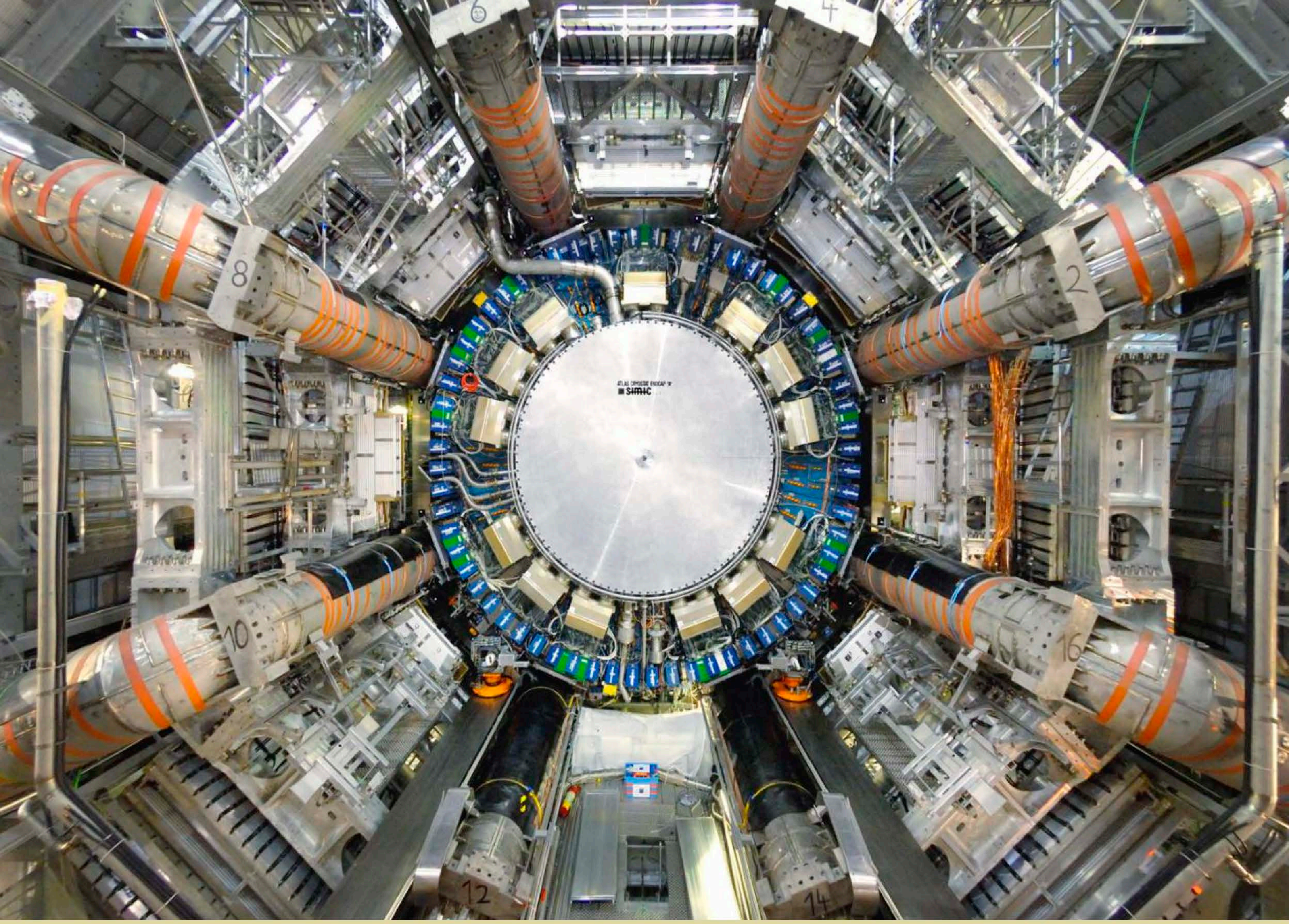


David Lopes Fernandes

Measurement of rare top quark decays in ATLAS/LHC

This dissertation is submitted for the degree of *Mestrado em Física*,
Supervised by Prof. Dr. Filipe Veloso, Presented to the Department of Physics of the Faculty of Science and Technology University of Coimbra

September 2018

· U C ·

UNIVERSIDADE DE COIMBRA

Measurement of rare top quark decays in ATLAS/LHC

David Lopes Fernandes

Supervisor: Prof. Dr. Filipe Veloso

Department of Physics of the Faculty of Science and Technology
University of Coimbra

This dissertation is submitted for the degree of
Mestrado em Física

2018

I would like to dedicate this thesis to my loving girlfriend, Rita, without whom this would have been finished a lot sooner.

Acknowledgements

I would like to thank my supervisor, Filipe Veloso, for all the guidance that he gave me throughout the year. Without him, my work would have succumbed to chaos and I would still be stuck in that one ROOT bug. Also to be acknowledged is Bruno Galhardo. If not for his assistance, this thesis would have taken three times as much to be completed.

Next on the list are the guys with whom I shared so many lunch hours. Your discussions (barring those that involved Pokémon) taught me much and made me laugh many times.

My fellow jiujitsukas are also to be thanked, especially my Sensei, Rui Costa. Who would have thought that getting beaten up twice a week could release so much tension?

A sincere thanks also goes to those friends I hold close to my heart: Tomás and Ricardo. They had to put up with my sense of humor many times when it was not up to par. Their honesty is always appreciated.

Finally, how could I not thank my parents? Without them, I would literally not be here. But without them, I also wouldn’t be half the person that I am today.

Abstract

In this thesis, the ATLAS detector sensitivity for the measurement of the top quark decay $t \rightarrow sW$ branching ratio is presented. Data collected by the ATLAS detector during the first two years of Run II (2015-2016) from proton-proton collisions at the Large Hadron Collider (LHC), with a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 36.2 fb^{-1}, scaled to $150\,\text{fb}^{-1}$, are analysed. As signal, top-quark pair-decay events are used, where one of the quark decays through the $t \rightarrow sW$ channel and the other through the main decay of the top quark, $t \rightarrow bW$. In the events chosen, both W bosons decay leptonically, with one decaying to an electron and another to a muon, of opposite charges. The branching ratio estimated is $\text{BR}(t \rightarrow sW) = 0.00162 \pm 0.00013$ (stat) ± 0.00014 (syst). Additionally, the branching ratio of $t \rightarrow bW$ is estimated to be $\text{BR}(t \rightarrow bW) = 0.9983 \pm 0.0060$ (stat) ± 0.0006 (syst). This result was obtained with a signal significance of 6.0 ± 0.9 (statistical uncertainty only). This leads to the conclusion that a measurement might be possible with the full Run II dataset.

Resumo

Nesta tese, a sensibilidade do detector ATLAS para a medida da fração de decaimento do quark top $t \rightarrow sW$ é apresentada. Os dados analisados foram recolhidos pelo detetor ATLAS durante os primeiros dois anos da Run II (2015-2016), em colisões de protão-protão no Grande Colisionador de Hadrões (LHC), com uma energia de centro de massa de $\sqrt{s} = 13$ TeV, correspondendo a uma luminosidade integrada de $36.2\,\mathrm{fb}^{-1}$, escalada para $150\,\mathrm{fb}^{-1}$. Como sinal, eventos de decaimento de pares de quarks top são usados, onde um dos quarks decai pelo canal $t \rightarrow sW$ e o outro pelo decaimento principal do quark top, $t \rightarrow bW$. Nos eventos escolhidos, ambos os bosões W decaem leptonicamente, com um a decair para um electrão e o outro para um muão, de cargas opostas. A fração de decaimento medida é $\mathrm{BR}(t \rightarrow sW) = 0.00162 \pm 0.00013$ (stat) ± 0.00014 (sist). Foi também medida a fração de decaimento de $t \rightarrow bW$, tendo sido obtido $\mathrm{BR}(t \rightarrow bW) = 0.9983 \pm 0.0060$ (stat) ± 0.0006 (sist). Este resultado foi obtido com uma significância do sinal de 6.0 ± 0.9 (apenas incerteza estatística). Isto leva a conclusão que uma medida pode ser possível com os dados da Run II completa.

Contents

List of figures

List of tables

Nomenclature

Roman letters

$\mathcal{L}$	instantaneous luminosity
$\mathscr{L}$	Lagrangian density
L	integrated luminosity

Greek letters

σ	cross section

Acronyms / Abbreviations

AI	Artificial Intelligence
ATLAS	A Toroidal LHC ApparatuS
AUC	Area under ROC curve
BR	Branching Ratio
CERN	Conseil Européen pour la Recherche Nucléaire
CKM	Cabbibo-Kobayashi-Maskawa
CR	Control Region
CSG	Cathode Strip Chambers
DL	Deep Learning
EF	Event Filter
FCal	Forward Calorimeter

GPU	Graphical Processing Unit
HEC	Hadronic Endcap Calorimeter
ID	Inner Detector
JER	Jet Energy Resolution
JES	Jet Energy Scale
JVF	Jet-Vertex Fraction
LAr	Liquid Argon
LHC	Large Hadron Collider
MC	Monte Carlo
MDT	Monitored Drift Tubes
ML	Machine Learning
NNLL	Next-to-next-to leading logarithmic
LO	Leading order
NLO	Next-to-leading order
NNLO	Next-to-next-to-leading order
PDF	Parton Distribution Function
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
MS	Muon System
RMS	Root-mean-squared
RPC	Resistive Plate Chambers
SCT	Semiconductor Tracker
SGD	Stochastic Gradient Descent
SM	Standard Model of Particle Physics

SR	Signal Region
TGC	Thin Gap Chambers
TileCal	Tile Calorimeter
TRT	Transition Radiation Tracker
VR	Validation Region

Chapter 1

Introduction

The top quark is the heaviest elementary particle known and, for that reason, is an excellent probe for the Standard Model of Particle Physics (SM). Its mass is close to the energy scale of the electroweak symmetry breaking and so can provide information on symmetry breaking mechanisms, as well as the Higgs mechanism. This would allow the study of Physics Beyond the SM, by looking for any differences between what is predicted by this model and what is in fact measured. However, to do so, the properties of the top quark must be precisely measured. These include its mass, the production mechanisms and effective cross sections, its electric charge, spin, the asymetries in its decays, its rare decays, amongst others.

According to the SM, the top quark decays almost exclusively to a W boson and a down-type quark, mainly a bottom (b) quark. The corresponding decay branching ratios (BR) are related to the Cabbibo–Kobayashi–Maskawa (CKM) mixture matrix, two of which elements have not yet been measured, V_{ts} and V_{td}, but only determined based on the assumption that the CKM matrix is unitary. A direct measurement of these elements would allow the testing of the assumptions associated with the properties of this matrix as, for example, the number of quark families in the SM.

This thesis aims to present the work done to study the sensitivity for the direct measurement of one of these elements, V_{ts}. To do that, data collected by the ATLAS detector during the first two years of Run II (2015-2016) from proton-proton collisions at the Large Hadron Collider (LHC), with a center-of-mass energy of $\sqrt{s} = 13$ TeV and a total integrated luminosity of 36.2 fb^{-1}, scaled to 150 fb^{-1} (the expected integrated luminosity of the full Run II), are analysed. The events where this decay happens are searched for in $t\bar{t}$ production, with one top quark decaying through the SM dominant decay, to a W boson and a b-quark, and the other to a strange (s) quark and a W boson. One W boson then decays to an e^- or a e^+ and the other to a μ^+ or μ^-.

The outline of the thesis is as follows. Chapter 2 gives an introduction to the SM, giving particular focus to the top quark and its properties. An overview of the LHC and the ATLAS detector is given in Chapter 3, with each sub-system of the detector described. On Chapter 4, the measurement of the top quark decay to sW is presented, which includes descriptions of the background and signal samples used, the strategy implemented to control and measure the backgrounds, a linearity study performed to estimate the b-tagging efficiency uncertainty, the systematics involved in the analysis and, finally, the calculation of $t \rightarrow sW$ and $t \rightarrow bW$ branching ratios. In the end, in Chapter 5, the final conclusions are presented.

Throughout this thesis, natural units will be used ($\hbar = c = 1$), and energies, masses and momenta will be expressed in GeV, unless otherwise stated.

Chapter 2

Theoretical Framework

This chapter gives a brief overview of the Standard Model of Particle Physics (SM). Besides listing the various particles in the SM and their interactions, some of the theory behind these interactions will be described. Finally, the top quark and its properties will be discussed. What is presented here is based on [1, 2]. For the top quark sections, [3, 4] were used.

2.1 Particles of the Standard Model

So far, all the experimental observations and discoveries indicate that all matter is formed by twelve fundamental particles, named leptons and quarks, which are all fermions (with spin 1/2). Both quarks and leptons are in turn divided into three families (or generations). For the leptons, there are the electron (e) and electronic neutrino (ν_e), the muon (μ) and muonic neutrino (ν_μ) and the tau (τ) and the tauonic neutrino (ν_τ). In the case of the quarks, the up (u) and down (d) quarks form the first generation of elementary particles, the strange (s) and charm (c) the second, and the bottom (b) and top (t) quarks the third. All of these particles, with the exception, perhaps, of the neutrinos, have their corresponding anti-particle. They also have both a left-handed and a right-handed component, which is important if they are to have mass, as will be shown.

The fermions interact with one another. Mediating these interactions are other particles, the gauge bosons (with spin 1). The SM describes three of these interactions. The electromagnetic interaction, which exists between electrically charged particles, is mediated by the photon (γ). As the photon is massless, the electromagnetic interaction has an infinite range. The weak interaction is mediated by three particles, the $W^{\pm}$ and the Z. Both of these interactions were merged into the electroweak theory in the 1960s, by Glashow, Weinberg and Salam [5]. Finally, the strong interaction acts on particles with a color charge and is mediated by the gluon, of which there are eight. These interactions will be further described later on.

A fundamental interaction that is missing from the SM is gravity. While being well described by the General Relativity Theory, there is no known particle mediating it. One possibility is the graviton, which has spin 2.

An overview of the fermions of the SM is given on Table 2.1. Besides the electric charge, Q, and the color charge, C, the hypercharge, Y, the isospin and its third component, T and T_3, respectively, are also indicated. Similar information for the gauge bosons is given on Table 2.2.

A third class of fundamental particles in the SM is formed by the Higgs boson. With spin 0, this particle is not a gauge boson and does not mediate any interaction. Instead, this particle is credited with assigning mass to elementary particles. The Higgs mechanism, as it is called, will be discussed on a later section.

Table 2.1 Fermions in the Standard Model and their quantum numbers.

Generations			Quantum Numbers				
1	2	3	T	T_3	Y	Q	C
	Leptons						
$\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_L$	$\begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_L$	$\begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_L$	$\begin{pmatrix}1/2\\ 1/2\end{pmatrix}$	$\begin{pmatrix}+1/2\\ -1/2\end{pmatrix}$	$\begin{pmatrix}-1\\ -1\end{pmatrix}$	$\begin{pmatrix}0\\ -1\end{pmatrix}$	—
e^-_R	μ^-_R	τ^-_R	0	0	-2	-1	—
	Quarks						
$\begin{pmatrix}u\\ d\end{pmatrix}_L$	$\begin{pmatrix}c\\ s\end{pmatrix}_L$	$\begin{pmatrix}t\\ b\end{pmatrix}_L$	$\begin{pmatrix}1/2\\ 1/2\end{pmatrix}$	$\begin{pmatrix}+1/2\\ -1/2\end{pmatrix}$	$\begin{pmatrix}+1/3\\ +1/3\end{pmatrix}$	$\begin{pmatrix}+2/3\\ -1/3\end{pmatrix}$	$\begin{pmatrix}r,g,b\\ r,g,b\end{pmatrix}$
u_R	c_R	t_R	0	0	$+4/3$	$+2/3$	r,g,b
d_R	s_R	b_R	0	0	$-2/3$	$-1/3$	r,g,b

Table 2.2 Gauge bosons in the Standard Model. Masses taken from [2].

Boson	Interaction	Mass M	Electric Charge Q
Photon γ	Electromagnetic	0	0
$W^\pm$	Weak	(80.385 ± 0.015) GeV	± 1
Z	Weak	(91.1876 ± 0.0021) GeV	0
Gluon g	Strong	0	0

2.2 Interactions of the Standard Model

To describe the SM mathematically, the Lagrangian formalism is used. With it, one can describe the interactions and masses of a theory using a Lorentz invariant called Lagrangian density, represented by $\mathscr{L}$. Ensuring that the the physical system is locally gauge invariant results, according to Noether's theorem [6], in a conserved current. This in turn introduces the force carriers (the gauge bosons that mediate the interaction) in the theory.

The local gauge invariance can be seen as a symmetry. For every interaction of the theory there is a corresponding symmetry, which, in the mathematical formalism, is represented by a symmetry group. The total symmetry group of the SM is $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. The first part, $SU(3)_C$, is related to the strong interaction, described by the Quantum Chromodynamics (QCD) theory. The last two parts are related to the electroweak force, the unification of the weak force $(SU(2)_L)$ and the electromagnetic one $(U(1)_Y)$. This symmetry is broken spontaneously by the Higgs mechanism. It is this breaking that gives rise to the Higgs boson.

The Lagrangian density of the SM is then given by

$$\mathscr{L}_{SM} = \mathscr{L}_{EW} + \mathscr{L}_{QCD} + \mathscr{L}_{\text{Higgs}}, \tag{2.1}$$

where $\mathscr{L}_{EW}$ describes the electroweak part of the theory, $\mathscr{L}_{QCD}$ the strong part and $\mathscr{L}_{\text{Higgs}}$ the Higgs mechanism.

2.2.1 Electromagnetic interaction

Firstly, the electromagnetic interaction was formalized by the Quantum Electrodynamics (QED) theory, developed in the 1940s. Beginning with the Dirac Lagrangian of the freely propagating fermion of mass m

$$\mathscr{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \tag{2.2}$$

where ψ is a 4-component spinor field representing the fermion and γ^μ are the Dirac matrices, it can be seen that it is invariant under a global transformation like

$$\psi \to e^{i\alpha}\psi, \tag{2.3}$$

with α being any real number. The set of these values make up the unitary transformation $U(1)$. However, when using a local transformation, where α is a function of the space-time coordinates, i. e.,

$$\psi \to e^{i\alpha(x)}\psi, \tag{2.4}$$

one needs to replace the derivative ∂_μ by a covariant derivative

$$D_\mu = \partial_\mu + iqA_\mu \tag{2.5}$$

to keep the Lagrangian locally invariant. Here, q is the charge of the fermion. The new vector field A_μ transforms like

$$A_\mu \rightarrow A_\mu - \frac{1}{q}\partial_\mu \alpha(x). \tag{2.6}$$

However, this Lagrangian is incomplete. To obtain the full Lagrangian, terms describing the kinetic energy and mass of the new particle introduced by this vector field must be included. As can be shown, the mass term would again violate the local gauge invariance and so it must be zero, leading then to the massless photon. The kinetic term is given by $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The full QED Lagrangian is then,

$$\begin{aligned} \mathscr{L}_{QED} &= \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} && (2.7)\\ &= \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - q\bar{\psi}\gamma^\mu\psi A_\mu, && (2.8) \end{aligned}$$

with the first two terms corresponding, respectively, to the kinetic and mass terms of the fermion, the third to the kinetic term of the photon and the last to the interaction term, with the strength being proportional to the electric charge q.

2.2.2 Strong Interaction

The strong interaction, described by QCD, happens between particles that have color charge: the quarks and the gluons. The quarks can have one of three colors – red (r), green (g) and blue (b) – as well as their anti-colors ($\bar{r}, \bar{g}, \bar{b}$), while gluons have one unit of color and another of anti-color (for example, $g\bar{r}$).

The Lagragian of a free quark is of the same form as Equation 2.2. However, in this case, it must be invariant under the $SU(3)_C$ (where C corresponds to color) transformations

$$\psi \rightarrow e^{i\alpha_a(x)T^a}\psi, \tag{2.9}$$

where $\alpha_a(x)$ are functions of space-time and T_a are the eight generators of $SU(3)$ (the Gell-Mann matrices are usually used). A sum over the repeated suffix a is implied.

The derivate ∂_μ is then replaced by the covariant derivative

$$D_\mu = \partial_\mu + igT_a\mathcal{A}^a_\mu, \tag{2.10}$$

where g is the coupling constant of the strong interaction and each $\mathcal{A}^a_\mu$ is one of the 8 gluon fields, transforming like

$$\mathcal{A}^a_\mu \to G^a_\mu - \frac{1}{g}\partial_\mu \alpha^a(x) - f^{ab}_c \alpha_b(x)\mathcal{A}^c_\mu, \tag{2.11}$$

where f_{abc} are the structure constants and come from $[T_a, T_b] = if^c_{ab}T_c$. The final QCD Lagrangian is then

$$\mathscr{L}_{QCD} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a - g(\bar{\psi}\gamma^\mu T_a\psi)\mathcal{A}^a_\mu. \tag{2.12}$$

Here, the first two terms are again the kinetic and mass terms of the fermions (in this case, quarks), the third is the gluon kinetic term (with $G^a_{\mu\nu} = \partial_\mu\mathcal{A}^a_\nu - \partial_\nu\mathcal{A}^a_\mu - gf^a_{bc}\mathcal{A}^b_\mu\mathcal{A}^c_\nu$) and the last is the interaction term.

While this Lagrangian seems similar to the QED one, the gluon kinetic term contains trilinear and quartic terms, allowing the gluons to self-interact, as opposed to the photons, that do not. This leads the strong interaction to act in different ways. At small distances, the force has a weak intensity. This is called the asymptotic free behaviour of the strong interaction. However, if the particles are pulled appart, the strong force increases to the point that it is more energetically efficient to create new particles than to separate them further. This is called the confinement regime. Consequently, only colourless particles (color/anti-color or all(-anti)-three-colors combinations) can be observed.

2.2.3 Weak Interaction

The weak force is responsible for interactions between particles with non-vanishing weak isospin. Its gauge bosons can only interact with left-handed fermions or right-handed anti-fermions. It can be described by the $SU(2)_L$ symmetry group, where L corresponds to left-handedness. The Lagrangian density of a free fermion is

$$\mathscr{L} = \bar{\psi}_L(i\gamma^\mu\partial_\mu - m)\psi_L. \tag{2.13}$$

The spinor must transform like

$$\psi_L \to e^{i\boldsymbol{\alpha}(x)\cdot\frac{\boldsymbol{\tau}}{2}}\psi_L, \tag{2.14}$$

where $\boldsymbol{\tau}$ are the generators of $SU(2)$ (the Pauli matrices). Once more, if the Lagrangian is to be locally gauge invariant, the derivative ∂_μ must be replaced by

$$D_\mu = \partial_\mu + ig\frac{\boldsymbol{\tau}}{2}\cdot \boldsymbol{W}_\mu, \tag{2.15}$$

where g is the weak interaction coupling constant and $\boldsymbol{W}_\mu$ are three new massless vector fields that transform as

$$\boldsymbol{W}_\mu \rightarrow \boldsymbol{W}_\mu - \frac{1}{g}\partial_\mu \boldsymbol{\alpha} - \boldsymbol{\alpha}\times \boldsymbol{W}_\mu. \tag{2.16}$$

The full Lagrangian, including the kinetic term of the gauge bosons, is

$$\mathscr{L}_{\text{Weak}} = \bar{\psi}_L(i\gamma^\mu\partial_\mu - m)\psi_L - g(\bar{\psi}_L\gamma^\mu\frac{\boldsymbol{\tau}}{2}\psi_L)\cdot \boldsymbol{W}_\mu - \frac{1}{4}\boldsymbol{W}_{\mu\nu}\cdot \boldsymbol{W}^{\mu\nu}. \tag{2.17}$$

One property of the weak interaction is that it allows fermions of one generation to decay to fermions of another one. In the case of the quarks, this can be formalized by writing the weak eigenstates of the quarks as a linear combination of their strong eigenstates. Connecting all three generations gives a unitary matrix (primed variables are weak interaction quarks)

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.18}$$

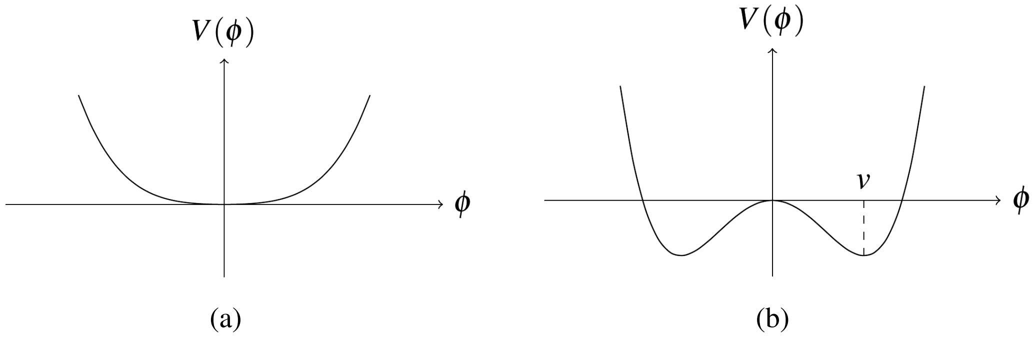
This is called the Cabbibo–Kobayashi–Maskawa (CKM) matrix. The absolute value squared of each element of the matrix is equal to the BR of one of the quarks decaying to the other. For example, $|V_{ts}|^2$ is the probability of the top quark decaying to the strange quark.

As in the case of the electromagnetic interaction, if the mass term of the gauge bosons was added to the Lagrangian density, it would stop being locally gauge invariant. Unlike the electromagnetic interaction, though, the gauge bosons of the weak interaction are not massless. Some way had to be found to incorporate the masses of the gauge bosons, while keeping the local gauge invariance.

2.2.4 Higgs mechanism

This is done by spontaneously breaking the symmetry. To illustrate this phenomenon, one may start with this Lagrangian density for scalar particles:

$$\mathscr{L} = T - V = \frac{1}{2}(\partial_\mu\phi)^2 - (\frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4), \tag{2.19}$$



(a) (b)

Fig. 2.1 The potential $V(\phi)$ for (a) $\mu^2 > 0$ and (b) $\mu^2 < 0$, and $\lambda > 0$.

with $\lambda > 0$. The potential $V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$ is plotted in Figure 2.1. In the case where $\mu^2 < 0$, the Lagrangian has a mass term of the wrong sign for the field ϕ, as the signs of the kinetic energy and the ϕ^2 term are the same. However, in this case, the minimum of the potential is not at $\phi = 0$, but at $\phi = \pm v = \pm\sqrt{-\mu^2/\lambda}$. If the field is slighly perturbed around one of these minima, which can be $\phi = v$, for instance, the substitution $\phi(x) = v + \eta(x)$ in the Lagrangian gives

$$\mathscr{L}' = \frac{1}{2}(\partial_\mu\eta)^2 - \lambda v^2\eta^2 - \lambda v\eta^3 - \frac{1}{4}\lambda\eta^4 + \text{const.} \tag{2.20}$$

Here, the mass term of η has the correct sign and is

$$m_\eta = \sqrt{2\lambda\eta^2} = \sqrt{-2\mu^2}. \tag{2.21}$$

Doing perturbation around one of the ground states, instead of at $\phi = 0$ (which is not a minimum) reveals that the scalar particles in fact have mass. This principle of spontaneous symmetry breaking is what is behind the Higgs mechanism.

2.2.5 Electroweak theory

The previous example, however, is too simple to fully explain the Higgs mechanism. To do so, the electroweak theory, the unification of the weak and electromagnetic interactions, will be described. Then, its symmetry will be spontaneously broken to obtain the gauge bosons masses, as well as the Higgs boson.

As said before, the electroweak is represented by the symmetry groups $SU(2) \otimes U(1)$, with the first one having as generator the operator $\boldsymbol{\tau}$ and the second one the operator Y. Since the first one can only be applied to left-handed fermions, the two transformations then are

$$\begin{aligned}\psi_L \to \psi_L' &= e^{i\frac{\boldsymbol{\tau}}{2}\cdot\boldsymbol{\alpha}(x)+i\beta(x)Y}\psi_L \quad \text{and} \\ \psi_R \to \psi_R' &= e^{i\beta(x)Y}\psi_R.\end{aligned} \tag{2.22}$$

The procedure now is the same as for the electromagnetic (with $A_\mu \to B_\mu$, and the electric charge being replaced by the hypercharge Y) and the weak interactions. The covariant derivative (for the left-handed part) is then

$$D_\mu = \partial_\mu + ig\frac{\boldsymbol{\tau}}{2}\cdot\boldsymbol{W}_\mu + ig'\frac{Y}{2}B_\mu \tag{2.23}$$

and the electroweak Lagrangian is

$$\begin{aligned}\mathscr{L}_{EW} =& \bar{\psi}_L\gamma^\mu[i\partial_\mu - g\frac{\boldsymbol{\tau}}{2}\cdot\boldsymbol{W}_\mu - g'\frac{Y_L}{2}B\mu]\psi_L \\ &+ \bar{\psi}_R\gamma^\mu[i\partial_\mu - g'\frac{Y_R}{2}B_\mu]\psi_R - \frac{1}{4}\boldsymbol{W}_{\mu\nu}\cdot\boldsymbol{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}.\end{aligned} \tag{2.24}$$

With this in mind, the Higgs mechanism must be formulated in such a way so that the $W^\pm$ and the Z become massive and the photon remains massless. In this case, four real scalar fields are necessary

$$\phi = \begin{pmatrix}\phi^+ \\ \phi^0\end{pmatrix}, \tag{2.25}$$

with

$$\begin{aligned}\phi^+ &= (\phi_1 + i\phi_2)/\sqrt{2} \\ \phi^0 &= (\phi_3 + i\phi_4)/\sqrt{2}.\end{aligned} \tag{2.26}$$

The Lagrangian in this case is

$$\mathscr{L} = \left|\left(i\partial_\mu - g\frac{\boldsymbol{\tau}}{2}\cdot\boldsymbol{W}_\mu - g'\frac{Y}{2}B_\mu\right)\phi\right|^2 - V(\phi), \tag{2.27}$$

where $|x|^2 = (x)^\dagger(x)$ and $Y = 1$. The vacuum expectation value, ϕ_0, chosen is

$$\phi_0 = \sqrt{\frac{1}{2}}\begin{pmatrix}0 \\ v\end{pmatrix}. \tag{2.28}$$

By slightly perturbing the field around this value, $v \to v + h(x)$, a term in order of h^2 appears. This corresponds to the mass term of the Higgs boson. The relevant term for the gauge bosons masses is

$$\begin{aligned}
&\left|\left(-ig\frac{\boldsymbol{\tau}}{2}\cdot\boldsymbol{W}_\mu - i\frac{g'}{2}B_\mu\right)\phi\right|^2 \\
&\quad = \frac{1}{8}\left|\begin{pmatrix} gW^3_\mu + g'B_\mu & g(W^1_\mu - iW^2_\mu) \\ g(W^1_\mu + iW^2_\mu) & -gW^3_\mu + g'B_\mu \end{pmatrix}\begin{pmatrix} 0 \\ v \end{pmatrix}\right|^2 \\
&\quad = \frac{1}{8}v^2g^2\left[(W^1_\mu)^2 + (W^2_\mu)^2\right] + \frac{1}{8}v^2(g'B_\mu - gW^3_\mu)(g'B^\mu - gW^{3\mu}) \\
&\quad = \left(\frac{1}{2}vg\right)^2 W^+_\mu W^{-\mu} + \frac{1}{8}v^2[gW^3_\mu - g'B_\mu]^2 + 0[g'W^3_\mu + gB_\mu]^2,
\end{aligned} \tag{2.29}$$

where $W^\pm = (W^1 \mp iW^2)/\sqrt{2}$. The first term can be identified as the mass term of the charged boson, $M_W^2 W^+W^-$, with

$$M_W = \frac{1}{2}vg. \tag{2.30}$$

The last two terms can be identified as mass terms of the Z and γ bosons, respectively. Their corresponding fields (normalized) and masses are

$$Z_\mu = \frac{gW^3_\mu - g'B_\mu}{\sqrt{g^2 + g'^2}}, \qquad \text{with} \quad M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}; \tag{2.31}$$

$$A_\mu = \frac{g'W^3_\mu + gB_\mu}{\sqrt{g^2 + g'^2}}, \qquad \text{with} \quad M_A = 0. \tag{2.32}$$

And so the symmetry is broken with the bosons associated with the weak interaction having mass and the photon being massless. Additionally, the Higgs boson is also revealed.

2.2.6 Yukawa couplings

However, so far, only bosons (with the exception of the photons) have mass. The fermions remain massless. Fortunately, the Higgs mechanism also introduces mass terms for them. Looking only at the case of the electron, this $SU(2) \otimes U(1)$ gauge invariant term may be included in the Lagrangian

$$\mathscr{L}' = -G_e \left[(\bar{\nu}_e, \bar{e})_L \phi e_R + \bar{e}_R \phi \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \right], \tag{2.33}$$

where G_e is known as a Yukawa coupling. By spontaneously breaking the symmetry and making the substitution

$$\phi = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{2.34}$$

the term becomes

$$\mathscr{L}' = -\frac{G_e}{\sqrt{2}} v (\bar{e}_L e_R + \bar{e}_R e_L) - \frac{G_e}{\sqrt{2}} v (\bar{e}_L e_R + \bar{e}_R e_L) h. \tag{2.35}$$

Choosing G_e so that $m_e = G_e v / \sqrt{2}$, the required electron mass is generated:

$$\mathscr{L}' = -m_e \bar{e} e - \frac{m_e}{v} \bar{e} e h. \tag{2.36}$$

2.3 Top quark

Having been discovered in 1995 by the CDF and the DØ experiments [7, 8] in $p\bar{p}$ collisions with centre-of-mass energy of 1.8 TeV at Fermilab's Tevatron collider, the top quark is the heaviest elementary particle known, with $m_t = 173.21 \pm 0.51 \pm 0.71$ GeV [2] (see Figure 2.2). Due to this, its lifetime is extremely small ($\approx 0.5 \times 10^{-25}$ s [2]) and so it is possible for it to decay before it enters the confinement regime. This allows the indirect study of isolated quarks by looking at the decay products of the top quark.

The top quark decays almost exclusively to a W and a b-quark (Figure 2.3), with $|V_{tb}| = 1.009 \pm 0.031$ [2] (see also Figure 2.4), obtained directly, without assuming unitarity of the CKM matrix, from the top-quark production cross-section. The second and third most likely decays are $t \to sW$ and $t \to dW$. The CKM matrix elements corresponding to these decays have not yet been measured. Their theoretical values are $|V_{ts}| = (40.0 \pm 2.7) \times 10^{-3}$ and $|V_{td}| = (8.2 \pm 0.6) \times 10^{-3}$ [2]. A direct measurement of these elements would allow the testing of the unitarity of the CKM matrix, as well as the possibility of the existence of other generations of quarks.

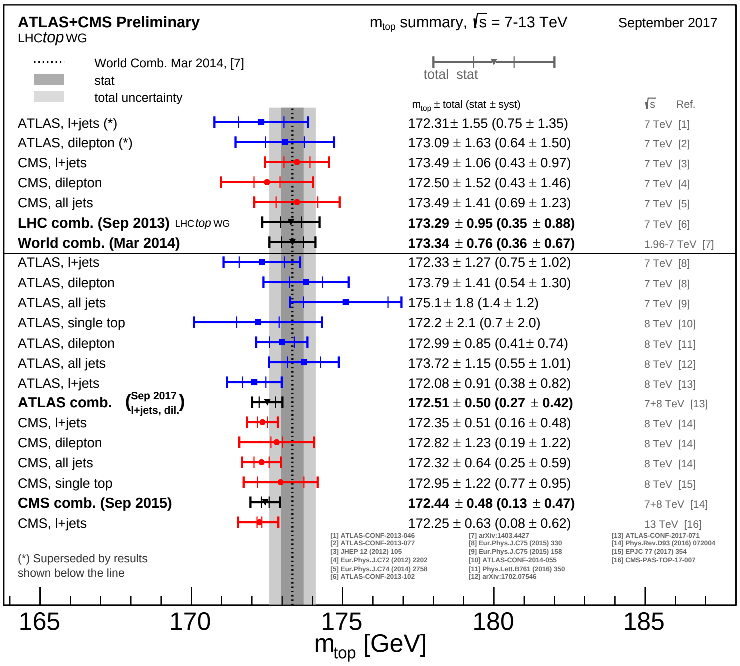


Fig. 2.2 Summary of the ATLAS and CMS direct m_t measurements. The results are compared with the LHC and Tevatron+LHC m_t combinations. Taken from [9].

