UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Search for a charged Higgs boson in $\tau\nu_\tau$ and $\mathrm{t\bar{b}}$ decays in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV with the CMS detector

Pietro Vischia

Orientador: Doutor João Manuel Coelho dos Santos Varela
Co-Orientador: Doutor Michele Gallinaro

Tese aprovada em provas públicas para obtenção do Grau de Doutor em

Física

Qualificação atribuída pelo Júri: Aprovado com Distinção

Júri

Presidente: Presidente do Conselho Científico do IST
Vogais:
Doutor Mário João Martins Pimenta
Doutor João Manuel Coelho dos Santos Varela
Doutor José Ricardo Morais Silva Gonçalo
Doutor Michele Gallinaro
Doutor Marco Pieri
Doutor Pedro Manuel Vieira de Castro Ferreira da Silva

2016

TÉCNICO LISBOA

UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

Search for a charged Higgs boson in $\tau\nu_\tau$ and $\mathrm{t\bar{b}}$ decays in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV with the CMS detector

Pietro Vischia

Orientador: Doutor João Manuel Coelho dos Santos Varela

Co-Orientador: Doutor Michele Gallinaro

Tese aprovada em provas públicas para obtenção do Grau de Doutor em

Física

Qualificação atribuída pelo Júri: Aprovado com Distinção

Júri

Presidente: Presidente do Conselho Científico do IST

Vogais:

Doutor Mário João Martins Pimenta, Professor Catedrático do Instituto Superior Técnico da Universidade de Lisboa;

Doutor João Manuel Coelho dos Santos Varela, Professor Associado (com agregação) do Instituto Superior Técnico da Universidade de Lisboa;

Doutor José Ricardo Morais Silva Gonçalo, Investigador Principal, Laboratório de Instrumentação e Física Experimental de Partículas (LIP), Lisboa;

Doutor Michele Gallinaro, Professor Auxiliar Convidado do Instituto Superior Técnico da Universidade de Lisboa;

Doutor Marco Pieri, Full Research Scientist, University of California, San Diego, USA;

Doutor Pedro Manuel Vieira de Castro Ferreira da Silva, Investigador, CERN - European Organization for Nuclear Research, Suiça.

Instituições Financiadoras

FCT - Fundação da Ciência e Tecnología, bolsa SFRH/BD/52067/2012 (programa IDPASC)

2016

To my grandmother Rina, that missed this last one.

Sit tibi terra levis.

“In almost all circumstances, and at all times,
we find ourselves in a state of uncertainty.
Uncertainty in every sense.”
– Bruno de Finetti

“That’s what I do.
I drink, and I know things.”
– Tyrion Lannister

Acknowledgements

Arrived at the end of this amazingly formative road, I thank:

the IDPASC program and the Portuguese FCT (Fundação da Ciência e Tecnología) for the financial support, under the FCT grant SFRH/BD/52067/2012; LIP (Laboratório de Instrumentação e Física Experimental de Partículas), for supporting financially my training (trips to physics schools) and my first steps into the scientific community (conferences and workshops) through the projects FCOMP-01-0124-FEDER-022817 (Ref. FCT CERN/FP/123601/2011) and CERN/FIS-NUC/0029/2015, and for providing a stimulating scientific environment through the organization of seminars on different disciplines; my advisors, João Varela and Michele Gallinaro, for the supervision, support, and guidance provided along these years; the weekly LIP-CMS meetings, for having been an endless source of anedoctes; Tommaso Dorigo, who is a friend since many years; André David, that showed me how a scientific discussion should be; Lauri Wendland, that showed me the value of pickiness, and that I am deeply honoured to have worked with; Loic Quertenmont, for always making sure that there are not too many executables in the repository; João Seixas, for understanding the needs of the students, and doing whatever he can to help; Nuno Almeida, for teaching me the advantages of silence – that I have never learnt completely; Olaf Behnke, Bob Cousins, Louis Lyons, and the other fellows of the CMS Statistics Committee, for being a continuous source of stimuli; Javier Cuevas, for believing in me.

I thank my parents, for having always let me follow my path, and Marta, for loving me and being patient now that I will fly away. I love you! I am still waiting for those hints on aquarela, though...

I also thank: Daniele, because we were united in the darkest hours; Ana Carolina, because of the endless exchanges of ideas; Tiago, because of the DnD sessions and the prestidigitation tricks; Nicola, for our endless chats; Alessandra and Mario, because even if we see each other once per year or less, it is always as if time did not pass; Cristóvão, for the awesome and very very precise discussions; the other people from the office (Alex, Bruno, Matilde), for sharing a variable amount of years and daily awesomeness together; the people from *Meccanica Quantistica Gruppo Serio*, because I met decent people and had always the possibility of having “flames” with many others; my cats, Maria João and Sborra, for providing an endless list of things to clean everyday, while still being so fluffy.

A special thank goes to: Alice, because if I am here it is because of her; Lara, because she showed me a side of life that I had forgotten, and that there are better places to be at and better people to hang with; Pedro, who taught me a lot and is a model of what a physicist should be; and to Maurizio Loreti, who opened my mind and is a true and everlasting inspiration for life and work.

Last but not least, I thank all the negative examples and situations, because in an ideal world they should not be there, but in the real world they make you grow, faster but bitter. Still, fuck you :V

Title: Search for a charged Higgs boson in $\tau\nu_\tau$ and $\mathrm{t\bar{b}}$ decays in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV with the CMS detector

Abstract

The Large Hadron Collider (LHC) started the first proton-proton collisions at a center-of-mass energy of 7 TeV in 2010. Soon thereafter, the experiments started collecting data and were able to rediscover the Standard Model (SM) in a few months, thanks to the very good understanding of the detectors, and their already precise calibrations. The LHC took data at $\sqrt{s} = 7$ TeV and 8 TeV in the years 2010-2011 and 2012, respectively: the peak of his intensive data taking has been, in 2012, the discovery, by the CMS and ATLAS experiments, of a neutral boson with a mass of approximately 125 GeV. The properties of the new boson are consistent with those predicted for the Standard Model (SM) Higgs boson, and models with an extended Higgs sector are constrained by the measured properties of the new boson: the discovery of another scalar boson, neutral or charged, would represent unambiguous evidence for the presence of physics beyond the SM.

Charged Higgs bosons are predicted in models consisting of at least two Higgs doublets, of which the simplest are the two-Higgs-doublet models (2HDM). Among the physical Higgs bosons predicted by these models, the charged ones (H^+ and its charge conjugate, H^-) have been chosen as topic for this Thesis.

At the LHC, the charged Higgs boson can be studied in top quark decays, if the mass of the charged Higgs boson is smaller than the mass difference between the top and the bottom quarks, i.e. $m_{\mathrm{H^+}} < (m_{\mathrm{t}} - m_{\mathrm{b}})$, or in associated production with a top quark if $m_{\mathrm{H^+}} > (m_{\mathrm{t}} - m_{\mathrm{b}})$. In the latter case, top quarks are also part of the decay chain of the charged Higgs boson.

In this Thesis, a charged Higgs boson is searched for across the whole mass range. In the case $m_{\mathrm{H^+}} < (m_{\mathrm{t}} - m_{\mathrm{b}})$, studied at a center-of-mass energy of 7 TeV, the lepton+tau final state is selected: the main background in that case is the contribution from events with jets misidentified as taus, hence a dedicated data driven technique is employed to improve the estimate of that contribution.

In the case $m_{\mathrm{H^+}} > (m_{\mathrm{t}} - m_{\mathrm{b}})$, where most BSM scenarios predict a charged Higgs boson decay into a top and a bottom quark, the dilepton final state is investigated, being more sensitive to the presence of a charged Higgs boson.

Two measurements of SM parameters, the top mass at 7 TeV and the top pair production cross section at 8 TeV, are presented first, as ways of obtaining a better understanding of the calibration of the CMS detector.

A search for a charged Higgs boson decaying into a tau lepton and a neutrino is presented next, in the final state characterized by one muon and one hadronically decaying tau. A previous publication by the CMS collaboration had set upper limits of the order of 5% for the branching ratio $\mathcal{B}(\mathrm{t} \to \mathrm{H^+\bar{b}})$, by using almost half of the data collected at 7 TeV by the CMS detector. That result has been improved in the context of this Thesis, by analyzing the full 7 TeV dataset and applying more advanced statistical methods (improving the data driven determination of the main background, and exploiting the polariza-

tion properties of the tau lepton in order to improve the sensitivity of the result): the improvement has been of the order of 50%, bringing the upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\overline{\mathrm{b}})$ down to $\sim 2 - -3$%.

A search for a charged Higgs boson with a mass larger than the top quark mass is presented next, that has been performed using the 8 TeV data using two main final states: the lepton+tau final state, that in this mass region is characterized by a very low sensitivity, and the dilepton final state. In this mass regime, for the majority of the theoretical scenarios, the charged Higgs boson decays either into a tau and a neutrino, or into a top quark and a b quark. The production mechanism changes as well, and is characterized by the associate production of a top quark and a charged Higgs, sometimes with an additional b quark. The decay into a top quark and a b quark, together with the production mechanism, justifies the use of the dilepton final state as a flagship for the search strategy. The result of this search represented the first direct search ever for a heavy charged Higgs boson decaying into a top quark and a b quark.

Finally, an overview is made of the perspectives for charged Higgs boson searches in the context of the ongoing LHC run at a center-of-mass energy of 13 TeV.

Keywords: CMS, Standard Model, Charged Higgs, MSSM, Awesomeness

Título: Busca por um bosão de Higgs carregado nos canais de decaimento $\tau\nu_\tau$ e $\mathrm{t\bar{b}}$, em colisões protão -protão com uma energia no centro de massa de $\sqrt{s} = 7$ e 8 TeV com o detetor CMS

Nome: Pietro Vischia

Doutoramento em: Física

Orientador: Professor Doutor João Manuel Coelho dos Santos Varela

Co-orientador: Doutor Michele Gallinaro

Resumo

O Large Hadron Collider (LHC) iniciou, no ano 2010, a colidir protões com protões com uma energia de 7 TeV no centro de massa. Logo depois, as collaboraç oes experimentais começaram a colecionar dados, e foram capazes de re-descubrir o Modelo Padrão (MP) em poucos meses: isso foi possível graças à excellente comprehensão do detector e a sua muito precisa calibração .

O LHC tomou dados com $\sqrt{s} = 7$ TeV e 8 TeV nos anos 2010-2011 and 2012, respectivamente: o pico desta operação intensiva de tomada de dados foi, no 2012, a descoberta, efectuada por as collaboraç oes experimentais CMS and ATLAS, de um bosão neutral com uma massa de aproximadamente 125 GeV. As propriedades deste novo bosão são compativeis com as do bosão de Higgs previsto por o MP: outros modelos, caracterizados por um setor de Higgs extendido, foram sujeitos a vínculos devido as propriedades medidas do novo bosão : a descoberta de um outro bosão scalar, seja isso neutro ou carregado, representaria uma evidência experimental muito clara da presenҫa de física fora do MP.

Bosões de Higgs carregados são previstos por modelos em que há pelo menos dois dubletos: os modelos deste genero mais simples são os two-Higgs-doublet models (2HDM), que prevêm a existência de cinco bosões, entre os quais os dois que são carregados – sendo um o conjugado de carga do outro – foram escolhidos como tópico principal desta Tese de doutoramento.

No LHC, pode-se estudar o bosão de Higgs carregado nos decaimentos de quarks top se a massa do Higgs carregado for minor da diferença entre a do top and do bottom quarks, ou seja $m_{\mathrm{H^+}} < (m_{\mathrm{t}} - m_{\mathrm{b}})$, ou em associação com um quark top se $m_{\mathrm{H^+}} > (m_{\mathrm{t}} - m_{\mathrm{b}})$. Neste último caso, os top quarks fazem também parte da cadeia de decaimento do bosão de Higgs carregado.

O tema principal desta Tese é a busca de um bosão de Higgs carregado, num alcance de massa que inclui os dois regimes. No regime $m_{\mathrm{H^+}} < (m_{\mathrm{t}} - m_{\mathrm{b}})$, que foi estudiado em colisões com uma energia no centro de massa de 7 TeV, foi selecionado o estado final caracterizado por um leptão (electrão o muão) e um tau: o fundo principal, neste caso, é a contribução de eventos com jatos incorrectamente identificados como taus, e consequentemente foi usada uma técnica *data driven* para melhorar a estimativa

da entidade desta contribuição .

No regime $m_{\mathrm{H}^+} > (m_{\mathrm{t}} - m_{\mathrm{b}})$, onde a maioria dos scenarios de física ao longo do modelo padrão (*beyond standard model* – BSM) prefire um decaimento do bosão de Higgs carregado em um quark top e um quark bottom, foi o estado final caracterizado por dois leptões (dois muões, dois electrões, ou um muão e um electrão) foi seleccionado, porque é tem mais sensitividade a presenạ de um bosão de Higgs carregado.

Duas medidas de parámetros do modelo padrão , a massa do quark top em dados a 7 TeV e a medida da secção eficaz de pares de quarks top em dados a 8 TeV, são apresentadas como medidas preliminares que permitiram de adquirir uma melhor comprehensão da calibração do detetor CMS.

Uma busca para um bosão de Higgs carregado, do canal de decaimento com um tau e um neutrino, é apresentada no estado final caracterizado por um muão e um tau sujeito a decaimento adrónico. Uma publicação anterior, por a colaboração CMS, pus limites superiores de acerca de 5% na fracção de decaimento $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\overline{\mathrm{b}})$, utilizando quase metade dos dados colecionados a 7 TeV com o detetor CMS. Nesta Tese, este resultado foi melhorado analizando o inteiro dataset a 7 TeV e aplicando métodos estatísticos mais avançados: a estimativa *data driven* do fundo principal foi melhorada, e a sensitividade do resultado foi melhorada utilizando a informação contida em forma de polarização do leptão tau. O resultado foi melhorado quase do 50%, e o limite superior em $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\overline{\mathrm{b}})$ passou a ser de $\sim 2 - -3\%$.

Uma busca por um bosão de Higgs carregado com massa maior da massa to quark top é apresentada, utilizando datos colecionados a 8 TeV, em dois estados finais: o estado final lepton+tau, que nesto regime de massa tem uma sensitividade muito baixa, e o estado final dilepton (dois muões, dois electrões, ou um muão e um electrão). Neste regime de massa, na maioria dos scenários teóricos, o bosão de Higgs carregado decai ou em tau+neutrino, ou em um quark top e um quark bottom. O mecanismo de produção também muda, e é neste regime caracterizado por a produção asociada de um quark top e de um bosão de Higgs carregado, as vezes com um bottom quark adicional. O decaimento num quark top e um quark bottom, junto com o mecanismo de produção , justifica o uso do estado final caracterizado por dois leptões como porta-estandarte da stratégia de busca. O resultado desta busca representa a primeira busca direita por um bosão de Higgs carregado no canal de decaimento em um quark top e um quark bottom.

Para concluir, as perspectivas para a busca de um bosão de Higgs carregado são apresentadas no contexto da tomada de dados com uma energia no centro de massa de 13 TeV, que está a acontecer neste momento no LHC.

Palavras-chave: CMS, Modelo Padrão , Higgs carregado, MSSM, Grandiosidade.

Título: Busca por um bosão de Higgs carregado nos canais de decaimento $\tau\nu_\tau$ e $\mathrm{t\bar{b}}$, em colisões protão -protão com uma energia no centro de massa de $\sqrt{s} = 7$ e 8 TeV com o detetor CMS

Resumo Alargado

O Large Hadron Collider (LHC) iniciou, no ano 2010, a colidir protões com protões com uma energia de 7 TeV no centro de massa. Logo depois, as collaboraç oes experimentais começaram a colecionar dados, e foram capazes de re-descubrir o Modelo Padrão (MP) em poucos meses: isso foi possível graças à excellente comprehensão do detector e a sua muito precisa calibração .

O LHC tomou dados com $\sqrt{s} = 7$ TeV e 8 TeV nos anos 2010-2011 and 2012, respectivamente: o pico desta operação intensiva de tomada de dados foi, no 2012, a descoberta, efectuada por as collaboraç oes experimentais CMS and ATLAS, de um bosão neutral com uma massa de aproximadamente 125 GeV. As propriedades deste novo bosão são compativeis com as do bosão de Higgs previsto por o MP: outros modelos, caracterizados por um setor de Higgs extendido, foram sujeitos a vínculos devido as propriedades medidas do novo bosão : a descoberta de um outro bosão scalar, seja isso neutro ou carregado, representaria uma evidência experimental muito clara da presenạ de física fora do MP.

Bosões de Higgs carregados são previstos por modelos em que há pelo menos dois dubletos: os modelos deste genero mais simples são os two-Higgs-doublet models (2HDM), que prevêm a existência de cinco bosões, entre os quais os dois que são carregados – sendo um o conjugado de carga do outro – foram escolhidos como tópico principal desta Tese de doutoramento.

No LHC, pode-se estudar o bosão de Higgs carregado nos decaimentos de quarks top se a massa do Higgs carregado for minor da diferença entre a do top and do bottom quarks, ou seja $m_{\mathrm{H^+}} < (m_{\mathrm{t}} - m_{\mathrm{b}})$, ou em associação com um quark top se $m_{\mathrm{H^+}} > (m_{\mathrm{t}} - m_{\mathrm{b}})$. Neste último caso, os top quarks fazem também parte da cadeia de decaimento do bosão de Higgs carregado.

O tema principal desta Tese é a busca de um bosão de Higgs carregado, num alcance de massa que inclui os dois regimes. No regime $m_{\mathrm{H^+}} < (m_{\mathrm{t}} - m_{\mathrm{b}})$, que foi estudiado em colisões com uma energia no centro de massa de 7 TeV, foi selecionado o estado final caracterizado por um leptão (electrão o muão) e um tau: o fundo principal, neste caso, é a contribução de eventos com jatos incorrectamente identificados como taus, e consequentemente foi usada uma técnica *data driven* para melhorar a estimativa da entidade desta contribuição .

No regime $m_{\mathrm{H^+}} > (m_{\mathrm{t}} - m_{\mathrm{b}})$, onde a maioria dos scenarios de física ao longo do modelo padrão (*beyond standard model* – BSM) prefire um decaimento do bosão de Higgs carregado em um quark top e um quark bottom, foi o estado final caracterizado por dois leptões (dois muões, dois electrões, ou um muão e um electrão) foi seleccionado, porque é tem mais sensitividade a presenạ de um bosão de Higgs carregado.

Duas medidas de parámetros do modelo padrão , a massa do quark top em dados a 7 TeV e a medida da secção eficaz de pares de quarks top em dados a 8 TeV, são apresentadas como medidas preliminares que permitiram de adquirir uma melhor comprehensão da calibração do detetor CMS.

Uma busca para um bosão de Higgs carregado, do canal de decaimento com um tau e um neutrino, é

apresentada no estado final caracterizado por um muão e um tau sujeito a decaimento adrónico. Uma publicação anterior, por a colaboração CMS, pus limites superiores de acerca de 5% na fracção de decaimento $\mathcal{B}(\mathrm{t} \rightarrow \mathrm{H}^{+}\overline{\mathrm{b}})$, utilizando quase metade dos dados colecionados a 7 TeV com o detetor CMS. Nesta Tese, este resultado foi melhorado analizando o inteiro dataset a 7 TeV e aplicando métodos estatísticos mais avançados: a estimativa *data driven* do fundo principal foi melhorada, e a sensitividade do resultado foi melhorada utilizando a informação contida em forma de polarização do leptão tau. O resultado foi melhorado quase do 50%, e o limite superior em $\mathcal{B}(\mathrm{t} \rightarrow \mathrm{H}^{+}\overline{\mathrm{b}})$ passou a ser de $\sim 2 - -3$%.

Uma busca por um bosão de Higgs carregado com massa maior da massa to quark top é apresentada, utilizando datos colecionados a 8 TeV, em dois estados finais: o estado final lepton+tau, que nesto regime de massa tem uma sensitividade muito baixa, e o estado final dilepton (dois muões, dois electrões, ou um muão e um electrão). Neste regime de massa, na maioria dos scenários teóricos, o bosão de Higgs carregado decai ou em tau+neutrino, ou em um quark top e um quark bottom. O mecanismo de produção também muda, e é neste regime caracterizado por a produção asociada de um quark top e de um bosão de Higgs carregado, as vezes com um bottom quark adicional. O decaimento num quark top e um quark bottom, junto com o mecanismo de produção , justifica o uso do estado final caracterizado por dois leptões como porta-estandarte da stratégia de busca. O resultado desta busca representa a primeira busca direita por um bosão de Higgs carregado no canal de decaimento em um quark top e um quark bottom.

Para concluir, as perspectivas para a busca de um bosão de Higgs carregado são apresentadas no contexto da tomada de dados com uma energia no centro de massa de 13 TeV, que está a acontecer neste momento no LHC.

Palavras-chave: CMS, Modelo Padrão , Higgs carregado, MSSM, Grandiosidade.

Contents

List of Tables

List of Figures

Glossary

ALICE A Large Ion Collider Experiment. 20

ATLAS A Toroidal LHC ApparatuS. 19

CERN Conseil Europeen pour la Recherche Nucleaire. 20

CMS Compact Muon Solenoid. 19, 20

LHC Large Hadron Collider. 19

LHCb Large Hadron Collider beauty experiment. 19

Chapter 1

Introduction

1.1 Physics at the LHC: the charged Higgs bosons in top quark physics

The Large Hadron Collider (LHC) started the first proton-proton collisions at a center-of-mass energy of 7 TeV in 2010. Soon thereafter, the experiments started collecting data and were able to rediscover the Standard Model (SM) in a few months. This impressive achievement was made possible thanks to the very good understanding of the detectors, and their already precise calibrations. The LHC took data at $\sqrt{s} = 7$ TeV and 8 TeV in the years 2010-2011 and 2012, respectively. This period of data-taking is commonly referred to as *Run 1*. In 2015 – after approximately 2 years of maintenance and upgrading the accelerator complex and the experimental apparata – the LHC restarted operations at the increased collision energy of $\sqrt{s} = 13$ TeV. The work performed in this thesis refers to the analysis of the *Run 1* data, i.e. at $\sqrt{s} = 7$ and 8 TeV collision energies.

In 2012, a neutral boson with a mass of approximately 125 GeV was discovered by the CMS and ATLAS experiments [14, 15] at the CERN LHC. The properties of the new boson are consistent with those predicted for the standard model (SM) Higgs boson [16, 17, 18, 19, 20, 21, 22]. Models with an extended Higgs sector are constrained by the measured mass, CP quantum numbers, and production rates of the new boson. The discovery of another scalar boson, neutral or charged, would represent unambiguous evidence for the presence of physics beyond the SM.

Charged Higgs bosons are predicted in models consisting of at least two Higgs doublets, of which the simplest are the two-Higgs-doublet models (2HDM) [23]. Two Higgs doublets result in five physical Higgs bosons: light and heavy CP-even Higgs bosons h and H, a CP-odd Higgs boson A, plus charged Higgs bosons H^+ and its charge conjugate. Charge conjugate states and processes are always implied throughout this manuscript. The minimal supersymmetric SM (MSSM) [24, 25, 26, 27, 28, 29, 30, 31] used as a benchmark in this thesis is a special case of a Type-II 2HDM scenario.

If the mass of the charged Higgs boson is smaller than the mass difference between the top and the bottom quarks, i.e. $m_{\mathrm{H}^+} < (m_{\mathrm{t}} - m_{\mathrm{b}})$, the top quark can decay via $\mathrm{t} \to \mathrm{H}^+\mathrm{b}$. This is referred to as the *light* charged Higgs scenario. In this case, the charged Higgs boson is produced most frequently

in $\mathrm{t\bar{t}}$ production. Furthermore, in the considered MSSM benchmark scenarios it preferentially decays to a τ lepton and the corresponding neutrino, $\mathrm{H}^+ \to \tau^+\nu_\tau$, for $\tan\beta > 5$, where $\tan\beta$ is defined to be equal to the ratio of the vacuum expectation values of the two Higgs boson doublet fields [32]. The diagram for a dominant production and decay mode for a low-mass charged Higgs boson is shown in Fig. 1.1 (a). The presence of the $\mathrm{t} \to \mathrm{H}^+\mathrm{b}$ with $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay modes would alter the τ yield in the decays of $\mathrm{t\bar{t}}$ pairs. The most sensitive 95 % confidence level (CL) upper limits on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b})$ have been determined by the ATLAS and CMS experiments. For the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode with the hadronic decay of the τ lepton and hadronic W boson decays (τ_h+jets) final state the current upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ is set to 1.0–0.2 % for $m_{\mathrm{H}^+} = 80$–160 GeV [2, 33]. For the $\ell\tau_\mathrm{h}$ (where ℓ=e, μ throughout the manuscript) and $\mathrm{e}\mu$ final states the 95 % CL upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b})$ is set to 3–4 % in the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode for $m_{\mathrm{H}^+} = 80$–160 GeV assuming $\mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1$ [2, 34, 35]. Also the $\mathrm{H}^+ \to \mathrm{c\bar{s}}$ decay mode with the ℓ+jets final states have been studied with the current 95 % CL upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b})$ set to 5–1 % for $m_{\mathrm{H}^+} = 90$–160 GeV by the ATLAS experiment with the assumption $\mathcal{B}(\mathrm{H}^+ \to \mathrm{c\bar{s}}) = 1$ [36]. The large electron-positron collider experiments determined a model-independent lower limit of 78.6 GeV on the H^+ mass [37, 38, 39, 40].

If the charged Higgs boson mass exceeds the mass difference between the top and bottom quark, i.e. $m_{\mathrm{H}^+} > m_\mathrm{t} - m_\mathrm{b}$, the charged Higgs boson is predominantly produced by the fusion of bottom and top quarks illustrated in Figs. 1.1 (b) and (c). The two production mechanisms, and the interplay between them, are described in better detail in Chap. 7. This is referred to as the *heavy* charged Higgs scenario.

In the MSSM benchmark scenarios considered, the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode dominates for $m_{\mathrm{H}^+} < 220$ GeV [32]. For large m_{H^+} and large $\tan\beta$ values the decay $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ becomes dominant but the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode remains experimentally viable. The $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode is probed with the best sensitivity in the final state characterized by the hadronic decays of both the tau and the associated W boson: the current upper limits on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ are set to 0.8–0.004 for $m_{\mathrm{H}^+} = 180$–1000 GeV [33].

The mass range around the top quark mass, $m_{\mathrm{H}^+} = 160$–180 GeV is characterized by interference terms between the charged Higgs boson production in top quark decays and the associated production of a charged Higgs boson together with a top quark. Preliminar theoretical prescriptions on how to generate simulated events for the charged Higgs signal in this interference region are still in discussion among theoreticians: for this reason, the interference mass range could not be included in the analyses that are described in this manuscript.

At the LHC collision energies, the production cross section of top quarks is large and, consequently, the LHC can be effectively considered a “top factory”. The properties of the top quark have been object of precise measurements from previous experiments, in particular CDF (the experiment that actually discovered the top quark) and D0. Since the very specific topology of its decay modes is well known and many of those properties (such as the cross section) are very well known from theory, the study of top quark production is also a powerful handle to cross-check the calibration of the CMS detector, for any run at any given center-of-mass energy, as well as for performing precision measurements for some of the SM parameters, the two most immediate ones being the top quark mass and the top quark pairs

production cross section. The variety, in terms of involved objects, of the final states originated from top quarks represents a unique opportunity for probing the same observables by exploiting different parts of the detector.

Furthermore, due to its large mass, the top quark is predicted to have couplings to BSM particles strong enough to permit the observation of BSM particles either in top quark decays, or decaying into top quarks, or produced in association to top quarks. Consequently, the top quark is a powerful handle in the search for New Physics.

Top quarks at the CERN LHC are mostly produced in pairs with the subsequent decays $\mathrm{t\bar{t}} \rightarrow \mathrm{W^{+}bW^{-}\bar{b}}$. The decay modes of the two W bosons determine the event signature. The decay channel labeled *dilepton* corresponds to the case in which both W bosons decay into leptons, where the term lepton usually refers to electrons or muons, as studied in Refs. [41, 42]. The decay channel labeled *tau dilepton* corresponds instead to the case in which one of the W bosons decay into a light lepton (electron or muon) and the other decays to a tau. The study of this decay mode is particularly important in the context of charged Higgs searches, because the charged Higgs, when produced in the decay of the top quark, would replace the W boson in the $\mathrm{t} \rightarrow \mathrm{W^{+}b}$ decay; this process can happen if the charged Higgs mass is lower than the top quark mass, and when this happens the MSSM predicts that the charged Higgs decay preferentially into a tau and a neutrino, in most scenarios. Consequently, the tau dilepton decay of the top quark represents a main background for the search for a light charged Higgs.

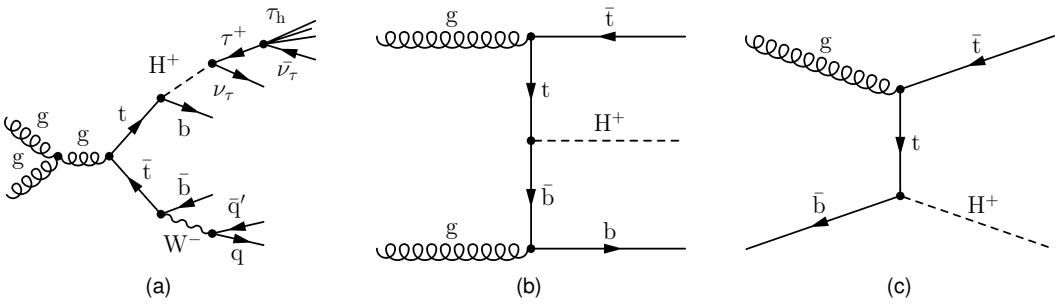


Figure 1.1: (a) Diagram for a dominant production mode for the light charged Higgs boson through $\mathrm{t\bar{t}}$ production with a subsequent decay to the τ_{h}+jets final state. (b–c) Direct production of the charged Higgs boson in the 4FS and 5FS, respectively.

1.2 Overview of the work

This manuscript collects work performed by its author across his career as a PhD student, and spans both the 7 TeV and the 8 TeV LHC runs. The thesis work focused on the search for a charged Higgs boson with the available data in Run 1, both in the *light* and *heavy* charged Higgs scenarios.

The 7 TeV data are first used to measure the top quark mass, which is a way of obtaining a better understanding of the calibration of the CMS detector, of the objects (electrons, muons, jets, etc.) characteristic of the final state examined, and a comprehensive verification of the data. The result was, at the moment in which it was made public as a preliminary result, the most precise measurement of the

top mass in the dilepton final state, and has been later published as a cross-check of an improved result, from another group, that makes uses of another method in Ref. [43]. The published main result has been later included by the CMS collaboration in a global fit for top mass measurement using all the available data (at 7 and 8 TeV) for all the available final states [44]. The top mass measurement is described in Sec. 5.1: the measurement method was developed previously by former elements of the group, for an integrated luminosity of about $36\ pb^{-1}$: the author of this manuscript joined the group at the moment of updating the result with $2.3\ fb^{-1}$, and had his initial training in producing the results and participate in the analysis by following all the various steps of the measurement. This study of the dilepton final state at 7 TeV has been also useful as a preparation to exploiting that final state in the search for a charged Higgs boson at 8 TeV described below.

The 7 TeV data have been then exploited in order to search for charged Higgs bosons decaying into a tau lepton and a neutrino, in the final state characterized by one muon and one hadronically decaying tau. A previous publication by the CMS collaboration [2] had set upper limits of the order of 5–10% for charged Higgs boson masses between 80 and 160 GeV for the branching ratio $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\overline{\mathrm{b}})$, by using almost half of the data collected at 7 TeV by the CMS detector. That result has been improved in the context of this thesis, by analyzing the full 7 TeV dataset and applying more advanced statistical methods (improving the data driven determination of the main background, and exploiting the polarization properties of the tau lepton in order to improve the sensitivity of the result): the improvement has been of the order of 50%, bringing the upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\overline{\mathrm{b}})$ down to approximately 2–5% in the charged Higgs mass region between 80 and 160 GeV. This improved result is described in Sec. 6. The author has been the responsible of this analysis, improved the background estimation method that was used in the previous publication, and improved the statistical analysis by introducing a shape-based computation of the upper limits.

The 8 TeV data have then been analysed, again starting from a precision measurement of a Standard Model property: the top quark pair production cross section has been measured in the lepton+tau final state, by applying an determination of the background due to misidentified tau leptons improved with respect to the 7 TeV charged Higgs search described in Sec. 6. This resulted in an improved top cross section measurement in the tau+lepton final state, with respect to previous results [45, 46, 47, 48, 49]. The author rewrote the analysis code inherited by earlier members of the group, to improve its efficiency, and adapted it to the 8 TeV datasets. The resulting measurement was published in Ref. [43] and is described in Sec. 5.2.6. The main contribution of the author to this measurement was the improvement of the misreconstructed tau background, which was determined directly from data. The study of this final state was useful to the author to familiarize with the lepton+tau final state, which is similar to the final state examined in the search for a *light* charged Higgs. In fact, the eventual presence of the charged Higgs boson in the data could be signaled by an increase of the top quark pair cross section in this final state. Therefore, an increased precision in the measurement gives an enhanced sensitivity to the charged Higgs searches.

Finally, a search for a charged Higgs boson with a mass larger than the top quark mass has been performed using the 8 TeV data using two main final states: the lepton+tau final state (characterized

by a lower sensitivity in this mass region), and the dilepton final state. In this mass regime, for the majority of the theoretical scenarios, the charged Higgs boson decays either into a tau and a neutrino, or into a top quark and a b quark. The production mechanism changes as well, and is characterized by the associate production of a top quark and a charged Higgs, sometimes with an additional b quark. With the subsequent charged Higgs decay into a top quark and a b quark, the final state contains a top quark pair and at least one b quark. This justifies the use of the dilepton final state as a flagship for the search strategy. The result of this search has been made public as a preliminary result [50], and represented the first direct search for a heavy charged Higgs boson decaying into a top quark and a b quark. Since then, other teams within the CMS collaboration have obtained results using other final states: the dilepton result has been combined with the later results, and published resulting in the first published direct search for a heavy charged Higgs boson [51]. The search for a heavy charged Higgs boson is described in Sec. 7. The author has been the responsible for the full analysis of the lepton+tau and dilepton final state, obtaining the first direct search for $\mathrm{H}^{+} \rightarrow \mathrm{t}\overline{\mathrm{b}}$ as a preliminary result, and has been the responsible of the paper and of the statistical combination of his results with the later results from other groups in the CMS collaboration presented in the same paper [51]. Those additional final states are described in App. A and B.

Chapter 2

Theoretical introduction

2.1 How experiments guided the discovery of the building blocks of our universe

In the beginning of the last century, as outlined in [52], the experimental knowledge about the constituents of matter was quite limited. Only a few particles were known: the ones constituting the atom (electron, proton, neutron), and the photon. Furthermore, the existence of the neutrino was conjectured in order to guarantee energy and momentum conservation in beta decays.

During the 1930s, cosmic ray physics started to yield shattering discoveries of new particles, mostly not forecast before. In the meanwhile, the first particle accelerators were being developed and put at work: a race, still ongoing, then started to increase the beam energy of the collisions, thus giving access to the production of heavier particles as a result of the collisions. Fig. 2.1 shows the evolution of the beam energies along the years.

A humungous number of new, unforecast, particles was discovered, and soon the problem of classification arose. A first attempt involved a classification based on the mass of the particle, yielding to categories such as leptons ("low" mass), mesons ("medium" mass) and baryons ("high" mass). With time, though, the focus shifted from the mass of the particle to its typical interactions with other particles: thus, currently some leptons (namely the τ) are heavier than baryons like the proton.

The current nomenclature calls "leptons" the fermions that do not interact strongly with the nuclei of atoms, and "baryons" the fermions that do. "Mesons" are currently bosons subject to strong interactions with the nuclei. Given their similar pattern of interaction, baryons and mesons are together called "hadrons".

While studying all these particles, some of their characteristic properties (that now we usually identify with some quantum number) were found to be conserved in interactions and decays: some properties always (e.g. electric charge), some others (e.g. the product of charge and parity) are usually conserved but sometimes the conservation is violated.

The study of conserved quantities received a large impulse by the demonstration by Emmy Noether that, when describing a system through the hamiltonian formalism, any symmetry of the hamiltonian

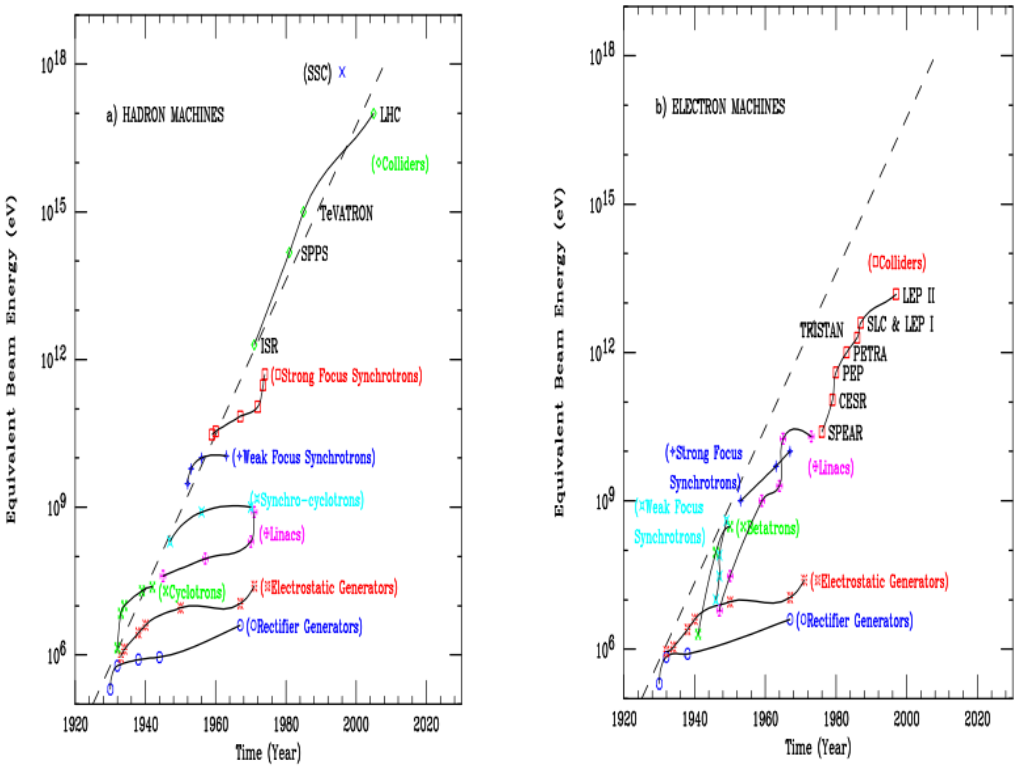


Figure 2.1: Livingston Plot: equivalent beam energy of colliders versus the year of their introduction, for (a) Hadron Machines and (b) for Lepton Machines. The figures are taken from Ref. [3]

(hence of the system) results in a conserved quantity. Group theory became a fundamental tool for studying the classification of elementary particles and their interactions through the study of the symmetry of particular groups such as the Unitary and the Special Unitary groups, whose matrix representation are particularly apt for studying hermitians operators such as the Hamiltonian: a detailed review of group theory, including a discussion of the (Special) Unitary groups, can be found in [53] in Italian, while a brief review in English can be found in [52] This fundamental result, once extended to quantum mechanical systems and for both continuum and discrete symmetries, permitted to study experimentally conserved quantities in terms of symmetries of the hamiltonian describing the particle. For example, Heisenberg himself proposed to formalize proton and neutron physics in terms of two states of a single particle called "nucleon", and characterized by a symmetry ("isospin" symmetry) with respect to rotation in the space defined by the two states. The short range nuclear interaction would be, in this framework, explained by the exchange of a massive mediator identified with the pion.

A classification based on isospin was set up also by Gell-Mann and Ne'eman for mesons and baryons (collectively called hadrons). Some particles characterized by large masses and large pair production cross section had been found to have lifetimes far larger than the ones forecast in case they were nuclear resonances: they are produced via strong interactions, but decay weakly. After postulating the existence of a new quantum number ("strangeness"), those particle have been grouped in multiplets based on isospin and strangeness. Furthermore, an empirical relationship has been identified by Gell-Mann and Nishijima, that links the electric charge Q, the third component of the isospin, I_3, and the hypercharge

Y defined as the sum of baryonic number[1] and strangeness:

$$Q = I_3 + \frac{1}{2}Y \tag{2.1}$$

When building the isospin octets for particles of the same spin and parity, one would expect to have in each multiplet particles with the exact same mass: experimentally, though, the particles that constitute the various multiplets (respectively: Kaons and pions; neutron, proton, Σ, Ξ) differ by 130–150 MeV. This is a sign that the isospin symmetry for hadrons is only an approximate one.

Following this realization, a fundamental representation for the hadron multiplets was proposed by Gell-Mann and Zweig in terms of a triplet of spin $\frac{1}{2}$ particles called "quarks", together with the conjugate representation (triplet of "antiquarks"). This classification sets the ground for introducing the so-called "flavour" symmetry, described via the corresponding $SU(3)_{flavour}$ group.

An additional symmetry (and the corresponding quantum number, called "colour") is the one described by the $SU(3)_{colour}$ group, and has been introduced to explain the symmetricity of the total wave function of some observed identical fermions characterized by a symmetric total wave function, in contradiction with the Pauli exclusion principle.

In this framework, quarks are represented by complex three-vectors, and are assigned a colour quantum number that can assume three values, usually referred to as "Red", "Green", "Blue". Antiquarks have the corresponding anticolours ("Antired", "Antigreen", "Antiblue"). The interactions between quarks happen as a result of the exchange of eight colour field bosons, called gluons, that correspond to the eight generators of the $SU(3)_{colour}$ group. The theory was originally developed for ground states (states with orbital angular momentum $l = 0$), but has been shown to work as well for the categorization of the hundreds of multiquark excited states that have been discovered in the decades.

Until the late 60s, quarks were thought to as mostly a mathematical construct (even by Gell-Mann himself), although in the 50s the experiments with electron scattering performed by Hofstader showed that the proton has an anomalous magnetic structure [54]: the beam energy was unfortunately too small to resolve the proton structure, but later experiments were able to confirm that the proton has internal structure.

The simple parton model, in which nucleons are constituted by three charged quarks (partons) accounting for the whole momentum of the nucleon, was probed via further deep inelastic scattering experiments, leading to the observation that the three partons account for only around 50% of the momentum of the nucleon: the model was thus modified to take into account the momentum carried by the bosons mediators of the strong force, the gluons. A more realistic picture takes into account also the fact that, in time scales hidden by the Heisenberg indetermination principle, within the nucleon there is a continuous production and annihilation of quark pairs: in this picture, the three partons are called "valence quarks", and the quarks resulting from pair production-annihilation are called "sea quarks".

[1] The baryonic number B is set to be 0 for mesons, $+1$ for baryons, and -1 for anti-baryons

2.2 The current picture

Currently, we model the world in terms of two types of particles [55]: the constituents of matter itself, that may be called *matter particles*, and the *interaction particles* that explain the way in which matter particles interact with each other.

Matter particles have spin $\frac{1}{2}$, thus being fermions: they are divided into leptons and quarks, and the ones known so far are globally distributed in three set scalled *families*.

Each family is constituted by two quarks and two leptons. The first family describes completely the ordinary matter (the matter that can be experienced by human beings without any instrument): the elements of the other two families have intra-family relationships that mimic the ones of the first family. The first family is constituted by one quark with have electrical charge $Q = +\frac{2}{3}$, thus being called *up quark*, and by another with $Q = -\frac{1}{3}$, called *down quark*. One lepton has electrical charge $Q = -1$ and is called *electron* (actually, in this manuscript electrical charge is expressed by purposefully setting the unitary charge to be that of the electron, which is a physics constant denoted by *e* [56]) whereas the other is electrically neutral ($Q = 0$) and is called neutrino.

As mentioned, the other two families mimic the structure of the first family, being characterized by one *up type* quark, one *down type* quark, one *electron-like* lepton and one neutrino. Each element is heavier than the corresponding element of the previous family, and has same electrical charge. Each family differs from the other by the increasing mass of the corresponding constituents.

Experimentally, the last family (the one made up by the heaviest constituents) is the one that differentiates itself the most from the other ones. The up-type quark of the third family, called *top quark*, decays by electroweak interaction before having the time to fragment via strong interaction. The down-type quark of the third family, called *bottom quark* still fragments, but due to its mass has a lifetime long enough that its experimental signature at a collider includes fragmentation products that come from a different point than the main collision vertex. The electron-like lepton, called *tau lepton*, decays mainly into hadrons, on the contrary of electron and muon.

As mentioned in the previous section, each quark have a colour quantum number, that can assume three values, bringing the actual number of quarks belonging to each family to six.

For each of the forementioned particles, an *antiparticle* with opposite electrical charge is also part of the family.

The interaction particles are tightly liked to the elementary interactions. Leaving the gravitational interaction aside (because it has not been successfully described by a field theory yet, and because anyways it is too weak to have any experimentally detectable effect in high energy collisions), all the interactions involved in particle physics are described as being mediated by the exchange of an intermediate particle, called *force carrier*, that happens to be a boson with spin 1. The photon, denoted by γ, is the mediator of the electromagnetic interaction; the eight gluons are the mediators of the strong interaction among quarks; and the three weak bosons, $\mathrm{W}^{\pm}$ and Z, are the mediators of the weak interaction responsible of β decays.

Finally, the recently discovered Higgs boson (with has spin 0, on the contrary of the forementioned

weak bosons), is the expression of the mechanism that confers mass to all elementary particles.

Neutrino were generally thought to be massless, until recent experiments confirmed that they are subject to flavour oscillation. Nevertheless, there are models that explain neutrino oscillations by allowing neutrinos to be massless: in such models, the oscillation would be generated by torsion of the tensor structure of the space-time [57].

2.3 The Standard Model in a nutshell

Formally, the Standard Model is a gauge field theory based on the $SU(3)_C \times SU(2)_L \times U(1)_Y$ group, that is composed by the $SU(3)_C$ group, that is the symmetry group governing the strong interactions, and by the $SU(2)_L \times U(1)_Y$ group, that governs the electroweak interactions. The group governing the electromagnetic interactions, $U(1)_{em}$, is formally a subgroup of $SU(2)_L \times U(1)_Y$.

The range of each interaction is governed by the mass of the mediator: the electromagnetic interaction has infinite range, being mediated by the massless photon; the weak interaction has a range of about 10^{-18} m, that corresponds to a mediator boson mass of ~ 100 GeV (which is indeed the mass range of the weak bosons); the strong interaction should have an infinite range, being mediated by the massless gluons, but has instead a range of about 10^{-15} m, due to a phaenomenon called *confinement*. This is a consequence of the colour structure of the quark sector: when separating two colour-charged particles, the colour fields between them favours the creation of a new particle-antiparticle pair in a short interaction with the originating particles instead of a long range interaction between the two parent particles.

However, the strength of the strong interaction is governed by the size of the strong coupling constant α_s, that varies from ~ 1 at low energies to 0 in the infinite energy limit. This implies that when quarks are observed at (infinitely) large energies (or equivalently at infinitely short distances) they behave as free particles, and is particularly relevant in modern particle colliders: in fact, the energies and the ranges involved in proton-proton collisions at energies of the order of the TeV permit to treat the quarks as free particles, not bound into hadrons.

The strength of the electromagnetic interaction is governed by the size of the electron charge, which is the electromagnetic coupling constant and can be equivalently expressed as $\alpha = \frac{e^2}{2\pi}$. At low energies, $\alpha \simeq \frac{1}{137}$. The strength of the weak interactions, at energies much lower than the mediator boson mass, is governed by the Fermi constant $G_F \simeq 10^{-5}$ GeV^{-2}.

The gauge sector of the SM is characterized by three massive weak gauge bosons, and by one massless electromagnetic gauge boson. The mass of the weak boson implies that $SU(2)_L \times U(1)_Y$ is not a symmetry of the vacuum, and therefore that it must be broken. The masslessness of the photon implies instead that $U(1)_{em}$ is a symmetry of the vacuum, and therefore it must be present after having broken the $SU(2)_L \times U(1)_Y$ one. For these reasons, the spontaneous symmetry breaking in the SM is

$$SU(3)_C \times SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \times U(1)_{em} \tag{2.2}$$

The symmetry is thought to be broken through a mechanims called Higgs mechanism, which generates the masses of the massive weak gauge bosons and of the fermions through their interaction with a new field, called Higgs field. The Higgs field is described by a spin-0 boson, the Higgs boson, that has been discovered recently at the LHC by the CMS and ATLAS collaborations.

As a historical remark, the Higgs mechanism has been theoretized by six physicists in the '60s: Brout, Englert, Guralnik, Hagen, Higgs, and Kibble. Since that time, the mechanism has been uniquely known and named after Peter Higgs. Only during the frenetic months before the experimental discovery, when hints that the discovery was close were already in the air, the issue of the full attribution of the "Higgs mechanism" was raised among the scientific community, leading to the forceful replacement of *Higgs boson* with *BEH boson* or *the scalar boson* in conference talks and proceedings. The author of this manuscript strongly thinks that, while the issue of attribution is a delicate one and must be settled in a fair way, it does not make any real sense to change a five-decades-long unanimous nomenclature; he also points out that, once the hype of the discovery settled out, the scientific community spontaneously resettled on calling the particle *Higgs boson* in articles and conferences. For these reasons, in this manuscript Higgs bosons are called Higgs bosons, and not BEH bosons or "the scalar bosons".

The lagrangian of the Standard Model of particle physics, and the Higgs mechanism, are beyond the scope of the current work. The interested reader can find a thorough description in Ref. [55]. The Higgs mechanism is especially examined in the wonderful review by Abdelhak Djouadi [58].

2.4 Extending the Standard Model

2.4.1 Introducing the Two-Higgs Doublet Models

There are many ways of extending the SM. The most immediate one consists in adding scalar doublets and singlets: extensions characterized by two doublets, generically called Two-Higgs Doublet Models (2HDM) are well motivated [1, 59]. In SUSY reference models (for example in the MSSM [24, 25, 26, 27, 28, 29, 30, 31]), the existence of a fermion/boson symmetry is postulated, leading each particle to have a partner with spin differing by half integer (hence turning fermions into bosons and vice versa): due to the chirality multiplet structure of such models, two doublets are needed in order to give mass to quarks with both charge $\frac{2}{3}$ and $-\frac{1}{3}$, and in order to have anomalies cancellation. Axion models (for example in the Peccei-Quinn model), exhibit CP-violation in $\mathcal{L}_{\mathcal{QCD}}$, that is known from experiments to be very small [1]: the CP-violation present in the theory can be eliminated if $\mathcal{L}_{\mathcal{QCD}}$ has $U(1)$ symmetry, which requires two doublets. Although many axion models have been ruled out, this mechanism is characteristic also of effective low-energy limits, that have not been ruled out yet. Furthermore, baryon asymmetry is unexplained in the SM, and it can be generated (with a size sufficient to give origin to the present asymmetry) in 2HDM, because of the flexible mass spectrum and the additional sources of CP violation provided by such models. Neutrino oscillations and the presence of dark matter constitute additional problems that can be solved via the introduction of an additional doublet in the SM. Finally, 2HDMs exhibit a rich vacuum structure, and various scalar potentials can be explored, leading to probing

even situations with charge-violating minima.

2.4.2 Building the 2HDM potential

In order to build the 2HDM potential, it is enough to add one scalar doublet to the Standard Model: no additional symmetries are imposed, which excludes the 2HDM from being cathegorized as a SUSY model. As mentioned above, though, it is worth to note that by converse SUSY models are usually characterized by two doublets. The most general Higgs potential is characterized by fourteen parameters, yielding many possibilities for the minima (CP-conserving, CP-violating, charge-violating, and so on). Some of these parameters can be rotated away via phase transformations, whereas others have physical implications.

When discussing 2HDM models, it is important to keep in mind that the results usually depend on a few simplifying assumptions that are often made. First of all, it is assumed that CP is not spontaneously broken, and that CP is conserved in the Higgs sector: this is necessary in order to be able to distinguish between scalars and pseudoscalars. This leads to the simplified potential for the doublets Φ_1 and Φ_2, each with hypercharge $+1$

$$\begin{aligned} V_{2HDM} = m_{11}^2\Phi_1^\dagger\Phi_1 + m_{22}^2\Phi_2^\dagger\Phi_2 - \Big[m_{12}^2\Phi_1^\dagger\Phi_2 + h.c.\Big] \\ + \frac{1}{2}\lambda_1\Big(\Phi_1^\dagger\Phi_1\Big)^2 + \frac{1}{2}\lambda_2\Big(\Phi_2^2\Phi_2\Big)^2 \\ + \lambda_3\Big(\Phi_1^\dagger\Phi_1\Big)\Big(\Phi_2^\dagger\Phi_2\Big) + \lambda_4\Big(\Phi_1^\dagger\Phi_2\Big)\Big(\Phi_2^\dagger\Phi_1\Big) \\ + \frac{1}{2}\lambda_5\Big[\Big(\Phi_1^\dagger\Phi_2\Big)^2 + h.c.\Big] \\ + \Big\{\Big[\lambda_6\Big(\Phi_1^\dagger\Phi_1\Big) + \lambda_7\Big(\Phi_2^\dagger\Phi_2\Big)\Big]\Big(\Phi_1^\dagger\Phi_2\Big) + h.c.\Big\} \end{aligned} \tag{2.3}$$

A further simplification is made, that the quartic terms that are odd in either of the doublets can be canceled via discrete symmetries. This leads to the final form of the simplified potential for the 2HDM

$$\begin{aligned} V_{2HDM} = m_{11}^2\Phi_1^\dagger\Phi_1 + m_{22}^2\Phi_2^\dagger\Phi_2 - \Big[m_{12}^2\Phi_1^\dagger\Phi_2 + h.c.\Big] \\ + \frac{1}{2}\lambda_1\Big(\Phi_1^\dagger\Phi_1\Big)^2 + \frac{1}{2}\lambda_2\Big(\Phi_2^\dagger\Phi_2\Big)^2 \\ + \lambda_3\Big(\Phi_1^\dagger\Phi_1\Big)\Big(\Phi_2^\dagger\Phi_2\Big) + \lambda_4\Big(\Phi_1^\dagger\Phi_2\Big)\Big(\Phi_2^\dagger\Phi_1\Big) \\ + \frac{1}{2}\lambda_5\Big[\Big(\Phi_1^\dagger\Phi_2\Big)^2 + h.c.\Big] \end{aligned} \tag{2.4}$$

The minimization of this potential in a given region of parameter space yields two vacuum expectation values

$$\langle\Phi_{1,2}\rangle_0 = \begin{pmatrix} 0 \\ \frac{v_{1,2}}{\sqrt{2}} \end{pmatrix} \tag{2.5}$$

where the two complex scalar SU(2) doublets can be described by eight fields

$$\Phi_{1,2} = \begin{pmatrix} \phi_{1,2}^+ \\ \frac{(v_{1,2}+\rho_{1,2}+i\eta_{1,2})}{\sqrt{2}} \end{pmatrix} \tag{2.6}$$

The application of the spontaneous symmetry breaking mechanism to this potential results in three degrees of freedom (three fields) "eaten" to give mass to the $\mathrm{W}^{\pm}$ and Z^{0} gauge bosons. The Higgs sector is characterized by five physical scalar Higgs fields: two neutral scalars (h,H), two charged scalars ($\mathrm{H}^{\pm}$), one pseudoscalar (A).

The mass terms of the lagrangian for the charged scalars are of the form

$$\mathcal{L}_{\phi^{\pm}mass} = \Big[m_{12}^2 - (\lambda_4+\lambda_5)v_1v_2\Big]\Big(\phi_1^-,\phi_2^-\Big)\begin{pmatrix} \frac{v_2}{v_1} & -1 \\ -1 & \frac{v_1}{v_2}\end{pmatrix}\begin{pmatrix}\phi_1^+\\ \phi_2^+\end{pmatrix} \tag{2.7}$$

The zero eigenvaleue corresponds to the charged Goldstone boson that results in the massive $\mathrm{W}^{\pm}$. Writing the mass term for the physical charged Higgs bosons results in their masses to be expressed as

$$m_+^2 = \Big[\frac{m_{12}^2}{v_1v_2} - \lambda_4 - \lambda_5\Big]\Big(v_1^2+v_2^2\Big) \tag{2.8}$$

The mass term for the pseudoscalars can be expressed as

$$\mathcal{L}_{\eta mass} = \frac{m_A^2}{v_1^2+v_2^2}\Big(\eta_1,\eta_2\Big)\begin{pmatrix} v_2^2 & -v_1v_2 \\ -v_1v_2 & v_1^2\end{pmatrix}\begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix} \tag{2.9}$$

that can be expressed in terms of the physical pseudoscalar, yielding a mass

$$m_A^2 = \Big[\frac{m_{12}^2}{v_1v_2} - 2\lambda_5\Big]\Big(v_1^2+v_2^2\Big) \tag{2.10}$$

It has to be noted that in case $m_{12}^2 = 0$ and $\lambda_5 = 0$, then the physical pseudoscalar is massless. Furthermore, in that limit, an additional U(1) global symmetry is broken.

