Orthogonality study for the S $\rightarrow Z_d Z_d \rightarrow$ 2l2ν/2l2j with the ATLAS detector at the LHC.

Doomnull Unwuchola[1], Lerothodi Leeuw[1], Michiel Veen[2], Loan Truong[3], Simon Connell[3], Rafael Coelho Lopes De Sa[2], Ketevi Assamagan[4]

[1]Department of Physics and Astronomy, University of the Western Cape, Cape-Town, South Africa.
[2]Department of Physics, University of Massachusetts, Massachusetts, U.S.A.
[3]Department of Mechanical Engineering Sciences, University of Johannesburg, Johannesburg, South Africa.
[4]Brookhaven National Lab.

E-mail: `doomnull@gmail.com`

Abstract. The hidden abelian Higgs model is used for a search for an additional scalar decaying to two Z-dark bosons (Z_d) to two leptons, two neutrinos (2l2ν). The search uses pp collision data collected with the ATLAS detector at the LHC with an integrated luminosity of 139 fb^{-1} at a centre-of-mass energy $\sqrt{s} = 13$ TeV. This is a follow-up to the study of the 4l final states. In our 2l2ν channel analysis, using the Run-2 data with the ATLAS detector to conduct the search for an additional scalar with a distinct mass from the Higgs boson demands study of signal overlap from the 2l2j channel. A technique is introduced to separate signal events of our 2l2ν channel from that of the 2l2j channel. We present the work and result of the orthogonality study done to achieve this.

1 Introduction

The Standard Model (SM) explains a wide range of phenomena but leaves unanswered questions such as the origin of dark matter and possible hidden sectors [1, 2, 3, 4, 5] . In our search for S $\rightarrow Z_d Z_d \rightarrow$ 2l2ν, we use the Hidden Abelian Higgs model [1] which could provide a viable dark matter (DM) candidate . Considering lepton universality for the decay of Z_d to leptons, our search is a follow up to the study of the 4l final states [6], where in our 2l2ν channel, more events and higher sensitivity expected because the Branching Ratio in the Standard Model is: BR(ZZ $\rightarrow$ 2l2ν) = 6*BR(ZZ $\rightarrow$ 4l). The main background is Z+jets with minor contribution from eμ ($t\bar{t}$, WW, Wt), dibosons and fakes. To suppress the background our chosen mass range is $m_S \leq 183$ GeV and $m_{Zd} < 70$ GeV. Considering the potential overlapping signatures due to misreconstructed jets and missing transverse energy (MET) involved in the 2l2j and the 2l2ν channels, clear separation (orthogonality) of the two channels is essential for an unbiased search. The mass points used in this study are [m_S , m_{Zd}] : [70, 20], [70, 35], [84, 31], [110, 30], [110, 55] and [183, 50] GeV.

The two channels as shown in the Feynman diagram_S in Figure 1 demands an orthogonality study so that the data does not get used twice. To achieve this, a likelihood function that will select events for 2l2j where the di-lepton mass matches with the di-jet mass is proposed to be used. This demands using, in the orthogonality study, only the final state with corresponding Feynman diagram_S for S $\rightarrow Z_d Z_d \rightarrow$ 2l2j production where its appropriate Z-dark can be reconstructed. This implies any final state where the second jets escaped as in $|\eta| > 4.9$ was not used in this study.

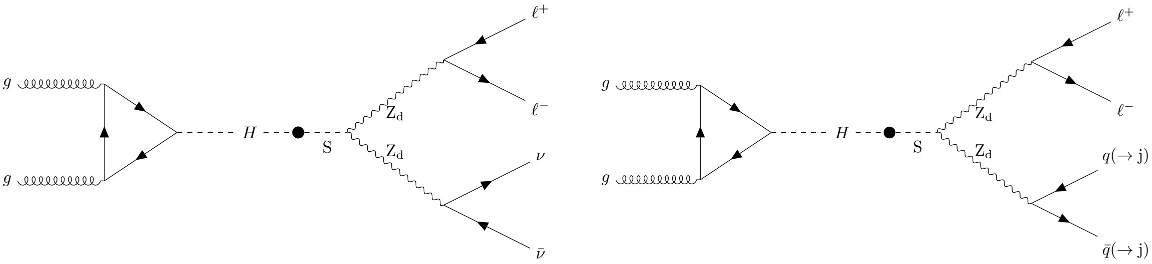

Figure 1: (*left*) The $2l2\nu$ channel. (*right*) 2l2j channel.