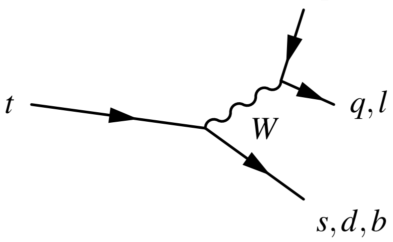


Fig. 2.3 Dominant top quark decay channel $t \to qW$, with $q = b, s, d$.

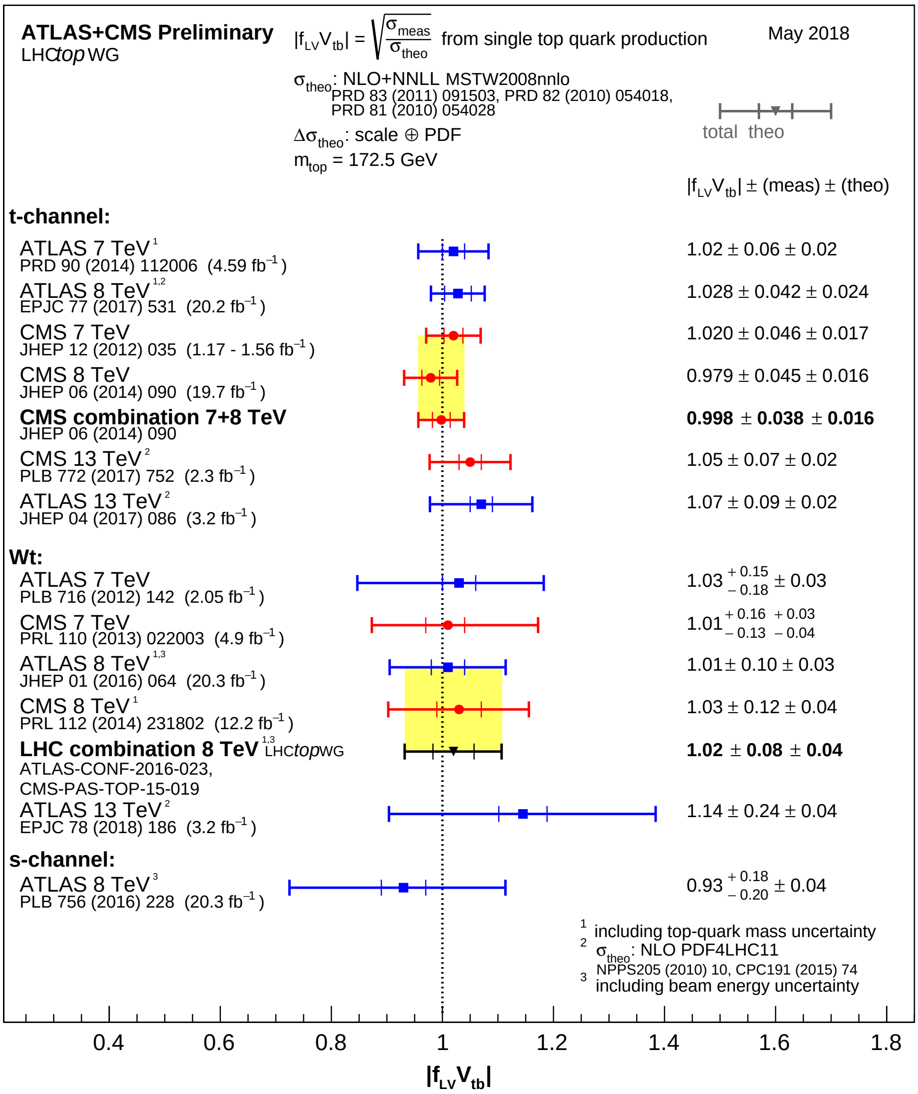


Fig. 2.4 Summary of the ATLAS and CMS extractions of the CKM matrix element V_{tb} from single top quark measurements. For each result, the contribution to the total uncertainty originating from the uncertainty on the theoretical prediction for the single top production cross-section is shown along with the uncertainty originating from the experimental measurement of the cross-section. Taken from [10].

2.3.1 Top-quark production

The top quark can be produced mainly in two ways: pair production ($t\bar{t}$) and single production (single top). The first one happens due to the strong interaction and the second one due to the weak. Leading-Feynman diagrams for these two processes are shown in Figures 2.5 and 2.6.

The top-quark pair production can happen through gluon fusion or $q\bar{q}$ annihilation. At the LHC with $\sqrt{s} = 13$ TeV, gluon fusion (Figure 2.5 (a), (b) and (c)) dominates, producing approximately 90% of $t\bar{t}$ pairs, while the $q\bar{q}$ annihilation (Figure 2.5 (d)) produces the remaining 10%. In Figure 2.7 the most recent public results from ATLAS and CMS of $t\bar{t}$ production cross-section measurement are summarized.

As was mentioned previously, the single-top production is due to the weak interaction. This can happen through three different channels: the t-channel with the exchange of a virtual W boson (Figure 2.6 (a)); the s-channel also with the exchange of a virtual W boson (Figure 2.6 (d)); and the associated production of a top quark and a W boson, the Wt-channel (Figure 2.6 (b) and (c)). The most recent results for this process cross-section measurement are summarized in Figure 2.8.

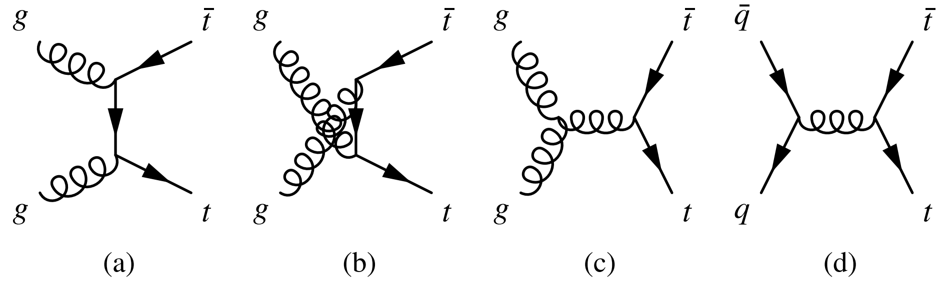


Fig. 2.5 Leading-order Feynman diagrams of the $t\bar{t}$ production. Gluon scattering processes (a), (b) and (c) are the dominant processes at LHC, while quark scattering (d) is the dominant one at Tevatron.

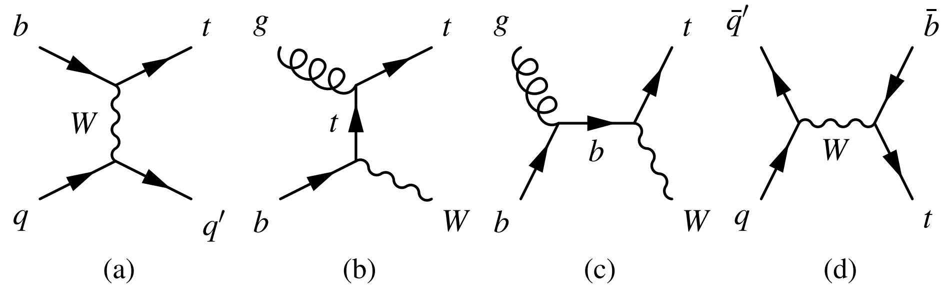


Fig. 2.6 Leading-order Feynman diagrams of the single-top production, (a) t-channel, (b) and (c) Wt-channel and (d) s-channel.

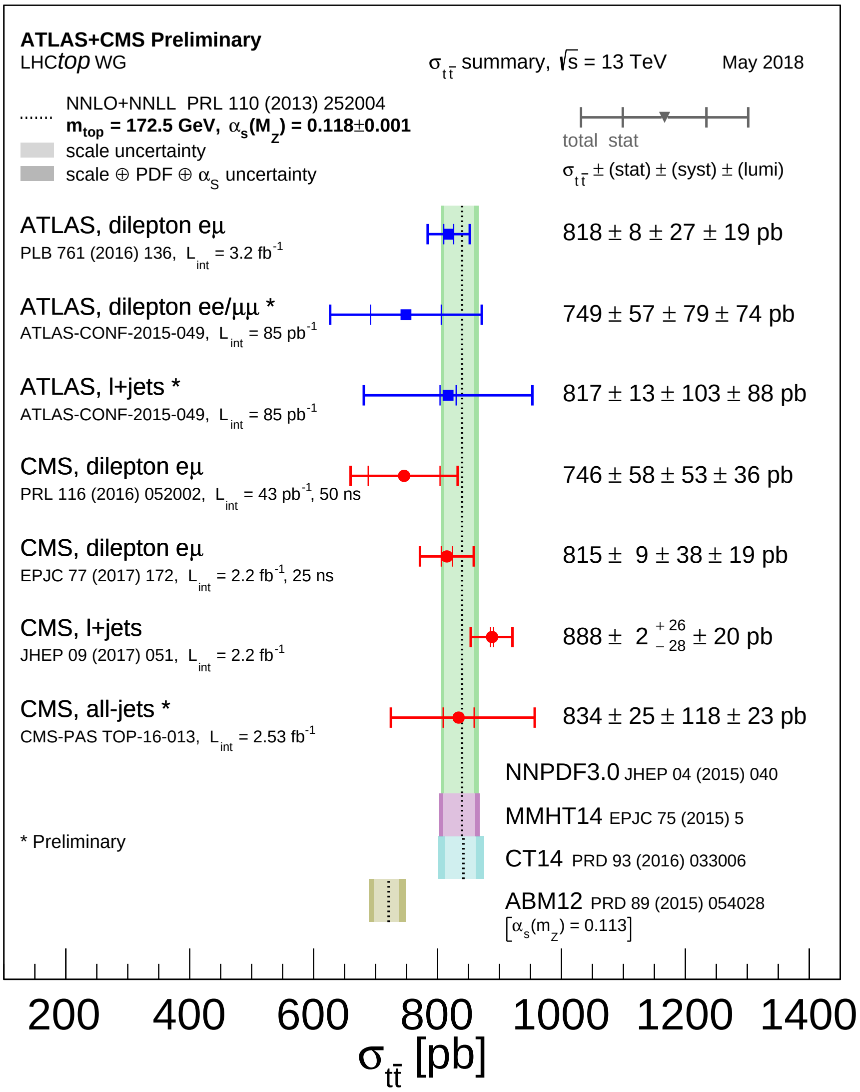


Fig. 2.7 Summary of measurements of the top-pair production cross-section at 13 TeV compared to the exact next-to-next-to-leading order (NNLO) QCD calculation complemented with next-to-next-to-leading logarithmic (NNLL) resummation (`top++2.0`). The theory band represents uncertainties due to renormalisation and factorisation scale, parton density functions and the strong coupling. The measurements and the theory calculation are quoted for $m_t = 172.5$ GeV. Taken from [11].

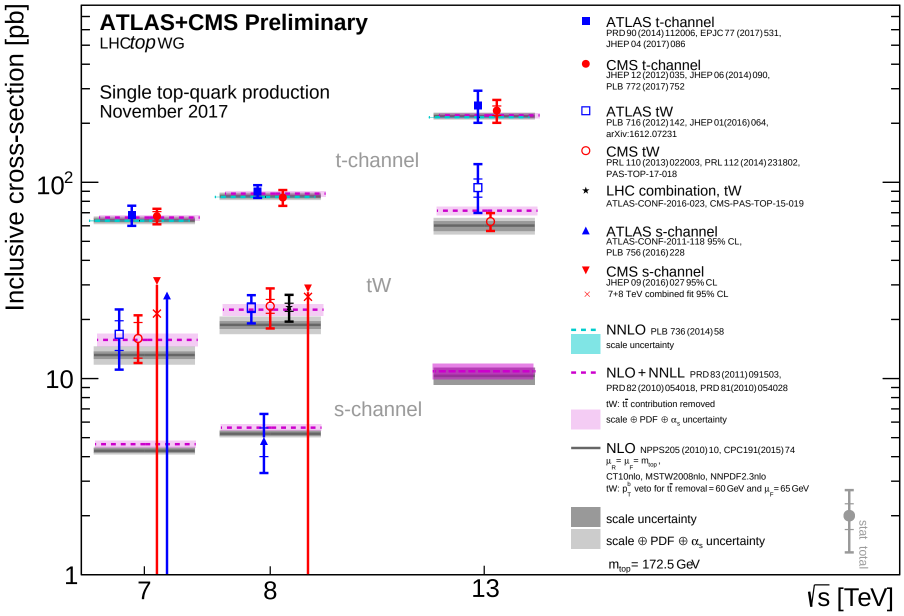


Fig. 2.8 Summary of ATLAS and CMS measurements of the single top production cross-sections in various channels as a function of the center of mass energy. The measurements are compared to theoretical calculations based on: NLO QCD, NLO QCD complemented with NNLL resummation and NNLO QCD (t-channel only). Taken from [12].

Chapter 3

Experimental apparatus

In this chapter, an overview of the the CERN laboratory is given, focusing mainly on the Large Hadron Collider and the ATLAS detector.

3.1 CERN

Founded in 1954, the European Organization for Nuclear Research (or CERN, for the French *Conseil Européen pour la Recherche Nucléaire*) is the world's largest particle physics laboratory, located near Geneva, hosting over twelve thousand visiting scientists and engineers from all over the world [13].

Being one of Europe's first joint ventures, it allowed its member states, of which there were twelve (as of 2018, it has twenty-two), to collaborate in nuclear research and share the cost of building and maintaing the research facilities. Over the years, as the center-of-mass energy available increased, the main focus of the CERN experiment shifted to particle physics. It has been responsible for many important achievements in particle physics, such as the discovery of the W and the Z bosons in the UA1 and UA2 experiments [14, 15] (1983), the determination of the number of neutrino families at the Large Electron–Positron Collider [16] (1989), the invention of the World Wide Web [17] (1989) and the observation of a new particle consistent with the SM Higgs Boson [18, 19] (2012).

3.2 Large Hadron Collider

The Large Hadron Collider (LHC) [20] is a two-ring-superconducting-hadron accelerator and collider installed in the existing 26.7 km tunnel that was constructed for the CERN LEP (Large Electron-Positron Collider) machine. It can either accelerate protons or heavy ions and the collisions it creates are used by four large and several small experiments, with various goals.

For protons, the LHC was designed for a center-of-mass energy of up to $\sqrt{s} = 14$ TeV. For the years comprising Run I (2011-2012), the LHC operated with $\sqrt{s} = 7$ TeV (for 2011) and $\sqrt{s} = 8$ TeV (for 2012). At the beginning of 2013, the LHC was shutdown to undergo maintenance work, restarting in 2015, starting Run II, with $\sqrt{s} = 13$ TeV.

To get to such high energies, the injection and acceleration chain, represented in Figure 3.1, is used. This chain starts with a Linear Accelerator (LINAC2), and then continues with the BOOSTER, Proton Synchrotron and Super Proton Synchrotron, before being injected into the LHC itself. As time passes (usually in the time scale of hours), the beam quality and luminosity deteriorates, until they reach a point where they are not useful and so have to be discarted and the acceleration process started anew.

As mentioned before, LHC supplies four large experiments, which are located in four different straight sections of the accelerator. These experiments are: ATLAS (A Toroidal LHC ApparatuS), CMS (Compact Muon Solenoid), ALICE (A Large Ion Collider Experiment) and LHCb (A Large Hadron Collider beauty Experiment). The first two are general purpose experiments while the last two had specific goals when created, specifically, to analyse heavy ions collisions, for ALICE, and study properties of the b-quark, for LHCb.

The design value of the beams' instantaneous luminosity is $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ for proton-proton collisions and, at this value, each beam has 2808 bunches (each with 1.15×10^{11} protons), spaced 25 ns apart. The luminosity of the beams is given by

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \varepsilon_n \beta^*} F, \tag{3.1}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency of the beam and γ_r is the relativistic gamma-factor. The parameters ε_n and β^* define the shape of the beam, with ε_n being the normalised emittance (which is the average spread of the particles in position-momentum space, normalised to beam energy) and β^* being the optical beta-function at the interaction point (the distance from the focus point of the beam to the point at which the beam width is twice as large). The crossing angle of the beams is included in F, the geometric luminosity reduction, defined as

$$F = \left[1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right]^{-\frac{1}{2}}. \tag{3.2}$$

Here, θ_c is the beam crossing angle, and σ_z and $\sigma*$ are the root-mean-squared (RMS) length and transverse RMS beam size at the interaction point. Then, to achieve such a high

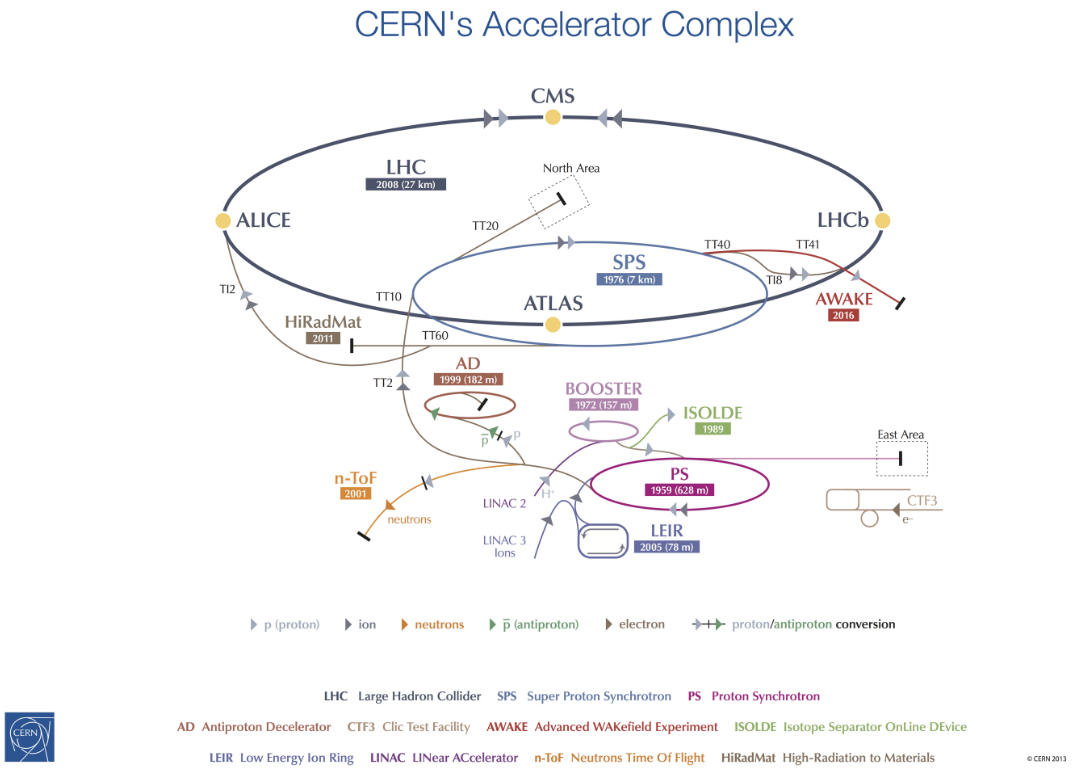


Fig. 3.1 Schematic representation of the CERN Accelerator Complex. Taken from [21].

luminosity, the LHC must have a high number of bunches and protons per bunch, as well as small ε_n and β^* – that is, highly squeezed bunches, with a small crossing angle at the interaction point. This is done by applying a strong magnetic field, which is produced by a system of 1232 superconducting dipole magnets. These bend the beams and keep them on course. Additionally, 392 main quadrupole magnets keep the beams focused and higher-order magnetic multipoles further correct the beam trajectory (see Figure 3.2). These magnets operate at an average temperature of 1.9 K, achieved with liquid helium constantly circulating around the main pipe.

The number of events (N) for a given process in a certain period of time is given by

$$N = \sigma \int \mathcal{L} \mathrm{d}t = \sigma L \tag{3.3}$$

where σ is the cross-section of the process and L is the integrated luminosity for that period. This last quantity is usually given in fb^{-1}, which is equivalent to $10^{-39}\,\mathrm{cm}^{-2}$.

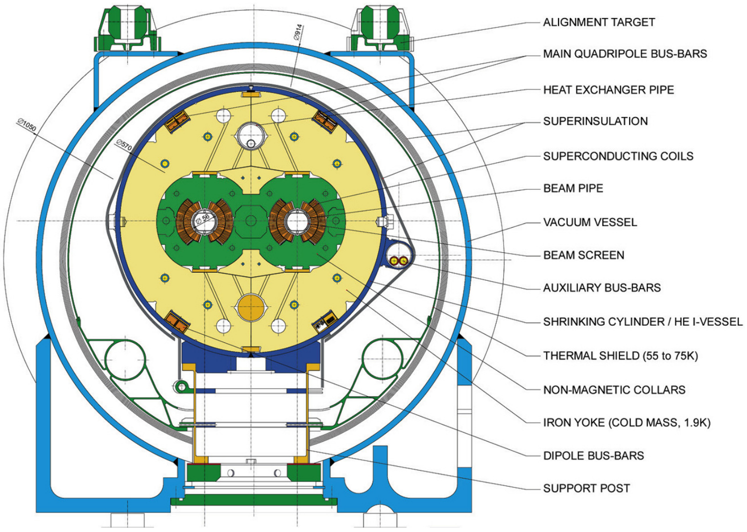


Fig. 3.2 Cross-section of the superconducting dipole magnet. Adapted from [22].

3.3 ATLAS detector

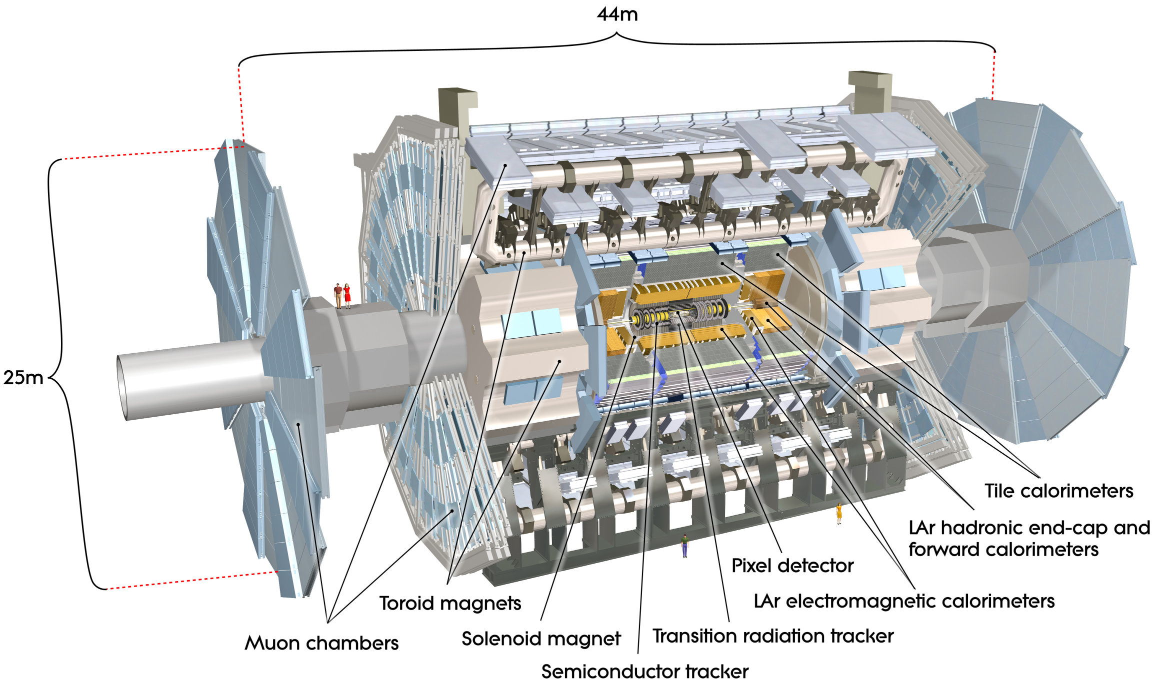


Fig. 3.3 Illustration of the ATLAS detector. Taken from [23].

The ATLAS detector [24], depicted in Figure 3.3, is a general purpose, hermetic detector that records the results of LHC's proton-proton collisions. It provides nearly full azimuthal coverage and is designed to reconstruct the properties of the particles produced during collisions. In total, it weights 7000 tonnes and is 44 m long with a diameter of 25 m.

ATLAS is composed of four major parts: the Inner Detector (ID), which tracks electrically charged particles and measures their momenta; the Muon Spectrometer identifies the muons and measures their kinematics; the Magnet System, which includes a solenoidal magnet outside of the ID and the toroid magnet located in the muon system, bends the charged particle trajectories; the Calorimeter System, which is used to measure the energy deposits of particles crossing the detector. By gathering information from the different parts of the detector, it is possible to reconstruct the event at the moment of collision. The signatures left in the detector by different types of particles are represented in Figure 3.6.

During the first three years of the Run II phase of the operations, the ATLAS detector recorded more than 80 fb^{-1} of data. Figure 3.4 shows the cumulative luminosity delivered by the accelerator, in green, recorded by ATLAS (yellow) and good for physics (blue) as a function of time, during stable beams for pp collisions at centre-of-mass energy of 13 TeV in 2015-2017.

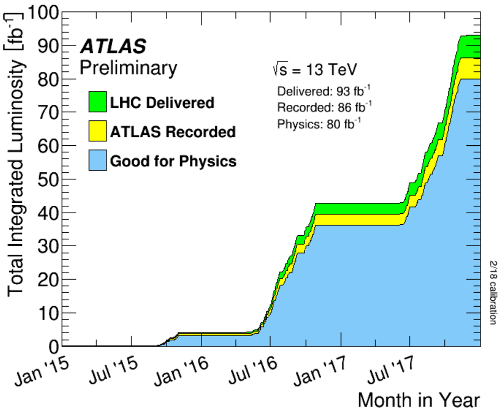


Fig. 3.4 Cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015-2017. Taken from [25].

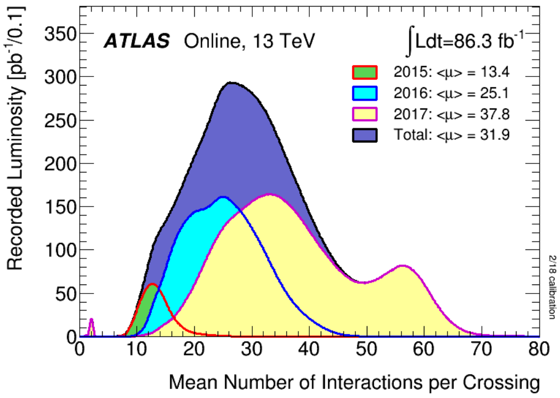


Fig. 3.5 Luminosity-weighted distribution of the mean number of interactions per crossing for the 2015-2017 pp collision data at 13 TeV centre-of-mass energy. Taken from [26].

3.3.1 Coordinate system

The coordinate system used in ATLAS is adapted to the detector's symmetries. It is a modified cylindrical coordinate system, with the z axis going along the beam. The y axis points upwards and the x axis into the center of the accelerator ring. The azimuth, ϕ, is chosen such that its null value corresponds to the positive x axis. Instead of using the polar angle, θ, the pseudo-rapidity, η, is chosen, defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{3.4}$$

For massless particles, this is identical to the rapidity $y = 1/2\ln\left[(E+p_z)(E-p_z)\right]$. Strictly speaking, this definition is only valid if $p_x, p_y \ll p_z$, which happens to be the case in most high-energy physics experiments. Differences in y are invariant under Lorentz boosts along the z-axis. For particles with negligible mass, such as electrons and muons, the same applies for the pseudo-rapidity η. Distances in the (η,ϕ) space are calculated as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

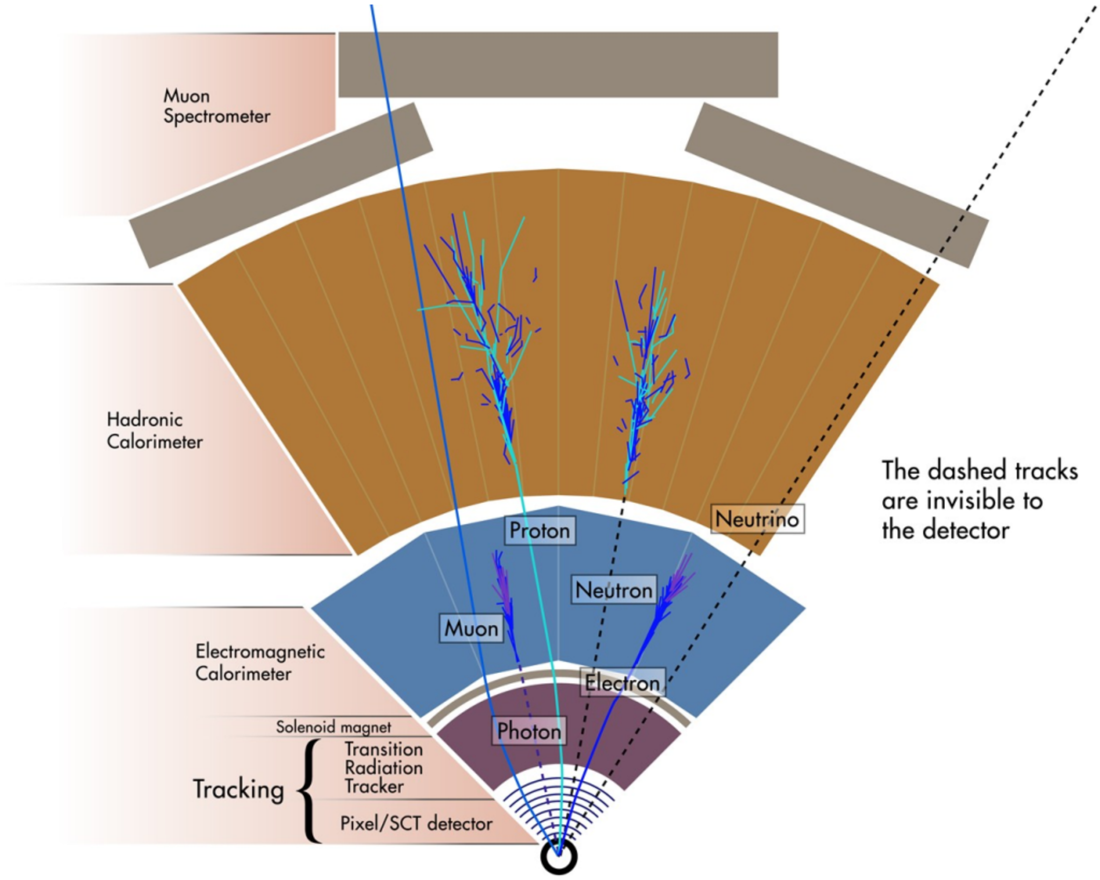


Fig. 3.6 Particle interactions with different layers of the ATLAS detector. Adapted from [27].

3.3.2 Magnet System

The magnet system of ATLAS [28] is composed of four large superconducting magnets, represented in Figure 3.7: three toroidal magnets (one along the barrel and two for each end)

Table 3.1 Table of resolution and acceptance requirements for ATLAS sub-detectors. Taken from [24].

Detector component	Required resolution	η coverage	
		Measurement	Trigger
Tracking (ID)	$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T \oplus 1\%$	± 2.5	
EM calorimetry	$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic calorimetry (jets)			
barrel and endcap	$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$	± 3.2	± 3.2
forward	$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\frac{\sigma_{p_T}}{p_T} = 10\%$ at $p_T = 1$ TeV	± 2.7	± 2.4

and one central solenoid around the ID. It is 26 m long and has a diameter of 20 m. The entire system is cooled by liquid helium, at a temperature of 4.5 K.

As mentioned before, the magnet system is used to bend the trajectories of charged particles. It does this in two different areas of the detector. The solenoidal magnet applies a 2 T field inside the ID, which is directed along the beam axis. The thickness of this magnet is the least possible so that particles crossing to the calorimeters lose the minimum amount of energy. The toroidal magnets do the same to muons in the outer part of the detector. The barrel toroid creates a 0.5 T field, while the endcap toroids generate a 1 T one each.

3.3.3 Inner Detector

The Inner Detector (ID) [29, 30], represented in Figure 3.8, is used for precise measurement of the trajectories of the particles produced in the LHC collisions, within the range of $|\eta| < 2.5$. To do that, it measures the momenta of the particles (only for $p_T > 0.5$ GeV) as well as assigning them to their corresponding production vertex (the primary vertex). For b-hadrons (particles that contain the b-quark), it also assigns them to their secondary vertices. This is specially useful for b-tagging algorithms. In total, the ID is 6.2 m long and has a diameter of 2.1 m. It is divided into three parts, described in the following paragraphs. These parts have in total more than 40 tracking space points.

The first part is the Pixel Detector, which is composed by three cylindrical layers, the barrels, and three endcap layers, with three discs on each side. The Pixel Detector has 1744 pixel modules, of dimensions 6.3×19 mm^2, with 47232 pixel elements per module. This gives it approximately 80 million readout channels. With a spatial resolution $(R - \phi) \times z =$

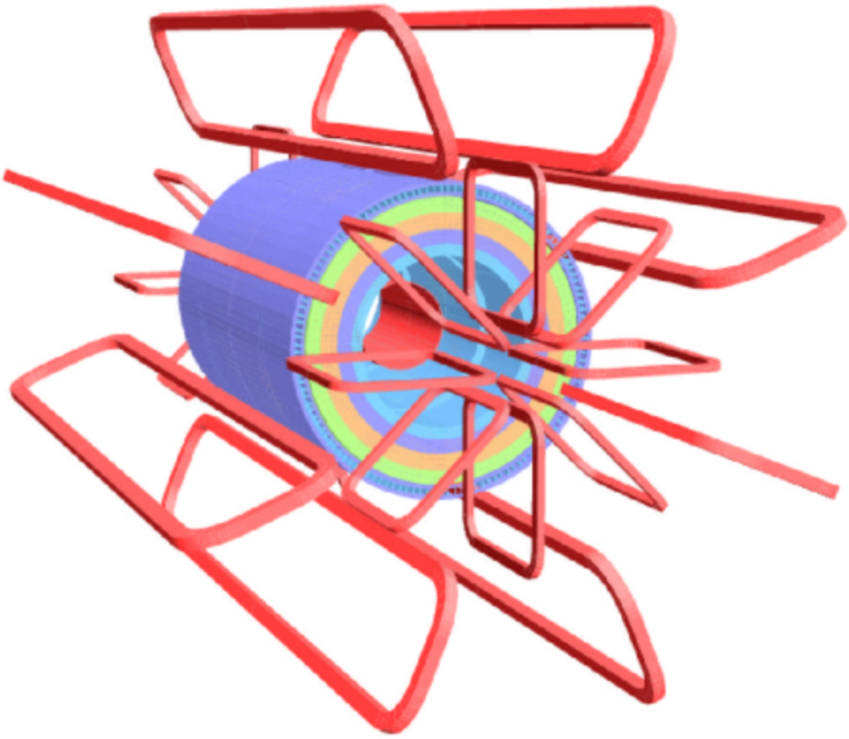


Fig. 3.7 Schematic representation the ATLAS magnet system. The toroidal magnets are in red and the solenoidal one is inside the cylindrical volume, representing the calorimeter. Taken from [24].

10×115 µm^2, its main aims are to determine the impact parameter resolution and find short-lived particles.

Next is the Semiconductor Tracker (SCT), with 4 barrels and 9 disks on each endcap, which has a total of 4088 modules of dimensions 6.36×6.40 cm^2. Each of them has four sensors (two on the bottom and two on the top) that are rotated by ± 20 rad to give two-dimensional information about each hit. The SCT has a total of around 6.3 million readout channels and a spatial resolution $(R-\phi) \times z = 17 \times 580$ µm^2. Its aims are to measure momenta, impact parameters and vertex positions.

