



University of  
Zurich <sup>UZH</sup>

# Bachelor Thesis

---

## Algorithms to identify high-energy $B$ hadrons via their hit multiplicity increase through pixel detection layers

**Author:**

Manuel Sommerhalder

**Supervised by:**

Prof. Dr. Benjamin Kilminster

Thea Arrestad

Dr. Yuta Takahashi

**Handed in:**

27.07.2018

**Abstract:**

Track-based  $b$  tagging algorithms are shown to suffer significant efficiency loss for jets with a transverse momentum above 1 TeV due to track reconstruction inefficiencies. In the same high-momentum region, a simple hit counting approach performs similarly or even exceeds the performance of the most commonly used  $b$  tagging algorithm in the CMS experiment, the Combined Secondary Vertex Tagger (CSV). Different implementations of cut-based discriminants, as well as neural network models are shown to yield a relative efficiency gain of up to 112% for jets with transverse momenta above 1.2 TeV for a mistag rate of 10% when used in addition to the CSV.



# Contents

1	Introduction	3
1.1	Motivation . . . . .	3
1.2	LHC, CMS & the pixel detector . . . . .	3
1.3	$b$ quarks . . . . .	5
1.4	CSV . . . . .	5
2	Preparation of data	7
2.1	Simulated samples . . . . .	7
2.2	Signal definition . . . . .	7
2.3	Pixel hit cluster matching . . . . .	8
3	Data characteristics	9
3.1	Jet properties . . . . .	9
3.2	Pixel hit distributions . . . . .	10
4	Results	12
4.1	Cut-based discriminants . . . . .	12
4.1.1	Discriminant and cone optimizations . . . . .	12
4.1.2	$p_T$ dependent cuts . . . . .	18
4.1.3	Stability with respect to PU . . . . .	22
4.2	Artificial neural networks . . . . .	24
4.2.1	Choice of model . . . . .	24
4.2.2	Comparison to cut-based approach and CSV . . . . .	28
4.2.3	Stability with respect to pileup . . . . .	30
5	Summary	34
6	Appendix	35
6.1	General concepts . . . . .	35
6.2	Hadrons vs quarks . . . . .	35
6.3	Determination of optimum cone sizes . . . . .	36
	References	40

# 1 Introduction

## 1.1 Motivation

At the CMS experiment at CERN's Large Hadron Collider, the successful tagging of high energy  $b$  jets (sprays of hadrons resulting from the decay of a  $B$  hadron) is important for studies of the standard model itself as well as physics beyond the standard model. Examples of phenomena that lead to such  $b$  quarks are:

- Higgs pair production and decay via  $HH \rightarrow b\bar{b}b\bar{b}$ , sensitive to Higgs trilinear couplings [1]
- graviton and radion decays to heavy fermions and bosons in warped extra dimension models [2]
- third-generation superpartners in supersymmetry [3]

Due to the long lifetime of the  $B$  hadron, it can travel up to several centimetres before decaying if its energy is high enough. Thus, the  $B$  hadron can traverse the first layer of pixel detector, resulting in a degradation in track reconstruction efficiency. This in turn results in a drop in  $b$  tagging efficiency for track-based taggers above a jet  $p_T$  of around 500 GeV [4]. This study builds on previous work [5] where the possibility of improving the efficiency of high energy  $b$  tagging by counting hits in the pixel detector instead of relying on a track-based algorithm is investigated.

The decay of highly boosted  $B$  hadrons into several daughter particles is expected to increase the number of hits inside a small angular region relative to the jet axis in the subsequent detector layers [5]. This increase in hit multiplicity is not expected for other hadrons, which decay close to the interaction point. Thus, the hit-based tagging algorithms discussed in this study aim at identifying  $B$  hadrons via increase in hit multiplicity increase through subsequent layers.

## 1.2 LHC, CMS & the pixel detector

The Large Hadron Collider (LHC) at CERN near Geneva is the largest and most powerful particle collider in the world with a circumference of 27 km [6]. It is primarily used for proton-proton collisions with a center-of-mass energy up to  $\sqrt{s} = 13$  TeV. It has four crossing points where the main detectors are located: ATLAS, CMS, LHCb and ALICE.

The Compact Muon Solenoid (CMS) detector is a large 14,000-tonne general-purpose

detector described in detail elsewhere [7]. With a diameter of 15.0 m and an overall length of 28.7 m it is relatively compact compared to other detectors of its type. The term "Muon" arises from the fact that it was designed to measure the muon momenta with high precision. This is achieved through a powerful superconducting solenoid magnet. The detector has a cylindrical onion-like architecture, consisting of five distinct concentric layers, as can be seen in Fig. 1.

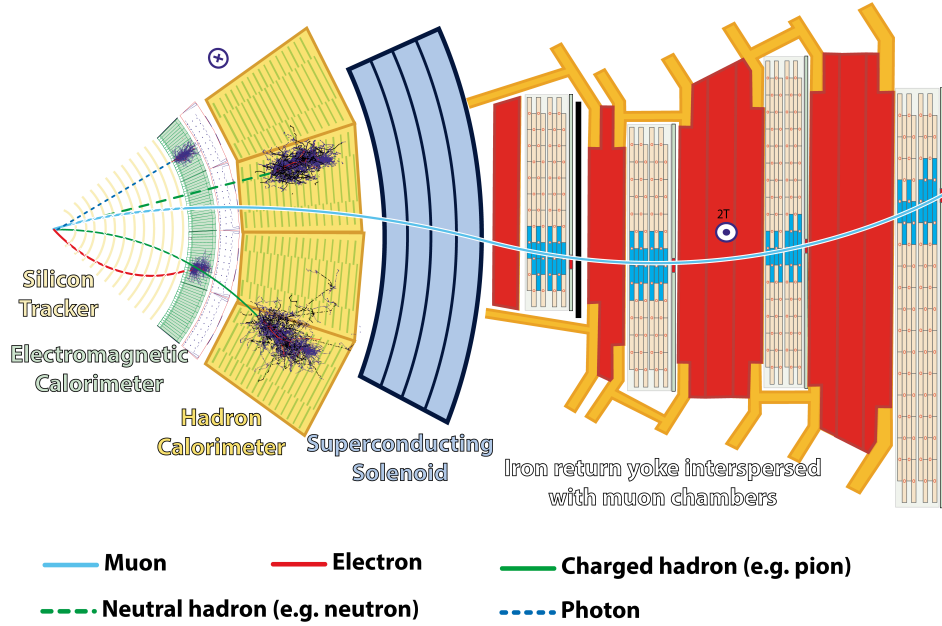


Figure 1: Transverse section of the CMS detector [8]

The closest layer to the interaction point is the tracker. Its silicon detector modules are designed to measure the position of traversing charged particles. The particle momenta can be deduced from the curvature of the trajectory due to the magnetic field. The innermost part of the tracker layer is the pixel detector, which will be discussed in more detail later, followed by microstrip detectors. The second layer is the electromagnetic calorimeter, which measures the energy of electrons, positrons and photons by stopping them completely. The third layer is a hadronic calorimeter, which measures the energy of charged and neutral hadrons. The fourth layer is the 3.8 T solenoid magnet. It consists of refrigerated superconducting niobium-titanium coils that allow a precise measurement of the particle momenta. The fifth and last layer contains the iron return yoke and muon detection chambers.

The CMS silicon pixel detector [9], which was upgraded in 2017, has an active surface of around 2 m<sup>2</sup> with a total amount of 124 million pixels. The pixels each have a size of 150 x 100  $\mu\text{m}$  and the thickness of the sensitive volume is 285  $\mu\text{m}$ . The barrel, which is the lateral surface of the detector, consists of four layers at radii from the interaction point of 2.9, 6.8, 10.9, 16.0 cm. The forward region along both directions parallel to the

beam axis, also called the endcaps, are each covered by three disks that lie at distances of 3.2, 3.9 and 4.8 cm from the interaction point. Each disk is composed of two rings of modules with average radii of 7.8 and 12.8 cm.

