Analyzing Electroweak ZZ + jj → 4l + jj Vector Boson Scattering with ATLAS

Elise Hinkle[1, 2]

[1] *Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*
[2] *Department of Physics, Brown University, Providence, Rhode Island 02912, USA*

Vector Boson Scattering (VBS) is a key process to test the Electroweak (EWK) Symmetry Breaking (EWSB) mechanism and search for anomalous Quartic Gauge Couplings (aQGC). The ZZ+jj→ 4l+jj VBS channel was studied using data from the ATLAS detector collected between 2015 and 2017 and theoretical models of signal and background processes from Monte Carlo (MC) simulations. These sets all correspond to an LHC center of mass energy ($\sqrt{s}$) of 13 TeV and an integrated luminosity ($\mathcal{L}$) of $79.91 fb^{-1}$. A combination of cut-based and multivariate analysis techniques were used to analyze these data and samples, resulting in a measurement for an inclusive (EWK and Quantum Chromodynamics, QCD) ZZ+jj cross section of $1.50 \pm 0.04(lumi) \pm 0.01(bkg) \pm 0.18(stat) \pm^{0.13}_{0.11}(exp)$ fb. Furthermore, the expected significance of the EWK ZZ+jj process was measured at $3.4 \pm^{0.5}_{0.4} \sigma$, excluding systematic uncertainties.

INTRODUCTION

The Standard Model (SM) is the framework that physicists use to describe the most basic building blocks of the universe and the forces through which they interact. While the SM is relatively effective in its ability to fulfill this function, there are still physical phenomena which suggest that the SM does not represent a complete view of fundamental particles and their interactions. For instance, one may look to the discovery of the Higgs Boson and its coupling to gauge bosons, specifically the weak force carriers, W± and Z bosons, which are also known as vector bosons [1]. This coupling is responsible for the non-zero mass of the W± and Z bosons as opposed to the massless nature of the electromagnetic force carrier, the photon (γ). While the electromagnetic and weak forces generally act in parallel ways and, in fact, are often discussed in unison as the electroweak (EWK) force, the discrepancy between the masses of their force carriers is an example of how the symmetry between the two forces is incomplete, or broken. The method for this phenomenon, known as electroweak symmetry breaking (EWSB), is not completely understood at this time and therefore is one avenue for further exploration within the SM.

Another related area in which knowledge of particle interactions can be advanced is that of higher order couplings of gauge bosons. As many theories of physics beyond the SM involve changes in the way that gauge bosons couple to each other, observing, for example, anomalous Quartic Gauge Couplings (aQGC) would point to new physics [1].

One fascinating process which can provide further knowledge regarding both EWSB and aQGC is Vector Boson Scattering (VBS). This class of interactions is therefore one of many studied through the A Toroidal LHC Apparatus (ATLAS) Experiment, one of two large experiments seeking to learn from proton-proton collisions at the Large Hadron Collider (LHC) at CERN. Thus, the potential to identify areas of new physics arises from comparing experimentally found cross sections of VBS processes to those expected by the SM.

While there are many VBS channels, this analysis is concerned solely with ZZ VBS in the presence of two jets with a fully leptonic (four lepton, or 4l) final state still in the presence of two jets. The goal of this analysis was therefore to measure an inclusive (EWK and Quantum Chromodynamics, QCD) cross section of ZZ+jj → 4l+jj processes and, furthermore, find the expected significance of the EWK ZZ+jj → 4l+jj process. Examples of Feynman diagrams showing the EWK and QCD ZZ+jj processes, as well as gluon-initiated background processes, can be seen in Figure 1.