The final part of the ID is the Transition Radiation Tracker (TRT), with a central barrel plus three endcaps on each side. The TRT has around 280K proportional drift tubes (straws) of 4 mm in diameter and 144 cm long. Each of these straws is filled with a gas that ionizes when a charged particles goes through it, producing a signal in a wire located at the straw's center. In-between the straws, there is a material that leads to the production of transition radiation for ultra-relativistic particles. This enables the TRT to distinguish electrons from pions, since

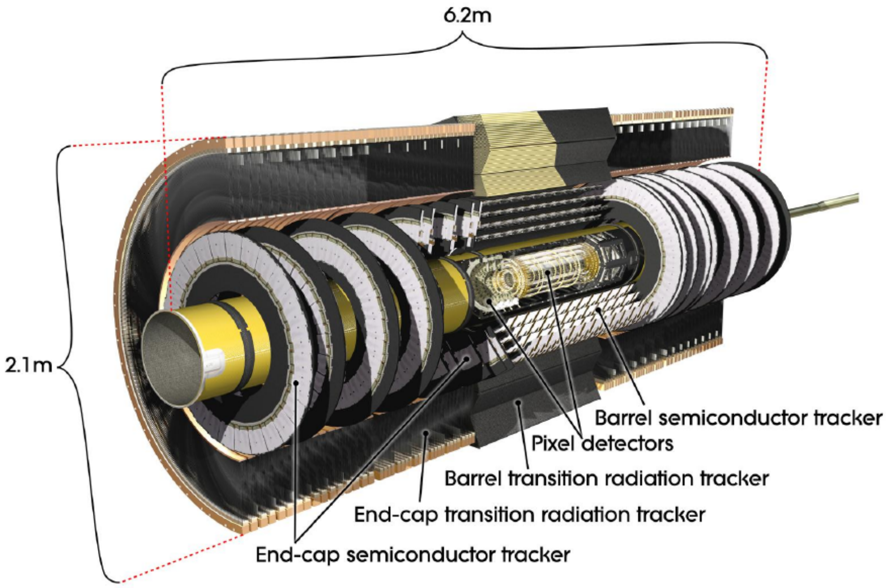


Fig. 3.8 Cut-away view of the ATLAS inner detector. Taken from [24].

the amount of transition radiation for particles with identical momentum decreases with their mass. The TRT has a total of around 350K readout channels.

3.3.4 Calorimeters

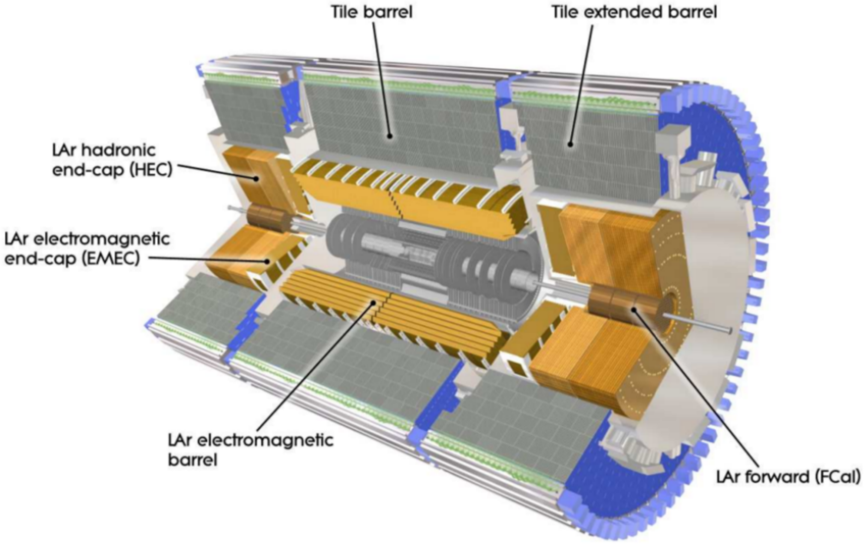


Fig. 3.9 Cut-away view of the ATLAS calorimetry system. Taken from [24].

The two calorimeters in ATLAS, shown in Figure 3.9, are used to measure the energy of charged particles (except muons, which are detected in the muon system) and neutral particles (except neutrinos, which are not detected at all). Besides this, it also shields the muon system from these particles, preventing the detection of fake muons. The two calorimeters are of two different types: one is electromagnetic, measuring the energy of electrons and photons, while the other is hadronic, which measures the energy of hadrons and mesons. Together, they cover 99.5% of a full sphere. The coverage of individual systems is given in Table 3.2.

Table 3.2 Summary of the different ATLAS calorimeter sub-systems. Taken from [24].

System	Coverage in $\lvert\eta\rvert$	Detector medium	Absorber medium
Electromagnetic Barrel	$0.0 < \lvert\eta\rvert < 1.475$	Liquid Argon	Lead
Electromagnetic Endcap	$1.375 < \lvert\eta\rvert < 3.2$	Liquid Argon	Lead
Hadronic Tile Barrel	$0.0 < \lvert\eta\rvert < 1.0$	Plastic	Steel
Hadronic Tile Extended Barrel	$0.8 < \lvert\eta\rvert < 1.7$	Plastic	Steel
Hadronic Endcap	$1.5 < \lvert\eta\rvert < 3.2$	Liquid Argon	Copper
Forward Calorimeter	$3.1 < \lvert\eta\rvert < 4.9$	Liquid Argon	Copper & Tungsten

The first of these two parts, the electromagnetic calorimeter [31–33], is a sampling calorimeter made of accordion-shaped lead absorbers, 1.1 mm to 2.2 mm thick. These are immersed in liquid argon (LAr) at a temperature of 82 K. This calorimeter is composed of three parts: one barrel, covering $|\eta| < 1.475$ and two endcaps, covering $1.375 < |\eta| < 3.2$. The first of these is in turn divided in two cylindrical halves, with a small "crack" between, where no particles are detected. These halves are 3.2 m long and 53 cm thick, and each is divided into 16 modules, with 3424 readout channels per module. The way the electromagnetic calorimeter measures the energy of electrons and photons is the following. First, these particles interact with the lead layers of the calorimeter, radiating and creating a shower of photons, electrons and positrons. These, in turn, ionize de atoms of argon that surround the lead absorbers, creating pairs of electrons and positive ions. By applying an electrical field, the electrons drift to electrodes located in gaps that are between the absorbers. This gives rise to currents that are proportional to the energy the particles lose. Measuring the charge deposited by these currents, it is then possible to obtain the original energy of the incident particles.

The second part, the hadronic calorimeter [31, 34, 35] is in turn divided into three components: the tile calorimeter (TileCal), the LAr hadronic endcap calorimeter (HEC) and the LAr forward calorimeter (FCal). The first of these surrounds the Electromagnetic Calorimeter, with one long central barrel (covering $|\eta| < 1.0$) and two extended barrels (covering $0.8 < |\eta| < 1.7$). These are divided into 64 modules in the azimuthal direction.

The TileCal is also a sampling calorimeter, but uses alternating layers of steel (which acts as the absorber) and scintillating plastic tiles (which act as the active material). It has a total of 5184 readout cells, with 9856 channels. When a high-energy hadron reaches TileCal, it interacts with the nuclei in the steel, producing hadronic showers, which may then enter the scintillators. This creates ultraviolet light with an intensity that is proportional to the incident energy, which then goes through wavelength shifting fibres, converting it to visible light. This in turn is captured by the PMTs, sensible only to this part of the spectrum. By looking at the the light intensity given by the PMTs, it is possible to obtain the initial energy.

The HEC and the FCal cover the regions of $1.5 < |\eta| < 3.2$ and $3.1 < |\eta| < 4.9$, respectively. They are both similar to the Electromagnetic Calorimeter, but HEC uses copper as an absorber, instead of lead, while the FCal is divided into three parts. One of them is electromagnetic and uses copper as the absorbing material, and the other two are hadronic, using instead tungsten. They both share the same cryostat as the Electromagnetic Calorimeter LAr endcaps.

3.3.5 Muon system

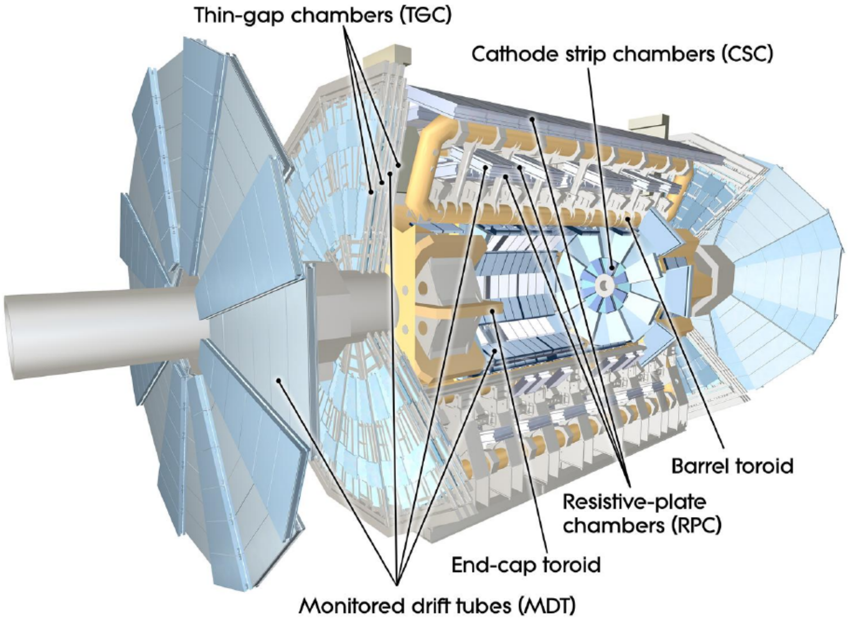


Fig. 3.10 Cut-away view of the ATLAS muon system. Taken from [24].

The ATLAS Muon System [36, 37], shown in Figure 3.10, is the detector's outer shell shielded by the inner parts from most of the particles. As its name suggests, it is used to detect muons and measure their momenta with high precision. To achieve that, it is helped by the magnet system, that deflects the tracks of the particles. The principle is similar to what happens in the ID. In the range $|\eta| < 1.4$, the magnetic field from the large barrel toroid is used to bend the tracks; in $1.6 < |\eta| < 2.7$, that is done by the two endcap magnets; in the range inbetween, a combination of the three is used.

There are three layers in the Muon System, three barrels (located at radii 5 m, 7.5 m and 10 m) and three endcaps on each side (at ±7.4 m, ±14 m and ±21.5 m). It is made up of four components, where two are responsible for the precise measurement of the momenta of the muons, the Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSG), and the other two for the triggering (signalling a few ns after a particle passes through them), the Resistive Plate Chambers (RPC) for the barrel and the Thin Gap Chambers (TGC) for the endcaps.

Covering most of the detector, the 1171 MDT chambers (giving around 355 thousand drift tubes) work in the following way. As the muon passes in one of the tubes, the gas in it is (a mixture of argon and CO_2) ionized. An electrical field (due to a potential difference of 9.8 kV) makes the electrons from the ionization drift towards the wire located at the center of the chamber. With the kinetic energy gained from this electrical field, the electrons then free more of their kin. The RPC's and the TGC's measure the time taken by the electrons to reach the wire, allowing the determination of the distance between the particle and the wire. Connecting the drift radii from each MDT hit by a tangent line to each drift circle, it is possible to reconstruct the track segment.

3.3.6 Trigger and data acquisition system

At the LHC nominal bunch crossing, approximately one billion collisions per second are produced (40 MHz × around 25 interactions per collision). Since the readout capacity of ATLAS is only 1 kHz, some mechanism must be implemented to filter the collisions produced. This is called the Trigger System [38–41] and it was originally divided into three parts: Level-1 (L1), Level-2 (L2) and event filter (EF). For Run II, the last two were merged into a single event processing High Level Trigger (HLT) farm. These systems are represented in Figure 3.11.

The L1 trigger is able to reduce the event rate to about 100 kHz. It does this by using coarse detector representation to select events with interesting features, such as muons/electrons/photons with high transverse momenta, τ-leptons decaying to hadrons or jets, or with large missing transverse energy. This is done with a maximum decision time of 2.5 μs after the bunch crossing. The L1 trigger identifies regions of interest in (η, ϕ), containing selected features that get passed to higher level triggers. The HLT then uses these regions of interests, full data information and sophisticated algorithms to reduce the event rate to 1 kHz, with a processing time of about 200 ms.

The remaining events then get transmitted to the CERN computer centre for storage. This amounts to around 4 petabytes per year. To be distributed efficiently across the world to the thousand of physicists that are a part of CERN, these resources were integrated into the Worldwide LHC Computing Grid [42]. This is divided into four tiers.

Tier 0 is CERN's data centres, located in Geneva, Switzerland and Budapest, Hungary. These two sites are connected by 100 Gbit/s data links. This tier passes the raw and reconstructed data to the 13 Tier 1 centers across the world using optical-fiber links working at 10 Gbit/s. These then share with the more than 150 Tier 2 centres, located at universities and scientific institutes, which contribute with further processing and storage of analysis-specific data as well as production and reconstruction of simulated events. Finally, the Tier 3 sites are where individual scientists can access and analyse the data.

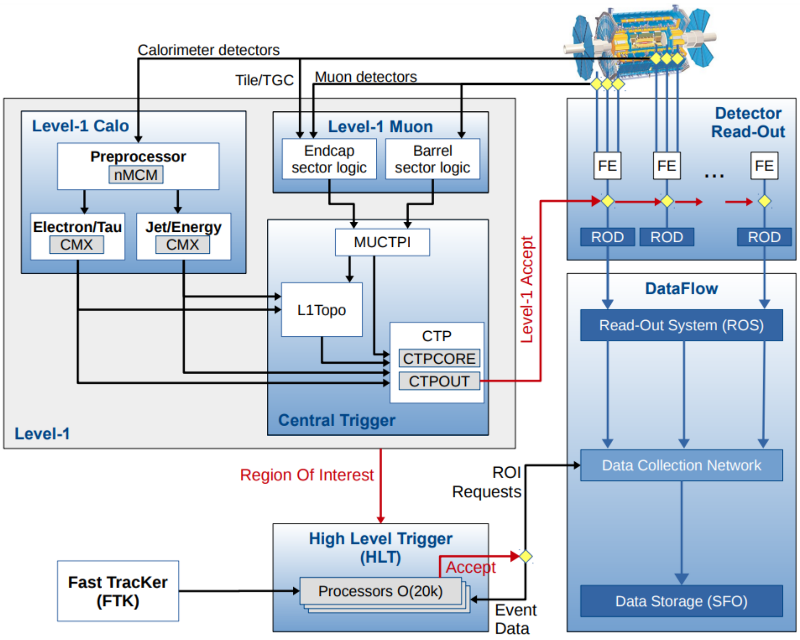


Fig. 3.11 Schematic layout of the ATLAS Trigger and Data Acquisition System in Run II. Taken from [41].

Chapter 4

Measurement of top quark decay to sW

4.1 Data and simulated samples

4.1.1 Data samples

This analysis uses data from pp collisions recorded, in the 2015-2016 run, with the ATLAS detector, with a center-of-mass energy of $\sqrt{s} = 13$ TeV. The dataset contains only events in which the detector systems and triggers were operating correctly, passing the quality requirements of the ATLAS Good for Physics list. The integrated luminosity of the dataset is $36.2\,\mathrm{fb}^{-1}$, which was then scaled to $150\,\mathrm{fb}^{-1}$, the expected integrated luminosity of the full Run II, from 2015 to 2018.

4.1.2 Event simulation

Simulated samples, obtained using Monte Carlo (MC) statistical techniques, are used to compare data with the theory predictions. Since the typical pp collision is extremely complex, involving the hard process of the collision, initial and final state radiation, hadronisation, as well as decays of unstable particles, the collision events cannot be computed completely from theory, requiring modelling from phenomelogical approaches. Having simulated all of this, the samples are in the what is called truth-level. This information is then sent to the detector simulator, where the passage of the particles through the detector is simulated. The samples are then in the reconstruction-level, which can be used in analyses.

The simulated samples used in this analysis are shown in Table 4.1, along with information about the generators, parton shower and parton distribution functions (PDFs) used to simulate them. The generation of the $t\bar{t}$ samples were done with `Powheg-Box v2` [43], interfaced to `Pythia 8` [44] with the `A14` [45] set of tuned parameters and the `NNPDF3.0NLO` [46] and

Table 4.1 Generators, parton shower simulation, parton distribution functions and tune parameters used to produce simulated samples for this analysis. The acronyms ME and PS stand for matrix element and parton shower, respectively. References are given in the text.

Sample	Generator	Parton Shower	ME PDF	PS PDF	Tune parameters
$t\bar{t} \to sWbW$	`Powheg-Box v2`	`Pythia8`	`NNPDF3.0NLO`	`NNPDF2.3LO`	`A14`
$t\bar{t} \to bWbW$	`Powheg-Box v2`	`Pythia8`	`NNPDF3.0NLO`	`NNPDF2.3LO`	`A14`
Single Top Wt-channel	`Powheg-Box v2`	`Pythia6`	`CT10`	`CT10`	`Perugia2012`
Single Top t-channel	`MadGraph`	`Pythia8`	`NNPDF3.0NLO`	`NNPDF2.3LO`	`A14`
Z+jets	`Sherpa 2.2.1`	`Sherpa 2.2.1`	`NNPDF3.0NNLO`	`NNPDF3.0NNLO`	Sherpa default
ZZ	`Sherpa 2.2.1`	`Sherpa 2.2.1`	`NNPDF3.0NNLO`	`NNPDF3.0NNLO`	Sherpa default
WZ	`Sherpa 2.2.1`	`Sherpa 2.2.1`	`NNPDF3.0NNLO`	`NNPDF3.0NNLO`	Sherpa default
WW	`Sherpa 2.2.1`	`Sherpa 2.2.1`	`NNPDF3.0NNLO`	`NNPDF3.0NNLO`	Sherpa default

`NNPDF2.3LO` [47] sets. The single top Wt-channel sample was generated using `Powheg-Box v2` interfaced to `Pythia 6` [48], with the `Perugia2012` tuned parameters [49] and `CT10` set [50], while the t-channel samples differs from the $t\bar{t}$ in the generator used, `MadGraph` [51]. The other samples, Z+jets and diboson, were generated at LO using `Sherpa 2.2.1` [52], with the `NNPDF3.0NNLO` sets. At the time of writing, only truth-level samples for the signal $t\bar{t} \to sWbW$ were available, so those were used.

4.2 Object reconstruction

In this section, the definition of each object present in this analysis is given. This is based on their track signatures left on the inner detector (for charged particles) as well as the energy they deposit in the calorimeter. For this analysis, the primary objects are electrons, muons, jets (with the special case of the b-tagged jets) and missing transverse momentum.

4.2.1 Electrons

The electron reconstruction and identification [53] is done using the energy deposits in the electromagnetic calorimeter that are matched with tracks from the ID. The electrons must have $p_T > 15$ GeV and the pseudorapidity of the cluster (energy deposit) associated with the electron candidate must be $|\eta_{\text{cluster}}| < 2.47$. The candidates in the transition region between the central and end cap regions, with $1.37 < |\eta| < 1.52$, are excluded due to poor energy resolution.

To reduce the background from non-prompt electrons, several criteria are applied. These include $|d_0|/\sigma(d_0) < 5$, where d_0 is the transverse impact parameter measured with respect to the beam axis and $\sigma(d_0)$ is its uncertainty; and $|z_0 \sin\theta| < 0.5$ mm, where z_0 is the distance from the transverse impact parameter point along the beam line to the primary vertex. Additionally, criteria associated with the energy sum of the clusters within a cone of size $\Delta R = 0.2$ around the electron candidate and associated with the p_T sum of the tracks within $\Delta R = 0.3$, giving a 90% isolation efficiency, are applied [54].

4.2.2 Muons

Muons are reconstructed and identified [55] using the tracks from the ID and the MS. They must have $p_T > 15$ GeV and $|\eta| < 2.5$, as well as, $|d_0|/\sigma(d_0) < 5$ and $|z_0 \sin\theta| < 0.5$ mm. Also, the p_T sum of the tracks around the muon candidate within a cone of $\Delta R = \min(10\ \text{GeV}/p_T^{\mu}, 0.3)$ must be less than 5% of the p_T of the muon candidate.

4.2.3 Jets

Jets are reconstructed from topological clusters in the calorimeters. These are noise-supressed and calibrated to the electromagnetic scale [56] using the anti-k_t algorithm [57] with radius parameter $R = 0.4$. Firstly, the average additional energy due to pile-up is subtracted from the energy measured in the calorimeters. Then, the direction of the jet is also corrected, so that it originates from the main primary vertex of the interaction and not from the geometric center of the detector. Finally, finer corrections in η and p_T are applied.

To reduce the number of selected jets that originate from pile-up, the Jet-Vertex Fraction (JVF) technique is used. With it, the fraction of the sum of the p_T of the jets that originate from the primary vertex must be at least 50% of the p_T sum of all the jets.

Jets containing b-hadrons are identified (b-tagged) [58] using an algorithm based on multivariate techniques. It uses the impact parameters of displaced tracks (that are displaced from the primary vertix, due to the long lifetime of b-hadrons) as well as topological properties of secondary and tertiary vertices reconstructed within the jet. Using $t\bar{t}$ events, the tagging efficiency is determined to be 77%. The rejection factors (the inverse of the fraction of jets that are mis-identified) for light flavours and charm jets are 133 and 6.2, respectively.

4.2.4 Missing Transverse Momentum

The components of the missing transverse momentum E_T^{miss} are defined as

$$E_{x(y)}^{\text{miss}} = -\left(\sum_{\text{selected}} p_{x(y)} + \sum_{\text{soft}} p_{x(y)} \right), \tag{4.1}$$

where the first term corresponds to the momentum of all selected and calibrated objects in the event and the second term to the energy from soft particles in the event that is not associated to any of the selected particles [59, 60]. This last term includes only tracks that are associated with the primary vertex, to prevent contamination from pile-up events.

The magnitude of the missing transverse momentum is then

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}. \tag{4.2}$$

4.3 Event Selection

To reliably measure $|V_ts|^2$, the backgrounds should be suppressed as far as possible, while maximizing the signal significance $s/\sqrt{s+b}$, where s and b are the number of signal and background events, respectively.

This is done by defining a signal region (SR), with several steps. In every step, a certain cut is performed that increases the signal significance. This is done to both data and simulated samples. The defined SR has three phases:

- The preselection chooses the events based on basic requirements, which include lepton multiplicity and kinematic criteria, as well as number of jets. For details see subsection 4.3.1.
- A kinematic reconstruction is performed assuming several hypotheses: that the event is signal or one of the main backgrounds. See subsection 4.3.2 for details.
- Finally, the remaining events are used to train a neural network, to optimally separate signal and background (for an overview of neural networks, see Appendix A). The details are described in subsection 4.3.3.

4.3.1 Event Preselection

The cuts in the preselection phase are based on the ATLAS analysis on the measurement of the $t\bar{t}$ production cross-section with $e\mu$ final state [61]. Requiring exactly one electron and one muon of opposite charges significantly reduces the backgrounds from the decays $Z/\gamma^* \to e^+e^-/\mu^+\mu^-$. These leptons must have the following properties:

- $p_T > 25\,\text{GeV}$;
- $\Delta\eta(e,\mu) > 0.15$;
- $\Delta\phi(e,\mu) > 0.15$ and

- The invariant mass of the two lepton system must be greater than 40 GeV.

Additionally, the following jet cuts are also applied:

- At least two jets;
- At least one non-b-tagged jet and
- At most two b-tagged jets.

The jet cuts are motivated by the $t\bar{t} \to sWbW$ final state topology. The histograms representing the number of b-tagged jets before and after the preselection are shown in Figure 4.1. As expected, most Z+jets events are supressed. Due to the preliminary nature of this study, no data points are shown in the SR, so that the strategy of the analysis is not tailored to them.

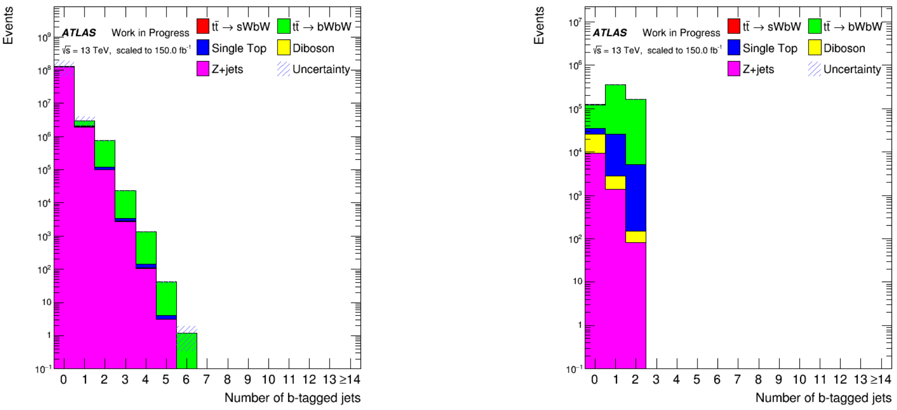


Fig. 4.1 Number of b-tagged jets before (left) and after (right) the preselection, in the SR.

4.3.2 Kinematic Reconstruction

To further improve the signal efficiency, several kinematic reconstructions are performed, assuming different kinds of event topology. In one these hypotheses, $t\bar{t}$ decaying to $qWqW$, to obtain the kinematic properties, the following expression is minimised without constraints:

$$\chi^2_{t\bar{t}} = \frac{\left(m^{\mathrm{reco}}_{j_a\ell_a\nu_a} - m_{t_1}\right)^2}{\sigma^2_{t_1}} + \frac{\left(m^{\mathrm{reco}}_{j_b\ell_b\nu_b} - m_{t_2}\right)^2}{\sigma^2_{t_2}} + \frac{\left(m^{\mathrm{reco}}_{\ell_a\nu_a} - m_W\right)^2}{\sigma^2_W} + \frac{\left(m^{\mathrm{reco}}_{\ell_b\nu_b} - m_W\right)^2}{\sigma^2_W}. \quad (4.3)$$

In this case, $m^{\text{reco}}_{j_a\ell_a\nu_a}$ is the reconstructed mass of a qW system, while $m^{\text{reco}}_{j_a\ell_a\nu_a}$ is the reconstructed mass of the other qW system. Finally, $m^{\text{reco}}_{\ell_a\nu_a}$ and $m^{\text{reco}}_{\ell_b\nu_b}$ are the reconstructed masses of the two $\ell\nu$ systems. To get the minimum χ^2, every jet combination is tested. The leptons are also switched. Since the two neutrinos from the decays of each top quark are undetected, their four-momentum must be estimated. This means that the minimisation of χ^2 corresponds to obtaining the most probable values of $p_z^{\nu_a}$, $p_x^{\nu_a}$, $p_y^{\nu_a}$ and $p_z^{\nu_b}$. The other two components of the momentum of ν_b can be obtained by assuming that the missing transverse energy has only contributions from these two neutrinos and then calculating

$$p_x^{\nu_b} = E_x^{\text{miss}} - p_x^{\nu_a} \quad \text{and} \tag{4.4}$$

$$p_y^{\nu_b} = E_y^{\text{miss}} - p_y^{\nu_a} \tag{4.5}$$

The central value for the masses and the widths of the top quarks and W bosons were obtained from [62]. These values are $m_{t_1} = m_{t_2} = 167.2\,\text{GeV}$, $\sigma_{t_1} = \sigma_{t_2} = 24.0\,\text{GeV}$, $m_W = 81.2\,\text{GeV}$ and $\sigma_W = 15.1\,\text{GeV}$.

Other main background is the single-top events (Wt), where the top decays to a quark and a W boson. Then, another hypothesis for the kinematic reconstruction assumes these kinds of events, with every jet tested. The expression that is minimised in this case is

$$\chi^2_{Wt} = \frac{\left(m^{\text{reco}}_{j_a\ell_a\nu_a} - m_{t_1}\right)^2}{\sigma^2_{t_{bW}}} + \frac{\left(m^{\text{reco}}_{\ell_a\nu_a} - m_W\right)^2}{\sigma^2_W} + \frac{\left(m^{\text{reco}}_{\ell_b\nu_b} - m_W\right)^2}{\sigma^2_W}, \tag{4.6}$$

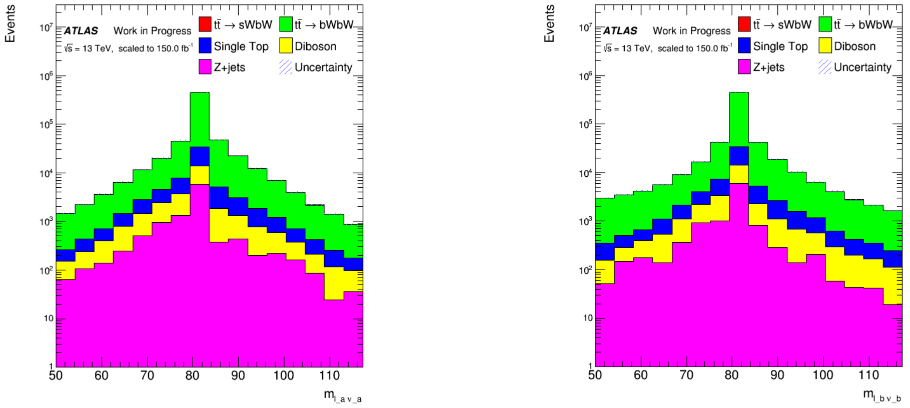
where the central values and widths are the same as the ones previously given.

Finally, while the preselection cut that requires one electron and one muon of opposite charges reduces drastically the contributions of $Z \to ee/\mu\mu$, another leptonic decay of the Z boson, $Z \to \tau\tau$, still contaminates the SR. This happens with the decays $\tau_1 \to e\nu\nu$ and $\tau_2 \to \mu\nu\nu$. The final hypothesis tries to reconstruct these kinds of events, by minimising the following expression

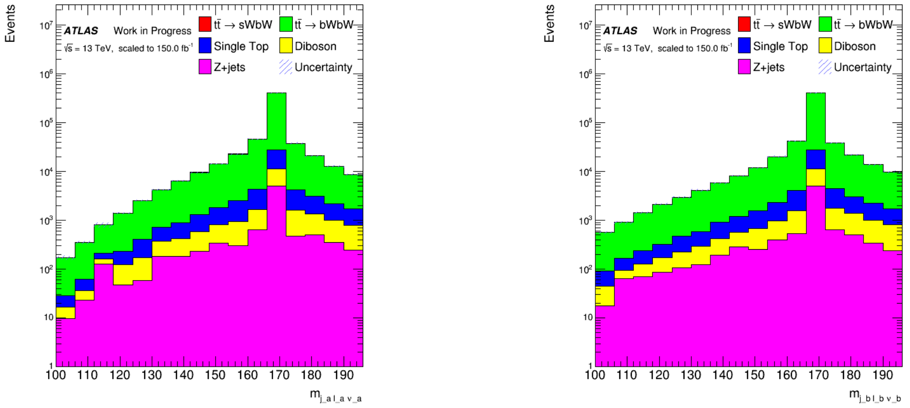
$$\chi^2_Z = \frac{\left(m^{\text{reco}}_{\ell_a\ell_b\nu_a^1\nu_a^2\nu_b^1\nu_b^2} - m_Z\right)^2}{\sigma^2_Z} \tag{4.7}$$

where $m^{\text{reco}}_{\ell_a\ell_b\nu_a^1\nu_a^2\nu_b^1\nu_b^2}$ is the reconstructed mass of the $\ell\ell\nu\nu\nu\nu$ system, and $m_Z = 91.1876\,\text{GeV}$ and $\sigma_Z = 3\,\text{GeV}$. These values were obtained from [62]. In this case, despite the reconstructed system having four neutrinos, only one particle is being reconstructed, the Z boson. Then, by assuming that the missing energy corresponds only to these four neutrinos, the minimisation of Equation 4.7 is equivalent to finding the most probable value of the z component of the missing energy.

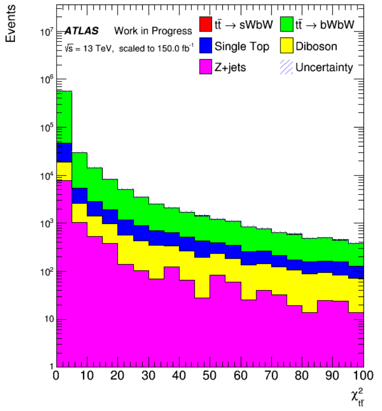
The reconstructed masses and χ^2 for all the hypotheses, in the SR, are represented in Figures 4.2 to 4.4. To further increase the signal efficiency, one additional cut was introduced, related to these kinematic reconstructions. To further decrease the contribution of the Z+jets events, it was required that $m_{\ell_a \ell_b \nu_a^1 \nu_a^2 \nu_b^1 \nu_b^2}$ be less than 80 GeV or greater than 100 GeV. The number of b-tagged jets after this cut is shown in Figure 4.5. The other hypotheses, while not giving a clear differentiation between the signal and background samples, were used as variables in the training of the neural network (see subsection 4.3.3).



(a) $m^{\text{reco}}_{\ell_a \nu_a}$ (b) $m^{\text{reco}}_{\ell_b \nu_b}$

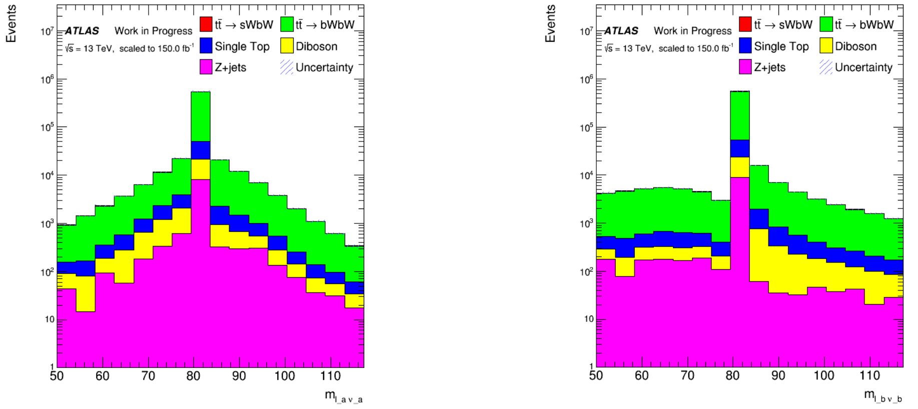


(c) $m^{\text{reco}}_{j_a \ell_a \nu_a}$ (d) $m^{\text{reco}}_{j_b \ell_b \nu_b}$

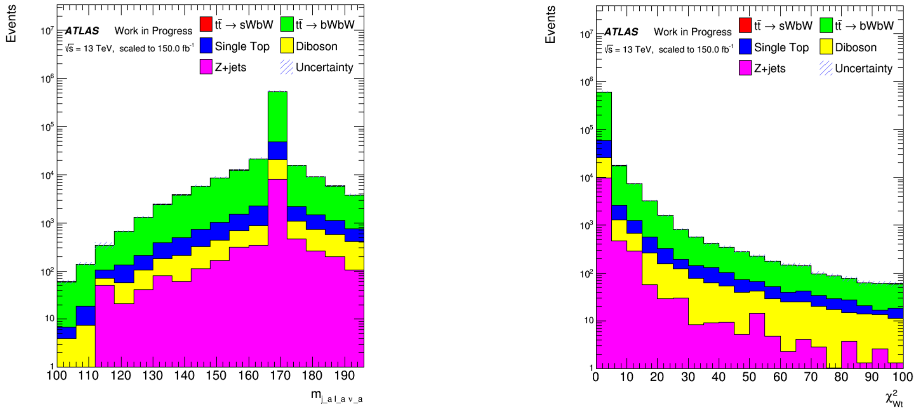


(e) $\chi^2_{t\bar{t}}$

Fig. 4.2 Reconstructed masses and χ^2 for the $t\bar{t}$ hypothesis, in the SR.

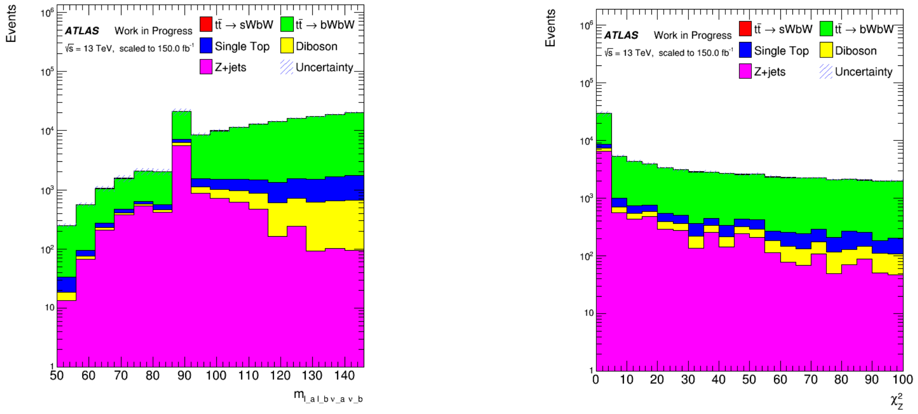


(a) $m^{\text{reco}}_{\ell_a \nu_a}$ (b) $m^{\text{reco}}_{\ell_b \nu_b}$



(c) $m^{\text{reco}}_{j_a \ell_a \nu_a}$ (d) χ^2_{Wt}

Fig. 4.3 Reconstructed masses and χ^2 for the Wt hypothesis, in the SR.



