Study of VBS ZZ production associating two jets with the ATLAS experiment

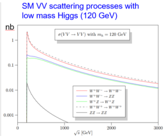
Shuzhou Zhang

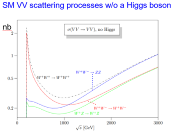
01 Aug, 2019

University of Michigan

The VBS Processes at the LHC

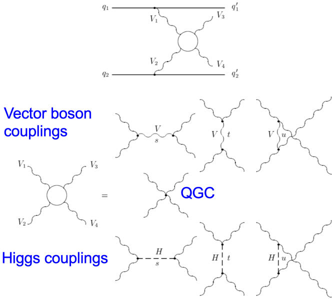
Vector Boson Scattering (VBS) is a key process to probe the mechanism of electroweak symmetry breaking, and close connection with the SM Higgs.





JHEP11(2008)010

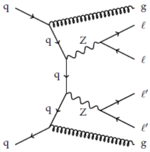
- Involving Quartic Gauge Couplings (QGCs) which is sensitive to new physics.



Phys. ReV. D 55, 7165-Published 1 June 1997

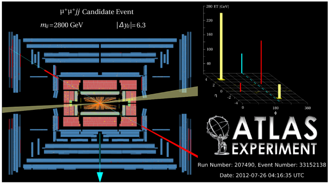
Experimental Signatures of VBS

- Two intermediate vector bosons radiated from two incoming quarks.
- Final state with two vector bosons plus two outgoing jets.
- In general, two "tag" jets in forward region with large rapidity separation and large invariant mass.
- EW VBS has relatively smaller cross-sections, suffer from irreducible QCD VV + 2jets events.



First evidence of these processes

Candidate VBS event from ssWW



Phys. Rev Lett. 113, 141803

Analysis in the ZZ Channels

- Measurement of inclusive ZZ+2j cross-section (QCD and electroweak production of ZZ+2j) in sensitive phase spaces for 4l and llvv.
- Observation of electroweak production of ZZ+2j, combining 4l and llvv.
- Data: full Run II data, 2015-2018, 139 fb^{-1}.
- Signal sample: MG.
- QCD background: Sherpa.

Analysis Strategy

- First step inclusive cross-section measurements with cut-based analysis.
- MVA (BDT)-based analysis is used then to extract the EW VBS ZZ signal from background.
- Interference between EW and QCD is treated as systematic on the EW VBS ZZ production measurement.

- Study of EW VV→ZZ (V=W, Z) processes is a good way to probe vector boson scattering. Analysis is optimized to the on-shell ZZ region so VBF H is minor.

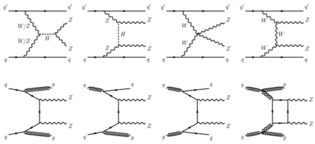


Figure 1: Typical diagrams for the production of $ZZjj$, including the relevant VBS diagrams (first row) and the QCD diagrams (second row).

Analysis Overview

ZZ$\rightarrow 4l$ channel:

- Very clean experimental signature, small background contribution (3%) from misidentified leptons (Z+jets, $t\bar{t}$, WZ).
- While extrapolating the EW VBS processes, the QCD 4ljj production becomes the major background. EW/QCD is around 20% level overall, MVA is needed

ZZ$\rightarrow 2l2\nu$ channel:

- Larger background compared to the 4l channel. Mostly from WZ,WW,$t\bar{t}$ and irreducible QCD ZZjj.
- EW/background is around 15% level, MVA is used as well to extrapolate EW VBS ZZ processes, but background components are more complicated than the 4l channel.