2 The likelihood definition

The likelihood function L is defined to evaluate the compatibility between the dijet and dilepton systemm_S :

$$L = \max_{j_1,j_2} \max_{E_1,E_2} \mathrm{G}(m_{jj}(E_1, E_2) - m_{\ell\ell} \mid 0, 5) \times \mathrm{G}(E_1 \mid E_{j_1}, \sqrt{E_{j_1}}) \times \mathrm{G}(E_2 \mid E_{j_2}, \sqrt{E_{j_2}}) \tag{1}$$

G is the Gaussian distribution $\mathrm{G}(x \mid \mu, \sigma)$; E_{j_1}, E_{j_2} represent the measured jet energies; E_1, E_2 represent the hypothetical jet energies tested in maximization; and $m_{\ell\ell}$ is the measured dilepton invariant mass. m_{j1j2} is the measured or initial dijet mass, computed from the four-momenta of jet 1 and jet 2 before any energy rescaling. It is a fixed value, given by the invariant mass formula:

$$m_{j1j2} = \sqrt{(E_{j1} + E_{j2})^2 - (\vec{p}_{j1} + \vec{p}_{j2})^2} \tag{2}$$

Here, E_{j1}, E_{j2} are the measured jet energies, and $\vec{p}_{j1}, \vec{p}_{j2}$ are the three-momenta of the jets. m_{j1j2} is the fixed, measured dijet mass, based on detector-reconstructed jet momenta and energies. Where:

$$m_{jj}(E_1, E_2) = m_{j_1j_2} \sqrt{\frac{E_1}{E_{j_1}}} \sqrt{\frac{E_2}{E_{j_2}}} \tag{3}$$

$m_{jj}(E_1, E_2)$ is the modified dijet mass used during likelihood maximization. It varies with the chosen values E_1 and E_2, and rescales m_{j1j2}.

For a given masspoint 2l2j sample as shown in Figure 2, before applying any cut, the negative logarithm likelihood fit function was plotted with a varying line demarcation in preparation for the classification of events as for 2l2j and $2l2\nu$ channels. This implies $(100\ \%)_{2l2\mathrm{j}}$ samples = [$(\%)_{<2\mathrm{j}}$ + $(\%)_{2l2\mathrm{j}}$ + $(\%)_{2l2\nu}$] events. Subsequently, events selected to be evaluated by the likelihood function where cuts will be applied are those with jets ≥ 2. The standard deviation of the modified dijet and dilepton invariant mass term of 5 GeV was found to be efficient.

3 Orthogonality Strategy

The main goal is to loop over all jet pairs $(\mathrm{j}_1, \mathrm{j}_2)$ and energies (E_1, E_2) in order to maximize the likelihood L by adjusting hypothetical jet energies E_1, E_2 to best match the dijet system to the dilepton system. This involves penalising mismatch between adjusted dijet mass and observed dilepton mass which also considered a Gaussian distribution for the correspondence between the reconstructed mass from the jets and the leptons. This distribution I centred at 0 with width $\sigma = 5\,\mathrm{GeV}$ expressed by the term $\mathrm{G}(m_{jj}(E_1, E_2) - m_{\ell\ell} \mid 0, 5)$. Concurrently, using the probabilistic model of the "true" jet energy E_i given the measured energy E_{j_i} with a Gaussian distribution centered at E_{j_i}, width $\sigma = \sqrt{E_{j_i}}$ (reflects calorimeter resolution) expressed by the term $\mathrm{G}(E_i \mid E_{j_i}, \sqrt{E_{j_i}}) \quad \text{for } i = 1, 2$. For the adjusted dijet mass expressed as $m_{jj}(E_1, E_2) = m_{j_1j_2}\sqrt{\frac{E_1}{E_{j_1}}}\sqrt{\frac{E_2}{E_{j_2}}}$, the invariant mass is rescaled based on hypothetical energies with $m_{j_1j_2}$ representing the measured invariant mass of jets j_1 and j_2.