(a) $m^{\text{reco}}_{\ell_a \ell_b \nu_a^1 \nu_a^2 \nu_b^1 \nu_b^2}$ (b) χ^2_Z

Fig. 4.4 Reconstructed masses and χ^2 for the $Z \rightarrow \tau\tau$ hypothesis, in the SR.

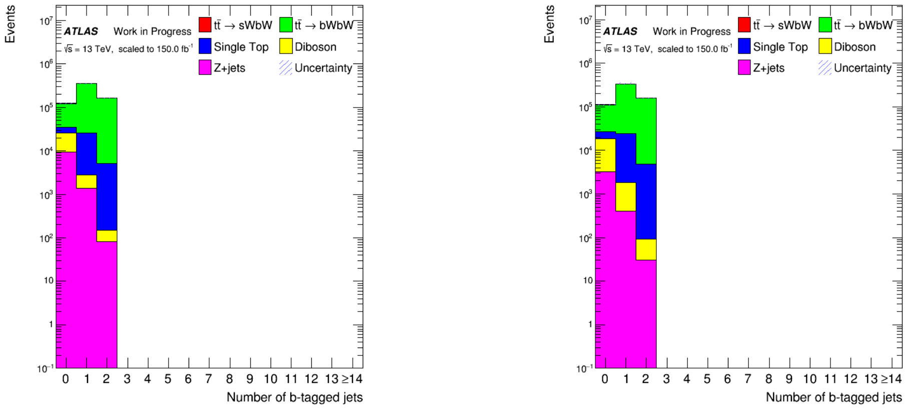


Fig. 4.5 Number of *b*-tagged jets after preselection (left) and after kinematic reconstruction cuts (right), in the SR.

4.3.3 Neural Network Results

While the previous cuts decreased the contamination of some backgrounds significantly, one background remains dominant in the signal region: $t\bar{t} \to bWbW$. This is due to the fact that its topology is very similar to the signal and to the much higher probability for the top quark to decay to a b-quark. To further increase the signal significance, a more sophisticated approach must be used, namely, a multivariate one. The one used for this analysis was a neural network. Besides distinguishing between signal and $t\bar{t} \to bWbW$, it also distinguishes it from other backgrounds.

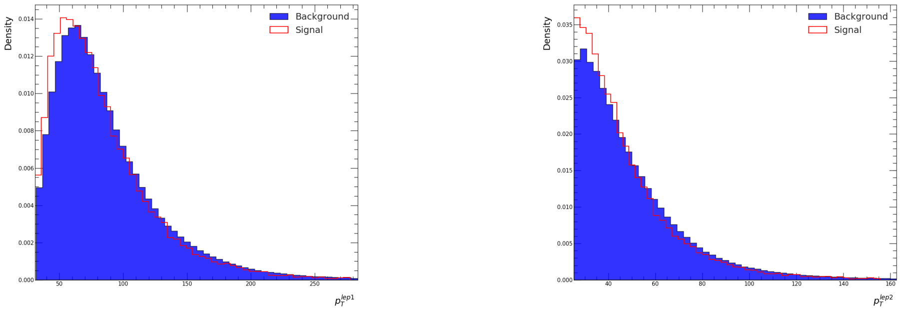
To train the neural network, features had to be selected that allowed it to distinguish the signal from the backgrounds. The features chosen are represented in Figures 4.6 and 4.7. A list of them is presented in Table 4.2. Other possible features are those related to the angular coordinates of the leptons. However, since the signal sample is at truth-level, it has events where its leptons have η coordinate corresponding to the "crack" of the detector (see subsection 3.3.3), while the other samples do not. This would allow the neural network to be unrealistically able to distinguish the signal from the backgrounds.

Using the `Keras` framework, the network architecture is defined, with several models being tested. The one with the best results had four layers. The first three had a ReLu activation function (see Appendix A), with 100 neurons each, while the last one had a sigmoid function, since the output was binary. To train the network, k-fold cross-validation was used, with 10 folds. With this, the training data is split into k parts. There are then k training iterations, where one of the parts is used as a validation set (used to see if the network generalizes to unseen data) and the other $k-1$ are used for training. Having k different networks trained, they are all then used to classify unseen data (the test dataset), giving a greater weight to those networks that had a better result during the training. This can slightly improve the results, at zero cost. The Receiver Operating Characteristic (ROC) curves for the training and test sets are shown in Figure 4.8. This curve shows the true positive rate versus the false positive rate. The greater the area under this curve (AUC), the better the classifying capabilities of the trained network. The area under the test ROC curve is slightly lower than the one for the training set, since the network does not completely generalize.

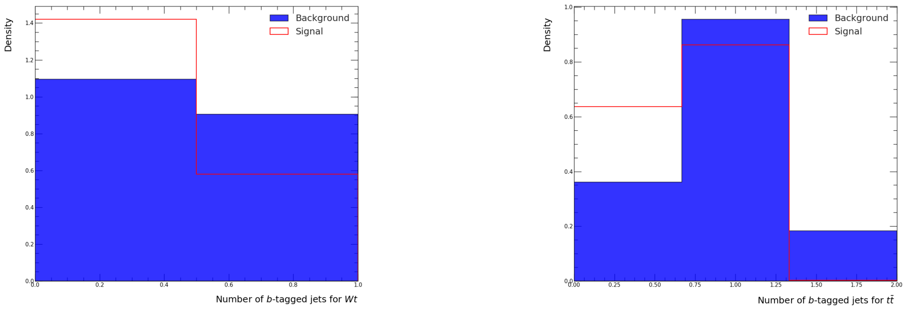
The cut in the neural network output chosen was the one that maximized the signal significance for the entire simulated set. The significance is plotted in Figure 4.9 with the $\pm 1\sigma$ bands. For a neural network output bigger than 0.67, the event is considered signal. This gives a signal significance of 5.98 ± 0.89. The signal and background density distributions from the output of the neural network are shown in Figure 4.10.

Table 4.2 Variables used to train the neural network

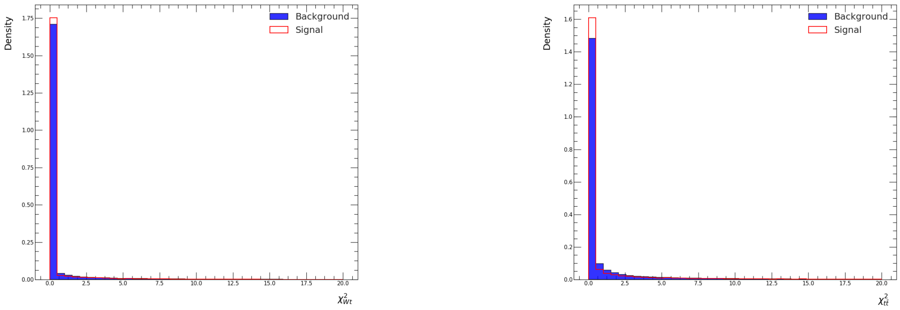
	Variable
Leptons	p_T of leading lepton p_T of second lepton
Jets	p_T of leading jet p_T of second jet Mass of leading jet Mass of second jet η of leading jet η of second jet $\Delta R(j_1, j_2)$
Kinematic Reconstruction	$\chi^2_{t\bar{t}}$ χ^2_{Wt} χ^2_{Z} Number of b-tagged jets for Wt hypothesis Number of b-tagged jets for $t\bar{t}$ hypothesis $\Delta R_{t\bar{t}}(j_1, j_2)$
Missing energy	E_T^{miss}



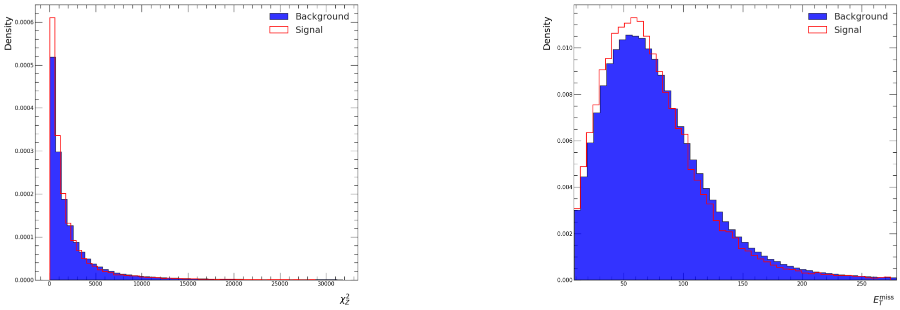
(a) p_T of leading lepton (b) p_T of second lepton



(c) Number of b-tagged jets for Wt hypothesis (d) Number of b-tagged jets for $t\bar{t}$ hypothesis

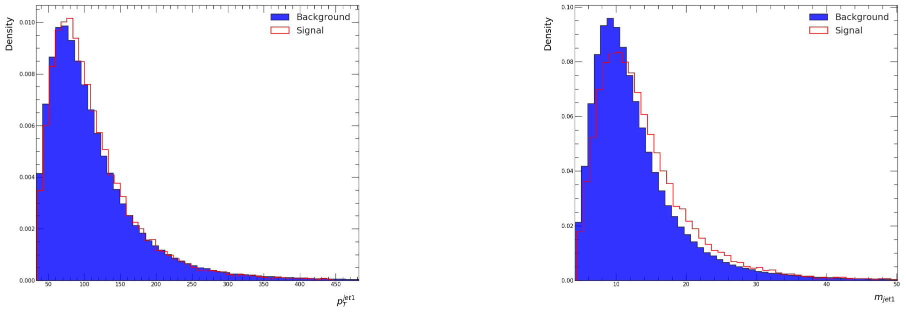


(e) χ^2_{Wt} (f) $\chi^2_{t\bar{t}}$



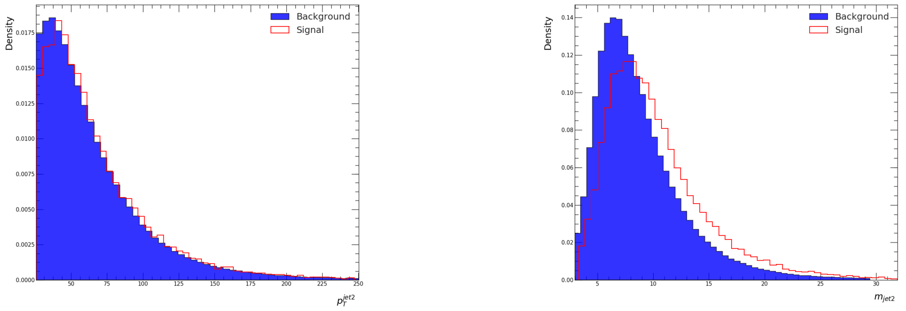
(g) χ^2_Z (h) E_T^{miss}

Fig. 4.6 Probability density functions of variables used to train the neural network.



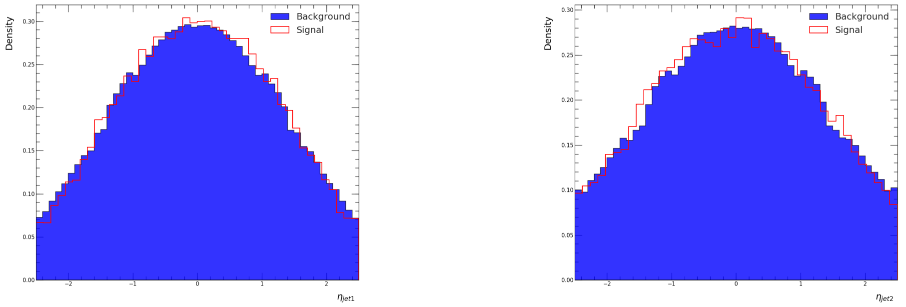
(a) p_T of leading jet

(b) Mass of leading jet



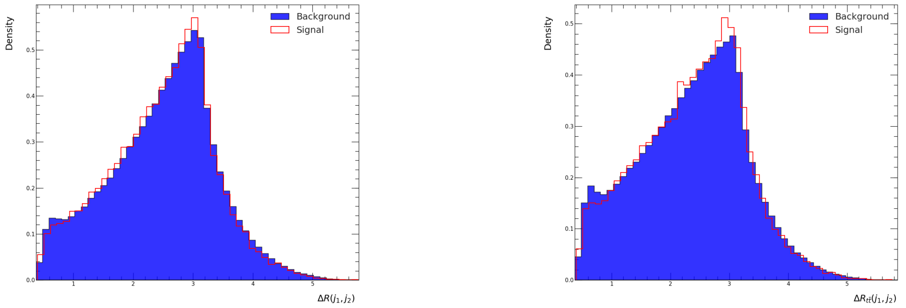
(c) p_T of second jet

(d) Mass of second jet



(e) η of leading jet

(f) η of second jet



(g) $\Delta R(j_1, j_2)$

(h) $\Delta R_{t\bar{t}}(j_1, j_2)$

.

Fig. 4.7 Probability density functions of variables used to train the neural network (continuation).

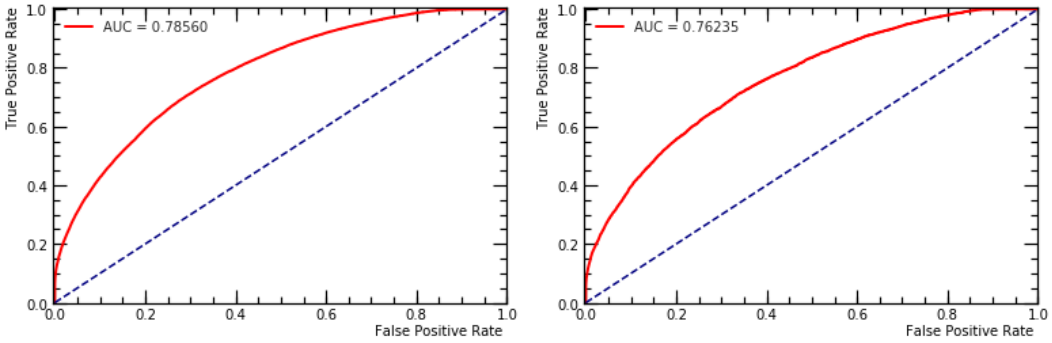


Fig. 4.8 ROC curve for the training set (left) and test set (right), in the SR.

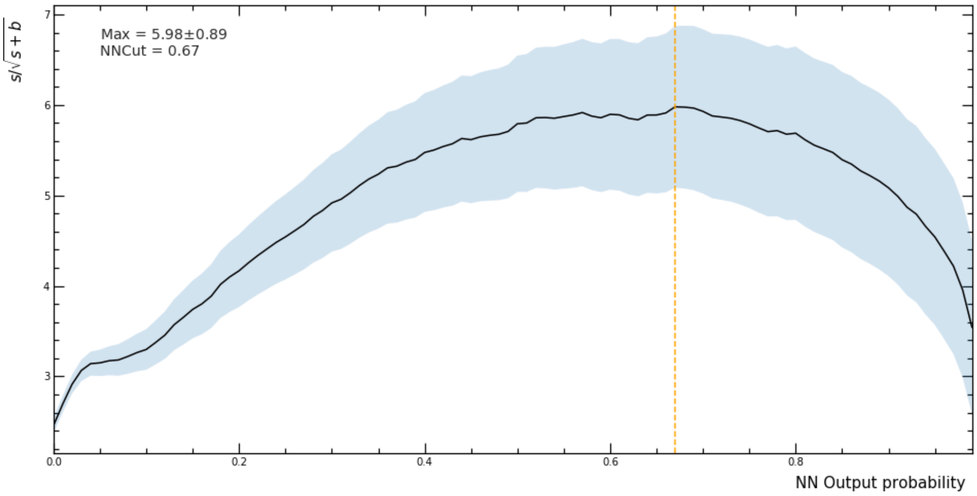


Fig. 4.9 Signal significance with statistical $\pm 1\sigma$ bands, in the SR.

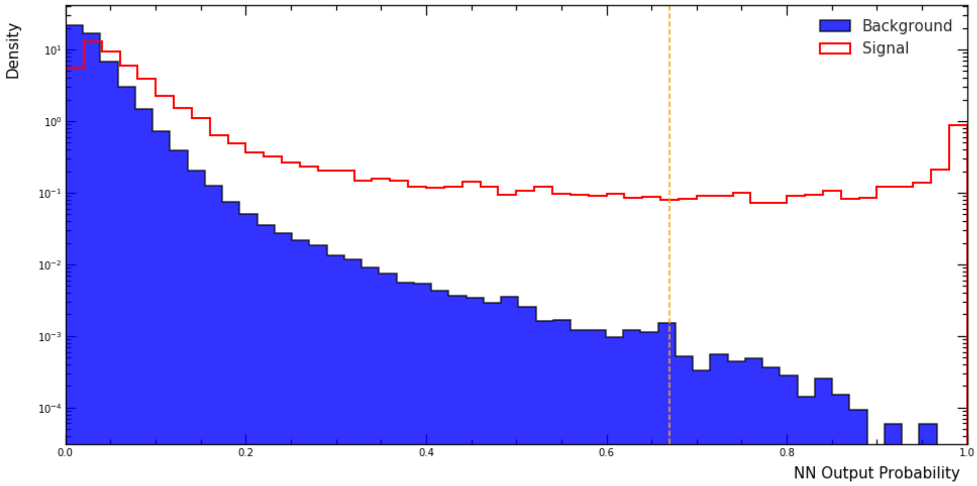


Fig. 4.10 Neural network classifier distribution, in the SR.

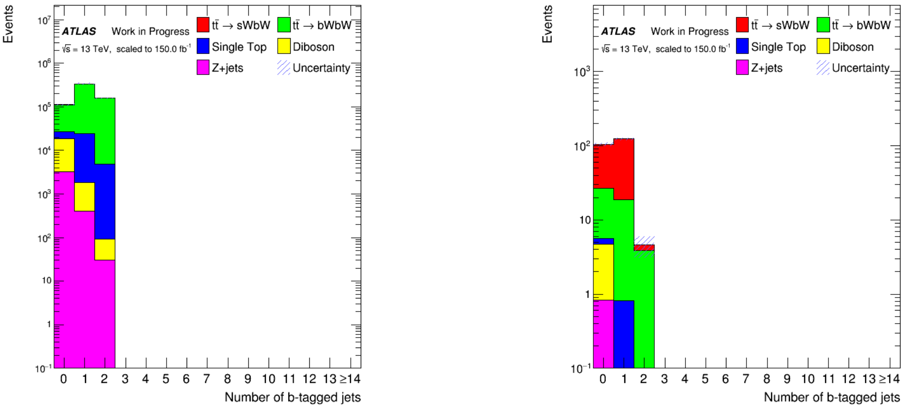


Fig. 4.11 Number of *b*-tagged jets after kinematic reconstruction cuts (left) and after neural network output cut (right), in the SR.

4.4 Background Estimation

To obtain a better measurement of $|V_{ts}|^2$, it is important to have a good estimation of the backgrounds that still contaminate the SR. To do that, specific regions, called Control Regions (CR), are defined that try to isolate each one of those backgrounds. Additionally, other regions can be defined which are only used to validate the analysis, that is, to see if the simulation matches the data even in regions where there is no fit involved. These are called Validation Regions (VR). For this analysis, one CR and two VR were defined. They are described in this section.

4.4.1 $t\bar{t} \to bWbW$ control region

The biggest background by far is $t\bar{t} \to bWbW$ due to the great similarity between its final state topology and of the signal. For this reason, the preselection and kinematic reconstruction cuts are exactly the same as those of the SR.

To try to further isolate this background, a neural network was used one more time. The method and variables where exactly the same as in the SR, only this time the $t\bar{t} \to bWbW$ events were used as signal and the other samples as background. The results are shown in Figures 4.12, 4.13 and 4.14. However, by looking at Figure 4.13, it can be seen that using a neural network does not improve the $t\bar{t} \to bWbW$ efficiency. In fact, it slightly decreases it. For that reason, a neural network specific to this region was not used. Instead, the opposite neural network cut of the SR was applied, that is, it was required that the neural network output of the SR be less than 0.67. The number of b-tagged jets after this cut is shown in Figure 4.15. Since this analysis also aimed to measure BR($t \to bW$), the plots of this region do not show the data points.

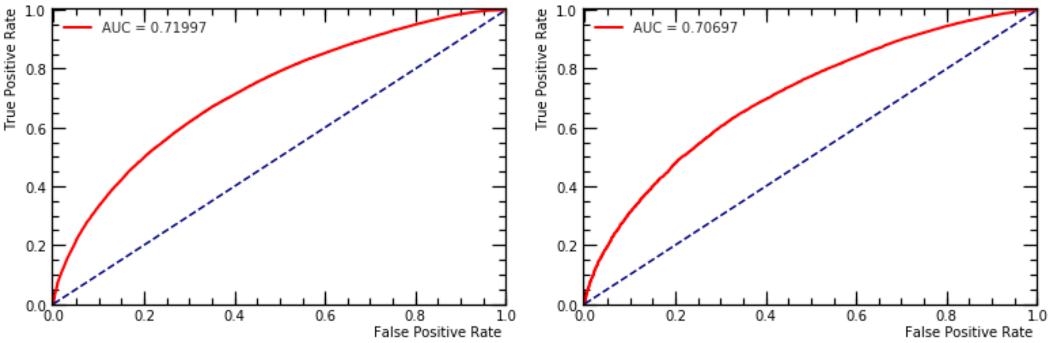


Fig. 4.12 ROC curve for the training set (left) and test set (right) in the $t\bar{t} \to bWbW$ CR.

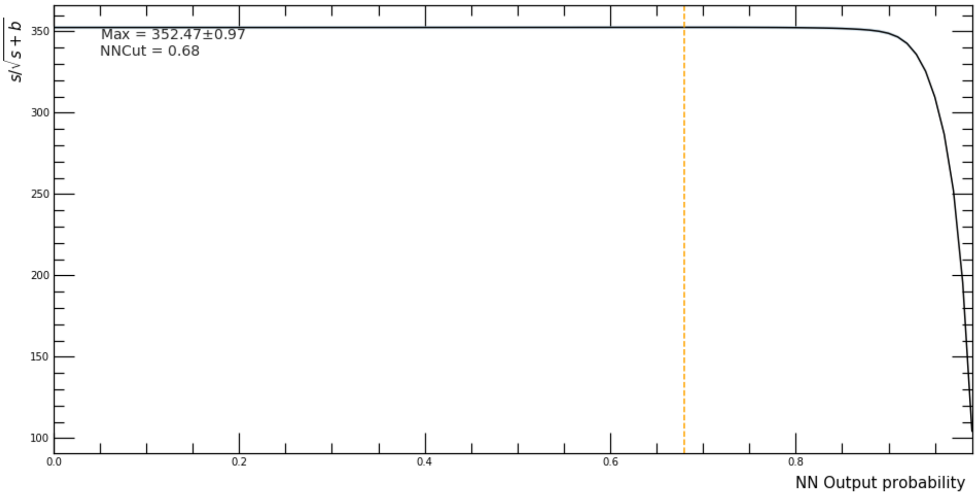


Fig. 4.13 Signal significance with statistical ±1σ bands, in the $t\bar{t} \to bWbW$ CR.

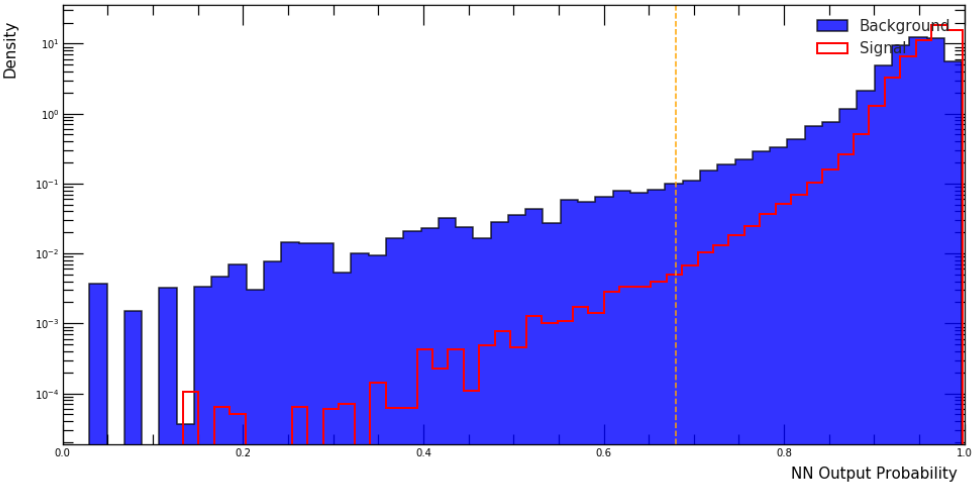


Fig. 4.14 Neural network classifier distribution, in the $t\bar{t} \to bWbW$ CR.

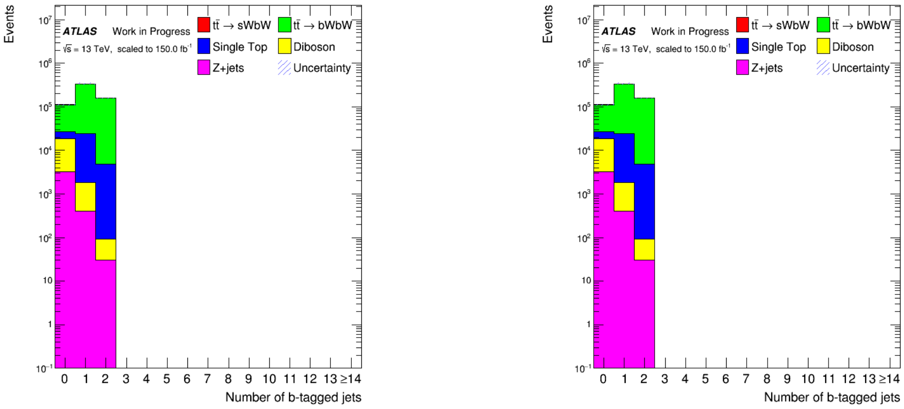


Fig. 4.15 Number of b-tagged jets after kinematic reconstruction cuts (left) and after neural network output cut (right), in the $t\bar{t} \to bWbW$ CR.

4.4.2 Single top validation region

To validate the analysis, one VR was defined with a greater contribution of the background due to the single top, Wt, with the W boson decaying leptonically and the top quark decaying to a b-quark and a W boson, which in turn also decays leptonically. Once more, both the preselection and kinematic reconstruction phases are the same as the SR. The number of b-tagged jets before and after the cuts are shown in Figure 4.16. Since some of the plots are similar to the SR and the $t\bar{t} \to bWbW$ CR, only the final plot of this region shows the data points.

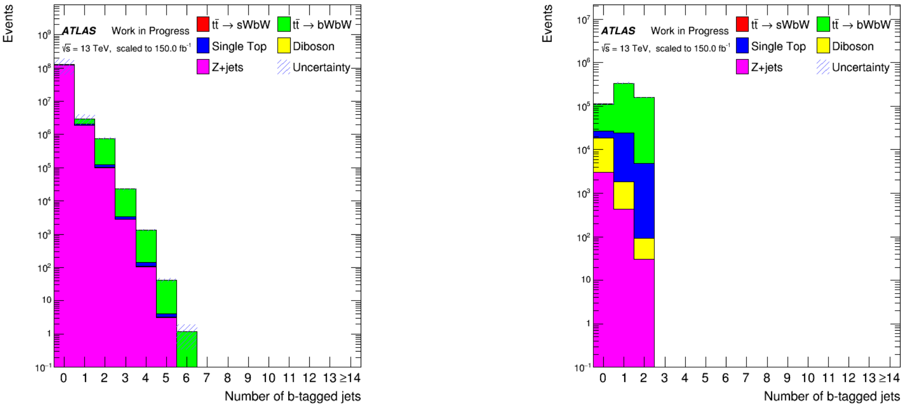


Fig. 4.16 Number of b-tagged jets before any cuts (left) and after kinematic reconstruction cuts (right), in the Wt VR.

To further isolate this background, another neural network is used. The model and variables are the same. The difference is that the Wt events are used as signal. The results are shown in Figures 4.17, 4.18 and 4.19. The chosen neural network output cut is 0.03.

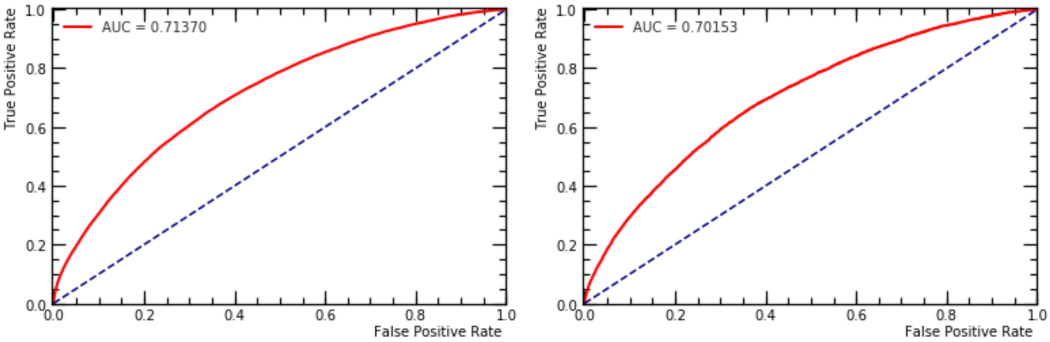


Fig. 4.17 ROC curve for the training set (left) and test set (right) in the Wt VR.

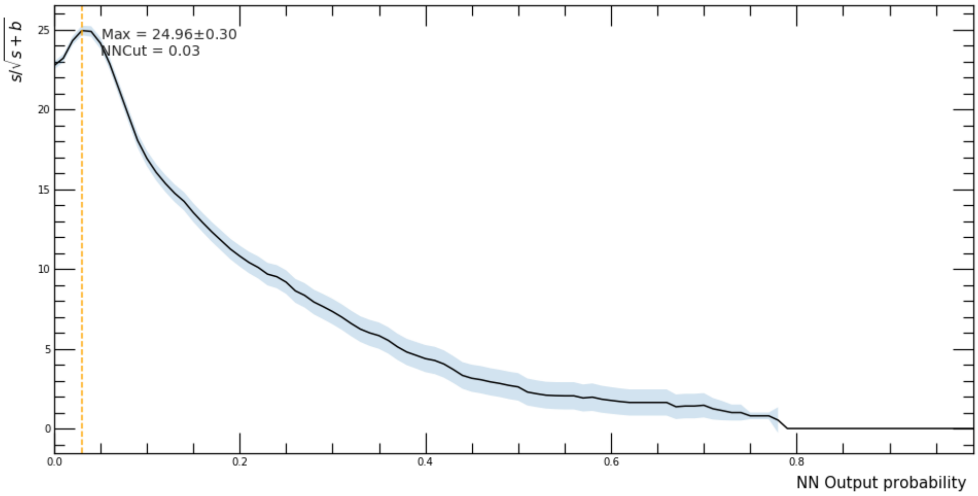


Fig. 4.18 Signal significance with statistical $\pm 1\sigma$ bands, in the Wt VR.

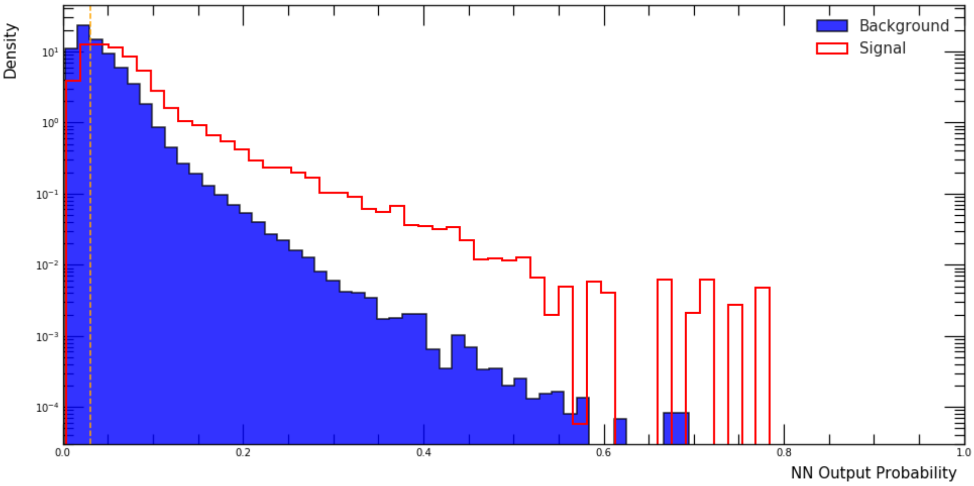


Fig. 4.19 Neural network classifier distribution, in the Wt VR.

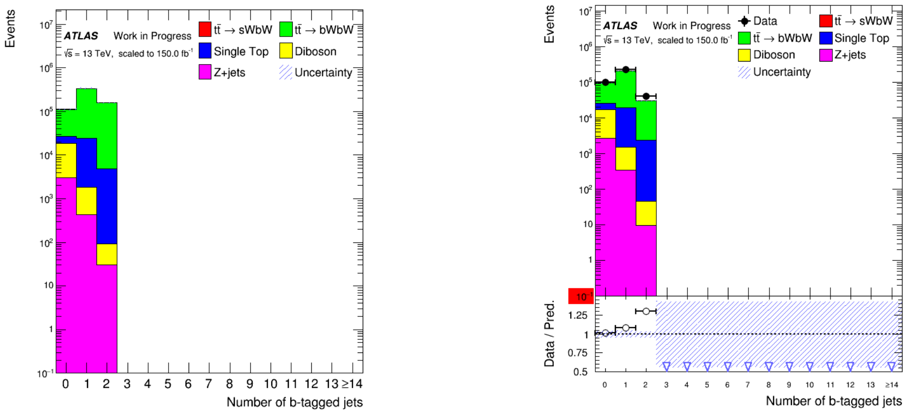


Fig. 4.20 Number of b-tagged jets after kinematic reconstruction cuts (left) and after neural network cut (right), in the Wt VR.

4.4.3 Z+jets validation region

The second and final validation region defined has a greater contribution from Z+jets events. This region requires that $m_{\ell_a \ell_b \nu_a^1 \nu_a^2 \nu_b^1 \nu_b^2}$ from the kinematic recostruction of the $Z \to \tau\tau$ hypothesis be within 80 GeV and 100 GeV.

To increase the agreement between these samples and the data, two fits were performed, using the `TRexFitter` tool: one that required that the leptons be two electrons of opposite charges and another with two muons of opposite charges. For these fits, the Z+jets samples were separated in the three possible jet filters: with b-quarks, charm quarks and light (u, d and s) quarks. Six normalization factors were then obtained. They are shown in Table 4.3. The plots before and after the fit are shown in Figures 4.21 and 4.22.

The $Z \to \tau\tau$ events were corrected by multiplying each type of sample by $2\sqrt{\varepsilon_e^j \varepsilon_\mu^j}$, where ε_ℓ is the normalization factor of the events with two leptons ℓ (electron or muon) and j is the filter of the sample (b, charm or light). The reason for this multiplication is the assumption that each lepton ℓ contributes $\sqrt{\varepsilon_\ell}$ to the total normalization factor in each fit. Since in the Z+jets VR the two τs from the $Z \to \tau\tau$ decay to one electron and one muon, there is one contribution from each. The multiplication by 2 comes from the fact that there are two combinations where this happens, depending on which τ decays to which ℓ. The results are shown in Figure 4.23.

Table 4.3 Normalization scale factors obtained in both Z+jets VR fits.

Selection	Z+light	Z+charm	Z+bottom
ee opposite charge	0.7412 ± 0.0009	1.215 ± 0.004	1.059 ± 0.008
$\mu\mu$ opposite charge	1.1866 ± 0.0005	0.9390 ± 0.0024	2.218 ± 0.004

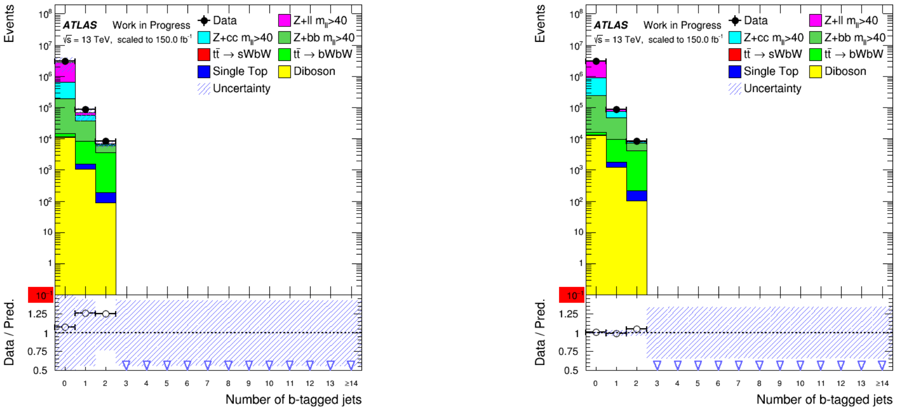


Fig. 4.21 Number of *b*-tagged jets, for events with two electrons of opposite charge, before the fit (left) and after the fit (right), in the *Z*+jets VR.

