Master's Thesis

Background estimation methods in the split-boosted 1-lepton $X \to SH \to b\bar{b}WW$ analysis for Run 3 at ATLAS

prepared by

Lena Schulz

from Wolfenbüttel

at the II. Physikalischen Institut

Thesis number:	II.Physik-UniGö-MSc-2024/12
Thesis period:	29th May 2024 until 16th October 2024
First referee:	Prof. Dr. Stanley Lai
Second referee:	apl. Prof. Dr. Jörn Große-Knetter

Abstract

The known incompleteness of the Standard Model gives rise to many theories of physics beyond the Standard Model. Many of those theories predict the existence of additional particles, motivating the search for new signatures.
This thesis will focus on the resonant production of a Higgs boson in conjunction with an additional scalar particle S in the $X \rightarrow HS \rightarrow b\bar{b}WW$ decay channel with one lepton in the final state. Depending on the masses of the resonance X and the particle S, multiple possible topologies arise for this decay channel. The topic of this thesis is the split-boosted topology, which is characterized by a sufficiently large separation between the two W bosons, allowing for the resolution of the final-state lepton from the jet of the hadronically decaying W boson, whilst the hadronic decay products of the Higgs boson and the W boson are too collimated to be resolved.
For the analysis of this decay channel, a lepton study is performed to determine the optimal choice of particle identity (ID) and isolation working points. It is shown that for electrons, the tight ID in combination with the `PFlowTight_VarRad` isolation give the best signal efficiency while rejecting a majority of the non-prompt background. For muons, the medium ID in combination with the `Tight_VarRad` is the best performing working point. This selection is then used in the definition of signal leptons of the analysis.
Furthermore, the non-prompt control region for the split-boosted analysis is defined and unblinded. Comparing the modelling to the data in the control region, it can be seen that the shape of the distributions is modelled sufficiently by the Monte Carlo simulation, while the absolute values feature an offset that needs to corrected in the fitting process later in the analysis. To assess systematic uncertainties, the closure uncertainty of the modelling is calculated. The Monte Carlo modelling provides a sufficient understanding of the non-prompt background, making a data-driven approach unnecessary.

Contents

1. Introduction

The Standard Model (SM) of particle physics [1] is one of the great achievements of modern physics. Encompassing the fundamental particles and the forces that govern their interactions, except for gravity, the SM offers a comprehensive framework to describe and understand the building blocks of nature. Developed in the latter half of the 20th century, the SM has undergone extensive experimental testing. The results of these tests align with the predictions of the SM with remarkable precision. One of the most recent breakthrough discoveries was the discovery of the Higgs boson at the LHC in 2012 [2, 3]. The SM combines the electromagnetic, strong and weak interaction into one quantum field theory. The only force not explained in the SM is gravity, which is described by general relativity [4]. This limitation means that the SM, as it stands, is incomplete and cannot fully describe the universe's fundamental interactions. Beyond gravity, there are several other phenomena not accounted for by the SM, such as neutrino masses, dark matter, and more [5, 6].
This known incompleteness leads to a variety of theories beyond the Standard Model (BSM). Many BSM theories predict the existence of additional fundamental particles. Examples of such BSM theories are Extended Two Higgs Doublet models [7] or the Two Real Singlet model [8], which both predict two additional scalar bosons. If these two additional scalar bosons X and S have masses such that $m_X > m_S > m_H$ and $m_X - m_S \geq m_H$, the resonant decay of a scalar boson X into a Higgs boson H in conjunction with a scalar boson S becomes possible. The H and S bosons then decay further. In this thesis, the $b\bar{b}WW$ decay channel is analyzed in the 1-lepton final state. For a thorough analysis of this channel, it is important to understand the background processes and reduce their impact on the results. Therefore, control regions for the different background contributions are defined, in which the modelling can be checked and background estimation can be performed, if necessary. The focus of this thesis lies on the non-prompt background contributions in the $X \to SH \to b\bar{b}WW$ decay channel. These contain non-prompt leptons, that have their origin in hadronic processes and can fake the properties of the signal leptons.

First, an overview over the Standard Model and the Higgs mechanism within and beyond the Standard Model is given in Chapter 2. This is followed by a Chapter 3 on the experimental background of the LHC and a description of the setup of ATLAS detector at CERN. Then, a more detailed picture of the split-boosted 1-Lepton analysis of the $X \to SH \to b\bar{b}WW$ channel is drawn in Chapter 4. A study of the working points for lepton identification and isolation is presented in Chapter 5, followed by Chapter 6, which describes the definition of a non-prompt background enhanced control region. This chapter also contains the comparison between Monte Carlo simulation of the background processes and the unblinded data, as well as an assessment of the systematic uncertainties associated with this modelling. Chapter 7 gives a conclusion of the results and an outlook into the next steps of the analysis and beyond the scope of this thesis.

2. Theoretical Background

In nature, there are four fundamental interactions: the electromagnetic, the strong and the weak interaction, as well as gravity. The first three can be summarized into a consistent theory by the Standard Model (SM) of particle physics. The SM is a largely successful theory and has shown its ability to describe nature with great accuracy in the past decades. However, the SM has some shortcomings which will be described in Section 2.2. To look into these Beyond the Standard Model (BSM) scenarios, one should first understand the Standard Model itself, which is discussed in the Section 2.1. In Section 2.1.2, the Higgs pair production is explained followed by a Section 2.3 detailing the $X \rightarrow SH \rightarrow \; b\bar{b}WW$ decay channel.

2.1. The Standard Model of Particle Physics

The SM consists of four types of bosons mediating the three fundamental interactions and twelve fermions plus the Higgs boson. All the fundamental particles of the SM are represented in Figure 2.1. Moreover, there are the corresponding antiparticles to all the particles pictured in the figure. These can be interpreted as particles moving back in time or as particles with a negative sign on the charges. Uncharged particles are their own antiparticles.

All the mediating bosons have a spin of 1 and mediate one of the fundamental forces, each coupling to different properties of the particle. The photon γ couples to the electrical charge, mediating the electromagnetic interaction. Strong interactions are mediated via the gluon, which couples to the color charge of particles. The $W^{\pm}$ boson and the Z^0 boson are the mediating bosons for the weak interaction. The fermions in the SM have spin-$\frac{1}{2}$ and some can be classified as quarks, which are the only fermions with a color charge and thus take part in the strong interaction. The up-type quarks have an electric charge of $+\frac{2}{3}$ and down-type quarks have a charge of $-\frac{1}{3}$. The other fermions are the leptons, which are the electron, the muon and the tauon as well as their corresponding neutrinos. The Higgs boson is the only scalar boson in the SM (see Section 2.1.1).

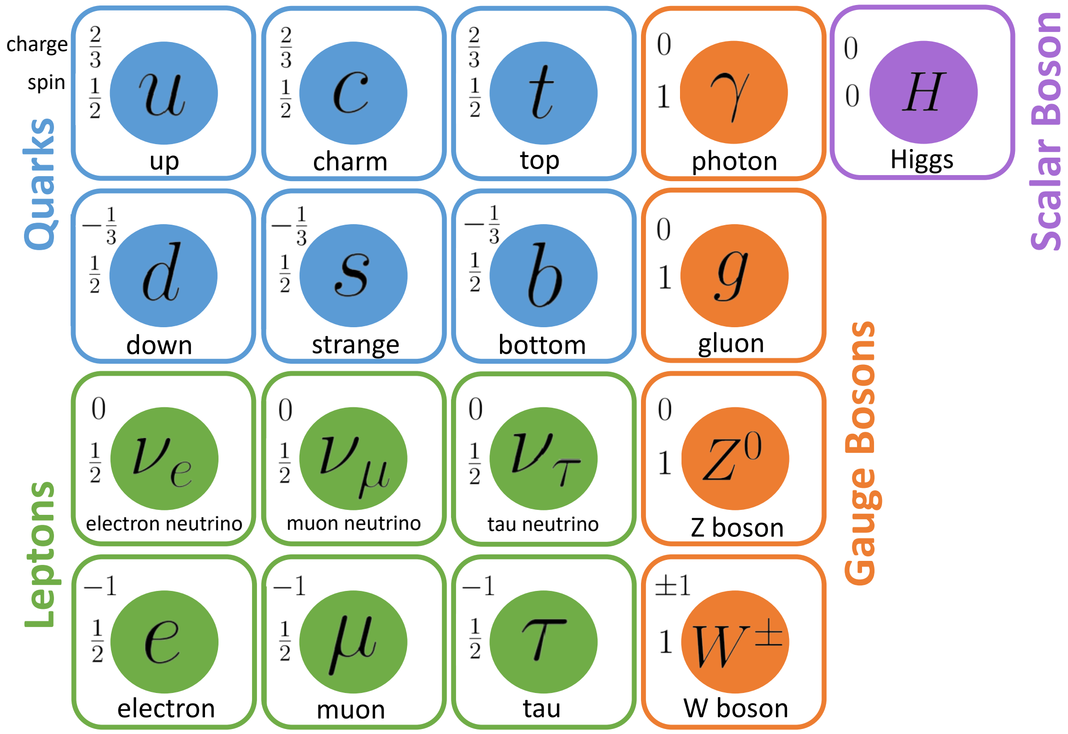


Figure 2.1.: Overview over the particle content in the Standard Model. Spin and charge values taken from [9].

The Standard Model possesses a $SU(3)_\mathrm{c} \otimes SU(2)_\mathrm{L} \otimes U(1)_Y$ symmetry. Here the c refers to the color charge, the L refers to the left-chiral nature of the weak interaction and the Y indicates the weak hypercharge (see Section 2.1).

Quantum Electrodynamics

The electromagnetic interaction is described by quantum electrodynamics (QED) [10, 11]. This abelian gauge theory imposes a local gauge invariance under the U(1) symmetry group. The QED Lagrangian is given by [1]

$$\mathcal{L}_{EM} = \underbrace{\overline{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x)}_{\text{free fermion field}} + \underbrace{e\overline{\psi}(x)\gamma^\mu\psi A_\mu}_{\text{fermion-photon interaction}} - \underbrace{\frac{1}{4}F^{\mu\nu}F_{\mu\nu}}_{\text{free photon field}} \quad , \tag{2.1}$$

where $\overline{\psi} = \psi^\dagger\gamma^0$ is the adjoint spinor and γ^μ denotes the gamma matrices. Furthermore, m stands for the particle mass and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\nu$ is the field strength tensor.

Quantum Chromodynamics

Quarks have a property which is called color charge. It can take three values: red, green and blue. The mediating boson of the strong interaction, the gluon, couples to this color charge.

The underlying theory of the strong interaction is called quantum chromodynamics (QCD) [12] and the Lagrangian of QCD is given by

$$\mathcal{L}_{QCD} = \overline{q}_i (i\gamma^\mu (D_\mu)_{ij}) q_j - \frac{1}{4} G^{\mu\nu} G_{\mu\nu} \tag{2.2}$$

$$\text{with} \quad G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^a_\nu \tag{2.3}$$

$$\text{and} \quad (D_\mu)_{ij} = \partial_\mu \delta_{ij} - i g_s (T_a)_{ij} G^a_\mu \quad . \tag{2.4}$$

Here g_s denotes the coupling strength and $T_a = \frac{1}{2}\lambda_a$ are the generators with the Gell-Mann matrices λ_a. While f^{abc} are the structure constants, the indices i and j correspond to the color states. Equation 2.3 expresses the kinematic terms of the gluon fields G^a_μ and Equation 2.4 is the covariant derivative. This Lagrangian is invariant under a local $\mathrm{SU}(3)_c$ symmetry.

A special property of the strong interaction is quark confinement, which means that color-charged quarks cannot exist in isolated states but only in color neutral states. A quark and an antiquark can form a bound state called a meson when they have opposite color charges. The other option for a color neutral state are three quarks or antiquarks bound together into a state called a baryon. For baryons, the three quarks must have different color charges, so either red, green and blue or antired, antigreen and antiblue.

Together with the asymptotic freedom of QCD, which refers to the increase in coupling strength with increasing distance, the confinement leads to the characteristic hadronization when color neutral states get separated.

Quantum Flavordynamics

The weak interaction has some unique properties. In contrast to the other interactions, it does not couple equally to particles of left- and right-handed chirality. Rather, the $W^\pm$ boson, the mediator of the charged weak interaction, only couples to left-chiral particles and right-chiral antiparticles. Theoretically, this is expressed by the third component of the isospin having a value of $I_3 = \pm\frac{1}{2}$. Right-chiral particles and left-chiral antiparticles consequently have $I_3 = 0$.

The charged weak interaction can be described by the Lagrangian [13, 14]

$$\mathcal{L}_{CC} = -\frac{g_2}{\sqrt{2}}\Big[\bar{d}_j\gamma^\mu\frac{1-\gamma^5}{2}V^{ij}_{\text{CKM}}u_i + \bar{\nu}_i\gamma^\mu\frac{1-\gamma^5}{2}\ell_i\Big]W^+_\mu \tag{2.5}$$

$$-\frac{g_2}{\sqrt{2}}\Big[\bar{u}_i\gamma^\mu\frac{1-\gamma^5}{2}V^{ij}_{\text{CKM}}d_j + \bar{\ell}_i\gamma^\mu\frac{1-\gamma^5}{2}\nu_i\Big]W^-_\mu \quad . \tag{2.6}$$

with u and d being up- and down-type quarks and ℓ and ν representing the charged leptons and the neutrinos. Here γ^5 is given by $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ and g_2 denotes the coupling strength. V_{CKM} is the Cabibbo-Kobayashi-Maskawa matrix [15], which connects the quark flavor eigenstates to the mass eigenstates and describes the mixing between quark flavors.
For the neutral weak interaction the Lagrangian takes the form of

$$\mathcal{L}_{NC} = \frac{g_2}{\cos\theta_W}\Big[\overline{\psi}\gamma^\mu\frac{1-\gamma^5}{2}\psi - \sin^2\theta_W Q\overline{\psi}\gamma^\mu\psi\Big]Z_\mu \quad , \tag{2.7}$$

with the electrical charge Q and the Weinberg angle θ_W. The Z boson is the mediator of the neutral weak interaction. It couples to a combination of weak isospin and the electrical charge.

Electroweak Unification

In quantum flavor dynamics, three gauge fields associated with the associated with the $SU(2)_\text{L}$ are introduced: two associated with $W^\pm$ and one giving rise to a neutral current. This current is not simply the Z^0 boson, as the Z^0 couples to both right-handed particles and left-handed antiparticles.
The neutral weak interaction and the electromagnetic interaction have in common that they are mediated by a electrically neutral boson. This motivated Glashow [16], Salam [17] and Weinberg [18] to create a theory in which electromagnetism and the neutral weak interaction are unified. In their theory, they replace the $\text{U}(1)_e$ symmetry of QED with a new $\text{U}(1)_Y$ local gauge symmetry, where Y refers to the so-called hypercharge which is given by [1]

$$Y = 2\Big(Q - I_3\Big) \quad . \tag{2.8}$$

Here, Q is the electric charge and I_3 denotes the third component of the weak isospin.
The physical Z^0 boson is then acquired from a mixing of the third gauge field of quantum flavordynamics and the gauge field of the $\text{U}(1)_Y$ symmetry and can be written as

$$Z_\mu = -B_\mu\sin\theta_\text{W} + W^{(3)}_\mu\cos\theta_\text{W} \quad . \tag{2.9}$$

Similarly, the photon can be written as

$$A^{\mu} = +B_{\mu}\cos\theta_{\mathrm{W}} + W_{\mu}^{(3)}\sin\theta_{\mathrm{W}} \quad . \tag{2.10}$$

The Lagrangian of electroweak unification takes the form of

$$\mathcal{L}_{NC} = g_Z \bar{f}\gamma_{\mu}\frac{1}{2}\left(c_V^f - c_A^f\gamma^5\right) f Z^{\mu} \quad , \tag{2.11}$$

with the coupling constant of the weak hypercharge g_Z. c_V and c_A denote the vector and axial-vector couplings which can be expressed in terms of electrical charge Q and third component of the isospin I_3 via

$$c_V = I_3 - 2Q\sin^2\theta_W \tag{2.12}$$

$$c_A = I_3 \quad , \tag{2.13}$$

with the Weinberg angle θ_W, which is a free parameter of the SM.

2.1.1. The Brout-Englert-Higgs mechanism

The Brout-Englert-Higgs mechanism [19–24] is a crucial part of the SM. It explains how gauge bosons can acquire mass.

In this mechanism, a complex scalar doublet Φ is introduced, which can be written as

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} , \tag{2.14}$$

with a charged component Φ^+ and a neutral component Φ^0. This field follows a potential of the form [1]:

$$V(\Phi) = \mu^2(\Phi\Phi^{\dagger}) + \lambda(\Phi\Phi^{\dagger})^2. \tag{2.15}$$

For $\mu^2 < 0$, the potential has the typical form of the "mexican hat", that can be seen in Figure 2.2. It has a non-zero expectation value and an infinite set of minima [1]. Since the minimum of the potential is not at the origin, the electroweak symmetry is spontaneously broken. The system chooses a vacuum state $v = \sqrt{\frac{-\mu^2}{\lambda}}$ and therefore violates symmetry. One can now rewrite the field Φ in the unitary gauge, so that it takes the form

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} . \tag{2.16}$$

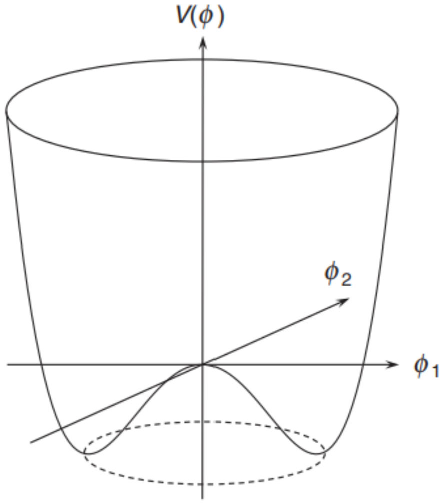


Figure 2.2.: The plot shows the potential given by Eq. (2.15) with $\mu^2 < 0$ and $\lambda > 0$. The picture is taken from Ref. [1].

The result can then be inserted into the Lagrangian. By insisting on the $SU(2)_\mathrm{L} \otimes U(1)_Y$ symmetry, one will obtain mass terms for the gauge bosons in the Lagrangian. The final Standard Model Lagrangian is then given by

$$\mathcal{L} = \underbrace{\frac{1}{2}\partial_\mu H \partial^\mu H}_{\text{kinematic term}} - \underbrace{\lambda v^2 H^2}_{\text{mass term}} - \underbrace{\lambda v H^3 - \frac{1}{4}\lambda H^4}_{\text{self-interactions}} + \underbrace{\frac{g_2^2 v^2}{4} W_\mu^- W^{+\mu} + \frac{v^2}{8(g_2^2+g_1^2)} Z_\mu Z^\mu}_{\text{mass terms for gauge bosons}} \tag{2.17}$$

$$+ \underbrace{\frac{g_2^2 v}{2} W_\mu^- W^{+\mu} H + \frac{g_2^2}{4} W_\mu^- W^{+\mu} H^2 + \frac{v}{4(g_2^2+g_1^2)} Z_\mu Z^\mu H + \frac{1}{8(g_2^2+g_1^2)} Z_\mu Z^\mu H^2}_{\text{interactions between the scalar and gauge fields}} . \tag{2.18}$$

The Higgs mass is dependent on the parameter v of the Higgs potential and through experimental determination of the Higgs mass the shape of the Higgs potential can be probed. The Higgs mass is measured to be 125.25 GeV [9].

The Higgs mechanism can also explain the fermion masses, by adding the term

$$\mathcal{L}_\mathrm{d} = -\lambda_f \left(\overline{\psi}_L \Phi \psi_R + \overline{\psi}_R \Phi^\dagger \psi_L\right) \tag{2.19}$$

to the Lagrangian. This term is invariant under $SU(2)_\mathrm{L} \otimes U(1)_Y$ transformations and in unitary gauge takes the form of

$$\mathcal{L}_\mathrm{d} = -\frac{\lambda_f v}{\sqrt{2}}\left(\overline{D_L} D_R + \overline{D_R} D_L\right) - \frac{\lambda_f}{\sqrt{2}} H \left(\overline{D_L} D_R + \overline{D_R} D_L\right) \quad . \tag{2.20}$$

Here, λ_f refers to the Yukawa coupling [1]. In this way down-type fermions can acquire their mass and also an interaction term between the fermions and the Higgs boson is generated. For up-type quarks the formalism takes a different form, because the vacuum state of the Higgs potential must be in the neutral component of the doublet. So, the conjugate complex scalar is needed, which can be written as

$$\Phi_C = -i\sigma_2\Phi^* = \begin{pmatrix} -\phi_{0*} \\ \phi^- \end{pmatrix} \quad , \tag{2.21}$$

and transforms the same as the normal doublet. After symmetry breaking this results in the Lagrangian term for up-type fermions looking like

$$\mathcal{L}_\mathrm{d} = \lambda_f \left(\overline{\psi}_L \Phi_C \psi_R + \overline{\psi}_R \Phi_C^\dagger \psi_L\right) \tag{2.22}$$

$$= -\frac{\lambda_f v}{\sqrt{2}}\left(\overline{U_L} U_R + \overline{U_R} U_L\right) - \frac{\lambda_f}{\sqrt{2}} H \left(\overline{U_L} U_R + \overline{U_R} U_L\right) \quad . \tag{2.23}$$

The fermion masses are then given by

$$m_f = \frac{v\lambda_f}{\sqrt{2}} \quad . \tag{2.24}$$

2.1.2. Higgs boson pair production

The Higgs boson can undergo self-interactions in the SM. This allows for Higgs boson pair production, but only in a non-resonant mode. The dominant production mechanism for Higgs boson pairs at hadron colliders is gluon-gluon fusion (ggf). The possible processes can be seen in Figure 2.3. But the two processes interfere destructively leading to a very small cross-section of $\sigma(pp \to \ HH) = \ 31.05\,\mathrm{fb}$ for $\sqrt{s} = 13\,\mathrm{TeV}$ [25]. This cross-section will increase with the center-of-mass energy. For a centre-of-mass energy of $\sqrt{s} = 14\,\mathrm{TeV}$, a cross-section of $\sigma(pp \to \ HH) = \ 36.69\,\mathrm{fb}$ is expected.

Figure 2.3.: Dominant SM processes producing two Higgs bosons at hadron colliders.

This cross-section can be enhanced by beyond the Standard Model processes (Section 2.2). Two Higgs bosons can then also be produced in decays of a heavy resonance X that has a mass larger than the Higgs mass. The corresponding Feynman diagram is pictured in

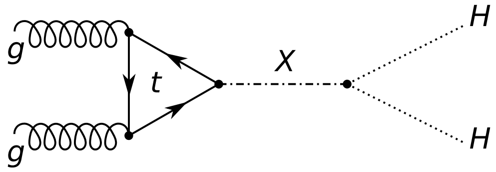


Figure 2.4.: Higgs pair production in a resonant decay of a BSM particle X.

	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.07%	
$\gamma\gamma$	0.26%	0.1%	0.02%	0.01%	$<0.001\%$

Table 2.1.: Listed are the branching ratios for a Higgs boson pair with a Higgs mass of $m_H = 125\,\text{GeV}$ [26]

Figure 2.4. The two Higgs bosons then decay further, resulting in the branching ratios listed in Table 2.1. The highest branching ratio has the decay channel where both Higgs bosons decay into a $b\bar{b}$ pair. The $b\bar{b}WW$ decay channel occurs in a quarter of these Higgs boson pair productions.

2.2. Beyond the Standard Model Theories

Although the SM is largely successful theory, it is obvious that it has some shortcomings and thus cannot be the ultimate theory of nature. One of these shortcomings is the fact that in the SM neutrinos are massless particles. Contrary to this, neutrino oscillations have been observed [5], which imply non-zero neutrino masses. This suggests a possible BSM mechanism that generates these masses.
Another example for the incompleteness of the SM is the fact, that the $\mathcal{CP}$ violation present in the SM cannot explain the asymmetry between matter and antimatter in the universe [27]. Moreover, there is no explanation for gravity or dark matter [6] in the Standard Model, which is needed for it to be an all encompassing theory.

This incompleteness of the Standard Model motivates a variety of BSM theories. Two examples focusing on the Higgs sector will be illustrated in the following sections.

2.2.1. Two-Higgs-Doublet Model

The Two-Higgs-Doublet Model (2HDM) [28] is an extension of the SM, where an additional complex scalar doublet is introduced to the potential. This potential then takes the form

$$\begin{aligned} V(\Phi_1, \Phi_2) &= m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 (\Phi_1^\dagger \Phi_2^\dagger + \Phi_2^\dagger \Phi_1^\dagger) && (2.25) \\ &+ \frac{\lambda_1}{2}(\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2}(\Phi_2^\dagger \Phi_2)^2 + \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 && (2.26) \\ &+ \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2}\Big[(\Phi_1^\dagger \Phi_2)^2 + (\Phi_2^\dagger \Phi_1)^2\Big] \quad . && (2.27) \end{aligned}$$

Here Φ_1 and Φ_2 are the two complex scalar doublets with hypercharge $Y = +1$. The potential reaches its minimum for

$$\langle \Phi_1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \quad \langle \Phi_2 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 \end{pmatrix} \quad . \tag{2.28}$$

This theory results in eight fields, where three are assigned to the $W^\pm$ and Z^0 boson masses. The remaining five fields result in physical states. Hence, in this theory there are two scalar bosons h and H, two charged bosons $H^\pm$ and a pseudoscalar particle A.
The 2HDM can be extended by other additions such as another real singlet [7], leading to the generation of three scalar bosons X, S and H. The potential is then written as

$$V = V^{\text{2HDM}} + \frac{m_S^2}{2} S^2 + \frac{\lambda_S}{8} S^4 + \frac{\lambda_{S1}}{2}\Big(\Phi_1^\dagger \Phi_1\Big) S^2 + \frac{\lambda_{S2}}{2}\Big(\Phi_2^\dagger \Phi_2\Big) S^2 \quad , \tag{2.29}$$

where V^{2HDM} is the potential from Equation 2.25. Expanded around the minima, the fields take the form

$$\Phi_1 = \begin{pmatrix} \phi_1^+ \\ \frac{v_1 + \rho_1 + i\eta_1}{\sqrt{2}} \end{pmatrix}, \quad \Phi_1 = \begin{pmatrix} \phi_2^+ \\ \frac{v_2 + \rho_2 + i\eta_2}{\sqrt{2}} \end{pmatrix}, \quad S = v_S + \rho_S \quad . \tag{2.30}$$

Here, v_1, v_2 and v_S are real expectation values with $v = \sqrt{v_1^2 + v_2^2} \approx 246\,\text{GeV}$.
Two $\mathbb{Z}_2^S$ symmetries are imposed in the following way

$$\begin{aligned} &\mathbb{Z}_2 : \Phi_1 \to \Phi_1,\ \Phi_2 \to -\Phi_2,\ S \to S, \quad , && (2.31) \\ \text{and} \quad &\mathbb{Z}_2^S : \Phi_1 \to \Phi_1,\ \Phi_2 \to \Phi_2,\ S \to -S \quad . && (2.32) \end{aligned}$$

If $v_S = 0$ the second $\mathbb{Z}_2^S$ symmetry is conserved and S becomes a dark matter candidate and if $v_S \neq 0$ S mixes with other $\mathcal{CP}$ even scalar particles.
This leads to 3 physical states that couple to each other allowing the decay $X \to SH$. The scalar boson S could then decay into two W bosons.

