GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für Physik

Master's Thesis

Studien zur Diskriminierung zwischen prompten und hadronischen Fake-Photonen unter Benutzung neuronaler Netze

Studies of the discrimination between prompt photons and hadron fakes using neural networks

prepared by

Benedikt Völkel

from Hildesheim

at the II. Physikalischen Institut

Thesis number: II.Physik-UniGö-MSc-2017/07

Thesis period: 2nd April 2017 until 1st October 2017

First referee: Prof. Dr. Arnulf Quadt

Second referee: Prof. Dr. Stan Lai

Zusammenfassung

Machine Learning Algorithmen wie Neuronale Netze finden heutzutage auf vielen unterschiedlichen Gebieten Anwendung und werden zur Lösung verschiedener Probleme genutzt. Auch in der Hochenergiephysik wächst ihr Anwendungsbereich stetig.

Diese Studie legt einen typischen Anwendungsfall dar, in dem die Diskriminierung zwischen Photonen aus harten Parton-Parton Interaktionen und solchen Objekten, die deren Detektorsignatur imitieren können, untersucht wird.

Im Gegensatz zu Likelihood-Methoden oder der Anwendung von Schnitten wird die Güte der Diskriminierungseigenschaften Neuronaler Netze grundsätzlich nicht durch das Vorhandensein von Korrelationen zwischen Eingangsvariablen reduziert. Daher können die sogenannten Shower Shapes von Photonen, die im elektromagnetischen Kalorimeter gemessen werden und korreliert sind, als Eingangsvariablen genutzt werden. Wie sich zeigt, eignen sich diese für die gewünschte Separierung, die insbesondere für Präzisionsmessungen, die Photonen beinhalten, entscheidend ist.

Die Implementierung in die Analyse der Top-Quark Paarproduktion zusammen mit einem Photon bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$ mit dem ATLAS Detektor wird erläutert. Dies beinhaltet weitere Studien der Leistungsfähigkeit des Neuronalen Netzes sowie die Ableitung seiner systematischen Unsicherheiten.

Abstract

Machine learning algorithms such as neural networks are nowadays implemented in a large variety of fields and are used to solve problems of different kinds. Also in high energy physics their field of application grows steadily.

This thesis presents a typical use case by studying the discrimination between photons originating from hard parton-parton interactions and other objects which can potentially fake their detector signatures.

Neural networks do in general not suffer from correlated input features as it is the case for likelihood approaches or cuts. Hence, correlated photon shower shapes obtained in the ATLAS calorimeter system can be used as input variables and a high separation is shown to be achieved. This is especially crucial for precision measurements involving photons.

The implementation into the analysis of top-quark pair production in association with a photon at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector is explained. This also includes further performance studies as well as the derivation of systematic uncertainties of the neural network.

Contents

1. Introduction

The Standard Model of elementary particle physics (SM) is a local gauge invariant and renormalisable quantum field theory [1–5]. Many of its predictions were and are probed by a large number of different experiments. To date it turned out to be very successful in describing the properties and fundamental interactions of the smallest constituents of the known universe. Elementary particles are nowadays often studied at facilities like the Atlas [6] detector at the Large Hadron Collider (Lhc) [7] at Cern near Geneva, Switzerland. The latest grand achievement was the discovery of a new resonance in the invariant mass spectrum of a diphoton system at the Atlas [8] and Cms [9] experiments in 2012. The measured properties of the resonance are in agreement with the SM Higgs boson which would complete the SM.

However, it is known that the SM lacks in certain aspects. A candidate for a *dark matter* particle is missing as well as a description for what is called *dark energy*. Further, it does not incorporate gravity. Neutrinos are considered to be massless which is in contradiction to the observed neutrino flavour oscillation (see, for instance, [10]). In particular, this deviation shows the need to further investigate also those particles and interactions which are already experimentally established. High precision measurements may uncover even more deviations because some small effects could still be covered by experimental uncertainties.

Many of those precision measurements are part of the research programs of experiments at the Lhc. Different approaches are used to become more precise: a better understanding of the detector and other systematic uncertainties, a larger amount of data to decrease statistical uncertainties and the improvement of analysis techniques.

This thesis focuses on the latter point investigating possible improvements of analysis techniques with respect to the discrimination between photons and so-called *hadron fakes* using a neural network (NN). Conventionally, hadron fakes are defined as either photons from final state hadron decays or hadrons which are misidentified as photons. They are one of the dominant backgrounds in many analyses involving photons and improving their discrimination enables for more precise measurements in these analyses. The NN developed in this thesis is implemented in the Atlas analysis measuring the cross section of top-quark pair production in association with a photon at a centre-of-mass energy of 13 TeV (also only referred to simply as $t\bar{t}\gamma$ analysis in the following).

The next Chapter provides an introduction to the theoretical background of the SM and introduces the top-quark. The final part of that Chapter contains the summary of the analyses of $t\bar{t}\gamma$ production at 7 TeV and 8 TeV as benchmarks for the current analysis at 13 TeV. Since it is the experimental setup needed for this analysis, the LHC and the ATLAS detector along with its components and purpose are briefly described in Ch. 3. Photons are the most important objects in this thesis and thus, Secs. 3.3 and 3.4 provide a dedicated introduction to photon reconstruction and identification at ATLAS as well as to the definitions of prompt photons and hadron fakes. This is followed by an introduction to the basic concepts of machine learning in Ch. 4 which are necessary for the development of the NN. After summarising the data, object definitions and event selection criteria in Ch. 5, the central part of this thesis is contained in Ch. 6 which explains the data preparation and training of the NN as well as the final implementation in the $t\bar{t}\gamma$ analysis. The summary of the achievements and an outlook can be found in Ch. 7.

2. Theoretical background

2.1. The Standard Model of elementary particle physics

The Standard Model of elementary particle physics (SM) describes the fundamental particles together with their interactions. Mathematically, it is formulated as a local gauge and Lorentz invariant Lagrangian density function with the underlying gauge group $\mathrm{U}(1)_{\mathrm{Y}} \times \mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{SU}(3)_{\mathrm{C}}$ which can be shown to be fully renormalisable [5]. The number of generators of the individual gauge groups corresponds to the number of so-called gauge fields with integer spin 1. They mediate interactions between the fermions, which are fundamental spin-$\frac{1}{2}$ matter fields, by coupling to the *charges* the latter ones carry. Some gauge fields are charged as well such that some gauge bosons can couple to one another. Different from the gauge fields the number of fermions is not predicted by the theory.

The charge or another property a certain gauge boson couples to is indicated by the subscript Y (hypercharge), L (left handed fermions only) or C (colour charge) of the individual gauge groups. Since some gauge bosons carry charges themselves, the respective bosons can also interact among each other. The SM is completed by the spin-0 Higgs field. A particle which is consistent with the excitation of the SM Higgs field was discovered in 2012 by the ATLAS [8] and CMS [9] experiments at the LHC at CERN. The mechanism of spontaneous symmetry breaking predicts that the $W^{\pm}$ and Z boson acquire their masses by coupling to the Higgs field. Also fermions can acquire their masses by coupling to the Higgs field.

A sketch showing the SM particles can be found in Fig. 2.1. There are 12 fermions of different so-called *flavour* grouped in 3 generations of quarks and leptons with increasing mass from first to third generation. The quark and lepton sectors are each divided into up-type and down-type fermions, respectively, which differ by one unit in the third component of the *weak isospin* I_3. Up-type fermions have $I_3 = +\frac{1}{2}$ and the down-type fermions have $I_3 = -\frac{1}{2}$. All fermions are therefore affected by the weak force mediated by the heavy $W^{\pm}$ and Z bosons.

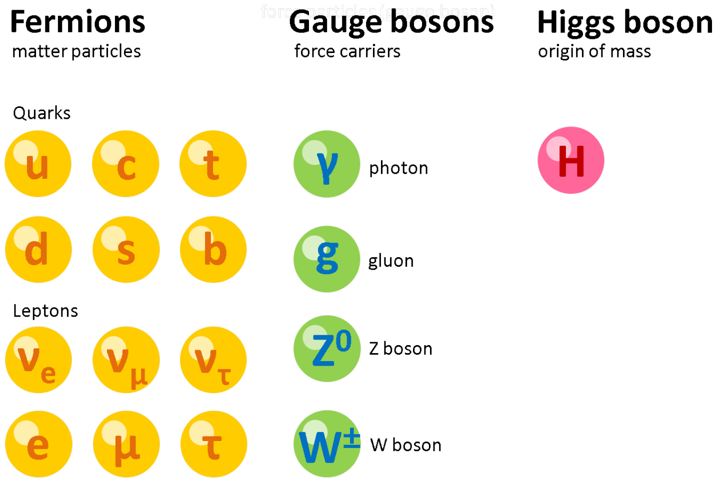


Figure 2.1.: Sketch of the SM particles.

The difference in units of electric charge Q between up-type and down-type fermions in each generation is also 1. Up-type quarks have $+\frac{2}{3}$ whereas their down-type partners have units of electric charge of $-\frac{1}{3}$. In the lepton sector the so-called charged leptons have electric charge of -1 and their up-type partners, the neutrinos, are electrically neutral. Hence, all fermions but the neutrinos participate in the electromagnetic interaction mediated by the photon. In addition to electric and weak charge, quarks carry so-called colour which is the charge of Quantum Chromodynamics (QCD). Due to the $\mathrm{SU}(3)_\mathrm{C}$ group structure the SM incorporates 8 gluons coupling to colour. These gluons also carry colour themselves and hence couple to one another via triple and quartic gauge couplings. In the electroweak sector quartic gauge couplings are described for $W^\pm$ which have $I_3 = \pm 1$. Triple gauge coupling is possible in the case of a $W^\pm$ pair interacting with a Z or a photon, respectively. The Higgs boson couples to all massive particles proportional to their masses and hence to itself via triple and quartic interaction. The SM does not describe couplings between the Higgs boson and photons, gluons and neutrinos, respectively.

2.1.1. The electroweak theory and the Higgs mechanism

Due to the requirement of local gauge invariance and renormalisability of the SM, all gauge bosons have to be massless. However, the $W^\pm$ and Z bosons are found to be massive with $m_Z = 91.1876 \pm 0.0021$ GeV and $m_W = 80.385 \pm 0.015$ GeV [11]. Hence, these bosons cannot directly correspond to the generators of any of the gauge groups. A theoretical solution to this problem was first proposed by Glashow [1], Weinberg [2] and Salam [3], nowadays known as GWS theory. It keeps the group structure as a whole but entangles the $\mathrm{U}(1)_\mathrm{Y} \times \mathrm{SU}(2)_\mathrm{L}$ group as a consequence of *spontaneous symmetry breaking*.

This mechanism was first described by Brout and Englert [12], Higgs [4] and Guralnik, Kibble and Hagen [13] for the general case of non-Abelian gauge theories and gives a possible explanation of how gauge bosons can acquire mass in a local gauge invariant theory. In the SM this is implemented by introducing a complex valued isospin doublet

$$\mathcal{H} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.1}$$

The starting point of the derivations is the part of the SM Lagrangian representing the coupling between the Higgs field and the gauge fields according to

$$\mathcal{L}_{\mathcal{H}} = \left(\mathcal{D}_\mu \mathcal{H}\right)^\dagger \left(\mathcal{D}^\mu \mathcal{H}\right) - \underbrace{\left[-\mu^2 \mathcal{H}^\dagger \mathcal{H} + \lambda \left(\mathcal{H}^\dagger \mathcal{H}\right)^2\right]}_{\text{Higgs potential } V(\mathcal{H})},$$

where $\mathcal{D}_\mu$ is the so-called covariant derivative given as

$$\mathcal{D}_\mu = \left(\partial_\mu - i g_W W_\mu^a \tau^a - \frac{i}{2} g' B_\mu\right).$$

It is $\tau^a = \sigma^a/2$ (with Pauli matrices σ^a), W_μ^a $(a = 1, 2, 3)$ are the gauge fields of $\mathrm{SU(2)_L}$ and B_μ is the gauge field of $\mathrm{U(1)_Y}$. Summing over the index a is implied. g_W and g' are independent coupling constants[1] and the gauge fields of different gauge groups commute with each other. For $\mu^2 > 0$ and $\lambda \neq 0$ the Higgs potential has a global minimum at $v^2 = \phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 > 0$, where v is called vacuum expectation value (VEV). The minimum is the most probable and hence expected field configuration. It can be shown that the coupling to the gauge bosons of *unbroken* $\mathrm{SU(2)_L}$ and $\mathrm{U(1)_Y}$ symmetries to the Higgs doublet in the so-called *unitary gauge* can be written as

$$\mathcal{L}_h = \begin{pmatrix} 0 & v+h \end{pmatrix} \left(g_W W_\mu^a \tau^a + \frac{1}{2} g' B_\mu\right) \cdot \left(g_W W^{b\mu} \tau^b + \frac{1}{2} g' B^\mu\right) \begin{pmatrix} 0 \\ v+h \end{pmatrix}, \tag{2.2}$$

where h is the physically visible and potentially massive Higgs boson of the theory. To see how the gauge fields acquire masses Eq. (2.2) has to be evaluated for v. The Lagrangian density can then be written as

$$\mathcal{L}_v = \frac{v^2}{8} \left[g_W^2 \left(W_\mu^1\right)^2 + g_W^2 \left(W_\mu^2\right)^2 + \left(-g_W W_\mu^3 + g' B_\mu\right)^2\right].$$

[1] These parameters are no constants in the strict sense but their value depends on the energy scale they are evaluated at. This energy dependence will be discussed for the coupling in the strong interaction later on.

Re-writing this in terms of mass eigenstates yields 2 oppositely electrically charged fields

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right), \qquad \text{with mass} \qquad m_W = g_W \frac{v}{2},$$

one electrically neutral and massive field

$$Z_\mu = \frac{1}{\sqrt{g_W^2 + g'^2}}\left(g_W W_\mu^3 - g' B_\mu\right), \qquad \text{with mass} \qquad m_Z = \sqrt{g_W^2 + g'^2}\frac{v}{2}$$

and a massless photon field

$$A_\mu = \frac{1}{\sqrt{g_W^2 + g'^2}}\left(g' W_\mu^3 - g_W B_\mu\right).$$

The charged weak bosons are a superposition of only $\mathrm{SU(2)_L}$ fields whereas the massless photon field A_μ and the massive Z_μ are a superpositions of $\mathrm{SU(2)_L}$ *and* $\mathrm{U(1)_Y}$ gauge fields. Since the latter two are orthogonal to each other, this immediately reflects the fact that there is no coupling between the photon and the Z boson. The superposition arises after a change of basis which can be written as

$$\begin{pmatrix} Z \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} W^3 \\ B \end{pmatrix},$$

where θ_w is the *electroweak mixing* or *Weinberg* angle with

$$\cos\theta_w = \frac{g_W}{\sqrt{g_W^2 + g'^2}} \qquad \text{and} \qquad \sin\theta_w = \frac{g'}{\sqrt{g_W^2 + g'^2}}.$$

Besides entangling of $\mathrm{SU(2)_L}$ and $\mathrm{U(1)_Y}$ groups which causes massive gauge bosons, Eq. (2.2) also gives rise to triple and quartic couplings between the Higgs and the massive $W^\pm$ and Z, where the coupling strength is proportional to the squared mass of the corresponding gauge boson. On the other hand, no coupling to the photon emerges.

2.1.2. CP violation and fermion masses

In addtition to the absolute value of the coupling strength between certain particles, the *structure* is characteristic as well. In 1956, Wu studied the spins of electrons originating from the β^- decay in $^{60}\mathrm{Co} \longrightarrow ^{60}\mathrm{Ni} + e^- + \bar{\nu}_e$ [14]. The spin orientations of the Cobalt nuclei were aligned by cooling and placing them in a magnetic field.

It was found that the spins of the emitted electrons were oriented opposite to the spins of the Cobalt in almost all cases. This means that the charged weak interaction, which underlies the β^- decay as it is known today, is *maximally parity violating*. $W^\pm$ bosons couple *only* to fermions (anti-fermions) in left (right) handed *chirality* state which is encoded in the subscript L in $SU(2)_L$. Exchanging a $W^\pm$ boson changes the third component of the weak isospin I_3 by ± 1 unit and hence, left handed fermion states come in isospin doublets of *Dirac spinors* representing the fermion fields

$$\begin{array}{llll} \text{leptons:} & \begin{pmatrix} \nu_e \\ e \end{pmatrix}_\mathrm{L}, & \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_\mathrm{L}, & \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_\mathrm{L}, \\ \text{quarks:} & \begin{pmatrix} u \\ d' \end{pmatrix}_\mathrm{L}, & \begin{pmatrix} c \\ s' \end{pmatrix}_\mathrm{L}, & \begin{pmatrix} t \\ b' \end{pmatrix}_\mathrm{L}. \end{array}$$

Right handed chiral states are singlets with respect to the charged weak interaction and have $I_3 = 0$. The SM contains

$$\begin{array}{llll} \text{leptons:} & e_\mathrm{R}, & \mu_\mathrm{R}, & \tau_\mathrm{R}, \\ \text{quarks:} & u_\mathrm{R}, & c_\mathrm{R}, & t_\mathrm{R}, \\ & d_\mathrm{R}, & s_\mathrm{R}, & b_\mathrm{R}. \end{array}$$

Right (left) handed neutrinos (anti-neutrinos) have not been observed yet and are not included in the SM. First evidence for that was given by the famous Goldhaber experiment in 1957 [15][2].

The left handed doublets shown above are eigenstates of the charged weak interaction. After the discovery of the charm-quark at Slac [16] and at Bnl [17] the parton model was taken more seriously and quarks were accepted as physical entities. For instance, by observing the decay of $K_s^0 \to \pi^0\pi^0$ it could be inferred that quark flavour is not conserved since no strange quark is present in the final state. Hence, mass eigenstates of quarks and those of charged weak interaction do not coincide. At the time when the third generation of quarks has not been found, Cabbibo [18] proposed a change of basis between the respective eigenstates by introducing a unitary 2×2 matrix and the so-called Cabbibo angle θ_C. Later, the idea was extended to a 3×3 matrix by Kobayashi and Maskawa [19] which implies a third generation of quarks.

[2] The experiment actually studied the neutrino *helicity*. In the SM neutrinos are massless which means that helicity and chirality are equivalent. Therefore, the SM does not incorporate left (right) chirality states of neutrinos (anti-neutrinos).

With this idea they were able to incorporate the phenomenon of violation of charge *and* parity (CP-violation) which was observed in neutral Kaon decays by Christenson, Cronin and Fitch in 1964 [20]. CP-violation is possible in this Ansatz due to a complex phase which is a free parameter and allows for different squared matrix elements of two CP transformed processes. Although the diagonal entries are close to unity, there are small contributions on the off-diagonals reflecting the fact that weak and mass eigenstates of quarks are not identical. This unitary transformation matrix also incorporating the CP-violating complex phase is known since then as the CKM[3] matrix [11].

The parity violating property of weak interactions can mathematically be implemented by a $(\mathrm{V}-\mathrm{A})$ vertex structure[4] proportional to

$$\frac{\gamma^{\mu}\left(c_{\mathrm{V}}-c_{\mathrm{A}}\gamma^{5}\right)}{2}, \tag{2.3}$$

where the γ^{μ} matrices ($\mu = 0, 1, 2, 3$) represent the operators satisfying the Dirac algebra (see also [21]) with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. In the case of the maximally parity violating $W^{\pm}$ exchange c_{V} and c_{A} are equal to unity. This is proportional to the operator which projects out the left handed chiral part f_{L} of a fermion spinor f according to

$$P_{\mathrm{L}} = \frac{1}{2}\left(1-\gamma^{5}\right) \qquad \Rightarrow \qquad P_{\mathrm{L}}f = f_{\mathrm{L}}$$

and hence the interaction vertex of $W^{\pm}$ boson has the form

$$\frac{-ig_W}{\sqrt{2}}\frac{\gamma^{\mu}\left(1-\gamma^{5}\right)}{2}.$$

The Z boson exchange is not maximally parity violating and the coefficients are given by superpositions of charges $c_{\mathrm{V}}^{f} = I_3^f - 2Q_f \sin^2\theta_w$ and $c_{\mathrm{A}}^{f} = I_3^f$ depending on the fermion f. This directly corresponds to the mixing of the gauge fields of $\mathrm{U}(1)_{\mathrm{Y}}$ and $\mathrm{SU}(2)_{\mathrm{L}}$ as discussed in the previous Section and the complete Z vertex structure is given by

$$-i\frac{g_W}{\cos\theta_w}\frac{\gamma^{\mu}\left(c_{\mathrm{V}}^{f}-c_{\mathrm{A}}^{f}\gamma^{5}\right)}{2}. \tag{2.4}$$

[3]Named after Cabbibo, Kobayashi and Maskawa.
[4]V − A means 'vector minus axial-vector'.

Electromagnetic interactions mediated by the photon are *parity conserving* and the interaction vertex is therefore only vector-like and proportional to

$$iQ_f e\gamma^\mu, \tag{2.5}$$

where the coupling strength is $e = g_W \sin\theta_w$. From the GWS theory it follows that the electric charge Q_f is a superposition of hypercharge Y_f and the third component of weak isospin I_3^f according to

$$Q_f = I_3^f + \frac{Y_f}{2}.$$

Since left and right handed chiral states transform differently under $SU(2)_\mathrm{L}$ transformation, fermion mass terms of the form $m_f\bar{f}f = m_f(\bar{f}_\mathrm{L}f_\mathrm{R} + \bar{f}_\mathrm{R}f_\mathrm{L})$ cannot be explicitly added to the Lagrangian density function. However, it is possible to incorporate fermion masses by making use of spontaneous symmetry breaking mechanism. A term proportional to

$$\lambda_e \begin{pmatrix}\nu_e & e\end{pmatrix}_\mathrm{L} \mathcal{H}\, e_\mathrm{R} + \mathrm{h.c.}$$

reproduces the desired mass term of electrons in the Lagrangian. This can be done in an analogous way for the other fermions as well. A fermion f acquires a mass of

$$m_f = (\lambda_f v)/\sqrt{2}, \tag{2.6}$$

where λ_f is the Yukawa coupling constant.

2.1.3. Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the quantum field theory describing the strong force with underlying group $\mathrm{SU}(3)_\mathrm{C}$. The interaction is mediated by 8 gluons coupling to colour charge. Quarks carry one colour while gluons carry colour and anti-colour which leads to gluon self-coupling. While $\mathrm{U}(1)_\mathrm{Y}$ and $\mathrm{SU}(2)_\mathrm{L}$ are spontaneously broken, the $\mathrm{SU}(3)_\mathrm{C}$ symmetry remains unbroken. Hence, there is no interaction between the Higgs and gluon fields which are therefore massless.

Running coupling, asymptotic freedom and confinement

The coupling α_S of QCD is not a constant but depends on the energy scale Q it is evaluated at as shown in Fig. 2.2. Its values cannot be calculated from first principles but it can be extrapolated to scales Q after α_S has been experimentally determined for a reference scale M according to

$$\alpha_S(Q) = \frac{\alpha_S(M)}{1 + \frac{b_0 \alpha_S(M)}{6\pi} \ln \frac{Q}{M}}, \tag{2.7}$$

where $b_0 = 11N_C - 2n_f$. N_C is the number of colours which is 3 in the SM and n_f is the number of approximately massless quarks at the energy scale Q. At energy scales explored so far 6 quarks are known and by definition it is always $b_0 > 0$. Therefore, it is expected that α_S decreases as $1/\ln(Q)$ and hence with increasing energy. At small energy scales where $\alpha_S \sim 1$ perturbative expansions in terms of α_S are not reliable. Extrapolations according to Eq. (2.7) near energy scales of order 1 GeV have to be done carefully and are not valid below such scales. The fact that perturbation theory as an expansion in the strong coupling can be applied in QCD at large enough energy scales is known as *asymptotic freedom*. For instance, at the mass of the Z boson it is measured to be $\alpha_S(m_Z) = 0.1182 \pm 0.0012$ [11]. According to the measurements presented in Fig. 2.2 it is well below 1 for energies above 10 GeV.

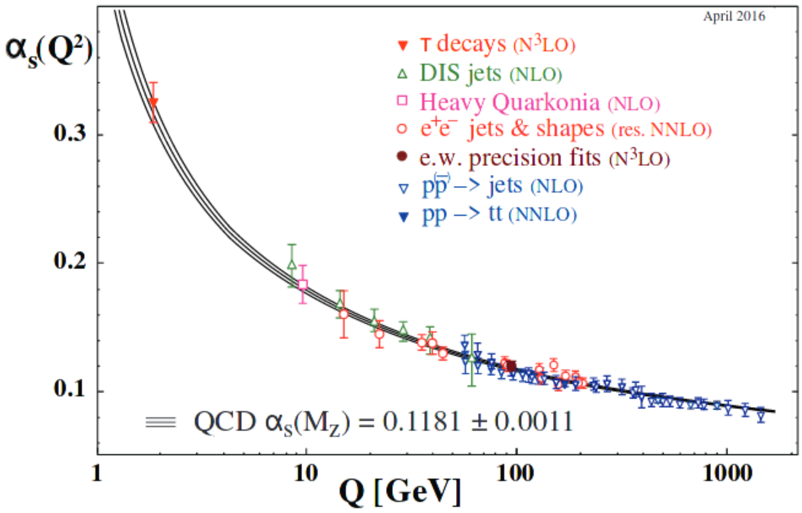


Figure 2.2.: Measurements of running α_S [11] at different experiments and energy scales. From its value at the reference scale of the Z mass the theoretical extrapolation according to Eq. (2.7) with its uncertainty band shows a good agreement with the data.

At energy scales experimentally reached so far no coloured particle has been observed yet. All observed objects are colour singlets, so-called *hadrons*, and non of those is affected by the strong force as such. This property of QCD is called *confinement*. Confined states observed so far are either bound states of 3 quarks (like uud in the case of a proton) or those of a quark-antiquark pair (like $u\bar{d}$ in the case of a π^+). Also tetra or penta quarks which are bound states of four and five quarks/anti-quarks, respectively, could theoretically exist. In 2015, the LHCb experiment [22] at CERN observed a resonance consistent with a bound state of 5 quarks [23]. In lattice QCD studies it can be shown that confinment can be a property of non-Abelian gauge theories which is equivalent to self-interacting gauge fields. Confinement arises at the similar energy scales where asymptotic freedom breaks down[5]. Lattice QCD approaches provide hence a field for studying the non-perturbative regime of QCD (see also [24]). Although colour confinement has not been fully understood yet, a more descriptive approach is to think about two separating quarks being connected by a colour field as sketched in Fig. 2.3. This colour field occurs because of the self-interaction of gluons. The colour tube between the two quarks has a constant energy density and hence, drawing them further apart increases the energy in that tube. This acts against the separation of the two quarks which decreases their kinetic energy. If the energy in the tube is large enough, quark/anti-quark pairs can be produced out of the vacuum which decreases the potential energy. At some point the kinetic energies of the quarks are low and the attracting force keeps them together in colourless bound states.

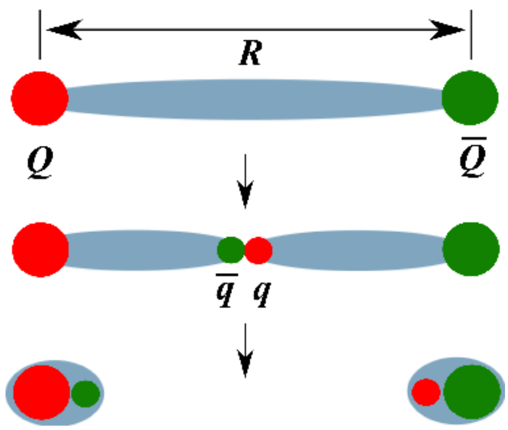


Figure 2.3.: Sketch of colour confining tubes in QCD.

[5]That does not imply that the source of asymptotic freedom and confinement is the same and both phenomena can be in gerenal independent.

Deep inelastic scattering and parton distribution functions

Probing Gold atoms by shooting α particles onto a Gold foil [25], Rutherford found that atoms have a sub-structure. The conclusion was that they are made of a heavy electrically positively charged nucleus and a cloud of electrons moving around it. The model of atoms has evolved further in the following years and today it is known that also protons and neutrons are bound states of quarks. This was first observed in electron-proton collisions where it was found that the energy and angular distribution of the scattered electron was not compatible with the assumption of point-like protons. These *deep inelastic scattering* experiments were then used to determine the sub-structure of protons [26]. A set of parton distribution functions (PDF) summarises this sub-structure by providing the probability to find a certain parton (quark or gluon) carrying a specific fraction of the total energy of a proton. Fig. 2.4 shows an example of a PDF. Gluons are most likely to carry low momentum fraction. Further, there are two maxima for u- and d-quarks at similar momentum fractions. Since the value of the maximum of the u-quark is roughly twice as large as the one of d-quarks, this is compatible with the earlier interpretation that a proton is a bound state of the form uud which are called *valence* quarks.

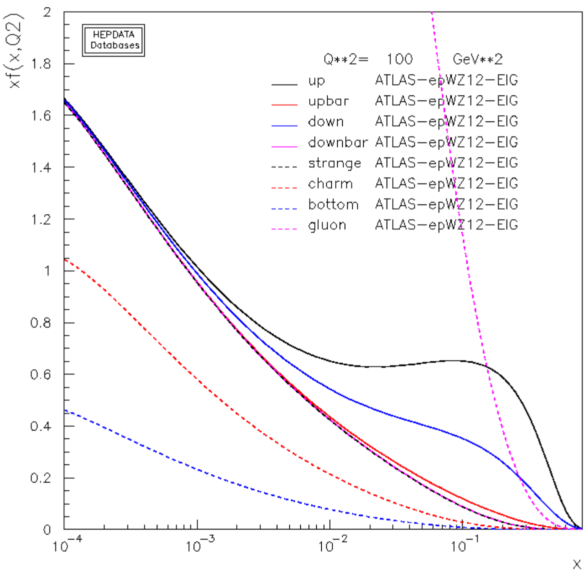


Figure 2.4.: Parton densities in a proton at a scale of 10 GeV extracted using [27].

Due to the uncertainty principle partons can be produced out of the vacuum on short time scales and it follows that the parton distributions look different depending on the energy scale the proton structure is probed at. However, once determined at a scale Q^2, the PDFs can be calculated for other scales using the so called $DGLAP^6$ equations [28–30].

[6]Named after the authors Dokshitzer, Gribov, Lipatov, Altarelli and Parisi.