### 1.3 $b$ quarks

The bottom ( $b$ ) quarks are, with a mass of around 4.2 GeV, the most massive down-type quark flavour [10]. Due to color confinement in quantum chromodynamics (QCD), they cannot be observed in Nature as free particles, but form bound states, the so-called hadrons, with other quarks [11]. Because of their large mass compared to the scale of the strong force, their decay can be approximated as free  $b$  quarks whose ground states decay via the weak interaction, mostly into a charm ( $c$ ) quark and an off-shell  $W^-$  boson that subsequently decays either hadronically (into a quark and an antiquark) or leptonically (into a lepton and its corresponding anti-neutrino) [12]. An important property of  $B$  hadrons is their lifetime, which is around  $\tau = 1.5$  ps [10]. This allows decay lengths of a few millimeters, up to several centimeters for highly energetic hadrons.

A hypothesized source of high momentum  $b$  quarks is the decay of the so called  $Z'$  boson. A large number of models exists that predict such a boson, which is assumed to be neutral, colorless, self-adjoint and more massive than the standard model  $Z$  boson [13]. This  $Z'$  boson can decay into two  $b$  quarks as shown in the Feynman diagram in Fig. 2. Because of the large mass of the  $Z'$  boson,  $b$  quarks resulting from its decay can have very high  $p_T$ .

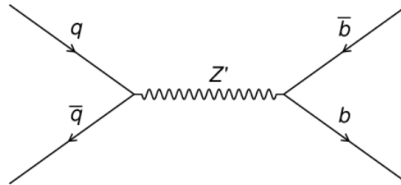


Figure 2: Decay of the hypothetical  $Z'$  boson to  $b\bar{b}$  [14]

### 1.4 CSV

The most commonly used algorithm for  $b$  tagging in CMS is the Combined Secondary Vertex (CSV) Tagger [4]. This algorithm relies on tracks, which are the trajectories of electrically charged particles, reconstructed from hits in the silicon detector using an iterative procedure with a Kalman filter. Several constraints are imposed on the tracks in order to be taken into account by CSV, e.g. they are required to have  $p_T > 1$  GeV and leave at least one hit in the pixel layers of the tracker detector. From these tracks, the secondary vertices are reconstructed using the algorithms AVR and IVF. Each jet is then associated with one of the following three categories: RecoVertex, PseudoVertex or NoVertex, according to whether one or more secondary vertices could be found or not.

The label PseudoVertex belongs to jets where no secondary vertex could be found but a set of at least two tracks were found to have a 2D impact parameter significance above two. The reconstructed secondary vertex data and the variables shown in Fig. 3 are put into a neural network that outputs the estimated probability of each jet to originate from a  $B$  hadron.

Input variable	Run 1 CSV	CSVv2
SV 2D flight distance significance	x	x
Number of SV	—	x
Track $\eta_{rel}$	x	x
Corrected SV mass	x	x
Number of tracks from SV	x	x
SV energy ratio	x	x
$\Delta R(SV, jet)$	—	x
3D IP significance of the first four tracks	x	x
Track $p_{T,rel}$	—	x
$\Delta R(track, jet)$	—	x
Track $p_{T,rel}$ ratio	—	x
Track distance	—	x
Track decay length	—	x
Summed tracks $E_T$ ratio	—	x
$\Delta R(summed tracks, jet)$	—	x
First track 2D IP significance above c threshold	—	x
Number of selected tracks	—	x
Jet $p_T$	—	x
Jet $\eta$	—	x

Figure 3: Input variables used for the two versions of the CSV algorithm where the symbol "x" ("—") denotes that the variable is (not) used. [4]

The central problem is that track-based algorithms tend to lose efficiency at high  $p_T$  due to the following :

1. For high energy jets, more tracks are collimated into a small angle, resulting in a higher hit density and thus more ambiguous association of hits with tracks.
2. Due to their long lifetime, an increasing fraction of high momentum  $B$  hadrons decays after the first layer of the pixel detector. This can simply lead to a smaller number of hits available for track reconstruction, but can in the worst case lead the tracking algorithm to misinterpret a hit from the more densely populated inner layer as part of the track.



## 2 Preparation of data

### 2.1 Simulated samples

The following study of hit-based taggers is performed using two different pileup scenarios. First, the tagging performance using simulated data with only one primary vertex per event is evaluated.

In a realistic setting at CMS, multiple collisions will be measured at once, resulting in a higher number of primary vertices (PV) per event. The second step of this study therefore addresses this phenomenon, commonly called pileup (PU), by testing the tagging performance as a function of the number of primary vertices. The pileup profile used here corresponds to the expected amount of pileup in 2017.

The signal samples used in this study were generated privately in a simulation of  $Z'$  decaying to  $b\bar{b}$  with a  $Z'$  mass of  $M = 2$  TeV and  $M = 4$  TeV using the PYTHIA 8 Monte Carlo generator [15]. For the background, officially produced QCD simulations were used. The sample without pileup is:

*/QCD\_Pt-15to7000\_TuneCUETP8M1\_Flat\_13TeV\_pythia8/RunIISummer17DRStdMix-NZSNoPU\_92X\_upgrade2017\_realistic\_v10-v3/GEN-SIM-RECO.*

The sample with pileup:

*/QCD\_Pt-15to7000\_TuneCUETP8M1\_Flat\_13TeV\_pythia8/RunIISummer17DRPremix-NZS\_92X\_upgrade2017\_realistic\_v10-v3/GEN-SIM-RECO.*

### 2.2 Signal definition

The first part of the study consists of distinguishing the signal events (the actual  $b$  jets) from the background (non  $b$  jets). The signal events were taken to be those jets in the signal sample whose axis lies inside a cone of size  $\Delta R < 0.3$  around the momentum 4-vector of an outgoing  $B$  meson or baryon. Similarly, the background jets were defined as the jets in the background sample that could be matched to the momentum 4-vector of a non  $b$  quark from the hardest subprocess. The reason for using the outgoing hadrons for the selection of  $b$  jets is their relatively long lifetime. The idea behind instead using the hard process quarks in the case of the background was their easier implementation. This reasoning is further expanded in Section 6.2.

A further constraint, which was imposed in both signal and background, is a particle transverse momentum threshold of at least 350 GeV, since the focus of the analysis lies on particles with a momentum high enough to traverse the first pixel detector layer. When an excited particle emits a photon during de-excitation, the resulting state is treated as a distinct particle in the simulation. Thus, special care had to be taken in order to count such particles only once.

## 2.3 Pixel hit cluster matching

Pixel clusters, which are for simplicity also called "hits" in this study, are formed from adjacent pixels in the silicon pixel detector whose charge exceeds a pre-defined readout threshold due to traversing particles. A more thorough description of how the shape and position of pixel clusters are determined can be found in Ref. [16].

Since the simulated samples only contained spatial information about the clusters, the geometrical matching of fired clusters to jets had to be performed explicitly. The jet axis was manually reconstructed by using the primary vertex of the simulated event as the point of origin and the momentum 3-vector yielded the further direction of travel. This axis was then used as the axis of a cone with a size of fixed  $\Delta R$  inside which the fired clusters of the detector modules must lie in order to be matched to the corresponding jet. Originally, the same procedure was performed with respect to the particle trajectory, as determined from the Monte Carlo generator, instead of the jet axis. The jet axis was then observed to yield similar results and was henceforth used since it corresponds to the actual observable in the detector, contrary to the particle trajectory. The procedure of matching clusters to a reconstructed jet axis is illustrated in Figs. 4a and 4b on a small selection of sample events.

