University of California
Santa Barbara

Standard model four-top quark production in the all-hadronic final state in proton-proton collisions at 13 TeV with the CMS experiment

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Physics

by

Melissa Kathryn Quinnan

Committee in charge:

Professor Joseph Incandela, Chair
Professor Jeffrey Richman
Professor David Berenstein

March 2022

The Dissertation of Melissa Kathryn Quinnan is approved.

Professor Jeffrey Richman

Professor David Berenstein

Professor Joseph Incandela, Committee Chair

March 2022

Standard model four-top quark production in the all-hadronic final state in proton-proton collisions at 13 TeV with the CMS experiment

by

Melissa Kathryn Quinnan

To Thomas F. Quinnan.

Acknowledgements

I can honestly say that my Ph.D. journey would not have been possible without the love and support of my incredible colleagues, friends, and family. I am truly privaleged and grateful to have had you in my life over the course of the past five and a half years.

I would like to take this opportunity to thank those who have worked with me over the course of this very long and difficult analysis. Valentina, for working with me practically non-stop, meeting with me at all hours of the day and night and going over many a silly question and mistake. It was truly a pleasure working with you. Joe, for your always-useful insight and experience, and for not giving up on me when things got tough. Huilin, for your kindness and constant willingness to answer questions or have a chat. Jeff and David, for checking in on me and giving advice over the years. Phillip, who has been lovely to get to know and work with these past couple of months. Loukas, for your great support and advice on top taggers. Suyong, Jae, and Hayoung, for working with me from across the world, and troubleshooting many a complex issue these past few years. Meenakshi, Freya, Bob, Steve, Emmanuele, Nick, and Vichayanun, for working with me as colleagues on the combined four-top analysis and for your support in applying to postdoc positions and other opportunities. To you and my other four-top and CMS colleagues - this wouldn't be possible without your efforts, and for that you have my sincere gratitude.

I would also like to thank my friends, who have been at my side bringing joy and support throughout this time. Nicole and Gabi, for always being there both for adventures and as a shoulder to cry on. Sophia, Alex, Edoardo, Bruno, and Anders for being fantastic flatmates and friends, and Katja, Franzi, Vanessa, and others for the same at my home-away-from-home in Freiburg. Sean, Mason, Isa, Emily, Nathan, Kathryn, Jannicke, and Josh for your friendship and weekends spent skiing or relaxing by the lake, and

equivalently Alex, Zach, Emilie, Nick, Jamie, Dylan, and Bryant, for the weekends at the beach and hiking in the California sun. And of course Zack, Jaanki, Ramya, Katie, Sam, Kim, and Olivia for your love and support that extends before my Ph.D. and beyond. I look forward to continuing to have you all in my life in the years to come.

Finally, I would like to thank my family. My grandfather, who was always there to help with my math homework as a child and who I wish was still here to see me finish my graduate school experience, I hope you would be proud. My grandmother, for being my best friend and always being there for a chat and a laugh over the years. My parents, for loving and supported me from the very beginning. Kaleigh and Pat, who I have loved getting to know as adults. Anne, Ulf, Ole, and the Reyers and Bergmanns, for welcoming me into their family with open arms. And Max, for being there for this entire journey, struggles, pandemic, and all, and for being a constant source of stability and joy thoughout all of it, I owe you so much. Thank you.

Curriculum Vitæ

Melissa Kathryn Quinnan

Education

2022	Ph.D. in Physics (Expected), University of California, Santa Barbara.
2019	M.A. in Physics, University of California, Santa Barbara.
2016	B.S Physics (with honors), The Pennsylvania State University

Publications

- TOP-21-005 "Measurement of four-tops production in multiple final states using full Run 2" [https://cms.cern.ch/iCMS/analysisadmin/cadilines?line=TOP-21-005] (publication under review, anticipated 2022).
- M. G. Aartsen et al. [IceCube], "PINGU: A Vision for Neutrino and Particle Physics at the South Pole," J. Phys. G 44, no.5, 054006 (2017) doi:10.1088/1361-6471/44/5/054006 [arXiv:1607.02671 [hep-ex]].
- M. G. Aartsen et al. [IceCube], "The IceCube Neutrino Observatory: Instrumentation and Online Systems," JINST 12, no.03, P03012 (2017) doi:10.1088/1748-0221/12/03/P03012 [arXiv:1612.05093 [astro-ph.IM]].

Abstract

Standard model four-top quark production in the all-hadronic final state in proton-proton collisions at 13 TeV with the CMS experiment

by

Melissa Kathryn Quinnan

Standard model (SM) four-top quark production, $pp \rightarrow t\bar{t}t\bar{t}$, is a rare process with great potential to reveal new physics. Measurement of the cross section is not only a direct probe of the top quark Yukawa coupling with the Higgs, but an enhancement of this cross section is predicted by several beyond the standard model (BSM) theories. This process is studied in fully-hadronic proton-proton collision events collected during Run II of the CERN LHC by the CMS detector, which corresponded to an integrated luminosity of 137 fb^1 and a center of mass energy of 13 TeV. In order to optimize signal sensitivity with respect to significant and challenging backgrounds, several novel machine-learning based tools are applied in a multi-step and data-driven approach. Boosted decision tree (BDT) and deep neural net (DNN) based hadronic top taggers are used to identify hadronically decaying top quark candidates with moderate and high transverse momenta, respectively, in order to suppress backgrounds and categorize events by the multiplicity of reconstructed top tags, and an event-level kinematic BDT distribution is subsequently used to extract the signal. Control regions inspired by the "ABCD" method are used to obtain a data-driven estimate of the background, and data distributions in these control regions are given as inputs to a DNN in order to estimate the event-level BDT discriminant distributions of the major backgrounds. The observed signal strength μ, defined as the ratio of the observed four-top production rate to the standard model expectation, is measured to be $\mu = 5.1^{+2.3}_{-2.0}$. The corresponding observed (expected) significance and

limit times the SM cross section are 2.25 (0.43) σ and 8.39 (4.88) respectively. A combination of this result with multiple final states is in progress as of the writing of this thesis, and future BSM interpretations to investigate this larger-than-expected signal strength are planned.

Contents

Chapter 1

Introduction

This thesis summarizes an analysis of standard model four-top production ($pp \to t\bar{t}t\bar{t}$) in the fully-hadronic final state. Four-top quark production ($\mathrm{t\bar{t}t\bar{t}}$) is a rare process that serves as an important probe of both standard model (SM) and beyond the standard model (BSM) physics. Representative leading-order (LO) Feynman diagrams for $\mathrm{t\bar{t}t\bar{t}}$ production are shown in Fig. 1.1. The predicted cross section for SM $\mathrm{t\bar{t}t\bar{t}}$ production is $12.0^{+2.2}_{-2.5}$ fb at next-to-leading order (NLO) [1], which is very small compared to for example the $\mathrm{t\bar{t}}$ SM cross section of about 832^{+55}_{-64} pb [2]. It is sensitive to the Yukawa coupling of the top quark to the Higgs boson, but is also independent of the Higgs boson decay, making it an ideal process to reveal any deviations from SM predictions of the top quark Yukawa coupling [3]. Deviations of the four-top production cross section from SM predictions can also be caused by undiscovered heavy scalar (or pseudo-scalar) bosons decaying to top quarks, which are predicted in many simple extensions of the SM including two-Higgs doublet models (2HDM) [4].

Top quarks decay into a bottom quark and a W boson. The W then decays either hadronically, into quarks, or leptonically, producing a lepton and a neutrino. A CMS search targeting the same-sign dilepton and multi-lepton final states [5], which consti-

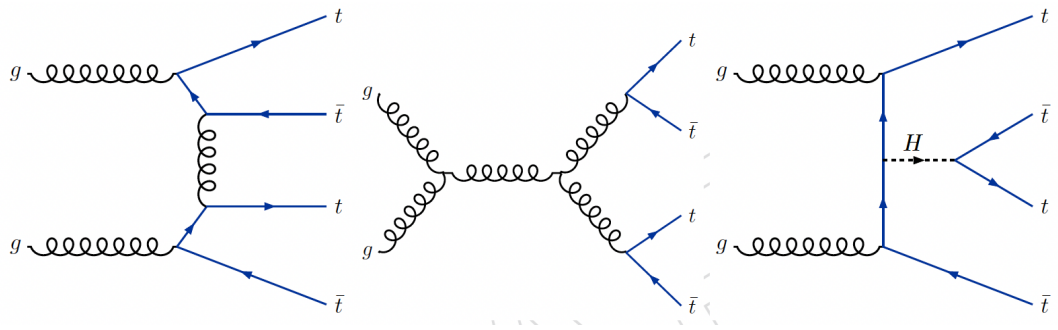

Figure 1.1: Representative LO Feynman diagrams for SM $t\bar{t}t\bar{t}$ production.

tute a combined branching ratio of roughly 10%, observed a significance of 2.7 standard deviations for SM $t\bar{t}t\bar{t}$ production using the Run II data set. The single-lepton and opposite-sign dilepton final states have also previously been explored with the CMS 2016 dataset [6]. The final state with four fully-hadronic top quark decays constitutes about 20% of all four-top quark events, and was studied for the first time in the analysis presented by this thesis.

Significant backgrounds from top quark pair ($t\bar{t}$) and QCD multijet production make the $t\bar{t}t\bar{t}$ search in the all-hadronic final state particularly challenging. Specialized machine-learning based hadronic top quark tagging tools to target the all-hadronic $t\bar{t}t\bar{t}$ decay mode are therefore exploited. Top quarks in the "boosted" (high-p_{T}) regime, characterized by collimated jets originating from the hadronization of top quark decay products, are identified with the DeepAK8 boosted object tagger [7]. Those in the "resolved" regime with moderate p_{T}, in which the hadronization of the top quark decay products results in distinct jets, are identified with a dedicated BDT-based resolved top tagger that was developed for this analysis. The search strategy relies on the classification of events into distinct categories based on the number of reconstructed tops and H_{T} (the scalar sum of jet p_{T}). An event-level BDT based on kinematic variables is then used for the extraction of $t\bar{t}t\bar{t}$ signal relative to background. Data-driven techniques are used to predict the main backgrounds in this search, which originate from $t\bar{t}$ and QCD

multijet production.

1.1 Permissions and Attributions

The analysis discussed in this thesis is the product of a collaboration between the author, Valentina Dutta and Joseph Incandela of UCSB, and Suyong Choi, Chang Whan Jung, Hayong Oh, and Jae Hyeok Yoo of Korea University. It is also part of a larger multi-channel four-top combination that involves Ulrich Heintz, Daniel Li, Meenakshi Narain, Nikolas Pervan, Sinan Sagir, Emanuele Usai, and Wenyu Zhang of Brown University, Steve Wimpenny, Bob Clare, and Nick Manganelli of the University of California Riverside, Freya Blekman of Vrije Universiteit Brussel/DESY, and Norraphat Srimanobhas and Vichayanun Wachirapusitanand of Chulalongkorn University.

Chapter 2

An Overview of Standard Model Physics

This section introduces the Standard Model (SM) and physics related to $t\bar{t}t\bar{t}$ production and particle collisions in general. It gives a brief overview of fundamental particles and their interactions as well as an introduction to relevant Beyond-the-Standard-Model (BSM) concepts. This chapter was written with reference to [8], [9], and [10].

2.1 An Introduction to Matter and Interactions

Particle physics aims to understand the nature and behavior of the tiniest irreducible building blocks of the universe. The current theory used to describe the fundamental particles that constitute matter and the way they interact is the Standard Model (SM). It is a consistent, finite gauge quantum field theory that describes particles and interactions in terms of quantized fields. While it has reliably predicted practically all experimental results to date, it cannot be a complete description of nature. It does not, for example, justify why neutrinos have mass, or provide an explanation for dark matter. As such,

much of particle physics research is dedicated to searching for new physics that could reveal a broader, more fundamental theory.

Figure 2.1 represents the SM in terms of the particles it predicts. Matter is made up of fundamental, point-like, massive spin 1/2 particles called fermions. These fermions interact (and experience forces) via the exchange of spin 1 force mediating bosons. Such interactions explain fundamental forces except gravity, which is extremely weak on sub-atomic scales.

The weakest force in the SM is, fittingly, the weak force. All fermions experience the weak force and so can interact via the exchange of neutral Z or charged W bosons. The electromagnetic force, well known for its generation of the electric fields and charges found in modern electronic devices, is mediated by the photon. All fermions that are electrically charged can interact via electromagnetism. Finally, the aptly named strong force is the strongest of these fundamental forces, and is (for example) the force that binds together atomic nuclei. Fermions affected by the strong force interact via the exchange of gluons as force mediators. The fundamental forces, their relative strengths, and relevant bosons are summarized in Table 1.1. The Higgs Boson is not a force mediator. It is a scalar boson with a mass of approximately 125 GeV [11] and a spin of 0, and is only subject to weak and self interactions. It is a part of the mechanism in electroweak theory by which particles, namely the W and Z bosons, acquire mass, as detailed later in this chapter.

Fermions can be subcategorized into quarks and leptons. Quarks can be "up-type", having a positive electrical charge of $+2/3$ times the charge of an electron e (the up, charm and top quarks) or "down-type" with a negative electrical charge of $-1/3$ e (the down, strange and bottom quarks). Given their electric charges, quarks interact electromagnetically, and like all fermions they participate in weak interactions. Quarks are also subject to the strong interaction, and so also have color charges ("red", "green" or

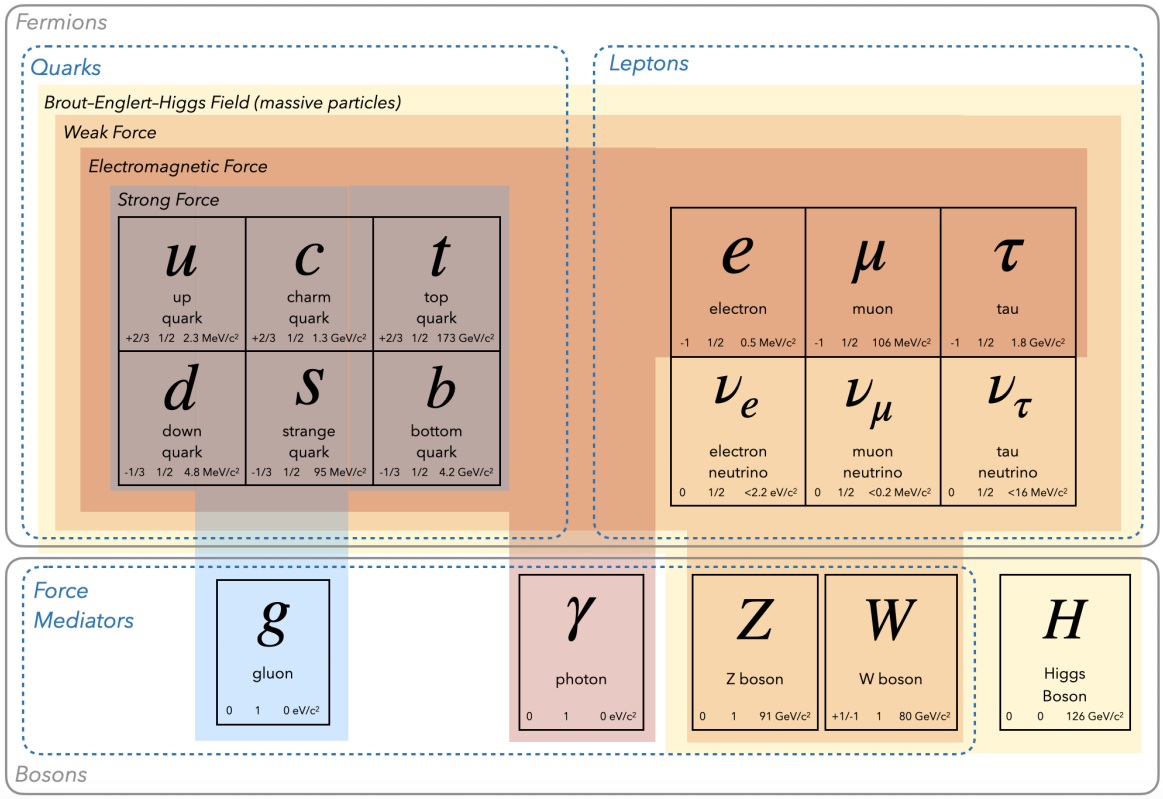

Figure 2.1: Particles of the Standard Model. The symbol, name, electrical charge, spin and mass of each particle is shown. Bosons and fermions are indicated by solid gray lines, and further categorized by dashed blue lines into force mediators and quarks/leptons, respectively. Particles that interact with the Brout-Englert-Higgs Field (and thus have mass) are contained within the yellow shaded box. Likewise, particles that interact with the strong, electromagnetic, and weak forces via the exchange of gluons, photons, and W/Z bosons are indicated with blue, red, and orange shaded boxes respectively. Fermions also have corresponding antiparticles with opposite quantum numbers (not shown).

"blue" as described by Quantum Chromodynamics, or QCD). Quarks form bound states of color neutral hadrons of two or more quarks. Most commonly these are bound states of a quark-antiquark pair ("mesons") or of three valence quarks ("baryons"). The top quark is unique in that it is the only quark with a short enough lifetime (about 5×10^{-25} seconds) [11] to decay before hadronization can occur, and so is not part of any bound state. It is also the heaviest particle in the SM, with a mass of 174 GeV [11]. These properties make the top quark an interesting probe of Higgs interactions (which couple

Table 1.1: Forces of the Standard Model. The four known fundamental forces and their associated force mediating bosons are shown. Interaction strengths are given as approximate values corresponding to two particles separated by the distance of roughly the radius of a proton (10^{-15}m [11]). Gravity's theoretical force carrier, the graviton, is shown, however gravity is only detectable at much larger scales and so no evidence of the graviton has yet been proven. Due to its extremely weak nature compared to the other forces, gravity is not relevant on the scale of subatomic particles.

Force	Strength	Boson	Mass (GeV)	Spin	Electrical Charge (e)
Strong	1	gluon	0	1	0
Electromagnetic	10^{-3}	photon	0	1	0
Weak	10^{-8}	W boson Z boson	80.4 91.2	1 1	+1/-1 0
Gravity	10^{-37}	graviton (?)	0	2	0

more strongly to heavier particles) as well as an opportunity to study an independent quark outside of a bound state. Another curiosity in the quark sector is the fact that the down quark is heavier than the up quark. In other generations the "up-type" quarks are heavier than their "down-type" partners. The reversal of this mass hierarchy SM pattern for first generation quarks is as of yet unexplained.

Leptons are another class of fermions, of which there are three generations of electrically charged leptons and neutral neutrinos. Charged leptons (electrons, muons, and taus) are impacted by weak and electromagnetic forces, and have electrical charges of -1e. Of these, the electron is the only stable lepton, although the relatively long lifetime of the muon (about 2×10^{-6} seconds [11]) means that at high energies it can be considered stable in an experimental context. The tau has a much shorter lifetime of about 3×10^{-13} seconds, and a mass larger than that of the lightest charged hadron, the pion (140 MeV) [12]. This means that taus, unlike other charged leptons, can also decay into hadrons via charged weak interactions involving the W boson.

Neutrinos are neutral leptons that have been observed to have very small masses (on the scale of a few eV [11]), despite the fact that the SM does not explicitly predict them

to be massive particles. This fact can be considered to be an observed deviation from SM expectations. There are three generations of neutrinos (the electron, muon and tau neutrinos) that correspond to their charged lepton partners. Neutrinos do not have color or electrical charges, and so only interact weakly.

Matter as we experience it is mostly made up of first generation particles (as opposed to antiparticles). The dominance of first generation matter is explained by the decays of more massive 2nd or 3rd generation fermions to their lighter 1st generation counterparts. However, it is not understood why (or indeed if) there are exactly three generations of fermions. The dominance of matter over antimatter in the known universe is also unexplained by the SM, which predicts by symmetry that both should have been initially produced in equal amounts in the big bang.

The SM can be represented mathematically as a quantum field theory using a Lagrangian density $\mathcal{L}_{SM}$, which is a function that summarizes fields and dynamics in a system. The SM Lagrangian $\mathcal{L}_{SM}$ obeys the gauge symmetry

$$SU(3) \times SU(2) \times U(1) \tag{2.1}$$

Here $SU(3)$ is the symmetry group for QCD describing strong interactions and $SU(2) \times U(1)$ corresponds to electroweak theory, which unifies the weak and electromagnetic interactions. The SM Lagrangian can be written in the condensed form

$$\mathcal{L}_{SM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\gamma_\mu D^\mu\psi + (y_{ij}\bar{\psi}_i\psi_j\phi + h.c.) + |D_\mu\phi|^2 - V(\phi) \tag{2.2}$$

The first term, $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ is the scalar product of the field strength tensor $F_{\mu\nu}$ for a gauge field A_μ with Lorentz indices μ and ν representing spacetime coordinates.

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu \tag{2.3}$$

This term leads to the existence of force mediating boson fields and their interactions (including self-interactions) and is formulated to be invariant under gauge transformations. Feynman diagrams for basic interaction vertices that follow from this term and others are summarized in Figure 2.2.

In the second term ($i\bar{\psi}\gamma_\mu D^\mu\psi$), ψ is the field of a fermion, $\bar{\psi}$ refers to its transposed complex conjugate, γ are the gamma matrices, and D is the covariant derivative, which preserves gauge invariance:

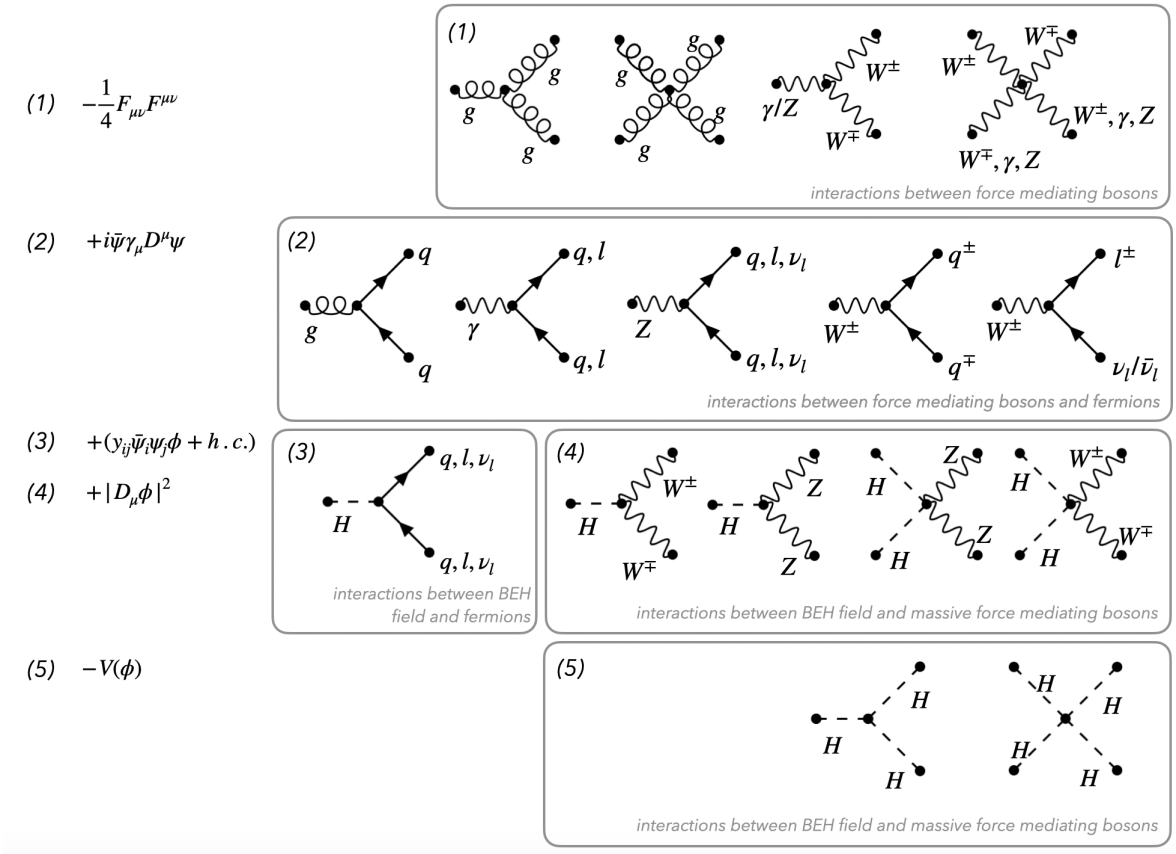

Figure 2.2: Interactions in the the standard model are shown according to their description as part of equation (2). Here g, γ, Z, and $W^\pm$ indicate force mediating bosons, H refers to the higgs boson, q to quarks (u,d,c,s,t,b), l to charged leptons (e, μ, τ), and ν_l to neutrinos. Charges and antiparticles are shown where necessary. Conjugates of each listed vertex (reversing the direction of arrows) is allowed.

$$D_\mu = \partial_\mu - igA^a_\mu t_a \tag{2.4}$$

Here t_a is the group generator in the chosen representation. This term describes interactions between force mediating bosons and matter particles (fermions).

The third term ($y_{ij}\bar{\psi}_i\psi_j\phi + h.c.$) decribes how fermions ψ couple with the Brout-Englert-Higgs (BEH) field, ϕ, and thus obtain mass. Here y_{ij} is the Yukawa matrix of couplings to the BEH field, which are directly related to particle masses. In this term $h.c.$ refers to the hermitian conjugate of $y_{ij}\bar{\psi}_i\psi_j\phi$, which describes the interactions between force mediators and antifermions.

The fourth term ($|D_\mu\phi|^2$) describes force mediators that couple to the BEH field (and thus are massive), which only applies to the W and Z bosons. Finally, the fifth and last term, $-V(\phi)$, describes the BEH potential and Higgs boson self-interactions. The Higgs boson itself is a quantized excitation of this field.

As the heaviest particle in the SM, the top quark also has the largest coupling to the BEH field. This makes processes containing top quarks ideal for measuring such couplings, namely the top-Higgs Yukawa coupling. More details on this coupling and the mechanics of Higgs physics is described in the next section.

2.1.1 Electroweak Theory and Spontaneous Symmetry Breaking

Electroweak (EW) theory unifies the description of electromagnetic and weak interactions and demonstrates how fermions and massive bosons (the W and Z bosons) acquire mass through spontaneous symmetry breaking. The EW Lagrangian can be written as

$$\mathcal{L}_{EW} = \mathcal{L}_{sym} + \mathcal{L}_H + \mathcal{L}_Y. \tag{2.5}$$

Here $\mathcal{L}_{sym}$ is the symmetric part describing EW gauge and fermion fields, $\mathcal{L}_H$ is the Higgs Lagrangian contribution that leads to the W and Z mass terms through spontaneous symmetry breaking (SSM), and $\mathcal{L}_Y$ is the Yukawa term that describes interactions between the BEH field and fermions by which fermions acquire mass. As noted in equation (1), this Lagrangian is the product of two symmetry groups $SU(2) \times U(1)$, and thus has two coupling constants and two sets of generators. The generator of the $U(1)$ group is the the weak hypercharge operator Y, and the generators of the $SU(2)$ group are the weak isospin operators T_1, T_2, and T_3. Quantum numbers for fermions are defined by eigenstates of these operators and satisfy the relation $Q = T^3 + \frac{Y}{2}$.

EW theory is a chiral theory where fermion fields are split into left and right-handed components such that left-handed fermions and right-handed antifermions form weak isospin doublets with (T,T_3) eigenvalues of (1/2,$\pm$1/2) and right-handed fermions form weak isospin singlets with an eigenvalue of $T = 0$. This leads to the absence of right-handed neutrinos or left-handed antineutrinos in the SM. The symmetric term of the electroweak Lagrangian is

$$\mathcal{L}_{sym} = \sum_j i\bar{\psi}_L^j \gamma_\mu D_L^\mu \psi_L^j + \sum_k i\bar{\psi}_R^k \gamma_\mu D_R^\mu \psi_R^k - \frac{1}{4}\sum_{a=1}^{3} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{2.6}$$

where $W_{\mu\nu}^a$, and $B_{\mu\nu}$ can be written in terms of the gauge fields of SU(2) and U(Y) (W_μ^a and B_μ, respectively), and D_L^μ, and D_R^μ are the covariant derivatives for the left and right-handed field components and can be written in terms of the Y and T generators and the two SU(2) and U(1) coupling constants (g and g', respectively):

$$W_{\mu\nu}^a = \partial_\mu B_\nu^a - \partial_\nu B_\mu^a - g\epsilon_{abc} W_\mu^b W_\nu^b, \tag{2.7}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.8}$$

$$D_\mu^L = \partial_\mu + ig \sum_{a=1}^{3} T_L^a W_\mu^a + ig' \frac{Y_L}{2} B_\mu, \tag{2.9}$$

$$D_\mu^R = \partial_\mu + ig' \frac{Y_R}{2} B_\mu \tag{2.10}$$

The gauge fields W_μ^a and B_μ lead to the definition of the physical bosons of EW interactions, namely the photon (γ), the W boson (W), and the Z boson (Z). They can be written using a SU(2) representation using the Pauli matrices such that $T_a = \sigma_a/2$. The W boson can be defined as a mixture of these gauge fields, requiring the electrical charge of $Q = \pm 1e$ and the weak hypercharge of $Y = 0$:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \pm iW_\mu^2) \tag{2.11}$$

Likewise, the photon and Z fields can be described by the orthogonal, normalized mixture of B_μ and W_μ^3 as a function of the weak mixing angle θ_W given by $tan(\theta_W) = g'/g$:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} cos(\theta_W) & sin(\theta_W) \\ -sin(\theta_W) & cos(\theta_W) \end{pmatrix} . \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{2.12}$$

While the symmetric term of the EW Lagrangian $\mathcal{L}_{sym}$ describes electroweak interactions and the physical photon, W and Z bosons, it does not indicate that fermions or any gauge bosons have mass. This contradicts observations that quarks and leptons and the W and Z bosons are massive. The mass terms for these gauge bosons (excepting the massless photon) arise from the Higgs Lagrangian $\mathcal{L}_H$ and "spontaneous symmetry breaking" (SSB).

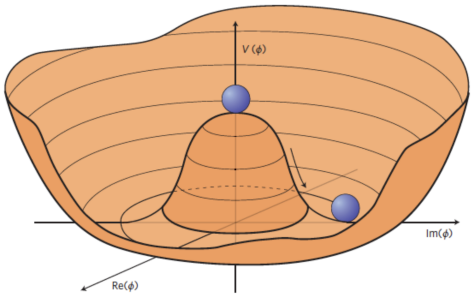

Figure 2.3: An illustration of the Higgs potential $V(\phi)$. Choosing any point at the potential's minimum (visualized by the purple sphere moving from the 'sombrero' top to the lower point) has the effect of breaking the symmetry. [13]

The Higgs Lagrangian is defined as:

$$\mathcal{L}_H = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) D_\mu = \partial_\mu + i\frac{g}{2}\sum_{a=1}^{3}(\sigma_a W_\mu^a) + i\frac{g'}{2}B_\mu \tag{2.13}$$

Where $V(\phi)$ is the Higgs potential:

$$V(\phi) = -\frac{\mu^2}{2}\phi^\dagger\phi + \frac{\lambda}{4}(\phi^\dagger\phi)^2 \tag{2.14}$$

Here μ^2 and λ are positive constants and ϕ represents non vanishing scalar fields. As visualized in Figure 2.3, the symmetrical shape of this potential is spontaneously broken (hence SSB) when the fields ϕ are at a minimum, which occurs when

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} v = \frac{2\mu}{\sqrt{\lambda}} \tag{2.15}$$

Here v is the vacuum expectation value. After perturbative expansions about v the scalar fields and Higgs potential can be written in terms of the real Higgs field H:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v+H \end{pmatrix} V(H) = \mu^2 H^2 + \frac{\mu^2}{v}H^3 + \frac{\mu^2}{4v^2}H^4 \tag{2.16}$$

Here the mass of the Higgs boson can be identified as $m_H = \sqrt{2}\mu^2$ and the triple and quadruple Higgs vertices are (m_H^2/v and m_H^2/v^2). While the mass of the Higgs is established here, the first part of the Higgs Lagrangian ($(D_\mu\phi)^\dagger(D^\mu\phi)$) post-symmetry-breaking is what leads to the masses of the Z and W bosons being defined as:

$$m_Z = \frac{1}{2}\sqrt{g^2 + g'^2}v \tag{2.17}$$

$$m_W = \frac{1}{2}gv = cos(\theta_W)m_Z \tag{2.18}$$

This leaves the last part of the Lagrangian that leads to fermion masses through interactions with the scalar Higgs field: the Yukawa Lagrangian $\mathcal{L}_Y$. This can be written in terms of left-handed weak isospin doublets (ψ_L) and right handed singlets (ψ_R) and M, the non-diagonal fermion mass matrix generated by the coupling to the Higgs field. This is discussed in terms of quarks here but leptons behave similarly.

$$\mathcal{L}_Y = -\bar{\psi}_L M \psi_R - \bar{\psi}_R M^\dagger \psi_L - \bar{\psi}_L \frac{M}{v}\psi_R H - \bar{\psi}_R \frac{M^\dagger}{v}\psi_L H \tag{2.19}$$

After diagonalizing the mass matrix M and transforming the doublets and singlets, this leads to a mixing of mass eigenstates between left handed fermion doublets and charged gauge fields (represented as amplitudes $|V_{ij}|$) by the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.20}$$

While the matrix shown here is for quark flavor mixing, an analogous matrix exists

for leptons.

As an example, the emergence of fermion masses can be demonstrated more clearly for top and bottom quarks in particular while neglecting flavor mixing. Here, given the left-handed weak isospin doublets and right handed singlets of top and bottom quarks, the Yukawa coupling strengths $\lambda_{t,b}$, and a modified Higgs doublet ψ_u for up-type quarks $\psi_u = i\sigma_2\psi*$ the Yukawa Lagrangian can be written as

$$\mathcal{L}_{Y,t/b} = -\lambda_t(\bar{\psi}^t_L, \psi^b_L)\phi_u\psi^t_R - \lambda_b(\bar{\psi}^t_L, \psi^b_L)\phi\psi^b_R \tag{2.21}$$

Notably, the modified Higgs doublet ψ_u is replaced by two separate doublets in many extentions of the SM, including in the minimal supersymmetric SM (MSSM) and other "two-higgs-doublet-models" (2HDM) [4].

Applying instead the SSB Higgs doublet reveals the mass terms of the top and bottom quarks m_t and m_b as a function of the vaccuum expectation value v and the top and bottom Yukawa couplings λ_t and λ_b.

$$\mathcal{L}_{Y,t/b} = -\frac{1}{\sqrt{2}}(\lambda_t\bar{\psi}^t\psi^t + \lambda_b\bar{\psi}^b\psi^b)(v + H) \tag{2.22}$$

$$= -\lambda_t\frac{v}{\sqrt{2}}\bar{\psi}^t\psi^t - \lambda_b\frac{v}{\sqrt{2}}\bar{\psi}^b\psi^b - \lambda_t\frac{1}{\sqrt{2}}\bar{\psi}^t\psi^t H - \lambda_b\frac{1}{\sqrt{2}}\bar{\psi}^b\psi^b H \tag{2.23}$$

$$= -m_t\bar{\psi}^t\psi^t - m_b\bar{\psi}^b\psi^b - \lambda_t\frac{m_t}{v}\bar{\psi}^t\psi^t H - \lambda_b\frac{m_b}{v}\bar{\psi}^b\psi^b H \tag{2.24}$$

Here it can be seen that the mass terms are proportional to the Yukawa couplings $m_{t/b} = \lambda_{t/b}\frac{v}{\sqrt{2}}$. With the largest mass in the SM, this is why the top quark is such a powerful experimental tool for probing the top-Higgs Yukawa coupling and testing and measuring the mechanics of the SM.

2.1.2 Quantum Chromodynamics and the Fine Structure Constant

While EW theory gives rise to most of the particles predicted by the SM, it is missing a description of the strong force that leads to interactions between quarks in order to form bound hadronic states. The mathematical formulation of strong interactions in quantum field theory is called Quantum Chromodynamics (QCD). Like EW theory, QCD involves generators following from a symmetry group, but in this case the group is a non-abeliean unitary group $SU(3)_C$ that unlike EW theory with a single electric charge, has three analogous color charges named "red", "green", and "blue". This leads to eight generators T_C expressed by noncommuting Gell-Mann matrices λ_C as $T_C = \lambda_C/2$.

For a particular quark with color charge c and mass m the QCD Lagrangian $\mathcal{L}_{QCD}$ is

$$\mathcal{L}_{QCD} = \bar{\psi}_c(i\gamma^\mu D_\mu - m)\psi_c - \frac{1}{4}G^c_{\mu\nu}G^{\mu\nu}_c \tag{2.25}$$

Where $D_\mu = \partial_\mu - ig_s T_C G^C_\mu$ is the covariant derivative, $G^C_{\mu\nu} = \partial_\mu G^C_\nu - \partial_\nu G^C_\mu - g_s f^C_{ab} G^a_\mu G^b_\nu$ are the eight masslass gluon fields, f^C_{ab} are the antisymmetric $SU(3)_C$ structure constants, and g_s is the strong coupling constant. The first part of this Lagrangian ($\bar{\psi}_c(i\gamma^\mu D_\mu - m)\psi_c$) leads to quark and gluon propagators and the gluon-quark-antiquark ($g\bar{q}q$) vertex. The second part ($-\frac{1}{4}G^c_{\mu\nu}G^{\mu\nu}_c$) leads to gluon self-couplings and corresponding triple and quadruple gluon vertices.

The intrinsic strength of the strong coupling constant compared to that of electromagnetism or the weak force is usually reported in terms of a dimensionless coupling strength α. Thus interaction probabilities include a single α per interaction vertex and the value of α independent of units used. For electromagnetism, this coupling strength is the fine structure constant $\alpha = e^2/(4\pi\epsilon_o\hbar c) = 1/137$ [8]. The strong force has a greater coupling strength of approximately $\alpha = 1$ at low energies. The weak force actually has a

greater coupling strength than electromagnetism (about $\alpha = 1/30$) but the large mass of the W boson means that at low energy scales the weak interaction is much weaker than electromagnetic interactions.

Perturbation theory is required to derive physics predictions for interactions at high energy scales (high momentum transfer Q^2 at short distances) for QCD, due to the nature of the $SU(3)_C$ symmetry and the gluon self-couplings. This means that calculations require regularization and normalization and therefore an arbitrary mass scale μ_R. In practice this means that at high energy scales the effective "running" coupling strength decreases with increasing energy scales Q^2, as α_s is proportional to $1/ln(Q^2)$. This is commonly referred to as "asymptotic freedom". At lower energies on the other hand, where perturbation theory is not available, the coupling strength is so large that the amount of energy needed to create bound quark pairs is lower than that needed to escape the QCD potential. This leads to the lack of free quarks observed in nature and is referred to as color-confinement. It also means that at in high energy particle collisions like those that take place at the LHC, quarks are initially produced in quasi-free states near primary proton-proton collision vertices (PV) but within a short space and time they transition from a high-energy state of asymptotic freedom to a low-energy bound state of color confinement as they hadronize within particle detectors.

2.2 Beyond the Standard Model

This section gives a brief introduction to Beyond the Standard Model (BSM) physics that is interesting in the context of four-top production. Previous sections already hinted that considering two Higgs doublets rather than one is the basis of the MSSM and 2HDMs. This will be discussed briefly here.

As mentioned earlier in this chapter, the SM offers no explanation of a few notable

phenomena. The concepts of dark matter and energy, for example, arise from astronomical observations that roughly 85% of the mass in the universe is matter or energy that cannot be directly observed [8]. As of yet, no explanation for this extra mass has been proven. Another curiosity unexplained by the SM is the fact that neutrinos have mass [8]. The observation of neutrino oscillations necessitates that contradictory to SM predictions, neutrinos must be massive, albeit with only small masses. There are other inelegant aspects of the SM, for example its failure to incorporate gravity, and SM parameters need to be "fine-tuned" in order to be consistent with the cutoff at which gravity becomes relevant and the SM no longer holds (the Planck scale at $10^{19} GeV$ [8]). As a result, measurements that test the limits of the SM or probe for new physics are of particular interest in particle physics.

Supersymmetry (SUSY) [14] is one proposed solution to many of these problems. It supposes there is a symmetry between fermions and bosons, and thus that every SM particle has a corresponding superpartner with a spin differing by 1/2 between fermions and bosons. The existence of such particles would to some extent address the issue of fine-tuning and may also provide dark matter candidates.

The simplest supersymmetric extension of the SM is the MSSM, which requires two Higgs doublets as an example of a 2HDM [4]. Because in supersymmetric theories scalars and their complex conjugates belong to chiral multiplets of opposite chirality that cannot couple together in the Lagrangian, a single Higgs doublet would not be able to give mass to up and down type quarks simultaneously and so an additional Higgs doublet would be required. This would result in five Higgs-like scalar or pseudoscalar bosons rather than one, which may have not yet been detected at current energy scales probed by particle accelerators. However, if these extra bosons decay into other particles, for example into multiple top quarks, an excess of signal in relevant SM processes could provide evidence for their existence. An excess of four-tops beyond SM expectations, for example, could

suggest the validity of such 2HDMs and so is a common BSM physics objective of four-top production measurements. The scope of this thesis, however, will focus on four-top production in the context of the SM.

Chapter 3

The LHC and the Compact Muon Solenoid

This chapter introduces the CMS (Compact Muon Solenoid) experiment and its collection of particle collision data, including the functionalities and responsibilities of its various subdetectors and reconstruction capabilities.

3.1 The Physics of Hadron Colliders and the LHC

Particle colliders like the Large Hadron Collider (LHC) collide particles within detectors in order to study the products of these collisions. These collisions are typically induced by two beams of charged particles that are accelerated in opposite directions until they reach a desired kinetic energy, at which point they are allowed to cross at the desired collision point within particle detectors. The LHC, specifically, is a circular collider that accelerates bunches of protons through the use of magnets that accelerate and focus the beams. These bunches contain billions of protons each that are accelerated along a 27 kilometer ring and cross each other in detectors at rates of up to 40 million

times per second [15]. The LHC is the largest and highest-energy particle accelerator in the world, and is located at CERN (the European Center for Nuclear Research) on the border between France and Switzerland. Data taking during Run II of the LHC, which is relevant for this thesis, took place between 2015 and 2018 with a center of mass energy of $\sqrt{s} = 13TeV$.

A core quantity in accelerator physics is the concept of the cross section of a given process. Qualitatively, it can be thought of as essentially a measurement of the probability that a certain process will occur, reported as a quantity that is independant of the energy or intensity of particle beams and so can be compared for a given process across experiments [16], [17]. Processes that are more likely to occur in particle collisions therefore have larger cross sections compared to those less likely to occur. The units of cross sections are in area (barns). When referring to a cross section most often it is the total cross section that is refered to, or the sum over all possibilities for a given process, including all scattering angles and other variables. A differential cross section, on the other hand, is a cross section measurement that is given as a differential limit for a particular final state variable such as a particular energy or scattering angle.

