Universidade do Minho
Escola de Ciências

Emanuel Demétrio Mendes Gouveia

Search for ttH production at the LHC

junho de 2016

Universidade do Minho
Escola de Ciências

Emanuel Demétrio Mendes Gouveia

Search for ttH production at the LHC

Tese de Mestrado
Mestrado em Física - Física Fundamental

Trabalho realizado sob a orientação de
Doutor José Ricardo Morais Silva Gonçalo
Professor Doutor António Joaquim Onofre de A. R. Gonçalves

junho de 2016

Nome: Emanuel Demétrio Mendes Gouveia
Endereço electrónico: eddygouveia91@gmail.com
Telefone: +351 918967796
Número do Bilhete de Identidade: 14179870
Título dissertação (EN): Search for ttH production at the LHC
Título dissertação (PT): Pesquisa por produção de ttH no LHC
Orientador: Doutor José Ricardo Gonçalo (LIP – Laboratório de Instrumentação e Física Experimental de Partículas)
Co-orientador: Doutor António Onofre (Departamento de Física da ECUM; LIP)
Ano de conclusão: 2016
Designação do Mestrado: Mestrado em Física - Física Fundamental

Declaração

Eu, Emanuel Demétrio Mendes Gouveia, declaro que é autorizada a reprodução integral desta dissertação apenas para efeitos de investigação, mediante declaração escrita do interessado, que a tal se compromete.

Universidade do Minho, ___ /___ /_____
Assinatura: _____

Acknowledgements

I start by ackowledging the support and guidance of my advisors, Ricardo Gonçalo and António Onofre. I am grateful to both of them for allowing me to take part in this project, and for the invaluable knowledge and experience it gave me. In particular, I am thankful to Ricardo for his diligent reviews and crucial comments during the writing of this document. I am thankful to António for numerous fruitful discussions about physics and for his persistent enthusiasm, regardless of how exhasperating the problems at hand may seem.

I am grateful to Rui Santos, for his kindness in reviewing the theoretical sections of this thesis.

I thank the researchers at LIP-Minho, namely Nuno Castro, Raul Sarmento and Juanpe Araque, for their promptness in giving me helpful advice, and for their interest in how the work was progressing. I am thankful for the good company of my fellow students with whom I worked together in $t\bar{t}H$ studies at some point: Susana, André, Rui Jorge, João Tiago, Lia and Eduardo.

A word of gratitude is due to my friends in Azeituna, for all the adventures they provided me during the last years. I also thank Xico, Palabra, São and Nácio, my housemates for most of the time during which this work was done, for their friendship. I am thankful to Eduarda for the permanent motivation and care, and for always being so proud of me.

Most of all, I am grateful to my family, and to my mother in particular, for all their effort that allows me to pursue my education in physics.

Search for $t\bar{t}H$ production at the LHC

Abstract

The recent discovery of a Higgs boson with a mass close to $125\,\mathrm{GeV}$, by the ATLAS and CMS experiments at the CERN LHC, represented a remarkable accomplishment of the Standard Model of particle physics. Studying this newly discovered particle provides insight into the electroweak symmetry breaking mechanism, by which all fundamental particles acquire mass. The mass of fermions, in particular, is proportional to the corresponding Yukawa couplings. The top quark Yukawa coupling is of order unity, much larger than that of any other fermion. A precise measurement of this parameter is a stringent test to the validity of the Standard Model. In proton-proton collisions, such as the ones occuring at the LHC, the production of a Higgs boson in association with a pair of top quarks ($t\bar{t}H$) is the only process allowing a direct measurement of the top quark Yukawa coupling.

In this thesis, a search for the production of $t\bar{t}H$ at the LHC is discussed. In the addressed channel, the Higgs boson decays to $b\bar{b}$ ($H \to b\bar{b}$) and the decay of the $t\bar{t}$ system is dileptonic ($t\bar{t} \to b\bar{b}W^+W^- \to b\bar{b}\nu\bar{\nu}\ell^+\ell^-$). An event reconstruction algorithm is described, along with a large set of new angular observables: the motivation for the interest in such observables is given, as well as the instructions for their computation.

Two complementary studies employing the new angular variables are presented. The first study demonstrates that these observables provide discrimination between samples of $t\bar{t}H$ and the irreducible background process $t\bar{t}b\bar{b}$. In the second study, the observables are tested as discriminants between $t\bar{t}H$ and $t\bar{t}A$, where A stands for a non-Standard Model pseudoscalar Higgs boson. When compared to a set of observables suggested for this purpose found in a literature review, the new angular observables show remarkable discrimination.

Pesquisa por produção de $t\bar{t}H$ no LHC

Resumo

A descoberta recente de um bosão de Higgs com uma massa próxima dos 125 GeV, pelas experiências ATLAS e CMS, no LHC, CERN, representou um feito assinalável do Modelo Padrão da física de partículas. O estudo desta partícula recentemente descoberta permite compreender o mecanismo de quebra espontânea de simetria, através do qual todas as partículas fundamentais obtêm massa. A massa dos fermiões, em particular, é proporcional aos acoplamentos de Yukawa correspondentes. O acoplamento de Yukawa do quark top é da ordem de grandeza da unidade, muito maior do que o de qualquer outro fermião. Uma medida precisa deste parâmetro é um teste rigoroso à validade do Modelo Padrão. Em colisões protão-protão, tais como as que ocorrem no LHC, a produção de um bosão de Higgs associado a um par de quarks top ($t\bar{t}H$) é o único processo que permite uma medida directa do acoplamento de Yukawa do quark top.

Nesta tese, discute-se uma nova estratégia de pesquisa por produção de $t\bar{t}H$ no LHC. No canal de decaimento abordado, o bosão de Higgs decai para $b\bar{b}$ ($H \to b\bar{b}$) e o decaimento do sistema $t\bar{t}$ é dileptónico ($t\bar{t} \to b\bar{b}W^+W^- \to b\bar{b}\nu\bar{\nu}\ell^+\ell^-$). Descreve-se um algoritmo de reconstrução de acontecimentos e um conjunto extenso de novas observáveis angulares é apresentado: é dada a motivação para o interesse nessas observáveis, bem como as instruções para o seu cálculo.

Dois estudos complementares das novas observáveis angulares são apresentados. O primeiro estudo demonstra que estas observáveis proporcionam discriminação entre as amostras de $t\bar{t}H$ e o fundo irredutível $t\bar{t}b\bar{b}$. No segundo estudo, as observáveis são avaliadas enquanto discriminantes entre $t\bar{t}H$ e $t\bar{t}A$, onde A representa um bosão de Higgs pseudoescalar, não previsto no Modelo Padrão. Quando comparadas com um conjunto de observáveis anteriormente propostas para esse efeito, encontradas numa revisão da literatura, as novas observáveis exibem discriminação notável.

Contents

Outline of the thesis

The Higgs boson was predicted in 1964 by three independent groups of physicists [1–6], as the experimentally verifiable consequence of a mechanism responsible for the spontaneous symmetry breaking in electroweak interactions. One of the main goals set for the operation of the Large Hadron Collider (LHC) [7], at CERN, was a better understanding of the electroweak symmetry breaking mechanism and, in particular, the discovery of this boson. It remained as the only undiscovered fundamental particle predicted by the Standard Model (SM) of particle physics, since the top quark was discovered in 1995, at the Tevatron [8, 9]. In July 2012, the CMS and ATLAS experiments at the LHC announced the discovery of a new particle with a mass close to $125 \, \text{GeV}$ [10, 11], as a result of their searches for the predicted boson. Since this discovery, many analyses have been carried out with the purpose of studying the properties of the new particle. Continuing these studies exhaustively is of fundamental importance to reveal the extent to which this boson is compatible with the SM Higgs boson. In this sense, these properties studies are stringent tests to the validity of the SM. Ultimately, they may reassure our confidence in the SM as the most complete theory of particle physics, or may uncover divergences from the theory expectations, leading to exciting new developments.

In the SM, the mass of elementary fermions is given by a term proportional to their Yukawa coupling, which also characterizes the interaction of fermions with the Higgs boson. A verification of this proportionality between fermion mass and interaction with the Higgs boson is an important test to the electroweak symmetry breaking mechanism. The top quark, being by far the most massive elementary fermion, has the largest Yukawa coupling, which is close to 1. The fact that this coupling is so large means that, if the SM is a valid theory up to the Planck scale ($M_P = 1.22 \times 10^{19} \, \text{GeV}$), this coupling gives important contributions to the Higgs potential, and has an impact on the allowed region of vacuum stability scenarios [12, 13]. A direct measurement of the top quark Yukawa coupling can be obtained from the cross-section of associated production of a Higgs boson with a pair of top quarks ($t\bar{t}H$) at the LHC. The presence of an interaction vertex between a top quark and a Higgs boson in this process also gives it the desirable property

of being sensitive to the CP (charge-parity) nature of the Higgs boson [14]. This means that from the reconstruction of the final state of this process, one could be able to determine which of the following scenarios corresponds to the produced Higgs boson: a scalar (CP-even) particle, as in the SM; a pure pseudoscalar (CP-odd) particle, which means the Higgs field would change sign under a combined transformation of charge and parity; or a mixed state (CP-violating) particle.

This thesis presents a phenomenological study in which the search for $t\bar{t}H$ production at the LHC, as well as the measurement of a possible pseudoscalar component of the Higgs boson, are discussed. The proposed search focuses on the final state in which the $t\bar{t}$ system decays dileptonically ($t\bar{t} \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}\nu\bar{\nu}\ell^+\ell^-$, where ℓ stands for either a muon or electron), and the Higgs boson decays to $b\bar{b}$. Two features differentiate this analysis. The first is a tool for reconstructing the undetectable neutrinos and the intermediate particles of the $t\bar{t}$ decay chain. The second is an extensive set of new angular observables, proposed as discriminants between $t\bar{t}H$ and the expected backgrounds at the LHC, as well as discriminants between two different Higgs boson CP scenarios (scalar and pseudoscalar).

In section 1, an introduction to the physics of the Higgs boson is given. The SM of particle physics is reviewed, emphasizing the need for an electroweak symmetry breaking mechanism, and its inevitable consequence: the existence of a Higgs boson. Some theoretical predictions for Higgs boson production at the LHC, namely rates of production and decay, are discussed. The ATLAS and CMS analyses which resulted in the 125 GeV Higgs boson discovery are presented. A set of results summarizing the experimental state of the art in studies of Higgs boson properties is also discussed. Finally, some arguments are given for the importance of $t\bar{t}H$ production in studies of Higgs boson properties and in tests to the validity of the SM.

In section 2, the experimental setup in which the observation of $t\bar{t}H$ production can currently take place is discussed. The LHC operation is described, as well as the ATLAS [15] and CMS [16] detectors. The ATLAS experiment, in particular, is described in considerable detail, as an example of a modern, general-purpose particle physics experiment. Since the present study was based on simulated events, the tool used for detector simulation, DELPHES [17], is also described in this section.

The body of the analysis and the different studies performed are described in

section 3, starting with the generation of the simulated event samples and the parameters used for the fast simulation of a generic detector. The expected final state of signal events is defined and the main expected background processes at the LHC are listed. Then, the developed reconstruction routine is described, mostly focusing on the kinematic fit performed in order to reconstruct the neutrinos and the intermediate heavy particles in signal events. The description also includes the selection cuts applied and the truth-matching procedure used to match generated particles to the corresponding reconstructed objects.

The use of new angular observables is a central feature of the proposed analysis. The motivation for the observables is given, as well as the details for their computation. A procedure for tagging one of the top quarks in each event as having radiated the Higgs boson is proposed, which allows the definition of a whole new set of different angular observables. A study of the discrimination obtained between signal and background using the new observables is described. This study relies on a series of multivariate analyses in order to compare the signal sensitivity provided by different sets of observables. Finally, some observables proposed as discriminants between different Higgs boson CP scenarios in publications prior to this work are presented and compared to the new angular observables.

In section 4, the results of the studies described before are shown and discussed. Regarding the search for $t\bar{t}H$ production, the distributions of the new observables which exhibit best discrimination between $t\bar{t}H$ events and background events are shown, as well as the results of the different multivariate analyses performed. Regarding the measurement of a pseudoscalar component of the Higgs boson, a selection is also presented of distributions of the new angular observables showing the best discrimination between different CP scenarios. The observables collected from the literature review are also replicated, and the effects on discrimination, from simulating a detector and performing a kinematic fit, are shown.

The conclusions of this work are given in section 5. It is argued that reconstruction of the $t\bar{t}$ system may be advantageous in searches for $t\bar{t}H$ production. Many of the new angular observables are remarkably discriminant between $t\bar{t}H$ and backgrounds. The multivariate analysis tests show that combinatorial background (events in which the assignment of reconstructed jets to quarks is done incorrectly) currently limits the potential gain in signal sensitivity. Many of the

new angular observables proved to be good discriminants between the two different Higgs CP scenarios studied. Only a few of the observables collected from prior studies remain as good as those, after the detector effects and the kinematic fit. Still, all the ones that remain good discriminants require reconstruction of the $t\bar{t}$ system. Some proposals for improving the work presented here and comments about applying it to future data analyses are also provided.

1   Introduction

In the first part of this introductory section, the theoretical framework leading to the prediction of the Higgs boson is discussed. A brief review of the SM of particle physics is presented. The particle content of this theory is exposed, as well as its underlying group structure and the lagrangian density, which governs the particle interactions and kinematics. The most important limitations of the SM are pointed out. The need for a mechanism leading to the electroweak symmetry breaking is explained, and illustrated by a simplified example of a Higgs mechanism.

In the second part of this section, experimental aspects are addressed. The main Higgs boson production processes at the LHC are presented, as well as its main decay channels. Then, the analyses which eventually led to the discovery of the 125 GeV Higgs boson in 2012 are summarized. Some results of other analyses are also discussed, regarding the properties of the new particle, including production and decay rates, couplings to other particles, and spin-parity configuration.

In the end of the section, a summary of properties of the top quark is presented. The importance of a precise measurement of the top quark Yukawa coupling is given as the leading motivation for studying the $t\bar{t}H$ production process.

1.1   Standard Model of particle physics

The SM of particle physics is a theory that describes elementary particles and all their interactions, with the exception of gravity [18–21]. Fundamental particles in the SM are divided into fermions, gauge bosons and the Higgs boson.

Fermions are spin-$^1/_2$ particles and are divided into leptons and quarks, each set organized in three families. The first family of leptons is composed of the electron (e^-), with negative electric charge $-e$ and the electron neutrino ν_e, with zero charge. The first family of quarks is composed of the up (u) quark, with charge $+^2/_3|e|$ and the down (d) quark, with charge $-^1/_3|e|$. The second and third families of leptons are each composed of a charged lepton and a neutrino. Analogously, each of the three families of quarks is composed of an up-type quark and a down-type quark. Quarks in the second and third families have the same

charge as their counterparts in the first family.

Fermion masses (or equivalently, Yukawa couplings) are free parameters of the SM. The masses of neutrinos are an exception to this, since neutrinos are assumed to be massless in the SM. This assumption was recently proven to be wrong. Albeit small, neutrino masses must be non-zero to accomodate for the experimental observation of neutrino oscillations [22]. For this experimental result, the 2015 Nobel Prize in Physics was awarded to Takaaki Kajita and Arthur B. McDonald, who led the Super-Kamiokande and SNO collaborations, respectively, where the oscillations were first observed [23]. Concerning the fermions which are massive in the SM, a mass hierarchy is evident among quark families, and among leptonic families, with every fermion in the third family being more massive than the corresponding one in the second family, which in turn is more massive than its first family partner.

The gauge bosons are spin-1 particles and, within the SM, they play the role of force carriers. The photon (γ) is massless, electrically neutral and is responsible for carrying the electromagnetic force, thus interacting with all electrically charged particles. The gluons (g) are also massless and electrically neutral, and they mediate the strong nuclear force. The gluons interact with all particles carrying colour charge. This includes all quarks, but also gluons themselves. The weak force carriers are the Z, the W^+ and the W^- bosons. The Z boson is electrically neutral, while the $W^\pm$ bosons have charge $\pm|e|$. They are massive particles, which explains the short range of the weak force.

The Higgs boson is the only scalar (i.e., spin-0) fundamental particle in the SM. It is electrically neutral and is not a force carrier. Instead, it plays a central role in electroweak symmetry breaking (EWSB) and generates the masses of fundamental particles, in the context of a mechanism further exposed in the next section.

For every elementary particle, a corresponding anti-particle exists, with the same mass and opposite-sign quantum numbers. A scheme of the particle content of the SM is shown in figure 1.

Formally, the SM is a quantum field theory based on the local gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. This gauge group includes the symmetry group of the strong interactions, $SU(3)_C$, and the symmetry group of the electroweak interactions, $SU(2)_L \times U(1)_Y$. The gauge fields corresponding to the $SU(3)_C$ group are

6
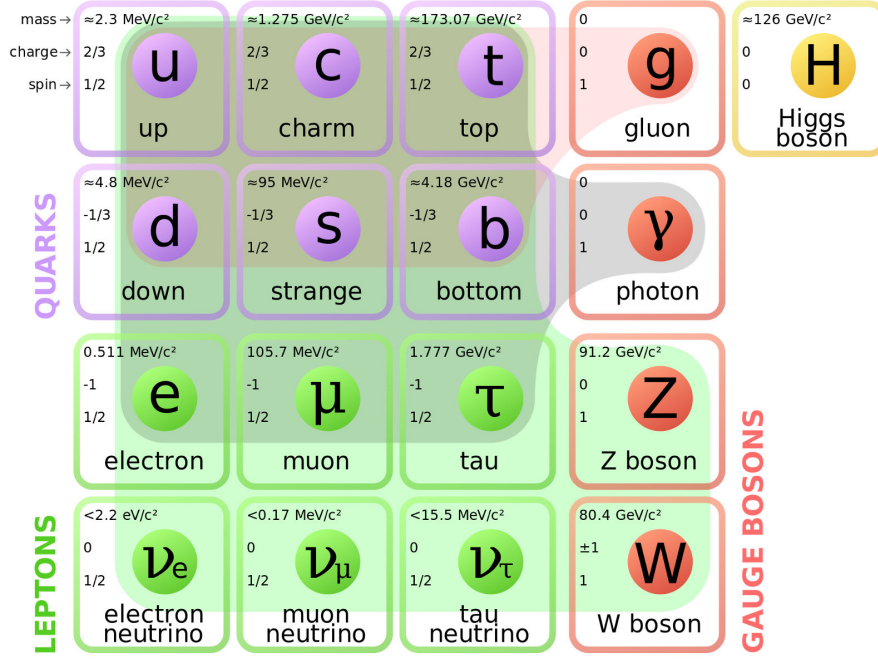
Figure 1: Particle content of the SM, with the three families of fermions, the gauge bosons, and the Higgs boson. Shaded areas surround each gauge boson and the fermions that interact with it [24]. The mass hierarchy in the fermionic families is evident.

G_μ^α ($\alpha = 1, ..., 8$), and can be identified with the gluon fields, one for each of the 8 independent bicoloured kinds of gluons. The gauge fields of the $SU(2)_L$ group are W_μ^i (index i can be $1, 2$ or 3) and the field of the $U(1)_Y$ group is B_μ. After the EWSB, linear combinations of the fields W_μ^1, W_μ^2, W_μ^3 and B_μ can be identified with the physical photon, Z boson and $W^\pm$ bosons. It is in this sense that the electromagnetic and the weak interactions are said to be unified in the SM.

The full SM lagrangian density (usually referred to simply as the 'lagrangian') can be written as

$$\mathcal{L}_{SM} = \mathcal{L}_{Gauge} + \mathcal{L}_f + \mathcal{L}_{SSB} + \mathcal{L}_{Yukawa}. \tag{1}$$

$\mathcal{L}_{Gauge}$ contains the kinetic and self-interacting terms of the gauge fields:

$$\mathcal{L}_{Gauge} = -\frac{1}{4} G_{\mu\nu}^\alpha G_\alpha^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \tag{2}$$

This is written in terms of the field strength tensors

$$G^{\alpha}_{\mu\nu} = \partial_{\mu}G^{\alpha}_{\nu} - \partial_{\nu}G^{\alpha}_{\mu} + g_s f^{\alpha\beta\gamma}G_{\mu\beta}G_{\nu\gamma} \tag{3}$$

$$W^{i}_{\mu\nu} = \partial_{\mu}W^{i}_{\nu} - \partial_{\nu}W^{i}_{\mu} + g\epsilon^{ijk}W^{j}_{\mu}W^{k}_{\nu} \tag{4}$$

$$B_{\mu\nu} = \partial_{\mu}B_{\nu} - \partial_{\nu}B_{\mu}, \tag{5}$$

where $f^{\alpha\beta\gamma}$ and ϵ^{ijk} are the structure constants of the respective groups. They are necessary in order to preserve gauge invariance whenever the generators of a group do not commute.

$\mathcal{L}_f$ contains the terms that couple each fermion to the gauge fields. It can be written as:

$$\begin{aligned}
\mathcal{L}_f =& \bar{L}\gamma^{\mu}(i\partial_{\mu} - g\frac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{W_{\mu}} - g'\frac{Y}{2}B_{\mu} - g_s\boldsymbol{T}\cdot\boldsymbol{G_{\mu}})L \\
&+ \bar{R}\gamma^{\mu}(i\partial_{\mu} - g'\frac{Y}{2}B_{\mu} - g_s\boldsymbol{T}\cdot\boldsymbol{G_{\mu}})R,
\end{aligned} \tag{6}$$

where L and R are the left-handed and right-handed chirality components of the fermion fields, respectively, and a sum is implicit over all leptons and quarks. The $i\partial_{\mu}$ term is the kinetic term of the fermion field. The second term rules the interaction with W^{i}_{μ}, which are the weak fields of the $SU(2)_L$ group. The corresponding coupling constant is g and $\frac{\tau_i}{2}$ are the generators of the group, also called the components of the isospin operator, in analogy to the spin operator. The L in $SU(2)_L$ stands for "Left", meaning that only left-handed chirality fermions transform as isospin doublets under action of the group. Right-handed fermions transform as isospin singlets and, as a result, they don't interact with the W^{i}_{μ} fields. After the EWSB, this leads to $W^{\pm}$ bosons only coupling to left-handed fermions. The third (and the sixth) term governs the interaction with the B_{μ} field. The associated coupling constant is g', and Y is the generator of the group, known as the hypercharge quantum number of the fermion field. The isospin, the electric charge and the hypercharge quantum numbers for the elementary fermions are listed in table 1. The fourth (and the last) term in $\mathcal{L}_f$ contains the interaction with the gluon fields G^{α}_{μ}. The strong coupling constant is g_s, and T_{α} are the corresponding generators of the group. Quarks transform as triplets of colour under $SU(3)_C$. Since leptons transform as singlets ($\boldsymbol{T} = \boldsymbol{0}$), they do not take part

8

Table 1: Quantum numbers of elementary fermions [18]: τ and τ_3 are the total isospin and the third component of the isospin, respectively, Q is the electric charge in units of $|e|$, and Y is the hypercharge. The symbol e stands for any charged lepton, ν for any neutrino, u for any up-type quark and d for any down-type quark.

Fermion	τ	τ_3	Q	Y
ν_L	1	1	0	-1
e_L	1	-1	-1	-1
e_R	0	0	-1	-2
u_L	1	1	$\frac{2}{3}$	-1
d_L	1	-1	$-\frac{1}{3}$	-1
u_R	0	0	$\frac{2}{3}$	-1
d_R	0	0	$-\frac{1}{3}$	-1

in strong interactions.

$\mathcal{L}_{SSB}$ is the spontaneous symmetry breaking lagrangian:

$$\mathcal{L}_{SSB} = |(i\partial_\mu - g\frac{1}{2}\boldsymbol{\tau} \cdot \boldsymbol{W_\mu} - g'\frac{1}{2}B_\mu)\phi|^2 + \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2, \qquad (7)$$

where ϕ is the Higgs field. The terms inside the squared absolute value include the Higgs kinetic term and its interactions with the electroweak fields. The SM Higgs field is an isospin doublet, thus interacting with the W_μ^i fields. The negative of the last two terms is collectively referred to as the Higgs potential (μ^2 and λ must be real and > 0). This potential is such that it allows for the EWSB to occur, giving the Higgs field a non-zero vacuum expectation value, as explained below.

$\mathcal{L}_{Yukawa}$ is the lagrangian confering mass to all fermions. It can be written as

$$\mathcal{L}_{Yukawa} = -(K_1\bar{L}\phi R + K_2\bar{L}\phi_C R + \text{h.c.}), \qquad (8)$$

where $\phi_C = -i\tau_2\phi^*$ and transforms identically to ϕ. The couplings K_1 and K_2 are chosen such that, after EWSB, this originates mass terms for all fermions. Another consequence of this term is an interaction term between each fermion and the Higgs boson, proportional to the fermion mass.

The SM has proven very successful at making predictions up to energies in the few TeV scale. One remarkable result is the agreement between the quantum electrodynamics (QED) prediction and the experimental determination of

the electron magnetic moment. This quantity has been measured with a precision of 0.28×10^{-12} and the difference to the QED computed value is smaller than 5×10^{-12} [25]. Within the electroweak theory, the prediction of the masses of the Z and $W^{\pm}$ bosons before they were discovered is also noteworthy [26, 27].

In spite of its success, the SM is not a complete description of all fundamental interactions in nature. It does not account for gravity at all and, even within the scope of the interactions it describes, there are phenomena that remain at least partially unexplained. A few of those are enumerated here.

The experimentally verified oscillations and non-zero masses of neutrinos [22] are not predicted by the SM. Extensions of the SM exist that do accomodate these features, but it is not consensual which one best describes the structure of the neutrino fields. Another unsolved issue is the strong CP problem [28]. According to the SM, the strong force is allowed to violate the CP symmetry, i.e., it is not necessarily invariant under a combined transformation of charge and parity. However, CP violation has never been observed experimentally in strong interactions, and very small upper limits have been set to the free parameter of the SM which dictates its rate, suggesting that either the parameter is very finely tuned, or that an additional mechanism is required that imposes CP conservation in the strong sector. Also related to the issue of CP violation is the asymmetry between existing matter and antimatter in the Universe. If equal amounts of matter and antimatter existed immediately after the Big Bang, CP violation could provide an explanation for the present abundance of matter in the Universe, in contrast to the scarcity of antimatter. However, the currently measured rates of CP violation in weak interactions from neutral kaon decay do not seem to be consistent with the verified asymmetry [29].

Another puzzling issue is the naturalness of the Higgs boson mass. The SM predicts corrections to the Higgs boson mass proportional to the square of the renormalization scale of the theory. If the SM is a complete theory up to the Planck scale ($M_P = 1.22 \times 10^{19}\,\text{GeV}$), at which gravity becomes important in particle interactions, these very large corrections must nearly cancel each other out in order for the observed Higgs boson mass to lie in the weak scale ($\sim 10^2\,\text{GeV}$). If one wishes to avoid the unnaturalness of this near cancellation, the renormalization scale must be chosen at a few TeV. This is only justifiable if new physical

10

processes beyond the Standard Model (BSM) are introduced at this scale, with additional symmetries "protecting" the Higgs boson mass from larger corrections. In fact, this naturalness problem has motivated BSM theories such as supersymmetry (SUSY) [30], technicolour [31], and theories predicting the existence of extra dimensions [32].

1.2 Electroweak symmetry breaking and the Higgs mechanism

After the established success of QED in the 1950s, gauge theories were extensively studied as models for elementary particles and their interactions. Salam, Weinberg and Glashow [33–35] are credited for having unified the electromagnetic and weak forces.

A challenge posed by this unification is the origin of mass of the weak gauge bosons. If mass terms for the gauge bosons are introduced 'ad-hoc' in the lagrangian of the theory, the gauge invariance is lost.

An alternative mechanism to avoid introducing symmetry-breaking terms directly into the lagrangian is a spontaneous symmetry breaking. Such a symmetry breaking occurs when a theory has a degenerate set of minimal energy (i.e. vacuum) states. The symmetry breaking occurs because the system evolves into one particular state among the degenerate set. That particular vacuum state is no longer invariant under the same gauge transformations as the original lagrangian. The lagrangian can then be re-written, replacing the original fields in the theory with their constant vacuum expectation values and a perturbation relative to those values. In this way, the fundamental lagrangian keeps the original symmetry, and the re-written lagrangian contains the desired mass terms for the gauge bosons. However, if the broken symmetry is continuous, the symmetry breaking should give rise to Goldstone particles, which are massless bosons, not consistent with observation [36–38].

In 1964, three independent groups (Peter Higgs [1–3]; Robert Brout and François Englert [4]; Gerald Guralnik, C. R. Hagen and Tom Kibble [5, 6]) proposed a mechanism to explain the spontaneous breaking of the electroweak symmetry. This mechanism introduced the now called Higgs field, a scalar field with a non-zero

expectation value in the vacuum, thus breaking the gauge invariance of the ground state of the theory. This mechanism ensures that the degrees of freedom of the Goldstone bosons are "absorbed" by the longitudinal polarization states of the vector bosons, allowing them to acquire mass, and leaving an additional degree of freedom which corresponds to the mass of the Higgs boson itself. Through the Yukawa interaction terms, the Higgs field also gives mass to all fundamental fermions. The existence of the Higgs boson, a massive and scalar particle, is a verifiable prediction of the mechanism. Decisive proof that the electroweak theory is renormalizable came only later, from 't Hooft and Veltmann [39].

1.2.1 Spontaneous breaking of U(1) symmetry

In this section, the example of spontaneous breaking of an $U(1)$ local gauge symmetry is given [18]. In the SM, the relevant broken symmetry is $SU(2)$, and the minimal Higgs mechanism requires a scalar doublet. The $U(1)$ case requires only a singlet and can be presented more briefly, while illustrating the key aspects of the Higgs mechanism. We consider a scalar complex field $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$, where ϕ_1 and ϕ_2 are real. The lagrangian of the theory is

$$\mathcal{L} = (\partial_\mu \phi)^*(\partial^\mu \phi) + \mu^2 \phi^* \phi - \lambda(\phi^* \phi)^2, \tag{9}$$

where λ and μ^2 are assumed real and positive in order to generate the spontaneous symmetry breaking. If we impose invariance under the local transformation $\phi \to e^{i\alpha(x)}\phi$, the introduction of a gauge field A_μ is necessary. The derivative ∂_μ must be replaced with the covariant derivative $D_\mu = \partial_\mu - ieA_\mu$ and the gauge field must transform as $A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha$. The gauge-invariant lagrangian becomes

$$\mathcal{L} = (\partial^\mu + ieA_\mu)\phi^*(\partial_\mu - ieA_\mu)\phi + \mu^2 \phi^* \phi - \lambda(\phi^* \phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{10}$$

where $F_{\mu\nu}$ is the field strength tensor defined from A_μ as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The Higgs potential is $V(\phi) = -\mu^2|\phi|^2 + \lambda|\phi|^4$, which is minimum along the line $|\phi|^2 = \mu^2/2\lambda \equiv v^2$. An example of a Higgs potential is plotted in figure 2.

In terms of ϕ_1 and ϕ_2, we can write $\phi_1^2 + \phi_2^2 = v^2$. Without loss of generality, let us consider the case where $\phi_1 = v$ and $\phi_2 = 0$. Expanding the field ϕ about
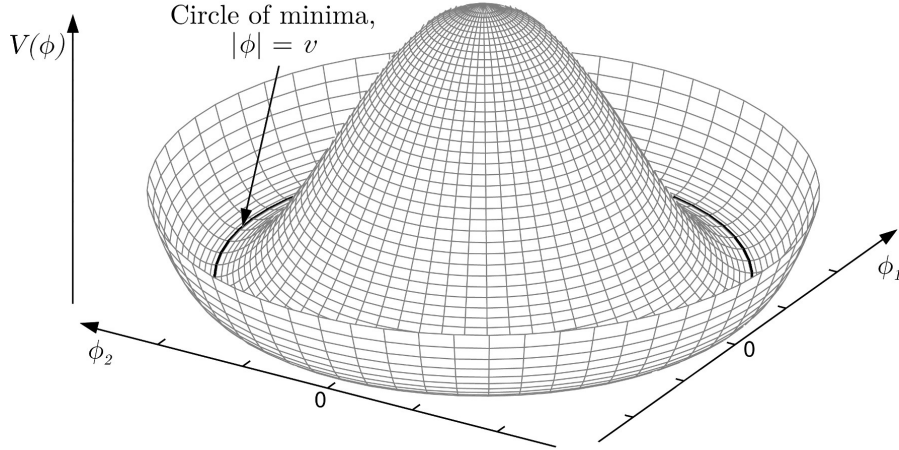
Figure 2: Higgs potential $V(\phi)$, often called the "Mexican hat potential", due to its shape.

the vacuum, using the real fields η and ξ, yields

$$\phi(x) = \frac{1}{\sqrt{2}}[v + \eta(x) + i\xi(x)], \tag{11}$$

and for the lagrangian,

$$
\begin{aligned}
\mathcal{L}' = {} & \frac{1}{2}(\partial_\mu \xi)^2 + \frac{1}{2}(\partial_\mu \eta)^2 - v^2\lambda\eta^2 + \frac{1}{2}e^2 v^2 A_\mu A^\mu \\
& - ev A_\mu \partial^\mu \xi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \text{ interaction terms.}
\end{aligned}
\tag{12}
$$

At this point, the mass term for the η field is evident, with $m_\eta = \sqrt{2\lambda v^2}$. More importantly, the symmetry breaking was successful in generating the mass term for the vector boson A_μ, with $m_A = ev$. However, ξ remains massless, as expected from a Goldstone boson. By confering mass to the vector boson, we allow it to have a longitudinal polarization, thus increasing by 1 the number of polarization degrees of freedom of the theory. Such a change should not be allowed by simply making a transformation of the field variables. The off-diagonal term proportional to $A_\mu \partial^\mu$ motivates us to make a different gauge transformation. To first order in ξ,

$$\phi = \frac{1}{\sqrt{2}}(v + \eta + i\xi) \simeq \frac{1}{\sqrt{2}}(v + \eta)e^{i\xi/v}. \tag{13}$$

13

This suggests we write ϕ in terms of new fields h and θ, and choose the gauge in which ϕ and A_μ transform as

$$\phi \to \frac{1}{\sqrt{2}}[v + h(x)]e^{i\theta(x)/v} \quad \text{and} \quad A_\mu \to A_\mu + \frac{1}{ev}\partial_\mu\theta, \tag{14}$$

with $\theta(x)$ uniquely determined so that h is real, anticipating the vanishing of θ from the theory. Substituting in the original lagrangian gives

$$\mathcal{L}'' = \frac{1}{2}(\partial_\mu h)^2 - \lambda v^2 h^2 + \frac{1}{2}e^2 v^2 A_\mu^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (h, A_\mu) \text{ interaction terms.} \tag{15}$$

Finally, it is clear that the Goldstone boson does not appear as a physical particle: the apparent degree of freedom corresponded to the choice of gauge transformation. The lagrangian describes only two interacting particles: a massive scalar h and a massive vector A_μ. It is in this sense that the Goldstone boson is said to be "absorbed" by the longitudinal polarization of the gauge vectors.

It has been mentioned that fermions acquire mass through the Yukawa couplings to the Higgs field. Before symmetry breaking, these terms have the form shown in equation (8) for $\mathcal{L}_{Yukawa}$. But when the scalar field acquires a non-zero vacuum expectation value v, the term for a fermion field ψ can be written as

$$\mathcal{L}'_{Yukawa} = -\frac{y_f}{\sqrt{2}}\bar{\psi}\psi(v + h) = -\frac{y_f v}{\sqrt{2}}\bar{\psi}\psi - \frac{y_f}{\sqrt{2}}\bar{\psi}\psi h \tag{16}$$

where y_f is the Yukawa coupling of the corresponding fermion. The first term is the mass term of the fermionic field, while the second is the interaction term between the fermion and the Higgs boson. The mass of the fermion is given by $m_f = y_f v/\sqrt{2}$. The interaction term can then be written as $-(m_f/v)\bar{\psi}\psi h$, making more clear the proportionality between fermion mass and interaction with the Higgs field.

1.3 Higgs boson physics at the LHC

The Higgs boson production in proton-proton (pp) collisions at the LHC is expected to occur through four main processes [40–42]: the one with the largest cross-section is gluon fusion (ggF), the next is vector boson fusion (VBF), followed
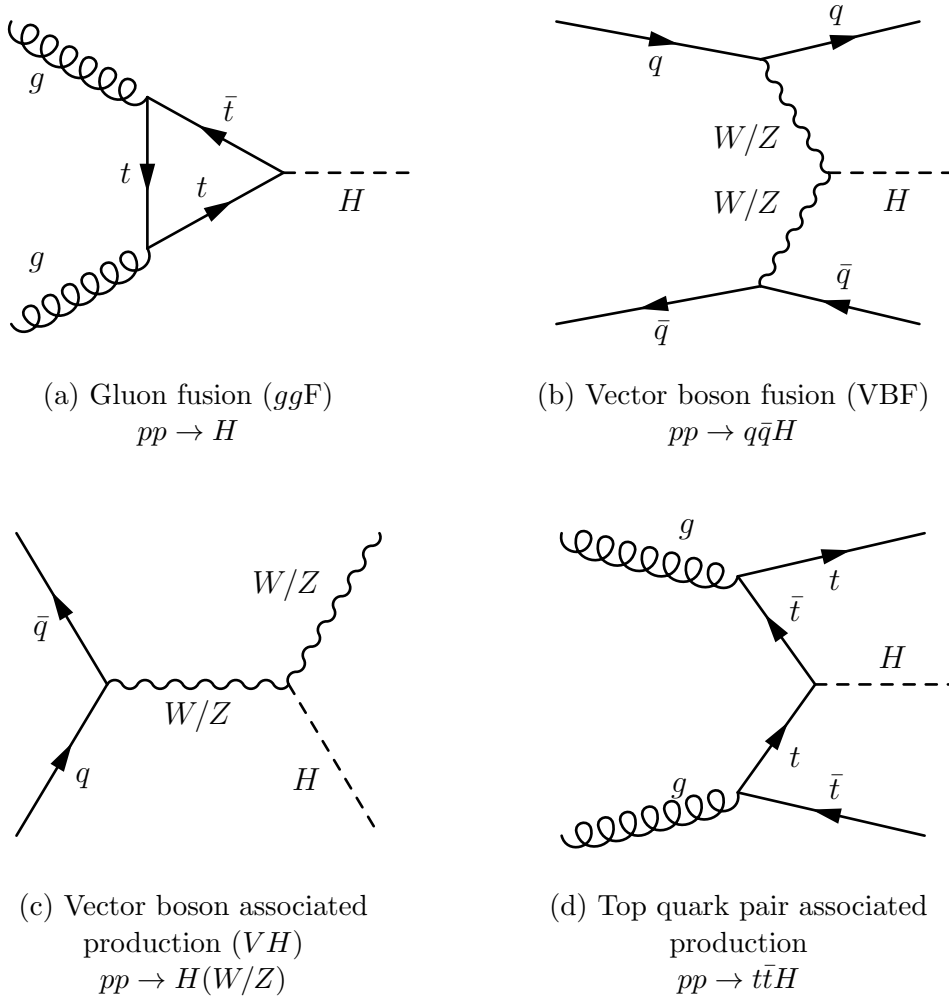
(a) Gluon fusion (ggF)

$pp \to H$

(b) Vector boson fusion (VBF)

$pp \to q\bar{q}H$

(c) Vector boson associated

production (VH)

$pp \to H(W/Z)$

(d) Top quark pair associated

production

$pp \to t\bar{t}H$

Figure 3: Main processes for Higgs boson production at the LHC. In **(a)**, although the top quark dominates, there are contributions from loops of every massive fermion.

in turn by associated vector boson production (ZH and WH) and, finally, associated top quark pair ($t\bar{t}H$) production. Examples of typical LHC leading order production diagrams are shown in figure 3. The theoretical cross-section for each process in pp collisions is plotted as a function of the centre-of-mass energy in figure 4 [43], assuming a SM Higgs boson with mass $m_H = 125\,\text{GeV}$.

With a mass of 125 GeV, the Higgs boson branching ratios (BR) are 57.7% to a b-quark pair ($b\bar{b}$), 21.5% to WW, 8.57% to a gluon pair (gg), and 6.32% to $\tau\tau$ [43]. Less common but relevant decay modes include $H \to ZZ$ (2.64%) and $H \to \gamma\gamma$ (0.228%). Theoretical values for the most important branching ratios are listed in table 2 [43].

15
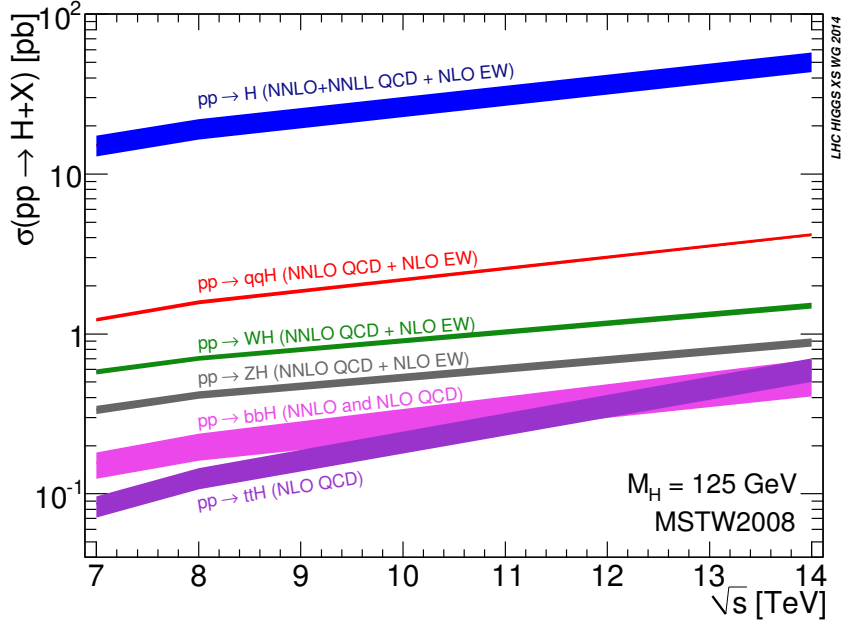
Figure 4: Theoretical cross-sections for the different processes for SM Higgs boson production in *pp* collisions as a function of the centre-of-mass energy [43]. A Higgs boson a mass of 125 GeV is assumed. The line widths correspond to theoretical uncertainties.

Table 2: Theoretical values for the Higgs boson branching ratios, assuming a SM Higgs boson mass of 125 GeV [43].

Decay channel	$b\bar{b}$	WW	gg	$\tau\tau$	$c\bar{c}$	ZZ	$\gamma\gamma$	Others
BR (%)	57.7	21.5	8.57	6.32	2.91	2.64	0.228	< 0.2

Both the production cross-sections and the decay branching ratios shown above were calculated for a Higgs boson mass of $125\,\mathrm{GeV}$. It should be stressed, however, that before the discovery of the $125\,\mathrm{GeV}$ boson, there was no particular reason to believe that the Higgs boson mass was close to this value. Expected production and decay rates change with the Higgs boson mass, and searches have to be inclusive enough to accomodate for all possibilities. For example, for a Higgs boson with a mass just above twice the W boson mass, the decay into WW would be the dominant one, and the $b\bar{b}$ decay fraction would be much smaller than it is for a light Higgs boson.

Even for a light Higgs boson, the $H \rightarrow b\bar{b}$ decay is not the preferred one for discovery or mass measurements at the LHC. The products of the decay are indistinguishable from the multijet background, a problem that can be partially overcome if the analysis is performed in one of the rare associated production channels (VH or $t\bar{t}H$), resulting however in a much smaller number of events. Besides, the experimental energy resolution of jets is worse than that of leptons and photons, resulting in a weaker resolution of the reconstructed Higgs boson mass. As for $H \rightarrow WW$, even if one or both of the W bosons decay leptonically, there are neutrinos in the final state that go undetected and can only be reconstructed from missing transverse energy in the calorimeters, a process that brings the same resolution issue. The decay channels $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ (4ℓ stands for final states containing 2 pairs of same-flavour charged leptons) are relatively rare. However, they produce unique final states, with leptons and photons, particles which are detected with excellent energy resolution. Therefore, $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ provide the best resolution in Higgs boson mass measurements [44, 45], and the new particle was expected to produce a narrow peak in the reconstructed mass, over the smooth background. For this reason, they were used in the Higgs boson searches at the LHC.

1.4 Higgs boson discovery and experimental status

The discovery of the Higgs boson and, more generally, a better understanding of the electroweak symmetry breaking, is one of the primary goals for the operation of the Large Hadron Collider [7] and of the general-purpose LHC experiments

ATLAS [15] and CMS [16]. On the 4th of July 2012, both the ATLAS and CMS collaborations announced, as a result of their searches for the SM Higgs boson, the discovery of a new resonance with mass close to 125 GeV, consistent with the predicted particle [10, 11]. In 2013, the Nobel Prize in Physics was granted to Peter Higgs and François Englert "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider" [46].

Following this discovery, the properties of the new particle are being studied on its many production and decay channels, to further check its compatibility with the SM Higgs boson, and also to constrain SM extensions. The results of some of these measurements are presented below, all of which showing that the discovered particle is so far consistent with the SM predictions for the Higgs boson.

The $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ channels were analysed by both CMS and ATLAS to perform measurements of the Higgs boson mass [44, 45]. Figure 5 shows two histograms of the invariant mass of the particles expected as products of the Higgs boson decay. In both plots, the excess of events in the Higgs boson mass region is evident. The plot on the left is from the CMS analysis of the $H \rightarrow ZZ^* \rightarrow 4\ell$ channel [47], and the plot on the right is from the ATLAS analysis of the $H \rightarrow \gamma\gamma$ channel [44].

The collaborations later combined the results to obtain the currently most precise measurement [48]:

$$m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.}) \, \text{GeV}.$$

Searches for higher mass resonances decaying to a photon pair were performed on the data delivered during the LHC Run 2, at a centre-of-mass energy of 13 TeV. Both ATLAS [49] and CMS [50] presented results hinting at the existence of a resonance with a mass of about 750 GeV. The local significance of this signal is at the 3.6σ level in the case of the ATLAS collaboration, and 2.6σ for the CMS collaboration. Such significances are far from excluding the possibility of statistical fluctuations. Nevertheless, the fact that both experiments reported an excess of
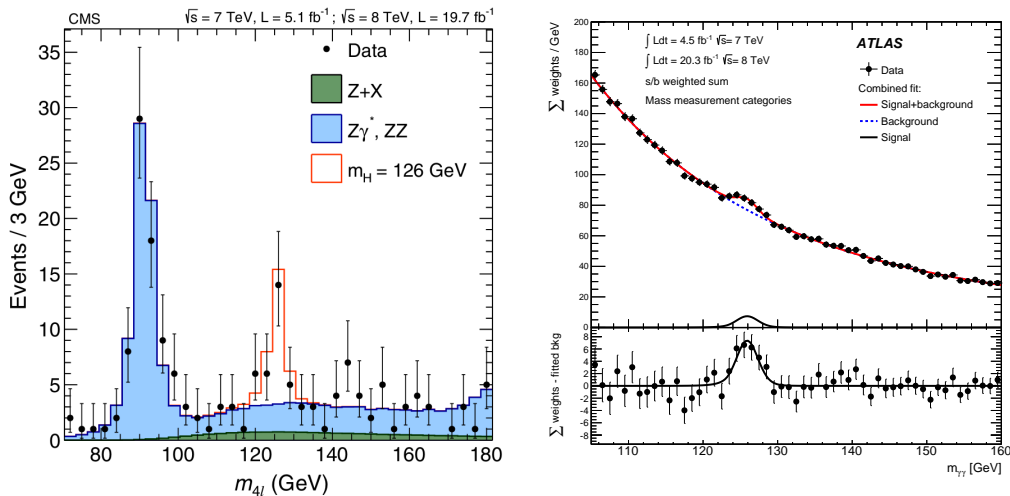
Figure 5: **Left:** CMS analysis of the $H \to ZZ^* \to 4\ell$ channel [47]. Distribution of the four-lepton reconstructed mass for the sum of the $4e$, $2e2\mu$, and 4μ channels for the mass region $70 < m_{4\ell} < 180$ GeV. Points with error bars represent the data, shaded histograms represent the backgrounds, and the unshaded histogram represents the signal expectation for a mass hypothesis of $m_H = 126$ GeV. Signal and the ZZ background are normalized to the SM expectation; the $Z + X$ background to the estimation from data. **Right:** ATLAS analysis of the $H \to \gamma\gamma$ channel [44]. Invariant mass distribution of 7 TeV and 8 TeV samples combined, showing weighted data points with errors, and the result of the performed fit. The fitted signal (s) plus background (b) is shown, along with the background-only and signal-only components of this fit. Events were selected for different categories, and summed together with a weight given by the s/b ratio in each category. The bottom plot shows the difference between the summed weights and the background component of the fit.

data in the same mass region has triggered numerous attempts by model-builders in finding SM extensions that predict the diphoton resonance. Interestingly, the excess is not straightforward to explain within the most commonly considered frameworks for BSM physics, like two-Higgs-doublet models or the minimal supersymmetric standard model [51].

In order to assert that the particle discovered in 2012 is an SM-like Higgs boson, it is not sufficient to discover a signal in the predicted production and decay channels. Alternative scenarios regarding particle properties such as spin and parity must be excluded with reasonable confidence. Besides, the production cross-sections, as well as the decay branching ratios, must be consistent with the values predicted by the SM, given the Higgs boson mass m_H.

In the SM, the Higgs boson is a spin-zero, CP-even particle ($J^P = 0^+$). For the Higgs boson discovered at the LHC, the spin-1 scenario is excluded by the Landau-Yang theorem [52, 53], which states that a diphoton decay is forbidden for a massive spin-1 particle. Measurements have been performed to further constrain

the spin-parity quantum numbers of the discovered Higgs boson. Analyses using the whole dataset of LHC Run 1 (with approximately $5\,\mathrm{fb}^{-1}$ of data at $\sqrt{s} = 7\,\mathrm{TeV}$ and $20\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 8\,\mathrm{TeV}$) were carried out by ATLAS [54] and CMS [55].

The ATLAS analysis used the decay channels $H \to \gamma\gamma$, $H \to ZZ^* \to 4\ell$ and $H \to WW^* \to e\nu\mu\nu$. It tested the SM hypothesis against spin-2 models and non-SM spin-0 models. All the alternative models were excluded at a confidence level greater than 99.9%. However, assuming the spin-0 scenario, the Higgs boson can be in a mixed state of the SM CP-even state and a non-SM CP-odd state. Using the $H \to ZZ^* \to 4\ell$ and $H \to WW^* \to e\nu\mu\nu$ channels, ATLAS gave an upper limit to the CP-odd fraction of the Higgs boson of $f < 0.41$, with a 95% confidence level. The CMS analysis used the decays $H \to ZZ^*$, $Z\gamma$, $\gamma^*\gamma^* \to 4\ell$, $H \to \gamma\gamma$ and $H \to WW^* \to \ell\nu\ell\nu$. It excluded the spin-two hypotheses with confidence levels above 99%. The pure pseudoscalar hypothesis was excluded at a 99.98%. The upper limit for the pseudoscalar Higgs fraction was set at $f < 0.43$ with a 95% confidence level, using the $H \to ZZ^* \to 4\ell$ events.

The CMS and ATLAS collaborations combined the results of measurements of Higgs boson production and decay rates using the available data from LHC Run 1 [56]. This comprehensive study included the production processes ggF, VBF, WH, ZH and $t\bar{t}H$, and the decay modes $H \to \gamma\gamma$, $H \to ZZ$, $H \to WW$, $H \to b\bar{b}$, $H \to \tau\tau$ and $H \to \mu\mu$.

The signal strength parameter is commonly used to express signal yields in such measurements. Signal strengths can be defined for a given production process i or for a decay channel f as

$$\mu_i = \frac{\sigma_i}{(\sigma_i)_{\mathrm{SM}}} \quad , \quad \mu^f = \frac{\mathrm{BR}_f}{(\mathrm{BR}_f)_{\mathrm{SM}}}, \tag{17}$$

where σ_i is the measured cross-section of the production process i ($i = gg$F, VBF, $WH, ZH, t\bar{t}H$) and BR_f is the measured branching ratio of the decay mode f ($f = H \to \gamma\gamma$, $H \to ZZ$, $H \to WW$, $H \to b\bar{b}$, $H \to \tau\tau$, $H \to \mu\mu$). The subscript SM refers to the respective SM predictions.

The number of events in which the Higgs boson is produced through process i and decays through channel f is proportional to $\mu_i \times \mu^f$, and the μ_i and μ^f values cannot be separately obtained from the number of such events alone. Even

the selection of events resulting from a particular production process i and decay channel f is not always possible. In the ATLAS and CMS combined analyses, events were classified in mutually exclusive categories, defined by selection cuts based on the final state objects. The number of signal events in each category k is given by

$$n_{signal}(k) = \sum_{i,f} \mu_i \times \mu^f N_i^f(k), \qquad (18)$$

where $N_i^f(k)$ is the theoretically expected number of events resulting from the production process i and decay channel f to be accepted in the k category. It is obtained from the luminosity, the detector acceptance, the analysis efficiency for that particular process, the predicted cross-section $(\sigma_i)_{\text{SM}}$ and the predicted branching ratio $(\text{BR}_f)_{\text{SM}}$.