2.2.2. Two-Real-Singlet Model

To obtain a BSM theory with extra scalar bosons, one can add N singlets ϕ to the potential of the SM

$$V(\phi_i, \Phi) = V_{SM}(\Phi) + V_{\text{Singlets}}(\phi_i, \Phi) \tag{2.33}$$

$$\text{with} \quad V_{\text{Singlets}}(\phi_i, \Phi) = a_i\phi_i + m_{ij}\phi_i\phi_j + T_{ijk}\phi_i\phi_j\phi_k + \lambda_{ijkl}\phi_i\phi_j\phi_k\phi_l \tag{2.34}$$

$$+ T_{iHH}\phi_i(\Phi^\dagger\Phi) + \lambda_{ijHH}\phi_i\phi_j(\Phi^\dagger\Phi) \quad . \tag{2.35}$$

In the case of the Two-Real-Singlet Model (TRSM) [8] two singlets X and S are added to the SM. These singlets are real and transform under the $\mathbb{Z}_2$ symmetry as

$$\mathbb{Z}_2^S : S \to -S,\ X \to X,\ \Phi \to \Phi \quad , \tag{2.36}$$

$$\text{and} \quad \mathbb{Z}_2^X : X \to -X,\ S \to S,\ \Phi \to \Phi \quad . \tag{2.37}$$

Under these assumptions, the potential in Equation 2.34 simplifies to

$$V(\Phi, X, S) = \mu_\Phi^2(\Phi^\dagger\Phi) + \lambda_\Phi(\Phi^\dagger\Phi)^2 + \mu_S^2 S^4 + \mu_X^2 X^2 + \lambda_X X^4 \tag{2.38}$$

$$+ \lambda_{\Phi S}(\Phi^\dagger\Phi)S^2 + \lambda_{\Phi X}(\Phi^\dagger\Phi)X^2 + \lambda_{SX}S^2X^2 \tag{2.39}$$

Using the unitary gauge, the fields can be expressed as

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + \phi_H \end{pmatrix}, \quad X = \frac{v_x + \phi_X}{\sqrt{2}}, \quad S = \frac{v_S + \phi_S}{\sqrt{2}} \tag{2.40}$$

where v is the SM expectation value of the Higgs potential.
When v_s and v_X take non-zero values, the $\mathbb{Z}_2$ symmetry is softly broken resulting in three physical states h_1, h_2, h_3 with $m_1 < m_2 < m_3$ that are a mixture between ϕ_H, ϕ_X and ϕ_S. One of these physical states h_i has the couplings of the Higgs boson predicted by the SM. Thus one can set this state equal to the discovered scalar boson, h_1. The other two physical states are $\mathcal{CP}$ even scalar particles. With the Higgs mass m_H and the parameter $v \approx 246\,\text{GeV}$ set by measurement, the TRSM contains seven free parameters. These free

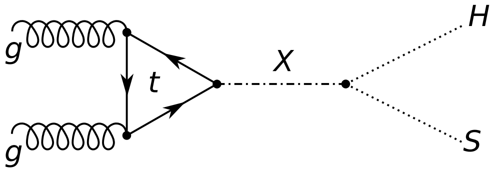


Figure 2.5.: Feynman diagram of the decay of a BSM scalar boson X decaying into a Higgs boson and a scalar boson S.

parameters are the masses of the BSM particles m_2 and m_3, the vacuum expectation values v_x and v_S but also the mixing angles θ_{HS}, θ_{HX}, θ_{SX}, that describe the mixing between the fields forming the mass eigenstates.
One important phenomenological implication of this model is that for $m_2 < 2\,m_H$ the $b\bar{b}WW$ decay channel has the highest branching ratio. This would mean that it is mostly likely that S decays into two W bosons, while the Higgs boson decays into a $b\bar{b}$ pair.

2.3. The $\mathbf{X \rightarrow SH \rightarrow b\bar{b}WW}$ decay channel

In some BSM theories, the production of a Higgs boson in conjunction with a BSM scalar particle S in the decay of a heavy resonance X is possible. In this analysis, the masses of the bosons fulfill $m_X > m_S > m_H$. If all of the scalar bosons are produced on-shell, another mass constraint emerges, which is $m_X - m_S > m_H$. The Feynman diagram of this decay is depicted in Figure 2.5. If S has a Higgs boson-like couplings but a different mass than the Higgs boson, the $b\bar{b}WW$ decay channel could have the highest branching ratio with $S \rightarrow WW$ being the dominant decay mode for S.

In the 1-lepton final state, one of the W bosons decays hadronically, while the other one decays leptonically. This final state is the best compromise between a reasonable branching ratio and a controllable background. Depending on the masses of X and S different topologies of this decay channel arise. If the mass of the resonance X is large in comparison to m_S and the Higgs mass, the decay products have more kinetic energy leading to boosted topologies, where not all decay products can be resolved. All possible topologies in the 1-lepton final state can be seen in Figure 2.6. The simplest topology is the resolved topology (see Cat. 1 in Figure 2.6), which occurs for low m_S and m_X. In this topology, all decay products appear seperated in the detector and can be resolved. For higher m_X and m_S, the split-resolved and the semi-boosted topology become dominant. In both of them one of the hadronic decay products is boosted. In the split-resolved case the hadronically decaying W boson is boosted, whilst for the semi-boosted case the Higgs boson is boosted.

Above $m_X = 1\,\text{TeV}$ the dominant topologies are the split-boosted and the boosted topology. The boosted topology is defined by collimated decay products of the Higgs boson as well as the hadonically decay W bosons. Additionally, the final-state lepton is emitted so closely to the hadronic W boson, that it cannot be resolved and might even overlap with it.

In this thesis, the focus lies on the split-boosted topology (see Cat. 2 in Figure 2.6), which is dominant for $m_X \gtrsim 1\,\text{TeV}$ and $m_S \gtrsim 0.3\,m_X$. The characteristics of the split-boosted topology are the collimated hadronic decay products of the Higgs boson and the hadronically decaying W boson and the large separation between the two W bosons. This results in the final state lepton being separated from the jets such that it is possible to be resolved in the detector.

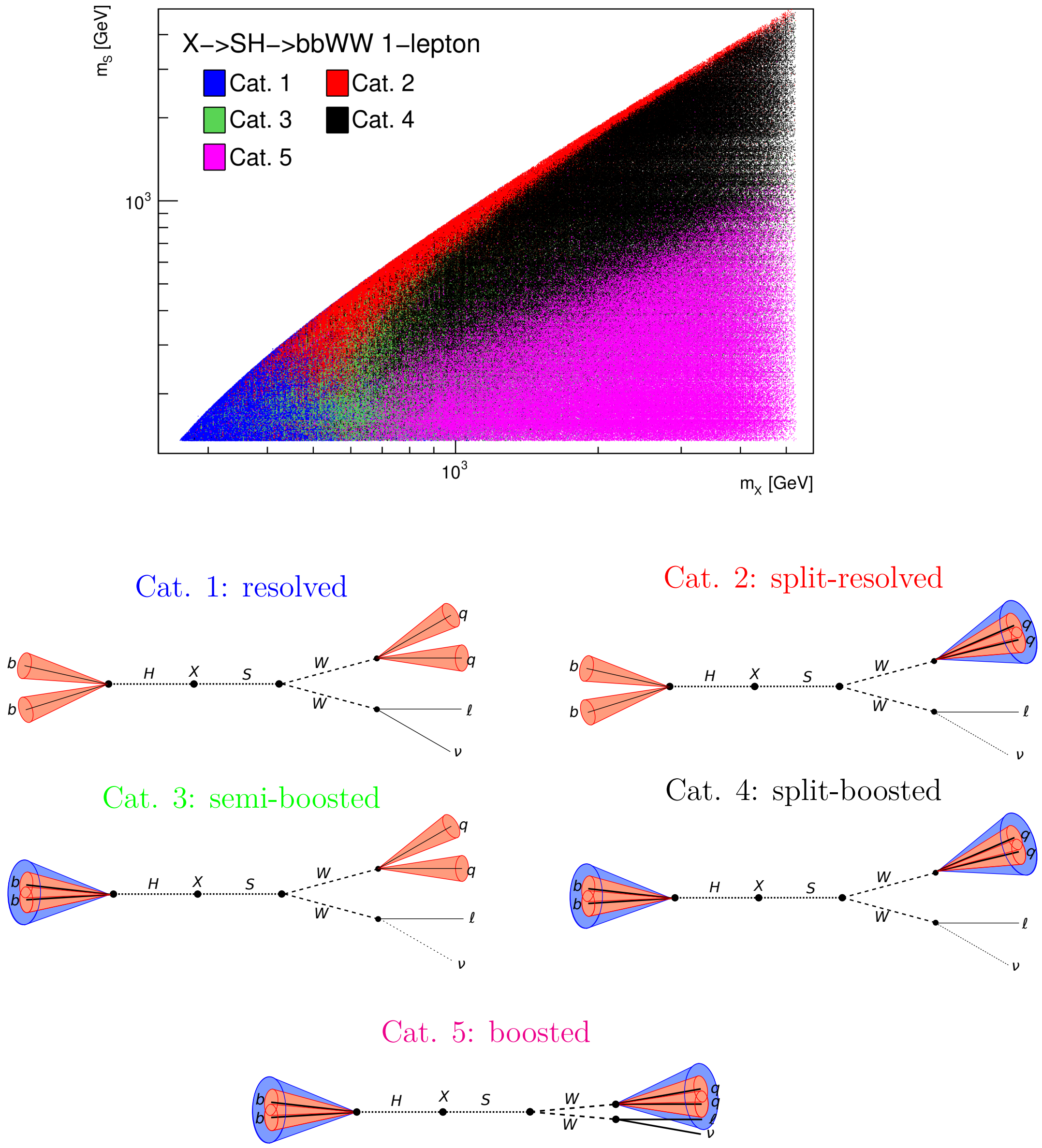


Figure 2.6.: All topologies of the $X \rightarrow SH \rightarrow b\bar{b}WW$ decay channel in the 1-lepton final state. The plot on top shows which of the topologies is dominant depending on the mass of the BSM particles X and S.

3. LHC and the ATLAS detector

The following chapter illustrates the setup of the Large Hadron Collider and the Atlas experiment, which is used in the data collection for the studies performed in this thesis.

3.1. Large Hadron Collider

The Large Hadron Collider (Lhc) at Cern is a ring accelerator, in which both protons and heavy ions can be collided. It was build in the old tunnel of the Large Electron-Positron Collider (LEP). As of now it is the largest and most powerful particle collider in the world. At the Lhc, bunches of up to 10^{11} protons collide 40 million times per second [29].

In Run 3, which is the data collection period from 2022 to 2025, the Lhc operates at an center-of-mass energy of $\sqrt{s} = 13.6\,\text{TeV}$. Besides the center-of-mass energy, the achieved luminosity is an important parameter, which quantifies the amount of collision data. The Atlas experiment collected a luminosity of $140.07\,\text{fb}^{-1}$ in Run 2, the last data collection period. For Run 3, the goal is a luminosity of $250\,\text{fb}^{-1}$.

The acceleration of the protons is done successively using a series of smaller accelerators. First, negative hydrogen ions are accelerated to 160 MeV by the Linear Accelerator 4 (LINAC4) and then injected into the Proton Synchrotron Booster (PSB), where they are stripped of their electrons. This leaves only protons, which are accelerated to 2 GeV. Next the Proton Synchrotron brings them up to energies of 26 GeV. Then they enter the super Proton Synchrotron, which accelerates the protons to an energy of 250 GeV before they end up in the Lhc collider ring. The overview of this accelerator chain can be seen in Figure 3.1.

Four large experiments are located along the collider ring of the Lhc: Alice [31], Atlas [32], Cms [33] and Lhcb [34]. In the following section the Atlas experiment will be explained in detail.

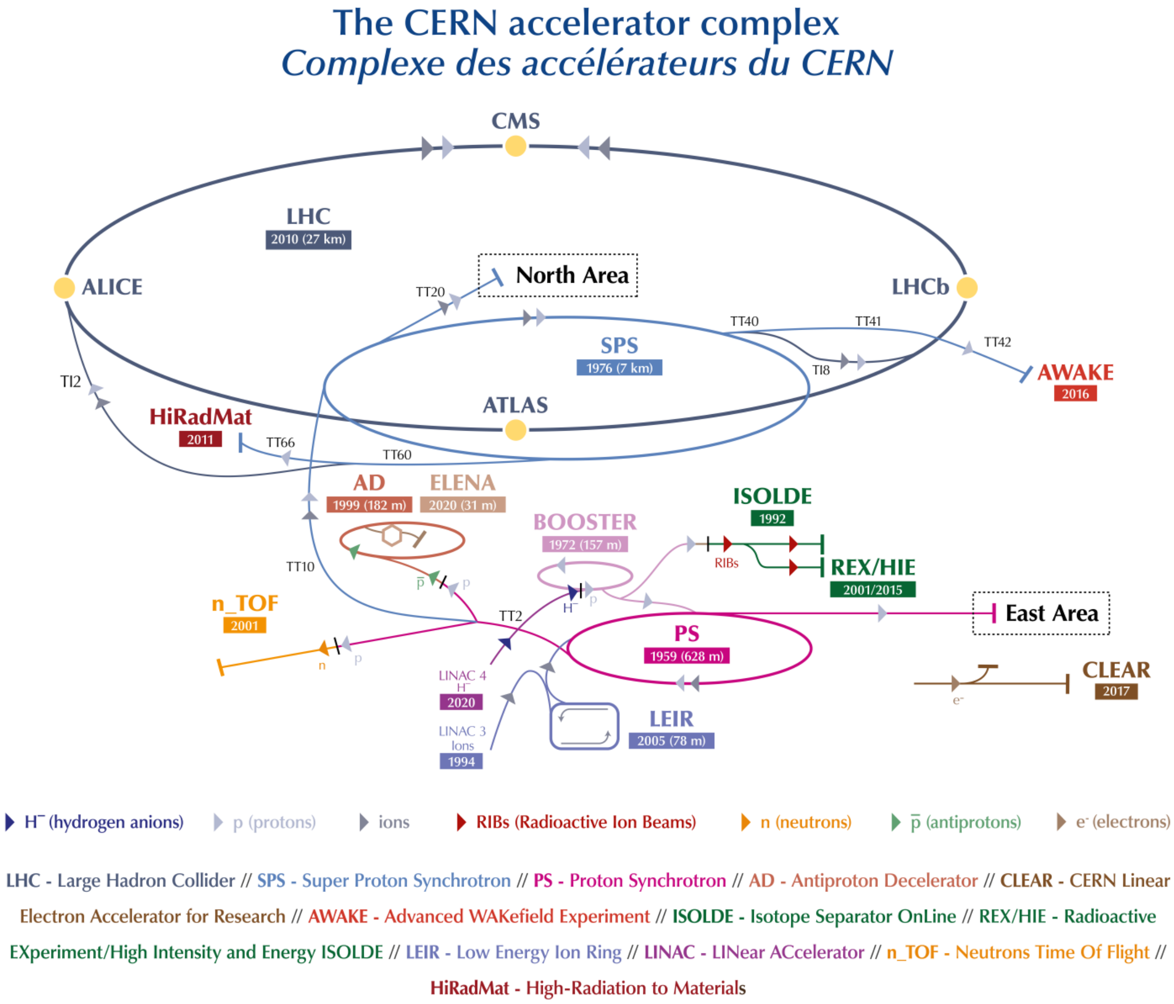


Figure 3.1.: Overview of the accelerators at CERN [30].

3.2. ATLAS detector

The ATLAS detector [32] is one of the experiments located at the LHC at CERN. ATLAS is an acronym for A Toroidal Lhc ApparatuS. The detector is a general purpose detector and has a cylindrical shape with a length of 46 m and a diameter of 25 m. An overview of the setup of ATLAS can be seen in Figure 3.2.

The ATLAS detector is made up of several layers of instrumentation, each fulfilling a specific purpose in the measurement. The Inner Detector is closest to the interaction point, which is described in more detail in Section 3.2.1. After that an Electromagnetic Calorimeter is placed in the ATLAS setup, followed by the Hadronic Calorimeter [29]. To measure muons more precisely the outer layer of the ATLAS detector is equipped with the

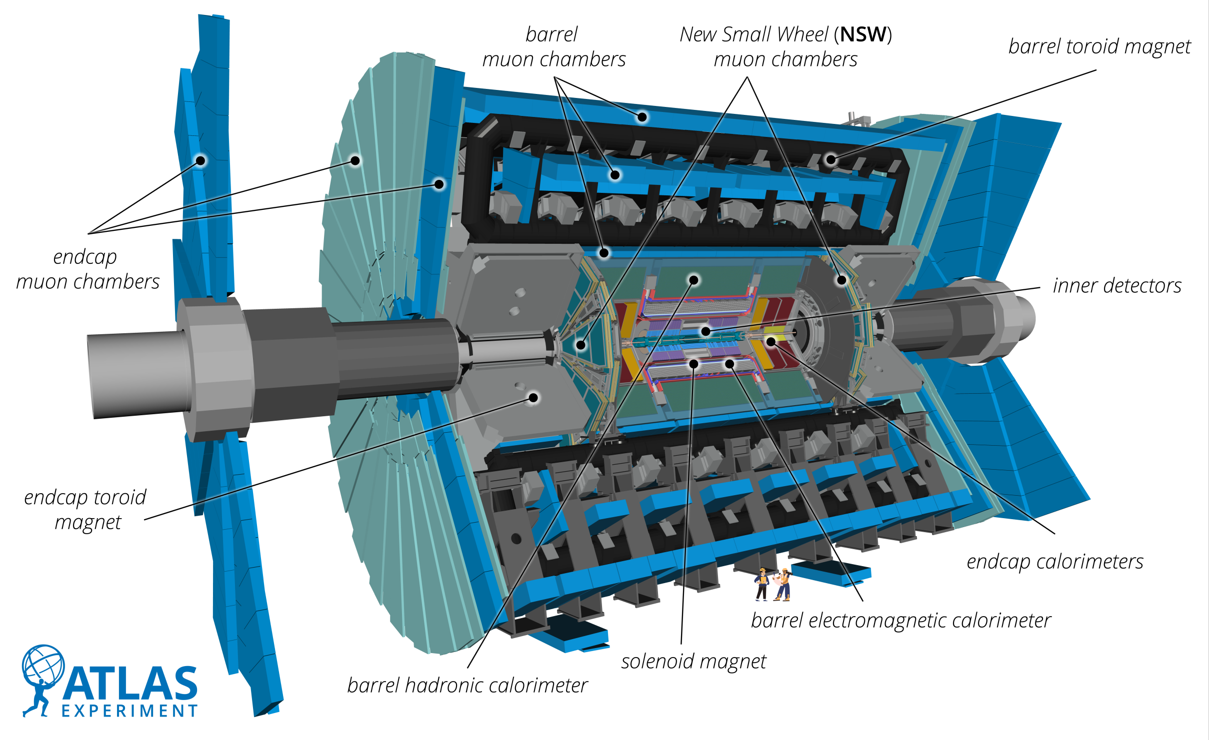


Figure 3.2.: Overview of the setup of the ATLAS detector at CERN. The image was taken from Ref. [35].

Muon Spectrometer.

The coordinate system in the ATLAS detector can be seen in Figure 3.3. The origin of the coordinate system is placed at the interaction point. The x-axis points to the center of the LHC, while the y-axis is defined upwards and the z-axis follows the beamline such that the coordinate system is right-handed. Due to the cylindrical form of the detector, it is useful to apply a cylindrical coordinate system, where the azimuthal angle ϕ goes

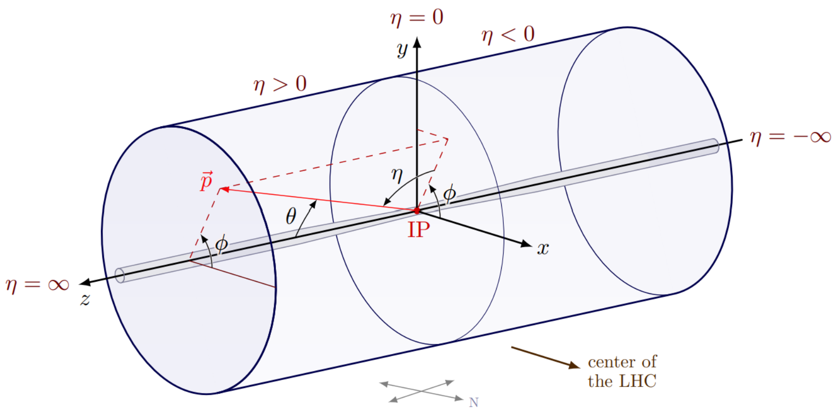


Figure 3.3.: The coordinate system used in the ATLAS detector. [36]

around the z-axis and the polar angle θ lies between the z-axis and the x-y plane. Another useful quantity is the pseudorapidity η which is defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \quad . \tag{3.1}$$

Furthermore, the distance between the momentum vectors of two objects in the detector can be described by

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad . \tag{3.2}$$

The momentum of an object in the detector is typically expressed as the transverse momentum

$$p_T = \sqrt{p_x^2 + p_y^2} \quad . \tag{3.3}$$

3.2.1. Inner Detector

The Inner Detector is, as the name suggests, the detector closest to the interaction point. As a tracking detector, its main purpose is to reconstruct particle tracks and measure charged particles. The basic principle of this detector relies on the Lorentz force. With the help of a magnetic field, the particles are deflected by the Lorentz force, according to the equation:

$$R = \frac{p_T}{B \cdot q} \tag{3.4}$$

Here, p_T is the transverse momentum of the particle, B is the magnetic field strength, q is the charge of the particle and R is the radius of the particle's trajectory. Thus, the radius of the trajectory is related to the ratio of the p_T and q, which are properties of the particle measured. From the direction of the curvature of the trajectory, one can also determine the particle's charge[1]. Since this kind of detector relies on the ionization, only charged particles can be detected. In the Inner Detector, a solenoidal magnetic field of 2 T is used [29]. The Inner Detector is composed of three subdetectors. These subdetectors are the silicon pixel detector, the semiconductor tracker and the transition radiation tracker [29]. The silicon pixel detector, the innermost part of the Inner Detector, is made up of about 92 million silicon pixels, which are arranged in 4 barrel layers and 3 end cap layers. All pixels have a size of $50 \times 400\,\mu\text{m}^2$ except for the pixels in the Insertable B-Layer (IBL). The IBL was added before Run 2 to improve vertex detection and B-hadron identification. It was inserted between the B-Layer (the innermost layer of the pixel detector in the barrel region) and a beam pipe with a smaller radius. The pixels of the IBL have a size of $50 \times 250\,\mu\text{m}^2$. If a charged particle hits a pixel, it ionizes the atoms in the material and creates electron-hole pairs, which travel to the electrodes resulting in a measurable current. These hits can then be reconstructed to a full trajectory of the particle.

In the semiconductor tracker, silicon strips replace the arrays of pixels, while having overall the same measurement principle as the silicon pixel detector. The semiconductor tracker contains over 6 million strips in 4 barrel layers and 9 end cap disks. The silicon strips have a mean pitch of about $80\,\mu$m. Combined the silicon pixel detector and the semiconductor tracker have a pseudorapidity coverage of $|\eta| < 2.5$.

The transition radiation tracker (TRT) is located at the outermost part of the Inner Detector. The basic principle of this detector relies on transition radiation, which is emitted when a charged particle traverses an inhomogeneous medium. The radiation excites gas atoms creating a current. The TRT is especially useful for the identification of electrons. Since transition radiation is highly dependent on the Lorentz factor, light particles such as electrons emit more transition radiation than heavier particles, thus contributing to particle identification. In the TRT, the particles travel through drift tubes filled with a mixture of xenon, carbon dioxide and molecular oxygen. It contains 300000 of these tubes, giving the detector a coverage of $|\eta| < 2.0$. All in all, the spatial resolution of the TRT equates to $170\,\mu m$.

The overall momentum resolution of the Inner Detector is $\sigma_{p_T}/p_T = 0.05\% \cdot p_T \oplus 1.0\%$ [GeV].

3.2.2. Calorimeters

Within a calorimeter, traversing particles interact with the detector material, initiating a cascade of secondary particles in a process known as a particle shower. The shower develops through repeated interactions, including radiation, ionization, scattering, and, in the case of hadrons, nuclear interactions. These processes lead to energy loss and particle multiplication until the secondary particles fall below the energy threshold required for further interactions, at which point the shower gradually ceases [1]. Calorimeters can be distinguished into electromagnetic and hadronic calorimeters. The Electromagnetic Calorimeter (ECAL) is designed specifically to measure the energy deposition of particles that interact electromagnetically (electrons, positrons and photons). Moreover, there are hadronic calorimeters (HCAL) dedicated to the measurement of hadron energy. Electromagnetic and hadronic showers differ in their properties in the detector. In general, electromagnetic showers are expected to be shorter and narrower than their hadronic counterpart.

In the ATLAS Detector, sampling calorimeters are used. They are built from alternating layers of active and absorbing material.

Liquid argon (LAr) calorimeters are utilized in the ATLAS detector and placed around the

Inner detector. They can be divided into a calorimeter covering the barrel region as well as the LAr electromagnetic endcaps, two LAr hadronic endcaps and three LAr forward calorimeters also for hadronic showers. The choice of passive material of the calorimeter depends on the placement in the detector. While the barrel and end-cap part of the ECAL use lead as a passive material, copper is used in the end-cap of the HCAL and in the forward region in combination with tungsten. The HCAL, that spans around the barrel region, is made up of plastic scintillating tiles as the active and steel as the passive material. The ECAL in the barrel region and endcaps have an η coverage of $|\eta| < 3.2$. The same η coverage is achieved by the HCAL except for the forward region, where the η coverage is given by $3.1 < |\eta| < 4.9$ [29].

The calorimeters are designed such that they contain the showers and prevent them from going into the muon system. Therefore, the calorimeter thickness is carefully chosen. The ECAL has a total thickness of > 22 radiation lengths (X_0) in the barrel region and $> 24\ X_0$ in the end-caps. In interaction lengths (λ), the thickness can be expressed as $9.7\,\lambda$ for the barrel region and $10\,\lambda$ for the end-caps.

The energy resolution of the ECAL is designed to be $\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$ [GeV] for electromagnetic showers. For the HCAL the energy resolution equates to $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ [GeV] for the barrel and the end cap parts. In the forward part of the HCAL the energy resolution is given by $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$ [GeV][29].

3.2.3. Muon Spectrometer

Muons are minimally ionizing particles and hence exit the Inner Detector and the calorimeters without depositing a significant fraction of their energy. Therefore, a dedicated tracking detector is placed around the calorimeters, the so-called Muon Spectrometer [29]. The Muon Spectrometer can be divided into two parts: a large barrel toroid, which covers $|\eta| < 1.4$, and the end-caps, which is equipped with two smaller magnets and covers $1.6 < |\eta| < 2.7$. The region between $1.4 < |\eta| < 1.6$ is commonly referred to as the transition region. In this region muons are deflected by the barrel and the endcap magnets [29].

The measurement of the tracks is done using Monitored Drift Tubes (MDT) for the majority of the η-range. For $2.0 < |\eta| < 2.7$ tracks are detected with Cathode Strip Chambers (CSC). These are multiwire proportional chambers with cathodes, that are segmented into strips and are more robust to a higher rate and rougher background conditions.

For triggering, Resistive Plate Chambers are installed in the barrel region, whilst Thin Gap Chambers are used in the end-caps. Together the trigger system covers $|\eta| < 2.4$.

Overall the muon spectrometer reaches a resolution of $\sigma_{p_T} = 10\%$ for $p_T = 1$TeV muons

[29].

