

Search for pair production of heavy particles decaying to a top quark and a gluon and development of ML-based L1 trigger strategies at the CMS experiment

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik,
Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von
Finn Jonathan Labe

Hamburg
2024



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Gutacher/innen der Dissertation:	Prof. Dr. Johannes Haller Prof. Dr. Christian Schwanenberger
Zusammensetzung der Prüfungskommission:	Prof. Dr. Günter Sigl Prof. Dr. Johannes Haller Prof. Dr. Christian Schwanenberger Dr. Roman Kogler Prof. Dr. Konstantinos Nikolopoulos
Vorsitzende/r der Prüfungskommission:	Prof. Dr. Günter Sigl
Datum der Disputation:	09.05.2025
Vorsitzender des Fach- Promotionsausschusses Physik:	Prof. Dr. Wolfgang J. Parak
Leiter des Fachbereichs Physik:	Prof. Dr. Markus Drescher
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

In dieser Arbeit präsentiere ich eine Suche nach einer potentiellen Signatur neuer Physik: der Existenz einer Resonanz t^* , die in ein Top-Quark und ein Gluon zerfällt. Ich verwende 138 fb^{-1} an Daten von Proton-Proton Kollisionen, die mit dem Compact Muon Solenoid (CMS) Detektor am Large Hadron Collider (LHC) gemessen wurden. Ich analysiere Ereignisse mit einem Lepton, Jets und fehlendem transversalen Impuls und reiche potentielle Signal-Ereignisse mithilfe eines Deep Neural Networks (DNN) an. Ich nutze eine aktive, zweistufige Dekorrelationsprozedur um sicherzustellen, dass das DNN unabhängig von der sensitiven Variable S_T der Suche ist. Diese ist als skalare Impulssumme aller betrachteten Objekte definiert. Ich analysiere S_T -Verteilungen auf der Suche nach Paarproduktion von t^* mit Spin- $\frac{1}{2}$ oder Spin- $\frac{3}{2}$. Meine Resultate stimmen mit den Vorhersagen des Standardmodells überein, so kann ich die Paarproduktion von t^* mit Spin- $\frac{1}{2}$, je nach Masse des t^* , für Wirkungsquerschnitte zwischen 120 fb für $m_{t^*} = 700 \text{ GeV}$ und 0.8 fb für $m_{t^*} = 3000 \text{ GeV}$ ausschließen. Für ein t^* mit Spin- $\frac{3}{2}$ liegen die Ausschlussgrenzen zwischen 15 fb für $m_{t^*} = 700 \text{ GeV}$ und 1.0 fb für $m_{t^*} = 2750 \text{ GeV}$. Unter der Annahme, dass alle t^* als $t^* \rightarrow tg$ zerfallen, kann die Existenz eines t^* unterhalb einer Masse von 1050 GeV für Spin- $\frac{1}{2}$ bzw. 1700 GeV für Spin- $\frac{3}{2}$ ausgeschlossen werden. Dies sind die bisher stringentesten Ausschlussgrenzen für dieses Teilchen.

In einem zweiten Teil präsentiere ich die Entwicklung von Algorithmen für den Level-1 Trigger (L1T) von CMS, die mithilfe von maschinellem Lernen versuchen eine Verbesserung gegenüber aktuellen Algorithmen zu erzielen. Aufgrund der hohen Kollisionsrate in CMS und den daraus resultierenden immensen Datenströmen muss der L1T mit einer Rate von 40 MHz entscheiden, ob ein Ereignis gespeichert wird. Insbesondere in Hinblick auf die zukünftige "High-Luminosity"-Phase des LHC sind solche neuartigen Techniken unabdingbar für CMS, um die Effektivität des Trigger-systems zu gewährleisten. Ich führe Studien zur Optimierung, Hardwareintegration und Nutzung eines Triggers, der auf einem Neuronalen Netz basiert durch. Ich zeige, dass Ereignisse des Prozesses $HH \rightarrow bbWW$, die ein einzelnes Muon enthalten, mithilfe eines kleinen neuronalen Netzwerkes effektiver selektiert werden können als mit klassischen Algorithmen. Ich demonstriere, dass dieses Netzwerk in die Hardware des L1T von CMS integriert werden kann.

In this thesis, I present a search for a potential signature of new physics: the existence of a new resonance t^* , characterized by its decay into a top quark and a gluon. I use 138 fb^{-1} of proton-proton collision events gathered with the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). I analyze events with a single lepton, multiple jets, and missing transverse momentum and enrich potential signal events using a deep neural network (DNN). I utilize an active, two-step decorrelation procedure to ensure that the DNN is independent of the sensitive variable S_T of the search. This variable is defined as a scalar energy sum of all considered objects. I analyze distributions of S_T in search for pair production of t^* with either spin- $\frac{1}{2}$ or spin- $\frac{3}{2}$. The results agree with Standard Model predictions. Therefore, I set exclusion limits on the production cross section of t^* . In the spin- $\frac{1}{2}$ case, these are between 120 fb for $m_{t^*} = 700 \text{ GeV}$ and 0.8 fb for $m_{t^*} = 3000 \text{ GeV}$. For t^* with spin- $\frac{3}{2}$, the exclusion limits are between 15 fb for $m_{t^*} = 700 \text{ GeV}$ and 1.0 fb for $m_{t^*} = 2750 \text{ GeV}$. Assuming all t^* decay as $t^* \rightarrow tg$, I exclude the existence of t^* below a mass of 1050 GeV for spin- $\frac{1}{2}$ and 1700 GeV for spin- $\frac{3}{2}$. These are the most stringent exclusion limits for this particle to date.

In a second part, I present the development of algorithms for the Level-1 trigger (L1T) of CMS, which utilize machine learning to improve on currently used algorithms. Due to the high collision rate in CMS and the resulting immense data streams, the L1T needs to decide whether to store or discard an event at a rate of 40 MHz. Especially towards the high-luminosity phase of the LHC, novel techniques will be crucial for CMS to ensure optimal performance of the trigger system. I present studies on the optimization, hardware integration and usage of a trigger based on a neural network. I show that events of a $HH \rightarrow bbWW$ (single muon) process can be selected more effectively than with classical algorithms using a lightweight neural network. I demonstrate that this network can be integrated into the hardware of the CMS L1T.

LIST OF OWN CONTRIBUTIONS

This thesis gives an overview on my high energy particle physics research. To put it into a cohesive context, I add other information, such as on the theoretical background or experimental setup. All publicly available information is cited from the appropriate sources, including publications of the specific work I present in this thesis. Additionally, in this section, I give a summary of my personal contributions to all presented information.

Search for pair production of t^*

I was the primary analyzer of the search that I document in chapter 5, which was published in [1]. I presented the results as part of a summary presentation at ICHEP2024 [2], and composed proceedings on that presentation [3]. A tabulated form of the analysis results is available as a HEPData entry in [4]. Preliminary studies on the general analysis strategy were presented in my master's thesis [5], based on which I worked on the following items:

- Managing the generation of simulated signal samples for the analysis and studying the difference between spin- $\frac{1}{2}$ and spin- $\frac{3}{2}$ t^* pair production in them.
- Optimization of the cut-based selection steps.
- Optimization of the DNN-based signal discrimination and developing its two-step decorrelation.
- Implementing a data-driven background estimation strategy for non- t background sources.
- Including many centrally derived corrections and scale factors in the analysis.
- Calculating electron trigger and b -tagging yield scale factors specific to the analysis.
- Accounting for all relevant systematic uncertainties and performing the statistical analysis, including statistical tests and the step by step unblinding of the signal region.
- Managing the collaborative work within CMS and the internal review:
 - Regular exchange with other groups working on t^* searches.
 - Regular presentations of the analysis status in the CMS B2G subgroup.
 - Writing analysis documentation for internal review.
 - Writing and editing the analysis publication draft.
 - Presenting the analysis for internal pre-approval and approval.
 - Managing the communication with the analysis review committee and publication committee during the internal review.
- Answering referee comments during peer review after submitting the publication to the European Physical Journal C.

The analysis was primarily supervised by Dr. Roman Kogler. I supervised a bachelor's thesis performing studies on t^* decaying to a t quark and a photon [6].

Development of ML-based L1T algorithms

As part of a team of researchers, I worked on the development of ML-based L1T seeds, the results of which are presented in chapter 6. I participated in this project since its beginning and worked on the following items:

- Initial conceptualization and feasibility studies of supervised NNs in the L1T.
- Planning and implementation of a computing framework for ML-based L1 trigger development and evaluation, used by multiple other people within and outside of the (Universität Hamburg) UHH group.
- Optimization of a L1 NN targeting the $HH \rightarrow bbWW$ (single muon) process.
- Integration of a L1 NN in the L1T hardware, tested during 2023 data taking.
- Integration of a L1 NN in two different offline analysis frameworks, including preparation of samples containing the relevant L1 objects:
 - UHH2 [7], a framework for analyzing Run 2 data implemented in C++.
 - columnflow [8], a python-based framework for analysis of Run 3 data.
- Studies on the HLT usage of L1 NNs, the effect on offline analysis, and efficiency measurement approaches.

The team was led by Dr. Artur Lobanov. In the context of this work I supervised two master's theses, which are accessible in [9, 10], and one bachelor's thesis [11].

Other work

Aside from these two main projects, my tasks included other contributions to the CMS Collaboration, which I list below:

- For about two years, I served as *trigger contact* for the B2G subgroup. In this position, I managed the communication between the B2G and trigger subgroups, and provided expertise about trigger usage and development to physics analyses in B2G. Specifics of the work included:
 - Regular presentations in B2G and trigger subgroup meetings.
 - Reviewing B2G analyses on proper trigger usage during internal review.
 - Implementing and updating data quality monitoring and release validation code for B2G triggers, as documented in section 3.2.3.4.
 - Monitoring the performance of B2G triggers during Run 3 data taking, as documented in section 3.2.3.4 for new soft drop mass triggers. These results are published in [12].
- I developed a web-based software release validation tool for HLT paths under the supervision of Dr. Sam Harper, which is briefly introduced in section 3.2.3.4.
- As part of the shift crew, I contributed to CMS operation as *trigger shifter*, monitoring the trigger performance during data taking and updating prescales when necessary.
- I presented various aspects of my work at several conferences, including the yearly meeting of the German Physical Society, yearly meeting of german CMS groups, as well as different CMS-internal and national workshops and meetings.

1	Introduction	1
2	Theory	2
2.1	The Standard Model of particle physics	2
2.1.1	Fundamental forces	2
2.1.2	Elementary particles	2
2.1.3	Mathematical description	4
2.1.4	Cross section calculations	10
2.2	Physics beyond the Standard Model	12
2.2.1	Observational shortcomings of the SM	12
2.2.2	Structural shortcomings of the SM	14
2.2.3	Extensions of the SM	15
3	Experiment	18
3.1	Large Hadron Collider	18
3.1.1	The LHC pre-accelerator chain	18
3.1.2	Particle acceleration at the LHC	18
3.1.3	Proton collider physics	21
3.2	Compact Muon Solenoid experiment	22
3.2.1	Detector overview	22
3.2.2	Data collected by CMS	28
3.2.3	Data acquisition	32
3.2.4	Event reconstruction	43
4	Computing methods	50
4.1	Monte Carlo simulations	50
4.2	Machine Learning	51
4.2.1	Introduction to neural networks	51
4.2.2	Activation functions	52
4.2.3	Loss functions	53
4.2.4	Optimizers and backpropagation	53
4.2.5	Types of neural networks	55
4.2.6	Performance evaluation	56
4.2.7	Training stability	56
5	Search for pair production of a heavy new particle t^*	58
5.1	Theoretical background of t^*	58
5.1.1	Mathematical description: spin- $\frac{1}{2}$ scenario	59
5.1.2	Mathematical description: spin- $\frac{3}{2}$ scenario	59
5.1.3	Production of t^* at proton-proton colliders	60
5.1.4	Decay channels of t^*	61
5.2	Previous searches	61

5.3	Analysis	62
5.3.1	Analysis overview	63
5.3.2	Monte Carlo simulated samples	64
5.3.3	Signal process studies	64
5.3.4	Object definitions	67
5.3.5	Cut-based event selection	73
5.3.6	Corrections and quality selections	78
5.3.7	Selection results	80
5.3.8	Deep neural network	86
5.3.9	Estimation of SM backgrounds	97
5.3.10	Systematic uncertainties	102
5.4	Results	113
5.4.1	Statistical analysis	113
5.4.2	Discussion and future prospects	121
6	Development of ML-based L1 trigger strategies	125
6.1	Limitations of cut-based triggers	125
6.2	Trigger optimization	127
6.2.1	Neural network input variables	128
6.2.2	Neural network architecture	132
6.3	Hardware integration	132
6.3.1	Compressing the neural network	135
6.3.2	Testing under running conditions	137
6.4	Trigger usage and analysis integration	139
6.4.1	HLT strategies	140
6.4.2	Efficiency measurement	143
6.5	Outlook	145
7	Summary	148
	Appendix	149
A.1	Description of the experiment and methods	149
A.2	Search for pair production of t^*	152
A.3	ML-based L1 trigger algorithms	157
	Bibliography	158

Our knowledge about elementary particles, the fundamental building blocks of the universe, is constantly increasing. The first truly elemental particle was found with the discovery of the electron through cathode rays in 1897 [13], which also was the first sub-atomic particle discovered. Research continued with the discovery of the atomic nucleus in the Rutherford scattering experiments in 1911 [14], and the fact that it is composed from protons (in 1919 [15]) and neutrons (in 1932 [16]). Following discoveries of the muon in 1936 [17] and many other subatomic particles, the proton and neutron proved to not be truly fundamental: a substructure was proposed in form of the quark model in 1964 [18–20] and discovered in deep inelastic scattering experiments in 1969 [21, 22]. This discovery of the up-, down- and strange-quarks was followed by the charm quark in form of its bound state J/ψ [23, 24] in 1974. Many more findings were made to form the currently accepted model of elementary particles: the Standard Model (SM) of particle physics. A third lepton generation was discovered with the tau lepton around 1975 [25], and a third quark generation in form of the bottom quark in 1977 [26] and the top quark in 1995 [27, 28]. Interaction bosons were found as well: The gluon was discovered in three-jet events in 1978 [29] by the PLUTO experiment in Hamburg. Furthermore, the W and Z bosons were discovered in 1983 [30–33]. Most notably, the last missing piece of the theory was obtained with the discovery of the Higgs boson in 2012 [34–36] at CERN.

The SM, while being exceptionally successful in its prediction of particle and interaction properties, is not the ultimate model of reality, as it is known to have deficiencies. I outline the SM and its particle content, and motivate a range of theories proposing physics beyond the SM, in chapter 2. Many of these theories can be studied by trying to discover new elementary particles they predict. I investigate one of these potential new particles, labeled t^* and characterized by its decay $t^* \rightarrow tg$. To probe for its existence, I use proton-proton collision data gathered by the Compact Muon Solenoid (CMS) [37] experiment at the CERN Large Hadron Collider (LHC). After an introduction to the experimental setup in chapter 3 and a description of computing methods in chapter 4, I analyze data gathered between 2016 and 2018 in chapter 5.

Studies on elementary particles and their interactions will also continue in the future, with more and more data being produced at the LHC. This requires the continuous development of various aspects of the CMS detector. In chapter 6, I perform studies towards the usage of machine-learning based algorithms in the very first layer of the CMS trigger system. Implemented in hardware on field programmable gate arrays, this system accepts or rejects proton-proton collision data in real time at a rate of about 40 MHz. It is used to handle the immense amount of data produced by CMS and ensures that optimal data is available for analysis.

In order to research elementary particles, it is crucial to formulate a mathematical description able to make predictions on their properties and behavior, which can then be compared to experimental observations. This underlying mathematical formalism of particle physics is introduced in this chapter. Starting with the Standard Model of particle physics in section 2.1, I present a description of the currently known elementary particles and their interactions. Then, I outline the shortcomings of the theory, and potential solutions in the form of physics models beyond the Standard Model, in section 2.2.

2.1 The Standard Model of particle physics

The *Standard Model of particle physics* (SM) describes our current knowledge of the properties and interactions of all known elementary particles. It follows the formalism of a *quantum field theory* (QFT), describing quantum wave fields where particles are represented by quantum excitations of those fields. Below, I introduce the fundamental forces, the elementary particles themselves, and their mathematical description. All information in this section is based on [38] and [39], unless otherwise noted.

2.1.1 Fundamental forces

Four fundamental forces are known, of which three are described by the SM: the *electromagnetic*, *weak* and *strong interactions*. The fourth force is gravity, which is not included in the SM. For each force, there is a corresponding charge: the *electric charge* of the electromagnetic interaction, the *weak charge* and *weak isospin* of the weak interaction and the *color charge* of the strong interaction. Forces only act on particles that have a non-zero value of the specific charge corresponding to that force. In the SM, these interactions are described as an exchange of a mediator particle.

2.1.2 Elementary particles

Figure 2.1 shows all particles known to the SM, uniquely characterized and grouped by their mass, charges and spin. Notably, for each particle, a corresponding antiparticle exists. An antiparticle has the same properties as its regular counterpart, except for all charges being inverted. In some cases, like for the photon, this results in an antiparticle that is identical to the particle: the photon is its own antiparticle.

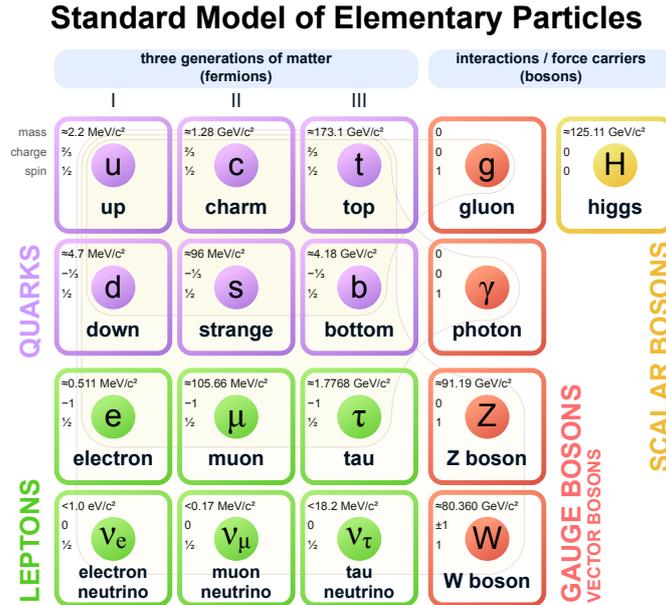


Figure 2.1: Overview of the particle content described by the SM. Split into fermions and bosons, the mass, electric charge and spin of all particles is given (following the 2019 update of [40]). Fermions are split into quarks and leptons, and bosons into gauge and scalar bosons. Brown rectangles indicate which particles can interact with each other (excluding the Higgs boson, which interacts with all massive particles). Taken from [41].

2.1.2.1 Fermions

With half-integer spin, fermions make up the majority of the SM's particle content. All of them have non-zero *weak isospin* and *weak charge*, allowing them to interact via the weak interaction. They are grouped into three so-called *generations*, and are further divided into *quarks* and *leptons*: quarks, massive particles carrying color charge, can interact via the strong interaction, whereas leptons can not. In order of increasing masses, the quarks are the up quark (u), the down quark (d), the strange quark (s), the charm quark (c), the bottom quark (b) and the top quark (t). With a mass of about 173 GeV , the t quark is the heaviest fundamental particle known to the SM. All quarks have non-zero electric charge and can thus interact electromagnetically. The other six fermions are the electron (e), muon (μ) and tau (τ), and a corresponding neutrino (ν_e , ν_μ and ν_τ) for each of them. Electron, muon and tau are massive, whereas the three neutrinos are massless in the SM. Similarly, the neutrinos do not carry electric charge, whereas the electron, muon and tau do.

2.1.2.2 Bosons

Bosons have integer spins, and can be subdivided into two groups differentiating spin-1 and spin-0 particles. With spin-1, the *gauge bosons* are the mediator particles of the elementary forces. The *photon* (γ) mediates the electromagnetic interaction, is massless and carries no charges. In contrast, the W and Z bosons, mediators of the weak interaction, are massive. The W bosons carry electric charge themselves, and are thus able to interact with other electrically charged particles via the

electromagnetic interaction. Finally, the massless *gluon* (g) is the mediator particle of the strong interaction. It carries color charge itself, which means that a gluon can exchange a gluon with another particle to interact strongly: gluon self-interaction is possible. The last boson in the SM is the *Higgs boson* (H), which has spin-0. As part of the Higgs mechanism (see section 2.1.3), it is a required part of the SM formalism allowing for boson masses. However, it is not a mediator particle of any interaction.

2.1.3 Mathematical description

Now that all elementary particles have been described, the mathematical formalism of the SM can be introduced. First, I will present a simplified formalism for a single interaction as an example, before extending it to the full SM. The SM is a QFT, based on a *Lagrangian density* (which will be called *Lagrangian* \mathcal{L}) in the following. It describes the behavior of a quantum field for all points in spacetime \mathbf{x} . A simple Lagrangian for a massless fermion can be written as

$$\mathcal{L} = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi \quad (2.1)$$

following from the Dirac equation. Here, Ψ is the *wave function* of a relativistic quantum field, $\bar{\Psi}$ its complex conjugate, γ^μ the Dirac matrices and ∂_μ a four-dimensional spacetime derivative. A summation over the index μ is implied, and the position of μ distinguishes covariant (subscript) and contravariant (superscript) vectors.

2.1.3.1 Gauge transformations

To construct a description of the behavior of these wave functions, I require them to be invariant under *gauge transformations*. A similar concept is found in classical mechanics in the form of Noether's theorem [42], where a theory is constructed by requiring invariance against translations in space or time, resulting in momentum and energy conservation, respectively. Assuming that a field Ψ has some intrinsic, non-observable *phase*, any observable quantity (like a charge) should be invariant under global transformations with an arbitrary phase θ , defined as

$$\Psi \rightarrow e^{i\theta}\Psi. \quad (2.2)$$

This gauge transformation for a single phase is called $U(1)$ gauge symmetry. Invariance trivially follows from equation 2.1 as the two phase transformations have opposite sign for Ψ and $\bar{\Psi}$ and thus cancel out. However, I generalize the requirement to *local gauge invariance*, defined as

$$\Psi \rightarrow e^{i\theta(\mathbf{x})}\Psi. \quad (2.3)$$

where θ is now spacetime dependent. If Ψ is independent of the choice of gauge, it should also be independent on locally different choices. Also, requiring a global gauge invariance could be understood as a gauge transformation acting at all positions in spacetime simultaneously, including outside of the light cone: this would violate the causality principle. The Lagrangian as introduced in equation 2.28 is not invariant under local gauge transformation. However, invariance can be ensured through introduction of a massless field A_μ . This field is defined to transform as $A_\mu \rightarrow A_\mu - \partial_\mu\theta(\mathbf{x})$, and the derivative ∂_μ in the Lagrangian is replaced by the *covariant derivative* D_μ :

$$D_\mu = \partial_\mu + ig_e q A_\mu. \quad (2.4)$$

with a free parameter g_e . This allows to write an updated Lagrangian as

$$\mathcal{L} = i\bar{\Psi}\gamma^\mu D_\mu\Psi = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - g_e\bar{\Psi}\gamma^\mu q A_\mu\Psi \quad (2.5)$$

It can be shown that this new Lagrangian is invariant under local gauge transformation, and with Noether's theorem that the described charge q is conserved. However, when adding the new field A_μ , another invariant term can be added as well, completing the Lagrangian:

$$\mathcal{L}_{QED} = i\bar{\Psi}\gamma^\mu D_\mu\Psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} = \underbrace{i\bar{\Psi}\gamma^\mu\partial_\mu\Psi}_{\text{fermion kinetic}} - \underbrace{g_e\bar{\Psi}\gamma^\mu q A_\mu\Psi}_{\text{interaction}} - \underbrace{\frac{1}{4}F^{\mu\nu}F_{\mu\nu}}_{\text{boson kinetic}} \quad (2.6)$$

The three resulting terms can be interpreted as follows: the first is the initial kinetic term, describing the behavior of the fermion. Through addition of the A_μ field, an interaction term arises, describing the interaction between the fermion and the new field with the interaction strength g_e , and the generator q (which is just a real number for $U(1)$). Finally, a kinetic term for the new field is present as well, which is written compactly as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Finding solutions for A_μ , it can be shown that this field must have three degrees of freedom, corresponding to three polarizations of a spin-1 boson.

This example shows that it is possible to construct a locally gauge-invariant formalism describing a fermion, if one adds a boson that interacts with it. Understanding the Ψ as, for example, an electron and the A_μ as a photon, this is *quantum electrodynamics* (QED), the theory of the electromagnetic interaction. Here, the interaction strength g_e would be the elementary electric charge e and q the electric charge of a particle in units of e . This can then be expressed as a fine structure constant α_{em} , given as

$$\alpha_{em} = \frac{g_e^2 Q}{4\pi}, \quad (2.7)$$

which is measured at about $\frac{1}{137}$. However, through *renormalization*, the coupling is not actually constant, but changing with the energy scale Q it is probed at: While $\frac{1}{137}$ is true at $Q = 0$, at the scale of the Z -boson $\alpha_{em} \approx \frac{1}{127}$ is found. Thus, it increases with Q : $\frac{\partial\alpha_{em}}{\partial Q} > 0$.

This introduction only includes a single particle and interaction, but can be trivially extended to include all fermions. Notably, in reality QED does not exist on its own, but was found to be unified with the weak interaction (as I explain in section 2.1.3.3). Thus, this introduction merely is an overview of the mathematical principle of gauge theories.

2.1.3.2 Quantum chromodynamics

Quantum chromodynamics (QCD) is the theory of the strong interaction. A main difference to QED are the different degrees of freedom found in the color charges: while there is only one electric charge (which can be positive or negative), there are three color charges, often labeled as red, green and blue, and the corresponding negative charges anti-red, anti-green and anti-blue. A local gauge invariance for this

configuration of charges can be described by a $SU(3)_C$ symmetry group, where the C labels this as resulting from color charge. With a similar derivation as for QED, the QCD Lagrangian can be written as

$$\mathcal{L}_{\text{QCD}} = \sum_q \underbrace{(i\bar{\Psi}_q \gamma^\mu \partial_\mu \Psi_q)}_{\text{quark kinetic}} - \underbrace{g_s \bar{\Psi}_q \gamma^\mu T^a G_\mu^a \Psi_q}_{\text{interaction}} - \underbrace{\frac{1}{4} G_{\mu\nu}^a G^{\mu\nu,a}}_{\text{gluon}}. \quad (2.8)$$

Here, the index q runs over all six quark flavors, resulting in an individual kinetic and interaction term for each of them. By the introduction of a new field G_μ^a and a covariant derivative as

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_s T^a G_\mu^a \quad (2.9)$$

local gauge invariance is ensured. In this definition of D_μ , a coupling strength g_s appears, as well as the generators T^a of the $SU(3)_C$ group, which can be realized by the Gell-Mann matrices. This group has eight generators, symbolized by the index a , and results in eight different gluon fields G_μ^a . An interaction term between gluons and quarks arises, as well as a term describing the gluon. A crucial difference is found in comparison to QED, where this term only contained a kinetic part:

$$G_{\mu\nu}^a = \underbrace{\partial_\mu G_\nu^a - \partial_\nu G_\mu^a}_{\text{gluon kinetic}} - \underbrace{g_s f^{abc} G_\mu^b G_\nu^c}_{\text{self-interaction}}. \quad (2.10)$$

An additional term describing the gluon self-interaction, with the structure constant f^{abc} , is present here. Another difference to QED is found in the coupling strength g_s . Similar as in QED, one can write this as

$$\alpha_s = \frac{g_s}{4\pi}. \quad (2.11)$$

However, in contrast to QED, α_s is decreasing with energy: $\frac{\partial \alpha_s}{\partial Q} < 0$. This feature of QCD results in so-called *asymptotic freedom*: at very high probing energies (equivalent to very small distances), color-charged particles can be treated as free objects, allowing a description using perturbation theory. In contrast, for smaller energies or higher distances, free color-charged particles can not be observed. Instead, they hadronize forming color-neutral bound states like *mesons* (one quark, one anti-quark) or *baryons* (three quarks or anti-quarks). An example of a meson is the *pion* $\pi^+ = \{u\bar{d}\}$, and well-known baryons are the *proton* $p^+ = \{uud\}$ and *neutron* $n^0 = \{udd\}$. This so-called *color confinement* is also the reason for hadronization processes: assuming a color-charged particle with high momentum is produced in some interaction, it will create other colored objects around it, which combine to form color-neutral hadrons. This spray of particles, which are all moving in approximately the same direction of the initial particle, is called a *jet*.

2.1.3.3 Weak interaction and electroweak unification

The third and final fundamental force in the SM, the weak interaction, is described by another local symmetry group: $SU(2)_L$. However, it is not considered alone, as it can be unified with the the electromagnetic interaction into a single theory, combining $U(1)_Y$ and $SU(2)_L$. This combined interaction, called *electroweak* (EW), is described by the following Lagrangian:

$$\begin{aligned}
 \mathcal{L}_{\text{EW}} = & \sum_f \underbrace{(i\bar{\Psi}_f \gamma^\mu \partial_\mu \Psi_f)}_{\text{fermion kinetic}} - \underbrace{\frac{g'}{2} \bar{\Psi}_f \gamma^\mu Y B_\mu \Psi_f}_{U(1)_Y \text{ interaction}} - \underbrace{\frac{g}{2} \bar{\Psi}_f \gamma^\mu T_i W_\mu^i \Psi_f}_{SU(2)_L \text{ interaction}} \\
 & - \underbrace{\frac{1}{4} B_{\mu\nu} B^{\mu\nu}}_{U(1)_Y \text{ field kinetic}} - \underbrace{\frac{1}{2} \epsilon_{ijk} W_{\mu\nu}^i W^{\mu\nu,j}}_{SU(2)_L \text{ field kinetic}}.
 \end{aligned} \tag{2.12}$$

As previously introduced, one obtains a kinetic term for all fermions (denoted by the index f). By introducing two new fields in the covariant derivative as

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + i\frac{g'}{2} Y B_\mu + i\frac{g}{2} T_i W_\mu^i, \tag{2.13}$$

new interaction terms and kinetic terms of these fields arise. Here, Y is the *weak hypercharge*, the generator of the $U(1)_Y$ group, and T_i the weak isospin (with $i \in [1, 2, 3]$), generator of the $SU(2)_L$ group and realized by the Pauli matrices. Notably, there is a relation between the electric charge q and these newly introduced charges:

$$q = T_3 + \frac{1}{2} Y. \tag{2.14}$$

The fields introduced in this formalism do not correspond directly to the gauge bosons introduced in section 2.1.2. In the context of spontaneous symmetry breaking and the Higgs mechanism (which is introduced below), they can be combined to obtain the physical bosons. With an angle θ_W , the *weak mixing angle* or *Weinberg angle*, the neutral γ (represented by the field A_μ) and Z bosons are obtained, as well as the charged W^\pm bosons:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2), \quad \begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}. \tag{2.15}$$

The weak mixing angle also relates the interaction strengths by

$$\tan \theta_W = \frac{g}{g'}. \tag{2.16}$$

In order to complete the description of the electroweak interaction, the concept of *chirality* needs to be defined. In the context of gauge theories, this is done using projection operators $P_R = \frac{1}{2}(1 + \gamma^5)$ and $P_L = \frac{1}{2}(1 - \gamma^5)$ for so-called *right-handed* and *left-handed* states, respectively. Only left-handed fermions (and right-handed anti-fermions) participate in the weak interaction, denoted by the subscript L in the symmetry group $SU(2)_L$. They can be ordered as doublets where the upper component has $T_3 = \frac{1}{2}$ and the lower component has $T_3 = -\frac{1}{2}$:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \quad \begin{pmatrix} u \\ d' \end{pmatrix}_L, \quad \begin{pmatrix} c \\ s' \end{pmatrix}_L, \quad \begin{pmatrix} t \\ b' \end{pmatrix}_L. \tag{2.17}$$

In contrast, all right-handed fermions are singlets with $T_3 = 0$.

Due to the inequality of mass and flavor eigenstates of quarks, transitions between quarks of different doublets are possible. Mass and flavor eigenstates are related via the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (2.18)$$

Here, (d, s, b) are the quarks' mass eigenstates and (d', s', b') their interaction eigenstates. The square of the magnitudes V_{ij} describes the transition probability between two quarks with flavors i and j . Some elements of the CKM matrix can be measured individually, and most precise results are then obtained from a fit to all measurements [43], yielding

$$|V_{\text{CKM}}| = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00369 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix}. \quad (2.19)$$

2.1.3.4 Higgs mechanism

In all previous sections, mass terms for fermions and bosons were omitted, despite some fermions and some gauge bosons being measured to be massive. However, simply adding these terms to the respective Lagrangians violates the initially required gauge invariance. Instead, masses are described with the *Brout-Englert-Higgs* mechanism, often just called *Higgs mechanism*. A new complex scalar field is added to the SM, the *Higgs field* ϕ . It is defined as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (2.20)$$

and is a T_3 doublet with $Y = 1$. It is described by the Lagrangian

$$\mathcal{L}_\phi = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \quad (2.21)$$

where D_μ is the covariant derivative introduced in the electroweak unification (eq. 2.13). It contains a potential $V(\phi)$, which is given by

$$V(\phi) = \mu^2 |\phi|^2 + \lambda |\phi|^4 \quad (2.22)$$

with two free, real parameters μ^2 and λ . To ensure vacuum stability, meaning a finite minimum of $V(\phi)$, $\lambda > 0$ is required. However, different configurations of μ^2 are possible. With $\mu^2 > 0$ a potential with a ground state at $|\phi| = 0$ would be obtained. Choosing $\mu^2 < 0$ results in a potential with a different minimum, as illustrated in figure 2.2.

At the origin, $U(1)_Y$ symmetry is still fulfilled. However, this symmetry is *spontaneously broken* when a point in at minimum is chosen as ground state. This point v is called *vacuum expectation value* and is defined as

$$v = \sqrt{\frac{-\mu^2}{\lambda}}. \quad (2.23)$$

As a continuum of degenerate ground states exists, one of them can be chosen arbitrarily and the Higgs field can be rewritten as

$$\phi_{\text{vacuum}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \quad (2.24)$$

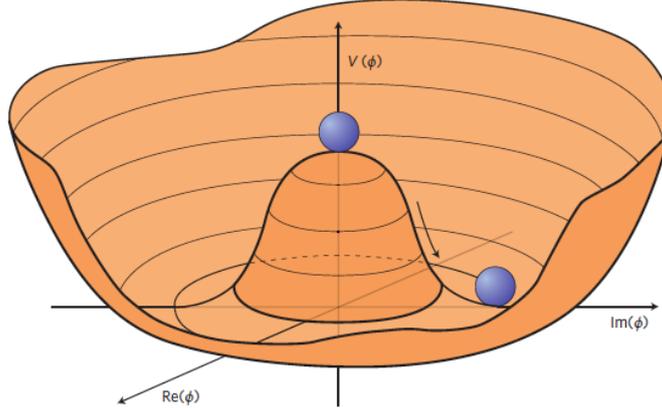


Figure 2.2: Illustration of the shape of a Higgs potential with $\mu^2 < 0$, where two blue spheres show ground states before and after spontaneous symmetry breaking. Taken from [44].

where H can be interpreted as the field of a physical particle, the *Higgs boson*. The potential is then found to be

$$V(\phi_{\text{vacuum}}) = \underbrace{-\mu^2 H^2}_{\text{Higgs mass term}} + \underbrace{\lambda v H^3 + \frac{1}{4} \lambda H^4}_{\text{Higgs self-interaction}} \quad (2.25)$$

which adds three terms to the Lagrangian: the mass term of the Higgs boson which has $m_H = -\mu^2$, being a free parameter of the theory, and two terms describing self-interactions of three and four Higgs bosons. Including the Higgs mechanism in the electroweak Lagrangian results in the following additional terms:

$$\begin{aligned} \mathcal{L}_H = & \underbrace{\frac{1}{2}(\partial_\mu H)(\partial^\mu H)}_{\text{Higgs kinetic}} + \underbrace{\frac{g^2 v^2}{4}((W_\mu^+)^2 + (W_\mu^-)^2)}_{W^\pm \text{ mass term}} + \underbrace{\frac{g^2 v^2}{8 \cos^2 \theta_W} Z_\mu^2}_{Z \text{ mass term}} \\ & + \underbrace{\frac{g^2}{8}((W_\mu^+)^2 + (W_\mu^-)^2)(2vH + H^2)}_{W^\pm H \text{ interaction term}} \\ & + \underbrace{\frac{g^2}{16 \cos^2 \theta_W} Z_\mu^2(2vH + H^2)}_{Z H \text{ interaction term}}. \end{aligned} \quad (2.26)$$

One obtains mass terms for the W^\pm and Z bosons, where $m_W = \frac{gv}{2}$ and $m_Z = \frac{m_W}{\cos \theta_W}$. Additionally, coupling terms between the Higgs boson and the W^\pm and Z bosons arise, with coupling strengths proportional to the boson masses.

Now, the only missing part of the mathematical description of the SM concerns the fermion masses. These are added in the so-called *Yukawa sector* of the SM Lagrangian, written as

$$\mathcal{L}_{\text{Yukawa}} = \sum_f \underbrace{-\frac{\lambda_f v}{\sqrt{2}} \bar{\psi}_f \psi_f}_{f \text{ masses}} - \underbrace{\frac{\lambda_f}{\sqrt{2}} H \bar{\psi}_f \psi_f}_{f H \text{ interactions}}. \quad (2.27)$$

Here, mass and interaction terms between the H boson field and the fermions are visible. The fermion masses are given by $m_f = \frac{\lambda_f v}{\sqrt{2}}$, where λ_f is a free parameter, the *Yukawa coupling*.

With these introductions, I have presented an overview on the full SM Lagrangian, which is

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} + \mathcal{L}_H + \mathcal{L}_{\text{Yukawa}}. \quad (2.28)$$

Note that only one kinetic term for quarks is included, which was presented as part of both \mathcal{L}_{QCD} and \mathcal{L}_{EW} above for simplicity.

2.1.4 Cross section calculations

A specific interaction of elementary particles, characterized by the incoming and outgoing particles, is often called a *process*. Using the mathematical formalism introduced above, it is possible to predict the probability of a process to occur when elementary particles interact. This probability is usually presented in form of a *cross section* σ , which has the unit of an area: m^2 . More often however, it is given in the unit *barn*, defined as $1 \text{ b} = 10^{-28} \text{ m}^2$. In principle, the cross section can depend on the scattering angle and momentum, therefore an absolute cross section can be calculated by integrating the *differential cross section* $d\sigma$ over all angles Ω and possible momenta p :

$$\sigma = \int_p \oint_{4\pi} \frac{d\sigma}{dp d\Omega}. \quad (2.29)$$

The differential cross-section itself can be calculated. Following *Fermi's Golden Rule*, it is proportional to the square of the so-called *matrix element* \mathcal{M} and a *phase space factor*, parametrizing the number of allowed quantum states of the final state particles (e.g. three different color states for processes resulting in quarks). The matrix element describes the dynamics of an interaction and is what can ultimately be calculated from the SM formalism introduced above. It consists of terms parametrizing the interactions involved, including the coupling constants of the respective forces, and *propagators* of intermediate particles. Using the so-called *Feynman rules*, processes can be visualized as *Feynman diagrams* and the corresponding matrix element approximated through perturbation theory. Expanding the dependency on the SM coupling constants as a power series, processes involving the smallest number of couplings (for which the result is non-zero) are referred to as *leading-order* (LO) processes. From here, one can define *next-to-LO* (NLO) and *next-to-NLO* (NNLO), considering additional terms with increasing powers of the coupling constant.

Figure 2.3 shows the cross sections of various SM processes in particle collisions at different centre-of-mass energies. As one can see, SM processes span a very wide range of cross sections. Notably, these are predictions for proton-proton (or proton-antiproton) interactions, which are not elementary particles themselves: more information on the physics of protons is presented in section 3.1.3.

Renormalization and factorization scales Assuming the cross section of some process $ab \rightarrow cd$ is calculated, as stated above this requires an integration over the matrix element of that process: $\sigma \propto \int_p \mathcal{M}(ab \rightarrow cd)$. If divergences appear in the perturbative expansion of the matrix element, the resulting cross section will be

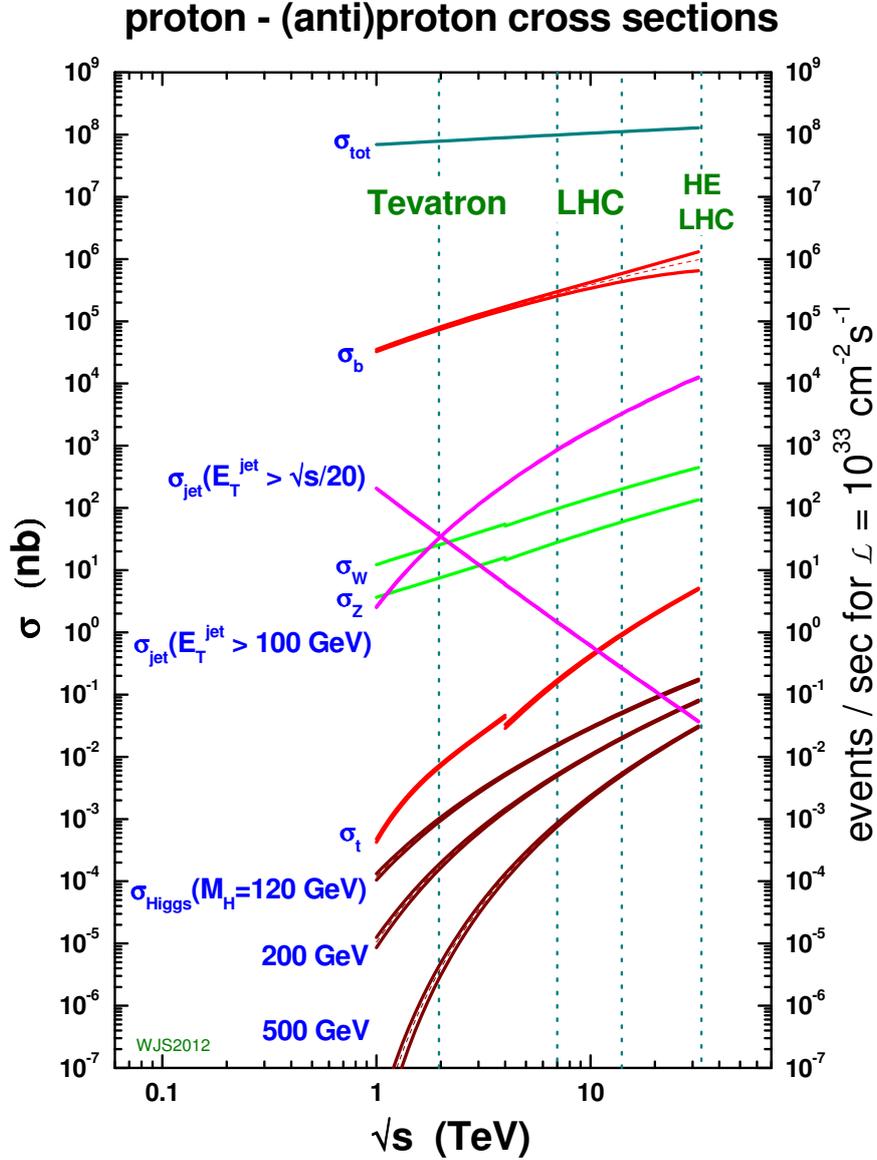


Figure 2.3: Cross sections of various SM processes at particle colliders, as a function of their centre-of-mass energy \sqrt{s} . Steps in some lines result from a switch from $p\bar{p}$ collisions (Tevatron) to pp collisions (LHC). On the right y-axis, the resulting event rates assuming an instantaneous luminosity (introduced in section 3.2.2.1) of $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ are shown. This figure predates the discovery of the Higgs boson, and therefore shows σ for multiple potential Higgs boson masses. Taken from [45].

divergent [46]. One distinguishes between ultraviolet (UV) and infrared (IR) divergences, where in the SM UV divergences occur due to loop contributions with high momenta, and IR divergences appear when particles with zero momentum are described. In order to address these issues, *renormalization* and *factorization* are employed, in which a dependency of the coupling constant on two new parameters, the renormalization scale μ_r and factorization scale μ_f , is assumed. The renormalization is linked to the so-called *running* of the coupling constants: $\alpha_s(\mu_r)$. The factorization scale has an influence on the parton distribution functions which I introduce in section 3.1.3. As μ_r and μ_f are unphysical parameters (which only result from the limited-order approximation of the matrix element), physical results should be independent of the choice of μ_r and μ_f .

To complete this approach, in what is called *regularization*, the SM formalism is assumed to only be valid up to a cutoff scale Λ . This makes it possible to cancel out the divergent terms (which are now dependent on Λ) and results in a non-divergent theory. Like μ_r and μ_f , the choice of cutoff scale Λ should not influence any physical properties.

2.2 Physics beyond the Standard Model

The SM is a highly successful description of reality. It accurately predicts results of particle physics experiments, the most notable example being the discovery of the Higgs boson in 2012 [34–36]. Additionally, it shows great internal consistency in global parameter fits, combining many individual measurements [47, 48]. However, despite these successes, the theory has limitations: some effects observed by experiments can not be explained, and there are some intrinsic structural issues yet to be solved. These shortcomings of the SM motivate the continuous work on both the experimental and theoretical side of particle physics. Thus, they also motivate the work presented in this thesis, which is why I present an overview on the various shortcomings in the following sections. Motivated by these shortcomings, I continue with an introduction on various potential extensions of the SM, which could solve one or more of its issues.

2.2.1 Observational shortcomings of the SM

There are several experimental results in which deviations from the SM predictions on specific parameters were measured. Examples include a measurement of the W mass by the CDF collaboration [49], deviating from the SM by 7.0σ ; differences in respect to the SM prediction in measurements of the anomalous magnetic dipole moment of the muon by the muon $g-2$ experiment [50]; or lepton flavor anomalies in meson decays at the LHCb experiment [51–53]. Some deviations might disappear under further study due to increased available statistical power or currently unknown errors in the analysis techniques. An example are the lepton flavor anomalies, where recent LHCb results are consistent with SM predictions [54–59]. However, there are some more general effects that the SM can not explain on a conceptual level. A major shortcoming is the exclusion of gravity from the SM. Some other issues are introduced below.

2.2.1.1 Neutrino oscillations and masses

One of the most striking shortcomings of the SM is the existence of neutrino masses. As I mentioned above, these particles are assumed to be massless in the context of the SM. However, this is contradicted by the discovery of *neutrino oscillations*, in which neutrinos are able to change flavor, similar to the quark flavor changes described by the CKM matrix. First hints were found in a deficit of solar neutrinos in 1968 [60], and first evidence for atmospheric neutrino oscillations was found in 1998 [61] by the Super-Kamiokande experiment [62]. Neutrino oscillations, described by the *Pontecorvo–Maki–Nakagawa–Sakata* (PMNS) matrix, would only be possible in case of non-zero neutrino masses, resulting in an inequality of mass and interaction eigenstates. Assuming that three neutrino flavors exist, non-zero mass differences can be experimentally determined in neutrino oscillation experiments: $\Delta m_{21}^2 = (7.53 \pm 0.18) \times 10^{-5} \text{ eV}^2$ and $\Delta m_{32}^2 = (2.437 \pm 0.033) \times 10^{-3} \text{ eV}^2$ [43]. The absolute neutrino masses are investigated by dedicated experiments like KATRIN [63], which found $m_\nu < 0.7 \text{ eV}$ [64]. Until now only upper limits on the absolute neutrino mass could be set.

2.2.1.2 Dark matter and dark energy

Already in 1933, observations of the velocity dispersion in the Coma galaxy cluster motivated the existence of *nonluminous matter* to explain the required mass density, which could not be fulfilled by visible matter alone [65]. Since then, a large amount of research has been performed on this so-called *dark matter* [66]: Rotation curves of the orbital velocity of gas and stars in galaxies provided more evidence for dark matter, as well as gravitational lensing effects, analyses of the cosmic microwave background and cosmological calculations. Today, in the Λ CDM model, it is assumed that about $(26.5 \pm 0.7) \%$ of the energy in the universe is dark matter, in contrast to only $(4.93 \pm 0.06) \%$ ordinary, baryonic matter [43]. The SM does not include any particle that could explain this prevalence of invisible matter. Dark matter candidates would need to be massive particles that are neutral in respect to the electromagnetic and strong interactions. Another large part of the energy density in the universe, the remaining $(68.5 \pm 0.7) \%$ according to the Λ CDM model, is assumed to be *dark energy* [43], motivated by the increasing acceleration of the expansion of the universe. Similar to dark matter, dark energy is not described by the SM.

2.2.1.3 Baryon asymmetry

Another cosmologically motivated shortcoming of the SM is the *baryon asymmetry*. Today, an abundance of matter over antimatter is observed. However, assuming a genesis of the universe from a vacuum, some process must have resulted in this imbalance, which has fulfill the three so-called *Sakharov conditions*: violation of the baryon number conservation, CP-violation and interaction outside of thermal equilibrium. CP-violation, the violation of *charge* (C) and *parity* (P) conservation, was measured in the Wu experiment [67] (P-violation) and Fitch–Cronin experiment [68] (CP-violation), and is described by the SM as a complex phase in the CKM matrix, or in the strong interaction. However, no evidence of CP violation in the SM that is large enough to explain the observed matter-antimatter asymmetry has

been found. Additionally, the SM does not allow for violation of the baryon number.

2.2.2 Structural shortcomings of the SM

In addition to the above listed observational shortcomings, there are arguments for the SM being incomplete or wrong from a structural side. The SM is parametrized by a relatively high number of 19 free parameters. Additionally, the existence of exactly three generations of fermions, or simply the construction as a gauge theory, can be seen as arbitrary. Some structural issues are related to the so-called *naturalness* of the SM. A variety of issues arise from the large difference in strength between the three forces described by the SM, and gravity. Naturalness describes the desire for a theory that explains reality with parameters of approximately the same size. A theory with parameters very far from unity might be correct, but can be viewed as flawed. Two specific examples where strong *fine-tuning* would be required to explain observations within the SM are explained below.

2.2.2.1 Strong CP problem

CP -violation in the weak sector of the SM has been discovered in many of the predicted decays [69–73] by several experiments, and its strength has been found to agree well with the SM predictions. However, the SM also would allow for a CP -violating phase in the strong sector [74, 75]. Accessible through the electric dipole moment of the neutron, which has been measured to be $n_{\text{EDM}} < 1.8 \times 10^{-26} e \text{ cm}$ [43], an upper limit on the CP -violating phase $\bar{\theta}$ in the strong sector can be found. Depending on the way of calculation, the n_{EDM} in the SM is predicted around $n_{\text{EDM}} \approx \bar{\theta} * \mathcal{O}(10^{-16} e \text{ cm})$ [76, 77], which means that $\bar{\theta} < \mathcal{O}(10^{-10})$ would be required: a clear example for fine-tuning in the SM.

2.2.2.2 Hierarchy problem

When calculating quantum corrections to the Higgs boson mass in the SM, quark loops as shown in figure 2.4 result in divergent contributions to m_H [78, 79].

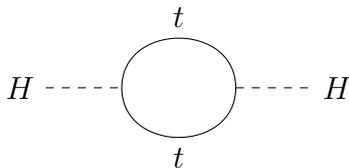


Figure 2.4: Top quark loop diagram contributing to the Higgs boson mass, following [80].

Assuming that the measurable Higgs boson mass is defined as $m_H^2 = m_{H,0}^2 + \Delta m_H^2$ where $m_{H,0}^2$ is the bare Higgs mass before renormalization, one can investigate the corrections Δm_H^2 . The dominant contribution is resulting from the t quark Yukawa coupling λ_t and scales as

$$\Delta m_H^2 \propto -|\lambda_t|^2 \Lambda^2. \quad (2.30)$$

In this equation, Λ is the cutoff scale of the SM introduced in section 2.1.4. As no physics beyond the SM has been found yet, no experimental constraints on this

energy scale can be set, except for the *Planck scale* of 1.2×10^{19} GeV [43], which is the energy scale where gravitation becomes relevant. Given that this contribution is huge, substantial fine-tuning would be required to explain the Higgs mass of around 125 GeV that was measured experimentally.

2.2.3 Extensions of the SM

To address the shortcomings of the SM, many different theories describing physics *beyond the Standard Model* (BSM) are being created. These aim to provide a mathematical description of reality, similar to the SM, that can solve one or multiple of the SM's issues. Ideally, proposed theoretical models offer a handle to experimentally support or exclude them, like the existence of some new particle, or changes to the SM predictions of measurable quantities. I give a non-comprehensive overview of common BSM theory types in the following.

2.2.3.1 Supersymmetry

The construction of the SM as a gauge theory, as introduced above, is based on symmetries. When studying potential extensions to solve its shortcomings, basing these extensions on an additional symmetry is natural, and the core principle of *supersymmetry* (SUSY) [81–84]. As a category of new physics models, it includes all theories that propose a symmetry between fermions and bosons: an invariance under spin transformation.

The most simple SUSY model is the *Minimal Supersymmetric Standard Model* (MSSM) [85]. In it, for each fermion there is a boson with the same properties (same couplings, charges, but different spin), and vice versa. Crucially, the Higgs sector is extended with another complex isospin doublet, which leads to the existence of five physical Higgs particles, one of which is the light SM-like Higgs.

The new Higgs fields participate in the electroweak symmetry breaking and mix with the neutral SM fields to form the so-called *neutralinos*. Therefore, SUSY introduces a promising dark matter candidate in form of the *lightest supersymmetric particle*, which is realized by the lightest neutralino in most SUSY models. Additionally, SUSY models would provide a simple solution to the hierarchy problem, as the contribution of fermion and scalar loops would cancel out (assuming both particles have the same mass).

SUSY models can be searched for at particle colliders as they predict the existence of detectable new particles and the fact that these should be relatively light (TeV scale). However, no signs of SUSY have been found at the LHC or previous experiments, disfavoring SUSY models assuming mass degeneracy of SM and SUSY particles. Nevertheless, research for SUSY models with heavier particles is ongoing.

2.2.3.2 Grand Unified Theories

As I outline in section 2.1, the electromagnetic and weak interactions are combined into the electroweak interaction. *Grand Unified Theories* (GUTs) [86] propose a further unification of all three fundamental forces of the SM into a single, higher-dimensional symmetry. This could be realized by a $SU(5)$ group, as first proposed in 1974 [87], or even higher-order groups. Offering handles to explain dark matter, the relation between quark and fermion generations or non-zero neutrino masses,

GUTs are an exciting field of study. Investigating the energy-dependence of the SM interactions, their coupling strengths will be almost the same at around 10^{15} GeV, a behavior than could be improved to an exact match assuming the MSSM were true, providing additional motivation for study of GUTs. Results on proton decays, neutrino masses or other precision measurements could give experimental access to the existence of GUTs. Additionally, theories proposing detectable particles exist as well: if a GUT combines quarks and leptons into n -plets (instead of separate doublets as in the electroweak formalism), this would motivate the existence of *leptoquarks* which can mediate transformations between quarks and leptons. Leptoquarks are studied in various experiments and hints towards their existence have been observed [88], but no discovery has been made yet. Unified theories that include gravity are studied as well, usually called *theories of everything*.

2.2.3.3 Composite particles

Under the label of *compositeness* [89], various BSM theories are grouped, all of which predict some of the particles known to the SM to not actually be fundamental. When quarks were first proposed, a motivation was the reduction of the large number of mesons and baryons discovered at the time into a lower number of fundamental particles. Similarly, some compositeness models suggest that quarks and leptons have substructure and are built from so-called *preons*, fundamental point-like particles. These kind of models could explain the existence of the generations as well as the relations between charges and masses of fermions, neutrino oscillations and dark matter [90]. However, with the discovery of the Higgs boson, existence of preons as a way to describe particle masses without need for a Higgs field was disfavored. Still, many other BSM models predict composite quarks to exist in order to provide solutions to the SMs shortcomings [91–98].

Composite Higgs models are still highly relevant. They assume the Higgs boson to not be a fundamental particle, but instead a bound state of particles from a BSM sector. This can be realized in two different ways: a Higgs boson could be created as a “light scalar (dilaton-like) particle of the new strong dynamics or it arises similar to pions in QCD as a pseudo Nambu-Goldstone boson” [99]. Composite Higgs models provide a solution to the hierarchy problem, and usually predict the existence of new particles at $\mathcal{O}(\text{TeV})$ providing a handle to search for composite Higgs models. Composite Higgs models introducing a light Higgs as a Goldstone boson are also called little Higgs models [100–103]. Searches for composite Higgs models, accessible through the Higgs boson properties, or predictions of other new particles, are performed at various experiments, but have not resulted in a discovery so far.

2.2.3.4 Models with extra dimensions

Another category of BSM models try to solve the SMs shortcomings through the addition of *extra dimensions*, mostly targeting a solution to the hierarchy problem and an inclusion of gravity into the SM, but also providing solutions to the strong CP problem. In models with *large extra dimensions* [104, 105], gravity acts on higher dimensions than the other fundamental forces, resulting in the weaker strength observed in our $3 + 1$ (space + time) dimensional world. Another example are Randall–Sundrum models [106–108], in which a fifth dimension is added, where

our reality sits on a so-called *brane* on which gravitation acts only weakly. Extra dimension models are accessible through predictions of new various particles and are researched at particle collider experiments. Particle physics models from the field of *string theory* [109, 110], describing elementary particles not as point-like, but as one-dimensional strings, require the existence of higher dimensions and are thus related to extra dimension models.

In this thesis, I present a search for a potential signature of BSM physics. In order to perform this kind of research, an experimental apparatus is needed. Therefore, in the following, I present the infrastructure I use to obtain the results I detail subsequently: In section 3.1, I introduce the Large Hadron Collider, and in section 3.2 the CMS detector.

3.1 Large Hadron Collider

The *Large Hadron Collider* (LHC) [111] is a particle accelerator located at the *European Organization for Nuclear Research* (CERN) near Geneva, Switzerland. As a hadron collider, it is designed to increase the kinetic energy of hadrons in order to study the physics of highly energetic particle collisions. Most of the time, protons are accelerated within the LHC, so I will focus on this operation mode in the following. However, it is also possible to accelerate heavy ions, like lead [112].

3.1.1 The LHC pre-accelerator chain

The LHC is not able to accelerate protons from rest. Instead, a chain of pre-accelerators, shown in figure 3.1, is used to subsequently increase the proton energy before their injection into the LHC. First, hydrogen atoms are ionized by passing them through an electric field, obtaining free protons. These are accelerated to 50 MeV using *Linac2* [113], a linear accelerator equipped with *radio frequency* (RF) cavities. After leaving the linear accelerator, three circular accelerators are used to further increase the proton energy: the *Proton Synchrotron Booster* (BOOSTER) [114], made up of four superimposed synchrotron rings, accelerates them to 1.4 GeV; the *Proton Synchrotron* (PS) [115] to 26 GeV; and the *Super Proton Synchrotron* (SPS) [116] to 450 GeV. At this energy, the protons are then injected into the LHC itself. Some changes to the pre-accelerator chain were made after 2018, which I summarize in appendix A.1.1.

3.1.2 Particle acceleration at the LHC

The LHC is a circular accelerator with a circumference of 26.7 km, located in an underground tunnel which is between 45 m and 170 m below the surface. It reuses the tunnel previously housing the *Large Electron-Positron Collider* (LEP) [118], which collided electrons and positrons at energies up to 209 GeV.

At the LHC, two proton beams, traveling in opposite directions around the ring, are kept on a circular trajectory using a magnetic field with a field strength of 8.3 T, created using superconducting dipole magnets. In addition to these, several types

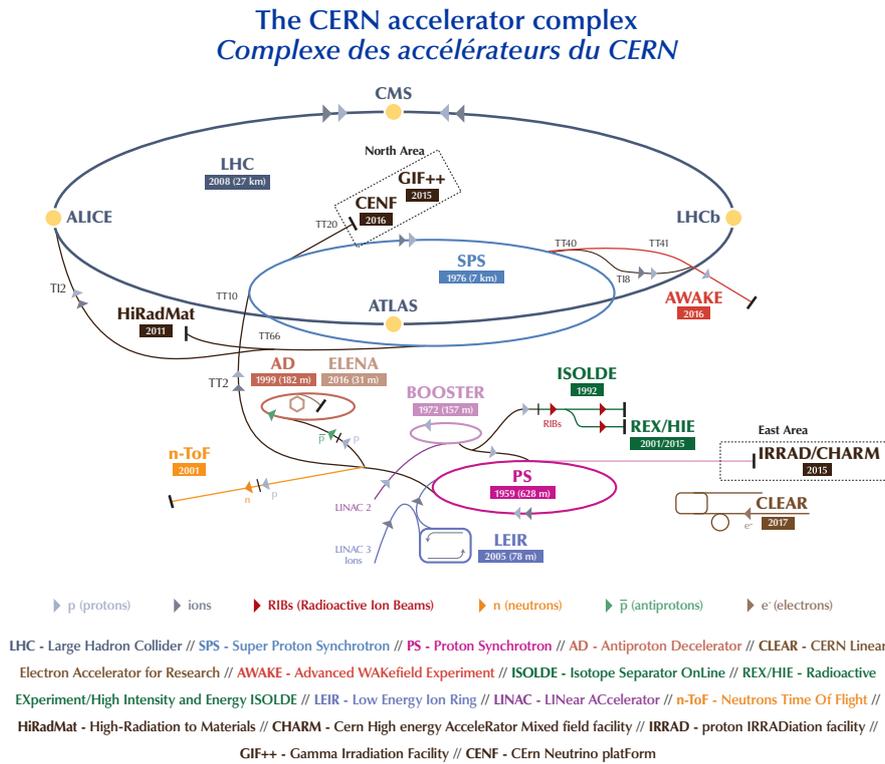


Figure 3.1: An overview of the CERN accelerator complex in 2018, including the LHC and all its pre-accelerators: the Linac2, PSB, PS and SPS. Other accelerators not part of the LHC pre-accelerator chain are also shown, as well as several experiments making use of the available acceleration infrastructure. Taken from [117].

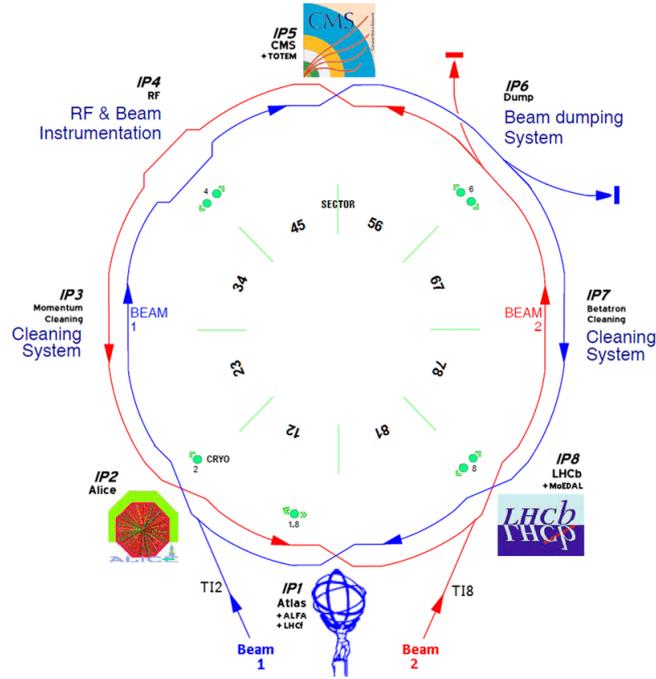


Figure 3.2: Schematic overview of the LHC, showing the beams (in blue and red) circulating around the accelerator ring. Displayed are all eight IPs, housing the four experiments of the LHC, as well as the machinery required for the particle acceleration: the RF system, beam instrumentation and cleaning system, as well as the beam dump used to remove the protons from the experiment. Taken from [122].

of smaller superconductive quadrupole magnets are used to focus the proton beams, and higher order magnets are used for further beam stabilization.

To accelerate the protons, 16 *superconducting RF cavities* [119] are used, each increasing the proton beam energy by applying a maximum voltage of 2 MV [120]. Within about 20 minutes, the protons beams are brought to an energy of 6.5 TeV. Due to the pulsing acceleration structure of the RF cavities, the proton beams are not continuous, instead they consist of many proton *bunches*, each containing about 1.15×10^{11} protons [121]. The two proton beams are collided at four *interaction points* (IP), at a rate of around $f = 40$ MHz, with a *centre-of-mass* energy of $\sqrt{s_{pp}} = 6.5 \text{ TeV} + 6.5 \text{ TeV} = 13 \text{ TeV}$. Each instance of a single collision is referred to as a collision *event*.

Four experiments measure the collision events delivered by the LHC, as shown in figure 3.2. The CMS experiment, topic of this thesis, is introduced in detail in section 3.2. Constructed similarly to CMS is ATLAS (A Toroidal LHC Apparatus) [123], designed as a general-purpose particle detector able to investigate all particle collisions produced by the LHC. LHCb (LHC-beauty) [124] targets the physics of B mesons and is characterized by its asymmetric detector layout, focusing on particles with Lorentz boosts in the direction of the beam axis. Finally, ALICE (A Large Ion Collider Experiment) [125] is specifically designed to precisely measure heavy ion collisions using sophisticated particle identification techniques.

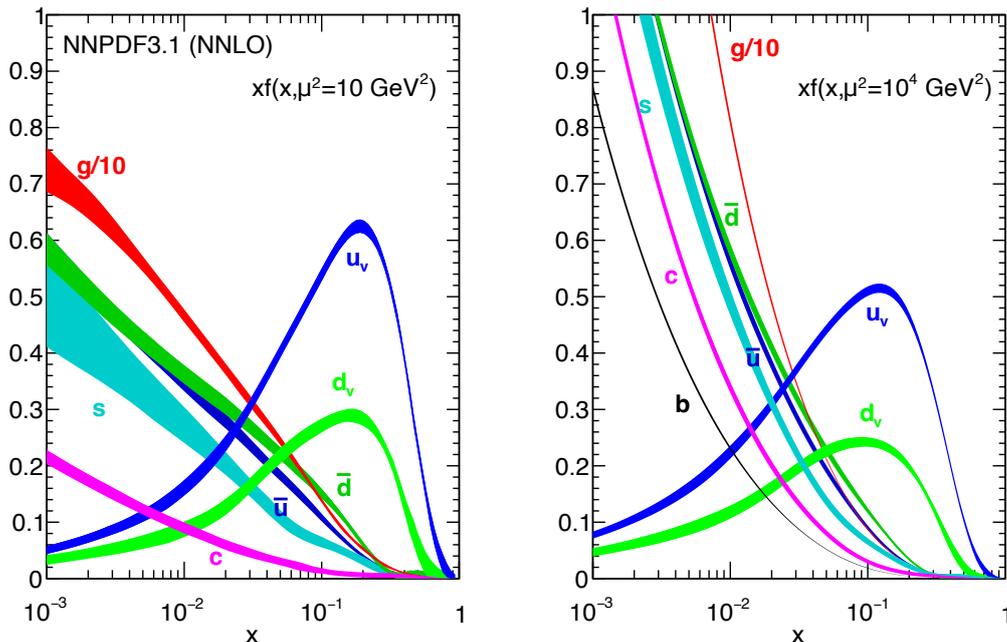


Figure 3.3: Left: PDFs evaluated at $\mu^2 = 10 \text{ GeV}^2$, showing the probability to find a specific parton as a function of its x . Right: PDFs evaluated at $\mu^2 = 10^4 \text{ GeV}^2$. Figure taken from [129], which presents NNLO precision PDF calculations.

3.1.3 Proton collider physics

The LHC accelerates and collides protons. In contrast to other particle accelerators like LEP [118] or SuperKEKB [126] (a currently existing electron-positron collider), the collided particles are not elementary. Thus, physics at a proton collider has some special features, which I outline here. Protons are made up from two u -quarks and one d -quark, which are called *valence quarks*. However, beyond these three particles, there are more to be found inside a proton: gluons, mediating the strong interaction holding together the valence quarks, are present, as well as so-called *sea-quarks*. These quarks, which can have any flavor, result from gluons splitting into short-lived quark-antiquark pairs. All possible constituents of the proton are called *partons* [127]. Assuming an accelerated proton, each parton carries a certain fraction of the total proton momentum: $p_{\text{parton}} = x p_{\text{proton}}$, where x is the so-called *Bjorken scaling variable* [128]. The inner structure of protons is described by *parton distribution functions* (PDF) $f(x, \mu^2)$, which parameterize the probability of finding a parton with a specific energy fraction x when probing the proton at an energy μ^2 . In figure 3.3, I show the results of theory calculations determining the PDFs at different μ^2 . There, one can see that valence quarks dominate for large x , whereas other partons like gluons become more relevant for smaller x , or larger μ^2 .