In the ideal case, each category k would exclusively select signal events from one production process i and one decay channel f. For the decay channels, this ideal case is approached, since the definition of categories is based on the detected decay products. In the case of the production processes, however, the categories are much less pure, and there is important cross-contamination in most channels. Still, the μ_i and μ^f can be used as parameters to fit the obtained number of signal events in each category k.

A global signal strength can be obtained by requiring all production and decay signal strengths to be equal among themselves, as in the SM (in which, by definition, they are equal to 1). This reduces the number of fit parameters to a single one, μ. The value for μ obtained from the fit was:

$$\mu = 1.09^{+0.11}_{-0.10} = 1.09^{+0.07}_{-0.07}(\text{stat.})^{+0.09}_{-0.08}(\text{syst.}),$$

which shows remarkable agreement with the SM. However, the requirement that all signal strengths are equal makes this measurement strongly model dependent. To obtain values for each production and decay signal strength in a more model-independent way, the fit parameters were partially relaxed. In one fit, all the decay signal strengths were set to the SM value $\mu^f = 1$, and no condition was imposed on the production signal strengths μ_i. In another fit, all μ_i were set to 1 and the μ^f were the fit parameters. The results are shown in figure 6.
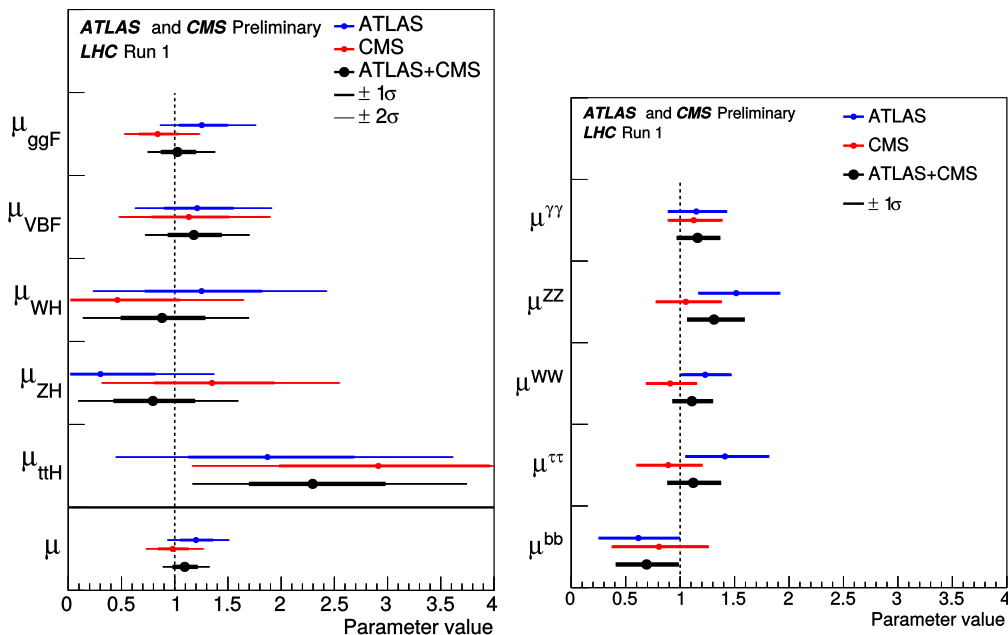
Figure 6: Best-fit results for the signal strengths for the combination of ATLAS and CMS, and the results for each experiment [56]. Error bars indicate the 1σ (thick) and 2σ (thin) intervals. **Left:** Production signal strengths. The measurements of the global signal strength μ are also shown. **Right:** Decay signal strengths.

A similar fit procedure was also used to obtain values of the Higgs couplings to other particles. This is accomplished by writing the production cross-sections and decay rates in terms of the individual couplings, at lowest order.

In the first fit, the parameters were the following couplings (normalized to the SM prediction): κ_Z, κ_W, κ_t, κ_b, κ_τ, κ_g, κ_γ and $\mathrm{BR_{BSM}}$. In this fit, κ_W and κ_Z were constrained to values ≤ 1. The $\mathrm{BR_{BSM}}$ parameter was introduced to account for possible undetected BSM decays of the Higgs boson. Besides, κ_g and κ_γ were included as effective couplings, since the Higgs does not couple at tree-level to neither gluons nor photons. Thus, these effective couplings are sensitive to contributions from BSM particle loops. A second fit was performed with the same parameters, but instead setting $\mathrm{BR_{BSM}}{=}0$ and letting the κ_W and κ_Z free. This fit is not sensitive to BSM decays, but could still reveal BSM loop contributions through κ_g and κ_γ. The result of these fits is shown on the left side of figure 7.

The third fit performed is even more constraining, expressing κ_g and κ_γ in terms of the fermion and weak boson couplings to the Higgs, as predicted from the SM loops. No BSM decay parameter is included. The coupling to the muon is
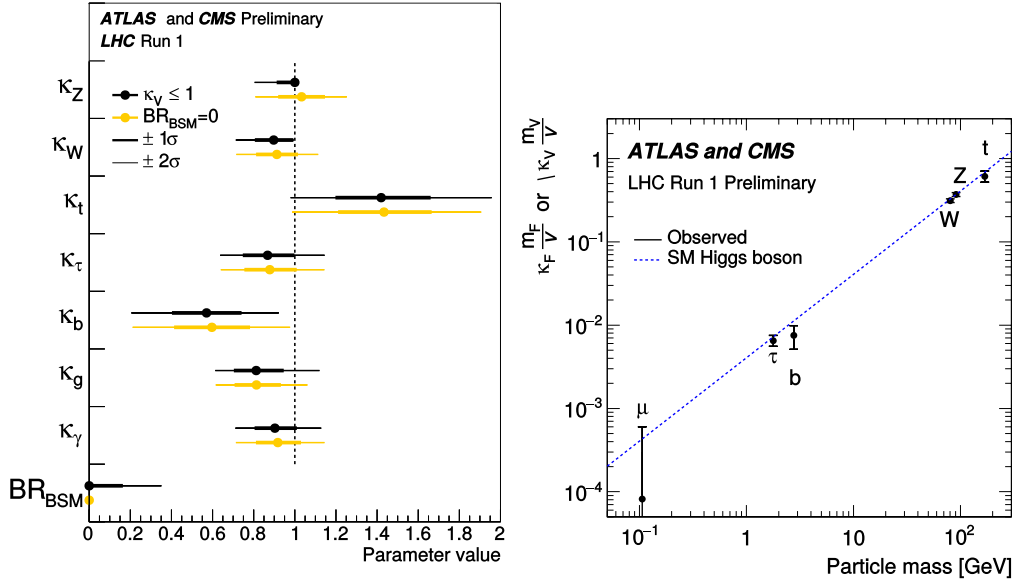
Figure 7: **Left:** Fit results for the two parameterisations allowing BSM loop couplings. In black, results with $\kappa_V \leq 1$, where V stands for Z or W, and in yellow, without additional BSM contributions to the Higgs boson width, i.e. $\mathrm{BR_{BSM}} = 0$. The measured results for the combination of ATLAS and CMS are reported together with their uncertainties [56]. Error bars indicate the 1σ (thick) and 2σ (thin) intervals. **Right:** Fit results for the combination of ATLAS and CMS in the case of the parameterisation with reduced couplings ($y_F = \kappa_F \frac{m_F}{v}$ for fermions and $y_V = \sqrt{\kappa_V} \frac{m_V}{v}$ for vector bosons) as a function of the particle mass [56]. The dashed line indicates the SM predicted dependence on the particle mass.

added as a parameter, resulting in the set of parameters: κ_Z, κ_W, κ_t, κ_b, κ_τ, κ_μ. This parameterisation is composed of the couplings to the massive vector bosons and to some of the heaviest fermions. From the Higgs mechanism, a linear relation between fermion couplings to the Higgs boson and fermion masses is expected, as well as a linear relation between vector boson couplings to the Higgs boson and the square of vector boson masses. Reduced couplings are defined for fermions as $y_F = \kappa_F \frac{m_F}{v}$, and are defined for vector bosons as $y_V = \sqrt{\kappa_V} \frac{m_V}{v}$, so that the same $\frac{y}{m} = v$ is expected for both fermions and vector bosons. The result of the fit is presented on the right side of figure 7.

1.5 Top quark

The top quark was the last discovered quark and completes the third family of elementary fermions. It is a spin-1/2 particle and has charge $+^2\!/_3|e|$, being an up-type quark. Both the top and bottom quarks were first theoretically proposed

by Makoto Kobayashi and Toshihide Maskawa, in a model aiming to describe CP-violating weak interactions in hadrons [57]. Top quark production was first observed in 1995 at the Tevatron by the DØ and CDF experiments [8, 9]. It is the most massive elementary particle in the SM. The world combination of top quark mass measurements [58] from DØ, CDF and the 2011 datasets of ATLAS and CMS yielded $m_t = 173.34 \pm 0.27(\text{stat.}) \pm 0.71(\text{syst.})$ GeV. In September 2015, CMS published a combination of its top quark measurements from the whole Run 1 dataset [59], yielding $m_t = 172.44 \pm 0.13(\text{stat.}) \pm 0.47(\text{syst.})$ GeV.

The top quark decay width has been measured by the DØ collaboration [60], with the result $\Gamma_t = 2.00^{+0.47}_{-0.43}$ GeV. Using the Heisenberg relation $\Delta E \Delta t \geq \hbar$, the top quark mean decay time can be quickly estimated to be of order 3.3×10^{-25} s. In fact, the DØ collaboration obtained $\tau_t = 3.29^{+0.90}_{-0.63} \times 10^{-25}$ s, which translates in an upper limit for the top quark lifetime of $\tau_t < 4.88 \times 10^{-25}$ s, at 95% confidence level. Such a lifetime is much shorter than that of any other quark and shorter than typical hadronization time ($\tau_{had} = \hbar/\Lambda_{QCD} \simeq 3 \times 10^{-24}$ s), which means the top quark is the only quark that decays from its "bare" state, instead of hadronizing.

The main decay channel for a top quark is the charged current $t \to W^+ b$ ($\bar{t} \to W^- \bar{b}$), with a probability very close to 1. Classification of the top quark decay is based on the decay of the resulting W boson. The decay is said to be leptonic if the W decays to a lepton and a neutrino, and hadronic if the W decays to a quark and anti-quark pair.

The short decay time of the top quark ensures that its spin state can be measured from the decay products. In the leptonic decay in particular, the resulting lepton in the final state preserves most of the spin information of the decaying top quark. This is in contrast with any other quark decay, in which the spin information is quickly degraded during hadronization, as hadronic bound states emerge and decay in succession [61]. Consequently, the polarization of a top quark pair at production level is fully transferred to the decay products of the system. Although $t\bar{t}$ pairs are unpolarized in production at hadron colliders, their spins are strongly correlated [62].

1.6 Importance of the $t\bar{t}H$ production process

The way fundamental fermions acquire mass in the Higgs mechanism is through Yukawa terms in the lagrangian that couple the fermion fields to the Higgs field. The top quark, being the most massive fermion, is then expected to have the largest Yukawa coupling to the Higgs field. This ensures that Higgs production channels involving fermion loops, such as that in figure 3 (a) will get their main SM contribution from top quark loops. However, this kind of process is not ideal for measuring Yukawa couplings, because there exist contributions from loops of every massive fermion, which does not exclude possible BSM contributions. Besides, even assuming SM-only loops, the total cross-section of the gluon-fusion process is sensitive to several orders of perturbation, resulting in an important systematic uncertainty.

An alternative process that allows direct measurement of this coupling is $t\bar{t}H$ production. In this process, there is always one vertex of interaction between the Higgs boson and a t or $\bar{t}$ in the leading production diagrams. Examples of such diagrams are shown in figure 8. After $t\bar{t}H$ production, the decay of the $t\bar{t}$ system can either be fully-hadronic, semi-leptonic or dileptonic if, respectively both, one or none of the top quarks decays hadronically.

If the SM is assumed to be a valid theory for energies up to the order of the Planck scale, the Higgs quartic self-coupling λ, as defined in section 1.2, must be corrected for loop contributions, which become important at higher energy scales [13]. In particular, the value of λ is very sensitive to the Higgs mass parameter m_H and to the top quark Yukawa coupling, since it is the only Yukawa coupling of order unity. For corrections near the Planck scale, a large top Yukawa coupling could drive λ to zero and even negative values. This change of sign would have dramatic consequences for the Higgs potential $V(\phi)$. Depending on λ, the SM vacuum could either be stable (global minimum), metastable (local minimum) or unstable (not a minimum at all). The requirement that the SM vacuum is a stable one results in a constraint between the Higgs boson and the top quark masses. The currently measured mass values fall into the metastable region, but are still compatible, within uncertainties, with the stable scenario [12].

As already mentioned, the observed asymmetry in the Universe between mat-
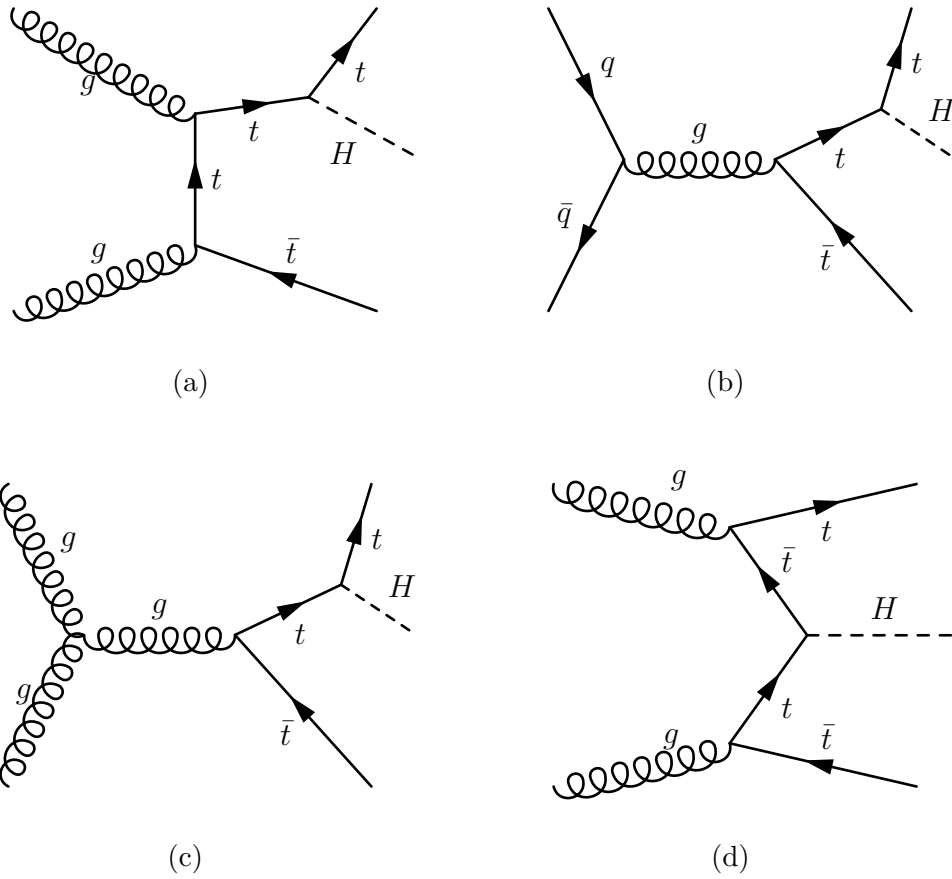
Figure 8: The four main processes for $t\bar{t}H$ production at the LHC.
(a): Gluon scattering through the exchange of a t-channel top quark.
(b): Quark fusion. **(c)**: Gluon fusion. **(d)**: Top quark fusion. An important remark is that **(a)**, **(b)** and **(c)** are simply processes of $t\bar{t}$ production followed by radiation of a Higgs boson from one of the top quarks.

ter and antimatter is in need of an explanation, usually believed to require new CP-violating interactions beyond the SM. One possibility would be to introduce CP violation in the Higgs sector, by allowing the Higgs boson to be in a mixed state of CP-even and CP-odd components. In popular BSM models such as supersymmetry and the two Higgs doublet model, the existence of CP-odd Higgs bosons is predicted.

The SM Higgs boson is a CP-even particle, which means the Higgs field is invariant under a combined charge and parity transformation. The field of a CP-odd Higgs boson would instead flip sign under such a transformation. Analyses have been conducted to exclude alternative CP scenarios for the Higgs boson, using the decays to ZZ and WW [54, 55]. However, a pseudoscalar boson is not allowed to couple at tree-level to gauge bosons [14], which means that decays of such a particle into WW and ZZ can only occur through loops, and are expected to be highly suppressed, when compared to the scalar scenario.

On the other hand, the tree-level coupling to fermions is allowed for a pseudoscalar Higgs boson, resulting in the effective interaction lagrangian

$$\mathcal{L}_{Hf\bar{f}} = -\frac{y_f}{\sqrt{2}}\bar{\psi}_f(a_f + ib_f\gamma_5)\psi_f h, \tag{19}$$

where a_f and b_f are respectively the scalar and pseudoscalar components of the Higgs boson, for the interaction with the fermion f. The SM is recovered by setting $a_f = 1$ and $b_f = 0$, a pure pseudoscalar has $a_f = 0$ and $b_f \neq 0$, whereas a mixed CP state has both $a_f \neq 0$ and $b_f \neq 0$.

In the most general case, it is possible that the values of b_f are different for different fermions, and the Higgs boson coupling to all fermions should be addressed. However, if these b_f are similar (as the a_f in the SM), the large top quark Yukawa coupling makes $t\bar{t}H$ production a privileged process for studying the Higgs boson CP nature. Measuring the total production cross-section is obviously not enough, as it will only be sensitive to $a_f^2 + b_f^2$, and not to the relative size of a_f and b_f. Sensitivity to the relative size of the components can be obtained by exploiting the kinematics of the $t\bar{t}H$ decays. Several discriminating observables have been proposed in [14, 63–65], and in the present work a set of new angular observables for this effect are also suggested.

2 Experimental setup

In this section, the current experimental conditions available for studying Higgs boson physics are discussed. An overview of the LHC operation is given, and the general-purpose detectors in which the Higgs boson was discovered are described.

The ATLAS detector is described in detail as an example of such a detector. It is one of the multi-purpose detectors at the LHC, along with CMS. Knowledge of the ATLAS experiment and its very complete set of subsystems is an excellent starting point to understand other general-purpose and even specialized particle detectors, through comparison of the technologies and construction strategies employed. It was chosen for a description with greater detail because the studies presented here were performed within a team mostly composed of members of the portuguese ATLAS group. Following the phenomenological work presented here, this team will engage in the Run 2 data analysis efforts in ATLAS. A summarized description of the CMS detector is also presented, highlighting the features that differentiate it from ATLAS.

The studies presented in this thesis required the simulation of a generic multi-purpose detector. In the end of this section, the package used for this fast detector simulation, DELPHES, is also described.

2.1 Large Hadron Collider

The Large Hadron Collider is the world's largest particle accelerator [7]. It is located at CERN's accelerator complex, in Geneva, crossing the French-Swiss border, and was first started on the 10th September 2008. The LHC consists of a 27 km ring of superconducting magnets buried at approximately 100 m underground, with a number of accelerating stages to boost the energy of the particles. Inside the accelerator, two high-energy beams travel in opposite directions, in separate beam pipes.

The beam pipes are kept at ultra-high vacuum and the beams are guided around the accelerator ring by a strong magnetic field maintained by the superconducting electromagnets. The magnetic dipoles operate at a temperature of 1.9 K. For this reason, much of the accelerator is connected to a distribution system of liquid

helium, which cools the magnets.

For each detector along the ring, there is a precisely defined interaction point. As the beams approach the detectors, they are radially confined ("squeezed") and crossed to induce particle collisions in these interaction points. Particles circulating in the LHC are distributed in bunches along the beam direction. In proton collisions during the LHC Run 2, the spacing between consecutive proton bunches is such that bunches cross every 25 ns at the interaction points.

The LHC was designed to collide proton beams with a centre-of-mass energy of up to 14 TeV and a luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, and also to collide heavy (Pb) ions with an energy of 2.8 TeV per nucleon and a peak luminosity of $10^{27}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The LHC integrated luminosities delivered to ATLAS and CMS, as of March 2016, are summarized in table 3 [66, 67].

Table 3: Integrated luminosities delivered by the LHC to the ATLAS and CMS detectors [66, 67]. Delivered luminosity is not totally recorded by the experiments, due in part to detectors being started only after the stable beams are circulating in the beam pipe. In collisions involving lead (Pb) ions, $\sqrt{s}$ is the nucleon-nucleon centre-of-mass energy.

	pp			pPb	PbPb
	$\sqrt{s} = 7\,\mathrm{TeV}$	$\sqrt{s} = 8\,\mathrm{TeV}$	$\sqrt{s} = 13\,\mathrm{TeV}$	$\sqrt{s} = 5\,\mathrm{TeV}$	$\sqrt{s} = 2.76\,\mathrm{TeV}$
ATLAS	$5.46\,\mathrm{fb}^{-1}$	$22.8\,\mathrm{fb}^{-1}$	$4.2\,\mathrm{fb}^{-1}$	$31.2\,\mathrm{nb}^{-1}$	$175.69\,\mu\mathrm{b}^{-1}$
CMS	$6.10\,\mathrm{fb}^{-1}$	$23.30\,\mathrm{fb}^{-1}$	$4.22\,\mathrm{fb}^{-1}$	$36.14\,\mathrm{nb}^{-1}$	$184.07\,\mu\mathrm{b}^{-1}$

The high luminosity delivered by the LHC also poses an experimental challenge. The design luminosity is such that, in every bunch crossing, the average number of inelastic pp interactions expected to occur is 24. This overlap of events is known as pile-up. In figure 9, distributions of the mean number of inelastic pp collisions per bunch crossing is shown, corresponding to the data collected by the ATLAS experiment during the LHC Run 1 [66].

Between February 2013 and March 2015, the LHC has undergone its first long shutdown period for maintenance and upgrade [68] [69]. A large engineering effort took place in order to strengthen the accelerator, with the main purpose being to consolidate the 10 170 high-current connections between superconducting magnets. This renovation work made the LHC safer from magnet quenching incidents for current operation, with proton beams circulating with the energy of 6.5 TeV
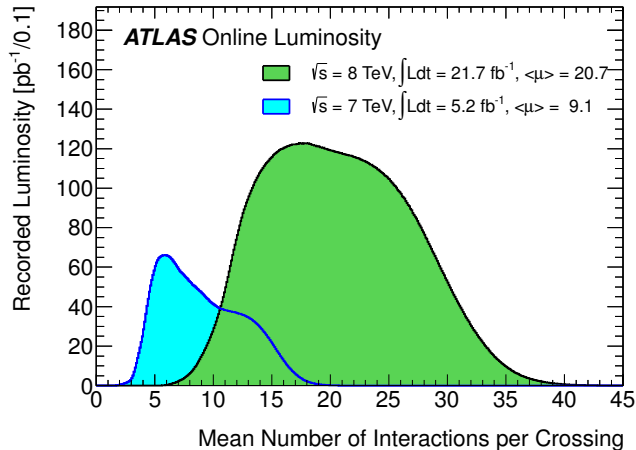
Figure 9: Distributions of the mean number of inelastic *pp* collisions per bunch crossing in data collected by ATLAS during the LHC Run 1 [66]. The average number of collisions in data collected at $\sqrt{s} = 8\,\text{TeV}$ was 20.7, close to the design goal average of 24.

per proton and colliding at a centre-of-mass energy of $13\,\text{TeV}$, unprecedented in accelerator high-energy physics.

2.2   ATLAS detector

ATLAS was designed and built for probing *pp* and PbPb collisions, being able to perform a wide range of precise SM measurements and having sensitivity for new physics processes [15, 70, 71]. A set of the new phenomena expected to occur at the TeV scale defined the requirements for the detector features.

In particular, the search for the SM Higgs boson established the minimum performance requirements for many sub-systems of ATLAS. The decay of the Higgs boson into a photon pair should be a relatively clean channel, requiring good electromagnetic calorimetry. On the other hand, the decay $H \to b\bar{b}$, dominant for Higgs masses below twice the Z boson mass, is expected to have a large QCD background and to require good *b*-tagging efficiency, only attained with fine vertex reconstruction. Associated production of the Higgs boson, as in $t\bar{t}H$, WH and ZH should be relatively clean thanks to leptons from $t\bar{t}$ or vector boson decays. These decays involve neutrinos in the final state, which adds the requirement for good missing transverse energy reconstruction.

Other important physics processes which motivated ATLAS features include

Table 4: General performance goals of the ATLAS detector [15]. The units of energy (E) and transverse momentum (p_T) are in GeV. The symbol $\oplus$ means a sum in quadrature.

Detector component	Required resolution	η coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\%\, p_T \oplus 1\%$	± 2.5	
EM calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	$\pm\,3.2$	$\pm\,2.5$
Hadronic calorimetry (jets)			
barrel and end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	$\pm\,3.2$	$\pm\,3.2$
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < |\eta| < 4.9$	$3.1 < |\eta| < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1\,\text{TeV}$	± 2.7	$\pm\,2.4$

heavier Higgs bosons ($m_H \gtrsim 600\,\text{GeV}$), new heavy gauge bosons W' and Z', supersymmetric particles and experimental signatures of the existence of extra dimensions.

The ATLAS detector must provide a mechanism to deal with pile-up, which again requires high precision tracking, for resolving the different interaction vertices in a single bunch crossing.

The performance goals imposed by the constraints mentioned before are summarized in table 4.

ATLAS uses a right-handed reference system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the beam pipe. The polar angle θ is the angle from the beam axis and the pseudorapidity η is defined as $\eta = -\ln[\tan(\theta/2)]$. Distances in (η, ϕ) space are given as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

ATLAS is a forward-backward symmetric, multi-layered detector, with 25 m in height and 44 m in length, weighing approximately 7 000 tonnes. Starting from the inner region, its subsystems are: the inner detector (ID), which is immersed in a 2 T uniform magnetic field and reconstructs the charged particle tracks; the electromagnetic (EM) calorimeter, a setup of lead and liquid argon (LAr) accordion-shaped layers where photons and electrons are contained and where EM-interacting particles leave energy deposits; the hadronic calorimeters, also
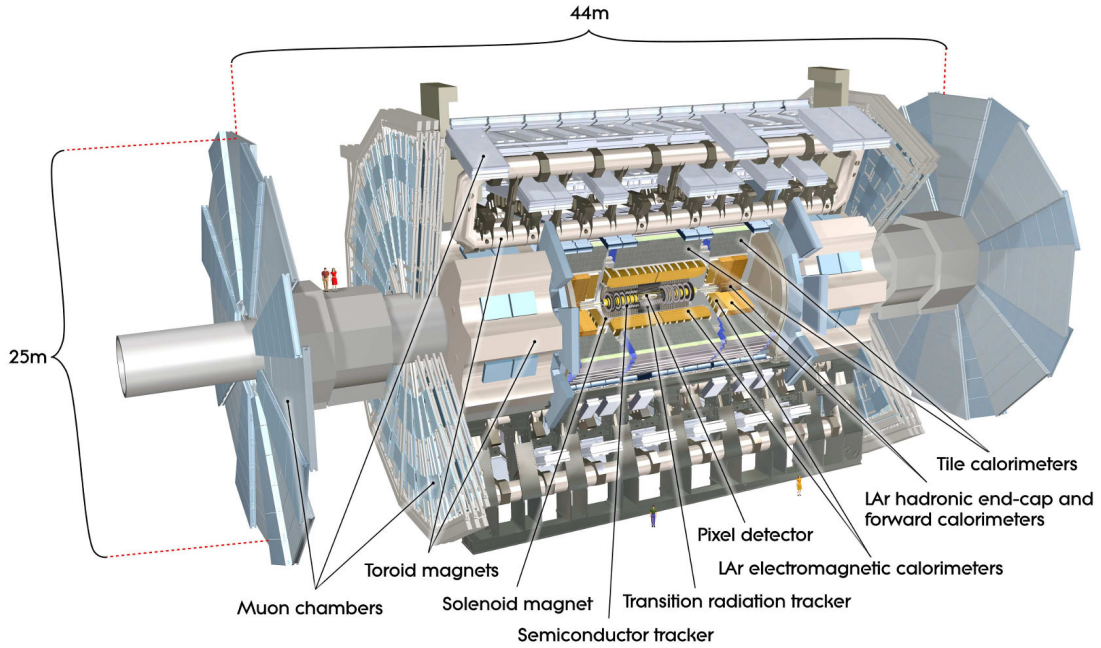
Figure 10: Cutaway view of the ATLAS detector [15]. The subdetectors and main systems are identified, and two persons are visible on the image for a notion of scale.

LAr-based in the forward ($|\eta| > 1.5$) region and composed of scintilating tiles in the central ($|\eta| < 1.7$) region, where hadrons are contained and also leave energy deposits; and the muon chambers, immersed in a magnetic field created by toroid magnets, which track the curved trajectories of muons. A cutaway view of the ATLAS detector with its subdetectors exposed is shown in figure 10. During the LHC long shutdown, improvements were made to the ATLAS experiment, notably with the insertion of an additional pixel layer in the inner detector, the inner B-layer (IBL) [72].

At the LHC, the nominal bunch crossing rate is $40\,\mathrm{MHz}$. With the design goal of 24 interactions per crossing, ATLAS gets an incoming interaction rate of $\simeq 1\,\mathrm{GHz}$. The ATLAS trigger system was designed to reduce the event rate from $40\,\mathrm{MHz}$ down to an average (peak) of $\simeq 200\,\mathrm{Hz}$ ($400\,\mathrm{Hz}$), such that the event data could be stored during the LHC Run 1. During the second run of the LHC, the final event recording rate increased to $1\,\mathrm{kHz}$. The trigger system must be very efficient and fast at eliminating the large background rate, while recording the rare events with potentially interesting physics.

2.2.1 Inner detector

The inner detector (ID) is the innermost subdetector of ATLAS. Being so close to the interaction point where the beams collide, it must deal with a very high density of particle tracks. The ID is a tracking system for charged particles, which allows momentum measurements and precise reconstruction of interaction vertices. The momentum measurement relies on bending the particle trajectories in a magnetic field peaking at $2\,\mathrm{T}$ and pointing along the z-axis, provided by a solenoid magnet placed outside the ID. The inner detector is composed by three subcomponents: the pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT), each with structures covering the barrel region and the end-cap regions.

The pixel detector along the barrel region is arranged in a set of concentric cylinders, starting as close as $31\,\mathrm{mm}$ from the beam axis (after the insertion of the IBL). On the end-cap region, it is disposed in circular disks, perpendicular to the beam axis. The high granularity of the pixel detector is crucial for the vertexing performance of ATLAS. The semiconductor tracker (SCT) is composed of silicon strips disposed in four concentric cylinders around the beam axis. The end-cap region is covered by nine disks on each side, with the strips aligned radially. The transition radiation tracker (TRT) is based on straw tube detectors with $4\,\mathrm{mm}$ in diameter with a gold-plated wire running inside. Electron identification capability is added by employing Xenon gas to detect transition radiation photons created in a radiator between the straws. The maximum straw length is $144\,\mathrm{cm}$ in the barrel, which contains about 50 000 straws. The end-caps contain 320 000 radial straws. Each channel provides a drift time measurement, giving a spatial resolution of $170\,\mu\mathrm{m}$ per straw. Each end-cap consists of 18 wheels.

The inner detector is connected to 80.3 million readout channels for the pixels, 6.3 millions for the SCT and 351 000 for the TRT. Figure 11 shows a cutaway view of the ID where all of its components are visible.

2.2.2 Electromagnetic calorimeter

The electromagnetic calorimeter is the detector layer in which electromagnetically interacting particles leave energy deposits, and where most photons and
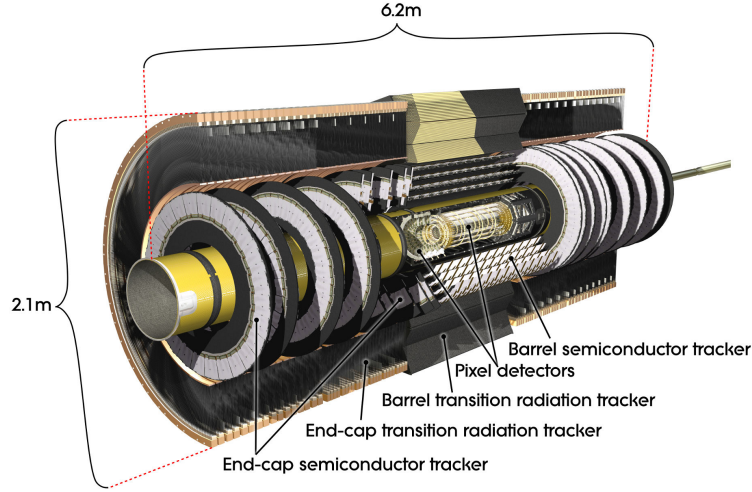
Figure 11: Cutaway view of the inner detector, with all of its components labeled. [71] The image portrays the ID before the insertion of the innermost pixel layer, the IBL.

electrons are contained. It is divided into a barrel part ($|\eta| < 1.475$) and two end-cap parts ($1.375 < |\eta| < 3.2$). In the barrel region, the EM calorimeter shares the vacuum vessel with the central solenoid. It consists of two half-barrels, with a $4\,\mathrm{mm}$ separation at $z = 0$.

Each of those half-barrels is composed by accordion-shaped layers of lead, liquid argon (LAr) and electrodes (copper deposited in kapton). Figure 12 shows a stack of accordion-shaped layers that belong to the barrel EM calorimeter. The accordion geometry provides complete and symmetrical coverage in ϕ, avoiding gaps. The high density lead plates act as the absorber elements of the calorimeter. They induce the production of particle showers and give a large effective depth to the calorimeter, assuring that particles such as electrons and photons do not punch through into the outer layers of the detector. As they cross the LAr layers (active element), the particles in the shower ionize the argon atoms, thus creating a current in the electrodes which is then detected and recorded. Only a fraction of the energy of the particles in the shower is deposited in this way. Using this fraction of energy as a sample and through calibration of the calorimeter, the measurement of the total energy deposited and its spatial distribution is obtained.

Each end-cap calorimeter is mechanically divided into two coaxial wheels: an outer wheel covering the region $1.375 < |\eta| < 2.5$, and an inner wheel covering the
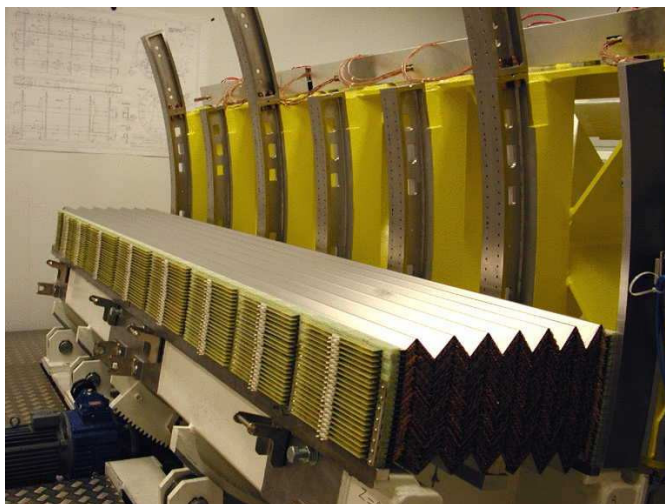
Figure 12: Photograph of a partly stacked barrel electromagnetic LAr module, where the accordion geometry is visible [15].

region $2.5 < |\eta| < 3.2$.

2.2.3 Hadronic calorimeter

The hadronic calorimeter is the subdetector where hadrons not contained by the EM calorimeter are ultimately contained and deposit their energy, which is then measured through sampling. This calorimeter is composed by the tile calorimeter (placed directly outside the EM calorimeter), the LAr hadronic end-cap calorimeter (consisting of two wheels per end-cap) and the LAr forward calorimeter, a high density cylinder which provides both EM and hadronic calorimetry in a region of larger $|\eta|$.

The tile calorimeter is, in its turn, composed of three barrels: the central barrel, which roughly matches the EM barrel in its z extension and covers the region $|\eta| < 1$; and two extended barrels, covering the range $0.8 < |\eta| < 1.7$. The central barrel, as well as the extended barrels, are divided azimuthally in 64 modules, and radially range from an inner radius of 2.28 m to an outer radius of 4.25 m. This calorimeter uses steel as the absorber medium and scintillating plastic tiles as the active material. The scintilating tiles emit light as they are crossed by ionising particles. The emmited light is collected by two optical wavelength-shifting fibres at the edges of each tile. The fibres are arranged in bunches and attached to
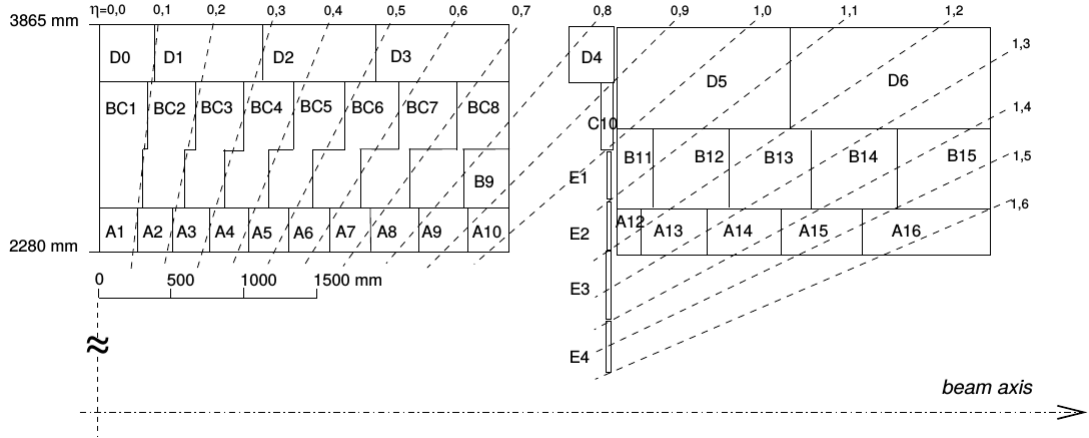
Figure 13: Segmentation in depth and η of the tile-calorimeter modules in the central (left) and extended (right) barrels.

photomultiplier tubes (PMT) which are placed at the outer edge of each module. The wavelength-shifting fibres correct the spectrum of the scintillator to match the sensitivity of the PMT. The readout cells corresponding to the bunches of fibres leading to the photomultipliers are represented in figure 13, where it is visible that they are approximately projective towards the interaction point.

The two wheels in each end-cap of the LAr hadronic calorimeter share the same cryostat as the EM end-cap calorimeter and are placed directly behind it. They cover the region with $1.5 < |\eta| < 3.2$, overlapping with the η coverage of both the tile calorimeter ($|\eta| < 1.7$) and the LAr forward calorimeter ($|\eta| > 3.1$). Each wheel is built from 32 azimuthal modules. The wheels are composed of parallel copper plates, 25 mm thick in the inner wheel and 50 mm thick in the outer wheel. In both cases, the plates are alternated with 8.5 mm LAr gaps, which serve as the active medium.

The LAr forward calorimeter (FCal) is composed of a cylindrical structure placed inside each of the end-cap wheels of the LAr hadronic calorimeter, sharing the same LAr cryostat. It has three modules in depth, with the innermost module made of copper, optimized for EM calorimetry and the two outermost modules made of tungsten and optimized for hadronic calorimetry. The electrode structure consists of concentric rods and tubes parallel to the beam axis. The sensitive medium is the LAr gap between each tube and the rod inside it.

37

2.2.4 Muon spectrometer

Muons are the only particles, besides the undetectable neutrinos, that are not contained by the calorimeters. The ATLAS muon system serves the purpose of tracking and measuring the momentum of these particles. It is based on a complex arrangement of toroid magnets which deflect the muon trajectories, a dedicated muon trigger system and high-precision tracking chambers.

There are three large air-core toroids in the ATLAS magnet system: the central barrel toroid, which provides magnetic bending over the range $|\eta| < 1.4$ and two end-cap toroids, performing the same function in the region with $1.6 < |\eta| < 2.7$. Each of the toroids is composed of eight superconducting coils of an Al-stabilised Nb/Ti/Cu alloy, which operate at a temperature of $4.6\,\mathrm{K}$. The coils are assembled symmetrically in planes of constant ϕ. The end-cap toroids are rotated in ϕ by an angle of $\pi/8$ with respect to the barrel toroid to enable radial overlap of the toroids and to optimize the bending power in the transition range of $1.4 < |\eta| < 1.6$. Each coil of the barrel toroid is housed in its own cryostat, while the end-cap toroid coils are assembled inside one large cryostat for each end-cap. The field produced in the central region is approximately $0.5\,\mathrm{T}$ on average and $3.9\,\mathrm{T}$ at its peak. In the end-caps, the field is approximately $1\,\mathrm{T}$ on average and $4.1\,\mathrm{T}$ at its peak. Figure 14 shows the 3D configuration of all the magnet coils in ATLAS.

The precision tracking chambers in the barrel region are placed in three concentric cylindrical layers around the beam axis at radii of approximately $5\,\mathrm{m}$, $7.5\,\mathrm{m}$, and $10\,\mathrm{m}$, with the innermost and outermost ones roughly on the limits of the barrel toroid and the central one inside it. These layers exhibit the same eight-fold symmetry as the toroid magnets and have a small component and a large one at each octant. Projections of the ATLAS muon system on the transverse plane and on an axial plane are represented in figure 15. In the two end-cap regions, muon chambers form large wheels, perpendicular to the z-axis and located at distances of $|z| \simeq 7.4\,\mathrm{m}$, $10.8\,\mathrm{m}$, $14\,\mathrm{m}$, and $21.5\,\mathrm{m}$ from the interaction point. In each end-cap, the wheel closer to the interaction point is placed in front of the end-cap toroid, another wheel is placed around the toroid, and the other two are placed behind it, as can be seen in the bottom image of figure 15.

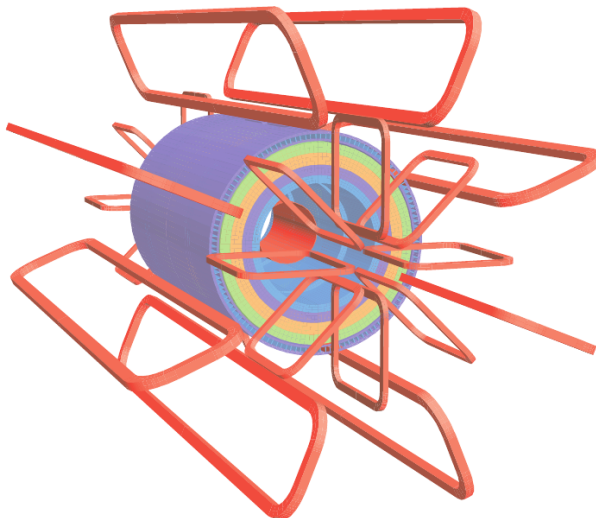The precision momentum measurement is performed by the monitored drift

Figure 14: Geometry of magnet windings and tile calorimeter steel. The eight barrel toroid coils, with the end-cap coils tilted relative to the barrel ones are visible. The solenoid winding lies inside the calorimeter volume. [15]

tube (MDT) chambers. They cover the pseudorapidity range $|\eta| < 2.7$ (coverage is limited to $|\eta| < 2.0$ in the innermost end-cap layer). These chambers are composed of three to eight layers of drift tubes with an operation similar to the TRT in the inner detector. The working absolute pressure inside the tubes is $3\,\mathrm{bar}$, and these chambers achieve an average resolution of $80\,\mu\mathrm{m}$ per tube, or about $35\,\mu\mathrm{m}$ per chamber. In the forward region ($2.0 < |\eta| < 2.7$), cathode-strip chambers (CSC) are used in the innermost tracking layer, since they can handle higher hit rates and have finer time resolution. The CSC are multiwire proportional chambers in which cathode planes are segmented into strips running in perpendicular directions. This allows measurement of two coordinates from the distribution of the induced charge. The resolution of a chamber is $40\,\mu\mathrm{m}$ in the plane containing the beam axis and the track (bending plane) and about $5\,\mathrm{mm}$ in the transverse (non-bending) plane. A high-precision optical alignment system monitors the positions and internal deformations of the MDT chambers; it is complemented by track-based alignment algorithms.

The precision tracking chambers are complemented by a system of fast trigger chambers. In the barrel region ($|\eta| < 1.05$) this is accomplished with Resistive Plate Chambers (RPC), while in the end-cap ($1.05 < |\eta| < 2.4$) thin gap chambers
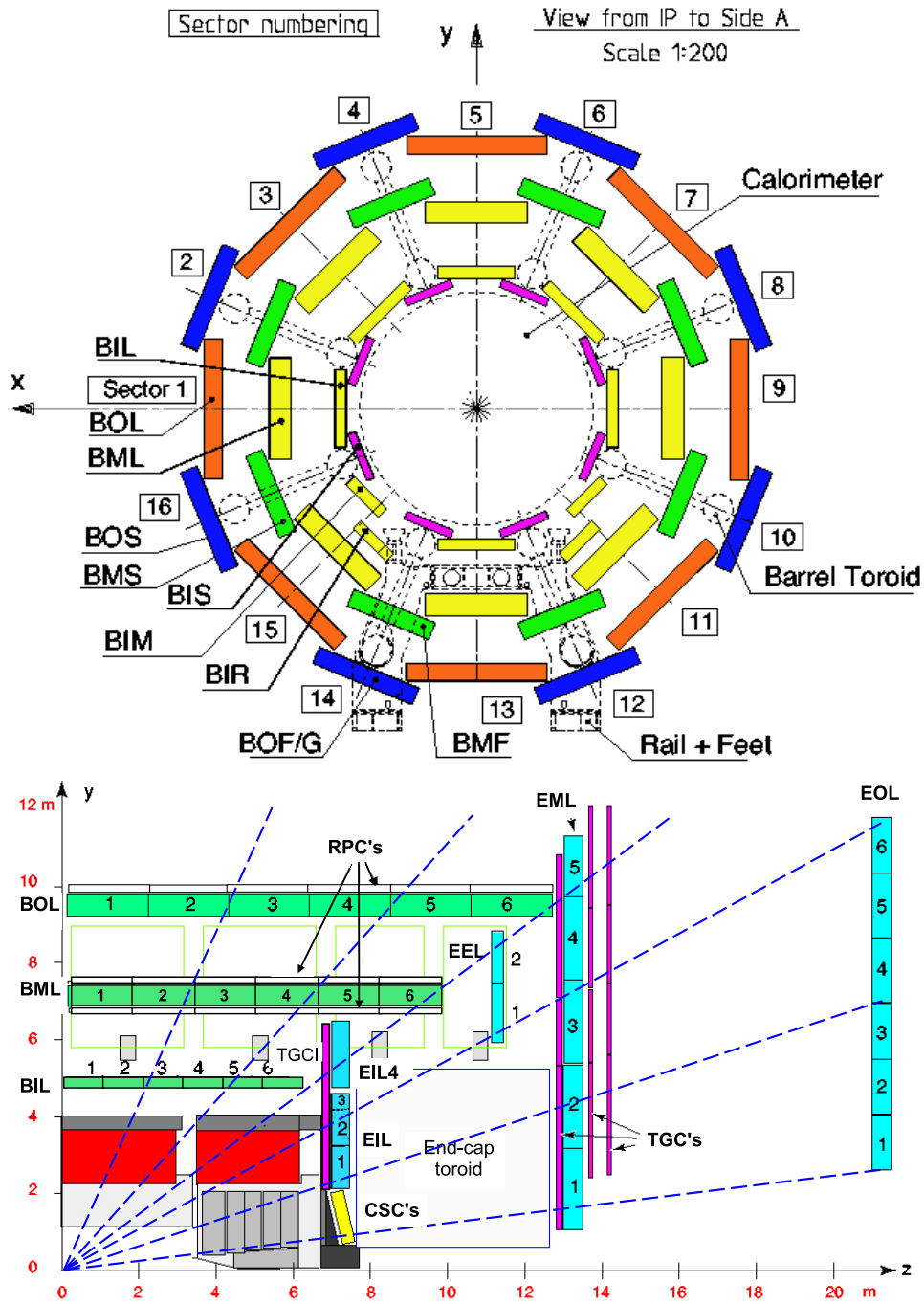
39

Figure 15: Labels: B/E - barrel/end-cap, I/M/O - inner/medium/outer, S/L/F - small/large/feet. **Top:** Cross-section of the barrel muon system perpendicular to the beam axis (non-bending plane), showing three concentric cylindrical layers, each comprising eight large and eight small chambers. The outer diameter is about 20 m [15]. **Bottom:** Cross-section of the muon system in a plane containing the beam axis (bending plane). Infinite-momentum muons would propagate along straight trajectories which are illustrated by the dashed lines and typically traverse three muon stations [15].

40

(TGC) are used. The design goal was to minimize time contributions from signal propagation and electronics to allow efficient identification of the beam crossing. Both chamber types deliver signals with a spread of 15–25 ns, allowing each individual chamber to tag the beam-crossing with efficiency $\geq 99\%$. Besides triggering, TGC provide a measurement of both coordinates of the track, one in the bending plane and one in the non-bending plane. In particular, the measurement in the non-bending plane is used as the second coordinate of the track, instead of the value from the MDT, which has a larger uncertainty.

2.2.5 Trigger system

The ATLAS trigger system is structured in two trigger levels: the Level-1 (L1) trigger and the higher-level trigger (HLT).

The L1 trigger is based in custom-made electronics which process information from the trigger chambers in the muon spectrometer and from all the calorimeters, with reduced-granularity information from the latter. It searches for high transverse momentum (p_T) muons, electrons, photons, jets and τ-leptons decaying hadronically. It also looks for large missing transverse energy and large total energy in the event. L1 takes up to $2.5\,\mu s$ to select the events and define some regions of interest (RoI) in the detector, while the full data is stored in the pipeline memory. The RoI information includes the η and ϕ coordinates of the interesting feature, the type of feature (electromagnetic, hadronic, muon...) and the reached threshold. This trigger level delivers an output event rate of up to $100\,\text{kHz}$.

The HLT takes the previously defined RoI information as input, and has access to full granularity and precision information from all the detectors withtin those regions (approximately 2% of the total event data). The HLT selection is mostly implemented using offline analysis procedures, which reduce the event rate, allowing permanent storage of the data. The average output rate of the HLT is $1\,\text{kHz}$. The event processing average time at this level is of the order of $4\,\text{s}$, and the final event size is approximately $1.3\,\text{Mbyte}$.

For the second run of the LHC, the ATLAS trigger system experienced important upgrades [73]. The Level-1 trigger output rate of $100\,\text{kHz}$ resulted from an upgrade, in which it was increased from the Run 1 rate of $70\,\text{kHz}$. The previously

existing Level-2 trigger and Event Filter were effectively merged into the HLT, to avoid the overhead of transmitting the same event information at least twice, to different devices. The final event recording rate was also increased from an average of 400 Hz during Run 1 to the already mentioned rate of 1 kHz.

2.2.6 Object reconstruction

Electrons and photons

The reconstruction of electrons and photons starts from an energy deposit (cluster) in the electromagnetic calorimeter. All the reconstructed tracks in the inner detector are searched, looking for one that loosely matches that cluster. Also in the inner detector, a search for a matching photon conversion to e^-e^+ takes place. Based on the result of these searches, the cluster is identified either as an electron candidate or a photon candidate: electrons are required to have an associated track, but no associated conversion, and photons are defined as not having a matched track, or as being matched to a conversion. Electron identification results from the combination of reconstruction properties. These include the ratio of energy (measured by the calorimeter) to momentum (measured at the ID), the difference between the (η,ϕ) coordinates reconstructed from the cluster and the ones obtained from the track extrapolation into the calorimeter, and transition radiation hits on the track.

The energy of high-p_T electrons is obtained from the calorimeter measurement. However, the direction (i.e. η and ϕ) is more precisely determined from the matched track. The identification for isolated high-p_T electrons is based on cuts on the shapes of particle showers and on information from the associated track. Depending on the requirements of each particular physics analysis, different sets of cuts can be applied, resulting in different electron identification efficiencies and jet rejection rates: "loose cuts" consist of simple shower-shape cuts and very loose matching between reconstructed track and calorimeter cluster; "tight cuts", on the other hand, tighten the track-matching criteria and the cut on the energy-to-momentum ratio. Additionally, calorimeter energy isolation beyond the cluster itself can be required, ensuring further isolation of the electron.

For photons, both the energy and the (η,ϕ) direction are derived from the

calorimeter. Unlike isolated energetic electrons, such as those expected from the decay of a vector boson, photons are much harder to differentiate from the jet background. A single set of cuts for photon identification, analogous to the electron "tight cuts", has been optimized. It relies on calorimeter shower shapes and focuses on separating single π^0 mesons from photons, using the very fine granularity in η of the silicon strip layer. A track isolation requirement has also been added to further improve the jet rejection, while preserving most of the converted photons.

Muons

ATLAS identifies and measures muons in the muon spectrometer, and combines information from the calorimeters and the inner detector to improve identification efficiency and resolution of the momentum measurement. The muon spectrometer also has the capability of efficiently triggering on muons, for pseudorapidities $|\eta| <$ 2.4 and over a wide range of energies.

