UNIVERSITY OF TARTU

Faculty of Science and Technology

Institute of Physics

Joosep Pata

MEASUREMENT OF TOP QUARK POLARISATION IN T-CHANNEL SINGLE TOP PRODUCTION WITH THE CMS DETECTOR

Supervisors:

Mario Kadastik, PhD

prof. Andrea Giammanco, PhD

prof. acad. Martti Raidal, PhD

Master's thesis

Tartu 2014

Contents

Preliminaries

In this thesis, the Einstein summation convention is used, with repeated indices corresponding to an implicit sum over those indices:

$$a^2 = a^i a_i = \sum_i a^i a_i. \tag{1}$$

Greek letters represent space-time indices $\mu \in \{0, 1, 2, 3\}$, with time being the 0-th component. The metric is chosen according to the mostly minus convention

$$\eta_{\mu\nu} = \text{diag}\,(+1, -1, -1, -1). \tag{2}$$

In this work, the natural system of units is used where $\hbar = c = 1$, meaning that 1 GeV $= 1.602 \times 10^{-10}$ J is the only fundamental unit, with the length and time of GeV^{-1} corresponding to 0.1975×10^{-15} m and 0.659×10^{-24} s.

Introduction

This thesis describes my work in the measurement of top quark polarisation in t-channel production of single top quarks [1]. The measurement was carried out using approximately $L_{\text{int.}} = 20$ fb^{-1} of proton-proton collision data at a centre-of-mass energy of $\sqrt{s} = 8$ TeV. The data were collected during the 2012 run of the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). The results of this measurement have been incorporated into the Particle Data Group review of the top quark [2].

The top quark (the $Q = 2/3$, $T_3 = +1/2$ member of the weak-isospin doublet containing the bottom quark [2]) is the most massive elementary particle discovered so far, with $M_{\text{top}} = 173.34 \pm 0.27$ (stat.) $\pm$ 0.71 (syst.) GeV as measured independently by the CMS, ATLAS, D0 and CDF collaborations and reported in a recent combination [3]. For comparison, a gold atom, consisting of 79 protons, each containing 3 valence quarks, has a mass of 183 GeV. Since M_{top} is close to the electroweak symmetry breaking scale set by the vacuum expectation value of the Higgs field $v = 246$ GeV, the top quark plays an important role in the determining the properties of the recently-discovered Higgs boson. The stability of the the Standard Model (SM) Higgs potential and the phenomenology of the top quark are mainly driven by the large mass [4].

Additionally, the top quark is the only quark that can decay into a real W boson and which has a Yukawa coupling to the Higgs boson (Y_t) of order unity. Thus, it plays an important role in the SM and its extensions and an accurate determination of the properties of the top quark is crucial in shaping the development of our understanding of the underlying physics [5].

The electroweak decay of the top quark proceeds on time-scales of $\tau_t = 0.5 \times 10^{-24}$ s, an order of magnitude shorter than typical quantum chromodynamics (QCD) time-scales of $\tau_{\text{QCD}} = 10^{-23}$ s. The top quark decays predominantly by $t \to W^+b$, before it can form top-flavoured hadrons or $t\bar{t}$ bound states. Therefore, the decay products W^+, b ($W^-, \bar{b}$) of the (anti-)top quark retain information about the spin orientation of the top quark, which would otherwise be randomized in the hadronisation [6]. Thus, the top quark is a powerful tool in experimental studies of the structure of the Wtb vertex.

In the SM and in the t-channel single-top production process, discovered in 2009 at Tevatron [7,

8], considered in this thesis, the top quark is produced with nearly 100% spin polarisation P_t, meaning that the top quark population is left-handed. This is a result of the vector-axial-vector ($V-A$) structure of the Wtb coupling vertex, which can directly be probed by measuring the angular distributions of the decay products of the top quark [9]. The aim of this work is to experimentally determine the polarisation of the top quark. This measurement was not possible with precision before the start of the LHC, since at the Tevatron collider, the statistics were not sufficient for a differential measurement, with a production cross-section $\sigma = 1.98 \pm 0.25$ pb and an integrated luminosity of around half of that of the 2012 of LHC [10, 11].

Therefore, this work summarises the first successful experimental measurement of the polarisation in the single-top channel with $\sim 50\%$ precision. Presently, polarized top quarks can only be produced in the Tevatron collider at Fermilab, which is now shut down, and at the LHC of CERN. Previously, the polarisation measurement has been attempted by the CDF collaboration, however, sufficient precision to draw conclusions about the polarisation was not achieved[12].

In this analysis, the polarisation was measured directly and in a model-independent way by selecting a pure sample of single top t-channel collision events and reconstructing the (anti)-top quark candidate from the decay $t \to W^+b \to l^+\bar{\nu}_l b,\ l = \mu, e$*. The polarisation was measured through the differential angular distribution $d\Gamma/d\cos\theta^* = \Gamma(1/2 + A\cos\ \theta^*)$, where A is the asymmetry defined as $A = \frac{N_\uparrow - N_\downarrow}{N_\uparrow + N_\downarrow}$ and Γ the partial decay width. Here, $N_\uparrow$ ($N_\downarrow$) are the number of events where the projection of the charged lepton momentum onto the spectator jet axis is (anti-)parallel to the spectator jet direction in the top quark rest frame, defined through the angle θ^* between the charged lepton and the spectator jet as $N_\uparrow = N(\cos\ \theta^* > 0)$ [$N_\downarrow = N(\cos\ \theta^* \leq 0)$].

The asymmetry is directly related to the polarisation through $A = 1/2\alpha P_t$, with α being spin analysing power of the charged lepton[13]. The spin analysing power relates the top quark spin to the angular distribution of the charged lepton, measuring the sensitivity of the angular distribution of the daughter particle to the spin state of the parent. The charged lepton has near-maximal ($\alpha \simeq \pm 1$) spin analysing power in the SM and several extensions of it and in this analysis, it is assumed to be 1. The spin analysing power is defined though the expectation value $\langle\cos\theta\rangle = \langle \boldsymbol{p}\cdot\boldsymbol{P}\rangle$, where $\boldsymbol{p}$ is the 3-momentum of the spin analyser and $\boldsymbol{P}$ the polarisation vector.

*charge conjugation is implied unless otherwise noted

The selection of t-channel events is performed by physically motivated selection criteria validated in t-channel cross-section measurements and by an additional multi-variate machine learning technique. The yields of the signal and background processes are determined in-place by a binned maximum likelihood fit on the differential distribution of the output of the multivariate classifier.

In general, detector effects and an imperfect selection and reconstruction efficiency inhibit the analysis of all relevant t-channel events and introduce a distortion to the sensitive distribution $d\Gamma/d\cos\theta^*$. The value of A is extracted from the distribution corrected for the known detector effects by a technique called unfolding, which allows the true distribution at parton-level to be extracted by applying inversion with regularization. The use of unfolding allows P_t to be directly measured in the absence of detailed particle physics models with non-SM values of P_t [14, 15] and the results to be directly comparable to theoretical predictions.

The measurement is performed independently in the muon and electron decay channels of the W boson, corresponding to a total branching ratio of approximately 22%. The hadronic branch is not used due to difficulties in triggering on a fully-hadronic final state, whereas the τ-lepton branch would suffer from comparatively large systematic uncertainties. The muon and electron channels serve as cross-checks to each other and the final result is given by a statistical combination of the independent results through the use of the best linear unbiased estimator (BLUE) method [16].