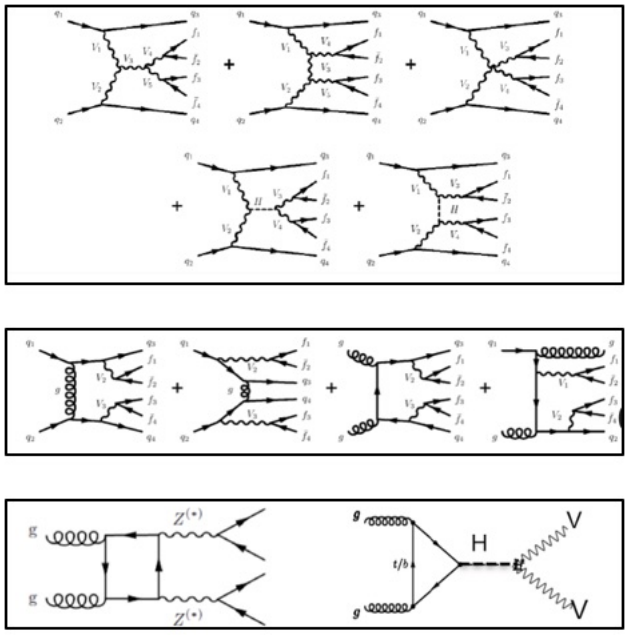


FIG. 1: Feynman diagrams of VBS processes including EWK ZZ+jj → 4l+jj (top), QCD ZZ+jj → 4l+jj (middle), and gluon-mediated irreducible background processes (bottom) [2, 3].

DATA COLLECTION AND SAMPLES USED

All data and samples used were collected by or simulated based on data from the ATLAS experiment. The ATLAS detector is located on the LHC at CERN in Geneva, Switzerland, and consists of five main concentric layers of detection materials. From the inside out, these layers include tracking equipment, a solenoid magnet, an electromagnetic calorimeter, a hadronic calorimeter, and a muon spectrometer [4]. Each piece of the detector is essential in the collection of information used to accurately recreate the particle interactions resulting from initial proton-proton collisions in the LHC.

The particular ATLAS detector data used in this analysis was taken between 2015 and 2017 and corresponds to a center of mass energy ($\sqrt{s}$) of 13 TeV and an integrated luminosity ($\mathcal{L}$) of $79.91 fb^{-1}$.

In addition to detector data, Monte Carlo (MC) samples of different ZZ+jj processes, derived from SM calculations and driven by 2017 detector data, were used to enhance understanding of the detector data. In this analysis, the MC samples describing the EWK ZZ+jj process were processed using MadGraph, while samples for EWK ZZ+jj background processes such as QCD ZZ+jj, ggZZ, ttZ, and WWZ were processed using Sherpa. So-called "fake" events were data-driven, but the fake factor used to identify them was from MC samples generated by Sherpa (Z+jets) and Powheg (ttbar). All of these MC samples also have $\sqrt{s}$ equal to 13 TeV and an $\mathcal{L}$ of $79.91 fb^{-1}$.

ANALYSIS TECHNIQUES

Two main analysis techniques were used to understand the data and MC samples studied. While these tools, cut-based analysis and multivariate analysis, each have merit separately, a combination of the two techniques was found to be the best way to separate signal and background processes. Thus, these combined methods led to increased understanding of the VBS processes in question.

Cut-Based Analysis

Cut-based analysis involves using spectra of individual kinematic variables in detector data and MC samples to identify variable values more likely to correspond to signal or background processes. Thus, cuts can be made on these variables to isolate regions where there is more signal versus background.

Before making cuts related to the identification of signal and background processes, cuts must be made to account for detector geometry and physical considerations. The initial cuts necessary for this purpose can be seen in Figure 2. In order to understand this figure, it is important to know a few key variables. For instance, η represents the pseudorapidity of a physical entity, which is related to the angle between the particle(s) or jet(s) and the initial collision axis (i.e. of the proton-proton collisions in the LHC), p_T refers to transverse momentum, and m represents mass. Also, subscripts of j refer to jets while sub- or superscripts of l refer to leptons which, in this case, are either electrons or muons.

MAIN CUTS MADE
$\eta_{J1} * \eta_{J2} < 0.$
$\lvert \Delta\eta_{JJ} \rvert > 2.$
$m_{JJ} > 300.\,GeV$
$p_T^{\ell} > 20., 20., 10.,$ and $7.$ GeV
$\lvert \eta_{\ell} \rvert < 2.5$ (2.7) for e(μ)
$66\ GeV < m_{\ell\ell} < 116\ GeV$

FIG. 2: Table showing the main cuts made on data and MC samples to account for detector geometry and physical limitations.

Further event selection was also completed to make sure that the events studied all reflected the characteristics and topology of possible ZZ+jj processes. This involved making cuts related to the lepton and jet characteristics of the events. First, selections were made based on general lepton (electron and muon) characteristics. Next, the number and form of jets were considered. Then, groups of four leptons, or quadruplets, were identified in the events as plausible decay products of two Z bosons. As a final state of the ZZ+jj $\rightarrow$ 4l+jj process has one set of four leptons, final events were chosen based on the identification of one quadruplet that was the most likely to have arisen from a ZZ+jj process. These large sets of event selection cuts are not listed in this paper for the purpose of clarity and to provide emphasis on the work done by the author, which did not involve choosing these relatively extensive cuts.