Cross sections of hadron collisions

The underlying interaction in hadronic collisions takes place between partons. Due to the factorisation theorem (see for instance [31]) the proton-proton fully *inclusive* cross section for the final state Y can be computed according to

$$\sigma(\sqrt{s})_{pp\to Y+X} = \sum_{a,b\,\in\,\{\text{partons}\}} \int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b\; f_a(x_a,\mu_\mathrm{f}^2) f_b(x_b,\mu_\mathrm{f}^2)\cdot\sigma(\sqrt{\hat{s}},\mu_\mathrm{f},\mu_\mathrm{R})_{ab\to Y+X}. \tag{2.8}$$

μ_f is the *factorisation* scale where the PDFs are probed at and μ_R is the *renormalisation* scale. To obtain the full cross section it has to be integrated over all possible momentum fractions $x_{a/b}$ and summed over all partons a and b. $\sqrt{s}$ is the centre-of-mass energy of the proton-proton system while $\sqrt{\hat{s}}$ is that of the partonic system. A short calculation shows that the centre-of-mass energy in the parton-parton system can be approximated by $\hat{s} = x_a x_b s$ at high energies where transverse momenta of the partons can be neglected.

Jets

The production of partons in hard interactions leads to so called *jets* which are collimated sprays of final state particles. Monte Carlo (MC) simulations use the fact that additional parton radiation is enhanced in the collinear and soft phase space regions of the emitting parton which leads to the approach of *parton showers*. It is used to describe the transition between single high energetic partons produced in the hard interaction to many low energetic partons which eventually organise in hadrons due to confinement. A description of the latter transition derived from first principles has not been found yet and hence, so-called *fragmentation* models (reviewed in [32]) are applied. In the entire chain not only hadrons but also some leptons and photons can be produced. First evidence for jet structure has been found at the Spear experiment at Slac [33, 34]. Since then jets have been investigated extensively, for instance in 3-jet events at Lep, in order to study the gluon self-coupling and the strong coupling [35, 36].

It is not feasible to kinematically describe every single particle originating from parton production in a hard scattering. Clustering algorithms [37] group those particles into jets and provide a well defined object. One heavily used at hadron collider experiments such as Atlas or Cms is the anti-k_T algorithm [38]. It has been shown that it is infrared and collinear safe which means that additional soft or collinear emission of particles in the jet evolution do not alter the clustering of particles and the obtained jet cone. Furthermore, soft particles around a high energetic one yield a nearly circular jet shape in the $\eta-\phi$ plane.

2.2. The top-quark

The value of $m_t = 172.44 \pm 0.13(\text{stat}) \pm 0.47(\text{syst})$ GeV [39] is the best single measurement of the top quark mass and it is thus the heaviest elementary particle known today. It was discovered in 1995 at the collider experiments CDF [40] and DØ [41] at the TEVATRON. Due to its high mass it could not be produced in any collider experiment before and hence there are 18 years between the discovery of the bottom [42] and top quark. All measured properties are in agreement with the SM.

The theoretically predicted value of the top quark width, taking next-to-leading order QCD and electroweak corrections into account [43], can be written as

$$\Gamma_t = \frac{G_\mathrm{F} m_t^3}{8\pi\sqrt{2}} \left(1 - \frac{m_W^2}{m_t^2}\right)^2 \left(1 - \frac{2m_W^2}{m_t^2}\right) \left[1 - \frac{2\alpha_\mathrm{S}}{3\pi} \left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right],$$

leading to $\Gamma_t \approx 1.33\,\text{GeV}$ by assuming $m_t \approx 172.44\,\text{GeV}$ [39], $m_W \approx 80.4\,\text{GeV}$ and $\alpha_\mathrm{S}(m_Z) \approx 0.118$ [11]. The width corresponds to a lifetime of the order of 10^{-25} s which is approximately 2 orders of magnitude smaller than the time scale of hadronisation[7]. Hence, it decays before it forms hadronic bound states transferring its spin and kinematic properties directly to its daughter particles. This makes it possible to infer back on the top-quark's bare properties by studying the decay products which is unique in the quark sector. All other quarks form bound states before their decay and hence their bare kinematics get decorrelated after forming bound states.

2.2.1. The top-quark in the strong interaction

In the strong interaction the top-quark can only be produced in $t\bar{t}$-pairs because it is flavour conserving. The top-quark decay is also not possible via the strong force for the same reason. In $p\bar{p}$ collisions at the TEVATRON, a $t\bar{t}$ pair was in approximately 85 % of the cases produced via annihilation of a $q\bar{q}$ pair because they are available as valence quarks in the proton and anti-proton, respectively. At $\sqrt{s} = 1.96$ TeV their contributions dominate the other PDFs at $x_{q/\bar{q}} \approx 0.6$ when assuming a minimum of $2m_t = \sqrt{\hat{s}} = \sqrt{x_q x_{\bar{q}} s}$ in the parton-parton frame[8]. This is different at the LHC where protons collide at centre-of-mass energies of more than 7 TeV. Anti-quarks are not available as valence quarks and due to the PDFs the dominant production mechanism is gluon-gluon fusion in approximately 90 % of the cases and even larger ratios for higher energies.

[7] Hadronisation takes place on time scales of the order of 10^{-23} s. This can be estimated by taking the inverse of the energy scale at which confinement occurs which is $\approx 1\,\text{GeV}$.

[8] That is a naive approach since both quarks are assumed to carry the same momentum fraction. However, it gives the idea of how to work with PDFs to get at least rough estimates.

2.2.2. Top quark in the weak interaction

Since the strong force is flavour conserving and no flavour changing neutral currents have been observed yet [44–50], the only possible way to decay to a lighter particle is via the charged weak force. Given the absolute value $|V_{tb}| = 1.009 \pm 0.031$ [11] of the CKM matrix the mass eigenstates in the third generation almost coincide with the eigenstates of the charged weak force. Hence, the decay of the top quark almost exclusively happens via

$$t \rightarrow W^+ b,$$

where b is meant to be a mass eigenstate of the bottom-quark. This decay is always assumed for the top quark in the following.

Single top-quark production is possible since mediating a $W^\pm$ changes the flavour. It was first observed at CDF [51] and DØ [52] in 2009. Studying single top production gives a deeper insight into the Nature of the interaction of the weak charged gauge bosons.

The $W^\pm$ boson can decay either hadronically into a $q\bar{q}'$ (in approximately 67 % of the cases) or leptonically into $l\nu_l$ (in approximately 33 % of the cases) [11]. The latter scenario yields missing transverse energy in a hadron-hadron collision since neutrinos traverse the detector experiments considered in the following without interacting with the material. For the b-quark only decays *across* quark generations are possible since the top-quark is too heavy. However, these contributions are highly suppressed as indicated by the corresponding entries in the CKM matrix of $|V_{cb}| \approx 5 \cdot 10^{-2}$ and $|V_{ub}| \approx 4 \cdot 10^{-3}$ [11]. Therefore, the decay of a hadron containing a bottom-quark is also suppressed which leads to a relatively large lifetime compared to other heavy hadrons. This property is used in collider experiments to tag jets containing a b-hadron by looking for decay vertices which are displaced from the primary interaction vertex due to a longer lifetime of the b-hadron. By doing so, jets which are likely to be initiated by b-quarks can be detected.

The decay mode of the top-quark can thus be characterised by that of the $W^\pm$. In $t\bar{t}$ production this leads to 3 specific final state topologies which are classified as *full-hadronic*, *single lepton* or *dilepton* with respective fractions shown in Fig. 2.5. In 46 % of the cases both bosons decay hadronically where at least 2 b-jets and at least 6 jets in total are expected. Without b-tagging there are 6! possibilities to assign a parton to a jet and furthermore, the hadronic final state is hardly distinguishable from QCD multi-jet production. The dilepton channel has a clean signature with at least 2 b-jets and 2 oppositely charged leptons. One drawback is the production of 2 neutrinos because their separate contributions to missing measured momentum cannot be disentangled on an event-by-event basis since the system is underconstrained.

Although the corresponding branching ratio is only 9 %, it is possible to incorporate it in current analyses due to the growing amount of data[9]. Typically, only final states with electrons and muons are considered which leaves approximately half of the full dilepton branching ratio. However, it still has to be dealt with the possible leptonic decay of tuas causing the appearance of electrons and muons as well.

The final state topology which is most often worked with is the single lepton channel. At least 2 b-jest are expected as well as at least 4 jets in total together with a charged lepton and a significant amount of missing transverse momentum. The kinematic reconstruction can be done in a more straightforward way compared to other channels. Knowing the mass of the $W^{\pm}$, the z-component of the missing momentum can be reconstructed. By measuring energies and angles it is also possible to assign the jets to the respective matching top-quark since the combinatorics decreases significantly by roughly a factor of $5 \cdot 6 = 30$ compared to the full-hadronic channel. Typically, only the decay products of the τ lepton can be observed and it has to be considered separately. The term *leptonic* refers to $l = e, \mu$ in following.

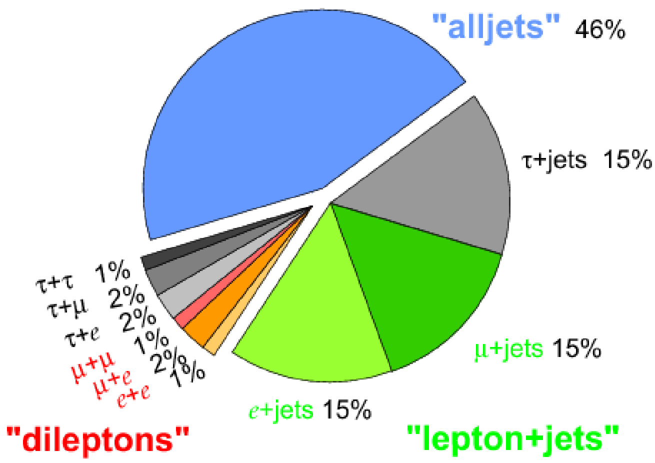


Figure 2.5.: Branching ratios of possible final state topologies of the $t\bar{t}$ system.

An interesting process with respect to top quark pair production is $pp \to t\bar{t}Z$ where a final state top quark or an initial state quark radiates a Z boson. It is of particular interest since the coupling to the Z boson contains information about the third component of the weak isospin of the top quark according to Eq. (2.4).

Although this process has already been observed [53] to be consistent with the SM, the property I_3 of the top-quark has not been measured yet. Finding it to be $+\frac{1}{2}$ would strongly support the hypothesis that this quark is indeed the weak isospin partner of the bottom-quark in the left handed chiral doublet.

[9]For that reason, the $e\mu$-channel provides the most precise measurements in current analyses due to the production of two differentiable charged leptons.

2.2.3. Top quark and the Higgs boson

Since the coupling strength of the Higgs to fermions is proportional to the mass of the fermions (see Eq. (2.6)), it is expected to be approximately 1 in the case of the top-quark. This is the largest Yukawa coupling in the SM and the top quark is therefore of particular interest and a good candidate for physics beyond the SM, such as supersymmetry or other models with more than one Higgs boson. The coupling has been observed indirectly in the Higgs production and decay via virtual triangular loops as shown in Fig. 2.6. Studies to measure the top-Higgs coupling directly in the production of $pp \to t\bar{t}H$ are ongoing.

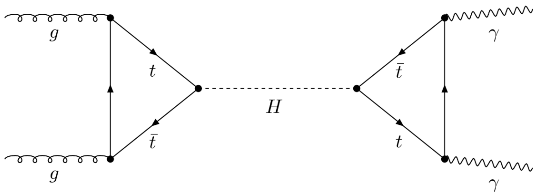


Figure 2.6.: Higgs boson production and decay via virtual triangular top quark loops. The contribution of the top-quark in the loop is the dominant one in the fermionic sector due to its large mass.

2.2.4. Top quark in the electromagnetic interaction, $pp \to t\bar{t}\gamma$

The process of $pp \to t\bar{t}\gamma$ was already established at ATLAS at $\sqrt{s} = 7$ TeV [54] and the total cross section is in agreement with the SM. The leading order (LO) contributions to this process are given by gluon-gluon fusion and quark-antiquark annihilation as depicted in Fig. 2.7 and 2.8, respectively. The photon is radiated either off a top quark in the final state or an initial state quark. In the SM the coupling between the photon and the top-quark is expected to be a vector coupling as given in Eq. (2.5). Further precision measurements are necessary in order to get sensitive to possible deviations from that expectation. Since the cross section is directly related to the coupling strength e which is in turn related to the electric charge, this process is sensitive to both. In some models beyond the SM, contributions from tensor couplings [55, 56] are included by terms proportional to

$$iQ_t e \left[\gamma_\mu \left(W_1^{\mathrm{V}} + W_1^{\mathrm{A}}\gamma^5\right) + i\frac{\sigma_{\mu\nu}}{2m_t} q_\nu \left(W_2^{\mathrm{V}} + W_2^{\mathrm{A}}\gamma^5\right)\right].$$

$W_1^{V/A}$ are the form factors of vector/axial-vector coupling and $W_2^{V/A}$ encode the electric/magnetic dipole moment of the top quark. In the SM all contributions but W_1^{V} vanish at LO. Neither the cross section nor the top quark charge measurements have shown any disagreement with the SM [57–59].

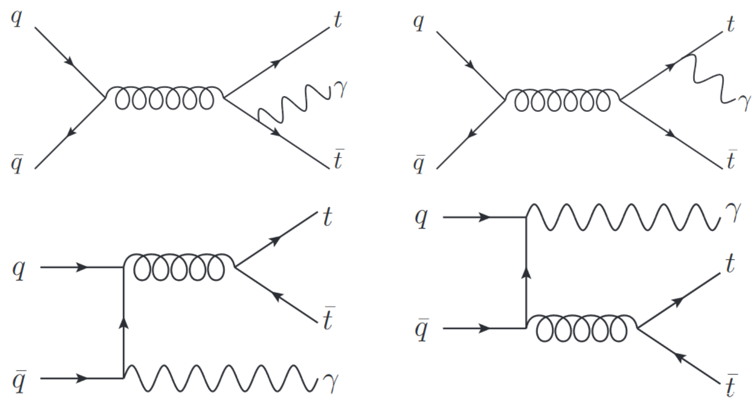


Figure 2.7.: Leading order Feynman diagrams of the production of a top-quark pair in association with a photon via quark-antiquark annihilation.

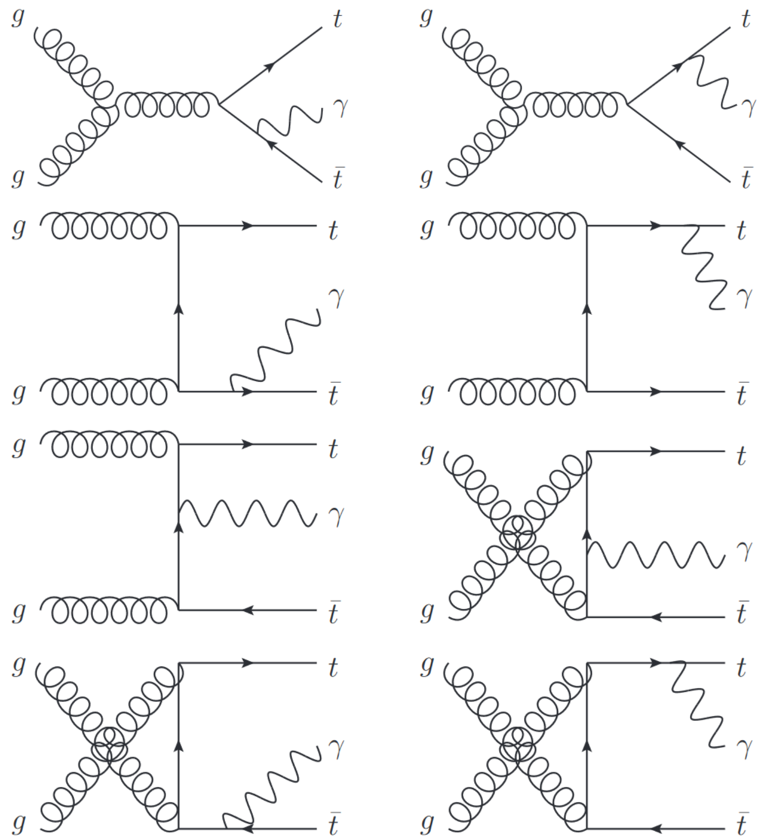


Figure 2.8.: Leading order Feynman diagrams of the production of a top-quark pair in association with a photon via gluon-gluon fusion.

Furthermore, possible contributions of flavour changing neutral currents might become visible with higher precision. In order to precisely measure the $t\gamma$ coupling, photons radiated off by a charged decay product of the top quark have to be taken into account as well as shown in Fig. 2.9.

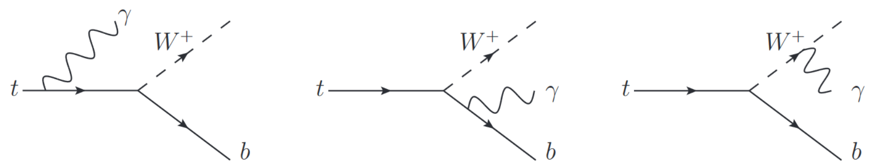


Figure 2.9.: Possible photon radiation off a top-quark or its charged decay products. Branching off a final state charged lepton originating from a leptonic $W^{\pm}$ decay can also be included.

2.2.5. Summary of the $t\bar{t}\gamma$ analysis strategy and results at 7 and 8 TeV centre-of-mass energy

The $t\bar{t}\gamma$ analyses at 7 TeV [54] and 8 TeV [60] are measurements of the total production cross section of the process $pp \to t\bar{t}\gamma$ in the single lepton channel in a fiducial kinematic region. At 7 TeV the cross section is found to be $\sigma_{\text{fid}} = 63 \pm 8(\text{stat.})^{+17}_{-13}(\text{syst.}) \pm 1(\text{lumi.})$ fb. This result is in agreement with theoretical prediction of 48 ± 10 fb obtained in QCD next-to-leading order calculations in the decay. It was the first observation of this process with a significance of 5.3 standard deviation away from the null-hypothesis. The fiducial cross section at 8 TeV is measured to be $\sigma_{\text{fid}} = 139 \pm 7(\text{stat.}) \pm 17(\text{syst.})$ fb which is in agreement with the theoretical prediction of 151 ± 24 fb. In addition to the total cross section measurement, differential cross sections in 5 bins of transverse momentum and pseudo rapidity of the photon are included.

Hadrons misidentified as photons or photons originating from hadronic decays in the final state yield a dominant background contribution. In both analyses the strategy to estimate these so-called *hadron fakes* is the same, namely deriving templates for *prompt photons* originating from the hard interaction and hadron fakes[10] using the isolation variables $p_{\text{T}}^{\text{cone20}}$ or $p_{\text{T}}^{\text{iso}}$. It is defined as the scalar sum of momentum in the transverse plane measured in a cone with radius $R = 0.2$ around the photon candidate in the inner detector where the contribution of the photon itself is subtracted. At 7 TeV both the prompt photon and hadron fake template were derived from data whereas the prompt photon template in the 8 TeV analysis was derived from simulated events. The templates are then fitted to data using a maximum likelihood fit in order to extract the respective contributions. As an example, Fig. 2.10 shows the templates fitted to data in the single electron channel of the 7 TeV analysis. Prompt photons are well isolated whereas hadron fakes show a poor isolation and in general a large contribution to the total data yield. The fit yields 52 ± 14 signal events and 38 ± 26 events where a hadron fake was selected.

[10]Since there is also a contamination of prompt photons from other background processes such as $W\gamma$, it is accounted for those as well to disentangle their contribution from that of the $t\bar{t}\gamma$ process.

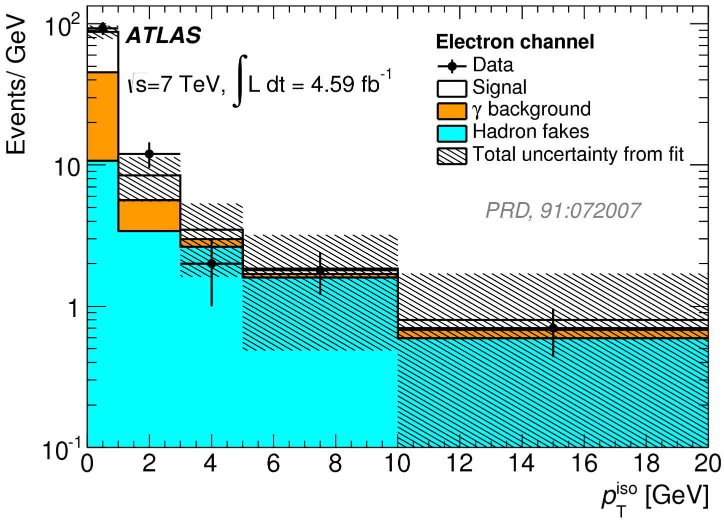


Figure 2.10.: Templates of prompt photons from $t\bar{t}\gamma$ signal events, prompt photons from background events and hadron fakes fitted to data to estimate the $t\bar{t}\gamma$ production cross section in the single electron channel at 7 TeV centre-of-mass energy [54].

3. The ATLAS experiment at the LHC

3.1. The LHC

In 2009, the Large Hadron Collider (LHC) [7] based at CERN started operating. It is placed around 100 m underground near Geneva, Switzerland, and has a circumference of $\approx$ 27 km. It was built to collide protons with protons or lead nuclei and is designed for a maximum centre-of-mass energy of $\sqrt{s}$ = 14 TeV in the proton-proton system. Two separate beam pipes with magnetic fields pointing in opposite directions are needed in order to provide proton-proton collisions. A field strength of $|B| = 8.33\,\mathrm{T}$ is required to bend the high energetic particles onto a circular orbit. These field strengths are reached by the installation of superconducting magnets which operate at temperatures of 1.9 K using liquid Helium. The protons are grouped into bunches containing $1.2 \cdot 10^{11}$ protons each and the LHC is designed to guide 2808 of these bunches through each beam pipe. In March 2010, 7 TeV were reached for the first time and data in 2016 was taken at $\sqrt{s} = 13\,\mathrm{TeV}$.

The LHC makes use of the older accelerating facilities at CERN for pre-acceleration. It is required since the magnet system and high frequency cavities of the LHC are not capable to cover the entire acceleration process. Fig. 3.1 shows a schematic picture of the accelerating chain. The starting point of protons is the linear accelerator LINAC2. After an initial acceleration up to 50 MeV, the protons reach the BOOSTER where they are further accelerated to 1.4 GeV and injected into the Proton Synchrotron (PS). They leave it with an energy of 26 GeV, are injected into the Super Proton Synchrotron (SPS) and finally reach the LHC with an energy of 450 GeV. They are accelerated until they reach the desired energy. There are 4 interaction points where the 2 beam pipes cross each covered by a detector experiment to observe the final states of the collisions. These are ATLAS [6], CMS [61], ALICE [62] and LHCb [22]. The former two are multi-purpose detectors whereas the latter ones are specialised on heavy-ion and b-physics, respectively.

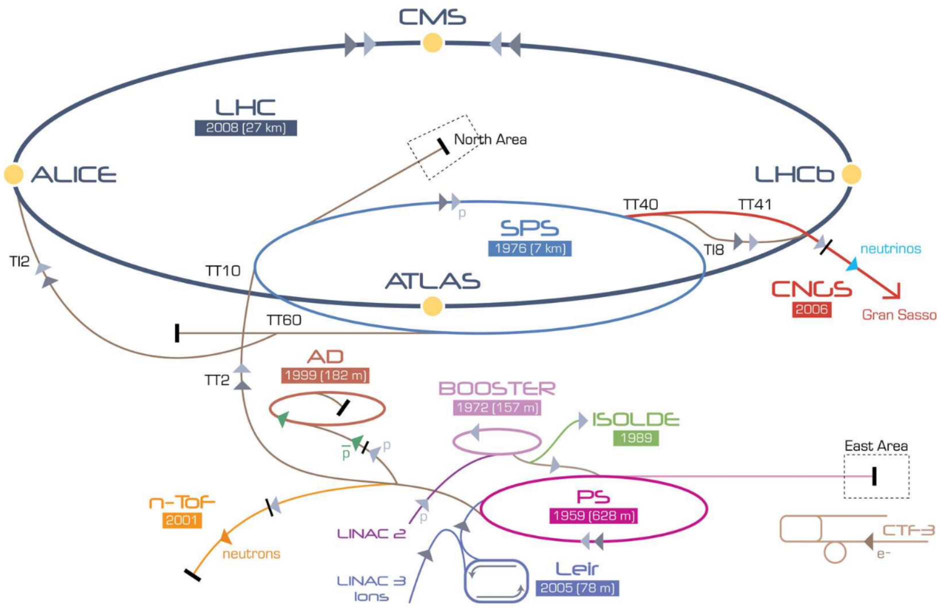


Figure 3.1.: Chain of accelerating facilities starting from LINAC2 and finally reaching the LHC.

One of the most characteristic parameters of the LHC is the *instantaneous luminosity*

$$\mathcal{L} = f \frac{n_b n_1 n_2}{4\pi\sigma_x\sigma_y},$$

where n_b is the number of proton bunches, $n_{1/2}$ are the numbers of protons in two colliding bunches, f is the frequency the bunches travel around the LHC and $\sigma_{x/y}$ are a measure of the lateral widths perpendicular to the velocity. Due to the collisions $n_{1/2}$ and $\sigma_{x/y}$ change in time and hence, $\mathcal{L}$ itself is time dependent. The reason for the importance of the luminosity is that it can be directly related to the theoretically computable quantum physical cross section σ of a certain process $pp \to X$. This is done by measuring the event rate $\mathrm{d}N/\mathrm{d}t$ of interest and using the relation

$$\sigma = \frac{N}{A \int \mathcal{L}\mathrm{d}t}.$$

Cross section measurements come down to counting events of interest whose number N has to be corrected for the *acceptance* and *efficiency* of the respective experiment encoded in the factor A. Fig. 3.2 shows the performance of the LHC in terms of the integrated luminosity $\int \mathcal{L}\mathrm{d}t$. Large luminosities imply larger probabilities of proton-proton interaction. The larger the number of observed events N of a certain process the lower is the statistical uncertainty on the corresponding measurement and the more events of rare processes can be collected.

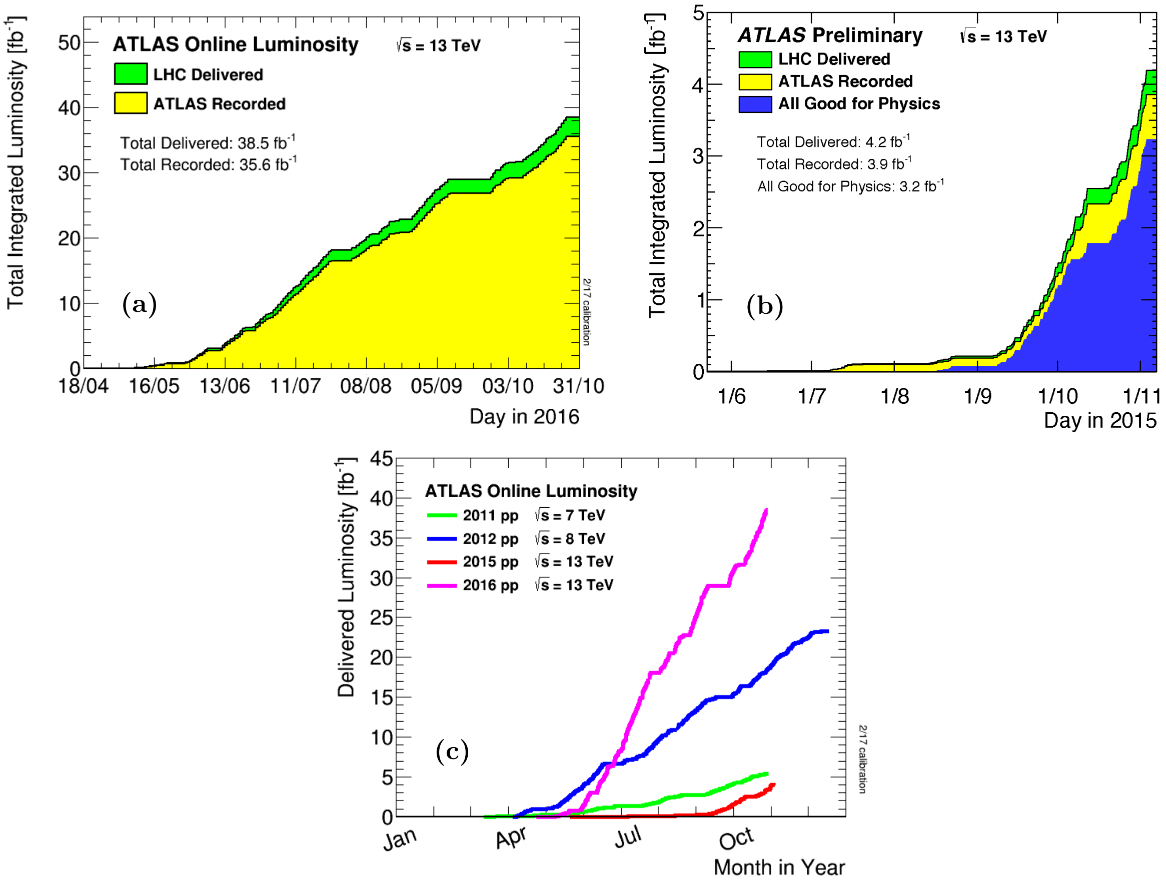


Figure 3.2.: (a) Total integrated luminosity in the year 2016 and (b) in 2015 where the fraction of recorded events good for physics at ATLAS is also shown. The total recorded luminosity increased roughly by a factor of 9 from 2015 to 2016. (c) summarises the integrated luminosities delivered by the LHC in the past operating periods.

An increase of the luminosity was achieved successfully over the last years as seen in Fig. 3.2 (c). In the next years the centre-of-mass energy will be increased up to 14 TeV and from the year 2024 onwards the final upgrades for the so-called High-Luminosity LHC are planned reaching even larger luminosity values.

3.2. The ATLAS experiment

The ATLAS detector [6] is a multi-purpose detector designed to study a huge variety of different processes to precisely test the SM and to look for physics beyond that. Fig. 3.3 shows a sketch of the detector and its main components. It is symmetrically placed around the beam pipes and almost covers the solid angle of 4π to detect as many particles originating from proton-proton collisions as possible.

An understanding of its components is crucial since the detector provides the data which is the basis for all physics analyses. Especially the discussion of the electromagnetic calorimeter system is important since it is used to measure observables which are incorporated in the development of the NN later on.

To identify objects and work with their kinematics or to study the events in detail, a reasonable choice of observables is needed. A reference frame (also called laboratory frame) is established by introducing a right handed coordinate system with its origin at the nominal interaction point. The z-axis points in the direction of the beam pipe, the y-axis points up and the x-axis points to the centre of the LHC ring.