The luminosity of a particle accelerator is a quantity that measures an accelerator's ability to produce the required number of particle interactions. It is, in other words, the proportionality factor between the interaction rate (the number of events per second, or $\frac{dN_{process}}{dt}$) and the cross section of a given process ($\sigma_{process}$) and can be written as follows [18]:

$$\frac{dN_{process}}{dt} = L \cdot \sigma_{process} \tag{3.1}$$

As hinted by this formula, luminosity as a quantity is a function of area and time, with the luminosity of LHC beams on the order of $10^{-34} cm^2 \cdot s^{-1}$ [15].This can also be

written in terms of barns, which are a unit of area defined as 1 barn $(b) = 10^{-24} cm^2$. In these terms, the LHC has a design luminosity on the order of $100 fb^{-1} \cdot s^{-1}$.

The quantity (L) is sometimes specified as the instantaneous luminosity rather than just "luminosity" to distinguish it from the total luminosity aquired over time, which is instead called the integrated luminosity L_{int}. Its formula is the integral of the instantaneous luminosity over time:

$$L_{int} = \int L \cdot dt \tag{3.2}$$

It follows, then, that units of integrated luminosity are often given in inverse barns b^{-1}. This is a useful quantity for calculating the number of events of a given process $(N_{process})$ that can be expected in particle collisions:

$$N_{process} = L_{int} \cdot \sigma_{process} \tag{3.3}$$

Figure 3.1 shows the integrated luminosity vs. time collected by the LHC and the CMS detector during Run II. This thesis focuses on data taken by the CMS experiment between 2016 and 2018, which corresponds to an integrated luminosity of $137 fb^{-1}$.

3.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) experiment is one of two general purpose detectors at the LHC (the other being the ATLAS detector) specialized to detect high energy collisions at the LHC. It involves one of the largest scientific collaborations in the world, which includes over 5000 scientists and engineers from about 200 institutes and nearly 50 countries [20]. The detector itself is pictured in Figure 3.2. It is located about 100 meters underground at a point on the LHC near Cessy, France. It is cylindrical in shape,

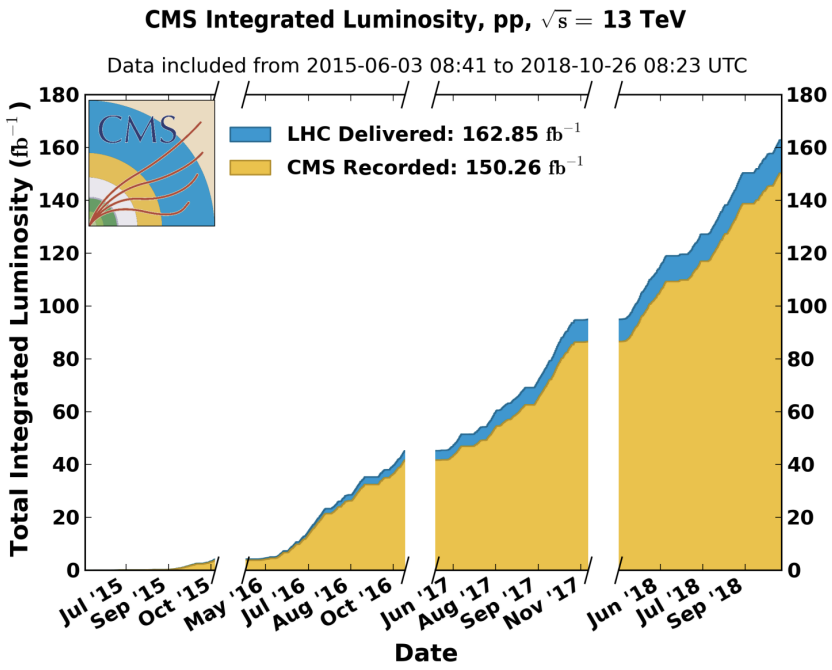

Figure 3.1: Integrated Luminosity recorded in Run II (2015-2018) vs. time recorded by the LHC (blue) and by the CMS experiment (yellow). Gaps indicate shutdowns during which no data was taken. [19]

with a length of 21.6 meters and a diameter of 14.6 meters and a weight of roughly 14 thousand tons [20], [21]. The detector is designed to measure the energies and trajectories of particles resulting from LHC collisions within dedicated subdetectors that are arranged radially around the beam pipe and which are designed to identify particle types and properties. The functionality of each of these subdetectors will be discussed briefly in the following subsections. The CMS technical design report [20], [21] was used as a reference for these sections.

3.2.1 The Solenoid Magnet

One of the main design features of CMS is its 3.8T superconducting solenoid magnet, which is the largest of its kind in the world. The "compact" aspect of CMS refers to the fact that many of the subdetectors (like the tracker and calorimeters) are contained

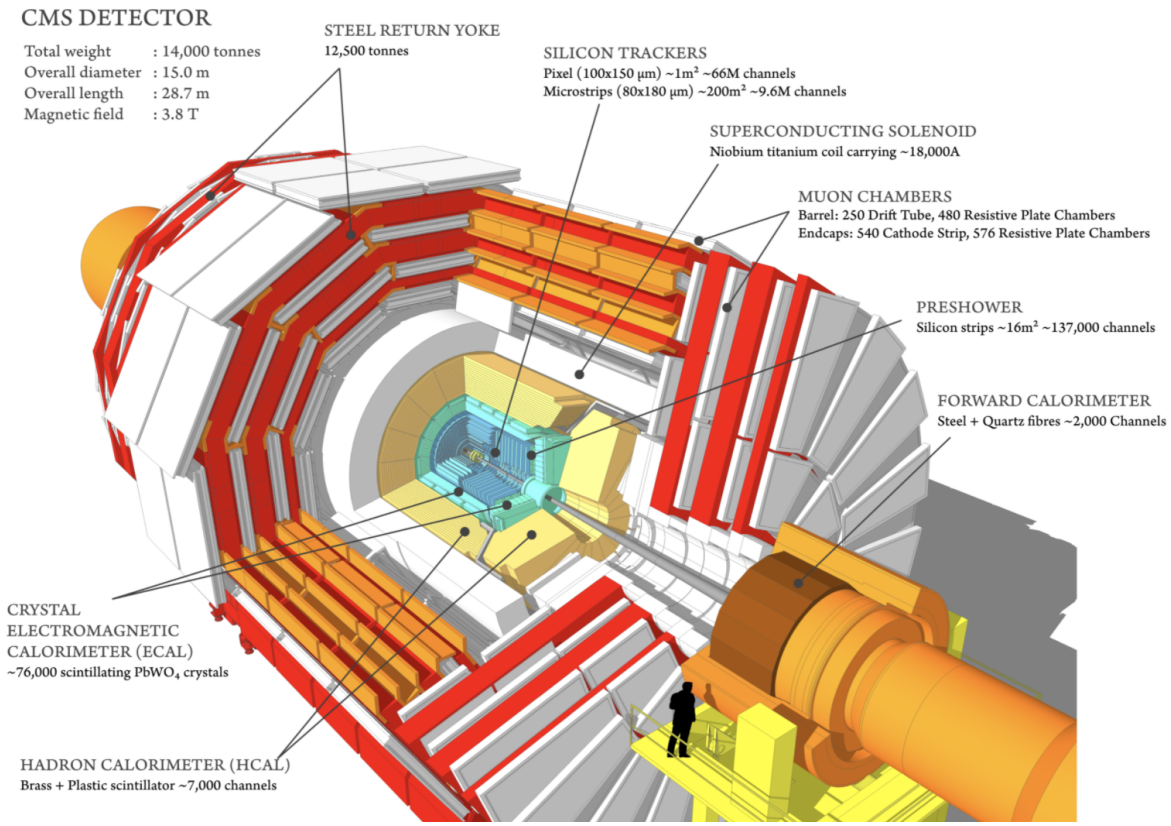

Figure 3.2: Schematic diagram of the CMS detector with labeled subdetectors. [20]

within this magnet, which bends the trajectories of charged particles in order to gain information about their charges and momenta. It has the shape of a hollow cylinder of length 12.5 meters and a diameter of 6.0 meters, and is formed out of a superconducting niobium-titanium coil to generate the magnetic field. This is surrounded by three layers of iron return yoke layers to better contain the magnetic flux within the detector volume. The coils must be kept at a chilly 4.7K temperature in order to maintain their zero-resistance superconducting electrical properties [20], [21].

3.2.2 The Silicon Tracking System

The first subdetector encountered by particles produced in LHC collisions within the CMS detector is the tracking system. It operates based on the formation of electron-hole pairs in silicon semiconductors that form when charged particles pass through. Currents from these electron-hole pairs are digitized by readout electronics, allowing for the algorithmic reconstruction of the the precise location and direction of charged particle tracks and vertices. As it is so close to the interaction point where LHC beams collide, this system must cope not only with very high particle fluxes and therefore complex track reconstruction requirements, but also with large doses of radiation experienced during long operation periods. As such, the silicon detectors are designed to have high resolution and efficiency and to be radiation hard, which is achieved in part by cooling silicon detector components to -10 degrees Celsius.

The subdetector is itself made up of several parts as shown in Figure 3.3. The innermost part of the tracking system is the pixel detector. This included two endcap disk layers and three inner layers between transverse cylindrical radii of 4.4 and 10.2 cm until 2017, when they were replaced by three endcap layers and four inner layers between transverse cylindrical radii of 3.0cm and 16.0cm for better resolution. As the closest detector to the interaction point of colliding LHC beams, the pixel detector requires very high granularity in order to resolve the large number of simultaneously transversing particles into reconstructable tracks and vertices. To accomplish this, this innermost subdetetector is made up of $66 \cdot 10^6$ tiny silicon pixels of size $100 \cdot 150 \mu m^2$. All together, these pixels form a total active area of about a square meter.

The outer part of the tracking system surrounding the pixel detector is made up of silicon microstrips rather than pixels. It includes four parts: the tracker inner barrel (TIB) encapsuling the pixel detector, the tracker outer barrel (TOB) around that, and

on either side the inner disks (TID) followed by the endcaps (TEC). These subdetectors still require enough resolution to reconstruct particle tracks and vertices coming from the nearby beam line but do not encounter particle fluxes as high as the inner pixel detector, hence the use of strips rather than pixels. These strips number a total of $9.3 \cdot 10^6$ and cover an active area of roughly 198 square meters. The TIB has four layers and is about 130m long. Just outside of it (radially) is the 6 layer TOB. In the forward regions, the TID has three layers and the TOB nine. All together the CMS tracking system has an active area of about 200 square meters and forms a cylinder that is about 5.4m long with a diameter of about 2.4m, making it the largest silicon device ever constructed.

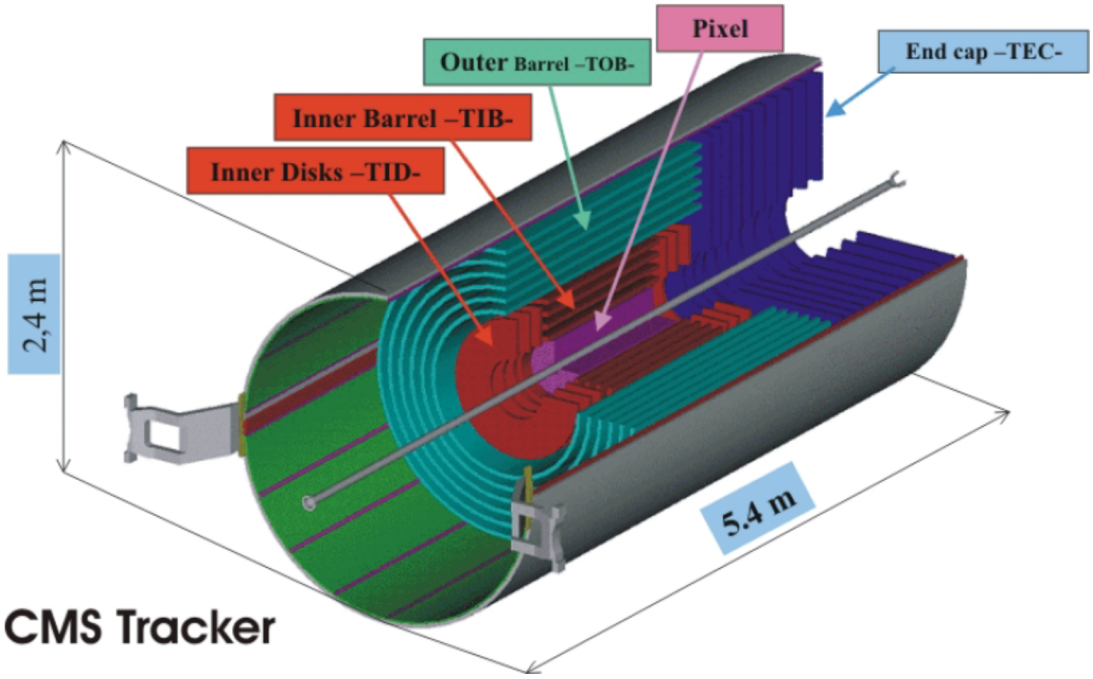

Figure 3.3: Diagram of the CMS tracker, including pixel detector, inner barrel (TIB) and inner disk detectors (TID), the outer barrel (TOB), and endcap trackers (TEC- and TEC+). Note that the endcaps exist on both sides (TEC+ is not shown). The gray line in the center represents the beam pipe [22].

3.2.3 The Electromagnetic Calorimeter

Just outside the tracking system and also enclosed by the solenoid magnet is the electromagnetic calorimeter (ECAL). This subdetector is made up of nearly 70 thousand lead tungstate ($PbWO_4$) clear scintillating crystals of size $2.86 \cdot 2.86 \cdot 22.0$ cm^2 in the endcap region and $2.2 \cdot 2.2 \cdot 23.0$ cm^2 in the barrel region of the ECAL. When charged particles pass through these crystals they give off scintillation light with a brightness corresponding to the energy of the particle. The very dense crystals also seek to slow and eventually absorb the particles. When the light is amplified by photodiodes/phototriodes and digitized this allows for an electronic readout of the location and energy of charged particle deposits. The small Moliere radius of electromagnetic showers (about 2.2 cm) and the small radiation length of lead tungstate (0.85 cm, or about 25 per crystal) means that electromagnetic showers are often longitudinally contained within a single layer of crystals and transversally contained within a single crystal. This enables high resolution and efficiency in detecting charged particles, especially photons and electrons.

3.2.4 The Hadronic Calorimeter

The next subdetector, sandwiched mostly between the solenoid and the ECAL, is the hadronic calorimeter. This subdetector is designed to absorb and meaure the energy of strongly interating particles, both charged and uncharged. It is a sampling calorimeter orginized into four regions: the barrel calorimeter (HB), the endcap calorimeter (HE), the outer calorimeter (HO) and forward calorimeter (HF). All of these regions except the HF are made up of alternating layers of brass absorbers and plastic scintillators. Particles create hadronic showers in the brass absorber layers which in turn cause scintillation light in the plastic scintillator layers which is detected and read out in order to detect the location and energy deposit of hadronic showers. The HF, which is located on both

sides of the detector outside of the solenoid and muon endcaps, is made of steel and quartz fibers and is designed to produce narrow hadronic showers in the forward regions along the axis of the beam line. The HO is the only other part of the HCAL situated outside of the solenoid magnet and is tasked with detecting the tails of any hadronic showers that entend beyond the solenoid. The HB and HF detect hadronic showers inside of the solenoid magnet, and are situated around the radial area just outside of the ECAL (HB) as well as capping both ends of the cylinder outside of the ECAL endcap.

3.2.5 The Muon Chambers

The muon chambers are located outside of the solenoid and are the outermost subdetector of CMS. They are designed to detect the charge and momenta of muon tracks that can be matched up to information from the tracker. The chambers detect muons using three different types of gaseous detectors, which output electrical signals when charged particles (namely muons) pass through and ionize the gas. These are the drift tubes (DTs) in the barrel region, cathode strip chambers (CSCs) in the endcaps, and resistive plate chambers (RPCs) in both the barrel and endcap regions.

There are four layers of DTs in the barrel, arranged in concentric cylinders and made up of a total of 250 individual chambers. They use an ionizing gas mixture of argon (Ar) and carbon dioxide (CO_2) and detect ionization using gold plated steel wires. These wires alternate in direction to be perpendicular or parallel to the beam line in order to localize muon signals in that plane as they traverse the different radial detector layers. The DT layers are interspersed with five RPC layers, which are simple gaseous detectors involving a potential and two parallel metal plates. While not as good at spatial resolution as the DTs and CSCs, these RPCs provide excellent time resolution on the order of a few nanoseconds.

There are a total of 486 CSC detectors on the endcaps, which deliver excellent spatial resolution (on the order of 100 μm) of charged muons traveling in the direction of the beamline. These are gaseous chambers containing 6 anode wire layers alternated with 7 cathode strip panels in each. The 5 layers of CSCs on each side are also interspersed with 3 RPC layers for better timing resolution.

3.2.6 The Trigger System

Coping with data rates in CMS is a daunting task. It would be impossible to store information from every collision within CMS: with about one billion proton-proton collisions occuring per second, this would mean about 1000 terabytes of data per second [21], [20], [23]. The CMS trigger system is therefore tasked with filtering out less interesting events in order to reduce this data rate by a factor of 10 million to about 100 megabytes per second. It employs two tiers: a level-1 (L1) hardware trigger and a more sophisticated high-level-trigger (HLT).

The L1 trigger is based on FPGAs and calculates basic momentum and position information for objects such as muons, electrons, jets, and energy sums in order to make quick decisions to keep or reject events. Its goal is to reduce data rates by a factor of about a million to a rate of about 100 kHz or less. To do this the L1 trigger hardware makes decisions in less than 1 μs [21].

Once passing the L1 trigger, events are passed to the HLT trigger, which uses more precise and complex trigger decisions that involve partial event reconstruction and information from the inner tracker. This is a software trigger that runs on a computer farm close to the detector. The goal of the HLT is to reduce data rates to less than one kHz. Events that pass the HLT trigger are subsequently passed to data storage and processing for reconstruction and analysis.

3.3 Reconstruction and Definitions of Physics Objects

The raw output of the CMS experiment involves data events filtered by the L1 and HLT triggers and containing particle locations ("hits") and energy deposits from the various subdetectors. This information must be reconstructed to form a hypothesis of the particle types, momenta, energies, and tragectories within an event. This is done using reconstruction algorithms, namely the Particle Flow (PF) algorithm. This algorithm reconstructs the primary interaction vertex (PV) of proton-proton collisions, the sum of the present and missing transverse momenta (MET) in an event, as well as analysis objects such as electrons, photons, muons, hadronically decaying taus, and jets. Other more complex analysis objects, such as hadronically decaying bottom and top quarks, are reconstructed using other higher-level algorithms.

The PF algorithm is described in detail in reference [24]. It first collects and groups information from the various subdetectors into preprocessed PF objects, including tracks of charged particles from the tracker, clusters of energy deposits in the ECAL and HCAL, and tracks within the muon chambers. It then uses an algorithm to spatially link this information, before deriving reconstructions of analysis objects, for example particles, primary verticess, and jets.

3.3.1 Primary Vertices (PVs)

The primary vertex (PV) refers to the reconstructed location of the proton-proton collision of an event of interest. In order to reconstruct these vertices, first tracks in a given event are identified and selected by a track finding algorithm based on their goodness of fit and the number of hits in the tracker. Tracks are then clustered and

matched to the closest collision vertices along the beam line. The PV is identified as the vertex that, when fit to these clustered tracks, yields the highest sum of associated scalar transverse momenta. Other unimportant proton-proton interaction vertices are discarded, along with their associated particles and tracks.

3.3.2 Muons

Muons are distinctive particles with long-lived charged particle tracks that curve within the magnetic field from the solenoid magnet and often travel through the entire detector volume. They are reconstructed using information from the tracking system and the muon chambers using two different reconstruction techniques. The first type of PF muon is the tracker muon. These must have charged particle tracks in the silicon tracking system that are identified as likely muon candidates that can be matched to tracks in the muon system. The second type are global muons. Rather than starting with the tracker and matching to tracks in the muon chambers, these start with tracks in the muon chambers that are extrapolated back to tracks in the tracker previously unidentified as muons. Additional selections are subsequently applied to suppress fakes. Using these two definitions, PF is able to reconstruct 99% of the muons that pass through the CMS detector.

3.3.3 Electrons and Photons

Electrons often appear as tracks that bend in the magnetic field according to their charge in the tracking system before decaying into energy clusters in the ECAL. Photons, similarly, appear as charged particle energy clusters in the ECAL, but as neutral particles they lack charged particle tracks in the tracker. Both particles involve the reconstruction of energy clusters in the ECAL, which start from cells with higher-energy deposits that are

grouped with neighboring cells into clusters of energy deposits. These clusters are further grouped together to form larger "superclusters" (SCs). PF then evaluates the likelihood of these SCs as coming from electrons vs. photons or charged hadrons by matching them to tracks in the tracker and corresponding clusters in the HCAL. Unlike muons, which have clear tracks and their own dedicated muon detectors, the reconstruction of energy deposits in the ECAL and potential overlap with other particles make electron and photon reconstruction more challenging (and less efficient) than muon reconstruction.

3.3.4 Jets

Jets are hadronic collimated clusters of energy deposits reconstructed from hits in the HCAL. These can be associated with charged or neutral hadrons, which in the charged case can be associated with tracks and possibly showers in the tracker and ECAL respectively. Jets are identified by PF using the anti-k_T algorithm [24], [25], [26] implemented using FastJet [27]. As a cluster algorithm this works, in essence, by defining a distance between hits and using it to iteratively group clusters of hits until the resulting clusters are far enough apart from each other. The optimal HCAL hit clusters are then defined as jet candidates [28].

Jets are defined using a distance parameter R related to the thickness of the jet. Objects referred to as "jets" have a distance parameter of $R = 0.4$, whereas in this thesis "fat jets" refer to objects with a distance parameter of $R = 0.8$. Jet reconstruction efficiencies are limited by the accuracy of the clustering algorithm and jet reconstruction, due in part to the difficulty of disentangling jet overlap.

3.3.5 Missing transverse energy (MET)

MET is an important quantity measured in proton-proton collisions that offers infor-

mation about particles that are not detected, which can include for example neutrinos (which are very light and only interact weakly) and potential BSM particles. Assuming the momenta of reconstructed final state particles are balanced in the transverse (not along the beamline) plane, any missing transverse momentum is indicative of undetected particles. This is heavily reliant on reconstruction efficiencies and therefore the PF algorithm. Mathematically, MET, or p_T^{miss}, is defined as the magnitude of the negative vectorial sum of all reconstructed particle momenta MET$= |-\sum_i^{particles} p_{T,i}|$.

3.3.6 b-tagging

Jets coming from bottom quarks (b-jets) are distinct in that after being produced at a primary vertex, they can travel for a relatively large distance (on the order of centimeters for energies on the order of a hundred GeV) in the detector before decaying at a secondary vertex, as visualized in Figure 3.4. This characteristic is very useful in the identification of b-jets by identification ("b-tagging") algorithms. While there are several b-tagging algorithms used by CMS analyses, this thesis uses "DeepJet" [29] with a medium working point that corresponds to a 1% mistidentification rate of b-jets in QCD multijet simulated events. DeepJet is based on a deep neural network (DNN) that inputs low-level PF inputs (pre-complex reconstruction) from as many jet constituents as possible, unlike other algorithms that use fewer, higher-level jet inputs. It performs better than the next-best algorithm (DeepCSV) both compared to c-flavored and lighter flavored jets as demonstrated in Figure 3.5. This algorithm was chosen due to this gain in performance.

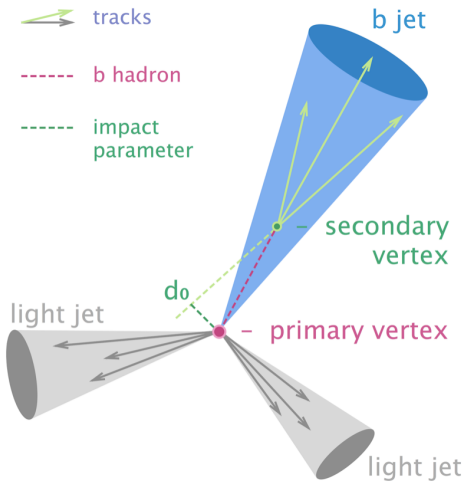

Figure 3.4: Schematic of a reconstructed b-jet, showing the primary and secondary vertices that result from their relatively long lifetime compared to lighter jets. This characteristic is often exploited in b-jet identification algorithms along with other kinematic information. [30]

3.4 Other Important Physical Concepts and Definitions

This section briefly describes some key concepts used in the analysis described by this thesis.

In order to describe the position and trajectories of measured particles and objects CMS uses a coordinate system with its origin at the center of the detector, the z axis along the beamline in the counterclockwise direction, the y axis upwards and the x axis towards the center of the LHC ring. The transverse angle ϕ is defined between the x and y axes, and the polar angle θ is defined between the z and y. However, rather than using θ, the Lorentz-invariant psuedorapidity η is used. This is defined in terms of the the absolute momentum of a particle p and the z component of this momentum p_z:

$$\eta = -ln(tan(\frac{\theta}{2})) = -\frac{1}{2}ln(\frac{p+p_z}{p-p_z}) = arctanh(\frac{p_z}{p}) \tag{3.4}$$

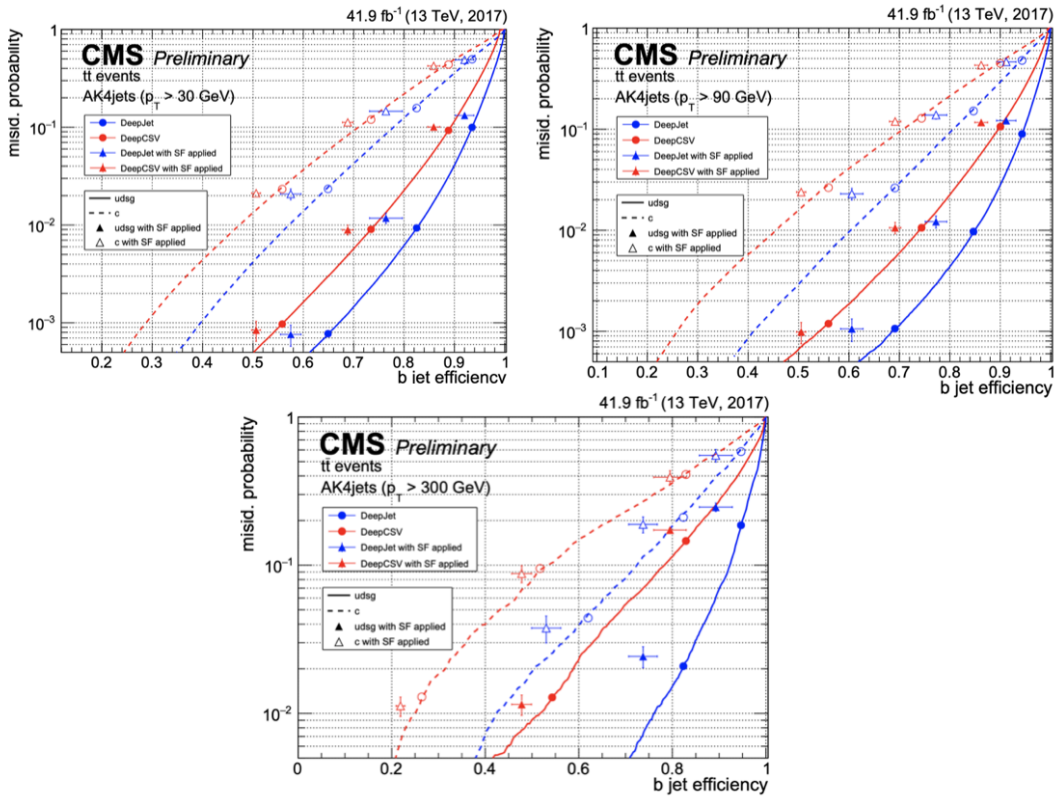

Figure 3.5: Performance of DeepJet vs. DeepCSV in simulated $t\bar{t}$ events in various p_T ranges and shown vs. light jets (udsg) and charm jets (c). For reference, the analysis in this thesis uses a working point of a 1% mistag rate, which corresponds to tagging efficiencies of between 40 and 60 % depending on the flavor and p_T range. [29]

This coordinate system is shown in Figure 3.6. The transverse plane in this system is the x-y plane that is not in the direction of the beamline. The quantity p_T is the transverse component of the absolute particle momentum p, and H_T is the scalar sum of all transverse momenta in an event. If particle momenta are evenly distributed in all directions in the x-y plane, H_T can be expected to be low compared to if there was more momentum in a particular direction. The latter tends to happen when there is a heavier particle (such as a top quark) produced in an event, and so H_T can serve as a good discriminating variable when looking for heavy hadronic signatures.

As revealed in future chapters, the modeling of proton-proton collisions is complex

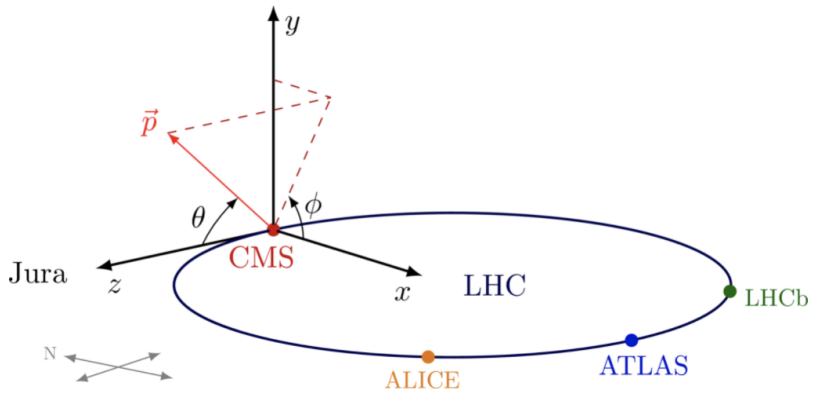

Figure 3.6: The coordinate system used by the CMS detector. [31]

and requires many effects and uncertainties to be taken into account. One such effect is the "underlying event" (UE). In a proton-proton collision, "leftover" partons that do not engage in the hard scattering process also hadronize and decay into the detector. The resultant extra hits and tracks are the UE. Another important consideration is pileup (PU). It is possible (and indeed likely, at high luminosities corresponding to high rates of rare events) for multiple hard scattering events to occur in one bunch crossing, leading to multiple events with large momentum transfer that overlap and interfere with each other. Both the UE and PU must be taken into account when simulating collision events in CMS and when considering systematic and experimental uncertainties.

Chapter 4

Statistical Methods

This section describes important concepts used to describe statistical results in this analysis. To summarize, a binned maximum likelihood fit is performed simultaneously in signal region (SR) categories and using systematic uncertainties as nuisance parameters in order to extract a signal strength based on a profile likelihood method. This closely follows standard CMS statistical recommendations and uses standard statistical software packages [32], [33], [34], [35].

4.1 Statistical Overview

In particle physics, the properties of signal processes (like four-top production) are based on probabilistic models of how theoretical calculations manifest under experimental conditions. Statistical analysis, in a general sense, allows for the evaluation and validation of these probabilistic models given experimental data. There are two schools of thought when it comes to statistical inference: frequentist statistics and Bayesian statistics [36]. Both frequentist and Bayesian interpretations motivate useful tools to quantify the validity of a signal model and to describe statistical results in a particle

physics analysis. Frequentist statistics interprets probability as the frequency of a given outcome in a repeated experiment. Bayesian statistics adds subjectivity to the concept of probability, in that Bayesian methods describe the probabilities of not only the measured outcomes, but also of the original hypothesis. This is demonstrated by Bayes' theorem, which states that the conditional probability of observing some model m given data D is equal to the conditional probability of observing the data D given the model m multiplied by the probability of m divided by the probability of D:

$$P(m|D) = \frac{P(D|m)P(m)}{P(D)} \tag{4.1}$$

In particle physics, this is perhaps more usefully written as:

$$P(theory|data) \propto P(data|theory)P(theory) \tag{4.2}$$

Where "theory" is the theoretical model of a given signal and "data" is the observed data collected by the experiment. Here $P(theory)$ reflects the subjective probability of the experimental hypothesis prior to the experiment and $P(data|theory)$ is the "likelihood", or the probability to have measured that data given that particular theory. The likelihood is an important metric in quantifying the reliability of the theoretical model and assumptions made in the experimental setup. This version of Bayes theorem is reported as a proportionality relation because it must be summed over all possible hypotheses.

In particle physics analyses, the theoretical model is made up of a set of parameters, including the parameter of interest that describes the result (usually the signal strength), and other "nuisance" parameters, which typically describe the "error" of statistical or systematic uncertainties in the analysis. The signal strength, often represented by μ, refers to how much signal is observed compared to what was predicted by theoretical

models and simulation (μ_{obs}/μ_{SM}). The set of nuisance parameters is represented by θ.

This section has introduced three concepts core to statistical analysis: the likelihood $P(data|theory)$, the signal strength μ, and nuisance parameters θ. The next section describes how a likelihood is constructed in practice, in terms of a toy "counting experiment".

4.2 Constructing the Likelihood

A counting experiment can help to elucidate how to quantify and mathematically define likelihoods which, in the context of particle physics, can be thought of as the probability to get observed data given a theoretical model and experimental setup. The likelihood depends on the parameter of interest μ and nuisance parameters θ. In a counting experiment, the probability of counting some number of events x can be written following a Poisson distribution with an expected total event yield λ:

$$Poisson(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \tag{4.3}$$

In practice, distributions of observed variables in an experiment are binned in histograms, for example in individual histogram and analysis category bins i over a total of N bins. The probability of measuring x events given a statistical model λ that follow this Poisson distribution (in other words the likelihood of x given λ) is:

$$p(x|\lambda) = \prod_{i}^{N} Poisson(x_i|\lambda_i) \tag{4.4}$$

This model λ describes all assumptions and uncertainties in the experiment, including experimental detector uncertainties and efficiencies, uncertainties related to the theoretical modeling of signal and background events in simulation, and other systematic

uncertainties. In practice, it is useful to express the number of expected events λ_i per histogram and category bin i as the signal strength μ times the number of signal events in that bin given nuisances $s_i(\theta)$ plus the number of background events in that bin given nuisances:

$$\lambda_i = \mu \cdot s_i(\theta) + b_i(\theta) \tag{4.5}$$

The likelihood itself can be written in terms of a given measurement ("*data*") as a function of the signal strength μ and nuisance parameters θ over N histogram bins :

$$L(data|\mu,\theta) = \prod_{bin=i}^{N} Poisson(data_i|\lambda_i) = \prod_{bin=i}^{N} \frac{(\mu \cdot s_i(\theta) + b_i(\theta))^{data_i}}{data_i!} e^{-(\mu \cdot s_i(\theta) + b_i(\theta))} \tag{4.6}$$

Here $data_i$ is the number of data events observed in bin i. An important note is that nuisance parameters θ may depend on a given histogram bin i, not just on a given process (signal or background). For example, for background, the number of events is equal to the sum over background processes of the nominal expected number of background events for that process in that bin, times the set of probabilities for each bin and process and nuisance parameter (n) for the background:

$$b_i(\theta) = \sum_{p}^{processes} b_{p,i} \cdot \prod_{n}^{dim(\theta)} P_{n,p,i}(\theta_n) \tag{4.7}$$

The probability distributions for all nuisances $P_{n,p,i}(\theta_n)$ can be dependent or constant per nuisance n, process p and bin i. This leads to different types of nuisance parameter probability distributions and different corresponding correlations between them. In this analysis, nuisances encountered are "rate-based", "shape-based", or "statistical". If a nuisance affects the normalization of one or more processes in some or all bins (thus

impacting the number of events or scale of some observable variable distribution rather than its shape) this nuisance is described as a "rate" or "normalization" uncertainty, and varying that nuisance parameter can change the number of events but not the shape of distributions. If, on the other hand, observable distributions are affected, this nuisance is a "shape" uncertainty. Nuisances can also impact both the shape and normalization of a given distribution. Lastly, in order to describe the statistics of simulated or predicted events that infer the yield of a given process in a given bin, an additional nuisance parameter per bin is assigned to account for statistical uncertainties [34], [35]. All of these nuisances can be correlated or uncorrelated between processes or bins depending on experimental and theoretical assumptions, and so must also be treated accordingly.

In particle physics, when searching for the presence or absence of a physical process (signal) it is customary to discuss two likelihood models in an experiment: the background-only (or "null") hypothesis such that $\mu = 0$, and the signal hypothesis such that $\mu > 0$. Generally speaking, a signal process is "discovered" when measured data is found to be incompatible with the background only hypothesis. The next section will define additional concepts that help to describe and quantify such results.

4.3 Profile likelihood ratios and other important statistical definitions

A profile likelihood ratio is a powerful quantity when it comes to comparing two competing likelihoods. It can be written in terms of the parameter of interest (in this case the signal strength μ) as

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} \tag{4.8}$$

Here the "hat" symbol denotes particular variables that maximize the likelihood L. In other words, this equation reads that the profile likelihood is equal to the likelihood at the value of $\theta = \hat{\theta}(\mu)$ for which the likelihood is maximized for some μ, divided by the likelihood at the value of μ and θ for which it is maximized. This constructed so as to be independent of the nuisance parameters θ, allowing the likelihood to be a function of the signal strength alone.

Maximizing the likelihood can often be more usefully written as minimizing the negative log likelihood $-ln(L)$. This and the profile likelihood definition form the basis of a central definition in statistics [37], [38] which defines a test statistic $\tilde{q}(\mu)$ that asymptotically converges towards a non-central χ^2 distribution with one degree of freedom:

$$\tilde{q}(\mu) = \begin{cases} -2ln\tilde{\lambda}(\mu), & \hat{\mu} \leq \mu \\ 0, & \hat{\mu} > \mu \end{cases}$$

In other words, when the signal strength μ is less than or equal to the value of μ that maximizes the profile likelihood $\hat{\mu}$, this test statistic is equal to a nonzero quantity $-2ln\tilde{\lambda}(\mu)$. Otherwise, this test statistic is zero. $\tilde{q}(\mu)$ is therefore a quantity that differentiates between the signal vs. background-only hypotheses in a way independent of nuisances θ given a sufficiently large sample size. As visualized in Figure 4.1, any one measurement leads to one observed value of $\tilde{q}(\mu) = \tilde{q}_{obs}$. A set of possible measurements, for example sampled as toys from a statistical model or from an Asimov dataset (an approximate dataset with free parameters set to expected values), allows for the determination of the probability $f(\tilde{q}(\mu)|\mu)$ of getting an observed test statistic $\tilde{q}_{obs}$ given μ . These quantities ultimately allow for the useful frequentist definition of upper limits CL_S.

Upper limits express the compatibility of a statistical model with the signal strength μ given an observed test statistic $\tilde{q}(\mu)$. The upper limit of the full S+B model CL_{S+B}, for example, is $CL_{S+B} = \int f(\tilde{q}(\mu)|\mu) \cdot d\tilde{q}(\mu)$. If one assumes the background-only case where there is no signal (and so $\mu = 0$) this upper limit becomes $CL_B = \int f(\tilde{q}(\mu)|0) \cdot d\tilde{q}(\mu)$. Most interesting is the measurement of the upper limit of the signal strength alone. This can be expressed as $CL_S = \frac{CL_{S+B}}{CL_B}$ and is shown schematically in Figure 4.2. Upper limits allow signal strengths to be quantified based on observation, and therefore are commonly used as a measure of sensitivity. For example, the observed upper frequentist limit on the value of μ above which the signal strength is excluded by the measurement $\tilde{q}_{obs}$ with 95% confidence is calculated by setting $CL_S = 0.05$. The "expected" limit can also be calculated by sampling $\tilde{q}_{exp}$ from a statistical model assuming no signal $\mu = 0$. If $CL_{S(exp)} > 1$ in this case would mean that the background-only model would be expected to explain the measurement as a gauge of the sensitivity of a blinded analysis.

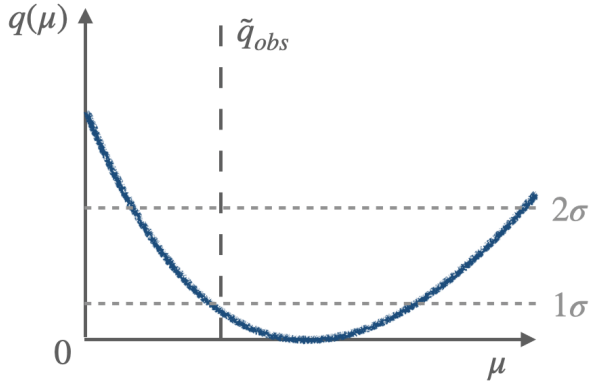

Figure 4.1: Visualization of a measurement of an example test statistic $\tilde{q}_{obs}$ as a probing of the distribution $q(\mu)$ at some value $\tilde{\mu}$. Note that $q(\mu)$ approaches zero at a certain $\mu = \hat{\mu}$ that mamimizes the profile log likelihood, corresponding to the minimum on the curve $q(\mu)$. Example lines are also shown for one and two standard deviations from this maximum profile log likelihood.

The "discovery" of a physical process can be defined using similar quantities in order to construct the significance of an observation. Qualitatively, this is the statistical fluctuation in a background-only probability distribution $f(\tilde{q}(\mu)|0)$ required to explain some

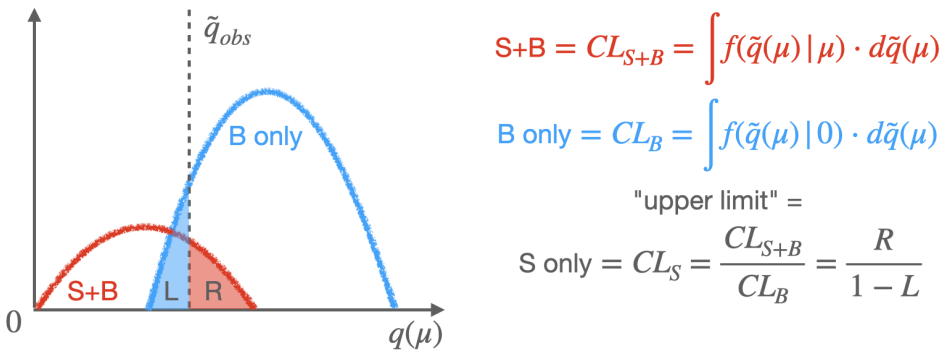

Figure 4.2: Schematic visualizing the definition of CL_{S+B} and CL_B as the upper limits of the signal+background and background-only hypotheses as a function of the signal strength μ, defined as integrals over probability distributions $f(\tilde{q}(\mu)|\mu)$. An example line for $\tilde{q}_{obs}$ is shown, as well as the integrals L and R that make up the signal-only frequentist upper limit CL_S.

observation $\tilde{q}(0)$. The significance is thus often expressed in terms of Gaussian standard deviations σ. Mathematically, this means that $\tilde{q}(0)$ is defined for the background-only case as:

$$\tilde{q}(0) = \begin{cases} -2ln\tilde{\lambda}(\mu), & \hat{\mu} > 0 \\ 0, & \hat{\mu} \leq 0 \end{cases}$$

and the probability of observing $\tilde{q}(0)$ given $\mu = 0$ is then

$$p(\tilde{q}(0)|0) = \int f(\tilde{q}(0)|0) \cdot d\tilde{q}(0) \tag{4.11}$$

In terms of Gaussian standard distributions this can be written as the following and solved for the significance s:

$$p(\tilde{q}(0)|0) = \int_s \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot dx \tag{4.12}$$

In particle physics a significance $s > 3\sigma$ is often quoted as the threshold for "evidence" of a given physical process, and a significance $s > 5\sigma$ is often quoted as the threshold for a "discovery".

