2 and $\widetilde{\Phi}_2$, and the $SU(2)_L$ triplet Φ_3. SLQs contribute to DY dilepton processes via *t*-channel production at the tree level, which is initiated by the fusion of two quarks and produces two leptons, and the interaction is mediated by a leptoquark.

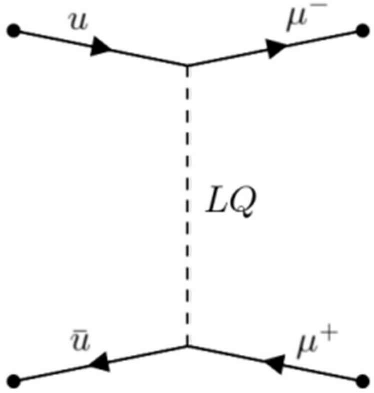

Figure 1: Example of leptoquark contribution to dimuon production at the tree level, initiated by the fusion of two up quarks.

The largest contributors to the background SM processes are *s*-channel Z-boson production and *s*-channel photon production. Important to note is that the different SLQ interfere differently with the SM background. The fields Φ_1, Φ_2, and Φ_3 lead to constructive interference, while $\widetilde{\Phi}_1$ and $\widetilde{\Phi}_2$ interfere destructively with the SM DY background.

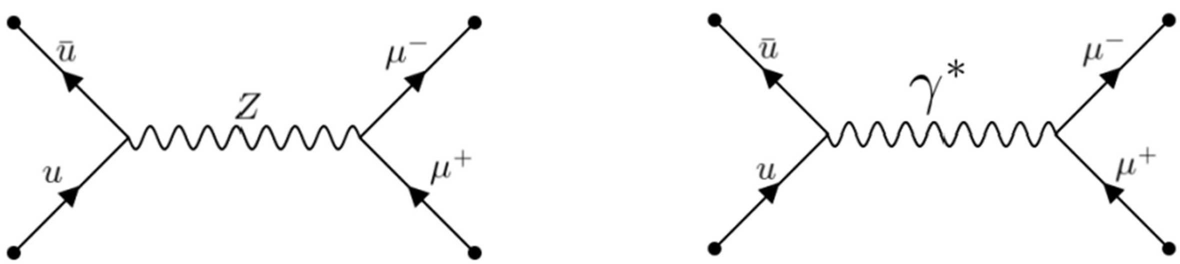

Figure 2: Background SM DY processes. The left Feynman diagram represents s-channel Z boson production, while the right Feynman diagram represents s-channel photon production.

SLQ Model Considered

The simulations were performed using MadGraph5 software and the default SM effective field theory included in the software package. To then include SLQ contributions to DY dilepton production, the SLQRules-UFO-CKM MadGraph model[5] was imported with the default couplings set to unity, and the SLQ masses set to the following values.

Table 1
SLQ Masses

SLQ Representation	Mass (GeV)
Φ_3	5000
Φ_2	3000
$\widetilde{\Phi}_2$	3000
Φ_1	1000
$\widetilde{\Phi}_1$	1000

METHOD

Data Collection

The data in this study was collected from MC simulations of $pp \rightarrow \mu^+\mu^-$ production performed using Madgraph5 software only up to the Leading Order (LO) level, for both the SM and the SM+SLQ effective field theories. No showering effects were included in this study. Only dimuon production was considered due to LHCb being better equipped to detect muons rather than electrons. The SM and SM+SLQ simulations were performed for both the LHCb detector model and also a model Central Acceptance detector, like the ATLAS or CMS collaborations, with the Run 3 expected luminosities of these collaborations in mind. The following is a table summarizing the generation level parameters used as well as the luminosities considered for normalization.

Table 2
Generation-level simulation parameters and luminosities used.

Generation Parameter	LHCb	Central Acceptance
Total number of events N	10^6	10^6
Collision energy (TeV)	13.6	13.6
Lepton pT (GeV)	> 10	> 10
Lepton Rapidity η	$2 < \eta < 5$	$-2.5 < \eta < 2.5$
$M_{\ell\ell}$ (GeV)	> 500	> 500
Luminosity L (fb^{-1})	50	500

The kinematics that were analyzed in this study include the lepton transverse momentum p_T, the rapidity of each lepton η, the rapidity difference between the two leptons $\Delta\eta$, the mean lepton invariant mass $M_{\ell\ell}$, and the angle between the leptons in the boosted of one of the leptons, $\cos(\theta^*)$.

