Applications of Deep Neural Networks in a Top Quark Mass Measurement at the LHC

(Anwendung Neuronaler Netze in einer Topquarkmassenmessung am LHC)

Master Thesis in Physics

by

Torben Lange

(6444869)

(born: 06.06.1994)

angefertigt im

Institut für Experimentalphysik

vorgelegt der

Fakultät für Mathematik, Informatik und Naturwissenschaften

1. Gutachter : Prof. Dr. Peter Schleper

2. Gutachter : Jun.-Prof. Dr. Gregor Kasieczka

Hamburg, den 03.04.2018

1. Gutachter : Prof. Dr. Peter Schleper

2. Gutachter : Jun.-Prof. Dr. Gregor Kasieczka

Datum des Einreichens der Arbeit: 03.04.18

Zusammenfassung

In der vorliegenden Arbeit wird die Verwendung von Neuronalen Netzen für eine verbesserte Ereignisselektion in der Topquarkmassenmessung im myon+jets Endzustand für Ereignisse am CMS Experiment für den Run II des LHC mit einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV untersucht. Es wurde gezeigt, dass die Zusammensetzung der selektierten Ereignisse bezüglich Permutationstypen der Jetzuordnung in $t\bar{t}$ Ereignissen einen starken Einfluss auf die systematische Unsicherheit der Topquarkmassenanalyse hat. Eine Selektion basierend auf einem neuronalen Netzwerk, dass auf die Klassiefizierung von Ereignispermutationen des $t\bar{t}$ myon+jets Enzustandes trainiert wurde, konnte die systematische Unsicherheit der bisherigen Topquarkmassenmessung in diesem Kanal von etwa 630 MeV auf 560 MeV verbessern.

Abstract

In this analysis the usage of deep neural networks for an improved event selection for the top-quark-mass measurement in the $t\bar{t}$ muon+jets channel for events at the CMS experiment for the LHC run II with a center of mass energy $\sqrt{s} = 13$ TeV was investigated. The composition of the event selection with respect to different jet-assignment permutation types was found to have a strong influence on the systematic uncertainty of the top-quark-mass measurement. A selection based on the output of neural network trained on classifying event permutations of the $t\bar{t}$ muon+jets final state into these permutation types could then be used to improve the systematical uncertainty of the current mass measurement from a systematical uncertainty of around 630 MeV to 560 MeV.

Acknowledgments

Contents

1 Introduction

Particle physics is the science that describes the interaction of the fundamental particles, the building blocks of nature. Thus research in the area of particle physics gives us insight in processes of all scales of nature, from the interaction of the sub atomic quarks inside the atomic nuclei up to the long term evolution of the universe, cosmology. The Standard Model of particle physics is a consistent theory that describes todays knowledge in this field and seems to be in very good agreement with most of the experimental data gathered in the last decades. Its predictions of new particles are all satisfied since the discovery of the higgs boson in 2012 [23]. Nontheless there is still a lot that needs to be explained. The Standard Model gives no candidate for dark matter, which is motivated by astronomical and cosmological observation, it does not allow massive neutrinos, despite the fact that we observed neutrino oscillations [32]. It does not cover gravity. The internal structure of the Standard Model also raises some questions: Why do we have three generations of particles? Why are the fermion masses distributed over several orders of magnitude? Can we explain the hierarchy problem or the lack of CP violation?

All these questions are a good motivation for extending the Standard Model. Therefore one of the tasks of modern particle physics is to search for more deviations between nature and the Standard Model. In this thesis data from the CMS[1] experiment at CERN is used to measure the mass of the top-quark, the heaviest fundamental particle we found so far. The top-quark mass is an important parameter of the Standard Model of particle physics as well as models beyond. Therefore a precise measurement of this quantity is an important step to find and understand new physics. To improve the current measurement approaches, new methods are needed, therefore this thesis focuses on using machine learning methods in general and Deep Artificial Neural Networks (Deep ANN's) in particular. This "deep learning" is computational expensive but delivers a powerful tool when dealing with complicated classification or regression problems. The goal of this work was to adapt ANN techniques for selection purposes in the top- quark-mass analyses.

---

[1]**C**ompact **M**uon **S**olenoid

2 The Standard Model of Particle Physics

The Standard Model of particle physics (SM) is the foundation of modern particle physics. The following chapter gives a short overview of the main aspects of the theory based on [69], [48] and [45].

2.1 Introduction to the Standard Model

The SM is a quantum field theory with a $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry. A more detailed theoretical description can be found in subsection 2.1.2. The particle content and basic interactions are described in the following section 2.1.1

2.1.1 Particle Content and Interactions in the SM

The SM describes the interaction of the known fundamental particles via the fundamental forces, except gravity, in a quantum field theory. Figure (1) gives a good overview of the particle content so the SM. It can be divided in force mediating bosons with spin 1, fermions with spin $\frac{1}{2}$ as building blocks of nature and the famous higgs boson with spin 0. The higgs boson arises from the higgs mechanism that gives mass to the fundamental particles, as described in section 2.1.2. Depending on how they interact through the strong, weak and electromagnetic force, the fermions are further divided in quarks, charged and neutral leptons. The charges of a particle are quantum numbers that determine how a particle can interact with other particles. Each fermion comes with a partner, its anti particle with inverse charge with respect to the three forces, but otherwise same properties like the particle mass.

The charge most noticeable in everyday life is the electric charge. All SM fermions with the exception of the neutral leptons, the neutrinos, are electrically charged and can therefore interact via the electromagnetic force. The associated force carrier is the massless photon. Particles with electric charge interact with each other by exchanging photons. The strong force that holds atoms cores together is coupling to the color charge. There are three color charges: red, green and blue. Quarks are the only fermions holding color charge. The corresponding force carrier is the massless gluon. The gluon itself is charged, holding color and anti color and therefore also interacts with itself. Stable states are found to be always colorless. This means that bound states are either comprised of particles with canceling colors and anti colors (e.g a blue quark and an anti blue anti quark) or comprised of particles with all three colors (e.g. a red, a green and a blue quark). The weak force that is mainly responsible for particle decays couples to the so called weak isospin. The weak isospin of a particle directly connects to another particle property, the chirality. The chirality is defined mathematically by the response of a particle to the chirality projection operators $P_L = \frac{1}{2}(1 - \gamma^5)$ and $P_R = \frac{1}{2}(1 + \gamma^5)$. For particles with a velocity below the speed of light the chirality is a superposition of left handed and right handed states, depending on the helicity of the particle. For velocities $v \approx c$ helicity and chirality approach equality. Left handed particles correspond to a spin projection anti parallel to the particle mo-

mentum. In the weak force only left handed particles (and right handed anti particles) interact.
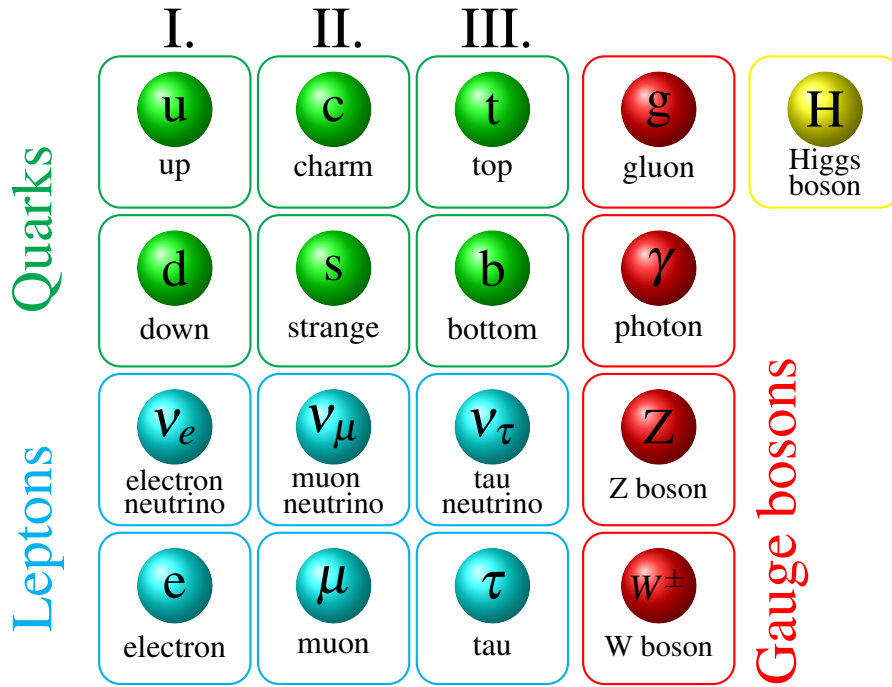
Figure 1: Elementary particles of the Standard Model [70].

The fermions are given by left handed dublets and right handed singlets. For the doublets, the third component of the weak isospin $I_3 = +\frac{1}{2}$ for one and $I_3 = -\frac{1}{2}$ for the other particle. Right handed singlets have $I_3 = 0$. This allows dividing the quarks in up-type and down type quarks and leptons in neutrinos and charged leptons with $I_3 = \pm\frac{1}{2}$ for their left handed variants. In the SM neutrinos are massless, traveling at the speed of light, therefore the SM excludes right handed neutrinos since they are also electrically and color neutral and would not interact with any other SM particle. The weak force has three force carriers in total. The Z boson is uncharged and interacts with all particles with weak charge. The two electrically charged W bosons allow another kind of interaction. They couple to the fermion doublets, allowing transitions between up and down type particles as well as neutrinos and charged leptons. Due to the non zero mass of the W and Z bosons, the range of the weak force is limited at low energy scales and weak interactions are suppressed. Fermions come in three different generations. For every fermion there are two other variants with the same quantum numbers with respect to the three forces but a different mass. The forces described by the SM with exception of the weak force conserve additional quantum numbers, that define the flavour of a particle given by the corresponding generation and weather it is up or down type. Only the weak force allows transitions between different flavors and generations. An overview of all fundamental particles and their quantum numbers can be found in table 1.

Table 1: List of all particles in the SM with their masses and quantum numbers [52].

		name	symbol	mass $\left[\frac{\text{GeV}}{c^2}\right]$	electric charge [e]	color charge	weak isospin
		fermions spin=$\frac{1}{2}$					
up-artige	quarks	up quark (1st. generation)	u	$0.0022^{+0.0006}_{-0.0004}$	+2/3	red/green/blue	+1/2, 0
		charm quark (2nd. generation)	c	1.27 ± 0.03	+2/3	red/green/blue	+1/2, 0
		top quark (3rd. generation)	t	174.2 ± 1.4^a	+2/3	red/green/blue	+1/2, 0
down like	quarks	down quark (1st. generation)	d	$0.0047^{+0.0005}_{-0.0004}$	-1/3	red/green/blue	-1/2, 0
		strange quark (2nd. Generation)	s	$0.096^{+0.008}_{-0.004}$	-1/3	red/green/blue	-1/2, 0
		bottom quark (3rd. Generation)	b	$4.18^{+0.04}_{-0.03}$	-1/3	red/green/blue	-1/2, 0
neutri-	nos	electron neutrino (1st. generation)	ν_e	0^b	0	-	+1/2
		muon neutrino (2nd. generation)	ν_μ	0	0	-	+1/2
		τ neutrino (3rd. generation)	ν_τ	0	0	-	+1/2
charged	leptons	electron (1st. generation)	e	$5.1099 \cdot 10^{-4}$	-1	-	-1/2, 0
		muon (2nd. generation)	μ	0.1057	-1	-	-1/2, 0
		τ lepton (3rd. generation)	τ	1.777	-1	-	-1/2, 0
		gauge bosons spin=1					
bosons		photon (electromagnetic force)	γ	0	0	-	0
		W^+-Boson (weak force)	W^+	80.385 ± 0.015	+1	-	+1
		W^--Boson (weak force)	W^-	80.385 ± 0.015	-1	-	-1
		Z-Boson (weak force)	Z^0	91.1876 ± 0.0021	0	-	0
		Gluon (strong force)	g	0	0	color + anti color	0
		higgs boson spin=0					
		higgs boson	h	125.09 ± 0.24	0	-	0

[a]The pole mass of the top quark. "pole mass"

[b]Massles in the SM.

2.1.2 Electroweak unification and gauge theory of the SM

The phenomenology of the SM described in the previous subsection arises from the equations of motion received from the lagrangian density of the SM [69, 60]. The lagrangian of the SM contains the kinetic and potential energy terms of the different fermions. For the SM the lagrangian density is required to be invariant under transformation of a SU(3)×SU(2)×U(1) symmetry. This requires the introduction of gauge fields that will resemble the force carriers of the electromagnetic, the weak and the strong force. Interaction terms with those fields lead to the description of the respective forces. Finally the higgs mechanism from which the higgs boson arises is introduced to allow mass terms for the introduced gauge fields without breaking the gauge invariance. In the following the composition of the SM lagrangian density is motivated. To start we can look at the U(1) symmetry group. The kinetic and mass terms of the fermionic fields Ψ would result in the following lagrangian density

$$\mathscr{L}_{\text{fermion}} = \bar{\psi}(i\gamma^\mu \partial_\mu)\psi - m\bar{\psi}\psi \,.$$

With γ^μ the gamma or dirac matrices and m the fermions mass. A local U(1) phase transformation would change the dirac fields in the form of

$$\psi(x) \to \psi'(x) = e^{ig\frac{Y}{2}\chi(x)}\psi(x) \ .$$

With a coupling constant g, the Hypercharge Y and a local phase $\chi(x)$. Now we would like to require that the lagrangian density is invariant under local phase transformations of the U(1) symmetry group

$$\mathscr{L}' = \mathscr{L} \ .$$

If we use the U(1) phase transformation $\mathscr{L}$ is transformed to

$$\mathscr{L} \to \mathscr{L}' = ie^{-ig\frac{Y}{2}\chi}\bar{\psi}\gamma^\mu[e^{ig\frac{Y}{2}\chi}\partial_\mu\psi + ig\frac{Y}{2}(\partial_\mu\chi)e^{ig\frac{Y}{2}\chi}\psi] - me^{-ig\frac{Y}{2}\chi}\bar{\psi}e^{ig\frac{Y}{2}\chi}\psi$$

$$= \mathscr{L} - g\frac{Y}{2}\bar{\psi}\gamma^\mu(\partial_\mu\chi)\psi \ .$$

To fulfill the requirement of local gauge invariance, gauge fields are introduced which introduce new terms that cancel the deviations between $\mathscr{L}$ and $\mathscr{L}'$. For the U(1) symmetry we can simply introduce a new field B_μ and replace the partial derivative in the lagrangian density with the so called covariant derivative

$$\partial_\mu \to D_\mu = \partial_\mu + ig\frac{Y}{2}B_\mu$$

where B_μ changes under the U(1) local phase transformation like

$$B_\mu \to B'_\mu = B_\mu - \partial_\mu\chi \ .$$

If we look at the covariant derivative we see that a new interaction term is introduced:

$$g\frac{Y}{2}\gamma^\mu B_\mu\psi$$

This term will later describe the electromagnetic interactions and B_μ will, after a mixing with another gauge field, represent the photon. Similar to B_μ, other gauge fields are introduced to allow for invariance of the lagrangian density under local phase transformations of the SU(2) and SU(3) symmetry groups. Those fields then together with B^μ represent the Z and the W bosons as well as the gluons. It is easily possible to write down a invariant kinetic term for the B_μ

$$\mathscr{L}_{EM} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$$

with $F^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu$. But if we try to introduce a mass term for the B_μ we encounter a problem:

$$\frac{1}{2}m^2 B_\mu B^\mu \to \frac{1}{2}m^2(B_\mu - \partial_\mu\chi)(B^\mu - \partial^\mu\chi) \neq \frac{1}{2}m^2 B_\mu B^\mu$$

The introduced mass term is not invariant under transformation. This is not a problem of electromagnetism since the photon is massless but similar terms arise for the fields of the weak force even though the Z and W bosons are measured as massive particles. This problem is solved by introducing the higgs mechanism to the SM. This is done in the Salam-Weinberg model where electromagnetic and weak force are unified to the electroweak force. The higgs field of the SM consists of a doublet of two complex scalar fields

$$\phi = \begin{pmatrix} \phi+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} .$$

A new term is added to the lagrangian density

$$\mathscr{L}_{higgs} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) ,$$

with $V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2$ the higgs potential. If μ^2 is negative, the higgs potential has an infinite number of minima with non zero expectation value satisfying $\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}$, with v the vacuum expectation value of the higgs field. Since the photon is massless the vacuum should be electrically neutral. Therefore only the neutral part of the higgs doublet is expected to have a non zero expectation value. The higgs field can be expanded around its minimum at $\frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ v + \eta(x) + i\phi_4(x) \end{pmatrix}$$

The introduction of the higgs doublet introduced four new degrees of freedom. In the Salam-Weinberg model a gauge is chosen, fixing three of the higgs fields, which leads to three massive bosons for the weak interaction and one physical higgs field $h(x)$ giving rise to the higgs boson. The higgs doublet in this gauge is given by:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$$

Similar as for the U(1) symmetry we can define a covariant derivative for a SU(2)×U(1) symmetry: $\partial_\mu \rightarrow \partial_\mu = \partial_\mu + ig_W \mathbf{T} \cdot \mathbf{W}_\mu + ig' \frac{Y}{2} B_\mu$. With fields $\mathbf{W}_\mu$ for the weak interaction, $\mathbf{T} = \frac{1}{2}\sigma$ the generators of the SU(2) symmetry group, a coupling g_W and the already for the U(1) symmetry introduced field B_μ. This leads to the gauge invariant lagrangian density for the higgs field $\mathscr{L}_{higgs} = (D_\mu \phi)^\dagger (D_\mu \phi) - V(\phi)$ with:

$$(D_\mu \phi)^\dagger (D_\mu \phi) = \frac{1}{2} (\partial_\mu h)(\partial^\mu h) + \frac{1}{8} g_W^2 (W_\mu^{(1)} + iW_\mu^{(2)})(W^{(1)\mu} - iW^{(2)\mu})(v+h)^2$$

$$+ \frac{1}{8} (g_W W_\mu^{(3)} - g' B_\mu)(g_W W^{(3)\mu} - g' B^\mu)(v+h)^2$$

From this the terms quadratic in the gauge fields can be associated with potential energy terms and thus with the mass terms. For the W bosons we get terms like:

$$\frac{1}{8}v^2 g_W^2 W_\mu^{(1)} W^{(1)\mu} = \frac{1}{2}m_W^2 W_\mu^{(1)} W^{(1)\mu}$$

with $m_W = \frac{1}{2}g_W v$. The terms of the photon and Z boson are intertwined. To arrive at the physical field a rotation can be performed resulting in the photon field A_μ and the field corresponding to the Z boson Z_μ:

$$A_\mu = \frac{g' W_\mu(3) + g_W B_\mu}{\sqrt{g_W^2 + g'^2}}$$

$$Z_\mu = \frac{g' W_\mu(3) - g_W B_\mu}{\sqrt{g_W^2 + g'^2}}$$

with $m_Z = \frac{1}{2}v\sqrt{g_W^2 + g'^2}$. Due to the mixing of B_μ and $W_\mu^{(3)}$ the coupling to the Z boson then is not given by g_W and thus the weak isospin, but by the so called weak hypercharge $2Q - I_W^{(3)}$ with the electrical charge Q. The mixing terms between the physical higgs field and the weak fields $W^{(i)}$ can be interpreted as interaction terms describing interactions between the gauge bosons and the higgs with a coupling proportional to their mass. A term quadratic in the higgs field $h(x)$ gives us the mass term of the higgs boson leading to a higgs mass given by $\sqrt{2\lambda}v$ while terms with h^3 and h^4 represent self interaction terms of the higgs boson. Since the weak force we observe in nature only couples to left handed fermions, the mass term that we introduced in the beginning of this subsection $m\bar{\psi}\psi = -m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$ will not be invariant under local gauge transformations of the $SU(2)_L$ symmetry group. To add mass terms for the fermions, the Yukawa term

$$-g_f[\bar{L}\phi R + (\bar{L}\phi R)^\dagger]$$

with L a left handed fermion doublet, R a fermion singlet and $g_f = \sqrt{2}\frac{m_f}{v}$ with $\phi^+ = 0$ and $\phi^0 = v + h$ can be added. This results in a mass term for fermions and a coupling term of the higgs to fermions. Before we can arrive at the full lagrangian density of the SM another property of the weak interaction that has to be discussed is flavor mixing. As discussed in section 2.1.1, there are three generations of fermions. The lefthanded doublets the $W^{(1)}$ and $W^{(2)}$ fields couple to, are mixed versions of the doublets in the mass base given by the six quark flavors. For quarks this mixing is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [43], that describes a mixing of the down type quarks to flavor eigenstates:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

This allows for flavor changing interactions between the generations through the exchange of W bosons. A similar matrix, the Pontecorvo-Maki-Nakagawa-Sakata matrix (PMNS [54, 46]) exist for the leptonic sector describing the mixing of the neutrino flavours. The last components of the lagrangian density of the SM are given by the terms describing the strong interactions of QCD. Those arise from requiring invariance under local phase transformations of the SU(3) symmetry group. Similar to the electromagnetic and weak interaction new fields are introduced who manifest themselfs as the bosons called gluon. The SU(3) symmetry group has 8 linear independent generators given by the Gell-Mann matrices resulting in 8 gluons representing different color/anti-color states e.g. in the basis of $r\bar{g}$, $g\bar{r}$, $r\bar{b}$, $b\bar{r}$, $g\bar{b}$, $b\bar{g}$, $\frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g})$ and $\frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b})$. If the potential and kinetic terms of the higgs field including the interaction terms with gauge bosons and their mass terms, the Yukawa terms of the fermions, the kinetic energy terms of the gauge bosons and the kinetic energy terms for the fermions with the appropriate covariate derivative giving us the interactions of electromagnetism, weak and strong interaction are combined, we arrive at the final lagrangian density [60] of the Standard Model of particle physics which is invariant under transformations of the $SU(3)_c \times SU(2)_L \times U(1)_Y$ symmetry:

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{8}tr(\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu}) - \frac{1}{2}tr(\mathbf{G}_{\mu\nu}\mathbf{G}^{\mu\nu})$$ The kinetic terms for the gauge fields

$$+(\bar{v}_L, \bar{e}_L)\tilde{\sigma}^\mu iD_\mu \begin{pmatrix} v_L \\ e_L \end{pmatrix} + \bar{e}_R \sigma^\mu iD_\mu e_R + \bar{v}_R \sigma \mu iD_\mu v_R + (h.c.)$$ Kinetic and interaction terms of the leptons

$$-\frac{\sqrt{2}}{v}\left[(\bar{v}_L, \bar{e}_L)\phi M^e e_R + \bar{e}_R \bar{M}^e \bar{\phi} \begin{pmatrix} v_L \\ e_L \end{pmatrix}\right]$$ Mass and higgs interaction terms of the charged leptons

$$-\frac{\sqrt{2}}{v}\left[(-\bar{e}_L, \bar{v}_L)\phi M^\nu v_R + \bar{v}_R \bar{M}^\nu \bar{\phi} \begin{pmatrix} -e_L \\ v_L \end{pmatrix}\right]$$ Mass[2] and higgs interaction terms of the neutrinos

$$+(\bar{u}_L, \bar{d}_L)\tilde{\sigma}^\mu iD_\mu \begin{pmatrix} u_L \\ d_L \end{pmatrix} + \bar{u}_R \sigma \mu iD_\mu u_R + \bar{d}_R \sigma^\mu iD_\mu d_R + (h.c.)$$ Kinetic and interaction terms of the quarks

$$-\frac{\sqrt{2}}{v}\left[(\bar{u}_L, \bar{d}_L)\phi M^d d_R + \bar{d}_R \bar{M}^d \bar{\phi} \begin{pmatrix} u_L \\ e_L \end{pmatrix}\right]$$ Mass and higgs interaction terms of the down type quarks

$$-\frac{\sqrt{2}}{v}\left[(-\bar{d}_L, \bar{u}_L)\phi M^u u_R + \bar{u}_R \bar{M}^u \bar{\phi} \begin{pmatrix} -d_L \\ u_L \end{pmatrix}\right]$$ Mass and higgs interaction terms of the up type quarks

$$+\overline{D_\mu \phi} D^\mu \phi - m_h^2[\bar{\phi}\phi - v^2/2]^2$$ Higgs kinetic and self interaction terms

$$-V(\phi)$$ Contribution of the higgs potential

Here $F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$, $\mathbf{W}_{\mu\nu} = \partial_\mu \mathbf{W}_\nu - \partial_\nu \mathbf{W}_\mu + ig_W(\mathbf{W}_\nu \mathbf{W}_\mu - \mathbf{W}_\nu \mathbf{W}_\mu)/2$ and $\mathbf{G}_{\mu\nu} = \partial_\mu \mathbf{G}_\nu - \partial_\nu \mathbf{G}_\mu + ig_s(\mathbf{G}_\nu \mathbf{G}_\mu - \mathbf{G}_\nu \mathbf{G}_\mu)$. With $\sigma^\mu = [I, \sigma_1, \sigma_2, \sigma_3]$ and $\tilde{\sigma}^\mu = [I, -\sigma_1, -\sigma_2, -\sigma_3]$ with the Pauli matrices σ_i. e, μ and u, d are the Dirac fields of the fermions with $f_L = (1 - \gamma^5)f$ and $f_R = (1 + \gamma^5)f$. The mass matrices M^f hold the fermion masses as well as the information of the flavor mixing matrices (PNMS and CKM). The covariant derivative D_μ includes the interaction terms with the different gauge fields and acts on quarks and leptons as well as left and right handed particles differently:

$$D_\mu \begin{pmatrix} v_L \\ e_L \end{pmatrix} = \left[\partial_\mu - \frac{ig_Y}{2}B_\mu + \frac{ig_W}{2}\mathbf{W}_\mu\right] \begin{pmatrix} v_L \\ e_L \end{pmatrix},$$

---

[2]Originally in the SM, the neutrino masses were assumed to be zero but due to the observations of neutrino oscillations we indeed know that neutrinos have mass.

$$D_\mu \begin{pmatrix} u_L \\ L \end{pmatrix} = \left[\partial_\mu + \frac{i g_Y}{6} B_\mu + \frac{i g_W}{2} \mathbf{W}_\mu + i g_s \mathbf{G}_\mu \right] \begin{pmatrix} u_L \\ d_L \end{pmatrix},$$

$$D_\mu v = \partial_\mu v,$$

$$D_\mu e_R = [\partial_\mu - i g_Y B_\mu] e_R,$$

$$D_\mu u_R = \left[\partial_\mu + \frac{i 2 g_Y}{6} B_\mu + i g_s \mathbf{G}_\mu \right] u_R$$

$$D_\mu d_R = \left[\partial_\mu - \frac{i g_Y}{6} B_\mu + i g_s \mathbf{G}_\mu \right] d_R$$

and $D_\mu \phi = \left[\partial_\mu + \frac{i g_Y}{2} B_\mu + \frac{i g_W}{2} \mathbf{W}_\mu \right] \phi$.

2.1.3 Limitations of the Standard Model

Until today the Standard Model of particle physics has proven to describe most experimental data. With the discovery of the higgs boson in 2012, all fundamental particles predicted by the Standard Model were found and seem to be SM like [23]. Experimental data suggests [58] that there is no 4th generation of fermions to be expected in a mass range close to the other fermions. Nonetheless there are a few disadvantages of the SM that makes it most certain that there is physics beyond our current Standard Model that has to be discovered. In the following some of the motivations for physic beyond the SM will be discussed.

Dark Matter   Many different observations indicate that there must be a different kind of stable matter not described by the particle content of our Standard Model (e.g. [21]). The matter we are able to describe seems to only make 5% of the energy density of the universe while a far bigger amount of 23% seems to be a new kind of matter that only sparsely interacts non gravitationally with SM particles or itself with interaction strength possibly consisted with those of the weak interaction in the SM. In contrast to neutrinos of the SM, those particles should have high masses to explain for example the structure formation of galaxies.

Matter- Anti-Matter Asymmetry   According to our current cosmological models, the universe that we live in was at the beginning consisting of a dense and hot plasma. Particles, anti-particles and radiation were in a constant state of interaction with particle creation and annihilation. When the universe expanded, the energy density dropped, the universe got colder and depending on their mass the creation of particles in the thermal equilibrium eventually stopped. Since in the Standard Model of particle physics matter- and anti-matter always is produced in pairs, one would expect that when particle creation is no longer possible, the particles would annihilate with the anti particles, converting their energy to electromagnetic radiation. A small amount of the needed inbalance between matter- and anti-matter can be explained with a small amount of CP violation allowed in the SM[3]. For the current models of the early universe this would not be enough to explain the asymmetry we observe [56].

---

[3]In the PNMS and CKM matrix that describe flavor mixing, a parameter for CP violating phase is included.

The hierarchy problem   The particle masses in the SM are depending on the energy scale and undergo corrections from loop interactions with other particles. This leads especially for the higgs bosons to high mass corrections that are quadratic in the energy scale. If the theory is build to hold up to the highest energy scales, a huge amount of fine tuning is needed to keep the higgs mass as small as 125 GeV at the low energy regime we observe at modern particle colliders. This fine tuning is considered unnatural and other theories like a supersymetric extension of the SM deliver more natural ways to explain the relatively low higgs-boson mass.

Neutrino masses   As mentioned in section 2.1.1 the neutrinos are assumed massless in the SM to avoid the problem of right-handed neutrinos. Today we know that neutrinos have mass due to the observation of neutrino oscillations [32]. It is possible to describe massive neutrinos in the framework of the Standard Model, but the neutrino masses are constrained to be in the sub electron volt regime. Therefore the SM offers a fermion mass range spread over 11 orders of magnitudes from the sub electron volt neutrinos to the top quark with a mass of 172.5 GeV. The SM does not deliver a natural explanation for this huge mass range.

Unification   As discussed in section 2.1.2 the electromagnetism and the weak force can be described in a unified theory. In fact if the energy scale is high enough interactions with photons and Z bosons become indistinguishable. A theory that could combine all three forces would give a more natural description of nature. Such a unification of forces is possible in other physics models outside the scope of the SM. E.g., in theories with a SU(5) symmetry [34] the couplings of the different forces unify and that delivers additional theoretical motivation for a model beyond the SM.

