

A Tale of two Higgs: Search for pair  
production of Higgs bosons in the  $b\bar{b}b\bar{b}$   
final state using proton–proton collisions at  
 $\sqrt{s} = 13$  TeV with the ATLAS detector

A DISSERTATION PRESENTED  
BY  
BAOJIA TONG  
TO  
THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY  
IN THE SUBJECT OF  
PHYSICS

HARVARD UNIVERSITY  
CAMBRIDGE, MASSACHUSETTS  
MAY 2018

©2017-2018 – BAOJIA TONG  
ALL RIGHTS RESERVED.

# A Tale of two Higgs: Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

## ABSTRACT

This thesis presents a search for Higgs boson pair production, with the  $b\bar{b}b\bar{b}$  final state. This search uses the full 2015 and 2016 data collected by the ATLAS Collaboration at  $\sqrt{s} = 13$  TeV, corresponding to  $3.2 \pm 0.2 \text{ fb}^{-1}$  of 2015 and  $32.9 \pm 1.1 \text{ fb}^{-1}$  of 2016  $pp$  collision data. Improvements with respect to the previous analysis come from the increased dataset, detailed background estimation and additional signal regions. Search sensitivity is specially enhanced for the resonance signals between 2500 GeV and 3000 GeV. The data is found to be compatible with the Standard Model predictions, and no signs of new physics have been observed. The results are interpreted in the context of the bulk Randall-Sundrum warped extra dimension model with a Kaluza-Klein graviton  $k/\bar{M}_{\text{Pl}} = 1.0$  or  $2.0$  decaying to  $hh$ , and the Type 2 two-Higgs doublet model (2HDM) where the neutral heavy CP-even  $H$  scalar decays to  $hh$ .

# Contents

o INTRODUCTION	I
1 THEORY AND MOTIVATION	5
1.1 The Standard Model and the Higgs Boson . . . . .	5
1.2 Standard Model di-Higgs production . . . . .	7
1.3 Beyond the Standard Model physics di-Higgs production . . . . .	9
1.4 Di-Higgs decay and LHC previous search results . . . . .	13
1.5 Resolved and Boosted . . . . .	15
2 LHC AND ATLAS	17
2.1 The Large Hadron Collider . . . . .	18
2.2 A Toroidal LHC ApparatuS . . . . .	20
3 RECONSTRUCTION AND OBJECTS	26
3.1 ID Tracks and Vertices . . . . .	27
3.2 Jets . . . . .	28
3.3 Flavor Tagging . . . . .	31
3.4 Leptons . . . . .	34
4 DATA AND SIMULATION	37
4.1 Data . . . . .	37
4.2 MC . . . . .	39
4.3 Signal . . . . .	40
5 EVENT SELECTION	42
5.1 Data Cleaning . . . . .	42
5.2 Trigger . . . . .	43
5.3 Object Selection . . . . .	44
5.4 Resolved Veto . . . . .	50
5.5 2D Higgs Mass Cut . . . . .	53

5.6	Number of $b$ -tagging requirement . . . . .	54
5.7	Signal efficiency and cutflow . . . . .	55
6	BACKGROUND ESTIMATION	59
6.1	Background Estimation . . . . .	61
7	CONCLUSION	65
	REFERENCES	67

# Listing of figures

1.1	Fermions and bosons of the Standard Model and their properties <sup>1</sup> , where all the values are measured experimentally. . . . .	6
1.2	Leading order Feynman diagrams contributing to di-Higgs production via gluon-gluon fusion, through the Higgs-fermion Yukawa interactions 1.2a and the Higgs boson self-coupling 1.2b. Only Figure 1.2b probes $\lambda_{hhh}$ . . . . .	7
1.3	Leading order Feynman diagrams contributing to Higgs pair production via VBF. . . . .	8
1.4	Total cross sections (y-axis) at the NLO in QCD for the six largest di-Higgs production channels at p-p colliders at different energy (x-axis). Gluon-gluon fusion, VBF, top-pair, $W^\pm, Z$ and single-top associated di-Higgs productions are shown. The thickness of the lines corresponds to the uncertainties added linearly. $H$ refers to the SM Higgs. . . . .	8
1.5	BSM Higgs boson pair production: non-resonant production proceeds through changes in the SM Higgs couplings in 1.5a and 1.5b, resonant production proceeds through 1.5c an intermediate resonance, X. $H$ and $b$ both refers to the SM Higgs. . . . .	10
1.6	Total cross sections (y-axis) at the LO and NLO in QCD for di-Higgs production channels, at the $\sqrt{s} = 14$ TeV LHC as a function of the self-interaction coupling $\lambda$ (x-axis). The dashed (solid) lines and light- (dark-) color bands correspond to the LO (NLO) results and to the scale and PDF uncertainties added linearly. The SM values of the cross sections are obtained at $\frac{\lambda}{\lambda_{SM}} = 1$ , indicated by the red vertical line. $H$ refers to the SM Higgs. . . . .	10
1.7	Parton luminosity ratios as a function of resonance mass $M_X$ for $13/8$ TeV <sup>2</sup> . 1, 2, and 3 TeV $M_X$ ratios are indicated by red lines. For a 2 TeV $X$ , the luminosity ratio is almost 10. . . . .	12
1.8	Summary of di-Higgs final states and their ratios. Top left, $b\bar{b}b\bar{b}$ , has the largest branching ratio. . . . .	13

1.9	The observed and expected 95% CL upper limits of $\sigma(gg \rightarrow H) \times BR(H \rightarrow bb)$ at $\sqrt{s} = 8$ TeV as functions of the heavy Higgs boson mass $m_H$ , combining resonant searches in Higgs boson pair to $b\bar{b}\tau^+\tau^-$ , $W^+W^-\gamma\gamma$ , $b\bar{b}\gamma\gamma$ , and $b\bar{b}b\bar{b}$ final states. The expected limits from individual searches are also shown. The green and yellow bands represent $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty ranges of the expected combined limits. The improvement above $m_H = 500$ GeV reflects the sensitivity of the $b\bar{b}b\bar{b}$ analysis. The results beyond 1 TeV are only from the $b\bar{b}b\bar{b}$ final state alone. . . . .	14
1.10	Cartoon for $X \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ , with resolved event topology (left) and boosted event topology (right). . . . .	15
2.1	A schematic view of the LHC ring <sup>3</sup> . LINAC <sub>2</sub> , Booster, PS, SPS, and LHC accelerate the protons in order. Four main experiments are located at interaction points along the ring. ATLAS and CMS are general purpose experiments, while ALICE focuses on heavy ion collisions and LHC <sub>b</sub> is dedicated to $B$ physics. . . . .	18
2.2	The luminosity-weighted distribution of the mean number of interactions per crossing for the 2015 and 2016 $p\bar{p}$ collision data at 13 TeV centre-of-mass energy. <sup>4</sup> . . . . .	20
2.3	A detailed computer-generated image of the ATLAS detector and its systems. . . . .	21
2.4	Geometry of IBL, PIXEL and SCT detectors in Run2. . . . .	23
2.5	The overall layout of the ATLAS Muon Spectrometer. . . . .	24
2.6	Cumulative luminosity vs. time delivered to (green) and recorded by ATLAS (yellow) during stable beams for $p\bar{p}$ collisions at 13 TeV centre-of-mass energy. . . . .	25
3.1	Illustration of particle interactions in ATLAS. . . . .	27
3.2	Uncalibrated (dashed line) and calibrated (solid line) reconstructed jet mass distribution 3.2a, and the jet mass resolution vs jet $p_T$ 3.2b for calorimeter-based jet mass, $m_{calo}$ (red), track-assisted jet mass $m_{TA}$ (black) and the invariant mass of four-vector sum of tracks associated to the large-radius calorimeter jet $m_{track}$ (blue) for W/Z-jets <sup>5</sup> . . . . .	30
3.3	Secondary vertex reconstruction rate and $MV_{2c10}$ output for b-jets (solid blue), c-jets (dashed green) and light-jets (dotted red) evaluated with simulated $t\bar{t}$ events. . . . .	32
3.4	$b$ -jet efficiency for the fixed cut working point with a $b$ -jet efficiency of 77% as a function of the jet $p_T$ for the comparison between the $MV_{2c10}$ $b$ -tagging algorithm employed for the 2016 analyses (2016 config) and the previous version of the tagger, $MV_{2c20}$ (2015 config), which has 15% $c$ -fraction in the training . . . . .	34

3.5	Light-flavour jet and $c$ -jet rejection as a function of jet $p_T$ for the previous (2015 config) $MV2c20$ and the current $MV2c10$ configuration (2016 config). A fixed cut at 77% $b$ -jet efficiency operating point is used <sup>6</sup> . . . . .	35
5.1	2015 ATLAS boosted di-Higgs event candidate's event display. The data is taken in run 282712, lumiblock 503 and has event number 595661746. ID is in grey, ECAL is in green, HCAL is in red and MS is in blue. Only tracks with $p_T > 25$ GeV are shown. Jets are gray cones, and ID tracks are colored lines in the ID. There are 2 $b$ tagged tracks within each of the large- $R$ jets. The one large- $R$ jet has 119 GeV mass and 543 GeV $p_T$ , and the other one has 127 GeV mass and 413 GeV $p_T$ . The $\Delta R$ between the two large- $R$ jets is 3.54 and the invariant mass of the two large- $R$ jets is 1336 GeV. . . . .	43
5.2	Different trigger efficiencies as a function of the signal resonance mass with respect to all events with no selection (left) and with respect to events passing the two large- $R$ jets $p_T > 400$ GeV and leading/subleading jet $p_T > 250$ GeV (right). For 1.4 TeV signal, the trigger efficiency is about 98%. . . . .	44
5.3	Large- $R$ jet trigger efficiencies, defined as the fraction of events fired trigger with a given highest large- $R$ jet $p_T$ , measured in 2015 Data (HLT_j360_a10_lcw, left) and 2016 Data (HLT_j420_a10_lcw, right) and MC. . . . .	45
5.4	Percentages of truth Higgs to large- $R$ jet $\Delta R < 1.0$ matching as a function of $G_{KK}^*$ mass. Both Higgs almost never match to the same large- $R$ jet. . . . .	46
5.5	Normalized $\Delta R$ between the truth Higgs (leading on left, subleading on right) and the truth children $b$ -quarks for $G_{KK}^*$ MCs. Lines are drawn at $\Delta R = 0.4$ ( $R$ of small- $R$ jets) and $\Delta R = 1.0$ ( $R$ of large- $R$ jets). . . . .	47
5.6	Percentage of $\Delta R < 0.2$ matching truth $b$ 's to track jets (leading Higgs on the left, sub-leading Higgs on the right) for different $G_{KK}^*$ mass. The cases listed in the legend are orthogonal to each other. The cases not listed on the legend (including when a truth $b$ is not contained in the large- $R$ jet) happen in total at most 1.6% of the time for a given $G_{KK}^*$ mass. . . . .	48
5.7	Kinematics of 1 TeV $G_{KK}^*$ before and after muon-in-jet corrections. The reconstructed Higgs masses (top row, left for leading large- $R$ jet, right for subleading large- $R$ jet) are closer to 125 GeV after the correction, which improves the signal efficiency for the signal region selection by $\sim 10\%$ (bottom row, left for $m_{JJ}$ , right for event distribution differences on the leading-subleading large- $R$ jet mass plane.). . . . .	49

5.8	After large- $R$ jet requirements, normalized $\Delta\eta_{JJ}$ distribution in 1.5 TeV $G_{KK}^*$ and data, where the data consists of mostly multijet events ( $> 90\%$ ). The background multijet event is flatter in $\Delta\eta_{JJ}$ distribution. . . . .	50
5.9	Event display of the same event using 5.9a resolved and 5.9b boosted topologies. The resolved reconstruction gives a $m_{4J}$ of 873 GeV, and the boosted reconstruction gives $m_{2J}$ of 852 GeV. . . . .	51
5.10	For RSG $c = 1.0$ samples, number of events as a function of leading Higgs candidate mass and subleading Higgs candidate mass, for 1.2 TeV (left) signal and 2 TeV (right) signal samples. The red dotted line in the center correspond to the signal region, passing $X_{hh} < 1.6$ . . . . .	53
5.11	Signal fraction in different $b$ -tag categories (left) and detailed fraction in different number of track jet and $b$ -tag categories (right) as a function of signal resonance mass hypothesis for selection cuts. The efficiencies are relative to the total number of events passing the 2D mass cut. . . . .	56
5.12	Absolute (left) and relative (right) signal efficiency as a function of RSG $c=1.0$ signal resonance mass hypothesis for selection cuts. The relative efficiency is defined from the previous cut, where the order of cuts is given by the legend. PassTrig means the event passes the trigger selection; PassDiJetPt means the event passes the leading and sub-leading jet $p_T$ cuts; PassDiJetEta means the event passes the leading and sub-leading jet $\eta$ cuts; Pass-DeltaH means the events passes the $ \Delta\eta  < 1.7$ cut; PassBJetSkim means the event contains at least two $b$ -tagged track jets, inclusive of 2 $b$ , 2 $bs$ , 3 $b$ and 4 $b$ configurations; PassSignal means the event passes the signal region cut $X_{hh} < 1.6$ . . . . .	57
6.1	Values of the $X_{hh}$ and $R_{hh}$ variables, which are shapes in the two-dimensional plane of the large- $R$ jet masses used to defined signal, control, and sideband regions. For both variables, a smaller value indicates the jets are closer to the Higgs mass. . . . .	63
6.2	Detailed signal efficiency in different signal/control/sideband regions as in 2 $bs$ (top left), 3 $b$ (top right), 4 $b$ (bottom left) and inclusive $b$ -tagged regions, which include 2 $b$ , 1 $b$ and 0 $b$ as well, (bottom right) as a function of signal resonance mass hypothesis for selection cuts. The efficiencies are relative to the total number of events in the preselection. . . . .	64

# Listing of tables

2.1	LHC nominal and operational parameters . . . . .	21
5.1	Physics objects and their technical names in the boosted analysis. . . . .	45
5.2	The selection efficiency for $G_{KK}^* \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events ( $c = 1.0$ ) at each stage of the event selection. Uncertainties are the MC stat uncertainty only. . . . .	57
5.3	The selection efficiency for $G_{KK}^* \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events ( $c = 2.0$ ) at each stage of the event selection. Uncertainties are the MC stat uncertainty only. . . . .	58
5.4	The selection efficiency for $H \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events at each stage of the event selection. . . . .	58

VERITAS SHALL MAKE YOU FREE.

# Acknowledgments

THANKS TO EVERYONE WORKING AT CERN. CERN is a unique and special place. I love how the streets are named by physicists. Without the support from the IT department, I could not log into lxplus, check the twikis and finish my work. CERN user's office also made traveling to Europe a much nicer and easier experience for me.

THANKS TO EVERYONE WORKING ON THE LARGE HADRON COLLIDER. Without a fully functional accelerator, there would have been no data for me to study. The LHC performed outstandingly since 2015, and all the valuable data in this thesis is produced from it.

THANKS TO EVERYONE THE ATLAS COLLABORATION, who has supported this remarkable program and has contributed to every bit of the result in my thesis. I am standing on the ATLAS(member)'s shoulders—eight floors high and don't shrug. Without the excellent work on detector design, commissioning, operational works, reconstruction, data processing, performance studies and recommendations, software support, computing support, and analysis discussions and guidance, I could not have completed this thesis. I sincerely thank all ATLAS members for their work.

*It was the best of times, it was the worst of times, it was  
the age of wisdom, it was the age of foolishness, it was  
the epoch of belief, it was the epoch of incredulity, it  
was the season of Light, it was the season of Darkness,  
it was the spring of hope, it was the winter of despair,  
we had everything before us, we had nothing before us,  
we were all going direct to Heaven, we were all going  
direct the other way—in short, the period was so far like  
the present period, that some of its noisiest authorities  
insisted on its being received, for good or for evil, in the  
superlative degree of comparison only.*

Charles Dickens

# 0

## Introduction

In 2012, the Higgs boson was discovered by the ATLAS and CMS experiment at the LHC. The particle physics community faces a period just like at the beginning of *A Tale of Two Cities*.

After Run 1 of the LHC, with the existence of the Higgs now firmly established, the focus shifted to searches for physics beyond the Standard Model.

In particular, searches for high mass resonances benefit from the LHC’s increase to  $\sqrt{s} = 13$  TeV in Run 2. The cross section for a generic gluon-initiated resonance with a mass of 2 TeV increases tenfold in Run 2, making searches for high mass resonances a high priority. The newly discovered Higgs can be used as a tool in these searches. After the discovery, the Higgs boson provides a large swath of unmeasured phase space where new physics could be discovered. Higgs pair production in the Standard Model has a low cross section that requires large datasets (on the order of the

LHC's lifetime) for full measurement. However, new physics can modify this cross section, especially through new resonances which decay to two Higgs bosons. Such high mass resonances also produce difficult to recognize final state topologies due to the merging of decay products from high momentum Higgs bosons. A search for Higgs pair production in the  $HH \rightarrow b\bar{b}b\bar{b}$  final state was performed with  $3.2\text{fb}^{-1}$  collected with ATLAS at  $\sqrt{s} = 13$  TeV in 2015. The results are presented in this dissertation with a focus on a dedicated signal region for boosted final states. This signal region uses new techniques for recognizing jet substructure and  $b$ -tagging to the improve signal acceptance of high mass resonances.

The discovery of the Standard Model (SM) Higgs boson ( $h$ )<sup>2</sup> at the Large Hadron Collider (LHC) motivates searches for new physics using the Higgs boson as a probe. In particular, many models predict cross sections for Higgs boson pair production that are significantly greater than the SM prediction. Resonant Higgs boson pair production is predicted by models such as the bulk Randall–Sundrum model<sup>7,8</sup>, which features spin-2 Kaluza–Klein gravitons,  $G_{KK}^*$ , that subsequently decay to a pair of Higgs bosons. Extensions of the Higgs sector, such as two-Higgs-doublet models<sup>9,10</sup>, propose the existence of a heavy spin-0 scalar that can decay into  $hpairs$ . Enhanced non-resonant Higgs boson pair production is predicted by other models, for example those featuring light coloured scalars<sup>11</sup> or direct  $t\bar{t}hh$  vertices<sup>12,13</sup>.

Previous searches for Higgs boson pair production have all yielded null results. In the  $b\bar{b}b\bar{b}$  channel, ATLAS searched for both non-resonant and resonant production in the mass range of 400–3000 GeV using  $3.2\text{ fb}^{-1}$  of 13 TeV data<sup>14</sup> collected during 2015. CMS searched for the production of resonances with masses of 750–3000 GeV<sup>2</sup> using 13 TeV data and with masses 270–1100 GeV with 8 TeV data<sup>2</sup>. Using 8 TeV data, ATLAS has examined the  $b\bar{b}b\bar{b}$ <sup>15</sup>,  $b\bar{b}\gamma\gamma$ <sup>16</sup>,  $b\bar{b}\tau^+\tau^-$  and  $W^+W^-\gamma\gamma$  channels, all of which were combined in Ref.<sup>2</sup>. CMS has performed searches using 13 TeV data for the  $b\bar{b}\tau^+\tau^-$ <sup>2</sup> and  $b\bar{b}\ell\nu\ell\nu$ <sup>2</sup> final states, and used 8 TeV data to search for  $b\bar{b}\gamma\gamma$ <sup>2</sup> in addition to a search in multilepton and multilepton+photons final states<sup>2</sup>.

The analyses presented in this paper exploit the dominant  $b \rightarrow b\bar{b}$  decay mode to search for Higgs boson pair production in both resonant and non-resonant production. Two analyses are presented, which are complementary in their acceptance, each employing a unique technique to reconstruct the Higgs boson. The “resolved” analysis is used for  $hh$  systems in which the Higgs bosons have Lorentz boosts low enough that four  $b$ -jets can be reconstructed. The “boosted” analysis is used for those  $hh$  systems in which the Higgs bosons have higher Lorentz boosts, which prevents the Higgs boson decay products from being resolved in the detector as separate  $b$ -jets. Instead, each Higgs boson candidate consists of a single large-radius jet, and  $b$ -decays are identified using smaller-radius jets built from charged-particle tracks.

Both analyses were re-optimized with respect to the former ATLAS publication <sup>14</sup>; an improved algorithm to pair  $b$ -jets to Higgs boson candidates is used in the resolved analysis, and in the boosted analysis an additional signal-enriched sample is utilized. The dataset comprises the 2015 and 2016 data, corresponding to  $27.5 \text{ fb}^{-1}$  for the resolved analysis and  $36.1 \text{ fb}^{-1}$  for the boosted analysis, with the difference due to the trigger selections used. The results are obtained using the resolved analysis for a resonance mass between 260 and 1400 GeV, and the boosted analysis between 800 GeV and 3000 GeV. The main background is multijet production, which is estimated from data; the sub-leading background is  $t\bar{t}$ , which is estimated using both data and simulations. The two analyses employ orthogonal selections, and a statistical combination is performed in the mass range where they overlap. The final discriminants are the four-jet and dijet mass distributions in the resolved and boosted analyses, respectively. Searches are performed for the following benchmark signals: a spin-2 graviton decaying into Higgs bosons, a scalar resonance decaying into a Higgs boson pair, and SM non-resonant Higgs boson pair production.

This dissertation begins by discussing the status of di-Higgs. Chapter 1 gives an overview of double Higgs production in the Standard Model and beyond. Chapter 2 and 3 present details regarding

the Large Hadron Collider and the ATLAS experiment. Chapter 4 provides an overview of object reconstruction in ATLAS, with a focus on Muon Segment Seeding. A brief interlude in Chapter 5 on the ATLAS Muon Data Quality, as this has been a focus of my graduate work.

The rest of the dissertation presents a search for Higgs pair production in the  $HH \rightarrow b\bar{b}b\bar{b}$  channel. Chapter 6 presents an overview of physics object selection, where the Higgs pairs are the result of the decay of a heavy resonance. Chapter 7 discusses the background estimation techninics in detail, followed by Chapter 8, Systematics. Chapter 9 presents the results, and Chapter 10 shows the limits between the boosted regime and the resolved regime, which is sensitive to lower mass resonances and non-resonant Higgs pair production. Finally, the work is summarized a conclusion and brief outlook of future Higgs physics with ATLAS.