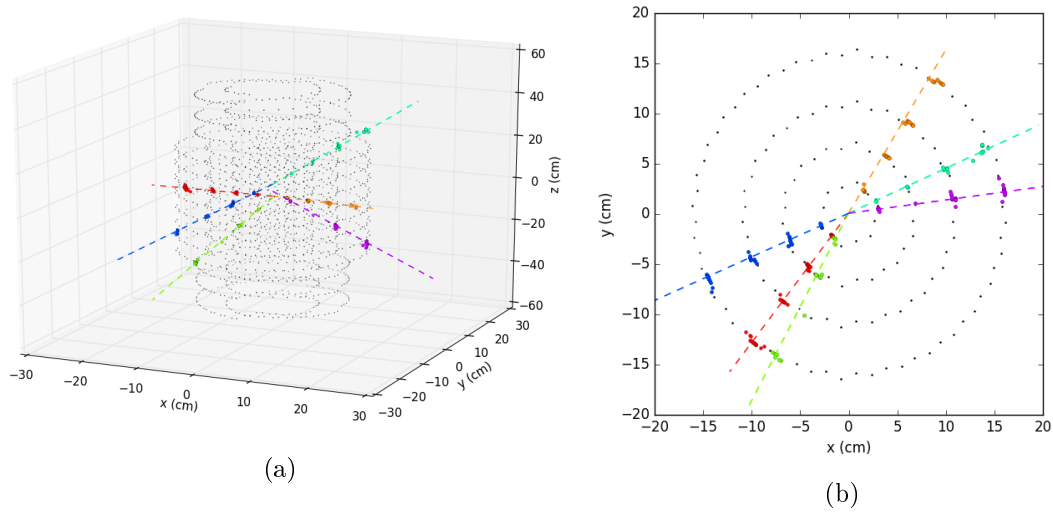


Figure 4: Sample events illustrating the matching of fired clusters (colored dots) to reconstructed jet axes (dashed lines). The black dots correspond to the positions of detector modules. The three-dimensional plot (a) is also shown projected onto a two-dimensional plane (b).



### 3 Data characteristics

#### 3.1 Jet properties

The jet  $p_T$  spectrum for signal and background jets is shown in Fig. 5a. The background jet  $p_T$  distribution yields a truncation around 200 GeV because the particle matched to the jet is required to pass a  $p_T$  threshold of 350 GeV. While the peaks of the three distributions are clearly distinct from each other, care is taken such that methods for distinguishing signal and background are not biased according to jet  $p_T$ .

In order to check the assumption that a non-negligible number of  $B$  hadrons decay after the first detector layer, the distances  $R$  from the  $z$  axis to the particle decay vertices in the  $x$ - $y$  plane were plotted in a normalized histogram in Fig. 5b. These distributions show clearly that a large fraction of  $B$  hadrons decay only after the first detector layer, located at  $R = 2.9$  cm. A small fraction of  $B$  hadrons even decays after a later detector layer or traverses the pixel detector completely. This is not true for the background, where  $R$  does not exceed 1 mm because the particles, to which these jets are matched, are the hardest subprocess quarks and gluons. However, because of mis-reconstruction effects, the background will still yield jets that pass standard  $b$  tagging selection, and such jets are referred to as mis-tagged.

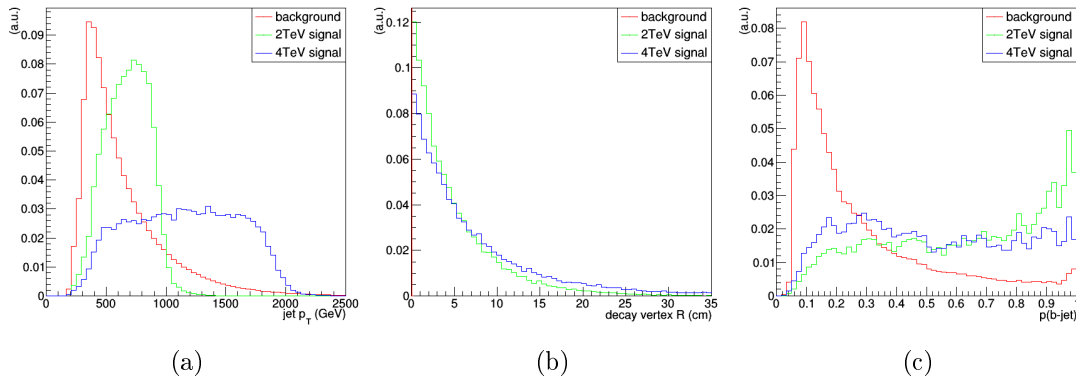


Figure 5: Normalized histograms showing the distribution of jet  $p_T$  (a), the decay vertex  $R$  (b) and the CSV  $b$  tag value (c) of the three data samples.

The CSV  $b$  tagging discriminant is shown in Fig. 5c. The  $b$  tag value lies within a range from 0 to 1 and corresponds to the probability that a jet is a  $b$  jet. While there is a desirably large and distinct peak in the background distribution at low  $b$  tag values, the 2 TeV signal peak at high values is more diffuse with a large tail to the left. The 4 TeV signal, on the other hand, is almost uniformly distributed throughout most of the spectrum. This can be seen as an indication that the current CSV  $b$  tagging algorithm

loses performance for the higher  $p_T$  jets of the 4 TeV signal sample.

### 3.2 Pixel hit distributions

In order to capture whether an increase in hit multiplicity between consecutive detector layers is observable, we look at the total number of hits in each layer. Only hits inside a cone of size  $\Delta R < 0.1$  around a matched jet were considered. The lower plots of Figs. 6a to 6c were made using only events where the particle had a transverse momentum of at least 1 TeV, while the full jet  $p_T$  spectrum was counted in the plots above.

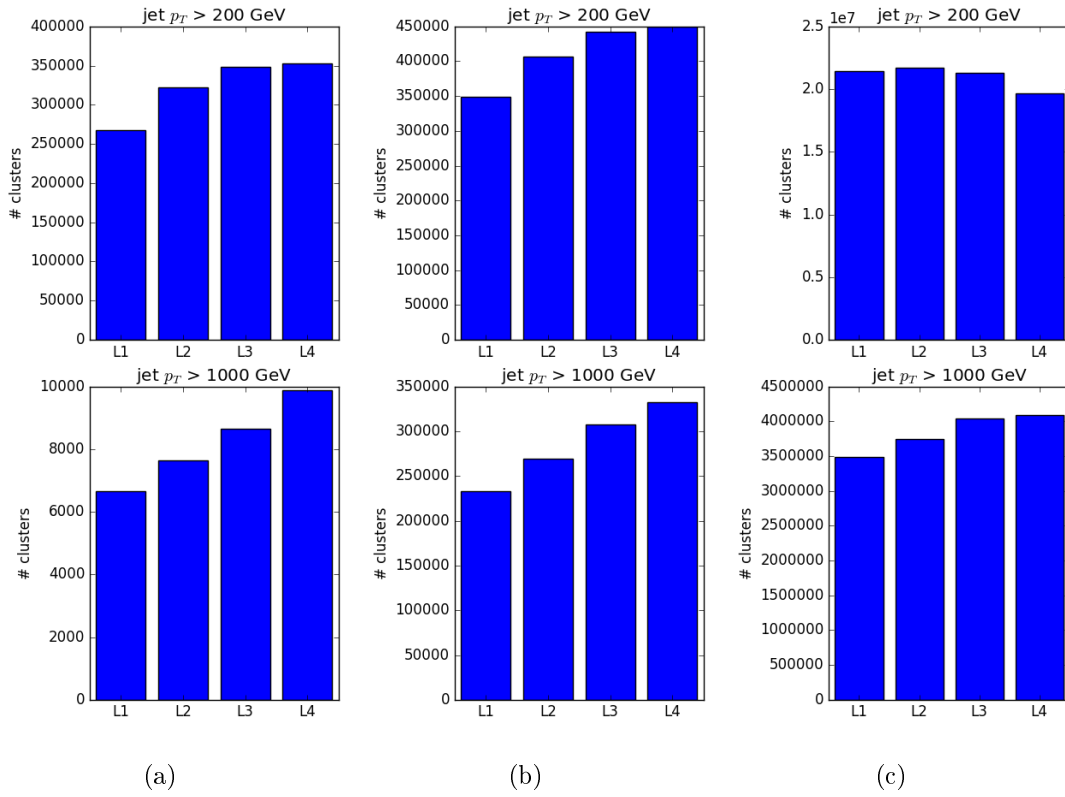


Figure 6: Total number of fired clusters in each layer in the 2 TeV sample (a), the 4 TeV sample (b) and the background sample (c), evaluated for all jets with  $p_T > 200$  GeV (above) and above a jet  $p_T$  threshold of 1200 GeV (below).