Momentum measurements are performed using combined tracks from the inner detector and the muon spectrometer. The best resolution of the muon spectrometer measurement is obtained for $p_T \simeq 100\,\text{GeV}$. For lower p_T muons, the energy losses in the calorimeters determine the resolution, and for higher p_T, it is the spatial resolution of drift chambers. Momentum resolution of muons with p_T below $100\,\text{GeV}$ benefit significantly from combining measurements from both subsystems. Below $30\,\text{GeV}$, the resolution is dominated by the inner detector resolution, which is much better than the spectrometer in this p_T region. Even at the largest values of $|\eta|$, excellent momentum resolution is assured by the toroidal field.

When performing the momentum measurement, muon tracks in the spectrometer are propagated back to the interaction point. This requires a momentum correction due to the energy lost by the muon in the calorimeters. An algorithm is used to perform this correction, by means of a parameterised expected energy loss, obtained from simulation. A measured energy loss may be used instead, if the muon track is isolated and if the measured loss is much larger than expected from the parameterisation.

Jets

Jet reconstruction from calorimeter signals relies on the use of a clustering

algorithm. The choice of algorithm depends on the particular physics analysis. Currently, almost every ATLAS analysis uses the anti-k_t algorithm [74]. Some algorithms, including anti-k_t, must be given a ΔR parameter. Typical values are $\Delta R = 0.4$ for narrow jets and $\Delta R = 0.6$ for wide jets. The application of the clustering algorithm is followed by a calibration step. Jet-clustering algorithms can take different object collections as input for finding jets, as long as those objects have defined directions and energies. Most ATLAS analyses use topological clusters as the objects on which the algorithm is performed.

Topological cell clusters result from attempting the reconstruction of three-dimensional calorimeter energy depositions. For every cell, σ_{cell} is defined as the total noise expected in that cell, both from electronics and pile-up events. First, a seed cell is identified if a significant absolute signal above the major seed threshold $4\sigma_{cell}$ is present. Then, the nearest neighbours of each seed cell are collected, independently of the amplitude of their signal. They can be considered secondary seeds, if the absolute value of their signal magnitude is above the secondary seed threshold $2\sigma_{cell}$. If this is the case, their nearest neighbours are also added to the cluster. Finally, if there are not additional secondary seeds among these nearest neighbours, all the cells surrounding the ones collected so far, with signal above a very low threshold are added to the cluster. The resulting cluster is searched for local signal maxima. If more than one maximum is found, the cluster is split into smaller ones, along the boundaries determined by signal valleys between the maxima.

Missing E_T

In hadron colliders, the inelastic physical processes studied occur between constituents of the colliding protons, called partons. These partons do not necessarily carry the same fraction of momentum from the parent protons. An important consequence of this is that, in events resulting from hadron collisions, the sum of the momenta of resulting particles does not add up to zero. However, that is the case for the transverse component alone. For this reason, it is usual to use transverse quantities, such as p_T, in hadron collider experiments. Missing transverse energy (missing E_T, or $\not{E}_T$) is defined as the negative of the vectorial sum of the p_T of all detected particles. If there are energetic undetected particles in the event, such as

44

neutrinos, this quantity should roughly correspond to the transverse component of the sum of their momenta.

The reconstruction of $\not{E}_T$ in ATLAS relies primarily on the calorimeter cell energies, after calibration, and on the reconstructed muons. The muon term in $\not{E}_T$ is calculated from the muon momentum measurement from the muon spectrometer alone. This means that energy losses of muons in the calorimeters are not counted twice, since they only contribute to $\not{E}_T$ through the calorimeter term. Sometimes, in events with very energetic jets, many jet constituents reach the muon spectrometer and a 'fake' muon is identified. In order to reduce possible contributions of 'fake' muons to the $\not{E}_T$ muon term, only muons matched to an inner detector track are considered.

After this primary calculation, the $\not{E}_T$ reconstruction introduces the cryostat term, which accounts for the energy lost by particles in the cryostat between the barrel electromagnetic calorimeter and the tile hadronic calorimeter.

A final calibration of $\not{E}_T$ is performed through the following matching procedure. All reconstructed high-p_T objects in the event are carefully ordered and, for each one of them, the calibrated calorimeter cells close to that object are associated to it, unless they were already associated to a previous object. The calibration of $\not{E}_T$ consists in replacing the initial term from globally calibrated cells by terms from the high-p_T objects corresponding to those cells. The cells which are not rejected by a noise cut and are not matched to any reconstructed object are also accounted for in the $\not{E}_T$ calculation, keeping the previous global calibration.

b-tagging

The ability to tag hadronic jets arising from heavy flavours is an important asset for many physics analyses, such as precision measurements in the top quark sector and searches for Higgs bosons or other new physics signatures. The basic b-tagging algorithms use charged particle tracks to produce a set of variables which discriminate between different jet flavours [75]. ATLAS uses three distinct basic b-tagging algorithms, which provide complementary information. First, tracks are associated to a calorimeter jet, based on their angular separation ΔR (track, jet). The ΔR association requirement varies as a function of the jet p_T, resulting in a narrower cone for jets at high p_T, which are more collimated. Then, tracks are

required to pass a quality selection, which is different for each of the b-tagging algorithms.

The impact parameter-based algorithm makes use of the impact parameter of the tracks matched to the jet. The impact parameter is the distance of the track to the primary vertex at the point of closest approach. The sign is defined positive (negative) if this point is in front of (behind) the primary vertex with respect to the jet direction. The inclusive secondary vertex reconstruction algorithm aims to explicitly reconstruct an inclusive displaced secondary vertex within the jet. The decay chain multi-vertex reconstruction algorithm exploits the topological structure of weak b and c hadron decays inside the jet and tries to reconstruct the hadronic decay chain: primary vertex $\rightarrow b \rightarrow c$.

The output of these b-tagging algorithms are later combined in a multivariate discriminant, which provides the best separation between the different jet flavours. Depending on the physics analysis, a cut on this discriminant variable defines the b-tagging working point. For example, a typical b-tag working point for LHC Run 2 requires a 70% b-tagging efficiency, resulting in rejection factors of 8.1 for c-jets, 26 for τ-jets and 440 for light-flavoured jets [75].

Pile-up removal

With an average number of inelastic pp collisions per bunch crossing close to 20, every hard-scattering process that triggers an event in ATLAS is expected to be detected simultaneously with many pile-up interactions. These pile-up interactions are uncorrelated with the hard-scattering process, and are mostly composed of soft hadronic activity, resulting in small energy depositions [76]. Thus, the most severe effects of pile-up are on the reconstruction of jets: it can affect the reconstructed jet energies and structure, and it can even lead to the reconstruction of additional jets.

The time space between bunch crossings at the LHC was 50 ns for most of the Run 1, and it is 25 ns during Run 2. The response time of many subdetectors is much larger than this time interval, which can result in the association of a detection to the wrong bunch crossing. This effect, called out-of-time pile-up (as opposed to the in-time pile-up explained above), is especially important in the electromagnetic LAr calorimeter.

Many strategies have been developed to remove the effects of pile-up in AT-LAS. The most imediate strategy consists in adjusting the σ_{cell} parameter used in the construction of topological clusters. This parameter defines a threshold, and if it is kept above the typical pile-up noise in the corresponding calorimeter cell, the number of topological clusters formed solely of pile-up depositions is significantly reduced. The limitation for increasing this parameter comes from the loss of efficiency in reconstructing desirable jets. The expected pile-up noise can be obtained from simulation, given the mean number of interactions per bunch crossing $\langle \mu \rangle$. For data collected at $\sqrt{s} = 8\,\mathrm{TeV}$, for example, the parameter σ_{cell} used corresponded to $\langle \mu \rangle = 30$.

A correction is performed on the transverse momentum of reconstructed jets, according to

$$p_T^{jet} \to p_T^{jet} - \rho \times A^{jet},$$

where ρ is the pile-up p_T density and A^{jet} is the area of the jet, in (η, ϕ) space. The pile-up p_T density ρ is calculated separately for each event. The k_t algorithm is used to find jets in the event, without a minimum p_T threshold, and then ρ is defined as the median p_T density of all those jets. The k_t algorithm is particularly sensitive to low energy radiation, and thus adequate to describe pile-up behaviour.

The subtraction procedure described above removes a significant fraction of in-time pile-up. However, it is not efficient at removing out-of-time pile-up, especially in the forward regions of the detector. After the subtraction of $\rho \times A^{jet}$, a residual subtraction is performed, with one term proportional to the number of reconstructed pile-up vertices (to address the remaining in-time pile-up), and another term proportional to $\langle \mu \rangle$. Because $\langle \mu \rangle$ is a function of the instantaneous luminosity, it is a good estimator of out-of-time pile-up activity, and actually the subtraction of the term proportional to $\langle \mu \rangle$ is effective in mitigating out-of-time pile-up. Both these terms are derived from parton-level jets in simulated dijet events.

After the p_T corrections are applied, energy contributions of pile-up to jets from the hard scattering are greatly minimized. Besides, the majority of reconstructed pile-up jets are removed because, after the $\rho \times A^{jet}$ subtraction, they fall below the p_T threshold of $20\,\mathrm{GeV}$. Still, it is possible that some pile-up jets remain in
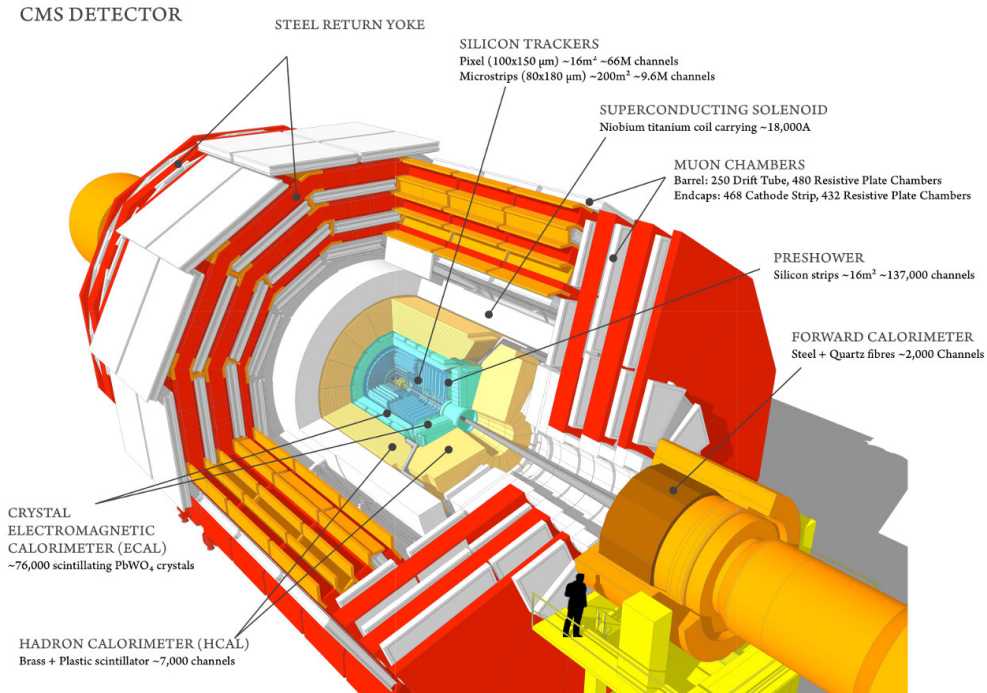
Figure 16: Cutaway view of the CMS detector, with its subdetectors exposed [77]. The calorimeters fit inside the solenoid, unlike in ATLAS.

the event. They can be either genuine QCD jets from a pile-up vertex, or jets reconstructed from a local fluctuation in the energy depositions coming from pile-up. Removal of such jets relies on reconstructed tracks and vertices in the inner detector. The tracks from each vertex are propagated into the calorimeter cells, and discriminant quantities are used to classify each jet either as coming from the hard-scattering vertex or from a pile-up vertex.

2.3 CMS detector

The design of CMS (Compact Muon Solenoid) [16] was driven by the same physics goals and experimental challenges as the design of ATLAS. The main distinguishing features of CMS are a high-field solenoid and a homogeneous electromagnetic calorimeter based on scintillating crystals. CMS is 21.6 m long, has a 14.6 m diameter, and a total weight of 12 500 tonnes. A cutaway view of CMS is shown in figure 16.

48

2.3.1 Superconducting magnet

The central element of CMS is a superconducting solenoid, 12.5 m long and with 6 m of inner diameter, large enough to accomodate for the inner tracker and the calorimeters inside it. It produces a 4 T uniform field pointing along the z direction (the CMS coordinate system is defined in the same way as the ATLAS system), and provides a bending power of 12 Tm. At full current, the total stored energy is of 2.6 GJ. The coil consists of a distinctive 4-layer winding, made from a stabilised reinforced NbTi conductor, and has a cold mass of 220 tonnes.

On the outside of the solenoid, the magnetic flux is returned by a 10 000 tonne iron yoke, comprising 5 wheels and 2 end-caps, each end-cap composed of three disks. Part of the CMS muon system is placed within the volume of this return yoke.

2.3.2 Inner tracking system

The CMS tracking system is designed to provide precise and efficient measurement of the trajectories of charged particles, and to enable reconstruction of secondary vertices. It spans a cylindric volume with 5.8 m of length and a 2.5 m diameter. It consists of three layers of silicon pixels close to the interaction point, with radii from 4.4 cm to 10.2 cm, and 10 layers of silicon strips, extending out to 1.1 m in radius. Additionally, each side of the end-cap region is covered by 2 disks in the pixel detector and 12 disks in the silicon strip tracker. These extend the acceptance of the tracking system to $|\eta| < 2.5$. The CMS tracker is the largest silicon tracker ever built, with about $200 \, \text{m}^2$ of active silicon area.

2.3.3 Calorimeters

The electromagnetic calorimeter (ECAL) in CMS is a hermetic and homogeneous detector, based on scintillating crystals of lead tungstate (PbWO_4). It is composed of 61 200 crystals in the central barrel region, and 7324 crystals in each of the end-caps. Avalanche photodiodes are used to detect the scintillation light in the barrel region, and vacuum phototriodes are used in the end-cap regions. In front of the end-cap ECAL, a preshower detector system is installed to provide π^0

rejection. The ECAL covers the pseudorapidity range $|\eta| < 3.0$, and its thickness is larger than 25 radiation lengths.

The hadronic calorimeter (HCAL) is a sampling scintillator calorimeter, using brass as the absorber medium. Scintillation light travels through wavelength-shifting fibres before being detected by hybrid photodiodes. The HCAL barrel extends from an inner radius of 1.77 m to an outer radius of 2.95 m, covering the pseudorapidity range $|\eta| < 3.0$. Its thickness varies from 7 to 11 interaction lengths, depending on η. The outer radius of the HCAL barrel is limited by the magnet coil. In order to increase the amount of material crossed by hadronic showers in the central region and thus increase the HCAL stopping power, a "tail catcher" calorimeter is placed outside the magnet coil. Including this "tail catcher", the total thickness of HCAL varies from 10 to 15 interaction lengths.

In the forward region ($|\eta| > 3.0$), a calorimeter using quartz fibres as the active medium, and steel as an absorber, placed at 11.2 m away from the interaction point, provides coverage up to $|\eta| = 5.2$. This calorimeter uses photomultipliers to detect the Cherenkov radiation produced by the forward particles travelling inside the quartz fibres.

Regions of even higher pseudorapidity are covered by dedicated detectors: the CASTOR and ZDC calorimeters, and the TOTEM tracking detectors. These are especially designed for studying heavy ion collisions, as well as diffractive pp collisions.

2.3.4 Muon system

The middle name of CMS is "Muon", hinting at the central importance that the muon system had for its design. The muon detection system of CMS serves three purposes: muon identification, muon momentum measurement, and muon triggering. It comprises 4 muon stations, structurally integrated with the iron yoke. The latter also serves as a hadron absorber, resulting in cleaner muon detection. The muon stations are shaped as concentric cylinders in the barrel region and parallel disks in the end-caps. Each muon station consists of several layers of aluminium drift tubes (DT) in the barrel region and cathode strip chambers (CSC) in the end-cap region, complemented by resistive plate chambers (RPC). The muon

system has a total area of detection planes of about $25\,000\,\mathrm{m}^2$.

The barrel region DT chambers cover the pseudorapidity range $|\eta| < 1.2$. The first (innermost) muon station fits inside the iron yoke, the second and third stations are among the return yoke plates, and the fourth station is outside the yoke and limits the CMS detector volume. Each station is segmented in 12 sectors along ϕ and in 5 wheels along z, a segmentation imposed by the structure of the iron yoke. In order to preserve efficiency, the segmentation in ϕ is such that the unavoidable gaps in each station do not overlap with the gaps in other stations.

In the first 3 stations, each one of the $12 \times 5 = 60$ segments contains one DT chamber. Each DT chamber contains 12 layers of drift tubes: 4 inner layers, with tubes parallel to the beam pipe, which give a ϕ measurement; 4 layers in the middle with tubes perpendicular to the beam pipe, which give a z measurement; and 4 outer layers, again parallel to the beam pipe, to provide the best possible angular resolution. A depiction of such a chamber is shown in figure 17. In the fourth station, some segments contain 2 chambers, resulting in a total of 70 chambers for this station, and 250 chambers for the whole muon system. The fourth station chambers do not contain the z measurement layers of tubes.

Muon detection in the end-caps is provided by CSCs, which are radiation resistent and have fast response times. These CSCs are placed among the flux return plates and they identify muons in the range $0.9 < |\eta| < 2.4$. The cathode strips in each chamber run radially outward and provide a precise measurement of ϕ. The anode wires are approximately perpendicular to the strips, and are used to obtain measurements of η and beam-crossing times. Each CSC is composed of 6 layers, which enables good rejection of non-muon background and matching of tracks to other stations and to the CMS inner tracking system.

The muon system efficiency for muon identification is above 95% for almost the entire range $|\eta| < 2.4$. The muon transverse momentum is measured independently by the inner tracker and by the muon system. Combining these measurements, the obtained resolution is of about 3% for a muon with $p_T = 100\,\mathrm{GeV}$ and about 10% for a muon with $p_T = 1\,\mathrm{TeV}$, assuming muons with high pseudorapidity ($1.2 < |\eta| < 2.4$), for which the resolution is worse.

The trigger function of the muon system is provided by resistive plate chambers (RPC). They produce a fast response with good time resolution, but coarser
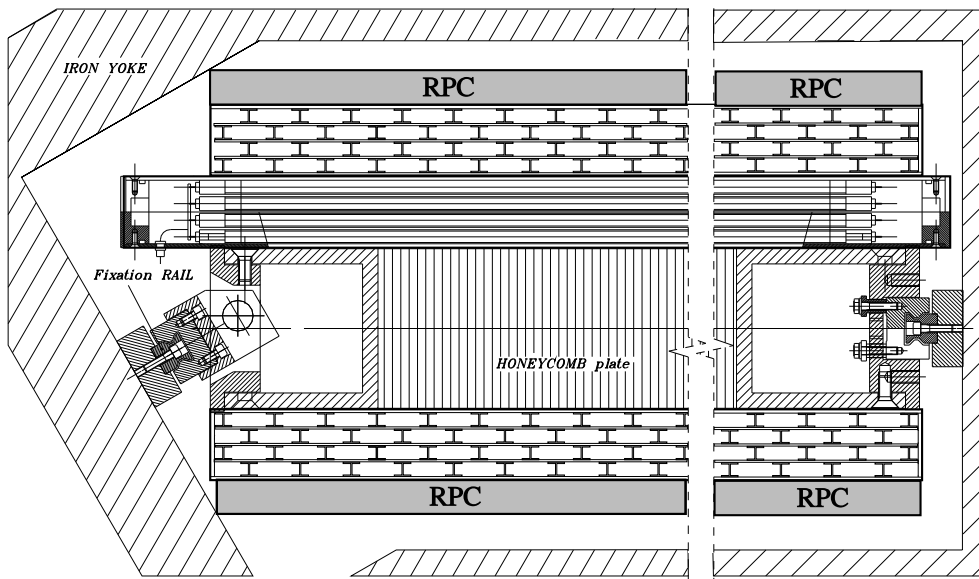
Figure 17: Transverse cut of a CMS barrel muon chamber [16]. The chamber is fixed to the plates of the iron yoke. In this figure, two sets of 4 ϕ-measuring layers are visible, as well as the 4 z-measuring layers between them. The honeycomb plate has a structural purpose. The presence of two RPC layers indicates that this chamber belongs to one of the first 2 muon stations.

position resolution than the DTs or CSCs. They also help to resolve ambiguities in attempting to make tracks from multiple hits in a chamber. A total of 6 layers of RPCs were placed in the barrel muon system: 2 in each of the first 2 stations and 1 in each of the last two stations. In the end-cap muon system, 1 layer was introduced in each of the first 3 stations. The redundancies in the placement of these trigger chambers improve the beam-crossing time identification, as well as the background rejection rate.

2.4   Delphes framework

When physics analyses or detector studies require a high level of accuracy, it is usual for collaborations to rely on tools that fully simulate a detector response. Such tools must reproduce the particle interactions with the matter content of the detector, the electronic response of each detector component, and the complex algorithms used in the real experiments to obtain the observables from the electronic output of the detector. Such a complete framework requires computational resources not usually available to individual researchers or small collaborations.

In phenomenological studies such as the ones presented here, a full detector simulation is not required. It is then customary to use fast detector simulators, which parameterise the response of the detector, while keeping a reasonable degree of realism. The DELPHES [17] package, used for the work presented here, is one example of such a tool. DELPHES simulates the response of a multi-purpose detector, composed of an inner tracker immersed in an uniform magnetic field, electromagnetic and hadronic calorimeters, and a muon detection system. All are organized concentrically with cylindrical symmetry around the beam axis. The detector active volume, calorimeter segmentation and strength of the uniform magnetic field, which is directed along the beam axis, are parameters which can be set by the user.

2.4.1 Particle propagation

The first phase of detector simulation simply propagates the particles resulting from the physics process (generation and hadronization) through the inner tracker. Neutral particles travel along straight lines and charged particles describe helicoidal curves, according to their momentum and to the magnetic field strength. Particles are propagated until they reach a calorimeter cell. Particles originating outside the tracker volume are not considered in this phase.

Tracking is assumed to have perfect angular resolution, which is not far from real conditions. The user can define the resolutions of energy and p_T as functions of particle type, p_T and η. Also the probability for charged particles to be reconstructed as tracks is a user-defined parameter.

2.4.2 Calorimeters

The calorimetry in DELPHES is performed by two calorimeters: the electromagnetic calorimeter ECAL, and the hadronic calorimeter HCAL. These calorimeters are segmented in a rectangular grid in (η, ϕ). The size of the rectangles can be defined by the user and can depend on η. Segmentation in ϕ is uniform, and the granularity is the same for ECAL and HCAL.

Each particle has parameters f_{ECAL} and f_{HCAL}, representing the fraction of its total energy which is deposited in the ECAL and HCAL, respectively. By default,

electrons and photons have $f_{ECAL} = 1$ and hadrons have $f_{HCAL} = 1$, except for kaons and Λ particles for which $f_{ECAL} = 0.3$ and $f_{HCAL} = 0.7$. Neutrinos and muons do not deposit any energy in the calorimeters. Assuming that hadrons do not deposit any energy in the electromagnetic calorimeter and that muons do not deposit any energy at all are rough approximations. However, these are default parameters which can be changed to more adequate values.

Since ECAL and HCAL are equally segmented, a straight line coming from the interaction point crosses one ECAL cell and one HCAL cell covering precisely the same region in (η, ϕ). These pairs of cells are called calorimeter towers, and are used in the object reconstruction, together with tracks.

The detected energy in each tower is given by a sum over all particles travelling through that tower. Each particle contributes with an ECAL term and an HCAL term. These are equal to the energy deposited in the corresponding calorimeter, after the application of a smearing. For these smearings, the energy resolution used is a function of the particle total energy and η, which is different for the ECAL and for the HCAL terms.

2.4.3 Charged leptons and photons

Muons

For muons, the user can define a global probability of reconstruction and the momentum resolution, which is a function of the muon p_T and η. The momentum measurement is then obtained from a gaussian smearing of the original muon 4-momentum, according to the defined resolution. The muon reconstruction efficiency is zero outside the tracker acceptance and for muon momenta below a certain threshold, to avoid looping particles.

Electrons

Typical electron identification requires combining information from the tracking system and the calorimeter. DELPHES avoids this necessity by parameterising the electron reconstruction efficiency as a function of energy and η. The electron energy resolution is a combination of the tracker and the ECAL resolutions, such that the tracker resolution dominates at low energy, while at high energy,

the calorimeter resolution dominates. Electron identification efficiency vanishes outside the acceptance of the tracker and below a certain energy threshold.

Photons

The reconstruction of photons relies solely on the ECAL. The final photon energy is obtained from applying the ECAL resolution to the original photon. Photon conversions into e^+e^- pairs are neglected, and electrons with no reconstructed track that reach the ECAL are reconstructed as photons.

Currently, DELPHES does not include a fake rate for electrons, muons or photons. The fake rate parameterises the possibility of a certain object (for example, a jet) being misidentified as a lepton or photon. In physical analyses with multi-lepton final states, the lepton fake rates are important to correctly determine the expected contribution of each background process to the analysis yield.

For a lepton or photon to be reconstructed, an isolation criterion must be met. The isolation variable I of a particle P is defined as the sum of the p_T of all particles with p_T above a threshold p_T^{min} and within a cone of $\Delta R < R$ around that particle, normalized to the p_T of P. The particle is said to be isolated if the condition $I < I_{min}$ is verified. The default values of the parameters are $p_T^{min} = 0.1\,\mathrm{GeV}$, $R = 0.5$ and $I_{min} = 0.1$.

2.4.4 Particle-flow reconstruction

The particle-flow approach aims to obtain the best measurements, using information available from all subdetectors. In real experiments, the momentum of a charged particle can be estimated from the particle track or from the calorimeter. The preferred measurement depends on an energy threshold, below which the momentum resolution obtained from the track is better. Above the threshold, the calorimeter energy deposit is more reliable to estimate the momentum. In the particle-flow phase of the simulation, if a track exists for a certain particle, information from the track is always preferred.

The particle-flow algorithm creates two sets of 4-vectors, which will serve as input for the subsequent reconstruction of jets and $\not{E}_T$. These 4-vectors include

particle-flow tracks and particle-flow towers. A particle-flow track is created for every track in the inner tracker. Then, for each calorimeter tower, the energy deposits originating from particles with reconstructed tracks are subtracted. If the remaining energy E_{tower} is positive, a particle-flow tower is created with this energy and with the direction of the tower (η, ϕ) coordinates.

This definition implies that particle-flow tracks include charged particles, measured with good resolution, and that particle-flow towers include a combination of neutral particles, charged particles without a track, and excesses in deposits originating from the smearing process of the calorimeters, all measured with worse resolution. While very simple when compared to what is actually required in real experiments, this algorithm reproduces well the performance achieved at LHC experiments.

2.4.5   Jets

Jet reconstruction can be performed using one of three different collections of objects as input: the long-lived particles resulting from parton shower and hadronization, the calorimeter towers, or the particle-flow tracks and particle-flow towers. DELPHES integrates the FastJet [78] package, making it possible to choose one among the most common jet clustering algorithms, and setting the corresponding parameter values. A minimum p_T threshold for a jet to be stored in the final jet collection can also be set. In order to avoid double-counting, DELPHES automatically removes jets which have already been reconstructed as leptons or photons.

b and τ jets

The algorithm for b and τ jet identification is purely parametric. A jet can potentially be identified as a b or τ candidate if its direction is within a certain ΔR cone relative to a generated b or τ, respectively. Given this condition, the probability for the jet to be identified as a b or τ will be given by user-defined parameterisation of the tagging efficiency. A mis-tagging efficiency can also be introduced, leading to the realistic possibility that a particle other than b or τ can be identified as such.

2.4.6   Missing E_T and H_T

The total transverse energy H_T is defined as the scalar sum of the p_T of all detected particles.

Just as was the case with jets, the missing E_T and the H_T can be estimated from one of three object collections: generated particles, calorimeter towers, or particle-flow tracks and particle-flow towers.

3   Analysis

The spin correlation of top quark pairs in $t\bar{t}$ production in the dilepton topology has been extensively studied with the use of angular observables [62]. It has been suggested that observables of this kind provide an useful separation of $t\bar{t}H$ signal and irreducible backgrounds ($t\bar{t}b\bar{b}$ and $t\bar{t}\gamma\gamma$), with a gain in signal-to-background ratio of up to 15%, depending on the particular observable [79].

The work presented here is motivated partially by that suggestion. It consists in a phenomenological study, with the primary goal of finding a set of angular observables with potential for discriminating $t\bar{t}H$ events, in which the $t\bar{t}$ system decays dileptonically and the Higgs boson decays through $H \to b\bar{b}$, from the dominant background, $t\bar{t}b\bar{b}$, under the experimental conditions of the LHC Run 2. The discrimination between SM $t\bar{t}H$ and non-SM $t\bar{t}A$ (where A stands for a pure pseudoscalar Higgs boson) is also a strong motivation for such a study.

In the present section, the main steps of this study are exposed, and a description of the employed tools is given.

Firstly, the generation of the simulated event samples used and the general-purpose detector simulation are described. Secondly, the signal and background processes are identified. Then, the analysis routine is explained in detail. This includes the event selection cuts, the truth-matching procedure, used to match generated particles to the corresponding reconstructed objects, and the kinematic fit, which is a distinctive feature of this analysis used to reconstruct the undetected neutrinos and the intermediate heavy particles in the events. Strategies for reducing combinatorial background are also discussed.

After describing the analysis routine, a motivation is given for the type of angular observables studied, using the helicity formalism. The computation of these observables is also explained. Also, a review is presented of observables proposed in previous studies as potentially discriminating between $t\bar{t}H$ and $t\bar{t}A$ production.

Finally, a set of angular observables is chosen, and five different multivariate analyses are performed using TMVA (Toolkit for Multivariate Data Analysis with ROOT) [80]. The goal of the TMVA analyses is to quantify the potential gain in signal sensitivity coming from the use of the proposed variables. The first TMVA

analysis is performed with the set of variables used in the correponding ATLAS analysis at 8 TeV [81]; the second analysis is performed with the set of angular observables chosen previously; the third analysis is performed with the same observables, but using truth-matched objects to separate out the effect of combinatorial errors; the fourth analysis is performed using both the ATLAS variables and the chosen angular observables, to evaluate the sensitivity gain from using the new variables; finally, the fifth analysis also uses these variable sets, but it is run over a background sample including all the main SM backgrounds. It serves as a validation for the initial choice of only considering the $t\bar{t}b\bar{b}$ background.

3.1 Generation of samples

The generation of signal and background simulated event samples was done using MADGRAPH5 [82]. The decay chains of heavy particles (Higgs bosons, top quarks and vector bosons) were handled by the MADSPIN [63] module. This module ensures that spin correlation effects are preserved during particle decays. The resulting events were then passed to PYTHIA6 [83] for simulation of parton-showers and hadronization. Parton radiation is described differently in the event generator and in the parton-shower and hadronization software. In order to avoid phase-space overlap from these different descriptions in multijet events, a matching procedure was done in the MLM scheme [84], which does so by rejecting events lying on the region of overlap.

The events were generated for pp collisions at a centre-of-mass energy of 13 TeV. The Higgs boson and the top quark masses were set to $m_H = 125$ GeV and $m_t = 173$ GeV, respectively.

The $t\bar{t}H$ and $t\bar{t}A$ signal samples, as well as the $t\bar{t}b\bar{b}$ background, were generated at next-to-leading order (NLO) in QCD. For the signal samples $t\bar{t}H$ and $t\bar{t}A$, 2.5×10^6 events were generated. For the background $t\bar{t}b\bar{b}$, 1.6×10^6 events were generated.

Additional backgrounds were generated at leading order (LO) in QCD, with up to 4 light-flavoured jets (not b-jets), to account for additional radiation contributions. A list of all the generated samples is shown in table 5, with details about the generation, including the maximum number of additional jets gener-

Table 5: Generated event samples. For each sample, the table lists the order in QCD at which the sample was generated, the maximum number of additional light-flavoured jets allowed in the production, the enabled decays in MADSPIN, and the product of cross-section and branching ratio returned by MADGRAPH5. Leptonic decays include only electrons and muons, and exclude taus.

Process	QCD order	No. of jets	Enabled decays	$\sigma \times$BR (pb)
$t\bar{t}H$	NLO	0	$H \to$ all, $t\bar{t} \to$ dileptonic	2.28×10^{-2}
$t\bar{t}A$	NLO	0	$A \to$ all, $t\bar{t} \to$ dileptonic	9.12×10^{-3}
$t\bar{t}b\bar{b}$	NLO	0	$t\bar{t} \to$ dileptonic	7.90×10^{-1}
$t\bar{t} + jets$	LO	3	$t\bar{t} \to$ dileptonic	3.15×10^{1}
$t\bar{t}V + jets$ $(V = W, Z)$	LO	1	$t\bar{t} \to$ dileptonic, $V \to$ all	6.18×10^{-2}
Single t	LO	0	$t \to$ leptonic	9.63×10^{1}
$Z + jets$	LO	4	$Z \to$ leptonic	3.12×10^{3}
$Zb\bar{b} + jets$	LO	2	$Z \to$ leptonic	1.23×10^{2}
$W + jets$	LO	4	$W \to$ leptonic	3.45×10^{4}
$Wb\bar{b} + jets$	LO	2	$W \to$ leptonic	2.89×10^{2}
$VV + jets$ $(V = W, Z)$	LO	3	$V \to$ all	1.33×10^{2}

ated, the decay channels enabled in MADSPIN, and the product of cross-section and branching ratio returned by MADGRAPH5.

3.2   Detector simulation

A fast simulation of a general-purpose detector was performed in DELPHES [17], using the default ATLAS parameters, without pile-up simulation. A summary of those parameters is given in the present section.

The particle propagation phase takes place in a cylinder with radius $r = 1.15\,\text{m}$ and length $l = 7.02\,\text{m}$, under a magnetic field of $2\,\text{T}$.

Particle tracking efficiencies are defined in segments of $|\eta|$ and p_T , and separately for charged hadrons, muons and electrons. All tracking efficiencies are zero for $p_T \leq 0.1\,\text{GeV}$ or $|\eta| > 2.5$. They remain relatively small for charged particles with $0.1\,\text{GeV} < p_T < 1\,\text{GeV}$, ranging from 50% to 75%. The efficiency for tracking charged hadrons with higher p_T is always larger than 85%. For electrons, it is always larger than 83%, and for muons always larger than 98%. Momentum resolution for charged tracks is defined in segments of $|\eta|$ and p_T, and it is never larger than 5%.

61

The energy resolution δE for electrons is defined as a function of energy of the form

$$\left(\frac{\delta E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{20}$$

where the parameters S, N and C are defined in segments of $|\eta|$ and energy. For $E \leq 25\,\text{GeV}$ and $|\eta| \leq 2.5$, $C = 0.015$ is the only non-zero term, which means constant fractional resolution of 1.5%. For electrons with $E > 25\,\text{GeV}$ and $|\eta| \leq 2.5$, or any energy and $2.5 < |\eta| \leq 3.0$, the parameters S, N and C are such that the resolution matches the 1.5% value at $E = 25\,\text{GeV}$, and drops asymptotically to 0.5% for higher energies. For forward ($|\eta| > 3.0$) electrons, the high-energy asymptotical limit for the energy resolution is 10.7%.

The muon momentum resolution is defined in segments of $|\eta|$ and p_T. In general, the resolution is worse for higher p_T and higher $|\eta|$, with its maximum at 10%, for $p_T > 100\,\text{GeV}$ and $1.5 < |\eta| \leq 2.5$.

Calorimeters are segmented in (η, ϕ) rectangular cells. In the region with $|\eta| \leq 2.5$, the cells have dimensions $(\eta, \phi) = (0.1, 10^\text{o})$, and for $|\eta| > 2.5$, their size is $(\eta, \phi) = (0.2, 20^\text{o})$. The ECAL and HCAL resolutions are defined as a function of energy, of the same kind as in equation (20). Parameters S, N and C are defined in segments of $|\eta|$. The energy deposition fractions f_{ECAL} and f_{HCAL} are defined for electrons, photons and π^0 mesons as $f_{ECAL} = 1$, $f_{HCAL} = 0$. For kaons and Λ particles, $f_{ECAL} = 0.3$ and $f_{HCAL} = 0.7$. For muons and neutrinos, both energy deposition fractions are zero. For any other long-lived particle, $f_{ECAL} = 0$, $f_{HCAL} = 1$.

Photon, electron and muon identification efficiencies are parameterised in segments of $|\eta|$. All identification efficiencies are 95% in the central region $|\eta| \leq 1.5$, 85% in the intermediate region $1.5 < |\eta| \leq 2.5$ (2.7 for muons), and zero for $|\eta| > 2.5$ (2.7 for muons) or $p_T < 10\,\text{GeV}$. The parameters used for checking the isolation of electrons, photons and muons, as defined in section 2.4, are set to $p_T^{min} = 0.5$, $R = 0.5$ and $I_{min} = 0.1$.

Jet reconstruction is performed by the anti-k_t algorithm [74], with the ΔR parameter set to 0.6. Only jets with $p_T > 20\,\text{GeV}$ are stored in the event information. The efficiency for b-tagging is given separately for b-jets and c-jets, as an asymptotically increasing function of p_T, in two segments of $|\eta|$. For b-jets (c-jets),
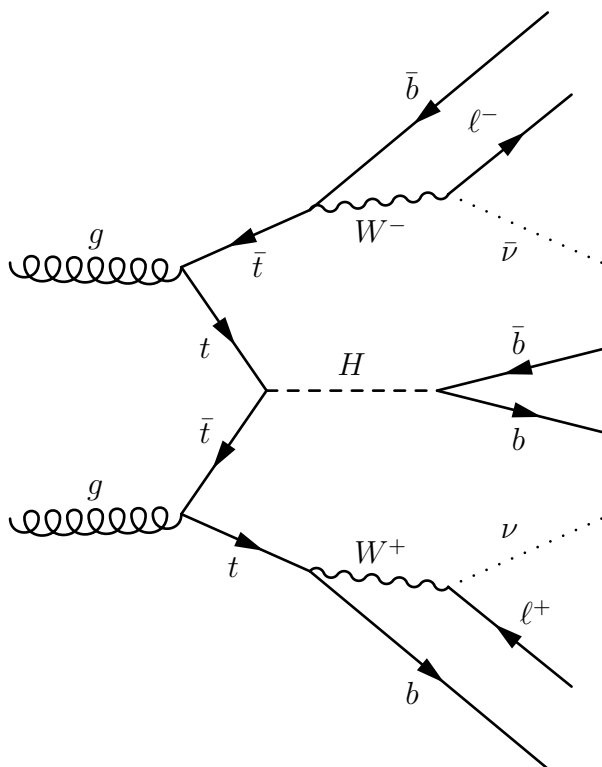
Figure 18: Signal process: $t\bar{t}H$ production (in this example, through $t\bar{t}$ fusion), followed by dileptonic decay of the $t\bar{t}$ system and decay of the Higgs boson into $b\bar{b}$. The resulting final state includes 2 charged leptons, 4 b-jets and 2 neutrinos.

the b-tagging efficiency is limited to 50% (20%) in the $|\eta| \leq 1.2$ segment and to 40% (10%) in the $1.2 < |\eta| \leq 2.5$ region. It is zero for jets with $p_T \leq 10\,\mathrm{GeV}$ or $|\eta| > 2.5$. A default constant b-tagging efficiency is set to 0.1% for any other jet, to simulate a misidentification rate. In most of the studies presented below, b-tagging information was not used at all. It was used only in the TMVA analyses, to define the signal-rich region in which the analyses are performed.

3.3   Signal and background identification

The analysis presented here is a search for Higgs bosons produced in association with a $t\bar{t}$ pair, with the Higgs boson decaying through $H \to b\bar{b}$ and the $t\bar{t}$ system decaying through the dileptonic channel. An example of a signal event diagram is shown in figure 18. Each of the final state leptons must be either an electron or a muon. Decays into τ-leptons are not considered as signal in the present analysis.

The τ-lepton decays into hadrons or into a less massive lepton and 2 neutrinos. Either one of those decays "spoils" the desired resolution in a leptonic final state.

The $t\bar{t}H$ production is a rare process, with a cross-section of about 129 fb in pp collisions at a centre-of-mass energy of 8 TeV and about 509 fb in collisions at 13 TeV [43]. On top of that, only about 4% of $t\bar{t}$ decays are dileptonic (already excluding final states with τ). However, this final state with isolated leptons is very distinct from the multijet background, and the relative lack of statistics is compensated by the improved signal-to-background ratio. Additionally, if a kinematic fit is to be performed, the relatively lower number of jets in the final state minimizes the chances of making a mistake during the necessary assignment of reconstructed jets to the decay products of H, t and $\bar{t}$.

The production of $t\bar{t}A$, with the $t\bar{t}$ system decaying dileptonically and the pure pseudoscalar Higgs boson decaying as $A \to b\bar{b}$, is also studied. The expected final state objects are the same as with the scalar Higgs boson. The pure pseudoscalar scenario has been excluded with a 99.98% confidence level for the observed 125 GeV Higgs boson [54, 55]. Nevertheless, observables discriminating between the pure scalar and pure pseudoscalar scenarios are still of great interest. Since there is a possibility that the Higgs boson is in a mixture of a CP-even and a CP-odd state, these observables can be used to measure and set limits on this possible pseudoscalar component.

The main source of background for this search comes from the production of top quark pairs, decaying through the dileptonic mode, in association with at least two jets. This source can be separated in two categories: the irreducible background $t\bar{t}b\bar{b}$, in which the $t\bar{t}$ pair is produced together with a correctly identified $b\bar{b}$ pair, resulting in the same exact final state objects as the signal; and $t\bar{t} + jets$, where the produced jets come from $c\bar{c}$ production or lighter flavour quark pairs and are misidentified as b-jets. Background processes besides $t\bar{t} + jets$ include all production processes that may result in a dileptonic final state, together with energetic jets. The most important ones are $t\bar{t}V$, where the $t\bar{t}$ pair is produced in association with a vector boson ($V = W, Z$), $W + jets$, $Wb\bar{b} + jets$, $Z + jets$, $Zb\bar{b} + jets$, diboson production (ZZ, WZ and WW) and single top production.

Table 6 shows the expected cross-sections of signal and background processes from proton collisions at 13 TeV, and the corresponding expected number of events

in a data sample with an integrated luminosity of $20\,\mathrm{fb}^{-1}$. The cross-sections were extracted from the event simulation, as described in section 3.1, with the exception of $t\bar{t} + jets$, single t, and $t\bar{t}H$, for which theoretical values are available, more accurate than the ones from the MADGRAPH5 generation at LO (NLO for $t\bar{t}H$ and $t\bar{t}b\bar{b}$). For $t\bar{t} + jets$, the cross-section is predicted to be 832^{+40}_{-46} pb for a top quark mass of $172.5\,\mathrm{GeV}$. It has been calculated at next-to-next-to-leading order (NNLO) in QCD, including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms with Top++2.0 [85]. For single-t, the reference values from NLO calculations recommended by the LHC Top Physics Working Group were used [86]. For $t\bar{t}H$, the cross-section obtained from the MadGraph5 NLO generation was multiplied by a scale factor obtained from NLO calculations with NLL resummation corrections [87]. Number of events is also shown after the requirement of 2 leptons (e or μ) in the final state. No events from the $W + jets$ background survive this requirement because we are only considering 'prompt' leptons in the final state. In reality, the leading contribution of $W + jets$ to dilepton final states requires the leptonic decay of W and the reconstruction of one 'fake' lepton. A 'fake' lepton can result from a leptonic decay inside a jet, for example, in which the lepton is radiated in a direction far enough from the cone of the jet for it to meet isolation criteria. The effect of further requiring 4 b-quarks in the final state is also shown. Again it should be noted that parton-level b-quarks are considered, neglecting the inefficiency in b-tagging jets and the possible appearence of 'fake' b-jets. In order to obtain the total production cross-sections from the product $\sigma \times BR$ returned by MADGRAPH5, as well as to obtain the expected number of events with leptons and b-quarks in the final state, the branching ratios listed by the Particle Data Group [19] were used.

Further requiring the invariant mass of two opposite-sign leptons ($m_{\ell\ell}$) to be outside a window of $20\,\mathrm{GeV}$ around the Z boson mass peak ($\sim 91\,\mathrm{GeV}$) rejects most of the $Z + jets$ background. This exercise assumes perfect lepton identification and b-tagging, which is far from the real experimental conditions. However, two valid conclusions can be retained from it. The first is that requiring dilepton final states with more than 2 b-jets and with $m_{\ell\ell}$ far from the Z boson mass peak ($|m_{\ell\ell} - m_Z| > 10\,\mathrm{GeV}$) mitigates most of the background processes at the LHC. The second is that, after such a selection, $t\bar{t}$ is the dominant background, with

Table 6: Cross-sections of signal and background processes, from proton collisions at $\sqrt{s} = 13\,\text{TeV}$, and the corresponding expected number of events in $20\,\text{fb}^{-1}$ of data. Cross-section values were obtained from the MadGraph5 simulation, with the exception of $t\bar{t} + jets$, single t, and $t\bar{t}H$, for which more accurate values were used, respectively from [85], [86] and [87]. Also shown is the expected number of events with at least 2 leptons in the final state, and also with at least 2 leptons and 4 b-quarks.

Process	σ (pb)	N (total)	$N\ (\geq 2\ell)$	$N\ (\geq 2\ell, \geq 4b)$
$t\bar{t}H$	5.50×10^{-1}	1.10×10^{4}	8.69×10^{2}	2.89×10^{2}
$t\bar{t}b\bar{b}$	1.74×10^{1}	3.48×10^{5}	1.58×10^{4}	1.58×10^{4}
$t\bar{t} + jets$	8.54×10^{2}	1.71×10^{7}	7.76×10^{5}	0
$t\bar{t}V + jets$	1.36×10^{0}	2.72×10^{4}	3.06×10^{3}	1.07×10^{2}
Single t	3.03×10^{2}	6.06×10^{6}	6.60×10^{4}	0
$Z + jets$	4.66×10^{4}	9.31×10^{8}	6.24×10^{7}	0
$Zb\bar{b} + jets$	1.84×10^{3}	3.67×10^{7}	2.46×10^{6}	2.46×10^{6}
$W + jets$	1.62×10^{5}	3.24×10^{9}	0	0
$Wb\bar{b} + jets$	1.36×10^{3}	2.71×10^{7}	0	0
$VV + jets\ (V = W, Z)$	1.33×10^{2}	2.66×10^{6}	1.56×10^{5}	0

a much larger expected number of events than $t\bar{t}H$. In table 6, $t\bar{t}b\bar{b}$ is the only component of $t\bar{t}$ remaining after the selection. Experimentally, this is not the case, since the misidentification rate of light jets as b-jets is enough for $t\bar{t} + jets$ to be of the same magnitude as $t\bar{t}b\bar{b}$, even after the requirement of more than 2 b-tagged jets.

The background mainly addressed here is the dominant irreducible background $t\bar{t}b\bar{b}$, which is the most challenging one. In a first approach, the other background processes will not be discussed at all and it will be assumed that the methods found to deplete the $t\bar{t}b\bar{b}$ background will be effective on a complete set of backgrounds. Eventually, this assumption is justified by conducting the final analysis on an inclusive bakground sample, and confirming that it yields results similar to the $t\bar{t}b\bar{b}$ sample alone.

3.4 Reconstruction routine

The main analysis tool used in this work is a reconstruction routine implemented in MadAnalysis5 [88] expert mode. This routine reads all the generated and simulated events. It processes events one by one and, for each event, it stores all the variables necessary for the later physics studies. This record of event information is organized in three levels, with the same set of variables replicated for each level. MadAnalysis5 outputs files in Root format, and the Root [89] framework is used for further data processing and visualization.

The information stored in the first level (generator level, or parton level) is obtained directly from the generated particles, before quarks and gluons hadronize and before any interaction with the detector takes place. At this level, the particles in the $t\bar{t}H$ or $t\bar{t}b\bar{b}$ decay chain are identified, from t and $\bar{t}$ (and H in $t\bar{t}H$ events) down to the neutrinos, leptons and b-quarks, and including the intermediate W bosons. Besides, for a given final state particle, the decay chain it came from is also known. This means, for example, that the b-quark resulting from the t-quark decay is immediately identified as such, and will not be misidentified as the b-quark from the H decay. At this level, the momenta of the neutrino and the anti-neutrino resulting from the $t\bar{t}$ dileptonic decay are directly obtained from the generator information.

In the second level (reconstruction with truth-match), particles are reconstructed using combined information from the detector objects (leptons, jets, $\not{E}_T$) and the generated particles. This combination consists of a truth-matching process. Each parton-level particle is matched to a detector object, using a proximity criterion in (η, ϕ) space. However, the four-momenta of the particles used are the ones obtained from the detector objects. The expected result in this level is to reproduce the resolution effects and inefficiencies of the detector on reconstructed particles, but still to assign correctly, in most of the events, the reconstructed leptons and jets to the corresponding generated leptons and quarks. The momenta of the neutrino and the anti-neutrino are lost (only the resulting missing transverse energy is reconstructed), and these particles, along with all the intermediate particles in the $t\bar{t}$ decay chain, must be reconstructed using a kinematic fit, described later in this section.

The third level (reconstruction without truth-match) does not require information from the generated particles at all. The lepton charge must be correctly identified (which, in DELPHES, is the case 100% of the time) and 4 out of all the reconstructed jets must be assigned to the 4 b-quarks in the final state of the $t\bar{t}H$ or $t\bar{t}b\bar{b}$ event. The kinematic fit is also used to reconstruct the $t\bar{t}$ decay chain, while at the same time it excludes possible choices of 4 jets that are not consistent with the decay of the $t\bar{t}$ system. The reconstruction level without truth-match is the one that most closely reproduces experimental conditions, and this study is based on the premise that any variable or set of variables that provide discrimination between samples at this level may be of interest to study with real data.

3.4.1 Selection cuts

In order to assure that it is possible to apply the reconstruction routine successfully, and because signal events are expected to result in high-energy final state particles, events are required to have at least two leptons with transverse momentum not below $20\,\text{GeV}$ and $|\eta| \leq 2.5$, and at least four jets meeting the same p_T and η requirements. At this stage, no requirement on the b-tagging of jets is made.

If a data analysis were to be carried out instead of this phenomenological study, event selection would begin at the trigger. Since the signal has a dileptonic final state, events in which the single lepton trigger was activated should be used. This trigger is highly efficient, especially in dileptonic events, and rejects most of the events with purely hadronic final states, abundant at the LHC. The requirement of some b-tagged jets would also be an effective way of suppressing important backgrounds, such as $t\bar{t} + jets$ and vector bosons, which produce mainly light-flavoured jets.

In the official ATLAS analysis of the same channel at $\sqrt{s} = 8\,\text{TeV}$ [81], the events considered were collected using a single-electron or single-muon trigger, and only while the beam conditions were stable and all subdetectors were operational. Only jets reconstructed with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$ are considered, and events must have at least 2 b-tagged jets. Furthermore, in that analysis, events are separated in 3 categories according to the flavour of the leptons: ee, $e\mu$ and $\mu\mu$. In the ee and $\mu\mu$ categories, additional cuts are performed on the invariant mass

Table 7: Summary of selection cuts used, compared to the necessarily more sophisticated ATLAS analysis of the same channel [81].

Variable	Condition	ATLAS selection condition
No. of leptons	≥ 2	$= 2$ (opposite charge)
No. of jets	≥ 4	≥ 2
Lepton p_T	$\geq 20\,\text{GeV}$	$> 25\,\text{GeV}$ (for the highest-p_T lepton) $> 15\,\text{GeV}$ (for the other lepton)
Lepton $|\eta|$	≤ 2.5	< 2.5 for muons only
Jet p_T	$\geq 20\,\text{GeV}$	$> 25\,\text{GeV}$
Jet $|\eta|$	≤ 2.5	< 2.5
No. of b-tagged jets	-	≥ 2
$m_{\ell\ell}$	-	$> 15\,\text{GeV}$ and $|m_{\ell\ell} - m_Z| > 8\,\text{GeV}$ for ee and $\mu\mu$
H_T	-	$> 130\,\text{GeV}$ for $e\mu$

of the 2-lepton system, $m_{\ell\ell}$, to exclude resonances decaying to same-flavour lepton pairs. Requiring $m_{\ell\ell} > 15\,\text{GeV}$ excludes hadronic resonances such as J/ψ and Υ, and $|m_{\ell\ell} - m_Z| > 8\,\text{GeV}$ excludes the Z boson resonance. In the $e\mu$ category, the scalar sum of the transverse energies of leptons and jets, H_T, is required to be above 130 GeV.

In table 7, the selection cuts used in the present work are summarized, and are compared to the ATLAS analysis. All the studies discussed in the following sections were performed after these selection cuts were applied. The fraction of generated events that pass the selection cuts is 16.3% in the $t\bar{t}H$ event sample (resulting in 4.08×10^5 events), 17.1% in the $t\bar{t}A$ event sample (resulting in 4.28×10^5 events), and 11.0% in the $t\bar{t}b\bar{b}$ event sample (resulting in 1.76×10^5 events).

