Signal Modeling Studies for VH4bLLP Exotics Analysis with the ATLAS Experiment

Helen Asuküla
University of Tartu

Supervised by:
Manfredi Ronzani
New York University

Abstact

This report describes a study performed in the VH4bLLP exotic decay model. An exotic decay of the Higgs boson, produced in association with the W-boson, into a pair of Long Lived Particles (LLP), is considered. Invariant di-jet masses of the LLPs are reconstructed and analysed from Monte Carlo simulation of the considered process.
The first part of the study compares FullSim and AtlFast-II Monte Carlo simulation methods by reconstructing invariant di-jet masses of LLPs. The Kolmogorov–Smirnov goodness-of-fit test, a statistical test for comparing distributions, is also performed. The second half of the analysis looks at reconstructed invariant masses in simulations with different input mean mass m_{LLP} and mean proper lifetime $c\tau$ of the LLPs. This signal result is compared with the background estimations from Monte Carlo simulations.

CERN Summer Student Programme
3 June - 26 July 2019

Contents

1 Theoretical overview

The Standard Model (SM) is the current theory of fundamental particles and interactions [1]. It is in agreement with many experimental tests, but several limitations of the SM have been shown by a number of experimental observations and theoretical arguments, such as the lack of description for the gravitational force, the predominance of the matter against the antimatter, and the unsolved problem of the existence of the Dark Matter [2]. Therefore, there are many pursuits to go beyond the Standard Model [3]. A way of expanding the Standard Model is to look for "hidden sectors" – new particles, which may have not yet been detected due to their weak couplings with the SM particles [2][3].

The Higgs boson could be a possible mediator between the SM and a hidden sector with possible exotic decays. Long decay times arising from the weak coupling lead to the search of Long Lived Particles (LLP) [2]. The VH4bLLP model theorizes an exotic decay of the Higgs into two spin-zero LLPs called the a-boson, $H \to aa$. The a-scalar boson is then decaying into a pair of opposite charge quarks, which can be of the light and strong flavours. In this model, the Higgs boson is created alongside a vector boson, either a W-boson or a Z-boson. Figure 1 represents the process occurring in association with the W-boson, called the WH channel [4].

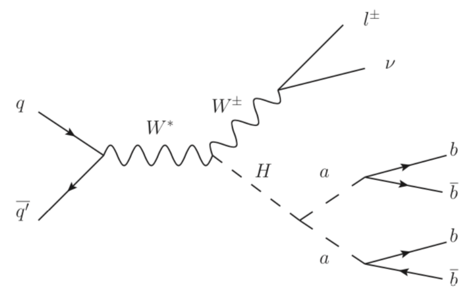


Figure 1: Feynman diagram for the WH production process with the subsequent decay H→aa→4b [4].

2 The ATLAS Experiment

The ATLAS experiment is a multi-purpose experiment operating at the Large Hadron Collider (LHC) [6]. The ATLAS detector uses a coordinate system, which originates from the interaction point. The z-axis points along the beam line and the x-y plane is transverse to the beam − the x-axis points towards the centre of the LHC ring and the y-axis points upwards [5]. In cylindrical coordinates, the azimutal angle ϕ is measured in the x-y plane around the beam. The polar angle θ, which is measured from the beam axis, is used to define pseudorapidity

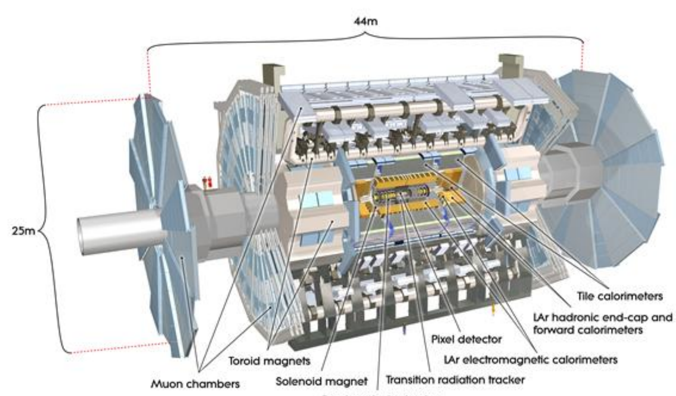


Figure 2: Image of the ATLAS detector [5]

$$\eta = -\ln\left\{\tan(\theta/2)\right\}. \tag{1}$$

The transverse momentum p_T is the projection of momentum onto the x-y plane [4].