From Figs. 6a to 6c, it can already be seen that the number of hits increases between every layer in the signal samples, while the background has a flatter distribution with a slight decrease in hits between some layers. In the case of the upper plots of Figs. 6a to 6b, the the largest increase in hits can be observed between the first two layers. On the plots below, the largest increase is found between the second and third layer.

While the total hit counts yield valuable insight on the general trend, more information can be found in Fig. 7, where the hits per detector layer are recorded separately for each matched jet and filled into a histogram.

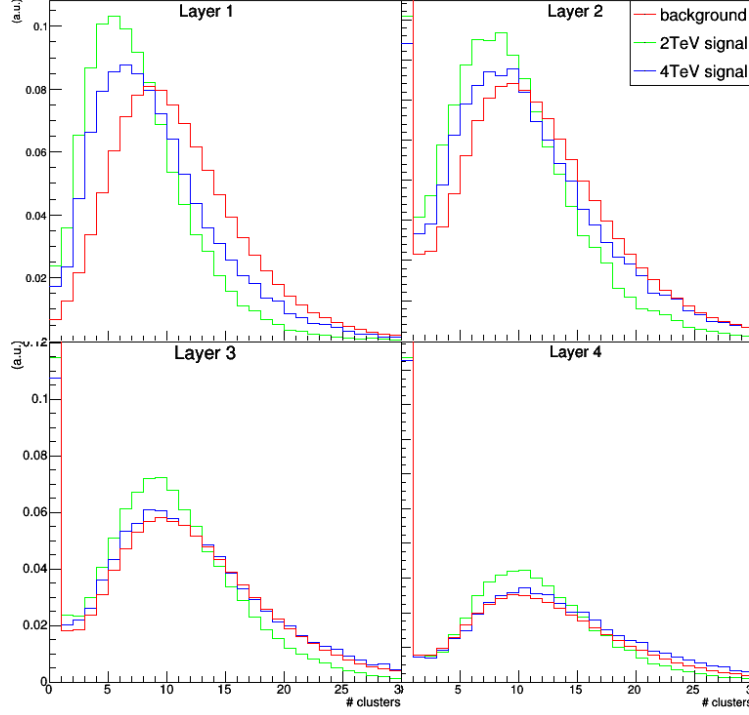


Figure 7: Normalized histograms of the number of fired clusters per layer for each matched jet.

The non-zero peak of the background sample remains approximately 9 hits for each of the four layers. The signal distributions, on the other hand, are significantly shifted to lower numbers of hits with respect to the background distribution in the first layer and gradually shift to the right in the subsequent layers until they overlap in the fourth layer. This behavior is characteristic for the signal consisting of high- $p_T$   $B$  hadrons that decay to some extent only after the first detector layer. The distinct peaks in the first detector layer also provide a hint that a hit-counting approach should be able to distinguish the signal from the background.

## 4 Results

### 4.1 Cut-based discriminants

#### 4.1.1 Discriminant and cone optimizations

Given that variables related to the number of hits per layer can distinguish background jets from high- $p_T$   $b$  jets, we proceed to use this information to develop a method for  $b$  tagging that is independent of variables used in the standard CSV tagger. First, an optimization study is performed in order to find the best discriminant and the corresponding jet-hit association cone. Since the hit multiplicity is expected to increase for layers further away from the beam axis in the case of the signal, natural choices of discriminant variables are the ratios  $L_i/L_j$  and the differences  $L_i - L_j$ , where  $L_i$  and  $L_j$  refer to the number of hits per layer,  $i$  refers to layers 2, 3, and 4 and  $j$  refers to layers 1, 2, 3, such that  $i > j$ . Thus, all such combinations were tested. Figures 8a to 8f show the distribution of all combinations of ratios between the number of hits between subsequent layers  $L_i/L_j$  inside a cone of size  $\Delta R < 0.1$ . Figures 9a to 9f show the same for all combinations of  $L_i - L_j$  inside a cone of size  $\Delta R < 0.04$ . These two cone sizes are found to yield the most optimal discriminants, as described in the following discussion.

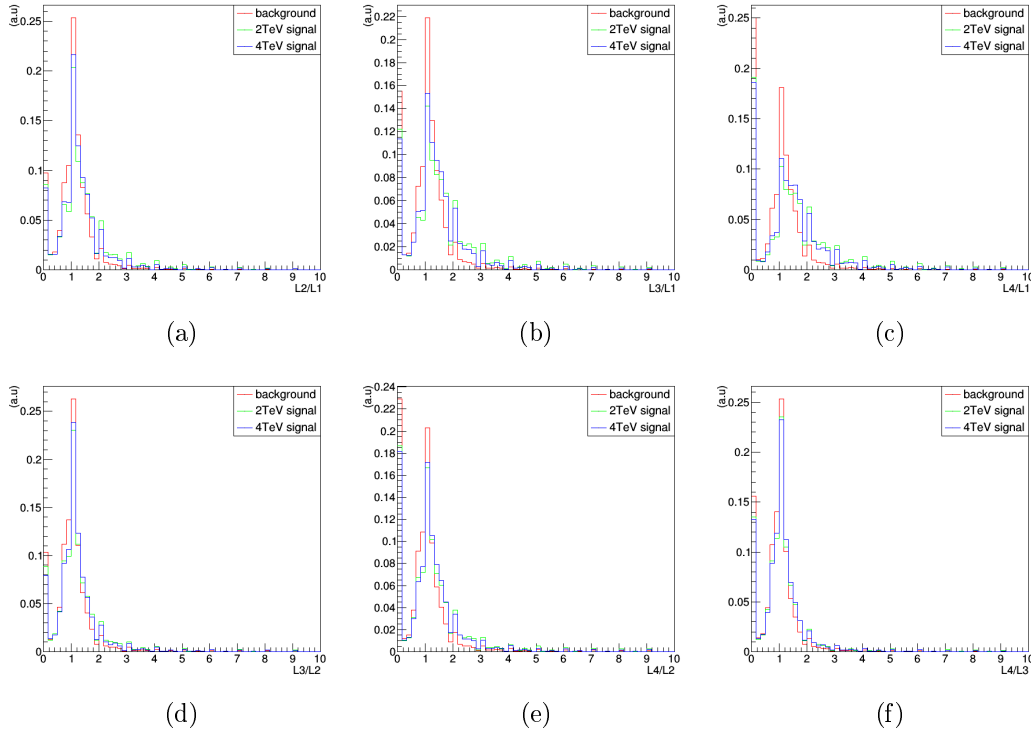


Figure 8: Normalized histograms showing the distribution of every possible ratio-based discriminant on each sample (a: L2/L1, b: L3/L1, c: L4/L1, d: L3/L2, e: L4/L2, f: L4/L3).