So the key components are to iterate over the event data and find combination that maximizes the likelihood L, match modified dijet mass to dilepton mass ($m_{jj} \approx m_{\ell\ell}$), and compare hypothesized jet energies E_1, E_2 with measured E_{j_1}, E_{j_2}. The interpretation is that a higher L implies more likely the event matches $2\ell 2$j topology and thus, implies a lower $-\ln L$ results to the event being chosen as $2\ell 2$j. Finally, we apply a cut on $-\ln L$ to separate (orthogonalize) event types ($2\ell 2$j vs $2\ell 2\nu$).

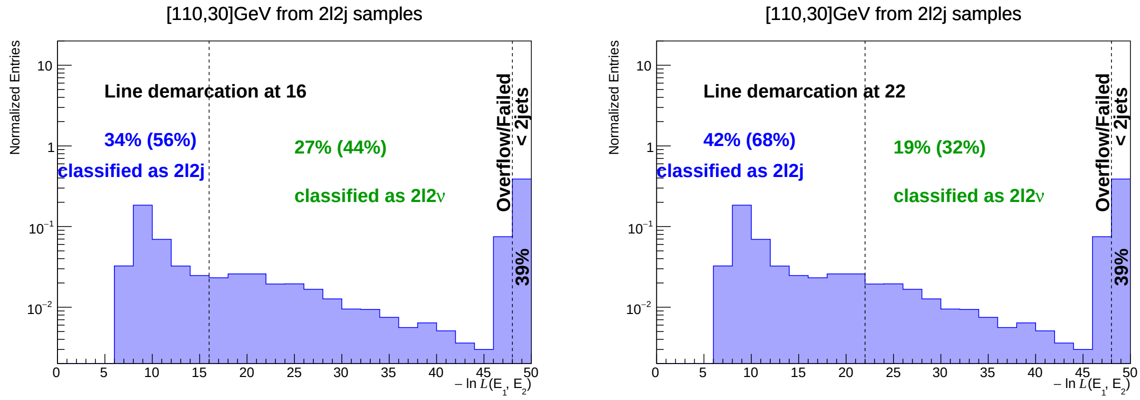

Figure 2: (*left*) The plot of the negative logarithm of the likelihood function. [m_S, m_{Zd}]=[110, 30] GeV after excluding "$<$ 2j" events, classified 2l2j vs. 2l2ν with varying line demarcation this time at 16. (*right*) The plot of negative logarithm of the likelihood function. [m_S , m_{Zd}]=[110, 30] GeV after excluding "$<$ 2j" events, classified 2l2j vs. 2l2ν with varying line demarcation this time at 22.

3.1 Representation of the Likelihood function in the Code

The mathematical equation of the fit applied in the code is based on a **Gaussian likelihood** minimization. The full likelihood function is then:

$$L(E_1, E_2) = P_{\text{lep}}(M_{\text{combined}}) \cdot P_{\text{jet1}}(E_1) \cdot P_{\text{jet2}}(E_2) \tag{4}$$

Each term is modeled as a Gaussian distribution. E_1 and E_2 are the energies of `jet1` and `jet2`, M_{combined} is the invariant mass of the combination of jets and leptons.

$$L(E_1, E_2) = \exp\left(-\frac{(M_{\text{combined}} - M_{\text{lep}})^2}{2 \cdot \sigma_{\text{lep}}^2}\right) \cdot \exp\left(-\frac{(E_1 - \mu_{\text{jet1}})^2}{2 \cdot \sigma_{\text{jet1}}^2}\right) \cdot \exp\left(-\frac{(E_2 - \mu_{\text{jet2}})^2}{2 \cdot \sigma_{\text{jet2}}^2}\right) \tag{5}$$

μ_{jet1} and μ_{jet2} are the mean energies of `jet1` and `jet2`, i.e., the original jet energies,σ_{jet1} and σ_{jet2} are the standard deviations of the jet energies, taken as $\sqrt{E_1}$ and $\sqrt{E_2}$ in the code. M_{lep} is the invariant mass of the dilepton system, σ_{lep} is fixed to 5 GeV.