*Knowledge knows no bounds.*

Creator

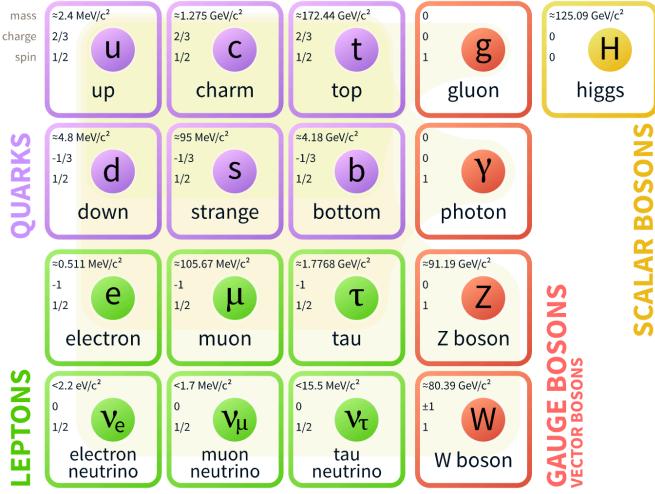
# 1

## Theory and Motivation

### 1.1 THE STANDARD MODEL AND THE HIGGS BOSON

The Standard Model(SM)<sup>1,17,18,19</sup> is a quantum field theory describing the interactions of fundamental particles through the strong, weak and electromagnetic forces. The elementary particles are shown in Figure 1.1. So far, the SM predictions of particle interactions agree extremely well with experimental observations.

In the SM, the Higgs mechanism introduces a complex scalar Higgs field,  $\varphi$ , with nonzero vacuum expectation values. The scalar Higgs potential is  $V(\varphi) = -v^2\lambda_v\varphi^\dagger\varphi + \lambda_v(\varphi^\dagger\varphi)^2$ . Through spontaneous symmetry breaking,  $W^\pm$  and  $Z$  bosons acquire their masses. This process also predicts an extra scalar, the Higgs boson. The SM Lagrangian containing Higgs couplings,  $\mathcal{L}_{\text{Higgs}}$ , is shown



**Figure 1.1:** Fermions and bosons of the Standard Model and their properties<sup>1</sup>, where all the values are measured experimentally.

in Eq 1.1.

$$\mathcal{L}_{\text{Higgs}} = -\lambda_{h\bar{f}f} h\bar{f}f + \delta_V V_\mu V^\mu (\lambda_{hVV} b + \lambda_{hVV} b^2) + \lambda_{hh} b^2 + \lambda_{hhh} b^3 + \lambda_{hhhh} b^4 \quad (1.1)$$

where

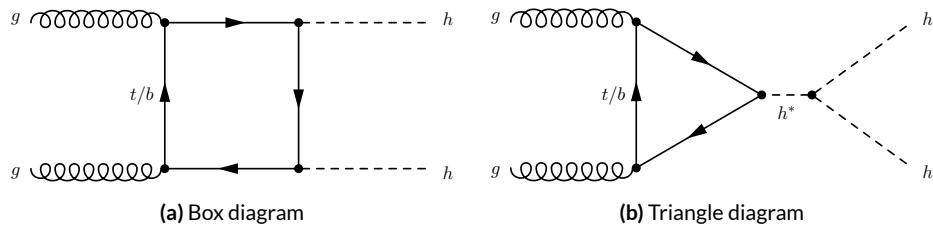
- $v \sim 246 \text{ GeV}$ , is the non-zero expectation value of the Higgs field;
- $\lambda_v \sim -0.13$ , the coefficient for the quartic potential term, is constrained from Higgs mass;
- $m_h = \sqrt{2\lambda_v v} = 125.09 \pm 0.24 \text{ GeV}$ , is the Higgs mass; this was discovered in 2012<sup>20,21</sup>;
- $V = W^\pm$  or  $Z$ ,  $\delta_W = 1$ ,  $\delta_Z = \frac{1}{2}$ ;
- $\lambda_{h\bar{f}f} = \frac{m_f}{v}$ , is the Higgs to fermion coupling;  $m_f$  is the mass of the fermion;
- $\lambda_{hVV} = \frac{2m_V^2}{v}$ , is the Higgs to boson coupling;  $m_V$  is the mass of the boson;
- $\lambda_{hh} = \frac{m_h^2}{v^2}$ , is the Higgs-Higgs to boson-boson coupling;
- $\lambda_{hh} = \frac{m_h^2}{2}$ , is the Higgs mass term;

- $\lambda_{hhh} = \frac{m_h^2}{2\nu} = \lambda_{\nu}$ , or  $\lambda_{hhh}$ , is the Higgs self-coupling;
- $\lambda_{hhhh} = \frac{m_h^2}{8\nu^2}$ , is the Higgs quartic-coupling.

What's particularly interesting and has not been measured experimentally is  $\lambda_{hhh}$ . The SM predicts  $\lambda_{hhh} = \frac{m_h^2}{2\nu}$ , which is referred as  $\lambda_{SM}$  in this thesis. Measuring  $\lambda_{hhh}$  directly probes the Higgs potential. Also, the  $\lambda_{hhh} h^3$  term shows one way for double Higgs production within the SM. Double Higgs production is also known as *di-Higgs* or *Higgs pair production*.

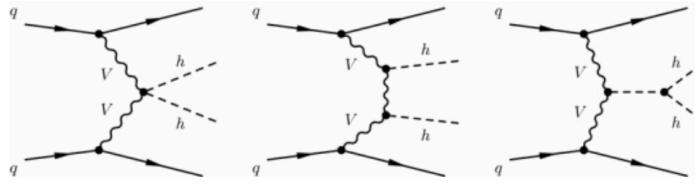
## 1.2 STANDARD MODEL DI-HIGGS PRODUCTION

There are two main production diagrams of di-Higgs at the LHC, both shown in Figure 1.2. In the gluon-gluon fusion process, di-Higgs are produced through a box or a triangle loop. Only the triangle loop 1.2b probes the  $\lambda_{hhh}$ . In the triangle diagram, the middle Higgs boson acts as a propagator (off-shell), and the two Higgs bosons in the final state are on-shell. The diagram with an on-shell middle Higgs and two off-shell Higgs bosons in the final state is strongly disfavored<sup>1</sup>. The box and triangle diagrams interfere destructively, which makes the overall production rate smaller than what would be expected in the absence of a  $\lambda_{hhh}$  term.

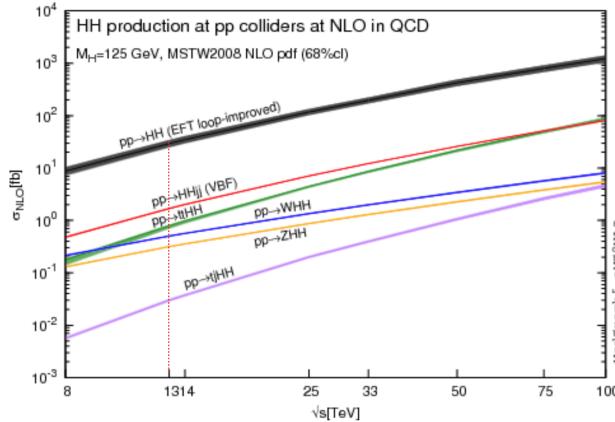


**Figure 1.2:** Leading order Feynman diagrams contributing to di-Higgs production via gluon-gluon fusion, through the Higgs-fermion Yukawa interactions 1.2a and the Higgs boson self-coupling 1.2b. Only Figure 1.2b probes  $\lambda_{hhh}$ .

Many other different production modes of di-Higgs exist, but gluon-gluon fusion is the dominant one. Figure 1.4<sup>22</sup> compares the cross sections of gluon-gluon fusion, Vector Boson Fusion (VBF) 1.3, top-pair,  $W^\pm$ ,  $Z$  and single-top associated di-Higgs production.



**Figure 1.3:** Leading order Feynman diagrams contributing to Higgs pair production via VBF.



**Figure 1.4:** Total cross sections (y-axis) at the NLO in QCD for the six largest di-Higgs production channels at p-p colliders at different energy (x-axis). Gluon-gluon fusion, VBF, top-pair,  $W^\pm$ ,  $Z$  and single-top associated di-Higgs productions are shown. The thickness of the lines corresponds to the uncertainties added linearly.  $H$  refers to the SM Higgs.

For p-p collisions at  $\sqrt{s} = 13$  TeV, the total cross section for SM di-Higgs production<sup>23</sup> is evaluated at next-to-next-to-leading order (NNLO) with the summation of logarithms at next-to-next-leading-logarithm (NNLL) accuracy and including finite top-quark mass effects at NLO. The uncertainties are from energy scale and parton distributions functions (PDF):

- Gluon gluon fusion:  $\sigma_{gg \rightarrow HH} = 33.49^{+4.3\%}_{-6.0\%} \pm 2.1\% \text{ fb.}$
- Vector boson fusion:  $\sigma_{VBF \rightarrow HH} = 1.62^{+2.3\%}_{-2.7\%} \pm 2.3\% \text{ fb.}$
- Gluon gluon fusion to Triple-Higgs:  $\sigma_{gg \rightarrow HHH} = 0.06332^{+16.1\%}_{-14.1\%} \pm 3.4\% \text{ fb.}$

This means inside 2015 and 2016 ATLAS 36  $\text{fb}^{-1}$  data, the SM expectation is around one thousand di-Higgs events and only two triple Higgs events.

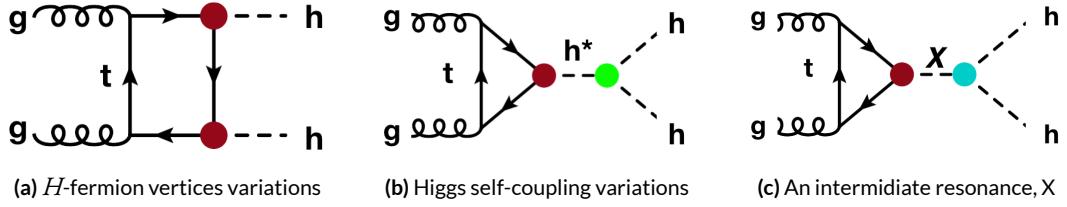
### 1.3 BEYOND THE STANDARD MODEL PHYSICS DI-HIGGS PRODUCTION

The SM works extremely well, yet the Higgs boson mass at 125 GeV requires extreme fine-tuning for radiative corrections. The presence of new physics at the TeV scale would help solve the naturalness problem.

BSM physics could significantly enhance the production of di-Higgs at the LHC. This is separated into two categories: non-resonant and resonant production. The non-resonant production generally refers to modifications of the Higgs couplings, either the Higgs self-coupling or the Higgs-top couplings. Resonant production refers to the case when a particle with mass greater than twice the Higgs mass decays directly into two Higgs bosons. The non-resonant and resonant production also differ from the invariant mass distribution of the di-Higgs at the truth level. In the non-resonant case, the invariant mass distribution has no clear peak, whereas in the resonant case, the invariant mass distribution usually forms a peak with model dependent width.

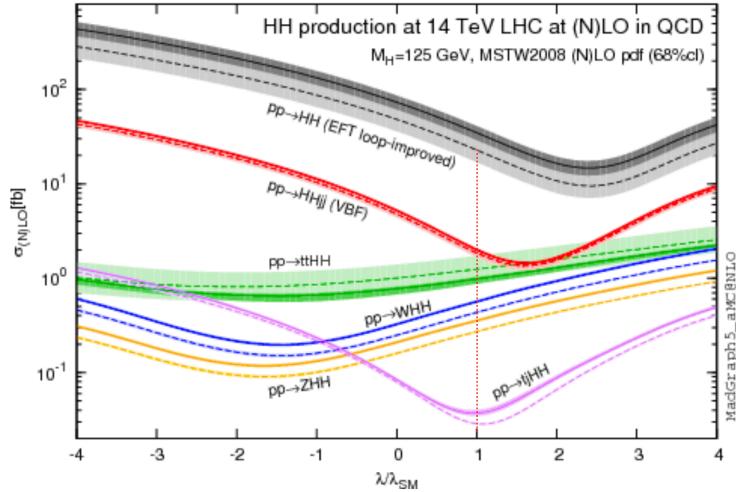
#### 1.3.1 BSM NON-RESONANT DI-HIGGS

Enhanced non-resonant Higgs boson pair production is predicted by many models. BSM models featuring direct  $t\bar{t}hh$  vertices<sup>12,13</sup> or new light colored scalars<sup>14</sup> could augment vertex strength, shown as the red dots in Figure 1.5. A direct modification of the Higgs self-coupling coefficient in Eq 1.1, from  $\lambda_{sm} hhh$  to  $\lambda hhh$ , is also possible. This is shown as the green dot in Figure 1.5b.



**Figure 1.5:** BSM Higgs boson pair production: non-resonant production proceeds through changes in the SM Higgs couplings in 1.5a and 1.5b, resonant production proceeds through 1.5c an intermediate resonance,  $X$ .  $H$  and  $h$  both refers to the SM Higgs.

The non-resonant di-Higgs enhancement is usually described by the di-Higgs cross section ratio between a BSM  $\lambda$  coupling scenario and the  $\lambda_{\text{SM}}$  coupling scenario.  $\frac{\lambda}{\lambda_{\text{SM}}}$  indicates the ratio between the BSM model  $\lambda$  and  $\lambda_{\text{SM}}$ . From the SM electroweak measurements, the self coupling term is constrained to  $-14 \leq \frac{\lambda}{\lambda_{\text{SM}}} \leq 17.4$ <sup>24</sup>. Variations of  $\lambda$  have a non-trivial effect on di-Higgs production cross section, shown in Figure 1.6<sup>22</sup>. In the regime of relatively high trilinear coupling, the observation will be an excess of di-Higgs events with respect to the expected background. A simple limit can be set in this case.



**Figure 1.6:** Total cross sections (y-axis) at the LO and NLO in QCD for di-Higgs production channels, at the  $\sqrt{s} = 14$  TeV LHC as a function of the self-interaction coupling  $\lambda$  (x-axis). The dashed (solid) lines and light- (dark-) color bands correspond to the LO (NLO) results and to the scale and PDF uncertainties added linearly. The SM values of the cross sections are obtained at  $\frac{\lambda}{\lambda_{\text{SM}}} = 1$ , indicated by the red vertical line.  $H$  refers to the SM Higgs.

### I.3.2 BSM RESONANT DI-HIGGS

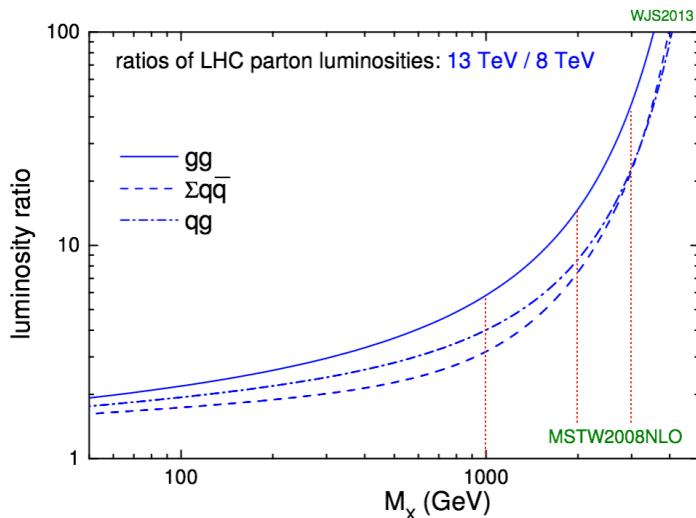
Resonant Higgs boson pair production is also predicted by many models. Extensions of the Higgs sector, such as two-Higgs-doublet models(2HDM)<sup>9,10</sup>, propose the existence of a heavy spin-0 scalar  $H$  that can decay into di-Higgs. The bulk Randall-Sundrum model<sup>7,8</sup>, which features spin-2 Kaluza-Klein gravitons,  $G_{KK}^*$ , could also subsequently decay to pairs of Higgs bosons. These proposed heavy particles, heavy CP-even scalar  $H$  and  $G_{KK}^*$ , are represented as X in Figure I.5c.

The 2HDM is a simple extension of the SM which has large resonance effects<sup>23</sup>. The 2HDM consists of 5 physical Higgs bosons:  $h$  (light scalar Higgs),  $H$  (heavy scalar Higgs),  $A$  (heavy pseudoscalar Higgs), and  $H^\pm$  (two charged Higgs). The 2HDM can introduce tree level flavor changing neutral currents. To avoid this, models impose discrete symmetries in which the charged fermions only couple to one of the Higgs doublets. One version is type II 2HDM, in which all positively charged quarks couple to one doublet and the negatively charged quarks and leptons couple to the other. The type II model is the Minimal Supersymmetric Standard Model(MSSM)'s Higgs sector.

Resonant di-Higgs production in 2HDM models can proceed through decays of the heavy CP-even Higgs  $H \rightarrow hh$ . The branching ratio for  $H \rightarrow hh$  depends on the model type as well as the values of  $\tan\beta$  and  $\cos(\beta - \alpha)$ .  $\tan\beta = \frac{v_{\text{doublet}}}{v_{\text{SM}}}$  is the ratio of the vacuum expectation values of the two Higgs doublets.  $\alpha$  is the mixing angle between the heavy  $H$  and light  $h$  fields. The limit where  $\cos(\beta - \alpha) = 0$  is called the alignment limit, and in this limit the light Higgs  $h$  has the same couplings as a SM Higgs. Near the alignment limit there is some unprobed phase space depending on the exact models and values of  $\tan\beta$ , and they are particularly interesting to be searched for at the LHC.

The Randall-Sundrum model proposes a five-dimensional warped spacetime that contains two manifolds: one where the force of gravity is very strong and a second manifold at the TeV scale

corresponding to the known SM sector. The experimental consequence of this theory is a series of widely mass-spaced Kaluza-Klein graviton resonances,  $G_{\text{KK}}^*$ . In cases where the fermions are localized to the SM brane, production of gravitons from fermion pairs is suppressed and the primary mode of production of  $G_{\text{KK}}^*$  is gluon fusion. These gravitons further decay to di-Higgs, with branching fraction ranging from 6.43% for gravitons with a mass of 500 GeV to 7.66% at 3 TeV. Randall-Sundrum models have two free parameters - the mass of the graviton and  $c = k/\bar{M}_{\text{pl}}$ , where  $\bar{M}_{\text{pl}}$  is the reduced Planck mass and  $k$  is the curvature scale of the extra dimension. The width of the graviton increases with both mass and  $c$ .



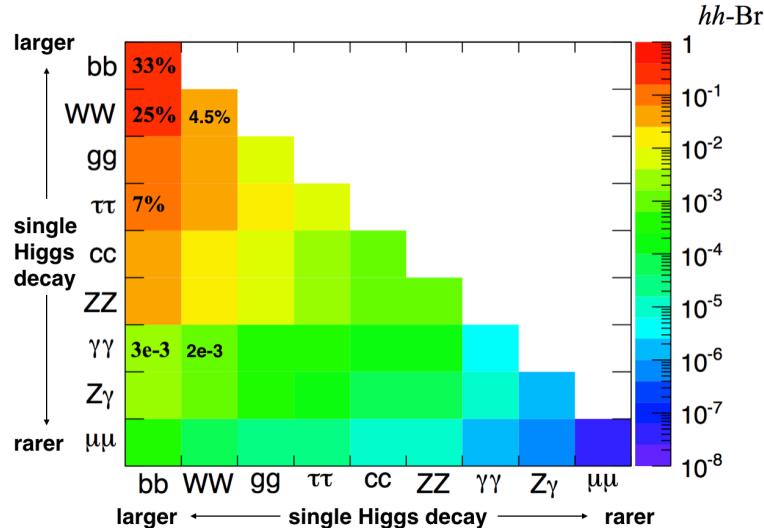
**Figure 1.7:** Parton luminosity ratios as a function of resonance mass  $M_X$  for 13/8 TeV<sup>2</sup>. 1, 2, and 3 TeV  $M_X$  ratios are indicated by red lines. For a 2 TeV  $X$ , the luminosity ratio is almost 10.

In model dependent searches, based on fixed assumptions of the resonance particles' branching ratio, other search channels like resonance  $VV$  or  $t\bar{t}$  are more sensitive compared to di-Higgs<sup>25</sup>. In order to constrain more BSM physics phase space, di-Higgs search results need to be interpreted in different baseline models, covering both narrow and wide resonances.

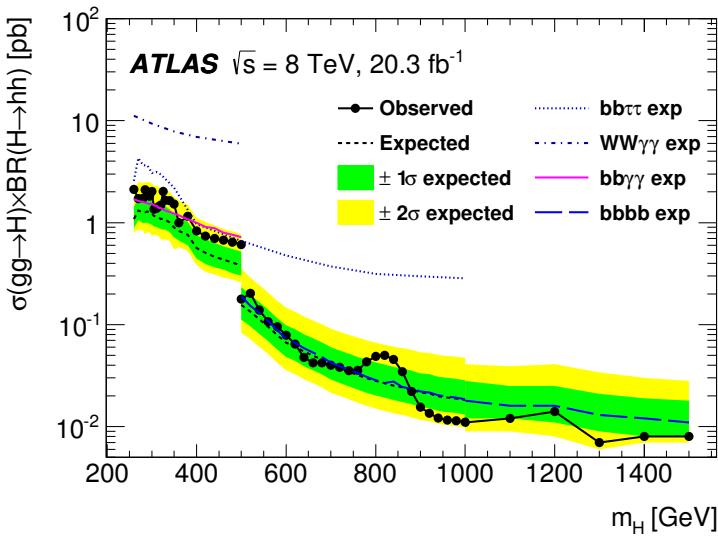
Generally, it is easy theoretically for new heavy resonance particles to interact with the SM through the Higgs as a mediator, resulting in resonance di-Higgs production. With the increased center of mass collision energy from 8 TeV to 13 TeV in LHC Run 2, the production cross section for particles with TeV mass grows, as shown in Figure 1.7. In the case of a 2 TeV  $M_X$ , the cross section gain is almost a factor of 10. Therefore, it was particularly important to focus on resonant searches above TeV region.

#### 1.4 DI-HIGGS DECAY AND LHC PREVIOUS SEARCH RESULTS

Double Higgs decays into other elementary particles in  $1.56^{-22}$ s. Therefore, searching for di-Higgs requires reconstructing the Higgs boson from the decay products. Di-Higgs decay is a combination of single Higgs decays. The coupling terms to fermions and bosons are shown in Eq 1.1. The branching ratio of the di-Higgs final state is shown in Figure 1.8.



**Figure 1.8:** Summary of di-Higgs final states and their ratios. Top left,  $b\bar{b}b\bar{b}$ , has the largest branching ratio.



**Figure 1.9:** The observed and expected 95% CL upper limits of  $\sigma(gg \rightarrow H) \times BR(H \rightarrow hh)$  at  $\sqrt{s} = 8$  TeV as functions of the heavy Higgs boson mass  $m_H$ , combining resonant searches in Higgs boson pair to  $b\bar{b}\tau^+\tau^-$ ,  $W^+W^-\gamma\gamma$ ,  $b\bar{b}\gamma\gamma$ , and  $b\bar{b}b\bar{b}$  final states. The expected limits from individual searches are also shown. The green and yellow bands represent  $\pm 1\sigma$  and  $\pm 2\sigma$  uncertainty ranges of the expected combined limits. The improvement above  $m_H = 500$  GeV reflects the sensitivity of the  $b\bar{b}b\bar{b}$  analysis. The results beyond 1 TeV are only from the  $b\bar{b}b\bar{b}$  final state alone.