For the neutral scalars the mass term has the form

$$\mathcal{L}_{\rho mass} = -\Big(\rho_1,\rho_2\Big)\begin{pmatrix} m_{12}^2\frac{v_2}{v_1} + \lambda_1 v_1^2 & -m_{12}^2 + \lambda_{345}v_1v_2 \\ -m_{12}^2 + \lambda_{345}v_1v_2 & m_{12}^2\frac{v_1}{v_2} + \lambda_2 v_2^2\end{pmatrix}\begin{pmatrix}\rho_1\\ \rho_2\end{pmatrix} \tag{2.11}$$

where $\lambda_{345} := \lambda_3 + \lambda 4 + \lambda 5$. This mass-squared matrix can be diagonalized, and an angle α can be defined to be the rotation angle that performs the diagonalization. The charged scalars and pseudoscalar mass-squared matrices can be diagonalized as well via an additional rotation angle β, which is related to the vacuum expectation values through its tangent

$$tan\beta := \frac{v_2}{v_1} \tag{2.12}$$

The two angular parameters α and β determine the interactions of the five Higgs fields with the vector bosons and (given the fermion masses) the fermions. The Yukawa couplings of, e.g., the $Q = -\frac{1}{3}$ quarks are

$$\mathcal{L}_Y = y_{ij}^1\bar{\psi}_i\psi_j\Phi_1 + y_{ij}^2\bar{\psi}_i\psi_j\Phi_2 \tag{2.13}$$

where i, j are generation indexes. The corresponding mass matrix is

$$\mathcal{M}_{ij} = y^1_{ij}\frac{v_1}{\sqrt{2}} + y^2_{ij}\frac{v_2}{\sqrt{2}} \tag{2.14}$$

In the SM, the diagonalization of the mass matrix yields automatically the diagonalization of the Yukawa interactions. In 2HDMs, instead, in general y^1 and y^2 are not simultaneously diagonalizable, and consequently the Yukawa couplings are not in general flavour-diagonal. This implies that there will be tree-level flavour-changing neutral currents (FCNCs) with implications that are phenomenologically problematic (e.g. $K - \bar{K}$ tree level mixing with an exchanged 10 TeV scalar). In order to avoid these tree-level FCNCs, the Paschos-Glashow-Weinberg theorem can be exploited: this theorem states that a necessary and sufficient condition for the absence of tree-level FCNCs is that all fermions of a given charge and helicity transform according to the same irreducible representation of SU(2), correspond to the same eigenvalue of T_3, and that a basis exists in which they receive their contributions in the mass matrix from a single source. In layman's terms, this means that FCNCs are avoided if all fermions with the same quantum numbers (i.e. the ones that are capable of mixing) couple to the same Higgs multiplet. In 2HDMs, it is possible to satisfy the Paschos-Glashow-Weinberg theorem in two ways:

- all quarks couple to just one of the Higgs doublets (conventionally Φ_2). This situation is denoted as Type I 2HDM;

- the $Q = \frac{2}{3}$ RH quarks couple to one Higgs doublet (conventionally Φ_2), and the $Q = -\frac{1}{3}$ RH quarks couple to the other (Φ_1). This situation is denoted as Type II 2HDM;

In order to realize these conditions, it is necessary, in case of the Type I 2HDMs, to introduce a $\Phi_1 \to -\Phi_1$ discrete symmetry. In case of the Type II 2HDMs, it is necessary to introduce instead a $\Phi_1 \to -\Phi_1$, $d^i_R \to -d^i_R$ discrete symmetry. It is worth to note that Peccei-Quinn and SUSY models (e.g. MSSM) have the same Yukawa couplings as the Type II 2HDMs, but in order to satisfy the Paschos-Glashow-Weinberg theorem continuous symmetries are introduced, instead of discrete ones.

The picture is further complicated by noting that Type I and II 2HDMs are defined by assuming that right-handed leptons behave and couple in the same way as d^i_Rs: the Paschos-Glashow-Weinberg theorem then yields two more possibilities, illustrated in Table 2.1, taken from Ref. [1].

Table 2.1: Models which lead to natural flavour conservation. The superscript *i* stands for the generation of the quark and leptons. By convention, u^i_R always couples to Φ_2. This table is taken from Ref. [1].

Model	$u_R i$	$d_R i$	$e_R i$
Type I	Φ_2	Φ_2	Φ_2
Type II	Φ_2	Φ_1	Φ_1
Lepton-specific	Φ_2	Φ_2	Φ_1
Flipped	Φ_2	Φ_1	Φ_2

2.4.3 The MSSM

The MSSM aims to be the minimal (i.e. most general) SUSY model that can be built [60]. It is constituted by the Standard Model and by the minimal particle content that is needed to be added in order to feature the breaking of a super-symmetry.

Gauge bosons are promoted to multiplets. Quarks and leptons are promoted to chiral supermultiplets, but in order to do so it is first necessary to charge-conjugate the right-handed fields in order to have all fields described as left-handed Weil spinors. The Higgs sector is analogous to the 2HDM, and the lightest neutral Higgs is predicted to have a mass $m_{h^0} < m_Z$. However, one-loop level effects modify this prediction (that would have already ruled out experimentally) by pushing the lightest neutral Higgs boson mass to about 130 GeV (for a scalar top mass of up to 1 TeV), which is compatible with the observed Higgs boson $h(125)$. Taking into account these experimentally-driven developments, the Higgs sector of the MSSM can be described, after including the one-loop corrections, by only two parameters [60]: one of them is normally the ratio of the v.e.v.s, $tan\beta$, and the other can be the mass of the charged Higgs boson. Furthermore, to a good approximation, the Higgs sector of the MSSM can still be described by two parameters even when the full set of radiative corrections to the Higgs masses at the two–loop level is included, as demonstrated in Ref. [61].

Overall, apart from the Higgs sector, the full MSSM still needs many parameters to be fixed in order to obtain an experimentally testable scenario. The main benchmark scenarios relevant to experimental searches at the LHC are described in Ref. [11], and probed in Sec. 6 and Sec. 6, although the searches object of this manuscript aim to be as model independent as permitted by each final state that has been investigated.

2.4.4 Main charged Higgs productions modes

The most prominent production mechanism for the search of a charged Higgs is the one related with top quark physics, as mentioned. When the charged Higgs is lighter than the top quark mass, $m_{\mathrm{H}^+} < (m_{\mathrm{t}} - m_{\mathrm{b}})$, the charged Higgs is produced within the top quark decay, where a W boson is substituted by the charged Higgs boson. When the charged Higgs is heavier than the top quark mass, $m_{\mathrm{H}^+} > (m_{\mathrm{t}} - m_{\mathrm{b}})$, it is produced in association with a top quark, and sometimes a bottom quark. This production mechanism is discussed in Sec. 7.1.

A charged Higgs boson can also be produced in the frame of vector boson fusion or decay. Cascade decays of a MSSM heavy neutral Higgs decaying into a charged Higgs and a W boson have been studied recently by the ATLAS Collaboration [62]

Vector boson fusion, in which a W boson and a Z boson from the beam fuse into a charged Higgs boson, have been recently studied as well by the ATLAS Collaboration [63]. This production mechanism and subsequent decay mode is particularly interesting because it is possible to intepret it in the framework of the Georgi-Machacek triplet model [64, 65] in terms of fraction of m_W^2 and m_Z^2 generated by the triplet vacuum expectation value (fraction which is proportional to the charged Higgs cross section and width).

2.4.5 Phenomenology of the tau lepton

In most MSSM scenarios, the dominant decay mode for a charged Higgs boson with mass $m_{\mathrm{H}^+} < (m_{\mathrm{t}} - m_{\mathrm{b}})$ is to a tau lepton and a neutrino. This decay mode is also the second-dominant decay mode for a heavy charged Higgs boson, $m_{\mathrm{H}^+} > (m_{\mathrm{t}} - m_{\mathrm{b}})$, surpassed only by the decay into a top and a bottom quark. The phenomenology of the tau lepton can thus be exploited in order to achieve a good discrimination power between Standard Model events and events characterized by a charged Higgs boson decaying into a tau lepton and a neutrino.

The hadronic decay of the tau lepton is dominated by the 1-prong decay mode [66, 67] (i.e. the decay mode with one charged partic le as decay product, plus some neutrals), that accounts alone for 50% of the full tau decay width (being 65% the decay width into hadrons). In that decay, the center-of-mass angular distribution of the decay depends on the polarization of the tau lepton: for the decays into a pion or into a vector meson v (ρ or a_1),

$$\frac{1}{\Gamma_\pi}\frac{d\Gamma_\pi}{dcos\theta} = \frac{1}{2}(1 + P_\tau cos\theta)\frac{1}{\Gamma_v}\frac{d\Gamma_{vL,T}}{dcos\theta} = \frac{\frac{1}{2}m_\tau^2, m_v^2}{m_\tau^2 + 2m_v^2}(1 \pm P_\tau cos\theta) \tag{2.15}$$

Here, L and T denote the longitudinal and transverse polarization states for the vector meson. One can then write the expression for the fraction of the tau momentum carried by the decay meson, in the collinear approximation in which $p_\tau >> m_\tau$,

$$x = \frac{1}{2}(1 + cos\theta) + \frac{m_{\pi,v}^2}{2m_\tau^2}(1 - cos\theta) \tag{2.16}$$

Taus coming from the decay of a charged Higgs boson maintain its polarization, i.e. $P_\tau = 1$, whereas SM background processes exhibit taus with $P_\tau = -1$, and fake taus from misidentified jets have no preferentiality. From Eqs. 2.15, 2.16 it can be seen easily that the positive polarization contribution comes from pions and longitudinally polarized vector mesons, whereas the negative tau polarization component comes from transversally polarized vector mesons.

The longitudinal and transverse polarization states for vector mesons have an interesting characteristic, though: the transversally polarized ρ_T and a_{1T} decay favouring an even sharing of momentum among their decay pions, whereas the longitudinally polarized ρ_L and a_{1L} decay favouring a very uneven distribution: the charged pion carries either very little, or most of the momentum.

Consequently, variables like

$$R_\tau = \frac{p_{\pi^\pm}}{p_{\tau-jet}} \tag{2.17}$$

that are defined between 0 and 1, are expected to exhibit a broad peak in the middle for background processes, including jets misidentified as taus, and to be peaked at either 0 or 1 for a charged Higgs signal process [66].

However, the peak near 0 is not detectable, because tau reconstruction at CMS operates a minimum cut on the transverse momentum of the charged tracks for tau identification (1 GeV, see Sec. 3.4): the distribution for the signal is then expected to peak only at 1.

Among the background processes, the one with taus coming from top quark pair decays are expected to have the broad peak in the middle of the distribution, corresponding to the negative polarizations of the tau, as well as a peak at 1, corresponding to the positive polarization of the tau, that is still present (although not dominant) in $\mathrm{t\bar{t}}$ decays.

This variable, R_τ, is exploited in Chap. 6 and Chap. 6 for separating the charged Higgs signal from the SM backgrounds.

Chapter 3

From the accelerator to the laptop

3.1 The LHC and the CMS detector

This thesis makes use of an experimental setup constituted by a proton-proton collider machine and a detector covering almost all the solid angle around the collision point.

Protons are accelerated and brought to collision by the Large Hadron Collider (LHC), located outside Geneva (Switzerland) and spanning the Swiss-French border. Protons are accelerated in smaller linear and cyclical accelerators before being injected into the LHC, into two different beams circulating in opposite directions through the ring. The protons are brought to collision at four spots along the LHC by intersecating the beams. Surrounding one of these spots is the Compact Muon Solenoid (CMS) detector, which consists of multiple subsystems which work together to identify signatures of different types of particles.

3.2 Large Hadron Collider

A full description of the LHC can be found in [68]; in this section, its main structure is briefly outlined. The LHC is a two-ring superconducting synchrotron designed to collide particles at high energy and high luminosity. It sits in a 26.7 km tunnel located 45–170 m underneath the Swiss-French countryside outside of Geneva, Switzerland. The LHC can produce collisions with either protons or heavier ions. This leads to three possible operational modes, proton-proton, ion-ion, and proton-ion. The production of top quarks and possibly of charged Higgs bosons can be probed with great precision in the proton-proton mode, which is consequently the only mode discussed in this thesis.

The LHC was designed to accelerate protons to an energy of 7 TeV and collide them at a center–of–mass energy ($\sqrt{s}$) of 14 TeV. The protons are brought to a collision at four points along the LHC beam line. Surrounding two of these interaction points sit the general purpose detectors labelled CMS and A Toroidal LHC ApparatuS (ATLAS). These detectors are designed to record collisions at the highest instantaneous luminosity the LHC can supply: the design value for the instantaneous luminosity is 10^{34} cm^{-2} s^{-1}. The other interaction points are surrounded by the special purpose detectors Large

Hadron Collider beauty experiment (LHCb) and A Large Ion Collider Experiment (ALICE) and are designed to have instantaneous luminosities of 2×10^{29} cm^{-2} s^{-1} and 10^{27} cm^{-2} s^{-1}, respectively. This thesis takes into account only data collected by the CMS detector.

The acceleration of protons to their final energy of 7 TeV is done in series of steps employing smaller accelerators located on the Conseil Europeen pour la Recherche Nucleaire (CERN) site. The protons originate in the linear accelerator Linac2, which accelerates them to an energy of 50 MeV. From there, they are passed through a series of synchrotron accelerators: the Proton Synchrotron Booster, the Proton Synchrotron, and the Super Proton Synchrotron, with their energy raised to 1.4 GeV, 25 GeV, and 450 GeV, respectively. After passing through the Super Proton Synchrotron the protons are injected into the LHC. According to the design specification, the LHC should finally accelerate the protons to their final design energy of 7 TeV, however only an energy of 6.5 TeV has been achieved so far.

The beams are designed to contain proton bunches spaced such that collisions at the interaction points occur every 25 ns. The LHC can hold a total of 2,808 bunches; in some places it is designed to have gaps larger than 25 ns between bunches to allow for dumping of the beam without harming the LHC. Each 25 ns time window is referred to as a bunch crossing window, whether there are proton bunches colliding in CMS or not. The time resolution of the detector is larger than the time the bunches takes to cross by $\mathcal{O}(1)$ ns: as a consequence, each collision between the proton bunches can result in more than one proton-proton collision. This results in CMS seeing numerous proton-proton collisions overlaid on one another. At collision, the bunches have a longitudinal length of 9 cm and radius of 20 μm [56], both numbers are RMS values. This results in the collisions in a bunch crossing being spread over a time period of a few tenths of a nanosecond.

The commissioning of the LHC saw it run at progressively higher energies building towards the design energy. In 2008, the LHC ran at $\sqrt{s} = 900$ GeV and for a short period at 2.36 TeV. Then after further work on the LHC, the center–of–mass energy was raised to 7 TeV for both 2010 and 2011 and then to 8 TeV in 2012. This manuscript only covers the data collected at 7 and 8 TeV in 2011 and 2012. It is planned to raise the center–of–mass energy to its design goal of 14 TeV through additional work on the LHC and the injector system. At the time of writing of this manuscript, the LHC has started running at a center–of–mass energy of 13 TeV, and will keep doing so for about one more year.

Similarly, the instantaneous luminosity was ramped up during the commissioning phase. In 2010, the maximum instantaneous luminosity achieved [69, 70, 71] was 2.1×10^{32} cm^{-2} s^{-1} and in 2011 it was 3.5×10^{33} cm^{-2} s^{-1} During the 2012 running, the maximum instantaneous luminosity reached was 7.7×10^{33} cm^{-2} s^{-1}. The spacing between the collisions in CMS has been of 50 ns in 2011 and 2012 collisions. At 13 TeV, after a first test run of about 100 pb^{-1} that has been performed with a spacing of 50 ns, the accelerator is being run with a 25 ns bunch spacing.

3.3 Compact Muon Solenoid

The CMS detector is built around one of the interaction points of the LHC. A full description of CMS be found in references [72, 73].

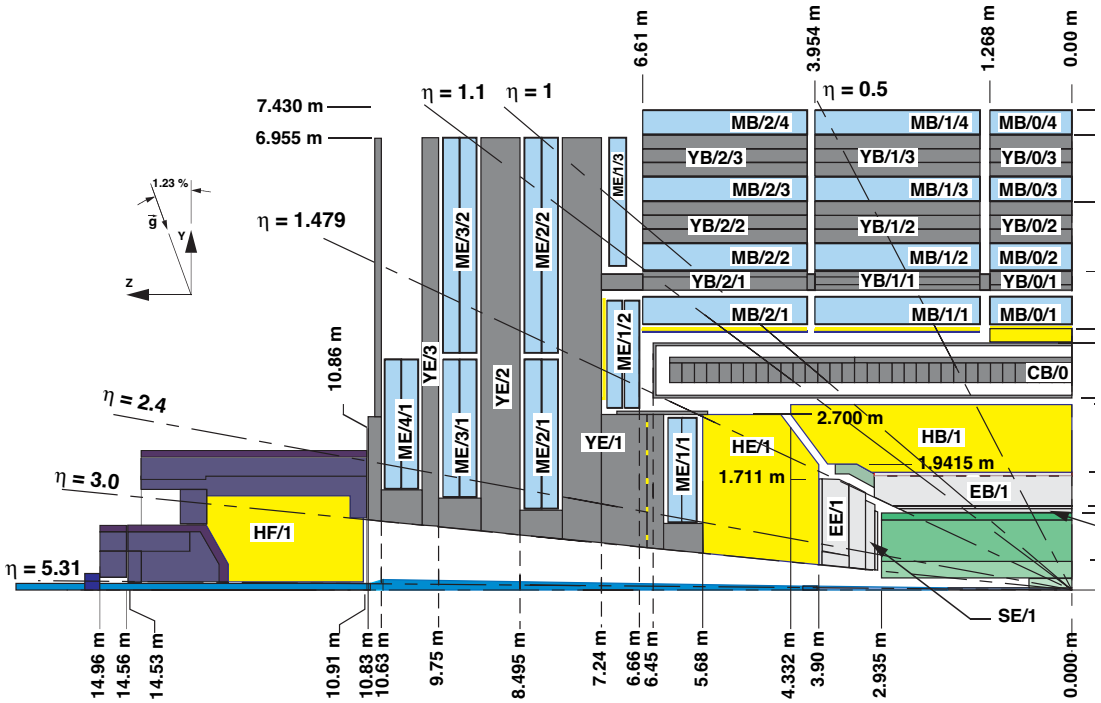


Figure 3.1: Cross-sectional view in r-z of a quarter of the CMS detector. Protons enter CMS along the bottom of the figure and are brought to a collision in the bottom right corner. The inner silicon tracker is in the bottom right in green. The electromagnetic calorimeter and hadronic calorimeter are outside of the silicon tracker in light gray and yellow, respectively. The muon detectors are located on the outside of the detector and are in blue.

CMS was designed to be a general purpose detector that would have sensitivity to a wide range of physics. The central feature of CMS is a superconducting solenoid magnet with a 6 m diameter and 13 m length that provides a 3.8 T magnetic field. The return field from the solenoid is powerful enough to saturate 1.5 m of steel, which is used to guide the return magnetic field. CMS has a cylindrical shape with an onion like design where inner subdetectors are nested inside of outer ones. From inside out, these subdetectors are an all silicon tracker, an electromagnetic calorimeter, a hadronic calorimeter, the magnet, and finally the muon system. Figure 3.1 shows a cross-sectional quarter view of the CMS detector. The signatures of SM particles in CMS are shown in Fig. 3.2.

CMS employs a right handed coordinate system with the x-axis pointing to the center of the LHC ring, the y-axis pointing vertically upward, and thus making the z-axis be along the beam line pointing in the counter-clockwise direction if looking at the LHC from above. The azimuthal angle θ is defined relative to the z-axis. The variable pseudorapidity η is defined as $\eta = -\ln[\tan(\theta/2)]$. The polar angle ϕ is defined relative to the x-axis, meaning that vertically upward (downward) has a ϕ value of $\pi/2$ ($-\pi/2$).

3.3.1 Subdetectors

The innermost part of CMS is an all silicon tracker. Closest to the interaction point are pixel detectors with three barrel layers and two endcap disks, totaling 1,440 modules. Outside of this are strip detectors with ten barrel layers and twelve endcap disks. The tracker extends up to a pseudorapidity range of

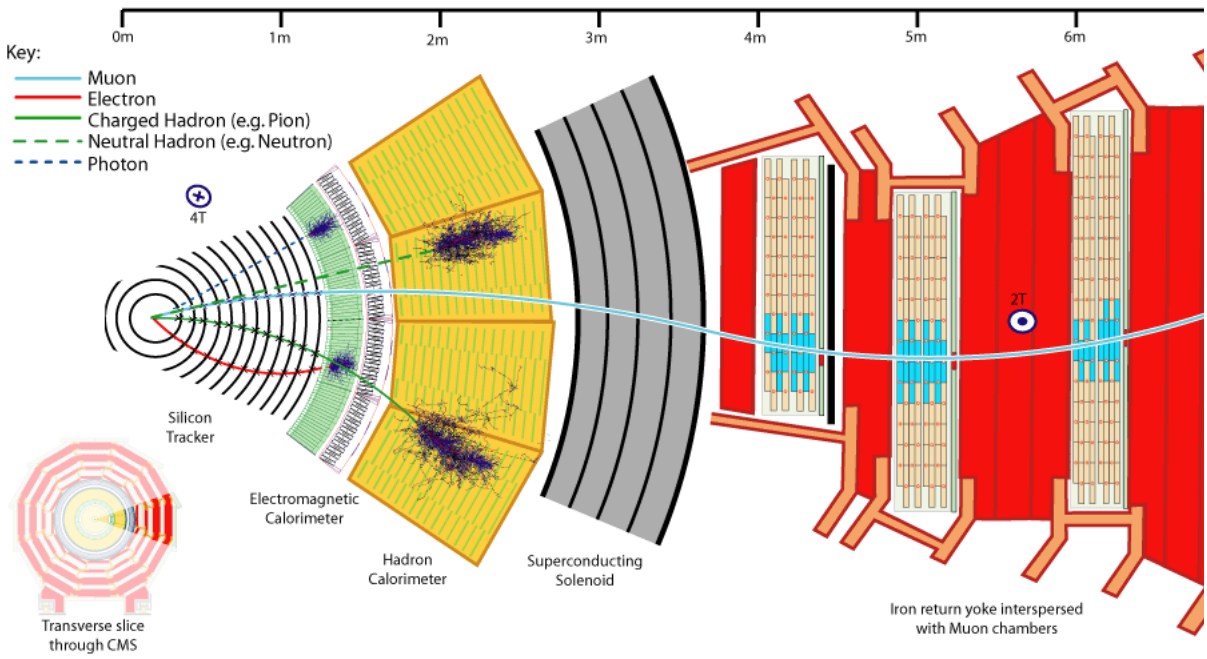


Figure 3.2: A drawing of a r-ϕ cross section of CMS at η=0 along with the expected interactions of SM particles as they propagate through CMS. Shown are a muon as solid line in light blue, an electron as solid line in red, a charged hadron as a solid line in green, a neutral hadron as a dashed green line, and a photon as a dashed line in dark blue.

2.5 with the resolution on track p_T being approximately 1.5% for a 100 GeV/c particle at $|\eta| = 1.6$ and growing larger at high $|\eta|$ due to the bending lever arm. Both the strips and the pixels have an analog readout of the deposited charge with a maximum readout of roughly three times the charge expected to be deposited by a muon. Charge from particles traversing the inner tracker is expected to be spread out among multiple sensors in the same layer (of pixels or strips), allowing the position of the particle to be calculated more precisely than simply the center of the sensor. The charge sharing also allows the possibility to identify hits where two particles have overlapped.

Outside of the inner tracker is the calorimeter. The purpose of the calorimeter is to measure the energy of particles and aid in their identification by stopping particles at different points in the calorimeter. The calorimeter is split into an inner electromagnetic calorimeter (ECAL) and an outer hadronic calorimeter (HCAL). The ECAL is made of 75,848 lead tungstate ($PbWO_4$) crystals split between the barrel and endcap. As particles lose energy in the ECAL the crystals emit scintillation light which is collected by photodetectors. The HCAL consists of plates of brass absorbers interleaved with scintillator detectors. Electrons and photons are likely to stop in the ECAL where they deposit all of their energies. Hadrons, electrically charged or neutral, will deposit some energy in the ECAL but will deposit most in the HCAL where they are very likely to come to a rest. High-momentum muons will deposit of the order of two GeV of energy in the calorimeter and are generally the only charged SM particles that are able to exit the calorimeter.

The outermost part of the detector is the muon system [74] which is split into three parts: Cathode Strip Chambers (CSC), Drift Tubes (DT), and Resistive Plate Chambers (RPC). The CSC cover the

forward part of the detector with $|\eta| > 0.9$ while the DT and the RPC cover the barrel portion extending up to $|\eta|$ of 1.2 and 1.6, respectively. The muon system consists of four stations of chambers with the steel for the magnet return yoke located between the stations. The magnet return yoke provides a magnetic field in the muon system.

The CSC chambers have a trapezoidal shape with six layers of cathode strips and anode wires arranged in a nearly orthogonal pattern. The strips run radially away from the beam line and measure the ϕ of hits while the wires measure the radial position of hits. The cathode strips are segmented so that charge that collects on them will be spread over multiple strips. The amount of charge on the strips is read out every 50 ns. A precise measurement of the ϕ location of the hit can be made by charge interpolation of adjacent strips. Charge collected on the wires is passed to a constant fraction discriminator which outputs a 40 ns pulse. The pulse is sampled every 25 ns and this sampling is readout. The CSCs are laid out with four stations with increasing z from the interaction point and rings of increasing radial distance from the beam line.

Each DT chamber has two or three superlayers which themselves are composed of four layers of drift cells which are staggered by half a cell. Each cell reads out a time measurement. The ensemble of time measurements can be converted into a position measurement and time of particle passage. All of the DT chambers have two superlayers oriented parallel to the beam line, these superlayers measure the position of particles in the $r - \phi$ plane. The three inner stations additionally have a superlayer running perpendicular to the beam line to measure the position of particles in the $r - z$ plane.

The RPC chambers are gaseous parallel plate detectors that can provide a time resolution of 2 ns, which is much smaller than the design LHC bunch spacing of 25 ns allowing for a very high efficiency to correctly tag hits with the correct bunch crossing. The spatial resolution is sufficient to be able to associate RPC hits with hits from the other muon subdetectors.

3.3.2 Trigger and Computing

The rate of bunch crossings, or events, inside of CMS is too large for all of them to be readout and stored offline. To deal with this, CMS employs a two level trigger that selects interesting events online. The level one (L1) trigger must reduce the rate of events readout to less than 100 kHz in less than 4 μs requiring a completely electronics based approach. Events are selected by a variety of algorithms but most of them look for a high momentum track in the muon system, large amount of energy in the ECAL or HCAL, or a combination of these. Multi-object triggers are specifically important for the isolation of new physics signals, that often result in complex final states that can be separated from some SM processes already at trigger level. Signals from these systems trigger the readout of data from the entire detector through the data acquisition system.

As the LHC was designed to operate with 25 ns bunch spacing, many of the subsystems, the tracker especially, only readout the data in the 25 ns window associated with the event. This means that triggers that pre- or post-fire will not contain much of the data from the event.

Once the data are readout by the data acquisition system after an L1 trigger, it is passed to a comput-

ing farm located above CMS. The next step in the trigger, the High Level Trigger (HLT), then runs on the computing farm. The HLT must reduce the number of events to a few hundred Hertz. The HLT is split into two different phases, Level 2 (L2), and Level 3 (L3). The L2 step is mostly concerned with confirming the L1 decision using more robust algorithms, the detector finest granularity, and reducing the rate so that more complex and time consuming reconstruction can be performed in the L3 step within the time restrictions. The L3 step reconstructs tracks in the inner silicon tracker and match them to objects in other parts of the detector, such as tracks found in the muon system or energy deposits in the electromagnetic calorimeter. This permits to identify electrons and muons at HLT with efficiencies above 90% for transverse momenta of about 20–30 GeV, which are the typical lower thresholds used in the analyses which are part of this manuscript. Those analyses rely heavily on top quark pair decay topologies involving light leptons: because of this reason, the performance of triggers based on jet and missing transverse energy, as well as that of "cross-object" triggers, is not crucial. Cross-object triggers are triggers that use more than one physics object for triggering, and are often used in searches for rarer topologies that are involved in the decay of supersymmetric particles. The search for a charged Higgs boson in the fully hadronic final state, described in App. B, actually makes use of a cross tau/missing transverse energy trigger, because of the pecularities of the all-hadronic final state, but only the statistical treatment of the results pertains to this manuscript.

The CMS HLT, being composed of all the single-object triggers and the cross-object triggers, is characterized by a wide variety of different signatures that are searched for; if any are found, the data are passed to computers located at CERN and throughout the world for storage and further analysis.

CMS maintains a software package, CMSSW, which is responsible for taking the raw data readout from CMS, and reconstructing what was happening in the event. This includes applying calibration constants, finding tracks, and identifying particles. The package includes the implementation of the HIT as well. After this reconstruction, the data size is at the scale of petabytes, which is too large for offline analyzers to run over frequently. To deal with this skims of the data are produced, dropping lower level quantities and selecting only events that a particular analysis is interested in studying.

CMSSW is also tasked with simulating how particles, coming from both SM processes and new physics, would interact with the detector so that this can be used to compare against data. A few steps are performed before the simulation has the same format as data readout from the detector, at that point it follows the same chain as data. The first step is the simulation of the proton-proton collision and the particles that are created from it; the detector is not used at all in this step. The simulation is done by separate event generators, such as PYTHIA [75] or ISAJET [76], which then provide a list of final-state particles to be used in the next step. The simulation of how these particles will interact as they pass through CMS is handled by the GEANT program [77]. Finally the behavior of the detector electronics, including the L1 trigger, is handled by CMSSW. After this point, the simulation is handled the same as data.

3.4 The reconstruction of physics events

In the data collected during 2011, an average of 9 proton-proton interactions occurred per LHC bunch crossing. During 2012, due to the increased instantaneous luminosity [78], the average number of proton-proton interactions per LHC bunch crossing increased to around 21. To select the primary interaction vertex, the squared sum of the transverse momenta of the charged particle tracks, $\sum p_{\mathrm{T}}^2$, associated to each interaction vertex is calculated using a Kalman-Filter technique [79]. In the analyses presented in this manuscript, the interaction vertex with the largest $\sum p_{\mathrm{T}}^2$ value is taken as the primary interaction vertex in the event. The other pp collisions occurring in the same event are referred to as pileup.

Events are reconstructed with the particle-flow (PF) algorithm [80, 81, 82], which combines information from all sub-detectors to identify and reconstruct individual electrons, muons, photons, charged, and neutral hadrons. Electrons are reconstructed from clusters of ECAL energy deposits matched to hits in the silicon tracker [83]. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The reconstruction efficiency and misidentification rate for electrons have been estimated in internal documentation of the CMS collaboration, and vary between about 0.91 and 1.0 at 7 and 8 TeV.

Muons are reconstructed by performing a simultaneous global track fit to hits in the silicon tracker and the muon system [84]. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for zero-suppression effects and for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energy. The reconstruction efficiency and misidentification rate for muon have been estimated in internal documentation of the CMS collaboration, and vary between about 0.97 and 0.99 at 7 and 8 TeV.

The composite physics objects, such as jets, taus, and missing transverse energy are reconstructed from these PF particles.

Jets are reconstructed from the PF candidate particles clustered by the anti-k_{t} algorithm [85, 86] with a size parameter of 0.5. Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found in the simulation to be within 5 % to 10 % of the momentum of the corresponding jet at generator level over the whole p_{T} spectrum and detector acceptance. An offset correction is applied to take into account the extra energy clustered in jets arising from additional proton-proton interactions within the same bunch crossing. Jet energy corrections are derived from simulation, and are confirmed with in situ measurements with the energy balance of dijet and photon+jet events [4]. The correction factors are derived from the ratio of distributions as average correction factors binned in p_{T} and η.

Jet energy corrections are applied sequentially in a fixed order: first the *L1-PileUp "offset"* corrections are applied: the goal is to remove from the jet the energy coming from pile-up events. This should

remove any dependence on the luminosity (because the different pileup conditions would be already accounted for by the correction). These offset corrections are determined from simulation using QCD dijet events processed with and without adding pileup (see Sec. sec:sim for an overview of the way pileup is simulated in MonteCarlo samples). Then *L2-L3 MC-truth* corrections are applied. Historically, these were two different corrections, *L2-Relative* accounting for η dependence and *L3-Absolute* accounting for p_T dependence. These corrections are based on the simulation of QCD dijet samples and have the purpose of making the response uniform in jet η and jet p_T. Finally, only for data, the *L2L3-Residuals* corrections are applied. Those are meant to correct for remaining small differences (of the order of $\sim 1\%$) between the jet response in data and MC. The *L2-Residuals* eta-dependent corrections are determined with dijet events, relative to a jet of similar p_T in a barrel reference region. These corrections include a p_T dependence of the jet energy scale relative to the jet energy scale of the barrel jet. The *L3-Residuals* corrections act on the jet absolute scale, i.e. on its p_T dependence. These corrections are determined, for barrel jets, using $Z(\mu\mu, ee)$+jet, photon+jet, and multijet events, and are described in detail in Ref. [87]

In addition to the *L1-Pileup* corrections, pileup is dealt with at clustering level by using the so-called *charged hadron subtraction* technique: this subtraction technique consistes in ignoring, when clustering the jets, charged hadrons originating from pileup vertices, and is described in detail in Ref. [88].

Fig. 3.3, taken from Refs. [4, 5], shows the performance of the jet energy corrections in terms of the uncertainty on the total jet energy scale for particle flow jets, for 7 TeV (a) and 8 TeV (b). Additional selection criteria are applied to each event to remove spurious jet-like features originating from isolated noise patterns in certain HCAL regions.

The pileup corrections described above had as a target the correction to the energy of jets that come from the presente of pileup deposits clustered into the jet. A second effect of pileup is the presence of jets that are entirely made of pileup deposits, and is accounted for by an identification algorithm. In all analyses, jets originating from pileup interactions are removed by a multivariate jet identification algorithm [89]. This algorithm expoits two classes of variables: vertexing related variables enable to distinguish between charged particles coming from the primary vertex from the ones coming from pileup, whereas shower-shape related variable enable to estimate the origin of deposits coming from neutral particles. The latter, being neutral, cannot indeed be studied using vertexing related variables, because of the absence of tracking information. All the variables are fed to a BDT algorithm (BDT algorithms are described in Sec. 4.2) trained in different pseudorapidity regions in order to maximize the classification performance with respect to the calorimeters and tracker performance, that vary with the pseudorapidity.

Jets from the hadronization of b quarks (b-tagged jets) are identified with the combined secondary vertex algorithm [6, 90]. The algorithm is based on a likelihood-based discriminator which uses information on reconstructed decay vertices of short-lived mesons and transverse impact parameter measurements of charged particles. The act of setting a threshold on the discriminator fixes the performance of the algorithm for that working point, in terms of probability of tagging a quark coming from a b quark (b-tagging probability) and of probability of tagging a quark coming from a light quark, i.e. u, d, s, or$\mathrm{c}$ (mis-tagging probability). In practice, the choice of a working point for the b-tagging discriminators starts from deciding an acceptable mistag rate: this defines the cut on the discriminator, which in turn enables

to retrieve the b-tagging efficiency for that working point. The performance of the b-tagging algorithm used in the analyses included in this work is shown for the 7 TeV run in Fig. 3.4, with respect to the misidentification probabilities for light jets and for $\mathrm{c}$jets. The corresponding performance for 8 TeV is similar, although it has not yet been released as a public plot. The difference in b-tagging efficiency between data and simulation is corrected for by applying data-to-simulation scale factors dependent on the jet momentum and pseudo-rapidity.

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\,\text{miss}}$ is defined as the projection on the plane perpendicular to the beams of the negative vector sum of the momenta of all reconstructed PF particles associated with the primary vertex in an event. Its magnitude is referred to as $E_{\mathrm{T}}^{\text{miss}}$. The $E_{\mathrm{T}}^{\text{miss}}$ reconstruction is improved by propagating the jet energy corrections to the reconstructed $E_{\mathrm{T}}^{\text{miss}}$: since the jet energy corrections include corrections due to pileup, their propagation to $E_{\mathrm{T}}^{\text{miss}}$ results in accounting for pileup effects in the reconstruction of the $E_{\mathrm{T}}^{\text{miss}}$ object. Further filter algorithms are applied to reject events with anomalously large $E_{\mathrm{T}}^{\text{miss}}$ resulting from instrumental effects.

The "hadron-plus-strips" algorithm [7, 91] is used to reconstruct hadronically decaying τ leptons. The algorithm uses the constituents of the reconstructed jets to identify individual τ decay modes with one charged and up to two neutral pions, or three charged pions. The neutral pions are reconstructed by clustering the reconstructed photons in narrow strips along the azimuthal angle direction taking into account possible broadening of calorimeter depositions from photon conversions. The τ_{h} candidates compatible with electrons or muons are rejected. Jets originating from the hadronization of quarks and gluons are suppressed by requiring that the τ_{h} candidate is isolated as described below. The τ_{h} identification efficiency depends on $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ and $\eta^{\tau_{\mathrm{h}}}$, and is on average 50 % for $p_{\mathrm{T}}^{\tau_{\mathrm{h}}} > 20\,\text{GeV}$ with a probability of approximately 1 % for hadronic jets to be misidentified as a τ_{h}.

Fig. 3.5, taken from Ref. [7], shows the performance of the τ_{h} identification algorithms for the 8 TeV run. The values for the 7 TeV run are similar.

Electrons and muons assumed to originate from W or Z boson decays or from τ leptonic decays, as well as hadronically decaying τ leptons, are required to be isolated from other particles by considering photon, neutral hadron, and charged hadron transverse momenta in a cone $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, where ϕ is the azimuthal angle and η the pseudorapidity, around the charged lepton candidate momentum direction. The isolation for electrons, muons, and τ_{h} is defined as:

$$I^{\mathrm{e}} = \sum_{\text{charged}} p_{\mathrm{T}} + \max\left(0, \sum_{\text{neut. hadr.}} p_{\mathrm{T}} + \sum_{\gamma} p_{\mathrm{T}} - \rho_{\text{neutral}} \times A_{\text{eff.}}\right), \tag{3.1}$$

$$I^{\mu} = \sum_{\text{charged}} p_{\mathrm{T}} + \max\left(0, \sum_{\text{neut. hadr.}} p_{\mathrm{T}} + \sum_{\gamma} p_{\mathrm{T}} - 0.5 \sum_{\text{charged,pileup}} p_{\mathrm{T}}\right), \tag{3.2}$$

$$I^{\tau_{\mathrm{h}}} = \sum_{\text{charged}} p_{\mathrm{T}} + \max\left(0, \sum_{\gamma} p_{\mathrm{T}} - 0.46 \sum_{\text{charged,pileup}} p_{\mathrm{T}}\right), \tag{3.3}$$

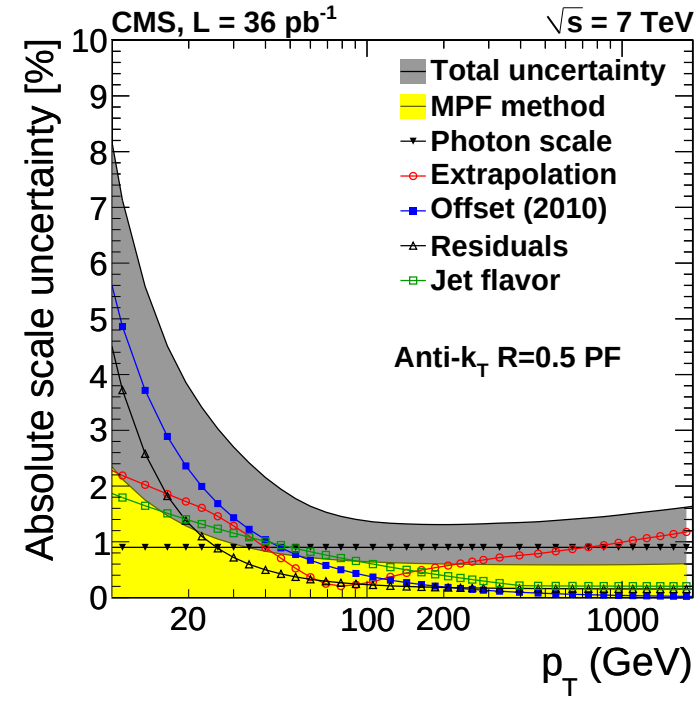
where $\sum_{\text{charged}} p_{\mathrm{T}}$ is the scalar sum of the transverse momenta of charged hadrons, electrons, and muons from the primary interaction vertex, $\sum_{\text{neut. hadr.}}$ is the scalar sum of neutral hadrons, and $\sum_{\gamma} p_{\mathrm{T}}$

is the scalar sum of photons in the cone ΔR around the charged lepton candidate momentum direction.

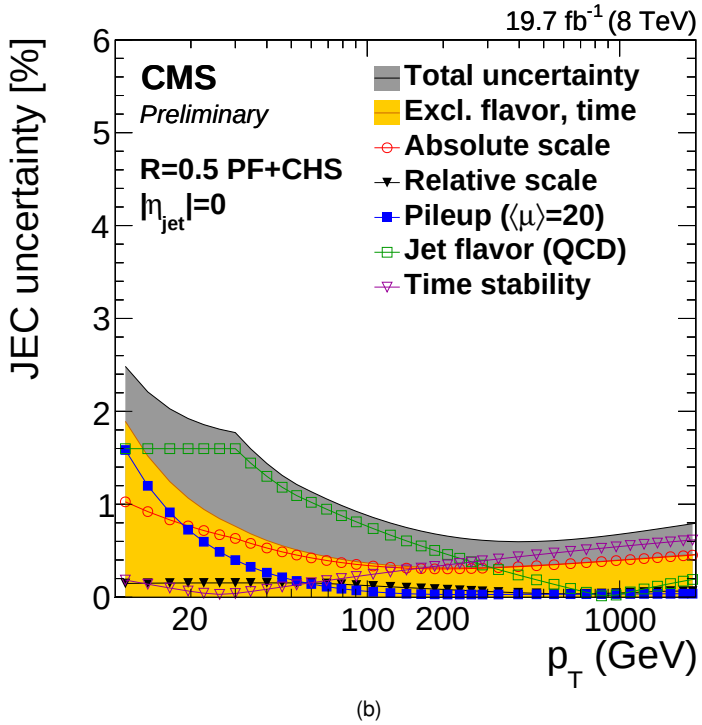
The requirement of being matched with a primary vertex is applied to charged hadrons only: this is because neutral hadrons are reconstructed only in the calorimeters, and the CMS calorimeter system is not a layered system, making a trajectory fit (and hence an match between a trajectory and the primary vertex) impossible. For comparison, the ATLAS detector features layered calorimeters, enabling the event reconstruction algorithms of that experiment to try to assign neutral hadrons to the primary vertex. The presence of particles from pileup events is taken into account depending on the charged lepton type. For electron candidates, the scalar sum of the p_{T} of photons and neutral hadrons from pileup events in the isolation cone is estimated as the product of the neutral-particle transverse momenta density and the effective cone area, $\rho_{\mathrm{neutral}} \times A_{\mathrm{eff.}}$. The ρ_{neutral} component is evaluated from all photons and neutral hadrons in the event, and represents the event-specific average pile-up energy density per unit area in the (ϕ, η) place. The effective area $A_{\mathrm{eff.}}$ is an effective area specific to any given type of isolation, and is modeled by constant values, usually expressed as a function of the pseudorapidity η, that are derived CMS-wise for each dataset of real or simulated events.

For muons and hadronically decaying τ leptons, the scalar sum of the p_{T} of photons and neutral hadrons from pileup events is estimated from the scalar sum of the transverse momenta of charged hadrons from pileup events in the isolation cone, $\sum_{\mathrm{charged,pileup}} p_{\mathrm{T}}$, by multiplying it with the average ratio of neutral to charged hadron production in inelastic pp collisions. This factor is slightly smaller for the τ_{h} isolation than for the muon isolation since the contribution from neutral hadrons is ignored in the isolation variable.

For electrons, the isolation cone size of $\Delta R = 0.3$ or 0.4 is used depending on the analysis. For muons and hadronically decaying τ leptons, the isolation cone sizes of $\Delta R = 0.4$ and 0.5 are used, respectively. Electrons and muons are selected if the relative isolation $I^{\ell}_{\mathrm{rel}} = I^{\ell}/p^{\ell}_{\mathrm{T}}$, where $\ell = \mathrm{e}, \mu$, is lower than 10–20 % depending on the analysis. Hadronically decaying τ leptons are considered isolated if $I^{\tau_{\mathrm{h}}} < 1\,\mathrm{GeV}$.

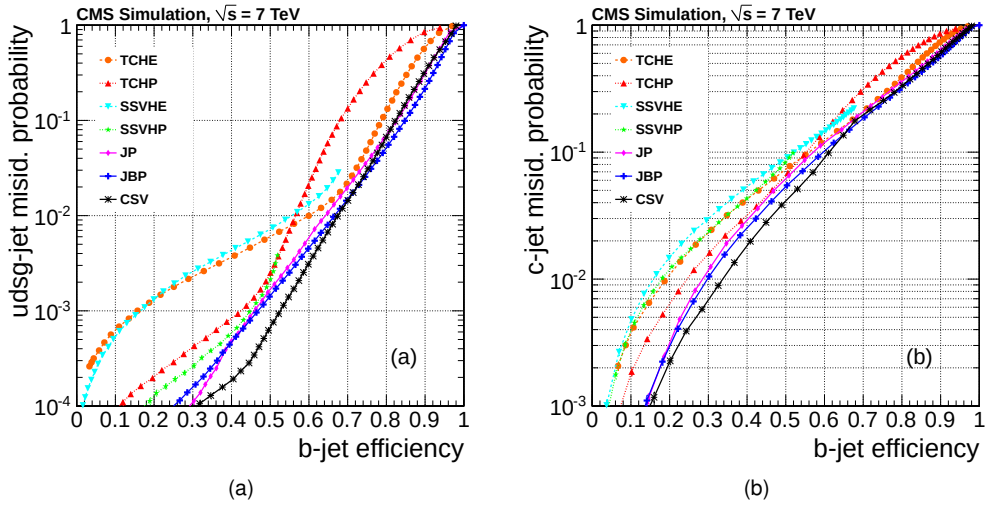


(a)



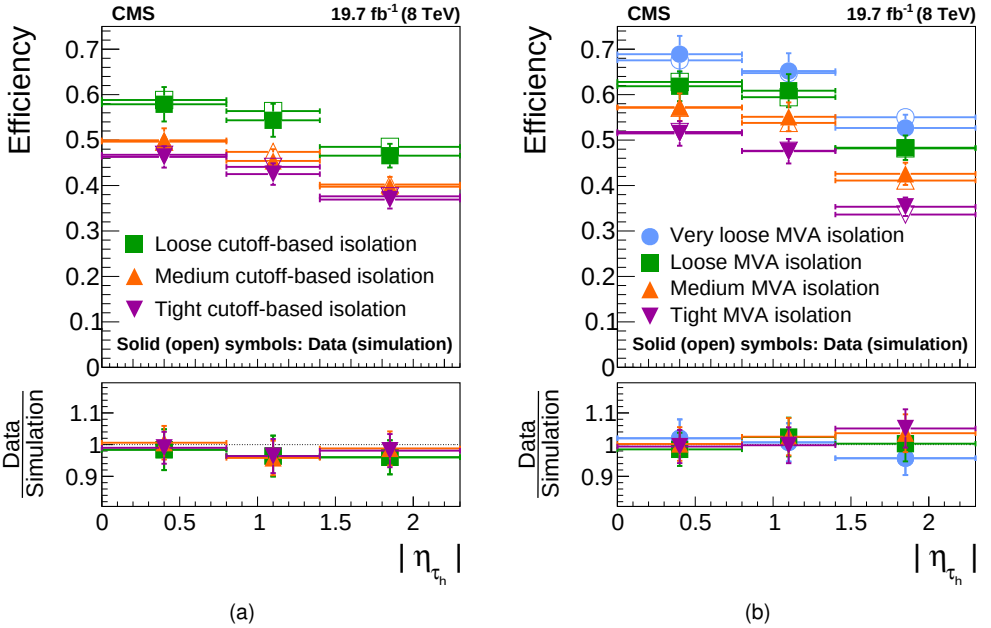
(b)

Figure 3.3: Absolute jet energy scale uncertainty as a function of the jet transverse momentum for Particle Flow jets, for 7 TeV (a) and 8 TeV (b). The figures are taken from Refs. [4, 5].



(a) (b)

Figure 3.4: Performance curves obtained from simulation for the algorithms currently tested at CMS. (a) lightparton- and (b) c-jet misidentification probabilities as a function of the b-jet efficiency. Jets with transverse momentum larger than 60 GeV in a sample of simulated multijet events are used to obtain the efficiency and misidentification probability values. The figures are taken from Ref. [6].



(a) (b)

Figure 3.5: Tau identification efficiency measured in $Z/\gamma^* \to \tau\tau \to \mu\tau_{\mathrm{h}}$ events as function of the tau pseudorapidity η, for different cutoff-based and MVA-based τ_{h} isolation discriminants, compared to the MonteCarlo expectation. The figures are taken from Ref. [7].

3.5 Simulation of physics processes

The proton-proton interactions and the subsequent decays can be represented by a complex multi-dimensional probability density function (p.d.f.) that models the probability of each interaction to occur in each of the points of the phase space defined by the kinematics of the objects involved in the event. A method commonly used to easen the task of dealing with multidimensional p.d.f.s is the Monte Carlo (MC) method. Events are generated by sampling each quantity involved in the determination of the event from the p.d.f. given by the theoretical model for the given interaction (i.e. picking a random value for that quantity based on the relevant p.d.f.). This process results in a set of events which populates the phase space discreetly, but according to the joint probability of each phase space point. A number of software packages have been developed, which detail the structure of the expected final products of different type of collisions.

- **MADGRAPH [92, 93, 94]:** this is a FORTRAN software framework for computing Leading-Order Standard Model cross sections, generating hard process events, and interfacing the result with further generators (e.g. hadronizers like PYTHIA). Models different than the Standard Model can be easily probed via parameter cards containing the list of particles predicted by the model and their decays. Recently the framework has been updated and integrated with the aMC@NLO Next-to-Leading-Order generator, yielding the version 5 of the framework, giving the possibility of computing NLO cross section and generating NLO hard process events. The CMS Experiment used the LO version (version 4) for the simulation of proton-proton collisions at 7 TeV, and the NLO version for collisions at 8 TeV. However, due to the high computing time needed to generate a NLO event, most of the saples at 8 TeV have been produced at leading order, and the resulting distributions have been normalized to the next-to-leading-order cross sections computed separately. The matrix elements (ME) computed within MADGRAPH provide up to four additional partons, including $\mathrm{b}$quarks. Parton showering and matching of the soft radiation with the contribution from the ME are performed separately in a following step using PYTHIA, using the k_{T}-MLM approach [95];

- **POWHEG [96, 97, 98, 99, 100]:** this is a computer framework for implementing NLO calculations in shower Monte Carlo programs according to the POWHEG method. The CMS Experiment has been employed this generator to generate single top events, for which the leading order approximation is not modeling well the top quark spectra;

- **PYTHIA [75]:** this is a FORTRAN software framework for the generation of high-energy collisions. It supports a set of physics models for the evolution of an event from a few-body hard process to a multi-hadronic final state. Initial- and final-state parton showers, multiple parton-parton interactions, beam remnants, string fragmentation and particle decays are supported through several physics libraries. The full suite has recently being rewritten in C++, resulting in the version 8 of the suite. The CMS Experiment, though, still used the FORTRAN version (PYTHIA 6) for the simulation of proton-proton collisions at 7 and 8 TeV;

- **TAUOLA [101]:** this is a package for simulating τ lepton decays. These decays can be well separated from their production process because of their narrow width: spin degrees of freedom connect these two parts of the physics process of interest for high energy collision experiments. The package can be interfaced with events generated by other programs, and replaces τ lepton decays computed by those programs with full decays computed from scratch: the package fully takes into account the effects of spin, genuine weak corrections or of new physics, depending of the physics model supplied to the production chain. In particular, the τ lepton polarization is accurately simulated.

At 7 TeV, the charged Higgs signal processes are generated with PYTHIA 6.4 Background events for $\mathrm{t\bar{t}}$, W+jets, Z+jets are generated with MADGRAPH 5 interfaced with PYTHIA 6.4. The diboson production processes WW, WZ, and ZZ are generated by PYTHIA. Single-top-quark production is generated with POWHEG [97]. The signal processes, $\mathrm{t\bar{t}} \to \mathrm{H^{\pm}bH^{\mp}\bar{b}}$ and $\mathrm{t\bar{t}} \to \mathrm{W^{\pm}bH^{\mp}\bar{b}}$, are generated by PYTHIA. The TAUOLA package is used to simulate τ decays in all cases. Generated events are processed through the full detector simulation based on GEANT4 [102, 103], followed by a detailed trigger emulation (including digitization and out-of-time pileup effects) and the CMS event reconstruction. Several minimum-bias events are superimposed upon the hard interactions to simulate pileup. Simulated events are reweighted to reproduce the number of pileup interactions expected in data. The expected pileup is computed based on the instantaneous luminosity and the pp inelastic cross section. The PYTHIA parameters for the underlying event are set according to the "Z2" tune [104], which uses the CTEQ6L [105] parton distribution functions (PDFs). In this context, a "tune" is a set of parameters that determine the way the event is modeled by the generator. These parameters area adjusted to better fit the underlying event description measured in data. An estimate of the performance of different tunes can be found in Ref. [106].

At 8 TeV, the charged Higgs signal processes are generated with PYTHIA 6.4. The $\mathrm{t\bar{t}}$, W+jets, and Z+jets backgrounds are generated using the MADGRAPH 5.1 event generator, and the generated events are interfaced with PYTHIA to provide the showering of the partons and to perform the matching of the soft radiation with the contributions from the ME. The single-top-quark production is generated with POWHEG 1.0 and the QCD multijet and diboson production processes WW, WZ, and ZZ are generated using PYTHIA. Both the MADGRAPH and POWHEG generators are interfaced with PYTHIA for parton shower and hadronization. The TAUOLA package is used to simulate τ decays for the simulated signal as well as background samples.

The generated events are passed through the full process described above for the simulation of the CMS detector and of the trigger and event reconstruction. Pileup effects are similarly included, and weights are applied to simulated events to match the pileup conditions present in the 8 TeV data. The PYTHIA parameters for the underlying event are set according to the "Z2*" tune [107], which uses the CTEQ6L [105] PDFs and is an evolution of the "Z2" tune mentioned above for the 7 TeV simulation.

Chapter 4

Statistical Methods for Data Analysis

The act of analyzing data coming from high energy collisions is a multi-step process: at any of these steps, data are treated using more or less complex statistical methods that are mostly oriented to allowing estimating most probable values for physical quantities one is interested in and, perhaps more importantly, in estimating intervals representing a reasonable amount of uncertainty around those estimated values.

In this chapter are described the main methods that have been employed specifically for the analyses reported in this thesis:

- the profile likelihood ratio method that have been used in particular for the measurement of the top quark mass 5.1 is described in Sec. 4.1;
- a multivariate analysis method called “k-Nearest-Neighbour” have been used for estimating the major background in the top pair cross section measurement 5.2.6 and for the light and (to a minor extent) the heavy charged Higgs searches **??**. Such method is described in Sec. 4.2;
- the CL_s method for setting upper limits is described in Sec. 4.3: it has been used for testing the Standard Model hypothesis against the New Physics hypothesis in the light and heavy charged Higgs searches **??**.

4.1 Finding most probable values and throwing intervals: the profile likelihood ratio

The profile likelihood ratio method is a technique for estimating the most probable value, and a reasonable confidence interval around it, for a physical parameter θ in presence of multiple sources of uncertainty. The unidimensional case is often referred to as *likelihood ratio method*, *MINOS method* (so-called after the MINOS computational routine developed by James and Roos [108]), or *likelihood based confidence intervals*.

The parameter θ is in general an n-dimensional vector that can be arbitrarily divided into a vector of *parameters of interest* θ' and a vector of *nuisance parameters*. The parameters of interest are the

parameters one wants to quote and construct confidence interval for, that is the interesting physics: the nuisance parameters are the parameters that "get in the way" of those determinations (in the sense that those estimates depend also on the values of nuisance parameters) and that one needs to evaluate and take into account when estimating the parameters of interest and their confidence interval. This distinction is somehow arbitrary, in the sense that depends on which are the parameters one wants to construct: for example, usually the energy calibration of hadronic jets is considered to be a nuisance parameter that gets in the way of the determination of some "more important" parameter, such as the mass of the top quark or the top pair production cross section. However, in some cases it is convenient to consider the energy calibration as a parameter of interest together with the "main" one, and find best fit value and confidence inverval for both symultaneously (see for example the symultaneous top mass and jet energy scale measurement described in Ref. [44]).

Given a set of N random variables **X**, each with probability distribution function $f(X_i|\theta)$, their joint probability distribution function (*joint* simply means that the p.d.f. is a function of more than one variable) is:

$$P(\mathbf{X}|\theta) = P(X_1, ..., X_N|\theta) = \prod_{i=1}^{N} f(X_i|\theta) \tag{4.1}$$

The set of variables **X** can be replaced by the observed data $\mathbf{X^0}$: in this situation, P is no longer a p.d.f. and is no longer a function of random variables: to avoid confusion, it is usually denoted by $L(\mathbf{X^0}|\theta)$ and called *likelihood function*. Some authors, in order remark the dependence of the likelihood on the parameter θ only, write it as $L(\theta|\mathbf{X^0})$, notation that in frequentist statistics is still acceptable: in bayesian statistics, however, this would result in an improper probability inversion. This is immediate when expressing the Bayes theorem in the form

$$P(\theta|\mathbf{X^0}) = \frac{P(\mathbf{X^0}|\theta)P(\theta)}{P(\mathbf{X^0})} \tag{4.2}$$

The likelihood, there denoted by $P(\mathbf{X^0}|\theta)$, effectively represents the probability of obtaining the observed data, as a function of the parameter θ. The issue with probability inversion, and the different interpretations of the likelihood function, are discussed in detail in the wonderful Ref. [109].

Finally, since the observed data are fixed, many authors (for example Ref. [110]) omit specifying them, and write the likelihood function simply as $L(\theta)$.

In order to estimate a parameter, it is necessary to choose a function of the observations: this function is called *estimator*. The numerical value yielded by the estimator for a particular choice of the set of observations is called *estimate*. An important property that is requested for an estimator is *consistency*: an estimator is a *consistent estimator* when the estimates converge to the true value of the parameter as the number of observations increases. This is a desirable property if the purpose of using an estimator is to extract information about the true value of the parameter θ. Conventionally, the estimator of the parameter θ is denoted by $\hat{\theta}$.

A consistent estimator can be constructed using the weak version of the law of large numbers: given a function $a(X)$ of the observations for a single random variable, with finite variance $V\big[a(X)\big]$, this states

that

$$N^{-1}\sum_{i=1}^{N} a(X_i) \xrightarrow{N\to\infty} E\big[a(X)\big] = \int a(X) f(X,\theta) dX_i \tag{4.3}$$

that is that if the variance is finite, then the sample mean is a consistent estimator of $a(\mathbf{X})$.

Three vastly used methods of estimation that make use of this construction are the *moments method*, the *least squares method*, and the *maximum likelihood method*. The moments method starts from expressing the expectation value of $a(X)$ as a function $h(\theta)$ of θ

$$h(\theta) := E\big[a(X)\big] = \int a(X) f(X,\theta) dX_i \tag{4.4}$$

and assuming that for the true value $\theta_\mathbf{0}$ an inverse function exists such as $h^{-1}\big[h(\theta_\mathbf{0})\big] = \theta_\mathbf{0}$. If this is the case, the estimator $\hat{\theta}$ can be written as a function of the observations X, rather than as a function of the unknown p.d.f. $f(X,\theta)$. In the particular case in which $a_j(X_i) := X^j$, the functions $h_j(\theta)$ become the moments of the p.d.f., giving thus the name to the full method.