One final important statistical definition is the concept of "pull". Until now the determination of the parameter of interest (the signal strength μ) was discussed in the context of finding the "best fit" value $\hat{\mu}$ that maximizes the profile likelihood (or, equivalently, minimizes the negative profile log likelihood). Physics analyses also often seek to quantify the nuisance parameters that maximize the profile log likelihood $\hat{\sigma}(\hat{\mu})$. This is done by temporarily considering each nuisance parameter the parameter of interest and finding its best fit value (with other nuisances held constant such that the fit is independent of them). This can be done before ("pre-fit") and after ("post-fit") the observation $\tilde{q}_{obs}$ that finds the observed signal strength μ. Pulls refer to the change of values and probability distribution widths $\Delta\theta$ of nuisance parameters before and after the fit in order to gauge whether underlying assumptions about individual nuisances in the statistical model make sense.

$$pull(\theta) = \frac{\theta_{post} - \theta_{pre}}{\Delta\theta_{pre}} \tag{4.13}$$

Chapter 5

Analysis Technologies

This chapter gives brief overviews of some important technological and computational tools used in this analysis. This includes machine learning algorithms used to discriminate between signal and background processes, simulation technologies used to simulate particle physics processes in the CMS detector, data formats and structures for data sample management, and statistical packages for performing profile maximum likelihood fits in order to obtain sensitivity results.

5.1 Introduction to Multivariate Analysis and Machine Learning

Generally, multivariate analysis (MVA) involves a function that given a set of input features (or multiple variables), makes a prediction about that input based on example data. Machine Learning (ML) refers to computer models that construct the mathematical functions that perform these mappings from inputs to predictions. These models are often extremely complex, involving many mathematical operations and parameters that can be optimized to organize complex datasets based on patterns often unrecognizable

by the human eye. ML algorithms can be supervised (trained on labeled training data), unsupervised (training data is not labeled), or semi-supervised (some training data is labeled). In supervised algorithms, optimization involves "training" the model by minimizing an objective loss function that compares predicted and "true" expected outputs. This involves a "training" set of data that contains both inputs and true categorization information. Supervised learning can be classification or regression, depending on whether an algorithm is predicting discrete quantities, or a function of continuous values, respectively. In particle physics ML is commonly used for classification tasks, where events are classified as signal or background based on their characteristics (for example kinematic information). In these cases, the training set often takes the form of a simulated dataset containing generator-level "truth" information regarding whether events are actually coming from signal or background. The performance of ML algorithms is often checked using another set of "testing" data that the algorithm has not encountered before, in order to evaluate performance and the need for further optimization. This section will briefly discuss ML algorithms relevant to this analysis, including boosted decision trees and neural networks. Sources [39], [40], were used as a reference for this information, and contain more details and examples of other types of algorithms.

5.1.1 Boosted Decision Trees

A decision tree is a supervised ML model that works by iteratively partitioning data in a flowchart-like (or branch-like) fashion (see figure 5.1). Decision trees can be used for both classification and regression. A problem with decision trees is that they tend to overfit data: they become very good at correctly labeling the training set but are not able to generalize well to the test set. Boosted decision trees (or BDTs) help remedy this, by sequentially training several decision trees that learn from the mistakes of the previous

trees via a loss function. Models can be added sequentially until no further improvements are made. A further innovation is gradient boosting, where each decision tree takes into account and predicts the errors of previous trees, and the combination of the new prediction and the prediction from the previous tree is what is propagated forward to the final prediction. This technique uses gradient descent as the loss function. XGBoost [41] is a common, powerful, and easily implementable gradient boosting agorithm that is used in this analysis for a top-tagging algorithm, as described in Chapter 7.

Another BDT algorithm used in this analysis (in the event-level BDT described in Chapter 9) is the CatBoost algorithm [42]. While also a gradient boosting algorithm, this has the innovation of allowing what is called "one-hot-encoding" in classification problems. In simple terms, this allows for the algorithm to be egalitarian when treating different categories rather than ranking them. In practice this means labeling categories in a binary fashion as "001", "010", and "100" rather than as "1", "2", and "3". When doing this, the CatBoost algorithm also permutes the training set in random order, so that each time it assembles a tree the labels are not necessarily the same. The effect of this is often a reduction of overfitting. While older and simpler than many other ML algorithms, decision trees are still relevant as very powerful algorithms that can often achieve very similar performance with less computation time than more complex techniques like neural networks. Of course, problems and datasets tend to lend themselves to specific algorithms and optimization strategies, so an important part of building an algorithm is the choice of one that is best suited to the problem at hand.

5.1.2 Neural Networks and Normalizing Flows

Like decision trees, neural networks (NNs, or deep neural networks, "DNNs") are machine learning algorithms that are good at modeling data that has nonlinear relation-

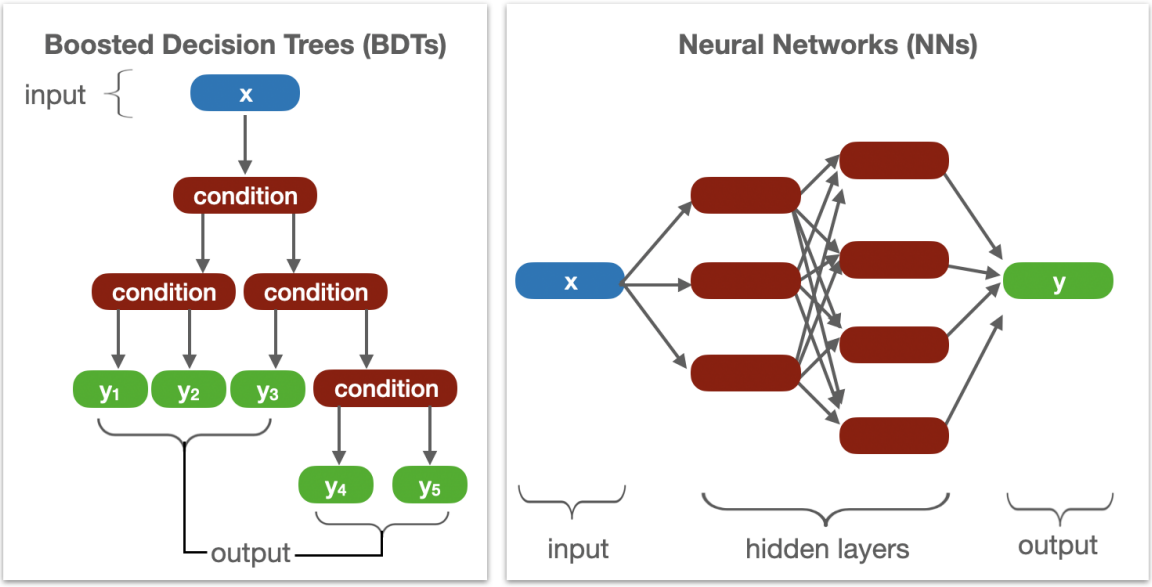

Figure 5.1: Schematic comparing the structure of a BDT (left) vs. a NN (right) given inputs x and outputs y. Note that these are just example structures and that x and y can be vectors of multiple inputs and outputs.

ships between features. Rather than categorizing data like branches on a tree, neural networks are structured like a circuit of neurons, with multiple layers of "nodes" linked by complex connections. As such, they can be thought of as similar, but more complex algorithms. They can be supervised or unsupervised, and supervised models can be used for classification or regression.

On a fundamental level, a neuron processes one or several inputs in order to produce a single output. Inputs are weighted, combined, and passed through an activation function (in order to ensure output is in a predictible, bounded form). Mathematically, for example, a neuron could take two inputs x_1 and x_2, apply a bias b and weights w_1 and w_2, and apply an activation function F in order to get output y. If the inputs and weights are written as vectors x and w, this is:

$$y = F(w_1 \cdot x_1 + w_2 \cdot x_2 + b), \tag{5.1}$$

$$y = F(w \cdot x + b). \tag{5.2}$$

When neurons are connected together into layers, such that the outputs of neurons in one layer become the inputs of neurons in another, the result is a NN (see Figure 5.1). Any layer of neurons between the input and output layers is a "hidden" layer, and the depth of a neural network refers to the number of hidden layers (and thus the complexity of the network). Like BDTs, neural networks also optimize their performance using loss functions. During training, a NN seeks to minimize its loss function evaluated by comparing its outputs to truth-level output information from training data. There are several types of loss function that can be chosen based on the data and problem encountered. Negative log likelihood as described in Chapter 4, for example, can be used as a loss function. A loss function that is used in this analysis is the Maximum Mean Discrepancy (MMD) loss [43], [44], which computes the differential between the feature means of a predicted distribution (generated by a network) and an observed distribution (of data within a mini-batch, or small subset of the training data). In practice this loss function is typically applied using the formula:

$$L(x, y) = \sum_{i=1}^{N}\sum_{j=1}^{N} k(y_i, y_j) - 2\sum_{i=1}^{N}\sum_{j=1}^{N} k(y_i, x_i) + \sum_{i=1}^{N}\sum_{j=1}^{N} k(x_i, x_j) \tag{5.3}$$

Where x and y are the true and predicted outputs, N is the number of outputs in the x and y vectors, and k is the gaussian distribution with standard deviation σ given by $k(a, b) = e^{\frac{-(a-b)^2}{\sigma}}$.

The loss is minimized by calculating its partial derivative (in terms of the weights w and the bias b for example) and using that to update the variables. For example, given a learning rate η and stochastic gradient descent as an optimization function, after one execution of a neural network weights w would be updated according to the loss function

L such that

$$w_{new} = w_{old} - \eta \frac{\partial L}{\partial w}. \tag{5.4}$$

The optimization function in used in this analysis is the common "Adam" optimizer [45] which is based on higher-order stochastic gradient descent. After updating the weights and bias, the NN has completed one epoch. It will continue training to minimize the loss for a set number of epochs, or until performance has reached some desired threshold.

A type of neural network relevant to this thesis is neural autoregressive flow (NAF) [46]. This is an autoregressive NN that inputs information not just from the current epoch, but also training information from previous epochs as additional inputs to the model. A NAF learns an invertible, bijective function that expresses a transformation between variables, namely sets of input and output variables x and y. It is therefore able to predict distributions based on a complex set of input variables, which makes it handy for predicting data-driven distributions as described in Chapter 10.

5.2 Monte Carlo simulation

Monte Carlo simulation, in simple terms, is a computerized way of modeling possible outcomes given the probability of each of those outcomes occuring. It calculates possible results by sampling from input probability distributions in order to produce distributions of possible outcome variables. In the context of particle physics, Monte Carlo (MC) generators simulate the results of particle collisions given theoretical calculations and detector conditions from the initial collision event to the readout of "data" in various subdetectors [12], [47]. These generated events depend on theoretical probability distri-

butions (for example related to parton distribution functions (PDFs), factorization and renormalization scales, and the order of perturbative calculations), as well as detector conditions (including for example detection and reconstruction efficiencies and the effects of pileup). Therefore, simulated events are subject to various corrections to ensure they behave in the same way as real data and systematic uncertainties related to the setup of the MC generators used. These corrections and uncertainties are described in later chapters.

5.3 Data format and organization

The size and storage of real and simulated CMS datasets is an important consideration in CMS analyses, as processing time and storage requirements are often what slows down or prevents an analysis from moving forward. In order to reduce file size and processing time as well as to promote ease of use, a compact data format called "NanoAOD" [48] stores event information in numpy array format [49] at the expense of some information loss, especially computationally expensive generator-level MC simulation information. This analysis was specifically designed to be compatible with NanoAOD datasets.

5.4 Statistical packages

In order to apply the statistical methods discussed in Chapter 4 and to stay true to standard CMS recommendations and procedures in the calculation of results, the "combine" tool was used [32], [33], [34], [35]. This is a software package developed by the CMS experiment to perform statistical computations, and is very commonly used in CMS analyses.

Chapter 6

The Physics of Four-Top Production

So far, a general overview of the physics, methods, and technologies of CMS particle physics analyses have been presented, including how different particle signatures can be detected and signal strength measurements can be extracted from collision data. This section now focuses on the particular process that is the main topic of this thesis, SM $t\bar{t}t\bar{t}$ production in the all-hadronic final state. An overview of $t\bar{t}t\bar{t}$ final states and backgrounds are discussed, as well as previous $t\bar{t}t\bar{t}$ results.

6.1 Standard Model Four-Top Production

Four-top production is a rare SM process that can occur in proton-proton collisions through gluon fusion or (rarely) quark-antiquark annihilation. This can be mediated via gluons (QCD), photons (electromagnetism), Z bosons (weakly), or a Higgs boson, as shown in Figure 1.1. The predicted cross section for SM $t\bar{t}t\bar{t}$ production is $12.0^{+2.2}_{-2.5}$ fb at next-to-leading order (NLO) [1], which thousands of times smaller compared to for example the $t\bar{t}$ SM cross section of about 832^{+55}_{-64} pb [2]. Each of the four tops in four-top production decay into a W boson and a bottom quark. The bottom quark hadronizes

and may manifest as a b-jet in the CMS detector, and the W boson further decays either hadronically, into a quark-antiquark pair that can hadronize into jets in the detector, or leptonically, into a lepton and its partner neutrino. The neutrino passes through the detector and can only be reconstructed as MET. If the lepton is a muon or electron, it can produce tracks or showers respectively, but taus pose a challenge as they can also decay hadronically and are difficult to reconstruct. For this reason, tau final states are often unpopular final states in four-top analyses, and "leptons" in the context of four-top analyses refer only to muons or electrons.

Depending on how all four tops decay, there are various possible final states (or "channels") in four-top production, as outlined in Figure 6.1. Four-top analyses seek to find evidence of a particular final state signature relative to backgrounds. Each of the channels has its own likelihood of occurring and thus anticipates a given number of events, and they each have backgrounds that have signatures like that of signal and thus must be measured and discriminated against. In order for a given $t\bar{t}t\bar{t}$ signal to be detected it must have enough events that are statistically significant with respect to backgrounds. A major background in all channels, due to its close resemblance to signal, is $t\bar{t}$, or top-antitop pair production. While its signature is different from signal in that it contains half the number of expected top quarks, discrimination is complicated by reconstruction efficiencies, the potential radiation of extra jets or overlap with pileup. The most sensitive four-top channels are those that most reduce this background, namely signatures where two or more of the four tops, of which two are (anti)tops, decay leptonically into at least two leptons of the same sign (so two (anti)electrons, two (anti)muons, or an (anti)muon-(anti)electron pair). Because leptonic $t\bar{t}$ would always decay into two opposite-sign leptons (a lepton and an antilepton), this eliminates $t\bar{t}$ as a background except in the case of lepton sign misreconstruction, making such "same-sign dilepton and multilepton" final states the most sensitive despite not having the largest branching ratios. To date,

this is also the only final state published for full Run II data in CMS, as shown in Figure 6.2. This analysis observed a signal significance of 2.6σ compared to an expected significance of 2.7σ in this channel [50]. This comes close to the 3σ threshold often quoted as "evidence" of the existence of a process. The most recent Run II ATLAS results included several final states and reported an observed significance of 4.7 sigma, close to the 5 sigma discovery threshold, compared to an expected significance of 2.6 sigma [51]. The analysis central to this thesis is part of an effort by CMS to study $t\bar{t}t\bar{t}$ production in Run II data for multiple final states, in order to better compare against ATLAS's result and to be the first to study $t\bar{t}t\bar{t}$ production with an expected significance of more than 3σ. This effort will add three additional final states to the same-sign-dilepton result [50] in order to combine four analyses in total.

Two of these analyses are an improvement and extention of a 2016 result involving the opposite-sign-dilepton and single-lepton final states [52]. These channels have to contend with $t\bar{t}$ as a major background without the advantage of using lepton signs as a discriminatory tool. The third addition, the all-hadronic channel, is the subject of this thesis. This final state is absent of leptons and thus its $t\bar{t}t\bar{t}$ signature involves only jets. It therefore not only has to contend with hadronic $t\bar{t}$ as a major background, but also with QCD multi-jet backgrounds resembling signal. Lepton-containing signatures greatly reduce these QCD backgrounds, which have enormous cross sections relative to signal ($1{\cdot}10^{20}$ times that of signal) and are difficult to simulate: the large cross section means the rates of rare QCD events are on the same scale as the $t\bar{t}t\bar{t}$ signal. The computational power required to produce enough of these rare events to effectively simulate QCD is so immense as to be impossible, and so these backgrounds need to be estimated using alternative data-driven means. This makes all-hadronic $t\bar{t}t\bar{t}$ a very challenging final state when it comes to achieving signal sensitivity. The remaining sections of this thesis describe an analysis that seeks to do just that as the first all-hadronic $t\bar{t}t\bar{t}$ final state

measurement. The results will be a part of the aforementioned upcoming CMS Run II multichannel paper, which is likely to be published in 2022.

The all-hadronic $t\bar{t}t\bar{t}$ analysis described here is significant in that it is the first analysis of its kind that achieves sensitivity using novel analysis techniques that are potentially applicable to other hadronic analyses. It is also important as part of a combined effort to find and measure 3σ evidence of $t\bar{t}t\bar{t}$ production, and to compare with the unexpectedly large signal significance seen in the similar ATLAS result [51]. As mentioned in Section 2, a measured excess of $t\bar{t}t\bar{t}$ signal could be evidence of BSM physics, and final states involving tops are well-poised to measuring the top-Higgs Yukawa coupling. The measurement of the cross section of SM four-top production is therefore an important test of the SM, both in terms of measuring SM parameters, and probing for hints of physics beyond it.

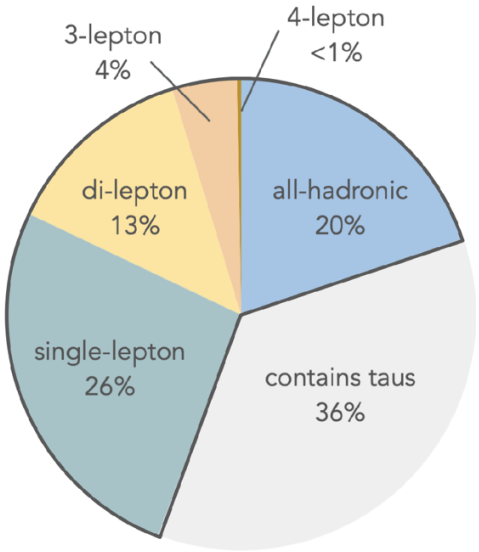

Figure 6.1: Branching ratios of various SM $t\bar{t}t\bar{t}$ final states. Possible final states and their likelihood of occuring are shown as percentages.

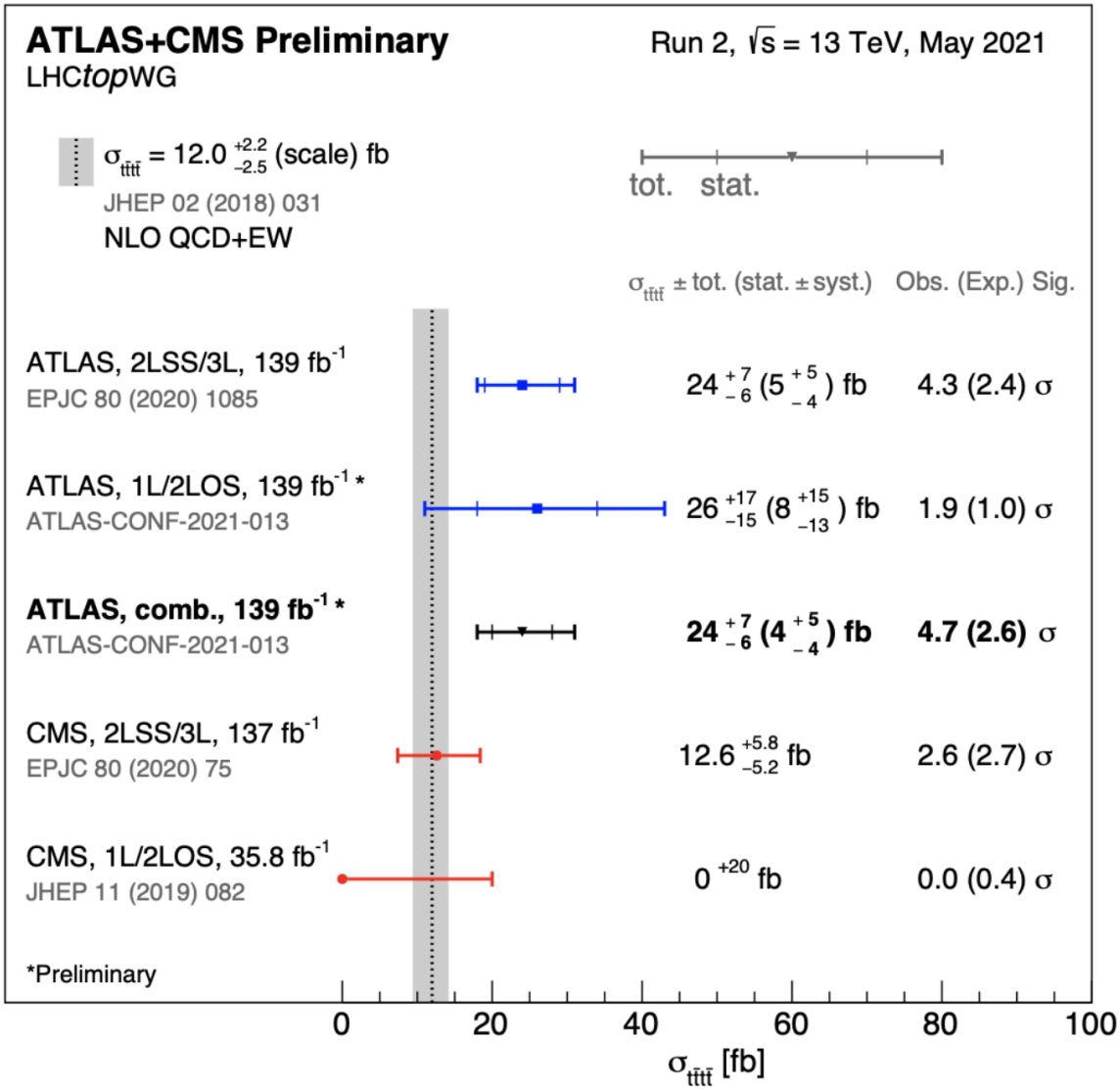

Figure 6.2: Summary of previous $t\bar{t}t\bar{t}$ results in ATLAS and CMS.

Chapter 7

Object Selection

This chapter defines objects used in the all-hadronic $t\bar{t}t\bar{t}$ analysis that is the topic of this thesis. Events are defined so as to contain no leptons and according to quality preselection requirements for jets and vertices. In addition to PF reconstructed objects, additional objects are defined, including b-tagged jets and hadronically decaying top quarks. These definitions are discussed in the following sections, with particular attention paid to top tagging as a new algorithm that was developed for this analysis.

7.1 Vertex selection

As is typically recommended in CMS analyses, the following standard selection criteria are applied to interaction vertices reconstructed in an event:

- The vertices must come from fits to trajectories of reconstructed particle tracks with positive χ^2 values.
- There are at least 5 degrees of freedom in the vertex fit.

- The distance, in absolute z, along the beam line from the nominal center of the detector is less than 24 cm.
- The transverse displacement, ρ, from the beam line is less than 2 cm.

Selected events are required to have at least one vertex fulfilling these requirements. The vertex which satisfies these criteria and has the highest $\sum p_{\mathrm{T}}^2$ of tracks associated to it is taken to be the primary vertex (PV) from which the physics objects used in this analysis originate.

7.2 Leptons

As this is an all-hadronic final state, leptons (muons and electrons) must be identified in order to remove ("veto") events containing leptons from consideration. Standard CMS criteria are used to identify electron and muon candidates. Events with electron or muon candidates identified with these criteria are vetoed from the analysis. The EGamma POG "Fall17-noIso-v2" MVA with a "loose" working point [53] is used to select electron candidates with $p_{\mathrm{T}} > 15\,\mathrm{GeV}, |\eta| < 2.5$ (excluding those in the barrel-endcap transition region with $1.442 < |\eta| < 1.566$). The "loose" muon definition recommended by the Muon POG [54] is used to select muon candidates with $p_{\mathrm{T}} > 15\,\mathrm{GeV}, |\eta| < 2.5$. Both electron and muon candidates are required to be isolated from hadronic activity by requiring their relative "mini-isolation" quantity to be less than 0.4. The muon and electron selections were chosen to be orthogonal to the single-lepton $\mathrm{t\bar{t}t\bar{t}}$ channel, in order to ease combination with this final state that is also mostly hadronic and so has the highest potential for overlap.

7.3 Jets

The jets selected for use in this analysis are PF jets clustered with the anti-k_{T} algorithm [55] with a distance parameter of 0.4, and using standard jet energy correction and pileup correction procedures [56]. These corrections are applied in order to translate jet energy measured in the detector to the true particle or parton energy. For quality control, selected jets are also required to have $p_{\mathrm{T}} > 35$ GeV and to be within the tracker volume ($|\eta| < 2.4$). Jets must also satisfy loose PF Jet identification criteria for data collected in 2016 and tight PF Jet identification criteria for data collected in 2017 and 2018 as recommended by the JetMET POG [57],

7.3.1 b-tagging

As noted in Chapter 3, b-tagged jets are identified using the DeepJet algorithm [58]. The medium working point is used, corresponding to thresholds of 0.3093, 0.3033, and 0.2770 for 2016, 2017, and 2018, respectively.

7.3.2 Top- and W-tagging

The reconstruction of hadronically decaying top quarks involves reconstructing decay products that result from b-quarks and other lighter quarks coming from W bosons from jets in the hadronic calorimeter. As shown in Figure 7.1, at lower energies (below about $400GeV$) jets resulting from top decays are seperately reconstructable, or "resolved", but at higher energies these decay products become columnated into larger boosted "fat-jet" objects. As indicated in Figure 7.2, most tops are expected to be in the resolved regime in the $\mathrm{t\bar{t}t\bar{t}}$ all-hadronic channel, however some sensitivity can be gained by also reconstructing boosted tops as many leading and subleading tops are expected to be boosted. For this reason, this analysis employs both boosted and resolved top tagging

algorithms.

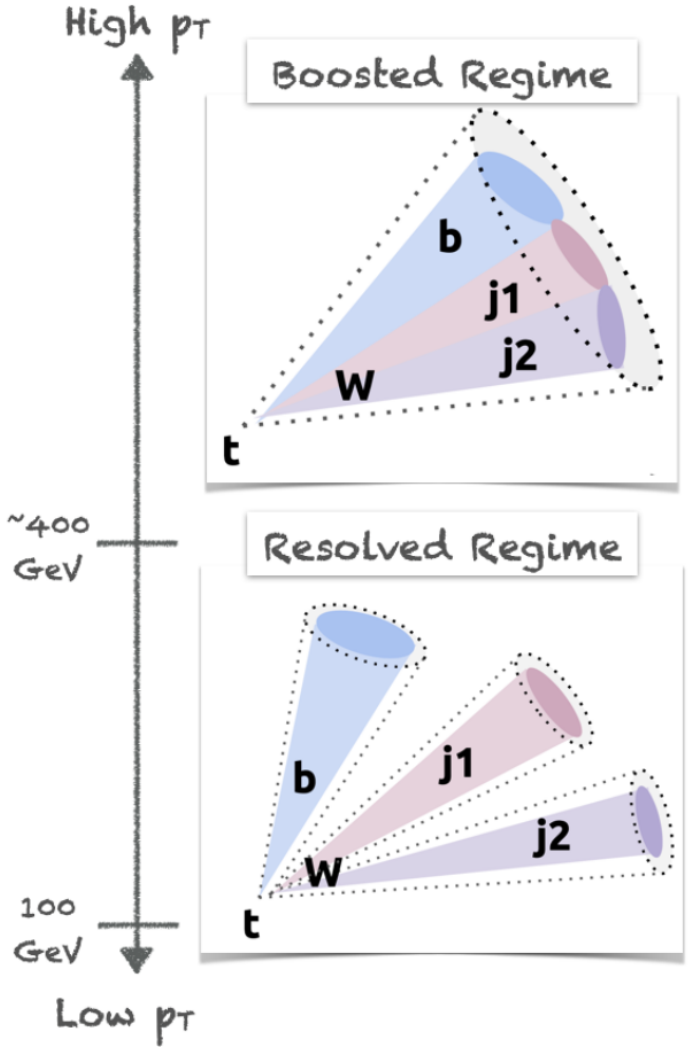

Figure 7.1: Illustration of resolved vs. boosted hadronic top decay.

The decay products of boosted top quarks with $p_{\mathrm{T}} > 400$ GeV, or of W bosons with $p_{\mathrm{T}} > 200$ GeV, are expected to be contained within a ΔR radius of 0.8. The DeepAK8 algorithm is applied to jets clustered with the anti-k_{T} algorithm using a distance parameter (R) of 0.8, i.e. twice the value used for the standard CMS jet collection, to identify boosted hadronically decaying top quark and W boson candidates, within $|\eta| < 2.4$. This algorithm was chosen because it is broadly and easily usable and has excellent perfor-

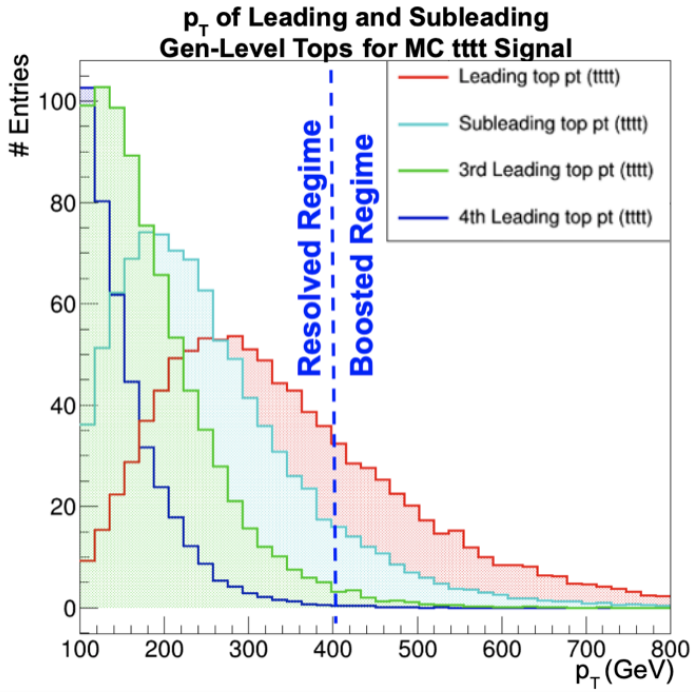

Figure 7.2: p_T distribution of gen-level simulated all-hadronic four-top events. The p_T of each of the four tops in the event is shown.

mance. This algorithm uses boosted top and W candiates passing a medium working point corresponding to a 1% mistag rate in simulation. Boosted top quark and W boson candidates are required to be separated by ΔR of at least 0.8 from the resolved top quark candidates identified as described below to prevent overlap.

A dedicated BDT was developed for this search to identify moderately boosted top quarks, whose decay products can be resolved into separate anti-k_{T} $R = 0.4$ jets, and is based on a resolved top quark tagger that was originally developed for a search for supersymmetric partners to the top quark in the all-hadronic final state [59]. This algorithm was developed to be compatible with centrally-produced reduced-data-format NanoAOD samples (as opposed to larger MiniAOD samples) in order to significantly reduce time and space needed for computing. It is the first resolved top tagger compatible with these centrally-produced samples. It was also updated compared to the original algorithm to

use DeepJet b-tag information. Three-jet resolved top candidates are formed starting from one of up to four jets in the selected $R = 0.4$ jet collection with the highest DeepJet b-tag values, which is designated the "b" jet candidate. Two additional jets are identified from all unique two-jet combinations in the collection, excluding the already identified b subjet, and are designated the "W" subjets. To reduce the combinatorial background before making any tight selections, the choice of the two W subjets is subject to the loose conditions that the mass of the system formed by the W subjets is within 40 GeV of the true W boson mass and the mass of the combined three-jet system is within 80 GeV of the true top quark mass. These three-jet combinations are candidates for the reconstruction of resolved top quarks. A lower energy threshold of $100GeV$ is also applied because of reduced tagging efficiency in that region. The following variables form the inputs to the BDT:

- The mass of the jet designated as the b constituent jet
- The pairwise invariant masses of the b jet with each of the W subjets.
- The mass of the top candidate (the top candidate four-vector is the sum of the four-vectors of the constituent jets).
- The mass of the W candidate (the W candidate four-vector is the sum of the four-vectors of the jets designated as the W subjets).
- The product of the top candidate p_{T} and ΔR between the b jet and the W candidate.
- The product of the W candidate p_{T} and ΔR between the W subjets.
- The "soft-drop condition" from the soft-drop declustering algorithm [60] reinterpreted as a variable over the two W subjets: $\frac{\min(p_{\mathrm{T}1}, p_{\mathrm{T}2})}{p_{\mathrm{T}1}+p_{\mathrm{T}2}} \Delta R_{j_1,j_2}^{-2}$. This measure tends

to reject relatively soft collinear jets.

- The DeepJet b-tag scores of all three constituent jets.
- The DeepJet c-tag scores of the W subjets.
- The quark-gluon likelihood scores of the W subjets [61].
- The jet constituent multiplicities of the W subjets.

Input variables were chosen based on the original algorithm, and modified and optimized to achieve best performance given NanoAOD compatability. The resultant BDT is trained using the "XGBoost" package, with genuine and fake hadronic top candidates taken from single-lepton and di-lepton $\mathrm{t\bar{t}}$ simulation samples respectively, using $\approx 100,000$ simulated 2016 events from each sample. Resolved top candidates for the analysis are selected using a medium working point, corresponding to BDT thresholds of 0.9988, 0.9992, and 0.9994 for 2016, 2017, and 2018, respectively. This working point corresponds to a misidentification rate of roughly 10% for jets from QCD multi-jet production, as shown in Fig. 7.3. The candidates are required to have a $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ and $|\eta| < 2.4$. If candidates overlap with one another within a ΔR of 0.4, the lower scoring candidate is rejected.

Control regions are used to derive scale factors to correct the efficiencies and misidentification rates in simulation to match data. These correction factors are parameterized in the candidate p_{T}, and uncertainties on measured scale factors are propagated as uncertainties in the simulation-based estimates.

A zero-lepton, $N_{\mathrm{b}} = 1$ region is used to derive the correction factors for the misidentification rate. The misidentification scale factor region was selected to be enriched in QCD multijet events, and is orthogonal to the signal region via an exclusively 1-b-tagged jet requirement. The strategy and selections used are modeled after those used to measure

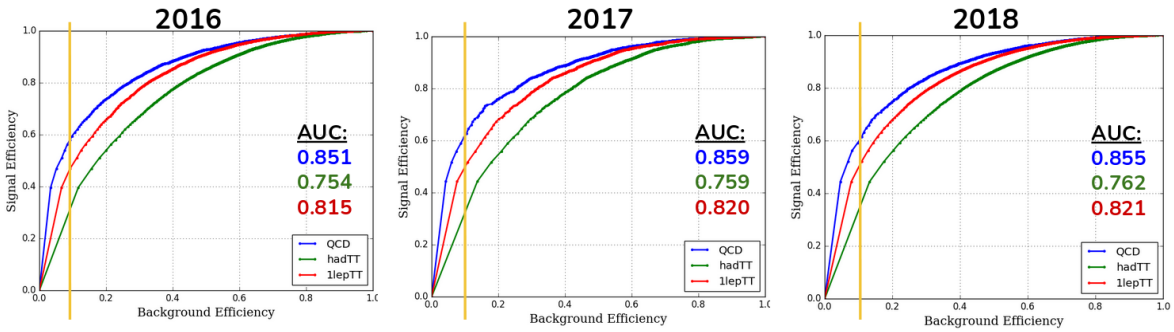

Figure 7.3: Resolved top candidate efficiency vs misidentification rate calculated for $t\bar{t}t\bar{t}$ signal vs QCD multijet, all-hadronic $t\bar{t}$, and single-lepton $t\bar{t}$ simulated events for 2016 (left), 2017 (center), and 2018 (right). The working point in each year (corresponding to a roughly 10% mistag rate in QCD) is indicated by a yellow line.

correction factors for the original version of the tagger. Events in this region are selected with pure H_T triggers, with further requirements of $N_b = 1, N_j \geq 6$, and $H_T > 1200\,\text{GeV}$. In this region, simulation is first renormalized to data. Figures 7.4-7.5 show the composition of events in this region. We correct for the estimated contamination from genuine tops by subtracting the number of candidates matched to top quarks at the generator level in simulation. Then, mistag rates are measured as a function of resolved top candidate p_T by calculating the fraction of 3-jet candidates passing the resolved top tagger working point. The scale factors are calculated as the ratio of the mistag rates measured in data (after the subtraction of the estimated contamination from genuinely tagged tops), and in simulation for candidates not matched to genuine tops. The scale factors binned in candidate p_T are shown in Fig. 7.6.

A single-muon control region is used to derive the correction factors for the tagging efficiency. This region was selected to be enriched in semi-leptonic $t\bar{t}$ events, with similar kinematics to the SR. The contamination from $t\bar{t}t\bar{t}$ signal is negligible in this region ($< 0.05\%$ of all candidates). The strategy used to measure these scale factors was inspired by the method used to measure scale factors for the DeepAK8 algorithm [62]. Events in this region are selected with single-muon triggers, with further requirements of $N_b \geq$

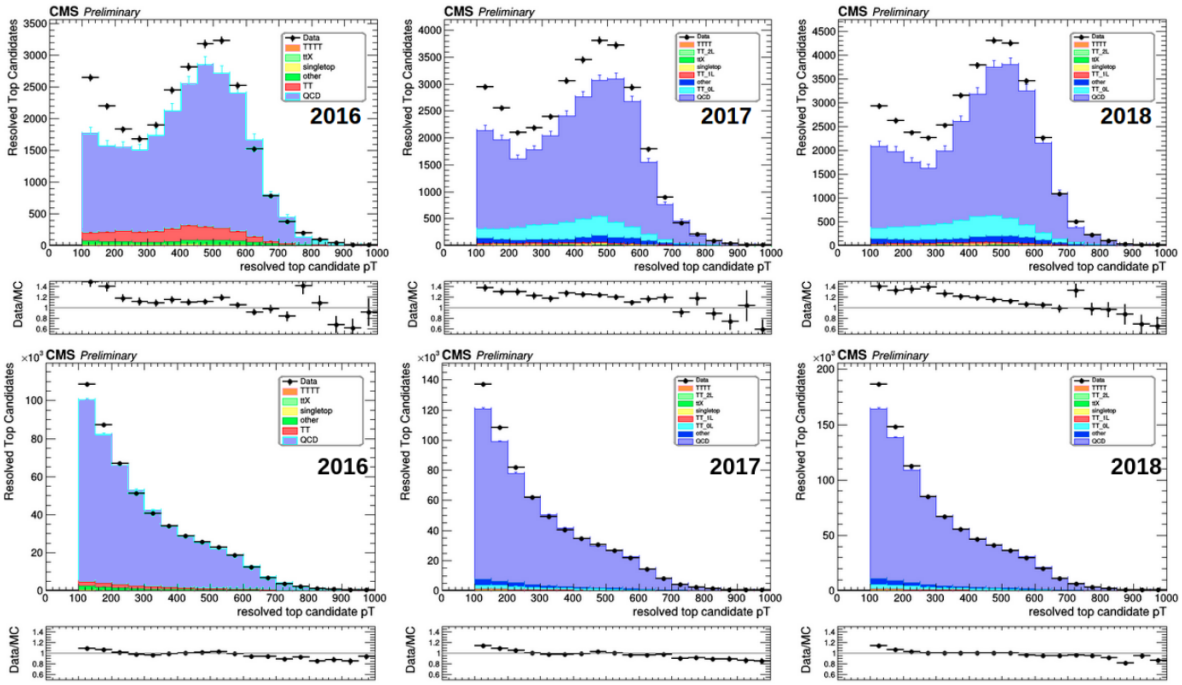

Figure 7.4: Distributions of resolved top candidate p_{T} in data and simulation for candidates passing the top tagger working point (upper row) and for all candidates (lower row) for 2016 (left), 2017 (middle), and 2018 (right), in the region used to derive the mistag rate scale factors. Contributions from different processes as estimated from simulation are shown in the stacked histograms. The event yield in simulation is scaled to match data inclusively in this region, prior to the application of the top tagger working point.

$2, N_{\mathrm{j}} \geq 4, p_{\mathrm{T}}^{\mathrm{miss}} > 75\,\mathrm{GeV}$, and muon $p_{\mathrm{T}} > 50\,\mathrm{GeV}$. Jets are cleaned against leptons and top candidates against muons (using ΔR of 0.4). The lepton selections used for the single-lepton control region are the same as those used in the single-lepton $\mathrm{t\bar{t}t\bar{t}}$ search [63].The scale factors are calculated using a template fit of the top candidate mass distribution to data in this region, for candidates passing and failing the top tagger working point. As in the mistag scale factor region, simulation is first normalized to data. The mistag scale factors described above are applied to candidates passing the working point but that are not matched to top quarks at the generator level. Candidates are split into three p_{T} regions: low $(100 - 300\,\mathrm{GeV})$, medium $(300 - 500\,\mathrm{GeV})$, and high $(> 500\,\mathrm{GeV})$. Two templates of simulated candidates binned in candidate mass

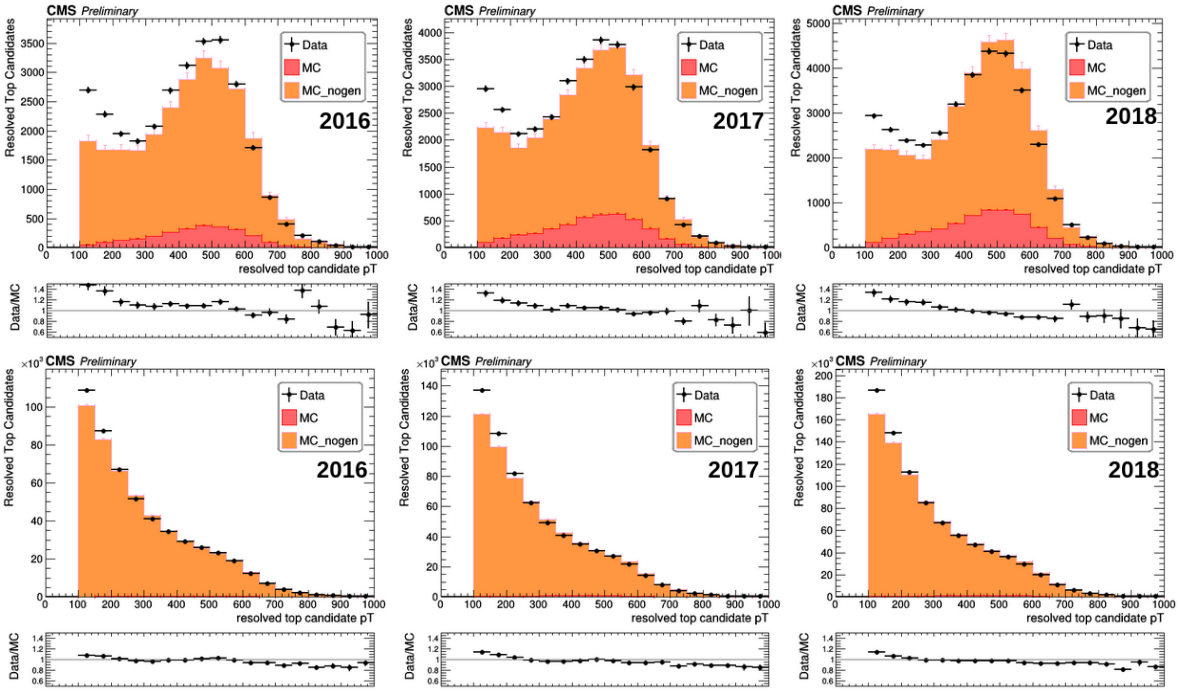

Figure 7.5: Distributions of resolved top candidate p_{T} in data and simulation for candidates passing the top tagger working point (upper row) and for all candidates (lower row) for 2016 (left), 2017 (middle), and 2018 (right), in the region used to derive the mistag rate scale factors. Contributions from candidates matched and not matched to top quarks at the generator level in simulation are shown in the stacked histograms. The event yield in simulation is scaled to match data inclusively in this region, prior to the application of the top tagger working point.