The ATLAS experiment has several simulation softwares that run event simulation, detector-simulation and event reconstruction. The full simulation (FullSim) features a detailed detector-simulation, which gives topmost accuracy, but is also highly consuming in terms of CPU and time resources [6]. The fast simulation (AtlFast) can be used for large Monte Carlo (MC) simulations as this method is less CPU-intensive [7], but it produces less accurate results due to its use of parametrisation of particle showers in the calorimeters among others [6]. As of 2012, the simulation method AtlFast-II is available with further reduced simulation time [7].

3 Software overview

The analysis in this project was done using the scientific software toolkit ROOT. ROOT has its own Jupyter Notebook kernel, which allows the user to write a project in both languages [8]. The project presented here was written in the SWAN (**S**ervice for **W**eb based **AN**alysis) environment, which has a web-based Jupyter Notebook interface.

The project used simulation and data samples, which were processed with the VH4bLLP Exotics Analysis project's algorithm called 'Hto4bLLPAlgorithm'. It was hosted on CERN's LXPLUS service and run on ATLAS's PanDA Production and Distributed Analysis System.

4 Comparing FullSim and AtlFast-II simulation methods

The VH4bLLP Exotics Analysis project intends to use Monte Carlo simulations for the signal modeling. As AtlFast-II (referred to as just AtlFast from now on) is faster and takes less resources, it would be the preferred method. However, it is necessary to check that the differences from the FullSim method would not affect the results in any significant way. Some preliminary studies using di-jet masses for comparison have been performed within the analysis group providing no conclusive answer. The task of this project is to firstly repeat the study, also as a learning tool.

The comparison uses the WH channel exotic decay simulations with parameters set to: Higgs boson mass $H = 125$ GeV, a-boson mean mass at $m_a = 55$ GeV and a-boson mean proper lifetime $c\tau_a = 100$ mm. The results combined W^--boson and W^+-boson scenarios.

4.1 Invariant di-jet mass calculation

For mass calculation, some requirements were introduced to the reconstructed jets. Firstly, their transverse momentum had to satisfy $p_T > 20$ GeV and pseudorapidity $|\eta| < 2.1$. For jets passing those criteria, a Lorentz four-vector (p_T, η, ϕ, E) was created. The jet was counted only if it was matched to the a_1 or a_2 boson using a matching procedure which takes advantage of the available information on the particle decay chain for simulated events. The jets originating from the showering of particles, identified as coming from the a-scalar decay, are further selected if their transverse position is within a cone of 0.2 from the a-scalar one. This information is stored in the Ntuple in a vector called "jet_LLP_index".

For each event, the masses of both of the LLPs were tentatively reconstructed. A mass would only be calculated if there were two or more jets associated with that LLP. The two most energetic jets associated to the LLP were selected and their Lorentz vectors added together. The invariant mass of the jets system is then calculated.

4.1.1 Results

A total of 42 673 events were present in FullSim and 43 553 events in AtlFast. Out of those events, 12 965 and 12 915 LLP masses were calculated for FullSim and AtlFast respectively. A preliminary study showed that both LLPs behaved similarly, therefore the two LLP reconstructed masses has been counted in the same histogram, allowing the doubling of the statistics.

In Figure 3, the di-jet mass distributions from AtlFast and FullSim can be observed. The mass distributions peak at around 55 GeV as expected from the simulation input a-boson mean mass. The shapes are vaguely Gaussian around the peak, but have a longer tail towards higher masses. As first attempt of comparison, the ratio between the two distributions has been fit with a linear function in the higher statistics range.