2.2   Top physics in the Standard Model and beyond

As motivated in section 2.1.3 the Standard Model of particle physics is not a complete theory and extensions beyond the SM are highly motivated. Checking the internal consistency of the SM and the most precise measurement of its parameters is one of the possibilities to see where new physics might arise. In those consistency test for example done by the Gfitter group [15], the dependencies of the different Standard Model parameters on each other are used to constrain the other parameters like for example the higgs mass. By comparing the fit results for one parameter with its experiental measured value, it can be tested if the SM describes particle physics in a self consistent way. To reasonably do this, precision measurements of the different parameters are needed. One of the most influential parameters is the mass of the heaviest Standard Model particle, the top quark. It heavily contributes to the loop corrections on the higgs mass and can for example in combination with the W boson mass be used to check compatibility with the mass of the higgs as can be seen in figure (2).
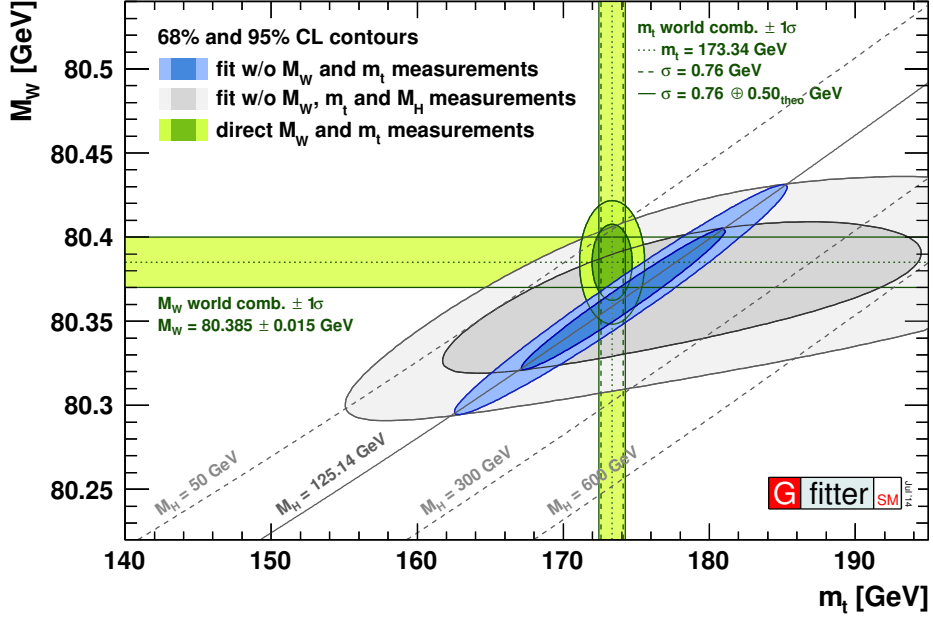
Figure 2: The 68% and 95% confidence level contours for electroweak fits including and excluding the higgs mass measurement as well as the experimental measurements of top quark and *W* boson mass [15].

Besides tests of the SM, the top-quark mass can also be used to examine the form of the higgs potential. It determines if our universe is stable with respect to the minimum of the higgs potential. The dependence of the vacuum stability on higgs and top mass as well as their current measurements can be seen in figure (3).
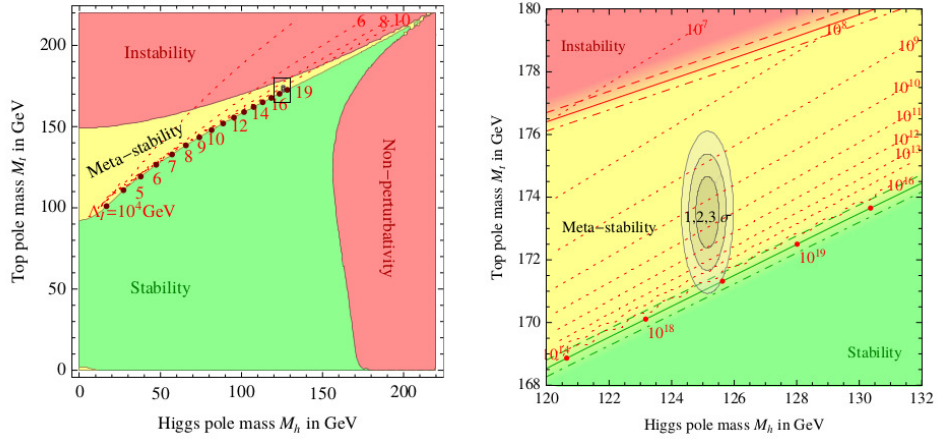

Figure 3: The stability/instability dependence of the higgs potential on top-quark and higgs mass and the error ellipse of the current mass measurements [20].

Further the mass of the top quark and its coupling to other particles have far reaching implications on theories beyond the Standard Model, for example in supersymmetric theories that minimize the influence of top-quark-loop corrections on the higgs mass.

2.3 The t$\bar{\text{t}}$ process

Since the top-quark mass is a possible indicator for new physics, this thesis aims to enable a high precision measurement. The idea of this thesis is based on the top-quark-mass measurement of the CMS experiment discussed in section 3.2 in the semi-leptonic decay channel of top anti-top pair creation events (t$\bar{\text{t}}$). Therefore the following discusses the phenomenology of t$\bar{\text{t}}$ processes. For this thesis we look at t$\bar{\text{t}}$ pairs produced at proton-proton collisions at the Large Hadron Collider (LHC), discussed in section 3. The proton is a baryon consisting of three valence quarks, two up and one down quark bound together by gluon exchange. However if probed at high energy scales at the LHC, the protons substructure of gluons and virtual quark-anti-quark pairs is seen. Therefore the collision of two protons at high energies is the collision of multiple particles from two particle crowds. The average momentum fraction x that the different proton constituents hold is described by the parton distribution functions. It represents the number of particles at a given x, see figure(4).
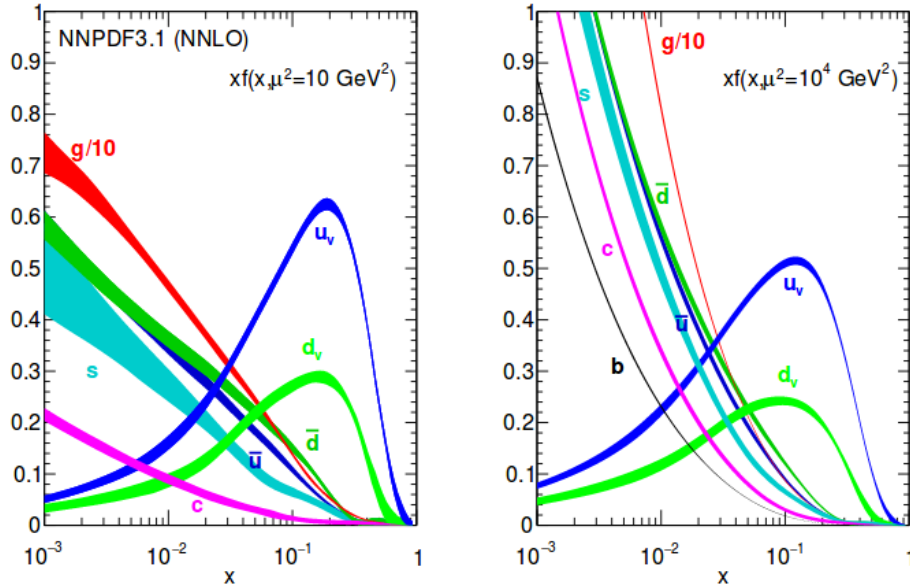

Figure 4: The parton distribution functions of gluons, valence- and sea quarks inside the proton at scales of Q=10 GeV^2 and 10^4 GeV^2[17].

At the LHC the collision energy is high enough that gluons and sea quarks that are mainly found at low x have enough energy to produce a top-quark pair. Therefore the leading processes for t$\bar{\text{t}}$ production at the LHC are given by gluon-gluon fusion and quark-anti quark annihilation shown in figure (5)[59].
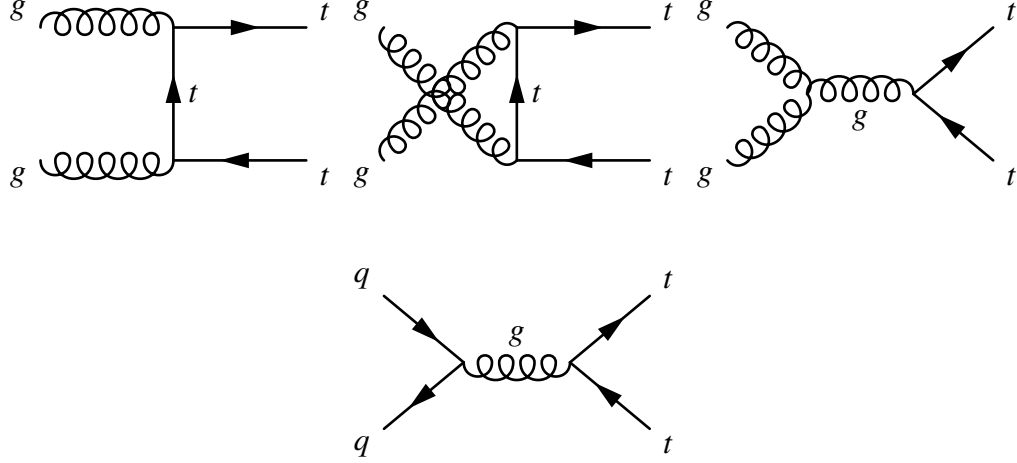
Figure 5: The leading order feynman diagrams of t$\bar{\text{t}}$ pair production at the LHC with gluon gluon fusion (top) and quark anti-quark annihilation.

The top quark has a very short lifetime ($\tau_t \approx 5 \times 10^{-25}$s) and decays via a weak interaction before forming bound states with other particles. The decay is a flavor changing decay via the irradiation of a W boson. The CKM matrix element for couplings between third generation quarks as the top quark and first or second generation quarks is very small. The branching ratio of a decay to a bottom quark is $\text{BR}(t \to bW) \approx 99.8\%$. The W boson then decays further either hadronically in a pair of one quark and a different anti quark or leptonically in a lepton and a neutrino. This thesis focuses on the lepton+jets channel where one of the two W bosons produced in the decay of the t$\bar{\text{t}}$ pair decays hadronically in two light quarks and the other leptonically in a neutrino and either a electron or a muon.
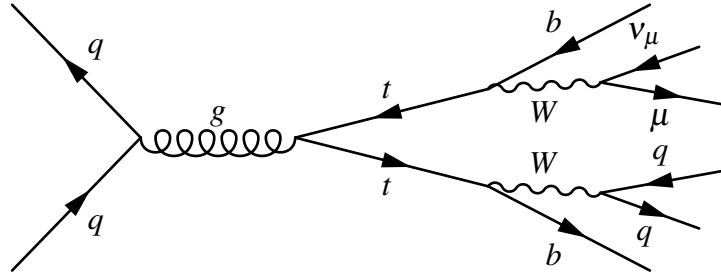

Figure 6: The feynman diagram for the muon+jets final state of t$\bar{\text{t}}$ events.

This channel has two advantages. First a rather high branching fraction since the hadronic decay mode of the W boson is more likely. Second the leptonically decaying W bosons gives the event a rather clear signature, since pure hadronic events have more background processes that produce similar signatures in the detector. Colored states like quarks can not exist freely. If an uncolored state is broken up, a cascade of particles is created until only colorless states remain. Therefore the two expected light quarks as well as the two b quarks produce so called jets consisting of mesons, baryons, leptons and photons collimated in a cone-like structure in the

detector. For this process we expect an event signature consisting of four jets, one lepton and some not reconstructed momentum carried away by the neutrinos. Two of the jets are expected to be reconstructed as b-jets. The feynman diagram describing the $t\bar{t}$ decay with an example production can be seen in figure (6). Processes similar to the one that we aim to observe that contribute as background events include WZ, W+jets and ZZ production as well as single top production tW and QCD multijet events [4].

2.4   Natural Units

Particle physics uses mostly the unit system of natural units. In natural units, [kg, m, s] of the SI system are expressed in terms of [$\hbar$, c, GeV]. By then choosing $\hbar = c = 1$, most equations of particle physics are simplified, since energy, momentum and mass now all can be expressed in terms of GeV [69]. Natural units are used in this thesis.

3 About the Experiment

This thesis uses simulated event samples and experimental data from the CMS[4] experiment inside the LHC[5] at CERN[6] near Geneva in Switzerland. This section gives an overview about the experiment and the data acquisition in CMS.

3.1 The Large-Hadron Collider

The LHC is a circular particle collider in the tunnel of the former LEP[7] experiment at CERN. It accelerates either protons or lead nuclei and collides them at four interaction points at the ATLAS[8], ALICE[9], LHCb[10] and CMS detectors. The data for this thesis is the 2016 data from proton-proton collisions in the run II of the LHC with a center of mass energy of $\sqrt{s} = 13$ TeV and a peak Luminosity of $\mathscr{L} = 1.2 \times 10^{34} \mathrm{cm}^{-2}\mathrm{s}^{-1}$. The protons are supplied by a chain of accelerators, the CERN accelerator complex that can bee seen in figure (7).

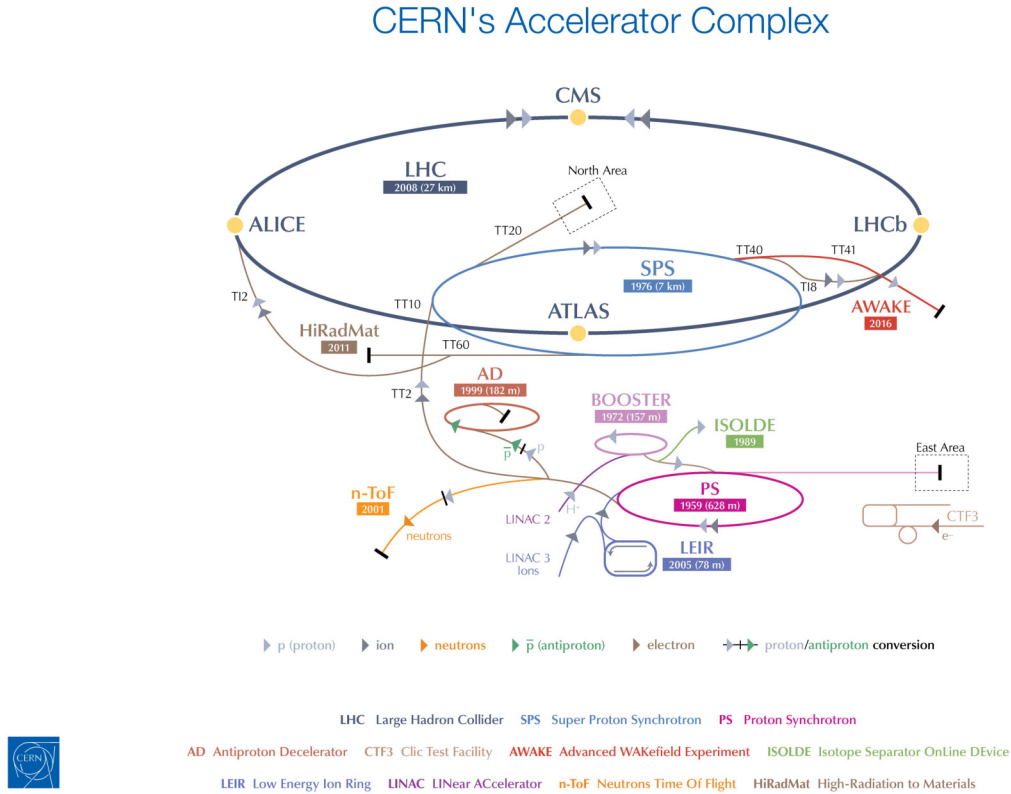**Figure 7:** Schematic overview of the CERN accelerator complex with the pre accelerators of the LHC and the main detector experiments [47].

---

[4]**C**ompact **M**uon **S**olenoid
[5]**L**arge **H**adron **C**ollider
[6]**C**onseil **e**uropéen pour la **r**echerche **n**ucléaire, the European Organization for Nuclear Research
[7]**L**arge **E**lectron **P**roton collider
[8]**A** **T**oroidal **LHC** **A**pparatu**S**
[9]**A** **L**arge **I**on **C**ollider **E**xperiment
[10]**LHC** **b**eauty experiment

The accelerator chain includes the linear accelerator (LINAC), the BOOSTER, the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) accelerators which preaccelerate the protons to an energy of 450 GeV. The LHC ring itself then accelerates the protons to their final energy of 6.5 TeV. This results in a center of mass energy of $\sqrt{s} = 13$ TeV, at the four interaction points where two proton bunches with an average proton number of $1.1 \cdot 10^{11}$ collide every 25 ns [19, 18].

3.2 The CMS Detector

The CMS experiment is a multi purpose detector at the LHC. Similar to the ATLAS experiment, it was designed to discover the higgs boson and search for all kinds of deviations from the Standard Model of particle physics. The CMS experiment consists of four subsystems, that are the name giving muon system, the 3.8 T solenoid magnet, a calorimetric system and a tracking system. A slice of the detector can be seen in figure [8].
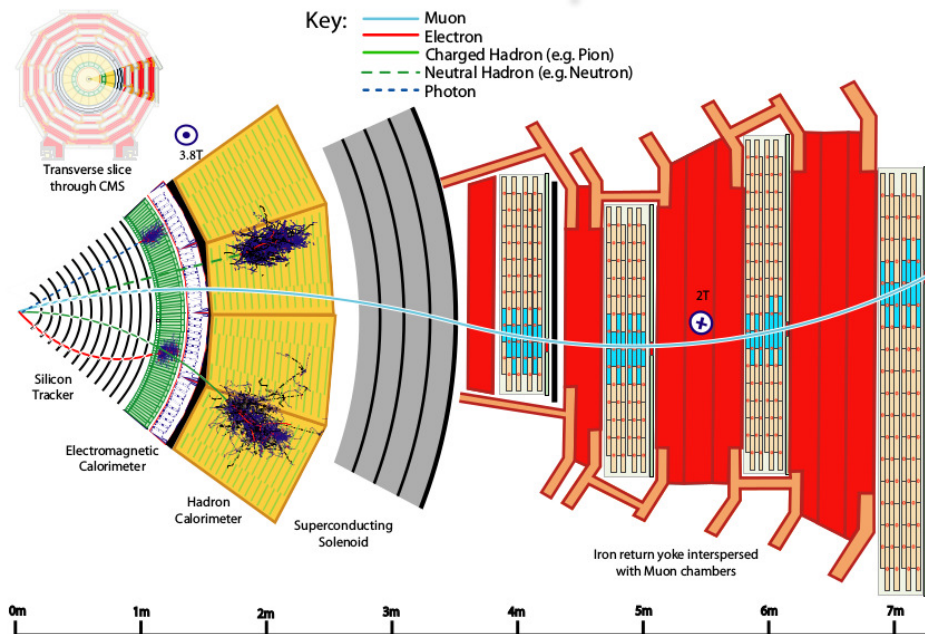

Figure 8: A slice of the CMS detector, showing how different particles interact with the sub systems[62].

The slice shows how particles in CMS are reconstructed in general:

- Charged particles leave tracks in the tracking system situated at the center of the detector. Their tracks are bend in the magnetic field of the superconducting solenoid magnet, allowing for charge reconstruction and momentum determination.

- Mainly electromagnetic interacting particles like electrons and photons, as well as short lived particles decaying into those like the π^0 meson, deposit most of their energy in the electromagnetic calorimeter, allowing for energy measurements.

- Particles that mainly interact via the strong force like protons and neutrons deposit their energy in the hadron calorimeter.

- Muons, beeing minimally ionizing particles, are not stopped in the inner parts of the detector. Their tracks are reconstructed with the muon system in the outer part of the detector.

A schematic overview of the whole detector and the position of the different subsystems can be seen in figure (9). In the following sections the separate subsystems are described in more detail.
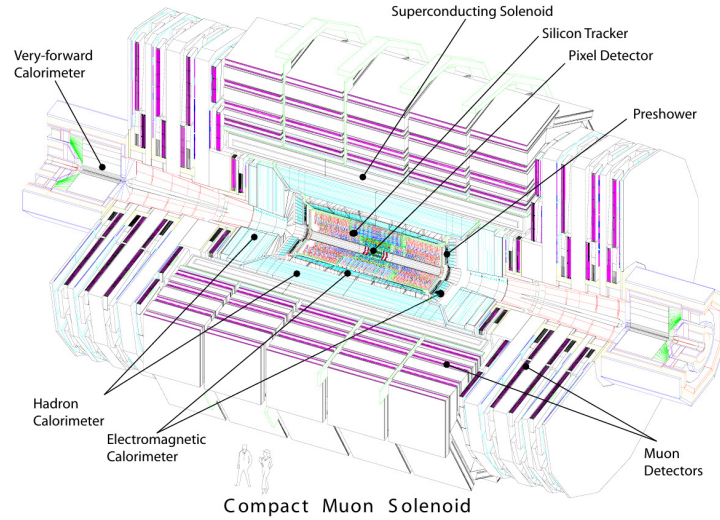

Figure 9: A schematic overview of the CMS detector showing the layered structure of different subsystems including the tracking system, the calorimetry, the solenoid magnet and the muon system [22].

3.2.1 Coordinate system in CMS

To discuss the structure of CMS it is helpful to know the coordinate system of CMS and a few basic terms. The CMS is build as a cylindrical structure around the beam line, with the interaction point in the center. CMS uses a right handed coordinate system with its origin at the interaction point. The x axis points toward the center of LHC and the y axis points up, while the z axis is given by the counterclockwise beam direction. Cylindrical coordinates are used. The radial distance r is the minimal distance from the z axis. The azimuthal angle φ measured in the $x - y$ plane and the polar angle θ measured from the positive z-axis. The latter is often used in form of the pseudo rapidity $\eta = \ln\tan(\theta/2)$ measured from the positive z axis, to give an Lorentz invariant measure $\Delta\eta$ for the angle [59]. Most of CMS subsystems are comprised of a barrel and a end cap part. The barrel parts are rotation symmetrical structures around the z axis and form, together with the end caps which are perpendicular to the z axis, a cylindric structure.

3.2.2 Tracking systems

A tracking system like the CMS tracker [22] has to fulfill several requirements. It measures the momentum, using the bending of charged particles in the 3.8 T magnetic field of the solenoid. Therefore a high granularity is needed to enable the measurement of the curvature even for highly energetic particles with marginally bent tracks. A good spatial resolution is needed

especially close to the interaction point, to allow for determination of secondary and displaced vertices. Being close to the interaction point the tracker has to be radiation hard, while having a low material budget. For a precise energy measurement the particles should lose their energy in the calorimeter system, not the tracker. CMS uses silicon trackers, ensuring a good resolution and radiation hardness. The total material budget is between 0.4 X_0 and 1.8 X_0. With X_0 the radiation length of charged particles. Good vertex reconstruction is ensured with the highly granular pixel detector. A number of up to 12 layers in the strip detector, together with the three pixel layers, ensures a good transverse momentum reconstruction of $\frac{\Delta p_T}{p_T} = $ 1-2% up to $|\eta| > 1.6$ at $p_T \approx 100$ GeV.

The pixel tracker covers the innermost part of the detector with three layers in the barrel region of the detector (BPix) at radii between 4.4 cm and 10.2 cm from the beam line. It has two pixel layers each in the end cap region (FPix). The silicon strip detector has 10 barrel layers and 3 + 9 disk layers in the end cap region. All together barrel and end cap cover a detector region up to a pseudo rapidity of $|\eta| < 2.5$. The pixel detector that can be seen in figure (10) is comprised of (48 + 18) million 100×150 μm pixels in the two subsystems BPix and FPix. Using the effect of charge sharing a spatial resolution of 15-20μm can be achieved. This allows for a 3D vertex reduction that is essential for b quark and τ lepton reconstruction.
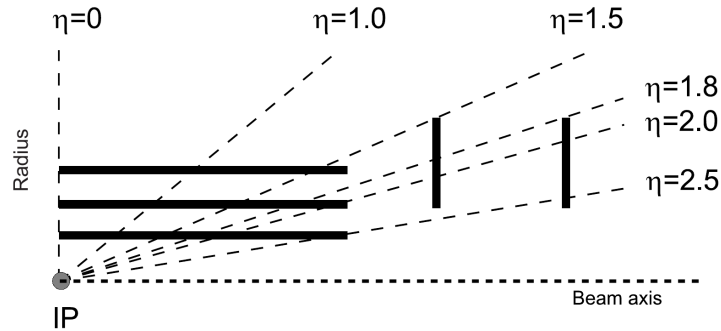

Figure 10: Schematic of the inner tracking layers of CMS with the pixel detector layers in the barrel region (BPix) and the end cap region (FPix), as well as orientation lines for the η coverage of the detector planes [22].

The silicon-strip detector is comprised of four sub-systems, the TIB[11] and TOD[12] in the barrel region and TID[13] as well as TEC[14] in the end cap region (see figure (11)).

---

[11]**T**racker **I**nner **B**arrel
[12]**T**racker **O**uter **B**arrel
[13]**T**racker **I**nner **D**isks
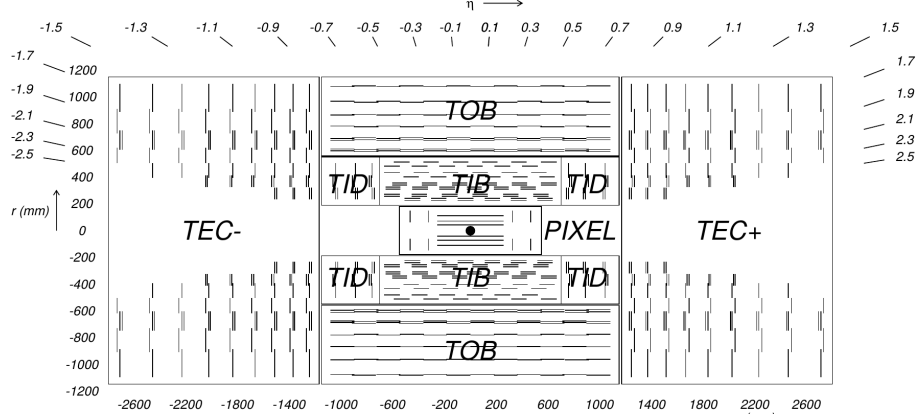[14]**T**racker **E**nd **C**ap

Figure 11: Schematic overview of the different subsystems of the tracking system in CMS with emphasis on the strip detectors. The radial and longitudinal positions of the pixel detectors in forward/end cap and barrel part of CMS, FPix and BPix (Pixel) and the strip detectors: the Tracker Inner Barrel (TIB), Tracker Outer Barrel (TOB), Tracker Inner Disks (TID) and Tracker End Cap (TEC) are given together with information about the η coverage of the detector planes [22].

They are made of 9.3 million strip sensors with a thickness of 320 μm (all) and 500 μm (TEC). In TIB the pitch between the strips is 80 μm and 120 μm in layers 1 + 2 and 3 + 4 respectively leading to a single point resolution of 23 μm and 35 μm. The TOB has pitches of 183 μm in the first four layers and 122 μm in the outer two layers resulting in a resolution of 53 μm and 35 μm.

3.2.3 Calorimeters

Calorimeters use dense materials to further particle showering, eventually stopping the particles and measuring the particle energy deposited in the calorimeter material. There are mainly two different types of particle showers, namely electromagnetic and hadronic showers dominated by interactions of the electroweak force and the strong force respectively. For particles mainly showering via the electroweak force like electrons and photons, the characteristic shower scale is given by the radiation length X_0, driven by the electron density of the showering material. The characteristic shower scale of hadronically showering particles like protons, the interaction length λ_1 is driven by the nucleon density of the showering material. This suggests two calorimetric systems, one with a high Z material, measuring electromagnetic showering (ECAL) and one system with dense material to measure hadronic showering (HCAL). While protons and other strongly interacting particles can travel through the ECAL without loosing to much energy, electromagnetically interacting particles would be stopped in the HCAL, therefore the ECAL lies radially inside the HCAL. Since electrically neutral particles can not be measured in the tracking system, the calorimetric system has to provide good granularity and energy resolution to reconstruct neutral particles like photons. In CMS the ECAL and HCAL lie between the tracking system and the solenoid magnet. This provides the advantage that the particle energy can be measured, without depositing energy in the massive magnet. On the other hand both calorimeters have to be dense to deliver enough radiation length or interaction length respec-

tively to stop the particles in the constrained space inside the solenoid.

ECAL   The CMS ECAL [22] uses lead tungsten crystals $PbWO_4$ as showering material. The advantage of $PbWO_4$ is that it can also be used as a scintillator. Charged particles, like those produced in an electromagnetic shower, that travel through scintillator material produce low energy photons that then again can be measured by photo detectors attached to the crystals. Measuring the number of photons, which is proportional to the deposited energy, allows for an energy measurement of the primary particle that induced the particle shower. The CMS ECAL is composed of 61200 crystals in the barrel region and another 7324 crystals in the end cap region. The barrel part of the ECAL covers particles up to a pseudo rapidity of $|\eta|$ < 1.479 with an granularity of 360-fold in φ and (2×85) fold in η, resulting in an average cross section of 0.0174×0.0174 in $\eta - \varphi$. The crystal length of the barrel crystals of 230 mm corresponds to 25.8 X_0. The crystals in the end caps cover a range of $1.479 < |\eta| < 3.0$. The end cap crystals have a length of 220 mm corresponding to 24.7 X_0. In addition to the crystals, the ECAL consists of a pre-shower detector covering $1.653 < |\eta| < 2.6$. The pre-shower detector detects neutral pions in the end caps and helps to distinguish electrons from minimally ionizing particles, while enhancing the position determination of electrons and photons with a high granularity. The pre-shower detector is build as a sampling calorimeter. This means that the calorimeter is composed of an passive material that initiates showering in alternating layers with the actual active detector material that measures the deposited energy. The pre-shower detector consists of two layers lead radiators as showering material with silicon strip sensors after each radiator to measure the shower energy. The two strip layers are orthogonal to each other allowing for a 2D position measurement. The pre-shower detector has a length of 20 cm, corresponding to 2 + 1 X_0 before the first and second detector layer, resulting in 95% of the photon induced showers starting before the second plane. This is for identify e.g. neutral pions that rapidly decay to two photons, which need to be identified as two separate showers to identify the pion. For particles below an energy of 500 GeV, the energy resolution of the ECAL can be described with a stochastic term S, a noise term N and a constant term C :

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + (C)^2$$

A typical resolution to be found for energies E in GeV is:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2$$

HCAL   The HCAL [22] is comprised of several subsystems that can be seen in figure (12).
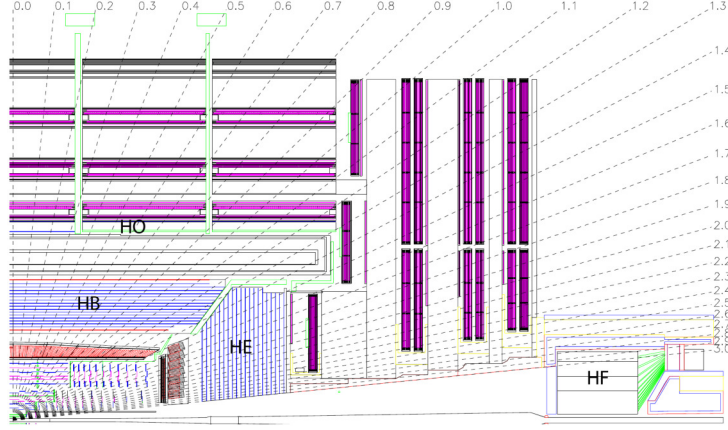
22

Figure 12: Schematic of the CMS detector with emphasis to the subcomponents of the HCAL. Positions of the HB, HO detectors in the barrel part of CMS as well as the HE and HF detectors in end cap and forward region are marked [22].

The barrel part of the HCAL HB is situated between the ECAL at $r = 1.77$ m and the solenoid at $r = 2.95$ and covers an η range up to $|\eta| = 1.3$. It is a sampling calorimeter composed of a front and a back steel absorber plate and 8 layers of flat brass absorber plates as showering material. Plates of plastic scintillator serve as detector material between the different absorber plates. The scintillator plates deliver a detector segmentation of $(\Delta\eta, \Delta\varphi) = (0.087, 0.087)$. The total number of interaction lengths in the absorber material is $5.82\lambda_1$. The effective thickness increases under the polar angle θ to up to a maximum of $10.6 \, \lambda_1$ at $|\eta| = 1.3$. The ECAL adds another $1.1 \, \lambda_1$ of material to the total budged.