When calculating the cross section of SM processes at the LHC (as introduced in section 2.1.4), the proton substructure needs to be taken into the account:

$$\sigma = \sum_{a,b} \int \int \underbrace{f_a(x_1, \mu_F^2)}_{\text{PDF of parton } a} \underbrace{f_b(x_2, \mu_F^2)}_{\text{PDF of parton } b} \underbrace{\sigma_{a,b}(s_{a,b}, \mu_R^2, \mu_F^2)}_{\sigma \text{ from matrix element}} dx_1 dx_2 \quad (3.1)$$

Here, the total cross section is a sum of all possible combinations of partons a, b ,

scaled by the likelihood of their respective occurrence, as predicted by the PDFs. The total cross section results from an integral over all possible momentum fractions x_1 and x_2 , and dependencies on the renormalization and factorization scales are visible (see section 2.1.4). Crucially, as x_1 and x_2 can be different, the centre-of-mass energy $\sqrt{s_{a,b}}$ of the interaction is not equal to the proton-proton centre-of-mass energy $\sqrt{s_{pp}}$. Instead, it is given as

$$\sqrt{s_{a,b}} = \sqrt{x_1 x_2 s_{pp}}. \quad (3.2)$$

Due to the inequality of x_1 and x_2 , the decay products of proton-proton collisions can have non-zero Lorentz boosts along the beam axis. This is a crucial difference between hadron and lepton colliders, and necessitates the usage of transverse variables, as introduced below.

3.2 Compact Muon Solenoid experiment

The Compact Muon Solenoid (CMS) [37] experiment is a general-purpose detector, targeting measurements of all SM and potential BSM physics processes. I give an overview of the detector parts in section 3.2.1, and discuss luminosity and data taking runs in section 3.2.2. Afterwards, I introduce the infrastructure needed for data acquisition in section 3.2.3, including a detailed description of the CMS trigger system. Finally, I give an overview on the reconstruction of physics objects from the detector readout in section 3.2.4.

3.2.1 Detector overview

To describe the CMS detector and the particles within it, a coordinate system needs to be defined. For this, the nominal collision point of the two proton beams, located in the center of the detector, serves as the origin. A z -axis then points along the beam axis, where positive values of z refer to the direction of the counter-clockwise beam. A radius r from this axis is defined, and an azimuthal angle ϕ and polar angle θ complete the coordinate system. A full definition is available in [37]. As I introduced in section 3.1.3, the substructure of the colliding photons leads to unknown Lorentz boosts of the decay products along the beam axis. This affects the angles θ of objects in the detector, and assigning a physical meaning to θ differences between different objects is challenging. Therefore, the *pseudorapidity* η , defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \quad (3.3)$$

is used instead. Pseudorapidity differences $\Delta\eta$ are invariant under Lorentz transformations in z -direction, making them suitable variables to describe polar distances between to objects in the detector. The *angular distance* ΔR between two objects is then given as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \quad (3.4)$$

Finally, due to the potential unknown Lorentz-boost along the z -axis, the total momentum of an object is usually not of interest. Instead, the *transverse momentum* p_T is used, defined as

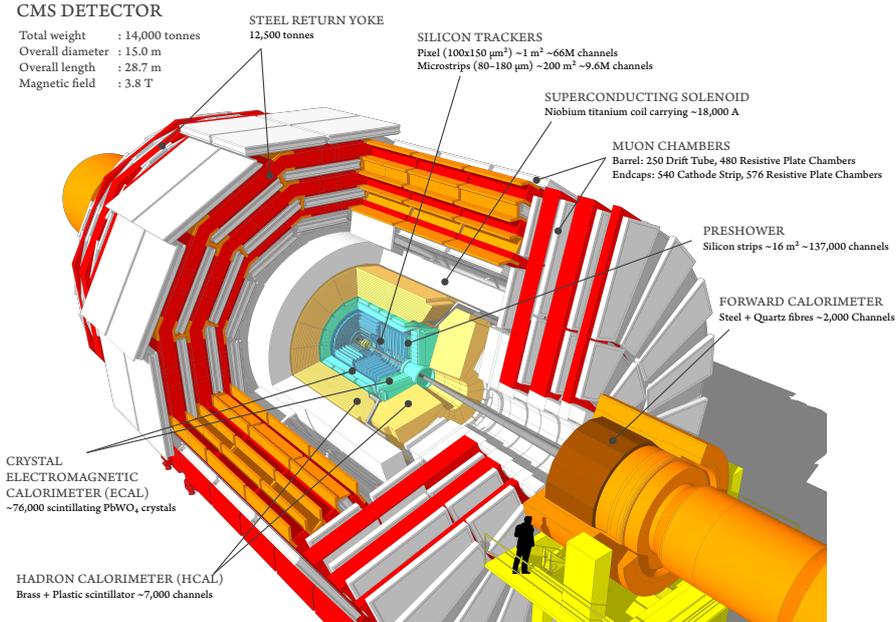


Figure 3.4: Technical drawing of CMS, cut open to show the various detector parts, and some key machine parameters. Taken from [130].

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (3.5)$$

As the total transverse momentum prior to the collision is zero by design, due to momentum conservation the p_T sum of all objects produced in a collision must sum up to zero. Therefore, a particle escaping the detector without interacting is visible as *missing transverse momentum* p_T^{miss} .

With the coordinate system defined, I will introduce the individual parts of CMS next. It has a cylindrical layout with different layers arranged around the central *beam pipe*. All detector parts are shown in figure 3.4.

3.2.1.1 Tracking system

The innermost part of the CMS detector is the *tracking system* [131]. It is designed to precisely measure the trajectories of charged particles moving through it, while keeping the reduction of their energy to a minimum. This is achieved with multiple layers of semiconducting silicon detectors, covering a region of $|\eta| < 2.5$. A schematic drawing of the tracking system is shown in figure 3.5.

The tracking system consists of different parts. At the very center, directly outside of the beam pipe, a *silicon pixel detector* (PIXEL) is located. In 2016, it consisted of three barrel layers and two endcap discs. It provides a three-dimensional measurement of the point where particles hit the silicon sensors, with a hit efficiency of $> 99\%$ [132]. The position resolution of each hit is found to be $9.4 \mu\text{m}$ in the $r\phi$ -direction, and between 20 and $45 \mu\text{m}$ in z -direction, depending on the angle at which the particle passes through the sensor [132]. Between 2016 and 2017, the PIXEL was replaced by an upgraded version [133], figure 3.6 shows a comparison between the two configurations. It now consists of four barrel layers and three endcap discs. The upgraded PIXEL is "designed to cope with the higher instantaneous luminosities that have been achieved by the LHC" [133]. The high hit efficiency is

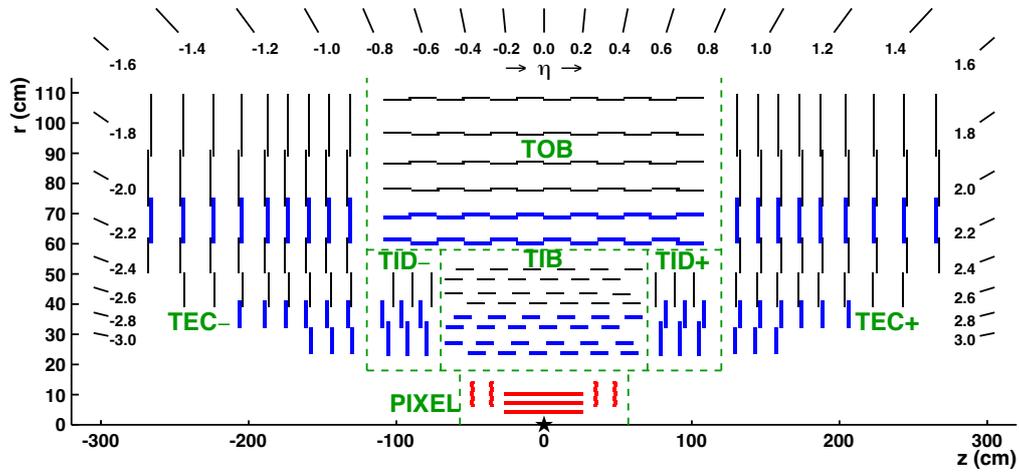


Figure 3.5: The 2016 configuration of one half of the CMS tracking system. The barrel and endcap parts of the central PIXEL system are shown in red. Different parts of the strip detectors are shown in blue (two back-to-back strip modules) and black (single strip modules). The strip detector is divided into different regions by green dashed lines. Taken from [132].

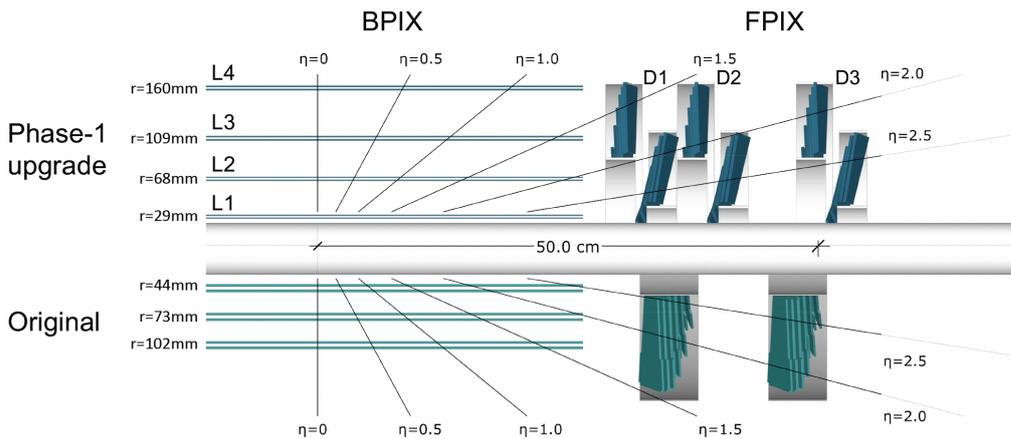


Figure 3.6: Schematic drawing of the CMS PIXEL detector setup, showing the setup used in 2016 (lower half) and the upgraded setup used in 2017 and 2018 (upper half). The detector is split into a barrel (BPIX) and endcap (FPIX) part. Taken from [133].

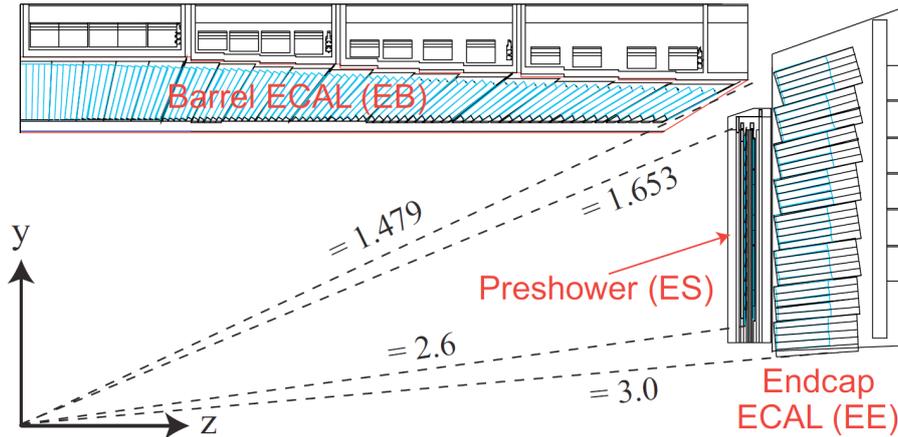


Figure 3.7: Schematic drawing of one quarter of the CMS ECAL, showing its barrel and endcap parts, as well as the preshower system. Diagonal lines show the η coordinate, visualizing the ECAL reach of $|\eta| < 3$. Taken from [136].

mostly retained and the hit resolution is measured (for the third barrel layer as an example) to $9.5\ \mu\text{m}$ in $r\phi$ - and $22.2\ \mu\text{m}$ in the z -direction [133]. Additionally, the upgraded PIXEL has less material in the endcap region, thus lowering the effect on the energy of particles passing through.

Outside of the PIXEL, a *silicon strip detector* is situated, split into four parts: The *tracker inner barrel* (TIB), consisting of four layers; and the *tracker inner discs* (TID) consisting of three endcap discs, cover $r < 55\ \text{cm}$ and $|z| < 118\ \text{cm}$. Outside of these, the *tracker outer barrel* (TOB) with six layers and the *tracker endcaps* (TEC) with nine discs extend the total coverage up to $r < 116\ \text{cm}$ and $|z| < 282\ \text{cm}$. The overall hit efficiency in the strip detector is close to 100%. The position measurement resolution varies from 13–38 μm in the inner to 18–47 μm in the outer part of the strip detector [132, 134].

3.2.1.2 Electromagnetic calorimeter

Moving further outside from the particle collision point, the next subsystem of CMS is the *electromagnetic calorimeter* (ECAL) [135]. This detector part is designed to measure the energy of light electromagnetically interacting particles, for example photons or electrons. When passing through material, these particles can produce an *electromagnetic shower*, ultimately depositing their energy their energy in the detector material, thus allowing a measurement of the initial particles energy.

Similar to the tracker, one distinguishes between the barrel and endcap parts of the ECAL. In total, the ECAL covers an area of $|\eta| < 3$, as shown in figure 3.7. Build as a hermetic homogeneous calorimeter, the ECAL is made from scintillating lead tungstate crystals. Energy deposited in this material is visible as *scintillation light* and detected by photodetectors. About 99% of the produced light is collected within 100 ns [136], making the ECAL a very fast detector. The initial performance of the CMS ECAL was measured with a test beam of electrons between 20 and 250 GeV [137]. From these measurements, the *energy resolution* $\frac{\sigma_E}{E}$ of the ECAL can be parameterized as

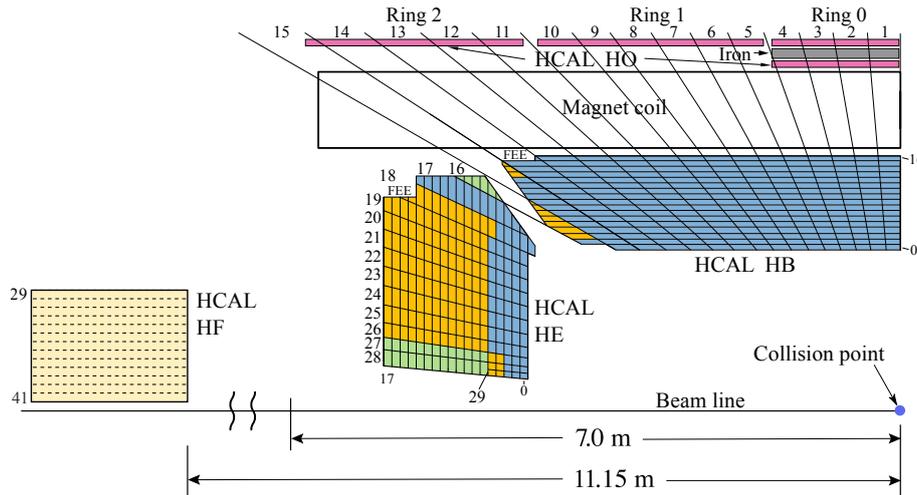


Figure 3.8: Schematic overview of a quarter of CMS HCAL, showing its different parts: the hadron barrel (HB), hadron endcap (HE), hadron outer (HO), and hadron forward (HF) calorimeters. Taken from [140].

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E[\text{GeV}]}} \oplus 0.3\% \oplus \frac{12\%}{E[\text{GeV}]} \quad (3.6)$$

The first term describes the stochastic development of the electromagnetic shower, scaling with $\frac{1}{\sqrt{E}}$. The second term, which is constant in energy, describes the resolution from limited accuracy in the calibration of each ECAL cell and is dominant for high energies. The third term, scaling with $\frac{1}{E}$ and thus being dominant for low energies, describes effects from background *noise*. The ECAL performance is constantly monitored, understanding and correcting, for example, for crystal transparency changes. Thanks to this, the energy resolution is kept as low as 2% in the center of the ECAL barrel, and below 2% in the endcap [138].

3.2.1.3 Hadronic calorimeter

Similar to electromagnetically interacting particles, it is important to measure the energy of hadronically interacting objects in the detector. These deposit their energy in material in a *hadronic shower*, similar to the electromagnetic shower introduced above, but driven by hadron production, excitation and deexcitation of nuclei; and particle decays. Notably, all hadronic showers have an electromagnetic component. It is not possible to catch the full energy of a heavy hadron in the ECAL due to its limited depth. Therefore, a second calorimeter, the *hadronic calorimeter* (HCAL) [139] is used to measure hadronic particles energies at CMS. In contrast to the homogeneous ECAL, the HCAL is a *sampling calorimeter*, made from high density brass or steel absorbers and lower density plastic scintillation material. Only about 7% of the total energy deposited in the calorimeter is visible as scintillation light. However, knowing this fraction allows to still determine the initial particles energy. The HCAL structure is displayed in figure 3.8, the showing different parts covering a total range of $|\eta| < 5.19$.

Similar to the ECAL, also the HCALs performance is monitored. Assuming pions that do not interact significantly in the ECAL, the energy resolution is found to be

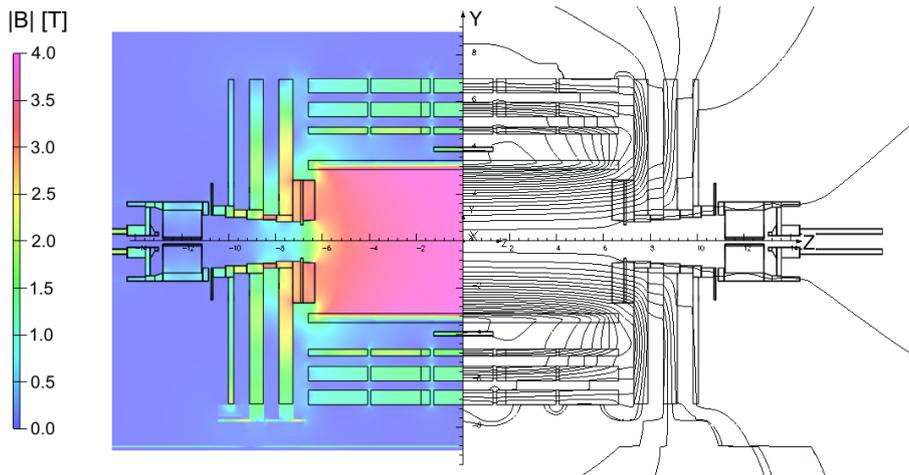


Figure 3.9: Field strength of the CMS magnetic field, shown in color in the left half, and as field lines in the right half of the figure. Taken from [142].

$$\frac{\sigma_E}{E} = \frac{91\%}{\sqrt{E[\text{GeV}]}} \oplus 4\% \quad (3.7)$$

in test beam studies [139]. In practice, the energy resolution in data taken with CMS is usually monitored for jets, which are reconstructed using combined tracker and calorimeter information and are discussed in detail in section 3.2.4.2.

3.2.1.4 Solenoid magnet

Precisely reconstructed particle trajectories from the tracking system can be used to determine the momentum of charged particles. When moving through a magnetic field, charged particles are affected by the Lorentz force, resulting in a curved trajectory. Assuming a known magnetic field magnitude B , the radius of the curvature allows to access the momentum of the particle. Additionally, the direction of the curvature allows to distinguish positively and negatively charged particles. In order to induce a magnetic field into the tracking system volume, a *superconducting solenoid magnet* [141] is used, which is located outside of the calorimeters. Producing a magnetic field with a maximum of 3.8 T, the magnet consists of a superconducting coil and a magnet yoke. The magnetic field is mostly homogeneous within the tracking system, as can be seen in figure 3.9, and wraps around in the muon chambers. For accurate reconstruction (and simulation) of objects in the detector, precise knowledge of the magnetic field map is crucial. It can be determined using measurements of cosmic muon rays [142].

3.2.1.5 Muon system

The outermost part of CMS is the muon system [143]. Muons pass through the calorimeters and the solenoid magnet without being stopped, thus a dedicated detector system is used to measure these particles. An overview is given in figure 3.10.

Similar to the tracking system, a measurement of the trajectory of muons passing through the muon system is targeted. Detector hits are measured using gas-filled

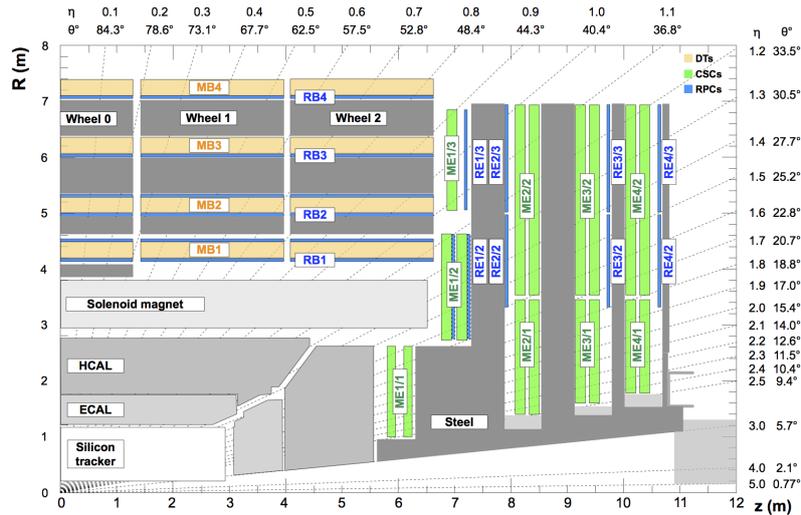


Figure 3.10: Schematic cutout of the CMS detector, showing a quarter of the muon system. Drift tubes in the barrel are shown in yellow and labeled MB. Cathode strip chambers in the endcap regions are shown in green and labeled ME. Resistive plate chambers are shown in blue and labeled RB (in the barrel) and RE (in the endcap), respectively. Taken from [144].

drift tubes (DT) in the barrel region and *cathode strip chambers* (CSC) in the endcaps, allowing to determine the two-dimensional hit position. These detectors are accompanied by *resistive plate chambers* (RPC) [145] which are primarily used for triggering (see 3.2.3), thanks to their fast response time. As the muon system is outside of the solenoid magnet, the magnetic field wraps around and curves muon trajectories into the opposite direction compared to the tracking system, which allows momentum measurement like in the tracking system. The muon system is able to measure muons with a high hit efficiency of 94 to 99%. Additionally, it has very precise hit spatial resolution of 50 to 300 μm and momentum resolution of 1 to 3% for muons with $p_T < 100 \text{ GeV}$, which stays below 7% for muons with $p_T < 1 \text{ TeV}$ [144].

3.2.2 Data collected by CMS

The LHC delivers collision data to CMS, which is measured using the detector parts I described above. Here, I briefly discuss quantification of this data in terms of *luminosity* and the different eras of data taking.

3.2.2.1 Luminosity

In section 2.1.4, I show the cross section of various physics processes. The number of events per second $\dot{N}(t)$ of a process occurring in LHC proton-proton collisions is calculated as

$$\dot{N}(t) = L(t) \sigma \quad (3.8)$$

where $L(t)$ is the so-called *instantaneous luminosity*, describing the number of proton-proton collisions produced by the LHC per area and time. To estimate

the total amount of collision events N of a process in a given time, the *integrated luminosity* L is used:

$$N = \sigma L = \sigma \int L(t) dt. \quad (3.9)$$

Often, the integrated luminosity is given in inverse barn (b^{-1}), analogous to cross sections given in barn. To predict how many events of a given process are present in the data gathered at one of the LHCs experiments, luminosity measurements are crucial. Therefore, dedicated luminosity measurements are performed. At CMS, the absolute luminosity scale is obtained through *van der Meer scans*. By varying the transverse separation of the two proton beams, it is possible to obtain information on the beam size σ_x and σ_y perpendicularly to the beam direction. Then, the instantaneous luminosity can be calculated as

$$L = N_B f \frac{N_1 N_2}{4\pi\sigma_x\sigma_y} \quad (3.10)$$

where N_B is the number of bunches in the LHC, f is their collision frequency and N_1 and N_2 are the number of protons per bunch. However, as this method requires dedicated scans and thus can not be performed during data taking, different techniques are needed to track changes in luminosity over time. For this, a combination of different measurements is employed, both using the main parts of the CMS detector (as introduced above) and dedicated subsystems (called *luminometers*). Each of these measures some rate observable averaged over $\mathcal{O}(s)$. Scaling this rate with some constant parameter describing the relation between the observable and the instantaneous luminosity, which is derived during van der Meer scans, then yields a measurement of the instantaneous luminosity over time, without the need for continuous dedicated scans. More information on luminosity measurements at CMS can be found in [146].

3.2.2.2 Data taking runs

The LHC is not running continuously, instead its operation is divided into multiple *runs*, each containing multiple years of data taking. Between runs, so-called *long shutdowns* (LS) allow time to maintain and upgrade accelerator and detector technology. The full LHC schedule, including past runs and future plans, is shown in figure 3.11. Data taking started in 2011 with *Run 1*, gathering a total 30 fb^{-1} of data at a $\sqrt{s_{pp}}$ of 7 TeV and 8 TeV.

The second data taking period, *Run 2*, started in 2015 after the end of LS1. However, as the data taken in that year was mostly used for machine commissioning, in the context of this thesis, I use the term Run 2 to refer to the time period between 2016 and 2018. From Run 1 to Run 2, the most important change to the LHC was an increase of $\sqrt{s_{pp}}$ to 13 TeV. During Run 2, the LHC delivered 159.3 fb^{-1} integrated luminosity to CMS (not considering 4.3 fb^{-1} in 2015) [148], an overview of which is given in figure 3.12. Of this data, 146.82 fb^{-1} was recorded by CMS, split into 38.25 fb^{-1} from 2016, 44.9 fb^{-1} from 2017, and 63.67 fb^{-1} from 2018 (not accounting for 3.86 fb^{-1} in 2015). Ideally, CMS would measure all delivered collision events, the reduced recorded luminosity results mainly from temporarily unavailable detector subsystems or a busy DAQ system [149]. The quality of all recorded data is thoroughly examined and the performance of all CMS subsystems is evaluated.

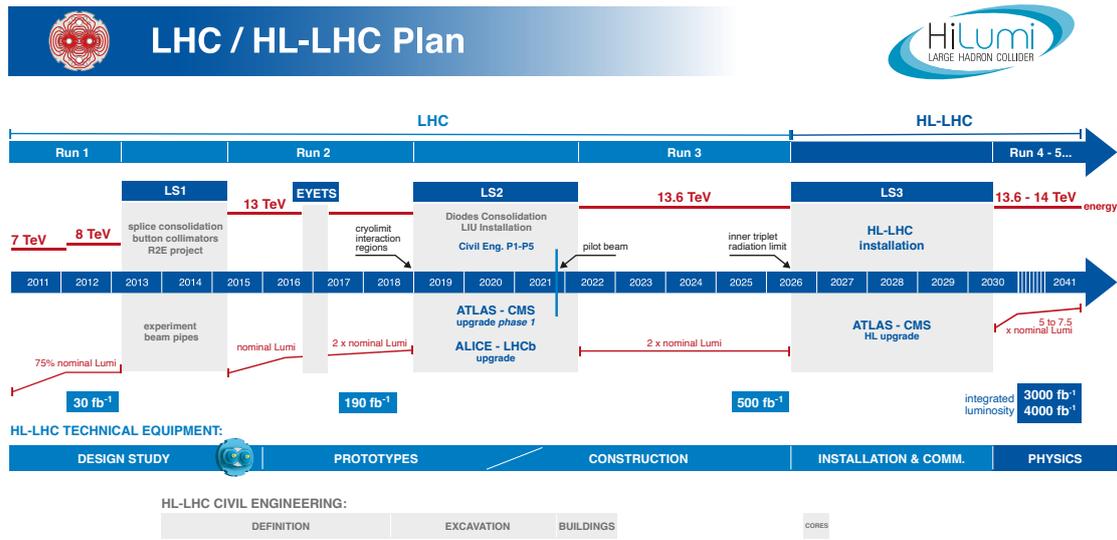


Figure 3.11: Overview of the LHC schedule as of 2024, showing past runs and plans for the future, including the HL-LHC. Taken from [147]

Only if all subdetectors perform as expected, data is certified as usable for physics analysis. This further lowers the luminosity, so that ultimately in Run 2 137.62 fb^{-1} are available for usage, split into 36.31 fb^{-1} from 2016, 41.48 fb^{-1} from 2017, and 59.83 fb^{-1} from 2018 (not accounting for 2.27 fb^{-1} in 2015). The comparison between delivered, recorded and certified luminosity during Run 2 is visualized in figure 3.13. During 2016, the strip detector system saw a decreased signal-to-noise ratio for technical reasons, resulting in a drop in hit efficiency. This was fixed after about half of the 2016 dataset was collected by changing the *feedback preamplifier bias voltage* (VFP). As the detector difference might lead to changes in observed results, the 2016 dataset is split into the *pre-VFP* and *post-VFP* eras containing 19.5 fb^{-1} and 16.8 fb^{-1} of integrated luminosity, respectively [150].

All results and information in this thesis, unless otherwise noted, assume a Run 2 setup, including the description of the LHC above. In 2022 (after LS2 and some delays due to the Covid-19 pandemic), *Run 3* started with another increase of $\sqrt{s_{pp}}$ to 13.6 TeV, closely approaching the design energy of 14 TeV. I present some studies for Run 3 of CMS in this thesis, and therefore list relevant changes to CMS in appendix A.1.2. It is planned to continue until 2026, after which another LS will take place (LS3), followed by the *high luminosity* phase of the LHC (HL-LHC) [152]. Characterized by a planned strong increase in instantaneous luminosity and a $\sqrt{s_{pp}}$ of 14 TeV, it will require substantial upgrades not only to accelerator infrastructure, but also detector components. In appendix A.1.3, I provide some information on the HL-LHC phase of the CMS experiment, and list any differences that are relevant to this thesis.

3.2.2.3 Pileup collisions

The high instantaneous luminosity that is achieved at the LHC results in very high interaction probabilities. Therefore, multiple proton-proton interactions happening

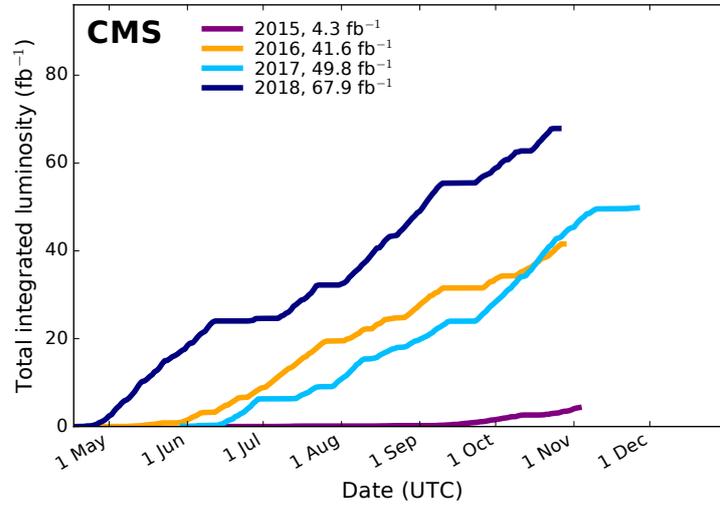


Figure 3.12: Integrated luminosity delivered to the CMS detector by the LHC in the years from 2015 to 2018. Taken from [148].

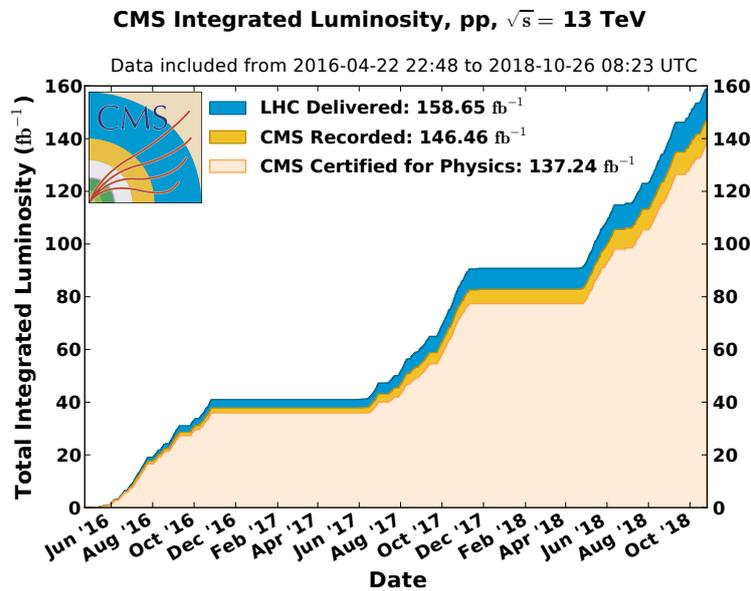


Figure 3.13: Cumulative integrated luminosity during Run 2, comparing delivered, recorded and certified data, taken from [151]. Note that the total certified luminosity number differs slightly from the number I list in the text, as these results are obtained using a different (older) data reconstruction campaign than the one I use in this thesis.

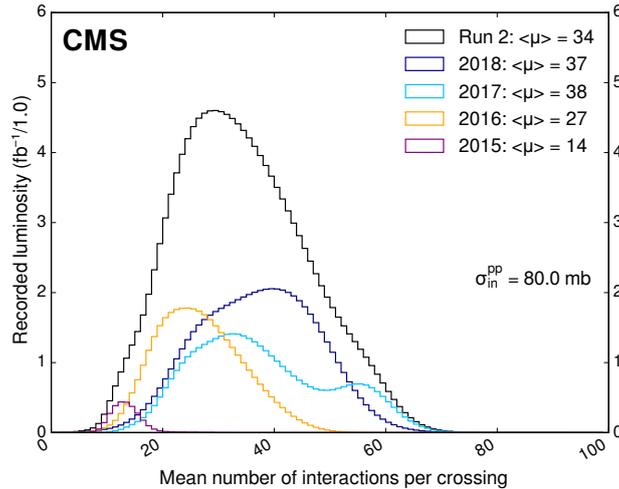


Figure 3.14: PU distribution at CMS during Run 2, shown combined and split by data taking year. Taken from [148].

per bunch crossing are likely, which leads to an overlap of different collision processes in the detector. Assuming some main interaction that is under study, all other overlapping interactions are called *pileup* (PU) [153]. Following equation 3.8, one can see that pileup is expected to scale approximately linearly with instantaneous luminosity. Also, from the cross section overview in section 2.1.4, it is expected that PU almost entirely consists of low energy QCD processes. An overview of the PU distribution in CMS data is given in figure 3.14. There, one can see that the mean number of PU interactions during Run 2 is about 30 to 40, but can reach as high as 60.

3.2.3 Data acquisition

As discussed above, collisions within CMS happen at a rate of about 40 MHz, being detected by the various CMS subsystems. The *data acquisition system* (DAQ) [154, 155] is used to read out the data measured by these. However, it can not read out events at a rate of 40 MHz. Additionally, with an average offline event size of 1 MB, this would correspond to a data stream of about 40 TB s^{-1} , which can not feasibly be transferred or stored. Therefore, a two layer *trigger system* is used at CMS. It aims to lower the rate at which events are read out with as little impact on the physics performance as possible. The cross section overview in figure 2.3 shows that the majority of hard scattering SM process are very rare compared to basic QCD processes, like (elastic or inelastic) scattering or multi-jet production. Additionally, it can be assumed that any potential BSM physics that is within the energy reach of the LHC is even more rare, as otherwise it could have been already detected. Therefore, the goal of the trigger system is to identify and trigger on events from rare, interesting processes while rejecting less relevant, more common background events.

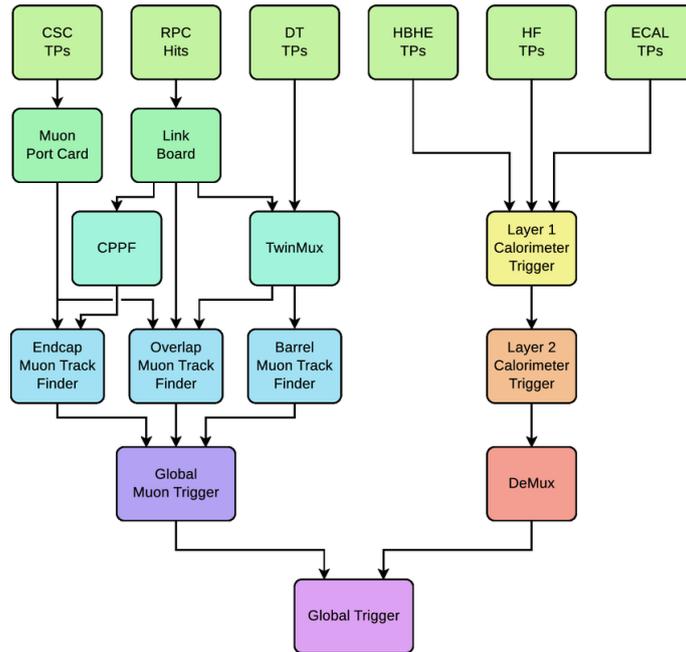


Figure 3.15: Schematic diagram of the CMS L1T in its Run 2 configuration. Data from the different input parts, shown at the top, flows through different parts until it reaches the GT at the bottom of the figure. Taken from [158].

3.2.3.1 Level 1 trigger

The *Level 1 trigger* (L1T) [156] is the first layer of the trigger system, which acts before detector readout through the DAQ system. It works synchronously to the LHC clock of 40 MHz and is fully implemented in hardware, using *field-programmable gate arrays* (FPGA) and *application-specific integrated circuits* (ASIC). Data is held in buffers for up to 4 μ s awaiting readout. A schematic overview of the Run 2 L1T is given in figure 3.15, which is a completely new L1T setup compared to Run 1 [157]. In order to distinguish interesting events from background events, physics objects need to be reconstructed from the raw detector data. For this, information from fast detector parts, using a reduced resolution, is used.

In the CSCs and DTs, so-called *trigger primitives* (TP) are built, which contain information on p_T , coordinates, timing and detector hit quality. Depending on the pseudorapidity region, the TPs and RPC hits are combined to muon tracks by three different track finders: the *barrel muon track finder* (BMTF), *overlap muon track finder* (OMTF) and *endcap muon track finder* (EMTF). The resulting muon candidates contain information on track quality, p_T and charge and are passed to the *global muon trigger* (GMT). Here, duplicates are removed, which might occur from muons passing through overlapping regions of different track finders. After improving the tracks by extrapolating to the nominal interaction point and performing p_T -sorting, muons are passed from the GMT to the *Global Trigger* (GT).

Similar to the muon systems, in the HB, HE, HF and ECAL, TPs are constructed from energy deposits, which are then passed to the first layer of the calorimeter trigger. Here, calibrations accounting for time-dependent changes in calorimeter responses are applied, before the TPs are passed to the second layer, and physics objects are reconstructed. As information from the tracking system is not available in the L1T, photons and electrons can not be distinguished and are reconstructed

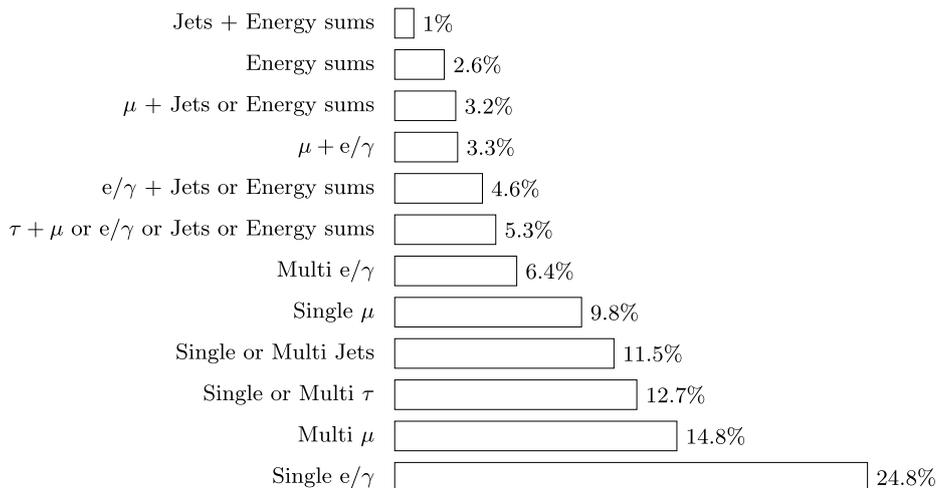


Figure 3.16: Fraction of L1T rate allocated to triggers targeting different L1 objects or combinations of those, taken from [158].

as combined e/γ objects. Clustering energy around a local energy maximum, some quality criteria are applied to low momentum e/γ objects to distinguish them from hadronically induced calorimeter showers. Also, one distinguishes between isolated and non-isolated e/γ objects depending on the surrounding energy. A dedicated reconstruction algorithm is employed to identify the detector signature of τ leptons decaying to pions. Having similar properties as QCD-induced jets, they can be distinguished as they form separate calorimeter clusters. As for e/γ objects, isolation properties are evaluated. Finally, jets are reconstructed from calorimeter clusters using a sliding window algorithm around a local energy maximum. Using the available information, energy sum variables like the total jet momentum sum H_T or missing transverse energy p_T^{miss} can be calculated. For all L1T calorimeter reconstruction algorithms, dedicated measures are employed to minimize the influence of PU. Finally, all calorimeter-based objects are passed to the GT.

I refer to the objects reconstructed in the L1T as L1 objects in the following, to differentiate them from objects used in offline analysis (see section 3.2.4).

The GT contains the so-called L1 *menu*, a collection of different L1 *triggers*. Every trigger in the menu is a combination of one or multiple *trigger conditions*, which result in a single boolean value for that trigger. Typically, trigger conditions are thresholds on p_T , η or quality criteria of a L1 object. The majority of triggers in the L1T are single- or double-object triggers which cover a wide range of physics processes. So-called *cross triggers*, combining information from L1 objects of different types, exist, as well as more complex triggers targeting specific signal signatures. A logical OR is calculated from the results of all triggers, and if it is **true**, a signal is sent to the DAQ system to trigger the detector readout. The output rate of the L1T is therefore limited to the maximum DAQ readout rate of 100 kHz, which is shared by all triggers. A typical setup of triggers during Run 2 is shown in figure 3.16.

In order to include triggers which would have too high rates in the L1 menu, *prescales* are used. A prescaled trigger is only outputting **true** every N th time its conditions are fulfilled, where N is the triggers prescale factor.

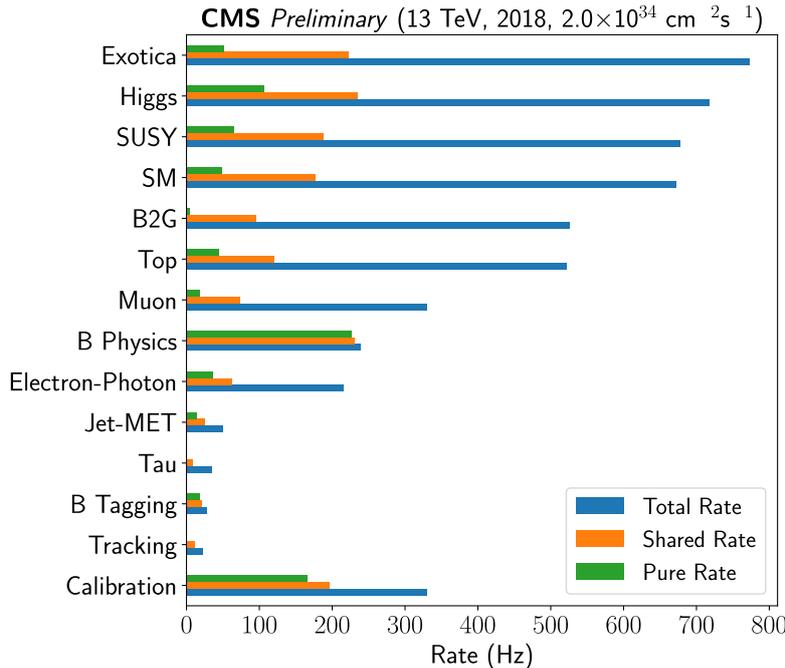


Figure 3.17: Combined rate of triggers belonging to different physics groups in CMS during 2018 data-taking. Listed are physics analysis groups (Exotica, Higgs, SUSY, SM, B2G, Top, B Physics) which usually implement more complex, analysis-specific triggers, and physics object groups (Muon, Electron-Photon, Jet-MET, Tau, B Tagging, Tracking) which implement basic single- and multi-object triggers. Taken from [160].

3.2.3.2 High Level Trigger

At the L1T acceptance rate of about 100 kHz, the DAQ can now read out the triggered events. However, the resulting output stream of about 100 GB s^{-1} of data is still too much to store. Therefore, a second trigger layer is needed. Through DAQ system, the information measured in every detector sub-system is transferred to the *event builder* via optical links, which combines them per collision event and passes them on to the *high level trigger* (HLT) [154, 159]. In contrast to the L1T, this system has access to the full detector information, and is running on a regular commercial server farm. Events are buffered for up to two minutes while the HLT makes a decision whether to pass on or discard them.

Running a software very similar to the offline reconstruction software but optimized for fast execution, the HLT consists of a menu of many different *trigger paths*. Each path consists of a sequence of reconstruction and filtering modules, where the first module always is one or multiple L1 triggers, called the *seeds* of the path. A reconstruction module provides the object(s) required by a following filter module. In the following, I refer to objects reconstructed as part of a HLT path as HLT objects. Modules in a path are executed consecutively, and as soon as a filter condition is not fulfilled, the execution of that path is stopped. Therefore, in contrast to the L1T, there is no fixed latency in the HLT. As long as the average timing is low enough that the storage buffers do not fill up, evaluation can take up to $\mathcal{O}(\text{seconds})$ for some events. The targeted HLT output rate is about 1 kHz. Among the HLT trigger paths are simple single- and double-object triggers, but also many more

complex algorithms targeting specific detector signatures. To compare the relative rate contributions, figure 3.17 shows the rate share of triggers sorted by the CMS physics groups managing the respective triggers.

Multiple triggers might select the same event, thus when adding a new trigger to the menu, the total menu rate will only increase due to events that the new trigger selects that are not selected by any other trigger. This added rate of a single trigger relative to the rest of the menu is called *pure rate*, which is also shown in figure 3.17. Like in the L1T, a logical OR of all HLT paths is determined and the selected events are passed on for long-term storage and offline reconstruction.

3.2.3.3 Trigger rate monitoring

It is crucial to constantly monitor and improve the performance of the trigger system, as all data that is not triggered is lost and can not be recovered. Monitoring the rate of L1 triggers or HLT paths ensures the overall L1 or HLT rate budget is not exceeded and helps to detect possible problems with individual triggers or detector parts. These could become visible in unstable, too high or too low rates.

Trigger rates are monitored while CMS is collecting data using the *Online Monitoring System* (OMS) [161]. This web-based tool provides information about detector status, luminosity and trigger rates, both live and for past data taking periods. Rates are calculated per *luminosity section* (LS), which is a unit of data subdivision and corresponds to about 23 seconds of data taking [162]. Additionally, trigger rates are evaluated for longer periods of data taking, in order to determine changes due to different detector conditions.

Changes in the rate of a specific trigger are most strongly driven by changes in instantaneous luminosity: as the expected number of all processes scales linearly with the instantaneous luminosity (see equation 3.8), so do trigger rates. However, this behavior can be altered by PU. In figure 3.18, I show rates per LS against the average PU in that LS, showing that the slope of the PU-dependent increase varies between trigger algorithms. Usually, muon triggers are close to ideal, whereas triggers based on hadronic activity in the detector (like H_T or E_T^{miss}) show a stronger PU-dependency.

3.2.3.4 Trigger efficiency monitoring

Beyond rates, the efficiency of a trigger is an important measure of its performance, however it is not as easy to measure. In a general sense, the efficiency ϵ of a trigger is defined as

$$\epsilon(\text{studied trigger}) = \frac{N(\text{events passing studied trigger})}{N(\text{all events})}. \quad (3.11)$$

A trigger efficiency gives a measure on how well a given trigger (or combination of triggers) is suited to select events of interest. However, by design, not all events are stored, thus the value of $N(\text{all events})$ is inaccessible. There are multiple ways how this can be circumvented to allow what is called *efficiency measurement* in data, which I introduce in the following. In contrast, in simulated events (which I introduce in section 4.1), efficiencies can simply be calculated by using equation 3.11, as all events are available.

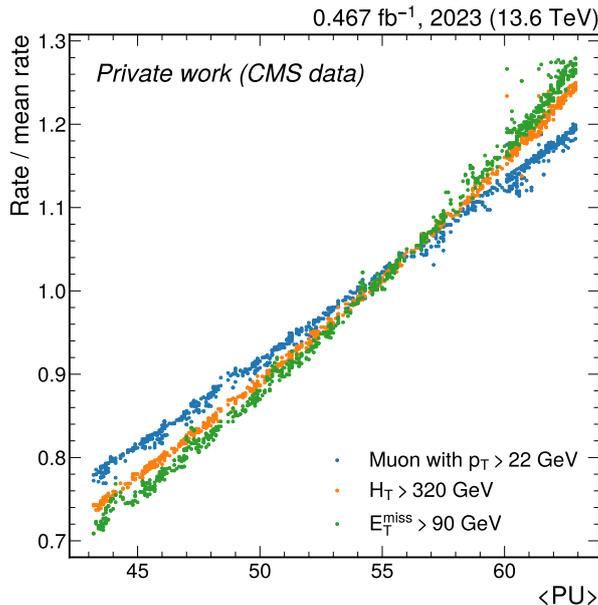


Figure 3.18: L1T rates per LS against the average PU in that LS for data from a single run during 2023 data taking. Rates are normalized with the average rate of each respective trigger to make the PU-dependencies comparable.

Measurement using random triggers CMS takes data using so-called *random triggers*, where the trigger condition is not dependent on any measured quantity, but fully random. As the resulting dataset is unbiased, measuring a triggers efficiency can be done trivially following equation 3.11. However, as the overall efficiency (relative to all collisions in CMS) of any trigger must be low to fit the trigger rate constraints, only few events from such a random dataset will pass the analyzed trigger condition, resulting in very large statistical uncertainties on the efficiency, especially when calculating it differentially, or after applying some other selections to the data. Therefore, this method is usually only used for initial studies or very high rate L1 triggers.

Measurement in an orthogonal dataset To obtain a dataset where enough events pass the studied trigger so that an efficiency can be properly measured, this dataset can be gathered by some other trigger, called the *orthogonal trigger*. Then, the trigger efficiency is defined as

$$\epsilon(\text{studied trigger}) = \frac{N(\text{events passing studied trigger and orthogonal trigger})}{N(\text{events passing orthogonal trigger})} \quad (3.12)$$

Assuming there is no correlation between the orthogonal trigger and the studied trigger, this relation becomes equal to equation 3.11. To investigate this, the parameter α is defined as

$$\alpha = \frac{\epsilon(\text{studied trigger}) * \epsilon(\text{orthogonal trigger})}{\epsilon(\text{studied trigger and orthogonal trigger})}. \quad (3.13)$$

The value $|1 - \alpha|$ describes the strongest possible error on the measured efficiency originating from correlation between the datasets. Thus, a good orthogonal trigger should have α close to one. Often, its value is then assigned as an uncertainty onto the measured efficiency. The value of α can be tested in simulated events.

Measurement using a reference trigger Another method to obtain a dataset where enough events pass the studied trigger uses a so-called *reference trigger*. When using a dataset gathered with a trigger that is a strict super-set of the studied trigger, an efficiency measurement can be performed without any intrinsic bias. Assuming that the efficiency of the reference trigger is known, the overall efficiency can then be obtained as

$$\epsilon(\text{studied trigger}) = \epsilon(\text{reference trigger}) * \epsilon(\text{relative}) \quad (3.14)$$

with

$$\epsilon(\text{relative}) = \frac{N(\text{events passing studied trigger and reference trigger})}{N(\text{events passing reference trigger})}. \quad (3.15)$$

One can easily see that equation 3.11 is recovered if all events pass the reference trigger. The reference trigger method is often used for triggers including multiple filters, where the reference trigger is identical except for the removal of one or more filters. Often, triggers used as reference triggers are prescaled.

Measurement using tag and probe The final method of efficiency measurement uses an inherently different approach compared to the other three: instead of somehow trying to define an unbiased dataset to measure a per-event efficiency in, instead a per-object efficiency is measured. In the *tag and probe* method, events containing two independent objects (for example two muons from a Z decay) are analyzed. Events gathered with a single-object trigger are used, and an offline object of the same type is required to be near the object that the trigger used. Then, a secondary offline object is selected, and the efficiency of the studied trigger to select on this object is measured. For a single-object trigger, the resulting per-object efficiency is identical to a per-event efficiency. For multi-object triggers, efficiencies can be multiplied assuming that the different selection criteria of the trigger factorize.

Trigger efficiency studies Using these measurement techniques, the efficiency of L1T and HLT triggers is constantly monitored. This is often done differentially, for example as a function of offline object p_T or η . Due to differences in online and offline reconstruction, the efficiency of a trigger including a momentum requirement is smeared out near that threshold. The rise in efficiency is referred to as the *turn-on* of the trigger, after which a *plateau* of constant, high efficiency follows. Performance results of the most common L1 triggers can be found in [158], and HLT performance results in [163], where the excellent efficiency of the CMS trigger system is documented.

I perform efficiency studies for two new HLT paths that are added for Run 3 data taking (differences to the Run 2 configuration of CMS are listed in appendix A.1.2). A first path requires a single jet with $p_T > 420$ GeV and $m_{SD} > 30$ GeV. A second path requires two jets, both having $p_T > 270$ GeV and $m_{SD} > 30$ GeV. The jet mass

m_{SD} is defined using the soft drop algorithm [164], as the triggers replace similar ones with an older jet mass calculation method. The jet mass condition lowers the trigger rate and thus allows lower jet momentum thresholds compared to jet triggers without a mass requirement.

I measure efficiencies in a dataset gathered during 2022, corresponding to 34.4 fb^{-1} . To target a set of events relevant to analyses using these HLT paths (for example a search for a hypothetical particle X decaying as $X \rightarrow HH \rightarrow 4b$), I impose some event selection criteria: Only events with at least two large-radius jets with $p_T > 300 \text{ GeV}$ (250 GeV) of the leading (second leading) jet are considered. Both jets are required to have a jet mass $m_{SD} > 70 \text{ GeV}$, and a single muon with $p_T > 30 \text{ GeV}$ is required as well. I define an orthogonal dataset using a single muon trigger, requiring either an isolated muon with $p_T > 24 \text{ GeV}$ or a non-isolated muon with $p_T > 50 \text{ GeV}$. The resulting efficiencies are shown in figure 3.19. I include a single jet trigger without mass requirement, having a p_T threshold of 500 GeV , and a H_T -based trigger requiring 1050 GeV of total hadronic momentum, for reference. One can see that the jet mass triggers help increase the efficiency at low jet p_T and H_T . When combining all tested HLT paths in a logical OR, as usually done in offline analysis, high efficiency over a wide range of momenta is visible. I perform further efficiency checks, splitting the available data in parts before and after an update of HCAL calibrations which happened during 2022. In figures 3.20 and 3.21, I show that the trigger efficiency in the turn-on increases after the HCAL calibration update. These results showcase how trigger efficiency studies can be used to evaluate changes in detector conditions and their effect on the data gathered by CMS. These results are published in [12] in form of a CMS detector performance note.

Software release validation Another application of trigger efficiency measurement is release validation of the *CMS software* (CMSSW). For each new release, simulated events are passed through the CMSSW software stack automatically and HLT efficiencies are compared to the previous release in order to immediately spot potential issues. The implementation of the efficiency calculation is managed by the physics group responsible for a path.

I develop a central HLT release validation tool for usage in Run 3. This additional validation is designed to be a centrally available tool that does not rely on the individual physics groups, is easy to use and configure and produces results that can be investigated in a short time. Therefore, the tool consists of two parts: The first part calculates efficiencies using a generic approach that is applicable to most triggers. To not be dependent on object definitions, instead of using offline object properties, efficiencies are calculated as function of generator-level object properties (which I introduce in section 4.1). These are matched to the object passing a HLT path using a ΔR criterion. Provided by the user in a simple configuration file, many settings are available, including...

- ...the HLT path to evaluate.
- ...the generator-level object type to use.
- ...the variables to evaluate efficiencies against and their binning.
- ...selection requirements, both generic and variable-specific.
- ...the ΔR threshold to match generator-level and trigger objects.

This way, the tool can be used quickly by only selection a few settings. Additionally,

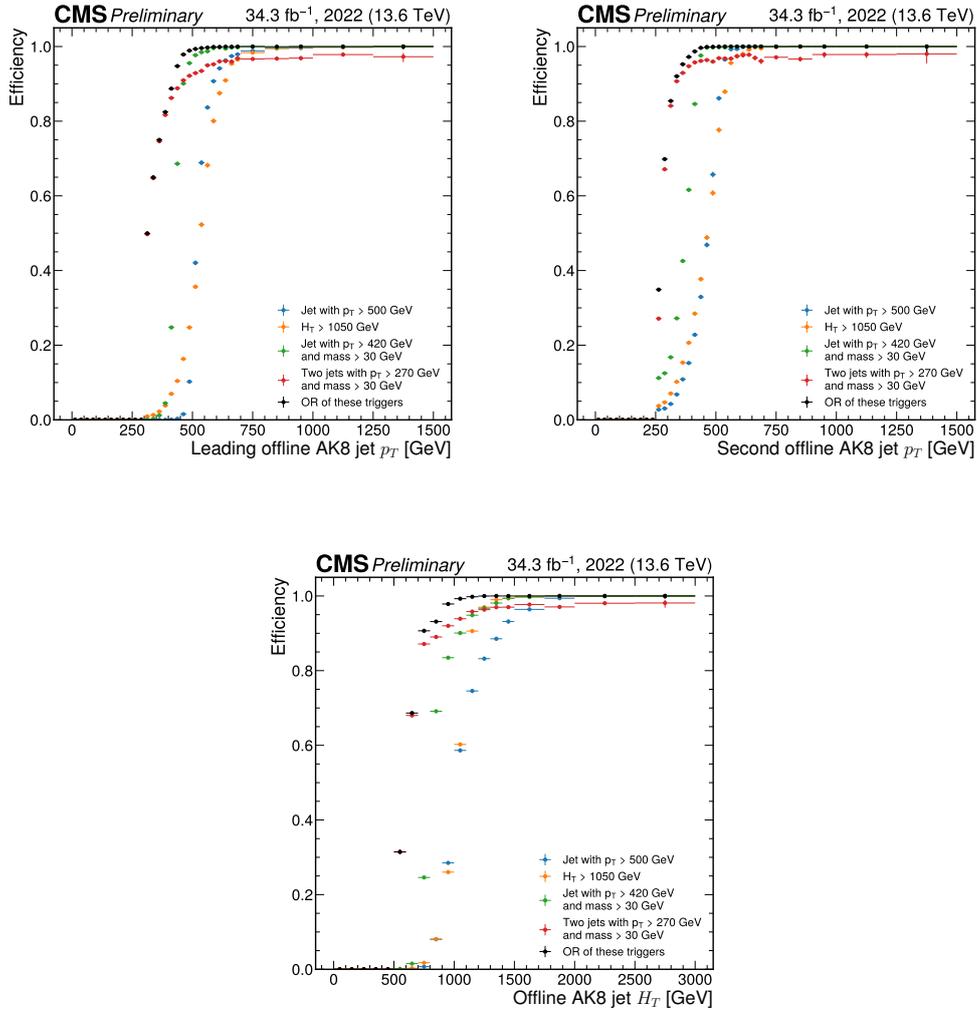


Figure 3.19: Trigger efficiencies as a function of the leading (upper left), second leading (upper right) offline jet p_T and H_T (bottom). Shown are existing jet and H_T triggers requiring a single jet with $p_T > 500$ GeV and $H_T > 1050$ GeV, respectively. Additionally, the efficiency of two new triggers are shown: One requires a single jet with $p_T > 420$ GeV and $m_{SD} > 30$ GeV, the other two jets with $p_T > 270$ GeV and $m_{SD} > 30$ GeV each. These results are published in [12].

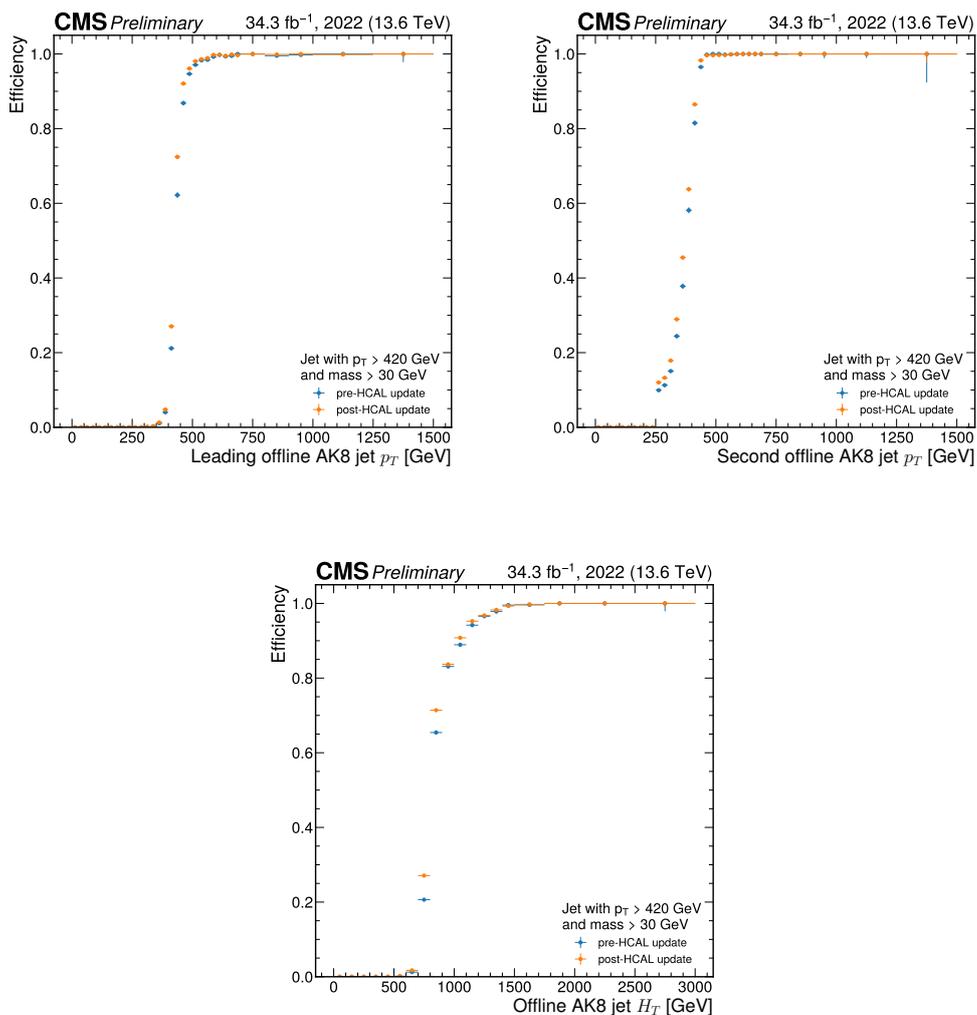


Figure 3.20: Trigger efficiencies as a function of the leading (upper left), second leading (upper right) offline jet p_T and H_T (bottom) for a trigger requiring a single jet with $p_T > 420$ GeV and $m_{SD} > 30$ GeV. Results before and after a change in HCAL calibrations are compared. These results are published in [12].

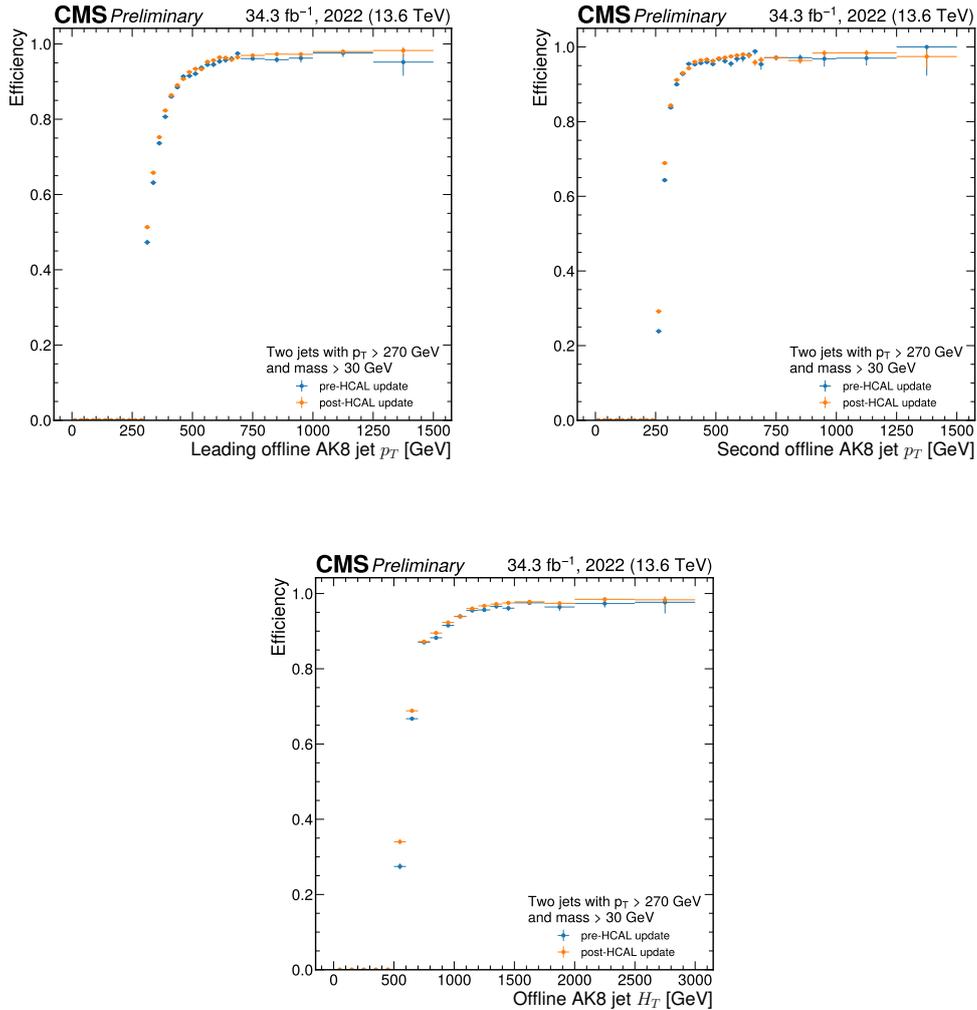


Figure 3.21: Trigger efficiencies as a function of the leading (upper left), second leading (upper right) offline jet p_T and H_T (bottom) for a trigger requiring a two jets with $p_T > 270$ GeV and $m_{SD} > 30$ GeV. Results before and after a change in HCAL calibrations are compared. These results are published in [12].

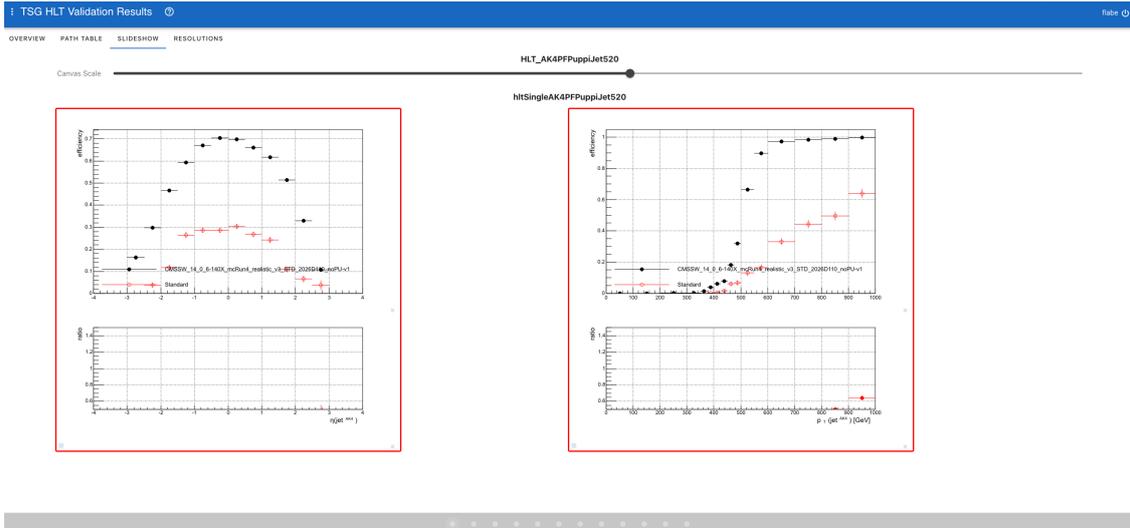


Figure 3.22: Representative screenshot of the web-based trigger validation interface. A slideshow view of trigger efficiencies is provided for all tested triggers, evaluating the efficiency as a function of different variables. When a baseline result is selected, differences to it are automatically detected and marked with a red border.

it can be automatically executed regularly with a default set of parameters. Efficiencies are calculated and the results stored. They can then be easily accessed using a web-based interface, the second part of the validation tool. An example is shown in figure 3.22. The main view option is quick overview of the results, showing only the final filter per HLT path in a slideshow view. Alternatively, the full filter-by-filter efficiencies of all paths can be displayed. In both view configurations, the user can compare validation results to a baseline, any deviations will automatically be marked so that it can be detected quickly. This way, the tool can be used by trigger experts to quickly ensure that no major issues arise in a new software release. If an issue is visible, enough information is available to point to the problematic path, object or filter, so further investigation can be made.

3.2.4 Event reconstruction

To perform offline analyses like the one I present below, physics objects are reconstructed from the raw detector measurements, based on the signature they have in the experiment. This signature varies for different particles, as is visualized in figure 3.23.

3.2.4.1 Particle Flow reconstruction

Most objects interact not only with one part of the detector. Therefore, a *particle flow* (PF) algorithm is used in CMS [166]. With PF, physics objects are reconstructed using all relevant information, combining different detector subsystems.