Normalization.

The simulations results were normalized using the normalization constant c, where σ is the event cross section calculated by the MC simulation, L is the luminosity, and N is the number of events that were originally generated:

$$c = \frac{\sigma * L}{N}$$

Significance.

A simple significance estimator was used to assess the signal/background differentiation being observed and thus how sensitive the detector is to t-channel SLQ production. In the following equation, $(SM + SLQ)$ is the combination of Standard Model and SLQ production events generated by the simulation with the SLQ model imported, while SM is the events generated by the Standard Model simulation alone.

$$S = \frac{(SM + SLQ) - SM}{\sqrt{SM}}$$

In the following plots, a significance scan technique is used, where the total number of $(SM + SLQ)$ and SM events occurring before the current location along the x-axis, called n, is summed and then the significance calculation is performed, so that the above equation then becomes

$$S = \frac{\sum_{i=0}^{n}(SM + SLQ) - \sum_{i=0}^{n} SM}{\sqrt{\sum_{i=0}^{n} SM}}$$

This significance scan process is also repeated but summing in the opposite direction, so that is it the events occurring after n that are being summed rather than those occurring before:

$$S = \frac{\sum_{n}^{x_{max}}(SM + SLQ) - \sum_{n}^{x_{max}} SM}{\sqrt{\sum_{n}^{x_{max}} SM}}$$

ANALYSIS

Data

The cross sections calculated by the MC simulation with the generation parameters specified in the above section are summarized in the following table.

Table 3

Cross sections

Detector	SM σ (fb)	SM+SLQ σ (fb)
Central Acceptance	65.6	76.5
LHCb	1.39	1.51

Higher cuts were made on the $M_{\ell\ell}$ kinematic in succession at > 600, 800, and 1000 GeV to explore if such cuts could increase the significance being achieved. Since the contributions of the SLQ t-channel production are enhanced at higher energies, it was believed that at higher energies and therefore higher cuts on $M_{\ell\ell}$, more of the background processes would be eliminated and the more of the signal would be preserved, intensifying it relative to the background. The following two tables summarize the effects of these cuts on the fraction of SM and SM+SLQ events kept as well as the total number of signal events, $(SM + SLQ) - SM$, and background (SM alone) events kept. Recall that one of the original event generation parameters was that the $M_{\ell\ell}$ for all events would be greater than 500 GeV (see Table 2).

Table 4

Effect of $M_{\ell\ell}$ cuts made on Central Acceptance detector results.

$M_{\ell\ell}$ Cut (GeV)	Fraction of SM Events	Fraction of SM+SLQ Events	Signal	Background
> 600	0.017	0.021	4440	17000
> 800	0.006	0.008	2660	5660
> 1000	0.002	0.004	1520	2240

Table 5

Effect of $M_{\ell\ell}$ cuts made on LHCb results.

$M_{\ell\ell}$ Cut (GeV)	Fraction of SM Events	Fraction of SM+SLQ Events	Signal	Background
> 600	$5.63*10^{-6}$	$6.64*10^{-6}$	0.70	5.94
> 800	$4.76*10^{-7}$	$5.60*10^{-7}$	0.09	0.48
> 1000	$3.66*10^{-8}$	$5.00*10^{-8}$	0.01	0.04

Already from the information presented in the tables, it is clear that LHCb will have a very low sensitivity to t-channel SLQ production. Even after the first $M_{\ell\ell}$ cut, less than one signal event is expected for LHCb. If no further cut is made on the $M_{\ell\ell}$ kinematic after the generation parameter of $M_{\ell\ell}$ > 500 GeV, only six signal events can be expected. The very low number of expected events can partly be attributed to LHCb's lower luminosity, which is ten times smaller than the Central Acceptance detector, and can also be attributed to LHCb's unique design as a forward detector, which places strict limits on the detector's η acceptance. LHCb is designed to detect particles produced in the forward region, which is very effective for studying the physics of B-mesons, however the majority of SLQ t-channel production events occur in the central rapidity region, and so the majority of the SLQ events are not able to be detected by LHCb.