The end cap part HE of the HCAL covers a large eta range of $1.3 < |\eta| < 3$. The HE is also designed as a sampling calorimeter using brass plates and plastic scintillator. Together with the ECAL, it has a total length of 10 interaction length (λ_1). The granularity is $(\Delta\eta, \Delta\varphi) = (0.087, 0.087)$ for $|\eta| < 1.6$ and $(\Delta\eta, \Delta\varphi) = (0.17, 0.17)$ for $|\eta| \geq 1.6$. In the central region $|\eta| < 1.3$ the ECAL and HB are expanded by the outer calorimeter HO laying outside of the solenoid and before the muon system. This is done since the combined stopping power of HB and ECAL is not strong enough to stop all hadron showers. The HO adds another $\frac{1.4}{\sin(\theta)}$ interaction length λ_1 to the calorimeter. Thus the minimum absorber length is extended to 11.8 λ_1 at minimum. The tiles in the HO are made to match those of the HB, having a granularity of $(\Delta\eta, \Delta\varphi) = (0.087, 0.087)$. It consists of two layers of scintillator in the central region from -2.686m to +2.686m from the interaction point in *z*-direction. This corresponds to ring 0 of the muon system. Outside the central region one layer of scintillator is used. The HCAL is completed by a forward calorimeter HF, covering a range up to $|\eta| = 5$. The HF has to survive large doses of radiation, therefore quartz fibers were chosen as the active medium. A signal is generated for shower particles over the Cherenkov threshold. Since the threshold is low for light particles like electrons (190 keV) and higher for heavy particles, the calorimeter is mostly sensitive to electromagnetic subcomponents of the hadron showers. The calorimeter consists of steel absorber plates, with the quartz fibers embedded in the absorber. The structure is divided into $(\Delta\eta, \Delta\varphi) = (0.175, 0.175)$ towers. The whole length of the detector is approximately 10

λ_1.

3.2.4 Solenoid

The CMS detector contains a massive solenoid [22] with a length of 12.5 m, able to create a magnetic field of up to 4 T inside its 6.3 m diameter. The magnetic flux is returned in a 10000 t iron return yoke that contains the muon system. The solenoid itself has a four layer winding of NbTi conductor, weighing 220 t.

3.2.5 Muon system

One of the goals of CMS is to achieve a very high muon detection efficency and a very precise muon energy resolution, since the $h \rightarrow Z \rightarrow 4\mu$ is one of the golden channels for finding the Higgs boson. Therefore the CMS detector has an additional dedicated muon system [22] situated outside the solenoid magnet, embedded in the return yoke of the magnet. The muons are the only Standard Model particles that traverse the whole detector, except for neutrinos, which can not be measured by the CMS detector. This allows for a good muon identification when identifying them in the tracker as well as in the muon system (global muons). The muon system consists of gaseous particle detectors inside the plate of the return yoke, that serve as absorber for muon identification. In the barrel region ($|\eta| < 1.2$), drift tube chambers (DT) are used. In the end caps cathode-strip chambers (CSC) are used, covering a range of $|\eta|$ between 0.9 and 2.4. In both the barrel and the end cap region the muon system is extended with fast resistive plate chambers (RPC), dedicated for muon triggering. The DT system, that can be seen in figure(13), consists of four cylindric structures consisting of 60 chambers each in the first three cylinders and 70 chambers in the fourth with in total over 172 000 sensitive wires.
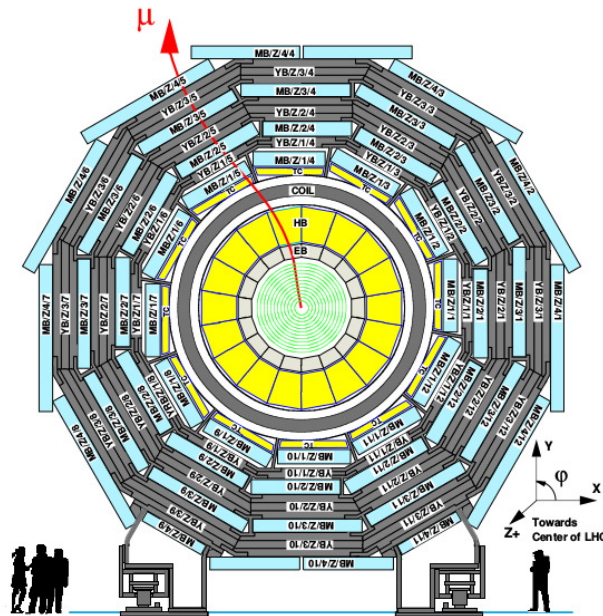

Figure 13: Schematic of the CMS detector with emphasis on the muon system. The radial positions of the ECAL and HCAL barrel detectors, the solenoid magnet and the different components of the iron return yoke and muon system are given[22].

There are layers of wires parallel to the beam line allowing for a track measurement in the (r-φ) plane, as well as wires orthogonal to the beam allowing for a position measurement along the z axis. The single cell spatial resolution was determined in test beam studies to be 170 μm on average. The CSC system, shown schematically in figure (14) consists of 468 chambers of different sizes, covering either $10°$ or $20°$ in φ. Muons in $1.2 < |\eta| < 2.4$ cross three to four CSC. In the overlap region $0.9 < |\eta| < 1.2$ muons are detected by both DT and the CSC system.
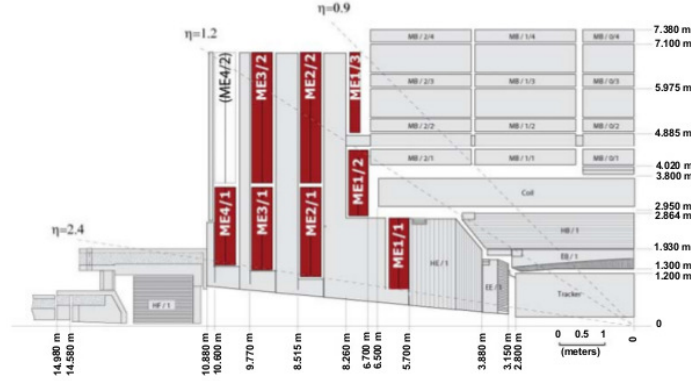
Figure 14: Schematic view of the muon system of CMS with an emphasis on the CSC detectors [22]. The CSC have an expected spatial resolution of ≈ 80 μm.

The RPCs are gaseous parallel-plate detectors. They deliver a time resolution of much smaller than the 25 ns between two bunch crossings and are therefore usable as a source of trigger information [22].

3.3 Event reconstruction in CMS

3.3.1 Trigger system

The LHC creates a huge number of proton proton interactions to find even the rarest types of particle interactions, colliding bunches of protons every 25 ns, resulting in a event rate of over 1 GHz. Processing and storing the resulting amount of data is not feasible. Therefore the CMS has a trigger system, selecting possibly interesting events for storage and further analysis. The selection process has to be very fast, since even a fraction of the > 1 GHz event rate would be impossible to store. Fot this reason, the CMS trigger system is comprised of a fast hardware based trigger, the L1 trigger, that dramatically reduces the event rate to 100 kHz and a software trigger, the HLT[15], that reduces the final event output rate of the CMS experiment to an average rate of 400 Hz [41].

The L1 trigger The L1 trigger has 4 μs to decide whether an event is further processed or discarded. It is comprised of several sub trigger components, which use the faster subsystems of CMS, the calorimetry and the muon system. The L1 trigger tries to identify events with high energetic muon, electron, photon, τ lepton or jet candidates, as well as events with a high amount

---

[15]**H**igh **L**evel **T**rigger

of missing transverse energy[16]. To give an example, the muon triggers use all three subsystem of the muon system, the DT, the CSC and the RPC. They identify track segments from hits in multiple detector planes in a single measurement station. Pattern recognition algorithms then identify muon candidates and their momentum. All candidates are assigned with a p_T, a quality code and a (η, φ) position in the muon system. In the final step of the L1 trigger system, up to 128 separate selections are implemented, using the information of the different sub trigger systems [41].

HLT The HLT consists of a server farm that allows for an event reconstruction similar to the offline event reconstruction. The event rate reduction of the L1 allows the HLT to accommodate a event processing time of up to 175 ms. In this time, for every event, physics objects like muons, electrons and jets are reconstructed. This allows the application of specific identification criteria to select the events of interest that are then further processed for offline reconstruction and permanent storage [41].

3.3.2 Reconstruction of physic objects

For offline analysis the information collected by the CMS detector has to be translated to particle candidates, allowing the reconstruction and analysis of the recorded events. This is done by assigning the signatures to physics object related event structures like tracks in the tracking system or energy clusters in the calorimeters. Combining the information of the different subsystems in the particle flow algorithm [62], those structures are combined to higher level physics objects like muon candidates or jets.

Particle flow Algorithm The particle flow algorithm [62] is used to reconstruct events, combining information of all CMS subsystems. The idea is to use the measurements with the highest energy resolution for the specific particle, disentangle overlapping particles and avoid double counting of particles. For example, to reconstruct an electron, a track reconstructed in the tracking system would be extrapolated to find corresponding energy clusters in the ECAL and HCAL and pre-shower detectors. Those would then be combined with candidate clusters of bremsstrahlungs photons to the final electron candidate.

Muon reconstruction Since muons can leave tracks in the tracking system as well as in the muon system, there are several possible definitions for reconstructed muons, based on which subsystems are emphasized in their reconstruction. For standalone muons, hits in every DT and CSC detector are used as the seed for pattern recognition algorithms that gathers all hits along the muon trajectory in the DT, CSC and RPC detectors. Tracker muons are reconstructed from tracks in the tracking system with a $p_T > 0.5$ GeV that are extrapolated to the muon system

---

[16]The negative sum of all energy deposited in the detector

and accounted as a tracker muon if at least one muon segment matches the extrapolated track. Finally, global muons, that are used in this thesis, are standalone muons that can be matched to a track in the inner detector [62].

Electron reconstruction   Electrons are reconstructed from clusters in the ECAL with a transverse energy $E_T > 4$ GeV. Those clusters are extrapolated to possible hits in the innermost tracking layers. To account for possible bremsstrahlungs photons, the ECAL cluster is extended to a super cluster searching for possible energy depositions from photons irradiated from the electron. A multivariate analysis is then used to identify electrons from possible electron candidates [62, 59].

Jet reconstruction   Jet reconstruction requires an infrared- and collinear safe reconstruction algorithm. This means that the jet definition has to be independent from infinitesimally energetic radiation like soft gluon irradiation and has to remain unchanged if the jet is split into multiple particles with nearly the same direction. For this analysis the anti-k_T algorithm is used. This algorithm is part of a set of algorithms defined by the two distances:

$$d_{ij} = \min(p_{T,i}^{2k}, p_{T,j}^{2k})\frac{\Delta_{i,j}^2}{R^2}$$
$$\text{and } d_{iB} = p_{T,i}^{2k}$$

with $\Delta_{i,j}$ the distance of the two objects i, j in the η, φ plane, R a radius parameter that defines the characteristic clustering scale and k defining the behavior of the clustering algorithm. Both distances are calculated for all pairs of input objects. If a d_{ij} is the smallest, both objects i, j are merged, if a d_{iB} is the smallest, i is considered a jet and removed from further clustering. The clustering stops when all remaining objects are considered jets and therefore no objects are left for clustering. Different values of k result in different clustering algorithms with different resulting jet shapes. For $k = -1$ the anti-k_T algorithm is found. The anti-k_T algorithm results in cone shaped jets with a maximum radius given by R. As input to the clustering algorithm, particle flow candidates are used. Charged particles are excluded if they originate from another vertex than the selected primary vertex. Jet candidates have to be composed of more than two constituents, one of them has to be either a charged hadron, a muon or an electron. The jets are required to have an energy fraction of charged hadrons smaller than 0.99 but larger than 0.0 and an contribution from neutral hadrons and photons, smaller than 0.99. On these particle flow jets, jet energy corrections [26] (JEC) are applied, determined on Monte Carlo data by comparing reconstructed jets and corresponding generator jets. Further residual corrections based on reference objects like leptonically decaying Z bosons are applied. Based on the jets, additional algorithms for particle identification are used. This process is called tagging and is used for example to identify jets induced by a *b* quark. For the *b*-tagging, the properties of the *b* quark like its relative hight mass of $\simeq 5$ GeV and its lifetime of $\mathscr{O}(10^{-12})$ s, are used. The rather long lifetime leads to typical *b* quark flight distances of ≈ 7 mm and therefore often to displaced

secondary vertices for the *b* quark decay. Flight distances to the secondary vertex, invariant mass and quantities like the number of tracks with large impact parameters with respect to the primary vertex are used in different algorithms to compute a likelihood based discriminator value that can then be used to separate *b* jets from other jet types. [59]

Missing transverse energy Neutrinos as well as possible new particles like dark matter candidates only interact via the weak force and will therefore not interact with the CMS detector. Nonetheless they leave a signature in the recorded events because of momentum and energy conservation. The momentum and energy carried by those undetectable particles can be tracked when looking at the overall momentum balance in the detector. This is described by the so called missing transverse energy (momentum), it is defined as the negative vectorial sum of the p_T of all reconstructed particles in the detector. This missing energy (MET) gives a handle on invisible particles like neutrinos in W boson decays and momentum and energy uncertainties of particles used in kinematic fits [59].

4 The top-quark-mass measurement with the ideogram method

The main part of this work is based on the CMS top quark mass measurement in the lepton + jets channel of $t\bar{t}$ events with the so called ideogram method. This chapter is aimed at describing the analysis for the 2016 $\sqrt{s} = 13$ TeV data. With the most recent result given in CMS-PAS-17-007 [4]. The following follows closely the analysis note [4], the CMS internal analysis note CMS AN-16-327[27], a phd thesis [44] , and the master thesis found in [33].

4.1 Dataset and event selection

For the most recent analysis the 2016 dataset of CMS, with an integrated luminosity of 35.9 fb^{-1}, was used. For the mentioned analysis, the events have to either pass a single muon trigger with an isolated[17] muon p_T greater than 24 GeV or an single electron trigger with an isolated[18] electron and p_T greater than 32 GeV. The selection is aimed at identifying $t\bar{t}$ events. Events with either exactly one isolated muon and $p_T > 26$ GeV and $|\eta| < 2.4$ or exactly one isolated electron with $p_T > 34$ GeV and $|\eta| < 2.1$ are selected if they have in addition at least four jets with $p_T > 30$ GeV and $|\eta| < 2.4$, of which two have to be b-tagged. The jets are clustered with the anti-k_T algorithm using particle-flow objects and a distance parameter of 0.4. (see section 3.3) In Table (2) the Monte Carlo generators and parton showering algorithms used in simulation are summarized.

Table 2: Monte Carlo generators and parton showering algorithms used for the signal and background samples.

Sample	Generators & parton showering	Annotations
$t\bar{t}$ signal	PYTHIA 8.219 [64, 65] with the CUETP8M2T4 tune [66]	m_t [GeV] $\in$ {166.5, 169.5, 171.5, 172.5, 173.5, 175.5, 178.5 }
WZ/ W+jets/ ZZ($\to 2l2q$)	MADGRAPH_aMC@NLO v2.2.2 [11] with FxFx merging [30] @ NLO	Drell-Yan with MLM [12] @ LO
WW/ ZZ($\to 2l2\nu$)/ tW	POWHEG v2 [49, 55, 9, 31, 14]	single top with POWHEG v2 and MadSpin

The default parametrization of the parton density functions, NNPDF3.0 NLO with the a strong coupling constant of $\alpha_s = 0.118$ is used for all datasets. In addition for all datasets a GEANT4 [8, 10] based simulation of the CMS detector including the effects of additional proton-proton collisions is applied.

4.2 Kinematic fit

To ensure the compatibility of the selected events with the $t\bar{t}$ hypothesis and to improve the resolution of the reconstructed quantities like the top-quark mass, a kinematic fit is applied that was first used at the DØ experiment [6]. The fit assumes the production of two heavy particles of equal mass, each decaying into a W boson with a mass of 80.4 GeV and a b quark with a mass of 4.7 GeV. Since we look at the lepton+jets channel of $t\bar{t}$ decays, we expect one W boson to decay hadronically and one to deacay leptonically. The fit uses the two non-tagged jets as

---

[17]The complete muon definition can be found in the appendix, section 9
[18]The complete electron definition can be found in the appendix, section 9

candidates for the *W* boson decay products of the hadronically decaying *W* boson. The *b* tagged jets are used as *b* quark candidates. The part of the missing transverse momentum p_T^{miss}, calculated as the negative vectorial sum of the transverse momentum of all particle flow candidates in the event that is uncorrelated to the other four jets is used to constrain the neutrino. The selected charged lepton is used as candidate for the muon/electron in the lepton+jets event. The fit then varies the transverse momentum components of the neutrino and the momenta of the other particle candidates within the given resolution and the kinematic and mass constraints minimizing $\chi^2 := (\mathbf{x} - \mathbf{x}^m)^T G(\mathbf{x} - \mathbf{x}^m)$ with $\mathbf{x}$ the vector of fitted observables, the vector of measured observables $\mathbf{x}^m$ and the inverse error matrix G of the observables. The masses of the neutrino, the lepton and the light quarks are approximated to be zero. With two possible assignments of *b* quarks and *W* bosons and two possibilities for the solution of the longitudinal momentum of the neutrino, each event has four possible input permutations for the fit. In the case of t$\bar{\text{t}}$ events, these can be classified as correctly paired permutations (cp) if the jet and neutrino assignment is correct, or wrongly paired if the assignments are switched. A third category is given for permutations or events where at least one quark or lepton of the t$\bar{\text{t}}$ decay cannot be matched with one of the selected jets or lepton within a distance in the $(\eta - \varphi)$ plane of $\Delta R < 0.4$[19]. To give an example, this can occur if there are more than four jets in the event and one (or more) of the signal jets are not selected due to detector acceptance or due to failing the selection. Those permutations are called unmatched (un). The wrongly paired and the unmatched permutations contribute to the measurement as an additional combinatorial background. Hence, to improve the fraction of correctly paired permutations, an additional cut on the goodness of fit probability, given by the integral over the χ^2 probability distribution is applied. The fit has two degrees of freedom, which simplifies P_{gof} to $P_{gof} = \exp\left(-\frac{1}{2}\chi^2\right)$. In Figure 15, the distribution of the three permutation types in χ^2 and P_{gof} can bee seen for muonic events. For the analysis, a cut of P_{gof} > 0.2 is applied. This reduces the contribution from non-t$\bar{\text{t}}$ background in the baseline selection from 7.6% to 4.3%. Additionally weighting the permutations by their P_{gof} value enhances the fraction of cp permutation from 14.9% to 48.0%, while mostly discarding unmatched permutations.

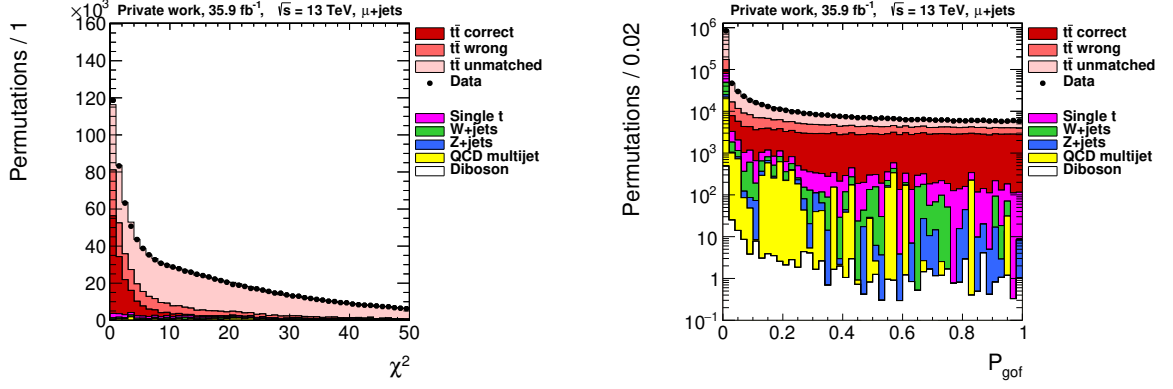[19] ΔR is definded as $\Delta R = \sqrt{\Delta\varphi^2 + \Delta\eta^2}$

Figure 15: The χ^2 (left) and P_{gof} (right) distributions of the kinematic fit for the baseline selection described in ssection 4.1 (Here: muon+jets events only).

4.3 Ideogram method

After the baseline selection described in section 4.1 and after the kinematic fit described in section 4.2 is applied, the top-quark mass is determined with the ideogram method [38]. The ideogram method is a template fit method[20] that extracts the top-quark mass simultaneously with an additional jet-energy-scale factor (JSF). The templates are extracted from the top-quark mass and JSF sensitive distributions of the fitted top-quark mass m_t^{fit} and the reconstructed W-boson mass m_W^{reco} (see figure (16)).
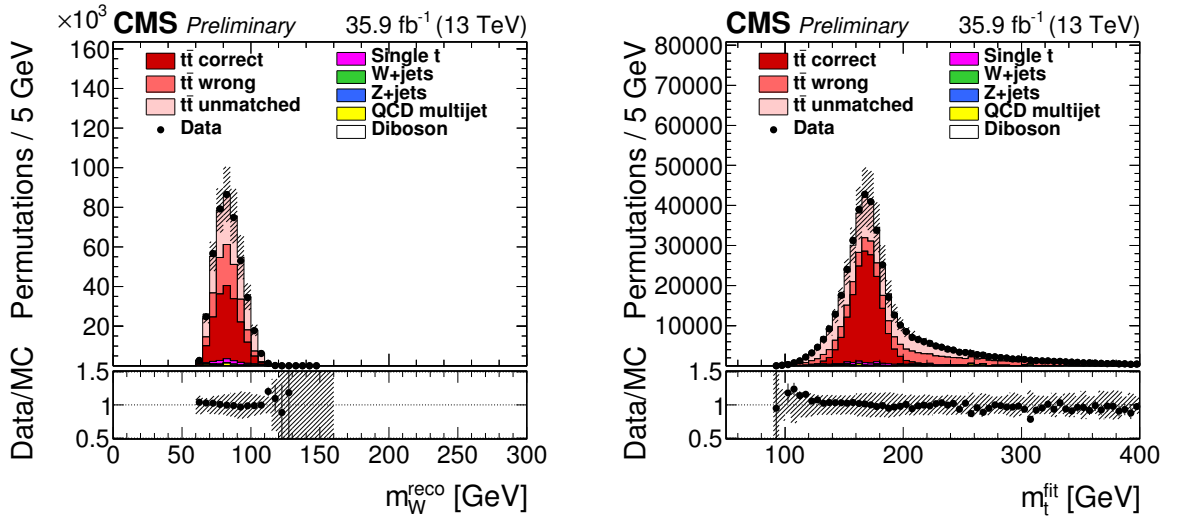
Figure 16: The m_t^{fit} and m_W^{reco} distributions after the baseline selection, P_{gof} cut and the application of P_{gof} as a weight[4].

For this, the two distributions are fitted on Monte Carlo samples for different generator values of the top-quark mass and the JSF. The templates are calculated separately for the different permutation types (cp, wp, un). For the m_W^{reco} distributions, an asymmetric Gaussian is assumed:

---

[20] The fitted quantaties are extracted, comparing distributions in measured data with templates extracted from simulation data, describing the dependence on the generator values of the extracted quantaties.

$$AG(x,\ \mu,\ \sigma_l,\ \sigma_r) = \begin{cases} \frac{1}{\sigma_l\sqrt{2\pi}\cdot\exp}\left(\frac{(x-\mu)^2}{2\sigma_l^2}\right) & ,\text{for}(x-\mu)\ <\ 0 \\ \frac{1}{\sigma_r\sqrt{2\pi}}\cdot\exp\left(\frac{(x-\mu)^2}{2\sigma_r^2}\right) & ,\text{for}(x-\mu)\ >\ 0 \end{cases}$$

The m_t distributions are fitted with a Voigt distribution, a Breit-Wigner distribution convoluted with a Gaussian function for the correctly paired permutations:

$$V(x,\ \mu,\ \sigma,\ \Gamma) = \int dx'\, BW(x',\ \mu,\ \Gamma) \times G(x-x',\sigma),$$

$$\text{with } BW(x,\mu,\Gamma) = \frac{1}{2\pi}\cdot\frac{\Gamma}{(x-\mu)^2+\frac{\Gamma^2}{4}},$$

$$\text{and } G(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\cdot\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

For the wrongly paired and unmatched permutations, a crystal-ball function is used:

$$C(x,\mu,\sigma,\alpha,n) = N\cdot \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{x-\mu}{\sigma} > -\alpha \\ \left(\frac{n}{|\alpha|}\right)^n\cdot\exp\left(-\frac{|\alpha|^2}{2}\right)\cdot\left(\frac{n}{|\alpha|}-|\alpha|-\frac{x-\mu}{\sigma}\right), & \text{for } \frac{x-\mu}{\sigma} \le -\alpha \end{cases}$$

The templates are then extracted by fitting the different functional parameters like x, μ, σ_l, σ_r in the asymmetric Gaussian for the different $(m_t^{\text{gen}}, \text{JSF})$ samples to determine their top-quark mass and JSF dependencies. These templates are then used to derive probability density functions $P_j(m_t^{\text{fit}}|m_t, \text{JSF})$ and $P_j(m_W^{\text{reco}}|m_t, \text{JSF})$ for the different permutation types denoted by j. Then the probability density functions can be used to calculate a likelihood to find the measured data for a given m_t and JSF:

$$\mathscr{L}(\text{sample } | \ m_t,\ \text{JSF}) = P(\text{JSF})\cdot\prod_{\text{events}}\left(\sum_{i=1}^{n}P_{\text{gof}}(i)\left(\sum_j f_j P_j(m_{t,i}^{\text{fit}}|m_t,\text{JSF})\times P_j(m_{W,i}^{\text{reco}}|m_t,\text{JSF})\right)\right)^{w_{\text{evt}}}.$$

Here n denotes the number of selected permutations in one event, j denotes the permutation type (cp, wp, un), f_j the fraction of the given permutation type derived on a Monte Carlo sample and P(JSF) the prior probability for JSF. Minimizing the negative log likelihood $-2\ln[\mathscr{L}(\text{sample } | \ m_t,\ \text{JSF})]$ by varying m_t and JSF then results in the measured top-quark mass and JSF. This measured top-quark mass can therefore be seen as the generator top-quark mass that would be needed to reproduce the data distributions with Monte Carlo samples.

Three different choices can be made for the prior P(JSF) :

1. Fixing JSF to 1, reducing $P_j(m_{W,i}^{\text{reco}}|m_t,\text{JSF})$ to a constant, since m_W^{reco} does not depend on the top-quark mass. This approach is called **1D approach**. This is equivalent to the original approach before introducing the JSF extraction to the top-quark-mass extraction.

2. Leaving JSF free in the minimization, enabling the extraction of m_t and JSF. This is called the **2D approach** and motivates the likelihood above. The additional fit of JSF

was introduced to reduce the systematic uncertainty coming from the standard jet energy scale (JEC) calibration.

3. The third approach was developed by combining the results of the two approaches above. This was mainly done to reduce the systematic uncertainties on the jet energy corrections (JEC), since a combination of the 1D and the 2D approach leads to a partial cancellation of uncertainties.

   In this **hybrid approach**, the JSF is constrained by a Gaussian centered at 1. The width of this Gaussian is motivated from prior knowledge on the JES systematic uncertainty: $\sigma_{\text{prior}} = \delta\text{JSF}^{\text{2D}}_{\text{stat}}\sqrt{1/w_{\text{hyb}} - 1} = 0.0012$, with the statistical uncertainty on the JSF measurement in the 2D approach $\delta\text{JSF}^{\text{2D}}_{\text{stat}}$ [21] and the hybrid weight $w_{\text{hyb}} = 0.3$, chosen to minimize the systematic uncertainty. The hybrid weight determines the contribution of 1D and 2D approach to the hybrid result with $m_t^{\text{hyb}}(w_{\text{hyb}} = 0) = m_t^{\text{2D}}$ and $m_t^{\text{hyb}}(w_{\text{hyb}} = 1) = m_t^{\text{1D}}$.

4.4 Calibration of the template fit Method

To compensate for possible biases in the top-quark-mass extraction originating from the choice of samples that is used for the template extraction[22], a calibration step is applied to the results of the template fit to compensate for a possible bias. Therefore pseudo experiments are generated for each of the 7×5 points of the Monte Carlo samples in the (m_t, JSF) plane. The difference between input and measured top-quark mass and JSF are calculated for the 7×5 samples and then fitted. This is then used to compute a correction function that is applied after the template fit to correct for biases. The effect of the calibration can be seen in figure (17).
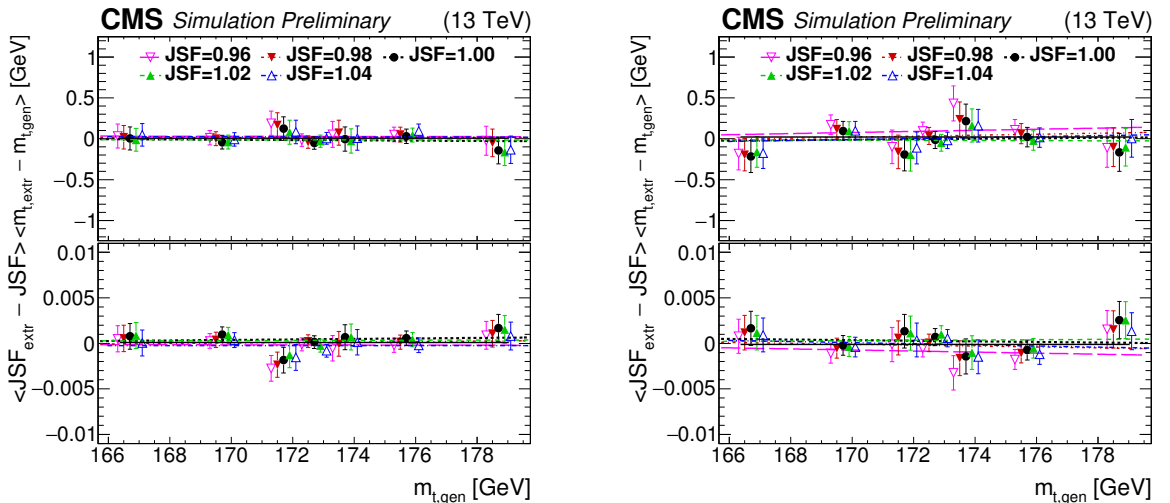
Figure 17: The difference between generator mass and the extracted 2D mass(top) and the extracted and applied JSF (bottom) for the different Monte Carlo samples after calibration for muons (left) and electrons (right). [4]

---

[21]expected value: $\delta\text{JSF}^{\text{2D}}_{\text{stat}}$=0.0008

[22]7×5 samples with m_t values 166.5, 169.5, 171.5, 172.5, 173.5, 175.5, 178.5 and JSF values 0.96, 0.98, 1.00, 1.02, 1.04 were used

4.5 Systematic uncertainties

For the determination of most of the systematic uncertainties, the top-quark-mass measurement is performed on Monte Carlo samples where parameters that represent the systematic uncertainty are scaled up and down. The resulting top-quark mass is then compared with the top-quark-mass result on the default Monte Carlo sample with a top-quark mass of 172.5 GeV and a JSF of 1.00. From the shifts between systematic samples and the default Monte Carlo sample, in case of an up-/down scaling, the larger shift is used as the systematic uncertainty contribution from the parameter that was varied. Table (3) gives an overview of the contributions of the different systematic uncertainties.