3.2.4. Trigger system

The recorded data in the Atlas detector causes a data stream of more than 60 TB/s. Due to limited processing speed and storage capacity, it is impossible to store all the information and a selection of the recorded data is necessary. For this, the trigger system is used.

At the Atlas experiment, the recorded data is sorted through a two-level trigger system. First, the Level-1 (L1) trigger uses calorimeter information and information from the muon spectrometer to search for high transverse-momentum muons, electrons, photons, jets, and τ-leptons decaying into hadrons, as well as large missing and total transverse energy. In this hardware based trigger, the information is compared to trigger objects and based on the similarity they are kept in storage or are discarded. With this selection the event rate is reduced from 40 MHz to about 100 kHz. The L1 trigger also defines Regions-of-Interest (RoIs). These are regions that contain interesting features according to the algorithm [29].

After that, the high level trigger (HLT) uses these RoIs as seeds for further software based selections. Other than the L1 trigger it also uses the information from the Inner Detector. This trigger performs its seletion based on high-level objects, which is why this trigger is also called high-level trigger (HLT). After the HLT the trigger rate is reduced to 3.5 kHz [29].

4. The split-boosted 1-Lepton Analysis

The analysis presented in this thesis specifically focuses on the split-boosted topology of the $X \to SH \to b\bar{b}WW$ decay channel in the 1-lepton final state. The characteristics of this topology are described in Section 2.3. A sketch of this topology is shown below.

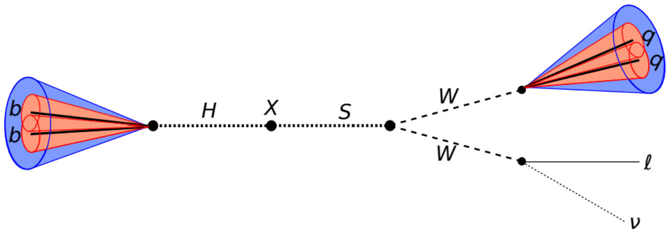


Figure 4.1.: Split-boosted topology of the $X \to SH \to b\bar{b}WW$ decay channel in the 1 lepton final state.

To illustrate the general approach of this analysis, the following sections will detail some of the overarching definitions relevant for this thesis.

4.1. Samples

The samples used for the studies in this thesis will be described in the following sections.

4.1.1. Signal Samples

The generation of $X \to SH$ samples is performed at leading order in α_S using PYTHIA 8.309, applying the A14 tune and the NNPDF2.3LO parton distribution functions (PDF) set. The simulation of hadronization and parton showering is also carried out with PYTHIA 8.309, while EVTGEN 2.1.1 is used to handle the decays of heavy-flavor particles.
Figure 4.1 shows the available signal mass points for X and S for the $X \to SH \to b\bar{b}WW$ in the 1 lepton final state. The dataset names can be found in Appendix A.2.

m_X in TeV / m_S in GeV	1	1.5	2	2.5	3
400	✓				
500	✓	✓			
750	✓	✓	✓		
1000		✓	✓	✓	✓
1500			✓	✓	✓
2000				✓	✓
2500					✓

Table 4.1.: Overview of the available mass points for the signal samples.

4.1.2. Background Samples

The background samples used in this thesis can be separated into two categories: prompt and non-prompt backgrounds. The prompt backgrounds are backgrounds that contain prompt leptons, which have their origin in leptonic decays of heavy bosons or top quarks. In hadronic processes, so-called non-prompt leptons emerge. These are typically not-well isolated, and their vertex is distinguishable from the primary vertex. Hence, hadronic backgrounds are referred to as non-prompt backgrounds. In this thesis, all-hadronic $t\bar{t}$, all-hadronic W+jets and dijet processes make up the non-prompt background. The prompt backgrounds considered are $t\bar{t}$, W+jets, Z+jets, single top and diboson.

The samples used for $t\bar{t}$ with one lepton in the final state are generated using the Powheg v2. The parton shower is modeled using Pythia 8.308, whilst heavy flavor generation is done with Evtgen 2.1.1. The W+jets samples are modeled using Sherpa 2.2.14 a tune NNPDF3.0nnlo at next-to-next-to-leading order. The same configuration is used for the generation of the Z+jets samples. For the single top contribution samples generated with Powheg v2 are used. Their parton shower is simulated with Pythia 8.308 and for heavy flavor decays Evtgen 2.1.1 is used. The samples used for the diboson background are generated using Sherpa 2.2.14 with a NNPDF3.0 tune at next-to-next-to-leading order. For the non-prompt samples, all-hadronic $t\bar{t}$ are generated with Powheg v2. The parton showera are simulated using Pythia 8.308, while heavy flavor decays are modeled with Evtgen 2.1.1. The all-hadronic W+jets samples are produced with Sherpa 2.2.14 and a NPDF3.0 tune at next-to-next-to-leading order. The parton distribution functions LHA is applied. The generation of the dijet samples uses Pythia 8.308 with a A14 tune at leading oreder and the NNPDF2.3LO parton distribution functions. Heavy flavor decays are simulated using Evtgen 2.1.1.
A complete list of the datasets used can be found in Appendix A.3.

4.2. Analysis Objects

The objects used in this analysis are defined as listed in Table 4.2.

	Electrons	Muons	Small-R jets	Large-R jets
Input	-	-	EM PFlow objects	UFO jets
Radius	-	-	0.4	1.0
p_T^{min}	10 GeV	10 GeV	50 GeV	250 GeV
$\|\eta\|^{\text{max}}$	2.47	2.5	4.5	2.0
Identification	`Tight`	`Medium`	-	-
Isolation criteria	`Tight_VarRad`	`PFlowTight_VarRad`	-	-

Table 4.2.: Overview of the objects used in the analysis.

Here, it is important to remark that the identification and isolation requirements for leptons are the results of the lepton selection study in Chapter 5. For further information on identification and isolation, the reader is encouraged to look into the lepton selection study.

Tau leptons are not considered in this analysis, since they decay further leptonically or hadronically after a short amount of time. The hadronically decaying tau leptons are neglected in the analysis, while the leptonically decaying ones are reflected in the analysis through their decay products.

In Run 3, EM PFlow objects (Electromagnetic scale Particle Flow) are used as input for the small-R jets. These objects combine charged particle tracks and calorimeter energy deposits for improved jet accuracy. To be more precise this analysis uses the `AntiKt4EMPFlowJets` collection, which contains jets clustered by the anti-k_t algorithm with a radius parameter of $R = 0.4$ [37]. For large-R jets unified flow objects (UFO) from the `AntiKt10UFOCSSKSoftDropBeta100Zcut10Jets` container are used [38]. The jets in this container are again clustered by the anti-k_t algorithm using a radius parameter of $R = 1.0$ and afterwards are groomed using the SoftDrop [39] algorithm with $\beta = 1.0$ and $z_{\text{cut}} = 0.1$. Similar to the EM PFlow objects, they use a combination of charged particle tracks and calorimeter energy deposits.

4.3. Object Reconstruction

To reconstruct the final-state objects of the $X \rightarrow SH \rightarrow b\bar{b}WW$ decay channel, different techniques are used. First, an overlap removal ensures that objects are not double counted in the analysis (see Section 4.3.1). Furthermore, important objects in the analysis such as the H_{bb} and W_{had} candidate are found using taggers, which are described in Section 4.3.2. These taggers can determine the origin of a given jet, which is also used to distinguish jets stemming from decaying to quarks.

4.3.1. Overlap Removal

Since the object reconstruction algorithms run independently of each other on all tracks and energy clusters from the calorimeter, geometrically overlapping objects might be double counted leading to a distortion of the analysis. To avoid this instance, an overlap removal between electrons, muons and jets is implemented. The tool used in this analysis [40] removes overlaps between electrons and muons that share a track. For the overlap removal between muons and jets, first non-b-tagged jets that overlap with muons are removed in a cone around the muon with a size of $\Delta R = 0.2$. After that, an outer cone with a size of $\Delta R = 0.4$ is defined around all the remaining non-pileup jets and all overlapping muons are discarded. The overlap removal between electrons and jets is done analogously with the same cone sizes [41].

4.3.2. Tagging

In this analysis multiple flavor taggers are used to assist in the preselection as well as the region definitions. The following sections will describe these taggers.

Small-R Jet b-tagging

The b-Tagging of the small-R jets uses the `DL1d` tagger [42, 43] with the 77% working point. This tagger uses a deep feed-forward neural network based on several input features. Besides the jet p_T and $|\eta|$, it also uses outputs from other subtaggers as input, namely IP2D/IP3D, SV1, JetFitter and JetFitter c-tagging. IP2D/IP3D is a tagging algorithm that uses a likelihood ratio between different flavor jet hypotheses based on track impact parameters to determine the jet flavor. The SV1 algorithm and the JetFitter algorithm both reconstruct secondary vertices. For additional properties of the secondary vertices the JetFitter c-tagging algorithm can be used.

The `DL1d` tagger uses in contrast to other `DL1` taggers also the output of the Deep Impact Paramet Set sub-tagger (DIPS) as input variables. From these variables, the network calculates a discriminant defined as

$$D_{\text{DL1}} = \log \frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_u} \quad , \tag{4.1}$$

where p_b, p_c and p_u are the probabilities how likely the jet is to be a b-jet, a c-jet or a light-flavour jet, respectively. f_c is the effective c-jet fraction in the background training sample.

H_{bb} *b*-tagging

The identification of the Higgs boson decaying into a pair of b quarks is crucial for a proper event selection. In this analysis, the latest GN2X algorithm [44] from ATLAS is used for the b-tagging of the large-R jets. With the help of this transformer neural network based tool, it is determined whether a large-R jet originates from a boosted Higgs boson decay into $b\bar{b}$ pairs. Jets are required to have $m > 50\,\text{GeV}$, $p_T > 250\,\text{GeV}$ and $|\eta| < 2.0$.
Using the information of the charged particle tracks of the large-R jet, the network calculates probalities on how likely the large-R jet is an H_{bb} jet, H_{cc} jet, top jet, or a multijet event. From these probabilities a discriminant is defined as :

$$D_{\text{GN2X}}^{\text{Hbb}} = \ln\left(\frac{p_{Hbb}}{f_{Hcc} \cdot p_{Hcc} + f_{\text{top}} \cdot p_{\text{top}} + (1 - f_{Hcc} - f_{\text{top}}) \cdot p_{\text{QCD}}}\right) \tag{4.2}$$

with the probabilities p_{Hcc}, p_{Hcc}, p_{top}, p_{QCD}. f_{Hcc} and f_{top} being free parameters, which the weight of the probabilities of a H_{cc} jet or a top jet with respect to a multijet event. The optimized values for f_{Hcc} and f_{top} are 0.02 and 0.25. The discriminant is pictured in Figure 4.2.
This tagger offers a variety of working points which correspond to cuts on the discriminant. These working points are given in percentages and refer to the H_{bb}-tagging efficiency [44].

Newest developments of this tagger allow the tagger to be mass-dependent. This feature was introduced to avoid mass sculpting of the efficiency. Instead of the fixed cut working point given in percentages, this tagger uses flat mass working points, which are given in decimal numbers that correspond to the flat efficiency [45, 46].

WZ Tagger

The WZ Tagger [47] in the version `SmoothedWZ` is used for the identification of hadronically decaying W bosons in large-R jets in this analysis. This tagger uses rectangular cuts on the jet mass, the number of ID tracks with $p_T > 500\,\text{MeV}$ associated with the jet before grooming and the energy-correlation function D_2 [48] with $\beta = 1$. Large-R jets need to fulfil $p_T > 200\,\text{GeV}$, $m > 40\,\text{GeV}$ and $|\eta| < 2.0$ to be considered by the tagger. The tagger then provides several working points that can be selected by the user.

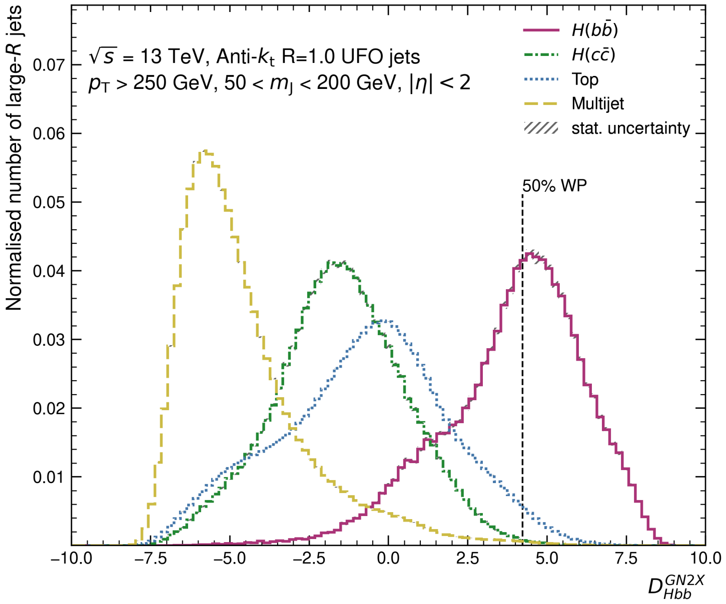


Figure 4.2.: Discriminant of the GN2X tagger with $f_{top} = 0.25$ and $f_{Hcc} = 0.02$ for the four jet classes. Working points are derived by imposing a threshold on the discriminant distribution of the H_{bb} jets to achieve the desired signal efficiency. The discriminant is defined in Equation 4.2. This Figure is taken from Ref. [44].

Top Tagger

In this analysis, the ATLAS `DNNTopTagger` [49] is used to distinguish jets that originate from top quarks from other large-R jets. This tagger is based on a deep neural network (DNN), which is optimized to determine which jets were produced by the hadronic decay of a top quark, specifically for contained tops.

For the training, only jets with a groomed reconstructed jet mass larger than 40 GeV and a number of constituents of at least 3 are considered in the DNN. To improve the distinction between top-jets and quark-/gluon-initiated jets, the DNN training uses jet substructure moments as input. The tagger requires large-R jets to fulfill $p_T > 350\,\text{GeV}$, $m > 40\,\text{GeV}$ and $|\eta| < 2.0$. This tagger is especially useful for this analysis to veto top background in region where they are not useful, for instance in the QCD and W+jets control regions and also in the definition of the top control region.

4.4. Event Selection

4.4.1. Preselection

The preselection for the split-boosted 1-lepton analysis starts by checking whether the events pass the triggers. Here, a large-R jet trigger is used. This trigger fires, if at least one large-R jet is present in the event. After that, the transverse momentum of the leading large-R jet is required to be larger than 500 GeV offline. This cut ensures that the triggers work properly and are well within the trigger plateau. Furthermore, every event is required to have exactly one signal lepton. The signal lepton is defined as listed in Table 4.2. Next, a cut on the number of large-R jets is placed, which requires events to have at least two large-R jets. This is needed in the next step, where the H_{bb} and the W_{had} candidate are classified. This classification first assigns the H_{bb} candidate as the jet with a mass closest to the Higgs mass $m_H = 125\,\text{GeV}$ and then the W_{had} candidate is found using the `SmoothedWZ` tagger at an 80% working point [47]. To further refine the selection, the preselection demands a H_{bb} and a W_{had} candidate to be present in the event.

To reflect the characteristic large separation between the signal lepton and the W_{had} jet of the split-boosted topology, a cut on the ΔR between the lepton and the closest jet is placed, which demands $\Delta R(\ell, \text{closest jet}) > 1$.

The background composition after preselection is shown in Figure 4.3 for the partial Run 3 dataset. It can be clearly seen that the background is largely dominated by W+jets and $t\bar{t}$ contributions, followed by the non-prompt contribution. Also, the background yields after preselection are listed in Table 4.3. For signal, a cross-section of 100 fb was assumed and the resulting yields are pictured in Figure 4.5. The values of the predicted signal yields can additionally be found in Appendix A.6. These yields only reflect the part of Run 3, that was recorded during 2022 with a luminosity of $26.07\,\text{fb}^{-1}$. For the expected total luminosity of Run 3 of $250\,\text{fb}^{-1}$ the yields need to be multiplied by a factor of 9.5. The wedge of the pie chart named “non-prompt” is composed of dijet as well as all-hadronic $t\bar{t}$ and all-hadronic W+jets. A full list of the used samples can be found in the Appendix A.3.

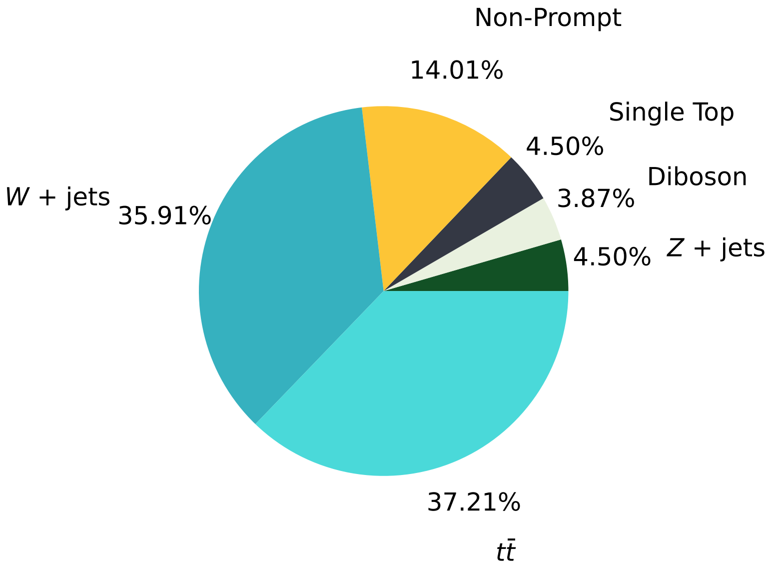


Figure 4.3.: Background composition after preselection for partial Run 3 for a luminosity of $26.07\,\mathrm{fb}^{-1}$. The non-prompt contribution is made up of Dijet, all-hadronic $t\bar{t}$ and all-hadronic W+jets.

	event yield
$t\bar{t}$	3627.20
W+jets	3500.00
Single Top	438.89
Z+jets	438.51
Diboson	377.32
Dijet	1325.07
all-hadronic $t\bar{t}$	32.51
all-hadronic W+jets	8.36

Table 4.3.: Background yields after preselection for partial Run 3 corresponding to a luminosity of $26.07\,\mathrm{fb}^{-1}$.

4.4.2. Preliminary Definition of the Signal Region

On top of the preselection explained in Section 4.4.1, the H_{bb} candidate is additionally required to pass the working point of the mass-dependent GNN Tagger, that is equivalent to 75% of efficiency (see Section 4.3.2) to fulfill signal region requirements. [1]

With this definition, the background composition of the signal region has the contributions shown in Figure 4.4. In the signal region, the dominating contribution is the $t\bar{t}$ background followed by single top and W+jets. The non-prompt background only contributes about 7% to the background of the signal region.

The yields in the signal region for the different background contributions can be found in Table 4.4. The signal yields are pictured in Figure 4.5 and a table with the values can be found in Appendix A.6. All the yields were calculated for partial Run 3 with a luminosity of 26.07 fb^{-1} and for the signal luminosity, a value of 100 fb was assumed.

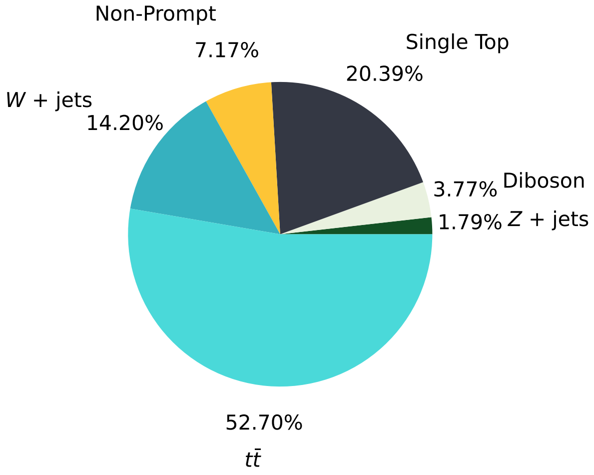


Figure 4.4.: Background composition in the signal region for partial Run 3 for a luminosity of 26.07 fb^{-1}. The non-prompt contribution is made up of Dijet, all-hadronic $t\bar{t}$ and all-hadronic W+jets.

---

[1] The 75% efficiency working point of the mass-dependent working point is named 0.94. For further information on this, the reader is referred to [45, 46].

	event yield
$t\bar{t}$	11.45
W+jets	3.08
Single Top	4.43
Z+jets	0.39
Diboson	0.82
Dijet	1.56
all-hadronic $t\bar{t}$	0.00
all-hadronic W+jets	0.00

Table 4.4.: Background yields in the signal region for partial Run 3 corresponding to a luminosity of 26.07 fb^{-1}.

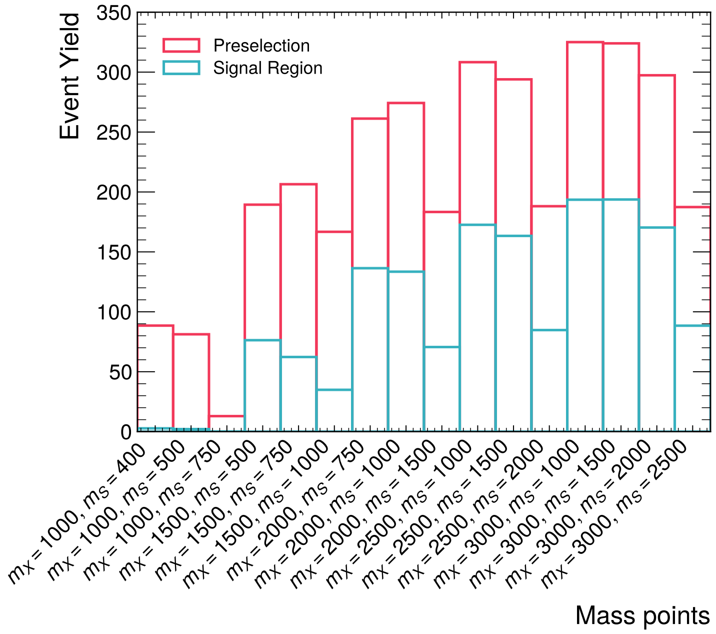


Figure 4.5.: Signal yields after preselection and in the signal region for partial Run 3 corresponding to a luminosity of 26.07 fb^{-1}. For the signal a cross-section of 100 fb was assumed. The numbers can also be found in Appendix A.6.

5. Study of the lepton identification and isolation working points

In this chapter a study of the lepton identification and isolation working points will be presented. The goal of this study is the optimization of the categorization of signal leptons based on lepton identification and isolation. This categorization is used throughout the analysis in the definition of electrons and muons as shown in Table 4.2.
Non-prompt leptons are leptons produced in the decay of hadrons. Since these hadrons have already traveled a distance in the detector when they emit the lepton, the vertex where the non-prompt lepton originates from, is distinguishable from the primary vertex. In contrast to prompt leptons, which stem from decays of heavy bosons and top quarks, non-prompt leptons are difficult to model via Monte Carlo simulations because of their hadronic origin.

5.1. Lepton Identification

One of the quantities used to define leptons is the lepton identification (ID). This quantity represents how likely it is that a given lepton candidate is in fact a muon or an electron, respectively. The lepton ID is defined differently for muons and electrons. For muons a cut-based method is used. The corresponding cuts to the working points are listed in Table 5.1 [50]. The variable $N^{\text{precision}}_{\text{layers}}$ refers to the number of precision layers hit by the muon candidate in the detector. Similarly, $N^{\text{precision}}_{\text{HoleLayers}}$ is the number of precision hole layers hit. `isAuthor(MuGirl)` and `isAuthor(MuTagIMO)` are boolean variables that describe whether or not the muon was reconstructed with the `MuGirl` and `MuTagIMO` algorithms, respectively. The q/p-significance $\Sigma_{q/p}$ is defined as [50]

$$\Sigma_{q/p} = \frac{|(q/p)_{\text{ID}} - (q/p)_{\text{MS}}|}{\sqrt{\sigma^2_{\text{ID}} + \sigma^2_{\text{MS}}}} \quad . \tag{5.1}$$

Here, $(q/p)_{\text{ID}}$ refers to the ratio of charge and momentum measured in the Inner Detector, whilst $(q/p)_{\text{MS}}$ is the ratio of charge and momentum measured in the muon spectrometer.

σ_{ID} and σ_{MS} are the corresponding uncertainties [50].

Working Point	Definition
Medium	$\Sigma_{q/p} < 7$ & $(N^{\mathrm{precision}}_{\mathrm{layers}} > 1$ or $(N^{\mathrm{precision}}_{\mathrm{layers}} == 1$ & $N^{\mathrm{precision}}_{\mathrm{HoleLayers}} < 2$ & $\|\eta\|<0.1$))
Loose	As for Medium or ($p_T < 7$ GeV & $\|\eta\| < 1.3$ & $N^{\mathrm{precision}}_{\mathrm{layers}} > 0$ & `isAuthor(MuGirl)` & `isAuthor(MuTagIMO)`)

Table 5.1.: Definitions of the muon ID working points [51].

For electrons the ID working points are based on a likelihood function, which is used to distinguish electrons stemming from prompt processes from light-flavor jets, electrons from photon conversions, and non-prompt electrons from the decay of hadrons. This likelihood uses different variables that describe properties of the showers in the calorimeter as well as properties in the tracking detector. Examples for that are the transverse impact parameter relative to the beam-line d_0, the number of hits in the pixel detector and variables describing the shower shape. A detailed list of the variables used as input for the likelihood function can be found in Ref. [52]. An updated description of the calorimeter variables that enter the likelihood function, is provided in Ref. [53].

5.2. Lepton Isolation

Isolation is a measure of how many particles are produced in the surroundings of a particle. This helps in the distinction between prompt and non-prompt leptons. Prompt leptons are expected to not have other particles close to them and are therefore expected to be well isolated. This, however, does not apply to non-prompt leptons [54]. Isolation variables can be defined based on calorimeter or tracking information or a combination of both.

5.2.1. Tracking isolation variables

The basic principal of a tracking based isolation variable consists of a cone of radius R around a particle track. The transverse momenta of all the tracks within the cone are summed up and divided by the transverse momentum of the particle by which the cone was defined. Apart from fixed radius variables, there are also isolation variables which use variable cones. In this case the radius is dependent on the transverse momentum of the particle such that the isolation cone shrinks with increasing transverse momentum.

The radius is then given by

$$R_{\text{var}} = \min\left(\frac{10\,\text{GeV}}{p_T}, R\right) \tag{5.2}$$

Variable cones are used in the isolation variables as `ptvarcone`, whereas fixed cones are labeled as `ptcone`. These are then followed by a two digit number such as 10,20,30, which refers to a radius $R = 0.1, 0.2, 0.3$, respectively [55].
Furthermore, the variable names include `Nonprompt_All_MaxWeightTTVA` where `TTVA` refers to the track-to-vertex association and `Nonprompt_All_MaxWeight` is the working point. This working point checks that the track has either been used in a vertex fit and has a maximum weight or has not been used in a vertex fit but fulfills $|\Delta z\,\sin\theta| < 5\,\text{mm}$ and $|d_0| < 5$. Here $|\Delta z\,\sin\theta|$ is the absolute value of the distance between the z_0 of the track and the primary vertex and $|d_0|$ is the tranverse impact parameter with respect to the beam line [55].
The variable name can also contain a requirement on the minimum tranverse momentum of the tracks, denoted as for instance `pt1000`, where the number indicates the minimum p_T in MeV. Additionally, some variables include the tag `LooseCone` which indicates that tracks that fall within a window of $|\Delta\eta| > 0.01$ are excluded [55].
Track based isolation variables take only charged particles into consideration [54].