3.4.2 Truth-matching

At the reconstruction level with truth-match, the matching between detector objects and generator level particles is done in the following way. First, the re-constructed leptons are matched to the generated ones. Leptons are required to lie inside a cone of $\Delta R < 0.1$ relative to the direction of the parton-level lepton to which they are being matched. A loop runs over all reconstructed leptons. The reconstructed lepton with the lowest ΔR relative to the generated positive lepton

is identified as the positive lepton. The information about the reconstructed lepton charge is not used. After this identification, the same procedure is used for the negative lepton.

Once the leptons are matched, a similar procedure is used to match the jets. A reconstructed jet is required to have $\Delta R < 0.5$ relative to the parton-level quark to which it is being matched. A loop runs over all possible ways of assigning 4 out of all the reconstructed jets to the 4 generated quarks. Each possibility can be represented by a vector of 4 distinct reconstructed jets, j_i, with $1 \le i \le 4$. If the vector of generated quarks is q_i, the matching (j_i, q_i) is done for the combination that minimizes $\sum_{i=1}^{4} \Delta R_{j_i,q_i}^2$.

The implemented truth-matching procedure is a very simple one. Two improvements that would lead to a more sophisticated truth-matching procedure, and possibly better results, are suggested here, but were not tested. The first one would be to include jet energies or p_T in the truth-matching. Currently, only the jet direction is considered, allowing for a parton-level jet to be matched to a reconstructed jet, even if their energies are very different. One possible way to include the jet energy E would be to define $\Delta R' \equiv \sqrt{\Delta\phi^2 + \Delta\eta^2 + (\Delta E/E_{\text{parton}})^2}$, and use this new distance in place of ΔR in the procedure explained above. The second improvement would be to allow more than one reconstructed jet to be matched to the same quark, since it is possible for a quark to radiate an energetic gluon before hadronizing into a jet, such that the gluon itself originates an additional jet. In that case, the momentum of the original quark would be best described by the system composed of the two resulting jets. Combinations of jets could be tried within the $\Delta R < 0.5$ cone around the parton-level quark, for example.

In spite of these limitations, truth-matched objects provide an estimate of the detector effects on reconstruction. In figure 19, the p_T resolution is shown for jets (left) and leptons (right). The resolution $\Delta p_T / p_T^{parton}$ is defined as the difference between reconstructed p_T and parton-level p_T, normalized to the parton-level value. Jet p_T resolution is worse than that of leptons, and it is expected to propagate to the reconstruction of intermediate particles. Furthermore, jets coming from the H decay have worse resolution than those coming from $t\bar{t}$ decay. Since DELPHES parameterises the resolution as a decreasing function of energy, this difference is likely due to the jets from a t or $\bar{t}$ decay being more energetic

70
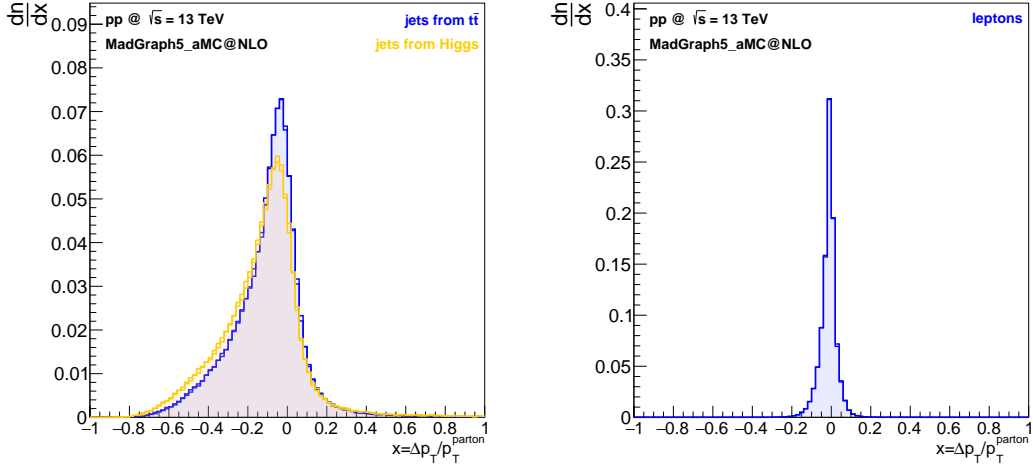
Figure 19: Resolution in p_T of the reconstructed b-jets (left) and leptons (right). The resolution $\Delta p_T / p_T^{parton}$ is defined as the difference between reconstructed p_T and parton-level p_T, normalized to the parton-level value. The bin content dn/dx is in units of fraction of total number of events per bin, where only events with successful truth-match and at least one kinematic fit solution are considered for the total. Jet p_T resolution is visibly worse and is the dominating the resolution effect when reconstructing intermediate particles, like t, $\bar{t}$ and H.

than the ones from a H decay.

3.4.3 Kinematic fit with truth-match

After the dileptonic decay of a $t\bar{t}$ system, the top and anti-top quarks can be reconstructed from their respective decay products. However, the neutrino and anti-neutrino momenta are not available from the detector, since neutrinos are not detectable in general-purpose experiments. It is assumed that the missing transverse energy corresponds to the transverse projection of the summed momenta of the neutrino and the anti-neutrino. That is,

$$
\begin{aligned}
p_\nu^x + p_{\bar{\nu}}^x &= \not{E}^x \\
p_\nu^y + p_{\bar{\nu}}^y &= \not{E}^y.
\end{aligned}
\tag{21}
$$

To find solutions in which all components of the momenta of the neutrino and anti-neutrino are fully determined (six quantities in total), at least four additional equations are neccessary. Conservation of momentum in a particle decay (for example $t \to W^+ b$) imposes that the momentum four-vector P of the decaying particle is equal to the sum of the momenta of the decay products (for example,

$P_t = P_{W^+} + P_b$). Using the relation $P^2 = m^2$ and setting the masses of W^+, W_-, t and $\bar{t}$ to fixed values, the four necessary equations are obtained:

$$
\begin{aligned}
(P_\nu + P_{\ell^+})^2 &= m_{W^+}^2 \\
(P_{\bar{\nu}} + P_{\ell^-})^2 &= m_{W^-}^2 \\
(P_{W^+} + P_b)^2 &= m_t^2 \\
(P_{W^-} + P_{\bar{b}})^2 &= m_{\bar{t}}^2,
\end{aligned}
\tag{22}
$$

where $P_{W^+} = P_{\ell^+} + P_\nu$ and $P_{W^-} = P_{\ell^-} + P_{\bar{\nu}}$, making clear the dependence on the unknowns P_ν and $P_{\bar{\nu}}$. The quadratic nature of these equations implies that there may be several solutions for the momenta of the neutrino and anti-neutrino, all meeting the mass constraints and resulting in the observed missing transverse energy.

To choose one particular solution among these, a likelihood method was used. We define $P_V(V')$ as the probability density function (p.d.f.) of the variable V, evaluated at $V = V'$. p.d.f.s can be generalized to more than one dimension. For example, in two dimensions, $P_{V,U}(V', U')$ is the probability density function of two variables, V and U, evaluated at $(V, U) = (V', U')$. The use of p.d.f.s of dimension two or larger is especially important when dealing with correlated variables. In most cases presented here, V' and U' will not take particular values, and the abbreviated notation $P(V') \equiv P_V(V')$ can be used without ambiguity.

For this kinematic fit, the solution chosen is the one that maximizes the following likelihood:

$$
\frac{1}{p_{T\nu} p_{T\bar{\nu}}} P(p_{T\nu}) P(p_{T\bar{\nu}}) P(m_t, m_{\bar{t}}) P(p_{Tt}) P(p_{T\bar{t}}) P(p_{Tt\bar{t}}).
\tag{23}
$$

The p.d.f.s used were obtained from generator-level $t\bar{t}H$ events, with the exception of $P(m_{b_H \bar{b}_H})$, which will be addressed below, in the discussion of the kinematic fit without truth-match. For the definition of a p_T p.d.f., the starting point is a 50-bin histogram filled with the parton-level p_T distribution. The relatively small number of bins means that statistical fluctuations are small in this histogram. However, if it were to be used as a p.d.f., abrupt changes in the probabitlity would appear on the edges of each bin. Besides, p_T distributions drop to zero rapidly

as p_T also goes to zero, a behaviour that cannot be captured by the wide bins in this histogram. In order to correct for these effects, another histogram is filled with a copy of the contents of the 50-bin histogram, but instead it has a total of 800 bins, such that each original bin is now subdivided in 16 bins with equal content. Then, the content of the first bin is set to 0. A smoothing procedure available in ROOT is applied 10^5 times to this histogram, and the result is used as the p.d.f.. The 2D p.d.f.s of the top quarks and W boson masses are simply filled with the parton-level distributions, in bins of $1\,\text{GeV} \times 1\,\text{GeV}$. In these p.d.f.s, the top quark masses range from $170\,\text{GeV}$ to $176\,\text{GeV}$, and the W boson masses range from $76.4\,\text{GeV}$ to $84.4\,\text{GeV}$. Outside these ranges, the p.d.f. is zero.

Figure 20 shows some of the p.d.f.s used. The p_T p.d.f.s are overlapped with the 50-bin histograms from which they were obtained. Outside the ranges represented, the p.d.f.s are set to zero. Units are not shown on the vertical axes, but they must be such that the p.d.f.s are normalized. $P(p_{T\bar{\nu}})$, $P(p_{T\bar{t}})$ and $P(m_{W^-}, m_{\bar{t}})$ are not shown, but they are similar to their conjugates.
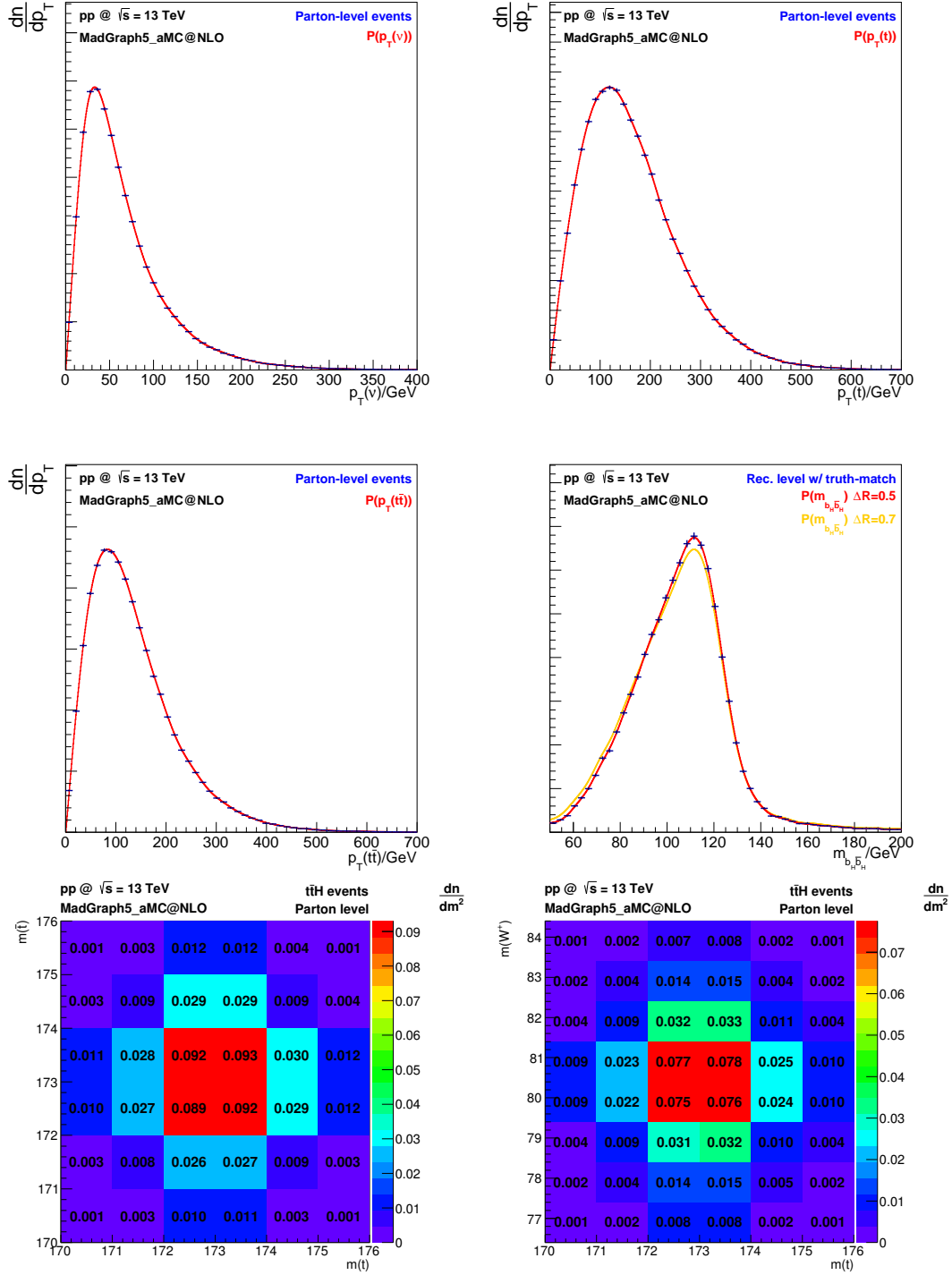
Figure 20: P.d.f.s used in the kinematic fit. The p_T p.d.f.s and $m_{b_H \bar{b}_H}$ are overlapped with the 50-bin histograms from which they were obtained. Units are not shown on the vertical axes, but they must be such that p.d.f.s are normalized. P.d.f.s are set to zero outside the ranges represented here. $P(p_{T \bar{\nu}})$, $P(p_{T \bar{t}})$ and $P(m_{W^-}, m_{\bar{t}})$ are not shown, but they are similar, respectively, to $P(p_{T \nu})$, $P(p_{T t})$ and $P(m_{W^+}, m_t)$. Also represented is the obtained p.d.f. for $m_{b_H \bar{b}_H}$ if the ΔR cut for truth-match were set to 0.7, instead of 0.5.

Each p.d.f. in the product (23) is introduced in order for the p_T and mass distributions of reconstructed events to resemble the ones at parton level. When only the p.d.f. factors are used, there is a bias towards solutions in which neutrinos have higher energies than at generator level. The origin of this effect is likely due to the next-to-leading-order generation of signal events. This order of generation allows a gluon to be radiated by any coloured particle during the event. If the gluon carries enough energy, it will originate a whole different jet in the event. The kinematic fit is designed in a leading-order philosophy, in the sense that it will not try to combine more than one jet to recover intermediate particles. The fit procedure must somehow compensate for this NLO radiation (as well as any detector losses) in order to reconstruct reasonable top quark and W masses. This compensation can only happen through the free parameters of the fit: the momenta of the neutrinos, which become systematically larger. The initial factor $1/(p_{T_\nu} \times p_{T_{\bar{\nu}}})$ is introduced to compensate for this effect, by favouring solutions with lower-p_T neutrinos.

To set the values used as the masses of the W^+, W^-, t and $\bar{t}$, mass variations must be allowed, but special care must be taken to preserve the expected phase-space density of states. For example, setting the mass value of one of the top quarks, immediately constrains the available phase-space for the event, and it is not straightforward to assume that the mass distribution of the other top quark is independent from this constraint. Likewise, the mass of a decaying top quark affects the allowed mass distribution of the resulting W boson.

The way this is implemented in the kinematic fit is the following. Firstly, the t and $\bar{t}$ masses are randomly generated, according to the 2D p.d.f., $P(m_t, m_{\bar{t}})$. This p.d.f., obtained from generator-level events, is expected to contain any existing correlations between the t and $\bar{t}$ masses. Then, the W^+ and W^- masses are generated according to the $P(m_{W^+}, m_t)$ and $P(m_{W^-}, m_{\bar{t}})$ p.d.f.s, respectively, considering only the "slice" of bins containing the value of the t or $\bar{t}$ mass generated previously.

After this initial sampling of the 4 masses, a fine random variation of the generated values is performed, within the resolution of the $P(m_{W^+}, m_t)$ and $P(m_{W^-}, m_{\bar{t}})$ p.d.f.s ($1\,\mathrm{GeV} \times 1\,\mathrm{GeV}$). Then, an attempt is made at solving the system of equations (21) and (22). First, using the initial set of masses, and then, the finely

75

variated one. The solution used to reconstruct the event is chosen among all of the solutions found in these two trials such that the quantity (23) is maximized. The fine variation procedure is used to increase the efficiency of the kinematic fit and also the "quality" of the solutions found.

Still, if none of the two trials return a solution, the procedure of generating new mass values is repeated from the initial sampling of $P(m_t, m_{\bar{t}})$, $P(m_{W^+}, m_t)$ and $P(m_{W^-}, m_{\bar{t}})$ p.d.f.s, up to 500 times.

The kinematic fit using truth-matched objects can find at least one solution for 26% of $t\bar{t}H$ events that pass the selection cuts (the truth-matching itself is performed successfully in 36% of the events, and at least a solution is found for 72% of those truth-matched events). Statistical uncertainties in percentages will not be represented whenever they are smaller than 0.5%, which is the case. Overall, a reasonably good reconstruction of the event kinematics is obtained through this method.

Figure 21 shows some p_T distributions of particles reconstructed with truth-match, compared with the corresponding p.d.f.s. This agreement between the distributions at reconstruction level with truth-match and at parton level is only obtained by combining the p.d.f.s and the $1/(p_{T_\nu} \times p_{T_{\bar{\nu}}})$ factors in the same likelihood. Using only either one or the other results in reconstructing neutrinos with systematically higher or lower p_T than at parton level.

In all the distribution histograms shown from this point on, unless otherwise stated, the event content dn/dx (or $dn/(dxdy)$ in 2D histograms) is always given in units of fraction of total number of events per bin. The total number of events is obtained after the application of selection cuts, and requiring at least one solution to the kinematic fit.
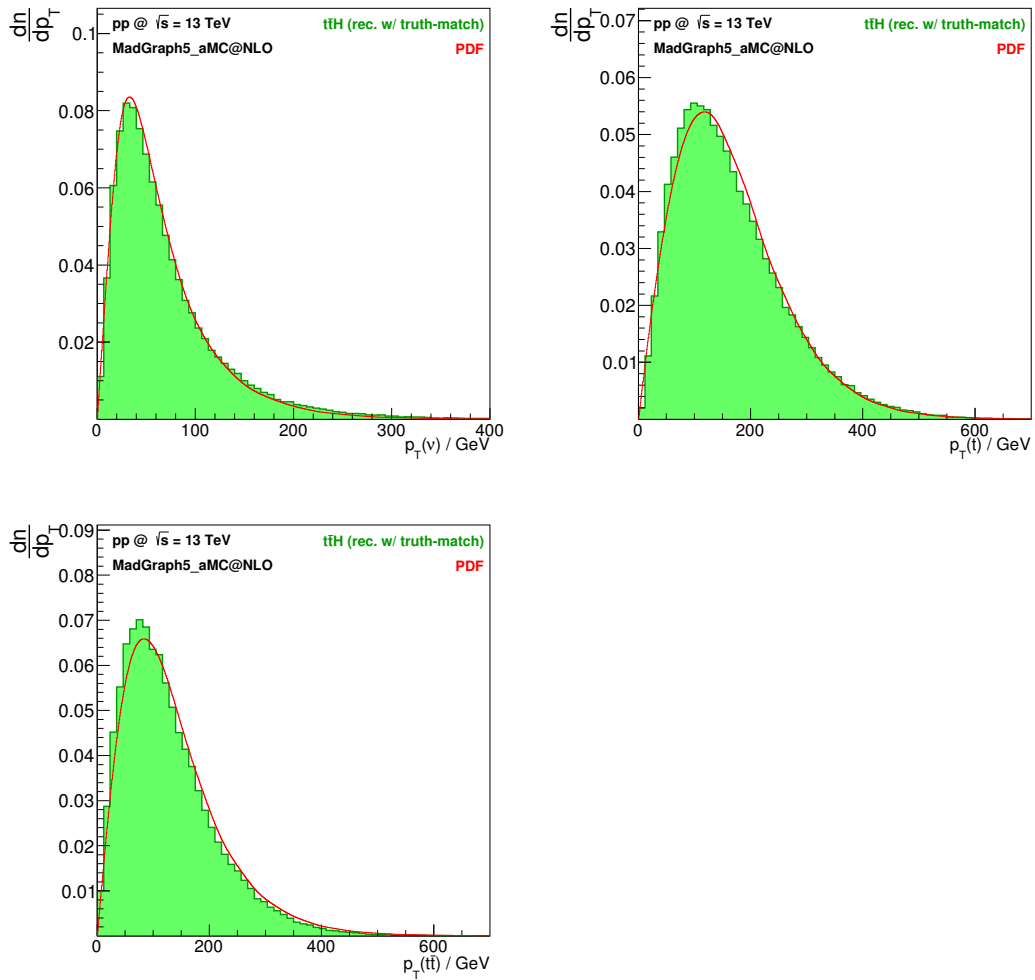
Figure 21: Distributions of the p_T of particles reconstructed with truth-match, compared with the corresponding p.d.f.s used in the fit. There is good agreement, which justifies the use of the $1/(p_{T\,\nu} \times p_{T\,\bar{\nu}})$ factor in the likelihood. Distributions of $p_{T\,\bar{\nu}}$ and $p_{T\,\bar{t}}$ are similar to their conjugates, but are not shown.

In figures 22, 23 and 24, the correlations between kinematic variables of the reconstructed particles and the same variables at parton level are shown. For these figures to be correctly interpreted, it should be noted that, while a kinematic fit procedure of this kind is expected to result in the recovery of the parton-level distributions, it is not expected to exactly recover the parton-level kinematics of each event. This is the case because the leptons, jets, and $\not{E}_T$ used for the fit are already affected by the detector resolution and inefficiencies, as well as by the limitations of the jet clustering algorithm employed. Furthermore, the signal events were generated at next-to-leading order in QCD, while the kinematic fit has a leading-order design, as it was explained above.
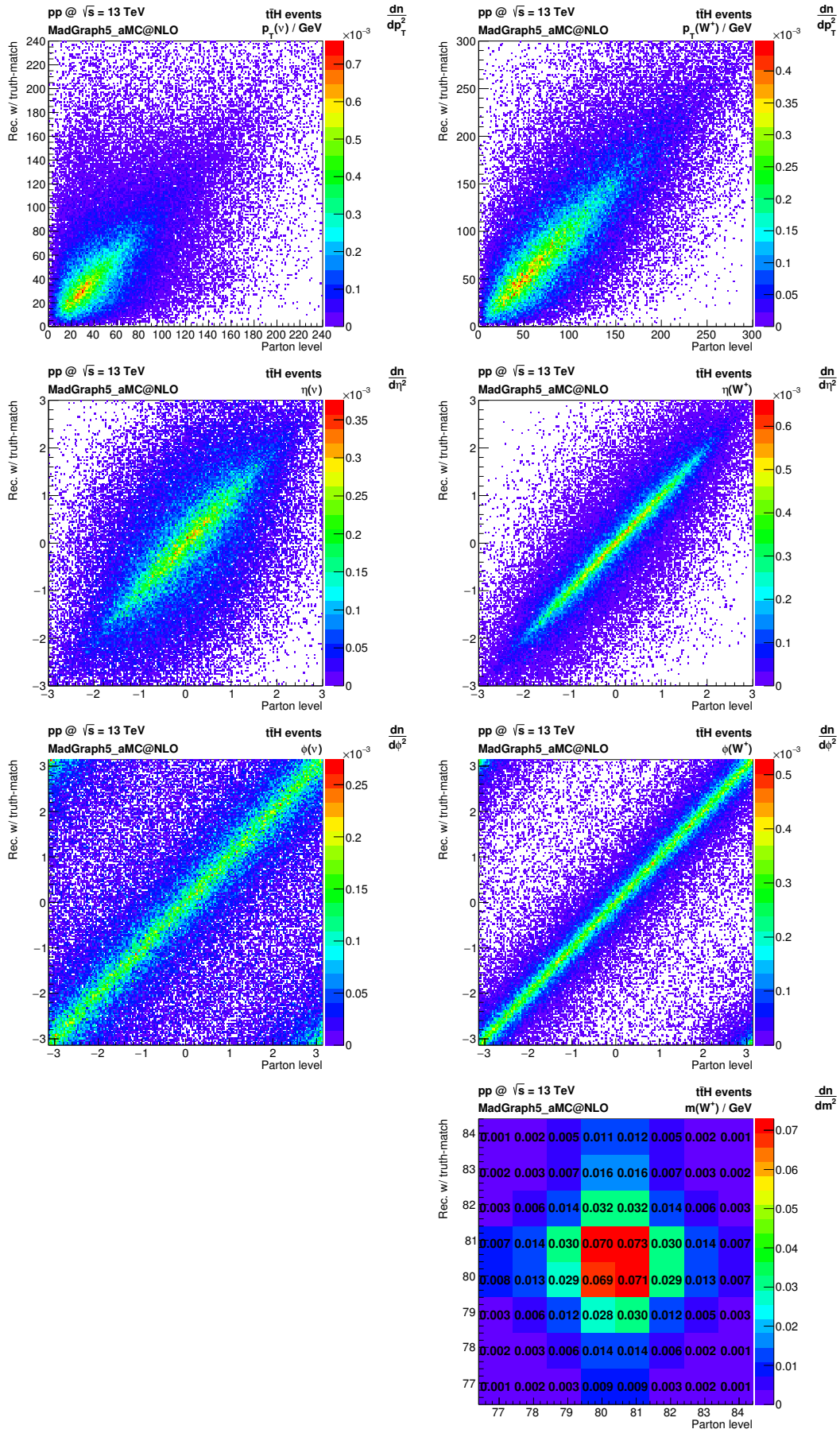
Figure 22: Two-dimensional distributions of kinematic variables in $t\bar{t}H$ events. The horizontal axes represent variables recorded at parton level, and the vertical axes represent the corresponding variables recorded at reconstruction level with truth-match. **Left:** Distributions of p_T, η and ϕ of the neutrino ν. Similar distributions are obtained for the anti-neutrino $\bar{\nu}$, but are not shown. The azimuthal angle ϕ is defined from $-\pi$ to π, which explains the peaks at the lower-right and upper-left corners of ϕ plots. **Right:** The same distributions for the W^+, with the addition of the mass distribution. The resolution in the mass plot is the same as in the p.d.f.s used in the mass trials during the kinematic fit. Similar distributions are obtained for the W^-, but are not shown.

79
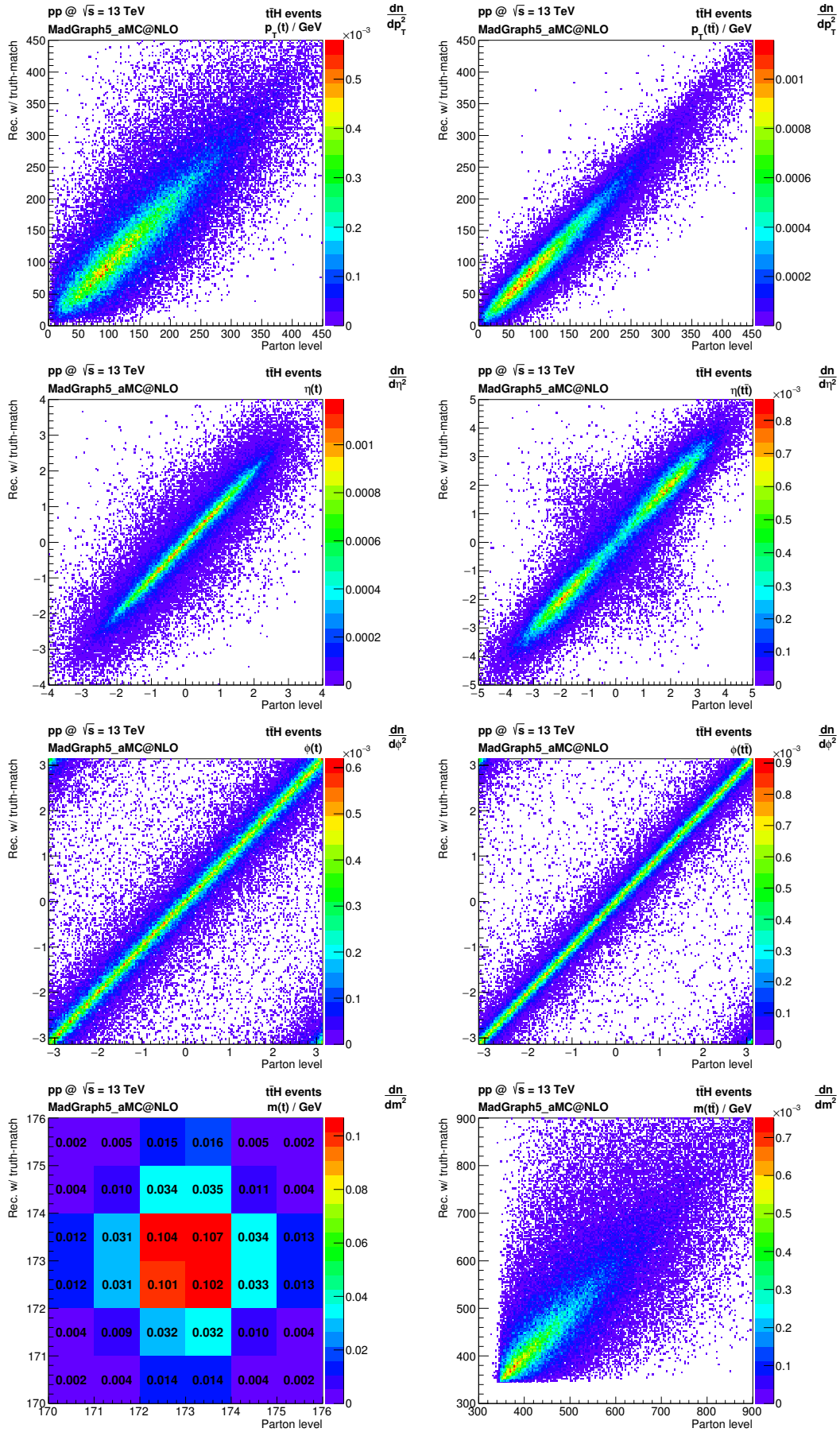
Figure 23: Two-dimensional distributions of kinematic variables in $t\bar{t}H$ events. The horizontal axes represent variables recorded at parton level, and the vertical axes represent the corresponding variables recorded at reconstruction level with truth-match. **Left:** Distributions of p_T, η, ϕ and mass of t. The resolution in the mass plot is the same as in the p.d.f.s used in the mass trials during the kinematic fit. Similar distributions are obtained for $\bar{t}$, but are not shown. **Right:** The same distributions for the $t\bar{t}$ system.
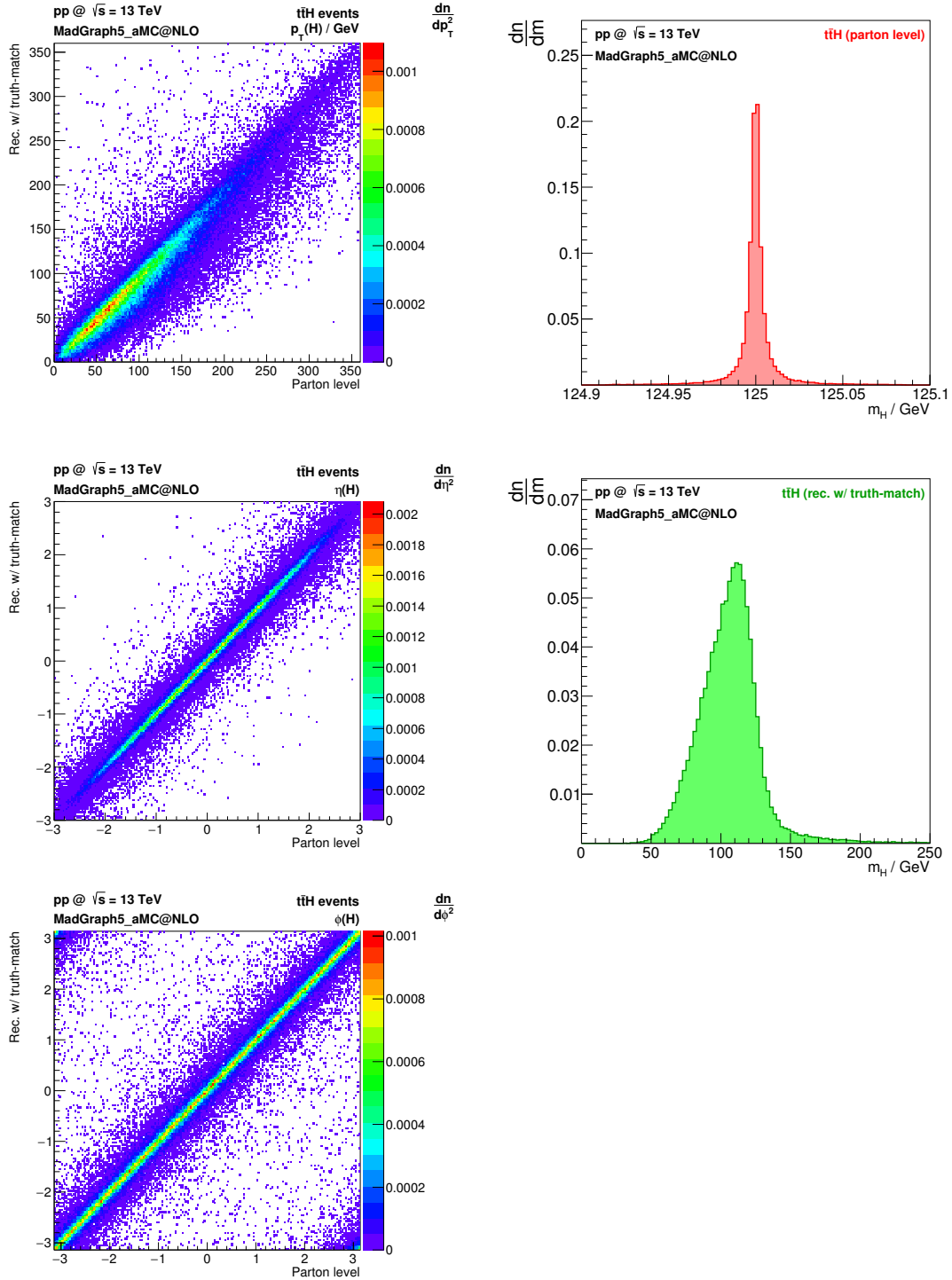
Figure 24: **Left:** Two-dimensional distributions, in $t\bar{t}H$ events, of p_T, η and ϕ of the Higgs boson H. The horizontal axes represent variables recorded at parton level, and the vertical axes represent the corresponding variables recorded at reconstruction level with truth-match. **Right:** Mass distributions of the Higgs boson H, in $t\bar{t}H$ events. In the upper plot, the parton-level mass. In the lower plot, invariant mass distribution of the $b\bar{b}$ system truth-matched to the H decay. The fact that $m_{b\bar{b}}$ is not nicely peaked around 125 GeV as in the $H \to \gamma\gamma$ decay is due to the relatively worse resolution in the energy of jets, compared to that of the energy of photons.

A general conclusion that can be drawn from all the 2D distributions is that the particle direction variables (η and ϕ) seem to be robust with respect to detector effects and NLO radiation. The reconstructed p_T, on the other hand, shows large deviations from its parton-level value. This better correlation in direction variables is in part artificially introduced by the requirement that truth-matched objects lie within a cone of $\Delta R < 0.5$ relative to the parton-level particle.

It is also evident that reconstructed particles are more alike the ones at parton level as we move backwards in the decay chain. The reconstructed neutrinos must compensate for all the energy losses and resolution effects in the event, and thus they are significantly different from the parton-level neutrinos. However, when the reconstructed neutrino is combined with the reconstructed lepton, the resulting W boson is not very different from the one at parton level. The similarity is even more pronounced in the top quark, obtained from combining the W boson with the jet matched to the corresponding b-quark. Finally, the reconstructed $t\bar{t}$ system is remarkably similar to the one at parton level.

The mass distributions of the top quarks and the W bosons within the allowed intervals are preserved, even though there is no correlation between the parton-level masses and the reconstructed ones. This happens because variations of these masses are allowed within a small interval ($6\,\text{GeV}$ for the top quarks and $8\,\text{GeV}$ for the W bosons) compared to the resolution in reconstructed energies of jets. The main concern is to preserve the mass distributions and the correlations between them at each level, and not so much the correlation between reconstructed and parton-level masses.

The reconstructed Higgs boson is very similar to the parton-level one with respect to the direction variables. The reconstructed p_T, however, shows an important asymmetric smearing, such that the Higgs boson is most of the times reconstructed with p_T lower than at parton level. The reconstructed mass also shows a large and asymmetric smearing. Large smearing in mass is expected since the natural Higgs boson width is very small (few MeV), and the energy resolution of jets is large ($1-10\,\text{GeV}$). Asymmetries in both the p_T and mass of the Higgs boson can be explained by energy lost from the jets matched to the b and $\bar{b}$ from the Higgs boson decay. Energy can be lost in out-of-cone losses, in which the b-quark radiates an energetic gluon, for example, that falls outside the cone

defined by the jet algorithm, or in decays of heavy particles in the jet that result in neutrinos or muons, and thus energy from the jet that is not deposited in the calorimeter.

3.4.4 Kinematic fit without truth-match

The previous section summarizes the reconstruction procedure with truth-match. However, when reconstructing without truth-match, additional choices must be made. It is neccessary to identify each of the reconstructed jets as originating from the b or $\bar{b}$ produced in the t and $\bar{t}$ decay, respectively, or from the b or $\bar{b}$ produced in the decay of the H (or $b\bar{b}$ system, in background events).

There is not an upper limit to the number of reconstructed jets in each event. In events with many jets, the corresponding number of jet-assignment possibilities can become overwhelming, increasing the probability of choosing one of the many wrong assignments. Each unique possibility is determined by the jet assigned to the b-quark from the t decay, the jet assigned to the $\bar{b}$ from $\bar{t}$ decay, and the two jets assigned to the H decay (or $b\bar{b}$ system, in the background). The result of exchanging the jets assigned to the b and $\bar{b}$ from the H decay is not considered a different assignment.

In order to limit the number of possibilities for the kinematic fit, a 'good' jet selection is applied, and only good jets can be assigned to the b-quarks from the $t\bar{t}H$ or $t\bar{t}b\bar{b}$ final state. A jet is considered 'good' if it satisfies $p_T > 20\,\text{GeV}$ and $|\eta| < 2.5$. It should be stressed that the selection cuts presented in 3.4.1 already include the requirement of at least 4 good jets in each event. This additional requirement ensures that any jet, besides these 4, can only be assigned to a quark from the decay of $t\bar{t}H$ or $t\bar{t}b\bar{b}$ if it also is a good jet.

Jets originated from the b-quarks of the $t\bar{t}H$ or $t\bar{t}b\bar{b}$ final state are expected to be some of the most energetic ones in each event. With this motivation, good jets ('good' as in the definition given above) were ordered by p_T, and the upper limit on the number of good jets to be used in the reconstruction without truth-match was set to 6. Thus, in events with 4, 5 or 6 good jets, all jet-assignment possibilities are considered. In events with more than 6 good jets, only the 180 possibilities using the 6 good jets with highest p_T are considered.

Table 8: Fraction of $t\bar{t}H$ events, after the requirement of successful truth-match, in which the n highest-p_T jets contained the 4 jets truth-matched to the b and $\bar{b}$ quarks, for $n = 4, 5, 6$. Also shown is the number of assignment possibilities using n jets.

n	n p_T-leading jets contain $b_t, \bar{b}_{\bar{t}}, b_H, \bar{b}_H$	Assignment possibilities
4	0.53	12
5	0.84	60
6	0.96	180

This choice of the 6-jet cutoff resulted from studying the $t\bar{t}H$ event sample. After requiring a successful truth-match of the events (i.e., 4 jets were truth-matched), the fraction of events in which the n highest-p_T jets contained the 4 jets truth-matched to the b and $\bar{b}$ quarks in the final state of $t\bar{t}H$ was calculated, for $n = 4, 5, 6$. The results are presented in table 8, together with the number of assignment possibilities using n jets. This study does not take into account whether the n jets are good or not ('good' having the same meaning as before).

In order to limit the number of meaningful possibilities for the assignment of jets to b and $\bar{b}$ quarks, information about the charge of the jets would be useful. From the experimental point of view, the charge of a jet is not an immediately available quantity. There are methods that can be used to estimate it, using a p_T-weighted average of the charges of the jet constituents [90]. However, their current ability to discriminate between a positively charged and a negatively charged jet is limited, and they will not be addressed here.

The use of b-tagging could also limit the number of jet-assignment possibilities, by excluding light jets originated from QCD radiation or from some other process taking place in the same event besides $t\bar{t}H$ (or $t\bar{t}b\bar{b}$). Since the b-tagging efficiency is also limited, and there is a non-negligible possibility of b-tagging light jets, such a criterion was not used here to decide the assignment of jets.

With the intent of optimizing the criteria used to choose among jet-assignment possibilities, the kinematics of truth-matched objects in signal ($t\bar{t}H$) events were studied. The focus was on quantities obtained directly from jets and leptons, such that these quantities could be used prior to the kinematic fit when reconstructing without truth-match. The studied variables included mostly ΔR, invariant masses
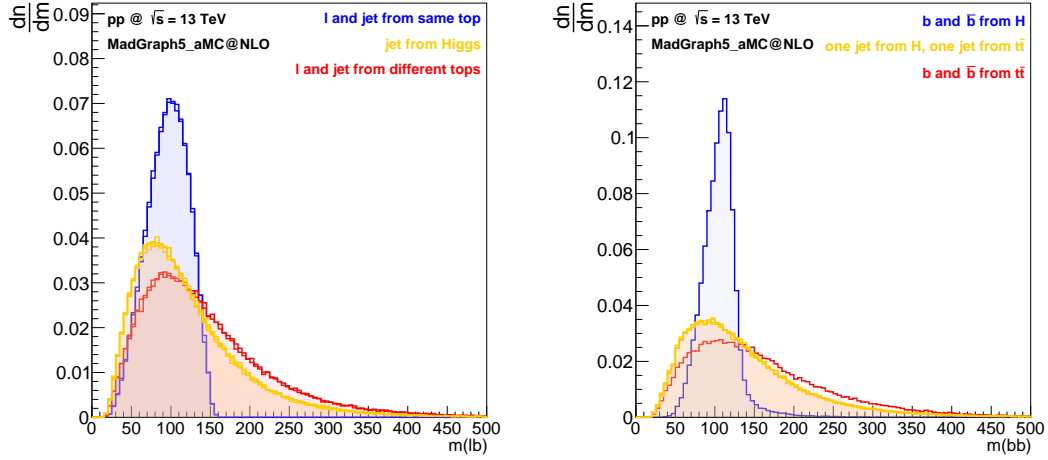
Figure 25: **Left:** Distribution of $m_{\ell b}$ for ℓb pairs originating from the same t or $\bar{t}$ decay (blue), for ℓb pairs in which the lepton and the jet come from different top quarks (red), and for ℓb pairs in which the jet comes from the Higgs boson decay (orange). **Right:** Distribution of m_{bb} for the two jets resulting from the H decay (blue), for the two jets resulting from the decay of the $t\bar{t}$ system (red), and for pairs of jets in which one jet comes from H and the other from $t\bar{t}$ (orange).

and angles between pairs of objects, where each object could either be a lepton or a jet. A few quantities proved useful at discriminating between correct and wrong assignments of jets.

The invariant mass of the system composed of a lepton and a b-jet, $m_{\ell b}$, exhibits a very different behaviour whether the lepton and the jet originate from the same t or $\bar{t}$ decay or not, as can be seen on the left side of figure 25. If they indeed originate from the same decay, $m_{\ell b}$ is very unlikely to be larger than $150\,\mathrm{GeV}$. Otherwise, there is a significant probability for $m_{\ell b} > 150\,\mathrm{GeV}$.

On the right side of figure 25, the invariant mass of two jets, m_{bb}, shows a similar behaviour, with most events at $m_{bb} < 150\,\mathrm{GeV}$ when the two jets originate from the H decay, and a more significant higher-mass tail for two jets originating from different intermediate particles.

Another variable to consider could be the angle $\theta_{\ell b}$, expected to be small if the lepton and jet originate from the same top quark decay. This angle should depend on the energy of the decaying top quark. If the decaying t or $\bar{t}$ is very energetic (momentum much larger than m_t), its decay products will be very collimated and $\theta_{\ell b}$ will be small. Otherwise, if the t or $\bar{t}$ energy is of order m_t or smaller, the allowed range for $\theta_{\ell b}$ should be larger. An estimation of the t or $\bar{t}$ momentum can

85
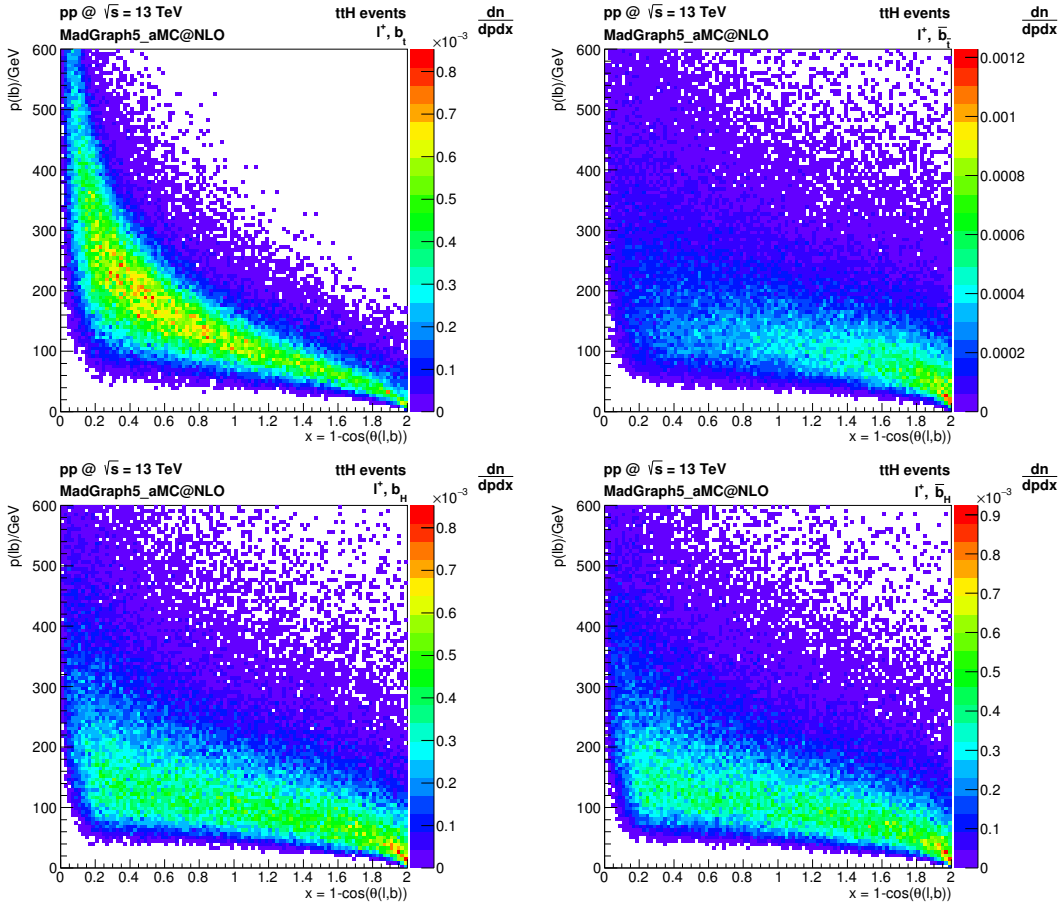
Figure 26: 2D distributions of $t\bar{t}H$ events with $1 - \cos(\theta_{\ell b})$ in the horizontal axis and $p_{\ell b}$ in the vertical axis. **Upper left:** For the correct pairing ℓ^+, b_t, where a strong correlation can be seen. **Upper right:** For $\ell^+, \bar{b}_{\bar{t}}$, where the event density is more uniform. **Bottom:** For ℓ^+ and each of the b-quarks from the H decay. Similar distributions are obtained with ℓ^-, but are not shown here.

be obtained from the momentum of the system composed of a lepton and a jet, $p_{\ell b}$. 2D histograms of $t\bar{t}H$ events were created with $p_{\ell b}$ in one axis and $1 - \cos(\theta_{\ell b})$ in the other. These histograms are shown in figure 26. While the correlation between the variables is evident when the lepton and the jet come from the same top quark decay, the density is more uniform when this is not the case.

Combining this information, the following selection is applied to the possibilities of jet-assignment: $m_{\ell^+ b_t} < 150\,\text{GeV}$ and $m_{\ell^- \bar{b}_{\bar{t}}} < 150\,\text{GeV}$, where b_t is the jet assigned to the b-quark from the t decay, and $\bar{b}_{\bar{t}}$ is the jet assigned to the $\bar{b}$-quark from the $\bar{t}$ decay; $m_{b_H \bar{b}_H} < 150\,\text{GeV}$ for the jets assigned to the b and $\bar{b}$ from the H decay.

86

The $\theta_{\ell^+ b_t}$ and $p_{\ell^+ b_t}$, as well as $\theta_{\ell^- \bar{b}_{\bar{t}}}$ and $p_{\ell^- \bar{b}_{\bar{t}}}$, were not used to constrain the possible jet assignments. In the future, this could be done by requiring that the jet assignment being tested verifies the correlation observed in the 2D plot for the truth-matched objects. In order to apply this requirement, two functions of $1-\cos(\theta_{\ell b})$ should be fitted to the upper and lower bounds of the densely populated region in the 2D plot of the correctly assigned ℓb pair. Trial jet assignments would then be required to lie between these two curves.

After applying the 6-jet cutoff and the kinematics selections as explained before, a reduced set of possibilities for the jet assignment is obtained for each event. For each possibility in this reduced set, the kinematic fit is performed. The masses of the W^+, W^-, t and $\bar{t}$ are generated as explained before, with up to 500 trials, each generating two finely different sets of mass values, until a trial returns at least one solution.

Then, all the solutions obtained from all the considered jet assignments are compared. The solution picked (along with the corresponding jet assignment) is the one with the highest value of the likelihood

$$\frac{1}{p_{T\nu} p_{T\bar{\nu}}} P(p_{T\nu}) P(p_{T\bar{\nu}}) P(m_t, m_{\bar{t}}) P(p_{Tt}) P(p_{T\bar{t}}) P(p_{Tt\bar{t}}) P(m_{b_H \bar{b}_H}), \qquad (24)$$

where $m_{b_H \bar{b}_H}$ is the invariant mass of the system composed of the two jets identified as originating from the H decay in that particular combination of jets. $P(m_{b_H \bar{b}_H})$ is obtained from the $m_{b_H \bar{b}_H}$ histogram at reconstruction level with truth-match, by the same smoothing procedure as the p_T p.d.f.s. The p.d.f. used is shown in figure 20, together with the 50-bin distribution from which it was obtained. Also represented is the p.d.f. that would be obtained in case the ΔR cut for truth-matching were set to 0.7, instead of 0.5. From $\Delta R = 0.5$ to $\Delta R = 0.7$, a slight increment of number of events is visible in the low mass region. In both cases, the asymmetric shape of the p.d.f. reflects the distorted distribution of the reconstructed Higgs boson mass, even with truth-match, and is expected to transfer that distortion to the reconstruction level without truth-match.

Because the Higgs width is much smaller than the experimental mass resolution, the parton-level distribution was not used, unlike for all the other p.d.f.s. If it were to be used, it would boost the probability of combinations where the Higgs boson

is reconstructed with an invariant mass close to $125\,\text{GeV}$, and fall abruptly to 0 outside a $0.2\,\text{GeV}$ range around that value. Such a behaviour would exclude most of the correct assignments, as can be seen from the $m_{b_H \bar{b}_H}$ distribution at reconstruction level with truth-match, on the right side of figure 24.

Although the $m_{b_H \bar{b}_H}$ criterion seems very reasonable for signal events, it is important to notice that there is no reason why in background events (such as $t\bar{t}b\bar{b}$) this should result in a correct identification of jets as originating from $t\bar{t}$ system or from the $b\bar{b}$ system. Thus, in background events, this method is expected to possibly result in wrong assignment of jets and to introduce a bias in the invariant mass distribution of the $b\bar{b}$ system, creating a peak similar to the one in $t\bar{t}H$ events, with no physical meaning. This should not be a reason to worry, since the focus of the kinematic fit is to correctly reconstruct the signal. By trying to reconstruct signal-like features in the events (masses of W bosons, top quarks and $m_{b_H \bar{b}_H}$, p_T of neutrinos, etc.) immediately some degree of background rejection will exist. The background events that somehow manage to be reconstructed by this fit must be truly irreducible, and at this point discriminant variables must be found that hopefully allow the extraction of a signal.