In contrast, at the Central Acceptance detector, over one thousand signal events are expected even after the highest $M_{\ell\ell}$ cut is made. Additionally, as expected, a higher proportion of background events than signal events are being removed as higher $M_{\ell\ell}$ cuts are made. However, as higher cuts are made, a large percentage of the signal is still being lost – between the > 600 and > 800 GeV $M_{\ell\ell}$ cuts, only 60% of the signal is kept while 33% of the background is kept. So while the signal over background enhancement that was expected at higher energies is indeed occurring, the higher than desired amount of signal events that are lost may cause the signal significance between cuts to stay approximately the same, or even to decrease. The following kinematic distribution plots demonstrate how distinguishable the signal is from the background at each detector and compute the significance that can be achieved.

Kinematic Distributions

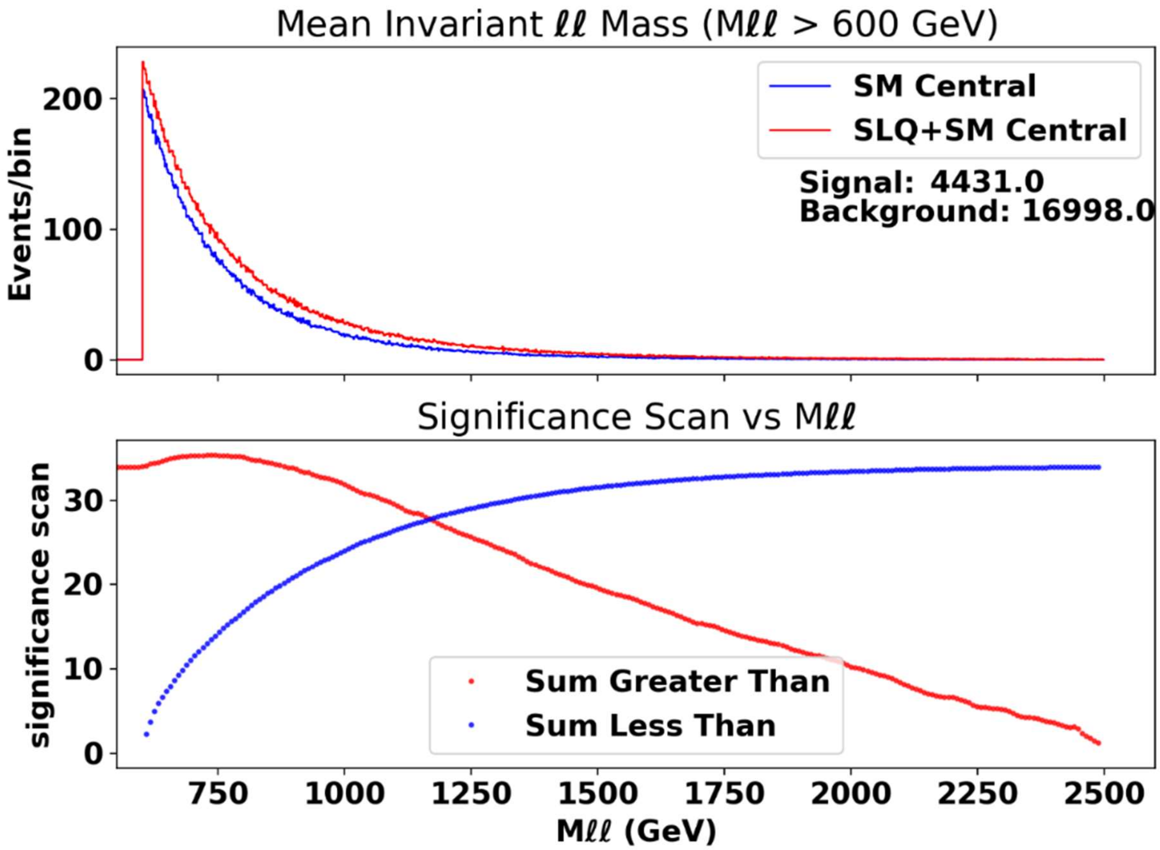

Figure 3: Histogram, significance scan of the $M_{\ell\ell}$ kinematic for the Central Acceptance detector.