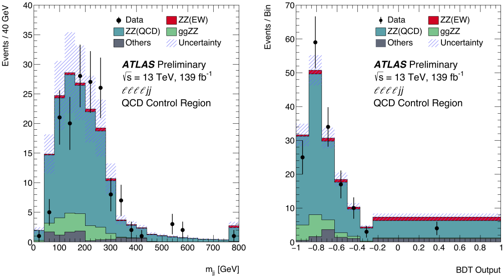
Object and Event Selection

	$\ell\ell\ell\ell jj$	$\ell\ell\nu\nu jj$
Electrons	$p_{\mathrm{T}} > 7$ GeV, $\lvert\eta\rvert < 2.47$ $\lvert d_0/\sigma_{d_0}\rvert < 5$ and $\lvert z_0 \times \sin\theta\rvert < 0.5$ mm	
Muons	$p_{\mathrm{T}} > 7$ GeV, $\lvert\eta\rvert < 2.7$ $\lvert d_0/\sigma_{d_0}\rvert < 3$ and $\lvert z_0 \times \sin\theta\rvert < 0.5$ mm	$p_{\mathrm{T}} > 7$ GeV, $\lvert\eta\rvert < 2.5$
Jets	$p_{\mathrm{T}} > 30$ (40) GeV for $\lvert\eta\rvert < 2.4$ ($2.4 < \lvert\eta\rvert < 4.5$)	$p_{\mathrm{T}} > 60$ (40) GeV for the leading (sub-leading) jet
ZZ selection	$p_{\mathrm{T}} > 20, 20, 10$ GeV for the leading, sub-leading and third leptons Two OSSF lepton pairs with smallest $\lvert m_{\ell^+\ell^-} - m_Z\rvert + \lvert m_{\ell'^+\ell'^-} - m_Z\rvert$ $m_{\ell^+\ell^-} > 10$ GeV for lepton pairs $\Delta R(\ell, \ell') > 0.2$ $66 < m_{\ell^+\ell^-} < 116$ GeV	$p_{\mathrm{T}} > 30$ (20) GeV for the leading (sub-leading) lepton One OSSF lepton pair and no third leptons $80 < m_{\ell^+\ell^-} < 100$ GeV No b-tagged jets $E_{\mathrm{T}}^{\mathrm{miss}}$ significance > 12
Dijet selection	Two most energetic jets with $y_{j_1} \times y_{j_2} < 0$ $m_{jj} > 300$ GeV and $\Delta y(jj) > 2$	$m_{jj} > 400$ GeV and $\Delta y(jj) > 2$

- Different quality and isolation requirements for two channels, due to different background level.
- Lepton favored overlap removal for 4l channel, where higher priority was given to leptons when overlapping with jets. Signal (background) yields increase by 19 (14)%.

Background Estimation-4l, QCD ZZjj

- Yield is constrained by dedicated QCD control region. QCD CR is defined by reverting m_{jj} or $\Delta\eta_{jj}$ cut to $m_{jj} < 300$ GeV or $|\Delta\eta_{jj}| < 2.0$.



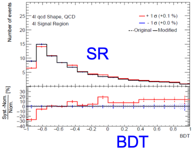
m_{jj} and BDT in QCD CR

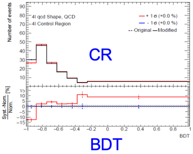
- Additional systematic uncertainties due to different generators and pile-up are also checked.

Background Estimation - 4l, QCD ZZjj, Additional Systematics

Modeling of the QCD ZZjj

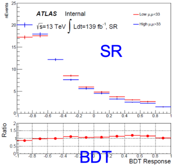
- The shape differences between different generators can not be covered by the standard QCD up/down systematics
- Additional shape systematic derived by comparing Sherpa with MG samples at truth level

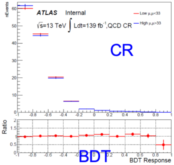




Check the pile-up modeling

- Have compared shape difference using low/high mu QCD MC events to cover the possible bias due to pile-up effect





Background Estimation - 4l, Fakes

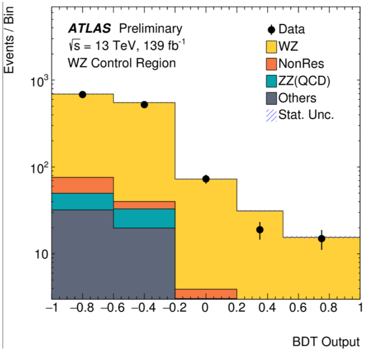
Fake factor method used, where different fake factors have been derived for Zjets and $t\bar{t}$ processes, from dedicated CRs.