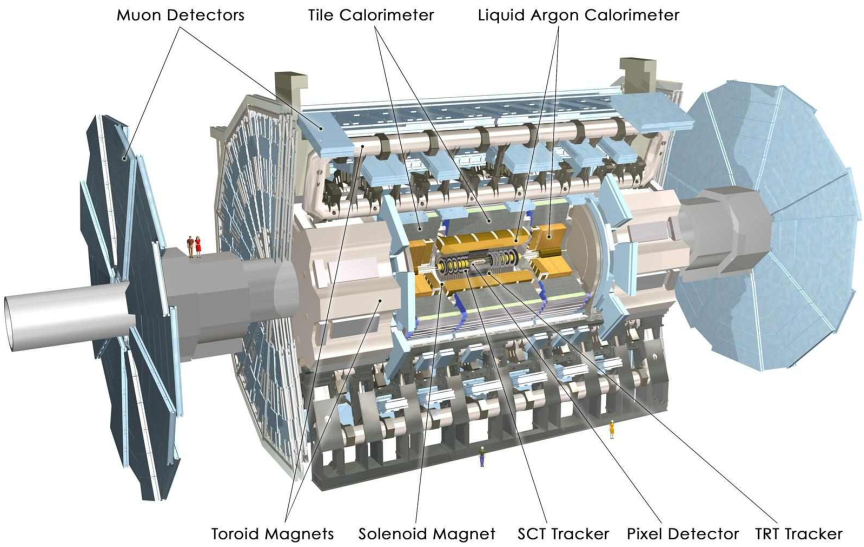


Figure 3.3.: Schematic picture of the ATLAS detector showing the most important substructures.

The transverse momentum of the interacting partons is neglected compared to their longitudinal momentum. Together with the conservation of energy and momentum this assumption implies that the total net transverse momentum of the final state topology vanishes as well. A deviation from that is measured as the missing transverse momentum (MET), $\not{E}_{\mathrm{T}}$, and can be a hint for particles which left the experiment without detection as it happens for neutrinos.

Most often Lorentz invariant observables are constructed since the results should be valid independent of the frame of reference. Different from the transverse momentum the longitudinal momentum of the interacting partons is unknown which potentially introduces a Lorentz boost of the final state along the z-axis in the laboratory frame. Measurements in the transverse plane, like the transverse momentum p_{T}, are not affected by that. For event reconstruction a measurement of angles is also crucial.

In the transverse plane this can be achieved by measuring the azimuthal angle ϕ with respect to a defined reference frame. However, the polar angle θ between the beam axis and an object found in the detector is not invariant under a boost along the z-direction. Therefore, the so called *rapidity*

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

is introduced where E and p_z are the total energy and the longitudinal momentum of the measured object. While this quantity itself is not invariant under a boost in z-direction, the rapidity *difference* Δy of two objects is. With this property a Lorentz invariant distance measure between two objects can be constructed according to

$$\Delta R = \sqrt{(\phi_1 - \phi_2)^2 + (y_1 - y_2)^2}. \tag{3.1}$$

For $E \gg m$, where m denotes the mass of the corresponding object, the rapidity can be approximated by the *pseudo rapidity*

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

Since energies at the LHC normally exceed the masses of the SM particles by orders of magnitude, the measurement of the pseudo rapidity is favoured over the rapidity because it only depends on the polar angle which can be measured directly.

Trigger

The LHC bunch crossing rate is approximately 40 MHz leading to more than 10^9 collisions per second. To handle the large amount of information and store events of interest, a trigger system is necessary. In Run II the ATLAS trigger system [63] consists of the hardware based level 1 (L1) trigger and the software based high level trigger (HLT). Based on energy measurements in small calorimeter cell clusters and on the muon chambers the L1 trigger builds so-called Regions of Interest (RoI). The L1 trigger reduces the rate to approximately 100 kHz. RoIs are passed to the HLT which can take full granularity information of the calorimeter system into account and based on this the rate is further reduced to approximately 1 kHz. The output of the HLT is written to disk and stored for further offline processing and physics analyses.

Inner detector

The ATLAS inner detector (ID) sketched in Fig. 3.4 (a) consists of 3 sub-components which are shown in more detail in Fig. 3.4 (b). It is placed in a 2 T magnetic field with field lines parallel to the beam pipe to bend charged particles on a circular orbit perpendicular to the beam pipe. By measuring their tracks the transverse momentum can be reconstructed with a resolution which gets worse with increasing p_T given as

$$\frac{\sigma(p_\mathrm{T})}{p_\mathrm{T}} \propto p_\mathrm{T}.$$

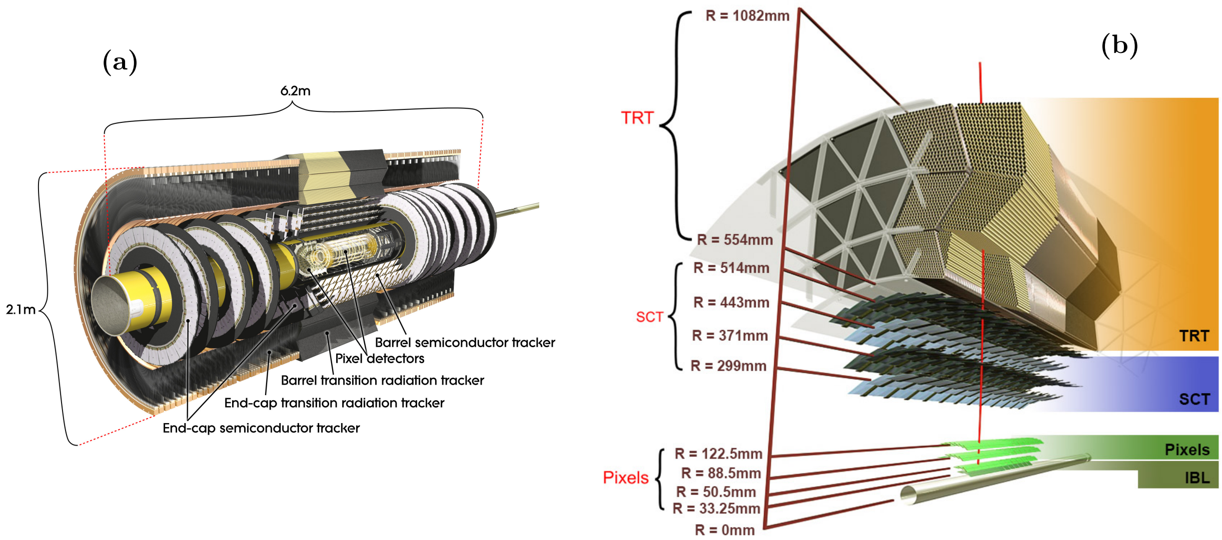


Figure 3.4.: (a) Overall sketch of the ATLAS inner detector showing spatial dimensions and different sub-components. (b) A detailed sketch of the different layers of the inner detector.

The innermost part of the ID is the pixel detector made of 4 layers. One main purpose is the reconstruction of primary and secondary vertices. Directly next to the beam pipe the Insertable B-layer (IBL) [64] is placed. It surrounds the beam pipe at a radius of approximately 3 cm. The dimension of the IBL pixels is $250 \times 50\,\mu\mathrm{m}^2$ and it is surrounded by 3 further pixel layers and there are 3 pixel disks in each of the two end-caps. The pixel size is $50 \times 400\,\mu\mathrm{m}^2$ and in total there are $8 \cdot 10^7$ readout channels.

The next part of the ID is the Semiconducter Tracker (SCT). Silicon strips are distributed over 4 barrel layers and 18 disks in the endcaps. The position of charged particles can be measured with an accuracy of 17 µm per layer perpendicular to the strips. In order to reduce the contribution of ghost hits the strips in one layer are rotated by an angle of 40 mrad with respect to the neighbouring ones.

The outermost part of the ID is the Transition Radiation Tracker (TRT) which consists of straw tubes filled with Xenon and provides approximately $351,000$ readout channels. The TRT can detect tracks up to $|\eta| = 2.0$ with a resolution of $130\,\mu\mathrm{m}$ in the $R-\phi$ plane. Due to their small mass electrons leave a significant amount of transition radiation when traversing the gas filled tubes which can therefore be used for electron identification.

Calorimeter System

A sketch of the ATLAS calorimeter system is shown in Fig. 3.5 (a). It covers the pseudo rapidity region of $|\eta| \leq 4.9$ and is sub-divided into the electromagnetic calorimeter (EMCal) and the hadronic calorimeter (HCal). They are used for energy measurements of the particles originating from the hard interactions of proton-proton collisions.

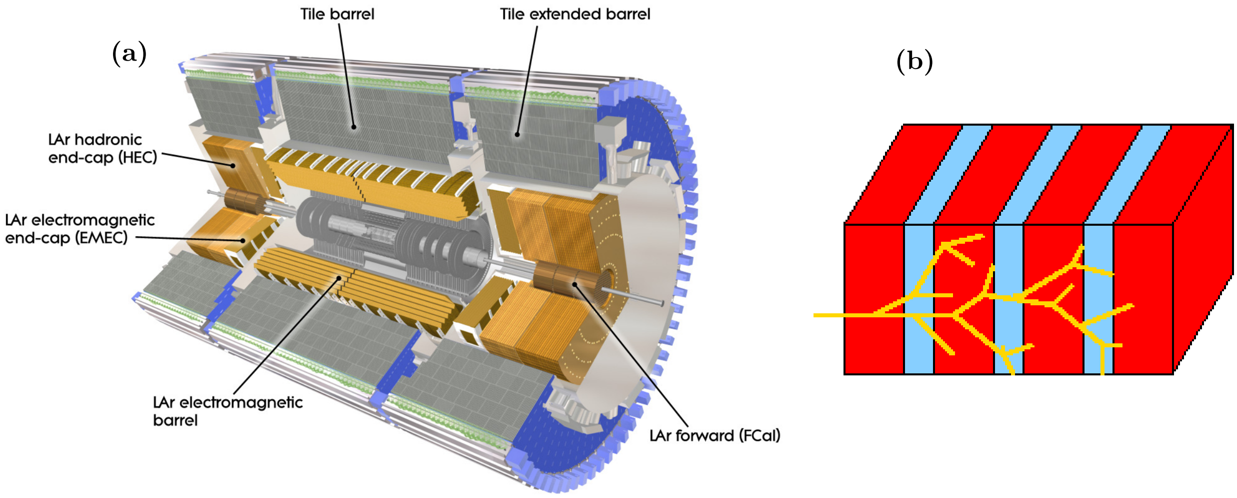


Figure 3.5.: (a) Sketch of the ATLAS calorimeter system showing sub-structures of the EMCal and HCal. (b) Principal configuration of a sampling calorimeter as used by ATLAS with an active and passive material (given in light blue and red).

It is desired that particles leave as much of their energy as possible in the calorimeter system. In that way uncertainties caused by the amount of energy which has not been measured are reduced. Furthermore, it prevents particles other than muons to enter the muon system. To achieve this the system has to consist of a material which has a large interaction probability with traversing particles. Both the EMCal and HCal are so called *sampling calorimeters* where a passive material used for stopping the particles and an active material measuring deposited energies are sampled as depicted in Fig. 3.5 (b).

Its energy resolution is given by

$$\frac{\sigma(E)}{E} = \frac{a}{E} + \frac{b}{\sqrt{E}} + c. \tag{3.2}$$

The first term on the right hand side is the *noise* term parametrising electronic noise of the system. At high energies this term is dominated by the other two terms but, for instance, plays a crucial role when energy measurements in small calorimeter clusters are of interest. The *stochastic* term $\propto 1/\sqrt{E}$ accounts for statistical fluctuations in the energy measurement. Systematic effects caused by dead material or energy leakages dominate at even higher energies and are encoded by the *constant* term c.

The electromagnetic calorimeter

The EMCal consists of a barrel part covering $|\eta| < 1.475$ and two end-cap parts. Each of these parts is divided into an outer wheel ($1.375 < |\eta| < 2.5$) and an inner wheel ($2.5 < |\eta| < 3.2$). Lead and liquid Argon are used as passive and active material, respectively. The *radiation length* X_0 of a material is defined as the length after which the energy of an electron decreases by the factor $(1/e)$ by radiating Bremsstrahlung. This is the dominating energy loss of high energetic electrons. On the other hand, high energetic photons preferentially convert into an e^+e^- pair when interacting with the detector material. Hence, electrons and photons are expected to initiate very similar signatures of *electromagnetic showers*. Since photons have a mean free path of $\approx \frac{9}{7}X_0$ before they undergo pair production, it is sufficient to approximate the thickness of the EMCal in units of X_0. The thickness of the EMCal barrel is $\gtrsim 22\ X_0$ and the thickness of the end-caps is $\gtrsim 24\ X_0$.

Fig. 3.6 (a) shows a partial sketch of the barrel EMCal. It is a three-layer calorimeter with an accordion shape in order to cover the entire ϕ angle without cracks. The innermost layer is the so-called *strip layer* and is finely segmented in η. In the end-caps its granularity gets coarser with increasing $|\eta|$. It can be used to discriminate a single photon from a neutral meson decaying to 2 photons. While both topologies shown in Fig. 3.6 (b) look very similar in the second layer, two maxima are visible in the strip layer for a $\pi^0 \to \gamma\gamma$. Most of the energy of electromagnetic interacting particles is deposited in the second layer. The third layer has a coarse segmentation in η because it is expected to collect only the tail of the electromagnetic shower and is used to estimate possible energy leakages.

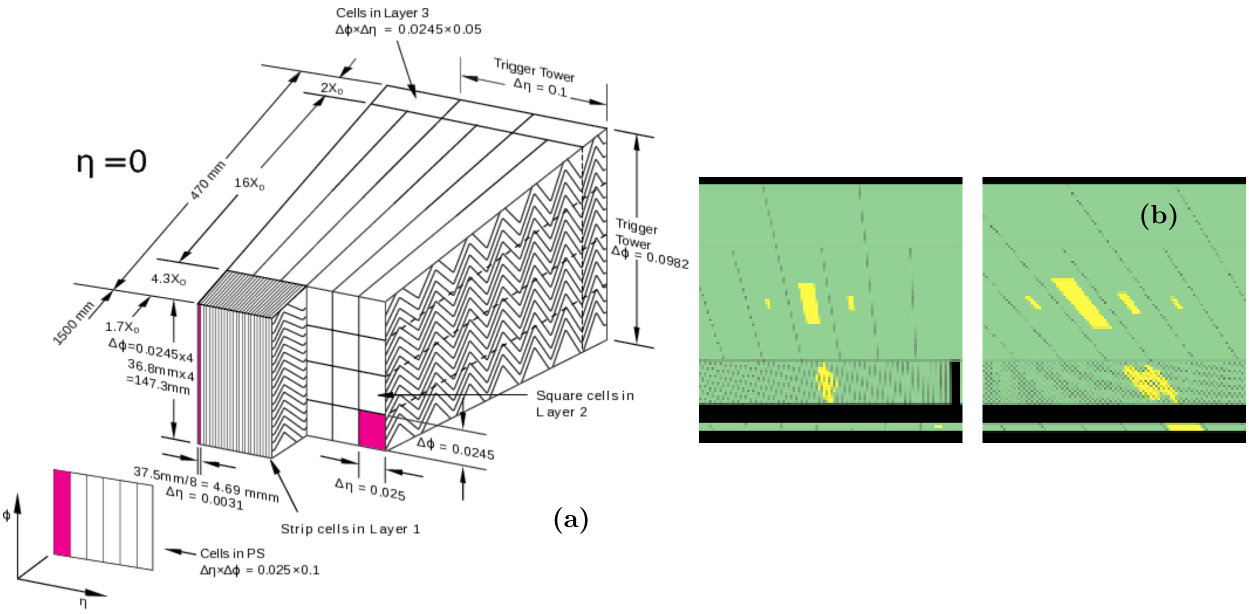


Figure 3.6.: (a) Sketch of a part of the barrel layer of the ATLAS EMCal showing the basic geometry and segmentation. (b) Using the detection of energy maxima in the strip layer can allow for the discrimination of photons originating from a $\pi^0 \to \gamma\gamma$ decay.

The hadronic calorimeter

Strongly interacting particles traverse the EMCal by only depositing a small amount of their energy. The HCal is hence made of material with a high number of nuclei to enhance the interaction probability with strongly interacting particles. It is made of 3 sub-components, the tile, the liquid Argon end-cap and the liquid Argon forward calorimeter. The thickness is measured in units of *interaction lengths* λ given as the depth of material after which the number of hadrons in a particle shower is reduced by the factor $(1/e)$.

The tile calorimeter of the HCal covers the EMCal and is sub-divided into 3 layers. Steel is provided as passive and scintillating tiles as active material. The layers have a thickness of 1.5, 4.1 and 1.8 interaction lengths in the barrel and 1.5, 2.6 and 3.3 in the extended barrel, respectively. Each end-cap of the EMCal is followed by 2 wheels of the liquid Argon end-cap calorimeter which in total cover a pseudo rapidity region up to $|\eta| = 3.2$. They are built using copper as passive and liquid Argon as active material. The liquid Argon forward calorimeter has 3 modules in the end-cap where the first one has Copper and the following two have Tungsten as passive material. It has a thickness of approximately 10 interaction lengths and extends to $|\eta| = 4.9$.

Muon system

The ATLAS detector is covered by the muon system marked as light blue in Fig. 3.3. The reason for this dedicated system is the fact that muons almost always traverse all former detector components. They are approximately 200 times as heavy as electrons which causes on one hand a small curvature in the ID providing a poor p_{T} measurement. That is the reason for a dedicated p_{T} measurement. On the other hand, the probability of emitting Bremsstrahlung scales as $1/m^2$ and muons traverse therefore the EMCal leaving only a small amount of their energy by ionisation. Furthermore, they do not interact strongly and pass the HCal as well. A signal in the muon chambers can hence be used for muon identification. Air-core toroidal magnets provide the magnetic field up to $|\eta| = 1.4$ whereas 2 end-cap magnets are used to bend the muons in the region $1.6 < |\eta| < 2.7$. In the transition region $1.4 < |\eta| < 1.6$ none of the fields can be neglected leading to a complex superposition.

3.3. Photon reconstruction and identification at the ATLAS experiment

In ATLAS, a 2-step procedure is applied to classify candidate topologies as initiated by photons, namely the *reconstruction* followed by the *identification*. Both procedures are briefly explained in the following since the further discrimination of photons originating from the hard interaction is the central problem addressed by this thesis. Detailed information on reconstruction and identification can be found in [65–67].

Photon reconstruction

Photon reconstruction is done in parallel with electron reconstruction because their signatures in the EMCal are expected to be very similar. ID track information and energy clusters in the second layer of the EMCal provide the discriminating observables the reconstruction procedure is based on. Seed clusters of $E_{\mathrm{T}} > 2.5\,\mathrm{GeV}$ built from 3×5 cells in the $\eta - \phi$ plane are searched using a sliding window algorithm [66]. If available, tracks in the ID are matched to the clusters. At the same time potential conversion vertices are searched where a photon has converted to an e^+e^- pair leaving 2 close by tracks in the ID which are also matched to seed clusters. Based on the track matching the algorithm decides whether a cluster is reconstructed as an electron, a photon or as both. Depending on the presence of conversion vertices photons can be reconstructed as being converted or unconverted.

Photon identification

Photon identification aims for an increase of purity of the set of reconstructed photons while keeping the efficiency high at the same time. This procedure is mainly applied to minimise the contamination of hadronic signatures which were incorrectly reconstructed as photons. Rectangular cuts are used on so-called *shower shapes* which are derived from clusters in the EMCal and energy leakages into the HCal. Corresponding cut values are derived by optimisation studies in 7 η regions according to $|\eta| < 0.6$, $0.6 < |\eta| < 0.8$, $0.8 < |\eta| < 1.15$, $1.15 < |\eta| < 1.37$, $1.52 < |\eta| < 1.81$, $1.81 < |\eta| < 2.01$, $2.01 < |\eta| < 2.37$. These regions are chosen to match the topology of the EMCal, especially the strip layer in the end-cap where the granularity gets coarser with each bin beyond $|\eta| = 1.8$ (see [6]). Fig. 3.7 shows a graphical overview of shower shape variables used at ATLAS and they are further explained in Tab. 3.1. All variables but ΔE are *ratios* of energy clusters. This is useful to minimise the influence of systematic uncertainties such as modelling issues and resolution effects.

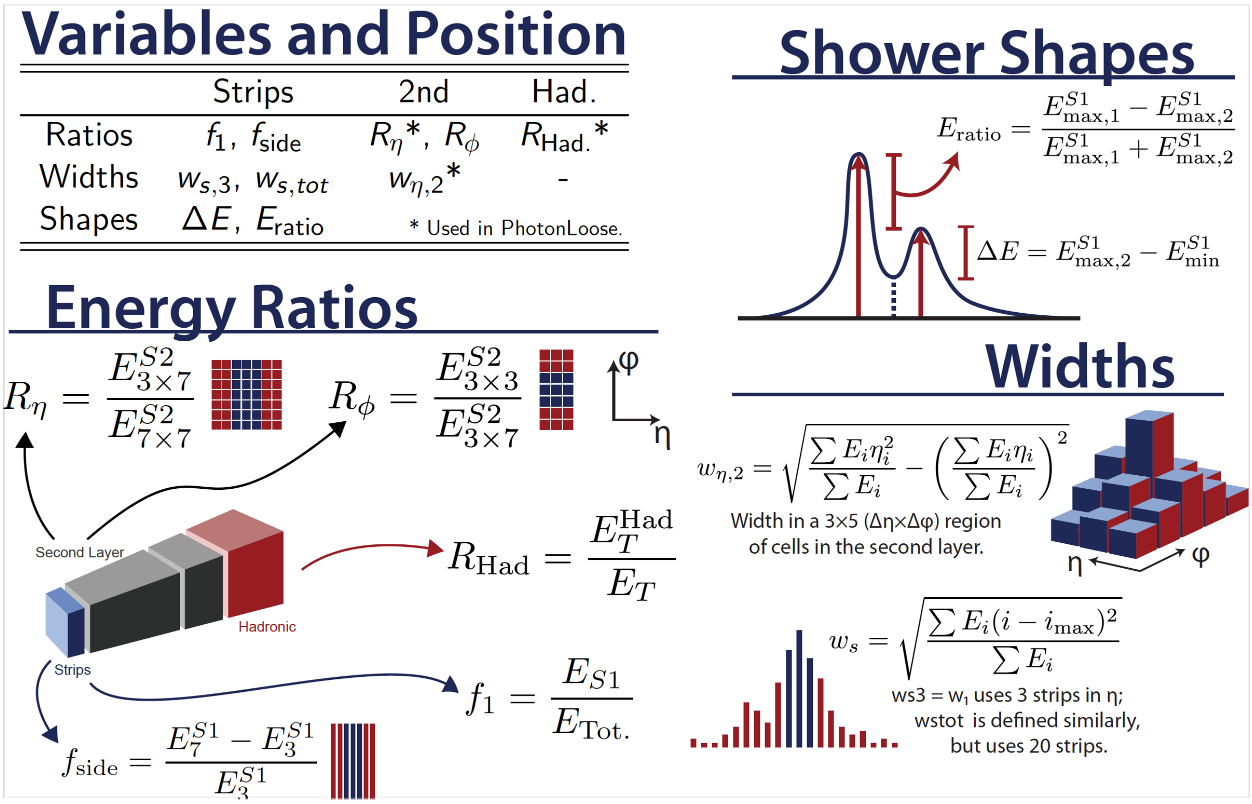


Figure 3.7.: Visualisation of shower shape variables used for photon identification at ATLAS.

There is a so-called `looseID` and a `tightID` menu and due to the method of applying rectangular cuts, the set of `tightID` photons is a sub-set of `looseID` photons. The `looseID` menu invokes shower shapes only obtained in the second layer of the EMCal as well as hadronic leakage, namely R_η, $w_{\eta,2}$ and R_{had}. Further cuts are applied in the `tightID` menu, especially those obtained from measurements in the strip layer. The identification efficiencies of `tightID` photons with $E_\mathrm{T} \approx 10\,\mathrm{GeV}$ are measured to be $50 - 65$ % ($45 - 55$ %) for unconverted (converted) photons and increases to $94 - 100$ % for photon energies of $E_\mathrm{T} > 100\,\mathrm{GeV}$. At $E_\mathrm{T} > 40\,\mathrm{GeV}$ the efficiency is larger than 90 % [65].

name	description
Hadronic leakage	
R_{had}	Transverse energy leakage in the HCal normalised to transverse energy of the photon candidate in the EMCal. In the region $0.8 \leq \lvert\eta\rvert \leq 1.37$ the entire energy of the photon candidate in the HCal is used and in the region $\lvert\eta\rvert < 0.8$ and $\lvert\eta\rvert > 1.37$ the energy of the first layer of the HCal is used
Energy ratios and width in the second layer of EMCal	
R_η	Energy ratio of 3×7 to 7×7 cells in the $\eta \times \phi$ plane.
R_ϕ	Energy ratio of 3×3 to 3×7 cells in the $\eta \times \phi$ plane.
$w_{\eta 2}$	Lateral width of cluster in $\eta \times \phi = 3 \times 5$: $\sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left(\frac{\sum_i E_i \eta_i}{\sum_i E_i}\right)^2}$
Energy ratios and widths in the first (strip) layer of EMCal	
$w_{\eta 1}(w_{s3})$	Energy weighted width in units of the number of strips using 3 strips around the maximum: $\sqrt{\frac{\sum_i E_i (i - i_{\mathrm{max}})^2}{\sum_i E_i}}$
$w_{\mathrm{tot,s1}}(w_s)$	Energy weighted width using 20 strips around the maximum, see $w_{\eta,1}$.
f_{side}	Energy within 7 strips without 3 central strips normalised to energy in 3 central strips.
E_{ratio}	Ratio between difference of first 2 energy maxima divided by their sum ($E_{\mathrm{ratio}} = 1$ if there is no second maximum).
ΔE	Difference between the second energy maximum and the minimum between first and second maximum ($\Delta E = 1$ if there is no second maximum).

Table 3.1.: Summary of shower shape variables used for cut-based photon identification at Atlas.

3.4. Prompt photons and hadron fakes

Because of the electromagnetic interactions of charged hadrons, electrons and photons, they can all leave similar signatures in the EMCal. The former two contributions can therefore fake photons if their signatures are incorrectly identified. The set of *hadron fakes* contains hadrons misidentified as photons as well as real photons originating from hadronic decays, for instance in a $\pi^0 \to \gamma\gamma$ decay. The reason why the latter are counted as hadron fakes is that these photons are decorrelated from the process of the hard parton-parton interaction and provide no information of the hard scattering amplitude[1]. Photons originating from the hard interaction are called *prompt photons*. As explained in Secs. 2.2.4 and 2.2.5 the estimation of the hadron fake contribution is crucial when studying for instance the $pp \to t\bar{t}\gamma$ process. A first discrimination between prompt photons and hadron fakes is already done in photon identification and especially shower shapes obtained in the strip layer provide strong separation. Further discrimination yielding a high hadron fake rejection and prompt photon selection efficiency can be desired, especially for precision measurements.

In simulation prompt photons and hadron fakes can be disentangled using truth information from the simulations providing the particle's ID in the Monte Carlo particle numbering scheme [11]. The ATLAS `MCTruthClassifier` scheme is briefly explained in App. A.7 for completeness.

[1] Measurements taking photons from hadron decays into account can be reasonable for instance to tune the modelling of hadrons and fragmentation in simulation.

4. Brief introduction to machine learning

Different machine learning algorithms are currently used in ATLAS. The most common ones among them are Boosted Decision Trees (BDTs) and neural networks (NNs). In the following, basic concepts of machine learning are introduced which will be needed for the development of a NN in Secs. 6.2 and 6.3.

First attempts in machine learning were made by introducing *linear* models by McCulloch and Pitts in 1948 [68] with an underlying function

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \cdot \mathbf{x} = \sum_i w_i x_i.$$

Each entry x_i of the vector $\mathbf{x}$ is called *feature* and each vector summarises the features of one *sample*. Feature i is weighted a factor by w_i which are parameters of the model and the function $f(\mathbf{x})$ assigns one value to each sample. A certain partition of the samples can be achieved, for instance, by grouping samples with similar function values. If a specific partition was desired all weights had to be set by hand accordingly. In 1958, Rosenblatt [69] came up with the first *perceptron* which was able to *learn* the values of the weights by presenting samples with known labels to the algorithm. However, Minsky and Papert [70] have shown that linear models could not learn the basic logic `exclusive OR` function which was understood as a major drawback. In 1986 Rumelhart et al. introduced the *backpropagation* algorithm [71] which solved that problem. In the following years many further developments were made and there are many datasets produced and collected which often serve as benchmarks when comparing the performance of different machine learning algorithms. One of these datasets is the MNIST dataset containing a large number of hand-written digits. A variety of algorithms was trained to recognise these digits and the performances of many of those is summarised in [72].

4.1. Basic building blocks

There are different purposes machine learning algorithms are aiming for. It can be used to learn the form of a function certain samples belong to which is called *regression*. Another common problem is that of *classification*. Discriminating between only 2 types, a *signal* and a *background*, is then a *binary* classification problem and the discrimination between prompt photons and hadron fakes is of this type. Everything in the following applies hence to binary classification problems[1].

There are 2 approaches of learning: In the case of *unsupervised* learning the algorithm will come up with a labelling and classification by itself. *Supervised learning*, as it will be used in this thesis, requires the knowledge of the true labels $\hat{\mathbf{y}}$ before learning also called *targets*. Generally, a machine learning algorithm with parameters w_j can be thought of as a function

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ \mathbf{x} &\longmapsto f(\mathbf{x}) = \mathbf{y}, \end{aligned}$$

which assigns a label $\mathbf{y}$ to a sample $\mathbf{x}$. In the case of a binary classification problem $m = 1$ is sufficient. Samples labelled with a value above the so-called *working point* are considered to belong to one class and everything below to the complementary class. During the *learning* or *training* phase, the parameters w_j are adjusted such that for each sample the assigned label $\mathbf{y}_i = f(\mathbf{x}_i)$ is as close to the true label $\hat{\mathbf{y}}_i$ as possible. A *measure* is required in order to quantify how well the classifier is performing. This measure is called *loss function* and is given as a scalar function $C\left(\{f(\mathbf{x}_i)\}, \{\hat{\mathbf{y}}_i\}\right)$. The terms *learning* or *training* then refer to the minimisation of the loss function by finding parameters w_{j0} satisfying

$$\nabla_w \left. C\right|_{\mathbf{w}_0} = 0.$$

Due to the potentially large number of free parameters and samples an analytic solution of this equation is not always possible. So-called *optimisers* provide a prescription of how the minimum is searched for numerically and how it is approached in small steps. This incorporates the calculation of the loss function's gradient at some point. For that reason algorithms such as backpropagation [71] were developed which are capable of computing the gradient numerically.