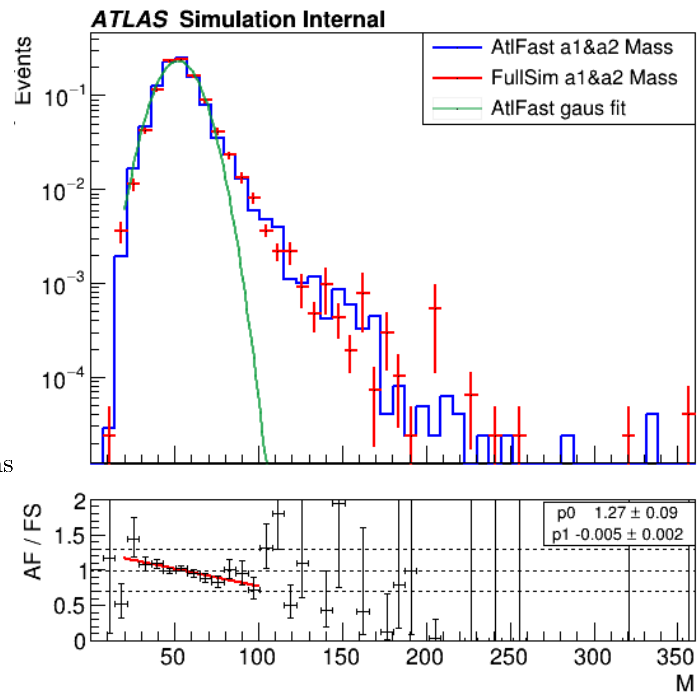


Figure 3: LLP mass comparison in FullSim and AtlFast, plots fitted in M range 20-100 GeV.

4.2 Considering LLP lifetime

Events have two LLP lifetimes reconstructed in the variable "LLP_R _coord", describing the distances the a-bosons travelled before decaying. The study has been binned in different ranges of the LLPs lifetime: 0-100 mm, 100-200 mm, 200-300 mm and over 300 mm. The radius of 300 mm is just above the Pixel Detector of ATLAS. The ranges do not follow precisely the radial division of the inner detectors as di-jet invariant

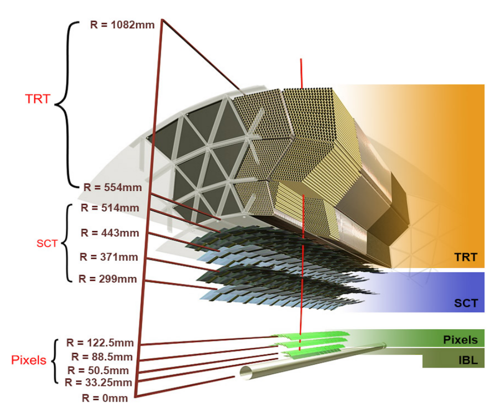


Figure 4: Schematic view of the ATLAS detector [9]

masses are calculated by using only calorimeter information.

The steps described in section 4.1 were taken separately for each range. The di-jet invariant mass distributions in the LLP lifetime ranges can be seen in Figure 5.