Basic particle flow elements The PF reconstruction starts with a reconstruction of basic PF *elements*: charged particle *tracks* and *calorimeter clusters*.

Charged particle tracks can be reconstructed using information from the tracking system and the muon chambers. The track reconstruction is an iterative procedure

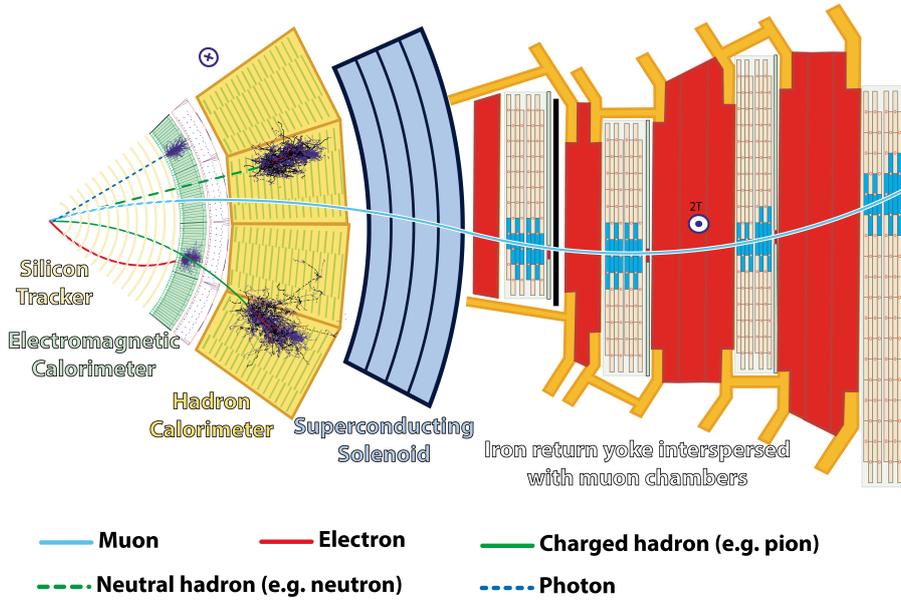


Figure 3.23: Detector signature of a photon, neutral hadron, muon, charged hadron and electron in a slice of the CMS experiment, taken from [165].

based on Kalman filtering [167]. A set of starting seeds is found, and a trajectory is extended from each of them. All hits along the trajectory are gathered and a fit is performed to them to obtain the origin, transverse momentum and direction of the particle inducing the track. All hits now associated to a track are removed and the procedure is repeated for a total of ten times, while gradually lowering the track quality requirements and changing the seed, targeting different types of tracks in each iteration. Seeds are initially based on the pixel detector, then strip information and finally are based on muon system information. From the reconstructed tracks, the primary interaction vertices (PV) of the collision can be determined [132], of which there might be multiple due to PU collisions. Extrapolating tracks to find ones originating from the same point in the detector (and requiring this point to be close to the nominal collision point) provides the information needed to determine the vertex position using a fit. Of these vertices, the one with the highest p_T^2 sum of its decay products is chosen as the main interaction vertex of the collision event. Calorimeter clusters are found separately in the different calorimeter parts. Cluster seeds are found in calorimeter cells with local energy maxima above a certain threshold, from which clusters are formed by combining with surrounding cells that have energy above twice the noise level. A multi-Gaussian fit then provides energy information of the cluster after a calibration process.

The link algorithm In a next step, the *link algorithm* combines different basic elements to create so-called PF *blocks*, which are the basis for the subsequent reconstruction and identification of physics objects. A link between a track and a cluster is created when the extrapolated trajectory of the track lies within the cluster. Multiple tracks can be combined, targeting either electron-photon conversion processes, or tracks originating from common vertices. Similarly, tracking system tracks can be combined with muon system tracks. Finally, multiple clusters can be combined if the cluster position of one cluster lies within the envelope of another. The algo-

rithm can in principle combine every pair of PF elements, however a restriction to the nearest neighbors is made to reduce the computational load.

Physics object reconstruction For each PF block, different identification and reconstruction steps are then executed sequentially to identify physics objects. First, muon candidates are obtained, for which different candidate types can be defined: *standalone muons* are only based on muon system tracks, *tracker muons* only based on tracking system tracks (matching at least one muon system hit) and *global muons* combine information from tracking system and muon chambers to obtain muon candidates. Then, certain quality selection criteria on global and tracker muons, for example depending on their isolation, provide the final collection of muons. After muons are reconstructed, all PF elements that were used to successfully reconstruct a muon are removed from further computation.

Next, electrons and photons are reconstructed. Electron seeds are either purely based on ECAL clusters (accounting for bremsstrahlung effects) or tracks. For cluster-based electron seeds, corresponding tracker hits are found based on the expected trajectory from the clusters energy and position. If no hits are found, the candidate is identified as a photon instead. Conversely, for tracker-based electron seeds, a cluster matching the track is found, for which the measured energy must match the tracks transverse momentum. Repeating the track fit with a Gaussian-sum filter [168] (instead of the Kalman filter used before) improves the results. As for muons, the reconstruction involves certain requirements on the quality and isolation of the considered PF elements, and after processing all elements that were used are removed.

Next, charged and neutral hadrons are identified, as well as non-isolated photons and (rarely) muons from hadron decays which were not identified by the previous algorithms. HCAL clusters not linked to tracks are identified as neutral hadrons, ECAL clusters without tracks are identified as photons, otherwise the result is a charged hadron, depending on the calorimetric energy relative to the associated track momenta. All resulting objects are passed on as so-called PF *particles*.

Finally, some post-processing steps are applied to avoid artificial large p_T^{miss} from various sources, like cosmic muons or misreconstructed objects.

3.2.4.2 Jet reconstruction

While muons, electrons and photons from proton-proton collisions reach the detector and can be measured as they are, this is not the case for hadronic particles. As I introduce in section 2.1.3.2, color-charged particles hadronize and form jets. The PF algorithm can reconstruct the individual objects that make up these jets, but to enable statements on the particle a jet originated from, it is of interest to reconstruct the jet itself as a physics object, instead of using its individual constituents. For this purpose, different jet reconstruction algorithms are used to reconstruct, or *cluster*, jets from the measured PF particles. For these algorithms, there are some crucial requirements: Aside from stability against PU, jet clustering algorithms need to be *collinear safe* such that the result is not affected by gluon splitting processes inside the jet. Similarly, *infrared-safety* is important, the clustering result should not change due to radiation of low energy gluons.

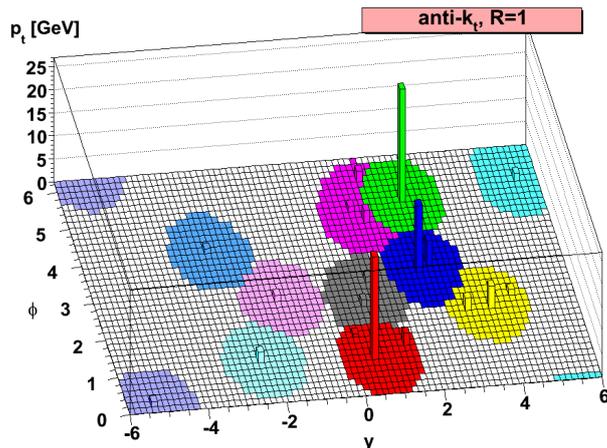


Figure 3.24: A parton-level event with additional lower- p_T background particles clustered with the anti- k_t algorithm. Different colors show different jets, with the shape effects of p_T differences between jets clearly visible. Taken from [169].

The anti- k_t algorithm A widely used jet clustering algorithm is the anti- k_t algorithm [169]. As a *sequential recombination algorithm*, it clusters jets by iteratively combining entities until a jet is reconstructed, in contrast to cone-based algorithms which reconstruct jets as a cone around some seed. A distance between two entities can be defined as

$$d_{ij} = \min(p_{T_i}^{2n}, p_{T_j}^{2n}) \frac{\Delta_{ij}^2}{R^2} \quad (3.16)$$

with $\Delta_{ij}^2 = (y_i + y_j)^2 + (\phi_i + \phi_j)^2$ where y_i is the rapidity (defined in appendix A.1.4, similar to the pseudorapidity η) of entity i . The parameters R and n are free and can be chosen when defining the algorithm. A second distance parameter is defined as

$$d_{iB} = p_{T_i}^{2n} \quad (3.17)$$

and describes the distance between entity i and the beam B . The anti- k_t algorithm is obtained when choosing $n = -1$, whereas other algorithms are obtained for other choices of n : $n = 0$ corresponds to the Cambridge/Aachen algorithm [170] and $n = 1$ to the k_t algorithm [171].

The entities used for the anti- k_t algorithm can vary. In the context of this thesis, PF jets based on PF particles are used, but other jet types exist within CMS. For all entities, all possible d_{ij} and d_{iB} are calculated and the smallest distance is found. If it is a distance between two entities (d_{ij}), they are combined to form a *pseudojet*, which will be considered an entity for the following iterations. If it is an entity-beam distance (d_{iB}) the entity is removed from the process and called a jet if it was a pseudojet. This procedure is repeated until all entities have been removed. A benefit of the anti- k_t algorithm is the shape of the resulting jets. Assuming well separated jets, these are mostly conical in a radius of R around the most energetic constituent. When two highly energetic particles are near each other, two jets will form, which can not both be conical. In that case, the jet with higher energy will be more conical than the other, as shown in figure 3.24. The anti- k_t algorithm

is collinear- and infrared-safe, and not influenced in shape by additional low- p_T particles.

Within CMS, different choices of radius parameter R are common: jets with $R = 0.4$ are called *small-radius* jets, and $R = 0.8$ results in *large-radius* jets. Small-radius jets target the decays of light quarks or gluons, whereas large-radius jets are well suited to reconstruct the decays of, for example, t quarks: Generally, the lifetime of top quarks is so small that they decay before hadronizing into a jet. Their decays almost always result in a b quark and a W boson, as follows from the CKM matrix introduced in section 2.1.3.3. Assuming that the W boson decays into two light quarks and the top quark was highly energetic, the b quark and the two light quarks can appear as a single large-radius jet.

The HOTVR algorithm A jet reconstruction algorithm specifically targeting the decay of heavy objects like t quarks is the *Heavy Object Tagger with Variable R* (HOTVR) [172]. It combines jet clustering with subjet finding and soft cluster rejection. Similar to the anti- k_t algorithm, HOTVR is a sequential recombination algorithm. It uses the same distance definitions I introduced above, with $n = 0$ and replacing the constant parameter R with a p_T -dependent parameter R_{eff} , defined as

$$R_{\text{eff}}(p_T) = \frac{\rho}{p_T} \quad (3.18)$$

with a tunable parameter ρ . For robustness, R_{eff} is limited to be between some R_{min} and R_{max} . HOTVR modifies the clustering procedure with a veto inspired by the *mass jump algorithm* [173]: If two entities are chosen by the algorithm, it is checked whether their invariant mass is greater than some threshold $m_{ij} > \mu$. If this is the case, the mass jump criterion is checked:

$$\theta m_{ij} > \max(m_i, m_j) \quad (3.19)$$

Only if this criterion is fulfilled and the momentum of both entities is larger than another threshold $p_{T_{i,j}} > p_{T_{\text{sub}}}$, the entities are combined. If they were pseudojets, they are stored as subjets of the combined jet. In this way, soft radiation is removed (in what is called *grooming*, making HOTVR collinear- and infrared-safe) and information on jet substructure becomes available. The nominal parameters for HOTVR jets are listed in table 3.1. I refer to jets clustered with this algorithm and these parameters as *variable-radius* jets in the following.

Parameter	Value	Description
R_{min}	0.1	Minimal value of R_{eff}
R_{max}	1.5	Maximal value of R_{eff}
ρ	600 GeV	Slope of R_{eff}
μ	30 GeV	Mass jump threshold
θ	0.7	Mass jump strength
$p_{T_{\text{sub}}}$	30 GeV	Minimum subjet p_T

Table 3.1: The HOTVR parameters used for variable-radius jets, corresponding to HOTVR in default t -tagging mode [172].

3.2.4.3 Jet tagging

When reconstructing jets, determining the flavor of the initial quark, or whether it was a gluon, is useful, and the goal of *jet tagging algorithms*. This is most often done for heavy quarks, and can either be implemented for some existing jet reconstruction algorithm (like anti- k_t) or utilize dedicated algorithms (like HOTVR). In this thesis, I utilize b -tagging and t -tagging and therefore will give an overview on these in the following.

b -tagging They decay of b quark to light up-type quarks is highly suppressed, as can be seen in the CKM matrix in section 2.1.3.3. However, the decay into a t quark is not possible due to the higher mass of the t quark. Therefore, b quarks have a larger lifetime than other quarks. They traverse a non-zero distance before hadronization processes begin, thus tracks from b quark decays will not originate from the primary, but some secondary vertex (SV). This makes it possible to uniquely identify jets originating from b quarks. In this thesis, I use the DeepJet [174] algorithm for this purpose. Like its predecessor DeepCSV [175], it utilizes machine learning (which I introduce in section 4.2). To classify a jet, “DeepJet uses approximately 650 input variables, divided into four categories: global variables, charged PF candidate features, neutral PF candidate features, and SV features associated with the jet” [174]. Global variables include information on the considered jet and the entire event. As the exact number of input variables can vary, an appropriate architecture is chosen: It combines convolutional, recurrent and dense layers to provide information to six different output nodes, allowing DeepJet to act as a multi-classifier performing not only b -tagging, but also c - and quark/gluon-tagging. In the context of this thesis, I use only the b -tagging mode, and refer to the respective output value as DeepJet *score* in the following. Importantly, during the DeepJet training, it is ensured that the network is not influenced by the p_T or η of the tested jet, to avoid sculpting of these variables. The excellent performance of DeepJet, and comparisons to DeepCSV, is shown in [174].

t -tagging t quarks have very short lifetimes due to their high mass. Therefore, they decay via the weak interaction before hadronization. t quarks decaying to a b quark and two light quarks can be reconstructed as a single jet if the t quark is Lorentz boosted, and the resulting decay products are therefore collimated. Such jets can be distinguished from other jets using substructure information: A jet originating from a t quark will have three so-called *prongs* (energy maxima), resulting from the three individual quarks the jet originates from. In contrast, other jets usually only have one (individual quarks and gluons) or two (boosted hadronically decaying W , Z or Higgs bosons) prongs. When analyzing a jet, the N -subjettiness variables τ_N [176] can be used to describe “to what degree it can be regarded as a jet composed of N subjets” [176]. Thus, for example, a t quark jet is expected to have some nonzero τ_3 , but higher τ_2 or τ_1 (as jets with three prongs also have two and one prong). This motivates the usage of N -subjettiness ratios, where the ratio τ_3/τ_2 is sensitive to boosted t quarks.

For jets reconstructed with the HOTVR algorithm, substructure information can be evaluated in form of the N -subjettiness variables and using the subjets defined in the HOTVR reconstruction process. Thus, a t -tag for HOTVR jets is defined by

the following criteria:

- $\frac{p_T(\text{leading subjet})}{p_T(\text{jet})} < 0.8$: jet p_T not exclusively in leading subjet.
- $N_{\text{subjets}} \geq 3$: three prongs expected for t quarks.
- $140 \text{ GeV} < m_j < 220 \text{ GeV}$: jet mass m_j around the known t quark mass.
- $m_{\text{min}} > 50 \text{ GeV}$: smallest pairwise mass of any subjet combination is large.
- $\tau_3/\tau_2 < 0.56$: ratio of N -subjettiness favors a three-prong jet.

Comparisons to other jet tagging algorithms are presented in [172], showing excellent performance of HOTVR over a wide range of jet momenta, thanks to the variable radius used for jet clustering.

3.2.4.4 Pileup removal

As I introduced in section 3.2.2.3, multiple collisions per bunch crossing result in considerable PU. Mostly consisting of low energy QCD events, PU results in additional particles that are clustered into jets, thus inducing a difference between the jet p_T and the actual p_T of the particle inducing the jet. Different techniques are used within CMS to address this issue. In *charged-hadron subtraction* (CHS) [166], vertex information is used to identify all charged hadrons not resulting from the main interaction vertex, removing them from the PF particle list used to reconstruct jets. This can only be done for charged particles. To also consider the PU contributions from neutral particles, an average neutral p_T contribution expected from PU is subtracted. In this analysis, I use a more sophisticated technique of PU suppression: Pileup Per Particle Identification (PUPPI) [177]. This algorithm combines the three different types of information that can be utilized to identify PU: "the event-wide pileup density, vertex information from charged tracks, and the local distribution of pileup with respect to particles from the leading vertex" [177]. PUPPI calculates a weight for each PF particle which is applied to the particles momentum: particles from PU receive low weights, whereas particles from the hard scattering process receive weights close to unity. Charged particle from the main interaction vertex are used as a starting point (and all other charged particles are discarded), relative to which the weight calculation is performed for all neutral particles. This procedure leaves only particles likely to be originating from a hard scattering process, and removes all others. Then, following reconstruction steps (like jet clustering) can be performed without any need for further PU consideration.

Due to the immense amount of data measured with the CMS experiment, sophisticated computing methods are used to process and understand that data. In this chapter, I give an overview on the creation of simulated samples of CMS collision events and the concept of Machine Learning, both of which are crucial parts of the results I subsequently present.

4.1 Monte Carlo simulations

To perform measurements at an experiment like CMS, it is crucial to understand how different physics processes behave in order to develop analyses and to interpret results. Therefore, *Monte Carlo* (MC) simulations of collision events are used in most CMS analyses. The MC method aims to predict a result numerically by random sampling, instead of calculating it analytically. Due to its random nature, it is named after the Monte Carlo casino in Monaco [178]. Below, I describe the different steps required to obtain simulations of proton collision events at CMS, and provide references to commonly used tools in CMS. Samples of simulated collision events are used in all following parts of this thesis, where I provide explicit recipes for each sample.

Matrix element calculation To obtain simulated events of a physics process, first the matrix element of the process needs to be calculated. This can be done using perturbation theory as described in section 2.1.4, yielding results on the cross section of the process, and branching fractions of potential decays. Tools to automate these calculations are available, like MADGRAPH5_aMC@NLO [179, 180] (MADGRAPH in the following) or POWHEG [181–183]. For some targeted combination of initial and final state particles, these tools provide information on all possible intermediate processes and their cross sections. From this information, they then repeatedly randomly sample to obtain a set of collision events. When the initial state particle is a composite particle, like in the case of proton-proton collisions, knowledge of the PDFs is required. Different so-called PDF *sets* describing the proton PDFs exist and can be used in the above mentioned programs, for example NNPDF [129].

Particle decay and hadron showering MADGRAPH and POWHEG only calculate what is called the *hard scattering process*, meaning the perturbative description of the physics process to a certain order. However, beyond that, to obtain a description of the particles we expect to reach the detector, decays, radiation processes, as well as hadronization of charged particles needs to be accounted for. This can be done using PYTHIA8 [184], which can simulate particle decays, hadron showering and radiation processes. For PYTHIA8, different so-called *tunes* exist, describing the configuration of various parameters relevant for the simulated processes. These

can be altered in order to optimize PYTHIA8s description of the simulated processes. An example used in this thesis is the CP5 tune [185].

Detector simulation Finally, with the objects that would reach CMS simulated, the last step is the simulation of the interaction between these particles and detector, and the resulting detector response. This is done using the GEANT4 [186] toolbox, a program designed to simulate interactions between elementary particles and matter. The resulting events can then be used in the same way as data events, and are passed through the CMS reconstruction software. The only difference to data events is the availability of the information about all particles that existed during simulation, from the hard scattering process to the hadronization products. I refer to these as *generator-level* particles.

4.2 Machine Learning

In the context of particle physics, *machine learning* (ML) approaches have been used since the 1990s, and have substantially increased in popularity since the 2010s [187]. For example being used in particle reconstruction and identification, as well as event classification, they play a crucial role in today's particle physics research. As I put a strong focus on ML techniques in this thesis, I give an introduction to ML in general and present some concrete examples, following [188]. Unless otherwise noted, all ML applications in this thesis are implemented using the `keras` [189] API for the `tensorflow` [190] backend.

In general, ML describes the development of algorithms that can learn to perform a given task without explicitly being programmed for it. This involves two separate steps: the *training* of the algorithm, where it learns to extrapolate a generalized behavior from some training data, and the *inference* where the algorithm is applied to data to perform its learned task. A distinction can be made between *supervised* and *unsupervised* training of ML algorithms (also called *models*). For a supervised training, the model is trained on some input data and the corresponding desired output values (so-called *labels*). Hereby, the model learns a high-dimensional function to map the input data onto the output values [191]. Supervised training is often used in particle physics application, for example in object or event classification, or regression tasks. In contrast, an unsupervised training does not use any labels: the model learns to identify patterns in the training data without outside guidance. There exist approaches in between these two extremes, for example *weak supervision*, where data is partially or imprecisely labeled.

4.2.1 Introduction to neural networks

Neural networks (NNs) are among the most popular ML algorithms. Deriving their name from *neurons*, the cells that make up the nervous systems of humans and animals [192], a NN is build from a number of *nodes* and connections between these. An example for a NN structure is given in figure 4.1, consisting of three layers: an input layer, receiving the data during training or inference, a single *hidden layer* which is not connected to the outside, and an output layer, from which the networks result (or *prediction*) can be obtained. This kind of NN is called a *fully-connected*

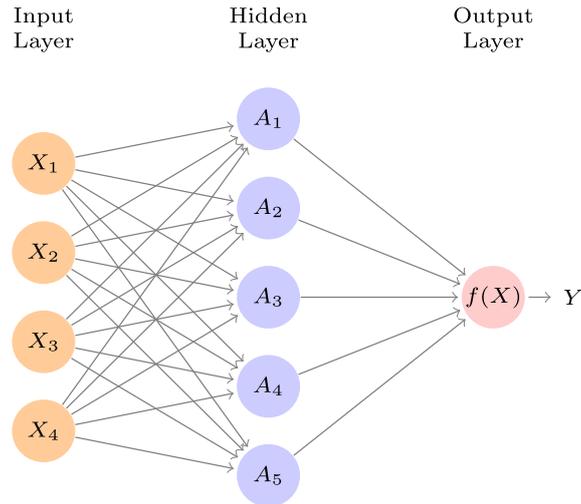


Figure 4.1: Example of a NN structure, showing nodes ordered in three layers, and the connections between them. Taken from [188].

network (FCN), and FCNs with more than one hidden layer are called *deep neural networks* (DNNs).

Describing NNs mathematically, “a neural network takes an input vector of p variables $X = (X_1, X_2, \dots, X_p)$ and builds a nonlinear function $f(X)$ to predict the response Y ” [188]. This function $f(X)$ is constructed from *activations* A_k at each hidden layer node, which are “nonlinear transformations of linear combinations of the inputs X_1, X_2, \dots, X_p ” [188]. Assuming an arbitrary number of hidden layers, they can be written as

$$A_k^i = g(w_{k0}^i + \sum_{j=1}^p w_{kj}^i A_j^{i-1}) \quad (4.1)$$

where A_k^i is the activation of node k in hidden layer i , g is some nonlinear function and w are freely tunable parameters (often called *weights*). In the case of the first hidden layer, A_j^{i-1} is replaced by the inputs X_j . The full model function $f(X)$ is then constructed similarly in the last layer, where an activation function can be used, but is optional and for example not used in regression tasks. In general, nonlinear activations functions are required, as any number of layers could otherwise be recombined into a single linear combination of the input variables. Additionally, these allow the model to learn complex, nonlinear relations in the data. It is possible to have multiple output nodes with multiple output functions $f_n(X)$.

4.2.2 Activation functions

A widely used activation function for hidden layers is the *rectified linear unit* (short ReLU) function. Defined as a piece-wise combination of a constant function and a linear function, it can be written as

$$g(z) = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{otherwise.} \end{cases} \quad (4.2)$$

It can be extended to the so-called **LeakyReLU** activation by adding a slope towards negative z values. Another option is the **sigmoid** function, defined as

$$g(z) = \frac{e^z}{1 + e^z}. \quad (4.3)$$

As it converts any input value in a probability between 0 and 1, it is often selected for the output node of a classifier NN with a single output score. Similarly, the `softmax` function can be used in the case of multiple output nodes, ensuring that the output values sum up to 1 for all data samples. It is defined as

$$g_n(z) = \frac{e^{z_n}}{\sum_{\ell} e^{z_{\ell}}} \quad (4.4)$$

where ℓ runs over all n output nodes. The last example for an activation function is the `tanh` function, simply defined using the hyperbolic tangent as

$$g(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \quad (4.5)$$

4.2.3 Loss functions

To train a supervised NN, first a metric parametrizing its performance is needed: the *loss*. This can be any function that takes the NN predictions Y_i^{pred} and labels Y_i^{true} of a set of n data points as arguments, and returns a number that becomes smaller the better the NN completes the desired task. In this thesis, I use NNs for event classification, and thus refer to these data points as events in the following. A basic example is *mean squared error* (MSE) loss. Also used in classical function fitting applications, it is defined as

$$L_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (Y_i^{\text{pred}} - Y_i^{\text{true}})^2. \quad (4.6)$$

MSE describes the mean deviation of the NNs prediction from its desired output, thus combining the errors from multiple events linearly. Another approach, the *binary cross-entropy* (BCE) [193] loss, also sometimes called *log loss*, is defined as

$$L_{\text{BCE}} = -\frac{1}{n} \sum_{i=1}^n Y_i^{\text{true}} \log(Y_i^{\text{pred}}) + (1 - Y_i^{\text{true}}) \log(1 - Y_i^{\text{pred}}). \quad (4.7)$$

In contrast to the MSE, BCE imposes some requirements onto Y_i^{true} and Y_i^{pred} . The labels are required to be either 0 or 1, thus the term *binary* in BCEs name. A variant for NNs with multiple output nodes exists, the *categorical cross-entropy*, where a single label can be 1 and all others are required to be 0. Accordingly, the predicted values need to be probabilities and are thus restricted to values between 0 and 1. These requirements originate from the construction of this loss function based on the *cross-entropy*, which is explained in appendix A.1.5.

4.2.4 Optimizers and backpropagation

Adjusting the weights w^i within the NN so that the loss is minimal is called the training of the NN. This can not be done analytically due to the high complexity of a NN. Instead, a statistical procedure is used, using algorithms called *optimizers*. The most basic example is *gradient descent* (GD). The NN is initialized with an arbitrary, usually random value for all weights and the loss is calculated for all

available events. From this point on the so-called *landscape* of potential loss values, a step that moves towards a minimum needs to be found. The vector of partial derivatives for all weights of the loss function is calculated, which gives the direction of steepest increase:

$$\Delta W = \rho \left. \frac{\partial L(W)}{\partial W} \right|_{W=W^m}. \quad (4.8)$$

Here, W^m is a vector of all weights w^i at step m and $L(W)$ the weight-dependent loss function. Assuming a somewhat smooth loss landscape, adjusting the NN weights by $-\Delta W$ likely results in decreasing loss. From this process, which is known as *backpropagation*, one obtains a change for each weight, scaled by the *learning rate* ρ , which determines the step size of the optimization procedure. The NN is evaluated with updated weights, and the procedure is repeated until a loss minimum is reached. GD can be adapted by not performing the model parameter update after evaluating the entire training data, but for every single training event, a process known as *stochastic gradient descent* (SGD). While this can be substantially faster than regular gradient descent, it may lead to instabilities in training. To achieve a balance between training acceleration and stability, random fractions, so-called *batches*, of the available data can be evaluated at the same time, which is called *mini-batch gradient descent*. When training a NN this way, training duration is measured in *epochs*, where a single epoch means that the NN has seen the equivalent of the full training dataset once.

All GD variants can have issues that might prevent the NN from converging into the desired global loss minimum. Due to fluctuations during training, the loss function can oscillate in different directions instead of moving straight towards the minimum, slowing down the training. To address this issue, *momentum* can be used in network training. By adding a term that includes the weight change from the previous weight update ΔW^{m-1} , the definition of the weight change is adapted to

$$\Delta W = \rho \left. \frac{\partial L(W)}{\partial W} \right|_{W=W^m} + \gamma \Delta W^{m-1}. \quad (4.9)$$

Like the name suggests, in analogy to a moving massive particle, considering momentum prevents abrupt small changes of the NNs trajectory in the loss landscape, as any weight change will also consider the previous change.

If ρ is the same for all trainable parameters, this can also cause issues. Some choices of ρ can be too large for the late steps of the training, preventing the NN from moving into a tight global minimum. However, lowering ρ can in turn make the training very slow, as large step sizes might be needed in the initial parts of the training to approach the general area of the global minimum. Also, the network could get caught in local minima instead of approaching the global minimum. This motivates changing the learning rate over time, which several optimizers do: using *Adagrad* [194], the learning rate is different for each parameter and training step. Its successor *Adadelta* [195] fixes a problem of Adagrad, which resulted in the training stopping due to constantly decreasing learning rate. *RMSprop* [196] penalizes features with very high impact on the change of the loss, so that other features can also influence the network, leading to faster and more stable training, but with fixed learning rate. An optimizer combining the advantages of Adagrad and RMSprop is *Adam* [197]. It utilizes “exponential moving averages of the gradient [...] and the squared gradient”

[197], thus taking into account the mean \hat{m}_t and variance \hat{v}_t of the gradient. The weight updates then are proportional to their ratio, as

$$\Delta W \propto \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t}} \quad (4.10)$$

with a stepsize setting α . This parameter can be understood as defining a *trust region* in which the loss landscape around the current point can be understood with the information provided from the local gradient, and thus needs to be chosen appropriately for the given problem. As towards a minimum, the ratio $\frac{\hat{m}_t}{\sqrt{\hat{v}_t}}$ gets smaller, the desired behavior of an adaptive learning rate is recovered. Adam is widely used as a state of the art optimizer, and outperforms the other optimizers mentioned here [197].

4.2.5 Types of neural networks

The above introduction of NNs is based on a specific architecture, the FCN. However, there are many other types of NNs relevant to high energy particle physics [187]:

Convolutional neural networks (CNN) are primarily used for classification of images and similarly structured data. Instead of working with a one-dimensional vector of input features per event, multi-dimensional inputs (like an array of pixels of an image) are evaluated. This is achieved using two specialized hidden layer types: *convolutional* and *pooling* layers. Made from different filters that combine information from (in the example of a picture) $N \times M$ pixels, and are moved (*convolved*) over the entire input picture, these layers highlight specific features in the input data. Combined with pooling layers to reduce the size of multidimensional data, convolutional networks are powerful tools, which are used to, for example, classify jets using calorimeter images.

Graph neural networks (GNN) [198] describe the input data as mathematical graphs: a set of points in a multidimensional euclidean space, with connections between different points. This representation is very natural for particle physics applications: an example are constituents of a jet, and relations between them, that can be described as a graph. This allows NNs to move away from fixed number of input values, which otherwise might limit their design. Using, for example, message-passing approaches [199], various NNs based on graphs can be built.

Recurrent neural networks (RNN) [200] or *recurrent layers* deviate from the strict sequential alignment of layers and allow for a feedback loop. In *recurrent units*, a hidden state is kept to allow for information to be retained over multiple steps.

Another set of NN architectures are *autoencoders*, specialized on unlabeled data. Input data is passed to the NN, and a set of layers is used to compress the data to a point in a *latent space* representation. From there, another set of hidden layers is used to reconstruct the initial format. The NN is trained to perform this compression and reconstruction as accurately as possible. If, during inference it is confronted with data very different to the training data, it will not be able to reconstruct it well, resulting in high loss. By selecting data with high loss, anomalies can be found. A variant is the *variational autoencoder* (VAE) [201], which compresses the data to a distribution instead of a point in the latent space.

4.2.6 Performance evaluation

After a NN has been trained, it is important to evaluate its performance. Evaluations based on the loss function provide a good measure of the change in performance of a specific algorithm during training, but are not comparable between different algorithms, especially should these use different loss functions or NN architectures. Therefore, other ways of comparing NNs need to be utilized. Assuming a classification NN, one of the most simple metrics for performance evaluation is the prediction *accuracy*. The accuracy is defined as the fraction of correctly predicted events when rounding the NN output to the nearest label. Thus, a fully random classifier will have an accuracy of 50% (assuming two classes of equal sizes), whereas a perfect classifier will achieve 100%, with real ones being somewhere in between. Accuracy is often evaluated during training, alongside the loss, and can provide additional information on how the training process is progressing. However, there are disadvantages to this metric: by simply rounding the results to the nearest label, which in the case of a binary network is equal to splitting the events at a NN score threshold of 0.5, an assumption is made. Mostly, when using a NN, the requirement on the output score will be different from 0.5, so the accuracy might contain an intrinsic bias. Such a bias can also come from the equal treatment of all classes of input events, which might not be in line with the actual target performance of the NN. Generalizing the concept of accuracy, assuming a binary classifier distinguishing a *signal* from some *background*, the *true positive rate* (TPR) and *false positive rate* (FPR) can be defined as the fraction of events from the signal (for the TPR) or background (for the FPR) class that are identified as signal when assuming a certain threshold. This approach avoids the issue of fixed thresholds, and allows investigation of signal and background separately. It can also be defined for multi-class NNs by using one of the output nodes. Evaluating the TPR and FPR as a function of thresholds between the lowest and highest possible threshold, this can be displayed as a so-called *receiver operating characteristic* (ROC) curve. Investigating ROC curves can provide useful information on the absolute performance of a NN, and this metric also allows relative comparisons of different networks. A derived measure of NN performance can be calculated, the *area under curve* (AUC), defined as the integral under the ROC curve. Higher AUC scores signal better NN performance. However, it is important to note that two NNs can have the same AUC despite performing differently. Therefore, when comparing networks, the relevant region of the ROC curve needs to be defined. Often, the metric of a TPR at a certain FPR (TPR@FPR) is used, comparing networks based on the assumption that they should achieve the highest possible TPR at a specific fixed FPR.

4.2.7 Training stability

The goal of a ML-based algorithm is to learn a generalized behavior, not specific details from the training dataset. However, it is possible that during training, the performance of a NN on the training dataset improves (and thus the loss decreases), but the performance on any other dataset gets worse. This behavior, called *over-training*, needs to be detected and avoided.

To detect overtraining, the total amount of data available for training is often split into multiple parts: a *training set*, a *validation set* and a *test set*. The training set is used to train the NN, as described above. After each training epoch, the loss (and

any other metrics monitored during training) are calculated for the validation set. Hereby, overtraining can be detected as soon as it arises during training, as it would result in the validation loss increasing. Finally, after the network is fully trained, a performance evaluation is performed on the test set. By testing on this third part of the data, which is completely independent from the training procedure, it can be ensured that the NN has not been biased towards the training or validation set.

Other approaches to divide the available data exist. In *k-fold cross validation*, the data is randomly divided into k parts of equal size. Then, k different NN trainings are performed, each time choosing a different one of the parts as test set, and training on the remaining $k - 1$ parts (where training refers to using this data, with some splitting, as training and validation set). This technique can be used when there is overlap between the data available for training and the data that the NN is later evaluated on, so that for each event, there exists a NN not biased towards it. Additionally, this technique can be used to evaluate the stability of a NN.

Overtraining can occur due to various factors: the network being too large (too many layers or nodes per layer) compared to the difficulty of the problem it is trained to solve, the training being too long, the learning rate being too high, and many more. Thus, it can be mitigated by choosing an appropriate NN architecture.

In addition to the basic NN components introduced above, additional layers and techniques can be used to reduce overtraining, stabilize and improve the NN training. If overtraining remains present even after selecting an appropriate architecture and training setup, *dropout* [202] can be used to further suppress it. A dropout layer, which can be inserted at any point in the network, sets a fraction of randomly chosen layer outputs to zero before passing them on. The effect on the training can be compared to the blurring of a picture, where small details are less visible as a result. Hereby, the network will be less likely to learn specific features of the training events, thus performance in learning more general relations improving and overtraining being reduced.

Still, training a NN for too many epochs can lead to overtraining (and even if not, unnecessarily use computing resources). Therefore, it is important to stop the NN training once the loss minimum has been reached. Using so-called *early stopping*, the NN performance (based on some metric, usually the loss), is evaluated after each epoch. If the performance has not improved for N epochs, where N is called *patience*, the training is automatically stopped.

When the inputs of a NN are of strongly varying sizes (for example angular variables of $\mathcal{O}(1)$ and momenta of $\mathcal{O}(50 - 1000)$), this can lead to an unstable NN training, as usually weights are expected to be of $\mathcal{O}(1)$. Thus, input variables often are normalized to a mean of 1 with a standard deviation of 0. Still, internal weights could systematically increase or decrease in size. To avoid this, and thus stabilize and speed up the training, *batch normalization* [203] can be used. Similar to the NN inputs, the inputs to each other layer are normalized for each batch.

In this chapter, I present the main topic of this thesis: a search for a potential signature of new physics, the existence of a heavy new particle labeled t^* . This partner particle of the t quark is well motivated by theory, as I detail in section 5.1. I summarize past searches for this particle in section 5.2, and then present my own analysis in section 5.3. Specifically, I perform a search for pair production of t^* decaying as $t^*\bar{t}^* \rightarrow t\bar{t}g$ in a single-lepton final state. I present the results of this work in section 5.4, where I also discuss future prospects.

5.1 Theoretical background of t^*

In section 2.2, I presented an overview of various BSM physics models. Among these, many predict the existence of new particles, which can be searched for at collider experiments to strengthen the believe in or disprove a model. From the perspective of experimental collider physics, a new particle is primarily characterized by its properties and interaction with SM particles, and less so by the underlying theory model.

A hypothetical new particle could have properties similar to the SM t quark, except for different mass or spin [204]. Such a t quark partner would provide a natural solution to the hierarchy problem. As I introduced above, this problem originates from the contribution of SM quark loops to the Higgs boson mass, dominated by the t quark. Adding a new, fermionic t quark partner T to the SM particle content would result in additional loop contributions, which are shown in figure 5.1. The additional contributions to m_H resulting from these t partner loops can cancel out the contribution from the SM t quark, thus avoiding fine tuning [80].

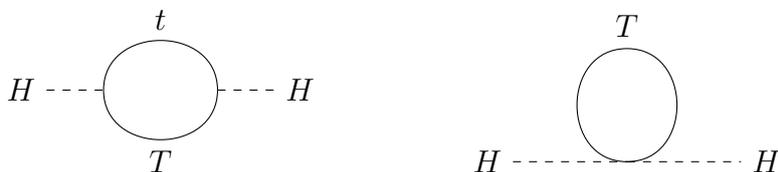


Figure 5.1: Loop diagrams including a fermionic t quark partner T contributing to the Higgs boson mass. Figure adapted from [80].

Importantly, the existence of fermionic t quark partners is not restricted to a single BSM model, they can arise in a wide range of theories. Among them are composite quark models, where a t quark partner could appear as an excited state of a non-fundamental t quark [91–98]. An excited t quark t^* would not be stable, but instead decay quickly into a t quark by radiating excess energy in form of, for example,

a photon or gluon: $t^* \rightarrow t\gamma$ or $t^* \rightarrow tg$. In little Higgs models [100–103], at least one new “electroweak singlet quark with charge $2/3$ ” [100] arises, similarly in composite Higgs models fermions could appear in a new gauge-fermion system [99]. Models with extra dimensions [106, 107, 110] can also predict fermionic top quark partners, for example in Randall-Sundrum models where “such resonance is the spin- $3/2$ excitation” [107] of the right-handed SM top quark.

Usually, searches for these generic fermionic t quark partners are performed targeting decays including W , Z or Higgs bosons. However, in case the mixing probability of t partner and SM t is small, these decay modes are suppressed and the decays $t^* \rightarrow tg$ and $t^* \rightarrow t\gamma$ become dominant [204, 205].

This plethora of theoretical models motivates the search for a particle defined by its decays $t^* \rightarrow tg$ and $t^* \rightarrow t\gamma$, with the experimental signature being independent of which theory model predicts it. I label this particle generically as t^* in the following. A t^* could exist in two different spin scenarios: spin- $\frac{1}{2}$ t^* or spin- $\frac{3}{2}$ t^* [204]. In order to be agnostic to the underlying theory, I use an *effective field theory* (EFT) approach to describe the addition of the t^* to the SM Lagrangian. Based on [204] and [108], I outline it in the following.

5.1.1 Mathematical description: spin- $\frac{1}{2}$ scenario

A mathematical description of a spin- $\frac{1}{2}$ t^* can be constructed very similarly to the one of the SM t quark: it is a singlet under $SU(2)_L$ and a triplet under $SU(3)_C$, thus it can interact strongly. Furthermore, it has a weak hypercharge Y of $\frac{2}{3}$. These properties can be described by a Lagrangian of the form

$$\mathcal{L} = \underbrace{i\bar{\psi}_{t^*(\frac{1}{2})}\gamma^\mu\partial_\mu\psi_{t^*(\frac{1}{2})}}_{\text{kinetic}} - \underbrace{\frac{g'}{2}\bar{\psi}_{t^*(\frac{1}{2})}\gamma^\mu Y B_\mu\psi_{t^*(\frac{1}{2})}}_{U(1)_Y \text{ interaction}} - \underbrace{g_s\bar{\psi}_{t^*(\frac{1}{2})}\gamma^\mu T^\alpha G_\mu^\alpha\psi_{t^*(\frac{1}{2})}}_{SU(3)_C \text{ interaction}} \quad (5.1)$$

where kinetic and interaction terms of the field $\psi_{t^*(\frac{1}{2})}$ are visible. In addition to a SM-like mass term and coupling terms to the Higgs boson, mixing terms with the SM t quark exist. If this mixing is negligible, after electroweak symmetry breaking, the resulting Lagrangian takes the form

$$\mathcal{L}_{\text{EFT}} = \underbrace{c_g\bar{\psi}_{t^*(\frac{1}{2})}\sigma^{\mu\nu}T^\alpha\psi_{t_R}G_{\mu\nu}^\alpha}_{t^*\rightarrow tg \text{ decay}} + \underbrace{c_\gamma\bar{\psi}_{t^*(\frac{1}{2})}\sigma^{\mu\nu}\psi_{t_R}A_{\mu\nu}}_{t^*\rightarrow t\gamma \text{ decay}} + \underbrace{c_Z\bar{\psi}_{t^*(\frac{1}{2})}\sigma^{\mu\nu}\psi_{t_R}Z_{\mu\nu}}_{t^*\rightarrow tZ \text{ decay}} \quad (5.2)$$

where the coefficients c_g , c_γ and c_Z describe the strengths of the different possible decay modes. This way, the interactions are simply parametrized by their strengths, without requiring a specific description of the interaction. Note that the negligible mixing is an assumption made for this analysis. It is driven by the results of many searches assuming mixing is allowed, in which no sign of t^* particles was found [206–212].

5.1.2 Mathematical description: spin- $\frac{3}{2}$ scenario

To add a spin- $\frac{3}{2}$ t^* to the SM Lagrangian, a different construction is needed. A spin- $\frac{3}{2}$ particle is represented by a Dirac spinor with a Lorentz index ($\psi_{t^*(\frac{3}{2})}^\mu$) in a

Rarita-Schwinger Lagrangian [213]. Following [108], where more detail can be found, for a t^* it can be constructed as

$$\mathcal{L} = \underbrace{\bar{\psi}_{t^*}^{\mu} \left(\frac{3}{2}\right) \gamma_{\mu\nu\rho} \partial^{\nu} \psi_{t^*}^{\rho} \left(\frac{3}{2}\right)}_{\text{kinetic}} - \underbrace{\frac{g'}{2} \bar{\psi}_{t^*}^{\mu} \left(\frac{3}{2}\right) \gamma_{\mu\nu\rho} Y B^{\nu} \psi_{t^*}^{\rho} \left(\frac{3}{2}\right)}_{U(1)_Y \text{ interaction}} - \underbrace{g_s \bar{\psi}_{t^*}^{\mu} \left(\frac{3}{2}\right) \gamma_{\mu\nu\rho} T^{\alpha} G^{\alpha,\nu} \psi_{t^*}^{\rho} \left(\frac{3}{2}\right)}_{SU(3)_C \text{ interaction}}. \quad (5.3)$$

As for the spin- $\frac{1}{2}$ t^* , mass terms and Higgs boson interactions are SM-like. Crucially, the different spin does not allow for mixing with the SM t quark. Therefore, the interactions can again be written as an EFT in a similar form as above:

$$\begin{aligned} \mathcal{L}_{\text{EFT}} = & \underbrace{i \frac{c_g}{\Lambda} \bar{\psi}_{t^*}^{\mu} \left(\frac{3}{2}\right) (g_{\mu\nu} + z \gamma_{\mu} \gamma_{\nu}) \gamma_{\rho} T^{\alpha} \psi_{t_R} G^{\alpha,\nu\rho}}_{t g \text{ decay}} \\ & + \underbrace{i \frac{c_{\gamma}}{\Lambda} \bar{\psi}_{t^*}^{\mu} \left(\frac{3}{2}\right) (g_{\mu\nu} + z \gamma_{\mu} \gamma_{\nu}) \gamma_{\rho} \psi_{t_R} A^{\nu\rho}}_{t \gamma \text{ decay}} \\ & + \underbrace{i \frac{c_Z}{\Lambda} \bar{\psi}_{t^*}^{\mu} \left(\frac{3}{2}\right) (g_{\mu\nu} + z \gamma_{\mu} \gamma_{\nu}) \gamma_{\rho} \psi_{t_R} Z^{\nu\rho}}_{t Z \text{ decay}}. \end{aligned} \quad (5.4)$$

Here, the same coefficients c_g , c_{γ} and c_Z that are used in the spin- $\frac{1}{2}$ scenario appear.

5.1.3 Production of t^* at proton-proton colliders

At a proton-proton collider, $t^* \bar{t}^*$ pair production is expected to be dominant over single t^* production, as pair production is driven by the strong interaction, while single t^* production happens via the weak interaction [107]. In figure 5.2, I show LO pair production Feynman diagrams.

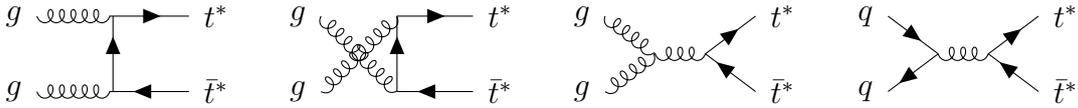


Figure 5.2: LO Feynman diagrams for the $t^* \bar{t}^*$ pair production, adapted from [107].

The calculation of the expected cross section depends on the considered spin scenario. For a spin- $\frac{1}{2}$ t^* , it can be calculated analogous to the SM $t \bar{t}$ pair production (just with different particle masses) from the gluon coupling part of the Lagrangian. In contrast, for spin- $\frac{3}{2}$ t^* pair production, production cross sections can be different, as follows from the different gluon coupling terms shown in equation 5.3. Estimations at $\sqrt{s_{pp}}$ of 14 TeV predict a higher cross section for the production of spin- $\frac{3}{2}$ excited top quarks compared to their spin- $\frac{1}{2}$ counterparts [204]. To obtain cross section predictions for the $\sqrt{s_{pp}} = 13$ TeV data I analyze in this thesis, I calculate cross sections using simulated signal samples, which I introduce in section 5.3.2. The results, compared to the 14 TeV cross section, are shown in figure 5.3. As a sanity check, I also show that the predicted cross section curve used in the previous CMS search is compatible with my calculations.

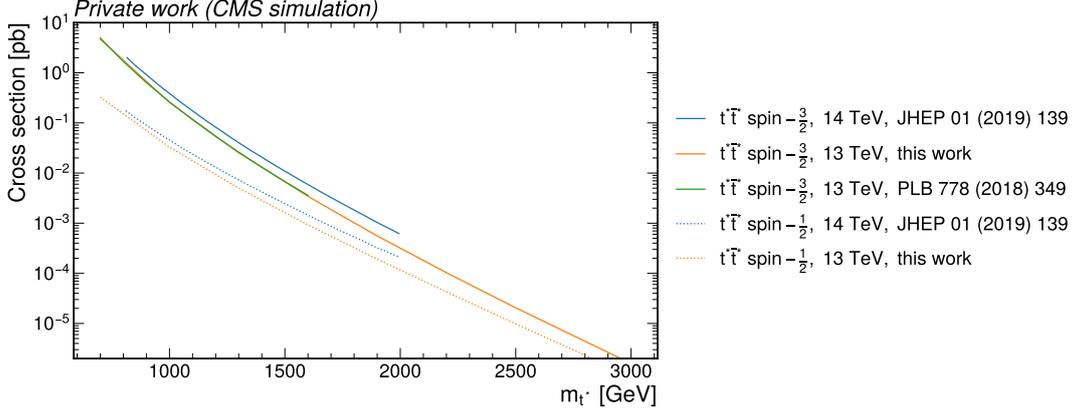


Figure 5.3: Production cross sections of spin- $\frac{1}{2}$ and spin- $\frac{3}{2}$ t^* pairs at LO, comparing the 14 TeV results from [204] to 13 TeV simulations. These are evaluated assuming an exclusive $t^*\bar{t}^* \rightarrow t\bar{t}g$ decay. Also shown is the expected cross section line from the results of the previous search for t^* pairs [214] to ensure comparability.

5.1.4 Decay channels of t^*

As mentioned above, I characterize the t^* by its decay modes. These are driven by the couplings to the SM t quark, as is visible in equations 5.2 and 5.4. The coefficients c_g , c_γ and c_Z parametrize the decays. They are the same for both spin scenarios, and given as

$$c_g = \frac{c_3 g_s}{\Lambda}, \quad c_\gamma = \frac{c_1}{\Lambda} q_t e \cos \theta_W, \quad c_Z = \frac{c_1}{\Lambda} q_t e \tan \theta_W, \quad (5.5)$$

where c_1 and c_3 are free coefficients, $q_t = \pm \frac{2}{3}$ is the electrical charge of the top quark, e the elementary charge, θ_W the weak mixing angle and Λ a scale of new physics. In a given theory predicting the existence of a t^* , c_1 and c_3 could take arbitrary values. In the EFT approach I present here, I assume that these are $c_1 = c_3 = 1$, thus considering SM-like couplings for the t^* . With that assumption (and $m_Z \ll m_{t^*}$), the branching fractions are found to be

$$BR(t^* \rightarrow tg) = 0.97, \quad BR(t^* \rightarrow t\gamma) = 0.021, \quad BR(t^* \rightarrow tZ) = 0.0060, \quad (5.6)$$

according to [204]. It is clearly visible that the gluon decay is dominant due to the strength of the strong interaction compared to the electroweak one.

5.2 Previous searches

At the LHC experiments, many searches for vector-like t quark partners are performed [206–212]. However, only few analyses targeting the t^* signature I analyze here exist. At CMS, searches for t^* particles were performed at 8 TeV and 13 TeV. In contrast, no analysis on the $t^*\bar{t}^* \rightarrow t\bar{t}g$ process has been performed by the ATLAS experiment yet.

Run 1 search at 8 TeV A search for pair production of spin- $\frac{3}{2}$ t^* was performed during Run 1 of CMS [215]. It used 9.5 fb^{-1} of proton-proton collision data at

a $\sqrt{s_{pp}}$ of 8 TeV. Exclusively targeting the $t^*\bar{t}^* \rightarrow t\bar{t}g$ process, and assuming a 100% branching fraction for it, events with a single isolated lepton, at least six jets and missing transverse momentum were selected. A mass reconstruction of the $t^*\bar{t}^*$ pair was performed to obtain the sensitive variable of the search. No deviation from the SM prediction was observed, thus cross section exclusion limits on the $t^*\bar{t}^*$ pair production were set. Comparing to theory predictions of the production cross section, a spin- $\frac{3}{2}$ t^* was excluded up to a mass of 803 GeV at 95% confidence. The results were recast onto a spin- $\frac{1}{2}$ scenario, assuming it behaves the same as spin- $\frac{3}{2}$, resulting in an upper mass exclusion limit of 512 GeV.

Run 2 (2016 only) search at 13 TeV Updating the previous CMS result, partial Run 2 data was used to perform a search for $t^*\bar{t}^*$ pair production [214]. Using 13 TeV collision data collected during 2016, corresponding to 35.9 fb^{-1} of integrated luminosity, the same final state as in the Run 1 analysis was targeted: pair production of spin- $\frac{3}{2}$ t^* , decaying exclusively to tg in a single-lepton final state. The analysis relied on a mass reconstruction approach. Using isolated leptons, small-radius jets and missing transverse momentum, the t^* mass was reconstructed. Estimating contributions from SM backgrounds with a log-normal distribution, an unbinned maximum likelihood fit yielded exclusion limits on the $t^*\bar{t}^*$ production cross section after no deviation from SM predictions was observed. The mass exclusion limit for spin- $\frac{3}{2}$ t^* was improved to 1.2 TeV. No analysis of the spin- $\frac{1}{2}$ t^* scenario was performed.

5.3 Analysis

In this thesis I present an updated search for t^* pair production. I investigate the same process as the two previous CMS analyses: $t^*\bar{t}^* \rightarrow t\bar{t}g$ in a single-lepton final state. A representative Feynman diagram of the signal process is shown in figure 5.4. Compared to the most recent CMS analysis using 2016 only data, I add data from 2017 and 2018 for a total of 138 fb^{-1} of integrated luminosity.

My choice to analyze the case where both t^* decay to tg is driven by the expected high branching fraction $\mathcal{B} \approx 97\%$ of that decay. For the resulting $t\bar{t}$ pair, choosing a single-lepton final state is natural to maximize the analysis sensitivity. A lepton can be well distinguished from the hadronic objects in the final state, but it comes at the cost of a neutrino which escapes CMS undetected. However, it can be reconstructed as missing transverse momentum p_T^{miss} , assuming all other objects have been detected. In contrast, while a two-lepton final state would include two easily reconstructible objects in form of the leptons, it has a lower branching fraction and information would be lost due to the presence of two neutrinos. Finally, a fully hadronic final state could also be analyzed. While this poses no issues of lost information, a very high multiplicity of hadronic objects would make this analysis challenging, losing the advantage of the well reconstructible lepton. In the SM, about 15% of $t\bar{t}$ pairs decay into a lepton and jets. In the context of the search I present here, the lepton + jets decay channel (and the name *lepton* ℓ) only refers to e +jets or μ +jets final states. The τ lepton itself is not stable and can decay either hadronically or into an electron or muon, where the latter case is included in the lepton + jets label.

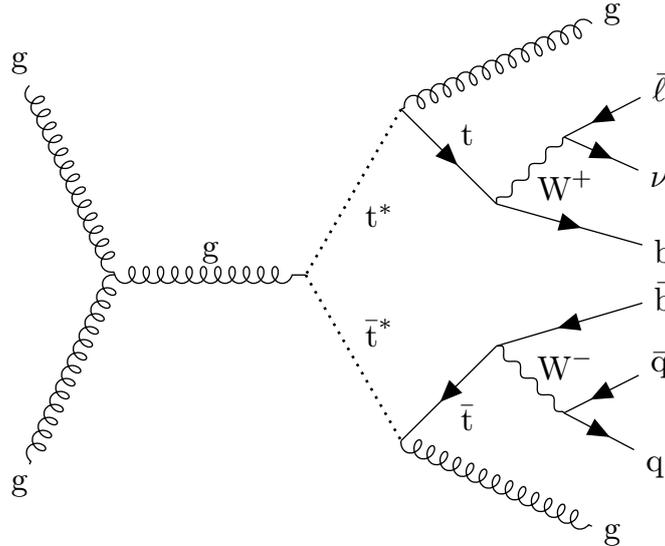


Figure 5.4: Example of a Feynman diagram of the $t^*\bar{t}^* \rightarrow t\bar{t}g$ process with a single-lepton final state. An arbitrary $t^*\bar{t}^*$ production mode is shown. This figure is published in [1].

5.3.1 Analysis overview

The SM predicts several processes that can result in detector signatures similar to that of signal events, called *background processes*. The most relevant of these is $t\bar{t}$ production, in which the only difference to the signal is the two missing gluon jets. However, additional jets not resulting from the hard $t\bar{t}$ event can appear in multiple ways: radiation of gluons from initial or final state particles or from particles in PU events. In addition to $t\bar{t}$, also processes with a single t quark contribute: either produced in association with a W boson, or a light quark. When one of the contained W bosons (either direct or from the t quark decay) decays to lepton and neutrino, the event can appear signal-like due to additional jets not originating from the hard process. These processes are called *t backgrounds*. In addition to these processes containing t quarks, other SM processes contribute as well, which I call *non-t backgrounds*: A production of a W boson with associated jets (W +jets), where the W decays to lepton and neutrino, can appear signal-like, as well as the production of two W or one W and one Z boson (*diboson* processes). Finally, the *Drell-Yan* (DY) process, as well as multijet QCD processes, can contribute when a jet is misidentified as a lepton.

Seeing that there are many potential sources of background, one of the main goals of this analysis is to distinguish potential signal events from these. Therefore, after defining the objects used in this analysis, I select only signal-like events in a *cut-based* selection procedure. Here, the term cut-based refers to low-dimensional combinations of selection criteria on the properties of considered objects. I then use a DNN-based approach to further enrich the analyzed dataset in signal-like events. Finally, the remaining data are analyzed in search for a $t^*\bar{t}^*$ signal.

5.3.2 Monte Carlo simulated samples

In order to understand the signal and background processes, I use simulated samples. They are generated separately for different eras of Run 2, reflecting changes in detector conditions. The production of $t\bar{t}$ pairs and the electroweak production of a single t quark are simulated at next-to-LO (NLO) with POWHEG v2. The cross section for the $t\bar{t}$ background is corrected to a prediction at NNLO accuracy, using a next-to-next-to-leading-logarithmic soft-gluon approximation, obtained with the TOP++ 2.0 program [216]. The strong production of a single t quark (simulated at NLO), as well as production of a W boson with associated jets (W +jets, simulated at LO), are generated with MADGRAPH 2.6.5. Quantum chromodynamics multijet (QCD) and diboson (VV) production are simulated with PYTHIA 8.240. The DY process is simulated at LO with PYTHIA 8.248. For all processes, PYTHIA with the CP5 tune is used for simulation of parton shower and hadronization. The NNPDF 3.1 [129] NNLO PDF sets are used for all samples, and the detector simulation is implemented using GEANT4. Additional inelastic pp collision events are simulated using PYTHIA and superimposed on all simulated events to model the effect of PU. Simulated events are assigned *event weights* so that their contribution to all considered observables follows the expected number of events based on their cross section. Additionally, I generate samples for $t^*\bar{t}^*$ signal events using MADGRAPH at LO. As the t^* mass m_{t^*} is a free parameter of the theory, signal samples using different mass assumptions are created, so that a range of t^* masses can be probed. Similarly, both spin scenarios are considered, thus generating samples with spin- $\frac{1}{2}$ and spin- $\frac{3}{2}$ t^* pair production. For the signal samples, only the $t^*\bar{t}^* \rightarrow t\bar{t}g$ decay mode is considered, but no restrictions are imposed on the decay of the t quarks. The width of the t^* mass distribution is automatically determined by MADGRAPH. The interactions with SM particles are described by the effective coupling approach I describe above. As for the SM backgrounds, I simulate showering and hadronization using PYTHIA, detector interactions using GEANT4 and superimpose PU. Full information on all simulated samples is provided in appendix A.2.1.

5.3.3 Signal process studies

The simulated $t^*\bar{t}^*$ signal samples can be used to study the signature of this process in order to optimize the analysis and understand the obtained results. In figure 5.5, I show some properties of generator-level t^* particles. Comparing the spin- $\frac{1}{2}$ and spin- $\frac{3}{2}$ t^* , some differences can be observed: while the particle content of the different spin scenarios is expected to be the same, the t^* properties are different. Spin- $\frac{3}{2}$ samples have, on average, higher t^* p_T , especially for low t^* masses. For higher masses, the momentum distributions become more similar. t^* are mostly produced centrally in the detector, and I observe no significant angular differences between the spin scenarios. In figure 5.6, I show distributions of the generator-level t quark pairs originating from the $t^*\bar{t}^*$ decay, and compare them to a SM $t\bar{t}$ pair (assuming single-lepton decay). As expected, the t quark momentum is much higher in $t^*\bar{t}^*$ events compared to $t\bar{t}$ events, and the same is true for the invariant mass $m_{t\bar{t}}$. The angular distance between the two t quarks is shown as well, where one can see further differences: for SM $t\bar{t}$, the t quarks are mostly back-to-back (at $\Delta R \approx \pi$), whereas the t quarks in $t^*\bar{t}^*$ events can be closer together. Finally, the decay products of the t quark pair can be investigated. As a single-lepton final state

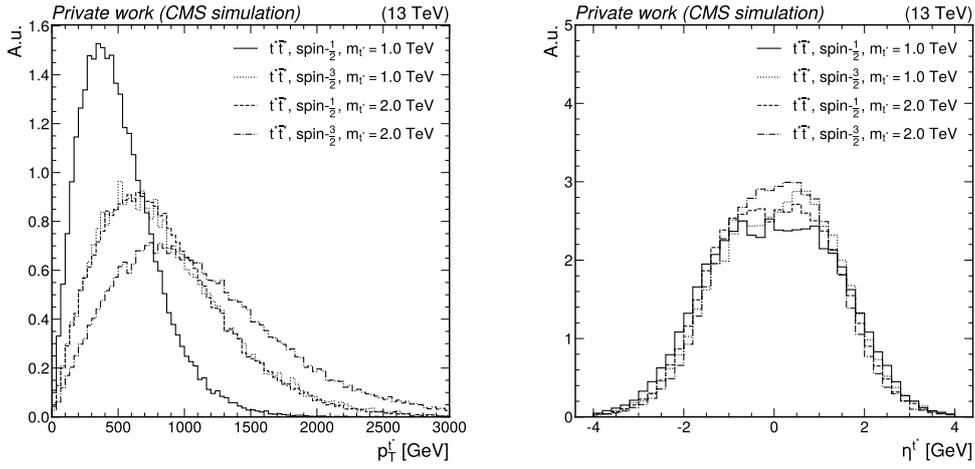


Figure 5.5: Left: p_T of generator-level t^* particles, assuming different masses and spin scenarios. Right: η of the same generator-level t^* particles.

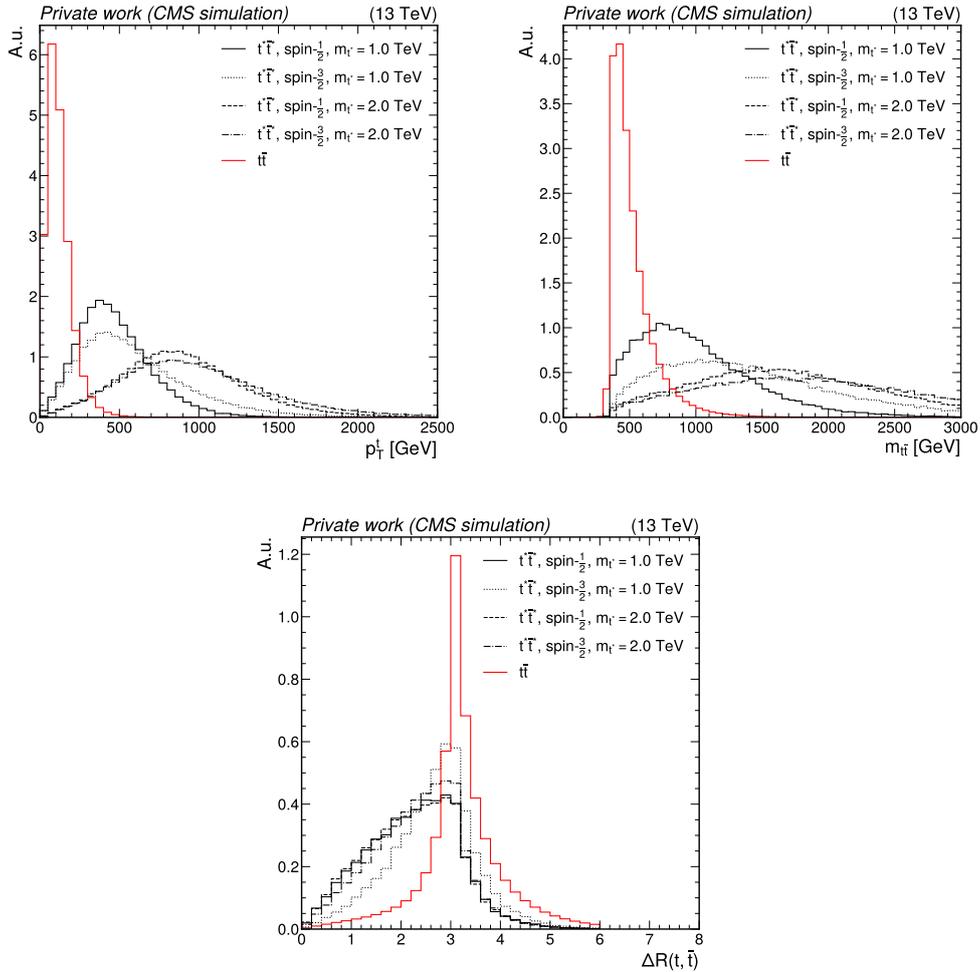


Figure 5.6: Upper left: Generator-level t quark p_T . Upper right: Invariant mass of the $t\bar{t}$ system. Bottom: Angular distance between t quarks (right). All figures compare $t^*\bar{t}^*$ signal samples of different t^* masses and spin scenarios to the SM $t\bar{t}$ process (assuming a single-lepton final state).

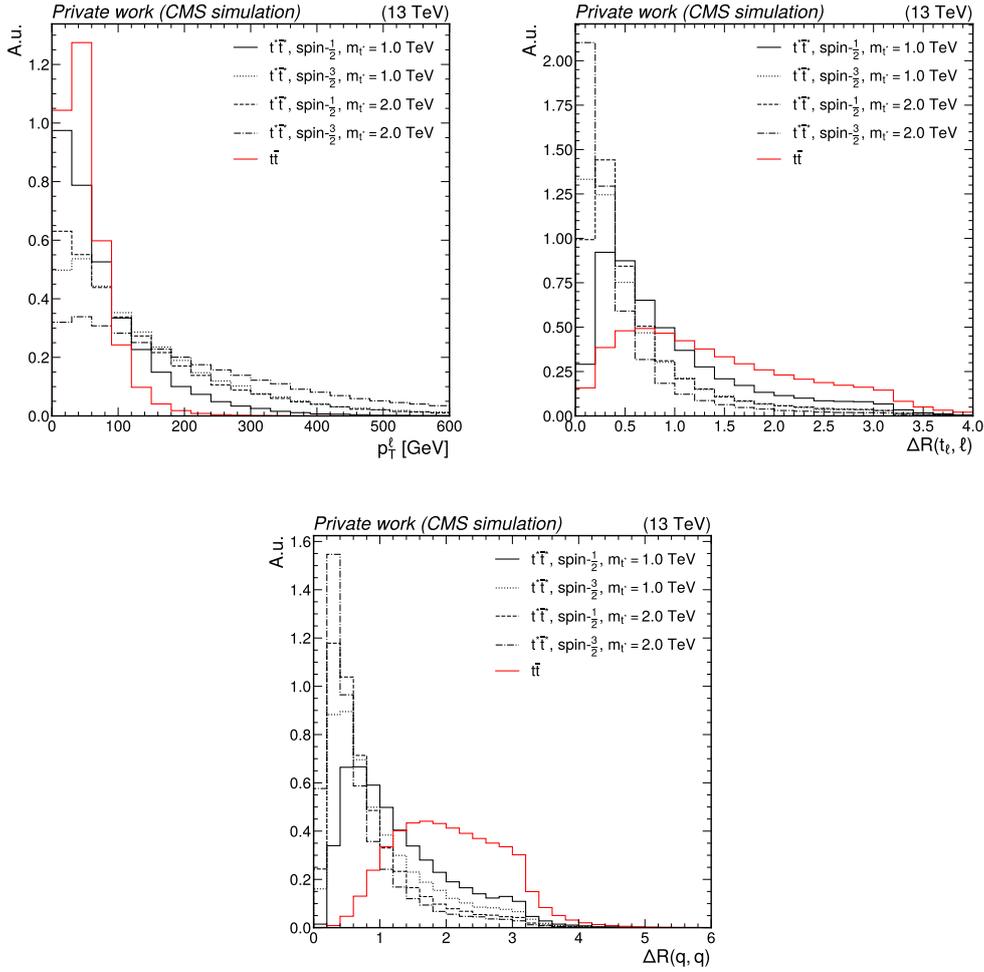


Figure 5.7: Upper left: Generator-level lepton p_T . Upper right: Angular distance between generator-level lepton and t quark. Bottom: Angular distance between two light quarks (right) from a W boson decay. All figures compare $t^*\bar{t}^*$ signal samples of different t^* masses and spin scenarios to the SM $t\bar{t}$ process (assuming a single-lepton final state).

is analyzed, I show some information on the generator-level lepton in figure 5.7: the lepton p_T and the distance of the lepton to the trajectory of the t quark it originated from. One can clearly see differences in lepton p_T between $t^*\bar{t}^*$ and $t\bar{t}$. Also, the lepton tends to be closer to the t quark in signal events. Additionally, I include the angular distance between the two light quarks from the other t quark decay in the figure, showing that these tend to be much closer to each other in $t^*\bar{t}^*$ events than in $t\bar{t}$.