Table 3: The different contributions to the systematical uncertinty for the combined measurement (electron and muon channel) for the three different top-quark mass approaches. [4]

	2D approach		1D approach	Hybrid	
	δm_t^{2D}	δJSF^{2D}	δm_t^{1D}	δm_t^{hyb}	δJSF^{hyb}
	[GeV]	[%]	[GeV]	[GeV]	[%]
Experimental uncertainties					
Method calibration	0.05	<0.1	0.05	0.05	<0.1
JEC (quad. sum)	0.13	0.2	0.83	0.18	0.3
– JEC: InterCalibration	(-0.02)	(<0.1)	$(+0.16)$	$(+0.04)$	(<0.1)
– JEC: MPFInSitu	(-0.01)	(<0.1)	$(+0.23)$	$(+0.07)$	(<0.1)
– JEC: Uncorrelated	(-0.13)	$(+0.2)$	$(+0.78)$	$(+0.16)$	$(+0.3)$
Jet energy resolution	-0.08	$+0.1$	$+0.04$	-0.04	$+0.1$
b tagging	$+0.03$	<0.1	$+0.01$	$+0.03$	<0.1
Pileup	-0.08	$+0.1$	$+0.02$	-0.05	$+0.1$
Non-t$\bar{\text{t}}$ background	$+0.04$	-0.1	-0.02	$+0.02$	-0.1
Modeling uncertainties					
JEC: Flavor (linear sum)	0.42	0.1	0.31	0.39	<0.1
– light quarks (uds)	$(+0.10)$	(-0.1)	(-0.01)	$(+0.06)$	(-0.1)
– charm	$(+0.02)$	(<0.1)	(-0.01)	$(+0.01)$	(<0.1)
– bottom	(-0.32)	(<0.1)	(-0.31)	(-0.32)	(<0.1)
– gluon	(-0.22)	$(+0.3)$	$(+0.02)$	(-0.15)	$(+0.2)$
b-jet modeling (quad. sum)	0.13	0.1	0.09	0.12	<0.1
– b frag. Bowler–Lund	(-0.07)	$(+0.1)$	(-0.01)	(-0.05)	(<0.1)
– b frag. Peterson	$(+0.04)$	(<0.1)	$(+0.05)$	$(+0.04)$	(<0.1)
– semileptonic B hadron decays	$(+0.11)$	(<0.1)	$(+0.08)$	$(+0.10)$	(<0.1)
PDF	0.02	<0.1	0.02	0.02	<0.1
Ren. and fact. scale	0.02	0.1	0.02	0.01	<0.1
ME/PS matching	-0.08	$+0.1$	$+0.03$	-0.05	$+0.1$
ME generator	$+0.19\pm0.14$	$+0.1$	$+0.29\pm0.08$	$+0.22\pm0.11$	$+0.1$
ISR PS scale	$+0.07\pm0.09$	$+0.1$	$+0.10\pm0.05$	$+0.06\pm0.07$	<0.1
FSR PS scale	$+0.24\pm0.06$	-0.4	-0.22 ± 0.04	$+0.13\pm0.05$	-0.3
Top quark p_T	$+0.02$	-0.1	-0.06	-0.01	-0.1
Underlying event	-0.10 ± 0.08	$+0.1$	$+0.01\pm0.05$	-0.07 ± 0.07	$+0.1$
Early resonance decays	-0.22 ± 0.09	$+0.8$	$+0.42\pm0.05$	-0.03 ± 0.07	$+0.5$
Color reconnection	$+0.34\pm0.09$	-0.1	$+0.23\pm0.06$	$+0.31\pm0.08$	-0.1
Total systematic	**0.72**	**1.0**	**1.09**	**0.62**	**0.8**
Statistical (expected)	0.09	0.1	0.05	0.07	0.1
Total (expected)	**0.72**	**1.0**	**1.09**	**0.62**	**0.8**

In the following, a brief description of the different sources of systematic uncertainty [4] and how the respective samples are created is given:

Experimental uncertainties:

- **Method calibration:**

  The quadratic sum of the statistical uncertainty and residual biases after the calibration is taken as the uncertainty for the calibration of the template method.

- **Jet energy corrections:**

  To take into account the influence of p_T- and η-dependent energy corrections on the extraction of the JSF, the energies of all jets were scaled up and down with respect to their data/simulation uncertainties [39, 1].

- **Jet energy resolution:**

  The jet energy resolution (JER) is measured in data. [39]. This is then used to smear out simulated events, so that they match the data. To propagate the uncertainty of the JER measurement, the smearing of the simulated events is varied by one standard deviation.

- **b-tagging:**

  To take the p_T dependent uncertainties of the b-tagging efficiency and the misidentification rate of the tagging algorithm into account, the weights of the default sample are reweighed for this uncertainty.

- **Pileup:**

  The uncertainty originating from the uncertainty on the number of pileup events is accounted for by varying the pp cross section by $\pm\,4.6\%$.

- **Non-$t\bar{t}$ background:**

  To account for the uncertainty in the theory prediction of the normalization of the non-$t\bar{t}$ background, the normalization is varied by $\pm100\%$ (QCD multijet samples), $\pm30\%$ (W+jets samples) and $\pm10\%$ (single top quark and Z+jets and di-boson samples). The luminosity uncertainty [3] of 2.5% is also included in this variations.

Modeling uncertainties:

- **JEC flavor dependence:**

  The difference in jet energy response in PYTHIA 6.4 and HERWIG++ 2.4 [16] is determined for each jet flavor [39]. The different uncertainties for each flavor are evaluated separately and then added linearly.

- **b-jet modeling**:

  Three terms are computed:

  1. The fragmentation into b hadrons is varied within the uncertainties of the Bowler-Lund fragmentation function [36, 7].

2. The Bowler-Lund fragmentation function is compared to the Peterson fragmentation function.

3. The uncertainty on the branching fraction of semileptonic B hadrons is included by varying the BR by -0.45% and +0.77%, the uncertainty obtained in the measurements of B^0/B^+ decays[53].

- **Parton distribution function (PDF):**
  To incorporate this uncertainty, a number of replicas for each event is created with NNPDF3.0 NLO PDF the RMS of those is used as uncertainty. Also the strong coupling α_s is scaled to 0.117 and 0.119 (default: 0.118).

- **Renormalization- and factorization scale:**
  A reweighting of the simulated events is done to match the shapes of events generated with renormalization and factorization scales that are both or separately scaled by factors of 0.5 and 2.

- **Matching:**
  In POWHEG the matching to parton showering described by PYTHIA is controlled by the parameter $h_{\mathrm{damp}} = 1.58^{+0.66}_{-0.59}$ [2]. To account for its contribution, it is varied within its uncertainty.

- **Parton shower: Scale of initial state radiation (ISR):**
  For the uncertainty in the ISR modeling, the scale used in the simulation of initial state radiation is scaled by factors of 0.5 and 2.

- **Parton shower: Scale of final state radiation (FSR):**
  For this uncertainty, the FSR scale of PYTHIA is scaled by a factors of $\sqrt{2}$ and $\frac{1}{\sqrt{2}}$ [66]. The momenta of all jets in the FSR samples are scaled such that the response for the light jets match the response of the default sample.

- **Choice of the Monte Carlo generator:**
  To estimate the effects from the choice of the Monte Carlo generator, the default sample with the POWHEG v2 generator is compared to a sample generated with MADGRAPH_aMC@NLO and FxFx matching.

- **Modeling of top p_{T} spectrum:**
  There is a suggestion from a recent calculation that the top p_{T} spectrum is strongly affected by next-to-leading-order effects [28]. Hence, the simulation of top p_{T} is varied to match the distribution measured in [40, 63].

- **Underlying event:**
  PYTHIA is tuned to match measurements [66, 2] of the underlying event for non perturbative QCD effects. The tune is varied within the uncertainty of its parameters

- **Early resonance decays:**

  Enabling early resonance decays (ERD) in PYTHIA allows the estimation of the effects of color reconnection on the top-quark decay products.

- **Color reconnection:**

  The default sample with PYTHIA8 is compared to samples with two alternative models of color reconnection, "QCD inspired" [25] and "gluon move" [13].

Since the 1D and 2D approaches of the mass extraction resemble special cases of the hybrid approach, it is understandable that the top-quark-mass shifts of the hybrid mass are between those of the 1D and 2D mass. In case of different signs of the top-quark-mass shifts in 1D and 2D approach this leads to a sort of partial cancellation effect and thus to lower systematic uncertainties. This fact is used in the derivation of the hybrid weight, so that the hybrid mass by construction reduces the systematic uncertainty.

4.6 Results of the lepton+jets analysis and implications on this thesis

The analysis yields the following results for the top-quark mass in the different approaches:

- m_t^{1D}=172.93±0.07(stat)±1.09(syst) GeV

- m_t^{2D}=172.40±0.09(stat+JSF)±0.68(syst) GeV

- JSF^{2D} = 0.994±0.001(stat)±0.010(syst)

- m_t^{hyb}=172.25±0.08(stat+JSF)±0.62(syst) GeV

- JSF^{hyb} = 0.996±0.001(stat)±0.008(syst)

As expected, the hybrid approach leads to the smallest uncertainties. It is to be noted that this analysis is limited by the systematic uncertainties and not the statistical uncertainty. Therefore, this thesis is aimed at searching for new techniques to improve the analysis summarized above and thus reduce the systematical uncertainties of the ideogram method. In the following chapters, possible approaches with neural networks that aim to solve this by an improved event selection, will be discussed. It should be mentioned that my work restricts itself to the muon channel. When only looking at the muon events, the top-quark-mass result and therefore the starting point of this thesis is: m_t^{hyb}=172.33±0.1(stat+JSF)±0.64(syst) GeV

5 Introduction to neural networks

Developments in machine learning in general and artificial neural networks in particular are growing more important everyday. The influence of machine learning in our everyday life ranges from image recognition that is used e.g. in modern cameras for facial recognition to the powerful algorithms of social media. But besides that it can also be used, and is used, in particle physics analysis. Machine learning allows us to write computer programs that can learn to see patterns in data and use the achieved knowledge to model nearly every function that maps the input data to an output value. This can be a continuous or integer valued function solving regression or classification problems. In the work of this thesis artificial neural networks were used for event categorization in three different categories, namely the cp, wp and un permutations that were described in chapter 4.2. To understand how this works, the following chapter explains the basics of machine learning with neural networks. The content in this chapter is if not noted other way based on [37] and [35].

5.1 The neuron model

The idea of a neural network is based on concepts that describe the biological processes in the human brain. The computational units of an artificial neural network, the neurons, are inspired by the biological neurons. This is often misunderstood in the sense that artificial neural networks model a human brain, but the truth is that the neurons in an artificial neural network are only a simplified mathematical description of a biological neuron and the structures how neurons are connected in the human brain are much more complicated. Nonetheless artificial neural networks have proven to be very useful (especially) outside biology.
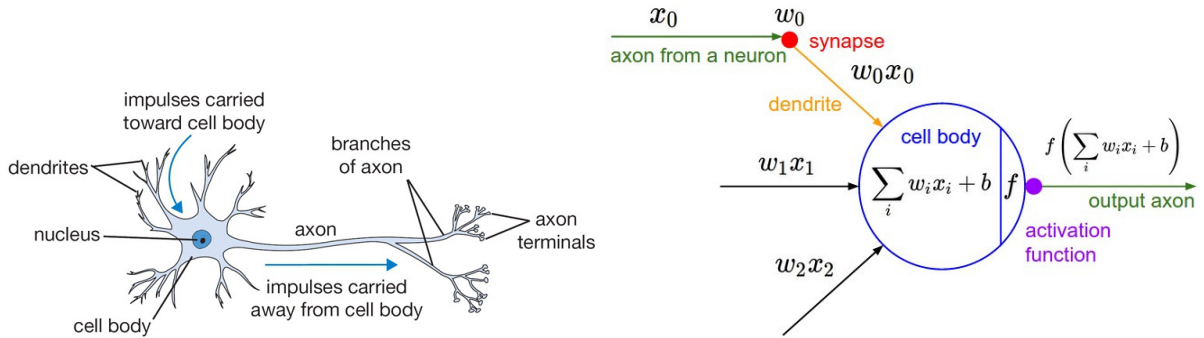
Figure 18: The biological (left) and mathematical model of a neuron (right) [37].

In figure 18 on the left we can see the biological model of a neuron. The neuron collects incoming signals from it dendrites which are connected to other neurons. Out of these signals it produces an outgoing signal along its axon, which then is again connected to the dendrites of other neurons. The mathematical description of this processes leads to the neuron model of an artificial neuron on the right of figure 18: The neuron computes a weighted sum of its inputs coming from other neurons and may add a bias value to it. The weighted sum of inputs (plus the bias) is then input to a nonlinear function, the so called activation function that computes the output signal of the neuron. Later we will see that the weights and a bias value that may be

added to the sum, are parameters that contain the information a neuron has learned. An artificial neural network is a more or less complex structure of these building blocks connected to each other. A trained network then maps a (multi dimensional) input to a (multi dimensional) output it was trained to compute.

5.2 Building blocks of a neural network

Before we can go on to explain how a neural network can be trained, we have to decide on how we connect the neurons to a network. Therefore the neurons are ordered in layered structures. In general the neurons in every layer take the outputs of the layer before as their input. The first layer of neurons gets its inputs from the input layer. Every node of the input layer is then one of the variables used as input in the neural network. The output of the last layer is then the output of our entire network. This can be a number, a vector or a tensor. One of the design choices one can make when building a neural network is how to connect the output of one layer to the input of the next, as well as the internal structure of the layers. Most commonly used are fully connected layers, this means that every neuron takes the output of every node of the layer before as an input. With this kind of layers we get a simple mathematical description of a neural network: Let x_0 be the vector of input values, then the output of the next layer of neuron is $h_1 = f(w_0 \cdot x_0 + b_0)$. Here the layer is denoted by h_1, since the layers between input and output layer are called hidden layers. $f()$ is the activation function of the neurons in this layer, w is the weight matrix containing a vector of weights for every neuron and b the bias vector with a bias value for each neuron. The output of the next layer would then be $h_2 = f(w_1 \cdot h_1 + b_1)$. This gives us a recursive mathematical definition of the whole network: $h_{i+1} = f(w_i \cdot h_i + b_i)$. A schematic depiction of a layered ANN is shown in picture 19.
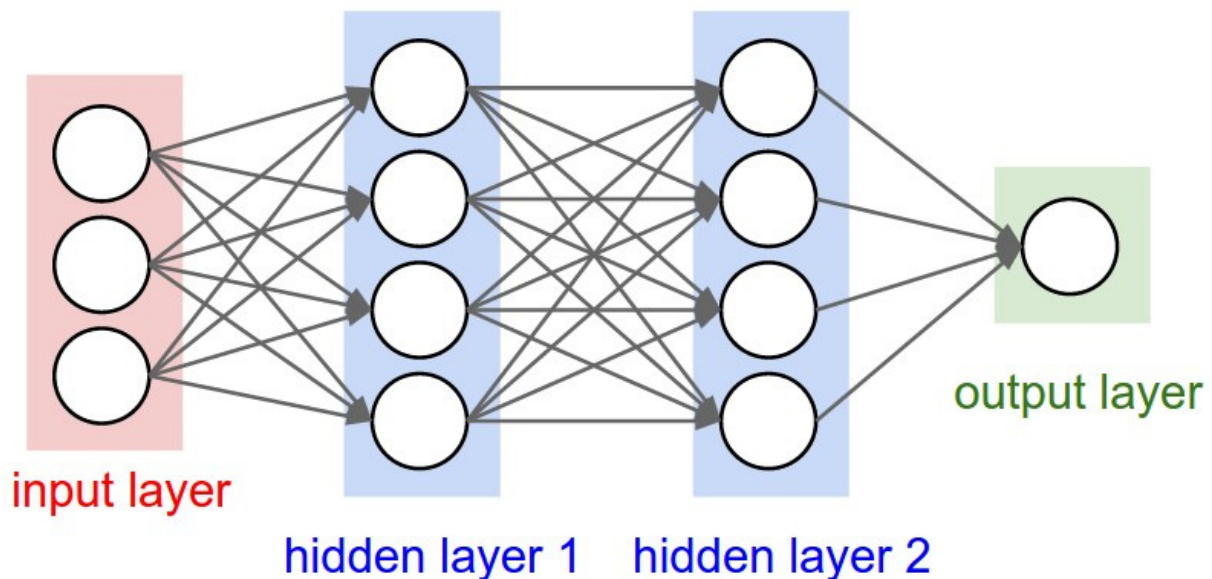
Figure 19: Schematic description of a densely connected feed forward network with two hidden layers [37].

This definition holds for all so called feed forward networks, i.e. networks without loops where

the output of a neuron has influence on its own input. Other layer types include pooling layers, convolutional layers, dropout layers and many more. Of those only the dropout [67] layer are used in the context of this thesis and are therefore discussed. Dopout layers are a technique for regularization and are discussed in subsection 5.5.

5.3 Activation functions

Another important step when designing a neural network is to choose wich activation function is used in the neurons. This is especially important for the output layer, since the activation function of the last layer gives the possible value range the output can have, i.e it determines the form of the network output that we can map the inputs to. In the following some of the most commonly used activation functions are shown and discussed.

Most commonly used are ReLu[23] units. They are defined as $f(z) = max(0, z)$, i.e the output of a ReLu neuron is simply $z = w_i[j] \cdot h_i + b_i[j]$ if the weighted input of the neuron $w_i[j] \cdot h_i$ is greater than a threshold $-b_i[j]$ or 0 otherwise. An advantage of ReLu units is, that they are easy to compute. On the downside ReLu units can "die", i.e. if their output gets zero, there is the possibility that a NN training algorithm does not find an other weight update that gets the neurons output non zero again, since most algorithms are gradient based and the gradient for negative $w_i[j] \cdot h_i + b_i[j]$ is simply zero.

Neurons with a step function like activation, the perceptrons, played an important role in the development of neural networks, since they can be used to build logical functions like AND, OR, and NAND. These are the basic functions out of which every other function on a computer is build up.

Sigmoid units are a more general version of a perceptron, that can have a non binary output. The sigmoid function is defined as $\sigma(z) = \frac{1}{1+e^{-z}}$ with z defined as for the ReLu units.
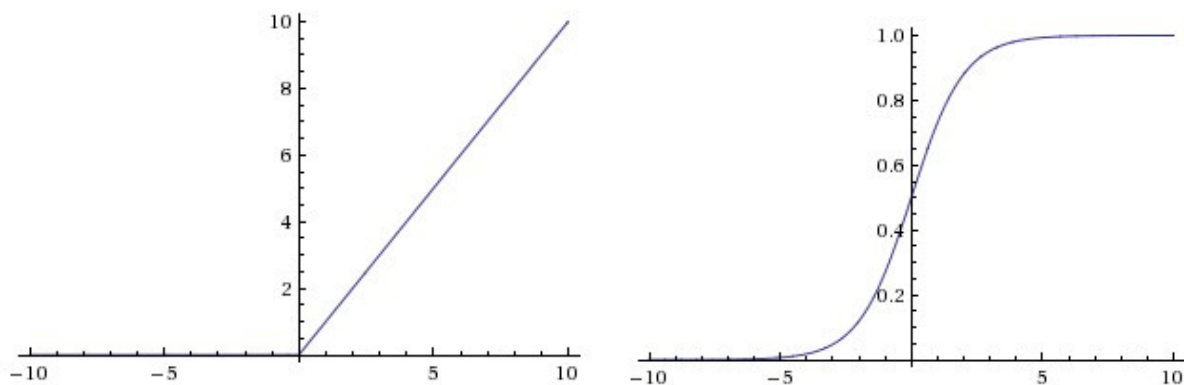
Figure 20: The ReLu (left) and sigmoid (right) activation function.[37].

The sigmoid maps its inputs to the value range [0,1]. A continuous function like the sigmoid function has the advantage of being better defined when calculating a gradient in training than a not differentiable function like the step function. The range [0,1] also allows the network to make a binary classification with a sigmoid in the output layer, i.e a input from type A would

---

[23]Rectified Linear Unit

be mapped to values ≈ 0 and inputs of another type B to values ≈ 1. A similar problem as for negative inputs of the ReLu function occurs also for very high or low inputs of the sigmoid, the gradient becomes very small and the output of the neurons becomes very insensitive to weight changes in the neural network training. To deal with this problem the inputs of the neural network like particle momenta were rescaled in this thesis to start in the gradient sensitive regions of the activations function. For the neural networks in this thesis, both ReLu and sigmoid were used. However, for the output layer a softmax classifier is used, that assures that the sum over all outputs in the output layer is one with, the single outputs also ranging between [0,1] as for the sigmoid neurons. It is defined as $f(x_j) = \frac{e^{x_j}}{\sum_k e^{x_k}}$.

5.4 Training of a neural network

The basic of idea of training a neural network is to compare the output of the network on a training input to a known target output and then modify the weights, biases and other trainable parameters of the neurons to improve the agreement of the computed and the target output. Therefore, the networks usually starts with a random initialization of the trainable parameters. A typical choice to do this would be to choose random values out of a Gaussian distribution around zero with width one. In this analysis, a data based approach, the layer-sequential unit-variance (LSUV) initialization described in [51], is used. In this initialization method the weights of every layer are iteratively scaled, so that the variance of the layer output on a test batch of data is forced to unity. This technique can help to improve the convergence behavior of the network in training, i.e. it improves the training time that is needed for training our network model. After the initialization a loss function is computed that describes the difference between the predicted and the target output. Then a usually gradient based algorithm is used to minimize this loss and hence improve the agreement between the neural network output and the wanted output on the training data. In the case of this thesis where a classification based on an output node for every class is used, the most common choice is the categorical cross entropy. It is defined as

$$-\frac{1}{n} \sum_x \sum_j [y_j \cdot \ln(z_j) + (1 - y_j) \cdot \ln(1 - z_j)]$$

where the z_i are the outputs of the different output neurons that classify our input and y_i the desired neuron outputs. The minimization algorithm then changes the weights and biases iteratively to minimize the loss. The simplest algorithms are based on gradient descent. This means that they calculate the derivative of the loss function with respect to the weights and bias parameters and change those in the direction of falling gradient with a variable step width. This step width is usually called the learning rate and is an adjustable parameter in the network design. Typically the learning rate is adapted over time, i.e. is reduced when approaching the minimum. One method tried in this thesis is ADADELTA. It is discussed in [71]. It is a gradient descent based method that chooses dynamically adapting learning rate for every weight dimension, so that no manual learning rate choosing is necessary. Other algorithms considered are the basic SGD, RMSprob, ADAM [42] and ADAGRAD [29].

5.5 Regularization

One problem that can occur when training a neural network is called overtraining. This basically means that the network learns patterns in the training data that are only of statistical nature and not existing in the overall data the network is trained to describe. To check the training process for this effect, normally a part of the training data is hold back to test the network on an independent data set while training it on the rest. If the loss function on the data used for training gets smaller than on the test data, the network starts to focus on statistical artifacts of the training data set. This happens eventually if the network is trained rather long and has learned most relevant features in the data or has an insufficient training data set for the given number of trainable parameters. A way to assist the network in handling over training is regularization. In the following some possible techniques are described: The commonly used L2/L1 regularizations and dropout layers. The first extends the loss function with a penalty term of the form: $-\frac{1}{2} \cdot \lambda \cdot w^2$, the squared sum of all weights and biases in the network multiplied with the regularization strength. This especially penalizes single large weights in the network. Those indicate the usage of a single input in parts of the network and therefore the ability of the network to focus on the distribution and fluctuations in one single input parameter instead of looking at overall patterns in the data. The idea of the L1 regularization is similar, here instead of the square a linear term of the form $\lambda \cdot w$ is used. The idea of dropout layers [67] takes a different aproach in the direction of network averaging. A dropout layer randomly disables nodes in the attached layer removing them temporally from the network. This way in every training iteration an other sub network of the original network is trained, basically training multiple networks with shared neurons at the same time. This has shown to prevent over training and helps improving the networks performing.

6 Analysis

The Analysis in this thesis mainly focuses on the usage of deep neural networks (DNN) in the context of the top-quark-mass extraction described in section 4. How DNNs work in general is explained in section 5. The following subsection 6.1 describes a first look at the application of neural networks in the context of the lepton+jets analysis. The knowledge and intuition gained in this first tests with neural networks is then used in section 6.2.1 to build a neural-network selector that can be used to tune the fractions of the different permutation types in the selection for the top-quark-mass extraction. The usability of this classifier for the top-quark-mass extraction is then discussed in section 6.3. The neural networks in this thesis were build with keras [24] using the theano [68] and tensorflow [5] backends. The data and Monte Carlo samples used here are the same that were used in the lepton+jets analysis described in section 4, however the work here is restricted to the muonic decay channel.

6.1 Starting with neural networks and classification problems

Before starting with the classification task of separating the different permutation types, some intuition is needed on how to build neural network architectures for physical classification tasks in general. We will start looking at Monte Carlo events from the lepton+jets analysis describing $t\bar{t}$ events at reconstruction level. Before we start distinguishing jet assignments we can start with a simpler task, namely the distinguishing of the di-jet systems in $t\bar{t}$. We take events that passed the baseline selection of the lepton+jets analysis having a correctly paired jet assignment. The two b-tagged jets will give the invariant mass of the bb system, m_{bb}. While the other two jets give the W-boson mass, m_W. To start with a two dimensional task we can try to separate both di-jet systems using their invariant masses. This will be a first step to distinguish jets and jet assignments and provides an easy understandable task of separating easyly distinguishable distributions. The distributions of m_W^{reco} and m_{bb}^{reco} that we want to separate can be seen in figure (21).
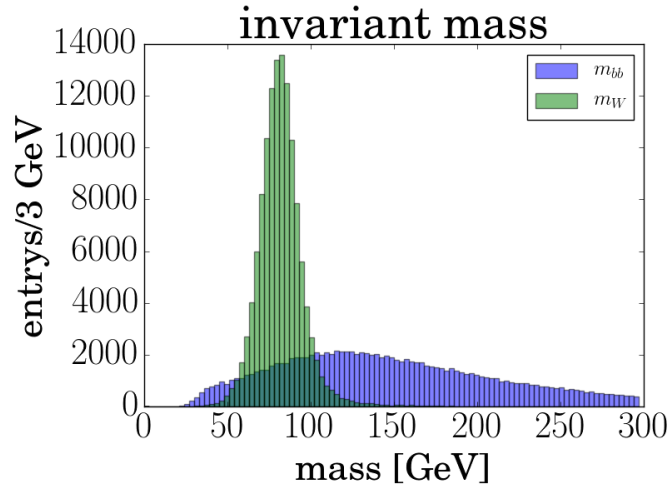
Figure 21: Distribution of the reconstructed di-jet masses m_W^{reco} (blue) and m_{bb}^{reco} (green) in GeV for simulated $t\bar{t}$ events with a generator top-quark mass of 172.5 GeV.

As can be seen the *W* mass distribution has a rather sharp peak around the W-boson mass of 80.4 GeV while the invariant mass of the bb system has a rather flat distribution that tends also to higher values of 300 GeV or more. In order to put the assignment problem in the form of a classification problem we give the network two inputs, supposed to be the *W* boson and the *bb* mass. Defining a true category with (m_W, m_{bb}) inputs and a false category with the reverse input (m_{bb}, m_W), the network then has to decide if the mass "assignment" was correct or not. This input mapping task the network has to learn is visualized in figure (22).
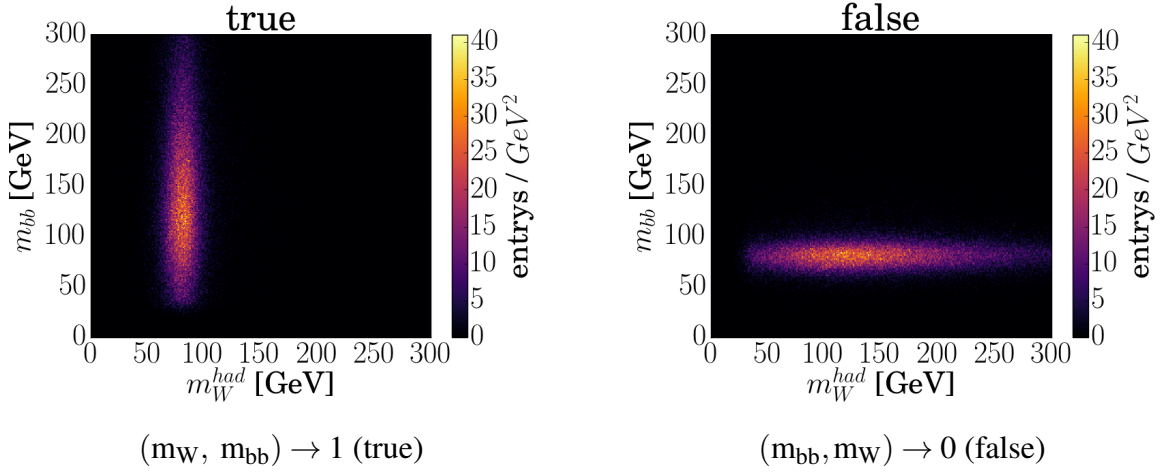
Figure 22: The input mapping a neural network has to learn to separate correct (true) and wrong (false) assignments of (m_W, m_{bb}), in the (m_W, m_{bb}) mass plane with masses in GeV.

Networks designed for classification should give an output usable to make binary decisions, if the different categories apply or not. In this case we have two categories, so we can use a single output to decide weather an input was classified true or false with predicted values near one and zero respectively. The training input is (m_W, m_{bb}) and (m_{bb}, m_W) with training labels (1) and (0). For the output we use a single output node with a sigmoid activation resulting in outputs between zero and one. The rest of the network architecture was optimized with a hyper parameter search determining how many layers and nodes are needed and which activation functions and optimizers to use. We also tested for overtraining, weather normal L1 or L2 regularization should be used or if dropout layers are helpful. Taking inputs with an output of > 0.5 as true and < 0.5 as false predicted, we achieved a 91.17% accuracy on a separate test data sample. A dense network with two hidden layers having 128 nodes each, intermediate dropout layers and a simplified version of the sigmoid function[24] as activation function was used. Considering that the distribution of two categories (true/false) in the m_W, m_{bb} are fairly different, 91.17% seems to be an insufficient result. However when looking at the distribution of both test samples in figure (23) an overlap between (m_W, m_{bb}) and (m_{bb}, m_W) inputs can be seen, where the bb mass is similar to the mass of the W-boson peak.

---

[24]The hard_sigmoid activation in keras, that approximates a sigmoid by a combination of linear functions, and therefore reduces computing time.
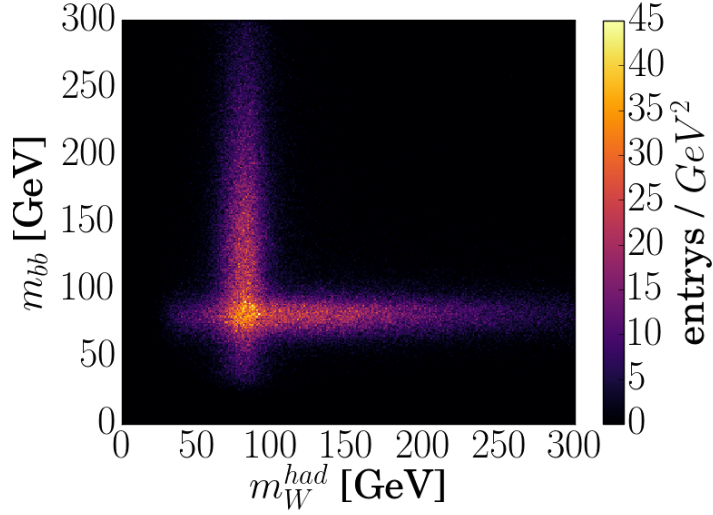
Figure 23: The test sampel of the NN training, with true (m_W, m_{bb}) and false (m_{bb}, m_W) permutations in the (m_W, m_{bb}) mass plane.