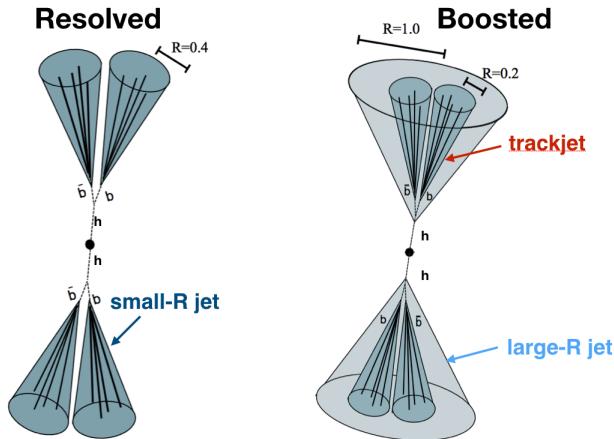
Previous searches for Higgs boson pair production have not found any significant signals. Using 8 TeV data, ATLAS has examined the  $b\bar{b}b\bar{b}$ <sup>15</sup>,  $b\bar{b}\gamma\gamma$ <sup>16</sup>,  $b\bar{b}\tau^+\tau^-$  and  $W^+W^-\gamma\gamma$  channels, all of which were combined<sup>26</sup>. The resonant search combination result is shown in Figure 1.9. The best non-resonant  $\sigma(pp \rightarrow hh)$  cross section limit in Run 1 is the ATLAS combination, at 0.69 pb. This corresponds to  $|\frac{\lambda}{\lambda_{SM}}| < 70$ . Different di-Higgs search challenges and perspectives are summarized below:

- $b\bar{b}b\bar{b}$ : Trigger limits the low mass resonance searches, but for high mass resonances above 500 GeV, the branching ratio of this channel provides a decisive advantage. It is great for non-resonant searches as well.
- $b\bar{b}W^+W^-$ : Despite the second largest branching ratio, large background from  $t\bar{t}$  limits this search sensitivity.
- $b\bar{b}\gamma\gamma$ : This channel benefits from good double photon trigger efficiency, a good photon

reconstruction efficiency and a low SM background. Most sensitive at low mass  $m_X \leq 350$  GeV. At higher masses, the smaller branching ratio and the merging of photons hurt the search sensitivity. Great for non-resonant searches.

- $b\bar{b}\tau^+\tau^-$ : An intermediate choice between  $b\bar{b}b\bar{b}$  and  $b\bar{b}\gamma\gamma$  for resonance searches. Yet this channel contributes to the non-resonant result significantly.
- $W^+W^-\gamma\gamma$ : Suffers from much lower branching ratio and lower reconstruction efficiency of the  $W^+W^-$  compared to  $b\bar{b}$ .
- $W^+W^-\tau\tau, W^+W^-W^+W^-$ ,  $b\bar{b}ZZ$ : There are no search results on these channels yet. But because of the relatively large branching ratio, it is likely that they would be explored in the future.

## 1.5 RESOLVED AND BOOSTED



**Figure 1.10:** Cartoon for  $X \rightarrow bb \rightarrow b\bar{b}b\bar{b}$ , with resolved event topology (left) and boosted event topology (right).

The thesis focuses on searching for a TeV scale resonance  $X \rightarrow bb \rightarrow b\bar{b}b\bar{b}$ . In order to separate the signal from the multi quark production backgrounds, it is important to fully reconstruct

the out-coming  $b$  quarks. When the Higgs bosons have Lorentz low boosts, four small- $R$   $b$ -jets with  $R = 0.4$  can be reconstructed. This final state is called the *resolved* state, shown on the left of Figure 1.10. The resolved strategy is effective for resonance  $M_X$  from 260 GeV up to 1.2 TeV. It is also good for non-resonance di-Higgs searches.

A different reconstruction strategy has to be used for di-Higgs systems with higher Lorentz boosts from heavier resonances. For a Higgs produced with momentum  $p_H$ , the angular separation between the  $b$  and  $\bar{b}$  quarks,  $\Delta R_{bb}$ , scales roughly as  $\frac{2m_H}{p_H}$ . For example, a 1.5 TeV resonance  $G_{KK}^*$  is produced roughly at rest in a  $p\bar{p}$  collision. In the lab frame, the two Higgs bosons each get  $\sim 625$  GeV momentum. The  $\Delta R_{bb}$  is around 0.4. This means that the standard  $R = 0.4$   $b$ -jets cannot resolve the  $b\bar{b}$  system. Therefore a different analysis strategy is required. Instead, each  $b\bar{b}$  system is reconstructed as a single large-radius jet. The large-radius jet contains the decay products of the Higgs boson. The presence of  $b$ -quarks is inferred using smaller-radius track jets built from charged-particle tracks. This final state is called the *boosted* state, shown on the right of Figure 1.10. This strategy works for resonance  $M_X$  from 1 GeV up to 3 TeV.

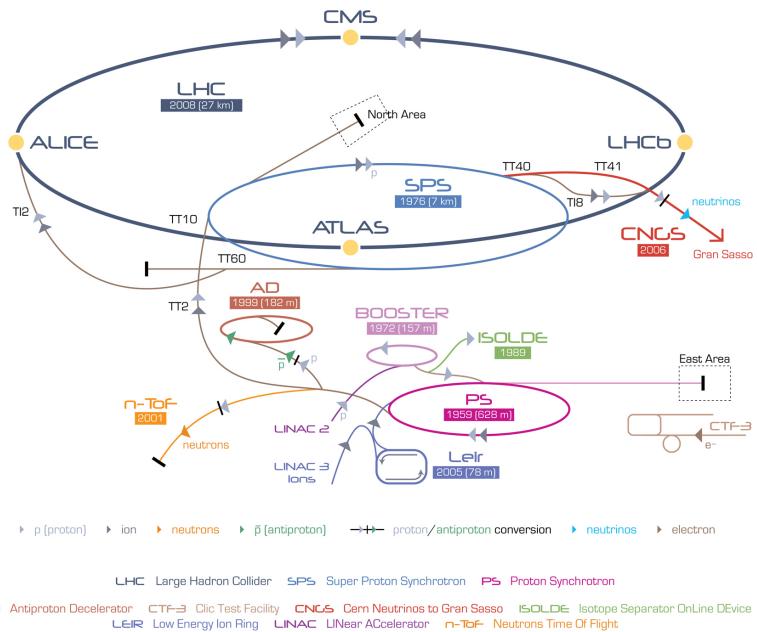
In summary, di-Higgs has a small production rate in the SM, but could be significantly enhanced in BSM scenarios. In particular, a heavy resonance spin-0 or spin 2 particle could decay into Higgs boson pair directly. The search sensitivity for massive resonances increases as the center of mass energy of the collision increases. For resonance signals above 1 TeV decaying into di-Higgs,  $b\bar{b}b\bar{b}$  channel has the best discovery potential in Run 2. In order to fully reconstruct these  $X \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ , boosted techniques has to be used. Therefore, searching for TeV scale resonance production of di-Higgs  $\rightarrow b\bar{b}b\bar{b}$  using the boosted analysis is the goal of thesis.

*Pain teaches lessons no scholar can.*

# 2

## LHC and ATLAS

The Large Hadron Collider (LHC) is a proton-proton ( $p\bar{p}$ ) collider at the European Organization for Nuclear Research (CERN) laboratory in Geneva, Switzerland<sup>27</sup>. ATLAS (A Toroidal LHC ApparatuS), CMS (the Compact Muon Solenoid), ALICE (A Large Ion Collider Experiment), and LHC $b$  (Large Hadron Collider beauty experiment)<sup>28,29,30,31</sup> are the four main experiments. They are located at the Interaction points(IPs) of the accelerator. Figure 2.1 shows a schematic of the LHC ring and its experiments.



**Figure 2.1:** A schematic view of the LHC ring<sup>3</sup>. LINAC2, Booster, PS, SPS, and LHC accelerate the protons in order. Four main experiments are located at interaction points along the ring. ATLAS and CMS are general purpose experiments, while ALICE focuses on heavy ion collisions and LHC $b$  is dedicated to  $B$  physics.

## 2.1 THE LARGE HADRON COLLIDER

Protons accelerated in the LHC are from a red bottle of hydrogen gas. The whole acceleration takes around 25 minutes in multiple steps:

- An electric field strips the electrons from the hydrogen to create protons;
- A linear particle accelerator, Linac 2, accelerates the protons to 50 MeV;
- The Proton Synchrotron Booster (PSB) accelerates the protons to 1.4 GeV;
- The Proton Synchrotron (PS) accelerates the protons to 25 GeV;
- The Super Proton Synchrotron (SPS) accelerates the protons to 450 GeV;
- The 16.7 kilometers LHC accelerates the protons in a series of Radio Frequency cavities to the final TeV energies. The LHC uses 1232 Niobium Titanium magnetic dipole for steering

the protons. The magnets are cooled by superfluid helium to 1.9 Kelvin, and can generate 8.33 Tesla magnetic field.

In proton-proton collisions, the rate of a certain physics process  $R_{\text{phy}} = L\sigma$ , where  $L$  ( $\text{m}^{-2}\text{s}^{-1}$ ) is the instantaneous luminosity, and  $\sigma(\text{m}^2)$  the cross section of physics process (like di-Higgs'  $\sigma$ , 1.2). For a Gaussian beam profile, the instantaneous luminosity is defined in Eq2.1<sup>3</sup>:

$$L = \frac{n_b N_b^2 f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (2.1)$$

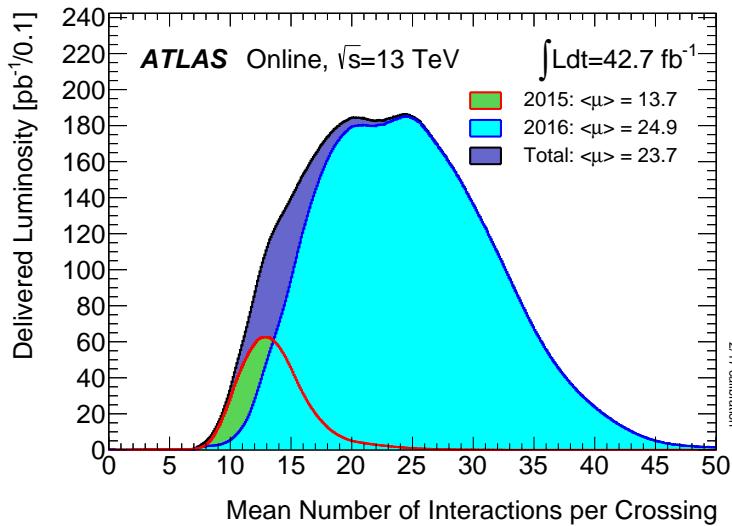
In the above Eq2.1:

- $n_b$  is the number of bunches per beam;  $n_b$  cannot be too large due to potential beam loss damages on the accelerator and detector;
- $N_b$  is the number of protons per bunch;
- $f_{\text{rev}}$  is the proton revolution frequency;
- $\gamma_r$  is the relativistic Lorentz factor for the protons;
- $\epsilon_n$  is the average beam spread length in the transverse plane;
- $\beta^*$  is the beam spread in the longitudinal direction; affected by focusing magnets;
- $F$  is a reduction factor for the angle beams are colliding; smaller crossing angles could cause larger spread in the longitudinal direction.

The instantaneous luminosity can also be written as the ratio of the rate of inelastic collisions to the inelastic cross section  $\sigma_{\text{inel}}$ <sup>32</sup>:

$$L = \frac{R_{\text{inel}}}{\sigma_{\text{inel}}} = \frac{\mu n_b f_{\text{rev}}}{\sigma_{\text{inel}}} \quad (2.2)$$

where,  $\mu$  is the number of interactions per bunch crossing. At each bunch crossing, multiple proton-proton collide, and the collisions without the highest center of mass energy are called “pileup” interactions. The target peak instantaneous luminosity for both the ATLAS and CMS experiments is  $L = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ <sup>27</sup>, which is already exceeded in 2016. This is partly due to the rising number of “pileup” interactions, shown in Figure 2.2. The main parameters of the LHC beam and performance are shown in Table 2.1.



**Figure 2.2:** The luminosity-weighted distribution of the mean number of interactions per crossing for the 2015 and 2016  $p\bar{p}$  collision data at  $13 \text{ TeV}$  centre-of-mass energy.<sup>4</sup>

## 2.2 A TOROIDAL LHC APPARATUS

The ATLAS experiment<sup>35</sup> at the LHC is a general-purpose particle detector with a near  $4\pi$  coverage in solid angle and a forward-backward symmetric cylindrical geometry. The ATLAS detector (Figure 2.3) consists of an inner tracking detector (ID) surrounded by a 2.3 m diameter thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic (EM) and hadronic calorimeters, and a muon spectrometer (MS). Three extra air-core toroid magnets generate the mag-

Parameter [unit]	Nominal design value	2015 Operating value	2016 Operating value
Beam Energy [TeV]	7	6.5	6.5
Peak L [ $10^{34} \text{ cm}^2 \text{ s}^{-1}$ ]	1	0.5	1.25
Bunch spacing [ns]	25	25	25
$f_{\text{rev}}$ [kHz]	11245	11245	11245
$n_b$ [ $10^{11}$ p/bunch]	1.15	1.15	1.12
$N_b$ [bunch]	2808	1825	2220
$\epsilon_n$ [mm mrad]	3.5	3.5	2
$\beta^*$ [cm]	55	20-40	40
$F$	0.84	0.84	0.59
$\langle \mu \rangle$	19	13	41

Table 2.1: LHC nominal<sup>27</sup> and operational parameters in 2015<sup>33</sup> and 2016<sup>34</sup>.

netic field in the MS.

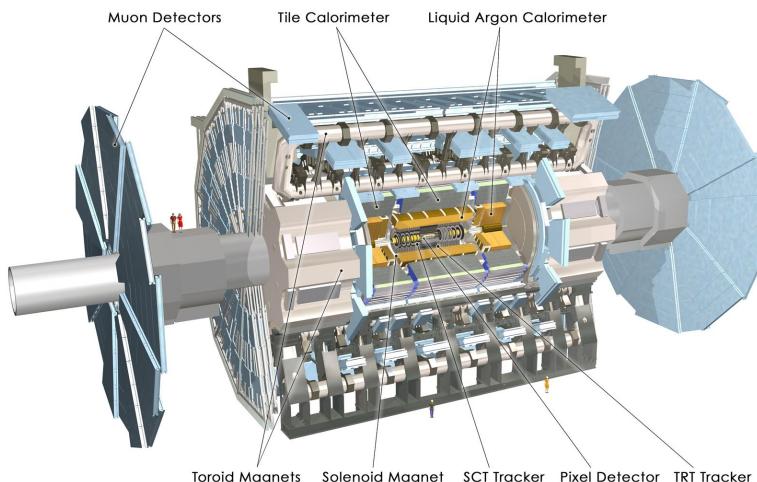


Figure 2.3: A detailed computer-generated image of the ATLAS detector and it's systems.

### 2.2.1 COORDINATE SYSTEM

ATLAS uses a right-handed coordinate system with its origin at the nominal IP in the center of the detector and the  $z$ -axis along the beam pipe. The  $x$ -axis points from the IP to the center of the

LHC ring, the  $y$ -axis points towards the sky, and the  $z$ -axis points straight (like bridges) towards the Geneva airport (A side), back from the Charlie pub in France (C side). Cylindrical coordinates  $(r, \varphi)$  are used in the transverse plane,  $\varphi$  being the azimuthal angle around the  $z$ -axis. The pseudorapidity is defined in terms of the polar angle  $\vartheta$  as  $\eta = -\ln(\tan(\vartheta/2))$ . It is the massless approximation of rapidity, the angle parameterizing special relativity's boosts. Most hadron productions are roughly constant in  $\eta$ , and for two massless particles traveling in different directions, their difference in  $\Delta\eta$  is invariant. Therefore, angular distance is measured in units of  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2}$ .

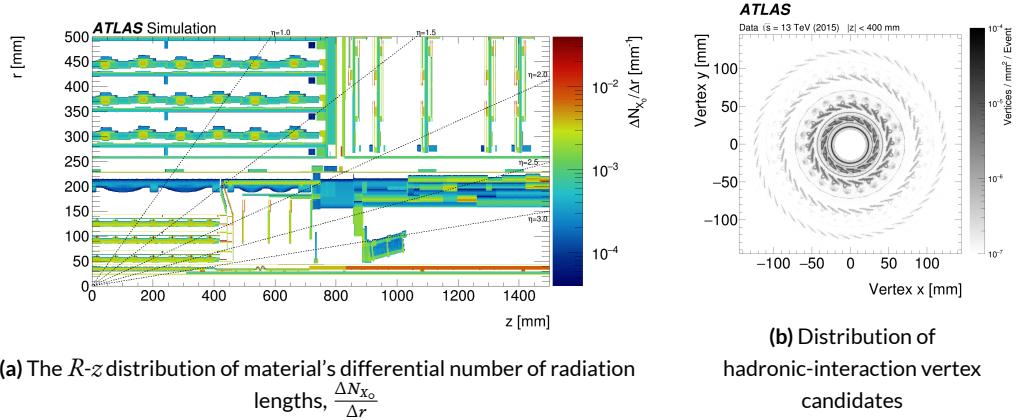
The region with  $|\eta| < 1.5$  is called “central”. It consists of the “barrel” elements, surrounding the beam line cylindrically. For  $|\eta| > 1.5$ , the region is called “endcap”, and the detector elements are arranged as disks perpendicular to the beam line. At high  $|\eta| > 2.5$ , the region is referred to as “forward”.

### 2.2.2 INNER DETECTOR

The ID covers the pseudorapidity range  $|\eta| < 2.5$ . It consists of three parts: silicon pixel (PIXEL), silicon microstrip (SCT), and straw-tube transition-radiation tracking (TRT) detectors. An additional pixel detector layer (IBL)<sup>36</sup>, inserted at a mean radius of 3.3 cm, is used in the Run-2 data-taking and improves the identification of  $b$ -jets<sup>37</sup>. A 10 GeV charged particle in the barrel region expect 1 + 3 IBL-Pixel hits, 8 SCT hits and 36 TRT hits.

ID is designed to provide charged particle momentum measurement with  $\sigma_{p_T}/p_T \sim 0.05\% p_T \oplus 1\%$  and vertex reconstruction. Because of this, each detector proves measurement accuracies of the order 10  $\mu\text{m}$  in  $R\text{-}\varphi$  and 100  $\mu\text{m}$  in  $z$ . Figure 2.4a<sup>38</sup> shows the  $R\text{-}z$  distribution of the material for a quadrant of the barrel region PIXEL and SCT. The intensity of a particle beam decreases exponentially in radiation length.  $I(x) = I_0 e^{-x/X_0}$ , where  $I$  is the intensity,  $x$  is the distance traveled, and  $X_0$

is the radiation length. Figure 2.4b shows the distribution of hadronic-interaction vertex candidates in  $|\eta| < 2.4$  and  $|z| < 400$  mm for 13 TeV data.



**Figure 2.4:** Geometry of IBL, PIXEL and SCT detectors in Run2.

### 2.2.3 CALORIMETER

Lead/liquid-argon (LAr) finely segmented sampling calorimeters provide EM energy measurements. A steel/scintillator-tile hadronic calorimeter covers the central pseudorapidity range ( $|\eta| < 1.7$ ). The endcap and forward regions are instrumented with copper/tungsten and LAr calorimeters for both the EM and hadronic energy measurements up to  $|\eta| = 4.9$ . The calorimeters also provide basic EM/Hadronic trigger information, with fast analogue summing in coarse granularity.

EM calorimeter (ECal) is designed to have  $> 22$  radiation lengths in the barrel and  $> 24$  in the endcap. It provides EM measurement with  $\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$ . The hadronic calorimeter (HCal) has approximately 9.7 interaction length in the barrel and 10 in the endcap. HCal provides hadronic measurement with  $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$  in the Barrel and Endcap regions, and  $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$  in the forward region.

#### 2.2.4 MUON SPECTROMETER

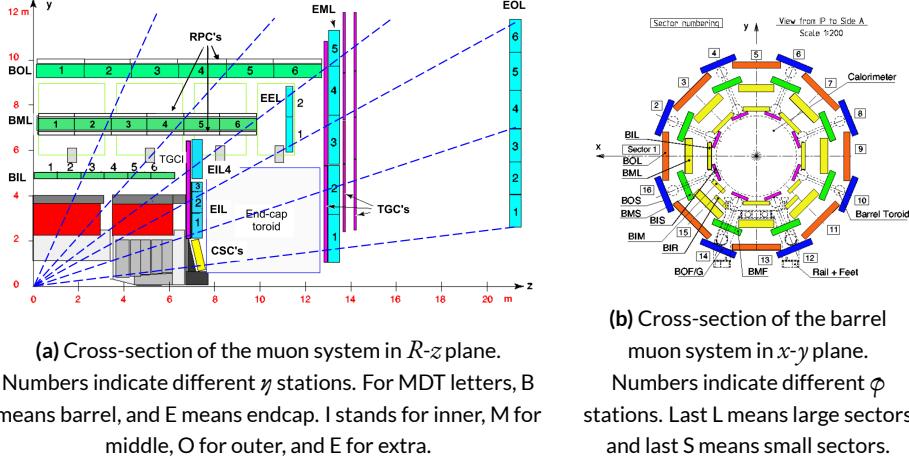


Figure 2.5: The overall layout of the ATLAS MuonSpectrometer.

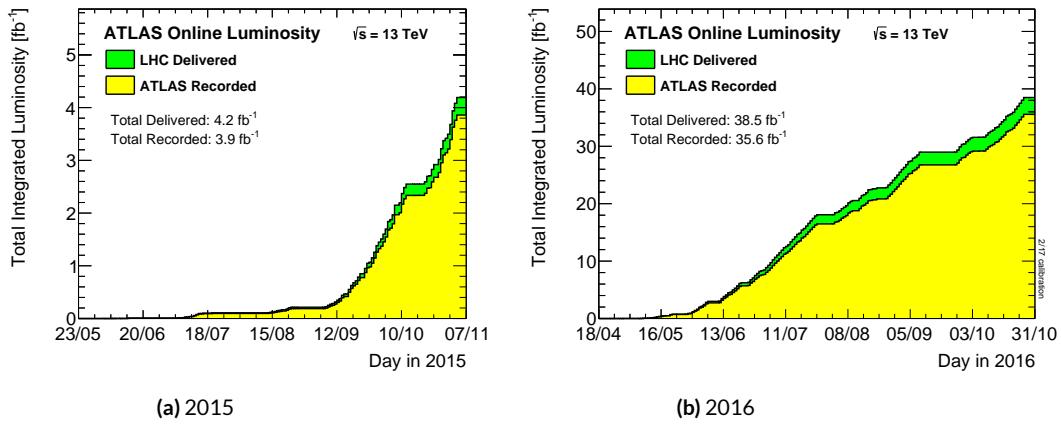
The muon spectrometer (Figure 2.5) surrounds the calorimeters and includes three large superconducting air-core toroids. The field integral of the toroids ranges between 2 and 6 T/m for most of the detector. Because of this bending power, the MS measures Muon momentum stand-alone, with  $\sigma_{p_T}/p_T \sim 10\%$  at  $p_T = 1\text{ TeV}$ . Muon Drift Tubes (MDT) and Cathode Strip Chambers (CSC) provide precision tracking. Each MDT has  $80\text{ }\mu\text{m}$  spacial resolution, with an alignment precision of  $30\text{ }\mu\text{m}$ . Resistive Plate Chambers (RPC) in the barrel and Thin Gap Chambers (TGC) provide triggering, with  $1.5\text{-}5\text{ ns}$  timing resolution. The muon spectrometer defines the overall dimensions of the ATLAS detector.