5.2.2. Calorimeter based isolation variables

For the calorimeter based isolation variables, a cone around a calorimeter cluster is defined. The transverse energy from clusters within this cone is summed up and divided by the transverse energy of the defining cluster.
For calorimeter based isolation variables there are different possibilities of using the calorimeter information. In `etcone` the transverse energy of all calorimeter cells is summed up. For `topoetcone` the sum of the topoclusters within the cone is used. Additionally there is the `eflowcone` which only uses clusters that could not be matched to jets.

5.2.3. Recommended isolation working points

The recommended working points [56] for muons and electrons are listed in Table 5.2. For muons all the recommended working points use a combination of tracking and calorimeter based variables. Similarly, the electron isolation working points `Loose_VarRad` and `Tight_VarRad` are using both tracking and calorimeter based isolation variables. As the name suggests, the `TightTrackOnly_FixedRad` and the `TightTrackOnly_VarRad` only rely on tracking based isolation variables, whilst `HighPtCaloOnly` only takes calorimeter information into account.

Isolation	
Muons	**Electrons**
Loose_VarRad	Loose_VarRad
Tight_VarRad	Tight_VarRad
PflowLoose_VarRad	TightTrackOnly_VarRad
PflowTight_VarRad	TightTrackOnly_FixedRad
-	HighPtCaloOnly

Table 5.2.: Recommended isolation working points for Run 3 [56].

The definitions of all these working points can be found in the Appendix A.1.
To illustrate the meaning of these working point definition, the example `Loose_VarRad` and `Tight_VarRad` will be explained in more detail. These two working points are defined for muons as

$$\texttt{Loose\_VarRad} = p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}} < 0.15 \& E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}} < 0.3 \tag{5.3}$$

$$\texttt{Tight\_VarRad} = p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}} < 0.04 \& E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}} < 0.15 \quad . \tag{5.4}$$

The distributions of the variable $p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}}$ and the variable $E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}}$ are shown in Figure 5.1 for a signal mass point of $m_X = 2000\,\mathrm{GeV}$, $m_S = 1500\,\mathrm{GeV}$, including the cuts used in the definition of the `Loose_VarRad` and `Tight_VarRad` working points. All muons below the cut fulfill the isolation working point.

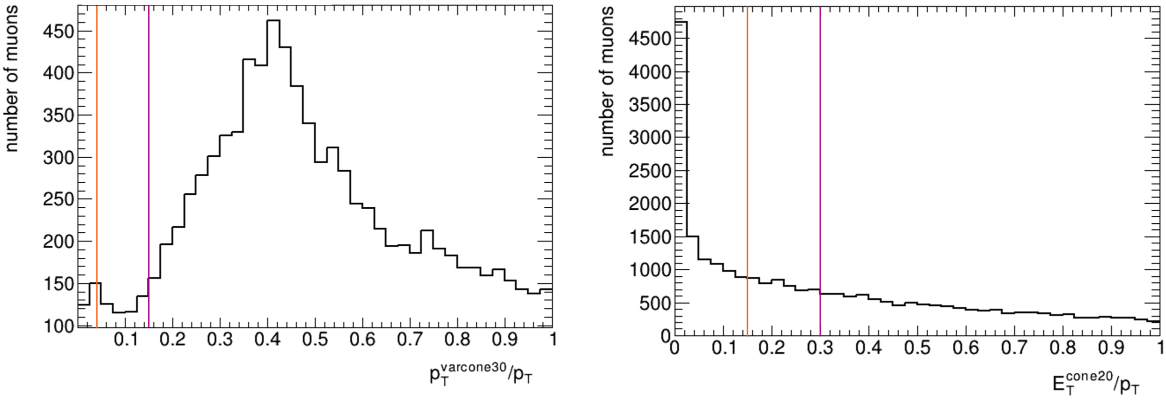


Figure 5.1.: On the left the distribution of $p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}}$ is plotted for muons of a signal sample for $m_X = 2000\,\mathrm{GeV}$, $m_S = 1500\,\mathrm{GeV}$. Analogously, the right plot shows the distribution of $E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}}$. The cuts from the working point definition are marked in the plot in purple (`Loose_VarRad`) and orange (`Tight_VarRad`) according to Equation 5.3.

5.3. Signal efficiencies

To study the effect the isolation and ID working points have on the signal, all available signal samples for 16 different mass points ranging from $m_X = 1\,\mathrm{TeV}$, $m_S = 0.4\,\mathrm{TeV}$ to $m_X = 3\,\mathrm{TeV}$, $m_S = 2.5\,\mathrm{TeV}$ were used. A list of all available mass points can be found in Section 4.1.1. For each sample, it is determined how many of the leptons pass the ID and isolation working points. The efficiency ε is calculated for every combination of ID and isolation working point by dividing the number of leptons passing the working point by the total number of leptons. The studied working points are listed in Table 5.2. In the beginning, no selection and no overlap removal are applied to the objects and events. Investigations with further selection criteria are detailed in Sections 5.5 and 5.6. The only cuts used in this study are p_T and $|\eta|$ from the object definitions (see Table 4.2) in Section 4.2.

In the signature of the $X \to SH \to b\bar{b}WW$ decay channel, one important feature is the signal lepton that originates from the decay of a W boson. To distinguish these prompt signal leptons from other non-prompt leptons, promptness is required for the signal efficiency. This is done with a tool that consideres the truth information from the Monte Carlo simulation. In Figure 5.2, the number of leptons passing each ID working point are depicted. Especially for electrons, it can be seen that the number of all electrons is very high in comparison to the ones passing the working points (see top left plot of Figure 5.2). In the `All` category lie all particles that are labeled as electrons without any identification applied. The high number of electrons in this bin indicates that a lot of these electrons are not signal leptons and therefore not useful for this study. Thus, it is useful to require the electrons to be prompt for the study of the signal. The same line of reasoning applies to the muon case as well, as it can be seen in the top right plot of Figure 5.2). The bottom row of Figure 5.2 show the number of prompt electrons and prompt muons passing the ID working points, respectively.

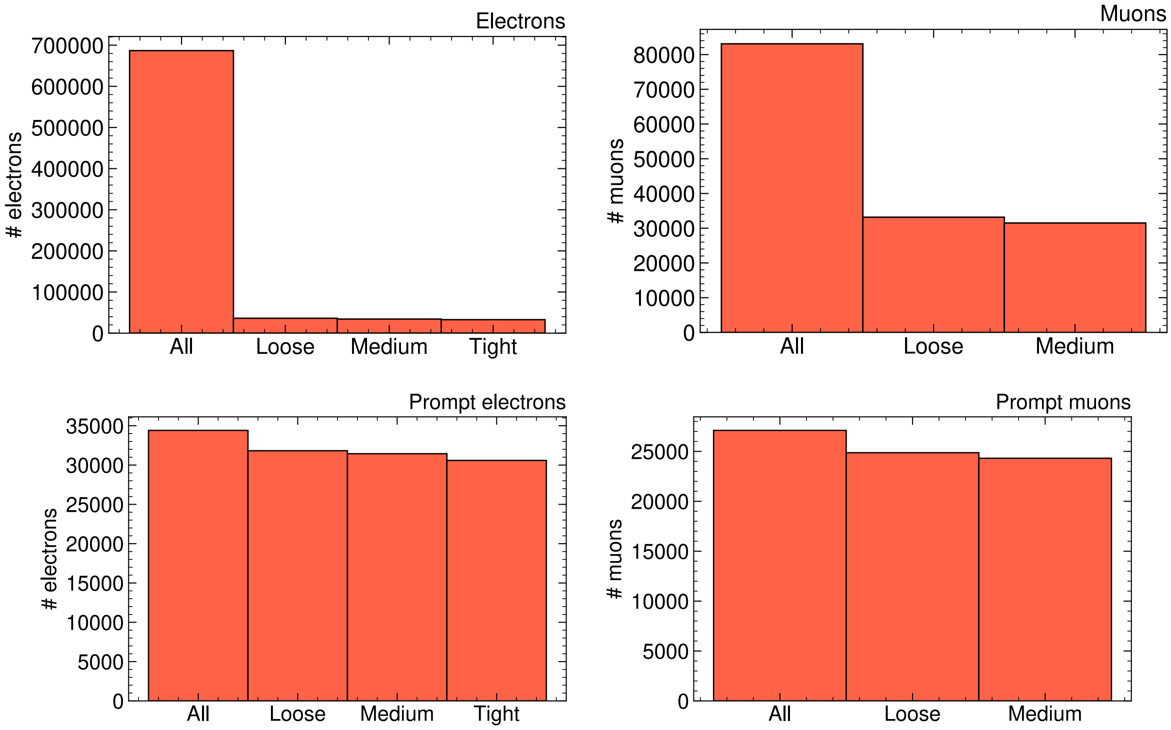


Figure 5.2.: Number of leptons passing each of the ID working points for all electrons (top left), all muons (top right), prompt electrons (bottom left) and prompt muons (bottom right). For the prompt leptons a tool was used to determine the promptness of the leptons.

The signal efficiencies for the different ID working points can be seen in Figure 5.3. Effiencies are also calculated for all combinations of ID and isolation working points. To summarize the main behaviors, Figure 5.4 shows the signal efficiencies for two example mass points for muons and electrons. The signal efficiencies for all mass points can be found in the Appendix A.4.

Furthermore, the kinematic dependency of the efficiencies was tested. For electrons, Figure 5.5 shows the kinematic dependency of the ID and isolation working points. For the right plot, the tight ID working point was applied. It can be clearly seen that the ID as well as isolation all plateau towards high tranverse momenta. In the electron isolation

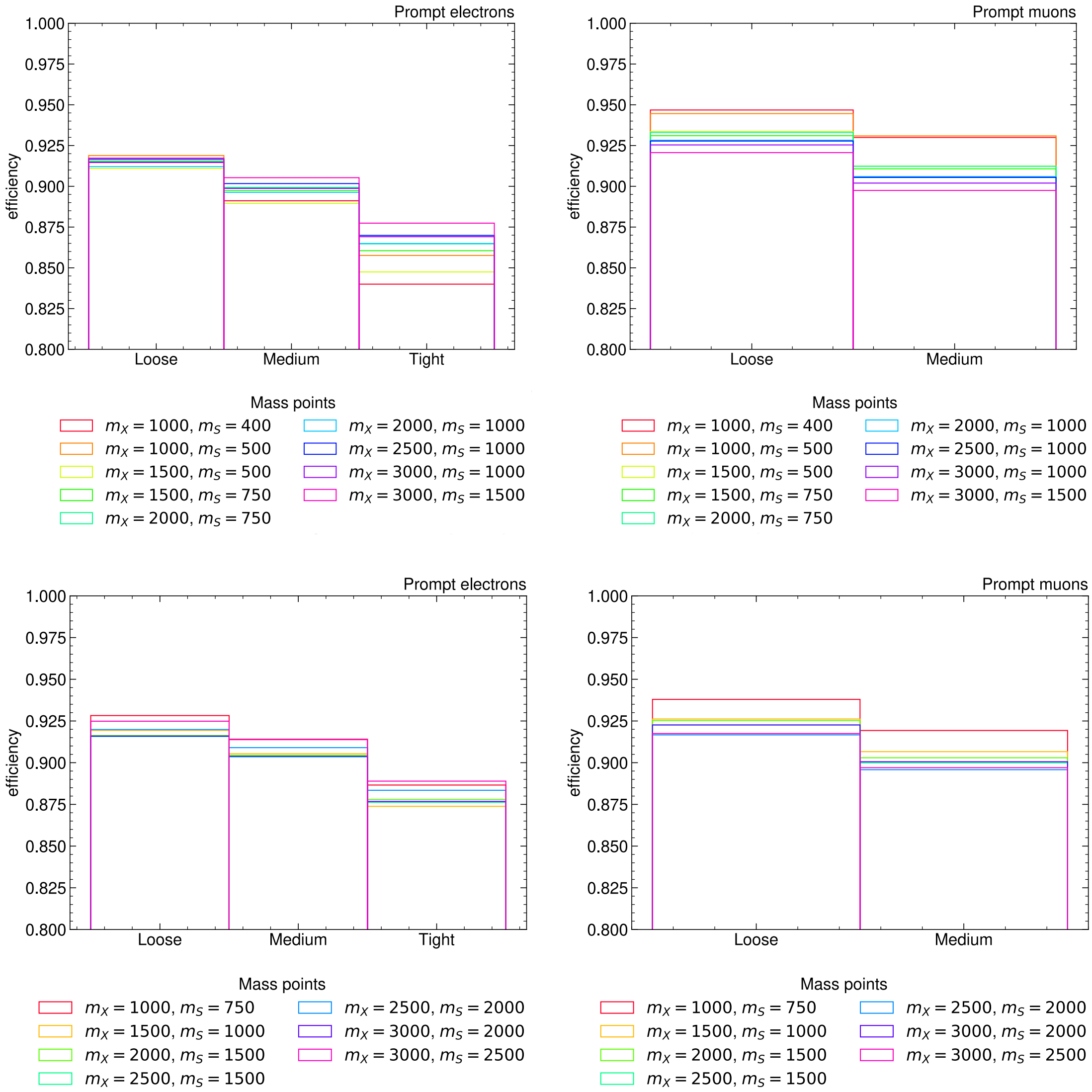


Figure 5.3.: The plots show the signal efficiencies for the different ID working points. The masspoints were split up into low mass ratio between S and X in the top row and high mass ratio in the bottom row. The left column depicts the electrons while the right column corresponds to muons.

it becomes obvious that the `HighPtCaloOnly` working point loses efficiency towards high transverse momenta, especially in comparison to other working points, which lie closely together. For muons, the equivalent distributions are depicted in Figure 5.6. To compare the isolation working points, a medium ID was applied. Again, both distributions plateau towards high transverse momenta. Differences between the working points are only small for muons in comparison to electrons.

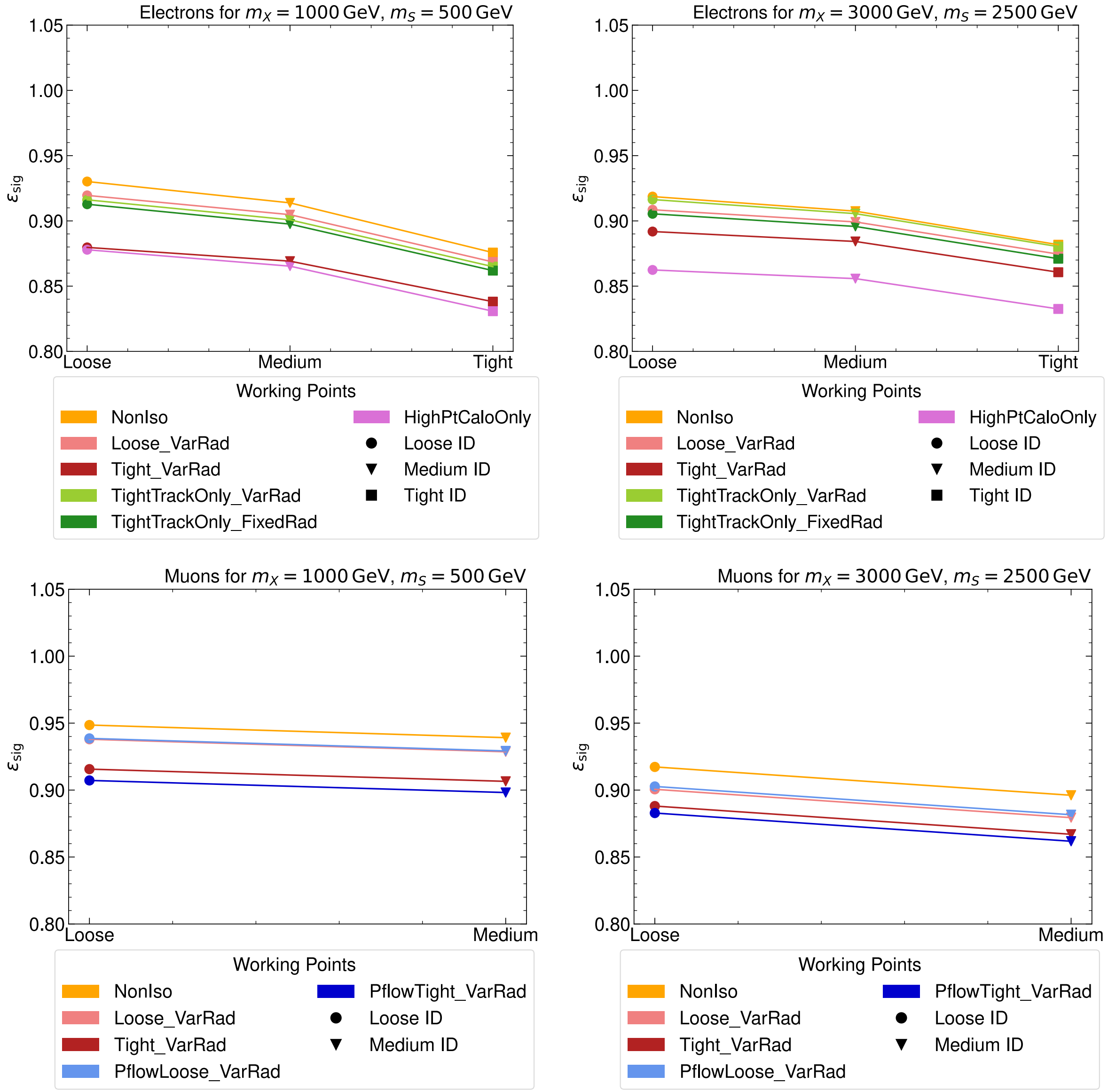


Figure 5.4.: Signal efficiencies for prompt electrons (upper row) and prompt muons (lower row). Two different masspoints are examined. The left column shows the mass point $m_X = 1\,\mathrm{TeV}$, $m_S = 0.5\,\mathrm{TeV}$, while the right column shows the mass point $m_X = 3\,\mathrm{TeV}$, $m_S = 2.5\,\mathrm{TeV}$. The x-axis displays the ID working points and the isolation working points are represented by color.

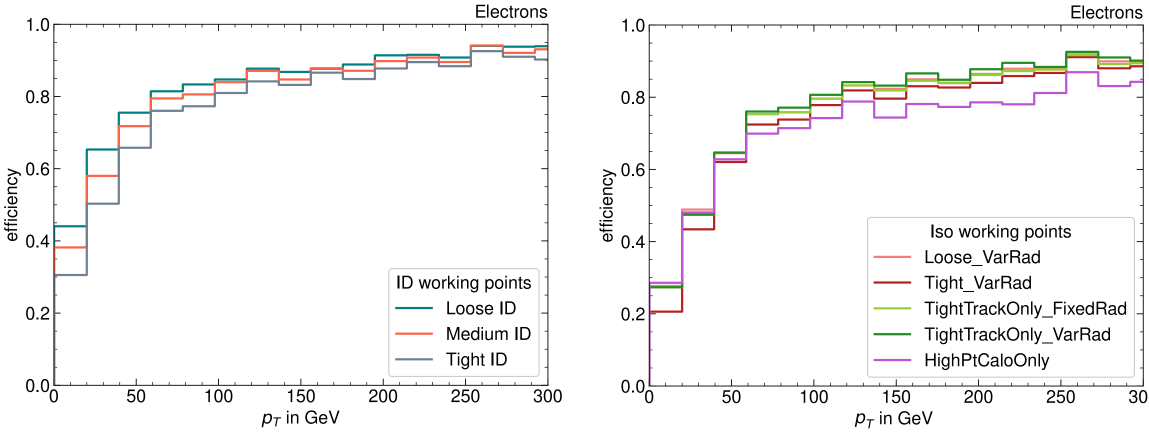


Figure 5.5.: Efficiency vs. the transverse momentum p_T of the ID working points (left) and of the isolation working points (right) for electrons for the mass point $m_X = 3\,\text{TeV},\ m_S = 2.5\,\text{TeV}$. To compare the isolation working points, the tight ID was applied.

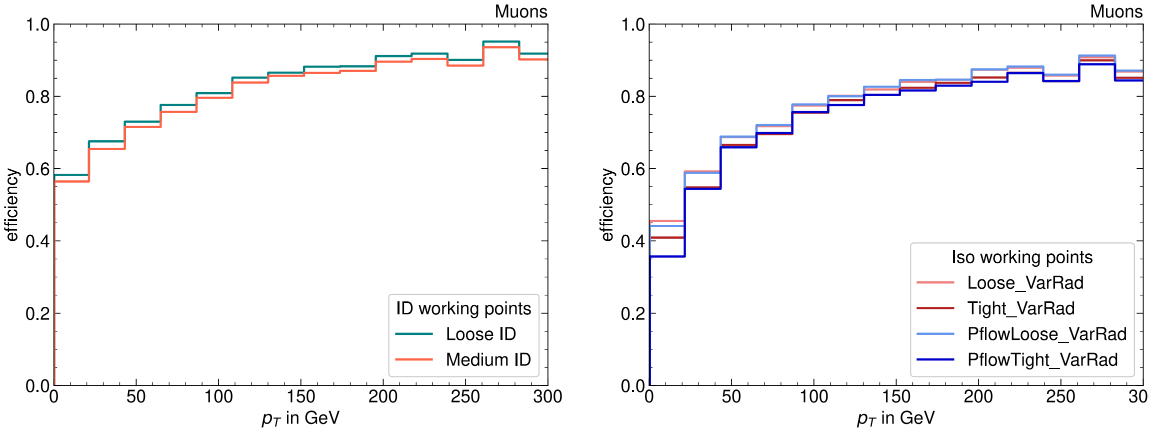


Figure 5.6.: Efficiency vs. the transverse momentum p_T of the ID working points (left) and of the isolation working points (right) for muons for the mass point $m_X = 3\,\text{TeV},\ m_S = 2.5\,\text{TeV}$. To compare the isolation working points, the medium ID was applied.

5.4. Background rejection

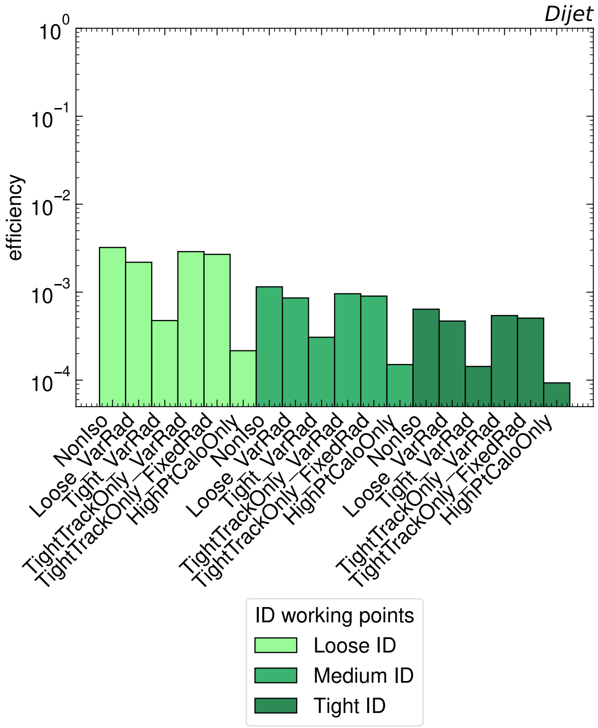
For a conclusive picture of the performance of the lepton working points, the background should also be investigated. For this study, three backgrounds are of interest: Dijet, $t\bar{t}$ and W+jets. In this thesis, the focus will be on the results of the Dijet background, since Dijet events include a significant number of non-prompt leptons. In contrast to that, the $t\bar{t}$ and W+jets samples also contain prompt leptons and therefore feature small background rejections. For completeness, background efficiency and background rejection plots for $t\bar{t}$ and W+jets backgrounds can be found in the Appendix A.5.

The background efficiencies ε_{bkg} are, similar to the signal efficiencies in Section 5.3, defined as the number of leptons passing the working point divided by the number of all leptons. In contrast to the signal, no promptness of the leptons is required in the dijet backgrounds, because this background contains almost no prompt leptons. This choice does not impact the results of the study, as long as the same conditions are applied for all working points.
The background efficiency is pictured in Figure 5.7. Background efficiencies for $t\bar{t}$ and W+jets can be found in Appendix A.5.

To see how much the ID and isolation working points surpress the background, the background rejection r_{bkg} is plotted in Figure 5.8. The background rejection is defined as

$$r_{bkg} = \frac{1}{\varepsilon_{bkg}} \quad , \tag{5.5}$$

where ε_{bkg} is the background efficiency.
To put the background rejection into perspective to the signal, Figure 5.9 depicts the background rejection as a function of the averaged signal efficiency. The averaged signal efficiency is given by the average of the signal efficiencies of all 16 mass points. This averaging can be done because the behaviour of the signal efficiencies does not change significantly between mass points, as can be seen in Figure 5.4.
The best performing working points can be found in Figure 5.9 by choosing a working point which combines high background rejection with a high signal efficiency. For muons this means that the Medium ID in combination with the `PflowTight_VarRad` isolation is the best choice to characterize signal muons. In the electron case, it was decided to use the `Tight_VarRad` isolation in combination with the Tight ID, even if the `HighPtCaloOnly` isolation has a higher background rejection. This choice comes for multiple reasons. The `HighPtCaloOnly` isolation working point does not perform as well as the other working



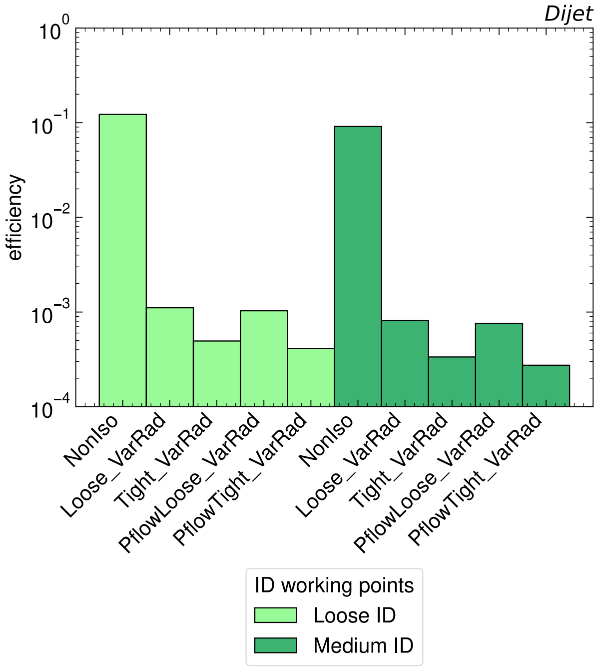


Figure 5.7.: Background efficiency for leptons in the dijet background for all the possible combinations between ID and Isolation working points. The left plot shows the background efficiencies for electrons, while the right plot pictures the efficiencies for muons.