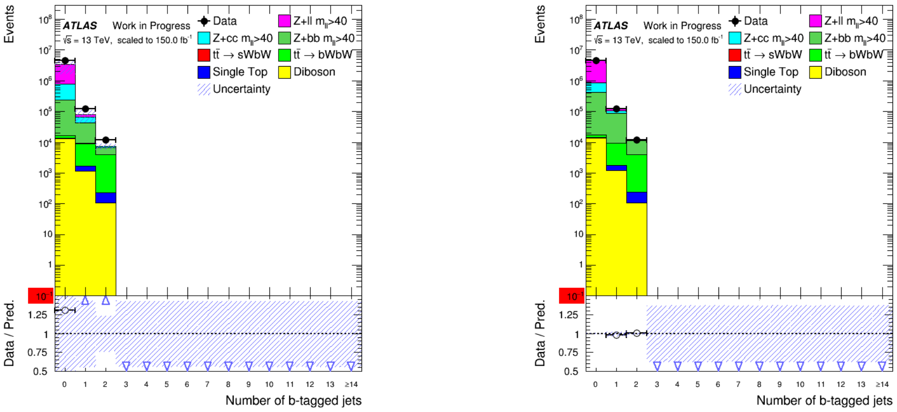


Fig. 4.22 Number of *b*-tagged jets, for events with two muons of opposite charge, before the fit (left) and after the fit (right), in the *Z*+jets VR.

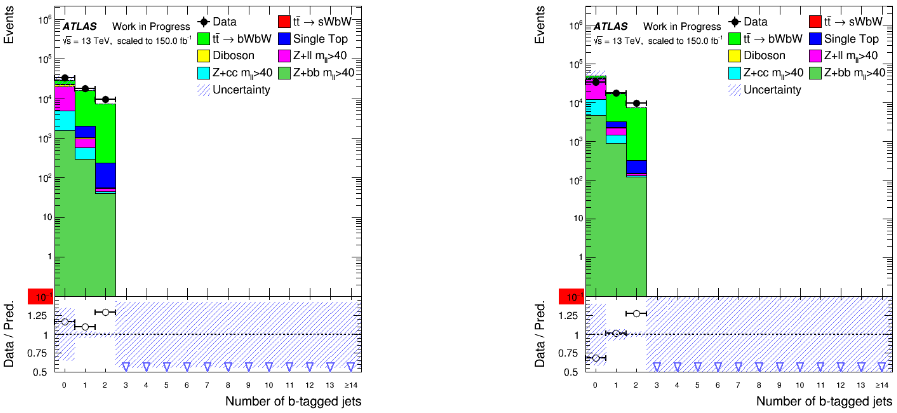


Fig. 4.23 Number of *b*-tagged jets, in the *Z*+jets VR, before (left) and after (right) the fits.

4.5 Linearity study

To verify that the signal fit to the data, done with the `TRexFitter` tool, was behaving properly, a linearity study was devised, with three steps. In all of these steps, the cross-section of $t\bar{t}$ production is kept at its SM value and it is assumed that the top quark decays only to the bottom and strange quarks and that the cross section of $sWsW$ is negligible. Then, the hypothesis of the cross-section of the signal is scaled by several factors:

$$\sigma_{sWbW} \to \sigma'_{sWbW} = \alpha\sigma_{sWbW} \tag{4.8}$$

with $\alpha = \{0.1, 0.5, 1, 5, 10\}$. The cross-section of the biggest background ($t\bar{t} \to bWbW$) must then be scaled as well:

$$\sigma'_{sWbW} + \sigma'_{bWbW} = \sigma_{t\bar{t}} \tag{4.9}$$

The corresponding transformation of σ_{bWbW} is

$$\sigma_{bWbW} \to \sigma'_{bWbW} = \beta\sigma_{bWbW}, \qquad \text{with} \qquad \beta = \frac{1-2\alpha|V_{ts}|^2|V_{tb}|^2}{|V_{tb}|^4} \tag{4.10}$$

where

$$\sigma_{bWbW} = \sigma_{t\bar{t}}|V_{tb}|^4 \tag{4.11}$$

$$\sigma_{sWbW} = 2\sigma_{t\bar{t}}|V_{ts}|^2|V_{tb}|^2 \tag{4.12}$$

were used.

From the fits, the new cross-sections are obtained, from which the branching ratios $|V_{ts}|^2$ and $|V_{tb}|^2$ can be calculated. By plotting the fitted values of both cross-sections and both branching ratios against their pre-fit values and then performing a linear fit, it is possible to see if the fitter behaves in a linear fashion. This is done by looking at the linear fit parameters. This test is also used to estimate the systematic uncertainty related to the b-tagging efficiency of the data, as well as the b-tagging calibration.

The way this test is designed is: firstly, the signal and the biggest background ($t\bar{t} \to bWbW$) are fitted to a simulated (Asimov) dataset (that is equal to the sum of the signal and every background), for each scale factor of the signal. The second step involves changing the default efficiency cut of the b-tagging algorithm, which is 77%, by ±10% and then fitting the signal and $t\bar{t} \to bWbW$ samples to an Asimov dataset, for each cut variation.

For both of these steps, the input cross-section and branching ratios are maintained after the fit, which can be seen from Figures 4.24, 4.25, 4.26, 4.27, 4.28 and 4.29. This shows that the fitting tool behaves in a linear fashion.

The last step is then used to estimate the uncertainty related to the b-tagging efficiency, along with the b-tagging calibration. In it, the simulated samples with the default cut are fitted, for the various scale factors, to the Asimov datasets with $\pm 10\%$ efficiency cut. The uncertainty estimates are then obtained from the linear fit parameters, that is, the biggest variations from unity of the slope parameters (for the cross-section fits) are used as a systematic uncertainties in the global fit (see section 4.7) for the signal and the $t\bar{t} \to bWbW$. These uncertainties are: 6.26% for signal and 76.9% for the background. The fit results are shown in Figures 4.30, 4.31, 4.32 and 4.33.

The pre- and post-fit histograms of this linearity study are shown in B.

4.5.1 Asimov data and simulated samples with b-tagging efficiency of 77%

The first step in this study fits the signal and the biggest background to the Asimov dataset in the SR and $t\bar{t} \to bWbW$ CR, keeping the other backgrounds constant. The b-tagging efficiency in this case is the one used throughout the analysis, 77%. The scale factors obtained with the fit are in Table 4.4. The linear fits of the branching ratios and cross-sections are presented in Figures 4.24 and 4.25.

Table 4.4 Scale factors (μ) obtained with the fit, with normal b-tagging efficiency, for different signal scale factors

α	μ_{sWbW}	μ_{bWbW}
0.1	1.0 ± 0.5	1.00 ± 0.04
0.5	1.00 ± 0.13	1.000 ± 0.004
1	0.99 ± 0.09	1.000 ± 0.032
5	1.00 ± 0.04	1.000 ± 0.011
10	1.000 ± 0.034	1.000 ± 0.010

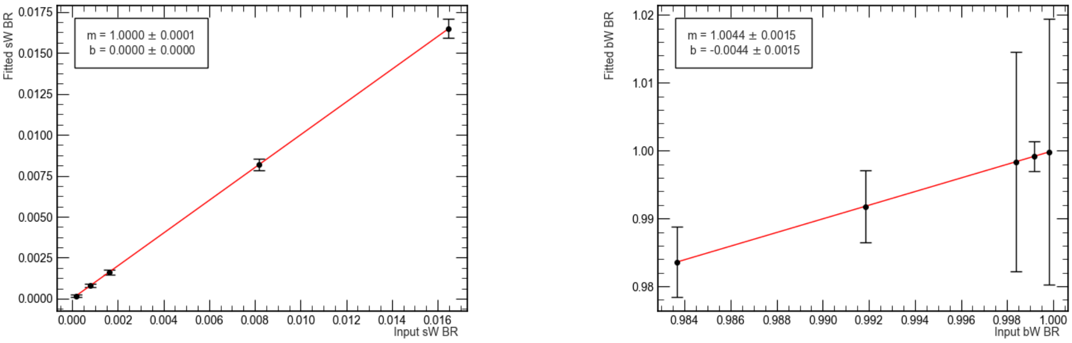


Fig. 4.24 Fitted branching ratios versus input branching ratios for normal b-tagging efficiency, for $t \rightarrow sW$ (left) and $t \rightarrow bW$ (right).

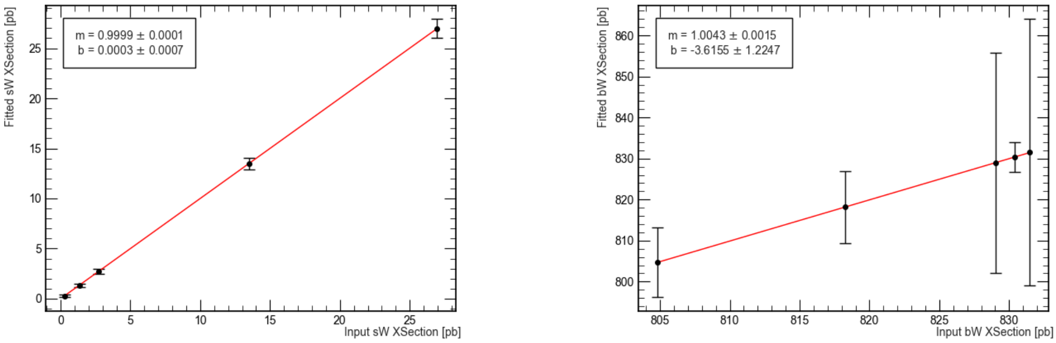


Fig. 4.25 Fitted cross-sections versus input cross-sections for normal b-tagging efficiency, for $t\bar{t} \rightarrow sWbW$ (left) and $t\bar{t} \rightarrow bWbW$ (right).

4.5.2 Asimov data and simulated samples with +10% b-tagging efficiency

The next step performs the same fits, but with a b-tagging efficiency of 87%. The scale factors obtained with the fit are in Table 4.5. The linear fits of the branching ratios and cross-sections are presented in Figures 4.26 and 4.27.

Table 4.5 Scale factors (μ) obtained with the fit, with +10% b-tagging efficiency, for different signal scale factors

α	μ_{sWbW}	μ_{bWbW}
0.1	1.0 ± 0.5	1.000 ± 0.010
0.5	1.00 ± 0.15	1.000 ± 0.034
1	1.00 ± 0.10	1.00 ± 0.04
5	1.00 ± 0.04	1.000 ± 0.012
10	1.000 ± 0.029	1.000 ± 0.004

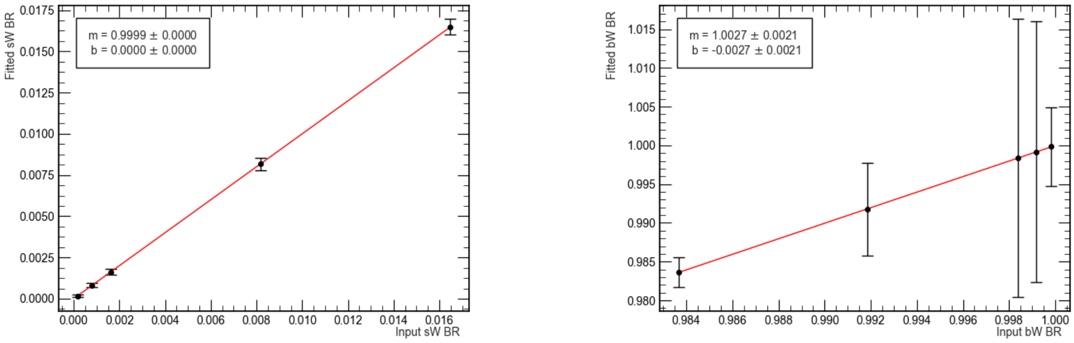


Fig. 4.26 Fitted branching ratios versus input branching ratios for +10% b-tagging efficiency, for $t \longrightarrow sW$ (left) and $t \longrightarrow bW$ (right).

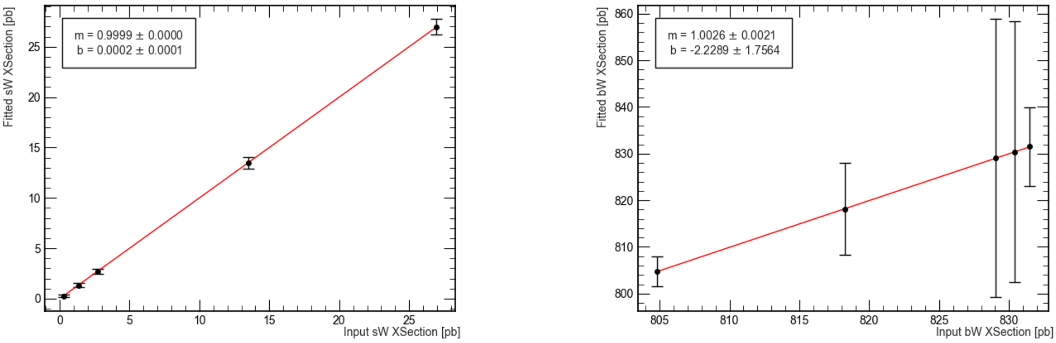


Fig. 4.27 Fitted cross-sections versus input cross-sections for +10% b-tagging efficiency, for $t\bar{t} \longrightarrow sWbW$ (left) and $t\bar{t} \longrightarrow bWbW$ (right).

4.5.3 Asimov data and simulated samples with -10% b-tagging efficiency

The next step performs the same fits, but with a b-tagging efficiency of 67%. The scale factors obtained with the fit are in Table 4.6. The linear fits of the branching ratios and cross-sections are presented in Figures 4.28 and 4.29.

Table 4.6 Scale factors (μ) obtained with the fit, with -10% b-tagging efficiency, for different signal scale factors

α	μ_{sWbW}	μ_{bWbW}
0.1	1.0 ± 0.5	1.00 ± 0.04
0.5	1.00 ± 0.14	1.000 ± 0.012
1	1.00 ± 0.09	1.00 ± 0.04
5	1.00 ± 0.04	1.000 ± 0.011
10	1.000 ± 0.029	0.999 ± 0.004

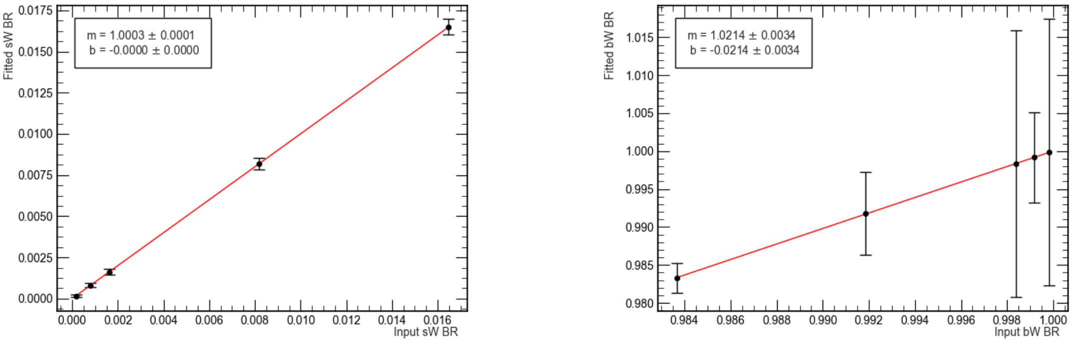


Fig. 4.28 Fitted branching ratios versus input branching ratios for -10% b-tagging efficiency, for $t \to sW$ (left) and $t \to bW$ (right).

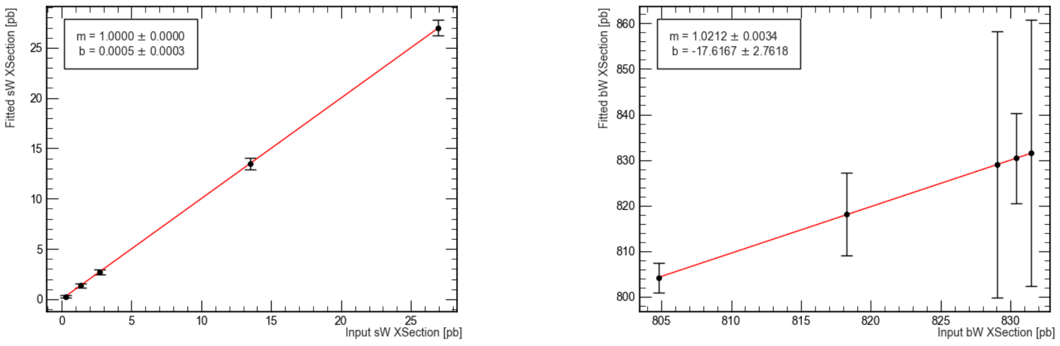


Fig. 4.29 Fitted cross-sections versus input cross-sections for -10% b-tagging efficiency, for $t\bar{t} \to sWbW$ (left) and $t\bar{t} \to bWbW$ (right).

4.5.4 Asimov data with +10% b-tagging efficiency and simulated samples with normal b-tagging efficiency

This step performs the fits using the simulated samples with a normal b-tagging efficiency, but with an Asimov dataset with +10% b-tagging. The scale factors obtained with the fit are in Table 4.7. The linear fits of the branching ratios and cross-sections are presented in Figures 4.30 and 4.31.

Table 4.7 Scale factors (μ) obtained with the fit, for Asimov dataset with +10% b-tagging efficiency and simulated samples with normal b-tagging efficiency, for different signal scale factors

α	μ_{sWbW}	μ_{bWbW}
0.1	1.2 ± 0.5	1.165 ±0.031
0.5	0.98 ± 0.14	1.164 ±0.030
1	0.95 ± 0.09	1.166 ±0.031
5	0.93 ± 0.04	1.177 ±0.032
10	0.941 ± 0.034	1.20 ±0.05

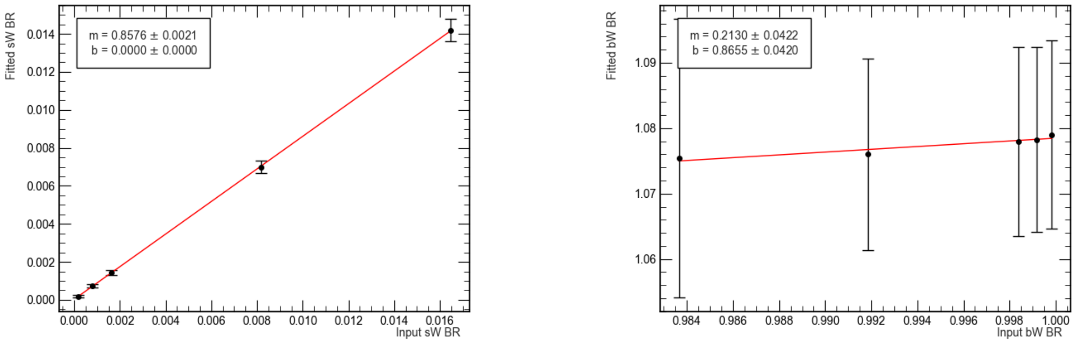


Fig. 4.30 Fitted branching ratios versus input branching ratios for Asimov dataset with +10% b-tagging efficiency and simulated samples with normal b-tagging efficiency, for $t \to sW$ (left) and $t \to bW$ (right).

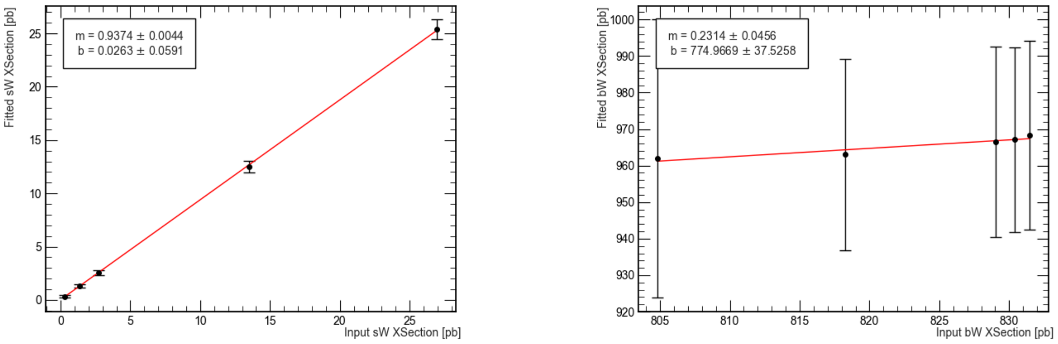


Fig. 4.31 Fitted cross-sections versus input cross-sections for Asimov dataset with +10% b-tagging efficiency and simulated samples with normal b-tagging efficiency, for $t\bar{t} \to sWbW$ (left) and $t\bar{t} \to bWbW$ (right).

4.5.5 Asimov data with -10% b-tagging efficiency and simulated samples with normal b-tagging efficiency

This final step performs the fits using the simulated samples with a normal b-tagging efficiency, but with an Asimov dataset with -10% b-tagging. The scale factors obtained with the fit are in Table 4.8. The linear fits of the branching ratios and cross-sections are presented in Figures 4.32 and 4.33.

Table 4.8 Scale factors (μ) obtained with the fit, for Asimov dataset with -10% b-tagging efficiency and simulated samples with normal b-tagging efficiency, for different signal scale factors

α	μ_{sWbW}	μ_{bWbW}
0.1	0.00 ± 0.24	0.454 ± 0.006
0.5	0.73 ± 0.14	0.551 ± 0.002
1	0.79 ± 0.09	0.428 ± 0.002
5	0.98 ± 0.04	0.547 ± 0.003
10	0.99 ± 0.04	0.463 ± 0.011

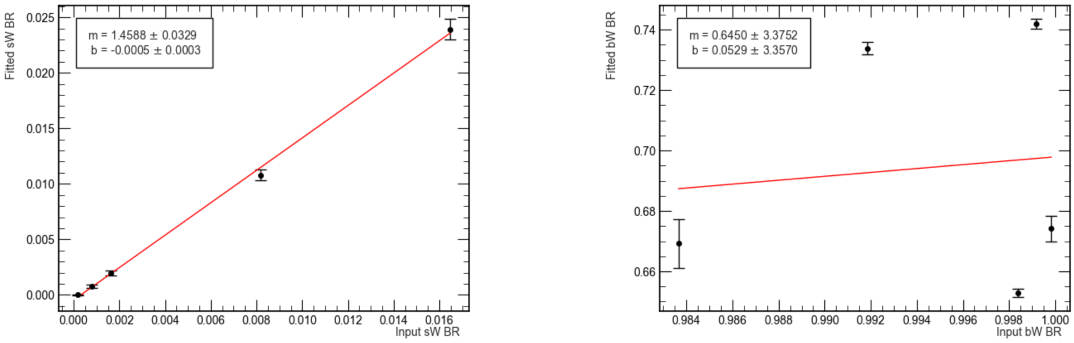


Fig. 4.32 Fitted branching ratios versus input branching ratios for Asimov dataset with -10% b-tagging efficiency and simulated samples with normal b-tagging efficiency, for $t \to sW$ (left) and $t \to bW$ (right).

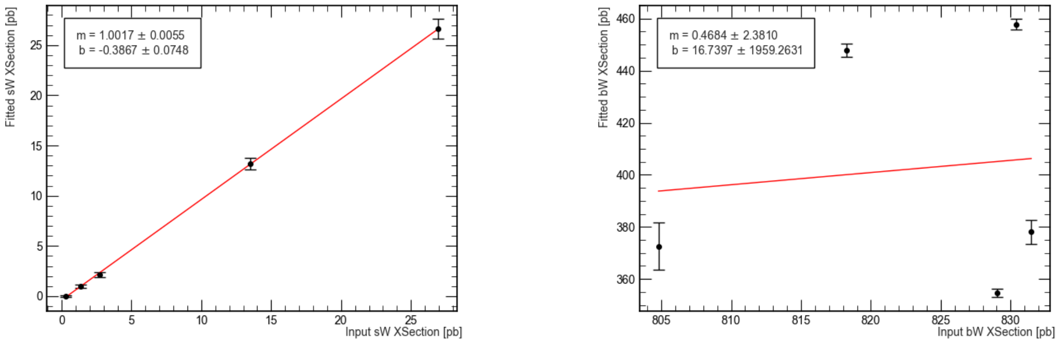


Fig. 4.33 Fitted cross-sections versus input cross-sections for Asimov dataset with -10% b-tagging efficiency and simulated samples with normal b-tagging efficiency, for $t\bar{t} \to sWbW$ (left) and $t\bar{t} \to bWbW$ (right).

4.6 Systematic Uncertainties

The systematic uncertainties present in this analysis are discussed here. There are two main types of them. The first one is experimental, related to the confidence with which each object's properties are measured. The other arises from theory, with the uncertainty of the modelling of each type of event.

These uncertainties affect different samples in separate ways. A full analysis takes this into account throughout it. At the time of writing, though, only a part the the systematic uncertainties were available. The missing ones were estimated mostly from [61]. Some were also estimated from [62].

4.6.1 Experimental Uncertainties

The effect of each type of experimental uncertainty is shown in Table 4.9. These include uncertainties related to the energy resolution of objects (jets, electrons and muons) as well as to reconstruction of the missing transverse energy. Additional uncertainties are associated with the lepton identification and trigger efficiencies. Finally, a prominent experimental uncertainty is that related to b-tagging efficiency, treated separately from the one obtained with the linearity study.

These uncertainties were not taken into account in the signal sample, since the one used was at truth-level. For that reason, their magnitude is indicated as 0% in the Table 4.9.

4.6.2 Modelling Uncertainties

Theory uncertainties arise from modelling the event generation process. To quantify these, different generators might be used for the different samples or the parameters of those generators changed. For this analysis, these uncertainties were estimated from similar analyses. The magnitudes are shown in Table 4.10. Also included here are the uncertainties from the Z+jets fit with ee and $\mu\mu$ (applied only to the Z+jets samples), as well as the linearity study uncertainty, which was quantified to be $\pm 6.26\%$ for the signal and $\pm 76.9\%$ for the $t\bar{t} \to bWbW$ sample.

Table 4.9 Experimental systematic uncertainties.

Source of uncertainty	$t\bar{t} \to sWbW$	$t\bar{t} \to bWbW$	Single top	Diboson	Z+jets
Jet Energy Scale (JES)[1]	0%	+4%	+4%	+4%	+4%
Jet Energy Resolution (JER)[1]	0%	-1.5%	-1.5%	-1.5%	-1.5%
JVT	0%	0%	0%	0%	±1.8%
b-tagging	0%	0%	0%	0%	±2.4%
Electron Momentum Scale[1]	0%	+0.20%	+0.20%	+0.20%	+0.20%
Electron Momentum Resolution[1]	0%	+0.05%	+0.05%	+0.05%	+0.05%
Electron Isolation	0%	0%	0%	±0.0023%	±0.4%
Muon Momentum Scale[1]	0%	-0.05%	-0.05%	-0.05%	-0.05%
Muon Momentum Resolution[1]	0%	+0.03%	+0.03%	+0.03%	+0.03%
Muon Identification	0%	0%	0%	0%	±5.0%
Muon Isolation	0%	0%	0%	0%	1.2%
Muon Trigger Efficiency	0%	0%	0%	±1%	±9.6%
E_T^{miss} Scale and Resolution[2]	0%	±3%	±3%	±3%	±3%
Pile-up	0%	0%	0%	±0.007%	0%
Luminosity[3]	0%	±2.1%	±2.1%	±2.1%	±2.1%

Table 4.10 Modelling systematic uncertainties.

Sample	Magnitude
Diboson[1]	±10%
Single top[1]	$^{+20\%}_{-50\%}$
Z+jets[1]	±50%

---

[1] Taken from [61].

[2] Taken from [62].

[3] Taken from [63].

4.7 BR measurement

To obtain the sensitivity to the measurement of the branching ratio $|V_{ts}|^2$, a global fit is performed. With it, each simulated sample is fitted to the data in its corresponding region: the signal in the SR and the $t\bar{t} \to bWbW$ in its CR. The Wt and Z+jets VRs are used to validate the fit. Given the preliminary nature of this study, in the SR and $t\bar{t} \to bWbW$ CR the samples are fitted to an Asimov dataset.

The fit is done by minimizing a global χ^2 that is built from all bins of all histograms. Besides scaling the normalization of the samples up and down, the fit also takes into account the systematic uncertainties. This allows the shape of the histograms to change, by moving events from one bin to another, to better minimize the χ^2.

The histogram used for the global fit was the number of b-tagged jets in each region, after the cut on the neural network output. These histograms, before and after the fit, are shown in Figure 4.34. The normalization scale factors obtained are shown in Table 4.11.

Table 4.11 Normalization scale factors (μ) obtained in the global fit

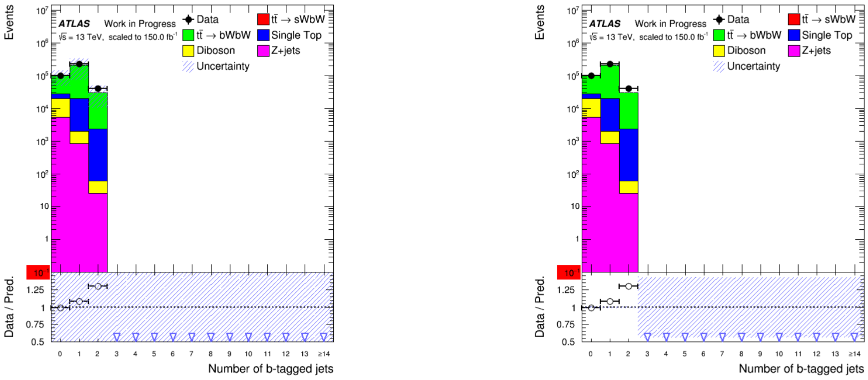
Samples	μ
$t\bar{t} \to bWbW$	$1.0000 \pm 0.0120(\text{stat}) \pm 0.0013(\text{syst})$
$t\bar{t} \to sWbW$	$1.00 \pm 0.06(\text{stat}) \pm 0.08(\text{syst})$

From the normalization scale factors, one can obtain the BR, from the theoretical ones used to simulate the samples. The two BR measured, both for $t \to sW$ and $t \to bW$ are

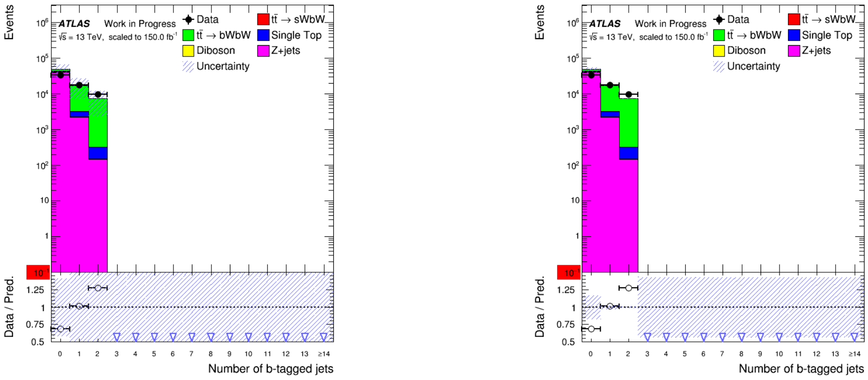
$$|V_{ts}|^2 = 0.00162 \pm 0.00013(\text{stat}) \pm 0.00014(\text{syst}) \tag{4.13}$$

$$|V_{tb}|^2 = 0.9983 \pm 0.0060(\text{stat}) \pm 0.0006(\text{syst}) \tag{4.14}$$

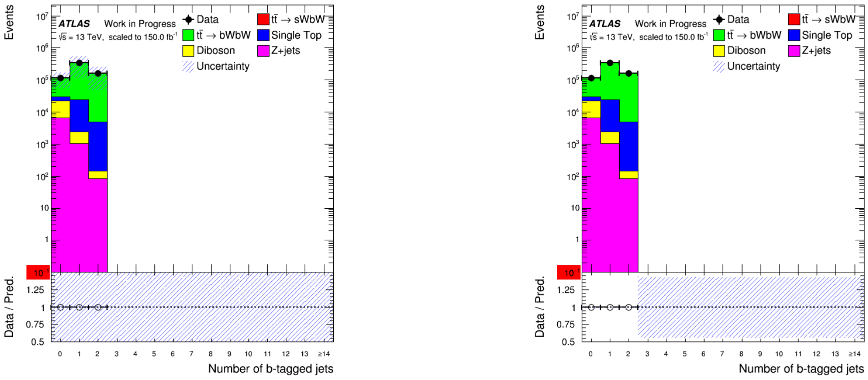
where the first uncertainty is statistical and the second is related to the systematics.



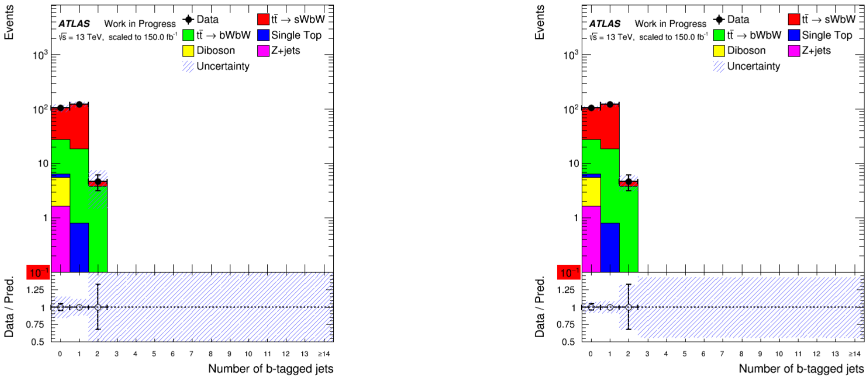
(a) Wt VR



(b) Z+jets VR



(c) $t\bar{t} \to bWbW$ CR



(d) SR

Fig. 4.34 Number of b-tagged jets in the several regions before (left) and after (right) the global fit.

Chapter 5

Conclusions

The sensitivity for the measurement of the branching ratio of $t \to sW$ was studied, using proton-proton collisions at $\sqrt{s} = 13$ TeV detected by ATLAS, based on the $36.2\,\mathrm{fb}^{-1}$ dataset from the 2015-2016 run, scaled to $150\,\mathrm{fb}^{-1}$.

The measurement is done by defining a signal-enriched region, with $t\bar{t}$ events, where one top quark decays through the sW channel and the other through the main top quark decay, the bW channel. A neural network is used to further supress background.

Additional regions were defined: one to control the main background, $t\bar{t} \to bWbW$, and two other to validate the fit, that were enriched with two other backgrounds, single top and Z+jets (specifically $Z \to \tau\tau$). To measure the branching ratio, a global fit was performed that fits $t\bar{t} \to bWbW$ in its control region, as well as the signal in its region, to an Asimov dataset, taking into account the systematic uncertainties, both experimental and theoretical. The branching ratio measured for $t \to sW$ was $|V_{ts}|^2 = 0.00162 \pm 0.00013$ (stat) ± 0.00014 (syst). Additionally, the branching ratio for $t \to bW$ was also measured, with the value $|V_{tb}|^2 = 0.9983 \pm 0.0060$ (stat) ± 0.0006 (syst) being obtained. This result was obtained with a signal significance of 6.0 ± 0.9 (the maximum chosen for the neural network output cut), meaning that the number of signal events is more than 5 times the statistical error of the sum of signal and background (assuming a Poisson distribution). This leads to the conclusion that a measurement of this branching ratio might be possible with the full Run II dataset.

It is necessary to note that this work is preliminary as, at the time of writing, only truth-level samples for the signal were available, as well as only part of the experimental systematic uncertainties. The next iteration of this work would include these systematics and use reconstruction-level signal samples. To further improve the measurement, a dedicated tagger for strange quarks (an s-tag) will have to be developed. This would use the information available inside the jets, such as the particle tracks, to indicate whether the origin of the jet is or not a strange quark, just like there is one for *b*-quarks. Finally, the final fit would use the full Run II dataset, instead of a partially simulated one.

Appendix A

Deep Learning

This appendix gives an overview of Deep Learning (DL) methods, starting with a historic prespective of Artificial Intelligence (AI) and where DL fits in. Then, a closer look of DL is given, with the several parts of a neural network discussed. The information presented here is based on [64].