5.3.4 Object definitions

I impose conditions onto reconstructed physics objects in order to use them in the following analysis steps. These usually involve requirements on p_T and η of an object, but also application of so-called *quality IDs*. These IDs are a set of requirements on various properties of the considered object and aim to reject objects that were misidentified, while keeping the acceptance efficiency of correct objects as high as possible. In general, it is crucial to ensure that the description of all relevant object properties in simulated samples agrees with what is seen in data. Due to the limitations of MC simulations or slightly different settings between MC generation and data-taking, this is not necessarily always the case. Similarly, detector effects and properties of the reconstruction algorithms might result in the measured data not perfectly corresponding to what actually happened in the detector. To alleviate these issues, I employ correction *scale factors* (SFs) beyond the calibrations already happening during event reconstruction, as I will describe below.

5.3.4.1 Leptons

Events with electrons and muons are considered in this analysis, which is divided into an electron and muon channel.

Electrons In order to be considered for this analysis, electrons need to have $p_T > 40$ GeV. This momentum threshold is mainly driven by the triggers used for this analysis, which I introduce below. Electrons must to be contained within $|\eta| < 2.4$ to be inside the tracker system coverage and consistent with the muon selection. Electrons in the barrel-endcap overlap region between $1.444 < |\eta| < 1.566$ are excluded. Different quality IDs are defined within the CMS reconstruction environment, trying to distinguish real prompt electrons from fake ones [217]. To find the optimal ID to use in this analysis, I perform an efficiency study. From a simulated signal sample (spin- $\frac{1}{2}$, $m_{t^*} = 1500$ GeV), I select events that contain two generator-level top quarks, of which one decays to a final state with a generator-level electron. The events are required to contain exactly one electron with $p_T > 40$ GeV and $|\eta| < 2.4$, which needs to be matched to the generator-level electron within $\Delta R < 0.2$. In this set of events, I measure the efficiency of this electron to pass a certain quality ID. The results are shown in figure 5.8.

I test three working points of a cut-based ID [217], selecting proper electrons using a combination of selection criteria on the properties of the electron, its track and calorimeter signature. The three working points corresponds to 90 % (loose), 80 % (medium) and 90 % (tight) average efficiency to select prompt electrons. Additionally, I test a dedicated cut-based high energy electron identification (HEEP) ID

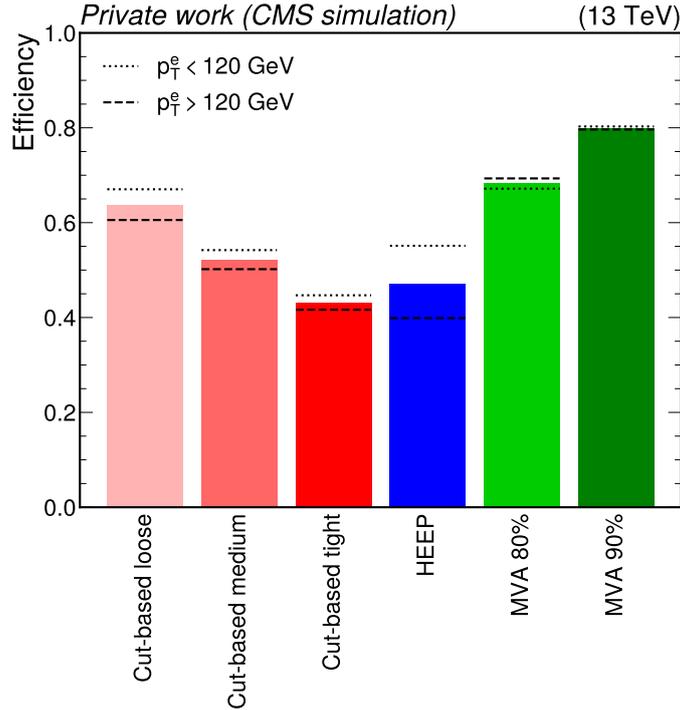


Figure 5.8: Efficiency of electrons in signal samples to pass a certain quality ID, with no restrictions on their momentum. Results for low- and high- p_T electrons are shown separately as black lines.

[217], as well as a multivariate-analysis (MVA) based ID [217] with 80 % and 90 % efficiency working points. The MVA IDs rely on boosted decision trees to combine information from multiple electron parameters into a single quality score, on which thresholds define the working point. It can be seen that the best performance is achieved using the MVA ID at its 90 % working point. I repeat the test separately for electrons above and below $p_T = 120$ GeV. While the results of the other IDs are mostly unchanged, the HEEP ID varies between 35 % and 55 % efficiency, but performs worse than the MVA IDs for all energies. Interestingly, it performs worse for highly energetic electrons: the reason for this most likely are intrinsic isolation requirements which lower the signal efficiency at high energies. Therefore, I chose the MVA ID at its 90 % working point for this analysis. For this test, no isolation criteria apart from those contained in the IDs themselves are considered: instead, I present detailed studies on lepton isolation below.

To ensure that the electron reconstruction efficiency is well described in simulation, I apply correction SFs. These are derived using a tag-and-probe approach on electrons from a Z boson decay: events with one well reconstructed electron are probed for the efficiency to reconstruct a second electron, requiring the invariant mass of the dielectron system to be around the Z mass. Comparing the results between data and simulation then allows a determination of the SFs. Similarly, the effect of the quality ID is evaluated using tag-and-probe in events with one well reconstructed electron and no restrictions on the second electron.

I refer to electrons below $p_T = 120$ GeV as low- p_T electrons, and electrons above that threshold as high- p_T electrons. I impose isolation requirements onto electrons, which I discuss below.

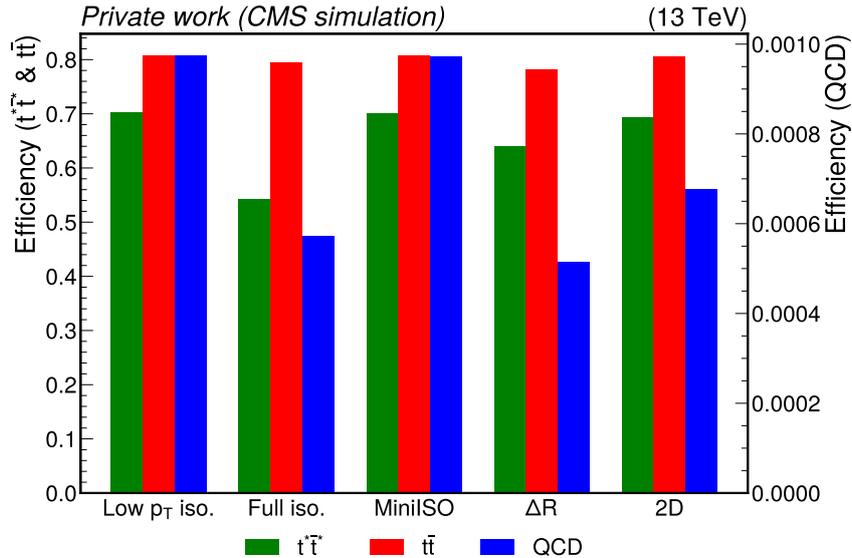


Figure 5.9: Efficiency of requiring exactly one electron fulfilling different isolation criteria in $t\bar{t}$, $t\bar{t}$ and QCD events. For $t\bar{t}$ and $t\bar{t}$, only events containing a pair of generator-level top quarks and a single generator-level electron are considered. QCD is shown on the right y -axis to keep it visible despite low efficiencies.

Muons I consider global muons with $p_T > 30$ GeV, motivated by the available HLT paths. Muons with $p_T < 55$ GeV are referred to as low- p_T muons, and muons above that threshold are high- p_T muons. All muons need to be contained within $|\eta| < 2.4$, according to the coverage of the muon system. For low- p_T muons, I impose an ID containing a set of requirements on the tracker and muon system signature to select prompt muons [218]. A different quality ID, specifically designed for high- p_T muons, is used for those [219]. Scale factors, measured using a tag-and-probe approach, are used to ensure proper selection efficiency description in simulation.

Lepton isolation QCD processes have a very high cross section and thus need to be strongly suppressed so that they not dominate this analysis. In general, no leptons are expected in QCD, so a large fraction of QCD events are already rejected when I require a lepton in each event. However, in some QCD events, leptons arise from misidentified jets or as decay products of mesons. Both cases have in common that the lepton is usually close to hadronic detector activity. Therefore, the QCD process can be strongly suppressed by requiring isolated leptons. Isolation can be described by the *relative isolation* variable I_{rel} , defined as the scalar sum of the transverse energy of all reconstructed particles contained within a cone of 0.4 (0.3) around a muon (electron) [144, 217, 220].

For low- p_T leptons, isolation is already imposed at trigger level to reduce rates, and thus can also be used offline without any avoidable loss of signal. Thus, I require low- p_T muons to fulfill $I_{\text{rel}} < 0.15$. For low- p_T electrons, isolation is included as part of the quality ID. Applying similar requirements to high- p_T leptons is problematic due to the expected lepton signature in signal events. In figure 5.7, I show that the generator-level lepton in signal events tends to be closer to the t quark than in $t\bar{t}$ events. The reason for this behavior is the Lorentz boost of the t quark,

resulting from the high m_{t^*} . Thus, high- p_T leptons in signal events are expected to be close to the b -quark induced jet from the t quark decay. Requiring isolation would therefore lead to a loss in signal efficiency. To gauge the strength of this effect, I perform isolation studies in the electron channel. In figure 5.9, I show the efficiency of requiring an electron fulfilling certain isolation criteria in different simulated samples.

Labeled as “low p_T iso.”, I present a setup where no isolation is required for high- p_T leptons. It can be understood as a baseline of this study, since it is the loosest possible isolation strategy. In contrast, “full iso.” labels the case where the low- p_T lepton requirements are extended to the full p_T range. While this lowers the contribution of QCD substantially, it also strongly reduces the signal efficiency. In contrast, $t\bar{t}$ is only slightly affected, showcasing the difference in behavior between $t\bar{t}$ and $t^*\bar{t}^*$ due to the Lorentz boost of the t quarks in signal events. Three other isolation strategies are shown. “MiniISO” describes an approach using a variant of the I_{rel} variable, where the cone size in which surrounding activity is considered is lowered the higher the lepton p_T is, from 0.2 below 50 GeV to 0.05 above 200 GeV. In the given example, a MiniISO threshold of 0.1 is tested, but does not show any substantial changes compared to the baseline scenario. Another option is a ΔR criterion between the lepton and the closest jet, which is tested at $\Delta R > 0.4$, based on the jet radius of small-radius jets. While this results in a strong QCD suppression, also signal is somewhat affected. The ΔR criterion can be extended to obtain the so-called *2D isolation criterion*: Requiring either $\Delta R > 0.4$ or a $p_T^{\text{rel}}(\ell, j) > 25$ GeV where $p_T^{\text{rel}}(\ell, j)$ is the p_T component of the lepton perpendicular to the axis of the closest small-radius jet, this criterion loosens the impact on signal events by not applying the ΔR criterion to all leptons. As the 2D criterion retains a high signal efficiency but lowers the QCD contribution, I chose it as the isolation criterion for this analysis and apply it to events with high- p_T leptons.

5.3.4.2 Jets

I use both small- and variable-radius jets in this analysis, as defined in section 3.2.4.2, where both cluster the full available set of PF candidates.

Small-radius jets Each small-radius jet needs to have $p_T > 30$ GeV to ensure efficient jet reconstruction. Additionally, they need to be contained within $|\eta| < 2.5$, the coverage of the tracking system. PU suppression is performed using the PUPPI algorithm. The resulting small-radius jets are well suited to reconstruct jets from light quarks and gluons. As the detector response to particles is not linear, the measured energy of a jet can not be trivially related to the true energy of the source particle. Therefore, calibration of jets to match the energy and momentum of the source particle is crucial. CMS utilizes a factorized approach, applying multiple correction steps sequentially, each accounting for a different effect. A schematic overview of the different steps is provided in figure 5.10.

The jet response, meaning the measured energy of a jet relative to generator-level information, is calculated in simulated events, and then correction factors are applied to both data and simulated events to obtain a uniform jet response for all p_T and η . Furthermore, two steps of residual corrections are applied to data events to correct for small differences in response between data and simulation.

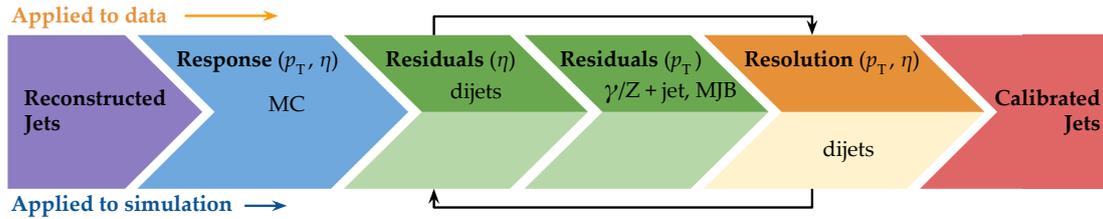


Figure 5.10: Overview of the jet calibration steps I use for PUPPI jets, showing corrections of the jet response (data and simulation), residual corrections (data only) and the jet resolution smearing (simulation only). The last three steps are applied iteratively until a stable result is reached, as they depend on each other. Taken from [221].

These corrections are referred to as *jet energy corrections* (JEC) in the following. It was found that the jet momentum resolution in data is worse than in simulation. Therefore, the jet resolution in simulated events is smeared so that data and simulation agree, which will be called *jet energy resolution smearing* (JER). The derivation of these jet corrections is documented in [222].

Variable-radius jets Variable-radius jets must have $p_T > 200$ GeV and be within $|\eta| < 2.5$. They are well suited for this analysis, as due to the different considered t^* masses, the Lorentz boost of the t quarks in signal events can vary strongly. An approach utilizing variable-radius jets can provide good performance over a wide range of t quark momenta, and splitting the analysis into boosted and resolved regions can be avoided. Variable-radius jets reconstructed with the HOTVR algorithm were first used in a search for excited b quarks [223]. There, it was shown that the same jet corrections used for small-radius jets can be applied to the subjets of HOTVR jets. Recombining these then yields a well calibrated variable-radius jet [224]. I use the same approach for this analysis and have validated that it still holds true.

Jet-lepton cleaning When clustering jets, all PF particles are considered. Thus, energy and momentum contributions from leptons can be included in the resulting jets. However, as I consider non-isolated leptons that can be close to or inside of jets, this would lead to a double-counting of energy. To avoid this, *jet-lepton cleaning* is employed. For all leptons considered in this analysis, the four-momentum is subtracted from any nearby jet if the lepton is inside that jet: $\Delta R(j, \ell) < R_j$. Here, R_j is 0.4 for small-radius jets and allowed to vary for variable-radius jets, where this subtraction is performed on subjet level. For small-radius jets, an additional criterion is required: the leptonic fraction of the jet energy is required to be compatible with the leptons energy within 10%, where the leptonic energy fraction is defined as the p_T sum of all lepton constituents of the considered jet.

b -tagging I use the DeepJet algorithm to identify jets originating from b quark decays. I refer to small-radius as b -jets if their DeepJet score is above a certain threshold, which is chosen at a 1% light jet misidentification rate. Additionally, the DeepJet score of jets is used directly later in the analysis. The DeepJet algorithm was trained on jets using CHS for PU suppression, in contrast to the PUPPI jets I use in this analysis. While no major effects of the PU suppression method on

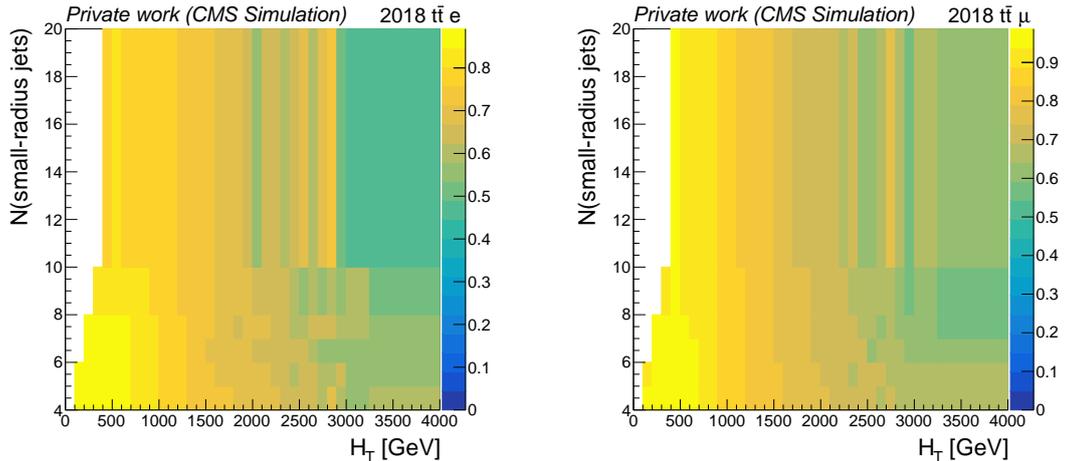


Figure 5.11: Example of the additional two-dimensional b -tagging yield SFs for $t\bar{t}$ events in the 2018 era, for the electron (left) and muon (right) channel.

DeepJet are expected, this can not be verified trivially. Therefore, I chose a hybrid jet approach to avoid any unexpected behavior. I reconstruct, clean and correct CHS jets in the same way as the PUPPI jets described above, with a minor difference: CHS jets require an additional jet correction step, correcting for PU before JEC and JER corrections [222]. Then, for each PUPPI jet, I find the CHS jet closest in ΔR , requiring $\Delta R < 0.2$ for two jets to be matched. If a matching CHS jet is found, the PUPPI jet is then assigned the DeepJet score of the matched CHS jet.

As the DeepJet score is not necessarily perfectly described in simulation, correction SFs are applied [175, 225]. These assign weights to each event based on the contained jets, where each jet contributes a factor depending on its p_T , η and DeepJet score. As these are derived for CHS jets, I apply them based on the CHS jets matched to the PUPPI jets I use in this analysis.

Crucially, for a given simulated sample, the applied weights should only change the shape of the DeepJet score variable, but not introduce an overall normalization effect. However, due to the analysis-dependent jet selection criteria and resulting composition of SM processes, this can not be ensured when creating the SFs, which are shared among many analyses in CMS. Therefore, I derive secondary, analysis-specific yield correction SFs to remove any normalization changes specific to the presented analysis. As these changes might vary based on the number of jets or total hadronic momentum H_T in an event, this needs to be taken into account, as well as differences between events containing an electron or muon. Therefore, I derive corrections separately for the electron and muon channels, and as a function of jet multiplicity and H_T . Examples of the yield SFs for $t\bar{t}$ events are given in figure 5.11. These show a systematic lowering of the SF towards high H_T and high jet multiplicity. Some statistical fluctuations from the limited size of the simulated dataset are visible.

To ensure that the yield correction scale factors have the desired effect, I evaluate different variables before any b -tagging related SFs, after application of the b -tagging correction SFs and after the secondary yield corrections. Results for $t\bar{t}$ events are shown in figure 5.12. A strong normalization change is induced by the b -tagging SFs in H_T and $N(\text{small-radius jets})$, especially towards higher values of these variables.

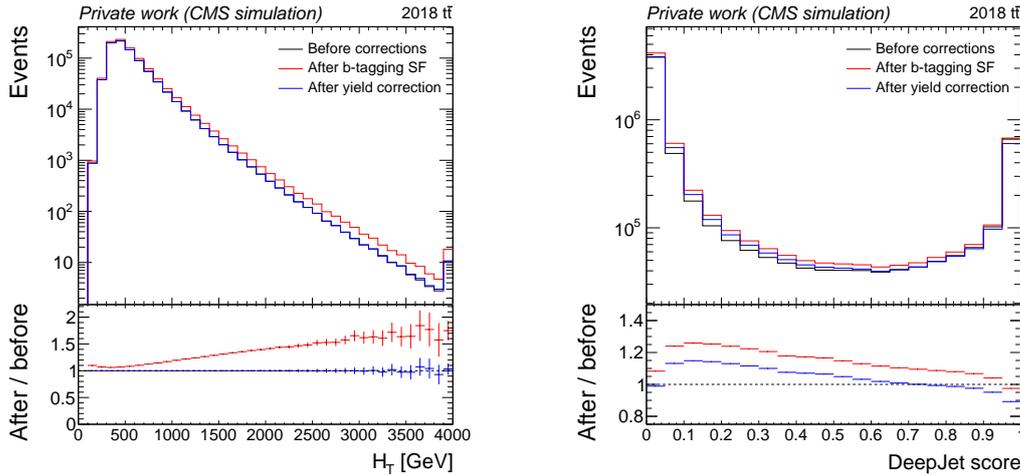


Figure 5.12: Distributions of the hadronic momentum sum H_T (left) and DeepJet score of all jets (right) at different steps of the b -tagging correction procedure. For these results, electron and muon channels are combined in the 2018 era.

The yield corrections are able to remove this effect, while only altering the DeepJet score in a constant manner, thus retaining the desired effect of the b -tagging SFs.

t -tagging In addition to a definition of b -jets, a HOTVR t -tag is used in the analysis, but only for validation purposes. The definition follows the HOTVR t -tag introduced in 3.2.4.3, with a τ_3/τ_2 threshold of 0.56. As this t -tag is only used for procedure validation, no SFs are used to correct the tagging results.

5.3.4.3 Missing transverse momentum

The missing transverse momentum p_T^{miss} is defined as the negative p_T sum of all PF candidates in an event, after PUPPI has been applied. The effect of JEC is taken into account, as well as JER smearing for the absolute value of p_T^{miss} .

5.3.5 Cut-based event selection

I apply several cut-based selection steps to data and simulated events, aiming to select signal-like events and suppress contributions from SM backgrounds.

5.3.5.1 Trigger selection

The first step of any analysis with CMS data is the selection of appropriate HLT paths. As I target t^*t^* events with a single-lepton final state, I rely on this property. Different triggers are used in the muon and electron channels, as documented in tables 5.2 and 5.1, respectively. There, I list the p_T thresholds of these triggers, which vary between years due to changes in instantaneous luminosity.

I trigger on low- p_T electrons using paths requiring a single, isolated electron. In addition to the momentum requirement, a set of quality criteria is used to select prompt isolated electrons [217]. Another algorithm, with higher p_T threshold, triggers on electrons without any isolation requirements. Quality requirements on the

year	all electron events
	Isolated electron with $p_T > 27$ GeV
2016	Electron with $p_T > 115$ GeV
	Photon with $p_T > 175$ GeV
	Isolated electron with $p_T > 35$ GeV
2017	Electron with $p_T > 115$ GeV
	Photon with $p_T > 200$ GeV
	Isolated electron with $p_T > 32$ GeV
2018	Electron with $p_T > 115$ GeV
	Photon with $p_T > 200$ GeV

Table 5.1: Electron triggers used in 2016, 2017 and 2018. All triggers within a year are combined using a logical OR.

year	$p_T^\mu \leq 55$ GeV	$p_T^\mu > 55$ GeV
2016	Iso. standalone μ with $p_T > 24$ GeV	Standalone μ with $p_T > 50$ GeV
	Iso. tracker μ with $p_T > 24$ GeV	Tracker μ with $p_T > 50$ GeV
2017	Iso. global μ with $p_T > 27$ GeV	Global μ with $p_T > 50$ GeV
		Tracker μ with $p_T > 100$ GeV
		Standalone μ with $p_T > 100$ GeV
2018	Iso. global μ with $p_T > 24$ GeV	Global μ with $p_T > 50$ GeV
		Tracker μ with $p_T > 100$ GeV
		Standalone μ with $p_T > 100$ GeV

Table 5.2: Muon triggers used in 2016, 2017 and 2018, split into low- and high- p_T muons. All triggers within a year and p_T range are combined using a logical OR.

calorimeter cluster and track associated with the electron ensure good selection efficiency of prompt electrons [217]. As this combination of triggers has an inefficiency for very high- p_T electrons due to too tight electron cluster shape thresholds, I add a photon trigger in a logical OR: As this trigger does not consider any tracking information, it is also sensitive to electrons and can recover the inefficiencies.

I use triggers requiring at least one isolated muon with low p_T thresholds, and complement these by triggers targeting muons without isolation requirements but higher p_T requirements. Different muon reconstruction types are used, as documented in section 3.2.4.1: In 2016, muon triggers relied on a combination of tracker muons and standalone muons. In 2017 and 2018, the main muon triggers used a HLT implementation of global muons, combined with tracker and standalone muons to ensure optimal efficiency [226].

The application of trigger selection criteria slightly differs between events with electrons and muons: where in the electron channel, a total combination of all considered triggers per era is used, the application is split by muon momentum in the muon channel. The reason lies in the availability of correction SFs for lepton properties, which are derived separately for low- and high- p_T muons, but combined for all electrons. I investigated this difference and found that the effect is negligible, as the lepton momentum thresholds that divide low- from high- p_T leptons are far enough away from the trigger p_T thresholds.

In the context of offline analysis, data gathered by different triggers is available in form of different *primary datasets*, combining data gathered by related triggers. For this analysis, I use the single muon primary dataset for muon events. Electron events are taken from a single electron or single photon primary dataset (2016 and 2017) or a combined e/γ primary dataset (2018). Notably, the same event might be contained in multiple primary datasets, for example when it was selected by both electron and muon triggers. Therefore, I avoid potential trigger double counting by sequentially selecting events from the primary datasets, only accepting those not taken from a previous stream. Here, the muon primary dataset is used first, then the electron and lastly the photon primary dataset.

To ensure a proper description of the collisions data by the simulated samples, the same trigger conditions used to gather data events also need to be considered in MC samples. As this might lead to slightly different results due to imperfections of the simulations, correction SFs are used to ensure a proper description. For the electron triggers, I measure dedicated trigger efficiency scale factors for this analysis. Using an orthogonal dataset method targeting $t\bar{t}$ events with two leptons, events are selected using the combination of muon triggers shown in table 5.2. Then, I apply all cut-based selection steps listed in this section, with two changes: I require two leptons, one muon and one electron, and I loosen the jet multiplicity selection to just require the presence of two small-radius jets. After these requirements, I calculate a trigger efficiency per year as a function of electron η in three bins of electron p_T . Efficiency results in the 2018 era, comparing the efficiency in data against the efficiency in $t\bar{t}$ MC, are shown as an example in figure 5.13a.

As can be seen, a difference between efficiencies in data and simulation exists. Therefore, I calculate a correction SF from the ratio of these efficiencies as

$$SF = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}. \quad (5.7)$$

Modifying the event weight of simulated events by the appropriate SF value cor-

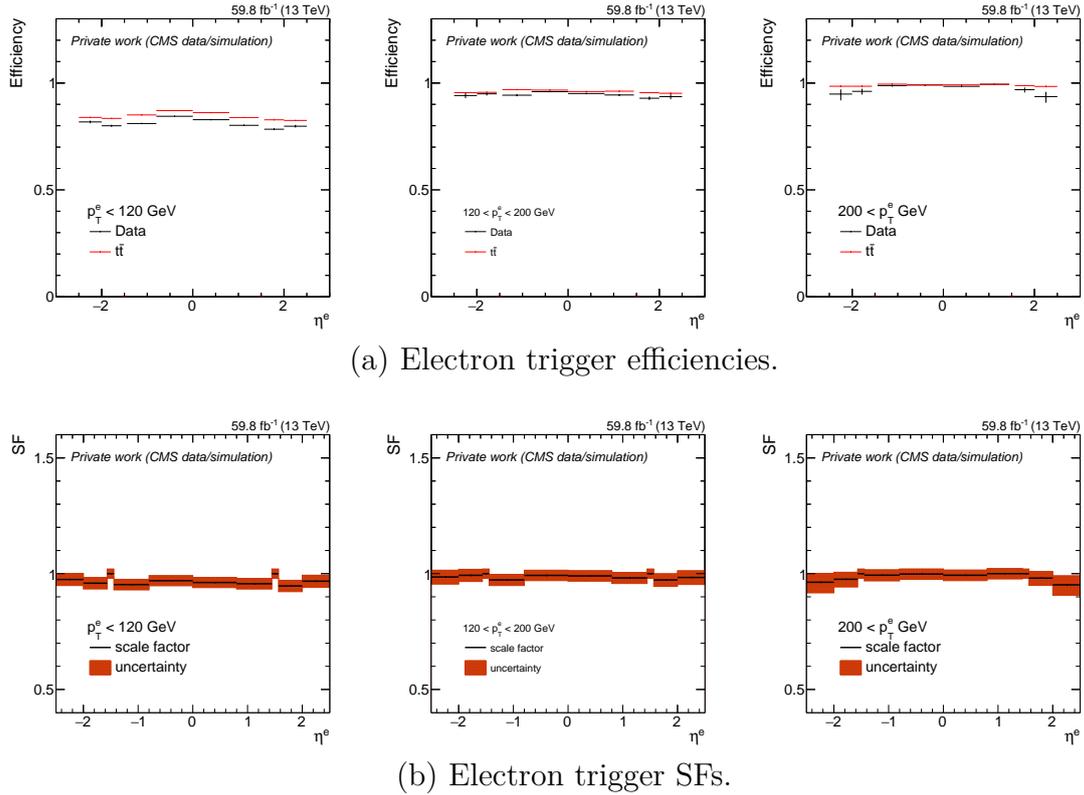


Figure 5.13: Electron trigger efficiencies (top) and resulting SFs (bottom), as used in this analysis. Efficiencies are measured in an orthogonal dataset from a single muon trigger. I compare data events to $t\bar{t}$ simulation results as a function of electron η , in three bins of electron p_T , using 2018 samples.

rects these events so that the efficiency is properly described. In order to assign a systematic uncertainty to this method, I take the available statistical power of the data and simulated samples used here into account and propagate it to the scale factors as

$$SF_{\text{variation}} = SF + \left(-\frac{\epsilon_{\text{data}}}{(\epsilon_{\text{MC}})^2} * \epsilon_{\text{MC}}^{\text{variation}} + \frac{1}{\epsilon_{\text{MC}}} * \epsilon_{\text{data}}^{\text{variation}} \right) \quad (5.8)$$

where “variation” denotes the up- and down-variation of the statistical uncertainty in the input samples, respectively. In addition, a flat 2% uncertainty is considered to cover potential non-closure. The resulting scale factors for the 2018 era are shown in figure 5.13b. Efficiencies and SFs for all eras are documented in appendix A.2.2.

For the muon triggers, efficiency SFs are measured in data and simulation using a tag-and-probe method [144] in DY events or targeting muons from the Z boson resonance, depending on muon momentum. Efficiencies are measured as a function of muon p_T and η , and validated by comparing to measurements in an orthogonal p_T^{miss} dataset. By dividing the efficiencies in data and simulation, a SF is obtained, and a systematic uncertainty can be assigned to it.

5.3.5.2 Lepton multiplicity

In order to target the single-lepton final state of the t quark pair decay, exactly one lepton is required, either a muon or electron. This mostly removes contribution from di-lepton $t\bar{t}$ decays, as well as fully hadronic events. At this step, events are split into an electron and muon channel.

5.3.5.3 Loose hadronic pre-selection

After the application of trigger condition and requiring a single lepton, I impose a *loose hadronic pre-selection*. The reason for this is of technical nature: removing some events early, before considering most corrections and SFs, substantially reduces the computing runtime and storage requirements of the analysis. Therefore, early in the analysis process, I demand the presence of at least two small radius jets (without any p_T or η requirements yet) and total object momentum above 450 GeV. The total momentum sum is calculated as a scalar sum of the p_T of all variable-radius jets, the lepton and p_T^{miss} . Similar, but more strict selection criteria are applied later. Thus, this loose pre-selection has no effect on the final result of the analysis. After this step, most correction SFs are applied.

5.3.5.4 Jet multiplicity selection

The jet multiplicity selection follows the objects expected in signal events. I require a single variable-radius jet. In signal events, at least one of these is expected to be present from the t quark decaying to a b quark and two light quarks (through a W boson). However, additional variable-radius jets can arise when light quark, b quark or gluon jets are reconstructed by the HOTVR algorithm. This is possible especially for highly energetic jets in the case of high m_{t^*} , as the HOTVR radius can then become small.

Additionally, I require at least four small-radius jets. Accounting for the two gluons, one isolated b quark jet and one small-radius jet overlapping with hadronically decaying t quark, this selection results in minimal loss of signal events, while suppressing the background contributions substantially.

No overlap removal between small-radius and variable-radius jets is applied, thus the same particle can induce jets of both types.

5.3.5.5 p_T^{miss} selection

A neutrino from the W decay in the leptonic leg of the $t\bar{t}$ decay escapes CMS undetected. To account for it, I demand at least 50 GeV of p_T^{miss} .

5.3.5.6 b -jet selection

Signal events are expected to contain jets originating from b quarks and t quarks. However, the main SM background process, $t\bar{t}$, also contains these same quarks. Thus, a selection on t -tagged jets would select approximately the same fraction of signal and $t\bar{t}$ background events, which is not beneficial to the sensitivity of the analysis. Therefore, I do not perform t -tagging. The same argument could, in principle, be made against b -tagging. However, suppressing backgrounds without b quarks, like W +jets or QCD, is crucial even when these are not the dominant

background. Thus, I employ a loose b -tagging selection, requiring at least one b -jet to be present in each event. This selection allows me to suppress non- t background contributions without losing too many signal events.

5.3.5.7 2D isolation selection

The 2D isolation criterion is introduced as a lepton property above. However, as it also considers jets, it can also be understood as an event-level selection requirement. As such, it is implemented here, at a late point in the selection procedure, where every event in which the lepton does not fulfill the 2D criterion is rejected.

5.3.5.8 Total transverse momentum S_T

An important variable I define for this analysis is the total transverse momentum S_T , given as

$$S_T = \sum p_T^{\text{variable-radius jet}} + p_t^\ell + p_T^{\text{miss}}. \quad (5.9)$$

Usually, signal events have high S_T due to the massive t^* that are present, compared to SM background events which are mostly at low S_T . Thus, this variable is sensitive to the presence of signal and serves as the observable used in the statistical analysis which I present in section 5.4.1. Additionally, it can be used to reject a region where no signal events are expected at $S_T < 500$ GeV. This reduces the amount of analyzed events and thus speeds up all following analysis steps, without losing any sensitivity to the $t^*\bar{t}^*$ signal.

5.3.6 Corrections and quality selections

Several minor procedures and SFs are used to correct for small effects that would influence the data or simulations and lead to disagreements.

5.3.6.1 MC PU reweighting

The PU distribution in simulated samples does not agree perfectly with the one observed in data, which I address using an event reweighting approach. Assuming a total inelastic cross section of 69.2 mb [227], the distribution of PU is determined in data and the MC PU distribution reweighted to match it.

5.3.6.2 HEM15/16 issue

The HEM15/16 issue refers to two HCAL modules whose power supply failed during data-taking in 2018. This outage affected a region in $-3.0 < \eta < -1.3$ and $1.57 < \phi < -0.78$ in the η - ϕ plane. The failure would lead to a miscalibration of jets in that area, as well as an increased rate of misidentified electrons. To address this issue, any data event from that part of data-taking that has an electron, jet or muon in the affected region is rejected. To account for this change in simulated samples, MC events with objects in this η - ϕ region are weighted according to the affected luminosity fraction: a factor of 0.352. Reweighting the simulated samples this way, instead of simply rejecting events, retains the full statistical power of the samples while properly describing the reduced amount of data in the affected region.

5.3.6.3 Trigger prefiring

As I introduce in section 3.2.3.1, TPs are built as part of the L1T reconstruction. Using timing information, each TP is assigned to a bunch crossing. Should a TP mistakenly be assigned to an earlier bunch crossing, its information is lost from the bunch crossing it actually originates from. Furthermore, should the TP result in a trigger firing, the correct bunch crossing can not be triggered on due to so-called *trigger rules*: When, based on the L1T decision, an event is accepted, these rules do not allow any event from the following two bunch crossings to be accepted.

The resulting loss in correctly triggered events is called *trigger prefiring*. During 2016 and 2017 data-taking, a timing shift of the ECAL was not properly corrected for when building TPs. Therefore, some events with ECAL-related objects were strongly affected by trigger prefiring, reaching up to 80% in some regions of p_T and η [138]. A similar effect occurred in the muon systems due to the limited time resolution of the muon detectors: an effect of up to 1.5% which was most pronounced in 2016 [226]. Gauging the effect of prefiring is challenging, as by the nature of the problem, the affected events are not triggered on, and thus can not be analyzed. However, events triggered exactly three bunch crossings after another event was selected can not be affected by prefiring. Thus, these events serve as a dataset not biased by trigger prefiring. It is used to derive SFs in order to account for prefiring in simulations, which do not describe the effect.

5.3.6.4 Noisy event filters

Several filters are applied to exclude events not fulfilling a wide range of quality criteria, the so-called *noisy event filters*. Interactions of the proton beams with gas particles or the beam pipe lead to the creation of a *beam halo* [228], particles flying through the LHC around the beam. High energy muons from the halo can interact with the CMS calorimeters and result in energy clusters there, thus events containing clusters likely to originate from the halo are excluded. Scintillator tiles in the HB or HE can sometimes produce fake noise signals. Different techniques allow to identify these and remove affected events [229]. Similarly, ECAL cells might rarely malfunction which can be seen from their energy measurement being close to saturation. Affected events are removed.

5.3.6.5 Top p_T reweighting

In NLO simulations of events containing t quarks, their p_T distribution is not perfectly modeled, especially in the high p_T region. This can be addressed by t quark p_T dependent SFs in order to improve the agreement of data and simulations in samples with t quarks. The scale factor is given by

$$SF(p_T^t) = e^{0.0615 - 0.0005 p_T^t} \quad (5.10)$$

obtained from $t\bar{t}$ cross section measurements [230, 231]. It is multiplied to the event weight of simulated samples with t quarks. For $t\bar{t}$ samples, where two t quarks are contained, the SFs are combined into a total weight w as

$$w = \sqrt{SF(p_T^t)SF(p_T^{\bar{t}})}. \quad (5.11)$$

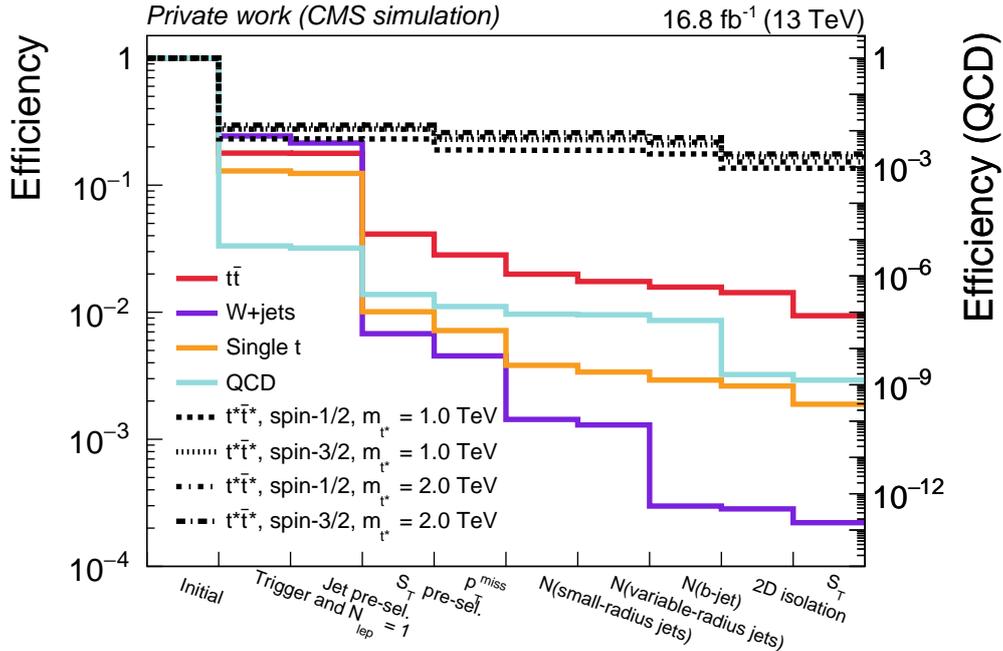


Figure 5.14: Step-by-step efficiencies of all cut-based selection parts, evaluated on full simulated samples of relevant SM background processes and representative signal samples. The QCD background is shown on the right y-axis due to its very low overall efficiency.

5.3.6.6 $V + \text{jets}$ theory weights

The $W + \text{jets}$ and DY simulated samples that are used in this analysis are simulated at LO only. To account for missing NLO QCD and electroweak contributions, corrections are applied as a function of the generator-level vector boson p_T . Derived in [232], these corrections result in a better agreement of data and simulation for these background sources. Note that their contribution is later be estimated from data, so I only use these SFs to better optimize the analysis, but they do not enter its final result.

5.3.7 Selection results

With all corrections applied, and all cut-based selection steps defined, I can evaluate the efficiency of the cut-based selection procedure. Using simulations, I display this relative to the full MC sample content in figure 5.14. Note that I use a secondary y-axis for the QCD contribution, which would otherwise not be visible.

One can see that the samples containing $t\bar{t}$ pairs are reduced to the appropriate fraction of events by the single-lepton requirement: about 30 % of $t\bar{t}$ pairs decay to a final state with a single electron or muon. Due to the required momentum thresholds, the efficiency in signal is a bit higher than in SM $t\bar{t}$, as signal leptons usually have higher p_T . All other selection steps have high signal efficiency and mainly reduce the contributions of SM backgrounds. The S_T pre-selection is of note here, suppressing all SM backgrounds without any signal loss. The $W + \text{jets}$ background is reduced by the jet multiplicity and b -tagging requirements, and QCD is strongly affected by the 2D isolation requirement. Overall, about 50 % of signal

events with a single lepton remain, with higher efficiencies for spin- $\frac{3}{2}$ compared to spin- $\frac{1}{2}$ signals, and rising efficiency towards higher m_{t^*} . The SM background contributions are suppressed to below 1%.

For the following steps of the analysis procedure, it is crucial to ensure that all relevant variables are well modeled by the simulations of SM backgrounds. Therefore, I present some control distributions in figures 5.15, 5.16, 5.17, and 5.18, demonstrating that the description of all relevant objects is good. Importantly, all variables I analyze here are subject to statistical variation in data and simulations, which I show in the distributions. Beyond this statistical uncertainty, systematic uncertainties are relevant as well, for example originating from the SFs introduced above. I do not include these effects in the control distributions, thus the uncertainty shown is an underestimation. More uncertainties might arise from the procedures I utilize below, so I discuss all of them together in section 5.3.10, where I also show their effect.

In general, good description of all object types is visible, with some fluctuations due to low statistical power, especially in the smaller datasets of 2016. Muons are usually better described than electrons, whereas no major differences are visible between small-radius and variable-radius jets. Some notable features are visible: Turn-on like behavior in lepton momenta results from a jump in events when non-isolated leptons are allowed in the high- p_T region. A similar effect in small-radius jets results from the S_T threshold that is imposed. In general, jet momenta show a downward trend towards high values, a known feature which is weakened, but not removed entirely, by top p_T reweighting.

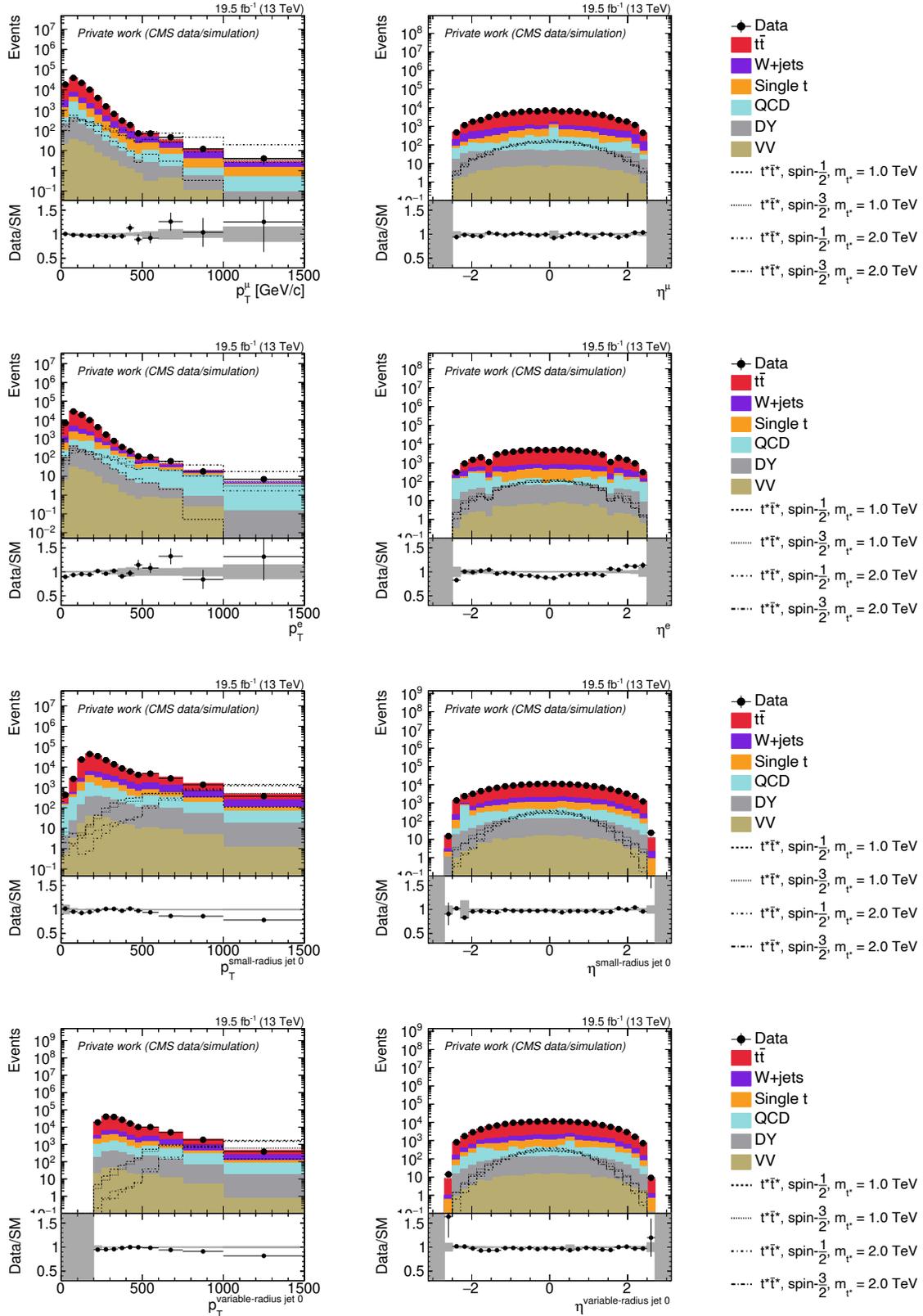


Figure 5.15: Control distributions of the 2016 pre-VFP era, showing p_T and η of the four different object types used in this analysis: muons (first row), electrons (second row), small-radius jets (third row) and variable-radius jets (fourth row). Only data and MC statistical uncertainties are shown and signal samples are scaled to 1 pb for visibility.

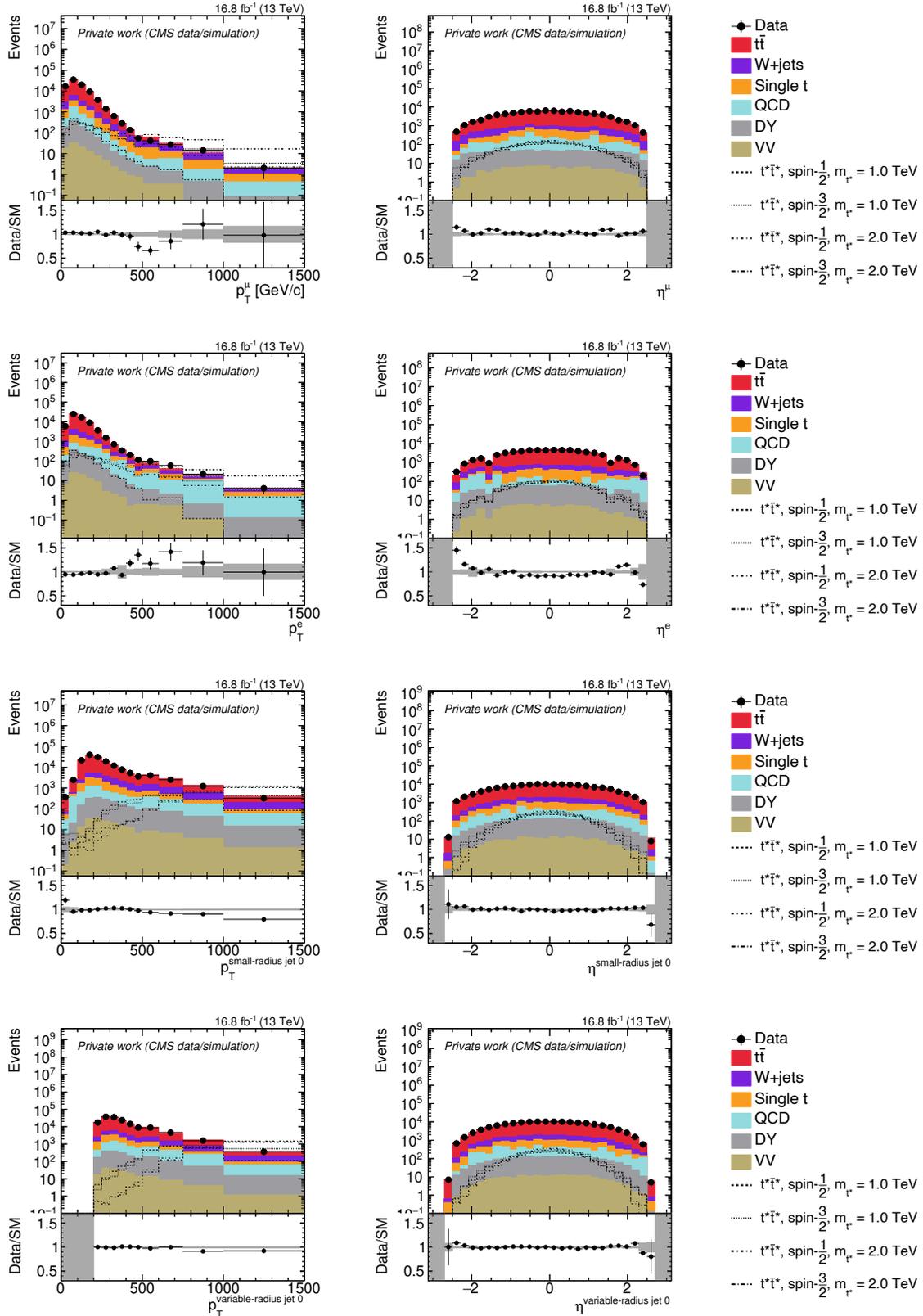


Figure 5.16: Control distributions of the 2016 post-VFP era, showing p_T and η of the four different object types used in this analysis: muons (first row), electrons (second row), small-radius jets (third row) and variable-radius jets (fourth row). Only data and MC statistical uncertainties are shown and signal samples are scaled to 1 pb for visibility.

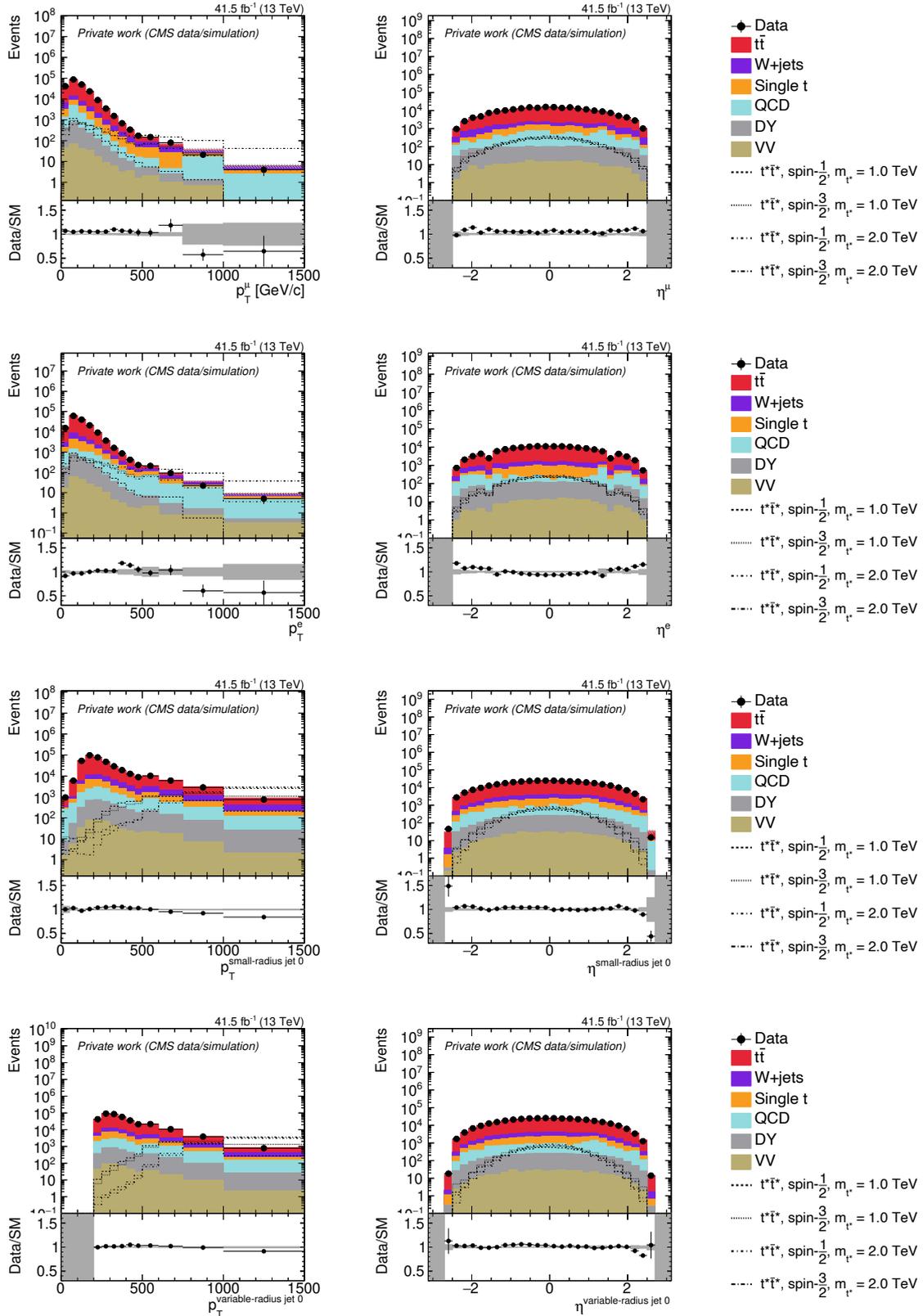


Figure 5.17: Control distributions of the 2017 era, showing p_T and η of the three four object types used in this analysis: muons (first row), electrons (second row), small-radius jets (third row) and variable-radius jets (fourth row). Only data and MC statistical uncertainties are shown and signal samples are scaled to 1 pb for visibility.

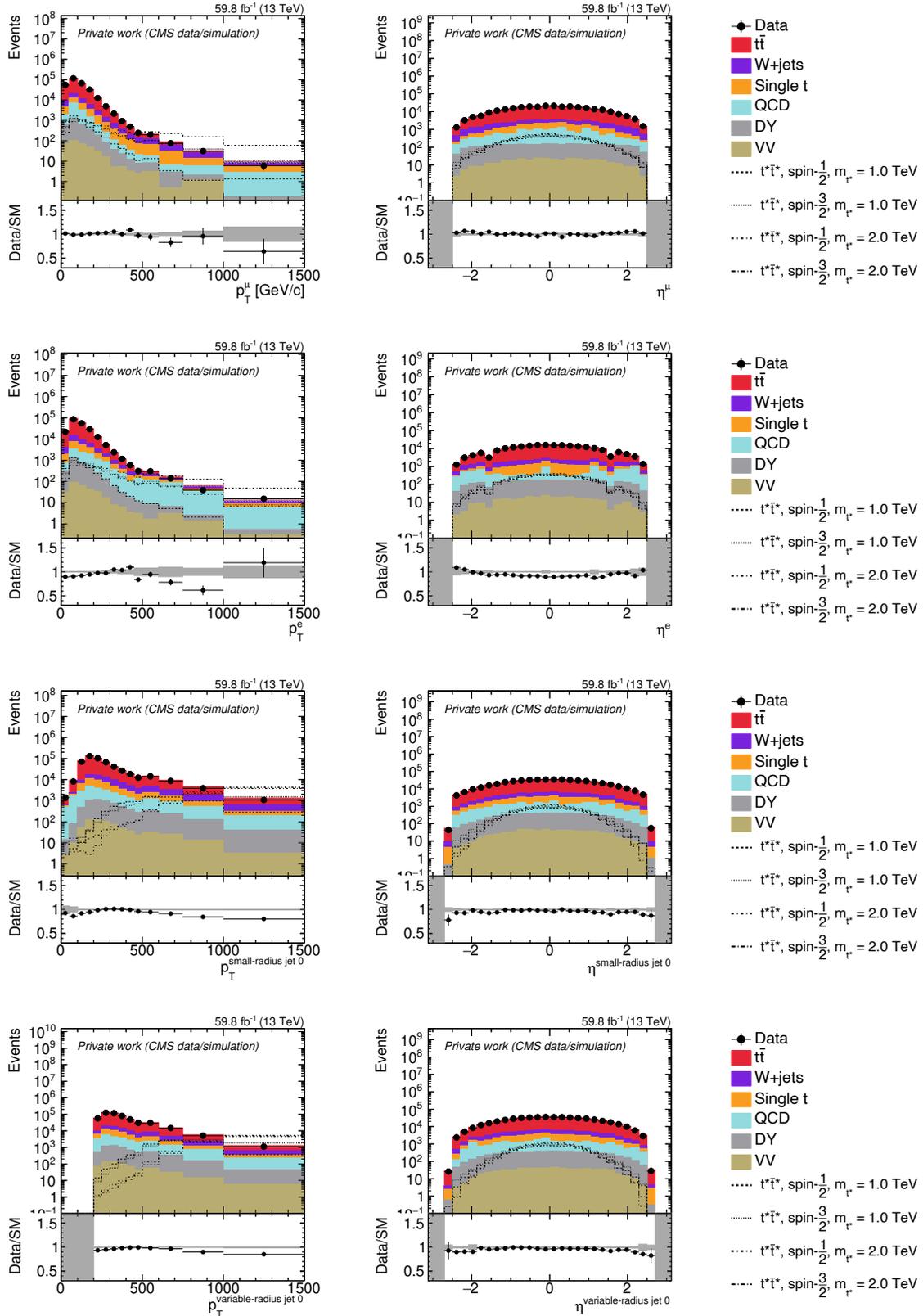


Figure 5.18: Control distributions of the 2018 era, showing p_T and η of the four different object types used in this analysis: muons (first row), electrons (second row), small-radius jets (third row) and variable-radius jets (fourth row). Only data and MC statistical uncertainties are shown and signal samples are scaled to 1 pb for visibility.

5.3.8 Deep neural network

Through the cut-based selection, I find a set of signal-like events. However, contributions from SM backgrounds are still large. To maximize the analysis sensitivity, it is crucial to analyze a subset of events where the contribution of SM background processes is low, and potential signal events are enriched. I define such a region using a DNN.

5.3.8.1 DNN overview

As the main SM contribution to the events remaining after the cut-based selection are $t\bar{t}$ events, I require the DNN to discriminate signal events from these. Therefore, I use a binary DNN, outputting a single score s_{DNN} , where high values denote signal-like events. Then, a threshold on that score can be used to define the SR. In the following, I first describe the DNN as I use it in the analysis, and after that outline the optimization procedure in which this DNN was obtained.

Input variables The DNN uses a set of 33 input variables, all of which are normalized to a mean of 0 with a standard deviation of 1 before being passed to the DNN for training or inference:

- 1 leading lepton: p_T , η , ϕ and I_{rel}
- 3 leading variable-radius jets: p_T , η , ϕ , τ_1 , τ_2 , τ_3 and N_{subjects}
- 1 b -jet (with highest DeepJet score): p_T , η , ϕ and DeepJet score
- p_T^{miss} and its ϕ direction
- $N(\text{small-radius jets})$, $N(\text{variable-radius jets})$

The majority of these variables are kinematic properties (p_T , η , ϕ) of the objects expected in signal-like events. With these, the DNN can learn the kinematic differences between t^*t^* and $t\bar{t}$ events I present in section 5.3.3, without any dedicated complex variables being passed as inputs. This is supported by the addition of jet substructure variables, which help the DNN to distinguish jets originating from Lorentz-boosted t quarks and gluon jets. Information about b -tagging and lepton isolation complete the set of input variables.

Some input variables cannot be properly defined for all events, as (for example) only a minimum of one variable-radius jet is required to exist in each event, but I use three for the DNN. Therefore, *zero-padding* is used, in which missing values are replaced by a value denoting a non-existent object. For p_T variables, this value is 0 GeV (as no objects with no momentum can exist), whereas for angular variables (η , ϕ) a value outside of the range of usual values is chosen, as a 0 here would have an actual physical meaning for non-empty objects.

DNN architecture The number of nodes in the first and last layer of the DNN are driven by the number of input (33) and output (1) variables. The hidden part of the DNN is optimized for sensitivity, consisting of four hidden layers with 25 nodes each. For all hidden nodes, I use a `tanh` activation function with initial weights following a random normal distribution. The output node instead uses a `sigmoid` activation function. For training, I utilize the RMSprop optimizer with a learning rate of 10^{-5} on a batch size of 4096 events, with an MSE loss function.

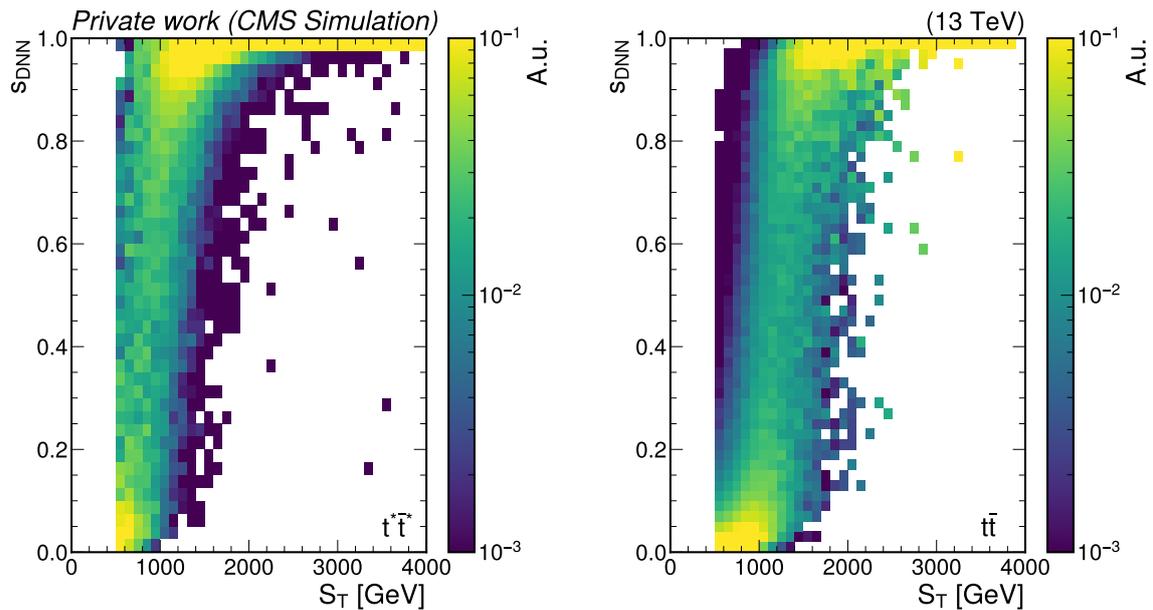


Figure 5.19: Two-dimensional distribution of S_T against s_{DNN} of a DNN with no active decorrelation, for t^*t^* (left) and $t\bar{t}$ (right) events. The content of each S_T bin is normalized to unity.

Decorrelation using input weights Several of the 33 input variables are highly correlated with S_T , for example the different individual object momenta. Therefore, training the DNN with the input variables and architecture described above leads to the DNN finding a loss minimum which just parametrizes an S_T threshold. I show this in figure 5.19, where a two-dimensional distribution of S_T against s_{DNN} is visible, comparing t^*t^* and $t\bar{t}$ events. Events with high S_T are classified as signal-like (high s_{DNN}) and events with low S_T score are classified as background-like (low s_{DNN}), independent of whether they are t^*t^* or $t\bar{t}$ events.

This is problematic, as the DNN should only use any information available beyond S_T to discriminate between t^*t^* and $t\bar{t}$, so that a threshold on s_{DNN} introduces minimal sculpting onto the S_T distribution. To address this, I remove the S_T information from DNN training. This can not be done by simply omitting the p_T inputs, as these are still important for the DNN to learn information on p_T relations between objects. Instead, I adjust the weight of events for training so that the S_T distribution of signal and background events is identical, as I show in figure 5.20.

DNN training With the reweighting strategy applied, I train the DNN. The signal class is composed from a combination of all available spin- $\frac{1}{2}$ signal samples after the cut-based selection, whereas the background class uses all $t\bar{t}$ events at that stage. All signal mass points contribute the same to the training, except for acceptance effects in the cut-based selection. Due to the S_T reweighting, it is ensured that the overall weight of the signal and background class is the same. In total, I use 851 850 simulated events, of which I exclude 20% from training to use as a test sample. Of the remaining 80% of events, I use 75% for training and 25% as validation set. The network is trained for 1000 epochs, and I track the loss and accuracy during the process, as shown in figure 5.21.

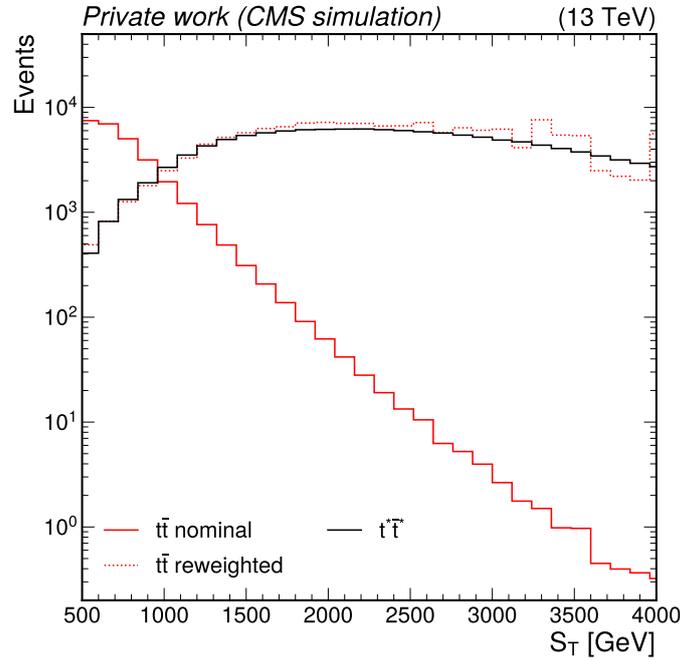


Figure 5.20: S_T distribution for $t\bar{t}$ before and after the reweighting procedure, compared to the t^*t^* S_T distribution.

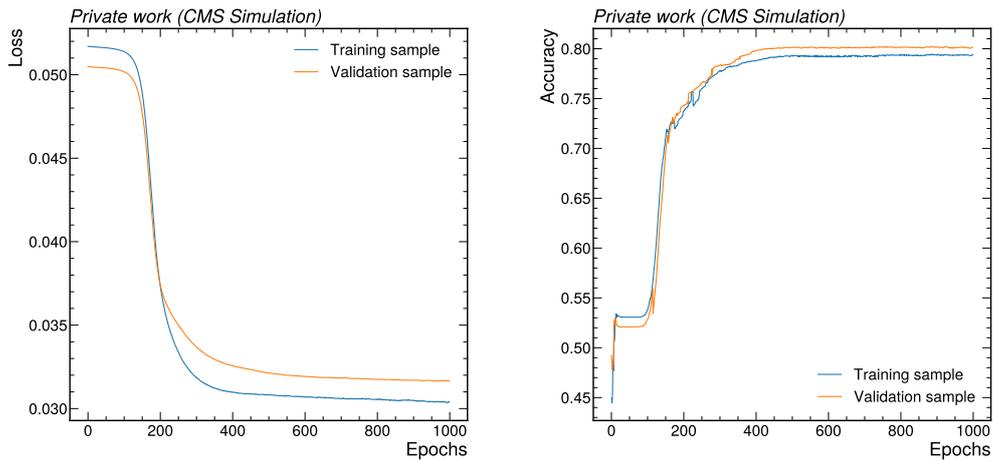


Figure 5.21: Loss function and accuracy of the neural network during training. Jumps in accuracy result from single events with high weights passing the 0.5 threshold used to calculate the accuracy.

As the loss (accuracy) of the validation sample does not increase (decrease) during training, no visible overtraining occurs. Still, a version of early stopping is used: when selecting the final DNN for the analysis, the DNN weights at the point with highest validation accuracy are used. Additionally, I compare the s_{DNN} distributions on training and validation samples to explicitly test for overtraining. After ensuring that the DNN is not overtrained, I introduce it into the analysis for inference and evaluate it on all simulated samples and data events in the regular weighting scheme. I show the resulting DNN score distribution in figure 5.22. There, one can clearly see good the performance of the DNN. Background events of all SM processes accumulate at low values of s_{DNN} , whereas signal events have their maximum at high values. Notable, the DNN does not use the full range of possible s_{DNN} values. This feature is introduced when adding the decorrelation procedure, but does not result in a degradation of the DNN performance. It can be seen that the data distribution of s_{DNN} is well described by the simulated samples. This is a crucial test in order to ensure that the DNN does not show unexpected behavior due to differences between data and simulation.