- Z+jets: two tag leptons under Z mass. Counting additional good and bad leptons to calculate fake factor.
- $t\bar{t}$: two tag leptons ($e\mu$ flavor) with additional btagging, MET and m_T^W cuts to enhance $t\bar{t}$ purity. Counting additional good and bad leptons to calculate fake factor.
- Bad lepton is defined by reverting the lepton quality, isolation or impact parameters.
- 4ljj data CR defined with at least two good leptons.

Final estimation is 2.3$\pm$1.6. Systematic includes variations on those MC subtraction, bad lepton definitions, fake factor binning, data MC difference, statistical uncertainty on fake factor due to limited data events.

WZjj Background Estimation - $2l2\nu$

- Normalization constraint by dedicated 3l, WZ dominant CR, with looser 2jet selections.
- Derived normalization factor is 0.85.
- Shape is estimated using simulation.



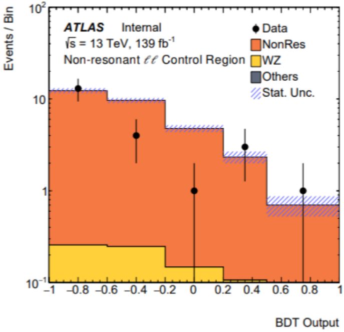
WW, top, Z$\rightarrow \tau\tau$ Background Estimation - $2l2\nu$

- Estimated with events in dedicated em data CR, following:

$$N^{Q,e\mu}_{SRee} = \frac{1}{2} \times \epsilon^{Q} \times N^{\mu\in Q,sub,bkg}_{e\mu CR}$$

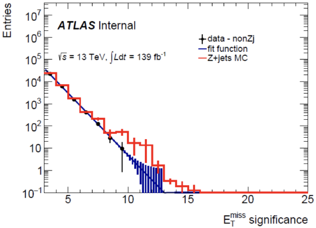
$$N^{Q,e\mu}_{SR\mu\mu} = \frac{1}{2} \times \frac{1}{\epsilon^{Q}} \times N^{e\in Q,sub,bkg}_{e\mu CR}$$

- e/m reconstruction and selection efficiency difference is taken care of by the epsilon factor (p_T, η dependent), calculated with data events from Z region.



Z+jets Background Estimation - $2l2\nu$

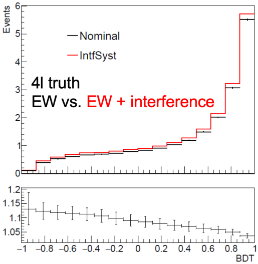
- Estimated by extrapolation (with Exponential function) from data events in low MET-significance region to high MET-significance region.
- Systematics include the MC and data-driven difference (dominant), different fitting functions, and different fitting range.



Final estimation: 0.3 + 1.5 – 0.3

Systematics on the EW VBS Signal

- Experimental systematics (based on CP recommendations).
- Theoretical (PDF, QCD scale, parton showering).
- Additional uncertainty due to the interference between EW and QCD processes, studied in truth level then convert to reconstruction level.
- Uncertainty due to interference:7(2)% in 4l ($2l2\nu$) channel. Difference mostly due to different m_{jj} cut between two channels



Summary of Signal and Background Yields

Process	$4l$ channel	$2l2\nu$ channel
EW $ZZjj$	20.6 ± 2.5	12.30 ± 0.65
QCD $ZZjj$	77 ± 25	17.2 ± 3.5
QCD gg$ZZjj$	13.1 ± 4.4	3.5 ± 1.1
Non-resonant-ll	-	21.4 ± 4.8
WZ	-	22.8 ± 1.1
Others	3.2 ± 2.1	1.15 ± 0.89
Total	114 ± 26	78.4 ± 6.2
Data	127	82

- Numbers shown in table are used to calculate the measured inclusive XS (pre-fit numbers).
- For minor backgrounds only the CP experimental systematics are included.

Inclusive Cross section Measurements

Cross sections are measured for the inclusive processes, in individual 4l and $2l2\nu$ channels in fiducial region.