The other two methods are both part of a class called *implicitly defined estimators*: the function $a(X)$ is chosen as a general function of both X and θ, $a(X,\theta)$. If the corresponding function $h(\theta) := E\big[a(X)\big]$ has a zero in correspondence of the true value of the parameter, then by the law of the large numbers the function $\xi(\theta) := N^{-1}\sum_{i=1}^{N} a(X_i,\theta)$ will have asymptotically the same roots as $h(\theta)$. As a consequence, one of the maxima or minima of the function $a(X,\theta)$ is a consistent estimator $\hat{\theta}$ of the true value $\theta_\mathbf{0}$. More often, instead of minimizing/maximizing $a(X,\theta)$, the search for the roots is performed for a function $g(X,\theta)$ such that

$$a(X,\theta) = \frac{\partial}{\partial\theta} g(X,\theta) \tag{4.5}$$

The maximum likelihood method makes indeed use of this auxiliary function. The joint likelihood for a single parameter θ can be written from the p.d.f. $f(\mathbf{X},\theta)$ of any observation X and from Eq. 4.1 as

$$L(\mathbf{X},\theta) = \prod_{i=1}^{N} f(X_i|\theta) \tag{4.6}$$

The auxiliary function for the maximum likelihood method is then

$$g(X,\theta) = lnf(X,\theta) \tag{4.7}$$

and the maximization is operated by finding the zeros of the logarithm if the likelihood:

$$\frac{\partial}{\partial\theta}\sum_{i=i}^{N} lnf(X_i,\theta) = \frac{\partial}{\partial\theta} lnL(\mathbf{X},\theta) = 0 \tag{4.8}$$

A couple caveates are needed. First, a necessary condition for the likelihood having a maximum is that the maximum does not correspond to a limit of the range of θ. Second, in practical cases the maximization is performed numerically: the numerical procedures have been found to be more stable

in minimizing rather than in maximizing, and consequently the practical usual choice is to minimize the additive inverse of the logarithm of the likelihood function, $-lnL(\mathbf{X}, \theta)$.

It can be demonstrated (see Ref [111] for the full demonstration) that one of the roots found by the maximization procedure is a consistent estimate for the parameter θ. However, the proof is true only asymptotically, for $N \to \infty$: this means that the maximum found for any finite N is not necesserily the absolute maximum. It is very possible that, as N increases, the relative heights of the different maxima of lnL change: this is a reflection of the fact that for any *finite* sample there will be always an intrinsic uncertainty on the true value θ_0.

The least squares method works in a similar way, but the minimization is done on a quadratic form, function of the deviation of the observations from their expectation values.

The maximum likelihood estimator $\hat{\theta}$ is consistent, which implies asymptotic unbiasness, and is asymptotically Normally distributed: its variance is the minimum variance possible, which is the one given by the Cramer-Ráo lower bound [110, 111]. It is also important to remark that the asymptotic limit is normally reached slowly: this means that often the likelihood function, evaluated for a finite N, will have more than one maximum. In this case, there is no way of deciding which of the maxima corresponds to the true value θ_0. The only case in which the maximum likelihood estimate has optimal properties even for N finite is the case in which the parent p.d.f. is an exponential of the form $f(X, \theta) = exp\big[\alpha(X)a(\theta) + \beta(X) + c(\theta)\big]$, because this form admits *sufficient statistics* (this means that data reduction can be performed without any loss of information: in the case at hand, it means that no information is gained in the passage from N finite to the limit N infinite).

One of the most important properties of the maximum likelihood estimate is its invariance for change of variables of the parameter θ: for a function $\tau(\theta)$, the maximum likelihood estimate is

$$\hat{\tau} = \tau(\hat{\theta}) \tag{4.9}$$

This is particularly important, because this means that the maximum likelihood estimate can be found for the parameter that has the simplest form of $L(\mathbf{X}, \theta)$, and the estimate can be immediately translated to the one for any transformation of that parameter. Moreover, this property applies not only to the maximum of the likelihood, but also to any relative values of the likelihood function: this is particularly important for interval estimation with the likelihood ratio method.

The *likelihood ratio ordering principle*, described in detail in Ref. [112], has been devised in order to avoid the issues that are present when building confidence intervals near to the physical boundary of the parameter θ. Confidence intervals are built, in frequentist statistics, around the notion of *probability content* of the interval. The goal is to find the interval $\theta_a \leq \theta \leq \theta_b$ that contains the true value with probability β. Since usually the true value is not known a priori, a function $Z(X, \theta)$ of the observations X and the parameter θ must be found, such that its p.d.f. is independent of the unknown θ. An interval that satisfies the desired requirement on its probability content is said to have the *coverage* property and is called *confidence interval*. For a given observation X sampled from a p.d.f. $f(X|\theta)$, the probability

content β of a region $[a, b]$ in the X space is:

$$\beta = P(a \leq X \leq b) = \int_a^b f(X|\theta)dX \tag{4.10}$$

This needs to be translated to the parameter space, by finding the confidence interval $[\theta_a, \theta_b]$ such that

$$P(\theta_a \leq \theta_0 \leq \theta_b) = \beta \tag{4.11}$$

The act of operating this probability inversion is very delicate: depending on the specific form of the p.d.f., the desired interval can be found starting from Eq. 4.10, either directly or via some auxiliary variable $Z(X, \theta)$. In some cases, the interval will intersecate the limit of the allowed range for the parameter θ, meaning that part of the interval will fall outside of the allowed range. This results in that the part of the interval that falls within the allowed range will have a probability content *lower* than the declared one, $\beta|_{allowed} < \beta$. In order to ensure proper coverage, the interval must be built from scratch, starting from each hypothetical value of the parameter θ and accumulating elements of probability until the total integral of the interval between its endpoints is β: this is referred to as Neyman construction [113]. An immediate issue with this construction is that there is no unique way of adding elements of probability to obtain the final interval for a given value of θ. This issue is solved by the forementioned likelihood ratio ordering principle [112], which states that, when determining the interval for some $\theta = \theta_p$, the elements of probability $P(X|\theta_p)$ that must be included are the ones with the largest value of the *likelihood ratio*

$$R(X) = \frac{P(X|\theta_p)}{P(X|\hat{\theta})} \tag{4.12}$$

where $\hat{\theta}$ is the value of θ that maximizes the likelihood $P(X|\theta)$ within the physical region of the parameter θ.

For likelihoods that have a normal distribution as a parent distribution, which is often the case asymptotically, the confidence interval for any given probability content can be found by translating the likelihood such that at the maximum $lnL = 0$: then, the most usual confidence intervals (the ones with probability contents of 68.3%, and 95.5%) can be found by finding the values corresponding to $lnL = -0.5$ (for $\beta = 68.3\%$) and $lnL = -2$ (for $\beta = 95.5$). In case the likelihood is not parabolic, it would be necessary to operate a change of variables in order to have a parabolic log-likelihood for some function $\tau(\theta)$ of the parameter but, thanks to the invariance property discussed above for the likelihood function, it is not necessary to perform the change of variables: the most probable value and the intervals can be read directly from the non-parabolic log-likelihood by looking at the values of the parameter corresponding to $lnL = 0$, $lnL = -0.5$, $lnL = -2$ and so on.

This treatment can be immediately extended to the case of a multidimensional parameter θ and using the likelihood ratio

$$\ell = \frac{L(\mathbf{X}, \theta)}{L(\mathbf{X}, \hat{\theta})}. \tag{4.13}$$

When the set of parameters is divided into parameters of interest and nuisance parameters, the nuisance parameters at the numerator of the likelihood ratio are the ones that maximize the likelihood function for a given set of values of the parameters of interest.

It is important to remark that, both in the single-dimensional and in the multi-dimensional case, this method is only approximative, since it relies on the asymptotic properties of the likelihood function. This means that the profile likelihood method is exact only in the limit $N \to \infty$ or, more specifically, that it will be exact only to the order $\frac{1}{N}$.

4.2 Classifying events using many variables at the same time: the k-Nearest-Neighbours method

Multivariate classifiers are algorithms used to separate a set of events into different classes, starting from the values assumed in each event by a set of variables, called *input variables*. The most simple case is the classification between two classes, A and B, than in high-energy physics applications are often indicated as *signal* and *background*. Each algorithm provides a general function for converting a set of input variables for a given event into a single number representing the classification of that event: the best values for the parameters of the conversion function are usually determined by fitting the function to a set of events whose classification is known a-priori. This process is referred to as *training*, and is part of a general class of problems called *supervised learning*. Algorithms that do not make use of training events classified a-priori are known as *unsupervised learning* algorithms, and are outside of the scope of this brief review.

The k-Nearest-Neighbour algorithm is a memory-based multivariate classifier [8]: this means no analytical model is required to be fit for finding the best parameters, since the model is constituted by the whole training data set and the fitting is performed either at evaluation or prediction time. For real-time applications, in which speed is a major concern, this can make this class of method not feasible. For "offline" applications, however, speed is not usually a concern, and it is consequently advantageous to use an algorithm that does not depend on the choice of an analytical model.

The basic idea is to start from building the N-dimensional parameters space, where N is the number of variables chosen for describing the problem, such that each set of symultaneous values of the N variables is described as a single point x in the parameters space. This point is represented by a N-dimensional vector. The space is then populated using the points from the *training set*, which is the set of points the algorithm must learn from: for the elements of the training set, the correct classification is known a-priori.

The *testing set* is another set of points for which the classification is known a-priori, and is used for evaluating the performance of the classifier by testing how the classifier, trained on the training set, classifies each element of the testing set.

The third important set is constituted of the points that need to be classified by the algorithm, for which the a-priori correct classification is unknown. This set will be referred to as *evaluation set*.

Given a probe point x_0, the k training points $x_r, r = 1, ..., k$ that are closest to the probe point are found: the probe point then is classified through a majority vote among the k neighbours. In case of a tie, which can happen if k is even, x_0 is classified at random.

In order to pick the k nearest neighbours for any given probe point, it is necessary to define a *distance* in the parameters space: the most immediate solution is to try the euclidean N-dimensional distance:

$$d_i^{euclidean} = \| x_i - x_0 \| \tag{4.14}$$

However, care must be taken when building a parameter space starting from variables which have different units. In this case, a variable with a wider distribution (i.e. with a larger variance) will contribute with a greater weight to the euclidean metric. A possible solution is to perform a transformation to the input variables prior to the construction of the parameter space, by transforming each one into a distribution with mean 0 and variance 1: the information given by the variable is preserved in the shape of the distribution, and the starting units do not influence the euclidean distance. Another solution is to rescale each variable i by a factor F determined from a width w_i of the distribution x_i for the combined sample of the events of the two classes: w_i is defined as the interval that contains the fraction F of x_i training values. Hence, the rescaled metric is:

$$d_i^{scaled} = \Big(\sum_{i=1}^{N} \frac{1}{w_i^2} |x_i - x_0|^2 \Big)^{\frac{1}{2}} \tag{4.15}$$

Both solutions are implemented in the TMVA package [114].

A peculiar feature of the k-Nearest-Neighbours algorithm that it takes into account of variations in the population density of the training sample: since k is constant, regions with sparse training points result in estimating the classification of a probe point using the properties of a wide region of the parameter space, whereas regions with dense training points result in a more local estimate that makes use of the properties of a smaller region of the parameter space. Another possible way of dealing with less populated regions of the parameter space is to generate toy events to populate those, in cases in which the multidimensional p.d.f. for the signal and background is known, as implemented in the Hyperballs algorithm [115].

The choice of k influences the smoothness of the decision boundary between the classes. In Fig. 4.1, taken from [8], can be seen the decision boundary between three classes (green, blue, and yellow) in a 2-dimensional parameter space as estimated by a nearest neighbour method with $k = 1$ (a) and $k = 15$ (b).

4.3 How to test an hypothesis against another one: the CL_s method

The CL_s method [116, 117] is part of a set of methods that are designed with the purpose of deciding whether the result of an experiment supports the hypothesis A or the hypothesis B. In high-energy physics, usually the two hypothesis are the *signal plus background* hypothesis and the *background-only* hypothesis.

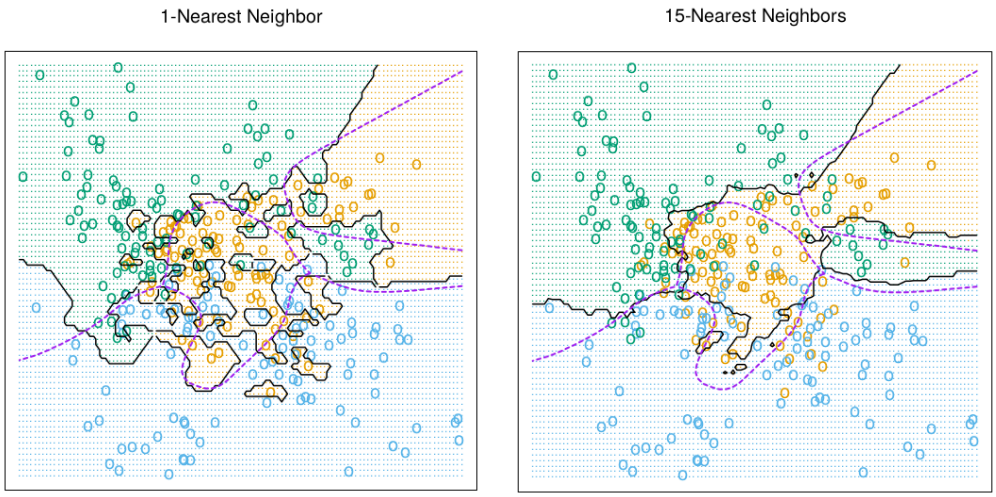


Figure 4.1: Decision boundaries for k-Nearest-Neighbour classifiers applied to a set of events picked from three classes (green, blue, and yellow points), in a 2-dimensional parameter space. The classifiers have been trained using $k = 1$ nearest neighbours (a) and $k = 15$ nearest neighbours (b), resulting in different smoothness for the decision boundary. The figures are taken from [8].

The likelihood function used to model the two hypotheses is given by the product of the poisson probabilities to observe n_i events in each signal region (or in each bin of the distribution used to separate the two hypothesis, equivalently: in the following, each signal region or category or bin will be generically referred to as *bin*)

$$L = \prod_i Poisson(n_i|\nu_i(\mu,\theta)) \cdot \prod_j Constraint(\theta_j, \tilde{\theta}_j) \tag{4.16}$$

where ν_i is the number of expected events for that bin, i.e. the sum of the expected signal events μs_i plus the expected background events b_i. The parameter μ, that multiplies the number of expected signal events, is called *signal strength modifier*, representing the (unknown) rate of the signal with respect to a reference cross-section (that determines s_i). The number of signal and the number of background events are both a function of a set of nuisance parameters θ, and can be written as $s_i(\theta)$ and $b_i(\theta)$. Each nuisance parameter θ_j represents one source of systematic uncertainty, the specific list depending on the details of each analysis. The part of the likelihood containing the product on the constraints on the nuisance parameters represents the knowledge one has about the nuisance parameters. Each *Constraint* term represents the probability for the true value of a nuisance parameter to be θ_j, given that the best estimate available for that parameter is $\tilde{\theta}_j$ (coming from another measurement, either real or imaginary). It must be noted that this probabilistic interpretation is somehow improper, since in frequentist statistics true values are fixed: in the frequentist approach, the expression *probability of the true value* does not make any sense.

The Poisson probability to observ n_i events in each bin is written as

$$Poisson(n_i|\nu_i(\mu,\theta)) = \frac{\nu_i^{n_i}}{n_i!} exp(-\nu_i) \tag{4.17}$$

The functional form of each *Constraint* term depends on the assumptions underlying the estimate of the systematic uncertainty associated to each nuisance parameter: the three functional forms that are used for the analyses object of this thesis are log-normal, gamma, and gaussian.

Log-normal constraints are used to represent uncertainties which correspond to multiplicative factors on the signal or background yields (e.g. theoretical errors on the cross section, efficiencies, etc.) and are of the form

$$\rho(\theta) = \frac{1}{\sqrt{2\pi} ln(k)} exp\Big(- \frac{(ln(\theta/\tilde{\theta}))^2}{2(ln(k))^2} \Big) \frac{1}{\theta} \tag{4.18}$$

The parameter k governs the size of the variation (for example, $k = 1.10$ represents the case in which the quantity of interest is known with an uncertainty of 10% relative to its nominal value).

Sources of systematic uncertainty that are statistical in origin, i.e. that originate from the statistical uncertainty on the number of events observed in a control region, are modeled by using a gamma distribution

$$\rho(n) = \frac{1}{\alpha} \frac{(\frac{n}{\alpha})^N}{N!} exp(-\frac{n}{\alpha}) \tag{4.19}$$

Here, the parameter N represents the number of events observed in the control region from which the expected event yield in the signal region, n, is extrapolated as $n = N \cdot \alpha$. The extrapolation factor α might be affected by an uncertainty: typically, the uncertainty on α is modeled by using an additional log-normal constraint.

The statistical fluctuation of the number of event in each bin are dealt with by using the Barlow-Beeston method [118, 119]. Each bin fluctuation is described by introducting a log-normal constraint acting on that bin only. This procedure ensures that the statistical fluctuations are properly treated as uncorrelated across all the different bins.

For the single lepton final state described in Appendix A, this full procedure resulted in a large number of nuisance parameters (almost 2500), yielding significantly long computation times and some instability in the determination of the values of the nuisance parameters that maximize the likelihood. To elimitate this problem, a trimming procedure that is part of the Barlow-Beeston approach is used: the statistical uncertainties of all backgrounds contributing to a given bin are added in quadrature (being statistically independent), and the resulting combined uncertainty is assigned to dominant background in that bin. Furthermore, the statistical uncertainties which are smaller than 10% times the background contribution expected in a given bin are neglected.

The process of setting limits on a physical quantity can be seen as a particular case of hypothesis testing, where several hypothesis for the parameter of interest are tested (e.g., in case of the cross section, the background-only hypothesis is tested against several signal-plus-background hypotheses

that differ one from each other by the size of the signal contribution, i.e. by the cross section of the signal).

In order to quantify to which degree the result of the experiment supports which hypothesis, it is necessary to start from a *test statistic*, i.e. from a function of the data. In the context of the searches performed in this thesis, the test statistics that is used is the *profile likelihood* that has been described in Sec. 4.1,

$$\tilde{q}_\mu = -2ln\frac{L(data|\mu,\hat{\theta}_\mu)}{L(data|\hat{\mu},\hat{\theta})},\ \text{constrainedby}\ 0 \leq \hat{\mu} \leq \mu \tag{4.20}$$

where *data* refers to the set of observations n_i. When only one signal region and no variable shape is exploited, this refers to the number of events observed in the signal region; when multiple signal regions or some shape is exploited this refers to the number of events observed in each category or each bin of the distribution. $\hat{\theta}_\mu$ is the set of nuisance parameters which maximize the likelihood function (Eq. 4.16 for a given value of the signal strength parameter μ. The set $(\hat{\mu},\hat{\theta})$, instead, is constituted by the set of values for the signal strength and the nuisance parameters that results in the global maximum for the likelihood function.

The constraint on $\hat{\mu}$ has two origins: $\hat{\mu} \geq 0$ guarantees that the signal strength has a physical interpretation (being it a multiplier of the signal cross section, it must be non-negative in order to yield a non-negative number of signal events); $\hat{\mu} \leq \mu$ constrains the test statistic forcing it to yield one-sided confidence intervals. One-sided confidence intervals are the only meaningful choice when testing a background-only hypothesis against a signal-plus-background hypothesis in terms of the signal cross section, in presence of sourcer of systematic uncertainty: intuitively, the signal would be detectable on top of the background only starting from a certain size up, whereas smaller signals would be as indetectable as the signal that results in the upper limit.

When computing the test statistic of Eq. 4.16 on the observed data, one obtains the *observed value* $\tilde{q}_\mu^{obs}$ given the value of the signal strength μ that is being tested.

In order to decide whether this observed value is more compatible with the background-only hypothesis or with the signal-plus-background hypothesis, it is necessary to compute the p.d.f.s $f(\tilde{q}_\mu)$ for each hypothesis. In order to do so, it is normally needed to generate toy events, grouped in *pseudo-datasets*, using MonteCarlo techniques. Each pseudo-dataset represents the possible outcome of one experiment, and separate pseudo-datasets are generated for the two hypotheses. First of all, the values of the nuisance parameters that maximize the likelihood function Eq. 4.16 are found: these are labeled as $\hat{\theta}_\mu^{obs}$ for the signal-plus-background hypothesis, and $\hat{\theta}_0^{obs}$ for the background-only hypothesis (remembering that formally the backgound-only hypothesis corresponds to the signal-plus-background hypothesis with null signal strength $\mu = 0$). Then, one pseudo-dataset at the time is generated by randomly varying all the nuisance parameters within their uncertainties: the values $\hat{\theta}_\mu^{obs}$ and $\hat{\theta}_0^{obs}$ that maximize the likelihood are used as starting values. The number of observed events is fluctuated as well, assuming the Poisson p.d.f. described in Eq. 4.17. Finally, the value $\tilde{q}_\mu^{toy}$ of the test statistic for the pseudo-dataset is computed using Eq. 4.20: for each pseudo-dataset, the values $\hat{\theta}_\mu^{obs}$ and $\hat{\theta}_0^{obs}$ are allowed to

fluctuate to values different from the values that maximized the likelihood used as starting points to generate the pseudo-dataset. The desired p.d.f. $f(\tilde{q}_\mu)$ for the two hypotheses is extracted from the distribution of the test statistic $\tilde{q}_\mu^{toy}$ for an ensemble of pseudo-datasets generated for each hypothesis: such p.d.f.s are denoted by $f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{obs})$ for the signal-plus-background hypothesis and $f(\tilde{q}_\mu|0,\hat{\theta}_0^{obs})$ for the background-only hypothesis. The two distributions are used to compute, according to the standard theory for hypothesis testing [111],the size and the power of the test:

$$\begin{aligned} p_\mu &= P(\tilde{q}_\mu > \tilde{q}_\mu^{obs}|signal-plus-background) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{obs})d\tilde{q}_\mu \\ 1-p_b &= P(\tilde{q}_\mu > \tilde{q}_\mu^{obs}|background-only) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|0,\hat{\theta}_0^{obs})d\tilde{q}_\mu \end{aligned} \tag{4.21}$$

The idea behind this is that if the value of the test statistics is larger than the value observed in data, $\tilde{q}_\mu^{obs}$, then the signal-plus-background is favoured. p_μ then represents the probability of the test statistic to fall into the critical region when the signal-plus-background hypothesis is true, whereas $1-p_b$ represents the probability of the test statistic to fall into the critical region when the background-only hypothesis is true: it is worth to notice that the latter implies that p_b represents the probability of the test statistic to fall outside the critical region (hence favouring the background-only hypothesis) when the background-only hypothesis is true.

Those two values are finally combined to obtain the CL_s value for a given value of μ

$$CL_s(\mu) = \frac{p_\mu}{1-p_b} \tag{4.22}$$

whose intepretation is then the ratio of the probability that the signal-plus-background hypothesis is favoured when it is indeed true, over the probability that the signal-plus-background hypothesis is favoured when the background-only hypothesis is true.

When a specific value of CL_s is chosen, $CL_s = c$, the value μ_f of μ such that $CL_s(\mu_f) = 1-c$ is denoted as "$100\cdot(1-c)$% confidence level (CL) upper limit".

In high energy physics at colliders, in the last couple decades the value $c = 0.05$ has become the standard choice to quote confidence levels, hence resulting in setting *95%CL limits* on the signal cross section.

The computation of the CL_s statistic using the procedure outlined above (i.e. based on the generation of MonteCarlo based ensembles of toys) is very computing intensive and scales quickly up with the number of nuisance parameters. It has been demonstrated [120] that, the profile likelihood test statistic and hence the p.d.f.s $f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{obs})$ and $f(\tilde{q}_\mu|0,\hat{\theta}_0^{obs})$ can be replaced with approximate formulas that are exact up to the order $\mathcal{O}(\frac{1}{\sqrt{N}})$ (yielding an error of ~ 20% when $N = 30$). Those formulas are based on the Wilks theorem, and the ensembles of pseudo-datasets used in the base method are replaced by one dataset, labeled *Asimov dataset*, that is defined as the dataset that, when used to evaluate the estimators for all the parameters, yields as a result the true values of the parameters: this works if the Asimov data are equal to their expectation values, which is true for large N, as shown in Sec. 4.1. Using a single dataset that makes use of expectation values of the relevant quantities yields a computing time

which is often several order of magnitudes smaller than the one taken by the full CL_s procedure based on ensembles of pseudo-datasets: when the amount of available events is large enough, consequently, a critical computational advantage can be obtained without any practical worsening of the final result.

Both the full and the asymptotic approach are implemented in the *Higgs Combination Tool* software package [121], based on the `RooFit/RooStats` packages [122], that has been developed in the context of the 2012 Higgs boson discovery.

Chapter 5

Precision tests of the Standard Model

5.1 Top mass precision measurement

5.1.1 Introduction

A measurement of the top quark mass, $m_{\rm top}$, in the dilepton decay channel $t\bar{t} \rightarrow (l^{-}\bar{\nu}_{l}b)\ (l^{+}\nu_{l}\bar{b})$ is presented, from proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC with the CMS detector. Events are selected from a data sample of 2.3 fb^{-1} by requiring two leptons, at least two jets, and missing transverse energy. The CMS collaboration developed two methods for the measurement of the top quark mass in the dilepton channel: one based on a full kinematic analysis (*KIN* method), and the other based on a matrix element method for weighting each event in the hypothesis that it is an event featuring a top quark pair decaying into dilepton (*MWT* method). The LIP group pursued the *KIN* method, slightly modified to exploit the b-tagging algorithms of CMS. The top quark mass is thus extracted by using likelihood template fits to data: this method, labeled *KINb*, is based on the full kinematical analysis and uses the kinematical characteristics of the events such as jet momenta and missing transverse energy to find an event by event estimate of m_{top}. A set of templates are constructed from simulated samples while the background contribution is modeled using data-driven techniques.

At the time of the study, the top quark mass $m_{\rm top}$ had been measured with a fantastic precision of approximately 1% at the Tevatron experiments [123]. With the 2010 data the CMS experiment measured, for the first time at the LHC, the top quark mass using the dilepton [124] and lepton+jets final states [125]. Experimentally, top quarks constitute an exception in the quark sector because no color singlet confined states containing top quarks have yet been observed. Due to its high mass ($m_{top} > m_W + m_b$) and to the preferred $t \rightarrow Wb$ decay, top quarks decay promptly without hadronizing. A precise determination of $m_{\rm top}$ may provide useful hints on unknown parameters of the standard model (SM), however a further improvement on the top quark mass measurement may be a challenging task at the LHC. In particular, a precise determination of $m_{\rm top}$ provides useful constraints, together with a determination of the mass of the W boson, on the mass of the Higgs boson, that at the time of the analysis was yet to be discovered. After discovery, the top quark mass determination still provides a useful consistency test for the SM.

In this manuscript, the measurement of the top quark mass is performed using the full dataset of

proton-proton collision at $\sqrt{s}$ =7 TeV data acquired by the CMS experiment until the end of 2011. The full kinematical (KINb) method is used to measure m_{top}; the study follows the previous results obtained in 2010 [124] making use of the improved method, originally developed by the CDF collaboration [126]. An alternative template fit to the lepton-jet invariant mass spectrum, used as a cross-check, is also briefly presented. This alternative method relies on a simple quantity without making use of the full kinematics of the event.

This section is organized as follows: in Sec. 5.1.2 the samples (data and simulation - MC) used are summarized and the event selection is discussed. The strategies adopted to reconstruct and measure m_{top} are presented in Sec. 5.1.3 as well as the procedures adopted to calibrate the two methods. The final results and the discussion of its associated uncertainties are discussed in Sec. 5.1.3. A summary is given in Sec. 5.1.4.

5.1.2 Dilepton event selection

In this section the datasets used are summarized, together with the trigger strategy, the offline event reconstruction and selection applied. Basic control distributions between data and MC are shown and a data-driven strategy used to estimate the number of background events from Drell-Yan (DY) is discussed. The event yields obtained in data and predicted for signal and MC are shown in the end of the Section.

Data samples, trigger and event selection

Samples

The first part of the data taken at a center-of-mass energy of 7 TeV, referred to as 2011A dataset, is used, which corresponds to a total integrated luminosity of 2165 pb^{-1}. A correction factor of 1.03 is applied to this estimate based on more recent results from [70]. For simplicity in what follows this number is rounded up to 2.3 fb^{-1}. The data samples are split into four categories, depending on the data taking conditions. In particular, the instantaneous luminosity changed during the full data taking period, yielding a different number of average pileup interactions in the events belonging to each part of the run period, ranging from 5.5 to 7.0, for an overall average of 6.2.

Reference simulated samples are used for comparison and for the purpose of evaluating biases, linearity and systematics. Table 5.1 summarizes the simulation samples used for all SM processes which are expected to contribute significantly to the $\mathrm{t\bar{t}}$ dilepton sample. The theoretical uncertainties on the cross sections are listed in the table and are taken from [127], [128], [129] and [130]. For single top (which is expected to contribute mainly through the tW channel) and for di-bosons (which are expected to contribute mainly through WW) the current experimental uncertainty for the measurement of the cross section at CMS is conservatively used. For single top, an uncertainty of 36% [131] and for di-bosons an uncertainty of 15% [132] are assigned. Other samples (mainly for the study of the systematics) are used in this study but will be referred throughout this section whenever necessary.

All the simulated samples have been generated using a pileup scenario that resulted to be different by the one observed in data: the procedure described in Sec. 3.5 is used to rescale the simulated events

to match the pileup conditions measured in data.

Table 5.1 summarizes the details of the simulated samples.

Table 5.1: List of the SM MC samples used in the comparison with 7 TeV data. For the different processes (signal and background) considered the expected cross sections and th. uncertainties are quoted. All samples have been generated using the PYTHIA tune "Z2", as described in Sec. 3.5.

Simulated process		
Process	Dataset	$\sigma \cdot BR \cdot k$ (pb)
W+jets	/WJetsToLNu_TuneZ2_7TeV-madgraph-tauola	$31314 \pm 407 \pm 1504$
$Z/\gamma^* \to \ell\ell$	/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/	$3048 \pm 34 \pm 128$
	/DYJetsToLL_M-10To50_TuneZ2_7TeV-madgraph/	12321
$\mathrm{t\bar{t}}$	/TTJets_TuneZ2_7TeV-madgraph-tauola/	$165\ ^{+4}_{-9}\ ^{+7}_{-7}$
Single top	/T_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/ (t)	$7.87 \pm 0.20\ ^{+0.55}_{-0.57}$
	/Tbar_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/ ($\bar{t}$)	$7.87 \pm 0.20\ ^{+0.55}_{-0.57}$
	/T_TuneZ2_t-channel_7TeV-powheg-tauola/ (t)	$41.92\ ^{+1.59}_{-0.21} \pm 0.83$
	/Tbar_TuneZ2_t-channel_7TeV-powheg-tauola/ ($\bar{t}$)	$22.6 \pm 0.50\ ^{+0.68}_{-0.91}$
	/T_TuneZ2_s-channel_7TeV-powheg-tauola/ (t)	$3.19 \pm 0.06\ ^{+0.13}_{-0.10}$
	/Tbar_TuneZ2_s-channel_7TeV-powheg-tauola/ ($\bar{t}$)	$1.44 \pm 0.01\ ^{+0.06}_{-0.07}$
Dibosons	/ZZ_TuneZ2_7TeV_pythia6_tauola/	5.9 ± 0.15
	/WW_TuneZ2_7TeV_pythia6_tauola/	$47.04\ ^{+2.0}_{-1.5}$
	/WZ_TuneZ2_7TeV_pythia6_tauola/	18.2 ± 0.7

Trigger

Double lepton triggers are used to select events in data. Events triggered by double electron, double muon or cross electron-muon triggers are selected exclusively for the ee, $\mu\mu$ and $e\mu$ channel, correspondingly. Due to the rapid increase of the instantaneous luminosity in 2011, the lepton transverse momentum thresholds were increased during the data taking period in order to keep the trigger rates within the data acquisition capabilities. The p_{T} thresholds ranged from 8 to 17 GeV, depending on the lepton flavor and data taking period.

Offline event reconstruction and selection

For the selection of the $\mathrm{t\bar{t}}$ dilepton sample, a standard event selection is followed that is commonly used by analyses performed in the context of top quark physics. Such selection is briefly summarized below. Details can be found in [133]. The event reconstruction, based on Particle Flow (PF) [80], is performed as described in Sec. 3.4. At least two prompt and isolated leptons with p_T >20 GeV/c and $|\eta|$ <2.4 are required. For the same flavor channels (i.e. $ee/\mu\mu$ events), the dilepton invariant mass is required to be $|M - M_Z|$ >15 GeV/c^2, reducing significatively the contamination from Drell-Yan events. For all channels it is also required that M >12 GeV/c^2, in order to exclude events in which the dilepton pair comes from heavy flavour resonances or low-mass Drell-Yan resonances. This veto removes almost no signal events.

For all channels it is further required that two jets with p_T >30 GeV/c and $|\eta|$ <2.4 are reconstructed. L1, L2, and L3 (as well as extra L2L3Residual for data) jet energy corrections, and the charged hadron subtraction scheme, have been applied as described in Sec. 3.4.

It is further required that E_T^{miss} >30 GeV: in the same-flavour channel in order to reduce the contamination from DY processes, in the opposite flavour channel in order to uniformize the expected resolution

between the same-flavour and opposite-flavour channels.

A slight excess in the region vetoed by the event selection is observed in the same-flavor channels. This excess is related to an extra contamination from DY events. In the next section a data-driven strategy to control this contamination is described.

Control of the Drell-Yan contamination in the selected sample

A method is applied based on a distribution which shows some discrimination power between Drell-Yan (DY) and $\mathrm{t\bar{t}}$ events, and which is less prone to mis-modelling of the dilepton invariant mass. This is particularly crucial if there are: mis-calibration effects as in the case of the ee channel, uncertainties in the low mass region ($M_{\ell\ell}$ <50 GeV/c^2) where a leading-order generator has been used (i.e. PYTHIA) or an incomplete modelling of other sources of prompt dileptons (e.g. from di-bosons).

The angle between the direction of the two leptons is chosen, as it is a fairly robust variable which is almost unaffected by the energy calibration and which reflects the fact that in DY the two leptons are produced after the decay of a single particle even if off-shell. In $\mathrm{t\bar{t}}$ events, the leptons originate from different top decay chains and the possible correlations are expected to be negligible. It is expected therefore that in $\mathrm{t\bar{t}}$ events the angle between the leptons (i.e. $\cos\theta_{\ell\ell'}$) is uniformly distributed, but not in DY events. The same is valid for the remaining most significant processes: tW, WW and lepton fakes. The angle between the two leptons is defined from the measured 3-momenta:

$$\theta_{\ell\ell'} = \arccos\left[\frac{\vec{p}_\ell \cdot \vec{p}_{\ell'}}{|\vec{p}_\ell||\vec{p}_{\ell'}|}\right] \tag{5.1}$$

The angle is shown in Fig. 5.1 in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ control region ($E_{\mathrm{T}}^{\mathrm{miss}}$ <30 GeV) and in the signal region for the same-flavor channels. A clear excess of events at small angles is found in data with respect to the simulation in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ region, pointing out to a need for correcting the simulation to account for this effect.

As in DY events $E_{\mathrm{T}}^{\mathrm{miss}}$ is due to instrumental effects and it is mostly expected to come from the mis-measurement of the jet energy, one can assume, in first approximation, that the boost of the dilepton pair is equivalent for $E_{\mathrm{T}}^{\mathrm{miss}}$ <30 GeV and for $E_{\mathrm{T}}^{\mathrm{miss}}$ >30 GeV for the same jet multiplicity bin. Under this assumption, the shape of the angle between the two leptons is assumed not to vary significantly by splitting the DY sample in the two $E_{\mathrm{T}}^{\mathrm{miss}}$ regions. Given the fact that the low $E_{\mathrm{T}}^{\mathrm{miss}}$ regions is mostly dominated by the DY contribution, the following procedure to fit the DY contribution in the signal region is applied: the shape is obtained from DY events in data, using the low $E_{\mathrm{T}}^{\mathrm{miss}}$ control region where the contributions from other processes are negligible; the $\mathrm{t\bar{t}}$ and non-DY background shapes are obtained by using the simulated events in the signal region; finally, the number of DY events (other processes), i.e. N_{DY} (N_{other}) are extracted in the $ee/\mu\mu$ sample by maximizing an extended binned likelihood where the two numbers are left to float freely in the fit.

Figure 5.2 shows the result of the fit used to extract the DY contribution.

The result of the fit is stable against variations of the template for the non-DY processes. After varying separately jet energy scale ($\pm1\sigma$), jet energy resolution and the pileup scenario ($\pm$0.6 interactions) the

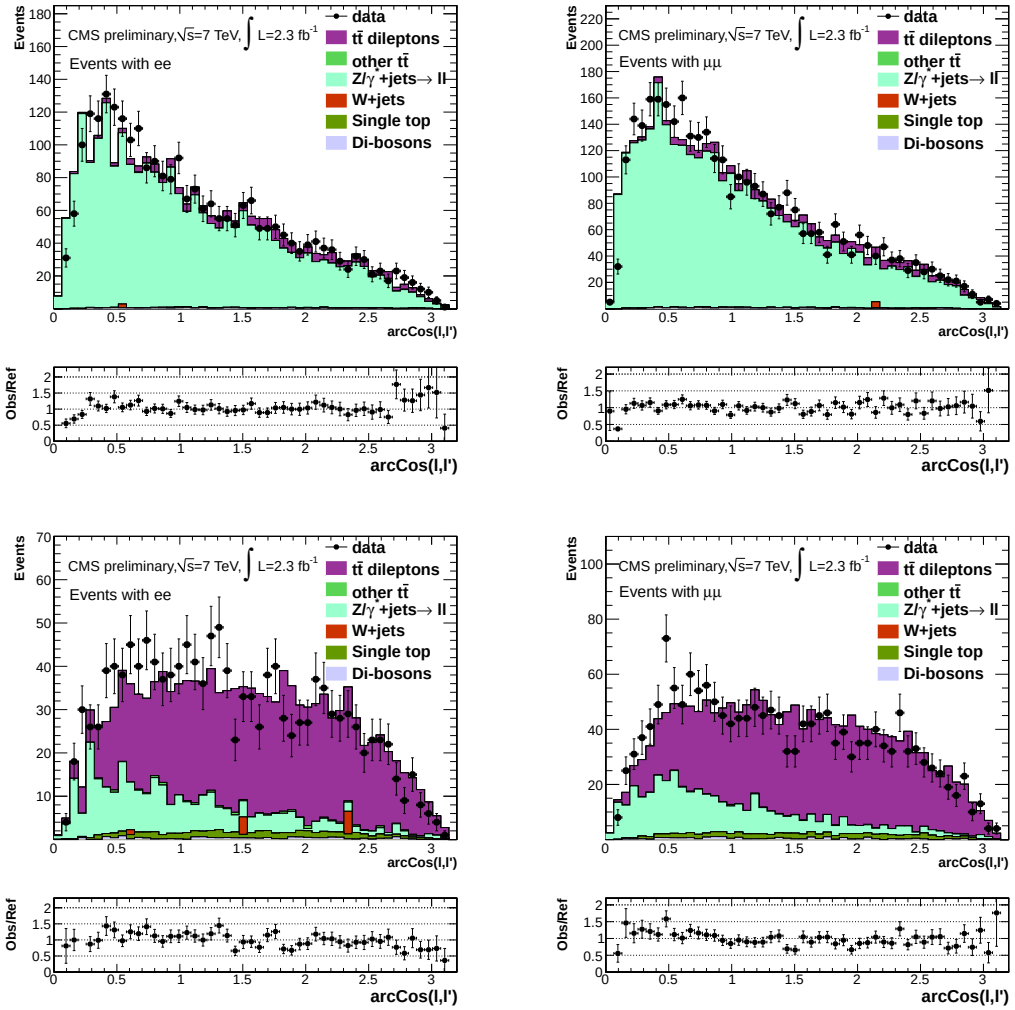


Figure 5.1: Angle between two electrons (*left*) or muons (*right*) for events in the low $E_{\mathrm{T}}^{\mathrm{miss}}$(<30 GeV) control region (*top*) and in the final $E_{\mathrm{T}}^{\mathrm{miss}}$ (> 30 GeV) selection region (*bottom*).

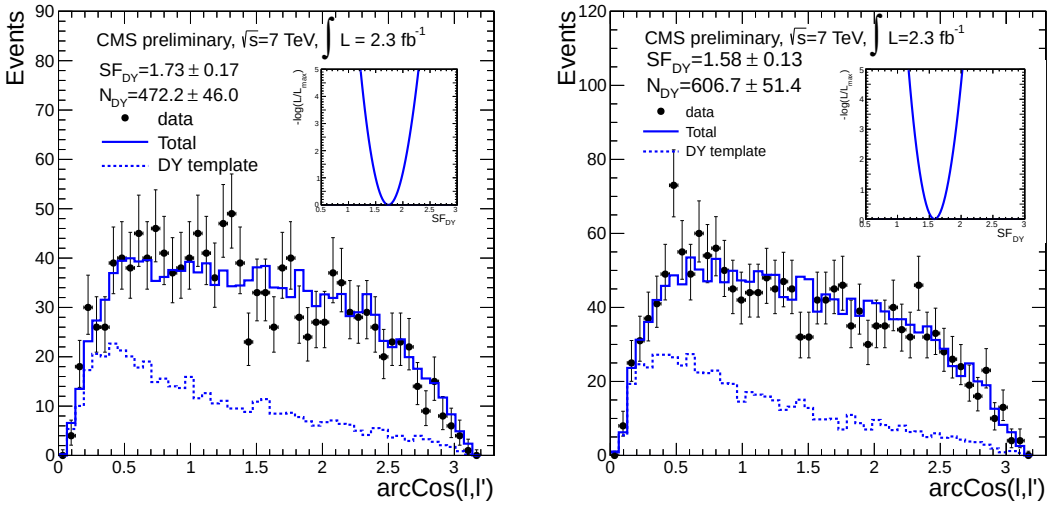


Figure 5.2: Result of the template fit to the angle between the two selected electrons (*left*) and muons (*right*). The inset shows the likelihood of the fit as function of the scale factor for the predicted yields from simulation for the DY contribution.

result of the fit is observed to vary by <1%. Likewise, any uncertainty on the luminosity, selection efficiency or theoretical cross sections are absorbed by the fact that the fit is done simultaneously on the number of DY and non-DY processes. The result above is not surprising as the variable chosen is fairly robust and independent of the uncertainties considered above. The main systematic may however be assigned to the shape chosen for the DY contribution. This is evaluated using MC closure tests for both the 1 jet bin (not considered for the measurement) and the $\geq$ 2 jets bin used to select the final sample. Results indicate a good agreement (within <5%) for all channels except the $\mu\mu$ with $\geq$2 jets where a bias of 15% is found. The details of these studies can be found elsewhere [134]. The closing of the MC test for events with $\geq$2 jets is partially affected by the statistical size of the sample and to the presence of a handful of events with significant weight assigned from the pileup re-weighing prescription, and an overall 15% systematic uncertainty is applied on the estimate of the DY contribution to the final sample.

In the $e\mu$ sample, extra contamination from DY is not expected to be as critical as in the same-flavor channel as $Z \to \tau\tau$, where both τs decay leptonically to e or μ, has a lower branching ratio and yields softer leptons. In this case a consistency check is done using a distribution which has some separation power to distinguish between $\mathrm{t\bar{t}}$ and $Z \to \tau\tau$: the sum of the transverse mass of each lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector [1]. The $\sum M_T$ variable is expected to reflect the fact that the leptons from τ decays tend to be softer than the leptons from prompt W decays and the neutrinos emitted after a τ decay are approximately collinear to the direction of the charged leptons. Therefore $\sum M_T$ is expected to have lower values with respect to the $\mathrm{t\bar{t}}$ dilepton channel. In this case, a consistency check is performed by using a data-driven generated sample of $Z \to \tau\tau$ events, obtained from a data sample of $Z \to \mu\mu$ candidates where the muons are replaced with simulated taus. The $\sum M_T$ distribution is obtained from the replacement sample and use it as a reference, instead of using the shape predicted from simulation. For the other processes the shape predicted by the simulation is used, and the same fit procedure as described for the same-flavour case is applied. As for the same-flavor channels, the variation of the templates for the other processes due to jet energy scale and resolution, pileup, luminosity or selection efficiency yields minor variations of the results (<1%). In this case the template is also less prone to systematic uncertainties from the definition of the control region, as the control region coincides with the signal region and the source of the template is data after the replacement of the di-muon candidate by a di-tau candidate. It can therefore be concluded that in this case the uncertainty of the estimate is driven by the statistical uncertainty.

5.1.3 Mass measurement: the Full Kinematic Analysis (KINb) method

For each selected event a full KINematical analysis (KINb) [124] is performed in order to reconstruct the mass of the candidate events. The method is based on the equations for the energy-momentum

[1] The transverse mass of a lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is computed as: $M_T(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) = 2p_T^{\ell} E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\phi)$ where p_T^{ℓ} is the transverse momentum of the lepton, $E_{\mathrm{T}}^{\mathrm{miss}}$ is the missing transverse energy and ϕ is the angle between the two vectors.

conservation of the decay products from each top quark and W boson which can be written as:

$$P(t) = P(b) + P(W^+) \qquad P(W^+) = P(l^+) + P(\nu_l)$$
$$P(\bar{t}) = P(\bar{b}) + P(W^-) \qquad P(W^-) = P(l^-) + P(\bar{\nu}_l) \tag{5.2}$$

Eqs. 5.2 have a total of 2×3×4=24 free parameters corresponding to the energy-momenta of the final state particles.

In the KINb method (originally from Ref. [126]), the dilepton and two jets are considered. In events with more than two jets the criteria for the selection is based on b-tagging information giving preference or strictly using the leading b-tagged jets as it will be detailed below. As b-tagging information is used to select the jets, the method is named *KINb*.

The usage of b-tagging information in order to select the jets is a key factor: the original *KIN* method prescribed to use the leading jets, regardless of their b-tagging status. However, the presence of initial or final state radiation might cause the leading jets to not be the ones originating from the decays of the top quarks. When using the b tagging information, though, the fraction of correctly assigned jets can be increased by 30% [135]. Consequently, the b-tagged jets are selected in an event even if they are not the leading jets. In case there are less than 2 b-tagged jets in an event, the selected b-tagged jet is supplemented by the leading jet. It is demonstrated using Monte Carlo simulation that the proportion of events in which the jets used for the reconstruction are correctly matched is increased significantly.

Beside the dilepton pair, the two selected jets, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement are also used. $E_{\mathrm{T}}^{\mathrm{miss}}$ is used to constrain the sum of the transverse momentum of the neutrinos produced from the two $W \to l\nu$ decays from the $\mathrm{t\bar{t}}$ system [2].

The remaining degree of freedom needed to solve the equations for the kinematics of the $\mathrm{t\bar{t}}$ system is the longitudinal imbalance, i.e. $p_z^{t\bar{t}}$. This quantity is expected to be almost independent of the top quark mass and of the MC generator used.

Table 5.2 summarizes the measured experimental variables and the constrains as well as the number of free parameters entering Eqs. 5.2. It can be concluded that the $\mathrm{t\bar{t}}$ system kinematics of the dilepton channel is fully specified by the KINb method.

Table 5.2: Parameters used to reconstruct the kinematics of a $\mathrm{t\bar{t}}$ decay in the dilepton channel.

Object	Parameter	Type	$N_{parameters}$	Total
jets	$(E, \vec{p})$	measurement	2×4	8
charged leptons	$(E, \vec{p})$	measurement	2×4	16
neutrinos	$\vec{E}_T^{miss}$	measurement	2	18
	$m_\nu = 0$	constraint	2	20
W boson mass	m_W =80.398±0.025 GeV/c^2	constraint	2	22
t mass	$m_t = m_{\bar{t}}$	constraint	1	23
$\mathrm{t\bar{t}}$ longitudinal balance	$p_z \sim \mathcal{G}aus(0, \sigma)$	hypothesis (MC)	1	24

[2] In the case of intermediate production of τ leptons three neutrinos are expected to be present in the final state. From the point of view of Eqs. 5.2, a single neutrino can be assumed with a 3-momentum corresponding to the sum of the momenta of the three neutrinos for the final state of the W which decayed to a τ lepton.

Given the resolution on $p_z^{t\bar{t}}$ and also on the jet (and consequently $E_{\mathrm{T}}^{\mathrm{miss}}$), the KINb method, attempts to solve the kinematics equations several times ($5{\cdot}10^4$ in our study) trying different $p_z^{t\bar{t}}$ hypotheses and smearing jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ resolutions. The $p_z^{t\bar{t}}$ hypotheses are assumed from the double gaussian fit while jet (and $E_{\mathrm{T}}^{\mathrm{miss}}$) resolutions are smeared according to the parameterization of the resolutions obtained from QCD simulation [136]. After solving numerically the equation for a given $p_z^{t\bar{t}}$ and jet resolution hypothesis, the result is accepted if the reconstructed top quark mass in both decay legs are within 3 GeV/c^2. Different top quark mass solutions may be found for different hypotheses and a further counting procedure of the solutions needs to be applied. The final mass estimate is extracted from the most probable value observed.

Notice also that for each selected event there are two possible lepton-jet assignments which can be used to reconstruct the top quark mass. If mass solutions are found for both assignments the choice of the lepton-jet assignment has to follow a pre-defined arbitration to apply on both data and MC. For this study the lepton-jet assignment is chosen from the following set of rules:

- lepton-jet assignments which yield a low number of solutions (i.e. $< 0.5\%$ of the total attempts to solve the equations) are discarded;
- the lepton-jet assignments which yields the higher number of solutions is preferred;
- if the number of solutions found for each lepton-jet assignment is similar (i.e. within a 90% difference) then the assignment with the highest number of counts for its most probable value is preferred.

Using "MC truth" to monitor the performance of this arbitration, it is found that 99 $\pm$ 1% of signal events (in which both b-jets are used as input for the mass reconstruction) are expected to yield valid solutions. For these events the correct lepton-jet pair assignment is chosen with a success rate of 74 $\pm$ 1%. An example of the distribution of the solutions found for a given event and how the best combination is chosen is shown in Fig. 5.3.

Event selection for the mass measurement

Before reconstructing the top quark mass, the event selection is further studied in order to maximize the fraction of correct assignments in the selected events, i.e. the number of events where both the b jets from the $t \to Wb$ decay have been reconstructed and selected. For this purpose solely the simulation of $\mathrm{t\bar{t}}$ dilepton events is relied on. In order to minimize further the effect of misassignments in events with more than two jets, the performance of different b-tagging algorithms is studied: the CSV algorithm, described in Sec 3.4, has been found to have the best double b-tagging efficiency in signal events (80%, against the $60 - 74\%$ of other algorithms probed). Scale factors are derived to correct for the difference in b-tagging efficiency between data and simulation, as described in Sec. 3.4.

The event selection described above is applied to data and simulation, and the KINb method is used to reconstruct the top quark mass using the jet selection based on the CSV ranking. The events with valid top quark mass solutions are categorized according to the b-tagging multiplicity which is corrected

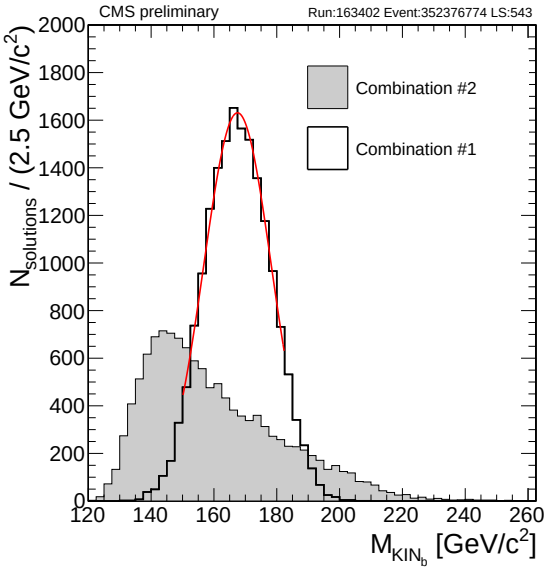


Figure 5.3: Top quark mass solutions for the KINb method for the two lepton-jet combinations after smearing the jet energy resolution for one selected event in data. The combination #2 is chosen in this case; the solid red line corresponds to the Gaussian fit used to estimate m_{KINb} (see text).

for simulation with the method previously proposed. Table 5.3 summarizes the inclusive event yields observed in data and in simulation.

Selected sample for the analysis

Table 5.3 summarizes the event composition of the selected inclusive sample after using the data-driven prediction for the DY yields. In computing these final yields, the trigger efficiencies for each channel are also applied. The average values applied to the simulation, in the form of data-to-MonteCarlo scale factors for reweighting simulated events, are $SF_{ee} = 0.96$, $SF_{e\mu} = 0.97$ and $SF_{\mu\mu} = 0.93$ for the dielectron, dimuon and electron-muon channels correspondingly. The values are taken from the reference $\mathrm{t\bar{t}}$ dilepton cross section measurement [137]. The data-driven data to MC scale factors for the offline selection efficiencies are not applied as they are mostly close to one, and a final systematic uncertainty is assigned to this choice.

Some of the events are rejected due to the fact that the KINb method does not find a valid solution. From simulation it is expected that 29$\pm$4% of background events are rejected, while 83$\pm$1% of the signal events are kept. In data, the fraction of events found with valid KINb solutions is compatible with the expectations from simulation. Figure 5.4 show the reconstructed top quark mass per b-tag multiplicity. An overall good agreement is found in both the core and the tail of the distribution between data and direct MC prediction for a top quark mass of 172.5 GeV/c^2. Different combinations will be studied in Sec. 5.1.3 and the expected yields will be used to constrain the background contribution in the fit as it will be detailed later.

Table 5.3: Number of expected events for the different background contributions are compared with data, after the selection requirements and after KINb reconstruction. The events with valid KINb solutions which have 1 or at least 2 b-tagged jets are shown in the leftmost columns. The total uncertainty (statistical and systematic) is shown. The systematic uncertainty reflects the uncertainty in the luminosity (4.5%), theoretical cross sections, jet energy scale and resolution, contamination from pile-up, trigger and selection efficiencies and the limited statistics in the MC samples.

Process	Pre-selection	KINb	=1 b-tags	$\geq$ 2 b-tags
Di-bosons	73 ± 14	55 ± 10	18 ± 4	4 ± 1
Single top	247 ± 92	182 ± 68	88 ± 33	76 ± 29
W+jets	22 ± 10	16 ± 8	8 ± 6	-
$Z/\gamma^* \to \ell\ell$	1091 ± 97	756 ± 71	238 ± 29	47 ± 11
other $\mathrm{t\bar{t}}$	32 ± 4	28 ± 3	11 ± 2	14 ± 2
$\mathrm{t\bar{t}}$ dileptons	5057 ± 463	4209 ± 385	1379 ± 127	2623 ± 240
total expected	6522 ± 482	5246 ± 398	1742 ± 134	2765 ± 242
data	6358	5047	1692	2620

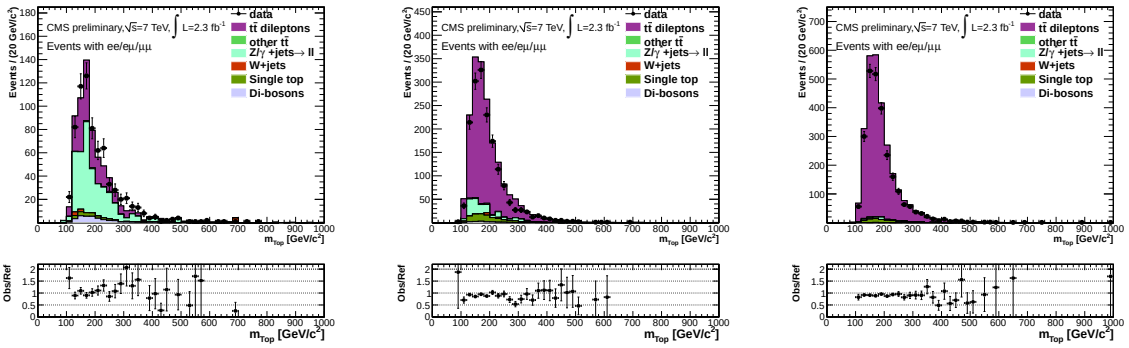


Figure 5.4: Reconstructed top quark mass for events with no (*left*), one (*center*) or at least two (*right*) b-tags.

Signal modeling

The signal component in the mass spectrum is modeled using MC simulation. Samples at different top quark mass values are generated using the MADGRAPH +PYTHIA generators and the full simulation of the CMS detector, as described in Sec. 3.5. The masses for the generated sample range from 161.5 GeV to 184.5 GeV, with steps of 3 GeV (except the first step, which is 2 GeV).

The mass distribution of the selected signal events is constructed for each sample using the KINb method. The set of the different distributions is fit simultaneously with a Gaussian+Landau template where the free parameters are assumed to be linear functions [3] of the top quark mass:

$$P_{signal}(m|m_t) = \frac{\alpha(m_t)}{I_l}\mathcal{L}_{\text{andau}}\left[m, \text{mpv}(m_t), \sigma_l(m_t)\right] + \frac{1-\alpha(m_t)}{I_g}\mathcal{G}_{\text{auss}}\left[m, \mu_g(m_t), \sigma_g(m_t)\right] \tag{5.3}$$

Figure 5.5 exemplifies the results of the fit of the generated mass distributions to Eq. 5.3 when using the b-tag information to rank the jets, for events with at least 2b-tagged jets, respectively. Signal is expected to dominate the $\geq 2b$-tag multiplicity bin, while a smaller fraction is expected to be found in the $1b$-tag bin. The 0 b-tags multiplicity bin is not considered as it is expected to be mostly dominated by background. It is shown, in both cases, that the average reconstructed top quark mass by the KINb method evolves linearly with the input top quark mass, pointing to the necessity of a calibration which is described in Sec. 5.1.3.

Background modeling

The mass distribution of the selected background events is constructed for each sample using the KINb method. Two distinct contributions are considered: DY and other background processes (which includes other $\mathrm{t\bar{t}}$, di-bosons, single top and W+jets). The DY is discussed separately as two data-driven methods can be used to model the shape of this specific background. For the other background processes the different distributions are weighted according to the relative expectations for the event yields (see Sec. 5.1.2), and then added together. Those are discussed in the next paragraphs.

The modeling of the background template is, in both cases, fit to a Gaussian+Landau shape, motivated by the distinction of the background shape that, on the contrary of the signal distribution, has a turn-on curve which is determined mainly by the selection cuts on the reconstructed objects.

The Non-DY fit is performed to the sum of the mass distribution of the background processes in the signal region, and its results are shown in Fig 5.6 (*left*): it can be observed that the Gaussian contribution tends to vanish in the fit, and is therefore neglected in the following. A Gaussian+Landau template is fit to the sum of the mass distribution of the background processes in the signal region: the result of the fit is shown in Fig 5.6 (*left*). It is observed that the Gaussian contribution tends to vanish after the fit and it is therefore neglected in the following.