For this network, the DNN output for $t^*\bar{t}^*$ and $t\bar{t}$ events can then again be displayed against S_T , which I show in figure 5.23. Here, one can see that the network is now able to learn a difference between the two processes beyond a simple S_T threshold: for all S_T , the full range of possible DNN score values is utilized. Still, some remaining S_T -dependency is visible by eye: the change in training weights has enabled the DNN to actually learn something beyond an S_T threshold, but has not fully decorrelated its output from S_T .

To further understand the DNN performance, I evaluate SHAP importance scores, as shown in figure 5.24. As expected, one can see a high relevance of jet multiplicity variables, but also jet substructure variables rank highly, showing that the DNN is using this information. The third leading variable-radius jet has high importance, as this is the first jet multiplicity number where signal and background differ. Among the lower ranking variables are primarily angular variables, specifically ϕ coordinates of objects.

5.3.8.2 Designing decorrelated taggers

Despite testing different reweighting schemes, all of them lead to non-negligible shaping of the background S_T distribution. A very similar problem of decorrelating a (ML-based) classifier from one or more parameters is given when developing jet tagging algorithms. Discriminating, for example, jets originating from the decay of a t quark from QCD jets, large differences are present in the p_T and jet mass distributions. Different techniques are used to avoid sculpting of these variables when identifying jet flavor. These include the training of an adversarial neural network, as in the case of DeepAK8s mass-decorrelated version [233] or methods like DisCo regularization [234]. Another option is the “designing decorrelated taggers” (DDT) technique, as described in [235]. Given that the decorrelation of the event classification DNN from S_T is similar to the decorrelation of jets from p_T , I use DDT for this analysis, and introduce it in the following. Instead of requiring a fixed threshold on s_{DNN} , DDT allows that threshold to vary depending on one or multiple variables the DNN should be decorrelated from. The variable threshold is chosen in a way that the background efficiency of the DNN is the same for all values of the variable(s). Conceptually, introducing a variable threshold on s_{DNN} is the same

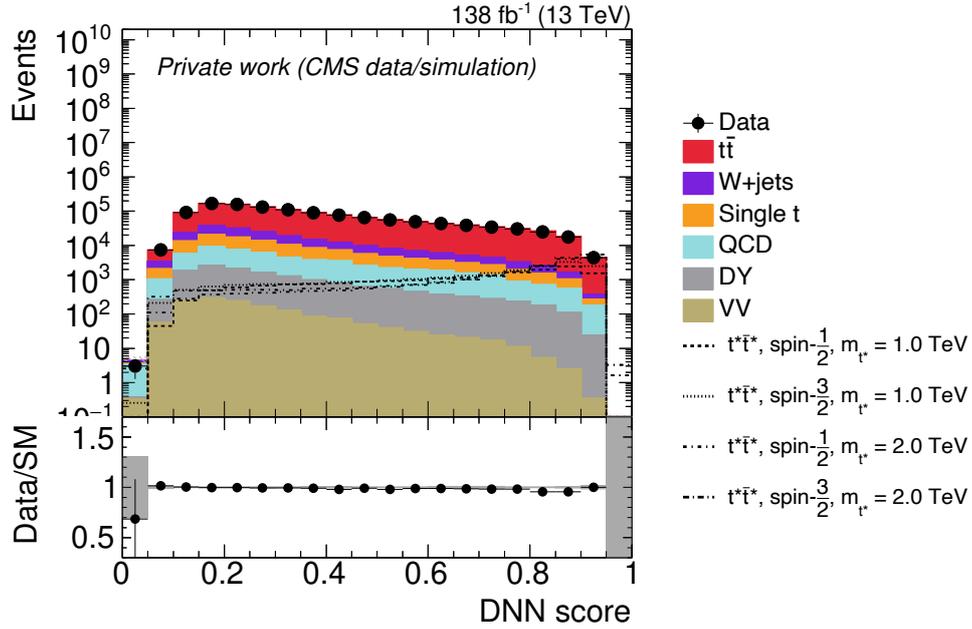


Figure 5.22: DNN output score for data, SM backgrounds and representative signal samples. Only data and MC statistical uncertainties are shown.

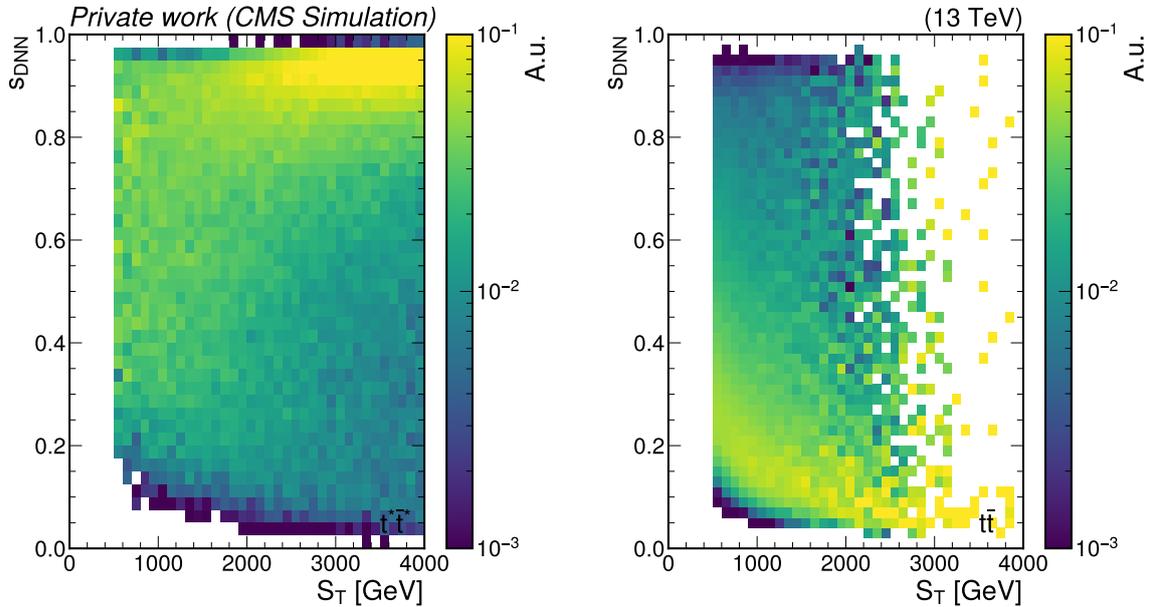


Figure 5.23: Two-dimensional distribution of S_T against s_{DNN} of a DNN with S_T reweighting applied, for $t^*\bar{t}^*$ (left) and $t\bar{t}$ (right) events. The content of each S_T bin is normalized to unity.



Figure 5.24: Feature importance scores of all DNN input variables, measured using mean absolute SHAP values.

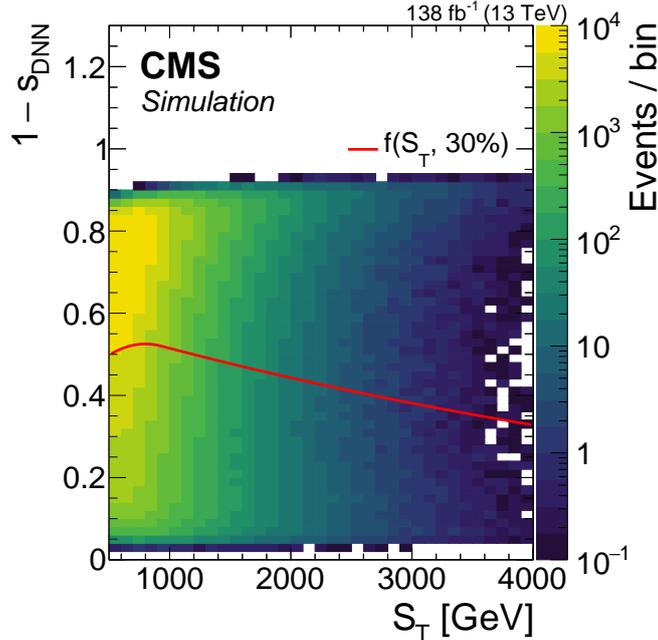


Figure 5.25: Two-dimensional distribution of S_T against $1 - s_{\text{DNN}}$ for $t\bar{t}$ events, combining the full dataset as well as electron and muon events. The red line shows the function $f(S_T, 30\%)$ used to define s_{DDT} and thus the SR.

as shifting s_{DNN} with some function f , which depends on the variable(s) s_{DNN} is to be decorrelated from, and then imposing a constant threshold. My goal is to decorrelate s_{DNN} from S_T , focusing on $t\bar{t}$ events. Thus, I define a new variable s_{DDT} as

$$s_{\text{DDT}} = s_{\text{DNN}} - f(S_T, \epsilon_{t\bar{t}}). \quad (5.12)$$

The function $f(S_T, \epsilon_{t\bar{t}})$ depends on S_T and a desired constant $t\bar{t}$ efficiency $\epsilon_{t\bar{t}}$, which is the tunable parameter of the method. To obtain the function $f(S_T, \epsilon_{t\bar{t}})$, I analyze a two-dimensional distribution of s_{DNN} against S_T , shown in figure 5.25, for $t\bar{t}$ events. I take the total number of events in some individual S_T bin of that distribution. Then, starting from the bottom of that bin, the content of s_{DNN} bins are summed up until $\epsilon_{t\bar{t}}$ of the total event number is reached. The s_{DNN} value where this occurs is the threshold one would need to impose on s_{DNN} to achieve a $\epsilon_{t\bar{t}}$ $t\bar{t}$ efficiency in that S_T bin. I chose a working point of $\epsilon_{t\bar{t}} = 30\%$ and perform the procedure for all bins of S_T to obtain a set of binned threshold values. To describe these by a continuous function $f(S_T, \epsilon_{t\bar{t}})$, I perform a fit to these thresholds with a mathematical function. The fit function that offers the best description of the points is a so-called *crystal-ball function*, which describes the turn-on behavior with a Gaussian core, and the tail with a power-law function. To help the fit converge, I assign an error to each of the points as \sqrt{N}/N , where N is the number of events in the considered S_T bin. Note that the choice of this error has no further physical meaning, it just ensures that S_T bins with more data have a stronger influence on the fit than S_T bins with less data. Figure 5.25 shows the resulting function $f(S_T, 30\%)$. Using that function, s_{DDT} can be calculated for all events. I show the results in figure 5.26, splitting all events into two bins. By design, 30% of $t\bar{t}$ events are found at $s_{\text{DDT}} > 0$, and imposing this threshold on all events results in a signal-enriched region while introducing very

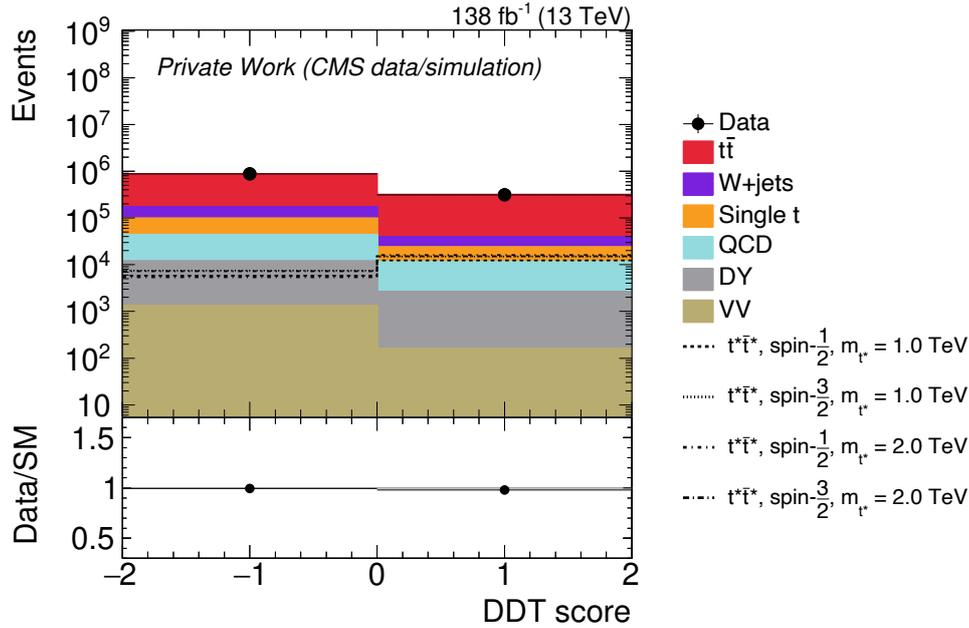


Figure 5.26: s_{DDT} for $\epsilon_{t\bar{t}} = 30\%$, showing which events enter the SR at $s_{\text{DDT}} > 0$. Evaluating data, SM backgrounds and signal samples, only data and MC statistical uncertainties are shown.

little sculpting onto S_T . This region is used in the statistical analysis of this search and is referred to as *signal region* (SR) in the following.

5.3.8.3 DNN optimization

The DNN configuration I presented above is obtained in a dedicated optimization process, which I summarize here. This process was performed during development of the analysis, which underwent some changes afterwards. While the DNN was retrained whenever changes to the analysis made this necessary (such as for changes of S_T threshold, jet calibrations or b -tagging strategy), the optimization procedure itself was not repeated. It was ensured that updating the DNN training due to analysis changes does not lead to changes in the overall sensitivity and final analysis results.

Ultimately, the best network is the one providing the best possible sensitivity in discovering or excluding the existence of $t^*\bar{t}^*$ pair production. However, as after the DNN application other steps that are dependent on it follow (the DDT method, data-driven background estimation and statistical analysis, the latter two I introduce below) it would not be feasible to perform all these steps for each possible network. Therefore, ROC curves and AUC scores are used as the metric to select the best performing DNN instead of, for example, exclusion limits.

Multiple DNN models might achieve similar performance, with differences smaller than the random fluctuations a single model can have for two different training iterations. These fluctuations arise from the inherent statistical nature of DNN training, and are emphasized here due to large event weights from the S_T reweighting. Therefore, below, I show the ROC curve of the optimized network as an envelope of five different training iterations with five different (fully random, non-exclusive) data splittings. Other DNN configurations are compared to this envelope.

I make two assumptions for the optimization procedure:

- The optimization of the DDT method is independent from the DNN optimization as long as the DNN does learn something beyond a simple S_T threshold.
- All hyperparameters are independent within the training uncertainties when all are close to their optimum.

It would not be feasible to optimize the DDT method for each network. Therefore, once a model with good performance is found, the DDT method is tested on it at a fixed DDT working point and the model is discarded in case DDT fails. An example of a network that would be dropped through the DDT test is the fully correlated network in figure 5.19, which results in an excellent AUC (0.99), but it is not possible to decorrelate it using DDT. I find that DDT is very stable against the choice of its working point, thus testing at a fixed working point is reasonable.

I start the optimization procedure by finding a well-performing and decently decorrelated model with a trial and error approach (taking into account past DNN studies for this analysis). This model serves as a starting point, from which individual parameters are systematically varied and an optimal model is found. I adjust the following hyperparameters during optimization:

- Number of hidden layers (between 2 and 5).
- Number of nodes per hidden layer (between 15 and 35).
- Training batch size (between 1024 and 8192).
- Learning rate (between 10^{-6} and 10^{-4}).
- Dropout percentage (off, or between 20% and 60%)
- Batch normalization (on or off)
- Different loss functions (MSE and BCE)
- Different hidden layer activation functions (`tanh`, `sigmoid`, `ReLU`, `LeakyReLU`)

In addition to these hyperparameters, I also investigate variants of altering the DNN input variables:

- Using three additional high-level input variables.
- Including an additional variable-radius jet.
- Passing the jet mass m_{SD} for variable-radius jets.

The baseline variables are chosen based on the expected behavior of the considered processes, as discussed above. I test additional high-level input variables that are chosen as they individually show sensitivity to discriminate the $t^*\bar{t}^*$ and $t\bar{t}$ processes: The momentum asymmetry between the leading and fourth small-radius jet in each event, as well as the ΔR between the two leading variable-radius jets or the lepton and closest variable-radius jet are evaluated. I also vary the number of variable-radius jets and add the jet mass (defined using the soft drop algorithm [164]). Additionally, I test whether the neural network benefits from an imbalance in training samples, training with twice the amount of background compared to signal, and vice versa.

The best network I describe above is found with this optimization procedure. Figure 5.27 shows the ROC curve band of this network and compares it to the ROC curves obtained by the other studied networks. All ROC curves of networks showing strong

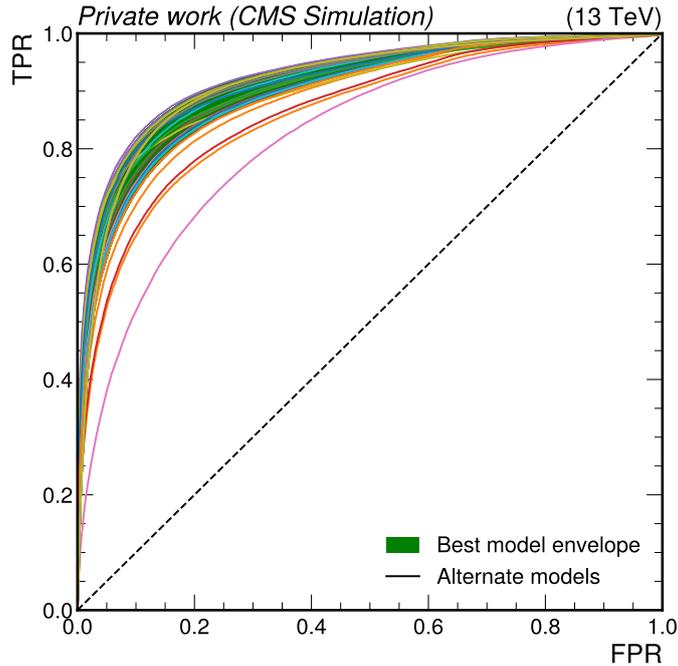


Figure 5.27: ROC curve band of the best performing DNN model, compared to ROC curves of various other models, including different variations of inputs and hyperparameters.

correlation, thus failing the DDT fit, are excluded here. From these results, it is clear that with the chosen model, a well-performing DNN is found. Also, it is shown that there are other models that offer similar performance: the DNN is not very sensitive to changes in the input variables or hyperparameters.

5.3.8.4 Training sample selection

The choice of training data could have an effect on the optimization results. To reduce the computational requirements, I train the DNN on data from a single era (2018 for the final DNN, 2016 post-VFP for the optimization, for technical reasons). To ensure that training on a specific era does not change the result, I train the final DNN configuration on all four eras of Run 2. Then, I compare the performance of the resulting DNN models when evaluated on all eras. Figure 5.28 shows the resulting ROC curves, where an uncertainty band is obtained from the standard deviation around a mean of five different trainings. It can be seen that the results are very comparable and independent of the year used for evaluation. If the DNN would be biased by training on a specific era, it would perform best when evaluating on the same era used for training, which is not the case. Thus, the DNN can be trained on a single era without any bias. I perform some more related tests: When training a DNN to identify events of a signal process that contains a particle of unknown mass (the t^* in this case), it's crucial to ensure that the DNN is not biased to the set of simulated masses it has seen during training. I train a DNN with a single t^* sample excluded and then test the performance on this specific sample. No changes compared to a DNN that does include $t^*\bar{t}^*$ events of that mass during training are observed. Thus, the DNN can also well identify $t^*\bar{t}^*$ events with t^* masses it has not seen during training.

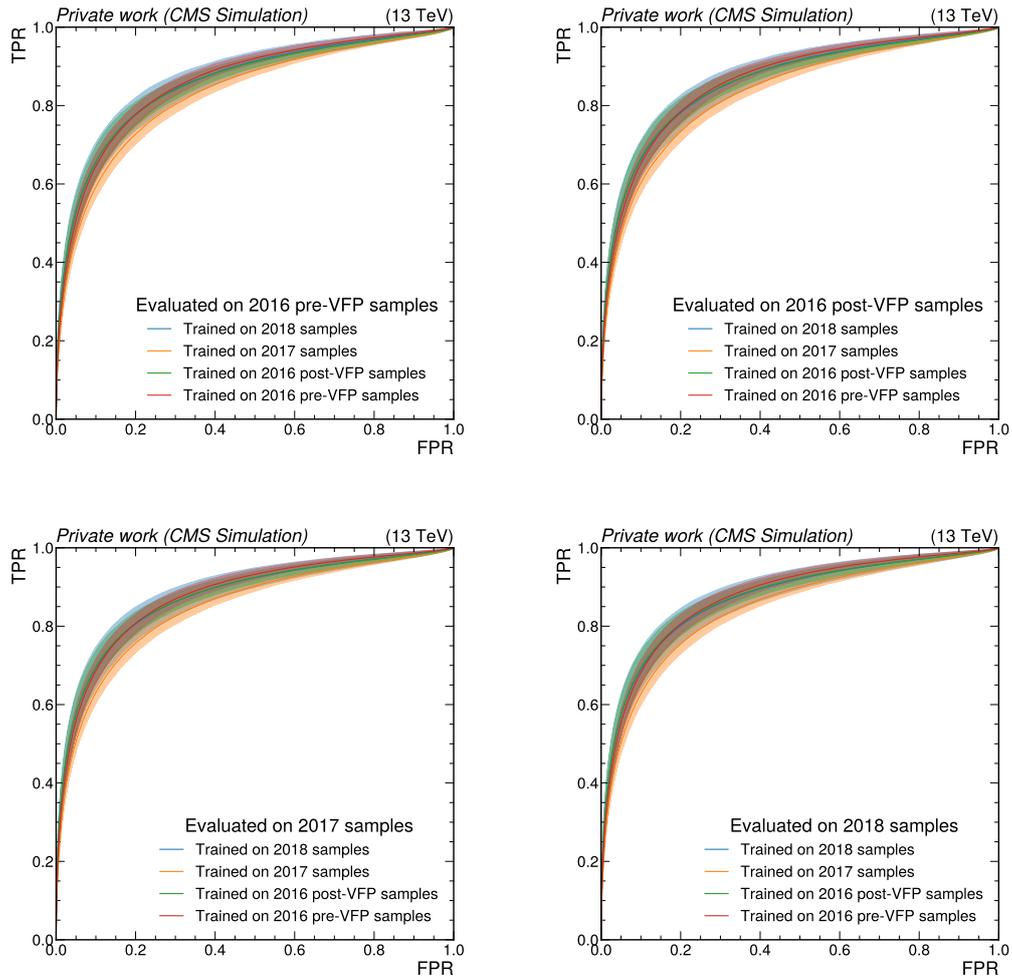


Figure 5.28: ROC curves of DNNs trained on samples from different eras, each evaluated on all four eras. An uncertainty band is obtained from multiple repeated training iterations.

5.3.8.5 Decorrelation optimization

In addition to the optimization of the DNN itself, I also perform studies on the optimal decorrelation. For the input reweighting, I compare the approach based on the signal shape to one where both $t^*\bar{t}^*$ and $t\bar{t}$ samples are weighted to be flat in S_T . Due to very large weights at the edges of the considered S_T range, this leads to unstable training results, and does not yield a better performance.

For DDT, I test the method for different values of the efficiency $\epsilon_{t\bar{t}}$ in increments of 5%. All resulting signal regions are passed through a statistical analysis similar to the one I describe below, but without considering systematic uncertainties. I compare sensitivity through expected exclusion limits on the existence of a t^* , and find optimal results for $\epsilon_{t\bar{t}}$ around 30%. Notably, the performance is stable in a wide range around that value, as long as no extreme choices of $\epsilon_{t\bar{t}}$ are made. Similarly, values around $\epsilon_{t\bar{t}} = 30\%$ result in low remaining sculpting of the S_T distribution. Overall, these results show that the procedure is very stable against the choice of $\epsilon_{t\bar{t}}$.

5.3.9 Estimation of SM backgrounds

In order to make a statement about the potential existence of $t^*\bar{t}^*$ pair production, precise knowledge of the SM processes entering the SR is required. For this, I use simulations for backgrounds with t quarks, and a data-driven approach for non- t backgrounds, which I detail in this section.

5.3.9.1 Analysis regions

As signal might be present in the SR, the description of SM processes can not be tested there. Instead, I define other analysis regions for this purpose, an overview of which I give in table 5.3.

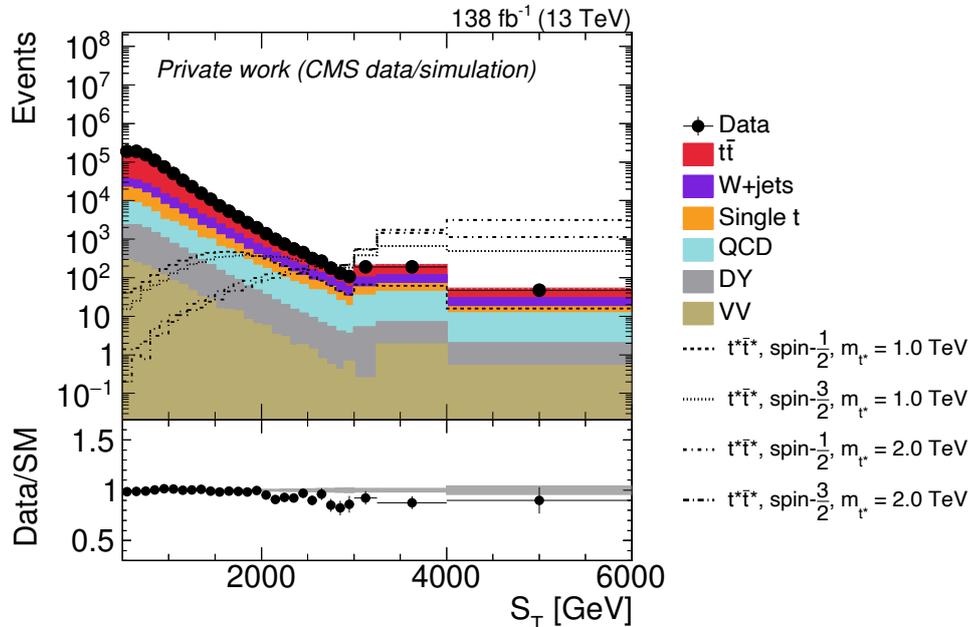
Low DDT validation region Estimating the performance of SM predictions in the SR is crucial, which motivates the creation of a region that is as similar as possible, but signal-suppressed. To do so, I keep all event selection requirements the same as in the SR, but invert the s_{DDT} requirement: all events with $s_{\text{DDT}} < 0$ belong to the *validation region* (VR). Thanks to the decorrelation techniques that I describe above, it is ensured that the DNN induces minimal S_T shape differences between the SR and VR, thus making the VR well suited to validate the agreement of data and SM background predictions. The S_T distribution in the VR is shown in figure 5.29. Comparing the data to predictions using simulated SM background samples in this region, a decent but not perfect agreement is observed.

$t\bar{t}$ validation region To specifically investigate the description of the $t\bar{t}$ process, I use a subset of the VR region: the $t\bar{t}$ VR. In addition to all other requirements of the VR listed above, I impose two more selection criteria:

- At least one variable-radius jet must fulfill the HOTVR t -tagging criterion.
- At least two b -jets must be present.

Region	Definition	Usage
Signal region (SR)	All cut-based selection steps and DDT score ≥ 0 .	Used in the statistical analysis presented in section 5.4.1.
Low DDT validation region (VR)	All cut-based selection steps and DDT score < 0 .	Not included in the statistical analysis, used to validate the treatment of SM backgrounds in this analysis.
$t\bar{t}$ validation region ($t\bar{t}$ VR)	Subset of VR, adding t -tagging and b -tagging requirements.	Not included in the statistical analysis, used to validate the description of $t\bar{t}$ events in this analysis.
b -veto control region (CR)	All cut-based selection steps, but inverting b -tagging requirements. Inclusive in DDT score.	Data from this region is used to estimate non- t backgrounds in the SR and VR.

Table 5.3: Overview on the regions used in this analysis.


 Figure 5.29: S_T distribution, combining electron and muon events, in the VR for the full Run 2 dataset, using simulations to predict the contributions of SM processes. Only data and MC statistical uncertainties are shown.

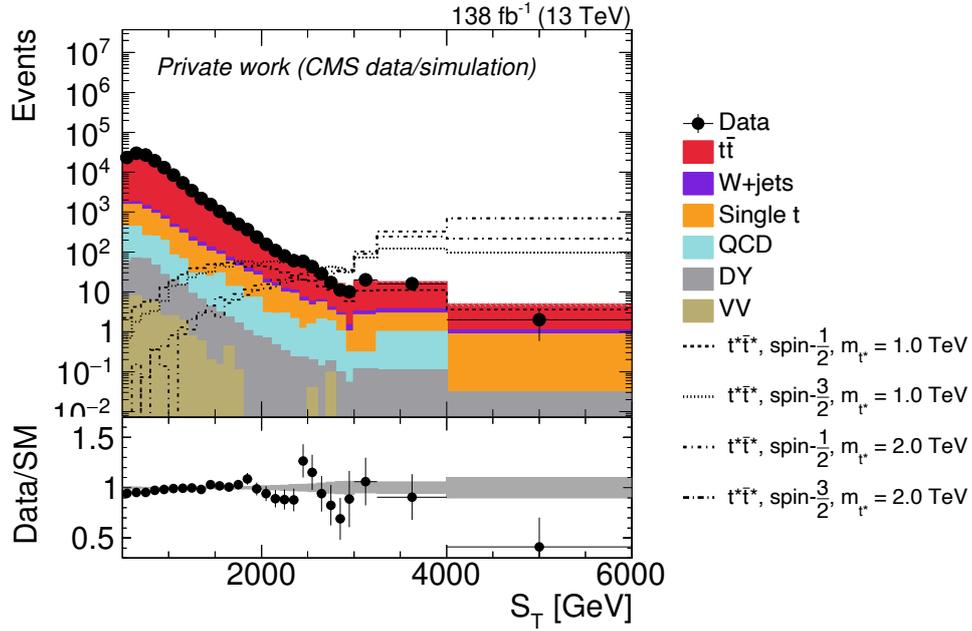


Figure 5.30: S_T distribution, combining electron and muon events, in the $t\bar{t}$ VR for the full Run 2 dataset, using simulations to predict the contributions of SM processes. Only data and MC statistical uncertainties are shown.

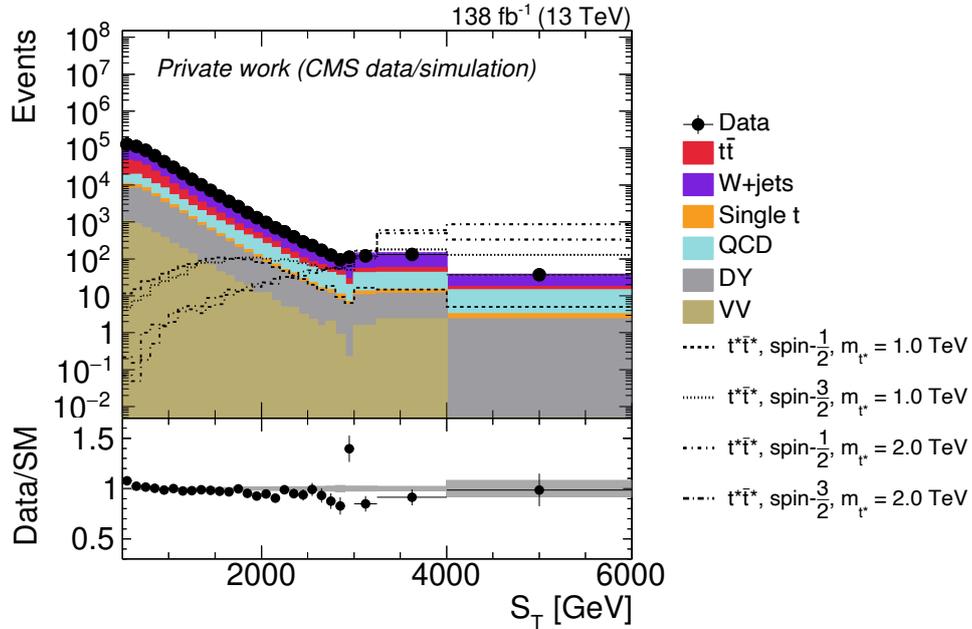


Figure 5.31: S_T distribution, combining electron and muon events, in the CR for the full Run 2 dataset, using simulations to predict the contributions of SM processes.

These additional requirements reduce the number of events dramatically when comparing to the VR, however the selection is now almost pure in $t\bar{t}$ events, as I show in figure 5.30. These results show that the description of the $t\bar{t}$ process by the simulated samples is very good.

***b*-veto control region** Both regions I present above are dominated by $t\bar{t}$ events. To study non- t backgrounds, I define another region enriched in these processes by rejecting events containing b -jets. The S_T distribution in this *b*-veto control region (CR), is shown in figure 5.31. No requirement on s_{DDT} is imposed for this region.

5.3.9.2 Data-driven estimation of non-top backgrounds

Backgrounds without t quarks, and especially QCD processes, are usually not well described in simulation. Addressing this with SFs is possible, but might not yield perfect results and comes with large systematic uncertainties. In order to more accurately describe the contribution of non- t backgrounds and reduce the systematic uncertainties associated with them, I estimate their contribution to the SR using a data-driven method. First, I calculate a bin-by-bin ratio α between events in the SR and CR based on simulation as

$$\alpha = \frac{N_{\text{SR}}(\text{non-}t \text{ bkg.})}{N_{\text{CR}}(\text{non-}t \text{ bkg.})} \quad (5.13)$$

where $N_{\text{SR}}(\text{non-}t \text{ bkg.})$ is the MC predicted number of non- t simulated events in a bin of the SR S_T distribution and similarly $N_{\text{CR}}(\text{non-}t \text{ bkg.})$ is the MC predicted number of non- t simulated background events in the same S_T bin of the CR. I derive this ratio separately for the electron and muon channel, as composition of non- t backgrounds is different between these. The resulting α -ratios are shown in figure 5.32. For each of them, I obtain a continuous transfer function $g_{\text{TF}}(S_T)$ through a fit to the α -ratio. This removes the influence of (potentially) too few simulated events in some S_T bins. To estimate and reduce the bias from the choice of mathematical fitting function, I find chose two different functions that fit the data well. One of these is a Landau function [236, 237], parametrizing a maximum with a broad tail using three free parameters. As a second function, I use a Gaussian distribution plus a constant offset, which has four free parameters. I show the results of fits with these functions as dashed and dotted lines in figure 5.32. From both functions, the mean is calculated, which is the main transfer function $g_{\text{TF}}(S_T)$. Multiplying the data in the CR by this function, after subtracting remaining contamination of t backgrounds in the CR using simulation information, results in a description of the non- t background contribution in the SR:

$$N_{\text{SR}}(\text{non-}t \text{ bkg.}) = g_{\text{TF}}(S_T) (N_{\text{CR}}(\text{data}) - N_{\text{CR}}(t \text{ bkg.})), \quad (5.14)$$

While the method relies on simulated information to derive the α -ratio, I expect that most systematic errors cancel out in the ratio, allowing a precise prediction of the non- t backgrounds in the SR. To validate the procedure, I repeat it in the same way in the VR, for which the α -ratios and fit functions are shown in figure 5.33. I show the S_T distribution in the VR, using the data-driven method to estimate non- t backgrounds, in figure 5.34. There, it can be seen that the data-driven background estimation works as intended.

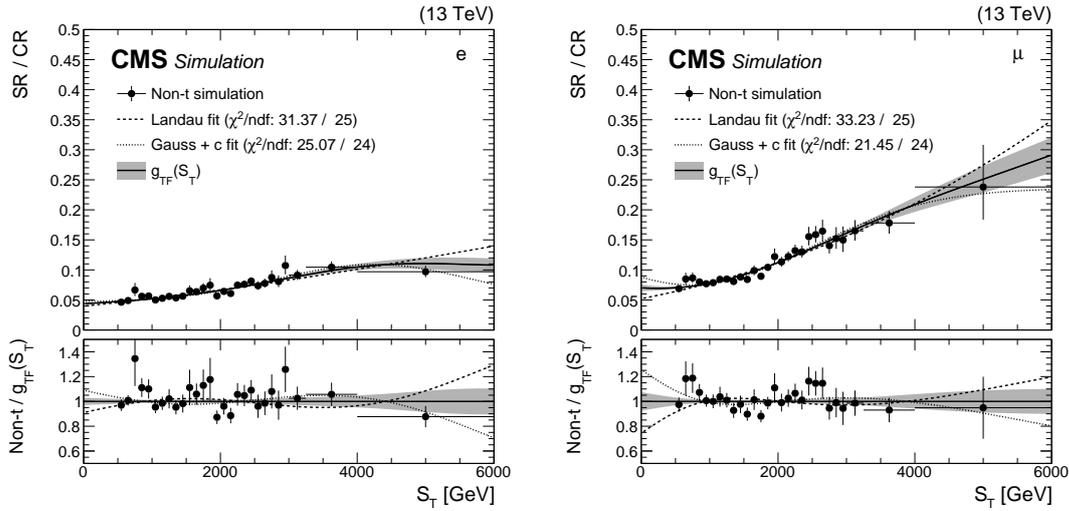


Figure 5.32: Background extrapolation procedure for the SR in the electron (left) and muon (right) channels. Black points in the upper part of the figures show the simulation-based α -ratios, which are fit with two functions shown as dashed lines. The mean of these is the final extrapolation function, showed as a black line with gray error band from the fit statistical uncertainty. The bottom part of the figure shows all entries divided by the mean fit function. These figures are published in [1].

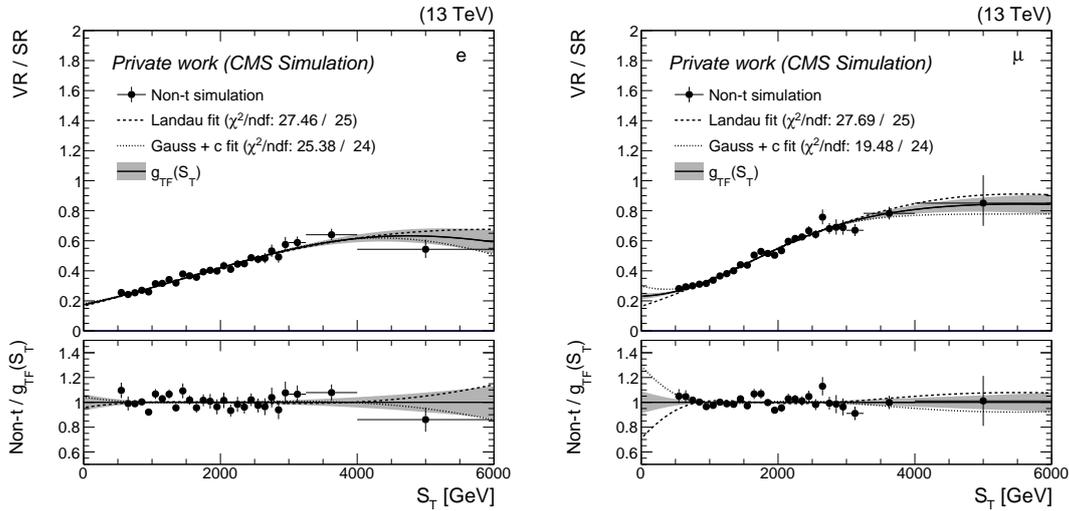


Figure 5.33: Background extrapolation procedure for the VR in the electron (left) and muon (right) channels. Black points in the upper part of the figures show the simulation-based α -ratios, which are fit with two functions shown as dashed lines. The mean of these is the final extrapolation function, showed as a black line with gray error band from the fit statistical uncertainty. The bottom part of the figure shows all entries divided by the mean fit function.

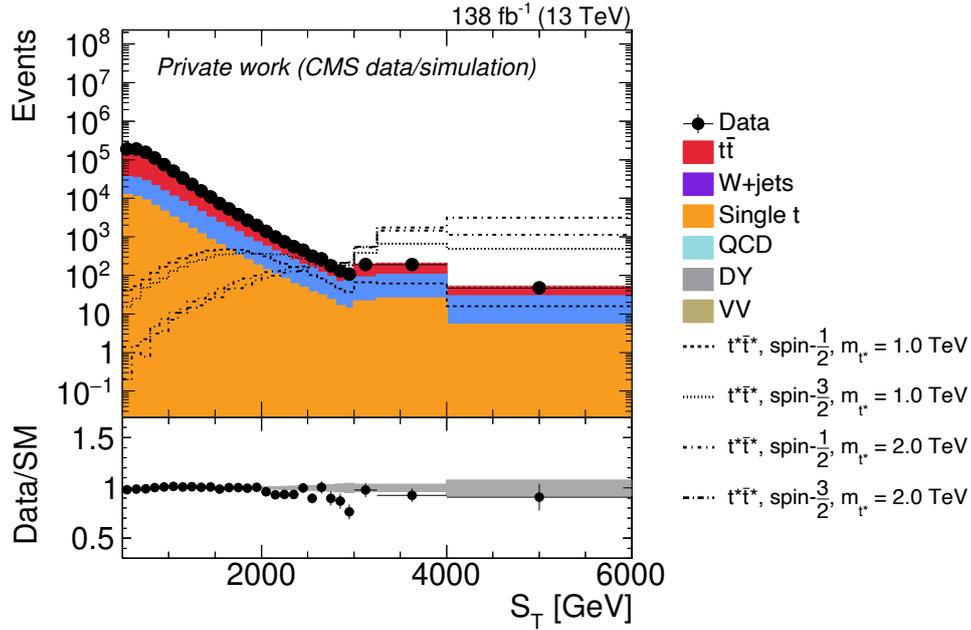


Figure 5.34: S_T distribution, combining electron and muon events, in the VR for the full Run 2 dataset, using simulations to predict top backgrounds, and data from the CR to predict non-top background contributions.

5.3.10 Systematic uncertainties

Several sources of systematic uncertainty affect this analysis. In table 5.4, I give an overview on all sources I consider, dividing them into uncertainties of experimental and theoretical nature. These uncertainties influence the S_T distribution of some or all processes considered in the analysis. For the experimental uncertainties, I show results on the size of the variation in figure 5.36 for a representative process. Similarly, I show the influence of theoretical uncertainties in figure 5.35. Below, I explain all sources in detail. Unless otherwise noted, uncertainties affect the muon and electron channels simultaneously.

5.3.10.1 Experimental uncertainties

Luminosity As I introduced in section 3.2.2.1, the luminosity of data collected with CMS is measured with several dedicated systems. The resulting integrated luminosity value of 137.62 fb^{-1} is associated with a total uncertainty of 1.6% [146, 238, 239]. This uncertainty affects all processes considered in the analysis, and is correlated between eras of data-taking.

JEC & JER In section 5.3.4.2, JEC and JER are introduced in order to correct for differences between data and simulation affecting jet energies. These corrections have uncertainties associated with them, which I independently vary within $\pm 1\sigma$. Importantly, different to most other uncertainty variations, these corrections do not affect event weights, but change jet properties (like their energy). I consider effects on small-radius and variable-radius jets, as well as p_T^{miss} , simultaneously and propagate it to S_T . For the JEC variations, due to limited statistical power in the simulated samples, strong artificial fluctuations become visible. These would lead

Uncertainty source	Affected samples	Era correlation
Luminosity	Signal, $t\bar{t}$, ST	Fully correlated
JEC	Signal, $t\bar{t}$, ST	Not correlated
JER	Signal, $t\bar{t}$, ST	Not correlated
b -tagging SFs	All samples	See below
b -tagging yield corrections	Signal, $t\bar{t}$, ST	See below
Prefiring weights	Signal, $t\bar{t}$, ST	Not correlated
PU weights	Signal, $t\bar{t}$, ST	Fully correlated
Lepton ID, isolation & trigger	Signal, $t\bar{t}$, ST	Not correlated
Decorrelation	$t\bar{t}$, ST	Fully correlated
Background estimation function choice	Non-top backgrounds	Fully correlated
Background estimation fit statistics	Non-top backgrounds	Fully correlated
PDF choice	Signal, $t\bar{t}$, ST	Fully correlated
μ_r and μ_f variations	Signal, $t\bar{t}$, ST	Fully correlated
Top p_T reweighting	Non-top backgrounds	Fully correlated

Table 5.4: Overview on the systematic uncertainties considered in this analysis, split into experimental sources (top part) and theoretical sources (bottom part).

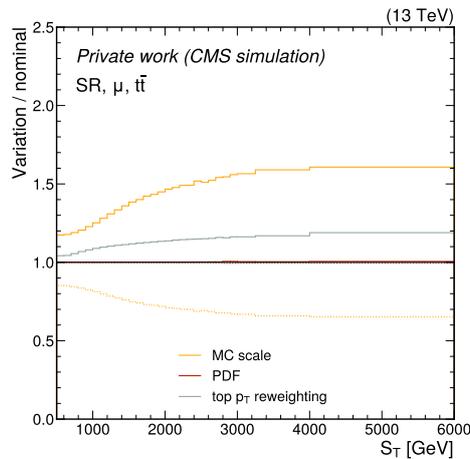
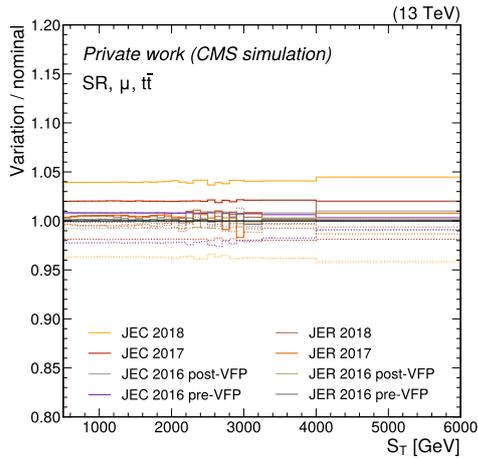
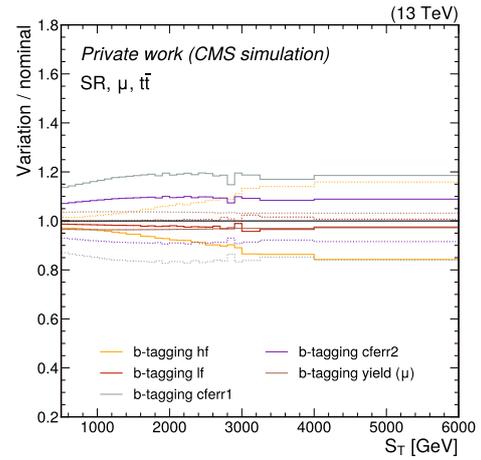
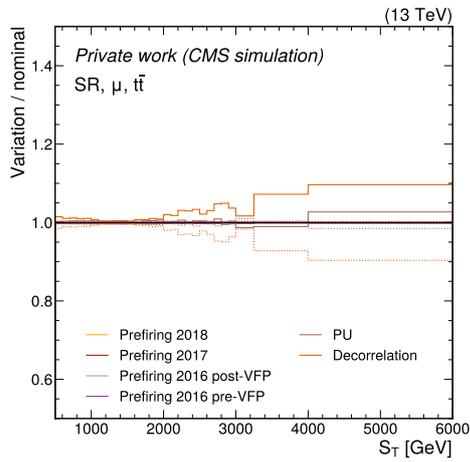


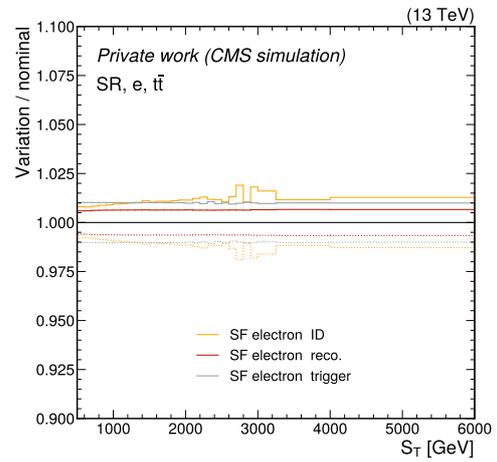
Figure 5.35: Variations of S_T for different theoretical uncertainties, divided by the nominal S_T distribution.



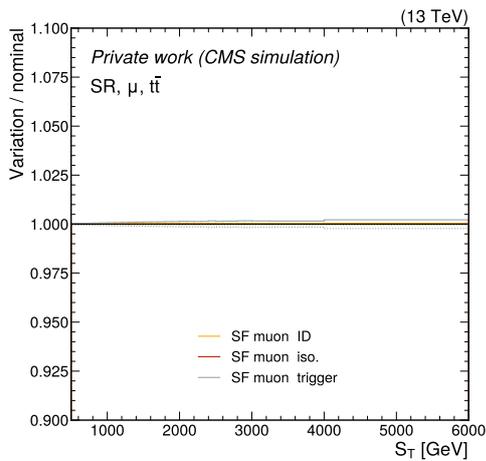
(a) Variations of JEC and JER.


 (b) b -tagging uncertainties.


(c) Misc. SF uncertainties.



(d) Electron-related uncertainties.



(e) Muon-related uncertainties.

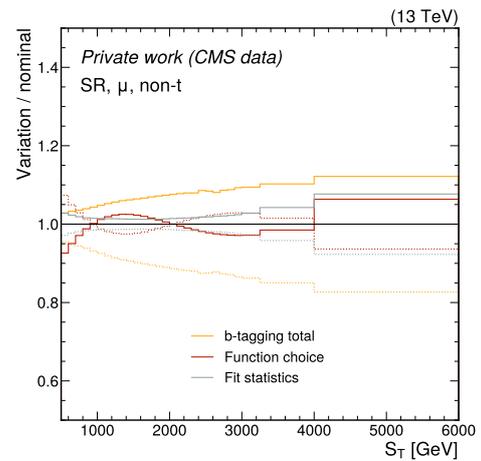

 (f) Non- t uncertainties.

 Figure 5.36: Variations of S_T for different experimental uncertainties, divided by the nominal S_T distribution. Different representative processes and channels are shown.

to issues in the statistical analysis, therefore I employ a smoothing procedure to address them. Figure 5.37 shows the up- and down-variations of JEC in one era and channel, with the variation being mostly constant against S_T . Therefore, I fit it with a constant function, which is then used to derive the new, smoothed JEC variations. JEC and JER are treated independently in each data-taking era, and all simulated samples are affected. Figure 5.36a shows the effect of the JEC and JER variations on S_T for the $t\bar{t}$ process.

***b*-tagging** Various uncertainties are associated with the *b*-tagging SFs affecting all simulated samples. I give an overview of the sources in table 5.5, which have different era correlation depending on their origin. The effect of the more relevant of these sources on the $t\bar{t}$ process is shown in figure 5.36b. Additionally, the *b*-tagging scale factors have an uncertainty that is varied together with the jet corrections, as these can change the *b*-tagging SFs as well. Because I observe a channel-dependent effect in the application of *b*-tagging scale factors, I consider an additional channel-specific uncertainty on the *b*-tagging yield corrections. It covers the difference between combined and channel-specific application of the secondary *b*-tagging yield corrections and covers the lepton p_T -dependent effect introduced by the *b*-tagging SFs. I label this uncertainty *b-tagging yield uncertainty*, and also show it in figure 5.36b.

Prefiring weights I vary the event weights correcting for trigger prefiring within $\pm 1\sigma$ of their uncertainty to obtain variations of the S_T distribution, which are shown in figure 5.36c for the $t\bar{t}$ sample. The uncertainty is treated as uncorrelated between eras of data-taking.

MC PU reweighting To estimate the uncertainty resulting from the correction of PU in simulated samples, the uncertainty on the total inelastic cross section of 4.6% is utilized [227]. To gauge the effect of this uncertainty onto the S_T distribution, I perform PU reweighting with two alternative PU distributions, resulting from the up- and down-variations of the total inelastic cross section. This results in the variation I show in figure 5.36c for $t\bar{t}$ samples. The resulting uncertainty is treated as fully correlated between eras of data-taking.

Lepton ID, isolation and trigger As I mention above, I use various SFs to ensure good description of lepton properties in the simulated samples. These SFs are associated with uncertainties, which I vary within $\pm 1\sigma$ to estimate their effect on the S_T variable. Their effect is treated as uncorrelated between eras of data-taking. I show results for electrons in figure 5.36d and for muons in figure 5.36e. Electron-related uncertainties do not affect the muon channel and vice versa.

Decorrelation Above, I use the VR to validate the background estimation procedure, and conclude that I expect it to therefore also be valid in the SR. While I take great care to ensure optimal decorrelation between the DNN and the S_T variable, some small S_T shape differences between SR and VR remain. I take these into account by assigning a *decorrelation uncertainty* to simulated processes in the SR. Per process, I calculate the ratio of (normalized) S_T distributions in SR and VR, and assign the deviation from unity as an uncertainty. As this is initially one-sided,

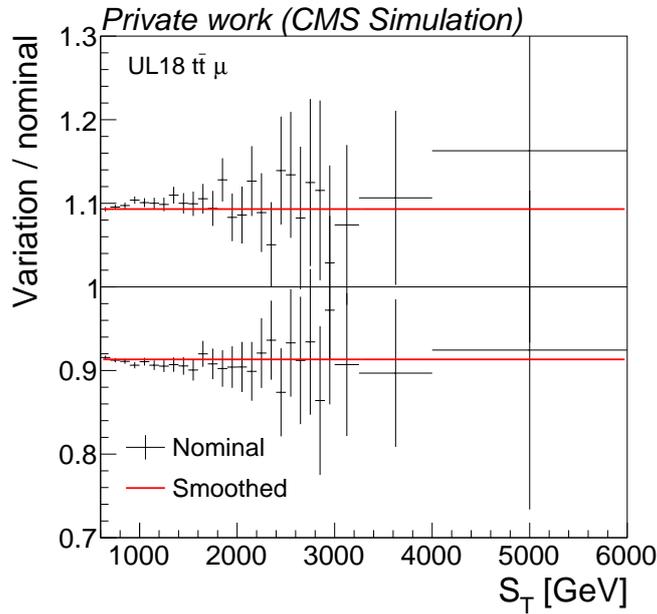


Figure 5.37: Example of the JEC smoothing from the 2018 muon channel, using $t\bar{t}$ events. The up- and down-variation is shown in black relative to the nominal S_T distribution, and two red lines show the constant fits used to derive the smoothed variations.

source	description	era correlation
hfstats1	statistical uncertainty on heavy flavor jets	uncorrelated
hfstats2	statistical uncertainty on heavy flavor jets	uncorrelated
lfstats1	statistical uncertainty on light flavor jets	uncorrelated
lfstats2	statistical uncertainty on light flavor jets	uncorrelated
hf	systematic uncertainty on heavy flavor jets	correlated
lf	systematic uncertainty on light flavor jets	correlated
cferr1	systematic uncertainty on charm jets	correlated
cferr2	systematic uncertainty on charm jets	correlated

Table 5.5: Description of all sources of b -tagging related uncertainties, splitting them into parts of statistical nature (uncorrelated) an systematic origin (correlated), acting on different jet flavors.

I mirror it to obtain a two-sided variation. To avoid jumps from a lack of simulated events in some bins, I employ a smoothing procedure, calculating the uncertainty in each bin as an average of itself and its neighboring bins. In principle, a shape difference between VR and SR does not invalidate any method, thus this uncertainty can be seen as artificial. Therefore, I expected it to be constrained to small values during statistical analysis. The uncertainty is treated as fully correlated between years of data-taking. I show its effect on S_T in figure 5.36c. For the non- t background components, this uncertainty is not considered as the usage of two different α -ratios for SR and VR covers any shape difference.

Background estimation Different systematic uncertainties are associated with the data-driven estimation of non- t background processes. To estimate the uncertainty resulting from the choice of fitting function, I repeat the procedure using the two individual fitting functions as transfer functions instead of the averaged $g_{\text{TF}}(S_T)$. The results of the background estimation with the individual functions provide an up- and down-variation of the nominal result, which I show in figure 5.36f. Additionally, each functional fit yields a statistical uncertainty from the minimization procedure. This is propagated to the mean fit function, assuming that the uncertainty is uncorrelated between the different functions. I show the resulting uncertainty band around $g_{\text{TF}}(S_T)$ in figures 5.32 and 5.33. Repeating the background extrapolation procedure with functions defined by the upper and lower border of this band yields an additional up- and down-variation of the nominal non- t S_T distribution, shown in figure 5.36f.

Furthermore, while it is expected that the contribution of most systematic effects cancels out when the α -ratio is built, some might still have an effect on the results. Therefore, I repeat the full background estimation procedure for variations of select systematic uncertainties: As the CR is defined by a b -tagging criterion, variations of the b -tagging correction SFs might influence the background estimation results. To avoid studying eight different sources, instead of using the individual sources of b -tagging SF uncertainty described above, I add their effects in quadrature to obtain a single, total b -tagging variation. In figure 5.38, I show two new α -ratios, derived for an up- and down-variation of the combined b -tagging uncertainty, which are different from the nominal one. Therefore, I repeat the fitting procedure described above for each of them and show the resulting average functions in figure 5.38. I perform the background estimation with these two functions and use the results as another source of systematic uncertainty on the estimation of non- t backgrounds. I test if it is relevant whether this source is correlated or uncorrelated with the other b -tagging uncertainties, and find no difference. I perform similar studies for the variations of JEC and JER scale factors, where the effect is much smaller, which is why I assume it to be covered by the other considered uncertainties.

The effect on S_T of all uncertainties from the data-driven background estimation procedure are shown in 5.36f.

5.3.10.2 Theoretical uncertainties

μ_r and μ_f variations In section 2.1.4, I introduce the renormalization and factorization scales μ_r and μ_f . In principle, physical results should be completely independent of these parameters. However, due to the limited degrees of pertur-

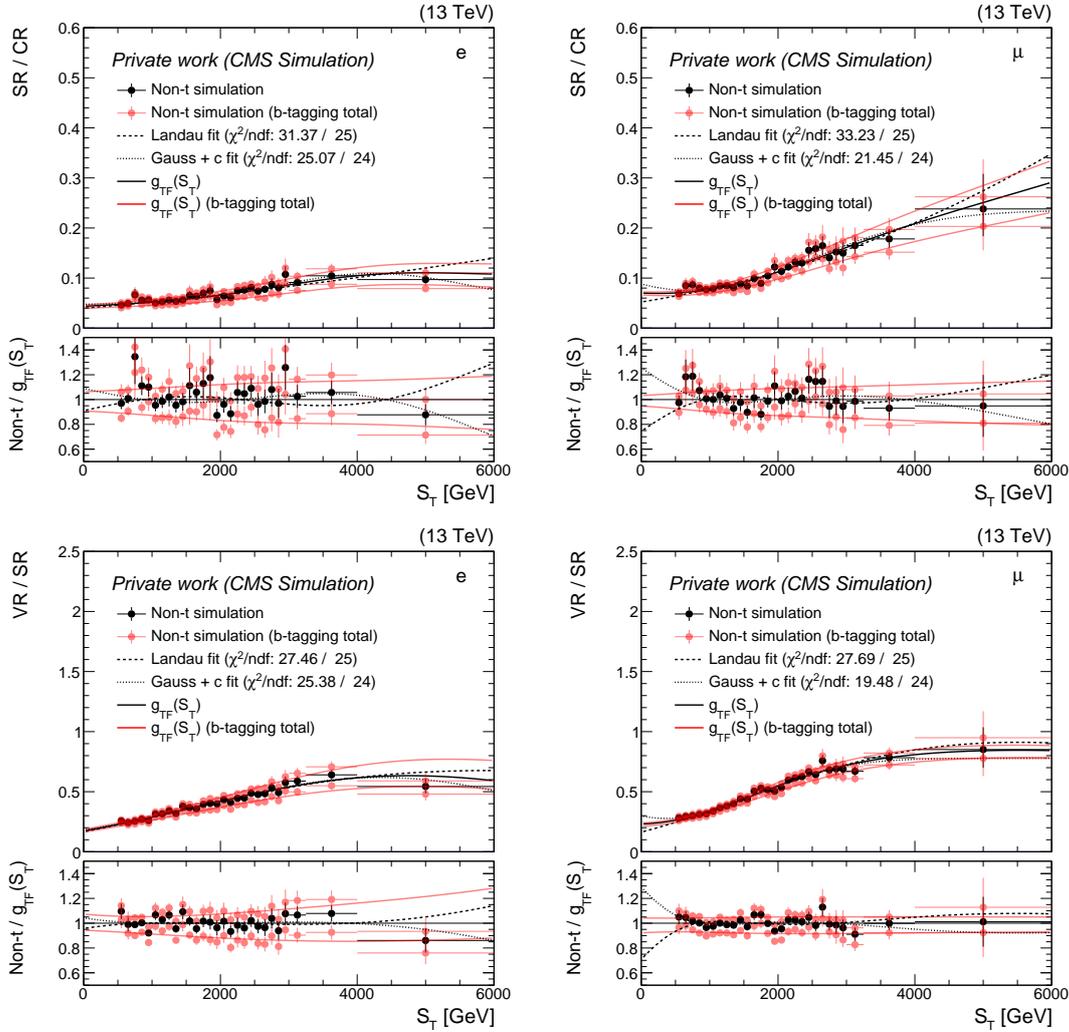


Figure 5.38: Background extrapolation procedure for the SR (top) and SR (bottom) in the electron (left) and muon (right) channels. Black points in the upper part of the figures show the nominal α -ratios, alongside the two fitting functions and their mean. For up- and down-variations of combined b -tagging SFs, α -ratios are shown in red, as well as the resulting mean fit functions. The bottom part of the figure shows all entries divided by the mean fit function.

bation theory used in simulation, this is not the case. Therefore, an uncertainty is assigned to account for this effect by varying both parameters up (down) by a factor 2 ($\frac{1}{2}$) individually. From these variations, I construct all possible combinations (except for the two unphysical up-down combinations). For each simulated process, I calculate an envelope around the resulting variations and use it as a single source of uncertainty on the S_T variable, which I refer to as MC *scale* uncertainty in the following. For the signal samples, where the initial amount of events is unknown and the statistical analysis later results in a relative signal strength, this uncertainty source must not change the normalization, therefore any normalizing change in these variations is removed. Being fully correlated between years of data-taking, this is one of the dominant uncertainties of this search, as shown in figure 5.35.

PDF choice As I introduce in section 5.3.2, the NNPDF 3.1 NNLO PDF set is used for simulation of the proton substructure. However, this specific choice might result in a bias. To account for this in form of an uncertainty, 100 replicas of the utilized PDF set are simulated and their effect on the S_T distribution is analyzed. I calculate the root mean squared (RMS) of these replicas compared to the nominal as

$$\text{RMS}_i = \sqrt{\frac{1}{N_{\text{replicas}}} \sum (b_i - b_{\text{nominal}})^2}, \quad (5.15)$$

where b_i is the bin content of some S_T bin b for replica i , and b_{nominal} is the nominal content of that bin. This RMS calculation results in a $\pm 1\sigma$ variation around the nominal S_T distribution for simulated samples. For signal samples, as for the μ_r and μ_f variations, any normalizing effect is removed. Overall, the effect of the PDF uncertainty, which is treated as fully correlated between eras, is minor in this analysis, as seen in figure 5.35.

Top p_T reweighting As described above, I use top p_T -reweighting to address some imperfections in the descriptions of high momentum t quarks. To ensure that this does not introduce any bias, I consider another uncertainty, defined as the difference between applying and not applying the reweighting. By design, this uncertainty is one-sided, as shown in figure 5.35. It is treated as fully correlated between years.

5.3.10.3 Analysis region results

With all uncertainties defined, distributions of S_T in the VR and SR can now be used to get a full picture of their effect. In figures 5.39 and 5.40, I show S_T split by eras, channel and regions. Overall, good agreement between SM prediction and data is visible in the VR. This shows that the utilized techniques work as intended. Due to the limited number of events towards high S_T , especially in the two parts of the 2016 dataset, strong fluctuations are visible. While information on the individual parts of Run 2 might be valuable, these fluctuations can worsen the stability of the statistical analysis procedure. Therefore, I combine all eras to obtain four final distributions: S_T in the VR and SR, split by lepton channel. Fluctuations observed in the individual eras cancel out and very good agreement is visible, as I show in figure 5.41.

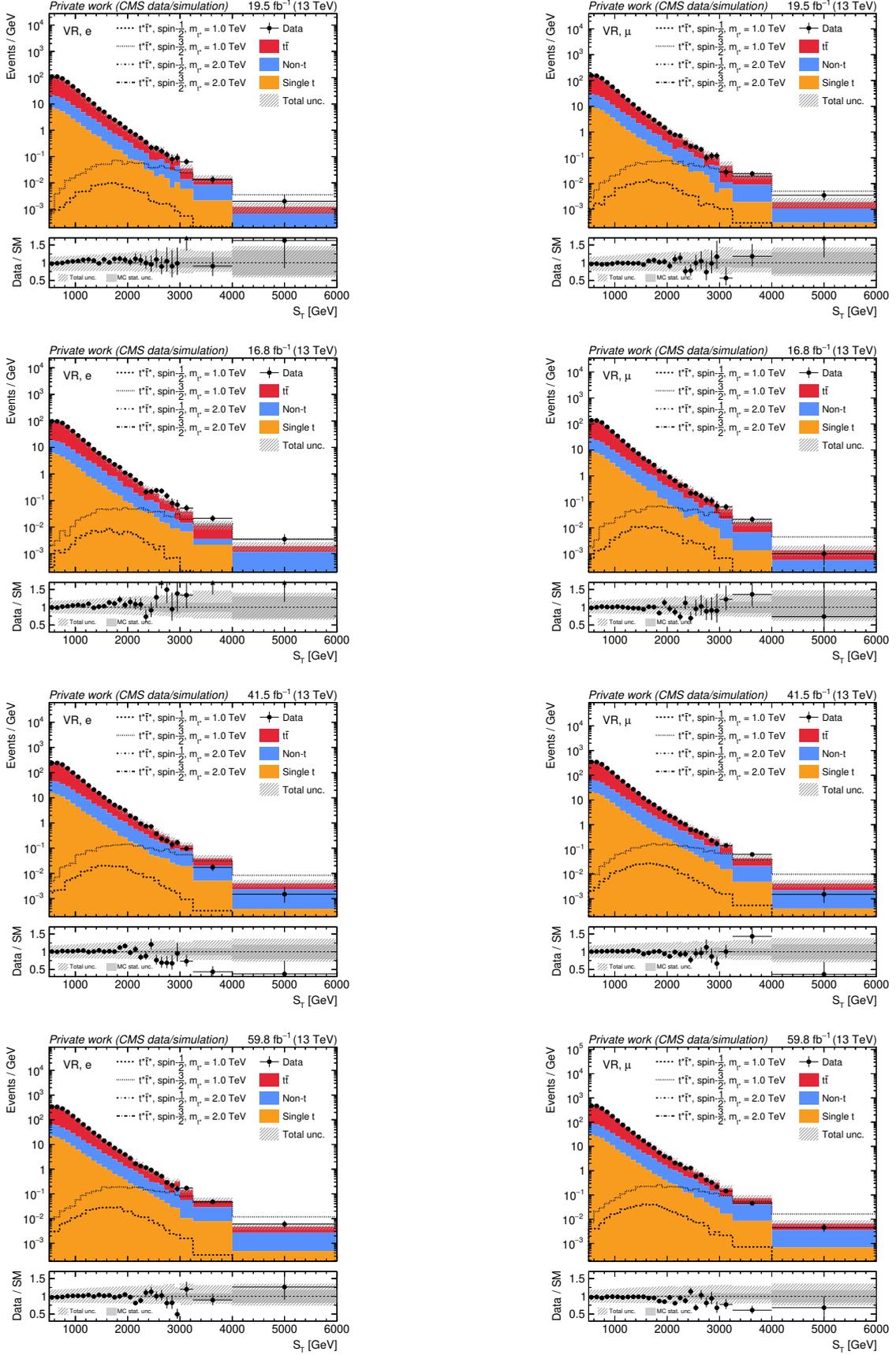


Figure 5.39: VR S_T distribution considering the full set of uncertainties for, from top to bottom, 2016 pre-VFP, 2016 post-VFP, 2017, and 2018.