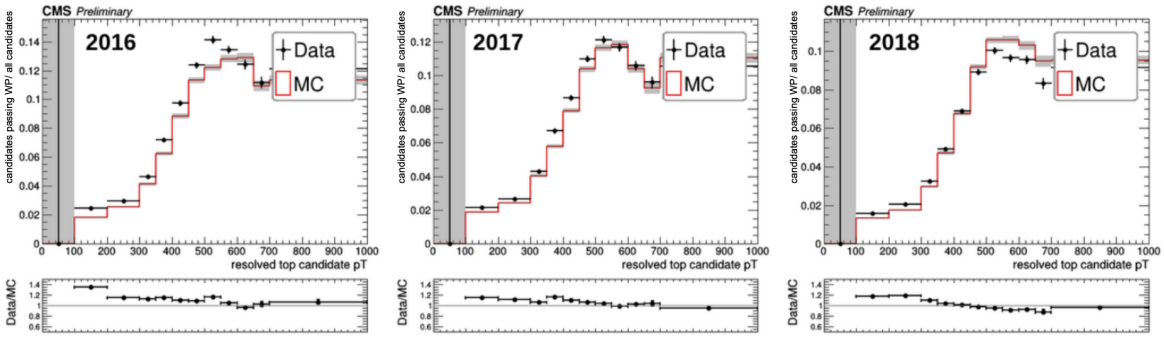

Figure 7.6: Mistag rates measured in data and simulation after subtracting the estimated contribution from genuine tops (ε = number of candidates passing the working point / total number of candidates) binned in candidate p_{T}. Mistag scale factors are defined as the ratio $\varepsilon_{\mathrm{data}}/\varepsilon_{\mathrm{MC}}$.

corresponding to candidates matched and unmatched to top quarks are fit to the data. The fit is performed simultaneously for candidates passing and failing the resolved top tagger working point. Scale factors are defined as $SF_{\text{eff}} = \varepsilon_{\text{data}}/\varepsilon_{\text{post-fit MC}}$ where ε is the number of candidates passing the working point divided by the total number of candidates and the post-fit simulation efficiency used for the efficiency scale factor is the one corresponding to matched candidates. Candidate mass is used for the fit distributions because it is independent of candidate p_{T}. Figures 7.7-7.8 show the composition of events in the region used to derive the scale factors. Figure 7.10 shows the results of the scale factor derivation from the simultaneous fit.

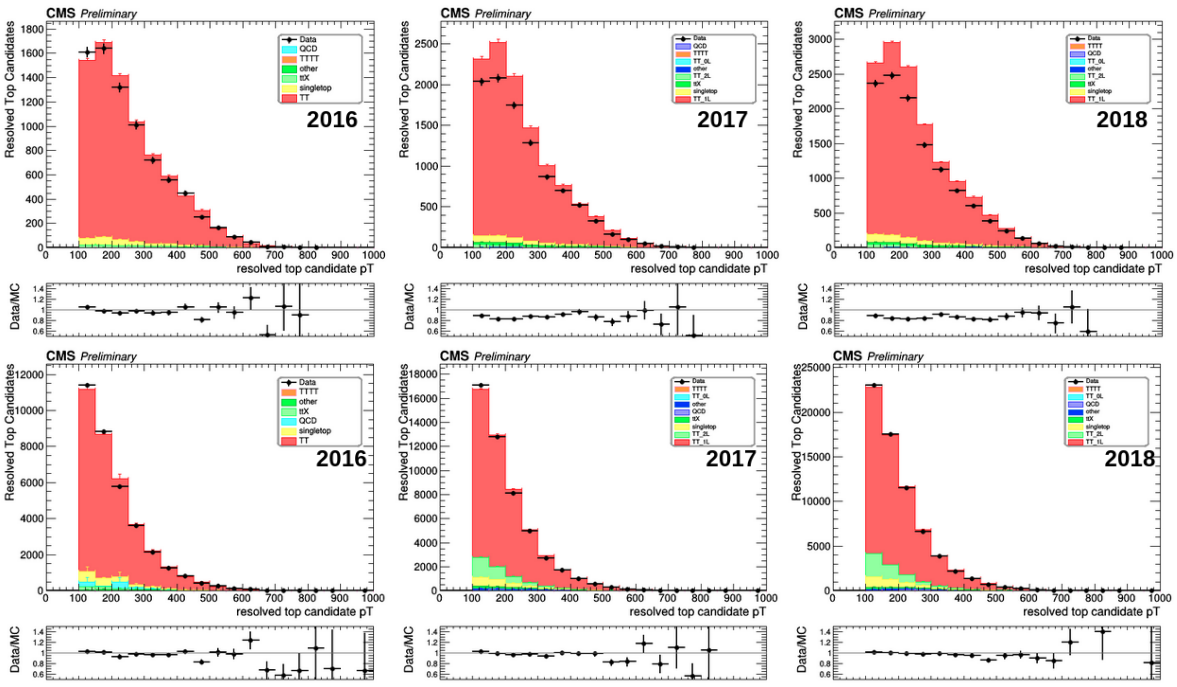

Figure 7.7: Distributions of resolved top candidate p_{T} in data and simulation for candidates passing the top tagger working point (upper row) and for all candidates (lower row) for 2016 (left), 2017 (middle), and 2018 (right), in the region used to derive the efficiency scale factors. Contributions from different processes as estimated from simulation are shown in the stacked histograms. The event yield in simulation is scaled to match data inclusively in this region, prior to the application of the top tagger working point.

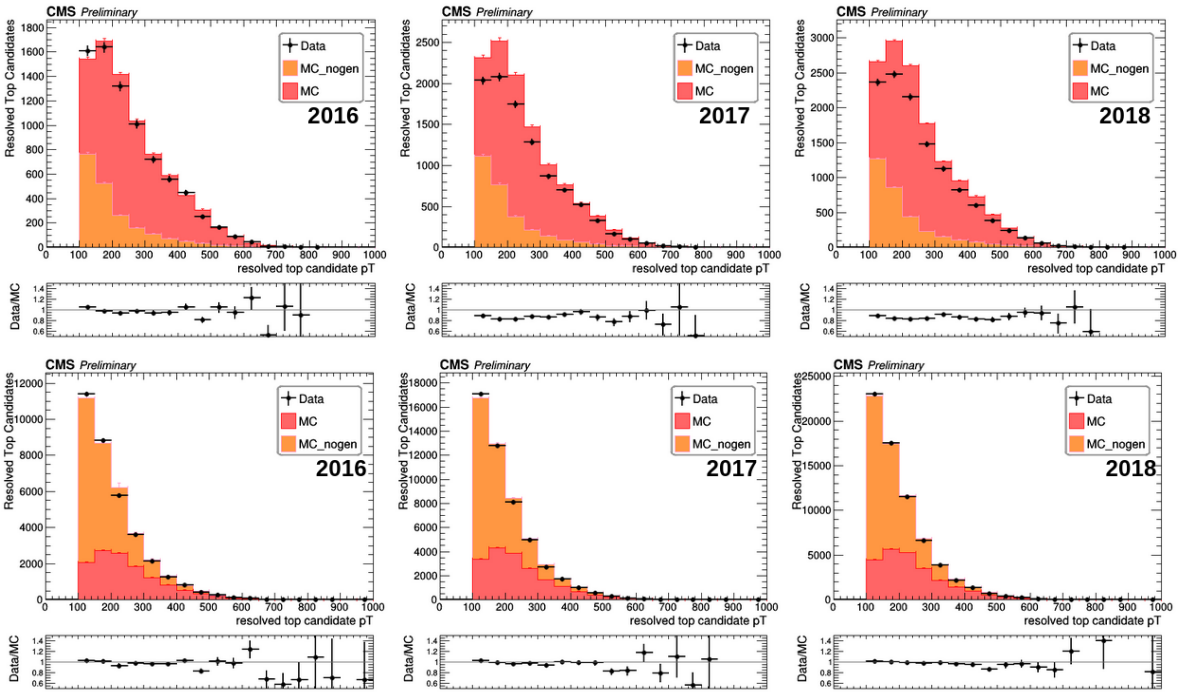

Figure 7.8: Distributions of resolved top candidate p_{T} in data and simulation for candidates passing the top tagger working point (upper row) and for all candidates (lower row) for 2016 (left), 2017 (middle), and 2018 (right), in the region used to derive the efficiency scale factors. Contributions from candidates matched and not matched to top quarks at the generator level in simulation are shown in the stacked histograms. The event yield in simulation is scaled to match data inclusively in this region, prior to the application of the top tagger working point.

7.4 Missing transverse energy

The raw PF $p_{\mathrm{T}}^{\mathrm{miss}}$ is computed as the negative vectorial sum of the transverse momenta of all PF candidates. Standard corrections [64] provided and recommended by the JetMET POG are applied to the $p_{\mathrm{T}}^{\mathrm{miss}}$. These corrections propogate jet energy corrections to $p_{\mathrm{T}}^{\mathrm{miss}}$.

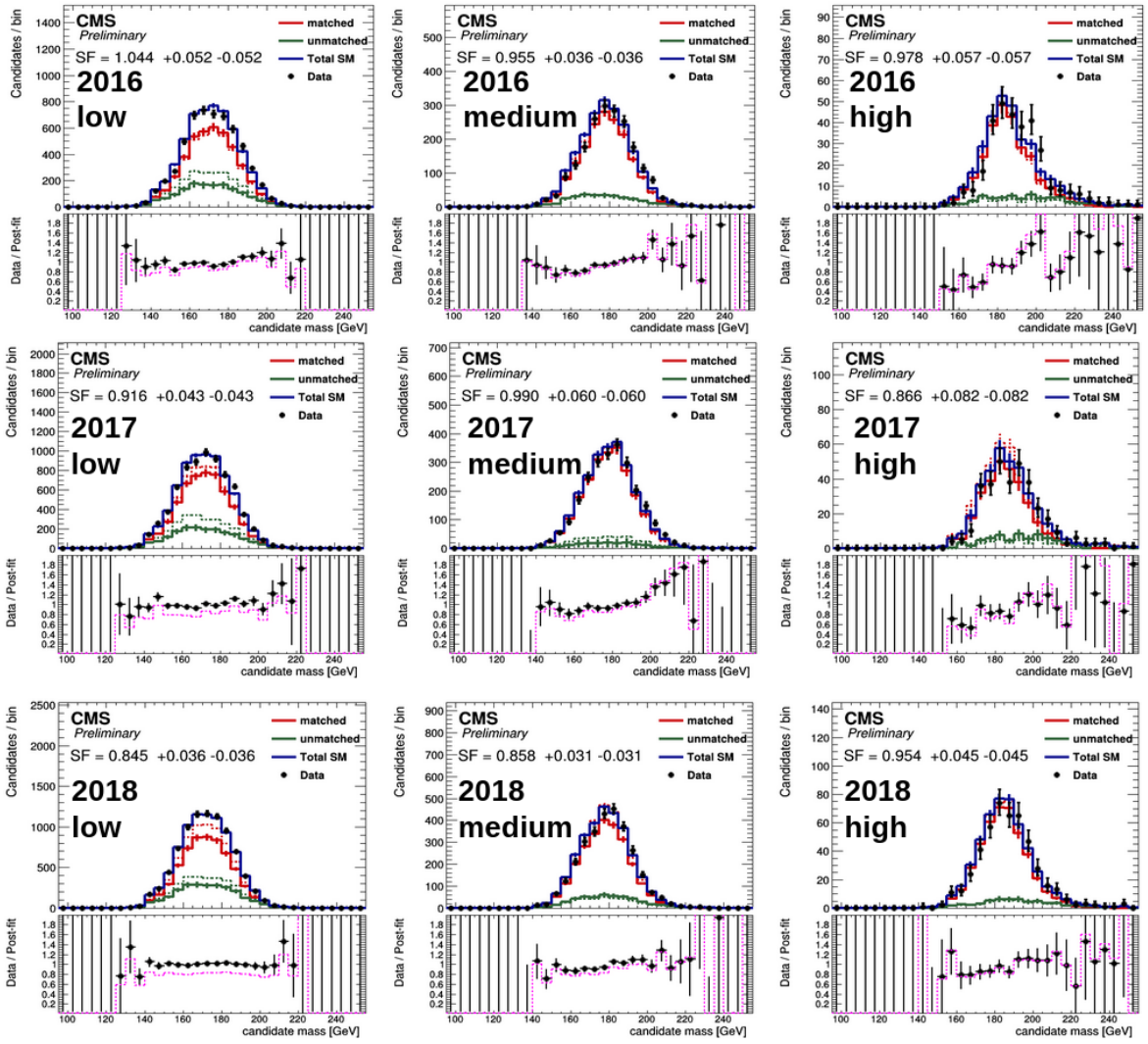

Figure 7.9: Data vs pre- and post-fit distributions in candidate mass for top candidates passing the working point in years 2016 (upper row), 2017 (middle row), and 2018 (lower row) for low (left column), medium (middle column), and high (right column) p_{T} ranges. Solid lines correspond to post-fit distributions and dashed lines to pre-fit distributions. Candidates matched to generator-level tops are shown in red, and the scale factor extracted from the fit is shown. Candidates unmatched to generator-level tops and total candidates are shown in green and blue respectively. The ratio of data to the total pre- and post-fit simulation is shown in the ratio panels.

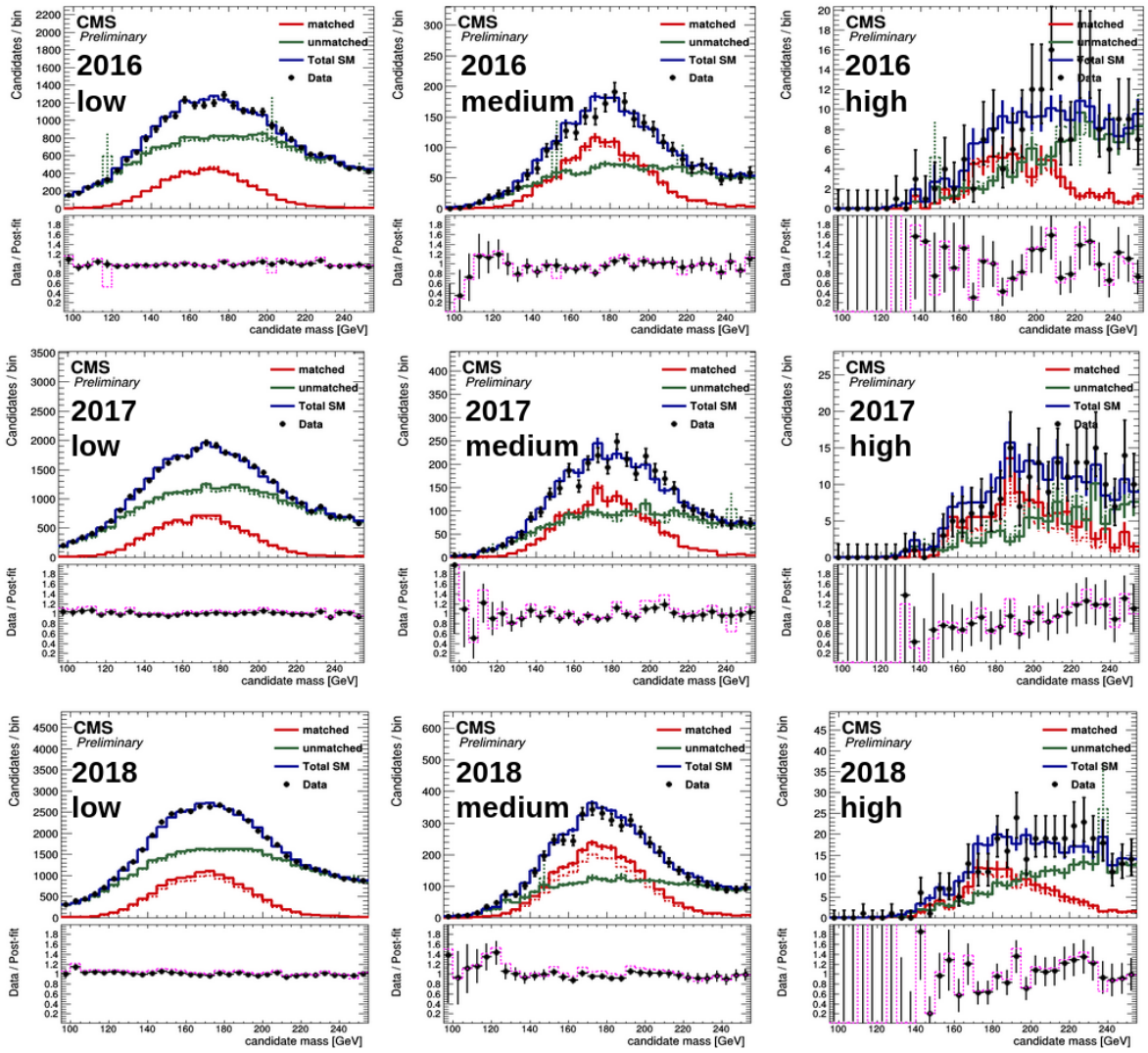

Figure 7.10: Data vs pre- and post-fit distributions in candidate mass for top candidates failing the working point in years 2016 (upper row), 2017 (middle row), and 2018 (lower row) for low (left column), medium (middle column), and high (right column) p_{T} ranges. Solid lines correspond to post-fit distributions and dashed lines to pre-fit distributions. Candidates matched to generator-level tops are shown in red, and the scale factors extracted from this fit are the same as in 7.9. Candidates unmatched to generator-level tops and total candidates are shown in green and blue respectively. The ratio of data to the total pre- and post-fit simulation is shown in the ratio panels.

Chapter 8

Triggers and datasets

This chapter gives an overview of the triggers, datasets, and simulated events used in the all-hadronic $t\bar{t}t\bar{t}$ analysis. Triggers are typically selected based on their likelihood of containing the process of interest while still efficiently representing the overall data collected in CMS. Datasets are the ensemble of experimental events selected given these triggers. Simulated events are used to model backgrounds and signal, and those utilized in this analysis are also discussed.

The data for this search are recorded using a suite of cross-triggers requiring the presence of $\geq$ 6 jets, $\geq$ 1 or $\geq$ 2 b-tagged jets, and large H_{T}. For part of the 2017 run, a 4-jet, 3-b jet, high H_{T} trigger is included as well in order to maximize the trigger efficiency. The HLT paths of the triggers used for the 2016, 2017, and 2018 data-taking periods are listed in Table 8.1. These triggers were selected to cater to the large H_{T}, and jet and b-tagged jet multiplicites expected for $t\bar{t}t\bar{t}$ signal in the all-hadronic final state.

Trigger efficiencies measured in data were computed and applied as corrections to simulated events in order to correctly represent detector performance. The efficiency of the triggers used in this analysis was measured in an independent sample selected with a single muon trigger (HLT_IsoMu24 or HLT_IsoMu27). The efficiency is measured

Table 8.1: HLT paths corresponding to the triggers used for the search in 2016, 2017, and 2018.

Data-taking period	HLT path
2*2016	HLT_PFHT400_SixJet30_DoubleBTagCSV_p056
	HLT_PFHT450_SixJet40_BTagCSV_p056
5*2017	HLT_PFHT380_SixJet32_DoubleBTagCSV_p075
	HLT_PFHT430_SixJet40_BTagCSV_p080
	HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2
	HLT_PFHT430_SixPFJet40_PFBTagCSV_1p5
4*2018	HLT_PFHT380_SixPFJet32_DoublePFBTagDeepCSV_2p2
	HLT_PFHT430_SixPFJet40_PFBTagDeepCSV_1p5
	HLT_PFHT400_SixPFJet32_DoublePFBTagDeepCSV_2p94
	HLT_PFHT450_SixPFJet36_PFBTagDeepCSV_1p59

in the 2D plane of N_{b} (number of b-jets) and N_{j} (number of jets) in the region where the efficiency of H_{T} leg is 100%. For each N_{b} and N_{j} bin in $N_{\mathrm{j}} \geq 7, N_{\mathrm{b}} \geq 2$, and $H_{\mathrm{T}} > 900$ GeV, the trigger efficiency was computed as follows:

$$\epsilon(N_{\mathrm{j}}, N_{\mathrm{b}}) = \frac{\text{Number of events passing OR of triggers and denominator selection}}{\text{Number of events that passed HLT\_IsoMu24 or HLT\_IsoMu27}}. \tag{8.1}$$

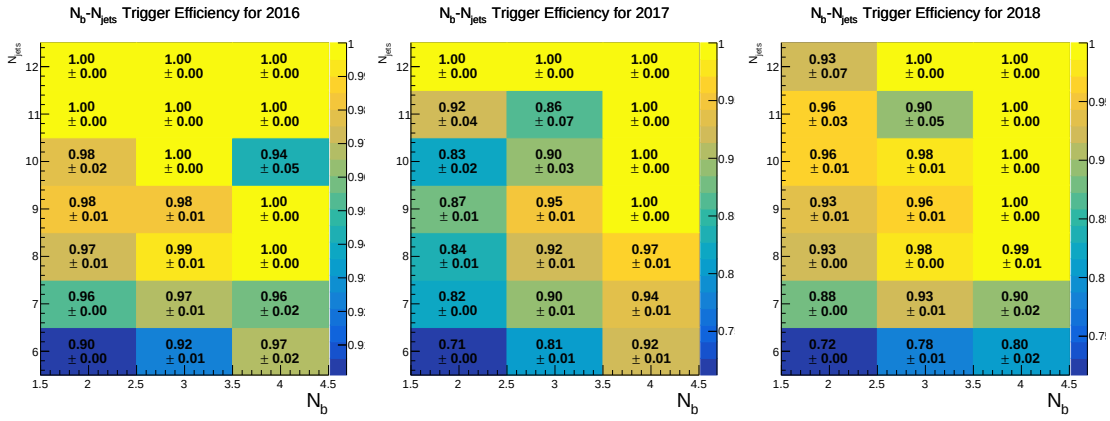

Figure 8.1: Efficiency measured for search triggers as a function of N_{j} and N_{b} for 2016 (left), 2017 (middle), and 2018 (right) data. The $N_{\mathrm{j}} = 6$ region used for the validation test is shown as well. Plots by Chang Hwan Jung and Jae Hyeok Yoo.

Figure 8.1 shows the measured efficiency in the N_{j} and N_{b} plane for each year. The dependence on N_{b} and N_{j} justifies the necessity of applying N_{j} and N_{b} dependant correc-

tions to simulated events in order to reflect trigger efficiencies in data. The bottom row corresponds to $N_{\mathrm{j}} = 6$ and shows low efficiencies due to the requirement of $N_{\mathrm{j}} \geq 6$ in the trigger paths. This region is used only for the validation of the background estimation methods, as discussed in Section 10.3.

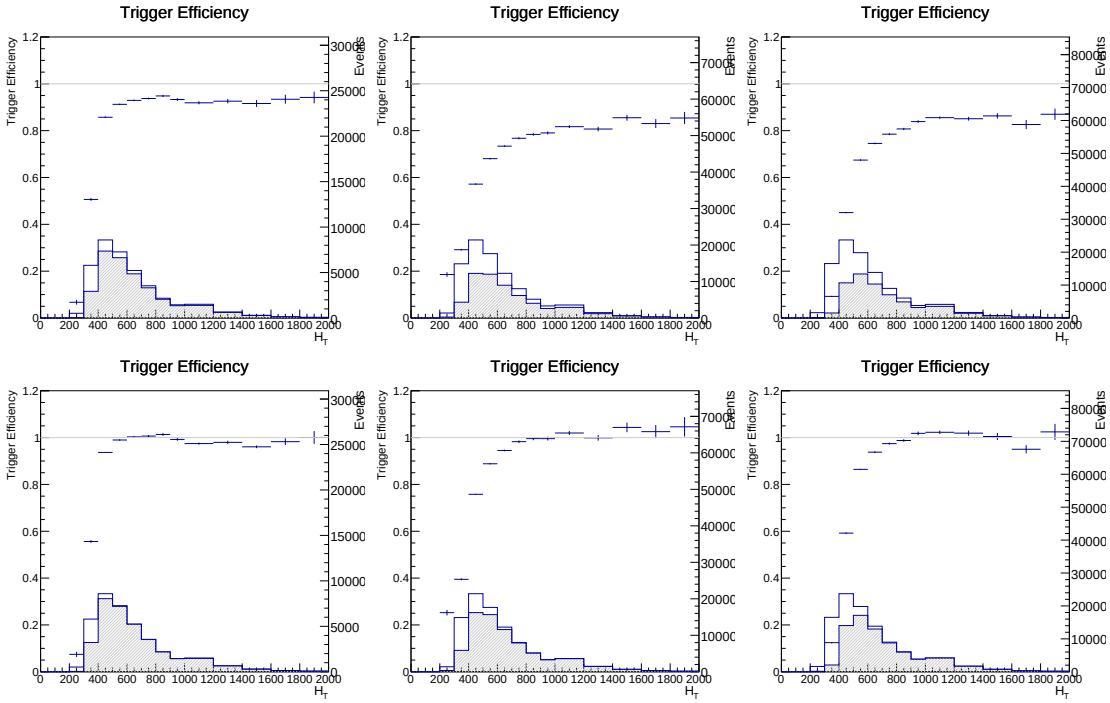

Figure 8.2: Trigger turn-on vs H_{T} before (top) and after (bottom) correcting for efficiencies measured as function of N_{j} and N_{b} for 2016 (left), 2017 (middle), and 2018 (right) data. Plots by Chang Hwan Jung and Jae Hyeok Yoo.

Figure 8.2 shows the H_{T} turn-on for each year before and after applying the N_{j}- and N_{b}-dependent corrections. Before the corrections, the efficiency in the plateau region ($H_{\mathrm{T}} > 900$ GeV) deviates from unity. After the corrections are applied, the plateau efficiency is close to 1. A correction for the residual inefficiency in the $700 < H_{\mathrm{T}} < 900$ GeV region is also applied. Trigger studies for this analysis were done by Chang Hwan Jung and Jae Hyeok Yoo.

Datasets used in this analysis are listed in Table 8.2. The simulated Monte Carlo (MC) samples used to model background processes and $\mathrm{t\bar{t}t\bar{t}}$ signal are listed in Tables

8.3-8.5. The NanoAODv6 data format is used for all datasets. Some samples, such as simulated $t\bar{t}$ and QCD datasets, are exclusively used for testing and validation.

The following corrections are applied to simulated samples in order to reduce discrepancies between data and simulation: b-tagging scale factors (calculated using the iterative fit method 1(d)) along with N_j and H_T dependent reweighting as described in Appendix C, reweighting of pileup distributions, trigger efficiency corrections, scale factors to correct the efficiencies for boosted W and top tagging using the DeepAK8 algorithm, and scale factors to correct the efficiency and misidentification rate for the resolved top tagger as described in Section 7.3.2. Scale factors to account for discrepancies in lepton veto efficiencies were found to be so close to 1 as to have a negligible effect.

Table 8.2: Datasets used in the analysis.

Year	Dataset
7*2016 ($35.9\,\text{fb}^{-1}$)	`/JetHT/Run2016B_ver2-Nano25Oct2019_ver2-v1`
	`/JetHT/Run2016C-Nano25Oct2019-v1`
	`/JetHT/Run2016D-Nano25Oct2019-v1`
	`/JetHT/Run2016E-Nano25Oct2019-v1`
	`/JetHT/Run2016F-Nano25Oct2019-v1`
	`/JetHT/Run2016G-Nano25Oct2019-v1`
	`/JetHT/Run2016H-Nano25Oct2019-v1`
5*2017 ($41.5\,\text{fb}^{-1}$)	`/JetHT/Run2018B-Nano25Oct2019-v1`
	`/JetHT/Run2018C-Nano25Oct2019-v1`
	`/JetHT/Run2018D-Nano25Oct2019-v1`
	`/JetHT/Run2018E-Nano25Oct2019-v1`
	`/JetHT/Run2018F-Nano25Oct2019-v1`
4*2018 ($59.7\,\text{fb}^{-1}$)	`/JetHT/Run2018A-Nano25Oct2019-v1`
	`/JetHT/Run2018B-Nano25Oct2019-v1`
	`/JetHT/Run2018C-Nano25Oct2019-v2`
	`/JetHT/Run2018D-Nano25Oct2019-ver2-v1`

Table 8.3: Simulated samples used for the analysis of 2016 data. The datasets used are from the RunIISummer16NanoAODv6 campaign.

Category	Sample name	Cross section [pb]
$t\bar{t}t\bar{t}$	TTTT_PSWeights_TuneCUETP8M2T4_13TeV-amcatnlo-pythia8	0.012
$t\bar{t}$	TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	831.76
QCD multi-jet	QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1710000
	QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	348000
	QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	32100
	QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6830
	QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1210
	QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	120
	QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	25.3
$t\bar{t}$X (X=W,Z,H)	TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.5297
	TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.2529
	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
	ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8	0.2934
	ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8	0.2151
	TTGJets_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	3.697
tZq	tZq_ll_4f_13TeV-amcatnlo-pythia8	0.0758
	tZq_nunu_4f_13TeV-amcatnlo-pythia8_TuneCUETP8M1	0.1379
Single-top	ST_s-channel_4f_InclusiveDecays_13TeV-amcatnlo-pythia8	10.32
	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.85
	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.85
	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	136.02
	ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	80.95
Other backgrounds	WJetsToQQ_HT400to600_qc19_3j_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	270
	WJetsToQQ_HT600to800_qc19_3j_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	59.1
	WJetsToQQ_HT-800toInf_qc19_3j_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	30.5
	WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	49.1
	WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	12.1
	WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.47
	WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.33
	WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.032
	ZJetsToQQ_HT400to600_qc19_4j_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	115
	ZJetsToQQ_HT600to800_qc19_4j_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	27.5
	ZJetsToQQ_HT-800toInf_qc19_4j_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	14.8
	ZJetsToNuNu_HT-400To600_13TeV-madgraph	3.59
	ZJetsToNuNu_HT-600To800_13TeV-madgraph	0.857
	ZJetsToNuNu_HT-800To1200_13TeV-madgraph	0.394
	ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.096
	ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.002
	WW_TuneCUETP8M1_13TeV-pythia8	110.8
	WZ_TuneCUETP8M1_13TeV-pythia8	47.13
	ZZ_TuneCUETP8M1_13TeV-pythia8	16.523
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	6077
	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	18610

Table 8.4: Simulated samples used for the analysis of 2017 data. The datasets used are from the RunIIFall17NanoAODv6 campaign.

Category	Sample name	Cross section [pb]
$t\bar{t}t\bar{t}$	TTTT_TuneCP5_13TeV-amcatnlo-pythia8	0.012
$t\bar{t}$	TTToHadronic_TuneCP5_13TeV-powheg-pythia8	377.96
	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	365.34
	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	88.29
QCD multi-jet	QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8	1550000
	QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8	323000
	QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8	30100
	QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8	6330
	QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8	1090
	QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8	99.1
	QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8	20.2
$t\bar{t}$X (X=W,Z,H)	TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
	TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8	0.2519
	TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8	0.5297
	ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8	0.2934
	ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8	0.2151
	TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	3.697
Single-top	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8	3.36
	ST_s-channel_4f_hadronicDecays_TuneCP5_13TeV-amcatnlo-pythia8	6.96
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
	ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	136.02
	ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	80.95
Other backgrounds	WJetsToQQ_HT400to600_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8	316
	WJetsToQQ_HT600to800_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8	68.7
	WJetsToQQ_HT-800toInf_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8	34.7
	WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	58
	WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	13.1
	WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	5.4
	WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.07
	WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.008
	ZJetsToQQ_HT400to600_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8	145
	ZJetsToQQ_HT600to800_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8	34.4
	ZJetsToQQ_HT-800toInf_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8	18.5
	ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.1
	ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.26
	ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.5
	ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.342
	ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.005
	WW_TuneCP5_13TeV-pythia8	110.8
	WZ_TuneCP5_13TeV-pythia8	47.13
	ZZ_TuneCP5_13TeV-pythia8	16.523
	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	6077
	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610

Table 8.5: Simulated samples used for the analysis of 2018 data. The datasets used are from the RunIIAutumn18NanoAODv6 campaign.

Category	Sample name	Cross section [pb]
$t\bar{t}t\bar{t}$	TTTT_TuneCP5_13TeV-amcatnlo-pythia8	0.012
$t\bar{t}$	TTToHadronic_TuneCP5_13TeV-powheg-pythia8	377.96
	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	365.34
	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	88.29
QCD multi-jet	QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8	1550000
	QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8	323000
	QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8	30100
	QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8	6330
	QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8	1090
	QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8	99.1
	QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8	20.2
$t\bar{t}$X (X=W,Z,H)	TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
	TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8	0.2529
	TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8	0.5297
	ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8	0.2934
	ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8	0.2151
	TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	3.697
Single-top	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8	3.36
	ST_s-channel_4f_hadronicDecays_TuneCP5_13TeV-amcatnlo-pythia8	6.96
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
	ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	136.02
	ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	80.95
Other backgrounds	WJetsToQQ_HT400to600_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8	316
	WJetsToQQ_HT600to800_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8	68.8
	WJetsToQQ_HT-800toInf_qc19_3j_TuneCP5_13TeV-madgraphMLM-pythia8	34.4
	WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	57.9
	WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	13
	WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	5.56
	WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.09
	WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.008
	ZJetsToQQ_HT400to600_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8	145
	ZJetsToQQ_HT600to800_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8	33.8
	ZJetsToQQ_HT-800toInf_qc19_4j_TuneCP5_13TeV-madgraphMLM-pythia8	18.7
	ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.1
	ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.26
	ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.5
	ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.349
	ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.005
	WW_TuneCP5_13TeV-pythia8	110.8
	WZ_TuneCP5_13TeV-pythia8	47.13
	ZZ_TuneCP5_13TeV-pythia8	16.523
	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	6077
	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610

Chapter 9

Event Selection

This chapter describes criteria used in this analysis to target a signal rich selection of data events. In order to isolate signal-like events in a statistically significant way, a general baseline selection is first applied to reduce background and ensure quality of data. This baseline selection is split into smaller "signal region" (SR) categories with stricter (more signal-like) selection requirements. These SR categories include both statistics rich but signal reduced categories and lower statistics but signal enhanced "sensitive" categories. Finally, within those SR categories, an event-level BDT is used to define a variable (BDT discriminant) that separates signal-like and background-like events by assigning each event a score between 0 ("background-like") and 1 ("signal-like"). The shape of this variable within SR categories is used in a profile maximum likelihood analysis to test the hypothesis of the presence of entirely hadronic four-top signal events vs. a background-only hypothesis. The baseline selection, SR categories, and BDT algorithm are defined and described in detail in this section.

9.1 Baseline and signal region selection

The baseline selection used in this analysis was chosen to target hadronic four-top events. This selection is applied to events passing the triggers described in in Chapter 8 and uses criteria and definitions described in Chapter 7 including requiring zero leptons. In order to reduce $t\bar{t}$ backgrounds, a high number of jets (at least 9) are required, at least 3 of which are b-tagged. A high H_T cut ($H_{\mathrm{T}} \geq 700\,\mathrm{GeV}$) is additionally applied to reduce QCD backgrounds, as heavy flavor decays (like the decay of top quarks) can be expected to have higher H_T compared to the light-flavor decays that largely dominate QCD events.

The signal region (SR) is defined using this baseline with the additional requirement of at least one tagged resolved top. Events in the SR are then subdivided into 12 categories based on the number of tagged resolved tops (N_{RT}), the number of tagged boosted tops (N_{BT}), and H_{T}. Table 9.1 defines these categories. These categories were optimized in order to maintain adequate statistics while maximizing signal sensitivity.

Table 9.1: Definitions of the SR categories based on the number of resolved tops (N_{RT}), number of boosted tops (N_{BT}), and H_{T}.

Top tags	H_{T} [GeV]							
$N_{\mathrm{RT}} = 1, N_{\mathrm{BT}} = 0$	700–800	800–900	900–1000	1000–1100	1100–1200	1200–1300	1300–1500	>1500
$N_{\mathrm{RT}} = 1, N_{\mathrm{BT}} \geq 1$	700–1400							>1400
$N_{\mathrm{RT}} \geq 2$	700–1100					>1100		

9.2 Event-level BDT

An event-level BDT, implemented using the "CatBoost" library [65], is used to further discriminate between signal and background in each SR category. The BDT is trained to discriminate $t\bar{t}t\bar{t}$ simulated events against simulated background events from $t\bar{t}$ and QCD multijet production after the baseline selection. The categorical features among

the input variables are treated using one-hot encoder in CatBoost. 2016 simulation samples are used for training. Several different algorithms were tested, but the CatBoost BDT algorithm was found to have the best performance compared to BDTs using the XGBoost package and neural net (NN) implementations. Additionally, training against a background composition including both $t\bar{t}$ and QCD multijet events was found to provide better discrimination than training against $t\bar{t}$ events alone.

Input variables for this BDT algorithm were chosen to take advantage of different kinematic properties of $t\bar{t}t\bar{t}$ signal compared to major ($t\bar{t}$ and QCD) backgrounds. The kinematics of jets, b-tagged jets and associated variables were tested as inputs for the BDT in order to take advantage of differing jet multiplicities and kinematics between signal and backgrounds. Variables similar to those used in the MVA developed for the CMS single-lepton $t\bar{t}t\bar{t}$ analysis with 2016 data [66] were tested. The number of boosted W candidates and the p_{T} of the leading resolved top candidates and b-tagged jets were also tested as inputs to help distinguish signal from QCD multijet and $t\bar{t}$ background events, which tend to have fewer boosted, heavy flavour, or high p_{T} objects. Furthermore, variables were tested in order to take advantage of differing signal and background event topologies. H_T variables were tested in order to distinguish $t\bar{t}t\bar{t}$ from QCD multijet events based on heavier vs. lighter flavor compositions as previously discussed. Hadronic $t\bar{t}t\bar{t}$ events should also differ from hadronic $t\bar{t}$ background events in that jets from $t\bar{t}$ events should have a less isotropic distribution due to their recoil from sources such as initial-state radiation (ISR). Event shape variables were thus also tested to potentially take advantage of this characteristic.

The number of resolved tops and boosted tops were not included as inputs in BDT training in order to use them as (somewhat) independent variables for binning events in SR categories. Resolved and boosted top discriminants were also disregarded as inputs to the BDT in order to avoid introducing a stronger dependence on the shapes of these

discriminants, which could potentially increase systematic uncertainties. In the end, inputs to the BDT were narrowed down to the following optimized set of variables:

- The number of jets present in the event, N_{j}
- The number of b-tagged jets present in the event, N_{b}
- The number of boosted W candidates
- The sum of the masses of $R = 0.8$ jets
- The missing transverse energy, $p_{\mathrm{T}}^{\mathrm{miss}}$
- The scalar sum of p_{T} of jets, H_{T}
- The scalar sum of p_{T} of b-tagged jets
- The $p_{\mathrm{T}}^{\mathrm{miss}}$ divided by square root of H_{T}
- The p_{T} of the leading b jet
- The p_{T} of the leading resolved top candidate
- The η difference between the leading and sub-leading jets
- The η difference between the leading and sub-leading b-tagged jets
- The absolute ϕ difference between the leading and sub-leading jets
- The absolute ϕ difference between the leading and sub-leading b-tagged jets
- The mean of the DeepJet b-tag scores of the b jets in the event
- The H_{T} of the six highest-p_{T} jets divided by the total H_{T} in the event
- The transverse momenta of the jet with the seventh-largest pT in the event

- Hadronic centrality (C), defined as the value of H_{T} divided by the sum of the energies of all jets in the event

- Event sphericity (S), calculated from all of the jets in the event in terms of the tensor $S^{\alpha\beta} = \sum_i p_i^{\alpha} p_i^{\beta} / \sum_i |\vec{p}_i|^2$, where α and β refer to the three-components of the momentum of the ı-th jet. The sphericity is then $S = (3/2)(\lambda_2 + \lambda_3)$, where λ_2 and λ_3 are the two smallest eigenvalues of $S^{\alpha\beta}$.

- Event aplanarity (A), defined as $A = (3/2)(\lambda_3)$

Tests were performed to validate and check the performance of this event-level BDT. Comparisons between input variable distributions for signal and background are shown in Fig. B.1 in Appendix B, although these comparisons use simulated QCD samples rather than data-driven QCD distributions like those used in this analysis. Additionally, as shown in Fig. 9.1, training and validation samples show similar distributions of BDT discriminant values. In Fig. 9.2, ROC curves of BDT discriminant values for events passing the baseline selection (left) and SR selection (right) using 2016 simulation samples are plotted. These demonstrate that the BDT has the ability to distinguish between signal the background even after the additional SR requirement of at least one resolved top relative to the baseline. The BDT discriminant distribution for simulated vs. data events is shown in Fig. 9.3 in a validation region requiring no resolved tops. ROC curves showing similar performance in all 3 years (2016, 2017 and 2018) can be found in Fig. B.2 in Appendix B. BDT studies were done as a collaborative effort between the author and Hayoung Oh of Korea University.

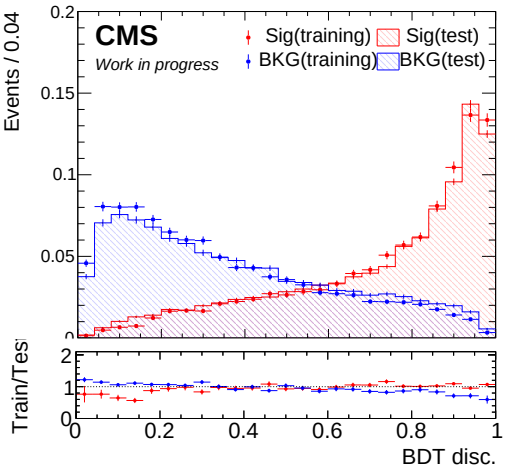

Figure 9.1: The discriminator distributions for the BDT classifier for signal and background in training and validation samples. Plot by Hayoung Oh.