The DY fit is performed separately in the same and opposite flavour final states: in the same flavour final state, events with at least two jets and $E_{\mathrm{T}}^{\text{miss}}$ >30 GeV, within the Z-mass peak (i.e. $|M_{ll} -$

[3] The linear functions are assumed in the form: $(m_t - 172) \cdot slope + intercept$.

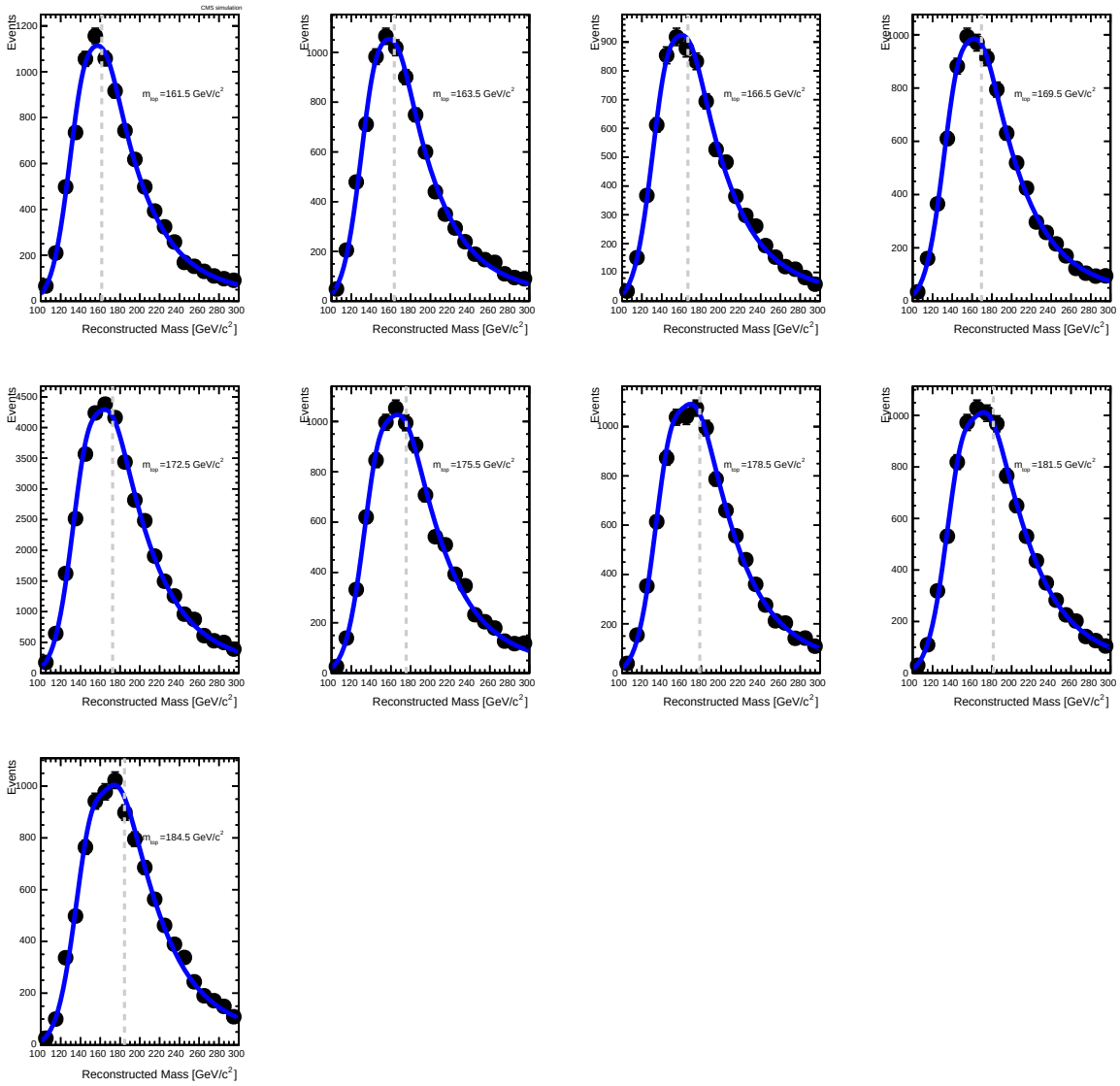


Figure 5.5: Top quark mass distributions for dilepton events generated with different input masses using b-tag information to rank the jets. The result of the combined fit to events with at least 2b-tags is superimposed on the distributions.

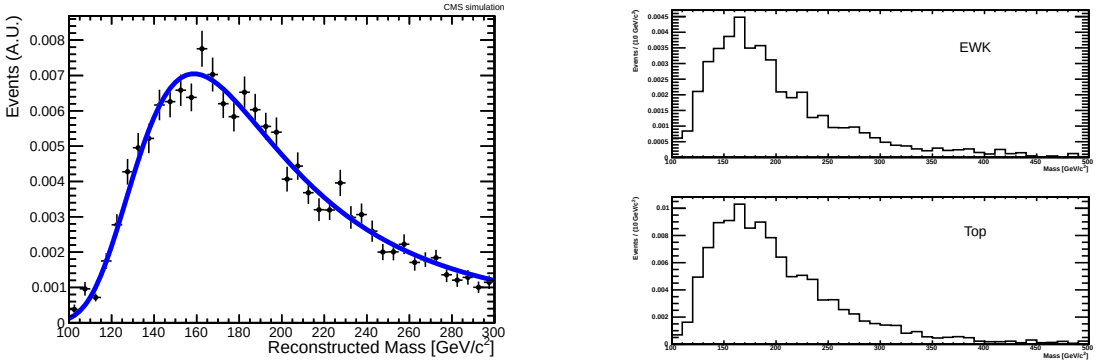


Figure 5.6: Top quark mass distributions for background processes. The result of the fit to the weighted sum for all the processes (*left*). The solid line shows the Landau component of the PDF fit to the reconstructed mass. The individual background contributions for the combined electroweak processes (i.e. di-bosons) (*right, top*), and for non-dilepton $\mathrm{t\bar{t}}$ and single top (*right, bottom*).

$M_Z| <15$ GeV/c^2) are used. Notice that these $ee/\mu\mu$ events are removed in the standard selection by the Z-veto (cf. Sec. 5.1.2). A further veto on the presence of b-tagged jets reduces by 95% the $\mathrm{t\bar{t}}$ contribution, while maintaining more than 60% of the Z candidates. The reconstructed top mass distribution is fit and, as in the Non-DY case, the Gaussian component is found to be vaninshing and is therefore dropped. The curves resulting from the fit are shown in Fig. 5.7. A fair agreement is found between the parameters in the control and signal regions in the MC and the result of the fit to data in the control region is therefore taken as the DY shape template

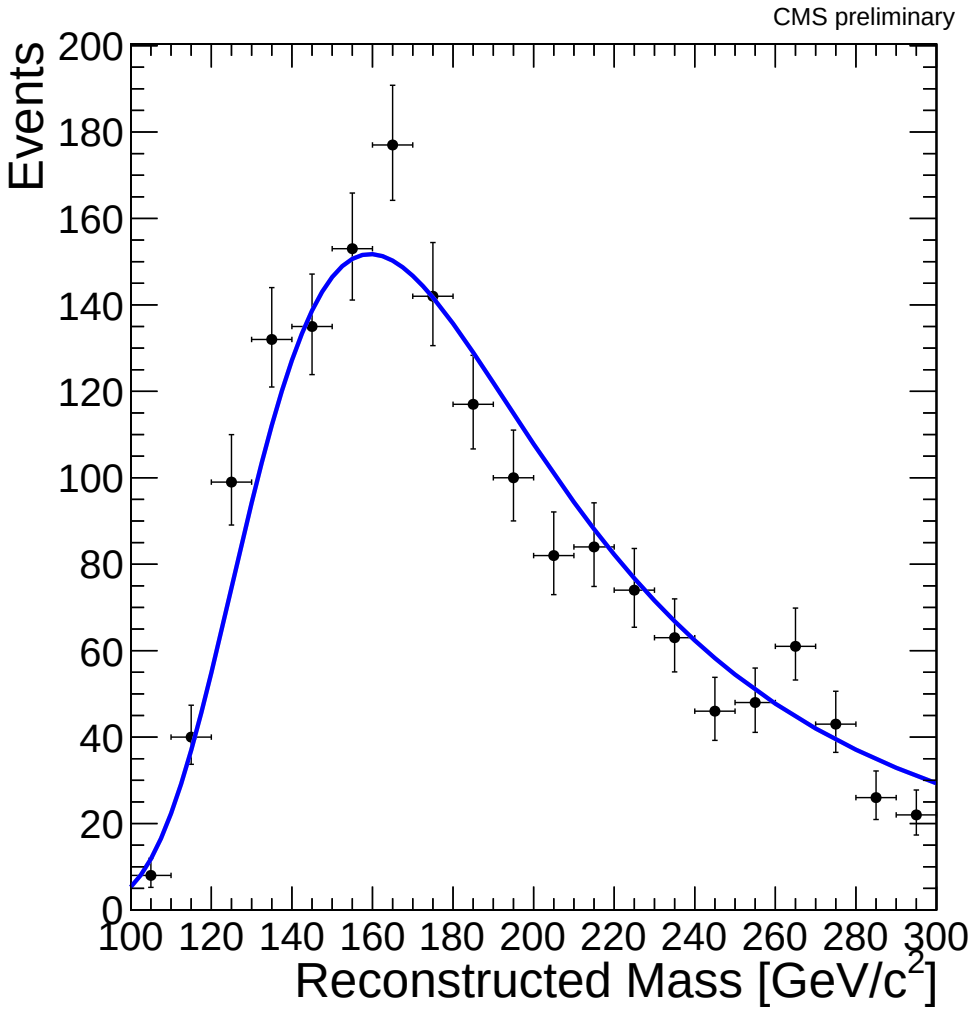


Figure 5.7: Distribution of the reconstructed top quark mass for same flavor events in the data-control regions with the template fit superimposed.

In the opposite flavor channel channel, the contamination from DY is expected to come mainly from di-τ decays which smear out the Z mass peak reconstructed from the two leptons, due to the associated production of neutrinos. In this case a data-driven generated sample of $Z \to \tau\tau$ events is used, obtained from the $Z \to \mu\mu$ candidates in data where the muons are replaced by simulated taus. The samples used for this purpose are the same as those used in Sec. 5.1.2. The full event selection (excluding trigger) is applied to these samples and the top quark mass is reconstructed in the selected events. Again, a Landau curve is observed to be sufficient to model the shape and the results are shown in

Fig. 5.8, for simulation and for data. In the opposite flavor channel the statistics is lower.

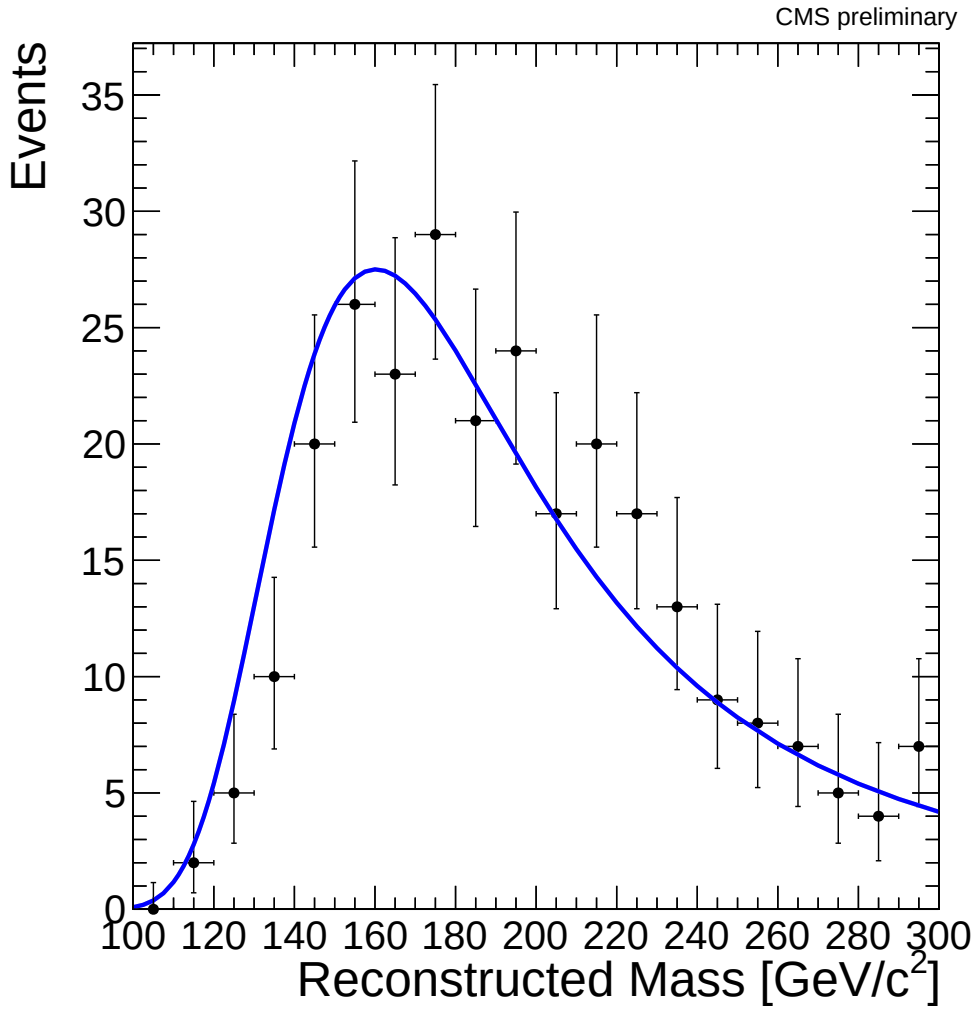


Figure 5.8: Distribution of the reconstructed top quark mass for $DY \to \tau\tau \to e\mu$ events with the template fit superimposed, obtained in the data-driven $\mu \to \tau$ replacement sample.

Top mass fit from the KINb reconstruction

Events with masses in the range 100-300 GeV/c^2 are used to fit the top quark mass. The top quark mass is measured by maximizing an unbinned likelihood:

$$\mathcal{L}(m_t) \;=\; \mathcal{L}_{\text{shape}}(m_t) \;\times\; \mathcal{L}_{n_b} \tag{5.4}$$

where

$$\mathcal{L}_{\text{shape}}(m_t) \;=\; \frac{e^{-(n_s+n_b)}(n_s+n_b)^N}{N!} \prod_{i=1}^{N} \frac{n_s \mathcal{P}_s(m_i|m_t) + n_b \mathcal{P}_b(m_i)}{n_s + n_b} \tag{5.5}$$

and

$$\mathcal{L}_{n_b} = \mathcal{G}_{\text{auss}}(n_b, \bar{n}_b, \sigma_{n_b}) \tag{5.6}$$

where n_s (n_b) is the estimated number of signal (background) events, $\mathcal{P}_s$ ($\mathcal{P}_b$) the signal (background) templates, and σ_{n_b} the uncertainty associated to the background distribution. The likelihood (Eq. 5.4) returns a value for the top quark mass (m_t), and the estimated number of signal (n_s) and background (n_b) events. For each event, the signal and background templates are used to assign a probability that each event is either signal or background based on the value of the top quark mass measured. Roofit [138] is used to implement Eq. 5.4 and to perform the fit to data.

The base likelihood described by Eq. 5.4 fits the top quark mass (and the event yields for signal and background) in an inclusive sample. The fit can however be extended in order to be performed simultaneously to exclusive categories of events - e.g. according to the flavor of the dilepton, the number of b-tags or a combination of the two previous. The extension to a combined fit is simple - a product of likelihoods with a common top quark mass as input (but exclusive signal and background event yields per category) is used instead of Eq. 5.4:

$$\mathcal{L}(m_t) = \prod_{k \in \{\text{categories}\}} \mathcal{L}^k_{\text{shape}}(m_t) \times \mathcal{L}^k_{n_b} \tag{5.7}$$

where the $\mathcal{L}^k_{\text{shape}}$ and $\mathcal{L}^k_{n_b}$ are given by Eqs. 5.5 and 5.6 correspondingly for the k-th category considered. In the following the results of two approaches are compared:

inclusive - fit to the inclusive dilepton channel combining the results of the fit to the 1 or $\geq$ 2 b-tags categories

combined - combination of the fit to the opposite flavor and same flavor channels with 1 or $\geq$ 2 b-tags categories

The next Section is devoted to calibrate the results of these fits using the background and signal templates previously derived. The bias and linearity of the fit are studied in detail with the purpose of calibrating the fit used to extract the top quark mass from data. For that purpose pseudo-data generated from MC pseudo-experiments are used.

Calibration of the fit

In order to minimize the residual bias resulting from the parameterization obtained from the combined fit procedures described in the previous Sections for both signal and background processes, pseudo-experiments are performed with dilepton event candidates from MC samples.

A residual bias might occur due to an imperfect modelling of the signal with Eq. 5.3, to the statistical size of the sample, to the statistical variation of the background contribution, etc..

Due to the limited size of the simulated signal samples, pseudo-experiments are build starting from the spectrum of the reconstructed mass derived using the full statistics available, for signal and back-

ground processes separately, and then drawing randomly measurements from a Poisson distribution with average corresponding to the expected number of events.

Each pseudo-experiment is then constructed by summing up the measurements obtained from each process to obtain a single top mass measurement distribution. The procedure is repeated as many times as needed.

After maximizing Eq. 5.7, the distribution of the values extracted for $m_{\rm top}$ is used to calibrate the measurement of the top quark mass, in particular the parametrization of the signal template.

The calibration of the fit consists in an additional linear fit which is performed for each category used in the fit. The corrected top quark mass $m_{\rm top}^{\rm corr}$ to be measured is obtained after the following transformation of variables $m_{\rm top} = {\rm bias} + {\rm slope} \cdot m_{\rm top}^{\rm corr}$, where m_t is used as the parameter for the signal and background PDFs. The bias term correspond to an overall shift of the mass measurement, wherease the slope is introduced to account for an eventual dependence of the correction on the input top quark mass. The top quark mass values from the *KINb* algorithm are shown in Fig. 5.9 (top) as a function of the generated top mass: the linear fit to the point is also shown. The bias is then defined as the difference between the actual reconstructed values (*y* coordinate of the black points in the plot) and the values returned by the linear fit in correspondence of each generated mass point (i.e. the difference in *y* coordinate between the black points and the grey line representing the linear fit). The bias is shown in Fig. 5.9 (bottom) as black dots, and a further linear fit to the bias as function of the input top quark mass is performed: even if the correction introduced by the previous linear fit should result in an average bias of zero, with no dependence on the generated mass, this additional linear fit is performed allowing for a non-null slope. The result of the additional linear fit is shown in Fig. 5.9 (bottom) as a red line, and shows that, after the bias correction, there is no residual dependence on the generated top mass.

The final bias is expected to be small ($\approx$0.4 GeV/c^2). The residual bias is 0.04$\pm$0.4 (0.1$\pm$0.4) for the inclusive (exclusive) fit. The envelope of the residual biases is conservatively assigned as an intrinsic KINb systematic uncertainty.

Results

The KINb method is applied to data and the results are summarized below. Figure 5.10 shows the results of the KINb fit to the 4 exclusive categories considered. The result of the fit in the different event categories agrees within the statistical uncertainties. It must also be noticed that the background yields obtained from the fit are in agreement with the prediction of Tab. 5.3. For the specific choice of the bin size in the figure a χ^2-like value can be computed and compared with the number of degrees of freedom. For an histogram with 20 bins and 1 estimated parameter, $NDOF = 19$. This approximate $\frac{\chi^2}{NDOF}$ values can be used to gauge the distance between the parametrization used in the fit and the data, where values of the order of unity are indication of a healthy agreement. The first bin of each distribution does not really satisfy the $N_{evts/bin} \geq 5$ requirement suggested in Ref. [139], and a probabilistic interpretation of the χ^2 values is avoided here. The $\frac{\chi^2}{NDOF}$ values obtained are: 0.98 and 2.79 for the same flavor channel with 1 or at least two b-tags correspondingly; 1.98 and 4.32 for the opposite flavor channel with 1 or at least two b-tags correspondingly. The values of the mass and number of background events fit

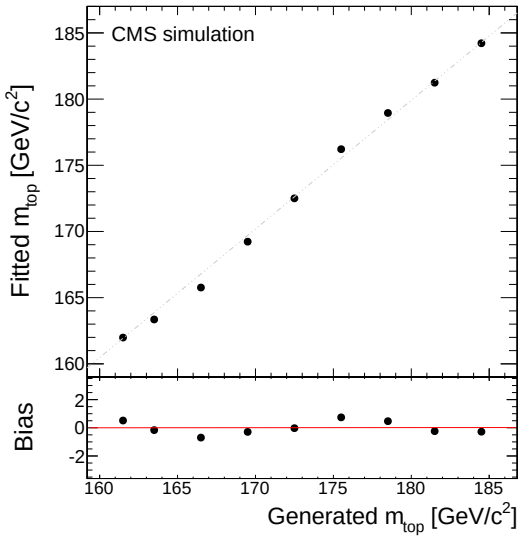


Figure 5.9: (Top) Fitted top quark mass values using the KINb algorithm from simulated pseudoexperiments, including signal and background processes, as a function of the actual top quark mass used in the simulation. A linear fit to the points is also shown. (Bottom) The difference (bias) between the linear fit and the actual reconstructed values from the pseudo-experiments. The bias is shown after calibrating the signal parametrization.

in each category are shown in the captions. The individual and the combined likelihoods are shown in Fig. 5.11. Table 5.4 summarizes the results of the fits to data with the KINb (exclusive fit) method. An additional method, based on the invariant mass of the lepton-bjets pairs and labeled "M_{lb} method" has been used as a crosscheck, yielding consistent results.

Table 5.4: Summary of the top quark mass fits. Uncertainties are statistical only. The statistical uncertainty for the Top quark masses has been corrected for the scale factor quoted in the text. The data column corresponds to the number of events used in the fit.

Method	Data	N_{signal}	$N_{background}$	m_{top} (GeV/c^2)
KINb	4312	3780.8 $\pm$ 70.4	518.3 $\pm$ 25.0	173.30 $\pm$ 1.23

In the next Section the systematic uncertainties are discussed, affecting our measurements of the top quark mass.

Systematic uncertainties

The sources of systematic uncertainty are summarized in Table 5.5. The dominant source of uncertainty on $m_{\rm top}$ is the overall jet energy scale (JES) which may introduce a global shift in the mass measured from data. A (p_T, η)-dependent variation of the JES is performed in order to estimate this uncertainty which is evaluated by scaling up and down the jet energy with respect to the nominal value [140]. The variation of the top quark mass fit obtained from each jet energy scaling is evaluated in MC pseudo-experiments. The average variation of the top quark mass is used to estimate the systematic uncertainty. Jet flavor-specific corrections, i.e. due to the fragmentation of b/c-quarks, are evaluated from samples based on the different response to quark and gluon jets. The uncertainty due to jet energy

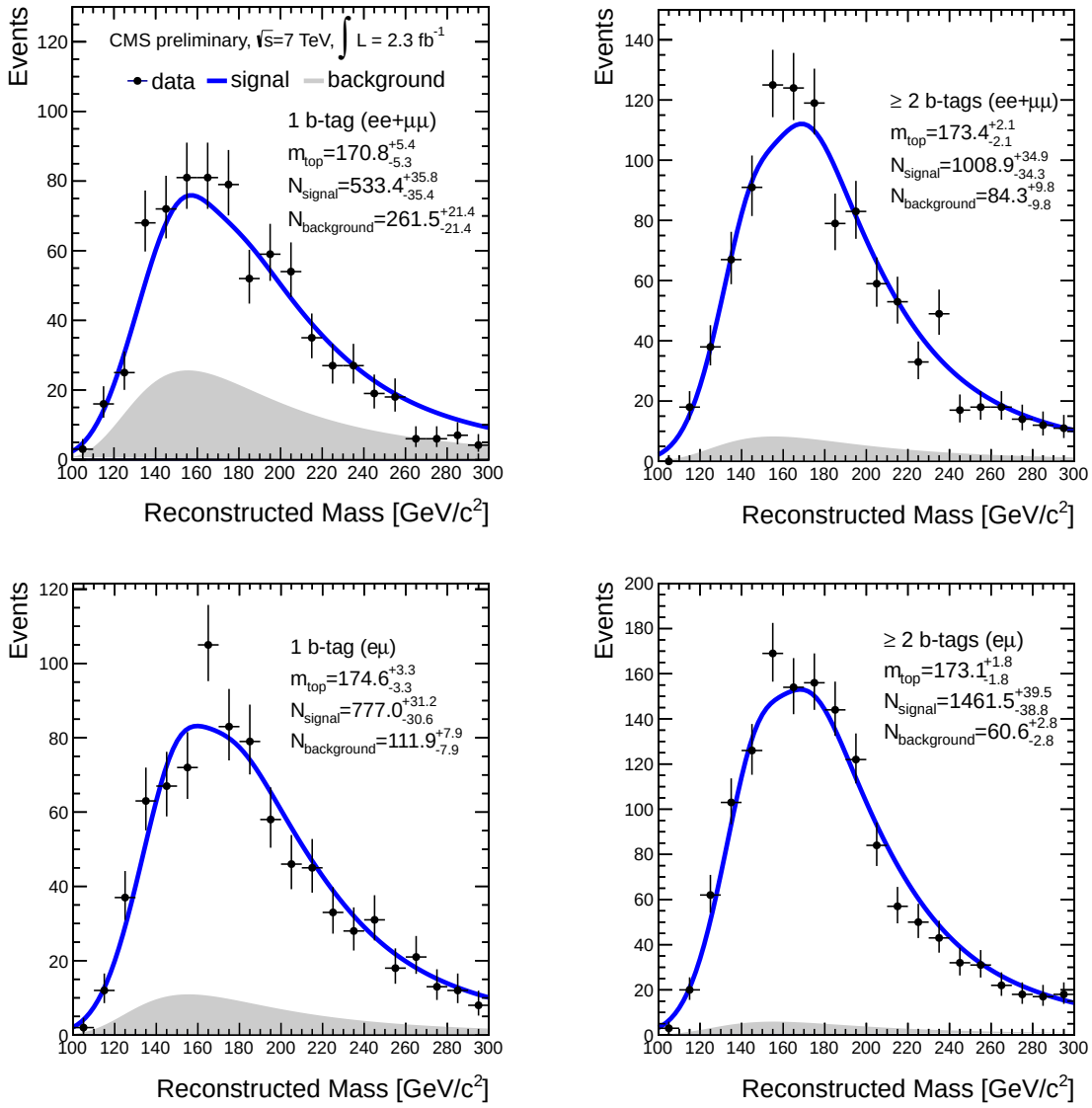


Figure 5.10: Results of the fit to the top quark mass with the KINb method in each event category considered.

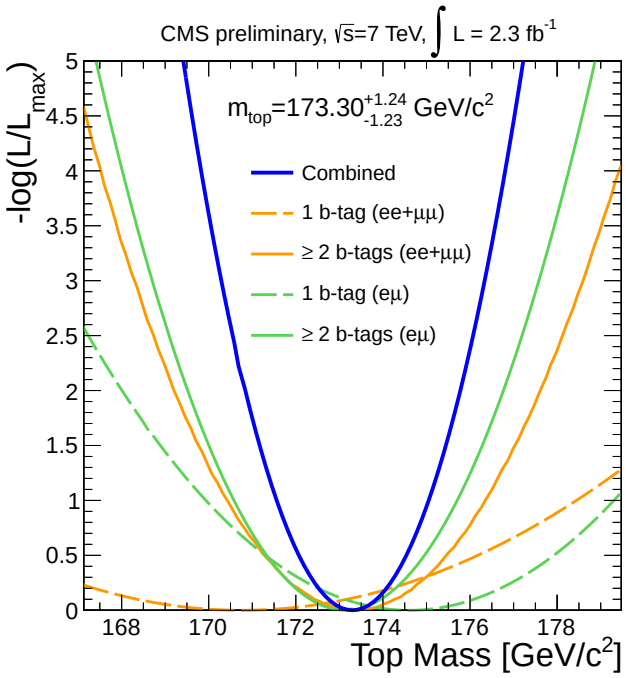


Figure 5.11: Variation of the likelihoods used to fit to the top quark mass with the KINb method in each event category considered. The combination of the likelihoods is also superimposed.

resolution (JER) is evaluated from pseudo-experiments by smearing the jet energies used in the $m_{\rm top}$ reconstruction up and down with respect to the default values. Lepton energy scale (LES) is observed to have almost negligible effect on the measurement of $m_{\rm top}$. The uncertainty in the $E_{\rm T}^{\rm miss}$ scale is propagated to the measurement of $m_{\rm top}$ after factorizing the clustered (i.e. jet energy) and leptonic components which are varied separately as already described above. The scale of the residual unclustered $E_{\rm T}^{\rm miss}$ is varied by $\pm$10% and the corresponding variation of the top quark mass fit is evaluated from MC pseudo-experiments. The uncertainty of the fit calibration is taken as the envelope of the residual bias of the fit. The systematic uncertainty due to the actual normalization of DY events is evaluated after varying the expected number of DY events by the uncertainty of the scale factors obtained from data. The variation of the top quark mass fit is assigned as a systematic uncertainty. The Q^2 factorization scale for top quark production is scaled up and down as well as the kinematic scale which is used to match clustered jets to partons (i.e. jet-parton matching scale) using with dedicated samples. The average number of pile-up interactions measured in data to be $\approx$6 is varied in the simulation by $\pm$ 0.6 interactions. Pseudo-experiments using the increased/decreased pile-up contamination of the sample are then used to fit $m_{\rm top}$ and to obtain the shift with respect to the nominal pile-up scenario used in the calibration. Uncertainties of b-tag/mistag rates may contribute to the systematic uncertainty in the measurement of $m_{\rm top}$ through a migration of events from one b-tagging multiplicity to the other, and are evaluated by varying the b-tagging and mistagging efficiencies independently according to the uncertainties in Ref. [135]. Additional uncertainties come from the modeling of the signal templates: MC generator and parton distribution functions (PDFs). For the MC generator the reference samples are compared to samples generated with the POWHEG [141] generator. The uncertainty in the modeling of

the PDF is studied from pseudo-experiments where the signal contribution is sampled after reweighing with the different sources of orthogonal variations (a total of 22) of the base CTEQ 6.6 PDF set [142]. For each source the difference found with respect to the nominal prediction is added in quadrature to the total PDF uncertainty. Non-pertubative QCD effects are taken into account by the Z2 parameter set used to model the underlying event in the $\mathrm{t\bar{t}}$ signal samples. The uncertainty associated to different modeling of the underlying event and the color reconnection [143] are still being studied and are therefore not reported in this manuscript. However the impact on the final result is expected to be small in the dilepton channel [144].

Table 5.5: Summary of the systematic uncertainties (in GeV/c^2) in the measurement of m_{top}.

Source	Δm_{top} (GeV/c^2)
JES	$^{+1.90}_{-2.00}$
flavor-JES	$^{+1.08}_{-1.13}$
JER	$\pm$ 0.30
LES	$^{+0.12}_{-0.18}$
Unclustered $E_{\mathrm{T}}^{\mathrm{miss}}$	$\pm$ 0.43
Fit calibration	$\pm$ 0.40
DY normalization	$\pm$ 0.40
Factorization scale	$\pm$ 0.41
Jet parton matching scale	$\pm$ 0.65
Pile-up	$\pm$ 0.19
b-tagging uncertainty	$\pm$ 0.30
mis-tagging uncertainty	$\pm$ 0.43
MC generator	$\pm$ 0.14
PDF uncertainty	$\pm$ 0.39
Total	$^{+2.52}_{-2.63}$

5.1.4 Summary

A sample of dilepton events has been selected from the 2011 data enriched in top quark pair events. The events are used to extract the top quark mass using the *KINb* method, based on a full kinematic reconstruction. The method yields a result in agreement with the Tevatron experiments and with the first results of the top quark mass measurement at the LHC as summarized in Fig. 5.12. The Tevatron measurements are taken from [145].

The result was the most precise measurement in the dilepton channel at the time in which it was made public by the CMS Collaboration.

Later, the measurement was completed by adding the full 7 TeV statistics available, amounting to about $5\,\mathrm{fb}^{-1}$, and compared with an analysis that used *MWT* method mentioned in the introduction. The method, based on a matrix element method for weighting each event in the hypothesis that it is an event featuring a top quark pair decaying into dilepton, yielded a compatible result, but with a better overall uncertainty. The mass resolution degradation for the *KINb* measurement is due to the fact that, a choice is made for the assignment of the lepton-jet pairs in the event, instead of reweighting the solutions based on expectations for the kinematic properties. Furthermore, only events with solutions to the full system

equation are used, as described in Sec. 5.1.3. The *MWT* method, instead, not only starts from a looser event selection (whose relatively higher background contamination is more than balanced by the higher expected precision of the method), but also makes full use of all the solutions found, reweigthing them according to a p.d.f. based on an analitical solution to the kinematics of the events: this result in a smaller statistical uncertainty on the final result. The idea of combining the two results had been considered, but nothing was to be gained from it, since the results are well compatible with each other, and most of the systematic uncertainties are fully correlated across the two measurement. In the end, the *KINb* was published as a cross-check measurement to the *AMWT* measurement [43].

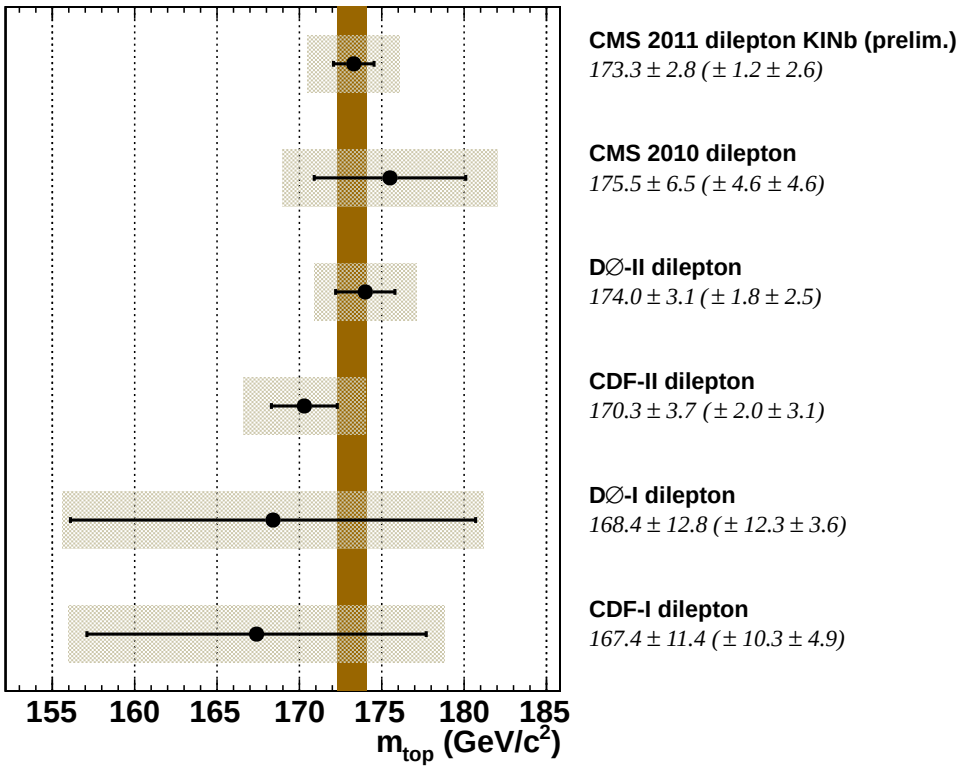


Figure 5.12: Top mass measurements in the dilepton channel.

5.2 Top-antitop pair production cross section measurement

5.2.1 Introduction

The top-quark pair production cross section is measured in final states with one electron or muon and one hadronically decaying τ lepton from the process $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\tau\nu_\tau)\mathrm{b\bar{b}}$, where $\ell = \mathrm{e}, \mu$. The data sample corresponds to an integrated luminosity of 19.6 fb^{-1} collected with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV.

In the current section of this manuscript, we measure the production cross section of top-quark pairs by considering dilepton decays where one W boson promptly decays into $\ell\nu_\ell$, with $\ell = \mathrm{e}$ or μ,

and the other decays into $\tau\nu_\tau$, $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\tau\nu_\tau)\mathrm{b\bar{b}}$. The expected fraction of these events is 4/81 of all $\mathrm{t\bar{t}}$ decays. The τ lepton is identified by means of its hadronic decay products, with a branching fraction $\mathcal{B}(\tau \to \text{hadrons} + \nu_\tau) \simeq 65\%$, to produce a narrow jet with a small number of charged hadrons, denoted as τ_h. The cross section is measured by counting the number of $\ell\tau_\mathrm{h} + X$ events consistent with originating from $\mathrm{t\bar{t}}$ production, after subtracting the contributions from other processes, and correcting for the efficiency of the event selection. A similar method was used in pp collisions at a centre-of-mass energy of $\sqrt{s} = 7$ TeV [146]. This "τ dilepton" channel is of particular interest because it is a natural background process to the search for a charged Higgs boson [147, 148] with a mass smaller than that of the top quark. In this case, the production chain $\mathrm{t\bar{t}} \to \mathrm{H^+bW^-\bar{b}}$, with $\mathrm{H^+} \to \tau^+\nu_\tau$ (or the corresponding charge-conjugate particles) could give rise to differences with respect to the standard model (SM) prediction of the number of $\mathrm{t\bar{t}}$ events with a τ lepton [149]. The present measurement is based on data collected by the CMS experiment in pp collisions at $\sqrt{s} = 8$ TeV corresponding to an integrated luminosity of 19.6 fb^{-1}. The relative accuracy of this measurement improves over previous results [45, 46, 47, 48, 49], thanks to the inclusion of additional data and improved analysis techniques.

Some details on the simulated samples is given in Section 5.2.2, complementing the information present in Section 3.5. A brief description of the event reconstruction and event selection is given in Section 5.2.3. The descriptions of the background determination and the systematic uncertainties are given in Sections 5.2.4 and 6.4, respectively. The measurement of the cross section is discussed in Section 5.2.6, and the results are summarised in Section 5.2.7.

5.2.2 Data and simulation samples

Events are selected online by a trigger requiring a single isolated electron (muon) with transverse momentum $p_\mathrm{T} > 27\,(24)$ GeV and $|\eta| < 2.5\,(2.1)$.

This measurement makes use of simulated samples of $\mathrm{t\bar{t}}$ events as well as other processes that mimic the $\ell\tau_\mathrm{h}$ decay signature. These samples are used to optimise the event selection, to calculate the acceptance for $\mathrm{t\bar{t}}$ events, and to estimate some of the backgrounds in the analysis.

The signal acceptance and $\mathrm{t\bar{t}}$ dilepton background are evaluated using a version of MADGRAPH which includes the effects of spin correlations [93, 150]. The number of expected $\mathrm{t\bar{t}}$ events is estimated with the next-to-next-to-leading-order (NNLO) SM cross section of $251.7^{+6.3}_{-8.6}\,(\text{scale}) \pm 6.5\,(\mathrm{PDF})$ pb [151, 152, 153, 154, 155] for a top-quark mass of 172.5 GeV, where the first uncertainty is due to renormalisation and factorisation scales, and the second is due to the choice of parton distribution functions (PDFs).

5.2.3 Event selection

Events are reconstructed with the particle-flow (PF) algorithm [80, 156], which combines information from all sub-detectors to identify and reconstruct individual electrons, muons, photons, charged and neutral hadrons. The primary collision vertex is chosen as the reconstructed vertex with the largest $\sum p_\mathrm{T}^2$ of the associated tracks. Electrons are identified with a multivariate discriminant combining several quan-

tities describing the track quality, the shape of the energy deposits in the electromagnetic calorimeter, and the compatibility of the measurements from the tracker and the electromagnetic calorimeter [157], and are reconstructed with an average efficiency of approximately 95%. Muons are identified with additional requirements on the quality of the track reconstruction and on the number of measurements in the tracker and the muon systems [158], and are reconstructed with an average efficiency of approximately 96%. Charged and neutral particles provide the input to the anti-k_{T} jet clustering algorithm with a distance parameter of 0.5 [159]. The jet momentum is determined from the vector sum of particle momenta in the jet. After jet energies are corrected for additional pileup contributions and for detector effects, they are found in simulations to be within 5–10% of the actual jet momentum [160]. The missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as the magnitude of the vector sum of momenta from all reconstructed particles in the plane transverse to the beam.

In addition, higher-level observables such as b-tagging discriminators and lepton isolation variables are used. The lepton relative isolation is defined as the transverse energy contributions deposited by charged hadrons ($E_{\mathrm{T,\,ch}}$), neutral hadrons ($E_{\mathrm{T,\,nh}}$), and photons ($E_{\mathrm{T,\,ph}}$) in a cone of radius $R = \sqrt{(\Delta\varphi)^2 + (\Delta\eta)^2} = 0.4$ centered on the lepton candidate track, relative to the lepton's transverse momentum (p_{T}), $I_{\mathrm{rel}} = (E_{\mathrm{T,\,ch}} + E_{\mathrm{T,\,nh}} + E_{\mathrm{T,\,ph}})/p_{\mathrm{T}}$. An electron (muon) candidate is considered to be non-isolated and is rejected if $I_{\mathrm{rel}} > 0.1$ (>0.12).

The hadronic products of the τ-lepton decay are reconstructed using a jet as the initial seed, and are then classified as having one or three charged hadrons with the "hadron-plus-strips" algorithm [161, 162]. In the "hadron-plus-strips" algorithm, calorimeter energy deposits clustered along strips in the φ direction are used for neutral pion identification. Then, the decay modes, four-momenta, and isolation quantities of the τ_{h} are determined, and the following categories are considered: single hadron, hadron plus a strip, hadron plus two strips, and three hadrons. These categories together encompass approximately 95% of hadronic τ-lepton decays. The sum of the charged hadron charges provides the τ_{h} charge. The τ_{h}-jet momentum is required to match the direction of the original jet within a maximum distance $R = 0.1$. Isolation criteria require that there be no additional charged hadrons with $p_{\mathrm{T}} > 1.0\,\mathrm{GeV}$ or photons with transverse energy $E_{\mathrm{T}} > 1.5\,\mathrm{GeV}$ within a cone of size $R = 0.5$ around the direction of the τ_{h} jet. Electrons and muons misidentified as τ_{h} are suppressed using algorithms that combine information from the tracker, calorimeters, and muon detectors [163]. The τ_{h} identification efficiency is defined as the ratio of the number of selected τ_{h} candidates divided by the number of hadronic τ-lepton decays in $\mathrm{t\bar{t}}$ events; the ratio depends on p_{T} and η of the τ_{h}, and is on average 50% for $p_{\mathrm{T}}^{\tau_{\mathrm{h}}} > 20\,\mathrm{GeV}$, with a probability of approximately 1% for generic jets to be misidentified as a τ_{h} jet.

The combined secondary vertex (CSV) algorithm [164] is used to identify jets originating from the hadronisation of b quarks. The algorithm combines the information about track impact parameters and secondary vertices within jets into a likelihood discriminant to provide separation between b jets and jets originating from light quarks, gluons, or charm quarks. The output of this CSV discriminant has values between zero and one; a jet with a CSV value above a certain threshold is referred to as being "b tagged". We choose a working point where the b-tagging efficiency is approximately 60%, as measured in a data sample of events enriched with jets from semileptonic b-hadron decays. The misidentification

rate of light-flavour jets is estimated from inclusive jet studies and is measured to be about 0.1% for jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$.

Events are preselected by requiring exactly one isolated electron (muon) with transverse momentum $p_{\mathrm{T}} > 35\,(30)\,\mathrm{GeV}$ and $|\eta| < 2.5\,(2.1)$, at least two jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, and one additional jet with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. The selected jets must be within $|\eta| < 2.4$. The electron or muon is required to be separated from any jet in the (η, φ) plane by a distance $R > 0.4$. Events with any additional loosely isolated, $I_{\mathrm{rel}} < 0.2$, electron (muon) of $p_{\mathrm{T}} > 15\,(10)\,\mathrm{GeV}$ are rejected. Further event selection requirements include $E_{\mathrm{T}}^{\mathrm{miss}} > 40\,\mathrm{GeV}$ and only one τ_{h} with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$. The τ_{h} and the lepton are required to have electric charges of opposite sign (OS). At least one of the jets is required to be identified as originating from b-quark hadronisation (b-tagged).

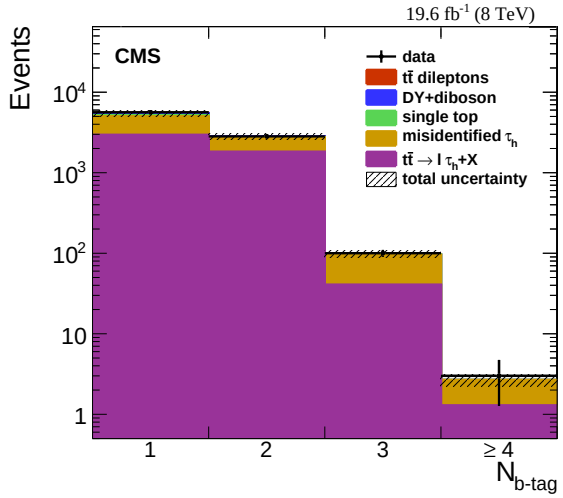


Figure 5.13: The b-tagged jet multiplicity after the full event selection. The simulated contributions are normalised to the SM predicted values. The hatched area shows the total uncertainty.

Figure 5.13 shows, for the sum of the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states, a comparison between data and simulation of the number of b-tagged jets in each event $N_{\text{b-tag}}$ after all the selection criteria have been applied. The distributions of the τ_{h} p_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ after the final event selection are shown in the left and right panels of Fig. 5.14, respectively. The distributions show agreement between the observed numbers of events and the expected numbers of signal and background events obtained from the simulated distributions normalised to the integrated luminosity of the selected data sample.

Following the final selection, additional kinematic features of the $\mathrm{t\bar{t}}$ events are studied to evaluate the agreement between the observed data and the predicted sum of signal and background. For each event, two invariant mass combinations are reconstructed by pairing the τ_{h} with the two candidate b-jets: (1) in events with two or more b-tagged jets, the two combinations are based on the two b-tagged jets with the highest value of the discriminator; (2) in events with one b-tagged jet, this is used for the first combination, while the non-b-tagged jet with the highest p_{T} is used to form the second combination. For the two combinations, the invariant mass with the lowest value is shown in Fig. 5.15 (left), for the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels combined.

For each event, the top-quark mass m_{top} is reconstructed using the KINb algorithm [43, 165]. Due to

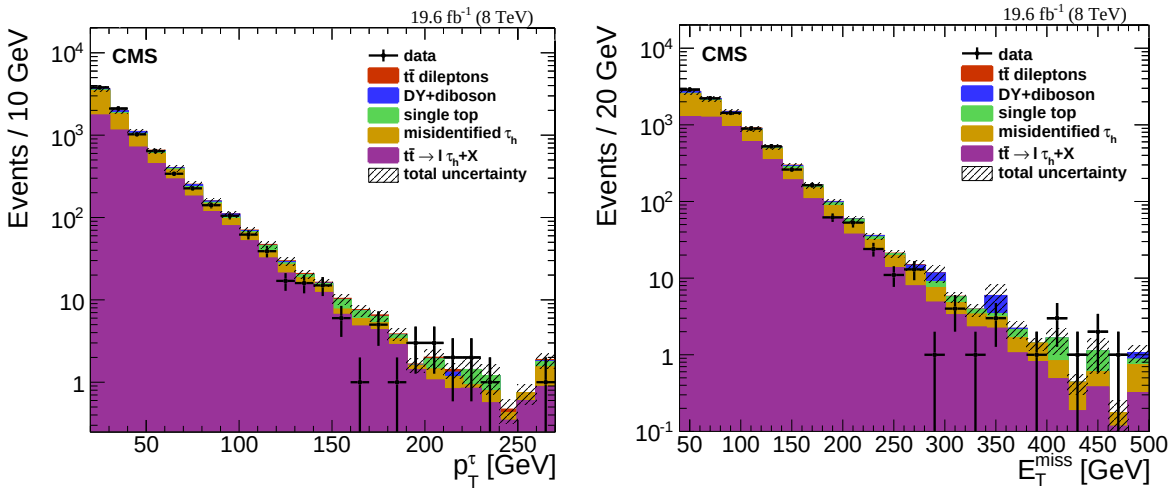


Figure 5.14: Distribution of the τ_{h} p_{T} (left) and $E_{\mathrm{T}}^{\mathrm{miss}}$ (right) after the full event selection, for the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels combined. The simulated contributions are normalised to the SM predicted values. The hatched area shows the total uncertainty. The last bins include the overflow events.

the multiple neutrinos in the event, the reconstruction of m_{top} leads to an underconstrained system. The KINb algorithm applies constraints on the W boson mass, the mass difference between the top and anti-top quark, and the longitudinal momentum of the $\mathrm{t\bar{t}}$ system. For each event, solutions to the kinematic equations are evaluated, varying the jet momenta and the direction of $E_{\mathrm{T}}^{\mathrm{miss}}$ within their resolutions. For each set of variations and each lepton-jet combination, the kinematic equations allow up to four solutions; the one with the lowest $\mathrm{t\bar{t}}$ invariant mass is accepted if the mass difference between the two top quarks is less than 3 GeV. For each event, the accepted solutions corresponding to the two possible lepton-jet combinations are counted and the combination with the largest number of solutions is chosen and m_{top} is obtained by fitting the peak of this distribution. The events in which solutions are found are shown in Fig. 5.15 (right). Data are in agreement with the expected sum of signal and background events.

5.2.4 Background estimate

The main background (misidentified τ_{h}) comes from events with one lepton (electron or muon), significant $E_{\mathrm{T}}^{\mathrm{miss}}$, and three or more jets, where one jet is misidentified as a τ_{h} jet [149]. The dominant source is $\mathrm{t\bar{t}}$ lepton+jet events. The misidentified τ_{h} background accounts also for events with W bosons produced in association with jets, either genuine W+jet or single-top-quark production, and for QCD multijet events. In order to estimate this background from data, the misidentification probability $w(\text{jet} \to \tau_{\mathrm{h}})$ is parameterised as a function of the jet p_{T}, η, and width (R_{jet}). The quantity R_{jet} is defined as $\sqrt{\sigma_\eta^2 + \sigma_\varphi^2}$, where σ_η (σ_φ) expresses the extent in η (φ) of the jet cluster [160].

The probability $w(\text{jet} \to \tau_{\mathrm{h}})$ is evaluated from two control samples:

- $w_{W+\text{jets}}$: from a W+jet event sample, selected by requiring one isolated muon with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.1$, and at least one jet with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$;
- w_{QCD}: from a QCD multijet sample, triggered by one jet with $p_{\mathrm{T}} > 40$ GeV, selected by requiring

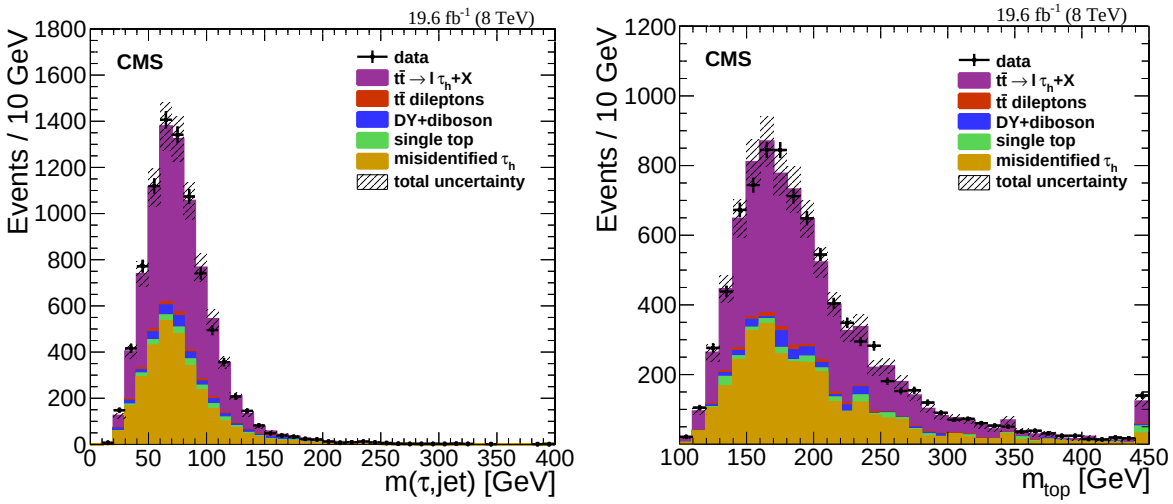


Figure 5.15: (left) Minimum invariant mass reconstructed by pairing the τ_{h} with either a b-tagged jet or with the highest p_{T} non b-tagged jet, as described in the text. (right) Distribution of the reconstructed top-quark mass m_{top} for the $\ell\tau_{\mathrm{h}}$ candidate events after the full event selection. Data (points) are compared with the sum of signal and background yields, for the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels combined. The simulated contributions are normalised to the SM predicted values. The hatched area shows the total uncertainty. The last bins include the overflow events.

events to have at least two jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$, where the triggering jet is removed from the misidentification rate calculation to avoid a trigger bias.

The full description of the algorithm, that has been developed for the charged Higgs searches in the lepton+tau final state, can be found in Sec. 6.3.1.

Both probabilities are evaluated in simulated events as well as in data, with good agreement found between the results from simulation and data [161].

The number of events containing misidentified τ_{h} candidates is then determined as

$$N^{\text{misid}} = \sum_{i}^{M} \sum_{j}^{m} w_i^j(\text{jet} \to \tau) - N^{\text{other}}, \tag{5.8}$$

where j is the jet index of event i, and m is the number of jets in each event and M is the total number of events. The quantity N^{other} is the expected $\simeq$20% contamination from signal and other processes to the misidentified background as estimated from simulated samples. The value of N^{other} is evaluated by applying the procedure described above to simulated events of $\mathrm{Z}/\gamma^* \to \tau\tau$, single-top-quark production, diboson production, and the $\mathrm{t\bar{t}}$ processes included in the misidentified τ_{h} background estimation.

Jets in QCD multijet events originate mainly from gluons, while in W+jet events they are predominantly from quarks. The quark and gluon composition in the misidentified τ_{h} events lies between these two control samples. As $w_{\mathrm{QCD}} < w_{\mathrm{W+jets}}$, the actual N^{misid} value is under- (over-) estimated by applying the w_{QCD} ($w_{\mathrm{W+jets}}$) probability. We determine from data the rate for the misidentification of a jet to be identified as a τ_{h}, and from simulation the quark/gluon composition in the W+jet and multijet samples. From these quantities we derive the following combination:

$$\langle N^{\text{misid}} \rangle = SF_{\mathrm{W+jet}} \times N^{\text{misid}}_{\mathrm{W+jet}} + SF_{\mathrm{QCD}} \times N^{\text{misid}}_{\mathrm{QCD}}, \tag{5.9}$$

where the misidentification rates, extracted from the data control samples discussed above, are combined with the scale factors SFs determined from the set of equations describing the quark/gluon composition of the samples: $SF_{\mathrm{QCD}} = 0.83$ and $SF_{\mathrm{W+jet}} = 0.17$. The corresponding systematic uncertainty is obtained from Eq. (5.9) by weighting the relative deviations of $N^{\mathrm{misid}}_{\mathrm{W+jet}}$ and $N^{\mathrm{misid}}_{\mathrm{QCD}}$ from $\langle N^{\mathrm{misid}} \rangle$ with the related scale factors. This results in an uncertainty of 7% for both $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels.

The efficiency of the OS requirement $\varepsilon_{\mathrm{OS}}$ is determined from simulated lepton+jet $\mathrm{t\bar{t}}$ events and is applied in order to obtain the misidentified τ_{h} background after the final event selection $N^{\mathrm{misid}}_{\mathrm{OS}}$, where $N^{\mathrm{misid}}_{\mathrm{OS}} = \varepsilon_{\mathrm{OS}} \cdot N^{\mathrm{misid}}$. We find values of $\varepsilon_{\mathrm{OS}} = 0.729 \pm 0.002\,\text{(stat)} \pm 0.004\,\text{(syst)}$ for the $\mathrm{e}\tau_{\mathrm{h}}$ selection and $\varepsilon_{\mathrm{OS}} = 0.731 \pm 0.002\,\text{(stat)} \pm 0.003\,\text{(syst)}$ for the $\mu\tau_{\mathrm{h}}$ selection, where all sources of systematic uncertainty are accounted for in the modelling of the simulated $\mathrm{t\bar{t}}$ lepton+jet events.

5.2.5 Systematic uncertainties

Several sources of systematic uncertainty are considered and listed in Table 5.6. They are related both to the signal reconstruction efficiency, background determination, and luminosity measurement (Experimental uncertainties) and to the theoretical assumptions on the $\mathrm{t\bar{t}}$ production (Theoretical uncertainties). In Table 5.6 and in what follows, relative values refer to the cross section uncertainty unless explicitly stated otherwise.

Table 5.6: List of systematic uncertainties in the cross section measurement, and their combination. Lepton reconstruction uncertainties are uncorrelated, while all other uncertainties are assumed 100% correlated.

Source	Uncertainty [%]		
	$\mathrm{e}\tau_{\mathrm{h}}$	$\mu\tau_{\mathrm{h}}$	Combined
Experimental uncertainties:			
τ_{h} jet identification	6.0	6.0	6.0
τ_{h} misidentification background	4.3	4.3	4.3
τ_{h} energy scale	2.4	2.5	2.5
b-jet tagging, jet misidentification	1.6	1.6	1.6
jet energy scale, jet energy resolution, $E^{\mathrm{miss}}_{\mathrm{T}}$	1.9	1.9	1.9
lepton reconstruction	0.8	0.6	0.5
other backgrounds	0.6	0.7	0.7
luminosity	2.6	2.6	2.6
Theoretical uncertainties:			
matrix element-parton shower matching	1.7	1.3	1.5
factorisation/renormalisation scale	2.9	2.9	2.9
generator	1.5	1.5	1.5
hadronisation	1.7	1.7	1.7
top-quark p_{T} modelling	0.7	0.5	0.6
parton distribution functions	0.8	0.7	0.7
total systematic uncertainty	9.6	9.5	9.5

Experimental uncertainties

Regarding the τ_{h} reconstruction, the uncertainty associated to the identification efficiency amounts to 6%, while the contribution relative to the τ_{h} jet energy scale is 2.4% (2.5%) for the $\mathrm{e}\tau_{\mathrm{h}}$ ($\mu\tau_{\mathrm{h}}$) channel, as estimated by varying the p_{T} of the τ_{h} jet by 3% [161, 162]. The uncertainty in the τ_{h} identification efficiency includes the uncertainty in charge determination which is estimated to be smaller than 1%. The uncertainty related to the misidentified τ_{h} background process, discussed in Section 5.2.4, is obtained by propagating the 7% uncertainty on $\langle N^{\text{misid}} \rangle$ to the cross section determination and results in 4.3% for both channels. It also includes the uncertainty in the OS efficiency determination.