1 Theoretical background

This section will elucidate the necessary conceptual and mathematical framework for working with top quark polarisation in the context of the SM of particle physics.

1.1 The Standard Model of particle physics

Particle physics aims to describe the elementary constituents of matter and the known interactions between them. The SM provides a quantitative and predictive model for much of the sub-nuclear world, and is remarkably successful in explaining, with some accommodation, all of the observed phenomena of experimental particle physics. It is an accumulation of iterative developments, which started with the discovery of the electron by Thompson in 1897 [17]. The discovery of the Higgs boson in 2012 represents the present state-of-the-art [18].

The SM is a re-normalizable theory of electroweak and strong interactions, based on the principles of symmetry. The major milestones in the development in the SM are the Salam[1]-Glashow[2]-Weinberg[3] unification of the electromagnetic and the weak force into the electroweak theory in the 1960-s [19, 20, 21], the development and incorporation of the mechanism for mass generation in gauge theories by Brout[4], Englert[5], Higgs[6], Kibble[7] [22, 23], the development of QCD to model strong interaction by, among others, Fritzsch[8], Gell-Mann[9], Leutwyler[10], Gross[11] and Wilczek[12] in the 1970-s [24]. Additionally, the important concepts of quark-gluon confinement in hadrons [25] and asymptotic freedom by Weinberg [26], [27] have proved to be accurate descriptions of the structure of hadrons. The re-normalizability of the strong and electroweak theories was shown by 't Hooft[13] and Veltmann[14] in the 1970-s, providing predictive power to the theory [28, 29].

The SM is sufficient to model roughly 5% of the observed matter-energy density of the Universe, the bulk of which is dark matter and dark energy, and all the fundamental forces except for gravity, the quantum theory of which is non-renormalisable. The only indisputable hints for physics beyond the SM presently come from astrophysical and cosmological observations.

1.1.1 Fundamental objects

In quantum field theory (QFT), the fundamental objects are complex fields ϕ depending on the space-time coordinates x. The fields are classified based on their transformation properties under the symmetry transformations of the theory, providing SM with the deep formal structure of group theory. A particularly useful classification is based on the Lorentz transformation corresponding to the group $\mathrm{SO}^+(3,1)$, first formulated by Minkowski [30].

The group of proper ortochronous Lorentz transformations consists of all Λ^μ_ρ which satisfy

$$\Lambda^\mu_{\ \rho}\eta_{\mu\nu}\Lambda^\nu_{\ \sigma} = \eta_{\rho\sigma},\ \det\ \Lambda = +1,\ \Lambda^0_0 = +1 \tag{3}$$

with

$$\eta^{\mu\nu} = \mathrm{diag}(+1,-1,-1,-1) \tag{4}$$

being the metric tensor of Minkowski.

The irreducible* representations $L(\Lambda)$ defined through the transformation rule of fields ϕ_A under a Lorentz transformation Λ

$$U(\Lambda)^{-1}\phi_A(x)U(\Lambda) = L^B_A\phi^B(\Lambda^{-1}x) \tag{5}$$

of the Lorentz group are most easily classified according to the correspondence $\mathrm{SO}^+(3,1) \simeq \mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$, that is according to the angular momenta (j_1, j_2) of the left and right handed $\mathrm{SU}(2)$ components of the decomposition. The fundamental objects used in the SM transform as scalars $(j_1 = 0, j_2 = 0)$, left and right handed Weyl spinors $(1/2, 0)$, $(0, 1/2)$ or vectors $(1, 1)$ under the Lorentz transformations.

The connection between massive particle states with spin and the Poincaré[15] group, which is the Lorentz group plus translations $x^\mu \to \Lambda^\mu_\nu x^\nu + a^\mu$ was first elucidated by Wigner[16] [31]. One particle states are the unitary representations, classified according to the eigenvalues of the Casimir invariants corresponding to mass $P^2 = P^\mu P_\mu \to m^2 > 0$ and spin j from $W^2 = W^\mu W_\mu \to -m^2 j(j+1)$ for massive particles and helicity $\lambda(P) = \boldsymbol{J}\cdot\boldsymbol{P}/|\boldsymbol{P}|$ and momentum P^μ for massless

*A representation is reducible if there exists a unitary transformation U acting on the representation $L_A \to U^{-1}L_A U$ such that the matrix representation L_A can be put into a block-diagonal form. A representation that is not reducible is irreducible.

particles. Here, $P^\mu = +i\partial^\mu$ is the standard 4-momentum of a particle and $W^\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}J^{\nu\rho}P^\sigma$ is the Pauli-Lubanski pseudovector. $J^{\mu\nu} = M^{\mu\nu} + S^{\mu\nu}$ is the angular momentum tensor, where the orbital angular momentum $M^{\mu\nu} = x^\mu p^\nu - x^\nu p^\mu$ and intrinsic angular momentum or spin $S^{\mu\nu}$ is independent of the space-time coordinate.

1.1.2 Gauge transformations as interactions

Gauge theories provide a description of fundamental interactions by requiring the invariance of the equations of motion, derived using the principle of minimal action, of the theory under local transformations of the fields $\phi(x) \to U(x)\phi(x)$.

Such continuous and differentiable transformations U form Lie groups G, the utility of which lies in the fact that continuous non-infinitesimal transformations can be represented as the exponential $U = \exp(-i\theta_j(x)H_j)$ of a finite number of Hermitian[†] operators: the generators $H_j,\ j = 1 \ldots n$, which form the Lie algebra g with dimension n that is equipped with the inner product $[H_j, H_k] = H_jH_k - H_kH_j = ih_{jkl}H_l$. The Lie algebra is completely described by specifying the structure constants h_{jkl}. The gauge parameters $\theta_j(x)$ are real and specify $U(x)$ in terms of the expansion in generators.

Let us take the example of an SU(N) gauge transformation, having $n = N^2 - 1$ generators. The invariance of the equations of motion of the theory under a local[‡] gauge transformation is achieved by the minimal substitution, replacing partial derivatives with respect to the space-time coordinates ∂_μ by their covariant counterparts

$$\partial_\mu\phi \Rightarrow \mathcal{D}_\mu\phi = (\partial_\mu + igA^j_\mu H_j)\phi. \tag{6}$$

The covariant derivative arises from the observation that the fields ϕ transform as representations R of the gauge group G by $\phi(x) \to U_R(x)\phi(x)$ where $U_R(x) = \exp(ig\theta_j(x)H_{j,R})$, but the partial derivatives

$$\partial_\mu\phi(x) \to U_R(x)\partial_\mu\phi(x) + \partial_\mu[U_R(x)]\phi(x) \tag{7}$$

[†] A matrix representation of H is Hermitian iff. $H^\dagger = (H^T)^* = H$

[‡] the expansion coefficients $\theta_j(x)$ of a generic element of the group are coordinate-dependent

do not. In order to guarantee that the kinetic terms $\partial_\mu\phi(x)$ transform as ϕ under the gauge symmetry, the additional terms must be counteracted. For this, gauge fields $A_\mu = A^j_\mu H_j$ corresponding to the generators II_j are introduced, which transform as

$$A_\mu \to U_R A_\mu U_R^\dagger - \frac{i}{g}(\partial_\mu U_R)U_R^\dagger \tag{8}$$

under the representation R. The covariant derivative defined in eq. 6 transforms in the same way as the fields ϕ: $\mathcal{D}_\mu\phi \to U_R\mathcal{D}_\mu\phi$.