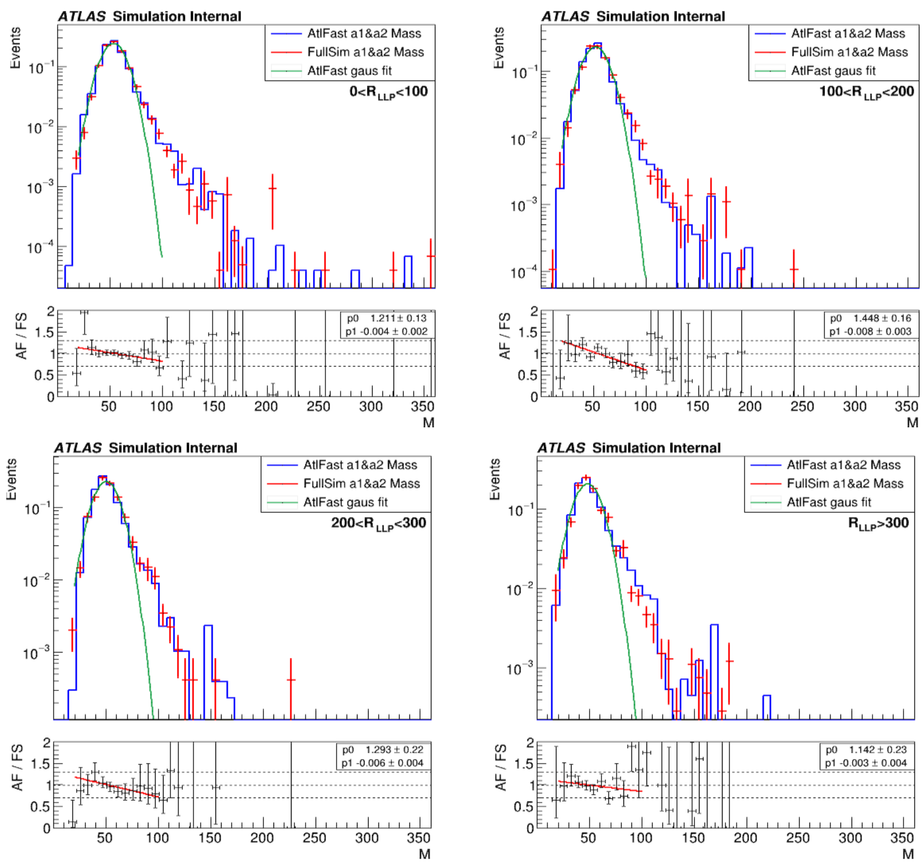


Figure 5: LLP masses in lifetime ranges of 0-100 mm (top left), 100-200 mm (top right), 200-300 mm (bottom left) and 300+ mm (bottom right), plots fitted in M range 20-100 GeV.

4.3 Kolmogorov–Smirnov Goodness-of-Fit Test

The Kolmogorov–Smirnov test can be used to compare a sample to a known distribution [10]. The ROOT offers a Kolmogorov–Smirnov test for histograms, but notes that the test is intended for unbinned data. The test returns the probability that two histograms are of the same distribution. The TH1 class documentation notes that the result with binned data will be shifted slightly higher than expected.

4.3.1 Statistical test for all events

The histogram in Figure 3 has 50 bins. A reference histogram was created with 100 000 bins to check the effects of binning on the result. In both cases normalization was also compared. The results were:

50 bins	100 000 bins
$p = 0.302$	$p = 0.128$

Table 1: Kolmogorov–Smirnov test results for AtlFast and FullSim.

The Kolmogorov–Smirnov test was also run for LLP lifetime ranges of 0-100 mm, 100-200 mm, 200-300 mm and 300+ mm.

R_{LLP}	50 bins	100 000 bins
0-100 mm	$p = 0.612$	$p = 0.549$
100-200 mm	$p = 0.270$	$p = 0.111$
200-300 mm	$p = 0.257$	$p = 0.209$
300+ mm	$p = 0.739$	$p = 0.299$

Table 2: Kolmogorov–Smirnov test results for AtlFast and FullSim in different LLP ranges.

4.3.2 Statistical test for filtered events

The previous comparison was done using all available events for improved statistics. Other studies performed on the signal modeling showed an improvement of the signal efficiency if a preliminary filter would be applied to the dataset. The filter takes advantage of some kinematic features of the LLP signal which cannot be discussed here in details. As the filter will be applied to all the events used in the analysis, it is interesting to perform the test over the filtered signal sample. The filter reduced the number of entries in histograms to 6197 for FullSim and 6161 for AtlFast. The Kolmogorov-Smirnov test results in this case were:

50 bins	100 000 bins
$p = 0.501$	$p = 0.110$

Table 3: Kolmogorov–Smirnov test results for AtlFast and FullSim for all LLP lifetimes with filtered events.

R_{LLP}	50 bins	100 000 bins
0-100 mm	$p = 0.780$	$p = 0.531$
100-200 mm	$p = 0.169$	$p = 0.115$
200-300 mm	$p = 0.983$	$p = 0.976$
300+ mm	$p = 0.388$	$p = 0.370$

Table 4: Kolmogorov–Smirnov test results for AtlFast and FullSim in different LLP ranges with filtered events.