The reconstruction of a light flavour jet as a b quark is defined as mistagging. The uncertainty due to b (mis)tagging is estimated to reflect the data-to-simulation scale factors and corresponding uncertainties for b-tagging and mistagging efficiencies [164]. When propagated to the cross section measurement, they amount to 1.6% for both $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels.

The jet energy scale (JES) uncertainty is estimated [160] by varying the jet energy within the p_{T}- and η-dependent JES uncertainties per jet, and taking into account the uncertainty due to pileup and parton flavour. The jet energy resolution (JER) is estimated by smearing the jet energy in simulation within the η-dependent JER uncertainties per jet. The JES and JER uncertainties are propagated in order to estimate the uncertainty of the $E_{\mathrm{T}}^{\text{miss}}$ scale. In addition, modelling of the $E_{\mathrm{T}}^{\text{miss}}$ component, which is not clustered in jets, is also considered. The resulting uncertainty from propagating these effects to the cross section measurement is 1.9% for both the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels.

Uncertainties due to trigger, lepton identification, isolation, and lepton energy scale are calculated from independent samples with a "tag-and-probe" method [157, 158], and yield 0.8% (0.6%) for the $\mathrm{e}\tau_{\mathrm{h}}$ ($\mu\tau_{\mathrm{h}}$) channel.

An overall 0.6% (0.7%) uncertainty for the $\mathrm{e}\tau_{\mathrm{h}}$ ($\mu\tau_{\mathrm{h}}$) channel is due to other minor backgrounds, accounting for the uncertainties related to the theoretical cross sections, JES, and b-tagging in these simulated samples, and the $\ell \to \tau_{\mathrm{h}}$ ($\ell = \mathrm{e}, \mu$) misidentification in the $\mathrm{Z}/\gamma^* \to \ell^+\ell^-$ and $\mathrm{t\bar{t}}$ dilepton processes.

Finally, the integrated luminosity is known with 2.6% accuracy [78].

Theoretical uncertainties

The theoretical uncertainty due to the matrix element (ME) and parton shower (PS) matching is estimated by varying up and down by a factor of two the threshold between jet production at the ME level and via PS, and it results in 1.7% (1.3%) for the $\mathrm{e}\tau_{\mathrm{h}}$ ($\mu\tau_{\mathrm{h}}$) channel.

The modelling uncertainty in the signal acceptance due to the factorisation and renormalisation scale choices is estimated by varying them simultaneously up and down by a factor of two from the nominal value equal to the Q^2 in the event, with an uncertainty of 2.9% found for both channels.

The uncertainty due to the choice of the generator is estimated as the relative difference between the acceptances evaluated with MADGRAPH and POWHEG [98, 141, 166, 167] after the full event selection and results in 1.5%. In a similar way, the uncertainty in the hadronisation scheme is evaluated from the

relative differences between the acceptances from POWHEG+PYTHIA and POWHEG+HERWIG samples, estimated prior to the b-tagging or τ_{h} jet requirement, resulting in a 1.7% uncertainty.

We consider the uncertainty related to the top-quark p_{T} scale modelling by varying the top-quark p_{T} spectrum and evaluating the change in the signal acceptance, resulting in 0.6%, and the uncertainty related to the PDF variations following the PDF4LHC prescriptions [153], resulting in 0.7%.

5.2.6 Cross section measurement

The number of expected signal and background events as well as the number of observed events after all selections are summarised in Table 5.7. The statistical and systematic uncertainties are also shown. The $\mathrm{t\bar{t}}$ production cross section measured from τ dilepton events is $\sigma_{\mathrm{t\bar{t}}} = (N - B)/(L \cdot A_{\mathrm{tot}})$, where N is

Table 5.7: Number of expected events for signal (assuming $m_{\mathrm{top}} = 172.5\,\mathrm{GeV}$) and backgrounds. The background from misidentified τ_{h} is estimated from data, while the other backgrounds are estimated from simulation. Statistical and systematic uncertainties are shown.

Source	$\mathrm{e}\tau_{\mathrm{h}}$	$\mu\tau_{\mathrm{h}}$
misidentified τ_{h}	$1341 \pm 3 \pm 94$	$1653 \pm 3 \pm 116$
$\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\ell\nu_\ell)\mathrm{b\bar{b}}$	$55 \pm 1 \pm 3$	$68 \pm 2 \pm 4$
$\mathrm{Z}/\gamma^* \to \mathrm{ee}, \mu\mu$	$11 \pm 5 \pm 5$	$12 \pm 5 \pm 5$
$\mathrm{Z}/\gamma^* \to \tau\tau$	$85 \pm 14 \pm 8$	$166 \pm 20 \pm 18$
single top quark	$104 \pm 7 \pm 9$	$133 \pm 8 \pm 10$
dibosons	$15 \pm 1 \pm 1$	$19 \pm 1 \pm 1$
total expected background	$1611 \pm 17 \pm 95$	$2051 \pm 22 \pm 118$
expected signal yield	$2134 \pm 9 \pm 170$	$2632 \pm 11 \pm 212$
data	3779	4767

the number of observed candidate events, B is the estimate of the background and L is the integrated luminosity. The total acceptance A_{tot} is the product of the branching fractions, geometrical and kinematic acceptance, trigger, lepton identification, and the overall reconstruction efficiency. It is evaluated with respect to the inclusive $\mathrm{t\bar{t}}$ sample. After the OS requirement and assuming a top-quark mass $m_{\mathrm{top}} = 172.5\,\mathrm{GeV}$, we obtain:

$$A_{\mathrm{tot}}(\mathrm{e}\tau_{\mathrm{h}}) = 0.04333 \pm 0.00017\,(\mathrm{stat}) \pm 0.00300\,(\mathrm{syst})\,\%;$$
$$A_{\mathrm{tot}}(\mu\tau_{\mathrm{h}}) = 0.05370 \pm 0.00021\,(\mathrm{stat}) \pm 0.00376\,(\mathrm{syst})\,\%.$$

The statistical uncertainties are due to the limited number of simulated events and the systematic uncertainties are estimated by accounting for all sources listed in Table 5.6. The statistical and systematic uncertainties listed in Table 5.7 are propagated to the final cross section measurements:

$$\sigma_{\mathrm{t\bar{t}}}(\mathrm{e}\tau_{\mathrm{h}}) = 255 \pm 4\,(\mathrm{stat}) \pm 24\,(\mathrm{syst}) \pm 7\,(\mathrm{lumi})\,\mathrm{pb};$$
$$\sigma_{\mathrm{t\bar{t}}}(\mu\tau_{\mathrm{h}}) = 258 \pm 4\,(\mathrm{stat}) \pm 24\,(\mathrm{syst}) \pm 7\,(\mathrm{lumi})\,\mathrm{pb}.$$

The BLUE method [168] is used to combine the cross section measurements in the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels, yielding weights of 0.47 and 0.53, respectively. Lepton reconstruction uncertainties are uncorre-

lated, while all other uncertainties are assumed 100% correlated. With this method we obtain a combined result of $\sigma_{\mathrm{t\bar{t}}} = 257 \pm 3\,(\mathrm{stat}) \pm 24\,(\mathrm{syst}) \pm 7\,(\mathrm{lumi})$ pb, in agreement with the NNLO expectation of $251.7\,^{+6.3}_{-8.6}\,(\mathrm{scales}) \pm 6.5\,(\mathrm{PDF})$ pb. Following the most recent conventions for the treatment of PDF and scale uncertainties the same calculation yields $252.9\,^{+6.4}_{-8.6}\,(\mathrm{scale}) \pm 11.7\,(\mathrm{PDF} + \alpha_{\mathrm{S}})$ pb [151, 152, 153, 154, 155]. The dependence on the top-quark mass has been studied for the range 160–185 GeV and is well described by a linear variation. If we adjust our result to the current world average value of 173.3 GeV [169], we obtain a cross section that is lower by 3.1 pb.

5.2.7 Summary

A measurement of the $\mathrm{t\bar{t}}$ production cross section in the channel $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\tau\nu_\tau)\mathrm{b\bar{b}}$ is presented, where ℓ is an electron or a muon, and the τ lepton is reconstructed through its hadronic decays. The data sample corresponds to an integrated luminosity of 19.6 fb^{-1} collected in proton-proton collisions at $\sqrt{s} = 8$ TeV. Events are selected by requiring the presence of one isolated electron or muon, two or more jets (at least one of which is b-tagged), significant missing transverse energy, and one τ. The largest background contribution is estimated from data and consists of $\mathrm{t\bar{t}}$ events with one W boson decaying into jets, where one jet is misidentified as a τ. The measured cross section is $\sigma_{\mathrm{t\bar{t}}} = 257 \pm 3\,(\mathrm{stat}) \pm 24\,(\mathrm{syst}) \pm 7\,(\mathrm{lumi})$ pb for a top-quark mass of 172.5 GeV. This measurement improves over previous results in this decay channel, and it is in good agreement with the standard model expectation and other measurements of the $\mathrm{t\bar{t}}$ cross section at same centre-of-mass energy.

Chapter 6

Search for a light charged Higgs boson

6.1 Introduction

A search for a light charged Higgs boson using proton-proton collisions at 7 TeV has been performed by the CMS collaboration [2] using the first part of the data collected by the CMS detector, corresponding to an integrated luminosity of about 2.2 fb^{-1}. In this chapter, an update is presented in which the $\mu\tau_{\mathrm{h}}$ final state has been updated to use the full 7 TeV statistics available, corresponding to an integrated luminosity of 4.9 fb^{-1}, and to exploit the different polarization status of the τ leptons coming from the charged Higgs decay with respect to the one of the τ leptons coming from the decay of the top quark. The resulting limits are combined with those in Ref. [2] for the other final states.

Results are presented in the search for a light charged Higgs boson that can be produced in the top quark decay $\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}$ with subsequent decay of H^{+} into $\tau^{+}\nu_{\tau}$. The search is sensitive to the decays of the top quark pairs $\mathrm{t\bar{t}} \to \mathrm{H}^{\pm}\mathrm{W}^{\mp}\mathrm{b\bar{b}}$ and $\mathrm{t\bar{t}} \to \mathrm{H}^{\pm}\mathrm{H}^{\mp}\mathrm{b\bar{b}}$. The results in the $\mu\tau + \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$+jet signature are updated with an integrated luminosity of 4.9 fb^{-1} recorded in proton-proton collisions at $\sqrt{s} = 7$ TeV by the CMS experiment at the LHC, and are combined with results of a previous search in other decay channels, based on an integrated luminosity of about 2 fb^{-1} .

The minimal supersymmetric extension of the standard model (MSSM) requires the introduction of two Higgs boson doublets in order that the superpotential can contain appropriate terms for giving masses to both up and down type quarks [24, 25, 26, 27, 28, 29, 30, 31]. This leads to the prediction of five elementary Higgs particles: two CP-even (h,H), one CP-odd (A), and two charged ($\mathrm{H}^{\pm}$) states [147, 148]. The lower limit on the charged Higgs boson mass is 78.6 GeV, as determined by LEP experiments [37, 38, 39, 40]. If the mass of the charged Higgs boson is smaller than the difference between the masses of the top and the bottom quarks, i.e. $m_{\mathrm{H}^{+}} < m_{\mathrm{t}} - m_{\mathrm{b}}$, the top quark can decay via $\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}$ Figure 6.1 shows the predicted branching fractions for a charged Higgs boson of mass $m_{\mathrm{H}^{+}} = 100$ GeV.

For values of $\tan\beta > 5$, the charged Higgs boson preferentially decays to a τ lepton and a neutrino, $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$, where $\tan\beta$ is defined as the ratio of the vacuum expectation values of the two Higgs boson

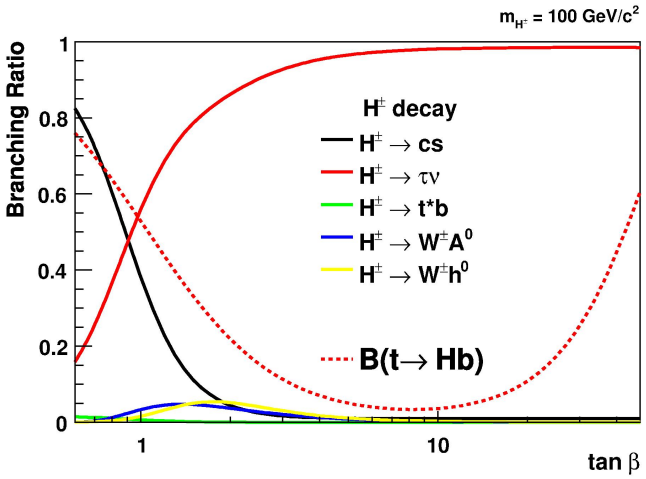


Figure 6.1: Predicted branching ratios for a charged Higgs boson of mass $m_{\mathrm{H}^+} = 100$ GeV. The figure is taken from [9]

doublets. In deriving the experimental limits we assume that the branching fraction $\mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1$.

The presence of the $\mathrm{t} \to \mathrm{H}^+\mathrm{b}$, $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay modes alters the τ lepton yield in the decay products of $\mathrm{t\bar{t}}$ pairs compared to the standard model (SM) expectations. The upper limit on the branching fraction, $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) < 0.2$, has been set by the CDF [170] and D0 [171] experiments at the Tevatron for m_{H^+} between 80 and 155 GeV, assuming $\mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1$. More recently, results from the ATLAS experiment at the LHC set upper limits on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b})$ between 5% and 1% for a charged Higgs boson in the mass range 90–160 GeV [172].

The dominant process of production of top quarks at the Large Hadron Collider (LHC) is $\mathrm{pp} \to \mathrm{t\bar{t}}$+X via gluon gluon fusion. The search for a charged Higgs boson is sensitive to decays of top quark pairs via $\mathrm{t\bar{t}} \to \mathrm{H}^\pm\mathrm{bH}^\mp\mathrm{\bar{b}}$ and $\mathrm{t\bar{t}} \to \mathrm{W}^\pm\mathrm{bH}^\mp\mathrm{\bar{b}}$ where each charged Higgs boson decays into a τ lepton and a neutrino. Throughout this manuscript, these two decay modes are referred to as WH and HH, respectively.

The final state studied includes a hadronically decaying τ lepton produced in association with a muon (labeled $\mu\tau_\mathrm{h}$). Figure 6.2 shows a representative diagram. We use the full data sample recorded by the Compact Muon Solenoid (CMS) experiment until the end of 2011, corresponding to an integrated luminosity of 4.9 fb^{-1}. This analysis has been updated with respect to Ref. [2] for the $\mu\tau_\mathrm{h}$ final state alone, and the resulting limits are combined with those in Ref. [2] for the other final states.

6.2 Reconstruction and simulation

Physics objects are reconstructed as described in Sec 3.4

The number of produced $\mathrm{t\bar{t}}$ events is estimated from the SM prediction of the $\mathrm{t\bar{t}}$ production cross section, 165^{+4}_{-9}(scale)$^{+7}_{-7}$(PDF) pb [129, 173, 174, 175], where the first uncertainty is due to renormalization and factorization scales, and the second is due to the PDF uncertainty. The theoretical prediction agrees with the cross section measured at the LHC [176, 177].

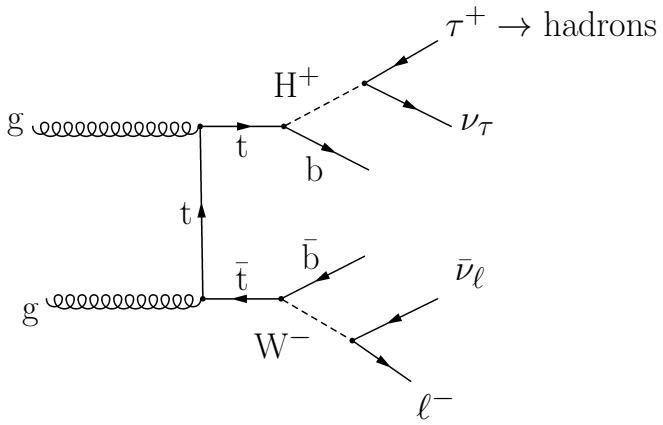


Figure 6.2: Representative diagram for the $\ell\tau_{\mathrm{h}}$ final state. The current analysis is performed for the muon channel, i.e. $\ell = \mu$.

6.3 Event selection and background determination

The event selection used is similar to that used in the measurement of the top quark pair production cross section in dilepton final states containing a τ described in Chapter 5.2.6.

A single-muon trigger is used to select the events. The threshold for the trigger changed from 17 to 24 GeV during the data taking period, due to the increased instantaneous luminosity. The amount of data analyzed corresponds to an integrated luminosity of $4.9 \pm 0.1\,\mathrm{fb}^{-1}$.

The events are selected by requiring one isolated muon with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.1$. The event must contain one τ_{h} with $p_T > 20$ GeV within $|\eta| < 2.4$, at least two jets with $p_{\mathrm{T}} > 30$ GeV within $|\eta| < 2.4$, with at least one jet identified as originating from the hadronization of a b quark, and $E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV. The τ_{h} and the muon are required to have opposite electric charges. The isolation of the muon candidate is measured by summing the transverse momenta of the reconstructed particles within a cone of radius $\Delta R = 0.3$ around the muon's direction. The contribution from the muon itself is excluded. If the value of this sum divided by the muon p_{T}, labeled I_{rel}, is less than 0.2, the muon is considered to be isolated. The muon is required to be separated from any selected jet by a distance $\Delta R >$ 0.3. Events with an additional electron (muon) with $I_{\mathrm{rel}} < 0.2$ and $p_T > 15\,(10)$ GeV are rejected.

The backgrounds arise from two sources, the first with misidentified τ_{h} from generic jets, which is estimated from data, and the second with genuine τ_{h} (or from other remaining sources), which is estimated from simulation. The background due to Drell-Yan or $\mathrm{t\bar{t}}$ events with one electron or muon misidentified as a τ_{h} is small and it is estimated from simulation, and it is also included in the second category.

The misidentified τ_{h} background comes from events with one muon (μ), $E_{\mathrm{T}}^{\mathrm{miss}}$, and three or more jets with at least one identified b quark jet (labelled "$\mu+ \geq 3$ jets" events), where one jet is misidentified as a τ_{h}. The dominant contribution to this background comes from W+jets, and from $\mathrm{t\bar{t}} \to \mathrm{W^+W^-b\bar{b}} \to \mu\nu\ q\bar{q}'\mathrm{b\bar{b}}$ events. The background contribution from misidentified τ_{h} is estimated by applying the probability that a jet mimics a τ_{h} to every jet in "$\mu+ \geq 3$ jets" events, as event weights. The probability that

a jet is misidentified as a τ_{h} is measured from data as a function of jet p_{T} and η using W+jets and QCD multijet events [2, 178]. The estimate makes use of the k-Nearest-Neighbours method described in Sec. 4.2

6.3.1 Estimating the tau fake background using the k-Nearest-Neighbour method

The $\ell+ \geq 3$ jet events are selected with the same selection as the signal selection except the τ selection. It is required that at least one of the jets with p_{T} >30 GeV is b-tagged (CSVM). The candidate jets in an event are selected as follows. If there are only two jets with p_{T} above 30 GeV, they are excluded; However, if there are more than two jets above 30 GeV, all of them are included. A similar approach is also used to include b-tagged jets: b-tagged jets are included only if there is more than one b-tagged jet in the event.

As a cross-check of the modeling of the W boson transverse momentum, for the same $\ell+ \geq 3$ jet events the W boson p_{T} is reconstructed as the sum of the transverse momenta of the muon and $E_{\mathrm{T}}^{\mathrm{miss}}$, yielding a good agreement between data and simulation.

The data driven method used makes use of the k-nearest-neighbour (kNN) multivariate algorithm for classifying jets faking taus [179], described in Sec. 4.2. The parametrization of τ fake rate is built as a function of jet p_{T}, jet η and jet radius $R_{\mathrm{jet}} = \sqrt{\sigma^2_{\eta\eta} + \sigma^2_{\phi\phi}}$. The jet radius parametrization improves the determination of the individual contributions of the quark jets and gluon jets (narrower and wider, respectively [180]), separately. The probability that a jet fakes a τ-jet ($w(jet \rightarrow \tau)$) is estimated as a function of three variables (jet p_{T}, jet η and R_{jet}) and it is applied to every jet in $W+ \geq 3$ jet events. Thus the expected number of τ-fake background is obtained as:

$$N^{\tau-\mathrm{fake}} = \sum_{i}^{N} \sum_{j}^{n} w_i^j(jet \rightarrow \tau) - N_{non-\tau-fake}, \tag{6.1}$$

where j is the jet index of the event i. The $N_{non-\tau-fake}$ is the small ($\simeq$ 20%) contamination of genuine tau contribution (i.e., non-τ-fake background) inside τ-fake background, which is estimated from MC. This is mostly due to the presence of real τ-jets in the $W+ \geq 3$ jet sample. In order to estimate this contribution, the same data driven method is applied to MC events of $Z/\gamma^* \rightarrow \tau\tau$, single top production, di-bosons, and the part of the SM $t\bar{t}$ background not included in the τ-fake background.

Examples of the parametrization of the $jet \rightarrow \tau$ fake rate probability can be found in Sec.7.4.

The parametrization as a function of R_{jet} is used in order to reduce the sensitivity to quark or gluon jets in the starting samples, either W+jets or QCD multijets. The goal is to account for the different probabilities for a quark (narrower) or gluon (wider) jet to fake a tau. Figure 6.3 shows the sensitivity of the R_{jet} variable in separating the quark and gluon jets components.

For the evaluation of $jet \rightarrow \tau$ fake probability, the QCD multijet and W+≥ 1 jet enriched samples are selected as in the following. The QCD multijet enriched events are selected by requiring events to pass the HLT_Jet30_v1 trigger and have at least two jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$. The jet matching to the HLT jet is removed from the fake rate calculation in order to avoid a trigger bias. This is done as in the following: if only one jet is matching to the HLT jet, then it is marked as ”tag” jet and others are

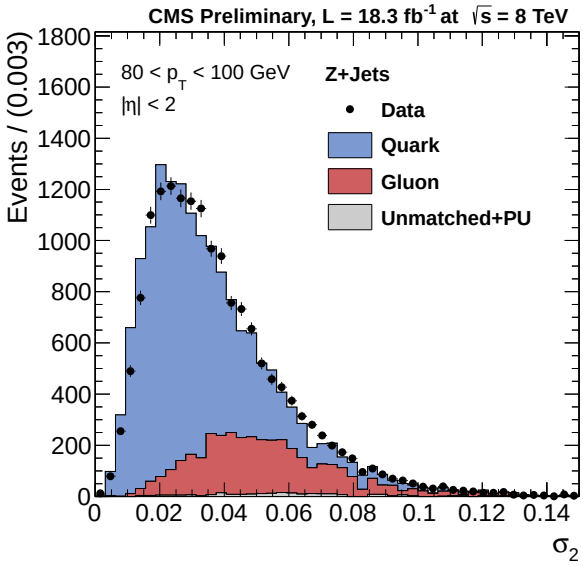


Figure 6.3: Distribution of R_{jet} for quark and gluon jets. The figure is taken from [10]

marked as ”probe” jets. However, all jets are marked as probe jets if more than one jet is matched to the trigger jet. Only the probe jets are used for the fake rate calculation. The $W+ \geq 1$ jet enriched events are selected by requiring only one isolated muon with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.1$ and at least one jet with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$. The muon isolation is described in Sec. 3.4. The separation between the muon and the closest jet is required to be $\Delta R_{\mu-jet} > 0.5$.

It has been shown in dedicated simulation studies (internal documentation of the CMS collaboration) that the quark and gluon jet compositions in $\ell + E_{\mathrm{T}}^{\mathrm{miss}} \geq 3$ jet events is different in QCD multi-jet and W+≥ 1 jet events, the number of τ fake events ($N^{\tau-\mathrm{fake}}$) is expected to be different in the two samples. In particular, the $N^{\tau-\mathrm{fake}}$ value will be under- (over-) estimated by applying the w_{QCD} (w_{W+jets}) probability. In order to properly account for the correct sample composition, the $N^{\tau-\mathrm{fake}}$ is estimated by re-weighing the number of events from each sample in order to take into account the quark and gluon jet composition as derived from MC. The systematic uncertainty is taken from the maximum variation between the two unweighted values:

$$N^{\tau-\mathrm{fake}} = W_{W+jets} \times \sum_{i}^{N} \sum_{j}^{n} w^{j}_{W+jets,\,i} \; + \; W_{QCD} \times \sum_{i}^{N} \sum_{j}^{n} w^{j}_{QCD,\,i} \tag{6.2}$$

$$\Delta N^{\tau-\mathrm{fake}} = \frac{\sum_{i}^{N} \sum_{j}^{n} w^{j}_{W+jets,\,i} \; - \; \sum_{i}^{N} \sum_{j}^{n} w^{j}_{QCD,\,i}}{2} \tag{6.3}$$

The difference between the “jet$\rightarrow \tau_{\mathrm{h}}$” probabilities measured in W+jets and multijet events is taken as systematic uncertainty, amounting to around $\sim 15\%$.

The backgrounds with genuine τ leptons are Drell–Yan $\tau\tau$, single-top-quark production, dibosons, and the SM $\mathrm{t\bar{t}}$ events in which a τ is produced from a W decay. The $\mathrm{Z}/\gamma^{*} \rightarrow \mathrm{ee}$, $\mu\mu$ and $\mathrm{t\bar{t}} \rightarrow \mathrm{W^{+}W^{-}b\bar{b}} \rightarrow \ell^{+}\nu\ell^{-}\overline{\nu}\mathrm{b\bar{b}}$ events may also contain electrons or muons misidentified as τ_{h}. The event yields for these backgrounds are estimated from simulation.

6.3.2 Yields

Data and the simulated event yield at various stages of the event selection are shown in Fig. 6.4. The backgrounds are normalized to the SM prediction obtained from the simulation. A good agreement is found between data and the SM background. The QCD multijet background contribution, evaluated from MC, is negligible at the final selection step. The expected event yield in the presence of $\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}$, $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decays is shown as a dashed line for $m_{\mathrm{H}^{+}} = 120\,\mathrm{GeV}$ under the assumption that $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}) = 0.05$.

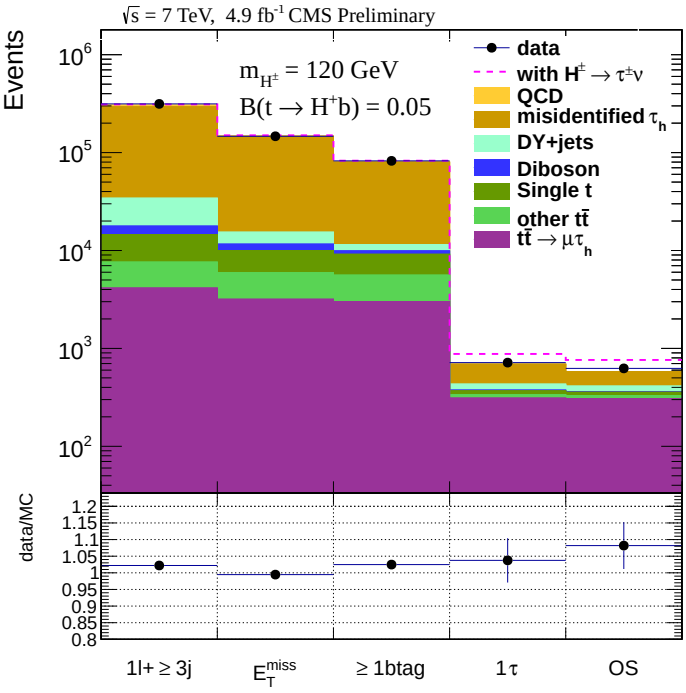


Figure 6.4: The event yields after each selection step, where OS indicates the requirement to have opposite electric charges for a τ_{h} and a μ. The backgrounds are estimated from simulation and normalized to the standard model prediction. The expected event yield in the presence of the $\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}$, $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decays is shown as a dashed line for $m_{\mathrm{H}^{+}} = 120\,\mathrm{GeV}$ and under the assumption that $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}) = 0.05$. The bottom panel shows the ratios of data over background with the total uncertainties. Statistical and systematic uncertainties are added in quadrature.

The observed number of events after the full event selection is shown in Table 7.3 along with the expected numbers of events from the various backgrounds, and from the charged Higgs boson signal processes WH and HH for $m_{\mathrm{H}^{\pm}} = 120\,\mathrm{GeV}$. The misidentified τ_{h} background measured from data, 222.0 ± 11.4 (stat. + syst.) events, is consistent with the expectation from simulation. After the final event selection, the distributions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and of the τ_{h} transverse momentum after the full event selection are shown in Fig. 6.5.

6.4 Systematic uncertainties

The sources and the size of the systematic uncertainties are listed in Table 6.2. The following effects are taken into account:

Table 6.1: Numbers of expected events in the $\mu\tau_{\mathrm{h}}$ final state for the backgrounds and the charged Higgs boson signal from WH and HH processes at $m_{\mathrm{H}^+} = 120\,\mathrm{GeV}$, and the number of observed events after the final event selection.

Source	$\mathrm{N}_{\mathrm{events}}$ ($\pm$ stat. $\pm$ syst.)
HH+HW, m_{H^+}=120 GeV, $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b})$=0.05	$179.3 \pm 8.7 \pm 22.1$
τ fakes (from data)	222.0 ± 11.4
$t\bar{t} \to WbWb \to (\mu\nu b)\ (\tau_h\nu b)$	$304.7 \pm 2.8 \pm 25.9$
$t\bar{t} \to WbWb \to (\ell\nu b)\ (\ell\nu b)$	$21.4 \pm 0.7 \pm 6.9$
$Z/\gamma^* \to ee, \mu\mu$	$0.4 \pm 0.4 \pm 0.1$
$Z/\gamma^* \to \tau\tau$	$50.6 \pm 17.6 \pm 20.7$
Single top	$26.6 \pm 1.2 \pm 3.3$
VV	$4.4 \pm 0.5 \pm 0.7$
Total expected from SM	$630.1 \pm 17.9 \pm 46.9$
Data	620

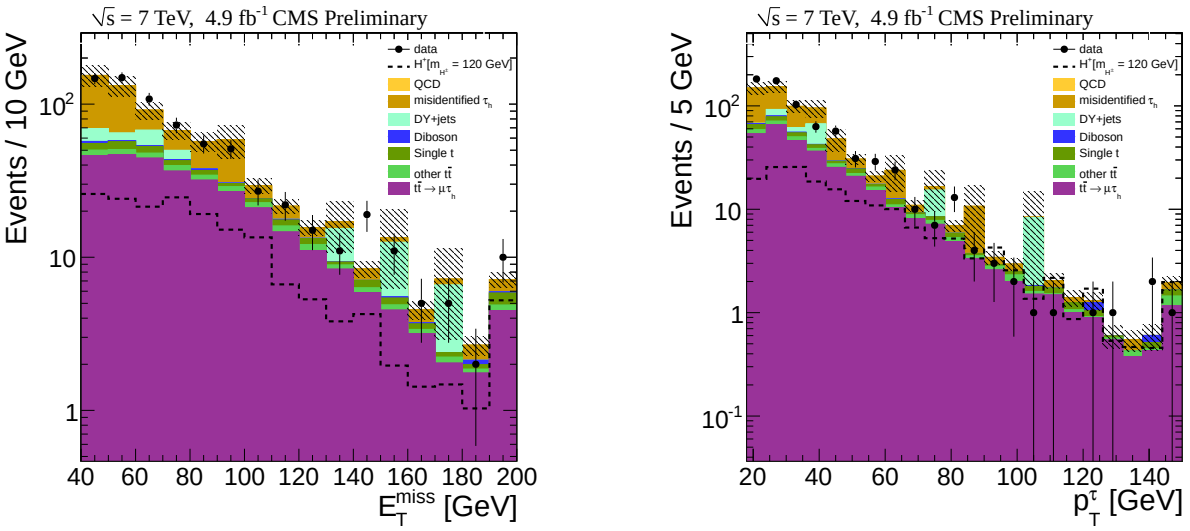


Figure 6.5: Distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ (left) and τ_{h} transverse momentum (right) after the full event selection. Distributions obtained from data (points) are compared with simulation. The simulated contributions are normalized to the SM predicted values. The expected event yield in the presence of the $\mathrm{t} \to \mathrm{H}^+\mathrm{b}$, $\mathrm{H}^+ \to \tau^+\nu_\tau$ decays is shown as a dashed line for $m_{\mathrm{H}^+} = 120\,\mathrm{GeV}$ and under the assumption that $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) = 0.05$. The last bin includes the overflow. The hatched area shows the total uncertainty.

- the uncertainty on the efficiency of τ identification, estimated to be 6% [178];
- the uncertainty on the rate of misidentification of a lepton as a τ_{h}, estimated to be 30% [178];
- The uncertainties on the jet energy scale (JES), jet energy resolution (JER), and $E_{\mathrm{T}}^{\mathrm{miss}}$ scale are estimated according to the prescription described in Ref. [181]; an uncertainty of 3% on the τ_{h} energy scale is included. These uncertainties also take into account the uncertainty due to JES dependence on the parton flavor. The uncertainty on JES is evaluated as a function of jet p_{T} and jet η. The JES and JER uncertainties are propagated in order to estimate the uncertainty of the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale;
- the uncertainty on the efficiency of b tagging, 5.4% [182];
- the uncertainty on the rate of misidentification of a jet as a b quark, 10% [135];
- the uncertainty on pileup modelling due to the reweighing of simulated events according to the measured distribution of the number of vertices;
- the uncertainty on the reconstruction, identification, and isolation efficiency of a muon is taken into account. The contribution from these sources is estimated to be $\simeq$2–3%.
- the uncertainty in the estimation of the misidentified τ_{h} ("τ fakes") background has two sources: the limited number of events for the measurement of the τ_{h} misidentification rate and the difference in the τ_{h} misidentification rates for jets originating from a quark with respect to jets originating from a gluon;
- the theoretical uncertainties on the signal and background cross sections;
- the uncertainty due to the limited number of events available in the simulated samples (MC stat.);
- finally, an estimated 2.2% uncertainty in the integrated luminosity [70];

The full sets of systematic uncertainties are used as input to the exclusion limit calculation.

Table 6.2: The systematic uncertainties on event yields (in percent) for the background processes and for the Higgs boson signal processes WH and HH for $m_{\mathrm{H}^+} = 120\,\mathrm{GeV}$.

	HH	WH	$t\bar{t}_{\ell\tau}$	$t\bar{t}_{\ell\ell}$	τ fakes	Single top	VV	DY(ee,$\mu\mu$)	DY($\tau\tau$)
τ-jet id	6	6	6			6	6		6
jet, $\ell \to \tau$ mis-id				30				30	
JES+JER+MET	6	4	5	4		6	11	100	21
b-jet tagging	6	5	5	5		7			
jet$\to$b mis-id							9	9	9
pile up	4	2	2	8		2	3	25	4
lepton selection	2	2	2	2		2	2	2	2
τ fakes					5				
cross-section	$^{+7}_{-10}$					8	4	4	
MC stats	4	5	1	3		4	11	100	35
luminosity	2.2					2.2			

6.5 Evaluation of limits on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$

The expected number of $\mathrm{t\bar{t}}$ events, after final event selection, is shown in Fig. 6.6 as a function of the branching fraction $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$ for $m_{\mathrm{H}^{+}} = 120\,\mathrm{GeV}$. Expectations are shown separately for contributions from WH, HH, and $\mathrm{t\bar{t}} \to \mathrm{WWb\bar{b}}$ (WW) processes. The total $\mathrm{t\bar{t}}$ event yield ($N_{\mathrm{t\bar{t}}}^{\mathrm{MSSM}}$) from WW, WH, and HH processes is larger than the yield from the standard model $\mathrm{t\bar{t}} \to \mathrm{WWb\bar{b}}$ process ($N_{\mathrm{t\bar{t}}}^{\mathrm{SM}}$). This is due to the fact that the branching fraction for the Higgs boson decay into $\tau\nu_{\tau}$ is larger than the corresponding branching fraction for the W boson decay.

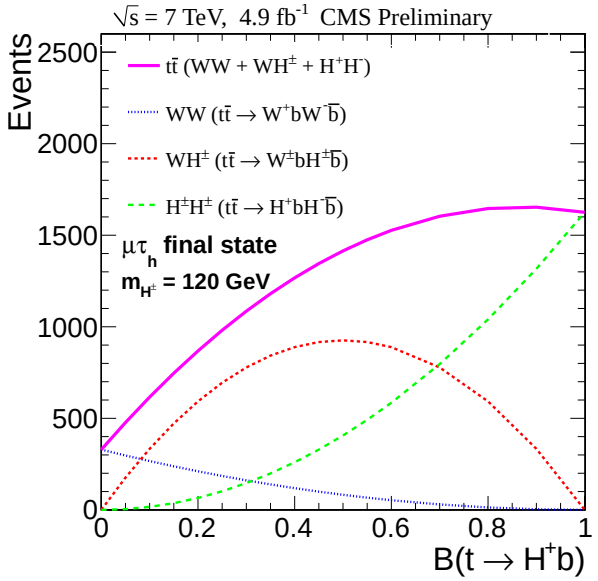


Figure 6.6: The expected number of $\mathrm{t\bar{t}}$ events after event selection as a function of the branching fraction $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$ for $m_{\mathrm{H}^{+}} = 120\,\mathrm{GeV}$. Expectations are shown separately for the WH, HH, and WW contributions.

Assuming that any excess of events in data, when compared with the expected background contribution, is due to the $\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}$, $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decays, the value of $x = \mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$ is related to the difference ΔN between the observed number of events and the predicted background contribution through the following equation:

$$\Delta N = N_{\mathrm{t\bar{t}}}^{\mathrm{MSSM}} - N_{\mathrm{t\bar{t}}}^{\mathrm{SM}} = 2x(1-x)N_{\mathrm{WH}} + x^{2}N_{\mathrm{HH}} + [(1-x)^{2} - 1]N_{\mathrm{t\bar{t}}}^{\mathrm{SM}}. \tag{6.4}$$

In this equation N_{WH} is estimated from simulation forcing the first top quark to decay to $\mathrm{H}^{\pm}\mathrm{b}$ and the second to $\mathrm{W}^{\mp}\mathrm{b}$, and N_{HH} forcing both top quarks to decay to $\mathrm{H}^{\pm}\mathrm{b}$. $N_{\mathrm{t\bar{t}}}^{\mathrm{SM}}$ is evaluated from simulation, as given by the $\mathrm{t\bar{t}}$ background in Table 7.3. The other backgrounds cancel out in the difference.

The CL_{s} method, described in Sec.4.3, is used to obtain an upper limit, at 95% confidence level (CL), on x using Eq. 6.4. The background and signal uncertainties described in Section 6.4 are modeled with a log-normal probability distribution function and their correlations are taken into account. The $R = p^{lead.track}/E_{\tau}$ distribution (Fig. 6.7), i.e. the ratio of the momentum of the leading track of the τ ($p^{lead.track}$) and the τ's energy (E_{τ}), is used in a binned maximum-likelihood fit in order to extract a

possible signal. This variable is linked to the polarization of the tau lepton, as discussed in Sec. 2.4.5.

Both statistical and systematic uncertainties are propagated to the limits by including the corresponding shapes of the R distribution calculated with up and down variations ($\pm 1\sigma$) separately for each component, and then feeding those distributions to the limit calculations. The upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$ as a

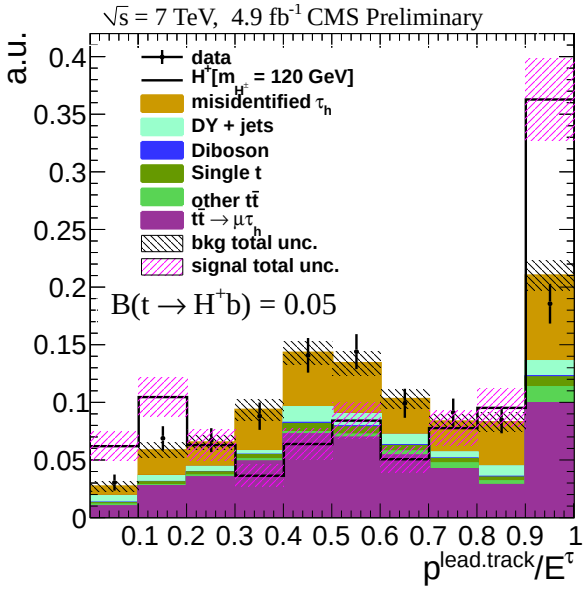


Figure 6.7: Distribution of $R = \frac{p^{lead.track}}{E_{\tau}}$. Data points, total SM backgrounds and charged Higgs boson signal yields are separately normalized to unity. The "misidentified τ_h" component is estimated using the data-driven method. The area shaded with lines indicates the total uncertainty.

function of $m_{\mathrm{H}^{+}}$ is shown in Fig. 6.8 (left), showing a 50% improvement on the results from the previous analysis of the same final state from Ref. [2]. The limits for the $\mu\tau_h$ final state are then combined with those obtained in the $e\tau_h$, $e\mu$ and τ_h+jet final states from Ref. [2] and the results are shown in Fig. 6.9. The combined upper limit is obtained using the procedure described in [183]. Table 6.3 summarizes the values of the median, $\pm 1\sigma$, and $\pm 2\sigma$ expected and the observed 95% CL upper limit for $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$ as a function of $m_{\mathrm{H}^{+}}$ for the combination of the fully hadronic, $\mathrm{e}\tau_{\mathrm{h}}$, $\mu\tau_{\mathrm{h}}$, and $\mathrm{e}\mu$ final states. The observed limits are above the expected limits due to the enhancement in the τ_{h}+jet final state alone [2]. All other final states yield observed limits within 1σ from expectations.

Table 6.3: The expected range and observed 95% CL upper limit for $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$ as a function of $m_{\mathrm{H}^{+}}$ for the combination of the $\mu\tau_{\mathrm{h}}$ (this analysis), fully hadronic, $\mathrm{e}\tau_{\mathrm{h}}$, and $\mathrm{e}\mu$ (from Ref. [2]) final states.

95% CL upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$						
$m_{\mathrm{H}^{+}}$ (GeV)	Expected limit					Observed limit
	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$	
80	0.0109	0.0159	0.0209	0.0284	0.0386	0.0291
100	0.0113	0.0144	0.0187	0.0260	0.0346	0.0278
120	0.0103	0.0138	0.0177	0.0238	0.0321	0.0247
140	0.00778	0.00938	0.0122	0.0172	0.0249	0.0183
150	0.00788	0.00954	0.0121	0.0174	0.0252	0.0214
155	0.00787	0.00851	0.0117	0.0171	0.0249	0.0209
160	0.00809	0.00806	0.0106	0.0149	0.0216	0.0173

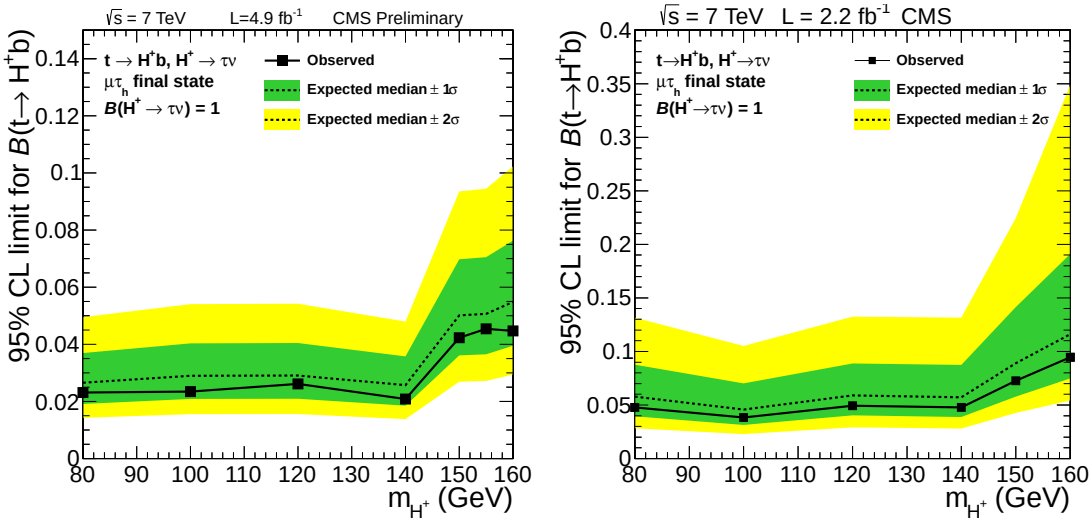


Figure 6.8: Upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$ as a function of $m_{\mathrm{H}^{+}}$ for the $\mu\tau_{\mathrm{h}}$ final state for the updated full-statistics analysis (left) and the previous result from Ref. [2] (right). The $\pm 1\sigma$ and $\pm 2\sigma$ bands around the expected limit are also shown.

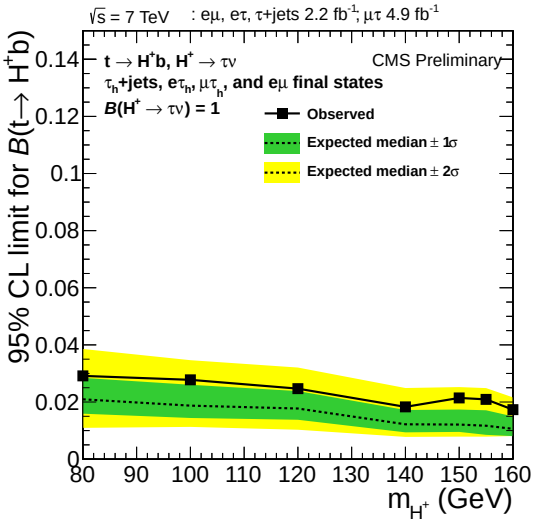


Figure 6.9: Upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$ as a function of $m_{\mathrm{H}^{+}}$ for the combination of the full-statistics $\mu\tau_h$ final state with the $e\tau_h$, $e\mu$, and τ_h+jet final states from Ref. [2]. The $\pm 1\sigma$ and $\pm 2\sigma$ bands around the expected limit are also shown.

6.6 Summary

Updated results have been presented in the search for a light charged Higgs boson produced in top quark decays $\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}$, with subsequent decay of H^{+} into $\tau^{+}\nu_{\tau}$. The results in the $\mu\tau_{\mathrm{h}}$ final state have been updated with a data sample corresponding to an integrated luminosity of 4.9 fb^{-1} collected in proton-proton collisions at $\sqrt{s} = 7$ TeV in 2011, and are combined with those previously published in Ref. [2] for the eτ_h, eμ and τ_h+jet final states. Upper limits on the branching fraction $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b})$ in the range of 2–3% are established for a charged Higgs boson with a mass between 80 and 160 GeV, under the assumption that $\mathcal{B}(\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}) = 1$. The results of this analysis represent a 50% improvement on the sensitivity of the analysis to the presence of a charged Higgs boson decaying into a tau and a neutrino, with respect to the previous result described in Ref. [2].

Chapter 7

Search for a heavy charged Higgs boson

7.1 Introduction to heavy charged Higgs boson

In this chapter, a search for the charged Higgs boson is performed in a data sample recorded by the CMS experiment at $\sqrt{s} = 8$ TeV and corresponding to an integrated luminosity of $19.7 \pm 0.5\,\text{fb}^{-1}$. The charged Higgs boson decay modes and final states discussed in this thesis are summarised in Table 7.1. The $\mathrm{H}^+ \to \tau^+\nu_\tau$ and $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ decay modes are inclusively studied in the $\mu\tau_\mathrm{h}$, single lepton (e+jets and μ+jets), and $\ell\ell'$ final states for $m_{\mathrm{H}^+} > (m_\mathrm{t} - m_\mathrm{b})$. Combined limits for the $\mathrm{H}^+ \to \tau^+\nu_\tau$ and $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ decay modes are set on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+)$ by assuming the branching fraction of the considered decay mode to be 100 %. Additionally, model-independent limits without assumption on the charged Higgs boson branching fraction are set for both light and charged Higgs boons on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ and $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ with the analysis on the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode in the τ_h+jets final state for $m_{\mathrm{H}^+} < (m_\mathrm{t} - m_\mathrm{b})$ and $m_{\mathrm{H}^+} > (m_\mathrm{t} - m_\mathrm{b})$, respectively. All the considered decay modes and final states are used to set exclusion limits in the m_{H^+}-$\tan\beta$ parameter space for different MSSM benchmark scenarios [11, 12] by applying the scenario-specific branching ratio for the decay modes. This manuscript reports for the first time results on direct charged Higgs boson production for $m_{\mathrm{H}^+} > m_\mathrm{t} - m_\mathrm{b}$ in the $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ decay mode.

The searches in the single lepton and τ_h+jets final states have been performed by other groups, and are reported in App. A and B because the discussion of those results is relevant to the forementioned combination, of which the author of this manuscript was responsible.

The two main modes involved in the production of a high-mass ($m_{\mathrm{H}^+} > m_\mathrm{t} - m_\mathrm{b}$) charged Higgs boson at a proton-proton collider are characterized by the fusion of bottom and top quarks, as illustrated in Figs. 1.1 (b) and (c) and mentioned in Chap. 1 and referred to as *four-flavour scheme* (4FS) and *five-flavour scheme* (5FS), respectively. In the 4FS there are no b quarks in the initial state causing a different ordering of the perturbative terms at a finite order between the 4FS and 5FS [12]. The predictions of the 4FS and the 5FS cross sections calculated at next-to-leading order (NLO) are combined using the

Table 7.1: Overview of the charged Higgs boson production processes, decay modes, final states, and mass ranges analysed in this manuscript ($\ell = \mathrm{e}, \mu$). The "jets" in τ_h+jets and ℓ+jets refers to hadronic decay of the or a W boson, respectively. All final states contain in addition jets from the hadronization of b quarks and missing transverse energy from neutrinos.

Decay mode	Signatures for $m_{\mathrm{H}^+} < m_\mathrm{t} - m_\mathrm{b}$	Signatures for $m_{\mathrm{H}^+} > m_\mathrm{t} - m_\mathrm{b}$
	$\mathrm{t\bar{t}} \to \mathrm{bH^+\bar{b}H^-}$, $\mathrm{t\bar{t}} \to \mathrm{bH^+\bar{b}W^-}$	$\mathrm{pp} \to \mathrm{\bar{t}(b)H^+}$
$\mathrm{H^+} \to \tau^+\nu_\tau$	τ_h+jets	τ_h+jets, $\mu\tau_\mathrm{h}$, $\ell\ell$
$\mathrm{H^+} \to \mathrm{t\bar{b}}$	-	$\mu\tau_\mathrm{h}$, ℓ+jets, $\ell\ell$

"Santander matching scheme" [184]. Calculated to all orders in perturbation theory, the 4FS and 5FS yield identical production cross sections.

Cross sections and branching ratios predicted in the m_h^{mod+} benchmark scenario are shown in Fig. 7.1.

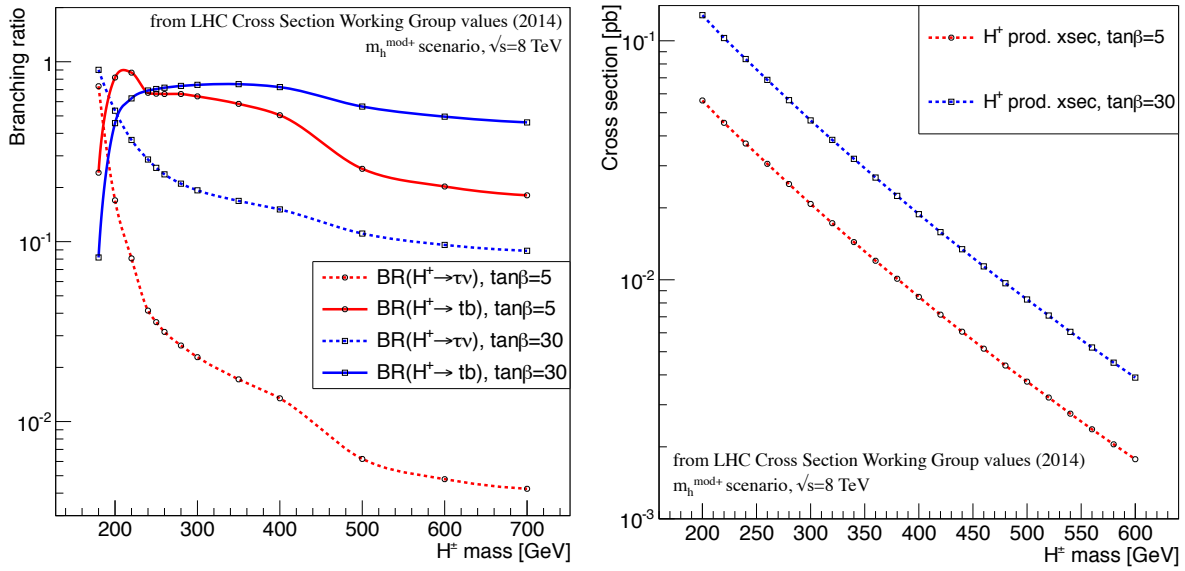


Figure 7.1: The heavy charged Higgs boson branching ratios (left) and cross sections (right) as predicted in the m_h^{mod+} benchmark scenario, as a function of the charged Higgs boson mass..

The event selection together with the background estimation is described in Sections 7.2, 7.6, and App. B, for the $\mu\tau_\mathrm{h}$, $\ell\ell'$, and τ_h+jets final states, respectively. The treatment of statistical and systematic uncertainties is described in Section 7.11. The results are presented in Section 7.12 and summarised in Section 7.13.

7.2 The $\mu\tau_\mathrm{h}$ final state for $\mathrm{H^+} \to \tau^+\nu_\tau$ and $\mathrm{H^+} \to \mathrm{t\bar{b}}$

In this analysis, a charged Higgs boson with a mass larger than the top quark is produced through $\mathrm{pp} \to \mathrm{\bar{t}(b)H^+}$ and searched for in the decay modes $\mathrm{H^+} \to \tau^+\nu_\tau$ and $\mathrm{H^+} \to \mathrm{t\bar{b}}$. In the first case, the τ decays hadronically and the final state is characterized by the leptonic decay into a muon of the W boson that originates from the $\mathrm{\bar{t}} \to \mathrm{\bar{b}W^-}$ decay. In the second, at least one of the W bosons from the

top quarks decays to a hadronically decaying τ. In this final state, the charged Higgs boson production is characterized by a larger number of b-jets than in the SM backgrounds, and consequently the shape of the b-tagged jet multiplicity distribution is used to infer the presence of a signal. The dominant SM background processes are from $\mathrm{t\bar{t}} \to \mu\tau_{\mathrm{h}} + X$ and other backgrounds where a jet is misidentified as a τ_{h} (mainly lepton+jet $\mathrm{t\bar{t}}$ events and W+jet production).

The number of produced $\mathrm{t\bar{t}}$ events is normalized to the predicted $\mathrm{t\bar{t}}$ production cross section $246.7^{+6.2}_{-8.4} \pm 11.4$ pb as calculated with the Top++2.0 program to next-to-next-to-leading order (NNLO) in perturbative QCD, including soft-gluon resummation to next-to-next-to-leading-logarithm (NNLL) order [152], and assuming a top-quark mass mtop = 173.34 GeV. The first uncertainty comes from the independent variation of the factorization and renormalization scales, μ_F and μ_R, while the second is associated to variations in the parton density functions (PDFs) and strong coupling constant α_{S}, following the PDF4LHC prescription with the MSTW2008 68 % confidence level (CL) NNLO, CT10 NNLO and NNPDF2.3 5-flavour FFN PDF sets [153, 154, 155, 185]. The predicted cross section is in good agreement with the experimental measurements by ATLAS and CMS [186]. The top quark p_{T} spectrum in data is found to be softer than predicted using the MADGRAPH MC generator [187]. To correct for this effect, the $\mathrm{t\bar{t}}$ events are reweighted to make the top quark p_{T} spectrum in simulation match that observed in data [188, 189].

The NNLO SM prediction calculated with FEWZ 3.1 is taken for the W+jets and Z/γ^* backgrounds [190, 191]. The cross section for the t-channel single top quark sample is calculated at next-to-leading order (NLO) in QCD with Hathor v2.1 [192, 193] with PDF and α_{S} uncertainties calculated using the PDF4LHC prescription [153]. For the single top quark s-channel and tW-channel cross section the SM prediction at NNLL in QCD is taken from Refs. [194, 195].

7.3 Event selection

The event selection is similar to that used in the measurement of the top quark pair production cross section in dilepton final states containing a τ_{h} [146, 196]. A single-muon trigger with thresholds of $p_{\mathrm{T}} > 24$ GeV and $|\eta| < 2.1$ is used to select the events.

Events are selected by requiring one isolated muon with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.1$, one hadronically decaying τ with $p_{\mathrm{T}} > 20$ GeV within $|\eta| < 2.4$, at least two jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$, with at least one jet identified as originating from the hadronization of a b quark, and $E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV. The τ_{h} and the muon are required to have opposite electric charges. The muon candidate is considered to be isolated if the relative isolation, as defined in Section 3.4, is $I_{\mathrm{rel}} < 0.12$. The muon and the tau are required to be separated from each other and from any selected jet by a distance $\Delta R > 0.4$. The choice of the radius matches the lepton isolation cone. Events with an additional electron (muon) with $I_{\mathrm{rel}} < 0.2$ and $p_{\mathrm{T}} > 15\,(10)$ GeV are rejected.

In this analysis, the b-tagging working point is loosened to a one per cent mistagging probability since the multijet background is smaller than in the τ_{h}+jets final state described in App. B . The corresponding probability to idenfity a b jet is about 70 %.

7.4 Background estimate

The backgrounds come from two main categories: the first with misidentified τ_{h} leptons, and the second with genuine τ_{h}s. Within the first category, the component with misidentified τ_{h} from generic jets is estimated from data: this includes contributions from W+jets and multijet events, as well as from $\mathrm{t\bar{t}} \to \ell$+jets events. The background from Z/γ^* or $\mathrm{t\bar{t}}$ dilepton events with one electron or muon misidentified as a τ_{h} is small and is estimated from simulation. Within the second category, the background due to $\mathrm{Z}/\gamma^* \to \tau\tau$ events with one τ_{h} and one τ decaying to a muon is estimated by taking the normalization from simulation and the shape from $\mathrm{Z} \to \mu\mu$ events in data where each muon has been replaced with reconstructed particles from a simulated τ lepton decay.