The new gauge fields A^j_μ are spin 1[§] and mediate the interaction between the fields ϕ with the interaction strength g through the kinetic terms $\bar{\phi}\gamma^\mu\mathcal{D}_\mu\phi$ which now contains interaction between ϕ and A_μ through the terms $gA^j_\mu\bar{\phi}\gamma^\mu H_j\phi$.

Quantum electrodynamics (QED) was formulated as a gauge theory by Tomonaga[17], Feynman[18] and Schwinger[19] [32, 33, 34] and is generated by the Abelian $U(1)$ group denoted as $U(1)_{EM}$. The generator of this group is the charge Q_f of the fermion (electron), the gauge field A^μ the photon. This is sufficient to generate Maxwell's equations of electromagnetism and to give very accurate predictions on observables such as the anomalous dipole moment of the electron $g-2$, which is accurate to within on part per billion [35].

The strong force and quarks are described by the $\mathrm{SU}(3)_C$ group, where C stands for a tri-valent quantum number called colour. Gluons are described by the 8 vector fields $G^a_\mu,\ a = 1\ldots 8$ corresponding to the generators of the $SU(3)_C$ group. As the structure constants h_{jkl} of this group do not vanish, it is non-Abelian and hence the gluons undergo self-interaction, which is the essence of the eponymous gauge theory first developed by Yang[20] and Mills[21] [36].

The unification of charged weak interactions and neutral electromagnetic interactions proceeded by observing that they could be grouped together into an $\mathrm{SU}(2)$ group, which provides 3 generators with the correct corresponding charges. However, the difficulty was that only the left-handed particles and right-handed anti-particles were seen to interact weakly, whereas electromagnetism is insensitive to chirality. By enlarging the left-handed $\mathrm{SU}(2)_L$ group with $U(1)_Y$, where Y stands for hyper-charge, the electroweak vector fields $W^j_\mu,\ j = 1, 2, 3$ and B_μ[¶] form the observed charged

[§]Transform as vectors under the Lorentz group

[¶]W^j_μ is a triplet and B_μ a singlet under the combined group. B_μ and the neutral W_μ can be interpreted in terms of

current bosons $W^{\pm}$ and the electromagnetic exchange boson γ. The observation of the newly predicted neutral current Z_μ confirmed that the true electroweak gauge group must encompass at least $\mathrm{SU}(2)_L \times U(1)_Y$.

The minimal group structure of the SM is hence given by

$$\mathrm{G}_{SM} = \mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y, \tag{9}$$

which contains strong and electroweak interactions. The SM fields have conserved charges under the combined group, consisting of the colour charge, weak isospin T and its projection T_3 of $\mathrm{SU}(2)_L$ and hypercharge Y of $\mathrm{U}(1)_Y$.

The complete covariant derivative has contributions from all these groups:

$$\mathcal{D}_\mu = \partial_\mu + ig_1 Y B_1 + ig_2 \tau_j W_\mu^j/2 + ig_3 \lambda_a G_\mu^a/2. \tag{10}$$

Here τ_j are the Pauli matrices of $SU(2)$ and λ_a the Gell-Mann matrices of $SU(3)$.

The minimal SM Lagrangian up to gauge-fixing and ghost terms is then

$$\mathcal{L} = \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{Yang-Mills}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}, \tag{11}$$

where

$$\mathcal{L}_{\text{fermion}} = \sum_{\Psi_L} \bar{\Psi}_L i\gamma^\mu \mathcal{D}_\mu \Psi_L + \sum_{\Psi_R} \bar{\Psi}_R i\gamma^\mu \mathcal{D}_\mu \Psi_R \tag{12}$$

describes the dynamics of the fermions, along with their interaction with gauge bosons and

$$\mathcal{L}_{\text{Yang-Mills}} = -\frac{1}{4} B^{\mu\nu} B_{\mu\nu} - \frac{1}{4} W_i^{\mu\nu} W^i_{\mu\nu} - \frac{1}{4} G_a^{\mu\nu} G^a_{\mu\nu} \tag{13}$$

which gives the interaction between gauge bosons through the field strength tensors according to their symmetry groups.

The inclusion of masses is problematic. For the fermion fields, the usual prescription of $m\bar{\Psi}\Psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$ is not possible due to the mass term not being an $\mathrm{SU}_L(2)$ singlet, and for spin-1 bosons the mass term $m^2 A_\mu A^\mu$ is not invariant under gauge transformations. This is circumvented by introducing the mass through the Higgs mechanism [22, 23], where a complex scalar field Φ (an

the photon and a new neutral current, the two charged W_μ as the charged weak currents.

$\mathrm{SU}_L(2)$ doublet) along with a $\mathrm{SU}_L(2)$ invariant potential $V(\Phi)$ is introduced and the spontaneous symmetry breaking of the lowest-energy state gives in effective mass for the weak gauge bosons terms from the following Lagrangian:

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}^{\mu}\Phi)^{\dagger}(\mathcal{D}_{\mu}\Phi) - V(\Phi^{\dagger}\Phi). \tag{14}$$

For fermions, the mass generation proceeds through the coupling of the left (Q_L, L_L) and right handed (u_R, d_R, l_R, ν_R) spinors with the $\mathrm{SU}_L(2)$ doublet Φ by

$$\mathcal{L}_{\text{Yukawa}} = -\bar{Q}_L Y_u u_R \tilde{\Phi} - \bar{Q}_L Y_d d_R \Phi + h.c. \tag{15}$$

$$-\bar{L}_L Y_\nu \nu_R \tilde{\Phi} - \bar{L}_L Y_l l_R \Phi + h.c., \tag{16}$$

where Q_L, L_L are the left-handed quark and lepton $\mathrm{SU}(2)_L$ doublets, u_R and d_R the right-handed up and down type quark singlets and l_R, ν_R the right handed charged lepton and neutrino singlets, with the existence of ν_R being true only in the case of non-vanishing neutrino masses ($Y_\nu \neq 0$). The coupling constants Y set the masses of the fermions through symmetry breaking to

$$m_f = \frac{v}{\sqrt{2}} Y_f, \tag{17}$$

where $v/\sqrt{2}$ is the vacuum expectation value of the scalar field

$$\Phi(x) = \frac{v + \phi(x)}{\sqrt{2}}$$

expanded around the minimum $\Phi^{\dagger}\Phi = -\mu^2/2\lambda$ of the quadratic potential

$$V(\Phi^{\dagger}\Phi) = \mu^2 \Phi^{\dagger}\Phi + \lambda(\Phi^{\dagger}\Phi)^2. \tag{18}$$

The existence of a non-zero and SU(2)-breaking vacuum expectation value leads to a spontaneous breaking of the SM gauge group

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \rightarrow \mathrm{SU}(3)_C \times \mathrm{U}(1)_{\text{em}}, \tag{19}$$

where the symmetry of 3 of the 4 generators of the electroweak group is broken, introducing masses for the fields $W^{\pm}, Z^0$. The fourth unbroken symmetry $\mathrm{U}(1)_{\text{em}}$ corresponds to electromagnetism mediated by massless photons. The massless degrees of freedom of the Higgs field, corresponding

to Goldstone bosons, are removed by a local gauge transformation, by which they become the missing longitudinal degrees of freedom of the massive gauge bosons. The charges of the Higgs field under $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ are chosen as $T = 1/2$, $T_3 = -1/2$, $Y = +1$, resulting in the conservation of the electromagnetic charge $Q = T_3 + Y/2$. This means that the vacuum state is invariant under $\mathrm{U}(1)_{\mathrm{em}}$ gauge transformations $\exp(i\alpha(x)Q)$. The mass terms $\frac{1}{2}m^2 A^\mu A_\mu$ for gauge bosons arise from the covariant derivative after the Higgs field ϕ settles in the vacuum state with $m = gv^2/2$. The mass terms of the fermions, which have the form $m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$, are created similarly from the Yukawa terms after symmetry breaking, with $m = -Y_i v/\sqrt{2}$.