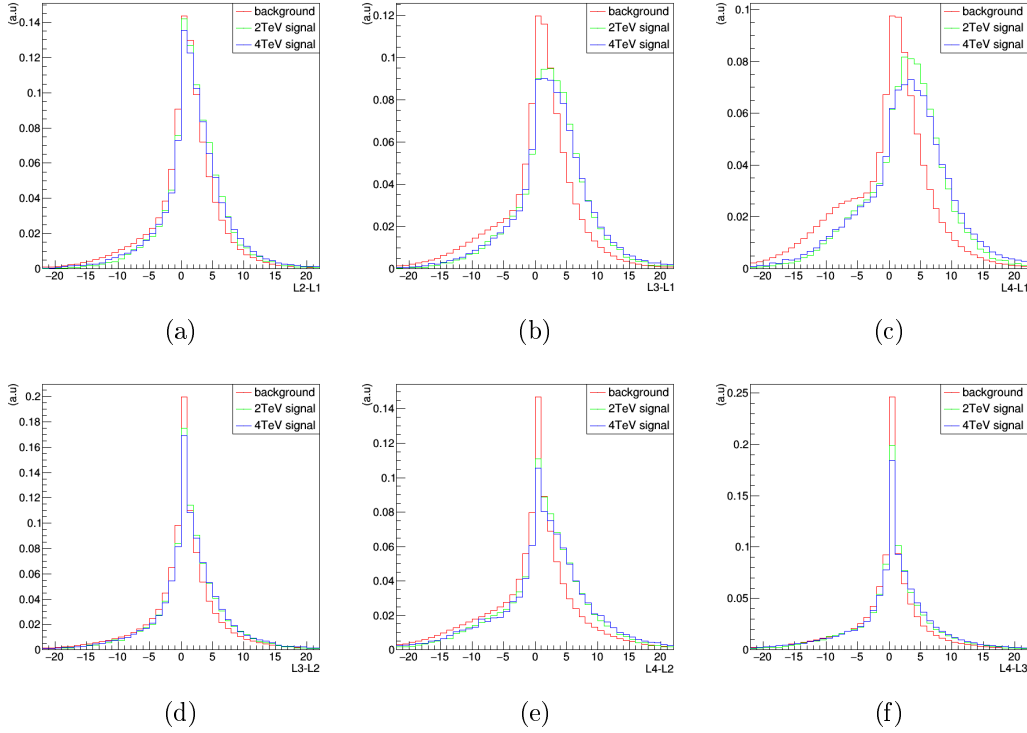


Figure 9: Normalized histograms showing the distribution every possible difference-based discriminant on each sample (a: L2/L1, b: L3/L1, c: L4/L1, d: L3/L2, e: L4/L2, f: L4/L3).

To see what value of  $\Delta R$  should be used as the cone size around the jet axis in which the fired clusters are counted, the number of clusters was plotted as a function of their position in  $\Delta R$  (see Section 6.3). Since this did not reveal a superior choice of  $\Delta R$ , different discriminants were constructed from clusters with cone sizes varying from 0.04 to 0.16, in order to determine which cone sizes had the best performance. To check the performance of every discriminant and cone size, ROC curves were produced of the efficiency (horizontal axis) and mistag rate (vertical axis). The mistag rate describes the normalized number of background jets that are falsely tagged as  $b$  jet. From these curves, a value of the efficiency (cut value) was selected in order to evaluate the total number of  $b$  jets and QCD jets that would be identified as  $b$  jets or rejected as mis-tagged jets by this cut value. Since a single cut was applied on the entire sample, a homogeneous jet distribution with respect to  $p_T$  is desirable. Thus, the following analysis was conducted on the 4 TeV signal sample alone.

The ROC curves for all possible combinations of ratios  $L_i/L_j$  are shown in Figs. 38a to 38e with varying cone sizes in the appendix Section 6.3. The cone size yielding the highest performance among these is  $\Delta R < 0.1$  and the corresponding ROC curves are shown here in Fig. 10a. In the case of the  $L_i/L_j$  ratios, there is a loss of efficiency due to

cases when there are zero hits in the inner of the two layers, such that the ratio cannot be computed. Analogously, the ROC curves for all combinations of  $L_i - L_j$  are plotted at varying cone sizes in the appendix Section 6.3 in Figs 38a to 38e. The choice of cone size with the highest performance was obtained to be  $\Delta R < 0.04$  and the corresponding ROC curves are shown in Fig. 10b.

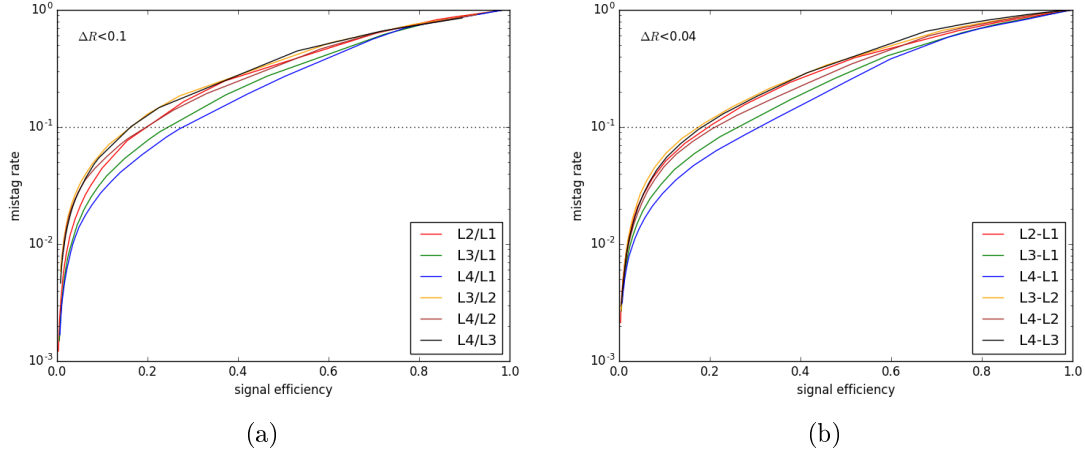


Figure 10: Best performing cone size choices for the ROC curves comparing every combination  $L_i/L_j$  (a) and  $L_i - L_j$  (b).

The black dotted line denotes the position corresponding to a 10% mistag rate. This point is used throughout the further course of the study as a working point to compare the efficiency of different discriminants among each other as well as to the efficiency of the CSV algorithm. Clearly,  $L4/L1$  and  $L4 - L1$  are the discriminants with the highest performance at almost any point in the spectrum of cut values. This is in agreement with what is seen in Fig. 7: The non-zero number of hits per event is on average almost the same for all samples in layer 4, but the non-zero mean values differ the most in layer 1. Thus, the relation between the number of hits of these layers reveals the highest discrimination power.

These discriminants are compared to the performance of the CSV algorithm in Fig. 11. While the CSV performs significantly better in the high-efficiency high-mistag region, the difference in efficiency becomes small at the 10% mistag point. Here CSV still has the highest efficiency, followed by  $L4 - L1$  and then  $L4/L1$ .

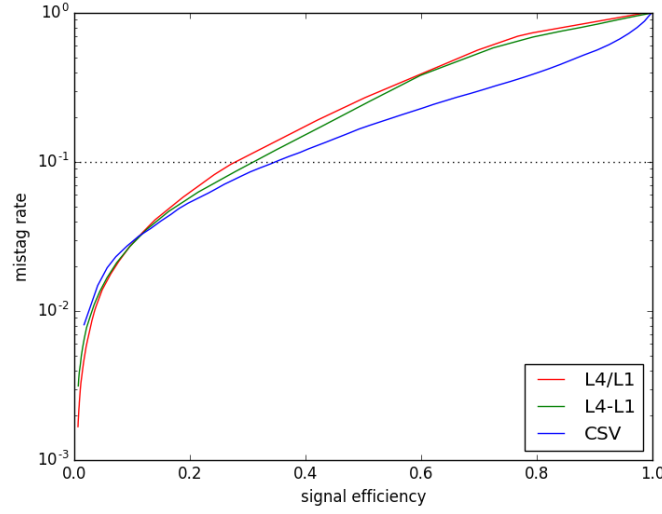


Figure 11: ROC curves comparing  $L4/L1$  with a cone size of  $\Delta R < 0.1$  and  $L4 - L1$  with a cone size of  $\Delta R < 0.04$  to the CSV algorithm.

The efficiency and the purity (1–mistag rate) as a function of discriminant cut for the  $L4/L1$ ,  $L4 - L1$  and CSV taggers are shown in Figs. 12a to 12c. From these plots, we find that a cut value of 1.833 for  $L4/L1$ , 5 for  $L4 - L1$  and 0.667 for CSV roughly corresponds to a mistag rate of 10%.