[1] This does not mean that explanations made here are always different for regression models. Some of them are also true for those and other machine learning algorithms.

4.1.1. Training and testing of a binary classifier

To check for the classification performance and whether a trained classifier can generalise to unseen data, it has to be *tested*. This can be done by drawing independent training and test sets. An evaluation of the classifier can be achieved by comparing its performance on the training as well as on the test samples. Different performance measures and techniques are introduced in the following.

Classification performance measures

Both the classification performance as well as the generalisation performance of a binary classifier can be calculated using the so-called *Receiver Operating Characteristics* (ROC) which yields the *background rejection* as a function of the *signal efficiency*. The background rejection is given by the fraction of background samples below the chosen working point and the fraction of signal samples above the working point yields signal efficiency. The ROC can be calculated separately for the training and the test and is visualised in Fig. 4.1. The agreement of the ROC curves indicates the generalisation capability to unseen data of the classifier given that the features distributions in both the training and test sets agree. A measure for the classification performance is the *Area Under Curve* (AUC). The perfect classifier has an AUC of 1 which is equivalent to classifying all samples correctly leading to a rectangular ROC curve with a signal efficiency and background rejection of 1. Such a high performance can only be reached asymptotically. If the AUC is 0.5 or lower, the classification performance is not better than that of tossing a coin.

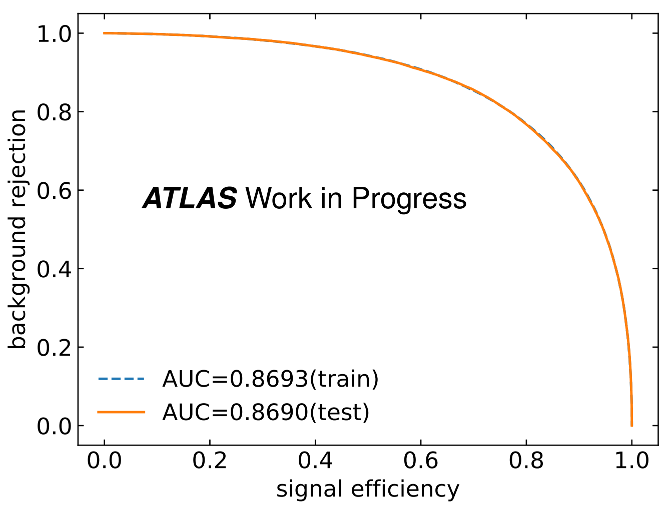


Figure 4.1.: Example of ROC curves of the training and test samples showing a good agreement.

Training on mini-batches and validation set

To save computation time the optimisation and backpropagation is often not applied on the entire training set. Training is then done on so-called *mini-batches* the training set is sub-divided into. In this way the gradient of the loss function is computed taking only a fraction of the samples into account. This is done for each mini-batch and when all of them are processed, an *epoch* is accomplished. A machine learning algorithm can be trained over many epochs to decrease the loss as much as possible.

To guarantee the generalisation capability already *while* the classifier is trained, a *validation set* can be used. It is an orthogonal set randomly drawn from the training samples and it is not used included in the calculation of loss minimisation. After each epoch the *validation loss* on the unseen validation set is calculated in addition to the training loss. A decreasing validation loss is hence a more reliable performance measure since the training loss decreases by construction while the former one does not necessarily.

Under- and overtraining and generalisation performance of a binary classifier

If the training phase is stopped far before approaching the minimum of the loss function, the classifier is said to be *undertrained* and the number of incorrectly classified samples is expected to be large. On the other hand, training over too many epochs can lead to so-called *overtraining*. In this scenario a classifier takes also statistical fluctuations (points in phase space with a small probability) in the training set into account. The test and validation sets are expected to contain samples with different fluctuations in other regions of phase space and hence, the generalisation capability of an overtrained classifier decreases. A hint for overtraining can be a decreasing loss along with an increasing validation loss during training. To check for overtraining completely independently of the training procedure, the ROC curves of training and test samples can be compared as described above. Significant deviations between the ROC curves are expected for an overtrained classifier.

Cross validation

To further investigate feature fluctuations in the training and test set and to check if the classifier is reproducible having a stable performance, the *n-fold cross-validation* can be applied. The entire data is equally split into n orthogonal sub-sets. Each sub-set i is once used as the test set after the classifier has been trained on the remaining $n-1$ sets. This yields n pairs of ROC curves. If the deviations between all resulting test ROC curves are large, it can be concluded that the training highly depends on the sample composition used in training and that the resulting classifier is hardly reproducible.

4.2. Neural networks

This Chapter is finalised by a brief summary of concepts dedicated to NNs which will also be needed for the development in the following. A sketch of the generic structure is shown in Fig. 4.2. The basic building blocks are *neurons* indicated by circles which are arranged in *layers*. In its simplest form each neuron in a layer is connected to all other neurons in the previous and following layers leading to a *feed-forward* NN.

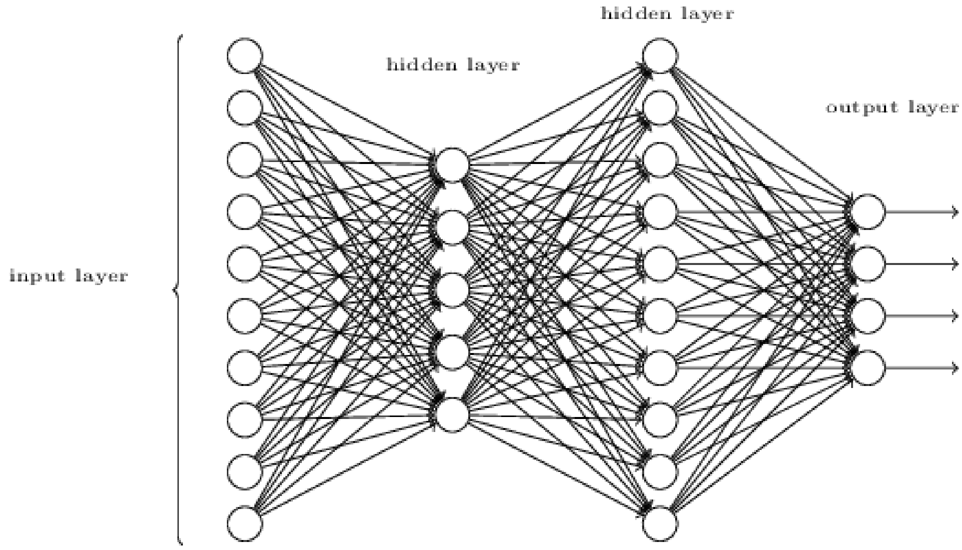


Figure 4.2.: Generic form of a NN with 2 hidden layers and an output layer containing 4 neurons.

As already mentioned, the first type of neurons was introduced by Rosenblatt [69] and called *perceptron*. Those come along with an *activation function*

$$\sigma(x) = \begin{cases} 0, & \text{if } x + b \leq 0 \\ 1, & \text{if } x + b > 0 \end{cases} \tag{4.1}$$

where the parameter b is called *bias* and is a property of the neuron. It can be thought of as a threshold which has to be overcome by the input in order to yield an output which is 1 in this case. An importance measure of single inputs j to a neuron is again provided by weights w_j.

Small changes of parameters in the NN should ideally cause small changes of the output to smoothly decrease the loss and avoid discontinuities of the output values during training. However, the output of a perceptron flips from 0 to 1 which is hence not a continuous function of the input. Commonly used continuous activation functions are the *sigmoid*, the *rectifier*, also called *relu* in the following, or the *softmax* activation functions defined as

$$\begin{aligned}\sigma_{\mathrm{sig}}(z) &= \frac{1}{1+e^{-z}},\\ \sigma_{\mathrm{relu}}(z) &= \max(0, z),\\ \sigma_{\mathrm{soft}}(\mathbf{z})_i &= \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}.\end{aligned} \tag{4.2}$$

The latter one is special in the sense that its output depends on all N neurons in a layer and not only on the inputs. Additional types of layers are also available. One of them is the so-called *batch normalisation* layer [73]. It is trained to scale the output of each neuron in a layer such that the distribution of all outputs has a distribution with mean 0 and standard deviation of 1. This is done to keep the parameters of the NN at values of similar orders of magnitude.

A commonly used loss function for binary classification problems is the *binary cross entropy* defined as

$$C_{\mathrm{BCE}}(y, \hat{y}) = -\sum_{i\in\mathrm{samples}} \left[\hat{y}_i \log y_i + (1-\hat{y}_i)\log(1-y_i)\right], \tag{4.3}$$

where $\hat{y}_i$ and y_i are the true and predicted label of sample i, respectively. In the implementation in this this thesis it is $\hat{y}_i \in \{0, 1\}$.

The concepts explained and summarised in this Chapter will be used later for the development of the NN for prompt photon and hadron fake classification. All activation functions in Eq. (4.2) are going to be implemented and the ROC and AUC are considered as the main performance metrics in the NN development.

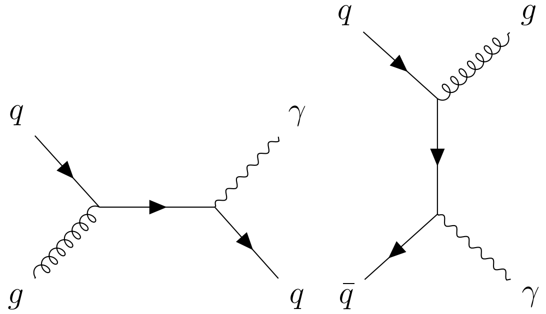
5. Simulated samples, object definitions and event selection

5.1. Monte Carlo samples

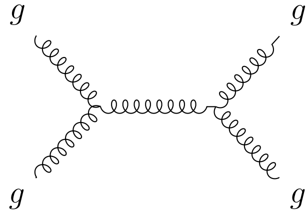
All Monte Carlo (MC) samples contain simulated events of proton-proton collisions at a centre-of-mass energy of 13 TeV.

Prompt photons and hadron fakes for `NN` training are selected in single photon and di-jet production. Leading order Feynman diagrams are shown in Fig. 5.1. Events of both processes are generated using PYTHIA8 [74] for the generation of the matrix element as well as for the parton showering using ATLAS A14 tune [75] and NNPDF2.3LO parton distribution function (PDF) [76]. A sample list can be found in Tab. A.2 in the Appendix.

The $t\bar{t}\gamma$ signal sample is produced with MADGRAPH_aMC@NLO generator [77] using the NNPDF2.3LO PDF set where the parton shower is simulated using PYTHIA8 with ATLAS tune A14. Inclusive $t\bar{t}$ events are simulated with POWHEG-BOX version 2 [78]. The parton shower, PDF sets and tune configurations are the same as in the simulation of $t\bar{t}\gamma$ samples. Leading order Feynman diagrams of $t\bar{t}\gamma$ production in gluon-gluon fusion can be seen in Fig. 2.7. $t\bar{t}$ final state topologies are discussed in Sec. 2.2.2.



(a) Prompt photon production via s- and t-channel.



(b) Hadron fake production in s-channel di-jet production via gluon-gluon fusion.

Figure 5.1.: LO Feynman diagrams showing leading contributions to single prompt photon and di-jet production in proton-proton collisions.

Further $t\bar{t}\gamma$ background processes are the production of a W or Z boson in association with a prompt photon or an additional jet. These events are simulated with SHERPA version 2.2.2 and 2.2.1 [79], respectively, both with PDF set NNPDF30NNLO.

Diboson production is another crucial background for $t\bar{t}$ and $t\bar{t}\gamma$ analyses. Up to the missing b-quarks the final states of WW, WZ and ZZ production look very similar to the signal process containing jets and charged leptons. Events of these processes are simulated using SHERPA 2.1 with PDF set CT10(NLO) [80]. Single top-quark production and Wt channel production are generated with POWHEG-BOX version 1 with PDF set CT10(NLO) interfaced with PYTHIA6 using Perugia2012 tune [81] for parton shower simulation. Additional photons are simulated with the PHOTOS package [82].

5.2. Object definitions and event selection

Photons

Reconstructed photon candidates are required to have a $p_{\mathrm{T}} > 20$ GeV in the $t\bar{t}\gamma$ samples and $p_{\mathrm{T}} > 25$ GeV in the case of single photon and di-jet samples. They have to pass a pseudo-rapidity cut of $|\eta| < 2.37$ excluding the crack region between $1.37 < |\eta| < 1.52$. Both converted and unconverted photon candidates passing `tightID` selection criteria are taken into account. In simulation prompt photons and hadron fakes are disentangled using MC truth information as explained in Sec. 3.4.

Electrons

The calibrated transverse energy [83] of electron candidates has to be above 25 GeV and within $|\eta| < 2.47$ excluding the region of $1.37 < |\eta| < 1.52$. `TightLH` identification [84] and `Gradient` isolation [85] are required. In the 2016 data taking period the cut on the transverse energy is raised to 27 GeV.

Muons

Muon candidates have to pass `Medium` identification [86] and `Gradient` isolation criteria [85] and have to fulfil $p_{\mathrm{T}} > 25$ GeV as well as $|\eta| < 2.5$. In 2016 data $p_{\mathrm{T}} > 27.5$ GeV is required.

Jets

Jet candidates in $t\bar{t}\gamma$ samples are reconstructed using the anti-k_{T} algorithm [37] with radius parameter $R = 0.4$ (see also Eq. (3.1)). A transverse momentum of $p_{\mathrm{T}} > 25$ GeV is required within the pseudo-rapidity range of $|\eta| < 2.5$. The Jet Vertex Tagger (JVT) discriminant [87] for jet candidates with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ is required to be smaller than 0.59 in order to reduce the contamination of jets from pile-up events.

b-jets
b-jet candidates are tagged using the `MV2c10` algorithm [88] at the working point of 77 % efficiency in $t\bar{t}$ events. The discriminating algorithm is based on a `BDT` where cutting on a certain value corresponds to different working points and hence efficiencies. It combines information from track properties as well as properties of secondary vertices if reconstructed.
Missing transverse momentum
Missing transverse energy $\not{E}_T$ in ATLAS [89] is mainly caused by the inability of detecting neutrinos[1]. Hence, their transverse momenta and energy components are not measured[2] and the total net transverse components in the final state are different from 0 as it is ideally the case before the proton-proton collision.

General event selection in $t\bar{t}\gamma$ signal and background samples

Any selection in the $t\bar{t}\gamma$ signal sample and the respective background samples has to be triggered by at least one of the single lepton triggers summarised in Tab. 5.1 where the triggering lepton must correspond to a selected lepton candidate in the event. Different triggers for 2015 and 2016 data taking are applied.

year	**electron trigger**	**muon trigger**
2015	`HLT_e24_lhmedium_L1EM20VH` or `HLT_e60_lhmedium` or `HLT_e120_lhloose`	`HLT_mu20_iloose_L1MU15` or `HLT_mu50`
2016	`HLT_e26_lhtight_nod0_ivarloose` or `HLT_e60_lhmedium_nod0` or `HLT_e140_lhloose_nod0`	`HLT_mu26_ivarmedium` or `HLT_mu50`

Table 5.1.: Electron and muon triggers in the event selections.

The event selection in the *single lepton* channel requires either exactly 1 electron ($e+$jets channel) or muon ($\mu+$jets channel) accompanied by at least 4 jets. At least one jet is required to be a b-jet and exactly 1 photon must be present. In the $e+$jets channel the invariant mass of the photon and electron candidate has to be outside the mass window of [85, 95] GeV to suppress electrons originating from a Z boson which are then incorrectly identified as photons.

[1] This is the case if only SM physics is assumed. If additional BSM physics is assumed, missing transverse energy can also be the result of particles and interactions in that BSM model.

[2] Their longitudinal components are also not measured. In general the longitudinal boost of the parton-parton system of a proton-proton collision is not known and does hence apply not only to neutrinos but to the entire event.

The *dilepton* channel selection criteria require exactly 2 oppositely charged leptons of the same flavour where electrons (ee channel) and muons ($\mu\mu$ channel) are considered. The invariant mass of the lepton pair has to be larger than 15 GeV and outside the mass window of [85, 96] GeV in order to suppress background events where a Z boson decays leptonically. For the same reason events are rejected where the invariant mass of the lepton pair together with the photon is in the mass window of [85, 95] GeV. Since 2 neutrinos are assumed to be present due to the leptonically decaying $W^{\pm}$ bosons, $\not{E}_T > 30$ GeV is required. At least 2 jets are required which ideally originate from the $t \to Wb$ decay and hence, at least one must be tagged as a b-jet. Finally, events have to contain exactly 1 photon.

Event selection in single photon and di-jet samples

Events have to pass at least one of the triggers listed in Tab. 5.2, good quality criteria in the detector are required and at least one `tightID` photon has to be present. No further event selection cuts are applied.

photon trigger	
HLT_g10_loose	HLT_g15_loose_L1EM7
HLT_g20_loose_L1EM12	HLT_g25_loose_L1EM15
HLT_g35_loose_L1EM15	HLT_g40_loose_L1EM15
HLT_g45_loose_L1EM15	HLT_g50_loose_L1EM15
HLT_g60_loose	HLT_g70_loose
HLT_g80_loose	HLT_g100_loose
HLT_g120_loose	HLT_g140_loose

Table 5.2.: List of photon triggers to be passed for event selection in `NN` training.

6. Prompt photon discrimination in ATLAS using a neural network

6.1. Studies of photon shower shapes

The following Section presents studies of shower shape variables summarised in Tab. 3.1 in different event topologies as well as for different kinematic regions of `tightID` photons. This is done since analyses can be different ranging from the final state topology, which is selected applying certain cuts, to different object definition criteria. The impact of these differences on photon shower shapes is studied focusing on prompt photons. The differences with respect to hadron fakes are then covered by studying the separation between the distributions of prompt photons and hadron fakes[1].

6.1.1. Shower shapes in different event topologies

Since photon shower shapes are computed based on energy measurements in cells of the EMCal (and also in the HCal in case of R_{had}), it is crucial to know to what extent additional activity influences these shapes. This is done by investigating shower shapes of prompt photons in the $t\bar{t}\gamma$ signal sample in 5 bins of different jet multiplicity[2]. Events with either exactly 2 or 3 jets are selected in the dilepton channel and events containing either exactly 4, 5 or ≥ 6 jets are selected from the single lepton channel. Since this is a pure MC study, all requirements on invariant masses are dropped as well as b-jet requirements. The remaining event selection cuts are summarised in Tab. 6.1.

Fig. 6.1 shows the shower shape distributions of R_η and f_{side} of prompt photons[3]. All distributions of different jet multiplicities agree well within their statistical uncertainties.

[1] This is not meant to be an exhaustive and quantitative study of photon shower shapes covering all properties in different kinematic regions. It is rather aimed for a qualitative understanding which can be used to develop and discuss a `NN` using these as features as explained later.

[2] It has to be noted that the given study is not meant to prove that shower shape variables are entirely independent of the event topology. A thorough study of shower shapes in different event topologies is beyond the scope of this thesis.

[3] Further distributions can be found in App. A.4.

channel	single lepton (e+jets, μ+jets combined)	dilepton (ee and $\mu\mu$ combined)
common	$== 1\, e$ `XOR` $== 1\, \mu$	$==$ 1 lepton pair: e^+e^- `XOR` $\mu^+\mu^-$
	trigger match	
photon	$== 1$	
jet	≥ 4	≥ 2

Table 6.1.: Event selection cuts in single lepton and dilepton channel to study the impact of jet multiplicity on photon shower shapes.

Around the maxima the distributions deviate in some cases slightly more than $1\,\sigma$ from one another. Since this appears to be the case for all shower shapes, it can be concluded that there are small systematic differences between event topologies. No further shape deviation can be observed.

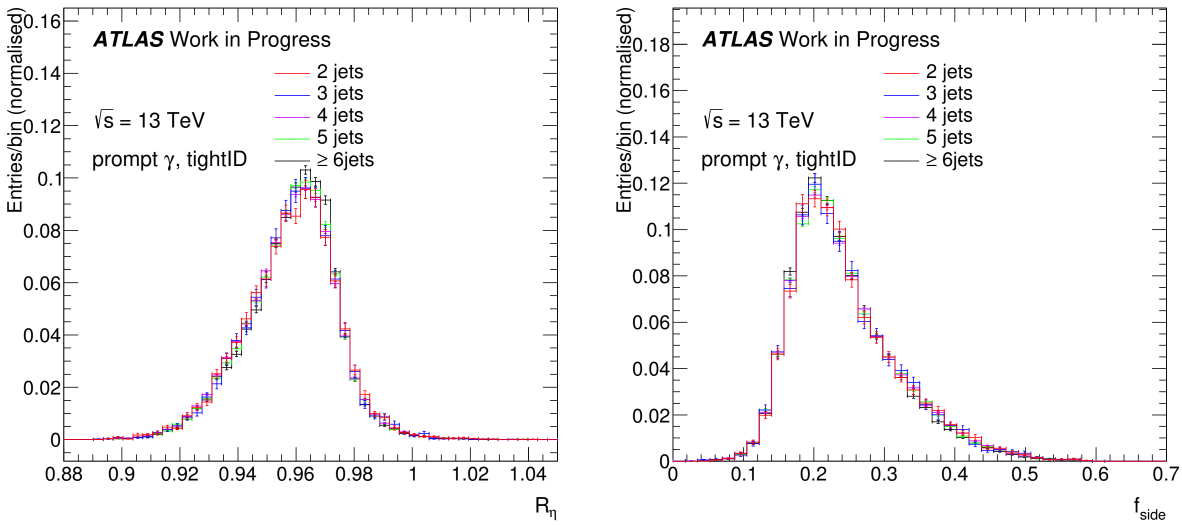


Figure 6.1.: Shower shape distributions of prompt photons in 5 different bins of jet multiplicity.

6.1.2. Shower shapes in different p_T and η bins

The topology of the ATLAS detector changes with pseudo rapidity η and the interaction between the photon and the material is in general energy dependent. Photon shower shapes can therefore be expected to depend on p_T and η. Because the number of photon candidates is significantly larger in single photon and di-jet MC samples, those are used for this study. Photons are selected in 3 bins of p_T, namely [25, 50) GeV, [50, 100) GeV and [100, ∞) GeV. The η binning is motivated by the topology of the EMCal. The first two bins of $|\eta| \in [0, 0.6)$ and $|\eta| \in [0.6, 1.37)$ are covered by the barrel part whereas the last bin with $|\eta| \in [1.52, 2.37)$ reflects the end-cap. The η region of (1.37, 1.52), where parts of the barrel and end-cap overlap and which contains a large amount of non-active material upstream, is excluded. Material upstream between the barrel and the end-cap as well as coarser granularities in larger η-regions (see also discussion in Sec. 3.2) in the strip layer are expected to have an impact on the shower development and its measurements.

The leakage of the transverse energy into the HCal is measured by R_had. The strongest peak can therefore be observed for high p_T prompt photons at values slightly above 0 as shown in Fig. 6.2. Prompt photons with lower p_T show hence more prominent tails relative to the maximum which decreases with p_T and approaches 0. Negative values are caused by subtraction of electronic noise in the calorimeter cells leading to a significant effect given small energy leakages (see also Eq. (3.2)). That also causes the distribution to become more symmetric with decreasing p_T. In the large η region the distribution gets significantly broader and more symmetric. The reason is that the fraction of the transverse energy with respect to the total energy decreases with increasing η.

R_η and R_ϕ measure energy fractions in the $\eta - \phi$ plane in the second layer of the EMCal. Distributions in different kinematic regions can be found in Fig. 6.2. For both observables the values above 1 are again caused by electronic noise[4]. The p_T dependence is stronger for R_η since this variable measures the leakage in the η direction. The relative size of the longitudinal energy component increases with decreasing p_T given two photons of the same energy and hence a larger leakage into neighbouring cells can be expected. The same argument applies to the distributions in different η bins. R_ϕ measures the leakage in the ϕ direction. A smaller boost in the $r - \phi$ plane leads to broader electromagnetic shower with decreasing p_T. Smaller R_ϕ values appearing for increasing η are mainly caused by a wider development of the electromagnetic shower due to the material upstream between the barrel and end-cap.

[4] In that case almost the entire energy is deposited in the smaller cell cluster and noise subtraction yields a slightly smaller energy determined in the larger cell cluster.

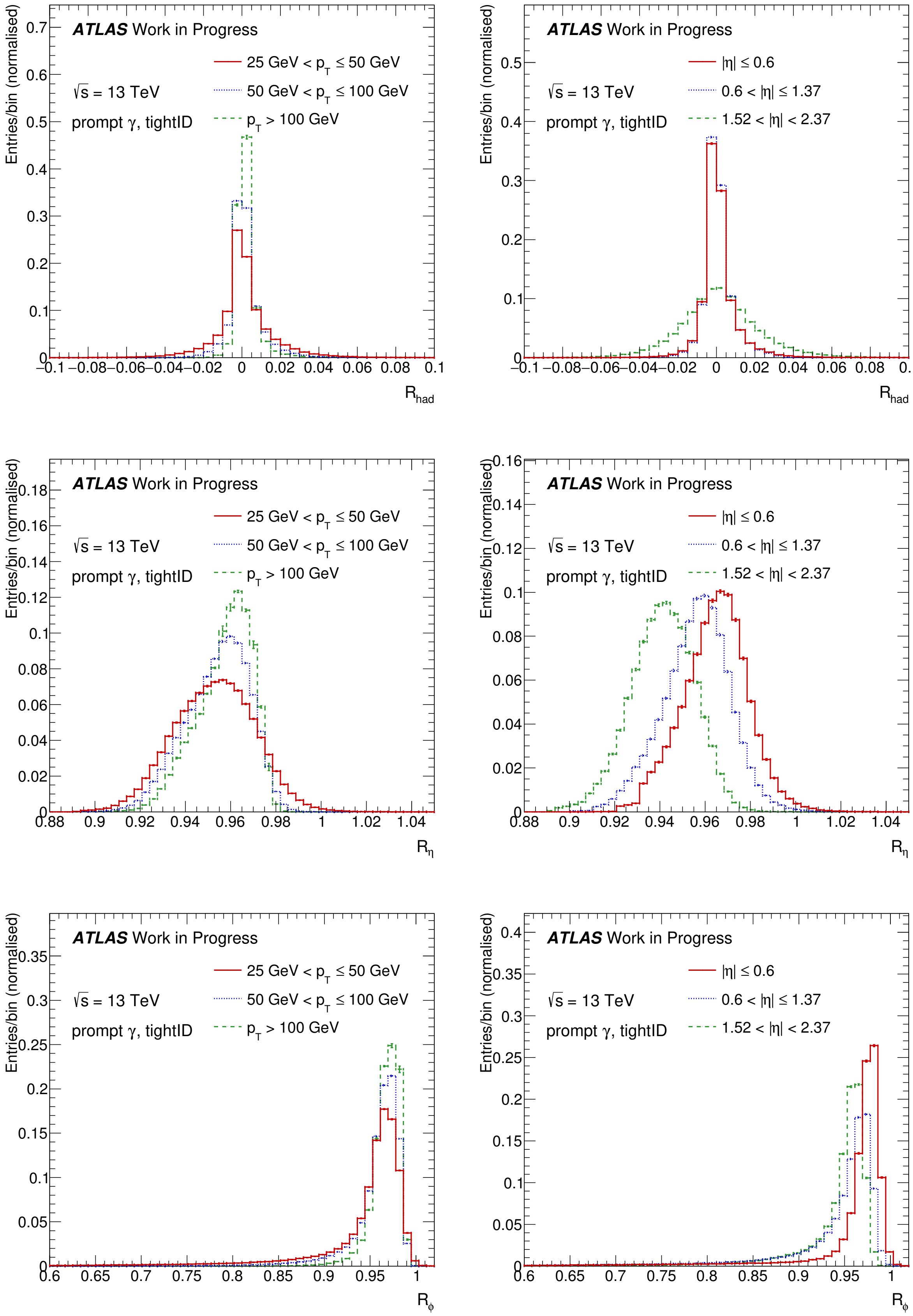


Figure 6.2.: Distributions of R_{had}, R_η and R_ϕ in 3 bins of p_{T} and η.

The variable $w_{\eta 2}$ measures the energy weighted shower width in the η direction in the second layer of the EMCal. That is expected to decrease with increasing p_{T} as seen in Fig. 6.3 for the same reasons R_{η} andR_{ϕ} decrease. The impact of the shower development in the transition region between the barrel and end-cap of the EMCal can be seen which shifts the maximum of the distribution to even larger values in the last η bin leaving only a smaller shoulder on the left of the maximum.

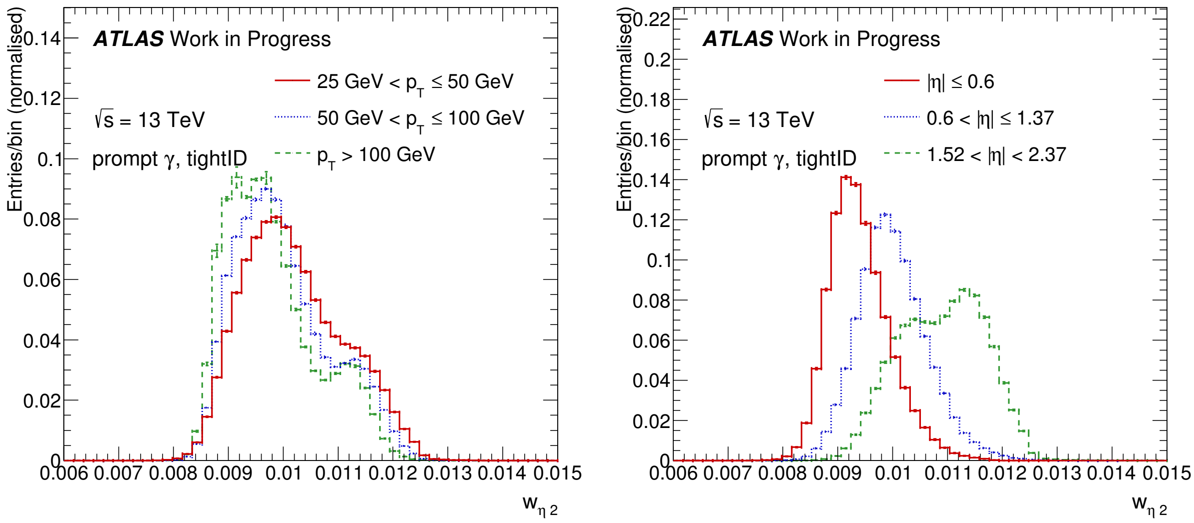


Figure 6.3.: Distributions of $w_{\eta 2}$ in 3 bins of p_{T} and η.

