Search for heavy Higgs bosons in conjunction with neural-network-driven reconstruction and upgrade of the Fast Beam Condition Monitor at the CMS experiment

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von
Jonas Rübenach

Hamburg
2023

Gutachter/innen der Dissertation:	Prof. Dr. Christian Schwanenberger Prof. Dr. Georg Weiglein
Zusammensetzung der Prüfungskommission:	Prof. Dr. Freya Blekman Prof. Dr. Jochen Liske Prof. Dr. Christian Schwanenberger Dr. habil. Christian Seifert Prof. Dr. Georg Weiglein
Vorsitzende/r der Prüfungskommission:	Prof. Dr. Jochen Liske
Datum der Disputation:	30.03.2023
Vorsitzender Fach-Promotionsausschuss PHYSIK:	Prof. Dr. Günter H. W. Sigl
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Eidesstattliche Versicherung / Declaration on oath

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Hamburg, den 4. Mai 2023

Jonas Ribenach

Unterschrift der Doktorandin / des Doktoranden

Abstract

This thesis presents a search for a heavy scalar and a heavy pseudoscalar Higgs boson decaying into a top quark and a top antiquark. The used data was collected by the CMS experiment at the LHC at an energy of $\sqrt{s} = 13\,\text{TeV}$ during 2016, 2017 and 2018 corresponding to an integrated luminosity of $138\,\text{fb}^{-1}$. The Higgs bosons are electrically neutral, have masses ranging from 400 GeV up to 1000 GeV and the top quarks, into which the Higgs bosons decay, are assumed to further decay dileptonically. The invariant mass of the top quarks together with two spin correlation observables are obtained from the full reconstruction of the two top four-momenta. The reconstruction has a resolution of around 25 % on the invariant mass. The obtained constraints on the coupling strength between the Higgs bosons and the top quarks depend on the Higgs boson mass and exclude coupling modifiers above the range 1.1 to 2.0 for the scalar boson and 0.95 to 1.5 for the pseudoscalar boson.

The limited resolution of the reconstruction is improved by a novel neural-network-based approach fully replacing the classic approach. The network is trained on SM data and specifically crafted data, devoid of assumptions from an underlying model. The network is seen to reach a resolution of around 15 % on the invariant mass, almost twice as good as the classic approach. It is shown that the crafted data training improves performance over a broad mass spectrum, where SM events are not prominent.

Furthermore, the thesis includes a detailed overview of the upgrade and commissioning of the Fast Beam Condition Monitor of the CMS detector, measuring luminosity and beam-induced background. The commissioning was finalized in 2022 together with the startup of the LHC. Its first results are shown. The luminosity measurement is expected to be used in future analyses of many years to follow.

Zusammenfassung

Diese Dissertation stellt eine Suche nach schweren skalaren und schweren pseudoskalaren Higgs-Bosonen, die in ein Top-Quark und ein Top Antiquark zerfallen, vor. Die Daten, die benutzt wurden, wurden vom CMS Experiment am LHC bei einer Energie von $\sqrt{s} = 13\,\mathrm{TeV}$ in den Jahren 2016, 2017 und 2018, was einer integrierten Luminosität von $138\,\mathrm{fb}^{-1}$ entspricht, aufgezeichnet. Die Higgs-Bosonen sind elektrisch neutral, haben Massen zwischen 400 GeV und 1000 GeV und es wird angenommen, dass die Top-Quarks, in die die Higgs-Bosonen zerfallen, weiter dileptonisch zerfallen. Die invariante Masse der Top-Quarks zusammen mit zwei Spinkorrelationsobservablen werden durch eine komplette Rekonstruktion der Top-Viererimpulse erhalten. Die Rekonstruktion hat eine Auflösung von etwa 25 % auf die invariante Masse. Die so erhaltenen Beschränkungen der Modifikation der Kopplungsstärke zwischen den Higgs-Bosonen und den Top-Quarks hängt von der Higgs-Boson-Masse ab und schließen Kopplungsmodifikationen in dem Bereich über 1.1 bis 2.0 für das skalare Boson und über 0.95 bis 1.5 für das pseudoskalare Boson aus.

Die begrenzte Auflösung der Rekonstruktion wird durch einen neuartigen, auf neuronalen Netzen basierenden Ansatz verbessert. Das Netzwerk wird auf SM-Daten und auf speziell angefertigten Daten, die frei von Annahmen zu einem zugrundeliegenden Modell sind, trainiert. Es wird festgestellt, dass das Netzwerk eine Auflösung von etwa 15 % auf die invariante Masse erreicht, was fast eine doppelte Präzision im Vergleich zur herkömmlichen Methode darstellt. Zusätzlich wird gezeigt, dass die angefertigten Daten die Ergebnisse über ein breites Massenspektrum, in welchen SM-Ereignisse nicht stark vorhanden ist, verbessert.

Ferner enthält diese Dissertation eine detaillierte Übersicht des Upgrades und der Inbetriebnahme des Fast Beam Condition Monitors des CMS-Detektors, welcher die Luminosität und den vom Strahl induzierten Untergrund misst. Die Inbetriebnahme wurde 2022 fertiggestellt, zusammen mit dem Anlauf des LHCs. Seine ersten Ergebnisse werden dargestellt. Es wird erwartet, dass die Luminositätsmessung in vielen zukünftigen Analysen der nächsten Jahre benutzt werden wird.

Contents

Chapter 1

Introduction

Questioning the underlying concepts of the universe is one of the key principles to gain knowledge and to open gateways to discoveries. The field of particle physics brings this to an extreme by seeking to study nature at its most fundamental scales. Particle physics understands matter and antimatter as fundamental particles and explains physics phenomena as forces between them. At the current state of research, the knowledge has been summarized within the standard model of particle physics, which has been proven to give astonishingly accurate and rigid predictions of experimental observations. Despite its performance, however, the model leaves many questions unanswered. One of these revolves around dark matter. Although its existence would explain a wide range of astrophysical observations, its nature remains a mystery and cannot be explained within the SM of particle physics. Therefore, various extensions of the SM with new particles have been proposed. A central building block of these theories are additional Higgs bosons, for example, as mediators that couple to both, known elementary particles and dark matter particles. The Higgs bosons come with masses possibly much higher than any of the known fundamental particles and they are thus referred to as "heavy". This thesis presents a search for heavy Higgs bosons.

To infer new knowledge, particle physics obtains observations from collisions of particles at high energies. These collisions occur in nature but can also be created artificially using particle accelerators to enable the study in laboratory conditions. The largest and highest-energy accelerator that has been built up to this date is the Large Hadron Collider (LHC). One of the two general purpose experiments that are receiving collisions from the LHC is the Compact Muon Solenoid (CMS). This thesis focuses on work both improving the performance of the detector itself and using its data of proton-proton collisions at a center-of-mass energy of 13 TeV.

One of the most fundamental quantities for a collider the like the LHC, that is measured by its experiments, such as the CMS experiment, is the luminosity. The luminosity describes the number of collisions that are observed and is a crucial input for measurements performed using the experimental data, including the search presented in this thesis. To provide a luminosity measurement the CMS detector possesses the Fast Beam Condition Monitor (BCM1F). Besides luminosity, BCM1F also monitors the particle beams of LHC for abnormalities. This thesis gives a detailed overview of the steps and studies that were performed to assemble and commission the subdetector, when it was upgraded to a design fully comprising of silicon sensors.

The search for heavy Higgs bosons in this thesis is performed by analyzing the combined data of the CMS experiment from the years 2016, 2017 and 2018 using the latest calibrations available. Specifically, the search targets Higgs bosons decaying into a top quark and a top antiquark (top pair). A standard model prediction is created with the currently highest available precision and systematic uncertainties are carefully evaluated. The search is facilitated by two different sources of information: One being the invariant mass of the top pair, which reflects the boson's high mass. The other source being correlations among the spins of the two top quarks, which depict properties intrinsic to the Higgs bosons that are searched for. These observables are reconstructed for every event based on the decay products of the top pair. The decays of the top pair that are studied involve four leptons, of which two are neutrinos, and 2 jets, with at least one originating from a bottom quark or antiquark. The neutrinos leave the experiment undetected, which poses an additional challenge to the reconstruction done in the analysis.

To improve the performance of the necessary reconstruction in the analysis, a deep neural network is designed and evaluated in detail. The neural network can fully reconstruct the four-momenta of the two top quarks the heavy Higgs bosons decay into. Moreover, the neural network is highly adaptable and can be exploited outside the analysis also on similar searches for BSM phenomena or for precision measurement of the standard model.

Chapter 2

Theory framework

2.1 Units used in particle physics

A common quantity in particle physics is the cross section, which is normally given in units of barn (symbol b). It has the dimension of area and typically one encounters the unit together with the prefixes femto or pico. With the exception of barns or if SI units are explicitly mentioned, this thesis uses natural units. The natural units are obtained by setting $c = \epsilon_0 = \hbar = 1$, where c is the speed of light, ϵ_0 is the vacuum permittivity and $\hbar$ is the reduced Planck constant. In natural units several quantities, such as mass and energy, use the same dimension. Masses and energies are usually given in units of GeV or other orders of electron volts. Other quantities might have units of GeV^{-1} or even become dimensionless. Table 2.1 shows the conversion factors for common quantities.

	SI	Natural units
Mass	$1\,\mathrm{kg}$	$5.6 \times 10^{26}\,\mathrm{GeV}$
	$1.8 \times 10^{-27}\,\mathrm{kg}$	$1\,\mathrm{GeV}$
Energy	$1\,\mathrm{J}$	$6.2 \times 10^{9}\,\mathrm{GeV}$
	$1.6 \times 10^{-10}\,\mathrm{J}$	$1\,\mathrm{GeV}$
Momentum	$1\,\mathrm{N\,s}$	$1.9 \times 10^{18}\,\mathrm{GeV}$
	$5.3 \times 10^{-19}\,\mathrm{N\,s}$	$1\,\mathrm{GeV}$
Elementary charge	$1.6 \times 10^{-19}\,\mathrm{C}$	0.30
Cross section	$1\,\mathrm{pb} = 10^{-40}\,\mathrm{m}^2$	$2.6 \times 10^{-9}\,\mathrm{GeV}^{-2}$
	$3.9 \times 10^{8}\,\mathrm{pb} = 3.9 \times 10^{-32}\,\mathrm{m}^2$	$1\,\mathrm{GeV}^{-2}$

Table 2.1: Conversion table between SI units and natural units.

2.2 Content of the standard model

The standard model (SM) of particle physics is a combination of quantum field theories describing physics at the most fundamental level observable to this day. In a quantum field theory physical phenomena are interpreted by the use of fields and interactions between these. The interactions can change the number of particles, which are described as excited states of their own specific type of field. One differentiates between three types of interactions, each arising from their own theory: the electromagnetic (EM) interaction, the strong interaction and the weak interaction. Not all fields take

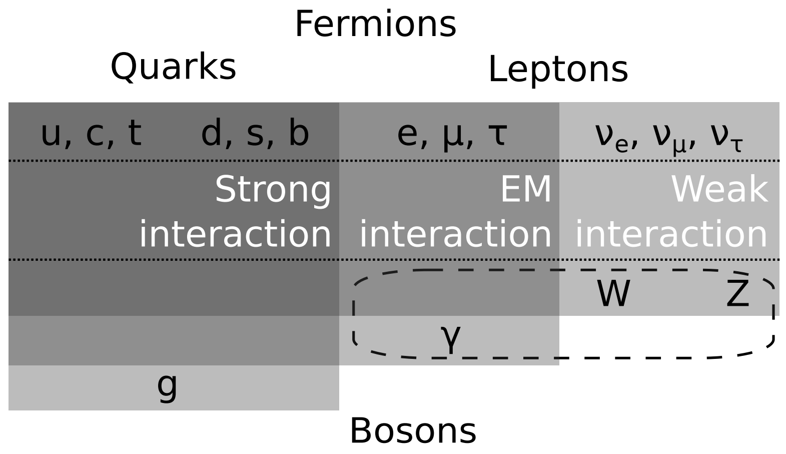


Figure 2.1: Fundamental SM interactions and the particles that participate in them. The strong interaction has the least number of different particles, only including the quarks and the gluon. On the other hand the weak interaction counts the most particles, with all fermions, the W and Z bosons. Additionally the γ, W and Z can interact, as they are all part of what is called electroweak interaction.

part in all of these interactions, as seen in figure 2.1, resulting in a variety of different behaviors for the particles. Furthermore particles are distinguished from another by a range of quantities, including their spin, charge and other quantum numbers. In table 2.2 the particles of the SM are summarized together with some of their associated quantum numbers.

The particles generally are divided into two categories using the magnitude of their spin: the fermions with a spin of 1/2 and the bosons with an integer spin. There are two types of fermions, quarks and leptons, with the difference that only the quarks take part in the strong interaction. Leptons are further divided into electrically charged ones and neutral ones (neutrinos). As the latter are electrically neutral, they also do not participate in the EM interaction. Similarly quarks can be considered as two groups: up-type quarks (u, c, t) and down-type quarks (d, s, b). The groups separated by their difference in electric charge and weak isospin. Each of these two groups then have three different flavors of quarks, with each flavor having a different mass. Furthermore, one can categorize the fermions into three generations. Each of the neutrinos is placed in a different generation, but together with its corresponding charged lepton partner. For the quarks, the generations are built from the pairs ud, cs and tb. The bosons with spin-1 on the other hand are each connected to one fundamental interaction, in which they participate. The bosons of the EM and weak interactions, namely the photon, W and Z boson, are able to interact with each other, while the gluon does not interact with the other spin-1 bosons. The only remaining particle, the Higgs boson, is a spin-0 particle and its field are associated to the mechanism that grants mass to the particles.

For the particles in the SM as shown in table 2.2 one can flip the sign of their charges, resulting in the corresponding antiparticle. Excluding the difference in charge, an antiparticle possesses the same properties as its particle partner, because all the other quantities, such as mass, are kept the same.

2.3 Handedness of particles: Helicity, Chirality and Parity

Elementary particles come with an intrinsic property called spin $\vec{S}$. Its magnitude is listed in table 2.2. For photons for example the spin relates to the classical phenomenon called polarization of light. On the other hand for fermions this property is more subtle. Fermions are described by the Dirac equation, which gives rise to four component functions, called spinors, as solutions. The spin

Category		Name	Symbol	Spin	Mass/GeV	Electric charge	Weak isopsin	PDG ID
Fermion	Quark	Down	d	1/2	4.67×10^{-3}	$-1/3$	$-1/2$	1
		Up	u		2.16×10^{-3}	$2/3$	$1/2$	2
		Strange	s		9.3×10^{-2}	$-1/3$	$-1/2$	3
		Charm	c		1.27	$2/3$	$1/2$	4
		Bottom	b		4.18	$-1/3$	$-1/2$	5
		Top	t		173	$2/3$	$1/2$	6
	Lepton	Electron	e^-		5.11×10^{-4}	-1	$-1/2$	11
		Electron neutrino	ν_e		0	0	$1/2$	12
		Muon	μ^-		0.106	-1	$-1/2$	13
		Muon neutrino	ν_μ		0	0	$1/2$	14
		Tau	τ^-		1.78	-1	$-1/2$	15
		Tau neutrino	ν_τ		0	0	$1/2$	16
Boson		Gluon	g	1	0	0	0	21
		Photon	γ		0	0	0	22
		Z	Z		91.2	0	0	23
		W	W^+		80.4	1	1	24
		Higgs	h	0	125	0	$-1/2$	25

Table 2.2: Particles in the Standard Model. Neutrinos are assumed to be massless. The column "weak isospin" denotes the third component of the isospin and indicates the value for the particles of left-handed chirality. Antiparticles, right-handed particles and different color states are omitted. Masses from reference [1].

operator and the Hamiltonian of the fermions do not commute. For this reason it is more convenient to look at the helicity h, whose operator commutes with the Hamiltonian and which is defined by

$$h = \frac{\vec{S} \cdot \vec{p}}{|\vec{p}|}, \tag{2.1}$$

with the momentum of the particle $\vec{p}$. As fermions are spin-1/2 particles, their helicities only have two values, one positive and one negative, just like the third component their spin. Fermions with a positive helicity are referred to as right-handed, otherwise they are left-handed.

A quantity closely related to the helicity is the chirality. Acting on the four-vectors of the spinors its operator can be defined by

$$\gamma^5 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \tag{2.2}$$

Similar to the helicity, the chirality only has 2 distinct eigenvalues, one positive and one negative, and again a right-handed chirality refers to the positive eigenvalues, while the a left-handed chirality refers to the negative one. The concept of chirality plays an important role for the weak interaction, as W bosons only interact with left-handed fermions and right-handed antifermions (see section 2.5.3). Helicity eigenstates can be written as a linear combination of chirality eigenstates. For example for the right-handed (helicity) fermion $u_\uparrow$ with momentum p, energy E and mass m and the chirality eigenstates fermions u_R and u_L (right- and left-handed, respectively):

$$u_\uparrow \propto (1 + \frac{p}{E+m})u_\mathrm{R} + (1 - \frac{p}{E+m})u_\mathrm{L}. \tag{2.3}$$

One can see that in the case of negligibly small mass ($E \gg m$), $u_\uparrow \propto u_\mathrm{R}$. Similar holds true for the other eigenstates, meaning in this limit helicity and chirality become equivalent.

Yet another quantity, that is defined by the either positive or negative eigenvalues of an operator, is the parity. The parity operator mirrors a system, meaning the spacial coordinates are changed from x to $-x$. It can be shown that fermions at rest have a positive parity, while antifermions at rest have a negative one. Bosons on the other hand cannot in general be associated with a parity. An exception is the SM Higgs boson, which has a positive parity. This is part of the reason why it is referred to as "scalar boson". One can also create models with Higgs bosons that have negative parity (see section 2.8), which then would be pseudoscalar bosons. Parity is conserved in many processes, similarly to how one would not expect physics to change from looking into a mirror. However, the weak interaction does not conserve parity (P violation) in general. Even more, the weak interaction does not even conserve the product of the charge conjugation (replacing every particle with its antiparticle) and the parity transformation, which is called CP violation.

2.4 Interaction rate and cross section

According to Fermi's golden rule, the rate Γ_{fi} at which a process from an initial state $|i\rangle$ to some final state $|f\rangle$ is occurring is

$$\Gamma_{fi} = 2\pi \left|T_{fi}\right|^2 \rho(E_i) \tag{2.4}$$

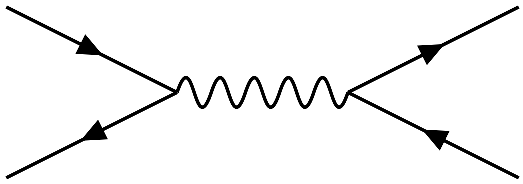


Figure 2.2: A Feynman diagram of two incoming particles, one intermediate particle and two outgoing particles, as often occurring in LO EM processes.

where $\rho(E_i)$ is the density of states at the initial energy E_i and $T_{fi} \propto \mathcal{M}_{fi}$, with the matrix element (ME) $\mathcal{M}_{fi}$ (and omitting a normalization factor) [2]. Utilizing perturbation theory the ME can be approximated via a series, assuming the magnitude of the terms decrease with their order:

$$T_{fi} = \langle f| H |i\rangle + \sum_{j\neq i} \frac{\langle f| H |j\rangle \langle j| H |i\rangle}{E_i - E_j} + \dots, \tag{2.5}$$

with the time-dependent interaction Hamiltonian H and energies E_j. If only the first term of equation 2.5 is used in an computation, it is referred to as leading order (LO). A computation including also the following term is called next to leading order (NLO). For increasing precision, even more terms can be added. The NLO term and further terms correspond to the process occurring via an intermediate state ($|j\rangle$ in the case of equation 2.5).

A rather convenient method to denote the long expressions that emerge when computing $\langle f| H |i\rangle$ and its higher orders, is the usage of Feynman diagrams. Fundamentally these are graphs, in which every vertex and edge corresponds to a factor that is part of the expression. At the same time, the diagrams provide a quick visualization of what can be understood to happen inside the process in terms of time and what particles are participating. By convention the time axis goes from left to right. Arrows of initial state fermions pointing away from a vertex indicate antiparticles, while arrows pointing towards a vertex indicate particles. The opposite is true for final state particles. Finally arrows on internal edges appear in the direction of the flow dictated by the initial and final state arrows. To put it simply: Arrows pointing to the left indicate fermions, while arrows pointing to the right indicate antifermions. A tilde-like line is used for photons, W and Z bosons, a spring-like line for gluons and a dashed line indicates Higgs bosons. An example diagram for an LO 2 to 2 particle process can be seen in figure 2.2. Internal edges (lines that are connected on both ends to other lines) represent virtual particles. These particles arise from the perturbation theory and can possibly be "off shell", meaning they violate the relationship $m^2 = E^2 - p^2$.

In general every vertex in a Feynman diagram introduces a factor that includes a coupling constant g, which is not fixed by theory and needs to be measured. Each of the fundamental interactions is associated with their own coupling constant. With increasing number of vertices at higher orders (the process includes one ore more intermediate states), the powers at which the coupling constants appear increases as well. Thus the series expansion given above for T_{fi} is an expansion in powers of g. For example the LO process of figure 2.2 will evaluate to an ME with g^2, while anything more complex will produce g^3 or higher. While g is a scalar and does not depend on individual particle momenta, the name "constant" still is misleading, as it depends on the energy it is measured at. For this reason it is often described as "running" coupling constant.

The rate Γ_{fi} is closely related to the cross section σ. While Γ_{fi} is still dependent on quantities like the flux of the incoming particles of the process, the cross section describes the underlying probability of the interaction. In a scattering process with two incoming particles with velocities v_a and v_b, the cross section is given by

$$\sigma = \frac{\Gamma_{fi}}{v_a + v_b}. \tag{2.6}$$

Note that here, relativistic factors are kept within Γ_{fi} [2].

An important observable is the so called decay width, which represents the decay rate of a specific particle. The total decay width of the particle is a sum of the (partial) decay widths of all its possible decay processes. The ratio of a partial decay width to the total one can be used to quantize the probability for one particular decay to happen and is called branching ratio (BR).

2.5 The fundamental interactions and bosons in the SM

Quantum field theories are generally formulated in terms of a Lagrangian density $\mathcal{L}$ (or simply "Lagrangian"). The Lagrangian plays an important role, for example in the Euler-Lagrange equations, which are used to describe the motion of a system. The Lagrangian can also serve as the foundation to derive the Hamiltonian H, that is part the computation of rates and cross sections. If the Lagrangian is invariant under a transformation, there is a quantity that stays constant during the process the Lagrangian describes. The existence of these conserved quantities, such as energy or momentum, is a result of Noether's Theorem. There is a range of conservation laws known to the SM, including the ones for charges, lepton numbers or baryon numbers, which all restrict the processes that are allowed.

2.5.1 Electromagnetic interaction

The EM interaction in the SM is described by quantum electrodynamics (QED) and is the underlying principle of electromagnetism. Its Lagrangian is required to be invariant under the local phase transformations of $U(1)$ (called the gauge transformation of the interaction). This means that multiplying a fermion wave function by a factor of

$$\exp\left(i\chi(x)\right) \tag{2.7}$$

does not change the Lagrangian. Here $\chi(x)$ is a position-dependent (hence "local") real value. This is accomplished by having the Lagrangian include a field (the EM field), which shows the same transformation property known from the classical EM field and its gauge freedom. This EM field then gives rise to the photon. Using this symmetry, Noether's theorem implies the existence of a conserved current, which translates to the continuity equation and the conservation of electrical charge known from classical electromagnetism. Due to the definition of the Lagrangian, the coupling constant of QED is made from the electric charge of the participating fermion. The charge of charged elementary particles is always either one, two or three thirds of the elementary charge e, as seen in table 2.2. As a conclusion one associates the fine structure constant

$$\alpha = \frac{e^2}{4\pi} \tag{2.8}$$

with the EM coupling constant, which is around $1/137$ at low energies [3] and increases at higher energies.

2.5.2 Strong interaction

Quantum chromodynamics (QCD) is the underlying theory of the strong interaction. The formalism of QCD is very similar to that of QED, with the difference that the gauge transformation here is $SU(3)$ instead of $U(1)$. Practically speaking this transformation give rise to factors in front of the quark wave functions of the form of

$$\exp(i\vec{\chi}(x) \cdot \hat{T}) \tag{2.9}$$

$\vec{\chi}(x)$ is an eight-dimensional vector of position-dependent real values and $\hat{T}$ is a vector of the eight Gell-Mann matrices. The Gell-Mann matrices are a set of eight linearly independent 3×3 traceless Hermitian matrices and span the Lie algebra of the $SU(3)$ group. QCD introduces eight new fields, which produce eight new particles. These are collectively referred to as gluons. In contrast to $U(1)$ from QED, $SU(3)$ acts on a three-dimensional vector space and because of this each quark takes part in the interaction as a mixture of three independent spinors. These three spinors are identified by three different types of color charges, named by the colors red, green and blue. In the case of antiquarks, which as usual have the opposite charge, the color charges are named antired, antigreen and antiblue. A system made from a particle and antiparticle charged with a color and its corresponding anticolor form a color neutral system in QCD. Additionally, three particles, each with a different color charge, also form a neutral system. For this reason the analogous naming scheme to colored light was chosen: Adding the three primary colors will produce white light. As gluons themselves carry color charge (in fact they carry a color and different anticolor), the emission of a gluon enables the transition from one color to another.

In QCD the coupling constant is expressed in terms of α_S. In great contrast to α from QED, it decreases with higher energies, while it drastically grows for lower energies. This behavior is shown in figure 2.4. Therefore and because terms of higher order have higher powers of α_S, for low energies any perturbation approach will fail. This energy regime is referred to as non-perturbative QCD. At high energies however, the coupling becomes weaker and weaker, leading to the so called asymptotic freedom, in which perturbation theory becomes applicable. The coupling α_S given at the mass of the Z boson is $\alpha_S = 0.1179 \pm 0.0010$, according to current measurements [1].

One consequence of the increasingly large coupling at low energies is the color confinement. The confinement states that particles that are non-neutral color-wise cannot be observed isolated, they are confined together with other color-charged particles to form neutral states. The confinement can also be seen from the form of the QCD potential. At close distances it has similar form as the Coulomb potential of QED, as it can be approximated with $V \propto 1/r$ [4]. A different picture is seen at increasing distances. With color confinement, already at distances above one fm pulling apart color charges requires enough energy to create new particles. The production of additional particles and creation of new bound states is called hadronization. At large distances it can be explained by adding a term proportional to the distance to the potential $V_{\text{cf}} \propto r$ [2].

Bound states of quarks are called hadrons. One differentiates between two types of hadrons: mesons (integer spin) and baryons (spin-1/2). While higher number of quarks are possible, mesons normally refer to an quark-antiquark particle, like the pion ($\text{u}\bar{\text{d}}$). Baryons refer to three quarks, like the well known proton (uud). However, the internal structure of these bound states turns out to be much more complex, thanks to the contribution from virtual particles. In figure 2.3 a Feynman diagram of the production of such virtual particles inside a pion can be seen. At high energies, which are reached for example at the LHC, these virtual particles can become the leading contribution, for

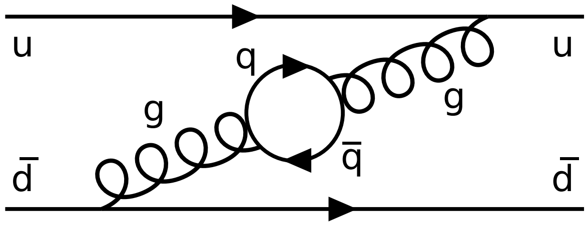


Figure 2.3: Feynman diagram of a pion (ud̄) with virtual gluons (g) and quarks (q).

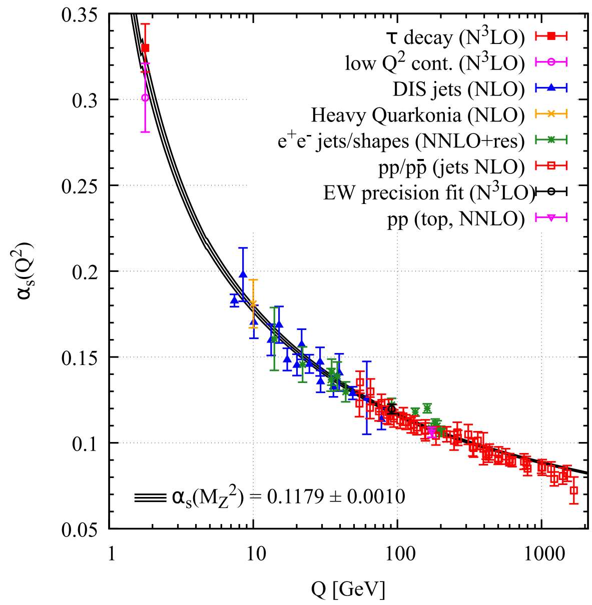


Figure 2.4: Dependence of the strong coupling α_S on the energy Q from many measurements. From [1].

example for the top quark pair production.

2.5.3 Weak interaction

The electroweak (EW) theory is another part of the SM. While it describes the weak interaction, it simultaneously provides a model for the EM interaction, thereby unifying EM and weak interaction into one theory. This is done by replacing the $U(1)$ invariance from QED with a new field, that is also invariant under $U(1)$, now denoted as $U(1)_Y$. Additionally a new invariance under the transformations of $SU(2)$ is added. This requires the addition of three new fields, amounting to a total of four fields. It turns out that the observed bosons W^+, W^- are linear combinations of two of the fields, while the Z and γ bosons are linear combinations of the other two.

As the $SU(2)$ transformations act on a two-dimensional vector space, the weak interaction acts on two-component vectors of spinors, called isospin doublets. This is similarly to QCD, where there were three components. Instead of a charge, the weak interaction is connected to a quantum number called isospin. The doublets combine different leptons or quarks: For example a doublet can be made up from a neutrino and its corresponding lepton partner or two quarks of the same generation. This enables the transition between up and down type or changes between neutrino and charged lepton via an interaction with a W boson.

There are two more refinements to the doublets. Firstly, they only contain the left-handed chirality states of the particle wave function (right-handed in the case of antiparticles). As a consequence right-handed chirality particles (and left-handed antiparticles) are not part of the weak interactions with the W boson and have a zero isospin. For this reason also the gauge transformation does not act on right-handed particles, hence it is often denoted as $SU(2)_L$. The second refinement is the linear combination of different flavors and is done because the W boson is seen to be able to change fermion flavor. The refinement implemented, by convention, on the down-type part of the doublet and means for example that the doublet in place of the down quark contains $d' = V_{\text{ud}}d + V_{\text{us}}s + V_{\text{ub}}t$, where V_{ij} are the elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. In the SM, this CKM matrix

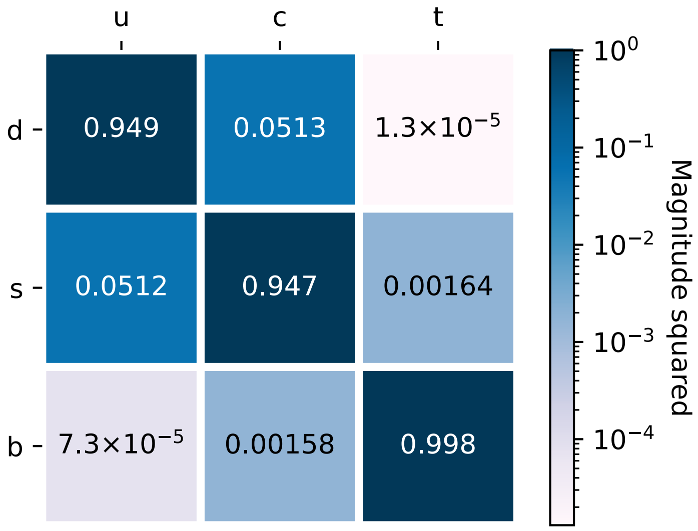


Figure 2.5: Magnitude squared of the elements of the CKM matrix. The larger the value, the more likely a transition between the flavors in the weak interaction is. Parameterization for the matrix taken from [1].

is unitary and its magnitude squared qualifies how likely a transition between the flavors is. The matrix is illustrated in figure 2.5 and its values have to be measured experimentally. Because of the structure of the matrix a decay of a top quark directly to a down quark via the emission of a W^+ boson is possible, albeit very rare.

One can understand the weak isospin as an analogy to the charges known from the other interactions. For example a top quark (third component of the weak isospin $T_3 = 1/2$) can emit a W^+ boson ($T_3 = 1$) to decay into a bottom quark ($T_3 = -1/2$). However, this idea is not as straight forward as in the case of the other interactions, as the EW theory is subject to the symmetry breaking by the Higgs mechanism, which allows processes in which the weak isospin is not conserved.

2.5.4 The Higgs field

So far all the particles as they are given by the Lagrangian would be massless. However, this is not what is observed in nature. For this reason there is another addition to the SM, which is the Higgs mechanism and a corresponding field. The field is made from a doublet ϕ with complex entries, affecting the $SU(2)_L$ symmetry of the weak interaction. The potential corresponding to the field is chosen so that it does not have a minimum at $\phi = 0$, but rather at some value $v \neq 0$, which is called the vacuum expectation value (VEV). This is a unique property of the Higgs field. As a consequence, the Lagrangian stops being invariant under the transformation of the electroweak theory, which is generally referred to as spontaneous symmetry breaking. Additionally it adds the necessary terms to it for the Z and W bosons to have a non-zero mass. Finally the field adds an additional particle, the Higgs boson, which then is the only elementary spin-0 particle in the SM.

After introducing the Higgs field with a non-zero VEV, not only the Z and W bosons gain a masses, also the fermions are no longer massless. This mechanism granting mass to the leptons works the form of a so-called Yukawa interaction between the Higgs boson and the leptons. Like the other interaction this interaction comes with its own coupling, called the Yukawa coupling. Interestingly, in place where the other couplings have charges, this Yukawa coupling depends on the masses of the particles the Higgs interacts with. For example, for the top quark, the coupling is

$$g_{\mathrm{ht}} = \sqrt{2}\frac{m_{\mathrm{t}}}{v}, \tag{2.10}$$

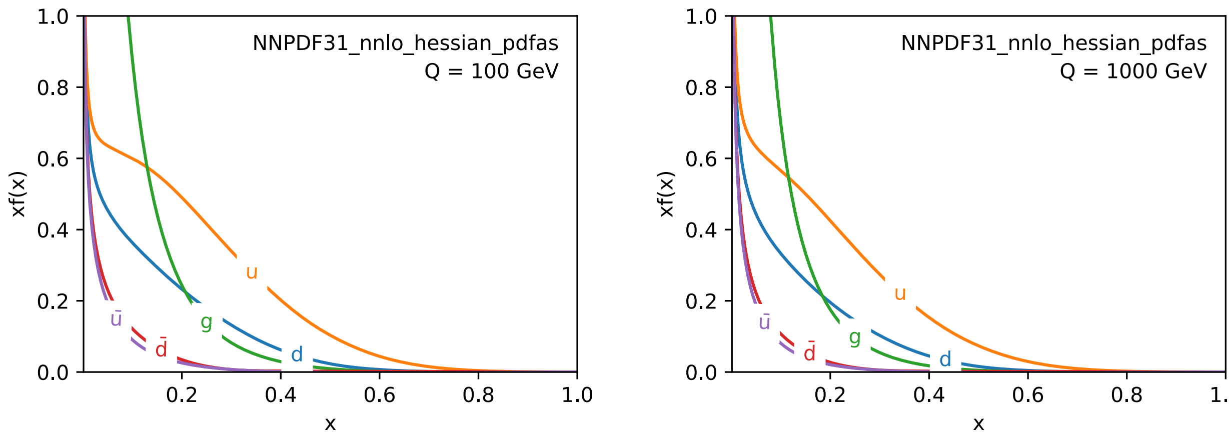
with the top quark mass m_{t}. Its value is surprisingly close to one 1, which has also been confirmed by measurements [5].

2.6 Prediction of proton-proton collisions

Collisions at particle accelerators like the LHC are made using hadrons as incoming particles, rather than bare elementary particles. As seen in section 2.5.2 the content of a hadron is rather complex: In addition to the quarks forming the bound state (valence quarks) there are virtual particles (sea quarks, gluons and other particles). Each of these so called partons carry a fraction of the total momentum of the hadron. For this reason the cross section for a process from two hadrons h_1 and h_2 to some final state X is given by [3]

$$\mathrm{d}\sigma(\mathrm{h}_1\mathrm{h}_2 \to X) = \sum_{a,b} \int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b \mathrm{d}\hat{\sigma}(ab \to X) f_a^{\mathrm{h}_1}(x_a) f_b^{\mathrm{h}_2}(x_b) \tag{2.11}$$

where $\sum_{a,b}$ is a sum over all possible partons inside h_1 and h_2, x_a and x_b are the ratios of the momentum carried by the partons a and b to the total momentum of h_1 and h_2, $\hat{\sigma}(ab \to X)$ is the partonic cross section, which depends on the ME, and $f_a^{\mathrm{h}_1}$ and $f_b^{\mathrm{h}_2}$ are the parton distribution functions (PDF). The PDF describes how likely it is to encounter a parton inside the hadron at a specific momentum fraction. They are derived from measurements. While not explicitly written in the formula, they also

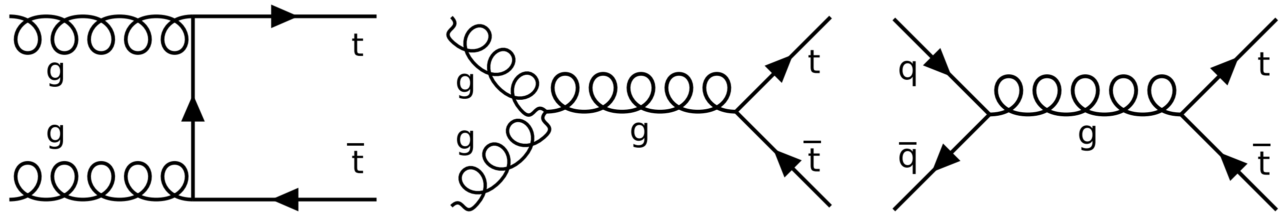


(a) (b)

Figure 2.6: Parton density functions for different quarks and for the gluon inside a proton at energies of 100 GeV and 1 TeV. Using NNPDF3.1 [6] and LHAPDF [7].

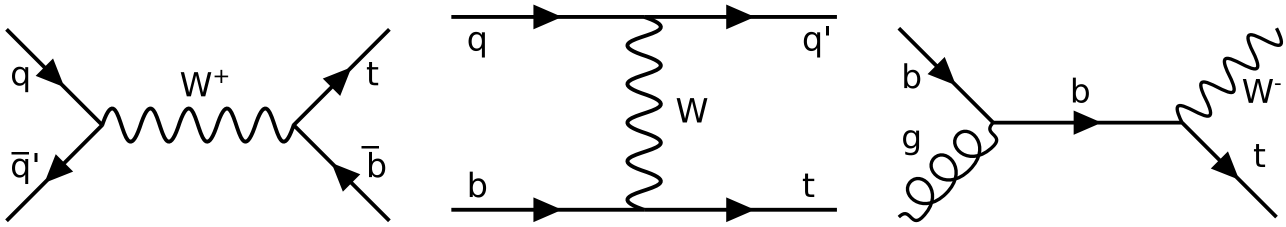
depend on the energy of the process Q. In figure 2.6 one can see that the up quark, which here is the sum of valence and sea contributions, inside the proton has the highest probabilities at higher momentum fractions. However, at low fractions, the gluon PDF sharply increases, vastly outweighing the contributions of the other partons.

After the initial process of the colliding partons, a long cascade of further lower energy ("softer")



(a) (b) (c)

Figure 2.7: Feynman diagrams for top pair production at LO for incoming gluons (a, b) and incoming quark/antiquark (c).



(a) (b) (c)

Figure 2.8: Feynman diagrams for single top production at LO. While the vertex including a charm or an up quark plus a top quark is possible, it is neglected here due to the very large CKM Matrix element of $V_{\mathrm{tb}} \approx 1$. (a) is often referred to as s-channel process, while (b) is referred to as t-channel process.

processes starts off. The simulation of this process is done by showering algorithms. The showering algorithm adds gluon radiation to the outgoing particles of the ME (final state radiation, FSR) and may also add them to the incoming particles (initial state radiation, ISR). The processes the algorithm simulates are referred to as parton shower. This is followed by particle decays as well as the formation of new hadrons (hadronization) from radiated quarks and gluons. These processes already start to take place long before the particles reach any instrumentation. The hadrons coming from the same initial particle will arrive in a detector as a jet, which is a multitude of particles, all arriving within a small solid angle.

Thus the computation of a collision is split into three parts, one that is described by the PDF, one that is described by the ME and one produced by soft processes. At orders above LO, the perturbation computation depends on two energies, referred to as factorization scale and renormalization scale. The factorization scale appeared in this paragraph within the PDF as Q and described the energy at which the hadron is being probed [8]. The renormalization scale is part of the mechanism to remove divergent terms that otherwise appear in the perturbation and is typically also set to Q.

2.7 Top Quark

With a mass of 172.76(30) GeV [1] the top quark is the heaviest elementary particle in the SM, making it an interesting handle to probe physics beyond the SM (BSM). The LHC provides a large variety of different top quark production processes. The main processes can be separated into production of a top quark plus antiquark pair (top pair) and into the production of a top (anti-)quark without additional top (anti-)quarks (single top). LO Feynman diagrams can be seen in figures 2.7 and 2.8.

With a measured value of $\sigma = 0.23\,\mathrm{nb}$ at the LHC with an energy of $13\,\mathrm{GeV}$, the t-channel has the highest cross section of the single top production channels [9]. Compared to that, the top pair production is more common by a factor of around 3.5 with a cross section of around 0.8 nb [9]. At LO the pair production primarily involves QCD interactions and due to the strong presence of gluons in the PDFs (see figure 2.6), the processes initiated by a pair of gluons dominate the production. The single top production at LO involves at least one vertex with a W boson and a top and a bottom quark. Due to the large CKM matrix element V_{tb} of almost 1 (see figure 2.5), this implies the presence of a bottom quark or antiquark in the initial or final state.

The decay width $\Gamma(\text{t} \to \text{qW}^+)$ at LO shows a top mass dependence of m_t^3 [2] and as a consequence, different from all other quarks, the lifetime of $1/\Gamma(\text{t} \to \text{bW}^+) \approx 5 \times 10^{-25}$ s is much shorter than what would be needed to observe bound states and hadronization of tops. Similar to the single top production, the fact that V_{tb} is almost equal to 1 means that the top quark practically always decays into a bottom quark and a W^+ boson, strongly suppressing decays into other down-type quarks. The W^+ boson in turn can either decay into a positively charged lepton and a neutrino (leptonic decay) or into a quark and an antiquark (hadronic decay). The corresponding Feynman diagrams can be seen in figure 2.9. Due to the three different possible color charges of the quarks, each of the decays of one

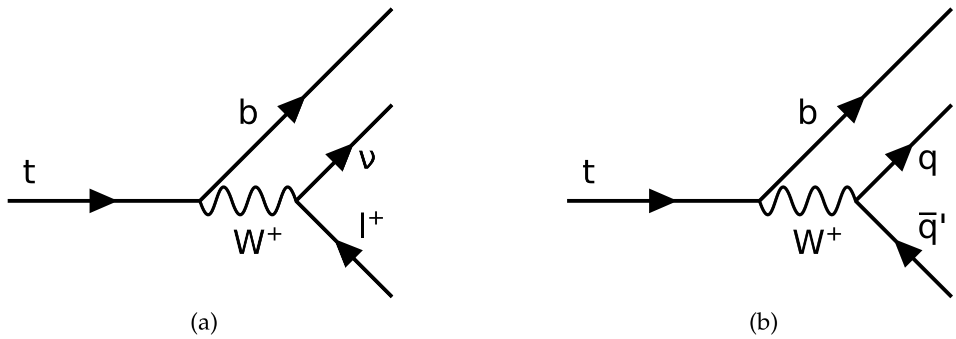


Figure 2.9: Feynman diagrams for the decay of the top quark. The top decays into a bottom quark and a W boson, which in turns decays either leptonically (a) or hadronically (b). Here, l^+ is a positively charged lepton, q is an up-type quark and $\bar{\text{q}}'$ is a down-type antiquark.

specific quark flavor is roughly three times as likely as the decay into a specific lepton flavor. Hence the branching ratio $BR(\text{t} \to \text{be}^+\nu_\text{e}) \approx 1/9$. In the case top quark pair production, one obtains three different cases: Both particles decaying hadronically (full hadronic decay), both particles decaying leptonically (dileptonic) or one hadronically and one leptonically (semileptonic).

The top quark is very well suited to get a handle on the spin distributions, thanks to its immediate decay via W^+ bosons. To explain this, one can look at the system in which the top is at rest and define the z axis to be in the direction of the bottom quark of the decay. Because the weak interaction only involves the fermions with left-handed chirality and because the mass of the bottom quark is relatively low ($m_\text{b} \ll m_\text{t}$), the decay will strongly favor left-handed helicity bottom quarks. As a consequence, the only non-zero ME at LO for a top quark with spin $S_z = -1/2$ involves longitudinally polarized W bosons, while for one with $S_z = +1/2$ the W boson involved is left-handed [2, 10]. Moreover, the decay of the W boson again defines preferred flight directions for the decay product depending on the polarization of the boson. In conclusion, the charged lepton will prefer to follow the z direction, in which the spin of the top quark pointed. Thus the angle distributions of the leptons will in turn indicate the spin distributions of the top quarks. This property is often indicated by the spin analyzing power of the leptons, which is maximal (equal to 1 at LO) [11]. By studying the top spin, one can obtain information about the production process, for example the processes shown in the following section.

2.8 Beyond the Standard Model: Additional Higgs Bosons

A favorable possible extension of the SM is the addition of further Higgs bosons. Additional Higgs bosons are key of several BSM theories, such as supersymmetric ones, especially the Minimal Supersymmetric Standard Model (MSSM) [12] or the Two Higgs Doublet Model (2HDM) [13]. Assuming no CP violation, both the MSSM and 2HDM have two vacuum expectation values and 5 Higgs bosons in total. One of them is associated to the Higgs boson known already from the SM. The different particles are summarized in table 2.3. As a consequence the remaining four Higgs bosons have higher

Symbol	Spin	Mass/GeV	Electric charge	Parity
h		125	0	+1
H		?	0	+1
A	0	?	0	-1
H^+		?	1	+1
H^-		?	-1	+1

Table 2.3: Higgs bosons in MSSM and 2HDM. The lowest mass one is associated to the Higgs boson known from the SM.

masses than the 125 GeV SM h and thus are referred to as heavy Higgs bosons. However, similar to how the mass of the SM h is not given by theory, the masses of the new bosons also are unknown and need to be fixed by an observation. Two of the heavy ones are electrically neutral, while the third is positively and the fourth negatively charged. One of the neutral ones, A, has the particularity of being a pseudoscalar boson. The other four all are scalar bosons. In the following, concepts of the two neutral heavy ones A and H will be discussed in more detail.

In all its generality, 2HDM allows flavor-changing neutral currents. These currents are suppressed in the SM and measurements agree with the SM prediction [14]. By constraining 2HDM to exclude the currents, one can in total obtain four different types of models. The models differ in which of the two Higgs doublets couple to which type of quark. By convention, the up-type quarks always couple to the doublet indexed by the number 2.

An often quoted free parameter in the MSSM and in the 2HDM is the ratio of the two VEVs, using an angle β:

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

where v_2 is the VEV of the doublet coupling to up-type quarks. As physical bosons are linear combinations of both Higgs fields, there is an inherent dependence on $\tan\beta$. For example the Yukawa coupling of the A to the up-type quarks is

$$g_{\mathrm{Au}} = \frac{1}{\tan\beta}. \tag{2.13}$$

Due to the high mass of the top quarks, the coupling between them and the Higgs bosons is likewise large. Thus at a collider like the LHC the production of neutral Higgs bosons via virtual top quarks could possibly be a comparatively common one. Now looking at the decay of the SM Higgs boson, producing a top pair is not possible, because the mass of h is so much lower than the required mass, which is twice the top mass. However, this changes with the introduction of the heavy Higgs bosons: Assuming a mass above roughly 350 GeV, this decay into a top pair becomes a highly likely one. A Feynman diagram for the combination of this production and decay can be seen in figure 2.10. Hav-

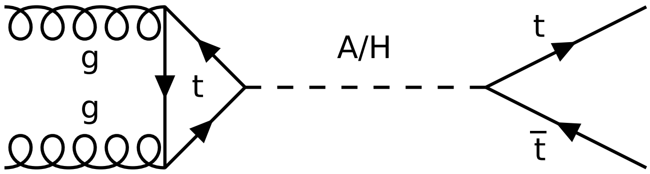
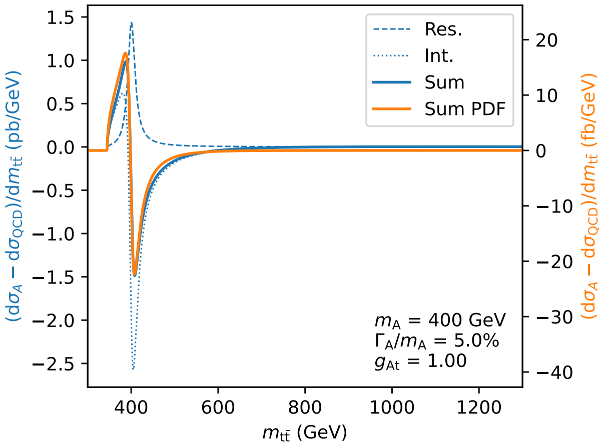
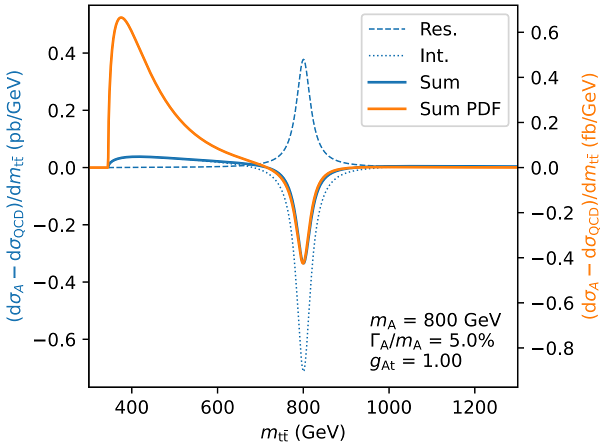


Figure 2.10: Feynman diagram for the heavy Higgs boson (scalar or pseudoscalar) production via a top loop and with a top decay.



(a)



(b)

Figure 2.11: Differential cross sections in terms of the invariant mass of the outgoing tops $m_{t\bar{t}}$ of the process $gg \to A \to t\bar{t}$ for the pseudoscalar A of masses $m_A = 400\,\text{GeV}$ (a) and 800 GeV (b), a relative decay width of 5% and a coupling modifier of 1. The QCD contribution $gg \to t\bar{t}$ is subtracted from the total cross section to obtain only the Higgs contribution. Shown are the resonance, interference and their sum in blue (left axis) and sum also accounting for the PDF (NNPDF3.1 [6], LHAPDF [7]) at the LHC energy of 13 TeV (right axis). The orange curve follows a different scale from the other curves. This scale is given on the right y axis.

ing identified the process $gg \to A/H \to t\bar{t}$, one will notice the SM already predicts a process without Higgs bosons but with the same initial and final states: $gg \to t\bar{t}$ (see section 2.7). As a consequence, in an experimental search for $gg \to A/H \to t\bar{t}$, there will always be an SM background, which interferes with the heavy Higgs signal [15]. Thus here the presence of the Higgs boson not only creates a positive resonance peak in the invariant mass spectrum at its own mass, but also an interference that contributes negatively on the same order. Differential cross sections for resonance and interference can be seen in figure 2.11. The resonance is explained in the same way as for other particle resonances: Once the available energy (here $m_{t\bar{t}}$) reaches the energy needed for the production (the Higgs boson mass m_A or m_H) the cross section sharply increases. However, the interference shows a more sophisticated structure, with a negative dip at around the same position as the resonance and with a long positive tail towards small invariant masses. This tail gets amplified by the contributions of the PDF, creating a large peak at low invariant masses when higher Higgs masses are assumed.

To be as model-independent as possible, in the following, the width is assumed to be a free parameter. However, the partial decay width is still bound by theory by (compare with [15])

$$\Gamma_{A/H \to t\bar{t}} = g_{A/H}^2 \frac{3 G_F m_t^2 m_{A/H}}{4\pi\sqrt{2}} \left(1 - \frac{4 m_t^2}{m_{A/H}^2}\right)^{x_{A/H}}, \tag{2.14}$$

with the Fermi constant G_F, the top quark mass m_t and

$$x_{\mathrm{A/H}} = \begin{cases} \frac{1}{2} & \text{for A} \\ \frac{3}{2} & \text{for H} \end{cases}. \tag{2.15}$$

A modification to the Yukawa coupling of A or H is denoted by the variable $g_{\mathrm{A/H}}$, which scales the coupling strength linearly. As this coupling appears as a factor in the particle decay width, there is an upper limit for the values of $g_{\mathrm{A/H}}$ in dependence of the assumed total decay width. At too high values the partial decay width exceeds the total one.

It should also be noted that when computing the cross section for gg $\to$ t$\bar{\mathrm{t}}$, because of the dependence of the heavy Higgs ME on the coupling ($M_{\mathrm{A/H}} \propto g^2_{\mathrm{A/H}}$), one obtains the following sum:

$$\sigma_{\mathrm{gg}\to\mathrm{t\bar{t}}} \propto |M_{\mathrm{A/H}} + M_{\mathrm{SM}}|^2 \tag{2.16}$$
$$= g^4_{\mathrm{A/H}} s_{\mathrm{res}} + g^2_{\mathrm{A/H}} s_{\mathrm{int}} + s_{\mathrm{SM}} \tag{2.17}$$

Here, s_{res}, s_{int} and s_{SM} are the contributions from the heavy Higgs resonance, interference and the SM respectively. Thus resonance and interference scale differently with respect to modifications of the Yukawa coupling.

Chapter 3

Experimental Setup

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular accelerator with a circumference of approximately 27 km located at a mean depth of 100 m underground on the border between Switzerland and France at the European Organization for Nuclear Research (CERN) [16]. It accelerates two counter-circulating beams of hadrons, either protons or lead ions, and makes them collide at its four large experiments: ATLAS [17], ALICE [18], CMS [19] or LHCb [20]. By definition, "beam 1" is the beam inside LHC that travels clockwise when looking at the LHC from above.

As the LHC is a collider with two beams at equal momentum amplitude, the center of mass energy that is available for the production of new particles upon collision is the sum of the energy of both beams. The final energy reached per proton beam has been increased in several steps over the years of LHC's operation, reaching 6.5 TeV in 2015 and 6.8 TeV in 2022 [21], adding up to 13 and 13.8 TeV. In order to arrive at these energies a chain of preaccelerators, constituting of linear and smaller circular accelerators, is required (see figure 3.1). Only at an energy of 450 GeV the preaccelerated protons are injected into the LHC.

At any time, may it be for normal operation or for safety considerations, the beams can be stopped and disposed of immediately by the LHC beam dumps. These are two 6 t devices containing steel barrels filled with graphite cores, onto which the beams can be shot to stop them [23].

The beams inside the LHC travel through an ultrahigh vacuum of 10^{-8} Pa to minimize the probability of the beam hitting stray particles. Many different systems of superconducting magnets are used at the LHC to control the particle beam. Among them are quadrupole magnets to shape and focus the beam and dipole magnets to bend its trajectory. The dipole magnets generate a magnetic field of up to 7.7 T, operate at a temperature of 1.9 K and are therefore constantly cooled by superfluid helium. A component that takes care of the beam's quality are the collimators installed along the LHC. These are devices made from a variety of robust materials, such as tungsten alloy [24], and can be moved close to the beam. This is done to restrict the area through which the beam can pass, thereby absorbing stray beam particles. The hadrons of the beam are accelerated using radio frequency (RF) cavities, which are made from niobium-coated copper and are operated at a superconducting temperature of 4.5 K [25]. These create an electromagnetic field in form of a standing wave with a frequency of 400.8 MHz [26]. For a beam particle to experience an accelerating force from the field, it should enter the cavity close to the point in time when the field has reached its maximum strength. For this reason the beam is not a continuous stream of particles but rather is made out of

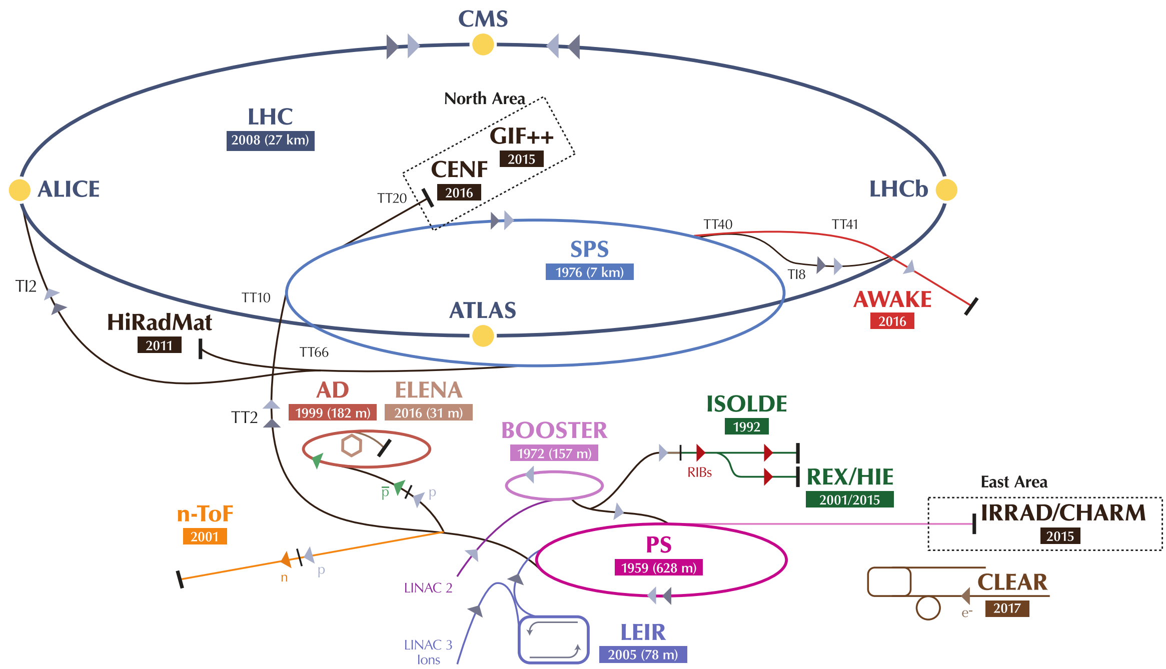


Figure 3.1: The CERN accelerator complex as of 2018. Shown are the accelerator's circumferences, the year of their first commissioning and experiments including the four large LHC experiments. From [22].

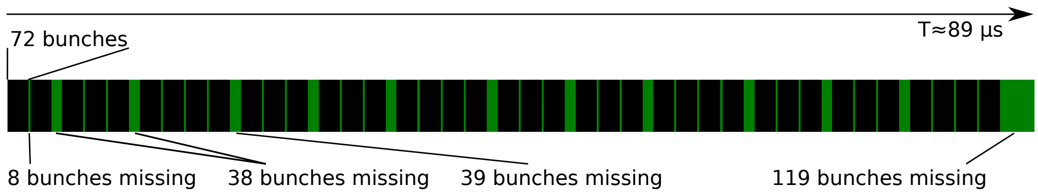


Figure 3.2: Possible filling scheme of the LHC, with proton bunches marked in black and gaps without bunches in green. The last big gap is the abort gap of around 3 µs. Based on [26].

bunches. Depending on the exact timing a particle enters the cavity's field, it experiences a stronger or weaker force, which aids keeping the particles in a bunch together, aligning their energies.

The LHC operates at a revolution frequency f_{LHC} of about 11.245 kHz. Its harmonic number, which is the number of oscillations of the RF cavities per revolution, is therefore 35640. As bunches must only enter the cavity at its maximum field strength, this also defines the upper limit of RF buckets one could possibly fill with particle bunches. Buckets are filled in the LHC with a minimal spacing of approximately 25 ns, which means a bunch every 7.5 m and implies a maximum of 3564 bunches. This number is further narrowed by several constraints, leading to a design number of bunches per beam of 2808 [27]. The exact number of bunches and their spacing is decided by the various filling schemes the LHC operates with. A possible filling scheme is visualized in figure 3.2. One of the constraints the filling schemes have to adhere to is the presence of an abort gap with 3 µs of empty buckets. This time is required for the kicker magnet to extract the beam into the beam dump, which is how the beam is disposed of so that LHC can ramp down and become ready for a new beam injection. A set of adjacent bunches without gap is called a train.

At injection a proton bunch inside the LHC contains around 1.2×10^{11} particles, which sums up to a total intensity of around 3×10^{14} protons per beam. Besides the beam intensity, in order to maximize the interaction probability of the particles, there is a variety of other parameters to take into account, one of which is the crossing angle. It describes the angle at which the beams collide and is required to be nonzero in order to have interactions only at the desired interaction point inside the detector. Another important parameter is the amplitude function β^*, which is connected to how focused the beam is and can be understood as the distance from the focus point to where the beam has reached twice the width [28]. The smaller the β^* the more squeezed together the particles are. Additionally the two beams can be separated from another, decreasing the probability of interaction. In order to keep number of simultaneous interactions, the pileup, low or to study the beam width, the separation can be increased.

To set up colliding beams that are used for physics measurements, the LHC goes through several different stages. The start is marked by beam particles from the preaccelerators at low energy arriving inside LHC, which is the beam injection. At this stage the two beams are still being positioned with high separation so that practically no collisions take place. The buckets of the LHC are gradually filled this way and once the targeted beam intensities are are reached, the acceleration can begin. The process of acceleration to the target beam energies is called ramp. The ramp is followed by two further stages, squeeze and adjust. In the former, the β^* is decreased, which focuses the beams, and in the latter the position of the beams is changed until they collide at a maximal rate. In order to determine the proper position in the last stage, the luminosity measurements reported by the experiments is made use of. These stages can take in total around one hour to complete. Concluding the adjustment, LHC declares stable beam and the physics measurements are being recorded. Due to the constantly occurring collisions and interactions, the beam intensities will decrease over time, which will also decrease the collision rates, thus making it necessary to inject new beams and to start anew. When the LHC control determines this to be the case, the left over beams are discarded by shooting them onto the beam dumps and the machine is made ready for a new injection. Often the beams stay in the machine in the stable beam mode for more than 12 hours. The process from injection to dump is called a fill and every fill is assigned its own fill number.

3.1.1 Luminosity

An important quantity that describes the performance of a particle collider is the luminosity. Together with the cross section σ of a process it describes the number of interaction of this process with

$$\frac{\mathrm{d}N}{\mathrm{d}t} = L\sigma. \tag{3.1}$$

Here L is the instantaneous luminosity, which is in contrast to the integrated luminosity L_{int}. The integrated luminosity is obtained by

$$L_{\mathrm{int}} = \int_{T_0}^{T_1} L\mathrm{d}t, \tag{3.2}$$

which is the integral over a time frame of the instantaneous luminosity.

The luminosity can be understood in the following way: With the cross section describing a probability for given particles to perform a process, the luminosity accounts for the amount of particles and how precisely they are brought into collision. Therefore it includes parameters, such as beam shape, collision angle or frequency, that are determined by the design and performance of the collider. Only the product of both, cross section and luminosity, expresses the number of interactions.

LHC reliably reached an instantaneous luminosity of around $20\,\mathrm{kHz}\,\mu\mathrm{b}^{-1}$ for proton-proton collisions in 2018, which was the highest luminosity delivered until then [29]. Within the same year the luminosity that was delivered by LHC integrated up to $67.86\,\mathrm{fb}^{-1}$.

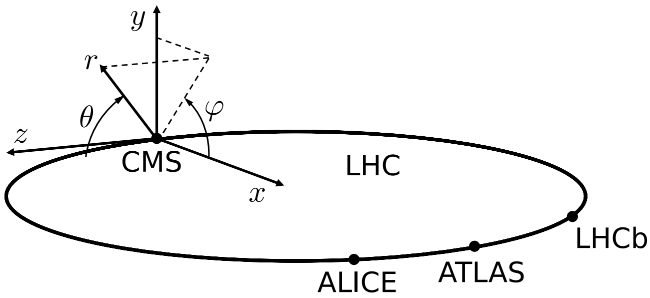


Figure 3.3: The CMS coordinate system. From [34], modified.

It is foreseen that the LHC will receive a major upgrade until 2029, which will allow a drastic increase of the luminosity [30]. This High-Luminosity LHC is planned to be operated throughout the 2030s and is designed to deliver 2.5 to 4 times the instantaneous luminosity reached in 2018. Some of the designated upgrade have already been put into place, such as a new beam pipe at the CMS experiment [31].

3.2 Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is the general purpose detector located at the interaction point 5 of LHC. It was designed with a broad spectrum of physics applications in mind. Its field of research includes the verification of the Standard Model, searches for particles and BSM phenomena or the search for extra dimensions. One of its initial goals was achieved in 2012 with the discovery of the Higgs boson [32]. The CMS collaboration, who takes care of the detector and all its data, consists of over 5000 people at over 200 institutes worldwide [33].

3.2.1 Coordinate system and variables

CMS uses a coordinate system that has its point of origin at the interaction point. From there, one defines a right-handed Cartesian system by having the x axis point into the center of LHC and the z axis pointing into the direction of where beam 1 is coming from (see figure 3.3). As a result, the y axis is pointing upwards. The x and y axes span a plane that is referred to as the transverse plane (as seen in relation to the beams). A common alternative is the usage of a spherical system of $\{r; \varphi; \theta\}$, which is obtained from transforming the Cartesian system. Here, following the usual convention, φ is the azimuthal angle and is the angle in the transverse plane.

An observable derived from the polar angle θ is the pseudorapidity η defined by

$$\eta = -\ln\tan\frac{\theta}{2}. \tag{3.3}$$

A pseudorapidity of 0 is a coordinate that is in the transverse plane going trough the point of origin, while high pseudorapidities are close to beams. In this system, one can define a distance denoted by ΔR using the differences in φ and η by

$$\Delta R = \sqrt{\Delta\varphi^2 + \Delta\eta^2}. \tag{3.4}$$

The rapidity y is equal to the pseudorapidity in the limit of high energies [2] and is defined by

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{3.5}$$

with the energy of a particle E and the z component of the momentum p_z of it.

Also in the context of momenta, the magnitude along the transverse plane is referred to as p_T that has the following relationship to the x and y components of a particle's momentum:

$$p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}. \tag{3.6}$$

3.2.2 Detector design

The CMS detector and ATLAS detector, another LHC detector located exactly at opposing sector to CMS, share many of their scientific goals. This seemingly duplication ensures that results of each of the two detectors can be validated by the other, which is made possible by the very different designs of the two detectors.

Compared to other detectors, such as the ATLAS detector, CMS is rather small with a height and width of 15 m and a length of 21 m, hence the "compact" in its name. However, having a weight of 12 500 t it is the heaviest detector at the LHC. A central component of CMS is the big solenoid magnet, that generates a large and nearly constant magnetic field of around 4 T inside its roughly 6 m diameter.

The objective of the detector is to measure and tell apart the particles coming from the collisions at its interaction point. For this purpose it consists of many different subsystems, each fulfilling different roles. The subsystems are ordered in layers around to pipe through which the beams are guided. The design results in a cylinder around the pipe, called barrel. To be able to reasonably measure particles leaving the interaction point at angles close to the beams (high η values), exiting the barrel before leaving a signal in all of the detector's layers, CMS incorporates so called endcaps on both of its sides. These are disc-shaped detector systems, layered in similar fashion as the barrel.