The above plot demonstrates that while at lower values of $M_{\ell\ell}$, the SM and SM+SLQ results differ, as $M_{\ell\ell}$ increases, the SM and SM+SLQ results converge. The significance scan also tells that we can expect an integrated significance on the order of 30σ. This is a very high expected significance, and so it is important to recall that this is a phenomenological study and that further refinement would be required to calculate a significance that can be expected at a real detector, however it is still promising that this kind of physics may be pursued at a Central Acceptance detector, like the ATLAS or CMS collaborations on the LHC.

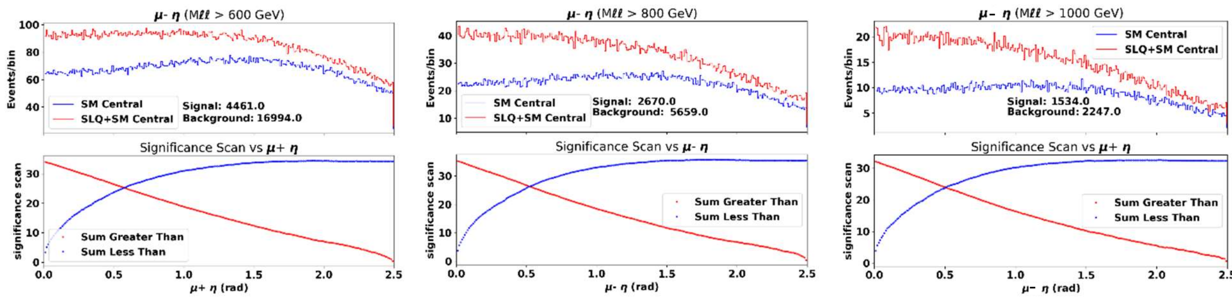

Figure 4: Histograms, significance scans of $\mu^- \ \eta$ kinematic distributions for the Central Acceptance detector, for the three $M_{\ell\ell}$ cuts made. From left: $M_{\ell\ell}$ > 600 GeV, $M_{\ell\ell}$ > 800 GeV, $M_{\ell\ell}$ > 1000 GeV.

The $\mu^- \ \eta$ kinematic distribution results are selected as a representative sample for the rest of the kinematic distributions studied, however these distributions are included in the index of this report. In the above distributions, it is shown that at higher $M_{\ell\ell}$ cuts, the visual differences between the SM and SM+SLQ distributions is enhanced. However, this does not lead to much, if any, significance enhancement due to the larger than desirable amount of signal events that are being lost with each cut. The Central Acceptance detector results indicate that while such a detector would have a high sensitivity to the detection of SLQ t-channel production, simple box cuts placed on the energy will not enhance the significance that can be expected and so do not assist in differentiating the signal and background events.