The kinematic fit without truth-match finds at least one solution for 78% of the signal events that pass the selection cuts. Figure 27 shows some p_T distributions of particles reconstructed without truth-match, as well as the reconstructed Higgs boson mass, all compared with the corresponding p.d.f.s. In figures 28, 29 and 30, the correlations between kinematic variables of the reconstructed particles without truth-match and the same variables at parton level are shown. In figure 31, 2D distributions of the W^+, t and $\bar{t}$ masses are represented, and it is clear that the mass trial procedure preserves the correlations between them.
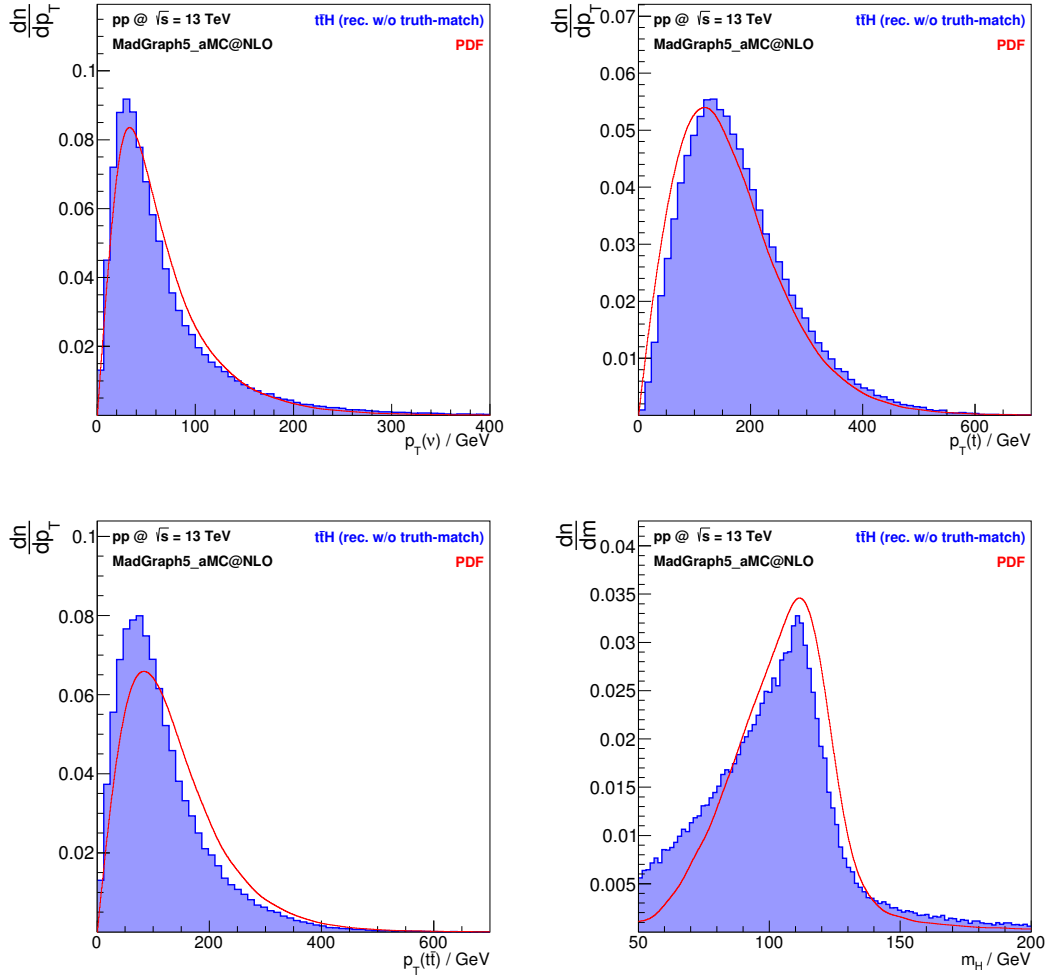
Figure 27: Distributions at reconstruction level without truth-match, compared with the corresponding p.d.f.s. Transverse momentum of the reconstructed particles, and Higgs boson mass. Once again, the kinematic fit results in distributions consistent with the p.d.f.s. The spectra of $p_{T\bar{\nu}}$ and $p_{T\bar{t}}$ are similar to their conjugates, but are not shown.
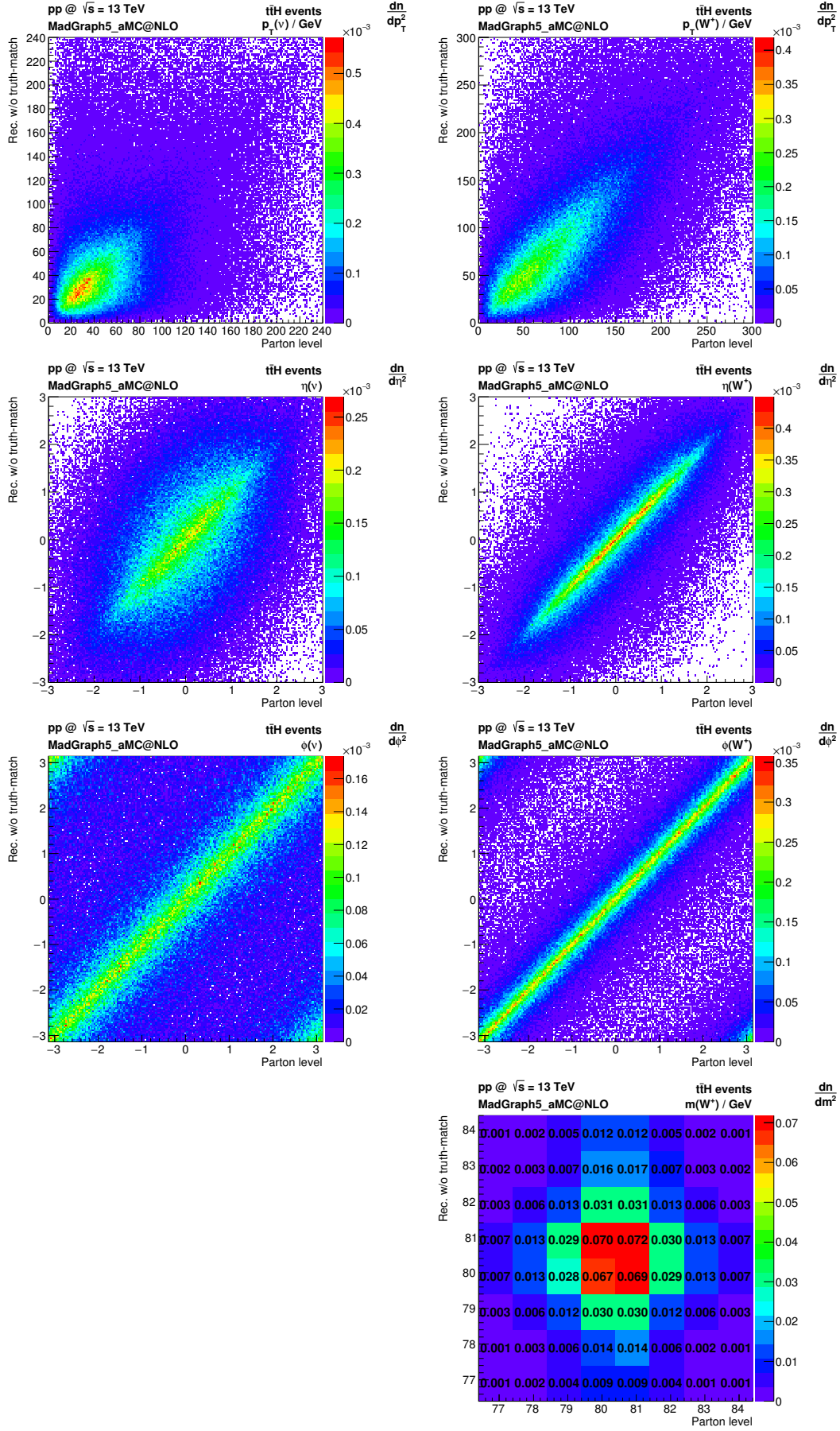
Figure 28: Two-dimensional distributions of kinematic variables in $t\bar{t}H$ events. The horizontal axes represent variables recorded at parton level, and the vertical axes represent the corresponding variables recorded at reconstruction level without truth-match. **Left:** Distributions of p_T, η and ϕ of the neutrino ν. Similar distributions are obtained for the anti-neutrino $\bar{\nu}$, but are not shown. **Right:** The same distributions for the W^+, with the addition of the mass distribution. The resolution in the mass plot is the same as in the p.d.f.s used in the mass trials during the kinematic fit. Similar distributions are obtained for the W^-, but are not shown.

Figure 29: Two-dimensional distributions of kinematic variables in $t\bar{t}H$ events. The horizontal axes represent variables recorded at parton level, and the vertical axes represent the corresponding variables recorded at reconstruction level without truth-match. **Left:** Distributions of p_T, η, ϕ and mass of t. The resolution in the mass plot is the same as in the p.d.f.s used in the mass trials during the kinematic fit. Similar distributions are obtained for $\bar{t}$, but are not shown. **Right:** The same distributions for the $t\bar{t}$ system.
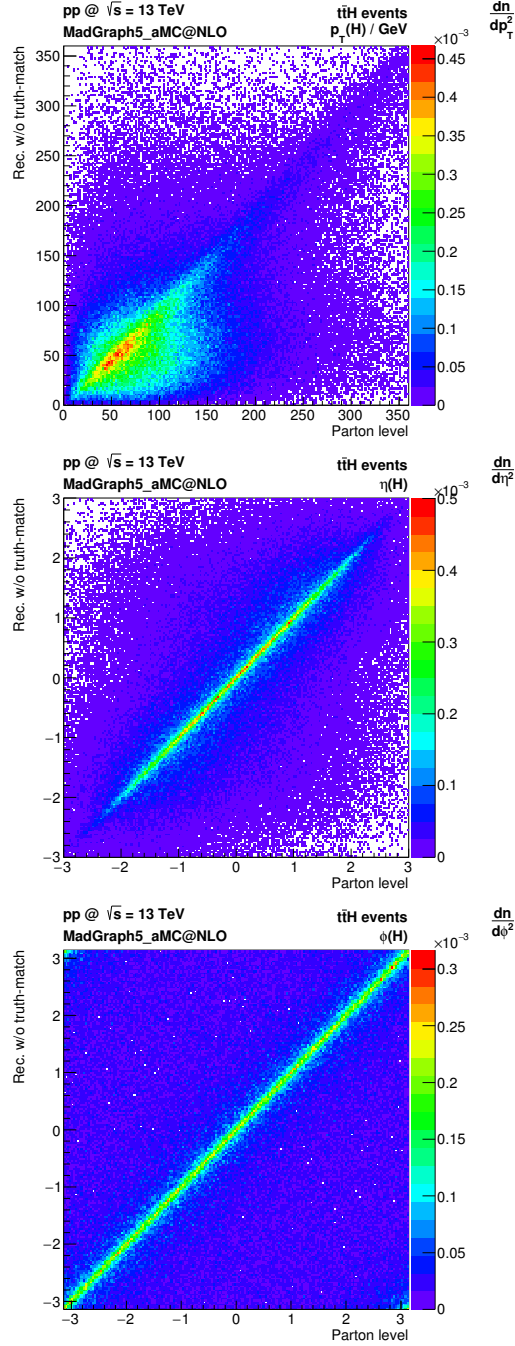
Figure 30: Two-dimensional distributions, in $t\bar{t}H$ events, of p_T, η and ϕ of the Higgs boson H. The horizontal axes represent variables recorded at parton level, and the vertical axes represent the corresponding variables recorded at reconstruction level without truth-match.
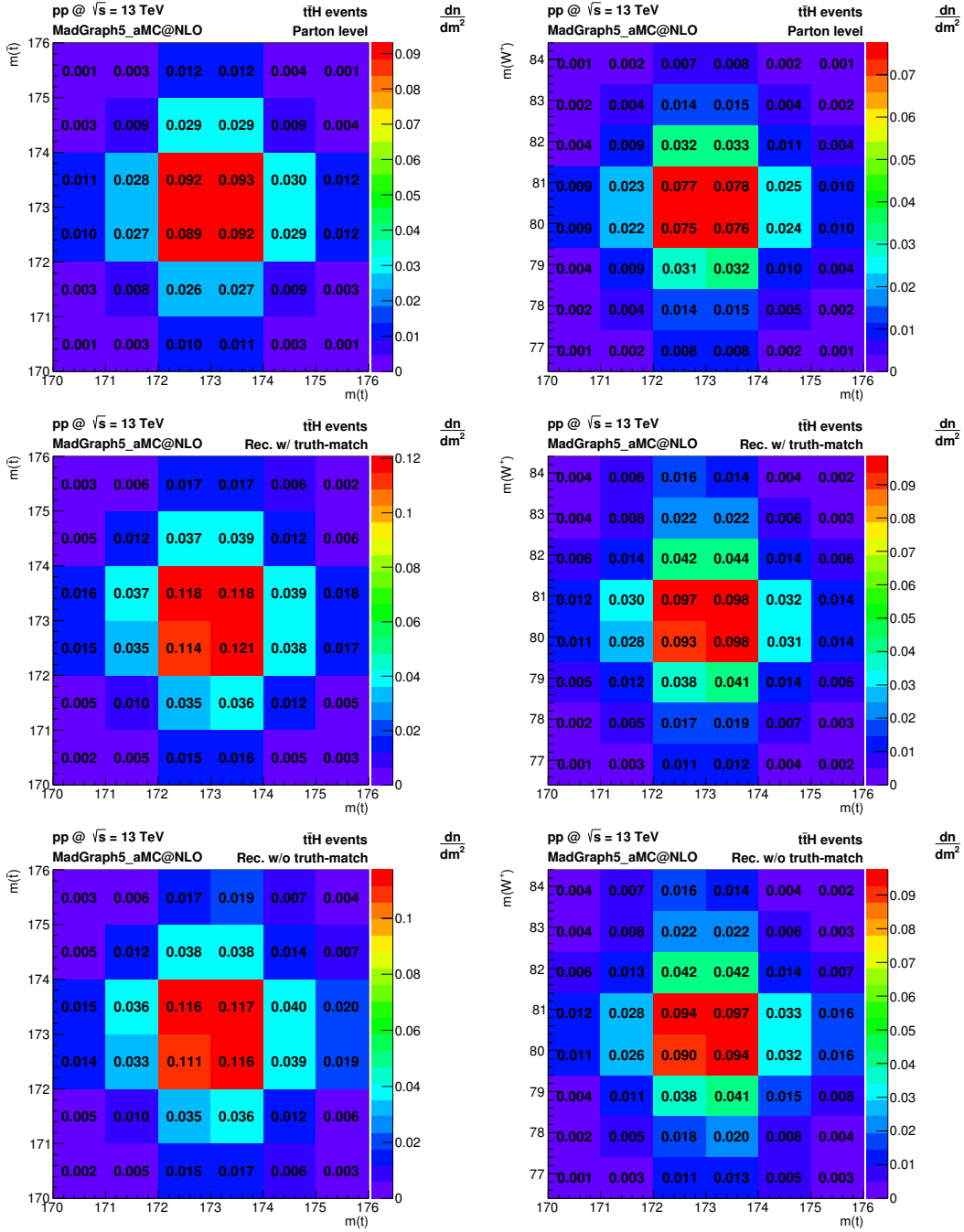
Figure 31: Two-dimensional distributions, in $t\bar{t}H$ events, of the t, $\bar{t}$ and W^+ masses, showing that mass correlations are respected in reconstructed events. **Left:** The horizontal axes represent the t mass and the vertical axes represent the $\bar{t}$ mass. The upper plot is the parton level distribution. It is the first 2D p.d.f. used in the mass trials of the kinematic fit. The middle (bottom) plot is the same distribution at reconstruction level with (without) truth-match. **Right:** The horizontal axes represent the t mass and the vertical axes represent the W^+ mass. The upper plot is the parton level distribution. It is the second 2D p.d.f. used in the mass trials of the kinematic fit. The middle (bottom) plot is the same distribution at reconstruction level with (without) truth-match. Similar distributions can be obtained for $\bar{t}$ and W^-, but those are not shown here.

At this level, the same general conclusions that were drawn at the reconstruction with truth-match are valid. Direction variables are reconstructed more closely to their parton-level values than p_T. Again, reconstructed particles in the earlier steps of the decay chain are more similar to parton-level particles.

In general, without truth-match, the reconstructed particle directions and p_T are less correlated to the ones at parton level, than when the reconstruction is done with truth-match. This is expected for two reasons. First, there is the effect of applying the $\Delta R < 0.5$ cut when truth-matching the particles, which immediately led to large correlation in direction variables. Then, there is the effect of combinatorial background, which is particularly evident in distributions of ϕ and in the Higgs boson p_T. In these distributions, the superposition of two populations can be observed: the correctly assigned events populate the band around $x = y$, while the combinatorial background is more uniformly distributed.

As expected, the reconstructed Higgs boson mass recovers the mass distribution obtained with truth-match, since it was used as a p.d.f. to weight the likelihood of the jet assignments.

In order to characterize the overall ability of the routine to correctly assign the jets, a migration matrix M was filled and is shown in table 9. M_{ij} is the number of $t\bar{t}H$ events with 4 truth-matched jets in which the jet truth-matched to a given quark i is assigned to the quark j. M_{ij} is normalized to the total number of events with 4 truth-matched jets, whether they have kinematic fit solutions or not. The numbers across a line or column do not sum to 1 because jets which were not truth-matched to any quark at all can be used in the reconstruction without truth-match, as long as they are good ($p_T > 20 \, \mathrm{GeV}$ and $|\eta| < 2.5$) and within the 6 p_T leading good jets. Also, it may happen that the truth-matched jets do not pass the good jet criteria or the 6-jet cutoff, thus being rejected at reconstruction level without truth-match.

Besides this migration matrix, a fraction was calculated that should be interpreted as follows: given a successfully truth-matched event returning at least one solution to the kinematic fit with truth-match, the kinematic fit without truth-match will return the same jet assignment as the truth-match procedure 33% of the time. It is important to notice once again that the exchange of the b and $\bar{b}$ from the Higgs boson decay is not considered to result in a different assignment.

94

Table 9: Migration matrix characterizing the ability of the reconstruction routine to correctly assign jets to the quarks in $t\bar{t}H$ events. Values are fractions of the total number of events with 4 truth-matched jets, regardless of whether they have kinematic fit solutions or not. Statistical uncertainties in these fractions are below 0.002.

		Assigned			
		b_t	$\bar{b}_{\bar{t}}$	b_H	$\bar{b}_H$
	b_t	0.53	0.10	0.13	0.15
Truth-matched	$\bar{b}_{\bar{t}}$	0.12	0.52	0.13	0.15
	b_H	0.16	0.16	0.28	0.30
	$\bar{b}_H$	0.16	0.16	0.28	0.30

If this fraction were to be computed using the full set of successfully truth-matched events, whether they have kinematic fit solutions or not, the value would be expected to drop. This happens because, if no solutions are found in the kinematic fit with the truth-match jet assignment, the kinematic fit without truth-match should not find a solution for that same jet assignment. There is a minor chance that it could happen because of the random procedure in the mass trials. However, since the number of trials is 500 when no solutions are found, the sampling of the phase-space is quite exhaustive, and the probability for a solution to appear after 500 unsuccessful mass trials is small.

3.5 Angular observables

In this work, several angular observables were tested for their potential as discriminating variables. The aim is twofold: to identify variables that distinguish between $t\bar{t}H$ and the irreducible background $t\bar{t}b\bar{b}$ and variables to discriminate between different parity scenarios of the Higgs boson in $t\bar{t}H$ production (scalar or pseudoscalar). For every variable, Kolmogorov-Smirnov (K-S) tests were run to check the statistical compatibility of the distributions of $t\bar{t}b\bar{b}$ and $t\bar{t}H$ (SM Higgs boson) events and the compatibility of $t\bar{t}H$ (SM Higgs boson) and $t\bar{t}A$ (pure pseudoscalar Higgs boson). In the present section, the tested observables are defined, and the motivation for this definition is given, within the scope of the helicity formalism applied to successive particle decays.

3.5.1 Helicity formalism in successive particle decays

The theoretical argument for the use of angular observables to increase the signal sensitivity in the search for $t\bar{t}H$ production rests largely on the fact that, in a particle decay, the probability of the decay products travelling in a certain direction is constrained by the spin states of the decaying particle and of the decay products. The detail of these constraints is illustrated below in the context of the helicity formalism [91].

The helicity operator is $h = \vec{S} \cdot \hat{p}$, that is, it is the projection of the spin operator along the direction of momentum of a particle. It is a convenient operator because it commutes with the total angular momentum operator J and with the momentum operator p. This ensures that particle states can be built which are simultaneously eigenstates of helicity and linear momentum or eigenstates of helicity and angular momentum.

Let us denote a two-particle eigenstate of helicity and linear momentum as $|\theta\varphi\lambda_1\lambda_2\rangle$, where the two particles have helicities λ_1 and λ_2 and, in the centre-of-mass of the two-particle system, particle 1 has momentum $\vec{p}$ with spherical components (p, θ, φ), with respect to a given z-axis (particle 2 must have momentum $-\vec{p}$).

Then, it is interesting to compute the quantity $\langle\theta\varphi\lambda_1\lambda_2|U|JM\rangle$, which is the amplitude for a state with total angular momentum J and angular momentum projection M on the z-axis to decay, through some interaction U, into a two-particle linear momentum and helicity eigenstate, where the angles θ and φ are defined with respect to the same z-axis.

To compute this quantity, we expand it in the complete basis of eigenstates of angular momentum with helicities λ_1 and λ_2, which yields

$$\langle\theta\varphi\lambda_1\lambda_2|U|JM\rangle = \sum_{jm}\langle\theta\varphi\lambda_1\lambda_2|jm\lambda_1\lambda_2\rangle\langle jm\lambda_1\lambda_2|U|JM\rangle. \qquad (25)$$

The fact that angular momentum is conserved in the process implies that the only non-vanishing component is the one with $j = J$ and $m = M$. This leaves

$$\langle\theta\varphi\lambda_1\lambda_2|U|JM\rangle = \langle\theta\varphi\lambda_1\lambda_2|JM\lambda_1\lambda_2\rangle\langle JM\lambda_1\lambda_2|U|JM\rangle. \qquad (26)$$

The matrix element $\langle JM\lambda_1\lambda_2|U|JM\rangle$ is rotationally invariant and will be denoted $A_{\lambda_1\lambda_2}$.

The momentum eigenstate in the first bra of the expression can be obtained from applying a rotation operator[1] $R(\varphi, \theta, -\varphi)$ to the momentum eigenstate with particle 1 travelling in the direction of the z-axis. The quantity can then be written as

$$\langle\theta\varphi\lambda_1\lambda_2|U|JM\rangle = \langle\theta=0, \varphi=0, \lambda_1\lambda_2|R^\dagger(\varphi, \theta, -\varphi)|JM\lambda_1\lambda_2\rangle A_{\lambda_1\lambda_2}. \qquad (27)$$

In general, the $|\theta\varphi\lambda_1\lambda_2\rangle$ states do not have definite values of J or M. However, the state $|\theta=0, \varphi=0, \lambda_1\lambda_2\rangle$ is unique because the particles are travelling along z, the quantization axis of M. Since the orbital angular momentum is $\vec{L} = \vec{r} \times \vec{p}$, its projection along z vanishes. Thus, the total angular momentum projection M is simply the spin projection, and this particular state is an eigenstate of M, with eigenvalue $\lambda = \lambda_1 - \lambda_2$. It can be expanded in the eigenstates of total angular momentum with angular momentum projection λ:

$$\langle\theta\varphi\lambda_1\lambda_2|U|JM\rangle = \sum_{J'} c_{J'\lambda}\langle J'\lambda|R^\dagger(\varphi, \theta, -\varphi)|JM\lambda_1\lambda_2\rangle A_{\lambda_1\lambda_2}, \qquad (28)$$

for some coefficients $c_{J'\lambda}$.

Because the rotation operator preserves the total angular momentum, the only non-vanishing term is the one with $J = J'$. The result of applying a rotation to an eigenstate of total angular momentum and angular momentum projection can be written in terms of Wigner D-matrices, yielding

$$\begin{aligned}\langle\theta\varphi\lambda_1\lambda_2|U|JM\rangle &= c_{J\lambda}\langle J\lambda|R^\dagger(\varphi, \theta, -\varphi)|JM\lambda_1\lambda_2\rangle A_{\lambda_1\lambda_2}\\ &= c_{J\lambda}\sum_{M'} D^J_{M',\lambda}{}^\dagger(\varphi, \theta, -\varphi)\langle JM'|JM\lambda_1\lambda_2\rangle A_{\lambda_1\lambda_2}\\ &= c_{J\lambda}D^J_{M,\lambda}{}^\dagger(\varphi, \theta, -\varphi)A_{\lambda_1\lambda_2}.\end{aligned} \qquad (29)$$

This result can be used to compute angular distributions in successive particle

---

[1]The arguments of the rotation operator are the conventional Euler angles.

Table 10: Elements of the d-matrices $d^J_{M,\lambda}(\theta)$, for $J = 0, {}^1/_2, 1$ [19]. All the missing elements can be obtained from the relations $d^J_{M,\lambda}(\theta) = (-1)^{M-\lambda}d^J_{\lambda,M} = d^J_{-\lambda,-M}$.

$J = 0$	$J = {}^1/_2$		$J = 1$	
$d^0_{0,0} = \quad 1$	$d^{1/2}_{1/2,1/2} =$	$\cos(\frac{\theta}{2})$	$d^1_{1,1} =$	$\frac{1+\cos\theta}{2}$
			$d^1_{1,0} =$	$-\frac{\sin\theta}{\sqrt{2}}$
	$d^{1/2}_{1/2,-1/2} =$	$-\sin(\frac{\theta}{2})$	$d^1_{1,-1} =$	$\frac{1-\cos\theta}{2}$
			$d^1_{0,0} =$	$\cos\theta$

decays. The Wigner D-matrices are given in terms of the (small) d-matrices as

$$D^J_{M,\lambda}(\alpha, \theta, \gamma) = e^{-i\alpha M}d^J_{M,\lambda}(\theta)e^{-i\gamma\lambda}. \tag{30}$$

The elements of the d-matrices are shown in table 10 for $J = 0, {}^1/_2, 1$, which are the only necessary ones in treating decays of fundamental particles [19]. Not all the elements are listed, but all the missing ones can be obtained from the relations $d^J_{M,\lambda}(\theta) = (-1)^{M-\lambda}d^J_{\lambda,M} = d^J_{-\lambda,-M}$.

To compute the decay amplitude of a particle which in its turn resulted from a previous decay, the rest frame of the particle should be considered, with the z-axis defined as the momentum direction of the particle in the rest frame of the previously decaying particle. This ensures that the particle is simultaneously in an eigenstate of total angular momentum (equal to its spin) and angular momentum projection along z (equal to its helicity, as measured in the rest frame of the previous particle).

For every decay in an event, the corresponding θ angle dependence will appear in the amplitude. This is the angle between the momentum direction of the decaying particle, in the rest frame of the particle that decayed into it (or an arbitrary axis in the case of the first particle in the decay chain), and the direction of momentum of one of the decay products in the rest frame of the decaying particle. Which of the two final particles is chosen to compute the angle is conventional since, in the decaying particle rest frame, they travel back-to-back.

The overall amplitude of a decay chain is obtained from the product of the
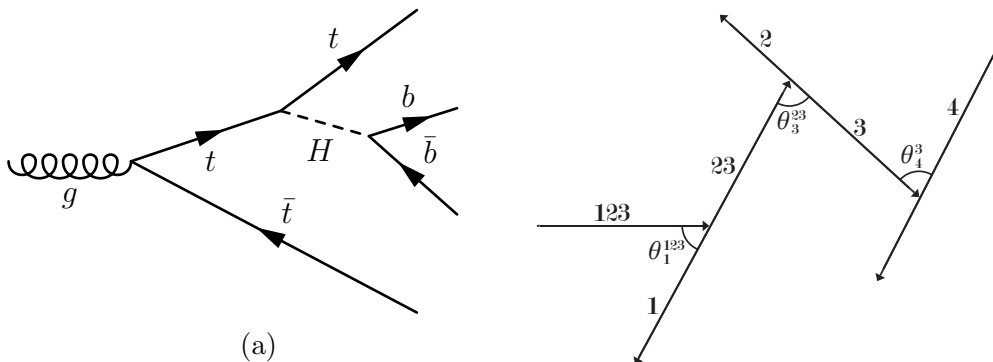
Figure 32: **Left:** $t\bar{t}H$ production through an s-channel gluon, which splits into a $\bar{t}$ and a t. In its turn, the t radiates a Higgs boson, which decays into $b\bar{b}$. **Right:** General decay chain with particles labeled 1, 2, 3 and 4.

amplitudes of the individual decays, where a sum must be made over all possible values of the helicities of the intermediate particles. Computing the square of the absolute value of the amplitude results in a cross-section, with multiple angular dependence.

In an experiment such as the ones at the LHC, the final state helicities are not measurable. In order to obtain an angular distribution comparable with experiment, it is necessary to sum the cross-sections over all possible helicities of final state particles.

3.5.2 Angles of the type θ_Y^X and derived observables

In the $t\bar{t}H$ production process shown on the left side of figure 32, three successive decays take place: the s-channel gluon decays into the $t\bar{t}$ system and then the t radiates a Higgs boson (in the case of the $t\bar{t}b\bar{b}$ background, similar diagrams are possible, with the radiation of a gluon, a photon or a Z boson). Finally, the Higgs boson decays to $b\bar{b}$. According to the formalism exposed in the previous section, three relevant angles will appear in the overall decay amplitude computation.

The first angle is between the direction of the s-channel gluon (kinematically equivalent to the whole $t\bar{t}H$ system) in the lab frame (arbitrary frame, since it is the first particle in the decay chain) and the direction of one of the top quarks in the rest frame of the $t\bar{t}H$ system. The second angle is between the direction of momentum of t, before it radiates the Higgs boson, in the rest frame of the $t\bar{t}H$ system, and the Higgs boson direction in the rest frame of this t. The third angle

is between the direction of H, in the rest frame of t before it radiates H, and the direction of b or $\bar{b}$ in the rest frame of H.

The decay chains of the final t and $\bar{t}$ also introduce additional angular dependences of potential interest, computed in a similar way.

The notation used from now on for angles of this type is θ_Y^X, meaning "the angle between the direction of particle Y in the rest frame of X and the direction of particle X". If X is the $t\bar{t}H$ system, its direction will always be calculated in the lab frame. For other particles, the rest frame in which the direction of X is considered will always be made clear in the context in which the angle appears.

Many of the angular observables distributions studied follow a single form, which can be easily related to the angles that appear in the computation of the decay amplitude as explained before. Considering a set of 4 particles labeled from 1 to 4, as on the right side of figure 32, two large families of functions of angles were addressed:

$$f(\theta_1^{123})g(\theta_4^3) \text{ and } f(\theta_3^{23})g(\theta_4^3).$$

Functions f and g are either sine or cosine. Particles 1 to 3 can either be t, $\bar{t}$ or H, without repetition. Particle 4 can be any of the products of the decay of the top quarks and the Higgs boson, including the intermediate W bosons. Labels with more than one particle symbol means the system composed by those particles. For example, $\bar{t}H$ means the system composed of the Higgs boson and the anti-top quark. If indeed the $\bar{t}$ radiated the Higgs boson, then $\bar{t}H$ is kinematically equivalent to the $\bar{t}$ before radiating. Direction of particle 3 was always computed in the rest frame of the system 23 and direction of the system 23 was always computed in the rest frame of system 123.

Some ambiguity exists in boosting particles from one rest frame to another. As an example, supposing θ_4^3 is to be computed, particle 4 must be boosted to the rest frame of particle 3. One possibility is for this to be done directly: the four-momentum of particle 3 is computed in the lab frame, and particle 4 is boosted from the lab frame according to this four-momentum of particle 3. The other possibility is to do it sequentially: particle 4 is boosted to the 123 frame according to the four-momentum of 123 in the lab frame, then it is boosted again to the 23 frame according to the 23 four-momentum in the 123 rest frame, then finally it is boosted

to the 3 rest frame according to the four-momentum of particle 3 in the frame of 23. The directions of particle 4 resulting from each of these different boosting procedures are also different, due to the so-called Wigner rotations, that result from successively applying Lorentz boosts in non-collinear directions [92]. All the new angular observables shown in this work were defined using the sequential prescription, which respects the reasoning from the helicity formalism. However, angular observables obtained with the direct prescription were also studied. The differences in the distributions obtained are not significant, but any future study of observables of this kind must take this choice into consideration.

With respect to the definition of new angular observables that can be tested as discriminants in the future, the possibilities are countless. Since the motivation for the construction of the observables studied in this work was a chain of 3 successive decays, one straightforward generalization would be to compute products of 3 functions of angles $(f(\theta_1^{123})g(\theta_3^{23})h(\theta_4^3))$. Some of these products of three functions were studied and, although interesting distributions are obtained, there are no major differences in the resulting discrimination of samples. The functions of angles themselves could be picked differently, for example as higher modes of oscillation ($\cos(n\theta)$ or $\cos^n\theta$). However, from the Wigner D-matrices definition and because the decaying particles have spins 0, $^1/_2$ or 1, all features in the distributions of the θ_Y^X angles should appear in modes with $n \leq 2$. A more diligent approach would be to compute the angular dependence of the whole decay chains of $t\bar{t}H$, $t\bar{t}b\bar{b}$ and $t\bar{t}A$ processes, using the rules exposed in 3.5.1. Doing such a calculation for the production through top quark fusion (as in figure 8 (d)) would require revisiting the helicity formalism since, besides the decay amplitudes, also a scattering amplitude must be computed. In the end, however, the full dependence of the production cross-sections on each angle θ_Y^X would be obtained, and a more insightful construction of independent angular observables could be done. This exercise falls beyond the scope of the present thesis, but should be addressed in the future.

Histograms of variables from both families were generated, permutating the particles assigned to each number and alternating the use of sine or cosine as the f and g functions. For each observable, two Kolmogorov-Smirnov (K-S) tests of statistical compatibility were performed, in order to assess the discriminating

101

power of the observable. One test was performed between the $t\bar{t}H$ and $t\bar{t}b\bar{b}$ samples, and another test between the $t\bar{t}H$ and $t\bar{t}A$ samples.

The following variant of the Kolmogorov-Smirnov test was employed. The cumulative distribution $F(x = x')$ is defined as the fraction of events in which the variable x is strictly smaller than x'. For each observable, the test compares the cumulative distributions $F_A(x)$ and $F_B(x)$, where the A and B subscripts stand for sample A and sample B that must be discriminated. The value returned by the test is the maximum of $|F_A(x) - F_B(x)|$ over the allowed x range.

The sorting of the many new angular observables from better discrimination to worse discrimination (between $t\bar{t}H$ and $t\bar{t}b\bar{b}$, as well as between $t\bar{t}H$ and $t\bar{t}A$), was performed according to the result of the Kolmogorov-Smirnov test between the two samples. This is a statistical compatibility test with some limitations. One limitation in particular is illustrated in the following example.

Let us suppose the test is being applied to two samples, A and B. We consider then an observable with simple distributions: an A-rich region ($A > B$) and a B-rich region ($B > A$), separated at some value of the observable. The absolute value of the cumulative difference between the sample distributions rises in the A-rich region, reaches its maximum at the point of population inversion (this maximum is the output of the K-S test), and decreases in the B-rich region, reaching 0 at the end of the x range. Let us consider another observable, in which the A-rich and B-rich region pattern repeats more than once (n times, for example), but the A and B relative populations in these regions are the the same as in the previous example. If this is the case, the cumulative difference will rise to its maximum at the first population inversion, then decrease to 0 at the following inversion. Since this pattern repeats itself n times, the maximum of the cumulative difference, and thus the K-S test result will be smaller than in the previous example by a factor n. However, both distributions would be equally constraining in a fit to data, for example. An alternative would be to use a different compatibility test, such as a chi-squared test. Of course, every statistical compatibility test has its own limitations, and possibly combining the results of two or three different tests could provide a better discriminant classification.

3.5.3 α-β assignment

From the reasoning presented before, it is clear that θ_1^{123}, θ_3^{23} and θ_4^3 can represent the angles appearing in the amplitude computation of the successive decays of the process on the left side of figure 32. In order for this to be the case, particle 1 should be assigned to $\bar{t}$, particle 2 and 3 should be assigned respectively to the final t and to the Higgs boson, and particle 4 should be either b or $\bar{b}$ from the H decay.

The reconstruction package developed for this analysis can take us as far early in the event as reconstructing the four-momenta of the t, $\bar{t}$ and H before they decay. In order to recover the angles resulting from the amplitude calculations, it is necessary to identify which of the top quarks has radiated the Higgs boson. Otherwise, we can, for example, label t as particle 1 in all events, and $\bar{t}$ and H as particles 2 and 3, respectively. Assuming production through diagrams in which H is radiated by t or $\bar{t}$, this implies that, in approximately one half of the events (those in which the $\bar{t}$ actually radiates the Higgs boson), the angles involved will have the correct physical significance derived from the helicity formalism. However, in the other half of the events, the interpretation is not valid and the same angular distributions are not expected.

In fact, the mixing of these two different distributions may be destructive for any pre-existing discernability between different samples. In figure 33, an example is given of a pair of observables at parton level which exhibit similar distributions for an inclusive set of events (left). If the events are restricted to those in which the Higgs boson (or $b\bar{b}$ system, in $t\bar{t}b\bar{b}$) is known to have been radiated by t (and not radiated by $\bar{t}$ or resulting from $t\bar{t}$ fusion, such as in figure 8 (d)), the distributions are very different (right), and both provide greater signal-background discrimination than in the inclusive set.

These different distributions could potentially be recovered in inclusive sets of events if one knew whether the t or $\bar{t}$ radiated the Higgs boson (or $b\bar{b}$ system, in $t\bar{t}b\bar{b}$) in each particular event. An attempt was made at creating a mechanism to reach for that information. Four samples, with 25 000 events each, were generated to allow this study: two samples of $t\bar{t}H$ events and two samples of $t\bar{t}b\bar{b}$ events. For each process, in one of the samples, all production diagrams were suppressed,
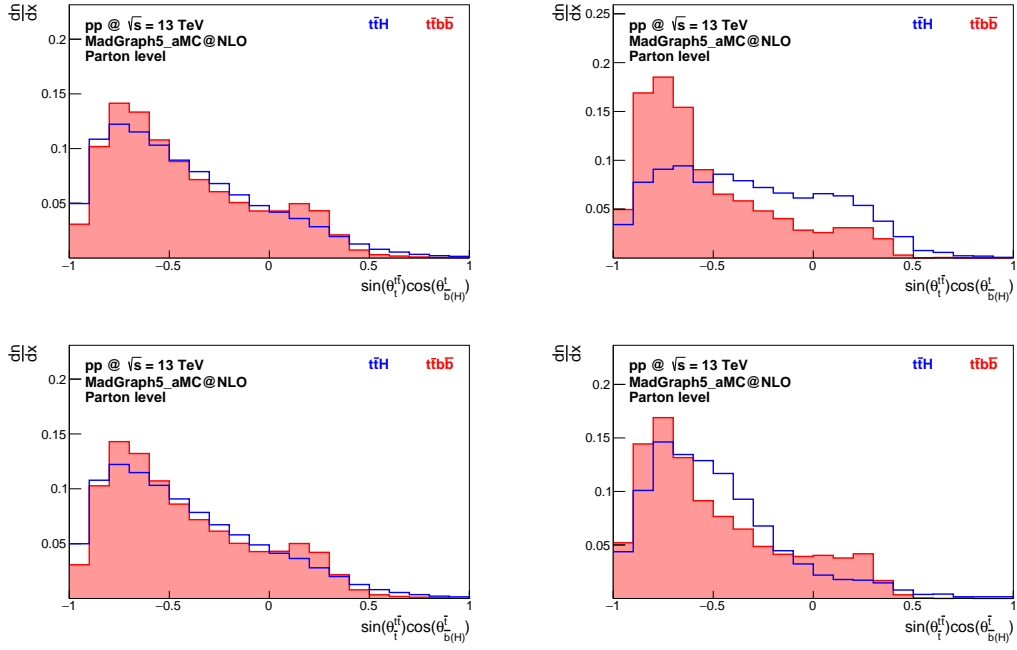
Figure 33: Distributions, at parton level, of $\sin(\theta_t^{t\bar{t}})\cos(\theta_{\bar{b}_H}^t)$ (top) and $\sin(\theta_{\bar{t}}^{t\bar{t}})\cos(\theta_{\bar{b}_H}^{\bar{t}})$ (bottom). The two observables are obtained from one another by exchanging t with $\bar{t}$. **Left:** For an inclusive set of $t\bar{t}H$ and $t\bar{t}b\bar{b}$ events, the variables have similar distributions. **Right:** For a sample of events in which the Higgs boson (or $b\bar{b}$ system, in $t\bar{t}b\bar{b}$) was exclusively radiated by t, it is clear that the t-$\bar{t}$ exchange symmetry is broken. The distributions are different from each other and better at discriminating signal from background.

Table 11: Conditions tested as indicators of whether t or $\bar{t}$ radiated the Higgs boson. The test was performed in signal events generated with the condition that the Higgs boson was radiated by t. The values are the fraction of events in which the condition was true. The uncertainties represented are statistical only.

Condition	Parton level	Rec. with truth-match	Rec. without truth-match
$\theta_{tH} < \theta_{\bar{t}H}$ ($t\bar{t}H$ frame)	0.73 ± 0.02	0.70 ± 0.03	0.57 ± 0.02
$m_{tH} < m_{\bar{t}H}$	0.73 ± 0.02	0.70 ± 0.03	0.57 ± 0.02
$\Delta R_{tH} < \Delta R_{\bar{t}H}$	0.68 ± 0.02	0.66 ± 0.03	0.57 ± 0.02
$\Delta\eta_{tH} < \Delta\eta_{\bar{t}H}$	0.67 ± 0.02	0.64 ± 0.03	0.57 ± 0.02

except those in which the t radiated the H (or $b\bar{b}$ system). The other sample included all production diagrams. The generation and simulation procedure were similar to the ones described in section 3.1, and the samples were submitted to the same analysis routine as the main samples.

Using the $t\bar{t}H$ events generated with the condition that t radiated the Higgs boson, a set of variables were tested for their efficiency at correctly identifying t, as opposed to $\bar{t}$, as the Higgs-radiating quark. This test is summarized in table 11. All variables resulted in similar performances, within the statistical uncertainty. The criterion chosen was to pick the top quark yielding the smallest angle between its direction and the H direction, in the $t\bar{t}H$ rest frame. At generator level, in $(72.9 \pm 2.0)\%$ of events, the direction of H was closer to the direction of t, than to the direction of $\bar{t}$. The efficiency drops to $(57.1 \pm 1.7)\%$ after reconstruction without truth-match.

This condition was used in all events of the main samples to identify which of the top quarks is the most likely to have radiated the Higgs boson (or $b\bar{b}$ system). This t or $\bar{t}$ is labeled t_α, and the other top quark is labeled t_β. All the particles in the decay chains of t_α and t_β also get the subscripts α and β, respectively.

A new set of angular observables was computed using the same prescription as in the previous section. The only difference is that particles 1 to 3 can either be t_α, t_β or H, without repetition. As before, particle 4 can be any of the products of the decay of the top quarks and the Higgs boson. However, if particle 4 results from the decay of one of the top quarks, it is identified using the α or β label.

For this new set of angular observables, histograms were filled with the $t\bar{t}H$,

$t\bar{t}A$ and $t\bar{t}b\bar{b}$ events. The statistical compatibility (Kolmogorov-Smirnov) tests were also performed, between $t\bar{t}H$ and $t\bar{t}b\bar{b}$ distributions and between $t\bar{t}H$ and $t\bar{t}A$ distributions.

3.6 Signal-background discrimination using TMVA

3.6.1 ATLAS analysis at $\sqrt{s} = 8\,\text{TeV}$

The official ATLAS analysis of the physical process addressed in this work ($t\bar{t}H$ production, with $t\bar{t} \to$ dileptonic and $H \to b\bar{b}$) using $8\,\text{TeV}$ data [81] classifies events in different categories, based on total number of jets and number of b-tagged jets. Only good ($p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$) jets are considered for this classification. The notation (nj,mb) is used to refer to the category of events with n jets and m b-tagged jets.

Six independent regions were considered. The signal-rich regions are ($\geq$4j,3b) and ($\geq$4j,$\geq$4b), where the analysis is performed. The regions (2j,2b), (3j,2b), (3j,3b) and ($\geq$4j,2b) are considered signal-depleted regions, since they are expected to exhibit signal-to-background ratios below 1%. These regions were used exclusively for background modelling, with the exception of (3j,3b), which is expected to have some sensitivity to the signal, and was also analysed. Figure 34 shows the expected fractions of background sources in each category. The dominance of $t\bar{t}b\bar{b}$ in the signal-rich regions is clear.

In the ATLAS analysis, a neural network (NN) was implemented in each signal-rich region to discriminate between signal and background, using the NeuroBayes [93] package. The choice of variables relied on the ranking procedure from NeuroBayes. Several classes of variables were considered: object kinematics, global event variables, event shape variables and object pair properties.

As already described in section 3.4.1, selected events must have exactly two opposite-sign leptons and two b-tagged jets. The analysis only considered jets with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$. Leptons were required to have $|\eta| < 2.5$, the p_T-leading lepton is required to have $p_T > 25\,\text{GeV}$ and the sub-leading one must have $p_T > 15\,\text{GeV}$. In the regions with ≥ 4 jets, a maximum of five jets were considered to construct the kinematic variables. The jets were ordered first using all the b-jets, ordered from highest to lowest p_T, and then the untagged jets, also
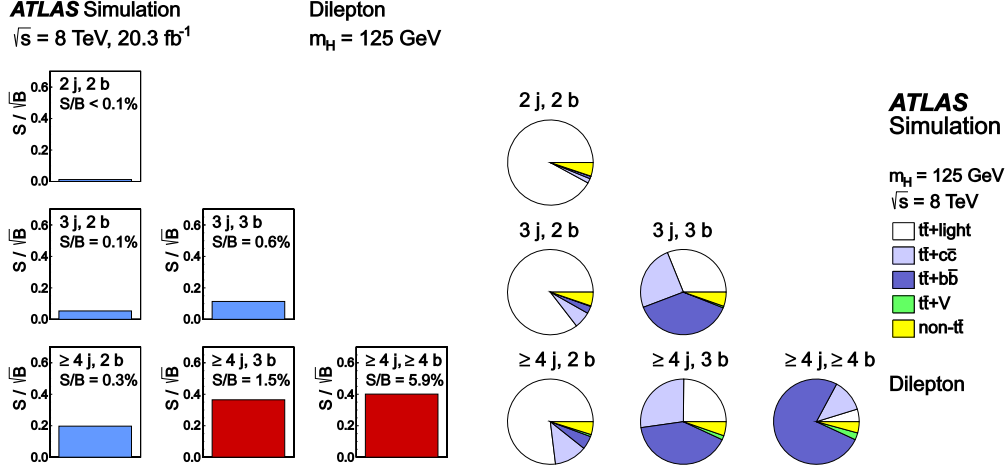
Figure 34: **Left:** Expected signal-to-background ratio (S/B) and signal significance ($S/\sqrt{B}$) for each of the event categories in the ATLAS 8 TeV analysis [81]. Signal-rich regions are painted red. **Right:** Fractions of different background sources in each region. In the signal-rich regions ($\geq$4j,3b) and ($\geq$4j,$\geq$4b), the dominant background process is $t\bar{t}b\bar{b}$.

from highest to lowest p_T.

When necessary for the computation of a variable, the two b-jets closest in ΔR to the leptons are assigned to the corresponding top quark decays and, among the remaining b-jets, the two b-jets leading in p_T are assigned to the H decay. None of the variables used require t or $\bar{t}$ reconstruction.

A set of 10 variables was used as input for the NN in each region. The full set of 17 variables is listed in table 12, with the respective variable definitions. The definitions in the table should be completed with information about the event shape variables $H4$ and Aplan$_{\text{jet}}$, where 'Aplan' is short for aplanarity. The momentum tensor $p^{\alpha\beta}$ and the fifth Fox-Wolfram moment $H4$ are defined as

$$p^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |p_i|^2} \quad , \quad H4 = \sum_{i,j} \frac{|p_i||p_j|}{E_{\text{vis}}^2} P_4(\cos\theta_{ij}), \tag{31}$$

where in the case of Aplan$_{\text{jet}}$, the sum is over jets only, and in the case of $H4$, it is over jets and leptons, and should include $i = j$ terms. E_{vis} is the total visible energy in the event and P_4 is the Legendre polynomial of order 4.

Figure 35 shows signal and background distributions, obtained by ATLAS, for some of the top-ranked variables in the ($\geq$4j,3b) and ($\geq$4j,$\geq$4b) categories [81]. Also shown are distributions of the neural network output for both categories. In

107

Table 12: Discriminant variables used as input to the neural networks (NN), and respecive definitions [81]. The variable rank in each of the signal–rich categories is also listed. For the analysis of each category, only the 10 best ranked variables were used.

Variable	Definition	NN rank		
		$\geq$4j, $\geq$4b	$\geq$4j, 3b	3j, 3b
1 $\Delta\eta_{jj}^{max\,\Delta\eta}$	Maximum $\Delta\eta$ between any two jets in the event	1	1	1
2 $m_{bb}^{min\Delta R}$	Mass of the combination of the two b-tagged jets with the smallest ΔR	2	8	–
3 $m_{b\bar{b}}$	Mass of the two b-tagged jets from the Higgs candidate system	3	–	–
4 $\Delta R_{hl}^{min\,\Delta R}$	ΔR between the Higgs candidate and the closest lepton	4	5	–
5 N_{30}^{Higgs}	Number of Higgs candidates within 30 GeV of the Higgs mass of 125 GeV	5	2	5
6 $\Delta R_{bb}^{max\,p_T}$	ΔR between the two b-tagged jets with the largest vector sum p_T	6	4	8
7 $Aplan_{jet}$	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built with all jets	7	7	–
8 $m_{jj}^{min\,m}$	Minimum dijet mass between any two jets	8	3	2
9 $\Delta R_{hl}^{max\,\Delta R}$	ΔR between the Higgs candidate and the furthest lepton	9	–	–
10 $m_{jj}^{closest}$	Dijet mass between any two jets closest to the Higgs mass of 125 GeV	10	–	10
11 H_T	Scalar sum of jet p_T and lepton p_T values	–	6	3
12 $\Delta R_{bb}^{max\,m}$	ΔR between the two b-tagged jets with the largest invariant mass	–	9	–
13 $\Delta R_{lj}^{min\,\Delta R}$	Minimum ΔR between any lepton and jet	–	10	–
Centrality	Sum of the p_T divided by sum of the E for all jets and both leptons	–	–	7
$m_{jj}^{max\,p_T}$	Mass of the combination of any two jets with the largest vector sum p_T	–	–	9
$H4$	Fifth Fox–Wolfram moment computed using all jets and both leptons	–	–	4
p_T^{jet3}	p_T of the third leading jet	–	–	6

these, the $t\bar{t}H$ signal and the different backrounds were fitted to data events, which are also represented.

3.6.2 Analyses at $\sqrt{s} = 13\,\text{TeV}$ using TMVA

For the study presented here, addressing signal-background discrimination at $\sqrt{s} = 13\,\text{TeV}$, the region considered was ($\geq$4j,$\geq$3b), which comprises the two signal-rich regions of the ATLAS analysis. Besides the selection cuts described in section 3.4.1, the cut $|m_{\ell\ell} - m_Z| > 10\,\text{GeV}$, used in data analyses to exclude the Z boson mass peak, was also applied.

The TMVA [80] package was used to implement the multivariate analysis. Several different methods were enabled: a likelihood method, a boosted Fisher method, two variants of the boosted decision tree (BDT) method, and a multi-layer perceptron (MLP) method, which is a neural network. Three sets of variables were defined:

- 'ATLAS' set: the set of discriminant variables used in the ATLAS 8 TeV analysis of the ($\geq$4j,3b) and ($\geq$4j,$\geq$4b) together, that is, the first 13 variables listed in table 12.

- 'Angular' set: set of angular observables of the kind $f(\theta_1^{123})g(\theta_4^3)$ or $f(\theta_3^{23})g(\theta_4^3)$, with or without α-β assignment. All angular observables of this kind were ordered by the output value of the statistical compatibility test between the respective $t\bar{t}H$ and $t\bar{t}b\bar{b}$ distributions, at reconstruction level without truth-match and after the ($\geq$4j,$\geq$3b) and $|m_{\ell\ell} - m_Z| > 10\,\text{GeV}$ cuts. The angular observable ranked first (the one with the highest result in the K-S test) enters this set. The second one enters this set if the linear correlation coefficient in signal events relative to the first one is not larger than 60%. The third one enters the set if the coefficients relative to the two variables already in the set are not larger than 60%, and so on, until 13 variables are selected. Two more observables were added to this set: the reconstructed Higgs mass as obtained from the kinematic fit (this is different from the ATLAS definition of the Higgs candidate mass) and $\theta_{W^+}^H$, which is the angle between the direction of W^+, in the H frame, and the direction of H, in the $\bar{t}H$ rest frame.