Since likelihood is maximized in principle, the code minimizes the negative log-likelihood.

$$-\ln L(E_1, E_2) = \frac{(M_{\text{combined}} - M_{\text{lep}})^2}{2 \cdot \sigma_{\text{lep}}^2} + \frac{(E_1 - \mu_{\text{jet1}})^2}{2 \cdot \sigma_{\text{jet1}}^2} + \frac{(E_2 - \mu_{\text{jet2}})^2}{2 \cdot \sigma_{\text{jet2}}^2} \tag{6}$$

Given the approximations, the standard deviations for jet energies are approximated by $\sigma_{\text{jet}} \approx \sqrt{E}$, and $\sigma_{\text{lep}} = 5.0$ GeV. The final form for minimisation is:

$$-\ln L(E_1, E_2) = \frac{(M_{\text{combined}} - M_{\text{lep}})^2}{2 \cdot (5.0)^2} + \frac{(E_1 - \mu_{\text{jet1}})^2}{2 \cdot E_1} + \frac{(E_2 - \mu_{\text{jet2}})^2}{2 \cdot E_2} \tag{7}$$

This is the expression to be minimized using a variable metric gradient method.

4 Results

The plots for the reconstruction of the scalar mass m_S from the 2l2j decay channel included simulated data contributions for the background represented by the Z+jets events (main background in the analysis channel) in order to also monitor how their contributions to the shape of the scalar mass distribution m_S varies and relate to one another as the cuts are applied. As this likelihood fit function is used in our search to select 2l2j events from those eventually considered for the 2l2ν analysis channel, we scan through the range of numerical cut values to optimise the negative logarithm of the likelihood function ($-\ln L$) final selected cut value. After certain cut values on the $-\ln$ L parameter, the number of signal events are not much less than the signal events without the cut implemented. This is the threshold final selected cut value. For $-\ln L <$ this threshold cut value, the events are for 2l2j channel and for $-\ln L$ otherwise, the events are made available for our 2l2ν channel as shown in Figure 3.

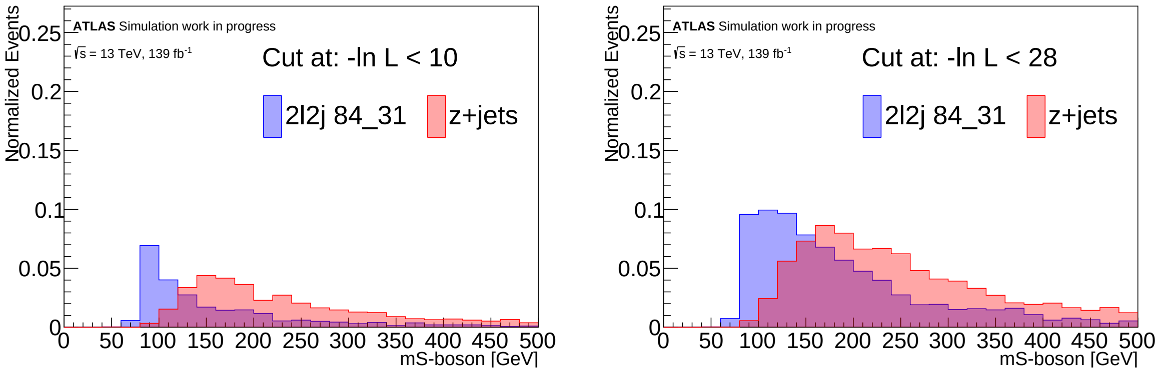

Figure 3: (*left*) As the negative logarithm of the likelihood function $(-\ln L) < 16$, events are made available for the 2l2j channel. (*right*) As $-\ln L$ not less than 16, the events are made available for 2l2ν channel.