- $4l$ channel fiducial region: Z window loose to [60, 120] GeV (is [66, 116] GeV for detector level) to reduce migration effect.
- $2l2\nu$ channel fiducial region: Lepton eta cuts harmonized to 2.5 for both electrons and muons. Truth MET $>$ 130 GeV instead of MET significance (difficult to define at truth level)

	Measured Fiducial σ [fb]	Predicted Fiducial σ [fb]
$\ell\ell\ell\ell jj$	1.27$\pm$0.12(stats)$\pm$0.02(theo)$\pm$0.07(exp)$\pm$0.01(bkg)$\pm$0.03(lumi)	1.14$\pm$0.04(stats)$\pm$0.20(theo)
$\ell\ell\ell\ell jj$	1.22$\pm$0.30(stats)$\pm$0.04(theo)$\pm$0.06(exp)$\pm$0.16(bkg)$\pm$0.03(lumi)	1.07$\pm$0.01(stats)$\pm$0.12(theo)

Uncertainties: statistic dominant.

MVA Analysis for EW Processes

- Gradient BDT is used in both channels.
- 4l channel trained in SR using EW vs. QCD events.
- $2l2\nu$ channel trained in SR using EW vs. all backgrounds except Zjets (due to minor contribution and large fraction of negative weight events).

Rank	$\ell\ell\nu\nu$ variables	$\ell\ell\ell\ell$ variables
1	$\Delta\eta(ll)$	m_{jj}
2	m_{ll}	leading p_T^j
3	$\Delta\phi(ll)$	subleading p_T^j
4	m_{jj}	$p_T(ZZjj)/H_T(ZZjj)$
5	$E_{\mathrm{T}}^{\mathrm{miss}} significance$	$Y(j1) \times Y(j2)$
6	$\Delta Y(jj)$	$\lvert\Delta Y(jj)\rvert$
7	$Y(j1) \times Y(j2)$	Y^*_{Z2}
8	HT	Y^*_{Z1}
9	$\Delta R(ll)$	p_T^{ZZ}
10	subleading p_T^j	m_{ZZ}
11	$E_{\mathrm{T}}^{\mathrm{miss}}$	p_T^{Z1}
12	subleading p_T^l	$p_T^{\ell 3}$
13	leading p_T^l	-

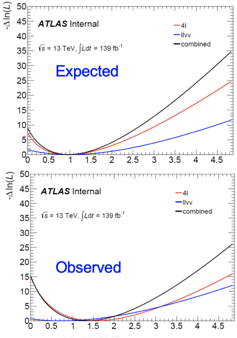
Table 48: Input variables for $\ell\ell\nu\nu$ and $\ell\ell\ell\ell$ channels.

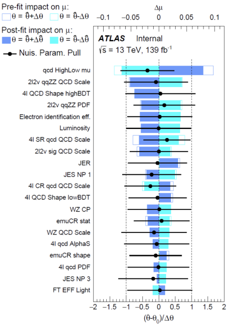
Input variables ordered by ranking

Statistical Fit for EW Processes

- BDT output is used as final discriminator for fitting.
- 3 regions enter the fit: 4l SR, 4l QCD CR and $2l2\nu$ SR.
- μ_{EW} as POI. μ_{QCD} (for 4l) also introduced as free parameter in the fit to constrain QCD normalization.

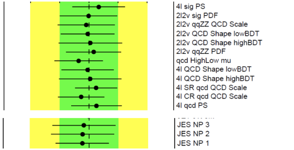
Fit Details





Observed (combined channel)

- ✓ Only the pulled ones shown here
- ✓ Full list in supporting note
- ✓ Pulls are understood as correlated to data-prediction differences in various bins.

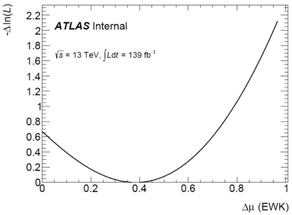


Result on EW Processes

	μ_{EW}	$\mu_{QCD}^{\ell\ell\ell\ell jj}$	Significance Obs. (Exp.)
$4l$	1.54 ± 0.42	0.95 ± 0.22	5.48(3.90) σ
$2l2\nu$	0.73 ± 0.65	-	1.15(1.80) σ
Combined	1.35 ± 0.34	0.96 ± 0.22	5.52(4.30) σ

- Cross section for EW processes (4l and $2l2\nu$) is derived ($\mu_{EW} \times \sigma_{SM}$) from combined fit, to be 0.82$\pm$0.21 fb (SM prediction0.61$\pm$0.03 fb), dominant by statistical uncertainty from data.