#### 2.2.5 TRIGGER AND DATA ACQUISITION

A dedicated trigger system is used to select events<sup>39</sup>. The first-level trigger ( $L_1$ ) is implemented in hardware and uses the calorimeter and muon detectors to seed regions of interest (RoI) and reduce

the accepted event rate to 100 kHz. This is followed by a software-based high-level trigger (HLT) that reduces the accepted event rate to 1 kHz on average. To avoid too high accept rates for certain triggers, the triggers are often prescaled, which means the accepted events get rejected at the prescale. For example, a prescale of two means only every second event passing all trigger conditions gets accepted.

Over 2015 and 2016, both the LHC and the ATLAS performed outstandingly <sup>4</sup>. The total data recording efficiency for ATLAS is around 92%, shown in Figure 2.6.



**Figure 2.6:** Cumulative luminosity vs. time delivered to (green) and recorded by ATLAS (yellow) during stable beams for  $p\bar{p}$  collisions at 13 TeV centre-of-mass energy.

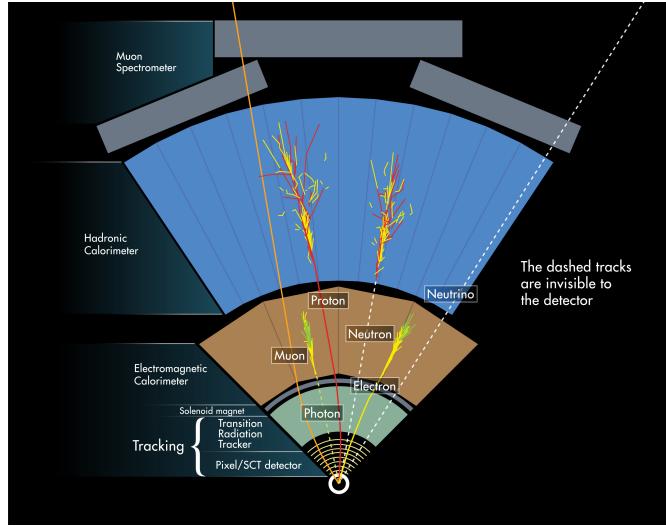
*“A picture is worth a thousand swords.”*

Tony

# 3

## Reconstruction and Objects

Reconstruction is the construction of particles from raw detector readouts. In each  $p\bar{p}$  collision recorded by ATLAS, charged particles bend in the magnetic field and leave tracks in the ID, electrons and photons deposit their energies in ECAL, hadrons are absorbed in HCAL, muons leave extra tracks in the MS, and neutrinos are inferred by the conservation of momentum in the transverse plane. Figure 3.1 gives an overview of the different sub-detectors that each type of particle will interact with in ATLAS. Quark reconstruction and identification is particularly important for this thesis, with di-Higgs decaying to  $b\bar{b}b\bar{b}$ .



**Figure 3.1:** Illustration of particle interactions in ATLAS.

### 3.1 ID TRACKS AND VERTICES

ID tracks originate from clusters based on PIXEL and SCT energy deposits. Three clusters form a seed, and the seeds are combined to build track candidates using a Kalman filter. After ambiguity solving, an artificial neural network is used to identify merged clusters. Finally, ID tracks are built from CPU intensive high resolution fits. The tracks are required to have  $p_T > 0.4 \text{ GeV}$  and  $|\eta| < 2.5$ .

Each  $p\bar{p}$  collision generates multiple vertices, and the vertices are reconstructed from the available ID tracks. The primary vertex (PV), or the hard-scatter vertex, is selected as the one with the largest  $\sum p_T^2$ , where the sum is over all tracks with transverse momentum  $p_T > 0.4 \text{ GeV}$  that are associated with the vertex. The ID tracks are usually required to have at least 1 PIXEL hit and 6 SCT hits, and to be tightly matched to the primary vertex.

The performance of track reconstruction is highly dependent on the momentum of the particle. With higher momentum, the decay tracks have smaller separations in the inner detector, hindering the resolving cluster process, and thus degrading the track identification efficiency. For a 1 TeV  $b$ -hadron, the reconstruction track efficiency is 83%, compared to 95% for a 200 GeV  $b$ -hadron <sup>40</sup>.

## 3.2 JETS

When a quark or gluon is produced in  $p\bar{p}$  collisions, it produces a spray of hadrons, which is known as a jet. Jets are built from topological clusters of energy deposits in calorimeter cells <sup>41</sup>, using a four-momentum reconstruction scheme with massless clusters as input. The directions of jets are corrected to point back to the primary vertex. Typically, jets are reconstructed using the anti- $k_t$  algorithm with different values of the radius parameter  $R$ .  $R$  appears in the denominator of the clustering distance metric. It determines the radial size of the jet in  $\eta$ - $\phi$  plane.

### 3.2.1 SMALL- $R$ JETS

The jets with  $R = 0.4$  (“small- $R$  jets”) are reconstructed from clusters calibrated at the electromagnetic (EM) scale. The jets are corrected for additional energy deposited from pile-up interactions using an area-based correction <sup>42</sup>. They are then calibrated using  $p_T$ - and  $\eta$ -dependent calibration factors derived from simulation, before global sequential calibration <sup>43</sup> is applied, which reduces differences in calorimeter responses to gluon or quark-initiated jets. The final calibration is based on in situ measurements in collision data <sup>44</sup>.

“Small- $R$  jets” are required to be consistent with the primary vertex, in order to avoid contamination from pileup interactions. The jet vertex fraction (JVF) is a useful variable for this purpose. It is the ratio of tracks associated with a primary vertex to the total number of tracks inside a jet. Jets from the PV should have most tracks consistent with the PV and therefore have a large JVF value.

### 3.2.2 LARGE- $R$ JETS

The jets with  $R = 1.0$  (“large- $R$  jets”) are built from locally calibrated<sup>43</sup> topological clusters.

They are trimmed<sup>45</sup> to minimize the impact of energy deposits from pile-up interactions. Trimming proceeds by reclustering the jet with the  $k_t$  algorithm<sup>46</sup> into  $R = 0.2$  sub-jets and then removing those sub-jets with  $\hat{p}_T^{\text{subjet}}/\hat{p}_T^{\text{jet}} < 0.05$ , where  $\hat{p}_T^{\text{subjet}}$  is the transverse momentum of the sub-jet and  $\hat{p}_T^{\text{jet}}$  that of the original jet. The energy and mass scales of the trimmed jets are then calibrated using  $p_T$ - and  $\eta$ -dependent calibration factors derived from simulation<sup>47</sup>.

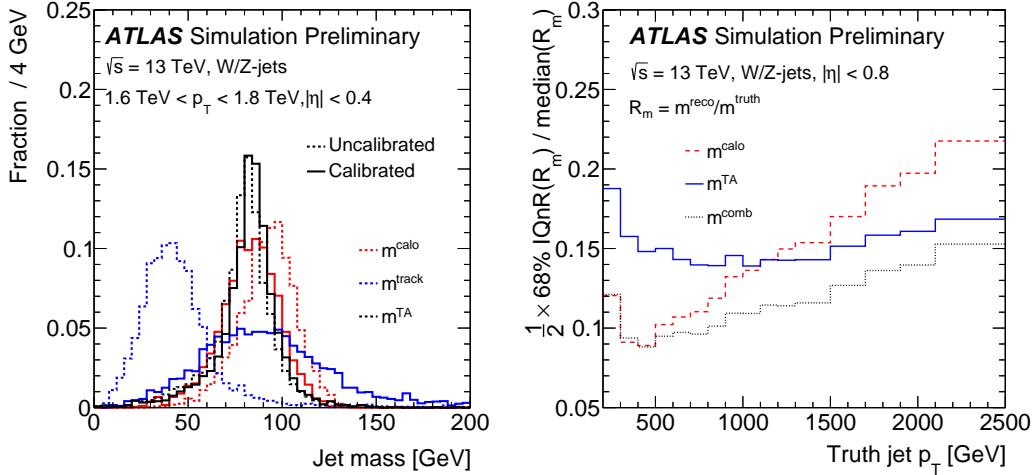
The calorimeter-based jet mass  $m^{\text{calo}}$  for a large-radius calorimeter jet  $J$  is computed from the calorimeter cell cluster constituents  $i$  with energy  $E_i$  and momentum  $\vec{p}_i$ :

$$m^{\text{calo}} = \sqrt{\left(\sum_{i \in J} E_i\right)^2 - \left(\sum_{i \in J} \vec{p}_i\right)^2}. \quad (3.1)$$

For a boosted massive particle, the angular spread in the decay products scales as  $\frac{1}{p_T}$ . For highly boosted cases, the spread is comparable with the  $\eta \times \varphi \sim 0.1 \times 0.1$  calorimeter granularity. Tracking information is used to maintain performance beyond the calorimeter granularity limit. The track-assisted jet mass,  $m^{\text{TA}}$ , is defined as:

$$m^{\text{TA}} = \frac{\hat{p}_T^{\text{calo}}}{\hat{p}_T^{\text{track}}} \cdot m^{\text{track}}. \quad (3.2)$$

where  $\hat{p}_T^{\text{calo}}$  is the transverse momentum of the large- $R$  calorimeter jet,  $\hat{p}_T^{\text{track}}$  is the transverse momentum of the four-vector sum of tracks associated to the large- $R$  calorimeter jet, and  $m^{\text{track}}$  is the invariant mass of this four-vector sum. This ratio corrects for charged-to-neutral hadron fluctuations, and therefore improves the resolution with respect to track-only jet mass.



(a) Uncalibrated (dashed line) and calibrated (solid line) jet mass distribution.

(b) The fractional jet mass resolution vs. the truth jet mass transverse momentum.

**Figure 3.2:** Uncalibrated (dashed line) and calibrated (solid line) reconstructed jet mass distribution 3.2a, and the jet mass resolution vs jet  $p_T$  3.2b for calorimeter-based jet mass,  $m_{calo}$  (red), track-assisted jet mass  $m_{TA}$  (black) and the invariant mass of four-vector sum of tracks associated to the large-radius calorimeter jet  $m_{track}$  (blue) for W/Z-jets<sup>5</sup>.

The above two mass definitions are only weakly correlated with each other, so they can be linearly combined to the combined mass,  $m^{comb}$ , by weighting the components with  $w$ :

$$m^{comb} = w \cdot m^{calo} + (1 - w) \cdot m^{TA}. \quad (3.3)$$

where  $w$  is determined for each large- $R$  jet from the resolution functions of the calibrated track and calo mass terms. This results in a smaller mass resolution and better estimate of the median mass value than obtained using only calorimeter energy clusters, as shown in Figure 3.2.

### 3.2.3 TRACK JETS

Track jets are essentially clustered charged hadron tracks. They are reconstructed from ID tracks using the anti- $k_t$  algorithm with a fixed  $R = 0.2$ . Once the track jet axis is determined, an extra step of track association is performed to select tracks with looser impact parameter requirements,

in order to collect the tracks needed for effectively running the  $b$ -tagging algorithms. Only track jets with at least two tracks are kept. Track jets are also required to have  $p_T > 10 \text{ GeV}$  and  $|\eta| < 2.5$ , in order to suppress track jets from light flavors.

Track jets are associated to large- $R$  jets using Ghost-Association. “Ghosts” are track jet 4-vectors, with each track jet’s  $p_T$  set to an infinitesimal amount, essentially keeping only the direction of the 4-vector. This ensures that Large- $R$  jets reconstruction is not altered by the ghosts when the calorimeter clusters plus ghosts are reclustered. Reclustering is then performed using the anti- $k_t$  algorithm with  $R = 1.0$ . The calorimeter jets after reclustering are identical to the parents of the trimmed jets used in this analysis, with the addition of the associated track jets retained as constituents. In addition, the track jets corresponding to the ghosts that survive the trimming procedure (and thus are clustered into one of the surviving sub-jets) are the track jets ghost-associated to the trimmed jet. The small radius parameter of the track-jets enables two nearby  $b$ -hadrons to be identified when their  $\Delta R$  separation is small, which is beneficial when reconstructing high- $p_T$  Higgs boson candidates.

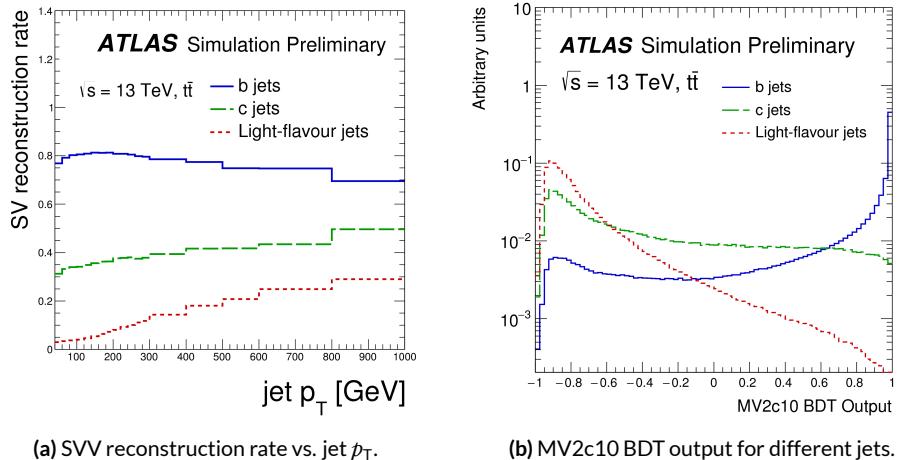
### 3.3 FLAVOR TAGGING

Some jets are formed from quarks with different flavors. Identifying the flavor of a jet is called flavor tagging. Jets originating from a  $b$ -quark is referred as a  $b$ -jet, from a  $c$ -quark is referred as a  $c$ -jet, and from a other quarks other than the  $t$ -quark is referred as a light jet.  $b$ -tagging is particularly useful for this analysis, since there are four  $b$ s in the final state.

$B$ -hadrons have a lifetime on the order of  $10^{-12}$  seconds, which makes a flight distance of  $0.5 \text{ mm}^{\textcolor{red}{1}}$ . This results in a displaced decay vertex that can be identified in vertex reconstruction. This allows some  $b$ -jets to be distinguished from other flavors of jets theoretically.

ATLAS uses three different basic  $b$ -tagging algorithms, which provide complementary information:

- Impact parameter (IP) based algorithm: it uses the transverse and longitudinal impact parameters  $d_o$  and  $z_o$  of the tracks inside a jet to determine their consistency with the primary vertex. The algorithm uses two or three dimensional templates for light,  $c$ , and  $b$  jets and evaluates the likelihood of the jet coming from each of these types.
- Inclusive secondary vertex (SV) reconstruction algorithm: it uses tracks inside the jet to fit for vertices that are displaced from the primary vertex. The algorithm provides information on the invariant mass of tracks pointing to same vertex, the number of two track vertices, and the angular separation between the jet and the PV  $\rightarrow$  SV direction.
- Decay chain multi-vertex reconstruction algorithm, or JetFitter (JF): it reconstructs the full flight path of the  $b$  by looking for multiple displaced vertices along the same direction. A Kalman filter is used to find common line for the  $b$  and  $c$  vertices, and hence the algorithm exploits the topology of the weak  $b/c$ -hadron decay chain.



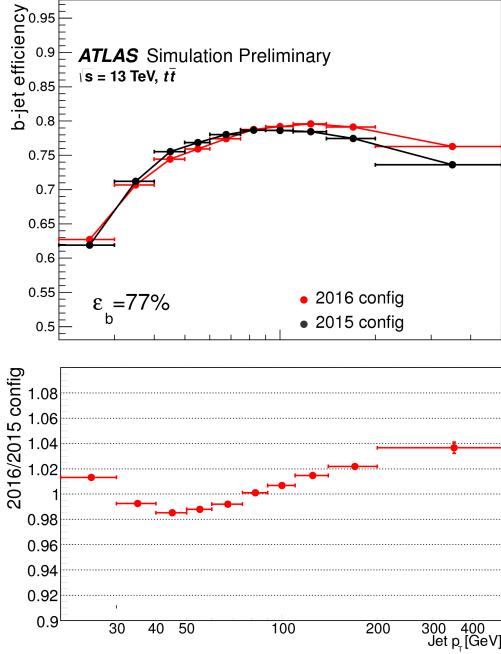
**Figure 3.3:** Secondary vertex reconstruction rate and MV2c10 output for  $b$ -jets (solid blue),  $c$ -jets (dashed green) and light-jets (dotted red) evaluated with simulated  $t\bar{t}$  events.

Jets containing  $b$ -hadrons are identified using a score value computed from a boosted decision tree(BDT) algorithm  $MV2c10$ <sup>6,48</sup>, which makes use of observables provided by three algorithms above. The  $MV2c10$  algorithm is trained on a sample with charm composition of 7%. It is applied to a set of charged-particle tracks that satisfy quality and impact parameter criteria and are matched to each jet. Hence either a small- $R$  jet or a track jet can be  $b$ -tagged.

The  $b$ -tagging working point (wp) is a fixed cut on the  $MV2c10$  value that lead to an efficiency of 70% for  $b$ -jets with  $p_T > 20$  GeV when evaluated in a sample of simulated  $t\bar{t}$  events. This working point corresponds to a rejection rate of jets originating from  $u, d$  or  $s$ -quarks or gluons of 380 for the jets with  $R = 0.4$  and 120 for the track jets. The rejection of jets from  $c$ -quarks is 12 for the  $R = 0.4$  jets and 7.1 for the track jets.

In this thesis, the track-jets have a wider  $p_T$  range, between 50 – 400 GeV, and the same working point leads to  $b$ -tagging efficiencies varying from 40% at low  $p_T$ , to 80% for  $p_T$  values of about 150 GeV, to 60% at high  $p_T$ . This can be seen in Figure 3.4. The increase of tracks from fragmentation in the high jet  $p_T$  region is the main reason for the performance degradation. As the jet  $p_T$  increases, the number of fake vertices is increasing, while the secondary vertex reconstruction efficiency for  $b$  and  $c$  jets decreases with jet  $p_T$ . This is shown in Figure 3.5. This non-trivial jet  $p_T$  dependence of  $b$ -tagging performance is one of the major challenge of this analysis.

Correction factors are applied to the simulated MC samples to compensate for differences between data and simulation in the  $b$ -tagging efficiency for  $b, c$  and light-jets. The correction for  $b$ -jets is derived from  $t\bar{t}$  events with final states containing two leptons, and the corrections are consistent with unity with uncertainties at the level of a few percent over most of the jet  $p_T$  range.

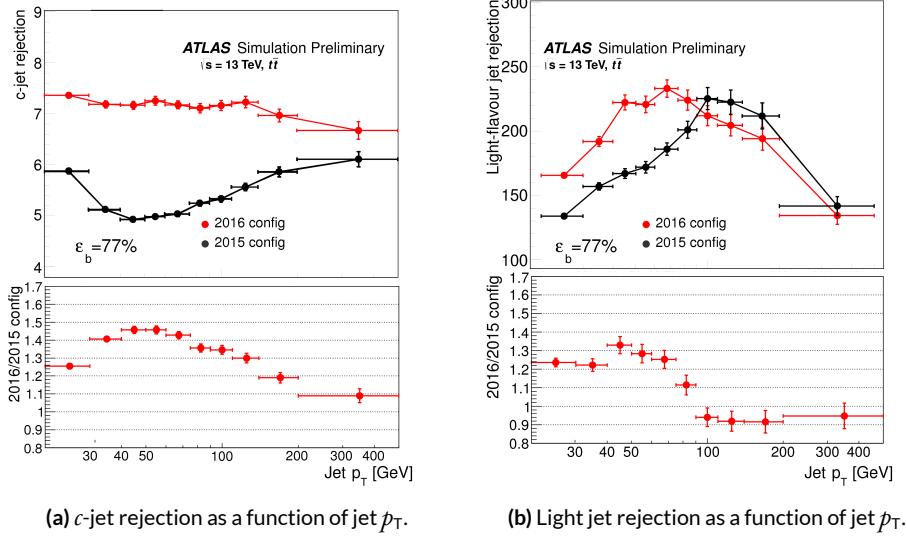


**Figure 3.4:**  $b$ -jet efficiency for the fixed cut working point with a  $b$ -jet efficiency of 77% as a function of the jet  $p_T$  for the comparison between the MV2c10  $b$ -tagging algorithm employed for the 2016 analyses (2016 config) and the previous version of the tagger, MV2c20 (2015 config), which has 15%  $c$ -fraction in the training

### 3.4 LEPTONS

Electron and photon identification is based on matching tracks to energy clusters in the ECAL and relying on the longitudinal and transverse shapes of the EM shower<sup>49</sup>. Well-reconstructed ID tracks matched to EM clusters are classified as electron candidates, while EM clusters without matching tracks are classified as unconverted photon candidates. Clusters matched to a reconstructed conversion vertex or to pairs of tracks consistent with the conversion hypothesis are classified as converted photon candidates. Electrons and photons are not used in this thesis.

Hadronic tau decays into a tau neutrino and one or three charged pions and up to two neutral pions<sup>50</sup>. Hence tau reconstruction is seeded by jets, and is matched to one or three associated tracks,



**Figure 3.5:** Light-flavour jet and  $c$ -jet rejection as a function of jet  $p_T$  for the previous (2015 config) MV2c20 and the current MV2c10 configuration (2016 config). A fixed cut at 77%  $b$ -jet efficiency operating point is used<sup>6</sup>.

with a total electric charge of  $\pm 1$ . A Boosted Decision Tree identification procedure, based on calorimetric shower shapes and tracking information is used to reject fakes from jets. Hadronic taus are not used in this thesis.