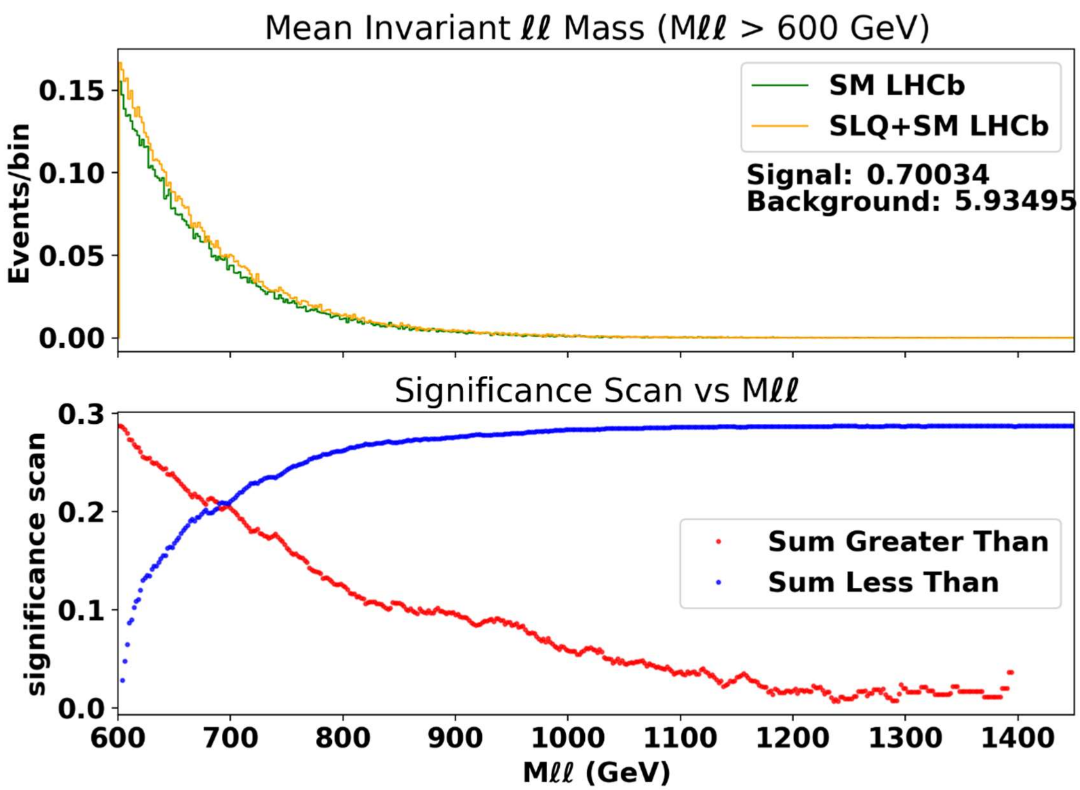

Figure 5: Histogram, significance scan of the $M_{\ell\ell}$ kinematic for LHCb.

Unfortunately, the results for detecting t-channel SLQ production at LHCb are not nearly as promising as the results for the Central Acceptance detector. Even when just selecting the $M_{\ell\ell}$ > 600 GeV, less than one signal event can be expected, and the maximum integrated significance is less than 0.3σ. If more parameters of the LHCb detector beyond its expected Run 3 luminosity and rapidity acceptance were taken into account, this expected significance would only continue to decrease, indicating that LHCb has a very low sensitivity to t-channel SLQ production.

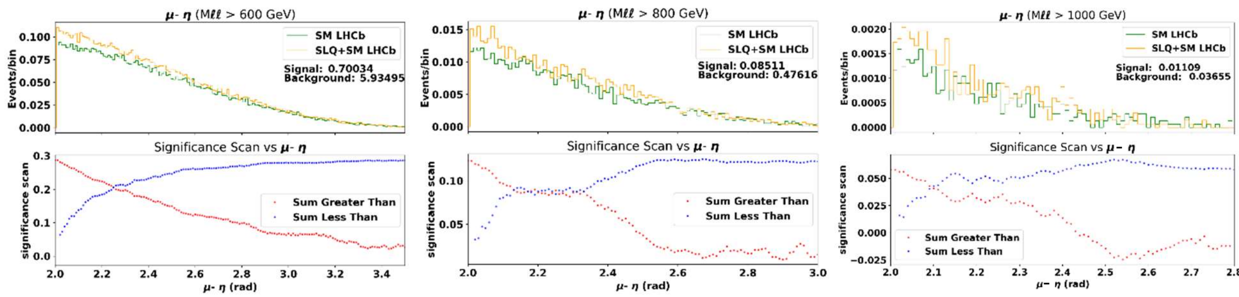

Figure 6: Histograms, significance scans of $\mu^{-}\ \eta$ kinematic distributions for LHCb, for the three $M_{\ell\ell}$ cuts made. From left: $M_{\ell\ell}$ > 600 GeV, $M_{\ell\ell}$ > 800 GeV, $M_{\ell\ell}$ > 1000 GeV.

As higher cuts are made on the $M_{\ell\ell}$ kinematic, the visual discrepancies between the SM and SM+SLQ results do not appear to be noticeable enhanced. The significance decreases with each cut, due to LHCb starting out with very few signal events to begin with. Even though with each cut the background is being eliminated at a higher rate than the signal, the signal that LHCb could be expected to detect is too small to begin with for cuts to be a viable option as no signal events can be afforded to be lost.