1.2 Top quark physics

At the LHC, single top quarks are produced in the t-channel process by the exchange of a space-like W boson. This is fundamentally an electroweak process and thus the top quarks are strongly polarized. In contrast, the production of top quark pairs, which proceeds mainly through $gg \to t\bar{t}$ at the LHC [37], a single spin configuration does not dominate the top quark pair ensemble [38]. Single tops can also be produced in the tW channel $pp \to bg \to tW^+$, where the top quark is produced in association with an emitted W boson in the final state, and the s-channel $pp \to tb$ where the W boson is timelike. The tW ($\sigma = 11.1$ pb) and s ($\sigma = 5.5$ pb) channel production modes are suppressed by relatively small cross-sections at the LHC, compared to t-channel ($\sigma = 87.2$ pb) [37]. The tW and s channels also have different kinematic distributions of the final state particles compared to the t-channel and are considered as minor backgrounds in this analysis. At first order, the production of top quarks in t-channel proceeds via the diagrams shown on figure 1, where the $2 \to 2$ scheme is used for cases where the initial b-quark has a low transverse momentum, defined as $p_T = \sqrt{p_x^2 + p_y^2}$ in the plane transverse to the beam axis, and the $2 \to 3$ process for high transverse momentum cases. The spin polarisation of the top quark is defined with respect to a spin basis $\boldsymbol{s}$. Using the helicity basis, where the spin is measured along the axis of motion of the top quark by

$$h = \boldsymbol{S} \cdot \hat{\boldsymbol{p}} \tag{20}$$

with $\hat{\boldsymbol{S}} = \frac{1}{2}(\tau_1, \tau_2, \tau_3)$ being the spin operator and $\hat{\boldsymbol{p}} = \boldsymbol{p}/|\boldsymbol{p}|$ a unit vector in the direction of

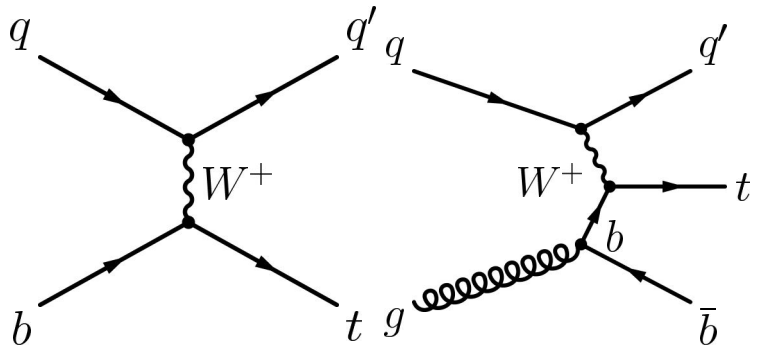

Figure 1: The leading-order Feynman diagrams for top quark production in the t-channel at the LHC. The left panel shows the $2 \to 2$ process (5-flavour scheme), where the b-quark is assumed to come from the 'sea' of quarks in a relativistic proton, whereas the right panel displays the $2 \to 3$ (4-flavour scheme) process which is logarithmically divergent. In the 5-flavour scheme, the proton contains a bottom quark in the 'sea', whereas in the 4-flavour scheme, the b-quark must come from the splitting of a gluon.

momentum is not the most effective choice, since the top quark produced at the LHC is not ultra-relativistic ($v/c \lesssim 0.6$), hence the helicity does not allow conclusions to be drawn about chirality, and additionally the helicity is not a manifestly Lorentz invariant quantity. This is important, since the natural zero-momentum frame is not perfectly reconstructable, thus, the helicity basis is subject to experimental effects. Instead, by using the spectator jet direction as a proxy for the d-type quark direction, the spin can be decomposed with an accuracy of $\sim 95\%$. The decay products of the top quark represent spin analysers, with a spin analysing power corresponding to the correlation between the direction of the decay product momentum and the spin of the top quark. The charged lepton has the highest spin analysing power at the LHC, with $\alpha \simeq 1$. A heuristic argument based on angular momentum conservation for the spin analysing power of the lepton is illustrated on figure 2, where the left-handed nature of the top decay will force the charged lepton direction to be correlated ($\cos\,\theta^* \sim +1$) to the top quark spin. The analysing power does not change significantly in several extensions of the SM, therefore, in this analysis it is fixed at $\alpha = 1$ [39].

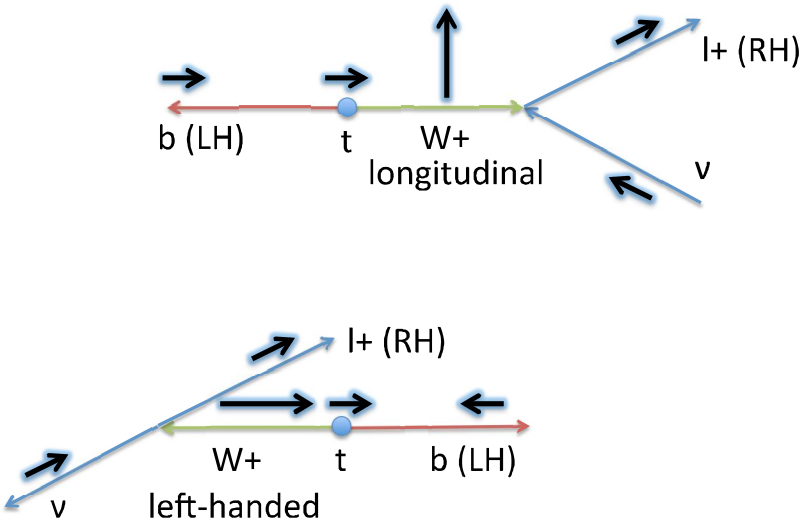

Figure 2: An adapted diagram [40] illustrating the correlation between the top quark spin and the charged lepton direction, based on the conservation of angular momentum, i.e. spin (dark, bold arrows). In case the W+ boson is longitudinal, it prefers to move in the same direction as the top quark spin, and the decay products align along the W^+ polarisation, with the charged lepton being emitted in the W^+ momentum direction due to the Lorentz boost. For the left-handed W^+ boson, the b quark must be left handed, forcing it to move in the direction of the top quark spin, and the W boson moving against this direction. The direction of the charged lepton at W decay is set by the chiral nature of the interaction, with the charged lepton being right-handed (RH) and the b-quark left-handed (LH). The charged lepton direction is correlated with the spin of the top quark in both cases.