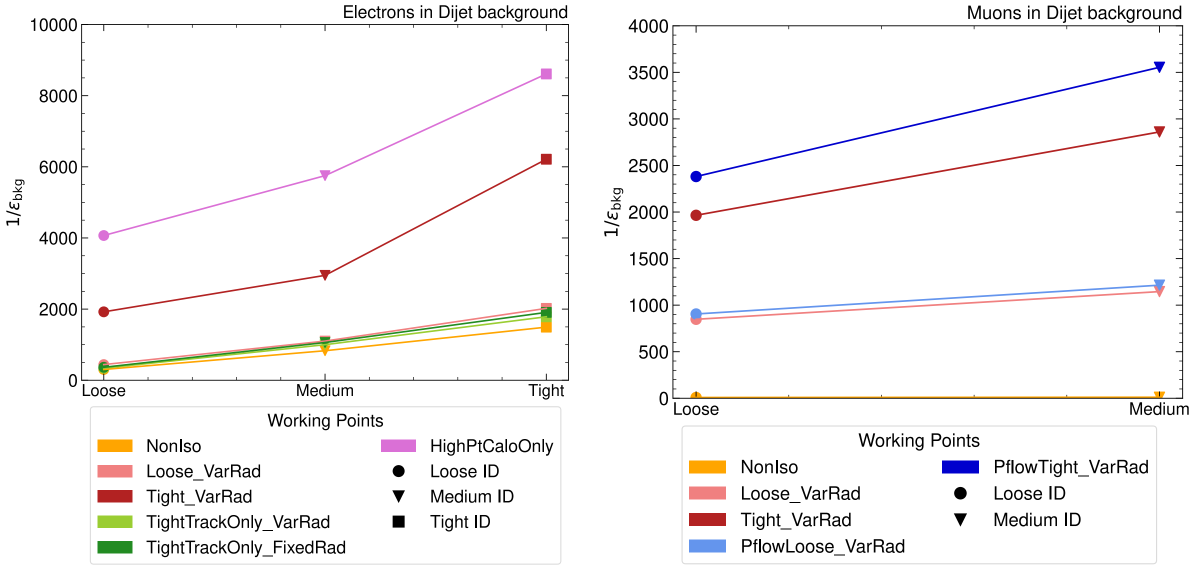


Figure 5.8.: The plots depicts the background rejection for the different ID and isolation working points in the dijet background. The background rejection is given by Equation 5.5. The left plot shows the rejection for electrons, while on the right side only muons are considered.

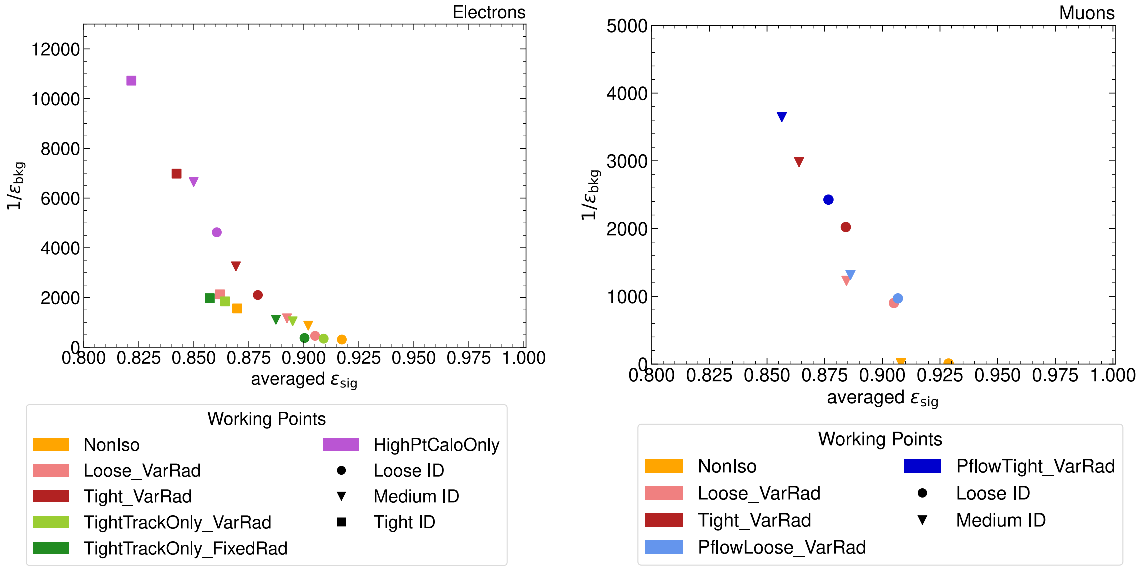


Figure 5.9.: Background rejection against the averaged signal efficiency for all combinations of ID and isolation working points. The left plot depicts the results for electrons, while the right plot shows the results for muons.

points for leptons with high transverse momenta in boosted scenarios. This behavior was investigated for the split-boosted scenario and can be observed in Figure 5.5. Another argument against this working point is the loss of signal due to the lower signal efficiency.

Due to the already very limited number of events in this decay channel, it was ultimately decided to use the safer choice of the `Tight_VarRad` isolation.

5.5. Overlap Removal

The study described up to this point was performed with no overlap removal for the leptons. To validate the results, it is necessary to repeat the study with overlap removal (OR) applied to all leptons. The criteria for OR were already explained in Section 5.5. Figure 5.10 shows the background rejection of the dijet background over the averaged signal efficiency and compares the results with and without overlap removal. Again the averaged signal efficiency refers to the average of the signal efficiencies for all 16 mass points. For muons, an applied overlap removal improves the signal efficiency, while also increasing the background rejection slightly as seen in the right plot in Figure 5.10. The same effect can be observed for electrons in the left plot in Figure 5.10 but with an even smaller impact. Both cases show that the behavior of the working points relative to each other is not impacted by the overlap removal verifying the results of the lepton study in Section 5.4.

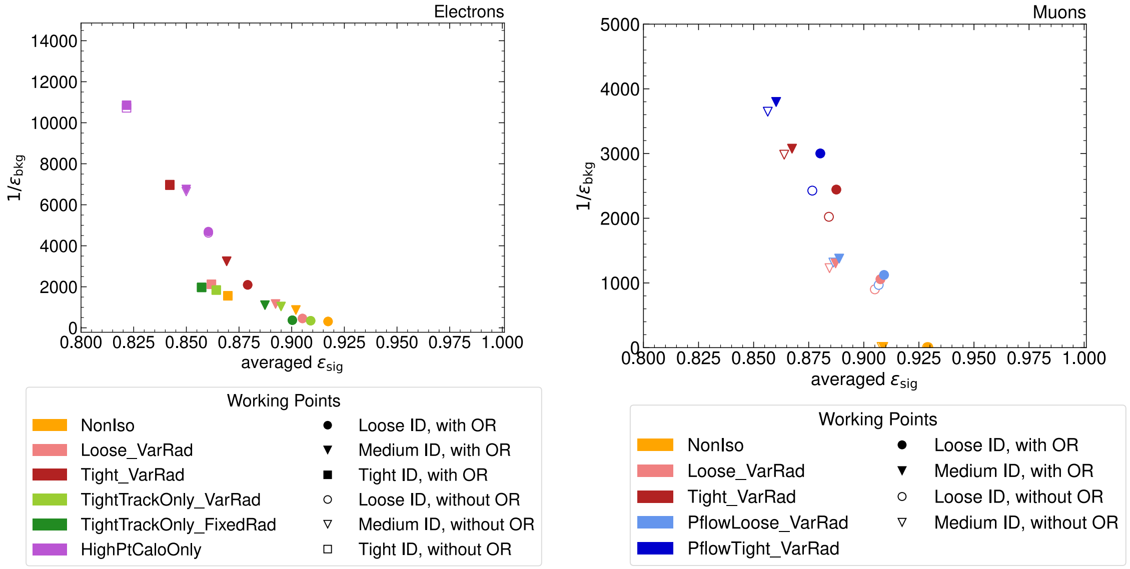


Figure 5.10.: The plots show the background rejection against the averaged signal efficiency for the available combinations of ID and isolation working points. Here the filled markers display the result with and the unfilled markers without overlap removal (OR). The left plot pictures the results for electrons and the right plot exhibits the results for muons.

5.6. Topology Specific Lepton Study

To optimize the previous study for the spit-boosted topology, a selection is implemented where the leptons are required to have a distance to the closest large-R jet of

$$\Delta R(\ell, \text{closest jet}) > 1 \quad . \tag{5.6}$$

With this cut the study described in Sections 5.3 and 5.4 is repeated. The results for the background rejections and averaged signal efficiencies are pictured in Figure 5.11. To solely focus on the effect the ΔR cut has an the performance of the working points, no overlap removal is applied.

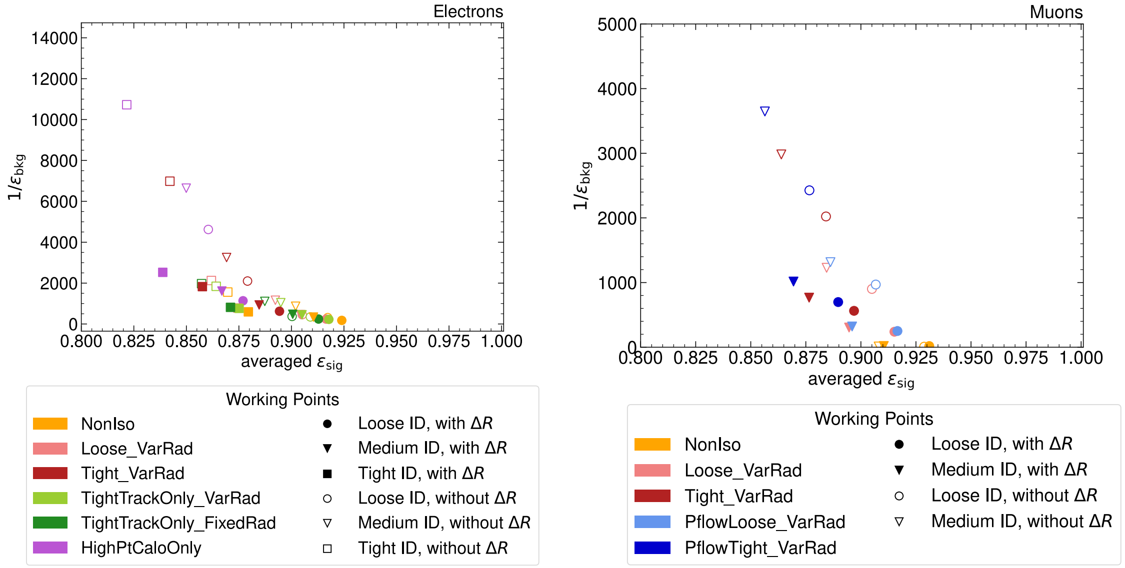


Figure 5.11.: Background rejection plotted against the average signal efficiency. The filled markers display the results with the ΔR cut (see Equation 5.6), and the unfilled markers show the results without this cut. The left plot depicts the results for electrons and the right plot shows the results for muons.

Again leptons in the signal are required to be prompt, whilst leptons in the background are not.

When working with this cut, an overlap removal for electrons needs to be implemented. Otherwise the electron itself is often falsely identified as the closest jet. This can be avoided by substracting the four-vectors of the electron and the large-R jet in case $\Delta R(\ell, \text{closest jet}) < 1$, and requiring that the jet still fulfills $p_T > 250\,\text{GeV}$, which is the threshold UFO jets need to pass.

When compared to the results in Section 5.4, the ΔR cut reduces the background rejection while increasing the signal efficiency (see Figure 5.11). The improvement of the signal efficiency is here even larger than in Section 5.5, where the effect of the overlap removal was studied.

Since the behavior of the working points relative to each other does not change, the choice for the lepton working points made in Section 5.4 is further verified.

6. Control Region Investigations for Processes with Non-Prompt and Fake Leptons

For a proper analysis, control regions for the dominant background processes need to be set up. These control regions can then be used to check the modelling and to perform background estimates, that can then be extrapolated to the signal region. The contribution of the background should be dominant or at least enhanced in the respective control region.

One important background in the split-boosted 1-lepton final state are non-prompt leptons, that fake the properties of the final-state lepton and are therefore hard to differentiate from the signal. These leptons stem from hadronic background processes and not from the decay itself. To reduce their impact on the analysis, an accurate non-prompt background estimate is desired.

For this estimate, one needs to find a control region with an enriched contribution of non-prompt leptons. This investigation is subject of Section 6.1. After the definition of the control region, the data within this region can be unblinded and the agreement between data and Monte Carlo prediction can be examined (see Section 6.2). Besides the statistical uncertainties covered in Section 6.2, some systematic uncertainties can also be inferred, which are discussed in Section 6.3.

6.1. Defining a control region

For the definition of a control region with an enriched contribution of non-prompt leptons, a set of variables is investigated to find properties that can distinguish the non-prompt background from the signal and other backgrounds. Furthermore, it is important to ensure orthogonality to other control regions and the signal region (see Section 4.4.2). The non-prompt background consists of dijet events and all-hadronic final-states from $t\bar{t}$ and W+jets processes. Other background processes that are considered in this study are $t\bar{t}$,

W+jets, Z+jets, single top and diboson production which contain at least one prompt lepton. A complete list of the used samples can be found in Appendix A.2 and A.3. In this study, a preselection is applied which is described in Section 4.4.1. For the definition of the analysis objects see Section 4.2.

In the Run 2 analysis, the following variables were used for the definition of the control regions [36] :

- Number of b-tagged TAR jets[1]
- transverse mass of the lepton and missing transverse energy combined (W_{lep})
- mass of the H_{bb} candidate and mass of the W_{had} candidate

The definitions of the control regions in Run 2 can be seen in Table 6.1.

CR	**Definition**
$t\bar{t}$	2 b-tagged TAR jets in the event, $m_{\text{TAR}}^{W_{\text{had}}} < 40\,\text{GeV}$
W+jets	0 b-tagged TAR jets in the event, $60\,\text{GeV} < m_T^{W_{\text{lep}}} < 120\,\text{GeV}$
QCD	0 b-tagged TAR jets in the event, $m_T^{W_{\text{lep}}} < 60\,\text{GeV}$ or $m_T^{W_{\text{lep}}} > 120\,\text{GeV}$

Table 6.1.: Definitions of the control region in the ATLAS Run 2 analysis [36].

The knowledge from the Run 2 analysis can be used as a starting point for Run 3. However, it should be acknowledged that object definitions changed between Run 2 and Run 3, which leads to the variables not being equivalent. Furthermore it is important that in Run 2 only the boosted topology was analyzed and region definitions might vary for the split-boosted topology.

The first useful variable is the transverse mass of the lepton and the missing transverse energy, which is defined as

$$m_T^{\text{lep}+\nu} = \sqrt{2 \cdot E_T^{\text{miss}} \cdot p_T^{\text{lep}} \cdot (1 - \cos\Delta\phi(\nu, \text{lep}))} \quad , \tag{6.1}$$

where E_T^{miss} is the missing transverse energy assumed to result from a neutrino, p_T^{lep} is the transverse momentum of the lepton and $\Delta\phi(\nu, \text{lep})$ is the angle between the E_T^{miss} and the lepton. Figure 6.1 depicts the distribution of this variable. For the non-prompt background the distribution peaks around small values, while the other backgrounds peak at

---

[1] Track assisted jets (TAR) jets are replaced by Unified Flow Objects (UFO) jets in Run 3.

higher transverse masses. The signal shows a flat distribution until $m_T^{\text{lep}+\nu} = 60\,\text{GeV}$ and slowly decreases afterwards. For the control region a cut of $m_T^{\text{lep}+\nu} < 50\,\text{GeV}$ is applied. In addition a cut on the ratio of the lepton transverse momentum and the transverse momentum of the leading large-R jet is placed. This distribution is shown in Figure 6.2. Without any cuts applied, the non-prompt background exhibits two peaks in this distribution (see left plot in Figure 6.2). However, the second peak disappears after the cut on the transverse mass of the lepton and the missing transverse energy. Thus, the cut $\frac{p_T \text{ of lepton}}{p_T \text{ of leading large-}R\text{ jet}} < 0.05$ can be used to separate the non-prompt contribution from the data and all the other backgrounds, which peak at higher values.

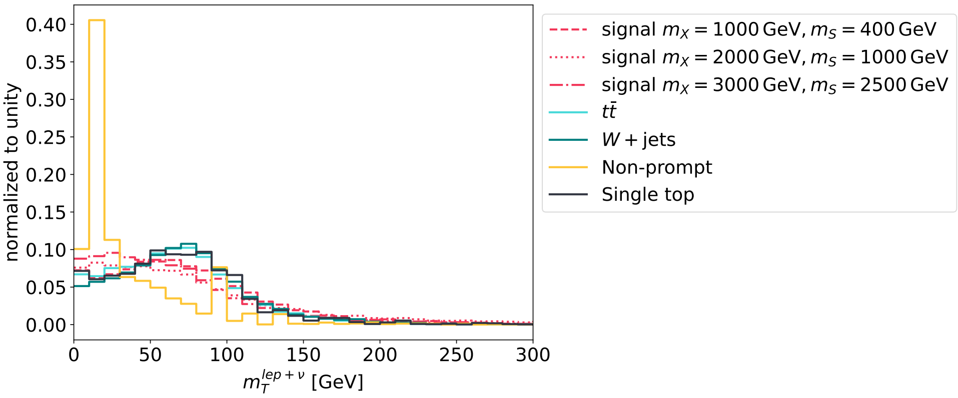


Figure 6.1.: Transverse mass of the lepton and the missing transverse energy.

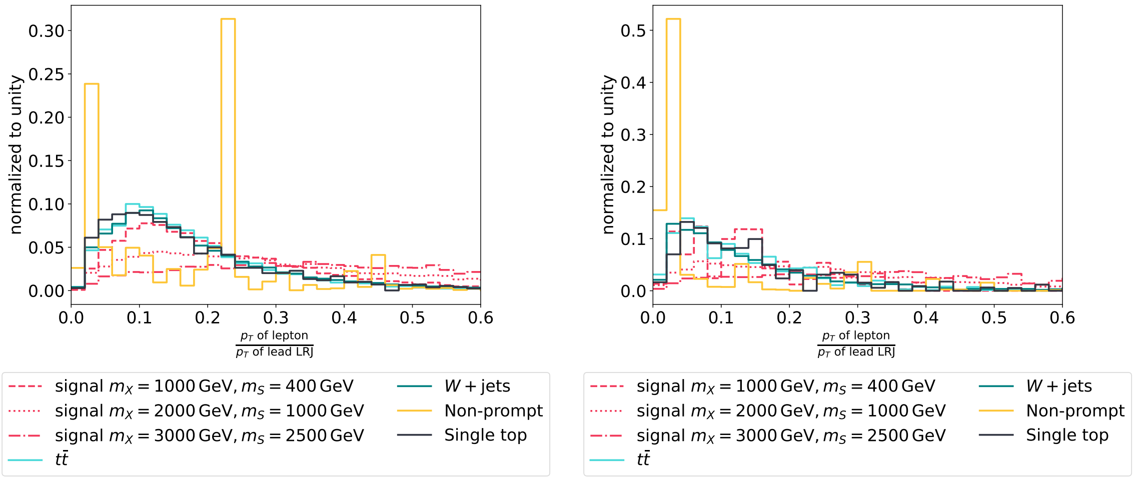


Figure 6.2.: Ratio of the lepton transverse momentum and the transverse momentum of the leading large-R jet. On the left the distribution is shown before any cuts are applied and on the right the distribution after the cut on the transverse mass of the lepton and the missing transverse energy is pictured.

Additionally, a veto on the H_{bb} b-tagging is applied to ensure orthogonality to the signal.

To provide orthogonality to the $t\bar{t}$ control region, a veto on the top tagging is implemented. The distribution of how many events pass or fail these working points is plotted in Figure 6.3.

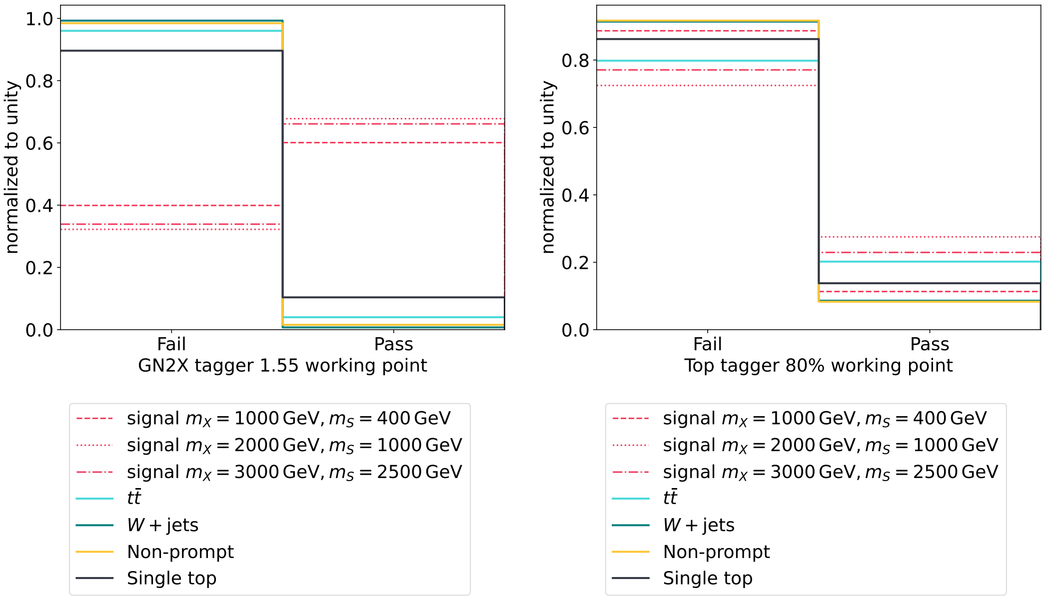


Figure 6.3.: Events passing or failing the tagging working points after preselection. On the left the GN2x tagger, which tags H_{bb} canidates, with a 1.55 working point is applied. This working point is equivalent to an efficiency of 85%. On the right the Top tagger is applied with a 80% working point.

The final definition of the non-prompt control region is thus:

- Preselection
- H_{bb} candidate fails GNN tagger the 85% working point [2]
- Top tagger fails the 80% efficiency working point
- $m_T^{\mathrm{lep}+\nu} < 50\,\mathrm{GeV}$
- $\frac{p_T \text{ of lepton}}{p_T \text{ of leading large-}R\text{ jet}} < 0.05$

The background composition in this non-prompt control region is depicted in Figure 4.4. It can be seen that the control is dominated by non-prompt contributions with about 48%. In the region, there are also small contributions from Single Top, Diboson and

---

[2] The 85% efficiency working point of the mass-dependent working point is named 1.55. For further information on this, the reader is referred to [45, 46].

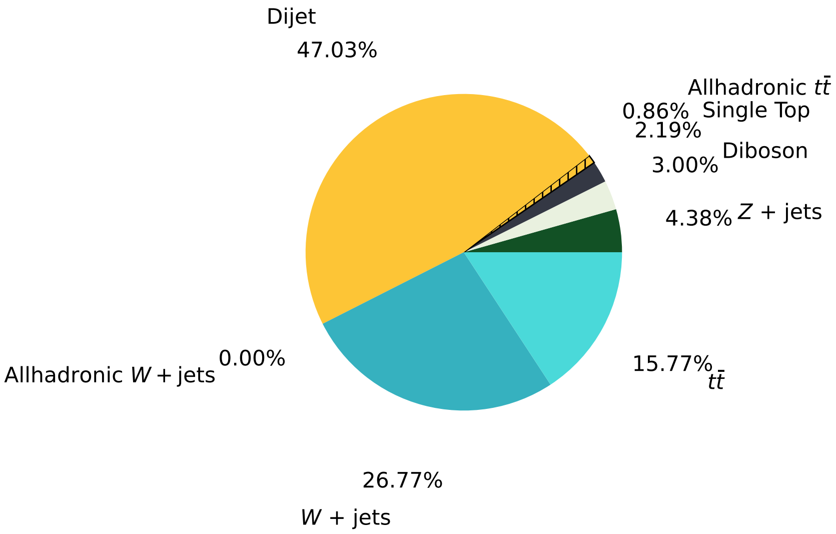


Figure 6.4.: Background composition in the control region. The non-prompt background is here broken down into dijet, all-hadronic $t\bar{t}$ and all-hadronic W+jets.

Z+jets. $t\bar{t}$ makes up about 15% of this region. The W+jets background is the hardest to distinguish from the non-prompt background and still contributes about 27% to the non-prompt control region. The yields of the respective background contribution for the partial Run 3 dataset are summarized in Table 6.2. It should be emphasized that these yields are only for 2022 data, corresponding to $26.07\,\mathrm{fb}^{-1}$, and for full Run 3 the yields will be approximately 8 times higher. Another important aspect is the signal contamination in the non-prompt control region. The expected signal events in the control region for the parial Run 3 dataset are listed in Table 6.3. Here, a cross section of 100 fb was assumed for the signal. Even in the worst case, which is about 2 signal events for $m_X = 3000\,\mathrm{GeV}$, $m_S = 1000\,\mathrm{GeV}$, signal contamination is so low that it makes up less

	event yield
$t\bar{t}$	28.81 ± 6.20
W+jets	48.92 ± 2.82
Single Top	4.01 ± 1.60
Z+jets	8.01 ± 0.40
Diboson	5.49 ± 0.45
Dijet	85.94 ± 20.27
all-hadronic $t\bar{t}$	1.57 ± 1.12
all-hadronic W+jets	0.00 ± 0.00

Table 6.2.: Background yields with statistical uncertainties in the signal region for partial Run 3 for a luminosity of $26.07\,\mathrm{fb}^{-1}$.

than 1% of the events in the control region.

mass point	event yield
$m_X = 1000\,\text{GeV}$, $m_S = 400\,\text{GeV}$	0.14 ± 0.07
$m_X = 1000\,\text{GeV}$, $m_S = 500\,\text{GeV}$	0.18 ± 0.09
$m_X = 1000\,\text{GeV}$, $m_S = 750\,\text{GeV}$	0.09 ± 0.07
$m_X = 1500\,\text{GeV}$, $m_S = 500\,\text{GeV}$	1.23 ± 0.26
$m_X = 1500\,\text{GeV}$, $m_S = 750\,\text{GeV}$	0.71 ± 0.23
$m_X = 1500\,\text{GeV}$, $m_S = 1000\,\text{GeV}$	0.43 ± 0.14
$m_X = 2000\,\text{GeV}$, $m_S = 750\,\text{GeV}$	1.33 ± 0.27
$m_X = 2000\,\text{GeV}$, $m_S = 1000\,\text{GeV}$	1.13 ± 0.26
$m_X = 2000\,\text{GeV}$, $m_S = 1500\,\text{GeV}$	0.46 ± 0.17
$m_X = 2500\,\text{GeV}$, $m_S = 1000\,\text{GeV}$	1.45 ± 0.30
$m_X = 2500\,\text{GeV}$, $m_S = 1500\,\text{GeV}$	0.80 ± 0.20
$m_X = 2500\,\text{GeV}$, $m_S = 2000\,\text{GeV}$	0.68 ± 0.22
$m_X = 3000\,\text{GeV}$, $m_S = 1000\,\text{GeV}$	1.96 ± 0.31
$m_X = 3000\,\text{GeV}$, $m_S = 1500\,\text{GeV}$	1.20 ± 0.27
$m_X = 3000\,\text{GeV}$, $m_S = 2000\,\text{GeV}$	0.86 ± 0.21
$m_X = 3000\,\text{GeV}$, $m_S = 2500\,\text{GeV}$	0.65 ± 0.20

Table 6.3.: Signal yields in the control region for all 16 available mass points. These yields were calculated for a luminosity of $26.07\,\text{fb}^{-1}$. For the signal a cross-section of 100 fb was assumed.