Neutrinos are inferred from the missing transverse momentum (MET), or  $E_T^{miss}$ . Neutrinos do not interact with the ATLAS. Their presence can only be deduced from the conservation of transverse momentum in each collision, as the incoming protons have no net momentum in the transverse plane. MET is calculated as the negative vectorial sum of the  $p_T$  of all fully reconstructed and calibrated physics objects. This procedure includes a soft term, which is calculated using the ID tracks that originate from the primary vertex but are not associated with reconstructed objects. MET is not used in this thesis.

Muons are identified by matching ID tracks with reconstructed MS tracks<sup>5</sup>. For this thesis, muons must have  $p_T > 4 \text{ GeV}$ ,  $|\eta| < 2.5$  and to satisfy “medium” muon identification criteria<sup>5</sup>.

Muons are used in this thesis, because  $b$  hadrons decay to muons with  $\sim 20\%$  probability. This will be demonstrated in the later chapters.

*Ugliness is in a way superior to beauty because it lasts.*

Serge Gainsbourg

# 4

## Data and Simulation

### 4.1 DATA

This analysis uses 2015 and 2016 LHC  $p\bar{p}$  collision datasets at  $\sqrt{s} = 13$  TeV recorded by the ATLAS experiment. Data were collected during stable beam conditions and when all relevant detector systems were functional. A Good Run List (GRL) is generated after gathering online and offline data quality reviews of the dataset after reconstruction. Typically, any  $> 10\%$  defect in any detector subsystem makes the corresponding Lumiblocks (LB) fail the GRL requirement. The integrated luminosity of the 2015 dataset passing the GRL is  $3.2 \text{ fb}^{-1}$ , and the 2016 dataset passing a different GRL is  $32.9 \text{ fb}^{-1}$ . These values are about 82% and 92% of the data ATLAS recorded, in Figure 2.6.

In the resolved analysis, a combination of  $b$ -jet triggers is used. Events are required to feature either one  $b$ -tagged jet with transverse momentum  $p_T > 225$  GeV, or two  $b$ -tagged jets, either both satisfying  $p_T > 35$  GeV or both satisfying  $p_T > 55$  GeV, with different requirements on the  $b$ -tagging. Some triggers require additional non- $b$ -tagged jets. Due to a change in the online  $b$ -tagging algorithm between 2015 and 2016, the two datasets are treated independently until they are combined in the final statistical analysis. After the selection described later, this combination of triggers is estimated to be 65% efficient for simulated signals with a Higgs boson pair invariant mass,  $m_{HH}$ , of 280 GeV, rising to 100% efficiency for resonance masses greater than 600 GeV.

During 2016 data-taking, a fraction of the data was affected by a bug. The movement of beam spot was not accounted for in the online vertex reconstruction. This reduced the efficiency of the algorithms used to identify  $b$ -jets. This reduces the integrated luminosity of the 2016 dataset for the resolved analysis to  $24.3 \text{ fb}^{-1}$ .

In the boosted analysis, events were selected from the 2015 dataset using a trigger that required a single anti- $k_t$  jet with radius parameter  $R = 1.0$  and with  $p_T > 360$  GeV. In 2016, a similar trigger was used but with a higher threshold of  $p_T > 420$  GeV. The efficiency of these triggers is 100% for simulated signals passing the jet requirements as described later, so the 2015 and 2016 datasets were combined into one dataset.

The data is further skimmed into the Derived Analysis Object Data (DAOD). The ATLAS offline software 20.7.8.7 derivation cache is used, with version  $p$ -tags  $p2950$ . The boosted slimming keeps events with at least two large- $R$  jets with  $p_T > 200$  GeV. The final input data file has name format:  $dataYR\_13TeV.periodPR.physics\_Main.PhysCont.DAOD\_EXOT8.grpYR\_vo1\_p2950$ , where YR is 15 and PR is DEFGHJ for 2015 data, and YR is 16 and PR is ABCDEFGIKL for 2016 data.

## 4.2 MC

All Monte Carlo (MC) samples used in this analysis are produced with full simulation. For all simulated samples, charm-hadron and bottom-hadron decays were handled by `EVTGEN 1.2.0`<sup>3</sup>. To simulate the impact of multiple  $p\bar{p}$  interactions that occur within the same or nearby bunch crossings (pile-up), minimum-bias events generated with `PYTHIA 8`<sup>52</sup> using the A2 set of tuned parameters<sup>53</sup> were overlaid on the hard-scatter event. The detector response was simulated with `GEANT 4`<sup>54,55</sup> and the events were processed with the same reconstruction software as that used for the data. Simulated data samples from the ATLAS MC15c campaign are used, corresponding to  $p$ -tags p2952–p2949.

### 4.2.1 BACKGROUNDS

A very small fraction of the background arises from  $Z + \text{jets}$  events. The  $Z+\text{jets}$  sample was generated using `PYTHIA 8.186` with the NNPDF2.3 LO PDF set.

The  $t\bar{t}$  background is modeled using large all-hadronic and non-all-hadronic samples that have both been generated with `POWHEG-BOX v1`<sup>56</sup> using the CT10 PDF set. The parton shower, hadronization, and the underlying event were simulated using `PYTHIA 6.428` with the CTEQ6L1 PDF set and the corresponding Perugia 2012 set of tuned underlying-event parameters<sup>57</sup>. The  $t$ -quark mass in both samples is set to 172.5 GeV. Higher-order corrections to the  $t\bar{t}$  cross section were computed with `Top++ 2.0`<sup>58</sup>. These incorporate NNLO corrections in QCD, including resummation of NNLL soft gluon terms. The  $t\bar{t}$  MC samples are normalized to the NNLO+NLL predicted inclusive  $t\bar{t}$  cross-section of 1821.87 pb multiplied by the all-hadronic branching ratio of 0.457 and non-all-hadronic of 0.543.

In order to keep statistical fluctuations small across the dijet invariant mass spectrum, especially for large values of  $m_{tt}$ , additional  $t\bar{t}$  samples are generated in slices of  $t\bar{t}$  invariant mass. The cross-section of the  $t\bar{t}$  process is normalized to NNLO+NNLL in QCD, as calculated by Top++ 2.0. Overlap with the inclusive  $t\bar{t}$  samples is removed by a fixed cut on the truth value of  $m_{tt}$  at 1100 GeV.

A PYTHIA dijet sample is used to understand the physical processes contributing to the multi-jet background and characteristics of the event selection. This MC sample is generated without a heavy flavor filter, hence is limited by the total generated number of events, given the high background rejection factors of the analysis selection.

#### 4.3 SIGNAL

In all signal samples, the mass of the Higgs boson ( $m_H$ ) was set to 125 GeV. The signal MC contains truth information, like the two Higgs and the b quark four-momentum before detector interactions. This enables a  $\Delta R$  truth matching between the reconstructed objects and the Higgs.

SM non-resonant production of Higgs boson pairs via the gluon–gluon fusion process was simulated at NLO with MG5\_aMC@NLO, using form factors for the top-quark loop from HPAIR<sup>66,67</sup>. The simulated events were reweighted to reproduce the  $m_{hh}$  spectrum obtained<sup>68,69</sup>, which calculated the process at NLO in QCD while fully accounting for the top-quark mass. Interference effects between di-Higgs resonant production and SM non-resonant di-Higgs production are not included in the simulated samples.

Signal  $G_{KK}^* \rightarrow hh \rightarrow b\bar{b}b\bar{b}$  events were generated at leading order (LO) with MG5\_aMC@NLO 2.2.2<sup>59</sup> interfaced with PYTHIA 8.186 for parton-showering, hadronization and underlying-event simulation. The NNPDF2.3 LO parton distribution function (PDF) set<sup>60</sup> was used for both MG5\_aMC@NLO and PYTHIA. The A14 set of tuned underlying-event parameters was used. These signal samples

were generated with  $k/\bar{M}_{\text{Pl}} = 1$  or  $2$ . Relative to the resonance mass, widths of the graviton signals range from  $3\%$  (at low mass) to  $13\%$  (at the highest mass) for  $k/\bar{M}_{\text{Pl}} = 1$ , and  $6\%$  to  $25\%$  for  $k/\bar{M}_{\text{Pl}} = 2$ . The graviton samples were normalized using fixed cross sections<sup>61</sup>.

Signal 2HDM Scalar  $\rightarrow HH \rightarrow b\bar{b}b\bar{b}$  events were generated at LO in QCD with MG5\_aMC@NLO 2.2.3 interfaced with HERWIG++<sup>62</sup> for parton-showering, hadronization and simulation of the underlying event. CT10<sup>63</sup> PDF sets were used for MG5\_aMC@NLO and CTEQ6L1<sup>64</sup> for HERWIG++. The UE-EE-5-CTEQ6L1 set of tuned underlying-event parameters<sup>65</sup> was used. The scalar signals were generated with a width of  $1$  GeV, which represent generic narrow-width scalar signals. Because the width and branching ratios depend on 2HDM parameters, each mass point generated with this fixed width corresponds to a different point in the 2HDM parameter phase space.

Resonant signal samples for the scalar and  $k/\bar{M}_{\text{Pl}} = 1$  models were produced in  $10$  GeV steps between  $260$  and  $300$  GeV, in  $100$  GeV steps up to  $1600$  GeV, in  $200$  GeV steps up to  $2000$  GeV, and in  $250$  GeV steps up to  $3000$  GeV. Signal samples for the  $k/\bar{M}_{\text{Pl}} = 2$  model were produced with the same spacings but omitting the masses of  $270$  GeV,  $290$  GeV and  $2750$  GeV due to the larger generated width. Unless specified, the MC signal sample used as benchmark is  $k/\bar{M}_{\text{Pl}} = 1$   $G_{\text{KK}}^*$ , due to its width is medium among the three signal models.

*You gotta have a swine to show you where the truffles are.*

Edward Albee

# 5

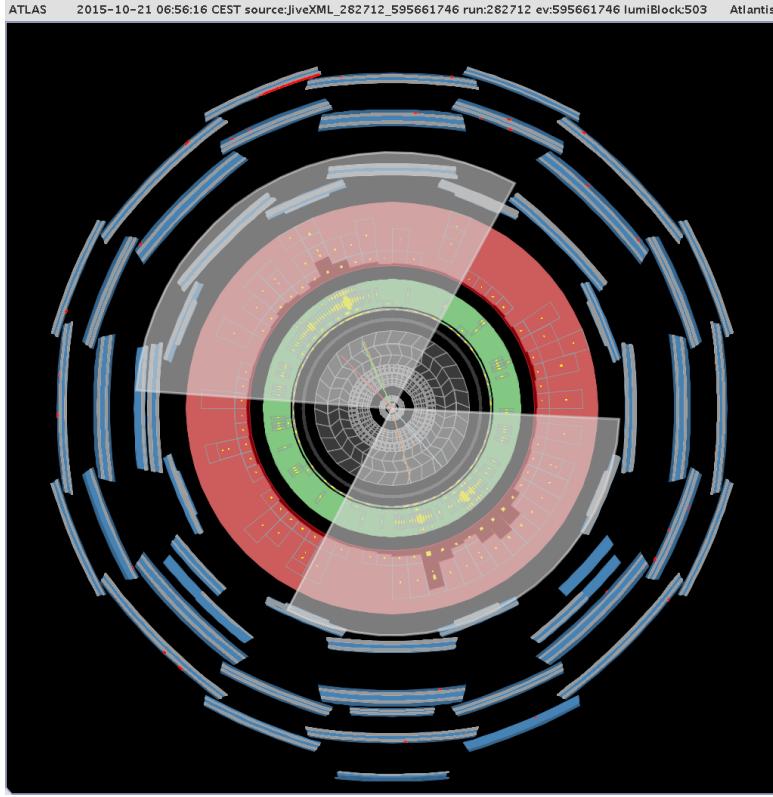
## Event Selection

An example of  $X \rightarrow hh \rightarrow b\bar{b}b\bar{b}$  boosted event is shown in Figure 5.1. This chapter explains how this event is selected as a  $X \rightarrow hh \rightarrow b\bar{b}b\bar{b}$  candidate.

### 5.1 DATA CLEANING

The following data cleaning requirements are made:

- Events with problems in SCT/TileCal/LAr are removed.
- Events that fail the jet cleaning procedure are removed. This is designed to exclude jets caused by detector noise, non-collision backgrounds and cosmic rays.
- Incomplete events are removed.



**Figure 5.1:** 2015 ATLAS boosted di-Higgs event candidate's event display. The data is taken in run 282712, lumiblock 503 and has event number 595661746. ID is in grey, ECAL is in green, HCAL is in red and MS is in blue. Only tracks with  $p_T > 25$  GeV are shown. Jets are gray cones, and ID tracks are colored lines in the ID. There are  $2b$  tagged tracks within each of the large- $R$  jets. The one large- $R$  jet has 119 GeV mass and 543 GeV  $p_T$ , and the other one has 127 GeV mass and 413 GeV  $p_T$ . The  $\Delta R$  between the two large- $R$  jets is 3.54 and the invariant mass of the two large- $R$  jets is 1336 GeV.

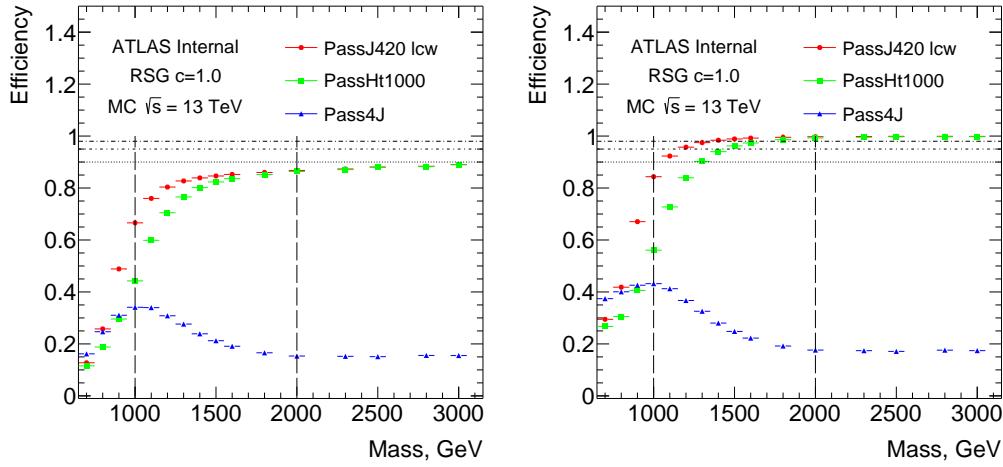
The analysis also runs over the debug stream, which contains events recorded that couldn't be reconstructed online due to CPU time constraints. No event passing the full signal selection is found.

## 5.2 TRIGGER

Events in data and MC are required to pass the lowest unprescaled large- $R$  jet trigger:

HLT\_j360\_a10\_lcw in 2015 and HLT\_j420\_a10\_lcw in 2016. The triggered jets are topocluster jets with local calibration weights and pile-up subtraction. They are seeded by the low-

est unprescaled L1 jet trigger, `L1_J100`. LCW cluster trigger is chosen, because the other option, reclustered large- $R$  jet trigger, has slower turn-on in multi-jet events. Other options such as the lowest unprescaled HT trigger, `HLT_ht1000`, has a much slower turn-on compared to large- $R$  jet triggers. Another trigger option is `HLT_4j100`, but because of the boosted jets merging, the trigger efficiency decreases rapidly as the signal mass increases. The results are shown in Figure 5.2.

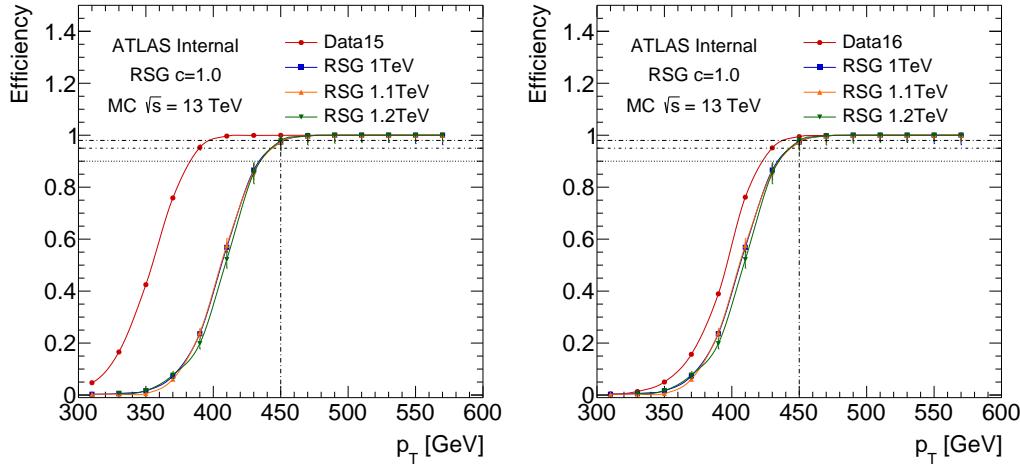


**Figure 5.2:** Different trigger efficiencies as a function of the signal resonance mass with respect to all events with no selection (left) and with respect to events passing the two large- $R$  jets  $p_T > 400$  GeV and leading/subleading jet  $p_T > 250$  GeV (right). For 1.4 TeV signal, the trigger efficiency is about 98%.

The selected large- $R$  triggers are found to have  $> 98\%$  efficiency for signals with mass above 1400 GeV. The trigger turn-on curve in 2015 and 2016 data, as a function of leading jet  $p_T$ , is shown in Figure 5.3.

### 5.3 OBJECT SELECTION

The specific physics objects used in the boosted analysis are described in previous sections and reiterated in Table 5.1.

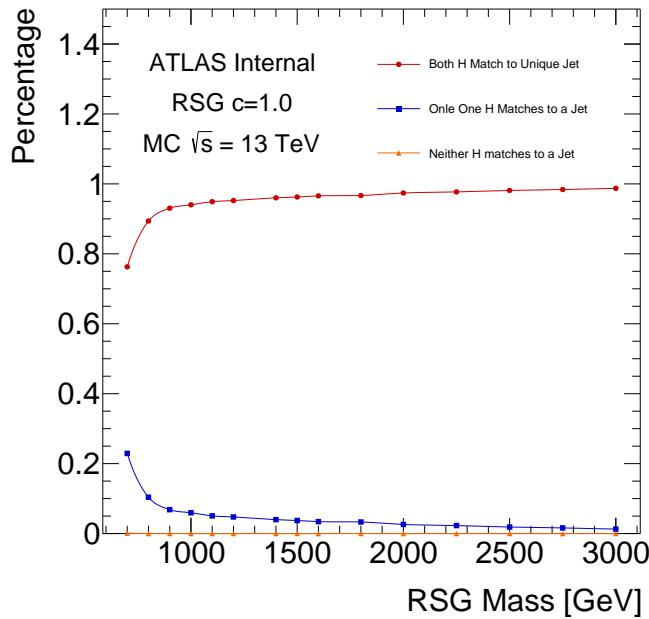


**Figure 5.3:** Large- $R$  jet trigger efficiencies, defined as the fraction of events fired trigger with a given highest large- $R$  jet  $p_T$ , measured in 2015 Data (HLT\_j360\_a10\_lcw, left) and 2016 Data (HLT\_j420\_a10\_lcw, right) and MC.

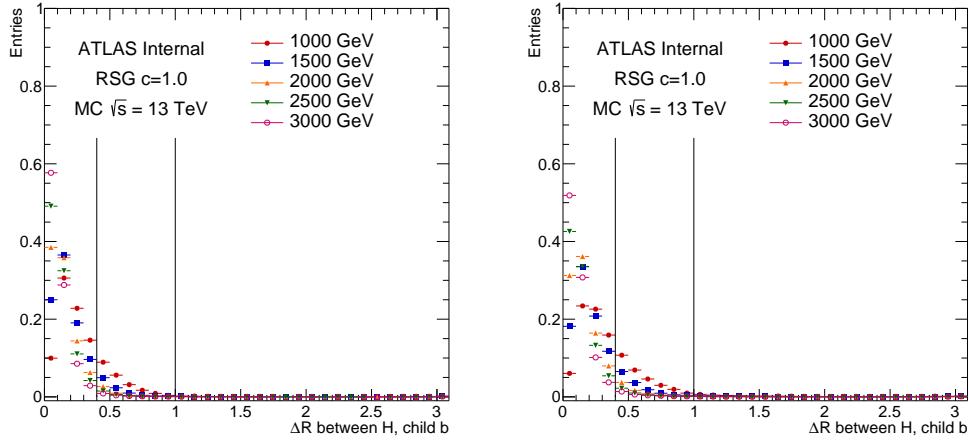
**Table 5.1:** Physics objects and their technical names in the boosted analysis.

object	technical name
large- $R$ calorimeter jets	AntiKt1oLCTopoTrimmedPtFrac5SmallR2oJets
small- $R$ track jets	AntiKt2PVoTrackJets
b-tagging	on track jets, MV2c10, 70% b-tagging wp

Each event must have at least two high momentum large- $R$  jets, each with at least one ghost associated track jets for  $b$ -tagging. They are sorted by  $p_T$ , and the highest  $p_T$  one is named as the leading large- $R$  jet, or the leading Higgs Candidate. The second highest  $p_T$  large- $R$  jet is the subleading large- $R$  jet, or the subleading Higgs Candidate. The large- $R$  jets are required to have  $p_T > 250$  GeV,  $|\eta| < 2$  to guarantee a good overlap with the tracking acceptance, and mass  $> 50$  GeV to avoid oversized derivations. The leading large- $R$  jet is also required to have  $p_T > 450$  GeV to be above the trigger turn on threshold. Only the leading and subleading large- $R$  jets are considered in the rest of this thesis. This selection is  $\sim 95\%$  efficient for 1.2 TeV signals, as shown in Figure 5.4.  $R = 1.0$  ensures that the two  $b$  quarks and their decay products are very likely to be contained within the large- $R$  jet, as shown in Figure 5.5.

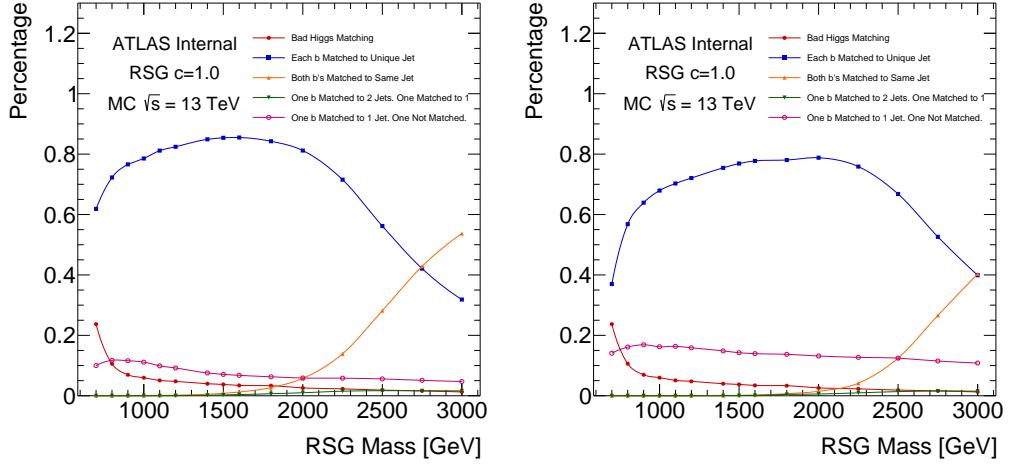


**Figure 5.4:** Percentages of truth Higgs to large- $R$  jet  $\Delta R < 1.0$  matching as a function of  $G_{KK}^*$  mass. Both Higgs almost never match to the same large- $R$  jet.