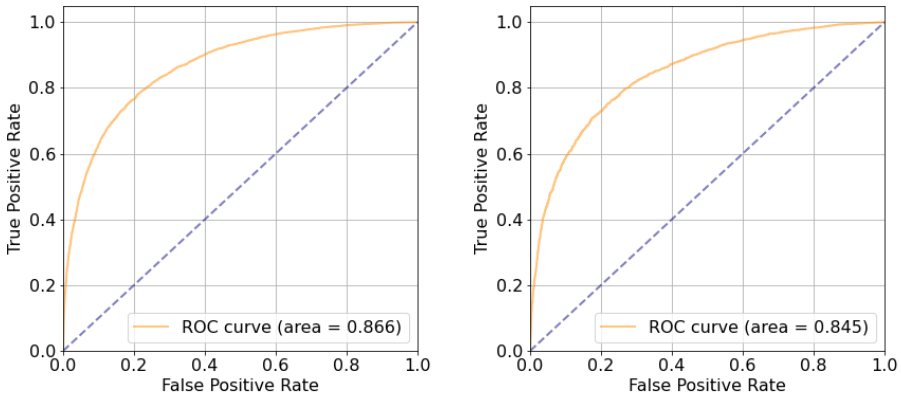

Figure 9.2: ROC curves of the BDT discriminant value for events passing the baseline selection (left) and SR selection (right). Plots by Hayoung Oh.

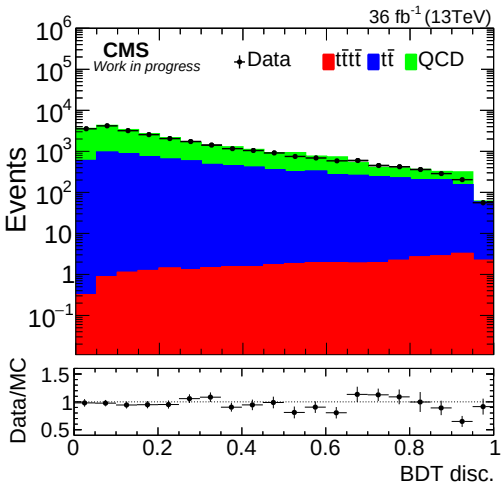

Figure 9.3: BDT discriminant distribution for 2016 data and simulation events for a 0-resolved top validation region, defined by requiring that events pass the baseline selection and do not have a resolved top candidate. Only statistical uncertainties are shown. Plot by Hayoung Oh.

Chapter 10

Background estimation

As noted in Chapter 6 QCD multijet and hadronic $t\bar{t}$ are the most important backgrounds in the four-top all-hadronic final state, largely due to their relatively large cross sections and similarity to signal (as visually demonstrated in Figure 10.1). Other potential backgrounds include hadronic $t\bar{t}$ +(H, W, Z, or gluons), single-top production, dibosons (WW, WZ or ZZ), W or Z +jets, or Drell-Yan. A chart representing the rough contribution of these backgrounds after baseline selection can be found in Figure 10.2. These backgrounds share many characteristics with $t\bar{t}t\bar{t}$ signal, in that they are largely made up of handronic activity and contain objects (tops, W hadrons, b-jets) that can be misreconstructed to look like tops or other objects. With the added complication of ISR and FSR jets many of these backgrounds can potentially very closely resemble signal. In the case of the dominant QCD multijet and hadronic $t\bar{t}$ backgrounds, the use of Monte Carlo (MC) simulation is problematic due to large uncertainties from next-to-leading order (NLO) calculations in strong interactions and limited statistics. Indeed, it was found that a MC simulation-based approach was unable to robustly predict QCD multijet and $t\bar{t}$ +jets backgrounds in signal-like high jet and b jet multiplicity regions. This problem most acutely impacts the QCD multijet background but it was found that the $t\bar{t}$ +jets

background could also greatly benefit from better modeling and statistics as the background most difficult to separate from $t\bar{t}t\bar{t}$ signal. Therefore, data-driven methods are used en lieu of simulation to predict both of these dominant backgrounds, while other, more minor, backgrounds are predicted with the simulated samples listed in Chapter 8. Two data-driven techniques are used to estimate 1)the absolute rate and 2)the shape of the data-driven QCD multijet and hadronic $t\bar{t}$ backgrounds (the "DD" backgrounds). Both methods are inspired by the common "ABCD" method, which identifies three orthogonal control regions (CRs) similar to the signal region (SR) and uses proportionality relations between those regions to make a prediction about the background in the SR given the background in the CRs. These methods are described in the following sections. Suyong Choi and Hayoung Oh were heavily involved in the development of these techniques.

10.1 The extended ABCD relation

Data-driven background estimation methods typically extrapolate information from at least one signal-depleted CR in order to predict backgrounds in the SR. By definition these CRs are distinct from the SR in that they are orthogonal in at least one variable but otherwise are very similar in order to justify this extrapolation. In this analysis, CRs are defined to be orthogonal to the SR in two variables: the number of jets N_{j} and the number of b-jets N_{b}. Selections are otherwise identical to the baseline selection. In the "usual" ABCD method of extrapolation, the phase space is divided into four regions, one of which is the SR and the remaining three CRs. The CRs are defined to have one fewer number of jets or b-jets than the SR. The information from the 3 sidebands, the A, B, and C CRs, is used to estimate the number of background events in the signal region, D through proportionality (see Fig. 10.3). In order for this to be the case, it is assumed

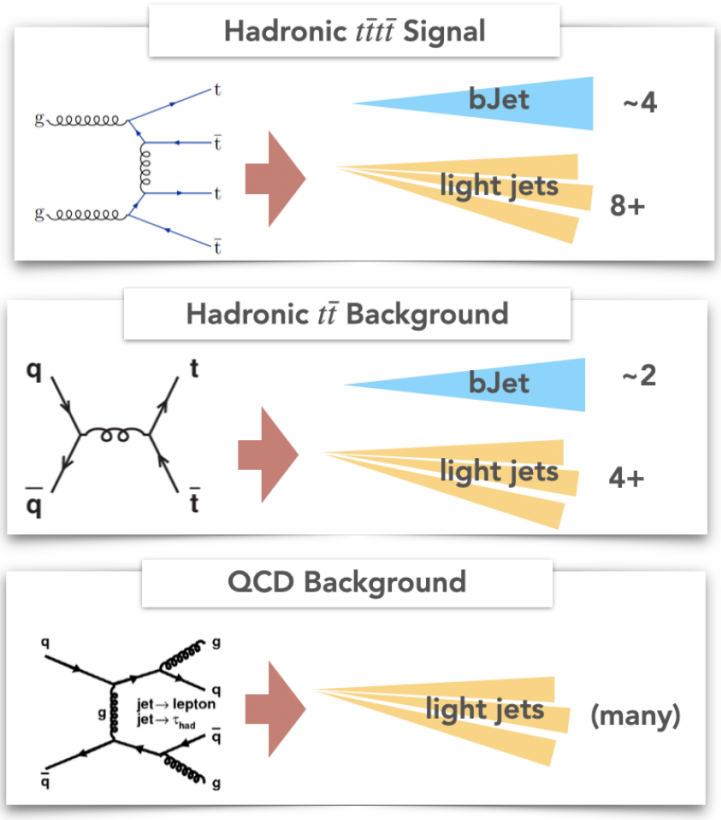

Figure 10.1: Schematic of expected final signatures of $t\bar{t}t\bar{t}$ signal vs. dominant backgrounds. Note these signatures, in practice, are impacted by detection and reconstruction efficiencies and ISR/FSR among other effects, making these signatures difficult to distinguish from one another. Furthermore, hadronic $t\bar{t}$ and QCD multijet backgrounds have cross sections roughly 70 thousand and 80 million times that of $t\bar{t}t\bar{t}$, respectively.

that the N_{j} and N_{b} distributions are mostly factorizable, that signal in the CRs is small compared to the SR, and that CRs are similar to the SR in terms of background shape and composition. Formally, the estimate of the number of background events in region D, D_{pred}, can be obtained by the following relation between the CRs: $D_{pred} = \frac{B \cdot C}{A}$.

It was found that the accuracy of the background yield estimates can be improved by expanding this proportionality relation to include two additional CRs [67]. This means that 5 CRs are used instead of the usual 3, by adding two CRs at lower jet multiplicity. The newly added control regions are labeled X and Y as shown in Fig 10.3.

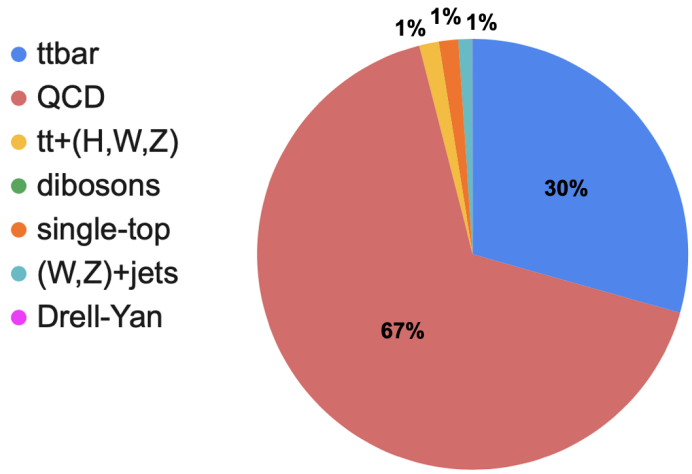

Figure 10.2: Approximate makeup of background by percent contribution, post-baseline selection (zero leptons, $H_{\mathrm{T}} \geq 700\,\mathrm{GeV}$, $N_{\mathrm{j}} \geq 9$, $N_{\mathrm{b}} \geq 3$)

The estimate for N_D, $\hat{N}_D$ can be expressed by the following relation between control regions : $D_{pred} = (\frac{B \cdot C}{A}) \cdot (\frac{C \cdot X}{A \cdot Y})$. In other words, it can be thought of as the usual ABCD method proportionality term multiplied by a higher-order term. This method is referred to as the "extended" ABCD method, and is used in this analysis to predict the yields of data-driven backgrounds.

To verify that this works, a closure test is performed using simulated event yields. Simulated 2016 $\mathrm{t\bar{t}}$ events requiring zero loose leptons and $H_{\mathrm{T}} > 700\,\mathrm{GeV}$ are used. In other words, using $\mathrm{t\bar{t}}$ yields in CRs, the $\mathrm{t\bar{t}}$ yield in region D is predicted and compared to the true $\mathrm{t\bar{t}}$ yield in D. The signal region is taken to be $N_{\mathrm{j}} \geq 9$ and $N_{\mathrm{b}} \geq 3$. CRs are defined by $N_{\mathrm{j}} = 7, 8, \geq 9$, and $N_{\mathrm{b}} = 2, \geq 3$, and are orthogonal to the SR. Table 10.1 shows that the predicted yield agrees within 7% of the actual simulated event yield, whereas the prediction from the standard ABCD method is off by 18%. This example justifies the use of the extended method over the "normal" ABCD relation.

In practice, in order to predict the yield of the QCD and hadronic $\mathrm{t\bar{t}}$ backgrounds using this method, the number of data events in each CR minus the number of simulated

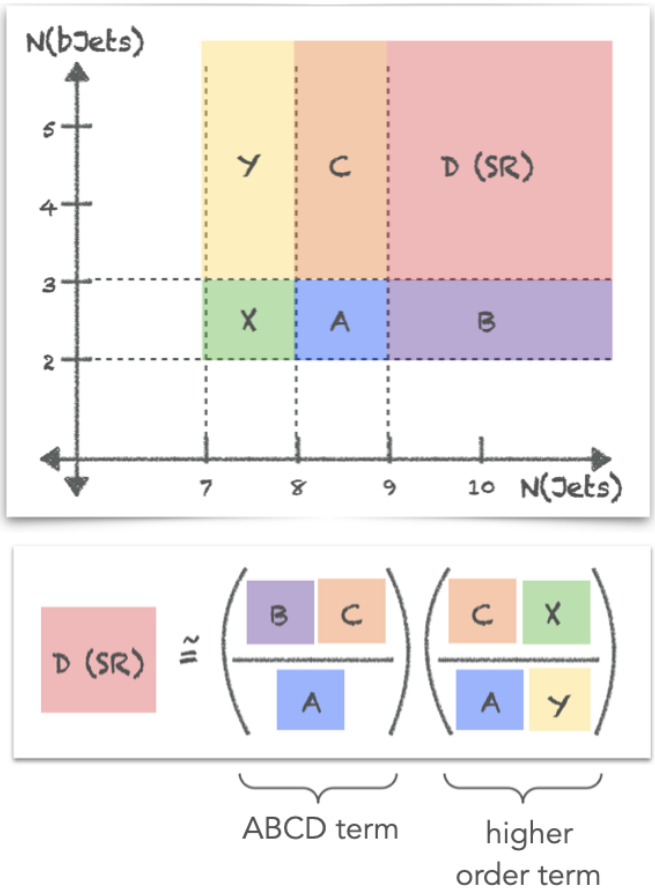

Figure 10.3: The control regions (A, B, and C) for the ABCD method and the added adjacent control regions (X and Y) for the extended ABCD method. For our method, the x axis corresponds to N_{j}, and the y axis to N_{b}.

Table 10.1: Event yields predicted by the extended ABCD method vs. traditional ABCD method and true event yields in 2016 simulation with baseline selection applied (0 leptons, $N_{\mathrm{j}} \geq 9$, $N_{\mathrm{b}} \geq 3$, and $H_{\mathrm{T}} > 700\,\mathrm{GeV}$). Only statistical uncertainties are shown. Table by Hayoung Oh.

Extended ABCD	ABCD	True
2042 ± 31.52	1684 ± 17.88	2202 ± 21.17

minor background events are propagated through the extABCD formula in order to predict the number of background events in the SR.

10.2 The ABCDnn method

The extended ABCD method proportionality relation discussed in section 10.1 can predict absolute rates of data-driven backgrounds, but does not address the shape of these backgrounds. Thus, another technique, the "ABCDnn" is introduced [68]. This uses the same five CRs defined in Section 10.1, but in this case uses the full DD background BDT discriminant distributions (data - minor MC contributions) rather than the event yields within those CRs. A deep neural network (DNN) is trained to transform input distributions (in this case, simulated $t\bar{t}$ distributions) to match these DD BDT distributions in each consecutive CR. The transformation is applied to an input distribution and morphs it to match the BDT distribution by translating, stretching, or squeezing of the phase space for the condition defined with control variables. After training to morph these $t\bar{t}$ input BDT distributions to match the DD background shape in the CR, the NN then predicts the shape of the DD background in the SR. These predicted distributions are then normalized to the yields derived in Section 10.1. Technically speaking, the NN used is a NAF [46], and transformations between the distributions of feature variables under different conditions are constructed as an invertible bijective function that is learned during training. The training proceeds in the direction of minimizing the maximum-mean-discrepancy [69] between the distributions of source (the input simulated $t\bar{t}$ distributions) and the target (the DD background BDT predictions). After training, the learned transformation between the source and the target is applied to the distribution of the $t\bar{t}$ simulation in the SR, which is subsequently morphed to predict the shape of the $t\bar{t}$ +QCD multijet DD BDT. Figure 10.4 represents the schematics of this method. This technique is powerful in that it can take into account complex correlations between feature variables in order to predict complicated multi-dimensional distributions (like a BDT discriminant).

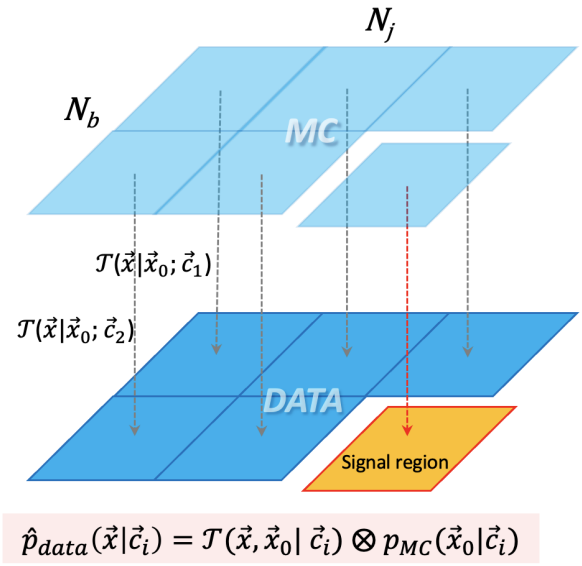

Figure 10.4: Method schematic for ABCDnn used in this study. Both simulation and real data are divided into 5 CRs and SR with control variables of jet multiplicity and b jet multiplicity. The transformation between data and simulation in the same condition, $\vec{c}$, is learned during training. The simulated and data events in the SR are not used in the training. Instead, the transformation under the SR condition is obtained from the training and applied to the distribution of simulation in the SR, which is morphed to predict the distribution of data in the SR. Figure by Suyong Choi.

A closure test was performed prior to applying this approach in data by morphing input simulated $t\bar{t}$ BDT distributions to predict $t\bar{t}X$ distributions and comparing these distributions to "true" simulated $t\bar{t}X$ distributions. Here simulated $t\bar{t}X$ and $t\bar{t}$ events for 2018 were used with a selection of of 0 loose leptons, $H_T > 700$ GeV and 1 or more resolved tops. The signal region was defined as having $N_j \geq 9$ and $N_b \geq 3$. For the ABCDnn training, the control regions are defined to exclude the SR with $N_j = 7, 8, or \geq 9$ and $N_b = 2, or \geq 3$. To predict the shape of the BDT discriminant in SR bins, the simulated $t\bar{t}X$ and $t\bar{t}$ events are separated into N_{RT} and N_{BT} categories, and the ABCDnn is applied to each. The H_T and BDT discriminant shapes of $t\bar{t}X$ and simulated $t\bar{t}$ in the CRs are used as inputs, and their distributions in the SR for $t\bar{t}X$ are predicted. The predicted H_T distribution is used to split the predicted BDT shape into bins corresponding to the

SR H_{T} binning in each top-tag category. Figure 10.5 shows the distributions of the BDT discriminants for $\mathrm{t\bar{t}}$ events before and after morphing to predict the distribution of $\mathrm{t\bar{t}X}$ events using ABCDnn, as well as the actual distributions of $\mathrm{t\bar{t}X}$ events, in the SR bins. Although the BDT distributions of the source and target are fairly different, the method is able to predict the target shape relatively well.

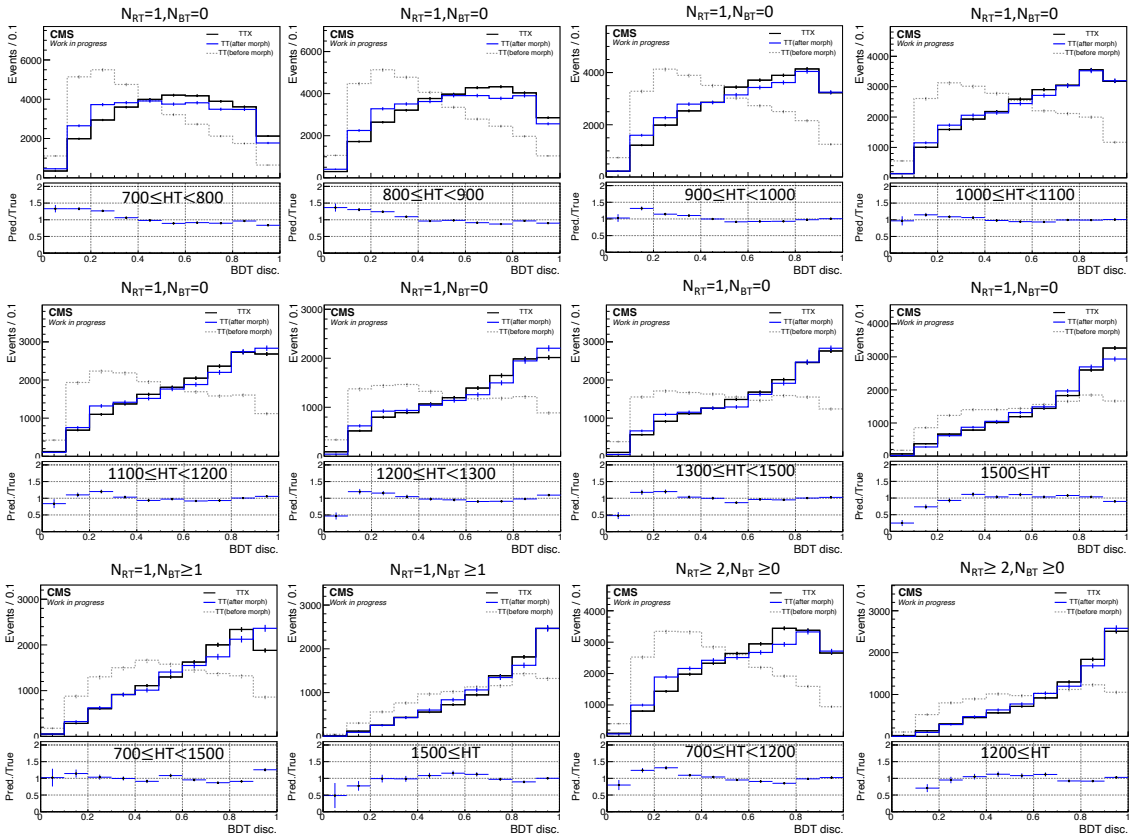

Figure 10.5: BDT discriminant shape for $\mathrm{t\bar{t}}$ events before morphing (gray), after morphing (blue) and for $\mathrm{t\bar{t}X}$ events (black) in the SR bins. The $\mathrm{t\bar{t}}$ BDT discriminant shapes before and after morphing are normalized to the yields from $\mathrm{t\bar{t}X}$. The lower panel shows the ratio of the predictions to their true values. The $\mathrm{t\bar{t}}$ distributions are morphed to predict the shape of the $\mathrm{t\bar{t}X}$ distribution using ABCDnn. The error bars indicate the statistical uncertainty. Plots by Hayoung Oh.

10.3 Validation of the background estimation methods

The background estimation strategy is validated in a validation region (VR) defined with the requirement of $N_{\mathrm{j}} = 8, N_{\mathrm{b}} \geq 3$ (10.6). The corresponding CRs used to estimate the $\mathrm{t\bar{t}}$ and QCD multijet background normalization and BDT shapes are defined by N_{j} and N_{b} for $N_{\mathrm{j}} = 6, 7, 8$, and $N_{\mathrm{b}} = 2, \geq 3$, but excluding the $N_{\mathrm{j}} = 8, N_{\mathrm{b}} \geq 3$ VR, which becomes the region "D" in the VR case. Figures 10.7-10.9 show predicted BDT discriminant distributions vs. data in N_{RT}, N_{BT}, and H_{T} VR categories corresponding to those defined for the SR (Table 9.1). The shapes of the $\mathrm{t\bar{t}}$ and QCD multijet background BDT distributions are predicted using the ABCDnn technique and normalized to the predicted yields obtained by applying the extended ABCD formula in each category. Other processes are estimated from simulation.

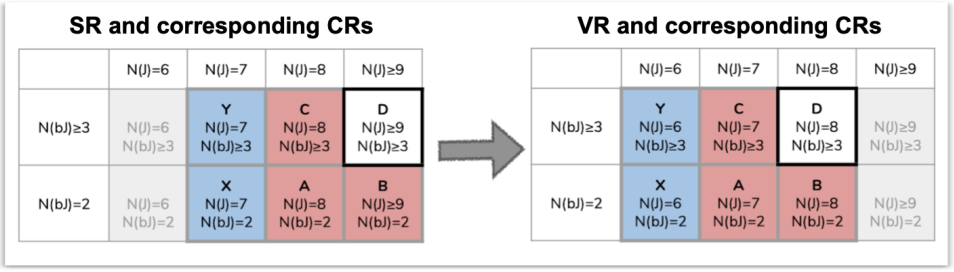

Figure 10.6: Schematic of VR vs. SR selections.

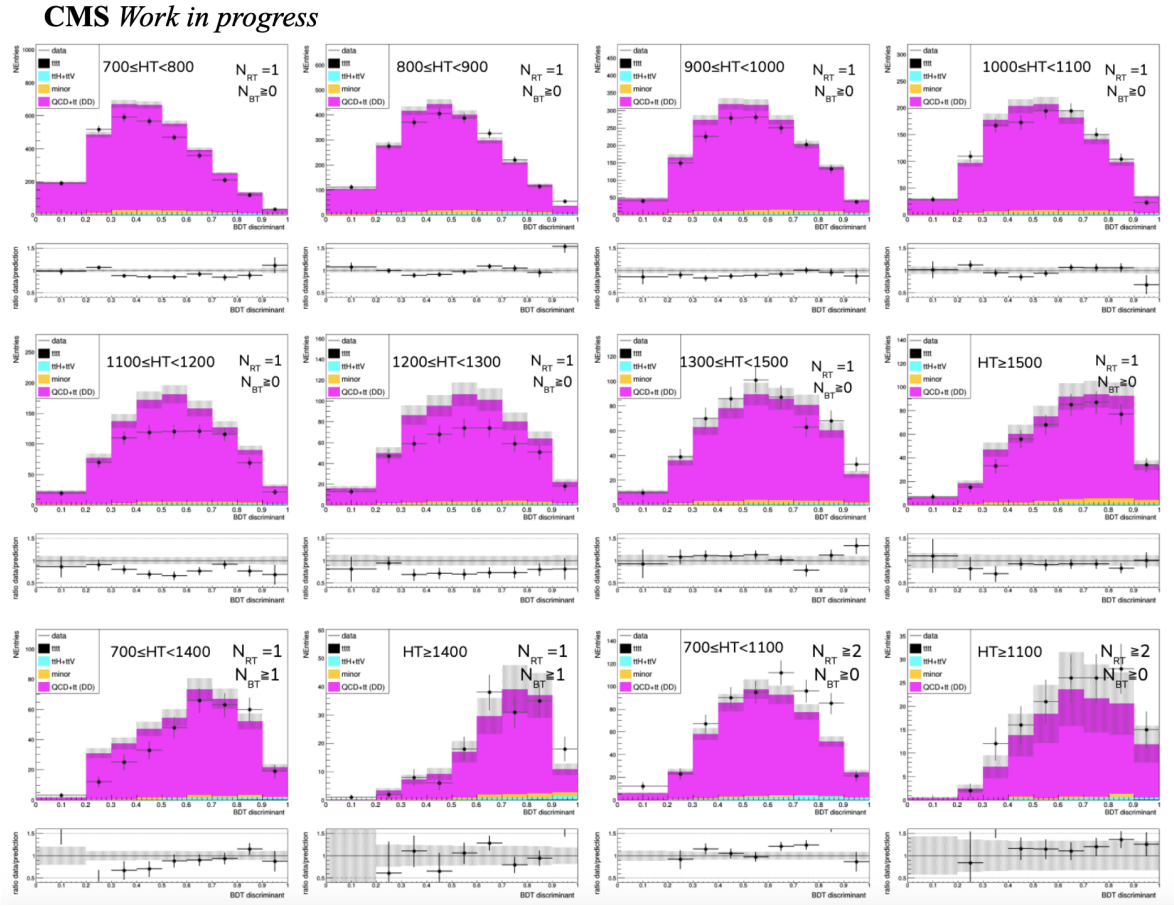

Figure 10.7: 2016 BDT shape predictions vs data corresponding to the SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the 8-jet validation region. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include only statistical uncertainties.

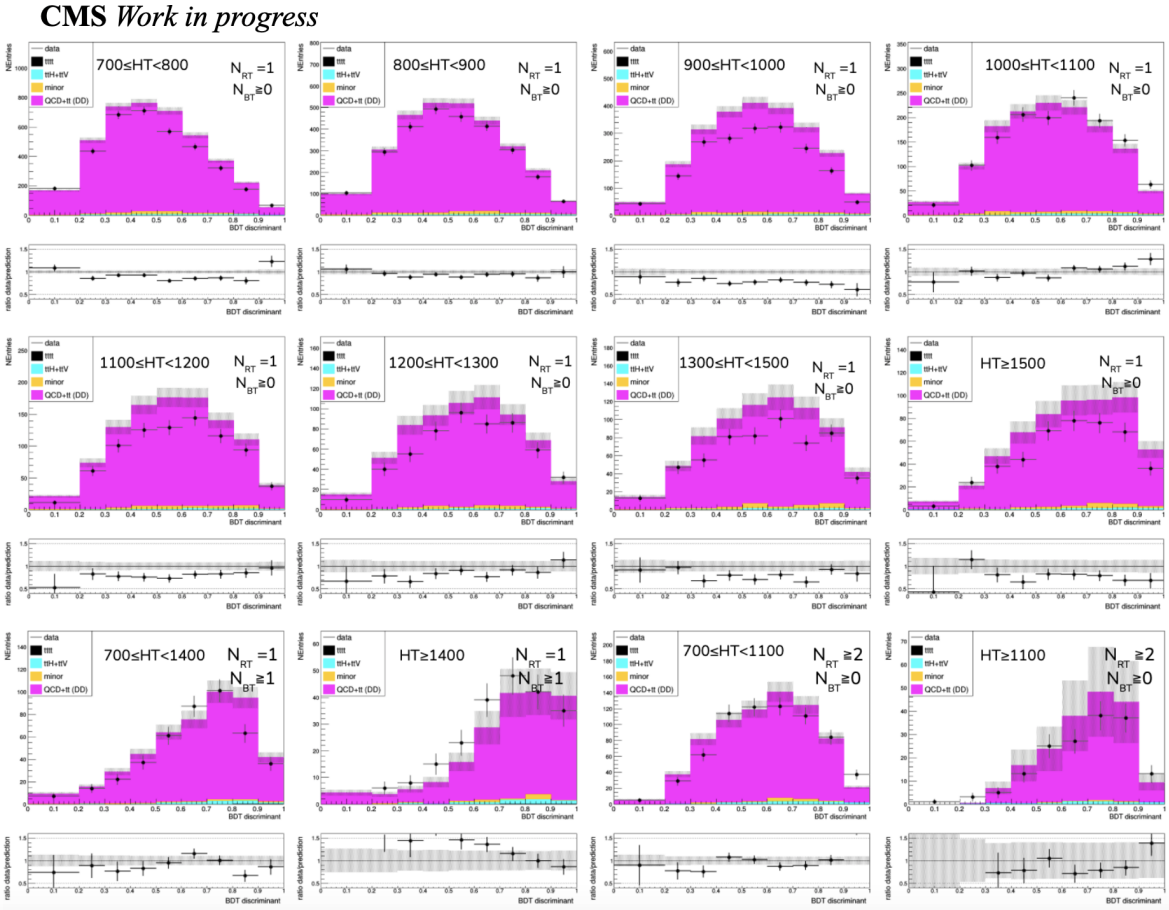

Figure 10.8: 2017 BDT shape predictions vs data corresponding to the SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the 8-jet validation region. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include only statistical uncertainties.

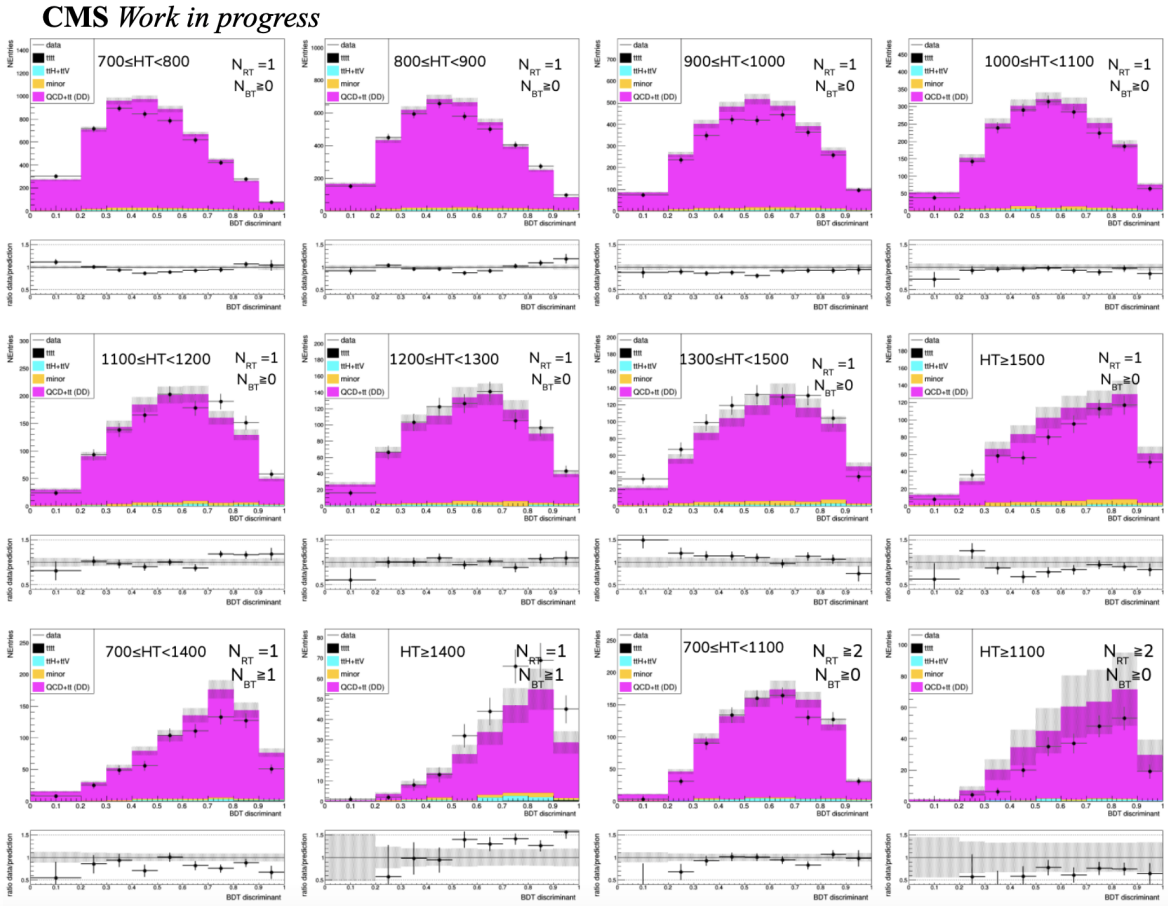

Figure 10.9: 2018 BDT shape predictions vs data corresponding to the SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the 8-jet validation region. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include only statistical uncertainties.

Chapter 11

Systematic Uncertainties

As discussed in Chapter 4, the overall goal of this analysis is to search for four-top production using a likelihood analysis that is a function of the measured data as well as nuisance parameters. These nuisances come from systematic uncertainties derived from theoretical, experimental, or statistical sources that affect individual processes, SR categories, or BDT distribution histogram bins differently and so must be treated accordingly. This chapter introduces relevant systematic uncertainties in this analysis, and discusses their treatment and importance.

11.1 Uncertainties of Data-Driven Backgrounds

As described in the previous section, $t\bar{t}$ and QCD multijet backrounds are data-driven, with the number of background events predicted by the extended ABCD formula, and the shape of BDT distributions predicted using the ABCDnn. This background estimation strategy is therefore subject to the following uncertainties:

- The statistical uncertainties of the estimated QCD multijet + $t\bar{t}$ yields in each CR, propagated through the extended ABCD formula.

- The statistical uncertainty of the simulated $\mathrm{t\bar{t}}$ events transformed to form each predicted BDT discriminant distribution.

- The statistical uncertainties from the trigger efficiency corrections, parameterized in N_{b} and N_{j}.

- A normalization uncertainty derived from the 8-jet validation region in order to account for any disagreement between the BDT prediction and data in each SR category. This is based on the level of disagreement observed between the data and predicted background in the VRs (Fig. 10.7-10.9). This uncertainty is calculated in each VR category as the sum in quadrature of two quantities: the deviation of the weighted mean (average) of events in that VR category from 1, and the weighted RMS of events in that VR category. The weighted mean $\langle f \rangle$ is defined as $\langle f \rangle = (\sum_i f_i w_i)/(\sum_i w_i)$, for each discriminant histogram bin i in the VR top-tag and H_{T} category of interest, where the weight w_i is the number of events in that histogram bin and f_i is the ratio of observed events to predicted events ($N_{\mathrm{data}}/N_{\mathrm{pred}}$) in that histogram bin. Thus, the mean is weighted to reflect the distribution of events in the histogram, and the deviation of this weighted mean from 1 ($1 - \langle f \rangle$) reflects the overall offset in normalization between the prediction and the data. Likewise, the weighted RMS is defined as $\sqrt{\langle f^2 \rangle - \langle f \rangle^2}$, and reflects the spread of the disagreement between the prediction and the data in the histogram bins. This method of quantifying the data-prediction discrepancy is based on recommendations from the CMS Statistics Committee. These uncertainties are derived from the VR region and applied to the SR.

- A shape uncertainty derived from the 8-jet validation region in order to account for any disagreement between the BDT prediction and data in each SR category based on the level of disagreement observed between the data and predicted background

in the VRs (Fig. 10.7-10.9). This uncertainty is calculated in each VR category as the percent shift in BDT discriminant values (up and down, for example for a variation of 2 percent this would correspond to BDT*1.02 for an up variation and BDT*0.98 for a down variation) required to account for any discrepancies between the data and prediction such that all data points agree with uncertainties in the VR in high BDT regions. These percent shifts were found to range between 1 and 3 percent, and a minimum shift of 1 percent is always applied. Once calculated in each VR category, the determined percent shift in BDT shape is applied to the corresponding SR category. These are taken to be correlated across SR categories. Additional tests and results of uncorrelating these uncertainties can be found in Appendix A.

11.2 Uncertainties of Simulated Processes

Monte Carlo simulated samples are used for the $t\bar{t}t\bar{t}$ signal predictions and for other minor backgrounds and so are subject to various systematic uncertainties related to the modeling of the experiment, particle kinematics, and theoretical models. The relevant sources of uncertainty affecting these processes are discussed below.

- Statistical uncertainties of the simulated samples.
- Uncertainties related to the boosted top and W tagging correction factors used to correct the performance of the DeepJet algorithm in simulation to match data.Two uncertainties are included: boosted top and boosted W uncertainties related to the non-mass-decorrelated scale factors. These uncertainties are correlated (simultaneous variations) between processes but uncorrelated between years [58].
- Tagging and misidentification uncertainties related to resolved top tagging correc-

tion factors used to correct the performance of the resolved top tagging algorithm in simulation to match data, as described in Chapter 7. Two uncertainties for both tagging efficiency and mistagging rate scale factors are considered, and uncertainties in these stem from statistical uncertainties that arise in calculating the scale factors. These are correlated between processes and uncorrelated between years.

- Uncertainties related to correction factors used to correct for the performance of the DeepJet b-tagging algorithm in simulation with respect to data. The DeepJet iterative fit/shape scale factors are used. Scale factors per event are calculated as the product of the scale factors for all jets in a given event passing pre-selection. The uncertainties are split into several sources, including those affecting HF (heavy-flavor) or LF (light-flavor) jets, uncertainties from charm jets (cferr1 and cferr2), and linear and quadratic statistical fluctuations (LFstats1, LFstats2, HFstats1, HFstats2). These are all considered as separate systematic variations, with statistical variations correlated between processes and uncorrelated between years (LFstats1, LFstats2, HFstats1, HFstats2) and others correlated between both processes and years (HF, LF, cferr1 and cferr2). Combined jet energy scale (JES) variations in the btagging scale factors are used when considering jet energy scale systematic uncertainties as described below [7].

- Statistical uncertainties in trigger efficiencies calculated in data and applied to simulation.

- Uncertainties associated to the pileup reweighting correction factor.

- Jet energy scale (JES) and resolution (JER) uncertainties impacting the reconstruction of jets, including $R = 0.4$ and $R = 0.8$ jets, $p_{\mathrm{T}}^{\mathrm{miss}}$, and tagged top, W, and b candidates and their correction factors. The JES systematic uncertainty is

currently applied using the combined JES uncertainty, which is the sum of all 20+ JES uncertainty sources added in quadrature. The JES and JER are varied up and down by their systematic uncertainties and propagate the effects through all analysis objects and selections. For example, for the "jerUp" variation the $R = 0.4$ and $R = 0.8$ collections obtained with the +1 standard deviation variation of the JER are used, and all relevant quantities using those variations are recalculated, including selecting jets and boosted objects and re-propagating the varied jets through the resolved top algorithm. For JES variations, the b-tagging scale factors with the jesUp/jesDown are also varied accordingly. JER and JES variations are correlated between processes but not between years.

- Uncertainties related to lepton selection correction factors applied to simulation to account for differences in lepton veto efficiencies between data and simulation.

- Uncertainties assigned to the integrated luminosity measured by the CMS experiment for the 2016 [70], 2017 [71] and 2018 [72] data-taking periods that affect simulation-based predictions.

- Theoretical uncertainties related to changes in renormalization (μ_R) and factorization (μ_F) scales, PDF, and strong coupling strength (α_S) predictions affect all simulated samples, but are most relevant for $t\bar{t}t\bar{t}$ signal predictions. The effect of μ_R and μ_F on predicted event counts are estimated by varying μ_R and μ_F up and down by a factor of 2 independently (i.e. just μ_R or just μ_F) and simultaneously (i.e. both μ_R and μ_F) and assigning an uncertainty envelope based on the upper and lower bounds of these uncertainty variations [73]. In calculating these variations, variations are first renormalized to the nominal sample. The uncertainty related to PDF choice and α_S is obtained from variations of the pdf sets stored in each NanoAOD sample, as recommended in [74]. These weights are renormalized to

the nominal sample. For these, the standard deviation of the renormalized weight variations is taken as the size of the uncertainty. The value of α_S in parton showers also leads to uncertainties related to the intial-state (ISR) and final-state radiation (FSR) emissions in QCD events. These uncertainties are handled by varying the renormalization scale of QCD (therefore α_S) up and down by a factor of two in both the ISR and FSR scales, using the variations stored as PSWeights in dedicated NanoAOD samples. These are varied independently, (for example varying ISR up and down and holding FSR constant). Overall, four separate theoretical uncertainties are considered: pdf, μ_R/μ_F (also referred to as ME for matrix element), ISR, and FSR. ME, ISR, and FSR variations are varied separately (uncorrelated) for both $t\bar{t}X$ (when PSWeight samples are available) and $t\bar{t}t\bar{t}$ processes but correlated between years. PDF variations are correlated between years and processes. Top mass uncertainties are not expected to be a significant source of uncertainty and are not considered.

- All of 2017 and some of 2016 data were affected by the L1 ECAL prefiring issue, especially in jet-heavy analyses. The effect of the prefiring probability in the signal region for 2016 and 2017 was investigated and a correction weight and uncertainty to address this for $t\bar{t}t\bar{t}$ signal and $t\bar{t}X$ processes (X=W,Z,H) was assigned.

- Uncertainties in the theoretical cross sections used to normalize $t\bar{t}X$ backgrounds are considered. These impact the number of $t\bar{t}X$ events in each SR bin. For $t\bar{t}X$ processes (X=W,Z,H), an uncertainty of 26% is assigned based on the largest deviation of the signal strength observed by CMS for any of these processes (relative to the SM prediction), which corresponds to the observation of $t\bar{t}H$ [75] (albeit with large uncertainties).

A summary of these uncertainties and their correlations and effects can be found

in Table 11.1. The most significant systematics in this analysis are the VR-derived shape and normalization uncertainties and those related to statistical uncertainties of the dominant hadronic $t\bar{t}$ and QCD multijet data-driven backgrounds. These backgrounds are most responsible for limiting the overall sensitivity of this analysis.

Table 11.1: Summary of systematic uncertainties and the ranges of their effects on signal and background yields. The uncertainty type (shape vs normalization only), affected processes, correlations (between processes and/or years), and effects on signal and background yields as % are shown. Systematic uncertainties are considered for all years unless otherwise indicated.