With $\approx 13\%$ of the events in the overlap region and an expected discrimination accuracy of 50% in this region, a total accuracy of $\approx 93\%$ is to be expected. This leaves the result with 91.17% as a reasonable result. For this low dimensional 2D case it is possible to look at the network output for the whole input space, as can be seen in figure (24).
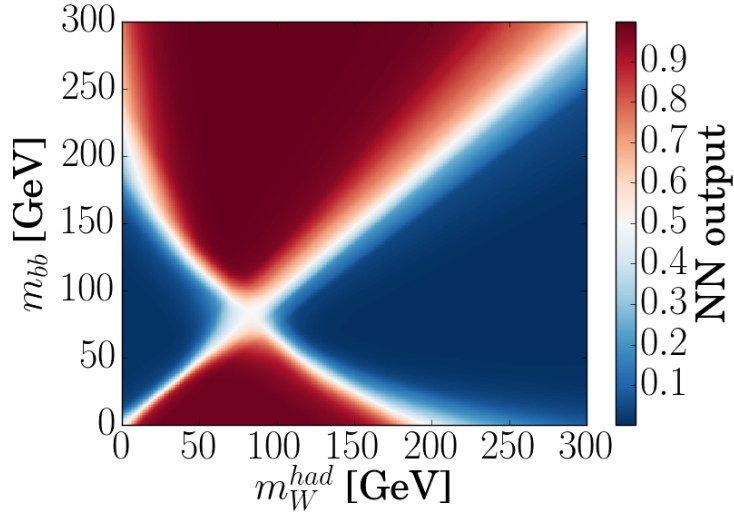
Figure 24: The output of a neural network for a scan in the (m_W, m_{bb}) input space.

This scan proves that the network is indecisive in the overlap region. It can also be seen what happens when the network tries to describe input regions that were not covered by the training input. The category predictions can and should work in regions where training data exists and the dataset was sufficiently big. However predictions outside the coverage of training data should be handled carefully.

To move closer to the physical problem of judging a jet assignment as the next step a neural network was challenged to solve the same task of classifying inputs of the di-jet systems, but this time using four momenta. The network has again to decide if an input has the right assignment,

with the momenta of the *W* system followed by those of the *bb* system, or a wrong assignment with the interchanged order of inputs. The network has to map:

$$(\vec{P}_{wp1}, E_{wp1}, \vec{P}_{wp2}, E_{wp2}, \vec{P}_{b1}, E_{b1}, \vec{P}_{b2}, E_{b2}) \rightarrow (1)$$

$$(\vec{P}_{b1}, E_{b1}, \vec{P}_{b2}, E_{b2}, \vec{P}_{wp1}, E_{wp1}, \vec{P}_{wp2}, E_{wp2}) \rightarrow (0)$$

If the NN understands the input well enough, we should receive a result at least as good as for the 2D Input, since the mass info can be calculated from the four vectors. For this task more layers proved to be helpful. Since the network has to use low level variables only, the need for more layers is to be expected. With four layers having 128 nodes, each using ReLu activations, intermediate dropout layers and a small L2 regularization term with strength 0.0001, an accuracy of 93.36% was achieved. This is fairly close to the best case accuracy we approximated for the 2D input. The additional information in the four vectors seems to be helpful, possibly allowing the network to improve its classification power in the mass overlap region, that is now accessible by the additional input information. To test how much physical insight the trained network contains, it was tested with rotated and lorenz-boosted data[25]. If the network has learned higher level physical quantities like the mass to separate both input types, the classification strength should be similar. An accuracy of only 68.31% shows that this is at least not completely applicable, even though in this case the network is still better than guessing which would lead to a accuracy of 50%. It is interisting to note that the network seemed to have learned non lorenz invariant quantities.

Since the input of the network has 16 dimensions, it is not obvious to find an accurate representation for the network output in the complete input space, as with the input scan for the 2D input. If we again calculate m_W and m_{bb} we can see where the network inputs classified as true and as false lie in the (m_W, m_{bb}) space, see figure (25).
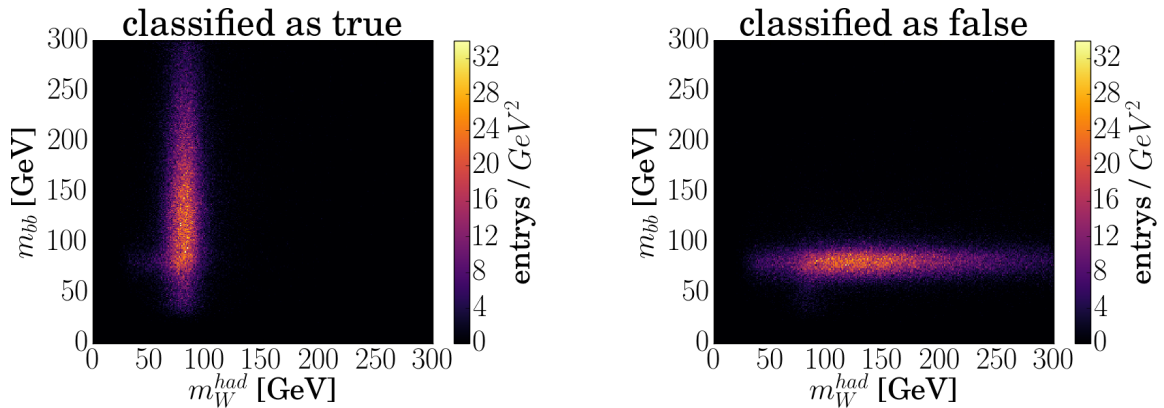
Figure 25: The input events classified true (left) and false (right) for the network trained with the 16 dimensional four vector input.

Except for some deviations in the overlap region, the recived distributions are close to the original mass distributions of the inputs the network had to separate.

---

[25]The Monte-Carlo events where rotated in η and boosted by a four vector, pointing in a random direction with a p_T between 0 and 30 GeV.

The next step is to combine low and high level variables. Combining the inputs of the two neural networks, the next problem tried was to map:

$$(\vec{P}_{wp1}, E_{wp1}, \vec{P}_{wp2}, E_{wp2}, \vec{P}_{b1}, E_{b1}, \vec{P}_{b2}, E_{b2}, m_W, m_{bb}) \rightarrow (1)$$

$$(\vec{P}_{b1}, E_{b1}, \vec{P}_{b2}, E_{b2}, \vec{P}_{wp1}, E_{wp1}, \vec{P}_{wp2}, E_{wp2}, m_{bb}, m_W) \rightarrow (0)$$

If the neural network is able to find more distinguishable features than the invariant masses, the result should be even better than for the 2D and 16D task. With the masses already provided, the best result was achieved with only one hidden layer using 128 nodes, ReLu activations and one dropout layer. The network achieved up to 95.44% accuracy. If again tested on lorenz boosted data, the accuracy drops to 63.4%.

In figure (26) we can see again the position of the events classified true (output > 0.5) and false in the (m_W, m_{bb}) space. The distribution looks now nearly completely similar to the original mass distribution in figure 22 .
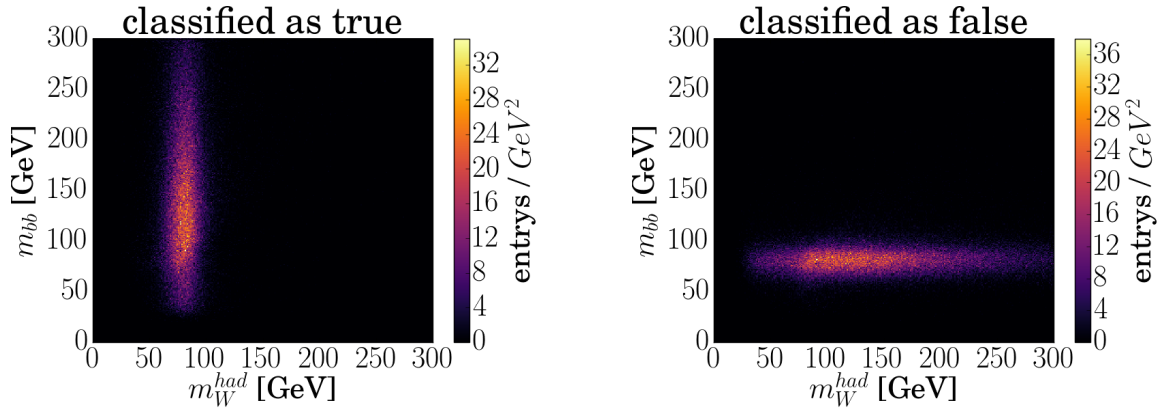
Figure 26: The input events classified true (left) and false (right) for the network trained with the 18 dimensional input including the four vectors and the two masses m_{bb} and m_W.

Until now, the separation problems were chosen to be easily solvable, having only two categories of which one had the reverse input of the other. We saw that small networks with up to four layers could solve the task sufficiently.

The next test is to expand this to a more realistic case. Instead of separating the two di-jet systems, the next task was to identify correctly paired event permutations using the definition in the lepton+jets analysis. In all the following test, dropout layers were found to be limiting the training capability of the network. Instead, small amounts of L2 regularization were used to prevent over training. If we look at the two event categories in the (m_W, m_{bb}) plane as shown in figure (27), we already can see that the W boson mass will be a helpful quantity. The correctly paired permutations have W boson masses close to the W boson peak of 80.4 GeV, while the other permutations are spread over a wide mass range.
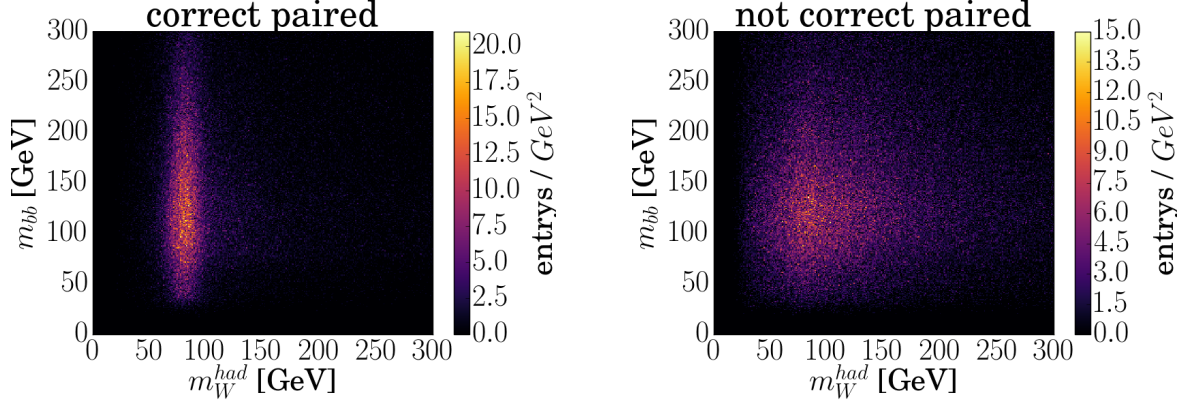
Figure 27: Correct paired (left) and not correct paired permutations in the (m_{bb}, m_W) plane.

Using the input that we used for the previous task, namely the four vectors of the leading jets and the invariant mass of the bb and light jet system we are able to separate correctly and not correctly paired permutations with an accuracy of 69% using four hidden layers with 128 nodes each using ReLu activations. It should be noted that due to computation complexity the hyper parameter search was limited to a maximum number of four hidden layers. It is likely that the optimal number of layers will be higher. To improve this result, the next step was to try different input and input representation e.g. using the (p_T, η, φ, M) representation of the four vectors. Using both the (E, $\vec{p}$) and (p_T, η, φ, M) representation as well as the two invariant masses resulted in a slightly better accuracy of 71%. As can bee seen in figure (28), the network is not completly able to classify events around the W-boson-mass peak.
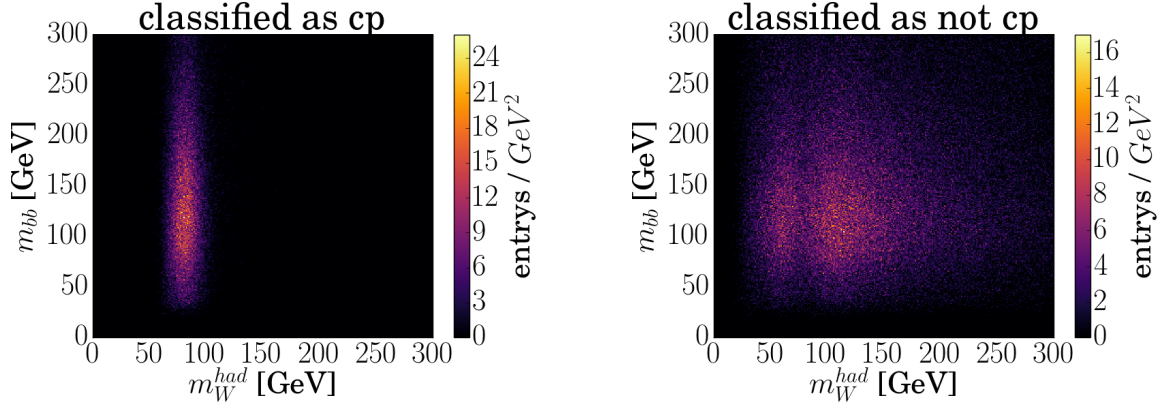
Figure 28: Permutations classified correct paired (left) and not correct paired in the (m_{bb}, m_W) plane with the 34 dimensional network input of four vectors in (E, $\vec{p}$) and (p_t, η, φ, M) representation and two invariant masses.

Conclusion:

For the task of identifying the permutation type a network with a dropout layer is not delivering promising results. The choice of the activation function in the hidden layers does not seem to have big effects on the training outcome. However the ReLu activation function seems to work best. The best results were achieved using the maximum number of hidden layers tested. Therefore a network with more layers could perform even better, provided the number of training data is sufficiently large, since the number of trainable parameters is growing linearly with

the number of layers and quadratically with the number of nodes per layer. Input for the NN that includes low and high level variables deliver good training results. The choice of input quantities is not optimized yet, but the 34 described observables give a good starting point for further optimization. It also should be tested if a classification in all permutation categories is helpful since the network then gets separate training insight into wrongly paired and unmatched permutations. During the test some stability issues concerning the training outcome were seen. If multiple networks with same architecture and the same training data were trained, it was observed that they converged differently due to random initialization. This effect could be reduced using more training data and the ADADELTA optimizer that dynamically adapts the learning rate.

6.2 A neural network discriminant for the classification of permutation types

As discussed in section 4.6, in the current version of the lepton+jets analysis, the top-quark-mass measurement is limited by systematic uncertainties. The current selection offers a event selection that is nearly free from non-$t\bar{t}$ background but has strong contributions from unmatched events and wrongly paired permutations. The correctly paired permutations of $t\bar{t}$ events resemble the signal. Despite the fact that permutations in the unmatched and wrongly paired categories are also described with the templates and their contribution is considered in the likelihood fit, the correctly paired permutations have a better top-quark-mass resolution and should be better described by Monte Carlo samples than the unmatched permutations. This suggests an approach focused on improving the jet assignment, reducing the contribution from wrongly paired permutations and unmatched events. As the following chapter shows, this can be done by a neural network classifier.

6.2.1 Network design and data processing

To solve the classification problem, we use a densely connected feed forward network that is based on the knowledge gained in section 6.1. There, we saw a first attempt for separating correctly paired permutations from unmatched and wrongly paired permutations. The following network setup is based on the networks used there.
The network uses 38 input variables, including :

- m_t^{fit}

- m_W^{reco}

- P_{gof}

- $p_{\text{jet}}^{\text{reco}}$ in (E,$\vec{p}$) and (M, p_T , η, φ) representation for the four leading jets (4×8 inputs)

- m_{bb}

- both b-tag values

This combines a reasonable amount of low and high level variables and provides good results for the classification task. To improve the training performance of the network, the inputs were rescaled, dividing every input by the typical scale of the input variable[26]. The network then consists of ten hidden layers with 128 nodes each, using ReLu activations. Adding Dropout layers was tested, but provided non satisfying results. Instead, a small L2 regularization term with a strength of 0.001 was used. Based on the studies in section 6.1, ADADELTA was used as optimizer during the training. The network has three output nodes with softmax activation, one for each permutation type. Categorical cross entropy was chosen as loss function.

hidden layers: $h_{i+1} = f(w_i \cdot h_i + b_i)$, 10×128 nodes
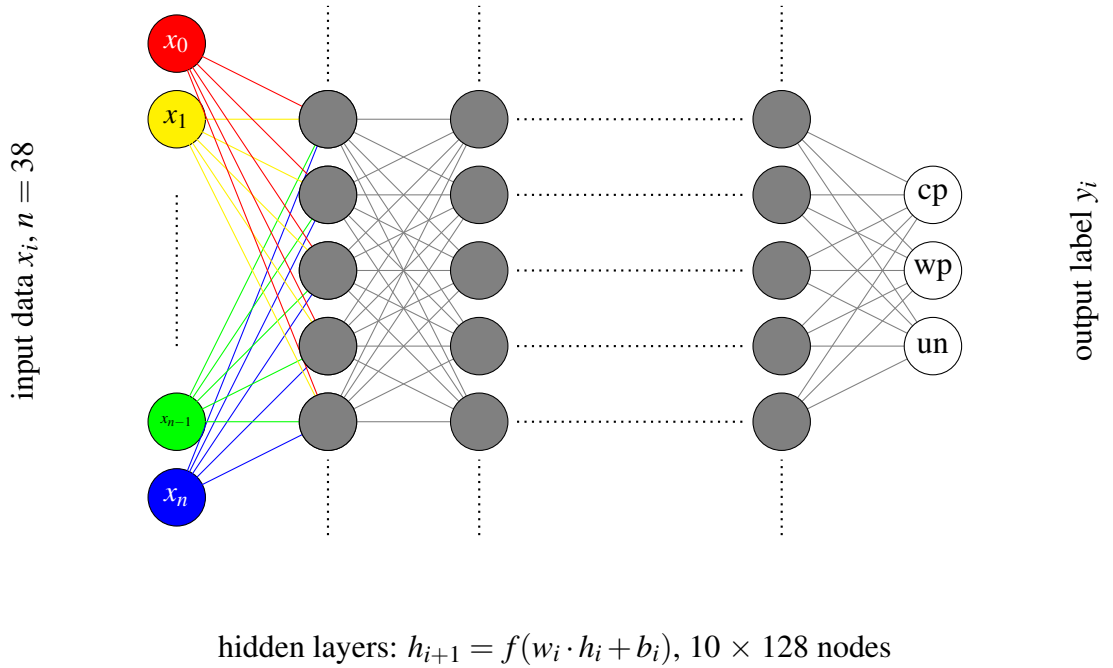
Figure 29: Scheme of the neural network layout used for separating the permutation types of the lepton+jets analysis.

The network was trained on subsets[27] with different m_t^{gen} and JSF of the Monte Carlo samples that are also used for the template calculation in the lepton+jets analysis. This was done to reduce the impact of the generator mass and JSF of the samples used for training and thus to reduce bias effects on the training results. Some more detail on possible bias effects is given in section 6.3.1. It should be mentioned that the design can be further optimized, testing other input combinations or different numbers of layers and nodes. The ten layers that are used give a fairly deep neural network. It can be assumed that also a smaller network would work reasonably well, but since the training time was reasonably low and over training was not a problem, this was not further optimized. The choice of activation functions, loss, optimizer and regularization motivated by the results of section 6.1 has proven to work well. The real optimization

---

[26]i.e. 100 for energies, masses and momenta in GeV and 2π for angular variables.

[27]500000 permutations from every of the 7×5 samples. Only the permutation with the best P_{gof} value in one event was used.

potential for this network lies in the choice of input variables. The current input is aimed at giving the network all information necessary for learning the different jet assignments. Up to now, the input is highly correlated and lacks other event informations, such as additional jets, as well as information about the lepton or the neutrino. Including these might also be helpful in the networks performance. One possible quantity to evaluate the impact of the currently used input variables is the mutual information [61] between the input distributions and the output distribution. It is defined as:

$$I(X;Y) = D_{KL}[p(x,y)||p(x)p(y)] = \sum_{x \in X, \, y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

With $D_{KL}[p||q]$ the Kullback-Liebler divergence of two distributions p and q, $p(x)$ the probability density distribution of x and $p(x, y)$ the joint probability density function of x and y. It describes how much information[28] in terms of bits, two distributions X, Y have in common and can therefore be used to rate the influence of the different input variables on the neural network output. Table (4) shows the mutual information between the different input variables and the neural-network output for the correct permutations category named NNoutput.

Table 4: The mutual information in bits between the input variables of the neural network and the network output for the correctly paired permutations category: NNoutput.

Input ϕ_i	$I(\phi_i; \text{NNoutput})$	Input ϕ_i	$I(\phi_i; \text{NNoutput})$
m_W^{reco}	0.933	$p_x(W_{had, \, jet2})$	0.139
m_t^{fit}	0.676	$M(b_1)$	0.139
P_{gof}	0.566	$E(b_2)$	0.133
$E(W_{had, \, jet1})$	0.282	$p_T(b_2)$	0.133
$p_T(W_{had, \, jet1})$	0.277	$M(b_2)$	0.128
$p_z(W_{had, \, jet1})$	0.232	$p_y(b_1)$	0.122
$p_T(W_{had, \, jet2})$	0.223	$p_x(b_1)$	0.122
$M(W_{had, \, jet1})$	0.221	$p_x(b_2)$	0.114
$E(W_{had, \, jet2})$	0.206	$p_y(b_2)$	0.114
$p_z(W_{had, \, jet2})$	0.182	$\eta(W_{had, \, jet1})$	0.08
$p_x(W_{had, \, jet1})$	0.181	$\eta(W_{had, \, jet2})$	0.07
$p_y(W_{had, \, jet1})$	0.181	$\eta(b_1)$	0.06
$M(W_{had, \, jet2})$	0.180	$\eta(b_2)$	0.06
m_{bb}	0.166	b-tag1	0.05
$p_T(b_1)$	0.154	b-tag2	0.04
$p_z(b_1)$	0.153	$\varphi(b_2)$	0.04
$p_z(b_2)$	0.153	$\varphi(W_{had, \, jet2})$	0.04
$E(b_1)$	0.148	$\varphi(b_1)$	0.04
$p_y(W_{had, \, jet2})$	0.139	$\varphi(W_{had, \, jet1})$	0.04

---

[28]Shannon information.

The results for the mutual information shows that the most important inputs seem to be m_W^{reco}, m_t^{fit} and P_{gof}. Also the momenta, energies and masses of the two light jets making up the W boson have a certain contribution while the b-jets and the angular information play a less significant role.

6.3 Performance of the neural network classifier

The neural network that is described in the previous section was trained in separating the three permutations types cp, wp and un. Cuts on the output node with the cp class score should now allow us to steer the fraction of correctly paired permutations entering the mass-extraction method. Since in the original lepton+jets analysis (see section 4.1) P_{gof} is used as selection criteria, it is reasonable to compare the selection strength of P_{gof} with that of the neural network output. In figure (30) we can see the distribution of the different permutation types in the NNoutput and P_{gof}. The included 2016 data with a center of mass energy of $\sqrt{13}$ TeV and a collected Luminosity of 35.9^{-1} fb shows that data agrees with the Monte Carlo simulation.
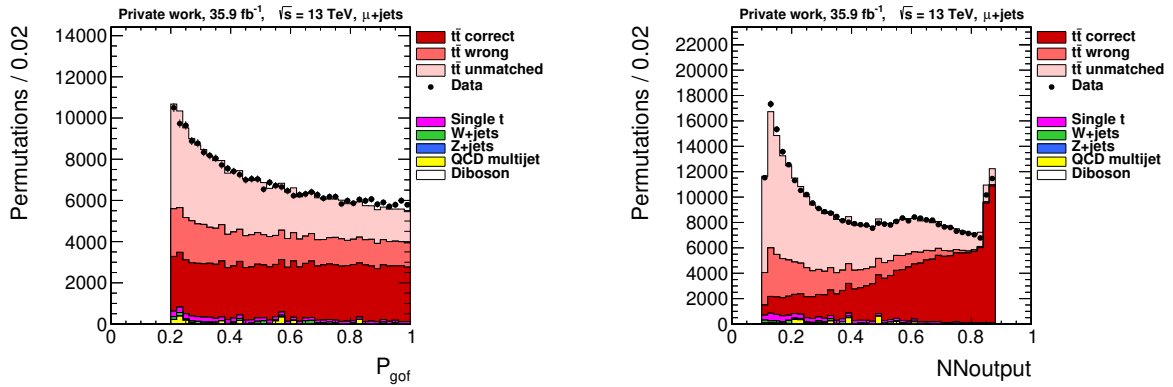
Figure 30: The P_{gof} (left) and NNoutput (right) distribution with the three permutation types cp, wp and un. For values above 0.2 in P_{gof} and 0.108 in the NNoutput respectively .

In both distributions a cut is already applied, those correspondent to the $P_{gof} > 0.2$ cut in the lepton+jets analysis and one of the working points NNoutput > 0.18 that was evaluated for the neural network cuts. Those cuts exclude rather large peaks around zero with huge contribution from unmatched events. The NNoutput distribution ends at ≈ 0.9. This indicates that the network could not find correctly paired permutations that have no overlap with permutations of the other types, optimizing the loss by giving even the clearest permutations a lower output value than one. In both distributions the unmatched permutations accumulate at low values, additionally for the NNoutput the correctly paired permutations accumulate at higher values. This makes the NNoutput a good separator for correctly paired permutations. In figure (31) we can see a comparison of the separation strength for P_{gof} and NNoutput.
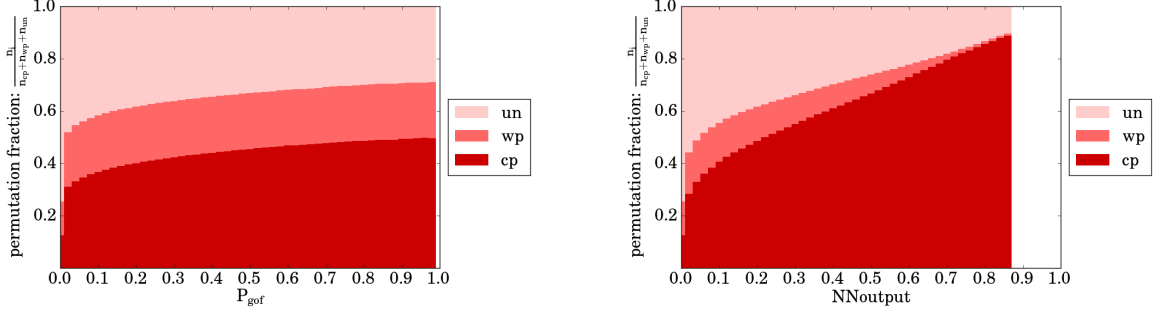
Figure 31: The composition of selected permutations for different cut values on P_{gof} (left) and the NNoutput (right).

In the figure we see the fraction of cp, wp and un permutations for different cut values between 0 and 1 on P_{gof} and NNoutput respectivly. For P_{gof} we see that above the 0.2 cut chosen in the lepton+jets selection the fractions do not change much, while for the NNoutput the fraction of correctly paired permutations increases monotonically with higher cut values, allowing fractions of over 80% for the correctly paired permutations. This shows that the neural network output provides a good classifier for the correct permutations. To test if this can be beneficial to the top-quark-mass measurement one should look not only at the fractions we can achieve with the neural network, but rather at the efficiency of the cp selection and the wp and un rejection, since a fraction of 80% correctly paired permutations would not be sufficient if only $\mathcal{O}(10)$ permutations survive the selection. In figure (32) we see ROC[29] curves comparing cuts on P_{gof} and NNoutput.
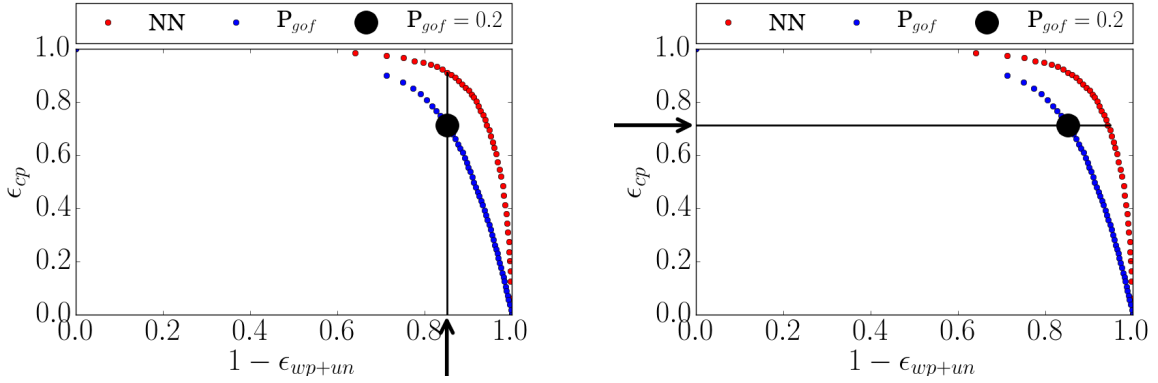
Figure 32: ROC curves showing the trade off between cp selection and wp/un rejection efficiency for different cut values on P_{gof} (blue) and the NNoutput (red). The (ε_{cp}, 1-ε_{wp+un}) point corresponding to the $P_{gof} > 0.2$ selection of the lepton+jets analysis, as well as two working point for a NNoutput cut with the same 1-ε_{wp+un} (left) and ε_{cp} (right) as for the $P_{gof} > 0.2$ selection are marked.

The ROC curve shows that the NNoutput is a way better selector for correctly paired permutations than P_{gof} and offers a better trade off between cp selection efficiency and wp/un rejection. The position of the $P_{gof} > 0.2$ cut from the lepton+jets analysis is marked in the figure. Two exemplary working points with either the same cp selection (right) or wp/un rejection (left)

---

[29]**R**eceiver **O**perating **C**haracteristic

efficiency ar marked. They correspond to cuts of 0.41 and 0.18 on the NNoutput respectively. In table (5) we can see the permutation fractions for baseline selection, P_{gof} cut and the two NNoutput working points.