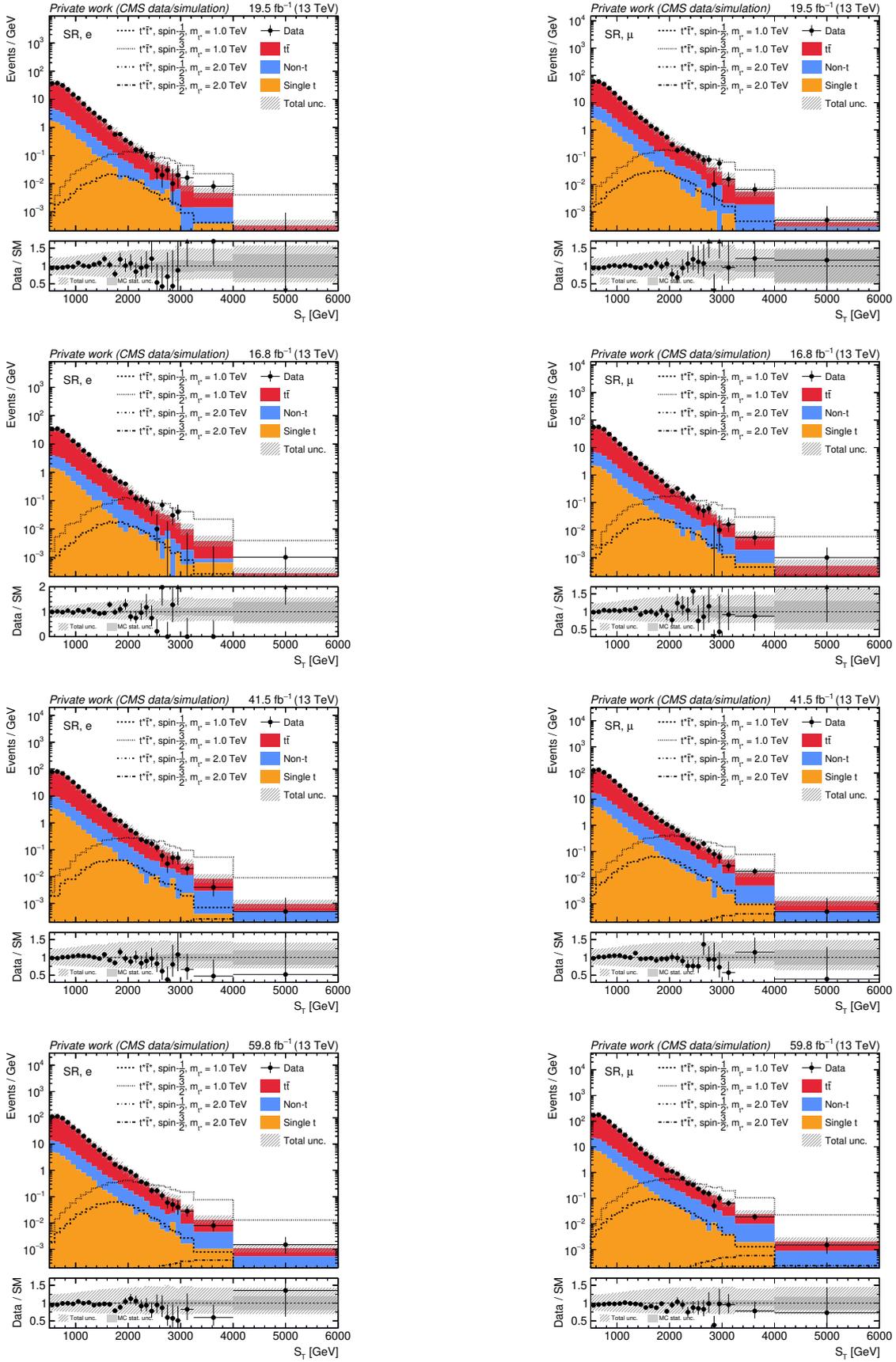


Figure 5.40: SR S_T distribution considering the full set of uncertainties for, from top to bottom, 2016 pre-VFP, 2016 post-VFP, 2017, and 2018.

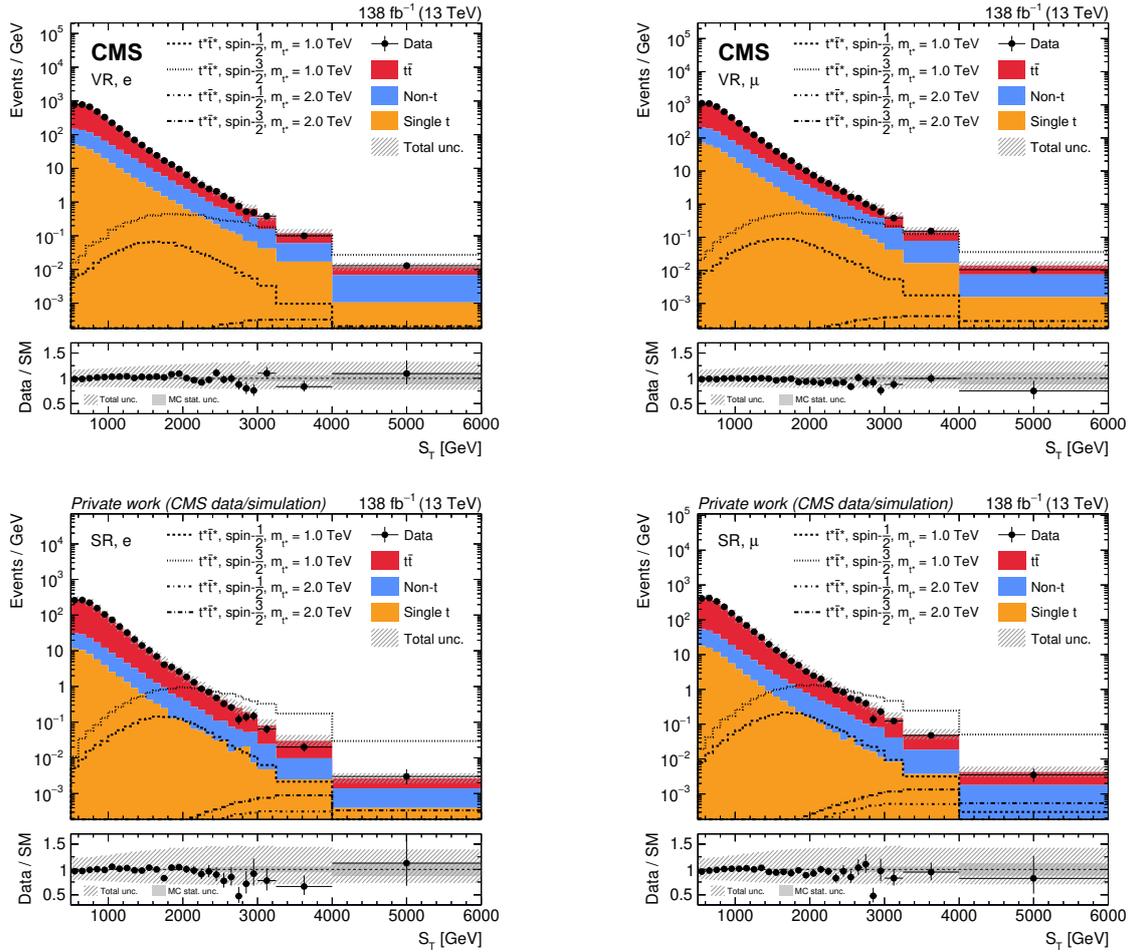


Figure 5.41: S_T distribution in the VR (top) and SR (bottom), combining all eras of data-taking and including full uncertainties. The upper two figures are published in [1].

5.4 Results

Having defined the SR and validated the description of SM processes there, the data in that region can now be probed for the existence of a $t^*\bar{t}^*$ signal. The development of the above steps was performed with a *blinded* signal region, not evaluating data there until after all procedures are fixed. The same is true for the statistical analysis I present below, which I performed on the VR first, before including the SR and the data therein.

5.4.1 Statistical analysis

I use the CMS statistical analysis tool COMBINE [240], which is based on the ROOFIT [241] and ROOSTATS [242] frameworks, for all methods described in the following. Ultimately, the goal of the statistical analysis is to make a statement about the presence of signal, parametrized by a signal strength modifier r . It describes the relative normalization of a signal sample, where $r = 1$ denotes a signal with exactly the input strength to the statistical procedure, which I chose to be 10 fb^{-1} . I use a *binned profile maximum likelihood* approach to study r . I construct a statistical model from the S_T distributions of simulated samples, data-driven non- t background estimation results and observed data, split into electron and muon channels. The model contains r as a single parameter of interest. All uncertainties detailed in section 5.3.10, as well as bin-by-bin statistical uncertainties, are included as nuisance parameters $\vec{\theta}$. A likelihood function \mathcal{L} is constructed as

$$\mathcal{L}(\vec{n}; r, \vec{\theta}) = \underbrace{\prod_b^{N_b} \mathcal{P}(n_b; \lambda_b(r, \vec{v}))}_{\text{Poisson}} \underbrace{\prod_k p_k(y_k; \theta_k)}_{\text{nuisances}}. \quad (5.16)$$

where \vec{n} describes the number of events in the different bins b of the input histograms. The first term describes the product of Poisson probability of n_b events being observed given some expectation $\lambda_b(r, \vec{v})$ dependent on r and the nuisance parameters \vec{v} . The second part describes the contribution of nuisance parameters θ_k , where the knowledge on these is parameterized by a vector of priors \vec{y} .

Post-fit distributions Using the MINUIT2 [243] minimizer with the MIGRAD algorithm, I determine the maximum likelihood estimator of a background-only fit to the data, fixing r to zero. The resulting so-called *post-fit* S_T distributions, where all nuisance parameters are set to their post-fit values $\hat{\theta}$, are shown in figure 5.42. The good agreement in the VR shows that the data is well described by the SM predictions and associated uncertainties. Likewise, the good agreement observed in the SR shows that the data can be well described by a SM-only hypothesis.

Nuisance parameter correlations Is it important to investigate potential correlations or anti-correlations between nuisance parameters. If these exist, the fitting procedure can be unstable, for example when multiple nuisance parameters counteract each other and thus prevent the fit from converging. I calculate a covariance matrix for all nuisance parameters and find the strongest correlation to be 54% between the electron channel b -tagging yield nuisance and the top p_T reweighting

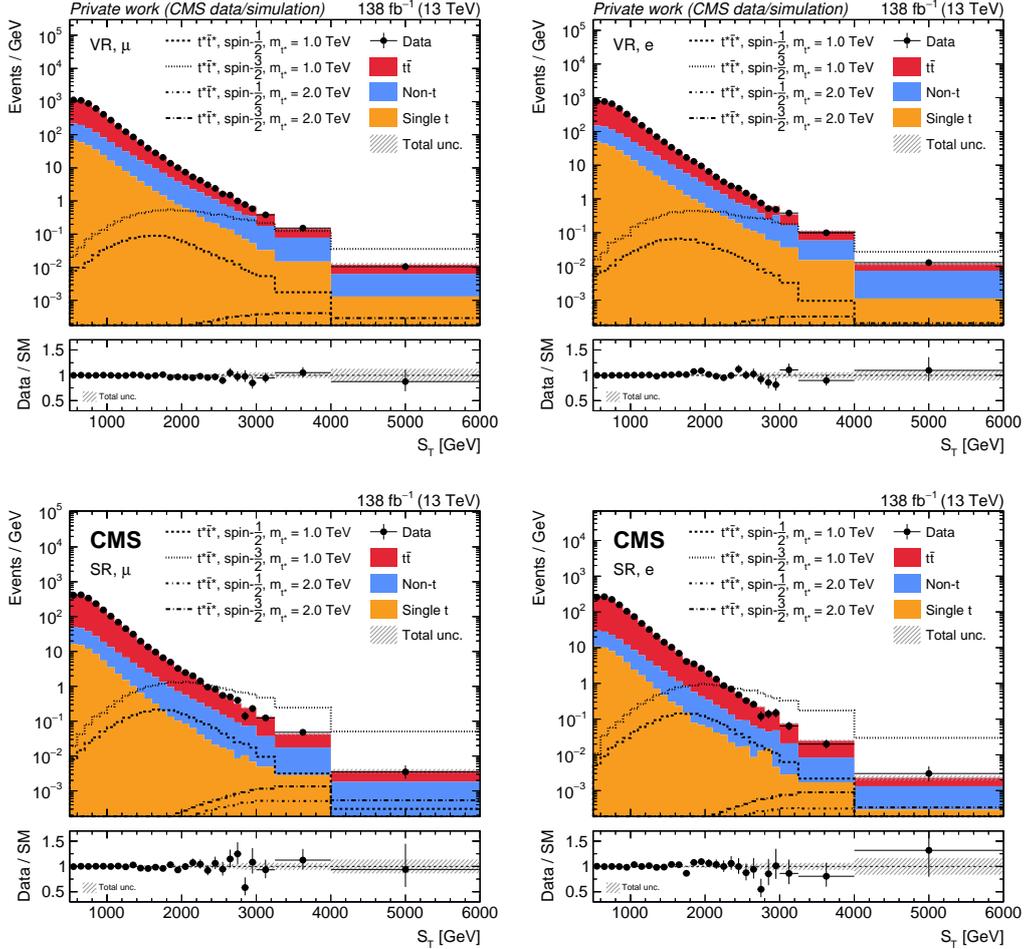


Figure 5.42: Post-fit S_T distributions in the VR (upper row) and SR (lower row) for a background-only fit to the data. The signal distributions are scaled to the cross section predicted by the theory. The figures in the lower row are published in [1].

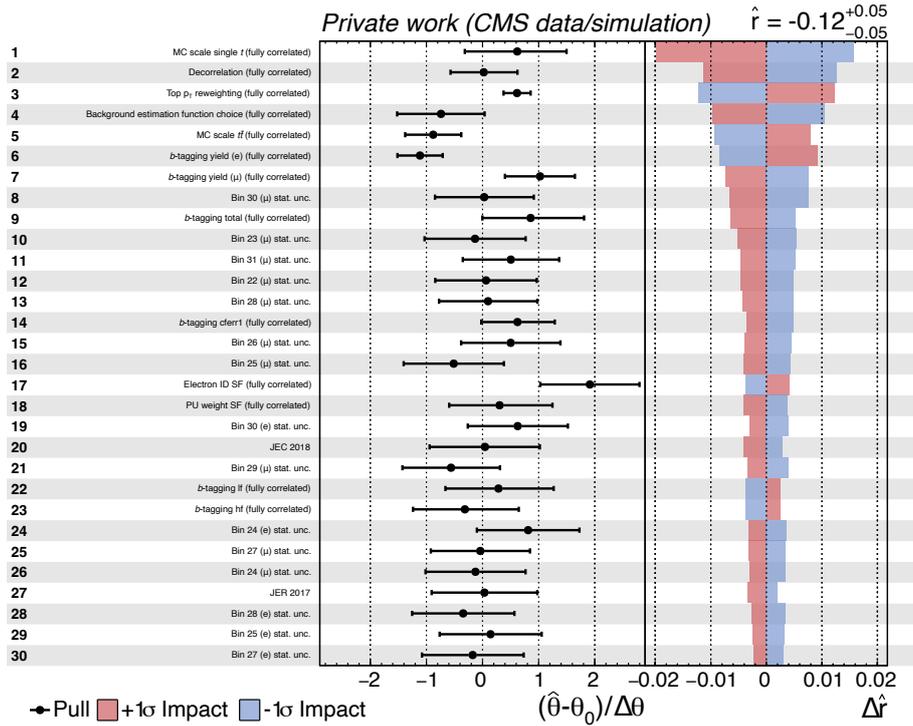


Figure 5.43: Pulls (left panel) and impacts (right panel) of all nuisance parameters in the VR. Nuisances are ordered by impact, only the 30 most relevant ones are shown.

nuisance. The correlation stems from these to nuisances having similar shapes in their variations, but is small enough to not be problematic.

Nuisance parameter impacts and pulls Before analyzing the parameter r in detail, it's crucial to understand the behavior of nuisance parameters in order to validate the fitting procedure. For this purpose, I investigate the so-called *impacts* and *pulls* of all nuisance parameters. The impact of a nuisance parameter θ on the parameter of interest r is the shift $\Delta\hat{r}$ that is induced when fixing θ to its $\pm 1\sigma$ post-fit values, and then profiling all other parameters as normal [244]. It provides information on the relevance of each nuisance to the final result of the analysis. The pull of a nuisance parameter is given as $(\hat{\theta} - \theta_0) / \Delta\theta$, where $\hat{\theta}$ is the post-fit and θ_0 the pre-fit value of the nuisance parameter, and $\Delta\theta$ its pre-fit uncertainty. If a parameter is strongly pulled towards a variation, this could signal that the central value of that parameter might not be optimally chosen. Also, uncertainties that are initially chosen too large can become visible by strongly constrained pulls, meaning that the post-fit uncertainty is much smaller than the pre-fit one. None of this necessarily means that something is wrong with the statistical model, but to avoid unstable fits or biases it is crucial to understand any unusual behavior.

In figure 5.43, I show impacts and pulls for the 30 most relevant nuisance parameters (ordered by their impacts), for a fit of the VR, investigating a signal with $m_{t^*} = 1200$ GeV. This fit results in a small post-fit value of the parameter of interest $\hat{r} = -0.12^{+0.05}_{-0.05}$, as expected due to the low amount of signal in the VR. MC scale uncertainties, b -tagging related nuisances, as well as the decorrelation uncertainty dominate. Most pulls are well centered and unconstrained, with some exceptions.

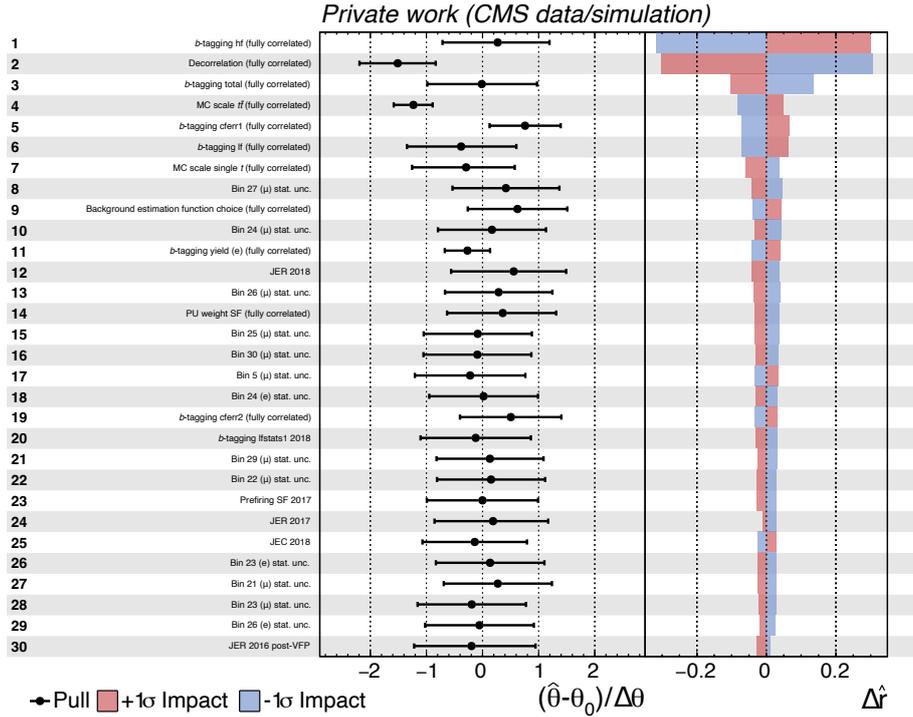


Figure 5.44: Pulls (left panel) and impacts (right panel) of all nuisances in the SR. Nuisances are ordered by impact, only the 30 most relevant ones are shown.

The top p_T uncertainty is strongly constrained, which is expected: it is not a real uncertainty where the data is expected to vary within the full range, instead there is a correct value of this parameter and the fit is allowed to find and measure it at a greater precision than assumed with the initial uncertainty. The b -tagging yield uncertainties are pulled away from their nominal value, which is also expected: I introduced them to enable the fit to describe differences between the electron and muon channels that are not properly modeled in simulation.

After ensuring that the general setup of nuisances works as intended in the VR, I repeat the same procedure in the SR and show the results in figure 5.43. I do this before unblinding \hat{r} in the SR, in order to first evaluate whether any problems are present in the nuisances. Thus, \hat{r} is not shown in the figure. The most relevant nuisances are similar to the ones in the VR: b -tagging related uncertainties, decorrelation and MC scale. Overall, most nuisances are unconstrained and well centered. One exception is the decorrelation uncertainty, which was introduced so that the fit is able to correct for any potential remaining bias originating from the DNN application. Thus, this uncertainty may be varied away from its central value without this being a sign for a problem with the fit. Another pulled nuisance parameter is the MC scale parameter related to the $t\bar{t}$ sample. As this parameter includes a large normalization effect on $t\bar{t}$ events, and the analysis measures a $t\bar{t}$ -dominated region, it is expected that the large pre-fit uncertainty can be constrained. Thus, overall, the statistical model does not show unexpected behavior and provides a good description of the VR and SR, respectively.

Goodness-of-fit tests Another important procedure to study the statistical model is a goodness-of-fit test (GoF), which evaluates how well a the statistical model de-

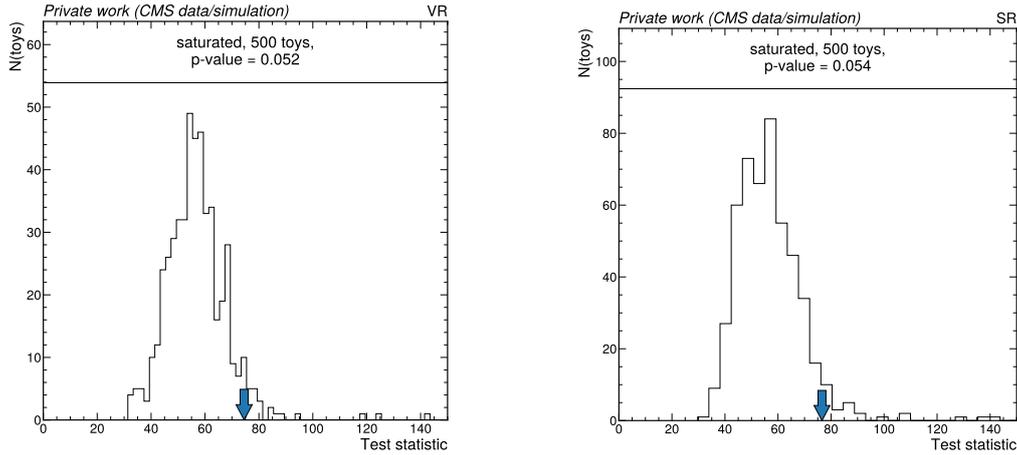


Figure 5.45: Results of GoF tests in the VR (left) and SR (right), showing the test statistic results for toy data as a histogram, and the value for data as an arrow.

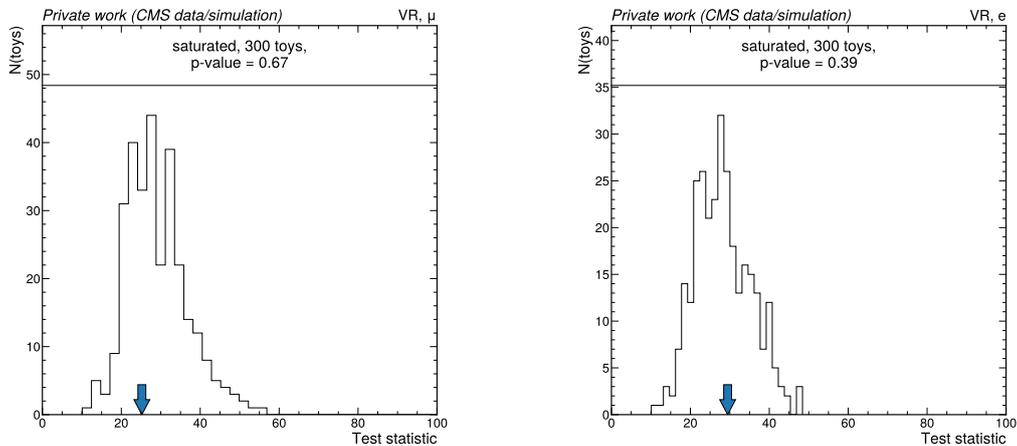


Figure 5.46: Results of GoF tests in the VR, split into muon (left) and electron channel (right). Shown are the test statistic results for toy data as a histogram, and the value for real data as an arrow.

describes the data. Several algorithms can be used for that, in this analysis I make use of the *saturated* model [245]. The test is based on the creation of toy data. I generate a number of toy datasets by randomly sampling from the probability distribution $p(\theta_k|y_k)$ of all nuisance parameters [240], and fixing r to zero. For each of these toy datasets, I calculate the value t^{sat} of the saturated test statistic. The same is done considering actual data to obtain (t_0^{sat}) . Then, a distribution $f(t^{\text{sat}})$ can be built for all toys, and the p -value of measuring the observed t_0^{sat} is calculated as

$$p^{\text{sat}} = \int_{t_0^{\text{sat}}}^{\infty} f(t^{\text{sat}}) dt^{\text{sat}}. \quad (5.17)$$

In figure 5.45, I show the results of a GoF test in the VR. Here, a background-only fit is expected to yield good agreement between prediction and data, where a p -value above 5% is seen as a passing test. This threshold is fulfilled in the VR, however the p -value is relatively small. During analysis development, I studied this feature to great detail in order to ensure that the statistical model is valid. Initially, I observed a failing GoF test in the VR, which resulted from shape differences between the two

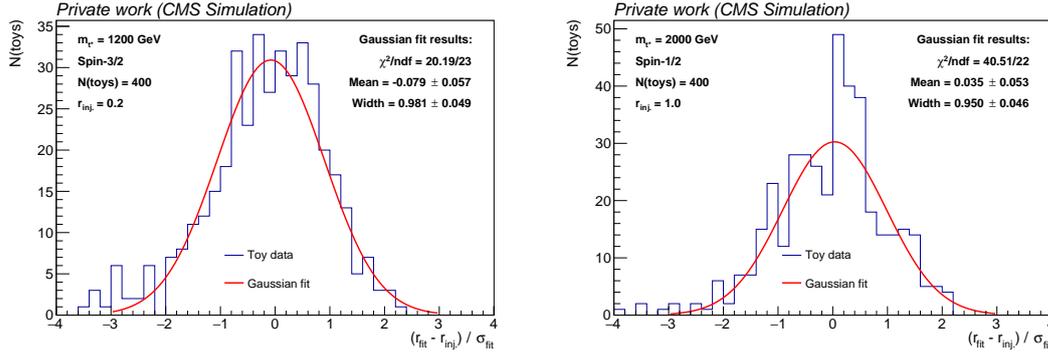


Figure 5.47: Two representative injection test results, showing the measured residuals for toy data, and a Gaussian fit to their distribution. The resulting fit parameters, including mean and width, are given in each figure.

lepton channels. As the fit had no freedom to properly describe this, the GoF failed, but was successful when performing it individually in both channels, as I show in figure 5.46. There, the p -values being near 50% signals very good description. As the main source of not well described shape difference between channels comes from the application of b -tagging scale factors, the introduction of a channel-specific b -tagging yield correction uncertainty fixed the issue and results in a passing GoF test in the VR. I then perform a GoF test in the SR. As shown in 5.45, the result is very comparable to the one seen in the VR, with the p -value being small but above 5%. This consistent result is expected, as the two regions do not differ in their background composition, thus showing that the DNN-based definition of the SR does not introduce any bias.

Signal injection tests All tests signal that a valid statistical model is found and the predictions of SM backgrounds are able to describe the data well. Before finally testing for the result of the analysis, the parameter r , it is crucial to ensure that the statistical model is also able to accurately determine that parameter. For this purpose, I perform *signal injection tests*. In these tests, toy data is generated using a fixed signal strength r_{inj} . Then, the signal strength is measured for each toy, obtaining a set of r_{fit} values. If the model is properly set up, the residual $r_{\text{fit}} - r_{\text{inj}} / \sigma_{\text{fit}}$ should follow a Gaussian distribution with width 1 centered at 0. I perform this test for two different potential t^* masses ($m_{t^*} = 1200$ GeV and 2000 GeV), considering both spin scenarios. To have a signal large enough to be measured, different than for all other tests, I chose $r_{\text{inj}} = 1$ to correspond to a cross-section of 1 pb, instead of 10 fb. All signal injection tests succeed, having a mean close to 0 with width close to 1. Two representative test results are presented in figure 5.47.

Exclusion limits Finally, after validating the statistical model, the post-fit signal strength modifier \hat{r} can be evaluated. As was already visible above, the data can be well described by the SM prediction, no deviation is observed and \hat{r} is found to be very small. Therefore, upper exclusion limits on the the signal strength, and thus the production cross section of the $t^* \bar{t}^* \rightarrow t \bar{t} g$ process, are set. For this, a *test statistic* $q_{\text{LHC}}(r)$ is defined as

$$q_{\text{LHC}}(r) = -2 \ln \left(\frac{\mathcal{L}(r, \hat{\vec{\theta}}(r))}{\mathcal{L}(\hat{r}, \hat{\vec{\theta}})} \right) \quad (5.18)$$

for a tested signal strength r (assuming $0 \leq \hat{r} \leq r$, see [240] for the full definition). Here, $\hat{\vec{\theta}}$ are the nuisance parameters corresponding to the maximum likelihood estimator of the best-fit value \hat{r} . $\mathcal{L}(\hat{r}, \hat{\vec{\theta}})$ can be understood as a normalizing factor, ensuring that the likelihood ratio is always smaller than one. Conversely, $\hat{\vec{\theta}}(r)$ are the nuisance parameter values maximizing the likelihood for the tested r . On this test statistic, p -values can be defined by integration over some distributions f as

$$p_r = \int_{q_{\text{LHC}}^{\text{obs}}(r)}^{\infty} f(q_{\text{LHC}}(r)|r) dq_{\text{LHC}} \quad (5.19)$$

and

$$p_b = \int_0^{q_{\text{LHC}}^{\text{obs}}(r)} f(q_{\text{LHC}}(r)|0) dq_{\text{LHC}}. \quad (5.20)$$

With these, the so-called CL_s criterion is defined as

$$\text{CL}_s = \frac{p_r}{1 - p_b}. \quad (5.21)$$

where the value of CL_s can be understood as a confidence level in the existence of a signal with strength r . I find upper limits on the signal strength at $\text{CL}_s = 95\%$. The probability distributions f could be obtained with pseudo-data randomly sampled from the likelihood distributions. However, using the `AsymptoticLimits` option of `COMBINE`, instead an asymptotic approximation as introduced in [246] is performed to speed up computation. From the upper limit on r , an upper limit on the production cross section times branching fraction squared $\mathcal{B}(t^* \rightarrow tg)$ can be obtained by considering the input normalization of the simulated signal samples.

I calculate *observed exclusion limits* on the $t^* \bar{t}^* \rightarrow tg \bar{t}g$ process as a function of m_{t^*} , separately for the spin- $\frac{1}{2}$ and spin- $\frac{3}{2}$ scenarios, using the respective signal samples. The resulting limits are shown in figure 5.48 for the spin- $\frac{1}{2}$ and figure 5.49 for the spin- $\frac{3}{2}$ case. In order to understand these results, *expected exclusion limits* are shown as well. These are calculated in the same way, but use a so-called *Asimov dataset*. To obtain this, pseudo-data is generated assuming the data exactly follows the SM predictions, with all nuisance parameters set to their best-fit value assuming $r = 0$. Any deviation from the SM-only hypothesis would be visible in a deviation of the observed from the expected limits, which is not the case.

For the spin- $\frac{1}{2}$ case, the observed exclusion limits are found to be between 120 fb (with 190 fb expected) for $m_{t^*} = 700$ GeV, and 0.8 fb (0.8 fb expected) for $m_{t^*} = 3000$ GeV. For the spin- $\frac{3}{2}$ scenario, I find upper exclusion limits between 15 fb (18 fb expected) at $m_{t^*} = 700$ GeV and 1.0 fb (0.9 fb expected) at $m_{t^*} = 2750$ GeV. I compare the results to theory predictions of the $t^* \bar{t}^*$ pair production cross section, assuming a 100% branching fraction of the $t^* \rightarrow tg$ decay. These theory predictions are obtained based on the same simulations used to generate signal samples, and were validated by comparing to the results of previous CMS analyses [214, 215] and theory results [204]. By investigating the crossing between my exclusion limits and theory predictions, mass exclusion limits are obtained. A spin- $\frac{1}{2}$ t^* is excluded up to m_{t^*} values of 1050 GeV, with 990 GeV expected. In the spin- $\frac{3}{2}$ scenario, the lower

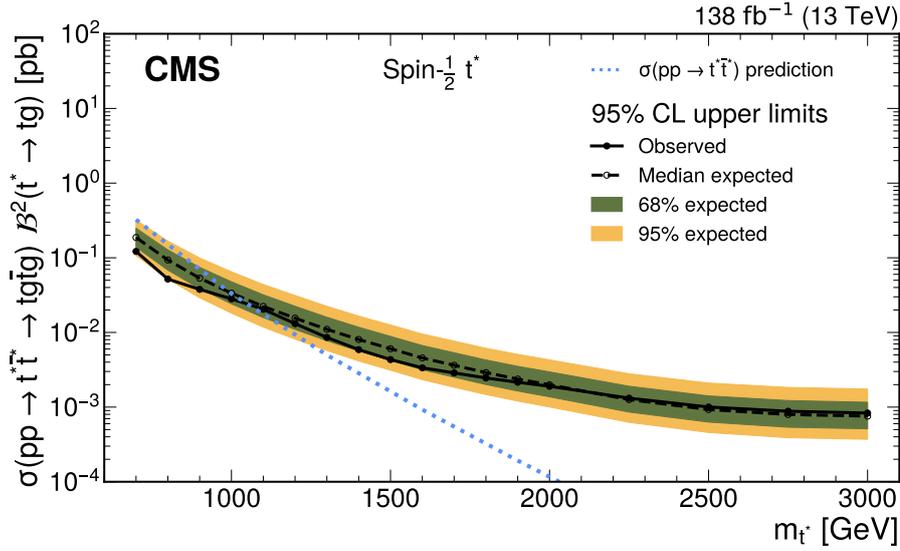


Figure 5.48: Expected and observed 95% CL upper exclusion limits on the $t^* \bar{t}^*$ production cross section times branching fraction squared $\mathcal{B}^2(t^* \rightarrow tg)$ for a spin- $\frac{1}{2}$ t^* as a function of m_{t^*} . The colored bands give the central probability intervals containing 68 and 95% of the expected upper limits, assuming the background-only hypothesis is true. The dashed line shows the $t^* \bar{t}^*$ pair production cross section predicted by theory, following the EFT approach introduced in [204], assuming $\mathcal{B}(t^* \rightarrow tg) = 1$. This figure is published in [1].

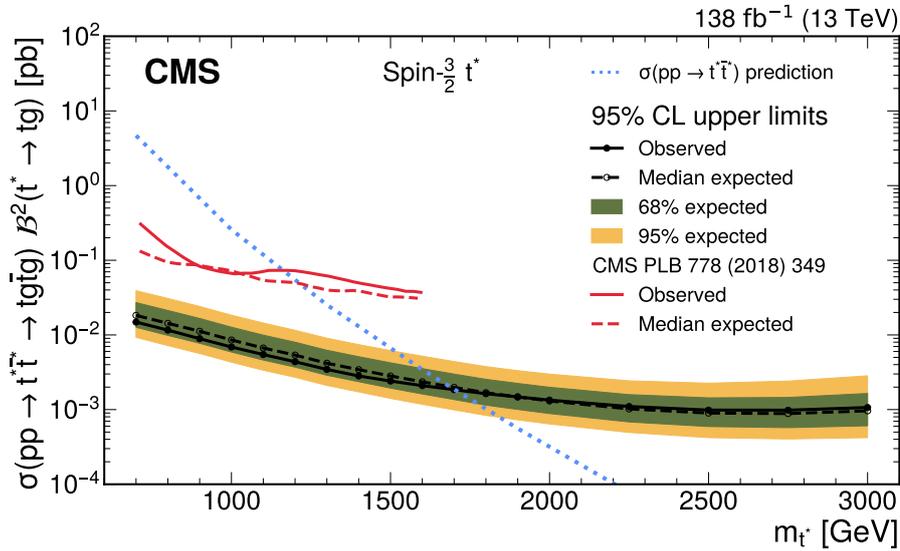


Figure 5.49: Expected and observed 95% CL upper exclusion limits on the $t^* \bar{t}^*$ production cross section times branching fraction squared $\mathcal{B}^2(t^* \rightarrow tg)$ for a spin- $\frac{3}{2}$ t^* as a function of m_{t^*} . The colored bands give the central probability intervals containing 68 and 95% of the expected upper limits, assuming the background-only hypothesis is true. The dashed line shows the $t^* \bar{t}^*$ pair production cross section predicted by theory, following the EFT approach introduced in [204], assuming $\mathcal{B}(t^* \rightarrow tg) = 1$. The results of the previous 13 TeV CMS analysis [214] are shown in red. This figure is published in [1].

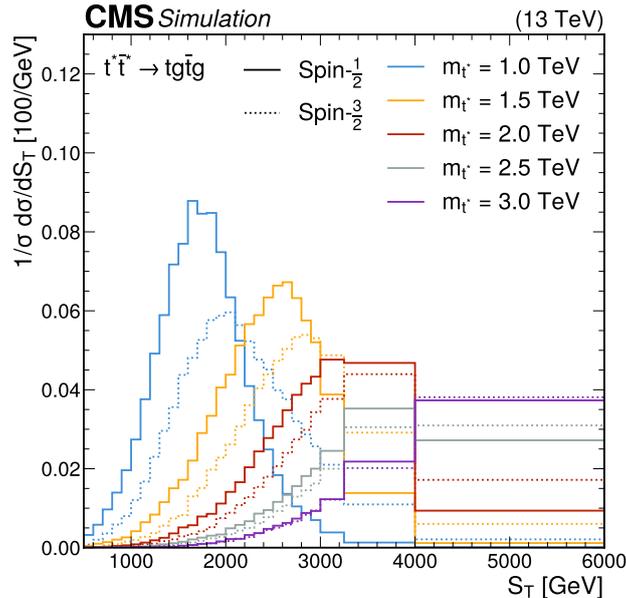


Figure 5.50: Distributions in S_T for $t^* \bar{t}^*$ signal samples in the SR, with different simulated values of m_{t^*} , for spin- $\frac{1}{2}$ (solid lines) and spin- $\frac{3}{2}$ (dashed lines) resonances. The distributions are normalized to the same area for each signal. This figure is published in [1].

mass limit is higher: The existence of a spin- $\frac{3}{2}$ t^* is excluded below m_{t^*} of 1700 GeV, with 1690 GeV expected.

5.4.2 Discussion and future prospects

The results on spin- $\frac{1}{2}$ t^* are the first ever 13 TeV exclusion limits, and the first ever study of this spin scenario with dedicated simulations at CMS. For a spin- $\frac{3}{2}$ t^* , the previous best mass exclusion limit, set by the 2016-only CMS search, was 1200 GeV, and is considerably improved by the work I present here. This improvement is not resulting from the increased size of the analyzed dataset: using 138 fb^{-1} instead of the previously used 36 fb^{-1} is estimated to yield about a factor of two improvement in sensitivity. The previous search used a SR definition relying on strict cut-based selection criteria, including a requirement of isolated leptons. As discussed above, this limits the selection efficiency on signal events, especially towards high m_{t^*} . The mass reconstruction approach the previous analysis followed made these strict requirements necessary, which I avoid by using S_T and a DNN-based approach in this work. Overall, this way I obtain an about five times higher efficiency of signal events to reach the SR, while increasing background yields only by about 10%. Thus, the updated analysis techniques result in an additional factor five improvement, resulting in an overall increase in sensitivity by about an order of magnitude. Notably, this change in analysis strategy is not without disadvantages. I studied the overall analysis strategy in my master's thesis [5], where I found the DNN- and S_T -based approach to be much more sensitive compared to a mass reconstruction. However, I now find that the mass resolution of the analysis diminishes: while the t^* samples do show distinctive maxima at different S_T values depending on m_{t^*} , as I show in figure 5.50, this behavior is not retained when evaluating exclusion limits.

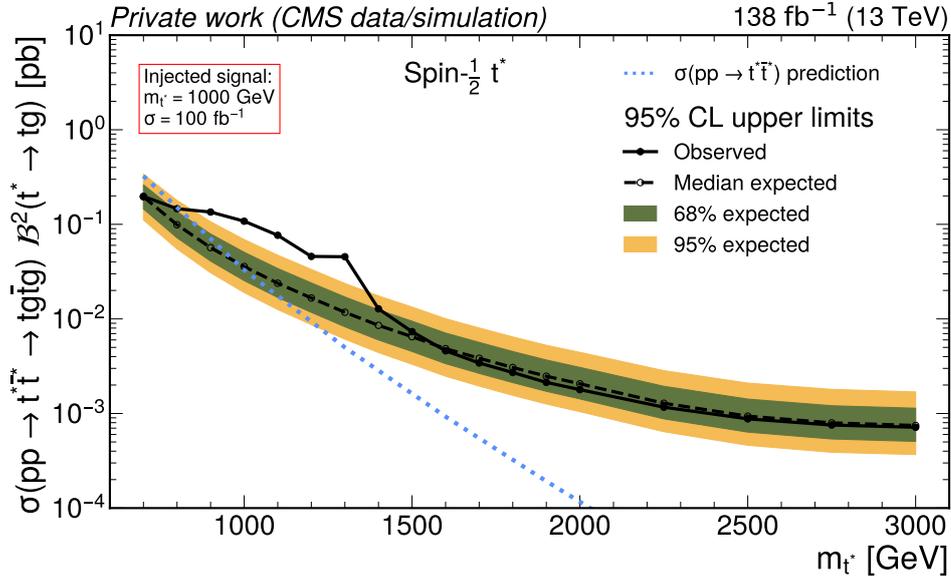


Figure 5.51: Expected and observed 95% CL upper exclusion limits, injecting an artificial signal of a $m_{t^*} = 1000$ GeV spin- $\frac{1}{2}$ t^* with $\sigma = 100$ fb⁻¹.

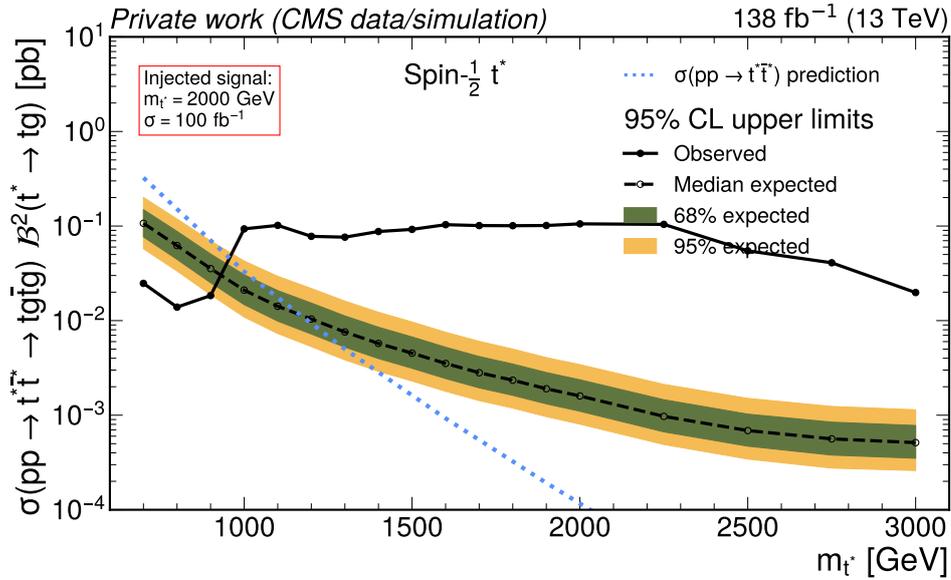


Figure 5.52: Expected and observed 95% CL upper exclusion limits, injecting an artificial signal of a $m_{t^*} = 2000$ GeV spin- $\frac{1}{2}$ t^* with $\sigma = 100$ fb⁻¹.

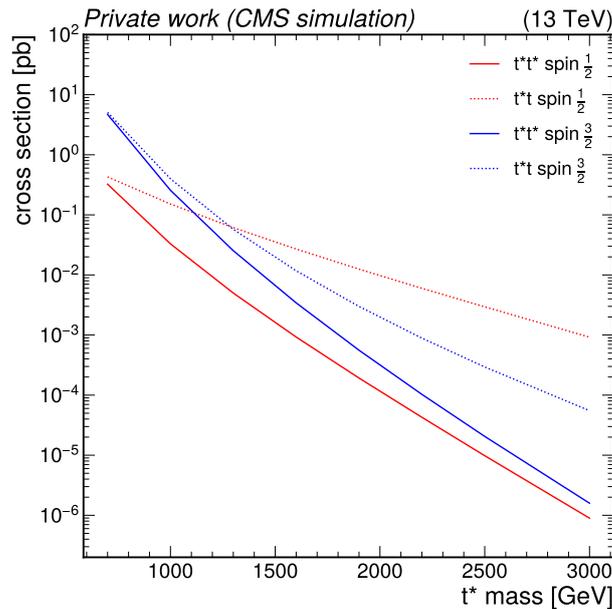


Figure 5.53: Predicted production cross sections of t^* pairs (solid lines) compared to production of a single t^* in association with a t quark (dashed lines). SM-like couplings are assumed for the t^*-t vertex.

Due to the large bin widths needed to ensure sufficient statistical power towards high S_T , t^* samples at different masses can not be well distinguished. This can be seen by artificially injecting a t^* signal on top of the actual data in the SR, and then calculating exclusion limits. I show the resulting observed limits in figure 5.51 for low m_{t^*} injection, where there still is some mass resolution left, and in figure 5.52 for a high m_{t^*} injection, where the mass resolution becomes very poor. This leaves room for possible improvements in figure iterations of the analysis, trying to recover some mass resolution sensitivity. In the context of the work I present here, I focus on optimal sensitivity, however I also implemented a mass reconstruction approach which could have been used to access m_{t^*} in case an excess is observed. My studies on the mass reconstruction approach can be found in my master thesis [5].

Another option to improve the analysis lies in the analyzed process itself. In my work, and also all previous ones at CMS, only $t^*\bar{t}^*$ pair production was considered. However, in principle, the production of a single t^* in association with a t quark is also allowed: a t^*t process (where either the t^* or the t is an antiparticle). This process could result in a final state identical to the one I analyze here, except for one fewer gluon jet. Especially thanks to the S_T -based approach, which does not require all final state objects to be individually reconstructed, the presented analysis most likely would be sensitive to this process, although less so than to the $t^*\bar{t}^*$ case: some sensitivity would be lost due to the overall lower S_T and greater similarity to $t\bar{t}$, which would make the DNN less sensitive. I study the predicted production cross section of t^*t in comparison to $t\bar{t}$ using simulations. The results are shown in figure 5.53.

These show that the t^*t process can actually dominate over $t^*\bar{t}^*$, especially towards high m_{t^*} and in the spin- $\frac{1}{2}$ case. Thus, a future analysis could benefit substantially

from an inclusion of this process. However, crucially, the coupling between t^* and t now not only enters in the t^* decay, but at the $g \rightarrow t^*t$ vertex as well. Therefore, different to the $t^*\bar{t}^*$ case, the production cross section depends on the assumptions made for the coupling parameters c_i . The results I present here assume SM-like couplings of the t^* , as presented above. However, this does not necessarily hold true, and would need to be taken into account when performing an analysis including the t^*t case.

As I mentioned in section 5.1, the branching fraction of the $t^* \rightarrow tg$ decay is predicted to be 97%. While this leaves only about 2% for the $t^* \rightarrow t\gamma$ decay, analyzing a final state with a photon might still be very promising [204]. Specifically, the $t^*\bar{t}^* \rightarrow tg\bar{t}\gamma$ process is of interest, as it offers reduced jet multiplicity and a clear splitting into a hadronic and photonic leg of the $t^*\bar{t}^*$ decay. This aids the t^* mass reconstruction, which can be a valid strategy for this process. Also, by introducing a photon veto to this analysis, a photonic search can be set up orthogonally, and by combining the searches an increased total sensitivity can be obtained.

Many analyses similar to the one I presented above are performed at CMS. All of them rely on the availability of the data they need, supplied by the CMS detector and selected by the trigger system. Therefore, the optimal performance of that system is crucial, and continuous development is needed. Especially in the context of the HL-LHC, sophisticated strategies beyond simple cut-based trigger algorithms will be necessary. The substantial increase in instantaneous luminosity will result in much higher PU, with an expected 200 simultaneous interactions. These conditions will pose additional challenges onto the trigger system. In this chapter, I present studies on the development of novel trigger algorithms for the L1T. Targeting a specific, currently trigger-limited process, I investigate the optimization, integration and usage of a fully supervised ML-based trigger strategy, as proposed in [156]. I study this in the context of the ongoing Run 3 of CMS, information about which are given in appendix A.1.2. While very important for the HL-LHC, novel techniques can already be tested and used now, both to learn more about their development and to aid current analyses.

6.1 Limitations of cut-based triggers

The currently used trigger algorithms, especially in the L1T, mostly rely on simple cut-based conditions, or combinations of these. For many applications, cut-based triggers are very suitable. However, there are some analyses for which cut-based triggers are not as efficient. An example for this is the process $HH \rightarrow bbWW \rightarrow bbqq\mu\nu$, which I write as $HH \rightarrow bbWW$ (single muon) in the following.

Final states of this process are characterized by low momentum muons, which can be seen in figure 6.1. There, using simulated samples, I compare generator-level muons from that process to ones from $t\bar{t}$ pair production. When triggering on this process with a cut-based strategy, usually a single muon trigger is used, requiring at least one L1 muon with a momentum threshold of around $p_T > 22$ GeV. It is clearly visible that with that approach many signal events are lost in the L1T. However, to improve the signal efficiency, it is not possible to simply lower the p_T threshold, as the trigger rate would increase substantially and might exceed the allowed rate budget. This can be seen in the right part of figure 6.1. The amount of low energy background events, which make up almost all of the collisions at the LHC, exponentially increases towards low muon momenta, and thus is the primary driver behind the rate of a trigger. This exponential rise is a general problem for cut-based triggers.

In figure 6.2, I show the signal efficiency of different cut-based trigger selections for the $HH \rightarrow bbWW$ (single muon) process as a function of their rate. Using simulated samples, I can calculate that efficiency without the need for any specific efficiency

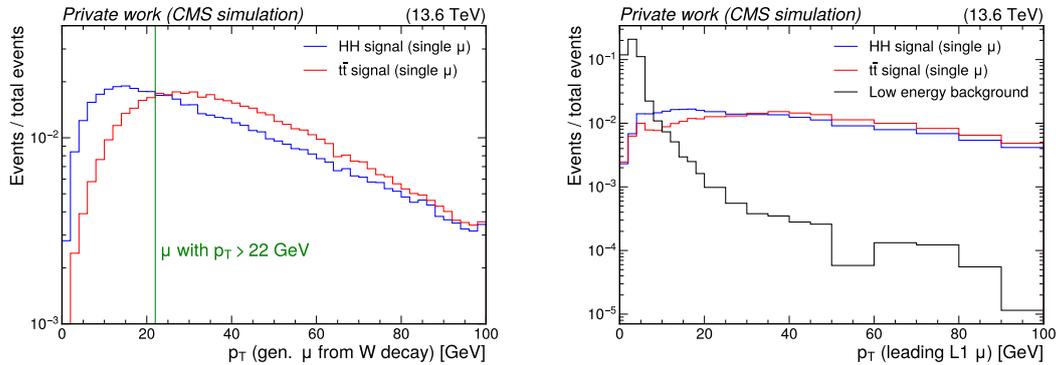


Figure 6.1: Left: Momentum distribution of the generator-level muon originating from a W -decay in $HH \rightarrow bbWW$ (single muon) and $t\bar{t}$ events, compared to the momentum threshold of a single muon trigger at 22 GeV. Right: Momentum distribution of the p_T -leading L1 muon in $HH \rightarrow bbWW$ (single muon) and low energy background events.

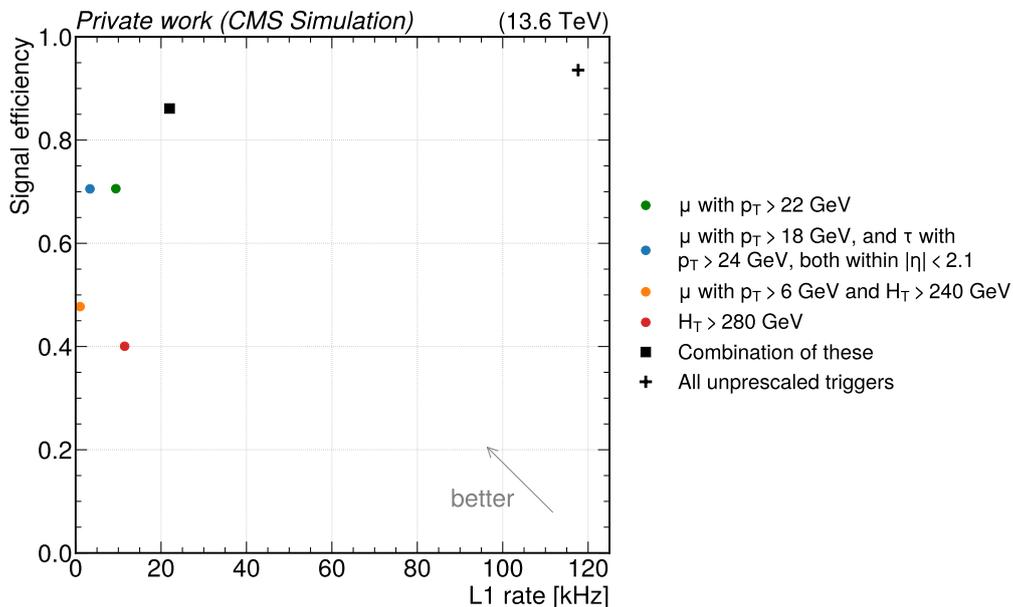


Figure 6.2: Signal efficiency against trigger rate of several cut-based triggers, evaluated on $HH \rightarrow bbWW$ (single muon) and low energy background events, respectively.

measurement technique. Similarly, I estimate trigger rates using a simulated sample of low energy background events. Multiplying the efficiency of a trigger in that sample by the total LHC rate of (about) 40 MHz allows access to the trigger rate. One can see that there are several cut-based triggers that provide good efficiency apart from the already introduced single muon trigger: A trigger requiring a muon with $p_T > 18$ GeV and a tau with $p_T > 24$ GeV (both within $|\eta| < 2.1$) is surprisingly efficient despite no tau lepton being expected in signal events. A trigger simply requiring 280 GeV of total hadronic momentum H_T is somewhat efficient as well, and some more sensitivity can be gained by combining a H_T threshold of 240 GeV with the presence of a muon with $p_T > 6$ GeV. Interestingly, a combination requiring that any of these triggers selects an event achieves a substantially higher signal efficiency (over 85%), and looking at the full menu of (unprescaled) L1 triggers, the overall efficiency is even higher. From these results, one can see that it is possible to somewhat increase the sensitivity to the $HH \rightarrow bbWW$ (single muon) process by combining multiple cut-based triggers, reaching above 90% efficiency when considering the full menu of unprescaled triggers. Unfortunately, using the full set of available L1 algorithms as seed for a single HLT path is not optimal due to timing constraints. The same is true for a combination of many, but not all, cut-based algorithms. Thus, finding a single algorithm that can match or improve the signal efficiency while keeping a lower overall rate would be beneficial.

In offline analysis, using ML to improve results beyond what can be achieved with classical approaches is widely adapted, as done in the $t^*\bar{t}^*$ search presented above. While ML has been used before in the HLT (for example in jet classification [247]), the L1T during Run 2 still relies exclusively on cut-based strategies. Trying to overcome the limitations of cut-based triggers, I study the development of a ML-based strategy for the L1T, targeting the $HH \rightarrow bbWW$ (single muon) process.

6.2 Trigger optimization

As a first step in the development of a ML-based trigger, a strategy needs to be defined. As I investigate a specific process, utilizing a supervised approach is natural. Thus, I develop a supervised, binary NN to identify signal events of the $HH \rightarrow bbWW$ (single muon) process. It should be situated in the GT, where than a threshold on its output can be used as a trigger condition. To find an optimal NN, I train different configurations and compare their performance. As I discuss in section 4.2, many different metrics to quantify the performance of a ML algorithm exist. In the context of trigger development, the most important one is the signal efficiency at a given rate, which is conceptually identical to a TPR@FPR. In contrast to commonly used ML applications, where an optimal FPR working point is in the intermediate range of possible values, trigger development requires a strong reduction in rates and thus very low FPR.

With these considerations, I find a well-performing NN in a dedicated optimization procedure. It involves a selection of input variables sensitive to the desired signal (section 6.2.1) as well as the evaluation of different network hyperparameters (section 6.2.2). All optimization steps need to be performed considering a later integration of the neural network in the trigger hardware: given equal performance, the smallest, fastest possible network is the optimal one. I perform the optimization using simulated $HH \rightarrow bbWW$ (single muon) events, assuming a SM-like Higgs trilinear

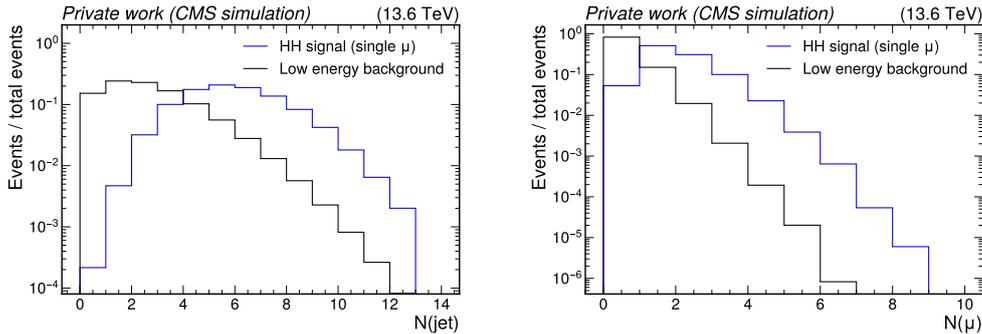


Figure 6.3: Number of L1 jets (left) and L1 muons (right) in $HH \rightarrow bbWW$ (single muon) compared to low energy background events.

coupling, as the signal class. As background class, simulated low-energetic collision events without any hard scattering process are used, as these make up the majority of collision events at the LHC, and thus should be rejected by the NN to suppress the trigger rate. Full information about all samples used in this section, alongside event counts and generators, is provided in appendix A.3.1.

Ultimately, a single NN is needed. However, to avoid biases from overtraining during optimization, I make use of 5-fold cross-validation. Additionally, I employ early stopping with a patience of 10 epochs for all NNs I consider. When a NN is found to be not overtrained this way, I then randomly chose one of the 5 resulting models as the main NN, and use the full available event count to further test it.

6.2.1 Neural network input variables

First, I study potential input variables by training NNs on different input configurations. As I do this prior to the hyperparameter optimization, which I present in section 6.2.2, I need a baseline NN configuration. I chose to use a NN with three hidden layers having 64, 32 and 32 nodes. A ReLU activation is used for all hidden layers, which feed into a single output node with sigmoid activation. I consider BCE loss with the Adam optimizer and a learning rate of 0.0005. To ensure that no overtraining occurs, I add 30% dropout in each layer.

The $HH \rightarrow bbWW$ (single muon) signal process is expected to contain at least one muon, jets and missing transverse energy, all of which are accessible in the GT as L1 objects. p_T and ϕ information is available for energy sums, and p_T , η and ϕ for muons and jets. No electrons or photons are expected, and thus I do not include $e\gamma$ objects in the study. All considered input variables are normalized to a distribution with a mean of 0 with a standard deviation of 1.

6.2.1.1 Number of input objects

First, I study the optimal number of objects to consider. An investigation of the number of muons and jets in signal and background events, as shown in figure 6.3, provides a starting point. As expected, signal events contain more objects than background events, visible for both muons and jets. Also, in general, events contain more jets than muons. It can be seen that the number of jets never exceeds 12, which is the maximal amount of jets passed to the GT by the calorimeter trigger.

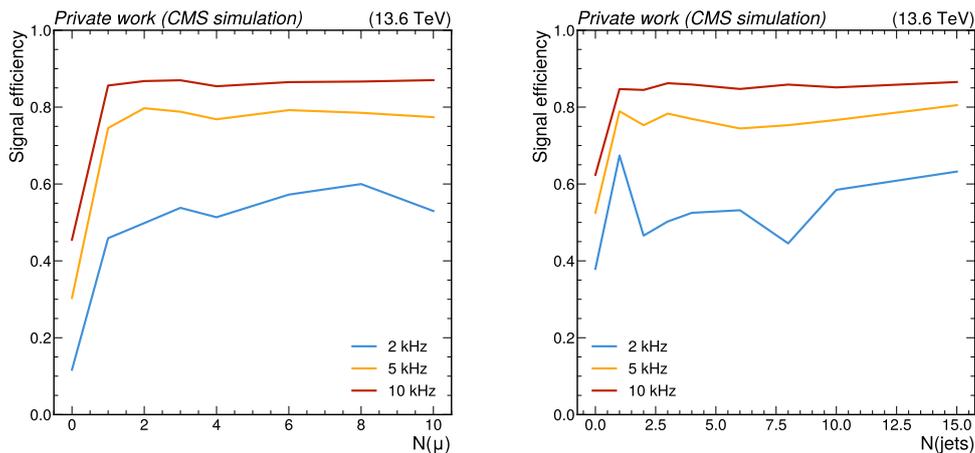


Figure 6.4: Signal efficiency at 2 kHz, 5 kHz and 10 kHz of rate for different muon (left) and jet (right) counts in training.

Signal events are expected to contain at least four jets: two of them from b quarks and two from light quarks from the decay of a W boson. However, the NN might benefit from utilizing more objects, or not need as many. Therefore, I test the NNs with different numbers of muons and jets. Passing p_T , η and ϕ , I fix one of the object counts (to 6 for jets, to 3 for muons) and vary the other. In addition, I include p_T^{miss} and H_T as inputs. For each configuration, I train a NN and ensure that no significant overtraining is visible. Then, the performance of the different object configurations can be compared. For that, I evaluate the signal efficiency of each NN at different rates in figure 6.4.

Overall, the 2 kHz working point shows somewhat unstable results, whereas the higher rates yield more stable curves. The reason for this behavior is limited statistical power in the background dataset, where for very strict requirements on the FPR only few events remain for rate estimation. It can be seen that the network needs at least a single muon to have any sensitivity. Adding a second muon yields a slight increase in performance, from that point the performance is mostly constant: Increasing the number of muons beyond two does not provide any useful additional information to the network. Therefore, for all following sections, I use two muons as input objects. For jets (where I remove the H_T input, as there is considerable overlap in information between it and the jets) the main performance gain also is introduced with adding the first jet. Beyond that, the efficiency is relatively constant. Thus, it seems possible to reduce the number of jets to very low values. However as following studies might show some added sensitivity from jet-related variables, and signal events are expected to contain at least four jets, I decide to keep the number of jets at four in the following.

6.2.1.2 Selection of per-object variables

When studying the number of input objects per type, I use the p_T , η and ϕ variables for all objects (where applicable). In a next step, I investigate whether this is necessary, or whether some information can be removed to further lower the number of input variables. For this, I evaluate SHAP importance scores for the input object

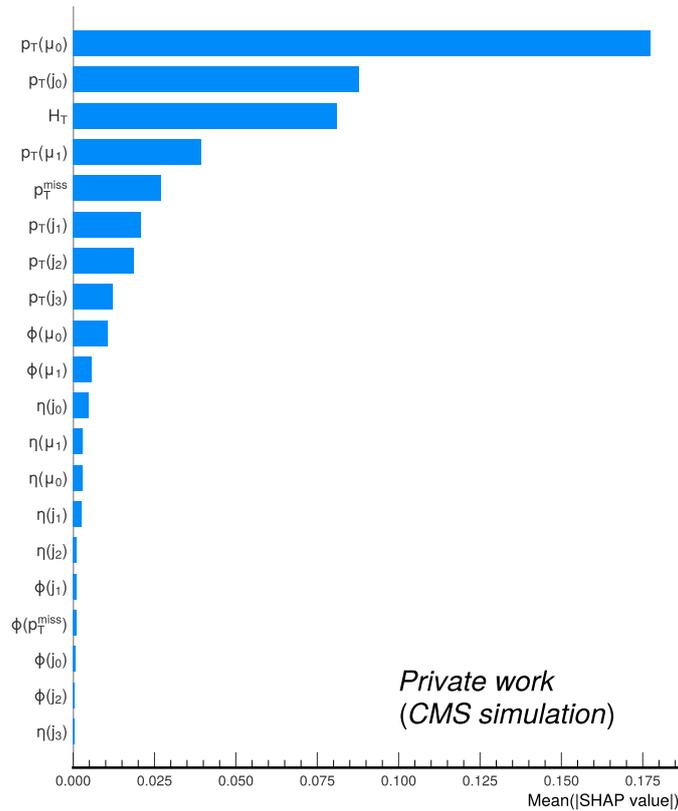


Figure 6.5: Mean absolute SHAP importance scores for a NN trained with four jets, two muons, H_T and p_T^{miss} . All objects are L1 objects.

configuration determined to be optimal in the last chapter, with the same network architecture. The importance scores are shown in figure 6.5.

From the importance scores, one can see that momentum information (p_T) overall is more relevant for the network output than angular information (η and ϕ). The angular input variable scoring the highest is the ϕ of the leading muon, whereas the angular variables of sub-leading jets score lowest. To further understand the influence of these input variables on the neural network, one can compare distributions between signal and background events, which I do in figure 6.6.

The sensitivity of the two displayed momentum variables is immediately visible. In contrast, the leading muon ϕ does not appear sensitive on its own, despite being rated relatively high in importance. To further understand the influence of these kind of angular variables on the NN performance, I train and compare NNs with different input configurations. ROC curve comparisons of these networks are presented in figure 6.7.

It can be seen that there are very small differences between the networks, which confirms the result visible in the SHAP scores: the NN mainly relies on p_T information. Still, in a small region of rates, the neural network without any angular information seems to perform slightly worse than the other two. Therefore, the next largest set of input variables will be used in the following, where learning is supported by usage of η information, but no ϕ variables are passed.

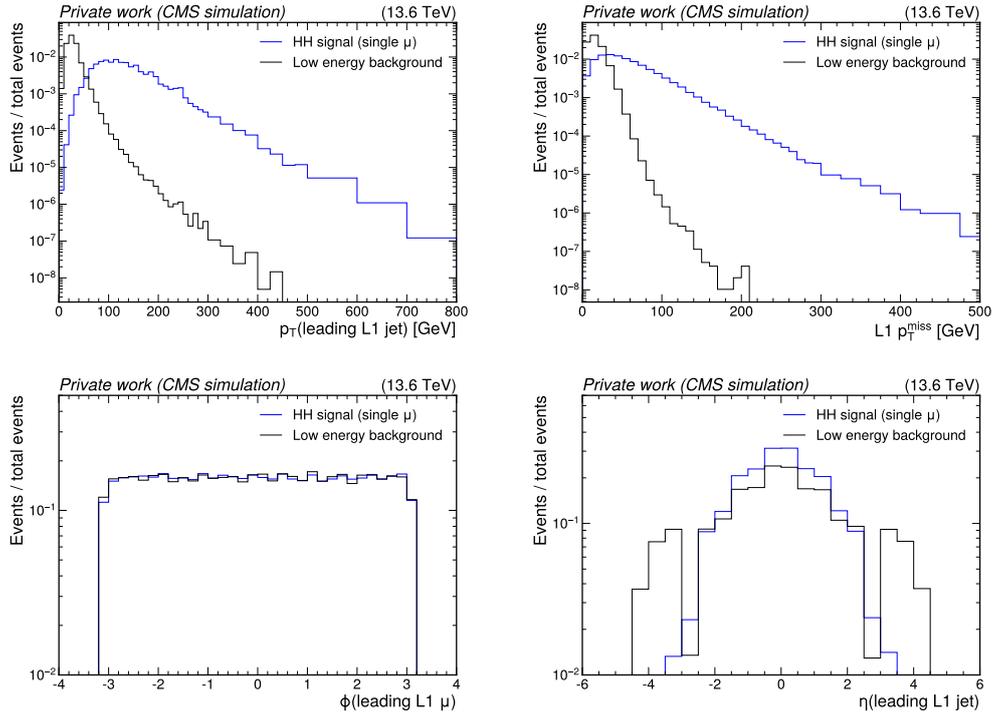


Figure 6.6: Example distributions for some L1 NN input variables, comparing background to $HH \rightarrow bbWW$ (single muon) signal events.

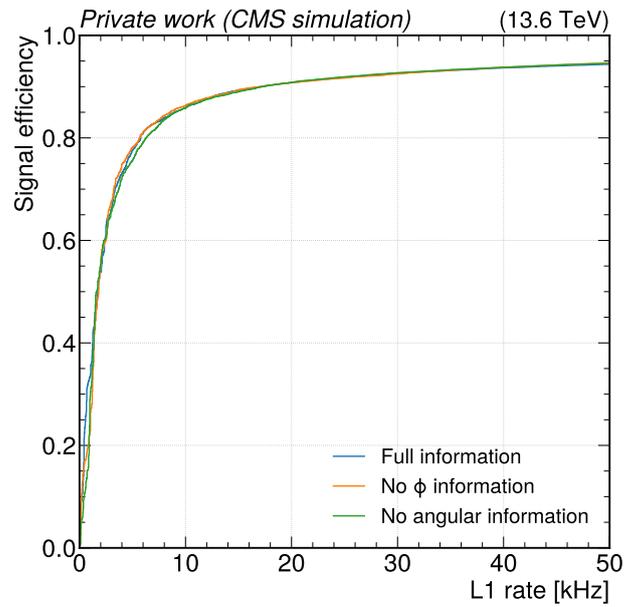


Figure 6.7: ROC curves of NNs trained on different configurations of input variables, from full angular to no angular information.

6.2.2 Neural network architecture

After finding a good set of input variables, I need to optimize the hyperparameters of the NN. To find the optimal number of nodes and layers for the neural network, both these parameters are studied individually first, assuming that they are independent. First, I vary the number of hidden layers between one and seven, while keeping the number of nodes per layer constant at 32. As before, I investigate the signal efficiency at different rates, as shown in figure 6.8. There, one can see that the NN already has very good performance for a single hidden layer, and degrades for high layer counts. In theory, with larger network size, a degradation is not expected, as the network only receives more degrees of freedom to learn with. However, effects of limited training data and overtraining can become more relevant and serve as an explanation for this observation. Still, this study shows that for the given problem, already very small networks provide good performance. I perform a similar test now varying the number of nodes per hidden layer, while keeping the layers fixed at three. Results for NNs with 5 to 150 nodes per hidden layer are shown in figure 6.8. There, one can see an increase in performance with rising node count, with the results stabilizing for NNs with above about 50 nodes per layer.