2 The experimental setup

2.1 The Large Hadron Collider

The LHC is a hadron collider situated in the cavern of the former Large Electron-Positron (LEP) collider at CERN, Geneva, Switzerland. The designed instantaneous luminosity of the collider is $L = 10^{34}$ cm^{-2}s^{-1}, with a centre-of-mass energy of $\sqrt{s} = 14$ TeV. The collider has been operating since April 2010, mainly in the mode of proton-proton collisions with short runs of lead-ion collisions. Presently, the LHC is the only available tool for directly producing top quarks, after the shut-down of the Tevatron in Fermilab in Batavia, Illinois, USA in September 2011. The acceleration of protons at LHC proceeds in stages by the linear accelerator (LINAC) 2 to 50 MeV, the proton synchrotron booster (PSB) to 1.4 GeV, by the proton synchrotron (PS) to 26 GeV and finally to 450 GeV by the super proton synchrotron (SPS) before insertion into the main ring in bunches, which are sustained for up to around 24 hours and collided in 4 interaction points (IP) with a frequency of 40 MHz. Two separate beam-pipes exist for the proton/ion beams, where particle beams are bent with the help of 1232 superconducting dipoles, along with 392 quadrupoles for focusing. The magnets are kept at an operating temperature of 1.9 K using 96 tons of liquid He-4.

During the data taking for this analysis, the collider operated at $\sqrt{s} = 8$ TeV with a bunch spacing of 50 ns and an instantaneous luminosity of $L = 7.7 \times 10^{33}$ cm^{-2}s^{-1}.

2.2 The CMS detector

Two general purpose detectors, A Toroidal LHC Apparatus (ATLAS) and CMS, and four smaller specialized experiments, A Large Ion Collider Experiment (ALICE), LHCb, Total Cross Section, Elastic Scattering and Diffraction Dissociation at the LHC (TOTEM), LHCf are placed on the LHC IP-s. ATLAS and CMS are designed for the validation of SM properties, the study of the electroweak symmetry breaking (EWSB) mechanism and possibly detecting new physics processes.

The CMS detector uses a right-handed coordinate system, formed by the z-axis in the anticlockwise beam direction, x-axis pointing to the centre of the LHC ring and y perpendicular to the ring plane

and upwards. The cylindrical coordinate system used in CMS is defined by the polar angle ϕ and the pseudo-rapidity $\eta = -\ln\tan(\theta/2)$ where θ is the azimuthal angle with respect to the z-axis.
The CMS detector consists of a superconducting solenoid with a 6-meter diameter, providing a magnetic field of about 3.8 T. The detector contains a tracking system consisting of silicon pixel and strip detectors. The silicon detector is designed to reconstruct the 3-dimensional tracks of all charged particles within the acceptance region $|\eta| < 2.5$. Reconstruction of these tracks allows the inference of the location of the primary interaction vertex (of the hard scattering) [41].
The tracker is surrounded by a lead tungstenate ($PbWO_4$) electromagnetic calorimeter (ECAL) with an acceptance of $|\eta| < 3.0$. The ECAL is important in determining the direction and energy of electrons and photons [42].
This is followed by the brass/scintillator hadronic calorimeter (HCAL), designed to measure the energy of charged and neutral hadrons. The solenoid is enclosed in gas ionisation detectors, drift tubes and cathode strip chambers, along with the steel return yoke of the solenoid. The calorimetric acceptance in the forward region, close to the beam pipe, is extended by an additional quartz-fibre and steel Cerenkov calorimeter [43].
The projected event rate of 40 MHz must be reduced to about 1 kHz for storage. This is done by a two-stage decision process, consisting of triggers, which must make a decision about storing the event in a matter of microseconds. The first stage (L1) accesses only the calorimeter and muon system information and is located partly on the detector in the form of dedicated hardware. This reduces the event rate to about 100 kHz. A cluster of commodity processors is then used to partially reconstruct the event in the high-level trigger (HLT), which decides if the event readout is kept. The HLT performs object reconstruction, jet clustering, b-tagging and calculation of invariant masses, reducing the event rate up to around 500 Hz for the main CMS physics program and 300-600 Hz for so-called parked data, which will be analysed in the future [44].

2.3 Simulation

The simulation of various physical processes taking place at LHC in the proton-proton collision is done in several stages. First, the hard interaction and the underlying Feynman diagrams, accounting

for loop corrections where they are known, is done using specialized software packages such as MADGRAPH [45], PYTHIA [46], SHERPA [47], POWHEG [48] or COMPHEP [49]. Typically, the parton-level hard process of proton scattering is modelled separately from the fragmentation and hadronisation (production of jets) of the quarks. In cases where there is ambiguity in the choice of model parameters, such as the choice of the renormalisation and factorization (Q^2) scale or the matching of the matrix element to the parton shower, the usual approach is to calculate the nominal response around the nominal value and use variations of the response as an estimation of systematic uncertainty.

The complete simulation of the CMS detector response for a particular event is very time consuming and is thus done centrally by the collaboration for the processes considered in this analysis. A fine-grained (resolution $\sim$ 100 μm) model of the CMS detector has been implemented by the collaboration using the GEANT4 tool-kit [50]. This model accounts for the known interactions between high-energy particles and matter, giving the 3-dimensional energy deposits in each region of the detector.

In general, a collision of the proton bunches at a beam crossing consists of hard interactions, where the quarks and gluons are well-described by perturbative QCD and are characterized by high-p_T remnants in the detector, and soft interactions which include among other processes diffraction, soft initial and final state radiation and beam-beam interactions and form the majority. We wish to study the hard interactions, therefore, they must be selected from the underlying data using a triggering system. However, the triggering criteria can introduce a selection bias by only considering events which pass some selection criteria, therefore, it is important to estimate the amount and properties of the soft contributions in the selected events.

The number of interactions per bunch crossing N_{int}, known as pile-up (PU), is a random variable distributed according to the Poisson distribution with the mean

$$\langle N_{\text{int}} \rangle = L\sigma\tau_b \tag{21}$$

where L is the instantaneous luminosity of the collider, σ the total inelastic scattering cross-section and τ_b the bunch spacing. In order to accurately model the PU conditions at the LHC, the number of interactions per bunch crossing is sampled from the Poisson distribution according to the minimal

bias cross-section, which is an experimentally-determined cross-section using triggers that minimize the selection bias from the kinematic cuts. The true and modelled distributions of the N_{int} are verified to be compatible within the uncertainty of the minimum bias cross-section.

The protons are composite objects, the contents of which are determined experimentally by measuring parton distribution functions (PDF) as a function of $\sqrt{s}$. Due to this, the mathematical description of the proton-proton collision has to use a particular PDF during the MC simulation of the processes. In this work, the MSTW08NNLO PDF set is used [51], with different PDF sets being used to evaluate the systematic uncertainty resulting from the proton composition.

3 The analysis method

3.1 Event selection and reconstruction

In order to analyse the properties of the weakly-produced top quark in t-channel production, the detector subsystem information and readouts must be interpreted in terms of physical particles and processes. As this study focuses on the leptonic decay mode of the W boson, signal events arising from the decay $t \to W^+b \to l^+\bar{\nu}_l b$ are identified by the presence of exactly one isolated electron or muon, missing transverse energy (defined below) from the neutrino in the leptonic W-decay, a b-jet from the top quark decay that is preferentially emitted in the central direction of the detector and an additional light quark from the hard scattering, which is generally emitted in the forward region of the detector. A secondary b-jet may be present as well, which in terms of the 4-flavour scheme arises from gluon splitting as seen on figure 1, however, it generally remains undetected due to a low transverse momentum. Events with additional muons or electrons are expressly vetoed to reduce the contamination by either $t\bar{t}$ or Drell-Yan type background processes. The event selection for this analysis follows closely the adopted best practices for other single-top analyses, in particular the measurement of single top t-channel cross-section and the $|V_{tb}|$ element of the CKM matrix at $\sqrt{s} = 8$ TeV [52].