Table 5: The permutation type wise composition of selected permutations for the baseline selection, the $P_{gof} > 0.2$ selection as in [27] and the NNoutput selections for the two working points 0.182, 0.411.

	baseline	$P_{gof} > 0.2$	NNoutput > 0.182	NNoutput > 0.411
ε_{cp}	1	71.3%	91.1 %	71.3 %
$1-\varepsilon_{wp,un}$	0	85.1%	85.1 %	94.8 %
(cp)	14.8 %	45.1 %	49.6 %	65.6 %
(wp)	8.1 %	15.7 %	14.3 %	5.9 %
(un)	77.1 %	39.3 %	36.1 %	28.5 %

The results in the table show that for low cut values the NNoutput selection behaves like the P_{gof} selection mostly discarding un permutations enhancing the fraction of cp and wp permutations. For higher values the NNoutput can be used to discard bigger fractions of wp permutations. In figure (33) we can see the m_t^{fit} distributions after baseline selection, for the P_{gof} and the two NNoutput cuts.
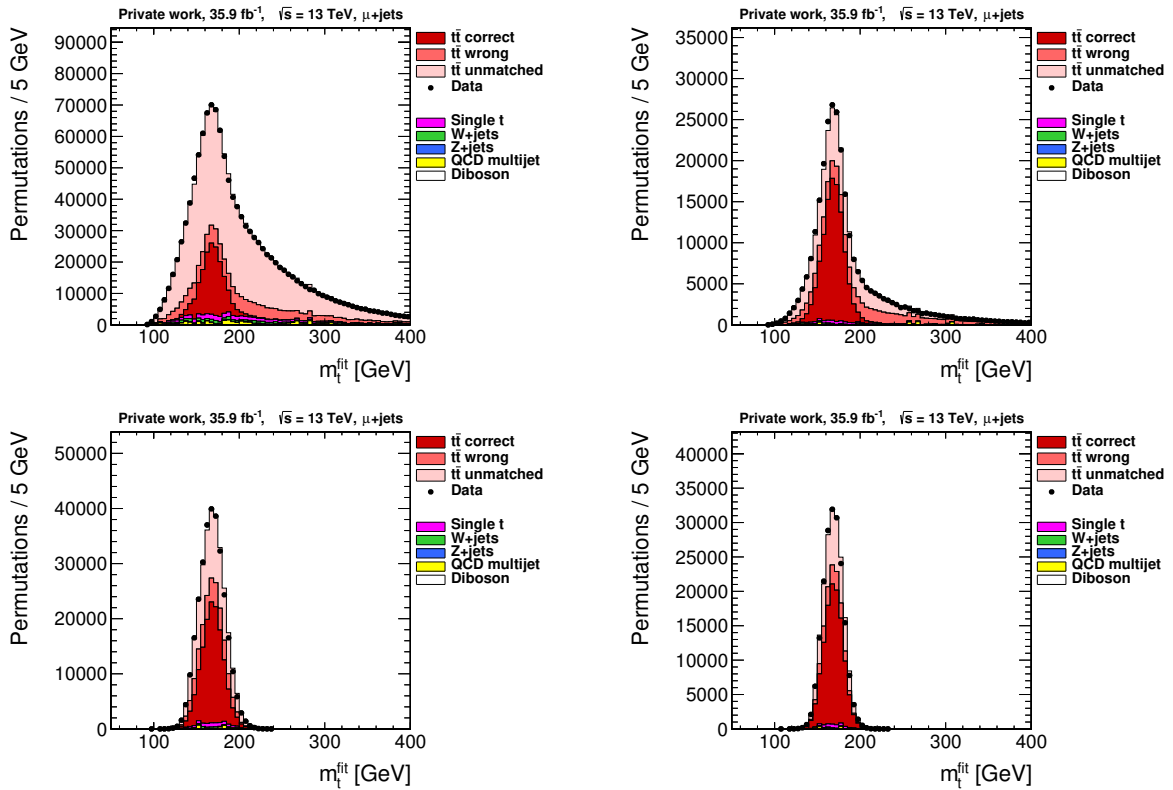
Figure 33: The distribution of correctly paired, wrongly paired and unmatched permutations as well as non-$t\bar{t}$ background in m_t^{fit} for the baseline selection (top-left), the $P_{gof} > 0.2$ selection and the two selected working points of the NN: NNoutput > 0.182 and NNoutput > 0.411.

We can see that for the NNoutput selections only permutations in a narrow peak around the top-quark mass survive. This excludes mainly unmatched permutations.

6.3.1 Controlling for bias effects

Sine the goal is to measure the top-quark mass as precisely as possible, the template method of the lepton+jets analysis is calibrated, compensating bias effects originating from the choice of Monte Carlo samples used for template extraction. Since the calibration mechanism is basically independent from the event selection, comparing generator information and fit results, it should also be able to compensate possible bias effects introduced by a neural network selector. Nonetheless, one should avoid large bias effects that might be difficult to compensate. To check for those, several networks were trained on Monte Carlo samples with different generator masses and JSF. To avoid performing the full template extraction in the optimization phase of the neural network, a simplified bias check was used. The m_t^{fit} distribution was reweighted with the NNoutput value for the different permutations and a simple Gaussian fit was performed to determine the position of the top-quark-mass peak for subsets of every mass sample. The position of the mass peak m_{NN} and the generator top-quark mass was then compared with the generator mass m_{true} in every sample.
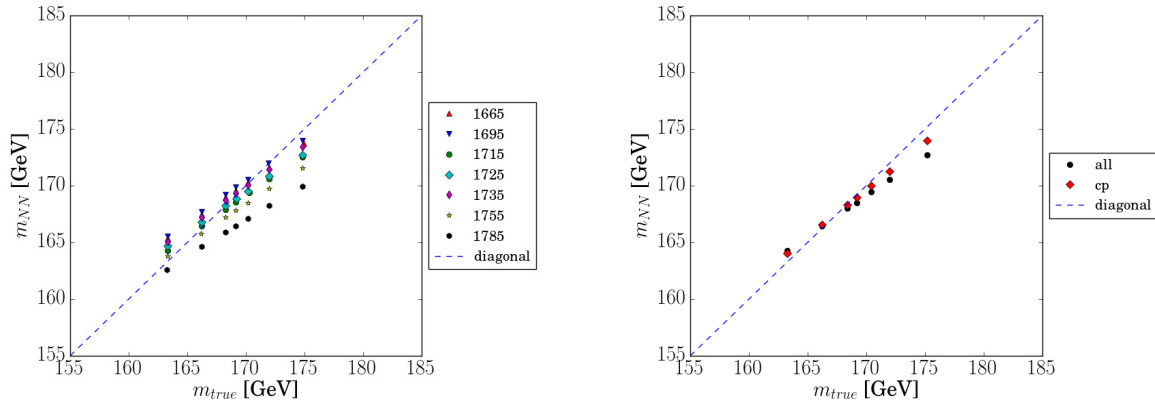
Figure 34: The position of the top-quark-mass peak m_{NN} in the m_t^{fit} distribution of different mass samples reweighted with the output of neural networks trained on samples with different generator-top-quark masses (left) and a network trained on a sample combining multiple generator masses (right) determined with a Gaussian fit plotted against the generator mass of the tested sample m_{true}. On the right plot the (m_{NN}, m_{true}) are also plotted for samples that only consist of correctly paired permutations (cp).

In the left part of figure (34) we can see the result. The (m_{NN}, m_{true}) points for every network lie on straight lines slightly tilted away from the diagonal. One can see that there is a strong dependence on the generator mass sample that was used for training as well as a deviation from the diagonal, leading to a mass dependent deviation of extracted and generator mass, that has to be corrected by a calibration. Similar dependencies were observed when looking at the JSF of the training samples. This led to the decision to train on samples with all generator masses and JSF values available drawn from the samples that are used for the template generation in the lepton+jets analysis. In the right part of figure (34) we can see a slight improvement when training on all samples. The (m_{NN}, m_{true}) points lie on a straight line close to the diagonal. If we look only at extracted masses from correctly paired event permutations, the dependence on the generator mass is smaller. The remaining bias then can be removed in the calibration step

of the template-fit method. This can be seen in figure (35) where the complete top-quark-mass extraction was performed for a selection on the workingpoint NNoutput > 0.41065.
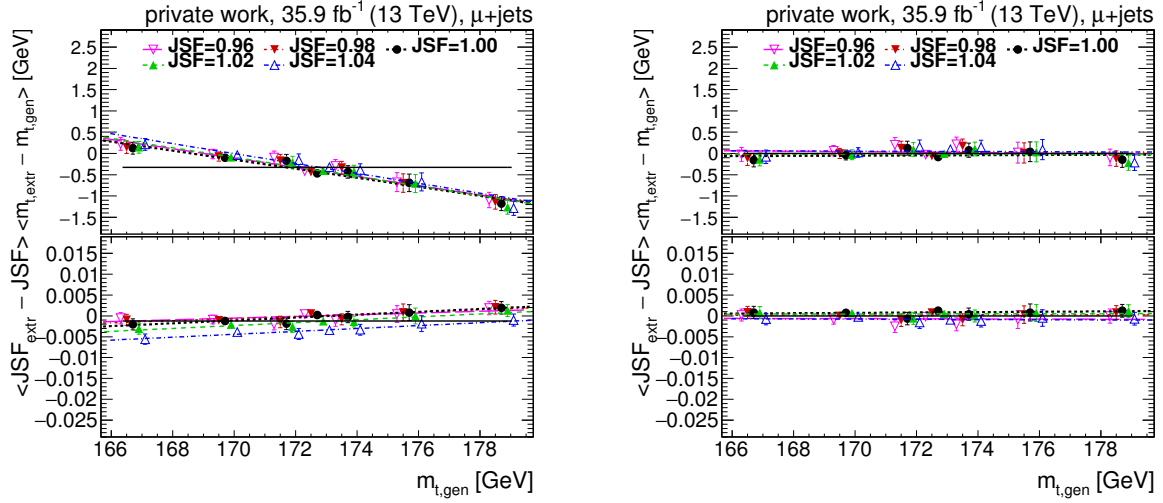
Figure 35: The difference in extracted and generator mass (top) and JSF (bottom) for a cut on the neural network output for the cp category: (NNoutput > 0.41065), before (left) and after calibration (right). Dependencies on m_t and JSF and thus possible biases in the mass extraction have been removed.

Another check that should be made, is to look if the new neural network based selection introduces a mass or JSF dependence of the permutation type fractions. These are presumed to be independent of mass and JSF for the $P_{gof} > 0.2$ selection. Thus in the lepton+jets analysis the fractions can enter the likelihood as derived on the "default" mass sample with a generator-top-quark mass of 172.5 GeV and a JSF of 1.00. To check if this assumption holds for a selection with the neural network classifier, the permutation fractions were calculated on subsets of the Monte Carlo samples for different working points of the neural network output. As can be seen in figure (36) the fractions are neither mass nor JSF dependent and the assumption holds.
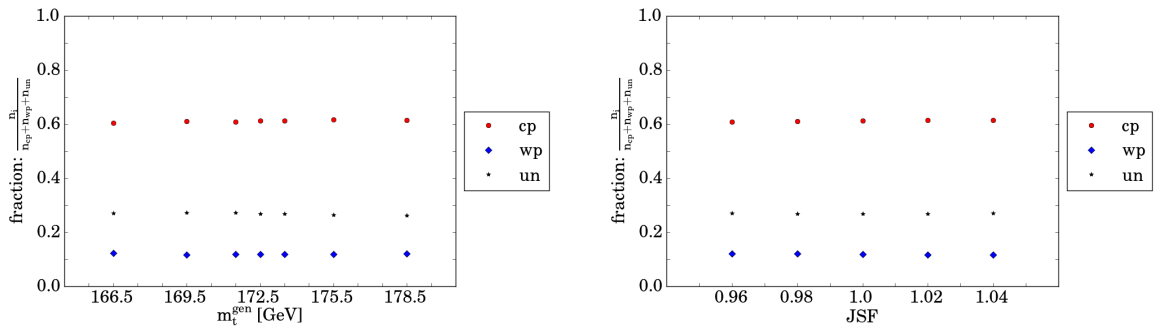
Figure 36: The fraction of correctly paired (cp), wrongly paired (wp) and unmatched permutations evaluated on samples with different generator-top-quark mass (left) and JSF (right) after a cut on the neural network output for correctly paired permutations: NNoutput > 0.41065.

6.3.2   Implications for systematical uncertainties in the top-quark-mass measurement

In the chapters 6.1-6.3 we discussed how we can achieve a better selection for correctly paired permutations. To see how well the cuts on the neural network output work in achieving our

primary goal to improve the precision of out top-quark-mass measurement, the complete top-quark-mass extraction method described in section 4.2 has to be redone, this time with cuts on the neural network output evaluated with respect to the systematical uncertainty. In table (6) we can see the shifts in the top-quark-mass extraction between the default Monte Carlo sample and the systematic samples for the biggest contributers to the systematic uncertainty for three different scenarios, a $P_{gov} > 0.2$ selection, a cut[30] on the network output of NNoutput > 0.411 and a cp-only selection.

Table 6: Shifts in the top-quark-mass extraction between a default sample with a generator top-quark mass of 175.5 GeV and JSF = 1.00 and several systematic samples, that represent the biggest contributers to systematic uncertainties in the top-quark-mass measurement in the lepton+jets channel. The table contains shifts for the 1D, 2D and hybrid extraction method evaluated for the $P_{gof} > 0.2$ selection, a selection with NNoutput > 0.411 and a cp-only selection that selects all correctly paired permutations of the baseline selection. In red/green larger deviations in the shifts for the hybrid mass between the $P_{gof} > 0.2$ and the other selections are marked.

shifts [GeV]	$P_{gof} > 0.2$			NNoutput >0.411			cp only		
	1D	2D	hyb	1D	2D	hyb	1D	2D	hyb
FSR	-0.55	0.31	0.05	-0.46	0.61	0.29 5	-0.5	0.45	0.17
ISR	0.1	0.13	0.11	0.08	0.13	0.12	0.11	0.1	0.1
JEC Flavour	-0.34	-0.42	-0.4	-0.34	-0.44	-0.41	-0.32	-0.38	-0.36
ME Gen.	0.23	0.28	0.27	0.2	0.1	0.13	0.51	0.32	0.38
CR def/"gluon move" (erd on)	0.21	0.29	0.27	0.17	0.27	0.24	0.25	0.4	0.36
CR def/"QCD inspired" (erd on)	-0.22	-0.15	-0.17	-0.29	-0.17	-0.21	-0.24	-0.11	-0.15
CR def ERD off / def ERD on	0.47	-0.15	0.04	0.5	0.06	0.19 4	0.45	-0.18	0.01
JEC (Quad sum)	(0.81)	(0.11)	(0.18)	(0.84)	(0.05)	(0.22)	(0.78)	(0.09)	(0.18)
B jet modl. (Quad sum)	(0.07)	(0.08)	(0.08)	(0.03)	(0.06)	(0.05)	(0.04)	(0.05)	(0.05)
Combined			0.59			0.65			0.69

If we compare the results for the P_{gof} and the NNoutput cut, we can see that some of the shifts for the hybrid mass are larger, some are reduced and the overall uncertainty grows slightly when exchanging the P_{gof} cut with the neural network selection. First of all, this means that this working point of the neural network will not improve the systematic uncertainty compared to the old selection. This can have at least three possible reasons:

1. The network classifier works not as anticipated and introduces bias effects that increase the overall systematic uncertainty.

2. The NN selector selects phase space regions that are especially vulnerable to the systematic variations.

3. The hybrid weight responsible for the hybrid mass is not optimized.

4. The approach of selecting more cp and less wp/un permutations is not helpful with regard to the systematic uncertainty of the top-quark-mass extraction.

Since no obviously unexpected effects could be seen in different control distributions for the $t\bar{t}$ selection as well as the fact that the calibration of the template method should compensate for

---

[30]This cut corresponds to the working point where the cp selection efficiency is as high as for the $P_{gov} > 0.2$ cut.

bias effects, the latter two reasons were studied. The optimization of the hybrid weight showed no qualitative difference in the observed lack of improvement. Therefore the efforts of understanding this result were directed at the systematic uncertainty in a cp only selection. To test the effectiveness of a purer selection, the top-quark-mass measurement was redone with a selection that artificially only selects correct permutations. This is solely possible on Monte Carlo events and is therefore no realistic scenario. But it offers the possibility to look at a selection that models a perfect neural network selection, selecting only cp permutations, while discarding wp and un permutations as well as background completely, without introducing any kind of bias. If we look at the result for the cp selection in table (6), we can see that the overall systematic uncertainty gets even larger. This means that the uncertainty would not improve if the neural network selector would work perfectly regarding to its classification goal and that the other permutation types are helping to make the top-quark-mass determination more stable against the systematical variations. The cp selection selects all permutations that are correctly paired regardless of their other properties, while the P_{gof} cut for example discards permutations that do not fit to the $t\bar{t}$ hypothesis. Therefore also combinations of the cp and the P_{gof} selection were tested. But those resulted all in systematic uncertainties higher than in the original $P_{gof} > 0.2$ selection. Therefore either the correct permutations are especially vulnerable to the systematic variations or the other permutation types have a stabilizing effect on the top-quark-mass extraction. In the following we will see that the latter is the case. The full top-quark-mass extraction was repeated again with selections of only wp and only un permutations. The resulting shifts were indeed higher for the wp and un category, meaning that the cp permutations behave better with respect to the systematical variations. To explain the overall better result when we select also non cp permutations, we have to look at the signs of the top-quark-mass shifts. For some of the systematic variations the shifts of the cp and wp or un permutations have a different sign. Thus having a selection that combines cp, wp and un permutation leads to a partial cancellation of the shifts and an overall smaller systematic uncertainty as observed for the P_{gof} selection. To support this explanation, the 1D shifts were reconstructed using the permutation fractions and the shifts that we got for the cp, wp, un only selections. The shifts measured for the cp, wp and un only selections were combined using a weight derived from the expected statistical uncertainty of the 1D top-quark-mass measurement determined in the pseudo experiments[31]:

$$w_i = \frac{\frac{1}{\sigma_i}}{\sum_j \frac{1}{\sigma_j}} \quad i,j \in [\text{cp, wp, un}] .$$

As can be seen in table (7), this reconstruction of the 1D shifts works for a top-quark-mass extraction on the baseline selection and to a certain degree also for the $P_{gof} > 0.2$ and the NNoutput > 0.411 case. The deviations between the measured and the reconstructed 1D shifts for the $P_{gof} > 0.2$ and the NNoutput > 0.411 cases are to be expected. The cp, wp and un only

---

[31]σ is the width in the top-quark-mass distribution for 1D masses extracted out of pseudo experiments, that emulate the real experiment with the number of $t\bar{t}$ events expected in the 2016 data.

selection include also permutation outside the $P_{gof} > 0.2$ and the NNoutput > 0.411 selection, while those selections influence more then the composition of cp, wp and un permutations.

Table 7: Shifts in the top-quark-mass extraction between a default sample with a generator top-quark mass of 172.5 GeV and JSF=1.00 and several systematic samples. The systematic samples represent the biggest contributers to systematic uncertainty in the top-quark-mass measurement with the ideogram method. The shifts for the 1D mass extraction case are shown for three different selections. Those include the baseline selection (top), the $P_{gof} > 0.2$ (middle) and the NNoutput > 0.411 selection (bottom) as well cp, wp, un only selections based on the baseline selection, the $P_{gof} > 0.2$ and the NNoutput > 0.411 selection. Also reconstructed values for the 1D shifts based on the shifts for the cp, wp and un-only selections and weights given by statistical uncertainty on the top-quark-mass extraction in those categories $w_i = \frac{\frac{1}{\sigma_i}}{\sum_j \frac{1}{\sigma_j}}$ $i, j \in$ [cp, wp, un] are shown. The line for the uncertainty on the chosen Monte Carlo generator is marked for the baseline selections, to show how the partial cancellation of the top-quark-mass shifts works.

shifts [GeV] (baseline)	cp	wp	un	1D rec.	1D
FSR	-0.5	-0.93	-0.73	-0.59	-0.62
ISR	0.11	-0.22	0.24	0.11	0.12
JEC Flavour	-0.32	-0.45	-0.52	-0.38	-0.34
ME Gen.	0.51	0.26	-2.58	-0.19	-0.18
CR def/"gluon move" (erd on)	0.25	0.24	0.22	0.24	0.23
CR def/"QCD inspired" (erd on)	-0.24	0.54	0.05	-0.1	-0.22
CR def ERD off / def ERD on	0.45	-0.58	0.67	0.4	0.45
JEC (Quad sum)	(0.78)	(0.93)	(1.06)	(0.85)	(0.82)
B jet modl. (Quad sum)	(0.04)	(0.36)	(0.67)	(0.14)	(0.12)
weight w_i	0.689	0.093	0.218		
shifts [GeV] ($P_{gof} > 0.2$)	cp	wp	un	1D rec.	1D
FSR	-0.5	-0.93	-0.73	-0.56	-0.55
ISR	0.11	-0.22	0.24	0.1	0.13
JEC Flavour	-0.32	-0.45	-0.52	-0.35	-0.34
ME Gen.	0.51	0.26	-2.58	0.17	0.23
CR def/"gluon move" (erd on)	0.25	0.24	0.22	0.25	0.21
CR def/"QCD inspired" (erd on)	-0.24	0.54	0.05	-0.15	-0.22
CR def ERD off / def ERD on	0.45	-0.58	0.67	0.39	0.47
JEC (Quad sum)	(0.78)	(0.93)	(1.06)	(0.82)	(0.81)
B jet modl. (Quad sum)	(0.04)	(0.36)	(0.67)	(0.04)	(0.07)
weight	0.819	0.077	0.104		

shifts [GeV] (NNoutput> 0.411)	cp	wp	un	1D rec.	1D
FSR	-0.5	-0.93	-0.73	-0.54	-0.46
ISR	0.11	-0.22	0.24	0.11	0.08
JEC Flavour	-0.32	-0.45	-0.52	-0.34	-0.34
ME Gen.	0.51	0.26	-2.58	0.26	0.2
CR def/"gluon move" (erd on)	0.25	0.24	0.22	0.25	0.17
CR def/"QCD inspired" (erd on)	-0.24	0.54	0.05	-0.18	-0.29
CR def ERD off / def ERD on	0.45	-0.58	0.67	0.42	0.5
JEC (Quad sum)	(0.78)	(0.93)	(1.06)	(0.81)	(0.84)
B jet modl.	(0.04)	(0.36)	(0.67)	(0.04)	(0.03)
weight	0.878	0.043	0.079		

These results show that there is no improvement in the systematic uncertainty of the top-quark-mass measurement to be gained from enhancing the fraction of correctly paired permutations. Nonetheless, the results give a deeper understanding of the connections between the permuta-

tion types and the systematic uncertainties and possibly offer new optimization potential since a selection with a optimized composition of the three permutation types might give a better result. The composition would then be a optimizable parameter like the hybrid weight in the determination of the hybrid mass.

6.3.3 Comparing the neural network approach and previous 2016 analysis

The tested working point on the neural network output NNoutput > 0.411 did not result in a better systematic uncertainty. In section 6.3.2 we understood why a purified selection of correctly paired permutations did not lead to an improved uncertainty, but we also saw that the uncertainty is closely connected to the composition of cp, wp and un permutations. Therefore, the full top-quark-mass extraction was performed multiple times to optimize the selection with the NNoutput variable. The mass extraction procedure including the determination of the permutation fractions, the template generation, the template calibration and the calculation of the systematic shifts for the hybrid mass was performed at four different working points. Assuming the correlation between systematic uncertainty and the cut value of the NNoutput can be described by a quadratic function, the results of the four working points can then be used to determine the optimal cut value for the NNoutput. To get comparable results with the results of the lepton+jets analysis, also the hybrid weight has to be optimized as well, requiring additional mass extractions for different hybrid weights. The result of the overall systematic uncertainty can be seen in figure (37).
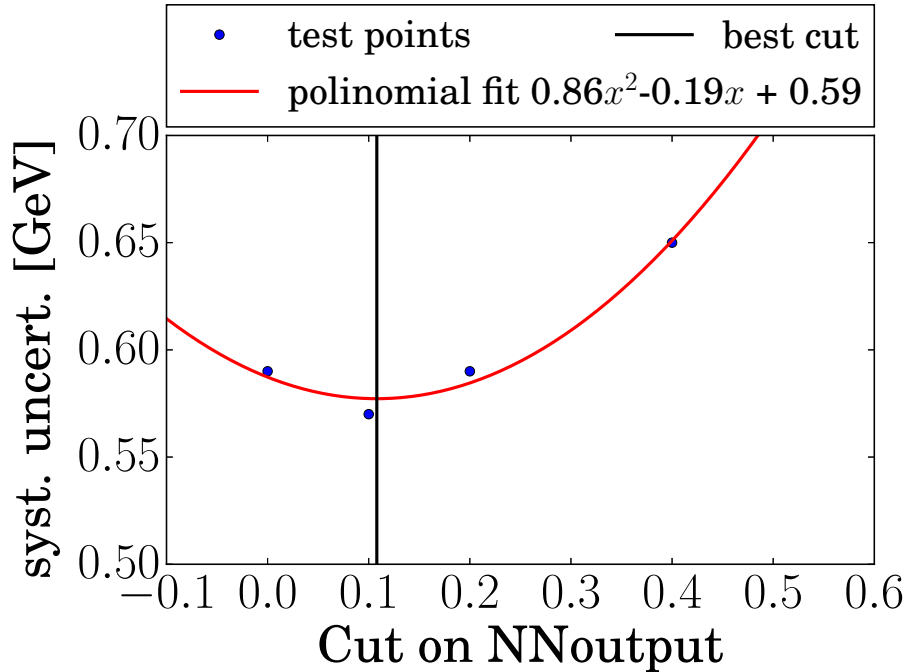
Figure 37: The systematic uncertainties of the top-quark-mass measurement for four different working-points on the neural network output. A quadratic fit was performed to find the optimal cut value at 0.108.

The quadratic fit lead to an expected optimal cut value of NNoutput > 0.108. If we again perform the top-quark-mass extraction for this selection and optimize the hybrid weight to $w_{\mathrm{hyb}} = 0.214$,

we get a systematic uncertainty of 0.56 GeV and a expected statistic uncertainty of 0.09 GeV. In this determination of the systematic uncertainty, some[32] of the smaller contributors to the systematic uncertainty of the $P_{gof} > 0.2$ selection were neglected. Therefore the result is to be taken with the necessary caution, since the neglected uncertainty sources might have a bigger contribution for the NNoutput selection. The complete list of systematic uncertainties for the NNoutput > 0.108 selection can be found in table (8). The same table for the $P_{gof} > 0.2$ selection can be found in table (9). For the $P_{gof} > 0.2$ selection a systematic uncertainty of 0.63 GeV and a expected statistical uncertainty of 0.1 GeV was achieved, in agreement with the original lepton+jets analysis using all uncertainty sources. Therefore the NNOutput can in fact be used as a better selector than the P_{gof} with regard to the systematic uncertainty.

Table 8: Final systematic uncertainties for the top-quark-mass extraction in $t\bar{t}$ events with the muon+jets final state for a selection with the NNoutput > 0.108

Uncertainty	δm_{1D} [GeV]	δm_{2D} [GeV]	δm_{hyb} [GeV]
Experimental uncertainties:			
JEC	0.81	0.1	0.11
Jet energy resulution	0.03	0.07	0.05
b tagging	0.01	0.03	0.02
Pileup	0.01	0.07	0.05
Modelling uncertainties:			
JEC Flavour	0.34	0.44	0.42
b-jet modelling	0.04	0.06	0.06
ME/PS matching	0.04	0.08	0.06
ME generator	0.21	0.10	0.09
ISR PS scale	0.05	0.05	0.05
FSR PS scale	0.57	0.38	0.17
p_t	0.01	0.01	0.01
Underlying event	0.09	0.07	0.06
Color reconnection	0.70	0.27	0.27
total systematic uncertainty	1.31	0.68	0.56

---

[32]The uncertainty on the non $t\bar{t}$ background, the PDF uncertainties, the uncertainty on the renormalization and factorization scale, the uncertainty due to early resonance decays and the method calibration.

Table 9: Systematic uncertainties in the muon+jets top-quark-mass analysis for a selection with $P_{gof} > 0.2$

Uncertainty	δm_{1D} [GeV]	δm_{2D} [GeV]	δm_{hyb} [GeV]
Experimental uncertainties:			
JEC	0.82	0.1	0.19
Jet energy resulution	0.03	0.08	0.05
b tagging	0.02	0.04	0.04
Pileup	0.02	0.09	0.06
Modelling uncertainties:			
JEC Flavour	0.35	0.43	0.4
b-jet modelling	0.05	0.07	0.06
ME/PS matching	0.05	0.10	0.08
ME generator	0.08	0.33	0.25
ISR PS scale	0.09	0.06	0.07
FSR PS scale	0.59	0.24	0.10
p_t	0.02	0.02	0.02
Underlying event	0.08	0.07	0.06
Color reconnection	0.69	0.33	0.30
total systematic uncertainty	1.30	0.72	0.63

To improve the trust in this result and to gain understanding about the physical background of this improvement we can look at how the input distributions of the neural network changed due to the selection NNoutput > 0.108 and how they compare to those of the P_{gof}>0.2 selection. In the following figures (38,39,40) the distributions can be seen for the most influential inputs m_W^{reco}, $m_t^{fit}$and P_{gof}. The other input distributions can be found in the appendix (section 9.1). When we look at these control distributions we can see that the distributions for the network selection look very similar to those of the P_{gof} selection. Only in the m_t^{fit}distribution one can see that the network selection cuts away most of the tail in the mass distribution while in the other distributions the network selection seems to be looser than the $P_{gof} > 0.2$ selection allwoing for more wp and un permutations. Despite this, no obvious differences between the distributions of the P_{gof} and the NNoutput selection can be seen in the 1D distributions of the network input variables.
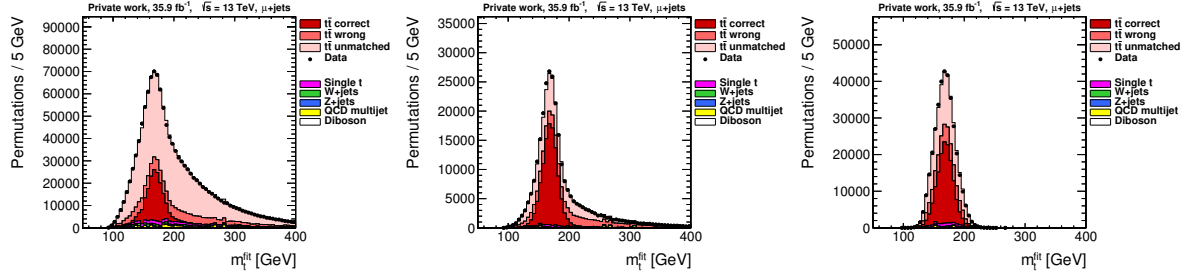
Figure 38: The m_t^{fit} distribution for the baseline selection (left), the P_{gof} selection (middle) and the best neural network selection (right).
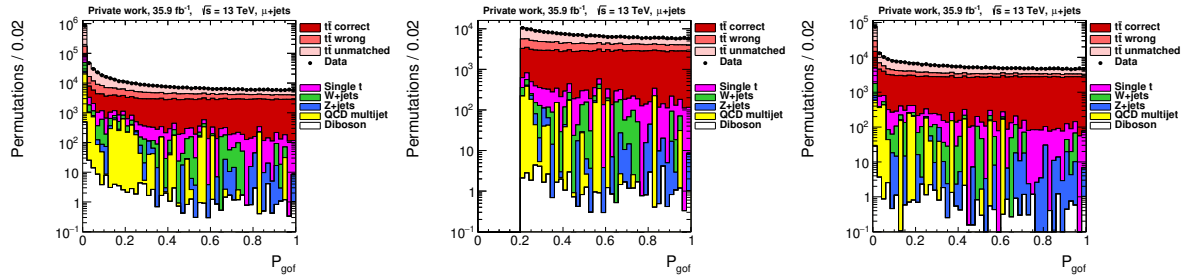
Figure 39: The P_{gof} distribution for the baseline selection (left), the P_{gof} selection (middle) and the best neural network selection (right).
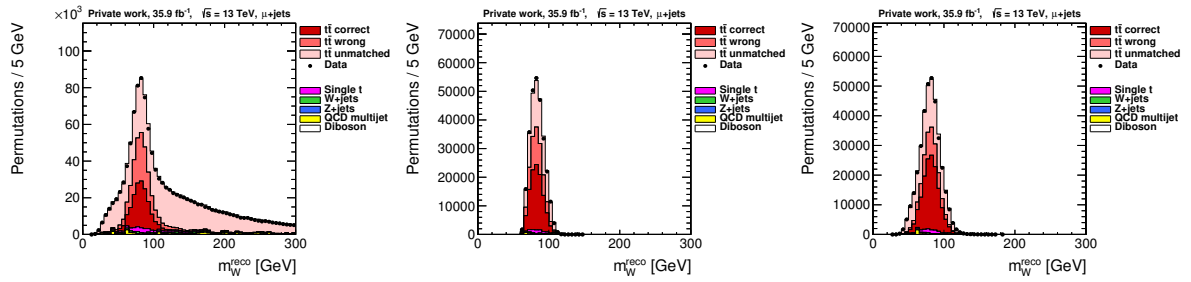
Figure 40: The m_W^{reco} distribution for the baseline selection (left), the P_{gof} selection (middle) and the best neural network selection (right).