The remaining SM backgrounds with genuine τ leptons are from single-top-quark production, di-bosons, and from $\mathrm{t\bar{t}}$ events in which a τ is produced from a W decay. The $\mathrm{Z}/\gamma^* \to \mathrm{ee}$, $\mu\mu$ and $\mathrm{t\bar{t}} \to \mathrm{W^+W^-b\bar{b}} \to \ell^+\nu\ell^-\overline{\nu}\mathrm{b\bar{b}}$ events may also contain electrons or muons misidentified as τ_{h}. The events from SM $\mathrm{t\bar{t}}$ production are labeled "other $\mathrm{t\bar{t}}$" in the following tables and plots. The yields for these backgrounds are estimated from simulation. The $\mathrm{t\bar{t}}$ events are categorized in order to separate the contribution from each decay mode The categorization is performed using the generated particles from simulation.

The misidentified τ_{h} background comes from events with one muon, $E_{\mathrm{T}}^{\mathrm{miss}}$, and three or more jets with at least one identified b quark jet (labelled "$\mu+ \geq 3$ jets" events), where one jet is misidentified as a τ_{h}. The dominant contribution to this background comes from W+jets, and from $\mathrm{t\bar{t}} \to \mathrm{W^+W^-b\bar{b}} \to \mu\nu\mathrm{q\bar{q}'b\bar{b}}$ events. The background contribution from misidentified τ_{h} is estimated by applying the probability that a jet mimics a τ_{h} to every jet in "$\mu+ \geq 3$ jets" events, as event weights. The probability that a jet is misidentified as a τ_{h} is measured from data as a function of jet p_{T}, jet η and jet radius using W+jet and multijet events [91, 196] In this document, the estimate of the τ fake background is improved with respect to the method used in Ref. [2] by weighing the raw estimates, and with respect to the method used in Chap. 6 by taking into account the quark and gluon jet compositions (from simulation) in the W+jet and multi-jet samples [196].

The probability $w(jet \to \tau)$ is evaluated using all jets (except the triggering jet) in the QCD multi-jet events (w_{QCD}) and all jets in the W+≥ 1 jet events (w_{W+jets}) as a function of jet p_{T}, η, R_{jet}, and is compared to the distributions expected from simulation in Figs. 7.4, 7.5, 7.6 (for the QCD multi jet events), and in Figs. 7.7, 7.8, 7.9 (for the W+≥ 1 jet events).

The improvement in the central value of the estimate is verified with a closure test with simulated events, and the systematics uncertainty improves by 30%.

The weights for the two estimates, W_{W+jets} and W_{QCD} are taken as the weights that make a linear combination of the two estimates to have the same quark and gluon jets composition as the final $\ell + E_{\mathrm{T}}^{\mathrm{miss}} \geq 3$ jet sample (as estimated in simulation).

The correspondence between the kNN estimate and the true quark and gluon jet composition in the

training samples is expressed by:

$$FR^{W+jet} = FR^{q-jets} \cdot Frac^{q-jets}_{W+1jet} + FR^{g-jets} \cdot Frac^{g-jets}_{W+1jet} \tag{7.1}$$

$$FR^{QCD} = FR^{q-jets} \cdot Frac^{q-jets}_{QCD} + FR^{g-jets} \cdot Frac^{g-jets}_{QCD} \tag{7.2}$$

Where FR^{W+jet} and FR^{QCD} are the fake rates extracted through the kNN estimate, FR^{q-jets} and FR^{q-jets} are the unknown fake rates for quark jets and gluon jets, respectively, which are applied to fractions of the sample expressed by $Frac^{q-jets}_{W+1jet}$ $Frac^{g-jets}_{W+1jet}$ for the "W+jet" sample and by $Frac^{q-jets}_{QCD}$ and $Frac^{q-jets}_{QCD}$ for the QCD sample. Finally, the weights are found by imposing that the final estimate should have the same result as applying the unknown fake rates for quark jets and gluon jets to the corresponding quark and gluon jet fractions in the final sample:

$$W_{W+jet} \times FR^{W+1jet} + W_{QCD} \times FR^{QCD} = FR^{q-jets} \cdot Frac^{q-jets}_{Wjets+t\bar{t}\to ljets} + FR^{g-jets} \cdot Frac^{g-jets}_{Wjets+t\bar{t}\to ljets} \tag{7.3}$$

The fraction of quark and gluon jets from MC are shown as a function of the transverse momentum of the jet in Fig. 7.2 (left) for the QCD sample and in Fig. 7.2 (right) for the W+jet sample. The fraction of quark and gluon jets from MC are shown as a function of the transverse momentum of the jet in Fig. 7.3 for the final $\ell + E_{\mathrm{T}}^{\mathrm{miss}} \geq 3$ jet sample.

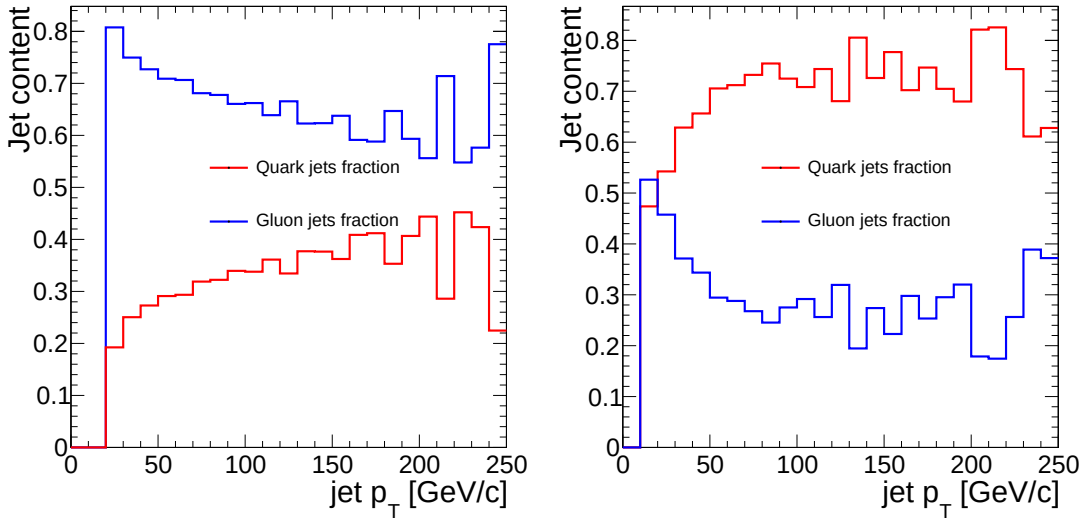


Figure 7.2: Fraction of quark (red) and gluon (blue) jets as evaluated from MC in the QCD training sample (left) and in the W+jet training sample (right).

The corresponding contribution of $N_{non-\tau-fake}$ described earlier are subtracted from Eq. 6.2. The uncertainties of w_{QCD} and w_{W+jets} (statistical uncertainty) are propagated to $N^{\tau-\mathrm{fake}}$ through Eq. 6.2. Finally, the efficiency ε_{OS}=0.69±0.04 of the OS requirement obtained from MC is applied to obtain the

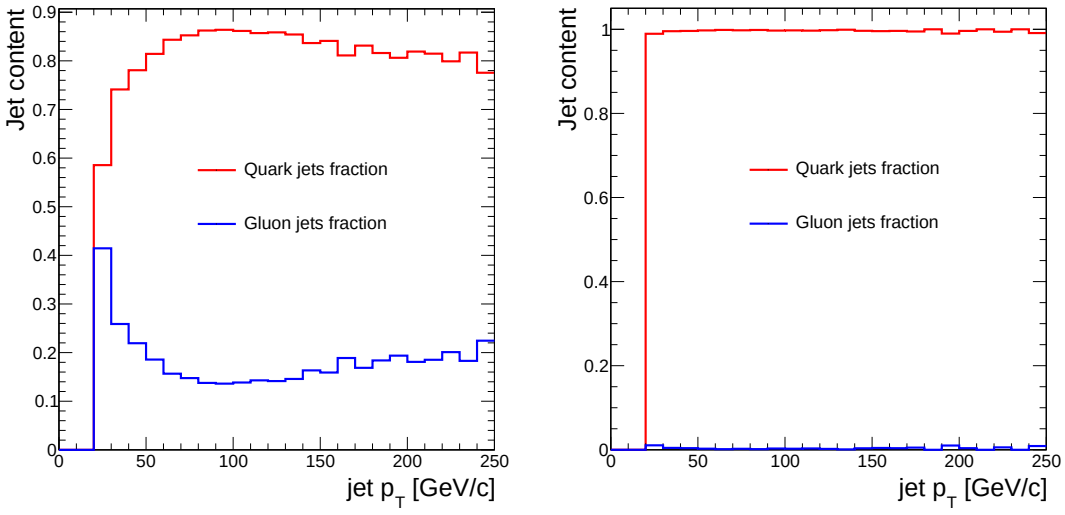


Figure 7.3: Fraction of quark (red) and gluon (blue) jets as evaluated from MC in the final $\ell + E_{\mathrm{T}}^{\mathrm{miss}} \geq 3$ jet sample (left) and in the signal sample ($M_{H^{\pm}} = 250\ GeV$) (right).

τ-fake background $N_{OS}^{\tau-\mathrm{fake}}$ as:

$$N_{OS}^{\tau-\mathrm{fake}} = \varepsilon_{OS} \times N^{\tau-\mathrm{fake}} \tag{7.4}$$

The estimated MC efficiency ε_{OS}=0.69$\pm$0.04 is in agreement with past results The estimation of ε_{OS} is cross checked in data using the W+≥ 3 jet type events. The W+≥ 3 jet control sample is selected as described in the earlier paragraph of this section, with anti-b-tagging to suppress $\mathrm{t\bar{t}}$ events. The MC OS efficiency has been calculated from $\mathrm{t\bar{t}} \rightarrow \ell + jets$ events.

The number of tau fakes $N^{\tau-\mathrm{fake}}$ is estimated using $w(jet \rightarrow \tau)$ measured from QCD multi-jet events and for "W+jet" events, by re-weighing according to the different quark and gluon jets composition of the training and final sample (Table 7.2). The column "estimated from data" gives the number of expected events when the data-driven method is applied; the data-driven background estimate is obtained after the τ identification selection step: $N_{\tau} = 1$. The column "Estimated from MC" contains the number of τ fake events predicted by the MC using the data-driven method when the $jet \rightarrow \tau_h$ fake rate has been obtained from the MC QCD multi-jet events or from the MC "W+jet" events, respectively. The column "MC expectation" gives the number of expected τ fake events from MC using the standard signal selection. This number is obtained as the sum of the number of MC events for the processes $t\bar{t} \rightarrow \ell + jets$, W+jets and QCD. It is to be observed that the number of events in the columns "MC expectation" and "Estimated from MC" are in well agreement within their uncertainties and thus validating the robustness of this background estimation method. As discussed earlier, "$\ell+ \geq 3$ jet" control sample also includes non-"tau fake" ($N_{non-\tau-fake}$) residual events from other processes. These residual background events include $\mathrm{t\bar{t}}$ events (dileptons, $\tau + jets$, $\tau\tau$, full had.), Z+jets, single top The contribution of residual background events is estimated from the MC and it is subtracted from the data-driven estimate, thus the numbers shown in column "estimated from data" of Table 7.2 are after the residual background sub-

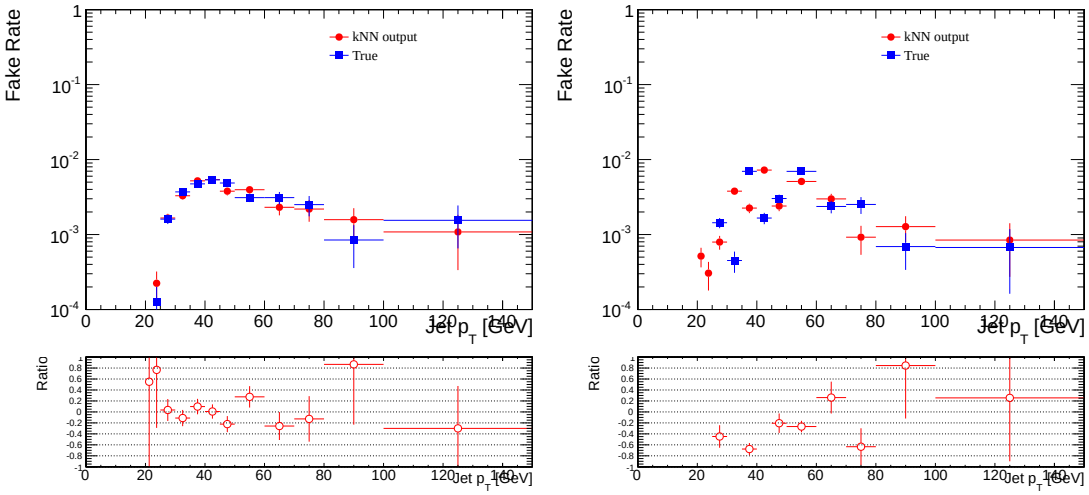


Figure 7.4: Comparison between the output of the algorithm and the expected distributions from the test set: tau fake rate distribution obtained from QCD multi-jet samples as a function of jet p_{T} for data (left) and simulation (right). The events are selected with HLT_Jet30 trigger and Run2011(A+B) samples. Pythia 6 with TuneZ2 (with pile-up) has been used for MC QCD multi-jet samples. The fake rates are calculated for the HPS ("medium") τ algorithm.

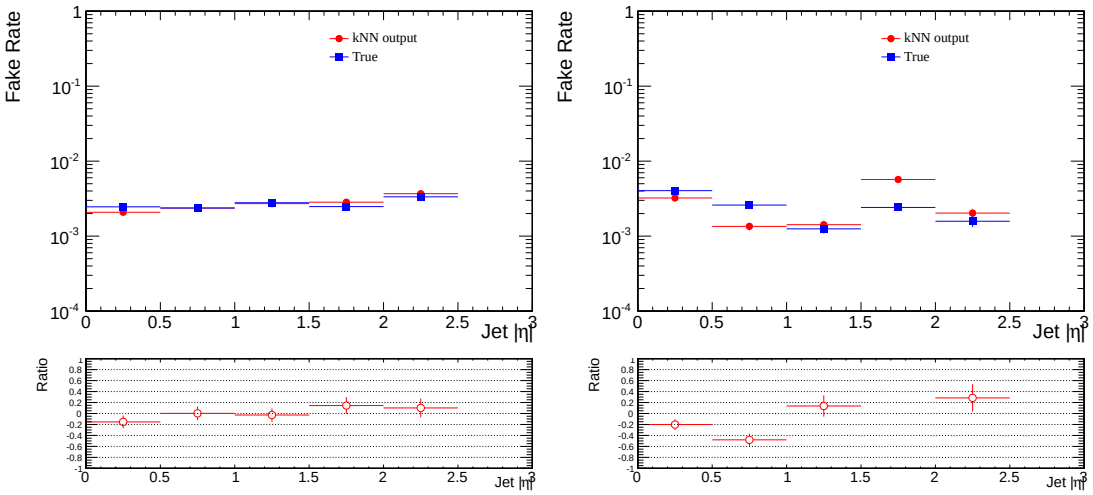


Figure 7.5: Comparison between the output of the algorithm and the expected distributions from the test set: tau fake rate distribution obtained from QCD multi-jet samples as a function of jet η for data (left) and simulation (right). The events are selected with HLT_Jet30 trigger and Run2011(A+B) samples. Pythia 6 with TuneZ2 (with pile-up) has been used for MC QCD multi-jet samples. The fake rates are calculated for the HPS ("medium") τ algorithm.

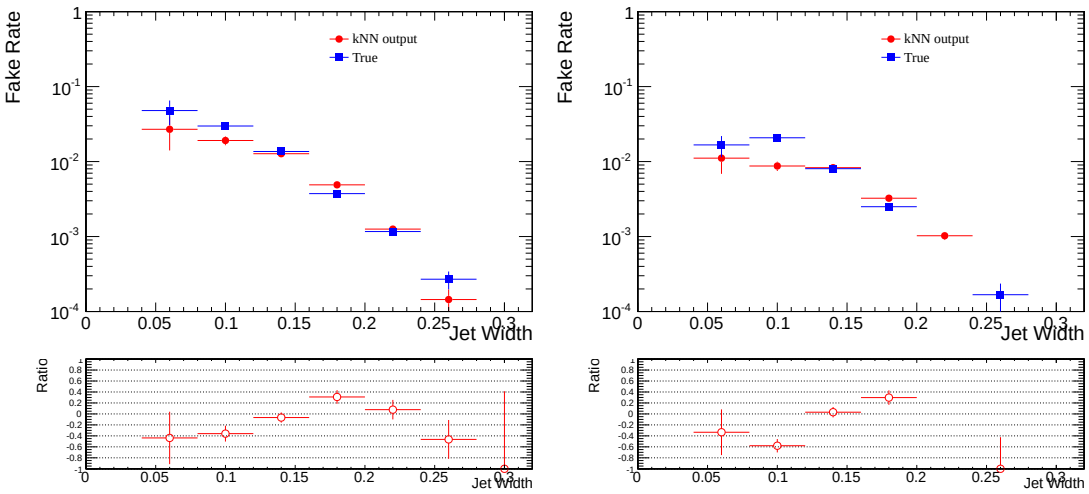


Figure 7.6: Comparison between the output of the algorithm and the expected distributions from the test set: tau fake rate distribution obtained from QCD multi-jet samples as a function jet radius for data (left) and simulation (right). The events are selected with HLT_Jet30 trigger and Run2011(A+B) samples. Pythia 6 with TuneZ2 (with pileup) has been used for MC QCD multi-jet samples. The fake rates are calculated for the HPS ("medium") τ algorithm.

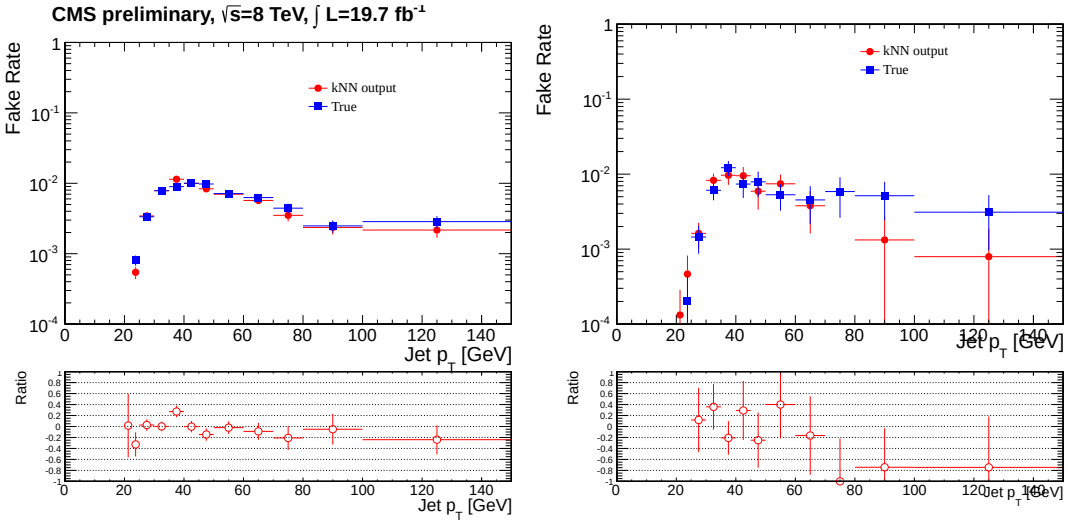


Figure 7.7: Comparison between the output of the algorithm and the expected distributions from the test set: tau fake rate distribution obtained from W+$\geq$1jet samples as a function of jet p_{T} for data (left) and simulation (right). The events are selected with HLT_Jet30 trigger and Run2011(A+B) samples. Pythia 6 with TuneZ2 (with pile-up) has been used for MC QCD multi-jet samples. The fake rates are calculated for the HPS ("medium") τ algorithm.

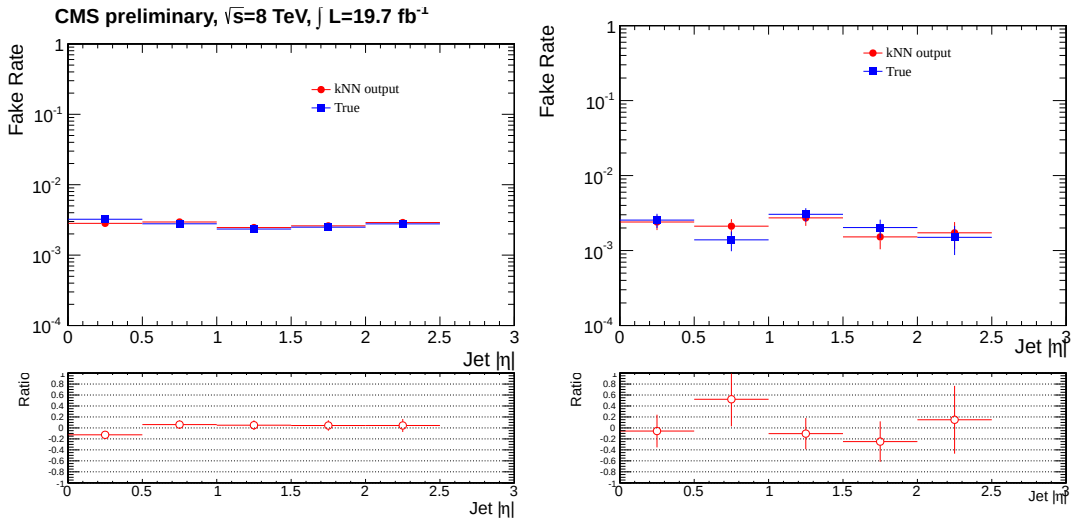


Figure 7.8: Comparison between the output of the algorithm and the expected distributions from the test set: tau fake rate distribution obtained from W+$\geq$1jet samples as a function ofjet η for data (left) and simulation (right). The events are selected with HLT_Jet30 trigger and Run2011(A+B) samples. Pythia 6 with TuneZ2 (with pile-up) has been used for MC QCD multi-jet samples. The fake rates are calculated for the HPS ("medium") τ algorithm.

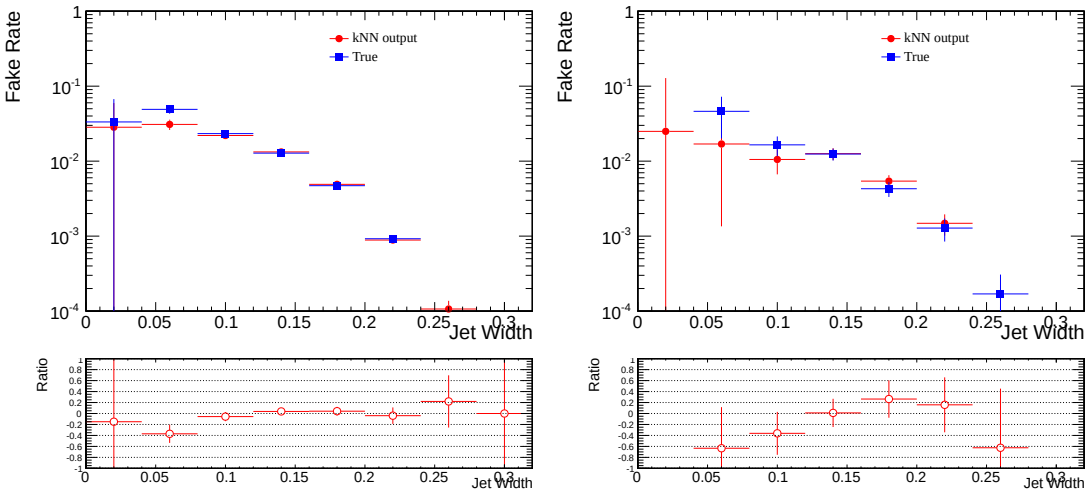


Figure 7.9: Comparison between the output of the algorithm and the expected distributions from the test set: tau fake rate distribution obtained from W+$\geq$1jet samples as a function of jet radius for data (left) and simulation (right). The events are selected with HLT_Jet30 trigger and Run2011(A+B) samples. Pythia 6 with TuneZ2 (with pile-up) has been used for MC QCD multi-jet samples. The fake rates are calculated for the HPS ("medium") τ algorithm.

traction. The unweighted average for the tau fake background estimate is computed (Table 7.2, row "unweighed average") before the OS cut. The proper weighing of the raw estimates, which takes into account the quark and gluon jets composition from MC is shown in the last row, "weighed average". In order to estimate the fake background after the OS cut, the number of background events estimated after the tau identification cut ($N_\tau = 1$) is then scaled by the efficiency, as described above, which is derived from MC.

Lepton pairs with additional jet production can also mimic $\mathrm{t\bar{t}}$ dilepton events. In particular, $DY \to \tau\tau$ events accompanied by additional jet production may appear as $\mathrm{t\bar{t}}$ events, when one tau decays leptonically and the other hadronically. This background is estimated from MC.

Table 7.2: The number of τ fake events, estimated using the data-driven method (column "estimated from data") described in the text, is compared to the expectations from MC (column "estimated from MC") obtained by applying the method to simulated events. The τ fake probability is obtained from QCD multi-jet and W+jet data samples; the last two rows show the average of the two results before and after re-weighing for the quark-gluon composition of the samples. Only statistical uncertainties are shown. The residual contribution of "non-fake" events is estimated from simulation and it is subtracted from the estimate of "τ-fake" events in order to avoid double counting.

Sample	MC expectation	Estimated from MC	Estimated from data
QCD multi-jet	2341$\pm$61	1983	1994
W+jets		2642	2499
Unweighted average		2312 ($\pm$14%)	2246 ($\pm$11%)
Weighted average		2095 ($\pm$14%)	2145 ($\pm$11%)

The fake rate estimation obtained using the data-driven method is affected by a total (stat. + syst.) uncertainty of 9%, which is one of the largest contribution to the systematic uncertainty (described in Sec. 6.4) associated with the determination of the expected SM yields. Still, the 9% represents a 30% improvement with respect to the estimate obtained in Ref. 6.

The background contribution from misidentified τ_h from electrons and muons is small and it comes from $\mathrm{t\bar{t}} \to \ell\ell$+jets and $Z \to \ell\ell$ events. This contribution is small and it is estimated from simulation.

The single lepton trigger efficiency and the muon isolation and identification efficiencies are corrected by multiplicative data-to-simulation scale factors dependent on the muon transverse momentum and pseudo-rapidity, derived using a "tag-and-probe" method [197, 198].

7.5 Event yields

Data and the simulated event yield at various steps of the event selection are shown in Fig. 7.10, left. Since the data-driven estimation is derived only after requiring one τ_h, the backgrounds are normalized to the SM prediction obtained from the simulation. A good agreement is found between data and the SM background expectations. The multijet background contribution is negligible at the final selection step. The expected event yields in the presence of $\mathrm{H^+} \to \mathrm{t\bar{b}}$ and $\mathrm{H^+} \to \tau^+\nu_\tau$ decays are shown as dashed lines for $m_{\mathrm{H^+}} = 250\,\mathrm{GeV}$, normalized to a cross section of 1 pb and assuming a branching fraction of

100 % for each decay channel.

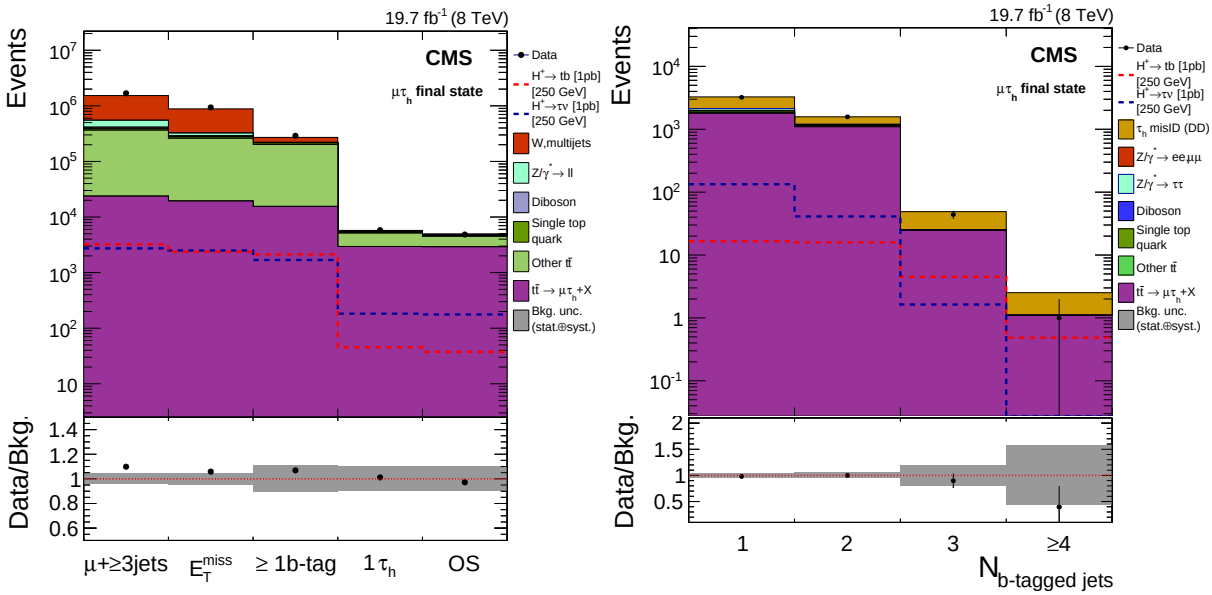


Figure 7.10: Left: event yields after each selection step, where OS indicates the requirement to have opposite electric charges for the τ_{h} and the μ. The backgrounds are estimated from simulation and normalized to the SM prediction. Right: the b-tagged jet multiplicity distribution after the full event selection. The "misidentified τ_{h}" component is estimated using the data-driven method and labeled "τ_{h} misID(DD))", while the remaining background contributions are from simulation normalized to the SM predicted values. For both distributions, the expected event yield in the presence of the $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ and $\mathrm{H}^+ \to \tau^+\nu_\tau$ decays is shown as dashed lines for $m_{\mathrm{H}^+} = 250\,\text{GeV}$. The signal yields are normalized to a cross section of 1 pb assuming $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t\bar{b}}) = 1$ and $\mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1$, respectively. The bottom panel shows the ratio of data over the sum of the SM backgrounds with the total uncertainties. Statistical and systematic uncertainties are added in quadrature.

The number of expected events for the SM backgrounds, the number of signal events from the $\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+$ process for $m_{\mathrm{H}^+} = 250\,\text{GeV}$ for the decay modes $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ and $\mathrm{H}^+ \to \tau^+\nu_\tau$, and the number of observed events after all selection cuts are summarized in Table 7.3. The misidentified τ_{h} background measured from data is consistent with the expectations from simulation. Statistical and systematic uncertainties evaluated as described in Section 7.11 are also shown. The number of signal events is normalized to a cross section of 1 pb, assuming 100 % branching fraction for each decay mode.

The b-tagged jet multiplicity is shown after the full event selection in Fig. 7.10, right. The ratio of the data to the sum of expected SM background contributions is shown in the bottom panel. Limits on the production of the charged Higgs boson are extracted by exploiting this distribution.

Table 7.3: Number of expected events in the $\mu\tau_{\mathrm{h}}$ final state for the SM backgrounds and in the presence of a signal from $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ and $\mathrm{H}^+ \to \tau^+\nu_\tau$ decays for $m_{\mathrm{H}^+} = 250\,\text{GeV}$ are shown together with the number of observed events after the final event selection. The signal cross section is normalized to a cross section of 1 pb, assuming a branching fraction $\mathcal{B} = 100\,\%$ for each decay channel.

Source	$\mathrm{N}_{\mathrm{events}}$ ($\pm$ stat. $\pm$ syst.)
$\mathrm{H}^+ \to \tau^+\nu_\tau$, $m_{\mathrm{H}^+} = 250\,\text{GeV}$	$176 \pm 10 \pm 13$
$\mathrm{H}^+ \to \mathrm{t\bar{b}}$, $m_{\mathrm{H}^+} = 250\,\text{GeV}$	$37 \pm 2 \pm 3$
$\mathrm{t\bar{t}} \to \mu\tau_{\mathrm{h}} + \mathrm{X}$	2913 $\pm$ 14 $\pm$ 242
misidentified τ_{h}	1544 $\pm$ 14 $\pm$ 175
$\mathrm{t\bar{t}}$ dilepton	101 $\pm$ 10 $\pm$ 27
$\mathrm{Z}/\gamma^* \to \mathrm{ee}, \mu\mu$	12 $\pm$ 3 $\pm$ 4
$\mathrm{Z}/\gamma^* \to \tau\tau$	162 $\pm$ 40 $\pm$ 162
single top	150 $\pm$ 12 $\pm$ 18
dibosons	20 $\pm$ 3 $\pm$ 2
total SM backgrounds	4903 $\pm$ 45 $\pm$ 341
data	4839

7.6 The dilepton ($\mathrm{ee}/\mathrm{e}\mu/\mu\mu$) final states for $\mathrm{H}^+ \to \tau^+\nu_\tau$ and $\mathrm{H}^+ \to \mathrm{t}\overline{\mathrm{b}}$

The $\ell\ell'$ final state is studied, where the full production and decay chain goes through $\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+$, with $\bar{\mathrm{t}}(\mathrm{b}) \to \ell\nu_\ell\overline{\mathrm{b}}(\mathrm{b})$ and $\mathrm{H}^+ \to \ell'\nu_{\ell'}\mathrm{b}\overline{\mathrm{b}}$. This is similar to the SM $\mathrm{t}\bar{\mathrm{t}}$ dilepton final state, with the addition of one or two b jets. The shape of the b-tagged jet multiplicity distribution is used to infer the presence of a charged Higgs boson signal. The dominant SM backgrounds are from $\mathrm{t}\bar{\mathrm{t}}$ and single top quark production.

7.7 Event selection

The event selection is similar to that used for the measurement of the SM $\mathrm{t}\bar{\mathrm{t}}$ cross section and of the ratio $\frac{\mathcal{B}(t\to Wb)}{\mathcal{B}(t\to Wq)}$ in the dilepton channel [197, 199]. Data are collected with double-lepton triggers (one electron and one muon, two electrons, or two muons) with p_T thresholds of 17 GeV for the leading lepton and 8 GeV for the other. After offline reconstruction, events are selected with two isolated, oppositely charged, leptons (one electron and one muon, or two electrons, or two muons) with $p_\mathrm{T} > 20$ GeV and $|\eta| < 2.5$ ($|\eta| < 2.4$) for electrons (muons), and at least two jets with $p_\mathrm{T} > 30$ GeV and $|\eta| < 2.4$. The relative isolation requirement is I_rel <0.15(0.20) for electrons (muons). Jets are required to be separated by a distance $\Delta R = 0.4$ from the isolated leptons. A minimum dilepton invariant mass of 12 GeV is required to reject SM background from low mass resonances. For the same flavour channels (ee, $\mu\mu$), events with dilepton invariant mass within 15 GeV from the Z boson mass are vetoed. in order to account for the presence of neutrinos, a $E_\mathrm{T}^\mathrm{miss} > 40$ GeV threshold is required. At least two b-tagged jets are required.

The b-tagging working point is further loosened, with respect to the lepton+tau final state described in Sec. 7.2, to a 10 % mistagging probability to enhance signal acceptance since the multijet background in this analysis is even smaller. The corresponding probability to identify a b jet is about 85 %.

7.8 Background estimate

All backgrounds are estimated from simulation. The dilepton trigger efficiency is corrected by a multiplicative data-to-simulation scale factor dependent on the final state, in order to provide agreement between data and simulation; the corresponding scale factors are computed using a "tag-and-probe" method [197, 198], and the resulting values are 0.97, 0.95, and 0.92 for the ee, $\mathrm{e}\mu$, and $\mu\mu$ final states, respectively.

The data-to-simulation scale factors for the lepton identification and isolation efficiencies are defined using a "tag-and-probe" method with $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-/\mu^+\mu^-$ events. For both electrons and muons with $p_\mathrm{T} > 20$ GeV, they are found to vary between 0.99 and 1.01, depending on the p_T range considered.

7.9 Event yields

The number of data events after each selection cut are in good agreement with SM background expectations, and are shown in Fig. 7.11, left, for the $\mathrm{e}\mu$ final state as a representative example.

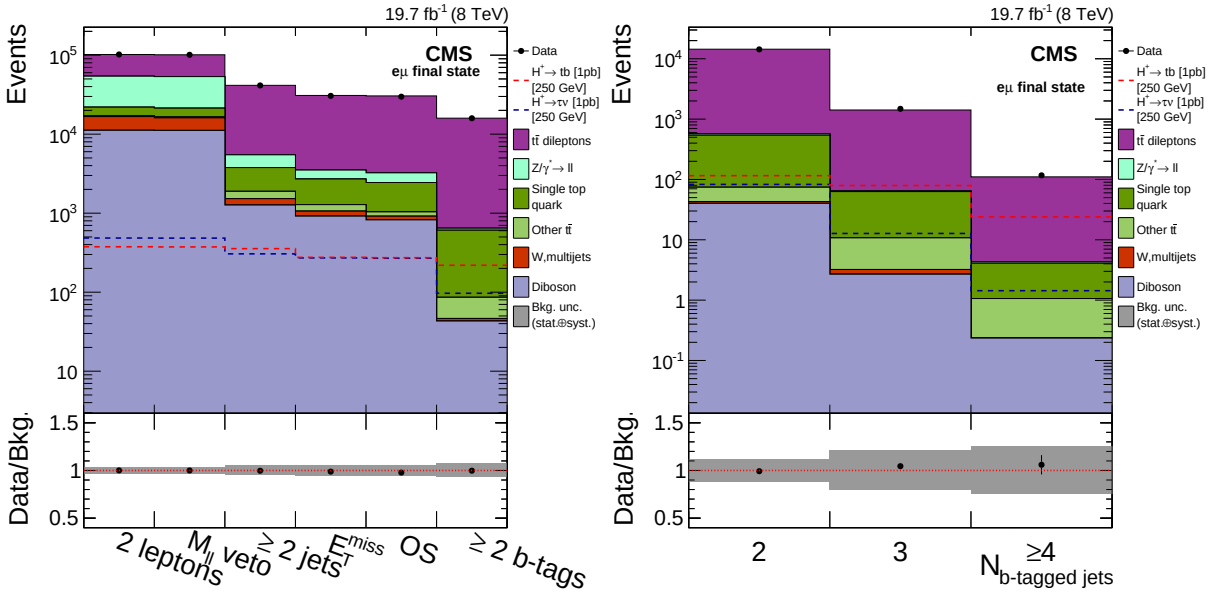


Figure 7.11: The event yields at different selection cut levels (left) and the b-tagged jet multiplicity after the full event selection (right) for the $\mathrm{e}\mu$ final state. The signal yields are normalized to a cross section of 1 pb assuming $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t\bar{b}}) = 1$ and $\mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1$, respectively. The bottom panel shows the ratio of data over the sum of the SM backgrounds with the total uncertainties.

The number of expected events after all selection cuts in the $\ell\ell'$ final state is summarized in Table 7.4 for the SM background processes and for a charged Higgs boson with a mass of $m_{\mathrm{H}^+} = 250\,\text{GeV}$. The main background comes from $\mathrm{t\bar{t}}$ production in the dilepton final state, including all three lepton flavours. Backgrounds from $\mathrm{t\bar{t}}$ production in the final states other than "$\mathrm{t\bar{t}}$ dilepton" (labelled "other $\mathrm{t\bar{t}}$ ") and other SM processes result in significantly smaller yields. Statistical and systematic uncertainties evaluated as described in Section 7.11 are also shown.

The b-tagged jet multiplicity distribution for the $\mathrm{e}\mu$ final state, shown after the full event selection in Fig. 7.11, right, is used to extract limits on charged Higgs boson production.

Table 7.4: Number of expected events for the SM backgrounds and for signal events with a charged Higgs boson mass of $m_{\mathrm{H}^+} = 250\,\text{GeV}$ in the ee, $\mathrm{e}\mu$, and $\mu\mu$ dilepton final states after the final event selection. The signal cross section is normalized to 1 pb, assuming a branching fraction $\mathcal{B} = 100\,\%$ for each decay channel. Event yields are corrected with the trigger and selection efficiencies. Statistical and systematic uncertainties are shown.

Source	ee	$\mathrm{e}\mu$	$\mu\mu$
$\mathrm{H}^+ \to \tau^+\nu_\tau$, $m_{\mathrm{H}^+} = 250\,\text{GeV}$	$39 \pm 3 \pm 3$	$97 \pm 4 \pm 5$	$40 \pm 3 \pm 3$
$\mathrm{H}^+ \to \mathrm{t\bar{b}}$, $m_{\mathrm{H}^+} = 250\,\text{GeV}$	$85 \pm 3 \pm 2$	$219 \pm 5 \pm 5$	$90 \pm 3 \pm 2$
$\mathrm{t\bar{t}}$ dilepton	$5692 \pm 17 \pm 520$	$15296 \pm 28 \pm 1364$	$6332 \pm 18 \pm 572$
other $\mathrm{t\bar{t}}$	$22 \pm 4 \pm 5$	$40 \pm 5 \pm 9$	$17 \pm 3 \pm 5$
$\mathrm{Z}/\gamma^* \to \ell\ell$	$96 \pm 7 \pm 35$	$36 \pm 2 \pm 7$	$139 \pm 10 \pm 42$
W+jets, multijets	$6 \pm 2 \pm 1$	$3 \pm 1 \pm 1$	< 1
single top	$199 \pm 10 \pm 21$	$522 \pm 15 \pm 54$	$228 \pm 10 \pm 26$
dibosons	$15 \pm 1 \pm 2$	$43 \pm 2 \pm 6$	$20 \pm 1 \pm 3$
total SM backgrounds	$6032 \pm 20 \pm 521$	$15941 \pm 32 \pm 1365$	$6736 \pm 23 \pm 575$
data	6162	15902	6955

7.10 Systematics

7.11 Systematic uncertainties

The uncertainties common to the analyses are presented in Section 7.11.1. The uncertainties specific to the individual analyses are discussed in Sec. 7.11.2, 7.11.3 for the dilepton and the muon+τ_{h} final states that constitute the main focus of this manuscript, and in App. B.3.1, A.3.1 for the all-hadronic and single lepton final states that entered the combination for the published result.

7.11.1 Uncertainties common to the analyses

The sources of systematic uncertainties common to the analyses (unless specified otherwise) and affecting simulated samples only are as follows:

- Uncertainties in the lepton trigger, identification, and isolation efficiencies are calculated from independent samples with a "tag-and-probe" method. The uncertainties in the single electron, single muon, and dilepton triggers amount to 2%, 2%, and 3%, respectively. For the τ_{h}+jets final state, the treatment is detailed in Section B.3.1;

- The uncertainty in the efficiency and identification of electrons is 2% (1%) for $p_{\mathrm{T}} > 20$ (30) GeV. For muons, the uncertainty in the efficiency and identification is 1%;

- The uncertainty in τ_{h} identification efficiency is estimated to be 6% [200];

- The misidentification uncertainty in events with an electron misidentified as the τ_{h} is 20% (25%) for the barrel (endcap); for events with a muon (jet) misidentified as the τ_{h} an uncertainty of 30% (20%) is estimated [200];

- The uncertainty in the τ_{h} energy scale (τ_{h} ES) is estimated by varying the τ_{h} momentum by ± 3% [200];

- The uncertainties in the jet energy scale (JES), jet energy resolution (JER), and the contribution to $E_{\mathrm{T}}^{\mathrm{miss}}$ scale from particles not clustered to jets ("unclustered $E_{\mathrm{T}}^{\mathrm{miss}}$ scale") are estimated independently according to the prescription described in Ref. [4], and found to within 1–6% for the signal and dominant simulated backgrounds in all the analyses. The variations of these quantities are also propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$.The uncertainty in JES is evaluated as a function of jet p_{T} and jet η, and takes into account JES variations due to parton flavour;

- The uncertainty arising from b tagging/mistagging efficiencies is estimated according to the description in Ref. [90]. Values of 3–20% are found in the different analyses;

- A 100% uncertainty is assumed for the reweighting of the top quark p_{T} spectrum of each top quark in simulated SM $\mathrm{t\bar{t}}$ events, discussed in Section 3.5. The reweighting and uncertainty depends on the top quark decay [201];

- The uncertainty in pileup event modelling is estimated by varying the total inelastic cross section used to infer the pileup distribution in data by $\pm 5\%$;
- Uncertainties in the theoretical cross section normalization described in detail in Section 3.5;
- For the $\mu\tau_{\mathrm{h}}$, ℓ+jets, and $\ell\ell'$ final states, the uncertainties due to ME and parton shower (PS) matching, and those due to the factorization and renormalization scale choices are applied only to the dominant simulated $\mathrm{t\bar{t}}$ backgrounds; they are estimated by varying by a factor of two the threshold between jet production at the ME level and via PS and by varying by a factor of four the nominal scale given by the momentum transfer of the hard process (Q^2) in the event;
- For the $\mu\tau_{\mathrm{h}}$ and $\ell\ell'$ final states, the uncertainty in the b-tagged jet multiplicity distribution shapes due to PDF variations is estimated separately for the dominant simulated $\mathrm{t\bar{t}}$ backgrounds by varying independently the components of the PDF parameterization;
- For the $\mu\tau_{\mathrm{h}}$ and $\ell\ell'$ final states, the uncertainty due to the modelling of the associated heavy-flavour production ($\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$) is taken into account by assigning to each bin of the b-tagged jet multiplicity distribution of the $\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$ events an uncorrelated bin-by-bin uncertainty of 44%. This uncertainty is based on the comparison between the observed and predicted ratios of $\frac{\sigma(\mathrm{t\bar{t}+b\bar{b}})}{\sigma(\mathrm{t\bar{t}+q\bar{q}})}$ [202];
- The uncertainty in the integrated luminosity is estimated to be 2.6% [78].

7.11.2 The $\mu\tau_{\mathrm{h}}$ final state for $\mathrm{H}^+ \to \tau^+\nu_\tau$ and $\mathrm{H}^+ \to \mathrm{t\bar{b}}$

The dominant sources of systematic uncertainties are the τ_{h} identification and misidentification, the top quark p_T modelling, and the prediction of the $\mathrm{t\bar{t}}$ cross section. In addition to the uncertainties described in Section 7.11.1, an uncertainty associated with the misidentified τ_{h} background estimated from data is evaluated as half of the maximum variation between the "W+jet" and "multijet" estimates discussed in Section 7.4. The statistical uncertainty associated with the number of events in the control region to which the final estimate is applied amounts to 1% and is taken into account in the limit computation.

The systematic uncertainties for the signal and background samples are summarized in Table 7.5. The diboson and Drell–Yan background yields are small compared to the uncertainty on the $\mathrm{t\bar{t}}$ background, and consequently are not used in the limit computation. Results are not sensitive to the inclusion of those backgrounds.

7.11.3 Dilepton ($\mathrm{ee}/\mathrm{e}\mu/\mu\mu$) final states for $\mathrm{H}^+ \to \tau^+\nu_\tau$ and $\mathrm{H}^+ \to \mathrm{t\bar{b}}$

The main sources of systematic uncertainties are the unclustered $E_{\mathrm{T}}^{\mathrm{miss}}$ scale, the b tagging efficiency, and the prediction of the $\mathrm{t\bar{t}}$ cross section.

The systematic uncertainties for signal and background events are summarized in Table 7.6. The diboson, Z/γ^* , "other $\mathrm{t\bar{t}}$", and W+jets backgrounds yields are small compared to the uncertainty on the $\mathrm{t\bar{t}}$ background, and consequently are not used in the limit computation. Results are not sensitive to the inclusion of those backgrounds.

Table 7.5: The systematic uncertainties for the $\mu\tau_{\mathrm{h}}$ final state (in %) for backgrounds, and for signal events from $\mathrm{H}^{+} \to \mathrm{t\bar{b}}$ decays for $m_{\mathrm{H}^{+}} = 250\,\mathrm{GeV}$. These systematic uncertainties are given as the input to the exclusion limit calculation. The uncertainties which depend on the b-tagged jets multiplicity distribution bin are marked with (S) and for them the maximum integrated value of the negative or positive variation is displayed. Empty cells indicate that an uncertainty does not apply to the sample. The uncertainties on the rows are considered to be fully correlated and the uncertainties on the columns lines are considered to be uncorrelated. The correlation for the uncertainties in on the rows marked with (T) are described in detail in the text. The uncertainties on the cross sections are to be considered uncorrelated for different samples and fully correlated for different final states of the same sample (e.g. the different $\mathrm{t\bar{t}}$ decays)

	Signal	$\mathrm{t\bar{t}} \to \mu\tau_{\mathrm{h}} + \mathrm{X}$	$\mathrm{t\bar{t}}$ dilepton	τ_{h} mis-id	single top
τ_{h} identification	6.0	6.0			6.0
Jet $\to \tau_{\mathrm{h}}$ mis-id			20		
$\mathrm{e} \to \tau_{\mathrm{h}}$ mis-id			3.0		
$\mu \to \tau_{\mathrm{h}}$ mis-id			3.0		
τ_{h} energy scale (S)	0.6	2.4	4.4		4.1
Jet energy scale (S)	2.5	1.9	2.6		3.9
Jet energy resolution (S)	0.8	0.1	1.6		0.2
Unclustered $E_{\mathrm{T}}^{\mathrm{miss}}$ energy scale (S)	0.8	0.1	1.8		0.2
b jet tagging (S)	1.8	1.8	2.7		3.2
udsg$\to$b mis-tagging (S)	$<$0.1	$<$0.1	$<$0.1		0.1
Single lepton trigger efficiency	2.0	2.0	2.0		
e identification	2.0	2.0	2.0		2.0
μ identification	1.0	1.0	1.0		1.0
Misidentified τ_{h}				11	
Top quark p_{T} modelling (S)		5.4	5.2		
τ embedding					
PDF effect on shape		shape only	shape only		
Heavy flavours (S)(T)		$<$0.1	$<$0.1		
Matching scale (S)(T)		12	5.1		
Q^2 scale (S)(T)		3.4	7.5		
Cross sections		$^{+2.5}_{-3.4} \pm 4.6$	$^{+2.5}_{-3.4} \pm 4.6$		8.0
Pileup modelling	4.0	2.0	8.0		2.0
Luminosity	2.6	2.6	2.6		2.6

Table 7.6: The systematic uncertainties (in %) for backgrounds, and for signal events from $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ decays for the dilepton channels for a charged Higgs boson mass $m_{\mathrm{H}^+} = 250\,\mathrm{GeV}$. The $\mathrm{e}\mu$ final state is shown as a representative example. These systematic uncertainties are given as the input to the exclusion limit calculation. The uncertainties which depend on the b-tagged jets multiplicity distribution bin are marked with (S) and for them the maximum integrated value of the negative or positive variation is displayed. Empty cells indicate that an uncertainty does not apply to the sample. The uncertainties on the rows are considered to be fully correlated and the uncertainties on the columns are considered to be uncorrelated. The correlation for the uncertainties in on the rows marked with (T) are described in detail in the text. The uncertainties on the cross sections are to be considered uncorrelated for different samples and fully correlated for different final states of the same sample (e.g. the different $\mathrm{t\bar{t}}$ decays)

	Signal	$\mathrm{t\bar{t}}$ dilepton	$\mathrm{Z}/\gamma^* \to \ell\ell$	single top
Jet energy scale (S)	1.4	1.1	1.7	1.4
Jet energy resolution (S)	0.3	0.3	0.4	0.4
Unclustered $E_{\mathrm{T}}^{\mathrm{miss}}$ energy scale (S)	1.3	2.1	11.7	2.6
b jet tagging (S)	2.4	3.7	10	4.3
udsg→b mis-tagging (S)	2.3	3.6	10	4.4
$\mathrm{e}\mu$ trigger efficiency	3.0	3.0	3.0	3.0
Electron identification	2.0	2.0	2.0	2.0
μ identification	1.0	1.0	1.0	1.0
Cross sections		$^{+2.5}_{-3.4} \pm 4.6$	4.0	8.0
Top quark p_{T} modelling (S)		3.8		
PDF shape		shape only		
Heavy flavours (S)(T)		<0.1		
Matching scale (S)(T)		7.7		
Q^2 scale (S)(T)		8.4		
Pileup modelling	0.6	0.4	1.2	1.2
Luminosity	2.6	2.6	2.6	2.6

7.12 Results

A statistical analysis of the m_{T} (Fig. B.2), b-tagged jet multiplicity (Fig. 7.10 (b) and Fig. 7.11 (b)), and H_{T} (Fig. A.2) distributions has been performed using a binned maximum likelihood fit. The data agree with the SM prediction and consequently 95% CL upper limits on charged Higgs boson production are derived using the modified frequentist $\mathrm{CL_s}$ criterion [116, 117] with a test statistic based on the profile likelihood ratio with asymptotic approximation [120, 203].

The systematic uncertainties described in Section 7.11 are incorporated via nuisance parameters following the frequentist paradigm. Correlations between the different sources of systematic uncertainty are taken into account. Uncertainties affecting the shape of the m_{T}, b-tagged jet multiplicity, or H_{T} distributions are represented by nuisance parameters whose variation results in a continuous perturbation of the distribution [204].

7.12.1 Ovierview of the expected contribution of the different final states to the limits on charged Higgs boson production

Each of the different final states described in this analysis has its main contribution to the combined limit in a different regime.

The τ_{h}+jets final state is sensitive only to the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode: any other hypotetical decay mode is included in the overall data driven background estimation.

The $\mu\tau_{\mathrm{h}}$ final state is sensitive mainly to the $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decay mode, but it is much less sensitive than the τ_{h}+jets final state, and it is hence expected to have a very small contribution when combined with it.

The $\ell\ell'$ and ℓ+jets final states are sensitive to both decay modes, but mainly to the $\mathrm{H}^{+} \to \mathrm{t\bar{b}}$ one.

The coupling structure of the MSSM (mutuated by the 2HDM-TypeII) causes different behaviours of the exclusion curves as a function of $tan\beta$

- the coupling to $\tau\nu$ is proportional to $m_{\tau}tan\beta$: hence, $\sigma \times \mathcal{B}$ is monotonic, and an upper limit on $tan\beta$ is set;
- the coupling to $\mathrm{t\bar{b}}$ is proportional to $m_{\mathrm{b}}tan\beta + m_{\mathrm{t}}cot\beta$: hence, $\sigma \times \mathcal{B}$ has a minimum, and both lower and upper limits are placed on $tan\beta$.

At values of $tan\beta$ approaching to 10, the $\mathrm{H}^{+} \to \mathrm{t\bar{b}}$ decay mode is still dominant on the $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decay mode: in particular,

$$\frac{\mathcal{B}(\mathrm{H}^{+} \to \mathrm{t\bar{b}})}{\mathcal{B}(\mathrm{H}^{+} \to \tau\nu_{\tau})} \sim 5, \text{ for } tan\beta > 8 \tag{7.5}$$

Since the $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decay mode sensitivity is still 4-14 times better than the $\mathrm{H}^{+} \to \mathrm{t\bar{b}}$ decay mode sensitivity, the upper limits on $tan\beta$ are still driven by the $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decay mode.

The lower limits, instead, are driven by the $\mathrm{H}^{+} \to \mathrm{t\bar{b}}$ decay mode.

7.12.2 Model-independent limits on charged Higgs boson production ($\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$)

In the analysis of the $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decay mode with the τ_{h}+jets final state no assumption on the charged Higgs boson branching fractions is needed because subtracting the background from "EW+$\mathrm{t\bar{t}}$ with τ_{h}" will remove any potential $\mathrm{H}^{+} \to \mathrm{t\bar{b}}$ and other such signals from data due to the embedding technique described in Section B.2.1. For $m_{\mathrm{H}^{+}} = 80$–160 GeV, the charged Higgs boson is produced most copiously through $\mathrm{t\bar{t}}$ production which can produce one ($\mathrm{t\bar{t}} \to \mathrm{bH}^{+}\mathrm{\bar{b}W}^{-}$) or two charged Higgs bosons ($\mathrm{t\bar{t}} \to \mathrm{bH}^{+}\mathrm{\bar{b}H}^{-}$) if $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}) > 0$. Furthermore, the presence of the charged Higgs boson suppresses the $\mathrm{t\bar{t}} \to \mathrm{bW}^{+}\mathrm{\bar{b}W}^{-}$ yield compared to the SM prediction. Consequently, the number of events in a given bin of the m_{T} distribution depends on the signal strength parameter μ according to:

$$N(\mu) = \mu^2 \times s(\mathrm{H}^{+}\mathrm{H}^{-}) + 2\mu(1-\mu) \times s(\mathrm{H}^{+}\mathrm{W}^{-}) + (1-\mu)^2 \times b(\mathrm{W}^{+}\mathrm{W}^{-}) + b, \tag{7.6}$$

where $\mu = \mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}) \times \mathcal{B}(\mathrm{H}^{+} \to \tau^{+}\nu_{\tau})$, $s(\mathrm{H}^{+}\mathrm{H}^{-})$ and $s(\mathrm{H}^{+}\mathrm{W}^{-})$ are the number of expected signal events for the $\mathrm{t\bar{t}} \to \mathrm{bH}^{+}\mathrm{\bar{b}H}^{-}$ and $\mathrm{t\bar{t}} \to \mathrm{bH}^{+}\mathrm{\bar{b}W}^{-}$ processes, respectively; $b(\mathrm{W}^{+}\mathrm{W}^{-})$ is the expected number of events from the portion of $\mathrm{t\bar{t}} \to \mathrm{bW}^{+}\mathrm{\bar{b}W}^{-}$ background that is estimated with simulation, and b is the expected number of other background events. The number of signal and $\mathrm{t\bar{t}} \to \mathrm{bW}^{+}\mathrm{\bar{b}W}^{-}$ background events is normalized to the SM predicted cross section and by setting $\mathcal{B}(\mathrm{t} \to \mathrm{H}^{+}\mathrm{b}) \times \mathcal{B}(\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}) = 1$ for a top quark decaying to a charged Higgs boson.

For $m_{\mathrm{H}^+} = 180$–600 GeV, the number of events in a given bin of the m_{T} distribution depends on the signal strength parameter according to:

$$N(\mu) = \mu \times \varepsilon_{\mathrm{s}} \mathcal{L} + b, \tag{7.7}$$

where $\mu = \sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$, ε_{s} is the event selection efficiency for signal events, $\mathcal{L}$ is the integrated luminosity, and b is the expected number of background events.

The upper limits on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ and on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ are shown in Fig. 7.12 for the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode with the τ_{h}+jets final state for the ranges $m_{\mathrm{H}^+} = 80$–160 GeV and $m_{\mathrm{H}^+} = 180$–600 GeV, respectively. The numerical values of the limits are given in Table 7.7, which is the reference. At $m_{\mathrm{H}^+} = 250$ GeV an excess of data is observed with a local p-value of 0.046 corresponding to significance of 1.7σ.