**Figure 5.5:** Normalized  $\Delta R$  between the truth Higgs (leading on left, subleading on right) and the truth children  $b$ -quarks for  $G_{KK}^*$  MCs. Lines are drawn at  $\Delta R = 0.4$  ( $R$  of small- $R$  jets) and  $\Delta R = 1.0$  ( $R$  of large- $R$  jets).

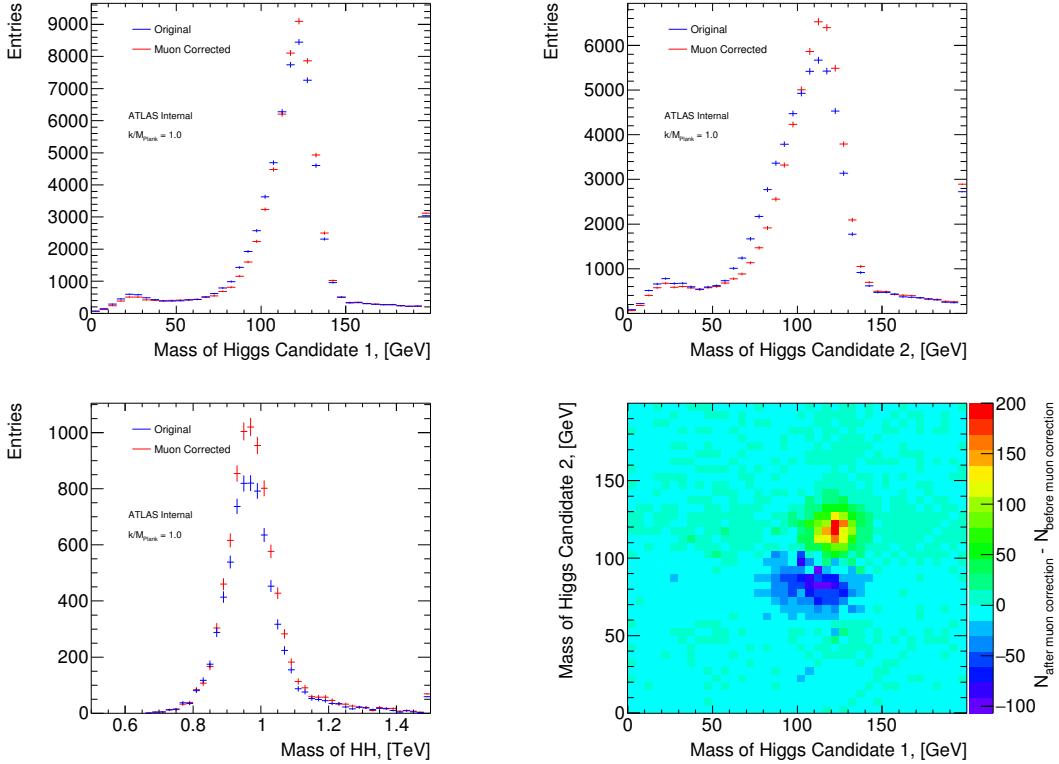
The track jets are required to have  $p_T > 10 \text{ GeV}$ ,  $|\eta| < 2.5$  and at least two tracks associated with it. A track jet is considered  $b$ -tagged if it has  $\text{MV}_{2\text{C}10} > 0.6455$ , see Section 3.3 for details. Each large- $R$  jet is required to have at least one, not two, track jet Ghost associated with it. This accounts for the  $R = 0.2$  track jets merging at really high boost, from signals above 2.5 TeV. If there are more than one track jet contained in the large- $R$  jet, they are also sorted by  $p_T$ . The highest  $p_T$  track jet is named as the leading track jet, and the second highest  $p_T$  one is named as the subleading track jet. Only the two highest  $p_T$  track jet is considered in this thesis. The one or two track jets  $\Delta R$  match to the truth  $b$  quarks 80%, as shown in Figure 5.6. In the Figure, red means the truth Higgs doesn't match the large- $R$  jet. Blue means both truth  $b$  matches the two track jets. Orange means two truth  $b$  matches one same track jet. Green indicates one  $b$  quark has  $\Delta R < 0.2$  for both leading and sub-leading track jet, and the other  $b$  is matched to one of the two track jets. Pink means one  $b$  quark matches to one of the two track jets, the other  $b$  doesn't match to the two leading track jets.



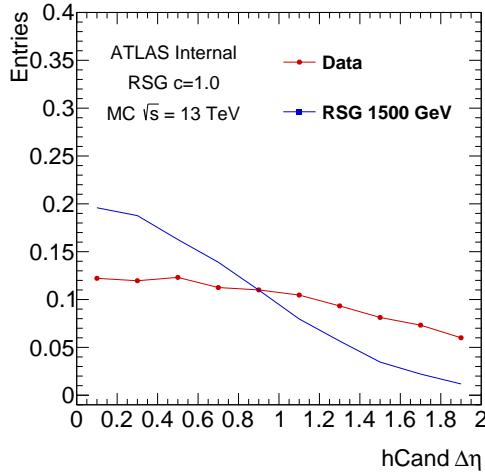
**Figure 5.6:** Percentage of  $\Delta R < 0.2$  matching truth  $b$ 's to track jets (leading Higgs on the left, subleading Higgs on the right) for different  $G_{KK}^*$  mass. The cases listed in the legend are orthogonal to each other. The cases not listed on the legend (including when a truth  $b$  is not contained in the large- $R$  jet) happen in total at most 1.6% of the time for a given  $G_{KK}^*$  mass.

A further muon correction accounts for energy loss due to leptonic  $b$ -hadron decays with a muon in the final state. The muon-in-jet corrections are applied only after the fiducial large- $R$  jet requirements on  $p_T$  and  $\eta$ . The muons are required have  $\Delta R < 0.2$  with the  $b$ -tagged track jets within each large- $R$  jet. In case more than one muon is found within a track jet, only the muon with the smallest  $\Delta R$  is considered. If two  $b$ -tagged track jets are found to have muons, both corrections are considered. The four-momenta of the matched muon is added to the large- $R$  jet four-momentum, with the muon calorimeter energy deposits subtracted. This correction is only applied to the calorimeter mass portion of the combined mass. The muon-in-jet correction improves the large- $R$  jet mass resolution by approximately 5%, and Figure 5.7 shows the impact of this correction on the 1 TeV  $G_{KK}^*$ .

Finally, the Higgs candidates (large- $R$  jets) are also required to have  $|\Delta\eta| = |\eta_{\text{leadJ}} - \eta_{\text{subJ}}| < 1.7$ . This is because the spin 2  $G_{KK}^*$  are produced mostly through s-channel, while the multijet events could also be produced through t-channels or u-channel. Figure 5.8 shows the distribution for sig-



**Figure 5.7:** Kinematics of 1 TeV  $G_{KK}^*$  before and after muon-in-jet corrections. The reconstructed Higgs masses (top row, left for leading large- $R$  jet, right for subleading large- $R$  jet) are closer to 125 GeV after the correction, which improves the signal efficiency for the signal region selection by  $\sim 10\%$  (bottom row, left for  $m_{jj}$ , right for event distribution differences on the leading-subleading large- $R$  jet mass plane.).



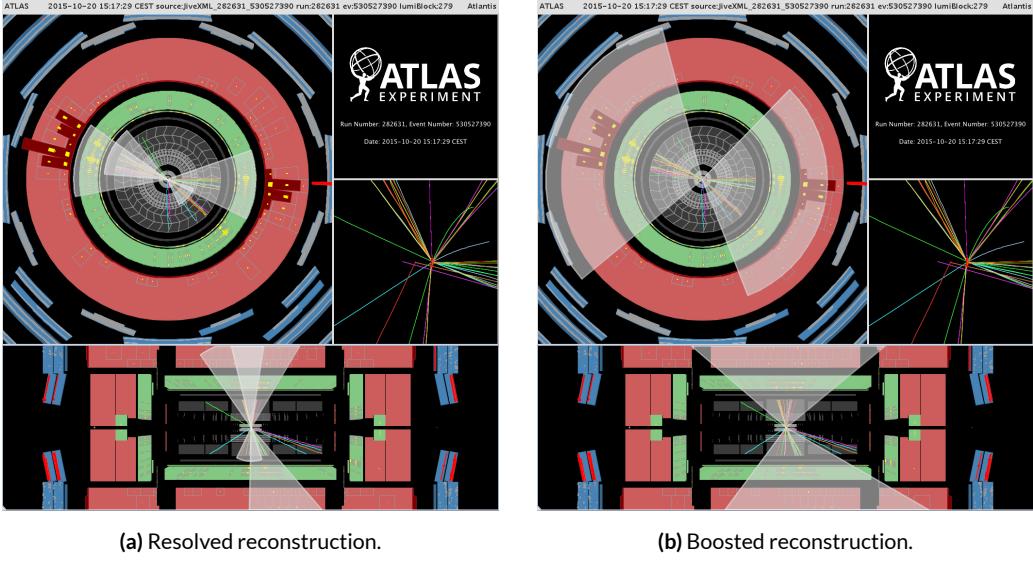
**Figure 5.8:** After large- $R$  jet requirements, normalized  $\Delta\eta_{JJ}$  distribution in  $1.5 \text{ TeV } G_{KK}^*$  and data, where the data consists of mostly multijet events ( $> 90\%$ ). The background multijet event is flatter in  $\Delta\eta_{JJ}$  distribution.

nal sample and data inclusive  $b$ -tag region  $\Delta\eta_{JJ}$  distribution. This cut is not entirely optimal for Scalar signals due to the different spin, yet it is fixed for both  $G_{KK}^*$  and Scalar selections.

#### 5.4 RESOLVED VETO

Sometimes one event can be reconstructed in both the resolved method as four small- $R$  jets and the boosted method as two large- $R$  jets with track jets. Figure 5.9 shows an example event display of collision data recorded in 2015.

In order to avoid events being reconstructed by both the resolved and the boosted analysis, events that pass the resolved signal region selections are vetoed in the boosted analysis. This is a political decision, and the effect of vetoing boosted event selection in the resolved analysis is never tested. The gain is a full statistical combination of the resolved and boosted result. For boosted analysis, it hurts the sensitivity up to  $1.5 \text{ TeV}$  for resonance signals. Hence it is necessary to introduce the resolved selection.



**Figure 5.9:** Event display of the same event using 5.9a resolved and 5.9b boosted topologies. The resolved reconstruction gives a  $m_{4j}$  of 873 GeV, and the boosted reconstruction gives  $m_{2j}$  of 852 GeV.

For resolved analysis, four small- $R$  jets with the highest  $b$ -tagging score are paired to construct two Higgs boson candidates. Each jet must have  $p_T > 40$  GeV,  $|\eta| < 2.5$ ,  $\text{MV2c10} > 0.8244$  (small- $R$  jet 70%  $b$ -tagging working point). Pairings of jets into Higgs boson candidates are only accepted if they satisfy the following requirements, where  $m_{4j}$  is expressed in GeV:

if  $m_{4j} < 1250$  GeV:

$$\frac{360 \text{ GeV}}{m_{4j}} - 0.5 < \Delta R_{jj}^{lead} < \frac{653 \text{ GeV}}{m_{4j}} + 0.475; \quad \frac{235 \text{ GeV}}{m_{4j}} < \Delta R_{jj}^{subl} < \frac{875 \text{ GeV}}{m_{4j}} + 0.35 \quad (5.1)$$

if  $m_{4j} > 1250$  GeV:

$$0 < \Delta R_{jj}^{lead} < 1; \quad 0 < \Delta R_{jj}^{subl} < 1 \quad (5.2)$$

In these expressions,  $\Delta R_{jj,lead}$  is the angular distance between jets in the leading Higgs boson candidate and  $\Delta R_{jj,subl}$  for the sub-leading candidate. The leading Higgs boson candidate is defined to be the candidate with the highest scalar sum of jet  $p_T$ . This requirement efficiently rejects jet-pairings

where one of the  $b$ -tagged jets is not consistent with that originating from a Higgs boson decay. The specific cut values in this and the following requirements were chosen to maximize the sensitivity to the signal.

Also, mass-dependent requirements are made on the leading Higgs boson candidate  $p_T$ , and the sub-leading Higgs boson  $p_T$ :

$$p_T^{\text{lead}} > 0.5m_{4j} - 105 \text{ GeV}, \quad p_T^{\text{subl}} > 0.33m_{4j} - 75 \text{ GeV} \quad (5.3)$$

where  $m_{4j}$  is again expressed in GeV.

A further ( $m_{4j}$ -independent) requirement is placed on the pseudorapidity difference between the two Higgs boson candidates,  $|\Delta\eta_{hh}| < 1.5$ , which rejects multijet events.

$$|\Delta\eta_{hh}| < 1.1 \quad \text{if } m_{4j} < 850 \text{ GeV}, \quad |\Delta\eta_{hh}| < 2 \times 10^{-3}m_{4j} - 0.6 \quad \text{if } m_{4j} > 850 \text{ GeV} \quad (5.4)$$

Events that have multiple Higgs boson candidates satisfying these requirements (which happens often when  $m_{4j} < 500$  GeV) necessitate an algorithm to choose the correct pairs. In the absence of energy loss through semi-leptonic decays, the optimal choice would be the combination most consistent with the decays of two particles of equal mass. To account for energy loss, the requirement of equal masses is modified. The distance,  $D_{hh}$ , of the pairing's leading and subleading Higgs boson candidate masses,  $(m_{2j}^{\text{lead}}, m_{2j}^{\text{subl}})$  from the line connecting  $(0 \text{ GeV}, 0 \text{ GeV})$  and  $(120 \text{ GeV}, 110 \text{ GeV})$  is computed, and the pairing with the smallest value of  $D_{hh}$  is chosen. The values of 120 GeV and 110 GeV are chosen because they correspond to the median values of the narrowest intervals that

contain 90% of the signal in simulations.  $D_{hh}$  can be expressed as follows:

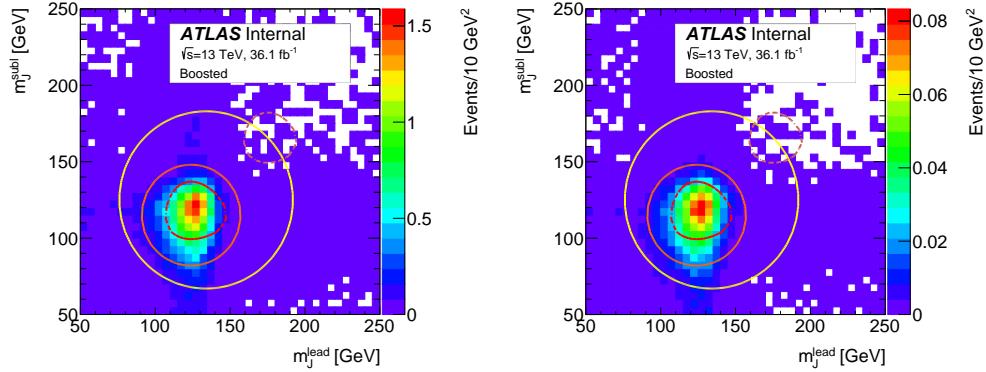
$$D_{hh} = \frac{\left| m_{2j}^{\text{lead}} - \frac{120}{\text{no}} m_{2j}^{\text{subl}} \right|}{\sqrt{1 + \left( \frac{\text{no}}{120} \right)^2}}. \quad (5.5)$$

A requirement on the Higgs boson candidates' masses is used to define the resolved signal region:

$$X_{hh-\text{resolved}} = \sqrt{\left( \frac{m_{2j}^{\text{lead}} - 120 \text{ GeV}}{0.1m_{2j}^{\text{lead}}} \right)^2 + \left( \frac{m_{2j}^{\text{subl}} - 110 \text{ GeV}}{0.1m_{2j}^{\text{subl}}} \right)^2} < 1.6, \quad (5.6)$$

where the  $0.1m_{2j}$  terms represent the widths of the leading and sub-leading Higgs boson candidate mass distributions, derived from simulation. The signal region is shown as the inner region of Figure ???. In summary, for any event can make a Higgs candidate through the Dhh minimization and passing through the resolved signal region  $X_{hh-\text{resolved}}$  cut, it is rejected in the boosted selection.

## 5.5 2D HIGGS MASS CUT



**Figure 5.10:** For RSG  $c = 1.0$  samples, number of events as a function of leading Higgs candidate mass and subleading Higgs candidate mass, for 1.2 TeV (left) signal and 2 TeV (right) signal samples. The red dotted line in the center correspond to the signal region, passing  $X_{hh} < 1.6$ .

To separate di-Higgs decays from background productions like QCD multi-jets and top, requirements on the leading and subleading large- $R$  jet masses are imposed. The signal region is defined using the expression 5.7:

$$X_{hh} = \sqrt{\left(\frac{m_j^{\text{lead}} - 124 \text{ GeV}}{0.1(m_j^{\text{lead}})}\right)^2 + \left(\frac{m_j^{\text{subl}} - 115 \text{ GeV}}{0.1(m_j^{\text{subl}})}\right)^2} \quad (5.7)$$

The denominator of each term in  $X_{hh}$  can be interpreted as a resolution on the reconstructed mass of 10% for the leading and subleading jets, hence  $X_{hh}$  can be interpreted as a  $\chi^2$  compatibility with the di-Higgs hypothesis. The subleading jet mass value of 115 GeV is chosen after investigating the signal jet masses in MC. The subleading large- $R$  jet typically has a reconstructed mass which is biased downward. This is partly due to the ordering of the large- $R$  jets in  $p_T$ , which makes the subleading jet towards lower energy. The energy losses from neutrinos in leptonic  $b$  decays, cracks in the calorimeter, and other effects also contributes. The signal region requires  $X_{hh} < 1.6$ . This cut gives nearly optimal performance. A more optimal signal region definition, using asymmetric signal jet mass resolution and a momentum dependent cut accounting for the higher mass resolution of larger  $p_T$  jets, can improve the overall sensitivity by 2 – 8%. Since the gain in sensitivity is small, the signal region is kept to be consistent with the  $X_{hh}-\text{resolved}$ . Figure 5.10 shows the  $G_{\text{KK}}^*$  MC  $m_{2j}^{\text{lead}}$ - $m_{2j}^{\text{subl}}$  2D distribution.

## 5.6 NUMBER OF $b$ -TAGGING REQUIREMENT

Passing basic object selection and  $X_{hh}$  cut, the signal region selection is defined by requiring multiple  $b$ -tags which are consistent with the di-Higgs decay. The presence of two  $b \rightarrow b\bar{b}$  decays in the final state naturally suggests requiring 4 track jets passing  $b$ -tagging requirements, and this is defined as the  $4b$  selection.

The  $4b$  requirement has an overall efficiency of roughly  $\varepsilon^4$ , where  $\varepsilon$  is the  $b$ -tagging efficiency chosen to be 70%. This means an overall  $0.7^4 \sim 0.24$  probability, but having one actual  $b$ -jet failing while the other three pass has probability  $3 \times 0.7^3 \times (1 - 0.7) \sim 0.31$ . Therefore, a  $3b$  selection is also introduced to recover the signal efficiency. An event with  $3b$ -tags must have at least  $3b$ -tagged track jets, but can have any number of additional un-tagged track jets. In  $4b$  and  $3b$ , each Higgs candidate can have at most two  $b$ -tagged track jets, hence  $\geq 3b$ -tagged trackjets cannot be in the same large- $R$  jet.

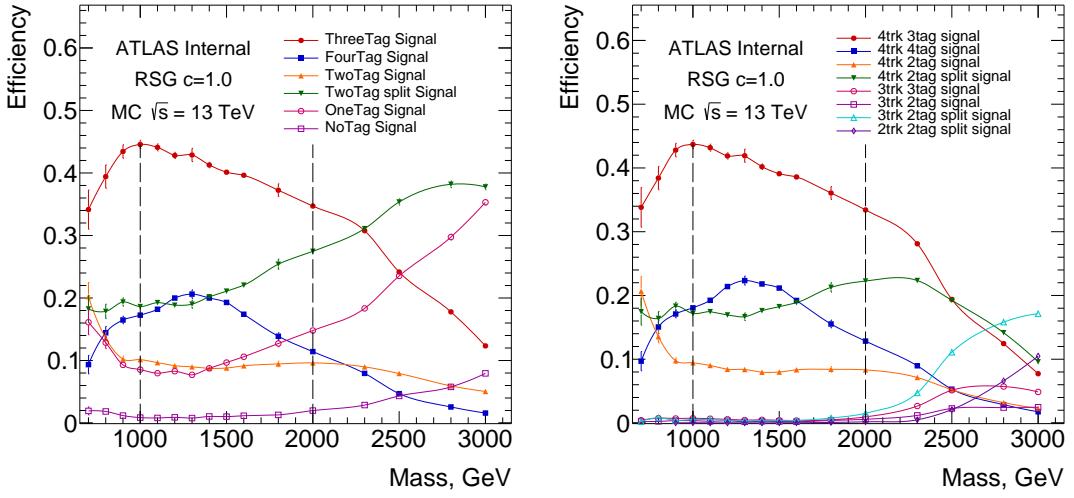
At the highest resonance mass, the Lorentz boost of the Higgs boson can be large enough to collimate the daughter  $b$ -quarks below the distance scale resolvable by the track jets ( $R = 0.2$ ). A third signal region is denoted by two-tag-split or simply  $2bs$ . It requires exactly one  $b$ -tagged track jet is found in each Higgs candidate, plus an arbitrary number of track jets that must fail the  $b$ -tag.

The other  $b$ -tagging situations are also sorted and studied, but not as signal regions due to their relatively small acceptance.  $2b$  region is defined as one large- $R$  jet has two  $b$ -tagged track jets, and the other large- $R$  jet has no  $b$ -tagged track jet.  $1b$  region is defined as one large- $R$  jet has one and only one  $b$ -tagged track jets, and the other large- $R$  jet has no  $b$ -tagged track jet.  $0b$  region is defined as both large- $R$  jets have no  $b$ -tagged track jet.

The MC events passing all signal region selections are sorted into different  $b$ -tagging categories. This is shown in Figure 5.11. For masses above 2.5 TeV, the  $2bs$  region (where each large- $R$  jet has exactly one  $b$ -tagged track jet) significantly improves the acceptance.