7 Outlook

7.1 Proof of principle: An evolutionary learning approach

We now used a neural network to manipulate the permutation fractions in the lepton+jets analysis with the goal to improve the systematic uncertainty of the top-quark-mass measurement. This worked and we found out that the permutation fractions have a strong influence on the systematic uncertainty of the mass extraction method. A more direct approach would be to directly train a network to make an event selection that reduces the systematic uncertainty. This idea has a few technical problems. Most learning algorithms are gradient based, trying to compute a gradient on a certain loss function to find weight updates that reduce the loss and therefore move closer to fulfill the learning goal. The systematic uncertainty of the top-quark-mass measurement is a complex quantity that takes around 2-3 days to compute for one event selection[33]. It is not an event based quantity and therefore difficult to use directly in a loss function. The neural network we used in the last sections had approximately 200000 parameters and therefore one would need to know for approximately 200000 weight changes how the systematic uncertainty would change to compute a gradient. A task that is simply not feasible. To possibly solve this, non gradient based machine learning methods were studied. A promising one is evolutionary training [57]. Evolutionary learning resembles a more sophisticated random based learning method. The idea is to create a population of random variants of a neural network model by adding noise to a base network. Those models are then evaluated and given a score how well they performed in the task they are trained to do. In the case of the context of this thesis, this score would represent how well a network based selection reduces the systematic uncertainty, thus giving high scores to models that resulted in low systematic uncertainty. After evaluating the different models, they will be combined to a new model based on their score, that is the base model for a new population of random models in the next epoch. For the actual implementation of this approach in the context of the top-quark-mass analysis the algorithm described in [57] is used:

Algorithm 1 Simple evolutionary learning

1: **Input:** Learning rate α, noise standard deviation σ, intial parameters θ_0, population size n
2: **for** $t = 0, 1, 2, ...$ **do**
3:      Sample random updates $\varepsilon_1, .., \varepsilon_n \sim N(0, I)$
4:      Compute returns $F_i = F(\theta_t + \sigma \varepsilon_i)$ for $i = 1, ... n$
5:      Set $\theta_{t+1} \leftarrow \theta_t + \alpha \frac{1}{n\sigma} \sum_{i=1}^{n} F_i \varepsilon_i$
6: **end for**

In every time step t random noise ε_i is created for the weights of every population member $1, ..., n$. Using the base weights of the current timestep θ_t and the noise scaled with respect

---
[33]Corresponding to $\approx$ 18000 h of CPU time.

67

to the given noise standard deviation σ, network models are created and evaluated by the call to $F(\theta + \sigma \varepsilon_i)$. Evaluated in our case means that the top-quark-mass analysis is being run on permutations that pass a selection based on the output of the newly created network. A permutation passes if the network output is higher than 0.5 using a network with a sigmoid activation function in the output layer. The return value F is the negative logarithm of the systematic uncertainty. This score function $-\ln(\mathrm{err}/2)$ punishes hight uncertainties and rewards low uncertainties. For this first test, the described algorithm is slightly modified based on an algorithm used in the python package evostra [50] was used. There, the scores F_i are rescaled before combining the weight updates, giving more gradient looking like update steps.

$$F_i = \frac{F_i - 1/2 \sum F_i}{RMS(F_i - 1/2 \sum F_i)} \ .$$

Since the usual estimation of the systematic uncertainties takes 2-3 days, a simplified and automated version of the top-quark-mass analysis was designed, trading the number of created pseudo experiments and therefore accuracy on the uncertainty estimation for a faster runtime of around 12-24h[34]. This allowed for the creation and evaluation of ≈ 5 neural networks in parallel. If an epoch is aimed to be run in around 2 days, this gives around 10-15 networks per epoch. Since only one direction in the weight space can be explored per population member, the used NN had to be much simpler than the network used in section 6.3. A network with only two hidden layers with ten nodes each and a total of 7 inputs was chosen. The fitted top mass m_t^{fit}, the reconstructed W mass m_W^{reco} and the fit probability P_{gof} were included, since they had a strong influence on the training for the network aimed at the permutation classification task. Also the azimuthal distances $\Delta\varphi$ between the b jets and the muon, as well as between the b jets themselves and between the bb system and the reconstructed neutrino were included to provide some information about the event topology. A schematic of the used network can be seen in figure (41).

---

[34]Corresponding to a CPU time of ≈ 1000 h.

hidden layers: $h_{i+1} = f(w_i \cdot h_i + b_i)$, 2×10 nodes
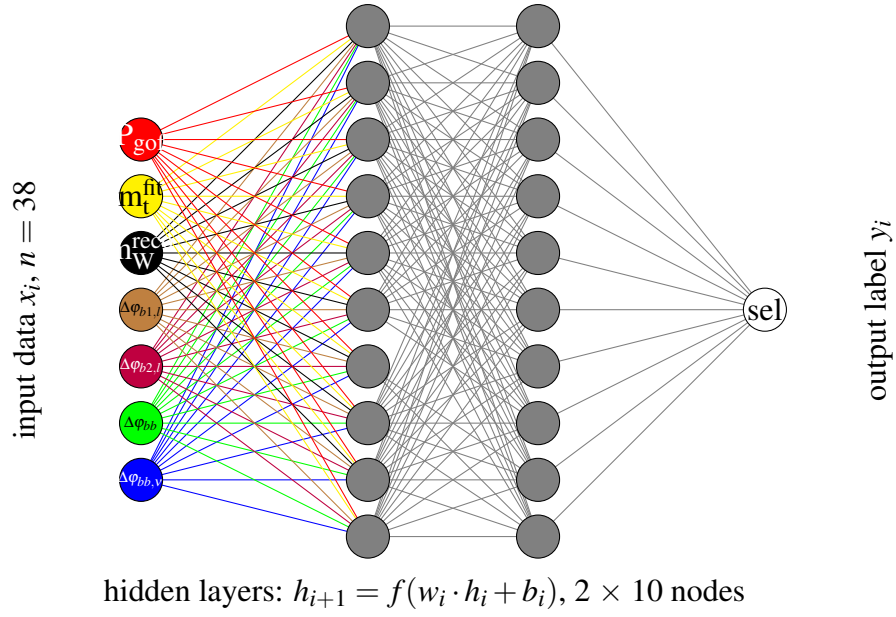
Figure 41: The schematic neural network layout used with the evolutionary training algorithm for the reduction of systematic uncertainties.

A first test showed that a pretraining was needed since a completly randomly initialized network was not allowing enough permutations to pass the selection. The network was pre trained for a few epochs on a classification task separating again correctly paired permutations from non cp permutations, before running 5 epochs of the evolutionary learning approach with populations of 10 | 10 | 14 | 9 | 9 networks. Despite the fact that the network still includes ≈ 250 parameters, an improvement of the systematic uncertainty could be seen during the training. The resulting learning curve can be seen in figure (42).
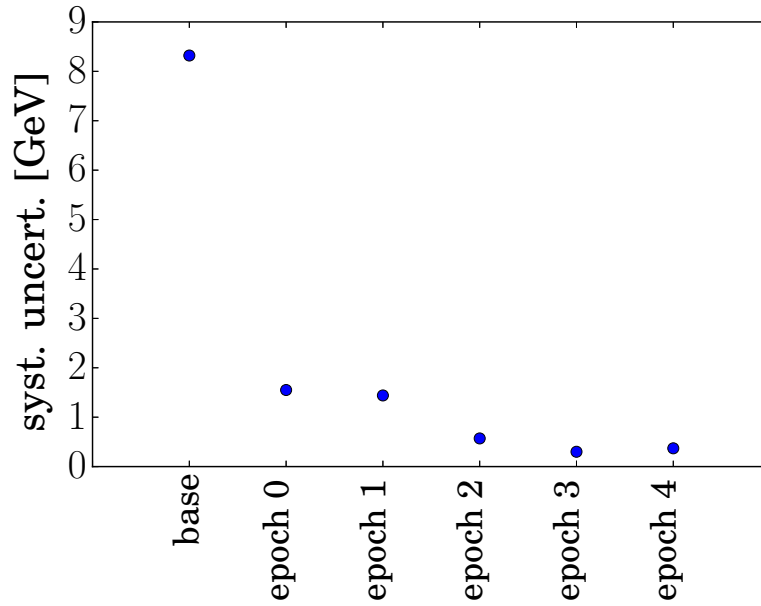
Figure 42: Learning curve of the neural network trained for the reduction of systematic uncertainty with an evolutionary learning approach. The combined systematic uncertainty of most contributors to the overall uncertainty estimated by the short analysis can be seen for the network model obtained from pre training (base) and after multiple epochs of evolutionary training (epoch0-epoch4).

For the learning process a noise standard deviation of $\sigma = 0.01$ and a learning rate of $\alpha = 0.001$ was used and was kept constant over the whole learning process. Even without optimizing the learning rate and the noise spread an incremental learning process can be observed. The best uncertainty observed was in the range of 300 MeV. However this result has to be taken with caution, since the uncertainties seen in figure (41) are those received in the simplified analysis. It was found that the templates used in the simplified analysis lead to significant differences in the result for the systematic uncertainty achieved with the final networks when compared to the full analysis, that only lead to an systematic uncertainty of ≈ 700 MeV. Therefore, the result achieved with this training is not yet directly applicable. Nonetheless it can be concluded that the training algorithm itself achieved a learning process for a complicated problem in a technically feasible manner. To validate the possible usage of a evolutionary training method in the context of systematic-uncertainty optimization, the training itself needs to be repeated with full treatment of systematic uncertainties including an adapted template extraction.

7.2 Further Ideas

A permutation selection that is independent from systematic variations, possibly achieved with adversarial neural networks was also discussed during the time of my thesis. However since the top-quark-mass measurement presented in this work is not event based, it is unclear if this would lead in fact to lower systematical uncertainties. The systematic uncertainties are given by shifts in the extracted top-quark-mass under systematical variations. This means that the extracted templates point to a different top-quark mass before and after a systematical variation. Therefore the systematical uncertainties arise from the description power of the templates when used on MC permutations after undergoing systematical variations.

If now the templates would be constructed so that they describe permutations independent from systematical variations, per construction the systematical uncertainties would be low. For this approach the systematical variations would need to be done for all 5×7 MC samples used in the current template creation and calibration. However the success of this approach will be limited if this results in loosend templates that do not describe the top-quark-mass dependence of the described distributions in a sufficient way only leading to a trade of between calibration and systematical uncertainty. Therefore the template creation could possibly be assisted or replaced by neural networks.

8 Summary

In this thesis two different approaches to reduce the systematic uncertainty of the top-quark-mass measurement with the help of neural networks were shown. The first approach focused on the three different permutation types found in the lepton+jets based top-quark-mass extraction. There are correctly paired permutations where all four leading jets of the $t\bar{t}$ event are assigned correctly, wrongly paired permutations where the two b-jets are interchanged and unmatched events, where at least one of the four leading jets does not originate from a top-quark decay. A deep neural network was successfully trained to classify jet permutations in the $t\bar{t}$ system as correctly paired, wrongly paired or unmatched permutations. With the help of this neural network, the tuning of the permutation fractions in the selection for the mass measurement is possible.

This thesis has shown that an enrichment of correctly paired permutations does in fact not lead to a smaller systematic uncertainty. Permutations of the other permutation types are reacting differently to the variations in the estimation process for the systematic uncertainty. This leads to a stabilization of the measured top-quark mass against systematic variations depending on the permutation type fractions in the event selection of the mass measurement. The strong influence of the permutation type event composition on the systematic uncertainty then could be used to find a selection that reduces the systematic uncertainty in the top-quark-mass measurement in the μ+jets channel to 0.56 GeV without loosing statistical accuracy. Compared to the result for the P_{gof}>0.2 selection this is an reduction of the systematic uncertainty of 12.5 % or 70 MeV.

The second approach is based on a randomized learning technique called evolutionary learning and was used to train a neural network directly as a selector that reduces the systematic uncertainty. The first test of this approach included a unfitting template extraction, so that the otherwise promising results are not yet directly applicable. Nontheless the tests performed for this thesis suggest that evolutionary learning in general is a technically feasible method, motivating further studies.

9 Appendix

Muon and electron definition

Muon Definition The selected muons have to pass the following requirements[27]:

- The muon fired a single muon trigger `HLT_IsoMu24_v*` or `HLT_IsoTkMu24_v*`

- $p_T > 26$ GeV and $|\eta| < 2.4$

- PFMuon ID `tight`, that means:

  - Selected muon is a global muon, reconstructed in both tracker and muon system
  - Found valid hits in at least 5 tracker layers, 1 pixel layer and at least one valid hit in the muon system were found the muon system
  - Global track fit resulted in a χ/ndf < 10
  - The impact parameter with regard to beam spot | dB | is < 0.02 cm
  - The distance to the primary vertex in z direction smaller than 0.5 cm

- The muon has to be isolated in a cone of $\Delta R < 0.4$:
  PF-based combined relative isolation with $\Delta\beta$ correction ($\sum p_T$ (ch.had from PV) + max($\sum E_T$ (neut.had) + $\sum E_T$ (photon) - $0.5 \cdot \sum p_T$ (ch. had from PU), 0))/p_T (μ)
  `(sumChargedHadronPt+max((sumNeutralHadronEt+sumPhotonEt-0.5*sumPUPt),`
  `0.0))`/$p_T < 0.15$

Electron Definition: The selected electrons have to pass the following requirements[27]:

- The electron fired the single electron trigger `HLT_Ele32_eta2p1_WPTight_Gsf_v*`

- $E_T > 34$ GeV and $|\eta| < 2.1$

- The supercluster η hast to be < 1.4442 or > 1.5660 to exclude the transition region between detector barrel and endcap

- The impact parameter with regard to beam spot | dB | has to be < 0.02 cm

- The impact parameters have to be: d0 < 0.05 cm and dz < 0.1 cm in the barrel region and d0 < 0.1 cm and dz < 0.2 cm in endcap region

- The electrons fulfill tight cut-based ID

- Photon conversion is vetoed

- The electron has to be isolated in a cone of $\Delta R < 0.4$:
  `chargedHadronIso+max((neutralHadronIso+photonIso-0.5*puChargedHadronIso),`
  `0.0))`/$p_T < 0.1$

## 9.1 Input distributions of the neural network selector
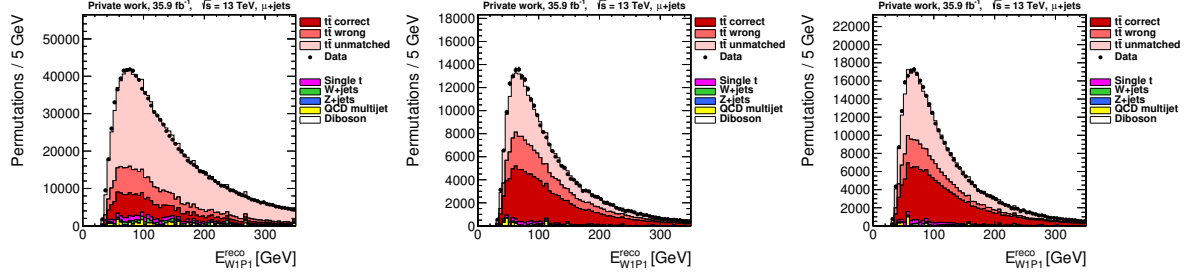
Figure 43: E of the 1st reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection
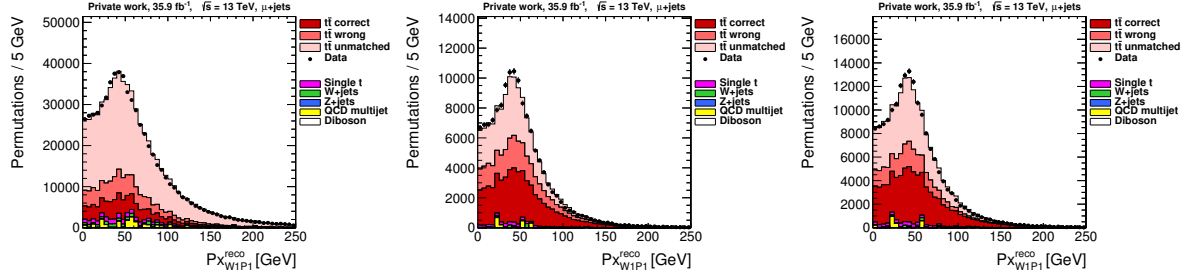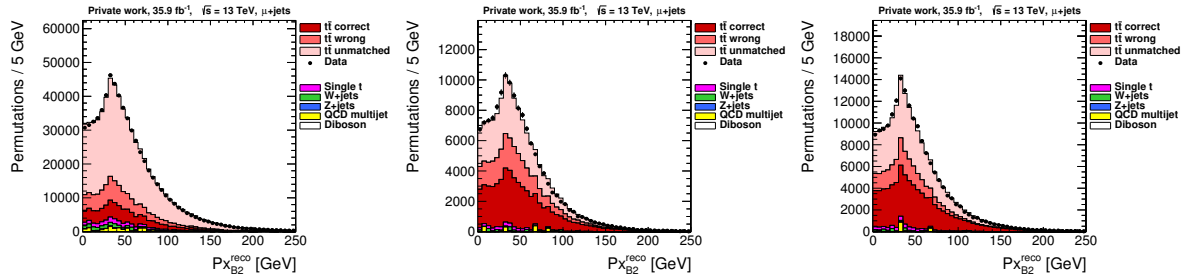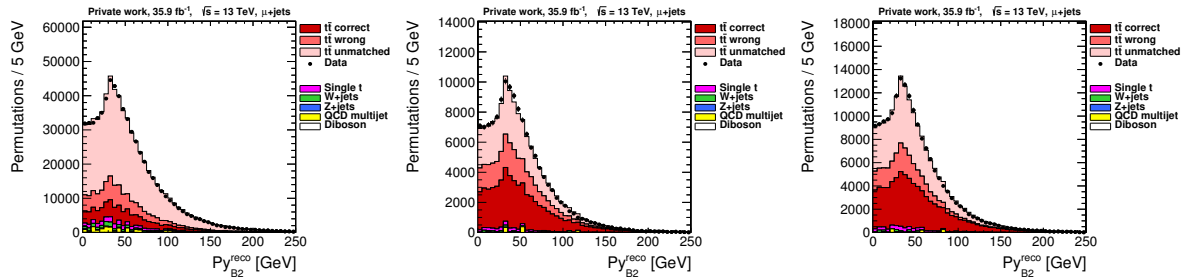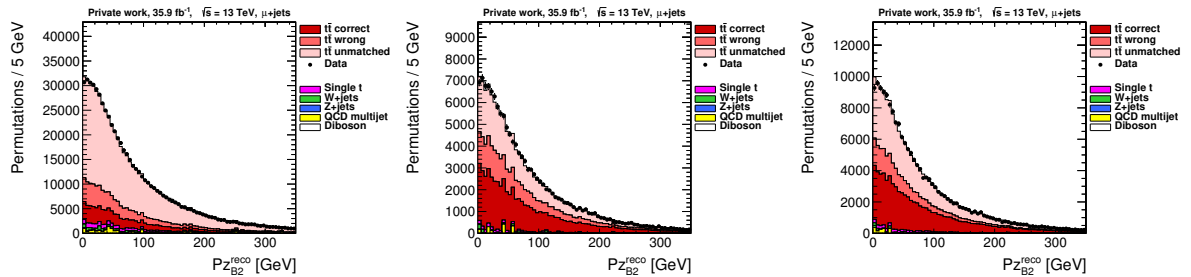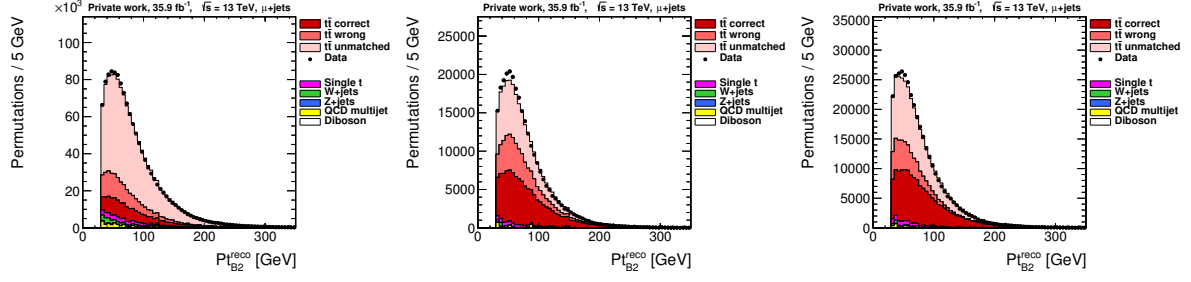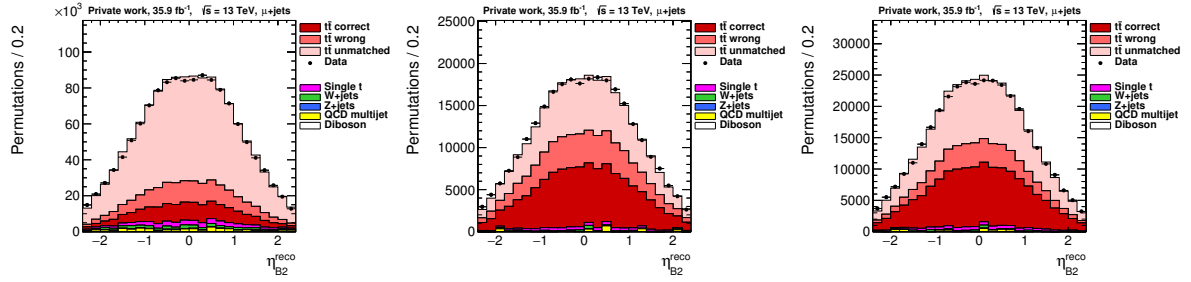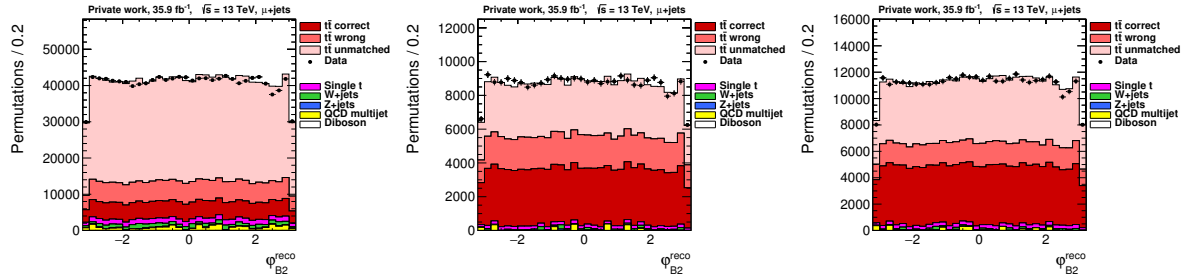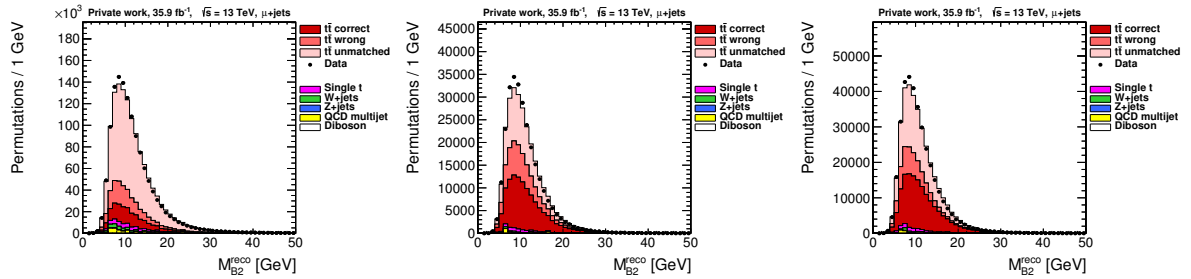

Figure 44: p_x of the 1st reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection
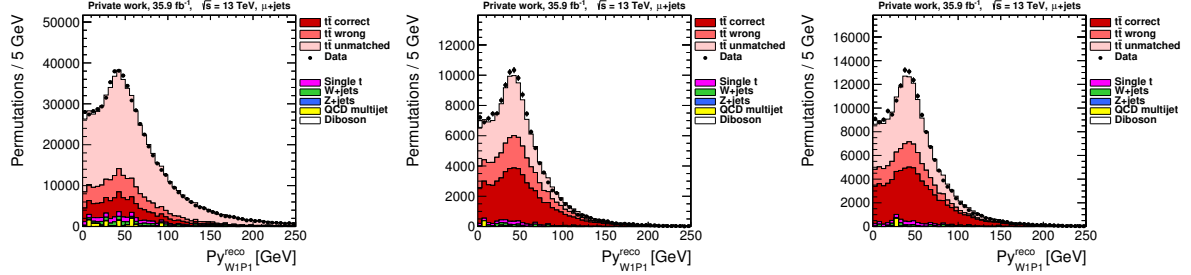
Figure 45: p_y of the 1st reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection
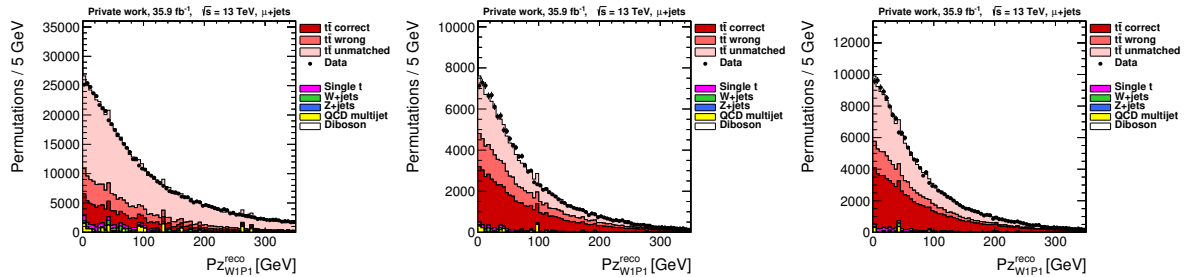
Figure 46: p_z of the 1st reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
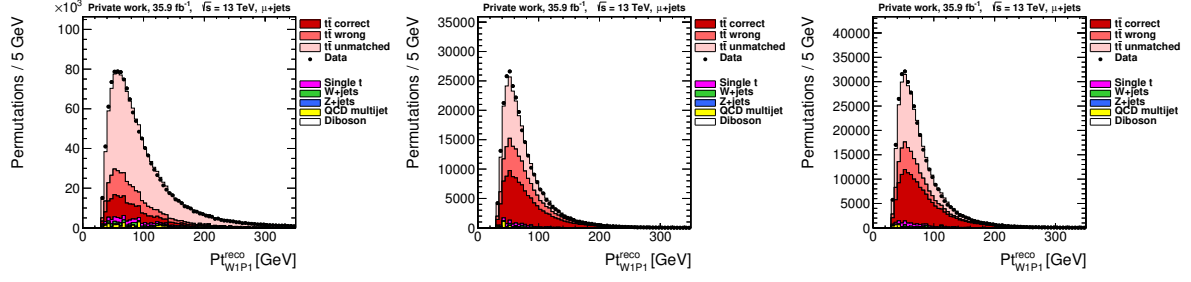
Figure 47: p_T of the 1st reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
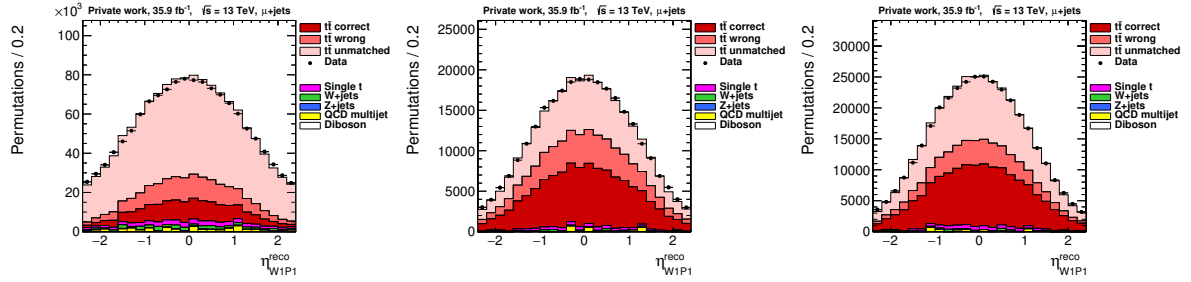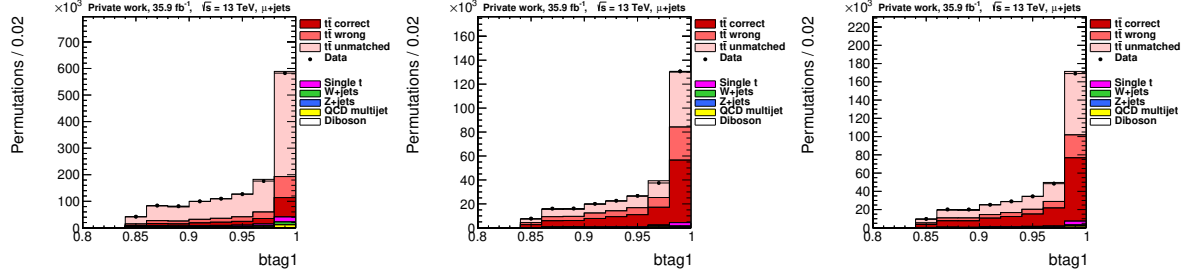