In the strip layer, $w_{\eta 1}$ and $w_{\mathrm{tot,s1}}$ measure the energy weighted width in units of strips using 3 and 20 strips around the maximum energy deposit, respectively. For the same arguments discussed before, the width gets smaller with decreasing p_{T} as shown in Fig. 6.4. A coarser strip granularity with increasing η in the end-cap shifts the maximum back to smaller values in the third η bin. When the strip widths gets larger, more energy is deposited in single strips which effectively decreases the strip width of the shower. The effect is more prominent for $w_{\mathrm{tot,s1}}$ since more strips are taken into account.

f_{side} measures the energy contained in 4 strips surrounding 3 strips around the maximum relative to these 3 strips. A more narrow shower is expected with increasing p_{T} and hence the leakage into the 4 outermost strips decreases which can be seen in Fig. 6.4. A coarser granularity in the largest η bin leads to a larger energy deposit within a smaller number of strips and hence the maximum is shifted back to smaller values with respect to the first two η bins.

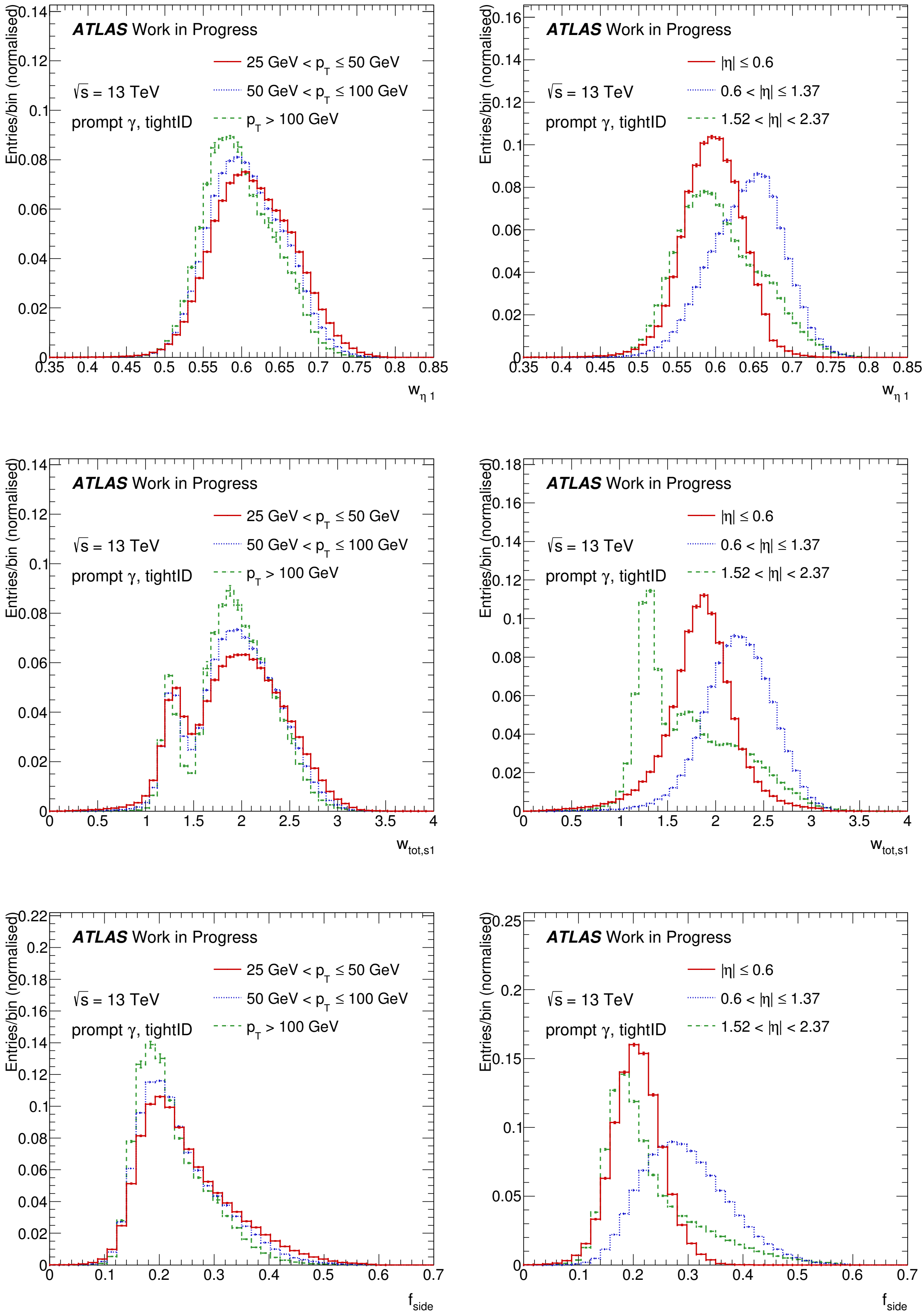


Figure 6.4.: Distributions of $w_{\eta 1}$, $w_{\mathrm{tot,s1}}$ and f_{side} in 3 bins of p_{T} and η.

Larger p_{T} values also cause larger values of E_{ratio} since a second energy maximum is either small or even not present which leads to a value of 1. With decreasing p_{T} the probability of a second energy maximum in the η direction increases which leads to a broader distribution and a smaller maximum at lower values as seen in Fig. 6.5. A coarser granularity in the third η bin also lowers the probability of a second maximum and hence the maximum is shifted back to larger values with respect to the former η bins.

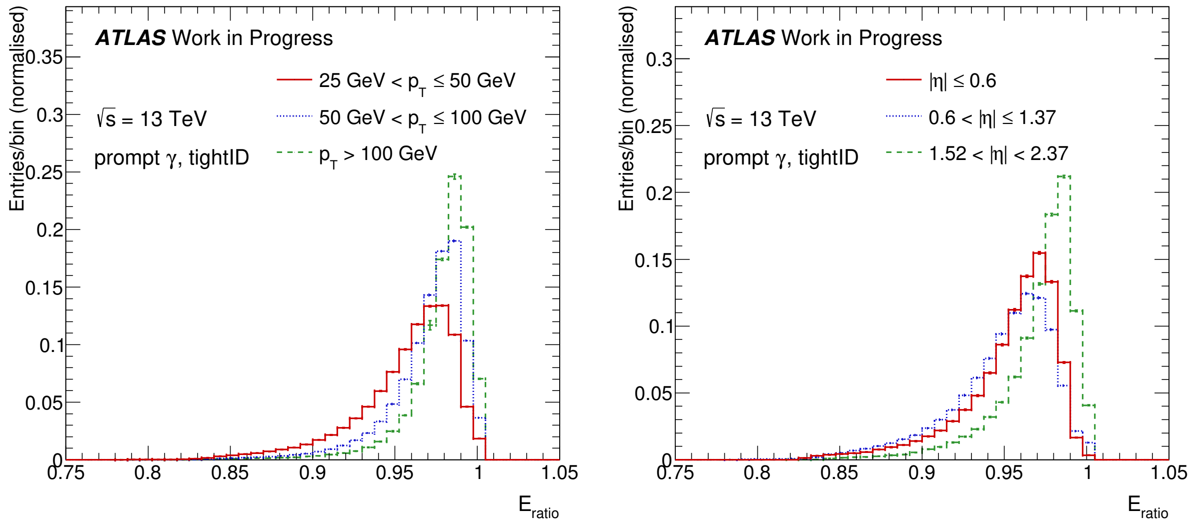


Figure 6.5.: Distributions of E_{ratio} in 3 bins of p_{T} and η.

6.1.3. Shower shapes of prompt photons and hadron fakes

The single photon and di-jet samples introduced in Sec. 5.1 yield $8,973,940$ prompt photons and $209,065$ hadron fakes, respectively. In the following, only 1,045,324 prompt photon samples are used which is ≈ 5 times the number of available hadron fakes and corresponds to ≈ 12 % of all available prompt photons. No information is lost and the shower shape distributions agree within their statistical uncertainties when comparing the shapes of all prompt photon samples to that of random 12 % split[5].

Hadrons misidentified as photons or photons originating from hadronic decays are expected to be accompanied by additional hadronic activity leading to additional energy deposits in the EMCal. Since hadrons loose only a small amount of their energy in the EMCal the effect is expected to be small. It is, however, visible on the scale of single calorimeter cells as shown in the following.

[5] App. A.2 summarises the corresponding distributions.

The *separation* between two normalised distributions is calculated according to

$$\mathcal{S} = \frac{1}{2} \sum_{i \in \text{bins}} \frac{(p_i - f_i)^2}{p_i + f_i}, \tag{6.1}$$

where p_i and f_i are the prompt photon and hadron fake contribution in bin i, respectively. Since the separation is calculated with respect to the nominal bin values, uncertainties must be negligible.

Separation plots of shower shapes R_{had}, R_η, R_ϕ, $w_{\eta 1}$, $w_{\eta 2}$, $w_{\text{tot,s1}}$, f_{side} and E_{ratio} can be found in Fig. 6.6. They show the normalised distributions of both prompt photons and hadron fakes where only statistical error bars are considered. All distributions show a separation of a few percent between prompt photons and hadron fakes. This separation is mainly caused by additional hadronic activity. Hence, the hadronic leakage of hadron fakes is larger than that of prompt photons and the widths in the second layer of the EMCal, measured by R_η, R_ϕ and $w_{\eta 2}$, are larger. The same can be observed for observables obtained in the strip layer. Broader shower shapes of hadron fakes lead to distributions which are shifted relative to those of prompt photons. Furthermore, the maxima are smaller and a larger contribution to the tails of the respective distributions can be observed. Tab. 6.2 summarises the separations calculated according to Eq. (6.1). The binning is included in the Table since a slightly different binning can change the separation slightly[6].

shower shape	**separation** [%]	**binning** (#bins, [low, up])
R_{had}	3.33	40, [-0.1, 0.1]
R_η	4.83	50, [0.88, 1.05]
R_ϕ	7.01	50, [0.6, 1.02]
$w_{\eta 2}$	2.01	50, [0.006, 0.015]
$w_{\eta 1}$	4.14	50, [0.35, 0.85]
$w_{\text{tot,s1}}$	4.55	50,[0.0, 4.0]
f_{side}	7.21	40, [0.0, 0.7]
E_{ratio}	1.89	40, [0.75, 1.05]
ΔE	0.336	50, [-10 MeV, 300 MeV]

Table 6.2.: Separation between prompt photons and hadron fakes of photon shower shapes calculated according to Eq. 6.1.

[6]For instance, the extreme case of only one single bin for prompt photons and hadron fakes leads to a separation of 0.

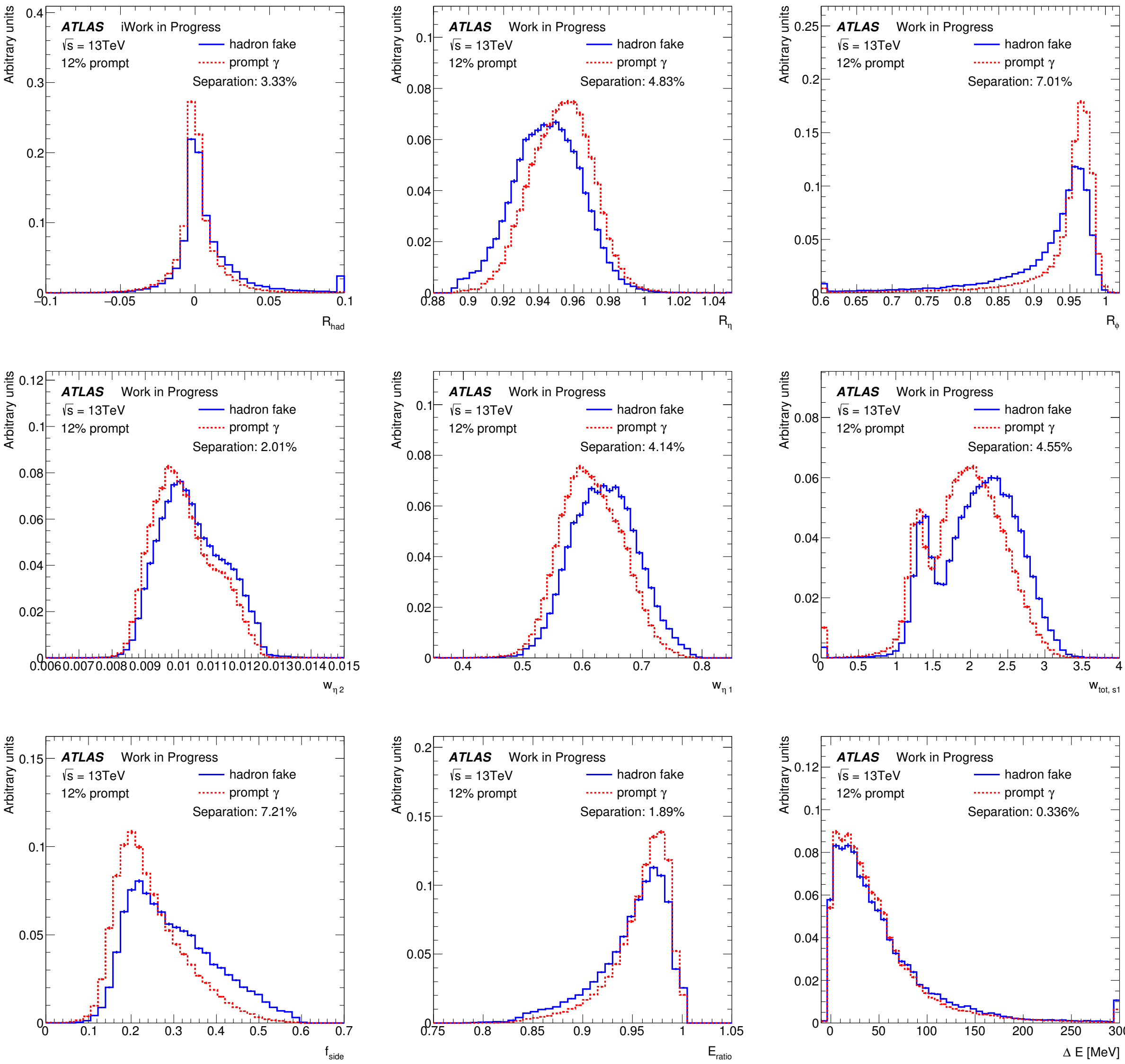


Figure 6.6.: Separation plots of prompt photons and hadron fakes showing the normalised distributions. The separation indicated in the plots is calculated according to Eq. (6.1).

6.2. Training and development of the PPT

The developed neural network (NN) aims for a purer selection of prompt photons along with a better rejection of hadron fakes *within* the set of tightID photons and will be called *prompt photon tagger* (PPT). Photon reconstruction and identification (see Sec. 3.3) is not altered. The PPT is a *binary* classification tool labelling tightID photon candidates with values between 0 and 1: Values towards 1 indicate prompt photons whereas values towards 0 indicate hadron fakes. It is trained using the Keras [90] library with Theano [91] as the backend. After training, the NN should be applicable in any analysis with photon candidates present. It is therefore developed as an analysis independent tool. Application of the PPT to photon candidates is managed using the LightWeight Neural Network library or LWTNN [92].

Feature selection

Observables which are measurable independently of an analysis have to be chosen in order to guarantee that the derived PPT can be used in any analysis involving photons. Depending on the chosen features the performance of the discrimination may change due to additional activity, for instance, when the number of jets changes which alters the hadronic activity in an event. Also, kinematic or other additional cuts can change the feature distributions which in turn changes the performance of a PPT relying on those features.

Photon shower shapes are potential feature candidates. They only depend on the reconstructed photon candidate's energy deposits in the calorimeter system and can therefore be measured in any analysis dealing with photons. A different number of jets was shown to have a minor impact, however, photon kinematic distributions have a significant impact on photon shower shapes as discussed in Sec. 6.1.3. Thus, the performance of a PPT using photon shower shapes is expected to depend on the analysis and its selection criteria.

The individual separation power of each shower shape variable was calculated in Sec. 6.1.3 showing significant values (see Tab. 6.2). However, it cannot be directly inferred on the individual impact of a certain variable on a multi-variate-analysis (MVA) tool since linear and higher order correlations between the variables play a role. Final conclusions on the impact of individual variables can therefore only be drawn after training an MVA tool. Fig. 6.7 shows significant linear correlations among the shower shape variables of prompt photons. NNs can deal with correlations and hence be used in the following to enhance the discrimination between prompt photons and hadron fakes based on shower shapes.

The set of potential features used in the following contains R_{had}, R_η, R_ϕ, f_{side}, $w_{\eta 1}$, $w_{\eta 2}$, $w_{\mathrm{tot,s1}}$ and E_{ratio}. Due to its small separation power of 0.304 % ΔE is not considered[7].

[7] As stated before it can not be inferred on the final performance from the individual separations. However,

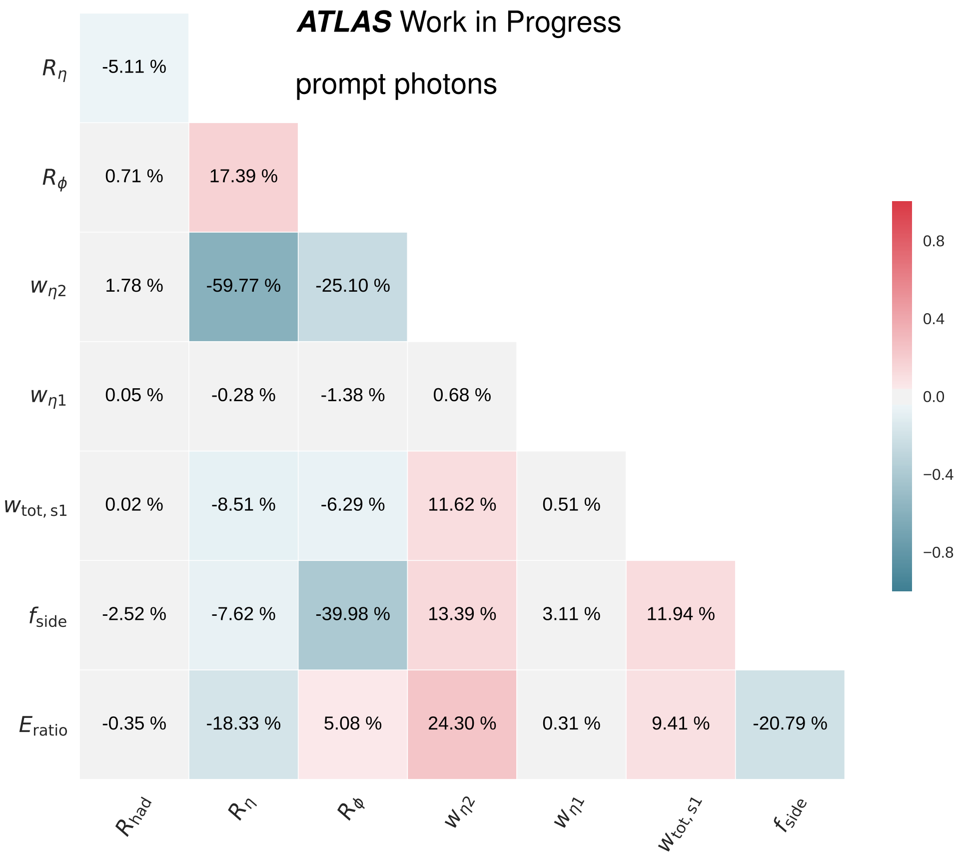


Figure 6.7.: Linear correlations between shower shape variables of prompt photons.

6.2.1. Data preparation using the `root2kerasPipeline`

The standard data structure of the `ROOT` analysis package[8] cannot be read by `Keras` and a preprocessing step is necessary. The `root2kerasPipeline`[9] tool saves the output in the `HDF5` format which can be read in a `Python` environment. In addition to the conversion, observables of interest based on desired cuts can be selected.

testing all combinations of feature sets and `NN` architectures goes beyond the scope of this thesis. Some reasonable decisions must therefore be made to decrease the set of possible configurations.

[8] `https://root.cern.ch/`, visited on October 2, 2017.

[9] `https://gitlab.cern.ch/bvolkel/root2kerasPipeline`, visited on October 2, 2017.

Two different conversion procedures are available to ensure that all samples after a conversion have the same size corresponding to the fixed number of input neurons of the `NN`[10].

1. If a `NN` should be used to discriminate between events of different processes, a maximum number for objects is specified. If there are more objects, those are cut away and if the number of objects is lower than the maximum number, remaining inputs corresponding to observables of these objects are filled with 0's.

2. If the `NN` should be trained to discriminate between objects, all observables of that particular object are assumed to have the same length per event. For instance, if there are n photons in an event, there will be n individual samples in the output.

6.2.2. Training

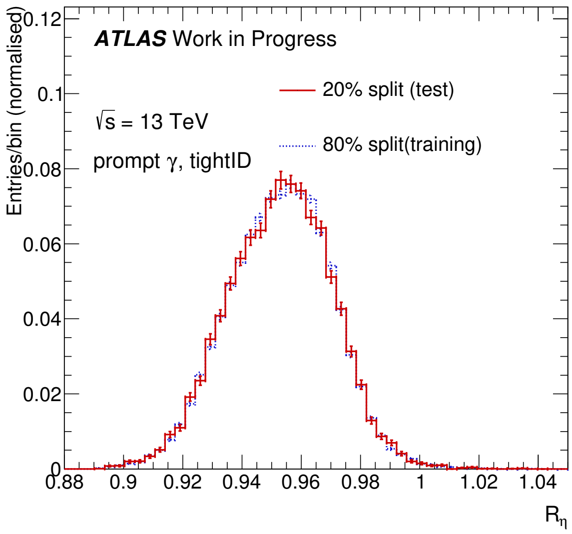
All `NN`s presented in the following have a feed-forward structure using the `adam` optimiser [93] and the binary cross entropy loss (see Eq. (4.3)) in training. 209,065 hadron fake samples derived from the di-jet simulations and 1,045,324 prompt photons are used.

Although it is generally possible, photon samples are not weighted in training due to their MC weights. That is done to reduce the impact of the underlying processes on the `PPT` performance. Effectively, the feature distributions in training are hence different from those in the single photon and di-jet processes. As discussed before the discrimination performance achieved in training is then expected to be different from that when the `PPT` is applied to prompt photons and hadron fakes weighted by their respective event weights.

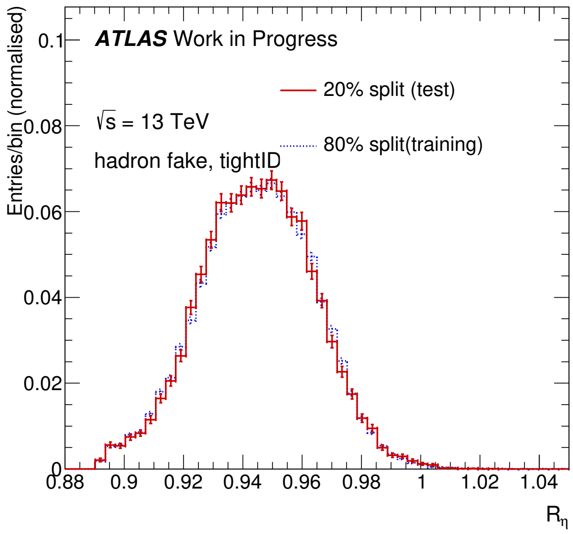
Fig. 6.8 shows the distributions of R_η, for prompt photons and hadron fakes[11] where the training samples are random splits of the original data into 2 orthogonal subsets containing 20 % and 80 % of the respective data. All distributions agree well within statistical uncertainties. The same feature distributions in the training and test set is a necessary condition to conclude on the generalisation capability of the `PPT` from overlaying ROC curves. Therefore, a split of data into 80 % training and 20 % test samples is reasonable and will be used in the following.

A training as well as a cross-validation run can be initiated passing a simple configuration file to the training script in the `root2kerasPipeline`. Various parameters can be set as summarised in App. A.1.

[10] There are approaches using so-called *recurrent* `NN`s. Information fed to those are passed back and can be added to further information fed afterwards. Effectively, it can be dealt with a varying number of inputs.
[11] Distributions of further shower shapes can be found in App. A.3.



(a) Prompt photons.



(b) Hadron fakes.

Figure 6.8.: Comparison of the shower shape distributions of R_η drawn from orthogonal 20 % and 80 % splits of 1,045,324 prompt photons and 209,065 hadron fakes.

6.2.3. PPT development

An effect similar to overtraining is possible with respect to MC modelling. In that case, imperfections in modelling and the fact that calculations cannot be done to infinite precision can lead to slightly different distributions of object distributions or event variables when changing, for instance, the underlying MC generator. When training and application is done using the same objects, the performance of the derived NN is potentially biased. This bias is minimised when the NN is trained using a statistically and systematically independent sample. This will be further discussed in Sec. 6.3.

The PPT performance will be evaluated in the training and testing considering three performance metrics:

1. As the measure of the discrimination performance the values of the AUC in the training and test set is considered. The higher the value the better the PPT's classification capability is considered to be.

2. The capability of generalising to unseen data is seen by overlaying training and test ROC curves.

3. To ensure that the derived NN is reproducible and that its classification and generalisation capabilities are stable, a k-fold cross-validation is applied.

In the first step different feature sets and architectures are compared to each other. After that a conclusion on the best performing PPT is drawn and a final cross-validation is conducted. Hadron fake samples are weighted twice as strong as prompt photon samples in the training to slightly increase their impact on the loss function. The initial feature set contains the 4 highest separating variables in single photon and di-jet samples which are R_η, R_ϕ, f_{side} and $w_{\eta 1}$. $w_{\text{tot,s1}}$ is not considered here since its values differ by one order of magnitude from the ones before which have values in (0, 1). R_{had}, E_{ratio} and $w_{\eta 2}$ are added step by step and in addition, 4 different architectures are considered according to Tab. 6.3. The first architecture only incorporates the first layer and in 3 further steps more layers are added while the input and output layers are always included. 16 different models are therefore studied using a batch size of 10,000 and training over 300 epochs.

number	**architecture**	**# layers**
1	`relu`, 64 neurons	1
2	+ `batchnormalisation` + `softmax`, 40 neurons	3
3	+ `batchnormalisation` + `softmax`, 52 neurons	5
4	+ `batchnormalisation` + `relu`, 8 neurons	7
output layer	`sigmoid`, 1 neuron	+1

Table 6.3.: 4 different architectures to be tested for the PPT. Starting with one layer further layers are added successively. The output layer is included in all cases. The different types of layers and activation functions can be found in Sec. 4.2.

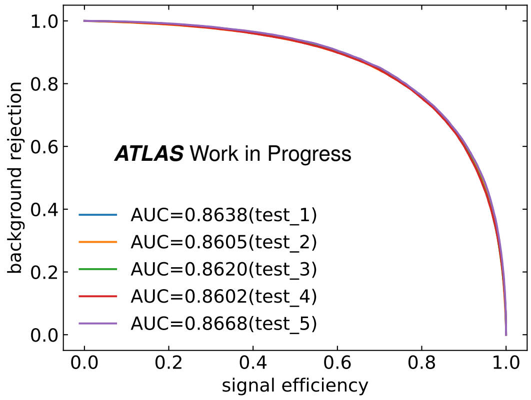
AUC values in the training and test run can be found in Tab. 6.4 and the corresponding ROC curves are summarised in App. A.6. For all features sets the AUC values of architecture 4 are lower than for architecture 3. However, the agreement of the training and test ROC curves as well as that of the AUC values indicates that neither over- nor undertraining is the reason for that. A possible explanation is the increase of trainable parameters, meaning weights and biases of the NN. It increases from 5,441 for architecture 3 to 6,333 for architecture 4. Training both architectures for the same number of epochs indicates that due to the larger number of free parameters those have not been well optimised. With an increasing number of parameters the number of epochs has to be increased as well.

The starting point of the development of the PPT for the $t\bar{t}\gamma$ analysis at $\sqrt{s} = 13\,\text{TeV}$ at that time was the set containing 4, 5 and 6 features, respectively. Hence, these were the ones considered for implementation.

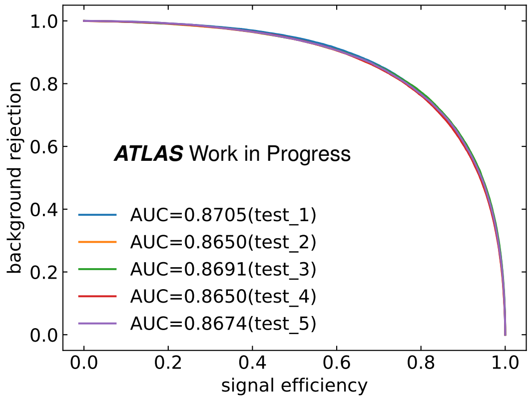
architecture	features	AUC (training)	AUC (test)
1	R_η, R_ϕ, f_{side}, $w_{\eta 1}$	0.8249	0.8266
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}	0.8414	0.8401
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$	0.8486	0.8477
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$, E_{ratio}	0.8728	0.8738
2	R_η, R_ϕ, f_{side}, $w_{\eta 1}$	0.8411	0.8419
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}	0.8668	0.8657
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$	0.8654	0.8649
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$, E_{ratio}	0.8792	0.8793
3	R_η, R_ϕ, f_{side}, $w_{\eta 1}$	0.8460	0.8452
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}	0.8631	0.8621
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$	0.8693	0.8690
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$, E_{ratio}	0.8846	0.8839
4	R_η, R_ϕ, f_{side}, $w_{\eta 1}$	0.8291	0.8297
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}	0.8595	0.8584
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$	0.8491	0.8475
	R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$, E_{ratio}	0.8742	0.8735

Table 6.4.: Summary of training and test AUC values for all 16 PPT configurations considered.