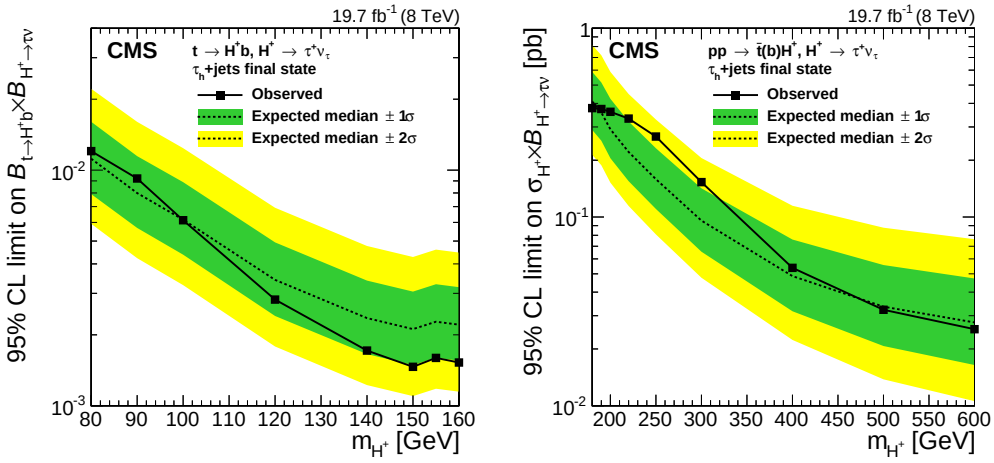


Figure 7.12: Expected and observed 95% CL model-independent upper limits on (a) $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ with $m_{\mathrm{H}^+} = 80$–160 GeV, and (b) on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ with $m_{\mathrm{H}^+} = 180$–600 GeV for the $\mathrm{H}^+ \to \tau^+\nu_\tau$ search in the τ_{h}+jets final state. The regions above the solid lines are excluded.

7.12.3 Limits on charged Higgs boson production with branching fraction assumed

In the presence of a charged Higgs boson and for $m_{\mathrm{H}^+} = 180$–600 GeV, the analyses of the $\mu\tau_{\mathrm{h}}$, ℓ+jets, and $\ell\ell'$ final states have sensitivity to both $\mathrm{H}^+ \to \tau^+\nu_\tau$ and $\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}}$ decays. Consequently, a model-independent limit can neither be provided for $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ nor for $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}})$. Nevertheless, one can test models by fixing $\mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ and $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}})$. In this section, results are reported for a model with $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}}) = 1$, to which the τ_{h}+jets analysis is blind because of the estimates of the backgrounds from data like described in Section B.2.1. For $\mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1$, the sensitivity of the $\mu\tau_{\mathrm{h}}$ and $\ell\ell'$ final states analyses is found to be substantially weaker than that obtained in the τ_{h}+jets analysis.

Table 7.7: Expected and observed 95% CL model-independent upper limits on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ for $m_{\mathrm{H}^+} = 80$–160 GeV (top), and on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ for $m_{\mathrm{H}^+} = 180$–600 GeV (bottom), for the $\mathrm{H}^+ \to \tau^+\nu_\tau$ search in the τ_h+jets final state.

m_{H^+}	Expected limit					Observed
[GeV]	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$	limit
95% CL upper limit on $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$						
80	0.0059	0.0079	0.0112	0.0160	0.0221	0.0120
90	0.0042	0.0057	0.0080	0.0115	0.0160	0.0092
100	0.0033	0.0044	0.0062	0.0089	0.0124	0.0061
120	0.0018	0.0024	0.0034	0.0049	0.0069	0.0028
140	0.0012	0.0017	0.0024	0.0034	0.0048	0.0017
150	0.0011	0.0015	0.0021	0.0031	0.0043	0.0015
155	0.0012	0.0016	0.0023	0.0033	0.0046	0.0016
160	0.0011	0.0016	0.0022	0.0032	0.0045	0.0015
95% CL upper limit on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ [pb]						
180	0.213	0.289	0.409	0.587	0.816	0.377
190	0.188	0.254	0.358	0.516	0.719	0.373
200	0.152	0.205	0.291	0.423	0.587	0.361
220	0.114	0.155	0.221	0.321	0.448	0.332
250	0.081	0.110	0.159	0.231	0.328	0.267
300	0.048	0.065	0.096	0.142	0.205	0.153
400	0.022	0.032	0.049	0.076	0.115	0.054
500	0.014	0.021	0.033	0.056	0.088	0.032
600	0.011	0.016	0.028	0.047	0.076	0.025

Equation (7.7) is used to derive the limits by counting the number of events in bins of the b-tagged jet multiplicity distribution for the $\mu\tau_\mathrm{h}$ and $\ell\ell'$ final states, and in bins of the H_T distribution for the ℓ+jets final state. The upper limits on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+)$ assuming $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}}) = 1$ are shown in Fig. 7.13 for the $\mu\tau_\mathrm{h}$ (a), ℓ+jets (b), and $\ell\ell'$ (c) final states.

The upper limit on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+)$ for the combination of the $\mu\tau_\mathrm{h}$, ℓ+jets, and $\ell\ell'$ final states is shown in Fig. 7.14. The numerical values are reported in Table 7.8. In the combination, the sensitivity is driven by the ℓ+jets final state.

Table 7.8: Expected and observed 95% CL upper limits on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}})$ assuming $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}}) = 1$ for the combination of the $\mu\tau_\mathrm{h}$, ℓ+jets, and $\ell\ell'$ final states.

m_{H^+}	Expected limit [pb]					Observed limit [pb]
[GeV]	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$	limit
95% CL upper limit on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+)$ with $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}}) = 1$						
180	1.07	1.43	2.01	2.81	3.78	1.99
200	0.87	1.16	1.62	2.27	3.07	1.52
220	0.62	0.83	1.16	1.64	2.20	0.99
250	0.49	0.66	0.93	1.31	1.78	0.89
300	0.33	0.45	0.62	0.88	1.18	0.54
400	0.22	0.29	0.40	0.57	0.76	0.33
500	0.15	0.20	0.28	0.39	0.52	0.21
600	0.10	0.14	0.19	0.27	0.36	0.13

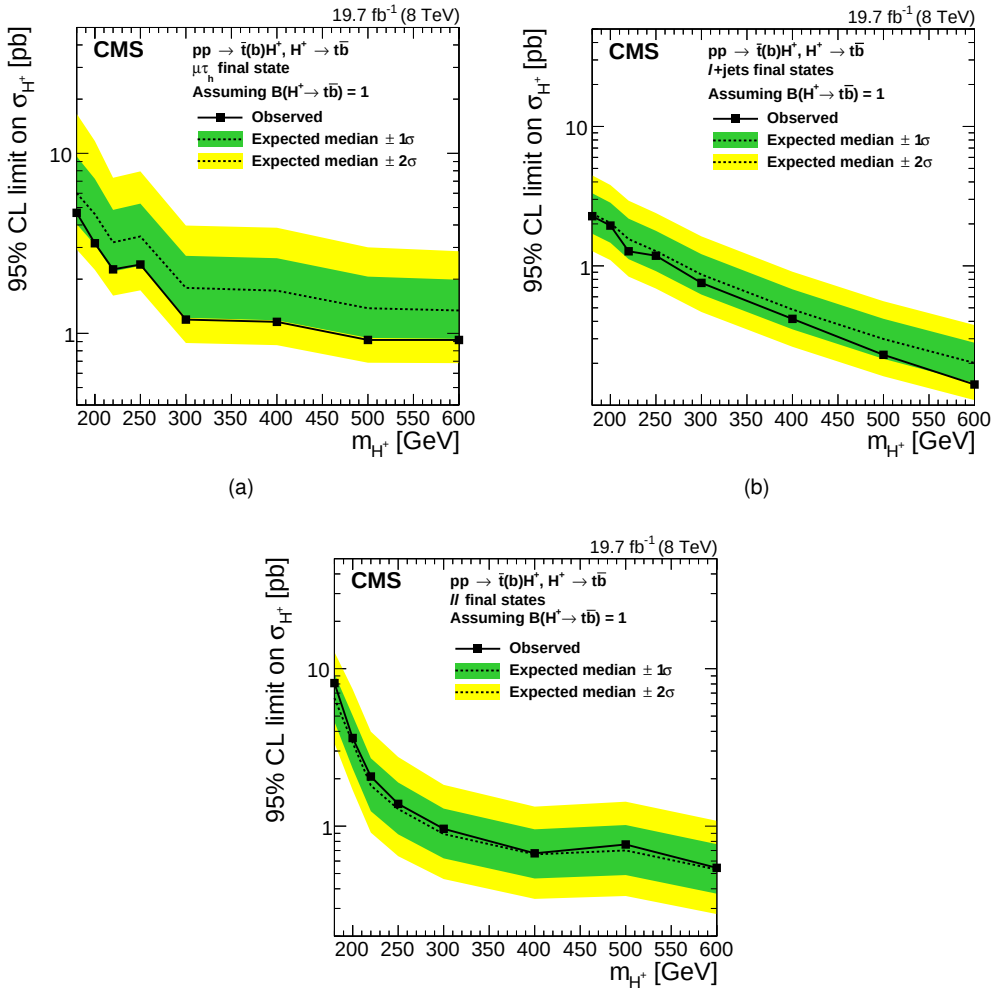


Figure 7.13: Expected and observed 95% CL upper limits on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+)$ for the $\mu\tau_\mathrm{h}$ (a), ℓ+jets (b), and $\ell\ell'$ final states (c) assuming $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}}) = 1$. The regions above the solid lines are excluded.

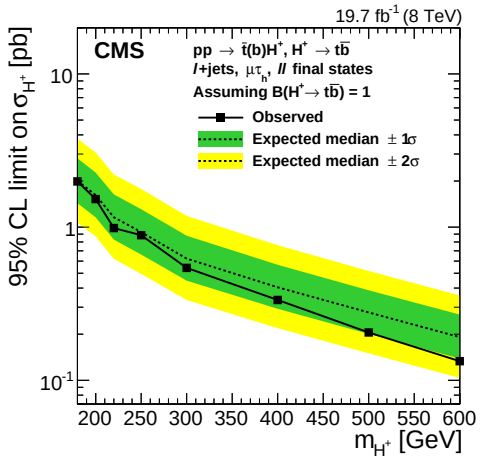


Figure 7.14: Expected and observed 95% CL upper limits on $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+)$ for the combination of the $\mu\tau_\mathrm{h}$, ℓ+jets, and $\ell\ell'$ final states assuming $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}}) = 1$. The region above the solid line is excluded.

7.12.4 Combined limits on $\tan\beta$ in MSSM benchmark scenarios

Using all decay modes and final states, exclusion regions have been set in the m_{H^+}–$\tan\beta$ plane according to the LHC Higgs cross section working group prescription for different MSSM benchmark scenarios [11, 12]: "updated $m_{\mathrm{h}}^{\mathrm{max}}$", "$m_{\mathrm{h}}^{\mathrm{mod+}}$", "$m_{\mathrm{h}}^{\mathrm{mod-}}$", "light stop", "light stau", "tau-phobic", and "low-M_{H}" scenarios. These MSSM benchmark scenarios are compatible with the properties of the recently discovered neutral scalar boson and with the current bounds on supersymmetric particle masses, and they are specified using low-energy MSSM parameters, i.e. no particular soft SUSY-breaking scenario is assumed. The updated $m_{\mathrm{h}}^{\mathrm{max}}$ scenario and $m_{\mathrm{h}}^{\mathrm{mod}}$ scenarios allow the discovered scalar boson to be interpreted as the light CP-even Higgs boson in large parts of the m_{H^+}–$\tan\beta$ plane. The light stop scenario leads to a suppressed rate for the Higgs boson production by gluon fusion, and the light stau scenario enhances the decay rate of the light CP-even Higgs boson to photons. A tau-phobic scenario has suppressed couplings to down-type fermions. In the low-M_{H} scenario, the discovered scalar boson is assumed to be the heavy CP-even Higgs boson and m_{A} is fixed to be 110 GeV causing m_{H^+} to be 132 GeV.

Figure 7.15 shows the limits on the updated $m_{\mathrm{h}}^{\mathrm{max}}$ and $m_{\mathrm{h}}^{\mathrm{mod-}}$ scenarios. For $m_{\mathrm{H}^+} = 90$–160 GeV, the analysis of the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode with the τ_{h}+jets final state described in Section B is taken as input. The mass range starts here from $m_{\mathrm{H}^+} = 90$ GeV, as the lower values of a charged Higgs boson mass are not accessible in the considered MSSM scenarios. For $m_{\mathrm{H}^+} = 200$–600 GeV, a combination of all decay modes and final states is used to set the limits. In this combination, the signal yields from the $\mathrm{H}^+ \to \tau^+\nu_\tau$ and $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ decay modes are defined by the branching fractions predicted by the model. If the limit on the charged Higgs boson production for a given m_{H^+}–$\tan\beta$ point is smaller than the cross section predicted by the model [12, 205, 206, 207], the point is excluded. The mass range is chosen to start from $m_{\mathrm{H}^+} = 200$ GeV to avoid the interference region where a charged Higgs boson is produced both from off-shell top quark decays and through direct production. In all these scenarios except for the low-M_{H} and light stop scenarios, a lower bound of about 155 GeV on the charged Higgs boson mass has been set assuming $m_{\mathrm{h}} = 125 \pm 3$ GeV. The light stop scenario is excluded for $m_{\mathrm{H}^+} < 160$ GeV assuming $m_{\mathrm{h}} = 125 \pm 3$ GeV. For $m_{\mathrm{H}^+} > m_{\mathrm{t}} - m_{\mathrm{b}}$, the $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ decay mode searches yield a lower limit on $\tan\beta$ while the upper limit on $\tan\beta$ is dominated by the results from the analysis of the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode with the τ_{h}+jets final state. The low-M_{H} scenario is completely excluded (Fig. 7.16) assuming the heavy CP-even MSSM Higgs boson mass is $m_{\mathrm{H}} = 125 \pm 3$ GeV.

In Figs. 7.15–7.16, theoretical systematic uncertainties affecting the expected signal event yields are added to the limit computation, modeled as nuisance parameters, in addition to the uncertainties discussed in Section 7.11. The uncertainty in the branching fractions of the charged Higgs boson is estimated from the decay width uncertainties as in Ref. [208] by scaling each partial width separately while fixing all others to their central values. This results in individual theoretical uncertainties for each branching fraction. The width uncertainties comprise the uncertainty from missing higher order corrections to beyond LO EW diagrams (5%), missing higher order corrections to NLO QCD (2%), and Δ_{b}-correction uncertainties (3%) [129]. The Δ_b-correction arises from the presence of squarks and gluino contributions in the charged Higgs boson Yukawa coupling to top and bottom quarks [209, 210].

For $m_{\mathrm{H}^+} = 90$–160 GeV, the theoretical uncertainties in the signal yield include the uncertainties in the branching fractions for $\mathrm{t} \to \mathrm{H}^+\mathrm{b}$ and $\mathrm{H}^+ \to \tau^+\nu_\tau$ totalling 0.1–5.0% depending on m_{H^+} and $\tan\beta$. Additionally, an uncertainty of 3% is added to the simulated $\mathrm{t\bar{t}}$ background to take into account higher order corrections to the $\mathrm{t\bar{t}}$ cross section. For $m_{\mathrm{H}^+} = 200$–600 GeV, the charged Higgs boson production cross section uncertainty and the uncertainty in the branching ratios are considered. The cross section uncertainty varies between 22–32% depending on m_{H^+}, $\tan\beta$, and the MSSM benchmark scenario. The uncertainty in $\mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau)$ varies between 0.4–5.0% for $\tan\beta = 10$–60 depending on m_{H^+} and the MSSM benchmark scenario. The $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t\bar{b}})$ uncertainty varies between 0.1–5.0% for $\tan\beta = 1$–10 depending on m_{H^+} and the MSSM benchmark scenario. The theoretical branching fraction uncertainties for a given m_{H^+}–$\tan\beta$ point are summed linearly according to the LHC Higgs cross section working group prescription [129, 208], but the cross section and branching fraction uncertainties are treated as independent nuisances. The expected limit improves by no more than 2% if the theoretical uncertainties are treated in the statistical model as independent sources.

7.13 Summary

A search is performed for a charged Higgs boson with the CMS detector using a data sample corresponding to an integrated luminosity of $19.7 \pm 0.5\,\mathrm{fb}^{-1}$ in proton-proton collisions at $\sqrt{s} = 8$ TeV. The charged Higgs boson production in $\mathrm{t\bar{t}}$ decays and in $\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+$ is studied assuming $\mathrm{H}^+ \to \tau^+\nu_\tau$ and $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ decay modes, using the τ_h+jets, $\mu\tau_\mathrm{h}$, ℓ+jets, and $\ell\ell'$ final states. Data are found to agree with the SM expectations.

Model-independent limits without an assumption on the charged Higgs boson branching fractions are derived for the $\mathrm{H}^+ \to \tau^+\nu_\tau$ decay mode in the τ_h+jets final state. Upper limits at 95% CL of $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1.2$–0.15% and $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 0.38$–0.025 pb are set for charged Higgs boson mass ranges $m_{\mathrm{H}^+} = 80$–160 GeV and $m_{\mathrm{H}^+} = 180$–600 GeV, respectively.

Assuming $\mathcal{B}(\mathrm{H}^+ \to \mathrm{t\bar{b}}) = 1$, a 95% CL upper limit of $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) = 2.0$–0.13 pb is set for a combination of the $\mu\tau_\mathrm{h}$, ℓ+jets, and $\ell\ell'$ final states for $m_{\mathrm{H}^+} = 180$–600 GeV. This is the first experimental result on the $\mathrm{H}^+ \to \mathrm{t\bar{b}}$ decay mode. Here, cross section $\sigma(pp \to t(b)H^\pm$ stands for the sum $\sigma(pp \to \bar{t}(b)H^+) + \sigma(pp \to t(\bar{b})H^-)$.

The results are interpreted in different MSSM benchmark scenarios and used to set exclusion limits in the m_{H^+}–$\tan\beta$ parameter spaces. In the various models, a lower bound on the charged Higgs boson mass of about 155 GeV is set assuming $m_\mathrm{h} = 125 \pm 3$ GeV. The light-stop scenario is excluded for $m_{\mathrm{H}^+} < 160$ GeV assuming $m_\mathrm{h} = 125 \pm 3$ GeV, and the low-M_H scenario defined in Refs. [11, 12] is completely excluded assuming $m_\mathrm{H} = 125 \pm 3$ GeV.

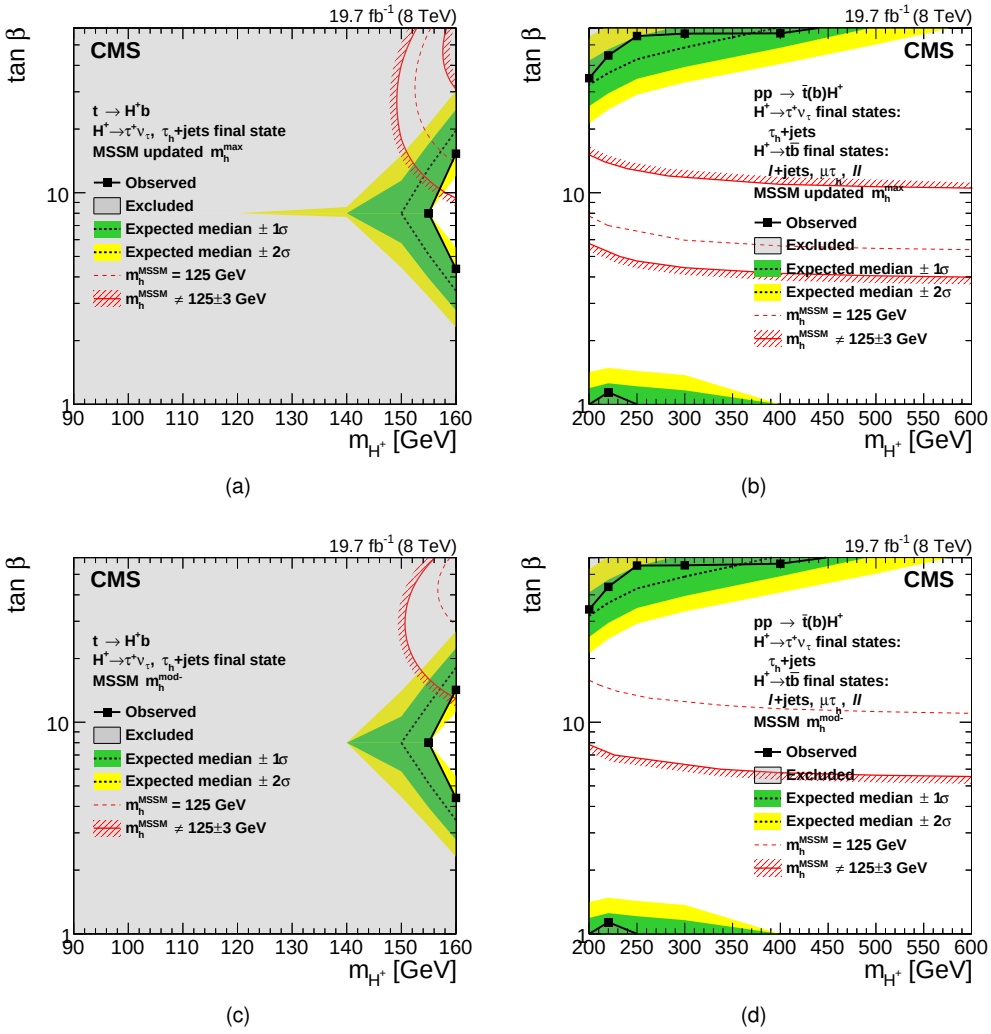


Figure 7.15: Exclusion region in the MSSM m_{H^+}–$\tan\beta$ parameter space for (a, c) $m_{\mathrm{H}^+} = 80$–160 GeV and for (b, d) $m_{\mathrm{H}^+} = 180$–600 GeV in the (a, b) updated MSSM $m_{\mathrm{h}}^{\mathrm{max}}$ scenario and (c, d) $m_{\mathrm{h}}^{\mathrm{mod-}}$ scenarios [11, 12]. In (a) and (c) the limit is derived from the $\mathrm{H}^+ \to \tau^+\nu_\tau$ search with the τ_{h}+jets final state, and in (b) and (d) the limit is derived from a combination of all the charged Higgs boson decay modes and final states considered. The $\pm1\sigma$ and $\pm2\sigma$ bands around the expected limit are also shown. The light-grey region is excluded. The red lines depict the allowed parameter space for the assumption that the discovered scalar boson is the lightest CP-even MSSM Higgs boson with a mass $m_{\mathrm{h}} = 125 \pm 3$ GeV, where the uncertainty is the theoretical uncertainty in the Higgs boson mass calculation.

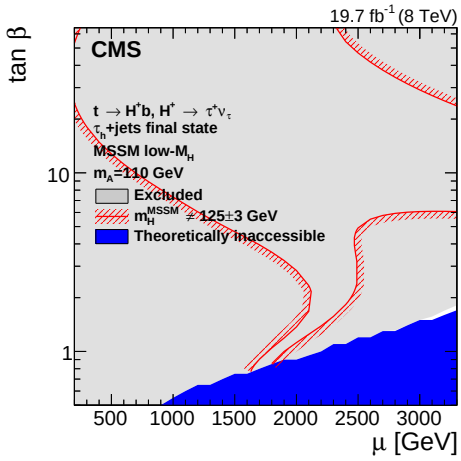


Figure 7.16: Exclusion region in the MSSM Higgsino mass parameter (μ) vs. $\tan\beta$ parameter space in the low-M_{H} scenario [11, 12] with $m_{\mathrm{A}} = 110$ GeV for the $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ search with the τ_{h}+jets final state. The light-grey region is excluded and the blue region is theoretically inaccessible. The area inside the red lines is the allowed parameter space for the assumption that the discovered scalar boson is the heavy CP-even MSSM Higgs boson with a mass $m_{\mathrm{H}} = 125 \pm 3$ GeV, where the uncertainty is the theoretical uncertainty in the Higgs boson mass calculation.

Chapter 8

Conclusions

Please let me rephrase.
– Michele Gallinaro

8.1 Achievements

This manuscript collected work performed by its author across his career as a PhD student, spanning both the 7 TeV and the 8 TeV LHC runs.

The top quark mass analysis, in the context of which the author had his initial training, has been – at the moment of publications and for some months thereafter – the most precise measurement of the top quark mass in the dilepton final state, and as such has been presented by the author, on behalf of the CMS collaboration, at the poster session of the March 2012 LHCC [211].

The various charged Higgs results have been presented by the author, on behalf of the CMS collaboration, at the 2012 and 2014 sessions of the *Prospects For Charged Higgs Discovery at Colliders – c$H^{\pm}$arged* workshop [212, 213].

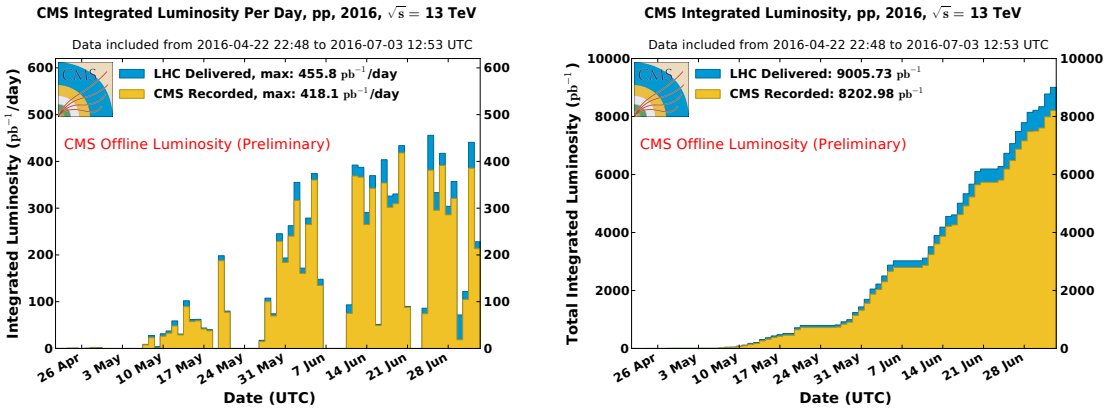
The results of the search for a heavy charged Higgs boson are the first direct search for a charged Higgs decaying into a top quark and a bottom quark ever published, and have been presented by the author, on behalf of the CMS collaboration, at the forementioned c$H^{\pm}$arged-2014 workshop and at the 2015 LHCP conference [214], in St. Petersburg (in this last conference, the author has also spoken on behalf of the ATLAS collaboration and present their results for other charged Higgs decay modes).

8.2 Future Work

The LHC started the first proton-proton collisions at a center-of-mass energy of 13 TeV at the end of 2015, collecting a few hundreds inverse picobarn of data with a bunch spacing of 50 ns, and about 2.3 fb^{-1} with a bunch spacing of 25 ns, the target bunch spacing for the LHC Run2. The maximum peak instantaneous luminosity achieved in the 2015 run has been of about 5.13 Hz/nb [13].

A few weeks before the submission of this thesis, collisions started again after the winter stop, and

the calibrations of the beginning of the year. So far (July 5th, 2016), the CMS detector has collected about 8 fb^{-1} of data, as shown in Fig. 8.1, out of which about 5 are already certified as "good runs for physics analyses". So far, the maximum peak instantaneous luminosity achieved in the 2016 run has been of 9.43 Hz/nb [13]. For comparison, the maximum peak instantaneous luminosity achieved in the 8 TeV run has been of 0.132 Hz/nb [13], that is a factor ~ 100 smaller.



(a) (b)

Figure 8.1: Offline luminosity versus day delivered to (blue), and recorded by CMS (orange) during stable beams and for p-p collisions at 13 TeV centre-of-mass energy in 2016 (left). The corresponding cumulative distribution is also shown (right), as well as the corresponding non cumulative distribution (right). The delivered luminosity accounts for the luminosity delivered from the start of stable beams until the LHC requests CMS to turn off the sensitive detectors to allow a beam dump or beam studies. Given is the luminosity as determined from counting rates measured by the luminosity detectors after offline validation. The figures are taken from Ref. [13].

The cross sections for a MSSM heavy charged Higgs boson production at 13 TeV are about 8 times higher than at 8 TeV, depending on the charged Higgs boson mass. Correspondingly, the main background for the dilepton final state, i.e. the SM $\mathrm{t\bar{t}}$ production, has a cross section increase of only a factor 3. This results in an increase in the signal-to-background ratioof between 100% and 400%, depending on the charged Higgs boson mass.

The about 100 times higher instantaneous luminosity in the 13 TeV run with respect to the 8 TeV run provides a large statistics, slightly less than 0.5 fb^{-1} of data per day. These conditions are ideal for improving the strategy of the search for a charged Higgs boson in the dilepton final state by exploiting an advanced multivariate analysis to increase the separation between the main SM $\mathrm{t\bar{t}}$ background and the charged Higgs signal.

The author has organized, in August 2015, a two-week summer stage for students of the Instituto Superior Técnico [215], which has been attended to by the excellent first-year students João Barata and Laura Lisboa. The stage consisted in a basic multi-variate analysis aiming to separate the SM $\mathrm{t\bar{t}}$ background from the charged Higgs signal in the dilepton final state, using simulated events at $\sqrt{s} = 8$ TeV, comparing a BDT method with a Neural Network approach. The results have been shown at an internal CMS meeting, and consist in an improvement of the sensitivity of the analysis with respect to a cut-and-count approach.

Such an approach, that from a statistical point of view consists in separating the events into many different categories (bins), requires a number of simulated signal and background events far larger than the one available at 8 TeV, in order to have a reasonable amount of events in each bin for the purposes of hypothesis testing: the production of such high statistics samples, both for the SM $\mathrm{t\bar{t}}$ background and the charged Higgs signal, is already taking place.

On the basis of these promising preliminary studies, the author is pursuing the search for a charged Higgs boson in the dilepton final state, using advanced multivariate analysis methods.

Bibliography

[1] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, and João P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1, 2012. doi: 10.1016/j.physrep.2012.02.002.

[2] Search for a light charged Higgs boson in top quark decays in pp collisions at $\sqrt{s}$ = 7 TeV. *JHEP*, 07:143, 2012. doi: 10.1007/JHEP07(2012)143.

[3] R. B. Palmer and J. C. Gallardo. High-energy colliders. In *Critical problems in physics. Proceedings, Conference celebrating the 250th Anniversary of Princeton University, Princeton, USA, October 31-November 2, 1996*, pages 247–269, 1997.

[4] S. Chatrchyan et al. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:P11002, 2011. doi: 10.1088/1748-0221/6/11/P11002.

[5] 8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb^{-1} of data in CMS. Oct 2013. URL `https://cds.cern.ch/record/1627305`.

[6] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8: P04013, 2013. doi: 10.1088/1748-0221/8/04/P04013.

[7] Vardan Khachatryan et al. Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS. *JINST*, 11(01):P01019, 2016. doi: 10.1088/1748-0221/11/01/P01019.

[8] J. Friedman T. Hastie, R. Tibshirani. *The Elements of Statistical Learning*. Springer-Verlag, 2009, 10th printing 2013.

[9] V. M. Abazov et al. D0 note 5175-conf.

[10] S. Chatrchyan et al. Performance of quark/gluon discrimination in 8 tev pp data. Technical report. URL `http://cdsweb.cern.ch/record/1599732`.

[11] M. Carena, S. Heinemeyer, o. Stål, C.E.M. Wagner, and G. Weiglein. MSSM Higgs Boson Searches at the LHC: Benchmark Scenarios after the Discovery of a Higgs-like Particle. *Eur. Phys. J. C*, 73:2552, 2013. doi: 10.1140/epjc/s10052-013-2552-1.

[12] S Heinemeyer et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 2013. doi: 10.5170/CERN-2013-004.

[13] Cms online luminosity. *https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults*.

[14] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012. doi: 10.1016/j.physletb.2012.08.020.

[15] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012. doi: 10.1016/j.physletb.2012.08.021.

[16] Georges Aad et al. Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC. *Phys. Lett. B*, 726:88, 2013. doi: 10.1016/j.physletb.2013.08.010.

[17] Georges Aad et al. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Phys.Lett.*, B726:120–144, 2013. doi: 10.1016/j.physletb.2013.08.026.

[18] Serguei Chatrchyan et al. Observation of a new boson with mass near 125 GeV in pp collision at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 06:081, 2013. doi: 10.1007/JHEP06(2013)081.

[19] Serguei Chatrchyan et al. Study of the Mass and Spin-Parity of the Higgs Boson Candidate via its Decay to Z Boson Pairs. *Phys. Rev. Lett.*, 110:081803, 2013. doi: 10.1103/PhysRevLett.110.081803.

[20] Serguei Chatrchyan et al. Measurement of the properties of a Higgs boson in the four-lepton final state. *Phys. Rev. D*, 88:081803, 2014. doi: 10.1103/PhysRevD.89.092007.

[21] Vardan Khachatryan et al. Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs. *Phys.Lett.*, B736:64, 2014. doi: 10.1016/j.physletb.2014.06.077.

[22] Vardan Khachatryan et al. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. Submitted to *Eur. Phys. J. C*, 2014.

[23] T.D. Lee. A Theory of Spontaneous T Violation. *Phys.Rev.*, D8:1226–1239, 1973. doi: 10.1103/PhysRevD.8.1226.

[24] Pierre Fayet. Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino. *Nucl. Phys. B*, 90:104, 1975. doi: 10.1016/0550-3213(75)90636-7.

[25] Pierre Fayet. Supersymmetry and weak, electromagnetic and strong interactions. *Phys. Lett. B*, 64:159, 1976. doi: 10.1016/0370-2693(76)90319-1.

[26] Pierre Fayet. Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions. *Phys. Lett. B*, 69:489, 1977. doi: 10.1016/0370-2693(77)90852-8.

[27] Savas Dimopoulos and Howard Georgi. Softly broken supersymmetry and SU(5). *Nucl. Phys. B*, 193:150, 1981. doi: 10.1016/0550-3213(81)90522-8.

[28] N. Sakai. Naturalness in supersymmetric GUTS. *Z. Phys. C*, 11:153, 1981. doi: 10.1007/BF01573998.

[29] Kenzo Inoue, Akira Kakuto, Hiromasa Komatsu, and Seiichiro Takeshita. Low-Energy Parameters and Particle Masses in a Supersymmetric Grand Unified Model. *Prog. Theor. Phys.*, 67:1889, 1982. doi: 10.1143/PTP.67.1889.

[30] Kenzo Inoue, Akira Kakuto, Hiromasa Komatsu, and Seiichiro Takeshita. Aspects of Grand Unified Models with Softly Broken Supersymmetry. *Prog. Theor. Phys.*, 68:927, 1982. doi: 10.1143/PTP.68.927.

[31] Kenzo Inoue, Akira Kakuto, Hiromasa Komatsu, and Seiichiro Takeshita. Renormalization of Supersymmetry Breaking Parameters Revisited. *Prog. Theor. Phys.*, 71:413, 1984. doi: 10.1143/PTP.71.413.

[32] S. Heinemeyer, W. Hollik, and G. Weiglein. FeynHiggs: A Program for the calculation of the masses of the neutral CP even Higgs bosons in the MSSM. *Comput. Phys. Commun.*, 124:76, 2000. doi: 10.1016/S0010-4655(99)00364-1.

[33] Georges Aad et al. Search for charged Higgs bosons decaying via $H^{\pm} \to \tau^{\pm}\nu$ in fully hadronic final states using pp collision data at $\sqrt{s}$ = 8 TeV with the ATLAS detector. *Submitted to JHEP*, 2014.

[34] Georges Aad et al. Search for charged Higgs bosons decaying via $H^{+} \to \tau\nu$ in top quark pair events using pp collision data at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 1206:039, 2012.

[35] Georges Aad et al. Search for charged Higgs bosons through the violation of lepton universality in $\mathrm{t\bar{t}}$ events using pp collision data at $\sqrt{s} = 7$ TeV with the ATLAS experiment. *JHEP*, 03:076, 2013. doi: 10.1007/JHEP03(2013)076.

[36] Georges Aad et al. Search for a light charged Higgs boson in the decay channel $H^{+} \to c\bar{s}$ in $t\bar{t}$ events using pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector. *Eur. Phys. J. C*, 73:2465, 2013. doi: 10.1140/epjc/s10052-013-2465-z.

[37] A. Heister et al. Search for charged Higgs bosons in $\mathrm{e^{+}e^{-}}$ collisions at energies up to $\sqrt{s}$ = 209 GeV. *Phys. Lett. B*, 543:1, 2002. doi: 10.1016/S0370-2693(02)02380-8.

[38] J. Abdallah et al. Search for charged Higgs bosons at LEP in general two Higgs doublet models. *Eur. Phys. J. C*, 34:399, 2004. doi: 10.1140/epjc/s2004-01732-6.

[39] P. Achard et al. Search for charged Higgs bosons at LEP. *Phys. Lett. B*, 575:208, 2003. doi: 10.1016/j.physletb.2003.09.057.

[40] G. Abbiendi et al. Search for charged Higgs bosons in $\mathrm{e^{+}e^{-}}$ collisions at $\sqrt{s} = 189 - 209$ GeV. *Eur. Phys. J. C*, 72:2076, 2012. doi: 10.1140/epjc/s10052-012-2076-0.

[41] Georges Aad et al. Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b -tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector. *Eur. Phys. J.*, C74(10):3109, 2014. doi: 10.1140/epjc/s10052-014-3109-7.

[42] Serguei Chatrchyan et al. Measurement of the $\mathrm{t\bar{t}}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8\,\mathrm{TeV}$. *JHEP*, 02:024, 2014. doi: 10.1007/JHEP02(2014)024.

[43] Serguei Chatrchyan et al. Measurement of the top-quark mass in $\mathrm{t\bar{t}}$ events with dilepton final states in pp collisions at $\sqrt{s} = 7\,\mathrm{TeV}$. *Eur. Phys. J. C*, 72:2202, 2012. doi: 10.1140/epjc/s10052-012-2202-z.

[44] 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.

[45] F. Abe et al. The $\mu\tau$ and $e\tau$ decays of top quark pairs produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. Lett.*, 79:3585, 1997. doi: 10.1103/PhysRevLett.79.3585.

[46] A. Abulencia et al. A search for $t \to \tau\nu q$ in $\mathrm{t\bar{t}}$ production. *Phys. Lett. B*, 639:172, 2006. doi: 10.1016/j.physletb.2006.06.030.

[47] V. M. Abazov et al. Measurement of the $\mathrm{t\bar{t}}$ production cross section and top quark mass extraction using dilepton events in $\mathrm{p\bar{p}}$ collisions. *Phys. Lett. B*, 679:177, 2009. doi: 10.1016/j.physletb.2009.07.032.

[48] Timo Antero Aaltonen et al. Study of top quark production and decays involving a tau lepton at CDF and limits on a charged Higgs boson contribution. *Phys. Rev. D*, 89:091101, 2014. doi: 10.1103/PhysRevD.89.091101.

[49] Georges Aad et al. Measurement of the top quark pair cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV using final states with an electron or a muon and a hadronically decaying τ lepton. *Phys. Lett. B*, 717:89, 2012. doi: 10.1016/j.physletb.2012.09.032.

[50] CMS Collaboration. Search for a heavy charged higgs boson in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ with the cms detector. CMS Physics Analysis Summary CMS-PAS-HIG-13-026, 2014.

[51] Vardan Khachatryan et al. Search for a charged Higgs boson in pp collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 11:018, 2015. doi: 10.1007/JHEP11(2015)018.

[52] M. J. M. Pimenta A. De Angelis. *Introduction to Particle and Astroparticle Physics*. Springer, 2015.

[53] A. Pascolini. *Metodi Matematici della Fisica*. Edizioni Libreria Progetto, Padova, 2003.

[54] The Nobel Foundation, editor. *Nobel Lectures, Physics 1942-1962*. Elsevier Publishing Company, 1964.

[55] M. Herrero. The Standard model. *NATO Sci. Ser. C*, 534:1–59, 1999. doi: 10.1007/978-94-011-4689-0_1.

[56] K Nakamura and Particle Data Group. Review of particle physics. *Journal of Physics G: Nuclear and Particle Physics*, 37(7A):075021, 2010. URL http://stacks.iop.org/0954-3899/37/i=7A/a=075021.

[57] Luca Fabbri and Stefano Vignolo. A torsional completion of gravity for Dirac matter fields and its applications to neutrino oscillations. *Mod. Phys. Lett.*, A31(03):1650014, 2016. doi: 10.1142/S0217732316500140.

[58] Abdelhak Djouadi. The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model. *Phys. Rept.*, 457:1–216, 2008. doi: 10.1016/j.physrep.2007.10.004.

[59] P. M. Ferreira, Renato Guedes, John F. Gunion, Howard E. Haber, Marco O. P. Sampaio, and Rui Santos. The CP-conserving 2HDM after the 8 TeV run. In *Proceedings, 22nd International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2014)*, 2014. URL https://inspirehep.net/record/1306634/files/arXiv:1407.4396.pdf.

[60] Hitoshi Murayama. Supersymmetry phenomenology. In *Particle physics. Proceedings, Summer School, Trieste, Italy, June 21-July 9, 1999*, pages 296–335, 2000. URL http://alice.cern.ch/format/showfull?sysnb=2177492.

[61] A. Djouadi, L. Maiani, G. Moreau, A. Polosa, J. Quevillon, and V. Riquer. The post-Higgs MSSM scenario: Habemus MSSM? *Eur. Phys. J.*, C73:2650, 2013. doi: 10.1140/epjc/s10052-013-2650-0.

[62] G. et al. Aad. Search for a multi-higgs-boson cascade in $W^+W^-b\bar{b}$ events with the atlas detector in pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Rev. D*, 89:032002, Feb 2014. doi: 10.1103/PhysRevD.89.032002. URL http://link.aps.org/doi/10.1103/PhysRevD.89.032002.

[63] G. Aad et al. Search for a charged higgs boson produced in the vector-boson fusion mode with decay $H^{\pm} \to W^{\pm}z$ using pp collisions at $\sqrt{s} = 8$ TeV with the atlas experiment. *Phys. Rev. Lett.*, 114:231801, Jun 2015. doi: 10.1103/PhysRevLett.114.231801. URL http://link.aps.org/doi/10.1103/PhysRevLett.114.231801.

[64] Howard Georgi and Marie Machacek. Doubly charged higgs bosons. *Nuclear Physics B*, 262(3): 463 – 477, 1985. ISSN 0550-3213. doi: http://dx.doi.org/10.1016/0550-3213(85)90325-6. URL http://www.sciencedirect.com/science/article/pii/0550321385903256.

[65] Kingman Cheung and Dilip Kumar Ghosh. Triplet higgs boson at hadron colliders. *Journal of High Energy Physics*, 2002(11):048, 2002. URL http://stacks.iop.org/1126-6708/2002/i=11/a=048.

[66] D. P. Roy. The hadronic tau decay signature of a heavy charged Higgs boson at LHC. *Phys. Lett. B*, 459:607, 1999. doi: 10.1016/S0370-2693(99)00724-8.

[67] D. P. Roy. Looking for the charged Higgs boson. *Mod. Phys. Lett.*, A19:1813–1828, 2004. doi: 10.1142/S0217732304015105.

[68] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008. URL http://stacks.iop.org/1748-0221/3/i=08/a=S08001.

[69] X. Zhen et al. The official cms luminosity calculation. *CMS twiki*, https://twiki.cern.ch/twiki/bin/view/CMS/LumiCalc, 2011.

[70] Luminosity calibration update with the full 2011 dataset. *CMS Physics Analysis Summary*, CMS PAS SMP-12-008, 2011. URL http://cdsweb.cern.ch/record/.

[71] CMS Collaboration. Luminosity calibration for the 2013 proton-lead and proton-proton data taking. CMS Physics Analysis Summary CMS-PAS-LUM-13-002, 2013. URL http://cdsweb.cern.ch/record/1643269.

[72] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 03:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.

[73] The CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006. URL https://cds.cern.ch/record/922757.

[74] *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. URL https://cds.cern.ch/record/343814.

[75] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006. doi: 10.1088/1126-6708/2006/05/026.

[76] Howard Baer, Frank E. Paige, Serban D. Protopescu, and Xerxes Tata. Isajet 7.69: A monte carlo event generator for pp, $\bar{p}p$, and e^+e^- reactions. 2003.

[77] Veronique Lefebure, Sudeshna Banerjee, and I. Gonzalez. CMS Simulation Software Using Geant4. CERN-CMS-NOTE-1999-072, 1999.

[78] CMS. Cms luminosity based on pixel cluster counting – summer 2013 update. CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013. URL https://cdsweb.cern.ch/record/1598864/.

[79] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[80] CMS. Particle–flow event reconstruction in CMS and performance for jets, taus, and $E_{\mathrm{T}}^{\mathrm{miss}}$. CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009. URL http://cdsweb.cern.ch/record/1194487.

[81] CMS. Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV. CMS Physics Analysis Summary CMS-PAS-PFT-10-002, 2010. URL http://cdsweb.cern.ch/record/1279341.

[82] CMS Collaboration. Commissioning of the particle flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV. CMS Physics Analysis Summary CMS-PAS-PFT-10-002, 2010. URL http://cdsweb.cern.ch/record/1279341.

[83] Vardan Khachatryan et al. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}$ = 8 TeV. 2015.

[84] Serguei Chatrchyan et al. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012. doi: 10.1088/1748-0221/7/10/P10002.

[85] 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.

[86] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J. C*, 72: 1896, 2012. doi: 10.1140/epjc/s10052-012-1896-2.

[87] K. Kousouris et al. Jet energy corrections. *CMS twiki*, https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookJetEnergyCorrections, 2011.

[88] CMS. Performance of charge hadron subtraction and puppi on jets and jet substructure. CMS Physics Analysis Summary CMS-PAS-JME-14-001, 2014. URL http://cdsweb.cern.ch/record/1751454.

[89] CMS Collaboration. Pileup jet identification. CMS Physics Analysis Summary CMS-PAS-JME-13-005, 2013. URL http://cdsweb.cern.ch/record/1581583.

[90] CMS Collaboration. Performance of b tagging at $\sqrt{s} = 8$ tev in multijet, $\mathrm{t\bar{t}}$ and boosted topology events. CMS Physics Analysis Summary CMS-PAS-BTV-13-001, 2013. URL http://cdsweb.cern.ch/record/1581306.

[91] Performance of τ-lepton reconstruction and identification in CMS. *JINST*, 07:P01001, 2012. doi: 10.1088/1748-0221/7/01/P01001.

[92] Fabio Maltoni and Tim Stelzer. MadEvent: Automatic event generation with MadGraph. *JHEP*, 02: 027, 2003. doi: 10.1088/1126-6708/2003/02/027.

[93] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: going beyond. *JHEP*, 06:128, 2011. doi: 10.1007/JHEP06(2011)128.

[94] 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. *J. High Energy Phys.*, 07 (arXiv:1405.0301. CERN-PH-TH-2014-064. CP3-14-18. LPN14-066. MCNET-14-09. ZU-TH 14-14):079. 158 p, May 2014. URL https://cds.cern.ch/record/1699128. Comments: 158 pages, 27 figures; a few references have been added.

[95] J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni, M. L. Mangano, M. Moretti, C. G. Papadopoulos, F. Piccinini, S. Schumann, M. Treccani, J. Winter, and M. Worek. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C*, 53:473, 2008. doi: 10.1140/epjc/s10052-007-0490-5.

[96] P. Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. doi: 10.1088/1126-6708/2004/11/040.

[97] S. Frixione, P. Nason, and C. Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. doi: 10.1088/1126-6708/2007/11/070.

[98] S. Alioli, P. Nason, C. Oleari, and E. Re. A general framework for implementing NLO calculations in shower monte carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010. doi: 10.1007/JHEP06(2010)043.

[99] S. Alioli, P. Nason, C. Oleari, and E. Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 09:111, 2009. doi: 10.1088/1126-6708/2009/09/111.

[100] E. Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J. C*, 71(2):1547, 2011. doi: 10.1140/epjc/s10052-011-1547-z.

[101] Z. Was. TAUOLA the library for tau lepton decay, and KKMC / KORALB / KORALZ status report. *Nucl. Phys. Proc. Suppl.*, 98:96, 2001. doi: 10.1016/S0920-5632(01)01200-2.

[102] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[103] John Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006. doi: 10.1109/TNS.2006.869826.

[104] Rick Field. Early LHC Underlying Event Data – Findings and Surprises. 2010.

[105] Pavel M. Nadolsky, Hung-Liang Lai, Qing-Hong Cao, Joey Huston, Jon Pumplin, Daniel Stump, Wu-Ki Tung, and C.-P. Yuan. Implications of CTEQ global analysis for collider observables. *Phys. Rev. D*, 78:013004, 2008. doi: 10.1103/PhysRevD.78.013004.

[106] Tune comparisons: deviation metrics per gen/tune and observable group. *http://rivet.hepforge.org/tunecmp/index.html*.

[107] Serguei Chatrchyan et al. Measurement of the underlying event activity at the LHC with $\sqrt{s} = 7$ TeV and comparison with $\sqrt{s} = 0.9$ TeV. *JHEP*, 09:109, 2011. doi: 10.1007/JHEP09(2011)109.

[108] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975. doi: 10.1016/0010-4655(75)90039-9.

[109] G. D'Agostini. *Bayesian Reasoning in Data Analysis*. World Scientific, 2003.

[110] F.C. Porter I. Narsky. *Statistical Analysis Techniques in Particle Physics*. Wiley-VCH, 2013.

[111] F. James. *Statistical Methods in Experimental Physics*. World Scientific, 2006 (2nd edition).

[112] Gary J. Feldman and Robert D. Cousins. Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57:3873–3889, Apr 1998. doi: 10.1103/PhysRevD.57.3873. URL http://link.aps.org/doi/10.1103/PhysRevD.57.3873.

[113] J. Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 236(767):333–380, 1937. ISSN 0080-4614. doi: 10.1098/rsta.1937.0005. URL http://rsta.royalsocietypublishing.org/content/236/767/333.

[114] A. Hoecker et al. Tmva - toolkit for multivariate data analysis. *ArXiv Physics e-prints*, 2007.

[115] T. Dorigo et al. S. Amerio. Narrowing the higgs mass peak: the hyperball algorithm. *CDF experiment internal note*, 2003.

[116] L. Read. Presentation of search results: the CL_s technique. *J. Phys. G*, 28:2693, 2002. doi: doi:10.1088/0954-3899/28/10/313.

[117] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999. doi: 10.1016/S0168-9002(99)00498-2.

[118] R. J. Barlow and C. Beeston. Fitting using finite Monte Carlo samples. *Comput. Phys. Commun,*, 77:219, 1993. doi: 10.1016/0010-4655(93)90005-W.

[119] J. S. Conway. Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. In *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, pages 115–120, 2011. doi: 10.5170/CERN-2011-006.115. URL https://inspirehep.net/record/891252/files/arXiv:1103.0354.pdf.

[120] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. doi: 10.1140/epjc/s10052-011-1554-0.

[121] Reference page for the higgs combination tool software package. *https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit*.

[122] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical Report CERN-OPEN-2012-016, New York U., New York, Jan 2012. URL https://cds.cern.ch/record/1456844.

[123] and others. Combination of CDF and D0 Results on the Mass of the Top Quark. 2010.

[124] Serguei Chatrchyan et al. Measurement of the $\mathrm{t\bar{t}}$ production cross section and the top quark mass in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 07:049, 2011. doi: 10.1007/JHEP07(2011)049.

[125] *CMS-PAS-TOP-10-007*, 2011.

[126] A. Abulencia et al. Measurement of the top quark mass using template methods on dilepton events in proton antiproton collisions at $\sqrt{s}$ = 1.96-TeV. *Phys. Rev.*, D73:112006, 2006. doi: 10.1103/PhysRevD.73.112006.

[127] R. Chierici et al. Standard model cross sections for cms at 7 tev. *CMS twiki*, https://twiki.cern.ch/twiki/bin/view/CMS/StandardModelCrossSections, 2010.

[128] J. Wagner et al. Single top cross sections (summer 2011 analyses). *CMS twiki*, https://twiki.cern.ch/twiki/bin/view/CMS/SingleTopSigma, 2011.

[129] S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. CERN Report CERN-2012-002, 2012. URL `http://cdsweb.cern.ch/record/1416519`.

[130] John M. Campbell, R.Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 1107:018, 2011. doi: 10.1007/JHEP07(2011)018.

[131] Search for single top tw associated production in the dilepton decay channel in pp collisions at sqrt s =7 tev. *CMS-PAS-TOP-11-022*, 2011.

[132] Measurement of the ww, wz and zz cross sections at cms. *CMS Physics Analysis Summary*, CMS PAS EWK-11-010, 2011. URL `http://cdsweb.cern.ch/record/1370067`.

[133] J. Andrea et al. Top pag reference selections. *CMS twiki*, https://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopRefEventSel, 2011.

[134] P. Silva, M Gallinaro, J. Varela, P. Vischia, and A. Alves. Probing the heavy flavor content of the $t\bar{t} \to (l^+\nu q)(l^-\bar{\nu}\bar{q})$ channel in proton-proton collisions at $\sqrt{s}$ =7 tev. *CMS Analysis Note*, CMS AN-2011/394, Nov 2011.

[135] CMS Collaboration. Performance of the b-jet identification in cms. CMS Physics Analysis Summary CMS-PAS-BTV-11-001, 2011. URL `https://cdsweb.cern.ch/record/1366061/`.

[136] R. Ciesielski, H. Held, P. Schieferdecker, and M. Zielinski. Jet energy resolutions derived from qcd simulation for the analysis of first $\sqrt{s}$=7 tev collision data. *CMS AN-2010/121*, 2010.

[137] Top pair cross section in dileptons. *CMS Physics Analysis Summary*, CMS PAS TOP-10-005, 2011. URL `http://cdsweb.cern.ch/record/1377323`.

[138] W. Verkerke and D. Kirkby. The roofit toolkit for data modeling. *Proc. Conference for Computing in High-Energy and Nuclear Physics (CHEP 03)*, 2003.

[139] G. Cowan. *Statistical Data Analysis*. Clarendon Press - Oxford, 1998.

[140] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002, 2011. URL `http://stacks.iop.org/1748-0221/6/i=11/a=P11002`.

[141] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. doi: 10.1088/1126-6708/2007/11/070.

[142] Pavel M. Nadolsky et al. Implications of CTEQ global analysis for collider observables. *Phys. Rev.*, D78:013004, 2008. doi: 10.1103/PhysRevD.78.013004.

[143] Daniel Wicke and Peter Z. Skands. Non-perturbative QCD Effects and the Top Mass at the Tevatron. *Nuovo Cim.*, B123:S1, 2008.

[144] 920234. Combination of CDF and D0 results on the mass of the top quark using up to 5.8 fb-1 of data. 2011. 16 pages, 2 figures.

[145] Tevatron Electroweak Working Group. Combination of cdf and do results on the mass of the top quark using up to 5.8 fb-1 of data. 2011.

[146] Serguei Chatrchyan et al. Measurement of the top quark pair production cross section in pp collisions at $\sqrt{s} = 7$ TeV in dilepton final states containing a τ. *Phys. Rev. D*, 85:112007, 2012. doi: 10.1103/PhysRevD.85.112007.

[147] John F. Gunion, Howard E. Haber, Gordon L. Kane, and Sally Dawson. *The Higgs Hunters's Guide*. Westview Press, 2000. ISBN 9780738203058.

[148] Abdelhak Djouadi. The anatomy of electro-weak symmetry breaking. II: The Higgs bosons in the Minimal Supersymmetric Model. *Phys. Rept.*, 459:1, 2008. doi: 10.1016/j.physrep.2007.10.005.

[149] Serguei Chatrchyan et al. Search for a light charged Higgs boson in top quark decays in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 07:143, 2012. doi: 10.1007/JHEP07(2012)143.

[150] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torielli, 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.

[151] Michal Czakon, Paul Fiedler, and Alexander Mitov. The total top quark pair production cross section at hadron colliders through O(α_S^4). *Phys. Rev. Lett.*, 110:252004, 2013. doi: 10.1103/PhysRevLett.110.252004.

[152] 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.

[153] Michiel Botje, Jon Butterworth, Amanda Cooper-Sarkar, Albert de Roeck, Joel Feltesse, et al. The PDF4LHC Working Group Interim Recommendations. 2011.

[154] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, et al. CT10 next-to-next-to-leading order global analysis of QCD. *Phys.Rev.*, D89(3):033009, 2014. doi: 10.1103/PhysRevD.89.033009.

[155] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, et al. Parton distributions with LHC data. *Nucl.Phys.*, B867:244–289, 2013. doi: 10.1016/j.nuclphysb.2012.10.003.

[156] Commissioning of the particle flow event reconstruction with the first LHC collisions recorded in the CMS detector. CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010. URL `http://cdsweb.cern.ch/record/1247373`.

[157] CMS. Electron reconstruction and identification at $\sqrt{s} = 7$ TeV. CMS Physics Analysis Summary CMS-PAS-EGM-10-004, 2010. URL `http://cdsweb.cern.ch/record/1299116`.

[158] Serguei Chatrchyan et al. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012. doi: 10.1088/1748-0221/7/10/P10002.

[159] 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.

[160] Serguei Chatrchyan et al. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:P11002, 2011. doi: 10.1088/1748-0221/6/11/P11002.

[161] Serguei Chatrchyan et al. Performance of tau lepton reconstruction and identification in CMS. *JINST*, 7:P01001, 2012. doi: 10.1088/1748-0221/7/01/P01001.

[162] Serguei Chatrchyan et al. Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons. *JHEP*, 05:104, 2014. doi: 10.1007/JHEP05(2014)104.

[163] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.

[164] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013. doi: 10.1088/1748-0221/8/04/P04013.

[165] Serguei Chatrchyan et al. Measurement of the $\mathrm{t\bar{t}}$ production cross section and the top quark mass in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1107:049, 2011. doi: 10.1007/JHEP07(2011)049.

[166] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. doi: 10.1088/1126-6708/2004/11/040.

[167] Simone Alioli, Sven-Olaf Moch, and Peter Uwer. Hadronic top-quark pair-production with one jet and parton showering. *JHEP*, 01:137, 2012. doi: 10.1007/JHEP01(2012)137.

[168] L. Lyons, D. Gibaut, and P. Clifford. How to combine correlated estimates of a single physical quantity. *Nucl. Instrum. Meth. A*, 270:110, 1988. doi: 10.1016/0168-9002(88)90018-6.

[169] ATLAS, CDF, CMS, and D0 collaborations. First combination of Tevatron and LHC measurements of the top quark mass. 2014.

[170] A. Abulencia et al. Search for charged Higgs bosons from top quark decays in $p\bar{p}$ collisions at $\sqrt{s} =$ 1.96-TeV. *Phys. Rev. Lett.*, 96:042003, 2006. doi: 10.1103/PhysRevLett.96.042003.

[171] V. M. Abazov et al. Search for charged Higgs bosons in top quark decays. *Phys. Lett. B*, 682:278, 2009. doi: 10.1016/j.physletb.2009.11.016.