Machine Learning Application

In an attempt to increase the signal/background differentiation beyond the limits of simple box cuts, a Machine Learning classification technique is introduced. The inputs into the ML algorithms are the kinematics studied in this report, and the output is classifying the event as either a signal or background event. The results of two classifier algorithms were explored, the first being the Histogram Gradient Boosting Classifier (HGBC) from the python sklearn library. The other classifier explored was a 5 hidden layer, fully connected neural network constructed for this study. To train the classifiers, a new sample data set of SLQ events alone, not including the SM background processes, was simulated. New simulation data was collected because the classifier would be unable to effectively distinguish between a data set that is the SM background alone and a data set that is still mostly made up of the SM background processes with a small proportion of SLQ events. However, this means that interference effects are not being taken into account in the algorithm's development of a decision function. Further study is required to determine how much of effect this would have on the validity of applying these such techniques to real physics data collected from a detector.

To train the algorithms, the data sets randomized and then were split into training and test samples. Both algorithms received the same training and test data. If the algorithm is successful in distinguishing the signal and background events, on a plot of the decision function, there should peak of the background distribution centered about 0 on the x-axis, while the signal distribution should have a peak centered around 1.

Classifier Results

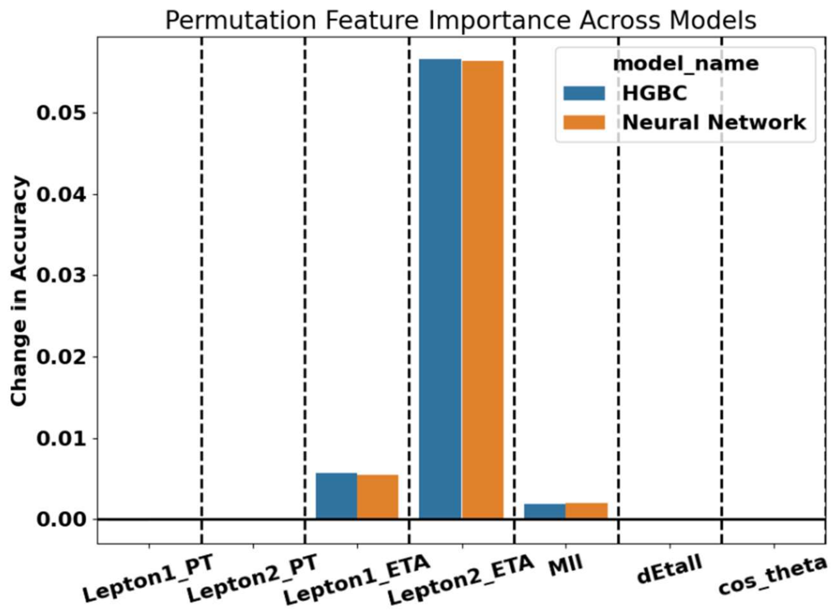

Figure 7: Visualization of which kinematics were most influential in development of decision function for HGBC and neural network algorithms trained on the Central Acceptance results.

From the above plot of the permutation feature importances for HGBC and neural network algorithms, it is apparent that the μ^{-} η kinematic was the leading kinematic in defining a decision function for both styles of algorithms, and many other kinematics were not considered by the algorithms at all.