A.1 A Brief History of Artificial Intelligence

Artificial Intelligence was started in the 1950s and it can be defined as the effort to automate intellectual tasks normally performed by humans. It has had two main phases: symbolic AI, which was dominant from the 1950s to the 1980s, and machine learning (ML), dominant from the 1990s until the present.

These two phases differ profoundly in their underlying principles. Symbolic AI tried to handcraft a large set of rules with which to perform the tasks. This was particularly useful in logical activities, such as playing chess. A ML system, on the other hand, is trained rather than explicitly programmed, by providing it with many examples relevant to the task. It then finds a statistical structure in these examples, giving the rules to perform the task. These two paradigms are represented in Fig. A.1.

Machine Learning itself has undergone a few phases, going from probabilistic modelling, by applying Bayes' Theorem or using logistic regression, through kernel methods, such as the support vector machines, to decision trees, random forests and gradient boosting machines. These methods try to find useful representations of the input data so that the algorithm output is close to the true output. But since they only focus on learning one or two layers of representations of the data, they are called shallow (as opposed to deep) learning methods.

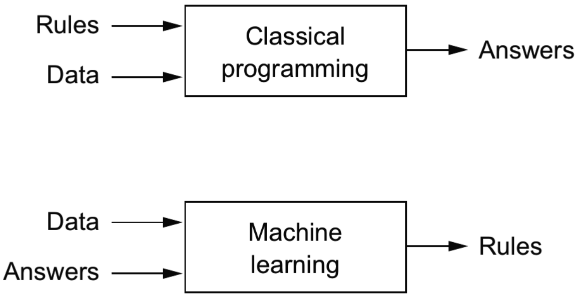


Fig. A.1 AI paradigms.

Deep Learning has been the prevailing phase of the last few years, due to a great increase in the amount of data, as well as better hardware, with the use of graphical processing units (GPU), and better algorithms used to train the neural networks used in DL. As other methods of ML, it tries to find useful representations, but DL can learn many (hundreds and even thousands) of layers that are learnt simultaneously. Hence the “deep”. It has been fundamental in the recent achievements of ML, such as near-human-level image classification and speech recognition.

A.2 Neural Networks

In DL, the layers of representations are learned via models called neural networks. Fig. A.2 shows the transformations of an image as it goes through the neural net. It can be seen as a succession of filters that purify the image so that the last representations are the most useful to the task required.

A.2.1 Anatomy of a Neural Network

A typical neural network is represented in Fig. A.3. It is made up of many layers of neurons, each transforming the input, X, in a certain way. The neural net is parametrized by a set of weights, which relates the neurons of a layer with the neurons of the next layer. The input of the neuron i of layer j, x_i^j, is related to the outputs of the layer $j-1$, $y_{i'}^{j-1}$, as

$$x_i^j = \sum_{i'} w_{i',i}^{j-1} y_{i'}^{j-1} + b^j, \tag{A.1}$$

where b^j is the bias term and $w_{i',i}^{j-1}$ is the weight associated with the connection of the neuron i' of the layer $j-1$ with the neuron i of the layer j.

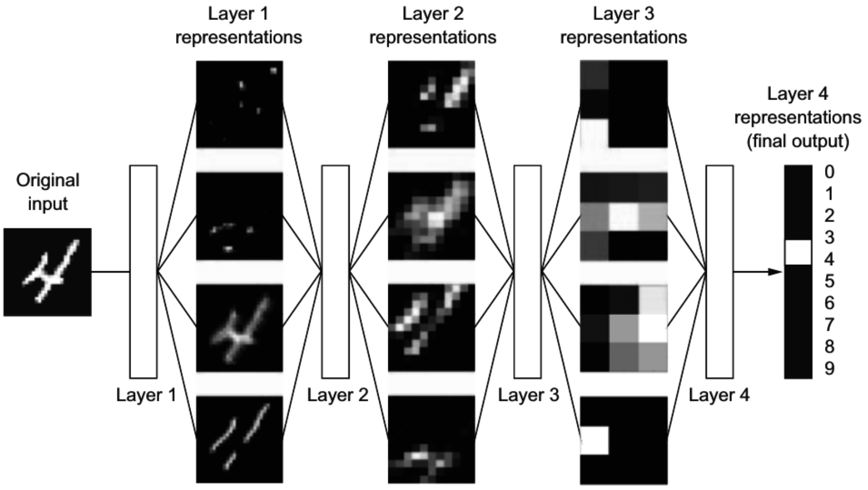


Fig. A.2 Deep representations learned by a digit-classification model.

The output of each neuron is obtained using an activation function. If no activation function is used, the net can only obtain linear relations between transformations. These functions, then, allow the learning of more complex relations. The output of neuron i of layer j, y_i^j, using an activation function f is

$$
\begin{aligned}
y_i^j &= f(x_i^j) && \text{(A.2)} \\
&= f\left(\sum_{i'} w_{i',i}^{j-1} y_{i'}^{j-1} + b^j\right) && \text{(A.3)}
\end{aligned}
$$

The two activation functions used in the analysis are represented in Fig. A.4.

The weights are normally randomly initiated and are adjusted during the training process so that the output of the neural net, Y', is as close to the true output, Y as possible. This is quantified using a loss function, which is chosen according to the task that the neural net is to perform. For example, if the neural net is supposed to do a regression, then a possible loss function is the mean squared error. To adjust the weights so that the loss function decreases, an optimizer is used. This is an algorithm which is usually a variation of the Stochastic Gradient Descent (SGD). The weights of iteration $t+1$, with a loss function J are

$$w^{t+1} = w^t - \eta \cdot \nabla J(w^t) \tag{A.4}$$

where η is the learning rate, which controls how much the weights are adjusted in each pass. This hyperparameter is intrumental in making sure that the weights do converge to a minimum (whether global or local). So that the weights converge to the global minimum

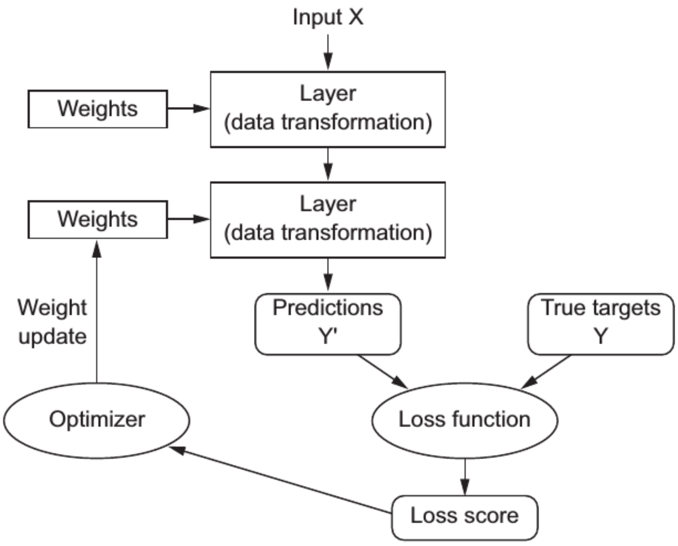
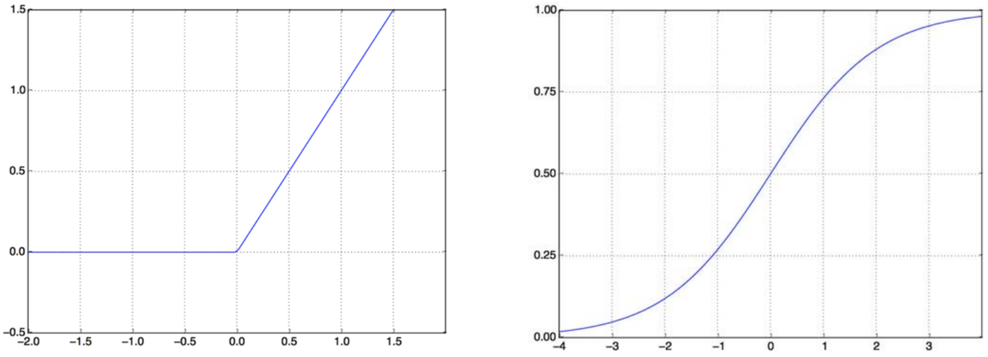


Fig. A.3 Schematic representation of a typical structure of a neural net.



(a) Rectified Linear Unit (ReLU) function

(b) Sigmoid function

Fig. A.4 Activation functions used in the analysis.

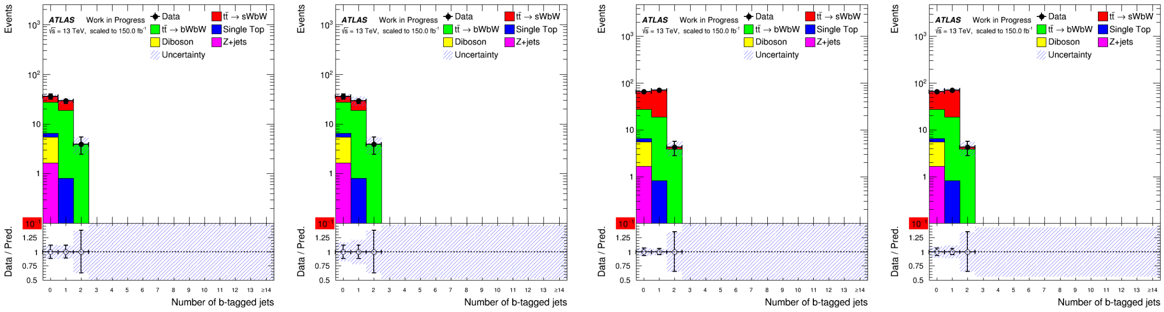
and not a local one, another hyperparameter, the momentum, is used. This can be seen as an equivalent to the physical momentum, that the ball of weights has and that allows it to overcome the bump after a local minimum, as it rolls down the gradient of the loss function (as akin to rolling down a hill). Some variations of SGD heuristically adapt these hyperparameters as the training progresses, for a faster convergence. The optimizer used in the analysis, Adam (for Adaptative Moment Estimation) does just that, and it is a common optimizer to use.

This optimization can be done in three different ways, depending on the number of output points that are used in each optimization. One way is to use all of them. This is called batch SGD. Each update using this process is more accurate. However, it is far more computationally intensive. At the other end of the spectrum is true SGD, where a single point is used in each optimization. This switches the advantage and disadvantage of the batch SGD. The efficient compromise between these two extremes is to use mini-batch SGD, where a reasonable amount of points is used in each optimization. A different, random, batch is used each time (hence the Stochastic), until no points are left to be used. This then concludes the training epoch. Normally, a neural network is trained over some tens of epochs.

Appendix B

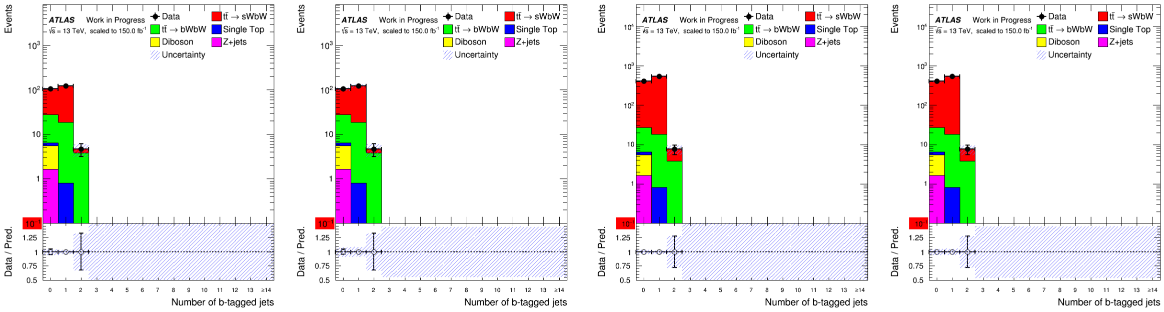
Linearity study pre- and post-fit histograms

B.1 Asimov data and simulated samples with b-tagging efficiency of 77 %



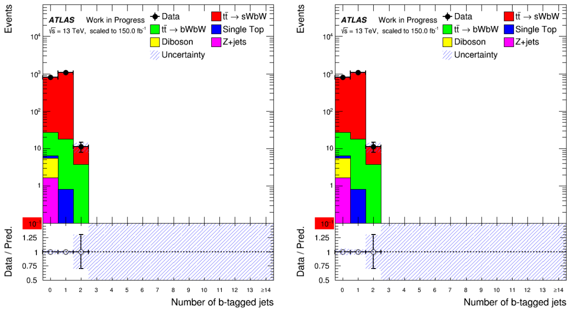
(a) Signal scale factor: 0.1

(b) Signal scale factor: 0.5



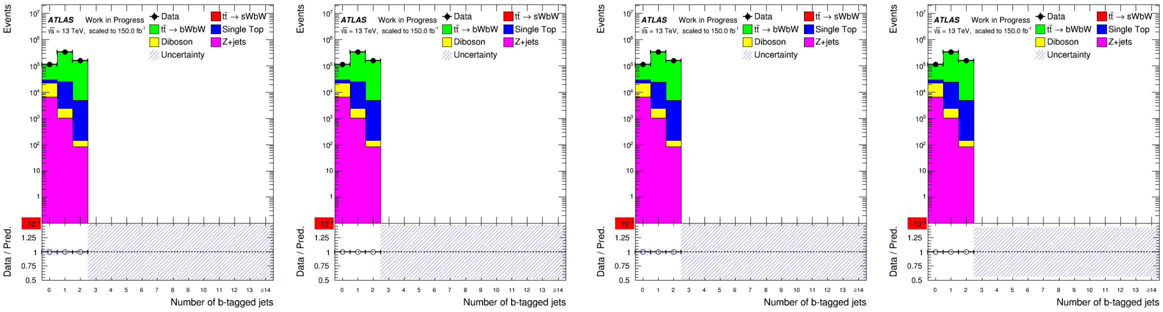
(c) Signal scale factor: 1

(d) Signal scale factor: 5

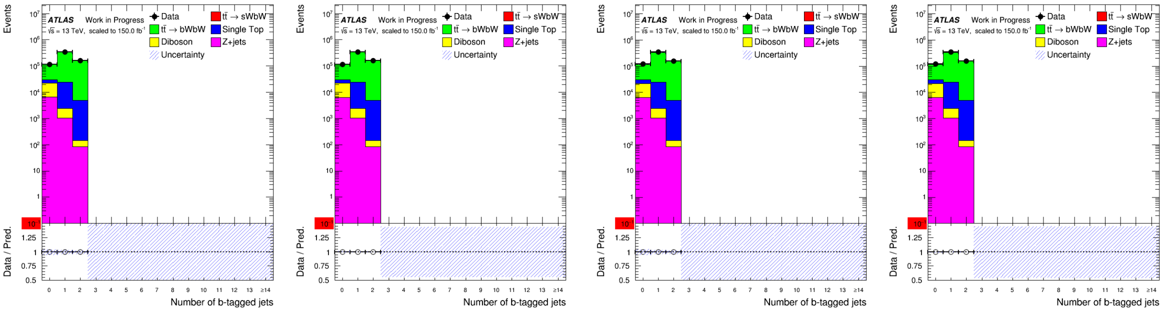


(e) Signal scale factor: 10

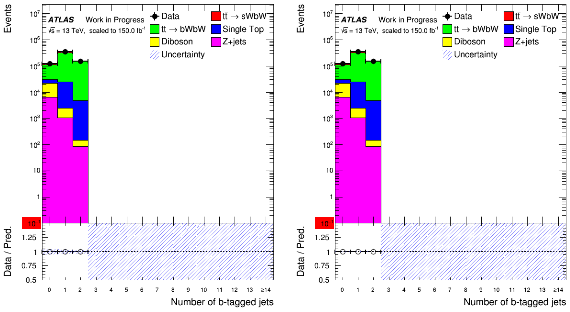
Fig. B.1 Number of b-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset and simulated samples with normal b-tagging efficiency, for different signal scale factors, in the SR.



(a) Signal scale factor: 0.1 (b) Signal scale factor: 0.5



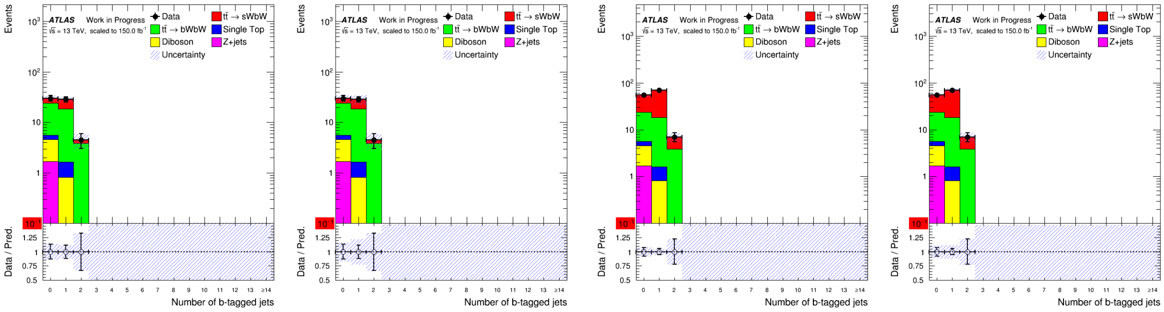
(c) Signal scale factor: 1 (d) Signal scale factor: 5



(e) Signal scale factor: 10

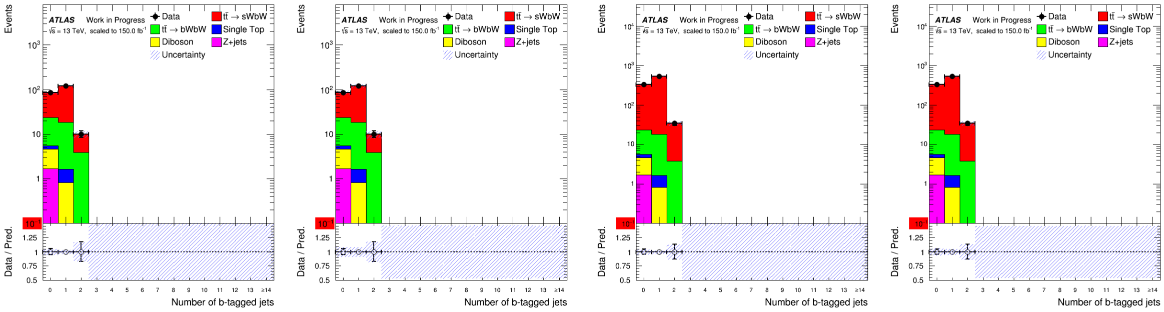
Fig. B.2 Number of *b*-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset and simulated samples with normal *b*-tagging efficiency, for different signal scale factors, in the $t\bar{t} \to bWbW$ CR.

B.2 Asimov data and simulated samples with +10 % b-tagging efficiency



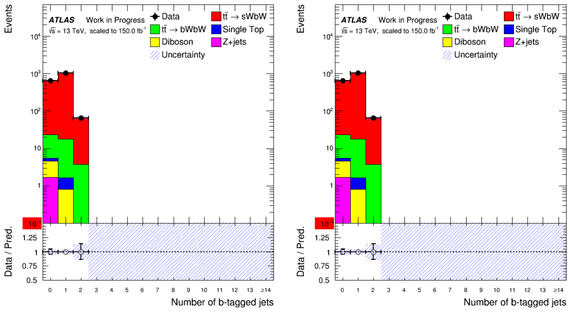
(a) Signal scale factor: 0.1

(b) Signal scale factor: 0.5



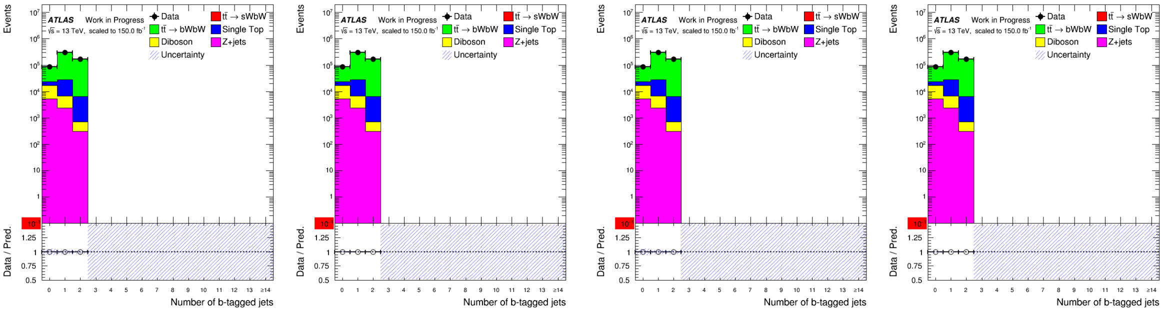
(c) Signal scale factor: 1

(d) Signal scale factor: 5



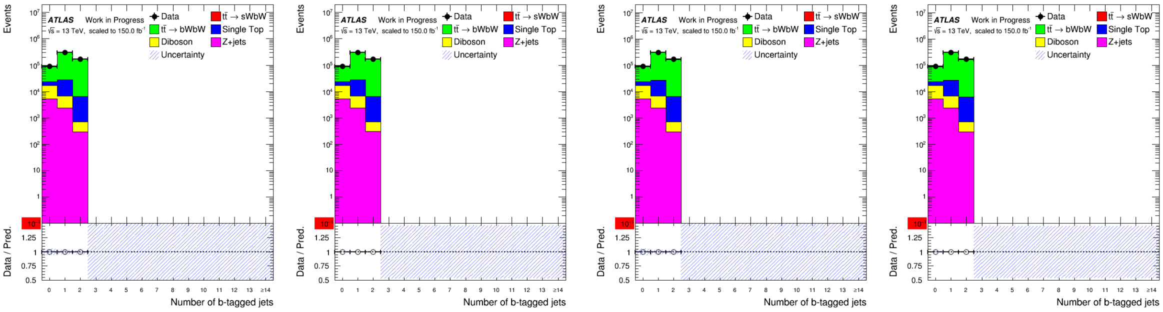
(e) Signal scale factor: 10

Fig. B.3 Number of b-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset and simulated samples with +10% b-tagging efficiency, for different signal scale factors, in the SR.



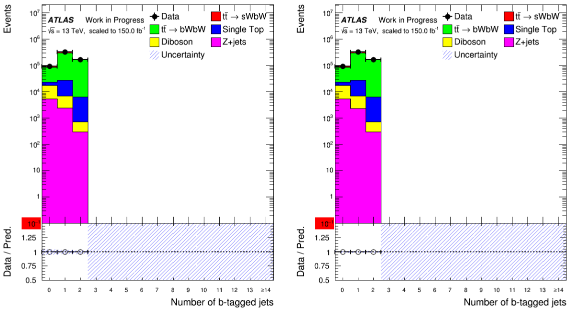
(a) Signal scale factor: 0.1

(b) Signal scale factor: 0.5



(c) Signal scale factor: 1

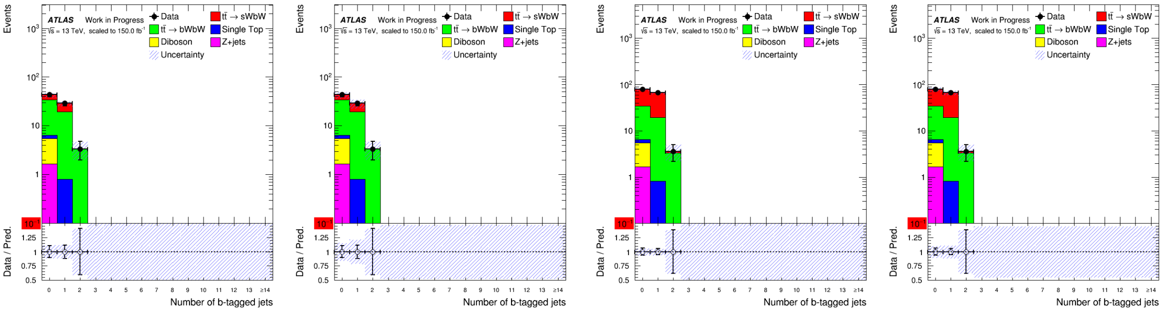
(d) Signal scale factor: 5



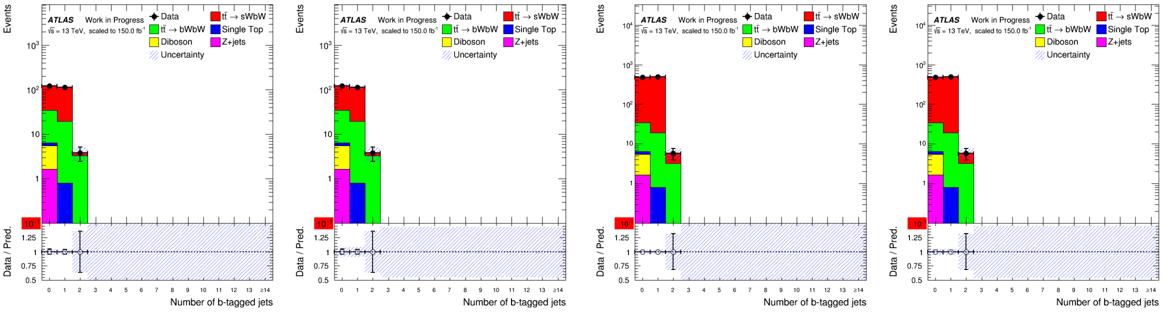
(e) Signal scale factor: 10

Fig. B.4 Number of *b*-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset and simulated samples with +10% *b*-tagging efficiency, for different signal scale factors, in the $t\bar{t} \rightarrow bWbW$ CR.

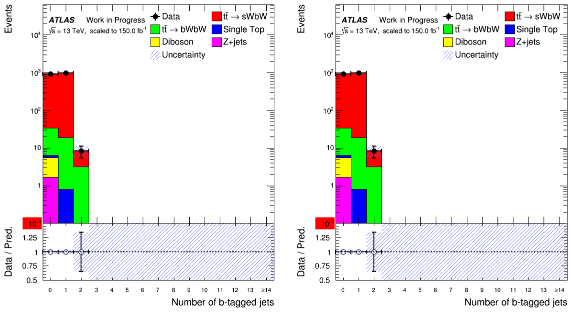
B.3 Asimov data and simulated samples with -10% b-tagging efficiency



(a) Signal scale factor: 0.1 (b) Signal scale factor: 0.5

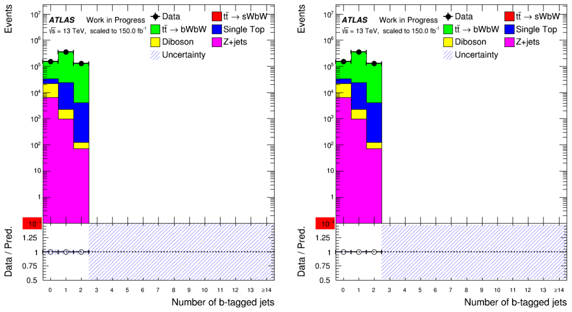


(c) Signal scale factor: 1 (d) Signal scale factor: 5

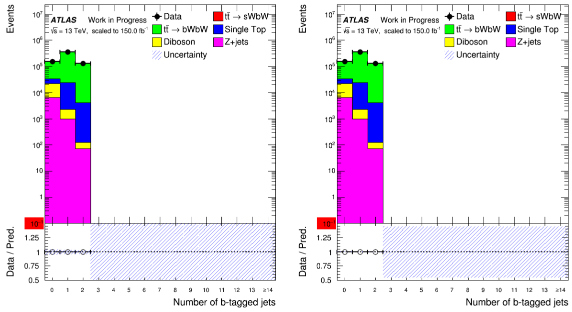


(e) Signal scale factor: 10

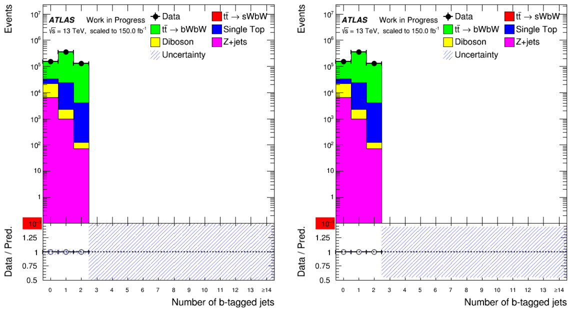
Fig. B.5 Number of b-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset and simulated samples with -10% b-tagging efficiency, for different signal scale factors, in the SR.



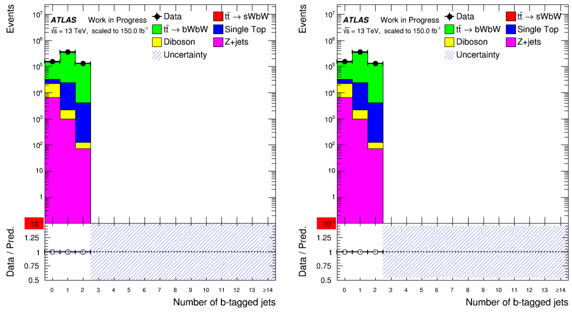
(a) Signal scale factor: 0.1



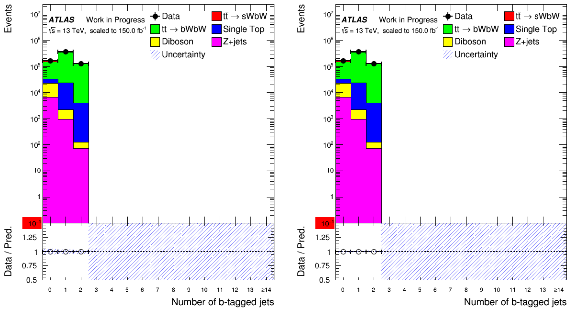
(b) Signal scale factor: 0.5



(c) Signal scale factor: 1



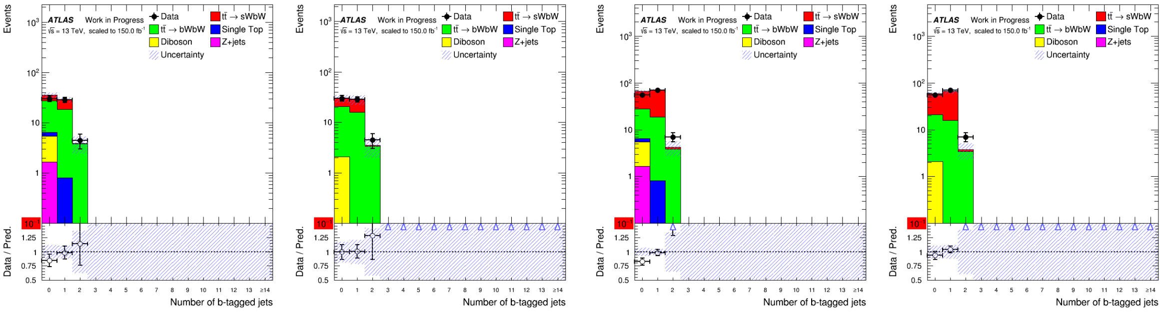
(d) Signal scale factor: 5



(e) Signal scale factor: 10

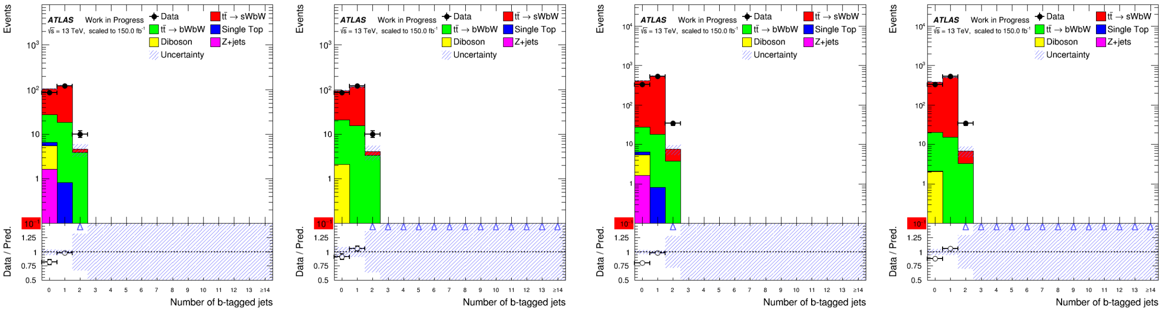
Fig. B.6 Number of *b*-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset and simulated samples with -10% *b*-tagging efficiency, for different signal scale factors, in the $t\bar{t} \rightarrow bWbW$ CR.

B.4 Asimov data with +10% b-tagging efficiency and simulated samples with normal b-tagging efficiency



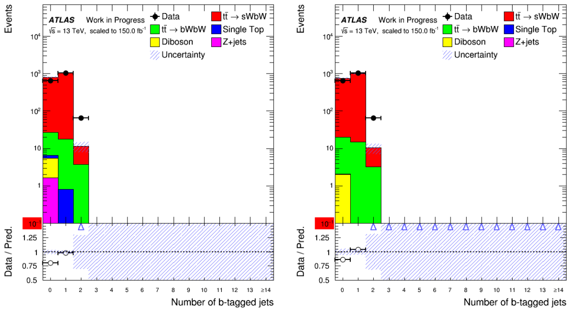
(a) Signal scale factor: 0.1

(b) Signal scale factor: 0.5



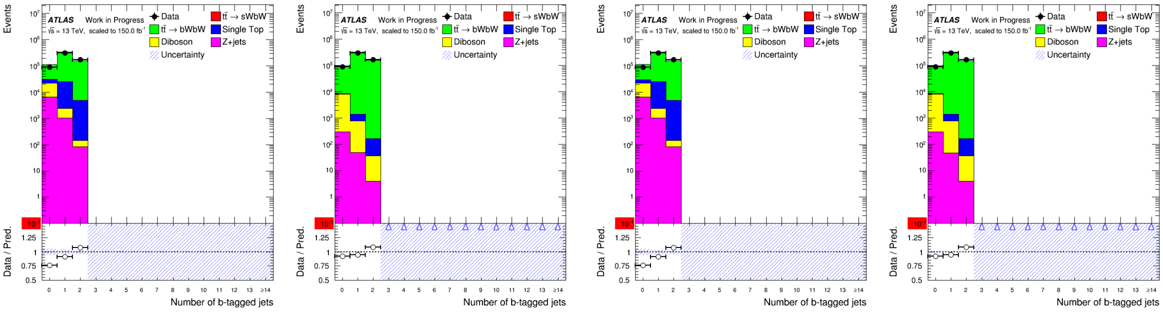
(c) Signal scale factor: 1

(d) Signal scale factor: 5



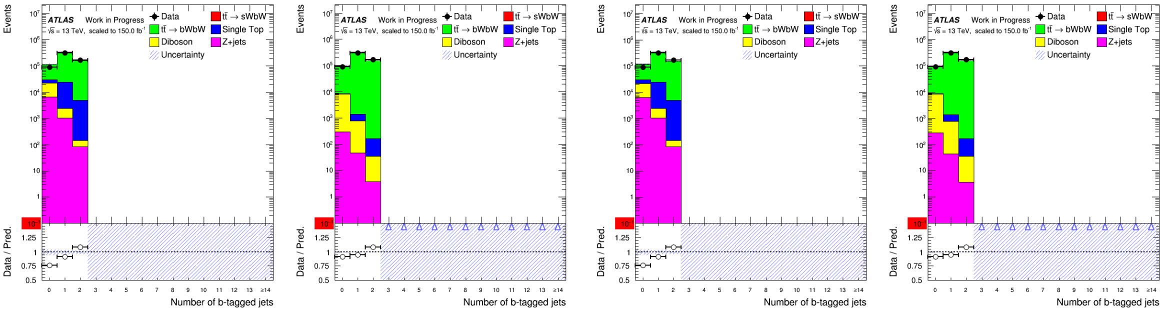
(e) Signal scale factor: 10

Fig. B.7 Number of b-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset with +10% b-tagging efficiency and simulated samples with normal b-tagging, for different signal scale factors, in the SR.