[172] Georges Aad et al. Search for charged Higgs bosons decaying via $H^+ \to \tau\nu$ in top quark pair events using pp collision data at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 1206:039, 2012. doi: 10.1007/JHEP06(2012)039.

[173] U. Langenfeld, S. Moch, and P. Uwer. New results for $\mathrm{t\bar{t}}$ production at hadron colliders. 2009.

[174] Nikolaos Kidonakis. Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution. *Phys. Rev. D*, 82:114030, 2010. doi: 10.1103/PhysRevD.82.114030.

[175] Matteo Cacciari, Stefano Frixione, Michelangelo L. Mangano, Paolo Nason, and Giovanni Ridolfi. Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC. *JHEP*, 0809:127, 2008. doi: 10.1088/1126-6708/2008/09/127.

[176] S. Chatrchyan et al. Measurement of the $t\bar{t}$ production cross section in pp collisions at 7 TeV in lepton $+$ jets events using b-quark jet identification. *Phys. Rev. D*, 84:092004, 2011. doi: 10.1103/PhysRevD.84.092004. URL `http://link.aps.org/doi/10.1103/PhysRevD.84.092004`.

[177] Georges Aad et al. Measurement of the cross section for top-quark pair production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector using final states with two high-pt leptons. Submitted to *JHEP*, 2012.

[178] CMS Collaboration. Performance of τ-lepton reconstruction and identification in cms. *Journal of Instrumentation*, 7(01):P01001, 2012. URL `http://stacks.iop.org/1748-0221/7/i=01/a=P01001`.

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

[180] Jason Gallicchio and Matthew D. Schwartz. Quark and Gluon Tagging at the LHC. *Phys. Rev. Lett.*, 107:172001, 2011. doi: 10.1103/PhysRevLett.107.172001.

[181] CMS Collaboration. Determination of the jet energy scale in cms with pp collisions at $\sqrt{s}$ = 7 TeV. CMS Physics Analysis Summary CMS-PAS-JME-10-010, 2010. URL `http://cdsweb.cern.ch/record/1308178`.

[182] CMS Collaboration. Measurement of b-tagging efficiency using $\mathrm{t\bar{t}}$ events. CMS Physics Analysis Summary CMS-PAS-BTV-11-003, 2011. URL `http://cdsweb.cern.ch/record/1421611`.

[183] ATLAS and CMS Collaborations, LHC Higgs Combination Group. Procedure for the LHC higgs boson search combination in Summer 2011. ATL-PHYS-PUB/CMS NOTE 2011-11, 2011/005, 2011. URL `http://cdsweb.cern.ch/record/1379837`.

[184] R. Harlander, M. Krämer, and M. Schumacher. Bottom-quark associated Higgs-boson production: reconciling the four- and five-flavour scheme approach. Technical Report CERN-PH-TH/2011-134, FR-PHENO-2011-009, TTK-11-17, WUB/11-04, 2011.

[185] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections. *Eur.Phys.J.*, C64:653–680, 2009. doi: 10.1140/epjc/s10052-009-1164-2.

[186] Combination of ATLAS and CMS top quark pair cross section measurements in the emu final state using proton-proton collisions at 8 TeV. CMS Physics Analysis Summary CMS-PAS-TOP-14-016, 2014.

[187] S. Chatrchyan et al. Measurement of differential top-quark pair production cross sections in pp collisions at $\sqrt{s}$ = 7 TeV. *Eur. Phys. J. C*, 73:2339, 3 2013. doi: 10.1140/epjc/s10052-013-2283-3.

[188] CMS Collaboration. Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV. CMS Physics Analysis Summary CMS-PAS-TOP-12-027, 2012. URL `http://cdsweb.cern.ch/record/1523611`.

[189] CMS Collaboration. Measurement of the differential top-quark pair production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8$ TeV. CMS Physics Analysis Summary CMS-PAS-TOP-12-028, 2012. URL `http://cdsweb.cern.ch/record/1523664`.

[190] K. Melnikov and F. Petriello. Electroweak gauge boson production at hadron colliders through o(α^2). *Phys. Rev. D*, 74:114017, 2006. doi: 10.1103/PhysRevD.74.114017.

[191] K. Melnikov and F. Petriello. The w boson production cross section at the LHC through o(α^2). *Phys. Rev. Lett.*, 96:231803, 2006. doi: 10.1103/PhysRevLett.96.231803.

[192] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, et al. HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR. *Comput.Phys.Commun.*, 182:1034–1046, 2011. doi: 10.1016/j.cpc.2010.12.040.

[193] P. Kant, O.M. Kind, T. Kintscher, T. Lohse, T. Martini, et al. HATHOR for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions. 2014.

[194] N. Kidonakis. Nnll resummation for s-channel single top quark production. *Phys. Rev. D*, 81: 054028, 2010. doi: 10.1103/PhysRevD.81.054028.

[195] N. Kidonakis. Two-loop soft anomalous dimensions for single top quark associated production with a w^- or h^-. *Phys. Rev. D*, 82:054018, 2010. doi: 10.1103/PhysRevD.82.054018.

[196] Vardan Khachatryan et al. Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV in dilepton final states containing one τ lepton. *Phys. Lett. B*, 739:23, 2014. doi: 10.1016/j.physletb.2014.10.032.

[197] Serguei Chatrchyan et al. Measurement of the $\mathrm{t\bar{t}}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 11:067, 2012. doi: 10.1007/JHEP11(2012)067.

[198] Vardan Khachatryan et al. Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1101:080, 2011. doi: 10.1007/JHEP01(2011)080.

[199] Vardan Khachatryan et al. Measurement of the ratio $B(t \to Wb)/B(t \to Wq)$ in pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett. B*, 736:33, 2014. doi: 10.1016/j.physletb.2014.06.076.

[200] Vardan Khachatryan et al. Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions. *JHEP*, 10:160, 2014. doi: 10.1007/JHEP10(2014)160.

[201] Vardan Khachatryan et al. Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C75(11):542, 2015. doi: 10.1140/epjc/s10052-015-3709-x.

[202] V. Khachatryan et al. Measurement of the cross section ratio in collisions at. *Physics Letters B*, 746:132 – 153, 2015. ISSN 0370-2693. doi: http://dx.doi.org/10.1016/j.physletb.2015.04.060. URL `http://www.sciencedirect.com/science/article/pii/S0370269315003263`.

[203] ATLAS and CMS Collaborations, LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in summer 2011. Technical Report ATL-PHYS-PUB-2011-011, CMS NOTE 2011/005, 2011. URL `http://cdsweb.cern.ch/record/1379837`.

[204] J. S. Conway. Nuisance parameters in likelihoods for multisource spectra. In H.B. Prosper and L. Lyons, editors, *Proceedings of PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, number CERN-2011-006, page 115. CERN, 2011. URL `http://cdsweb.cern.ch/record/1306523`.

[205] Martin Flechl, Richard Klees, Michael Kramer, Michael Spira, and Maria Ubiali. Improved cross-section predictions for heavy charged Higgs boson production at the LHC. *Phys. Rev. D*, 91: 075015, 2015. doi: 10.1103/PhysRevD.91.075015.

[206] S. Dittmaier, M. Krämer, M. Spira, and M. Walser. Charged-Higgs-boson production at the LHC: NLO supersymmetric QCD corrections. *Phys. Rev. D*, 83:055005, 2009.

[207] Edmond L. Berger, Tao Han, Jing Jiang, and Tilman Plehn. Associated production of a top quark and a charged Higgs boson. *Phys. Rev. D*, 71:115012, 2005. doi: 10.1103/PhysRevD.71.115012.

[208] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard model higgs-boson branching ratios with uncertainties. *Eur. Phys. J. C*, 71:1753, 2011. doi: 10.1140/epjc/s100052-011-1753-8.

[209] S. Dittmaier et al. Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. 2011. doi: 10.5170/CERN-2011-002.

[210] Lars Hofer, Ulrich Nierste, and Dominik Scherer. Resummation of tan-beta-enhanced supersymmetric loop corrections beyond the decoupling limit. *JHEP*, 0910:081, 2009. doi: 10.1088/1126-6708/2009/10/081.

[211] Poster session of the march 2012 lhcc. *https://indico.cern.ch/event/180683/*, .

[212] Prospects for charged higgs discovery at colliders – c$h^{\pm}$arged 2012. *https://indico.cern.ch/event/168858/*, .

[213] Prospects for charged higgs discovery at colliders – c$h^{\pm}$arged 2014. *https://indico.cern.ch/event/295196/*, .

[214] The third annual large hadron collider physics conference – lhcp 2015. *http://lhcp2015.com/*, .

[215] Estágios de verão 2015 – experiência cms no lhc do cern. *https://fenix.tecnico.ulisboa.pt/areacientifica/df/fpfn/ver-post/estagio-34hands-on-quantum-mechanics34*.

[216] Georges Aad et al. Measurement of the WZ production cross section and limits on anomalous triple gauge couplings in proton-proton collisions at sqrt(s) = 7 TeV with the ATLAS detector. *Phys.Lett.*, B709:341–357, 2012. doi: 10.1016/j.physletb.2012.02.053.

[217] John M Campbell and R Keith Ellis. $\mathrm{t\bar{t}}W^{+-}$ production and decay at NLO.

Appendix A

The single lepton (e/μ+jets) final states for $\mathrm{H}^+ \to \mathrm{t}\overline{\mathrm{b}}$

In this analysis, charged Higgs bosons produced in association with a top quark $\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+$, are searched for in the decay mode $\mathrm{H}^+ \to \mathrm{t}\bar{\mathrm{b}}$. The subsequent decays of the two W bosons produced in decays of the top quarks, $\bar{\mathrm{t}} \to \bar{\mathrm{b}}\mathrm{W}^-$, define the final state. In this search, one of the W boson decays leptonically, while the other W boson decays hadronically, leading to the final state signature of one lepton, jets, and MET. Thus, the full production and decay chain is given by $\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+ \to \ell\nu_\ell qq'\mathrm{b}\overline{\mathrm{b}\mathrm{b}}(\mathrm{b})$, where ℓ can be either an electron or a muon and can originate from either of the two top quarks. This final state signature is denoted as ℓ+jets. These final states are similar to the SM $\mathrm{t}\bar{\mathrm{t}}$ semi-leptonic final states, with the addition of one or two b jets. The H_T variable is defined as the scalar sum of the p_T of all selected jets. The distribution of this variable is used to check for the presence of signal in the data. The presence of at least one b-tagged jet is utilized further in conjunction with an H_T shape analysis to discriminate against and to gain sensitivity for the H^+ signal and discriminate against the dominant backgrounds from SM $\mathrm{t}\bar{\mathrm{t}}$, single top, and W+$\mathrm{c}\bar{\mathrm{c}}/\mathrm{b}\bar{\mathrm{b}}$ production.

A.1 Event selection

Events are selected with criteria consistent with the signal topology and final state signature. Data are collected by the single electron or a single muon trigger with p_T thresholds of 27 and 24 GeV, respectively. The offline event selection proceeds by requiring the presence of exactly one isolated electron (muon) with $p_\mathrm{T} > 30(27)$ GeV and $|\eta| < 2.5$ ($|\eta| < 2.4$). The electrons (muons) are required to be isolated with the relative isolation, I_rel, required to be less than $< 0.10(0.20)$. Events with additional leptons are rejected. In addition, the presence of at least two jets with $p_\mathrm{T} > 30$ GeV and $|\eta| < 2.4$ are required, with the highest p_T jet being above 50 GeV. At least one of the jets is required to be b-tagged. The missing transverse energy $E_\mathrm{T}^\mathrm{miss}$ must exceed 20 GeV to mimic the presence of a neutrino in the final event signature. To maintain orthogonality with the other analyses included in this paper, events with one or more hadronic tau decays with $p_\mathrm{T} > 20$ GeV and $|\eta_{\tau_\mathrm{h}}| < 2.4$ are rejected.

To account for differences in modelling of the lepton identification and trigger efficiency between simulation and data, η and p_{T}-dependent scale factors are applied. The single electron trigger correction factor is 0.973 (1.02) for $|\eta| \leq 1.5$ ($1.5 < |\eta| \leq 2.5$) and the single muon trigger correction factors vary between 0.96 and 0.99. The corrections to identification efficiency vary between 0.91(0.97) and 1.0(0.99) for electrons (muons).

The number of data events after each selection cut are compared to expectations from SM backgrounds and are shown in Fig. A.1 for both the electron and muon channels. Results are in good agreement with SM background expectations.

In this analysis, the b-tagging working point is loosened to a one per cent mistagging probability since the multijet background is smaller than in the τ_{h}+jets final state described in Sec. B . The corresponding probability to idenfity a b jet is about 70 %.

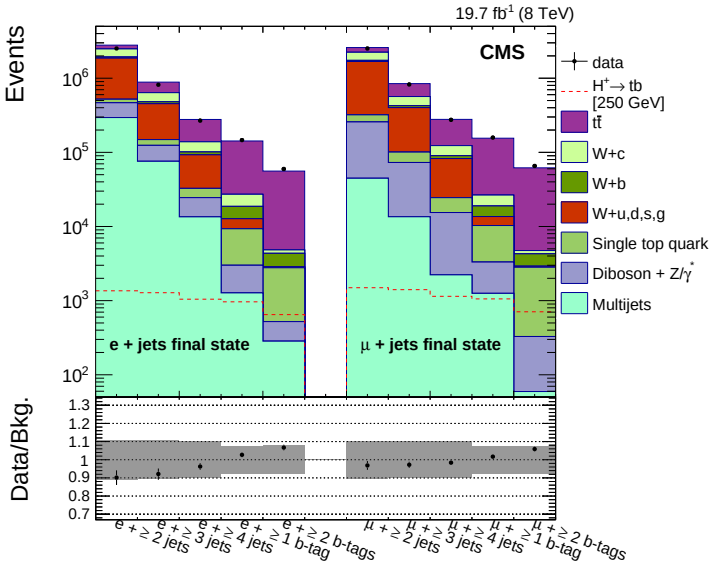


Figure A.1: Event yields at different selection cut levels for both the electron (rleft) and muon (right) final state. The bottom panel shows the ratio of data over the sum of the SM backgrounds with the total uncertainties.

A.2 Background estimate

The following background processes are considered: $t\bar{t}$+jets, single top, $t\bar{t} + Z$, $W(\rightarrow \ell\nu)$+jets, $Z/\gamma^{*}(\rightarrow \ell\ell)$+jets, and dibosons (WW, WZ, and ZZ). As observed in the $\mathrm{t\bar{t}}$ differential cross section measurement [188, 189], a mis-modelling of the top quark p_{T} distribution is present in the simulated $\mathrm{t\bar{t}}$ samples. The p_{T} distribution observed in data is softer than that predicted by simulation. Therefore the $\mathrm{t\bar{t}}$ simulation is reweighted using an empirical function based on the generated p_{T} of the top and anti-top quarks.

The backgrounds are subdivided into independent categories based on their yields and shapes in the signal region. The four samples: $\mathrm{t\bar{t}}$, W + light flavour (u, d, s, g) jets, W + heavy flavour (c, b) jets $(\mathrm{W} + \mathrm{c\bar{c}}/\mathrm{b\bar{b}})$, and QCD multijet are defined as independent categories. All H_{T} distributions are taken

Background	Electron channel	Muon channel
$\mathrm{SF}_{\mathrm{W+light}}$	1.08 ± 0.17	0.90 ± 0.09
$\mathrm{SF}_{\mathrm{W+c\bar{c}/b\bar{b}}}$	1.87 ± 0.18	1.52 ± 0.20
$\mathrm{SF}_{\mathrm{t\bar{t}}}$	1.05 ± 0.09	1.04 ± 0.09

Table A.1: Maximum-Likelihood fit results for normalization scale factors for the dominant backgrounds in the electron and muon channels. The uncertainty includes the statistical uncertainties obtained from the fit and a systematic component to account for the modelling of the CR.

from simulation and smaller background processes with similar shapes are grouped together and treated as one process. To this end, all diboson samples (WW, WZ, and ZZ) are merged together with Z+jets to make a single distribution, the "Z/γ^* /VV " background. Additionally, all of the single-top samples (t-,s- and tW-channel) are combined with the $\mathrm{t\bar{t}}$ + Z sample to create the "single top + $\mathrm{t\bar{t}}$+Z" background category. This results in a total of six background categories: $\mathrm{t\bar{t}}$, W + light flavour (u, d, s, g) jets, $\mathrm{W} + \mathrm{c\bar{c}}/\mathrm{b\bar{b}}$, QCD multijet, DY/VV, and single top. Additional contributions from $\mathrm{t\bar{t}}$ +Z and $\mathrm{t\bar{t}}$ +W are considered negligible and are not used.

For backgrounds which contribute small amounts to the signal region (single-top, diboson, Z+jets and QCD production), the normalizations are taken directly from the simulation. A more detailed procedure is developed for the three remaining processes which make up the majority of the background in the signal region: W + light flavour (u, d, s, g) jets, $\mathrm{W} + \mathrm{c\bar{c}}/\mathrm{b\bar{b}}$, and $\mathrm{t\bar{t}}$ production. The normalization for these backgrounds is initially taken from simulation, but is then determined by a maximum likelihood fit (ML) of the background distributions to the data.

First, preselected event samples are further subdivided into a few orthogonal regions, a signal region (SR) and a control region (CR). The CR is signal-depleted to derive data-driven normalizations for the dominant backgrounds, and check the background modelling. The SR is distinguished by its high jet multiplicity, and defined by the requirement $N_{\mathrm{jet}} \geq 4$. To enhance the overall signal sensitivity, the events in the SR are expected to have at least one b- ged jet and are divided into two mutually exclusive subsamples with $N_{\mathrm{b\ jet}} = 1$ and $N_{\mathrm{b\ jet}} \geq 2$. The background CR is defined by the low number of reconstructed jets (N_{jet}) and events are restricted to have $2 \leq N_{\mathrm{jet}} \leq 3$. The CR is further split into three subsamples defined by orthogonal b jet multiplicity requirements: $N_{\mathrm{b\ jet}} = 0$, $N_{\mathrm{b\ jet}} = 1$, and $N_{\mathrm{b\ jet}} \geq 2$. When considering that events are split into electron and muon channels as well, this leads to a total of four samples in the SR and six samples in the CR. All of the samples are mutually exclusive and do not have any common events.

A simultaneous ML fit of the background H_{T} distributions to the data in the two signal regions and in the three control regions for each lepton flavour final state is performed. Separate scale factors are assigned for the electron and muon channels. The six scale factors, $\mathrm{SF}_{\mathrm{W+light}}$, $\mathrm{SF}_{\mathrm{W+c\bar{c}/b\bar{b}}}$, and $\mathrm{SF}_{\mathrm{t\bar{t}}}$ (each for the electron and the muon channel), are left free to float during the ML fit so that the fit finds the best values for these scale factors. The values obtained for these normalization scale factors from the ML fit can be seen in table A.1.

A.3 Event yields

The number of expected events in each subsample can be seen in Table A.2. Events for data, SM background processes, and a charged Higgs boson with a mass of $m_{\mathrm{H}^+} = 250$ GeV are shown. The leading contributions to the SM background come from $t\bar{t}$ with a semi-leptonic final state and W boson production in association with heavy flavour jets. Statistical and systematic uncertainties are evaluated as described in Section 7.11 and are listed.

Table A.2: Number of expected events for the SM backgrounds and for signal events with a charged Higgs boson mass of $m_{\mathrm{H}^+} = 250$ GeV in the ℓ+jets final states after the final event selection. The signal cross section is normalized to 1 pb, assuming a branching fraction $\mathcal{B} = 100\,\%$ for $\mathrm{H}^+ \to \mathrm{t\bar{b}}$. Statistical and systematic uncertainties are shown.

Source	$N_{\mathrm{b\ jet}} = 1$	$N_{\mathrm{b\ jet}} \geq 2$	$N_{\mathrm{b\ jet}} = 1$	$N_{\mathrm{b\ jet}} \geq 2$
	Electron channel		Muon channel	
$\mathrm{H}^+ \to \mathrm{t\bar{b}}$, $m_{\mathrm{H}^+} = 250$ GeV	$315 \pm 4 \pm 17$	$647 \pm 6 \pm 34$	$348 \pm 5 \pm 19$	$707 \pm 7 \pm 37$
$\mathrm{t\bar{t}}$	$66714 \pm 77 \pm 6810$	$53132 \pm 68 \pm 5893$	$73790 \pm 81 \pm 7516$	$58846 \pm 72 \pm 6494$
$\mathrm{W{+}c\bar{c}/b\bar{b}}$	$11714 \pm 101 \pm 1702$	$1909 \pm 39 \pm 280$	$10756 \pm 91 \pm 1795$	$1836 \pm 38 \pm 314$
W+light flavour jets	$3054 \pm 41 \pm 656$	$83 \pm 6 \pm 22$	$2895 \pm 35 \pm 519$	$85 \pm 6 \pm 19$
single top + $\mathrm{t\bar{t}}$+Z	$4059 \pm 42 \pm 463$	$2253 \pm 30 \pm 274$	$4496 \pm 44 \pm 518$	$2493 \pm 32 \pm 292$
DY/VV	$1492 \pm 54 \pm 771$	$237 \pm 21 \pm 130$	$1792 \pm 60 \pm 941$	$269 \pm 22 \pm 139$
QCD multijet background	$987 \pm 274 \pm 1035$	$284 \pm 164 \pm 290$	$1215 \pm 476 \pm 1258$	$59 \pm 34 \pm 60$
Total SM backgrounds	$88020 \pm 312 \pm 7182$	$57898 \pm 186 \pm 5915$	$94944 \pm 498 \pm 7919$	$63588 \pm 97 \pm 6510$
Data	86580	59637	92391	65472

The H_{T} distributions for the four signal regions are shown in Fig. A.2. The ratio of data and the sum of expected SM background contributions is shown in the bottom panel. Limits on the production of the charged Higgs boson are extracted by exploiting these distributions.

A.3.1 Systematic uncertainties specific to the single-lepton (e/μ+jets) final states for $\mathrm{H}^+ \to \mathrm{t\bar{b}}$

In addition to the uncertainties described earlier in Sec. 7.11.1, the following systematic uncertainties specific to the ℓ+jets final states, affecting the simulated samples only, are as follows:

- The normalizations for $\mathrm{t\bar{t}}$, $\mathrm{W{+}c}$, $\mathrm{W{+}b}$, and W+light-flavour backgrounds are left unconstrained. Statistical and systematic uncertainties are applied to yields in the control regions described in Section 3.4. These uncertainties are based on deviations of the fitted normalization factor when varying multijet and Z/γ^*+jets contributions by a factor of two, signal contamination by a factor of five, and by requiring either two or three jets in the control region. The total uncertainty in the normalization factors ranges between 5–35%.

- A 50% uncertainty [130, 216, 217] is applied to the Z/γ^*+jets and diboson backgrounds due to their small contribution to the signal region;

- A 100% systematic uncertainty is applied to the QCD cross section normalization. This accounts for the maximal variation in the QCD normalization when left unconstrained in the background-only

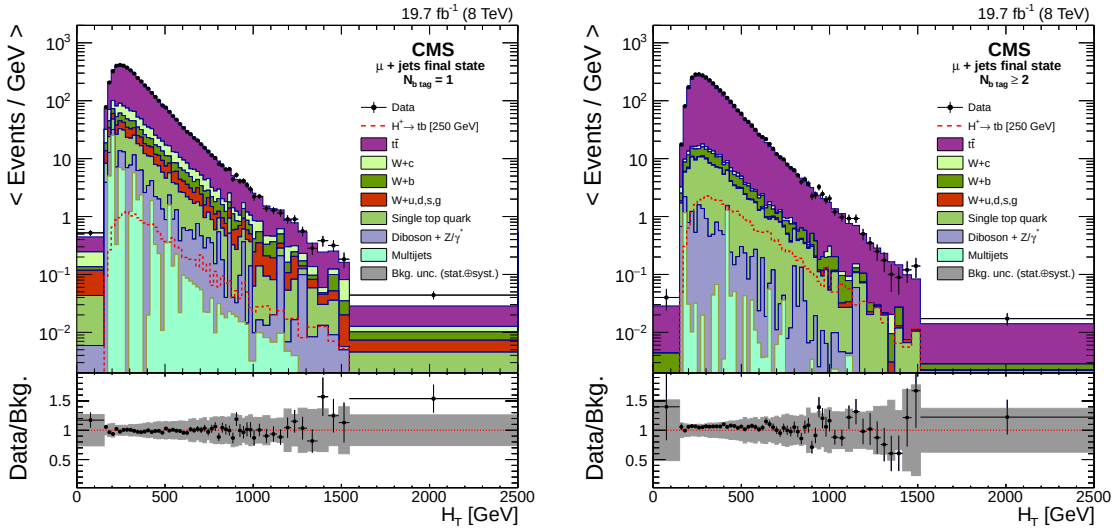


Figure A.2: The H_{T} distributions observed in data, and predicted for signal and background, for selected events in the muon (bottom) channel with $N_{\mathrm{b\ jet}} = 1$ (left) and $N_{\mathrm{b\ jet}} \geq 2$ (right). Normalizations for $\mathrm{t\bar{t}}$, W + light jets (u,d,s,g), and $\mathrm{W} + \mathrm{c\bar{c}}/\mathrm{b\bar{b}}$ are derived from data (see text). Normalizations for other backgrounds are based on simulation. Expectations for the charged Higgs boson for $m_{\mathrm{H^+}} = 250$ GeV, for the $\mathrm{H^+} \to \mathrm{t\bar{b}}$ decays, are also shown. The signal yields are normalized to a cross section of 1 pb. The bottom panel shows the ratio of data and the sum of the SM backgrounds with the total uncertainties. Expectations for the charged Higgs boson are for $m_{\mathrm{H^+}}$ = 250 GeV, in an assumption of the pp→t(b)H cross section being 1 pb and $\mathcal{B}(\mathrm{t} \to \mathrm{H^+b})$ =1. Bin contents are normalized to the bin's width.

fit to data while constraining normalizations for other backgrounds to their systematic uncertainties.

The systematic uncertainties for signal and background events are summarized in Table A.3.

Table A.3: The systematic uncertainties (in %) for backgrounds, and for signal events from $\mathrm{H^+} \to \mathrm{t\bar{b}}$ decays for the ℓ+jets channels for a charged Higgs boson mass $m_{\mathrm{H^+}} = 250$ GeV. The uncertainties which depend on the shape of the H_{T} distribution bin are marked with (S) and for them the maximum integrated value of the negative or positive variation is displayed. Empty cells indicate that an uncertainty does not apply to the sample. The uncertainties in the rows are considered to be fully correlated, with the exception of cross section and Data-driven normalization, which are considered to be uncorrelated. The uncertainties in the columns are considered to be uncorrelated. Uncertainties labelled with a * are left unconstrained and are fit to corresponding yields in the control regions (sec. A.2).

	$\mathrm{H^+} \to \mathrm{t\bar{b}}$	$\mathrm{t\bar{t}}$	$\mathrm{W{+}c\bar{c}/b\bar{b}}$	W+light	single top + $\mathrm{t\bar{t}}$+Z	Z/γ^* /VV	QCD multijet
Jet energy scale (S)	4.0	6.4	12.6	13.6	9.2	27.1	49.3
Jet energy resolution (S)	0.1	0.3	1.3	1.4	0.8	2.3	6.9
b tagging (S)	3.9	1.3	3.1	11.0	0.7	5.4	16.0
Pileup modelling (S)	1.2	0.7	0.4	0.4	0.7	3.7	7.0
Data-driven normalization, e+jets		*	*	*			
Data-driven normalization, μ+jets		*	*	*			
cross section					8.0	50.0	100
Top quark p_{T} modelling (S)		3.5					
Matching scale (S)		7.3					
$\mathcal{Q}^2$ scale (S)		7.6					
Luminosity	2.6				2.6	2.6	2.6
Electron trigger	2.0	2.0	2.0	2.0	2.0	2.0	2.0
μ trigger	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Electron identification	1.0	1.0	1.0	1.0	1.0	1.0	1.0
μ identification	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Appendix B

The τ_{h}+jets (fully hadronic) final state for $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$

In this analysis a charged Higgs boson is assumed to be produced through the $\mathrm{t\bar{t}} \to \mathrm{bH^{+}\bar{b}H^{-}}$, $\mathrm{t\bar{t}} \to \mathrm{bH^{+}\bar{b}W^{-}}$, and $\mathrm{pp} \to \mathrm{\bar{t}(b)H^{+}}$ processes and searched for in the $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decay mode with a hadronic decay of the tau and a hadronic decay of the W boson that originates from the associated $\mathrm{\bar{t}} \to \mathrm{\bar{b}W^{-}}$ decay. In these events the missing transverse momentum is expected to originate from the neutrinos in the decay of the charged Higgs boson, which allows for the reconstruction of the transverse mass, m_{T}, of the charged Higgs boson:

$$m_{\mathrm{T}} = \sqrt{2 p_{\mathrm{T}}^{\tau_{\mathrm{h}}} E_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\tau_{\mathrm{h}}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}))}, \tag{B.1}$$

where $\vec{p}_{\mathrm{T}}^{\,\tau_{\mathrm{h}}}$ denotes the transverse momentum vector of the hadronic tau and $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ its magnitude and $\Delta\phi$ the angle between the τ_{h} direction and the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ in the transverse plane. The search is performed as a shape analysis, using the transverse mass to infer the presence of a signal. The dominant background processes are the SM $\mathrm{t\bar{t}}$ and single top production as well as the electroweak (EWK) processes: W+jets, Z+jets, and dibosons (WW, ZZ, WZ). The multijet production constitutes a subleading background.

B.1 Event selection

The events are selected with a trigger that requires the presence of a τ_{h} and large $E_{\mathrm{T}}^{\mathrm{miss}}$. First the events are required to have calorimetric $E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV at the first level of the CMS trigger system. At the high-level trigger, the events are required to have calorimetric $E_{\mathrm{T}}^{\mathrm{miss}} > 70$ GeV, and a τ_{h} of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}} > 35$ GeV and $|\eta^{\tau_{\mathrm{h}}}| < 2.5$. The τ_{h} is required to be loosely isolated, to contain at least one track of $p_{\mathrm{T}} > 20$ GeV, and to have at most two tracks in total, targeting the tau decay modes into a single charged pion and up to two neutral pions. The probability for a signal event to pass the trigger amounts to 8–14 % in the $m_{\mathrm{H^{+}}}$ range of 80–160 GeV, and 19–44 % in the $m_{\mathrm{H^{+}}}$ range of 180–600 GeV.

The efficiency of the τ part of the trigger is evaluated from $\mathrm{Z}/\gamma^{*} \to \tau_{\mu}\tau_{\mathrm{h}}$ events, where τ_{μ} refers to

a muonic τ lepton decay, using a tag-and-probe technique as a function of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ [91]. The efficiency of the $E_{\mathrm{T}}^{\mathrm{miss}}$ part of the trigger is evaluated from events with a $\mathrm{t\bar{t}}$-like final state of τ_{h}+jets selected with a single tau trigger. The trigger efficiencies in simulated events are corrected with data-to-simulation scale factors applied as function of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ for the τ_{h} part of the trigger and as function of $E_{\mathrm{T}}^{\mathrm{miss}}$ for the $E_{\mathrm{T}}^{\mathrm{miss}}$ part of the trigger.

The b-tagging working point corresponding to one per mille probability for a jet arising from the fragmentation of light quarks or gluon to be misidentified as a b jet is used to suppress the multijet background. The corresponding probability to idenfity a b jet is about 50 %.

Selected events are required to have at least one τ_{h} with $p_{\mathrm{T}}^{\tau_{\mathrm{h}}} > 41$ GeV within $|\eta| < 2.1$ and matching to a τ_{h} trigger object. These thresholds are chosen to be compatible with the single muon trigger used for the data-driven estimate of backgrounds with hadronic tau decays as described in Section B.2.1. Only one charged hadron is allowed to be associated with the τ_{h} and its p_{T} is required to fulfill $p_{\mathrm{T}}^{\mathrm{h}} > 20$ GeV. Background events with $\mathrm{W} \to \tau\nu_{\tau}$ decays are suppressed by requiring $R_{\tau} = p^{\text{charged hadron}}/p^{\tau_{\mathrm{h}}} > 0.7$. The R_{τ} observable is sensitive to different polarizations of τ leptons originating from decays of W bosons (spin 1) and from decays of H^{+} (spin 0) [66].

A $\mathrm{t\bar{t}}$-like event topology is selected by requiring at least three jets of $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$ in addition to the τ_{h} and by requiring at least one of the selected jets to be identified as originating from the hadronization of a b quark. To select a fully hadronic final state, events containing identified and isolated electrons (muons) with $p_{\mathrm{T}} > 15$ (10) GeV are rejected. The electron (muon) candidates are considered to be isolated if the relative isolation $I_{\mathrm{rel.}}^{\mathrm{e}}$ ($I_{\mathrm{rel.}}^{\mu}$), as described in Section 3.4, is smaller than 15 % (20 %).

To suppress the multijet background, $E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV is required. The lower $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold on the particle-flow $E_{\mathrm{T}}^{\mathrm{miss}}$ compared to the calorimetric $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement applied at the trigger level improves the signal acceptance for the relatively light charged Higgs bosons. This approach can be used because of the improved resolution of the particle-flow $E_{\mathrm{T}}^{\mathrm{miss}}$ compared to the calorimetric $E_{\mathrm{T}}^{\mathrm{miss}}$.

In the multijet events selected with the τ+$E_{\mathrm{T}}^{\mathrm{miss}}$ trigger a hadronic jet is misidentified as the τ_{h} in the event and has typically a recoiling jet in the opposite direction. The $E_{\mathrm{T}}^{\mathrm{miss}}$ in these events arises from the mismeasurement of the energies of these jets with the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ direction thus aligned with the τ_{h} and recoiling jet momenta in the transverse plane causing either back-to-back or collinear systems of the $\vec{p}_{\mathrm{T}}^{\,\tau_{\mathrm{h}}}$ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. Such events with a back-to-back configuration of the $\vec{p}_{\mathrm{T}}^{\,\tau_{\mathrm{h}}}$ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ end up in the signal region in the m_{T} distribution thereby reducing sensitivity of the analysis. The signal events do not, however, contain a recoiling jet for the τ_{h} and therefore the correlations on the azimuthal angles between the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, τ_{h}, and selected jets can be used to suppress multijet events with minimal impact on signal acceptance. Events with a collinear configuration of the τ_{h} and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ are also suppressed to avoid events with misidentified $E_{\mathrm{T}}^{\mathrm{miss}}$. The best performance in terms of multijet suppression and signal acceptance is found to be obtained with two-dimensional circular cut variables instead of simple selections based on

azimuthal angle differences. The variables used for the azimuthal angle selections are defined as

$$R_{\mathrm{bb}}^{\mathrm{min}} = \min\left\{\sqrt{\left(\pi - \Delta\phi(\tau_{\mathrm{h}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})\right)^2 + \left(\Delta\phi(\mathrm{jet}_n, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})\right)^2}\right\}$$
$$R_{\mathrm{coll}}^{\mathrm{min}} = \min\left\{\sqrt{\left(\Delta\phi(\tau_{\mathrm{h}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})\right)^2 + \left(\pi - \Delta\phi(\mathrm{jet}_n, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})\right)^2}\right\}, \quad \text{(B.2)}$$

where the index n refers to any of the three highest-p_{T} jets in the event and $\Delta\phi$ denotes the azimuthal angle between the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ and the τ_{h} or one of the three highest-p_{T} jets. The selected events are required to satisfy $R_{\mathrm{coll}}^{\mathrm{min}} > 40^\circ$ and $R_{\mathrm{bb}}^{\mathrm{min}} > 40^\circ$.

The same event selection is used for all considered m_{H^+} values.

B.2 Background measurements

The background contributions arise from three sources:

1. Irreducible background from electroweak processes — W+jets, Z+jets, and dibosons — as well as SM $\mathrm{t\bar{t}}$ and single top production in the t- and tW-channels, where the selected τ_{h} originates from a hadronic decay of a τ lepton ("EWK+$\mathrm{t\bar{t}}$ with τ_{h}").

2. Reducible background from multijet events with large mismeasured $E_{\mathrm{T}}^{\mathrm{miss}}$ and jets that mimic hadronic τ decays.

3. Reducible background from EWK+$\mathrm{t\bar{t}}$ events, where an electron, muon, or a jet is misidentified as the τ_{h} ("EWK+$\mathrm{t\bar{t}}$ no τ_{h}").

The two largest backgrounds, "EWK+$\mathrm{t\bar{t}}$ with τ_{h}" and multijets, are measured with data-driven methods explained below in Sections B.2.1 and B.2.2. The contribution from "EWK+$\mathrm{t\bar{t}}$ no τ_{h}" is estimated from simulation and is described in Section B.2.3.

B.2.1 Measurement of the EWK+$\mathrm{t\bar{t}}$ with hadronic taus background

The m_{T} distribution for the "EWK+$\mathrm{t\bar{t}}$ with τ_{h}" background is modeled via the so-called embedding technique. First a control data sample of μ+jets events triggered with a single μ trigger is selected. The same jet selection as in the τ_{h} +jets sample is used, and events with electrons or further muons are rejected. Then the selected μ is replaced by a simulated τ lepton decay. The τ lepton momentum is the same as that of the μ, and the reconstructed τ decay products are merged with the original μ+jets event, from which the reconstructed muon is removed. In these hybrid events, the jets are reclustered and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is recalculated and then the events are subjected to the same event selection as the τ_{h}+jets sample, i.e. τ_{h} identification, b tagging, $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement, and the azimuthal angle selections are applied.

To obtain the estimate for the m_{T} distribution as if the events would have been selected with the τ+$E_{\mathrm{T}}^{\mathrm{miss}}$ trigger one needs to "undo" the effect of the muon trigger and the offline reconstruction efficiencies and to incorporate the efficiency of the τ+$E_{\mathrm{T}}^{\mathrm{miss}}$ trigger. First, the weight of each hybrid event is

increased by the inverse of the muon trigger and identification efficiencies. Then, the efficiency of the τ+$E_{\mathrm{T}}^{\mathrm{miss}}$ trigger is applied by weighting the events with the efficiencies of the τ part and of the first trigger level part of the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger and by placing a cut on a hybrid calorimetric $E_{\mathrm{T}}^{\mathrm{miss}}$ constructed from the original event and the simulated τ lepton decay. In addition, further corrections are applied. In a fraction of the selected μ+jets events the μ originates from a decay of a τ lepton, leading to an overestimation of the EWK+$\mathrm{t\bar{t}}$ background by a few percent. This bias is corrected for by applying to the hybrid events p_{T}^{μ}-dependent correction factors derived from simulated $\mathrm{t\bar{t}}$ events. A residual difference is seen in the m_{T} distribution between non-embedded τ+jets and embedded μ+jets events in simulated $\mathrm{t\bar{t}}$ events, which constitutes about 85 % of the "EWK+$\mathrm{t\bar{t}}$ with τ_{h} background". This difference is corrected by weighting the hybrid events by m_{T}-dependent correction factors derived from simulated $\mathrm{t\bar{t}}$ events.

It should be noted that the embedding technique allows for the separation of the charged Higgs boson signal from the $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ decay mode from other decay modes. In other decay modes than $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$, such as $\mathrm{H}^{+} \to \mathrm{t\bar{b}}$, the tau lepton originates from a W boson decay. Firstly, the acceptance for such signals is suppressed greatly by the thresholds on the τ+$E_{\mathrm{T}}^{\mathrm{miss}}$ trigger. And secondly, in the other charged Higgs boson decay modes than $\mathrm{H}^{+} \to \tau^{+}\nu_{\tau}$ taus and muons are produced at equal rates causing the embedding technique to measure such signals in a data-driven way as part of the "EWK+$\mathrm{t\bar{t}}$ with τ_{h}" background.

B.2.2 Measurement of the multijet background

The multijet background is measured with a "τ_{h} fake–rate" technique. An estimate of the multijet background in the signal region is obtained by measuring the probability of the τ_{h} candidate to pass the nominal and inverted τ_{h} isolation criterion. The fake-rate is measured in bins of τ_{h} transverse momentum, in an event sample that is obtained prior to applying the b tagging, $E_{\mathrm{T}}^{\mathrm{miss}}$, and $R_{\mathrm{bb}}^{\mathrm{min}}$ parts of the event selection described in Section B.1. The event sample that passes the nominal τ_{h} isolation selection contains a non-negligible contamination from EWK+$\mathrm{t\bar{t}}$ backgrounds with genuine and misidentified taus. Therefore, the number of multijet and EWK+$\mathrm{t\bar{t}}$ events is determined by a maximum likelihood fit of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. A fit is performed for each $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ bin. For multijet events, the $E_{\mathrm{T}}^{\mathrm{miss}}$ templates are obtained from the data sample with inverted τ_{h} isolation by subtracting a small contribution of simulated EWK+$\mathrm{t\bar{t}}$ events. The $E_{\mathrm{T}}^{\mathrm{miss}}$ templates for the EWK+$\mathrm{t\bar{t}}$ events are taken from simulation in the nominal region. The fake-rate probabilities w_j are defined as the ratio of the number of multijet events in the isolated sample and the inverted isolation sample. Their measured values vary between 0.050–0.061 depending on the $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ bin with a statistical uncertainty smaller than 3 %.

The measured τ_{h} fake-rate probabilities are then applied as weights to multijet events passing all nominal event selection criteria, except that the τ_{h} isolation criterion is inverted. The number of multijet events are obtained by subtracting the number of simulated EWK+$\mathrm{t\bar{t}}$ events from data. The estimate for the number of multijet events in a given bin i of the m_{T} distribution (N_i^{multijet}) is obtained by summing up

these weighted events over the $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ bins according to

$$N_i^{\text{multijet}} = \sum_j (N_{i,j}^{\text{data, inverted}} - N_{i,j}^{\text{EWK+t}\bar{\text{t}}\text{, inverted}}) w_j, \tag{B.3}$$

where N is the number of events and i and j denote m_{T} and $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ bins, respectively.

B.2.3 The EWK+$\mathrm{t\bar{t}}$ with misidentified taus background

The "EWK+$\mathrm{t\bar{t}}$ no τ_{h}" background originates almost solely from jets that are misidentified as the τ_{h} with a small contribution from electrons and muons misidentified as the τ_{h}. About 85 % of the "EWK+$\mathrm{t\bar{t}}$ no τ_{h}" background events come from $\mathrm{t\bar{t}}$ and the rest from single top production in the tW- and t-channels. The number of selected simulated events in the single top samples is small and therefore the m_{T} distribution for them is estimated with a procedure where the probability of each event to pass the b tagging is applied as a per-event weight instead of applying the b-tagging selection. The probability is evaluated for events with $\mathrm{t\bar{t}}$-like final state as function of jet p_{T} and flavour.

B.3 Event yields

Figure B.1 shows the event yield after each selection step starting from the requirement that a τ_{h}, no isolated electrons or muons, and at least three jets are present in the event. The multijet background and the "EWK+$\mathrm{t\bar{t}}$ with τ_{h}" background are shown as measured from the data while the "EWK+$\mathrm{t\bar{t}}$ no τ_{h} " background is shown as estimated from the simulation. The data agree with the sum of expected backgrounds within the total uncertainties.

The observed numbers of events after the full event selection are listed in Table B.1, along with those expected for the backgrounds and for the charged Higgs boson production. The systematic uncertainties listed in Table B.1 are discussed in Section 7.11. In the $m_{\mathrm{H^+}} > m_{\mathrm{t}} - m_{\mathrm{b}}$ region the limited number of background events in the high m_{T} tail is dealt with by fitting an exponential function of the form $p_0 e^{-p_1(m_{\mathrm{T}}-c)}$, where p_0 and p_1 are positive free parameters and where $c = 180$ GeV is the starting point of the fit, to the falling part of the m_{T} distribution for each background process. In the region of $m_{\mathrm{T}} > 160$ GeV the event yields for the backgrounds are taken from this exponential function. The m_{T} distributions obtained after all event selection criteria are applied are shown in Fig. B.2 for $m_{\mathrm{H^+}} < m_{\mathrm{t}} - m_{\mathrm{b}}$ and $m_{\mathrm{H^+}} > m_{\mathrm{t}} - m_{\mathrm{b}}$. Limits on the production of the charged Higgs boson are extracted with these distributions.

B.3.1 Systematic uncertainties specific to the τ_{h}+jets final state for $\mathrm{H^+} \to \tau^+\nu_\tau$

In addition to the uncertainties described earlier in Sec. 7.11.1, some uncertainties affect the τ_{h}+jets final state in a specific way. In particular, some of the systematic uncertainties related to simulated samples also affect the background measurements from data. In the multijet background, a small number

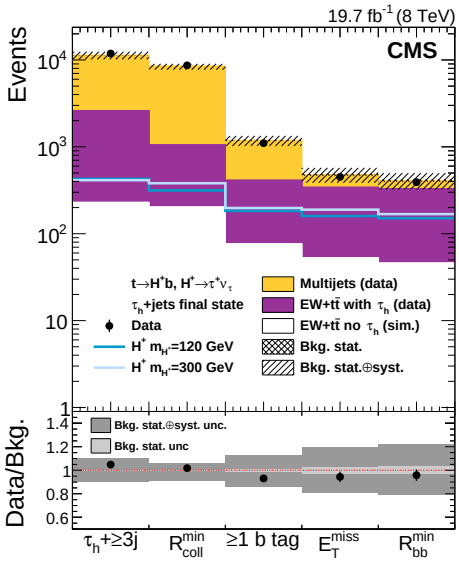


Figure B.1: The event yield after each selection step. For illustrative purposes, the expected signal yields are shown for $m_{\mathrm{H}^+} = 120$ GeV normalized to $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 0.01$ and for $m_{\mathrm{H}^+} = 300$ GeV normalized to $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1$ pb. The bottom panel shows the ratio of data over sum of expected backgrounds along with the uncertainties. The cross-hatched (light grey) area in the upper (lower) part of the figure represents the statistical uncertainty, while the collinear-hatched (dark grey) area gives the total uncertainty in the background expectation.

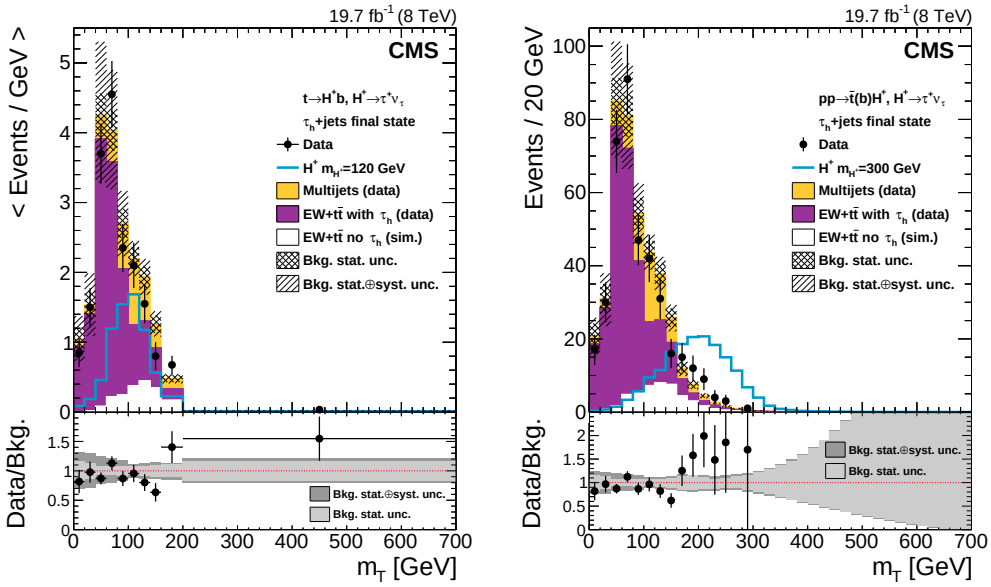
of simulated EW+$\mathrm{t\bar{t}}$ events is subtracted from the data to obtain the number of multijet events. The uncertainties affecting this small number of simulated events are taken into account, but their magnitudes are suppressed because they apply to only a fraction of the multijet background and a minus sign is assigned for them to denote anticorrelation. For the "EW+$\mathrm{t\bar{t}}$ with τ_h" background, uncertainties related to the simulated τ lepton decays are taken into account.

In addition to the uncertainties already described in Section 7.11.1, the following sources of systematic uncertainties are taken into account for the τ_h+jets final state:

- The uncertainties in the efficiencies of the τ part and $E_\mathrm{T}^\mathrm{miss}$ part of the τ+$E_\mathrm{T}^\mathrm{miss}$ trigger measured from data and simulation are considered separately. The simulated samples are affected by both

Table B.1: Numbers of expected events and their statistical and systematic uncertainties for the backgrounds and the number of observed events after the full event selection is applied. For illustrative purposes, the expected signal yields are shown for $m_{\mathrm{H}^+} = 120$ GeV normalized to $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 0.01$ and for $m_{\mathrm{H}^+} = 300$ GeV normalized to $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1$ pb.

	$N_\mathrm{events}(\pm\mathrm{stat.} \pm \mathrm{syst.})$
Signal, $m_{\mathrm{H}^+} = 120$ GeV	$151 \pm 4\,^{+17}_{-18}$
Signal, $m_{\mathrm{H}^+} = 300$ GeV	$168 \pm 2 \pm 16$
Multijet background (data)	$80 \pm 3\,^{+9}_{-10}$
EWK+$\mathrm{t\bar{t}}$ with τ_h (data)	$283 \pm 12\,^{+55}_{-54}$
EWK+$\mathrm{t\bar{t}}$ no τ_h (sim.)	$47 \pm 2\,^{+11}_{-10}$
Total expected	$410 \pm 12\,^{+57}_{-56}$
Data	392



(a) (b)

Figure B.2: The transverse mass distributions for the H^+ mass hypotheses of (a) 80–160 GeV and (b) 180–600 GeV. The event selection is the same in both (a) and (b) but in (b) the background expectation is replaced for $m_\mathrm{T} > 160$ GeV by a fit on the falling part of the m_T distribution. Since a variable bin width is used in (a) the event yield in each bin has been divided by the bin width. For illustrative purposes, the expected signal yields are shown for (a) $m_{\mathrm{H}^+} = 120\,\text{GeV}$ normalized to $\mathcal{B}(\mathrm{t} \to \mathrm{H}^+\mathrm{b}) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 0.01$ and for (b) $m_{\mathrm{H}^+} = 300$ GeV normalized to $\sigma(\mathrm{pp} \to \bar{\mathrm{t}}(\mathrm{b})\mathrm{H}^+) \times \mathcal{B}(\mathrm{H}^+ \to \tau^+\nu_\tau) = 1\,\text{pb}$. The bottom panel shows the ratio of data over sum of expected backgrounds along with the uncertainties. The cross-hatched (light grey) area in the upper (lower) part of the figure represents the statistical uncertainty, while the collinear-hatched (dark grey) area gives the total uncertainty in the background expectation.

sources of uncertainty, while the "EW+$\mathrm{t\bar{t}}$ with τ_h" background, obtained with the "embedding" procedure, is affected only by the uncertainty in the trigger efficiency measured in data. Furthermore, for the "EW+$\mathrm{t\bar{t}}$ with τ_h" background, the data part of the μ trigger efficiency is also considered, and a further 12% uncertainty is applied for approximating the $E_\mathrm{T}^\mathrm{miss}$ of the high-level trigger by offline calorimeter-based $E_\mathrm{T}^\mathrm{miss}$;

- The uncertainty in vetoing events with electrons and/or muons affecting only the simulated samples is estimated from the uncertainty in the electron and muon reconstruction, identification, and isolation efficiencies as 2% (1%) for electrons (muons);

- A 50% normalization uncertainty for the m_T distribution is assigned for the simulated single top quark samples in the "EW+$\mathrm{t\bar{t}}$ no τ_h" background for assigning as event weight the probability to pass b tagging instead of applying the b tagging condition;

- The uncertainties in the "EW+$\mathrm{t\bar{t}}$ with τ_h" background measurement method are described in the following. The uncertainty in the muon identification efficiency in data is found to be small. The contamination of the μ+jets control sample by multijet events is estimated with a μ enriched simulated multijet sample to be at most 2%, which is taken as a systematic uncertainty. The fraction of events with $\mathrm{W} \to \tau\nu_\tau \to \mu\nu_\mu\nu_\tau$, discussed in Section B.2.1, is evaluated from simulated events

and found to obey a functional form $1 - a \times p_T^{-b}$, where a and b are positive constants and p_{T} is the transverse momentum of the selected muon. The systematic uncertainty for correcting the event yield for this effect amounts to 1.2%. A 100% uncertainty is assumed on the event weights accounting for the difference between the τ+jets and embedded μ+jets events from simulated $\mathrm{t\bar{t}}$ events (denoted as "Non-emb. vs. emb. difference" in Table B.2) observed in the m_{T} distribution;

- The uncertainties in the multijet background measurement method are described in the following. The statistical uncertainty in the $E_{\mathrm{T}}^{\mathrm{miss}}$ template fit that is performed in each bin of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$, as described in Section B.2, is estimated to be 3% in each $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ bin. The difference in the m_{T} distribution shapes between the nominal sample and the sample with inverted τ_{h} isolation criterion is taken as a systematic uncertainty. It is evaluated from the ratio of the event yields of the samples with nominal and inverted τ_{h} isolation criterion as a function of m_{T} after requiring the other τ_{h} selection criteria, the veto against electrons and muons, at least three jets, and the requirement on $R_{\mathrm{coll}}^{\mathrm{min}}$. The statistical uncertainty of the ratio of the event yields is found to account for the difference in the shape and its magnitude is taken as the systematic uncertainty. Its value ranges between 5–15% depending on the bin of the m_{T} distribution.

A summary of the systematic uncertainties is shown in Table B.2.

In the region where the background yields are taken from the exponential fit on m_{T}, the statistical uncertainties in the background distributions are given by the uncertainties on the fit parameters while the relative values of the systematic uncertainties are kept the same like in the unfitted m_{T} distribution.

The dominant systematic uncertainties for signal arise from τ_{h} identification, τ_{h} energy scale, b tagging, and the theoretical $\mathrm{t\bar{t}}$ cross section uncertainty for $m_{\mathrm{H^+}} < (m_{\mathrm{t}} - m_{\mathrm{b}})$. For the backgrounds, the dominant uncertainties are those in τ_{h} identification, jet$\rightarrow \tau_{\mathrm{h}}$ misidentification, treatment of the $E_{\mathrm{T}}^{\mathrm{miss}}$ part of the trigger, and the difference between the transverse mass shapes of the τ+jets and embedded μ+jets events. In the region $m_{\mathrm{H^+}} > 300\,\mathrm{GeV}$ the sensitivity of the analysis is driven solely by the signal acceptance and the uncertainties in the signal.

Table B.2: The systematic uncertainties on event yields (in %) for the charged Higgs boson signal processes $\mathrm{t\bar{t}} \to \mathrm{bH^+\bar{b}H^-}$ ($\mathrm{H^+H^-}$), $\mathrm{t\bar{t}} \to \mathrm{bH^+\bar{b}W^-}$ ($\mathrm{H^+W^-}$), and $\mathrm{pp} \to \mathrm{\bar{t}(b)H^+}$ ($\mathrm{H^+}$) and for the background processes. The uncertainties which depend on the m_T distribution bin are marked with (S) and for them the maximum integrated value of the negative or positive variation is displayed. Empty cells indicate, that an uncertainty does not apply to the sample. The uncertainties on the rows are considered to be fully correlated and the uncertainties on the columns are considered to be uncorrelated. A minus sign in front of an uncertainty means anticorrelation with positive uncertainties on the same row.

	Signal $\mathrm{H^+H^-}$	Signal $\mathrm{H^+W^-}$	Signal $\mathrm{H^+}$	Multi-jets	EWK+$\mathrm{t\bar{t}}$ with τ_h	EWK+$\mathrm{t\bar{t}}$ no τ_h
τ part of trigger; data	1.5–1.8	1.3–1.5	1.8–3.0	-0.5	1.2	1.4
τ part of trigger; simulation	0.7–0.8	0.6–0.7	0.8–1.1	-0.2		0.8
$E_\mathrm{T}^\mathrm{miss}$ part of trigger; data	2.6–3.3	2.5–2.8	2.9–4.2	-1.2	2.5	2.8
$E_\mathrm{T}^\mathrm{miss}$ part of trigger; simulation	0.1	0.1	0.1	-0.1		0.4
Approximation in $E_\mathrm{T}^\mathrm{miss}$ part of trigger					12	
Single μ trigger; data					-0.1	
τ_h identification (S)	6.0	6.0	5.9–6.0	-0.8	6.0	
$\mathrm{e} \to \tau_\mathrm{h}$ mis-id (S)	$<$0.1	$<$0.1	$<$0.1	-0.1		3.3
$\mu \to \tau_\mathrm{h}$ mis-id (S)	$<$0.1	$<$0.1	$<$0.1	$<$-0.1		1.1
Jet$\to \tau_\mathrm{h}$ mis-id (S)	0.1	0.1–0.3	0.1	-6.9		17
Veto of events with e	0.1–0.2	0.2–0.3	0.2–0.3	$<$-0.1		0.4
Veto of events with μ	0.1	0.1–0.2	0.1	$<$-0.1		0.5
τ_h energy scale (S)	0.3–2.6	2.7–5.2	0.3–2.7	-1.8	5.8	2.0
Jet energy scale	2.6–5.2	2.0–3.0	1.6–2.1	-1.4		3.2
Jet energy resolution	1.1–1.8	0.5–1.3	0.7–1.5	-0.2		3.2
Unclustered $E_\mathrm{T}^\mathrm{miss}$ energy scale	0.1–0.4	0.1–0.9	0.1–0.4	-0.5		1.5
b-jet tagging (S)	5.9–20	4.7–5.3	4.6–5.4	-3.5		5.0
Probabilistic m_T in single top						6.8
Top quark p_T modelling (S)				$^{+5.6}_{-6.8}$		$^{+11}_{-6.6}$
Multijet m_T distribution shape (S)				4.6		
Multijet template fit				3.0		
μ identification; data					$<$-0.1	
Multijet contamination					2.0	
$\mathrm{W} \to \tau\nu_\tau \to \mu\nu_\mu\nu_\tau$ fraction					1.2	
Non-emb./emb. difference (S)					$^{+14}_{-12}$	
$\mathrm{t\bar{t}}$ cross section, scale	$^{+2.5}_{-3.4}$	$^{+2.5}_{-3.4}$		$^{+1.0}_{-0.7}$		$^{+2.2}_{-2.9}$
$\mathrm{t\bar{t}}$ cross section, PDF+α_S	4.6	4.6		-1.6		4.0
Single top cross section						1.0
W+jets, DY, VV cross section						0.1
Pileup modelling	0.1–0.9	0.1–0.8	0.1–0.6	-0.1		2.9
Luminosity	2.6	2.6	2.6	-0.8		2.6