4.4 Analysis of the results

The effect of the LLP lifetime does not have a substantial impact on the invariant di-jet masses of the LLPs, as the distributions remain relatively similar from Figure 3 to 5. According to the statistical test, AtlFast and FullSim are most compatible in the LLP lifetime range 0-100 mm with all events and and compatibility increases if selecting only filtered events. It seems safe to conclude that the following analysis may look at LLPs of all lifetimes.

There remain differences between the results FullSim and AtlFast simulation methods produce for invariant di-jet masses of the LLPs. The Kolmogorov–Smirnov test does not rate compatibility particularly high. As such, there has not been a definitive final conclusion on this matter at the time of writing. The discrepancy seems minimal enough that the analysis continues with the AtlFast simulation method.

5 Study on the distribution of the LLP mass for Signal and Background processes

Several other physics processes can lead to the same reconstrusted final state as the one produced by the exotics decay of the Higgs boson described in Chapter 1. This can happen either if those processes actually have the same final state, or because some information about their final state is missing or mis-reconstructed and they eventually appear as signal. In this part of the project, the signal-to-background discrimination properties of the LLP di-jet variable have been tested against the background processes that are most likely contributing to the final event yield.

This part of the study analyses a total of six AtlFast signal simulations. The different simulation parameters are mean a-boson mass of $m_a = 55$ GeV and $m_a = 15$ GeV and mean proper lifetime of $c\tau_a = 1$ mm, $c\tau_a = 10$ mm and $c\tau_a = 100$ mm.

The background consisted of AtlFast Monte Carlo simulations of two physics processes. Top-antitop quark pairs $t\bar{t}$ can decay into a pair of a W-boson and a b-quark, which can be mis-identified as the signal process. A W-boson can be created in association with jets and its decay can mimic the WH channel decays.

5.1 Invariant di-jet mass calculation

The background processes do not include a-bosons and therefore, the jets cannot be sorted into a_1 and a_2 as before. To calculate signal and background invariant di-jet mass comparably, a new method had to be used.

The jet variable CHF (Charged Hadron Fraction) has been taken in consideration. It is defined as the ratio between the transverse momentum of the tracks associated to the jets as calculated in the ATLAS tracker over the transverse momentum of the jets calculated with the energy deposits in the calorimeter.

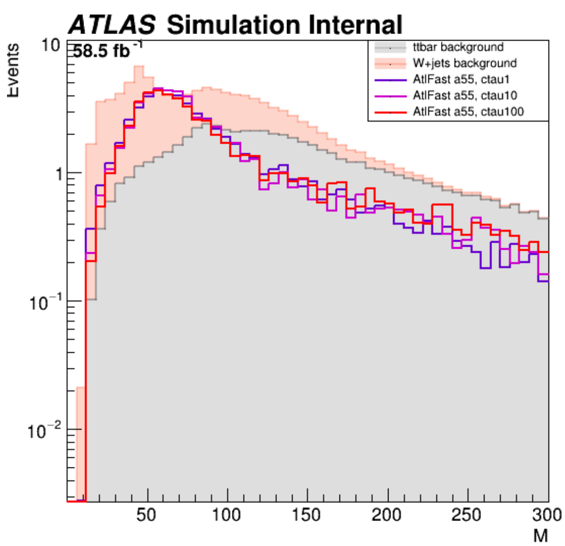
$$CHF_{\text{jet}} = \frac{\sum\limits_{\text{jet tracks}} (p_T)^{\text{jet track}}}{(p_T)^{\text{jet}}} \quad (2)$$

A low CHF would describe jets, whose creation has occurred after the tracker, more likely coming from the LLP decay. A value of CHF<0.3 has been applied as signal selection cut. The other requirements for mass calculation remained the same.