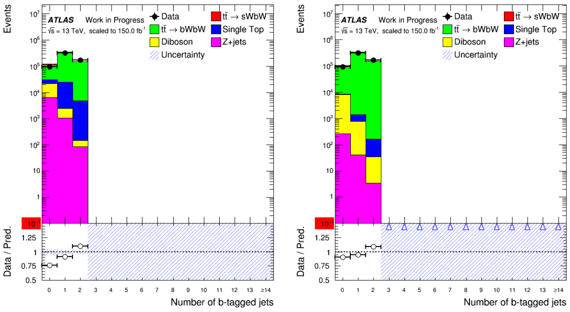
(a) Signal scale factor: 0.1

(b) Signal scale factor: 0.5



(c) Signal scale factor: 1

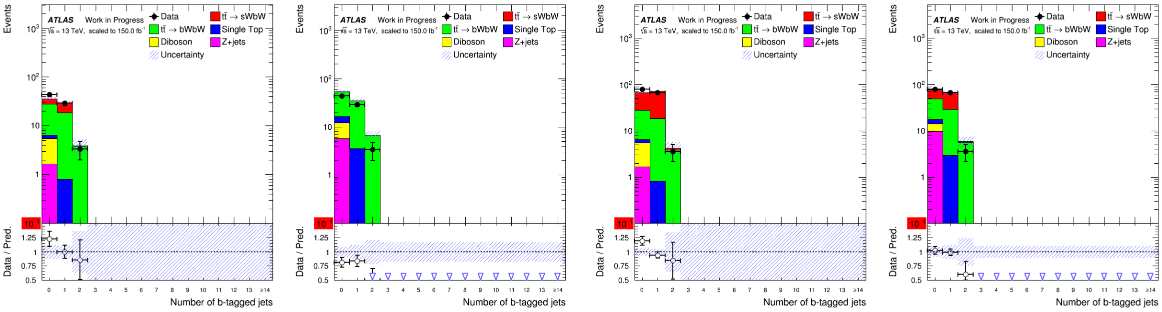
(d) Signal scale factor: 5



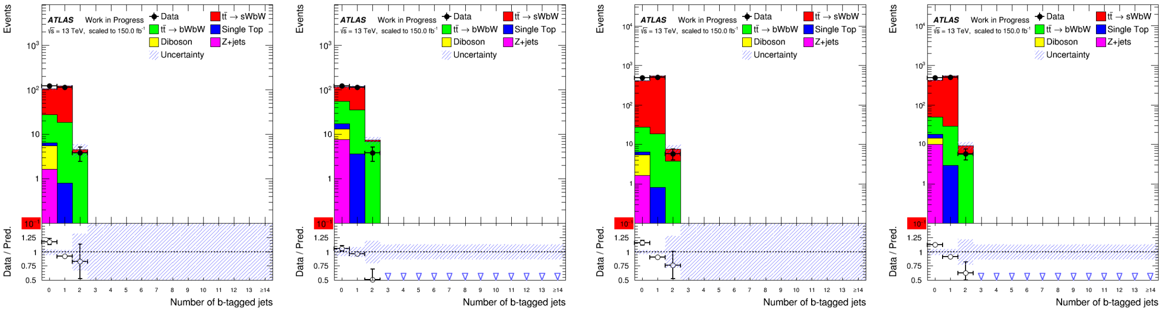
(e) Signal scale factor: 10

Fig. B.8 Number of *b*-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset with +10% *b*-tagging efficiency and simulated samples with normal *b*-tagging, for different signal scale factors, in the $t\bar{t} \rightarrow bWbW$ CR.

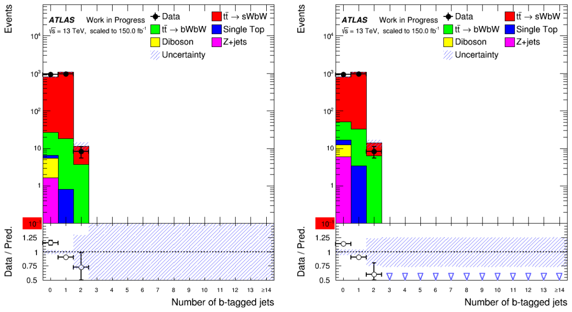
B.5 Asimov data with -10% b-tagging efficiency and simulated samples with normal b-tagging efficiency



(a) Signal scale factor: 0.1 (b) Signal scale factor: 0.5

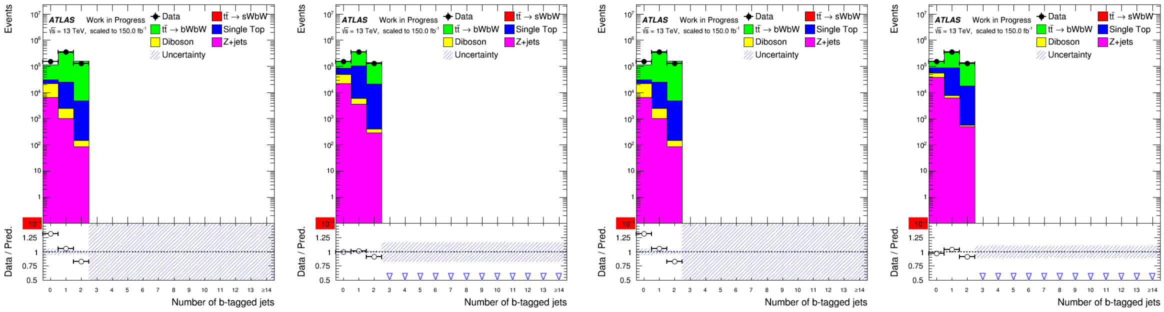


(c) Signal scale factor: 1 (d) Signal scale factor: 5



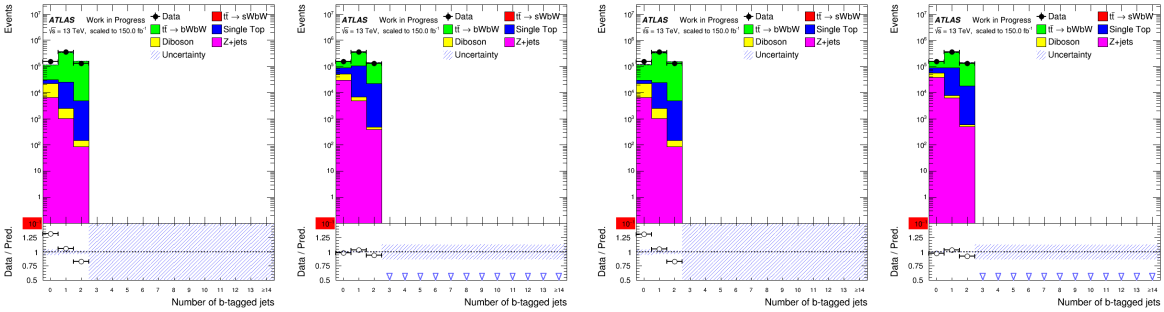
(e) Signal scale factor: 10

Fig. B.9 Number of b-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset with -10% b-tagging efficiency and simulated samples with normal b-tagging, for different signal scale factors, in the SR.



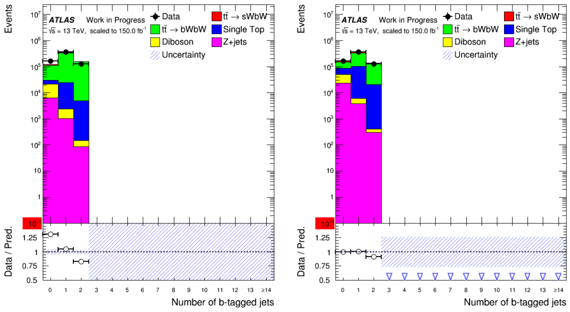
(a) Signal scale factor: 0.1

(b) Signal scale factor: 0.5



(c) Signal scale factor: 1

(d) Signal scale factor: 5



(e) Signal scale factor: 10

Fig. B.10 Number of *b*-tagged jets pre-fit (left) and post-fit (right), for Asimov dataset with -10% *b*-tagging efficiency and simulated samples with normal *b*-tagging, for different signal scale factors, in the $t\bar{t} \rightarrow bWbW$ CR.

Appendix C

Simulated datasets

All simulated samples used in this analysis are listed below. The signal samples are detailed in Table C.1 and all nominal background samples in Tables C.2 to C.5.

All samples are from mc16a campaign with the `DAOD_TOPQ1` derivation and tag `p3404` (`p3390` for the $t\bar{t} \to bWbW$ sample).

Table C.1 Signal samples

Name	σ [pb]	k-factor
410392.PowhegPythia8EvtGen_A14_ttbar_anti_s_hdamp258p75_nonallhad	379.242475	1.0
410393.PowhegPythia8EvtGen_A14_ttbar_s_hdamp258p75_nonallhad	379.242475	1.0

Table C.2 $t\bar{t} \to bWbW$ samples

Name	σ [pb]	k-factor
410503.PowhegPythia8EvtGen_A14_ttbar_hdamp258p75_dil	76.93	1.1392

Table C.3 Single top samples

Name	σ [pb]	k-factor
410015.PowhegPythiaEvtGen_P2012_Wt_dilepton_top	3.5835	1.054
410016.PowhegPythiaEvtGen_P2012_Wt_dilepton_antitop	3.5814	1.054
410560.MadGraphPythia8EvtGen_A14_tZ_4fl_tchan_noAllHad	0.24037	1.0

Table C.4 Diboson samples

Name	σ [pb]	k-factor
363356.Sherpa_221_NNPDF30NNLO_ZqqZll	15.563	0.13961
363357.Sherpa_221_NNPDF30NNLO_WqqZvv	6.7973	1.0
363358.Sherpa_221_NNPDF30NNLO_WqqZll	3.437	1.0
363359.Sherpa_221_NNPDF30NNLO_WpqqWmlv	24.717	1.0
363489.Sherpa_221_NNPDF30NNLO_WlvZqq	11.413	1.0
363494.Sherpa_221_NNPDF30NNLO_vvvv	0.60154	1.0
364250.Sherpa_222_NNPDF30NNLO_llll	1.2523	1.0
364254.Sherpa_222_NNPDF30NNLO_llvv	12.501	1.0
364255.Sherpa_222_NNPDF30NNLO_lvvv	3.2327	1.0

Table C.5 Z+jets samples

Name	σ [pb]	k-factor
364100.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV0_70_CVetoBVeto	1630.2243	0.9751
364101.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV0_70_CFilterBVeto	223.717472	0.9751
364102.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV0_70_BFilter	127.1799342	0.9751
364103.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV70_140_CVetoBVeto	75.0164716	0.9751
364104.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV70_140_CFilterBVeto	20.3477432	0.9751
364105.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV70_140_BFilter	12.3885125	0.9751
364106.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV140_280_CVetoBVeto	24.28530322	0.9751
364107.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV140_280_CFilterBVeto	9.2754186	0.9751
364108.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV140_280_BFilter	6.01361075	0.9751
364109.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV280_500_CVetoBVeto	4.77297475	0.9751
364110.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV280_500_CFilterBVeto	2.265570784	0.9751
364111.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV280_500_BFilter	1.491320988	0.9751
364112.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV500_1000	1.7881	0.9751
364113.Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV1000_E_CMS	0.14769	0.9751
364114.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV0_70_CVetoBVeto	1627.176708	0.9751
364115.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV0_70_CFilterBVeto	223.73136	0.9751
364116.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV0_70_BFilter	126.4502953	0.9751
364117.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV70_140_CVetoBVeto	76.292515	0.9751
364118.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV70_140_CFilterBVeto	20.3360066	0.9751
364119.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV70_140_BFilter	12.6227733	0.9751
364120.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV140_280_CVetoBVeto	25.03001412	0.9751
364121.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV140_280_CFilterBVeto	9.3719948	0.9751
364122.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV140_280_BFilter	6.08263138	0.9751
364123.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV280_500_CVetoBVeto	4.869231562	0.9751
364124.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV280_500_CFilterBVeto	2.279979034	0.9751
364125.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV280_500_BFilter	1.494370818	0.9751
364126.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV500_1000	1.8081	0.9751
364127.Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV1000_E_CMS	0.14857	0.9751
364128.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CVetoBVeto	1627.725872	0.9751
364129.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CFilterBVeto	223.881432	0.9751
364130.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_BFilter	127.7329554	0.9751
364131.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_CVetoBVeto	76.0261671	0.9751
364132.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_CFilterBVeto	20.212279	0.9751
364133.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_BFilter	12.29393	0.9751
364134.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CVetoBVeto	24.80341201	0.9751
364135.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CFilterBVeto	9.3282378	0.9751
364136.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_BFilter	5.47909362	0.9751
364137.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_CVetoBVeto	4.791190072	0.9751
364138.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_CFilterBVeto	2.275625215	0.9751
364139.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_BFilter	1.502837652	0.9751
364140.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV500_1000	1.8096	0.9751
364141.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV1000_E_CMS	0.14834	0.9751

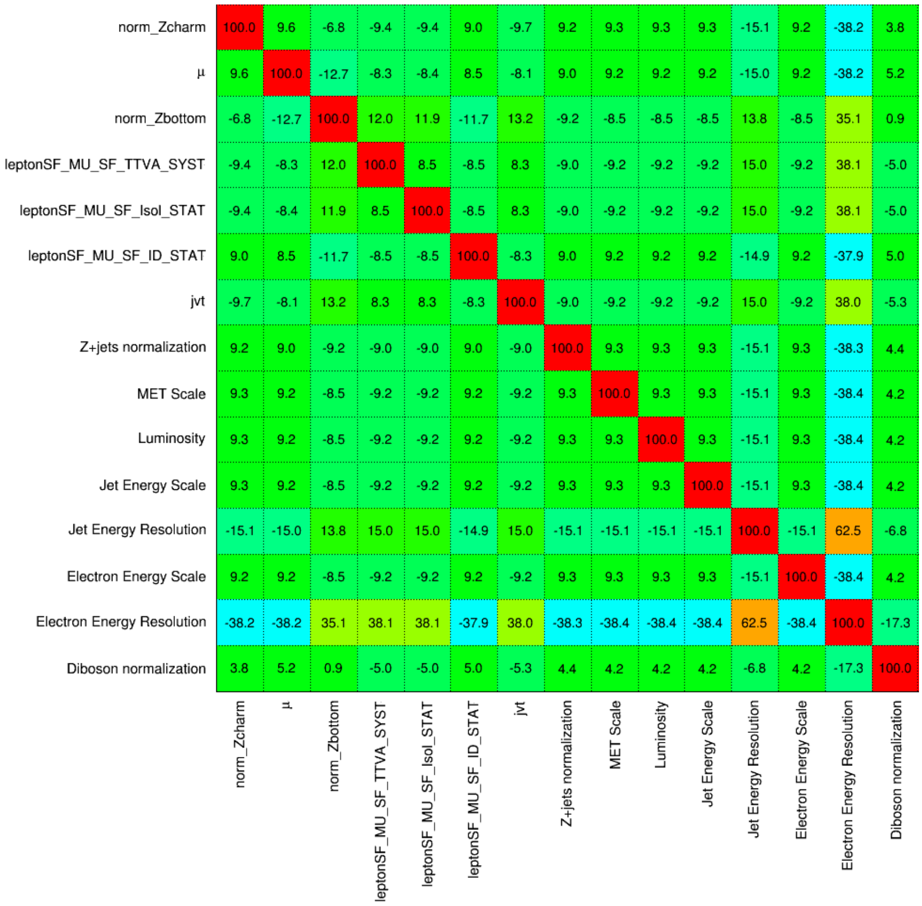
Appendix D

`TRexFitter` fit output

D.1 Z+jets VR with ee fit

Table D.1 Yields of Z+jets VR with ee (statistical uncertainty only).

Sample	Yields	
	Pre-fit	Post-fit
$t\bar{t} \to sWbW$	116.9 ± 3.9	116.9 ± 3.9
$t\bar{t} \to bWbW$	13021 ± 424	15218 ± 110
Single Top	982 ± 347	1189 ± 247
Diboson	12489 ± 1312	14627 ± 1040
Z+light	$2.20 \times 10^6 \pm 1.14 \times 10^6$	$2.13 \times 10^6 \pm 11\,170$
Z+charm	478239 ± 240785	709008 ± 5439
Z+bottom	206843 ± 104661	263145 ± 15692
Total	$2.91 \times 10^6 \pm 1.48 \times 10^6$	$3.13 \times 10^6 \pm 20111$
Data	3.14×10^6	



	norm_Zcharm	μ	norm_Zbottom	leptonSF_MU_SF_TTVA_SYST	leptonSF_MU_SF_Isol_STAT	leptonSF_MU_SF_ID_STAT	jvt	Z+jets normalization	MET Scale	Luminosity	Jet Energy Scale	Jet Energy Resolution	Electron Energy Scale	Electron Energy Resolution	Diboson normalization
norm_Zcharm	100.0	9.6	-6.8	-9.4	-9.4	9.0	-9.7	9.2	9.3	9.3	9.3	-15.1	9.2	-38.2	3.8
μ	9.6	100.0	-12.7	-8.3	-8.4	8.5	-8.1	9.0	9.2	9.2	9.2	-15.0	9.2	-38.2	5.2
norm_Zbottom	-6.8	-12.7	100.0	12.0	11.9	-11.7	13.2	-9.2	-8.5	-8.5	-8.5	13.8	-8.5	35.1	0.9
leptonSF_MU_SF_TTVA_SYST	-9.4	-8.3	12.0	100.0	8.5	-8.5	8.3	-9.0	-9.2	-9.2	-9.2	15.0	-9.2	38.1	-5.0
leptonSF_MU_SF_Isol_STAT	-9.4	-8.4	11.9	8.5	100.0	-8.5	8.3	-9.0	-9.2	-9.2	-9.2	15.0	-9.2	38.1	-5.0
leptonSF_MU_SF_ID_STAT	9.0	8.5	-11.7	-8.5	-8.5	100.0	-8.3	9.0	9.2	9.2	9.2	-14.9	9.2	-37.9	5.0
jvt	-9.7	-8.1	13.2	8.3	8.3	-8.3	100.0	-9.0	-9.2	-9.2	-9.2	15.0	-9.2	38.0	-5.3
Z+jets normalization	9.2	9.0	-9.2	-9.0	-9.0	9.0	-9.0	100.0	9.3	9.3	9.3	-15.1	9.3	-38.3	4.4
MET Scale	9.3	9.2	-8.5	-9.2	-9.2	9.2	-9.2	9.3	100.0	9.3	9.3	-15.1	9.3	-38.4	4.2
Luminosity	9.3	9.2	-8.5	-9.2	-9.2	9.2	-9.2	9.3	9.3	100.0	9.3	-15.1	9.3	-38.4	4.2
Jet Energy Scale	9.3	9.2	-8.5	-9.2	-9.2	9.2	-9.2	9.3	9.3	9.3	100.0	-15.1	9.3	-38.4	4.2
Jet Energy Resolution	-15.1	-15.0	13.8	15.0	15.0	-14.9	15.0	-15.1	-15.1	-15.1	-15.1	100.0	-15.1	62.5	-6.8
Electron Energy Scale	9.2	9.2	-8.5	-9.2	-9.2	9.2	-9.2	9.3	9.3	9.3	9.3	-15.1	100.0	-38.4	4.2
Electron Energy Resolution	-38.2	-38.2	35.1	38.1	38.1	-37.9	38.0	-38.3	-38.4	-38.4	-38.4	62.5	-38.4	100.0	-17.3
Diboson normalization	3.8	5.2	0.9	-5.0	-5.0	5.0	-5.3	4.4	4.2	4.2	4.2	-6.8	4.2	-17.3	100.0

Fig. D.1 Correlation plot for Z+jets VR with ee (only with correlations greater than 15%).

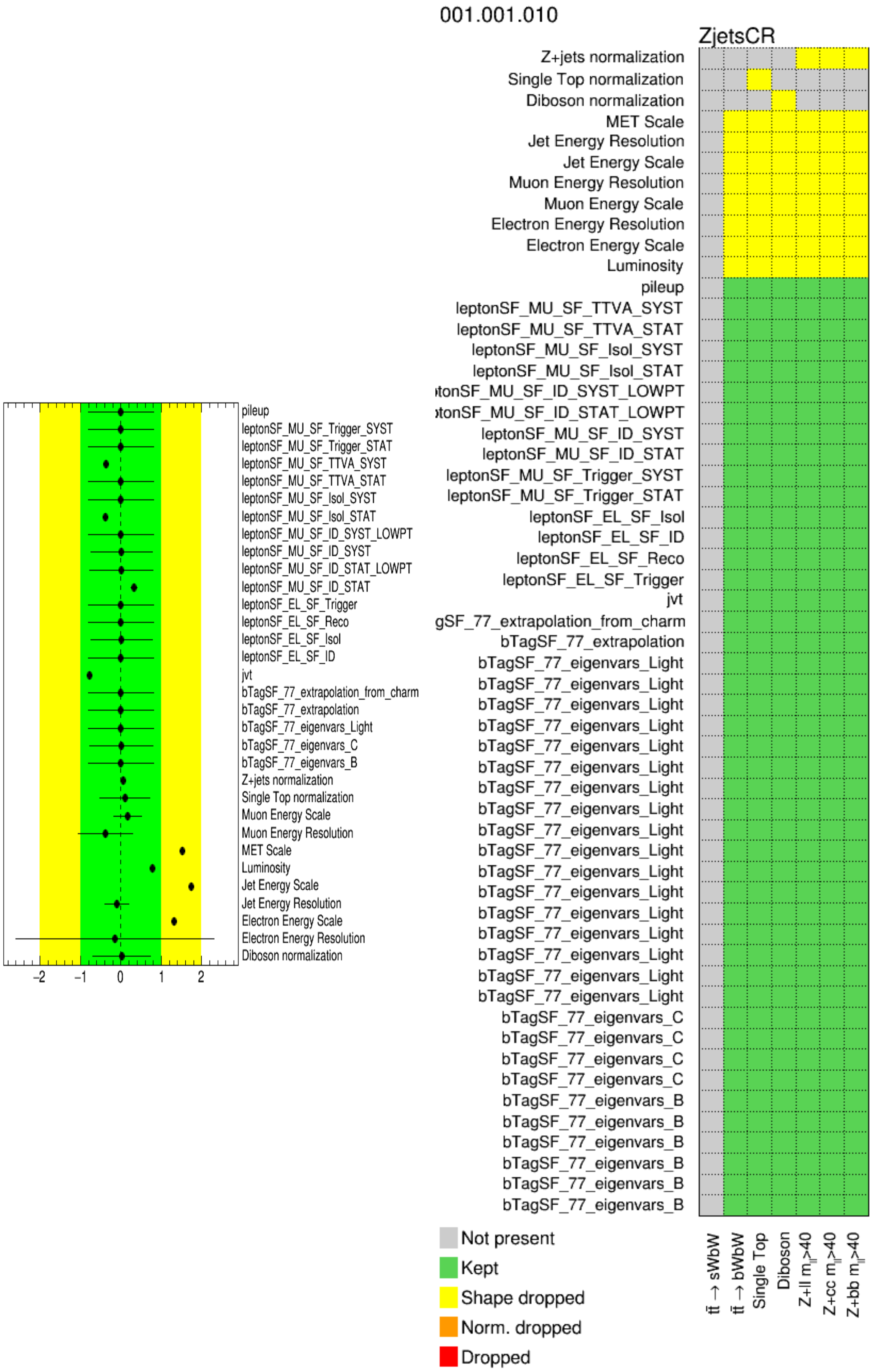
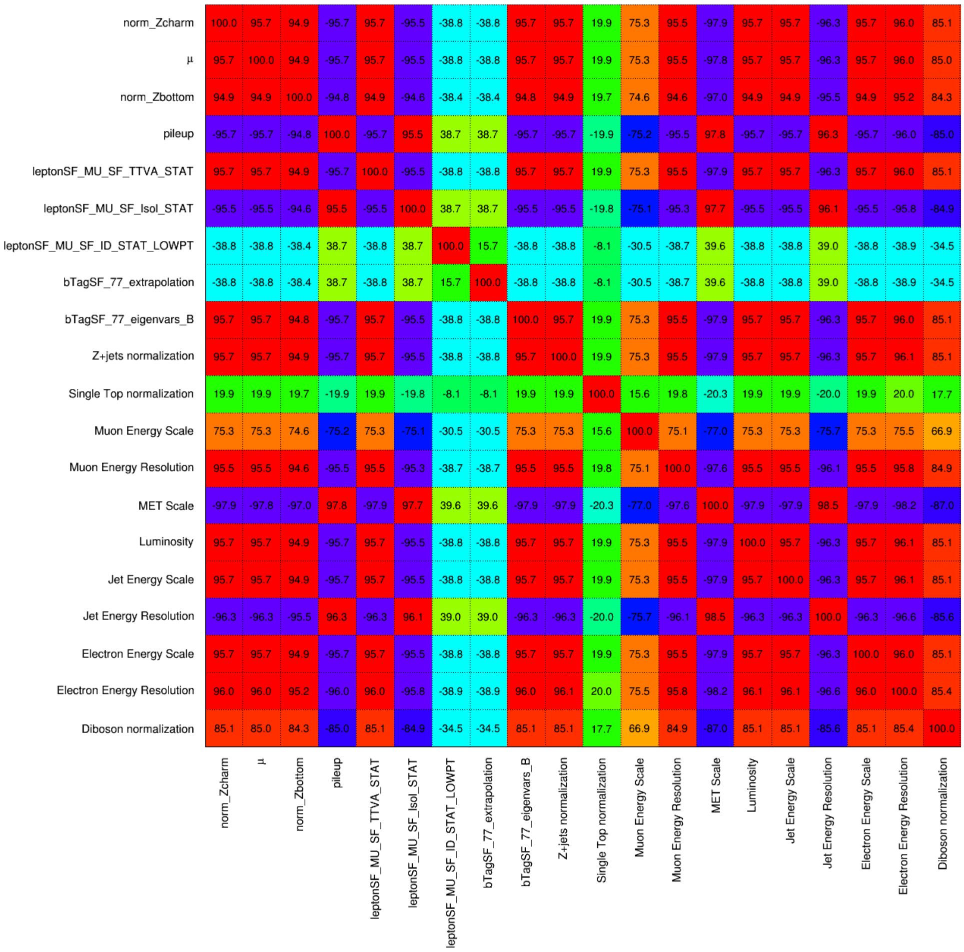


Fig. D.2 Pull plot (left) and pruning plot (right) for *Z*+jets VR with *ee*.

D.2 Z+jets VR with $\mu\mu$ fit

Table D.2 Yields of Z+jets VR with $\mu\mu$ (statistical uncertainty only).

Sample	Yields	
	Pre-fit	Post-fit
$t\bar{t} \to sWbW$	119.1 ± 3.9	119.1 ± 3.9
$t\bar{t} \to bWbW$	13936 ± 454	15184 ± 98
Single Top	1105 ± 390	1289 ± 40
Diboson	14453 ± 1518	15772 ± 246
Z+light	$2.63 \times 10^6 \pm 1.32 \times 10^6$	$3.42 \times 10^6 \pm 31705$
Z+charm	566062 ± 313842	529900 ± 6648
Z+bottom	261975 ± 338443	600690 ± 12041
Total	$3.49 \times 10^6 \pm 1.74 \times 10^6$	$4.58 \times 10^6 \pm 35733.7$
Data	4.60×10^6	



	norm_Zcharm	µ	norm_Zbottom	pileup	leptonSF_MU_SF_TTVA_STAT	leptonSF_MU_SF_Isol_STAT	leptonSF_MU_SF_ID_STAT_LOWPT	bTagSF_77_extrapolation	bTagSF_77_eigenvars_B	Z+jets normalization	Single Top normalization	Muon Energy Scale	Muon Energy Resolution	MET Scale	Luminosity	Jet Energy Scale	Jet Energy Resolution	Electron Energy Scale	Electron Energy Resolution	Diboson normalization
norm_Zcharm	100.0	95.7	94.9	-95.7	95.7	-95.5	-38.8	-38.8	95.7	95.7	19.9	75.3	95.5	-97.9	95.7	95.7	-96.3	95.7	96.0	85.1
µ	95.7	100.0	94.9	-95.7	95.7	-95.5	-38.8	-38.8	95.7	95.7	19.9	75.3	95.5	-97.8	95.7	95.7	-96.3	95.7	96.0	85.0
norm_Zbottom	94.9	94.9	100.0	-94.8	94.9	-94.6	-38.4	-38.4	94.8	94.9	19.7	74.6	94.6	-97.0	94.9	94.9	-95.5	94.9	95.2	84.3
pileup	-95.7	-95.7	-94.8	100.0	-95.7	95.5	38.7	38.7	-95.7	-95.7	-19.9	-75.2	-95.5	97.8	-95.7	-95.7	96.3	-95.7	-96.0	-85.0
leptonSF_MU_SF_TTVA_STAT	95.7	95.7	94.9	-95.7	100.0	-95.5	-38.8	-38.8	95.7	95.7	19.9	75.3	95.5	-97.9	95.7	95.7	-96.3	95.7	96.0	85.1
leptonSF_MU_SF_Isol_STAT	-95.5	-95.5	-94.6	95.5	-95.5	100.0	38.7	38.7	-95.5	-95.5	-19.8	-75.1	-95.3	97.7	-95.5	-95.5	96.1	-95.5	-95.8	-84.9
leptonSF_MU_SF_ID_STAT_LOWPT	-38.8	-38.8	-38.4	38.7	-38.8	38.7	100.0	15.7	-38.8	-38.8	-8.1	-30.5	-38.7	39.6	-38.8	-38.8	39.0	-38.8	-38.9	-34.5
bTagSF_77_extrapolation	-38.8	-38.8	-38.4	38.7	-38.8	38.7	15.7	100.0	-38.8	-38.8	-8.1	-30.5	-38.7	39.6	-38.8	-38.8	39.0	-38.8	-38.9	-34.5
bTagSF_77_eigenvars_B	95.7	95.7	94.8	-95.7	95.7	-95.5	-38.8	-38.8	100.0	95.7	19.9	75.3	95.5	-97.9	95.7	95.7	-96.3	95.7	96.0	85.1
Z+jets normalization	95.7	95.7	94.9	-95.7	95.7	-95.5	-38.8	-38.8	95.7	100.0	19.9	75.3	95.5	-97.9	95.7	95.7	-96.3	95.7	96.1	85.1
Single Top normalization	19.9	19.9	19.7	-19.9	19.9	-19.8	-8.1	-8.1	19.9	19.9	100.0	15.6	19.8	-20.3	19.9	19.9	-20.0	19.9	20.0	17.7
Muon Energy Scale	75.3	75.3	74.6	-75.2	75.3	-75.1	-30.5	-30.5	75.3	75.3	15.6	100.0	75.1	-77.0	75.3	75.3	-75.7	75.3	75.5	66.9
Muon Energy Resolution	95.5	95.5	94.6	-95.5	95.5	-95.3	-38.7	-38.7	95.5	95.5	19.8	75.1	100.0	-97.6	95.5	95.5	-96.1	95.5	95.8	84.9
MET Scale	-97.9	-97.8	-97.0	97.8	-97.9	97.7	39.6	39.6	-97.9	-97.9	-20.3	-77.0	-97.6	100.0	-97.9	-97.9	98.5	-97.9	-98.2	-87.0
Luminosity	95.7	95.7	94.9	-95.7	95.7	-95.5	-38.8	-38.8	95.7	95.7	19.9	75.3	95.5	-97.9	100.0	95.7	-96.3	95.7	96.1	85.1
Jet Energy Scale	95.7	95.7	94.9	-95.7	95.7	-95.5	-38.8	-38.8	95.7	95.7	19.9	75.3	95.5	-97.9	95.7	100.0	-96.3	95.7	96.1	85.1
Jet Energy Resolution	-96.3	-96.3	-95.5	96.3	-96.3	96.1	39.0	39.0	-96.3	-96.3	-20.0	-75.7	-96.1	98.5	-96.3	-96.3	100.0	-96.3	-96.6	-85.6
Electron Energy Scale	95.7	95.7	94.9	-95.7	95.7	-95.5	-38.8	-38.8	95.7	95.7	19.9	75.3	95.5	-97.9	95.7	95.7	-96.3	100.0	96.0	85.1
Electron Energy Resolution	96.0	96.0	95.2	-96.0	96.0	-95.8	-38.9	-38.9	96.0	96.1	20.0	75.5	95.8	-98.2	96.1	96.1	-96.6	96.0	100.0	85.4
Diboson normalization	85.1	85.0	84.3	-85.0	85.1	-84.9	-34.5	-34.5	85.1	85.1	17.7	66.9	84.9	-87.0	85.1	85.1	-85.6	85.1	85.4	100.0

Fig. D.3 Correlation plot for Z+jets VR with $\mu\mu$ (only with correlations greater than 15%).

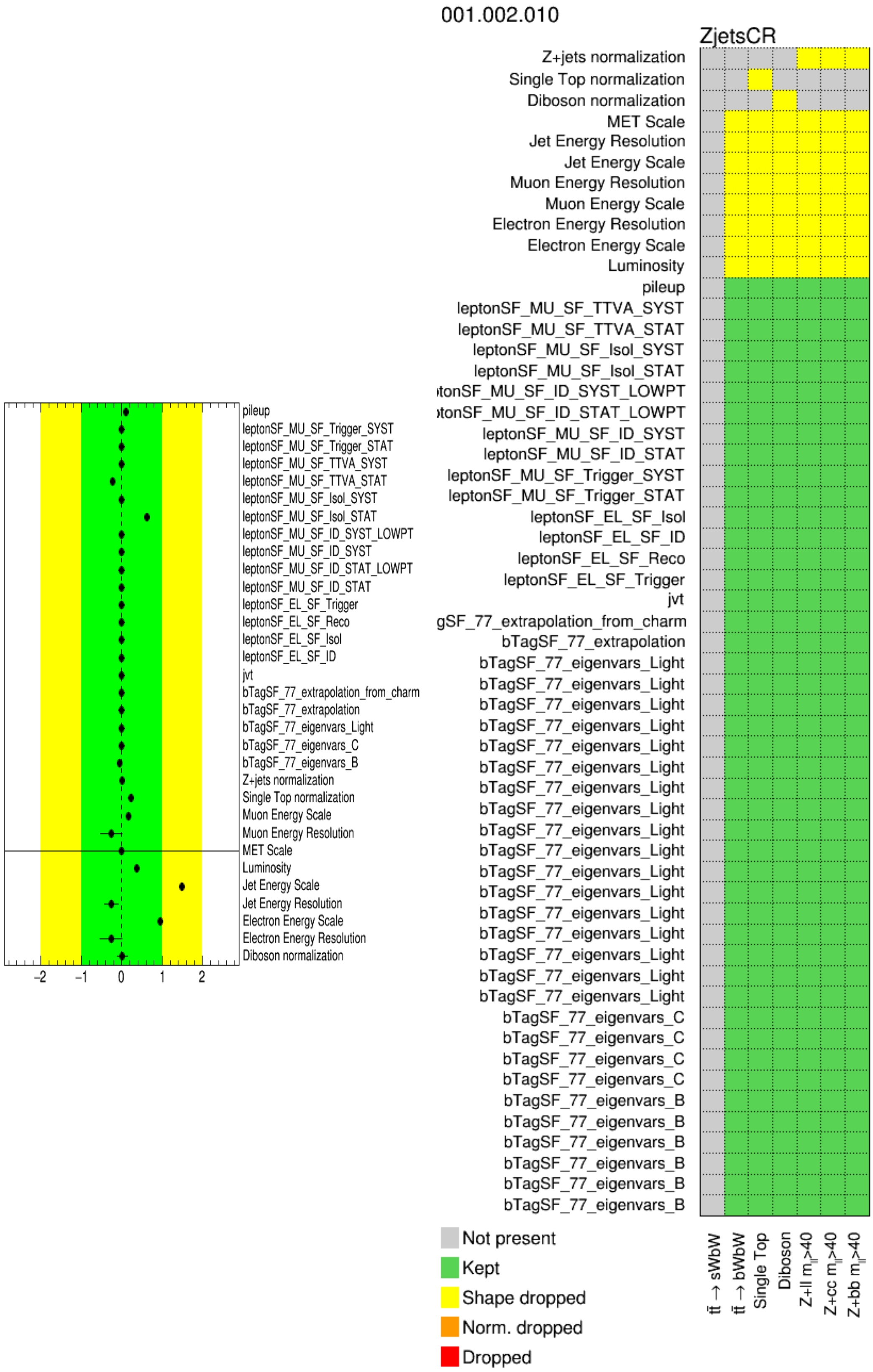
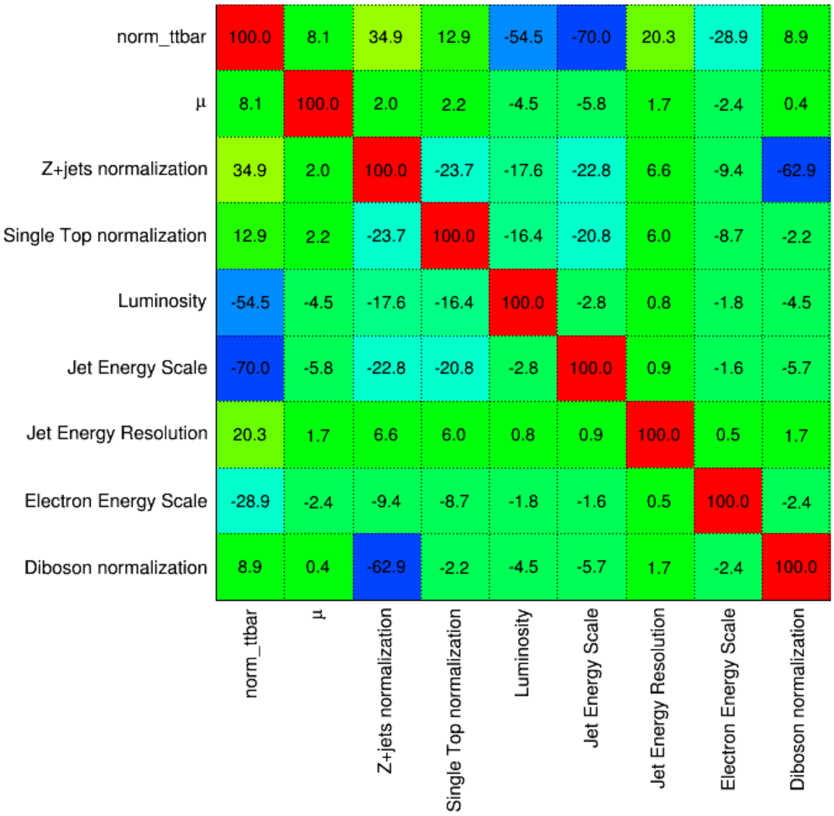


Fig. D.4 Pull plot (left) and pruning plot (right) for Z+jets VR with $\mu\mu$.