After accounting for physical and detector geometry limitations and selecting possible ZZ+jj events, the spectra of other variables were used to understand where the parts of the distribution where EWK ZZ+jj were highest in comparison to background processes (such as QCD ZZ+jj and gluon-mediated processes). While many variables were studied, two of the most insightful variables were m_{jj}, the invariant mass of the two jets in the event,

and $|\Delta\eta_{jj}|$, the absolute value of the difference in pseudo-rapidity between the two jets. Other variables that suggested possible further kinematic variable cuts included the p_T of each of the four leptons in the event and the separate and combined jet p_Ts.

Multivariate Analysis

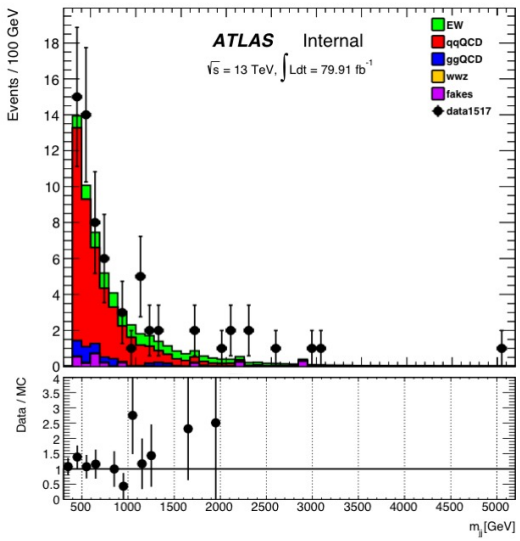
While cut-based analysis is useful, looking at various kinematic variables independently is not effective in fully separating events related to, for instance, EWK ZZ+jj interactions versus QCD ZZ+jj interactions. The idea behind multivariate analysis is therefore to use multiple kinematic variables simultaneously in order to separate one signal process from all of the background processes that could be mistaken for the signal. By using many variables at once, discrimination ability is therefore much stronger than using isolated cuts on variables.

The specific form of multivariate analysis used in this analysis was Gradient Boosted Decision Trees (BDTs) by way of the Toolkit for Multivariate Data Analysis with ROOT (TMVA). This algorithm relies on taking many simple binary decision trees, applying a gradient boosting algorithm to examine misclassified events, and combining the resulting outputs from all trees to obtain a single powerful output as the final discriminant. As the "forest" of trees can be set up in many different ways based on the number of trees, the number of layers per tree, and more complicated variables, many different combinations of settings were scanned in order to find the most effective BDT framework to separate EWK signal from QCD background.

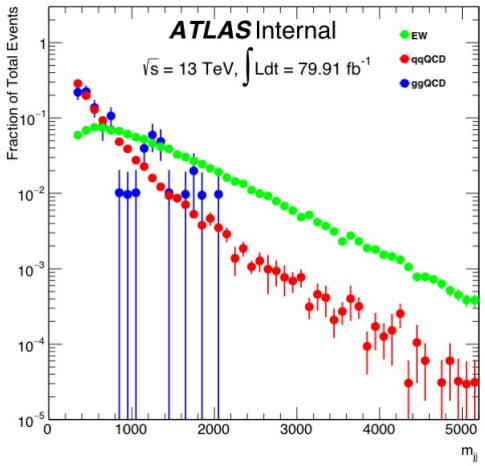
RESULTS

The first results considered in this analysis were kinematic variable distributions connected to cut-based analysis. While over thirty histograms were produced showing either general distributions of MC and data events and their event ratios or the relative distributions of major MC processes, only four are shown here as examples in Figures 3 and 4. Apart from the m_{jj} and $|\Delta\eta_{jj}|$ distributions shown here, other studied variables included lepton (leading, sub-leading, sub-sub-leading, and sub-sub-sub-leading) p_T, number of leptons in an event, lepton η, jet η, dijet and single jet p_T, and masses of paired leptons in events.

Following the cut-based analysis evaluations, event yields for each of the sub-processes simulated by MC algorithms and a total event yield from data were measured. These results can be seen in Figure 5a. The main forms of uncertainty included 2.1% uncertainty due to luminosity, less than 9% theoretical uncertainty from parton distribution functions (PDF), QCD scale factor,



(a) Distribution of invariant jet mass (m_{jj}) for ZZ+jj → 4l+jj VBS processes represented in data and MC samples, including the ratio of MC events to data events for each m_{jj} value.