A key idea for particle identification is the usage of detector systems that each receive a signal only from specific types of particles. One consequence is the layered structure of CMS, as can be seen in figure 3.4. The main systems responsible for these tasks from the inner post to the outer most are the following:

- **The silicon tracking detector** (tracker) is sensitive to charged particles and can measure their momenta and origins. It is the detector closest to the beam pipe and consists of 4 layers of pixel detectors and several layers of strip detectors made from silicon [36]. The four pixel layers contain 124 million pixels, each with a size of 100 µm by 150 µm, surrounded by four layers of strips in the barrel and three in the endcap. All this is called the inner tracker. Located around it are 6 barrel layers of strips and 9 endcap strip disks. The tracker is cooled to a temperature of −20 °C and its volume is sealed, keeping the humidity level stable. When a charged particle hits a pixel or a strip, a charge is induced in it, which creates a signal in the readout electronics. A particle signal together with the information of where the corresponding pixels and strips are located makes it possible to deduce the track the particle took through the detector. Thanks to the magnetic field of the solenoid, this trajectory will be bent, with its curvature depending on the particles momentum's magnitude. By extrapolating this track towards the position of

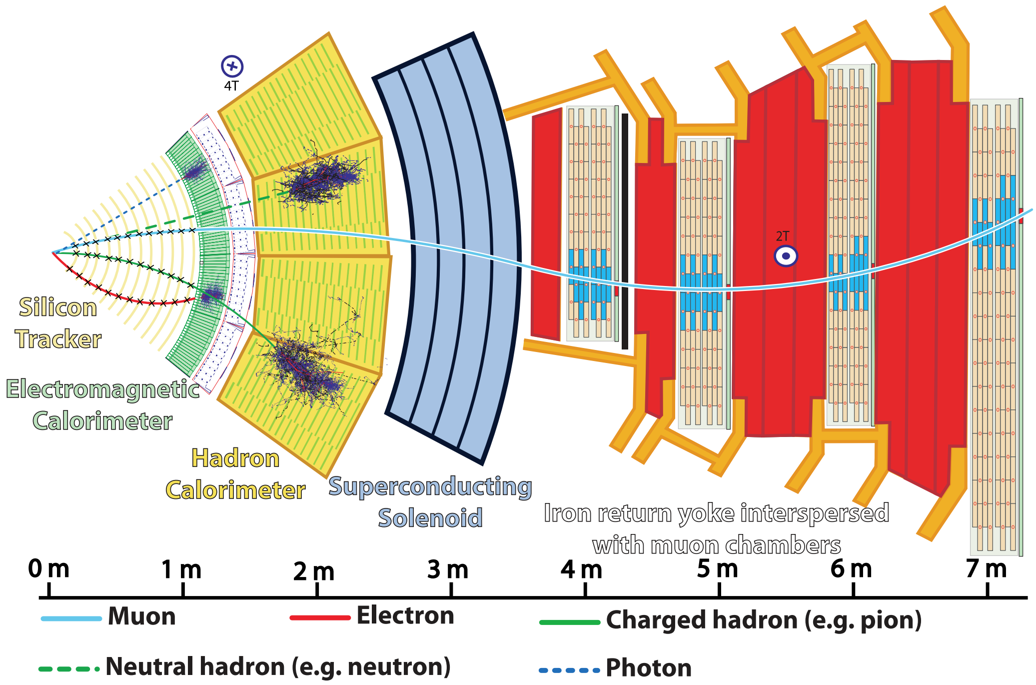


Figure 3.4: Cross section of the CMS detector with example particles traversing the different layers. Depending on the its type, a particle leaves a signal only in specific layers of the detector. From [35].

the beam, the point in space where the track originated from (vertex) can be reconstructed. The vertex resolution depends on several factors such as particle energy and direction. For vertices of proton collisions of interest for CMS it is on the order of 10 µm [37]. The pixel tracker is designed to cover a range of $|\eta| < 2.5$, in which particles hit all four layers [38], which is in alignment of the strip tracker, covering the same range.

- **The electromagnetic calorimeter** (ECAL) measures the energy of photons and electrons (and positrons). It consists of 75 848 lead tungstate crystals, that have a high density of $8.28\,\mathrm{g\,cm^{-3}}$ [39] and are transparent. The crystals have a length of 23 cm in the barrel (22 cm in the endcap) along the radial direction and have a square shape in the (η, φ) plane with an edge length of 2.2 cm (2.86 cm in the endcap). When an electron or photon hits the crystals it initiates an electromagnetic shower, a reaction in which more and more electron-positron pairs and photons are created. The amount of photons is dependent on the initial particles energy and is measured by avalanche photodiodes at the end of the crystals. Additionally the two-dimensional position is obtained by keeping track of the crystals from which the photons were observed. In most cases the electron or photon initiating the shower is stopped by the calorimeter and will not create a signal in any of the next layers. The crystals of the barrel ECAL cover a range of up to $|\eta| < 1.48$ and together with the endcap ECAL a range of $|\eta| < 3$ is achieved. The front of the ECAL in the endcaps is equipped with a special component, which is the Preshower. It helps initiating the shower in the thinner endcap calorimeter and is able to improve resolution on photons close to each other, utilizing silicon detectors [40].

- **The hadron calorimeter** (HCAL) again measures particle energy. In contrast to the ECAL, it is sensitive to hadrons, whereas electrons or photons will in most cases not even reach the HCAL, as they are stopped by the ECAL. The HCAL is a sampling calorimeter, meaning that it consists of interleaved layers of absorber plates and scintillator material. The absorber initiates hadron

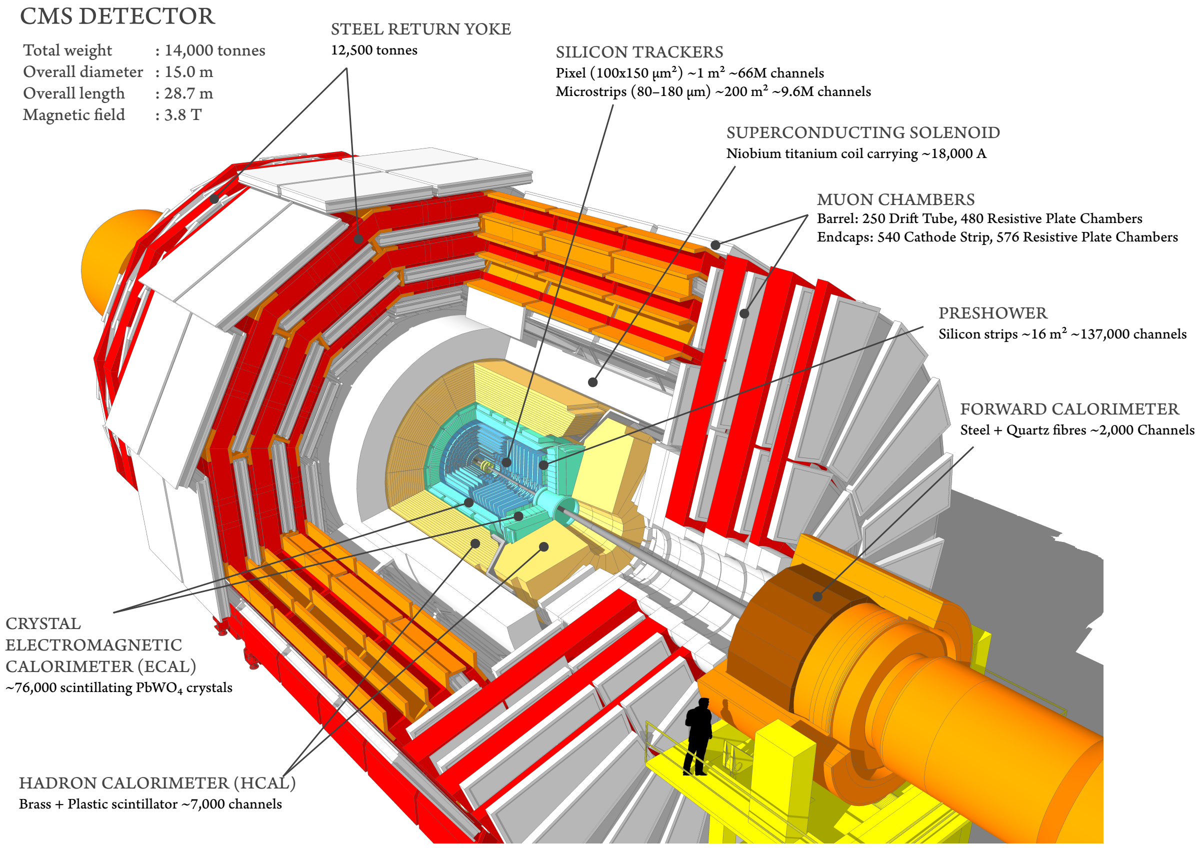


Figure 3.5: Cutaway diagram of the CMS detector with some of the main systems labeled. From [41].

showers, which creates many more particles from each incoming hadron. The scintillator is transparent, shifts the wavelength of the photons produced and guides them to the phototransducers, where the amount of photons is measured. The HCAL is separated into four sections at different locations: the barrel and outer barrel, the endcap and the forward sections (see figure 3.5). The sections differ in design: In the barrel and endcap sections brass absorbers are used, while the forward section uses steel absorbers. The barrel and endcap sections cover the range of up to $|\eta| < 1.4$ and $|\eta| < 3$ respectively [42]. The forward section extends to values of up to $|\eta| < 5$ and uses Cherenkov light, which is converted into an electrical signal by photomultiplier tubes [43].

- **The muon chambers** build the outer layer of the CMS detector and are responsible for detecting muons. All other particles that are directly detectable by the CMS generally are stopped by the previous layers, making muons the only particles that are able to reach the chambers and create a signal. This is further aided by the four massive layers of the iron yoke, mainly responsible to confine the magnetic field of the solenoid [44]. The muon chambers are divided into four different subsystems, each taking a different approach: drift tubes, cathode strip chambers, resistive plate chambers and gas electron multipliers. They all share the fact that they contain some form of gas, which is ionized once a muon passes through. For example the drift tubes use a gas mixture of argon and carbon dioxide. In them the charges from the ionization drift towards wires due to an electrical potential, which leaves a signal in the readout [45]. Combined with information about the signal timing, the crossing point of the muon can be determined. The drift tubes cover the lowest η values with a range of $|\eta| < 1.2$ [46]. Muons with higher

η values are detected by the cathode strip chambers ranging between $0.9 < |\eta| < 2.4$ and the resistive plate chambers with up to $|\eta| < 1.9$. Lastly the gas electron multipliers cover the range of $1.6 < |\eta| < 2.4$.

3.2.3 Trigger system

Thanks to LHC's high performance, proton interaction rates on the order of GHz are delivered to the CMS detector [47]. However, most collisions happen at low momentum transfers and for example only produce light particles at low energies, which are of very little interest to the physics program of CMS. Recording all of these interactions would also require immense computational effort. For these reasons CMS employs an event trigger system, made from two distinct levels. This system is responsible for selecting collisions of particular interest (an event), whose information will be further processed and made accessible for later offline analyses.

The first level (L1) of the trigger system is implemented using custom hardware and is reducing the event rate to 100 kHz. It uses the readout of several systems, including the CMS calorimeters, muon chambers or from the beam monitoring system (for an example see section 4.4). The information is used to obtain candidate objects that show features consistent with the ones of, for example, electrons or muons of interest. The L1 system is designed to operate at a high frequency and is able to reach a decision just 4 µs of a collision. There is a total of 512 L1 algorithms [48], of which at least one must give a positive result for a final positive decision.

From the L1 system the events are propagated to the high level trigger (HLT). The HLT has significantly more time available for a single event and is implemented in software running on about 26000 processor cores. It can utilize more sophisticated methods to build candidate objects and reduces the event rate to an average of 400 Hz.

A prescale can be assigned to any trigger algorithm. This prescale is a natural number n that makes makes the system treat the events as if an algorithm only gave a positive result in every nth time it actually did, effectively skipping events on a randomized basis. A prescale can be adjusted to reduce the rate of a trigger algorithm in the case it is determined a rate of a particular algorithm triggers too often.

3.2.4 Data processing

After an event has passed the triggers, its data is sent to the Worldwide LHC computing grid. The grid is distributed across many computing centers around the world and is divided into four tiers, starting with the Tier-0 center at CERN [49]. The Tier-0 receives the raw data and reconstructs first objects, such as electrons, that are of interests for further analysis. This from of data is called RECO. From there the Tier-1 and subsequent tiers take over. At the Tier-1 centers the so called Analysis Object Data (AOD) is created from the data coming from Tier-0. This AOD format contains all high-level physics objects plus a summary of the lower level information, which is sufficient information for a great majority of all CMS analyses. The AOD uses the ROOT file format to store its content, that is structured using the CMS Event Data Model. Reading this data requires the CMS software framework CMSSW [50].

Following the AOD, there are two more steps done centrally within the grid, which further narrow down the information content. The output formats are called MiniAOD [51] and NanoAOD, the former serving as input for generate the latter. The entire chain of data formats can be seen figure

3.6. NanoAOD is the most portable and lightweight format with a size of around 1 kB per event and does not require CMSSW for readout. Additionally most calibrations are already applied and it provides the output of many high-level algorithms, enabling a very optimized and performant analysis workflow.

Figure 3.6: Data formats employed by CMS and their approximate sizes per event. The data formats are produced in the same order from left to right: For example NanoAOD is produced from MiniAOD.

Already from the Raw stage, the events are categorized into different data sets, based on which HLT the event passed. Physics analyses then only need process the files belonging to the data sets with the triggers of interest, thereby greatly narrowing down the number of events being needed.

3.3 Object reconstruction

Computing variables, such as four momenta or charge, for physical objects like leptons from the raw detector data is what is called objects reconstruction and is the task of a range of algorithms employed by CMS. A central idea in the approach taken by CMS is the particle flow (PF) algorithm [52]. It correlates information from all the different layers of CMS instead of focusing on one particular part of the detector for each particle type. A building block of this approach is the link algorithm that takes the individual PF elements (these are tracks, vertices, electrons, muons and calorimeter clusters, see below) from different detector parts and creates groups from them, called PF blocks, if they fulfill specific criteria. For example, two elements are linked only when they are nearest neighbors in the (η, φ) plane. Normally the reconstruction algorithms for elements and objects start from characteristic features, such as a high, well located energy deposits, and try finding signals that can belong to the same object or element. These features are called seeds. The PF algorithm includes these elements and resulting objects (among others):

- **Tracks and vertices** of charged particles are fundamental to many quantities and are reconstructed using the silicon tracker. The raw data from the tracker provides location information about where a particle passed through a layer of the tracker called hits. To derive a track from the hits a three step approach of finding a seed, collecting hits for one particular track and performing a χ^2 fit is used. The seed of the first step is a property every track is required to have and can be for example a minimum number of hits in the pixel tracker. The procedure is iterated, while a hit can only ever be associated to one track, which reduces the number of hits available to each iteration. The seeds used in each iteration differ, going from seeds using only pixel hits, to including also strip hits and finally using information from the muon chambers. Primary vertices, which are the locations of the proton interactions, are reconstructed using the tracks [37]. This is done by imposing quality criteria on the tracks and clustering them using the closest approach to the interaction point along the z axis. x, y and z positions are obtained from fitting. In the fit each track is assigned a weight w_i between 0 and 1, corresponding to a likelihood of the track belonging to the vertex. Using the weight, the number of degrees of freedom of the track is defined as

$$n_{\mathrm{dof}} = -3 + 2\sum w_i, \tag{3.7}$$

giving a handle on the number of tracks coming from the vertex. Primary vertices are vertices at which the interaction of the beam particles took place and reconstructing them allows telling apart individual interactions within one bunch crossing. The number of primary vertices in a single bunch crossing corresponds to the amount of pileup. To differentiate between hard-scattering and uninteresting pileup vertices, the sum of the p_{T} of the tracks belonging to each vertex is used. The vertex with the highest sum is considered to be the hard-scattering vertex.

- **Calorimeter clusters** reflect energy deposits in the ECAL and HCAL together with a position. When objects enter the calorimeters they usually deposit energy in several neighboring calorimeter crystals. On top of that, multiple objects may create deposits in overlapping regions. To overcome this challenge, a clustering algorithm is employed that first finds seeds, then clusters all neighboring deposits into single clusters and finally determines one cluster per seed together with an energy value using a fit method. The seeds are crystals with deposits of certain size that are larger than any of their direct neighbors.

- **Electrons** (and positrons) usually start radiating photons when passing through the tracker volume. This results in multiple close clusters in the ECAL belonging to a single electron, which have to be grouped together. Such a group is called supercluster. Seeds for this method are given from clusters above a certain energy threshold. Additionally the clusters must fulfill isolation criteria, meaning they have be distanced from other calorimeter deposits, to make sure these electrons are not part of a jet. This ECAL-driven approach is complemented by a tracker-driven approach. This approach is seeded by any track with $p_{\mathrm{T}} > 2\,\mathrm{GeV}$ whose closest cluster has a compatible energy. Tracks and clusters form PF blocks, which are used to select the resulting electrons. These tracks and clusters may not be used for other objects further on.

- **Muons** are reconstructed using a three-way technique. This technique allows for standalone muons coming from tracks in the muon chambers. Such a muon then becomes a global muon if a compatible track in the inner tracker exists. Finally there can be tracker muons, which are made from tracks in the inner tracker, that are compatible with at least one hit in the muon chambers.

- **Hadrons** are built from clusters in the HCAL. This is done after the elements and objects (plus isolated photons) have been reconstructed and their inputs have been removed from the available PF items. A hadron can be either identified as neutral if the corresponding cluster does not have tracks linked to it, or as charged if the opposite is true. In the case the cluster lies outside the η region that is covered by the tracker, to be considered a hadron the cluster is required to be linked to an ECAL cluster. If tracks are linked to a cluster, they can aid in determining the energy of the hadron. CMS employs the pileup charged-hadron subtraction (CHS) method. This method removes hadrons that do not come from the hard-scattering vertex from the list of objects to use for further physics objects reconstruction.

Building upon the results of the PF algorithm, many other types of objects can be reconstructed, including:

- **Jets** appear in the detector when quarks and gluons are produced and hadronize, which means that they create a spray of hadrons as footprint of the original parton in the detector. During this hadronization process a multitude of new, lighter particles is created that hit the detector in a wider but still defined region. The anti-k_t algorithm (see reference [53]) is used to group all these particles, which can be any particle that is reconstructed by the PF algorithm. It is based on the two distance measures of any particles enumerated by i and j:

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{-2}, p_{\mathrm{T},j}^{-2}\right) \frac{\Delta_{ij}^2}{R^2} \tag{3.8}$$

$$d_{iB} = p_{\mathrm{T},i}^{-2} \tag{3.9}$$

with $\Delta_{ij}^2 = (y_i - y_j)^2 + (\varphi_i - \varphi_j)^2$ and where $p_{\mathrm{T},i}$, y_i and φ_i are the momentum, rapidity and azimuthal angle of the ith particle and R is a free parameter. In the algorithm if $d_{iB} > d_{ij}$ for a given i and all $j \neq i$, the particle i will be removed from the list of particles and will be considered a jet, otherwise the particle i will be grouped with the particle for which d_{ij} is the smallest and be considered one particle. This is repeated until there are no more particles left. R determines the allowed radius of the jet. A common value is $R = 0.4$, but also higher values find usage, for example to produce a single jet of multiple decay products with little separation [54].

  - **Bottom-flavored jets** are jets that are initiated by bottom quarks (or antiquarks). These kind of jets show several unique properties, which makes it possible to differentiate them from other jets stemming from other partons. Popular algorithms for this task in CMS are DeepCSV [55] and DeepJet [56], which both utilize deep neural networks to classify the jets. One important property is the displacement of the vertex of the bottom-flavored jet, allowing the reconstruction of a secondary vertex close to the primary vertex of the interaction. The cause of this displacement is the relatively long lifetime of the mesons that leave the interaction point and contain bottom quarks (B mesons) of 1.5 ps [1], allowing the meson to travel a distance of several mm. This property is illustrated in figure 3.7. Other useful properties are the high mass of the bottom quark resulting in large p_T values of its decay products and its probability of 20 % to decay into an electron or a muon.

- **Missing transverse energy** (MET) can be seen as the final product after all other objects have been reconstructed and calibrated. At the LHC the sum of the transverse momenta of the particles taking part in the interaction is approximately zero. Because momentum and energy is conserved, if all produced particles have been measured, the resulting sum of transverse momenta should still be zero. On the other hand, a non-zero value (the presence of MET) points to the presence of neutrinos or may indicate possible BSM physics. Assuming zero transverse momenta one usually expresses MET in terms of the three vector

$$\vec{p}_\mathrm{miss} = -\sum_i \vec{p}_i, \tag{3.10}$$

where $\vec{p}_i$ is the momentum three vector of the ith reconstructed particle [2]. Due to the before mentioned conservation, only the x and y components of this quantity are of interest (referred to as MET x and MET y in this thesis). Often the two dimensions are expressed in spherical coordinates, giving magnitude (MET p_T) and polar angle (MET φ).

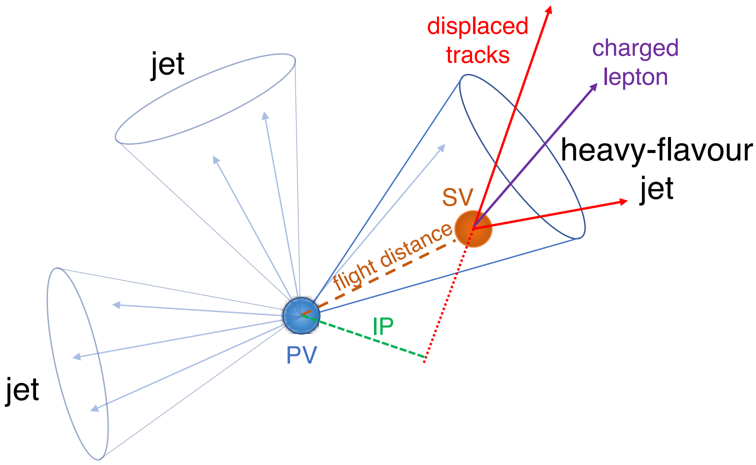


Figure 3.7: Illustration of an interaction producing a bottom-flavored jet (heavy-flavour jet). From a primary vertex (PV) several jets are emitted. The tracks for the heavy-flavor jet point to a secondary vertex (SV) that is displaced by a measurable distance from the PV as indicated by the nonzero impact parameter (IP). The phenomenon enables telling apart this type of jet. From [55].

Chapter 4

Measuring luminosity and beam-induced background

4.1 Beam Radiation, Instrumentation and Luminosity project

The Beam Radiation, Instrumentation and Luminosity (BRIL) project at CMS performs measurements and monitors, as well as simulates luminosity, beam induced background (BIB) and the radiation environment. The measured instantaneous luminosity serves as important input to the operation of the LHC. Integrated luminosity is a fundamental ingredient for analyses of the CMS data, such as cross section measurements. Continuous measurement of the background is required in order to detect possible deviations of the beams from their normal orbits and guarantee a safe environment for the operation of the CMS tracker. The amount of exposure to radiation is crucial for predicting the lifetime of individual parts of the CMS experiment. For these purposes, the BRIL project operates a range of detectors, located at various places throughout the CMS experimental cavern, some of which are shown in figure 4.1. The main BRIL detectors are the following:

- The **Fast Beam Condition Monitor (BCM1F)** is a detector that measures luminosity (a luminometer) and BIP and is described in detail in the following sections.
- The **Pixel Luminosity Telescope (PLT)** is another luminometer [58]. It is located right next to BCM1F and BCML1 and uses the same pixel sensors used in the past for the CMS pixel tracker. It operates 48 of these sensors that are arranged in three layers, building 16 telescopes with a length of 7.5 cm. The telescopes are placed in a circle at around $|\eta| = 4.2$ on both sides of the interaction point. Only when a coincidence of in all three layers is measured, it is counted as particle hit. The per-bunch-crossing counts are read out and can be used as a quantity that is expected to be proportional to the luminosity.
- The **Beam Conditions Monitor Leakage (BCML)** are two detectors responsible for protecting the tracker from abnormal beam conditions and its counts are reported as BIB to the other CMS and LHC systems [59]. The sensors of the first one, BCM1L, are placed right on top of the BCM1F board, while the second one, BCM2L, is located at a distance of $|z| = 14.4$ m. The systems use chemical vapor deposition (CVD) diamonds to count particles outside the beam axes. It is ensured the counts reported by BCML are below a certain threshold before the CMS tracker is powered on. Additionally, in the extreme beam-loss scenarios, that could potentially damage the tracker, BCML initiates an immediate beam dump.

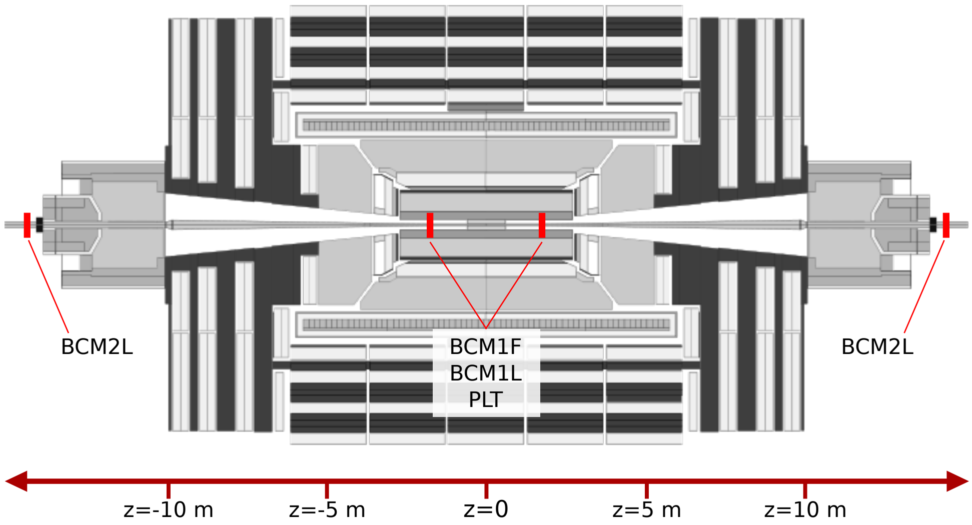


Figure 4.1: Locations of the three BRIL detectors BCM1F, PLT and BCML inside the CMS detector along the z axis. Using figure from [57].

- The **Beam Pick-up Timing for Experiments (BPTX)** is located outside the CMS experimental cavern at $|z| = 175$ m and uses a set of electrodes placed around the beam pipe [60]. It detects the incoming beams and can achieve a timing precision on the order of 50 ps. The obtained signal provides timing information about the beam bunches, is used as a zero-bias trigger with zero dead time to trigger several other CMS systems and is also part of the inputs to the L1 trigger system. One important parameter obtained is the ΔT, which is the time difference between two colliding bunches. This allows the prediction of the point at which the bunches actually collide along the z axis.

BRIL uses several other detector systems to derive luminosity measurements. This includes utilizing the pixel detector, the drift tubes of the muon chambers as well as the hadronic forward calorimeter. The measurement using the hadronic forward calorimeter can be done at the collision frequency and thus provides an additional per-bunch value during operation together with BCM1F and PLT.

4.2 Overview of the BCM1F detector

The BCM1F detector has two main purposes: Firstly it measures the luminosity that is delivered to CMS. Secondly it provides two background measurements, one for each beam. The frontend is located around the beam pipe on both sides of the interaction point with a distance along the z axis of 1.8 m (see figure 4.2). This distance is chosen on purpose as at this position incoming bunches and particles from the collision are separated by approximately 12.5 ns, which is half the spacing between incoming bunches [61]. The detector uses 48 silicon diodes as sensors arranged symmetrically around the beam pipe at a radius of around 7 cm. A high voltage (HV) is applied to the silicon diodes and they are operated in reverse bias. Charged particles hitting the diodes will induce a charge, leaving an electrical signal. The signal is amplified and send to the backend via an analog optical transmission, where the number of signals per time interval is counted. Counting is done in intervals of 6.25 ns and 4.17 ns (depending on the backend branch used) allowing 4 and 6 intervals every bunch crossing. The detector's location together with the high time granularity makes it possible to tell apart hits coming with the incoming bunches and hits resulting from the collisions. The former are used to calculate the background, the latter are used for the luminosity measurement.

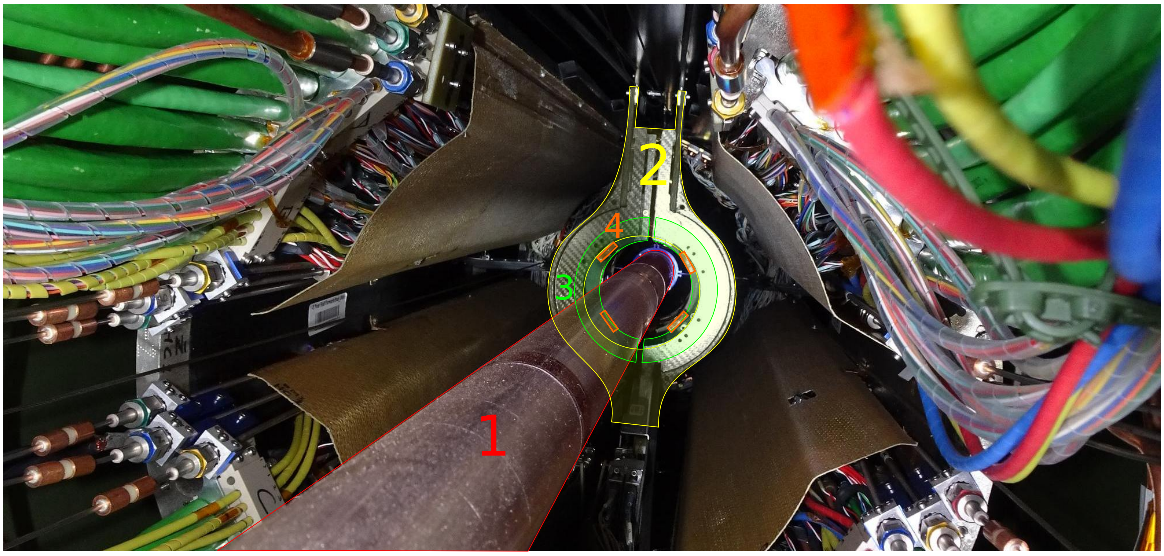


Figure 4.2: The BCM1/PLT detectors inside CMS. Marked is the LHC beam pipe (1, red), the supporting structure for the BRIL detectors (2, yellow), the BCM1F detector (3, green) and the BCM1L sensors (4, orange).

4.3 The BCM1F detector frontend and its upgrade

4.3.1 Frontend

The frontend of the BCM1F detector is located in the experimental cavern. Its main components are the so called C-Shapes, which consist of printed circuit boards shaped like the letter C, and the optoboards. Four quadrants, each containing one C-Shape, which can be seen in figure 4.3, and one optoboard, make up the frontend. Two quadrants are located on each side of the interaction point at the beam pipe, with the two C-Shapes forming a ring around the beam pipe. Each C-Shape has 6 double diodes, acting as sensors, resulting in a total of 12 individual readout channels. The diodes are made of silicon and were produced as part of the n-in-p wafers for the Phase 2 upgrade of the CMS tracking detector [62]. The diodes are connected via bond wires to a custom charge-sensitive amplifier chip (an ASIC, Application-specific integrated circuit) [63], that amplifies and shapes the signal, yielding voltage pulse with a full width at half maximum (FWHM) of around 10 ns as response to particle hits. The ASIC is also able to generate a pulse on request without the presence of a particle hit, which is used for testing the functionality of the frontend and is therefore referred to as test pulse. Via a flexible ribbon cable the voltage signals arrive at the optoboard of the quadrant, where Analog-Opto-Hybrids (AOH) [64] convert them to an optical signals. Laser diodes inside the AOH generate this signal. For the laser diodes to work in a proportional regime an offset current must be applied to the signal. The offset current, called AOH bias, can be configured in the AOH electronics via an I2C signal. This control signal is received through a Digital-Opto-Hybrid (DOH), which is also present on the optoboard. The frontend is located inside the same cooled volume that also contains the CMS tracker and the C-Shapes are cooled via cooling pipes connected to the tracker cooling. The nominal temperature of the connected cooling loop is $-20\,^{\circ}$C.

The design of the BCM1F frontend was upgraded several times over the duration of the operation of the LHC. Initially the detector used single-crystal CVD diamonds as sensors [65] and later changed to poly-crystal CVD diamond and silicon sensors [66]. The final upgrade to a full silicon-based detector was performed for the start of the LHC Run 3 in 2022. This transition to silicon promised a larger

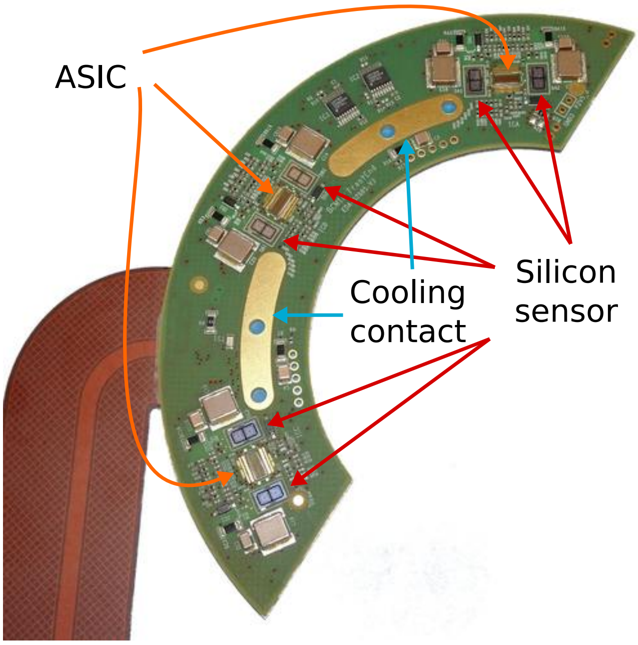


Figure 4.3: Photo of a C-Shape and the ribbon cable before assembly of the detector quadrant.

signal, therefore better separation from noise, and a very linear relationship between hit counts and luminosity. In order to compensate the heat created by the silicon sensors, in the same upgrade, the C-Shapes' connection to a cooling loop was added.

4.3.2 Silicon sensors

In particle physics, sensors made from silicon are widely used, with the CMS tracker or the BCM1F detector being just a few examples. Fundamentally these sensors are diodes and follow the same basic principles of pn-junctions: A p-doped silicon material is brought into contact with an n-doped material. As the p-doped one contains a majority of positive free charge carries, while the n-doped material contains a majority of negative charge carriers, the carriers close to the region of contact are exchanged and move into the oppositely doped material. This process depletes the contact region of carriers. By applying a voltage, called bias voltage, from n- to p-doped material an interesting behavior can be observed: In one polarity the voltage adds carriers, narrowing the depletion region until it is gone completely, and once this happens the diode is conductive. On the other hand, a voltage of the opposite polarity will widen the depletion region keeping the diode insulating. Such a voltage is called reverse bias. At too high values of reverse bias the diode will show a sudden electric breakdown, making it highly conductive again. The breakdown possibly damages the diode permanently.

While the diode can be seen as practically insulating with reverse bias, there still is a small current flowing through it. This current is called leakage current I_{L} and has the following property [67]:

$$I_{\mathrm{L}} \propto w \propto \sqrt{V} \tag{4.1}$$

where w is the width of the depletion region and V is the bias voltage.

In the case of silicon sensors for particle physics, the operation is always done with a reverse bias. When a particle hits the sensor, it creates free charge carriers in the depletion region, which then drift towards electrodes that apply the voltage. This effect created by the particle represents a current flow

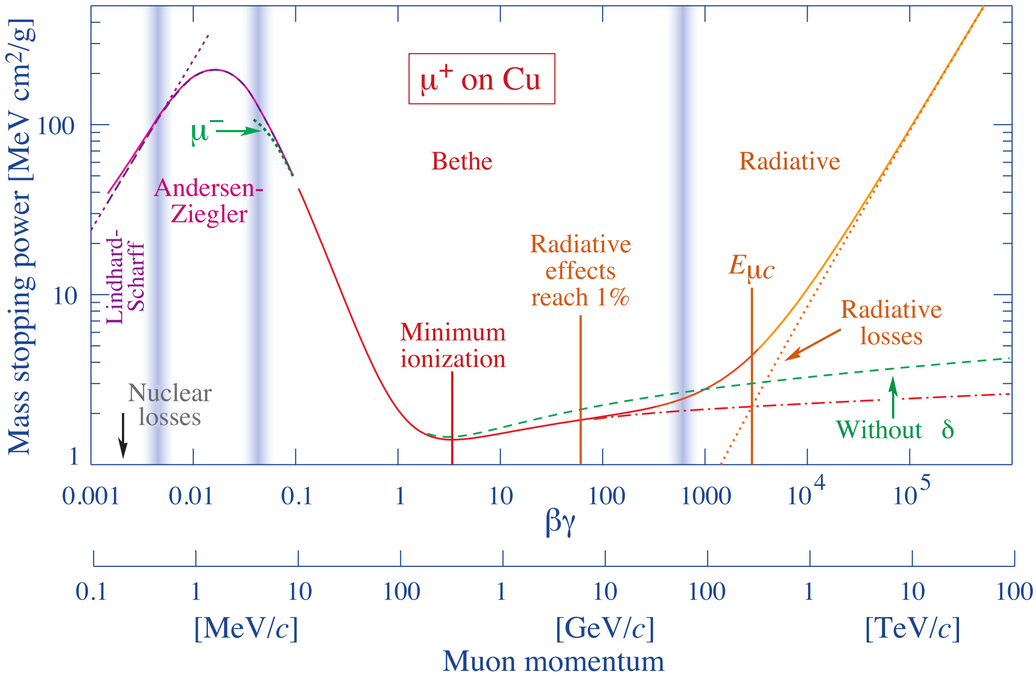


Figure 4.4: Graph of the Bethe-Bloch formula together with various corrections to it. Shown is the energy loss per density of a charged particle in copper. By dividing by the density, the dependence on the material (here copper) becomes minor, thus one can expect a very similar curve for silicon. From [1].

which can be measured. To have particles create signals that are as large as possible, the width of the depletion region should be maximal. The voltage at which the maximum width is achieved is called full depletion voltage V_{FD} and has the property

$$V_{\mathrm{FD}} \propto D^2 \tag{4.2}$$

where D is the sensor thickness.

The sensor can be seen as a plate capacitor, with the electrodes being the plates and the silicon the dielectric. Increasing the width of the depletion region is equivalent to increasing the space between the plates. Using the formula for the capacitance of a plate capacitor and the fact that the width is proportional to the square root of the applied reverse bias until the full depletion is reached, one obtains the following proportionality

$$C \propto \begin{cases} \frac{1}{\sqrt{V}} & V \leq V_{\mathrm{FD}} \\ \text{const} & V > V_{\mathrm{FD}} \end{cases}. \tag{4.3}$$

The size of the signal that is measured depends on the amount of energy and therefore the amount of charge an incident particle deposits in the depletion zone. The energy deposit from the ionization of charged particles traversing through a material is described by the Bethe-Bloch formula, whose graph a depicted in figure 4.4. For very low particle energies nuclear losses dominate, while for highly relativistic particles radiative losses are the dominant effect. For particles (in figure 4.4 shown for muons) with momentum on the order of 100 MeV to 10 GeV, which is the energy regime of many particles produced at the LHC, the energy loss is rather constant with momentum. Therefore most of the particles deposit very similar amounts of energy in the sensors. At around $\beta\gamma = 3$ the energy

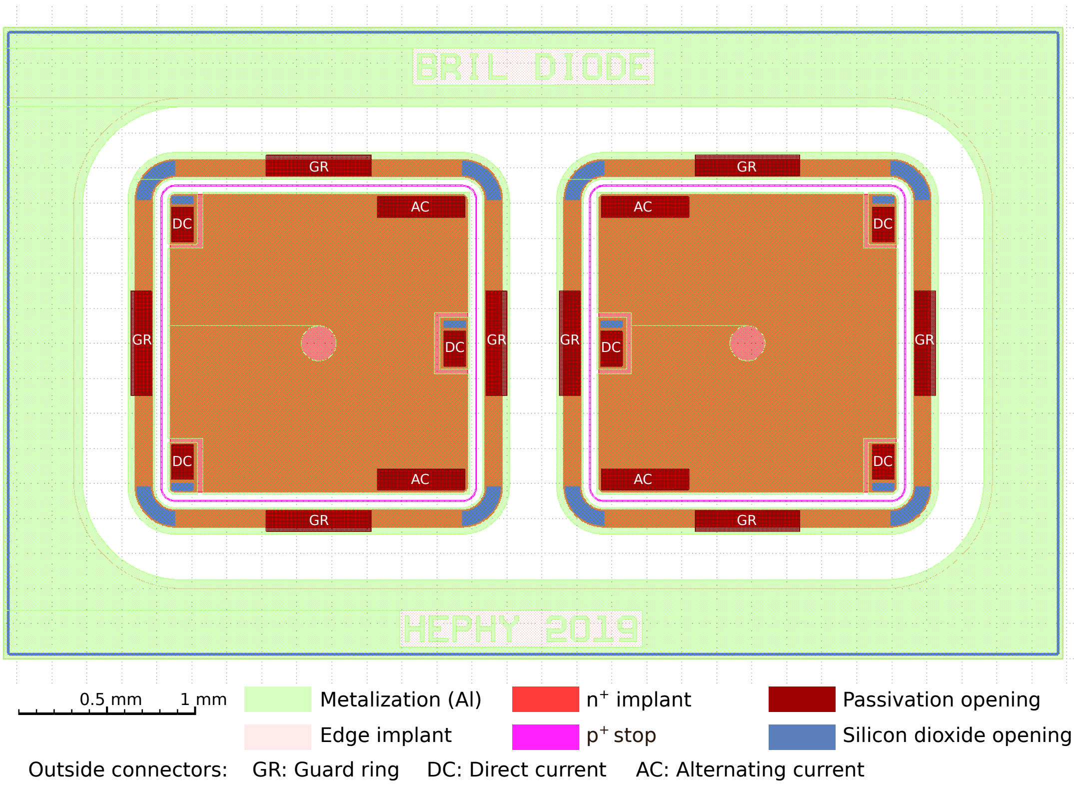


Figure 4.5: Layout of the sensors (silicon double diodes) used for the BCM1F detector. Metalization (made from aluminum) and n^+ implant layers overlap each other for most of the area, which is the orange area in the figure. Outside connections to the sensor are established via passivation openings. These openings are labeled in the figure by their purpose. The metalization and the n^+ implant inside the sensor are connected through the silicon dioxide openings.

loss is at its minimum. Particles with these momenta are called minimally ionizing particles (MIP) and it is crucial for the detector to be able to clearly separate MIP signals from noise.

The strong radiation a sensor is exposed to over its time of operation causes defects inside the lattice of the silicon crystal. The result are degradations in the sensor's properties, namely an increase of leakage current, change of the full depletion voltage and less separation between noise and signal. The cooling of the sensors carries away the heat that is produced due to the leakage current running through it. Because the amount of leakage current also increases with temperature, without proper cooling the sensor may enter the condition of "thermal runaway". In thermal runaway, the sensor is in a self-amplifying loop of increasing leakage current and heat generation, which prevents a stable operation of the sensor. Thermal runaway is especially a crucial aspect after radiation with increased bias voltage (because of higher V_{FD}) and already higher leakage currents.

4.3.3 Layout of the BCM1F sensors

The design of the BCM1F sensors can be seen in figure 4.5. In total BCM1F comprises 24 double diodes. Each of the diodes has a square shape with an area of $1.7 \times 1.7\,\mathrm{mm}^2$. The main component of these diodes is doped silicon, however several layers of other materials are also part of the diodes' design, which can be seen in figure 4.6. The diode has a width of 300 µm. The top of the diode is covered by a non-conductive layer of silicon dioxide, which acts as a passivation, protecting the layers underneath. Looking at the diode from the top, it can be seen that it is surrounded by a guard ring (GR). The GR prevents particle hits close to the diode from creating low efficiency hits. Between

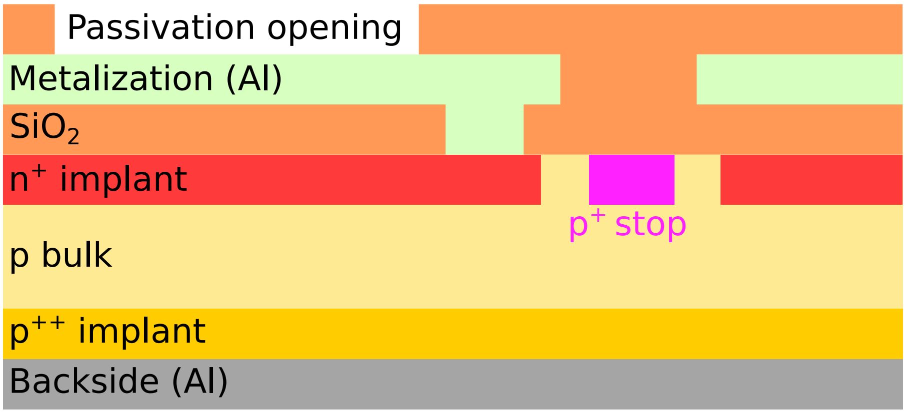


Figure 4.6: Side view (not to scale) of a slice through part of the diode of the BCM1F detector. The diode and the guard ring are separated by a p^+ stop. The conductive metaliaztion layer is only reachable for electric connections from the outside via the passivation openings. It connects to the n^+ implant only at specific parts of the sensor, where the silicon dioxide has an opening.

the GR and the actual diode a p^+ stop is located, which breaks possible electron accumulation layers that might be present below the insulating silicon dioxide layer. These electron accumulations could otherwise allow currents from the diode to its GR.

In operation an HV is applied from the conductive backside of the diode (which is the bottom in figure 4.6). Connections to the diode and its GR are established using wire bonds attached to the metalization layer through the passivation openings. The GR is connected to ground while the diode has two types of connections. One type is the direct current (DC) connection, which goes to ground. It is located on pads that are insulated from the rest of the metalization and connected to the n^+ implant via openings in the layer of silicon dioxide that otherwise separates them. The other type of connection is the alternating current (AC) one. Unlike the DC connection, the AC one is on top of the metalization that has no direct contact to the doped silicon. This connection is used as a capacitive coupling and the metalization together with the silicon can be viewed as a capacitor. While the leakage current will flow through the DC connection, the AC will only carry modulation on the current, which is the signal coming from particle hits. The AC signal is what is amplified and sent to the detector backend.

On the C-Shape, the DC connection features a $10\,\mathrm{M\Omega}$ resistor to ground. Therefore to not have the leakage currents go through the AC connection, the resistance between the metalization of the AC connection and the silicon is required to be substantially larger.

4.3.4 Sensor qualification

Before the assembly of the detector, the BCM1F sensors were qualified by performing three types of measurements.

- Voltage vs. current (IV) characteristic: The leakage current of the diode (via the DC connection) and the GR in dependence of the bias voltage is measured. If the breakdown voltage of the diode or the GR happens to be abnormally low, it will become apparent using the measurement. Applied voltages range from 5 V to 1000 V.

- Voltage vs. capacitance (CV) characteristic: The capacitance of the diode in dependence of the bias voltage is measured. This measurement allows to determine the full depletion voltage easily. Applied voltages range from 5 V to 500 V.

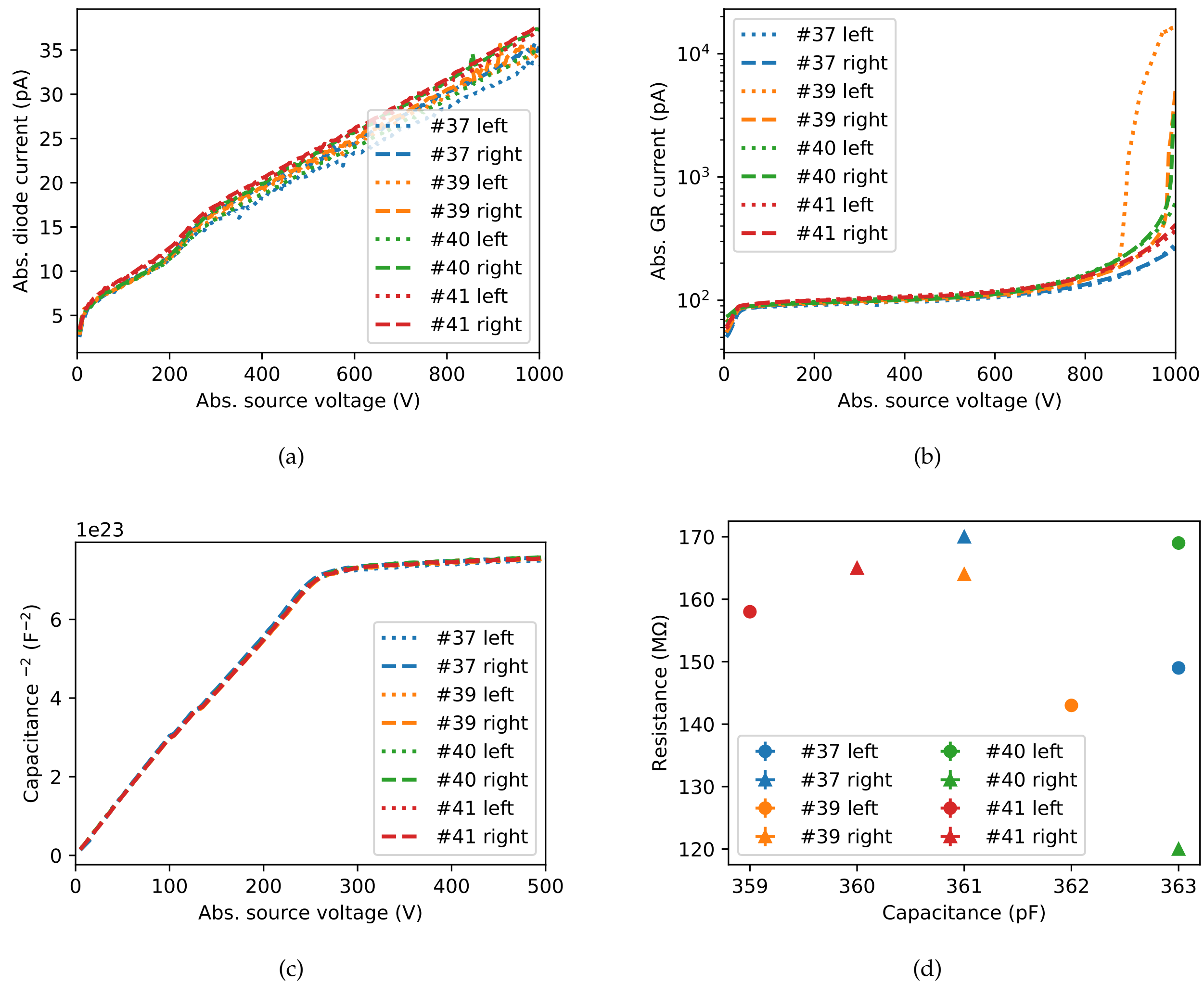


Figure 4.7: Results of the sensor qualification tests for selected sensors. (a) and (b) show the IV characteristic for the diode and the GR respectively, (c) shows the CV characteristic and (d) shows the resistance and capacitance measurement. All of the sensors are within expectation.

- Resistance and capacitance between AC and DC connections: The resistance and capacitance between the metalization layer (AC connection) and the silicon (DC connection) is measured. Possible defects in the silicon dioxide insulation layer can be probed for this way.

The tests are performed in a cleanroom environment. During the tests, connections to the sensor are established under a microscope using probe needles that can be positioned precisely. As these needles already inflict damage to the passivation and metalization layers, care is taken to only utilize connection areas that are not meant to be used for the wire bonds of the detector.

Figure 4.7 shows the results of the tests for four selected sensors, which have been considered good to take into operation. Firstly it is visible that the IV characteristics for the diodes show very low leakage currents, well below 0.1 nA and their curves loosely resemble the $\sqrt{V}$ shape. For the IV characteristic of the GR, the situation is more striking in that the current strongly increases above 900 V. This is an indication that the breakdown voltage is close. However, the current still stays extremely small, on the order of 10 nA. For a small number of sensors the IV characteristic has been measured also before the dicing of the sensor, which is the process in which the sensor is cut out from the silicon wafer. The process is done using a precision saw, but might inflict defects to the edges of

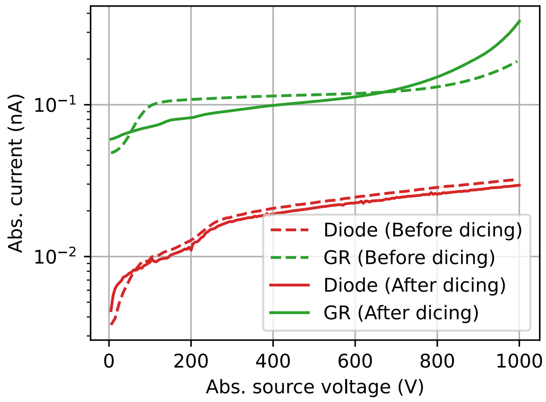


Figure 4.8: IV characteristic for one sensor before and after dicing. The behavior of the diode shows almost no change, while the guard ring's curve changed with a stronger increase of currents for very high voltages.

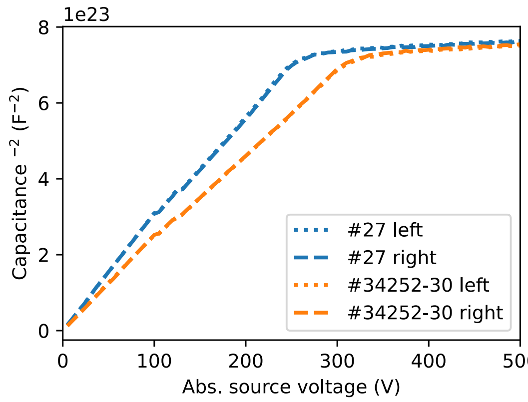


Figure 4.9: Differences in V_{FD} observed between different batches of wafer productions as seen from the CV characteristic. The earlier batch (sensor #27) has a lower V_{FD} than a later one (sensor #34252-30).

the sensor. As seen in figure 4.8, the characteristic of the diode itself practically stayed untouched, while the GR is affected very slightly. This effect is expected and within acceptable margins.

The CV characteristic in figure 4.7 is shown as Capacitance^{-2}. It shows a linear increase until V_{FD} and then suddenly becomes almost constant. This is consistent with what was described in section 4.3.2, where it was described that the voltage becomes constant at full depletion. V_{FD} can be identified to be around 260 V for most of the produced sensors. However, it was observed that sensors from one batch of wafers have an increased V_{FD} of around 300 V. A comparison between sensors from different batches can be found in figure 4.9. The capacitance of all normally performing sensors is determined to be approximately 1.16 pF at the full deletion voltage.

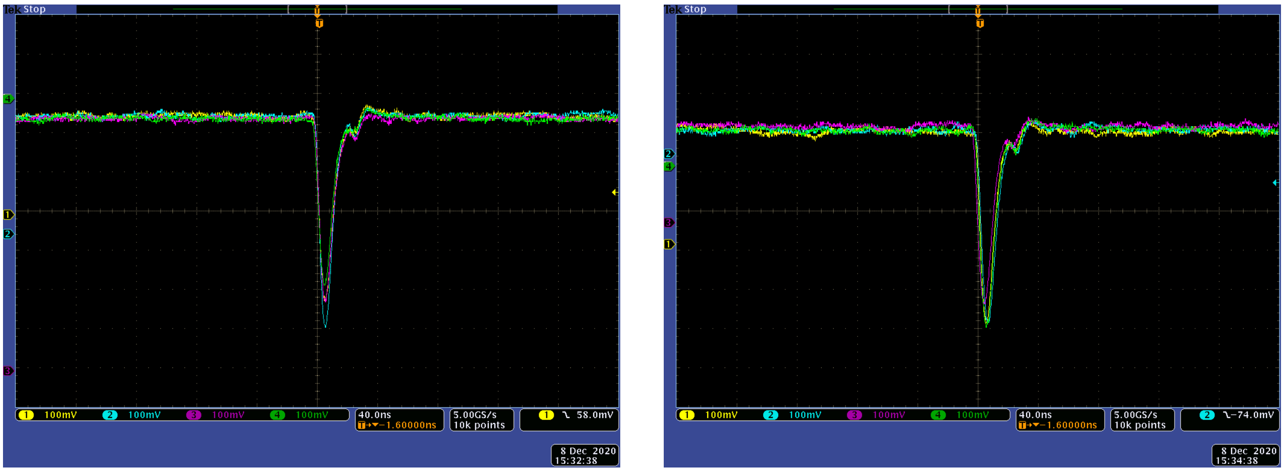
The capacitance of the normally performing diodes between the AC and DC connectors has a mean value of 361 pF, as can be seen in figure 4.7d. The resistance between the AC and DC connectors is measured to be well above 100 MΩ, fulfilling the requirements for the C-Shape circuit.

4.3.5 Frontend test pulses

After equipping the C-Shape PCBs with electrical components, the electronics are first tested by directly connecting the optical output signal to an oscilloscope and verifying the presence and shape of the test pulse generated by the ASICs. Figure 4.10 shows the test pulses obtained this way for 8 different channels on one C-Shape. All test pulses are clearly visible. A small ringing effect is following each of the test pulses for around 100 ns afterwards. Practically all of these oscillations are close to the magnitude of the noise and can therefore be disregarded. Each test pulse has a different amplitude, which is expected, as the different AOHs that are used for the different channels come with varying efficiencies. Overall the amplitudes are fairly large and also no performance issues are observed.

4.4 The BCM1F detector backend electronics

From the frontend, the optical signal arrives in the backend in the CMS service cavern, located underground, next to the experimental cavern. The backend consists of two branches, one using the VERSA Module Eurocard (VME) bus standard and another one using the Micro Telecommunications



(a) (b)

Figure 4.10: Test pulses as coming directly from the frontend for 8 channels on one tested C-Shape. The test pulses have been recorded by an 4 channel oscilloscope with an optical probe.

Computing Architecture (MicroTCA). While both aim to fulfill the same purposes, the MicroTCA has only been implemented over the recent years with the goal to replace old hardware in the future. For this reason the VME backend was used to provide the measurement throughout all of 2022, and the MicroTCA backend was operated in test mode. In the following the VME backend infrastructure is described in detail.

The signal goes through the following stages in the VME backend:

- Four CMS Opto Bahn receivers receive the optical signal of the 48 channels from the frontend (12 channels per receiver) and for each channel convert it into two duplicate electrical signals [68]. The only difference of the two signals is the polarity: One is using a negative voltage, while the other one is using a positive voltage. The positive signal is connected to the MicroTCA branch, while the negative one is used for the VME branch. For this reason peaks will appear inverted in the VME backend.

- The signal is duplicated again using 12 analog fan-in/out modules, each module comprises 4 individual fan-in/out channels. Each fan-out has four outputs of which one is AC coupled. The AC coupled one is used for the pulse discrimination.

- One copy of the signal is delivered to analog-to-digital converters (ADC). These ADCs are 6 CAEN V1721 modules and have a sampling rate of 500 MHz. The LHC clock is used to synchronize the ADCs and they are triggered either by the LHC orbit, the BPTX “B1 OR B2” signal (see section 4.5.3) or an internal threshold trigger. The modules are interconnected via a daisy chain using optical fibers and are read out by a dedicated optical VME link. The setup achieves transfer speeds of up to 80 MB s^{-1}, equaling to approximately 28 full orbits that can be recorded per second. The obtained data represents the voltage value of each channel measured every 2 ns and is used for calibration and diagnostics of the detector.

- The other copy of the signal arrives at three CAEN V895 discriminators (16 channels each). These discriminators output a pulse whenever the negative signal voltage goes below a threshold. The output pulse width and threshold can be adjusted. The discriminator is able to detect a threshold crossing only every 7 ns, limiting the single channel hit resolution to this value. The output pulse width is set to the smallest possible, which is 5 ns, to get a pulse that is smaller than the 7 ns resolution. The threshold is adjustable on a 1 mV precision and its determination is described in section 4.5.5.

- The pulses from the discriminators are delivered to 6 Real-Time Histogramming Units (RHU), which are custom made [69] for BCM1F within the CMS collaboration. The RHUs are FPGA based digital recorders that count the number of pulses per 6.25 ns time interval (called "bin"). The RHUs run an embedded Linux system and are connected via Ethernet, enabling the readout via TCP and access via SSH. The RHUs obtain central CMS information (such as fill number or run number) from the CMS Timing and Control Distribution System [70] and are timed by the LHC orbit trigger and bunch clock. Readout is done every 2^{12} orbits (one "nibble", approximately 0.3 s) and the counts are summed over all of these orbits. The summation discards the information of the exact orbit in which a hit was present but keeps the relative timing of the hit to the start of the orbit. While the RHU can also be configured to count each time a bin contains any part of the input pulse, it is set to count only on the rising edge of the pulse, to get an optimal timing resolution. A delay per RHU can be configured in units of bins, which eases the timing calibration described in section 4.5.6.

- The same pulses from the discriminators are also delivered to another module called Look-up-Table, which is a CAEN V1495. Its purpose is to provide logic and multiplicity operations. For example one of its outputs indicates whether particle hits were detected on both sides of the interaction point simultaneously, meaning within the same time window of 6.25 ns. It also provides signals for multiplicities, meaning at least 2, 4 or even more channels with simultaneous hits. These signals are provided to the BPTX backend, where they have to pass a gate. The gate ensures that the signal can only pass for non-colliding bunches (or the first colliding bunch) and it blocks signals outside of a gap of at least 900 ns according to the bunch filling scheme, as measured by BPTX. The result of the gated 2-multiplicity operation is delivered to the L1 trigger system and is the Beam Gas trigger.

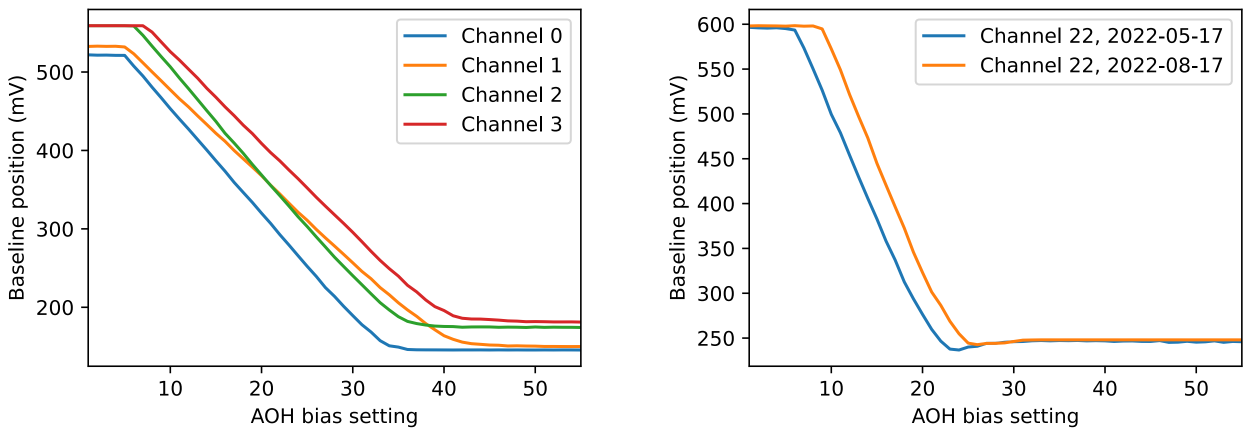
4.5 BCM1F commissioning and calibration

The following sections describe the commissioning and calibration of BCM1F in 2022 when LHC started its Run 3 period.

4.5.1 AOH bias

Signal transmission from the detector frontend inside the experimental cavern to its backend in the service cavern is done via an optical signal. The electrical signal from the sensors is translated into an analog optical one by the AOHs, utilizing laser diodes to produce the light. While the components used for the AOHs are designed to be radiation tolerant, the laser threshold of the diodes still changes over time with exposure to the radiation present inside CMS. For this reason, a bias current can be configured via the serial digital interface of the AOH, to assure that the diode is operated in a region with linear response (input and output are proportional). The BCM1F frontend design allows changing the AOH bias setting at any time.

For the BCM1F system, this linear region is determined using a scan over different AOH bias settings. If the bias is high enough and the diode is in its linear region, the overall signal strength will



(a) (b)

Figure 4.11: AOH bias scan results for four different channels (a) and for two different scans with 2 month in between (b). The baseline is computed by taking the mean signal over 32 orbits. In all cases the region of linear response is clearly visible, however its position and size depends on the channel and time. The baseline decreases (instead of increasing) due to the negative polarity of the signal reaching the ADCs.

change proportionally with respect to the bias. The overall signal strength corresponds to the position of the baseline as seen in the ADCs. In the BCM1F system another effect becomes visible: There exists an optical power at which the optical receiver in the backend saturates and a further increase of the bias settings leads to no further change of the signal. Figure 4.11 shows how the baseline position changes with respect to the bias setting. It can be seen that the linear region depends on the specific channel (due to differences between AOHs) and the date the scan was performed. The latter effect is shown for channel 22 in figure 4.11b, where for a later date the linear region corresponds to a range of higher bias values, possibly due to irradiation.

The bias value set as standard operation configuration is chosen to be close to the start of the linear region, in order to allow a broad spectrum of peaks to be visible in the backend. The value is determined automatically from a fit to the scan results with the function

$$f(x) = \min(\max(c_0 x + c_1, c_2), c_3) \tag{4.4}$$

where c_0 to c_3 are parameters determined by the fit. Start and end of the linear region can then be computed from $(c_3 - c_1)/c_0$ and $(c_2 - c_1)/c_0$ respectively. The bias is chosen such that it is at 10 % of the region's width above the start value.

4.5.2 Individual channel test pulses

The detector can be configured such that test pulses from the frontend are triggered once every orbit inside the abort gap. As there are no collisions during the gap, even during times of colliding beams, the test pulses can be clearly distinguished from actual particle signals. The test pulse can be used to verify whether the channel is connected and how strong its signal is. Figure 4.12 shows the shape for an example channel as received by the backend in the ADCs. The first visible dip in the signal is the actual test pulse. It is followed by a smaller dip at around 100 ns afterwards. The second dip is a technical effect of the pulse generation when the input signal to trigger the pulse turns off.

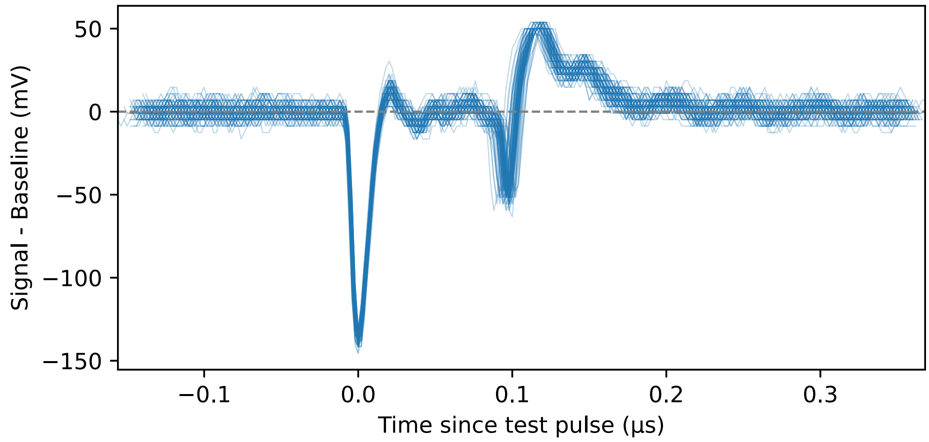
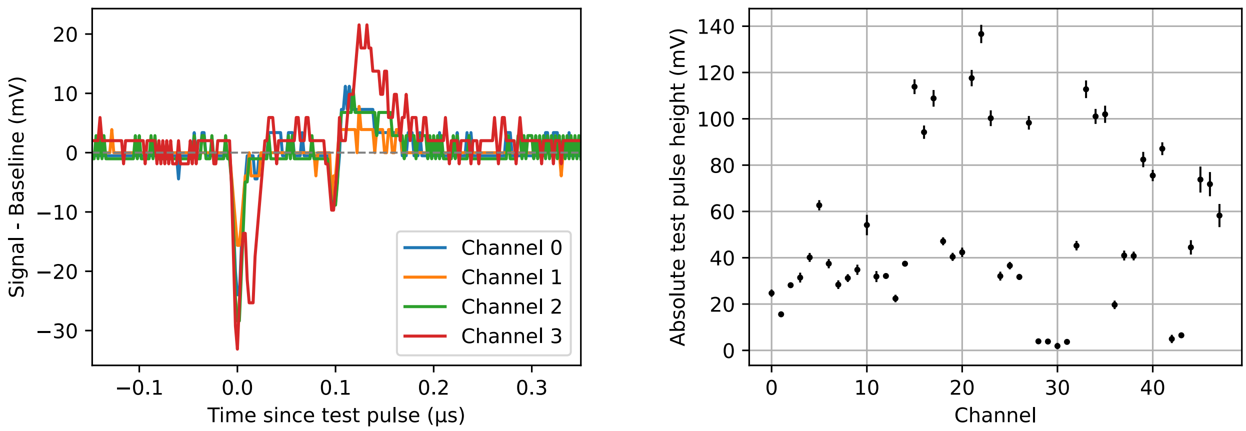


Figure 4.12: Overlay of 100 test pulses of channel 22 as seen in the ADC. Every test pulse is centered so that its peak is at 0 of the time axis.