In order to reduce the amount of data taking to feasible quantities (about 1kHz for storage), the data recorded from a collision event must pass the HLT based on the presence of at least 1 isolated muon (electron) over the transverse momentum threshold of $p_T > 24\ (27)$ GeV and a pseudo-rapidity of $|\eta| < 2.1\ (2.5)$. The performance of the muon and electron triggers is studied using tag-and-probe techniques, and is validated to not affect the analysis significantly [53].

The reconstruction and identification of muons, electrons, hadronic jet candidates and of the $\not{E}_T$ is done by the particle flow (PF) algorithm, which combines the signals of various subsystems of the detector into an interlinked description of the physics process. Typically, the stable constituents of collimated particle jets having $p_T \sim 100$ GeV which arise from decays of exotic particles are of the order $p_T < 10$ GeV. Therefore, it is essential to reconstruct as many of these relatively soft particles as possible with a low fake rate, and the PF algorithm is suited for that, improving the

expected performance over traditional methods [54, 55].

With the PF algorithm, the primary interaction vertex is reconstructed and the charged lepton is required to originate from it, restricting the contamination of non-prompt leptons. Additionally, the lepton is required to be isolated from other particles. This is achieved by requiring the sum of charged hadron, photon and neutral hadron transverse momenta in a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4(0.3)$ to be less than 12% (10%) for muons (electrons). The identification of electrons is further enhanced by a dedicated multi-variate identification criterion. Events with an additional lepton with $p_T > 10(20)$ GeV, $|\eta| < 2.5$ are rejected in the muon (electron) channel, reducing the contamination from $t\bar{t}$-events.

The presence of the neutrino from the leptonic decay of the W boson results in missing transverse energy ($\not{E}_T$), which is defined vectorially as

$$\not{\boldsymbol{E}}_{\boldsymbol{T}} = -\sum_{i \in PF} \boldsymbol{p}_{T,i} \tag{22}$$

where $\boldsymbol{p}_{T,i}$ is the transverse momentum of the i-th particle candidate identified by the PF algorithm. The scalar missing energy ($\not{E}_T$) is defined as the modulus $|\not{\boldsymbol{E}}_{\boldsymbol{T}}|$. Due to the presence of a real W boson in top-quark decay, as the top quark mass exceeds the masses of the W boson and b-quark, the reconstructed invariant mass of the W boson serves as a useful discriminator between the weak signal process and the background processes. The transverse mass m_T is defined as

$$m_T = \sqrt{\left(p_T^l + \not{E}_T\right)^2 - \left(p_x^l + \not{E}_{T,x}\right)^2 - \left(p_y^l + \not{E}_{T,y}\right)^2}. \tag{23}$$

where $p_T^l = |\boldsymbol{p}_{\boldsymbol{T}}^{\boldsymbol{l}}| = (p_x^l,\ p_y^l)$ is the transverse momentum of the charged lepton.

The criterion $m_T > 50$ GeV ($\not{E}_T > 45$ GeV) is used in the muon (electron) channel to reject background events arising from strong interactions i.e. QCD processes such as $pp \to jj'$, where one of the jets is or resembles a b-jet.

The reconstruction of jets proceeds by the clustering of the PF candidates by the anti-k_T (cone size 0.5) algorithm. The anti-k_T algorithm is a generalization of the usual sequential jet recombination algorithms such as k_T or the Cambridge-Aachen algorithm, where the distance measure between the subcomponents of the jet uses negative powers of the transverse momentum k_T^{2p}, with $p = -1$ corresponding to the anti-k_T method. This algorithm has been show to be robust against fluctuations

in the radiation of low-p_T particles[56]. Events with at least two jets passing $p_T > 40$ GeV, $|\eta| < 4.5$ are selected for this analysis, in order to reduce the multi-jet QCD background.

The main background for the single top t-channel process arises from the production of W bosons in association with jets, which can fake a true top quark that decays to a W boson. This process is denoted W+jets and its relative yield with respect to signal is reduced by requiring the presence of a jet which is likely to originate from a b-quark, i.e. a b-jet. For this analysis, the track counting (TC) algorithm is used, with the working point being optimized to have an approximately 0.1% probability of labelling jets not originating from b-quarks as b-jets (mis-tagging) and a selection efficiency of $\sim 40\%$ for true b-quarks [57].

In order to reject events where the untagged jet does not arise from the primary vertex, which arises mainly from pile-up, the events are required to pass

$$\Delta R_{rms} = \frac{1}{N}\sum_{i}^{N} \Delta R_i < 0.025, \tag{24}$$

where $\Delta R_i(\Delta\eta, \Delta\phi)$ is the angular separation between the i-th jet constituent and the jet axis and the sum is evaluated over all the N constituents.

The events are classified to signal and control regions based on the number of jets N_{jets} and the number of jets passing the b-tagging criterion N_{b-tags} and denoted as N_{jets}JN_{b-tags}T. The measurement is performed in the 2J1T region, with 2J0T serving as the control region for W+jets modelling and 3J1T, 3J2T as the control region for $t\bar{t}$ processes.

The presence of the true top quark can already be seen by calculating the invariant transverse mass of the neutrino ($\not{E}_T$), the b-jet and the charged lepton as a distinctive mass peak. However, in order to reconstruct the variable sensitive to polarisation ($\cos\ \theta^*$), the complete top quark candidate reference frame should be reconstructed. For this, the b quark is identified as the jet with the highest b-tagging discriminator value, which is important for disambiguation in the background-enriched control regions 2J0T, 3J1T, 3J2T. The recoil jet is identified as the one with the lowest b-discriminator value.

Additionally, the neutrino candidate from the leptonic decay of the W boson must be reconstructed in order to determine the top quark candidate reference frame. All of the decay products of $t \to bW \to bl\nu$ are reconstructable by the detector, but the longitudinal momentum p_z of the neutrino is

not measured directly. By making an assumption that the W boson is on the mass shell ($p_{W,\mu}p^{\mu}_{W} = m^2_W = (80.4\ \text{GeV})^2$) and that the entirety of $\not\!\!E_T$ is created by the undetected neutrino, momentum conservation implies

$$m^2_W = (E_\ell + \sqrt{\not\!\!E_T{}^2 + p^2_{z,\nu}})^2 - (\boldsymbol{p_{T,\ell}} + \boldsymbol{\not\!\!E_T})^2 - (p_{z,\ell} + p_{z,\nu})^2 \,. \tag{25}$$

which has 2 solutions

$$p_{z,\nu} = \frac{\Lambda \cdot p_{z,\ell}}{p^2_{T,\ell}} \pm \sqrt{\frac{\Lambda^2 \cdot p^2_{z,\ell}}{p^4_{T,\ell}} - \frac{E^2_\ell \cdot \not\!\!E_T{}^2 - \Lambda^2}{p^2_{T,\ell}}} \,, \tag{26}$$

with

$$\Lambda = \frac{m^2_W}{2} + \vec{p}_{T,\ell} \cdot \vec{\not\!\!E}_T \,. \tag{27}$$

In case two real solutions exist, the one with min $|p_{z,i}|$ is used.

If the equation admits no real solutions, which happens when $m_T > m_W$ and the discriminant becomes negative, an additional assumption has to be made. The $\not\!\!E_T$ polar angle ϕ is assumed to be measured with imperfect accuracy and is variated minimally to achieve a solution to the W-mass constraint. This is done analytically and leads to a cubic equation which always gives a solution.