From these studies I find that a NN with a single hidden layer having 64 nodes yields good results, and an increase in architecture size would not provide a strong gain in performance. With this configuration, I test whether dropout is actually needed to prevent overtraining, as naturally it worsens the NNs performance. As I show in figure 6.9, the NNs performance is best when considering no dropout. Investigating ROC curves of a NN with no dropout on the test and training sets, which I also show in figure 6.9, one can clearly see that no overtraining is present. Thus, the model with no dropout is chosen as the final, main model of this optimization. A full picture of this NNs training is given in figure 6.10. There, one can see the loss and accuracy of all folds, quickly reaching constant behavior. Furthermore, I show the output distribution split into signal and background events, where the discriminating power of the NN is clearly visible.

I demonstrate that good sensitivity to the $HH \rightarrow bbWW$ (single muon) can be achieved with a very small NN. In figure 6.11, I show the ROC curve of this NN together with the rate and signal efficiency of the cut-based triggers introduced above. There, one can clearly see that it is possible to improve upon the results of these with a ML-based approach. Importantly, the optimization procedure I present here is not comprehensive. It might very well be possible to create a larger, better-performing network utilizing more complex techniques. Instead, this section shows that a small NN like this can serve as a good starting point in a first implementation of ML in the CMS L1T.

6.3 Hardware integration

Having found a NN to work with, in a second step I integrate it into the L1T and create a trigger from it by imposing a threshold on the NN output. Integration is a non-trivial procedure, as the L1T is implemented in FPGAs, which can not execute Python code, in which the NN exists after being trained. To convert a Python-trained NN into an FPGA-compatible format, I use the `hls4ml` [248, 249] library. It is able to automatically convert NNs to FPGA firmware using *high level synthesis*

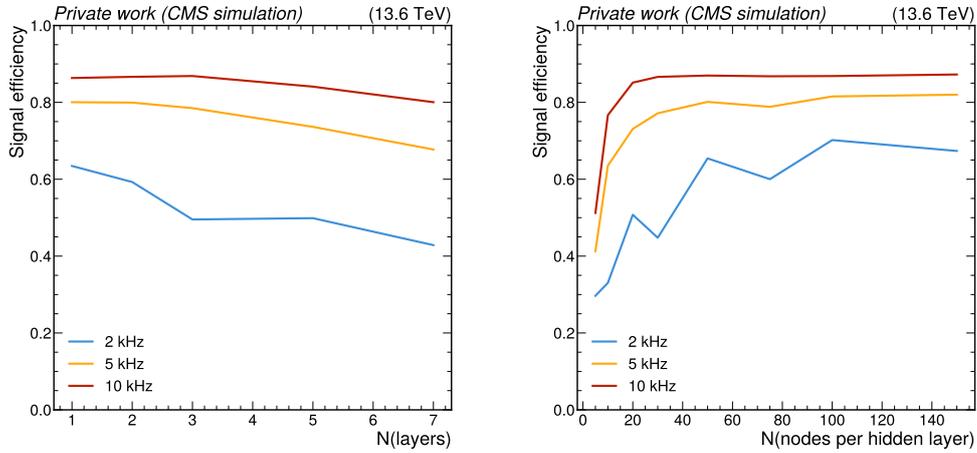


Figure 6.8: Signal efficiency at 2 kHz, 5 kHz and 10 kHz of rate as a function of the number of hidden layers (left) and nodes per hidden layer (right).

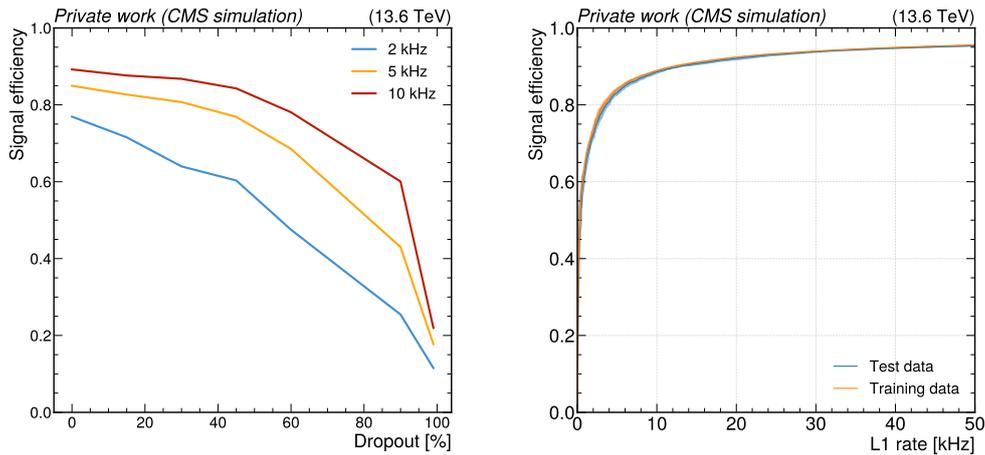


Figure 6.9: Left: Signal efficiency at 2 kHz, 5 kHz and 10 kHz of rate as a function of the dropout percentage. Right: ROC curve comparison between training and test data for a NN without dropout, showing no sign of overtraining.

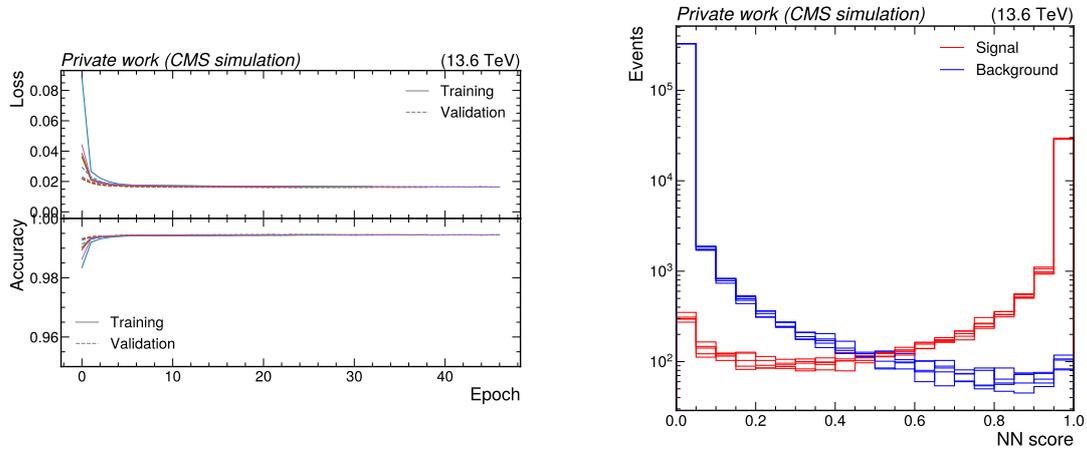


Figure 6.10: Left: Loss and accuracy of the NN trigger during training of all five folds. Right: NN output distribution for signal and background events, showing all five folds (right).

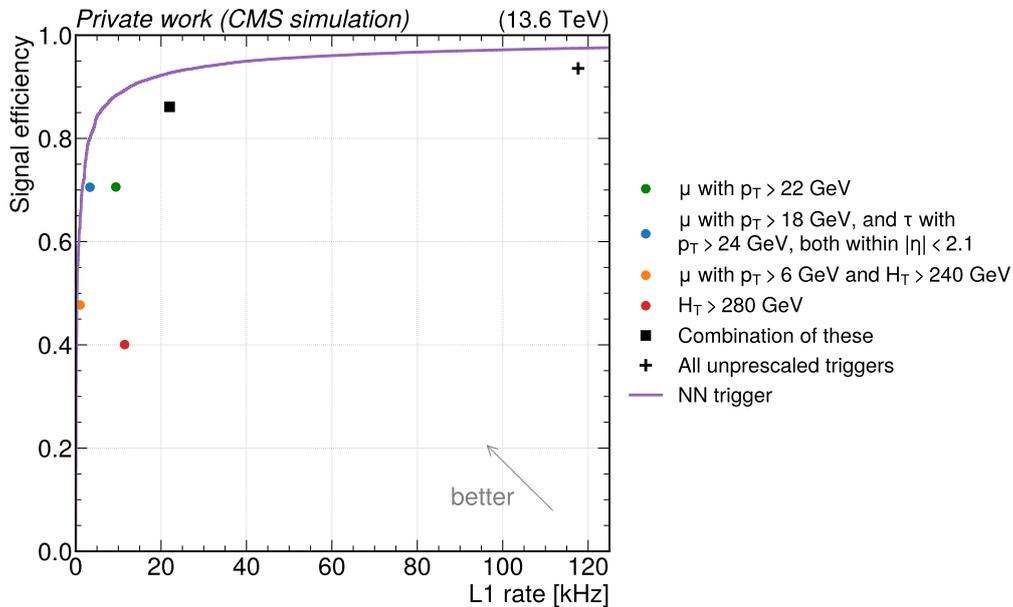


Figure 6.11: Signal efficiency and trigger rate of the NN trigger compared to several cut-based triggers, evaluated on $HH \rightarrow bbWW$ (single muon) and low energy background events.

(HLS) language. The firmware produced by `hls4ml` needs to fulfill the strict latency constraints of the CMS L1T. For the targeted Run 3 configuration, a GT algorithm has about 100 ns to make its decision, which corresponds to 4 clock cycles of the FPGA, which runs at the same clock of 40 MHz as the LHC.

FPGAs are equipped with limited resources of different types. Two basic elements of FPGAs are *lookup tables* (LUT) and *flip-flops* (FF). In LUTs, sets of input boolean variables can be mapped to a single boolean output using a customizable table. This way, complex relations can be calculated without dedicated implementations. FFs allow bit storage, analogous to regular processors. A more specialized part of FPGAs are *digital signal processors* (DSP). These can perform a specific, complex mathematical operation, for example a multiplication, much faster than an array of LUTs and FFs. Thus, they are especially important when implementing ML algorithms on FPGAs, as these mainly consist of multiplication operations. Finally, *block random access memory* (BRAM) provides memory on the FPGA itself. When deploying a firmware, the usage of each of these resources can be estimated. I need to ensure that the given algorithm fits the FPGA it is deployed to. Additionally, in CMS, multiple algorithms share the same board, thus making the algorithm as small as possible is crucial.

6.3.1 Compressing the neural network

A main factor that drives the resource usage and latency of a NN is its architecture, which is why the L1 NN is optimized to be as small as possible. Specifically, the latency mostly depends on the number of layers a network has: as these always have to be evaluated consecutively, the NN will always need at least one clock cycle per layer. Including output layer, the NN I optimize above passes the inputs through two layers, which determines its latency to be two clock cycles or 50 ns, fast enough to fit the given constraints. For applications where more time is available, `hls4ml` allows to tune the *reuse factor*, reducing resource usage at the cost of increasing latency by using a single FPGA element multiple times. As the latency needs to be as low as possible for the presented application, I do not use this method here.

Obviously, the resource utilization of a NN is also driven by its architecture, with smaller NNs requiring fewer resources. However, there are some other handles available to compress NNs when deploying to the L1T. FPGAs only perform fixed-point integer operations, using the so-called `ap_fixed` type. It represents a floating-point number using fixed number of total bits, of which another fixed number is used for the bits before the decimal point. An example is this number:

$$\begin{array}{ccc} \text{integer} & \text{fractional} & \\ \text{bits} & \text{bits} & \\ \underbrace{0100}_{\text{integer bits}} & \underbrace{.011110}_{\text{fractional bits}} & \\ \underbrace{\hspace{1.5cm}}_{\text{total bits}} & & \end{array} \quad (6.1)$$

It has four *integer bits*, six *fractional bits* and thus ten total bits. I refer to numbers in this format as *fixed-point* numbers in the following. All parameters of the deployed NN are expressed in fixed-point representation, a process known as *quantization*. The number of bits per number can be tuned, where smaller bit widths result in lower resource usage. As NN weights usually are of $\mathcal{O}(1)$, a single integer bit is enough. I test how the network performance changes when tuning the number of total bits. While this value, in principle, can be adjusted separately per parameter, or per layer,

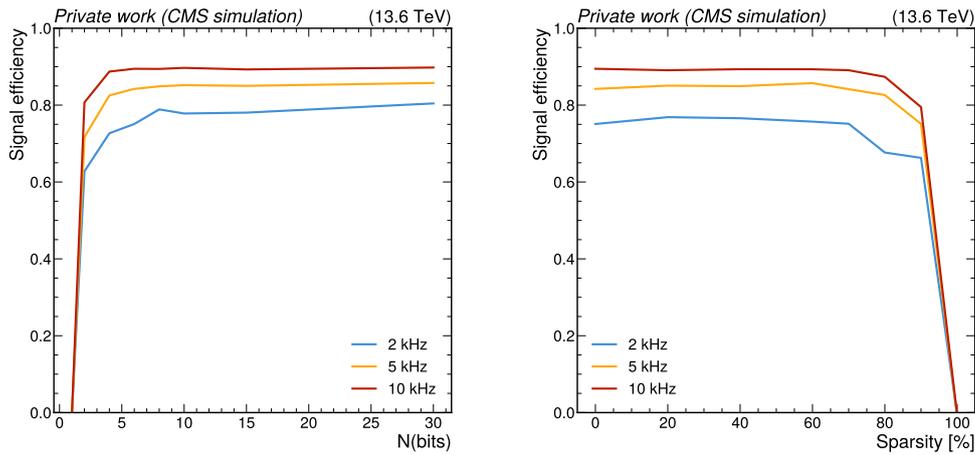


Figure 6.12: Signal efficiency at 2 kHz, 5 kHz and 10 kHz of rate as a function of total bits in a bit-fixed representation (left) and as a function of sparsity (right).

for simplicity I chose a single value for the full set of NN parameters. Crucially, performance loss might arise if a NN is trained in a floating-point representation and compression to fixed-point is done afterwards. Therefore, I utilize *quantization-aware* training using the `qkeras` [250, 251] library, including the quantization already during training to ensure the network only uses the given bit precision to learn. In figure 6.12, I show the NN performance, in terms of a signal efficiency at different rates, as a function of the number of total bits. It can be seen that only few bits are actually needed, the NN can be safely compressed down to six total bits without significant performance loss, which is what I use in the following.

NNs with a fully connected dense architecture contain a large number of connections between nodes. However, many of these might actually not contribute much to the final result of the NN, which is visible in small connection weights. This feature is exploited when *pruning* a NN. When pruning, some percentage of connections with low weights have their weights set to zero. This is very beneficial when deploying a NN onto an FPGA, as these connections then do not need to be calculated. I train the fixed-point NN with varying degrees of *sparsity*, and show performance results in figure 6.12. It can be seen that the NNs performance only significantly degrades for above 80% sparsity.

I estimate the resource usage for an AMD Virtex 7 FPGA, which is the one used in the Run 2 L1T. Before optimization, a single layer network with full bit precision (thus using 32 bits per number, the same as a float, with 6 integer bits) and no pruning applied requires about 68% of the available DSPs and 13% of LUTs. The usage of FFs and BRAM is negligible. Thus, the network is large and would take up the majority of one FPGA. Using the default `hls4ml` settings without quantization-aware training or pruning, compressing the NN to 16 total bits (6 integer) per value, still results in 21% of DSPs and 6% of LUTs being used. After fully compressing the NN as described above, the usage is reduced to about 1% (about 5000) of LUTs with negligible usage of other resources. This demonstrates that it is possible to deploy extremely lightweight models that still have good performance. Thus, targeting multiple NN-based triggers for different applications is feasible.

6.3.2 Testing under running conditions

Having trained and compressed a NN, its crucial to ensure that the network can cope with actual running conditions. Therefore, I implement the L1 NN in the so-called *test crate* of the CMS L1T. This system is identical to the GT, but it does not trigger detector readout. Thus, the rate of triggers can be tested here without disrupting the experiment if rates are too high or errors arise. I train a NN targeting the $HH \rightarrow bbWW$ (single muon) signal process based on the optimization procedure outline above: having a single hidden layer with 64 nodes and using four jets and two muons as input objects. For technical reasons, the ϕ information is not removed for this test, and two e/γ objects are passed as well: this allows using the same inputs as another algorithm tested at the same time. No impact on the model performance is expected from these changes. The NN is compressed to use six total bits with one integer bit and is pruned to 80% sparsity.

In order to create a trigger from the NN, a threshold on its output needs to be defined. In order to test the NN over a wide range of output rates, multiple instances are deployed and different thresholds chosen. Estimating the rate as above, I find NN output thresholds corresponding to different target rates, as shown in table 6.1. I convert these threshold to a fixed-point representation, as needed in the test create: an integer between 0 and 1024. Crucially, these rates are arbitrarily chosen, and most of them would be much too high for a single algorithm. However, testing the algorithm at high rates is interesting to understand potentially problematic behavior when triggering with high rates.

Target rate [kHz]	NN output threshold	Threshold fixed-point representation
1 kHz	0.999	1023
10 kHz	0.984	1007
50 kHz	0.883	904
1000 kHz	0.151	154
5000 kHz	0.025	25

Table 6.1: Target rates of the deployed triggers and the corresponding threshold in floating point and fixed-point representation.

With the algorithm and thresholds defined, it needs to be integrated into the GT environment. A schematic overview of the GT is given in figure 6.13. It shows the input information received from the second calorimeter layer and global muon trigger, as well as external condition information. Some preprocessing is performed before information is passed to the algorithms calculating cut-based trigger conditions. In parallel to those, the L1 NN is situated, using the same input information. Its output is passed to the combinatorial logic block, where it is combined with the other algorithms and passed to the DAQ system.

In figure 6.14 I display the test crate rates of the different NN-based triggers during a 2023 data-taking run. There, one can see that the overall development of rates over time is comparable between the different thresholds, showing that the NN is stable independent of the threshold it is evaluated at. The rates themselves are not constant against time: an initial rise is visible, followed by a period of relatively stable rates with a jigsaw pattern, and finally slowly decaying rates until the end of the test. This behavior is similar to that observed for a single muon

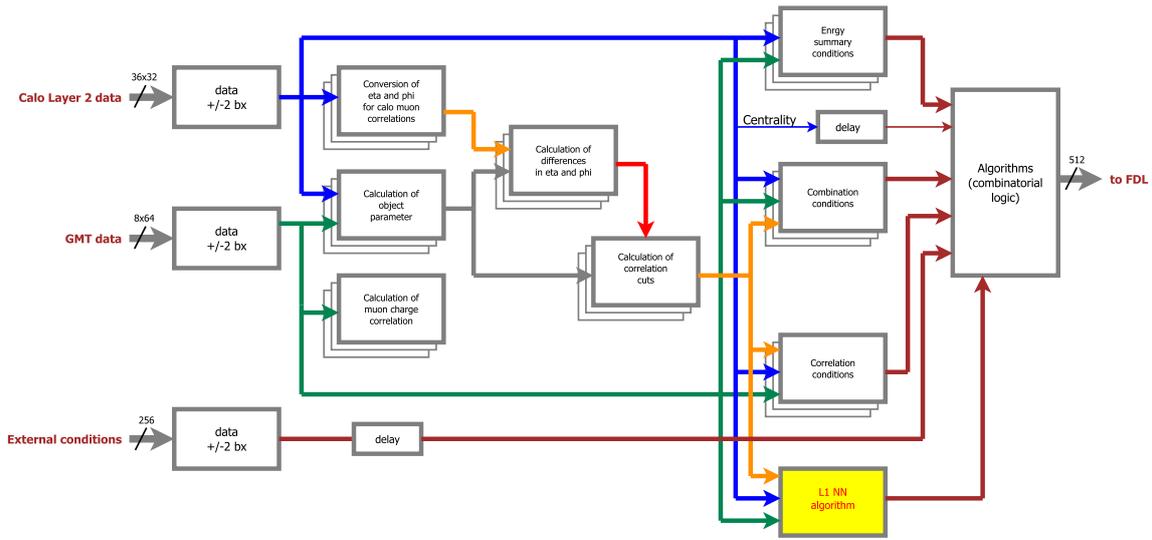


Figure 6.13: Schematic overview of the GT. The flow of data follows the colored arrows, with inputs shown on the left. The L1 NN I develop is shown in yellow, running in parallel to other trigger algorithm blocks. The GT output is seen on the right as a combination of the different algorithms. The figure is based on [252] with some adaptations.

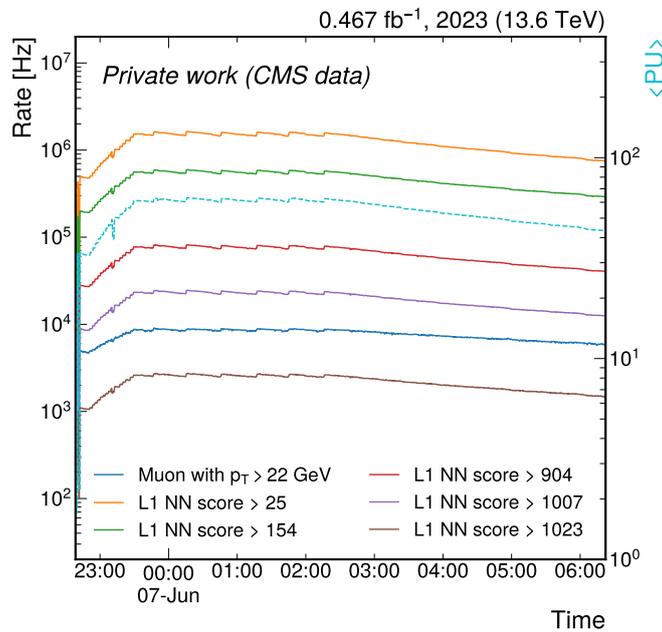


Figure 6.14: Test crate rates as a function of time for NN-based triggers at different thresholds on the L1 NN output score. For reference, a trigger requiring a single muon is shown in blue. Additionally, using the right y-axis, the average PU per LS is shown as a dashed line.

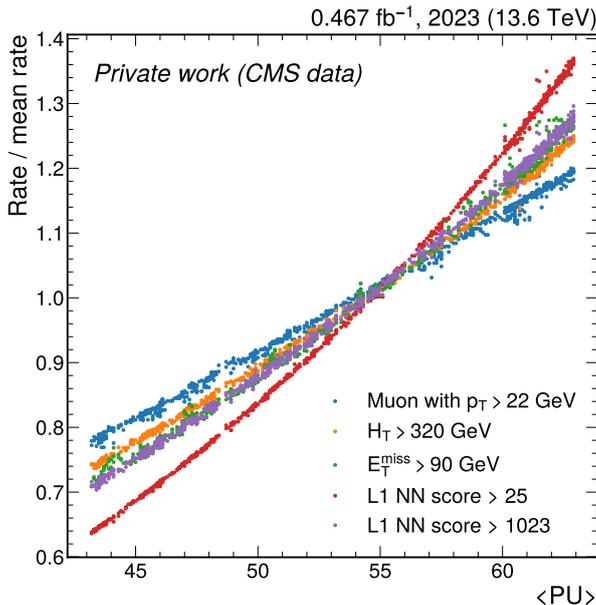


Figure 6.15: L1T rates per LS against the average PU in that LS for data from a single run during 2023 data-taking. Rates are normalized by the average rate of each trigger to make the PU-dependencies comparable.

trigger, which I include in the figure for reference. The trigger rates follow the changes in instantaneous luminosity that occur during a data-taking run: initially, the instantaneous luminosity increases until it has reached its maximum. It is held there in what is called *luminosity leveling*, where the natural decay of the beam is counteracted by periodically changed LHC settings, visible as a jigsaw pattern. Finally, after luminosity leveling, the instantaneous luminosity decays until the end of the run. All of this is visible in the per-LS PU mean $\langle \text{PU} \rangle$ that I also show in the figure, as PU scales (mostly) linearly with instantaneous luminosity, as explained in section 3.2.3.3. There, I presented results on the PU-dependency of classical trigger algorithms. In figure 6.15, I compare two thresholds of the NN-based trigger to the previously shown results. One can see that the PU-dependency of the L1 NN is close to linear. For a tight threshold on the NN score, the slope is comparable to a hadronic H_T -based trigger, whereas a loose threshold results in stronger PU-dependency. Overall, these results show that implementing a NN-based trigger is realistic, as the increase is not strongly exponential or erratic. Still, if PU dependency should become a problem, active decorrelation procedures could be used to reduce it, such as the one I use in the above search for t^* pair production.

6.4 Trigger usage and analysis integration

After demonstrating that a sensitive NN-based trigger can be developed and deployed in the CMS L1T system, it is equally important to understand how to use it in the context of the HLT and offline analysis. While the studies I present above merely are a first step towards a well-studied and optimized NN-based trigger, I discuss some ideas on its usage in this section. To test potential HLT strategies,

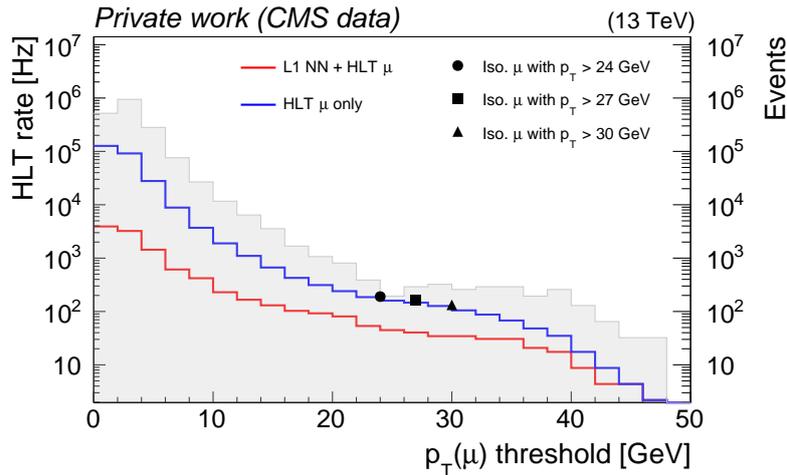


Figure 6.16: HLT rate estimate for a path requiring at least one isolated muon with p_T above some threshold, comparing a situation with (red) and without (blue) the L1 NN seeding that path. For reference, three actual single muon triggers from the HLT menu are shown as black points.

I implement a L1 NN trigger into a Run 2 $HH \rightarrow bbWW$ (single muon) analysis [253]. This way, I can evaluate the L1 NN together with offline objects, which I use as estimators for HLT objects to study different HLT strategies. While in principle, HLT objects can differ from what is reconstructed offline, they are sufficiently similar for these initial studies. As I am targeting a Run 2 analysis, I am not using the same network I introduced above. Instead, I work with a NN specifically trained on Run 2 samples following similar design principles. The NN is taken from the work in [9] and described in more detail there. I set a threshold on the NN output corresponding to a 10 kHz L1 rate in order to have a sufficient amount of events passing the trigger condition. I label the resulting trigger as *L1 NN* in the following.

6.4.1 HLT strategies

Different strategies to use the L1 NN at HLT are possible. Aside from simply passing on selected events with some prescale, or making use of the scouting or parking datastreams, I describe ideas on two HLT variants using the L1 NN below.

L1 NN as an intelligent trigger prescale The reconstruction of physics objects breaks down at very low momenta. Thus, while trigger-limited analyses like $HH \rightarrow bbWW$ (single muon) search would benefit from lower momentum thresholds, they do not require these to be removed entirely, only to loosen them. As discussed above, this is not possible due to the increased resulting trigger rates. However, this problem could be solved by seeding a single object HLT path with a L1 NN, which then would serve as an *intelligent prescale*. Similar to an actual prescale, it lowers rates and thus allows a trigger with an otherwise too loose selection criterion to be used. However, instead of randomly rejecting events, rates are suppressed while retaining high signal efficiency. This could allow, in the example of the $HH \rightarrow bbWW$ (single muon) process, a low- p_T extension of the currently used single muon trigger strategy. I study this idea by investigating the HLT rates resulting from different p_T

thresholds when requiring a single isolated muon. The results are shown in figure 6.16. I use a data sample gathered with random triggers for this study, which allows me to evaluate HLT rates. Notably, when requiring a muon with some p_T threshold, I do not consider any cut-based single muon L1 seed. This is valid, as L1 seeds usually are chosen so that they are almost fully efficiency for the events selected by a given HLT path, thus the difference is negligible. I demonstrate this by also including the rates of three actual HLT algorithms in the figure, requiring isolated muons with different p_T thresholds. These validate my rate estimation. Comparing the scenarios with and without using the L1 NN, one can see that it is possible to strongly suppress the HLT rate when seeding the HLT path with the L1 NN. Of course, as the regular single muon triggers are used by many analyses, the goal is not to replace them. Instead, this study shows that rate-wise, an additional muon trigger as a low- p_T extension could be feasible.

NN-based trigger at HLT While using a NN-based trigger as a seed of a classical cut-based HLT path is feasible, one might obtain a higher performance gain when also using a ML-based approach for the HLT. No special care on implementation is needed here: the HLT software is executed on regular computing hardware, which can execute ML models. I train a model using offline information as stand-in for HLT objects. As no strict latency and resource constraints are given, I use a larger architecture: three hidden layers with 25 nodes each. As the purpose of this test primarily is a demonstration of the workflow, no hyperparameter optimization is performed. The NN uses a single muon and four jets with full angular information, as well as p_T^{miss} and H_T as inputs. I train the network to discriminate signal from QCD multijet processes, with no selection criteria imposed. QCD events are similar to the unbiased low energy background events at CMS, but characterized by usually somewhat higher energies and more activity being simulated. I define a tight threshold onto the HLT NN targeting a reasonably low HLT rate. However, due to the limited statistical power of the utilized dataset, only a rough estimate of the rate can be made.

In figure 6.17, I show signal efficiencies of different trigger strategies for three different approaches. The currently used single muon strategy is visible, showcasing the issues it currently brings with it. I show the pure L1 NN without any HLT strategy, which has too high rates but can be used to see the effect of the HLT NN. Finally, I show the result of the HLT NN, which reduces the efficiency over the entire muon p_T range. This result does not appear promising, but does not take into account that an offline analysis always performs some cut-based selection steps to enrich signal and suppress backgrounds. Ultimately, the optimal trigger is efficient in the phase space the analysis is investigating. I therefore repeat the same study on events passing the cut-based selection of the analysis, which contains these steps:

- Exactly one muon fulfilling tight criteria on reconstruction and isolation is present within $|\eta| < 2.4$, with no p_T requirement.
- At least three small-radius jets contained in $|\eta| < 2.5$, having $p_T > 30$ GeV and fulfilling tight PUPPI criteria are present.
- At least one of these jets is b -tagged using DeepJet with a working point with 10 % misidentification rate of light jets.
- The angular distance between lepton and closest jet is $\Delta R > 0.2$.

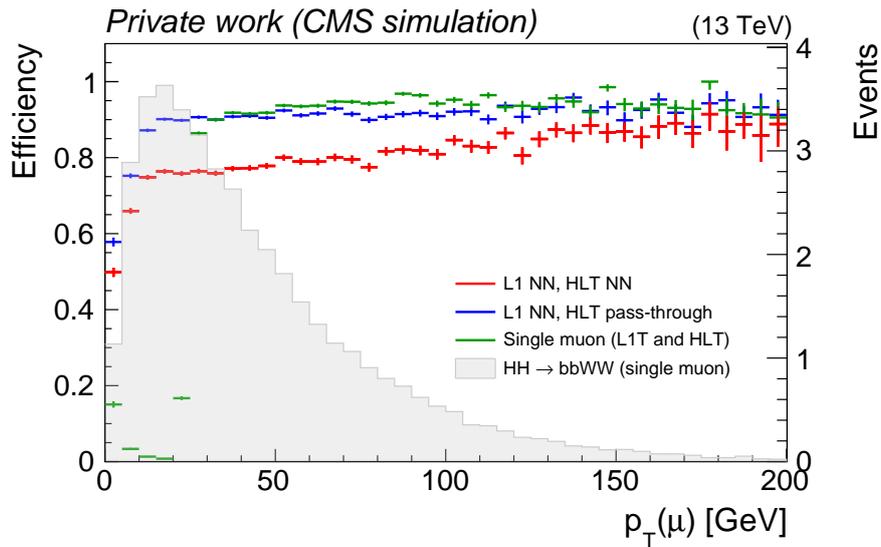


Figure 6.17: Efficiency of different trigger strategies as a function of offline muon p_T . A cut-based approach using a single muon trigger is shown in green. The blue curve shows the L1 NN assuming a HLT pass-through, and the red curve shows a setup with NNs in both the L1T and HLT. These results are obtained for simulated $HH \rightarrow bbWW$ (single muon) signal events without any selection requirements, except demanding a muon to be present. The muon p_T distribution in this process is shown in gray using the right y-axis, with arbitrary scaling.

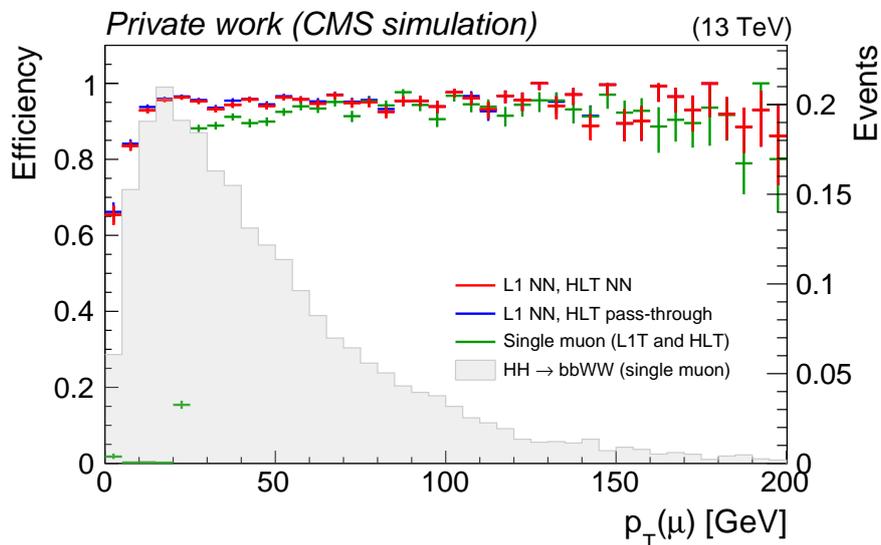


Figure 6.18: Efficiency of different trigger strategies as a function of offline muon p_T . A cut-based approach using a single muon trigger is shown in green. The blue curve shows the L1 NN assuming a HLT pass-through, and the red curve shows a setup with NNs in both the L1T and HLT. These results are obtained for simulated $HH \rightarrow bbWW$ (single muon) signal events after a cut-based selection. The muon p_T distribution in this process is shown in gray using the right y-axis, with arbitrary scaling.

- A transverse mass, defined as the invariant mass of the lepton, p_T^{miss} and the b jet on the leptonic b quark leg is greater than 60 GeV.

Full details on these selection steps can be found in [253]. The resulting efficiencies are shown in figure 6.18. Notably, the effects of the cut-based trigger and pure L1 NN do not change much. In contrast, the efficiency of the HLT NN strategy becomes much higher. The reason for this is the training of the HLT NN: it is designed to identify signal and suppress the QCD multijet process, which is also a goal of the cut-based selection. This shows that by utilizing the fact that offline analysis will anyway never use all available signal events, a HLT strategy that is highly efficient on the actually relevant events could be defined while still keeping low rates.

Notably, these studies are only a technical demonstration of investigating ML-based L1 and HLT trigger strategies in an offline analysis context. The 10 kHz threshold on the L1 NN is too high to actually be used, both due to the limited overall rate budget and the fact that a HLT NN would need to be evaluated for each event passing the L1 NN: This would most likely go beyond the available event processing time at HLT level. Further studies on improving the L1 NN and HLT strategy are needed, including porting this work to the currently ongoing analysis of Run 3 data. Also, a larger dataset or different strategy to estimate HLT rates is needed.

6.4.2 Efficiency measurement

For any given trigger, being able to measure its efficiency is crucial, which is also true for NN-based algorithms, regardless of whether a full L1T and HLT strategy or L1T only is considered. Aside from simply knowing whether a trigger is well-suited for a given task, a precise knowledge of trigger efficiencies in data is important to obtain correct results on cross sections and other parameters in offline analysis. In section 3.2.3.4, I introduce different methods to measure the efficiency of a trigger. However, not all of these are well-suited for NN-based triggers, as I briefly discuss in this section. Measuring a NN-based trigger rate using data gathered with random triggers is possible without further caveats. Unfortunately, the same limits as for all other triggers apply: usually, random triggers do not supply enough data for this method to work, especially when considering low-rate, unrescaled triggers used in offline analysis. The tag and probe is also not valid here, as it can only measure object-level efficiencies and thus can not be applied to a ML-based trigger.

This is different for the orthogonal datasets method, which is applicable as long as a truly orthogonal dataset can be found. For a NN trained on all available object types (jets, energy sums, e/γ s and muons), which is also under study, there is no independent object that could be used to define an orthogonal dataset. However, even when some objects are excluded from training, like e/γ s in case of the muon channel L1 NN I use above, orthogonality is not guaranteed. While for classical triggers, knowledge of what happens inside of the algorithm often allows to find an independent condition, NNs inherently can not be understood as easily. Even if a NN was not trained on e/γ objects, it might show correlation with their presence (or absence) in an event. I briefly study this based on an investigation of the α parameter describing the correlation of two triggers. To find a dataset that is similar to the process targeted by an analysis using the trigger, but has an electron in addition to the other objects, I utilize $t\bar{t}$ events with two leptons. I calculate the α parameter for different combinations of cut-based triggers and a threshold on the L1 NN, and

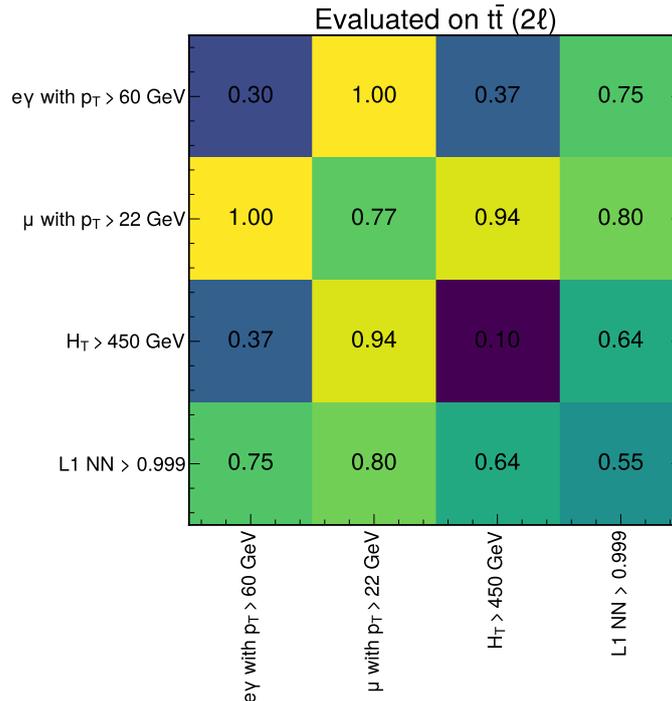


Figure 6.19: Correlation parameter α for different trigger combinations, evaluated on a simulated dataset of $t\bar{t}$ events with two leptons. The diagonal, which by design is fully correlated as a trigger is compared to itself, shows a property of α : it has a lower limit depending on the efficiency of the tested triggers.

show results in figure 6.19. One can see that little correlation is present between an e/γ and muon trigger, with an α of exactly one. For the NN-based trigger, the situation is worse: despite not being trained on this object, correlation with the e/γ trigger is visible. Notably, I do not impose any offline selection requirements, which might improve or worsen results. Similarly, I only investigate di-lepton $t\bar{t}$ events, which could introduce an inherent bias from the process itself. An example is the visible correlation between the e/γ and H_T trigger. When one of the t quarks is highly energetic, this will result in an energetic jet and thus high H_T , but also an efficiency triggered energetic e/γ . Ultimately, this shows that the orthogonal dataset method can not trivially be used for a NN-based trigger, even if unrelated objects are considered. Still, using decorrelation techniques like the one I use in the $t^*\bar{t}^*$ search in chapter 5, an implementation of the orthogonal dataset method potentially could be made possible, but might in return result in worse performance of the trigger algorithm.

The most promising approach to enable efficiency measurements of a NN-based trigger is the utilization of a reference trigger. As I mention above, offline analysis does not require momentum thresholds to be removed entirely. Thus, a trigger strategy that combines a momentum-threshold based trigger with a NN-based approach could be optimal: requiring a muon with some loose momentum threshold, which would have too high rates, and lowering these using a NN. When constructing a trigger like this, the efficiency of the NN requirement could then be measured relative to the single muon condition in a dataset gathered by a prescaled version of the single muon trigger.

Still, all of these techniques suffer from another issue: the efficiency of a NN could

vary depending on some unknown set of parameters. Usually, when measuring the efficiency of a trigger, this is performed in bins of relevant variables. For example, the efficiency of triggering on muons might vary depending on their momentum and η direction, thus SFs are measured in bins of muon p_T and η . If the efficiency of a DNN varies as function of some high-dimensional, unknown set of quantities, finding a proper binning strategy to avoid biases is challenging. Most likely, a better approach would be to measure efficiencies in each bin of the main observable of an analysis. Even if other correlations and biases exist, these will then not affect the final result of the analysis.

6.5 Outlook

In summary, I present studies towards the usage of ML-based trigger algorithms in the CMS L1T, which will be an important part of the HL-LHC CMS trigger strategy, as proposed in [156]. I demonstrate a large part of the workflow needed for a ML-based L1T algorithm: conceptualization, optimization, compression and deployment, as well as ideas towards usage in HLT and offline analysis. However, they merely are a starting point for future work in this area. While I demonstrate that a rather generic, very small network can achieve good results, more sophisticated techniques might yield even better performance. Further studies on the HLT strategy and efficiency measurement are needed in order to decide on a strategy and enable effective usage in the offline analysis. For this, ongoing analysis of Run 3 data can be used as a test environment, and an implementation of a L1 NN in Run 3 could provide more information and enable tests of efficiency measurement techniques.

Updates to the NN architecture could be made to optimize it for the L1T usage. An example is the output score, which currently is the result of a `sigmoid` activation function, ensuring it lies between 0 and 1. While this is very suitable for generic NN applications, which are usually operated at thresholds in the intermediate range of possible output scores, this is different for usage in the trigger system. Due to the strict latency constraints, the NN is operated in a regime of very low FPR, which correspond to NN thresholds very close to one. On one hand, this feature is not reflected in the loss function, thus during training the loss might decrease without any gain in the relevant region. Secondly, this imposes a problem when quantizing the output score: a large part of the available bit width is unused, which only allows for a large step size in the relevant rate range. Both issues could be solved by using a different activation function for the output layer which emphasizes the relevant region, for example with an exponential component.

Finally, supervised ML is not the only option at trigger level. In contrast to classical cut-based triggers, which usually are rather generic, the presented NN is only trained for a specific signal. While this is well motivated for highly important but trigger-limited processes like the production of two Higgs bosons, it might be a disadvantage compared to more generic cut-based triggers. To test whether a NN-based trigger is usable more generally, I study the efficiency of a network targeting the $HH \rightarrow bbWW$ (single muon) process on different simulated samples. I find good sensitivity also for other signals with similar final states, as I show in figure 6.20 for a single muon $t\bar{t}$ sample. More similar studies can be found in [9]. Still, the application is limited to certain final states, and issues of HLT strategy and efficiency measurement

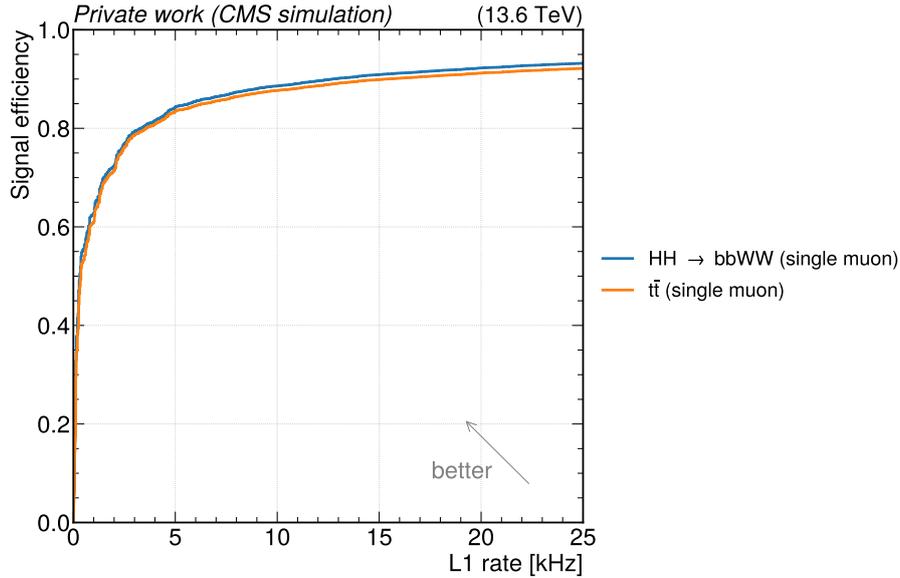


Figure 6.20: Comparison of ROC curves for the $HH \rightarrow bbWW$ (single muon) process, which the NN is trained on, and a single muon $t\bar{t}$ sample, previously unseen to the NN.

remain.

This motivates alternative strategies, multiple of which are under study in CMS: ML is well suited for anomaly detection applications, in which an unsupervised training allows model-independent searches for unusual signatures. An anomaly detection algorithm called AXOL1TL [254] takes data since May 2024 [255]. It is based on a VAE and situated in the GT, thus using the same input objects as the L1 NN I present here. Conceptually, the L1T is tasked to select relevant events from a large amount of similar, low-energy events. Therefore, extending it with ML-based anomaly detection is natural. In comparison to the L1 NN I present above, AXOL1TL is not limited to a specific signal and simply selects events that are different from the majority of events occurring in CMS. However, it shows a much stronger dependency on PU compared to the L1 NN, and analysis usage as well as efficiency measurement strategies are challenging. Further studies on this algorithm are needed and ongoing. Another complementary anomaly detection approach in the calorimeter trigger is also in development within CMS. The CICADA algorithm [256] uses a convolutional autoencoder approach based on calorimeter images, and selected anomalous hadronic signatures.

Especially towards the HL-LHC phase of CMS, as described in appendix A.1.3, ML has huge potential in the context of the L1T. The increased luminosity, going along with a targeted 200 average PU collisions, will pose additional challenges to trigger algorithms. These can be addressed by ML-based algorithms like the ones mentioned above, but also in other areas of the L1T. The L1T will have access to tracking information and use a PF algorithm similar to the one utilized offline. Thus, efficient vertex reconstruction and identification of the PV is crucial. An end-to-end NN approach, which combines regression of the PV position and determination of the likelihood of a track to originate from it, yields promising results [257]. Similarly, L1 object reconstruction and identification can be improved by using ML: muon track finding and momentum determination can be improved through usage of NNs

in the barrel [258] and endcap regions [259]. Identification of hadronically decaying tau leptons is challenging, and a novel approach with two NNs, the `TauMinator` algorithm, promises improved performance [260]. In offline analysis, identification of jets originating from b quarks is highly important. As this strongly relies on tracking and vertex information, it is currently impossible in the L1T. With tracking being added to the L1T, ML can be utilized to develop sensitive b -tagging algorithms for trigger usage [261]. These examples show the wide range of possibilities that ML in the L1T offers. The current studies can only be seen as a starting point into this realm, gathering more information and knowledge which will become crucial towards the HL-LHC phase of CMS.

In this thesis, I present my work on different aspects of high energy particle physics research with the CMS experiment. Well motivated by various models on BSM physics, I probe 138 fb^{-1} of proton-proton collision data at a centre-of-mass energy of 13 TeV for the existence of a potential new particle t^* . This particle, predicted to dominantly decay as $t^* \rightarrow tg$, can be produced in pairs at the LHC, motivating a search for the $t^*\bar{t}^* \rightarrow tg\bar{t}g$ process. I select events containing a single electron or muon, jets and missing transverse momentum. The momentum sum S_T , sensitive to the presence of t^* pair production, is used for statistical analysis. I define a signal region enriched in potential signal events using a DNN, which I carefully decorrelate against S_T in a two-step procedure. Results agree with SM predictions, therefore I set upper cross section exclusion limits at 95 % confidence level onto the production of $t^*\bar{t}^*$ pairs. Analyzing two different t^* spin scenarios, the upper limits range from 120 fb to 0.8 fb for a t^* with spin- $\frac{1}{2}$ and from 15 fb to 1.0 fb for a t^* with spin- $\frac{3}{2}$. Comparing to theory predictions of the $t^*\bar{t}^*$ pair production cross section, assuming a 100 % branching fraction of the $t^* \rightarrow tg$ decay, this corresponds to mass exclusion limits up to 1050 GeV and 1700 GeV for spin- $\frac{1}{2}$ and spin- $\frac{3}{2}$ t^* particles, respectively. These exclusion limits substantially improve previous results and thus further deepen our understanding of the smallest particles. Despite not discovering a sign of new physics here, with each tightening of an exclusion bound we are getting closer to a more fundamental understanding of reality. In addition to this search for a signature of new physics, I present my participation in the current operation and future improvement of the CMS experiment itself. I develop tools to validate current triggers and study newly added triggers during Run 3 data taking in order to ensure that optimal data is available for analysis. Furthermore, I present studies towards the usage of ML in the CMS L1T, which will become crucial especially in the HL-LHC phase of CMS. Using the ongoing Run 3 as a test environment, I demonstrate that it is possible to utilize a NN-based algorithm in the L1 trigger. It is able to select events of a currently trigger-limited process, $HH \rightarrow bbWW$ (single muon), more efficiently than classical approaches. This could allow more data to be used in offline analysis and thus improve the sensitivity to this highly relevant process. More importantly, in this context I demonstrate that very small NNs can yield good results, and that it is possible to integrate them into the L1T hardware environment. While these results are a step forward in high energy particle physics research, there are many opportunities for future work: Run 3 data, with higher centre-of-mass energy and more integrated luminosity most likely will enable an even more sensitive search for the t^* particle. This will be supplemented by more sophisticated analysis techniques being developed, and CMS reconstruction being improved. Similarly, great potential lies in the continuous development of ML-based L1 trigger strategies, using the presented work as a baseline: better, larger NN architectures, different approaches and further work on analysis integration of ML-based triggers are exciting prospects.

A.1 Description of the experiment and methods

In this appendix section, I provide some supplemental information on the experimental setup that is relevant to some parts of this thesis, and some additional detail on the methods used.

A.1.1 LHC Run 3 upgrades

After Run 2, some changes were made to the LHC and its pre-accelerator chain. Instead of Linac2, the newly built *Linac4* [262] accelerates the protons (H^- ions) as the first step towards supplying protons to the LHC since 2020.

The main other change in the Run 3 LHC configuration is the increase in $\sqrt{s_{pp}}$ to 13.6 TeV. Apart from that, preparations for the HL-LHC phase were made, as well as many smaller changes to ensure stable operation during Run 3. A comprehensive overview is given in [263]. A change relevant to CMS is the introduction of *luminosity leveling*. During a proton *fill* of the LHC, after ramping up the energy to 6.8 TeV per beam, the instantaneous luminosity is kept constant for a large part of the fill. This is achieved by periodically changing beam conditions, focusing the beam more strongly to counteract luminosity decrease due to the constantly lowering intensity of the beam.

A.1.2 CMS Run 3 configuration

The Run 3 configuration [264] of CMS is very similar to the Run 2 setup. The most notable change is the increase in $\sqrt{s_{pp}}$ from 13 TeV to 13.6 TeV. Most detector subsystem underwent some smaller updates in order to retain and improve their performance. Changes include the addition of a gas electron multiplier detector to the muon system, as well as upgrades to other detector parts, electronics and the trigger. However, none of these changes result in major differences to the Run 2 detector configuration I describe in section 3.2.

The reach of the HLT menu is extended by *parking* and *scouting* [265] in Run 3. A major factor in the limited HLT output rate is the inability to perform offline reconstruction for too many events. Thus, events are selected by the HLT, but not reconstructed immediately. Instead, this data is parked in a raw format for later reconstruction. For events in the scouting data stream, only a reduced event content is stored, but raw event information is not available. These two approaches allow to use the available output rate of the trigger system more effectively, and are a place where specialized triggers can enable previously impossible analyses.

A.1.3 CMS HL-LHC configuration

The HL-LHC [152] phase of CMS is planned to start around 2030. It is primarily characterized by an increase of $\sqrt{s_{pp}}$ energy to 14 TeV, as well as substantially higher instantaneous luminosity than currently achieved: targeting $7.5 \times 10^{-5} \text{ fb}^{-1} \text{ s}^{-1}$. Over the full duration of the HL-LHC phase, a total integrated luminosity of 3000 fb^{-1} of data is planned to be collected. To cope with the increase luminosity and the resulting increase in PU, upgrades to all detector subsystems are necessary.

For the tracking system [266], the ability to withstand radiation will be crucial. Fully replacing it, the reach of the tracking system will be extended to $|\eta| < 4$, with four PIXEL barrel layers, surrounded by six strip layers and a sophisticated endcap strategy. With smaller pixel sizes, granularity will be increased to cope with the large amount of tracks arising at 200 PU interactions, and enable precise vertex finding. The readout electronics of the calorimeters [267] and muon system [268] will be updated and measures taken to ensure their longevity. Additionally, a new calorimeter will be installed in the endcap region: the high granularity calorimeter (HGCAL) [269]. Close to the beam pipe, radiation is especially strong, and the currently installed calorimeters can not cope with the planned integrated luminosity of the HL-LHC. The HGCAL is constructed from a combination of silicon sensors (electromagnetic part) and plastic scintillator tiles (hadronic part). It will combine high radiation tolerance with high granularity, both laterally and longitudinally. Additionally, shower timing information will be available. Thus, the HGCAL is expected to have good energy resolution and will be able to aid in PU suppression and vertex identification. Another planned new detector part is a minimum ionizing particles (MIP) timing detector [270]. A precise measurement of the moment a particle reaches CMS can help to identify whether it originated from PU due to the non-zero longitudinal spread of colliding bunches. Therefore, the detector targets timing measurements to a precision of 30 ps to 40 ps. In addition to PU suppression, this detector could be used to identify long-lived particles or heavy charged particles with long flight times, which appear in some BSM theories.

Finally, in addition to these changes to the detector itself, the DAQ and trigger systems will be updated [271]. To enable appropriate performance of the L1T when confronted with 200 PU collisions, improvements are needed. I show a schematic of the updated L1T system in figure A.1. A major change to the current configuration is the availability of tracking information at this level, using the outer tracking layers. Information from the HGCAL is also included. The L1T output rate will be increased from 100 kHz to 750 kHz, and the latency increases to 12.5 μs . These changes allow for the inclusion of a particle flow reconstruction in the L1T in two dedicated layers, the *correlator trigger*. Faster and larger FPGAs will be used in form of the AMD Xilinx Virtex VU9P instead of the currently used AMD Virtex 7 architecture. Updates to the HLT and DAQ system [272] will allow for a higher HLT output rate (5 – 7.5 kHz). The total size of each event is also increased and more rate is allocated to scouting applications, with a total data stream of 60 TB s^{-1} being targeted.

A.1.4 Rapidity and pseudorapidity

In section 3.2.1, the pseudorapidity η is defined. As the anti- k_t jet clustering algorithm instead uses the rapidity y , I provide its definition and the relation to η here.

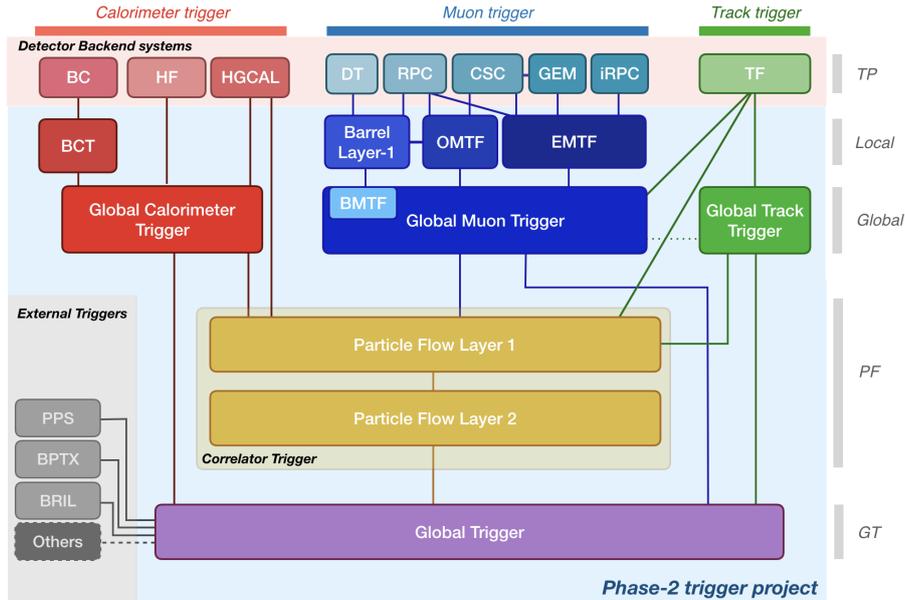


Figure A.1: Diagram of the HL-LHC setup of the CMS L1T. The calorimeter, muon and tracking inputs and preprocessing parts are shown in red, blue and green, respectively. The correlator trigger is visible in yellow, and the global trigger in purple, which is analogous to the GT in the Run 2 and Run 3 configuration of the L1T. Taken from [271], where more details on this figure are available.

The pseudorapidity can be written as a function of the 3-momentum of a particle as

$$\eta = \frac{1}{2} \ln \frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \quad (\text{A.1})$$

for an object with momentum vector \vec{p} with the component p_z in beam direction. In contrast, the rapidity y is defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \quad (\text{A.2})$$

Assuming the particle is highly relativistic,

$$m \ll |\vec{p}| \implies E \approx |\vec{p}| \quad (\text{A.3})$$

follows, where E is the particles energy. Thus, the rapidity and pseudorapidity are the same in the relativistic limit, which is valid for all particles at the LHC.

A.1.5 A loss function using cross-entropy

In this section I provide information on the construction of the BCE loss and show why the Y_i^{pred} values are required to be probabilities, based on [193]. Assuming some distribution $q = q(Y_i^{\text{true}})$ of Y_i^{true} values that the NN is supposed to predict, the entropy $H(q)$ of this distribution can be defined as

$$H(q) = - \sum_{c=1}^C q(Y_c) \log(q(Y_c)) \quad (\text{A.4})$$

where c is one of C classes (of which in the case of a binary network, there are two), and Y_c is the number of Y_i^{true} in class c . The entropy can be calculated if the underlying distribution q is known. If it is not, it can be approximated using another distribution p , as

$$H_p(q) = - \sum_{c=1}^C q(Y_c) \log(p(Y_c)) \quad (\text{A.5})$$

which is called the *cross-entropy*. Assuming p described q perfectly, $H(q)$ and $H_p(q)$ will be the same, otherwise, the cross-entropy will always be larger:

$$H_p(q) - H(q) \geq 0. \quad (\text{A.6})$$

Thus, by minimizing the cross-entropy $H_p(q)$, the optimal description p of the underlying true distribution q can be found, which is exactly the purpose of a NN. Rewriting the definition of the cross-entropy from equation A.5 for two classes that are 0 or 1, one obtains

$$\text{BCE} = -Y^{\text{true}} \log(p) + (1 - Y^{\text{true}}) \log(1 - p). \quad (\text{A.7})$$

from which, with $p = Y^{\text{pred}}$ and by summing over multiple n events i (and including a normalization term $1/n$), the BCE definition in equation 4.7 is obtained.

A.2 Search for pair production of t^*

In this appendix section, I provide details on the search for $t^*\bar{t}^*$ pair production that is not strictly necessary for an overview of the analysis, but given here in order to ensure reproducibility.

A.2.1 MC simulated samples

In table A.1, I provide information on all simulated SM processes, alongside their predicted cross sections and the number of events that are generated. Note that this number does not correspond to the actual number of events from these processes expected in a year, as usually more events are generated to avoid uncertainties due to low statistical power in the simulated samples, and weights are used to adjust their contribution accordingly. Similarly, information on signal samples is provided in table A.2.

A.2.2 Electron trigger scale factors

In section 5.3.5.1, I present the efficiency measurement and SF derivation for the electron triggers used in the analysis. In this appendix section, I provide the remaining efficiencies and resulting SFs for all years of data taking that are not shown in the main body, as these are not publicly available anywhere else. This includes all measured trigger efficiencies in figure A.2 and the resulting SFs in figure A.3.

Dataset	$\sigma[\text{pb}]$	$N(\text{events}) / 10^6$			
		2016 pre-VFP	2016 post-VFP	2017	2018
$t\bar{t} (1\ell)$	365.2	39 046.2	42 834.0	104 665.7	140 602.8
$t\bar{t} (2\ell)$	87.56	2680.29	3091.96	7545.28	10 319.11
$t\bar{t} (0\ell)$	381.1	30 289.3	33 636.2	72 581.8	105 390.9
single t ($tW, 1\ell$)	19.47	114.97	111.07	262.33	395.20
single t ($\bar{t}W, 1\ell$)	19.47	113.76	127.02	267.32	379.83
single t (tq, t -channel)	136	5839.19	6550.10	13 637.10	18 666.37
single t ($\bar{t}q, t$ -channel)	80.95	1951.75	1917.63	4382.09	6065.21
single top ($tq, 1\ell, s$ -channel)	3.364	19.34	19.26	48.68	67.08
W +jets ($1\ell, H_T \in [70, 100)\text{GeV}$)	1271	16.82	19.17	43.97	65.41
W +jets ($1\ell, H_T \in [100, 200)\text{GeV}$)	1253	21.27	19.48	46.73	50.90
W +jets ($1\ell, H_T \in [200, 400)\text{GeV}$)	335.9	17.59	14.68	41.67	57.39
W +jets ($1\ell, H_T \in [400, 600)\text{GeV}$)	45.21	2.47	2.09	5.42	7.24
W +jets ($1\ell, H_T \in [600, 800)\text{GeV}$)	10.99	2.29	2.19	5.37	7.53
W +jets ($1\ell, H_T \in [800, 1200)\text{GeV}$)	4.936	2.49	2.06	5.06	7.14
W +jets ($1\ell, H_T \in [1200, 2500)\text{GeV}$)	1.156	2.07	2.06	4.86	6.43
W +jets ($1\ell, H_T \in [2500, \infty)\text{GeV}$)	0.026 23	0.81	0.71	1.19	2.08
diboson (WW)	75.91	15.74	15.80	15.49	15.46
diboson (WZ)	27.56	7.91	7.54	7.79	7.87
QCD ($H_T \in [50, 100)\text{GeV}$)	1.86×10^8	35.73	11.08	26.03	38.23
QCD ($H_T \in [100, 200)\text{GeV}$)	2.36×10^7	65.50	72.64	53.30	82.21
QCD ($H_T \in [200, 300)\text{GeV}$)	1.55×10^6	17.97	42.72	42.32	56.30
QCD ($H_T \in [300, 500)\text{GeV}$)	3.24×10^5	13.59	45.50	42.91	60.99
QCD ($H_T \in [500, 700)\text{GeV}$)	3.03×10^4	55.50	15.07	35.75	48.64
QCD ($H_T \in [700, 1000)\text{GeV}$)	6440	15.24	13.71	33.65	47.93
QCD ($H_T \in [1000, 1500)\text{GeV}$)	1118	13.56	12.42	10.14	14.24
QCD ($H_T \in [1500, 2000)\text{GeV}$)	108	9.66	9.24	7.53	10.75
QCD ($H_T \in [2000, \infty)\text{GeV}$)	22	4.83	4.84	4.09	5.28
DY ($H_T \in [70, 100)\text{GeV}$)	140.1	6.57	5.85	11.97	16.65
DY ($H_T \in [100, 200)\text{GeV}$)	140.2	9.45	8.25	18.46	25.63
DY ($H_T \in [200, 400)\text{GeV}$)	38.4	5.75	5.58	12.23	17.92
DY ($H_T \in [400, 600)\text{GeV}$)	5.213	2.65	2.49	5.38	8.69
DY ($H_T \in [600, 800)\text{GeV}$)	1.266	2.63	2.25	5.18	6.92
DY ($H_T \in [800, 1200)\text{GeV}$)	0.5684	2.39	2.32	4.41	6.49
DY ($H_T \in [1200, 2500)\text{GeV}$)	0.1332	2.12	1.97	4.68	5.95
DY ($H_T \in [2500, \infty)\text{GeV}$)	0.002 98	0.72	0.70	1.36	1.90

Table A.1: Simulated SM background processes used in this analysis. Shown are the names of all datasets, alongside their cross section (including branching fraction in case a specific final state is simulated). Event counts for the different considered eras are given.

Dataset	$N(\text{events})$			
	2016 pre-VFP	2016 post-VFP	2017	2018
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 700$ GeV	108 000	92 000	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 700$ GeV	108 000	92 000	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 800$ GeV	108 000	92 000	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 800$ GeV	108 000	92 000	198 000	200 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 900$ GeV	108 000	92 000	197 000	200 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 900$ GeV	108 000	92 000	200 000	199 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1000$ GeV	108 000	92 000	200 000	191 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1000$ GeV	107 998	91 061	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1100$ GeV	108 000	92 000	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1100$ GeV	108 000	90 130	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1200$ GeV	96 000	92 000	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1200$ GeV	108 000	89 201	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1300$ GeV	108 000	92 000	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1300$ GeV	106 144	91 056	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1400$ GeV	108 000	92 000	200 000	197 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1400$ GeV	108 000	90 112	200 000	198 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1500$ GeV	108 000	92 000	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1500$ GeV	108 000	92 000	200 000	200 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1600$ GeV	162 000	138 000	276 000	300 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1600$ GeV	162 000	137 058	300 000	298 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1700$ GeV	162 000	138 000	297 000	297 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1700$ GeV	160 398	138 000	298 998	297 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1800$ GeV	138 000	127 000	281 000	300 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1800$ GeV	162 000	138 000	292 000	300 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 1900$ GeV	158 000	138 000	300 000	300 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 1900$ GeV	160 136	138 000	293 000	300 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 2000$ GeV	216 000	184 000	376 000	391 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 2000$ GeV	213 189	182 116	400 000	399 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 2250$ GeV	216 000	184 000	400 000	400 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 2250$ GeV	216 000	184 000	400 000	383 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 2500$ GeV	212 000	184 000	373 000	364 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 2500$ GeV	204 732	184 000	399 000	392 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 2750$ GeV	214 000	184 000	391 000	379 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 2750$ GeV	208 472	184 000	400 000	397 000
$t^*\bar{t}^*$, spin- $\frac{1}{2}$, $m_{t^*} = 3000$ GeV	212 000	184 000	364 000	391 000
$t^*\bar{t}^*$, spin- $\frac{3}{2}$, $m_{t^*} = 3000$ GeV	216 000	184 000	400 000	399 000

Table A.2: Simulated signal event counts per spin scenario and t^* mass combination.