Name	Type	Processes	Correlations	$t\bar{t}t\bar{t}$ Signal Uncertainty (%)	Background Uncertainty (%)
Statistics of data-driven backgrounds	shape	QCD+$t\bar{t}$	-	-	5-30
Statistics of transformed samples	shape	QCD+$t\bar{t}$	-	-	10
Statistics of simulated samples	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	-	0-20	0-20
Data-prediction normalization	lnN	QCD+$t\bar{t}$	-	-	7-37
Data-prediction shape	shape	QCD+$t\bar{t}$	-	-	n/a
DeepJet b-tag SF (HF, LF, cferr)	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes+years	0-10	0-10
DeepJet b-tag SF (stats)	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	0-10	0-10
Resolved top efficiency SF	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	0-5	0-5
Resolved top mistag SF	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	0-10	0-10
DeepJet boosted top SF	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	0-5	0-5
DeepJet boosted W SF	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	0-10	0-5
JER	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	0-20	0-20
JES	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	5-20	5-20
Pileup	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	0-5	0-5
Trigger efficiency	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	0-5	0-5
Lepton veto	lnN	$t\bar{t}t\bar{t}$, $t\bar{t}$X	processes	0-5	0-5
Luminosity	lnN	$t\bar{t}t\bar{t}$, $t\bar{t}$X, other	processes	2.3-2.5	2.3-2.5
PDF	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X	processes+years	0-10	0-10
ISR	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X	years	0-5	0-5
FSR	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X	years	0-20	0-20
μ_R, μ_F	shape	$t\bar{t}t\bar{t}$, $t\bar{t}$X	years	0-20	0-20
Cross section	lnN	$t\bar{t}$X	processes+years	-	26
Prefire	lnN	$t\bar{t}t\bar{t}$, $t\bar{t}$X (2016, 2017)	processes	< 1	< 1

Chapter 12

Results and interpretation

This chapter discusses the search for four-top signal in the all-hadronic $t\bar{t}t\bar{t}$ channel, given a SR separated into categories of resolved and boosted top multiplicity and H_{T}, and using event-level BDT discriminant distributions in order to perform a profile maximum likelihood fit. Four processes and their systematic uncertainties over the data taking period of 2016-2018 are considered: $t\bar{t}t\bar{t}$ signal, the data-driven dominant $t\bar{t}$ and QCD multijet backgrounds, the $t\bar{t}X$ background, and other minor backgrounds.

A binned likelihood analysis is carried out in the SR categories split by boosted and resolved top multiplicity and H_{T} as defined in Table 9.1. The systematic uncertainties described in Chapter 11 are incorporated as nuisance parameters. The BDT discriminant shapes for $t\bar{t}t\bar{t}$ signal, the $t\bar{t}$ and QCD multijet background predicted from data, and $t\bar{t}X$ and other minor backgrounds estimated from simulation are provided as inputs for each SR category. Up and down variations of the BDT discriminant shapes are likewise provided as inputs for each nuisance parameter corresponding to shape-based systematic variations (as listed in Table 11.1), while uncertainties affecting only the normalizations of processes are implemented using log normally distributed constraints on the simulation rates in each data-taking period. BDT discriminant shapes for the data-driven QCD and

$t\bar{t}$ background prediction, and other backgrounds are estimated from simulation in each SR category that are used as inputs for the combined fit are shown in Figs. 12.1-12.3.

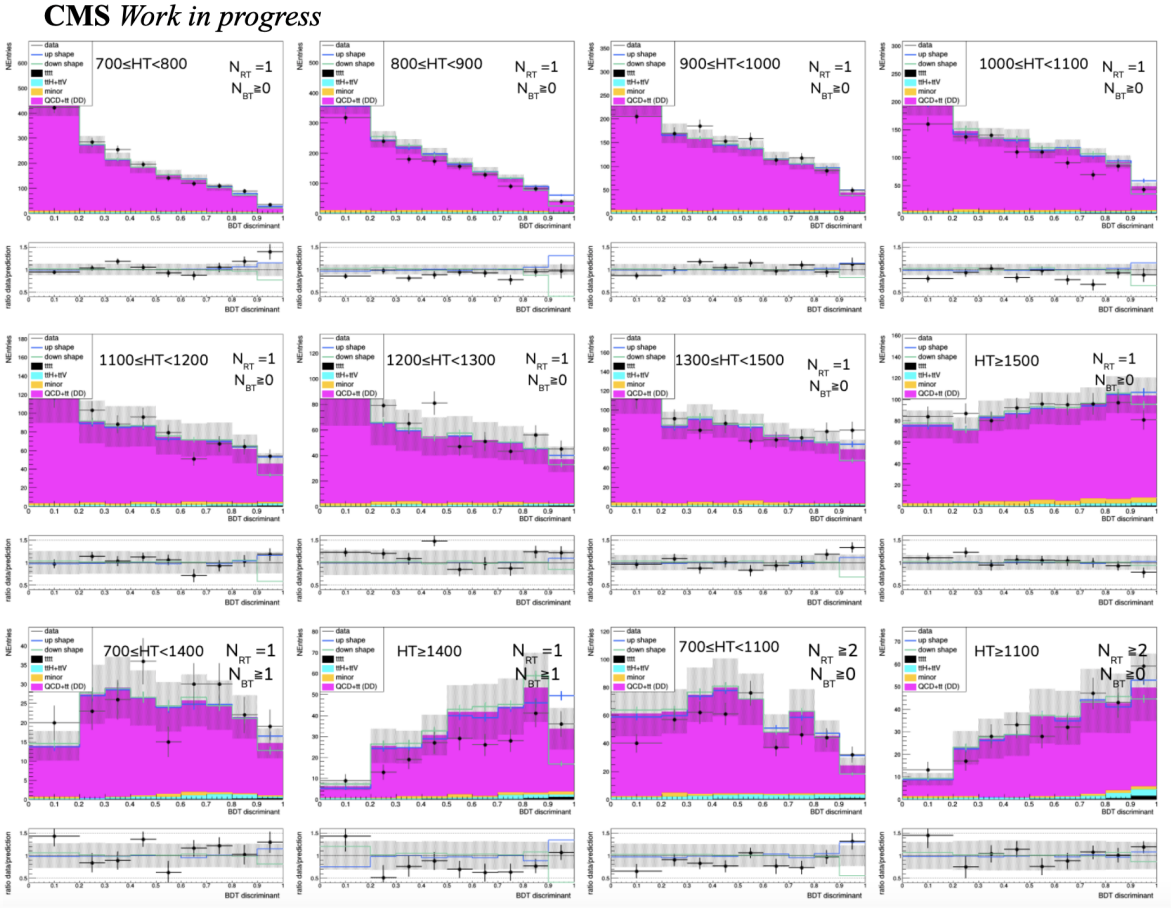

Figure 12.1: 2016 BDT prefit shape predictions in the 12 SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the signal region. The $t\bar{t}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $t\bar{t}t\bar{t}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include statistical uncertainties and uncertainties derived from the weighted mean and RMS to account for discrepancies between data and the data-driven background estimate observed in the VRs, as described in Section 11. Also shown are the up (blue) and down (green) shape uncertainty variations described in the VR, also described in Section 11.

After unblinding, the observed $t\bar{t}t\bar{t}$ signal significance relative to the background-only hypothesis and measured $t\bar{t}t\bar{t}$ cross section were calculated from a profile maximum-likelihood fit with the cross section taken as the parameter of interest. A 95% confidence

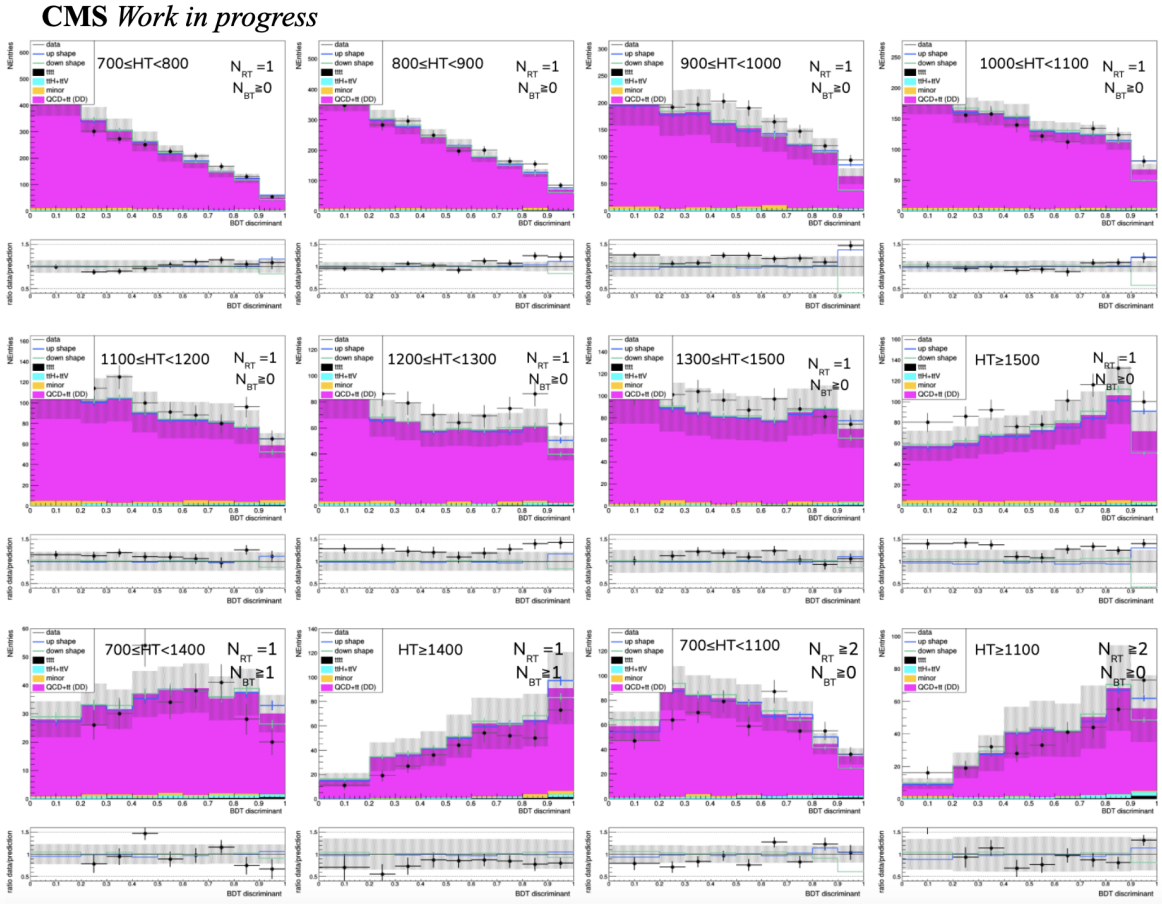

Figure 12.2: 2017 BDT prefit shape predictions shown in the 12 SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the signal region. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples.The error bands shown include statistical uncertainties and uncertainties derived from the weighted mean and RMS to account for discrepancies between data and the data-driven background estimate observed in the VRs, as described in Section 11. Also shown are the up (blue) and down (green) shape uncertainty variations described in the VR, also described in Section 11.

level (CL) upper limit will be set on the $\mathrm{t\bar{t}t\bar{t}}$ cross section using the profile likelihood ratio test statistic and asymptotic approximation and the modified frequentist approach [32, 76].

The signal strength (the ratio of the observed four-top production rate to the standard model expectation) was measured to be $\mu = 5.1^{+2.3}_{-2.0}$ (68% CL). The expected significance

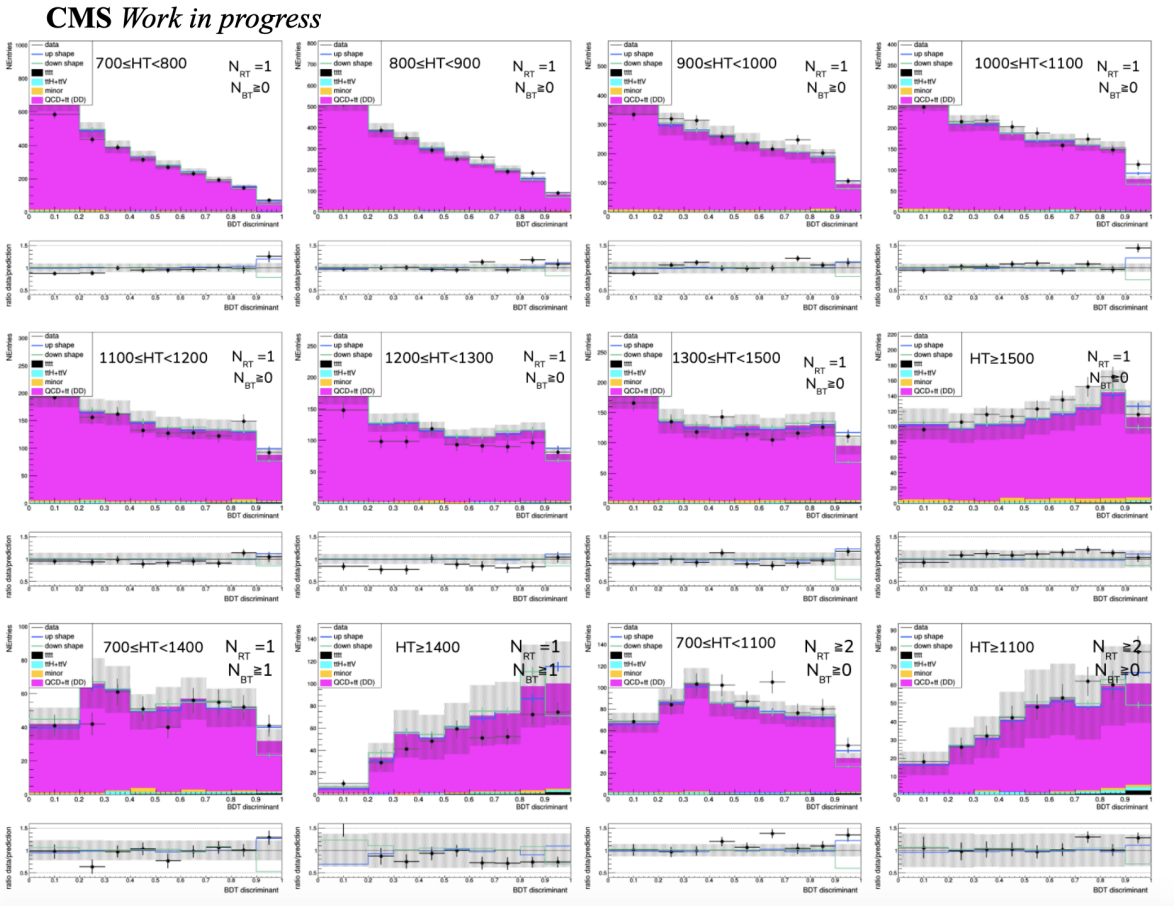

Figure 12.3: 2018 BDT prefit shape predictions shown in the 12 SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the signal region. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples.The error bands shown include statistical uncertainties and uncertainties derived from the weighted mean and RMS to account for discrepancies between data and the data-driven background estimate observed in the VRs, as described in Section 11. Also shown are the up (blue) and down (green) shape uncertainty variations described in the VR, also described in Section 11.

for $\mathrm{t\bar{t}t\bar{t}}$ production in the all-hadronic final state relative to the background-only hypothesis is 0.43 standard deviations. The expected 95% CL upper limit on the cross section from the all-hadronic channel alone is 4.88. The observed significance and 95% CL upper limit are 2.25 standard deviations and 8.39, respectively. This corresponds to a measured cross section of 60 ± 24 fb. Table 12.1 lists the expected and observed upper limits

Table 12.1: Expected and observed significances and limits times the standard model $t\bar{t}t\bar{t}$ cross section for each year and for 2016, 2017, and 2018 combined (full Run 2).

No fit to data

	2016 (36 fb^{-1})	2017 (42 fb^{-1})	2018 (60 fb^{-1})	Run II (137 fb^{-1})
Median expected limit:	8.91	8.97	8.31	4.88
Expected significance (s.d.)	0.25	0.23	0.26	0.43

With fit to data

	2016 (36 fb^{-1})	2017 (42 fb^{-1})	2018 (60 fb^{-1})	Run II (137 fb^{-1})
Median expected limit:	8.72	7.56	7.06	4.20
+1σ expected limit	13.2	11.1	10.4	6.05
+2σ expected limit	19.7	15.8	14.8	8.44
Expected significance (s.d.)	0.25	0.28	0.29	0.49
Observed limit:	12.9	14.0	10.2	8.39
Observed significance (s.d.)	1.10	1.76	1.15	2.25

"Limits" are on $\mu = \sigma/\sigma_{SM}$ **CMS** *Work in progress*

obtained for each data-taking period. Postfit shape distributions (signal+background) are shown in Figures 12.5- 12.7. Nuisance parameter impacts are shown in Fig. 12.4. Statistical uncertainties are included both as normalization uncertainties on the total rates in each category and bin-by-bin in the histograms of BDT distributions.

It can be noted that the fit to signal is larger than standard model expectations. While a larger than expected signal strength was observed by ATLAS in multiple channels [51], the most recent CMS Run II results (in the same-sign-dilepton channel) agreed with SM expectations [50]. A CMS Run II SM combined result, including the all-hadronic, single-lepton, and opposite-sign-dilepton channels combined with the already-published same-sign-dilepton result, is still underway as of the writing of this thesis. Future analyses are planned to explore potential BSM drivers of signal excesses.

The naming scheme for nuisance parameters in Figure 12.4 is as follows:

- RT#BT#htbin# - refers to the signal region bin. RT is for resolved tops and BT is for boosted tops, and together they refer to the 1 resolved top, 0 boosted tops (RT1BT0), 1 resolved top, 1 or more boosted tops (RT1BT1) and 2 resolved tops, 0

or more boosted tops (RT2BTALL) bins. htbin refers to ht bin splitting. Nuisance parameters with these labels refer to statistical uncertainties in the corresponding signal region bins.

- prop_binch#_bin# - these are the bin-by-bin statistical uncertainties from autoMCstats
- cross_section - uncertainty on theoretical cross sections used to normalize $t\bar{t}X$ backgrounds
- prefire - prefire uncertainty
- data-pred_norm_disagreement - normalization uncertainty to account for data to background prediction discrepancy in 8-jet validation region
- data-pred_shape_disagreement - shape uncertainty to account for data to background prediction discrepancy in 8-jet validation region
- lumi - luminosity uncertainty
- elveto - electron veto systematics
- muveto - muon veto systematics
- pileup - pileup systematics
- trigger - trigger efficiency systematics
- jer - jet energy resolution, combined
- jes - jet energy scale, combined
- isr - initial state radiation uncertainty (from PSWeight variations)

- fsr - final state radiation uncertainty (from PSWeight variations)
- ME - matrix element/ factorization and renormalization scale uncertainty (from LHEScaleWeight variations)
- pdf - pdf uncertainty (from LHEPdfWeight variations)
- btagHF - DeepJet btagging scale factor uncertianty (heavy flavour)
- btagLF - DeepJet btagging scale factor uncertianty (light flavour)
- btagHFstats1 - DeepJet btagging scale factor uncertianty (heavy flavour stats1)
- btagLFstats1 - DeepJet btagging scale factor uncertianty (light flavour stats1)
- btagHFstats2 - DeepJet btagging scale factor uncertianty (heavy flavour stats2)
- btagLFstats2 - DeepJet btagging scale factor uncertianty (light flavour stats2)
- btagCFerr1 - DeepJet btagging scale factor uncertianty (CFerror1 uncertainties from charm jets)
- btagCFerr2 - DeepJet btagging scale factor uncertianty (CFerror2 uncertainties from charm jets)
- DeepJetTopSF - DeepJet boosted top SF systematics (not mass decorrelated)
- DeepJetWSF - DeepJet boosted W SF systematics (not mass decorrelated)
- ResTopEff - custom NanoAOD resolved top tagger SF efficiency systematics
- ResTopMiss - custom NanoAOD resolved top tagger SF mistag systematics

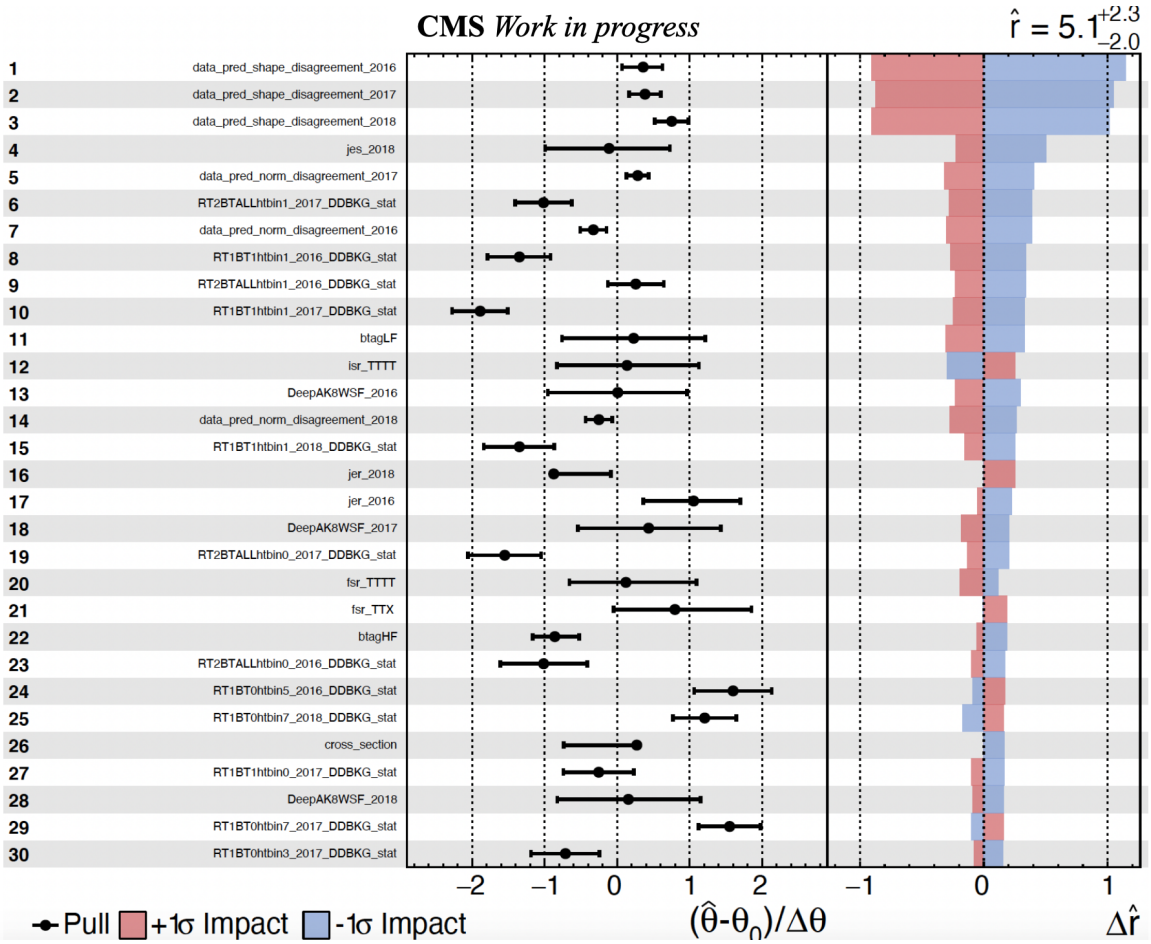

Figure 12.4: Impacts for first 30 nuisance parameters. Note that the VR-derived data-driven background uncertainties are the most important.

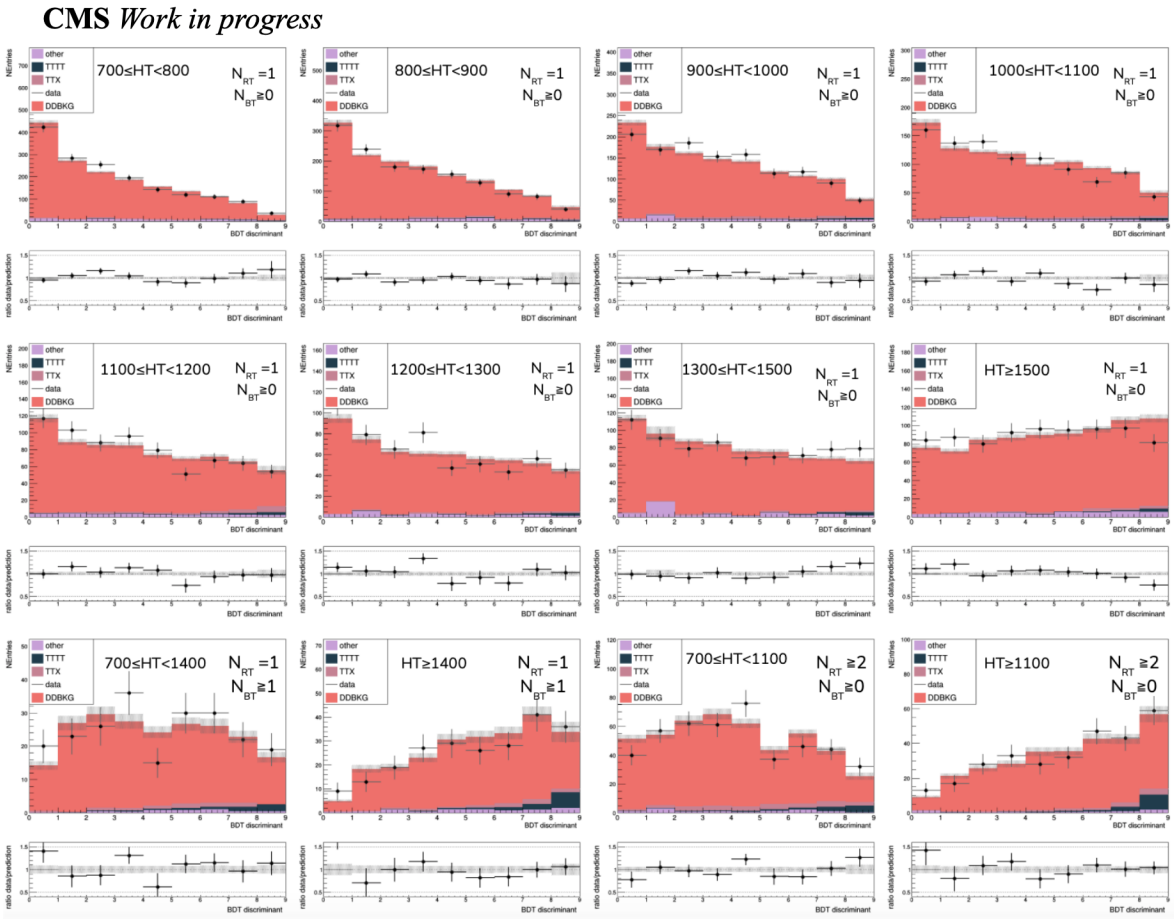

Figure 12.5: 2016 BDT postfit shape predictions for signal+background shown in the 12 SR bins defined by N_{RT}, N_{BT}, and H_{T}.

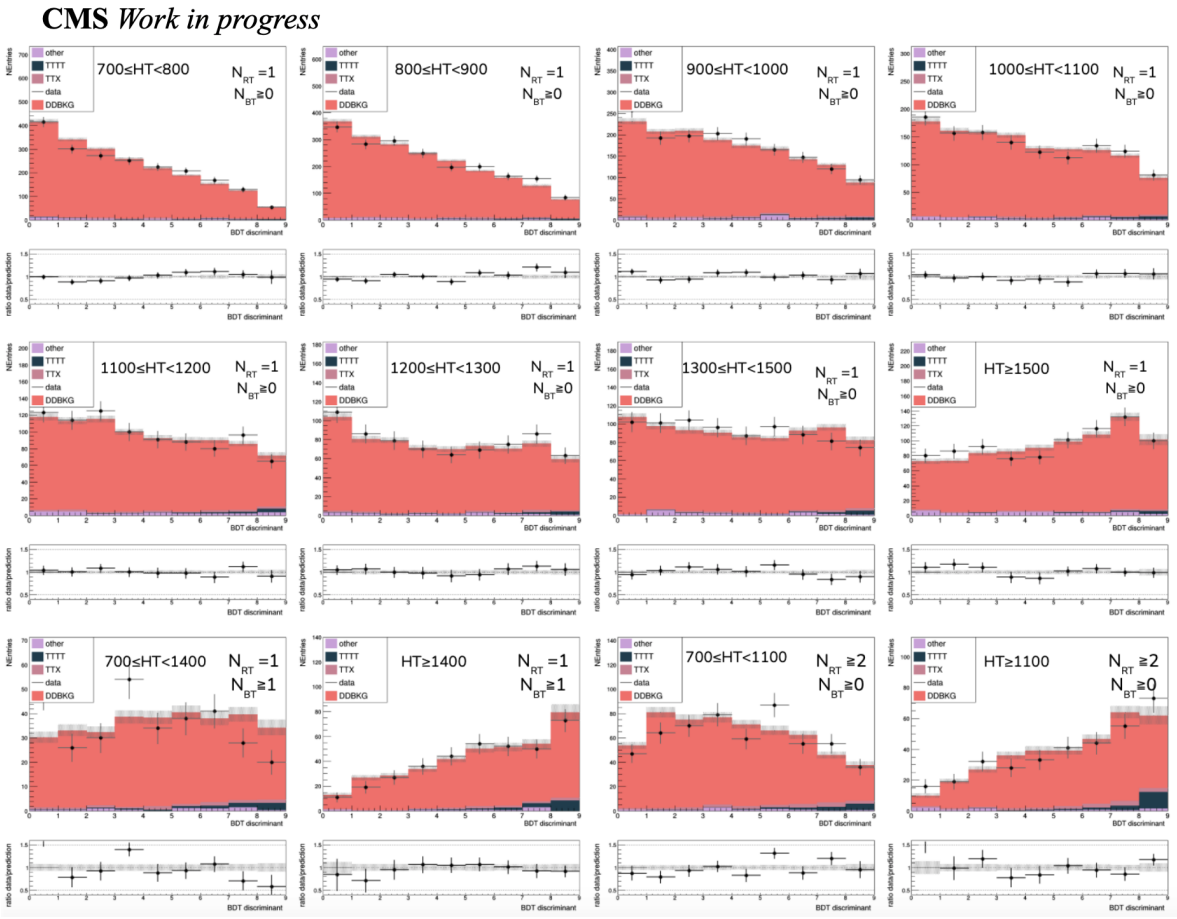

Figure 12.6: 2017 BDT postfit shape predictions for signal+background shown in the 12 SR bins defined by N_{RT}, N_{BT}, and H_{T}.

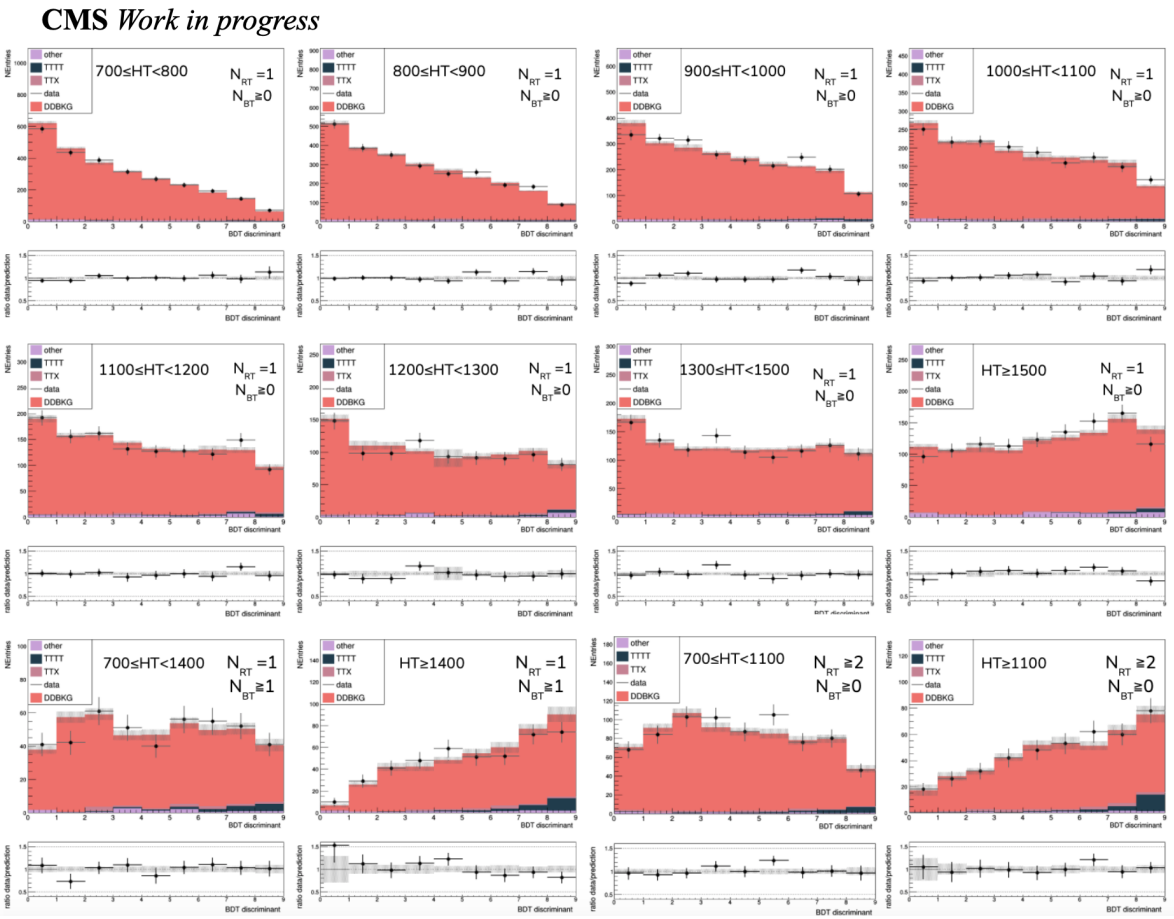

Figure 12.7: 2018 BDT postfit shape predictions for signal+background shown in the 12 SR bins defined by N_{RT}, N_{BT}, and H_{T}.

Chapter 13

Summary, Conclusions and Future Results

This thesis gave an overview of an first all-hadronic four-top analysis using Run II data with the CMS Experiment at the LHC. This analysis is significant as the first analysis of its kind, in its use of novel machine learning tools such as the NN-based data-driven background estimation and NanoAOD-based resolved top tagger, and as part of a larger combination of four-top analyses in multiple final states with CMS that expect to see evidence of four-top production corresponding to a significance of 3σ in a published result expected in 2022. This combination is still going through the review process as of the writing of this thesis. The all-hadronic analysis presented here observed a signal strength larger than SM expectations, of about $\mu = 5.1^{+2.3}_{-2.0}$. The observed (expected) significance and limit times the SM cross section were approximately 2.25 (0.43) σ and 8.39 (4.88) respectively.

Given that an excess of four-top production compared to the SM could provide hints of new BSM, for example evidence of heavy scalar or pseudo-scalar higgs-like bosons decaying to tops as predicted by many 2HDM and corresponding supersymmetric theories,

this result as well as the upcoming combined result will likely motivate further BSM searches for new physics. Future BSM interpretations in the all-hadronic final state and other final states are planned. In addition, future four-top analyses after more data-taking in CMS and ATLAS could ultimately lead to the discovery of four-top production at an observed significance greater than 5σ. With more statistics, future four-top analyses may also be sensitive to a precise measurement of the top-Higgs Yukawa coupling compared to SM expectations. In the meantime the tools and techniques developed by this analysis can be further developed and applied to other hadronic analyses to provide improved sensitivity and insight.

Appendix A

VR Data-Prediction Shape Based Uncertainties

This appendix includes various tests regarding the inclusion of shape-based uncertainties address data vs. prediction disagreements derived in the VR. As a reminder this is a horizontal % shift in BDT discriminant value (value = value+value*% for the up variation, for example), and shape variations are renormalized. The nominal version is a shape variation correlated between SR categories within a single year and with a % shift determined in the equivalent VR category such that, after applying the normalization uncertainty derived in the VR, any remaining discrepancies in high BDT score bins are accounted for. Distributions showing the % shift chosen and the corresponding uncertainty bands can be found in Figs. A.2–A.4.

	2016		2017		2018	
	VRnorm (%)	VRshape (%)	VRnorm (%)	VRshape (%)	VRnorm (%)	VRshape (%)
RT1BT0htbin0	12.4	1	13.8	1	8.9	1
RT1BT0htbin1	10.1	3	8.7	1	7.9	1
RT1BT0htbin2	11.2	1	23.1	3	12.1	1
RT1BT0htbin3	10.5	2	10.9	2	6.9	1
RT1BT0htbin4	24.9	2	20.7	1	11.8	1
RT1BT0htbin5	26.6	1	19.6	1	7.1	1
RT1BT0htbin6	14.5	2	24.2	1	13.0	2
RT1BT0htbin7	13.5	1	25.0	2	18.0	1
RT1BT1htbin0	24.6	1	18.8	1	20.3	2
RT1BT1htbin1	26.4	3	30.1	1	36.5	3
RT2htbin0	25.4	2	17.9	3	11.1	1
RT2htbin1	27.3	1	35.5	1	35.1	2

Figure A.1: Summary of the VRnorm and VRshape uncertainties applied in each SR category and year as derived in the VR. The VRnorm is derived using the weighted mean and RMS of BDT histograms in the validation region, and is reported as a percent error. The VRshape is derived as a horizonal percent shift in per-event BDT values.

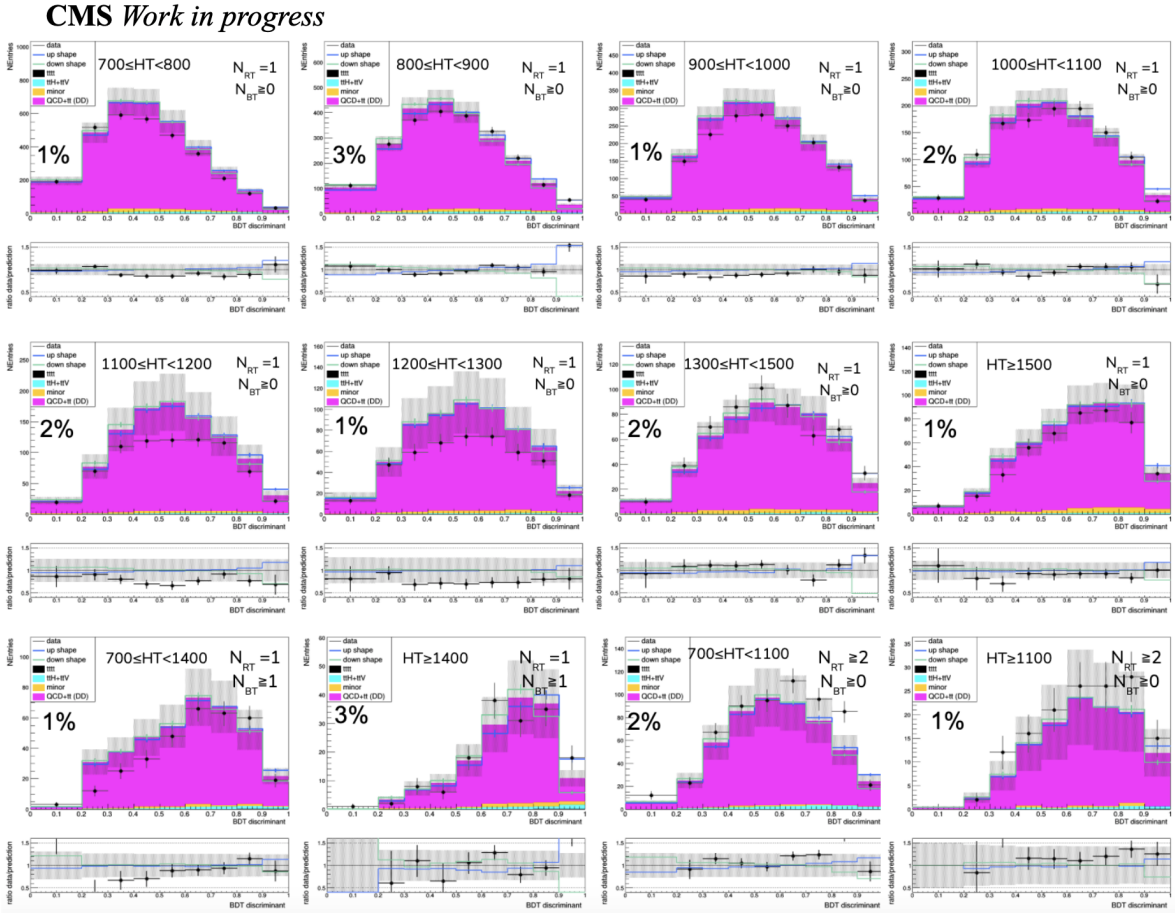

Figure A.2: 2016 BDT shape predictions vs data corresponding to the SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the 8-jet validation region. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include statistical uncertainties as well as data-pred_norm_disagreement and data-pred_shape_disagreement uncertainties to account for discrepancies between data and the data-driven background estimate observed in the VRs as described in Section 11 can be found in A.

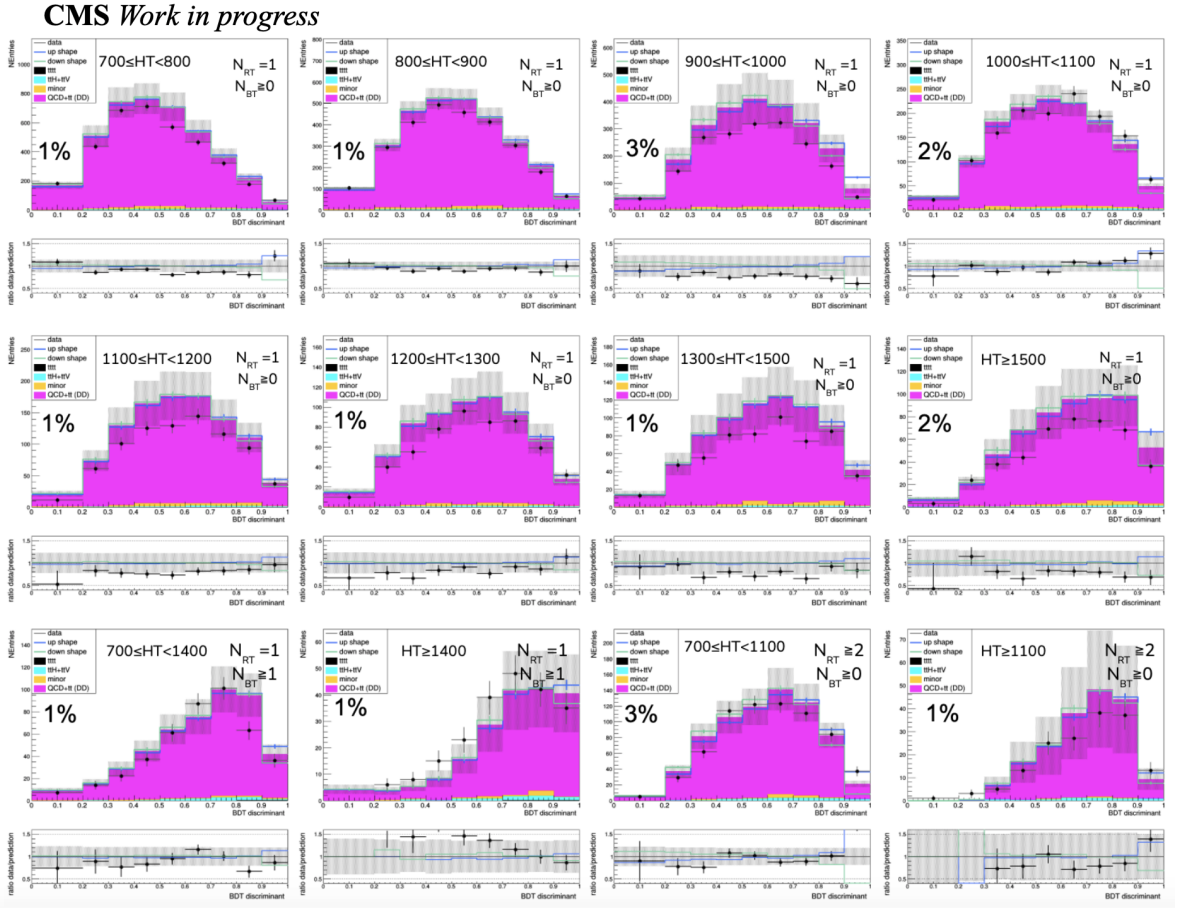

Figure A.3: 2017 BDT shape predictions vs data corresponding to the SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the 8-jet validation region. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include statistical uncertainties as well as data-pred_norm_disagreement and data-pred_shape_disagreement uncertainties to account for discrepancies between data and the data-driven background estimate observed in the VRs as described in Section 11 can be found in A.