109
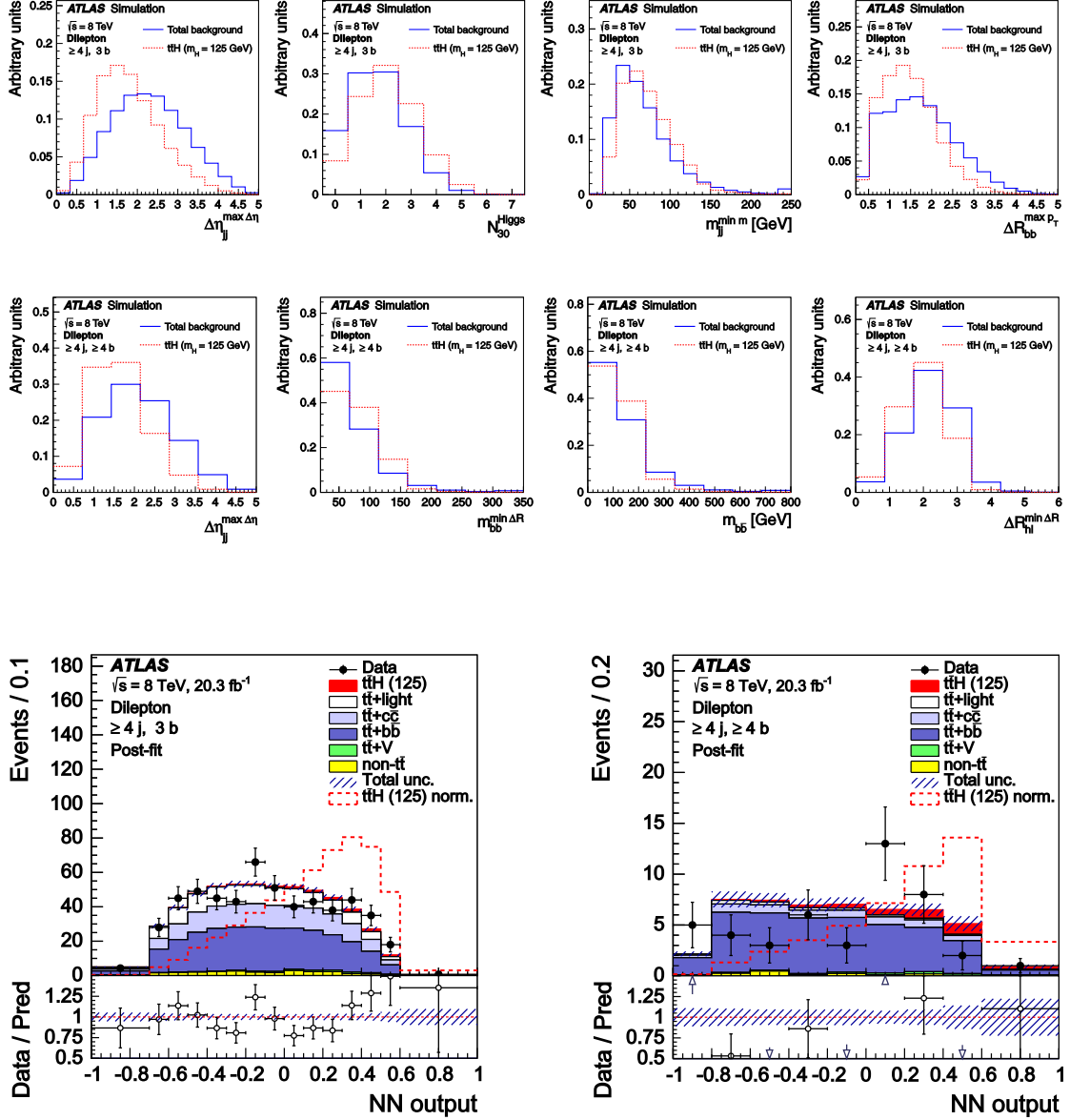
Figure 35: **Top:** Signal and background distributions of the four top-ranked variables in the (≥4j,3b) category. **Middle:** The same for the (≥4j,≥4b) category. **Bottom:** Distributions of the NN output, for the (≥4j,3b) category (left) and for the (≥4j,≥4b) category (right). The represented signal and backgrounds were fitted to data events, which are also shown. A normalized signal distribution is shown as a dashed histogram because, at the correct scale, the $t\bar{t}H$ distribution (filled in red) is small compared to the backgrounds.

- 'Truth-matched angular': the same set of 15 variables as in the 'Angular' set, evaluated using truth-matched objects.

The procedure employed in the choice of the 'Angular' set is motivated by the fact that, in a multivariate analysis, the use of highly correlated variables is undesirable. A variable correlated to the ones already in the set should not increase significantly the amount of information available and can compromise the performance of discrimination.

Five different analyses were performed, corresponding to different combinations of variables. The first three analyses correspond to each one of the variable sets defined before. The fourth analysis corresponds to a comprehensive set, combining the 'ATLAS' and 'Angular' variables. By comparing the performance of the comprehensive set against the 'ATLAS' set, an assessment can be made of the potential gain in signal-background discrimination resulting from the use of the angular observables being studied. The fifth analysis also used the comprehensive set of variables. It was run on a background sample containing not only $t\bar{t}b\bar{b}$, but the full set of generated SM background samples, as described in section 3.1.

3.7 Review of observables proposed for testing the Higgs boson CP nature

The reconstruction package developed was used to study a set of variables proposed by other authors as potentially useful for probing the CP nature of the Higgs boson in $t\bar{t}H$ production. This study had two goals: the first goal was to validate the new angular observables proposed here against the ones already published, and the second goal was to evaluate the impact of an experimental setting, with its inevitable degradation of information, in the discrimination power of the published observables, which were obtained from parton-level studies.

The angle between the direction of flight of ℓ^+ in the t rest frame and the direction of flight of ℓ^- in the $\bar{t}$ rest frame will be referred to as $\theta(\ell^+, \ell^-)$. In [63], $\cos[\theta(\ell^+, \ell^-)]$ was proposed as a discriminant between scalar and pseudoscalar Higgs boson production.

Azimuthal angle differences were proposed in [65] for the same effect. The observable $\Delta\phi^{t\bar{t}}(\ell^+, \ell^-)$ is defined as the azimuthal angle difference between the

ℓ^+ direction, in the t rest frame, and the ℓ^- direction, in the $\bar{t}$ rest frame. The $\Delta\phi^H(\ell^+, \ell^-)$ is the same azimuthal angle difference, but in this case the momenta of the leptons are both evaluated in the H frame. A third proposed angle is $\Delta\phi(t, \bar{t})$, where the t and $\bar{t}$ directions are evaluaed in the laboratory rest frame.

A more elaborate observable was proposed in [14]. The angle between the ℓ^+ and ℓ^- directions, projected onto the plane perpendicular to the H direction in the lab frame, is defined as $\theta^{\perp H}(\ell^+, \ell^-)$. Then, β is defined as the sign of $(\vec{p_b} - \vec{p_{\bar{b}}}) \cdot (\vec{p_{\ell^-}} \times \vec{p_{\ell^+}})$ (b and $\bar{b}$ result from the t and $\bar{t}$ decays, respectively) and β' is defined as the sign of $\vec{p_H} \cdot (\vec{p_{\ell^-}} \times \vec{p_{\ell^+}})$. The proposed observables are the products $\beta \cdot \theta^{\perp H}(\ell^+, \ell^-)$ and $\beta' \cdot \theta^{\perp H}(\ell^+, \ell^-)$. In a CP-violating scenario, the β and β' factors make these observables sensitive to the relative phase of scalar and pseudoscalar components of the Higgs boson.

Several observables involving projections of the t and $\bar{t}$ momenta were proposed in [64]. The transverse momentum vector is written as $\vec{p_T}$. The proposed observables are:

$$a_1 = \text{sign}(\vec{p_{Tt}} \cdot \vec{p_{T\bar{t}}}) \quad , \quad a_2 = \text{sign}(p_t^x p_{\bar{t}}^x)$$
$$b_1 = \frac{\vec{p_{Tt}} \cdot \vec{p_{T\bar{t}}}}{p_{Tt} p_{T\bar{t}}} \quad , \quad b_2 = \frac{\vec{p_{Tt}} \cdot \vec{p_{T\bar{t}}}}{p_t p_{\bar{t}}} \quad , \quad b_3 = \frac{p_t^x p_{\bar{t}}^x}{p_{Tt} p_{T\bar{t}}} \quad , \quad b_4 = \frac{p_t^z p_{\bar{t}}^z}{p_t p_{\bar{t}}}. \tag{32}$$

An important remark is that some of these observables require a reconstruction of the t and $\bar{t}$, such as the one performed in the kinematic fit. For the sake of an easier comparison, the distribution plots published along with the proposals are shown in the results section. All the observables described were computed at parton level and at reconstruction level, with and without truth-match. Histograms of the distributions for $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$ samples were produced, and the statistical compatibility tests were performed between the $t\bar{t}H$ and $t\bar{t}A$ distributions.

4   Results

In the current section, results of the studies described in section 3 are presented and discussed. The section is logically divided in two subsections, corresponding to the twofold motivation given for the use of new angular observables: discrimination between signal and background, and discrimination between Higgs CP scenarios (scalar and pseudoscalar).

The first part adresses discrimination between signal and background. Initially, distributions of some of the most interesting observables for this effect are shown and interpreted. Then, the results of integrating such observables in a TMVA analysis are presented. The obtainable signal efficiencies and background rejection rates are compared among the different sets of observables tested, including the set from the ATLAS analysis of this process. The importance of combinatorial background for signal sensitivity is discussed, and the earlier choice of only considering the $t\bar{t}b\bar{b}$ background is validated, by applying the TMVA analysis on a full set of SM backgrounds, and confirming that similar results are obtained.

The second part adresses discrimination between $t\bar{t}H$ and $t\bar{t}A$ production. Again, it starts with the presentation of distributions of new angular observables, in which a good discrimination is visible. Then, it focuses on the observables collected in the literature review: their distributions are shown and compared to the originally published ones. The effects of the detector simulation and the kinematic fit are discussed, and this collection of observables is compared to the new angular ones.

4.1   Discriminating between $t\bar{t}H$ and background

4.1.1   Angular observables from θ_Y^X angles

One particular class of observables, commonly used with angular variables, are forward-backward asymmetries (A_{FB}). Given a variable x that can take either positive or negative values, the forward-backward asymmetry $A_{FB}(x)$ is defined as

$$A_{FB} = \frac{N(x \geq 0) - N(x < 0)}{N(x \geq 0) + N(x < 0)},$$

(33)

where N is the number of events verifying the condition inside the brackets. A_{FB} ranges from -1 ($x < 0$ in all events) to +1 ($x > 0$ in all events), and is equal to 0 if events are symmetrically distributed.[2]

Considering one background and one signal sample only, the expected $A_{FB}(x)$ in data is given by

$$A_{FB}^{data}(x) = sA_{FB}^s(x) + bA_{FB}^b(x), \tag{34}$$

where s and b are the relative contributions of signal and background events. Since $A_{FB}^{data}(x)$ is an observable and $A_{FB}^s(x)$ and $A_{FB}^b(x)$ can be obtained from simulation, the forward-backward asymmetry can be used to fit the relative background and signal contributions to data. The larger the difference between $A_{FB}^s(x)$ and $A_{FB}^b(x)$, the more stringent the resulting fit will be.

Forward-backward asymmetries were computed for observables of the type $\cos(\theta_1^{123})\cos(\theta_4^3)$ and $\cos(\theta_3^{23})\cos(\theta_4^3)$ without α-β assignment. Some observables exhibiting remarkable asymmetry differences between $t\bar{t}H$ and $t\bar{t}b\bar{b}$ have been published in [94]. The published asymmetries correspond to the set of observables $\cos(\theta_H^{\bar{t}H})\cos(\theta_Y^H)$, where Y can be any of the leptons, W bosons, or b-quarks. These results are summarized in table 13, and the distributions for $Y = \ell^+$ and $Y = \ell^-$ are shown in figure 36. At the time of the referred publication, the samples used were generated at leading order (LO) in QCD and the procedure for reconstruction without truth-match was slightly different from the one described previously.

The search for new discriminant variables between $t\bar{t}H$ events and the irreducible background $t\bar{t}b\bar{b}$ resulted in an extensive set of 960 new angular observables. Many of these observables are of course highly correlated, namely those differing only in the particle chosen as particle 4, if this particle results from the same decay chain in both observables. For example, changing particle 4 from l^+ to W^+ or b_t results in a new observable highly correlated to the original one. Also, in observables without α-β assignment, the exchange of every particle and its anti-particle (H left unchanged) results in a new observable that, although not obviously correlated to the original one, exhibits similar distributions.

---

[2]It is elucidating to notice that the difference between signal and background forward-backward asymmetries can be defined in terms of the cumulative distributions as $-2(F_s(0) - F_b(0))$. In this sense, the output of the Kolmogorov-Smirnov test can be understood as a generalized asymmetry difference, computed at a value of x such that this difference is maximized.

Table 13: Forward-backward asymmetries A_{FB} of distributions of the type $\cos(\theta_H^{\bar{t}H})\cos(\theta_Y^H)$, for several Y [94]. The asymmetries are shown for $t\bar{t}H$ and $t\bar{t}b\bar{b}$ events and the difference $\Delta A_{FB} = A_{FB}^{t\bar{t}H} - A_{FB}^{t\bar{t}b\bar{b}}$ is also presented, at parton level before applying selection cuts and at reconstruction level without truth-match. These results were obtained using samples generated at leading order (LO) in QCD and using a slightly different procedure for reconstruction without truth-match.

Y	Parton level, no cuts			Rec. without truth-match		
	$A_{FB}^{t\bar{t}H}$	$A_{FB}^{t\bar{t}bb}$	ΔA_{FB}	$A_{FB}^{t\bar{t}H}$	$A_{FB}^{t\bar{t}bb}$	ΔA_{FB}
ℓ^+	-0.157	-0.137	-0.020	-0.141	-0.268	$+0.127$
ℓ^-	$+0.291$	$+0.056$	$+0.235$	$+0.331$	$+0.118$	$+0.213$
W^+	-0.154	-0.119	-0.035	-0.119	-0.275	$+0.156$
W^-	$+0.317$	$+0.067$	$+0.250$	$+0.348$	$+0.127$	$+0.221$
b_t	-0.155	-0.141	-0.014	-0.179	-0.306	$+0.127$
$\bar{b}_{\bar{t}}$	$+0.293$	$+0.053$	$+0.240$	$+0.334$	$+0.117$	$+0.217$
b_H	$+0.000$	$+0.001$	-0.001	$+0.086$	-0.048	$+0.134$
$\bar{b}_H$	$+0.000$	-0.001	$+0.001$	-0.086	$+0.048$	-0.134
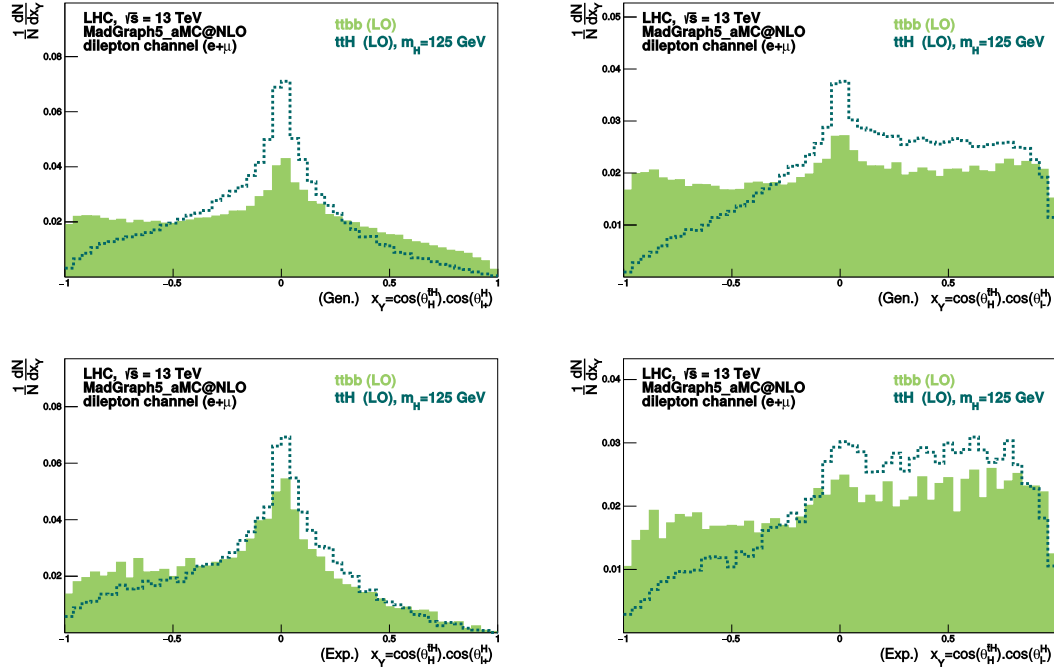
Figure 36: Angular observable distributions exhibiting forward-backward asymmetry differences between $t\bar{t}H$ and $t\bar{t}b\bar{b}$ [94]. **Left:** $\cos(\theta_H^{\bar{t}H})\cos(\theta_{\ell^+}^H)$. **Right:** $\cos(\theta_H^{\bar{t}H})\cos(\theta_{\ell^-}^H)$. The upper plots are the distributions at parton level, and the lower plots are the same distributions at reconstruction level without truth-match. These results were obtained using samples generated at leading order (LO) in QCD and using a slightly different procedure for reconstruction without truth-match. Similar asymmetry differences were obtained for W bosons instead of leptons.

Of particular interest for signal-background discrimination, are the observables with similar distributions in $t\bar{t}H$ and $t\bar{t}A$ events. Such observables have a signal-background discriminating power independent of the CP nature of the Higgs boson. In a data analysis, this would result in a model-independent constraint on the ratio between $t\bar{t}H$ production and background production.

Angular observables were separated in 4 different categories, explained below. Each of the following figures contains two examples of distributions among the best classified ones in each category. For the choice of the observables shown, some independence relatively to the Higgs CP scenario was also required, and the $t\bar{t}A$ distribution is also represented to demonstrate this feature. The distributions are shown at parton level and at reconstruction level, with and without truth-match. The events represented at parton level are restricted to those for which the kinematic fit without truth-match found at least one solution. Also visible is the output of the $(t\bar{t}H, t\bar{t}b\bar{b})$ Kolmogorov-Smirnov (K-S) test.

- Figure 37 shows 2 examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$ without α-β assignment;

- Figure 38 shows 2 examples of observables of the type $f(\theta_3^{23})g(\theta_4^3)$ without α-β assignment;

- Figure 39 shows 2 examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$ with α-β assignment;

- Figure 40 shows 2 examples of observables of the type $f(\theta_3^{23})g(\theta_4^3)$ with α-β assignment.
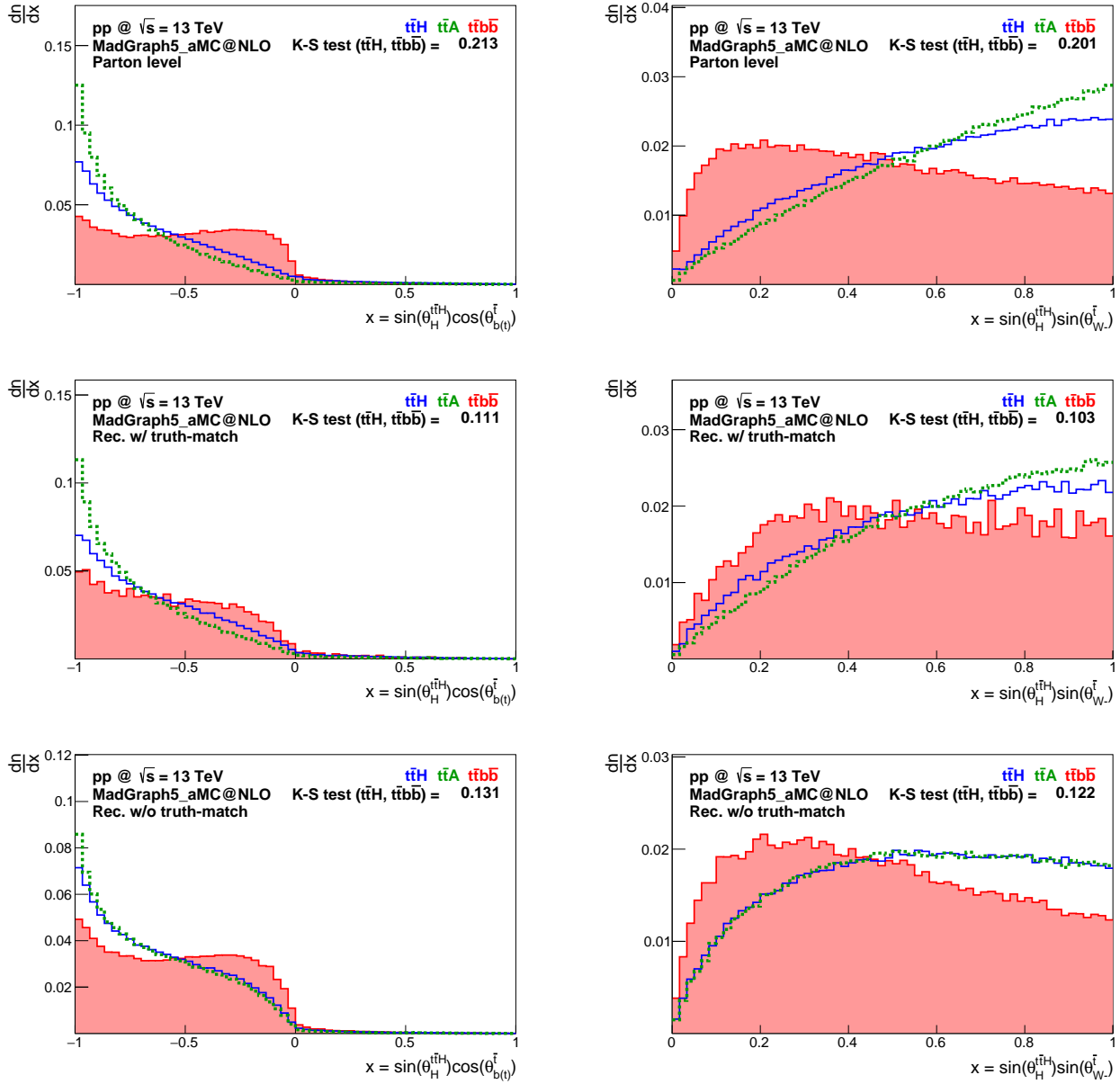
Figure 37: Two examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$, without α-β assignment, that performed well in the K-S test between $t\bar{t}H$ and $t\bar{t}b\bar{b}$ samples. Distributions are shown for $t\bar{t}H$, $t\bar{t}b\bar{b}$ and $t\bar{t}A$ events, at parton level (top), reconstruction level with truth-match (middle) and reconstruction level without truth-match (bottom). **Left:** $\sin(\theta_H^{t\bar{t}H})\cos(\theta_{b_t}^{\bar{t}})$. **Right:** $\sin(\theta_H^{t\bar{t}H})\sin(\theta_{W^-}^{\bar{t}})$.
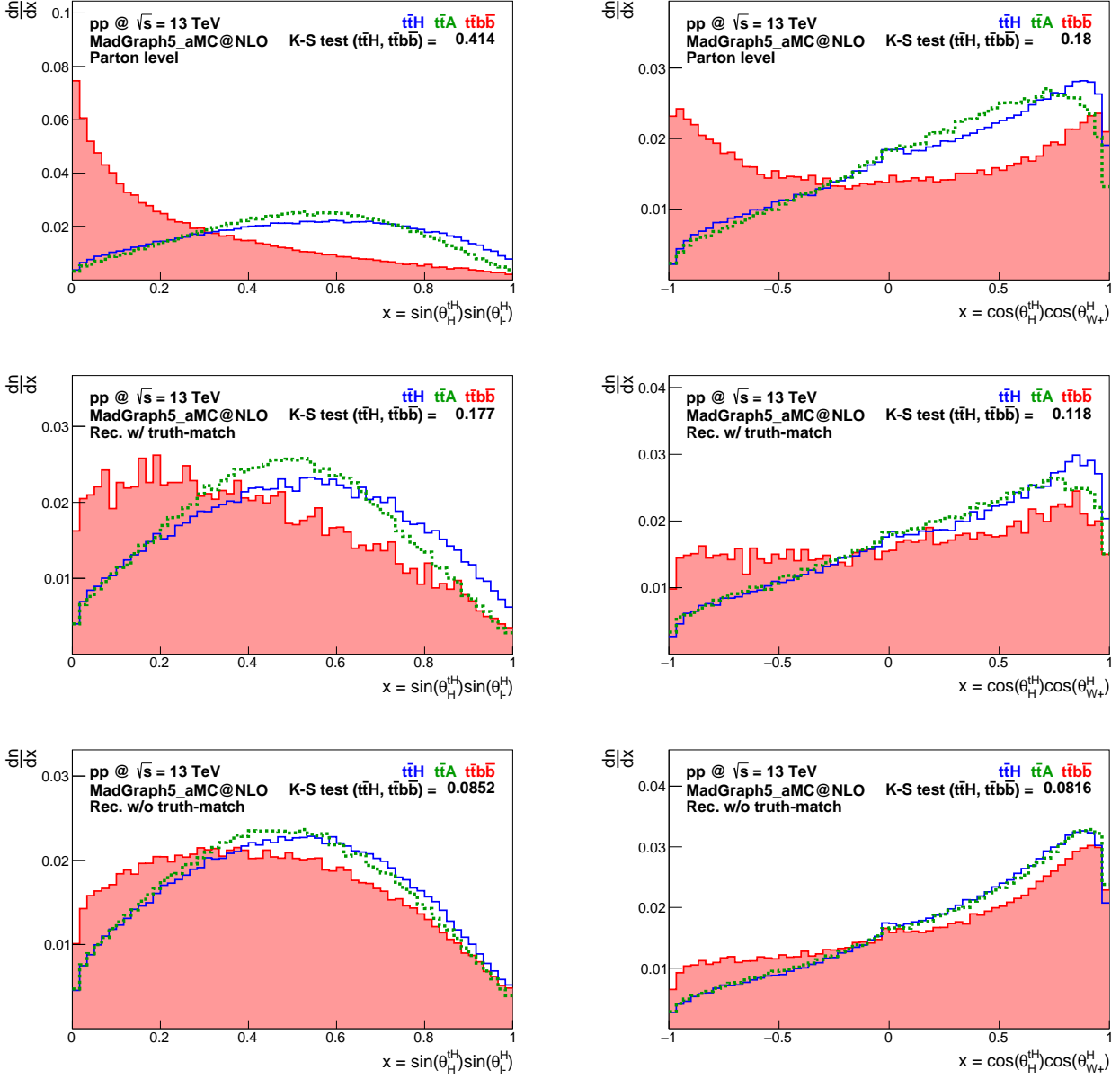
Figure 38: Two examples of observables of the type $f(\theta_3^{23})g(\theta_4^3)$, without α-β assignment, that performed well in the K-S test between $t\bar{t}H$ and $t\bar{t}b\bar{b}$ samples. Distributions are shown for $t\bar{t}H$, $t\bar{t}b\bar{b}$ and $t\bar{t}A$ events, at parton level (top), reconstruction level with truth-match (middle) and reconstruction level without truth-match (bottom). **Left:** $\sin(\theta_H^{tH})\sin(\theta_{\ell^-}^H)$. **Right:** $\cos(\theta_H^{tH})\cos(\theta_{W^+}^H)$.
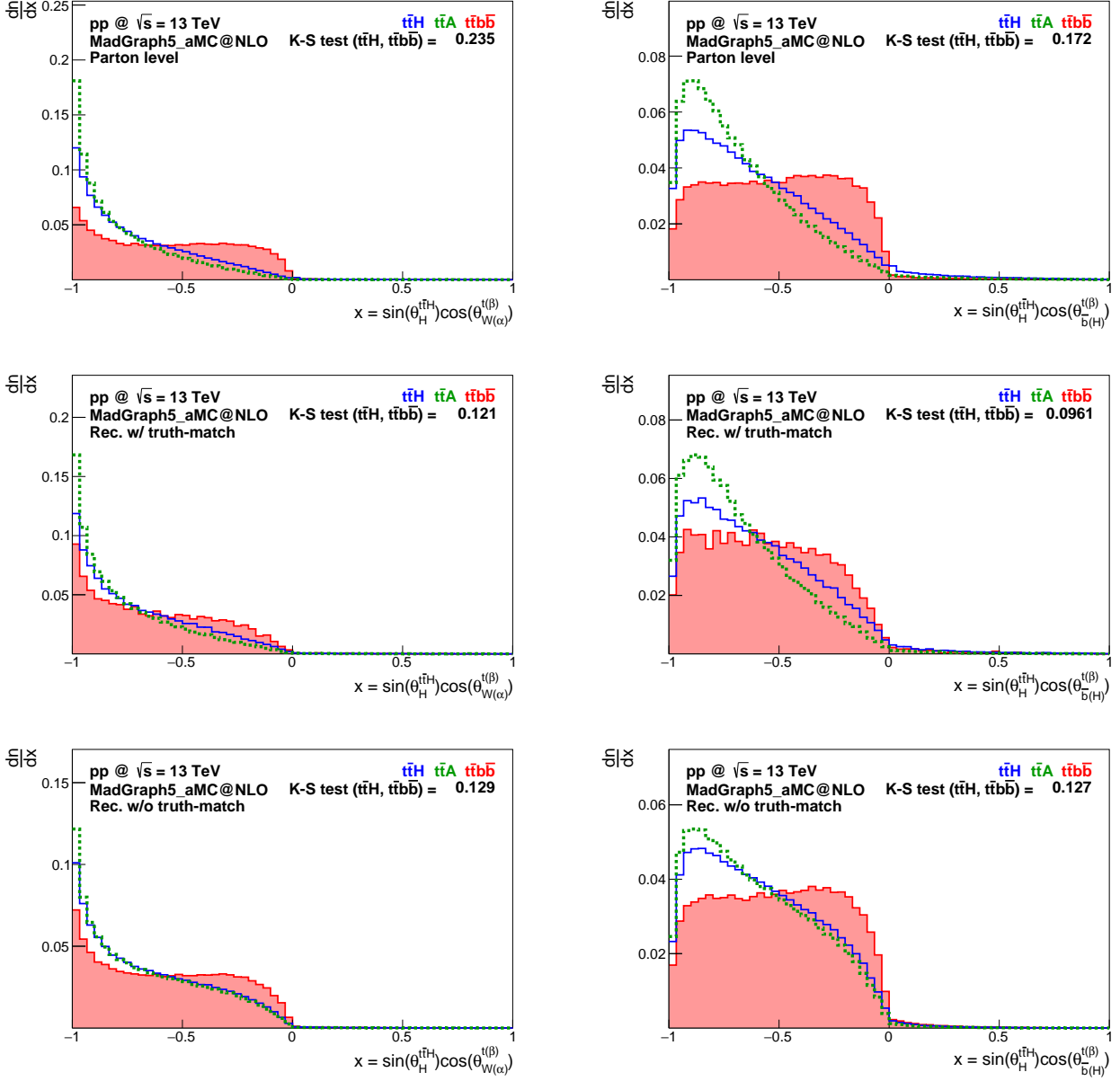
Figure 39: Two examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$, with α-β assignment, that performed well in the K-S test between $t\bar{t}H$ and $t\bar{t}b\bar{b}$ samples. Distributions are shown for $t\bar{t}H$, $t\bar{t}b\bar{b}$ and $t\bar{t}A$ events, at parton level (top), reconstruction level with truth-match (middle) and reconstruction level without truth-match (bottom). **Left:** $\sin(\theta_H^{t\bar{t}H})\cos(\theta_{W_\alpha}^{t_\beta})$. **Right:** $\sin(\theta_H^{t\bar{t}H})\cos(\theta_{\bar{b}_H}^{t_\beta})$.
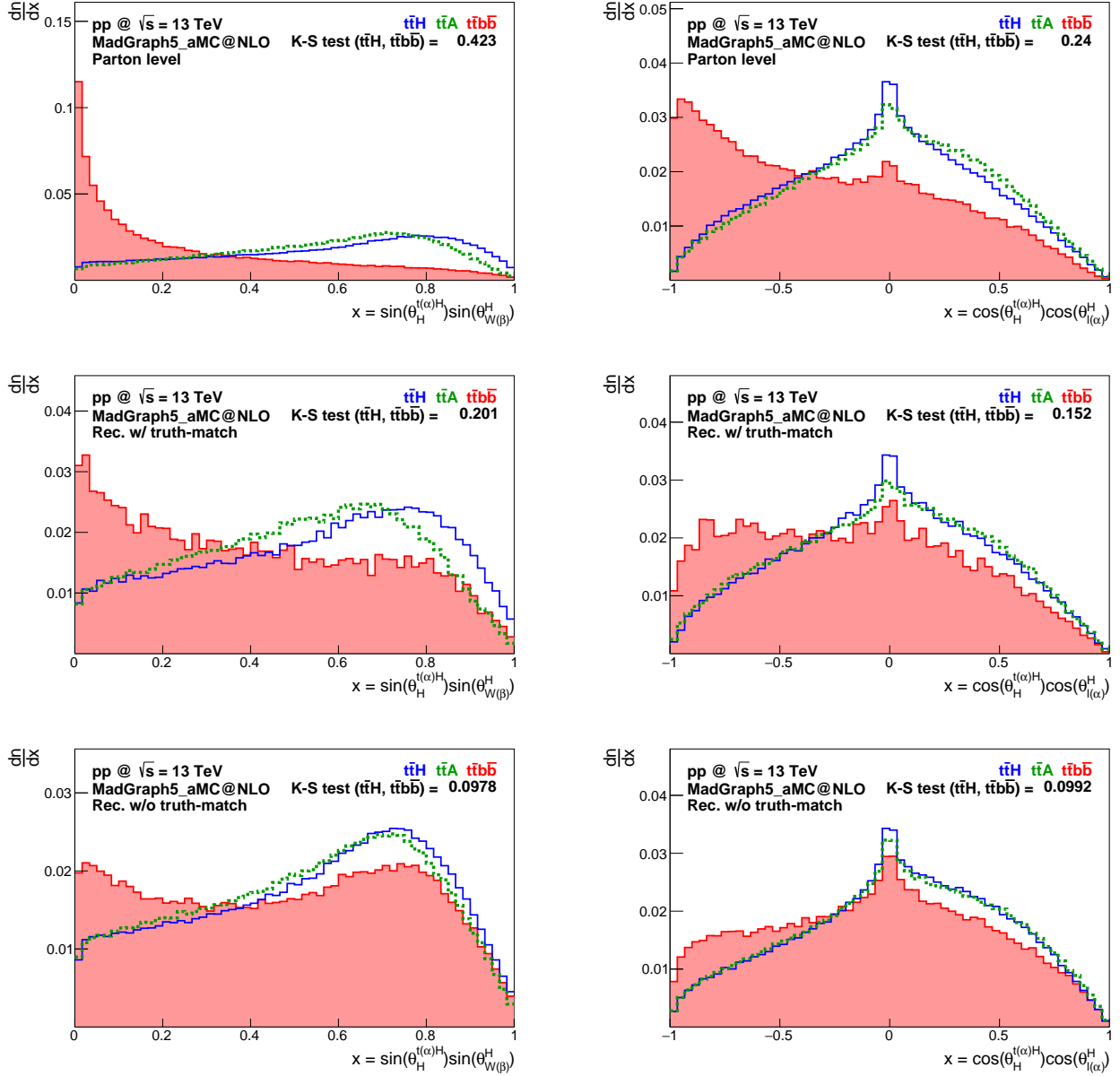
Figure 40: Two examples of observables of the type $f(\theta_3^{23})g(\theta_4^3)$, with α-β assignment, that performed well in the K-S test between $t\bar{t}H$ and $t\bar{t}b\bar{b}$ samples. Distributions are shown for $t\bar{t}H$, $t\bar{t}b\bar{b}$ and $t\bar{t}A$ events, at parton level (top), reconstruction level with truth-match (middle) and reconstruction level without truth-match (bottom). **Left:** $\sin(\theta_H^{t_\alpha H})\sin(\theta_{W_\beta}^H)$. **Right:** $\cos(\theta_H^{t_\alpha H})\cos(\theta_{\ell_\alpha}^H)$.

From the set of 960 new angular observables, many were found to exhibit good discriminating power between signal and background. In spite of the already mentioned correlations within this set, the selection represented above demonstrates that at least a few distinct families of distributions arise, with good signal-background discrimination and reasonable independence with respect to the CP nature of the Higgs boson. The angular observables with the best Kolmogorov-Smirnov test results yield values larger than 0.12 in that test.

4.1.2   Analyses at $\sqrt{s} = 13\,\text{TeV}$ using TMVA

Signal and background ($t\bar{t}b\bar{b}$ only) distributions of the variables in each set used for the multivariate analyses are shown in the following figures. The distributions shown were produced after the signal-region cuts. The 'ATLAS' and 'Angular' sets were used with all events reconstructed successfully without truth-match. On the other hand, the 'Truth-matched angular' set can only be used with events reconstructed successfully with truth-match, which results in a much smaller sample.

- Figure 41 shows distributions for the variables in the 'ATLAS' set, defined in table 12;

- Figure 42 shows distributions for the variables in the 'Angular' set and their definition;

- Figure 43 shows distributions for the variables in the 'Truth-matched angular' set.
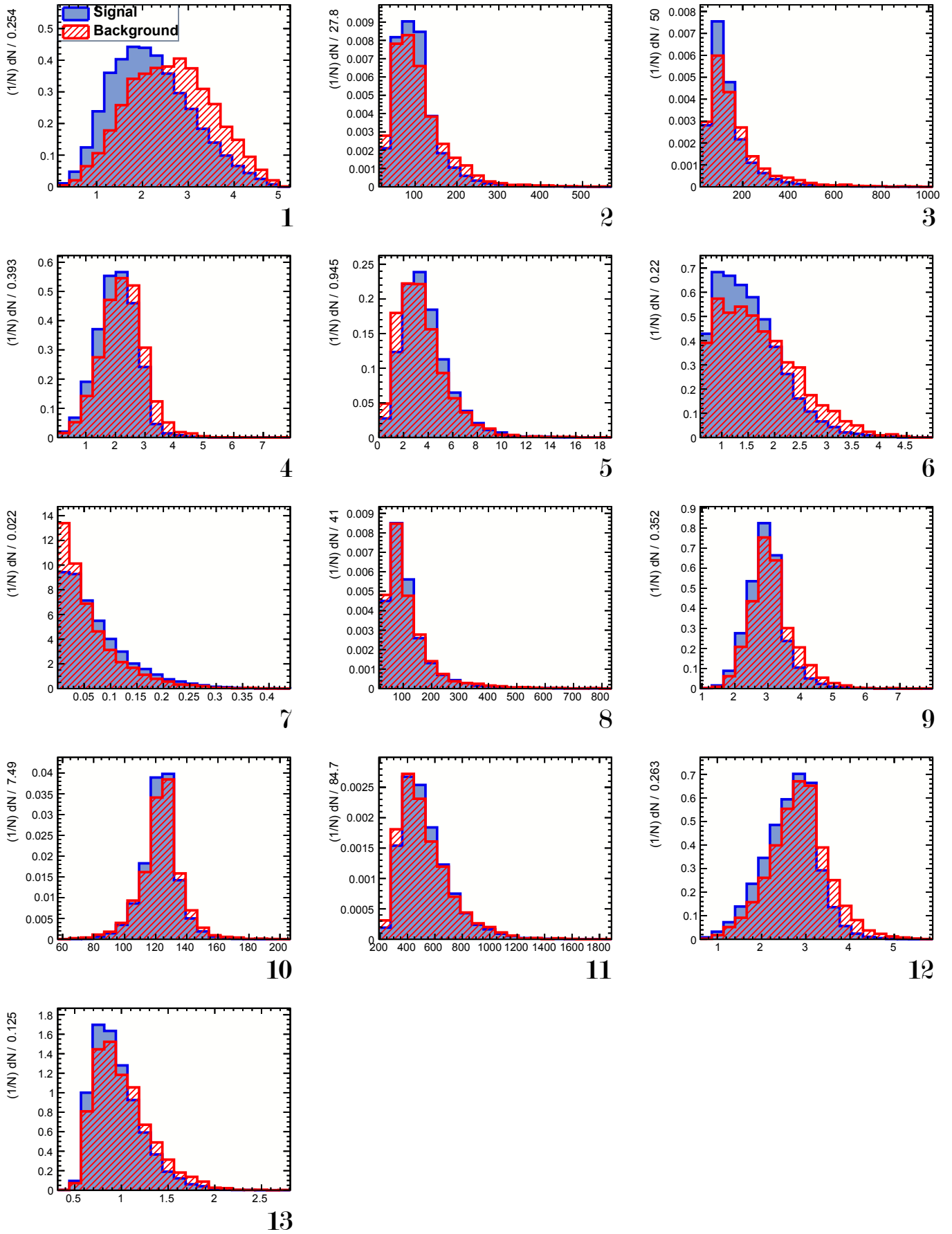
Figure 41: Signal and background ($t\bar{t}b\bar{b}$ only) distributions for the observables in the 'ATLAS' set.

1:	$\Delta\eta_{jj}^{\max\Delta\eta}$	**2:**	$m_{bb}^{\min\Delta R}$	**3:**	m_{bb}
4:	$\Delta R_{hl}^{\min\Delta R}$	**5:**	N_{Higgs}^{30}	**6:**	$\Delta R_{bb}^{\max p_T}$
7:	$\text{Aplan}_{\text{jet}}$	**8:**	$m_{jj}^{\min m}$	**9:**	$\Delta R_{hl}^{\max\Delta R}$
10:	$m_{jj}^{closest}$	**11:**	H_T	**12:**	$\Delta R_{bb}^{\max m}$
13:	$\Delta R_{lj}^{\min\Delta R}$				
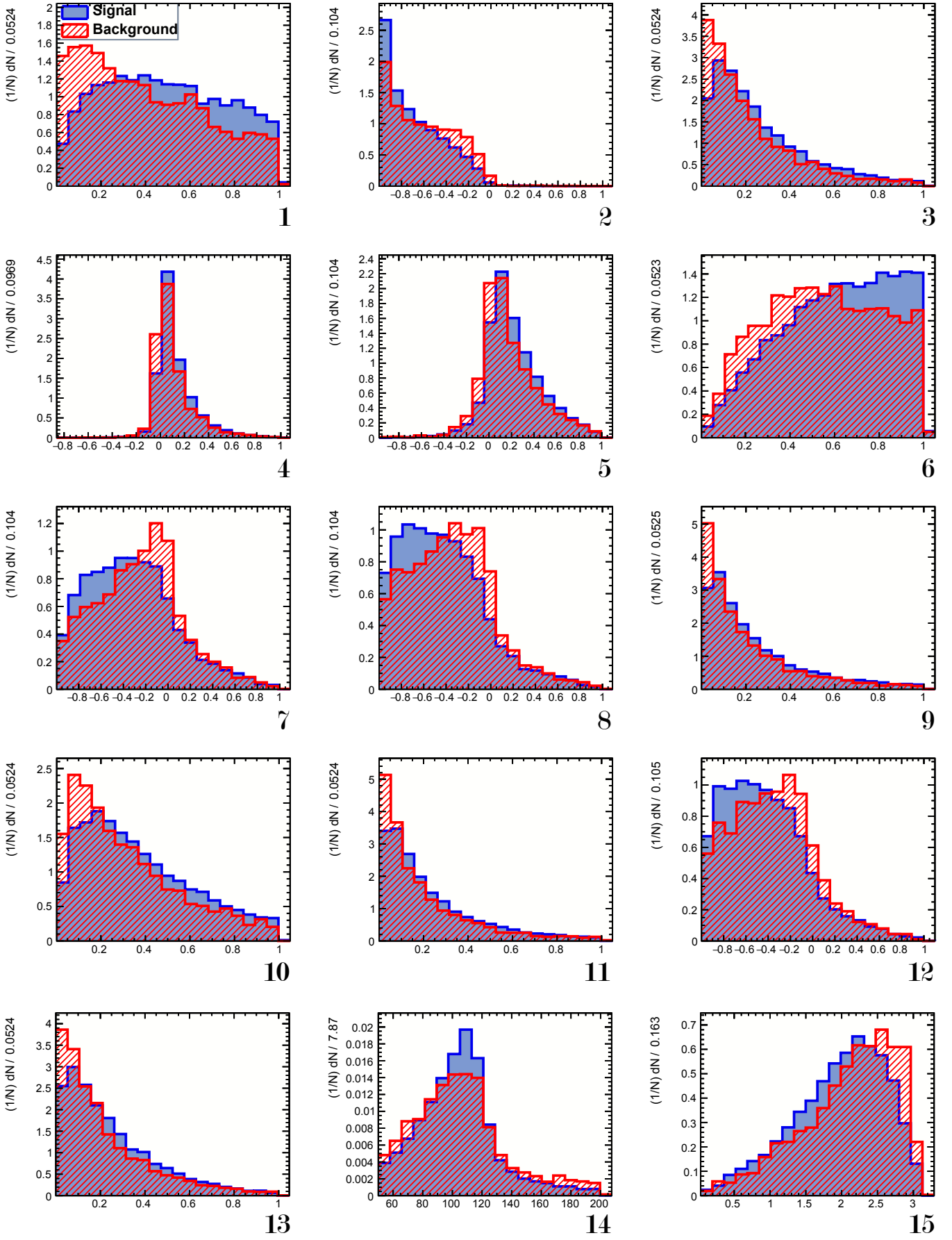
Figure 42: Signal and background ($t\bar{t}b\bar{b}$ only) distributions for the observables in the 'Angular' set.

1:	$\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{W_\beta}^{H})$	**2:**	$\sin(\theta_{t_\beta}^{t\bar{t}H})\cos(\theta_{W^-}^{t})$	**3:**	$\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{W_\alpha}^{H})$
4:	$\cos(\theta_{t_\beta}^{Ht_\beta})\sin(\theta_{\bar{b}_H}^{t_\beta})$	**5:**	$\cos(\theta_{t_\beta}^{Ht_\beta})\sin(\theta_{b_H}^{t_\beta})$	**6:**	$\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{b_H}^{H})$
7:	$\sin(\theta_{H}^{t\bar{t}H})\cos(\theta_{b_H}^{t_\alpha})$	**8:**	$\sin(\theta_{H}^{t\bar{t}H})\cos(\theta_{b_H}^{\bar{t}})$	**9:**	$\sin(\theta_{t}^{t\bar{t}H})\sin(\theta_{\bar{b}_H}^{\bar{t}})$
10:	$\sin(\theta_{\bar{t}}^{t\bar{t}H})\sin(\theta_{b_H}^{t})$	**11:**	$\sin(\theta_{\bar{t}}^{t\bar{t}H})\sin(\theta_{\bar{b}_H}^{t})$	**12:**	$\sin(\theta_{H}^{t\bar{t}H})\cos(\theta_{\bar{b}_H}^{\bar{t}})$
13:	$\sin(\theta_{t_\alpha}^{t\bar{t}H})\sin(\theta_{b_\beta}^{H})$	**14:**	$m_{b\bar{b}}$	**15:**	$\theta_{W^+}^{H}$

Figure 43: Signal and background ($t\bar{t}b\bar{b}$ only) distributions for the observables in the 'Truth-matched angular' set.

1:	$\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{W_\beta}^{H})$	**2:**	$\sin(\theta_{H}^{t\bar{t}H})\cos(\theta_{W^-}^{t})$	**3:**	$\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{W_\alpha}^{H})$
4:	$\cos(\theta_{t_\beta}^{Ht_\beta})\sin(\theta_{\bar{b}_H}^{t_\beta})$	**5:**	$\cos(\theta_{t_\beta}^{Ht_\beta})\sin(\theta_{b_H}^{t_\beta})$	**6:**	$\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{b_H}^{H})$
7:	$\sin(\theta_{H}^{t\bar{t}H})\cos(\theta_{b_H}^{t_\alpha})$	**8:**	$\sin(\theta_{H}^{t\bar{t}H})\cos(\theta_{b_H}^{\bar{t}})$	**9:**	$\sin(\theta_{t}^{t\bar{t}H})\sin(\theta_{\bar{b}_H}^{\bar{t}})$
10:	$\sin(\theta_{\bar{t}}^{t\bar{t}H})\sin(\theta_{b_H}^{t})$	**11:**	$\sin(\theta_{\bar{t}}^{t\bar{t}H})\sin(\theta_{\bar{b}_H}^{t})$	**12:**	$\sin(\theta_{H}^{t\bar{t}H})\cos(\theta_{\bar{b}_H}^{t})$
13:	$\sin(\theta_{t_\alpha}^{t\bar{t}H})\sin(\theta_{b_\beta}^{H})$	**14:**	$m_{b\bar{b}}$	**15:**	$\theta_{W^+}^{H}$

124

The procedure for constructing the 'Angular' set showed that a large number of discriminant observables can be chosen while requiring that correlations are not too high between any two of them, at least up to linear order. The comparison between distributions from the 'ATLAS' set in figure 41, and the 'Angular' set in figure 42, leads to the conclusion that most of the individual angular observables used show greater separation of $t\bar{t}H$ and $t\bar{t}b\bar{b}$ than the best observables in the 'ATLAS' set, with the exception of the first 'ATLAS' variable, $\Delta\eta_{jj}^{\max\Delta\eta}$.

The TMVA package provides information about the correlation between pairs of variables. Figure 44 shows the linear correlation coefficients (in %) between every pair of variables in the comprehensive set, for signal and background events. The 'Angular' set was built under the condition that these coefficients must be $< 60\%$ in signal events. Although in background ($t\bar{t}b\bar{b}$ only) events some pairs of variables exceed this value, the majority of correlation coefficients within the 'Angular' set are not large. The same is true within the 'ATLAS' set. Even pairs resulting from picking one variable from each set result in small correlations, which is remarkable, considering that the sets were constructed independently.
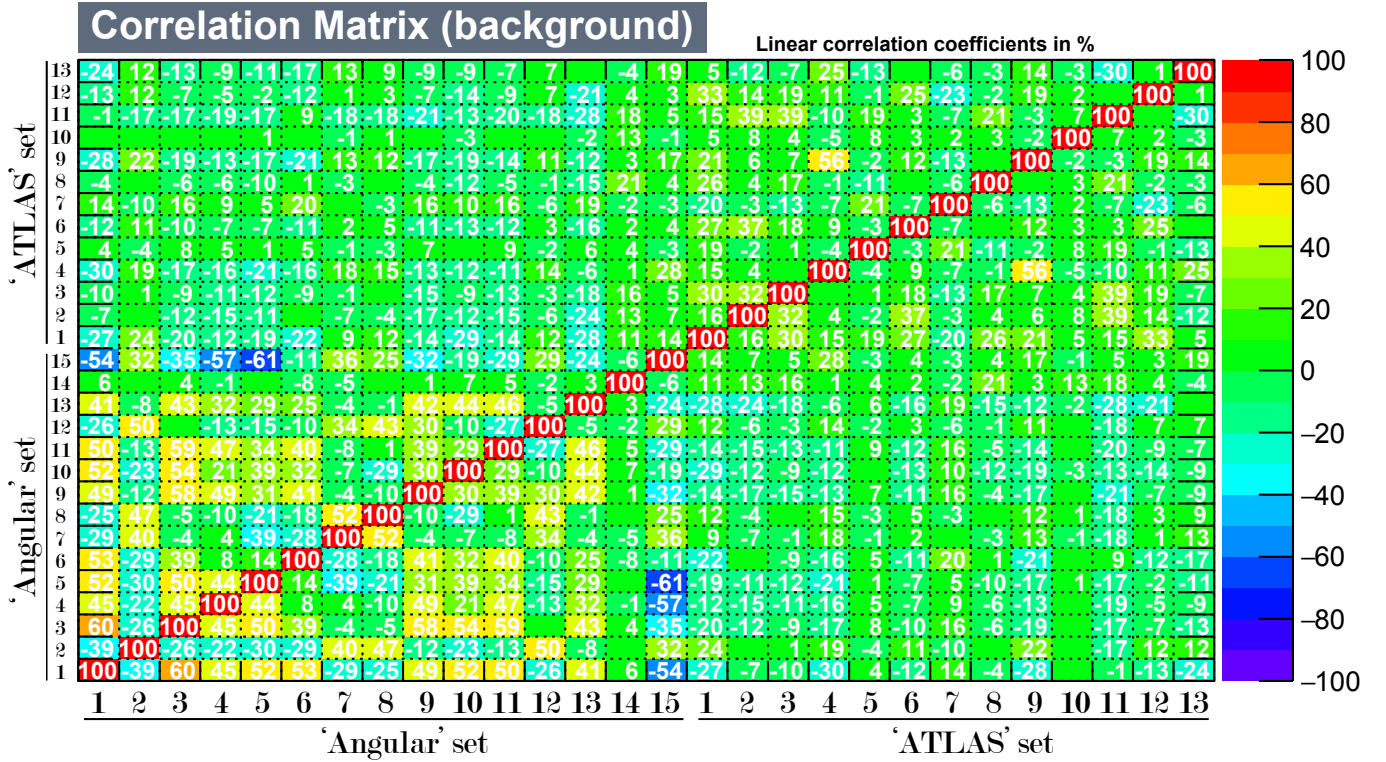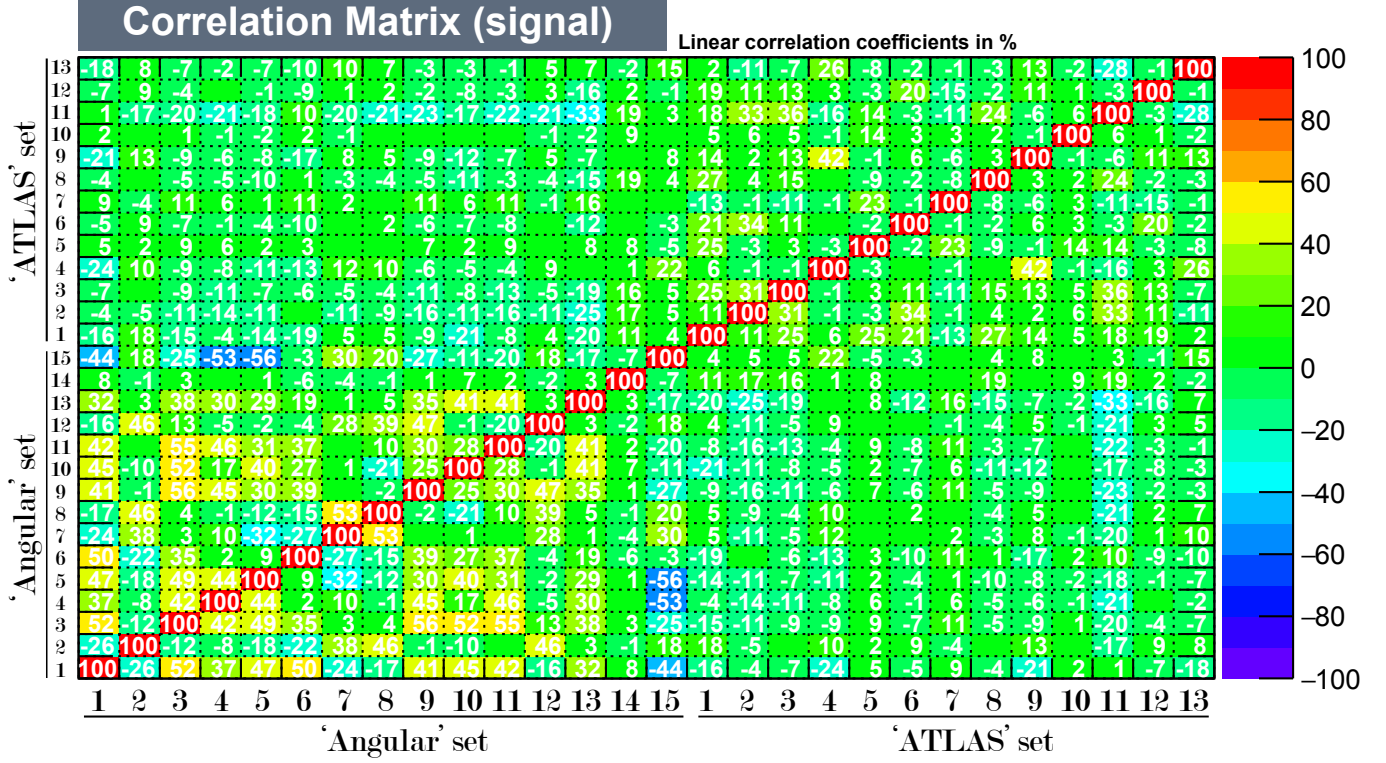
Figure 44: Linear correlation coefficients between every pair of variables in the comprehensive ('Angular'+'ATLAS') set. **Top:** Signal events. **Bottom:** Background ($t\bar{t}b\bar{b}$ only) events. The subsets were built independently in such a way that high correlations between variables are avoided within each one of them. Remarkably, the resulting correlation coefficients between variables of one subset and the other are also small.