(a) (b)

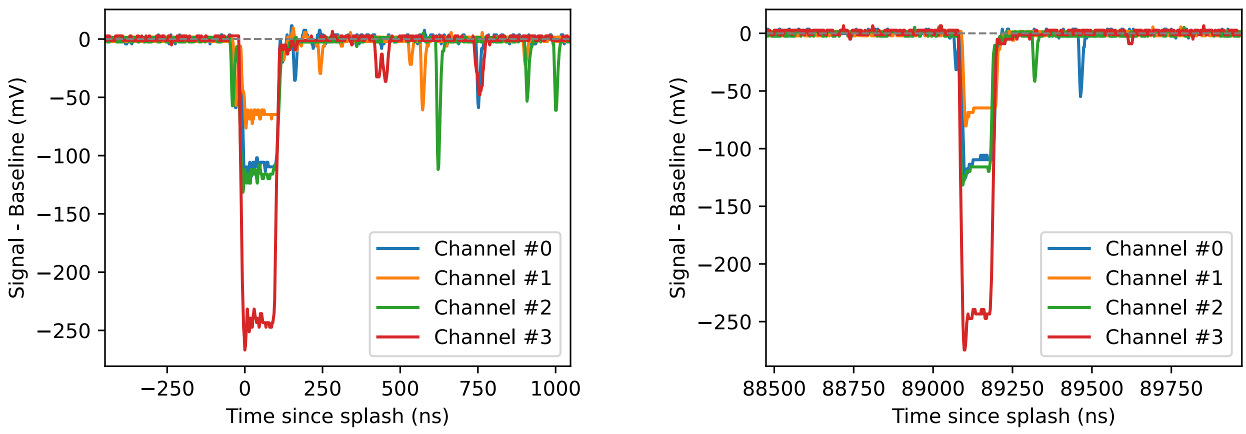
Figure 4.13: Test pulses for four different channels (a) and the height of the test pulses of all channels (b). There are six channels that do not show test pulses and their test pulse heights as shown in (b) are compatible with noise.

Additionally, slight ringing effects after the dips are visible, similar to what was observed in section 4.3.5. Shape and size of the test pulses are seen to have little variance.

In contrast to the consistency of a single channel's test pulses, the pulses can vary significantly across different channels. Figure 4.13 shows example test pulses for different channels and the overall test pulse heights for all channels. Six channels (channels 28, 29, 30, 31, 42, 43) do not show test pulses and their measured test pulse height here is compatible with noise. This indicates a connection or technical issue with these channels. Furthermore, channel 1 shows a rather small test pulse, which may point to bad signal transmission. These channels are further investigated in the following sections.

4.5.3 Splashes

The generation of so called splashes is part of the commissioning program of the LHC and was done in October 2021 and May 2022. A splash is generated by the LHC injecting a single bunch and closing



(a) (b)

Figure 4.14: Splash generated at the 7th of May 2022 at 11 am Geneva time. Figure (a) shows the primary splash, while (b) shows a secondary splash one orbit period after the primary splash. The peaks are cut off due to saturation in the readout.

the collimator closest to the experiment. The particles are stopped and a large shower of particles hitting the detector is created. This step serves as the first verification of the detector's response to particles outside the laboratory.

The splashes were recorded for analysis using the ADCs in conjunction with the BPTX "B1 OR B2" trigger signal. This signal is delivered to the ADCs via their external trigger input and fires for every bunch that is detected by BPTX, regardless of whether it is part of beam 1 or beam 2. Figure 4.14 shows the signals received for one splash for four BCM1F channels. First, the splash generates a signal strong enough to saturate the readout, creating a peak with a width of around 150 ns and a flattened out top. In the next few microseconds this is followed by particle hits in some of the sensors, which can be the result of several effects, including slower or scattered particles or possibly even activated material. Then a secondary splash signal is visible around 89 µs afterwards. This secondary one is created by the particles that were able to pass the collimator once and only interacted with it after circulating through LHC. Naturally the number of the particles in the secondary splash is smaller than that in the primary splash. It can be seen that even in the case of the secondary splash, the signal is strong enough to saturate all the channels.

The splashes have been observed on all channels expect for channel 42 and 43. These two channels, which are one double diode, are likely are not properly connected and cannot be used.

4.5.4 Peak distributions

To finalize the qualification of each channel of the detector, the distribution of the peak height a particle creates is highly useful. The overall height is expected to decrease with detector age, for example due to irradiation effects of the AOH or the sensor. Additionally, clear separation between electronic noise and signals coming from MIPs is required. The position of the MIPs in the height distribution depends on the size of the depletion region in the sensor diodes and therefore on the HV that is applied. Generally the detector should be operated with high enough voltage, so that the diodes are fully depleted and the MIPs are well separated from the noise.

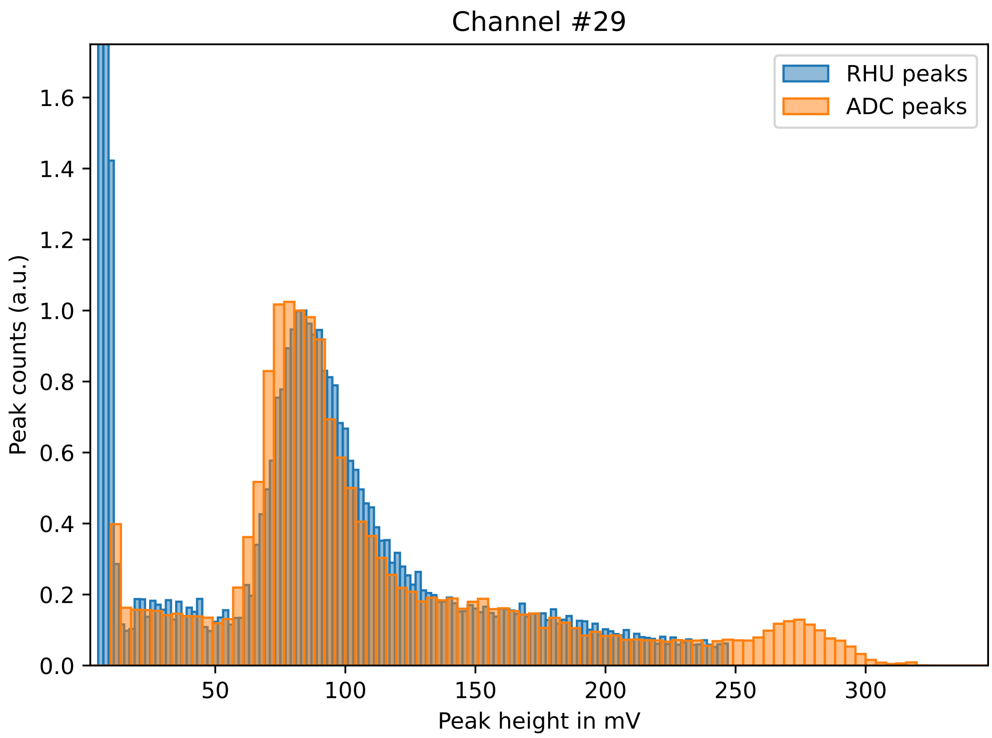
The peak height distribution can be obtained either by analyzing the data obtained from the ADCs or from the counts reported by the RHUs while performing a scan of many different discriminator threshold values.

- Using the data taken by the ADCs, the first step is to determine the per-channel baseline position by taking the mean of the data of each channel. Then all local minima are found (defined by having a larger amplitude than both their neighboring samples). Finally all local minima that are smaller than 7.8 mV (2 ADC counts) are discarded. The heights of all the remaining minima are filled into a histogram. The data needed for this method should be recorded at a time of stable beam inside LHC and high collision rates using the orbit trigger (at least 1000 colliding bunches per orbit for a measurement on the order of minutes). Alternatively, the internal ADC trigger can also be used in the case of low collision rates.

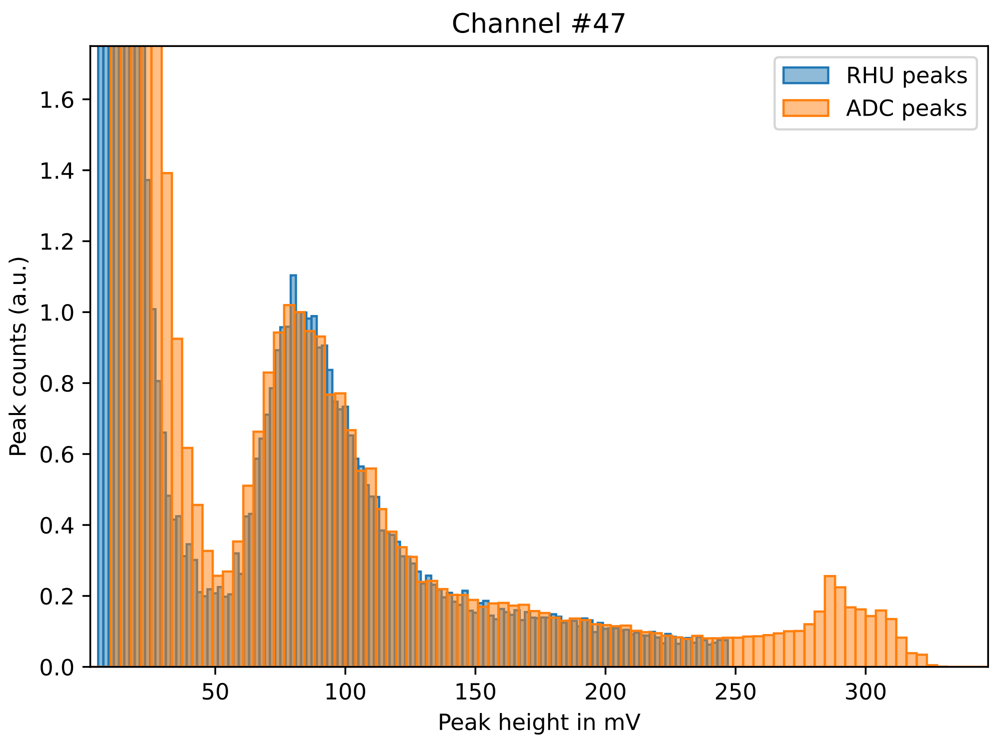
- Using the RHUs, counts are recorded for a range of different discriminator threshold voltages. By taking the difference between the counts obtained for two neighboring threshold values, one obtains the counts of hits at a specific peak height. This approach also requires a high collision rate. Counts taken from a longer time interval can be summed to decrease statistical uncertainty. However, as the instantaneous luminosity decreases over time, the measurement will be biased if performed over too long a time interval. This scan cannot be performed during luminosity/background measurement because for proper readout the discriminator threshold must be fixed to specific values (see section 4.5.5).

By performing both approaches at the same time, possible differences, for example due to technical issues, between the two backend branches can be probed. Additionally, the method based on RHU counts can deliver a more precise spectrum, as the discriminators are adjustable in units of 1 mV, while the ADCs are operating on an approximately 4 mV basis. On the other hand, the discriminators are limited to a maximum of 255 mV while the ADCs have a range of 1 V.

The resulting distributions are shown for two channels in figure 4.15. Both approaches, ADC and RHU overlap nicely and thus it can be verified both backend branches yield the same results. The first feature seen on the very left of the distribution is a maximum that increases for even smaller voltages. This maximum is merely the result of noise and is cut off due to the minimum peak height requirement of 6 mV (7.8 mV in the case of the ADCs). The next feature is located at around 80 mV for these particular two channels and is referred to as the MIP peak. It is created by MIPs, but also by particles of higher energies, crossing the sensor. At twice the voltage, at around 160 mV, a small bump in the distribution is visible. The peaks of that size are caused by two particles, in particular two MIPs, crossing the sensor at the same time. The last feature to the very right of the distribution is another maximum, which is caused by the optical receiver saturating and not producing voltages of higher magnitude. This property of the receivers makes peaks, that are of different heights in the front end, appear to be of same high in the backend. In this particular example the position of the MIP peak as well as the saturation maximum are at very similar positions for the two channels. However, in general these positions may vary significantly from channel to channel. What is visible is the difference in noise: Channel 47's noise comes close to the MIP peak, while there is clear separation between noise and MIP in the case of channel 29. Channel 47 will thus require a higher discriminator threshold setting during operation, as discussed in the next section.



(a)



(b)

Figure 4.15: Peak height distribution of BCM1F channel 29 and 47 taken using the ADCs and the RHUs. Both histograms have been normalized to allow comparison. The RHU data has been obtained with a scan from 6 mV to 250 mV, a step size of 2 mV and by summing 2^{15} orbits while there were around 1200 colliding orbits in LHC. The ADC data has been obtained simultaneously.

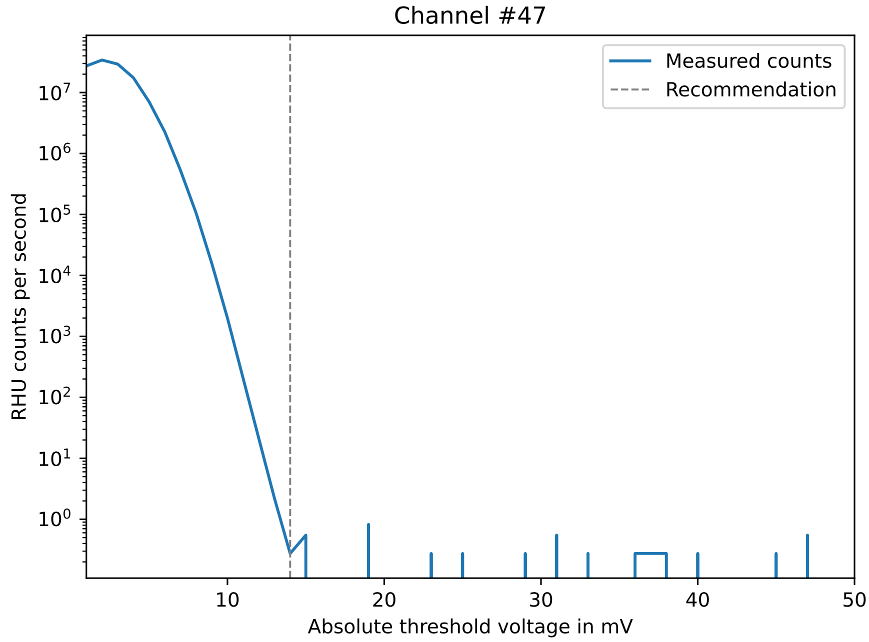


Figure 4.16: Threshold scan without beam presence to estimate the height of the noise for channel 47. The dashed gray line shows an automatically determined, "recommended" threshold value for which the noise counts are below 10 Hz. The counts were summed over an interval of $2^{12} \cdot 10$ orbits per threshold setting.

4.5.5 Discriminator threshold

Particle hit counts per time interval serve as the input for the background and luminosity measurements. Because peaks in the signal are not only the result of particles hitting the detector, but can also simply be the result of noise, it is necessary to differentiate between these two types of peaks. As described in section 4.5.4, noise peaks are generally of significantly lower height than particle peaks. While in general a more sophisticated approach is possible, as for example done in the MicroTCA backend branch, all peaks with heights above a certain per-channel threshold are considered to originate from particles. In the VME backend this is done using the discriminators. The thresholds are adjusted such that they are placing in the local minimum between the MIP peak and the noise (see figure 4.15)

In order to operate the detector before the instantaneous luminosity reached high enough values to perform the peak distribution analysis, a different approach was used. By scanning through the threshold values, as already described as the RHU approach in section 4.5.4, but without the presence of a beam, the height of the noise peaks can be estimated. The threshold is then set to a value where the counts from noise are below a certain rate, for example 10 Hz. Figure 4.16 shows the resulting noise count as a function of the threshold value for one example channel. A sufficiently low count can already be achieved at around 15 mV. The figure also shows small bumps for single scan points. This is the result of the scan being done in a limited amount of time, which limits the statistical precision of the low noise rates that appear at high threshold settings.

4.5.6 Timing

BCM1F is able to determine per-bunch luminosity and is able to tell apart background particles from collision products by their time of arrival. However, this technique requires a precise timing, so that

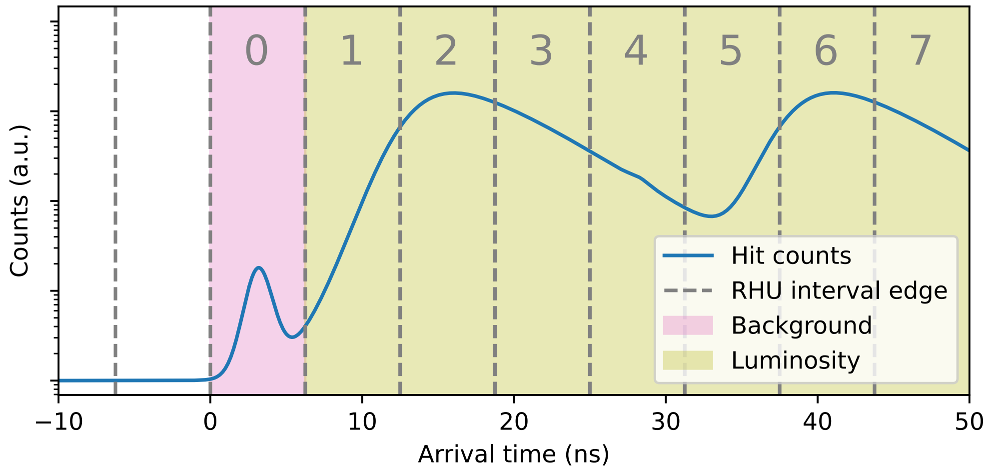


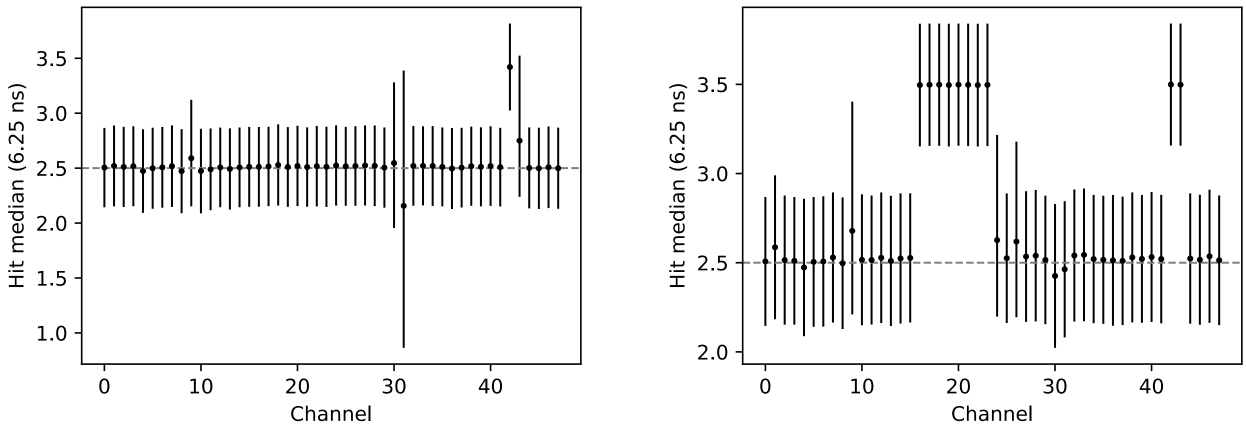
Figure 4.17: Illustration of an optimal hit timing in the RHUs. The blue line illustrates the signal counts coming from the first two bunches of a bunch train. The 0th RHU bin counts the BIB hits, while subsequent bins are used for the luminosity measurement.

particle signals arrive in the RHUs, where hits are associated with time information, at the correct moment. Figure 4.17 illustrates at which timing the signal should arrive. In reality, however, the arrival time of the signal can be shifted by an arbitrary amount and thus a timing calibration is necessary.

In order to achieve an optimal timing alignment of the detector, the following procedure is employed. The first step is to analyze data recorded from the RHUs during collisions or circulating beams (beams that are non-colliding). From the data, the median position of the hits within one bunch crossing (25 ns) is obtained, of which two example results for all channels can be seen in figure 4.18. The optimal timing is reached when the median is in the middle of the second RHU bin. Figure 4.18b shows an offset of one RHU bin for channels 16 to 23, corresponding to one RHU module. This can be accounted for by the RHU software configuration, which can add offsets in units of 6.25 ns. In the figure channel 42 and 43 only show noise and thus their alignment is arbitrary. The position of the first bunch crossing of the LHC filling scheme can also be determined from the data. Knowing these positions, the first timing correction is done by configuring the orbit trigger delay of each RHU, shifting the timing of all the channels per RHU module in units of RHU bins. In the next step a fine tuning of the timing for every channel is performed by inserting or removing cable length in the connections between backend modules. This way, adjustments even below ns level become possible. For this type of fine tuning, the ratio between the second RHU bin and its adjacent bins, as shown in figure 4.19, indicates channels that require improvement. For example it becomes apparent that channel 4 has is slightly misalignment.

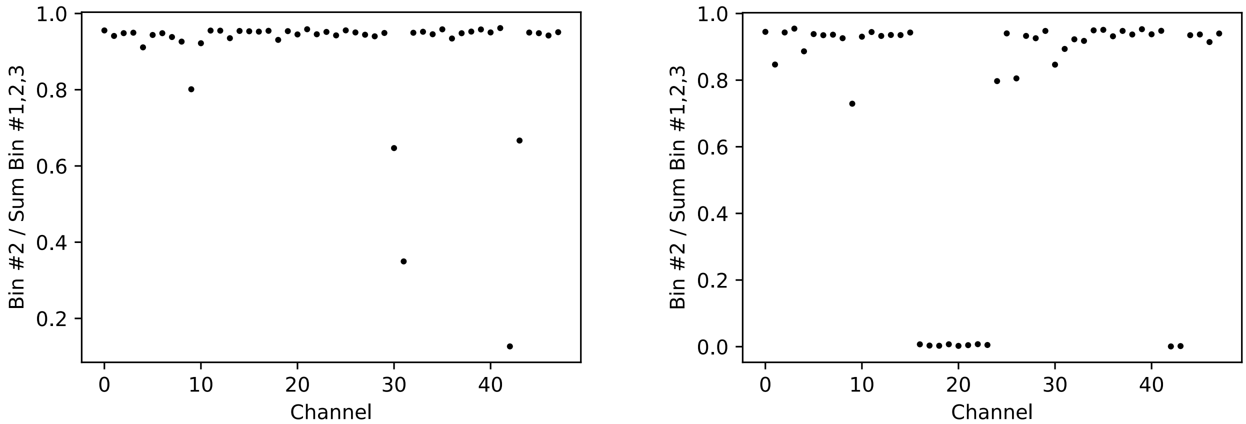
4.5.7 Channel efficiency correction and per-channel non-linearity

Over the duration of the detector operation it might become necessary to exclude a specific channel from the computation of the luminosity or the background. Because the results are presented as average over all channels, if that channel was over- or underefficient, this average will be affected. Additionally the background measurement depends on which side of the interaction point the sensors are positioned at. If all channels in a position were less efficient than in the other, the background measurement would become inconsistent. For these reasons the counts of each channel is scaled by



(a) (b)

Figure 4.18: Median hit positions from collisions. The error bars indicate 68 % intervals. One unit of the y-axis corresponds to 6.25 ns in time. The medians have been computed from a linear interpolation of the commutative sum of the counts given in the RHU intervals of 6.25 ns. a) with close to optimal timing adjustment and b) using data from a period when one RHU was misaligned.



(a) (b)

Figure 4.19: Ratio of the main RHU bin (the 2nd bin) and itself plus its adjacent bins. The detector configurations are the same as in figure 4.18 for (a) and (b) respectively. The ratio clearly shows even small timing shifts: Channel 9 can be easily observed to have counts outside the main bin.

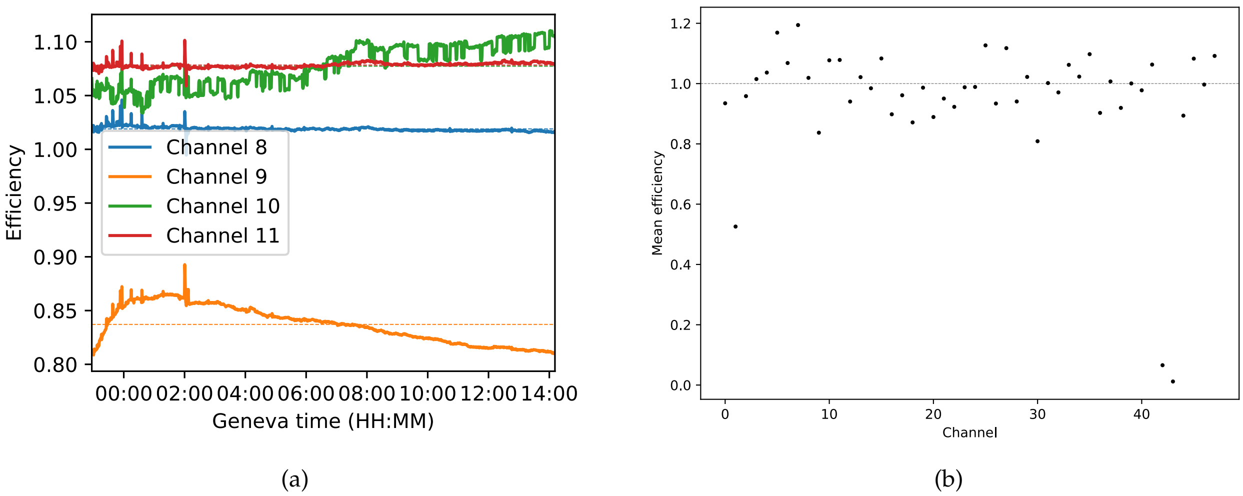


Figure 4.20: Channel efficiencies for one fill. (a) shows the evolution of the efficiencies of four selected channels over time. The data was taken on the 22nd and 23rd of July 2022. In this time the intensity of the beams is slowly decreasing as the beam particles constantly collide. The dashed lines indicate the mean efficiencies over this time. It can be seen that channel 9 and 10 deviate from their means and show non-linear effects. (b) shows the mean efficiencies of this time of all channels. Channel 42 and 43 only provide noise as stated before and therefore their efficiencies deviate strongly from 1.

its efficiency to correct for these differences.

The efficiency is derived by dividing the measured rate of each channel by the average rate of all channels. If a particular channel has a non-linear response to the particle flux, its efficiency will also show this dependency. This assumes that overall the channels show a linear response. Conversely, the non-linearity can be studied by comparing the evolution of the efficiency over the duration of one fill, as LHC's beams slowly loose intensity until new beam particles are filled. Figure 4.20 shows this evolution together with the efficiencies of all channels. Channel 9 and 10 clearly show a trend and noisy behavior and channel 1 is rather underefficient. However, overall the efficiencies are close to unity, which indicates that the channels are generally similar in response. The non-linear behavior observed here could be levitated by improving the discriminator thresholds, which decreased the amount of noise that was counted.

4.5.8 High voltage scan and radiation effects

The HV applied to the sensors should be larger than the full depletion voltage, as described in section 4.3.2. To ensure this for BCM1F in 2022 the HV was first set to 300 V and later increased to 400 V. To monitor the required voltage setting, HV scans were performed in which ADC data is recorded at different HV settings. By comparing the resulting spectra of the peak heights, the dependence on the HV can be observed. Before and at the beginning of the LHC run, no significant changes were observed above 300 V. However, as can be seen in figure 4.21, at a later point in time, after LHC delivered an integrated luminosity of approximately 30 fb^{-1}, there is a dependence of possibly up to 350 V. Before 350 V, the MIP peak height increases with voltage, while the peaks for 350 V and 400 V overlap almost perfectly. This clearly shows the increase of the full depletion voltage resulting from effects of accumulated radiation and it can be expected to further increase with more radiation.

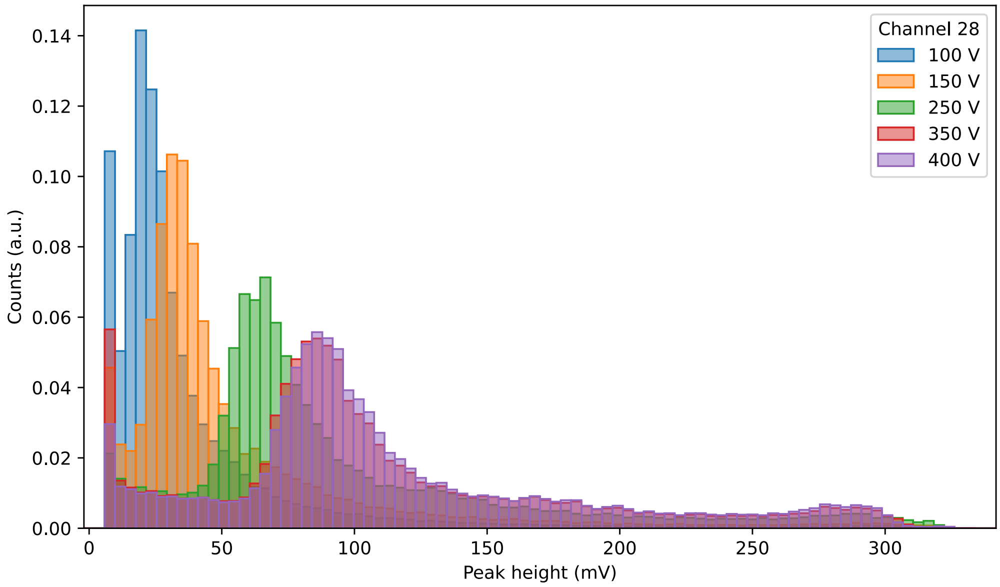


Figure 4.21: Peak height spectrum for different bias voltages, showing the dependence of the MIP peak position on the applied HV. Until the full depletion voltage is reached, the depletion region in the diode can extend, increasing the amount of charge a particle deposits in the diode. Note that the data was taken after approximately 30 fb^{-1} of integrated luminosity.

4.6 Background measurement

BCM1F is responsible for providing the two beam-induced background (BIB) measurements (commonly referred to as BKGD 1 and 2) that are also reported in real time to the CMS and LHC control rooms. This background arises from the operation of the LHC itself and its main source stems from interactions of the beam particles with residual gas inside the beam pipe [71]. A high BIB is an indication of bad beam quality, which can negatively influence the data taking of the CMS detector. In the worst case scenario, it could damage parts of the detector, such as the tracker. While a different system (BCML) provides protection in the case of intense BIB, BCM1F is highly sensitive to small increases of BIB and functions as an early warning system of adverse situations.

The background is measured by BCM1F by taking advantage of the fine time resolution of the BCM1F readout system. Before arriving at the interaction point for a collision, the beam particles, together with the BIB, cross the position of BCM1F. Due to the chosen location along the z axis for BCM1F, this happens 12.5 ns before the particles from the collision arrive. The RHUs divide this 25 ns interval into four 6.25 ns bins and the particle hit counts from the bin from 12.5 ns before the majority of the collision products arrive is used to compute the background. This is illustrated in figure 4.17, where it can be seen that a small peak in counts coming from the BIB, which is several orders of magnitude smaller than the counts of the collision, is located in the bin used for background. However, this method only works if there are no collisions from the bunch crossings before, as otherwise the collision products arriving at the same time vastly outweigh the BIB particles. For this reason, BIB is only measured in bins with filled buckets after 750 ns of no collisions, which corresponds to the injection gap between bunch trains in the usual filling scheme of LHC (see section 3.1). Because the background is measured in the bin right before the particles from the collision start arriving, which are much larger in number, is has a strong dependence on the correct timing alignment of the BCM1F system.

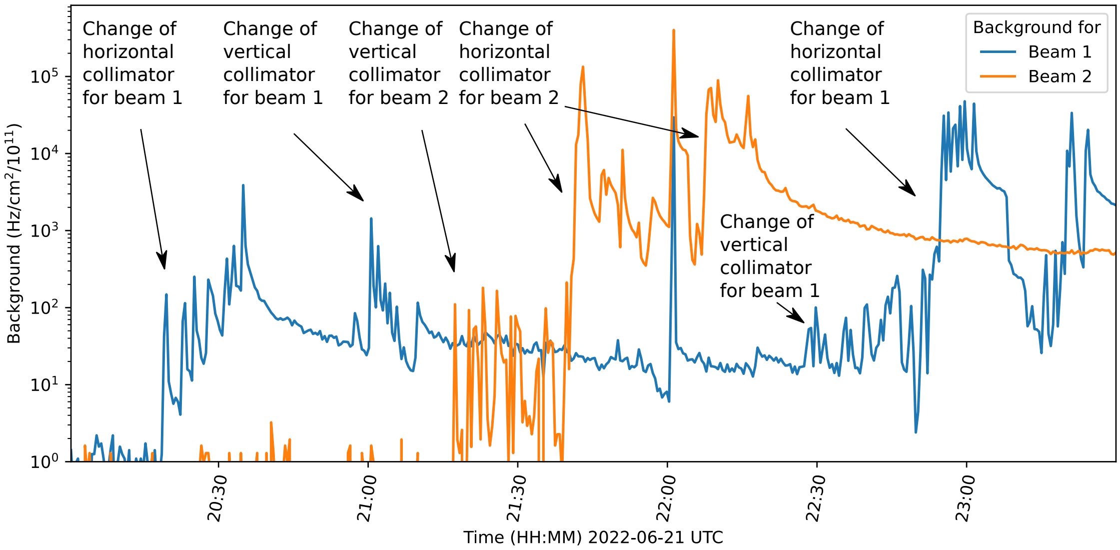


Figure 4.22: BIB measured by BCM1F during loss map measurement on the 21st of June 2022 (fill 7825). Indicated are all position changes of the tertiary collimators closest to CMS. Each change in position of a collimator has a strong impact on the background measurement.

Two independent background measurements are computed, one for each beam. For beam 1 this is done by using only the channels located on the side from which beam 1 is coming from and likewise for beam 2. The background is reported as a ratio to the beam intensity (in number of protons) and total BCM1F sensor size, as both quantities increase the measured counts linearly.

During the LHC commissioning program, so called loss maps are measured. During these loss map measurements, the collimator positions are systematically changed to determine their position with respect to the beam position. When a collimator comes very close to the beam, it creates a spray of particles that are measured as background by BCM1F. An example behavior during such a loss map measurement is shown in figure 4.22, where one can see that every time a collimator close to CMS is changing its position, a strong change of the BIB is measured.

4.7 Luminosity measurement

The measured instantaneous luminosity L is determined from the formula

$$L = \frac{\mu f_{\mathrm{LHC}}}{\sigma_{\mathrm{vis}}} \tag{4.5}$$

where μ is the per-bunch hit count from the detector, f_{LHC} is the LHC revolution frequency and σ_{vis} is the visible cross section, a proportionality constant that is determined from a dedicated measurement.

4.7.1 Visible cross section and scans

In order to obtain a high-quality luminosity measurement with a small uncertainty, the precise knowledge σ_{vis} is critical. To this end LHC allocates time to special beam configurations, called Van der Meer (VdM) scan programs, in which σ_{vis} and its uncertainties are determined. These programs usually take several days including preparations and include a range of different scans, in which the beams are not colliding head on, but rather are slightly separated from each other in several steps.

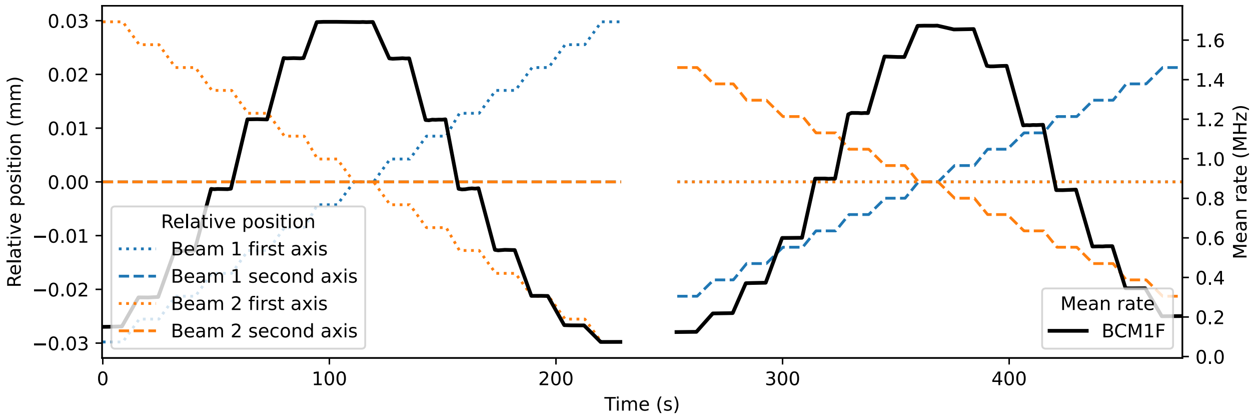


Figure 4.23: Emittance scan at the beginning of a fill. The transverse separation between the two beams is changed by synchronously changing the position of the beam axes in one direction (x or y). After this is done in one direction it is repeated in the orthogonal direction. At zero separation, the rate from the detector is at its maximum.

In each separation step the collision rate is measured. In the name giving VdM scan, the position of both beams are changed synchronously. This is done separately in the horizontal and vertical axes. Assuming head on collisions this procedure gives enough information to determine σ_{vis} on a bunch-by-bunch basis by using above formula and the relation

$$L = f_{\text{LHC}} \frac{n_1 n_2}{A_{\text{Overlap}}} \tag{4.6}$$

where n_1 and n_2 are the number of particles in each bunch and A_{Overlap} is the overlap area of the two beams when being head on (zero separation) [72]. A_{Overlap} describes the effective transverse area in which the collisions take place and is computed from the VdM scan. Depending on the transverse shape of the beams, the measured hit rate will decrease with increasing separation, making it possible to determine A_{Overlap} from the scan. However, the procedure requires large separations of the beams (at the LHC up to 6σ of the beam width, assuming Gaussian shapes). This was one reason why VdM scans were not used for example at the Sp$\bar{\text{p}}$S at CERN or the Tevatron at Fermilab, where the could not be separated as much due to technical limitations [73]. A range of corrections and sources of systematic uncertainties is associated to this procedure such as uncertainties in the true beam separation, a bias from the assumption that the beam shape can be factorized in horizontal and vertical parts, or from deflection occurring between the two beams [74]. The VdM program includes various scans of the beam positions from which most corrections and uncertainties are determined.

The program is rarely performed (usually only once a year) due to its large time requirement, another method to determine σ_{vis} is employed, which is utilizing emittance scans. Emittance scans can be seen as very short VdM scans and are performed by the LHC usually at the start of a fill and shortly before the regular end of it. Although the resulting σ_{vis} comes with much higher uncertainties, the short time required for the emittance scans makes it feasible to perform them regularly. This also enables regular performance monitoring of the luminometers as well as monitoring possible gradual changes. Like in a VdM scan, in an emittance scan the separation between the beams is gradually increased, first in one then in the other direction, as can be seen in figure 4.23. At the lowest separation the hit rate reported by the detector reaches its maximum, while it continues to decrease for higher separations. The resulting hit rate curve is fitted, usually with a Gaussian distribution. Using the widths obtained from the fit, A_{Overlap} can be calculated.

4.7.2 Detector hit count via zero counting

The luminosity is proportional to the amount of hits per time, passing the detector. One could directly take the counted hits and compute the luminosity from that. However, this quantity has a shortcoming: The detector has a dead time, meaning it is limited by how early another particle is allowed to arrive after a particle was already counted. The discriminators in the backend can only detect a hit every 7 ns and the FWHM of the signal peaks produced by the front end ASIC is around 10 ns. This means if two particles arrive simultaneously within 5 ns of each other, only one count will be registered.

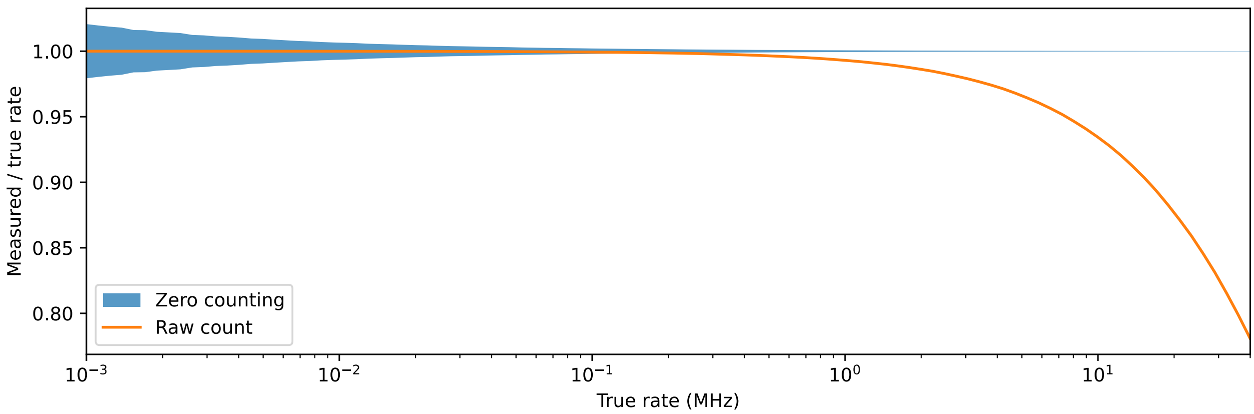


Figure 4.24: Raw detector counts and zero counting assuming a dead time of 7 ns and $n_{\text{max}} = 2^{14}$. The standard deviation of the zero counting result is indicated by line width. The raw count quickly deviates at higher rates, while the zero counting yields results close to the true rate. However, for very low rates the zero counting's uncertainty increases, as indicated by the line thickness.

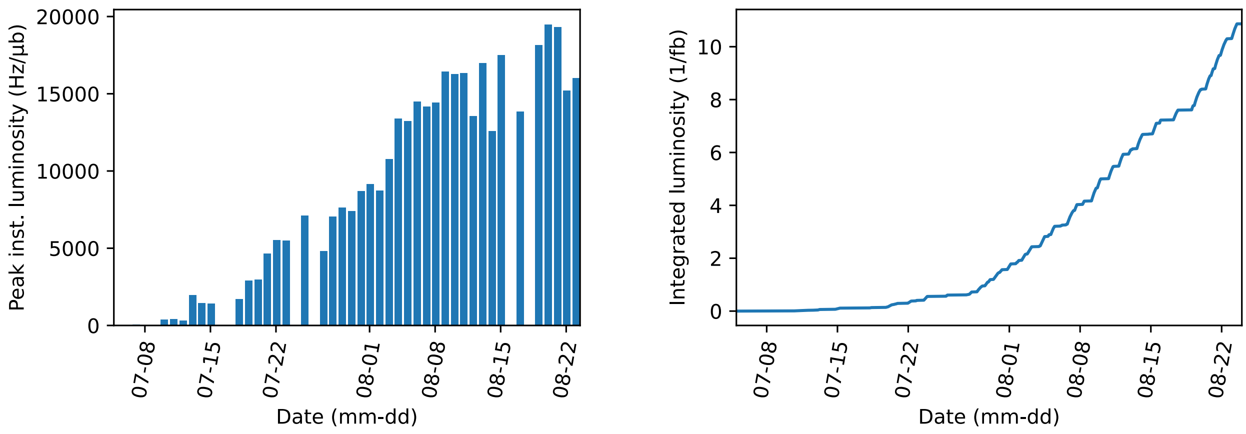
Instead of simply using the raw detector hit count, a better estimate can be obtained by using a technique called zero counting. Thanks to the fact that luminosity only needs to be reported every 2^{14} orbits ≈ 1.46 s (referred to as NB4), the zero counting can profit from the additional information coming from averaging over many orbits. The counts measured in a particular time interval are Poisson distributed and therefore the probability to get k counts when the expected value is μ can be calculated from

$$P(\text{counts} = k) = \frac{\mu^k \exp(-\mu)}{k!}. \tag{4.7}$$

The detector measures at most one count within its dead time. By taking the counts of many dead time windows, one can approximate the probability to get at least one count with

$$P(\text{counts} \geq 1) \approx \frac{n_{\text{detect}}}{n_{\text{max}}} \tag{4.8}$$

where for BCM1F n_{detect} is the count of a single dead time window over the entire NB4 and $n_{\text{max}} = 2^{14}$



(a) (b)

Figure 4.25: Luminosity measured by BCM1F as published online during running. The peak instantaneous luminosity on a day-by-day basis is shown in (a) while (b) shows the integrated luminosity.

is the number of orbits in an NB4. Plugging both relations, one obtains

$$\begin{aligned} 1 - \frac{n_{\text{detect}}}{n_{\text{max}}} &\approx 1 - P(\text{counts} \geq 1) && (4.9)\\ &= P(\text{counts} = 0) && (4.10)\\ &= \frac{\mu^0 \exp(-\mu)}{0!} && (4.11)\\ &= \exp(-\mu), && (4.12) \end{aligned}$$

and finally

$$\mu \approx -\ln\left(1 - \frac{n_{\text{detect}}}{n_{\text{max}}}\right). \quad (4.13)$$

Thus one can estimate the mean number of counts by zero counting, even if each detector channel can only count single hits. Figure 4.24 compares the raw detector counts to the result from zero counting for different orders of magnitude of hit rates. The raw detector rate quickly deviates from the true rate towards higher rates due to multiple particles arriving at once. On the other hand towards lower rates, the uncertainty of zero counting becomes significant. This is due to the rising imprecision of the approximation of $P(\text{counts} \geq 1)$ and gives a lower limit to when zero counting is applicable. Furthermore zero counting gives very precise results for high rates, where the raw rate is is not usable. However, in the limit of full saturation, when at least a hit is measured in each time interval, the detector response will not change anymore with increasing rates and thus the zero counting no longer be able to provide meaningful results.

4.7.3 Luminosity in 2022

The restart of the LHC was officially launched on the 5th of July 2022, after a shutdown of more than five years [75]. The run started with a very low number of only 3 buckets filled, of which 2 bunches were colliding at CMS. While continuously monitoring the machine, this number was gradually increased over the following months. The increase can be clearly seen in the instantaneous luminosity measured by BCM1F as shown in figure 4.25. While starting with just a few Hz μb^{-1}, the luminosity quickly rose to almost 20 000 Hz μb^{-1} during the last days of August 2022. At that point in time, LHC had already delivered an integrated luminosity of more than 10 fb^{-1}, since the start of July, and the BCM1F detector had continuously delivered precise measurements of the instantaneous luminosity into the CMS and LHC control rooms.

From the instantaneous luminosity one can calculate the pileup using

$$n_{\text{pileup}} = \frac{L\sigma_{\text{tot}}}{f_{\text{LHC}} n_{\text{bunches}}}, \tag{4.14}$$

where n_{bunches} is the number of colliding bunches and σ_{tot} is the total inelastic cross section. Figure 4.26 a to d use this relation to display the pileup in connection to the luminosity. In figure 4.26a the instantaneous luminosity over the entirety of the stable beam time of a fill is shown. "Stable beam" was declared shortly after 19:00, which is visible in the figure right after a sudden increase in luminosity. A that point in time LHC had adjusted the beams to collide head on. Right after luminosity is established and "Stable Beams" are declared, an emittance scan was started around 19:10, lasting a bit less than 10 minutes and is shown in detail in figure 4.26b. The "saw tooth" pattern that is visible during the hours after the emittance scan in figure 4.26a is caused by the β^*-leveling. Here, LHC reduces the β^* value in steps, ensuring that the luminosity stays approximately constant, at around $17\,500 \times 10^{30}\,\text{cm}^{-2}\,\text{s}^{-1}$, corresponding to a pileup of about 50 collisions per bunch crossing. To keep the luminosity constant while the beam intensities decrease, β^* is reduced in several steps until the minimum possible value of 30 cm is reached, which is the case at around 00:30. Afterwards the luminosity decreases exponentially with decreasing beam intensity. Over the time of the stable beam smaller dips in luminosity are visible in irregular intervals. These are short scans to optimize the beam parameters and a detailed view of one of them is shown in figure 4.26c. While for BCM1F the luminosity is computed based on 2^{14} orbits, each of the 25 ns bunch crossings is treated separately. This means from the 2^{14} orbits 3564 instantaneous luminosity values are derived, each corresponding to one specific time offset within the orbit. In the case of BCM1F, the instantaneous luminosity is determined separately for each of the 3564 bunch crossing time slices of 25 ns length each. Of the 3564 slices, in 2022 up to 2700 were filled. The resulting pattern with gaps at regular intervals displays the filling scheme. In figure 4.26d these individual instantaneous luminosities are shown. The bunch trains within the scheme as well as the abort gap at the end become clearly visible (compare to figure 3.2). Note that the filling scheme of this particular fill had only 2450 colliding bunches, hence the larger gap at around bunch crossing 2700.

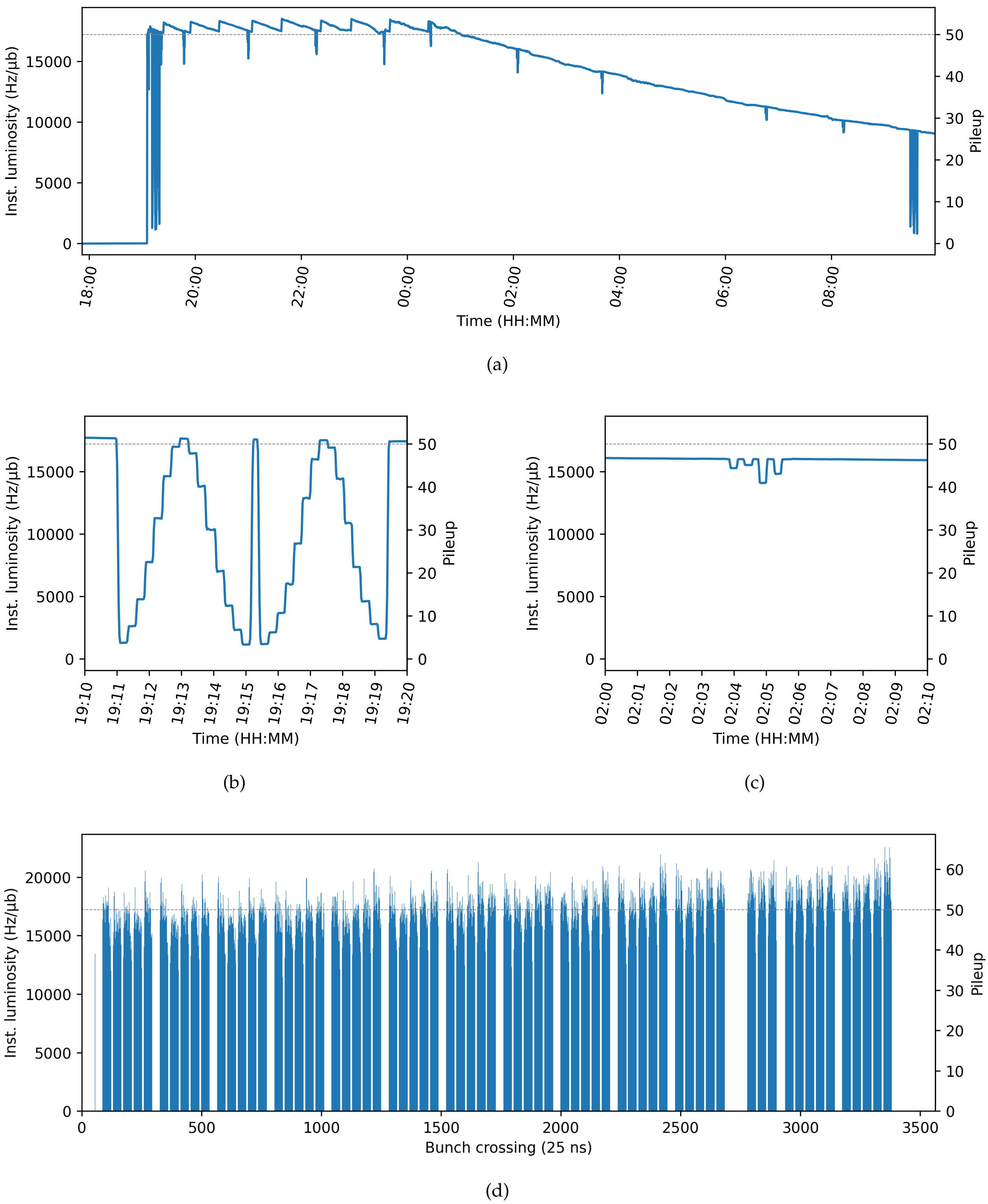


Figure 4.26: Instantaneous luminosity of fill 8333 on different time scales. (a) shows the evolution during the fill of around 16 hours. (b) shows the luminosity during one emittance scan of the fill. (c) shows the luminosity during one beam optimization scan. (d) shows the luminosity on the bunch crossing level. The pileup is computed using a total inelastic cross section of 80 mb.

Chapter 5

Search for heavy Higgs bosons

5.1 Overview

The question on the existence of multiple Higgs bosons has been and still is subject of ongoing research. As pointed out in section 2.8 a large range of BSM theories involve multiple Higgs bosons. So far only a single Higgs boson has been found, which up to now is consistent with the SM Higgs boson [32, 76]. In the context of BSM theories that include heavy Higgs bosons, such as the two Higgs doublet model (2HDM), the SM Higgs boson is associated with the lightest Higgs boson, thus declaring the other Higgs bosons in the models heavy. Searches for the heavy Higgs bosons must overcome a variety of challenges also due to the large available parameter space: Without favoring a particular model, generally the total decay widths, the branching ratios, as well as the masses are free parameters. The latter could be close to the SM Higgs boson mass of 125 GeV, but it could also be on the order of TeV, creating the requirement for very different analysis strategies. Generally, the searches up to this date are in agreement with the SM-only hypothesis, ruling out ever increasing portions of the parameter space. It is also worthwhile to note that there have been a few searches that showed slight tensions with the SM-only hypothesis. In particular, there have been analyses for $\mathrm{A/H} \to \tau\tau$ by the ATLAS and CMS collaborations showing tensions [77, 78]. Moreover, in a previous search for $\mathrm{A/H} \to \mathrm{t\bar{t}}$ CMS observed a slight excess was observed from A decaying to $\mathrm{t\bar{t}}$ with a mass of 400 GeV and a 5 % relative decay width [79]. This previous search used data recorded in the year 2016, corresponding to an integrated luminosity of 35.9 fb^{-1} and obtained a global significance of 1.9 standard deviations (3.5 ± 0.3 standard deviations locally) for the excess. The analysis described in this thesis adapts and evolves the methods previously utilized in the $\mathrm{A/H} \to \mathrm{t\bar{t}}$ publication.

The LHC Run-2, which mainly includes the operation time of 2016, 2017 and 2018, as of now, is the operation period yielding the largest amount of collision data at the, until then, highest center-of-mass energy of 13 TeV. In these three years CMS recorded an integrated luminosity for proton-proton collisions of 138 fb^{-1}[29], much larger than the 35.9 fb^{-1} of the previous analysis in reference [79]. A new analysis, that analyses the combined data of the three years, including a new analysis of the 2016 data, as presented in this thesis, can therefore exhaust the full statistical precision. Additionally in 2021 the CMS collaboration has made available its final and most precise calibration of the Run-2 data, commonly referred to as "Ultra Legacy". The utilization of this calibration is expected to significantly reduce systematic uncertainties.

The analysis in this thesis focuses on the neutral heavy Higgs bosons A and H decaying into a top quark and antiquark ($\mathrm{A/H} \to \mathrm{t\bar{t}}$, see section 2.8 for a Feynman diagram). In particular, the dileptonic decay channel of the top quark pair is probed. The channel gives rise to 2 bottom flavored jets, two neutrinos and two charged leptons as the final state, of which only the jets and charged leptons are visible in the detector. The task of the analysis is to correctly identify these decay products, remove unrelated events (background) as much as possible, and reconstruct the $t\bar{t}$ pair from the decay products. The reconstruction is especially challenging due to the missing information about the neutrinos. Furthermore the analysis has to include a validated description of the remaining background and has to quantify differences between data and background in terms of potential signal. Finally it must assess statistical and systematic uncertainties and derive an upper limit on the coupling strength modifier between the Higgs bosons and the top quark.

Besides the already mentioned improvements coming from the significantly increased amount of data available and the precise calibration, the analysis also profits from a range of new key features, such as the addition of a new search observable (see section 5.10) or the addition of systematic uncertainties for the SM Yukawa coupling. A future publication is planned that combines the results with the ones from the semileptonic decay channel, further improving the sensitivity to the Higgs boson signal.

In the following the two different Higgs bosons are often only denoted by A or H and are collectively referred to as A/H, where for example $m_{\mathrm{A/H}}$ refers to the mass of the pseudoscalar Higgs boson and to the one of the scalar Higgs boson. Additionally a Higgs boson at a specific mass and relative decay width is denoted by its symbol followed by it the mass and decay width values in parentheses.

5.2 Analysis software: The Pepper framework

A major chunk of the computational exercise within an analysis is the process that starts with reading the data provided centrally by the CMS collaboration and stretches until a final set of histograms or a similar functional dependence has been derived. Usually the analysis group itself is tasked with developing a software framework that implements the steps of the analysis. As a consequence there exists a manifold of analysis tools replicating functionality while still being extremely specific to the requirements of one particular analysis.

This analysis however is following a different approach: It is implemented using the Pepper framework [80]. The Pepper framework is a novel analysis framework, whose development began in 2019, initiated as part for this analysis. While the work on this analysis fueled the development, it now includes functionality that can support a wide range of analyses. It was designed with the following goals in mind:

- Easy to step up. No dependencies on heavy libraries such as CMSSW [81] or ROOT [82].
- Implemented in Python and following the design philosophy of Python.
- Adaptable to any analysis that uses NanoAOD-like input data.
- Reproducible results with easy to exchange analysis implementations.
- Configuration driven, including a wide range of configuration parameters.
- Provide a single reusable implementation for common steps that are part of most analyses.

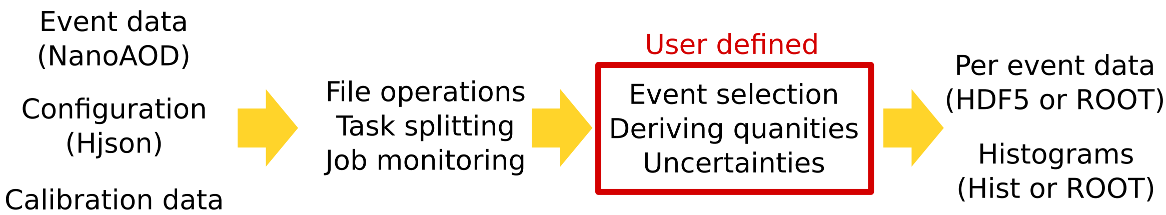


Figure 5.1: The Pepper workflow. Each of the shown stages can be adjusted via the configuration files. One stage, marked in red, is defined by the user, where Pepper provides ready-to-use implementations for common functionalities.

A driving factor that motivated to start the development of Pepper was the recent introduction of the new data format NanoAOD within the collaboration. While this format is independent from implementations within CMSSW (MiniAOD and other formats require CMSSW data structures), it also enables a new paradigm of analysis programming. In the past analyses were programmed on an event-by-event basis. The control flow would fully process the data of a single event before proceeding to start the procedure anew on the next event. Instead Pepper is using a columnar style [83], in which events are processed in chunks of 100, 000 to 1, 000, 000. As a result observables come as vector data types, and vector operations are used throughout the framework. This paradigm was motivated by the software library Coffea [84], whose features are also deployed within Pepper. Not only does the paradigm allow the consistent usage of the Python ecosystem for arrays, including libraries such as Numpy [85] or Matplotlib [86], but it also brings a significant performance boost and easy parallelizability. The latter goes in hand with one of the main tasks of the framework at runtime: Distributing the computational work across a large cluster of independent machines. In Pepper this is realized by the parallel programming library Parsl [87], that automatically submits jobs to a computing cluster and continuously pushes tasks to these jobs. The tasks are generated by Pepper via Coffea functionality and contain a list of instructions to execute (the analysis procedure) as well as the indices of events to process.

The main stages that a workflow goes through when executing Pepper are depicted in figure 5.1. Each stage is configured by the user and based on a configuration file implemented in Hjson [88]. The main inputs are NanoAOD files. However, anything similar in structure is in principle also supported. Pepper aims at separating implementation of computational tasks from particle-physics-related tasks. The latter tasks are implemented by the user in one specific stage. In the simplest case this is just a list of calls to functions already provided by Pepper. Nonetheless, the user is free to implement anything that is not provided by Pepper, for example to define quantities specific to the analysis. The final output is either in the form of histograms or in form of event-by-event data. The file formats used here are either motivated by the Python ecosystem (Hist [89] and HDF5 [90]) or are coming from ROOT. This enables compatibility with modern Python-based tools as well as established tools from particle physics.

In the following results will be presented that were achieved using Pepper.

5.3 Data sets and triggers

As the analysis is focusing on a process with two charged leptons in the final state, the selection of data sets and triggers aims to capture this type of event. The data sets used are labeled (following the naming scheme of the collaboration):

Year	Luminosity (fb^{-1})
2016pre	19.5
2016post	16.8
2017	41.5
2018	59.8
Total	138

Table 5.1: Integrated luminosities for the years used in the analysis.

- **MuonEG** containing events with at least both a muon and an electron.
- **SingleMuon** containing events with at least a single muon. While the inclusion of single lepton events seemingly contradicts the final state of the analysis, with two leptons, one lepton might have escaped the detector's acceptance. Hence an event of interest might only contain one lepton.
- **DoubleMuon** containing events with at least two muons.
- **DoubleEG** and **SingleElectron** containing events with at least two and one electron respectively. For 2018 data, these data sets have been merged to **EGamma** by the collaboration. The argument for usage of single electron events is the same as in the case for single muons.

Noticeably only electrons and muons are accounted for and not tau leptons. These are not further taken care of as they decay already before reaching the detector. The decay products leptonically decaying tau leptons are partially picked by the analysis nonetheless, while the hadronically decaying tau leptons would require a very different analysis approach.

The data used has been recorded in the years 2016, 2017 and 2018. In terms of data labeling, 2016 has been split by the collaboration into two time spans, roughly of equal size. This became necessary after an unforeseen saturation effect in the silicon strip tracker was observed during the former half of 2016 [91]. To account for this effect, two different simulations, one for each part of 2016, have been generated by the collaboration. In this thesis, the former part will be referred to as year "2016pre" and the latter as year "2016post".

The integrated luminosities of each of the years are shown in table 5.1. They are the result of combined measurements from the different luminometers, including BCM1F, [92–94] in conjunction with the data taking sections that has been validated and declared as high quality (internally referred to as "Golden JSON").

Each event is required to pass at least one HLT from a set of selected HLTs (see also section 5.6), ensuring that each event contains a minimum of objects that are relevant to the analysis. The HLT selection is based on another analysis that provides the trigger efficiencies (see section 5.7.6) [95]. Each of the HLTs is included in one of the data sets listed above. The triggers can be categorized into the flavors of the to final state leptons, resulting in the three channels ee, eμ and μμ. In general the HLTs make use of information from the calorimeter and tracker, such as isolation (minimum distance) to other objects or identification requirements. A central feature are p_T thresholds.

- For the ee channel an HLT that features a p_T threshold of 23 GeV (for the leading electron, with higher p_T) and 12 GeV (for the other one) is used. This is extended by a second HLT with a slightly higher threshold (different for each year) but less requirements on the tracks.

- For the eμ channel the analysis employs HLTs with thresholds of 8 GeV for the muon and 23 GeV for the electron, as well as 23 GeV for the muon and 12 GeV for the electron. Additionally in parts of 2016, due to HLT prescales, additional triggers are used (8 GeV and 17 GeV plus 23 GeV and 8 GeV)

- The μμ channel has HLTs with 17 GeV for the leading muon and 8 GeV for the other muon.

- In addition, for the ee channel or the eμ, single electron HLTs are used. Their p_{T} thresholds depend on the year. In 2016 triggers with 25 GeV, 27 GeV and 32 GeV are used. For 2017 the threshold is at 35 GeV, while it is at 32 GeV for 2018.

- Finally, for the eμ and μμ channels also single muon HLTs are used. Similar to the single electron HLT, their p_{T} thresholds vary with year. For 2016 trigger with 27 GeV and 22 GeV are used. For 2017 the threshold is at 27 GeV, while it is at 24 GeV for 2018.

The prescales of the HLTs and their corresponding L1 triggers have been studied. Some of the triggers have been found to be prescaled but the resulting effect on the event selection in this analysis has been determined to be negligibly small.

A full list of all trigger and data set names can be found in reference [96].

5.4 Simulated data

In order to obtain a prediction on what one would expect to see with and without the existence of additional Higgs bosons, simulated data is generated. The approaches used inside the simulation are touched upon in section 2.6 and is implemented using the Monte Carlo (MC) method. When using MC generators, random events are produced, whose quantities follow the probability distributions resulting from the physics theory (see equation 2.11). The software used to generate the MC data in this analysis are POWHEG [97], MadGraph5_aMC@NLO and MadSpin [98] and PYTHIA 8 [99]. POWHEG is providing single top and $\mathrm{t\bar{t}}$ simulation at NLO precision for this analysis. MadGraph5_aMC@NLO was used to simulate the other processes except for the two-boson ones. MadSpin is used in conjunction with MadGraph5_aMC@NLO to simulate the decays of the particles. Finally the parton shower in case of all data sets in addition to the full generation of the two-boson data sets is taken care of by PYTHIA 8. The simulation is augmented by a detector simulation using Geant4 [100]. After this simulation, one obtains event data that is ideally follows the same distributions as real event data. One piece of information that is always available in the MC data and not in real data however is the so called MC truth. The MC truth includes for example the information about which hypothesis was used (for example SM or BSM) or which particles were present before the decay.

Table 5.2 summarizes the SM processes whose simulations have been used in the analysis. The simulation is done centrally by the collaboration. These processes are usually referred to as background for the purpose of the analysis. Besides the $\mathrm{q\bar{q}} \to \gamma/\mathrm{Z} \to \mathrm{l\bar{l}}$ ($m < 50\,\mathrm{GeV}$) and the two-boson processes, which are simulated at LO, all processes are simulated at NLO. The processes have been selected on the basis of which contributions they have to the final event count after the event selection has been applied (described in section 5.6). Even processes with less than 0.1 % contribution (for example two-boson ones) to the overall event count are accounted for (see section 5.8). The number of events that are expected for each process and year can be computed with $N = L_{\mathrm{int}} \cdot \sigma$, where L_{int}

Process	Cross section (pb)	Generator
$\mathrm{q\bar{q}} \to \gamma/\mathrm{Z} \to \mathrm{l\bar{l}}$ ($m < 50\,\mathrm{GeV}$)	18610 (NLO [101])	MG (LO), PY
$\mathrm{q\bar{q}} \to \gamma/\mathrm{Z} \to \mathrm{l\bar{l}}$ ($m \geq 50\,\mathrm{GeV}$)	6077.22 (NNLO [102])	MG (NLO), PY
Single top quark (t-channel)	136.02 (NLO [103])	PH, PY
Single top antiquark (t-channel)	80.95 (NLO [103])	PH, PY
Single top quark or antiquark to leptons (s-channel)	3.74 (NLO [104])	MG (NLO), PY
Single top quark + W	35.85 (≈NNLO [101])	PH, PY
Single top antiquark + W	35.85 (≈NNLO [101])	PH, PY
$\mathrm{t\bar{t}}$ (semileptonic decay)	365.3452 (NNLO [105])	PH, PY
$\mathrm{t\bar{t}}$ (dileptonic decay)	88.2877 (NNLO [105])	PH, PY
$\mathrm{t\bar{t}} + (\mathrm{W} \to \mathrm{q\bar{q}'})$	0.4316 (NLO [104])	MG (NLO), PY
$\mathrm{t\bar{t}} + (\mathrm{Z} \to \mathrm{q\bar{q}})$	0.2432 (NLO [104])	MG (NLO), PY
$\mathrm{t\bar{t}l\nu}$	0.2149 (NLO [104])	MG (NLO), PY
$\mathrm{t\bar{t}ll'\nu\nu'}$	0.5104 (NLO [104])	MG (NLO), PY
WW	118.7 (NNLO [102])	PY
WZ	47.13 (NLO [101])	PY
ZZ	16.523 (NLO [101])	PY

Table 5.2: Background processes simulated using MC generators and the cross sections they are weighted according to. The last column denotes the simulation software used to generate the events. MG stands for MadGraph5_aMC@NLO (either LO or NLO), PH for POWHEG and PY for PYTHIA 8.

can be taken from table 5.1 and σ is the cross section as found in table 5.2. The number of events that are generated does not agree with this expected number. In fact it is purposefully kept higher. Because the final product that created from the events is always some form of histogram, the MC events are assigned a weight. This weight is used in the creation of the histograms and is computed from the ratio N/N_{MC}, where N_{MC} is the number of generated events. Note that the cross sections used here are for some processes from computations at orders higher than the actual simulation. This means that the analysis profits from an overall event count that is at even higher precision than the simulation. The simulation was done using the TuneCP5 [106] PYTHIA 8 setting and the PDF set NNPDF3.1 [6] at NNLO and $\alpha_{\mathrm{S}} = 0.118$.