6.2. Modelling in the control region

After the definition of the control region, the data in that region is unblinded. Part of the unblinded distributions will be shown and discussed in the following, while the rest can be found in Appendix A.7. In the following distributions, the full Run 2 dataset will investigated additionally to the partial Run 3 dataset. The reason for that is the limited amount of data from Run 3 which leads to a high statistical uncertainty on the results. For the event yields a luminosity of $140.07\,\text{fb}^{-1}$ was used for Run 2 and for partial Run 3 a luminosity of $26.07\,\text{fb}^{-1}$. The uncertainties plotted for the data points are the statistical uncertainties based on a Poisson distribution which is given by the square-root of the data yield itself. The plot at the bottom of the figure shows the ratio of the data and the Monte Carlo simulation. Here, the error bars are also divided by the Monte Carlo simulation. In all these plots, the signal yields are scaled up by a factor of 100 for better visibility.

To assess the quality of the Monte Carlo background fit, a weighted χ^2-test is performed. The resulting χ^2 value, which can be found in the upper right corner of each plot, is defined as

$$\chi^2 = \frac{\sum_i \frac{x_i^{\text{data}} - x_i^{\text{bkg}}}{\sigma_i^{\text{data}} + \sigma_i^{\text{bkg}}}}{N_{\text{dof}}} \quad . \tag{6.2}$$

Here, x_i^{data} are the datapoints of the data and σ_i^{data} are the corresponding errors. x_i^{bkg} are the datapoints of the Monte Carlo simulation with the corresponding errors σ_i^{bkg}. N_{dof} is the number of the degrees of freedom of each plot which is given by number of bins in the histogram.

The mass and the transverse momentum p_T of the H_{bb} candidate are pictured in Figure 6.5 and Figure 6.6. The distribution of the H_{bb} mass features as expected a peak around the Higgs mass for both the data and the Monte Carlo simulation.

Analogously, the mass and transverse momentum of the W_{had} candidate are depicted in Figure 6.7 and Figure 6.8.

The distributions of the transverse momentum of the H_{bb} and W_{had} share similar shapes with a peak around 500 GeV. This peak is the result of the cut on the transverse momentum of the leading large-R jet in the preselection. Since the leading large-R jet is not always the H_{bb} candidate, values smaller than 500 GeV are possible and there is no sharp cut in the distribution. The same applies for the W_{had} candidate.

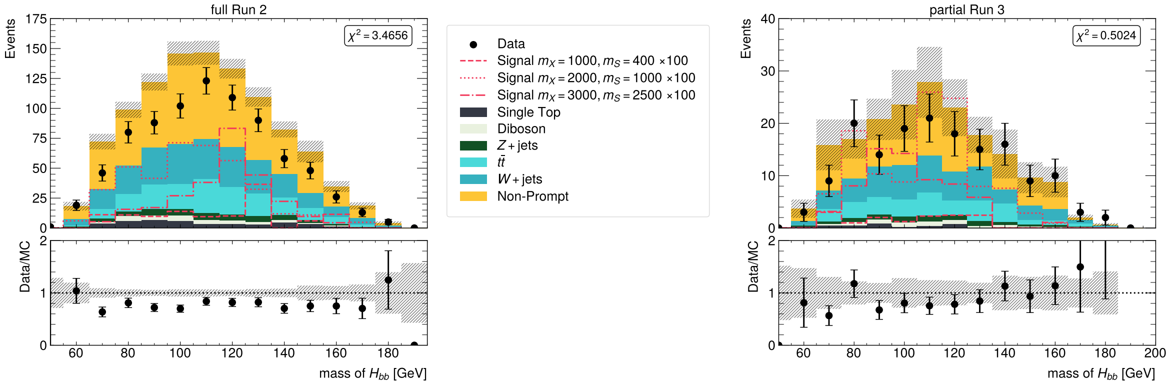


Figure 6.5.: Mass of the H_{bb} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

Another important object of the analysis is the final-state lepton. The unblinded data in comparison to the Monte Carlo prediction for the lepton tranverse momentum and pseudorapidity is shown in Figures 6.9-6.12. These plots distinguish between final states with electrons and muons as the signal lepton.

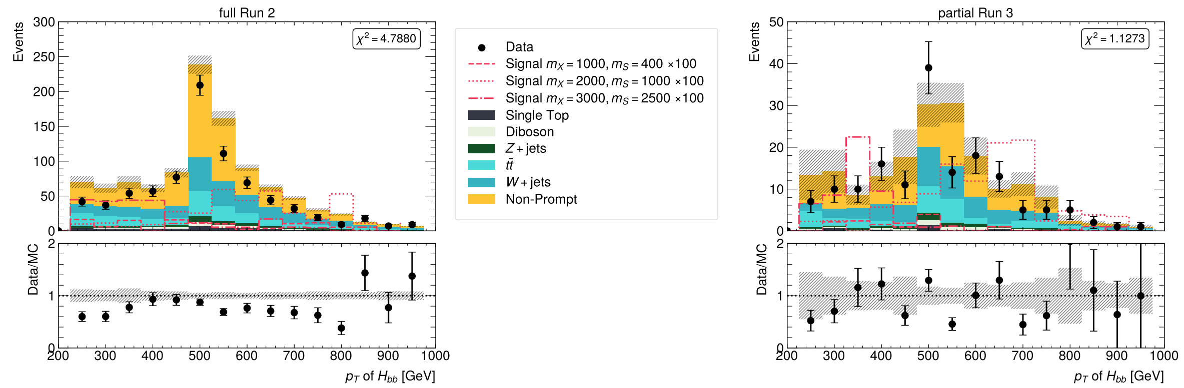


Figure 6.6.: Transverse momentum p_T of the H_{bb} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

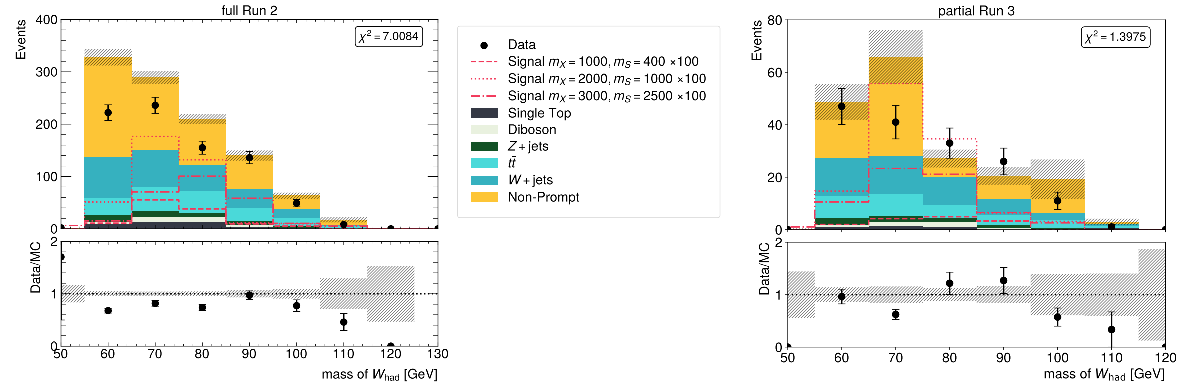


Figure 6.7.: Mass of the W_{had} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

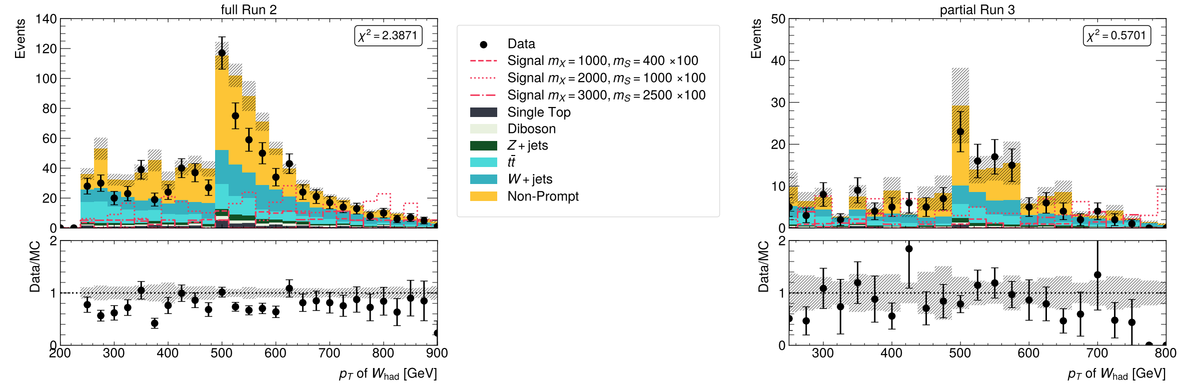


Figure 6.8.: Transverse momentum p_T of the W_{had} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

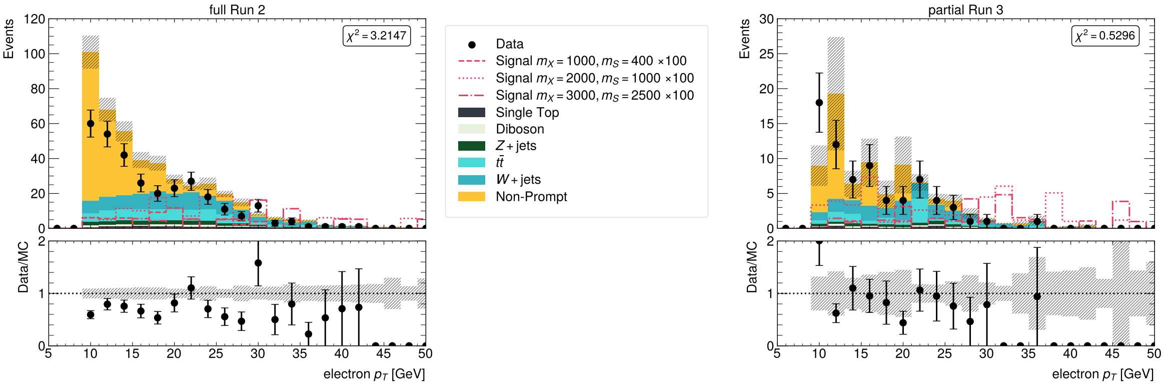


Figure 6.9.: Transverse momentum p_T of electrons on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

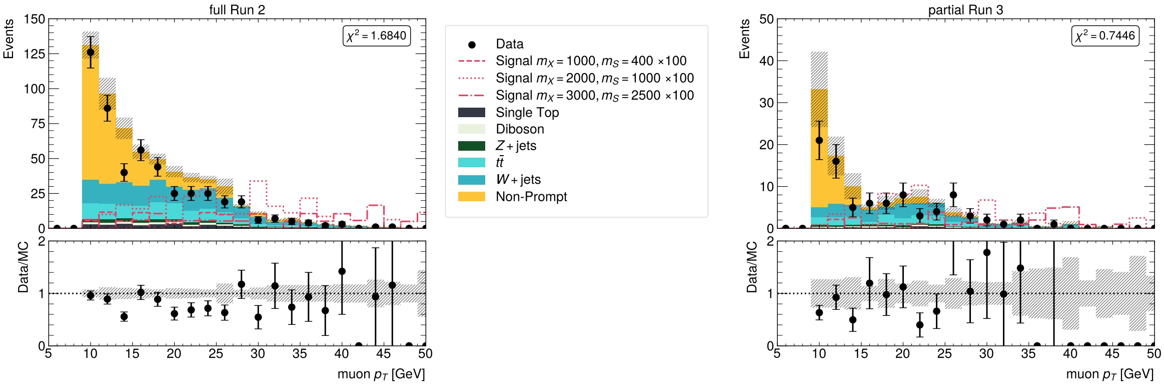


Figure 6.10.: Transverse momentum p_T of muons on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

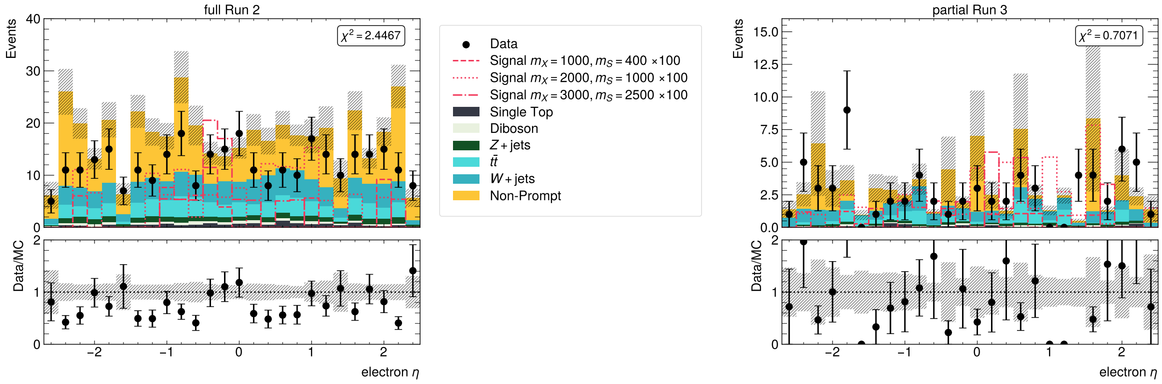


Figure 6.11.: Pseudorapidity η of electrons on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

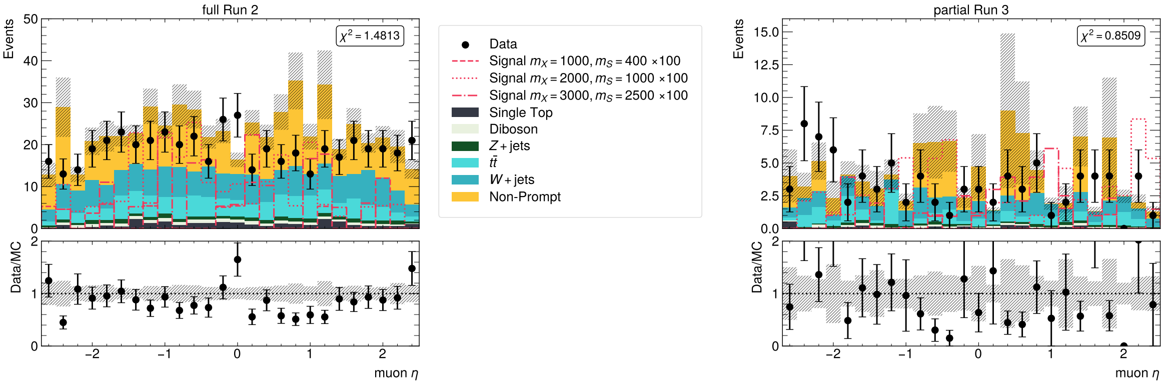


Figure 6.12.: Pseudorapidity η of muons on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

Figures 6.13, 6.14 and 6.15 depict the distributions of data and Monte Carlo for the distance between the different objects in the topology. In the distance between the W_{had} candidate and the lepton, the $\Delta R > 1$ cut from the preselection can be observed. Additionally, one can see that in the majority of the events the H_{bb} candidate and the W_{had} candidate are emitted back-to-back with a $\Delta R \approx \pi$ (see Figure 6.15). Due to this geometrical relation between the H_{bb} candidate and the W_{had} candidate, a cut on the ΔR between the H_{bb} candidate and the lepton is inferred from the preselection cut.

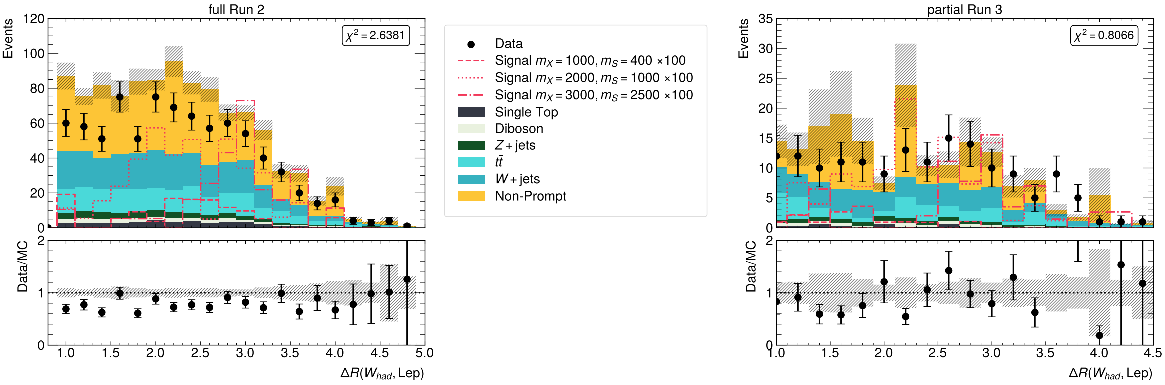


Figure 6.13.: Distance ΔR between the W_{had} candidate and the lepton on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

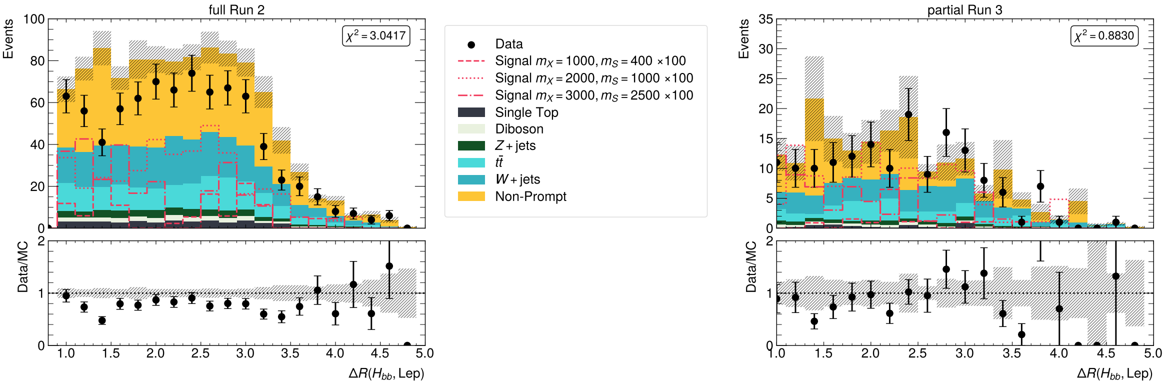


Figure 6.14.: Distance ΔR between the H_{bb} candidate and the lepton on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

The variables used to define the control region are plotted in Figure 6.16 and Figure 6.17. In these plots, the cuts from the definition of the control region are visible in these plots. Figure 6.18 pictures the number of b-tagged small-R jets. These jets are tagged as described in Section 4.3.2.

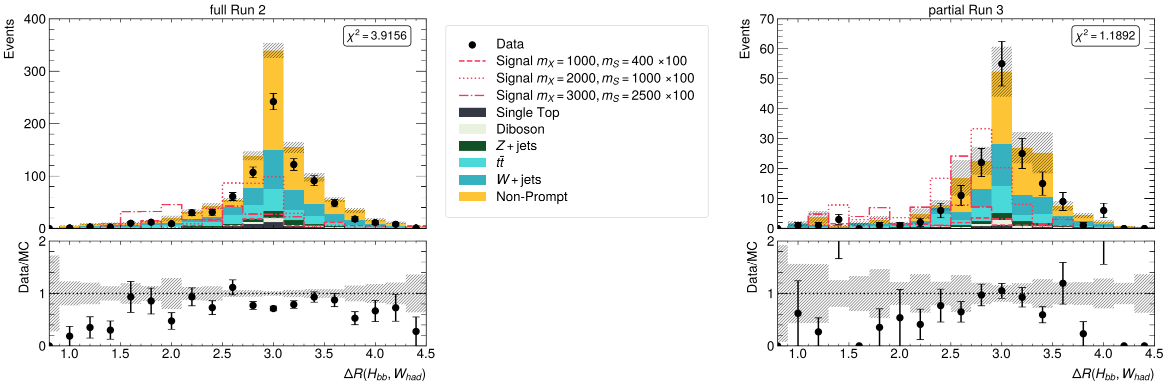


Figure 6.15.: Distance ΔR between the H_{bb} candidate and the W_{had} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

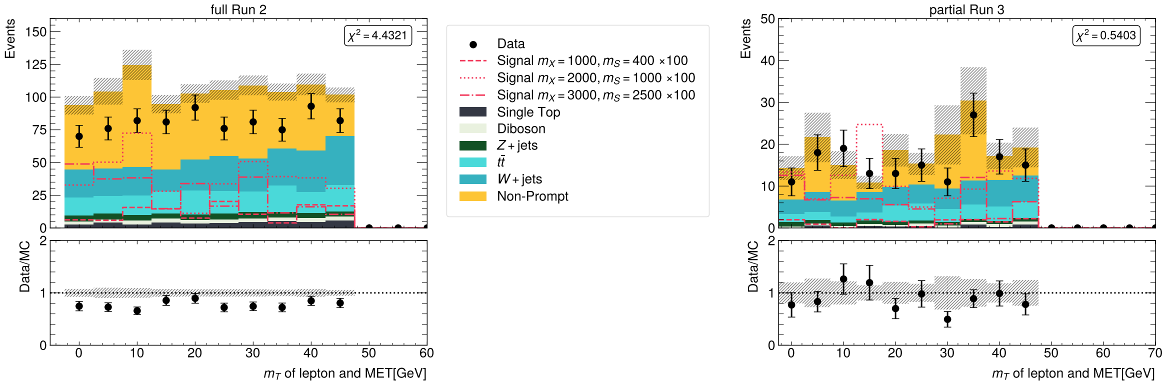


Figure 6.16.: Transverse mass of the lepton and the missing transverse energy on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

The final discriminant, which was used in Run 2 and will mostly likely also be used in Run 3, is defined as:

$$m_{\text{vis+met}} = \sqrt{(p^{H\to b\bar{b}})^2 + (p^{W_{\text{had}}})^2 + (p^{\ell})^2 + (p^{met})^2} \quad . \tag{6.3}$$

Its distribution is shown in Figure 6.19.

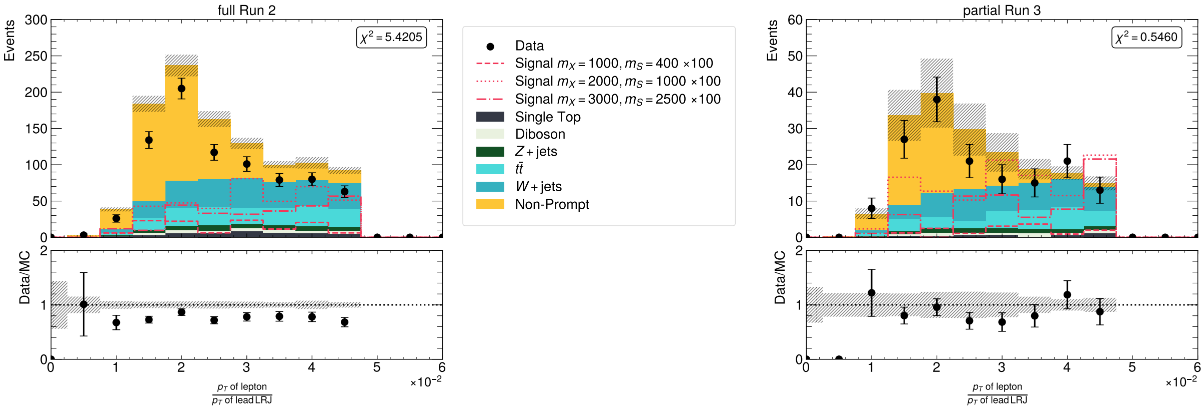


Figure 6.17.: Ratio of the tranverse momentum of the lepton and the transverse momentum of the leading large-R jet on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

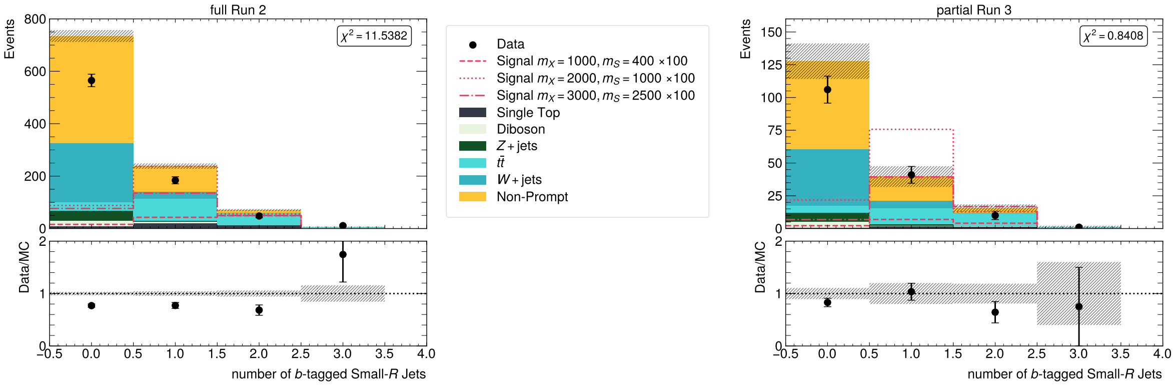


Figure 6.18.: Number of b-tagged small-R jets on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

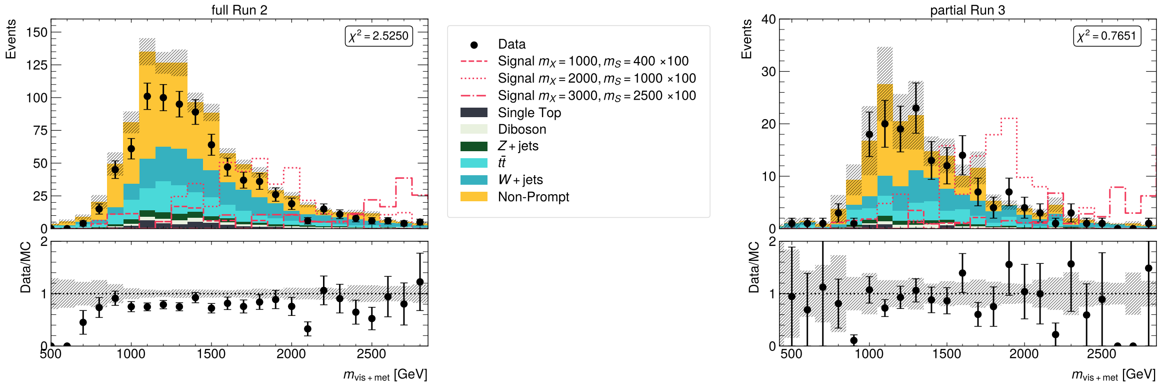


Figure 6.19.: The final discriminant $m_{\text{vis+met}}$ as defined in Equation 6.3 on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

6.3. Systematic uncertainties

To assess the quality of the Monte Carlo modelling in the control region, systematic uncertainties should also be investigated. From looking at the distribution of the final discriminant $m_{\text{vis+met}}$, one can see that the shapes of data and Monte Carlo simulation seem to be similar. However, there is a clear difference in the absolute number of events, which is especially visible in the Run 2 dataset. This trend can be observed in all of the distribution shown in the previous section.

This systematic offset can be quantified by summing up the events in the bins to the total number of events in both data and Monte Carlo simulation. The results are listed in Table 6.4.

Run period	Events in Data	Events in Monte Carlo	Difference
Run 2	800 ± 28.28	1034.26 ± 25.30	$(22.65 \pm 3.33)\%$
partial Run 3	158 ± 12.57	182.78 ± 14.89	$(13.56 \pm 9.84)\%$

Table 6.4.: Number of total events in data and Monte Carlo simulation for Run 2 and partial Run 3. The last column shows the relative difference between data and MC in percentages. For Run 2 a luminosity of $140.07\,\text{fb}^{-1}$ was used and for partial Run 3 a luminosity of $26.07\,\text{fb}^{-1}$.