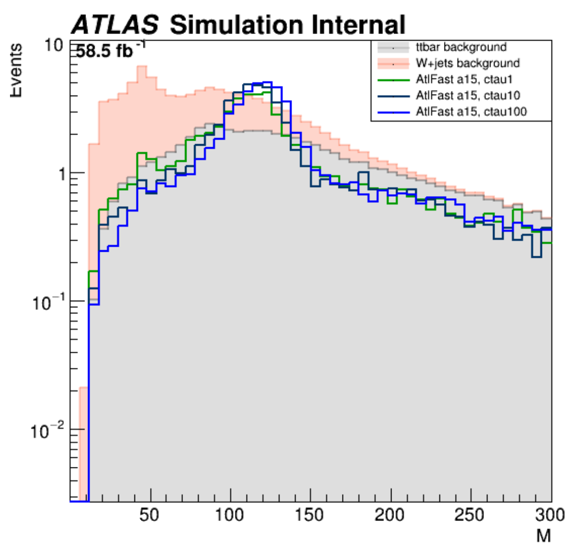
5.2 Results

The Figure 6a shows the invariant masses from simulations with $m_a = 55$ GeV, where $c\tau$ takes values of 1, 10 and 100 mm. The distributions peak at around 55 GeV, which is the expected a-boson mass.

In the $m_a = 15$ GeV case in Figure 6b, a peak around the mass of the Higgs (125 GeV) can be seen. This can be interpreted as the so-called "boosted scenario", where the a-boson has low mass and most of the energy is given to the outcoming jets, whose angular distribution is more collimated. As the result, the two most energetic jets will add up directly to the mass of the Higgs, instead of the mass of the individual intermediate a-boson.



(a) $m_a = 55$ GeV with background



(b) $m_a = 15$ GeV with background

Figure 6: Invariant di-jet mass distributions of signal with (a) $m_a = 55$ GeV, (b) $m_a = 15$ GeV and $c\tau = 1$, $c\tau = 10$ and $c\tau = 100$ mm with background estimations.

It should be noted that the normalization of the background requires further study. The background events may need different event weights for proper comparison with signal. Corrected normalization may yield better results, while currently the discriminating power of the reconstructed invariant mass variable seems rather limited.

6 Conclusion

The study compared FullSim and AtlFast Monte Carlo simulation methods by reconstructing invariant di-jet masses of Long Lived Particles. The mass distributions from FullSim and AtlFast simulations both peaked at around 55 GeV, but the ratio of AtlFast over FullSim was titled upwards in the smaller masses. A statistical Kolmogorov-Smirnov goodness-of-fit test was performed between FullSim and AtlFast mass distributions. The results were probability of 0.128 for all events and 0.110 for filtered events.

The masses were also compared according to their lifetime. The lifetime ranges considered were 0-100 mm, 100-200 mm, 200-300 mm and over 300 mm. The distributions remained similar to the previous comparison. The Kolmogorov-Smirnov test showed FullSim and AtlFast to be most compatible when considering filtered events in lifetime range 200-300 mm with a probability of 0.976 at 100 000 bins.

A deeper study of reconstructed LLP invariant di-jet masses from AtlFast simulations with different parameters was done. The mean LLP mass was set to either $m_a = 55$ GeV or $m_a = 15$ GeV, the mean proper lifetime was $c\tau = 1$ mm, $c\tau = 10$ mm or $c\tau = 100$ mm. A test on the bakground rejection power of the di-jet invariant mass has been performed using different background simulated processes which can mimic the WH 4b signal.

References

[1] ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 2*. Technical Design Report ATLAS. CERN, Geneva, 1999.

[2] David Curtin and Raman Sundrum. Flashes of Hidden Worlds at Colliders. 2017.

[3] Joseph D. Lykken. Beyond the Standard Model. In *CERN Yellow Report CERN-2010-002, 101-109*, 2010.

[4] The ATLAS collaboration. Search for the Higgs boson produced in association with a vector boson and decaying into two spin-zero particles in the H $\to$ aa $\to$ 4b channel in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector. *Journal of High Energy Physics*, 2018(10):31, Oct 2018.

[5] The ATLAS Collaboration et. al. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003–S08003, Aug 2008.

[6] ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1999.

[7] Wolfgang Lukas. Fast Simulation for ATLAS: Atlfast-II and ISF. *Journal of Physics: Conference Series*, 396(2):022031, Dec 2012.

[8] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[9] Karolos Potamianos. The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider. Technical Report ATL-PHYS-PROC-2016-104, CERN, Geneva, Aug 2016. 15 pages, EPS-HEP 2015 Proceedings.

[10] NIST/SEMATECH. e-Handbook of Statistical Methods, 2019.