D.3 Linearity study fit with +10% *b*-tagging efficiency and $\alpha = 1$

Table D.3 Yields of both regions for linearity study fit with +10% b-tagging efficiency and $\alpha = 1$ (statistical uncertainty only).

Sample	Yields			
	SR		$t\bar{t} \to bWbW$ CR	
	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t} \to sWbW$	171.2 ± 4.7	171.2 ± 17.0	3394 ± 21	3394 ± 21
$t\bar{t} \to bWbW$	38.8 ± 5.1	38.8 ± 5.1	505852 ± 16103	505853 ± 3410
Single Top	1.79 ± 1.42	1.79 ± 1.29	33165 ± 11657	33156 ± 3813
Diboson	3.72 ± 0.94	3.72 ± 0.92	16746 ± 1758	16753 ± 1606
Z+light	0.78 ± 0.96	0.78 ± 0.90	4502 ± 2367	4505 ± 1401
Z+charm	0.90 ± 1.01	0.90 ± 0.93	1770 ± 988	1772 ± 649
Z+bottom	0 ± 0	0 ± 0	1740 ± 897	1741 ± 507
Total	217.2 ± 7.3	217.2 ± 17.7	567168 ± 21823	567174 ± 1732
Data	217		567168	



	norm_ttbar	μ	Z+jets normalization	Single Top normalization	Luminosity	Jet Energy Scale	Jet Energy Resolution	Electron Energy Scale	Diboson normalization
norm_ttbar	100.0	8.1	34.9	12.9	-54.5	-70.0	20.3	-28.9	8.9
μ	8.1	100.0	2.0	2.2	-4.5	-5.8	1.7	-2.4	0.4
Z+jets normalization	34.9	2.0	100.0	-23.7	-17.6	-22.8	6.6	-9.4	-62.9
Single Top normalization	12.9	2.2	-23.7	100.0	-16.4	-20.8	6.0	-8.7	-2.2
Luminosity	-54.5	-4.5	-17.6	-16.4	100.0	-2.8	0.8	-1.8	-4.5
Jet Energy Scale	-70.0	-5.8	-22.8	-20.8	-2.8	100.0	0.9	-1.6	-5.7
Jet Energy Resolution	20.3	1.7	6.6	6.0	0.8	0.9	100.0	0.5	1.7
Electron Energy Scale	-28.9	-2.4	-9.4	-8.7	-1.8	-1.6	0.5	100.0	-2.4
Diboson normalization	8.9	0.4	-62.9	-2.2	-4.5	-5.7	1.7	-2.4	100.0

Fig. D.5 Correlation plot for linearity study fit with +10% b-tagging efficiency and $\alpha = 1$ (only with correlations greater than 15%).

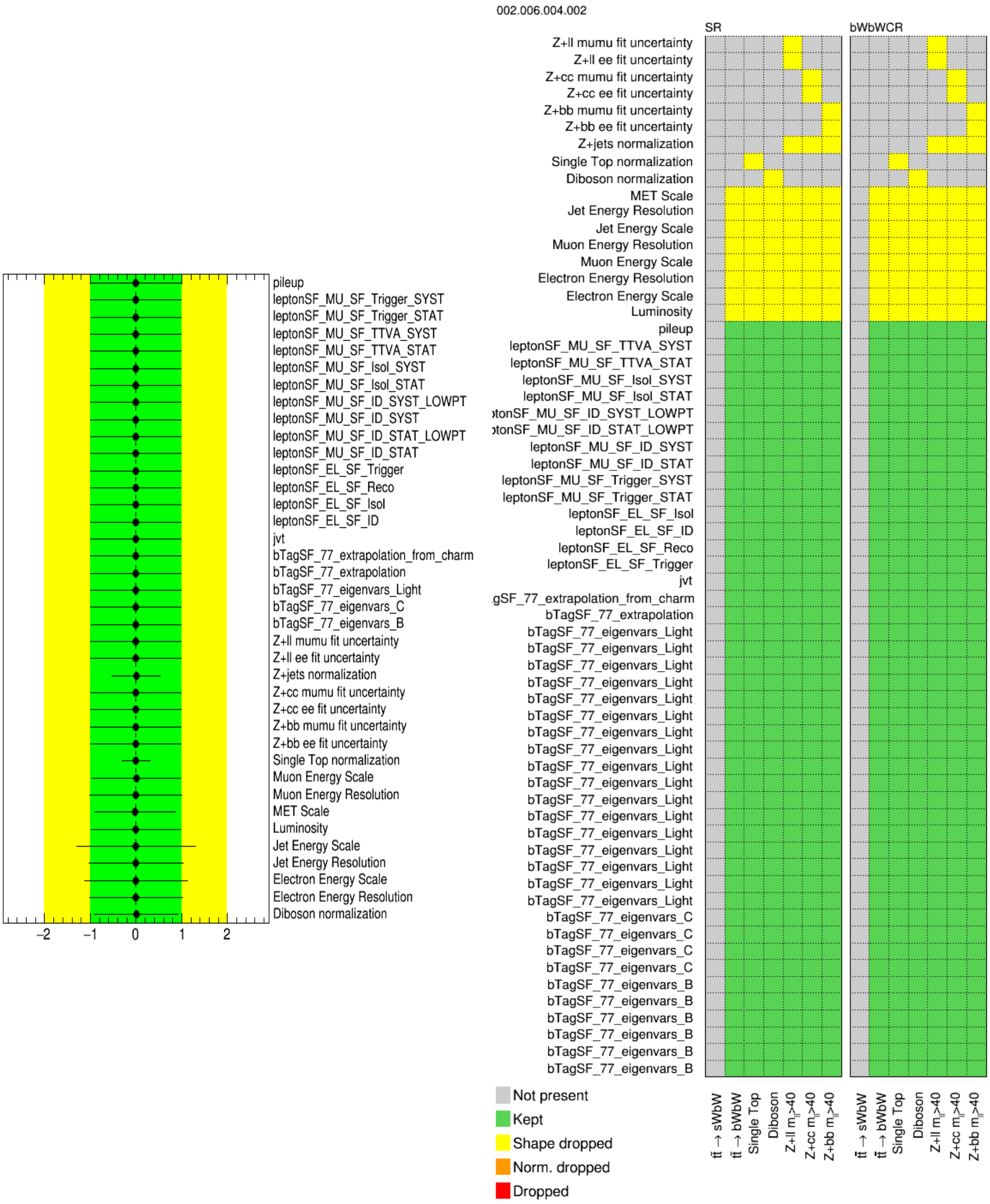
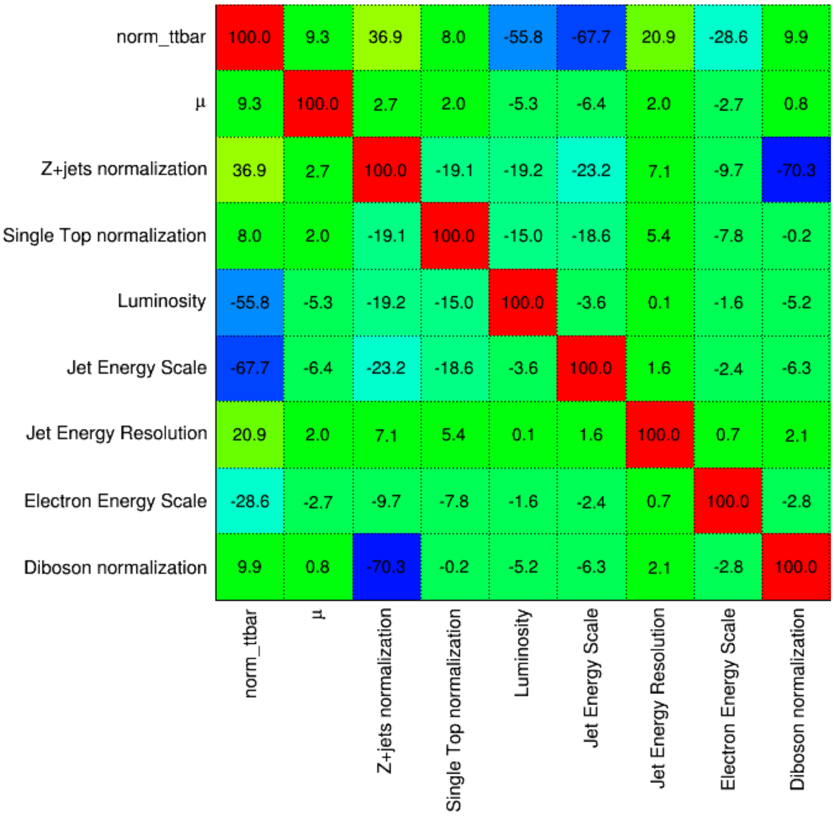


Fig. D.6 Pull plot (left) and pruning plot (right) for linearity study fit with +10% b-tagging efficiency and $\alpha = 1$.

D.4 Linearity study fit with -10% b-tagging efficiency and $\alpha = 1$

Table D.4 Yields of both regions for linearity study fit with -10% b-tagging efficiency and $\alpha = 1$ (statistical uncertainty only).

Sample	Yields			
	SR		$t\bar{t} \to bWbW$ CR	
	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t} \to sWbW$	184.7 ± 4.8	184.8 ± 17.9	3544 ± 21	3544 ± 21
$t\bar{t} \to bWbW$	49.9 ± 5.8	49.9 ± 5.8	575257 ± 18301	575262 ± 4122
Single Top	1.79 ± 1.42	1.79 ± 1.29	36353 ± 12778	36343 ± 4420
Diboson	3.84 ± 0.90	3.84 ± 0.88	16958 ± 1780	16962 ± 1595
Z+light	0.74 ± 0.82	0.74 ± 0.76	4312 ± 2367	4315 ± 1311
Z+charm	0.92 ± 1.03	0.92 ± 0.95	1819 ± 988	1820 ± 510
Z+bottom	0 ± 0	0 ± 0	1465 ± 897	1466 ± 422
Total	241.8 ± 7.9	241.9 ± 18.8	639708 ± 24284	639713 ± 1812
Data	242		639708	



	norm_ttbar	μ	Z+jets normalization	Single Top normalization	Luminosity	Jet Energy Scale	Jet Energy Resolution	Electron Energy Scale	Diboson normalization
norm_ttbar	100.0	9.3	36.9	8.0	-55.8	-67.7	20.9	-28.6	9.9
μ	9.3	100.0	2.7	2.0	-5.3	-6.4	2.0	-2.7	0.8
Z+jets normalization	36.9	2.7	100.0	-19.1	-19.2	-23.2	7.1	-9.7	-70.3
Single Top normalization	8.0	2.0	-19.1	100.0	-15.0	-18.6	5.4	-7.8	-0.2
Luminosity	-55.8	-5.3	-19.2	-15.0	100.0	-3.6	0.1	-1.6	-5.2
Jet Energy Scale	-67.7	-6.4	-23.2	-18.6	-3.6	100.0	1.6	-2.4	-6.3
Jet Energy Resolution	20.9	2.0	7.1	5.4	0.1	1.6	100.0	0.7	2.1
Electron Energy Scale	-28.6	-2.7	-9.7	-7.8	-1.6	-2.4	0.7	100.0	-2.8
Diboson normalization	9.9	0.8	-70.3	-0.2	-5.2	-6.3	2.1	-2.8	100.0

Fig. D.7 Correlation plot for linearity study fit with -10% b-tagging efficiency and $\alpha = 1$ (only with correlations greater than 15%).

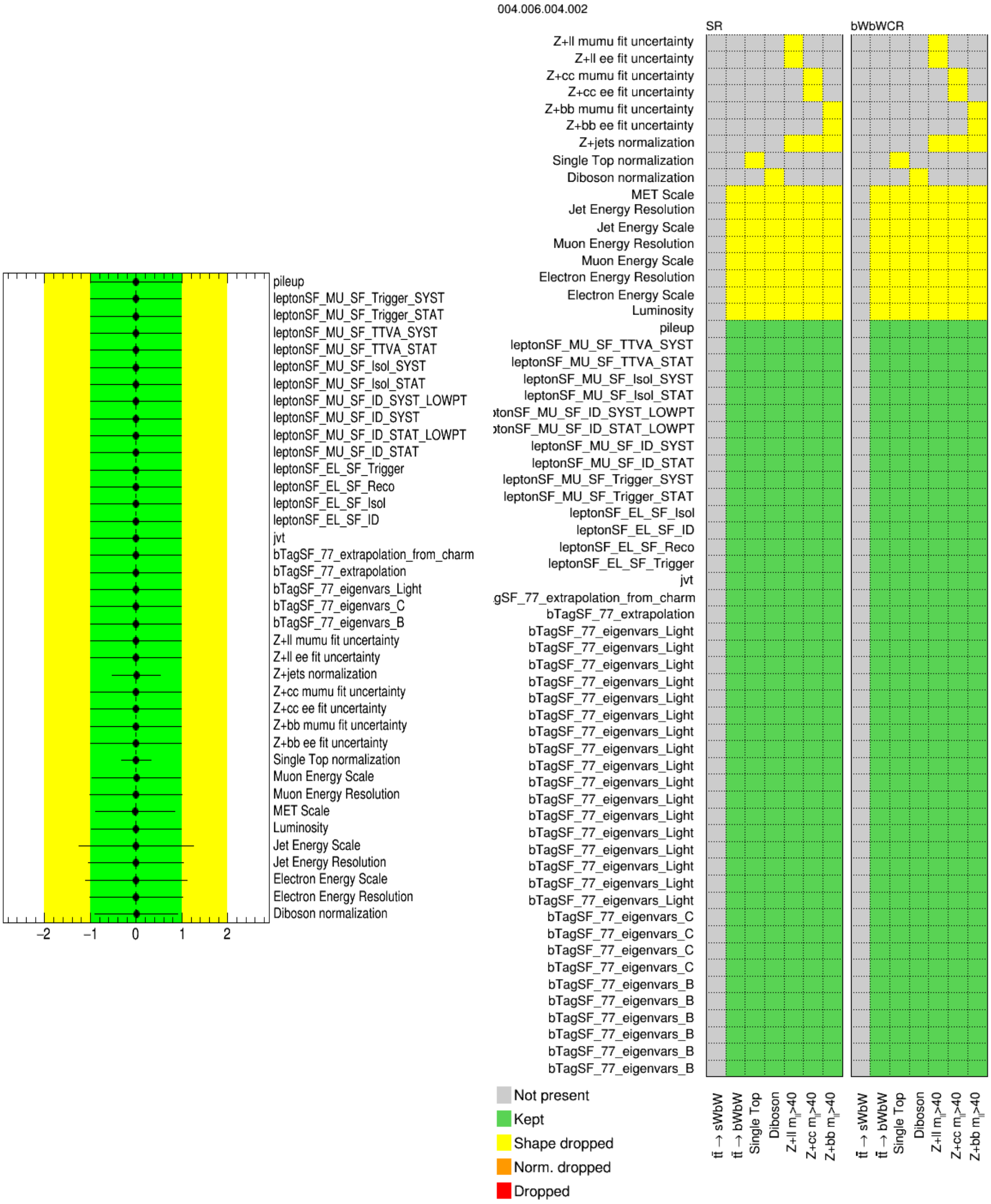


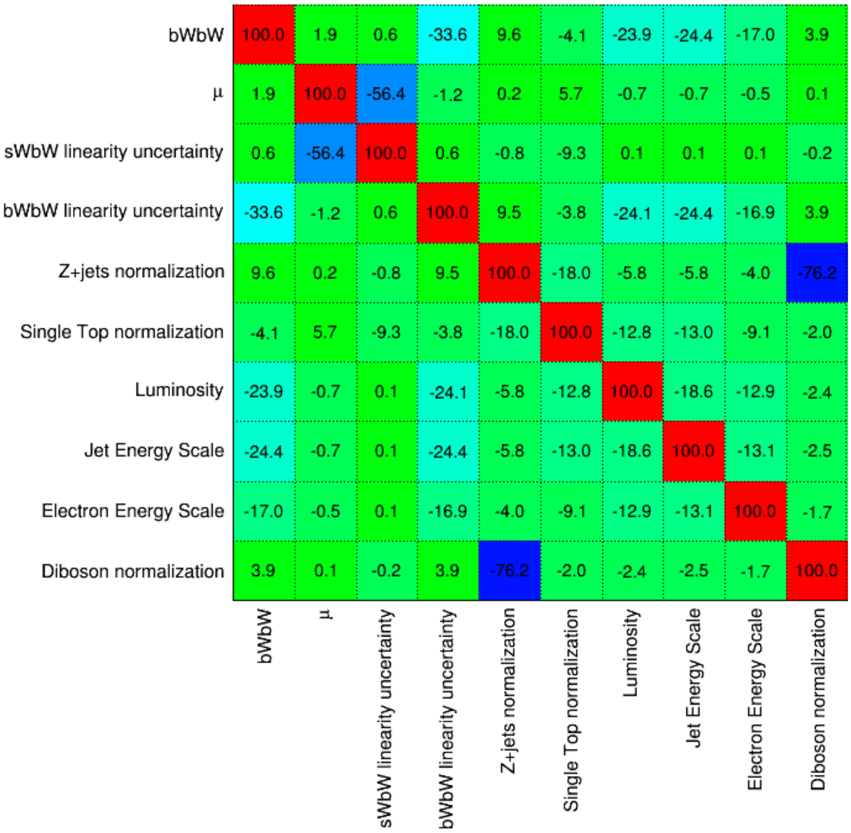
Fig. D.8 Pull plot (left) and pruning plot (right) for linearity study fit with -10% *b*-tagging efficiency and $\alpha = 1$.

D.5 BR measurement fit

Table D.5 Yields of all regions for BR measurement (statistical uncertainty only).

Sample	Yields			
	SR		Wt VR	
	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t} \to sWbW$	183.3 ± 12.6	183.3 ± 17.3	2061 ± 132	2061 ± 130
$t\bar{t} \to bWbW$	42.8 ± 31.0	42.8 ± 5.2	285364 ± 203194	285363 ± 3617
Single Top	1.79 ± 1.42	1.79 ± 1.28	28502 ± 10019	28505 ± 2539
Diboson	3.84 ± 0.90	3.84 ± 0.86	15726 ± 1651	15722 ± 1239
Z+light	0.74 ± 0.82	0.74 ± 0.75	3809 ± 2022	3808 ± 1050
Z+charm	0.92 ± 1.03	0.92 ± 0.94	1405 ± 730	1405 ± 355
Z+bottom	0 ± 0	0 ± 0	1148 ± 602	1148 ± 302
Total	233.4 ± 33.5	233.4 ± 18.1	338016 ± 203549	338032 ± 3826
Data	233		362033	

	$t\bar{t} \to bWbW$ CR		Z+jets VR	
	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t} \to sWbW$	3542 ± 227	3543 ± 222	229 ± 16	229 ± 15
$t\bar{t} \to bWbW$	549216 ± 391070	549259 ± 2998	26833 ± 19106	26834 ± 363
Single Top	35474 ± 12469	35424 ± 3158	1939 ± 683	1936 ± 178
Diboson	16954 ± 1780	16957 ± 1336	2598 ± 274	2598 ± 207
Z+light	4311 ± 2263	4312 ± 1138	28922 ± 14607	28931 ± 6421
Z+charm	1819 ± 1186	1819 ± 838	7980 ± 4033	7972 ± 1778
Z+bottom	1447 ± 746	1446 ± 356	5735 ± 2920	5737 ± 1326
Total	612763 ± 391380	612764 ± 1564	74233 ± 28789	74248 ± 9096
Data	612763		61095	



	bWbW	μ	sWbW linearity uncertainty	bWbW linearity uncertainty	Z+jets normalization	Single Top normalization	Luminosity	Jet Energy Scale	Electron Energy Scale	Diboson normalization
bWbW	100.0	1.9	0.6	-33.6	9.6	-4.1	-23.9	-24.4	-17.0	3.9
μ	1.9	100.0	-56.4	-1.2	0.2	5.7	-0.7	-0.7	-0.5	0.1
sWbW linearity uncertainty	0.6	-56.4	100.0	0.6	-0.8	-9.3	0.1	0.1	0.1	-0.2
bWbW linearity uncertainty	-33.6	-1.2	0.6	100.0	9.5	-3.8	-24.1	-24.4	-16.9	3.9
Z+jets normalization	9.6	0.2	-0.8	9.5	100.0	-18.0	-5.8	-5.8	-4.0	-76.2
Single Top normalization	-4.1	5.7	-9.3	-3.8	-18.0	100.0	-12.8	-13.0	-9.1	-2.0
Luminosity	-23.9	-0.7	0.1	-24.1	-5.8	-12.8	100.0	-18.6	-12.9	-2.4
Jet Energy Scale	-24.4	-0.7	0.1	-24.4	-5.8	-13.0	-18.6	100.0	-13.1	-2.5
Electron Energy Scale	-17.0	-0.5	0.1	-16.9	-4.0	-9.1	-12.9	-13.1	100.0	-1.7
Diboson normalization	3.9	0.1	-0.2	3.9	-76.2	-2.0	-2.4	-2.5	-1.7	100.0

Fig. D.9 Correlation plot for BR measurement fit (only with correlations greater than 15%).

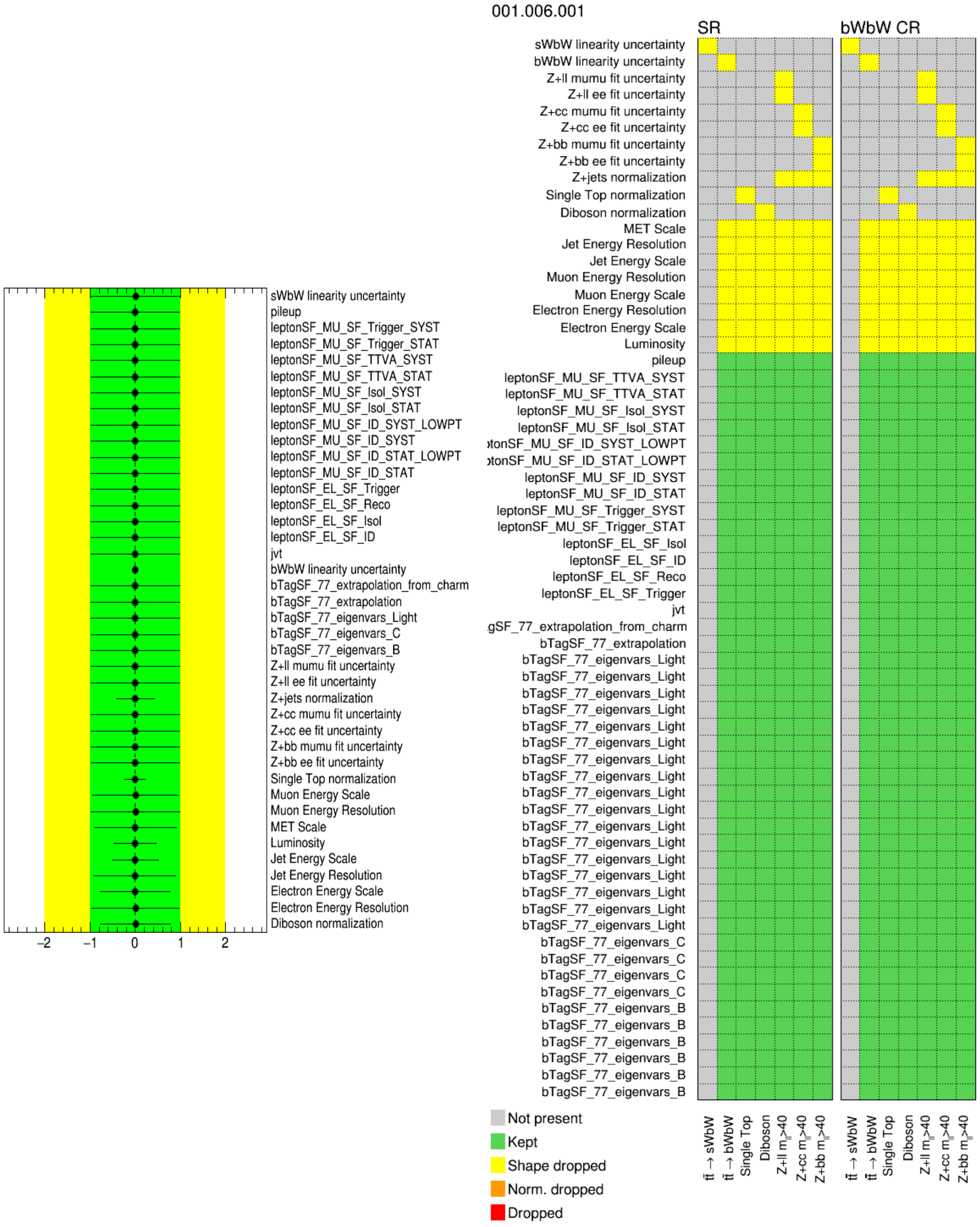


Fig. D.10 Pull plot (left) and pruning plot (right) for BR measurement fit.

References

[1] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course In Modern Particle Physics*. 1984.

[2] C. Patrignani *et al.*, "Review of Particle Physics," *Chin. Phys.*, vol. C40, no. 10, p. 100001, 2016.

[3] U. Husemann, "Top-Quark Physics: Status and Prospects," *Prog. Part. Nucl. Phys.*, vol. 95, pp. 48–97, 2017.

[4] M. Cristinziani and M. Mulders, "Top-quark physics at the Large Hadron Collider," *J. Phys.*, vol. G44, no. 6, p. 063001, 2017.

[5] T. W. B. Kibble, "History of electroweak symmetry breaking," *J. Phys. Conf. Ser.*, vol. 626, no. 1, p. 012001, 2015.

[6] E. Noether, "Invariant Variation Problems," *Gott. Nachr.*, vol. 1918, pp. 235–257, 1918. [Transp. Theory Statist. Phys.1,186(1971)].

[7] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions," *Phys. Rev. Lett.*, vol. 74, pp. 2626–2631, 1995.

[8] DØ Collaboration, "Observation of the top quark," *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, 1995.

[9] LHC*top* WG, "Summary of the atlas and cms direct m_{top} measurements," 2017. `https://twiki.cern.ch/twiki/pub/LHCPhysics/LHCTopWGSummaryPlots/LHC_topmass_sep2017.pdf`. [Online; accessed 2018-06-07].

[10] LHC*top* WG, "Summary of the atlas and cms extractions of the ckm matrix element v_{tb} from single top quark measurements," 2018. `https://twiki.cern.ch/twiki/pub/LHCPhysics/LHCTopWGSummaryPlots/singletop_vtb_lhc_may18.pdf`. [Online; accessed 2018-06-07].

[11] LHC*top* WG, "Summary of measurements of the top-pair production cross-section at 13 tev," 2018. `https://twiki.cern.ch/twiki/pub/LHCPhysics/LHCTopWGSummaryPlots/tt_xsec_lhc13_may18.pdf`. [Online; accessed 2018-06-07].

[12] LHC*top* WG, "Summary of measurements of the top-pair production cross-section at 13 tev," 2017. `https://twiki.cern.ch/twiki/pub/LHCPhysics/LHCTopWGSummaryPlots/singletop_allchannels_nov17.pdf`. [Online; accessed 2018-06-07].

[13] "About CERN," Jan 2012. `http://cds.cern.ch/record/1997225`. [Online; accessed 2018-07-26].

[14] UA1 Collaboration, "Further Evidence for Charged Intermediate Vector Bosons at the SPS Collider," *Phys. Lett.*, vol. B129, pp. 273–282, 1983.

[15] UA1 Collaboration, "Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider," *Phys. Lett.*, vol. B126, pp. 398–410, 1983.

[16] SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Group, L3 Collaboration, "Precision electroweak measurements on the Z resonance," *Phys. Rept.*, vol. 427, pp. 257–454, 2006.

[17] T. Berners-Lee, R. Cailliau, J.F. Groff and B. Pollermann, "World-wide Web: the information universe," *Electron. Netw.*, vol. 2, pp. 52–58, 1992.

[18] 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.*, vol. B716, pp. 1–29, 2012.

[19] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett.*, vol. B716, pp. 30–61, 2012.

[20] L. Evans and P. Bryant, "LHC Machine," *JINST*, vol. 3, p. S08001, 2008.

[21] J. Haffner, "The CERN accelerator complex. Complexe des accélérateurs du CERN," Oct 2013. `https://cds.cern.ch/record/1621894` [Online; accessed 2018-05-29].

[22] A. Team, "Diagram of an LHC dipole magnet. Schéma d'un aimaint dipôle du LHC." `https://cds.cern.ch/record/40524`. [Online; accessed 2018-07-26], Jun 1999.

[23] J. Pequenao, "Computer generated image of the whole ATLAS detector." `http://cds.cern.ch/record/1095924`. [Online; accessed 2018-05-29], Mar 2008.

[24] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, p. S08003, 2008.

[25] ATLAS Collaboration, "Total integrated luminosity and data quality in 2015-2017," 2017. `https://atlas.web.cern.ch/Atlas/GROUPS/DATAPREPARATION/PublicPlots/2017/DataSummary/figs/intlumivstimeRun2DQ.png`. [Online; accessed 2018-05-28].

[26] ATLAS Collaboration, "Number of interactions per crossing," 2017. `https://atlas.web.cern.ch/Atlas/GROUPS/DATAPREPARATION/PublicPlots/2017/DataSummary/figs/intlumivstimeRun2DQ.png`. [Online; accessed 2018-05-29].

[27] J. Pequenao and P. Schaffner, "An computer generated image representing how ATLAS detects particles." Jan 2013.

[28] ATLAS Collaboration, "ATLAS magnet system: Technical design report," 1997.

[29] ATLAS Collaboration, "ATLAS inner detector: Technical design report. Vol. 1," 1997.

[30] ATLAS Collaboration, "ATLAS inner detector: Technical design report. Vol. 2," 1997.

[31] ATLAS Collaboration, "ATLAS calorimeter performance Technical Design Report," 1996.

[32] ATLAS Collaboration, "ATLAS liquid argon calorimeter: Technical design report," 1996.

[33] ATLAS Collaboration, "Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions," *Eur. Phys. J.*, vol. C70, pp. 723–753, 2010.

[34] ATLAS Collaboration, "ATLAS tile calorimeter: Technical design report," 1996.

[35] ATLAS Collaboration, "Readiness of the ATLAS Tile Calorimeter for LHC collisions," *Eur. Phys. J.*, vol. C70, pp. 1193–1236, 2010.

[36] ATLAS Collaboration, "ATLAS muon spectrometer: Technical design report," 1997.

[37] ATLAS Collaboration, "Commissioning of the ATLAS Muon Spectrometer with Cosmic Rays," *Eur. Phys. J.*, vol. C70, pp. 875–916, 2010.

[38] K. Kordas *et al.*, "The ATLAS Data Acquisition and Trigger: Concept, design and status," *Nucl. Phys. Proc. Suppl.*, vol. 172, pp. 178–182, 2007.

[39] ATLAS Collaboration, "ATLAS first level trigger: Technical design report," 1998.

[40] ATLAS Collaboration, "ATLAS high-level trigger, data acquisition and controls: Technical design report," 2003.

[41] A. Ruiz-Martinez and A. Collaboration, "The Run-2 ATLAS Trigger System," Tech. Rep. ATL-DAQ-PROC-2016-003, CERN, Geneva, Feb 2016.

[42] J. Shiers, "The Worldwide LHC Computing Grid (worldwide LCG)," *Computer Physics Communications*, vol. 177, no. 1, pp. 219 – 223, 2007. Proceedings of the Conference on Computational Physics 2006.

[43] Ježo, Tomáš and Nason, Paolo, "On the Treatment of Resonances in Next-to-Leading Order Calculations Matched to a Parton Shower," *JHEP*, vol. 12, p. 065, 2015.

[44] Sjöstrand, Torbjörn and Ask, Stefan and Christiansen, Jesper R. and Corke, Richard and Desai, Nishita and Ilten, Philip and Mrenna, Stephen and Prestel, Stefan and Rasmussen, Christine O. and Skands, Peter Z., "An Introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.

[45] ATLAS Collaboration, "ATLAS Run 1 Pythia8 tunes," Tech. Rep. ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014.

[46] Ball, Richard D. and others, "Parton distributions for the LHC Run II," *JHEP*, vol. 04, p. 040, 2015.

[47] Ball, Richard D. and others, "Parton distributions with LHC data," *Nucl. Phys.*, vol. B867, pp. 244–289, 2013.

[48] Sjostrand, Torbjorn and Mrenna, Stephen and Skands, Peter Z., "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[49] Skands, Peter Zeiler, "Tuning Monte Carlo Generators: The Perugia Tunes," *Phys. Rev.*, vol. D82, p. 074018, 2010.

[50] Lai, Hung-Liang and Guzzi, Marco and Huston, Joey and Li, Zhao and Nadolsky, Pavel M. and Pumplin, Jon and Yuan, C. -P., "New parton distributions for collider physics," *Phys. Rev.*, vol. D82, p. 074024, 2010.

[51] Alwall, Johan and Herquet, Michel and Maltoni, Fabio and Mattelaer, Olivier and Stelzer, Tim, "MadGraph 5 : Going Beyond," *JHEP*, vol. 06, p. 128, 2011.

[52] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *JHEP*, vol. 02, p. 007, 2009.

[53] ATLAS Collaboration, "Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data," *Eur. Phys. J.*, vol. C74, no. 7, p. 2941, 2014.

[54] ATLAS Collaboration, "Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data," Tech. Rep. ATLAS-CONF-2016-024, CERN, Geneva, Jun 2016.

[55] ATLAS Collaboration, "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV," *Eur. Phys. J.*, vol. C76, no. 5, p. 292, 2016.

[56] ATLAS Collaboration, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1," *Eur. Phys. J.*, vol. C77, p. 490, 2017.

[57] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[58] ATLAS Collaboration, "Expected performance of the ATLAS b-tagging algorithms in Run-2," Tech. Rep. ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015.

[59] ATLAS Collaboration, "Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS," *Eur. Phys. J.*, vol. C72, p. 1844, 2012.

[60] ATLAS Collaboration, "Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV," 2013.

[61] ATLAS Collaboration, "Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 13$ TeV using $e\mu$ events with b-tagged jets," Tech. Rep. ATL-COM-PHYS-2016-021, CERN, Geneva, Oct 2016.

[62] ATLAS Collaboration, "Search for flavour-changing neutral current $t \to qZ$ top quark decays in proton-proton collisions at $\sqrt{s} = 13$ with the ATLAS detector," Tech. Rep. ATL-CONF-2017-070, CERN, Geneva, Sep 2017.

[63] ATLAS Collaboration, “Luminosity Determination in pp Collisions at $\sqrt(s) = 8$ TeV using the ATLAS Detector at the LHC,” *Eur. Phys. J.*, vol. C76, p. 653, 2016.

[64] F. Chollet, *Deep Learning with Python.* Greenwich, CT, USA: Manning Publications Co., 1st ed., 2017.