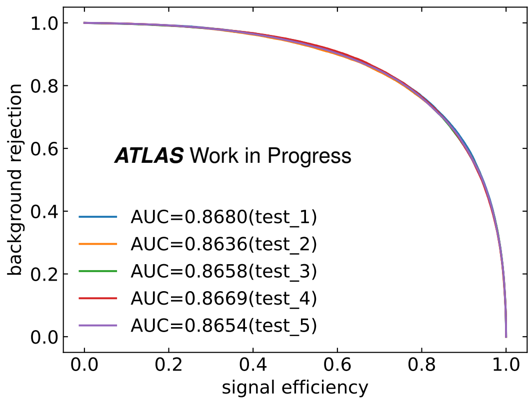
A 5-fold cross-validation is conducted on the configurations with the largest AUC values, which are those with 5 and 6 features having 3 and 5 hidden layers, respectively. A number of 5 folds was chosen to reproduce the sample ratio of 80 % and 20 % used in training. The plots containing all 5 test ROC curves of the respective cross-validation run can be found in Fig. 6.9. The highest average AUC value is 0.8686 in the case if architecture 3 involving 6 features whereas the lowest one is 0.8627 for architecture 2 with 5 features. The maximum difference between the AUC values is 0.90 %, which is considered to be small. All 4 scenarios show therefore a comparable performance and can be considered as possible implementations into the $t\bar{t}\gamma$ analysis. Since the architecture 3 with 6 features was studied and evaluated before the other ones and it was found to perform well, it was decided to implement it. In the following, the term ‘PPT’ always implies architecture 3 with 6 features.



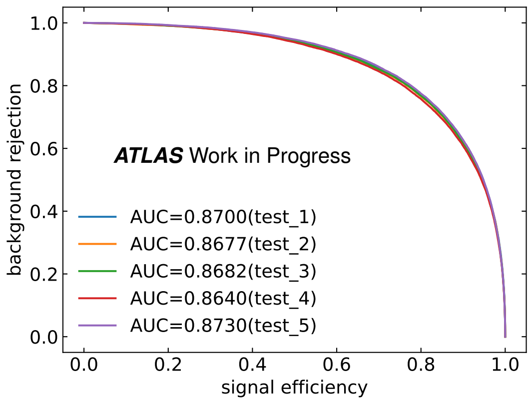
(a) Architecture 2 with features R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}.



(b) Architecture 2 with features R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$



(c) Architecture 3 with features R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}.



(d) Architecture 3 with features R_η, R_ϕ, f_{side}, $w_{\eta 1}$, R_{had}, $w_{\eta 2}$.

Figure 6.9.: ROC curves of each the test set in a 5-fold cross-validation for 2 architectures (see Tab 6.3) and 2 feature sets.

6.2.4. PPT in different bins of p_T and η

Fig. 6.10 shows the PPT shapes of prompt photons obtained from the single photon training sample using the same p_T and η bins as in the study of shower shapes in Sec. 6.1.2. The maximum of the PPT distribution for high p_T photons is significantly larger compared to those in the lower p_T bins. That can be expected since a high p_T leads to a more narrow shower in the $r-\phi$ plane which makes the shower shapes look more prompt photon like. A similar argument holds for a better discrimination in the lowest η bin. Central photons are less boosted along the beamline and therefore less smeared out in the η direction which yields narrower shower shape distributions.

It is also observed that the performance differs much less between the last two η bins which is different for the individual shower shapes. Especially the ones obtained in the strip layer of the EMCal show large difference in the outermost η bin which corresponds to the end-cap part and has a coarser granularity than the barrel.

Further studies of how the PPT performance depends on the photon kinematics can be conducted, for instance by probing a finer binning in η or by investigating correlations.

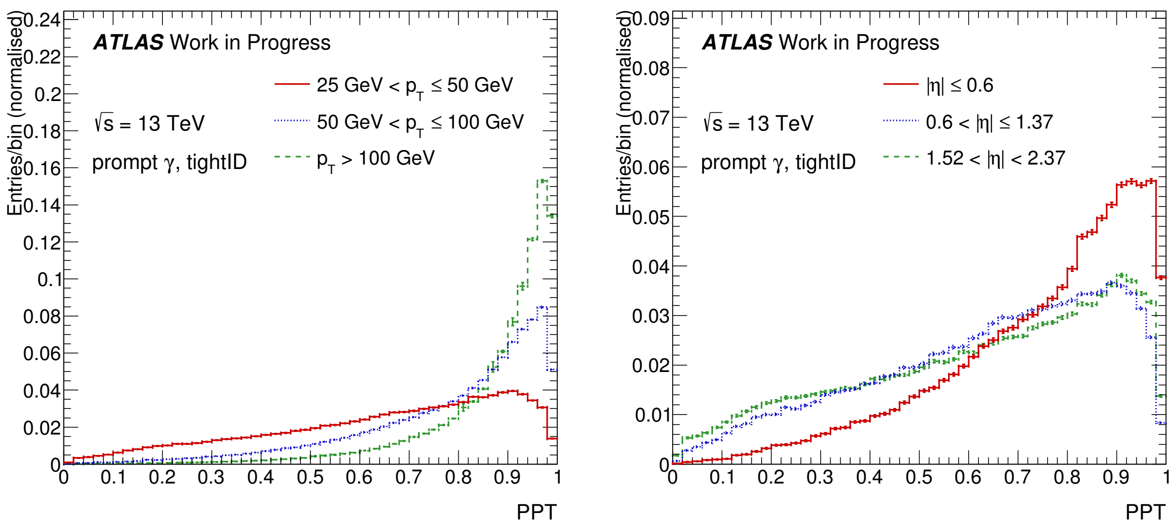


Figure 6.10.: PPT distribution of prompt photons in 3 different bins of p_T and η.

PPT in bins of different jet multiplicity

To study the impact of the jet multiplicity as it was done for shower shapes individually in Sec. 6.1.1, Fig. 6.11 shows the PPT distributions in the corresponding 5 bins. Up to a PPT value of ≈ 0.9 the distributions agree within their statistical uncertainties. However, the small deviations between the nominal distributions in the lower bins are summed up in the highest showing a statistically significant deviation between the ≥ 6 jets bin and the ones with lower jet multiplicity. Since the maximum region of the PPT is a convolution of the maxima of the shower shape distributions, the small deviations around the maxima observed in Sec. 6.1.1 potentially lead to the visible differences in the large PPT region.

Although 5 and 6 jets lead to the largest additional energy contamination, it has to be kept in mind that the 2 and 3 jet bins are obtained in the dilepton channel of the $t\bar{t}\gamma$ samples. In the case of forward boosted top-quarks also the radiated photon is expected to be boosted reaching larger η regions where the discrimination performance of the PPT gets poorer. Smaller longitudinal boosts cause the photons to be more central and in a region of better discrimination performance which could be the case for larger jet multiplicities. To draw a final conclusion further studies, for instance, investigating the relation between the jet multiplicity and the pseudo rapidity of the photon, are necessary.

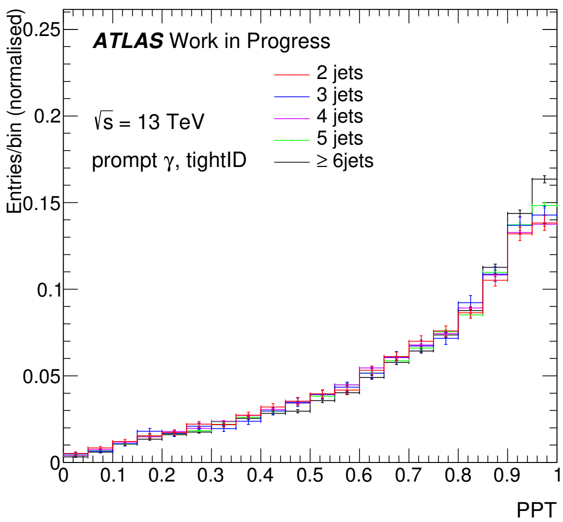


Figure 6.11.: PPT distribution of prompt photons in 5 bins of different jet multiplicities.

6.3. The PPT in the 13 TeV $t\bar{t}\gamma$ analysis

6.3.1. Analysis strategy at 13 TeV

The $t\bar{t}\gamma$ analysis at 13 TeV aims to increase the significance and reduce systematic uncertainties. Both total and differential production cross sections are measured in both the single lepton and dilepton channel. In addition to the object definitions in Sec. 5.2, photon candidates have to pass the `FixedCutTight` (`FCT`) isolation working point defined as $E_{\mathrm{T}}^{\mathrm{topocone40}} < 0.022\, p_{\mathrm{T}} + 2.45$ GeV as well as $p_{\mathrm{T}}^{\mathrm{cone20}} < 0.05\, p_{\mathrm{T}}$. It is used to minimise the hadron fake contamination since hadron fakes are expected to be less isolated than prompt photons. After this cut, hadron fakes are still a dominant background contribution in the single lepton channel. An overall scale factor for hadron fakes is derived in a data-driven approach. The value of the `PPT` assigned to the photon candidates in the selected events is used to discriminate between events with prompt photons and hadron fakes. Together with other parameters the `PPT` value is passed to another `NN` which is trained using $t\bar{t}\gamma$ signal and background MC samples introduced in Sec. 5.1. It is a binary classification tool trained to discriminate $t\bar{t}\gamma$ signal events against the background events inclusively and will be called *Event-Level-Disciminator* (`ELD`).

Tab. 6.5 summarises the event selection cuts of the signal region (SR) separately for all contributions in the single lepton and dilepton channel. Due to the expectation of at least 4 jets, the hadron fake contribution has a larger impact in the single lepton than in the dilepton channel. For that reason the `PPT` is used there to further discriminate events containing hadron fakes.

<table>
<tr><th>channel</th><th>$e+$jets</th><th>$\mu+$jets</th><th>ee</th><th>$\mu\mu$</th><th>$e\mu$</th></tr>
<tr><td rowspan="3">common</td><td>$== 1\, e$</td><td>$== 1\, \mu$</td><td>$== 1\, e^+e^-$</td><td>$== 1\, \mu^+\mu^-$</td><td>$== 1\, e^\pm\mu^\mp$</td></tr>
<tr><td colspan="5">trigger match</td></tr>
<tr><td colspan="2">-</td><td colspan="3">$m(\ell,\ell) > 15$ GeV</td></tr>
<tr><td>photon</td><td colspan="5">$== 1$</td></tr>
<tr><td>jet</td><td colspan="2">≥ 4</td><td colspan="3">≥ 2</td></tr>
<tr><td>b-jet</td><td colspan="5">≥ 1</td></tr>
<tr><td>$m(\ell,\ell)$</td><td colspan="2">-</td><td colspan="2">not in [85,96] GeV</td><td>-</td></tr>
<tr><td>$m(\ell,\ell,\gamma)$</td><td colspan="2">-</td><td colspan="2">not in [85,95] GeV</td><td>-</td></tr>
<tr><td>$\not{E}_T$</td><td colspan="2">-</td><td colspan="3">> 30 GeV</td></tr>
<tr><td>$m(\gamma,e)$</td><td>not in [85,95] GeV</td><td colspan="4">-</td></tr>
<tr><td>$\Delta R(\gamma,\ell)$</td><td colspan="5">> 1.0</td></tr>
</table>

Table 6.5.: Event selection cuts of the single and dilepton channel in the $t\bar{t}\gamma$ analysis at 13 TeV.

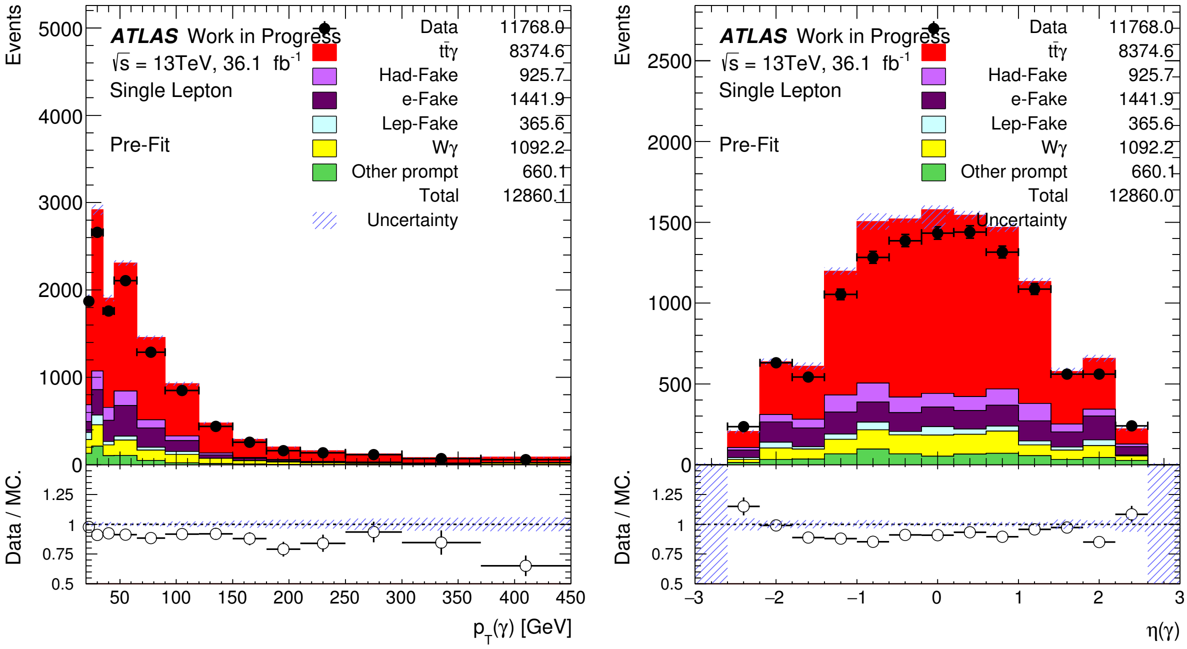


Figure 6.12.: p_{T} and η distributions of selected events in the signal region showing the contributions from $t\bar{t}\gamma$ signal events together with those containing a hadron fake and further important background contributions.

Fig. 6.12 shows the pre-fit plots of the p_{T} and η distribution of selected events in the single lepton channel. Hadron fakes are predominantly selected with $p_{\mathrm{T}} < 150$ GeV and prompt photons with transverse momenta of a few 100 GeV. This p_{T} range as well as the η range are covered by photon candidates contained in the single photon and di-jet samples and hence the PPT training accounts for the full kinematic phase space of photons in the $t\bar{t}\gamma$ analysis.

6.3.2. Implementation and performance of the PPT in the $t\bar{t}\gamma$ analysis

Fig. 6.13 shows the PPT distribution in the single lepton channel of the $t\bar{t}\gamma$ signal region. The signal events shown in red which contain a prompt photon are labelled preferably with large values. That is the desired and expected behaviour of the PPT. Furthermore, other events containing prompt photons get larger PPT values as well which is also expected since the PPT was only trained on photon shower shapes without taking any event variables into account. In most cases events with an electron faking the prompt photon are also labelled with larger values.

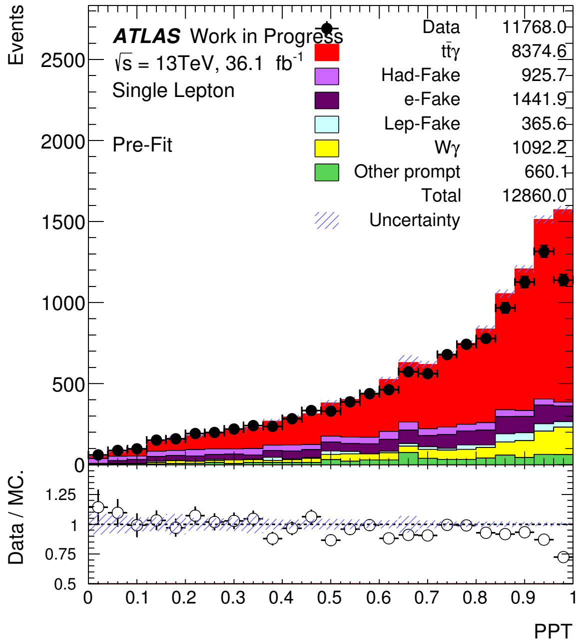


Figure 6.13.: PPT distributions of the single lepton channel in the $t\bar{t}\gamma$ signal region.

In the high PPT region the data/MC disagreement is found to be largest showing an overestimation in simulation. Since high p_{T} photons are more likely to be labelled with high PPT values as discussed in Sec. 6.2.4, the overestimation of prompt photons in the high p_{T} region as it can be seen in Fig. 6.12 can be one origin of the data/MC disagreement of the PPT seen in Fig. 6.13. Another origin caused by potential mismodelling in simulation will be discussed in the following Section and it is also shown that the disagreement is covered by systematic uncertainties.

The distribution of hadron fakes is mostly flat and looking at the separation plots in Fig. 6.14 it can be seen that the separation of 17.4 % between prompt photons and hadron fakes is mostly achieved by discriminating prompt photons well. The smallest separation is obtained between signal events and events containing prompt photons from the $W\gamma$ process or other events with prompt photons. In both cases the distributions agree within statistical uncertainties except for the PPT region above 0.9 in the case of other prompt photons. Furthermore, a fluctuation can be seen in the distribution of other prompt photons caused by large MC weights in the simulation. Although having a separation of 7.57 %, the distributions of prompt photons and Lep-Fakes agree within statistical uncertainties meaning that the separation quoted in that case is not reliable. A significant separation of 4.89 % can also be seen for electrons faking photons which, however, is still one order of magnitude below the separation of hadron fakes. The overall separation with respect to all background events combined is 3.46 %.

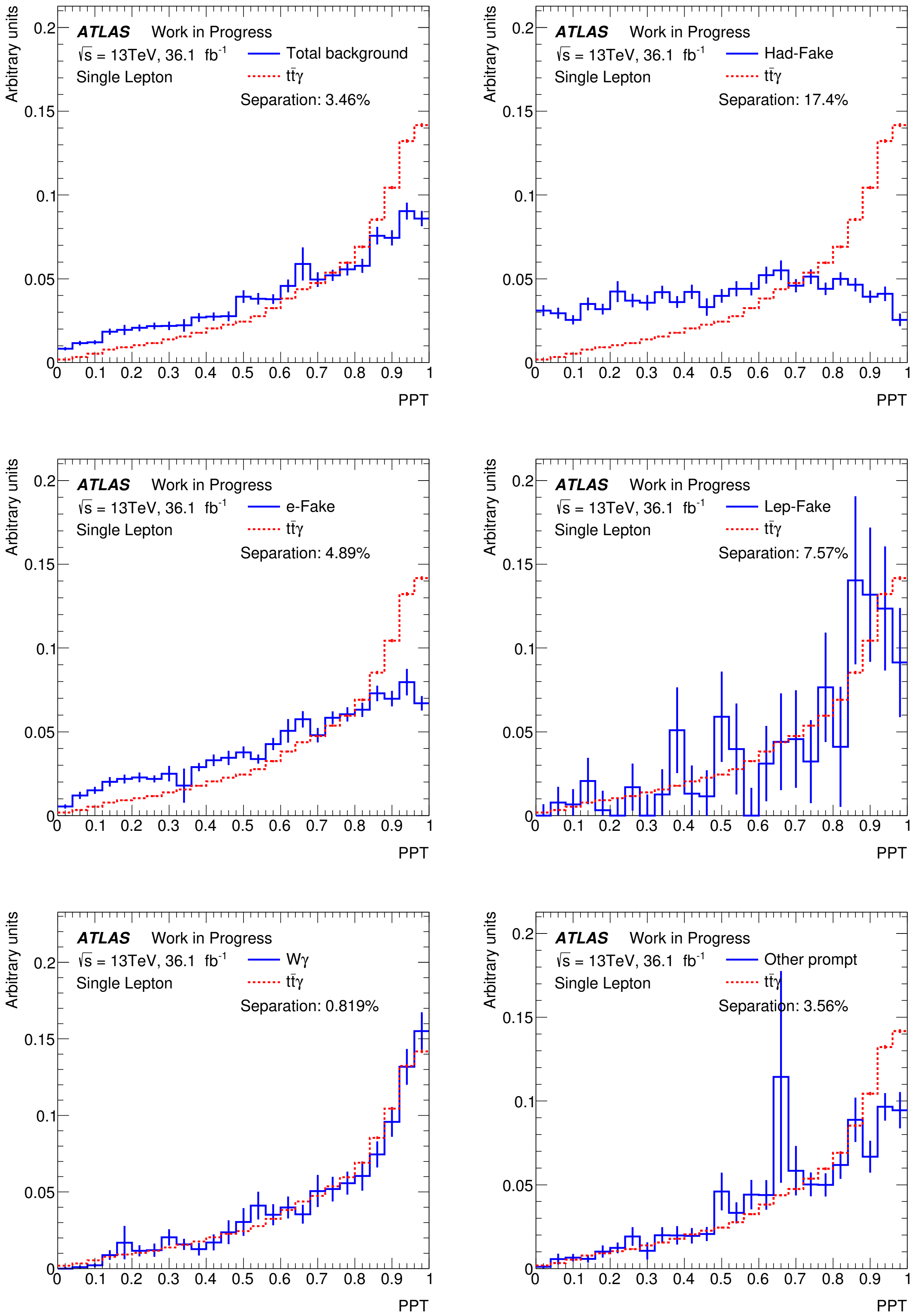


Figure 6.14.: Separation plots showing the PPT distributions of the $t\bar{t}\gamma$ signal events and those of various background contributions in the SR.

6.3.3. Systematic uncertainties of the PPT

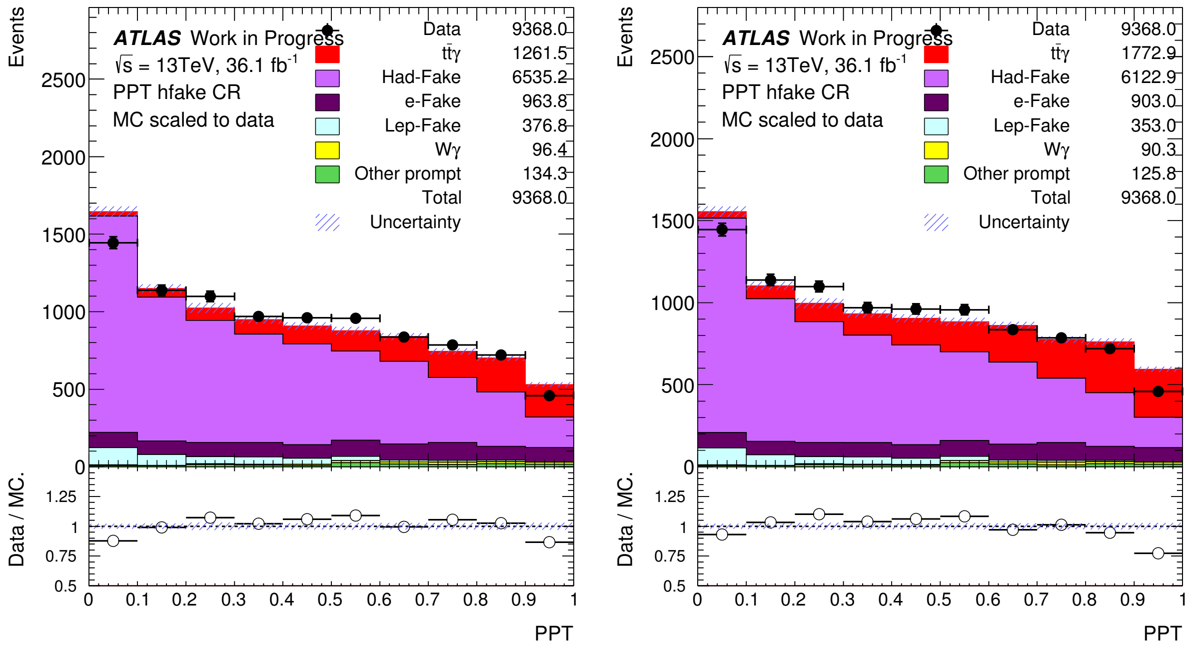
The final template fit of the ELD needs to include systematic uncertainties of the PPT. Since the PPT distribution as well as the discrimination performance depend on the kinematic distributions of the prompt photons and hadron fakes, the systematic uncertainties are derived directly in the $t\bar{t}\gamma$ analysis. To cover both an uncertainty on prompt photons and hadron fakes, two control regions (CRs) are defined. Shape uncertainties of the PPT are derived from the data/MC disagreement in these particular CRs. For that purpose MC yields are scaled to data before the ratio is extracted.

Hadron fake CR

The CR for hadron fakes is based on the selection cuts of the single lepton channel in the signal region summarised in Tab. 6.5. Different from the SR, photon candidates have to fail the FCT working point and be poorly isolated requiring $p_{\mathrm{T}}^{\mathrm{cone20}} > 3$ GeV. Since the shape deviation with respect to hadron fakes is of interest, a conservative approach of estimating uncertainties is to subtract the entire contribution of the signal sample. In the present study, the theoretical uncertainty on the signal sample is used to vary its contribution. Ongoing studies conducted by the analysis team show a relative uncertainty of ≈ 14 %. However, 20 % has been chosen to be conservative leading to a contribution of 80 % and 120 % of the nominal signal. After scaling the MC yields to data, the ratio with larger disagreement between data and MC is taken to be the systematics shape uncertainty. Fig. 6.15 shows both plots and the overall shape disagreement is found to be larger for 120 % of the nominal signal and hence, this data/MC ratio is chosen as the shape uncertainty. This uncertainty will be applied to all hadron fakes in simulation in the final fit.

Furthermore, it can be observed that the deviation between experimental and simulated data is largest in the outermost PPT bin towards 1. Because different configurations of MC generators, parton shower algorithms and PDF sets are used for the production of training samples and those in the $t\bar{t}\gamma$ analysis, small differences due to different modelling schemes are expected[12]. In the central region of the PPT these effects can be expected to be balanced between neighbouring bins but in the tails the impact gets stronger and is summed up. This can be interpreted as an overtraining effect in simulation. The effect is more prominent for prompt photons with labels towards 1 and a possible explanation for that is that 5 times as many prompt photon samples are used in training such that the PPT is more sensitive to their mismodelled properties. As discussed earlier, a mismodelling in higher p_{T} regions (see Fig. 6.12), which alters the shower shape distributions as shown before, can also contribute.

[12] That is also the reason why the derivation of MC modelling uncertainties is crucial for analyses in general.



(a) 80 % of the nominal $t\bar{t}\gamma$ signal contribution.

(b) 120 % of the nominal $t\bar{t}\gamma$ signal contribution.

Figure 6.15.: `PPT` distributions in the hadron fake CR where the MC yields are scaled to data. Only statistical uncertainties are included and the `PPT` shape uncertainty is derived from data/MC disagreement.

Prompt photon CR

The prompt photon CR is derived from a region enhanced by $Z\gamma$ events. In order to enhance the contribution of those events the invariant mass of the dilepton pair has to lie in the range of [60, 100] GeV. The cut on jets and b-jets is removed. Different from the hadron fake CR this one is not completely orthogonal to the SR since some of its events pass the invariant mass windows of [60, 85] GeV and [95, 100] GeV which are also included in the SR. However, the contribution from the signal MC sample is negligible compared to $Z\gamma$ events. Fig. 6.16 contains the data/MC plot of the `PPT` distribution in this CR showing a prominent and nearly linear slope. The outermost regions towards 0 and 1 show again the largest deviation between experimental data and simulation. As well as for the hadron fake CR the reason can be an overtraining effect with respect to modelling in simulation or a general modelling issue of shower shapes.

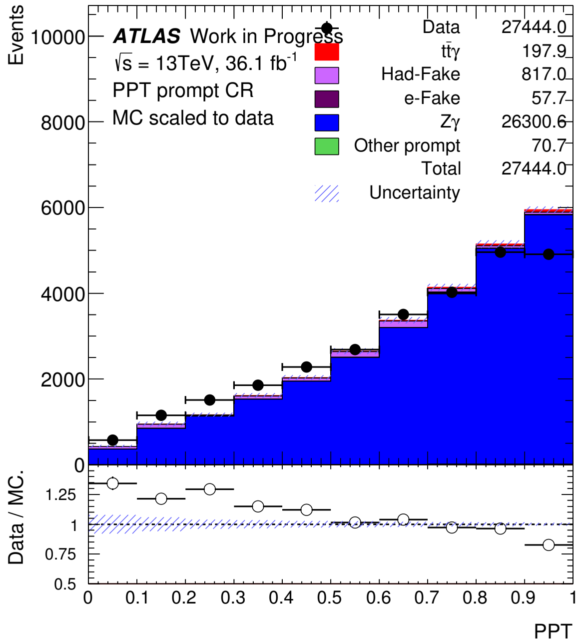


Figure 6.16.: PPT distributions in the prompt photon CR where the MC yields are scaled to data. Only statistical uncertainties are included and the PPT shape uncertainty is derived from data/MC disagreement.

6.3.4. Further PPT studies with the $t\bar{t}\gamma$ MC samples

If the FCT requirement of photons is dropped, the discrimination performance of the PPT increases significantly. Fig. 6.17 shows the PPT distributions in the single lepton channel applying all signal region selection cuts except for the requirement of FCT isolated photons. It can be seen that prompt photons and hadron fakes are well separated and as before, events of the $W\gamma$ process are as well labelled preferably with larger PPT values. Background events with other prompt photon sources or leptons faking photons are labelled more uniformly over the entire PPT range. Fig. 6.18 shows the corresponding separation plots and Tab. 6.6 summarises the separations along with those obtained when FCT isolated photons are required. After dropping the isolation requirement a gain of approximately a factor of 1.7 in the hadron fake separation is achieved. Prompt photons show a maximum towards 1 whereas hadron fakes have a peak at 0. This change can be explained by cutting on FCT isolated photons which alters the distributions of features and hence performance. In addition, the contribution of events with hadron fakes is reduced approximately by a factor of 9.5 whereas the signal contribution is reduced by less than 1.5. A large reduction of hadron fake events is the purpose of the FCT cut but it also decreases the space of different shower shape configurations of hadron fakes which causes a poorer separation in that case and a flat hadron fake distribution when the FCT cut is applied (see Fig. 6.14).

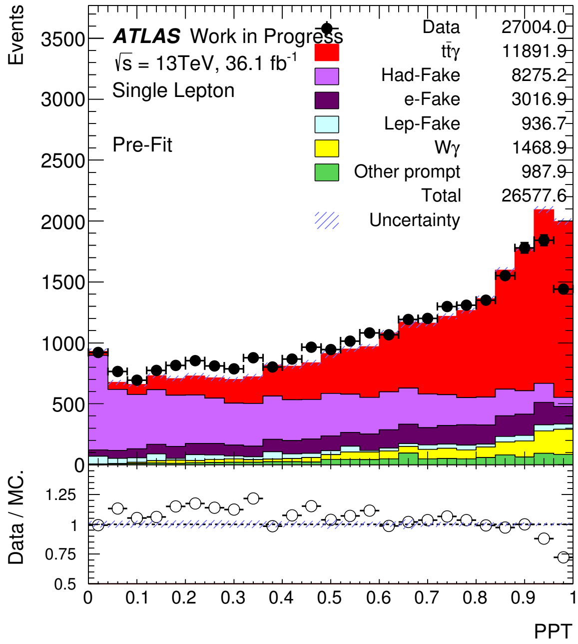


Figure 6.17.: PPT distributions in the single lepton channel after dropping the FCT isolation requirement for photon candidates.