126

In the multivariate analyses implemented, a fraction of the events in each sample is used as the training sample and the remaining events are used as the test sample. The training samples serve as input for the multivariate method to find, within the multidimensional space of the variables used, the features that better discriminate between signal and background. The phenomenon of overtraining is said to occur whenever the multivariate method 'picks up' features of the training sample that are not actually a feature of the global process. Overtraining results in worse discrimination performance of the analysis and can be diagnosed if the discrimination in the training sample is much better than the discrimination in the test sample. In order to avoid overtraining, the number of events in the training sample must be larger for analyses with a larger number of variables.

The result of a TMVA-based analysis is a discriminant variable for each one of the methods enabled. Then, events can be selected as signal or background by cutting on the obtained variable. Alternatively, the distributions of the discriminant variable could be used to perform a fit to data, allowing the extraction of a signal. In the case of a selection cut, the choice of the particular value where the cut is performed results in a certain signal efficiency ϵ and a certain background rejection rate r.

In order to compare the different methods and different analyses, the cut-values can be hidden altogether, and the background rejection rate can be plotted as a function of the signal efficiency. The TMVA package outputs such a plot for each one of the methods enabled in a multivariate analysis. In general terms, a discriminant variable is better if the area below the $r(\epsilon)$ curve is larger.

The following figures show the $r(\epsilon)$ curves from the four analyses performed, and the distribution of the best discriminant variable in each one. For all the analyses, the best dicriminant method was one of the BDT methods. These distributions show both the training sample and the test sample, in order to monitor the possibility of overtraining.

- Figure 45 shows the $r(\epsilon)$ curves and the best BDT discriminant distribution for the analysis using the 'ATLAS' variable set;

- Figure 46 shows the $r(\epsilon)$ curves and the best BDT discriminant distribution for the analysis using the 'Angular' variable set;
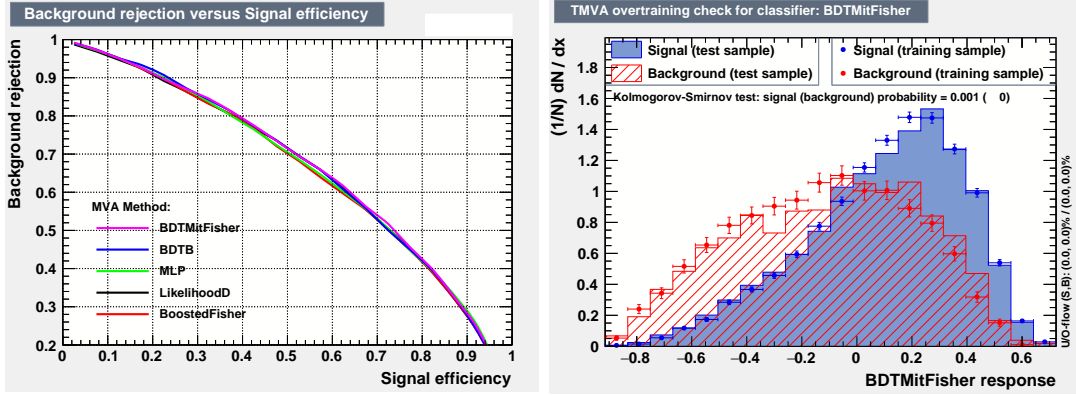
Figure 45: Results of the TMVA analysis with the 'ATLAS' set of variables. **Left:** $r(\epsilon)$ curve. **Right:** Distribution of the discriminant from one of the BDT methods with best results.
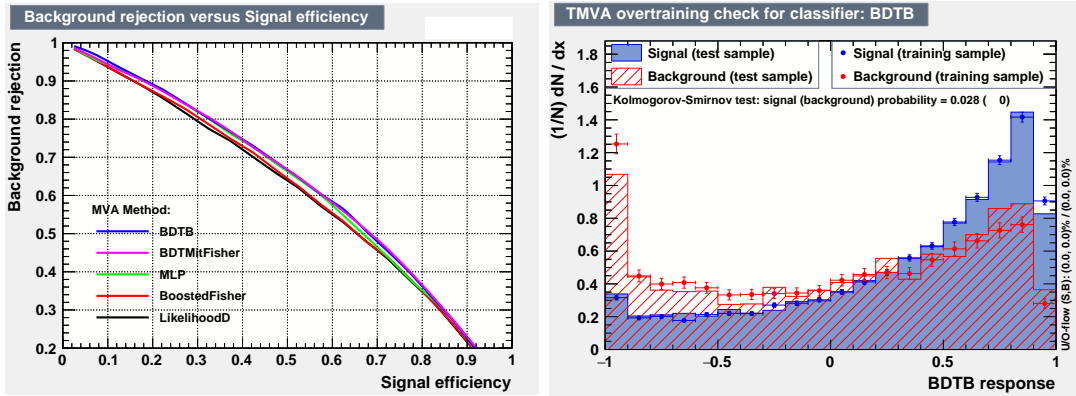
Figure 46: Results of the TMVA analysis with the 'Angular' set of variables. **Left:** $r(\epsilon)$ curve. **Right:** Distribution of the discriminant from one of the BDT methods with best results.

- Figure 47 shows the $r(\epsilon)$ curves and the best BDT discriminant distribution for the analysis using the 'Truth-matched angular' variable set;

- Figure 48 shows the $r(\epsilon)$ curves and the best BDT discriminant distribution for the analysis using the comprehensive set of variables.

- Figure 49 shows the $r(\epsilon)$ curves and the best BDT discriminant distribution for the validation analysis, which was run on the full set of generated backgrounds, using the comprehensive set of variables.

The better discrimination visible in the individual variables of the 'Angular' set with respect to the ones in the 'ATLAS' set does not seem to be reflected in the results of the multivariate analyses. The analysis using the 'Angular' set
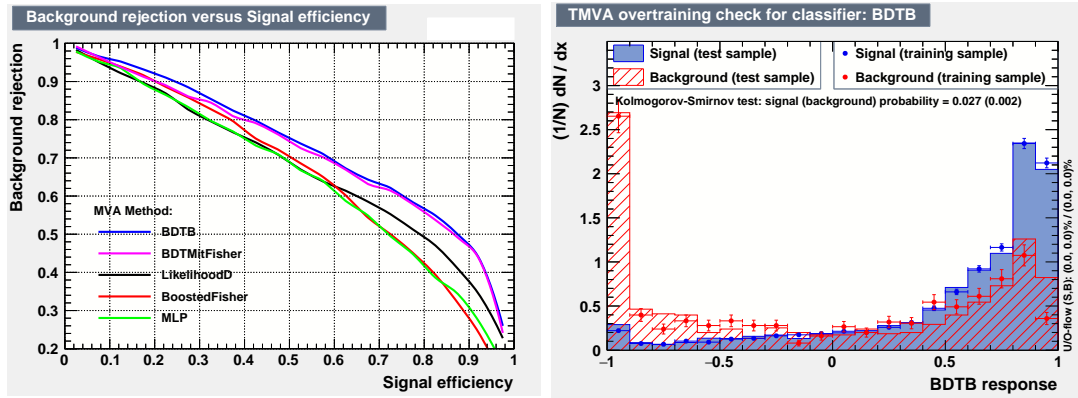
Figure 47: Results of the TMVA analysis with the 'Truth-matched angular' set of variables. **Left:** $r(\epsilon)$ curve. **Right:** Distribution of the discriminant from one of the BDT methods with best results.
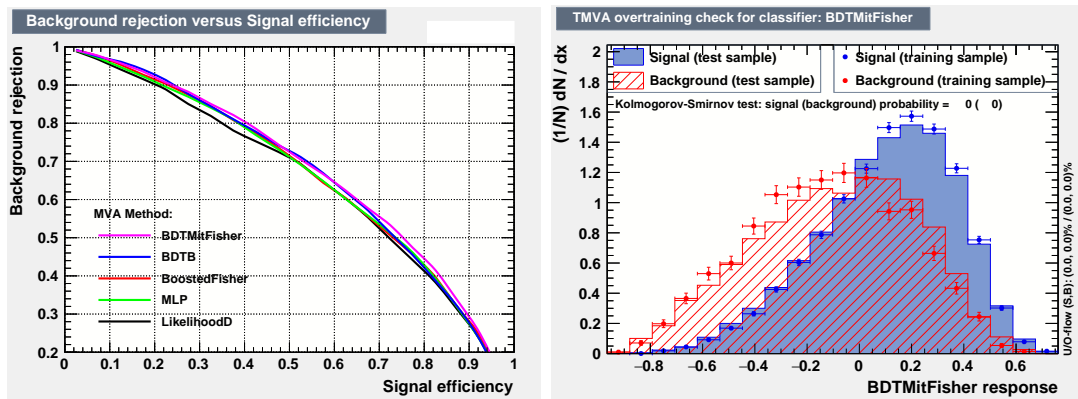
Figure 48: Results of the TMVA analysis with the comprehensive set of variables. **Left:** $r(\epsilon)$ curve. **Right:** Distribution of the discriminant from one of the BDT methods with best results.
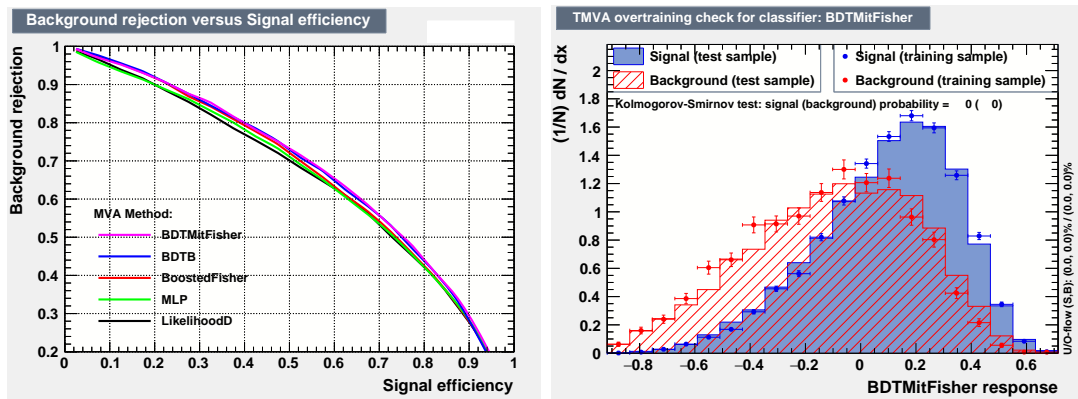
Figure 49: Results of the TMVA analysis performed on a full set of backgrounds, with the comprehensive set of variables. The choice of primarily focusing on the $t\bar{t}b\bar{b}$ background alone is validated by these results. **Left:** $r(\epsilon)$ curve. **Right:** Distribution of the discriminant from one of the BDT methods with best results.

129

alone performed slightly worse than the analysis using the 'ATLAS' set alone. With the two sets combined into one comprehensive set, some improvement in the performance of the analysis is visible, comparatively to using the 'ATLAS' set alone. One straightforward conclusion is that the 'ATLAS' set of observables contains more discriminating information than the 'Angular' set, which can be explained from the correlations within the 'Angular' set, which are generally larger than within the 'ATLAS' set.

Given these results, the very slight improvement in the analysis performance from including the angular observables should be put in perspective, with a comment regarding the difficulty in extracting a signal in the addressed channel. On the left side of figure 34, the signal-to-background ratio (S/B) and signal significance ($S/\sqrt{B}$, where $\sqrt{B}$ is the typical size of background-only statistical fluctuations) are shown for the different regions considered in the ATLAS analysis. For both the signal-rich regions, the signal significance is close to 0.4. Since this significance scales as $\sqrt{L}$, where L is the integrated luminosity, the observation of a $t\bar{t}H$ signal in each of these regions with enough significance to claim a discovery (5σ) would require $\sim (5/0.4)^2 \sim 150$ times more data than that collected at $\sqrt{s} = 8\,\text{TeV}$ during the LHC Run 1. This number reveals the difficulty of this search, the necessity of combining analyses of different decay channels, and also the importance of a gain in signal significance, even if a small one.

For this TMVA study, only events with at least one kinematic fit solution were considered. This requirement is in fact an additional cut, and must be considered together with the information in the $r(\epsilon)$ curve. At least one solution is found for 74% of signal events, and for 69% of background ($t\bar{t}b\bar{b}$ only) events, resulting in a 7.2% gain in signal-to-background ratio from making this requirement. While the computation of the new angular observables requires a successful kinematic fit, this is not the case for the ATLAS variables. If it is assumed that the performance of the ATLAS analysis is the same for the events rejected by this selection, paying the efficiency price may or may not be desirable, even though only 26% of signal events are lost. Selecting events with a kinematic fit solution is interesting because it provides access to many observables that are not otherwise available. The loss in efficiency must then be weighted against three advantages: the 7.2% gain in signal-to-background ratio from selecting events with at least one kinematic fit

solution; the possible gain in signal sensitivity from the otherwise unavailable observables (which, at the moment, is not yet significant, as can be concluded from comparing the TMVA runs using the 'ATLAS' set and the comprehensive set); and the physics results that can be extracted from new observables besides the signal yield, for example regarding the CP nature of the Higgs boson.

One possible explanation for the superior performance of the 'ATLAS' set is that its discrimination ability cannot be estimated from the discrimination of the individual distributions. Indeed, the procedure used by ATLAS for choosing the variables on which the $\sqrt{s} = 8\,\mathrm{TeV}$ analysis relied, was simply to let the multivariate method itself rank the variables. This led to a set of variables which is intrinsically multivariate, so it is possible that important discriminant features arise only in higher-dimensional distributions. On the other hand, the variables in the 'Angular' set were picked one by one, considering only the individual K-S test results and that linear correlation coefficients in signal events were not too large ($\geq 60\%$).

Considering this interpretation, a different method for constructing a set of angular observables to be used in the multivariate analysis is in order. The most obvious one would be to let the multivariate tool rank the best observables among all the ones available, including angular observables and the observables in the 'ATLAS' set. The ranking procedure is expected to avoid unwanted correlations and may find discrimination in combinations of observables, that would not qualify as good discriminants if they were classified individually.

Reconstructing the $t\bar{t}$ system enables the definition of new observables, that may be necessary to obtain better signal-background discrimination. Angular observables show some potential for this purpose, as can be seen in the results of the TMVA analysis using the 'Truth-matched angular' set, in figure 47. This set yields much better signal-background discrimination than any other of the sets used. Since they are defined with truth-matched objects, the variables in the 'Truth-matched angular' set are the same as the ones in the 'Angular' set, given that the jet assignment is done correctly at the reconstruction level without truth-match. In other words, the result of the TMVA analysis using the 'Truth-matched angular' set represents the potential gain in discrimination using the 'Angular' set alone, if the combinatorial background could be entirely removed.

Table 14: Expected number of events passing the signal-rich region cuts in $20\,\mathrm{fb}^{-1}$ of data, for the different generated backgrounds. Cross-sections used to obtain these numbers are the same as in table 6. Uncertainties are statistical only.

Process	N (signal-rich region)
$t\bar{t}H$ (SM)	6.34 ± 0.04
$t\bar{t}b\bar{b}$	68.3 ± 0.8
$t\bar{t} + jets$	115 ± 6
$t\bar{t}V + jets$	2.19 ± 0.15
Single t	2.48 ± 1.24
Total	194 ± 6

Completely avoiding combinatorial background is virtually impossible in a process with several jets from different sources in the final state. However, it appears from these results that even a partial reduction of wrong jet assignments could significantly improve the performance of the 'Angular' set, and consequently the performance of a comprehensive analysis.

The TMVA run in which a full set of backgrounds was used (figure 49) performed better than the run with $t\bar{t}b\bar{b}$ as the only background (figure 48). This is expected from the irreducible nature of the $t\bar{t}b\bar{b}$ background. Since $t\bar{t}b\bar{b}$ is the most signal-like background, any additional background component will be easier to discriminate from the signal. This result validates the initial choice of focusing on the discrimination between $t\bar{t}H$ and $t\bar{t}b\bar{b}$. In table 14, the expected number of events surviving the signal-rich region cuts ($\geq 4j, \geq 3b, |m_{\ell\ell} - m_Z| > 10\,\mathrm{GeV}$) and the requirement of a kinematic fit solution are shown, for the SM $t\bar{t}H$ signal and for the generated SM backgrounds, in the case of a data sample with $20\,\mathrm{fb}^{-1}$ of integrated luminosity. The cross-sections used for obtaining the numbers are the ones in table 6. The background processes $W + jets$, $Wb\bar{b} + jets$, $Z + jets$, $Zb\bar{b} + jets$ and VV ($V = W, Z$) are not represented since they yielded 0 events after the selection.

Distributions of 'fake' data were generated using distributions of the full SM simulation (signal plus background) as the p.d.f.s for the random generation, and

the total number of generated 'fake' entries corresponds to the expected number of total SM (signal plus background) events in $20\,\mathrm{fb}^{-1}$ of data. In figure 50, four such distributions are shown, with the 'fake' data distributions overlapped with the different backgrounds. The purpose of these distributions is to illustrate the expected data distributions after a $20\,\mathrm{fb}^{-1}$ data-taking at the LHC, and the difficulty in extracting a signal from such a sample.
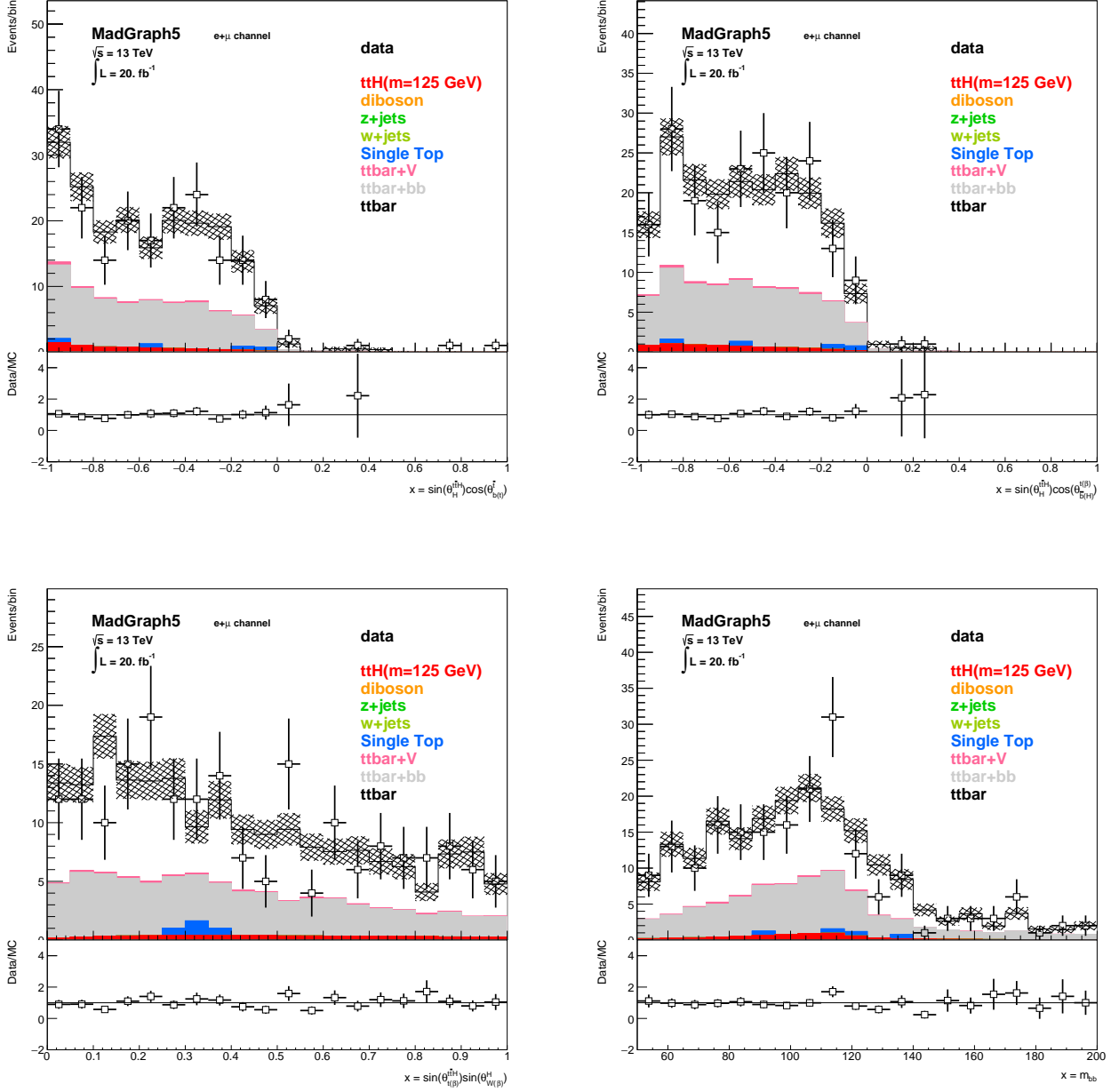
Figure 50: Distributions of angular observables and reconstructed mass of the Higgs candidate ($b\bar{b}$ system) for the different samples generated, compared with the generated 'fake' data distributions. Number of events were obtained after the signal-rich region cuts, and considering an integrated luminosity of $20\,\mathrm{fb}^{-1}$.

Top left: $\sin(\theta_H^{t\bar{t}H})\cos(\theta_{b_t}^{\bar{t}})$ (same observable as on the left side of figure 37).

Top right: $\sin(\theta_H^{t\bar{t}H})\cos(\theta_{\bar{b}_H}^{t_\beta})$ (same observable as on the right side of figure 39).

Bottom left: $\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{W_\beta}^{H})$ (same observable as on the right side of figure 54).

Bottom right: $m_{b\bar{b}}$ (same observable as on the bottom right plot of figure 27).

4.2 Testing the CP nature of the Higgs boson

4.2.1 Angular observables from θ_Y^X angles

The Kolmogorov-Smirnov test was applied to all angular observables, comparing the $t\bar{t}H$ and $t\bar{t}A$ samples, at reconstruction level without truth-match. A good discriminant between the CP-even and CP-odd Higgs production should be robust with respect to background fluctuations, since the background dominates the total number of events in a data analysis of this channel.

The result of a search for a possible CP-odd Higgs component could be expressed as an upper limit to the ratio between $t\bar{t}A$ and $t\bar{t}H$ production. For the estimation of this ratio to remain as constant as possible with respect to background fluctuations, the $t\bar{t}b\bar{b}$ overlaps with $t\bar{t}H$ and $t\bar{t}A$ should be small and similar to each other. Such a discriminant could be used in a measurement of the CP-odd component of the Higgs boson, complemented with observables exhibiting good signal-background discrimination, independently of the Higgs CP nature (such as the ones described in section 4.1.1). The latter would fix the signal-to-background ratio model-independently, and the first would fix the $t\bar{t}H$-to-$t\bar{t}A$ ratio. For this reason, distributions of $t\bar{t}b\bar{b}$ events are also shown in the figures below.

Angular observables of the type $f(\theta_3^{23})g(\theta_4^3)$ exhibit poor discrimination of $t\bar{t}H$ and $t\bar{t}A$ events. Therefore, unlike in section 4.1.1, only two categories of observables are considered. Four examples of distributions among the best classified ones in each category are represented in the following figures. Distributions are shown at parton level and at reconstruction level, with and without truth-match. Only events which were eventually reconstructed without truth-match are considered for the distributions at parton-level. Also visible is the output of the Kolmogorov-Smirnov test between $t\bar{t}H$ and $t\bar{t}A$ samples.

- Figures 51 and 52 show 4 examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$ without α-β assignment;

- Figures 53 and 54 show 4 examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$ with α-β assignment;
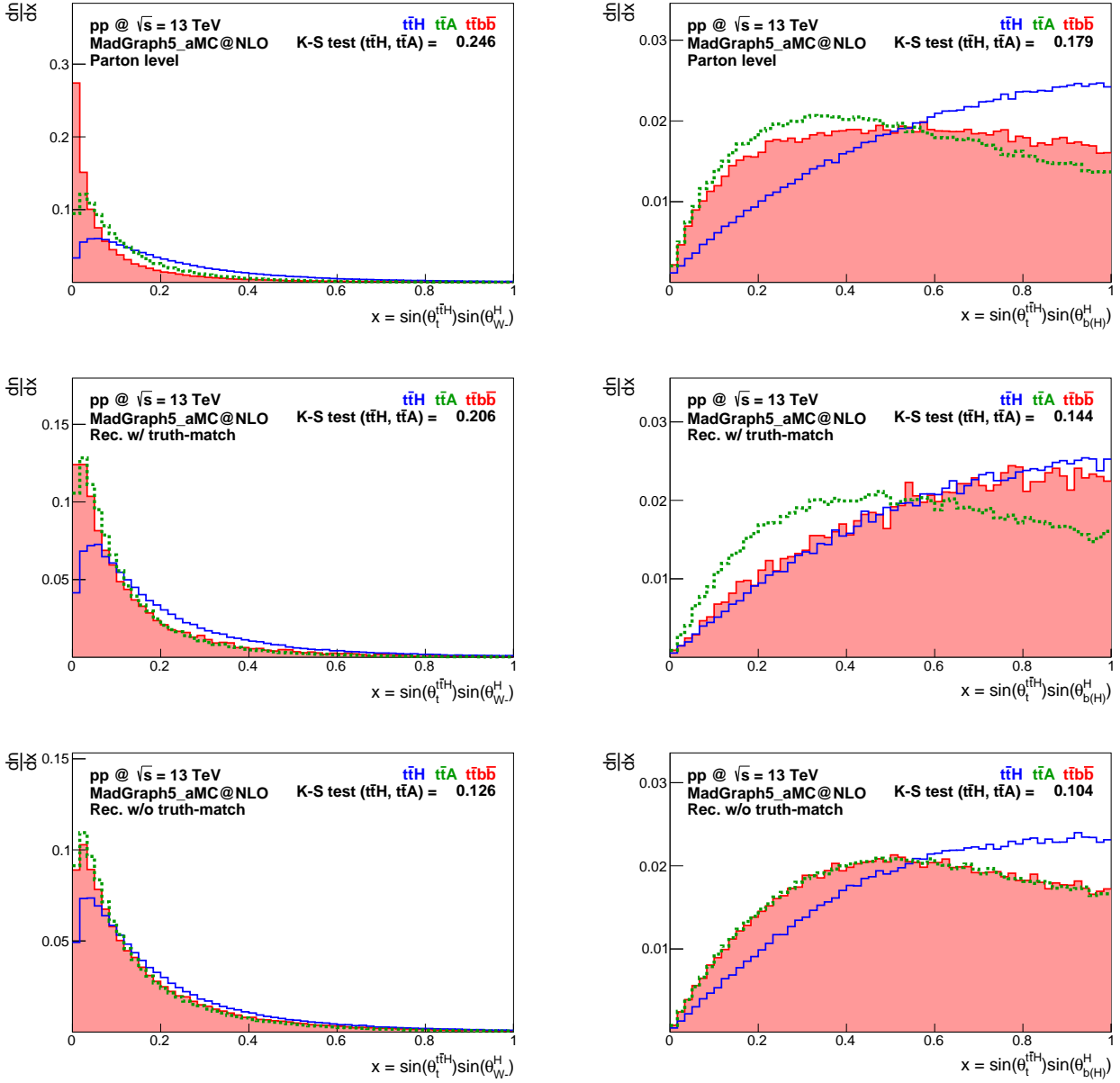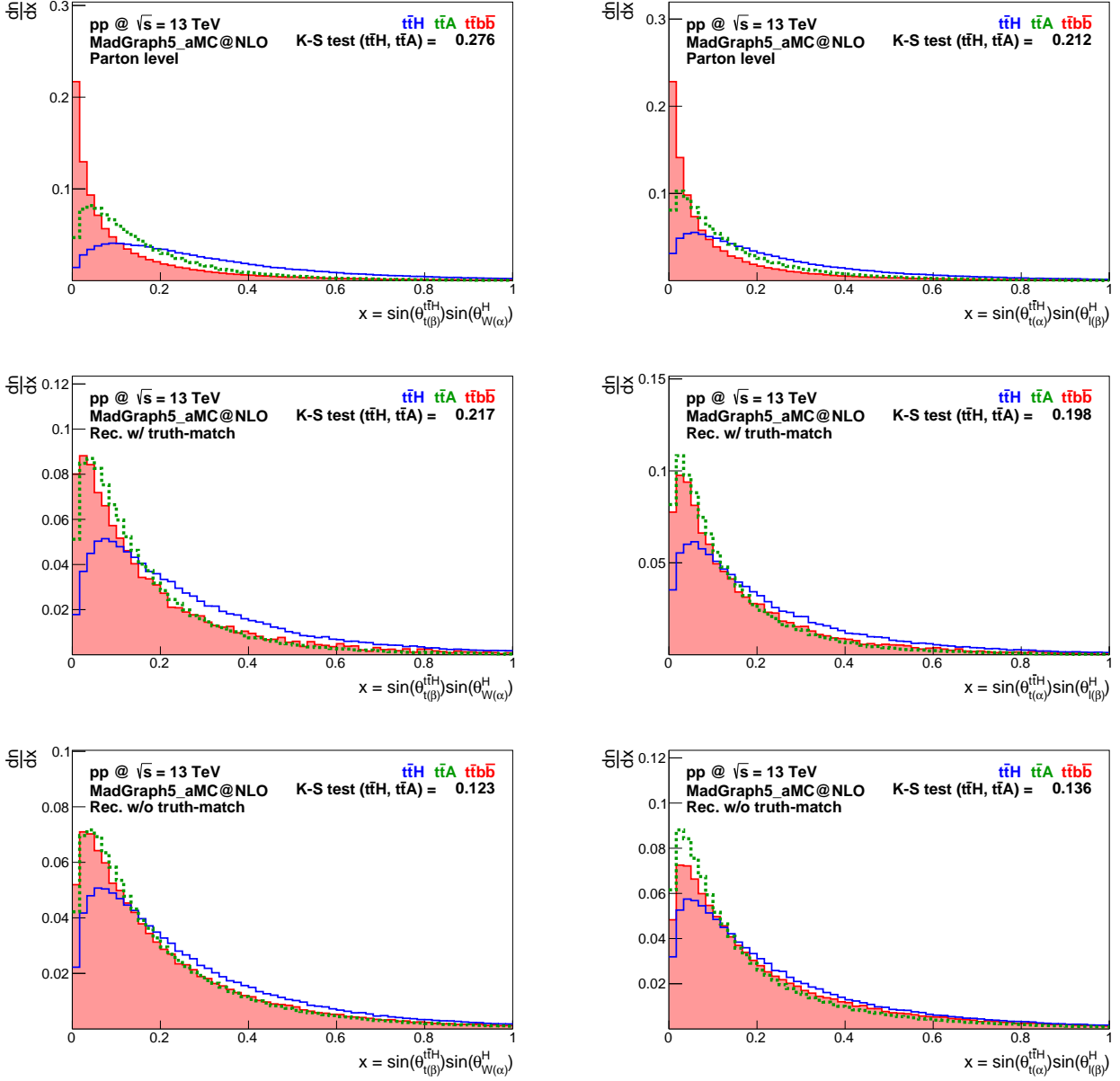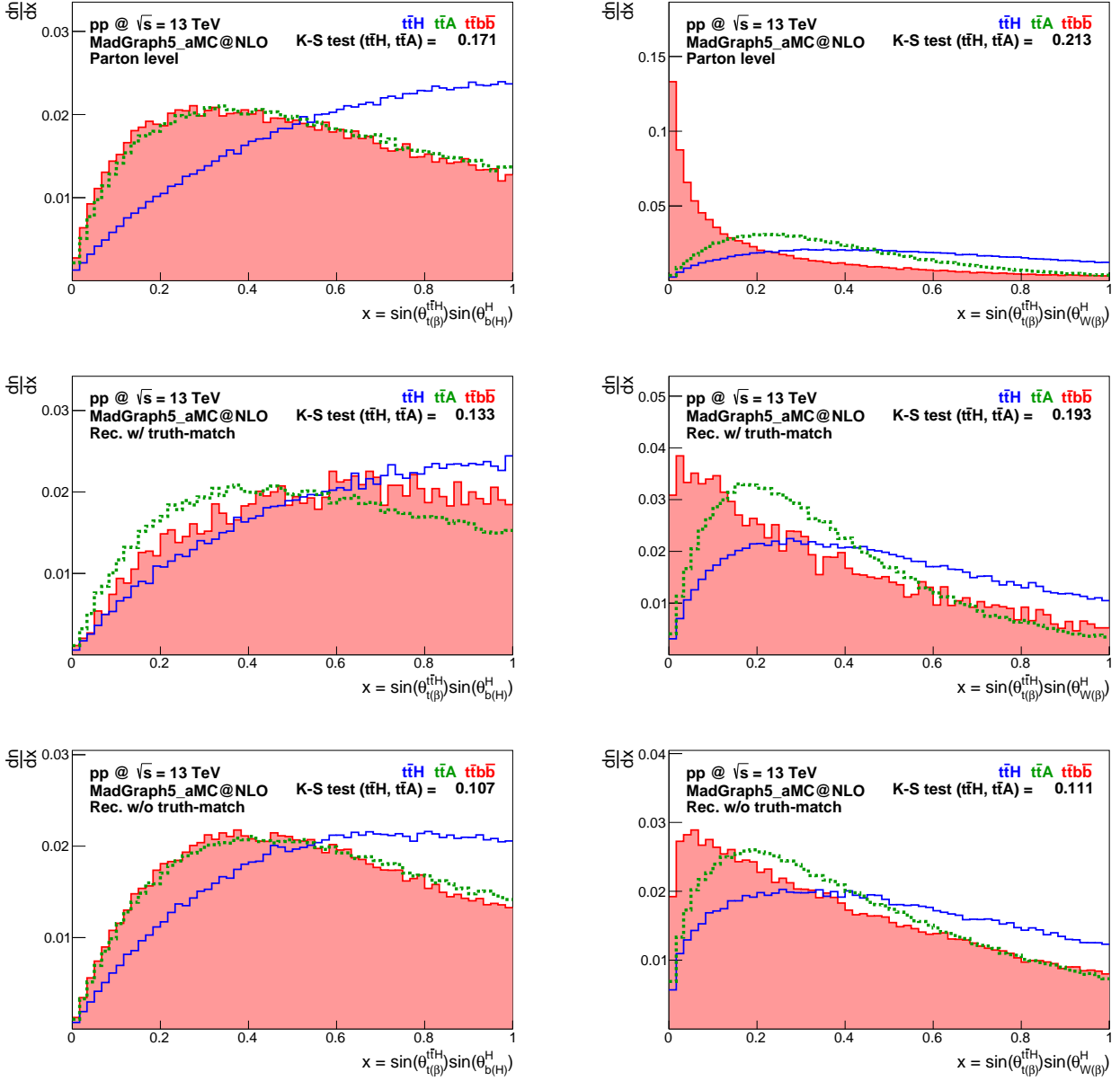
Figure 51: Two examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$, without α-β assignment, that performed well in the K-S test between $t\bar{t}H$ and $t\bar{t}A$ samples. Distributions are shown for $t\bar{t}H$, $t\bar{t}b\bar{b}$ and $t\bar{t}A$ events, at parton level (top), reconstruction level with truth-match (middle) and reconstruction level without truth-match (bottom). **Left:** $\sin(\theta_t^{t\bar{t}H})\sin(\theta_{W^-}^H)$. **Right:** $\sin(\theta_t^{t\bar{t}H})\sin(\theta_{b_H}^H)$.
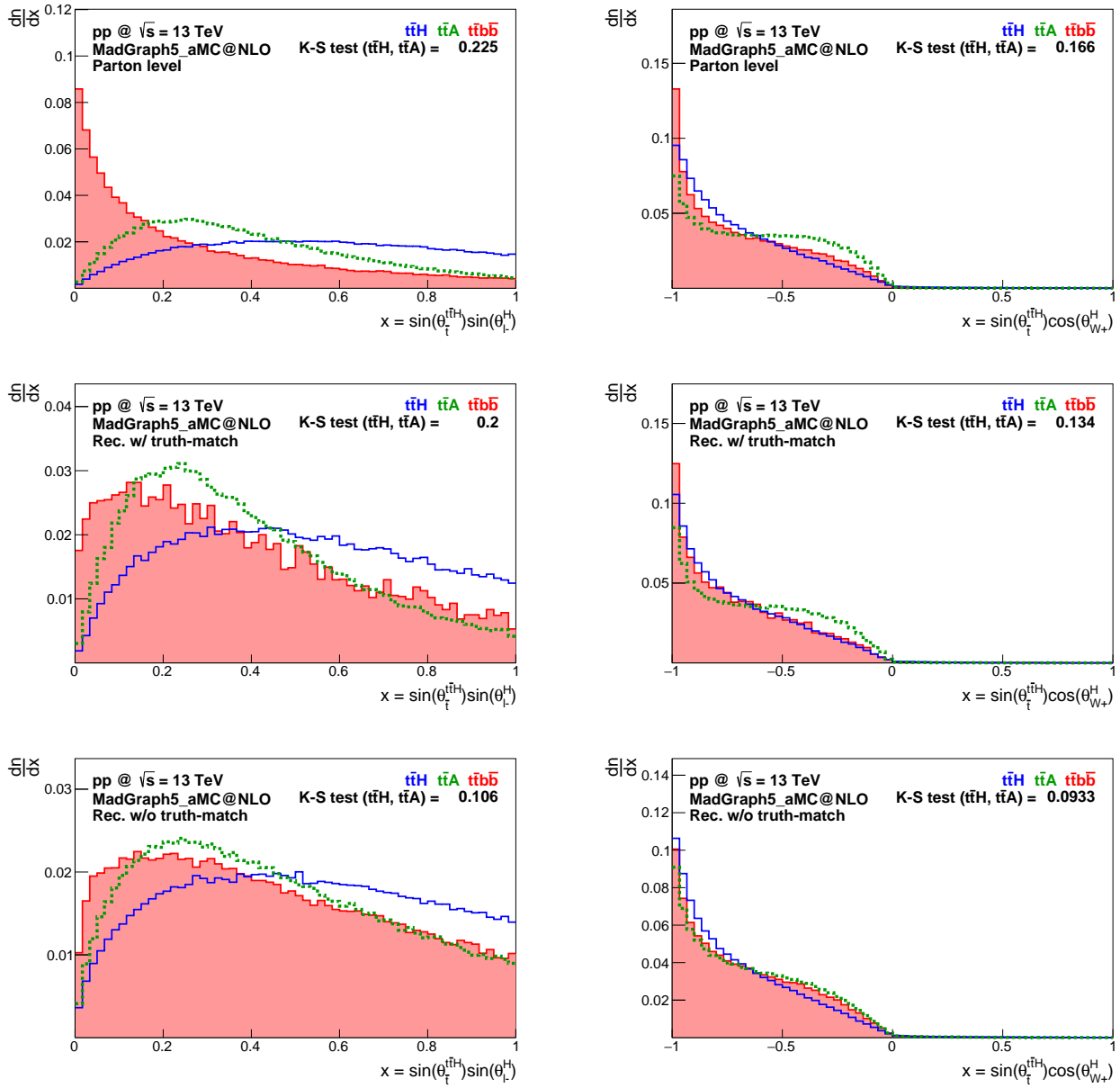
Figure 52: Two examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$, without α-β assignment, that performed well in the K-S test between $t\bar{t}H$ and $t\bar{t}A$ samples. Distributions are shown for $t\bar{t}H$, $t\bar{t}b\bar{b}$ and $t\bar{t}A$ events, at parton level (top), reconstruction level with truth-match (middle) and reconstruction level without truth-match (bottom). **Left:** $\sin(\theta_{\bar{t}}^{t\bar{t}H})\sin(\theta_{\ell^-}^H)$. **Right:** $\sin(\theta_{\bar{t}}^{t\bar{t}H})\cos(\theta_{W^+}^H)$.

Figure 53: Two examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$, with α-β assignment, that performed well in the K-S test between $t\bar{t}H$ and $t\bar{t}A$ samples. Distributions are shown for $t\bar{t}H$, $t\bar{t}b\bar{b}$ and $t\bar{t}A$ events, at parton level (top), reconstruction level with truth-match (middle) and reconstruction level without truth-match (bottom). **Left:** $\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{W_\alpha}^H)$. **Right:** $\sin(\theta_{t_\alpha}^{t\bar{t}H})\sin(\theta_{\ell_\beta}^H)$.

Figure 54: Two examples of observables of the type $f(\theta_1^{123})g(\theta_4^3)$, with α-β assignment, that performed well in the K-S test between $t\bar{t}H$ and $t\bar{t}A$ samples. Distributions are shown for $t\bar{t}H$, $t\bar{t}b\bar{b}$ and $t\bar{t}A$ events, at parton level (top), reconstruction level with truth-match (middle) and reconstruction level without truth-match (bottom). **Left:** $\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{b_H}^H)$. **Right:** $\sin(\theta_{t_\beta}^{t\bar{t}H})\sin(\theta_{W_\beta}^H)$.

4.2.2 Observables from the literature review

All the observables described in section 3.7 were studied. In the following figures, the distributions of $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$ samples are shown for each observable, at parton level and at reconstruction level, with and without truth-match. When available, the distributions originally published along with the proposal of the corresponding observable are also shown. Such distributions should be compared to the replicated parton-level distributions, since they don't include detection and reconstruction inefficiencies. Even at parton-level, the distributions are not expected to be equivalent, because the replicated distributions include the selection cuts described in section 3.4.1, unlike the originally proposed ones. The distributions at reconstruction level without truth-match demonstrate the expected discrimination between $t\bar{t}H$ and $t\bar{t}A$ events under experimental conditions.

- Figure 55 shows distributions of $\cos[\theta(\ell^+, \ell^-)]$, including the originally published distributions [63];

- Figures 56, 57 and 58 show distributions of $\Delta\phi^{t\bar{t}}(\ell^+, \ell^-)$, $\Delta\phi^H(\ell^+, \ell^-)$ and $\Delta\phi(t, \bar{t})$, respectively, including the originally published distributions [65];

- Figure 59 shows distributions of $\beta \cdot \theta^{\perp H}(\ell^+, \ell^-)$ and $\beta' \cdot \theta^{\perp H}(\ell^+, \ell^-)$, including the originally published distribution of $\beta \cdot \theta^{\perp H}(\ell^+, \ell^-)$ [14];

- Figures 60, 61 and 62 show distributions of the a_1, a_2, b_1, b_2, b_3 and b_4 observables proposed in [64].
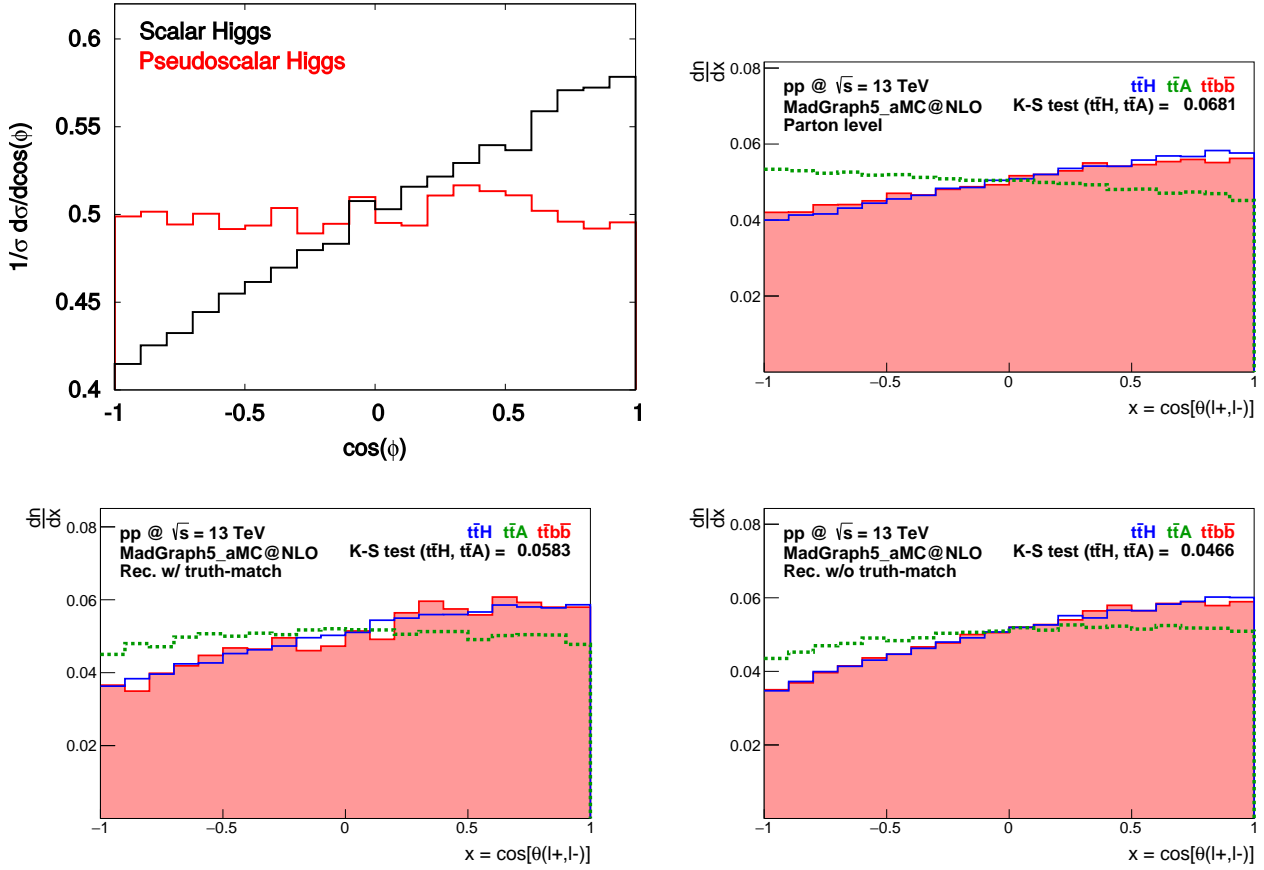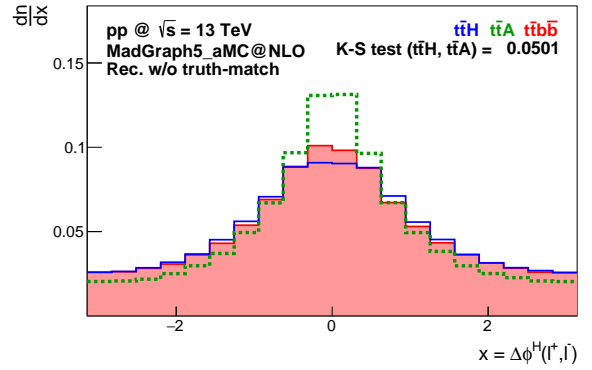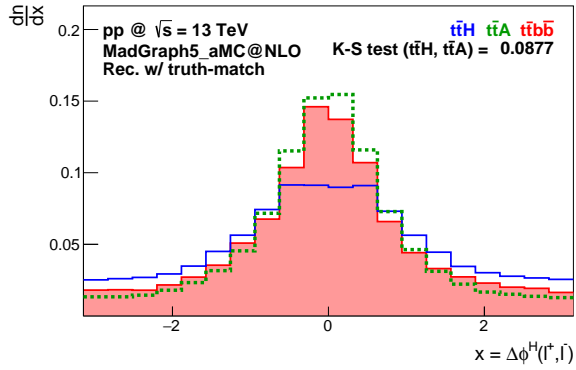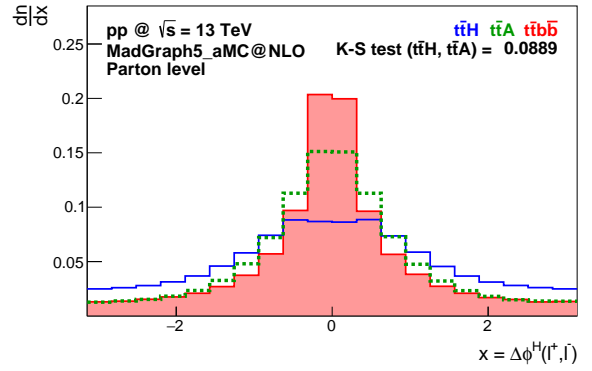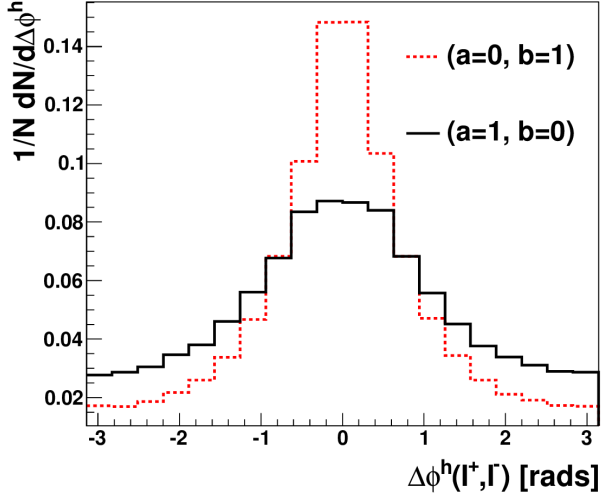
Figure 55: Distributions of $\cos[\theta(\ell^+, \ell^-)]$. **Top left:** Histograms published in [63], in which the observable is proposed as a discriminant. **Top right:** Distributions replicated in the present work at parton level, not only for $t\bar{t}H$ and $t\bar{t}A$ samples, but also for $t\bar{t}b\bar{b}$. There is reasonable agreement between the histograms. **Bottom left:** Same distributions at reconstruction level with truth-match. **Bottom right:** Same distributions at reconstruction level without truth-match.
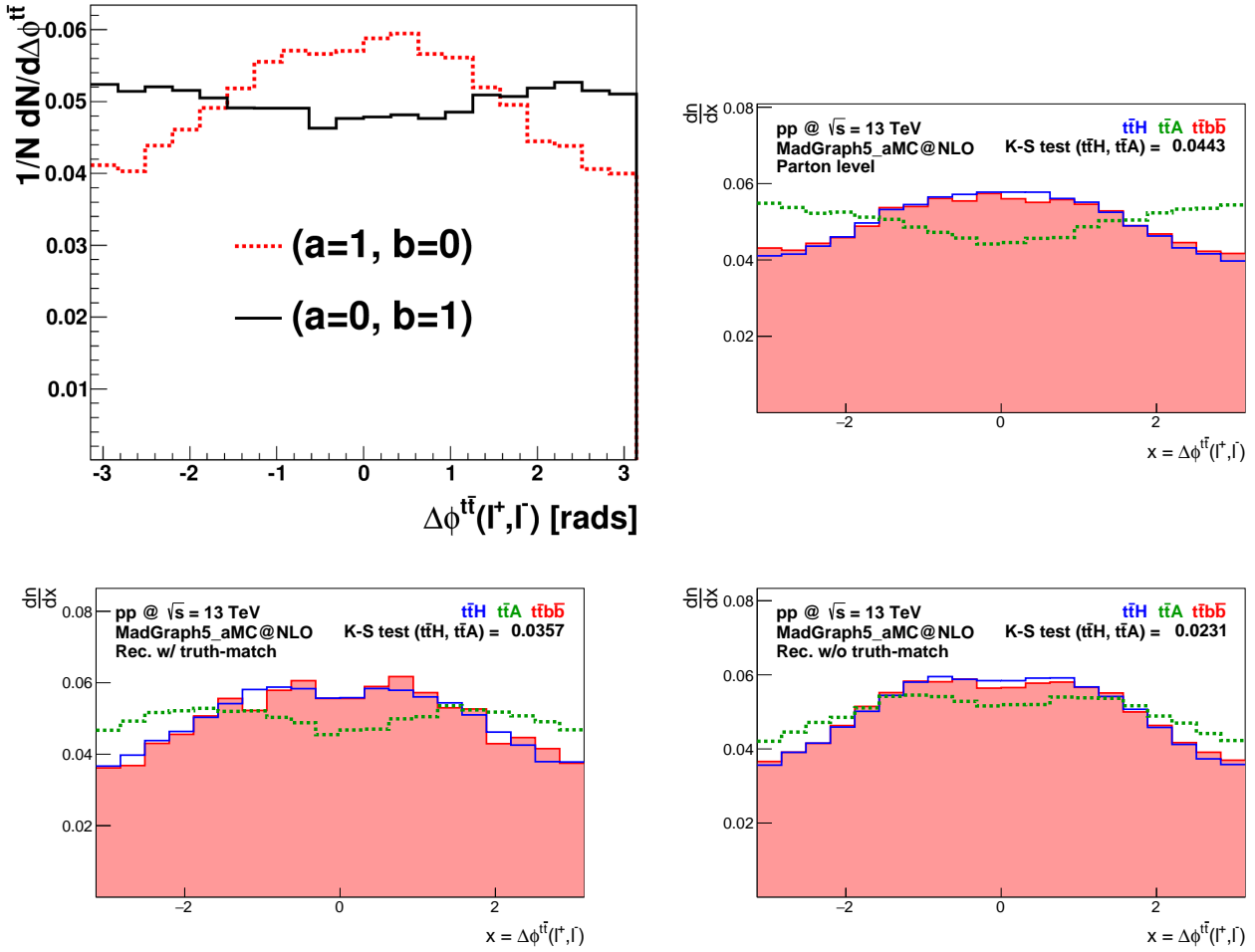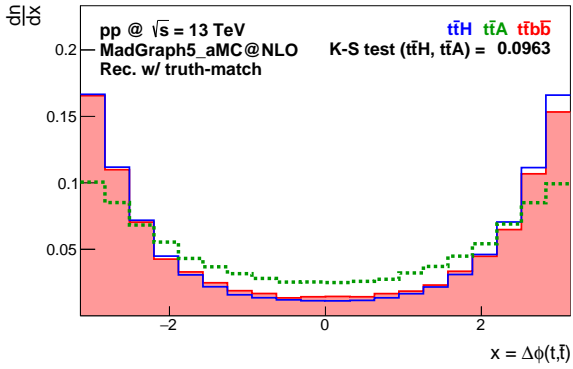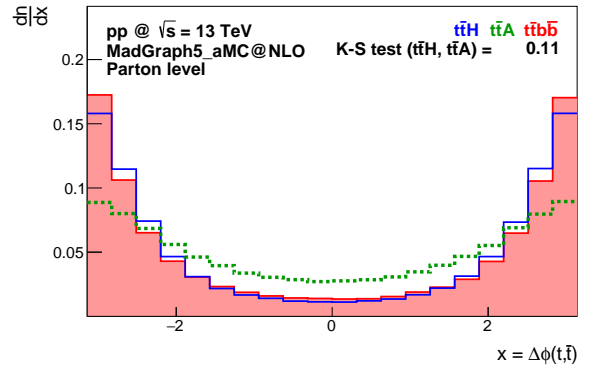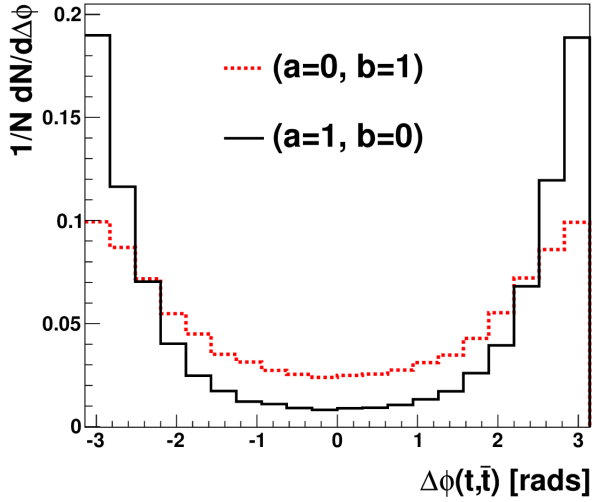
Figure 56: Distributions of $\Delta\phi^{t\bar{t}}(\ell^+, \ell^-)$. **Top left:** Histograms published in [65], in which the observable is proposed as a discriminant (a is the scalar component of the Higgs boson, and b is the pseudoscalar component). **Top right:** Distributions replicated in the present work at parton level, not only for $t\bar{t}H$ and $t\bar{t}A$ samples, but also for $t\bar{t}b\bar{b}$. The agreement between the histograms is good. **Bottom left:** Same distributions at reconstruction level with truth-match. **Bottom right:** At reconstruction level without truth-match, where a significant distortion of the original shapes of the histograms is visible.