Having reconstructed the neutrino 4-momentum p_μ, the top quark reference frame is found by adding the momenta of the neutrino and the charged lepton to for a W boson candidate, followed by the b-tagged jet momentum to reconstruct the top quark candidate.

Additional event selection is done via a multi-variate machine learning technique, which is able to capture the information content of several discriminating variables and therefore achieve a more optimal separation between the signal and background hypotheses. This is described in the following section.

3.1.1 Boosted decision trees

In this differential measurement, achieving a high purity and still keeping a reasonable amount of statistics is important, therefore, state-of-the-art machine learning techniques were applied to select collision events that have a high probability of arising from the single top t-channel production.

Boosted decision trees represent a way to combine the values of a large amount of variables $\boldsymbol{x} = (x_1, \ldots, x_n)$ corresponding to an observation into a response y, which can serve as a predictor for some feature of the observation, such as the underlying physical process (classification, e.g. $y \in$ signal, background) or some physical quantity which can only be indirectly accessed (regression, e.g. y corresponds to the momentum of an undetected particle).

The boosted decision tree algorithm lies in the class of supervised machine learning techniques, where some parameters corresponding to the model $F() \Rightarrow y$ are optimised with respect to the desired outcome, in this case, a perfect separation boundary in the input parameter space $\boldsymbol{x}$ and tested with respect to a statistically independent set of observations for which the true class is known. The training is continued until some stopping criteria, specific to the algorithm, are achieved [58].

Decision trees are in some ways the simplest classifiers that can be envisioned, where repeated binary decisions (nodes) are built on the input variables, based on a maximal information gain as measured by a suitable function of (mis)-classified observations, such as the Gini information coefficient defined through purity P on a node

$$P = \frac{\sum_{i \in \text{signal}} w_i}{\sum_i w_i} \tag{28}$$

as

$$G = \sum_i^n w_i P(1 - P) \tag{29}$$

where w_i is the weight corresponding to the i-th observation and n the number of observations on a particular node. The binary decision is chosen such as to maximize

$$C = G(\text{top}) - G(\text{left}) - G(\text{right}), \tag{30}$$

where top, left and right correspond to the parent node and the left and right nodes. The building of nodes is stopped when the purity has reached a certain threshold, the number of observations is below a statistically acceptable value for the tree has achieved a maximal node depth N_{depth} from the root node. The linked set of branches forms a decision tree which will give a binary classification for a particular observation.

Boosting involves repeatedly training trees with very few branches (base learners), with the training

observations receiving weights corresponding to correct or incorrect classification, thereby introducing feedback. The rationale is that a single decision tree is susceptible to fluctuations in the input data and that a single non-optimal split near the root can cause the complete tree to be inaccurate, whereas a large set of simple trees will be better able to capture the true underlying features of the training data.

In this analysis, gradient boosting is applied by constructing the model as $F(\boldsymbol{x}) = \sum_n^N \beta_n h_n(\boldsymbol{x})$ where β_n is a real and positive coefficient, $h_n(\boldsymbol{x}) : \boldsymbol{x} \to \{-1, +1\}$ the n-th decision tree out of an ensemble of N with -1 corresponding to background and +1 to signal prediction. The model is trained by iteratively solving

$$F^* = \arg\min L(y, F(\boldsymbol{x}|\boldsymbol{\beta})) \tag{31}$$

over the set of observation-class pairs $\{\boldsymbol{x}, y\}_i$ for the coefficients $\boldsymbol{\beta}$, the set of base learners $\boldsymbol{h} = \{h_1 \dots h_n\}$ with the loss function

$$L(y, F) = \log\left(1 + \exp\left(-2yF\right)\right). \tag{32}$$

This minimization is done numerically using the gradient method [59] using the TMVA package [60], which is free software. The BDT algorithm is visualized on figure 3.

The input variables $\boldsymbol{x}$ used for the training of the BDT are selected to be approximately independent of the coupling structure and to have some discriminatory power between single top t-channel and W+jets and $t\bar{t}$ processes, as verified by simulation and comparison of differential cross-sections in these variables.

Additionally, the list of variables is pruned by iteratively removing the ones that provide no significant benefit by the comparison of the signal efficiency vs. background rejection, i.e. receiver operating characteristic (ROC) curves of the discriminators with successive trainings.

The final list of variables is given below:

- the mass of the top-quark candidate, $M_{bl\nu}$;
- the pseudo-rapidity of the jet failing b tagging, $\eta_{j'}$;
- the "C-parameter" of the event (defined below)

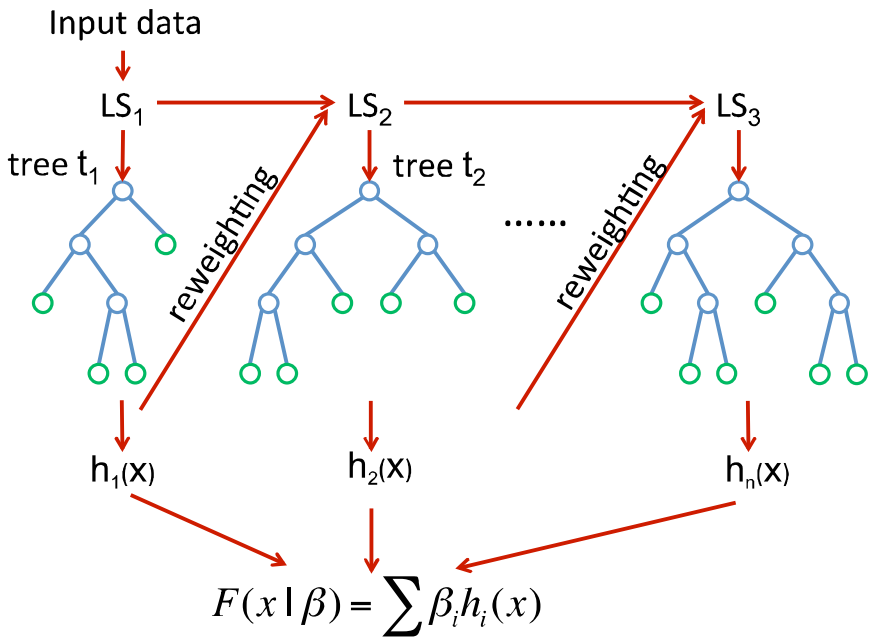

Figure 3: A visual description of the boosted decision tree algorithm. Each base learner (tree $h_i(x)$) is trained on successively re-weighted data pairs $\{\boldsymbol{x}, y\}$ to minimize the total loss function in terms of the weights $w_{i,j}$ of the i-th event under the j-th iteration and the combination coefficients β_i.

- the mass of the b-tagged jet, m_{bj};
- the mass of the untagged jet, m_{lj};
- the transverse momentum of the lepton, p_T^{lepton};
- the transverse energy of the b-tagged jet, p_T^{bj};
- the invariant transverse mass, m_T;
- the missing energy, $\not{E}_T$.

In general, the reconstructed top quark candidate mass $M_{bl\nu}$ and the pseudo-rapidity of the the recoil jet $\eta_{j'}$ were found to be very good discriminators. Intuitively, it can be understood that $M_{bl\nu}$ selects events that have at least one true top quark (single top t-channel or $t\bar{t}$, with the peak being slightly more smeared in $t\bar{t}$ due to the possible mis-identification of the individual top decay products under a single top quark hypothesis). Therefore, it will efficiently reject W+jets events and other backgrounds not containing a true top quark. Physically, the t-channel process gives an $\eta_{j'}$ distribution in the forward region due to the t-channel nature of the production, where the highly-energetic recoil quark is only somewhat deflected from the beam-line by the exchange of a W boson. In $t\bar{t}$, where the light quarks arise from e.g. a hadronic W decay or additional QCD radiation in a $gg \to t\bar{t} + g$ process, the forward region is of no special preference.