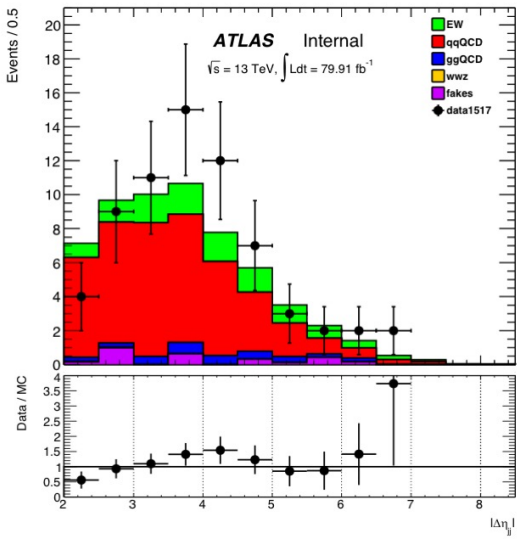


(b) Relative proportions of EWK, QCD, and ggZZ ZZ+jj → 4l+jj VBS processes over MC sample m_{jj} event values for these processes.

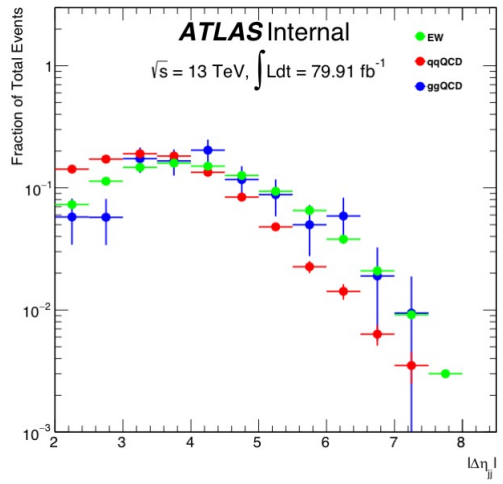
FIG. 3: Distributions showing m_{jj} in data and MC events, with major MC samples including EW = EWK ZZ+jj → 4l+jj VBS processes, qqQCD = QCD ZZ+jj → 4l+jj VBS processes, and ggQCD = ggZZ ZZ+jj → 4l+jj VBS processes.

and parton showering, and less than 10% experimental uncertainty from detector calibrations [2]. A more detailed breakdown of systematic uncertainties can be seen in Figure 5b.

Using these yield values, the inclusive ZZ+jj → 4l+jj cross section in the fiducial volume was then measured to be $1.50 \pm 0.04(lumi) \pm 0.01(bkg) \pm 0.18(stat)^{+0.13}_{-0.11}(exp)$



(a) Distribution of the difference in pseudorapidity of the two jets ($|\Delta\eta_{jj}|$) for ZZ+jj → 4l+jj VBS processes represented in data and MC samples, including the ratio of MC events to data events for each $|\Delta\eta_{jj}|$ value.



(b) Relative proportions of EWK, QCD, and ggZZ ZZ+jj → 4l+jj VBS processes over MC sample $|\Delta\eta_{jj}|$ event values for these processes.

FIG. 4: Distributions showing $|\Delta\eta_{jj}|$ in data and MC events, with major MC samples including EW = EWK ZZ+jj → 4l+jj VBS processes, qqQCD = QCD ZZ+jj → 4l+jj VBS processes, and ggQCD = ggZZ ZZ+jj → 4l+jj VBS processes.

fb, as compared to the inclusive cross section expected by theory of $1.31\pm^{0.21}_{0.15}(theo)$ fb. As the inclusiveness of the cross section measurement means that both quark- and gluon- initiated processes (EWK and QCD ZZ+jj processes) were used, the major background processes in this measurement are top and multi-boson processes. A description of the fiducial volume is given in Figure 6.

Region	Signal	Signal + Control
ZZ → 4l(signal)	11.39±0.04	13.72±0.05
qq → ZZ → 4l(QCD)	41.38±0.48	118.78±0.78
gg → ZZ → 4l	3.95±0.39	9.61±0.61
WWZ	0.17±0.02	0.48±0.04
tt̄Z	<0.02	<0.02
Other (includes fake leptons)	2.23±1.07	5.47±1.90
Expected	59.11±1.24	148.11±1.11
Data (2015-2017)	67	143

(a) Event yields for MC samples and data showing ZZ+jj → 4l+jj VBS processes. Signal region includes all events after cuts based on detector geometry, physical limitations, event selection, and kinematic variable distributions, while signal + control regions includes all detected (data) or sampled (MC) events.