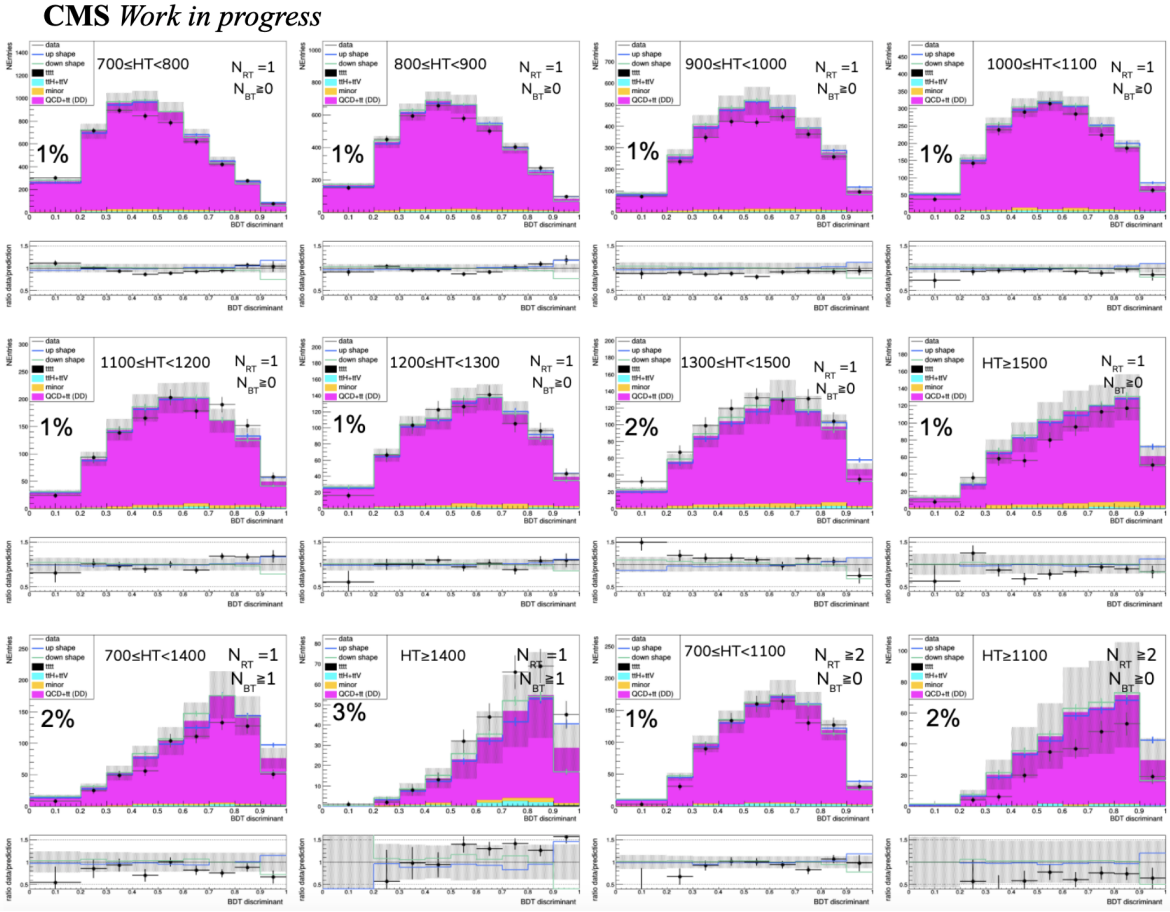

Figure A.4: 2018 BDT shape predictions vs data corresponding to the SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the 8-jet validation region. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include statistical uncertainties as well as data-pred_norm_disagreement and data-pred_shape_disagreement uncertainties to account for discrepancies between data and the data-driven background estimate observed in the VRs as described in Section 11 can be found in A.

No fit to data							
	optimized shape	**1% shape**	**2% shape**	**2.5% shape**	**3% shape**	**3.5% shape**	**4% shape**
Median expected limit: Expected significance (s.d.)	8.97 0.23	9.38 0.21	11.2 0.18	11.4 0.18	11.7 0.18	12.0 0.17	12.2 0.17

With fit to data							
	optimized shape	**1% shape**	**2% shape**	**2.5% shape**	**3% shape**	**3.5% shape**	**4% shape**
Observed limit: Median expected limit: +2σ expected limit Observed significance (s.d.)	14.3 9.16 18.4 1.25	16.6 9.25 18.5 1.71	13.8 10.5 21.5 0.48	13.8 10.7 21.8 0.44	14.2 10.9 22.1 0.43	13.9 11.1 22.6 0.28	13.0 11.1 22.8 0.0

"Limits" are on $\mu=\sigma/\sigma_{SM}$

Figure A.5: Results including various schemes of including correlated shape uncertainties, including the % variation needed to account for discrepancies in the VR, as well as blanket 1%, 2%, 3%, and 4% variations are shown. Observed and expected limits and significances for full Run II data in the less sensitive RT1BT0 categories are shown. The normalization uncertainty addressing data vs. prediction disagreement is included in all cases.

No fit to data				
	correlated norm only	**uncorrelated norm only**	**correlated norm+shape**	**uncorrelated norm+shape**
Median expected limit: Expected significance (s.d.)	5.08 0.40	5.05 0.40	8.97 0.23	7.22 0.28

With fit to data				
	correlated norm only	**uncorrelated norm only**	**correlated norm+shape**	**uncorrelated norm+shape**
Observed limit: Median expected limit: +2σ expected limit Observed significance (s.d.)	19.5 5.16 10.3 5.08	19.6 5.14 10.3 5.10	14.3 9.16 18.4 1.25	17.6 7.28 14.4 2.79

"Limits" are on $\mu=\sigma/\sigma_{SM}$

Figure A.6: Results including no VR shape uncertainty, correlated VR shape uncertainty and uncorrelated VR shape uncertainty. Observed and expected limits and significances for full Run II data in the less sensitive RT1BT0 categories are shown. The normalization uncertainty addressing data vs. prediction disagreement is included in all cases.

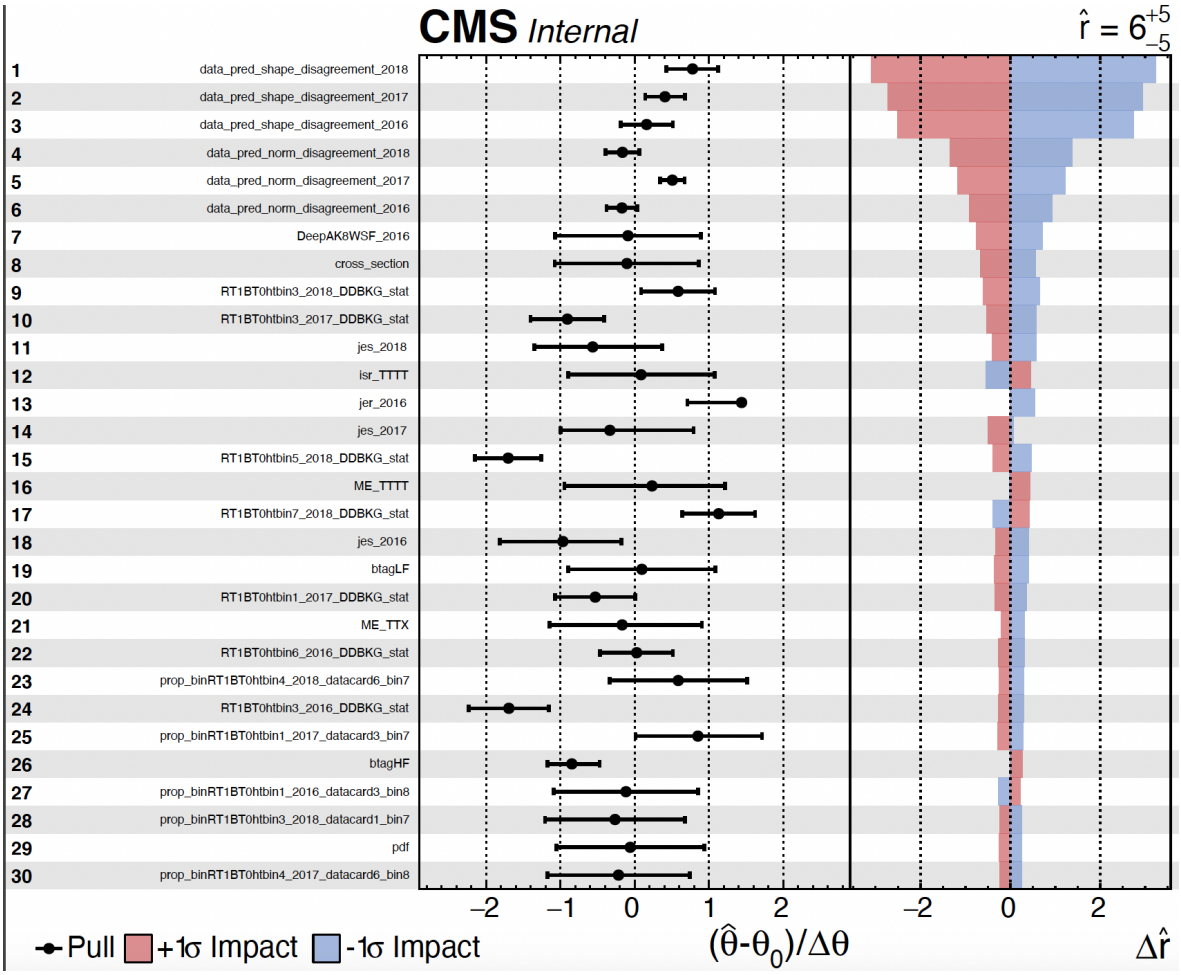

Figure A.7: 30 most important nuisance parameters for the case where shape uncertainties are included and correlated between SR categories. This is the case that is taken as nominal and used in Table 12.1 in the results section, although these impacts represent results including the first 8 less sensitive SR categories only.

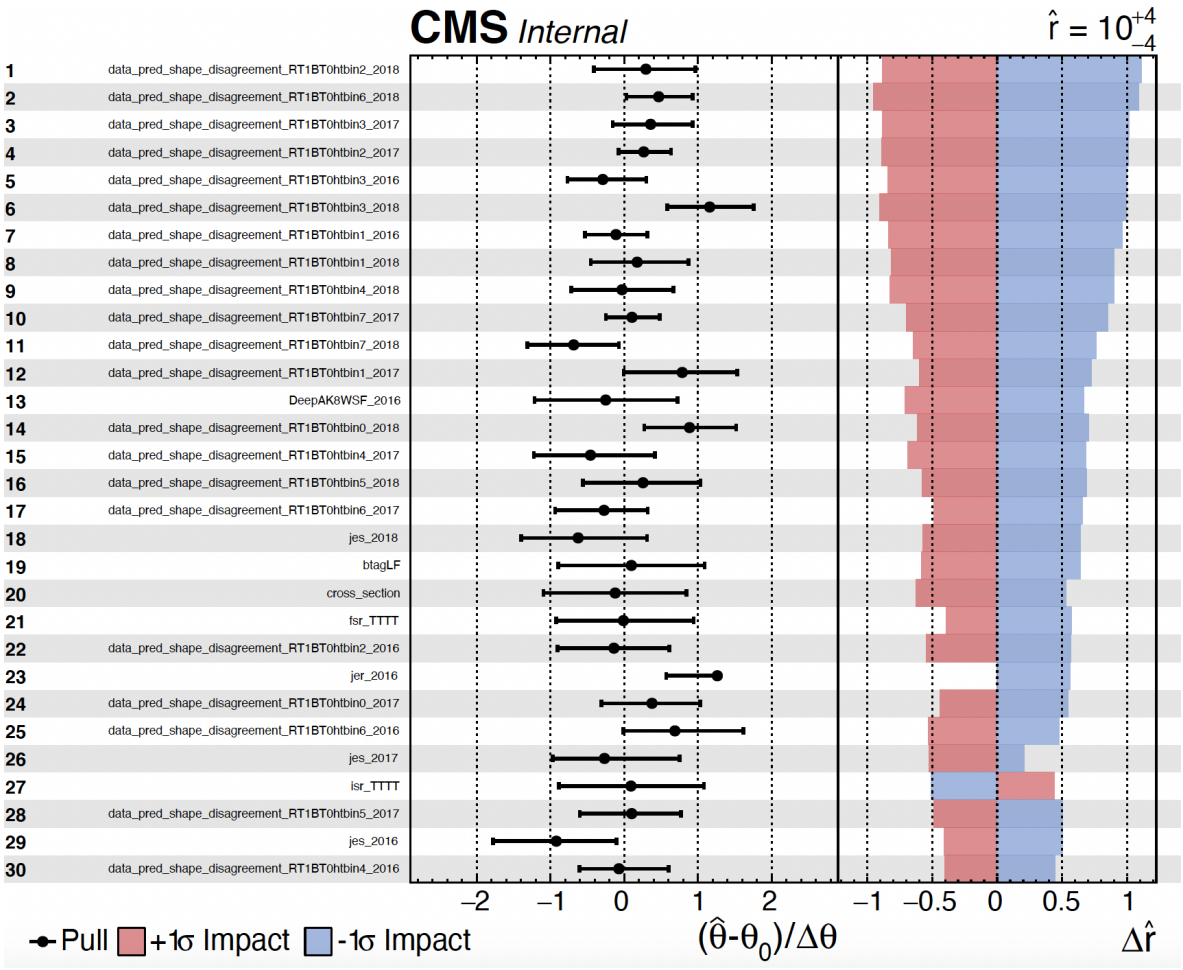

Figure A.8: 30 most important nuisance parameters for the case where shape uncertainties are included and uncorrelated between SR categories. These impacts represent results including the first 8 less sensitive SR categories only.

Appendix B

Distributions of BDT inputs and ROC curves

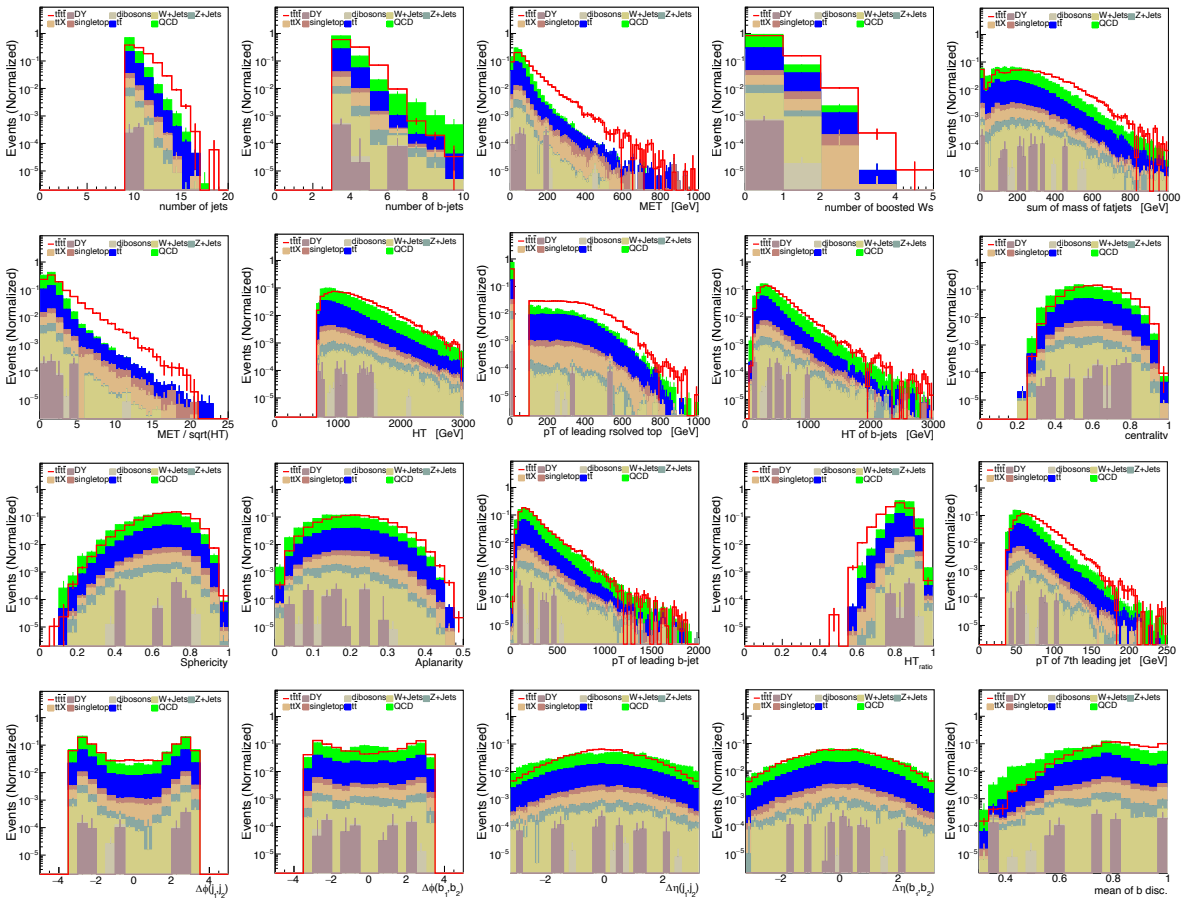

Figure B.1: Distribution of BDT input variables using 16 MC after baseline selection. The signal and backgrounds are normalized to 1. Plots by Hayoung Oh.

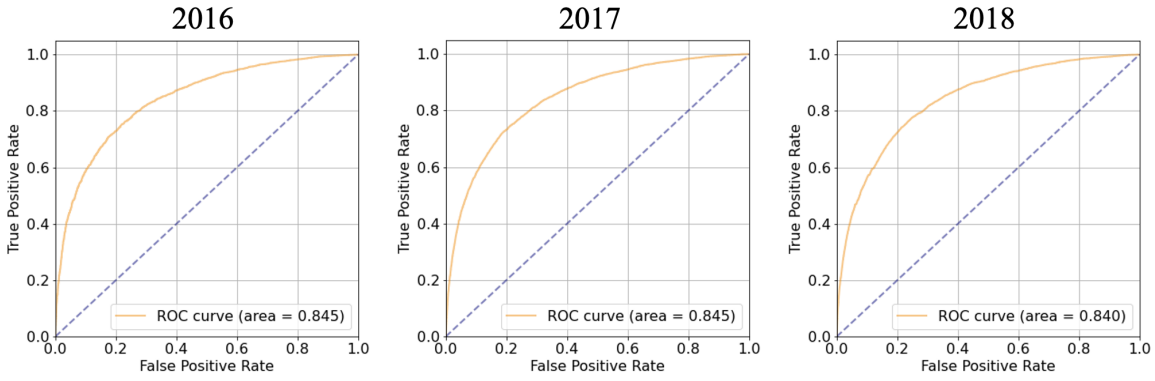

Figure B.2: ROC curves for event level BDT after SR selection for 2016 (left), 2017 (center) and 2018 (right). Plots by Hayoung Oh.

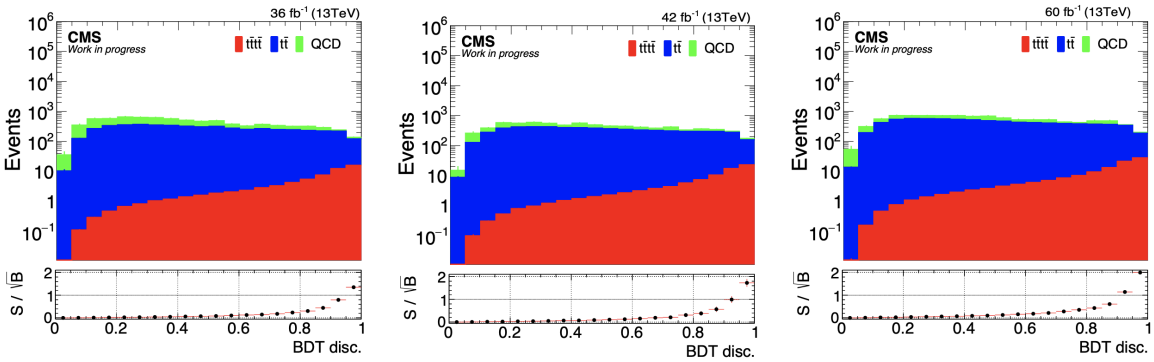

Figure B.3: Distribution of BDT discriminants for simulated data after SR selection for 2016 (left), 2017 (center) and 2018 (right). The lower inset shows signal divided by square root of background bin by bin. Plots by Hayoung Oh.

Appendix C

2D N_{j}/HT-dependent Corrections for B-Tagging Scale Factors

Following recommendations during the object review, we have added $N_{\mathrm{j}}/H_{\mathrm{T}}$-dependent corrections to b-tagging scale factors. These correction factors are derived and applied to $\mathrm{t\bar{t}t\bar{t}}$, $\mathrm{t\bar{t}X}$ and minor background processes estimated from MC. For each of these processes, the correction factors are derived as follows: two 2D distributions of event yields for each simulated process passing the baseline with an inclusive selection N_{b} (at least 9 jets, H_{T} >700GeV, and no leptons) are filled as a function of N_{j} and H_{T}. One 2D distribution has no b-tag scale factors applied (the numerator) and the other 2D distribution has the b-tag scale factors applied (the denominator). The ratio of these two distributions, binned in N_{j} and H_{T}, corresponds to the correction factor for a given process and year, as shown in Fig. C.1-C.9. This factor corrects for shape irregularities as a function of N_{j} and H_{T} as shown for example for $\mathrm{t\bar{t}t\bar{t}}$ in Fig. C.10. These correction factors are applied in tandem with the b-tagging scale factors per event and are varied in turn with each of the separately considered b-tagging systematics (HF, LF, HFstats1, LFstats1, HFstats2, LFstats2, CFerr1, CFerr2, and JEC). Bin sizes were chosen to ensure adequate statistics.

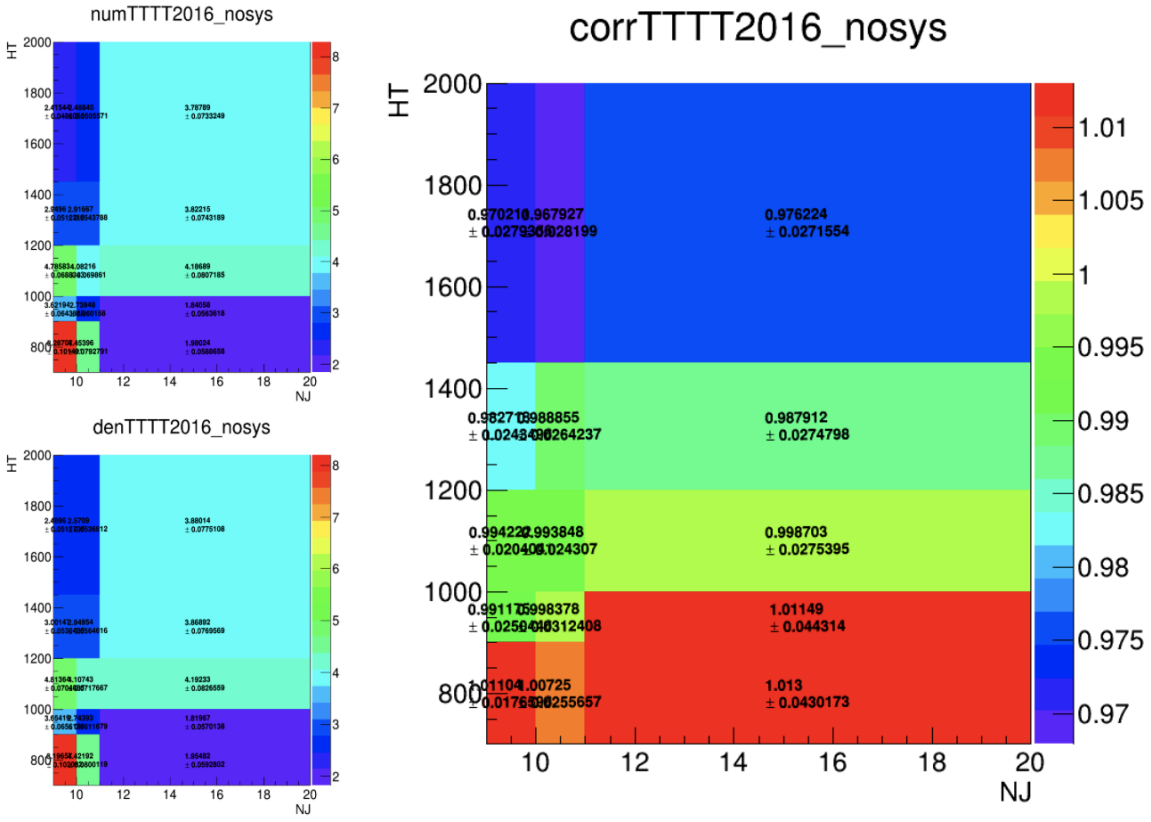

Figure C.1: 2D distributions of N_{j} and H_{T} dependent corrections to B-tagging scale factors for 2016 $\mathrm{t\bar{t}t\bar{t}}$.

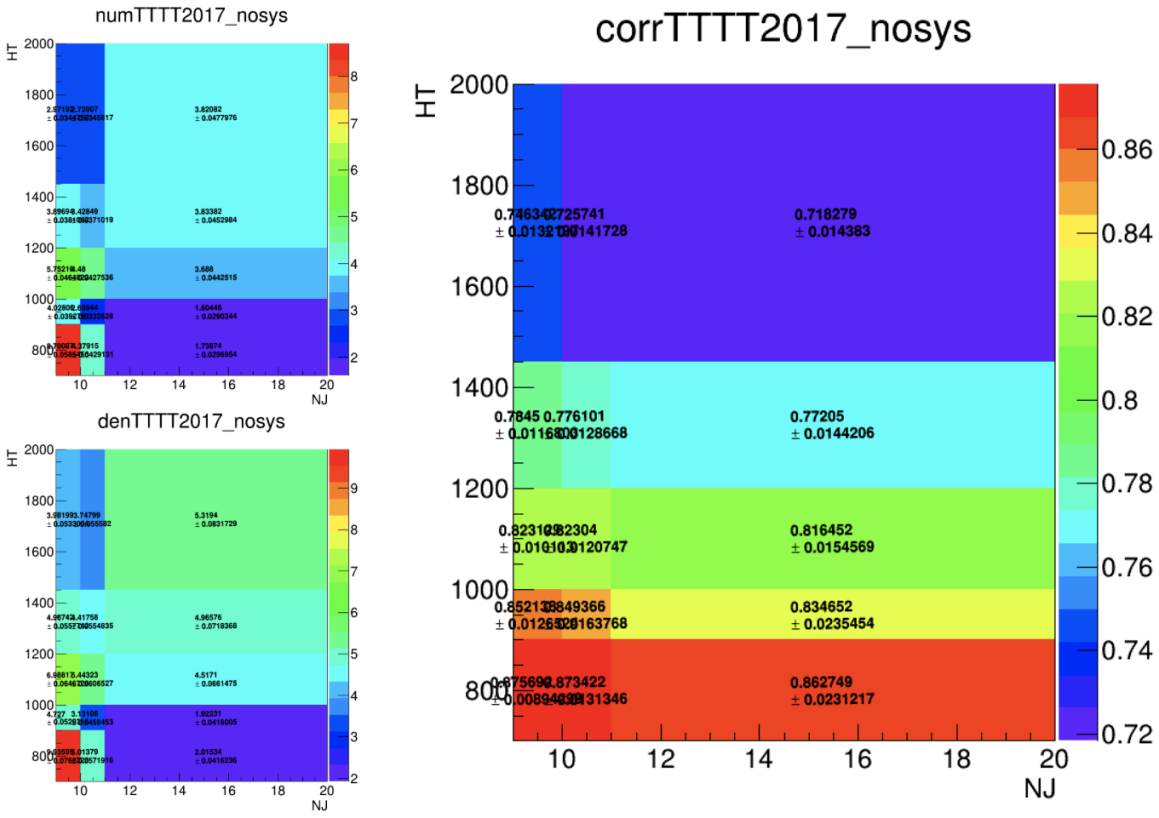

Figure C.2: 2D distributions of N_{j} and H_{T} dependent corrections to B-tagging scale factors for 2017 $\mathrm{t\bar{t}t\bar{t}}$.

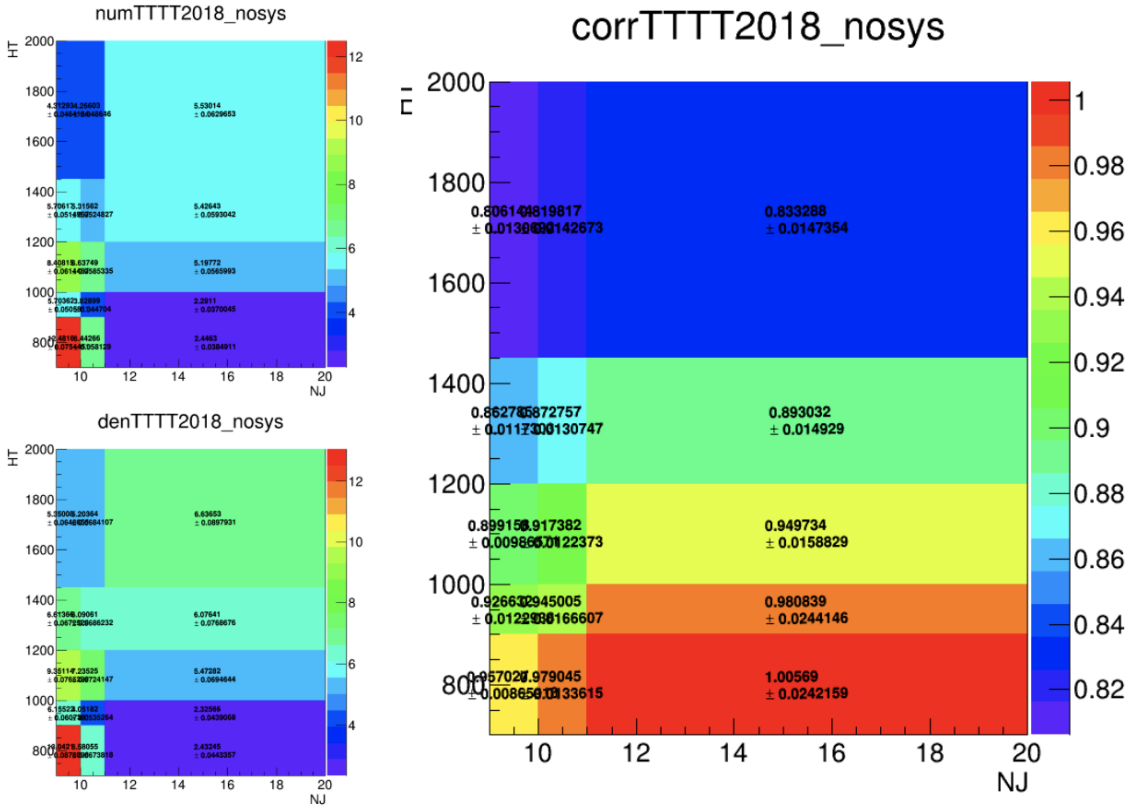

Figure C.3: 2D distributions of N_{j} and H_{T} dependent corrections to B-tagging scale factors for 2018 $\mathrm{t\bar{t}t\bar{t}}$.

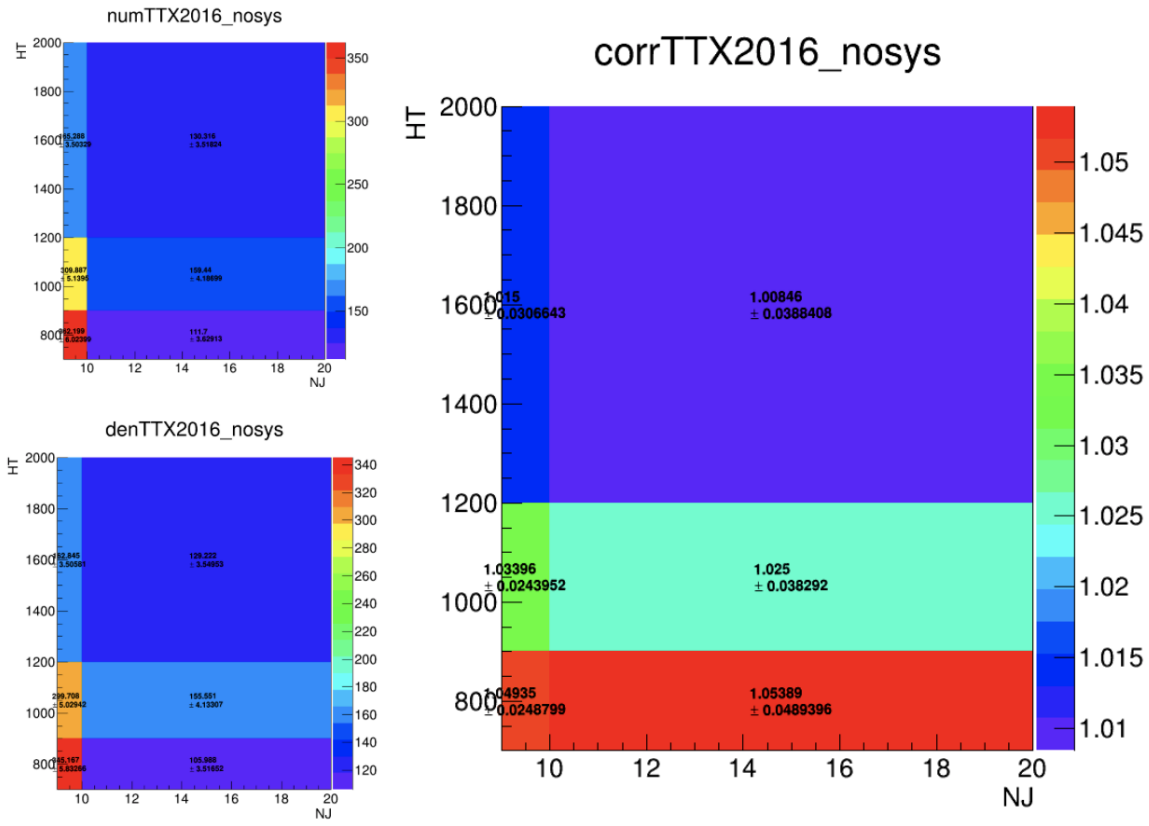

Figure C.4: 2D distributions of N_{j} and H_{T} dependent corrections to B-tagging scale factors for 2016 $\mathrm{t\bar{t}X}$.

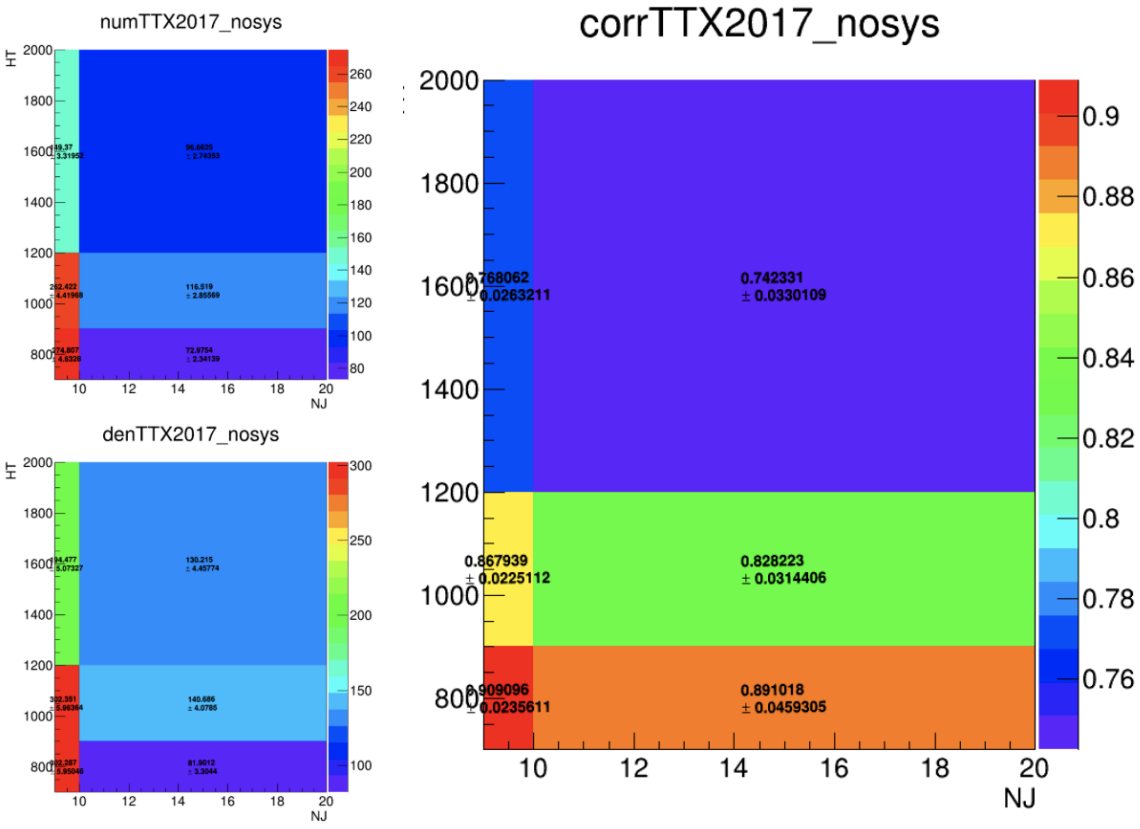

Figure C.5: 2D distributions of N_{j} and H_{T} dependent corrections to B-tagging scale factors for 2017 $\mathrm{t\bar{t}X}$.

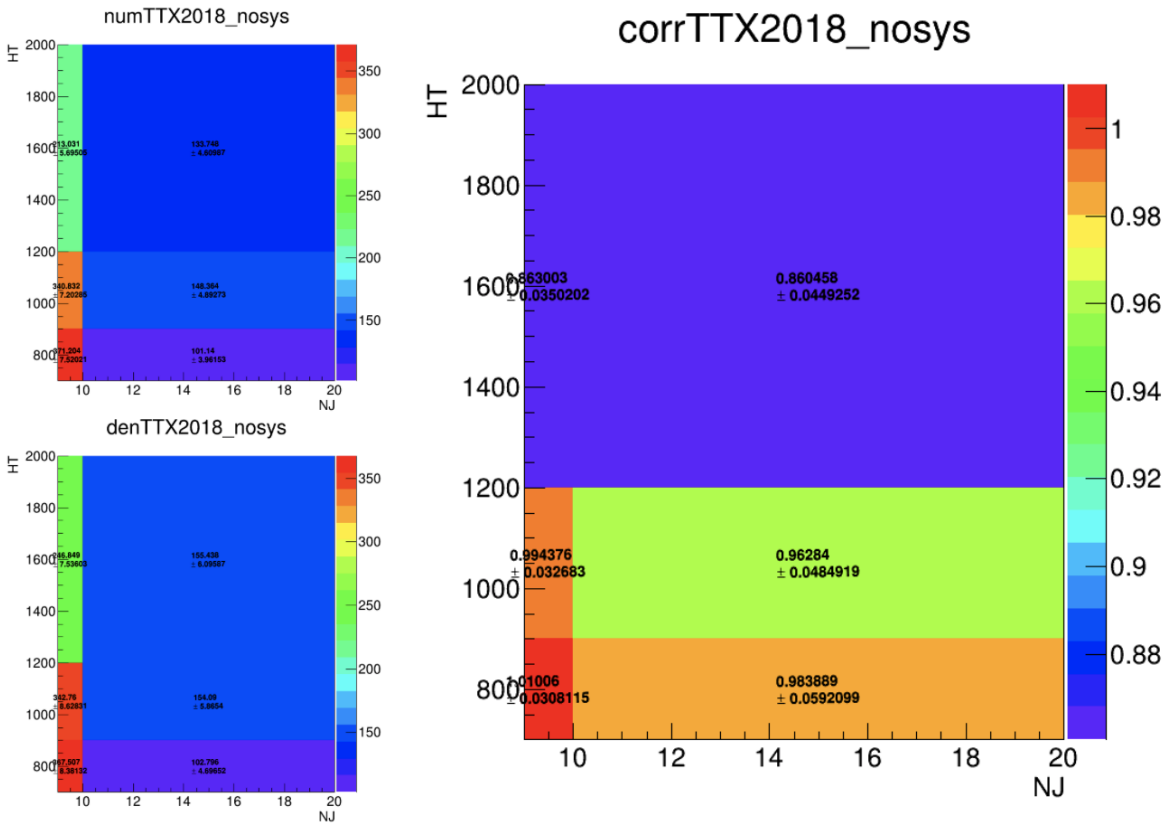

Figure C.6: 2D distributions of N_{j} and H_{T} dependent corrections to B-tagging scale factors for 2018 $\mathrm{t\bar{t}X}$.