In table 5.2 one can see the $\mathrm{q\bar{q}} \to \gamma/\mathrm{Z} \to \mathrm{l\bar{l}}$ process appearing twice, once with an invariant mass below 50 GeV and once above. The process is referred to as Drell-Yan process (DY) and has a comparatively high cross section, while its contribution to this analysis is minor, yet visible. This poses the challenge of a high statistical uncertainty if there are not enough generated events left after the event selection is applied. For this reason one needs a large quantity of generated events for the particular process. The collaboration provides a simulation that is a mixture of events with 0, 1 and 2 additional jets at ME level ("inclusive"). Additionally it provides simulations with only one specific number of additional jets ("binned"). This binned simulation sums up to roughly the same amount of generated events as in the inclusive one. For this reason the inclusive one is weighted so that it provides half of the expected event count for DY in the analysis, while the other half comes from the binned simulation. The binned ones are weighted with respect to each other so that their contribution matches the distribution of the number of jets in the inclusive one. This jet distribution is shown in figure 5.2. It can be seen that the inclusion of NLO effects causes the simulation to create events with three jets at ME level. These events sum up to a negative contribution due merging scheme used to

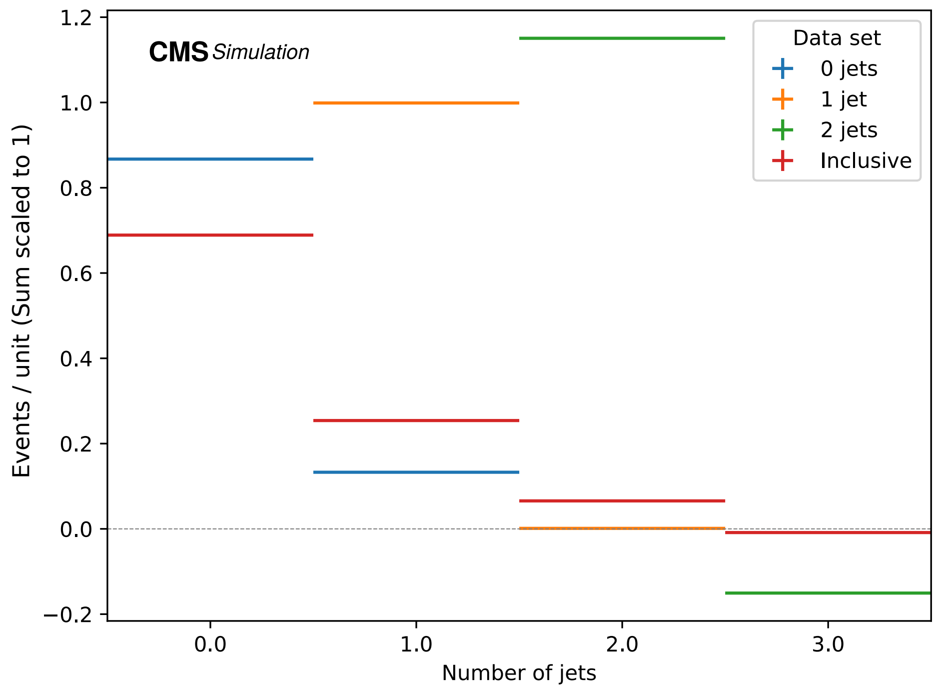


Figure 5.2: Distribution of the number of jets at ME level of the individual DY simulations before scaling them appropriately. Due to the fact the ME calculation is done at NLO, each of the simulations also contain a fraction of events with an additional jet.

combine ME simulation and parton shower. In other words, the parton shower will also add jets to the events and, to not double count parts of the three jet region, events with many jets at ME level are often negatively weighted.

The signal from the heavy Higgs bosons A and H is simulated in using the same set of software as the background. This is done at LO precision. In section 2.8 it was explained how the different parts of the signal (resonance and interference) scale differently with the coupling strength modifier $g_{\mathrm{A/H}}$. In order to be able to probe different values of $g_{\mathrm{A/H}}$ without having to simulate events for every value to probe, the two parts are simulated independently and then weighted according to the assumed value of $g_{\mathrm{A/H}}$. Furthermore, this analysis of the dileptonic decay also has to take the semileptonic decay into consideration as it amounts so roughly 2 % of the signal events after selection. Simulations are also done independently for $\mathrm{A} \to \mathrm{t\bar{t}}$ and $\mathrm{H} \to \mathrm{t\bar{t}}$ and for the two decay channels (semileptonic and dileptonic), resulting in eight independent simulations. This number is further multiplied by the number of chosen masses and decay widths. Masses and widths have been taken from a grid of

$$\{365\,\mathrm{GeV}, 400\,\mathrm{GeV}, 500\,\mathrm{GeV}, 600\,\mathrm{GeV}, 800\,\mathrm{GeV}, 1000\,\mathrm{GeV}\} \times \{2.5\,\%, 10\,\%, 25\,\%\},$$

where the second set denotes the decay widths relative to the assumed mass. The masses and widths have been chosen so that a broad spectrum of different hypotheses is covered. Additionally simulations have been done for 400 GeV and 5 % width, because here the analysis in reference [79] saw a disagreement. Finally simulations have been done for 9 % decay width and masses of 450 GeV, 550 GeV, 700 GeV and 900 GeV because of technical reasons in the signal ME reweighting, that is explained in section 5.7.5. The simulation has been performed centrally within the collaboration. The resonance cross sections used to weight the signal are taken from an NNLO computation resulting from SusHi [107]. The cross section of the interference is approximated by the geometric

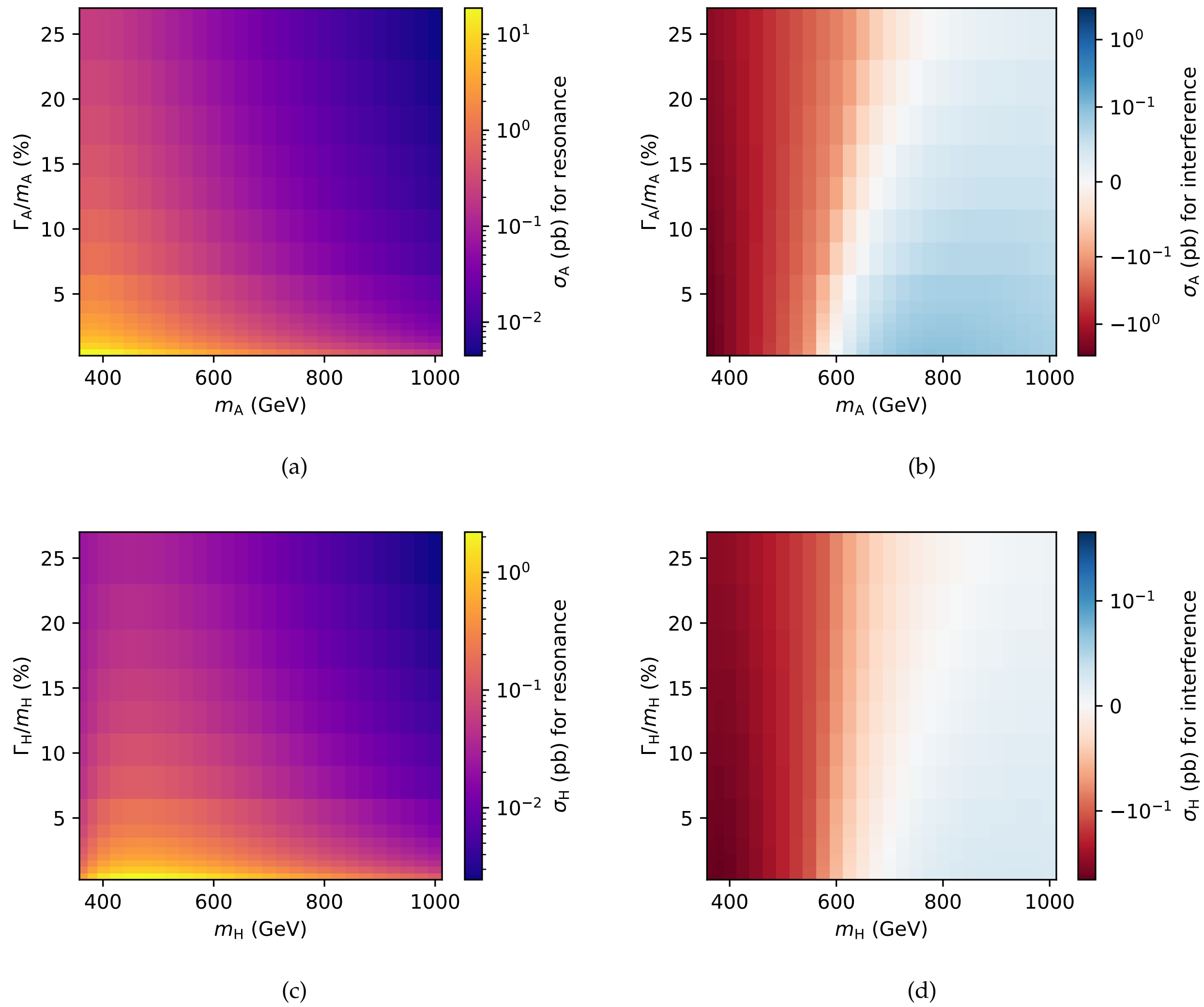


Figure 5.3: Cross sections used for the signal weighting. The upper row shows the ones for A, while the lower one shows the ones for H, with resonance on the left and interference on the right. The cross sections are taken from an NNLO computation.

mean of ratios for both, SM and resonance cross sections:

$$\sigma_{\text{NNLO,int}} = \sigma_{\text{LO,int}} \sqrt{\frac{\sigma_{\text{NNLO,res}}}{\sigma_{\text{LO,res}}} \frac{\sigma_{\text{NNLO+NNLL,SM}}}{\sigma_{\text{LO,SM}}}}, \tag{5.1}$$

where $\sigma_{\text{NNLO+NNLL,SM}}$ was taken from a computation resulting from Top++ [108]. The resulting cross sections can be seen in figure 5.3.

5.5 Physics objects

Due to the final state that is the focus of the analysis, the main interest lies on the following objects: electrons, muons, bottom flavored jets and MET. As described in section 3.3 these objects are already reconstructed centrally from the CMS particle flow algorithm. For the means of this analysis the quality criteria for the objects are tightened. Every object failing the criteria will be disregarded. The main idea behind the requirement is to ensure the objects are measured with a small uncertainty and proper calibrations are available. These criteria also contain the requirement of a certain identification algorithm (ID) to be passed. These IDs are essentially a set of further criteria that ensure the type of the object is correctly assigned. This is realized either as a simple requirement on a set of observables ("cut based"), as a multivariate analysis (MVA) including machine learning techniques such as boosted decision trees or even as a deep neural network. An important quantity for the ID is the efficiency, which is the fraction of objects that correctly pass the ID. In this context common terms are loose, medium or tight working points, which refer to different setups of the same ID, with higher to lower efficiency. Often connected to the ID are isolation criteria, which ensure there are no other objects in the vicinity, as otherwise there is a chance the objects could be mismeasured or faked due to the other object. The IDs used in the analysis are provided centrally by the collaboration.

Additionally corrections are applied to the objects' four-momenta. This can affect real data, if there are effects that are well understood and can be accounted for this way. However, at the stage of the analysis the corrections are mainly done on objects in MC only, as the simulation is not perfect and not all detector effects have already been accounted for. The procedure includes a correctional factor and a "smearing". Latter adjusts the resolution so that uncertainties in real data and MC match. The corrections ensure that the distributions seen in MC more closely depict what is observed from real data, albeit the calibrations are expected to be a very minor overall, with effects below the 1 % level on the overall event count.

For electrons, the following criteria are applied

- MVA identification with PF isolation at a 90 % efficiency [109]
- $p_{\text{T}} > 20\,\text{GeV}$
- $|\eta| < 2.4$
- Outside of the region between barrel and endcap ($1.444 < |\eta| < 1.566$)

The electron energy is corrected as provided already in NanoAOD, which is an effect in real data as well as simulation. [110]

For muons, the follow criteria are applied

- Cut based identification at a medium working point [111]

- PF isolation at a loose working point [111]
- $p_{\mathrm{T}} > 20\,\mathrm{GeV}$
- $|\eta| < 2.4$

Similar to the electrons, the muon energy is also corrected, affecting real data and simulation. This is done using the so called Rochester corrections [112].

The jets have to fulfill the following criteria

- Identification at a tight working point [113]
- $p_{\mathrm{T}} > 20\,\mathrm{GeV}$
- $|\eta| < 2.4$
- $\Delta R \geq 0.4$ to any muon or electron

Again, the jet energy is corrected using jet energy correction (JEC) and jet energy resolution (JER) [114]. Whether a jet is bottom flavored or not is determined using DeepJet at a medium working point.

Finally the MET is corrected by propagating the changes that arise from applying the corrections to the energy of the jets. This includes JEC as well as JER.

5.6 Event selection

The events that are used for the statistical evaluation in the analysis are selected following criteria that are meant to reduce the background events as much as possible and to exclude regions that are possibly not very well modeled by the simulation. Discarding events this way comes with the tradeoff of also removing events that come from the signal and which are meant to be kept. Thus one has to carefully define the requirements (cuts) on the observables, which are used for the selection.

Because the top quark reconstruction is done only at the very end of the event selection, observables based on the top quarks are not available to the selection. Instead an easily available observable that gives insight on the crucial invariant top quark mass $m_{\mathrm{t\bar{t}}}$ is the invariant mass of the two charged leptons fro the decay m_{ll}. m_{ll} is available already once two leptons have been identified.

The cuts used in this analysis have been optimized and appear in similar fashion already in several older analyses for the dileptonic decay of the $\mathrm{t\bar{t}}$ system [115–117]. The selection is done as follows in this order.

1. Events are discarded that are not part of the validated and certified event listing by CMS ("Golden JSON"). These events do not fulfill the strict quality criteria by the collaboration because they were for example taken during a time the CMS detector was operating in an abnormal state. This only applies to events in real data. MC events are accounted for by setting the integrated luminosity accordingly, as described in section 5.3.
2. The events are required to pass at least one of the HLTs named in section 5.3.
3. The events' MET must fulfill specific quality criteria defined by the collaboration [118]. This includes requirements on the primary vertex, the BIB, HCAL and ECAL noise and calibration and the quality of the electrons and muons.

4. Events are discarded that have more than two leptons, as defined in the previous section (electrons or muons of positive or negative charge). The presence of more leptons is a good indication that the event is not from the targeted dileptonic top decay. Signal events with more leptons could come from additional radiation.

5. Events are discarded that have less than two leptons, as defined in the previous section (electrons or muons of positive or negative charge). This brings all events that are left to exactly two leptons. The techniques applied in the following sections explicitly need two leptons and less than two leptons are not expected from the dileptonic decay.

6. If the two leptons do not have opposite charge, the event is discarded. As the analysis is looking for a top quark and antiquark pair, the charge of the leptons must be opposite and charge can be measured with high certainty by the detector.

7. All events are required to pass a trigger that belongs to the corresponding analysis channel. This means that an event from, for example, ee has to pass a two or single electron HLT.

8. One of the leptons must have a p_T of at least 25 GeV. This further removes low energy events, which are unlikely to come from top quark events.

9. The invariant mass of the lepton system m_ll must be above 20 GeV. This suppresses events from unrelated resonances of particles with lower masses with decays involving leptons and QCD events with multi jet production, namely the DY contribution.

10. In the same flavor lepton (ee and $\mu\mu$) channels, events with invariant lepton masses m_ll close to the mass of the Z boson are rejected. This especially diminishes the DY events, which may involve the decay of a Z boson into two leptons. In particular, the window in which events are discarded is $76\,\mathrm{GeV} \leq m_\mathrm{ll} \leq 106\,\mathrm{GeV}$ ($\pm 15\,\mathrm{GeV}$ around the Z boson mass).

11. The events are required to have at least two jets. Similar to the two lepton cut, the later computation requires the presence of two jets.

12. Two of the jets must have a p_T above 30 GeV, which further removes events of low energy. At this stage, the rejected events are mainly from the DY background.

13. At least one of the jets must be a bottom flavored one, as defined in the previous section. Again, this removes a big contribution of the DY background. While the target top decay includes two b jets, the probability to have both jets correctly detected as bottom flavored is rather small (less than 50 %), hence only one such jet is required.

14. For the same flavor channels, events are rejected that have a MET with p_T less than 40 GeV, which is the last step to remove a significant portion of the DY background. Additionally in this region of low MET p_T uncertainties are comparatively large due to limited statistical precision as well as an increased difficultly to model it. Hence, these events are not used.

15. The events are required to pass the top reconstruction algorithm as explained in section 5.9. While this is primarily a technical shortcoming of the algorithm, to a small degree this removes a higher background than signal fraction.

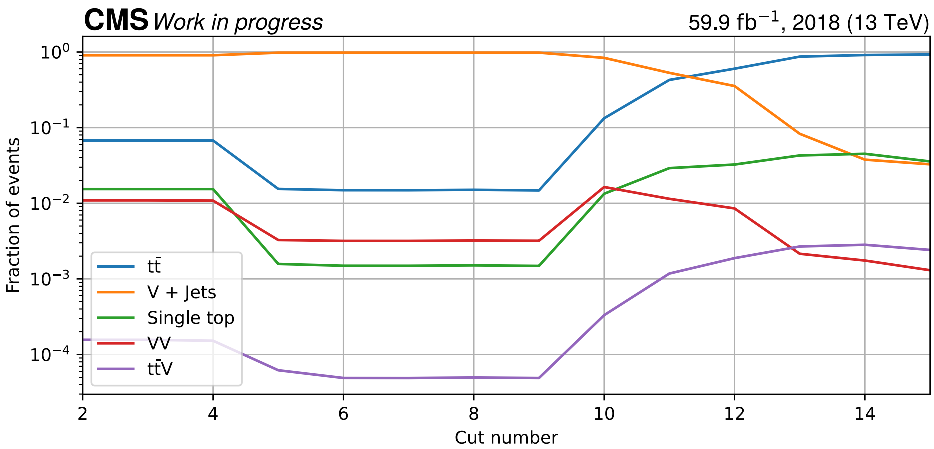


Figure 5.4: Fraction of events per background process group in simulated MC events for 2018. The other years show a very similar behavior. Note that the fourth cut removes large chunk of unrelated QCD events. These kind of events have a negligibly small contribution at the end of the cut chain and are thus not simulated.

Figure 5.4 shows the relative fractions of signal and backgrounds as a function of the applied cuts. Some of the cuts, including cuts 3, 7 and 8 induce very little change. For example only very few events are identified to have low quality MET, thus most of them pass cut 3. At the end of the selection chain, the $t\bar{t}$ background becomes by far the most significant contribution. This closely correlates with the signal contribution.

5.7 Reweighting techniques

5.7.1 The approach

In section 5.4 it was described how the MC events are weighted according to their process' cross section and the luminosity. By this the weights of the events from different processes become different, while within one process all events are assigned the same weight. The procedure is now used at a more exhausting level, where each MC event is assigned an individualized weight. The weight becomes a product of many subweights.

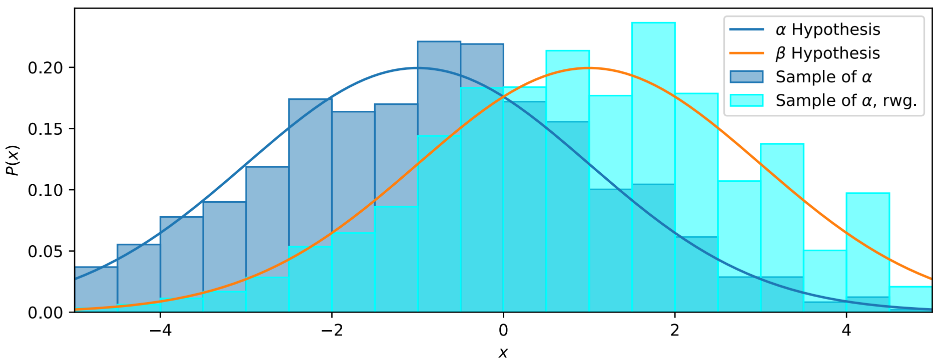


Figure 5.5: Reweighting a sample to another distribution. The blue and orange curves show the distribution of the same variable according to two different hypotheses. The bars show a single sample of the blue hypothesis. For the cyan bars the same sample is used, albeit reweighted according to the orange hypothesis.

This so called reweighting approach can be used to change a sample's distribution into another without the need to regenerate the sample. The approach provides an easy way to apply corrections to already generated events but also to change the physics assumptions that went into the generation. However, in the worst case, this is at the cost of higher uncertainties, as sparsely populated regions might get scaled up and their uncertainties likewise. As a simple example, the weights are derived by

$$w_i = \frac{P_{\text{new},i}}{P_{\text{orig},i}}, \tag{5.2}$$

where w_i is the weight an event at bin position i in the histogram will get, $P_{\text{new},i}$ is the distribution of the new hypothesis at position of bin i and $P_{\text{orig},i}$ is the distribution of which events were generated according to. This is demonstrated in figure 5.5. A proper example using such a ratio is the pileup correction, which is explained in the next section.

5.7.2 Pileup

The pileup distribution that served as an input during the MC generation does not necessarily match what is observed in the experiment. In fact the simulation is often prepared already at the start of data taking. Furthermore, the correct pileup distribution due to changing beam configurations strongly depends on the time the data was measured. Therefore the pileup distribution in the MC is corrected by reweighting.

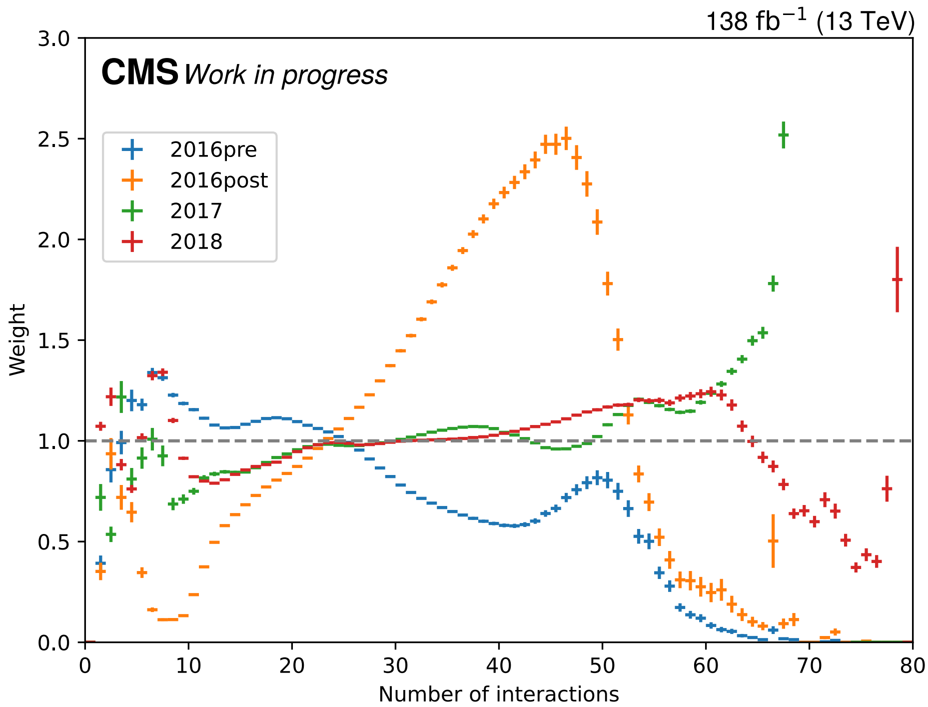


Figure 5.6: Weights used for each year to reweight the events to the pileup distribution observed in the experiment. The vertical bars indicate the statistical uncertainty. Each event is assigned the weight according to its pileup (number of interactions).

In order to derive the weights, the real pileup distribution is approximated from the luminosity, using the relationship given in section 4.7.3. As per recommendation of the collaboration, a total inelastic cross section of 69.2 µb is assumed [119]. The obtained distribution, a histogram of the number of events for each pileup value, is normalized to 1 (scaled so that the sum of all bin heights is

1). The same distribution is obtained and also normalized from the MC events using the true number of pileup that was simulated for each event. The weights are then obtained from the ratio of both, as indicated in the formula above. The results can be seen in figure 5.6. Most of the events are located at around 30, which is also where the weights are close to 1. This means that for a large fraction of events, not much changes and the difference imposed by the reweighting is on a fine level.

5.7.3 Drell-Yan reweighting

As can be seen in section 5.6, the DY background is strongly reduced by the event selection. However, it still persists as the third largest background. The DY simulation is assumed not to be fully accurate for the region selected in the analysis, especially when considering the distribution of the number of b jets, and thus a reweighting is applied to correct it. The weights used only depend on the analysis channel and year. While a more sophisticated dependence is possible, for example on the MET p_T, it was observed that this does not result is a significant change.

The weights are computed as ratios of the number of events in measured data and in simulation after the bottom flavored jet cut has been applied. However, simply reweighting the simulation to what is observed in measured data would bias the measurement. For this reason, the counts are computed from inside the window around the Z boson mass. As described in section 5.6 the events with an invariant lepton mass close to the Z boson mass are not used in the final evaluation due to the vast dominance of DY events in the region. Because the procedure uses counts from inside and outside the mass window it is often referred to as $R_{\mathrm{in/out}}$ method. It has been used already in many previous analyses [79, 120].

	2016pre	2016post	2017	2018
ee	1.16 ± 0.016	1.18 ± 0.017	1.11 ± 0.014	1.11 ± 0.009
eμ	1.21 ± 0.011	1.19 ± 0.011	1.13 ± 0.007	1.11 ± 0.006
μμ	1.23 ± 0.014	1.2 ± 0.014	1.15 ± 0.009	1.12 ± 0.008

Table 5.3: Weights from the Drell-Yan correction in all years with statistical uncertainties.

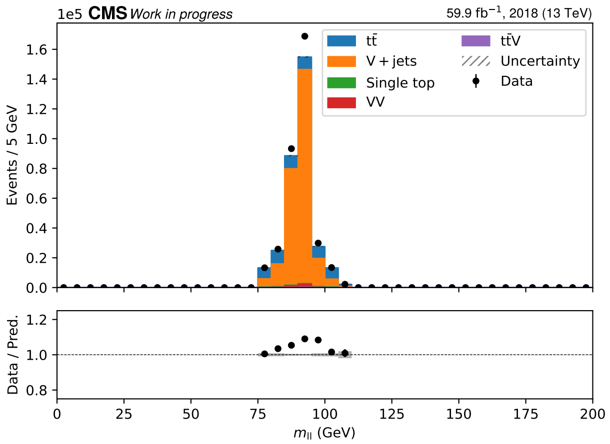
The resulting weights are found in table 5.3 for all years and channels. It can be seen that overall the event counts for the DY are corrected with an additional 10 % to 20 %. However, because the final contribution is only around 4 %, the effect is fairly small, as it is expected to be. As can be seen in figure 5.7, the agreement between data and simulation improves visibly inside the Z boson mass window. In the figure only statistical uncertainties are indicated, which is the reason why the data points still are visibly outside the uncertainties ranges even after the reweighting.

5.7.4 Electroweak and QCD correction

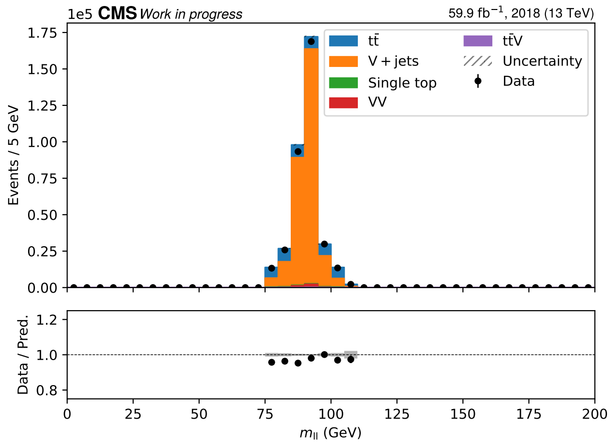
The $\mathrm{t\bar{t}}$ background is simulated at NLO accuracy in QCD. Using reweighting, the precision of the distributions obtained from the $\mathrm{t\bar{t}}$ simulation is improved to include also NNLO QCD as well as NLO EW effects. The required weights are the result of computations using MATRIX [121] (in the case of QCD) and HATHOR [122] (in the case of EW). The procedure for the correction of the EW effect is similar to what is used in reference [5].

For both types of corrections, the differential cross section

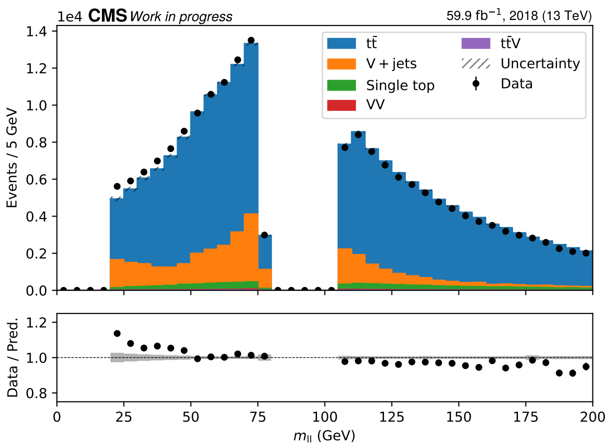
$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}m_{\mathrm{t\bar{t}}}\,\mathrm{d}\cos\theta^*} \tag{5.3}$$



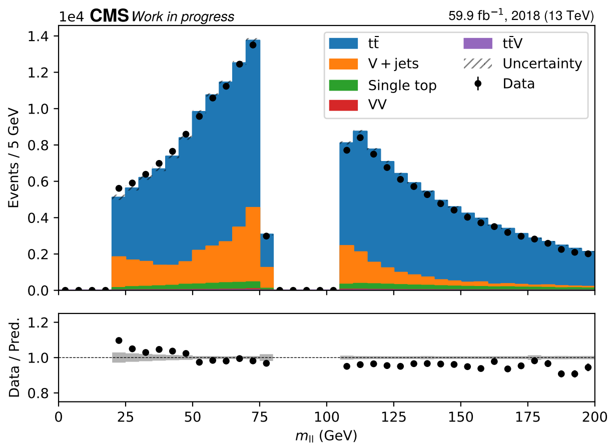
(a)



(b)



(c)



(d)

Figure 5.7: Distributions of the invariant mass of the lepton system before and after applying the DY reweighting for the μμ channel. (a) shows the distribution inside the Z boson mass window before applying, (b) shows it after applying, (c) and (d) show the distribution before and after applying but outside the window. The uncertainty shown is statistical only.

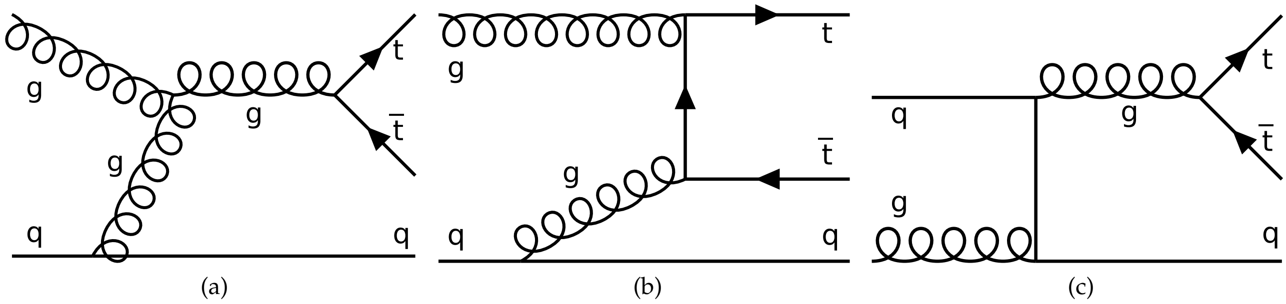


Figure 5.8: Feynman diagrams for $t\bar{t}$ production with an gq initial state. Here, q can be a quark as well as an antiquark. In (a) and (b) the initial quark radiates a gluon, creating either the s or t-channel known from the LO $t\bar{t}$ production via gg. In (c) the final state q comes from the gluon and the $t\bar{t}$ production resembles the LO one via $q\bar{q}$.

is calculated. Here, $\cos\theta^*$ is the cosine of the angle between the 3-momenta of the top quark in the $t\bar{t}$ rest frame and the $t\bar{t}$ system in the laboratory frame. $m_{t\bar{t}}$ is, as usual, the invariant mass of the $t\bar{t}$ system. This choice is made based on the search observables, which are introduced in section 5.10 and are used for the final evaluation. For the QCD correction, this cross section contains the NNLO effects, while the one for the EW correction contains some of the NLO EW effects and is otherwise assumed to be LO for QCD. The weight is then obtained by dividing it by the distribution that is obtained from the MC events in form of a histogram. The information is taken from the generator level (before the detector simulation). In the case of the QCD correction, the same MC NLO events are used as for the $t\bar{t}$ background. However, for the EW correction, in order to get an estimate on the EW effect only, LO MC events are used, that were generated using MadGraph5_aMC@NLO.

The procedure gives weights that depend on $m_{t\bar{t}}$ and $\cos\theta^*$ at ME level. It was observed that the weights for the EW correction show a significant dependence on the initial state (either gg, $u\bar{u}$ or $d\bar{d}$). For this reason the weights are obtained and applied also in dependence of the initial state. This poses one challenge though: Because the $t\bar{t}$ background simulation is done at NLO, the initial state can also be a gluon plus a quark or antiquark. As seen in figure 5.8, under this initial state, the quark can radiate a gluon, that together with the other gluon, creates the $t\bar{t}$ similar to the gg initial state. Another possibility is that the gluon decays into a quark antiquark pair, which enables the $t\bar{t}$ production similar to the $q\bar{q}$ initial state. Therefore the former is assigned the weights also used for gg, while the latter is assigned the ones of $u\bar{u}$ or $d\bar{d}$, depending on the flavor of the initial state. To make this distinction, the sign of p_z (the z component of the momentum) of the final state q at the ME level is used. Because the initial q and g travel in opposite directions on the z axis, if the sign agrees with the one of the p_z of the initial g, it is a strong hint that the g radiated the final q. Thus in this case the $q\bar{q}$ weights are used. In the other case, the final q most likely is the same as the initial q and thus gg weights are used.

The introduction of the EW correction opens up another very important aspect of the analysis. Because HATHOR allows setting the value for the top Yukawa coupling (to the SM Higgs boson), the weights can be produced for values different from the SM value. Therefore, the background can be reweighted to different coupling values, which is used in section 5.12 to predict the uncertainty in the analysis coming from the uncertainty on the Yukawa coupling.

The collaboration provides a similar form of reweighting, while targeting the top quark and antiquark p_T only. This top p_T reweigthing is motivated by the fact that CMS and ATLAS analyses observed a disagreement between measured data and simulation in form of a slope in the ratio [123].

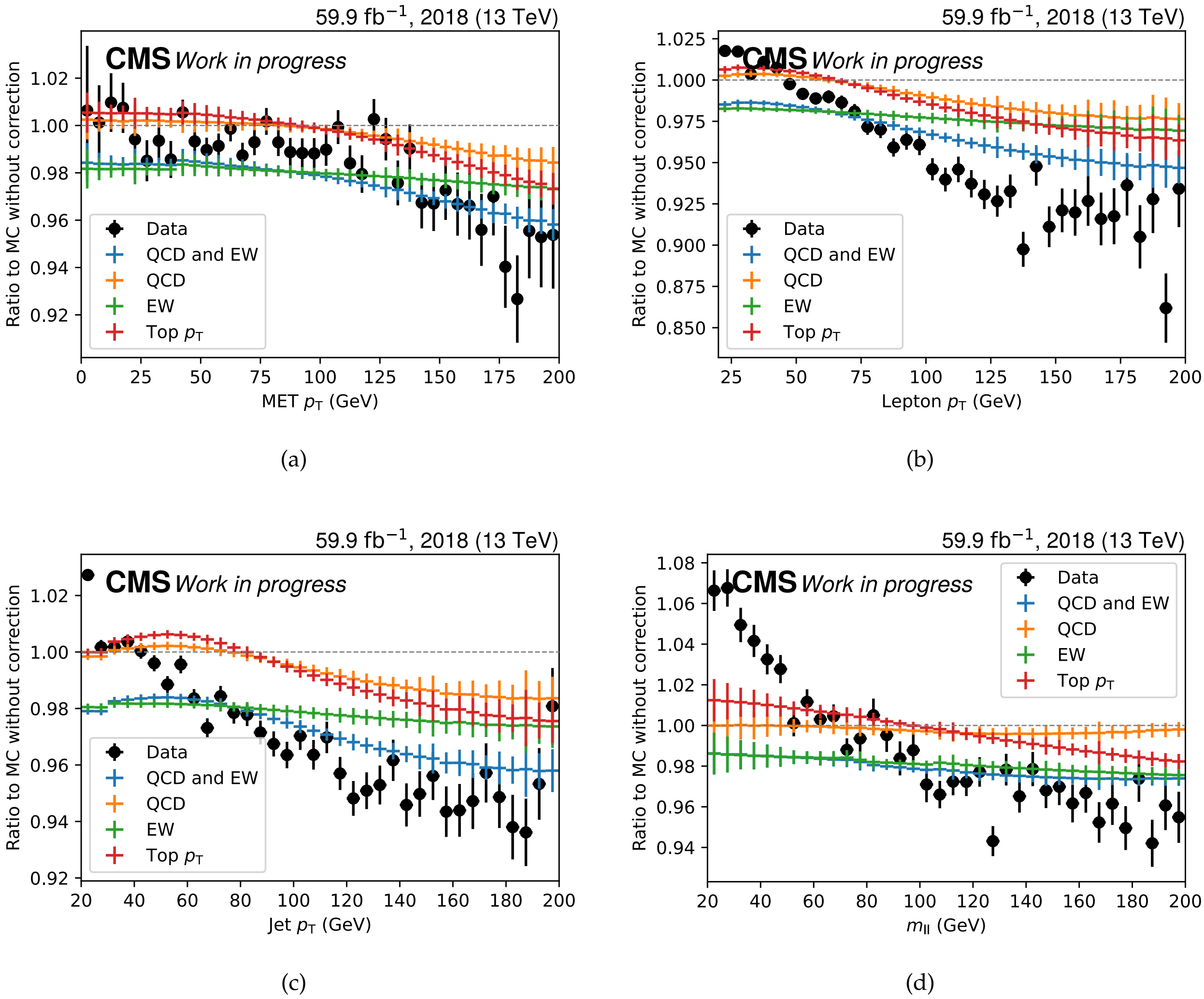


Figure 5.9: Comparison of the different corrections for the $t\bar{t}$ background. The QCD and EW ones are as used in the analysis and the top p_T one is provided by the collaboration. The used distributions are made after the MET p_T cut and include all background simulations (not just $t\bar{t}$ for which the correction applies). Vertical bars indicate the statistical uncertainty. Lepton and jet p_T distributions make use of all leptons/jets present in the event.

Figure 5.9 compares the effect of the different corrections to real data. Here, the top p_T reweighting is computed from

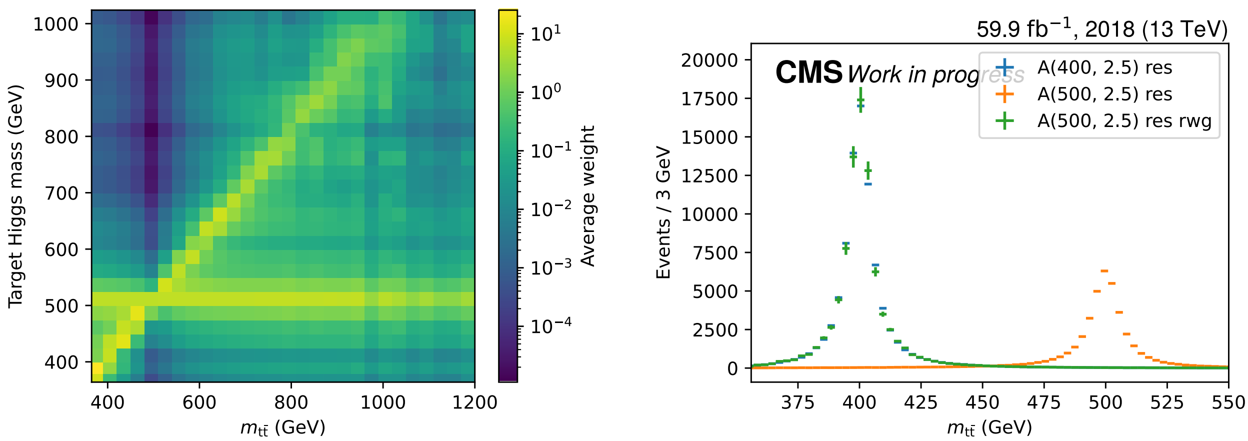
$$s(p_\mathrm{T}) = 0.103 \exp(-0.0118 p_\mathrm{T}) - 0.000134 p_\mathrm{T} + 0.973 \tag{5.4}$$

$$w = \sqrt{s(p_{\mathrm{T,t}}) s(p_{\mathrm{T,\bar{t}}})} \tag{5.5}$$

with the p_T of the top quark or antiquark and w being the weight. The MC simulation can be seen to have a deviation to the real data on the order below 10 %. The deviation still is within the uncertainties after adding systematic (see section 5.8), which have been omitted here. However, the corrections help lifting this deviation, as the follow uncertainties the exact same trend. It can also be seen that the corrections used in this analysis go into a very similar direction as the top p_T reweighting derived from data.

5.7.5 Matrix element reweighting

In section 5.4 it was specified for which choices of Higgs boson masses and decay widths simulations were performed. In total 23 different mass-width values have been picked. The selection was chosen in a way that the parameter space that is available to the analysis is covered consistently. The analysis in reference [79] used a morphing technique to generate additional histograms from the ones that are adjacent in the mass/width space. Therefore a tight grid of mass-width space was needed, which was also the initial motivation for the selection of this analysis. However, the analysis now replaces the morphing technique with a much more rigorous approach, which is an ME based reweighting.



(a) (b)

Figure 5.10: ME reweighting for the A resonance. (a) shows the average weights in dependence of the ME level $m_{t\bar{t}}$ and the target mass of A (resonance, 2.5 % decay width) for the A(500 GeV, 2.5 %) simulation. (b) shows the ME level $m_{t\bar{t}}$ distributions from simulations of the A resonance for 400 GeV and 500 GeV (blue and orange) and the 500 GeV one reweighted to 400 GeV (green). Only a small fraction of the generated events used in the analysis for the A signal is shown here. Vertical lines indicate the statistical uncertainty.

The ME reweighting enables a precise simulation of points of masses and widths without having to run the full simulation chain. Merely the ME needs to be recomputed. The weight for an event j is obtained from the ratio

$$w_j = \frac{\left|M_{\mathrm{target},j}\right|^2}{\left|M_{\mathrm{source},j}\right|^2}, \tag{5.6}$$

where $M_{\mathrm{target},j}$ is the ME for the event using the target mass-width point and $M_{\mathrm{source},j}$ is the ME the event was originally simulated for. The general behavior of the obtained weights can be seen in figure 5.10a. If the target mass and decay width of the reweighting is close to the mass inside the original simulation, the weights are generally constant (the horizontal line at the target mass of 500 GeV). It can also be seen how the reweighting creates the resonance peak at the mass of the Higgs boson in the $m_{\mathrm{t\bar{t}}}$ spectrum, which is indicated by the diagonal line in the figure. Finally if the target Higgs mass is far away from the original Higgs mass, the events in the resonance peak are assigned a very small weight (the vertical line in the figure at $m_{\mathrm{t\bar{t}}} = 500\,\mathrm{GeV}$).

The requirement for the approach to perform well is that in the phase space of the ME particles has to be covered well enough, meaning the source ME does not get too small compared to the target one. Otherwise the assigned weights will become overly large and the statistical uncertainty will rise. This effect can be seen in figure 5.10b, where just the simulation of the A 500 GeV, 2.5 % resonance has been reweighted to the one of 400 GeV. The resonance is very well localized around the Higgs boson mass. The single simulation used in the figure only equals to a small portion of the entire simulation used in the analysis and thus the statistical uncertainty visibly increases after reweighting. However, the well defined location of the resonance is well reproduced in the left side plot of the figure, where the weights clearly follow the resonance peak. The analysis reweights the simulated signal events to obtain predictions for new mass/width points, as well as to obtain more events for already simulated points. The new mass/width points then are as close as 25 GeV in $m_{\mathrm{A/H}}$ to each other.

5.7.6 Other reweightings used

A number of other reweightings are used to correct the MC events, which are explained in this section. Most of them are derived using the results from MC together with the measured data from the CMS detector. An exception to this is the generator weight.

- The **generator weight** is produced by the MC simulation and includes for example effects from the PDF or are of technical nature due to how the NLO simulation computation is implemented. Every MC event is assigned a weight by the simulation.

- **Trigger efficiencies** are corrected using weights provided by the collaboration [95]. These weights are given as functions of the p_{T} of the two leptons.

- The **electron and muon efficiency, identification and isolation** are corrected. The weights are provided by the collaboration separately for electrons and muons and depend on the electron and muon p_{T} and η [124, 125].

- Similar to the electron and muon reweighting, the **bottom flavored jets** are also corrected for. The approach that is followed is standardized within the collaboration and uses a ratio of the probability to observe an event with the number of (bottom flavored) jets in real data and MC [126]. The probability is calculated as a product of efficiency and, in the case of MC, scale factors. The scale factors are provided by the collaboration and depend on jet's flavor, p_{T}, η and the output value of the algorithm for bottom flavored jet discrimination. The efficiency is measured as part of the analysis as the ratio between the number of jets after and before the bottom flavored jet cut using histograms with jet p_{T}, η and flavor axes.

- The **ECAL "Prefiring" issue**, which was present in 2016 and 2017, is handled by reweighting. The issue was caused by a timing shift in the ECAL, which was not accounted for in the L1 triggers and therefore energy in the ECAL was incorrectly treated as being part of the previous bunch crossing [127]. The weights are derived by the collaboration and are supplied within NanoAOD as part of every event.

5.8 Data to simulation agreement

An important task during the analysis is the validation of the simulation and gain an understanding of the SM background and the measured data. This is done by comparing the different distributions of the background simulations to the real data. However, because the analysis is performed as a blind experiment, the accessibility to such comparisons is fairly limited. The approach of a blind experiment means the results shall not be viewed by the experimentalists until the analysis has been fully developed and is not changed anymore. Hence the experimentalists are "blind" to the results. This procedure is done to prevent a human to bias the results by implementing changes, even if doing so subconsciously. A commonly used approach to blind an analysis involves the usage of a control region, which is defined by an event selection that results in events very close to the ones of interest, but still completely depleted from signal events. The approach is not an option for this analysis as there exists no simple observable that separates the signal from the background.

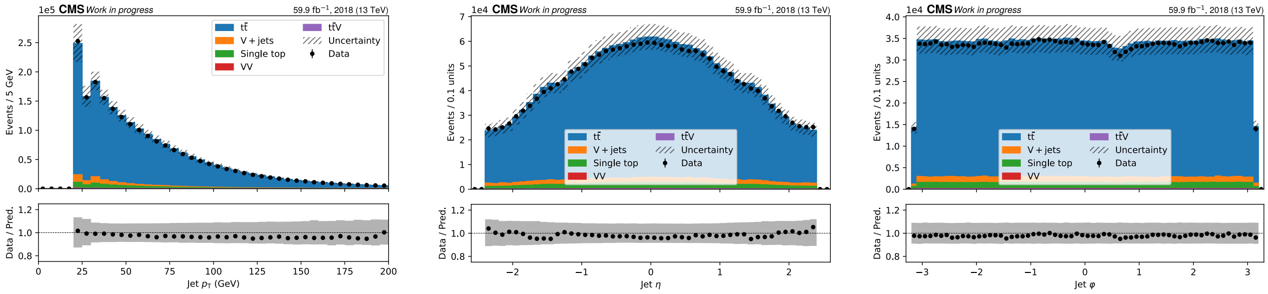
In the case of this analysis, the blindness is realized by not performing the reconstruction of the top quarks (see section 5.9) on real data. This also means that no top related observable can be used for the comparisons and it also means that the last cut cannot be applied. Luckily, for validating the performance of the simulation, it is sufficient to observe the wide range of available distributions.

The most important observables to verify are the momenta of the objects of interest (leptons, jets and MET) in terms of p_{T}, η and φ. These are shown in figures 5.11, 5.12 and 5.13. The figures show the statistical uncertainty together with the systematic uncertainties explained in section 5.12. One thing to note is that the event selection strongly reduces the non-$\mathrm{t\bar{t}}$ contributions to the background. Hence the agreement primarily depends on the $\mathrm{t\bar{t}}$ modeling. Overall, a very good agreement can be seen, with real data to MC simulation ratios being close to 1. Practically all deviations are seen only within the uncertainty bands. A slight slope of the ratio becomes visible in the p_{T} and m_{ll} distributions. This is an effect that is already known and is attributed to missing higher order effects [128, 129]. The QCD+EW correction shown in section 5.7.4 in fact mitigate this effect to a certain degree. There is further a sinusoidal shape visible in the ratio for the MET φ in the 2018 data. The collaboration provides a reweighting recipe to correct the effect [130]. However, this reweighting has not been applied in this analysis as the impact on the reconstructed top quarks was minimal.

For an extensive list of figures showing the data to simulation agreement, see appendix A.

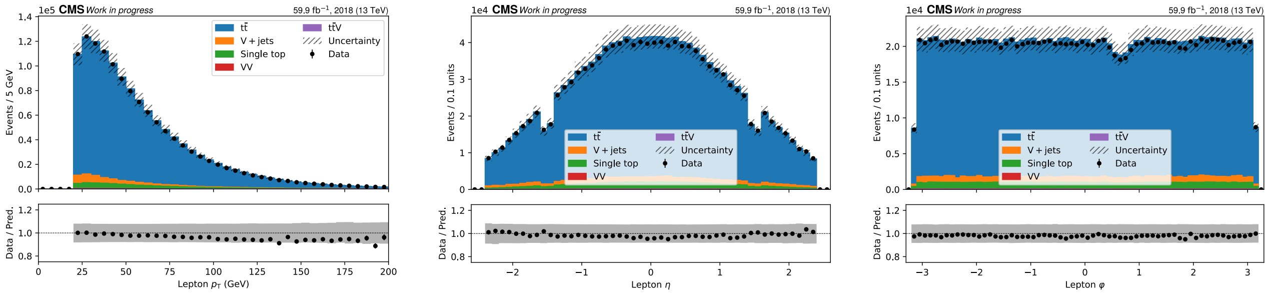
5.9 Reconstruction of the top quarks

To compute the observables used in the histograms for statistic evaluation and are detailed in the next section, one needs to know both four-momenta, the one of the top quark and the one of the antiquark. If all decay products were measured by the detector, reconstructing the top quarks would



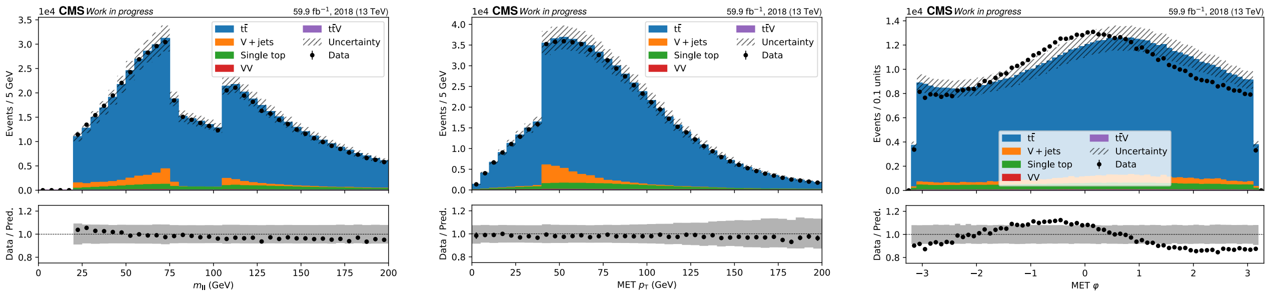
(a) (b) (c)

Figure 5.11: Comparison between real data and MC simulation (denoted as Pred. in the ratio plot) in 2018 for jet observables. All jets, as defined in section 5.5, found in the events are included, after full selection except for the last cut 15. Statistical as well as systematic uncertainties are included.



(a) (b) (c)

Figure 5.12: Comparison between real data and MC simulation (denoted as Pred. in the ratio plot) in 2018 for lepton observables. Both leptons, as defined in section 5.5, found in the events are included, after full selection except for the last cut 15. Statistical as well as systematic uncertainties are included.



(a) (b) (c)

Figure 5.13: Comparison between real data and MC simulation (denoted as Pred. in the ratio plot) in 2018 for the invariant lepton mass and MET. All cuts besides cut 15 are applied. Statistical as well as systematic uncertainties are included.

be a matter of adding four-momenta, thanks to the conservation of energy and momentum. However, the detector cannot directly measure the two neutrinos and they only show up as MET, together with everything else that was not measured or was mismeasured. For this reason a reconstruction algorithm is employed, that finds a solution based on the available information. To be precise: The four-momentum of the two bottom flavored jets, the two leptons and the MET.

The algorithm is based on the approach described in reference [131] and has found a wide usage in previous CMS and ATLAS publications [132–134]. The approach analytically solves the system of equations resulting from the conservation of energy and momentum. The eight degrees of freedom, arising from the two missing four-momenta of the neutrinos, are removed by several assumptions. Firstly, the masses of the neutrinos are set to 0, which is a reasonable choice, as the neutrino mass is anyway too small to affect the measurement. Secondly, the masses of the top quark and antiquark as well as the mass of the W boson, into which they decay, are assumed to be fixed values. This removes the ability to have off-shell top quarks or W bosons. Lastly, all the MET is assumed to be the result of the neutrinos, which associates the two observables p_{T} and φ of the MET purely to the neutrinos. This assumption is a good approximation in the regime of highly energetic neutrinos. Otherwise, the many other contributions to the MET and its associated uncertainty become significant. The obtained result is a quartic equation, with four possibly complex solutions.

For the implementation in this particular analysis, first the bottom quark, antiquark and the two oppositely charged leptons have to be identified from the measured objects. In the case of the leptons, the task is trivial, as the electric charge is measured without ambiguity and the event selection already restricts the events to exactly two oppositely charged leptons. On the other hand, the identification of the quark and antiquark with two jets in the event requires more steps that go as follows.

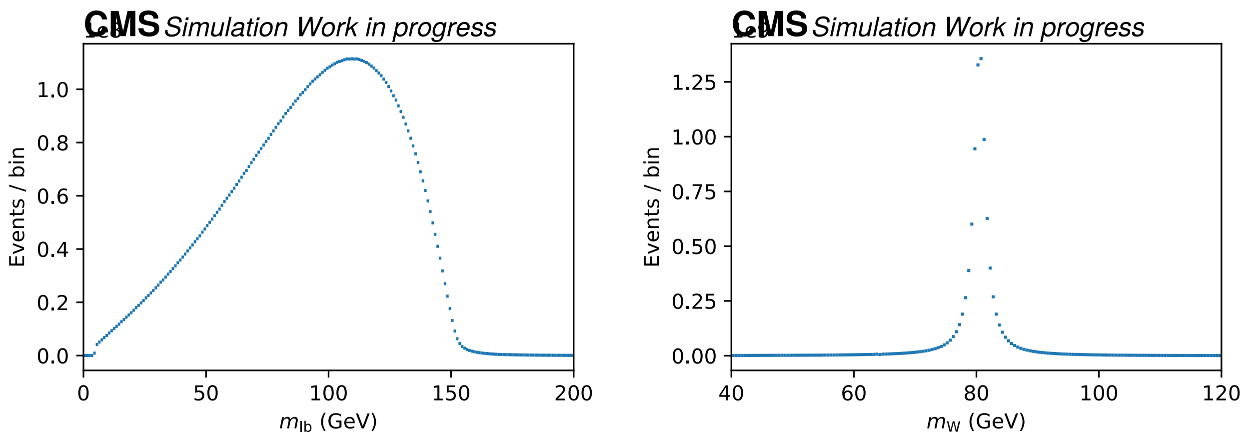
1. A set containing all jets is defined that have been determined to be bottom flavored.
    (a) If the event contains only one such jet, all other jets, which possibly are not bottom flavored, are also added to the set.
2. Associating an arbitrary jet from the set to the bottom quark and another one to the antiquark, the two invariant masses $m_{\mathrm{l^+b}}$ and $m_{\mathrm{l^-\bar{b}}}$ using the leptons are computed. This is done for all possible combinations.
3. The combination that maximize the probability

$$P_{\mathrm{lb}} = P(m_{\mathrm{l^+b}}) \cdot P(m_{\mathrm{l^-\bar{b}}}) \tag{5.7}$$

   is selected and the jets are assigned to the quark and antiquark accordingly.

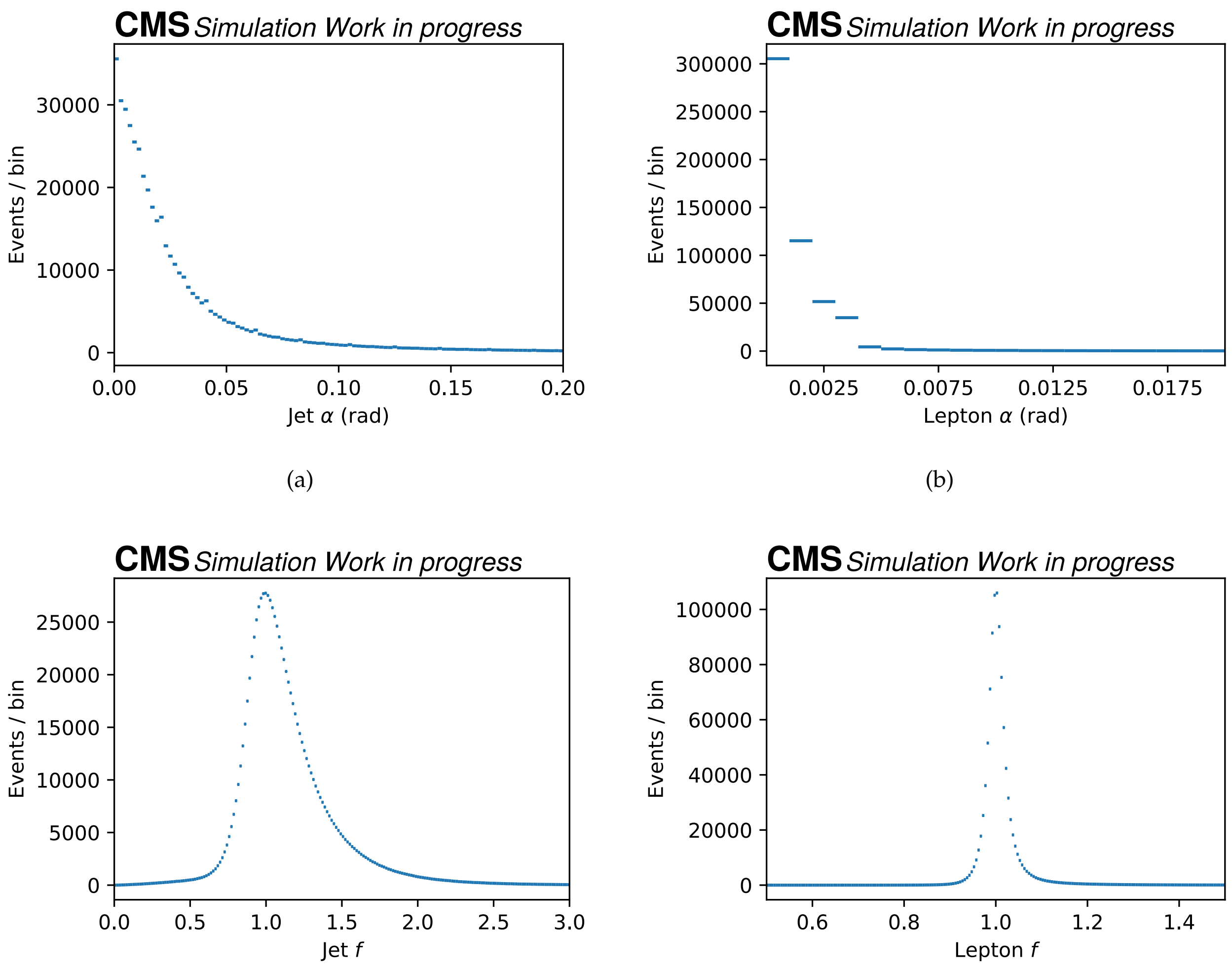
To compute the probability from above, a histogram of the m_{lb} distribution, that is derived from MC simulation before detector simulation, is used. The histogram can be seen in figure 5.14a. The procedure selects the two jets as the quark and antiquark that are most likely the ones coming from the respective top decay. A priority to the ones already seen as bottom flavored is given. Using the $\mathrm{t\bar{t}}$ simulation, the approach has been observed to pick the right jets in 69 % of the events, assuming the jets from both bottom quark and antiquark are present.

For the algorithm one has to set values for the top masses and W boson masses. The two top masses are set to 172.5 GeV, which agrees with the value set in the simulation. For the W^- and W^+ random values are picked according to the W boson mass distribution. The distribution is acquired from a histogram using the $\mathrm{t\bar{t}}$ (dileptonic decay) simulation. Figure 5.14b displays the histogram.



(a) (b)

Figure 5.14: Histograms used in the reconstruction algorithm of the top pair.



(a) (b) (c) (d)

Figure 5.15: Histograms used in order to determine random factors. The random factors are used to vary the inputs to the algorithm within their uncertainty.

The complex solutions from the quartic equation are ignored as being unphysical and only real solutions are kept. The numerical criteria used is Im(solution) $< 10^{-6}$. If there are multiple real solutions, the solution with the lowest $m_{t\bar{t}}$ is chosen. In around 60 % of the cases this is the best solution [135]. If an event does not have any real solutions, the reconstruction is being deemed as having failed and the event is discarded by the last cut in the event selection. Only about 50 % of the $t\bar{t}$ events pass the reconstruction this way. The reason why so many events fail it is because the system of equations holds true for ME level momenta, while the momenta that are measured also include all the effects from the particle shower and detector resolution. To solve the shortcoming, before applying the equations the values of the four-momenta of the leptons and jets that serve as input are varied randomly within their uncertainty. This is done in three steps: First the energy E is multiplied by a random factor f sampled from a histogram, which is shown in figure 5.15 on the lower row. However, the energy must not become smaller than the mass m and thus the resulting energy is

$$E' = \begin{cases} Ef & Ef \geq (1+\epsilon)m \\ (1+\epsilon)m & \text{else} \end{cases}, \tag{5.8}$$

where ϵ is set to 0.01. Secondly, the three-momentum is by a random angle α around a random axis orthogonal to the momentum. The random angle α is sampled from a histogram as well, which can be seen in the upper row of figure 5.15. Finally the three-momentum is scaled to keep the mass unchanged. The histograms for the distributions of f and α are produced for leptons and jets individually. To produce them, first all cut except cut 15 are applied to dileptonically decaying $t\bar{t}$ MC events. Then the jets and leptons are matched to particles before detector simulation (generator level). A jet and a generator level bottom quark or antiquark is considered matched if their ΔR is below 0.3 and if there is only one jet that matched the quark or antiquark. For a lepton to be matched it must have the same flavor and charge as the generator level lepton and has to also fulfill $\Delta R < 0.3$. To fill the f histogram, the ratio of energies of the generator level and the physical object is computed. For the α histogram, the difference in φ between both is computed. For every event, this variation procedure is repeated 100 times. After choosing the solution with the lowest $m_{t\bar{t}}$ for every input variation, as described before, the final solution, in form of four-vectors for the tops, is obtained from the up to 100 remaining solutions by computing the weighted average. The weight used here is P_{lb}, as given above, but calculated using the varied inputs. After the averaging, the energy of the four-vectors is adjusted so that the final mass still is 172.5 GeV.

The variation procedure reduces the amount of events that fail the reconstruction to about 10 % for dileptonically decaying $t\bar{t}$ events, to about 14 % for semileptonically decaying ones and to about 23 % for DY events. The difference in this efficiency for the different event types shows that the algorithm is slightly sensitive to whether the event truly is a dileptonic $t\bar{t}$ events or not. The goal of the algorithm however is to give a handle on top quarks under the assumption the event contains dileptonic $t\bar{t}$ process. The observed sensitivity is therefore an unexpected side effect. While the event selection due to the failure of the reconstruction slightly reduces the contribution of unrelated backgrounds, it is far from being a viable criterion to separate, for example, $t\bar{t}$ and DY. Additionally, the efficiency slightly decreases towards higher energies, as can be seen in its dependence on the lepton p_{T} in figure 5.16. The figure shows three efficiencies, one for no variation and a single fixed W boson mass, one for no variation and a sampled W boson mass and one for the case with input variation. Only the case with variation visibly increases the efficiency.

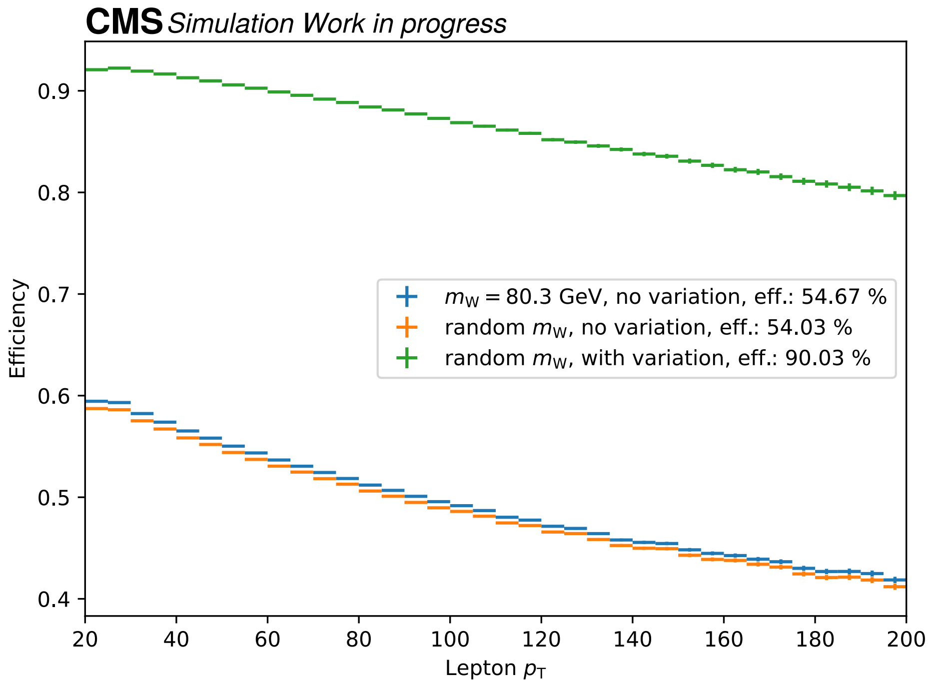
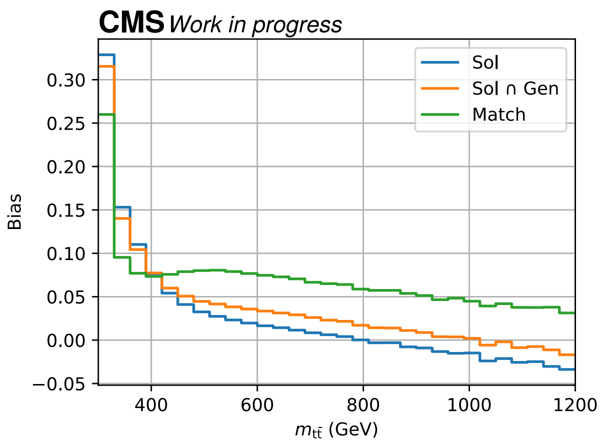
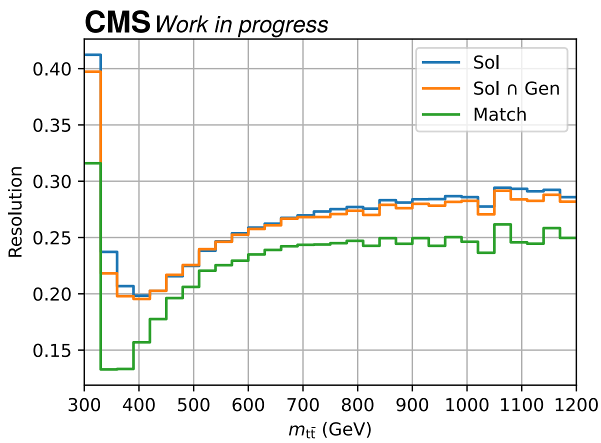


Figure 5.16: Efficiencies of the reconstruction in dependence of lepton p_T for different settings of the algorithm evaluated on dileptonically decaying SM $t\bar{t}$. The efficiency drops slowly towards higher p_T. Randomizing the W boson mass is seen to not result in a significant difference, while applying the variation procedure does.



(a)



(b)

Figure 5.17: Bias and resolution for the reconstruction algorithm for the reconstructed $m_{t\bar{t}}$ in dependence to the $m_{t\bar{t}}$ before detector simulation.