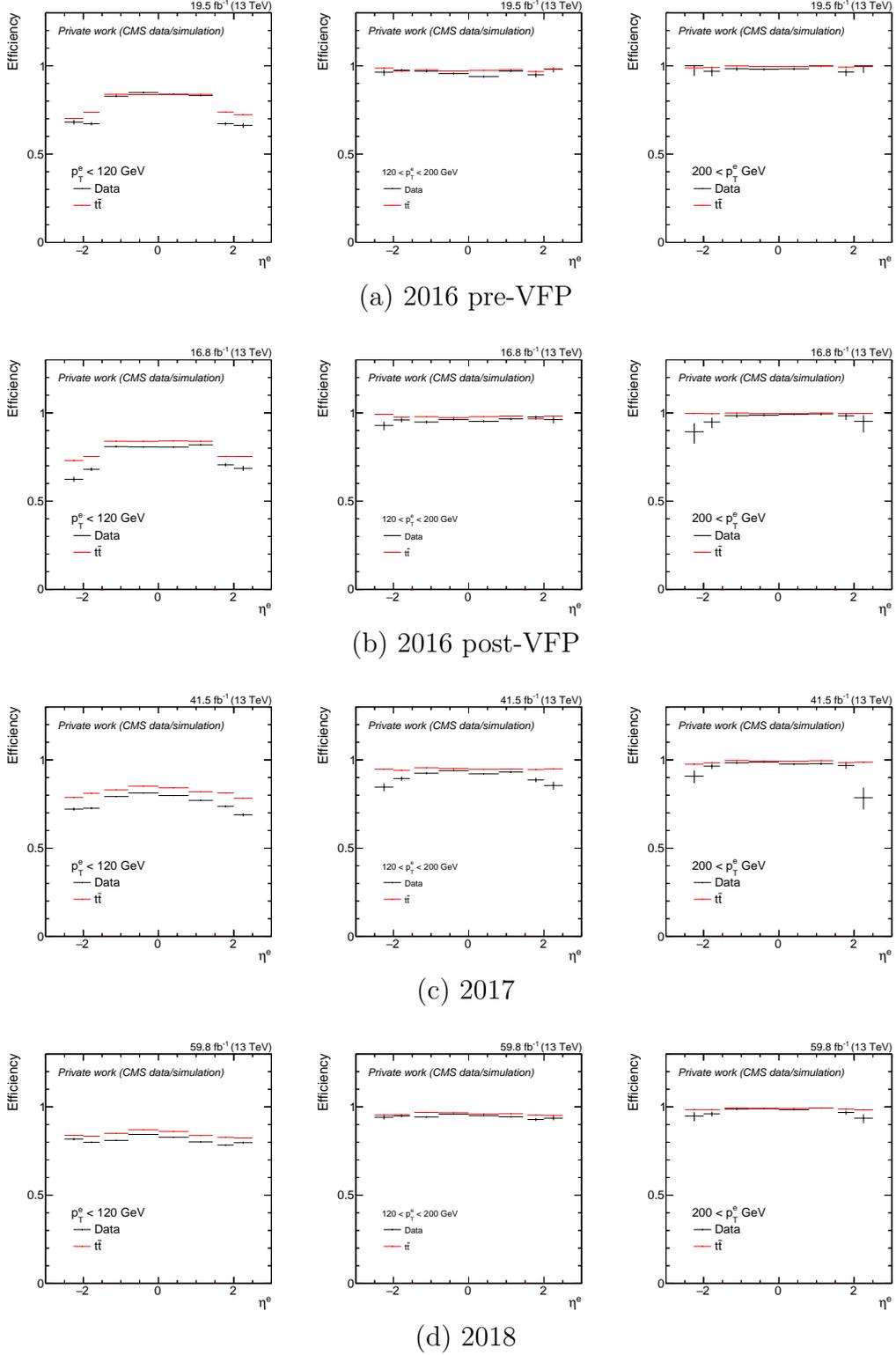
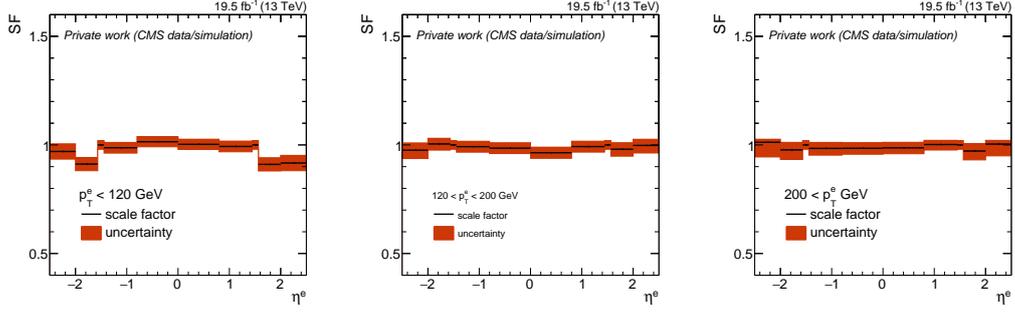
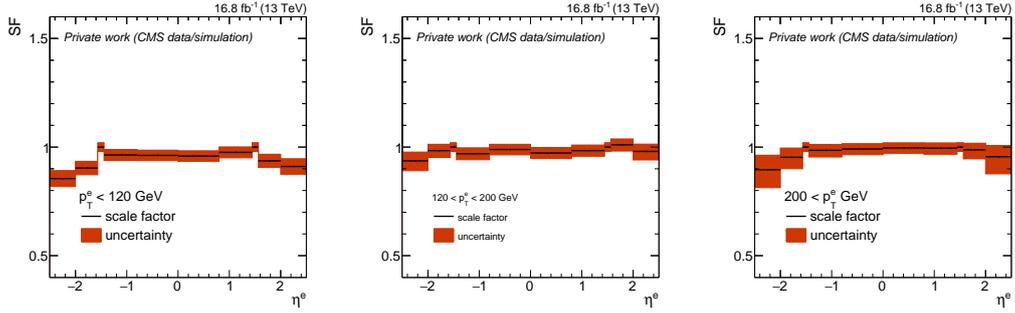


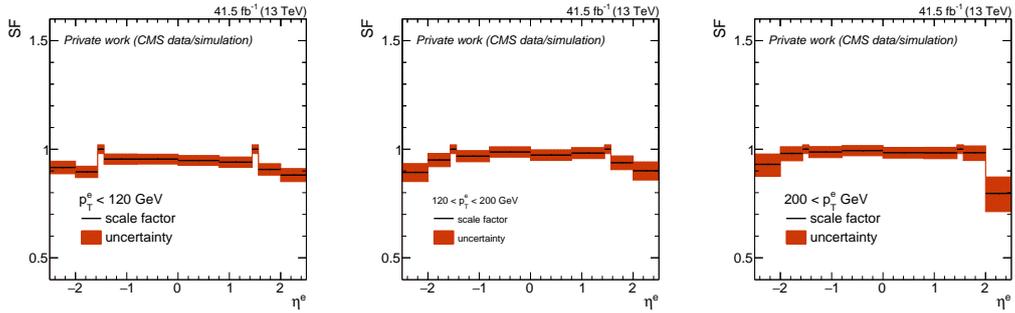
Figure A.2: Trigger efficiencies for the combination of electron triggers documented in section 5.3.5.1, measured in an orthogonal muon dataset. I compare data events to $t\bar{t}$ simulation results in three bins of electron p_T , binned in electron η .



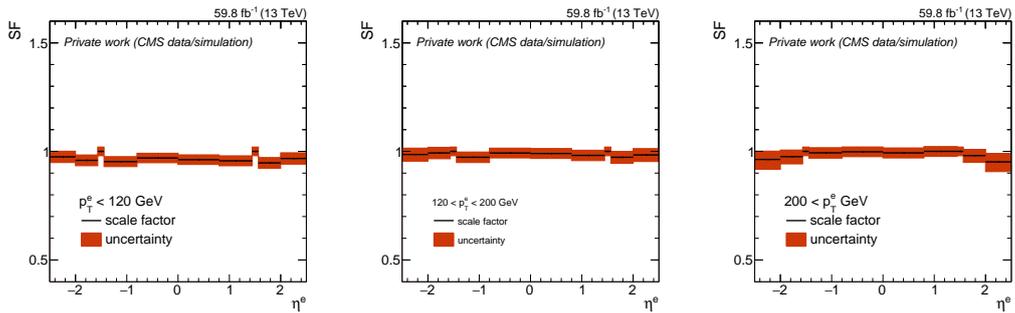
(a) 2016 pre-VFP



(b) 2016 post-VFP



(c) 2017



(d) 2018

Figure A.3: Electron trigger SFs in bins of electron p_T and η , obtained from the ratio of the efficiencies shown in figure A.2.

A.3 ML-based L1 trigger algorithms

In this appendix section, I provide some additional information on the development of ML-based L1 trigger algorithms to ensure reproducibility.

A.3.1 MC simulated samples

In table A.3, I provide information on all simulated samples I use for the various studies on ML-based L1T algorithms. Samples targeting the 2024 CMS configuration are used for the results in sections 6.1 and 6.2. For the NN deployed in the trigger system, a the 2022 background sample and a 2018 signal sample are used, as these were the only ones available at that time. Finally, 2018 samples are used to train the HLT NN.

Dataset	MC generators	Year	N(events)
$gg \rightarrow HH \rightarrow 2b 2W \rightarrow 2b 2q \mu\nu$	POWHEG	2024	167 002
$t\bar{t} \rightarrow 2b 2W \rightarrow 2b 2q \ell\nu$	POWHEG	2024	218 997
Low energy background (1ν)	none	2024	9 682 400
Low energy background (1ν)	none	2022	1 000 000
$gg \rightarrow HH \rightarrow 2b 2W \rightarrow 2b 2q \mu\nu$	POWHEG	2018	130 466
QCD multijet	PYTHIA8	2018	130 466

Table A.3: All MC simulated samples used for the results on NN-based L1 samples I present in this thesis. I give the generated process, alongside the matrix element generator that is used. All samples use PYTHIA8 with the CP5 tune for hadronization and showering, and GEANT4 for detector simulation. I also provide information on the detector conditions of which year the simulation target, and the number of events per sample I use.

I also use a sample of actual detector data: 24 055 000 events of randomly triggered data from 2018 for the Run 2 integration studies. This sample is chosen to be especially large in order to actually have some non-zero number of events passing a HLT selection.

- [1] CMS Collaboration, “Search for pair production of heavy particles decaying to a top quark and a gluon in the lepton+jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **85**, 3 (2025), p. 342, DOI: 10.1140/epjc/s10052-024-13729-y, arXiv: 2410.20601 [hep-ex].
- [2] F. Labe, “Searches for new resonances coupling to third generation quarks at CMS”, presentation at ICHEP 2024, URL: <https://indi.to/VzNjN> (visited on 09/10/2024).
- [3] F. Labe, “Searches for new resonances coupling to third generation quarks at CMS”, CMS-CR-2024-180 (2024), accepted by PoS, URL: <https://cds.cern.ch/record/2911730>.
- [4] CMS Collaboration, “Search for pair production of heavy particles decaying to a top quark and a gluon in the lepton+jets final state at $\sqrt{s} = 13$ TeV”, *HEPData record* (2024), DOI: 10.17182/hepdata.153717.
- [5] F. Labe, “Search for a pair production of excited top quarks at 13 TeV at the CMS detector”, master’s thesis (2020), URL: <https://www.physik.uni-hamburg.de/de/iexp/gruppe-haller/abschlussarbeiten/documents/masterarbeit-finn-labe.pdf>.
- [6] Y. J. Bantchien, “Search for excited top quarks”, bachelors’s thesis (2021).
- [7] “UHH2 framework”, URL: <https://github.com/UHH2/UHH2> (visited on 12/22/2024).
- [8] “columnflow”, URL: <https://github.com/columnflow/columnflow> (visited on 12/22/2024).
- [9] I. Komarov, “Development of neural-network based topological Level-1 triggers for the CMS experiment at the HL-LHC”, master’s thesis (2022), URL: <https://www.physik.uni-hamburg.de/iexp/gruppe-haller/scientific-output/documents/masterarbeit-ihor-komarov.pdf>.
- [10] S. Sepanlou, “Development of neural network based trigger algorithms for the Level 1 Trigger of the CMS experiment”, master’s thesis (2023), URL: <https://www.physik.uni-hamburg.de/en/iexp/gruppe-haller/scientific-output/documents/masterarbeit-shahin-sepanlou.pdf>.
- [11] K. Kleinbölting, “Development of a neural network for the CMS Level-1 trigger”, bachelors’s thesis (2022).
- [12] CMS Collaboration, “Performance of soft drop mass jet High Level Trigger at $\sqrt{s} = 13.6$ TeV in Run 3”, CMS-DP-2023-013 (2023), URL: <https://cds.cern.ch/record/2855342>.
- [13] J. J. Thomson, “XL. Cathode Rays”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44**, 269 (1897), pp. 293–316, DOI: 10.1080/14786449708621070.

-
- [14] E. Rutherford, “LXXIX. The scattering of α particles by matter and the structure of the atom”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **21**, 125 (1911), pp. 669–688, DOI: 10.1080/14786440508637080.
- [15] E. Rutherford, “LIV. Collision of α particles with light atoms. IV. An anomalous effect in nitrogen”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **37**, 222 (1919), pp. 581–587, DOI: 10.1080/14786440608635919.
- [16] J. Chadwick, “The existence of a neutron”, *Proc. R. Soc. Lond. A* (1932), pp. 692–708, DOI: 10.1098/rspa.1932.0112.
- [17] C. D. Anderson and S. H. Neddermeyer, “Cloud chamber observations of cosmic rays at 4300 meters elevation and near sea-level”, *Phys. Rev.* **50** 4 (1936), pp. 263–271, DOI: 10.1103/PhysRev.50.263.
- [18] M. Gell-Mann, “A schematic model of baryons and mesons”, *Physics Letters* **8**, 3 (1964), pp. 214–215, ISSN: 0031-9163, DOI: [https://doi.org/10.1016/S0031-9163\(64\)92001-3](https://doi.org/10.1016/S0031-9163(64)92001-3).
- [19] G. Zweig, “An SU(3) model for strong interaction symmetry and its breaking. Version 1”, CERN-TH-401 (1964), DOI: 10.17181/CERN-TH-401.
- [20] G. Zweig, “An SU(3) model for strong interaction symmetry and its breaking. Version 2”, *Developments in the quark theory of hadrons* **1** (1964), ed. by D. B. Lichtenberg and S. P. Rosen, pp. 22–101, DOI: 10.17181/CERN-TH-412.
- [21] E. D. Bloom et al., “High-Energy inelastic $e - p$ scattering at 6° and 10° ”, *Phys. Rev. Lett.* **23** 16 (1969), pp. 930–934, DOI: 10.1103/PhysRevLett.23.930.
- [22] M. Breidenbach et al., “Observed behavior of highly inelastic electron-proton scattering”, *Phys. Rev. Lett.* **23** 16 (1969), pp. 935–939, DOI: 10.1103/PhysRevLett.23.935.
- [23] E598 Collaboration, “Experimental observation of a heavy particle J ”, *Phys. Rev. Lett.* **33** (1974), pp. 1404–1406, DOI: 10.1103/PhysRevLett.33.1404.
- [24] SLAC Collaboration, “Discovery of a narrow resonance in e^+e^- annihilation”, *Phys. Rev. Lett.* **33** (1974), pp. 1406–1408, DOI: 10.1103/PhysRevLett.33.1406.
- [25] M. L. Perl et al., “Evidence for anomalous lepton production in $e^+ - e^-$ annihilation”, *Phys. Rev. Lett.* **35** (1975), pp. 1489–1492, DOI: 10.1103/PhysRevLett.35.1489.
- [26] E288 Collaboration, “Observation of a dimuon resonance at 9.5 GeV in 400 GeV proton-nucleus collisions”, *Phys. Rev. Lett.* **39** (1977), pp. 252–255, DOI: 10.1103/PhysRevLett.39.252.
- [27] CDF Collaboration, “Observation of top quark production in $\bar{p}p$ collisions”, *Phys. Rev. Lett.* **74** (1995), pp. 2626–2631, DOI: 10.1103/PhysRevLett.74.2626, arXiv: hep-ex/9503002.

- [28] D0 Collaboration, “Observation of the top quark”, *Phys. Rev. Lett.* **74** (1995), pp. 2632–2637, DOI: 10.1103/PhysRevLett.74.2632, arXiv: hep-ex/9503003.
- [29] B. R. Stella and H.-J. Meyer, “Y(9.46 GeV) and the gluon discovery (a critical recollection of PLUTO results)”, *Eur. Phys. J. H* **36** (2011), pp. 203–243, DOI: 10.1140/epjh/e2011-10029-3, arXiv: 1008.1869 [hep-ex].
- [30] UA1 Collaboration, “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV”, *Phys. Lett. B* **122** (1983), pp. 103–116, DOI: 10.1016/0370-2693(83)91177-2.
- [31] UA2 Collaboration, “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN anti-p p collider”, *Phys. Lett. B* **122** (1983), pp. 476–485, DOI: 10.1016/0370-2693(83)91605-2.
- [32] UA1 Collaboration, “Experimental observation of lepton pairs of invariant mass around 95 GeV/c² at the CERN SPS collider”, *Phys. Lett. B* **126** (1983), pp. 398–410, DOI: 10.1016/0370-2693(83)90188-0.
- [33] UA2 Collaboration, “Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN $\bar{p}p$ collider”, *Phys. Lett. B* **129** (1983), pp. 130–140, DOI: 10.1016/0370-2693(83)90744-X.
- [34] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett. B* **716** (2012), pp. 1–29, DOI: 10.1016/j.physletb.2012.08.020, arXiv: 1207.7214 [hep-ex].
- [35] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2012), pp. 30–61, DOI: 10.1016/j.physletb.2012.08.021, arXiv: 1207.7235 [hep-ex].
- [36] CMS Collaboration, “Observation of a New Boson with Mass Near 125 GeV in pp Collisions at $\sqrt{s} = 7$ and 8 TeV”, *JHEP* **06** (2013), p. 081, DOI: 10.1007/JHEP06(2013)081, arXiv: 1303.4571 [hep-ex].
- [37] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008), S08004, DOI: 10.1088/1748-0221/3/08/S08004.
- [38] D. Griffiths, “Introduction to elementary particles; 2nd rev. version”, Wiley, 2008, ISBN: 978-3-527-40601-2.
- [39] M. Thomson, “Modern Particle Physics”, Cambridge University Press, 2013, ISBN: 9781107034266.
- [40] Particle Data Group Collaboration, “Review of Particle Physics”, *Phys. Rev. D* **98** 3 (2018), p. 030001, DOI: 10.1103/PhysRevD.98.030001.
- [41] “File:Standard Model of Elementary Particles.svg”, URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (visited on 03/11/2024).
- [42] E. Noether, “Invariante Variationsprobleme”, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1918), pp. 235–257, URL: <http://eudml.org/doc/59024>.

- [43] Particle Data Group Collaboration, “Review of Particle Physics”, *PTEP* (2022), p. 083C01, DOI: 10.1093/ptep/ptac097.
- [44] J. Ellis, “Higgs Physics”, *2013 European School of High-Energy Physics*, 2015, pp. 117–168, DOI: 10.5170/CERN-2015-004.117, arXiv: 1312.5672 [hep-ph].
- [45] W. J. Stirling, private communication.
- [46] J. Zinn-Justin, “Quantum field theory and critical phenomena”, International Series of Monographs on Physics, Oxford University Press, 2021, ISBN: 978-0-19-850923-3.
- [47] Gfitter Group, “The global electroweak fit at NNLO and prospects for the LHC and ILC”, *Eur. Phys. J. C* **74** (2014), p. 3046, DOI: 10.1140/epjc/s10052-014-3046-5, arXiv: 1407.3792 [hep-ph].
- [48] Gfitter Group, “Update of the global electroweak fit and constraints on two-Higgs-doublet models”, *Eur. Phys. J. C* **78**, 8 (2018), p. 675, DOI: 10.1140/epjc/s10052-018-6131-3, arXiv: 1803.01853 [hep-ph].
- [49] CDF Collaboration, “High-precision measurement of the W boson mass with the CDF II detector”, *Science* **376**, 6589 (2022), pp. 170–176, DOI: 10.1126/science.abk1781.
- [50] Muon $g-2$ Collaboration, “Measurement of the positive muon anomalous magnetic moment to 0.20 ppm”, *Phys. Rev. Lett.* **131**, 16 (2023), p. 161802, DOI: 10.1103/PhysRevLett.131.161802, arXiv: 2308.06230 [hep-ex].
- [51] B. Capdevila et al., “Patterns of new physics in $b \rightarrow s\ell^+\ell^-$ transitions in the light of recent data”, *JHEP* **01** (2018), p. 093, DOI: 10.1007/JHEP01(2018)093, arXiv: 1704.05340 [hep-ph].
- [52] LHCb Collaboration, “Search for lepton-universality violation in $B^+ \rightarrow K^+\ell^+\ell^-$ decays”, *Phys. Rev. Lett.* **122**, 19 (2019), p. 191801, DOI: 10.1103/PhysRevLett.122.191801, arXiv: 1903.09252 [hep-ex].
- [53] LHCb Collaboration, “Test of lepton universality in beauty-quark decays”, *Nature Phys.* **18**, 3 (2022), pp. 277–282 [Addendum: *Nature Phys.* **19**, (2023)], DOI: 10.1038/s41567-023-02095-3, arXiv: 2103.11769 [hep-ex].
- [54] LHCb Collaboration, “Measurement of lepton universality parameters in $B^+ \rightarrow K^+\ell^+\ell^-$ and $B^0 \rightarrow K^{*0}\ell^+\ell^-$ decays”, *Phys. Rev. D* **108**, 3 (2023), p. 032002, DOI: 10.1103/PhysRevD.108.032002, arXiv: 2212.09153 [hep-ex].
- [55] LHCb Collaboration, “Test of lepton universality in $b \rightarrow s\ell^+\ell^-$ decays”, *Phys. Rev. Lett.* **131**, 5 (2023), p. 051803, DOI: 10.1103/PhysRevLett.131.051803, arXiv: 2212.09152 [hep-ex].
- [56] LHCb Collaboration, “Measurement of the ratios of branching fractions $\mathcal{R}(D^*)$ and $\mathcal{R}(D^0)$ ”, *Phys. Rev. Lett.* **131** (2023), p. 111802, DOI: 10.1103/PhysRevLett.131.111802, arXiv: 2302.02886 [hep-ex].

- [57] LHCb Collaboration, “Test of lepton flavor universality using $B_0 \rightarrow D^{*-} \tau^+ \nu_\tau$ decays with hadronic τ channels”, *Phys. Rev. D* **108**, 1 (2023), p. 012018 [Erratum: *Phys.Rev.D* 109, 119902 (2024)], DOI: 10.1103/PhysRevD.108.012018, arXiv: 2305.01463 [hep-ex].
- [58] LHCb Collaboration, “Test of lepton flavour universality with $B_s^0 \rightarrow \phi \ell^+ \ell^-$ decays”, LHCb-PAPER-2024-032 (2024), arXiv: 2410.13748 [hep-ex].
- [59] LHCb Collaboration, “Test of lepton flavour universality with $B^+ \rightarrow K^+ \pi^+ \pi^- \ell^+ \ell^-$ decays”, LHCb-PAPER-2024-046 (2024), arXiv: 2412.11645 [hep-ex].
- [60] R. Davis, D. S. Harmer, and K. C. Hoffman, “Search for neutrinos from the sun”, *Phys. Rev. Lett.* **20** 21 (1968), pp. 1205–1209, DOI: 10.1103/PhysRevLett.20.1205.
- [61] Super-Kamiokande Collaboration, “Evidence for oscillation of atmospheric neutrinos”, *Phys. Rev. Lett.* **81** (1998), pp. 1562–1567, DOI: 10.1103/PhysRevLett.81.1562, arXiv: hep-ex/9807003.
- [62] Super-Kamiokande Collaboration, “The Super-Kamiokande detector”, *Nucl. Instrum. Meth. A* **501** (2003), ed. by V. A. Ilyin, V. V. Korenkov, and D. Perret-Gallix, pp. 418–462, DOI: 10.1016/S0168-9002(03)00425-X.
- [63] KATRIN Collaboration, “The design, construction, and commissioning of the KATRIN experiment”, *JINST* **16**, 08 (2021), T08015, DOI: 10.1088/1748-0221/16/08/T08015, arXiv: 2103.04755 [physics.ins-det].
- [64] KATRIN Collaboration, “Direct neutrino-mass measurement with sub-electronvolt sensitivity”, *Nature Phys.* **18**, 2 (2022), pp. 160–166, DOI: 10.1038/s41567-021-01463-1, arXiv: 2105.08533 [hep-ex].
- [65] F. Zwicky, “Die Rotverschiebung von extragalaktischen Nebeln”, *Helv. Phys. Acta* **6** (1933), pp. 110–127, DOI: 10.1007/s10714-008-0707-4.
- [66] J. de Swart, G. Bertone, and J. van Dongen, “How dark matter came to matter”, *Nature Astron.* **1** (2017), p. 0059, DOI: 10.1038/s41550017-0059, arXiv: 1703.00013 [astro-ph.CO].
- [67] C. S. Wu et al., “Experimental test of parity conservation in β decay”, *Phys. Rev.* **105** (1957), pp. 1413–1414, DOI: 10.1103/PhysRev.105.1413.
- [68] J. H. Christenson et al., “Evidence for the 2π decay of the K_2^0 meson”, *Phys. Rev. Lett.* **13** (1964), pp. 138–140, DOI: 10.1103/PhysRevLett.13.138.
- [69] NA48 Collaboration, “A New measurement of direct CP violation in two pion decays of the neutral kaon”, *Phys. Lett. B* **465** (1999), pp. 335–348, DOI: 10.1016/S0370-2693(99)01030-8, arXiv: hep-ex/9909022.
- [70] BaBar Collaboration, “Measurement of CP violating asymmetries in B^0 decays to CP eigenstates”, *Phys. Rev. Lett.* **86** (2001), pp. 2515–2522, DOI: 10.1103/PhysRevLett.86.2515, arXiv: hep-ex/0102030.
- [71] Belle Collaboration, “Observation of large CP violation in the neutral B meson system”, *Phys. Rev. Lett.* **87** (2001), p. 091802, DOI: 10.1103/PhysRevLett.87.091802, arXiv: hep-ex/0107061.

- [72] LHCb Collaboration, “First observation of CP violation in the decays of B_s^0 mesons”, *Phys. Rev. Lett.* **110**, 22 (2013), p. 221601, DOI: 10.1103/PhysRevLett.110.221601, arXiv: 1304.6173 [hep-ex].
- [73] LHCb Collaboration, “Observation of CP Violation in charm decays”, *Phys. Rev. Lett.* **122**, 21 (2019), p. 211803, DOI: 10.1103/PhysRevLett.122.211803, arXiv: 1903.08726 [hep-ex].
- [74] D. Wu, “A brief introduction to the strong CP problem”, 1991, URL: <https://www.osti.gov/biblio/6260191>.
- [75] T. Mannel, “Theory and phenomenology of CP violation”, *Nucl. Phys. B Proc. Suppl.* **167** (2007), ed. by G. Borisov et al., pp. 115–119, DOI: 10.1016/j.nuclphysbps.2006.12.083.
- [76] S. Dar, “The Neutron EDM in the SM: A review” (2000), arXiv: hep-ph/0008248.
- [77] P. Cea and G. Nardulli, “A realistic calculation of the electric dipole moment of the neutron induced by strong CP violation”, *Physics Letters B* **144**, 1 (1984), pp. 115–118, ISSN: 0370-2693, DOI: [https://doi.org/10.1016/0370-2693\(84\)90187-4](https://doi.org/10.1016/0370-2693(84)90187-4).
- [78] M. Chaichian, R. Gonzalez Felipe, and K. Huitu, “On quadratic divergences and the Higgs mass”, *Phys. Lett. B* **363** (1995), pp. 101–105, DOI: 10.1016/0370-2693(95)01191-R, arXiv: hep-ph/9509223.
- [79] S. El Hedri, A. E. Nelson, and D. G. E. Walker, “Reducing the quadratic divergence in the Higgs boson mass squared without top partners”, *Phys. Rev. D* **98**, 3 (2018), p. 035029, DOI: 10.1103/PhysRevD.98.035029, arXiv: 1805.04523 [hep-ph].
- [80] J. Berger, J. Hubisz, and M. Perelstein, “A fermionic top partner: Naturalness and the LHC”, *JHEP* **07** (2012), p. 016, DOI: 10.1007/JHEP07(2012)016, arXiv: 1205.0013 [hep-ph].
- [81] J. Wess and B. Zumino, “Supergauge transformations in four-dimensions”, *Nucl. Phys. B* **70** (1974), ed. by A. Salam and E. Sezgin, pp. 39–50, DOI: 10.1016/0550-3213(74)90355-1.
- [82] H. P. Nilles, “Supersymmetry, supergravity and particle physics”, *Phys. Rept.* **110** (1984), pp. 1–162, DOI: 10.1016/0370-1573(84)90008-5.
- [83] H. E. Haber and G. L. Kane, “The search for supersymmetry: Probing physics beyond the Standard Model”, *Phys. Rept.* **117** (1985), pp. 75–263, DOI: 10.1016/0370-1573(85)90051-1.
- [84] R. Barbieri, S. Ferrara, and C. A. Savoy, “Gauge models with spontaneously broken local supersymmetry”, *Phys. Lett. B* **119** (1982), p. 343, DOI: 10.1016/0370-2693(82)90685-2.
- [85] S. Dimopoulos and H. Georgi, “Softly Broken Supersymmetry and $SU(5)$ ”, *Nucl. Phys. B* **193**, HUTP-81/A022 (1981), pp. 150–162, DOI: 10.1016/0550-3213(81)90522-8.
- [86] P. Langacker, “Grand unification”, *Scholarpedia* **7**, 10 (2012), p. 11419 revision #137135, DOI: 10.4249/scholarpedia.11419.

- [87] H. Georgi and S. L. Glashow, “Unity of all elementary particle forces”, *Phys. Rev. Lett.* **32** (1974), pp. 438–441, DOI: 10.1103/PhysRevLett.32.438.
- [88] LHCb Collaboration, “Test of lepton universality in beauty-quark decays”, *Nature Phys.* **18**, 3 (2022), pp. 277–282 [Addendum: *Nature Phys.* 19, (2023)], DOI: 10.1038/s41567-023-02095-3, arXiv: 2103.11769 [hep-ex].
- [89] “Compositeness”, URL: <https://home.cern/science/physics/compositeness> (visited on 03/21/2024).
- [90] I. A. D’Souza and C. S. Kalman, “Preons: Models of leptons, quarks and gauge bosons as composite objects”, World Scientific, 1992.
- [91] H. Georgi et al., “Effects of top compositeness”, *Phys. Rev. D* **51** (1995), p. 3888, DOI: 10.1103/PhysRevD.51.3888, arXiv: hep-ph/9410307.
- [92] B. Lillie, J. Shu, and T. M. P. Tait, “Top compositeness at the Tevatron and LHC”, *JHEP* **04** (2008), p. 087, DOI: 10.1088/1126-6708/2008/04/087, arXiv: 0712.3057 [hep-ph].
- [93] A. Pomarol and J. Serra, “Top quark compositeness: Feasibility and implications”, *Phys. Rev. D* **78** (2008), p. 074026, DOI: 10.1103/PhysRevD.78.074026, arXiv: 0806.3247 [hep-ph].
- [94] K. Kumar, T. M. P. Tait, and R. Vega-Morales, “Manifestations of top compositeness at colliders”, *JHEP* **05** (2009), p. 022, DOI: 10.1088/1126-6708/2009/05/022, arXiv: 0901.3808 [hep-ph].
- [95] D. Buarque Franzosi and A. Tonerio, “Top-quark partial compositeness beyond the effective field theory paradigm”, *JHEP* **04** (2020), p. 040, DOI: 10.1007/JHEP04(2020)040, arXiv: 1908.06996 [hep-ph].
- [96] G. Panico, “Top compositeness, flavor and naturalness”, *Proc. 10th Int. Workshop on Top Quark Physics* (2018), arXiv: 1801.03882 [hep-ph].
- [97] T. A. DeGrand et al., “Towards partial compositeness on the lattice: baryons with fermions in multiple representations”, *PoS LATTICE2016* (2016), p. 219, DOI: 10.22323/1.256.0219, arXiv: 1610.06465 [hep-lat].
- [98] A. Pierce and Y. Zhao, “Naturalness from a composite top?”, *JHEP* **01** (2017), p. 054, DOI: 10.1007/JHEP01(2017)054, arXiv: 1607.01318 [hep-ph].
- [99] O. Witzel, “Review on composite Higgs models”, *PoS LATTICE2018* (2019), p. 006, DOI: 10.22323/1.334.0006, arXiv: 1901.08216 [hep-lat].
- [100] N. Arkani-Hamed et al., “The lightest Higgs”, *JHEP* **07** (2002), p. 034, DOI: 10.1088/1126-6708/2002/07/034, arXiv: hep-ph/0206021.
- [101] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, “Electroweak symmetry breaking from dimensional deconstruction”, *Phys. Lett. B* **513** (2001), p. 232, DOI: 10.1016/S0370-2693(01)00741-9, arXiv: hep-ph/0105239.
- [102] M. Perelstein, “Little Higgs models and their phenomenology”, *Prog. Part. Nucl. Phys.* **58** (2007), p. 247, DOI: 10.1016/j.pnpnp.2006.04.001, arXiv: hep-ph/0512128.

- [103] A. Ahmed, M. Lindner, and P. Saake, “Conformal little Higgs models”, *Phys. Rev. D* **109** (2024), p. 075041, DOI: 10.1103/PhysRevD.109.075041, arXiv: 2309.07845 [hep-ph].
- [104] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, “The hierarchy problem and new dimensions at a millimeter”, *Phys. Lett. B* **429** (1998), pp. 263–272, DOI: 10.1016/S0370-2693(98)00466-3, arXiv: hep-ph/9803315.
- [105] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, “Phenomenology, astrophysics and cosmology of theories with submillimeter dimensions and TeV scale quantum gravity”, *Phys. Rev. D* **59** (1999), p. 086004, DOI: 10.1103/PhysRevD.59.086004, arXiv: hep-ph/9807344.
- [106] L. Randall and R. Sundrum, “A large mass hierarchy from a small extra dimension”, *Phys. Rev. Lett.* **83** (1999), pp. 3370–3373, DOI: 10.1103/PhysRevLett.83.3370, arXiv: hep-ph/9905221.
- [107] T. Gherghetta and A. Pomarol, “Bulk fields and supersymmetry in a slice of AdS”, *Nucl. Phys. B* **586** (2000), p. 141, DOI: 10.1016/S0550-3213(00)00392-8, arXiv: hep-ph/0003129.
- [108] B. Hassanain, J. March-Russell, and J. G. Rosa, “On the possibility of light string resonances at the LHC and Tevatron from Randall-Sundrum throats”, *JHEP* **07** (2009), p. 077, DOI: 10.1088/1126-6708/2009/07/077, arXiv: 0904.4108 [hep-ph].
- [109] P. Candelas et al., “Vacuum configurations for superstrings”, *Nucl. Phys. B* **258** (1985), pp. 46–74, DOI: 10.1016/0550-3213(85)90602-9.
- [110] S. B. Giddings, S. Kachru, and J. Polchinski, “Hierarchies from fluxes in string compactifications”, *Phys. Rev. D* **66** (2002), p. 106006, DOI: 10.1103/PhysRevD.66.106006, arXiv: hep-th/0105097.
- [111] L. Evans and P. Bryant, “LHC Machine”, *Journal of Instrumentation* **3**, 08 (2008), S08001, DOI: 10.1088/1748-0221/3/08/S08001.
- [112] “LHC report: make way for the heavy ions”, 2018, URL: <https://home.cern/news/news/accelerators/lhc-report-make-way-heavy-ions> (visited on 02/21/2024).
- [113] “Linear accelerator 2”, URL: <https://home.cern/science/accelerators/linear-accelerator-2> (visited on 02/21/2024).
- [114] “The Proton Synchrotron Booster”, URL: <https://home.cern/science/accelerators/proton-synchrotron-booster> (visited on 02/21/2024).
- [115] “The Proton Synchrotron”, URL: <https://home.cern/science/accelerators/proton-synchrotron> (visited on 02/21/2024).
- [116] “The Super Proton Synchrotron”, URL: <https://home.cern/science/accelerators/super-proton-synchrotron> (visited on 02/21/2024).
- [117] E. Mobs, “The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018”, 2018, URL: <https://cds.cern.ch/record/2636343>.
- [118] B. Richter, “Very High-Energy electron-Positron Colliding Beams for the Study of the Weak Interactions”, *Nucl. Instrum. Meth.* **136** (1976), pp. 47–60, DOI: 10.1016/0029-554X(76)90396-7.

- [119] D. Boussard et al., “The LHC superconducting cavities”, *IEEE Particle Accelerator Conference (PAC 99)* (1999), pp. 946–948, DOI: 10.1109/PAC.1999.795409.
- [120] “Accelerating: radiofrequency cavities”, URL: <https://home.cern/science/engineering/accelerating-radiofrequency-cavities> (visited on 02/22/2024).
- [121] “LHC report: full house for the LHC”, URL: <https://home.cern/news/news/accelerators/lhc-report-full-house-lhc> (visited on 02/22/2024).
- [122] J. Wenninger, “Machine protection and operation for LHC”, *2014 Joint International Accelerator School: Beam Loss and Accelerator Protection* (2016), pp. 377–401, DOI: 10.5170/CERN-2016-002.377, arXiv: 1608.03113 [physics.acc-ph].
- [123] ATLAS Collaboration, “The ATLAS experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008), S08003, DOI: 10.1088/1748-0221/3/08/S08003.
- [124] LHCb Collaboration, “The LHCb detector at the LHC”, *JINST* **3** (2008), S08005, DOI: 10.1088/1748-0221/3/08/S08005.
- [125] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008), S08002, DOI: 10.1088/1748-0221/3/08/S08002.
- [126] Y. Ohnishi et al., “Accelerator design at SuperKEKB”, *PTEP* **2013** (2013), 03A011, DOI: 10.1093/ptep/pts083.
- [127] R. P. Feynman, “The behavior of hadron collisions at extreme energies”, *Conf. Proc. C* **690905** (1969), pp. 237–258.
- [128] J. D. Bjorken, “Asymptotic sum rules at infinite momentum”, *Phys. Rev.* **179** 1547 (1969), DOI: 10.1103/PhysRev.179.1547.
- [129] NNPDF Collaboration, “Parton distributions from high-precision collider data”, *Eur. Phys. J. C* **77**, 10 (2017), p. 663, DOI: 10.1140/epjc/s10052-017-5199-5, arXiv: 1706.00428 [hep-ph].
- [130] T. Sakuma and T. McCauley, “Detector and event visualization with SketchUp at the CMS experiment”, *J. Phys. Conf. Ser.* **513** (2014), ed. by D. L. Groep and D. Bonacorsi, p. 022032, DOI: 10.1088/1742-6596/513/2/022032, arXiv: 1311.4942 [physics.ins-det].
- [131] CMS Collaboration, “The CMS tracker system project: Technical Design Report”, CMS-TDR-5 (1997), URL: <https://cds.cern.ch/record/368412>.
- [132] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9**, 10 (2014), P10009, DOI: 10.1088/1748-0221/9/10/P10009, arXiv: 1405.6569 [physics.ins-det].
- [133] CMS Tracker Group, “The CMS phase-1 pixel detector upgrade”, *JINST* **16**, 02 (2021), P02027, DOI: 10.1088/1748-0221/16/02/P02027, arXiv: 2012.14304 [physics.ins-det].
- [134] CMS Collaboration, “Operation and performance of the CMS outer tracker”, *PoS Vertex* **2017** (2018), p. 013, DOI: 10.22323/1.309.0013.

- [135] CMS Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report”, CMS-TDR-4 (1997), URL: <http://cds.cern.ch/record/349375>.
- [136] CMS Collaboration, “The CMS ECAL performance with examples”, *JINST* **9** (2014), ed. by P. Govoni et al., p. C02008, DOI: 10.1088/1748-0221/9/02/C02008.
- [137] P. Adzic et al., “Energy resolution of the barrel of the CMS electromagnetic calorimeter”, *JINST* **2** (2007), P04004, DOI: 10.1088/1748-0221/2/04/P04004.
- [138] CMS Collaboration, “Performance of the CMS electromagnetic calorimeter in LHC Run 2”, CMS-CR-2020-038 (2020), URL: <https://cds.cern.ch/record/2797453>.
- [139] CMS Collaboration, “The CMS hadron calorimeter project: Technical Design Report”, CMS-TDR-2 (1997), URL: <https://cds.cern.ch/record/357153>.
- [140] CMS Collaboration, “Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV”, *JINST* **15**, 05 (2020), P05002, DOI: 10.1088/1748-0221/15/05/P05002, arXiv: 1910.00079 [physics.ins-det].
- [141] CMS Collaboration, “The CMS magnet project: Technical Design Report”, CMS-TDR-1 (1997), URL: <https://cds.cern.ch/record/331056>.
- [142] CMS Collaboration, “Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays”, *JINST* **5** (2010), T03021, DOI: 10.1088/1748-0221/5/03/T03021, arXiv: 0910.5530 [physics.ins-det].
- [143] CMS Collaboration, “The CMS muon project: Technical Design Report”, CMS-TDR-3 (1997), URL: <https://cds.cern.ch/record/343814>.
- [144] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13**, 06 (2018), P06015, DOI: 10.1088/1748-0221/13/06/P06015, arXiv: 1804.04528 [physics.ins-det].
- [145] CMS Collaboration, “The CMS RPC detector performance and stability during LHC Run-2”, *JINST* **14**, 11 (2019), p. C11012, DOI: 10.1088/1748-0221/14/11/C11012, arXiv: 1808.10488 [physics.ins-det].
- [146] CMS Collaboration, “Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS”, *Eur. Phys. J. C* **81**, 9 (2021), p. 800, DOI: 10.1140/epjc/s10052-021-09538-2, arXiv: 2104.01927 [hep-ex].
- [147] “The HL-LHC project”, URL: <https://hilumilhc.web.cern.ch/content/hl-lhc-project> (visited on 12/16/2024).
- [148] “Public CMS luminosity information”, URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults> (visited on 02/28/2024).
- [149] “How does CMS measure luminosity?”, URL: <https://cms.cern/news/how-does-cms-measure-luminosity> (visited on 04/23/2024).

- [150] CMS Collaboration, “Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data”, CMS-DP-2020-045 (2020), URL: <http://cds.cern.ch/record/2740688>.
- [151] “Public CMS data quality information”, URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DataQuality> (visited on 04/23/2024).
- [152] I. Zurbano Fernandez et al., “High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report”, **10/2020** (2020), ed. by I. Béjar Alonso et al., DOI: 10.23731/CYRM-2020-0010.
- [153] CMS Collaboration, “Pileup mitigation at CMS in 13 TeV data”, *JINST* **15**, 09 (2020), P09018, DOI: 10.1088/1748-0221/15/09/P09018, arXiv: 2003.00503 [hep-ex].
- [154] CMS Collaboration, “CMS: The TriDAS project. Technical Design Report, vol. 2: Data acquisition and high-level trigger”, CMS-TDR-6-2 (2002), ed. by P. Sphicas, URL: <https://cds.cern.ch/record/578006>.
- [155] T. Bawej et al., “The new CMS DAQ system for Run-2 of the LHC”, *IEEE Trans.Nucl.Sci.* **62** 3 (2015), pp. 1099–1103, DOI: 10.1109/RTC.2014.7097437.
- [156] CMS Collaboration, “CMS: The TriDAS project. Technical Design Report, vol. 1: The trigger systems”, CMS-TDR-6-1 (2000), URL: <https://cds.cern.ch/record/706847>.
- [157] CMS Collaboration, “CMS Technical Design Report for the Level-1 trigger upgrade”, CMS-TDR-12 (2013), ed. by A. Tapper and D. Acosta, URL: <https://cds.cern.ch/record/2714892>.
- [158] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **15**, 10 (2020), P10017, DOI: 10.1088/1748-0221/15/10/P10017, arXiv: 2006.10165 [hep-ex].
- [159] CMS Trigger, Data Acquisition Group, “The CMS high level trigger”, *Eur. Phys. J. C* **46** (2006), pp. 605–667, DOI: 10.1140/epjc/s2006-02495-8, arXiv: hep-ex/0512077.
- [160] CMS Collaboration, “Early 2018 High-Level Trigger rates”, CMS-DP-2018-057 (2018), URL: <https://cds.cern.ch/record/2644379>.
- [161] CMS Collaboration, “Presentation layer of CMS Online Monitoring System”, *EPJ Web Conf.* **214** (2019), ed. by A. Forti et al., p. 01044, DOI: 10.1051/epjconf/201921401044.
- [162] M. Borisyak et al., “Towards automation of data quality system for CERN CMS experiment”, *J. Phys. Conf. Ser.* **898**, 9 (2017), ed. by R. Mount and C. Tull, p. 092041, DOI: 10.1088/1742-6596/898/9/092041, arXiv: 1709.08607 [physics.data-an].
- [163] “Run 2 High Level Trigger performance results”, URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/HighLevelTriggerRunIIResults> (visited on 11/20/2024).
- [164] A. J. Larkoski et al., “Soft Drop”, *JHEP* **05** (2014), p. 146, DOI: 10.1007/JHEP05(2014)146, arXiv: 1402.2657 [hep-ph].

- [165] D. Barney, “CMS detector slice”, CMS-PHO-GEN-2016-001 (2016), URL: <https://cds.cern.ch/record/2120661>.
- [166] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12**, 10 (2017), P10003, DOI: 10.1088/1748-0221/12/10/P10003, arXiv: 1706.04965 [physics.ins-det].
- [167] W. Adam et al., “Track reconstruction in the CMS tracker”, CMS-NOTE-2006-041 (2005), URL: <https://cds.cern.ch/record/934067>.
- [168] W. Adam et al., “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear and Particle Physics* **31**, 9 (2005), N9, DOI: 10.1088/0954-3899/31/9/N01.
- [169] M. Cacciari, G. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04** (2008), p. 063, DOI: 10.1088/1126-6708/2008/04/063, arXiv: 0802.1189 [hep-ph].
- [170] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross-sections in deep inelastic scattering”, *Workshop on Monte Carlo Generators for HERA Physics* (1998), pp. 270–279, arXiv: hep-ph/9907280.
- [171] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev. D* **48** (1993), pp. 3160–3166, DOI: 10.1103/PhysRevD.48.3160, arXiv: hep-ph/9305266.
- [172] T. Lapsien, R. Kogler, and J. Haller, “A new tagger for hadronically decaying heavy particles at the LHC”, *Eur. Phys. J. C* **76**, 11 (2016), p. 600, DOI: 10.1140/epjc/s10052-016-4443-8, arXiv: 1606.04961 [hep-ph].
- [173] M. Stoll, “Vetoed jet clustering: The mass-jump algorithm”, *JHEP* **04** (2015), p. 111, DOI: 10.1007/JHEP04(2015)111, arXiv: 1410.4637 [hep-ph].
- [174] E. Bols et al., “Jet flavour classification using DeepJet”, *JINST* **15**, 12 (2020), P12012, DOI: 10.1088/1748-0221/15/12/P12012, arXiv: 2008.10519 [hep-ex].
- [175] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13**, 05 (2018), P05011, DOI: 10.1088/1748-0221/13/05/P05011, arXiv: 1712.07158 [physics.ins-det].
- [176] J. Thaler and K. Van Tilburg, “Identifying boosted objects with N-subjettiness”, *JHEP* **03** (2011), p. 015, DOI: 10.1007/JHEP03(2011)015, arXiv: 1011.2268 [hep-ph].
- [177] D. Bertolini et al., “Pileup per particle identification”, *JHEP* **10** (2014), p. 059, DOI: 10.1007/JHEP10(2014)059, arXiv: 1407.6013 [hep-ph].
- [178] N. Metropolis, “The beginning of the Monte Carlo method”, URL: <https://api.semanticscholar.org/CorpusID:18607470>.
- [179] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014), p. 079, DOI: 10.1007/JHEP07(2014)079, arXiv: 1405.0301 [hep-ph].

- [180] R. Frederix et al., “The automation of next-to-leading order electroweak calculations”, *JHEP* **07** (2018), p. 185 [Erratum: *JHEP* **11**, 085 (2021)], DOI: 10.1007/JHEP11(2021)085, arXiv: 1804.10017 [hep-ph].
- [181] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004), p. 040, DOI: 10.1088/1126-6708/2004/11/040, arXiv: hep-ph/0409146.
- [182] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, *JHEP* **11** (2007), p. 070, DOI: 10.1088/1126-6708/2007/11/070, arXiv: 0709.2092 [hep-ph].
- [183] S. Alioli et al., “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010), p. 043, DOI: 10.1007/JHEP06(2010)043, arXiv: 1002.2581 [hep-ph].
- [184] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015), pp. 159–177, DOI: 10.1016/j.cpc.2015.01.024, arXiv: 1410.3012 [hep-ph].
- [185] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76**, 3 (2016), p. 155, DOI: 10.1140/epjc/s10052-016-3988-x, arXiv: 1512.00815 [hep-ex].
- [186] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003), pp. 250–303, DOI: 10.1016/S0168-9002(03)01368-8.
- [187] K. Albertsson et al., “Machine Learning in high energy physics community white paper”, *J. Phys. Conf. Ser.* **1085**, 2 (2018), p. 022008, DOI: 10.1088/1742-6596/1085/2/022008, arXiv: 1807.02876 [physics.comp-ph].
- [188] G. James et al., “An Introduction to Statistical Learning”, Springer, 2023, ISBN: 978-3-031-38747-0, DOI: 10.1007/978-3-031-38747-0.
- [189] F. Chollet et al., “Keras”, 2015, URL: <https://keras.io>.
- [190] Martín Abadi et al., “TensorFlow: Large-scale Machine Learning on heterogeneous systems”, 2015, URL: <https://www.tensorflow.org/>.
- [191] M. Mohri, A. Rostamizadeh, and A. Talwalkar, “Foundations of Machine Learning”, MIT Press, 2012, ISBN: 9780262018258.
- [192] NICHD, “What are the parts of the nervous system?”, 2018, URL: <https://www.nichd.nih.gov/health/topics/neuro/conditioninfo/parts> (visited on 01/30/2024).
- [193] D. Godoy, “Understanding binary cross-entropy / log loss: a visual explanation”, *Towards Data Science* (2018), URL: <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a> (visited on 01/31/2024).
- [194] J. Duchi, E. Hazan, and Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization”, *Journal of Machine Learning Research* **12**, 61 (2011), pp. 2121–2159, URL: <http://jmlr.org/papers/v12/duchi11a.html>.

- [195] M. D. Zeiler, “ADADELTA: An Adaptive Learning Rate Method” (2012), arXiv: 1212.5701, URL: <http://arxiv.org/abs/1212.5701>.
- [196] V. Bushaev, “Understanding RMSprop — faster neural network learning”, *Towards Data Science* (2018), URL: <https://towardsdatascience.com/understanding-rmsprop-faster-neural-network-learning-62e116fcf29a> (visited on 02/20/2024).
- [197] D. Kingma and J. Ba, “Adam: A method for stochastic optimization” (2017), arXiv: 1412.6980 [cs.LG].
- [198] F. Scarselli et al., “The Graph Neural Network model”, *IEEE Transactions on Neural Networks* **20**, 1 (2009), pp. 61–80, DOI: 10.1109/TNN.2008.2005605.
- [199] J. Gilmer et al., “Neural message passing for quantum chemistry” (2017), arXiv: 1704.01212 [cs.LG].
- [200] “Recurrent Neural Networks (RNNs)”, URL: <https://towardsdatascience.com/recurrent-neural-networks-rnns-3f06d7653a85> (visited on 12/22/2024).
- [201] D. P. Kingma and M. Welling, “Auto-encoding variational bayes” (2022), arXiv: 1312.6114 [stat.ML].
- [202] G. E. Hinton et al., “Improving neural networks by preventing co-adaptation of feature detectors” (2012), arXiv: 1207.0580 [cs.NE].
- [203] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift” (2015), arXiv: 1502.03167 [cs.LG].
- [204] H. Alhazmi et al., “Shedding light on top partner at the LHC”, *JHEP* **01** (2019), p. 139, DOI: 10.1007/JHEP01(2019)139, arXiv: 1808.03649 [hep-ph].
- [205] J. H. Kim and I. M. Lewis, “Loop induced single top partner production and decay at the LHC”, *JHEP* **05** (2018), p. 095, DOI: 10.1007/JHEP05(2018)095, arXiv: 1803.06351 [hep-ph].
- [206] CMS Collaboration, “Search for electroweak production of a vector-like T quark using fully hadronic final states”, *JHEP* **01** (2020), p. 036, DOI: 10.1007/JHEP01(2020)036, arXiv: 1909.04721 [hep-ex].
- [207] CMS Collaboration, “Search for pair production of vector-like quarks in leptonic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2023), p. 020, DOI: 10.1007/JHEP07(2023)020, arXiv: 2209.07327 [hep-ex].
- [208] CMS Collaboration, “Search for pair production of vector-like quarks in the $bW\bar{b}W$ channel from proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett. B* **779** (2018), pp. 82–106, DOI: 10.1016/j.physletb.2018.01.077, arXiv: 1710.01539 [hep-ex].
- [209] CMS Collaboration, “Search for single production of a vector-like T quark decaying to a top quark and a Z boson in the final state with jets and missing transverse momentum at $\sqrt{s} = 13$ TeV”, *JHEP* **05** (2022), p. 093, DOI: 10.1007/JHEP05(2022)093, arXiv: 2201.02227 [hep-ex].

- [210] ATLAS Collaboration, “Search for singly produced vectorlike top partners in multilepton final states with 139 fb^{-1} of pp collision data at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector”, *Phys. Rev. D* **109**, 11 (2024), p. 112012, DOI: 10.1103/PhysRevD.109.112012, arXiv: 2307.07584 [hep-ex].
- [211] ATLAS Collaboration, “Search for pair-produced vector-like top and bottom partners in events with large missing transverse momentum in pp collisions with the ATLAS detector”, *Eur. Phys. J. C* **83**, 8 (2023), p. 719, DOI: 10.1140/epjc/s10052-023-11790-7, arXiv: 2212.05263 [hep-ex].
- [212] ATLAS Collaboration, “Search for pair-production of vector-like quarks in pp collision events at $\sqrt{s} = 13\text{ TeV}$ with at least one leptonically decaying Z boson and a third-generation quark with the ATLAS detector”, *Phys. Lett. B* **843** (2023), p. 138019, DOI: 10.1016/j.physletb.2023.138019, arXiv: 2210.15413 [hep-ex].
- [213] W. Rarita and J. Schwinger, “On a theory of particles with half integral spin”, *Phys. Rev.* **60** (1941), p. 61, DOI: 10.1103/PhysRev.60.61.
- [214] CMS Collaboration, “Search for pair production of excited top quarks in the lepton + jets final state”, *Phys. Lett. B* **778** (2018), pp. 349–370, DOI: 10.1016/j.physletb.2018.01.049, arXiv: 1711.10949 [hep-ex].
- [215] CMS Collaboration, “Search for pair production of excited top quarks in the lepton + jets final state”, *JHEP* **06** (2014), p. 125, DOI: 10.1007/JHEP06(2014)125, arXiv: 1311.5357 [hep-ex].
- [216] M. Czakon and A. Mitov, “Top++: A Program for the calculation of the top-pair cross-section at hadron colliders”, *Comput. Phys. Commun.* **185** (2014), p. 2930, DOI: 10.1016/j.cpc.2014.06.021, arXiv: 1112.5675 [hep-ph].
- [217] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, *JINST* **16**, 05 (2021), P05014, DOI: 10.1088/1748-0221/16/05/P05014, arXiv: 2012.06888 [hep-ex].
- [218] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7\text{ TeV}$ ”, *JINST* **7** (2012), P10002, DOI: 10.1088/1748-0221/7/10/P10002, arXiv: 1206.4071 [physics.ins-det].
- [219] CMS Collaboration, “Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$ ”, *JINST* **15** (2020), P02027, DOI: 10.1088/1748-0221/15/02/P02027, arXiv: 1912.03516 [physics.ins-det].
- [220] CMS Collaboration, “Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7\text{ TeV}$ ”, *Phys. Lett. B* **713** (2012), pp. 68–90, DOI: 10.1016/j.physletb.2012.05.028, arXiv: 1202.4083 [hep-ex].
- [221] CMS Collaboration, “Jet energy scale and resolution measurements using prompt Run 3 data collected by CMS in the last months of 2022 at 13.6 TeV”, CMS-DP-2023-045 (2023), URL: <https://cds.cern.ch/record/2865846>.
- [222] CMS Collaboration, “Jet energy scale and resolution measurement with Run 2 legacy data collected by CMS at 13 TeV”, CMS-DP-2021-033 (2021), URL: <http://cds.cern.ch/record/2792322>.

- [223] CMS Collaboration, “Search for a heavy resonance decaying into a top quark and a W boson in the lepton+jets final state at $\sqrt{s} = 13$ TeV”, *JHEP* **04** (2022), p. 048, DOI: 10.1007/JHEP04(2022)048, arXiv: 2111.10216 [hep-ex].
- [224] A. Fröhlich, “Search for heavy resonances decaying to a top quark and a W boson with the CMS experiment”, PhD thesis, Universität Hamburg, 2021, URL: <https://www.physik.uni-hamburg.de/iexp/gruppe-haller/scientific-output/documents/thesis-alexander-froehlich.pdf>.
- [225] CMS Collaboration, “Performance summary of AK4 jet b tagging with data from proton-proton collisions at 13 TeV with the CMS detector”, CMS-DP-2023-005 (2023), URL: <https://cds.cern.ch/record/2854609>.
- [226] CMS Collaboration, “Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **16** (2021), P07001, DOI: 10.1088/1748-0221/16/07/P07001, arXiv: 2102.04790 [hep-ex].
- [227] CMS Collaboration, “Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2018), p. 161, DOI: 10.1007/JHEP07(2018)161, arXiv: 1802.02613 [hep-ex].
- [228] M. Rakic et al., “Updated analysis of beam halo measurements in LHC Run 2 and Run 3”, *JACoW IPAC2024* (2024), THPC67, DOI: 10.18429/JACoW-IPAC2024-THPC67.
- [229] CMS Collaboration, “Identification and filtering of uncharacteristic noise in the CMS hadron calorimeter”, *JINST* **5** (2010), T03014, DOI: 10.1088/1748-0221/5/03/T03014, arXiv: 0911.4881 [physics.ins-det].
- [230] “Measurement of the differential cross section for $t\bar{t}$ production in the dilepton final state at $\sqrt{s} = 13$ TeV”, TOP-16-011 (2016).
- [231] CMS Collaboration, “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. D* **95**, 9 (2017), p. 092001, DOI: 10.1103/PhysRevD.95.092001, arXiv: 1610.04191 [hep-ex].
- [232] J. M. Lindert et al., “Precise predictions for $V +$ jets dark matter backgrounds”, *Eur. Phys. J. C* **77**, 12 (2017), p. 829, DOI: 10.1140/epjc/s10052-017-5389-1, arXiv: 1705.04664 [hep-ph].
- [233] CMS Collaboration, “Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques”, *JINST* **15**, 06 (2020), P06005, DOI: 10.1088/1748-0221/15/06/P06005, arXiv: 2004.08262 [hep-ex].
- [234] G. Kasieczka and D. Shih, “Robust jet classifiers through distance correlation”, *Phys. Rev. Lett.* **125**, 12 (2020), p. 122001, DOI: 10.1103/PhysRevLett.125.122001, arXiv: 2001.05310 [hep-ph].
- [235] J. Dolen et al., “Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure”, *JHEP* **05** (2016), p. 156, DOI: 10.1007/JHEP05(2016)156, arXiv: 1603.00027 [hep-ph].
- [236] K. Kölbig and B. Schorr, “A program package for the Landau distribution”, *Computer Physics Communications* **31**, 1 (1984), pp. 97–111, ISSN: 0010-4655, DOI: [https://doi.org/10.1016/0010-4655\(84\)90085-7](https://doi.org/10.1016/0010-4655(84)90085-7).

- [237] K. Kölbig and B. Schorr, “Erratum to: “A program package for the Landau distribution” [Comput. Phys. Commun. 31 (1984) 97–111]”, *Computer Physics Communications* **178**, 12 (2008), p. 972, ISSN: 0010-4655, DOI: <https://doi.org/10.1016/j.cpc.2008.03.002>.
- [238] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV”, CMS-PAS-LUM-17-004 (2018), URL: <https://cds.cern.ch/record/2621960/>.
- [239] CMS Collaboration, CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV, tech. rep. CMS-PAS-LUM-18-002, 2019, URL: <https://cds.cern.ch/record/2676164/>.
- [240] CMS Collaboration, “The CMS statistical analysis and combination Tool: COMBINE”, *Comput. Softw. Big Sci.* **8**, 1 (2024), p. 19, DOI: 10.1007/s41781-024-00121-4, arXiv: 2404.06614 [physics.data-an].
- [241] W. Verkerke and D. P. Kirkby, “The RooFit toolkit for data modeling”, *eConf C0303241* (2003), ed. by L. Lyons and M. Karagoz, MOLT007, arXiv: physics/0306116.
- [242] L. Moneta et al., “The RooStats Project”, *PoS ACAT2010* (2010), ed. by T. Speer et al., p. 057, DOI: 10.22323/1.093.0057, arXiv: 1009.1003 [physics.data-an].
- [243] F. James and M. Roos, “Minuit: A system for function minimization and analysis of the parameter errors and correlations”, *Comput. Phys. Commun.* **10** (1975), pp. 343–367, DOI: 10.1016/0010-4655(75)90039-9.
- [244] ATLAS Collaboration, “Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector”, *JHEP* **01** (2015), p. 069, DOI: 10.1007/JHEP01(2015)069, arXiv: 1409.6212 [hep-ex].
- [245] R. D. Cousins, “Generalization of chisquare goodness-of-fit test for binned data using saturated models, with application to histograms” (2013), URL: <https://api.semanticscholar.org/CorpusID:5936965>.
- [246] G. Cowan et al., “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011), p. 1554 [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)], DOI: 10.1140/epjc/s10052-011-1554-0, arXiv: 1007.1727 [physics.data-an].
- [247] “Performance of the ParticleNet tagger on small and large-radius jets at High Level Trigger in Run 3”, CMS-DP-2023-021 (2023), URL: <https://cds.cern.ch/record/2857440>.
- [248] FastML Team, fastmachinelearning/hls4ml, version v0.8.1, 2023, DOI: 10.5281/zenodo.1201549, URL: <https://github.com/fastmachinelearning/hls4ml>.
- [249] J. Duarte et al., “Fast inference of deep neural networks in FPGAs for particle physics”, *JINST* **13**, 07 (2018), P07027, DOI: 10.1088/1748-0221/13/07/P07027, arXiv: 1804.06913 [physics.ins-det].

- [250] C. N. Coelho et al., “Automatic heterogeneous quantization of deep neural networks for low-latency inference on the edge for particle detectors”, *Nature Mach. Intell.* **3** (2021), pp. 675–686, DOI: 10.1038/s42256-021-00356-5, arXiv: 2006.10159 [physics.ins-det].
- [251] Google, qkeras, URL: <https://github.com/google/qkeras>.
- [252] H. Bergauer, private communication.
- [253] I. Komarov, “Search for non-resonant Higgs boson pair production in the lepton+jets final state of the bbWW decay mode at CMS”, master’s thesis (2022), URL: <https://www.physik.uni-hamburg.de/en/iexp/gruppe-haller/scientific-output/documents/masterarbeit-mathis-frahm.pdf>.
- [254] CMS Collaboration, “Anomaly detection in the CMS global trigger test crate for Run 3”, CMS-DP-2023-079 (2023), URL: <https://cds.cern.ch/record/2876546>.
- [255] CMS Collaboration, “2024 data collected with AXOL1TL anomaly detection at the CMS Level-1 trigger”, CMS-DP-2024-059 (2024), URL: <https://cds.cern.ch/record/2904695>.
- [256] CMS Collaboration, “Level-1 Trigger Calorimeter Image Convolutional Anomaly Detection Algorithm”, CMS-DP-2023-086 (2023), URL: <https://cds.cern.ch/record/2879816>.
- [257] CMS Collaboration, “Neural Network-based primary vertex reconstruction with FPGAs for the upgrade of the CMS Level-1 Trigger System”, *J. Phys. Conf. Ser.* **2438**, 1 (2023), p. 012106, DOI: 10.1088/1742-6596/2438/1/012106.
- [258] CMS Collaboration, “Deep learning fast inference on FPGA for CMS muon Level-1 trigger studies”, *PoS ISGC2021* (2021), p. 005, DOI: 10.22323/1.378.0005.
- [259] CMS Collaboration, “The CMS Level-1 endcap muon trigger at the High-Luminosity LHC”, *PoS ICHEP2020* (2021), p. 890, DOI: 10.22323/1.390.0890.
- [260] CMS Collaboration, “Hadronic tau reconstruction in the CMS Phase-2 Level-1 trigger using NNs with calorimetric information”, CMS-DP-2023-062 (2023), URL: <http://cds.cern.ch/record/2868783>.
- [261] CMS Collaboration, “Neural network-based algorithm for the identification of bottom quarks in the CMS Phase-2 Level-1 trigger”, CMS-DP-2022-021 (2022), URL: <https://cds.cern.ch/record/2814728>.
- [262] “LS2 Report: CERN’s newest accelerator awakens”, URL: <https://home.cern/news/news/accelerators/ls2-report-cerns-newest-accelerator-awakens> (visited on 12/20/2024).
- [263] G. Arduini et al., “LHC upgrades in preparation of Run 3”, *JINST* **19**, 05 (2024), P05061, DOI: 10.1088/1748-0221/19/05/P05061.
- [264] CMS Collaboration, “Development of the CMS detector for the CERN LHC Run 3”, *JINST* **19**, 05 (2024), P05064, DOI: 10.1088/1748-0221/19/05/P05064, arXiv: 2309.05466 [physics.ins-det].

- [265] CMS Collaboration, “Enriching the Physics Program of the CMS Experiment via Data Scouting and Data Parking”, CMS-EXO-23-007 (2024), arXiv: 2403.16134 [hep-ex].
- [266] CMS Collaboration, “The Phase-2 upgrade of the CMS tracker”, CMS-TDR-014 (2017), ed. by K. Klein, DOI: 10.17181/CERN.QZ28.FLHW.
- [267] CMS Collaboration, “The Phase-2 upgrade of the CMS barrel calorimeters”, CMS-TDR-015 (2017), URL: <https://cds.cern.ch/record/2283187>.
- [268] CMS Collaboration, “The Phase-2 upgrade of the CMS muon detectors”, CMS-TDR-016 (2017), URL: <https://cds.cern.ch/record/2283189>.
- [269] CMS Collaboration, “The Phase-2 upgrade of the CMS endcap calorimeter”, CMS-TDR-019 (2017), DOI: 10.17181/CERN.IV8M.1JY2.
- [270] CMS Collaboration, “A MIP Timing Detector for the CMS Phase-2 Upgrade”, CMS-TDR-020 (2019), URL: <https://cds.cern.ch/record/2667167>.
- [271] CMS Collaboration, “The Phase-2 upgrade of the CMS Level-1 trigger”, CMS-TDR-021 (2020), URL: <https://cds.cern.ch/record/2714892>.
- [272] CMS Collaboration, “The Phase-2 upgrade of the CMS Data Acquisition and High Level Trigger”, CMS-TDR-022 (2021), URL: <https://cds.cern.ch/record/2759072>.
- [273] OpenAI, “ChatGPT 4o”, URL: <https://chatgpt.com/>.

ERKLÄRUNG ZUR VERWENDUNG GENERATIVER KI DECLARATION ON THE USAGE OF GENERATIVE AI

Den Richtlinien guter wissenschaftlicher Praxis folgend gebe ich hier Auskunft über die Verwendung generativer KI im Kontext dieser Dissertation.

Ich habe vereinzelt das Large Language Model ChatGPT [273] von OpenAI für die Erzeugung von Programmcode verwendet. Dies geschah ausschließlich für visuelle Änderungen an Grafiken sowie visuelle Arbeit an diesem Dokument. Ich habe den von ChatGPT generierten Programmcode auf Richtigkeit überprüft und manuell adaptiert. Weder nutzte ich Programmcode von ChatGPT direkt für Datenanalyse, noch generierte ich Text für diese Dissertation.

Following the guidelines for good scientific practice, here I inform about the usage of generative AI in the context of this thesis.

I rarely used the large language model ChatGPT [273] by OpenAI to create code. This was only done for visual changes to figures and visual work on this document. I ensured that the code produced by ChatGPT is correct and adapted it manually. I neither used ChatGPT to generate code directly used for data analysis, nor to generate text for this thesis.

EIDESSTATTLICHE ERKLÄRUNG DECLARATION OF OATH

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben. Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated. If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

Hamburg, den 23.12.2024

Unterschrift des Doktoranden

ACKNOWLEDGMENTS

The work I present in this thesis was only possible thanks to many people I had the pleasure of collaborating with. First and foremost, I want to thank Prof. Johannes Haller for giving me the opportunity to work on these topics as part of his research group. I have very much enjoyed the positive atmosphere and supportive leadership in the group. I also want to thank Prof. Gregor Kasieczka and Prof. Christian Schwanenberger as part of my QURS advisory council and Christian as the secondary reviewer of this thesis. Thank you to Prof. Günter Sigl and Prof. Konstantinos Nikolopoulos for being part of my examination committee.

Thank you to Dr. Roman Kogler for the mentorship and joint work on the $t^*\bar{t}^*$ analysis, which I could never have brought to publication without your help. I very much enjoyed working with you and appreciate all the time you devoted to the analysis, even when moving on to DESY. Thank you for also being part of my examination committee. Thanks to Dr. Artur Lobanov for the supervision in the context of the development of ML-based trigger strategies. I am very much looking forward to seeing some L1TOP0 iteration running in the trigger, and to see what anomalies AXOL1TL might find. I want to thank Dr. Matthias Schröder for always being open for questions on all manner of things, from physics discussions to organizational questions. The same needs to be said for Dr. Matteo Bonanomi for helping me with my various (often COMBINE-related) issues. Thanks to Dr. Daniel Savoiu for many useful discussions, on physics but also topics beyond that (like the intricacies of different languages). Also, thanks to all of you for the valuable feedback you provided when proof-reading this thesis!

Thanks to the (newly minted) Dr. Alexander Paasch! I am very happy to have shared an office for several years, and really enjoyed both our scientific and private discussions, as well as our lunch break chess matches - even despite losing the majority of them. Thank you Dr. Tobias Quadfasel for enjoyable chats about work during our evening games. Also thank you to all other PhD, master's and bachelor's students I worked with during my time in the group. I want to thank Dr. Sam Harper for our joint work on the development of the web-based HLT validation tool, during which I learned a lot about C++ development for CMSSW, as well as all the intricacies of web design. Thanks also goes out to all conveners, contacts and members of the CMS B2G subgroup who have contributed to the publication of the $t^*\bar{t}^*$ analysis, and everyone else who participated in the internal review procedure. Thank you to Andrea Bremer for the support in all organizational matters, especially the preparation of my disputation.

Thank you to my parents for all the support, enabling me to pursue my scientific journey. Thanks to all other family members and all my friends for the enjoyable time spent away from work, letting me recharge whenever I needed to.

Finally, the most important thanks goes to my wife Clara, who supported me over this entire journey. I am certain that I could not have managed to actually finish this thesis without your support. We did it, Little Rock.