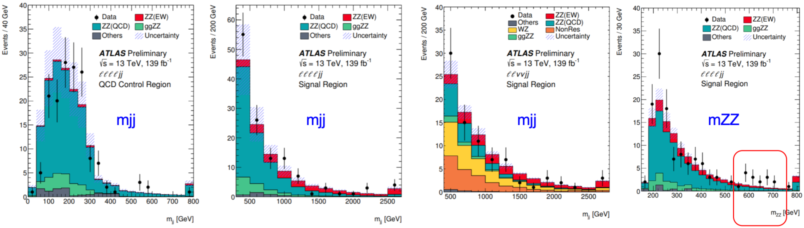
- Compatibility test: Two channels compatible within 1.2 σ.



$$\Delta\mu = \mu 4l - \mu_{2l2\nu}$$

Distribution: m_{jj} and m_{ZZ}

- Predictions have been scaled with (μ_{EW} and μ_{QCD}) from combined fit using BDT output.



Post-Fit BDT Distributions

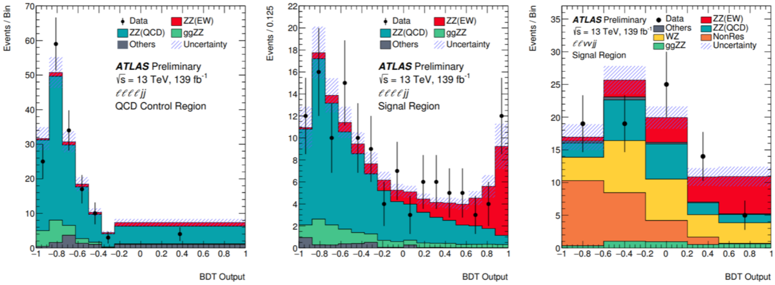


Figure 5: Observed and expected multivariate discriminant distributions after the statistical fit in the $\ell\ell\ell\ell jj$ QCD CR (left), and in the $\ell\ell\ell\ell jj$ (middle) and $\ell\ell\nu\nu jj$ (right) signal regions. The error bands include the experimental and theoretical uncertainties, as well as the uncertainties in μ_{EW} and $\mu_{\text{QCD}}^{\ell\ell\ell\ell jj}$. The error bars on the data points show the statistical uncertainty on data.

Summary

- Inclusive ZZ+2 jets cross section was measured.
- Observation of electroweak production of ZZ+2jet.
- This analysis completes observation of weak boson scattering.

Back Up

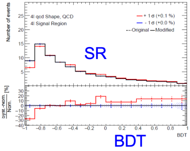
Samples Information

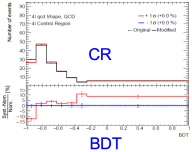
- Full Run 2 datasets

Process	Sample	Comments
EWK ZZjj	MG	Include VBF H (small contribution due to Z mass cuts)
QCD-qq ZZjj	Sherpa222	Compared with private MG sample at truth level for 4l channel
gg ZZjj	4l: Sherpa222 llvv: gg2VV	
Triboson, other diboson	Sherpa222	
Zjet	Sherpa221	
Top	Powheg, MG, aMC@NLO, Sherpa	

Modeling of the QCD ZZjj

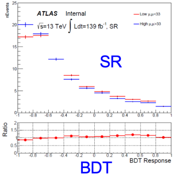
- The shape differences between different generators can not be covered by the standard QCD up/down systematics
- Additional shape systematic derived by comparing Sherpa with MG samples at truth level

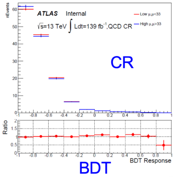




Check the pile-up modeling

- Have compared shape difference using low/high mu QCD MC events to cover the possible bias due to pile-up effect





QCD ZZjj Background Estimation-llvv

- Estimated from simulations.
- Experimental and theoretical systematics are included.
- Additional Sherpa vs. MG shape uncertainty is also included
  - Derived using 4l events at truth level in the eeμμ channel, but treat muon as neutrino