Another approach to understand the uncertainties of the modelling was taken by calculating the closure uncertainty. The closure uncertainty is defined as

$$\sigma_{\text{closure}} = \frac{x_{\text{data}} - x_{\text{MC}}}{x_{\text{MC}}} \tag{6.4}$$

The closure uncertainty binned in $m_{\text{vis+met}}$ is depicted in Figure 6.20. To account for the offset in absolute values, also a corrected closure uncertainty was computed. For that, the data and Monte Carlo simulation were first normalized to 1 and then Equation 6.4 was applied. The resulting plots can be found in Figure 6.21. The closure uncertainty distributions clearly show that a correction of the observed offset effectively reduces the uncertainties.

Overall, it is evident that the partial Run 3 dataset is significantly impacted by the low number of events leading to large statistical fluctuations. Similarly, statistical uncertainties increase for higher values of $m_{\text{vis+met}}$, where statistics also decrease. Nevertheless, with an appropriate normalization correction, a reasonable systematic uncertainty can be obtained for the shape of the final discriminant.

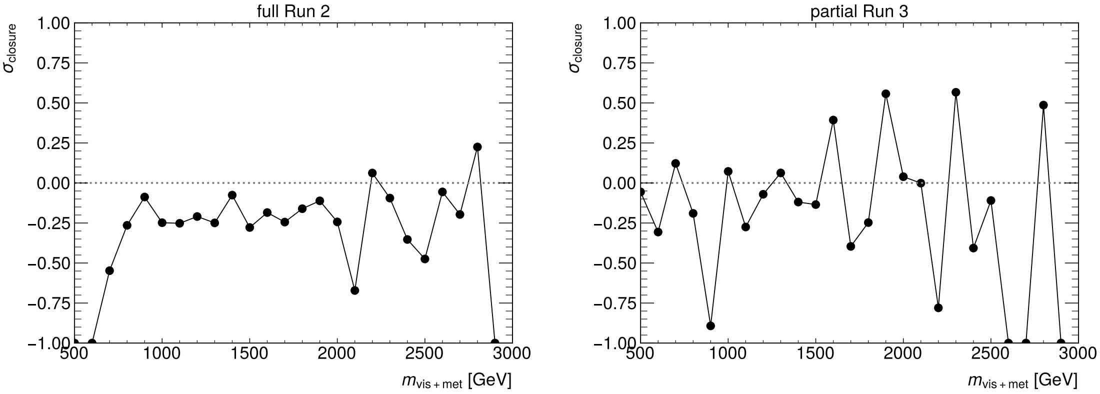


Figure 6.20.: Closure uncertainty σ_{closure} for the full Run 2 dataset on the left and partial Run 3 on the right. The definition of the closure uncertainty can be found in Equation 6.4.

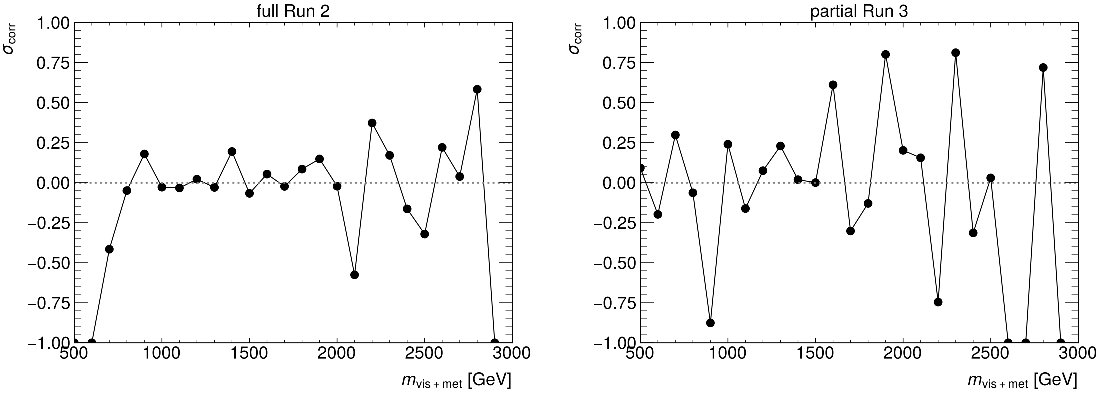


Figure 6.21.: Corrected closure uncertainty σ_{corr} for the full Run 2 dataset on the left and partial Run 3 on the right. To calculate the corrected closure uncertainty the data and Monte Carlo events are first normalized to unity and then Equation 6.4 is applied.

7. Conclusion and Outlook

The $X \to SH \to b\bar{b}WW$ decay channel offers a unique window to search for particles beyond the Standard Model (BSM). The search for BSM particles is highly significant, as it provides an oppurtunity for uncovering the existence of new, heavy bosons predicted by several extensions of the Standard Model, such as two-Higgs-doublet models or the two-real-singlet model. These new bosons, if discovered, would radically alter our understanding of fundamental interactions and open the door to a more comprehensive framework that could explain unresolved phenomena like dark matter, neutrino masses and many more.

The results of this thesis contribute to the object definitions and region definitions for the analysis of the split-boosted topology of the $X \to SH \to b\bar{b}WW$ decay channel in the 1-lepton final-state. These components are essential building blocks for the construction of a robust and precise analysis for Run 3 at ATLAS.
The following sections summarize and discuss the findings of this thesis and give an outlook into further steps that need to be taken by the analysis group.

7.1. Lepton Study

The lepton study performed in Section 5 establishes particle identification and isolation working points that generally feature high signal efficiencies for muons and electrons (see Figure 5.4). When comparing the signal efficiencies to the background rejection for the dijet background in Figure 5.9, one should look for the working point with the highest background rejection while keeping a reasonable signal efficiency. The resulting choice of working points for signal leptons is also listed in Table 7.1.
For electrons, the `Tight_VarRad` isolation was chosen, even though the `HighPtCaloOnly` working point has the highest background rejection. This choice was made due to the fact that it was at first unclear whether the latter working point should be used in the split-boosted topology. In boosted scenarios, the `HighPtCaloOnly` working point tends to not perform well for leptons with high momenta. Upon further investigation, this

behavior was found in the split-boosted although not as pronounced as in the boosted topology. It was nevertheless decided to go with the safer choice of the `Tight_VarRad` working point. Another reason for this decision was the higher signal efficiency and the overall higher statistics that are kept with `Tight_VarRad`. The choice for muons is rather uncontroversial, where the working point `PflowTight_VarRad` was chosen.

signal electrons	signal electrons
Tight ID	Medium ID
`PflowTight_VarRad`	`Tight_VarRad`

Table 7.1.: Definition of signal leptons based on ID and isolation.

To check the impact the overlap removal has on the choice of working points, the same study was performed with an overlap removal (as described in Section 4.3.1) applied leading to the plots in Figure 5.10. The changes in signal efficiency and background rejection due to the overlap removal are only minor, such that no adjustment of the choice of working points from above was needed.

Furthermore, a ΔR cut was introduced to the lepton study to adapt the study to the split-boosted topology (see Section 5.6). The results pictured in Figure 5.11 show in general, that this cut reduces background rejection and slightly increases signal efficiency. But ultimately, the performance of the working points relative to each other is not impacted, which can be seen in Figure 5.11. The previously chosen working points (see Table 7.1) are the best performing working points even when the cut is applied. Thus no alteration of the choice of working points is needed.

At the time this study was conducted, there were only Loose and Medium ID available for muons while the Tight and HighPt working points were not available for the study, since they could not be used in the analysis framework used for this thesis. So the results of this study should be further verified at a later point when the working points are available to use in the framework. However, the experience from Run 2 indicates that it is likely that the analysis can use the final results of this study.

7.2. Background Modelling in the Non-Prompt Control Region

In Chapter 6, a control region dedicated to the understanding of the non-prompt background was defined. The cuts used for the definition are :

- Preselection
- H_{bb} candidate fails GNN tagger at 85% working point [1]
- Top tagger fails at 80% efficiency working point
- $m_T^{\text{lep}+\nu} < 50\,\text{GeV}$
- $\frac{p_T \text{ of lepton}}{p_T \text{ of leading large-}R\text{ jet}} < 0.05$

This results in a non-prompt control region with a purity of about 47% (see Figure 6.4). Someone might argue that a higher purity in the control region would be beneficial to the analysis. A higher purity can certainly be achieved with additional cuts, but this comes with a drastic reduction in the number of events. Because the split-boosted 1-lepton analysis in general already has to deal with few events, it is advisable to not apply further cuts and thus keep more events. Moreover, a purity of 47% should be enough to check the modelling in the control region.

To compare the unblinded data points to the Monte Carlo background modelling in the control region, a set of kinematic variables was investigated. In general, it can be observed that the plots for the partial Run 3 dataset are largely dominated by statistical uncertainties leading to prominent fluctuations in the bins. These statistical uncertainties are reduced for the plots depicting Run 2 distributions and will also be reduced for Run 3 in the future, when more data is available.
A weighted χ^2-test is performed to assess the goodness of the fit. The χ^2 values are significantly larger for Run 2, due to the reduced statistical uncertainty, but overall lie within a reasonable range.
Another overall trend that is evident, is the good agreement in shape of the distributions. Moreover, there is a significant offset visible between the absolute values of data and the Monte Carlo simulation. The offset can be quantified by simply comparing the numbers of total events in data and in Monte Carlo. The differences between Monte Carlo and

---

[1] The 85% efficiency working point of the mass-dependent working point is named 1.55. For further information on this, the reader is referred to [45, 46].

data event yields in the control region is (22.65 ± 3.33)% for the Run 2 dataset and (13.56 ± 9.84)% for the partial Run 3 dataset. The differences of the Run 2 dataset and the Run 3 dataset agree within their errors and are therefore consistent, which shows that there is probably a normalization correction missing and not other additional effects at play. The large error on the difference for the partial Run 3 dataset stems from the prominent statistical fluctuations in this dataset.
At any rate, this offset can be corrected with normalization corrections in a fit. This will also reduce the values of the χ^2-test. For that, one first needs to find out which backgrounds need this normalization correction by looking into the other control regions. This task lies beyond the scope of this thesis but needs to be adressed in the future by the analysis group.
Besides the statistical uncertainty, systematic uncertainties also need to be considered. This was done in Section 6.3 by calculating the closure uncertainty in the bins of the final discriminant $m_{\text{vis+met}}$. Furthermore, a corrected closure uncertainty was calculated to account for the offset, which originates from a missing normalization correction, discussed above. For the partial Run 3 dataset, the results are significantly impacted by the low statistics resulting in large uncertainties. The same behavior can be seen in Run 2 for the higher values of $m_{\text{vis+met}}$, where statistics tend to get lower as well. Apart from that, we can see that with a normalization correction a reasonable systematic uncertainty can be derived for the final discriminant.

With these results, it can be concluded that the non-prompt background is modeled sufficiently well by Monte Carlo simulation and no data-driven approach to the background estimation is necessary. The contribution of the non-prompt background to the signal region rather small at 7%, such that a data-driven estimate would use time and person power on detail work that can be more beneficial in other places of the analysis.

Overall, it needs to be mentioned, that the ATLAS recommendations for the dijet reweighting have been adjusted. It should therefore be checked how much this impacts the results of this thesis and whether alterations for the split-boosted 1-lepton analysis are needed in the future.

With the ongoing data-taking at the ATLAS detector, statistical uncertainties will decrease and more data will be available to detect new signatures in the future. Such a discovery would be a breakthrough not only for Higgs physics but also physics beyond the standard model and help physicists to further understand the universe.

Bibliography

[1] M. Thomson, *Modern particle physics*, Cambridge University Press, New York (2013)

[2] 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), `1207.7214`

[3] 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), `1207.7235`

[4] A. Einstein, *The foundation of the general theory of relativity.*, Annalen Phys. **49(7)**, 769 (1916)

[5] Super-Kamiokande Collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81**, 1562 (1998), URL `https://link.aps.org/doi/10.1103/PhysRevLett.81.1562`

[6] G. Bertone, D. Hooper, J. Silk, *Particle dark matter: evidence, candidates and constraints*, Physics Reports **405(5)**, 279 (2005), URL `https://www.sciencedirect.com/science/article/pii/S0370157304003515`

[7] M. Muhlleitner, M. O. P. Sampaio, R. Santos, J. Wittbrodt, *The N2HDM under Theoretical and Experimental Scrutiny*, JHEP **03**, 094 (2017), `1612.01309`

[8] T. Robens, T. Stefaniak, J. Wittbrodt, *Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios*, Eur. Phys. J. C **80** (2020), URL `https://link.springer.com/article/10.1140/epjc/s10052-020-7655-x`

[9] R. L. Workman, et al. (Particle Data Group), *Review of Particle Physics*, PTEP **2022**, 083C01 (2022)

[10] R. P. Feynman, *The Theory of Positrons*, Phys. Rev. **76**, 749 (1949), URL `https://link.aps.org/doi/10.1103/PhysRev.76.749`

[11] R. P. Feynman, *Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*, Phys. Rev. **80**, 440 (1950), URL `https://link.aps.org/doi/10.1103/PhysRev.80.440`

[12] M. Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry* (1961)

[13] R. P. Feynman, M. Gell-Mann, *Theory of Fermi interaction*, Phys. Rev. **109**, 193 (1958)

[14] E. C. G. Sudarshan, R. e. Marshak, *Chirality invariance and the universal Fermi interaction*, Phys. Rev. **109**, 1860 (1958)

[15] M. Kobayashi, T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49**, 652 (1973)

[16] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22**, 579 (1961)

[17] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519**, 367 (1968)

[18] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19**, 1264 (1967)

[19] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12**, 132 (1964)

[20] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, 321 (1964)

[21] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 508 (1964)

[22] G. S. Guralnik, C. R. Hagen, T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**, 585 (1964)

[23] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145**, 1156 (1966)

[24] T. W. B. Kibble, *Symmetry breaking in nonAbelian gauge theories*, Phys. Rev. **155**, 1554 (1967)

[25] M. Grazzini, G. Heinrich, S. Jones, S. Kallweit, M. Kerner, J. M. Lindert, J. Mazzitelli, *Higgs boson pair production at NNLO with top quark mass effects*, JHEP **05**, 059 (2018), `1803.02463`

[26] D. de Florian, et al. (LHC Higgs Cross Section Working Group), *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector* **2/2017** (2016), `1610.07922`

[27] A. D. Sakharov, *Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34(5)**, 392 (1991), URL `https://dx.doi.org/10.1070/PU1991v034n05ABEH002497`

[28] T. D. Lee, *A Theory of Spontaneous T Violation*, Phys. Rev. D **8**, 1226 (1973), URL `https://link.aps.org/doi/10.1103/PhysRevD.8.1226`

[29] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3(08)**, S08003 (2008), URL `https://doi.org/10.1088/1748-0221/3/08/s08003`

[30] E. Mobs, *The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019* (2019), general Photo, URL `https://cds.cern.ch/record/2684277`

[31] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation **3(08)**, S08002 (2008), URL `https://dx.doi.org/10.1088/1748-0221/3/08/S08002`

[32] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3(08)**, S08003 (2008), URL `https://dx.doi.org/10.1088/1748-0221/3/08/S08003`

[33] CMS Collaboration, *The CMS experiment at the CERN LHC*, Journal of Instrumentation **3(08)**, S08004 (2008), URL `https://dx.doi.org/10.1088/1748-0221/3/08/S08004`

[34] LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation **3(08)**, S08005 (2008), URL `https://dx.doi.org/10.1088/1748-0221/3/08/S08005`

[35] R. M. Bianchi, ATLAS Collaboration, *ATLAS experiment schematic or layout illustration* (2022), general Photo, URL `https://cds.cern.ch/record/2837191`

[36] K. Abeling, *Search for resonant Higgs boson pair production in the $bbWW^*$ decay channel in the boosted 1-lepton final state using the full Run 2 ATLAS dataset*, Ph.D. thesis, Göttingen University, II.Physik-UniGö-Diss-2022/01 (2022)

[37] M. Cacciari, G. P. Salam, G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04**, 063 (2008), `0802.1189`

[38] ATLAS Collaboration, *Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions*, Eur. Phys. J. C **81(4)**, 334 (2021), `2009.04986`

[39] A. J. Larkoski, S. Marzani, G. Soyez, J. Thaler, *Soft Drop*, JHEP **05**, 146 (2014), `1402.2657`

[40] *OverlapAnalysisSequence* (2024), accessed: October 8, 2024, URL `https://gitlab.cern.ch/atlas/athena/-/blob/main/PhysicsAnalysis/Algorithms/AsgAnalysisAlgorithms/python/OverlapAnalysisSequence.py`

[41] *AssociationUtils* (2024), accessed: September 23, 2024, URL `https://gitlab.cern.ch/atlas/athena/-/blob/24.0/PhysicsAnalysis/AnalysisCommon/AssociationUtils/`

[42] *DL1d Tagger* (2024), accessed: September 23, 2024, URL `https://ftag.docs.cern.ch/algorithms/taggers/dl1/#dl1`

[43] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79(11)**, 970 (2019), `1907.05120`

[44] *Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS*, Technical report, CERN, Geneva (2023), all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2023-021, URL `https://cds.cern.ch/record/2866601`

[45] D. Battulga, *Flat Mass Discriminant of GN2Xv01* (2024), accessed: October 10, 2024, URL `https://gitlab.cern.ch/dabattul/ntup-histo-analysis-sh4b/-/blob/82b6791941ab5918b0797f4476342f8cbdf3edb2/GN2Xv01_FlatMassEff/FlatMassDiscriminant_GN2Xv01_25042024.csv`

[46] D. Battulga, *Gn2Xv01 boosted bb-jet tagger: Flat Mass mis-tag rate WPs* (2024), accessed: October 10, 2024, URL `https://indico.cern.ch/event/1408775/contributions/5920971/attachments/2846772/4977631/R24_GN2Xv01_tagger_FlatMass_qcd_April292024-1.pdf`

[47] *Performance of W/Z taggers using UFO jets in ATLAS* (2021), URL `https://cds.cern.ch/record/2777009`

[48] I. Moult, L. Necib, J. Thaler, *New Angles on Energy Correlation Functions*, JHEP **12**, 153 (2016), `1609.07483`

[49] ATLAS Collaboration, *Identification of hadronically-decaying top quarks using UFO jets with ATLAS in Run 2*, ATL-PHYS-PUB-2021-028 (2021), URL `https://cds.cern.ch/record/2776782`

[50] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **81(7)**, 578 (2021), `2012.00578`

[51] *ATLAS Isolation and Fake Forum* (2024), accessed: May 2, 2024, URL `https://atlas-mcp.docs.cern.ch/guidelines/muonselectiontool/index.html`

[52] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, JINST **14(12)**, P12006 (2019), `1908.00005`

[53] J. F. Beirer, *Novel Approaches to the Fast Simulation of the ATLAS Calorimeter and Performance Studies of Track-Assisted Reclustered Jets for Searches for Resonant* $X \rightarrow SH \rightarrow b\bar{b}WW^*$ *Production with the ATLAS Detector*, Ph.D. thesis, Göttingen University, II. Physik-UniGö-Diss-2023/01, CERN-THESIS-2023-096 (2023)

[54] D. Kar, *Experimental Particle Physics*, IOP Expanding Physics, Brisol, UK (2019)

[55] *ATLAS Isolation and Fake Forum* (2024), accessed: April 3, 2024, URL `https://atlas-iff.docs.cern.ch/isolationvars/index.html`

[56] *ATLAS Isolation and Fake Forum* (2024), accessed: April 6, 2024, URL `https://atlas-iff.docs.cern.ch/rel22recommendedisowps/index.html`

A. Appendix

A.1. Isolation Working Points

Electrons:

HighPtCaloOnly: topoetcone20 $<$ max(0.015 $\cdot$ p_T, 3.5 GeV),

Loose_VarRad: ptvarcone30_Nonprompt_All_MaxWeightTTVALooseCone_pt1000/pT $<$ 0.15 & topoetcone20/pT $<$ 0.2,

Tight_VarRad: ptvarcone30_Nonprompt_All_MaxWeightTTVALooseCone_pt1000/pT $<$ 0.06 & topoetcone20/pT $<$ 0.06,

TightTrackOnly_VarRad: ptvarcone30_Nonprompt_All_MaxWeightTTVALooseCone_pt1000/pT $<$ 0.06,

TightTrackOnly_FixedRad: ptvarcone30_Nonprompt_All_MaxWeightTTVALooseCone_pt1000/pT $<$ 0.06 (p_T $<$ 50 GeV), ptcone20_Nonprompt_All_MaxWeightTTVALooseCone_pt1000/pT $<$ 0.06 (p_T $>$ 50 GeV),

Muons:

Loose_VarRad: ptvarcone30_Nonprompt_All_MaxWeightTTVA_pt1000/pT $<$ 0.15 & topoetcone20/pT $<$ 0.3,

Tight_VarRad: ptvarcone30_Nonprompt_All_MaxWeightTTVA_pt1000/pT $<$ 0.04 & topoetcone20/pT $<$ 0.15,

PflowLoose_VarRad: (ptvarcone30_Nonprompt_All_MaxWeightTTVA_pt500 + 0.4 neflowisol20)/pT $<$ 0.16,

PflowTight_VarRad: (ptvarcone30_Nonprompt_All_MaxWeightTTVA_pt500 + 0.4 neflowisol20)/pT $<$ 0.045

A.2. Signal Samples

mc23_13p6TeV.800760.Py8EG_A14NNPDF23LO_XHS_X1000_S400_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.801856.Py8EG_A14NNPDF23LO_XHS_X1000_S500_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800762.Py8EG_A14NNPDF23LO_XHS_X1000_S750_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.801992.Py8EG_A14NNPDF23LO_XHS_X1500_S500_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800767.Py8EG_A14NNPDF23LO_XHS_X1500_S750_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800768.Py8EG_A14NNPDF23LO_XHS_X1500_S1000_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800773.Py8EG_A14NNPDF23LO_XHS_X2000_S750_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800774.Py8EG_A14NNPDF23LO_XHS_X2000_S1000_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800775.Py8EG_A14NNPDF23LO_XHS_X2000_S1500_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800781.Py8EG_A14NNPDF23LO_XHS_X2500_S1000_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855 mc23_13p6TeV.800782.Py8EG_A14NNPDF23LO_XHS_X2500_S1500_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800783.Py8EG_A14NNPDF23LO_XHS_X2500_S2000_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800789.Py8EG_A14NNPDF23LO_XHS_X3000_S1000_bbWWv_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800790.Py8EG_A14NNPDF23LO_XHS_X3000_S1500_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800791.Py8EG_A14NNPDF23LO_XHS_X3000_S2000_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

mc23_13p6TeV.800792.Py8EG_A14NNPDF23LO_XHS_X3000_S2500_bbWW_1lep.deriv.DAOD_PHYS.e8544_s4162_r14622_p5855

A.3. List of Background Samples

Dijet:

mc23_13p6TeV.801165.Py8EG_A14NNPDF23LO_jj_JZ0.deriv.DAOD_PHYS.e8514_s4162_r14622_p5980

mc23_13p6TeV.801166.Py8EG_A14NNPDF23LO_jj_JZ1.deriv.DAOD_PHYS.e8514_s4162_r14622_p5980

mc23_13p6TeV.801167.Py8EG_A14NNPDF23LO_jj_JZ2.deriv.DAOD_PHYS.e8514_s4162_r14622_p5980

mc23_13p6TeV.801168.Py8EG_A14NNPDF23LO_jj_JZ3.deriv.DAOD_PHYS.e8514_s4162_r14622_p5980

mc23_13p6TeV.801169.Py8EG_A14NNPDF23LO_jj_JZ4.deriv.DAOD_PHYS.e8514_s4162_r14622_p5980

mc23_13p6TeV.801170.Py8EG_A14NNPDF23LO_jj_JZ5.deriv.DAOD_PHYS.e8514_s4162_r14622_p5980

mc23_13p6TeV.801171.Py8EG_A14NNPDF23LO_jj_JZ6.deriv.DAOD_PHYS.e8514_s4162_r14622_p5980

mc23_13p6TeV.801172.Py8EG_A14NNPDF23LO_jj_JZ7.deriv.DAOD_PHYS.e8514_s4162_r14622_p5980

mc23_13p6TeV.801173.Py8EG_A14NNPDF23LO_jj _JZ8.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266

mc23_13p6TeV.801174.Py8EG_A14NNPDF23LO_jj_JZ9incl.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266

$t\bar{t}$:

mc23_13p6TeV.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.deriv.DAOD_PHYS.e8514_s4162_r15540_p6266

all-hadronic $t\bar{t}$:

mc23_13p6TeV.601237.PhPy8EG_A14_ttbar_hdamp258p75_allhad.deriv.DAOD_PHYS.e8514_s4162_r15540_p6266

W+jets:

mc23_13p6TeV.700777.Sh_2214_Wenu_maxHTpTV2_BFilter.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.700778.Sh_2214_Wenu_maxHTpTV2_CFilterBVeto.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.700779.Sh_2214_Wenu_maxHTpTV2_CVetoBVeto.deriv.DAOD_PHYS.e8514_s4162_r15540_p6266

mc23_13p6TeV.700780.Sh_2214_Wmunu_maxHTpTV2_BFilter.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.700781.Sh_2214_Wmunu_maxHTpTV2_CFilterBVeto.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.700782.Sh_2214_Wmunu_maxHTpTV2_CVetoBVeto.deriv.DAOD_PHYS.e8514_s4162_r15540_p6266 mc23_13p6TeV.700783.Sh_2214_Wtaunu_maxHTpTV2_BFilter.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.700784.Sh_2214_Wtaunu_maxHTpTV2_CFilterBVeto.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.700785.Sh_2214_Wtaunu_maxHTpTV2_CVetoBVeto.deriv.DAOD_PHYS.e8514_s4162_r15540_p6266

all hadronic W+jets:

mc23_13p6TeV.700843.Sh_2214_Wqq_ptW_200_ECMS.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

Single Top:

mc23_13p6TeV.601348.PhPy8EG_tb_lep_antitop.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.601349.PhPy8EG_tb_lep_top.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.601350.PhPy8EG_tqb_lep_antitop.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.601351.PhPy8EG_tqb_lep_top.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.601352.PhPy8EG_tW_dyn_DR_incl_antitop.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

mc23_13p6TeV.601355.PhPy8EG_tW_dyn_DR_incl_top.deriv.DAOD_PHYS.e8514_s4162_r14622_p5855