## 5.7 SIGNAL EFFICIENCY AND CUTFLOW

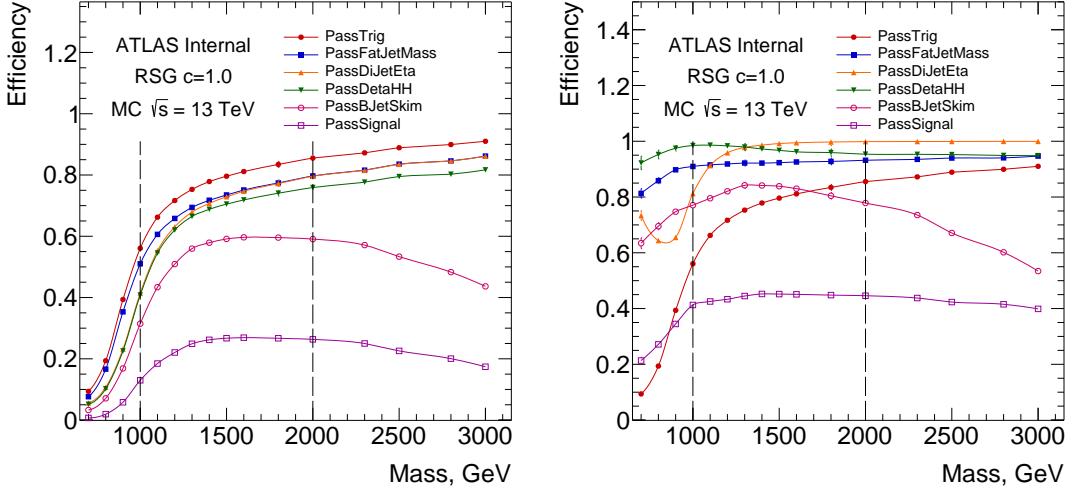
Acceptance means purely geometric fiducial volume of the detector. Efficiency refers to purely detector effectiveness in finding objects. The signal efficiency as a function of  $C_{KK}^*$  resonance mass



**Figure 5.11:** Signal fraction in different  $b$ -tag categories (left) and detailed fraction in different number of track jet and  $b$ -tag categories (right) as a function of signal resonance mass hypothesis for selection cuts. The efficiencies are relative to the total number of events passing the 2D mass cut.

is shown in Figure 5.12, both for the absolute signal efficiency and for the efficiency relative to the previous cut in the selection. Above a mass of  $\sim 1$  TeV, the reconstruction of high momentum large- $R$  jets with small  $\Delta\eta$  is efficient. Across the mass range considered, the signal jet masses requirement ( $X_{hh}$ ) and  $b$ -tagging requirements are  $\mathcal{O}(20\%)$  efficient relative to the previous cuts.

The selection efficiency at various stages for  $G_{KK}^*$  with  $c = 1.0$ ,  $G_{KK}^*$  with  $c = 2.0$ , and Heavy Scalar signal samples of all mass points can be found in Table 5.2, 5.3 and 5.4.



**Figure 5.12:** Absolute (left) and relative (right) signal efficiency as a function of RSG  $c=1.0$  signal resonance mass hypothesis for selection cuts. The relative efficiency is defined from the previous cut, where the order of cuts is given by the legend. PassTrig means the event passes the trigger selection; PassFatJetPt means the event passes the leading and sub-leading jet  $p_T$  cuts; PassDiJetEta means the event passes the leading and sub-leading jet  $\eta$  cuts; PassDataH means the events passes the  $|\Delta\eta| < 1.7$  cut; PassBJetSkim means the event contains at least two  $b$ -tagged track jets, inclusive of  $2b$ ,  $2bs$ ,  $3b$  and  $4b$  configurations; PassSignal means the event passes the signal region cut  $X_{hh} < 1.6$ .

**Table 5.2:** The selection efficiency for  $G_{KK}^* \rightarrow hh \rightarrow b\bar{b}b\bar{b}$  events ( $c = 1.0$ ) at each stage of the event selection. Uncertainties are the MC stat uncertainty only.

Resonance Mass [GeV]	Mini-ntuple Skimming	2 large-R jets	$\Delta\eta$	$X_{hh} < 1.6$	2bs SR	3b SR	4b SR
500	$317.31 \pm 6.0$	$295.75 \pm 5.79$	$164.5 \pm 4.32$	$8.45 \pm 0.99$	$1.08 \pm 0.37$	$2.14 \pm 0.52$	$0 \pm 0$
600	$269.07 \pm 3.64$	$247.94 \pm 3.5$	$136.31 \pm 2.59$	$11.31 \pm 0.76$	$2.57 \pm 0.37$	$3.84 \pm 0.45$	$0.66 \pm 0.19$
700	$253.68 \pm 3.35$	$226.93 \pm 3.16$	$124.83 \pm 2.35$	$16.79 \pm 0.86$	$3.74 \pm 0.42$	$6.99 \pm 0.56$	$1.91 \pm 0.29$
800	$286.26 \pm 2.28$	$245.36 \pm 2.11$	$129.2 \pm 1.53$	$24.41 \pm 0.67$	$5.11 \pm 0.31$	$11.27 \pm 0.46$	$4.13 \pm 0.27$
900	$306.51 \pm 1.61$	$275.57 \pm 1.52$	$158.03 \pm 1.15$	$40.72 \pm 0.59$	$8.81 \pm 0.28$	$19.76 \pm 0.41$	$7.5 \pm 0.25$
1000	$238.2 \pm 0.98$	$226.98 \pm 0.96$	$165.2 \pm 0.82$	$52.86 \pm 0.47$	$10.87 \pm 0.22$	$26.0 \pm 0.33$	$10.07 \pm 0.2$
1100	$164.5 \pm 0.63$	$160.94 \pm 0.63$	$132.53 \pm 0.57$	$45.26 \pm 0.34$	$9.55 \pm 0.16$	$21.88 \pm 0.23$	$9.03 \pm 0.14$
1200	$109.24 \pm 0.41$	$107.92 \pm 0.4$	$93.45 \pm 0.38$	$33.53 \pm 0.23$	$6.96 \pm 0.11$	$15.8 \pm 0.16$	$7.38 \pm 0.1$
1300	$72.72 \pm 0.59$	$72.2 \pm 0.59$	$63.74 \pm 0.56$	$24.19 \pm 0.35$	$5.02 \pm 0.17$	$11.33 \pm 0.24$	$5.45 \pm 0.16$
1400	$48.83 \pm 0.17$	$48.61 \pm 0.17$	$42.96 \pm 0.16$	$16.62 \pm 0.1$	$3.72 \pm 0.052$	$7.61 \pm 0.07$	$3.68 \pm 0.046$
1500	$33.13 \pm 0.12$	$33.02 \pm 0.12$	$29.25 \pm 0.11$	$11.31 \pm 0.07$	$2.67 \pm 0.036$	$5.08 \pm 0.047$	$2.44 \pm 0.031$
1600	$22.81 \pm 0.08$	$22.75 \pm 0.08$	$20.16 \pm 0.075$	$7.74 \pm 0.048$	$1.93 \pm 0.025$	$3.48 \pm 0.032$	$1.53 \pm 0.02$
1800	$11.2 \pm 0.1$	$11.18 \pm 0.1$	$9.93 \pm 0.094$	$3.71 \pm 0.059$	$1.1 \pm 0.034$	$1.6 \pm 0.038$	$0.6 \pm 0.022$
2000	$5.72 \pm 0.021$	$5.71 \pm 0.021$	$5.07 \pm 0.019$	$1.83 \pm 0.012$	$0.6 \pm 0.0072$	$0.76 \pm 0.0076$	$0.25 \pm 0.0041$
2250	$2.61 \pm 0.0088$	$2.61 \pm 0.0088$	$2.32 \pm 0.0083$	$0.78 \pm 0.005$	$0.31 \pm 0.0032$	$0.3 \pm 0.003$	$0.078 \pm 0.0014$
2500	$1.24 \pm 0.0054$	$1.24 \pm 0.0054$	$1.11 \pm 0.0051$	$0.33 \pm 0.0028$	$0.16 \pm 0.002$	$0.11 \pm 0.0016$	$0.021 \pm 0.00066$
2750	$0.6 \pm 0.0026$	$0.6 \pm 0.0026$	$0.54 \pm 0.0025$	$0.14 \pm 0.0013$	$0.081 \pm 0.00099$	$0.038 \pm 0.00065$	$0.0055 \pm 0.00024$
3000	$0.3 \pm 0.0011$	$0.3 \pm 0.0011$	$0.27 \pm 0.0011$	$0.058 \pm 0.00051$	$0.039 \pm 0.00041$	$0.013 \pm 0.00023$	$0.0016 \pm 8e-05$

**Table 5.3:** The selection efficiency for  $G_{KK}^* \rightarrow hh \rightarrow bbbb$  events ( $c = 2.0$ ) at each stage of the event selection. Uncertainties are the MC stat uncertainty only.

Resonance Mass [GeV]	Mini-ntuple Skimming	2 large-R jets	$\Delta\eta$	Xhh < 1.6	2bs SR	3b SR	4b SR
500	3705.15 ± 40.86	3479.44 ± 39.59	2325.18 ± 32.37	568.04 ± 16.17	122.56 ± 7.76	253.49 ± 10.78	100.7 ± 6.53
600	2549.14 ± 22.55	2374.01 ± 21.76	1591.92 ± 17.82	396.96 ± 9.03	89.01 ± 4.46	178.63 ± 6.05	74.31 ± 3.71
700	1928.4 ± 13.57	1782.85 ± 13.04	1183.86 ± 10.63	320.53 ± 5.62	71.38 ± 2.76	148.41 ± 3.81	59.31 ± 2.31
800	1595.14 ± 8.82	1457.71 ± 8.43	958.89 ± 6.84	268.75 ± 3.67	64.14 ± 1.86	123.48 ± 2.47	49.43 ± 1.51
900	1264.78 ± 5.77	1179.88 ± 5.58	819.75 ± 4.65	251.29 ± 2.61	55.63 ± 1.27	119.44 ± 1.79	48.72 ± 1.11
1000	891.0 ± 3.66	856.95 ± 3.59	662.54 ± 3.15	219.04 ± 1.84	49.45 ± 0.91	104.31 ± 1.26	42.97 ± 0.78
1100	595.58 ± 2.98	581.72 ± 2.95	481.67 ± 2.68	167.96 ± 1.61	37.64 ± 0.79	78.28 ± 1.09	34.59 ± 0.7
1200	390.84 ± 1.69	385.41 ± 1.68	330.23 ± 1.55	118.0 ± 0.94	26.23 ± 0.46	54.18 ± 0.64	25.34 ± 0.42
1300	257.66 ± 0.94	255.35 ± 0.94	222.37 ± 0.88	82.11 ± 0.54	19.04 ± 0.27	37.8 ± 0.37	16.99 ± 0.23
1400	172.09 ± 0.72	171.02 ± 0.71	150.22 ± 0.67	56.23 ± 0.42	13.6 ± 0.22	25.36 ± 0.28	11.78 ± 0.18
1500	116.25 ± 0.41	115.72 ± 0.41	101.92 ± 0.39	38.5 ± 0.24	9.94 ± 0.13	17.04 ± 0.16	7.64 ± 0.1
1600	80.09 ± 0.28	79.82 ± 0.28	70.48 ± 0.26	26.24 ± 0.16	7.01 ± 0.09	11.62 ± 0.11	4.92 ± 0.067
1800	38.99 ± 0.14	38.9 ± 0.14	34.39 ± 0.13	12.65 ± 0.081	3.82 ± 0.047	5.46 ± 0.053	2.02 ± 0.03
2000	19.94 ± 0.088	19.91 ± 0.088	17.68 ± 0.083	6.17 ± 0.05	2.15 ± 0.031	2.52 ± 0.032	0.85 ± 0.017
2250	9.02 ± 0.031	9.01 ± 0.031	7.99 ± 0.029	2.62 ± 0.017	1.03 ± 0.011	1.02 ± 0.01	0.28 ± 0.0051
2500	4.28 ± 0.016	4.28 ± 0.016	3.8 ± 0.015	1.13 ± 0.0083	0.52 ± 0.0058	0.4 ± 0.0048	0.098 ± 0.0022
3000	1.07 ± 0.004	1.07 ± 0.004	0.96 ± 0.0038	0.23 ± 0.0019	0.13 ± 0.0015	0.062 ± 0.00097	0.013 ± 0.00043

**Table 5.4:** The selection efficiency for  $H \rightarrow hh \rightarrow b\bar{b}b\bar{b}$  events at each stage of the event selection.

Resonance Mass [GeV]	Mini-ntuple Skimming	2 large-R jets	$\Delta\eta$	Xhh < 1.6	2bs SR	3b SR	4b SR
500	1557.94 ± 136.12	1022.77 ± 110.29	95.14 ± 33.64	11.69 ± 11.69	0 ± 0	11.69 ± 11.69	0 ± 0
600	3289.78 ± 123.99	2542.11 ± 108.99	485.99 ± 47.66	54.55 ± 15.77	18.73 ± 9.39	9.17 ± 6.49	0 ± 0
700	4655.21 ± 94.59	3855.64 ± 86.09	1237.8 ± 48.78	142.42 ± 17.03	28.55 ± 7.74	52.6 ± 10.75	7.69 ± 3.85
800	7506.31 ± 81.79	6020.56 ± 73.25	2150.64 ± 43.78	320.63 ± 17.02	67.63 ± 7.83	139.97 ± 11.23	47.57 ± 6.75
900	9732.89 ± 61.17	8400.91 ± 56.83	3574.63 ± 37.07	806.13 ± 17.76	188.71 ± 8.71	377.92 ± 12.12	127.7 ± 6.99
1000	7516.07 ± 37.72	7033.18 ± 36.49	4496.85 ± 29.18	1351.1 ± 16.2	303.71 ± 7.88	650.89 ± 11.19	234.2 ± 6.57
1100	4731.39 ± 21.54	4563.4 ± 21.15	3485.58 ± 18.49	1135.18 ± 10.7	251.77 ± 5.18	539.51 ± 7.36	215.39 ± 4.5
1200	2853.51 ± 12.23	2782.42 ± 12.07	2253.95 ± 10.87	786.92 ± 6.53	175.53 ± 3.21	366.01 ± 4.44	158.29 ± 2.8
1300	1700.83 ± 7.01	1668.05 ± 6.94	1362.91 ± 6.27	494.36 ± 3.84	107.0 ± 1.86	224.19 ± 2.58	107.55 ± 1.7
1400	1016.14 ± 4.03	999.98 ± 4.0	802.44 ± 3.59	296.46 ± 2.22	65.49 ± 1.1	133.86 ± 1.49	65.58 ± 0.99
1500	621.34 ± 2.41	613.46 ± 2.39	484.92 ± 2.13	179.29 ± 1.32	42.75 ± 0.68	79.32 ± 0.88	36.77 ± 0.56
1600	386.3 ± 1.46	382.44 ± 1.46	297.63 ± 1.28	109.99 ± 0.8	27.68 ± 0.42	49.3 ± 0.53	20.92 ± 0.32
1800	154.8 ± 0.58	153.5 ± 0.57	116.52 ± 0.5	42.24 ± 0.31	12.41 ± 0.18	18.2 ± 0.2	6.66 ± 0.11
2000	65.4 ± 0.24	65.02 ± 0.24	48.57 ± 0.2	16.89 ± 0.12	5.64 ± 0.076	7.01 ± 0.079	2.18 ± 0.041
2250	23.84 ± 0.085	23.73 ± 0.085	17.44 ± 0.073	5.57 ± 0.042	2.25 ± 0.028	2.11 ± 0.025	0.52 ± 0.012
2500	9.2 ± 0.032	9.17 ± 0.032	6.72 ± 0.028	1.9 ± 0.015	0.92 ± 0.011	0.64 ± 0.0086	0.11 ± 0.0034
2750	3.73 ± 0.013	3.73 ± 0.013	2.71 ± 0.011	0.63 ± 0.0054	0.37 ± 0.0042	0.17 ± 0.0027	0.021 ± 0.00093
3000	1.59 ± 0.0054	1.59 ± 0.0054	1.15 ± 0.0046	0.22 ± 0.0021	0.15 ± 0.0017	0.044 ± 0.0009	0.0038 ± 0.00025

*Even in the darkest night. Stars and angels still shine  
bright.*

Bear

# 6

## Background Estimation

### 6.0.1 OVERVIEW

The primary backgrounds to in  $4b$  signals region are QCD multi-jet production ( $\sim 95\%$ ),  $t\bar{t}$  ( $\sim 5\%$ ), and  $Z+jets$  ( $< 1\%$ ), where the percentages are the expected fraction of the background. In the  $3b$  signal region, the fractions are QCD ( $\sim 90\%$ ),  $t\bar{t}$  ( $\sim 10\%$ ), and  $Z+jets$  ( $< 1\%$ ). In the two  $b$ -jet split signal region, the fractions are QCD ( $\sim 80\%$ ),  $t\bar{t}$  ( $\sim 20\%$ ), and  $Z+jets$  ( $< 1\%$ ).

QCD is by far the dominant background. However, there is no reliable, high-statistics Monte Carlo simulation sample in this region of phase space (i.e with three or four  $b$ -jets collected into two high- $p_T$  large radius jets) and thus a data-driven background estimation is needed. (See Appendix ??.) For the  $t\bar{t}$ background, Monte Carlo simulation samples of reasonable size are available, and thus

can be used to guide an estimation of this background. The  $Z$ +jets background is small enough that we will rely on the Monte Carlo simulation of  $Z$ +heavy flavor jets.  $ZZ \rightarrow b\bar{b}b\bar{b}$  has been estimated to be completely negligible using a particle-level analysis, with less than one event expected after three  $b$ -tags are required, which will be further heavily suppressed by the  $X_{bb}$  requirement.

The QCD background prediction relies on finding a region which is similar enough in event properties that it can be used to estimate the shapes of the expected QCD background. This region is identical to the signal region defined by the full selection, with the exception that the events must have less  $b$ -tagged track jets:

- For the  $2bs$  category, the  $1b$  sample is used for modeling.
- For the  $3b$  and  $4b$  categories, the  $2b$  sample - where the two  $b$ -tagged trackjet are in the same large- $R$  jet - is used for modeling.

To prevent differences in the number of track jets from biasing the dijet mass distribution, the  $1b$ -tagged region requires that each large- $R$  jet has at least one track jet (to model  $2bs$ , ie.  $2b$  tag split). Similarly, the  $2b$ -tagged region requires that one large- $R$  jet has at least one track jet and the other one has at least two track jets (to model  $3b$ ), and each large- $R$  jet has at least two track jets (to model  $4b$ ).

However, this less  $b$ -tagged region only supplies the shapes of the expected background and not the total yield, and a second control sample, which we denote the *Sideband* region, is used to estimate the yield. The Sideband is obtained by doing the full analysis selection, except instead of the  $X_{bb}$  cut an alternative criteria on the large radius jet masses is used, that  $33 < R_{bb}$  and  $R_{bb}^{\text{high}} < 58 \text{ GeV}$ . To validate this approach, a third region, which we denote the *Control* region, is centered on the signal region in the plane of the two large radius jet masses but does not include the signal region, such that  $R_{bb} < 33 \text{ GeV}$ . The control region is used to validate the background estimations before unblinding. The control and sideband regions are optimized, as shown in the following sections, to accurately estimate the rate of the QCD background (and thus allow for an extrapolation

from the  $1b/2b$  estimate to a prediction in the  $4b/3b/2bs$  signal regions), whilst giving a control region which has kinematic properties similar to that of the signal region.

The  $t\bar{t}$  background shape is taken from MC. A data-based estimation of the  $t\bar{t}$  background yield is performed simultaneously with the QCD background yield estimation, by means of a binned likelihood fit. In the plane of the two leading large radius jet masses, the main contribution of the  $t\bar{t}$  background lies in the Sideband region. The data distribution in the Sideband region of the leading- $p_T$  large radius jet mass is fit simultaneously with the QCD shape estimate (from the less  $b$ -tagged sample) and with the  $t\bar{t}$  Monte Carlo shape. This fit is done separately in the  $4b$ ,  $3b$ , and  $2bs$  Sideband regions. From this fit, two terms are determined simultaneously:  $\mu_{QCD}$  and  $\alpha_{tt}$ .  $\mu_{QCD}$  is the ratio of the QCD event yield in the  $2bs/3b/4b$  regions to the amount in each corresponding less  $b$ -tagged region.  $\alpha_{tt}$  is the ratio of the fitted ttbar event yield to the yield predicted from ttbar MC. These two numbers are then used as multiplicative constants in other regions of the mass plane (i.e. the Control or Signal regions) to extrapolate from the rates of the less  $b$ -tagged regions to predictions of rates in the  $2bs/3/4$   $b$ -tagged regions, for estimating the amount of QCD, and to correct the rate of  $t\bar{t}$  production wrt. MC. Hence, the underlying assumption is that these scale factors are roughly constant over the 2D large-R jet mass plane for Sideband/Control/Signal regions, which has been verified by performing these fits in small bins across the 2D mass plane. This is shown in Appendix ???. The correction factors are derived separately for the  $4b$ ,  $3b$ , and  $2bs$  regions.

In this section, we describe this approach in more detail and show its validation in data.

## 6.1 BACKGROUND ESTIMATION

A similar circular variable can be defined in the two-dimensional mass plane,  $R_{bb}$ . The circular region  $R_{bb}$  has the same central values as  $X_{bb}$ , but without resolution terms in the denominators and is defined as:

$$R_{bb} = \sqrt{(m_J^{\text{lead}} - 124 \text{ GeV})^2 + (m_J^{\text{subl}} - 115 \text{ GeV})^2} \quad (6.1)$$

The region defined by  $1.6 < X_{bb}$  and  $R_{bb} < 33$  is the control region. It will be discussed in Section ???. The cut value was optimized to allow for a reasonable sized sample (twice the statistics as the signal region) in the control region with kinematics similar to the signal region, and avoiding the large contributions of the  $t\bar{t}$  sample when the large- $R$  jets have a mass near the top quark mass (with  $m_J > 160 \text{ GeV}$ ).