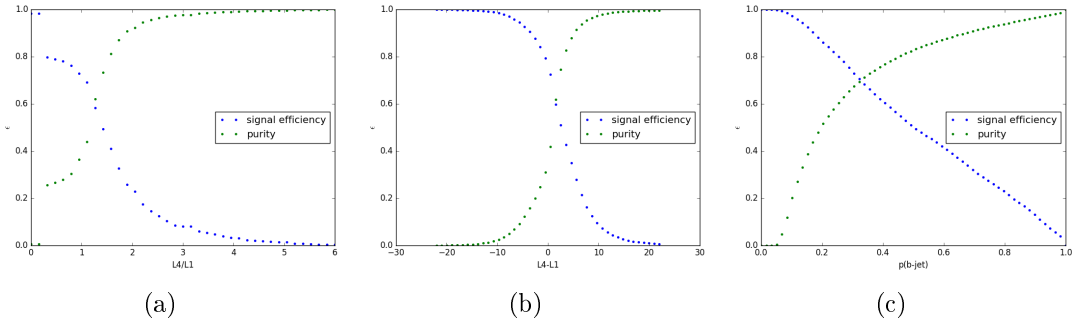


Figure 12: Signal efficiency and purity plotted as a function of possible cut values of  $L4/L1$  (a),  $L4 - L1$  (b) and the CSV  $b$  tag value (c).

The number of correctly tagged  $b$  jets corresponding to a 10% mistag rate is plotted as a function of  $p_T$  in Fig. 13a using  $L4/L1$  and in Fig. 13b using  $L4 - L1$ . In this representation, the correctly tagged  $b$  jets are separated into three categories: the ones tagged exclusively by the hit-based discriminant (red), the ones tagged exclusively by CSV (green) and those that are tagged by both (blue). The sum of the green and blue areas thus corresponds to the jets tagged by CSV, while the red area is the number of jets that are tagged additionally when also using the hit-based tagger in question. In

both cases, this area becomes increasingly large at higher values of  $p_T$ , matching the expectation of the hypothesis that predicts a decrease in performance of CSV and a higher efficiency of hit-based taggers. Above a  $p_T$  threshold of 1200 GeV, the number of correctly tagged  $b$  jets by any of the two hit-based discriminants becomes comparable to the one of CSV or even exceeds the latter.

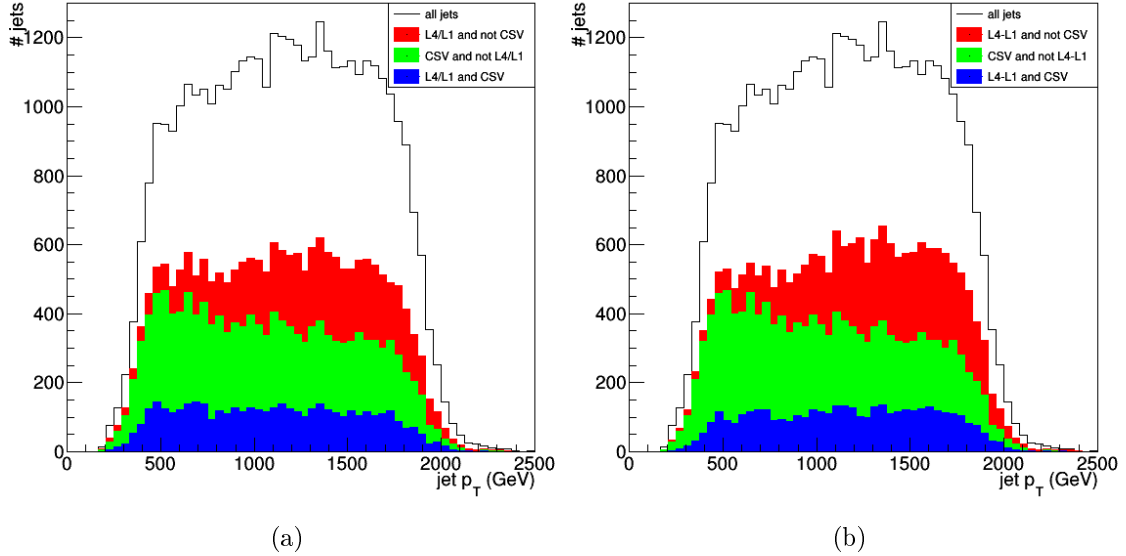


Figure 13: Number of  $b$  jets from the 4 TeV sample (black line) correctly tagged exclusively by CSV and those tagged only by  $L4/L1$  (a) or  $L4-L1$  (b) as a function of jet  $p_T$  at a mistag rate of approximately 10%.

At this point, it has to be kept in mind that so far, each discriminant is evaluated on the entire sample with jet  $p_T > 200$  GeV. Thus, the comparison in Fig. 11, as well as the number of correctly tagged  $b$  jets in Figs. 13a and 13b is ignorant with respect to the mistagging behavior of each discriminant at different values of  $p_T$ , as well as the different  $p_T$  profiles of the signal and background sample. The issue with using a single cut value, evaluated on the full sample, can be explicitly seen in Figs. 14a and 14b. Figure 14a shows the efficiency as a function of jet  $p_T$ . The increase in efficiency at higher  $p_T$ , particularly in the case of  $L4-L1$  looks promising at first sight but Fig. 14b reveals that the mistag rate increases as well at higher values of  $p_T$ . Consequently, the analysis introduced here, based on a single cut corresponding to a 10% mistag rate on the entire data sample, does not allow a meaningful comparison between different points on the  $p_T$  spectrum. It does, on the other hand, show that both  $L4/L1$  and  $L4-L1$  are taggers with a strong dependence on  $p_T$ . Figure 14b shows that the dependence of the mistag rate is very large in the case of  $L4-L1$  but not as large in the case of  $L4/L1$  where the mistag rate behaves similarly to the one of CSV. While the mistag rate is increasing as a function of  $p_T$  for all three taggers, the efficiency is increasing as well for both hit-based discriminants, contrary to CSV where the efficiency is decreasing at higher  $p_T$ .



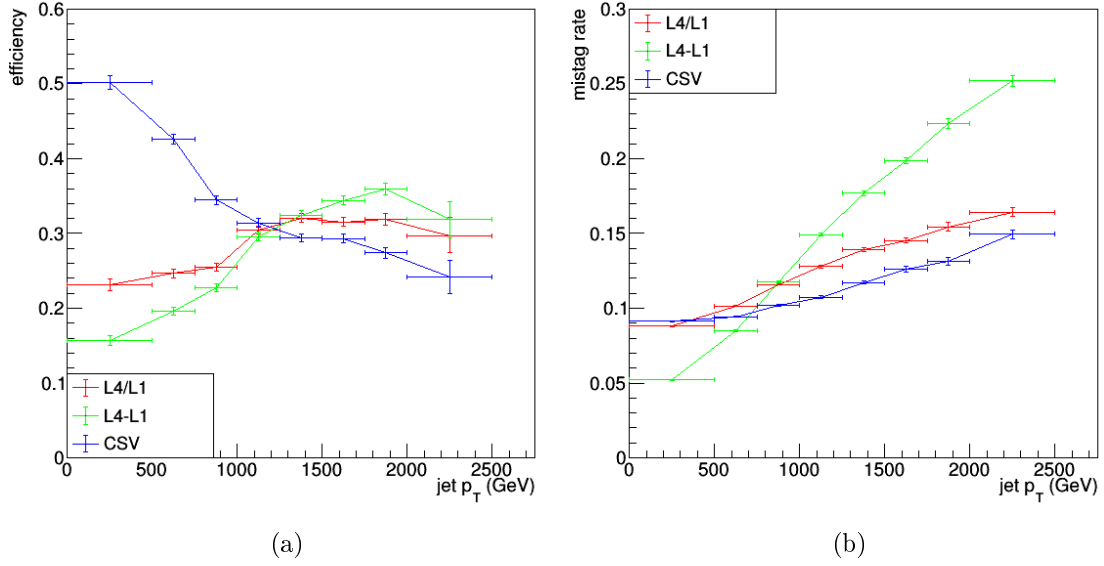


Figure 14: Efficiency (a) and mistag rate (b) of each discriminant as a function of jet  $p_T$  evaluated on the 4 TeV sample.

A first impression of how the discriminants behave at different values of  $p_T$  can be obtained by comparing the discriminants evaluated on the entire samples to those evaluated on high- $p_T$  events only. Taking into account only events above a jet momentum threshold of 1200 GeV, the distributions in Figs. 5c, 8c and 9c shift to those shown in Figs. 15c, 15a and 15b.