The separation between signal events and those containing leptonic fakes is approximately twice as large as before. The smallest separation can be observed for events with other prompt photon sources where the corresponding separation plots in Fig. 6.18 also show a shape agreement within statistical uncertainties over almost the entire PPT range. Large MC weights in the simulation process cause the fluctuation visible in the distribution. Due to lower number of data the statistical uncertainties of the Lep-Fake distribution are significantly larger compared to other distributions. The overall separation between signal events and all background events is 16 % which is almost 5 times as large as in the case of requiring FCT isolated photons.

background	FCT [%]	noFCT [%]
total	3.46	16
hadron fakes	17.4	29.4
electron fakes	4.89	7.54
lepton fakes	7.57	17.8
$W\gamma$	0.819	0.679
other prompt photons	3.56	2.38

Table 6.6.: Comparison of separations between $t\bar{t}\gamma$ signal events and various background with and without FCT isolation cut.

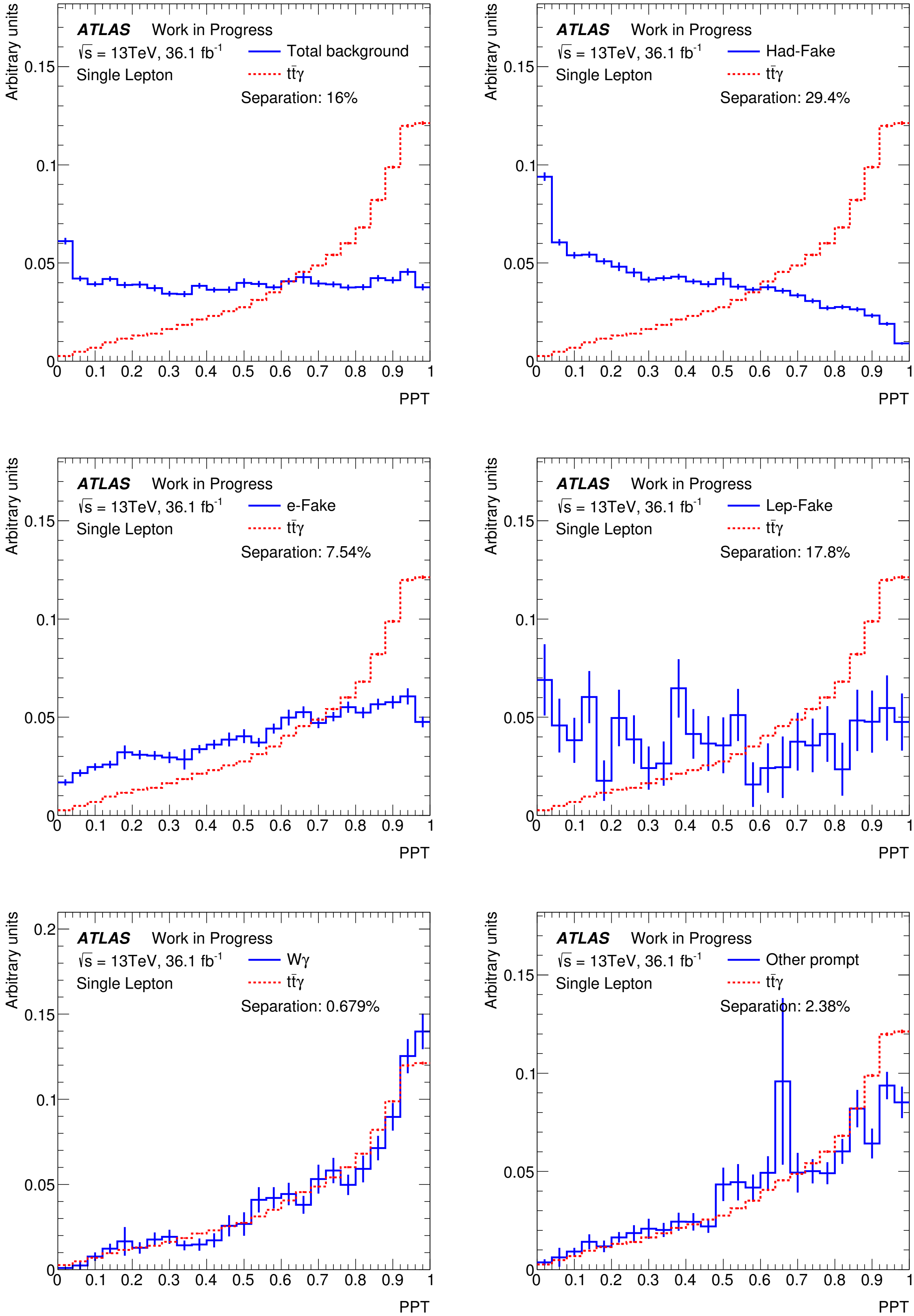


Figure 6.18.: Separation plots showing the PPT distributions of the $t\bar{t}\gamma$ signal events and those of various background contributions when the FCT cut is dropped.

6.4. Analysis independence of the PPT

The performance in the $t\bar{t}\gamma$ analysis has been discussed and the systematic shape uncertainties are derived for prompt photons and hadron fakes in two control regions. In the following, the separation performance of the PPT in the single photon and di-jet training samples is discussed and compared to the performance in the $t\bar{t}\gamma$ samples.

In Fig. 6.19 the separation plot showing the normalised PPT distributions of prompt photons and hadron fakes drawn from the single photon and di-jet sample, respectively, can be seen and a separation of 17 % is achieved. The separation is hence similar to that of 17.4 % in the $t\bar{t}\gamma$ analysis but with different shapes. Prompt photons have a larger tail into the central and low PPT region and the hadron fake distribution is not flat but has a prominent maximum towards 0.

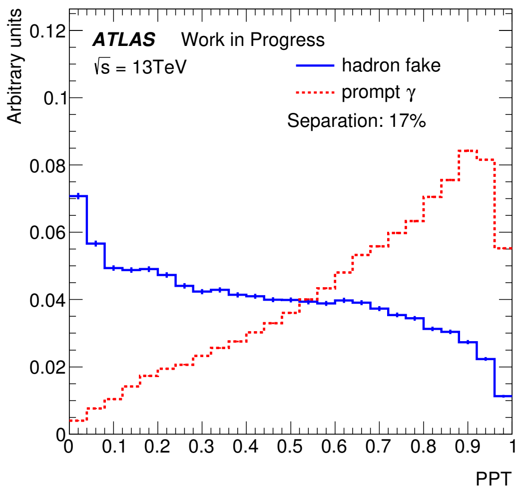


Figure 6.19.: Separation plot showing the PPT distributions of prompt photons and hadron fakes in the single photon and di-jet training samples, respectively.

These shape differences can be explained by different performances in different bins of p_{T} and η. Fig. 6.20 summarises the p_{T} and η distributions of prompt photons in the single photon and the $t\bar{t}\gamma$ signal samples. Most prompt photons of the single photon sample populate the medium p_{T} range between 30 and 50 GeV whereas the distribution of prompt photons in the $t\bar{t}\gamma$ signal sample is significantly broader. Prompt photons in the $t\bar{t}\gamma$ events are also more central and the pseudo rapidity distribution of prompt photons in the single photon sample is approximately flat. Due to these differences of photon kinematic variables, different PPT distributions in the single photon and $t\bar{t}\gamma$ signal sample are expected since the performance labelling prompt photons with high PPT values strongly depends in the kinematic regions as shown in Sec. 6.2.3.

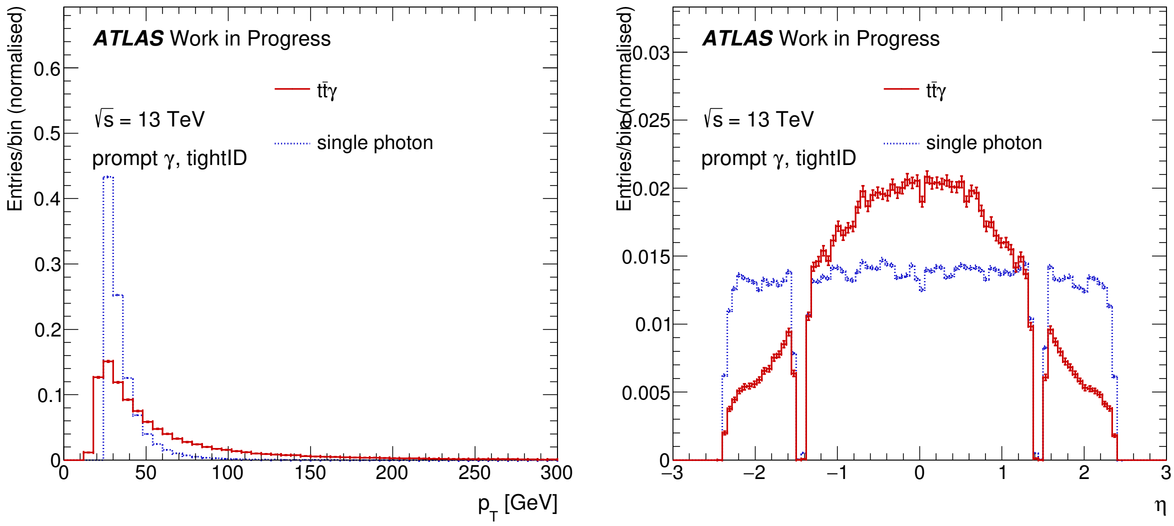


Figure 6.20.: p_{T} and η distributions of prompt photons in the $t\bar{t}\gamma$ signal process and single photon production, respectively.

7. Conclusion

7.1. Summary

The *prompt photon tagger* (PPT) whose development is described in this thesis was successfully implemented into the single lepton channel of the analysis of top-quark pair production in association with a photon at the ATLAS experiment at $\sqrt{s} = 13\,\text{TeV}$ ($t\bar{t}\gamma$ analysis). Together with other variables the PPT value is used as a feature in another NN to discriminate events of the $t\bar{t}\gamma$ signal process from background events.

Training was performed in simulated samples of single photon and di-jet production and the PPT relies on photon shower shape variables as features. These have been shown to provide separations of $\sim 1\,\%$ for `tightID` photon candidates and the separation of the derived PPT is $17\,\%$ on the same samples.

Among 16 configurations covering 4 different architectures and feature sets, 4 configurations have been observed to have comparable performances with AUC values of ≈ 0.86 and agreeable ROC curves in the 5-fold cross-validation. A PPT with 5 layers and 6 shower shape variables was chosen. Besides its good performance it was also the first one which was fully evaluated in the $t\bar{t}\gamma$ analysis including the determination of systematic shape uncertainties.

In the signal region of the analysis the PPT shows a nominal separation between signal and the total background of $3.46\,\%$ for `tightID` photons also passing the `FixedCutTight` isolation working point. Only considering the hadron fake background the separation is $17.4\,\%$ and the separation with respect to other single processes containing prompt photons is $< 4\,\%$ and hence one order of magnitude smaller as compared to hadron fakes. A separation of $\approx 7\,\%$ is observed between signal prompt photons and leptons faking a photon. However, the latter ones are not presented to the PPT during training and hence, it was not straightforward to derive expectations for that separation.

Shower shape variables used as features have been shown to depend on the photon kinematics and thus, the PPT's classification performance varies in different bins of p_{T} and η as it is shown explicitly for prompt photons. The discrimination performance increases with larger photon p_{T} and smaller $|\eta|$, which can be explained by a more narrow shower development averaging over all photon energies.

Correlations between shower shapes make a further improvement of `tightID` selection based on rectangular cuts cumbersome. A multivariate analysis (MVA) approach such as a `NN` can however take those correlations into account providing a single high discriminating variable between prompt photons and hadron fakes.

7.2. Outlook

An overshoot of simulated data in the high `PPT` region (see Fig. 6.13) has been observed. An increasing overshoot of simulated data with increasing p_{T} (see. Fig. 6.12) is potentially one contribution causing this since high p_{T} prompt photons are more likely to be labelled with larger `PPT` values as shown in Sec. 6.2.4. Another reason, which has not been studied, is a potential overtraining effect with respect to simulation. Specific modelling features in one simulation setup can be learned which then leads to a systematically different performance when the algorithm is applied to data obtained with another simulation setup. Although covered by systematic uncertainties both should be investigated further to minimise the impact of incorrectly modelled events on machine learning algorithms. This can also improve the reliability of systematic uncertainties avoiding their under- and overestimation.

Due to the large variety of `NN` architectures[1] in combination with the choice of different feature sets, the probability to find the most suitable (MVA technique solving a certain regression or classification problem is minimised. An improvement of the algorithms in upcoming analyses is therefore very likely by studying them in more and more depth. Already in the case of the `PPT` a set of 7 features outmatches that of 6 features at even less complex architectures as shown in Sec. 6.2.3.

For electron identification a likelihood discrimination has already been implemented in Atlas. Further investigations of MVA techniques can help to improve the classification performance of photons for both reconstruction and identification. Similar to b-tagging algorithms based on MVA techniques, performances and systematic uncertainties have to be derived for different working points which makes this investigation a long term study.

Because of the large number of free parameters in many machine learning algorithms, it might not be possible to completely avoid empirical studies such as comparing classification performances of different algorithms *after* they have been trained. However, one of the most important goals for the machine learning community in high energy physics should be an even better understanding of analytical properties of different algorithms. That enables faster and more reliable developments as well a decreasing the amount of computing resources for empirical studies.

[1] Also for `BDT`s there is a large number of different configurations available.

Acknowledgements

During the preparation and writing of this thesis I evolved a lot and got many new skills working on my first small contribution to the field of particle physics.

Many thanks, Prof. Dr Arnulf Quadt, for giving me this chance and providing a unique environment in the Göttingen ATLAS group for me. It was a pleasure to work in your group and I can look back to a year filled with enlightening discussions and great achievements.

This thesis also would have never seen any daylight without the large amount of help I got from Josh. I owe you a lot of time you spent to help me solving all kinds of physics and computational problems. Well, and I am not forgetting all the nice and sometimes hilarious conversations we had - not just about physics...

Thank you Boris for making the arrival and the first six month so easy for me. Another big 'thank you' goes to Lisa for your advice for some of the most essential parts of this thesis. Thanks, Royer, for proofreading everything and your very helpful comments.

Allora, Paolo, many gelato thanks for making my days a lot more fun. It was great to be your neighbour!

I want to thank all the people of the institute for the pleasant time, your help, for coffee and lunch breaks and many others things.

Special thanks go to my family for your support over all these years and providing always a safe and comfy home for me!

And most important: Thank you Imke! Your extraordinary patience is admirable. I guess especially the last month was not the most exciting one for you but without your backup I would not have reached all of that.

Bibliography

[1] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22(4)**, 579 (1961)

[2] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19**, 1264 (1967)

[3] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519**, 367 (1968)

[4] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 508 (1964)

[5] G. 't Hooft, M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. **B44**, 189 (1972)

[6] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3(08)**, S08003 (2008)

[7] L. Evans, P. Bryant, *LHC Machine*, JINST **3(08)**, S08001 (2008)

[8] 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(1)**, 1 (2012)

[9] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716(1)**, 30 (2012)

[10] Super-Kamiokande Collaboration, Y. Ashie, et al., *Evidence for an Oscillatory Signature in Atmospheric Neutrino Oscillations*, Phys. Rev. Lett. **93**, 101801 (2004)

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

[12] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, 321 (1964)

[13] G. S. Guralnik, C. R. Hagen, T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**, 585 (1964)

[14] C. S. Wu, et al., *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105**, 1413 (1957)

[15] M. Goldhaber, L. Grodzins, A. W. Sunyar, *Helicity of Neutrinos*, Phys. Rev. **109**, 1015 (1958)

[16] B. Richter, et al., *Discovery of a Narrow Resonance in e^+e^- Annihilation*, Phys. Rev. Lett. **33**, 1406 (1974)

[17] S. C. C. Ting, et al., *Experimental Observation of a Heavy Particle J*, Phys. Rev. Lett. **33**, 1404 (1974)

[18] N. Cabbibo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10**, 531 (1963)

[19] M. Kobayashi, T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, PTP **49(2)**, 652 (1973)

[20] J. H. Christenson, J. W. Cronin, V. L. Fitch, R. Turlay, *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett. **13**, 138 (1964)

[21] P. A. M. Dirac, *The Quantum Theory of the Electron*, Proc. R. Soc. A **117(778)**, 610 (1928)

[22] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3(08)**, S08005 (2008)

[23] LHCb Collaboration, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115**, 072001 (2015)

[24] T. R. Klassen, *QCD (lattice) potential and coupling: How to accurately interpolate between multiloop QCD and the string picture*, Phys. Rev. D **51**, 5130 (1995)

[25] E. Rutherford, *The scattering of α and β particles by matter and the structure of the atom*, Philosophical Magazine **21**, 669 (1911)

[26] M. Breidenbach, et al., *Observed Behavior of Highly Inelastic Electron-Proton Scattering*, Phys. Rev. Lett. **23**, 935 (1969)

[27] *The Durham HepData Project - Online PDF plotting and calculation*, Website (2017), visited on 31.03.2017, URL `http://hepdata.cedar.ac.uk/pdf/pdf3.html`

[28] G. Altarelli, G. Parisi, *Asymptotic freedom in parton language*, Nucl. Phys. B **126(2)**, 298 (1977)

[29] V. N. Gribov, L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15**, 438 (1972), [Yad. Fiz.15,781(1972)]

[30] Y. L. Dokshitzer, *Calculation of structure functions of deep-inelastic scattering and* e^+e^- *annihilation by perturbation theory in quantum chromodynamics*, JETP **46(3)**, 641 (1977)

[31] Collins, J. C., and Soper, D. E. , *The Theorems of Perturbative QCD*, Annual Review of Nuclear and Particle Science **37(1)**, 383 (1987)

[32] B. R. Webber, *Fragmentation and hadronization*, Int. J. Mod. Phys. **A15S1**, 577 (2000)

[33] G. Hanson, et al., *Evidence for Jet Structure in Hadron Production by* e^+e^- *Annihilation*, Phys. Rev. Lett. **35**, 1609 (1975)

[34] J. L. Siegrist, et al., *Hadron production by* e^+e^- *annihilation at center-of-mass energies between 2.6 and 7.8 GeV. I. Total cross section, multiplicities, and inclusive momentum distributions*, Phys. Rev. D **26**, 969 (1982)

[35] S. Bethke, α_S *2016*, Nucl. Part. Phys. Proc. **282-284**, 149 (2017), 19th International Conference in Quantum Chromodynamics

[36] S. Bethke, *QCD studies at LEP*, Phys. Rep. **403-404**, 203 (2004), CERN-the second 25 years

[37] R. Atkin, *Review of jet reconstruction algorithms*, JPCS **645(1)**, 012008 (2015)

[38] M. Cacciari, G. P. Salam, G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04**, 063 (2008)

[39] CMS Collaboration, *Measurement of the top quark mass using proton-proton data at* $\sqrt{s} = 7$ *and 8 TeV*, Phys. Rev. D **93**, 072004 (2016)

[40] F. Abe, et al. (CDF Collaboration), *Observation of Top Quark Production in* $\overline{p}p$ *Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74**, 2626 (1995)

[41] S. Abachi, et al. (D0 Collaboration), *Observation of the Top Quark*, Phys. Rev. Lett. **74**, 2632 (1995)

[42] S. W. Herb, et al., *Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39**, 252 (1977)

[43] M. Jezabek, J. Kühn, *QCD corrections to semileptonic decays of heavy quarks*, Nucl. Phys. B **314(1)**, 1 (1989)

[44] ATLAS Collaboration, *Search for FCNC single top-quark production at with the ATLAS detector*, Phys. Lett. B **712(4-5)**, 351 (2012)

[45] A. Bean, et al. (CLEO), *Improved Upper Limit on Flavor Changing Neutral-Current Decays of the b Quark*, Phys. Rev. D **35**, 3533 (1987)

[46] F. Abe, et al. (CDF), *Search for Flavor-Changing Neutral Current Decays of the Top Quark in $p\bar{p}$ Collisions at* $\sqrt{s} = 1.8$ TeV, Phys. Rev. Lett. **80**, 2525 (1998)

[47] A. Heister, et al. (ALEPH), *Search for Single Top Production in* e^+e^- *Collisions at* $\sqrt{s}$ *up to* 209 GeV, Phys. Lett. B **543**, 173 (2002)

[48] J. Abdallah, et al. (DELPHI), *Search for single top Production via FCNC at LEP at* $\sqrt{s} = 189 - 208$ GeV, Phys. Lett. B **590**, 21 (2004)

[49] P. Achard, et al. (L3), *Search for Single Top Production at LEP*, Phys. Lett. B **549**, 290 (2002)

[50] G. Abbiendi, et al. (OPAL), *Search for Single Top Production at LEP2*, Phys. Lett. B **521**, 181 (2001)

[51] T. Aaltonen, et al. (CDF Collaboration), *Observation of Electroweak Single Top-Quark Production*, Phys. Rev. Lett. **103**, 092002 (2009)

[52] V. M. Abazov, et al. (D0 Collaboration), *Observation of Single Top-Quark Production*, Phys. Rev. Lett. **103**, 092001 (2009)

[53] ATLAS Collaboration, *Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ production cross sections in multilepton final states using 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. **C77(1)**, 40 (2017)

[54] ATLAS Collaboration, *Observation of top-quark pair production in association with a photon and measurement of the $t\bar{t}\gamma$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. **D91(7)**, 072007 (2015)

[55] D. Atwood, A. Soni, *Analysis for magnetic moment and electric dipole moment form factors of the top quark via $e^+e^- \rightarrow tt^-$*, Phys. Rev. D **45**, 2405 (1992)

[56] J. Papavassiliou, C. Parrinello, *Gauge invariant top quark form-factors from e+ e- experiments*, Phys. Rev. **D50**, 3059 (1994)

[57] T. Aaltonen, et al. (CDF Collaboration), *Exclusion of an Exotic Top Quark with* $-4/3$ *Electric Charge Using Soft Lepton Tagging*, Phys. Rev. Lett. **105**, 101801 (2010)

[58] U. Baur, M. Buice, L. H. Orr, *Direct measurement of the top quark charge at hadron colliders*, Phys. Rev. D **64**, 094019 (2001)

[59] ATLAS Collaboration, *Measurement of the top quark charge in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, JHEP **2013(11)**, 31 (2013)

[60] ATLAS Collaboration, *Measurement of the* $t\bar{t}\gamma$ *production cross section in proton–proton collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, Technical Report ATL-COM-PHYS-2016-1750, CERN, Geneva (2016)

[61] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3**, S08004 (2008)

[62] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3(08)**, S08002 (2008)

[63] M. zur Nedden, *The LHC Run 2 ATLAS trigger system: design, performance and plans*, JINST **12(03)**, C03024 (2017)

[64] A. Miucci, *The ATLAS Insertable B-Layer project*, JINST **9(02)**, C02018 (2014)

[65] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, Eur. Phys. J. **C76(12)**, 666 (2016)

[66] J. Mitrevski, *Electron and Photon Reconstruction with the ATLAS Detector*, Nucl. Part. Phys. Proc. **273**, 2539 (2016)

[67] ATLAS Collaboration, *Photon identification in 2015 ATLAS data*, Technical Report ATL-PHYS-PUB-2016-014, CERN, Geneva (2016), (unpublished)

[68] W. S. McCulloch, W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bull. Math. Biophys. **5(4)**, 115 (1943)

[69] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, Psychol. Rev. **65(6)**, 386 (1958)

[70] M. Minsky, S. Papert, *Perceptrons*, MIT Press, Cambridge (1969)

[71] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Learning representations by back-propagating errors*, Nature **323**, 633 (1986)

[72] *The MNIST database of handwritten digits*, Website (2017), visited on 30.03.2017, URL `http://yann.lecun.com/exdb/mnist/`

[73] S. Ioffe, C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, CoRR **abs/1502.03167** (2015)

[74] T. Sjöstrand, et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191**, 159 (2015)

[75] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva (2014), URL `https://cds.cern.ch/record/1966419`

[76] J. Pumplin, et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07**, 012 (2002)

[77] 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**, 079 (2014)

[78] S. Frixione, P. Nason, C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11**, 070 (2007)

[79] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **0902**, 007 (2009)

[80] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, et al., *New parton distributions for collider physics*, Phys.Rev. **D82**, 074024 (2010)

[81] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82**, 074018 (2010)

[82] E. Barberio, B. van Eijk, and Z. Was, *Photos- a universal Monte Carlo for QED radiative corrections in decays*, Comput. Phys. Commun. 66 (1991) 155 (1991)

[83] A. Collaboration (ATLAS Collaboration), *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s}$ = 13 TeV*, Technical Report ATL-PHYS-PUB-2016-015, CERN, Geneva (2016)

[84] A. Collaboration (ATLAS Collaboration), *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, Technical Report ATLAS-CONF-2016-024, CERN, Geneva (2016)

[85] A. Collaboration, *ATLAS electron, photon and muon isolation in Run 2*, Technical Report ATL-COM-PHYS-2017-290, CERN, Geneva (2017)

[86] A. Collaboration (ATLAS), *Muon reconstruction performance of the ATLAS detector in protonproton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. **C76(5)**, 292 (2016)

[87] *Tagging and suppression of pileup jets with the ATLAS detector*, Technical Report ATLAS-CONF-2014-018, CERN, Geneva (2014)

[88] *Expected performance of the ATLAS b-tagging algorithms in Run-2*, Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva (2015)

[89] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s}$ = 13 TeV*, Technical Report ATL-COM-PHYS-2016-407, CERN, Geneva (2016)

[90] F. Chollet, et al., *Keras*, `https://github.com/fchollet/keras` (2015)

[91] *Theano*, `https://github.com/Theano/Theano/` (2017)

[92] D. H. Guest, M. Paganini, M. Kagan, J. W. Smith, M. Lanfermann, *lwtnn/lwtnn: Release for Athena v21* (2017), URL `https://doi.org/10.5281/zenodo.290682`

[93] D. P. Kingma, J. Ba, *Adam: A Method for Stochastic Optimization*, CoRR **abs/1412.6980** (2014)

A. Appendix

A.1. Configuration of NN training in the root2kerasPipeline

parameter name	description	required(✓)/ default
model_name	The name of function returning the desired architecture. Here it has to be decided which loss function and which optimiser should be used.	✓
hdf5_files	A list of paths pointing to HDF5 files the training and testing data is extracted from.	✓
features	A list containing the names of features. Those must be present in the HDF5 files.	✓
targets	A list containing the names of targets. Those must be present in the HDF5 files.	✓
sample_weight	Name of the column in the HDF5 files which contains the weights for each sample. These weights are only applied during training.	✓
sample_fraction	Only this fraction of samples contained in the HDF5 files is used for training and testing.	1.0
split_train_test	Use this option to decide whether samples should be split into a training and a test set. If it is False, no test is performed.	True
scale_samples	to decide whether samples are scaled individually to have mean 0 and standard deviation of 1	True
train_size	This is the ratio of samples used for training. The complementary set is used for testing. A value of 1 corresponds to split_train_test=False	0.8
nb_epoch	number of epochs	15
batch_size	number of samples per batch	10000
weight_sg	Weight assigned to all signal samples in training. This overwrites signal sample weights read from the HDF5 files.	None
weight_bkg	Weight assigned to all background samples in training. This overwrites background sample weights read from the HDF5 files.	None
do_cv	Whether or not to start a cross-validation run.	False
n_splits	Number of folds used in a cross-validation run.	5

Table A.1.: Summary of parameters which can be set in the training and testing of NNs in the root2kerasPipeline.

A.2. 12 % and 100 % splits of training single photon sample

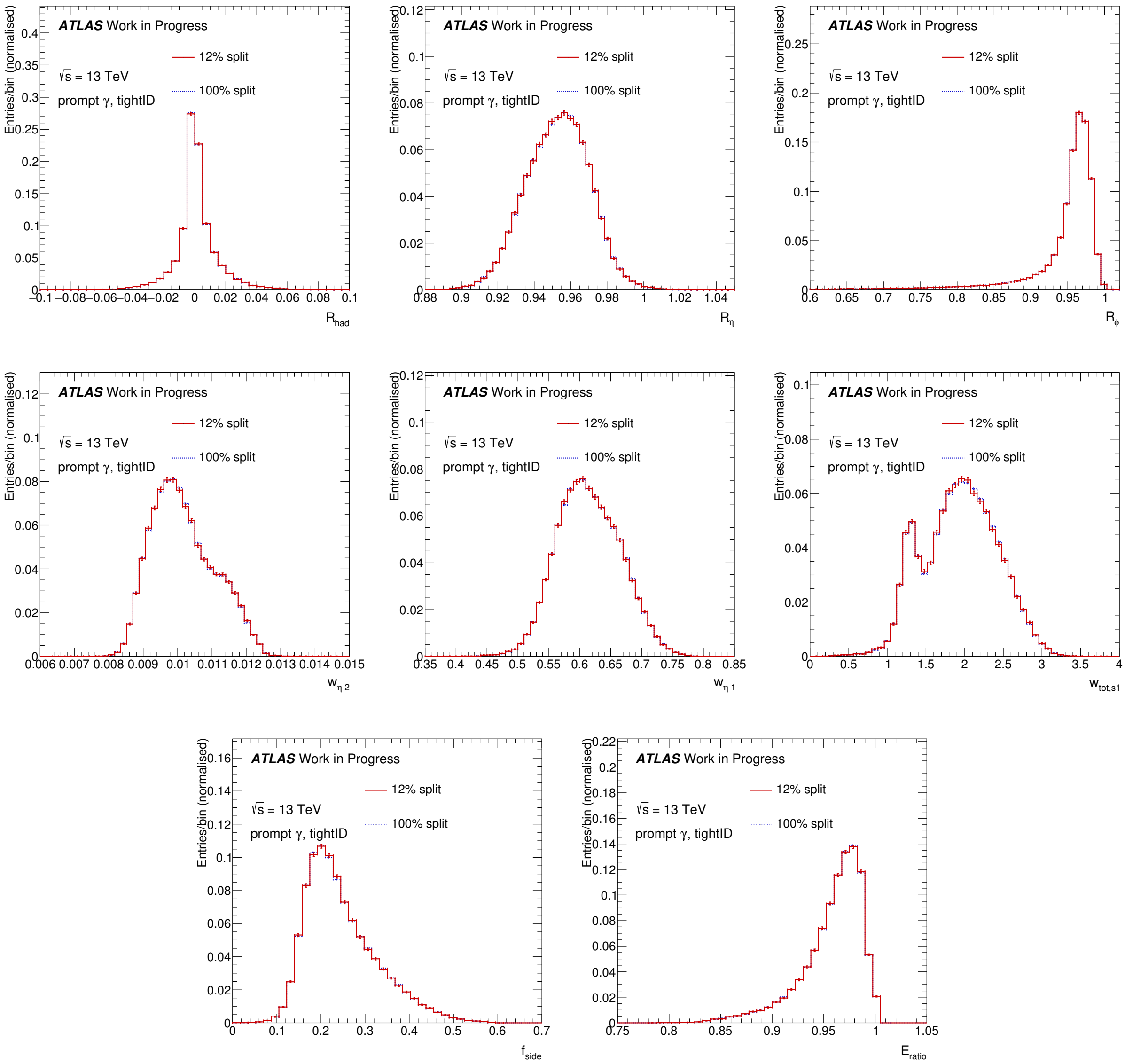


Figure A.1.: Shower shape distributions of prompt photons comparing the distributions of drawn 12 % randomly drawn to the distributions of the full single photon MC sample.