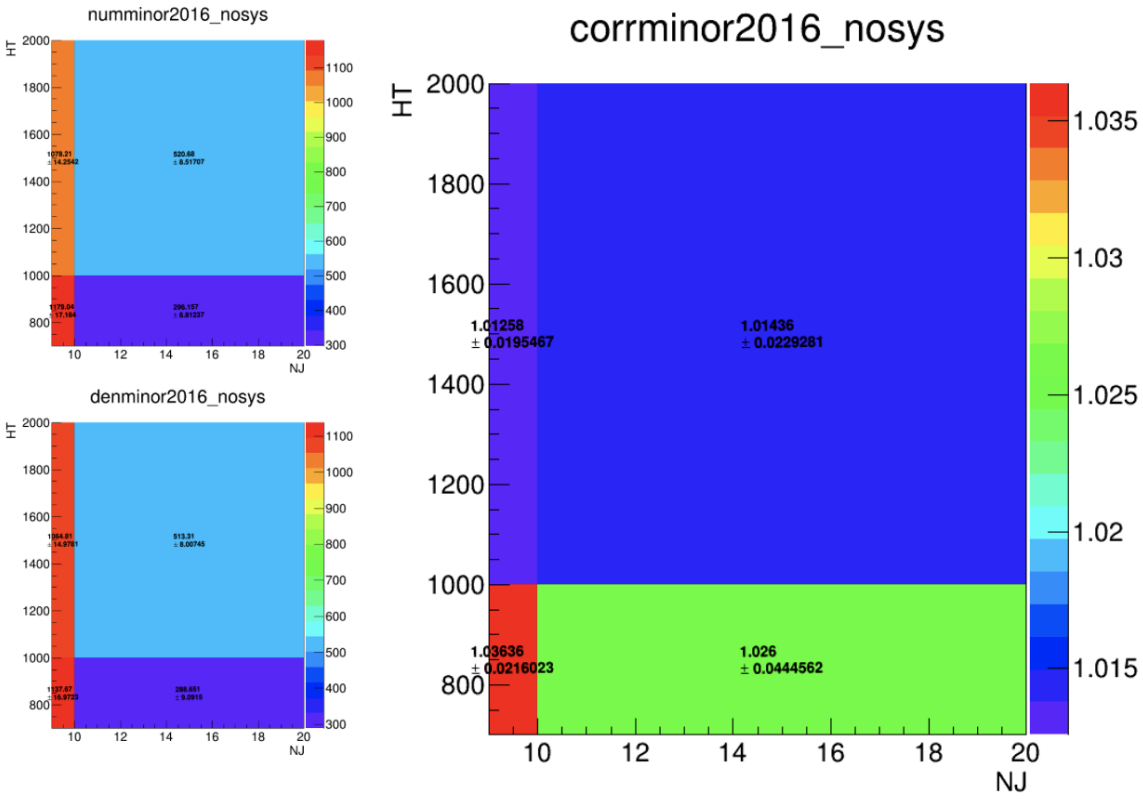

Figure C.7: 2D distributions of N_{j} and H_{T} dependent corrections to B-tagging scale factors for 2016 minor backgrounds.

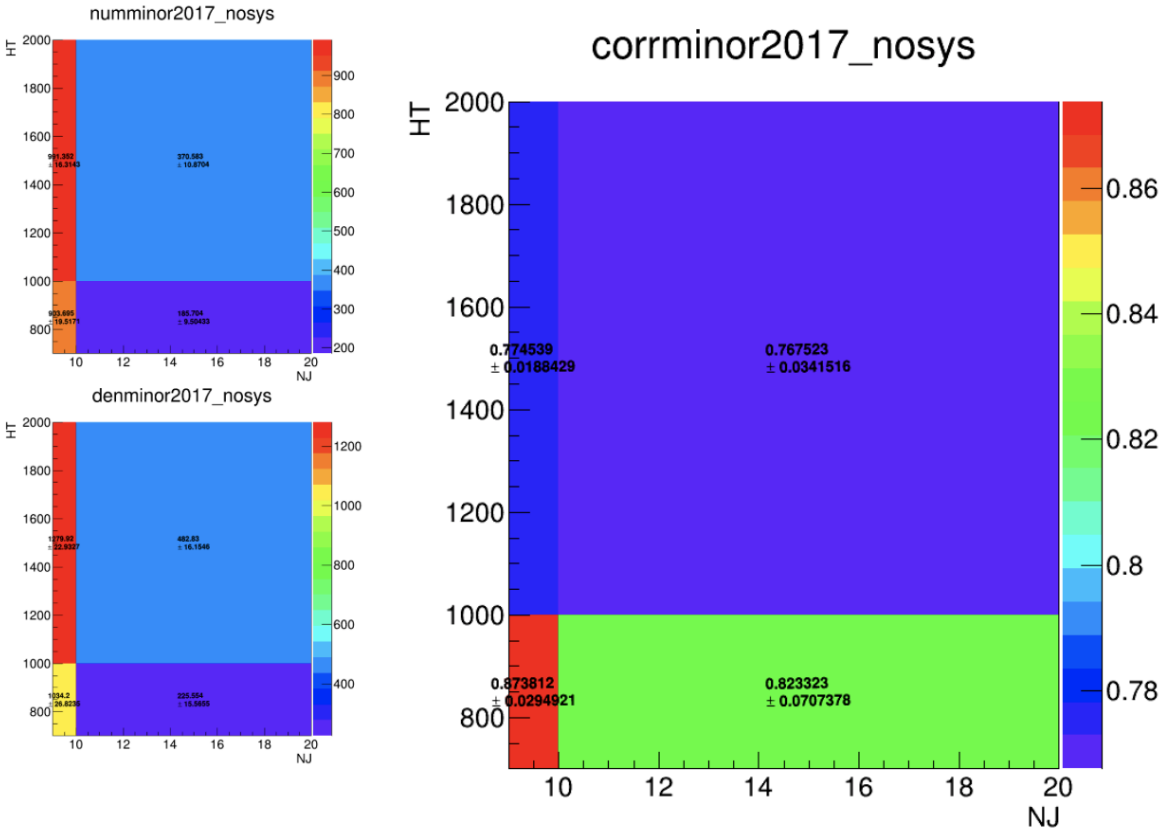

Figure C.8: 2D distributions of N_{j} and H_{T} dependent corrections to B-tagging scale factors for 2017 minor backgrounds.

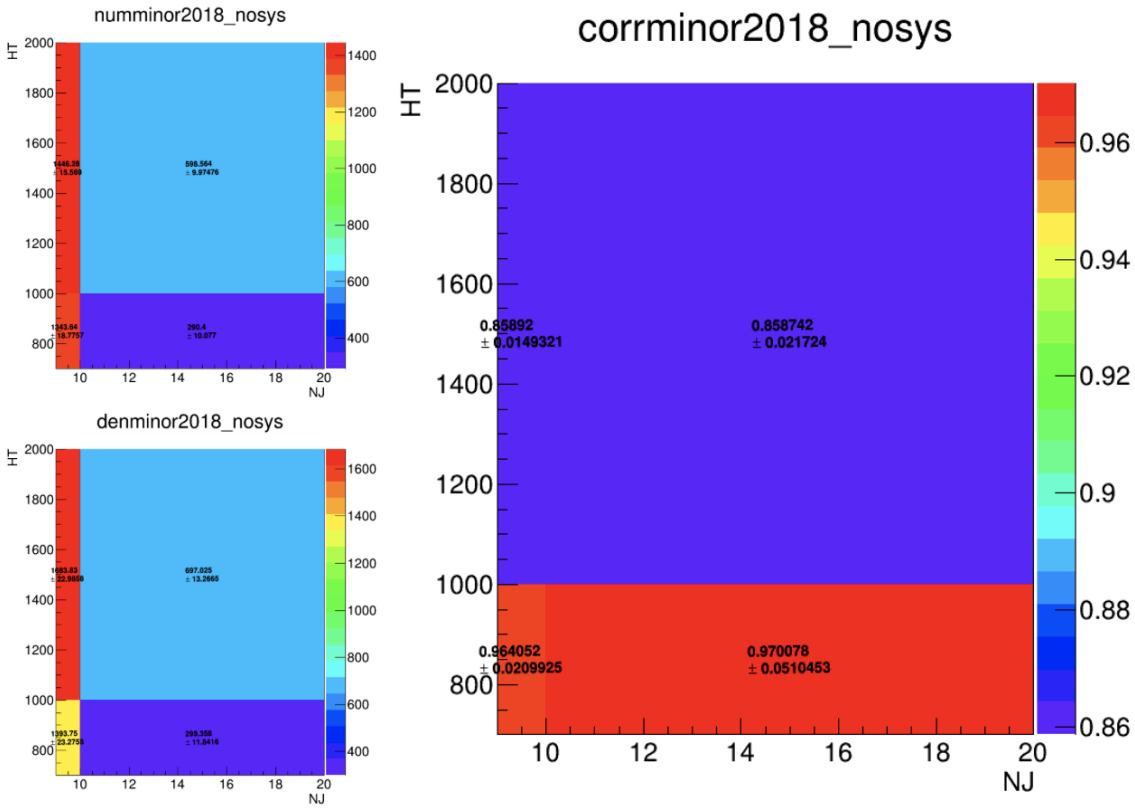

Figure C.9: 2D distributions of N_{j} and H_{T} dependent corrections to B-tagging scale factors for 2018 minor backgrounds.

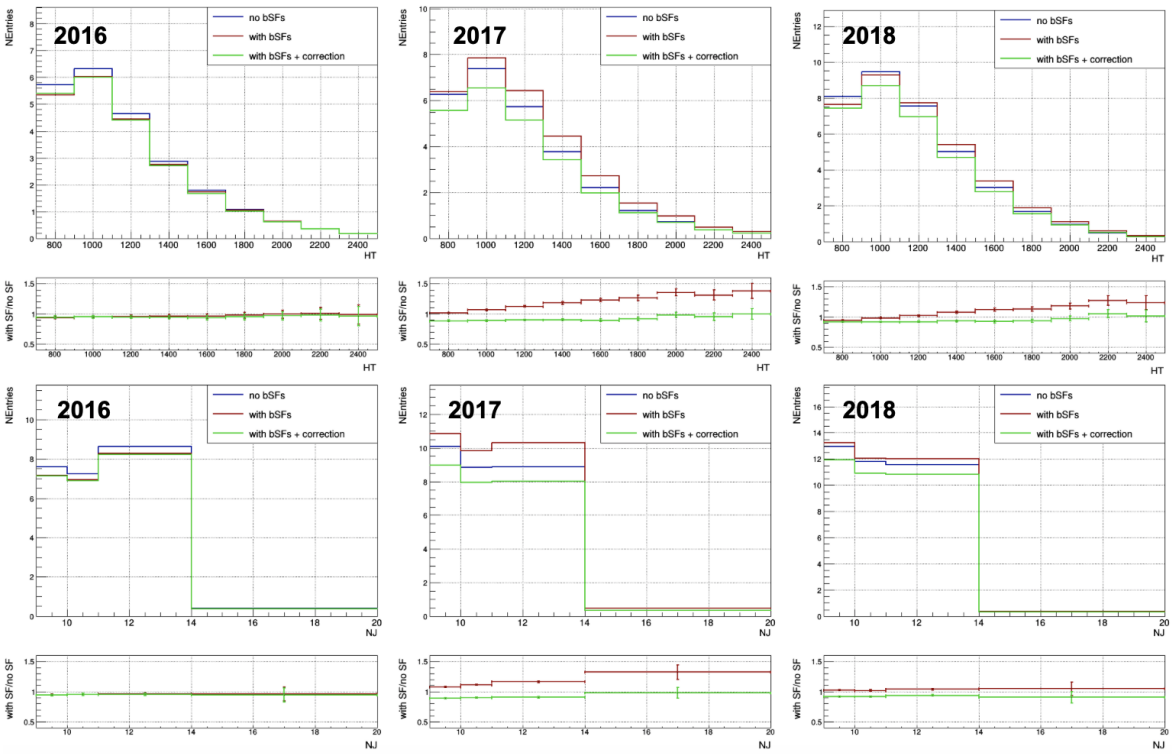

Figure C.10: Examples of b-tagging scale factors for 2016-2018 $t\bar{t}t\bar{t}$ 1D N_j and H_T distributions with and without corrections applied. Baseline selection with 1 or more resolved tops applied.

Appendix D

VR/SR Categorization Analysis

It was noted that for high H_T and high top multiplicity (categories with 1 resolved top and 1 or more boosted tops and 2 or more resolved tops) the 8-jet validation region has significantly lower statistics than the SR. This is evident in Figure 10.7-10.9, and impacts the data-prediction disagreement uncertainty propagated from VR to SR bins as described in 11. The effects of combining the H_T bins in the VR to better represent the SR were considered, but ultimately it was decided to split the bins in the same way as the SR for simplicity and because the overall sensitivity was not generally affected by these variations, despite the possibility that the propagated uncertainties are slightly too conservative. Effects of correlating or not correlating these uncertainties between SR categories were also considered, and were found to have a minimal impact on the sensitivity. We therefore decided to keep the uncertainty correlated between the different SR categories, in order to reflect the possible correlation of effects as a function of H_T and top multiplicity. The following figures contain details of studies performed.

Year		NRT=1 NBT≥1 HT<1400	NRT=1 NBT≥1 HT≥1400	NRT=1 NBT≥1 combined	NRT≥2 HT<1400	NRT≥2 HT≥1400	NRT≥2 combined
2016	VR	350.862 ± 36.514	129.763 ± 31.142	490.793 ± 46.331	474.538 ± 43.211	109.338 ± 38.209	584.268 ± 50.8
	SR	184.436 ± 25.015	263.052 ± 54.089	413.373 ± 45.458	469.543 ± 45.345	262.817 ± 55.318	722.601 ± 62.053
2017	VR	433.791 ± 46.29	170.154 ± 41.706	607.134 ± 58.436	662.076 ± 62.449	166.729 ± 70.523	821.76 ± 74.195
	SR	283.106 ± 36.718	428.509 ± 80.80	637.572 ± 66.215	548.215 ± 52.361	318.736 ± 64.432	868.582 ± 73.00
2018	VR	733.211 ± 62.743	181.385 ± 38.724	956.752 ± 74.631	872.975 ± 70.127	306.14 ± 103.922	1120.11 ± 85.528
	SR	433.497 ± 45.683	494.323 ± 82.33	853.048 ± 73.975	619.241 ± 52.859	350.532 ± 59.38	980.936 ± 73.123

Figure D.1: Predicted QCD+$t\bar{t}$ yields for low H_T, high H_T, and combined H_T bins in relevant SR and VR categories. Note that the high H_T bins (red) have more statistics in the SR than the VR bins.

Bin	2016	2017	2018
NRT=1 NBT≧1 HT<1400	0.277	0.227	0.188
NRT=1 NBT≧1 HT≥1400	0.300	0.293	0.460
NRT=1 NBT≧1 combined	0.217	0.172	0.112
NRT≧2 HT<1400	0.275	0.146	0.100
NRT≧2 HT≥1400	0.405	0.140	0.324
NRT≧2 combined	0.261	0.129	0.105

Figure D.2: Data-prediction disagreement uncertainties for low H_{T}, high H_{T}, and combined H_{T} bins in relevant VR categories (and propagated to corresponding SR categories).

Splitting of VR bins	2016 μ	2017 μ	2018 μ	RunII μ	RunII σ
NRT≧2 bin combined, NRT =1 NBT≧1 bin split All correlated	6.41	6.09	5.11	3.28	0.631
Both split All correlated	6.41	6.09	5.11	3.28	0.630
Both combined All correlated	6.41	6.09	5.08	3.27	0.632
Both split Not correlated	6.41	6.08	5.11	3.28	0.631
Both combined Not correlated	6.41	6.09	5.06	3.27	0.633

Figure D.3: Senstitivity (expected limits and expected significance) for different data-prediction discrepancy uncertainty scenarios related to splitting of VR H_{T} bins. "Both combined" and "Both split" refer to combined low and high H_{T} bins when calculating the uncertainty in relevant VR categories. "All correlated" and "Not correlated" refer to whether the propagated uncertainties are correlated or not correlated between SR categories in the datacards. The version used in 12 is "All correlated" "Both split".

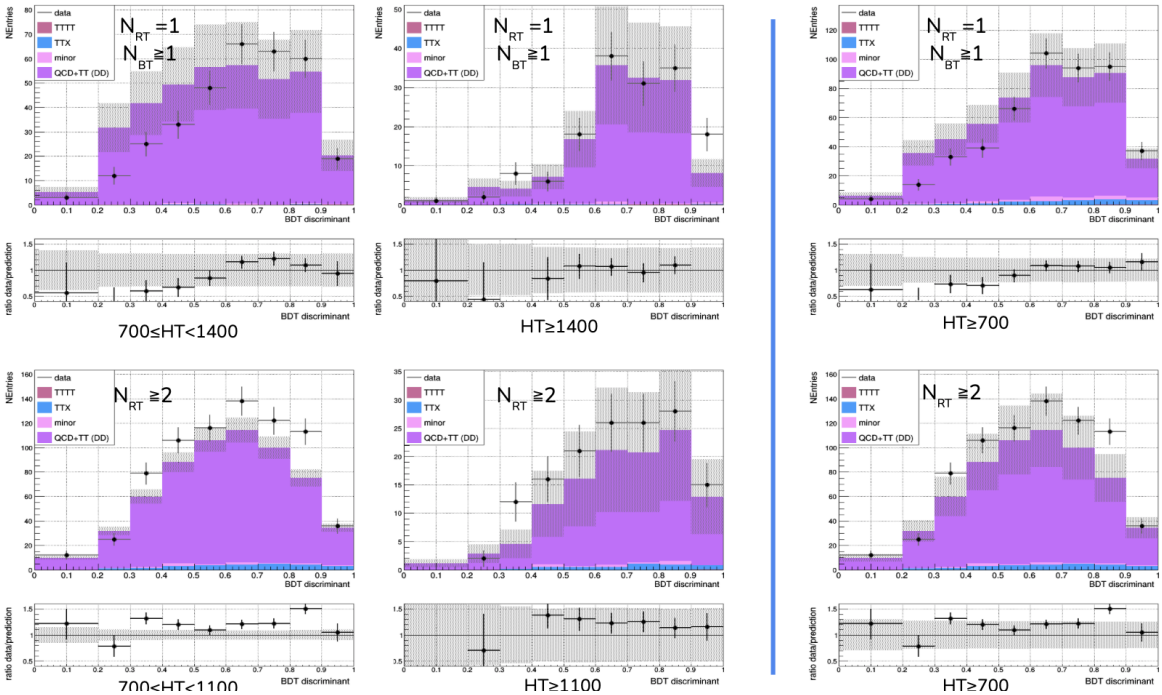

Figure D.4: 2016 BDT shape predictions vs data corresponding to the SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the 8-jet validation region, comparing split H_{T} (left) vs. combined H_{T} bins (right) in the two relevant top multiplicity categories. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include statistical uncertainties and uncertainties derived from the weighted mean and RMS to account for discrepancies between data and the data–driven background estimate observed in the VRs, as described in Section 11.

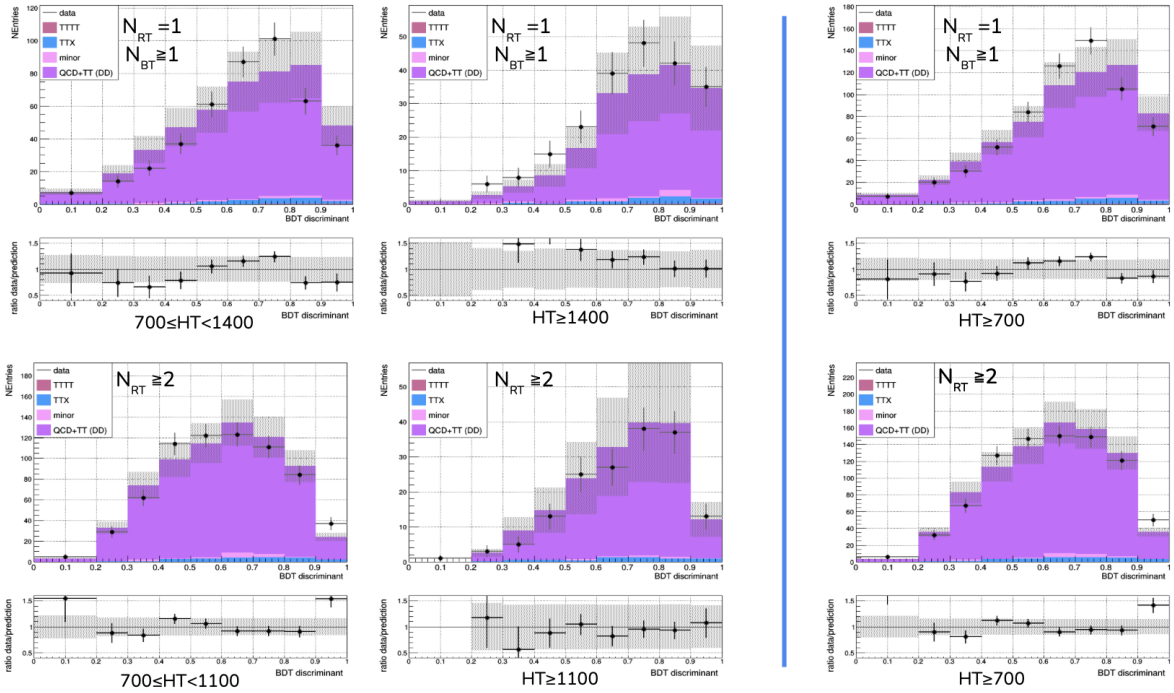

Figure D.5: 2017 BDT shape predictions vs data corresponding to the SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the 8-jet validation region, comparing split H_{T} (left) vs. combined H_{T} bins (right) in the two relevant top multiplicity categories. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include statistical uncertainties and uncertainties derived from the weighted mean and RMS to account for discrepancies between data and the data–driven background estimate observed in the VRs, as described in Section 11.

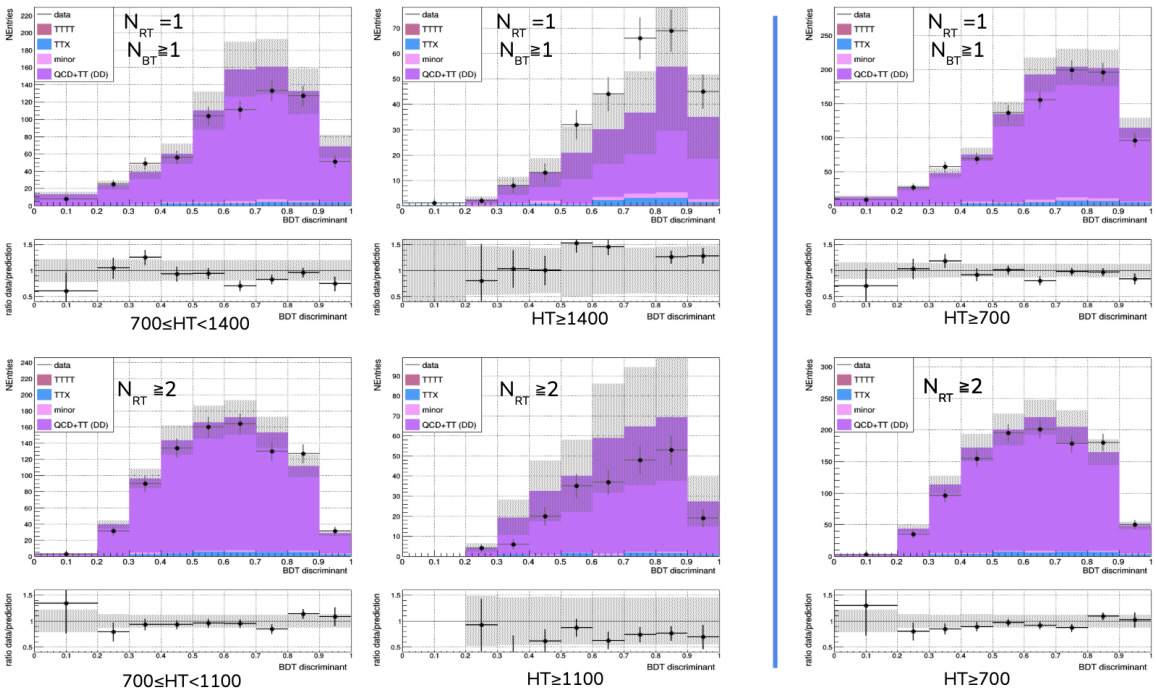

Figure D.6: 2018 BDT shape predictions vs data corresponding to the SR bins defined by N_{RT}, N_{BT}, and H_{T}, for the 8-jet validation region, comparing split H_{T} (left) vs. combined H_{T} bins (right) in the two relevant top multiplicity categories. The $\mathrm{t\bar{t}}$ and QCD multijet background BDT discriminant shape is predicted by the ABCDnn and normalized to the yields predicted by the extended ABCD method. Estimates for $\mathrm{t\bar{t}t\bar{t}}$ signal and other minor backgrounds are shown using simulated samples. The error bands shown include statistical uncertainties and uncertainties derived from the weighted mean and RMS to account for discrepancies between data and the data–driven background estimate observed in the VRs, as described in Section 11.

Bibliography

[1] R. Frederix, D. Pagani, and M. Zaro, *Large nlo corrections in $t\bar{t}w^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading ew contributions*, 1711.0211.

[2] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, *Computer Physics Communications* **185** (Nov, 2014) 2930–2938.

[3] Q.-H. Cao, S.-L. Chen, and Y. Liu, *Probing higgs width and top quark yukawa coupling from tth and tttt productions*, *Physical Review D* **95** (Mar, 2017).

[4] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, *Theory and phenomenology of two-higgs-doublet models*, 1106.0034.

[5] **CMS** Collaboration, A. M. Sirunyan *et. al.*, *Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at sqrt(s)= 13 TeV*, *Eur. Phys. J. C* **80** (2020), no. 2 75, [arXiv:1908.0646].

[6] **CMS** Collaboration, A. M. Sirunyan *et. al.*, *Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at sqrt(s)= 13 TeV*, *JHEP* **11** (2019) 082, [arXiv:1906.0280].

[7] **CMS** Collaboration, *Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment*, CMS Physics Analysis Summary CMS-PAS-JME-18-002, 2019.

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

[9] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*. Westview, Boulder, CO, 1995. Includes exercises.

[10] M. Rieger, *Search for Higgs Boson Production in Association with Top Quarks and Decaying into Bottom Quarks using Deep Learning Techniques with the CMS Experiment*, 2019. presented on 2019-06-07.

[11] **Particle Data Group** Collaboration, M. Tanabashi, K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, F. Takahashi, J. Tanaka, K. Agashe, G. Aielli, C. Amsler, M. Antonelli, D. M. Asner, H. Baer, S. Banerjee, R. M. Barnett, T. Basaglia, C. W. Bauer, J. J. Beatty, V. I. Belousov, J. Beringer, S. Bethke, A. Bettini, H. Bichsel, O. Biebel, K. M. Black, E. Blucher, O. Buchmuller, V. Burkert, M. A. Bychkov, R. N. Cahn, M. Carena, A. Ceccucci, A. Cerri, D. Chakraborty, M.-C. Chen, R. S. Chivukula, G. Cowan, O. Dahl, G. D'Ambrosio, T. Damour, D. de Florian, A. de Gouvêa, T. DeGrand, P. de Jong, G. Dissertori, B. A. Dobrescu, M. D'Onofrio, M. Doser, M. Drees, H. K. Dreiner, D. A. Dwyer, P. Eerola, S. Eidelman, J. Ellis, J. Erler, V. V. Ezhela, W. Fetscher, B. D. Fields, R. Firestone, B. Foster, A. Freitas, H. Gallagher, L. Garren, H.-J. Gerber, G. Gerbier, T. Gershon, Y. Gershtein, T. Gherghetta, A. A. Godizov, M. Goodman, C. Grab, A. V. Gritsan, C. Grojean, D. E. Groom, M. Grünewald, A. Gurtu, T. Gutsche, H. E. Haber, C. Hanhart, S. Hashimoto, Y. Hayato, K. G. Hayes, A. Hebecker, S. Heinemeyer, B. Heltsley, J. J. Hernández-Rey, J. Hisano, A. Höcker, J. Holder, A. Holtkamp, T. Hyodo, K. D. Irwin, K. F. Johnson, M. Kado, M. Karliner, U. F. Katz, S. R. Klein, E. Klempt, R. V. Kowalewski, F. Krauss, M. Kreps, B. Krusche, Y. V. Kuyanov, Y. Kwon, O. Lahav, J. Laiho, J. Lesgourgues, A. Liddle, Z. Ligeti, C.-J. Lin, C. Lippmann, T. M. Liss, L. Littenberg, K. S. Lugovsky, S. B. Lugovsky, A. Lusiani, Y. Makida, F. Maltoni, T. Mannel, A. V. Manohar, W. J. Marciano, A. D. Martin, A. Masoni, J. Matthews, U.-G. Meißner, D. Milstead, R. E. Mitchell, K. Mönig, P. Molaro, F. Moortgat, M. Moskovic, H. Murayama, M. Narain, P. Nason, S. Navas, M. Neubert, P. Nevski, Y. Nir, K. A. Olive, S. Pagan Griso, J. Parsons, C. Patrignani, J. A. Peacock, M. Pennington, S. T. Petcov, V. A. Petrov, E. Pianori, A. Piepke, A. Pomarol, A. Quadt, J. Rademacker, G. Raffelt, B. N. Ratcliff, P. Richardson, A. Ringwald, S. Roesler, S. Rolli, A. Romaniouk, L. J. Rosenberg, J. L. Rosner, G. Rybka, R. A. Ryutin, C. T. Sachrajda, Y. Sakai, G. P. Salam, S. Sarkar, F. Sauli, O. Schneider, K. Scholberg, A. J. Schwartz, D. Scott, V. Sharma, S. R. Sharpe, T. Shutt, M. Silari, T. Sjöstrand, P. Skands, T. Skwarnicki, J. G. Smith, G. F. Smoot, S. Spanier, H. Spieler, C. Spiering, A. Stahl, S. L. Stone, T. Sumiyoshi, M. J. Syphers, K. Terashi, J. Terning, U. Thoma, R. S. Thorne, L. Tiator, M. Titov, N. P. Tkachenko, N. A. Törnqvist, D. R. Tovey, G. Valencia, R. Van de Water, N. Varelas, G. Venanzoni, L. Verde, M. G. Vincter, P. Vogel, A. Vogt, S. P. Wakely, W. Walkowiak, C. W. Walter, D. Wands, D. R. Ward, M. O. Wascko, G. Weiglein, D. H. Weinberg, E. J. Weinberg, M. White, L. R. Wiencke, S. Willocq, C. G. Wohl, J. Womersley, C. L. Woody, R. L. Workman, W.-M. Yao, G. P. Zeller, O. V. Zenin, R.-Y. Zhu, S.-L. Zhu, F. Zimmermann, P. A. Zyla, J. Anderson, L. Fuller, V. S. Lugovsky, and P. Schaffner, *Review of particle physics*, *Phys. Rev. D* **98** (Aug, 2018) 030001.

[12] **Particle Data Group** Collaboration, M. Tanabashi, K. Hagiwara, K. Hikasa,

K. Nakamura, Y. Sumino, F. Takahashi, J. Tanaka, K. Agashe, G. Aielli, C. Amsler, M. Antonelli, D. M. Asner, H. Baer, S. Banerjee, R. M. Barnett, T. Basaglia, C. W. Bauer, J. J. Beatty, V. I. Belousov, J. Beringer, S. Bethke, A. Bettini, H. Bichsel, O. Biebel, K. M. Black, E. Blucher, O. Buchmuller, V. Burkert, M. A. Bychkov, R. N. Cahn, M. Carena, A. Ceccucci, A. Cerri, D. Chakraborty, M.-C. Chen, R. S. Chivukula, G. Cowan, O. Dahl, G. D'Ambrosio, T. Damour, D. de Florian, A. de Gouvêa, T. DeGrand, P. de Jong, G. Dissertori, B. A. Dobrescu, M. D'Onofrio, M. Doser, M. Drees, H. K. Dreiner, D. A. Dwyer, P. Eerola, S. Eidelman, J. Ellis, J. Erler, V. V. Ezhela, W. Fetscher, B. D. Fields, R. Firestone, B. Foster, A. Freitas, H. Gallagher, L. Garren, H.-J. Gerber, G. Gerbier, T. Gershon, Y. Gershtein, T. Gherghetta, A. A. Godizov, M. Goodman, C. Grab, A. V. Gritsan, C. Grojean, D. E. Groom, M. Grünewald, A. Gurtu, T. Gutsche, H. E. Haber, C. Hanhart, S. Hashimoto, Y. Hayato, K. G. Hayes, A. Hebecker, S. Heinemeyer, B. Heltsley, J. J. Hernández-Rey, J. Hisano, A. Höcker, J. Holder, A. Holtkamp, T. Hyodo, K. D. Irwin, K. F. Johnson, M. Kado, M. Karliner, U. F. Katz, S. R. Klein, E. Klempt, R. V. Kowalewski, F. Krauss, M. Kreps, B. Krusche, Y. V. Kuyanov, Y. Kwon, O. Lahav, J. Laiho, J. Lesgourgues, A. Liddle, Z. Ligeti, C.-J. Lin, C. Lippmann, T. M. Liss, L. Littenberg, K. S. Lugovsky, S. B. Lugovsky, A. Lusiani, Y. Makida, F. Maltoni, T. Mannel, A. V. Manohar, W. J. Marciano, A. D. Martin, A. Masoni, J. Matthews, U.-G. Meißner, D. Milstead, R. E. Mitchell, K. Mönig, P. Molaro, F. Moortgat, M. Moskovic, H. Murayama, M. Narain, P. Nason, S. Navas, M. Neubert, P. Nevski, Y. Nir, K. A. Olive, S. Pagan Griso, J. Parsons, C. Patrignani, J. A. Peacock, M. Pennington, S. T. Petcov, V. A. Petrov, E. Pianori, A. Piepke, A. Pomarol, A. Quadt, J. Rademacker, G. Raffelt, B. N. Ratcliff, P. Richardson, A. Ringwald, S. Roesler, S. Rolli, A. Romaniouk, L. J. Rosenberg, J. L. Rosner, G. Rybka, R. A. Ryutin, C. T. Sachrajda, Y. Sakai, G. P. Salam, S. Sarkar, F. Sauli, O. Schneider, K. Scholberg, A. J. Schwartz, D. Scott, V. Sharma, S. R. Sharpe, T. Shutt, M. Silari, T. Sjöstrand, P. Skands, T. Skwarnicki, J. G. Smith, G. F. Smoot, S. Spanier, H. Spieler, C. Spiering, A. Stahl, S. L. Stone, T. Sumiyoshi, M. J. Syphers, K. Terashi, J. Terning, U. Thoma, R. S. Thorne, L. Tiator, M. Titov, N. P. Tkachenko, N. A. Törnqvist, D. R. Tovey, G. Valencia, R. Van de Water, N. Varelas, G. Venanzoni, L. Verde, M. G. Vincter, P. Vogel, A. Vogt, S. P. Wakely, W. Walkowiak, C. W. Walter, D. Wands, D. R. Ward, M. O. Wascko, G. Weiglein, D. H. Weinberg, E. J. Weinberg, M. White, L. R. Wiencke, S. Willocq, C. G. Wohl, J. Womersley, C. L. Woody, R. L. Workman, W.-M. Yao, G. P. Zeller, O. V. Zenin, R.-Y. Zhu, S.-L. Zhu, F. Zimmermann, P. A. Zyla, J. Anderson, L. Fuller, V. S. Lugovsky, and P. Schaffner, *Review of particle physics*, *Phys. Rev. D* **98** (Aug, 2018) 030001.

[13] J. Ellis, *Higgs Physics*, arXiv:1312.5672. 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of High-Energy Physics, to appear as part of the

proceedings in a CERN Yellow Report.

[14] S. P. MARTIN, *A supersymmetry primer*, *Advanced Series on Directions in High Energy Physics* (Jul, 1998) 1–98.

[15] O. S. Bruening, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[16] J. Pivarski, *Physics in a nutshell: Cross section*, *Fermilab Today* (Friday, March 1, 2013).

[17] A. Korytov, *Note 5: Cross Section*. The University of Florida, 2008.

[18] W. Herr and B. Muratori, *Concept of luminosity*, .

[19] M. A. Shah, *Experiences from the rpc data taking during the lhc run-2*, .

[20] **CMS Collaboration** Collaboration, C. Collaboration, *The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment*, *JINST* **3** (2008) S08004. 361 p. Also published by CERN Geneva in 2010.

[21] **CMS Collaboration** Collaboration, C. Collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. CERN, Geneva, 2006.

[22] K. Gumus and N. Akchurin, *Search for new physics in the compact muon solenoid (cms) experiment and the response of the cms calorimeters to particles and jets*, .

[23] **CMS Collaboration** Collaboration, G. L. Bayatyan, M. Della Negra, Foà, A. Hervé, and A. Petrilli, *CMS computing: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 2005. Submitted on 31 May 2005.

[24] A. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, M. Flechl, M. Friedl, R. Frühwirth, V. Ghete, and et al., *Particle-flow reconstruction and global event description with the cms detector*, *Journal of Instrumentation* **12** (Oct, 2017) P10003–P10003.

[25] V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, M. Flechl, M. Friedl, R. Frühwirth, and et al., *Jet energy scale and resolution in the cms experiment in pp collisions at 8 tev*, *Journal of Instrumentation* **12** (Feb, 2017) P02014–P02014.

[26] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-ktjet clustering algorithm*, *Journal of High Energy Physics* **2008** (Apr, 2008) 063–063.

[27] M. Cacciari, G. P. Salam, and G. Soyez, *Fastjet user manual*, *The European Physical Journal C* **72** (Mar, 2012).

[28] D. P.-A. R. J. S. C. Coco, Victor and G. Soyez, *Jets and jet algorithms*, .

[29] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, *Jet flavour classification using deepjet*, *Journal of Instrumentation* **15** (Dec, 2020) P12012–P12012.

[30] **CMS Collaboration** Collaboration, *Identification of b quark jets at the CMS Experiment in the LHC Run 2*, tech. rep., CERN, Geneva, 2016.

[31] I. Neutelings, *Cms coordinate system*, Oct, 2021.

[32] **The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group** Collaboration, *Procedure for the LHC Higgs boson search combination in Summer 2011*, tech. rep., CERN, Geneva, Aug, 2011.

[33] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *The European Physical Journal C* **71** (Feb, 2011).

[34] J. S. Conway, *Incorporating nuisance parameters in likelihoods for multisource spectra*, 2011.

[35] R. Barlow and C. Beeston, *Fitting using finite monte carlo samples*, *Computer Physics Communications* **77** (1993), no. 2 219–228.

[36] G. Cowan, C. Patrignani, *et. al.*, *Statistics, ch. 39 in particle data group*, *Chin. Phys. C* **40** (2016) 100001.

[37] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *The Annals of Mathematical Statistics* **9** (1938), no. 1 60–62.

[38] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, *Transactions of the American Mathematical Society* **54** (1943), no. 3 426–482.

[39] A. Geron, *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Sebastopol, CA, 2017.

[40] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[41] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), pp. 785–794, ACM, 2016.

[42] A. V. Dorogush, V. Ershov, and A. Gulin, *Catboost: gradient boosting with categorical features support*, *CoRR* **abs/1810.11363** (2018) [arXiv:1810.1136].

[43] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. J. Smola, *A kernel method for the two-sample problem*, *CoRR* **abs/0805.2368** (2008) [arXiv:0805.2368].

[44] Y. Li, K. Swersky, and R. S. Zemel, *Generative moment matching networks*, *CoRR* **abs/1502.02761** (2015) [arXiv:1502.0276].

[45] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2015.

[46] C.-W. Huang, D. Krueger, A. Lacoste, and A. Courville, *Neural Autoregressive Flows*, arXiv:1804.0077.

[47] M. A. Dobbs, S. Frixione, E. Laenen, K. Tollefson, H. Baer, E. Boos, B. Cox, R. Engel, W. Giele, J. Huston, S. Ilyin, B. Kersevan, F. Krauss, Y. Kurihara, L. Lonnblad, F. Maltoni, M. Mangano, S. Odaka, P. Richardson, A. Ryd, T. Sjostrand, P. Skands, Z. Was, B. R. Webber, and D. Zeppenfeld, *Les houches guidebook to monte carlo generators for hadron collider physics*, 2004.

[48] Rizzi, Andrea, Petrucciani, Giovanni, and Peruzzi, Marco, *A further reduction in cms event data for analysis: the nanoaod format*, *EPJ Web Conf.* **214** (2019) 06021.

[49] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, *Array programming with NumPy*, *Nature* **585** (Sept., 2020) 357–362.

[50] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. E. Del Valle, M. Flechl, R. Frühwirth, M. Jeitler, and et al., *Search for production of four top quarks in final states with same-sign or multiple leptons in proton–proton collisions at sqrt(s)=13 tev*, *The European Physical Journal C* **80** (Jan, 2020).

[51] G. Aad, B. Abbott, D. Abbott, A. Abed Abud, K. Abeling, D. Abhayasinghe, S. Abidi, O. AbouZeid, N. Abraham, H. Abramowicz, H. Abreu, Y. Abulaiti, B. Acharya, B. Achkar, L. Adam, C. Bourdarios, L. Adamczyk, L. Adamek, J. Adelman, and L. Zwalinski, *Evidence for $tt^{-}tt^{-}$ production in the multilepton*

final state in proton–proton collisions at s=13 tev with the atlas detector, *The European Physical Journal C* **80** (11, 2020) 1085.

[52] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. Escalante Del Valle, M. Flechl, R. Frühwirth, M. Jeitler, and et al., *Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at sqrt(s)=13 tev*, *Journal of High Energy Physics* **2019** (Nov, 2019).

[53] `https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2`.

[54] `https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2`.

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

[56] `https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC`.

[57] `https://twiki.cern.ch/twiki/bin/view/CMS/JetID`.

[58] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, *Jet Flavour Classification Using DeepJet*, *JINST* **15** (2020), no. 12 P12012, [arXiv:2008.1051].

[59] **CMS** Collaboration, A. M. Sirunyan *et. al.*, *Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at sqrt(s)=13 TeV*, *JHEP* **10** (2017) 005, [arXiv:1707.0331].

[60] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, *Towards an understanding of jet substructure*, *Journal of High Energy Physics* **2013** (Sep, 2013).

[61] T. Cornelis and T. C. Collaboration, *Quark-gluon jet discrimination at cms*, .

[62] A. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, T. Bergauer, M. Dragicevic, J. Erö, A. E. D. Valle, M. Flechl, R. Frühwirth, and et al., *Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques*, *Journal of Instrumentation* **15** (Jun, 2020) P06005–P06005.

[63] S. Sagir *et. al.*, *Search for four-top production in the single-lepton final state using the full run2 dataset*, CMS Analysis Note 2020/020, 2021.

[64] `https://twiki.cern.ch/twiki/bin/view/CMS/MissingETRun2Corrections`.

[65] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, *Catboost: unbiased boosting with categorical features*, 2019.

[66] **CMS** Collaboration, A. M. Sirunyan *et. al.*, *Search for standard model production of four top quarks in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, *Phys.Lett.B* (2017) [11702.06164].

[67] S. Choi *et. al.*, *Improved Extrapolation Methods of Data-driven Background Estimation in High-Energy Physics*, arXiv:1906.1083.

[68] S. Choi *et. al.*, *Data-driven Estimation of Background Distribution through Neural Autoregressive Flows*, arXiv:2008.0363.

[69] M. J. R. B. S. A. S. Arthur Gretton, Karsten M. Borgwardt, *A kernel two-sample test*, *Journal of Machine Learning Research* (2012).

[70] **CMS** Collaboration, *CMS luminosity measurements for the 2016 data taking period*, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[71] **CMS** Collaboration, *CMS luminosity measurement for the 2017 data-taking period at* $\sqrt{s} = 13$ TeV, CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[72] **CMS** Collaboration, *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} = 13$ TeV, CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.

[73] S. Mrenna and P. Skands, *Automated parton-shower variations in pythia 8*, *Physical Review D* **94** (Oct, 2016).

[74] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. D. Roeck, J. Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, and et al., *Pdf4lhc recommendations for lhc run ii*, *Journal of Physics G: Nuclear and Particle Physics* **43** (Jan, 2016) 023001.

[75] A. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. Escalante Del Valle, and et al., *Observation of* $tt\bar{}h$ *production*, *Physical Review Letters* **120** (Jun, 2018).

[76] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *The European Physical Journal C* **71** (Feb, 2011).