In conclusion, the algorithm yields two four-vectors, one for the top quark and one for the top antiquark. From here, many other observables can be unambiguously determined, in particular $m_{t\bar{t}}$ or any of the other important ones for this analysis (see section 5.10). In figure 5.17 $m_{t\bar{t}}$ is used for a benchmark of the algorithm. Shown are the mean (figure 5.17a) and the standard deviation of the relative error to the true $m_{t\bar{t}}$ (figure 5.17b). The former is referred to as bias, while the latter is referred to as resolution. The true $m_{t\bar{t}}$ is defined as the $m_{t\bar{t}}$ before the detector simulation. For both, bias and resolution, values closer to 0 imply better performance. The event selection of the analysis is applied (denoted by "Sol"). Furthermore, additional cuts are applied to test the algorithm's performance in a particular well behaved region of events. For the "Sol $\cap$ Gen" curve, these additional cuts are applied:

- Lepton $p_T > 30\,\text{GeV}$ and $\eta < 2.4$, $\Delta R > 0.1$ between the two leptons
- Bottom $p_T > 30\,\text{GeV}$ and $\eta < 2.4$, $\Delta R > 0.1$ between the two bottom (anti)quarks
- Between bottom (anti)quarks and leptons $\Delta R > 0.4$

In addition, for the "Match" curve, the jets that are assigned to the bottom quarks and antiquarks are required to match the quarks and antiquarks on generator level with $\Delta R < 0.3$. The cuts have been taken from reference [136]. It can be seen that for the matched selection there persists a small bias towards higher $m_{t\bar{t}}$ values of around 5 %, while the "Sol" selection comes closer to 0. The resolution can be seen to worsen towards higher $m_{t\bar{t}}$ values, where it reaches 25 % to 30 %. With the match selection, the resolution overall is better. In general, the resolution is the best at $m_{t\bar{t}}$ values of around 400 GeV. At very low values (around two times to top quark mass), the top quarks are produced almost at rest, which makes a reconstruction increasingly difficult. This results in the bias and resolution going up at low values.

5.10 Search observables

The analysis uses three-dimensional histograms of three observables to optimize the sensitivity of the search. The first observable is the invariant mass of the top quark pair $m_{t\bar{t}}$. As shown in section 2.8, the heavy Higgs resonance creates a weighty peak on the $m_{t\bar{t}}$ distribution, while the interference is visible as well with its distinctive shape including the dip. The SM $t\bar{t}$ production cross section peaks at a $m_{t\bar{t}}$ value of around 400 GeV. The change from the heavy Higgs boson affects the entire $m_{t\bar{t}}$ range, starting from twice the top quark mass until shortly above the mass of the Higgs boson, as can be seen in figure 5.18. The presence of the dip coming from the interference is only expected to be visible at low masses. At higher masses, the cross section will be low and effect of the PDF will create the most noticeable contribution in form of a peak (compare with figure 2.11). In this case the dip might not survive the limited resolution of the top quark reconstruction, as its events are canceled out by part of the peak.

A clear criterion of distinction between the SM $t\bar{t}$ production, which at LO involves the s-channel with a gluon, and the signal, which includes an s-channel with a Higgs boson, is the particles' difference in spin and parity. This difference is expressed by changes in the distributions of spin correlation observables used in the analysis. The spin correlation observables are derived using the lepton's four-momenta and the ones of the top quark and antiquark, reconstructed using the algorithm from section 5.9. The leptons serve as an excellent handle on the spin of the tops as described in section

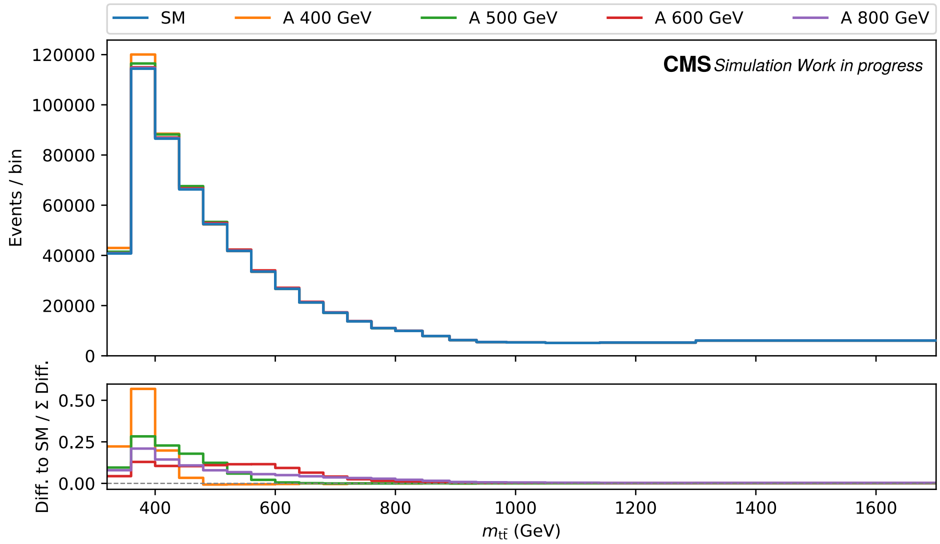


Figure 5.18: Histograms for the $m_{t\bar{t}}$ distribution of SM and different pseudoscalar heavy Higgs bosons with a relative decay width of 2.5 % after event selection and reconstruction. The upper panel shows the distribution of the $t\bar{t}$ production including a Higgs boson, while the lower panel shows its difference to the SM only hypothesis. The counts shown in the lower panel have been scaled to 1 by dividing by the sum of all counts for better visibility. In the lower panel, the change of the position of the signal due to the different Higgs masses becomes apparent. Due to the limited resolution of the top reconstruction, the dip of the Higgs interference is only faintly if at all visible.

2.7. The observables are evaluated using a $(\hat{k}, \hat{r}, \hat{n})$ coordinate system, that is constructed especially for the task of spin correlation in conjunction with $t\bar{t}$ [137]. The construction works as follows:

1. The top quark and top antiquark four-momenta are boosted into the zero-momentum frame (ZMF) of the $t\bar{t}$ system. $\hat{k}$ is defined as the unit vector in the flight direction of the top quark in this reference frame. By definition $-\hat{k}$ is the flight direction of the top antiquark.

2. $\hat{n}$ is defined as a unit vector orthogonal to $\hat{k}$ and $\hat{z}$ (the direction of the z axis, which is also the direction of one of the proton beams), and together with $\hat{k}$ spans the plane in which the tops are produced. $\hat{n}$ is defined using

$$\hat{n} = \begin{cases} \frac{1}{\sin\theta_t}\left(\hat{z} \times \hat{k}\right), & \cos\theta_t \geq 0 \\ -\frac{1}{\sin\theta_t}\left(\hat{z} \times \hat{k}\right) & \text{else} \end{cases} \tag{5.9}$$

where θ_t is the scattering angle of the top quark defined by $\cos\theta_t = \hat{z} \cdot \hat{k}$.

3. The coordinate system is completed by defining $\hat{r}$ as the remaining orthogonal vector, giving rise to a right-handed system.

4. The four momenta of the two charged leptons are boosted into the $t\bar{t}$ ZMF and from there into their respective top ZMF using the four-momenta obtained in step 1. This means the antilepton four-momentum is boosted along $\hat{k}$ and the lepton four momentum is boosted along $-\hat{k}$. The frame of motion constructed this way is referred to as helicity frame.

5. Define

$$\cos\theta^{\pm}_{\hat{a}} = \pm\hat{l}^{\pm} \cdot \hat{a}, \tag{5.10}$$

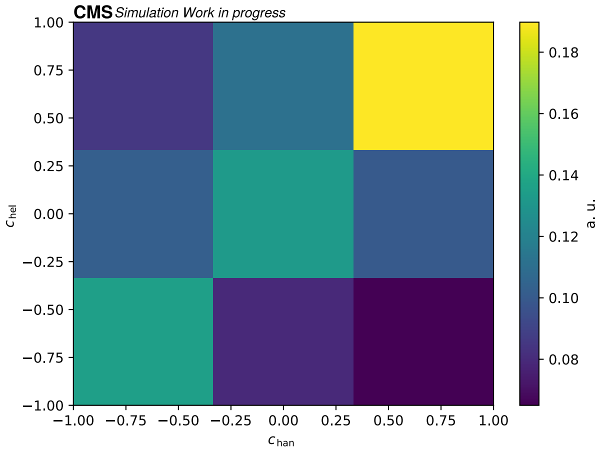
where $\hat{l}^+$ and $\hat{l}^-$ are the unit vectors in the flight direction of the antilepton and the lepton respectively in the helicity frame resulting from the previous step and $\hat{a}$ is one of $\hat{k}$, $\hat{r}$ or $\hat{n}$.

6. Define $\vec{\Theta}^{\pm}$ as the 3 element vectors with $\cos\theta^{\pm}_{\hat{a}}$ as entries and Λ as the 3 by 3 matrix with elements
$$\Lambda_{\hat{a}\hat{a}'} = \cos\theta^{+}_{\hat{a}}\cos\theta^{-}_{\hat{a}'}, \tag{5.11}$$
where, again, $\hat{a}, \hat{a}' \in \{\hat{k}, \hat{r}, \hat{n}\}$.

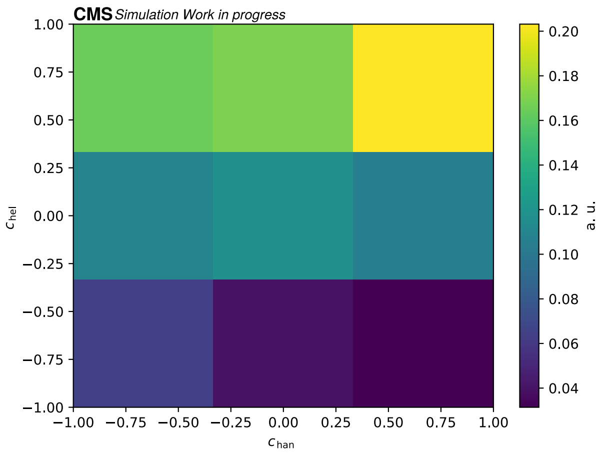
It can be shown that for the differential cross section of the $\mathrm{t\bar{t}}$ production with dileptonic decay adheres
$$\frac{1}{\sigma}\frac{\mathrm{d}^2\sigma}{\mathrm{d}\cos\theta^{+}_{\hat{a}}\ \mathrm{d}\cos\theta^{-}_{\hat{a}}} = \frac{1}{4}\left(1 + B^{+}_{\hat{a}}\Theta^{+}_{\hat{a}} + B^{-}_{\hat{a}'}\Theta^{-}_{\hat{a}'} - C_{ij}\Lambda_{\hat{a}\hat{a}'}\right),$$
where the vectors $\vec{B}^{\pm}$ contain information about the spin polarization of one of the top quarks, while the C matrix contains information on the spin correlation between both top quarks [116]. Thus by the usage of $\vec{\Theta}^{\pm}$ and Λ, one can infer this information. The two observables used in the analysis are defined by
$$c_{\mathrm{hel}} = -\mathrm{Tr}\Lambda = -\Lambda_{\hat{k}\hat{k}} - \Lambda_{\hat{r}\hat{r}} - \Lambda_{\hat{n}\hat{n}} \tag{5.12}$$
and
$$c_{\mathrm{han}} = \Lambda_{\hat{k}\hat{k}} - \Lambda_{\hat{r}\hat{r}} - \Lambda_{\hat{n}\hat{n}}. \tag{5.13}$$
Note the difference in the sign of the first term between the two definitions. It can be shown that c_{hel} is the cosine of the angle between the leptons in their helicity frames. Both, c_{hel} and c_{han}, are contained within the interval $[-1;1]$.

The two observables allow the distinction of the SM from the Higgs boson and between the two parities of the Higgs bosons A and H, as can be seen in figure 5.19. The SM has its maximum for both observables at 1. The opposite can be seen in the case of H, where both peak at -1. The distribution for A on the other hand shows a strong slope with respect to c_{hel} with maximum at 1, while it only has a light slope in the case of c_{han}. The reason for this is that c_{hel} and c_{han} are sensitive to the spin properties of the particles that are part of the spin production. In case of A or H, the tops are the product of the decay of a spin-0 boson (the heavy Higgs boson), while, for example, in the SM s-channel case, instead one finds a spin-1 gluon.

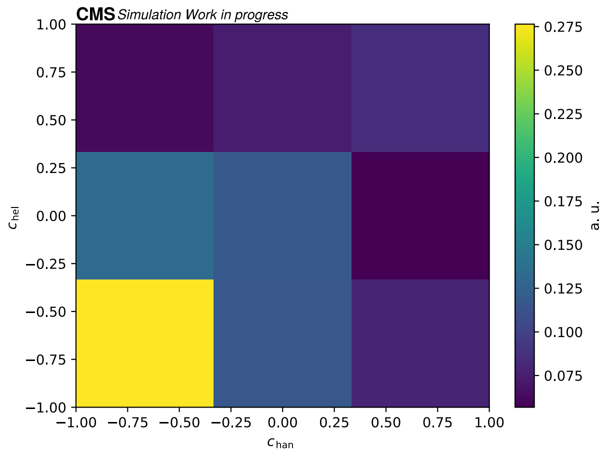
The three-dimensional histograms are constructed using these three observables together with a specific binning. The binning has been optimized to be used with A and H and has to solve a tradeoff problem between sensitivity and statistical uncertainty, as smaller bins imply larger relative statistical uncertainties. For the $m_{\mathrm{t\bar{t}}}$ it is crucial to be able to capture distinctive shape of the signal over the large range of different Higgs masses. In conclusion these $m_{\mathrm{t\bar{t}}}$ bin edges in GeV have been chosen
$$\{520; 560; 600; 640; 680; 720; 760; 800; 845; 890; 935; 985; 1050; 1140; 1300; 1700\},$$
amounting to a total of 15 bins. The last bin can be seen as an overflow bin, as it captures a wide range, that is above the largest Higgs mass and beyond of which practically no events are expected. For the spin correlation observables, an equal distance binning dividing the $[-1;1]$ range into three has been chosen, which results in 9 total bins for $c_{\mathrm{hel}} \times c_{\mathrm{han}}$. The full three-dimensional histogram therefore has 135 bins. The binnings for the individual observables have already been brought into action in figure 5.18 and 5.19. Additionally figure 5.18 includes five uniform bins with sizes of 40 GeV, which makes the characteristic peak-dip structure of the heavy Higgs boson for low masses visible. While



(a) SM



(b) A(400 GeV,2.5 %)



(c) H(400 GeV,2.5 %)

Figure 5.19: Histograms of the distribution of the two spin correlation observables c_{hel} and c_{han}. The histograms have been scaled to 1 for better visibility. The lower row shows the influence of the pseudoscalar and scalar Higgs bosons in terms of $\mathrm{t\bar{t}}$ production (including resonance and interference) minus SM $\mathrm{t\bar{t}}$ production.

the analysis has been developed also considering this region, these five bins are kept blinded for the analysis as presented in this thesis. Correctly modeling the $\mathrm{t\bar{t}}$ background in this low $m_{\mathrm{t\bar{t}}}$ region is part of ongoing research within the collaboration. A future publication is planned that will unblind also the region.

5.11 Limit calculation

The final result of the analysis is a statement on whether the data collected by CMS can rule out the coupling strength between the top quark and the heavy Higgs bosons at masses and decay width probed in the analysis for the specific decay channel. The statement is produced in form of a hypothesis test that constructs a confidence interval for the expected value under background-only for some parameter of interest (POI) and for the observed value of the POI. The expected one is determined from the simulation, while the observed one is coming from experimental data. If experimental data was identical to the background simulation, both intervals would agree. In particular physics, it has become standard to quote the upper limit of the POI, at which the probability to obtain a higher value is 5 % or lower. This is referred to as 95 % confidence level (CL).

Often analyses searching for new particles pick the cross section of the probed process as POI. In this case processes at a specific cross section with the new particle can be ruled out by an upper limit smaller on the cross section. However, in the case of this analysis, the cross section does not allow the formulation of a useful statement, as the cross section can also become close to 0 from increased contribution from destructive interference. Therefore the modification to the coupling strength modifier $g_{\mathrm{A/H}}$ is chosen as the POI. The contribution of the resonance grows with $g^4_{\mathrm{A/H}}$ and the one of the interference with $g^2_{\mathrm{A/H}}$ (see section 5.4), while a value of $g_{\mathrm{A/H}} = 0$ indicates the absence of signal.

The procedure used in the analysis is based on the statistical tests described in reference [138]. To formulate the hypothesis test, first the function L is computed. L is the likelihood that the expected or observed event count is seen within some uncertainty given the predicted event count obtained for a signal and background scenario. It is defined by

$$\begin{aligned} L\left(\mathrm{count}|g_{\mathrm{A/H}},\theta\right) &= \mathrm{Poisson}\left(\mathrm{count}|s(g_{\mathrm{A/H}},\theta) + b(\theta)\right)\cdot\nu(\theta) && (5.14)\\ &= \prod_i \frac{\left(s_i(g_{\mathrm{A/H}},\theta)+b_i(\theta)\right)^{n_i}}{n_i!}\exp\left(-s_i(g_{\mathrm{A/H}}) - b_i\right)\cdot\nu(\theta). && (5.15)\end{aligned}$$

Poisson denotes the probability mass function of the Poisson distribution, which has been written explicitly in the second line. As the analysis uses histograms, this is a product of the individual probabilities of each bin, indexed by i. The counts of each bin is given by b_i for the background simulation, s_i for the signal simulation and by n_i for the expected or observed event count. The counts for the signal simulation is defined by

$$s_i(g_{\mathrm{A/H}},\theta) = g^4_{\mathrm{A/H}}\cdot s_{\mathrm{res},i}(\theta) + g^2_{\mathrm{A/H}}\cdot s_{\mathrm{int},i}(\theta), \tag{5.16}$$

where $s_{\mathrm{res},i}$ is the count from the resonance simulation and $s_{\mathrm{int},i}$ is the count from the interference simulation. Both have a dependence on the coupling strength modifier $g_{\mathrm{A/H}}$ of the heavy Higgs boson. In these definitions, θ is a vector of nuisance parameters, which control the presence of the uncertainties. A value of $\theta_j = 1$ denotes an addition of one standard deviation from the jth uncertainty to the signal or background, while $\theta_j = -1$ denotes the subtraction of such. Extreme values of θ are made unlikely by the probability density ν, that is part of the likelihood.

Finally a test statistic is defined by the likelihood ratio

$$q(\mathrm{count}|g_{\mathrm{A/H}}) = -2\ln\frac{L\left(\mathrm{count}|g_{\mathrm{A/H}},\hat{\theta}(g_{\mathrm{A/H}})\right)}{L\left(\mathrm{count}|\hat{g}_{\mathrm{A/H}},\hat{\theta}\right)}, \tag{5.17}$$

where $\hat{\theta}(g_{\mathrm{A/H}})$ is the vector of nuisance parameter values that maximizes the likelihood given the value of $g_{\mathrm{A/H}}$ and $\hat{g}_{\mathrm{A/H}}$ and $\hat{\theta}$ are the values that give the global maximum of the likelihood. The CL is then obtained from

$$p_{g_{\mathrm{A/H}}} = \int_{q(\mathrm{count}|g_{\mathrm{A/H}})}^{\infty} f(q_{g_{\mathrm{A/H}}}|g_{\mathrm{A/H}})\mathrm{d}q_{g_{\mathrm{A/H}}}, \tag{5.18}$$

where $f(q_{g_{\mathrm{A/H}}}|g_{\mathrm{A/H}})$ is the probability density function of $q_{g_{\mathrm{A/H}}}$ given $g_{\mathrm{A/H}}$. The analysis is utilizing the CLs method, meaning it is using the following ratio to compute the CL:

$$\frac{p_{g_{\mathrm{A/H}}}}{1-p_0}, \tag{5.19}$$

where p_0 is $p_{g_{A/H}}$ with $g_{A/H}$ set to 0, to get the value for a background-only assumption. The CLs method is preferred due to its more reliable results at a low sensitivity to the signal [1]. The implementation of the calculation is done with the program package Combine [139].

Before the CL is computed, the nuisance parameters are fit to the data to obtain their most probable values as well as probable uncertainties. This creates two different definitions of the uncertainties, one that is "prefit" and another one that is obtained "postfit". The post-fit values are used in the CL computation. The values for the systematic uncertainties given in the next section denote the prefit values.

The probability density functions ν belonging to the uncertainties are selected based on the specific uncertainty. For statistical uncertainties the Barlow-Beeston method is used [140]. This imposes a Poisson distribution for every bin as an independent uncertainty. When considering systematic uncertainties it depends on the specific source, which is presented in the next section. However, most of the systematic uncertainties are based on additional histograms, in which case a Gaussian distribution is used.

5.12 Sources of systematic uncertainties

One source of uncertainty in the measurement of the heavy Higgs bosons is the statistical one coming from the limitation of the number of events. On the other hand, there are also the systematic uncertainties that are the effect of inherent inaccuracies in the way the measurement is performed. In fact, for this analysis the systematic uncertainties outweigh the statistical uncertainty. Each source of uncertainty is being approximated in form of two additional histograms and the procedure to define them goes as follows. First, every source is assigned one nuisance parameter with an uncertainty described by two possibly different standard deviations, one for positive and one for negative variations ("up" and "down"). When all nuisance parameters are at their central value, the search distribution is described by the histogram predicted by the standard MC simulation. When one of the parameters is at its plus or minus one standard deviation value, the expected distribution is described by one of the additional histograms for the uncertainty. The uncertainty histograms therefore have the same form as the histograms described in section 5.10 but contain counts different from the central histograms. The uncertainty histograms are referred to as up and down variations. All uncertainties are assumed to be independent from each other. Furthermore, in this analysis, if not otherwise stated, the uncertainties are assumed to be correlated across analysis channels but independent across years. The analysis uses a range of different uncertainty sources, which are explained in the following.

- The **top Yukawa coupling** Y_t is associated to an uncertainty. Its value has been measured to be $1.22^{+0.14}_{-0.12}$ with a systematic uncertainty of $+0.12$ and -0.10 [141]. As described in section 5.7.4, event weights correcting the EW effects in dependence of Y_t have been calculated. For the central prediction a value $Y_t = 1$ is used, while event weights for the up and down variations are made using the Y_t values of $1 + 0.12$ and $1 - 0.10$. The shape of the uncertainty can be seen in figure 5.20.

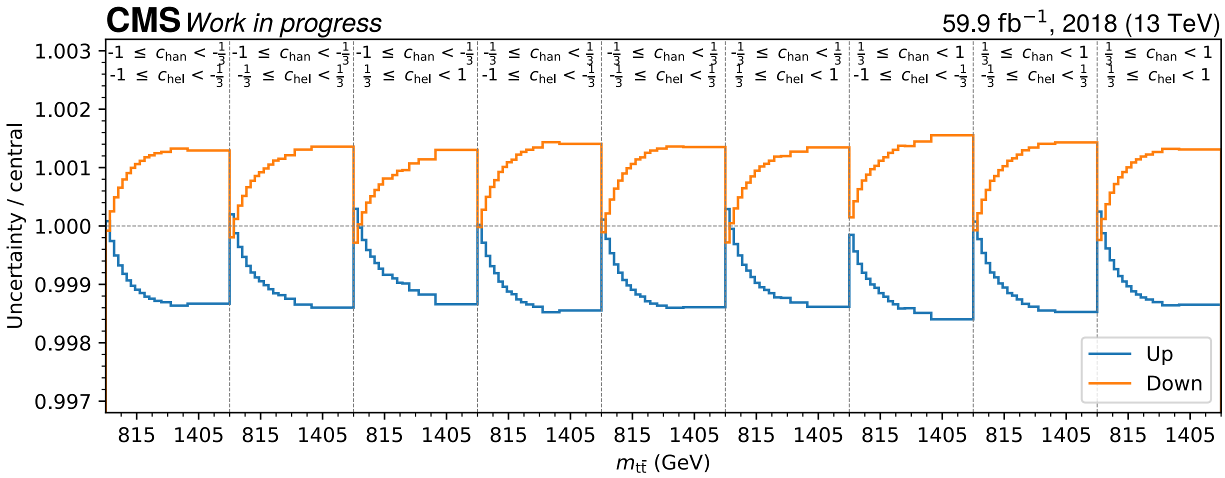


Figure 5.20: Ratio between the uncertainty variations for the top Yukawa coupling and the central histogram for SM $t\bar{t}$ in the 2018 data analysis.

- There is a **difference between multiplicative and additive approaches for the EW correction**. As described in section 5.7.4, the events are reweighted to correct for higher order EW effects by calculating the difference based on LO QCD simulation [142]. However, the background prediction is using NLO precision, giving rise to an ambiguity on how to interpret the weights. One possibility is to multiply them with the LO QCD contribution only and the NLO QCD contribution is added on top (additive approach). Another possibility is multiplying the weights with the sum of LO and NLO contributions (multiplicative approach). The difference between the approaches is a term $\delta_{\mathrm{NLO}} \cdot \delta_{\mathrm{EW}}(Y_t)$, where $1 + \delta_{\mathrm{EW}}(Y_t)$ is the weight for the EW correction and δ_{NLO} is the relative difference between the LO and NLO QCD. Including an uncertainty based on the difference from the two approaches, the new event weight becomes

$$(1 + \delta_{\mathrm{EW}}(Y_t)) \cdot (1 + \delta_{\mathrm{NLO}} \cdot \delta_{\mathrm{EW}}(Y_t))^{\phi}\,, \tag{5.20}$$

where $\phi = 0$ for the central histogram and $\phi = \pm 1$ for the up and down variations. δ_{NLO} is computed on a per bin basis using

$$\delta_{\mathrm{NLO},i} = \frac{n_{\mathrm{NLO},i} - n_{\mathrm{LO},i}}{n_{\mathrm{NLO},i}}, \tag{5.21}$$

where $n_{\mathrm{NLO},i}$ and $n_{\mathrm{LO},i}$ are the bin heights for the ith bin for NLO and LO-only, respectively. The shape of the uncertainty can be seen in figure 5.21.

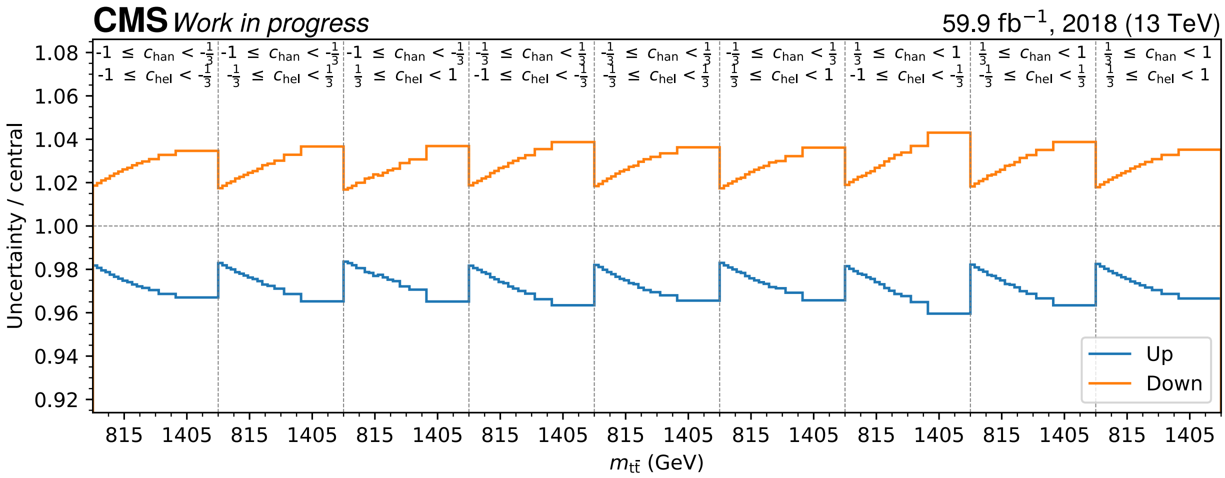


Figure 5.21: Ratio between the uncertainty variations for the multiplicative-additive uncertainty and the central histogram for SM in 2018.

- The **jet energy correction and resolution** come with uncertainties associated to it [114]. These are provided by the collaboration. The energy correction has a range of own sources, which are treated as independent in this analysis. Only the sources important to the jets that are within the analysis' selection are used. For example uncertainties that only affect jets with high η, that is excluded by the analysis selection, are not used. The sources are split into correlated and uncorrelated ones across years. The different sources change the energy of the jets, which means the computation of the varied histograms is done by rerunning the analysis implementation until histogram production.

- There is an uncertainty due to **MET coming from energy of particles that are not clustered in jets** by the PF algorithm. The uncertainty is provided by the collaboration in form of event weights.

- The **cross sections of the background processes** are computed up to a specific accuracy only. Exceptionally the uncertainties are implemented not using histograms but rather as a constant factor. This is done assuming the uncertainties are distributed according to a log-normal with a standard deviation depending on the process. The single top processes (each channel independently) are assigned a 15 % uncertainty. The diboson processes are assigned one uncertainty of 30 %. The DY process is given an uncertainty of 5 %. Finally, the most prevalent $t\bar{t}$ background is not included in this uncertainty source, as its uncertainty is accounted for by the ME renormalization and factorization scales.

- The **integrated luminosity** value, that is used to scale the MC events to, has an uncertainty. This is implemented as a log-normally distributed source, similar to the cross section uncertainties, and assumes a single uncertainty of 1.6 % [143].

- The value of the **strong coupling constant α_S used for the simulation inside the shower** has an uncertainty. It is approximated by scaling it by a factor of 2 (up variation) and 0.5 (down variation). The uncertainty is provided by the collaboration in form of event weights computed using PYTHIA.

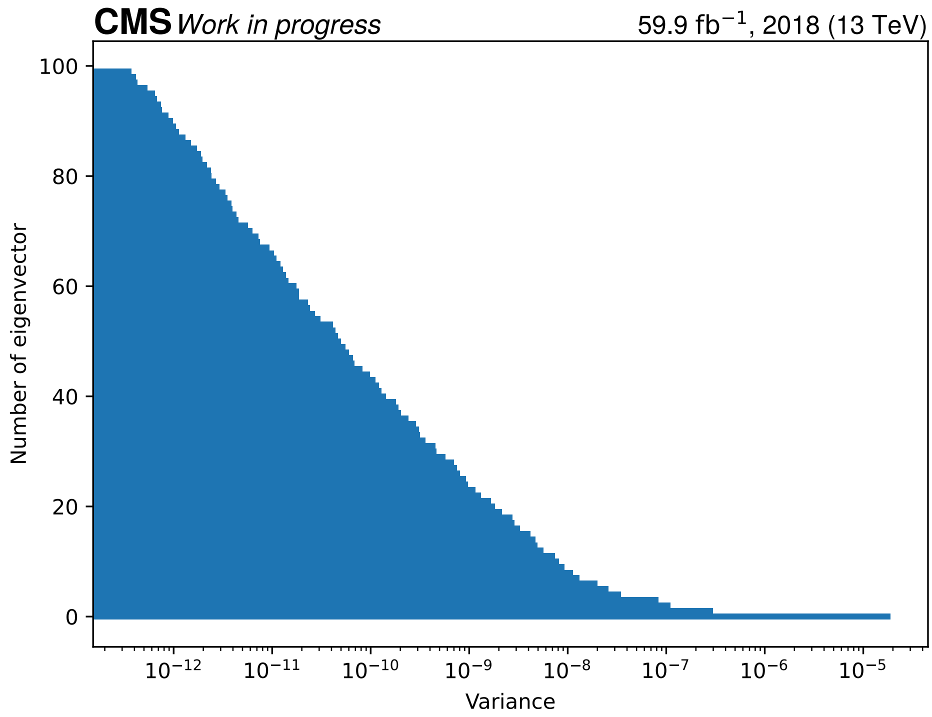
- The **parameter of the MC simulation $h_{\mathbf{damp}}$** is varied to give an uncertainty. h_{damp} is a parameter of POWHEG that influences the amount events with high p_T radiation. The parameter's value is determined from real data and has an uncertainty associated to it [106]. The uncertainty is approximated by dedicated simulation data sets of $\mathrm{t\bar{t}}$, which have been created with a varied value of h_{damp}. The analysis implementation is run on these data sets and the resulting histograms replace the $\mathrm{t\bar{t}}$ part of central histograms. The data sets are provided by the collaboration.

- Differences from different **color reconnection** models are included as an uncertainty. Variations from three alternative color reconnection models are included as uncertainty. The three models are called "QCD-inspired", "gluon move" and "early resonance decays" [144]. The uncertainties are addressed from dedicated data sets of $\mathrm{t\bar{t}}$ generated by the collaboration.

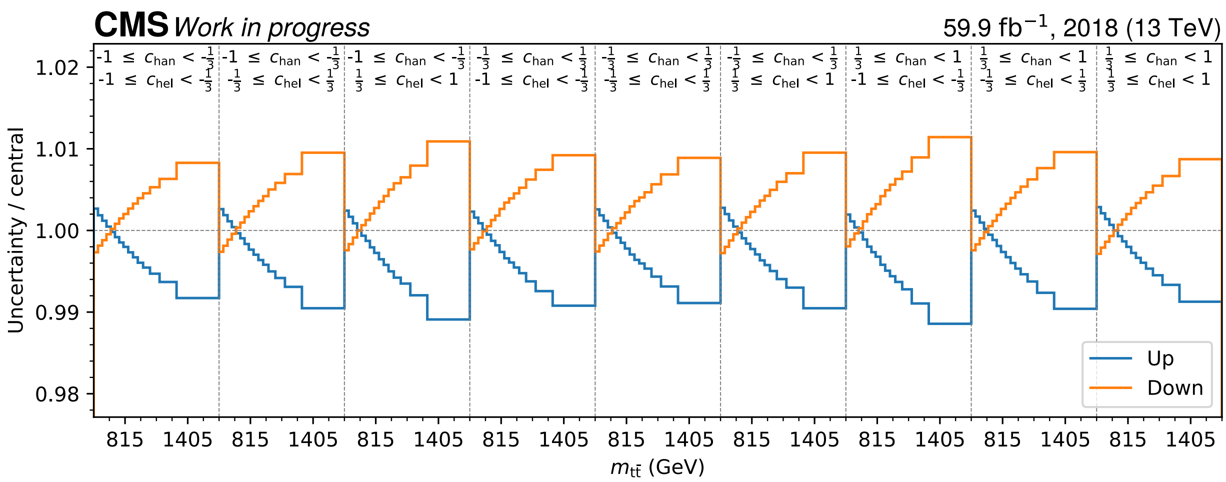
- Uncertainties due to changes in the **ME renormalization and factorization scales** for $\mathrm{t\bar{t}}$ and heavy Higgs boson simulation are included. In principle, these parameters should not change the result of the computation. However, due to the limited accuracy of the underlying perturbative calculations, clear effects are visible and thus this is included as one uncertainty of the simulation. The scales are multiplied by 0.5 and 2 separately and this is done using event weights provided by the collaboration. In this analysis the uncertainties also change the overall event count according to the NNLO QCD cross sections at these different scale values. This replaces a $\mathrm{t\bar{t}}$ cross section uncertainty.

- **PDF** uncertainties are provided by the authors of the used PDF [6] in form of 100 varied PDFs. These are given as event weights by the collaboration. A large majority of these 100 variations show practically no impact in this analysis, which is why a principal component analysis (PCA) is used to extract only the most important variations. The PCA is performed using the relative differences of the 100 resulting histograms with the central one and works using the following steps. Each bin difference is used to create a vector, giving in total $4 \times 3 \times 135$ (number of years times number of channels times number of search histogram bins) vectors. These vectors have 100 elements, one for each variation. The covariance matrix of these vectors is computed together with its 100 eigenvalues and eigenvectors. The eigenvectors now become the uncertainties, while the eigenvalue is the variances corresponding to the uncertainty. It turns out that only the uncertainty with the highest variance is important, because second largest is already two orders of magnitude smaller, as can be seen in figure 5.22a. The uncertainty is treated as an up variation and is multiplied by -1 to obtain a symmetric down variation. The result can be seen in figure 5.22b. Additionally an uncertainty of the PDF coming from the uncertainty on α_S is included in the analysis. The uncertainty is also provided by the collaboration as event weight.

- The **top quark mass** has an uncertainty, which is accounted for in the analysis. The collaboration provides dedicated simulated data sets for $\mathrm{t\bar{t}}$, in which the parameter for the top quark mass has been changed by $\pm 3\,\text{GeV}$. The analysis implementation is rerun on these data sets to produce varied histograms. The histograms replace the $\mathrm{t\bar{t}}$ contribution in the central one to give the uncertainty estimate. Because the top quark mass uncertainty is only at the level of $\pm 0.5\,\text{GeV}$, the final uncertainty is the result of scaling the difference between the varied histogram and the central one by $\frac{1}{6}$.



(a)



(b)

Figure 5.22: Results of the PCA done for the PDF. The upper figure shows the variance associated to each resulting eigenvector. Only the uncertainty corresponding to eigenvector 0 (shown in the lower figure) is kept, as its variance is larger by at least two orders of magnitude compared to the other.

- The **tune**, which is a set of assorted configuration parameters that go into the simulation, has been determined in reference [106] and was derived together with an uncertainty. The values of the parameters are changed accordingly and an additional SM $t\bar{t}$ simulation with the changed parameters is performed. For this purpose the collaboration provides dedicated simulated data sets, in which the parameters have been changed. The analysis implementation is rerun on these data sets and the result replaces the $t\bar{t}$ contribution in the central histogram to produce varied histograms.

- The total inelastic cross section used in the determination of the **pileup reweighting** is associated with an uncertainty. The cross section is varied by $\pm 4.6\,\%$ from its central value of 69.2 mb to produce two additional sets of event weights. The weight is scaled so that no change to the sum of event counts is done compared to the central weight for pileup reweighting.

- The uncertainties of the reweightings for the **electrons, muons and bottom flavored jets** are included in the analysis. The different types of reweightings (identification, isolation, reconstruction) are treated independently [109, 125]. The muon uncertainties are further split into statistical and systematic parts. The statistical part of the muon uncertainties are treated independently across years. The uncertainty for the bottom flavored jets algorithm is further split into a range of individual independent contributions. They are evaluated independently for jets that are actually bottom flavored in MC and for those that are not. All uncertainties are provided by the collaboration in form of event weights.

- The reweightings for the **trigger efficiencies and L1 trigger prefiring** are associated to uncertainties that cover effects including statistical uncertainties on the trigger efficiency evaluation or because of correlation between the triggers [127]. The uncertainties are provided by the collaboration as event weights.

5.13 Smoothing of the systematic uncertainties

As described in section 5.11 the systematic uncertainties based on histograms are fitted and included using a Gaussian distribution. A common issue of this procedure is that these uncertainties get artificially constrained. The cause is that the prediction of the uncertainties contains fluctuations on the per-bin level, meaning the histograms are noisy. This prompts the fitting algorithm to assign a very small standard deviation to the nuisance parameter, as the shape of the noisy histogram does not fit to the expected or observed data. As a consequence, the contribution from the uncertainty is very limited if present at all. The artificial constraints are especially prevalent in the case of systematic uncertainties that have been approximated by rerunning the analysis implementation. To mitigate this effect some of the uncertainty approximations are smoothed.

The smoothing is done using the LOWESS algorithm (locally weighted scatterplot smoothing) [145]. The procedure utilized in the analysis follows the recipe given in reference [146]. LOWESS obtains a smoothed function that still resembles the input data. It does so by performing a weighted least-squares fit considering all points within a sliding window. For the analysis the relative difference between the central histogram and the uncertainty histogram is smoothed. The size of the window is called bandwidth and is a parameter to be chosen before running the algorithm. As done in the reference, the bandwidths for the usage within the analysis are determined by a 10-fold cross-validation. This means the events are randomly split into 10 parts of equal size, producing 10 histograms. For all values from a set of possible bandwidths the smoothing is performed on the sum of 9 histograms. For each result the χ^2 value with the remaining histogram is computed. The χ^2 between two histograms a and b with N bins is given by

$$\chi^2 = \sum_i^N \frac{(a_i - b_i)^2}{\sigma^2_{\text{central},i} + \sigma^2_{\text{unc},i}}, \tag{5.22}$$

where $\sigma_{\text{central},i}$ and $\sigma_{\text{unc},i}$ are the statistical uncertainties of bin i of the central and the uncertainty histograms, from which the relative difference was computed. χ^2 is a measure for how similar a and b are. The process is repeated for all ways the histograms can be split up and the bandwidth giving the smallest average χ^2 is picked. The cross-validation is in turn repeated 500 times to be invariant to the random choices of event splittings. Again, a final χ^2 is computed from the average. Finally the bandwidth giving the smallest averaged χ^2 over the 500 repetitions is chosen as the best bandwidth. This is done for every uncertainty independently.

The algorithm is further constraint by the fact that up and down variations of one uncertainty must produce the same shape. Only their overall scale might differ.

To properly treat uncertainties for which the histograms are dominated by noise, a χ^2 is also computed for a fit with a flat line. If the χ^2 of the flat line is smaller (implying a better fit) the flat line is used to approximate the uncertainty instead of the result of the smoothing. This discards smoothed result, in which the algorithm picked up random features from the noise due to the actual features being vanishingly small.

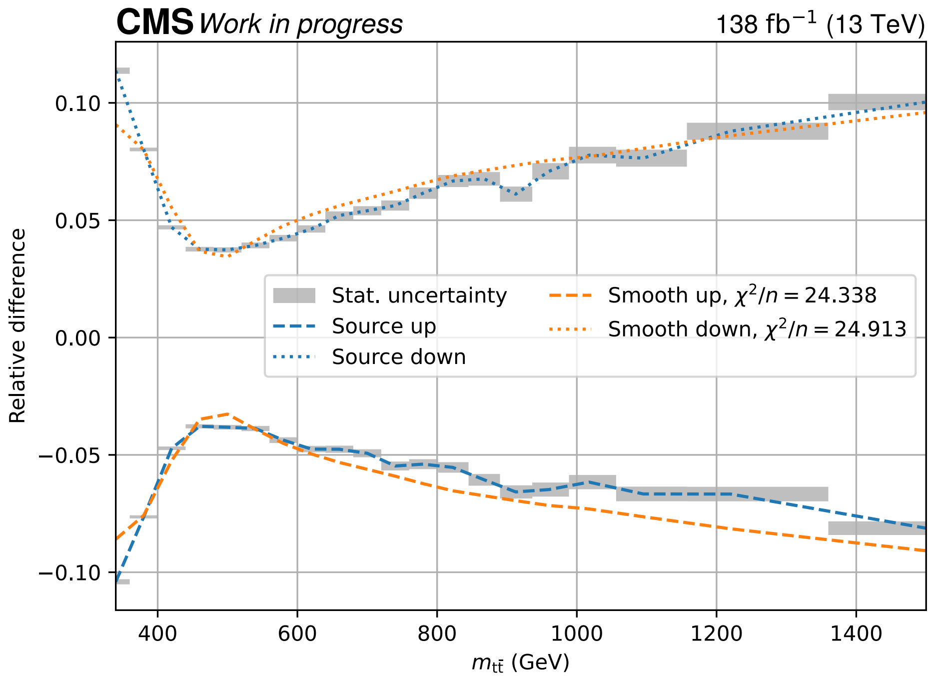
In the case of uncertainties of the background simulation, in order to decrease noise, the smoothing is performed on the sum of all background contributions.

Figure 5.23 show two examples for the smoothing that have been performed only on the $m_{\mathrm{t\bar{t}}}$ axis only. In practice, the smoothing is performed in three dimensions. It can be seen the uncertainty as predicted from simulation ("source" curve) is subject to fluctuations, while the smoothed curve is not. The top quark mass uncertainty in figure 5.23a has much less statistical uncertainty then the JEC uncertainty in figure 5.23b and thus requires much less smoothing. The algorithm is seen to perform well in both cases.

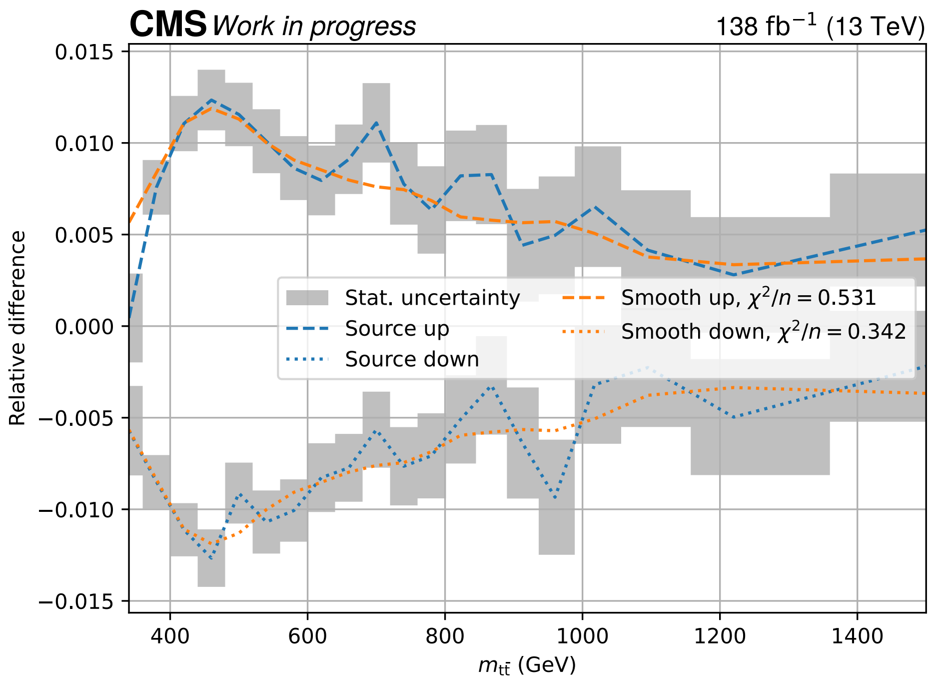
5.14 Results

As explained in the previous sections, in this analysis three-dimensional histograms of $m_{\mathrm{t\bar{t}}}$, c_{hel} and c_{han} are constructed and then used in an asymptotic likelihood fit to obtain upper limits on the coupling $g_{\mathrm{A/H}}$, also referred to as POI. Both Higgs bosons are analyzed independently from each other and without the assumption of BSM models such as MSSM or 2HDM. This way model independent results are obtained that lay the groundwork for further interpretations that are based on specific models. These models may even be in development during the time of the derivation of these results.

In figure 5.24 the histograms for one example mass/width point is shown. It already can be seen that the data from the experiment closely follows the predicted SM event yield. Only by taking a look at the fine differences between prediction and data (the lower panels in the figure), small deviations become visible. The deviations are within the uncertainties, which, as indicated here, are obtained after the fit to the data. In appendix **??** the results with also the low $m_{\mathrm{t\bar{t}}}$ region but evaluated on pseudo data only can be found.

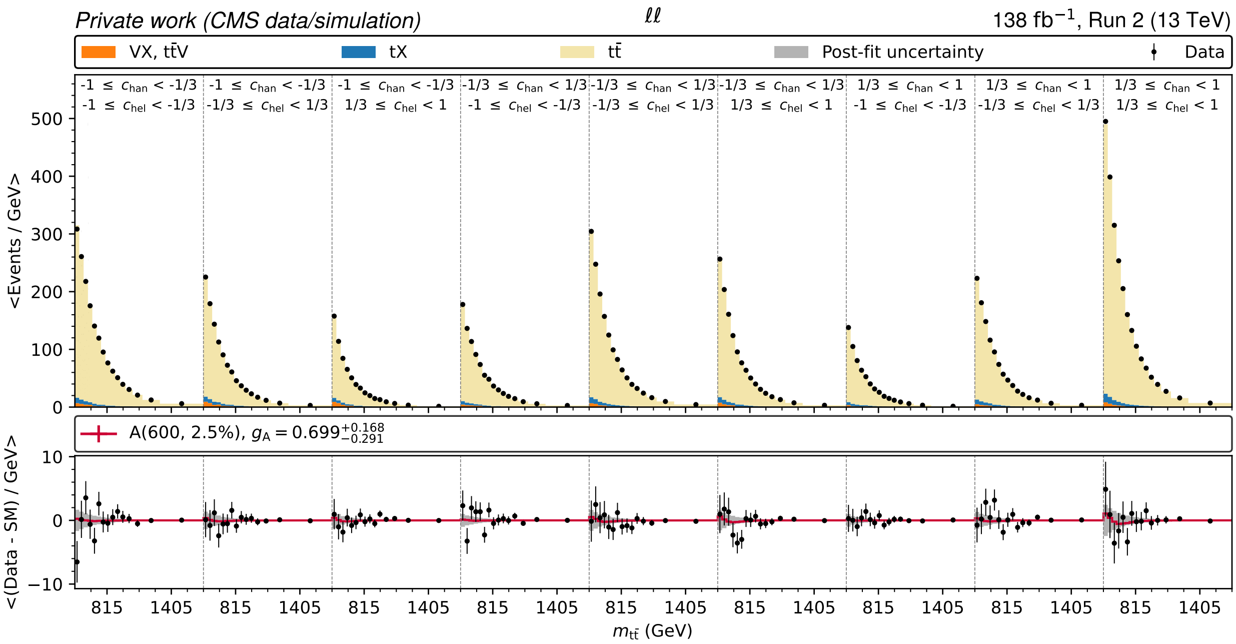


(a)

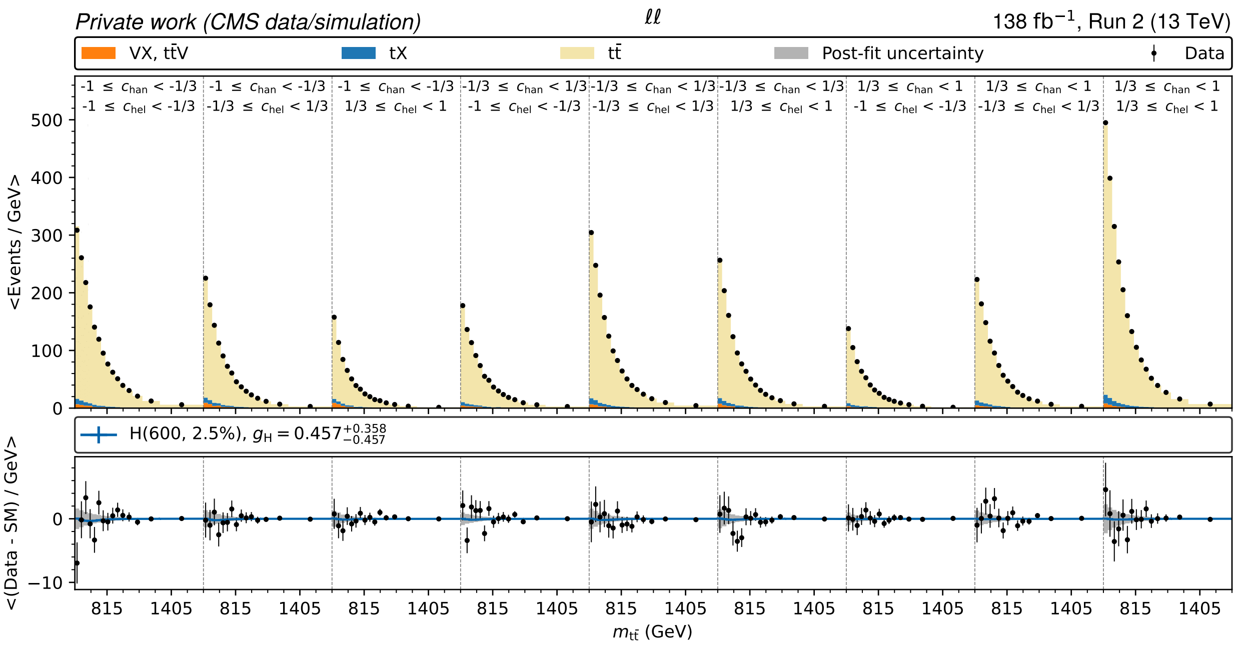


(b)

Figure 5.23: Smoothing performed on two difference sources of uncertainties. The upper figure shows the top mass uncertainty, while the lower shows one component of the jet energy correction. For visual simplicity, only one dimension ($m_{t\bar{t}}$) is used. The relative difference from the uncertainty as obtained from simulation (source, blue) is smoothed to obtain a prediction of the uncertainty that does not include statistical fluctuations (orange). The statistical uncertainty expected from a Poisson distribution spread on the source is shown as gray bands.



(a)



(b)

Figure 5.24: Distribution of the search observables (as described in section 5.10) for the pseudoscalar (upper figure) and scalar Higgs bosons (lower figure) at 600 GeV mass and 2.5 % relative decay width together with the SM background. The three dimensions of the histograms are shown in one dimension by slicing along the c_{hel} and c_{han} axes. The bin heights have been scaled so that they indicate the average counts per GeV. The uncertainties shown as the gray band include statistic uncertainties as well as systematic uncertainties. The indicated coupling $g_{A/H}$ is obtained from fitting the prediction to the (measured) data.

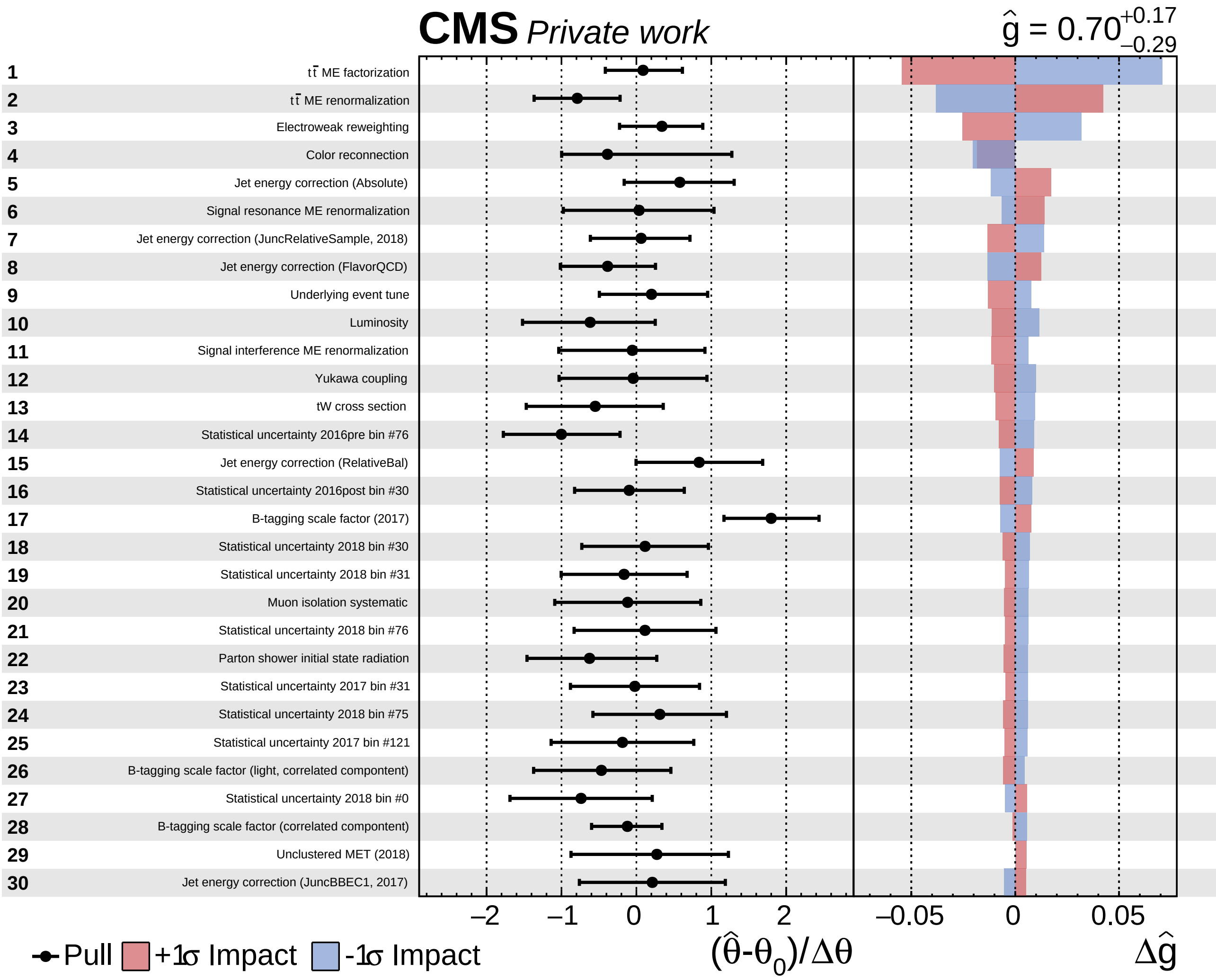


Figure 5.25: Size of individual sources of uncertainties in form of nuisance parameters in the scenario A(600 GeV, 2.5 %) as determined by the fit. The middle panel shows the change to the nuisance parameters after the fit. The right panel shows the change induced to the POI g_A by varying the nuisance parameter within its standard deviation determined from the fit. This is referred to as impact. Only the 30 sources with the largest impacts are shown.

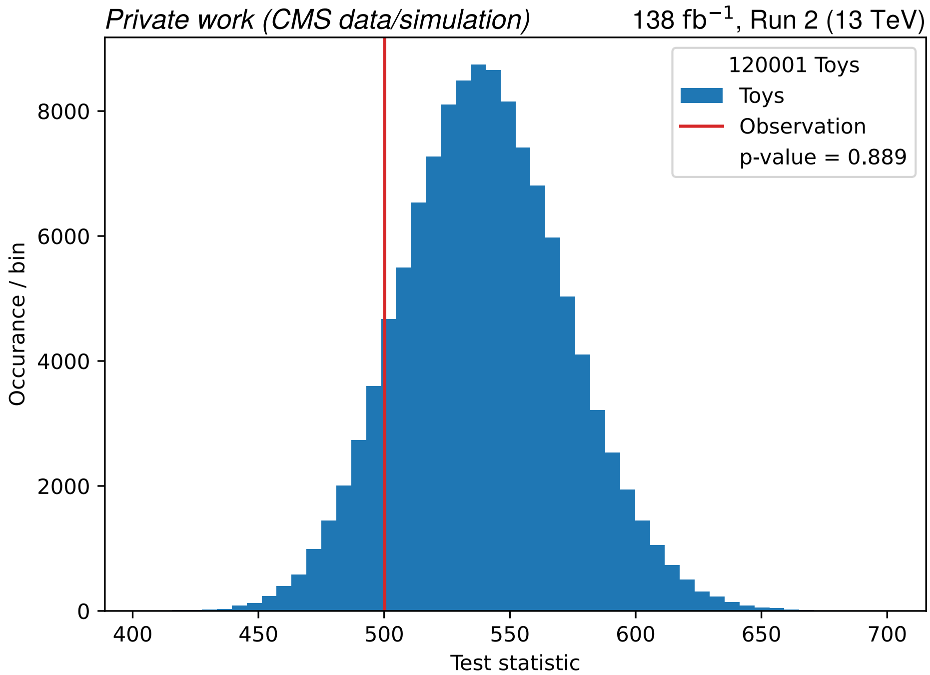


Figure 5.26: Goodness-of-fit test for the background-only hypothesis. The observed value is close to the most likely value, which is located at around 540.

One important type of quantity is represented by the impacts, defined as the size of change a nuisance parameter has on the POI. To study the impacts and the effect of the fit to the uncertainties, the results of the fit are validated before the final evaluation. The fit determines the maximum likelihood value of the nuisance parameter, which in turn determines the overall size of the uncertainty. Additionally an uncertainty of the nuisance parameter around its maximum likelihood value is obtained. This uncertainty is usually given in form of a one-standard-deviation interval. This is given in units such that a deviation of ± 1 indicates the fit determined the variation to be of the same size as the "up" and "down" variations given as input. Finally, the impact of a nuisance parameter is obtained by varying the value of the nuisance parameter within the standard deviation. The results of this study are visualized in figure 5.25 for the 30 leading nuisance parameters for the scenario $\mathrm{A}(600\,\mathrm{GeV}, 2.5\,\%)$. An extensive list can be found in appendix B. The sources of uncertainties are ordered by the size of their corresponding impacts. At the top with the largest impacts one dominantly finds uncertainties arising from the theory. In particular the leading ones are the ME factorization and normalization scales of the $\mathrm{t\bar{t}}$ background as well as the EW reweighting. The large impacts of the former can be explained by the fact that the $\mathrm{t\bar{t}}$ background is by far the largest background. Additionally the scale uncertainties are implemented in this analysis in a way that makes them include a change to the overall cross section of the $\mathrm{t\bar{t}}$ background. For this reason the scale uncertainties also take the role of accounting for the $\mathrm{t\bar{t}}$ cross section uncertainty. When keeping the overall cross section the same, the impact was seen to be much smaller.

To check if the prediction can actually describe the observed data, a goodness-of-fit test is performed. This is done by randomly varying the input histograms within the given uncertainties. One copy of a histogram created this way is referred to as a toy. From each toy the test statistic χ^2, as defined in reference [147] (saturated model) is computed. With a large number of toys, a histogram of test statistics of all toys can predict the distribution of the statistic. The χ^2 can be expected to follow a χ^2-distribution, where the degrees of freedom come from the number of bins. In this analyses there are 135 bins per histogram and in total 4 histograms (for the different eras and years). This amounts to 540 degrees of freedom, which is where the mean of the toy distribution should be located. In the next step the value of test statistic as observed in measured data is computed. An observed value that is reasonable close to the most likely value hints at a good agreement. A single value that described the goodness of fit is the p-value:

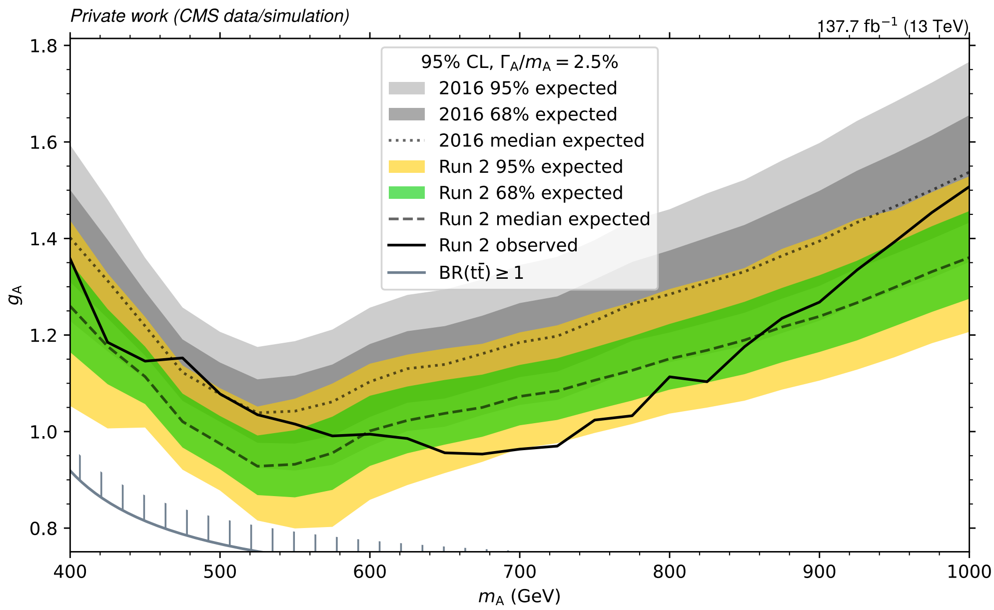
$$p = \int_{\text{Observed}}^{\infty} p(\chi^2)\mathrm{d}\chi^2, \tag{5.23}$$

where $p(\chi^2)$ is predicted from the histogram of all toys. The result using the background-only hypothesis can be seen in figure 5.26. Indeed, the mean value of the distribution is around 540. The observed value in experimental data is reasonable close to the most likely value and the obtained p-value is 0.9 and therefore a reasonable goodness of fit is expected.

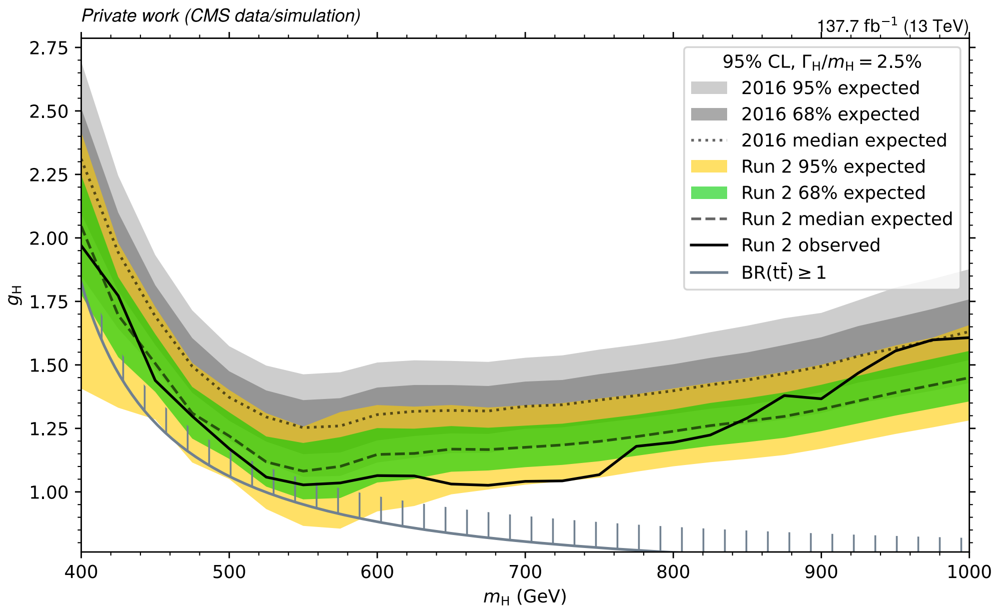
Finally upper limits on $g_{\mathrm{A/H}}$ are produced. This is done in two ways: one for the prediction for the background-only hypothesis using MC events and one for the data observed in the experiment. For the expected limits, the uncertainty on the limits is also derived. If the observation is not close to the prediction and far outside its uncertainty, the observation cannot be explained by the background-only assumption. In other words such an occurrence would hint at the presence of a heavy Higgs boson. Figure 5.27 shows the upper limits derived for the heavy Higgs bosons with a relative decay width of 2.5 %. The expected upper limits of the background-only hypothesis reach a minimum at a Higgs boson mass of around 520 GeV. The reason for this is straightforward: 520 GeV corresponds to the left edge of the lowest $m_{\mathrm{t\bar{t}}}$ bin and thus an increasing fraction of the contribution of Higgs bosons with lower masses will not be visible to the analysis. For higher Higgs boson masses the higher limits come from several sources. For one, the cross sections of the heavy Higgs boson production decreases with increasing masses (see section 5.4). Another reason is the increased decay width in absolute terms. For larger widths more positively weighted events of the signal cancel out with negatively weighted ones, as the resolution of the reconstruction algorithm is limited. Additionally the reconstruction algorithm also performs worse for higher values of $m_{\mathrm{t\bar{t}}}$ with a visibly worse resolution (see section 5.9). For almost all points shown in the figure, the observed limit is within the 95 % uncertainty of the expected limit. There is only one outlier from A with mass of 490 GeV that is however still very close to the 95 % uncertainty, which is why it can not be considered significant. The figure also indicates the expected upper limits for 2016 data only. Overall an improvement from combination of the three years of around two standard deviations can be seen. In conclusion, the search can exclude additional Higgs bosons with a coupling strength modifiers above the range 1.1 to 2.0 (for the scalar boson) and 0.95 to 1.5 (for the pseudoscalar boson).

5.15 Discussion

With the work presented in this chapter, a basis is given for the derivation of a large range of further results. A central building block is provided in the previous section by the model independent interpretation in form of constraints on the coupling. Starting from the obtained results interpretations using the framework of existing models, and possibly future ones, becomes possible. By combining the results with the analysis work that is being performed in other decay channels of the same heavy Higgs signal and together with the addition of a low $m_{\mathrm{t\bar{t}}}$ region in all channels, the sensitivity can further be increased significantly.



(a)



(b)

Figure 5.27: Upper limits of the coupling for a pseudoscalar Higgs boson (upper figure) and a scalar Higgs boson (lower figure) at masses above 400 GeV and a relative decay width of 2.5 %. The expected limit is determined from background simulation, while the observed one is based on measured data. The expected limit has an uncertainty associated to it, which is depicted as a colored band. Besides the expected upper limits for all the three years combined (Run 2), also the 2016-only result is shown. The limits obtained are slightly above the line above which the branching ratio to $t\bar{t}$ has reached one due to the high coupling strength (indicated by the gray hatched line).