Figure 57: Distributions of $\Delta\phi^H(\ell^+, \ell^-)$. **Top left:** Histograms published in [65], in which the observable is proposed as a discriminant (a is the scalar component of the Higgs boson, and b is the pseudoscalar component). **Top right:** Distributions replicated in the present work at parton level, not only for $t\bar{t}H$ and $t\bar{t}A$ samples, but also for $t\bar{t}b\bar{b}$. There is good agreement between the histograms. **Bottom left:** Same distributions at reconstruction level with truth-match. **Bottom right:** Same distributions at reconstruction level without truth-match.
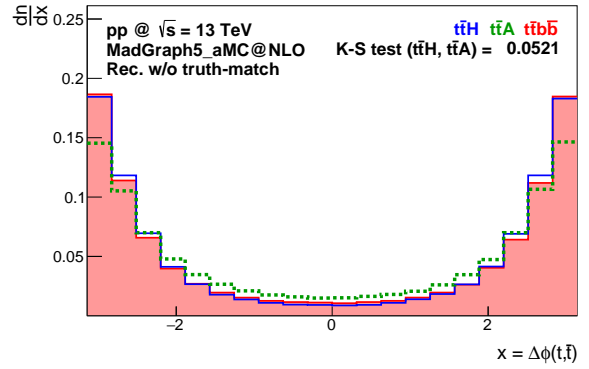
Figure 58: Distributions of $\Delta\phi(t,\bar{t})$. **Top left:** Histograms published in [65], in which the observable is proposed as a discriminant (a is the scalar component of the Higgs boson, and b is the pseudoscalar component). **Top right:** Distributions replicated in the present work at parton level, not only for $t\bar{t}H$ and $t\bar{t}A$ samples, but also for $t\bar{t}b\bar{b}$. There is good agreement between the histograms. **Bottom left:** Same distributions at reconstruction level with truth-match. **Bottom right:** At reconstruction level without truth-match, the distribution of the $t\bar{t}A$ sample is distorted and there is a significant loss of discrimination.
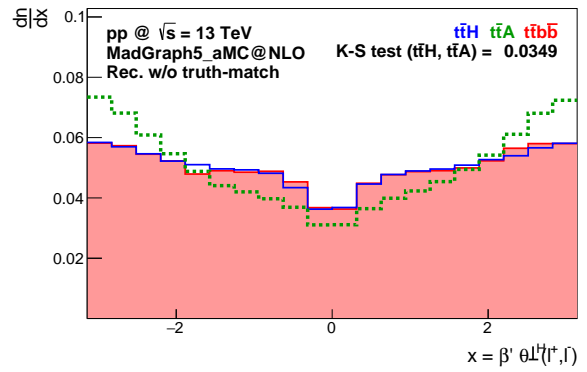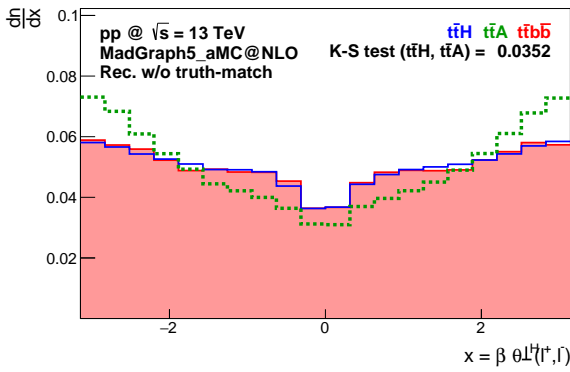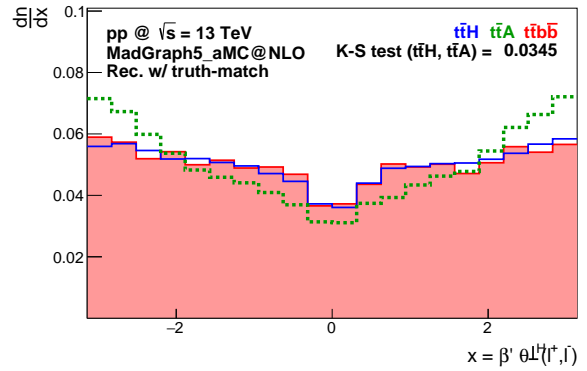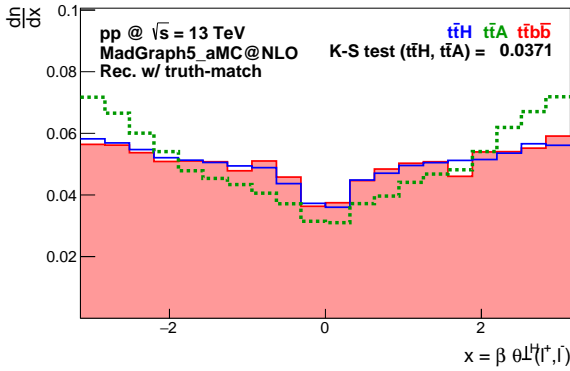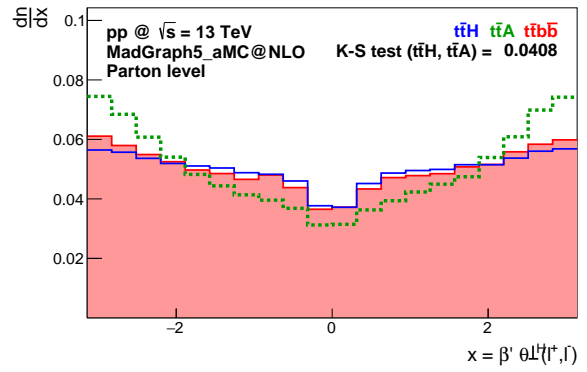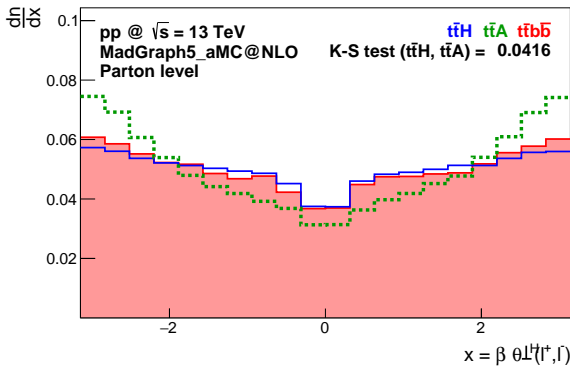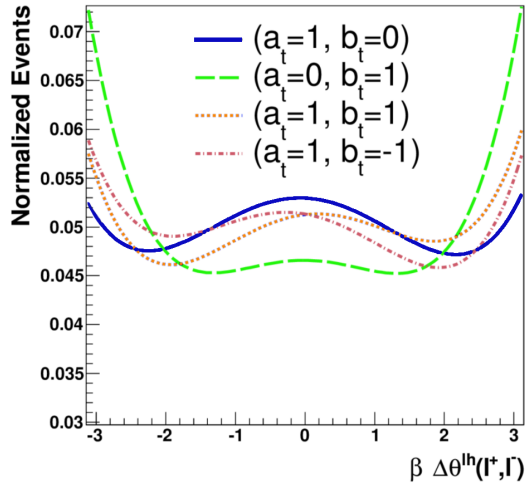
Figure 59: Distributions of $\beta \cdot \theta^{\perp H}(\ell^+, \ell^-)$ (left) and $\beta' \cdot \theta^{\perp H}(\ell^+, \ell^-)$ (right). **Top:** Distributions published in [14], in which the observable is proposed as a discriminant (a is the scalar component of the Higgs boson, and b is the pseudoscalar component). It was suggested that both $\beta \cdot \theta^{\perp H}(\ell^+, \ell^-)$ and $\beta' \cdot \theta^{\perp H}(\ell^+, \ell^-)$ should have similar distributions. **Upper middle:** Distributions replicated in the present work at parton level, not only for $t\bar{t}H$ and $t\bar{t}A$ samples, but also for $t\bar{t}b\bar{b}$. The application of selection cuts significantly distorts the distributions, but the $t\bar{t}H$ and $t\bar{t}A$ discrimination is present. **Lower middle:** Same distributions at reconstruction level with truth-match. **Bottom:** Reconstruction level without truth-match, at which the observables keep some degree of discrimination.
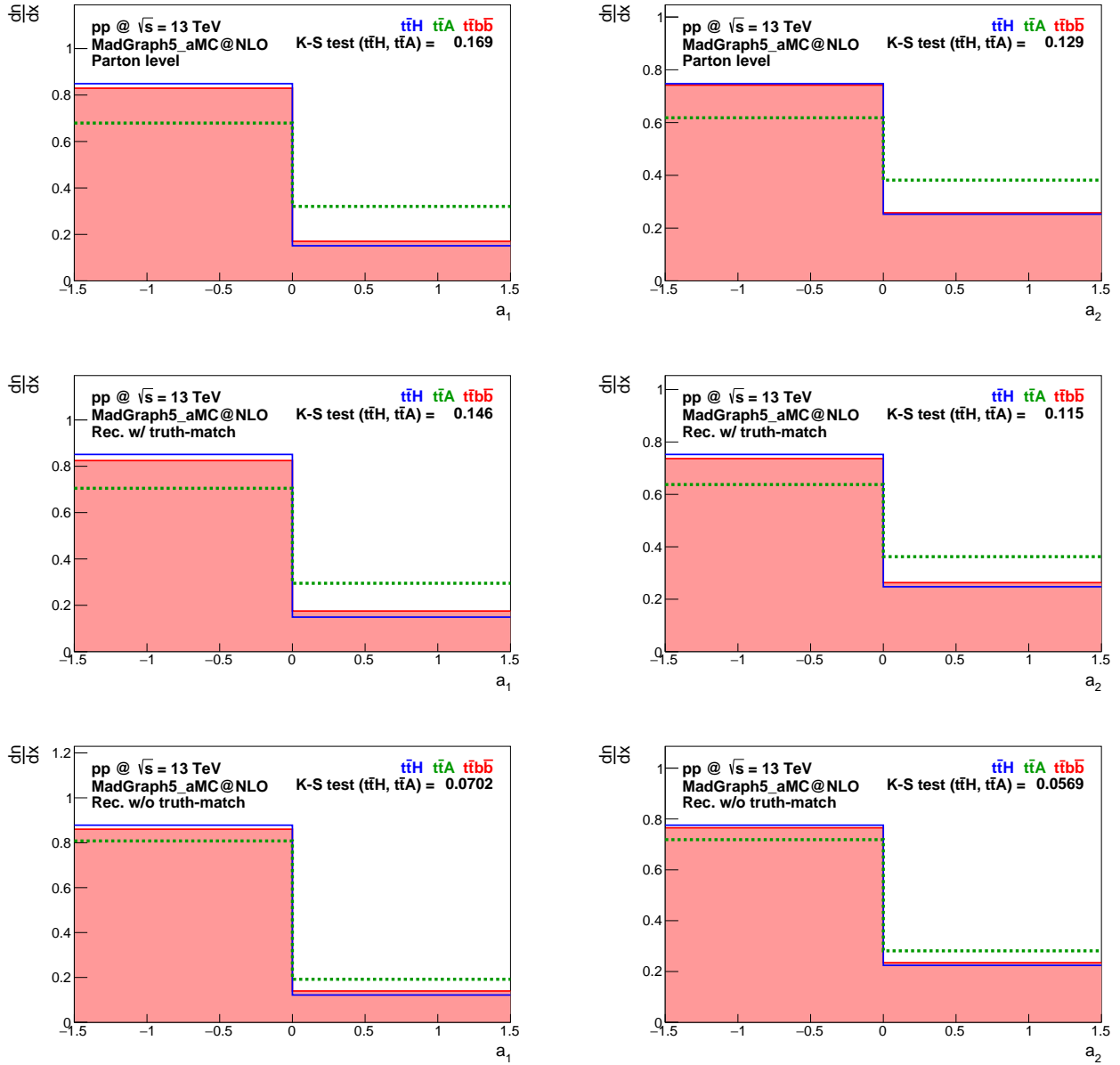
145

Figure 60: Distributions of a_1 (left) and a_2 (right), proposed as discriminants in [64]. In that proposal, there is no prediction of distribution shapes. **Top:** Distributions replicated in the present work at parton level, for $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$ samples. **Middle:** Same distributions at reconstruction level with truth-match. **Bottom:** Reconstruction level without truth-match.

Figure 61: Distributions of b_1 (left) and b_2 (right), proposed as discriminants in [64]. In that proposal, there is no prediction of distribution shapes. **Top:** Distributions replicated in the present work at parton level, for $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$ samples. **Middle:** Same distributions at reconstruction level with truth-match. **Bottom:** Reconstruction level without truth-match.
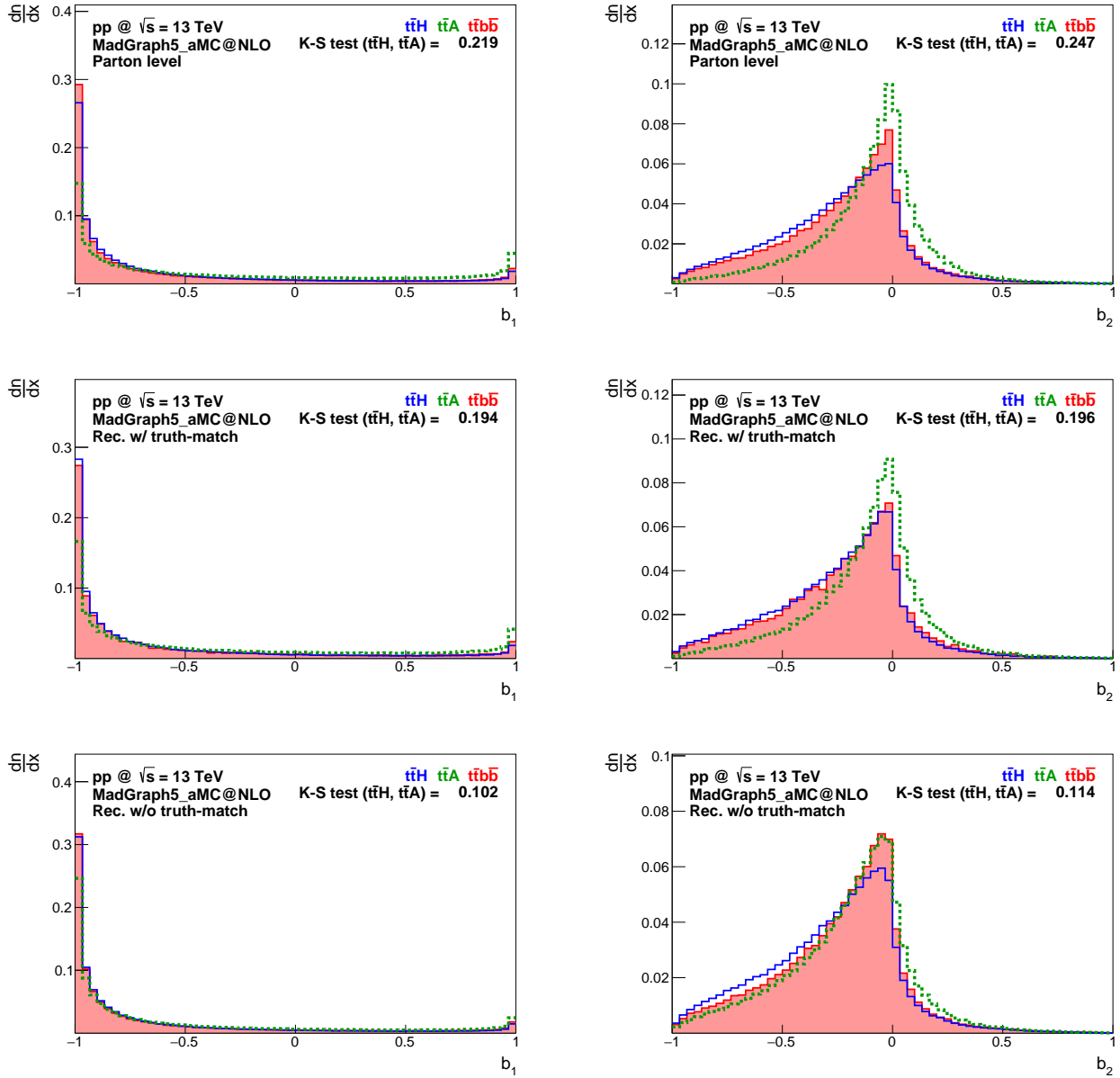
Figure 62: Distributions of b_3 (left) and b_4 (right), proposed as discriminants in [64]. In that proposal, there is no prediction of distribution shapes. **Top:** Distributions replicated in the present work at parton level, for $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$ samples. **Middle:** Same distributions at reconstruction level with truth-match. **Bottom:** Rec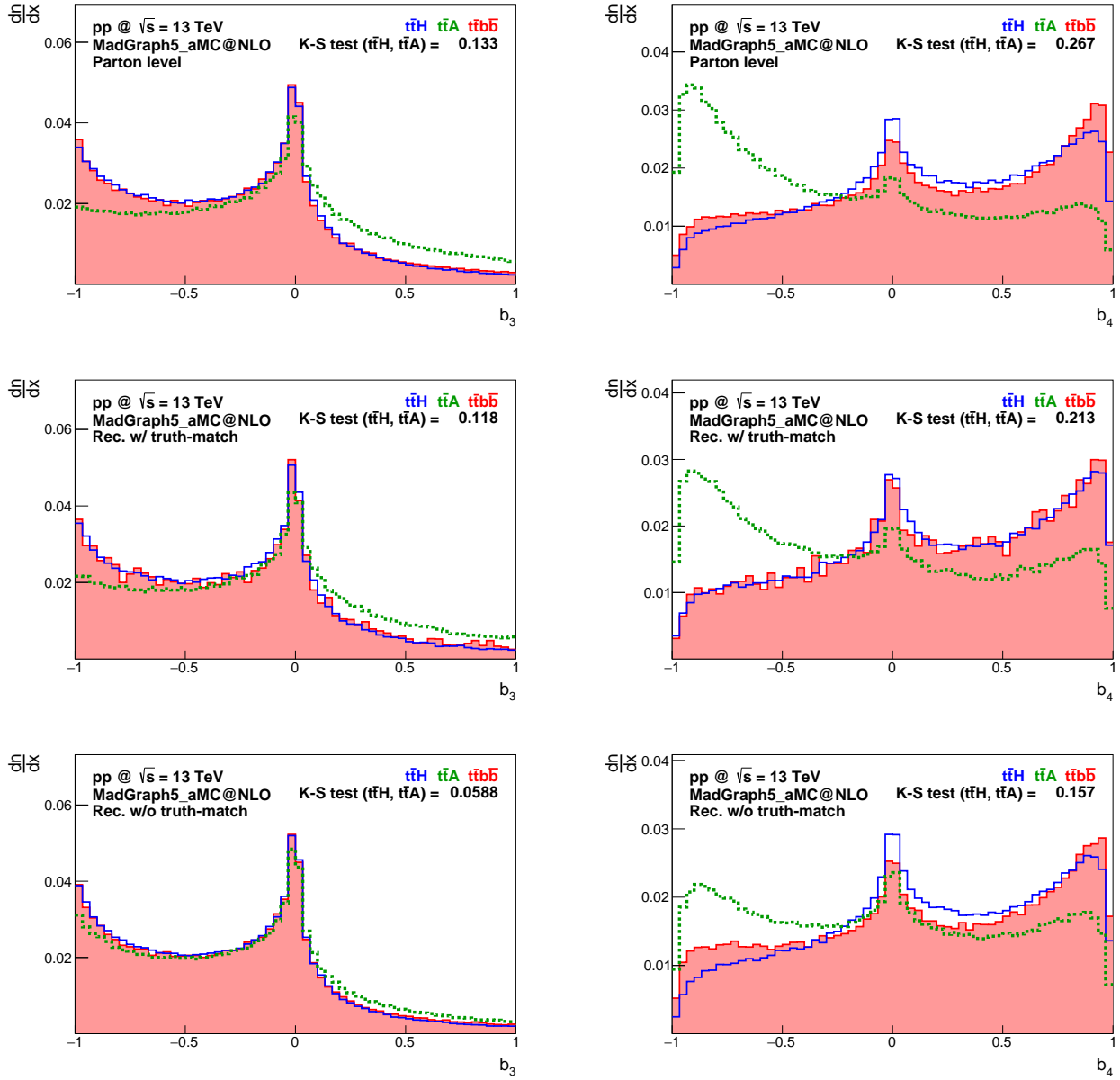onstruction level without truth-match. The b_4 observable preserves impressive discrimination even after the detection and reconstruction process.

Among the observables collected in the literature review, only a handful can be expected to preserve significant discrimination in experimental context. This is the case of $\Delta\phi^H(\ell^+, \ell^-)$ (figure 57) and $\Delta\phi(t, \bar{t})$ (figure 58) from [65] and b_1, b_2 (figure 61) and b_4 (figure 62) from [64]. These observables retain considerable discrimination between $t\bar{t}H$ and $t\bar{t}A$ at reconstruction level without truth match. The distribution of b_4, in particular, exhibits the largest asymmetry difference among all the tested observables, including the angular observables. The K-S test between $t\bar{t}H$ and $t\bar{t}A$ samples yields a result of 0.157 for this distribution. The K-S test result is not impressive neither for $\Delta\phi^H(\ell^+, \ell^-)$ nor $\Delta\phi(t, \bar{t})$. However, the difference in distribution shapes is visible, and the K-S test used systematically underestimates the statistical incompatibility in symmetrical distributions such as these.

The definition of b_1 is $b_1 = \frac{\vec{p_T}_t \cdot \vec{p_T}_{\bar{t}}}{p_{T\,t} p_{T\,\bar{t}}}$, which is equal to $\cos(\Delta\phi(t, \bar{t}))$. The observable b_2 has a similar definition to b_1, except that the denominator is replaced by the product of the total momenta $p_t p_{\bar{t}}$. The azimuthal angle difference between the leptons (in the Higgs boson rest frame), $\Delta\phi^H(\ell^+, \ell^-)$, is expected to be correlated to the azimuthal angle difference between the top quarks, $\Delta\phi(t, \bar{t})$. With these remarks, a relation is established between four out of the five most interesting reviewed observables. In terms of physical intuition, their respective distributions state the fact that, in $t\bar{t}A$ production, t and $\bar{t}$ directions tend to be closer in the transverse plane than in $t\bar{t}H$ production. On the other hand, the distributions of b_4, defined as $b_4 = \frac{p_t^z p_{\bar{t}}^z}{p_t p_{\bar{t}}}$, can be understood as stating that, in $t\bar{t}A$ production, the t and $\bar{t}$ often travel to opposite directions with respect to the beam axis, whereas in $t\bar{t}H$ events, they tend to travel both to the $z > 0$ or both to the $z < 0$ region.

When compared to the observables from the literature review, the new angular observables show remarkable discrimination between the two Higgs CP scenarios considered. Some of the best discriminating observables yield K-S test results above 0.12 and a few distinct distribution shapes are obtained. However, since the observables from the literature are themselves constructed mostly using angles between particle directions, they are likely to be correlated to the new angular observables. If a data analysis were to be implemented with the intent of measuring the CP-odd component of the Higgs boson, the choice of the observables to be used should take these correlations in consideration. It should also be stressed

that the most discriminating observables among the ones studied, and notably b_4, require reconstruction of the $t\bar{t}$ system, which makes this procedure of particular importance for this kind of searches.

The most discriminating new angular observables, both in signal-background and scalar-pseudoscalar tests, are observables defined using the α-β assignment. Furthermore, some of the distribution shapes obtained are evidently different from the distributions of observables without this assignment. This method succeeds in identifying the Higgs-radiating top quark in 57% of events reconstructed without truth-match, in which the Higgs actually is radiated by t or $\bar{t}$. Although this is not a very high rate, when truth-matched objects are used, the correct assignment increases to 70%, which motivates the belief that α-β assignment may lead to even more discriminating observables if one is successful in reducing the combinatorial background.

5   Conclusions and outlook

In summary, the work presented here tries to convey the message that full reconstruction of the $t\bar{t}$ system in searches for $t\bar{t}H$ production, with H decaying through $H \to b\bar{b}$ and the $t\bar{t}$ system decaying dileptonically, should not be overlooked. It opens countless possibilities for construction of new observables, otherwise inaccessible. Angular observables, in particular, demonstrate interesting signal-background discrimination and sensitivity to the CP nature of the Higgs boson. The combinatorial background arising from the 4 jets (at least) in the final state is a high price to pay, but the results of this study suggest that what is gained using the new observables may be worth this price. They also indicate where future work would be beneficial: the development of techniques to suppress combinatorial background and to improve reconstruction purity could lead to increased performance of this analysis method.

Some strategies for reducing combinatorial background that should be used in the future have already been pointed out. The results of b-tagging algorithms can be studied separately for b-jets coming from the t, $\bar{t}$ and H decay, as well as for jets from other sources (QCD or underlying event, for example). A similar study can take place with respect to the charge identification of jets. Exploiting the kinematics of the top quark decay, it is possible that constraints between the resulting lepton and b-jet provide additional rejection of wrong jet assignments. The correlation between $\theta_{\ell b}$ and $p_{\ell b}$, shown in figure 3.4.4, is an example of such a constraint, and could be used either to implement a selection cut or serve as input for calculating the likelihood of a given jet assignment.

Simply applying selection cuts to the result of these individual methods could compromise the overall efficiency of the analysis. Alternatively, they could be used as input for calculating a likelihood for each of the possible jet assignments, in a similar way to what is done with the use of p.d.f.s. Another possibility would be to train a multivariate analysis, combining information from all methods, to distinguish between correct and incorrect jet assignments.

So far, the kinematic fit itself stands as the best tool for reducing the combinatorial background. It requires that the neutrino and anti-neutrino are consistent with the reconstructed missing transverse energy, that intermediate particles are

reconstructed with the desired masses, and that the neutrino and anti-neutrino p_T p.d.f.s obtained from generated events are respected. This recipe is mainly responsible for the 33% rate of correct 4-jet assignments in events with at least one solution with truth-match. Although this rate is not likely to impress, it should be reminded that a completely random choice of jet assignment will be correct $1/12 \simeq 8.3\%$ of the time in events with 4 jets and merely $1/180 \simeq 0.56\%$ of the time in events with 6 jets. Relative to this 'blind' choice, the kinematic fit increases the rate of correct jet assignment by a factor somewhere between 4 and 60.

It has been made clear that the reconstruction routine presented here, as well as the choice of angular observables, can benefit from numerous improvements in the future. Nevertheless, the step that immediately will follow the work discussed here, ideally in parallel with the mentioned improvements, is the implementation of the developed tools in a search for $t\bar{t}H$ production, in a data analysis of LHC Run 2 pp collisions at $\sqrt{s} = 13\,\text{TeV}$. The goal is to perform this implementation in the context of the ATLAS experiment, complementing the existing Run 1 analysis with the reconstruction routine and with a selection of angular observables. Furthermore, the search should be extended to include a measurement of the pseudoscalar component of the Higgs boson.

Analogously to the ATLAS search in Run 1, the decay channel of $t\bar{t}H$ in which the Higgs boson decays through $H \to b\bar{b}$ and the $t\bar{t}$ system decays semileptonically must also be addressed. The branching ratio for this channel is approximately four times larger than for the dileptonic mode. However, the final state with a single lepton and 6 jets is more contaminated by background processes, and more prone to combinatorial background when attempting the reconstruction.

Moving from a phenomenological study into a data analysis meeting the standards of a large collaboration such as ATLAS greatly increases the complexity of the work involved. There is the requirement of using event samples in which a full simulation of the detector has been performed (instead of a fast simulation, as in DELPHES). Exhaustive studies of the systematic uncertainties must also take place. The simulated events must be reweighted to correct for inaccuracies of the background modelling or inefficiencies and resolution effects of the detectors. Regarding the data samples, the trigger options must be chosen appropriately, depending on the final state of the physical process being studied. In this case,

electron and muon triggers should be used. Additional selection cuts must be applied to ensure good working conditions of the detector, good object reconstruction, good lepton isolation, among other quality cuts. In such a data analysis, the amount of details requiring consideration can pose a challenge. Nevertheless, the possibility of working by analogy to a previously existing analysis of the same physical process surely makes the process less overwhelming.

The results of the work presented here demonstrate some advantages and difficulties of reconstructing the $t\bar{t}$ system in $t\bar{t}H$ events, with $H \to b\bar{b}$ and $t\bar{t}$ decaying dileptonically. From the discrimination obtained between signal and background using the angular observables proposed, and considering the wide room for improvement regarding the combinatorial background, a reconstruction procedure should not be readily discarded in any $t\bar{t}H$ search through this decay channel. The discrimination between $t\bar{t}H$ and $t\bar{t}A$ samples, obtained with the new observables, is similar to that of the best discriminants found in the literature, which in their turn also require reconstruction of the $t\bar{t}$ system.

During the LHC Run 2, $t\bar{t}H$ events are being produced at a higher rate than ever before. Careful analysis of these events during the next few years will result in the most stringent direct measurements of the top quark Yukawa coupling, and in measurements of the pseudoscalar component of the Higgs boson. This single process may be crucial for validating the SM as the ruling theory in particle physics, by finding the top quark Yukawa coupling to be just as predicted from the Higgs mechanism, and by setting smaller and smaller upper limits on the pseudoscalar Higgs component. If, however, the SM fails this validation, then the same physical process can become the experimental starting point for models of new physics in the Higgs sector. Either of these scenarios is a motivation for the work in $t\bar{t}H$ searches and a reason for great excitement.

6 Bibliography

[1] Peter W. Higgs. Broken symmetries, massless particles and gauge bosons. *Phys. Lett.*, (12):132, 1964. URL http://dx.doi.org/10.1016/0031-9163(64)91136-9.

[2] P.W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, (13):508, 1964. URL http://dx.doi.org/10.1103/PhysRevLett.13.508.

[3] P.W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, (145):1156, 1966. URL http://dx.doi.org/10.1103/PhysRev.145.1156.

[4] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, (13):321, 1964. URL http://dx.doi.org/10.1103/PhysRevLett.13.321.

[5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, (13):585, 1964. URL http://dx.doi.org/10.1103/PhysRevLett.13.585.

[6] T. W. B. Kibble. Symmetry breaking in non-Abelian gauge theories. *Phys. Rev.*, (155):1554, 1967. URL http://dx.doi.org/10.1103/PhysRev.155.1554.

[7] Lyndon Evans and Philip Bryant. LHC Macchine. *JINST*, (3):S08001, 2008. URL http://dx.doi.org/10.1088/1748-0221/3/08/S08001.

[8] CDF Collaboration. Observation of top quark production in $\bar{p}p$ collisions with the collider detector at Fermilab. *Phys. Rev. Lett.*, (74):2626, 1995. URL http://dx.doi.org/10.1103/PhysRevLett.74.2626.

[9] DØ Collaboration. Observation of the top quark. *Phys. Rev. Lett.*, (74):2632, 1995. URL http://dx.doi.org/10.1103/PhysRevLett.74.2632.

[10] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, (716):1, 2012. URL http://dx.doi.org/10.1016/j.physletb.2012.08.020.

[11] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, (716):30, 2012. URL http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[12] A. V. Bednyakov, B. A. Kniehl, A. F. Pikelner, and O. L. Veretin. Stability of the electroweak vacuum: Gauge independence and advanced precision. *Phys. Rev. Lett.*, 115:201802, Nov 2015. doi: 10.1103/PhysRevLett.115.201802. URL http://link.aps.org/doi/10.1103/PhysRevLett.115.201802.

[13] J.R Espinosa and M Quirós. Improved metastability bounds on the standard model higgs mass. *Physics Letters B*, 353(2–3):257 – 266, 1995. ISSN 0370-2693. doi: http://dx.doi.org/10.1016/0370-2693(95)00572-3. URL http://www.sciencedirect.com/science/article/pii/0370269395005723.

[14] Fawzi Boudjema, Rohini M. Godbole, Diego Guadagnoli, and Kirtimaan A. Mohan. Lab-frame observables for probing the top-Higgs interaction. *Phys. Rev.*, D92(1):015019, 2015. doi: 10.1103/PhysRevD.92.015019.

[15] ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider. *JINST*, (3):S08003, 2008. URL http://iopscience.iop.org/1748-0221/3/08/S08003.

[16] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, (3):S08004, 2008. URL http://iopscience.iop.org/1748-0221/3/08/S08004.

[17] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014. doi: 10.1007/JHEP02(2014)057.

[18] F. Halzen and Alan D. Martin. *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS.* 1984. New York, Usa: Wiley (1984) 396p.

[19] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014. doi: 10.1088/1674-1137/38/9/090001.

[20] Nico Kronberg. Standard model review. PhD/Diploma Student Seminar on Electroweak Symmetry Breaking and Supersymmetry, MPIK (Max Planck Institute for Nuclear Physics). URL `https://www.mpi-hd.mpg.de/lin/events/group_seminar/EW-SUSY/Nico.pdf`.

[21] M. J. Herrero. The standard model. NATO ASI 98 School, Techniques and Concepts of High Energy Physics.

[22] J. Beringer et al.(PDG). Neutrino mass, mixing and oscillations. *Phys. Rev. D*, (86), 2012. URL `http://pdg.lbl.gov/2012/reviews/rpp2012-rev-neutrino-mixing.pdf`.

[23] The Nobel Foudation. The Nobel Prize in Physics 2015 Presentation Speech. *Nobelprize.org. Nobel Media*, 2015. URL `http://www.nobelprize.org/nobel_prizes/physics/laureates/2015/presentation-speech`.

[24] User:MissMJ. Standard Model of Elementary Particles. *Wikimedia Commons, Creative Commons license CC-BY-3.0.* URL `https://upload.wikimedia.org/wikipedia/commons/0/00/Standard_Model_of_Elementary_Particles.svg`.

[25] D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse. Cavity control of a single-electron quantum cyclotron: measuring the electron magnetic moment. *Phys. Rev. A*, (83), 2011. URL `http://dx.doi.org/10.1103/PhysRevA.83.052122`.

[26] Ken-ichi Aoki, Zenrō Hioki, Rokuo Kawabe, Michiji Konuma, and Taizo Muta. Electroweak theory: Framework of on-shell renormalization and study of higher-order effects. *Progress of Theoretical Physics Supplement*, 73:1–226, 1982. doi: 10.1143/PTPS.73.1. URL `http://ptps.oxfordjournals.org/content/73/1.abstract`.

[27] ALEPH, DELPHI, L3, OPAL, SLD Collaborations, LEP Electroweak Working Group, SLD electroweak, and SLD heavy flavour group. Precision electroweak measurements on the *Z* resonance. *Physics Reports*, 427:257–454, 2006. URL `http://dx.doi.org/10.1016/j.physrep.2005.12.006`.

[28] Michael Dine. TASI lectures on the strong CP problem. In *Flavor physics for the millennium. Proceedings, Theoretical Advanced Study Institute in elementary particle physics, TASI 2000, Boulder, USA, June 4-30, 2000*, pages 349–369, 2000. URL `http://alice.cern.ch/format/showfull?sysnb=2232477`.

[29] Werner Bernreuther. CP violation and baryogenesis. *Lect. Notes Phys.*, 591:237–293, 2002. [,237(2002)].

[30] Stephen P. Martin. A Supersymmetry primer. 1997. doi: 10.1142/9789812839657_0001,10.1142/9789814307505_0001. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[31] Kenneth D. Lane. An Introduction to technicolor. In *Theoretical Advanced Study Institute (TASI 93) in Elementary Particle Physics: The Building Blocks of Creation - From Microfermius to Megaparsecs Boulder, Colorado, June 6-July 2, 1993*, 1993. URL `http://alice.cern.ch/format/showfull?sysnb=0175267`.

156

[32] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. The Hierarchy problem and new dimensions at a millimeter. *Phys. Lett.*, B429:263–272, 1998. doi: 10.1016/S0370-2693(98)00466-3.

[33] Sheldon L. Glashow. Partial-symmetries of weak interaction. *Nucl. Phys.*, (22):579, 1961. URL `http://dx.doi.org/10.1016/0029-5582(61)90469-2`.

[34] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, (19):1264, 1967. URL `http://dx.doi.org/10.1103/PhysRevLett.19.1264`.

[35] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc.*, C680519:367–377, 1968.

[36] Y. Nambu and G. Jona-Lasinio. Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I. *Phys. Rev.*, (122):345, 1967. URL `http://dx.doi.org/10.1103/PhysRev.122.345`.

[37] J. Goldstone. Field theories with "Superconductor" solutions. *Nuovo cimento*, (19):154, 1961. URL `http://dx.doi.org/10.1007/BF02812722`.

[38] Abdus Salam and Steven Weinberg. Broken symmetries. *Phys. Rev.*, (127):965, 1962. URL `http://dx.doi.org/10.1103/PhysRev.127.965`.

[39] G. 't Hooft and M. Veltman. Regularization and renormalization of gauge fields. *Nuclear Physics B*, 44 (1):189 – 213, 1972. ISSN 0550-3213. doi: http://dx.doi.org/10.1016/0550-3213(72)90279-9. URL `http://www.sciencedirect.com/science/article/pii/0550321372902799`.

[40] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables.* 2011. URL `http://dx.doi.org/10.5170/CERN-2011-002`.

[41] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions.* 2012. URL `http://dx.doi.org/10.5170/CERN-2012-002`.

[42] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties.* 2012. URL `http://dx.doi.org/10.5170/CERN-2013-004`.

[43] LHC Higgs Cross Section Working Group, . URL `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG`.

[44] The ATLAS Collaboration. Measurement of the Higgs boson mass from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data. *Phys. Rev. D*, (90):052004, 2014. URL `http://dx.doi.org/10.1103/PhysRevD.90.052004`.

[45] The CMS Collaboration. Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *The European Physical Journal C*, 75(212), 2015. URL `http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-015-3351-7`.

[46] The Nobel Foudation. The Nobel Prize in Physics 2013 Presentation Speech. *Nobelprize.org. Nobel Media*, 2013. URL `http://www.nobelprize.org/nobel_prizes/physics/laureates/2013/presentation-speech`.

[47] CMS Collaboration. Measurement of the properties of a higgs boson in the four-lepton final state. *Phys. Rev. D*, 89:092007, May 2014. doi: 10.1103/PhysRevD.89.092007. URL `http://link.aps.org/doi/10.1103/PhysRevD.89.092007`.

[48] ATLAS Collaboration and CMS Collaboration. Combined measurement of the higgs boson mass in pp collisions at $\sqrt{s} = 7$ and 8 tev with the atlas and cms experiments. *Phys. Rev. Lett.*, 114:191803, May 2015. doi: 10.1103/PhysRevLett.114.191803. URL `http://link.aps.org/doi/10.1103/PhysRevLett.114.191803`.

[49] Search for resonances decaying to photon pairs in 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2015-081, CERN, Geneva, Dec 2015. URL `http://cds.cern.ch/record/2114853`.

[50] Search for new physics in high mass diphoton events in proton-proton collisions at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-EXO-15-004, CERN, Geneva, 2015. URL `http://cds.cern.ch/record/2114808`.

[51] Andrei Angelescu, Abdelhak Djouadi, and Grégory Moreau. Scenarii for interpretations of the LHC diphoton excess: two Higgs doublets and vector-like quarks and leptons. *Phys. Lett.*, B756:126–132, 2016. doi: 10.1016/j.physletb.2016.02.064.

[52] L. D. Landau. On the angular momentum of a system of two photons. *Dokl. Akad. Nauk Ser. Fiz.*, 60(2): 207–209, 1948. doi: 10.1016/B978-0-08-010586-4.50070-5.

[53] C. N. Yang. Selection rules for the dematerialization of a particle into two photons. *Phys. Rev.*, 77:242–245, Jan 1950. doi: 10.1103/PhysRev.77.242. URL `http://link.aps.org/doi/10.1103/PhysRev.77.242`.

[54] ATLAS Collaboration. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. *Eur. Phys. J.*, C75(10):476, 2015. doi: 10.1140/epjc/s10052-015-3685-1,10.1140/epjc/s10052-016-3934-y. [Erratum: Eur. Phys. J.C76,no.3,152(2016)].

[55] CMS Collaboration. Constraints on the spin-parity and anomalous hvv couplings of the higgs boson in proton collisions at 7 and 8 tev. *Phys. Rev. D*, 92:012004, Jul 2015. doi: 10.1103/PhysRevD.92.012004. URL `http://link.aps.org/doi/10.1103/PhysRevD.92.012004`.

[56] The ATLAS and CMS Collaborations. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV. 2015. URL `http://inspirehep.net/record/1393544/`.

[57] Makoto Kobayashi and Toshihide Maskawa. CP-Violation in the renormalizable theory of weak interaction. *Prog. Theor. Phys.*, (49 (2)):652, 1973. URL `http://dx.doi.org/10.1143/PTP.49.652`.

[58] ATLAS, CDF, CMS, and DØ. First combination of Tevatron and LHC measurements of the top-quark mass. *ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014, D0-NOTE-6416, FERMILAB-TM-2582-E*, 2014. URL `arxiv.org/pdf/1403.4427`.

[59] Vardan Khachatryan et al. Measurement of the top quark mass using proton-proton data at $\sqrt{(s)} = 7$ and 8 TeV. *Phys. Rev.*, D93(7):072004, 2016. doi: 10.1103/PhysRevD.93.072004.

[60] DØ Collaboration. An improved determination of the width of the top quark. *Phys. Rev. D*, (85):091104, 2012. URL `http://dx.doi.org/10.1103/PhysRevD.85.091104`.

[61] Martin Beneke, I Efthymiopoulos, Michelangelo L Mangano, and J Womersley. Top Quark Physics. (hep-ph/0003033. CERN-TH-2000-100):111 p, Mar 2000. URL `http://cds.cern.ch/record/429963`.

[62] The ATLAS collaboration. Measurement of spin correlation in top–antitop quark events and search for top squark pair production in proton–proton collisions at $\sqrt{s}$=8 TeV using the ATLAS detector. *Report no. ATLAS-CONF-2014-056*, (56), 2014. URL `http://cds.cern.ch/record/1951325`.

158

[63] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015, 2013. doi: 10.1007/JHEP03(2013)015.

[64] John F. Gunion and Xiao-Gang He. Determining the CP nature of a neutral Higgs boson at the LHC. *Phys. Rev. Lett.*, 76:4468–4471, 1996. doi: 10.1103/PhysRevLett.76.4468.

[65] G. Brooijmans et al. Les Houches 2013: Physics at TeV Colliders: New Physics Working Group Report. 2014.

[66] ATLAS Collaboration. Luminosity Public Results. . URL `twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun1`.

[67] CMS Collaboration. Luminosity Public Results. . URL `twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[68] CERN. Long Shutdown 1: Exciting times ahead. *CERN webpage*, . URL `http://home.cern/about/updates/2013/02/long-shutdown-1-exciting-times-ahead`.

[69] CERN. CERN announces LHC restart schedule. *CERN webpage*, . URL `http://home.cern/about/updates/2014/06/cern-announces-lhc-restart-schedule`.

[70] ATLAS. Detector description. *ATLAS Webpage, CERN*, . URL `http://www.atlas.ch/detector.html`.

[71] ATLAS. ATLAS photos. *ATLAS Webpage, CERN*, . URL `http://www.atlas.ch/photos/index.html`.

[72] M Capeans, G Darbo, K Einsweiller, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne, and R Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, CERN, Geneva, Sep 2010. URL `http://cds.cern.ch/record/1291633`.

[73] 2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data. Technical Report ATL-DAQ-PUB-2016-001, CERN, Geneva, Mar 2016. URL `https://cds.cern.ch/record/2136007`.

[74] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04: 063, 2008. doi: 10.1088/1126-6708/2008/04/063.

[75] Expected performance of the ATLAS *b*-tagging algorithms in Run-2. Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015. URL `http://cds.cern.ch/record/2037697`.

[76] Georges Aad et al. Performance of pile-up mitigation techniques for jets in pp collisions with the ATLAS detector. *Nucl. Instrum. Meth.*, A824:367–370, 2016. doi: 10.1016/j.nima.2015.10.039.

[77] CMS. CMS detector design. *CMS Webpage, CERN*. URL `http://cms.web.cern.ch/news/cms-detector-design`.

[78] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012. doi: 10.1140/epjc/s10052-012-1896-2.

[79] Sanjoy Biswas, Rikkert Frederix, Emidio Gabrielli, and Barbara Mele. Enhancing the ttH signal through top-quark spin polarization effects at the LHC. *JHEP*, (07):020, 2014. URL `http://dx.doi.org/10.1007/JHEP07(2014)020`.

159

[80] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[81] Georges Aad et al. Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J.*, C75(7):349, 2015. doi: 10.1140/epjc/s10052-015-3543-1.

[82] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

[83] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006. doi: 10.1088/1126-6708/2006/05/026.

[84] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473–500, 2008. doi: 10.1140/epjc/s10052-007-0490-5.

[85] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185:2930, 2014. doi: 10.1016/j.cpc.2014.06.021.

[86] LHC Top Working Group. NLO single-top channel cross sections, . URL `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec`.

[87] Anna Kulesza, Leszek Motyka, Tomasz Stebel, and Vincent Theeuwes. Soft gluon resummation for associated ttH production at the LHC. *JHEP*, 03:065, 2016. doi: 10.1007/JHEP03(2016)065.

[88] Eric Conte, Benjamin Fuks, and Guillaume Serret. MadAnalysis 5, a user-friendly framework for collider phenomenology. *Computer Physics Communications*, 184(1):222–256, 2013. ISSN 0010-4655. doi: http://dx.doi.org/10.1016/j.cpc.2012.09.009. URL `http://www.sciencedirect.com/science/article/pii/S0010465512002950`.

[89] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997. doi: 10.1016/S0168-9002(97)00048-X.

[90] David Krohn, Matthew D. Schwartz, Tongyan Lin, and Wouter J. Waalewijn. Jet charge at the lhc. *Phys. Rev. Lett.*, 110:212001, May 2013. doi: 10.1103/PhysRevLett.110.212001. URL `http://link.aps.org/doi/10.1103/PhysRevLett.110.212001`.

[91] Jeffrey D. Richman. An Experimenter's Guide to the Helicity Formalism. *CALT-68-1148 DOE Research and Development Report*, 1984. URL `http://charm.physics.ucsb.edu/people/richman/ExperimentersGuideToTheHelicityFormalism.pdf`.

[92] Eugene P. Wigner. On Unitary Representations of the Inhomogeneous Lorentz Group. *Annals Math.*, 40:149–204, 1939. doi: 10.2307/1968551. [Reprint: Nucl. Phys. Proc. Suppl.6,9(1989)].

[93] M. Feindt and U. Kerzel. The NeuroBayes neural network package. *Nucl. Instrum. Meth.*, A559:190–194, 2006. doi: 10.1016/j.nima.2005.11.166.

[94] S. P. Amor dos Santos, J. P. Araque, R. Cantrill, N. F. Castro, M. C. N. Fiolhais, R. Frederix, R. Gonçalo, R. Martins, R. Santos, J. Silva, A. Onofre, H. Peixoto, and A. Reigoto. Angular distributions in $t\bar{t}h(h \to b\bar{b})$ reconstructed events at the lhc. *Phys. Rev. D*, 92:034021, Aug 2015. doi: 10.1103/PhysRevD.92.034021. URL `http://link.aps.org/doi/10.1103/PhysRevD.92.034021`.