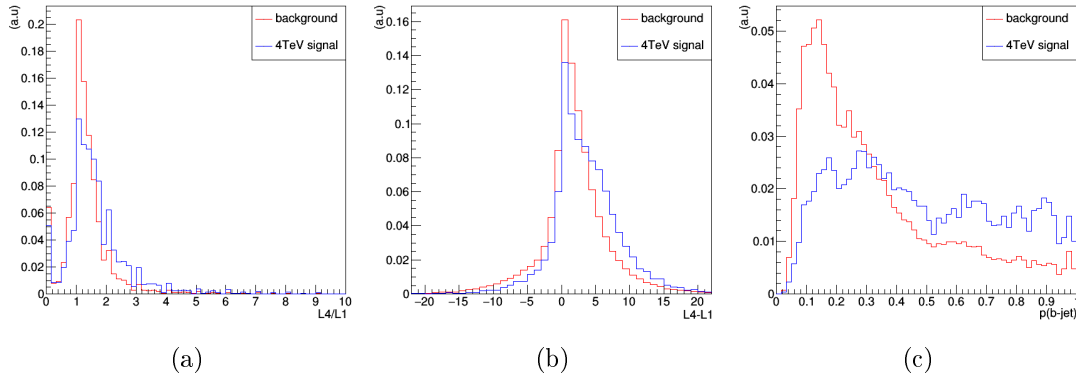


Figure 15: Normalized histograms showing the distribution of  $L4/L1$  (a),  $L4 - L1$  (b) and the CSV  $b$  tag value (c) at  $p_T > 1200$  GeV.

The distribution corresponding to the 2 TeV sample had to be omitted here because of a lack of data points above this  $p_T$  threshold, which can be seen in Fig. 5a. As before, these histograms were used to draw ROC curves, such that the performance of

the discriminants could be compared to the CSV tagger. This is shown for  $L4/L1$  in Fig. 16a and for  $L4 - L1$  in Fig. 16b where the curves corresponding to the full  $p_T$  spectrum are drawn as well.

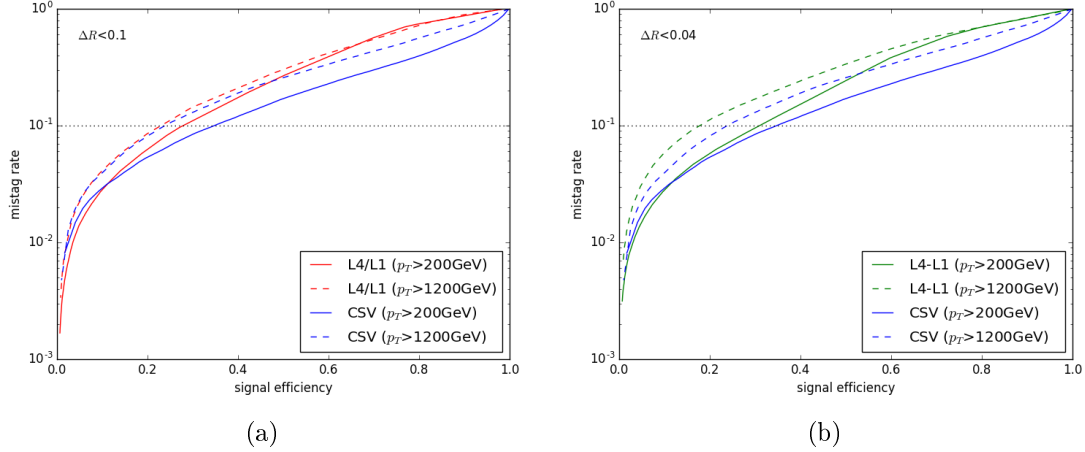


Figure 16: ROC curves comparing  $L4/L1$  (a) and  $L4 - L1$  (b) to CSV with different  $p_T$  thresholds.

While all three taggers,  $L4/L1$ ,  $L4 - L1$  and CSV, suffer a loss of efficiency above a momentum threshold of  $p_T > 1200$  GeV, this loss is notably lower in the case of  $L4/L1$ . In this region, the single-cut-based  $L4/L1$  has almost the same efficiency as CSV, whereas  $L4 - L1$  performs significantly worse.

#### 4.1.2 $p_T$ dependent cuts

Evaluating the cut-based discriminants above different  $p_T$  thresholds yields a deeper insight into the behavior at higher  $p_T$ , but neither does it resolve the  $p_T$  spectrum clearly, nor does it take into account the difference in  $p_T$  profiles between the samples. These issues were faced by evaluating each cut-based discriminant on specific  $p_T$  bins. Namely, bins with the following edges in GeV were used: 200, 500, 750, 1000, 1250, 1500, 1750, 2000, 2250, 2500. On each of these bins, the same procedure as in Section 4.1.1 was performed. Because on such narrow  $p_T$  bins, the evaluation becomes more indifferent with respect to the difference in  $p_T$  profiles between the samples, the 2 TeV sample was simply added to the 4 TeV sample used in the prior analysis in order to increase the amount of data points. The cuts corresponding to the  $p_T$  bins mentioned above are shown in Table 1.

$p_T$ bin (GeV)	$L4/L1$ cut	$L4 - L1$ cut	CSV cut
200-500	1.833	3	0.633
500-750	1.833	5	0.65
750-1000	2.0	6	0.667
1000-1250	2.0	6	0.683
1250-1500	2.0	7	0.7
1500-1750	2.167	8	0.717
1750-2000	2.167	8	0.733
2000-2250	2.167	9	0.767
2250-2500	2.167	9	0.767

Table 1: Cut values for each tagger corresponding to a 10% mistag rate in each  $p_T$  bin.

The efficiency as a function of  $p_T$  with the now  $p_T$  dependent cuts is shown in Fig. 17a. Figure 17b is left as a check that the mistag rate has been flattened to a value close to 10% over the entire  $p_T$  spectrum.

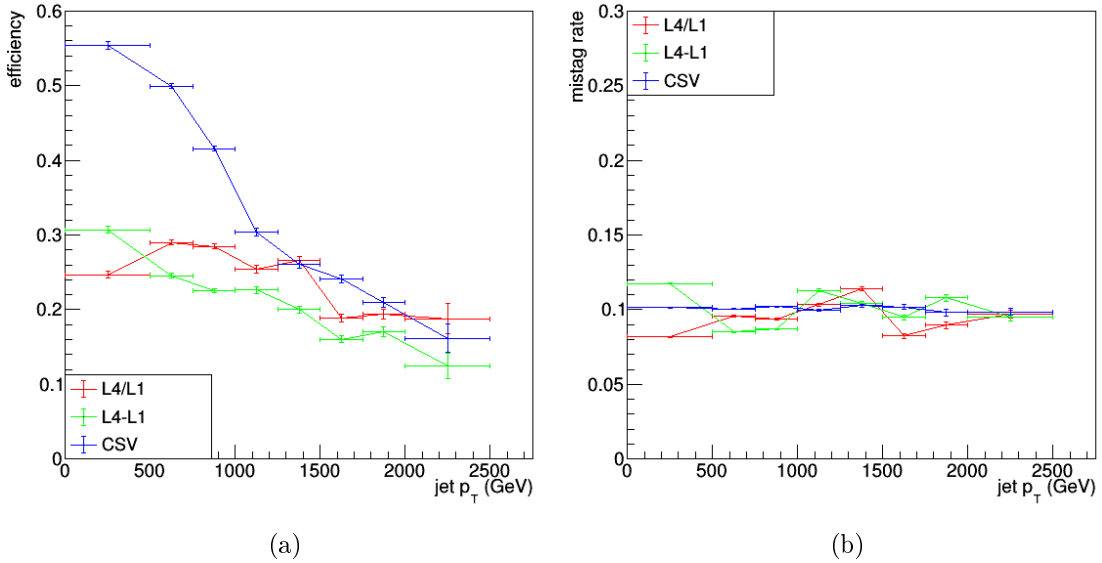


Figure 17: Efficiency (a) and mistag rate (b) of each tagger as a function of jet  $p_T$  evaluated with  $p_T$  dependent cut values.