		$ZZ \to 4\ell + 2j$	
Stat.	–	−0.4%	0.4%
Theo. Unc.	PDF	−6.2%	6.2%
	Scale	−5.4%	6.0%
	Total	−8.2%	8.7%
Exp. Unc.	Electron	-3.9%	4.0%
	Muon	-1.4%	1.4%
	Jet	-2.0%	2.0%
	E_T^{miss}		
	Pileup	-0.5%	0.5%
	Total	-4.6%	4.7%
Total		-9.4%	9.9%

(b) Event systematics for ZZ+jj → 4l+jj VBS processes [2].

FIG. 5: Event yields and systematics for MC samples and data for ZZ+jj → 4l+jj VBS processes.

Fiducial Volume:

$\lvert\eta^e\rvert$ <2.47, $\lvert\eta^\mu\rvert$ <2.7
p_T^l > 20, 20, 10, 7 GeV
$\Delta R(l,l)$ >0.2
$\lvert\eta^j\rvert$ <4.5
p_T^j >30 when $\lvert\eta^j\rvert$ <2.4
p_T^j >40 when $\lvert\eta^j\rvert$ <4.5
$\lvert\Delta\eta(j_1,j_2)\rvert$ >2.0

FIG. 6: Table showing the cut definitions of the fiducial volume in which the inclusive ZZ+jj → 4l+jj cross section was measured [2]. In general, fiducial volume measurements refer to those which take into account limitations due to detector geometry.

Finally, the expected significance of the EWK ZZ+jj → 4l+jj signal process versus the QCD ZZ+jj → 4l+jj background process was measured to be $3.4\pm^{0.5}_{0.4}\sigma$, only including statistical uncertainty. This value comes from

BDT trainings used to separate events from MC samples representing these two processes. These BDT trainings were then fitted to a binned likelihood function, resulting in the Poisson profile likelihood scan in Figure 7.

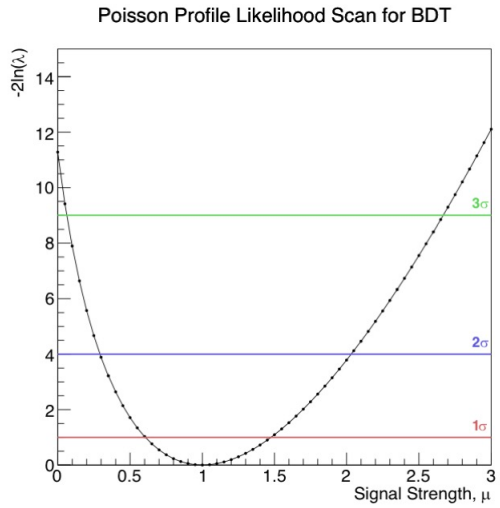


FIG. 7: Poisson profile likelihood scan for BDT results separating EWK ZZ+jj → 4l+jj signal process events from QCD ZZ+jj → 4l+jj background process events in MC samples. σ values represent the number of standard deviations of the measured value of total signal and background events from the completely background (no signal) hypothesis.

DISCUSSION

While not all histograms produced in relation to cut-based analysis of the processes studied have been shown here, the four provided in Figures 3 and 4 are some of the most valuable. The m_{jj} and $|\Delta\eta_{jj}|$ distributions in 3a and 4a show that, generally, the shape of the distributions of all of the MC events together mirrors that of data for these variables. Still, it is clear from the ratio plots underneath the main canvas of each of these histograms that, especially in the mid-high regions of the distributions, more data events were counted than combined MC events. In the m_{jj} plot in Figure 3a, we can also see that the vast majority of events were sampled or detected on the lower end of the spectrum. However, the fractional event total spectrum in Figure 3b shows that the proportion of EWK events is higher than that of QCD and ggZZ events for higher m_{jj}. A similar trend, albeit slightly more subtle, is seen with $|\Delta\eta_{jj}|$ from the two plots in Figure 4. From such considerations, signal regions of a data set can be honed by cuts on each variable so that there is a relatively high percentage of signal events in the final region studied.