Figure 48: η of the 1st reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
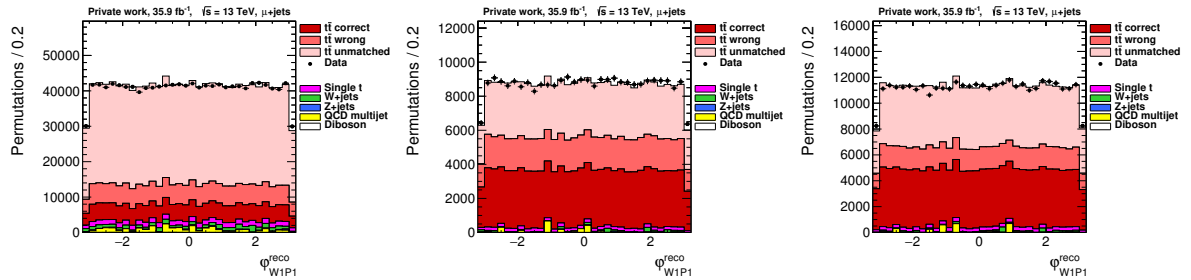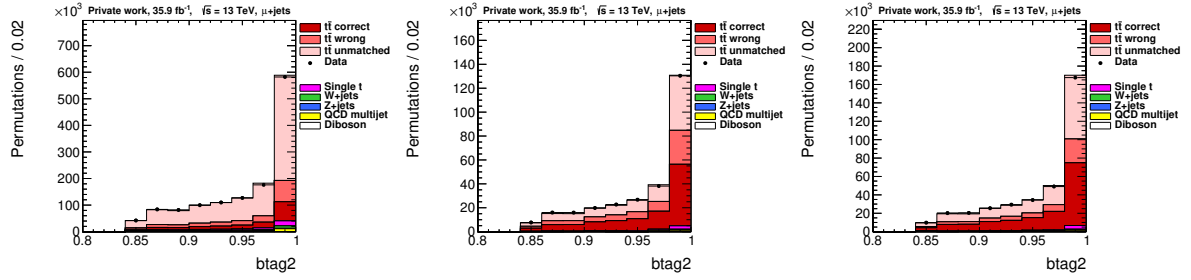

Figure 49: φ of the 1st reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
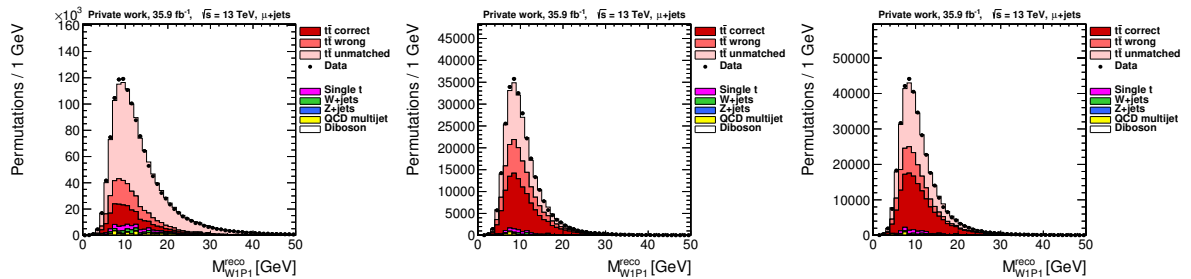
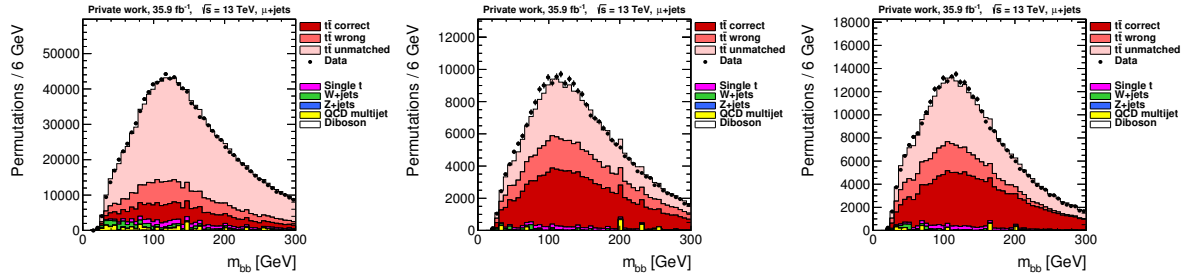**Figure 50:** Mass of the 1st reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
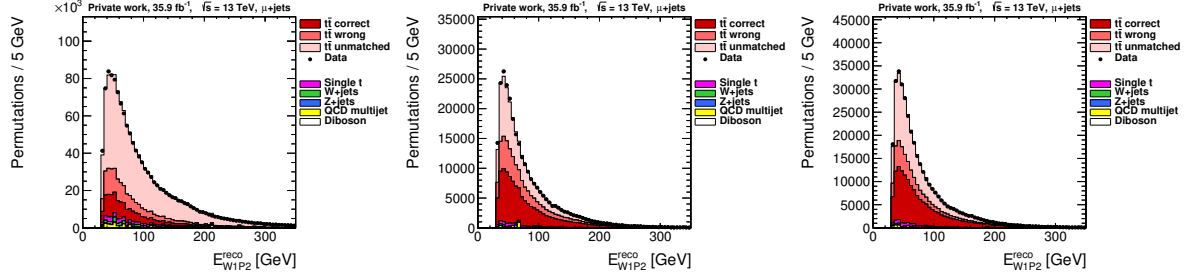
Figure 51: E of the 2nd reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
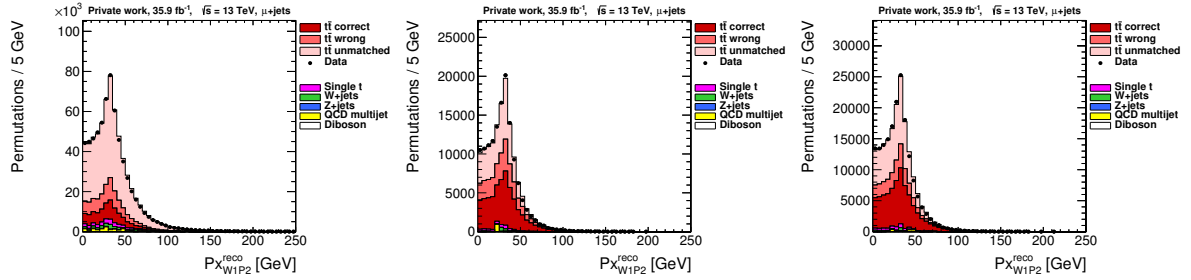
Figure 52: p_x of the 2nd reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
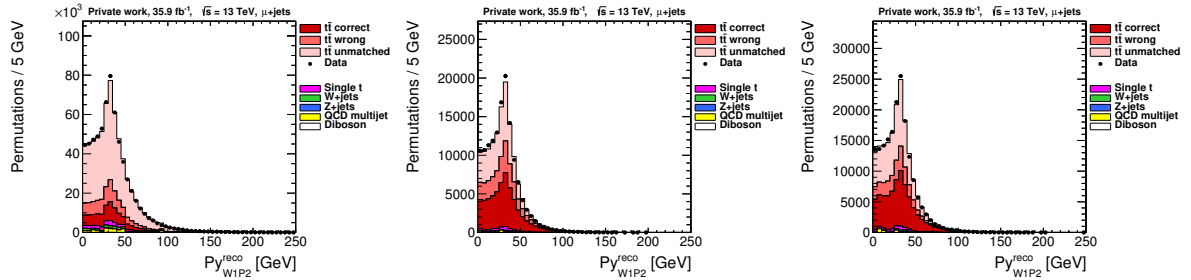
Figure 53: p_y of the 2nd reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
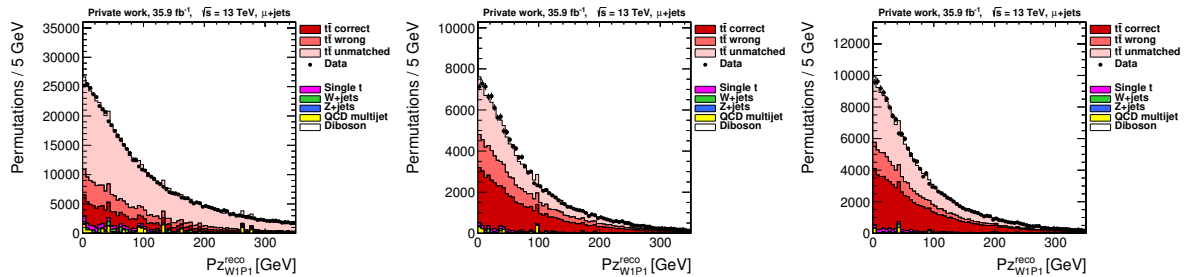
Figure 54: p_z of the 2nd reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
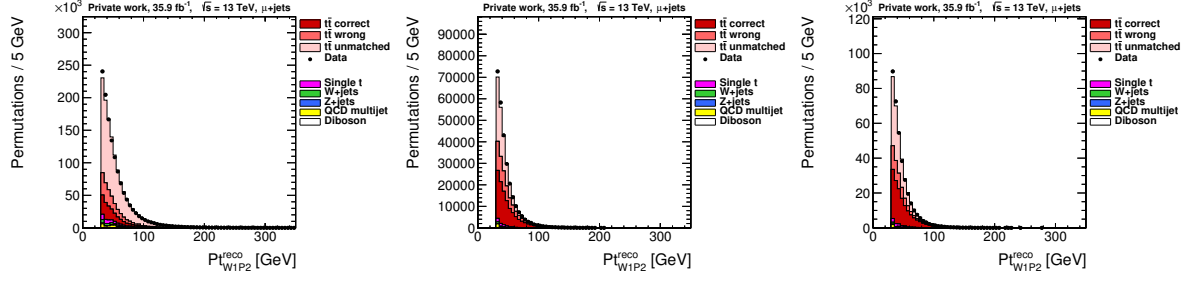
Figure 55: p_T of the 2nd reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
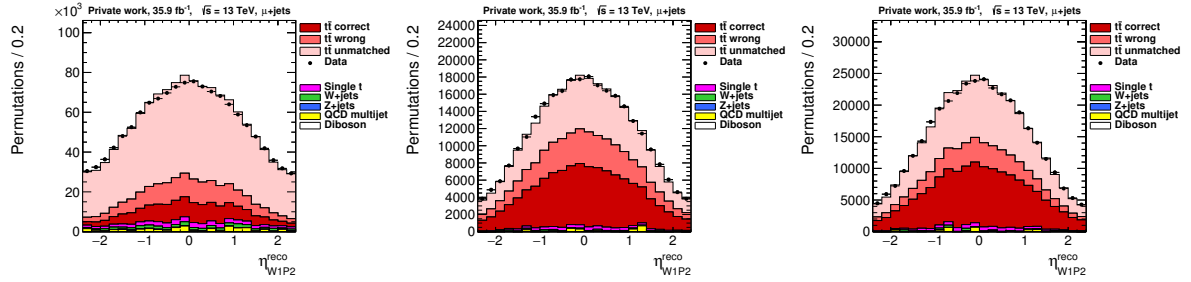
Figure 56: η of the 2nd reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
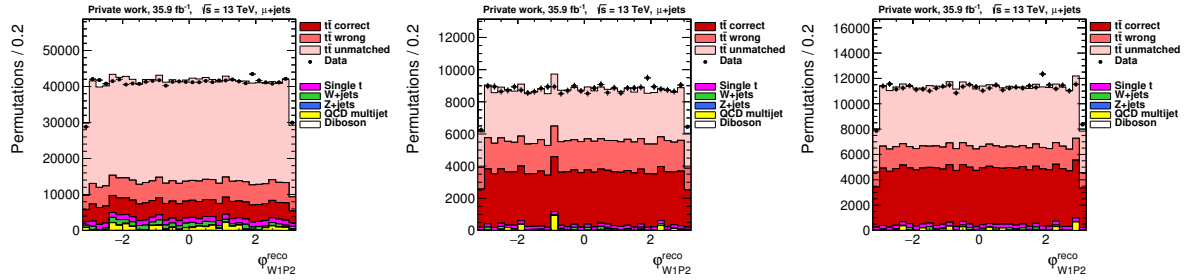
Figure 57: φ of the 2nd reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
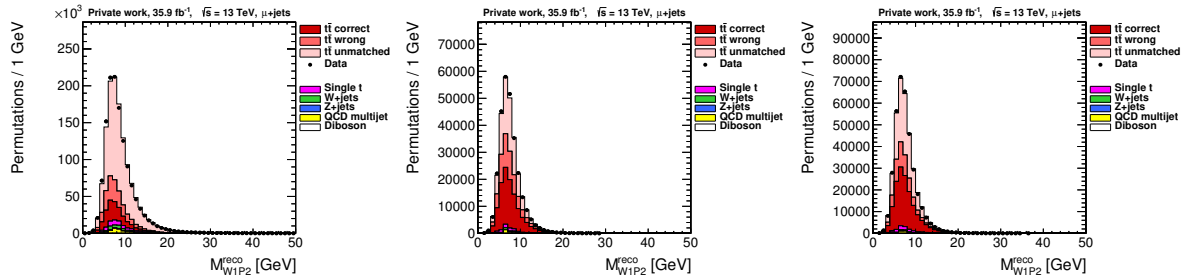
Figure 58: Mass of the 2nd reconstructed light jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
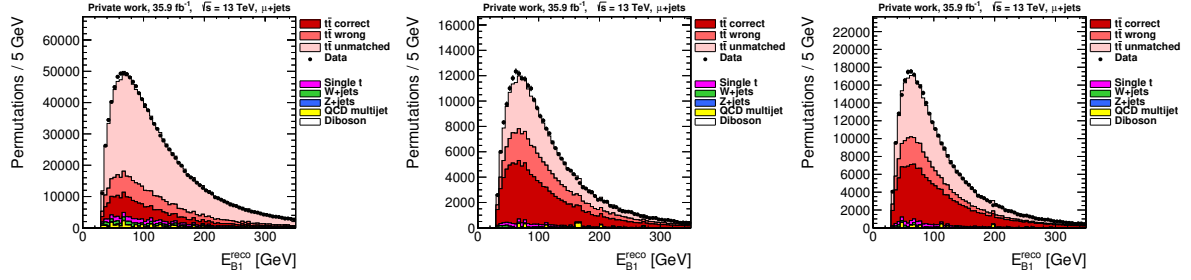
Figure 59: E of the 1st b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
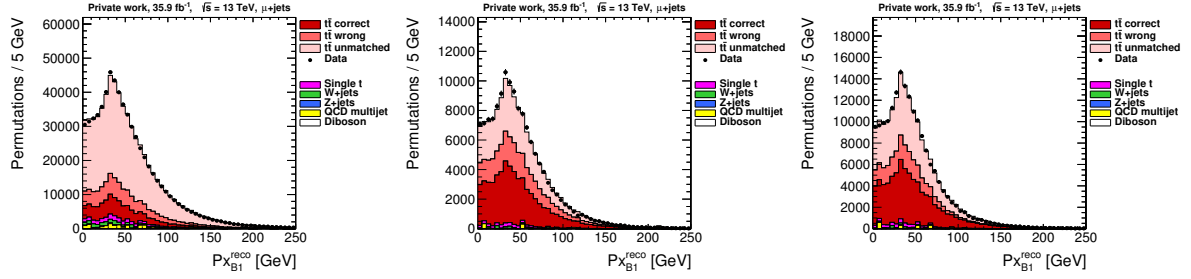
Figure 60: p_x of the 1st b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
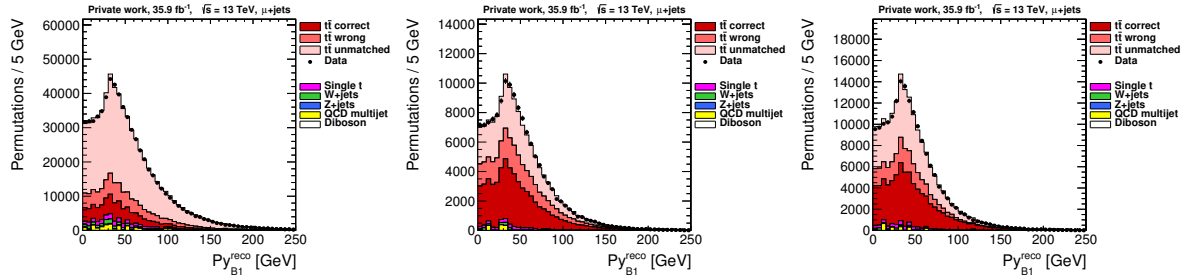
Figure 61: p_y of the 1st b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
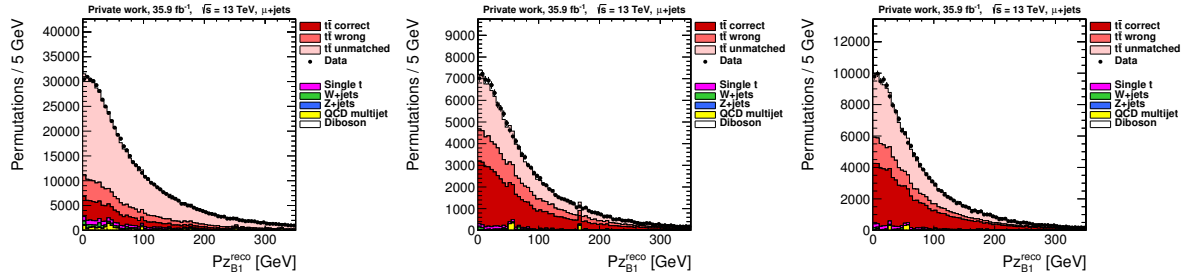
Figure 62: p_z of the 1st b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
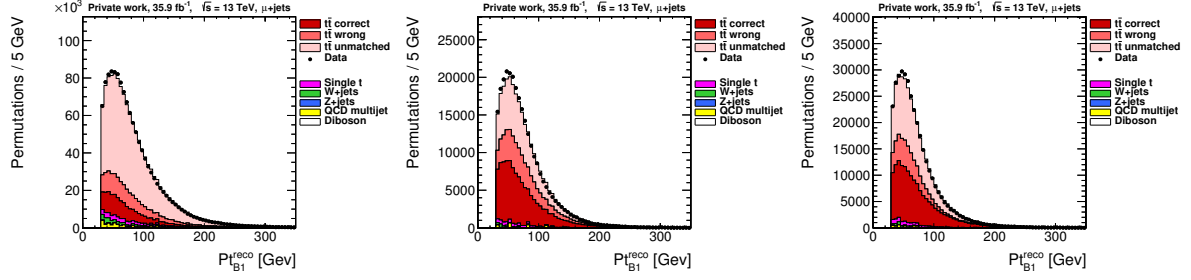
Figure 63: p_T of the 1st b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
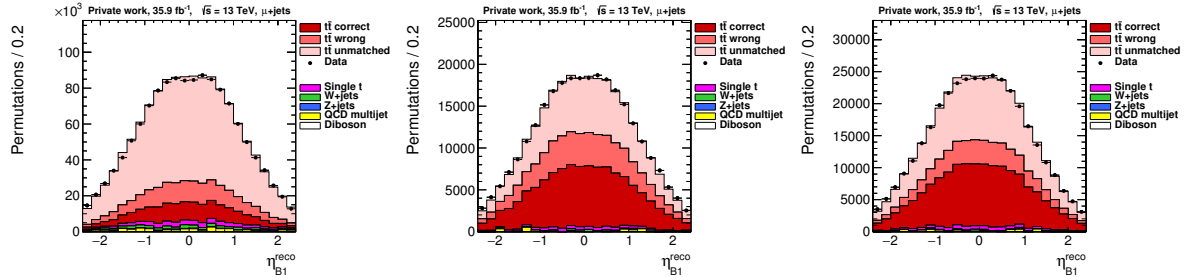
Figure 64: η of the 1st b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
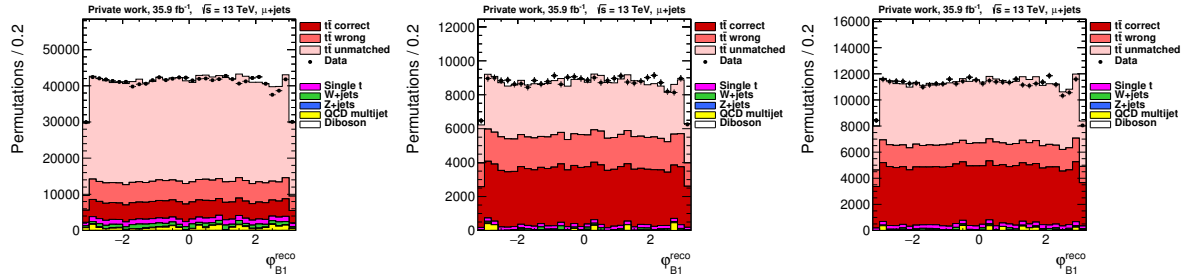
Figure 65: φ of the 1st b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
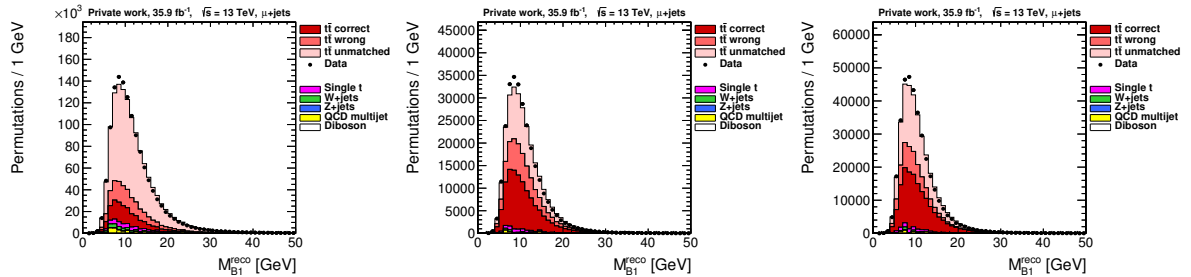
Figure 66: Mass of the 1st b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.
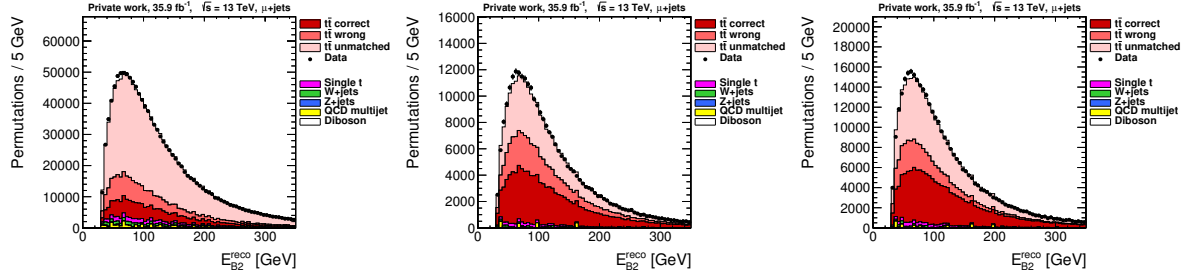
Figure 67: E of the 2nd b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 68: p_x of the 2nd b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 69: p_y of the 2nd b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 70: p_z of the 2nd b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 71: p_T of the 2nd b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 72: η of the 2nd b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 73: φ of the 2nd b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 74: Mass of the 2nd b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 75: The b-tag value of the 1st b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 76: The b-tag value of the 2nd b-jet, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

Figure 77: The invariant mass of the *bb* system, for the baselineselection(left), the $P_{gof} > 0.2$ selection (center) and the final NNoutput > 0.108 selection.

References

[1] Jet energy scale uncertainty correlations between ATLAS and CMS at 8 TeV. Technical Report CMS-PAS-JME-15-001. ATL-PHYS-PUB-2015-049, CERN, Geneva, 2015.

[2] Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV. Technical Report CMS-PAS-TOP-16-021, CERN, Geneva, 2016.

[3] CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[4] Measurement of the top quark mass with lepton+jets final states in pp collisions at $\sqrt{s} =$ 13 TeV. Technical Report CMS-PAS-TOP-17-007, CERN, Geneva, 2017.

[5] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[6] B. Abbott et al. Direct measurement of the top quark mass at D0. *Phys. Rev.*, D58:052001, 1998.

[7] J. Abdallah et al. A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole. *Eur. Phys. J.*, C71:1557, 2011.

[8] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.

[9] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 09:111, 2009. [Erratum: JHEP02,011(2010)].

[10] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce Dubois, M. Asai, G. Barrand, R. Capra, S. Chauvie, R. Chytracek, G. A. P. Cirrone, G. Cooperman, G. Cosmo, G. Cuttone, G. G. Daquino, M. Donszelmann, M. Dressel, G. Folger, F. Foppiano, J. Generowicz, V. Grichine, S. Guatelli, P. Gumplinger, A. Heikkinen, I. Hrivnacova, A. Howard, S. Incerti, V. Ivanchenko, T. Johnson, F. Jones, T. Koi, R. Kokoulin, M. Kossov, H. Kurashige, V. Lara, S. Larsson, F. Lei, O. Link, F. Longo, M. Maire, A. Mantero, B. Mascialino, I. McLaren, P. Mendez Lorenzo, K. Minamimoto, K. Murakami, P. Nieminen, L. Pandola,

S. Parlati, L. Peralta, J. Perl, A. Pfeiffer, M. G. Pia, A. Ribon, P. Rodrigues, G. Russo, S. Sadilov, G. Santin, T. Sasaki, D. Smith, N. Starkov, S. Tanaka, E. Tcherniaev, B. Tome, A. Trindade, P. Truscott, L. Urban, M. Verderi, A. Walkden, J. P. Wellisch, D. C. Williams, D. Wright, and H. Yoshida. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53(1):270–278, Feb 2006.

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

[12] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473–500, 2008.

[13] Spyros Argyropoulos and Torbjörn Sjöstrand. Effects of color reconnection on $t\bar{t}$ final states at the LHC. *JHEP*, 11:043, 2014.

[14] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015, 2013.

[15] M. Baak, J. Cúth, J. Haller, A. Hoecker, R. Kogler, K. Mönig, M. Schott, and J. Stelzer. The global electroweak fit at NNLO and prospects for the LHC and ILC. *Eur. Phys. J.*, C74:3046, 2014.

[16] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008.

[17] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J.*, C77(10):663, 2017.

[18] Roderik Bruce, Gianluigi Arduini, Hannes Bartosik, Riccardo De Maria, Massimo Giovannozzi, Giovanni Iadarola, John Jowett, Kevin Shing Bruce Li, Mike Lamont, Anton Lechner, Elias Metral, Daniele Mirarchi, Tatiana Pieloni, Stefano Redaelli, Giovanni Rumolo, Benoit Salvant, Rogelio Tomas Garcia, and Jorg Wenninger. LHC Run 2: Results and challenges. (CERN-ACC-2016-0103):MOAM5P50. 7 p, Jul 2016.

[19] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[20] Dario Buttazzo, Giuseppe Degrassi, Pier Paolo Giardino, Gian F. Giudice, Filippo Sala, Alberto Salvio, and Alessandro Strumia. Investigating the near-criticality of the Higgs boson. *JHEP*, 12:089, 2013.

[21] A. Carati. Gravitational effects of the faraway matter on the rotation curves of spiral galaxies. 2011.

[22] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[23] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[24] François Chollet et al. Keras. `https://github.com/keras-team/keras`, 2015.

[25] Jesper R. Christiansen and Peter Z. Skands. String Formation Beyond Leading Colour. *JHEP*, 08:003, 2015.

[26] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002, 2011.

[27] The CMS Collaboration. CMS draft analysis note CMS AN-16-327: Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in pp collisions at $\sqrt{s}$ = 13 TeV using 2016 data, November 2017. CMS internal.

[28] Michal Czakon, David Heymes, and Alexander Mitov. High-precision differential predictions for top-quark pairs at the LHC. *Phys. Rev. Lett.*, 116(8):082003, 2016.

[29] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, July 2011.

[30] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012.

[31] Stefano Frixione, Eric Laenen, Patrick Motylinski, and Bryan R. Webber. Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations. *JHEP*, 04:081, 2007.

[32] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[33] Christoph Garbers. Measurement of the Top Quark Mass in the Muon+Jets Final State at $\sqrt{s}$ = 13 TeV, October 2016. Master's Thesis.

[34] H. Georgi and S. L. Glashow. Unity of All Elementary Particle Forces. *Phys. Rev. Lett.*, 32:438–441, 1974.

[35] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[36] A. Heister et al. Study of the fragmentation of b quarks into B mesons at the Z peak. *Phys. Lett.*, B512:30–48, 2001.

[37] Andrej Karpathy. Cs231n convolutional neural networks for visual recognition. website: http://cs231n.github.io/, 2017. lecture note of the CS231 course of the stanford university.

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

[39] Vardan Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017.

[40] Vardan Khachatryan et al. Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV. *Phys. Rev.*, D95(9):092001, 2017.

[41] Vardan Khachatryan et al. The CMS trigger system. *JINST*, 12(01):P01020, 2017.

[42] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *ArXiv e-prints*, December 2014.

[43] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[44] Nataliia Kovalchuk. *Top quark mass measurement and color effects at the LHC*. PhD thesis, Universität Hamburg, 2018. (unpublished).

[45] T Lange. Dynamische Erweiterungen kinematischer Fits für die Higgsrekonstruktion am LHC, September 2015. Bachelor's Thesis.

[46] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, 1962.

[47] Fabienne Marcastel. CERN's Accelerator Complex. La chaîne des accélérateurs du CERN. October 2013. General Photo.

[48] B.R Martin and G. Shaw. *Particle Physics*. Wiley, 4th edition edition, 2017.

[49] Tom Melia, Paolo Nason, Raoul Rontsch, and Giulia Zanderighi. W+W-, WZ and ZZ production in the POWHEG BOX. *JHEP*, 11:078, 2011.

[50] Alireza Mika. evostra source code. {https://github.com/alirezamika/evostra}. version: commit cb1beb623b5f7f05dce5b2bfe0a5de3f61d3d753 (08.12.17).

[51] D. Mishkin and J. Matas. All you need is a good init. *arXiv preprint arXiv:1511.06422*, November 2015.

[52] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.

[53] C. Patrignani and Particle Data Group. Review of particle physics. *Chinese Physics C*, 40(10):100001, 2016.

[54] B. Pontecorvo. Inverse beta processes and nonconservation of lepton charge. *Sov. Phys. JETP*, 7:172–173, 1958. [Zh. Eksp. Teor. Fiz.34,247(1957)].

[55] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J.*, C71:1547, 2011.

[56] Brian Robson. The matter-antimatter asymmetry problem. 04:166–178, 01 2018.

[57] T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever. Evolution Strategies as a Scalable Alternative to Reinforcement Learning. *ArXiv e-prints*, March 2017.

[58] S. Schael et al. Precision electroweak measurements on the Z resonance. *Phys. Rept.*, 427:257–454, 2006.

[59] Markus Seidel. *Precise Measurement of the Top-Quark Mass at the CMS Experiment using the Ideogramm Method*. PhD thesis, University of Hamburg, August 2015.

[60] Jim Shifflett. Standard Model Lagrangian (including neutrino mass terms), 2015.

[61] Ravid Shwartz-Ziv and Naftali Tishby. Opening the black box of deep neural networks via information. *CoRR*, abs/1703.00810, 2017.

[62] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.

[63] Albert M Sirunyan et al. Measurement of normalized differential t-tbar cross sections in the dilepton channel from pp collisions at sqrt(s) = 13 TeV. 2017.

[64] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[65] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[66] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J.*, C74(8):3024, 2014.

[67] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[68] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, May 2016.

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

[70] Annika Vanhoefer. *Search for lepton-flavour violating decays of the Higgs boson at the LHC*. PhD thesis, U. Hamburg, Dept. Phys., Hamburg, 2017.

[71] M. D. Zeiler. ADADELTA: An Adaptive Learning Rate Method. *ArXiv e-prints*, December 2012.

List of Figures

List of Tables

Erklärung

Hiermit bestätige ich, dass die vorliegende Arbeit von mir selbständig verfasst wurde und ich keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe und die Arbeit von mir vorher nicht einem anderen Prüfungsverfahren eingereicht wurde. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich bin damit einverstanden, dass die Masterarbeit veröffentlicht wird.

....
Ort, Datum   Unterschrift