Of particular interest as a discriminator is the C-parameter, which measures the 3-jet structure of an event and vanishes for a perfect back-to-back two-jet event. C is constructed from the eigenvalues of the momentum tensor m_{ij} the elements of which are defined as

$$M_{ij} = \sum_n p_{n,i} p_{n,j}, \; m_{ij} = M_{ij} / \det M \tag{33}$$

by

$$C = 3 \cdot (q_1 * q_2 + q_1 * q_3 + q_2 * q_3), \tag{34}$$

where $0 \leq q_1 \leq q_2 \leq q_3$ are the eigenvalues of the momentum tensor m_{ij} and $p_{n,i}$ is the i-th component of the momentum 3-vector of the n-th particle.

The use of such derived variables provides the multivariate classifier with useful information about the event, which would otherwise be difficult to construct automatically, and allows the use of relatively shallow decision trees ($N_{\text{depth}} \sim 3 \ldots 5$, improving the speed and efficiency of training and classification. The use of other event shape variables such as aplanarity or centrality derived from the momentum tensor has been shown not to give a significant advantage for this analysis, since much of the information content of the other variables is already captured in C, however, they, along with the Fox-Wolfram moments [61, 62] become important for QCD or Higgs-physics studies. Additionally, a benefit of such variables defined through ratios of kinematic variables is that many systematic uncertainties in the input variables cancel, making the construct therefore more robust [63] than the inputs separately.

Future developments may include the incorporation of deep learning techniques to automatically and systematically find such combinations from input data. Such methods have been already studied in high-energy physics and shown to have great promise [64].

The working point α for the scalar discriminator value of the BDT to select signal events was chosen to yield a similar signal efficiency (defined as the ratio $N(\text{signal} \mid \text{BDT} > \alpha))/N(\text{signal}))$ to the t-channel cross-section measurement [52] for comparison purposes. In the future, a figure-of-merit for analysis sensitivity will be developed and the working point value will be optimized with respect to that value. In general, it is be expected that at low values of α, the analysis will be dominated by the uncertainties of the background estimation, whereas at very high values of α, the limited simulation statistics will inhibit precision, therefore an optimal point should lie somewhere in between.

The final determined values of α are given in table 1, along with the signal efficiency and background rejection.

lepton	α	ϵ_S	$1-\epsilon_B$
$\mu^{\pm}$	0.06	29.4%	97.2%
$e^{\pm}$	0.13	26.8%	97.2%

Table 1: Working points for selecting a signal-enriched phase space for the boosted decision tree in the muon and electron channel. The signal efficiency ϵ_S and background rejection $1-\epsilon_B$ are calculated before and after applying the $BDT > \alpha$ threshold on signal, $t\bar{t}$ and W+jets events.

The BDT was trained and tested on statistically independent samples, which is necessary to correctly estimate the amount of over-training, i.e. the sensitivity of the training to statistical fluctuations in the training data. Additionally, all the input variables were verified to be well-modelled in data in the 2J1T signal region by MC simulation.

3.2 Background processes

The validation of the selection and reconstruction procedures is done using statistically independent data. In particular, the lepton isolation criterion is inverted to determine the yield and distribution of multi-jet QCD-related backgrounds and the modelling of W+jets and $t\bar{t}$ is validated in the 2J0T and 3J1T, 3J2T control regions.

3.2.1 QCD

The vast majority of multi-jet QCD events are rejected by the m_T ($\not{E}_T$) requirement, however, are occurrences, e.g. the events populating the tails of the lepton p_T spectra, can contaminate the signal region. In order to reduce the sensitivity to modelling uncertainties in these extreme cases, the QCD model is derived directly from data, with MC simulations being used as cross-checks. The *ansatz* is made that data with the charged lepton isolation requirement inverted contain a large fraction of QCD events and the differential distributions of the kinematic variables of the QCD process are not significantly altered by the isolation requirement. This has been verified in simulation and found to be sufficiently accurate, with QCD processes comprising around 99% of the anti-isolated data, with the non-QCD processes being subtracted based on simulation.

To determine the yield of the QCD background, a binned maximum likelihood fit is performed, where the following likelihood is minimized numerically:

$$L(\alpha, \beta, \gamma \mid \boldsymbol{d}) = \prod_{i=1}^{N} \text{Poisson}(d_i|\lambda_i) P(\alpha, \beta, \gamma), \tag{35}$$

where

$$\lambda_i = \alpha N_{\text{QCD}}(i) + \beta N_{\text{W+jets}}(i) + \gamma N_{\text{other}}(i) \tag{36}$$

is the expected Poisson mean of the model specified by the scale factor α for the QCD processes and β, γ are treated as nuisance parameters. $P(\alpha, \beta, \gamma)$ is a function representing prior expectation about the parameters and for the case of QCD, is log-normal:

$$P(x = \gamma \mid \mu = 1, \sigma = 0.2) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log(x)-\mu)^2/2\sigma^2}. \tag{37}$$

The Poisson distribution is defined in the standard manner

$$\text{Poisson}(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}. \tag{38}$$

Here, $N_{QCD}(i)$ represents the i-th bin of the discretised anti-isolated data distribution in the reconstructed mass of the W boson m_T (missing energy $\not{E}_T$) for the muon (electron) channel. The N_{QCD} histogram is corrected for the anti-isolated background processes by subtracting the MC estimates. Similarly, $N_{\text{W+jets}}$ is the histogram of the W+jets processes as estimated by simulation in the isolated region and N_{other} contains all other relevant processes, in particular: $t\bar{t}$, single top (t, s and tW channel), Drell-Yan jets and diboson processes as estimated by simulation. $\boldsymbol{d}$ is the measured data distribution for the fit variable in the isolated region. Technically, the minimization is done using the publicly-available code `theta` [65].

The determined scale factors are used internally for the QCD model, and are assigned an uncertainty of 100% to cover the possible unaccounted non-QCD processes contaminating the anti-isolated distribution N_{QCD} and inaccuracies resulting from the fit.

Additionally, the statistical uncertainties in the MC-derived templates $N_{\text{W+jets}}$, N_{other} are accounted for by the Barlow-Beeston procedure, which considers the true bin means N as additional Poissonian random variables [66]. The priors and the inclusion of the limited MC statistics are however found to have a low impact on the final value of α.

3.2.2 W+jets

In general, the W+jets background results from generic production of W bosons in association with jets $pp \to W^{\pm} + qq' \to l^{\pm}\nu_l +$ jets or the production of $W^{\pm} + g$ or $W^{\pm} + q$ where an additional parton comes from initial state radiation. Such processes can easily fake a single top t-channel signal in case they pass the selection thresholds.

The W+jets background is very effectively reduced by the b-tagging 1T requirement, however the comparatively large production cross-section at $\sqrt{s} = 8$ TeV makes it a very significant background. Modelling studies of the W+jets background are performed in the 2J0T control region. Different MC generators are compared in their modelling of the $\cos\,\theta^*$variable, in particular, significant differences were found between the angular distributions as predicted by MadGraph [67]