The efficacy of this cut-based analysis in allowing for focus on regions of higher signal versus background processes can be seen in the event yields in Figure 5a. In this particular case, we see this in the large decrease in the number of simulated QCD ZZ+jj → 4l+jj events (the main background process to EWK ZZ+jj → 4l+jj) in the signal region versus the signal and control region, especially as compared to the decrease in EWK events. Percentage-wise, these two decreases are roughly 65% for the QCD process versus roughly 17% for the EWK process. By concentrating on regions in MC samples with a greater proportion of signal, we can therefore focus more easily on these events and get a better idea of their yields and distribution in data.

In terms of values measured, the inclusive expected EWK and QCD cross section for ZZ+jj → 4l+jj of $1.50 \pm 0.04(lumi) \pm 0.01(bkg) \pm 0.18(stat)\pm^{0.13}_{0.11}(exp)$ fb is consistent with the theoretical value from SM predictions of $1.31\pm^{0.21}_{0.15}(theo)$ fb. The expected significance of the EWK ZZ+jj → 4l+jj, found to be $3.4\pm^{0.5}_{0.4}\sigma$ without taking systematic uncertainties into account, is a strong indication that the EWK ZZ+jj → 4l+jj process occurs, as expected by theory, due to the deviation of greater than 3σ from the background-only hypothesis of zero EWK ZZ+jj → 4l+jj VBS processes. This particular measurement was only possible due to the multivariate BDT methods used, as complete separation of signal and background events in MC samples is quite difficult, or even impossible, when simply using cut-based analysis.

CONCLUSIONS

VBS processes and, specifically, the ZZ+jj → 4l+jj processes studied in this analysis, are key avenues for testing EWSB and searching for aQGC. By using data from the ATLAS Collaboration and data-driven MC samples based on this data, an inclusive QCD and EWK ZZ+jj → 4l+jj cross section was measured to be $1.50\pm0.04(lumi)\pm 0.01(bkg) \pm 0.18(stat)\pm^{0.13}_{0.11}(exp)$ fb, while the signal significance of the EWK process versus the QCD process was measured at $3.4\pm^{0.5}_{0.4}\sigma$ without taking systematic uncertainties into account. These values were measured at a center of mass energy ($\sqrt{s}$) of 13 TeV and an integrated luminosity ($\mathcal{L}$) of $79.91 fb^{-1}$. A combination of cut-based and multivariate analysis was used for this endeavor.

This work, completed as part of the University of Michigan Physics REU Program funded by the National Science Foundation (NSF) in summer 2018, represents a simplified portion of past and continuing work done by the ATLAS VBSZZ Analysis group. Certain considerations, such as incorporating systematic uncertainties into the signal significance value found for the EWK process, must be resolved before the work done in this paper could be used as part of the final analysis. Furthermore, more work regarding the relationship between ZZ+jj → 4l+jj processes and aQGC must be done, and all work related

to this analysis must pass in-depth ATLAS collaboration reviews before the release of a comprehensive paper on this topic. At the end of the REU program in August 2018, it was expected that a paper regarding these analyses would be released by ATLAS in fall of 2018.

The author found the REU program and this analysis to be a wonderful way to learn more about and be immersed in high energy physics analysis research. While the main products from the student's summer work are presented here, related endeavors, such as further scrutiny of some of the excluded kinematic variable distributions, were completed as well. The student was also able to attend University of Michigan ATLAS group meetings, as well as participate and present in ATLAS VBSZZ Analysis group meetings. Together, these components have given the author a well-rounded understanding of the general format and demands of analysis of a large high energy physics experiment. In short, this REU was an invaluable experience which the author would recommend to anyone interested in high energy experimental or other physics research.

ACKNOWLEDGEMENTS

The author would like to thank Professor Bing Zhou, Rongkun Wang, and the rest of the ATLAS University of Michigan and VBSZZ Analysis groups for their guidance and assistance throughout this project. She would also like to acknowledge all of the hard work of Professor Myron Campbell, Professor Jim Liu, and Grace Johnson in organizing the University of Michigan Physics REU for summer 2018. Finally, she would like to thank the NSF and the University of Michigan for funding this project and hosting this REU program through Grant 1559988.

[1] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, *et al.*, Physics letters B **774**, 682 (2017).

[2] J. Li, "Study of vbs zzjj production with zz $\rightarrow$ 4l, ll$\nu\nu$," Presented at US ATLAS Summer Workshop 2018 (2018).

[3] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, R. Aben, M. Abolins, O. AbouZeid, H. Abramowicz, H. Abreu, R. Abreu, *et al.*, The European Physical Journal C **75**, 335 (2015).

[4] ATLAS, "Atlas photos," (n.d.), accessed Aug. 9, 2018.