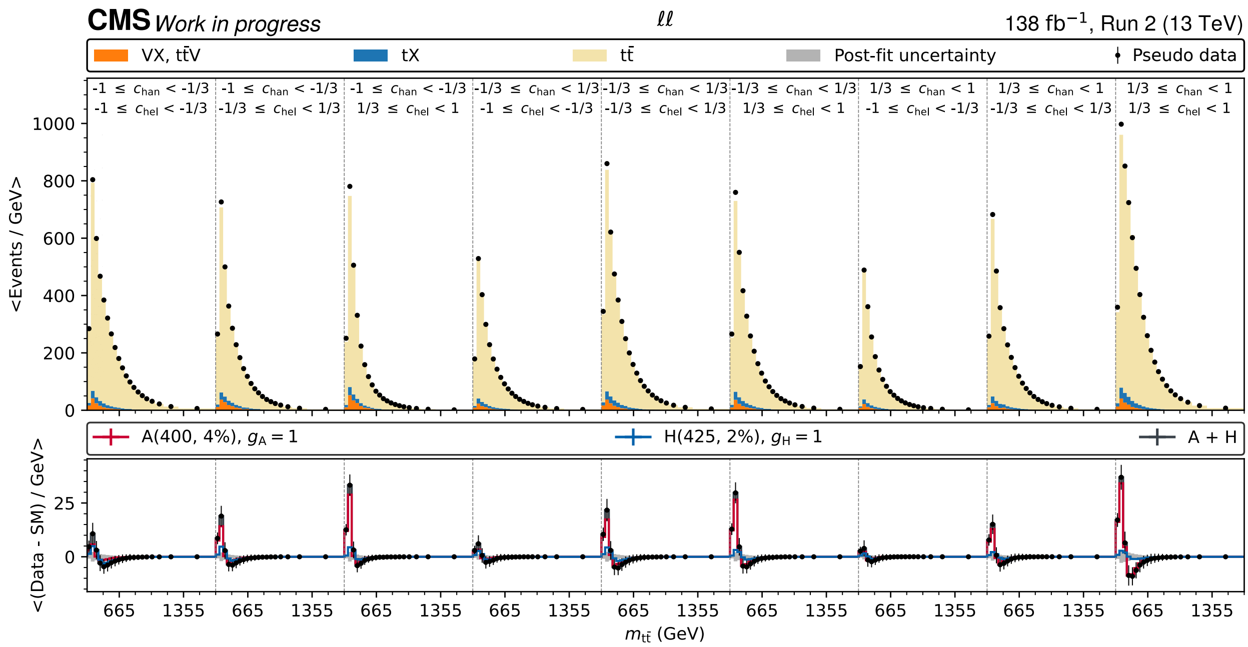


Figure 5.28: Histogram of the three search observables for the SM and signal simulation. Both, pseudoscalar and scalar Higgs bosons, are present, with a coupling modifier of 1. The black dots indicate pseudo data, representing the SM and signal simulation summed up.

Previously a different analysis on data from 2016 only was performed [79], which did not exclude the low $m_{t\bar{t}}$ region. For the dileptonic channel, the analysis excluded couplings above 1.4 for A(750 GeV, 2.5 %) [136], while this analysis, even with the excluded low $m_{t\bar{t}}$ region, excluded values above 1.1 (1.2 for 2016 only). It should be noted that even for a Higgs boson mass of 750 GeV, the majority of the signal events can be expected to appear within the low $m_{t\bar{t}}$ region due to effect of the PDF. The improvement seen indicates the increased sensitivity from the improved analysis strategy as well as the reduced statistical uncertainties.

Nonetheless, the exclusion of the low $m_{t\bar{t}}$ region is clearly limiting the full strength of the results. In figure 5.28 the histogram of the search observables can be seen with 5 additional bins at $m_{t\bar{t}} < 520$ GeV with a width of 40 GeV each to include the region. To keep the region blinded, only simulation and pseudo data is shown. Clearly the Higgs boson manifests as a peak-dip within the measured spectrum. For this reason a publication is planned that includes the region.

The uncertainties with largest impacts on the parameter of interest, $g_{A/H}$ arose from theory. Neither experimental uncertainties nor statistical uncertainties are the most impactful. This is thanks to the large amount of available data from the three years (2016, 2017, and 2018) combined combined with the "ultra legacy" calibrations of the CMS collaboration. For the former reason, the statistical uncertainties appear only at very low impact, while the latter decreased the systematic uncertainties arising from the experiment. The analysis sets a high benchmark in terms of uncertainties for future heavy Higgs boson searches in this channel.

The analysis performed in this chapter was faced with a range of challenges that were solved over the course of its execution. Major challenges were the statistical fluctuations that arose in the histograms made for the prediction of the systematic uncertainties. To extract shapes that were mostly free from statistical fluctuations, the LOWESS algorithm was employed to smooth out and remove noise. However, the application of the smoothing is computationally heavy due to its cross validation procedure ensuring that the best parameters for the algorithm are chosen. To make the analysis pipeline much more lightweight and adaptable, improving the cross validation and thereby reducing the need of computation power would be highly beneficial.

A different approach for reducing the statistical fluctuations is to increase the count of generated MC events. However, to be able to have a meaningful impact on the fluctuations, a drastic increase is required, possibly by a factor of two or more. For this reason the smoothing is the preferred approach to solve this issue. Another approach would be the migration away from dedicated data sets and rerunning the analysis procedure to event weights. Predicting systematic uncertainties using event weights in most cases makes the smoothing unnecessary as the predicted shape of the deviation is already smooth. However, even event weights are not able to provide smooth predictions in all cases. Furthermore, generating weights for the uncertainties, that are currently predicted by dedicated data sets, is an ongoing effort if possible at all.

A big step forward for this analysis was the introduction of reweighting to different masses and decay widths of the heavy Higgs signal. The reweighting facilitated the prediction of results at masses/widths for which no dedicated simulation had been done. It also decreased the statistical uncertainty coming from the signal prediction. This was because for one mass/width point not only the data set generated for the specific point could be used, but also all the other data sets. In hindsight one could have generated dedicated data sets only one or two width/mass points that are then reweighted to all the desired points.

Another addition to the signal prediction would be the inclusion of further higher order effects. The signal used in this analysis has been generated at LO precision. An NLO prediction of the heavy Higgs signal could be beneficial to the overall quality [148]. However, this comes at the cost of a significant addition of complexity of the analysis strategy. In particular one would have to reconsider the resosonance-interference split that is done in this analysis, as the additional terms in the ME computation make it difficult to differentiate between the two. Furthermore the assumption of no CP violation, which has been part of this analysis, could also possibly be lifted in the future. As a result one would be searching for superpositions of Higgs bosons instead of the pseudoscalar and scalar states.

A central limitation of the analysis is the resolution of the top four-momenta reconstruction. For the important observable $m_{t\bar{t}}$ the reconstruction can only deliver a resolution of around 30 %. This "smears" the peak-dip structure of the signal seen in the $m_{t\bar{t}}$ distribution so that a large fraction of the negatively weighted events of the dip, cancel out with the positively weighted signal events from the peak. So instead of creating the structure, often only a flattened out shape on top of the $m_{t\bar{t}}$ spectrum can be seen. Therefore the analysis is very depended on the $m_{t\bar{t}}$ resolution. At the same time the reconstruction is comparatively computationally heavy due to its feature of input variation. The variation tries to solve another shortcoming, which is the fairly common failure of the algorithm due to detector effects. A highly promising alternative approach for the reconstruction algorithm is the usage of deep neural networks. This idea is explored in the next chapter.

Chapter 6

Machine learning approach for top pair reconstruction

To obtain the most precise result profiting from all information of the full event topology in the search described in this theses, a new algorithm is developed to reconstruct the top quark system with a novel method utilizing deep learning. In the following, a neural network will be built from scratch that aims to outperform the classic approach. A major challenge is posed by the requirement for generality of the algorithm: The network should not only provide a good prediction for a sample of SM data, but also for data distributed according to another model's assumptions. One example would be the heavy Higgs signal of the search, which, depending on the Higgs mass, can even be located at extremely high scales, where very little to no SM events are located. Optimally the network behaves in such a way that independent of the true distribution the data follows, the network is able to reconstruct the top quarks at a high precision and accuracy. To achieve this the data used to train the network needs to be well selected. The network introduced in the following was initially inspired by attribute learning [149] and has been simplified down to its here described form without loss in performance.

6.1 Classic algorithms and developments in the field

A central building block of analyses featuring the production of a top pair with dileptonic decay is the reconstruction of said quarks. The analysis that is described in the previous chapter uses an algorithm that is based on the analytic solution of the equations of the decay. The algorithm is characterized in detail in section 5.9 and is based on the solutions given by Sonnenschein [131]. The key features of the algorithm that focused upon in this chapter are:

- It yields up to four complex solutions. If all of them have a numerically significant imaginary part, the algorithm is treated as being unable to reconstruct the event.
- MET is assumed to entirely be the product of the neutrinos. If this is not the case, for example due to detector resolution or due to additional BSM particles in the process, the chance to obtain complex solutions increases. Similarly, radiation of additional particles, even if they are detected, cannot be easily included.
- The top quark mass is fixed and cannot be determined by the algorithm.

Several other possibilities exist as well. Recent use have found the algorithm described in reference [150] by Betchart et al. and the "loose" reconstruction algorithm from reference [151]. Different to the Sonnenschein solution, Betchart's algorithm is utilizing a geometric approach, that allows the construction of a real solution even in the case where Sonnenschein's solutions would all be complex. The algorithm can be applied to the dileptonic decay as well as the semileptonic one. Compared to Sonnenschein's solutions in the case of the dileptonic decay, Betchart's algorithm produces similar results when real solutions are available. The loose reconstruction algorithm on the other hand meets different demands: Instead of giving a solution for the top quark and antiquark, it only gives a solution for the total $\mathrm{t\bar{t}}$ system (one four-momentum). This algorithm works without the prior fixation of the top quark mass. However, it cannot be used when one needs observables based on individual top four-momentum, like c_{han} and c_{hel} of the analysis of the previous chapter.

In addition to having to reconstruct the top quarks from bottom quarks and leptons, the analysis also faces the problem of having to define the bottom quarks. While the oppositely charged leptons generally are easy to identify in the CMS experiment, charge of the quarks is hidden from the detector. In the implementation of the reconstruction in section 5.9, the problem was solved by finding the assignment that has the highest probability given the one-dimensional quantity m_{lb}. In a test in which the analysis selection (section 5.6) is applied to dileptonically decaying $\mathrm{t\bar{t}}$ events and additionally only events are kept, in which the jets of both bottom quarks are present. This solution succeeds in only 69 % of the events. Small improvements can be added to the solution. For example if the analytic reconstruction fails, one can retry with a different combination of jets. However, the efficiency of identifying the bottom quark can only marginally be increased this way.

Deep learning has firmly established its position in particle physics experiments. It has also been exploited in the context of reconstructing hadronically or semileptonically decaying top quarks. For these decays a driving challenge is the correct assignment of decay products due to the abundance of jets coming from the final states [152–154]. In reference [155] single observables were reconstructed in the dileptonic decay channel using a deep neural network. This chapter is going to take a closer look at the possibilities within the dileptonic decay channel, where, in contrast to the other channels, a major challenge is arising from the fact two neutral leptons, the neutrinos, are not measured by the detector. The full four momenta of the top quark and antiquark will be reconstructed, enabling the derivation of any other top dependent observable.

6.2 Concepts for deep neural networks

In a way, neural networks can be understood with similar principles in mind as multi-dimensional histograms, which is one of the most common tools in particle physics. A neural network represents some function $f(x)$, which has been constructed so that it ties to minimize a differentiable function L, called loss function, given a probability distribution $P(x)$ for the inputs x and a desired output $t(x)$. The probability distribution is modeled according to a discrete sample, called the training set and in the ideal case f can be constructed so that $f(x) = t(x)$. One approach to construct f without neural networks would be to create histograms with axes for x and y, fill it according to the training set and compute the mean for every slice in x.

The challenge starts when the desired output also depends on variables whose values are unknown, meaning $t = t(x, y)$, where y cannot be determined. One example for this case is the problem of top quark reconstruction, where the neutrino momentum is unknown while the top quark momentum depends on it. Reference [156] looks at this problem of unknown values in conjunction with the mean squared error, a very common loss function. It is shown that the best possible function, the function that minimizes the mean squared error, is

$$f(x) = \frac{1}{P(x)} \int P(x,y)t(x,y)\mathrm{d}y, \tag{6.1}$$

which is simply the mean of t over y and a normalization $1/P(x)$. This prediction of f coincides with the approximation obtained from averaging histograms, given fine enough bins.

Thus the choice of a neural network instead of a histogram is a very natural one in particle physics, where probabilities and likelihood maximization is often executed using histograms. A histogram example in conjunction with the top quark reconstruction is the m_{lb} histogram used for assigning the jets to the bottom quarks, as described in section 5.9. The method only uses a single observable m_{lb} and therefore is limited in the amount of information it can use. Other parts of the algorithm also heavily rely on histograms as for example the variation of the inputs, due to detector uncertainties. With a neural network these problems are not present, as it can incorporate in principle any number of variables and always produces a smooth dependence.

Moreover, the strength of neural networks arises from their ability to approximate a vast range of function. In particular, it has been shown that given continuous function on a compact support the neural network will be able to uniformly approximate it to arbitrary accuracy [157]. The statement and other similar ones are known as uniform approximation theorems.

The main ingredient used in the neural networks presented in this thesis are dense layers. Each dense layer is one differentiable function. To make a network, a composition of layers is created. In the case the network is a composition of a large number of layers, one refers to it as deep network. One dense layer $\vec{d}$ is defined by

$$\vec{d}(\vec{x}, W, \vec{b}) = \vec{a}(W\vec{x} + \vec{b}) \tag{6.2}$$

where $\vec{x}$ is the input to the layer, W is the weight matrix, $\vec{b}$ is the bias vector and $\vec{a}$ is the activation function. The activation function is a non-linear function, for which a wide range of established choices exist. A common choice due to its high performance in deep neural networks [158] is the rectified linear unit (ReLU), which is defined by

$$\vec{a}_{\mathrm{ReLU}}(W\vec{x} + \vec{b}) = \max(0, W\vec{x} + \vec{b}), \tag{6.3}$$

where max is evaluated element-wise. The weight matrix has the dimensionality of $m \times n$, where n is the dimension of $\vec{x}$ and m is the desired output dimension. The values of this matrix and the ones of the m dimensional vector $\vec{b}$ are parameters that need to be determined by an algorithm. This algorithm tries to find the parameters that minimize the loss function by computing the derivative of the entire network. This process is called training. The training is a very similar task to what often appears in physics in form of function fitting and minimization. This thesis uses the ADAM optimizer [159] for training. A risk induced by the statistical uncertainty due to the limited size of the training set is overfitting. If overfitting occurs, the network's output resembles some or all of the statistical fluctuations of the training set, while underperforming on a sample independent from the training set. The phenomenon can be imagined as if the network is learning the training set by heart. The overfitting is mitigated by adding dropout layers between all dense layers [160]. The dropout layers are described by a simpler but similar form as the dense layer:

$$\vec{d}_{\text{dropout}}(\vec{x}) = W_\theta \vec{x}$$
$$W_{\theta,\text{ij}} = \begin{cases} 1 & i = j \text{ and with a probability } 1-\theta \\ 0 & \text{else} \end{cases}$$

The matrix W_θ is an identity matrix, where some of the 1s are randomly replaced by 0s. The probability for a 0 to appear is θ. The randomization is done at every step during the training. This randomly disables parts of the network, which makes its final output the average over several parts, evening out possible overfittings.

Before supplying input values to a network, the values are usually normalized. A straightforward approach to normalize the inputs is to compute the mean and the standard deviation of each input and shift and scale accordingly. The result is that all inputs the network sees are centered around 0 and, for a majority, are within the interval $[-1, 1]$. While the normalization in principle does not change the results of the network, it makes the training much more efficient and avoids possible numerical issues. This is because very differently sized inputs have a similar effect as differently weighted inputs, at least at the beginning of the training, possibly making the network pay more attention to uninteresting inputs. Similarly inputs that are biased towards one direction make it difficult for the training algorithm to find the optimum [161].

The parameters defined in this section, such as the layer width m, the dropout probability θ or simply the number of layers in the network are to be decided before the training and are called hyperparameters. This is in contrast to the network's trainable parameters, which include the weight matrix W and the bias $\vec{b}$. In this thesis the hyperparameters were chosen based on the fasted training and lowest final loss.

The network shown in this thesis is implemented within Keras in Tensorflow [162].

6.3 A network for dileptonically decaying top reconstruction

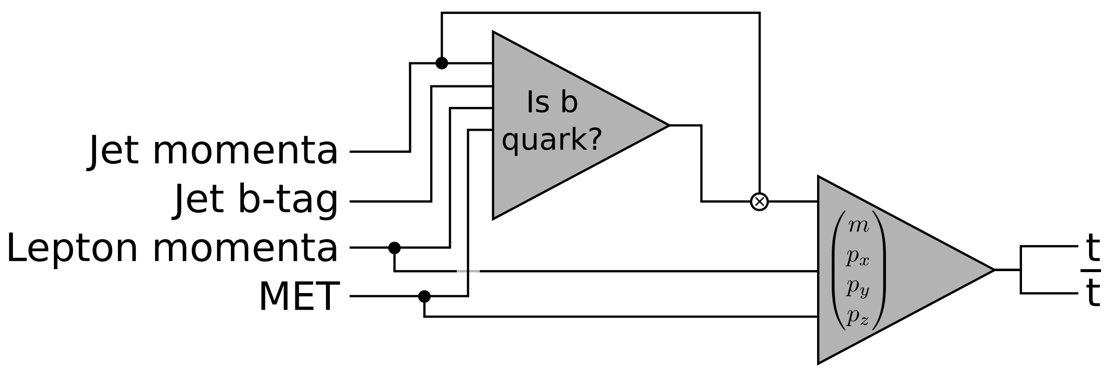


Figure 6.1: Scheme of the neural network used for reconstructing the top quark momenta. The network is made from two subnetworks.

The task of the neural network is to reconstruct the two four-momenta of the top quarks from top quark pair production given the basic objects of jets, leptons and MET. This is realized in two steps, each using their own separate neural network. The first step is to identify the jet of the bottom quark and the jet of the bottom antiquark. This information is then used to pipe the observables of the identified jets into the input for the bottom quark and antiquark of the network of the second step. The second step is responsible for the computation of the desired top momenta. This scheme is visualized in figure 6.1. The network produces momenta for every event and, different to the analytic method, does not fail for a fraction of the events.

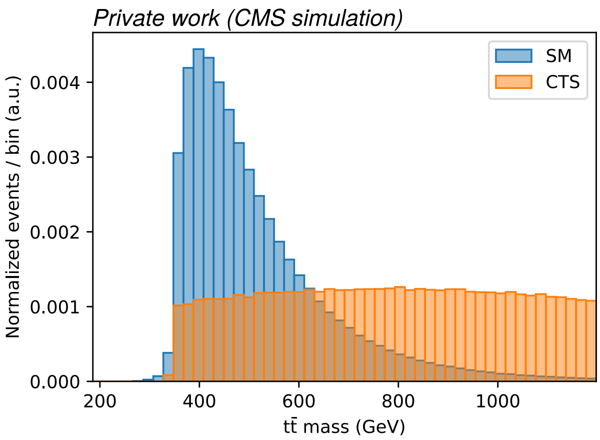
Instead of trying to reconstruct the top quarks independently (for example in two independent networks), the network treats the two top quarks as one object. This means an eight-dimensional vector is produced for every event. This choice has been made because the detector response to each of the two decays is inherently correlated with each other. Especially the MET, which is one important input object, has contributions from both decays.

A central question in building the network is the choice of the training set that should be used. As described in the previous section, the network is likely to produce a result that reflects the mean over the degrees of freedom that have not been constrained by the inputs. In this case, these come from the two neutrino four-momenta, which are only constrained by the two-dimensional MET. Training the neural network on a signal, to then use the network to search for said signal could potentially bias the search. For this reason, in this thesis the network is trained on two different data sets. One data set is a QCD NLO SM $\mathrm{t\bar{t}}$ simulation with dileptonic decays, while the other one has been specifically crafted to be devoid of an underlying process. This second data set is supposed to only contain information about the dileptonic decay, which the network is supposed to learn.

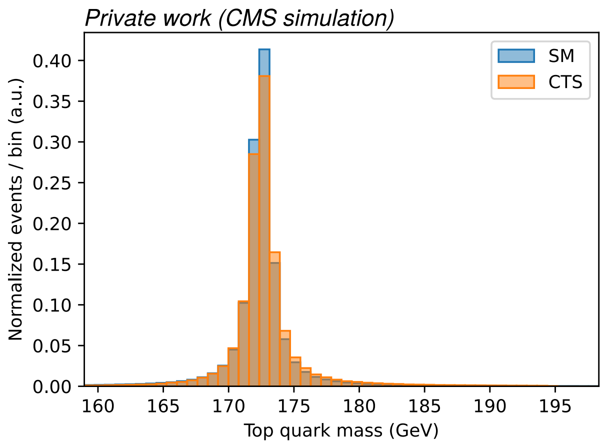
The crafted data set has been created by first producing pseudo $\mathrm{t\bar{t}}$ events. For this, the momentum of the $\mathrm{t\bar{t}}$ system is set to $p_x = p_y = 0$, while p_z and the mass m are chosen randomly. The $\mathrm{t\bar{t}}$ p_z is sampled from a Gaussian distribution centered at 0 and with a standard deviation of 500 GeV. The p_z scale of 500 GeV is chosen based on the width of the p_z distribution seen in SM $\mathrm{t\bar{t}}$ simulation. m is sampled from a uniform distribution starting at $2 \cdot 173$ GeV (twice the top quark mass) and ending at 1500 GeV. The choice for the mass distribution has been made so that the entire range of what is possible to be produced by the heavy Higgs boson signal in the analysis presented in this thesis is covered. The masses of the top quarks are set to 172.5 GeV. The direction of the top quark in the $\mathrm{t\bar{t}}$ frame is chosen randomly according to a three dimensional Gaussian distribution in x, y, z. This fully constrains the momenta of both top quarks. The events are assigned to either a $\mathrm{q\bar{q}}$ or a gg initial state with the same probability as the SM LO process. The momenta of the initial state particles are computed from the momenta of the top quarks. Next, on these so produced events, MadSpin is run to simulate the dileptonic $\mathrm{t\bar{t}}$ decay. MadSpin also adds decay width to the distribution of the top masses. The particles at this state are shown in figure 6.2 and a full overview of all input and target variables can be found in appendix C. The mass of the $\mathrm{t\bar{t}}$ system shows the flat spread over a large range, while the t mass shows the same peak as in the SM. p_{T} distributions of decay products are generally spread towards higher values. The p_{T} distribution of the $\mathrm{t\bar{t}}$ system of the crafted data is restricted to 0, which agrees with LO SM $\mathrm{t\bar{t}}$ simulation. The following steps of the simulation are the same as most of the CMS simulations, including parton shower and detector simulation.

An event selection is applied to the data sets that follows the event selection from section 5.6, but is much looser. In particular, no requirements are done on the lepton p_{T}, jet p_{T}, m_{ll} and MET p_{T}. Under this looser selection, the network is able to learn about the parameter space that is outside the particular analysis and thus has more adaptability.

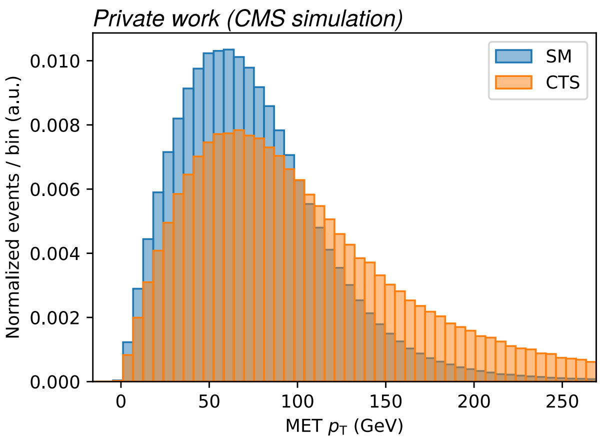
In preparation for the training and evaluation of the network, the data sets are split into two independent parts. The first part is used exclusively for training, while the other part is only used for evaluation and is called validation set. In the extreme case the network was merely memorizing



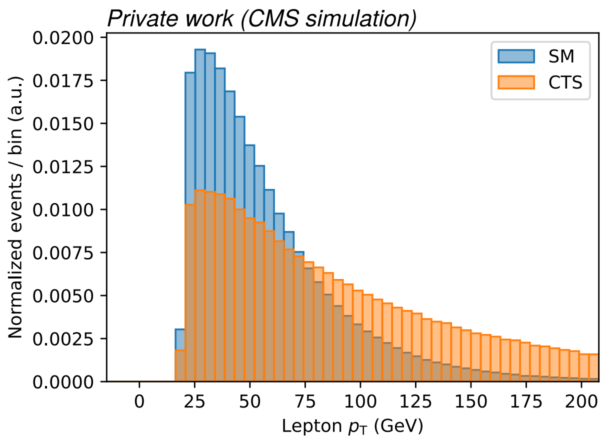
(a)



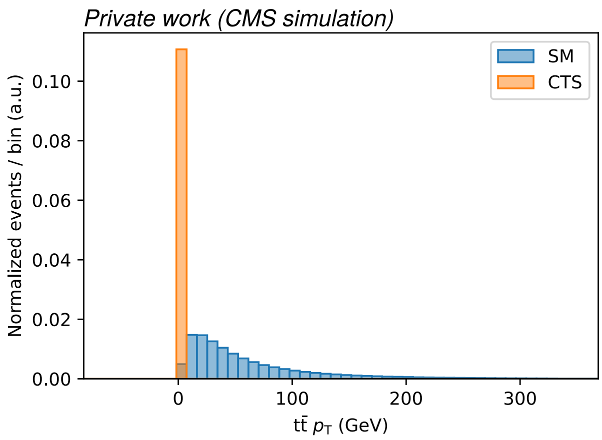
(b)



(c)



(d)



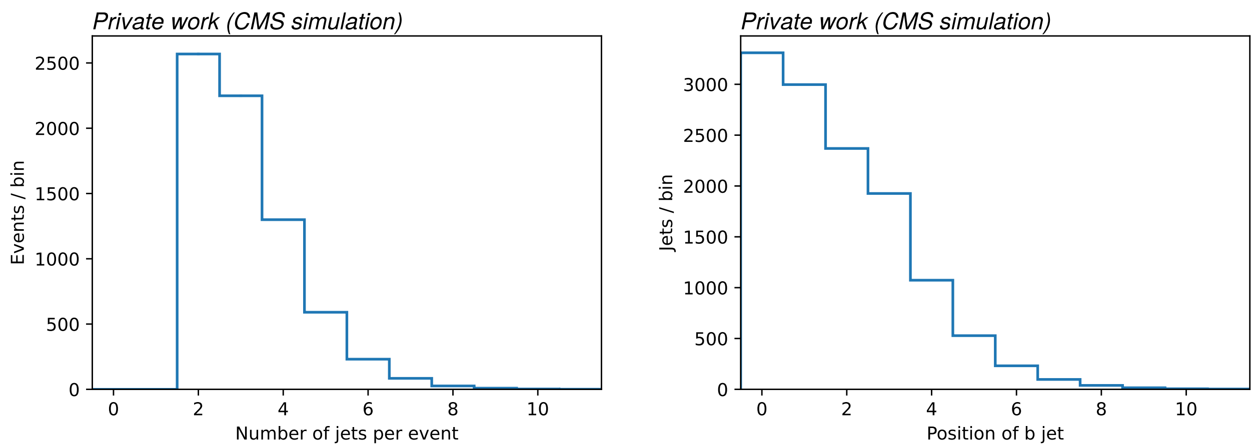
(e)

Figure 6.2: Distributions of selected observables in the SM tt̄ and the crafted tt̄ data set. The event counts have been scaled so that their sum is 1 for better visibility. The particles shown here are before parton shower and detector simulation.

its training set instead of learning how to predict the values, an evaluation using the validation set would immediately indicate the bad performance. The input and targets are normalized as described in the previous section.

6.4 Jet classification

The first step of the reconstruction using the network is the assignment of one jet to the bottom quark and another jet to the bottom antiquark. One of the challenges faced for this problem is the varying number of jets in each event. Some events only contain two jets while in some extreme cases there can even be 9 jets in a single event (see figure 6.3a). The jets are always kept sorted according to



(a) (b)

Figure 6.3: Number of jets per event for part of the SM $t\bar{t}$ training set and the position of the b jet within the jet collection. The jet collection is sorted according to the jet p_T and is truncated so that it includes not more than 7 jets. It can be seen the truncation only affects a very small portion of jets.

their p_T, going from highest to lowest. While low p_T jets may sometimes contain a small fraction of the momentum of the original $t\bar{t}$, they are by far not as significant to the overall event and can be neglected. Most of the b jets are within the leading p_T jets, many of them are the highest p_T ones even (see figure 6.3b). For this reason the challenge of varying jets is narrowed down by truncating the jet collection, so that no event has more than 7 jets.

The jets are being fed into the network with two dimensions, the first one enumerating the jets (timesteps) and the second one enumerating different jet observables (features). Here it becomes important to recall any event can also have less than 7 jets, making the array holding the jets irregular. One can imagine the array to be jagged. To obtain a regular array that is compatible with the matrix operations performed in the network, the array is padded with 0. This means that if the length along the first dimension is less than 7, 0s are appended until a length of 7 is reached. The padding is accounted for in the network by a masking layer. The masking layer adds the information that padded elements shall be skipped in subsequent layers. The subsequent layers are evaluated on each jet independently and the masking layer has the effect that the padded out jet positions will result in 0 values only.

Besides the jet inputs that go through the masking layer, the lepton, the antilepton and the MET of the event are used as input. All of these are provided in form of four-vectors in two coordinate

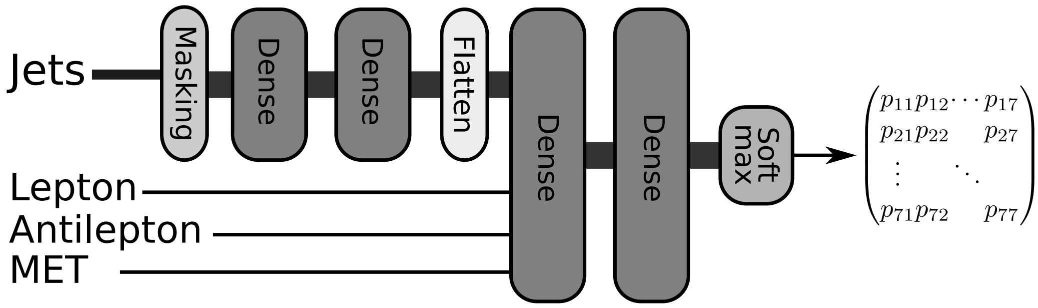


Figure 6.4: Scheme of the network for b jet classification. Due to their varying length the jets are fed into the network via a masking layer after padding with 0.

systems, one using the coordinates (p_x, p_y, p_z, E) and another one using $(p_\mathrm{T}, \varphi, \eta, m)$. In the case of MET, only the first two are provided, as MET is a two-dimensional quantity. While providing four-momenta in two coordinate systems induces redundancy to the inputs, it helps the network to grasp features faster. Both coordinate systems bring forth different relationships between objects, which is why both systems are in use in physics. However, conversion between them is not at all trivial and while a neural networks is in principle able to learn the conversion, it is much easier to provide it right away. For the jets additionally the discriminator value of the jet bottom flavor (b tag) is used as input. Similar as in the analysis of the previous chapter, the discrimination was done using the DeepJet algorithm, which in turn is a neural-network-based approach [56].

The scheme of the network part that is responsible for the b jet classification is shown in figure 6.4. It is assembled from four dense layers with an output dimensionality of 200 each and using the ReLU activation function. As written in section 6.2, each dense layer is equipped with dropout. The dropout probability θ is set to 0.25. The first two dense layers only take care of the jet information, while keeping the first dimension (enumerating the jets) intact. Afterwards the output matrix is flattened into a vector with $7 \times 200 = 1400$ entries. The last two dense layers combine information from the jet layers, the lepton inputs and the MET inputs.

The output of the network is a 49 dimensional vector produced by a final fifth dense layer. This layer uses the softmax activation function defined for the ith output dimension by [157]

$$a_i(\vec{x}) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \tag{6.4}$$

with the input vector $\vec{x}$. The activation function makes sure the output is properly normalized so that it can be interpreted as probability. This kind of output is required because next the output vector is reinterpreted as a 7×7 matrix, one row and column for every jet, where the matrix element p_{ij} represents the probability for the ith jet to be the bottom quark and for the jth jet to be the bottom antiquark.

The training of the classification network is done independently from the whole top reconstruction network. The training target (truth) are 7×7 matrices containing a single 1, marking the correct quark/antiquark jet position, and otherwise only 0s. The truth is generated from the parton flavor of the jet that is provided in NanoAOD, which is based on parton matching. The loss function used in the training is the cross-entropy

$$\sum_n \left(t_n \ln f_n(\vec{x}) + (1 - t_n) \ln\left(1 - f_n(\vec{x})\right)\right) \tag{6.5}$$

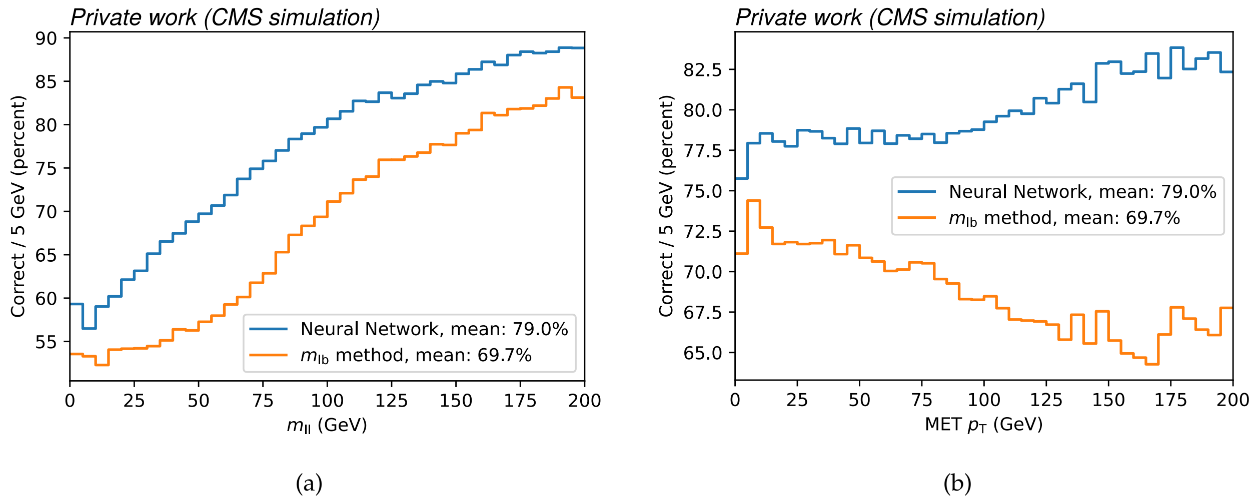


Figure 6.5: Performance of the jet classification via the neural network compared to the m_{lb} method explained in section 5.9. The vertical axis shows the fraction of correct assignments to b and $\bar{\mathrm{b}}$ in percent. For the figure, the events from the validation set are used together with the requirement that both jets are present.

where $f_n(\vec{x})$ is the nth element of the network output and t_n is the nth element of the target vector. For the loss function, target and output matrices are treated as flattened vectors. From the training set only events that contain both jets are used.

The classification network clearly outperforms the classic m_{lb} method, which was explained in section 5.9. This can be seen in figure 6.5. Here, both approaches were evaluated on the events of the validation set that contain both target jets. While the m_{lb} method only associates the jets correctly in 69.7 % of the events, the neural network is able to reach a correctness of 79 %. For both approaches low energetic events are difficult, as can be seen by the low performance at low m_{ll} values. However, even here the neural network clearly outperforms. The m_{lb} method is faced with increased difficulty at higher MET p_{T} values. On the other hand the neural network is even better in this high MET p_{T} region, underlining superiority. In conclusion, the reconstruction can now profit from a much better jet assignment compared to the m_{lb} method.

6.5 Momentum reconstruction

The next part of the network for reconstruction is using inputs similar to the inputs used by the first half. It uses the MET, the lepton and the antilepton, all of them given as four-vectors in two coordinate systems. The jets are not directly used as input. However, this part of the network additionally has inputs for the bottom quark and the bottom antiquark. Into these the two jets that have been selected by the first part of the network are being provided. This is realized by a custom network layer that takes the jets and the classification output and forwards the quark-classified jet as its first output and the antiquark-classified jet as its second. The jets are given also in terms of four-momenta.

The scheme of this part of the network is depicted in figure 6.6. The inputs besides the b jets are handled by two separate dense layers until also the b jets are taken into the computation with two further dense layers. The layers are made especially wide with an output dimension of 800. Similar to the dense layers of the classification network, the layers are using dropout with a probability of 0.25 and ReLU activation functions.

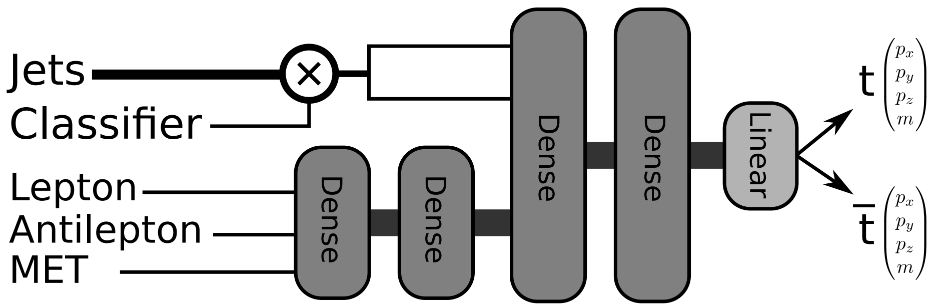


Figure 6.6: Scheme of the network for momentum reconstruction. Jets and classifier output are combined to give one jet for the bottom quark and one jet for the bottom antiquark.

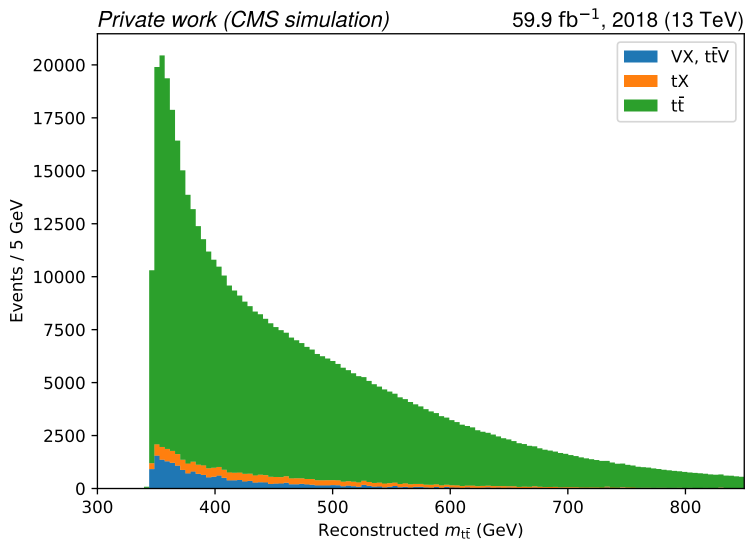


Figure 6.7: Distribution of the invariant $t\bar{t}$ mass as computed from the reconstructed four-momenta from the neural network. The figure includes not only the SM $t\bar{t}$ simulation used in the analysis of the previous chapter but also its other backgrounds scaled to the integrated 2018 CMS luminosity.

The output is generated by a final dense layer with a "linear" activation described by $a(\vec{x}) = \vec{x}$. The output is eight dimensional and models the four-momenta of the top quark and the top antiquark. An important property of the output is that these four-momenta are provided in form of (p_x, p_y, p_z, m). This way as long as $m > 0$ the network is by definition unable to generate unphysical results, as the components of three momentum in Cartesian coordinates are allowed to have any real value. The training is done on the reconstruction part of the network only, without changing anything of the classification network. The data set used for training is the same as used for the classification network. The target vectors are taken from the top quarks at ME level before parton shower. The employed loss function for the training is the mean squared error given by (omitting taking the mean):

$$\sum_n (f_n(\vec{x}) - t_n)^2 \tag{6.6}$$

where $f_n(\vec{x})$ is the nth element of the network output and t_n is the nth element of the target vector.

From the output in form of four-vectors all the top quark related quantities can be derived. Figure 6.7 shows the distribution of $m_{t\bar{t}}$ of the full background prediction of the analysis presented in the previous chapter. As expected from the SM the distribution peaks shortly before 400 GeV and the smaller backgrounds show a very similar behavior than the main $t\bar{t}$ background.

6.6 Performance and comparison

In the following, distributions of the reconstructed quantities will be shown. The figures are created using data independent from the validation sets. One important factor to note is that just by comparing shapes of distributions, it is not possible to tell which method of reconstruction performs better. As an example, one could take any observable whose histogram resembles the shape of a Gaussian distribution. Without making use of any information inside the event, one can now almost perfectly have a prediction recreate the shape by simply drawing from an independent Gaussian distribution. However, creating a much better prediction, that actually uses the information available, might distort the distribution seen from a histogram, so that it does not resemble the Gaussian distribution. For example, simply using the mean of the distribution as prediction for all events, gives a smaller error than the independent Gaussian distribution. For this reason, the figures also list the mean squared error (MSE). A smaller MSE indicates a better agreement with the truth. Furthermore, distortions in form of a general shift or bias have very little effect to analyses as presented in the previous chapter. In the analysis, histograms for different hypothesis are created. Only a change in the relative positions between the histograms can have an effect, a bias affecting both histograms similarly will not be visible to the final results.

Figure 6.8 compares the distributions of several top quark observables reconstructed by the analytic method of section 5.9 and by the neural network. Overall the reconstructed shapes are reasonably close to the true ones. Most importantly all the MSEs of the neural network are smaller than the ones of the analytic method. This indicates a better prediction from the neural network for all of the displayed observables. It can also be seen that the top quark mass distribution reconstructed by the neural network shows a width, which is close to the one of the SM. This is not at all the case for the analytic method, where the top quark mass is a fixed parameter. The reconstructed p_{T} and η distributions show a slightly deviating shape, while the MSE still is much better than the one of the analytic method. A slight bias towards lower p_{T} values is however visible. On the other hand, as explained before, the bias has only a minor role, if at all.

The MSE is the sum of the bias squared and the variance [163]. In the following bias and standard deviation (square root of the variance) will be shown. For some observables it makes more sense to look at relative errors instead of absolute errors. In this case relative bias and resolution are given. These quantities are defined by:

$$\text{bias} = \frac{1}{N}\sum_i^N (f(x_i) - t_i) \quad \text{std. dev.} = \sqrt{\frac{1}{N}\sum_i^N (f(x_i) - t_i)^2 - \text{bias}^2} \tag{6.7}$$

$$\text{relative bias} = \frac{1}{N}\sum_i^N \left(\frac{f(x_i) - t_i}{t_i}\right) \quad \text{resolution} = \sqrt{\frac{1}{N}\sum_i^N \left(\frac{f(x_i) - t_i}{t_i}\right)^2 - \text{relative bias}^2} \tag{6.8}$$

Here, $f(x)$ is the prediction and t represents the true value. For all of these variables, a value closer to 0 indicates a better prediction.

Two different trainings of the same reconstruction network have been tested: One that was trained on the SM $\mathrm{t\bar{t}}$ data (SM training set, SMTS), and one that was trained on the crafted data set (crafted training set, CTS). The results are shown in figure 6.9 for SM $\mathrm{t\bar{t}}$ data and in figure 6.10 for a pseudoscalar Higgs boson of the especially high mass of 1000 GeV. The mass has been chosen as it creates an increase of events at high values of $m_{\mathrm{t\bar{t}}}$, where there is very little data for the SM case.

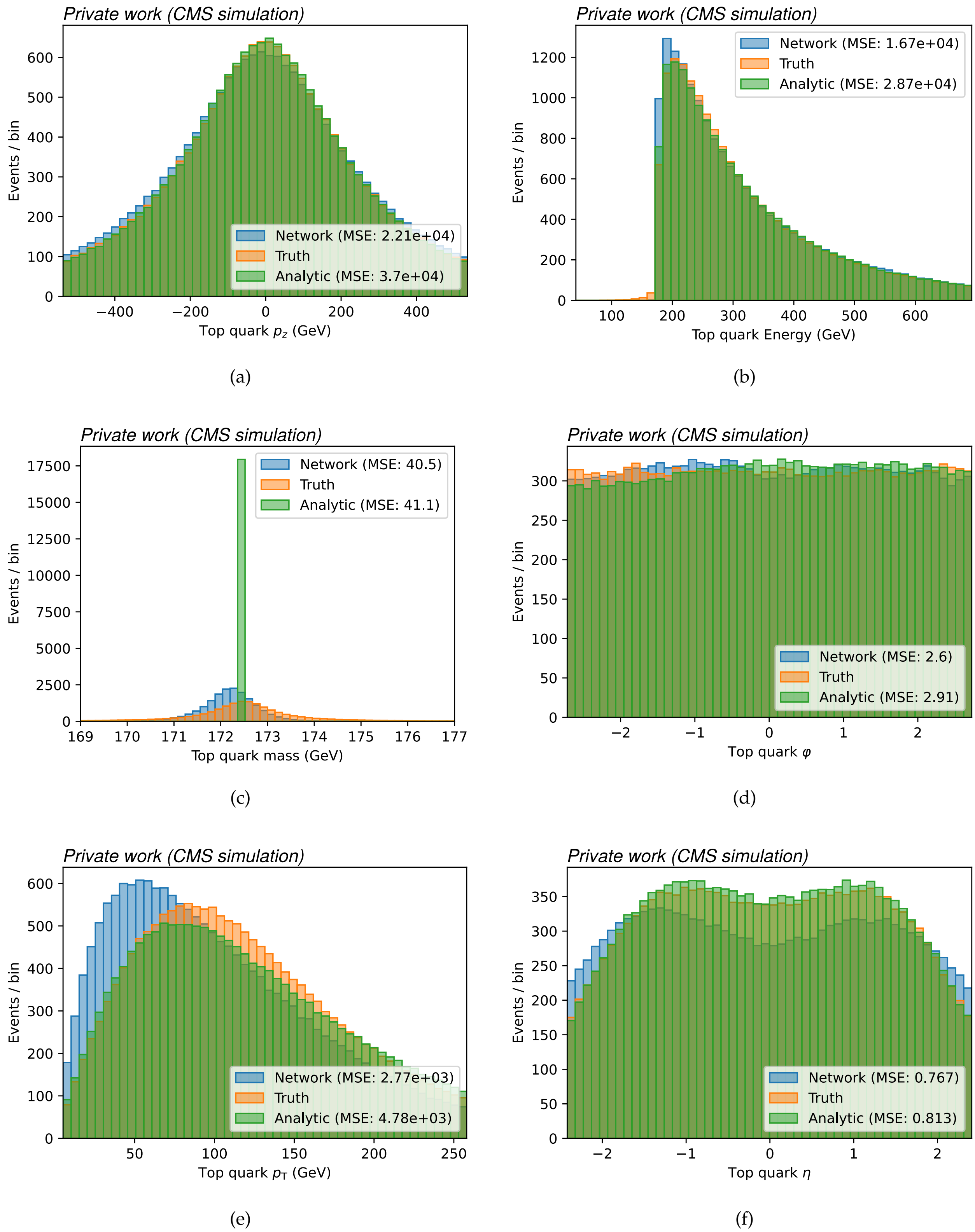
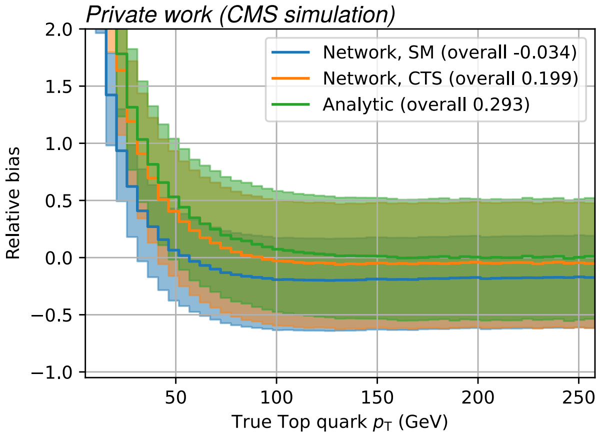
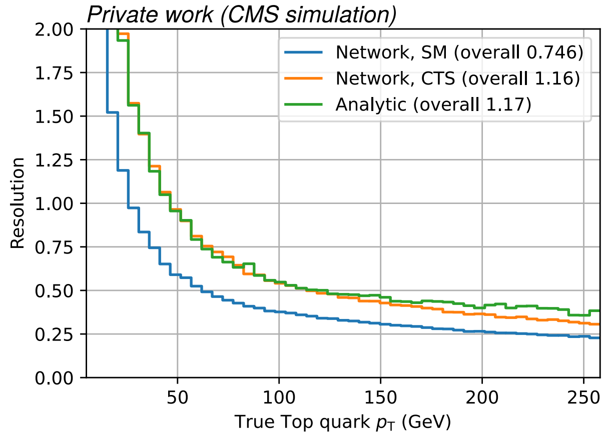


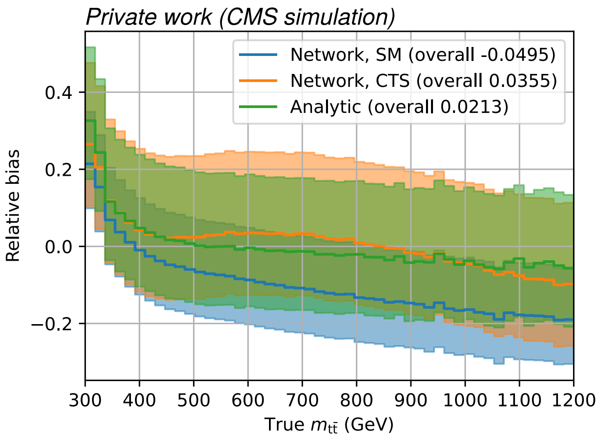
Figure 6.8: Reconstructed and true top quark observables in SM $t\bar{t}$ data. The network has been trained on a training set of SM $t\bar{t}$ data. The legend lists the mean squared error (MSE) between the reconstructed value and the truth (ME level value). A smaller value indicates a better reconstruction.



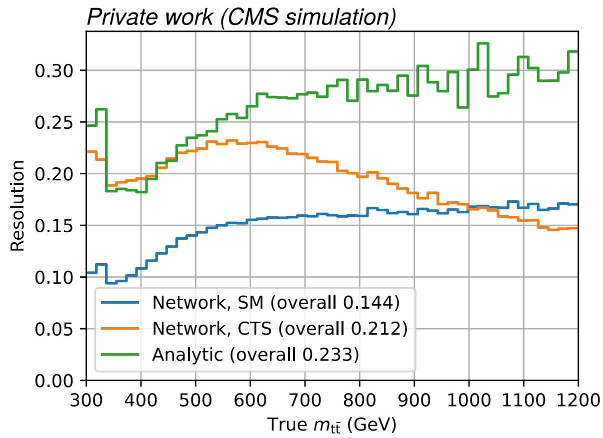
(a)



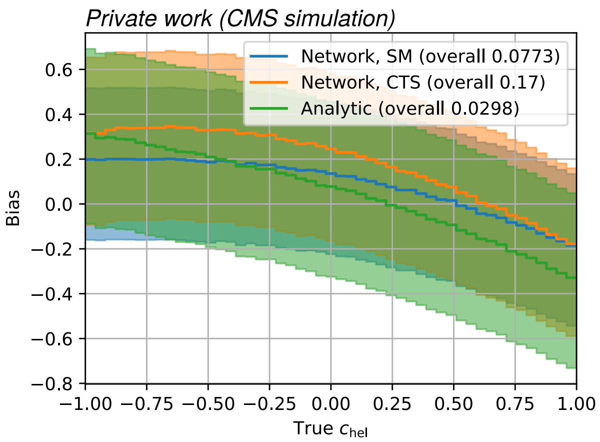
(b)



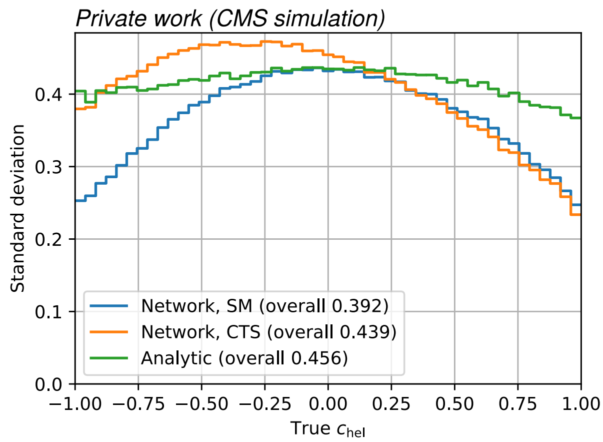
(c)



(d)



(e)



(f)

Figure 6.9: Biases and variances/resolutions for selected observables in SM $t\bar{t}$ data. The true value of the observable is shown on the x axis. The evaluated networks have been trained on SM and on the crafted training set (CTS). For the bias the band around the central line indicates the its standard deviation, which is equal to the resolution.

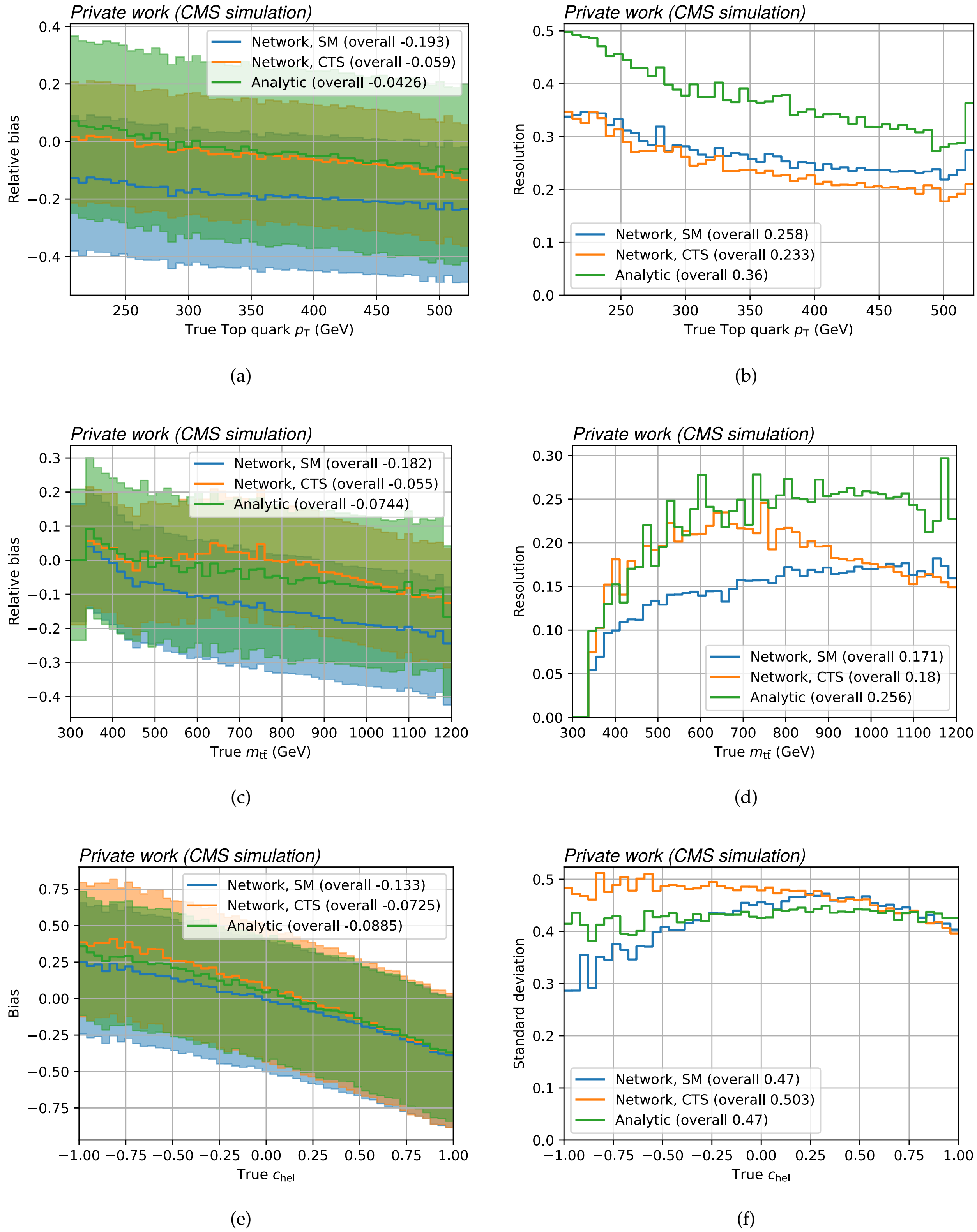


Figure 6.10: Biases and variances/resolutions for selected observables in an pseudoscalar Higgs boson at a mass of 1000 GeV and relative decay width of 2.5 %. The true value of the observable is shown on the x axis. The evaluated networks have been trained on SM and on the crafted training set (CTS). For the bias the band around the central line indicates the its standard deviation, which is equal to the resolution.

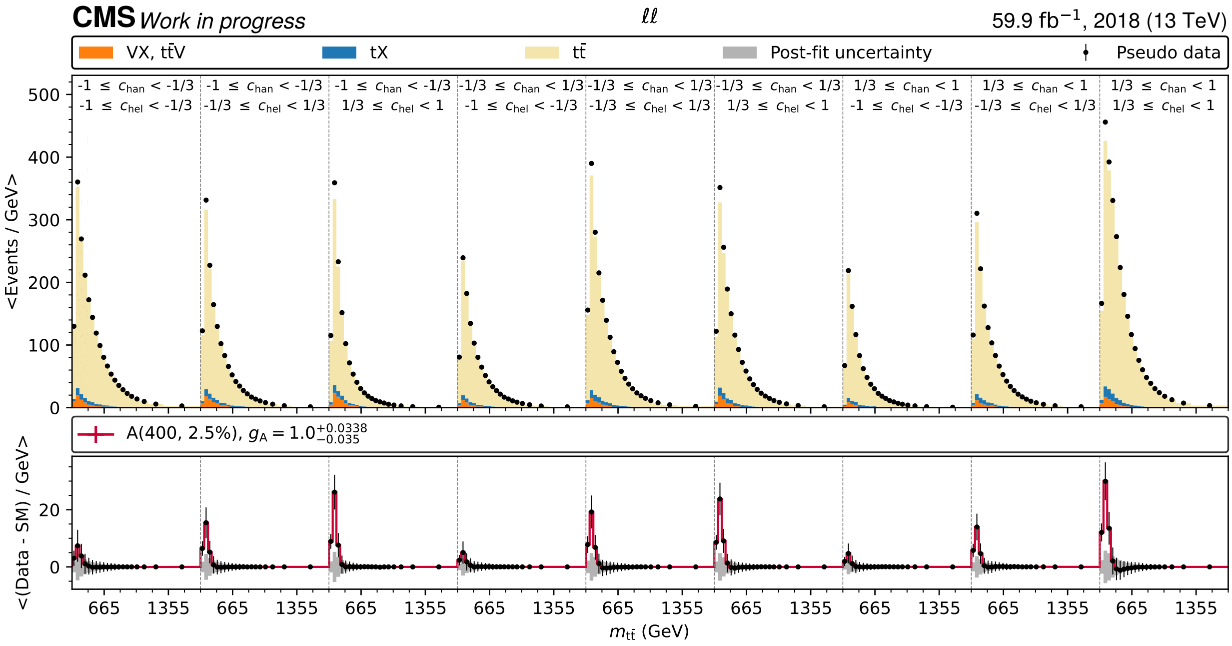
The figures show the dependence on top quark p_T, $m_{t\bar{t}}$ and c_{hel}. The latter two are search observables of the analysis presented in the previous chapter and therefore of special importance. When trying to reconstruct SM data, the SMTS yields the best results. The network shows a slight bias increase for $m_{t\bar{t}}$ and c_{hel} but has a clear improvement of the resolution and standard deviation. In the case of top quark p_T and $m_{t\bar{t}}$ the resolution even becomes almost twice as good. Looking at the CTS, the results are close to the analytic approach but still show an improvement. A different conclusion can be drawn from the results obtained by evaluating on the pseudoscalar Higgs boson data. There, overall the CTS gives the best results, especially in the case of top quark p_T and $m_{t\bar{t}}$ resolution. Again the SMTS is a close contender to the best result. The fact that the SMTS performs better for SM, while the CTS performs better for a high mass Higgs boson is expected. The CTS was provided a large number of events even at high $m_{t\bar{t}}$ during training and thus is paying more attention to this high mass region.

A study has been performed on the final improvement the network provides in case of a full analysis. For this, the strategy of the previous chapter was followed to obtain constraints on the coupling of the heavy Higgs bosons, once evaluated using the analytic method and once using the neural network method. For the study the binning of the $m_{t\bar{t}}$ axis of the histograms used in the analysis have been extended by five bins of a width of 40 GeV each so that the range covered goes as low as 320 GeV. This enables the comparison of the network versus the analytic solution also on very low values of $m_{t\bar{t}}$ where the analytic solution is expected to be particularly good. A comparison between the histograms of the search observables is shown in figure 6.11 for $\mathrm{A}(400\,\mathrm{GeV}, 2.5\,\%)$. While at this low width it is very difficult to resolve the full peak-dip structure, the distribution resulting from the neural network has a visible "dip" in several locations. The "dip" in the analytic method is only slightly visible and appears only for simultaneously high c_{hel} and c_{han} values, where the signal of A is especially prominent. Figure 6.12 shows the upper limits derived for both methods considering the simulation for the year 2018. The neural network is seen to improve the limits for all heavy Higgs masses. At high masses the difference is even close to one standard deviation. But even at the low masses, there is a clear improvement. Only for around $m_A = 600\,\mathrm{GeV}$ the results are very close. This is because of the very low interference cross section (see figure 5.3) around that mass, which means there is very little to none cancellation between interference and resonance. The cancellation is the main reason for the importance of the $m_{t\bar{t}}$ resolution in the analysis.

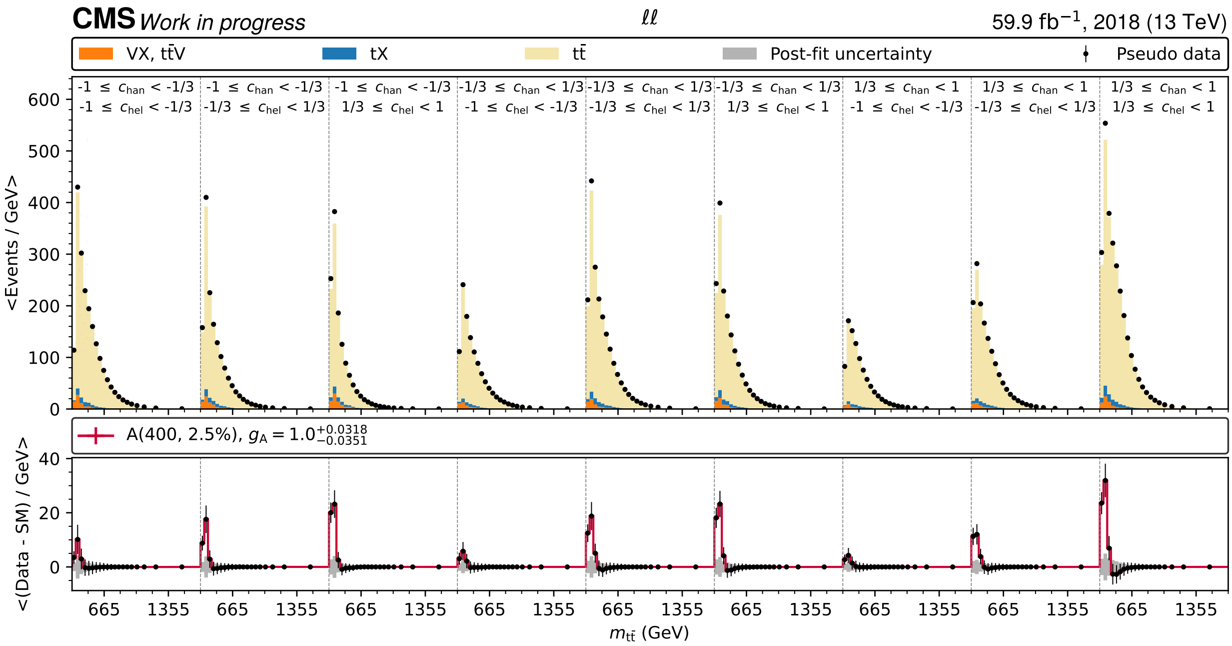
6.7 Discussion

This chapter introduced a neural network that fully replaces the top quark reconstruction used in many dileptonically decaying top pair analyses. Perhaps most noteworthy it fully replaces the classic method without any loss in functionality. Moreover the network significantly outperforms the classic method. Especially the invariant mass of the top pair $m_{t\bar{t}}$, which is the most important observable of the analysis in the previous chapter, is reconstructed with a resolution almost twice as accurate. In some cases the network was seen to have a slight increase in bias, while the overall mean squared error still is significantly smaller. It should be noted that for analyses like the one of the previous chapter, monotone bias is of little importance, as this only shifts the bins content consistently for prediction and observation. The analysis is searching for differences between prediction and observation and not for their absolute positions.

Another notable result is that the network that was trained on the crafted data, instead of SM data, produces an excellent reconstruction that is also seen to be better than the classic approach. The



(a) Analytic method



(b) Neural network (SM training)

Figure 6.11: Histograms of the search observables as described in section 5.10 including the low $m_{t\bar{t}}$ region to values of 320 GeV. Instead of measured data, pseudo data is used to perform a fit (see section 5.11) to determine the uncertainties. The pseudo data is obtained by summing the SM and heavy Higgs boson simulation. The upper figure shows the histogram using the analytic method for reconstruction, while the lower figure is using the neural network instead. The neural network is able to resolve the peak-dip structure better.

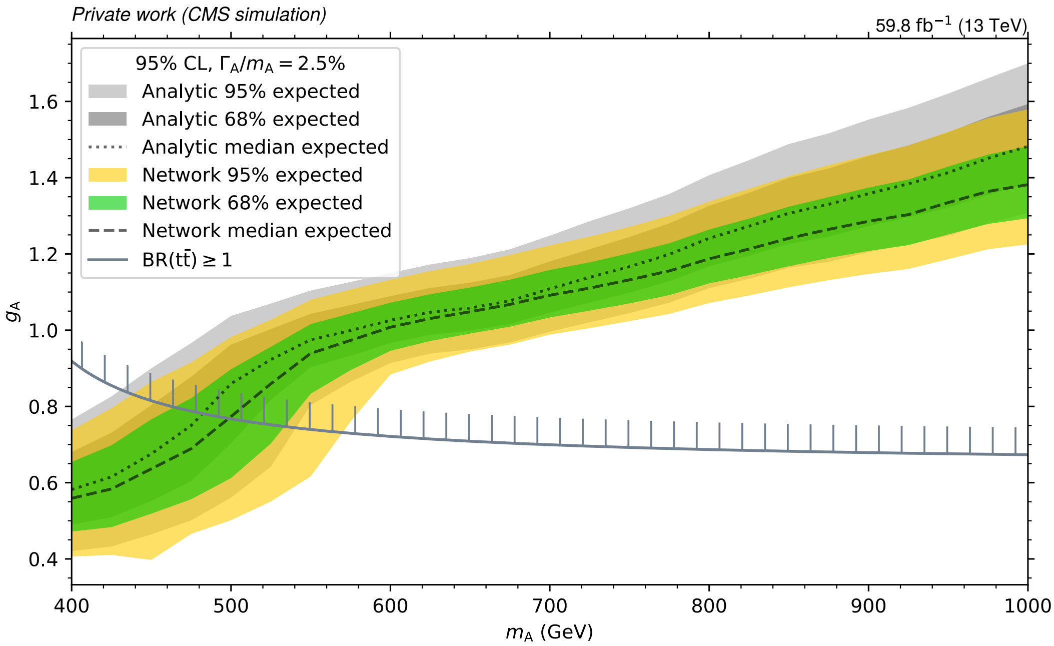


Figure 6.12: Expected upper limits of the coupling for a pseudoscalar Higgs boson for the SM-only hypothesis using the classic analytic reconstruction and using the neural network reconstruction. The network was trained on SM data. The strategy to derive the limits is the same as in chapter 5, except that the $m_{t\bar{t}}$ axis of the histogram was extended by five bins to 320 GeV and only simulation for the year 2018 was used.

crafted data set has been made so that it has no resemblance to any production process seen at LHC and only contains the top quark decay. This information appears to be sufficient to comprehend the top quark properties. However, crafted and SM trained network do not replicate each other, as it can be seen that the former network performs worse than the SM trained one when evaluating on SM. For this reason the choice of using the SM trained network is likely the best one for analyses within the framework of SM. On the other hand for BSM searches, where as little as possible assumptions need to be made, the crafted data set can be more favorable. It should also be noted that if the SM functions as null hypothesis in an analysis, using the SM as training can be considered the conservative approach, as in the worst case this merely brings the reconstructed distributions closer to the SM.

The crafted training set shows a few features, which are quite different than what is seen in the SM. Reconstructed results might be improved by incorporating more features which result from ME computations into the crafted data. An example would be the $t\bar{t}$ p_T, which is being restricted to 0, while in higher order computations the p_T distribution shows a long tail. Furthermore, depending on the analysis, the crafted data set could be equipped with additional sources of MET, which should tell the network to give the best results in presence of more than just neutrinos. An example use case poses a search for dark matter. Other possibilities include a wide constant spread of the top quark mass in the training set to gain sensitivity on this mass or the inclusion of additional particles alongside the top quarks.

One could consider to extend the network's architecture further to possibly obtain even better results. Recent years particle physics publications have proposed increasingly sophisticated network architectures. In reference [164] for example jet tagging has been facilitated via point clouds and a graph neural network. These ideas might also aid the reconstruction of top quarks.