For further clarity of the likelihood fit function results for the orthogonality study, the efficiencies for these six mass points used in the analysis are plotted. The fractional display of signal events expressed in percent of signal events with cut values on m_S is shown in Figure 4. Beyond a specific value of the cut, the percentage of signal events, remaining saturates in a plateau. This higher percentage yield is quite obvious in higher mass points. Using the threshold cut value of 16, the efficiencies of the mass points $[m_S\ ,\ m_{\mathrm{Zd}}]$: [70, 20], [70, 35], [84, 31], [110, 30], [110, 55] and [183, 50] GeV used in the orthogonality study are : 40%, 56%, 55 %, 64 %, 77 % and 82 % respectively.

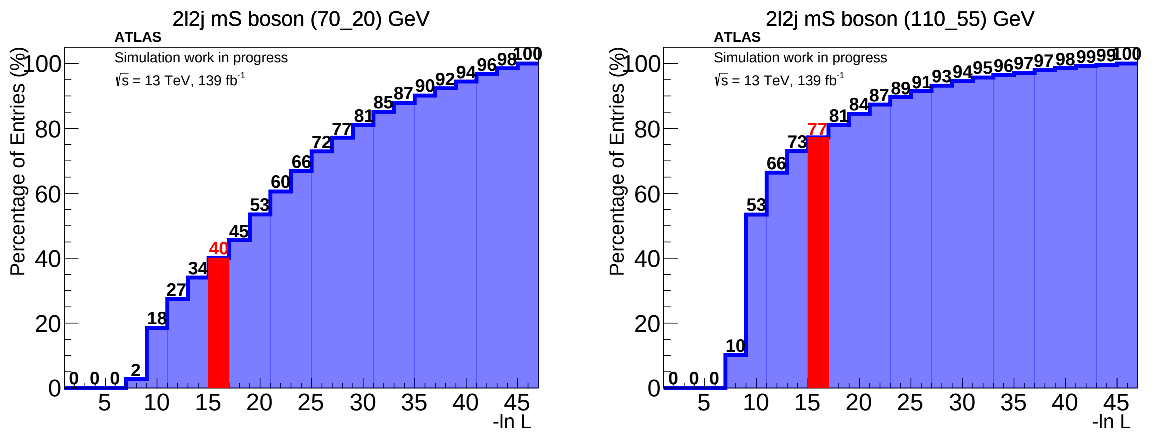

Figure 4: (*left*) [70, 20]GeV % of events tend to plateau. (*right*) Higher % yield more prominent for this masspoint.

5 Conclusion

We have developed a likelihood-based method to orthogonalize the events that are to be classified as belonging to the 2l2j and 2l2ν analyses channels. The aim is to select events for 2l2j that have a negative log likelihood below $(nll < 16)$. The 2l2j events that are cut away in this selection are not signal-like. That is the reconstructed mass for m_S from the di-lepton and di-jet mass of these events is not commensurate. A method has been presented which improves the reconstruction of the di-jet mass for the 2l2j analysis channel which can then be deployed to improve the signal/background separation.

References

[1] ATLAS Collaboration, "Search for Higgs boson decays to hidden-sector particles in the prompt lepton-jet channel with the ATLAS detector," 2025, aTL-PHYS-PROC-2025-029. URL: https://cds.cern.ch/record/2898771.

[2] atlas Collaboration, “Search for a new scalar resonance decaying into Z bosons in four-lepton final states with the ATLAS detector at the LHC,” *Journal of High Energy Physics*, vol. 2022, no. 7, p. 67, 2022.

[3] ATLAS Collab., “Search for dark photons produced in 13 TeV pp collisions with the ATLAS detector,” *Physical Review D*, vol. 104, no. 11, p. 112006, 2021.

[4] ATL. Collaboration, “Search for exotic Higgs boson decays to light resonances with the ATLAS detector,” 2020, aTL-PHYS-PUB-2020-012. [Online]. Available: https://cds.cern.ch/record/2725735

[5] A. Collab., “Search for invisible decays of a higgs boson produced via vector boson fusion in pp collisions at sqrt(s) = 13 tev with the atlas detector,” *Physical Review D*, vol. 100, no. 1, p. 012006, 2019.

[6] A. Collaboration, “Search for a new scalar decaying into new spin-1 bosons in four-lepton final states with the atlas detector,” arXiv:2410.16781 [hep-ex], URL: https://arxiv.org/abs/2410.16781, 2024, cERN-EP-2024-248.