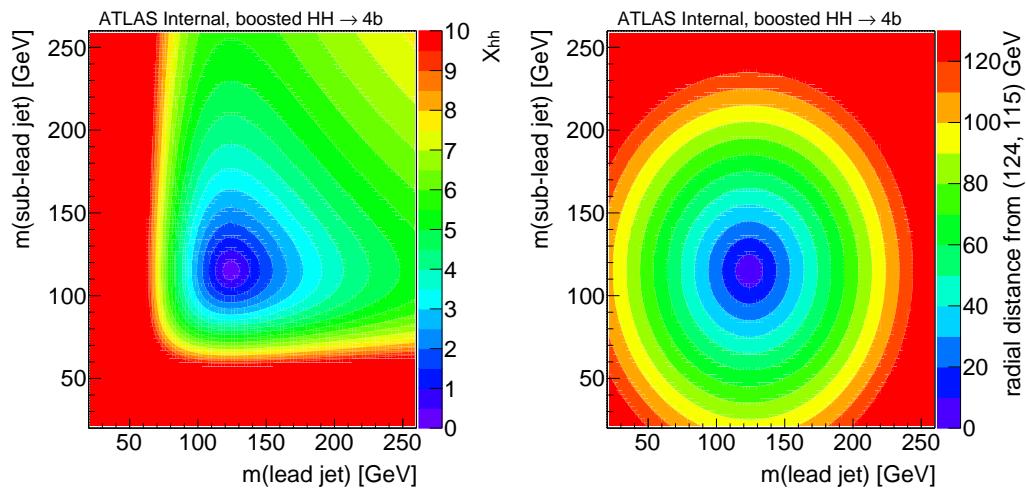
Similarly,  $R_{bb}^{\text{high}}$ , the circular region that has the shifted central values up by 10  $\text{GeV}$  is defined using the variable:

$$R_{bb}^{\text{high}} = \sqrt{(m_J^{\text{lead}} - 134 \text{ GeV})^2 + (m_J^{\text{subl}} - 125 \text{ GeV})^2} \quad (6.2)$$

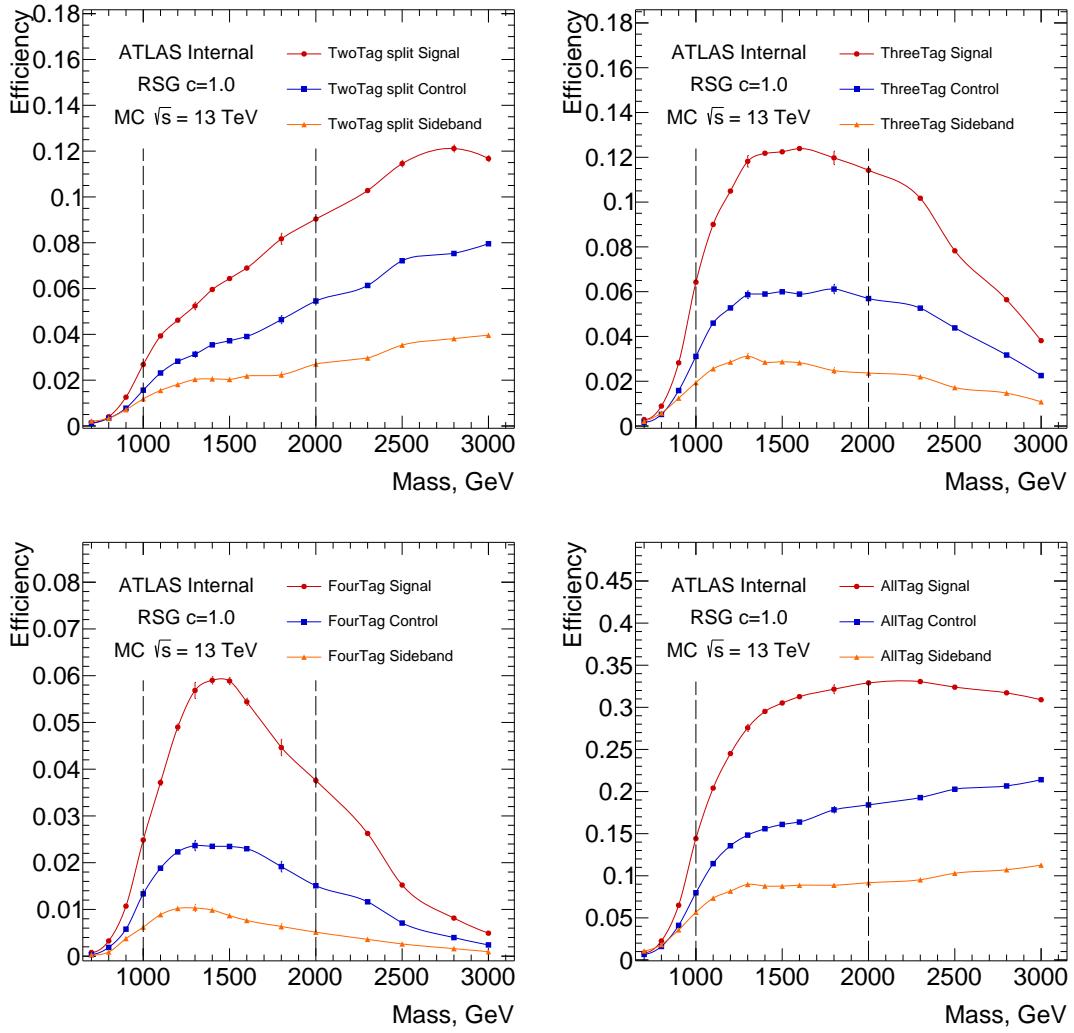
In Run-1, and in the the 2015 analysis, the sideband was defined to be all events not in the signal or control regions. However, the kinematics of events with very large and very small large- $R$  jet masses may not be the same as those within the signal region. To avoid biasing effects from extremely low mass or extremely high mass large- $R$  jets, the sideband region is also redesigned to be like the control region, but at  $R_{bb} > 33$  and  $R_{bb}^{\text{high}} < 58$ . The shift upwards helps to capture enough  $t\bar{t}$  events in the normalization estimates, as described in Section ??.

The values of the  $X_{bb}$  and  $R_{bb}$  variables can be seen graphically in Figure 6.1.

The number of events in the control region and sideband region as a function of Resonance mass is shown in Figure 6.2. For  $2bs$ ,  $3b$  and  $4b$ , each region has the number of events decrease from signal region to control region to sideband regions.



**Figure 6.1:** Values of the  $X_{hh}$  and  $R_{hh}$  variables, which are shapes in the two-dimensional plane of the large-R jet masses used to defined signal, control, and sideband regions. For both variables, a smaller value indicates the jets are closer to the Higgs mass.



**Figure 6.2:** Detailed signal efficiency in different signal/control/sideband regions as in  $2b$ s (top left,  $3b$  (top right),  $4b$  (bottom left) and inclusive b-tagged regions, which include  $2b$ ,  $1b$  and  $0b$  as well, (bottom right) as a function of signal resonance mass hypothesis for selection cuts. The efficiencies are relative to the total number of events in the preselection.

*It is a far, far better thing that I do, than I have ever  
done; it is a far, far better rest that I go to than I have  
ever known.*

Charles Dickens

# 7

## Conclusion

Di-Higgs search has a short history, but will have a long future. This thesis presents a search for both resonant and non-resonant production of pairs of Standard Model Higgs bosons has been carried out in the dominant  $b\bar{b}b\bar{b}$  channel, using 27.5–36.1  $\text{fb}^{-1}$  of LHC  $pp$  collision data at  $\sqrt{s} = 13$  TeV collected by ATLAS in 2015 and 2016. The search sensitivity of this analysis exceeds that of the previous analysis of the  $\sqrt{s} = 13$  TeV 2015 dataset<sup>3</sup> for non-resonant signal and also across the entire mass range of 260–3000 GeV for the resonance search, with significantly improvement in the high mass resonance sensitivities. The resolved analysis has each  $h \rightarrow b\bar{b}$  reconstructed as two separate  $b$ -tagged jets, and the boosted analysis has each  $h \rightarrow b\bar{b}$  reconstructed as a single large-radius jet associated with at least one small-radius  $b$ -tagged track-jet. The estimated background consists mainly of multi-jet and  $t\bar{t}$  events.

No significant excess is observed in the data. Upper limits on the production cross section times branching ratio to the  $b\bar{b}b\bar{b}$  final state are set for a narrow-width spin-0 scalar and for wider spin-2 resonances. The bulk RS model with  $k/\bar{M}_{\text{Pl}} = 1$  is excluded for masses between 313 and 1362 GeV, and the bulk RS model with  $k/\bar{M}_{\text{Pl}} = 2$  is excluded for masses below 1744 GeV. The 95% CL upper limit on the non-resonant production is 147 fb, which corresponds to 13.0 times the SM expectation.

This result confirms the great success of the Standard Model. The Higgs potential shape couldn't be very different from the SM predictions at TeV energy scale. Without any significant excess, the phase space for Beyond the Standard Model physics is further constrained.

Improvement with future Run 2 data could come  $b$ -tagging, especially efficiency increase in high  $p_T$  region. Other aspects include advanced trigger technologies, which can increase the signal event rate, and improved jet energy and mass resolution, which can increase the purity in selection. Together with the larger dataset, it is possible to double the current resonance search sensitivity. For non-resonance search, combined  $|\frac{\lambda}{\lambda_{\text{SM}}}| \sim 10$  is possible at the end of Run 2 in 2020.

For longer perspectives, di-Higgs searches and measurements will continue to be one of the most important analysis. It can constrain or hint the physics Beyond the Standard Model. In 2030 to 2040, the High Luminosity LHC will be able to constrain  $|\frac{\lambda}{\lambda_{\text{SM}}}| \sim 1$  with all the different channels from both ATLAS and CMS combined.

For even longer future developments in high energy experiments, all aspects—accelerator, detector, computation and theory—must advance together to answer the questions the Standard Model cannot answer, or find questions the Standard Model didn't ask<sup>70,71</sup>. Life is short for mankind, and the understanding of the universe is an endless journey. I am deeply honored to be a small part of this odyssey.

# References

- [1] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016. doi: 10.1088/1674-1137/40/10/100001.
- [2] W.J. Stirling.  $7/8$  and  $13/8$  TeV LHC luminosity ratios. 2013. URL [http://www.hep.physics.cern.ac.uk/~wstirlin/plots/lhclumi7813\\_2013\\_v0.pdf](http://www.hep.physics.cern.ac.uk/~wstirlin/plots/lhclumi7813_2013_v0.pdf).
- [3] Lyndon Evans. The Large Hadron Collider. *Annual Review of Nuclear and Particle Science*, 61(1):435–466, 2011. doi: 10.1146/annurev-nucl-102010-130438.
- [4] ATLAS Collaboration. ATLAS Luminosity Public Results, Run 2. 2015. URL <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [5] ATLAS Collaboration. Jet mass reconstruction with the ATLAS Detector in early Run 2 data. ATLAS-CONF-2016-035, 2016. URL <https://cds.cern.ch/record/2200211>.
- [6] ATLAS Collaboration. Optimisation of the ATLAS  $b$ -tagging performance for the 2016 LHC Run. *ATL-PHYS-PUB-2016-012*, (ATL-PHYS-PUB-2016-012), Jun 2016. URL <https://cds.cern.ch/record/2160731>.
- [7] Kaustubh Agashe, Hooman Davoudiasl, Gilad Perez, and Amarjit Soni. Warped gravitons at the CERN LHC and beyond. *Phys. Rev. D*, 76:036006, 2007. doi: 10.1103/PhysRevD.76.036006.
- [8] Liam Fitzpatrick, Jared Kaplan, Lisa Randall, and Lian-Tao Wang. Searching for the Kaluza-Klein graviton in bulk RS models. *JHEP*, 09:013, 2007. doi: 10.1088/1126-6708/2007/09/013.
- [9] T. D. Lee. A theory of spontaneous  $t$  violation. *Phys. Rev. D*, 8:1226–1239, Aug 1973. doi: 10.1103/PhysRevD.8.1226.
- [10] G.C. Branco et al. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1, 2012. doi: 10.1016/j.physrep.2012.02.002.
- [11] Graham D. Kribs and Adam Martin. Enhanced di-higgs production through light colored scalars. *Phys. Rev. D*, 86:095023, 2012. doi: 10.1103/PhysRevD.86.095023.

- [12] R. Gröber and M. Mühlleitner. Composite Higgs boson pair production at the LHC. *JHEP*, 06:020, 2011. doi: [10.1007/JHEP06\(2011\)020](https://doi.org/10.1007/JHEP06(2011)020).
- [13] Roberto Contino et al. Anomalous couplings in double Higgs production. *JHEP*, 08:154, 2012. doi: [10.1007/JHEP08\(2012\)154](https://doi.org/10.1007/JHEP08(2012)154).
- [14] ATLAS Collaboration. Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Rev. D*, 94: 052002, 2016. doi: [10.1103/PhysRevD.94.052002](https://doi.org/10.1103/PhysRevD.94.052002).
- [15] ATLAS Collaboration. Search for Higgs boson pair production in the  $b\bar{b}b\bar{b}$  final state from pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *Eur. Phys. J. C*, 75:412, 2015. doi: [10.1140/epjc/s10052-015-3628-x](https://doi.org/10.1140/epjc/s10052-015-3628-x).
- [16] ATLAS Collaboration. Search for Higgs Boson Pair Production in the  $\gamma\gamma b\bar{b}$  Final State Using  $pp$  Collision Data at  $\sqrt{s} = 8$  TeV from the ATLAS Detector. *Phys. Rev. Lett.*, 114: 081802, 2015. doi: [10.1103/PhysRevLett.114.081802](https://doi.org/10.1103/PhysRevLett.114.081802).
- [17] David Griffiths. *Introduction to elementary particles*. 2008.
- [18] Christopher G. Tully. *Elementary particle physics in a nutshell*. 2011.
- [19] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2014. ISBN 1107034736, 9781107034730.
- [20] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012. doi: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
- [21] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012. doi: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [22] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli, E. Vryonidou, and M. Zaro. Higgs pair production at the LHC with NLO and parton-shower effects. *Phys. Lett.*, B732:142–149, 2014. doi: [10.1016/j.physletb.2014.03.026](https://doi.org/10.1016/j.physletb.2014.03.026).
- [23] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016. doi: [10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002).

- [24] Graham D. Kribs, Andreas Maier, Heidi Rzezak, Michael Spannowsky, and Philip Waite. Electroweak oblique parameters as a probe of the trilinear Higgs boson self-interaction. *Phys. Rev.*, D95(9):093004, 2017. doi: 10.1103/PhysRevD.95.093004.
- [25] Viviana Cavaliere and Gabriel Facini. Summary of Limits on BSM Models from Diboson Searches. Technical Report ATL-COM-PHYS-2016-1071, CERN, Geneva, Sep 2016. URL <https://cds.cern.ch/record/2203605>.
- [26] Georges Aad et al. Searches for Higgs boson pair production in the  $hh \rightarrow bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$  channels with the ATLAS detector. *Phys. Rev.*, D92:092004, 2015. doi: 10.1103/PhysRevD.92.092004.
- [27] Lyndon R Evans and Philip Bryant. LHC Machine. *J. Instrum.*, 3:So8001. 164 p, 2008. URL <https://cds.cern.ch/record/1129806>. This report is an abridged version of the LHC Design Report (CERN-2004-003).
- [28] ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider. *JINST*, 3:So8003, 2008. doi: 10.1088/1748-0221/3/08/So8003.
- [29] CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):So8004, 2008. URL <http://stacks.iop.org/1748-0221/3/i=08/a=So8004>.
- [30] LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 3:So8005, 2008. doi: 10.1088/1748-0221/3/08/So8005.
- [31] ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):So8002, 2008. URL <http://stacks.iop.org/1748-0221/3/i=08/a=So8002>.
- [32] ATLAS Collaboration. Luminosity Determination in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV Using the ATLAS Detector at the LHC. *Eur. Phys. J.*, C71:1630, 2011. doi: 10.1140/epjc/s10052-011-1630-5.
- [33] Paul Collier for the LHC team. LHC Machine Status. CERN Resource Review Board, 2015. URL <https://cds.cern.ch/record/2063924/files/CERN-RRB-2015-119.PDF>.
- [34] Giulia Papotti for the LHC team. LHC Machine Status Report. CERN Resource Review Board, 2016. URL [https://indico.cern.ch/event/563488/contributions/2277292/attachments/1340292/2019570/20160921\\_LHCC.pdf](https://indico.cern.ch/event/563488/contributions/2277292/attachments/1340292/2019570/20160921_LHCC.pdf).

- [35] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi: 10.1088/1748-0221/3/08/S08003.
- [36] ATLAS Collaboration. ATLAS Insertable B-Layer Technical Design Report. CERN-LHCC-2010-013, ATLAS-TDR-19, Sep 2010. URL <https://cds.cern.ch/record/1291633>.
- [37] ATLAS Collaboration. Expected performance of the ATLAS  $b$ -tagging algorithms in Run-2. ATL-PHYS-PUB-2015-022, 2015. URL <https://cds.cern.ch/record/2037697>.
- [38] Morad Aaboud et al. Study of the material of the ATLAS inner detector for Run 2 of the LHC. *JINST*, 12(12):P12009, 2017. doi: 10.1088/1748-0221/12/12/P12009.
- [39] ATLAS Collaboration. Performance of the ATLAS Trigger System in 2015. *Eur. Phys. J. C*, 77:317, 2017. doi: 10.1140/epjc/s10052-017-4852-3.
- [40] M. Aaboud et al. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2. *Eur. Phys. J.*, C77(10):673, 2017. doi: 10.1140/epjc/s10052-017-5225-7.
- [41] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77:490, 2017. doi: 10.1140/epjc/s10052-017-5004-5.
- [42] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The catchment area of jets. *JHEP*, 04:005, 2008. doi: 10.1088/JHEP04(2008)05.
- [43] ATLAS Collaboration. Jet energy measurement with the ATLAS detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV. *Eur. Phys. J. C*, 73:2304, 2013. doi: 10.1140/epjc/s10052-013-2304-2.
- [44] ATLAS Collaboration. Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at  $\sqrt{s} = 13$  TeV. ATL-PHYS-PUB-2015-015, 2015. URL <https://cds.cern.ch/record/2037613>.
- [45] D. Krohn, J. Thaler, and L.-T. Wang. Jet trimming. *JHEP*, 02:084, 2010. doi: 10.1007/JHEP02(2010)084.
- [46] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160, 1993. doi: 10.1103/PhysRevD.48.3160.

- [47] ATLAS Collaboration. Performance of jet substructure techniques for large- $R$  jets in proton–proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector. *JHEP*, 09:076, 2013. doi: 10.1007/JHEP09(2013)076.
- [48] ATLAS Collaboration. Performance of  $b$ -jet identification in the ATLAS experiment. *JINST*, 11:P04008, 2016. doi: 10.1088/1748-0221/11/04/P04008.
- [49] Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data. Technical Report ATLAS-CONF-2016-024, CERN, Geneva, Jun 2016. URL <http://cds.cern.ch/record/2157687>.
- [50] Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using  $p\bar{p}$  collisions at  $\sqrt{s} = 13$  TeV. Technical Report ATLAS-CONF-2017-029, CERN, Geneva, May 2017. URL <http://cds.cern.ch/record/2261772>.
- [51] Georges Aad et al. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV. *Eur. Phys. J.*, C76(5):292, 2016. doi: 10.1140/epjc/s10052-016-4120-y.
- [52] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006. doi: 10.1088/1126-6708/2006/05/026.
- [53] ATLAS Collaboration. Summary of ATLAS Pythia 8 tunes. 2012. URL <http://cdsweb.cern.ch/record/1474107>.
- [54] S. Agostinelli et al. GEANT4: a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003. doi: 10.1016/S0168-9002(03)01368-8.
- [55] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823–874, 2010. doi: 10.1140/epjc/s10052-010-1429-9.
- [56] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010. doi: 10.1007/JHEP06(2010)043.
- [57] Peter Zeiler Skands. Tuning Monte Carlo generators: The Perugia tunes. *Phys. Rev. D*, 82:074018, 2010. doi: 10.1103/PhysRevD.82.074018.

- [58] Michał Czakon and Alexander Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Comput. Phys. Commun.*, 185:2930, 2014. ISSN 0010-4655. doi: [10.1016/j.cpc.2014.06.021](https://doi.org/10.1016/j.cpc.2014.06.021).
- [59] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [60] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244, 2013. doi: [10.1016/j.nuclphysb.2012.10.003](https://doi.org/10.1016/j.nuclphysb.2012.10.003).
- [61] A. Carvalho. Gravity particles from Warped Extra Dimensions, predictions for LHC. 2017.
- [62] M. Bahr et al. Herwig++ physics and manual. *Eur. Phys. J. C*, 58:639–707, 2008. doi: [10.1140/epjc/s10052-008-0798-9](https://doi.org/10.1140/epjc/s10052-008-0798-9).
- [63] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010. doi: [10.1103/PhysRevD.82.074024](https://doi.org/10.1103/PhysRevD.82.074024).
- [64] Pavel M. Nadolsky et al. Implications of cteq global analysis for collider observables. *Phys. Rev. D*, 78:013004, 2008. doi: [10.1103/PhysRevD.78.013004](https://doi.org/10.1103/PhysRevD.78.013004).
- [65] Michael H. Seymour and Andrzej Siodmok. Constraining MPI models using  $\sigma_{\text{eff}}$  and recent Tevatron and LHC Underlying Event data. *JHEP*, 10:113, 2013. doi: [10.1007/JHEP10\(2013\)113](https://doi.org/10.1007/JHEP10(2013)113).
- [66] S. Dawson, S. Dittmaier, and M. Spira. Neutral Higgs boson pair production at hadron colliders: QCD corrections. *Phys. Rev. D*, 58:115012, 1998. doi: [10.1103/PhysRevD.58.115012](https://doi.org/10.1103/PhysRevD.58.115012).
- [67] T. Plehn, M. Spira, and P.M. Zerwas. Pair production of neutral Higgs particles in gluon-gluon collisions. *Nucl. Phys. B*, 479:46, 1996. doi: [10.1016/0550-3213\(96\)00418-X](https://doi.org/10.1016/0550-3213(96)00418-X). [Erratum: *Nucl. Phys. B* 531 (1998) 655].
- [68] S. Borowka, N. Greiner, G. Heinrich, S. P. Jones, M. Kerner, J. Schlenk, U. Schubert, and T. Zirke. Higgs boson pair production in gluon fusion at NLO with full top-quark mass dependence. *Phys. Rev. Lett.*, 117(1):012001, 2016. doi: [10.1103/PhysRevLett.117.012001](https://doi.org/10.1103/PhysRevLett.117.012001).
- [69] S. Borowka, N. Greiner, G. Heinrich, S. P. Jones, M. Kerner, J. Schlenk, and T. Zirke. Full top quark mass dependence in Higgs boson pair production at NLO. *JHEP*, 10:107, 2016. doi: [10.1007/JHEP10\(2016\)107](https://doi.org/10.1007/JHEP10(2016)107).

- [70] Burton Richter. High Energy Colliding Beams; What Is Their Future? *Rev. Accel. Sci. Tech.*, 7:1–8, 2014. doi: 10.1142/9789814651493\_0001,10.1142/S1793626814300011.
- [71] Stephen Hawking and Gordon Kane. Should China build the Great Collider? 2018.