In conclusion the current state of the network already performs strikingly well and is able to provide a significant improvement to the analysis of the previous chapter. The network is ready to be used and can fully replace the classic approach. This holds true not only for heavy Higgs searches, but also to other types of analyses, including other BSM searches and SM measurements.

Chapter 7

Summary and Conclusion

The thesis described work done at the CMS experiment and presented an analysis using CMS data recorded between 2016 and 2018. In particular, the thesis gave an overview of the current state of the standard model of particle physics, an introduction to the beyond standard model particles that are searched for in the analysis and an overview describing the CMS experiment. In the following parts of the thesis the focus was put on the assembly and commissioning of one of the subdetectors of the CMS experiment, which was the luminometer BCM1F. Its previously measured luminosity was one input to the analysis presented subsequently. The analysis involves searches for additional Higgs bosons that are heavy with masses above 400 GeV and decay into top quarks. One limiting factor of the analysis was the resolution of the reconstruction algorithm employed for the top quarks. A novel approach to replace the algorithm using a deep neural network was introduced in the following.

The BCM1F detector was upgraded to its current design consisting of silicon sensors only. For this purpose several batches of sensors that were produced for the detector were qualified. After integration of said sensors into the detector electronics, the assembled detector was qualified as well. Together with the start up of the LHC in 2022 the detector was brought into its final commissioning stages. The newly upgraded frontend together with its backend electronics showed precise measurements not only for the luminosity but also for the background. This marked the successful deployment of the upgraded BCM1F detector, whose data is expected to be used for practically all CMS analyses of the following years.

The analysis featuring the search for heavy Higgs bosons decaying into a top quark pair was performed on CMS data that summed up to an integrated luminosity of 138 fb^{-1}. The heavy Higgs bosons were electrically neutral and either scalar or pseudoscalar and the final decay products included four leptons. While integrating previous research work, the analysis expanded past strategies by including an extra search observable in addition to two others. These three observables featured the invariant mass of the top quark pair $m_{\mathrm{t\bar{t}}}$ and spin correlations. The former reflected the mass of the heavy Higgs bosons, while the latter gave information on the spin and CP properties of the Higgs bosons. The observables were obtained using a reconstruction algorithm that was based on analytically solving the equations obtained from energy and momentum conservation. The algorithm showed a resolution of around 25 % on the invariant mass. The analysis was refined by the inclusion of several corrections and systematic uncertainties, which had not been considered in previous searches of this kind. Especially the NNLO QCD and electroweak corrections, which facilitated the evaluation of the SM top quark Yukawa coupling uncertainty, turned out to be crucial. Probed masses of the heavy Higgs bosons ranged from 400 GeV to 1000 GeV, where no discovery was observed. The search excluded coupling strength modifiers above the range 1.1 to 2.0 for the scalar boson and 0.95 to 1.5 for the pseudoscalar boson. A significant improvement of these constraints is expected to be seen in a soon to be published analysis, which additionally features a low $m_{t\bar{t}}$ region below 520 GeV and the semileptonic decay of the top quarks. Further improvements can be made by an improved theory description, as theoretic uncertainties were the sources with the highest impacts. A better understanding of the SM would contribute here, achievable for example through future high precision measurements, such as the ones expected after the high-luminosity upgrade of the LHC.

The resolution of the reconstruction algorithm turned out to be crucial, as the distinctive heavy Higgs boson signal was strongly affected by it. To improve in this regard, a deep neural network was designed that fully reconstructs the top quark four-momenta, independently of the analytic reconstruction algorithm. The network consisted of two parts, one for assigning jets to the bottom quarks of the decay of the top quarks, and one for the final reconstruction. Training of the network was performed on two different types of data sets, one was an SM simulation and the other one was a specifically crafted data set of top quark pair decay but without any specific assumption on the production of the top quarks. The key feature of the crafted data set was a uniform spread of invariant top pair masses. The network trained with SM data was observed to perform exceptionally well, with a resolution of around 15 % on the invariant mass, much better than the analytic reconstruction. Additionally, other probed quantities were also reconstructed with higher precision, indicated by a lower mean error than what was obtained from the analytic solution. The network trained with the crafted data did not outperform the other network for SM data but showed a better performance for the case of high mass Higgs bosons, thanks to the uniform spread over a large range of masses inside the data set. This demonstrated that the network pays more attention to details depending on which assumptions have been made inside its training data. The new method could replace the classic approach in the near future and become the new standard for analyses of dileptonically decaying top quarks.

Bibliography

[1] Particle Data Group et al. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020). 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. eprint: https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf.

[2] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[3] S. P. Martin and J. D. Wells. *Elementary Particles and Their Interactions*. Springer International Publishing, 2022. DOI: 10.1007/978-3-031-14368-7.

[4] D. Griffiths. *Introduction to elementary particles*. 2008. ISBN: 978-3-527-40601-2.

[5] CMS Collaboration. "Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the lepton+jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D 100, 072007 (2019)* (July 2, 2019). DOI: 10.1103/PhysRevD.100.072007. arXiv: 1907.01590v2 [hep-ex].

[6] NNPDF Collaboration et al. "Parton distributions from high-precision collider data". In: (June 1, 2017). DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428v2 [hep-ph].

[7] A. Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: (Dec. 23, 2014). DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420v2 [hep-ph].

[8] F. Maltoni et al. "Choosing the Factorization Scale in Perturbative QCD". In: (Mar. 14, 2007). arXiv: hep-ph/0703156v1 [hep-ph].

[9] CMS Collaboration. *Overview of CMS cross section results*. 2022. URL: https://twiki.cern.ch/twiki/pub/CMSPublic/PhysicsResultsCombined/CMSCrossSectionSummaryBarChart.pdf.

[10] T. M. P. Tait and C. P. Yuan. "Single Top Production as a Window to Physics Beyond the Standard Model". In: *Phys.Rev.D63:014018,2000* (July 26, 2000). DOI: 10.1103/PhysRevD.63.014018. arXiv: hep-ph/0007298v2 [hep-ph].

[11] A. Brandenburg, Z. G. Si, and P. Uwer. "QCD-corrected spin analysing power of jets in decays of polarized top quarks". In: *Phys.Lett. B539 (2002) 235-241* (May 2, 2002). DOI: 10.1016/S0370-2693(02)02098-1. arXiv: hep-ph/0205023v1 [hep-ph].

[12] A. Djouadi and J. Quevillon. "The MSSM Higgs sector at a high M_{SUSY}: reopening the low tanβ regime and heavy Higgs searches". In: *JHEP 1310 (2013) 028* (Apr. 5, 2013). DOI: 10.1007/JHEP10(2013)028. arXiv: 1304.1787v2 [hep-ph].

[13] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: (May 31, 2011). DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034v3 [hep-ph].

[14] CMS Collaboration. "Search for flavor-changing neutral current interactions of the top quark and Higgs boson in final states with two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. Lett. 129 (2022) 032001* (Nov. 3, 2021). DOI: 10.1103/PhysRevLett.129.032001. arXiv: 2111.02219v2 [hep-ex].

[15] D. Dicus, A. Stange, and S. Willenbrock. "Higgs Decay to Top Quarks at Hadron Colliders". In: *Phys.Lett. B333 (1994) 126-131* (May 1, 1994). DOI: 10.1016/0370-2693(94)91017-0. arXiv: hep-ph/9404359v1 [hep-ph].

[16] CERN. *LHC the guide*. 2017. URL: https://cds.cern.ch/record/2255762.

[17] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003.

[18] ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002.

[19] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004.

[20] LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005.

[21] CERN. *The third run of the Large Hadron Collider has successfully started*. 2022. URL: https://home.cern/news/news/cern/third-run-large-hadron-collider-has-successfully-started.

[22] E. Mobs. "The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018". In: (Aug. 2018). General Photo. URL: https://cds.cern.ch/record/2636343.

[23] J. Maestre et al. "Design and behaviour of the Large Hadron Collider external beam dumps capable of receiving 539 MJ/dump". In: *Journal of Instrumentation* 16.11 (Nov. 2021), P11019. DOI: 10.1088/1748-0221/16/11/P11019. URL: https://dx.doi.org/10.1088/1748-0221/16/11/P11019.

[24] CERN. *The new LHC protectors*. 2018. URL: https://home.cern/news/news/accelerators/new-lhc-protectors.

[25] D. Boussard et al. "The LHC superconducting cavities". In: vol. 2. Feb. 1999, 946–948 vol.2. ISBN: 0-7803-5573-3. DOI: 10.1109/PAC.1999.795409.

[26] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[27] R. Bailey and P. Collier. "Standard Filling Schemes for Various LHC Operation Modes". In: (Sept. 2003). URL: https://cds.cern.ch/record/691782.

[28] X. C. Vidal and R. Cid. *Beta & Emittance*. 2022. URL: https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.beta___emittance.

[29] CMS Collaboration. *CMS Luminosity - Public Results*. 2022. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[30] CERN. *The HL-LHC project*. 2022. URL: https://hilumilhc.web.cern.ch/content/hl-lhc-project.

[31] CMS Collaboration. *LS2 Report: Installation of the CMS beam pipe*. 2021. URL: https://home.cern/news/news/experiments/ls2-report-installation-cms-beam-pipe.

[32] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett. B 716 (2012) 30* 716.1 (July 31, 2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235v2 [hep-ex].

[33] CMS Collaboration. *People Statistics*. 2022. URL: https://cms.cern/collaboration/people-statistics.

[34] J. Lidrych. "Measurement of the differential cross section of W and Z bosons in the boosted region in proton-proton collisions with the CMS experiment". Dissertation, Universität Hamburg, 2021. Dissertation. Hamburg: Universität Hamburg, 2022, p. 201. DOI: 10.3204/PUBDB-2022-03557. URL: https://bib-pubdb1.desy.de/record/480191.

[35] S. R. Davis. "Interactive Slice of the CMS detector". In: (Aug. 2016). URL: https://cds.cern.ch/record/2205172.

[36] CMS Collaboration. *Silicon Pixels*. 2021. URL: https://cms.cern/detector/identifying-tracks/silicon-pixels.

[37] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST 9 (2014) P10009* 9.10 (May 26, 2014), P10009–P10009. DOI: 10.1088/1748-0221/9/10/p10009. arXiv: 1405.6569 [physics.ins-det].

[38] Tracker Group of the CMS collaboration. "The CMS Phase-1 pixel detector upgrade". In: *Journal of Instrumentation* 16.02 (Feb. 2021), P02027–P02027. DOI: 10.1088/1748-0221/16/02/p02027.

[39] C. Biino. "The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections". In: *Journal of Physics: Conference Series* 587 (Feb. 2015), p. 012001. DOI: 10.1088/1742-6596/587/1/012001.

[40] E. Tournefier. "The preshower detector of CMS at LHC". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 461.1 (2001). 8th Pisa Meeting on Advanced Detectors, pp. 355–360. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(00)01243-2.

[41] T. Sakuma. "Cutaway diagrams of CMS detector". In: (May 2019). URL: http://cds.cern.ch/record/2665537.

[42] C. Isik. *Phase 1 Upgrade of the CMS Hadron Calorimeter*. Tech. rep. Geneva: CERN, Mar. 2022. URL: https://cds.cern.ch/record/2810162.

[43] J. Mans et al. *CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter*. Tech. rep. Additional contact persons: Jeffrey Spalding, Fermilab, spalding@cern.ch, Didier Contardo, Universite Claude Bernard-Lyon I, contardo@cern.ch. Sept. 2012. URL: https://cds.cern.ch/record/1481837.

[44] CMS Collaboration. *Detector Overview*. 2022. URL: https://cms.cern/news/detector-overview.

[45] CMS Collaboration. *Muon Drift Tubes*. 2022. URL: https://cms.cern/detector/detecting-muons/muon-drift-tubes.

[46] N. Pozzobon. *The CMS Muon System performance during the LHC Run-2*. Tech. rep. Geneva: CERN, Sept. 2019. DOI: 10.1088/1748-0221/14/11/C11031. URL: https://cds.cern.ch/record/2701333.

[47] V. Khachatryan. "The CMS trigger system". In: *JINST 12 (2017) P01020* 12.01 (Sept. 8, 2016), P01020–P01020. DOI: 10.1088/1748-0221/12/01/p01020. arXiv: 1609.02366.

[48] M. Tosi. *The CMS trigger in Run 2*. Tech. rep. Geneva: CERN, Oct. 2017. DOI: 10.22323/1.314.0523. URL: https://cds.cern.ch/record/2290106.

[49] G. L. Bayatyan et al. *CMS computing: Technical Design Report*. Technical design report. CMS. Submitted on 31 May 2005. Geneva: CERN, 2005. URL: https://cds.cern.ch/record/838359.

[50] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. There is an error on cover due to a technical problem for some items. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/922757.

[51] G. Petrucciani, A. Rizzi, and C. Vuosalo. "Mini-AOD: A New Analysis Data Format for CMS". In: *Journal of Physics: Conference Series 664 (2015) 7 072052* 664.7 (Feb. 15, 2017), p. 072052. DOI: 10.1088/1742-6596/664/7/072052. arXiv: 1702.04685v2 [physics.ins-det].

[52] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST 12 (2017) P10003* 12.10 (June 15, 2017), P10003–P10003. DOI: 10.1088/1748-0221/12/10/p10003. arXiv: 1706.04965 [physics.ins-det].

[53] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP 0804:063, 2008* 2008.04 (Feb. 8, 2008), pp. 063–063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[54] CMS Collaboration. "Boosted Top Jet Tagging at CMS". In: (2014). URL: http://cds.cern.ch/record/1647419.

[55] CMS Collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST 13 (2018) P05011* 13.05 (Dec. 19, 2017), P05011–P05011. DOI: 10.1088/1748-0221/13/05/p05011. arXiv: 1712.07158v2 [physics.ins-det].

[56] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15 (Dec. 2020), P12012–P12012. DOI: 10.1088/1748-0221/15/12/P12012.

[57] CMS Collaboration. "Detector Drawings". CMS Collection. Mar. 2012. URL: https://cds.cern.ch/record/1433717.

[58] P. Lujan. "The Pixel Luminosity Telescope: A detector for luminosity measurement at CMS using silicon pixel sensors". In: (June 17, 2022). arXiv: 2206.08870v1 [physics.ins-det].

[59] M. Guthoff. "Radiation Damage to the diamond-based Beam Condition Monitor of the CMS Detector at the LHC". 2014. URL: https://cds.cern.ch/record/1977429.

[60] A. J. Bell. "Beam and Radiation Monitoring for CMS". In: *2008 IEEE Nuclear Science Symposium Conference Record*. IEEE, Oct. 2008. DOI: 10.1109/nssmic.2008.4774822.

[61] M. Guthoff. "Instrumentation for beam radiation and luminosity measurement in the CMS experiment using novel detector technologies". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 845 (2017). Proceedings of the Vienna Conference on Instrumentation 2016, pp. 565–569. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.06.028. URL: https://www.sciencedirect.com/science/article/pii/S0168900216305708.

[62] CMS Collaboration. *The Phase-2 Upgrade of the CMS Tracker*. Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.QZ28.FLHW. URL: https://cds.cern.ch/record/2272264.

[63] D. Przyborowski, J. Kaplon, and P. Rymaszewski. "Design and Performance of the BCM1F Front End ASIC for the Beam Condition Monitoring System at the CMS Experiment". In: *IEEE Trans. Nucl. Sci.* 63 (2016), 2300–2308. 9 p. DOI: 10.1109/TNS.2016.2575781. URL: http://cds.cern.ch/record/2256836.

[64] M. Friedl. "Analog optohybrids for the readout of the CMS silicon tracker". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 518.1 (2004). Frontier Detectors for Frontier Physics: Proceedin, pp. 515–518. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2003.11.073. URL: https://www.sciencedirect.com/science/article/pii/S0168900203029395.

[65] A. Bell et al. "Fast beam conditions monitor BCM1F for the CMS experiment". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 614.3 (Mar. 2010), pp. 433–438. DOI: 10.1016/j.nima.2009.12.056. URL: https://doi.org/10.1016%2Fj.nima.2009.12.056.

[66] M. Guthoff. *The new Fast Beam Condition Monitor using poly-crystalline diamond sensors for luminosity measurement at CMS. The new Fast Beam Condition Monitor using diamond and silicon sensors for luminosity measurement at CMS*. Tech. rep. Geneva: CERN, 2019. DOI: 10.1016/j.nima.2018.11.071. URL: http://cds.cern.ch/record/2799585.

[67] F. Hartmann. *Evolution of Silicon Sensor Technology in Particle Physics*. Springer International Publishing, 2017. DOI: 10.1007/978-3-319-64436-3.

[68] A. A. Zagozdzinska et al. "New Fast Beam Conditions Monitoring (BCM1F) system for CMS". In: *Journal of Instrumentation* 11.01 (2016), p. C01088. ISSN: 1748-0221. DOI: 10.1088/1748-0221/11/01/C01088. URL: https://bib-pubdb1.desy.de/record/316948.

[69] M. Penno. "A Real-time Histogramming Unit for Luminosity Measurement of each Bunch Crossing at CMS". In: Topical Workshop on Electronics for Particle Physics, Perugia (Italy), 23 Sep 2013 - 27 Sep 2013. Sept. 23, 2013. URL: https://bib-pubdb1.desy.de/record/154451.

[70] J. Hegeman et al. "The CMS Timing and Control Distribution System". In: *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. 2015, pp. 1–3. DOI: 10.1109/NSSMIC.2015.7581984.

[71] S. Orfanelli et al. "A novel Beam Halo Monitor for the CMS experiment at the LHC". In: *Journal of Instrumentation* 10.11 (Nov. 2015), P11011. DOI: 10.1088/1748-0221/10/11/P11011. URL: https://dx.doi.org/10.1088/1748-0221/10/11/P11011.

[72] S. White. "Luminosity Scans at the LHC. Luminosity Scans at LHC". In: (2011). URL: http://cds.cern.ch/record/1357865.

[73] P. Grafström and W. Kozanecki. "Luminosity determination at proton colliders". In: *Progress in Particle and Nuclear Physics* 81 (2015), pp. 97–148. ISSN: 0146-6410. DOI: `https://doi.org/10.1016/j.ppnp.2014.11.002`. URL: `https://www.sciencedirect.com/science/article/pii/S0146641014000878`.

[74] J. Knolle. "Measuring luminosity and the $t\bar{t}Z$ production cross section with the CMS experiment". Dissertation, Universität Hamburg, 2020. Dissertation. Hamburg: Universität Hamburg, 2020, p. 164. DOI: `10.3204/PUBDB-2020-03187`. URL: `https://bib-pubdb1.desy.de/record/442689`.

[75] CERN. *LHC Run 3: physics at record energy starts tomorrow*. 2022. URL: `https://home.cern/news/news/physics/lhc-run-3-physics-record-energy-starts-tomorrow`.

[76] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. DOI: `10.1016/j.physletb.2012.08.020`. URL: `https://doi.org/10.1016%2Fj.physletb.2012.08.020`.

[77] ATLAS Collaboration. "Search for Heavy Higgs Bosons Decaying into Two Tau Leptons with the ATLAS Detector Using pp collisions at $\sqrt{s} = 13$ TeV". In: *Physical Review Letters* 125.5 (July 2020). DOI: `10.1103/physrevlett.125.051801`. URL: `https://doi.org/10.1103%2Fphysrevlett.125.051801`.

[78] CMS Collaboration. "Searches for additional Higgs bosons and for vector leptoquarks in $\tau\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (Aug. 4, 2022). arXiv: `2208.02717v1 [hep-ex]`.

[79] CMS Collaboration. "Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2020.4 (Apr. 2020). DOI: `10.1007/jhep04(2020)171`.

[80] Pepper contributors. *Pepper - ParticlE Physics ProcEssoR*. 2023. URL: `https://gitlab.cern.ch/pepper/`.

[81] CMS Collaboration. *CMSSW*. 2023. URL: `https://cms-sw.github.io/`.

[82] R. Brun et al. *root-project/root: v6.18/02*. Version v6-18-02. Aug. 2019. DOI: `10.5281/zenodo.3895860`. URL: `https://doi.org/10.5281/zenodo.3895860`.

[83] Fermi National Accelerator Laboratory. *Coffea concepts*. 2019. URL: `https://coffeateam.github.io/coffea/concepts.html`.

[84] L. Gray et al. *CoffeaTeam/coffea: Release v0.7.20*. Version v0.7.20. Nov. 2022. DOI: `10.5281/zenodo.7295393`. URL: `https://doi.org/10.5281/zenodo.7295393`.

[85] C. R. Harris et al. "Array programming with NumPy". In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: `10.1038/s41586-020-2649-2`. URL: `https://doi.org/10.1038/s41586-020-2649-2`.

[86] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: `10.1109/MCSE.2007.55`.

[87] Y. Babuji et al. "Parsl: Pervasive Parallel Programming in Python". In: *28th ACM International Symposium on High-Performance Parallel and Distributed Computing (HPDC)*. 2019. DOI: `10.1145/3307681.3325400`. URL: `https://doi.org/10.1145/3307681.3325400`.

[88] C. Zangl. *Hjson, a user interface for JSON*. 2023. URL: https://hjson.github.io/.

[89] H. Schreiner, S. Liu, and A. Goel. *hist*. Version v2.6.3. Please cite the following works when using this software. Jan. 2023. DOI: 10.5281/zenodo.7532971. URL: https://doi.org/10.5281/zenodo.7532971.

[90] A. Collette et al. *h5py/h5py: 3.7.0*. Version 3.7.0. May 2022. DOI: 10.5281/zenodo.6575970. URL: https://doi.org/10.5281/zenodo.6575970.

[91] E. Clement. *Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/StripsAPVPerformance2020.

[92] CMS Collaboration. "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". In: *Eur. Phys. J. C 81 (2021) 800* (Apr. 5, 2021). DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927v2 [hep-ex].

[93] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-LUM-18-002. Geneva: CERN, 2019. URL: http://cds.cern.ch/record/2676164.

[94] CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2621960.

[95] S. Chenarani et al. *Measurement of the trigger efficiencies for a dilepton selection for a* $\mathrm{t\bar{t}}$ *analysis with the full run 2 dataset*. CMS Analysis Note AN-19-140. 2021.

[96] J. Rübenach. *Pepper data - HIG-22-013*. 2023. URL: https://gitlab.cern.ch/jrubenac/pepper_data/-/tree/HIG-22-013.

[97] S. Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP 1006:043,2010* (Feb. 12, 2010). DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581v1 [hep-ph].

[98] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP07(2014)079* (May 1, 2014). DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301v2 [hep-ph].

[99] T. Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: (Oct. 11, 2014). DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012v1 [hep-ph].

[100] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 2003), pp. 250–303. DOI: 10.1016/s0168-9002(03)01368-8.

[101] CMS Collaboration. *Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns. 2022.

[102] S. Tosi, T. Seva, and H. Y.-S. Tong. *Standard Model Cross Sections for CMS at 13 TeV*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat13TeV. 2014.

[103] B. Maier and O. M. Kind. *NLO single-top channel cross sections*. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec. 2015.

[104] CMS Collaboration. *XSDB*. URL: https://cms-gen-dev.cern.ch/xsdb/.

[105] *Cross sections at CMS*. https://wiki.physik.uzh.ch/cms/physics:crosssections.

[106] CMS Collaboration. *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements*. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2636284.

[107] R. V. Harlander, S. Liebler, and H. Mantler. "SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM". In: *Computer Physics Communications 184 (2013) pp. 1605-1617* (Dec. 13, 2012). DOI: 10.1016/j.cpc.2013.02.006. arXiv: 1212.3249v2 [hep-ph].

[108] M. Czakon, P. Fiedler, and A. Mitov. "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$". In: *Physical Review Letters* 110.25 (2013). DOI: 10.1103/physrevlett.110.252004. URL: https://doi.org/10.1103%2Fphysrevlett.110.252004.

[109] CMS Collaboration. *Multivariate Electron Identification for Run2*. 2022. URL: https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2#Recommended_MVA_Recipe_V2_for_re.

[110] CMS Collaboration. *Residual Scale and Smearing Corrections*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations#Residual_Scale_Smearing_Correcti.

[111] P. Bortignon. *Baseline muon selections for Run-II*. 2021. URL: https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonSelection.

[112] A. Bodek et al. "Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments". In: *The European Physical Journal C 2012, Volume 72, Number 10, 2194* (Aug. 18, 2012). DOI: 10.1140/epjc/s10052-012-2194-8. arXiv: 1208.3710v3 [hep-ex].

[113] D. Karasavvas. *Jet Identification for the 13 TeV UL data*. 2021. URL: https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVUL.

[114] CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST 12 (2017) P02014* (July 13, 2016). DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663v2 [hep-ex].

[115] CMS Collaboration. "Measurement of the $\mathrm{t\bar{t}}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C 79 (2019) 368* (Dec. 26, 2018). DOI: 10.1140/epjc/s10052-019-6863-8. arXiv: 1812.10505v2 [hep-ex].

[116] CMS Collaboration. "Measurement of the top quark polarization and $\mathrm{t\bar{t}}$ spin correlations using dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D 100, 072002 (2019)* (July 8, 2019). DOI: 10.1103/PhysRevD.100.072002. arXiv: 1907.03729v2 [hep-ex].

[117] CMS Collaboration. "Measurement of the top quark pole mass using $\mathrm{t\bar{t}}$+jet events in the dilepton final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV". In: (July 5, 2022). arXiv: 2207.02270v1 [hep-ex].

[118] CMS Collaboration. *MET Filter Recommendations for Run II*. 2022. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2.

[119] CMS Collaboration. *Utilities for Accessing Pileup Information for Data*. 2022. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData.

[120] CMS Collaboration. "Search for dark matter particles produced in association with a top quark pair at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. Lett. 122, 011803 (2019)* (July 17, 2018). DOI: 10.1103/PhysRevLett.122.011803. arXiv: 1807.06522v3 [hep-ex].

[121] M. Grazzini, S. Kallweit, and M. Wiesemann. "Fully differential NNLO computations with MATRIX". In: (Nov. 17, 2017). DOI: 10.1140/epjc/s10052-018-5771-7. arXiv: 1711.06631v1 [hep-ph].

[122] M. Aliev et al. "– HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR". In: *Comput.Phys.Commun.182:1034-1046,2011* (July 8, 2010). DOI: 10.1016/j.cpc.2010.12.040. arXiv: 1007.1327v1 [hep-ph].

[123] CMS Colalboration. *The modeling of the top quark p_t*. 2020. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting.

[124] CMS Collaboration. *EGamma Run 2 UL Recommendations*. 2023. URL: https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018.

[125] CMS Collaboration. *Reference muon id, isolation and trigger efficiencies for Run 2 legacy data.* 2020. URL: https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffsRun2Legacy.

[126] CMS Collaboration. *Methods to apply b-tagging efficiency scale factors*. 2022. URL: https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods.

[127] D. Kovalskyi and J.-F. Schulte. "CMS AN-2021/086 – Measurement of prefiring in the L1 muon trigger in Run 2". In: (2021). URL: https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2021/086.

[128] N. Kidonakis. "NNLL threshold resummation for top-pair and single-top production". In: *Phys.Part.Nucl.45:714,2014* (Oct. 29, 2012). arXiv: 1210.7813v2 [hep-ph].

[129] M. Czakon, D. Heymes, and A. Mitov. "High-precision differential predictions for top-quark pairs at the LHC". In: *Phys. Rev. Lett.* 116.8 (2016), p. 082003. DOI: 10.1103/PhysRevLett.116.082003. arXiv: 1511.00549 [hep-ph].

[130] CMS Collaboration. *MET Analysis - xy-Shift Correction*. 2019. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis#xy_Shift_Correction.

[131] L. Sonnenschein. "Analytical solution of ttbar dilepton equations". In: *Phys.Rev.D73:054015,2006; Erratum-ibid.D78:079902,2008* (Mar. 1, 2006). DOI: 10.1103/PhysRevD.73.05401510.1103/PhysRevD.78.079902. arXiv: hep-ph/0603011v3 [hep-ph].

[132] CMS Collaboration. *Measurement of the differential cross section for $t\bar{t}$ production in the dilepton final state at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2140061.

[133] CMS Collaboration. *First measurement of the differential cross section for ttbar production in the dilepton final state at sqrts = 13 TeV*. Tech. rep. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2047866.

[134] K. Kawade and K. Lie. *Measurement of top quark pair differential cross-sections in the dilepton channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV with ATLAS*. Tech. rep. 5 pages, TOP2016. Geneva: CERN, 2016. arXiv: 1612.01788. URL: https://cds.cern.ch/record/2237440.

[135] I. Korol. "Measurement of Double Differential $t\bar{t}$ Production Cross Sections with the CMS Detector". Universität Hamburg, Diss., 2016. Dr. Hamburg: Universität Hamburg, 2016, p. 236. DOI: 10.3204/DESY-THESIS-2016-011. URL: https://bib-pubdb1.desy.de/record/299582.

[136] A. A. Anuar. "Top Quark Spin and Polarization Properties in Searches for New Phenomena with the CMS Detector at the LHC". Dissertation, Universität Hamburg, 2019. Dissertation. Hamburg: Universität Hamburg, 2019, p. 283. DOI: 10.3204/PUBDB-2020-00203. URL: https://bib-pubdb1.desy.de/record/434642.

[137] W. Bernreuther, D. Heisler, and Z.-G. Si. "A set of top quark spin correlation and polarization observables for the LHC: Standard Model predictions and new physics contributions". In: (Aug. 21, 2015). DOI: 10.1007/JHEP12(2015)026. arXiv: 1508.05271v2 [hep-ph].

[138] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur.Phys. J.C71:1554,2011* (July 10, 2010). DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727v3 [physics.data-an].

[139] CMS Collaboration. *Combine*. 2023. URL: https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/.

[140] R. J. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.

[141] CMS Collaboration. "Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C 79 (2019) 421* (Sept. 27, 2018). DOI: 10.1140/epjc/s10052-019-6909-y. arXiv: 1809.10733v2 [hep-ex].

[142] E. Ranken et al. "Measuring the top quark Yukawa coupling using kinematic distributions of ttbar dilepton decays from the full run 2 dataset". In: (2020). URL: https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/236.

[143] CMS Collaboration. *Luminosity recommendations for Run 2 analyses*. 2022. URL: https://twiki.cern.ch/twiki/bin/view/CMS/LumiRecommendationsRun2.

[144] CMS Collaboration. "CMS PYTHIA 8 colour reconnection tunes based on underlying-event data". In: (May 5, 2022). arXiv: 2205.02905v2 [hep-ex].

[145] W. S. Cleveland. "Robust Locally Weighted Regression and Smoothing Scatterplots". In: *Journal of the American Statistical Association* 74.368 (1979), pp. 829–836. DOI: 10.1080/01621459.1979.10481038. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1979.10481038. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1979.10481038.

[146] A. Popov, J. Steggemann, and M. Verzetti. "Artificial constraints from statistical fluctuations in systematic variations". In: (2019). URL: https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2018/077.

[147] R. D. Cousins. "Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models , with Application to Histograms". In: 2013. URL: https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[148] D. B. Franzosi, E. Vryonidou, and C. Zhang. "Scalar production and decay to top quarks including interference effects at NLO in QCD in an EFT approach". In: (July 21, 2017). arXiv: 1707.06760v2 [hep-ph].

[149] O. Russakovsky and L. Fei-Fei. "Attribute Learning in Large-Scale Datasets". In: *Trends and Topics in Computer Vision*. Springer Berlin Heidelberg, 2012, pp. 1–14. DOI: 10.1007/978-3-642-35749-7_1.

[150] B. A. Betchart, R. Demina, and A. Harel. "Analytic solutions for neutrino momenta in decay of top quarks". In: (May 8, 2013). DOI: 10.1016/j.nima.2013.10.039. arXiv: 1305.1878v2 [hep-ph].

[151] CMS Collaboration. "Measurement of $t\bar{t}$ normalised multi-differential cross sections in pp collisions at $\sqrt{s} = 13$ TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions". In: *Eur. Phys. J. C 80 (2020) 658* (Apr. 10, 2019). DOI: 10.1140/epjc/s10052-020-7917-7. arXiv: 1904.05237v2 [hep-ex].

[152] S. Qiu et al. "A Holistic Approach to Predicting Top Quark Kinematic Properties with the Covariant Particle Transformer". In: (Mar. 11, 2022). arXiv: 2203.05687v2 [hep-ph].

[153] S. Yang et al. "Top quark pair reconstruction using an attention-based neural network". In: *Proceedings of 40th International Conference on High Energy physics — PoS(ICHEP2020)*. Sissa Medialab, Feb. 2021. DOI: 10.22323/1.390.0348.

[154] J. Erdmann et al. "From the Bottom to the Top – Reconstruction of $t\bar{t}$ Events with Deep Learning". In: *JINST 14 (2019) P11015* (July 25, 2019). DOI: 10.1088/1748-0221/14/11/P11015. arXiv: 1907.11181v2 [hep-ex].

[155] S. Wuchterl. "Top quark pole mass from $t\bar{t}$+jet using a machine learning based reconstruction for $t\bar{t}$ kinematics". Dissertation, Universität Hamburg, 2023. Dissertation. Hamburg: Universität Hamburg, 2023, p. 179. DOI: 10.3204/PUBDB-2023-00853. URL: https://bib-pubdb1.desy.de/record/573568.

[156] J. Brehmer et al. "A Guide to Constraining Effective Field Theories with Machine Learning". In: *Phys. Rev. D 98, 052004 (2018)* (Apr. 30, 2018). DOI: 10.1103/PhysRevD.98.052004. arXiv: 1805.00020v4 [hep-ph].

[157] C. M. Bishop. *Pattern Recognition and Machine Learning*. 2006.

[158] X. Glorot, A. Bordes, and Y. Bengio. "Deep Sparse Rectifier Neural Networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by G. Gordon, D. Dunson, and M. Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, Nov. 2011, pp. 315–323. URL: https://proceedings.mlr.press/v15/glorot11a.html.

[159] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". In: (Dec. 22, 2014). arXiv: 1412.6980v9 [cs.LG].

[160] N. Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[161] Y. A. LeCun et al. "Efficient BackProp". In: *Neural Networks: Tricks of the Trade: Second Edition*. Ed. by G. Montavon, G. B. Orr, and K.-R. Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–48. ISBN: 978-3-642-35289-8. DOI: 10.1007/978-3-642-35289-8_3. URL: https://doi.org/10.1007/978-3-642-35289-8_3.

[162] M. Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: `https://www.tensorflow.org/`.

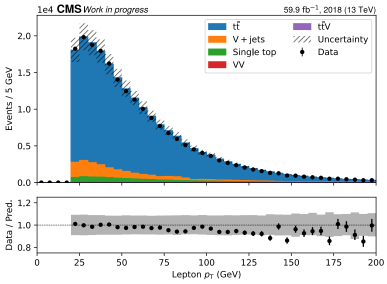
[163] G. James et al. *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013. URL: `https://faculty.marshall.usc.edu/gareth-james/ISL/`.

[164] H. Qu and L. Gouskos. "ParticleNet: Jet Tagging via Particle Clouds". In: *Phys. Rev. D 101, 056019 (2020)* (Feb. 22, 2019). DOI: `10.1103/PhysRevD.101.056019`. arXiv: `1902.08570v3 [hep-ph]`.

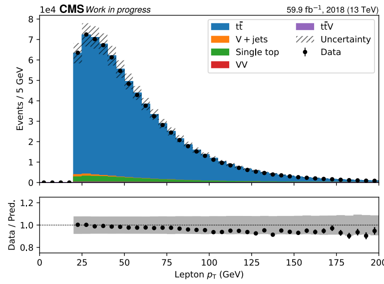
Appendix A

Data to simulation agreement

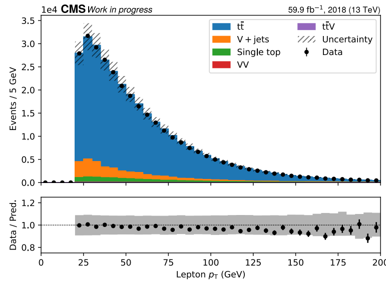
In the following, figures are given comparing the agreement between experimental data and simulation split for different years/eras and lepton flavors. The uncertainties given in the figures include statistical and systematic uncertainties before the fit. For the figures showing jets or lepton observables, all jets or all leptons that are present in the events are considered. All events after applying all cuts, except for cut 15, are used.



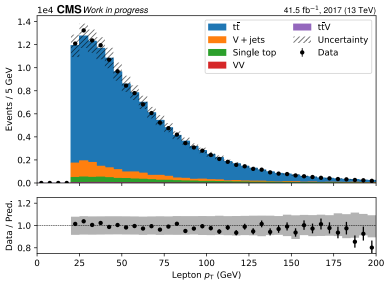
(a) ee



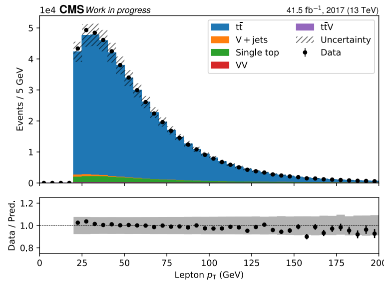
(b) eμ



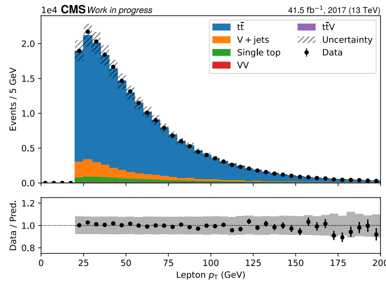
(c) μμ



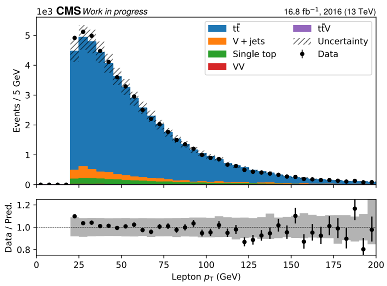
(d) ee



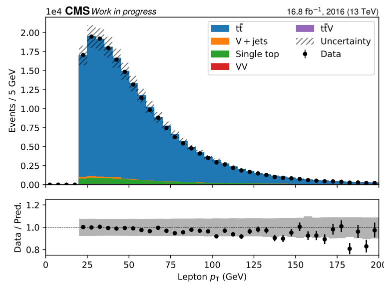
(e) eμ



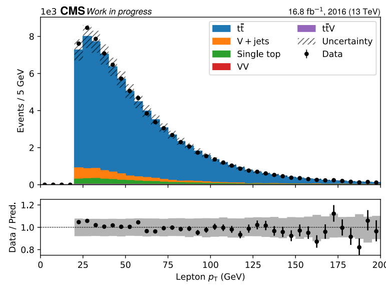
(f) μμ



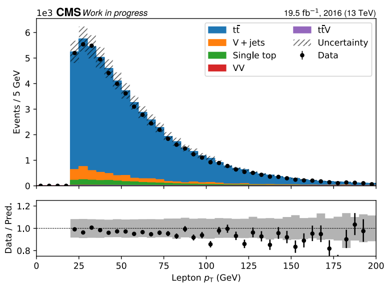
(g) ee



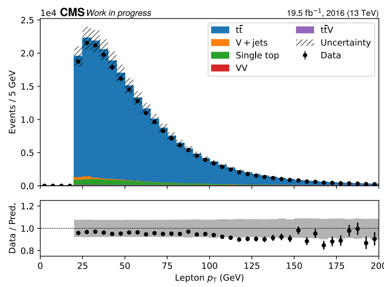
(h) eμ



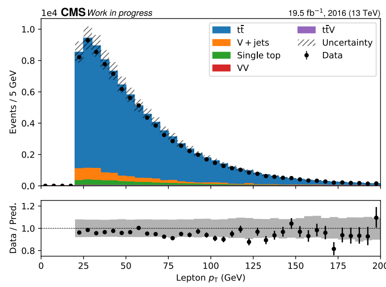
(i) μμ



(j) ee

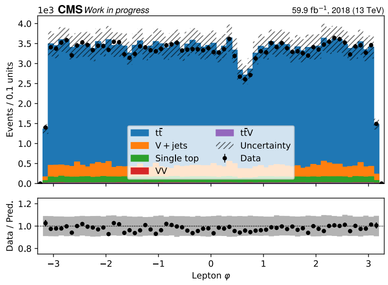


(k) eμ

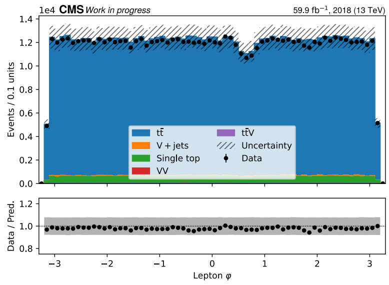


(l) μμ

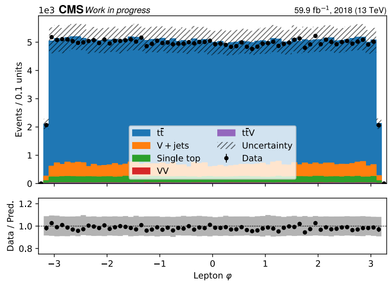
Figure A.1: p_T of the leptons for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



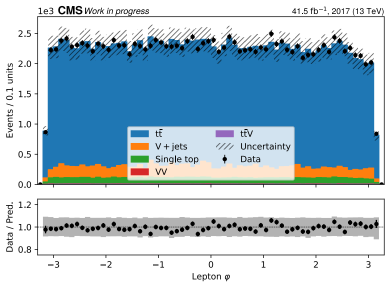
(a) ee



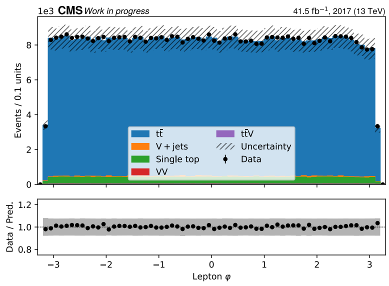
(b) eμ



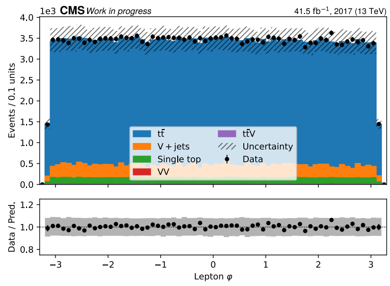
(c) μμ



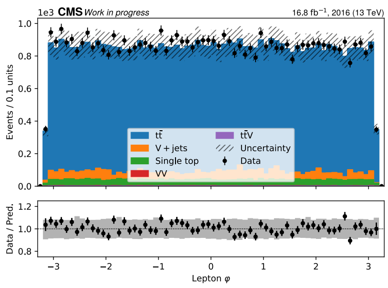
(d) ee



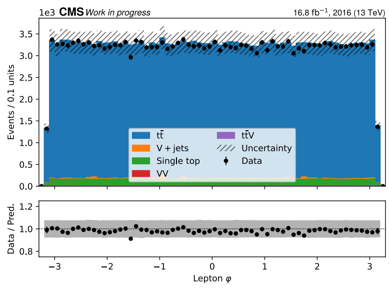
(e) eμ



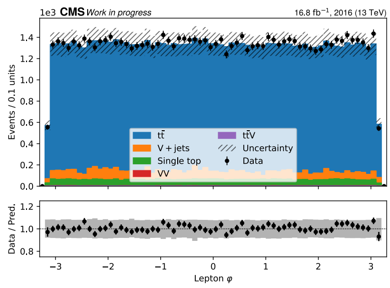
(f) μμ



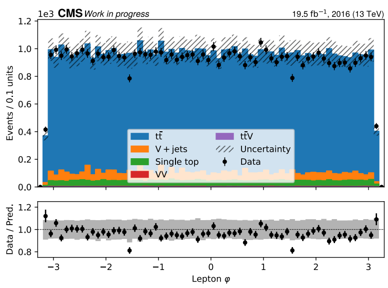
(g) ee



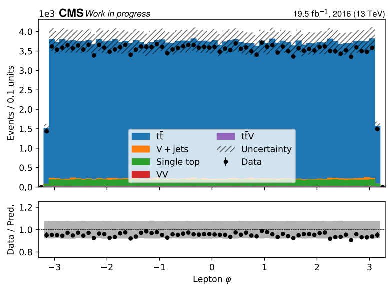
(h) eμ



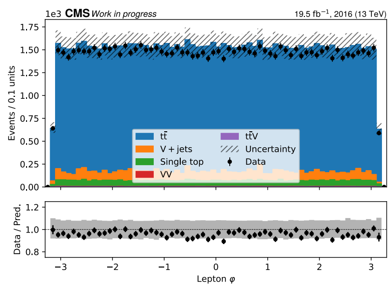
(i) μμ



(j) ee

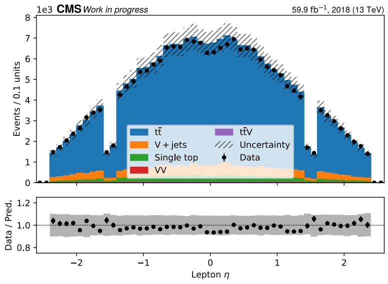


(k) eμ

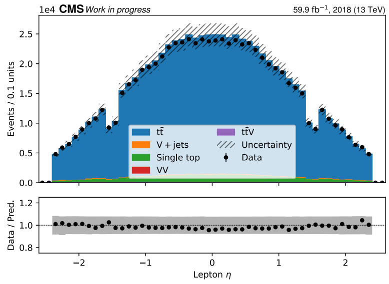


(l) μμ

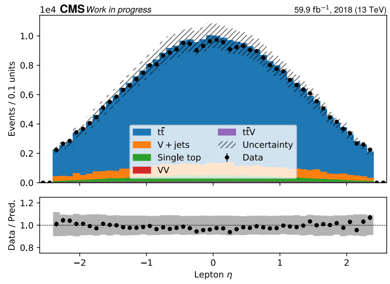
Figure A.2: φ of the leptons for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



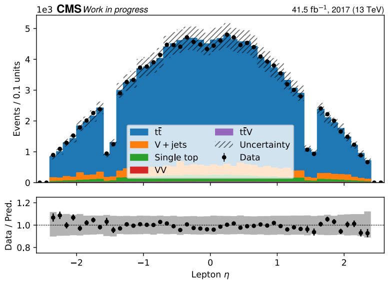
(a) ee



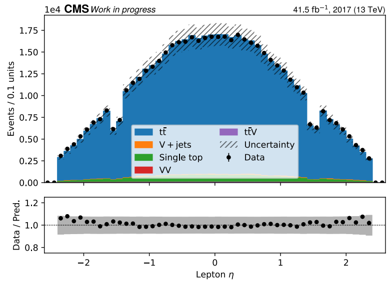
(b) eμ



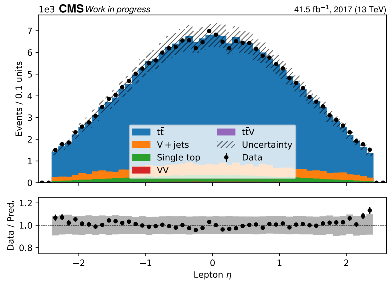
(c) μμ



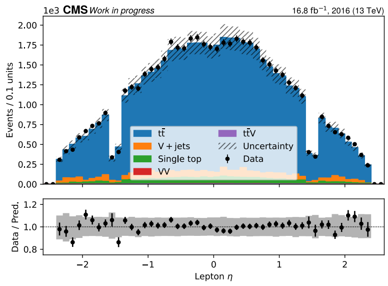
(d) ee



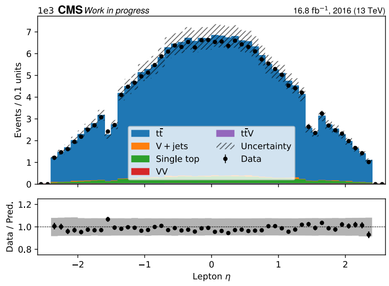
(e) eμ



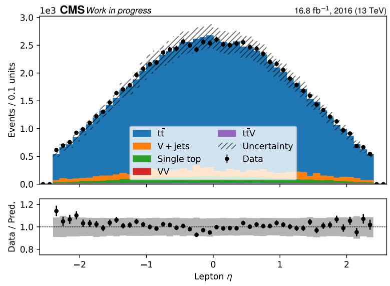
(f) μμ



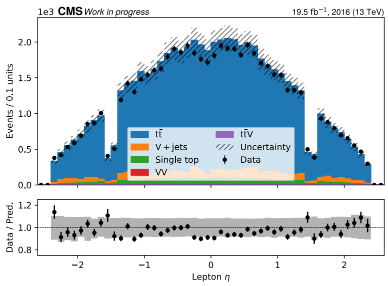
(g) ee



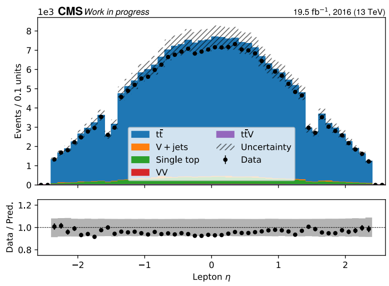
(h) eμ



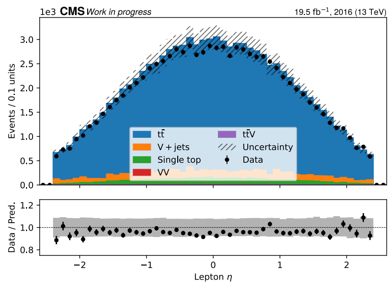
(i) μμ



(j) ee

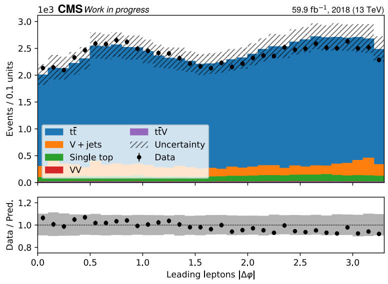


(k) eμ

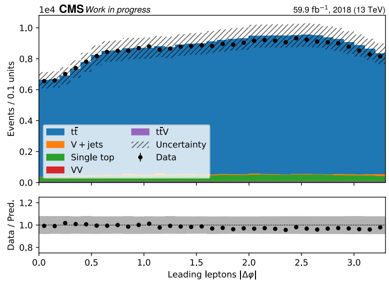


(l) μμ

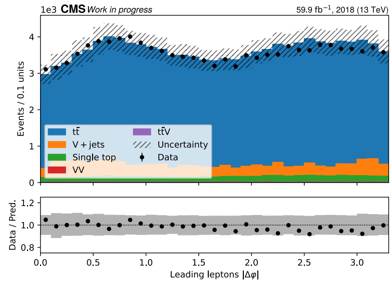
Figure A.3: η of the leptons for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



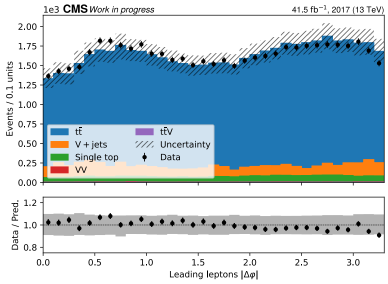
(a) ee



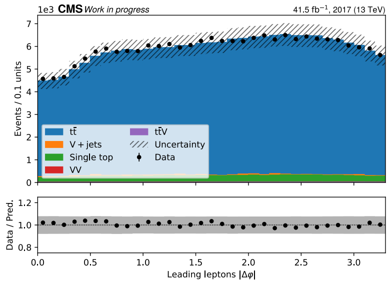
(b) eμ



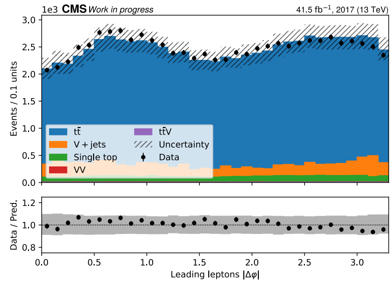
(c) μμ



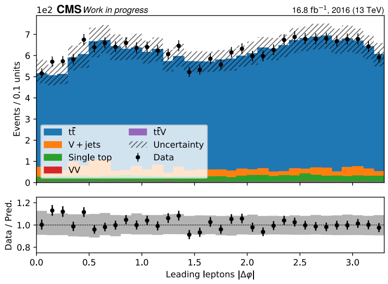
(d) ee



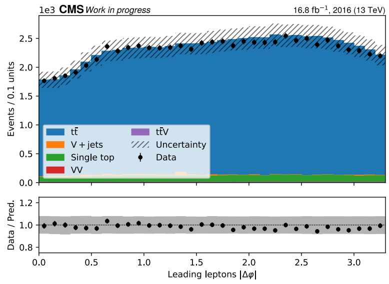
(e) eμ



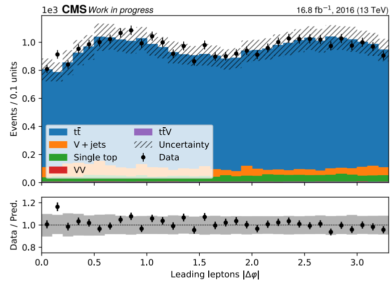
(f) μμ



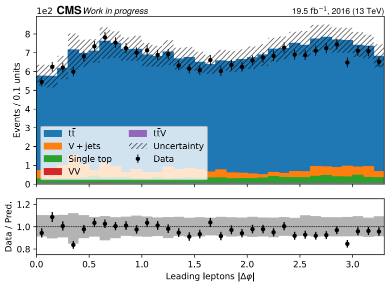
(g) ee



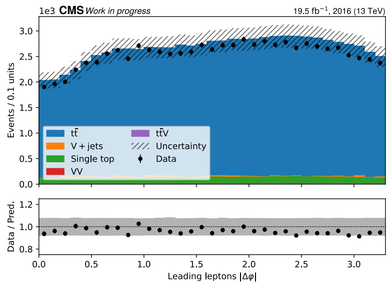
(h) eμ



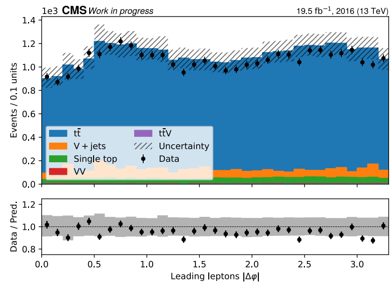
(i) μμ



(j) ee



(k) eμ



(l) μμ

Figure A.4: Difference between φ of the leptons for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).

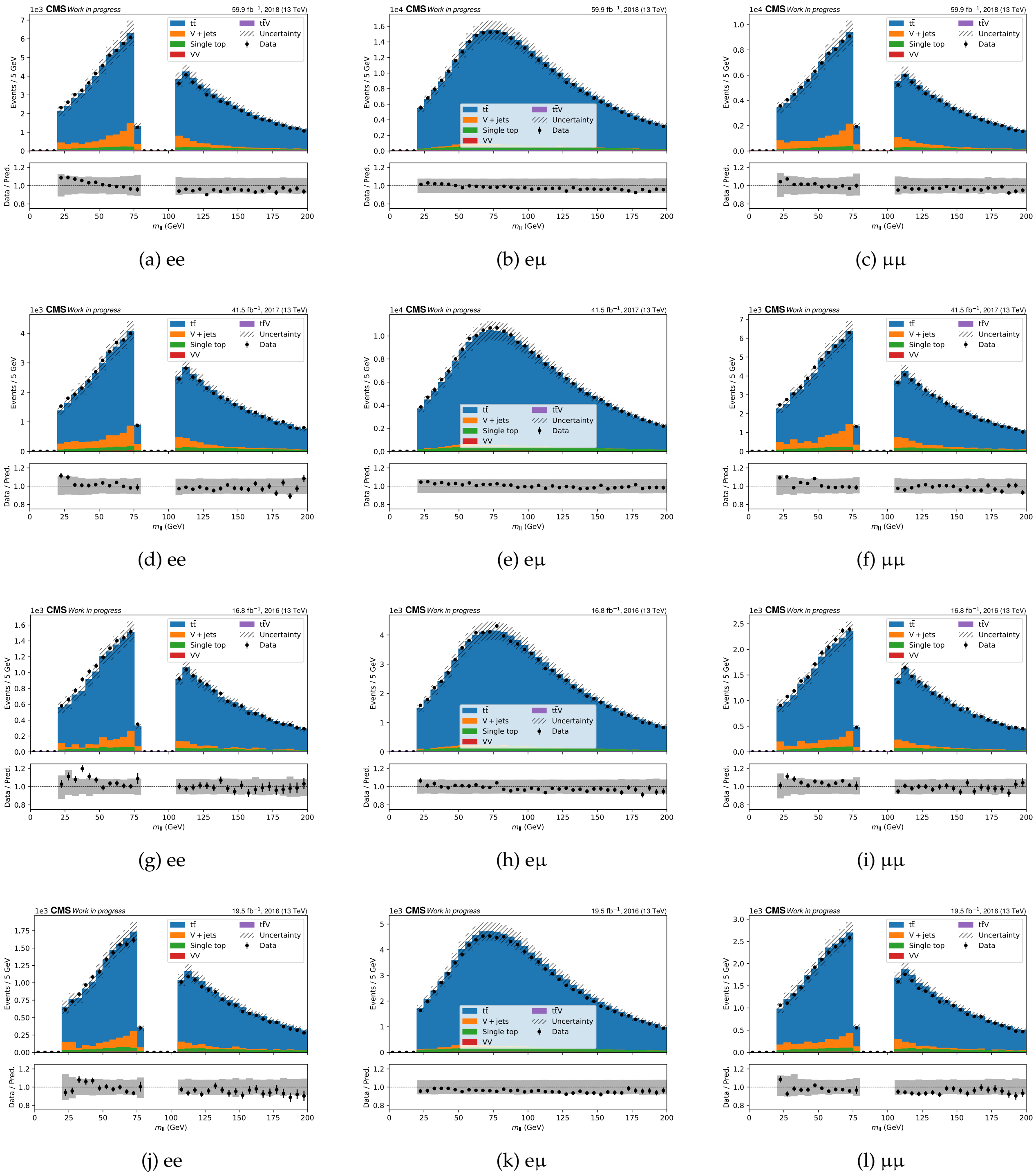
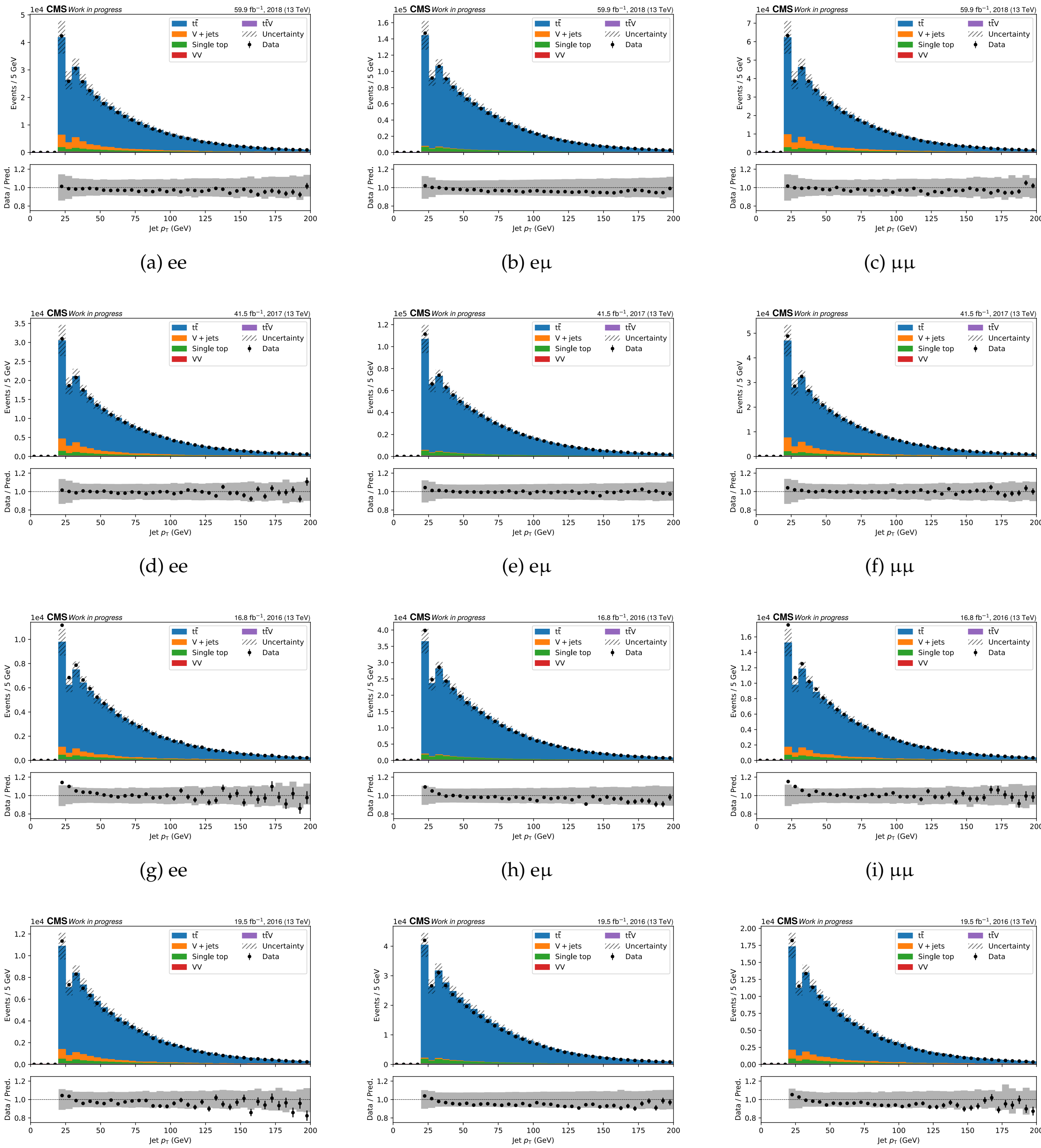


Figure A.5: Invariant mass of the leptons m_{ll} for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



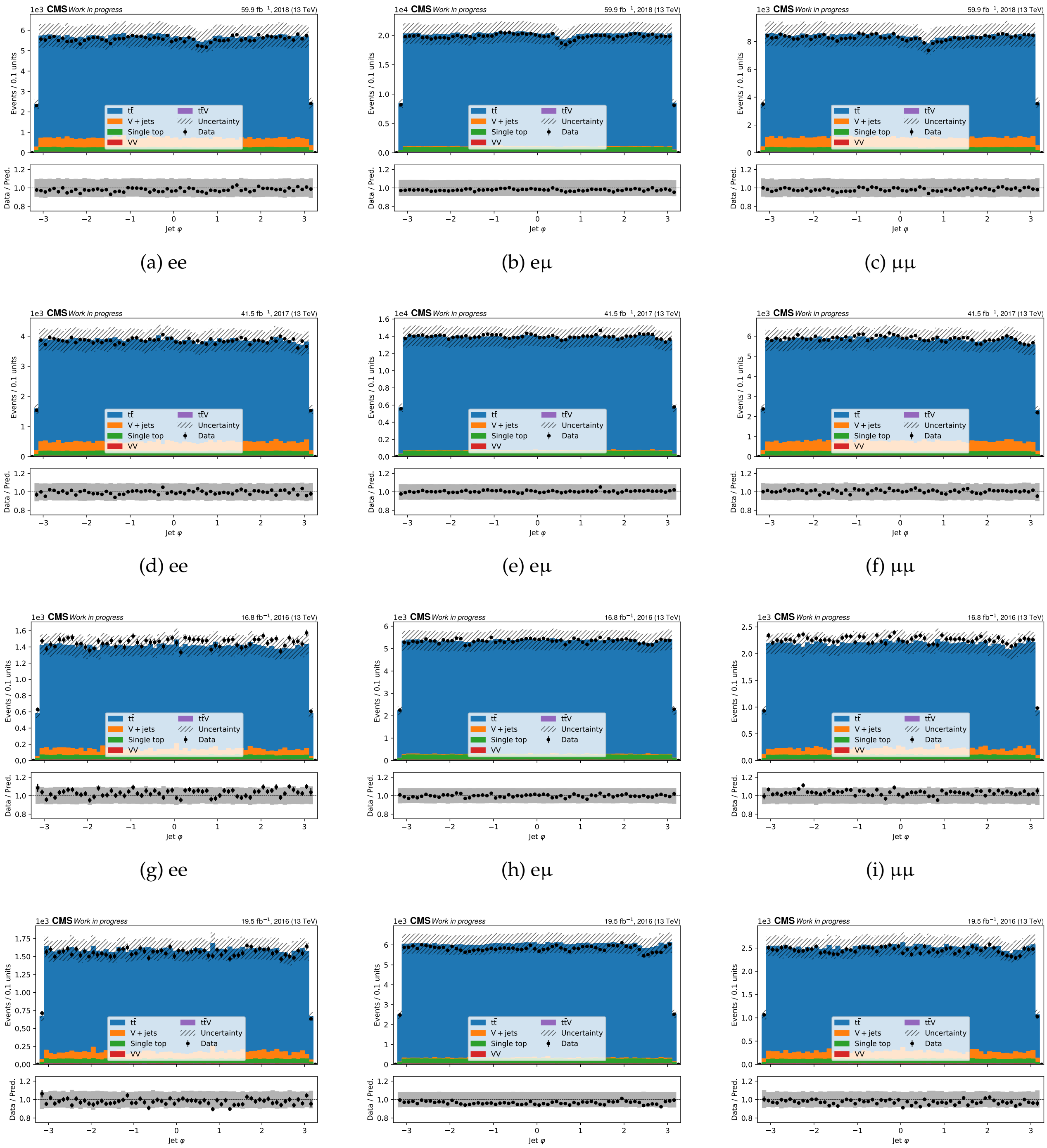
(a) ee (b) eμ (c) μμ

(d) ee (e) eμ (f) μμ

(g) ee (h) eμ (i) μμ

(j) ee (k) eμ (l) μμ

Figure A.6: p_T of the jets for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



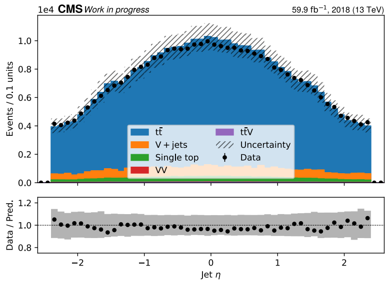
(a) ee (b) eμ (c) μμ

(d) ee (e) eμ (f) μμ

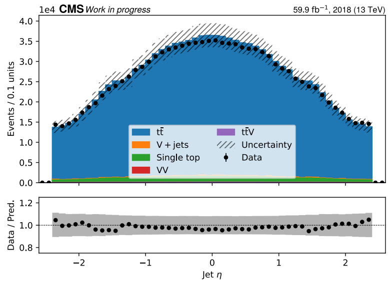
(g) ee (h) eμ (i) μμ

(j) ee (k) eμ (l) μμ

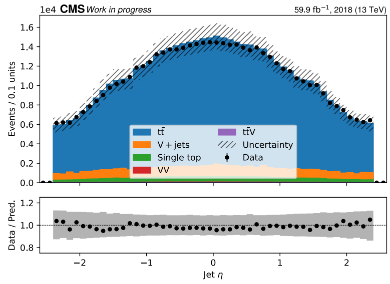
Figure A.7: φ of the jets for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