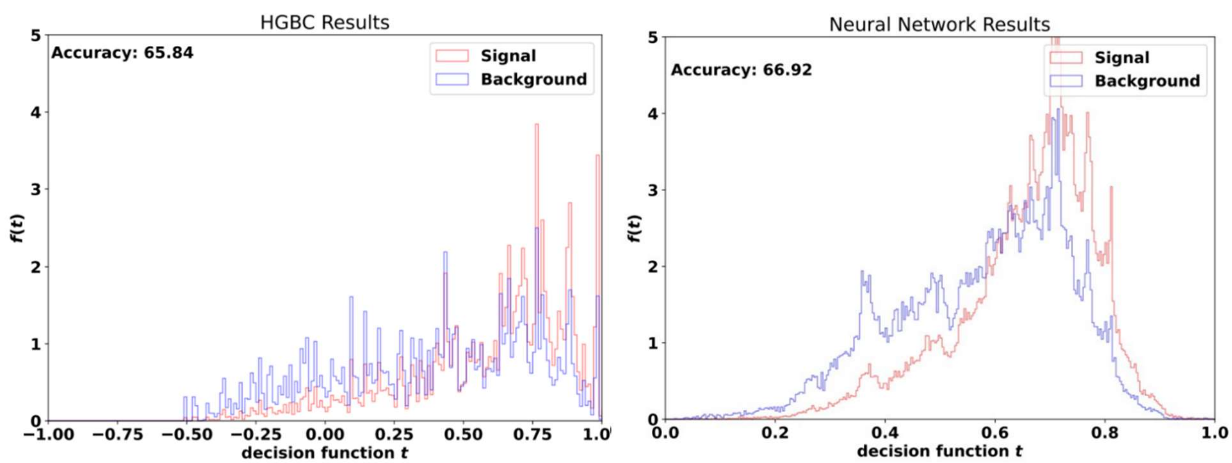

Figure 8: Plot of decision functions developed by HGBC algorithm (left) and neural network algorithm (right) with calculated accuracy shown.

While these results are highly preliminary, it is apparent that a neural network has the potential to assist in signal/background deconvolution beyond the limits of cuts placed on the kinematics and beyond the potential of the HGBC algorithm. Placing a cut at around 0.4 of the decision function of the neural network results, for example, would be eliminate a large amount of background while still maintaining the majority of the signal.

CONCLUSION

Summary

This study analyzed the possibility of detecting t-channel production of scalar leptoquarks at two different detector models, the LHCb forward acceptance detector and a Central Acceptance detector. Simulated data of $pp \rightarrow \mu^+\mu^-$ events was collected using MadGraph5 software and the dilepton kinematic distributions of the SM background processes alone and the SM plus SLQ t-channel production contributions were compared. It was found that unfortunately LHCb has a very low sensitivity to SLQ t-channel production, however a Central Acceptance detector such as the ATLAS or CMS collaborations are expected to have a much higher sensitivity to detecting this process. A machine learning classification technique was then introduced to determine if such methods could be applied to increase the signal/background differentiation beyond the limits of kinematic cuts.

Further Work

Opportunities for furthering this work include varying the free parameters in the SLQ model, such as altering the SLQ representation masses considered and exploring how such changes affect the expected dilepton distributions. The machine learning techniques introduced in this study can also be expanded upon by further adjusting the hyperparameters of the models to optimize the algorithms. This could be done using an optimization algorithm, such as the Bayesian Optimization algorithm.

Acknowledgements

I would first like to thank my advisors, Dr. Nate Grieser and Dr. Conor Henderson, for their guidance and support, and for making this project possible. I would also like to thank the other members of the LHCb collaboration for their hospitality this summer. I am also very grateful to the UM-CERN REU and CERN Summer Student Programme coordinators for their work in organizing this program and creating such an amazing opportunity. Finally, I am thankful to the other 2023 CERN summer students for building a wonderful community, and for making every moment of this summer unforgettable.

References

[2] Celani, S. (2021). Lepton Flavour Universality tests and Lepton Flavour Violation searches at LHCb. *16th International Workshop on Tau Lepton Physics (TAU2021)*.

[5] Crivellin, A., & Schnell, L. (2022(271)). Complete Lagrangian and Set of Feynman Rules for Scalar Leptoquarks. *Computer Physics Communications*, 108188.

[1] Fajfer, S. (2018). Scalar leptquarks: From GUT to B anomalies. *EPJ Web Conf., 192*, 00025.

[4] Haisch, U., Schnell, L., & Schulte, S. (2022(11)). On Drell-Yan production of scalar leptoquarks coupling to heavy-quark flavours. *Journal of High Energy Physics*, 1-21.

[3] Wang, Y. (2022). Study of flavour-changing neutral current and charged current B meson decays. *[Doctoral thesis, Hua-Zhong Normal U]. CERN Theses.*