A.3. 20 % and 80 % splits of training MC samples

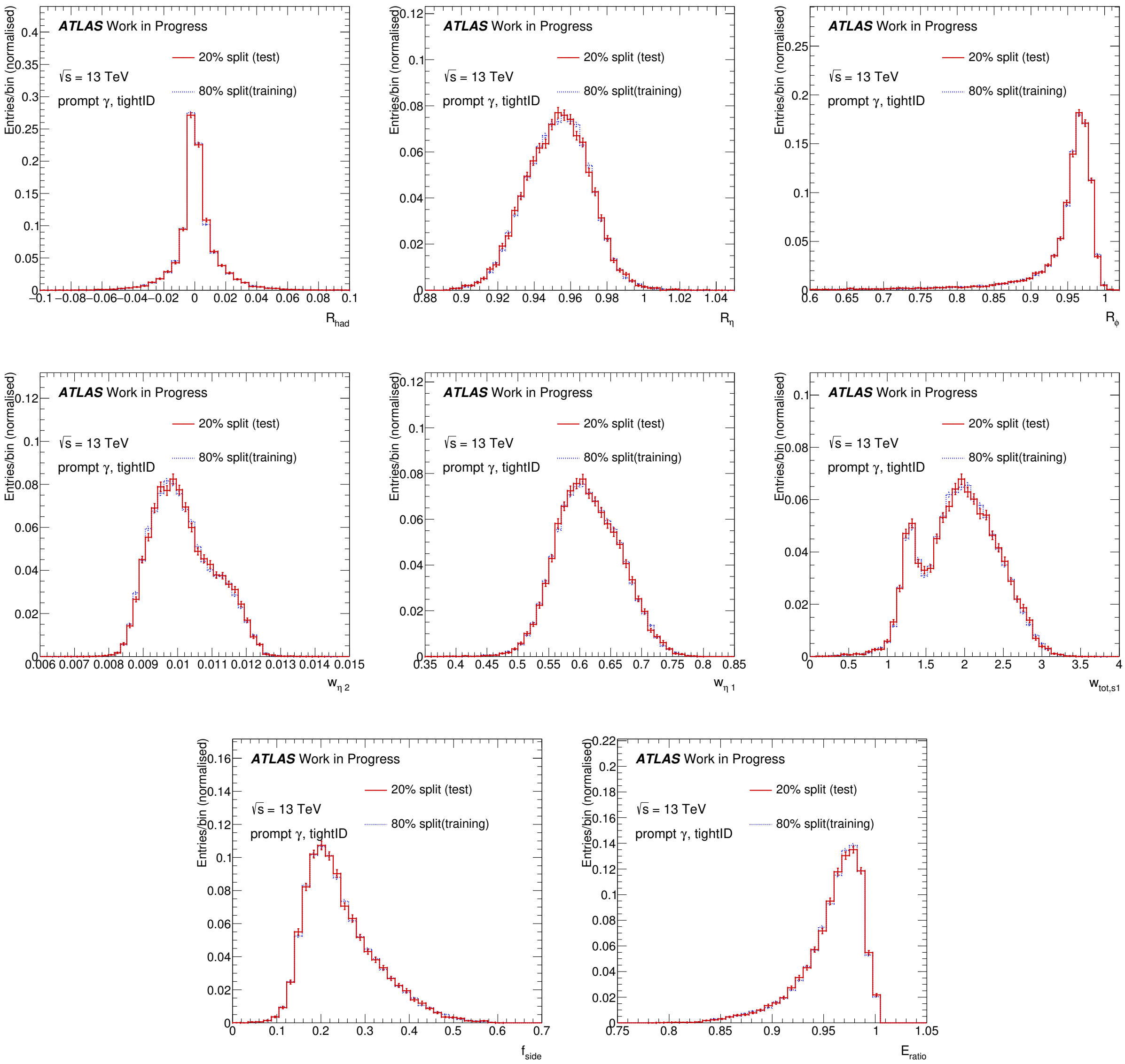


Figure A.2.: Shower shape distributions of prompt photons comparing the respective distributions drawn from orthogonal 20 % and 80 % split of the QCD-Compton MC sample.

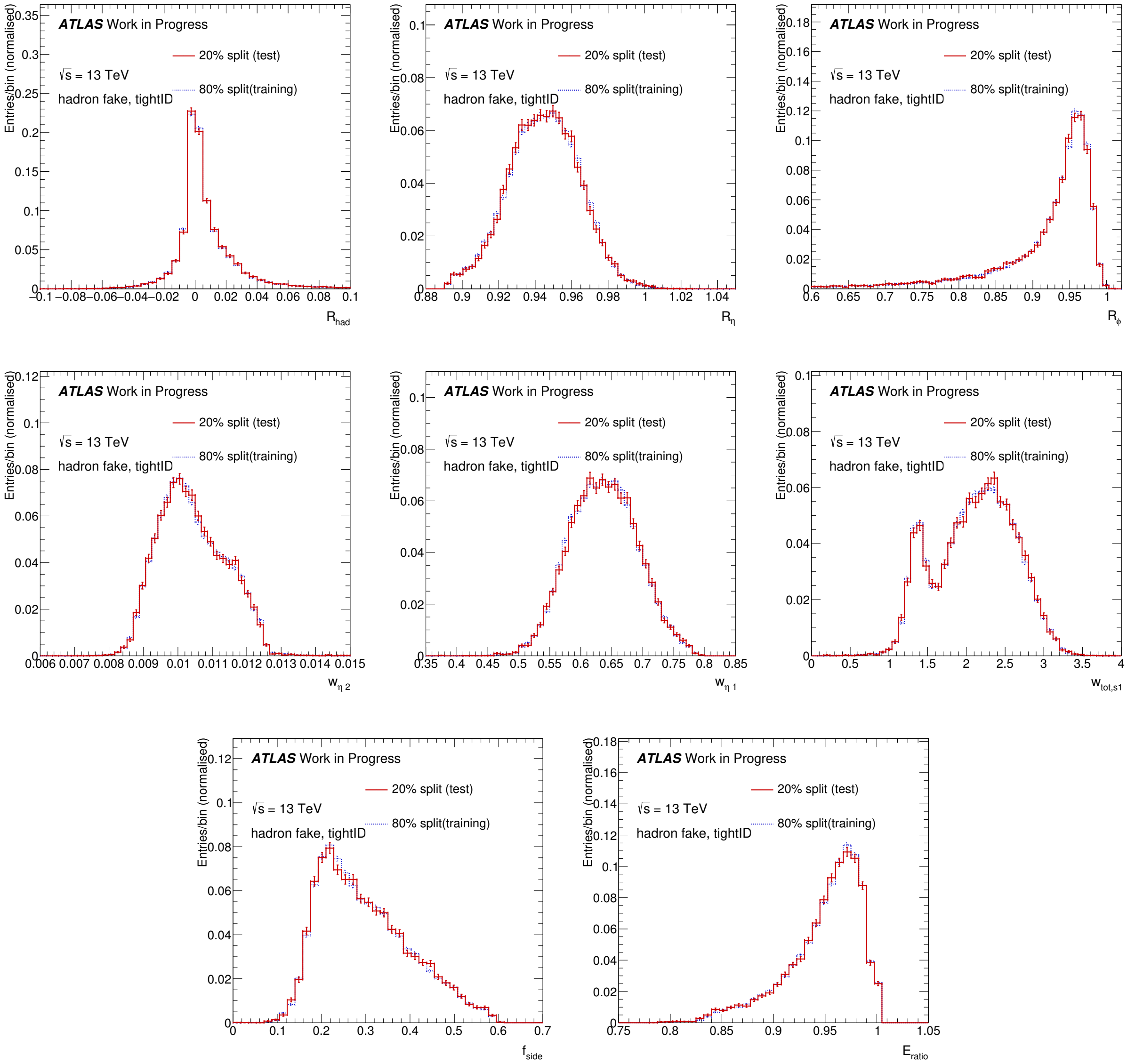


Figure A.3.: Shower shape distributions of hadron fakes comparing the respective distributions drawn from orthogonal 20 % and 80 % split of the di-jet MC sample.

A.4. Shower shapes of prompt photons for different jet multiplicities

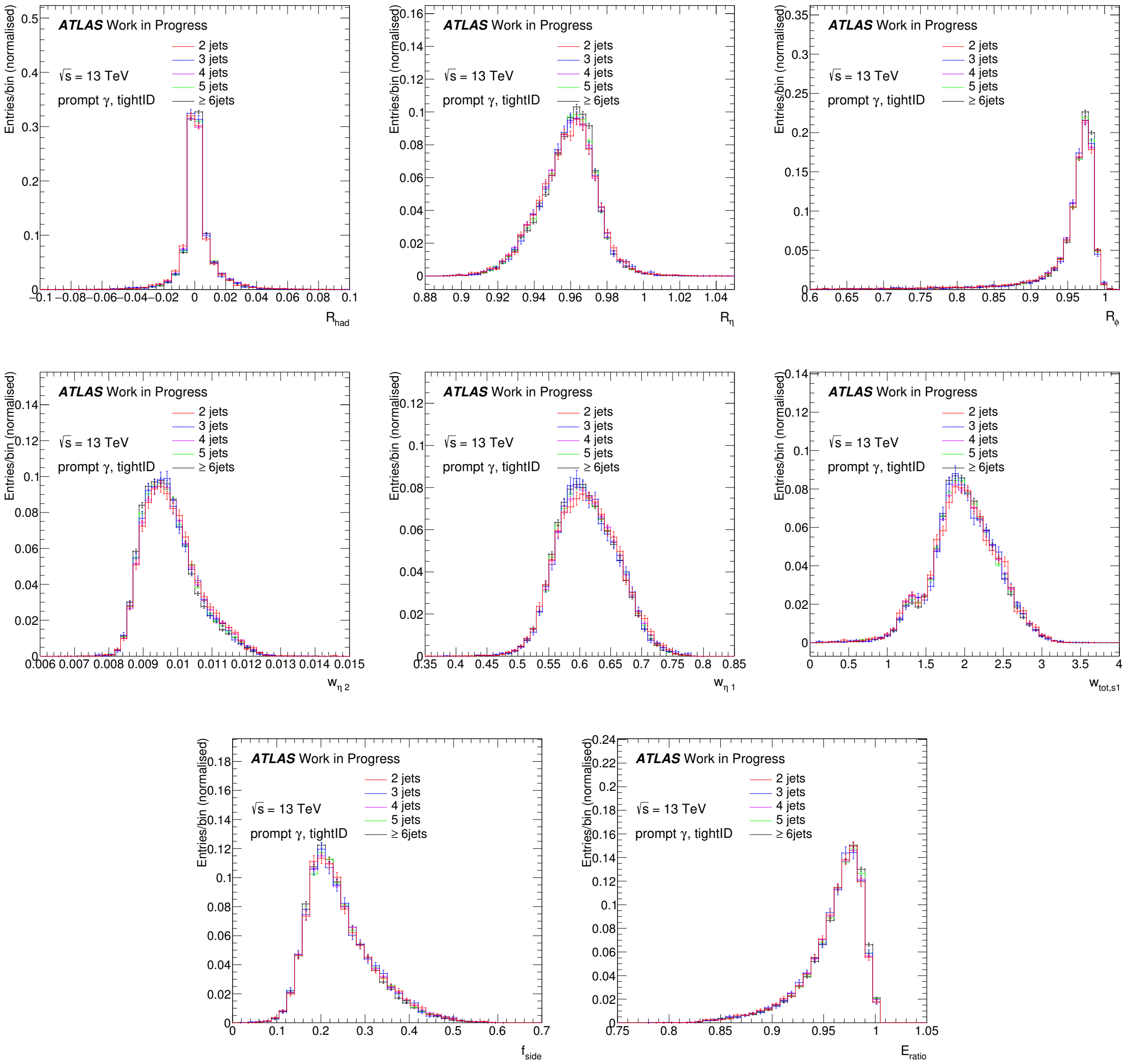


Figure A.4.: Shower shape distribution of prompt photons in 5 different bins of jet multiplicity.

A.5. Shower shapes of prompt photons in ϕ bins

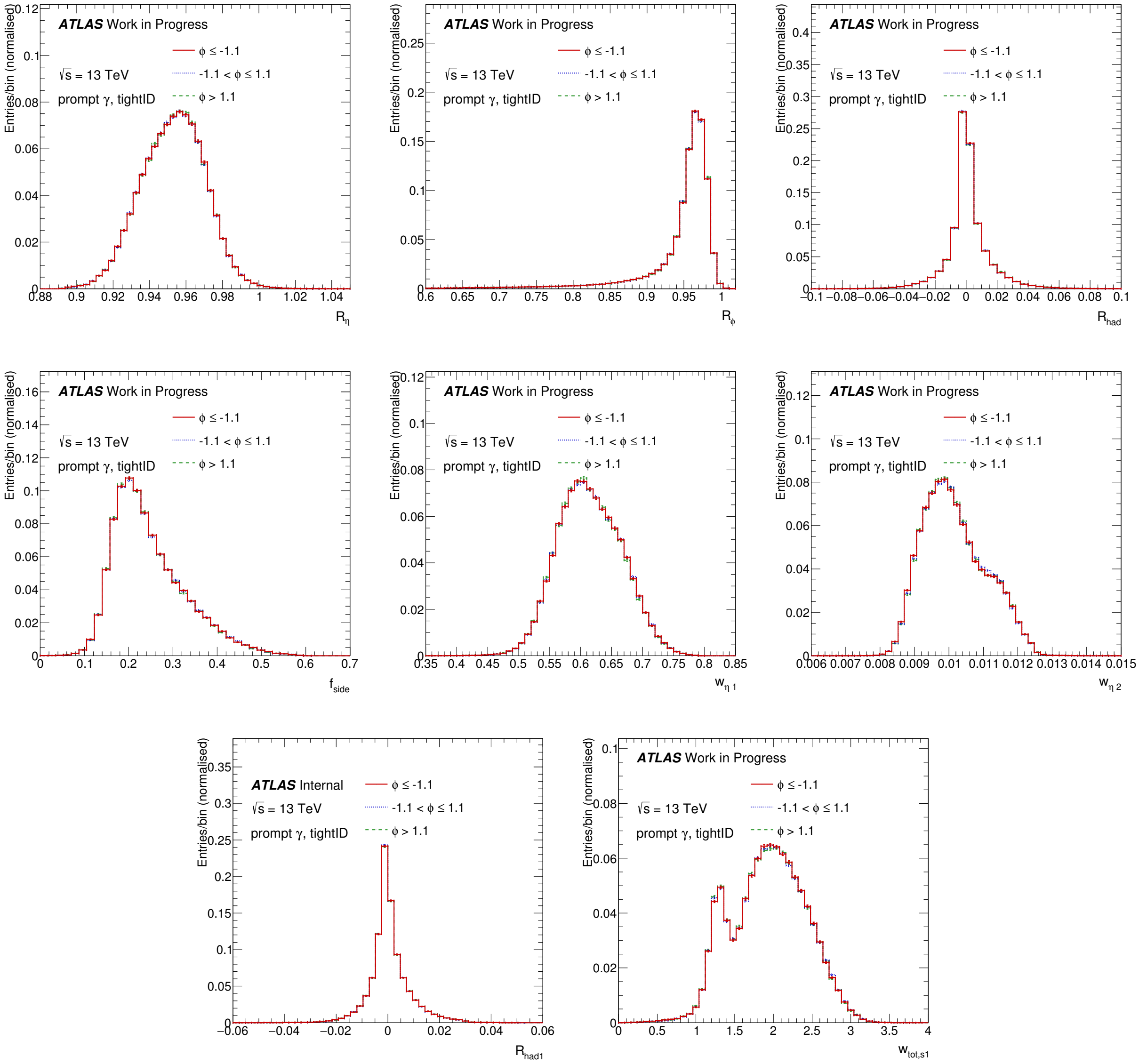
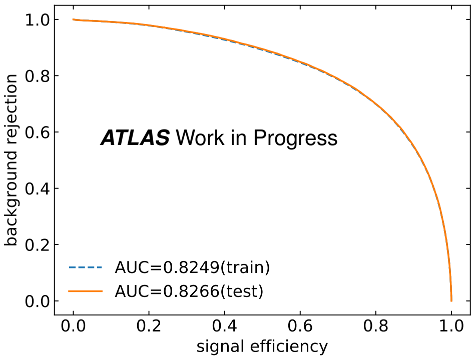
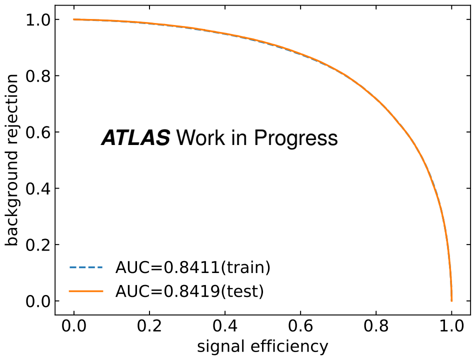
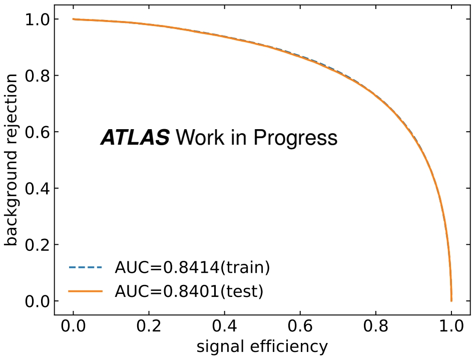


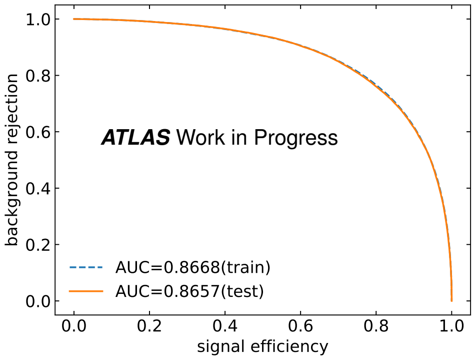
Figure A.5.: Shower shape distribution in 3 bins of ϕ.

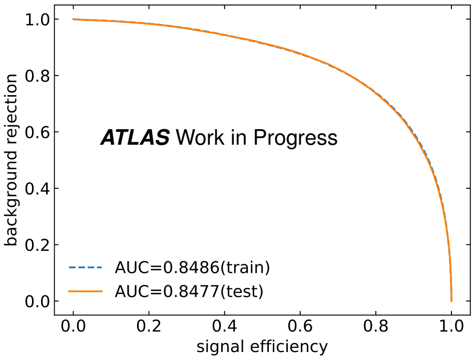
A.6. ROC curves of tested PPT configurations

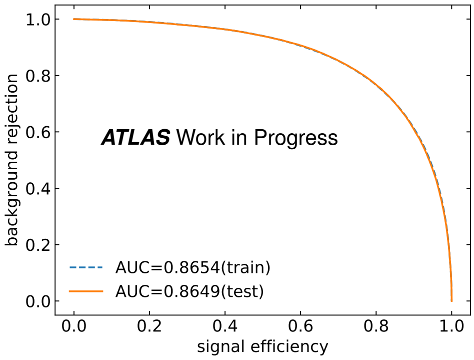


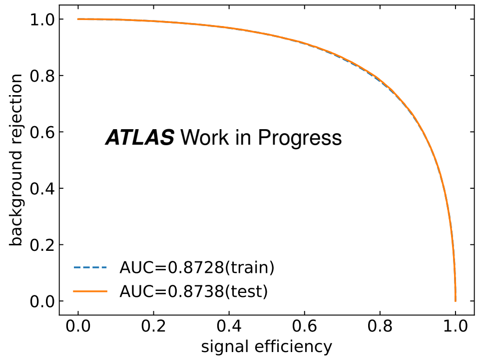




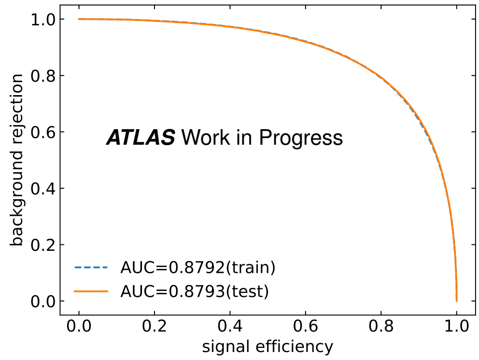






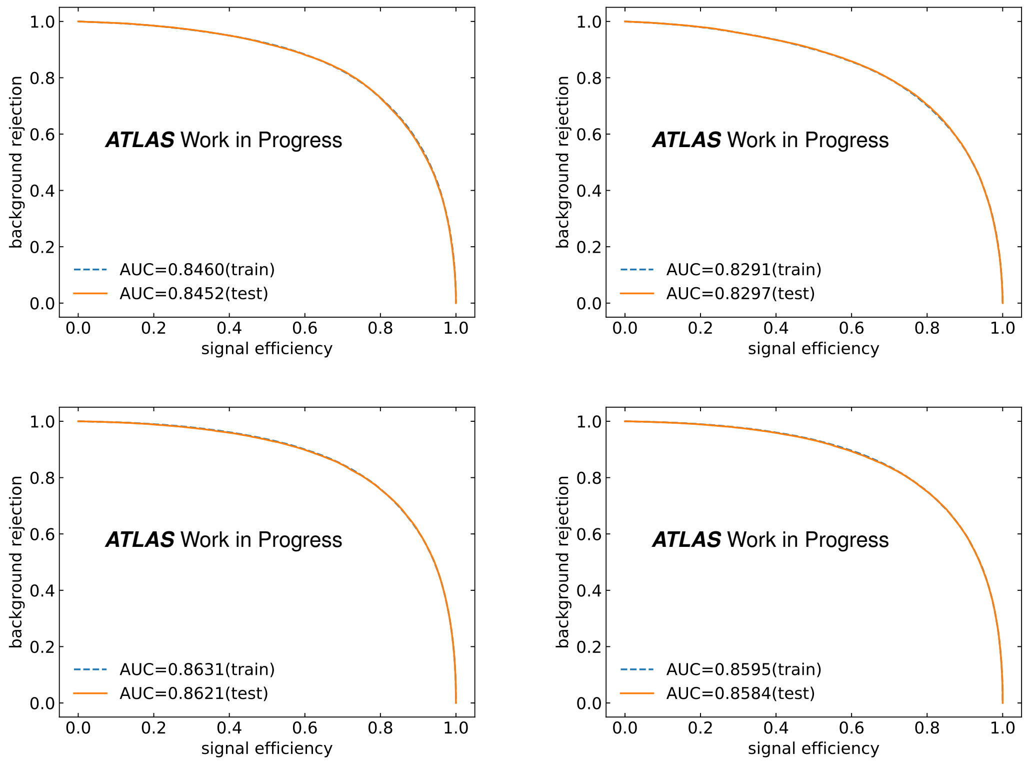


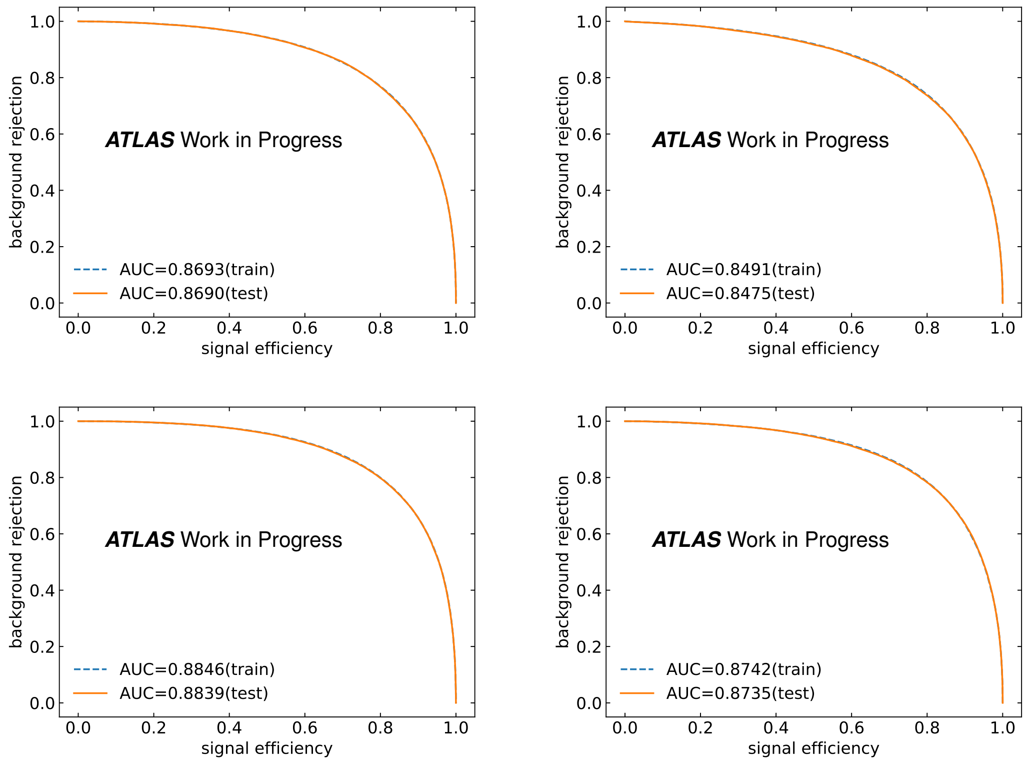
(a) Architecture 1.



(b) Architecture 2.

Figure A.7.: Two PPT architectures (see Tab. 6.3), starting with features R_η, R_ϕ, f_{side} and $w_{\eta 1}$. In the following rows R_{had}, $w_{\eta 2}$ and E_{ratio} are added successively.





(a) Architecture 3. *(b)* Architecture 4.

Figure A.9.: Two PPT architectures (see Tab. 6.3), starting with features R_η, R_ϕ, f_{side} and $w_{\eta 1}$. In the following rows R_{had}, $w_{\eta 2}$ and E_{ratio} are added successively.

A.7. MCTruthClassifier

To distinguish between prompt photons and hadron fakes, the particle's truth information and those provided by `MCTruthClassifier` are used. The latter one works as follows: If a photon candidate is reconstructed in simulation, all truth particles are extrapolated to the barycentre of the reconstructed candidate in the second layer of the EMCal. If available, the truth photon with highest p_T in a cone of $\eta \times \phi = 0.025 \times 0.050$ around the candidate is matched. If no truth photon is available, the next highest p_T truth particle is matched. If no other particle can be found in that radius and if also no particle can be found in a larger cone radius, particles produced in detector simulation are taken into account.

A.8. Monte Carlo derivations used for NN training

single photon
mc15_13TeV.423099.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP8_17.merge.DAOD_STDM2.e4453_s2726_r7725_r7676_p2669/
mc15_13TeV.423100.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP17_35.merge.DAOD_STDM2.e3791_s2608_s2183_r7725_r7676_p2669/
mc15_13TeV.423101.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP35_50.merge.DAOD_STDM2.e3904_s2608_s2183_r7725_r7676_p2669/
mc15_13TeV.423102.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP50_70.merge.DAOD_STDM2.e3791_s2608_s2183_r7725_r7676_p2669/
mc15_13TeV.423103.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP70_140.merge.DAOD_STDM2.e3791_s2608_s2183_r7725_r7676_p2669/
mc15_13TeV.423104.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP140_280.merge.DAOD_STDM2.e3791_s2608_s2183_r7725_r7676_p2669/
mc15_13TeV.423105.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP280_500.merge.DAOD_STDM2.e3791_s2608_s2183_r7725_r7676_p2669/
mc15_13TeV.423106.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP500_800.merge.DAOD_STDM2.e3791_s2608_s2183_r7725_r7676_p2669/
mc15_13TeV.423107.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP800_1000.merge.DAOD_STDM2.e4453_s2726_r7725_r7676_p2669/
mc15_13TeV.423108.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP1000_1500.merge.DAOD_STDM2.e4453_s2726_r7725_r7676_p2669/
mc15_13TeV.423109.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP1500_2000.merge.DAOD_STDM2.e4453_s2726_r7725_r7676_p2669/
mc15_13TeV.423110.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP2000_2500.merge.DAOD_STDM2.e4453_s2726_r7725_r7676_p2669/
mc15_13TeV.423111.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP2500_3000.merge.DAOD_STDM2.e4453_s2726_r7772_r7676_p2669/
mc15_13TeV.423112.Pythia8EvtGen_A14NNPDF23LO_gammajet_DP3000_inf.merge.DAOD_STDM2.e4453_s2726_r7772_r7676_p2669/
di-jet
mc15_13TeV.423300.Pythia8EvtGen_A14NNPDF23LO_perf_JF17.merge.DAOD_STDM2.e3848_s2608_s2183_r7725_r7676_p2666/
mc15_13TeV.423301.Pythia8EvtGen_A14NNPDF23LO_perf_JF23.merge.DAOD_STDM2.e3848_s2608_s2183_r7725_r7676_p2666/
mc15_13TeV.423302.Pythia8EvtGen_A14NNPDF23LO_perf_JF35.merge.DAOD_STDM2.e3848_s2608_s2183_r7725_r7676_p2666/
mc15_13TeV.423303.Pythia8EvtGen_A14NNPDF23LO_perf_JF50.merge.DAOD_STDM2.e3848_s2608_s2183_r7725_r7676_p2666/

Table A.2.: Summary of derivations used for training and evaluation of PPT.

Erklärung nach §17(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 1. Oktober 2017

(Benedikt Völkel)