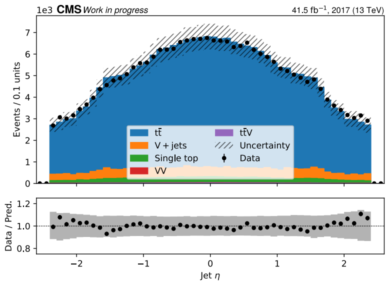
(a) ee



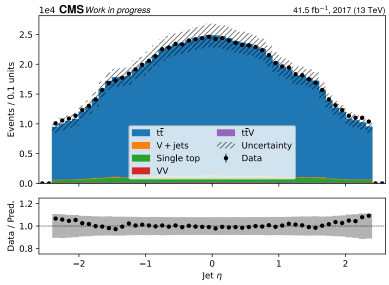
(b) eμ



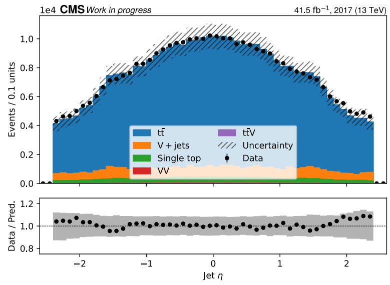
(c) μμ



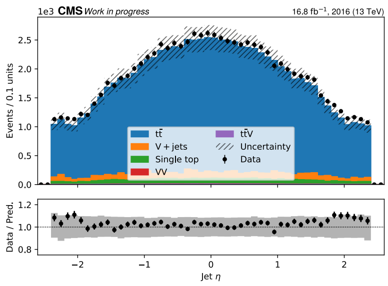
(d) ee



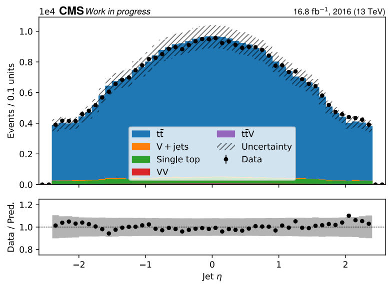
(e) eμ



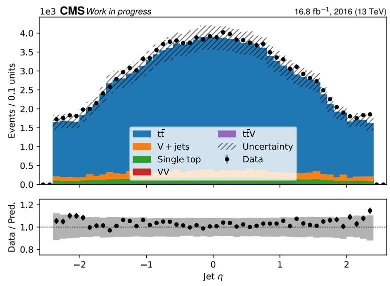
(f) μμ



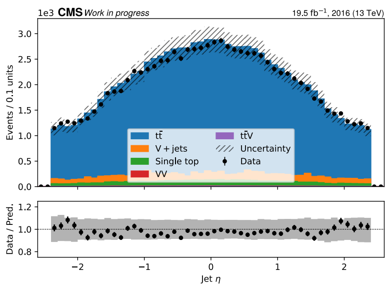
(g) ee



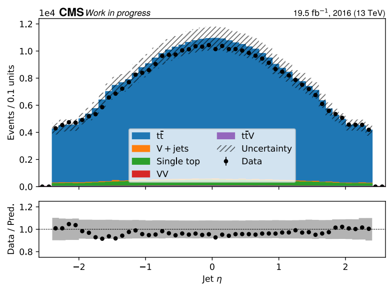
(h) eμ



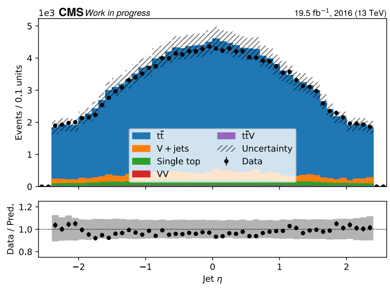
(i) μμ



(j) ee

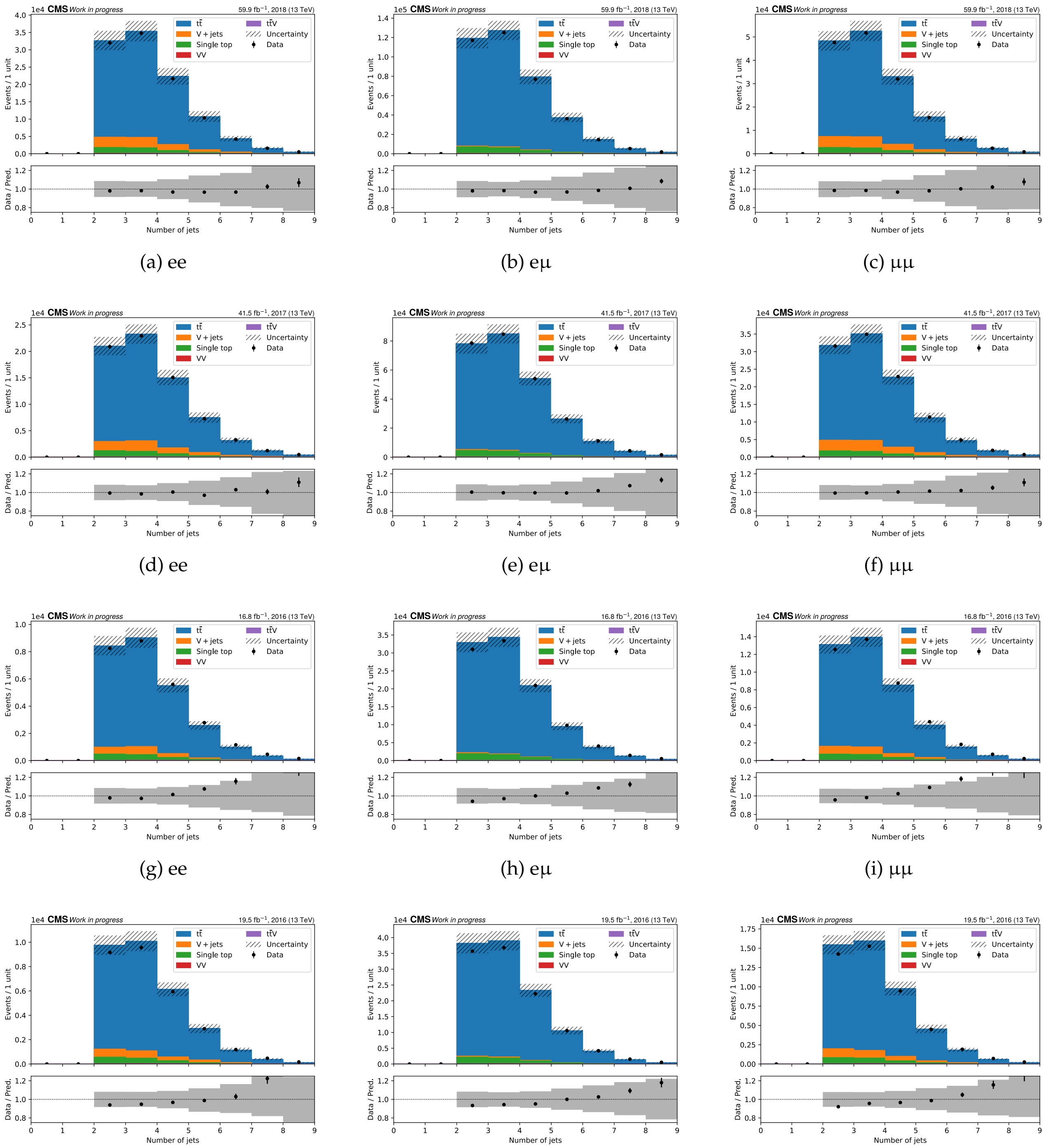


(k) eμ



(l) μμ

Figure A.8: η of the jets for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



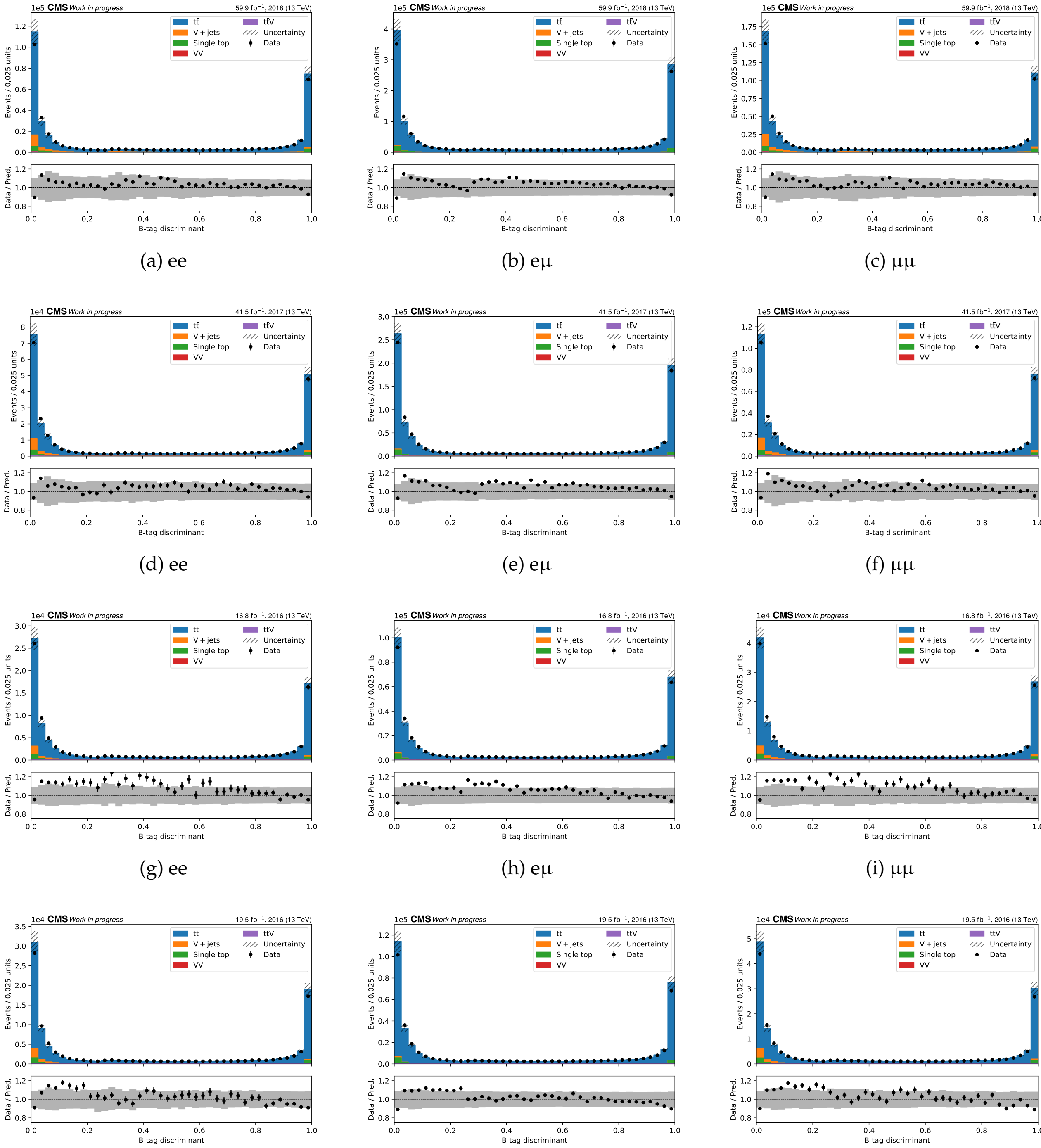
(a) ee (b) eμ (c) μμ

(d) ee (e) eμ (f) μμ

(g) ee (h) eμ (i) μμ

(j) ee (k) eμ (l) μμ

Figure A.9: Number of jets for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



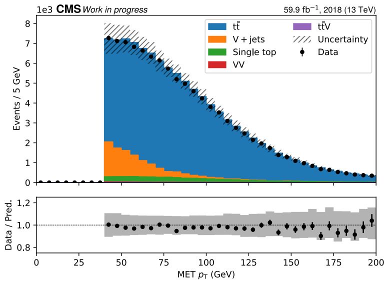
(a) ee (b) eμ (c) μμ

(d) ee (e) eμ (f) μμ

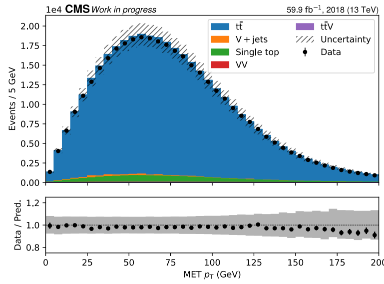
(g) ee (h) eμ (i) μμ

(j) ee (k) eμ (l) μμ

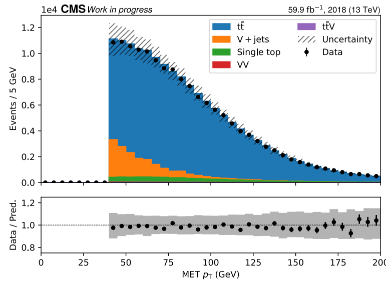
Figure A.10: Value of the bottom quark identification ("b-tag") discriminant of all jets for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



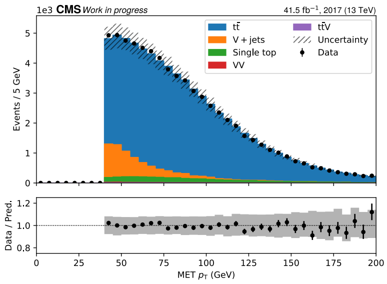
(a) ee



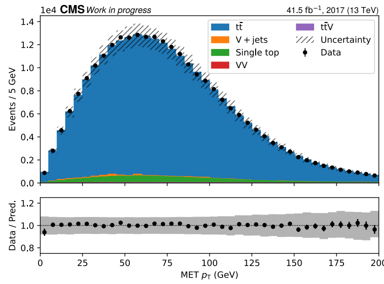
(b) eμ



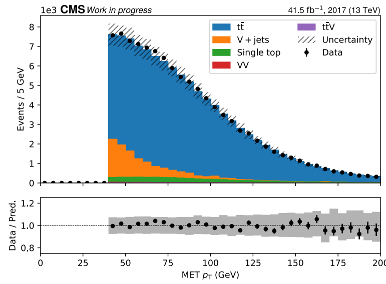
(c) μμ



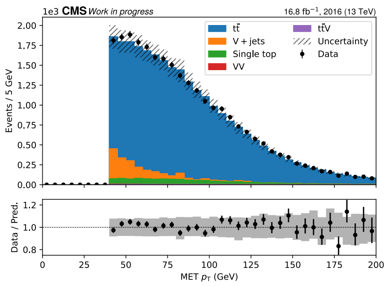
(d) ee



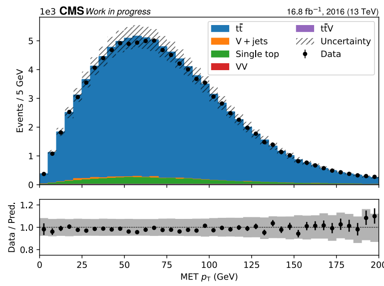
(e) eμ



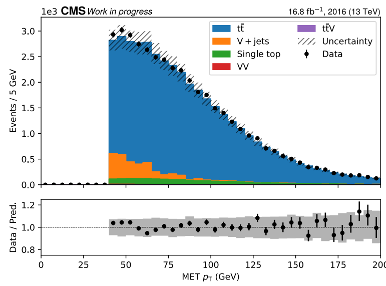
(f) μμ



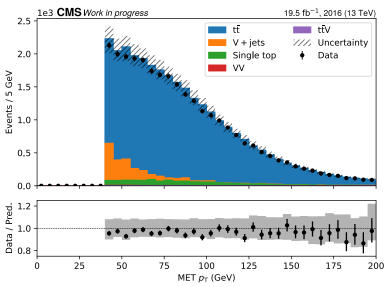
(g) ee



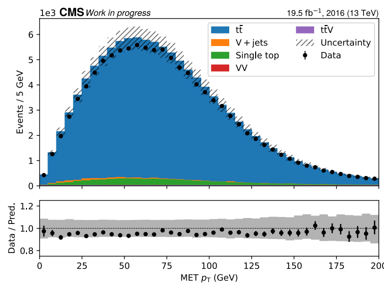
(h) eμ



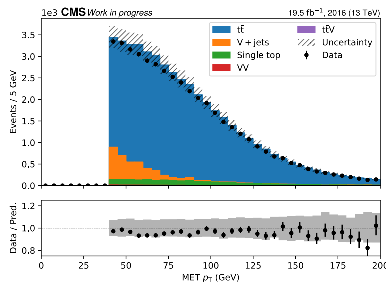
(i) μμ



(j) ee

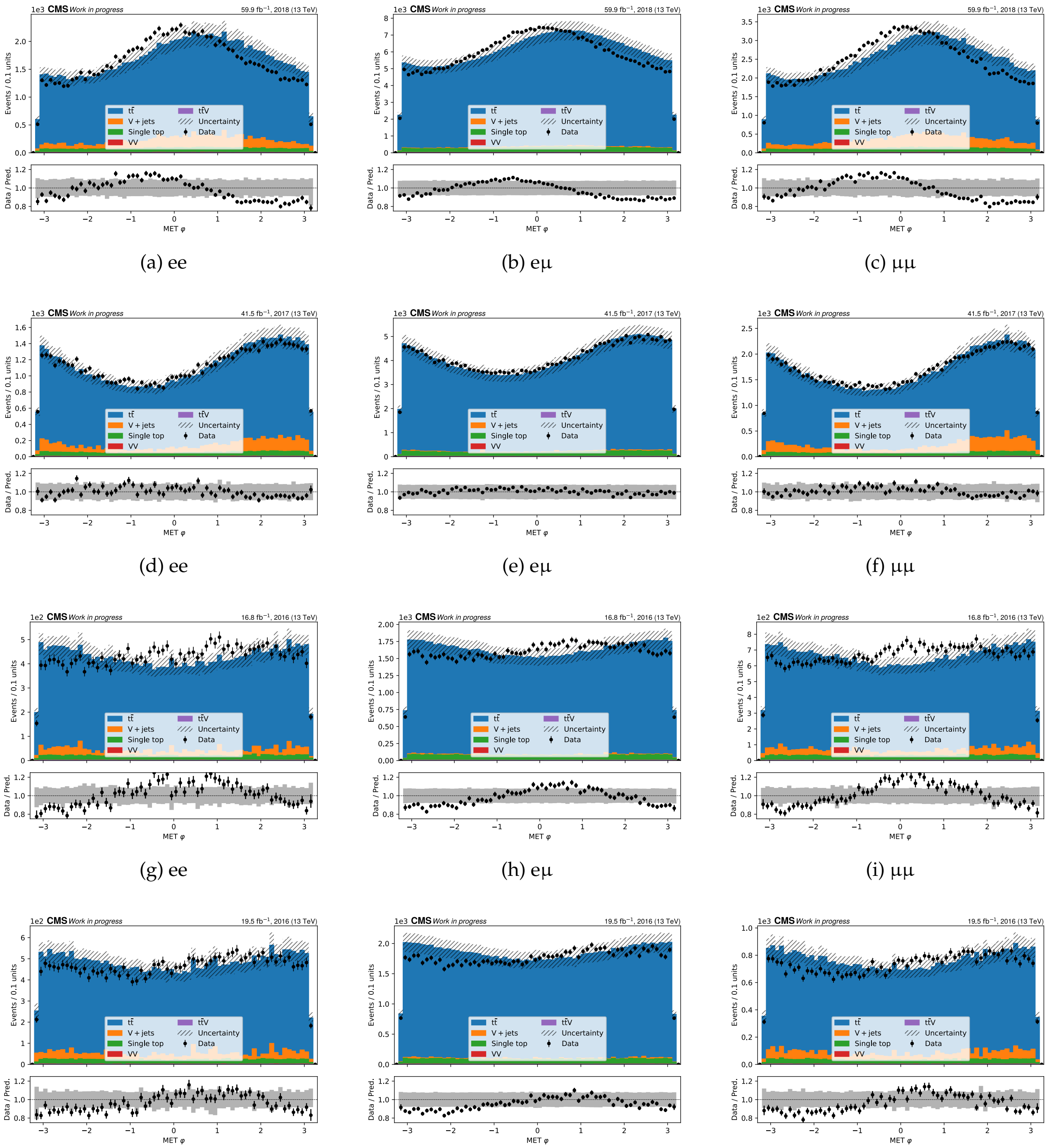


(k) eμ



(l) μμ

Figure A.11: p_T of the MET for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).



(a) ee (b) eμ (c) μμ

(d) ee (e) eμ (f) μμ

(g) ee (h) eμ (i) μμ

(j) ee (k) eμ (l) μμ

Figure A.12: φ of the MET for 2018 (first row), 2017 (second row), 2016post (third row) and 2016pre (fourth row).

Appendix B

Nuisance parameter impacts

The following figures display the sizes and impacts of the 150 leading nuisance parameters for A(600 GeV, 2.5 %) and H(600 GeV, 2.5 %). The middle panels show the change to the nuisance parameters after the fit. The right panels show the change induced to the POI g_{A} by varying the nuisance parameter within its standard deviation determined from the fit.

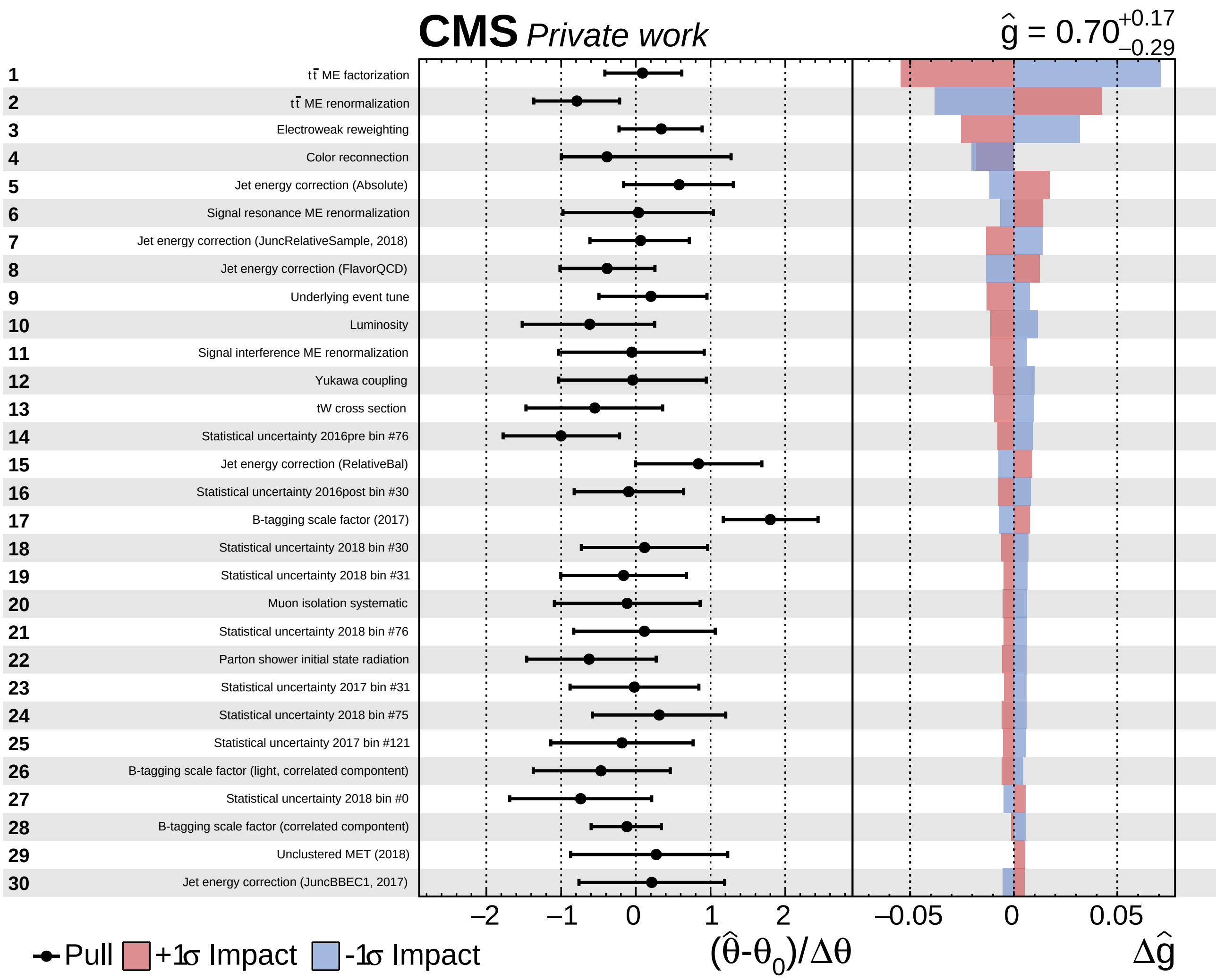


Figure B.1: Size of individual sources of uncertainties in form of nuisance parameters in the scenario A(600 GeV, 2.5 %) as determined by the fit.

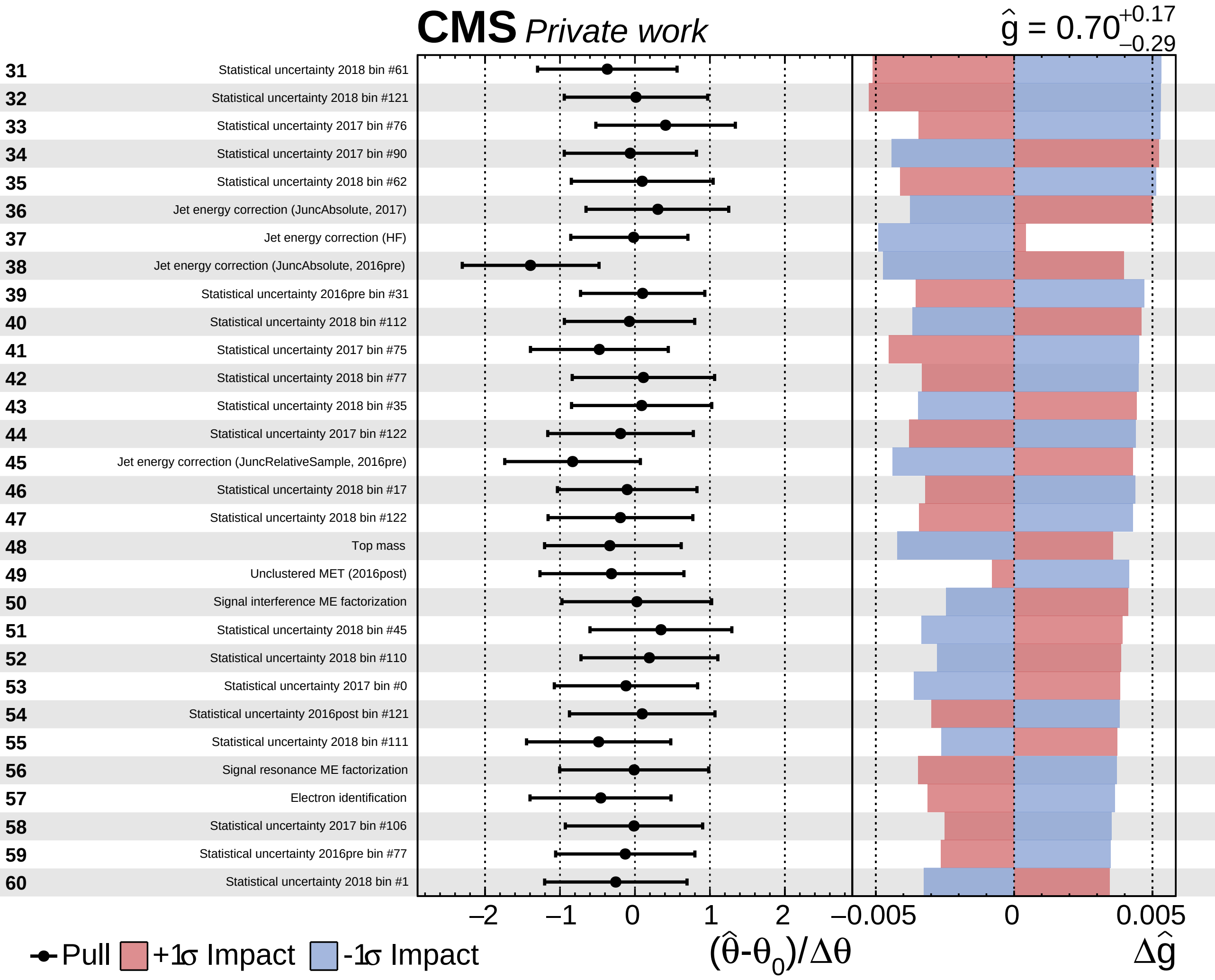


Figure B.2: Size of individual sources of uncertainties in form of nuisance parameters in the scenario A(600 GeV, 2.5 %) as determined by the fit.

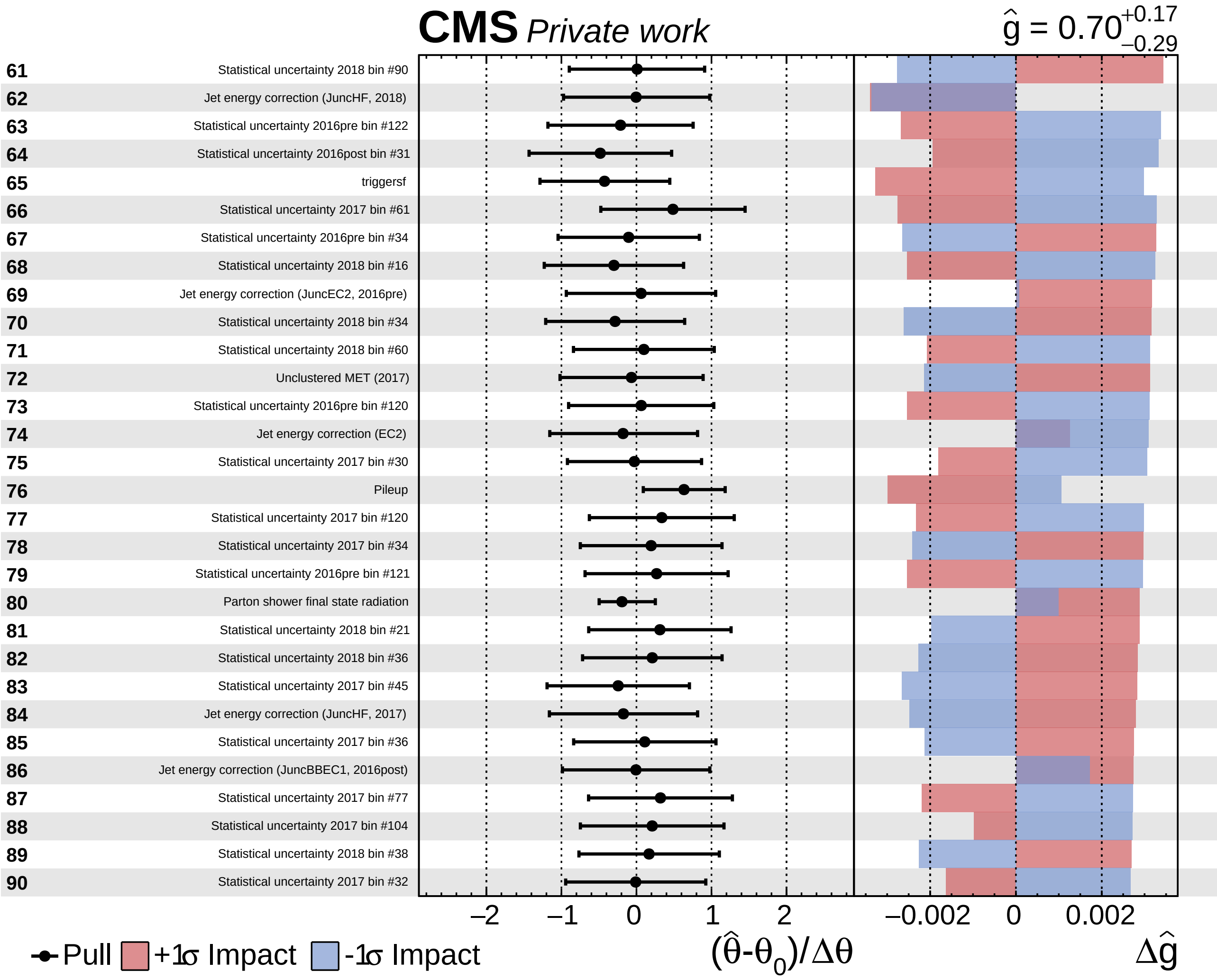


Figure B.3: Size of individual sources of uncertainties in form of nuisance parameters in the scenario A(600 GeV, 2.5 %) as determined by the fit.

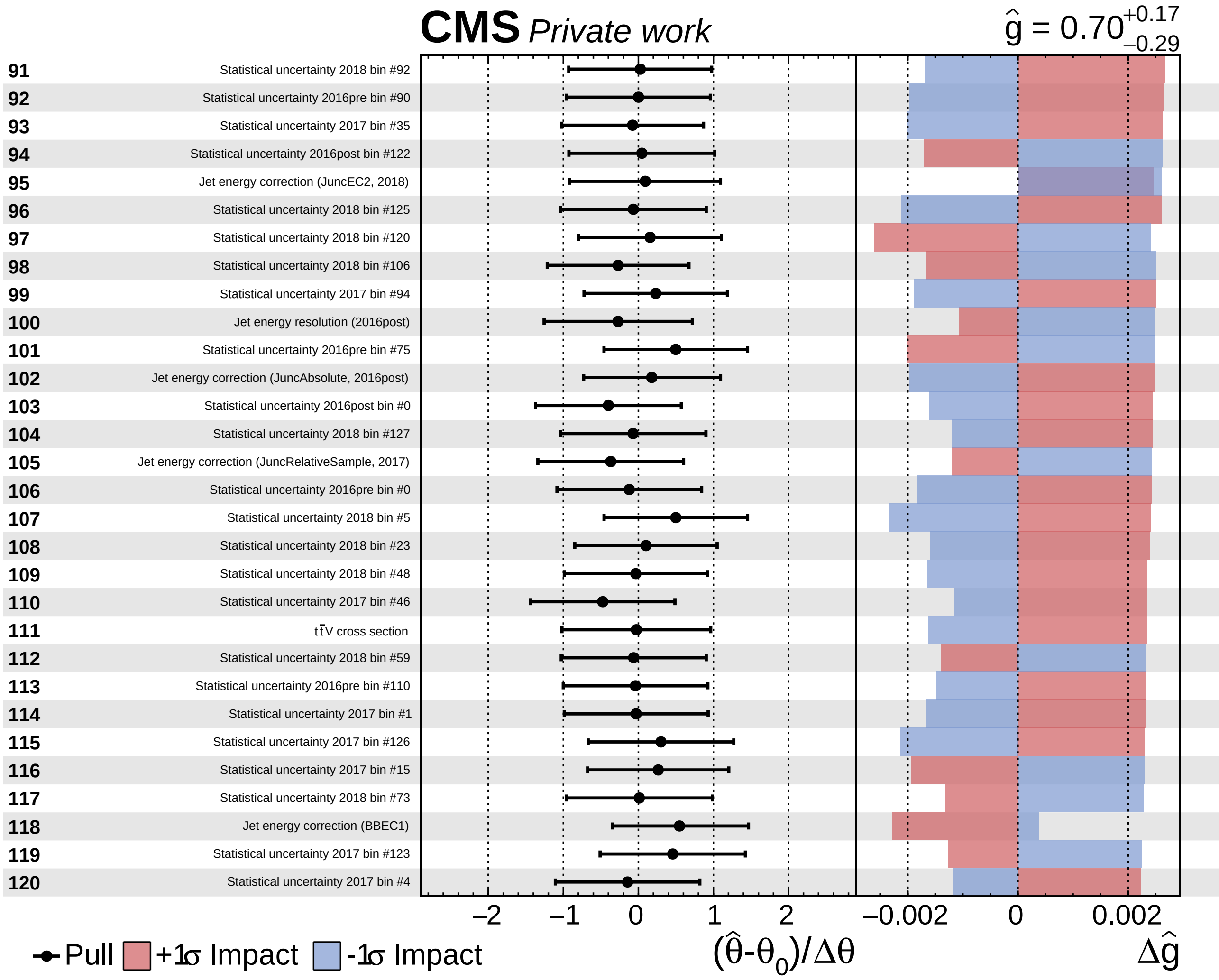


Figure B.4: Size of individual sources of uncertainties in form of nuisance parameters in the scenario A(600 GeV, 2.5 %) as determined by the fit.

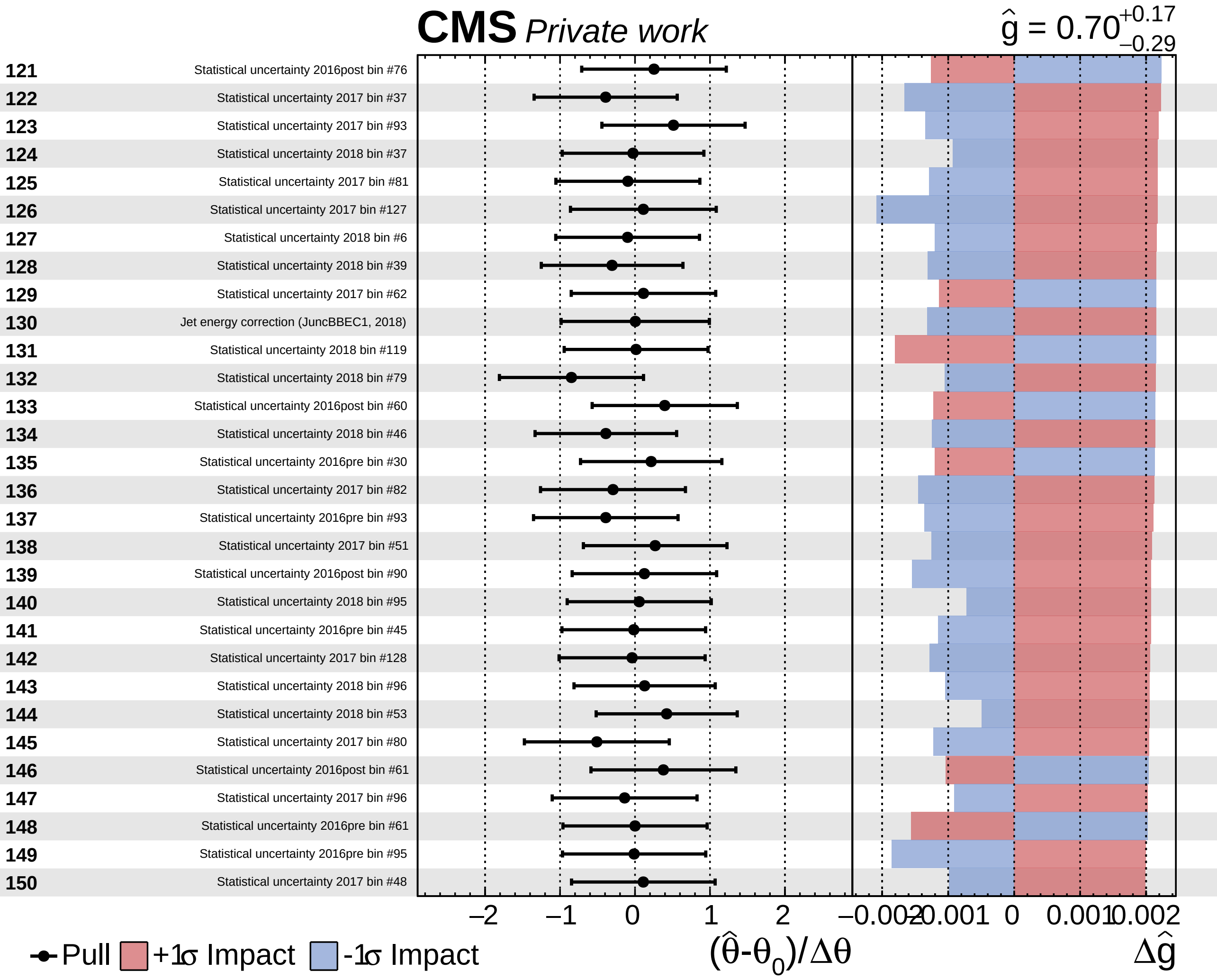


Figure B.5: Size of individual sources of uncertainties in form of nuisance parameters in the scenario A(600 GeV, 2.5 %) as determined by the fit.

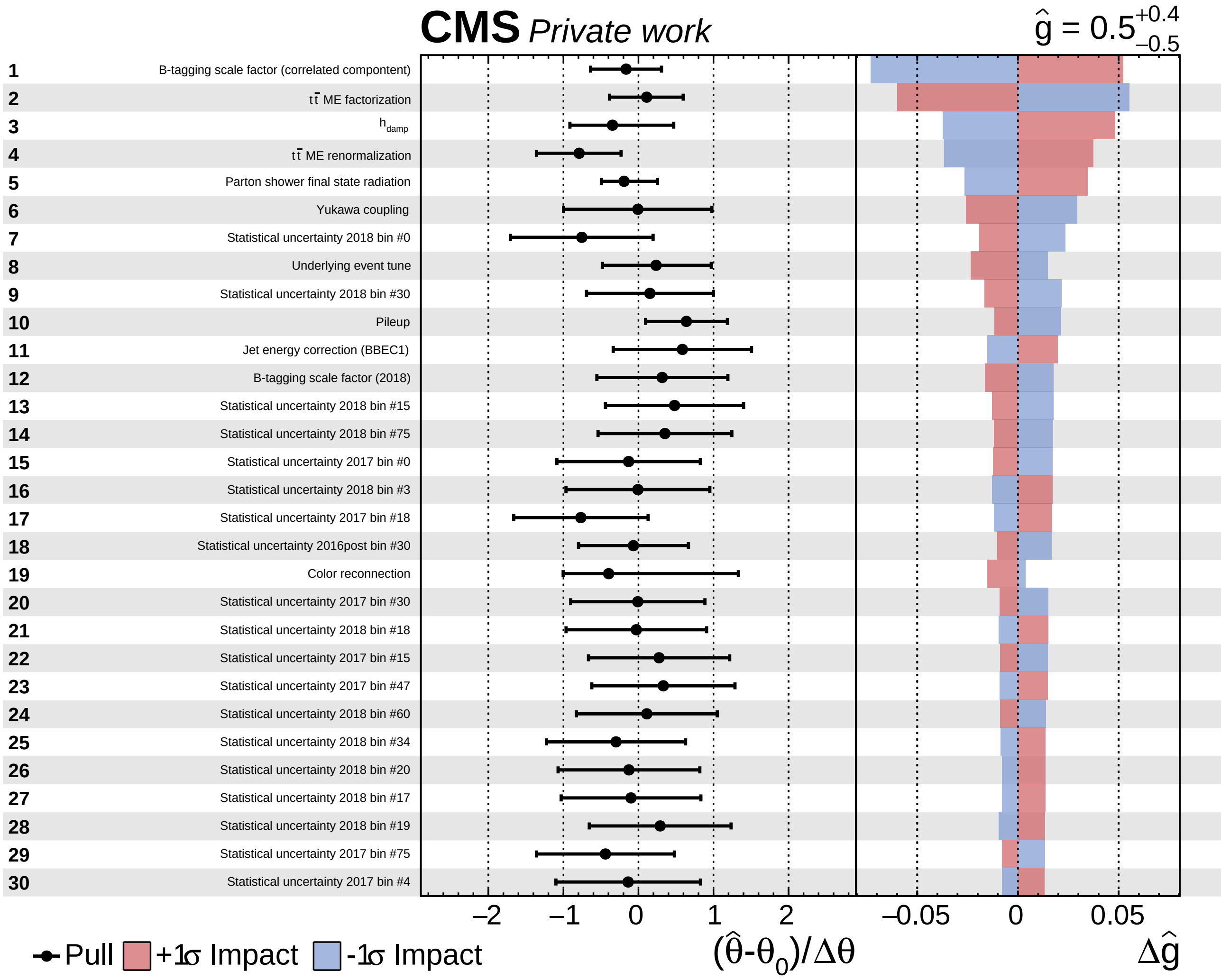


Figure B.6: Size of individual sources of uncertainties in form of nuisance parameters in the scenario H(600 GeV, 2.5 %) as determined by the fit.

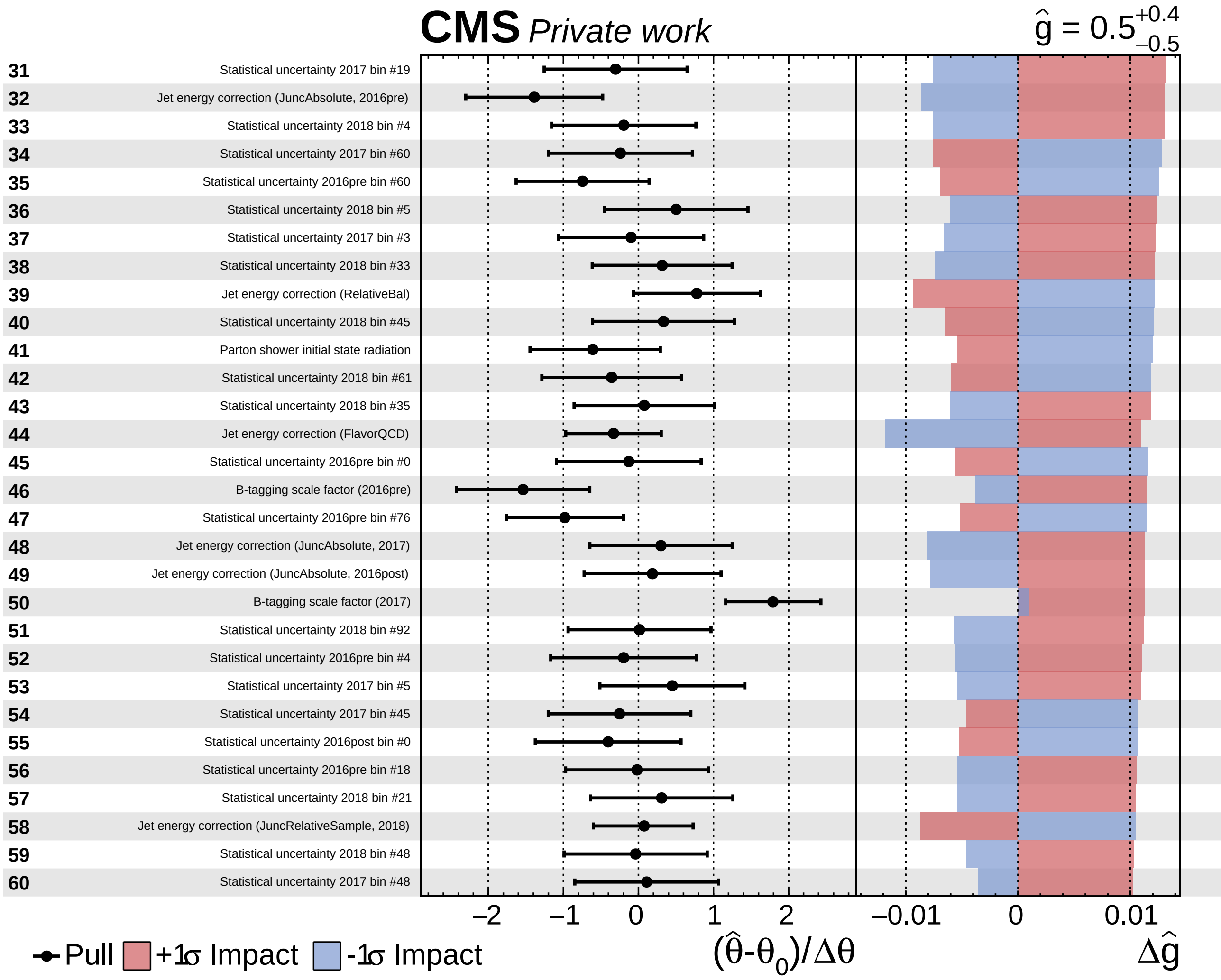


Figure B.7: Size of individual sources of uncertainties in form of nuisance parameters in the scenario H(600 GeV, 2.5 %) as determined by the fit.

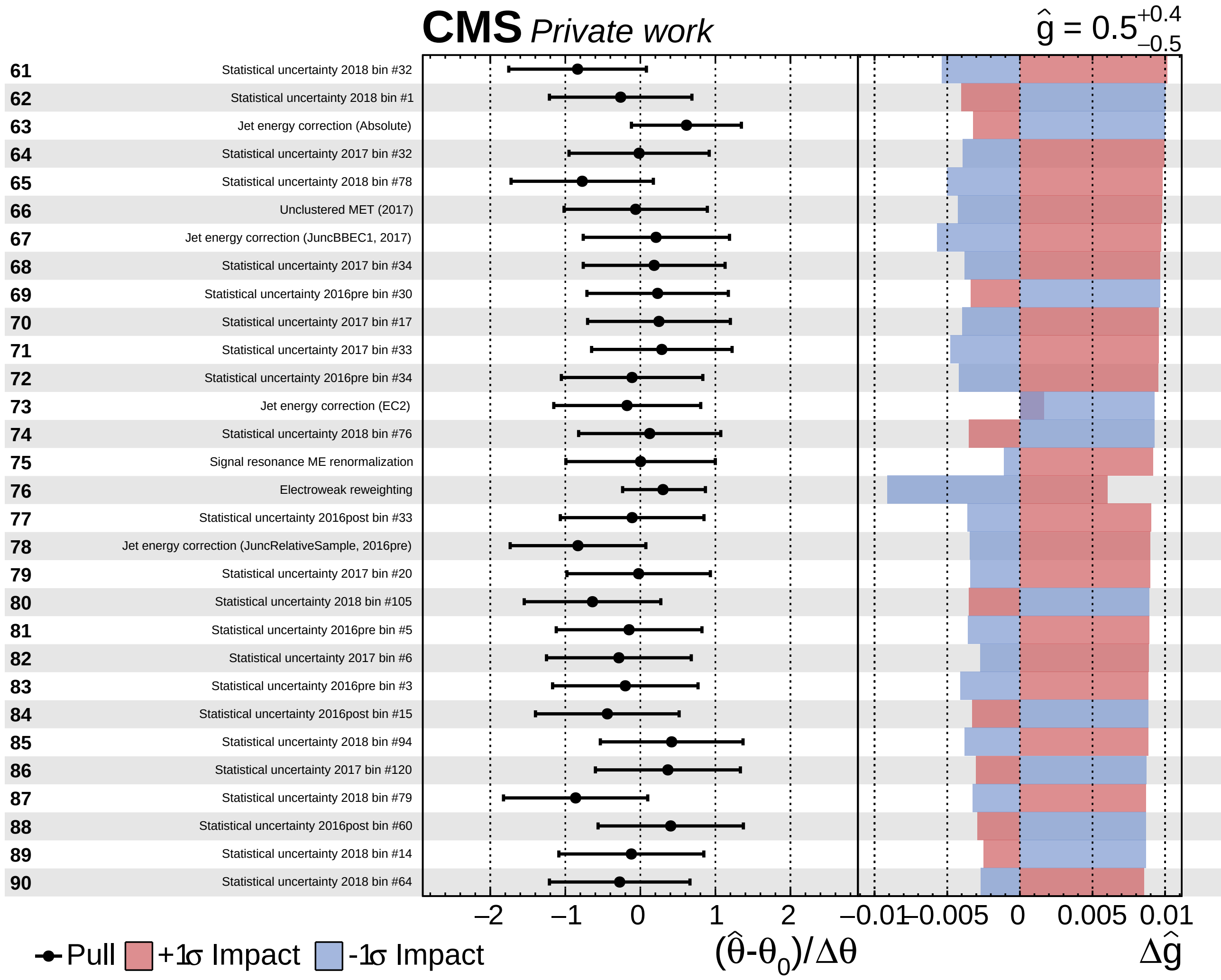


Figure B.8: Size of individual sources of uncertainties in form of nuisance parameters in the scenario $H(600\,\mathrm{GeV}, 2.5\,\%)$ as determined by the fit.

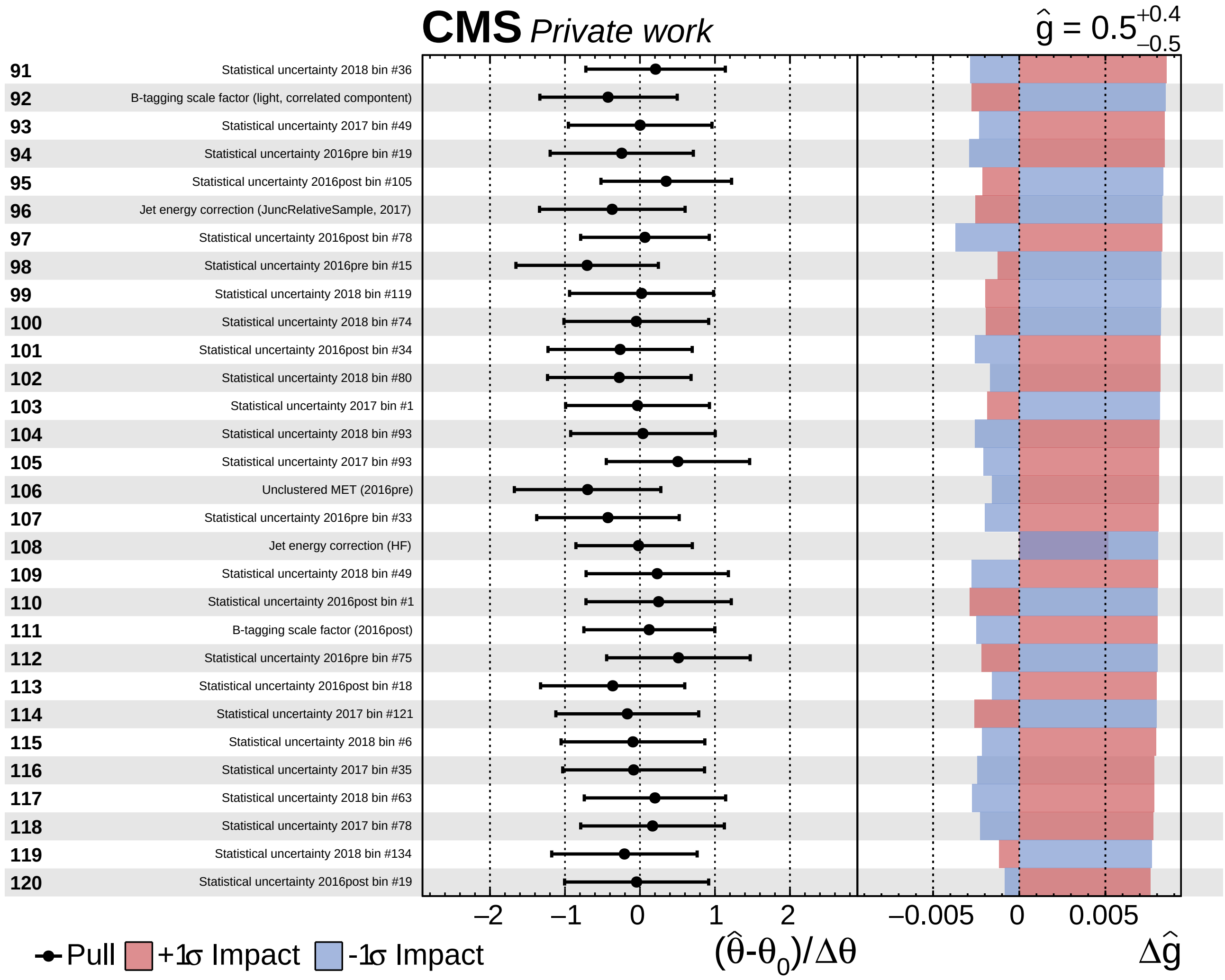


Figure B.9: Size of individual sources of uncertainties in form of nuisance parameters in the scenario H(600 GeV, 2.5 %) as determined by the fit.

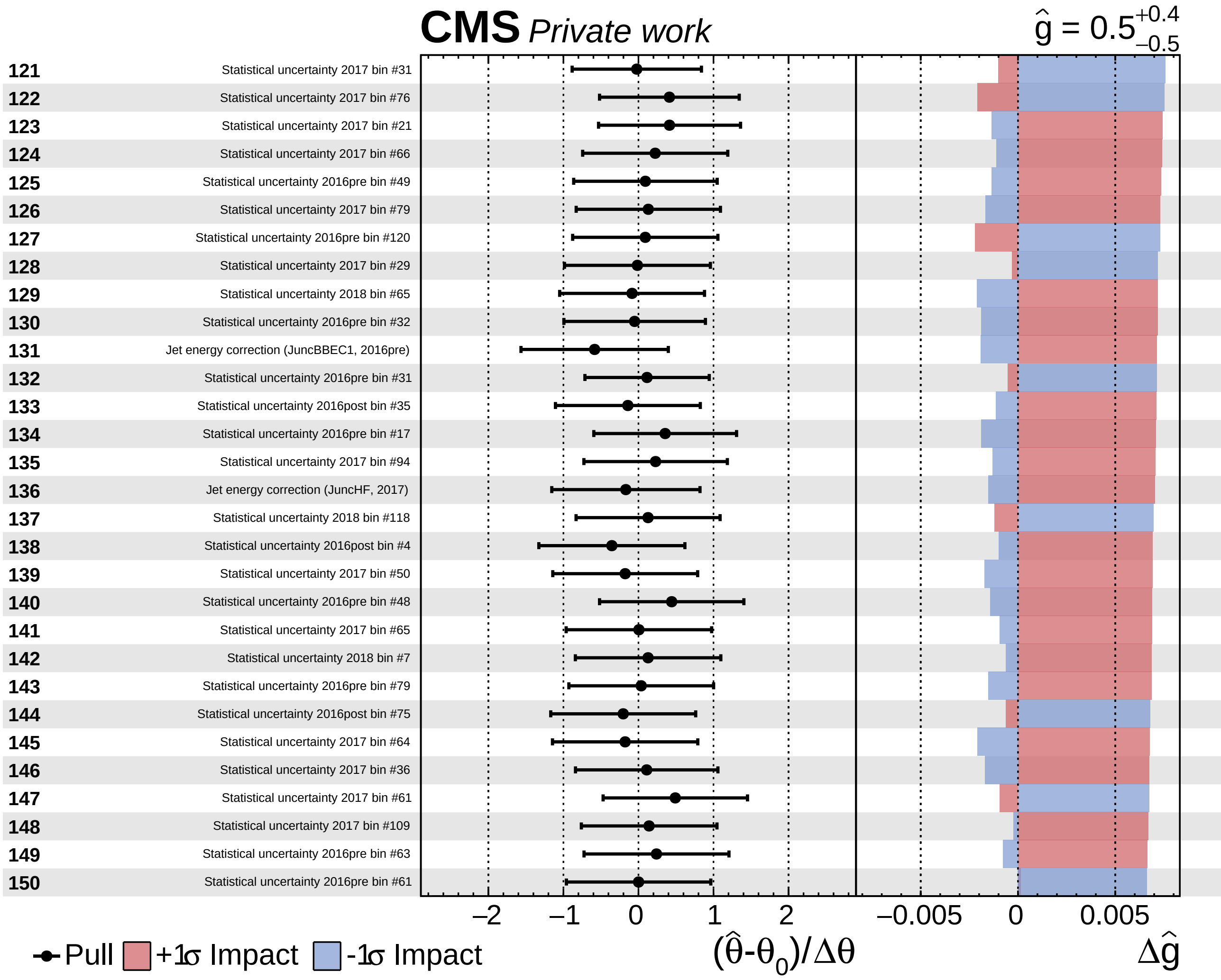


Figure B.10: Size of individual sources of uncertainties in form of nuisance parameters in the scenario H(600 GeV, 2.5 %) as determined by the fit.

Appendix C

Neural network input/target distributions

The following figures show the inputs and target distributions used to train the neural networks for the top quark reconstruction. The distributions for the SM simulation as well as for the crafted training set (CTS) are shown.

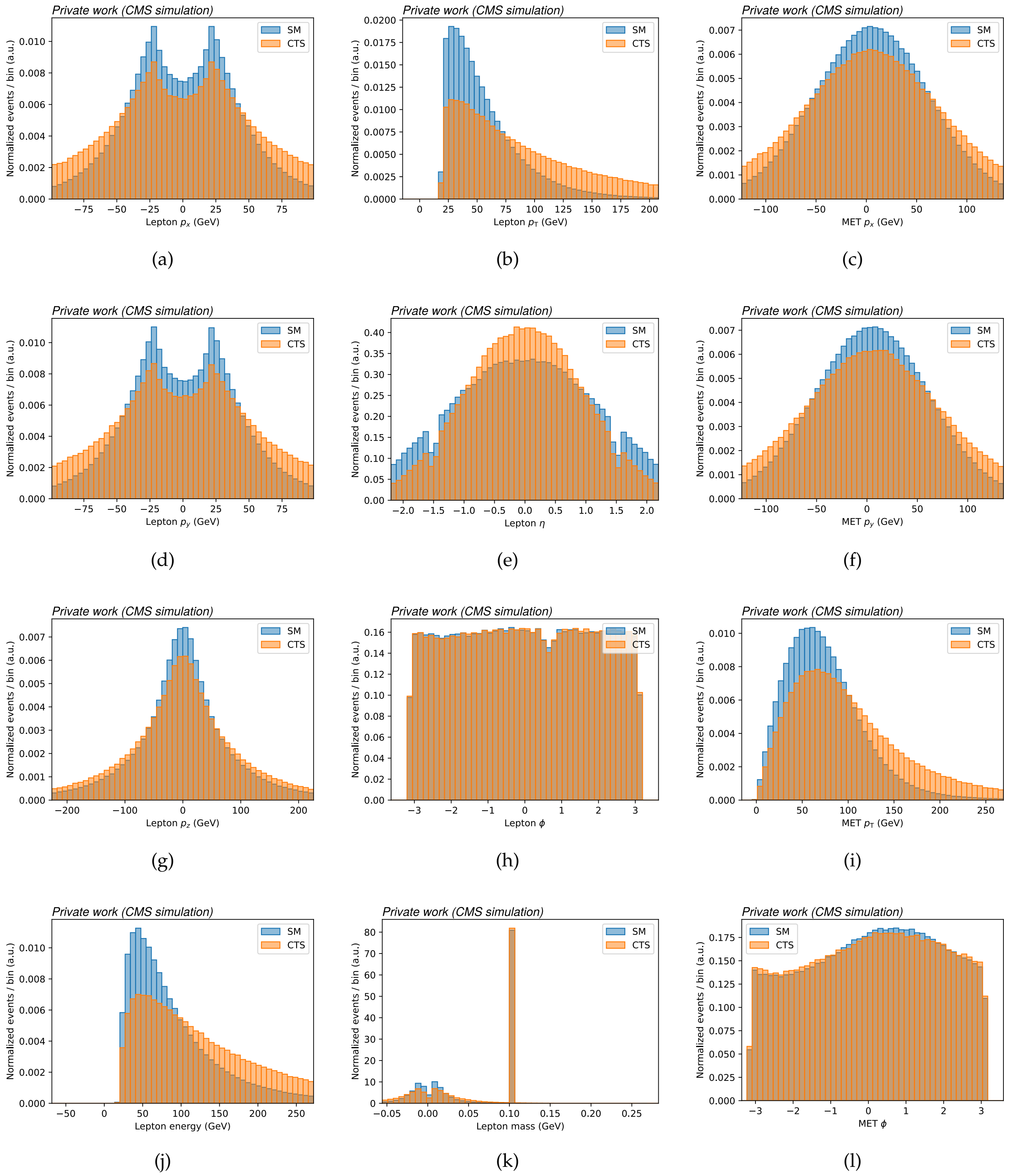


Figure C.1: Input distributions for the lepton and MET variables.

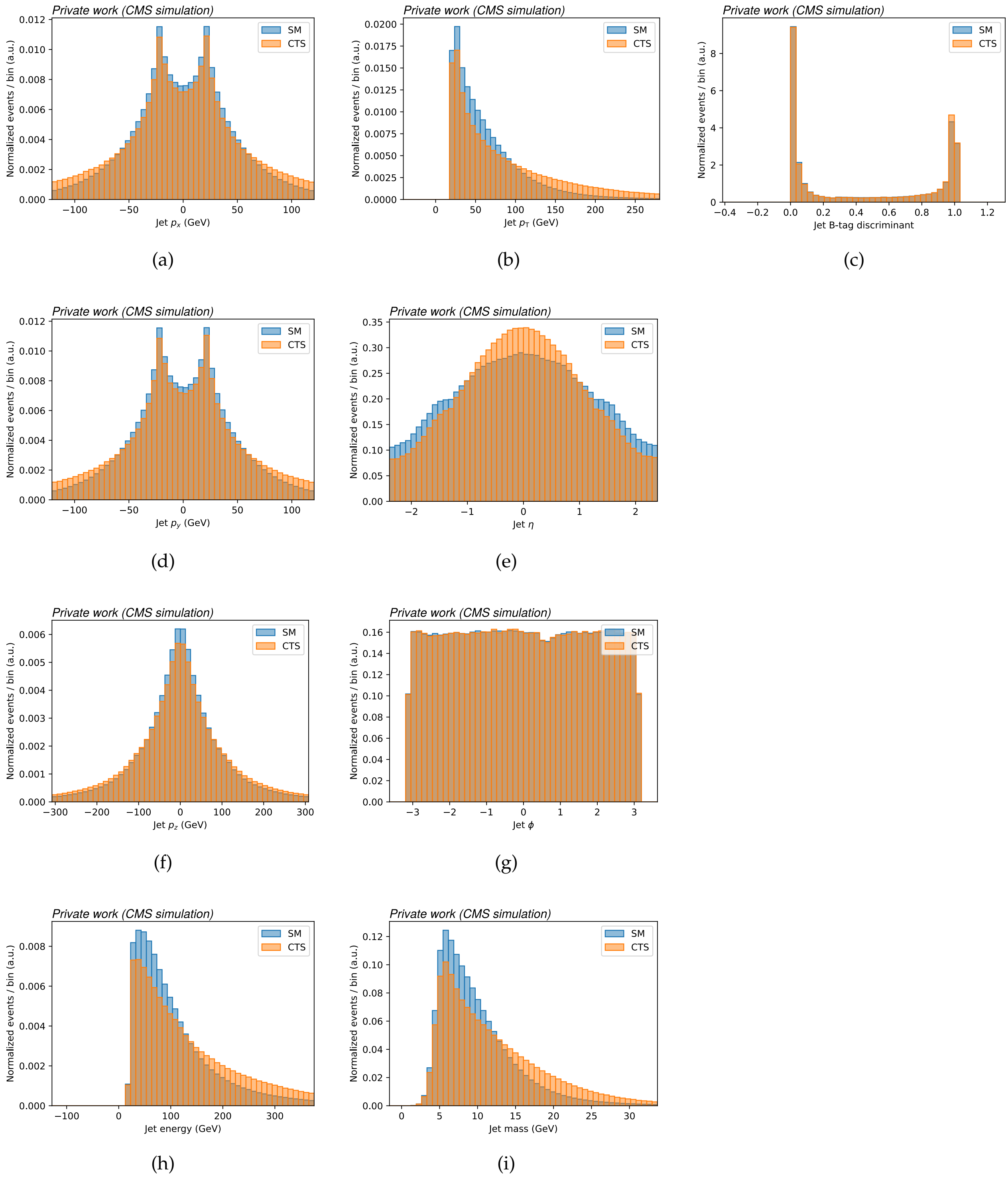


Figure C.2: Input distributions for the jet variables.

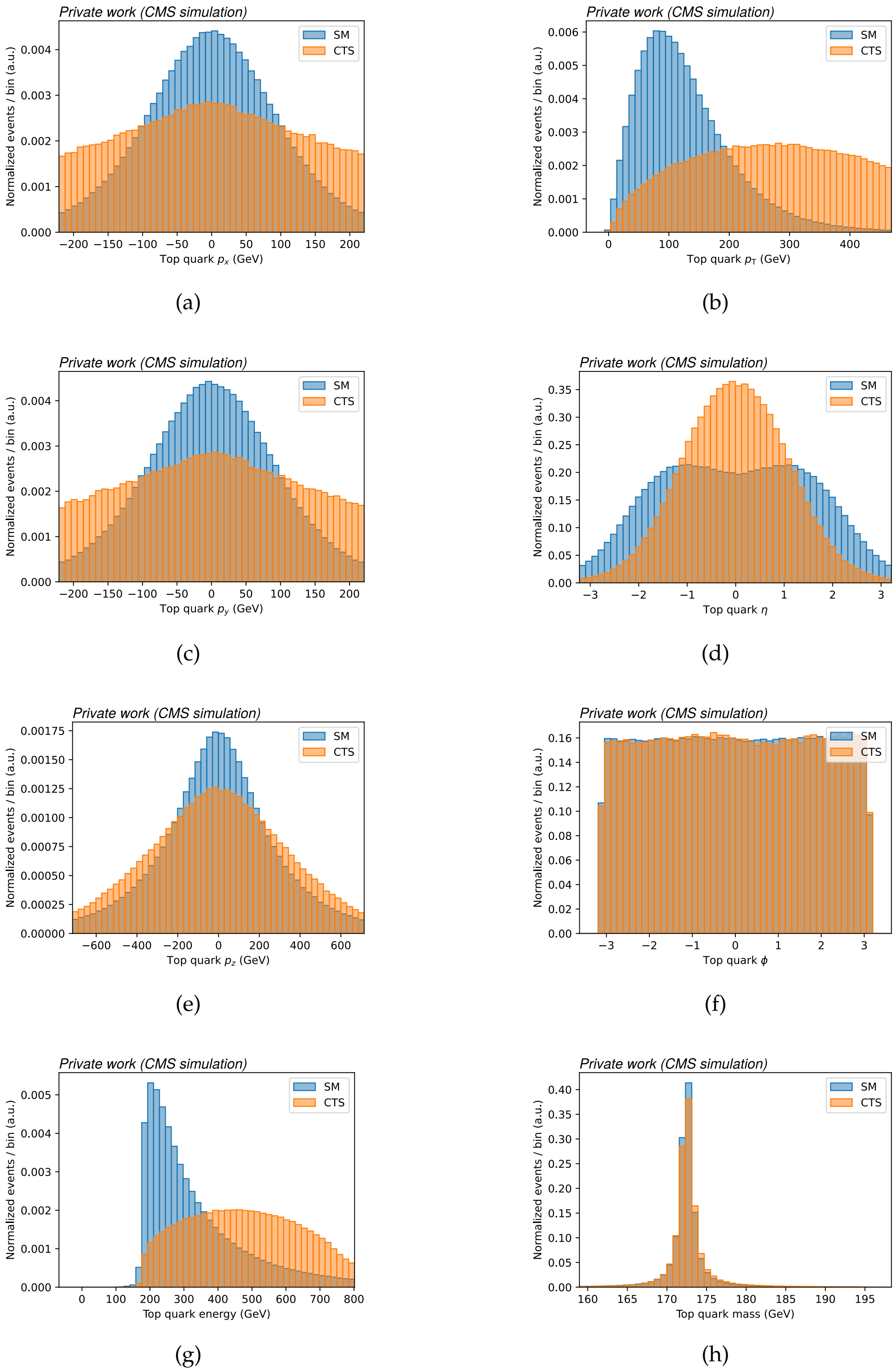


Figure C.3: Target distributions for the top quark variables.