Diboson:

mc23_13p6TeV.701085.Sh_2214_ZqqZll.deriv.DAOD_PHYS.e8543_s4162_r14622_p5855

mc23_13p6TeV.701090.Sh_2214_ZbbZll.deriv.DAOD_PHYS.e8543_s4162_r14622_p6266

mc23_13p6TeV.701095.Sh_2214_ZqqZvv.deriv.DAOD_PHYS.e8543_s4162_r14622_p6266

mc23_13p6TeV.701100.Sh_2214_ZbbZvv.deriv.DAOD_PHYS.e8543_s4162_r14622_p6266

mc23_13p6TeV.701105.Sh_2214_WqqZll.deriv.DAOD_PHYS.e8543_s4162_r14622_p6266

mc23_13p6TeV.701110.Sh_2214_WqqZvv.deriv.DAOD_PHYS.e8543_s4162_r14622_p6266

mc23_13p6TeV.701115.Sh_2214_WlvZqq.deriv.DAOD_PHYS.e8543_s4162_r14622_p6266

mc23_13p6TeV.701120.Sh_2214_WlvZbb.deriv.DAOD_PHYS.e8543_s4162_r14622_p6266

mc23_13p6TeV.701125.Sh_2214_WlvWqq.deriv.DAOD_PHYS.e8543_s4162_r14622_p6266

Z+jets:

mc23_13p6TeV.700786.Sh_2214_Zee_maxHTpTV2_BFilter.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266

mc23_13p6TeV.700787.Sh_2214_Zee_maxHTpTV2_CFilterBVeto.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266

mc23_13p6TeV.700788.Sh_2214_Zee_maxHTpTV2_CVetoBVeto.deriv.DAOD_PHYS.e8514_s4162_r15540_p6266

mc23_13p6TeV.700789.Sh_2214_Zmumu_maxHTpTV2_BFilter.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266

mc23_13p6TeV.700790.Sh_2214_Zmumu_maxHTpTV2_CFilterBVeto.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266

mc23_13p6TeV.700791.Sh_2214_Zmumu_maxHTpTV2_CVetoBVeto.deriv.DAOD_PHYS.e8514_s4162_r15540_p6266

mc23_13p6TeV.700792.Sh_2214_Ztautau_maxHTpTV2_BFilter.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266 mc23_13p6TeV.700793.Sh_2214_Ztautau_maxHTpTV2_CFilterBVeto.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266

mc23_13p6TeV.700794.Sh_2214_Ztautau_maxHTpTV2_CVetoBVeto.deriv.DAOD_PHYS.e8514_s4162_r15540_p6266

mc23_13p6TeV.700795.Sh_2214_Znunu_pTV2_BFilter.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266

mc23_13p6TeV.700796.Sh_2214_Znunu_pTV2_CFilterBVeto.deriv.DAOD_PHYS.e8514_s4162_r14622_p6266

mc23_13p6TeV.700797.Sh_2214_Znunu_pTV2_CVetoBVeto.deriv.DAOD_PHYS.e8514_s4162_r15540_p6266

A.4. Signal Plots

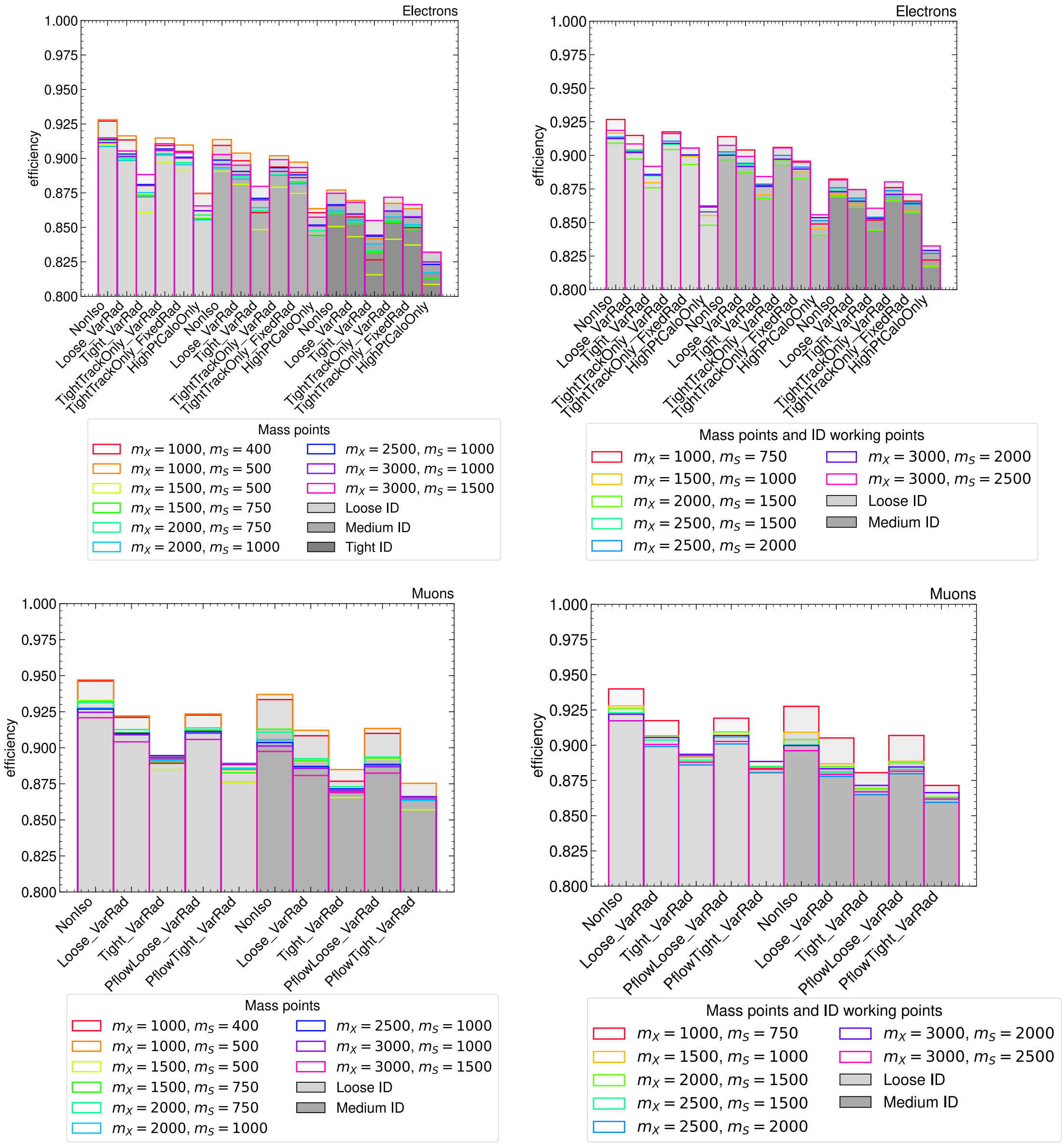


Figure A.1.: Signal efficiencies for prompt electrons (upper row) and prompt muons (lower row). The left side shows mass points with a low ratio between m_S and m_X and the right shows mass points with a high ratio. The bars show the different combinations of ID and isolation. The shades of grey represent the ID working points.

A.5. Background Plots

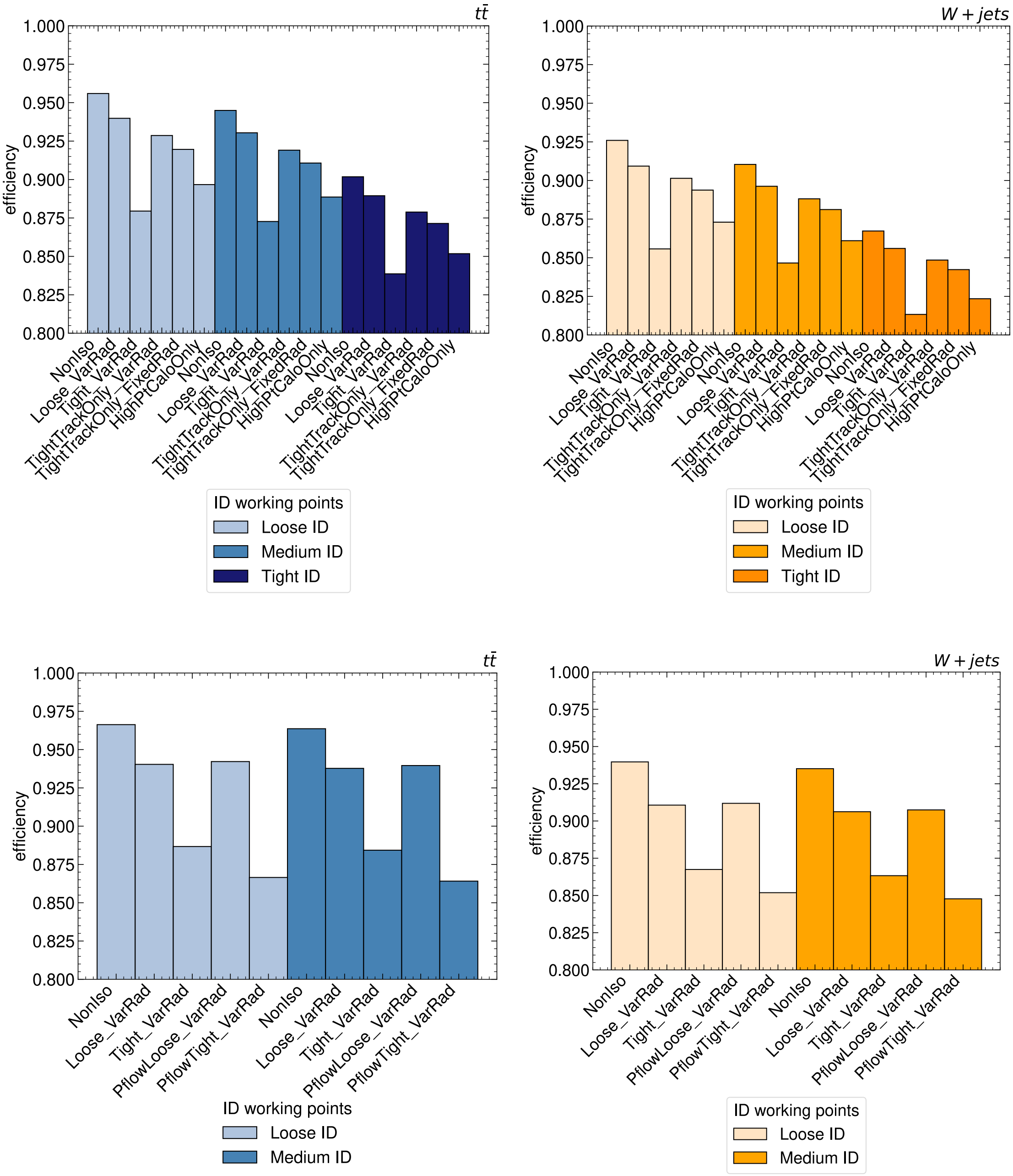


Figure A.2.: Background efficiencies for prompt electrons in the upper row and prompt muons in the lower row. The bars show the different combinations of ID and isolation. The shading of the bars indicates the applied ID working points. The left column shows the $t\bar{t}$ background, while the right column uses the W+jets background.

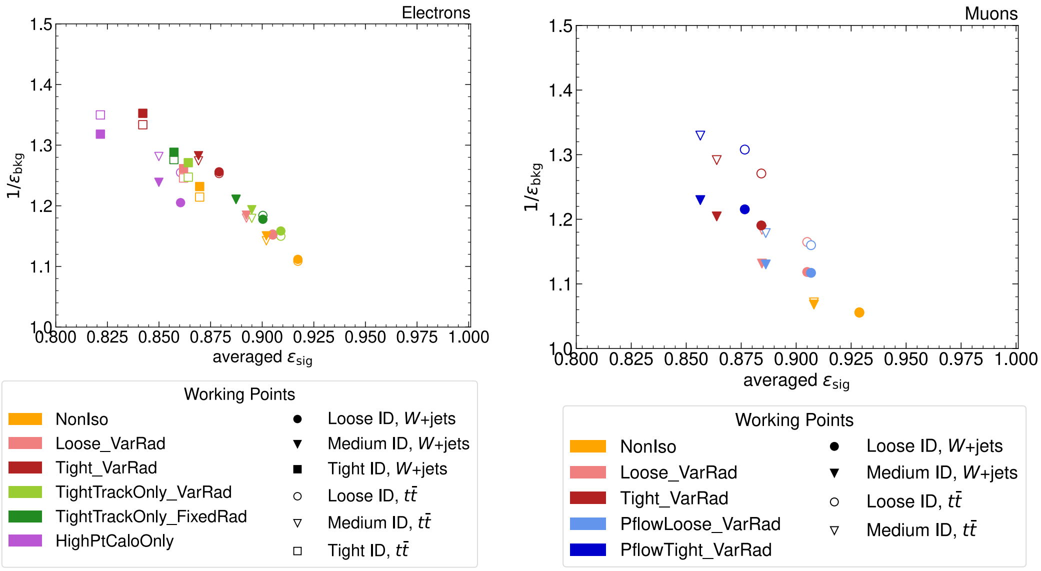


Figure A.3.: Background rejection over the signal efficiency for the W+jets and the $t\bar{t}$ background for electrons on the left and muons muons on the right.

A.6. Signal yields

Mass Point [GeV]	event yield	Mass Point [GeV]	event yield
$m_X = 1000$, $m_S = 400$	88.49	$m_X = 2000$, $m_S = 1500$	183.31
$m_X = 1000$, $m_S = 500$	81.25	$m_X = 2500$, $m_S = 1000$	308.31
$m_X = 1000$, $m_S = 750$	12.96	$m_X = 2500$, $m_S = 1500$	293.96
$m_X = 1500$, $m_S = 500$	189.40	$m_X = 2500$, $m_S = 2000$	188.11
$m_X = 1500$, $m_S = 750$	206.46	$m_X = 3000$, $m_S = 1000$	325.04
$m_X = 1500$, $m_S = 1000$	166.76	$m_X = 3000$, $m_S = 1500$	324.00
$m_X = 2000$, $m_S = 750$	261.22	$m_X = 3000$, $m_S = 2000$	297.37
$m_X = 2000$, $m_S = 1000$	274.25	$m_X = 3000$, $m_S = 2500$	187.39

Table A.1.: Signal yields after preselection for partial Run 3 corresponding to a luminosity of 26.07 fb^{-1}. For the signal a cross-section of 100 fb was assumed.

Mass Point [GeV]	event yield	Mass Point [GeV]	event yield
$m_X = 1000,\, m_S = 400$	2.77	$m_X = 2000,\, m_S = 1500$	70.60
$m_X = 1000,\, m_S = 500$	2.13	$m_X = 2500,\, m_S = 1000$	172.60
$m_X = 1000,\, m_S = 750$	0.15	$m_X = 2500,\, m_S = 1500$	163.35
$m_X = 1500,\, m_S = 500$	76.31	$m_X = 2500,\, m_S = 2000$	84.76
$m_X = 1500,\, m_S = 750$	62.29	$m_X = 3000,\, m_S = 1000$	193.54
$m_X = 1500,\, m_S = 1000$	34.91	$m_X = 3000,\, m_S = 1500$	193.70
$m_X = 2000,\, m_S = 750$	136.39	$m_X = 3000,\, m_S = 2000$	170.32
$m_X = 2000,\, m_S = 1000$	133.48	$m_X = 3000,\, m_S = 2500$	88.43

Table A.2.: Signal yields in the signal region for partial Run 3 corresponding to a luminosity of 26.07 fb^{-1}. For the signal a cross-section of 100 fb was assumed.

A.7. Unblinded distributions

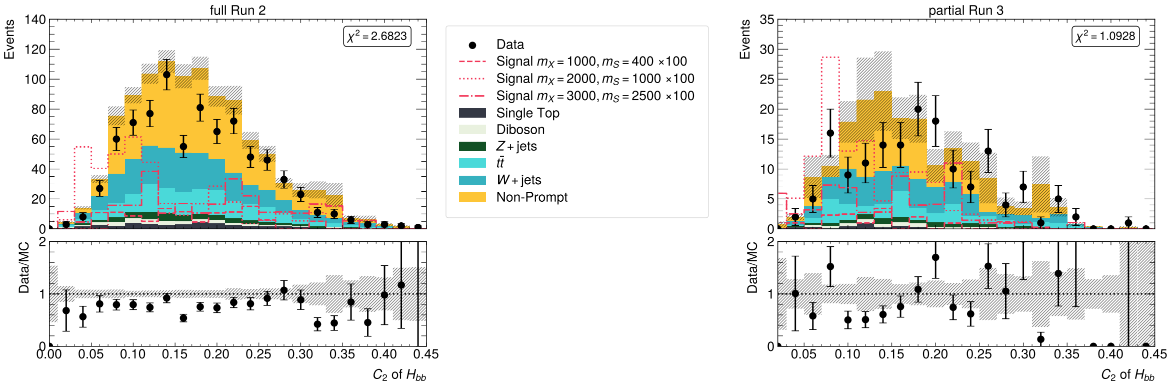


Figure A.4.: Jet substructure variable C_2 of the H_{bb} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

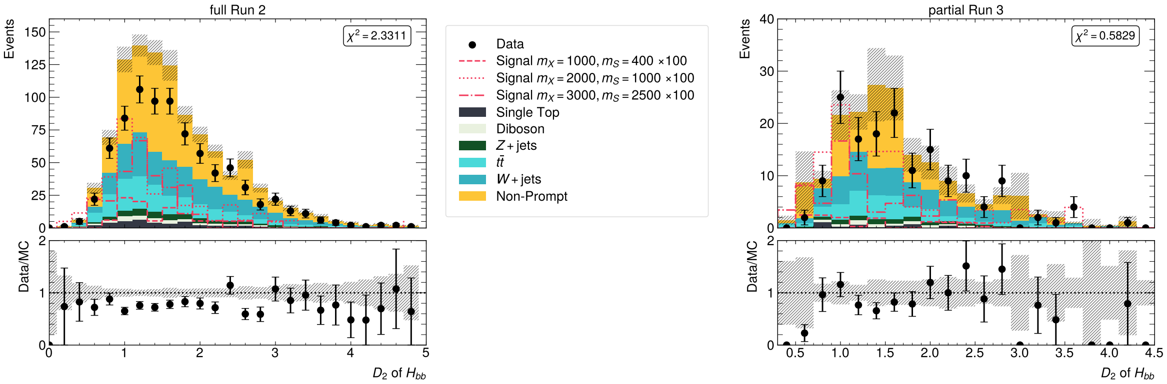


Figure A.5.: Jet substructure variable D_2 of the H_{bb} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

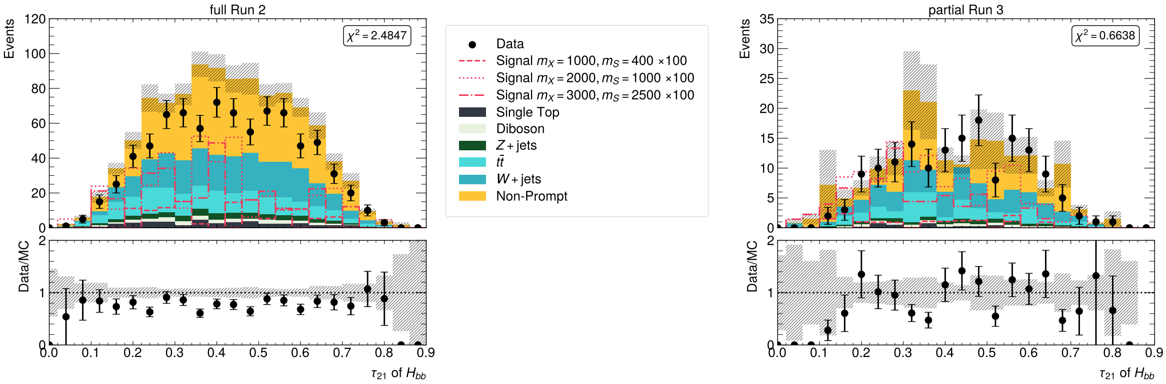


Figure A.6.: Jet substructure variable τ_{21} of the H_{bb} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

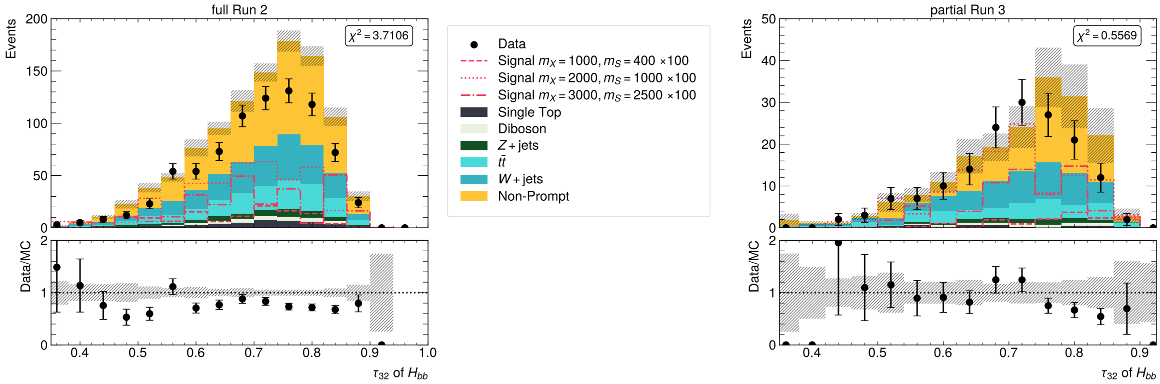


Figure A.7.: Jet substructure variable τ_{32} of the H_{bb} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

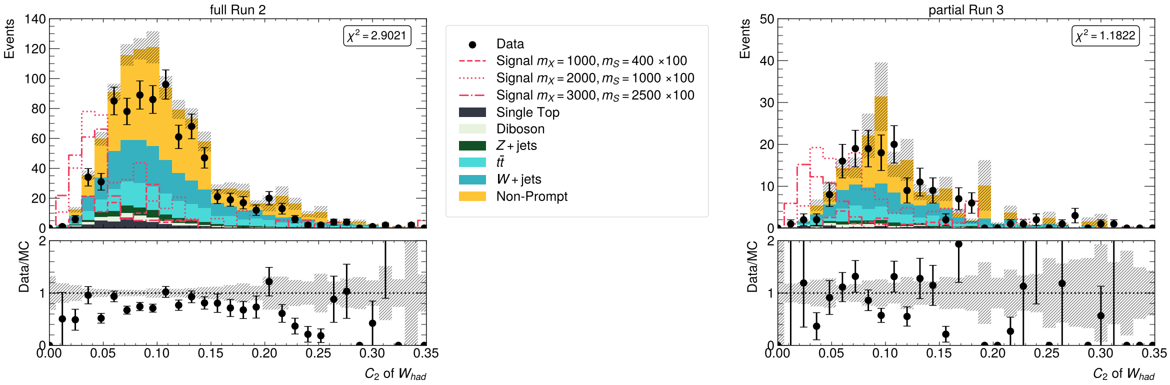


Figure A.8.: Jet substructure variable C_2 of the W_{had} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

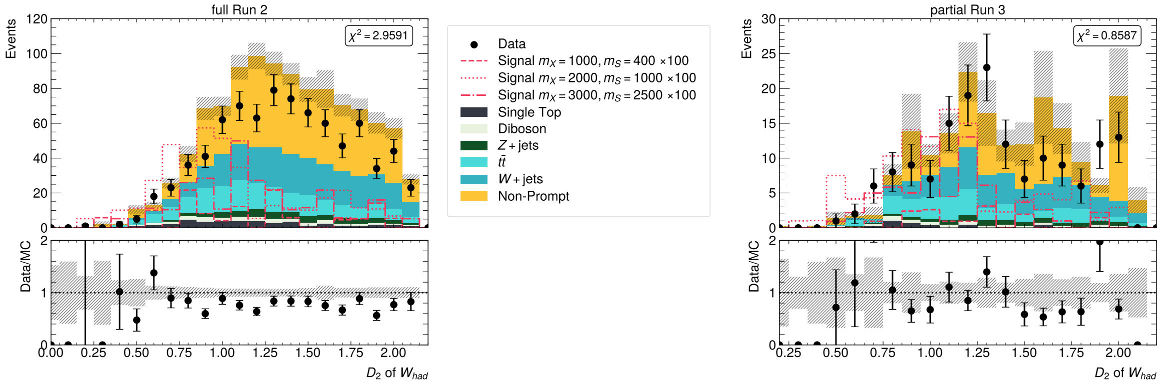


Figure A.9.: Jet substructure variable D_2 of the W_{had} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

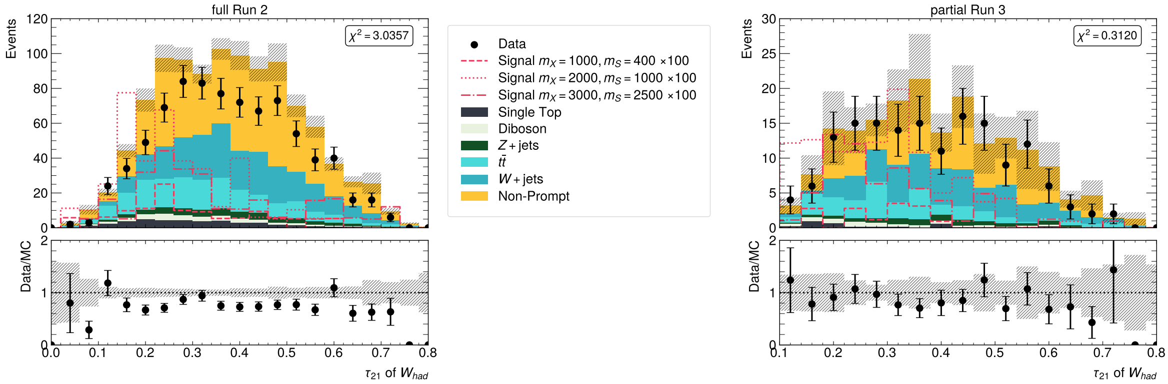


Figure A.10.: Jet substructure variable τ_{21} of the W_{had} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

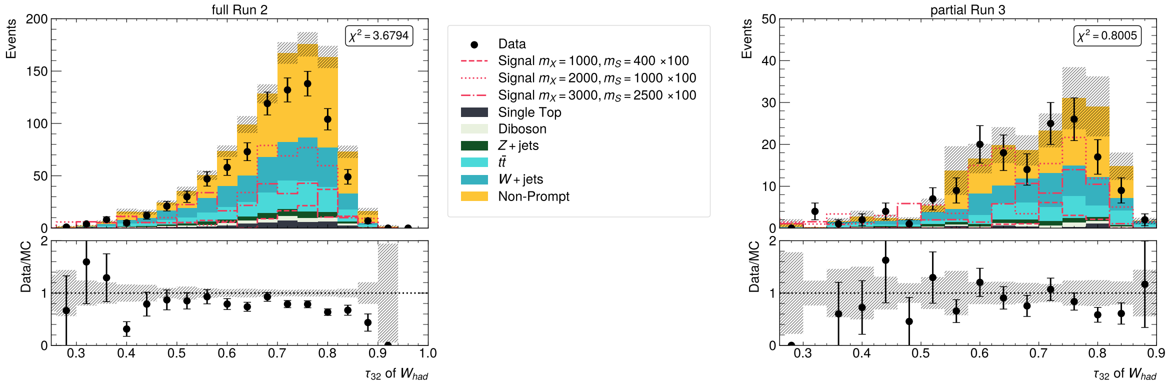


Figure A.11.: Jet substructure variable τ_{32} of the W_{had} candidate on the left for the full Run 2 dataset and on the right for the partial Run 3 dataset.

Thank you

First and foremost, I want to give a huge thanks to Prof. Stan Lai. This master's project was an amazing opportunity, and I am so grateful for the chance to learn, grow, and see what real research looks like. Working with you and being part of your group was always a joy, and I appreciate all your support along the way.

I also want to thank Jörn for being so kind and taking on the role of my second referee.

Thank you so much to my amazing supervisor, Kira. I honestly would not have made it through this without your endless knowledge and patience. Your guidance made all the difference, and I have learned so much from you. I wish you all the best for your future!

Also a big thank you to Lina! It was a great pleasure to work on my thesis at the same time as you. You were always there to help and it was very nice to share the wild ride of a master thesis with you. I am sure you will become a great teacher and wish you all the best for that!

To my wonderful office mate Zobi-thank you for making every day brighter. Sharing the office with you was one of the best parts of this year, and I could not have asked for a better friend by my side through it all. I have no doubt that you will achieve whatever you set your mind to, and I am wishing you all the best in everything that lies ahead. You've got this! And always remember our motto "work hard not smart"!

I also want to thank the rest of the Research Group for all the great times and laughs we had.

Finally, I extend my thanks to the $bbVV$ analysis group for their collaborative spirit and kindness. The questions and comments I received from this group greatly contributed to my learning over the past year.

Erklärung nach §17(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen: Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 23. Februar 2025

(Lena Schulz)