Because both  $L4/L1$  and  $L4 - L1$  describe relations between integers, the possible values and thus the possible cuts are more discrete than in the case of the CSV algorithm. For this reason, the mistag rate in Fig. 17b yields an irreducible deviation of up to 2% from an exact value of 10%. From Fig. 17a, it can be concluded that the efficiency of the hit-based taggers, particularly  $L4/L1$ , becomes comparable to the one of the CSV tagger at a jet  $p_T$  above 1200 GeV.

As before, the number of correctly tagged  $b$  jets is plotted as a function of jet  $p_T$  in Fig. 18a using  $L4/L1$  and in Fig. 18b using  $L4 - L1$ . Again, correctly tagged jets are separated into three categories referring to whether they were tagged exclusively by one of the taggers or both.

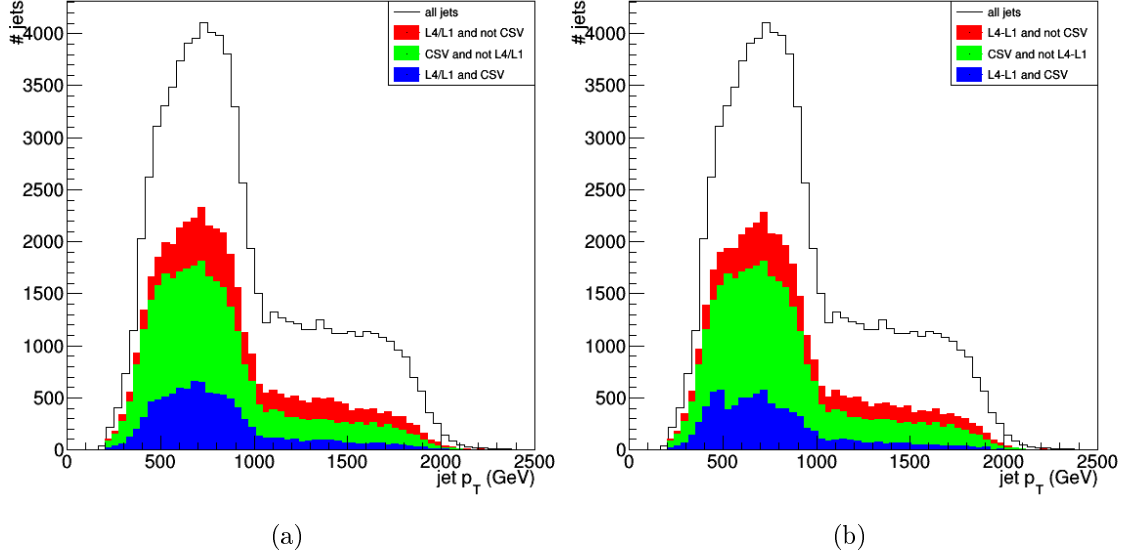


Figure 18: Number of  $b$  jets from the 2 TeV and 4 TeV sample (black line) correctly tagged exclusively by CSV and those tagged only by  $L4/L1$  (a) or  $L4 - L1$  (b) as a function of jet  $p_T$  at a mistag rate of approximately 10% using  $p_T$  dependent cuts.

The gain in efficiency compared to CSV when additionally using the hit-based taggers discussed above can be quantified by integrating the histograms corresponding to jets exclusively tagged by the tagger in question normalized by the number of jets correctly tagged by CSV in Figs. 18a and 18b. This is conducted in Fig. 19 as a function of jet  $p_T$  thresholds. Thus, on every point in the plot, the relative gain in efficiency is evaluated on the  $p_T$  spectrum above the  $p_T$  value indicated by the horizontal axis. A gain of approximately 32% in the case of  $L4/L1$  and approximately 29% in the case of  $L4 - L1$  could be achieved when evaluated on the entire sample. Above a  $p_T$  threshold of 1200 GeV, the gains increase to 63% for  $L4/L1$  and 54% for  $L4 - L1$ .

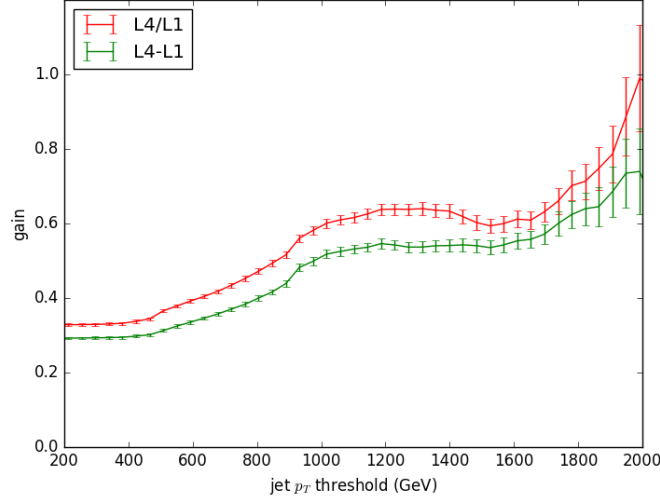


Figure 19: Efficiency gain when additionally using a hit-based tagger with respect to only using CSV, evaluated at mistag rate of 10%.

At last, also the ROC-curves are evaluated on separate  $p_T$  bins. Figures 20a and 20b show such evaluations on three different  $p_T$  bins where Fig. 20a explores the behavior of the  $L4/L1$  discriminant at different values of  $p_T$  and Fig. 20b does so for  $L4 - L1$ . While the former appears to perform similarly on the higher two  $p_T$  bins at a 10% mistag point as CSV, the latter fails to match the efficiency of CSV also at very high  $p_T$ . An explanation for the superior performance when using ratios rather than differences in hits between layers was hypothesized that the effect of showering would increase the number of hits as a function of the number of particles [5].

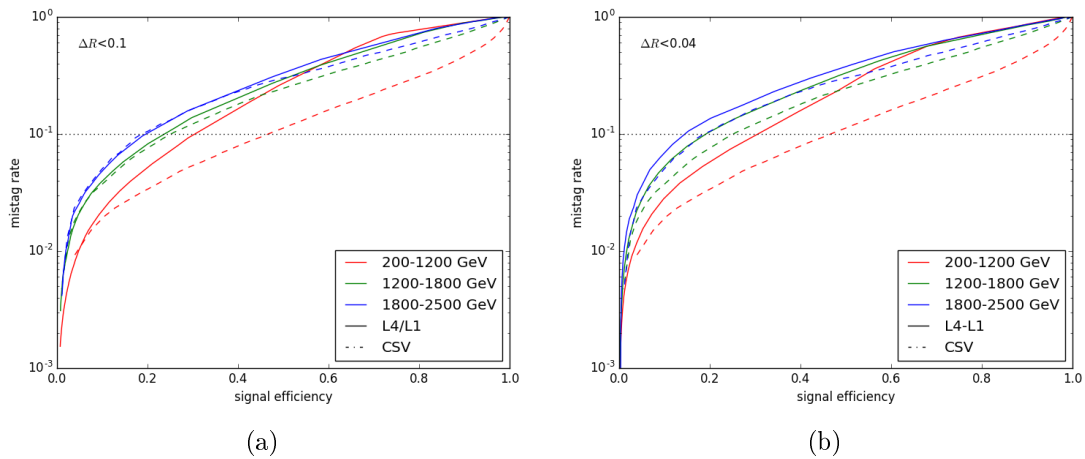


Figure 20: ROC curves comparing  $L4/L1$  (a) and  $L4 - L1$  (b) to CSV in three different  $p_T$  bins. Solid lines correspond to hit-based tagger, dashed lines to CSV.

### 4.1.3 Stability with respect to PU

Since in any realistic collision experiment at CMS, there will be pileup (PU), the cut-based discriminants discussed before are also analyzed for their stability with respect to an increased number of primary vertices. The data samples used for this analysis are introduced in Section 2.1 and their PU profiles are shown in Fig. 21.

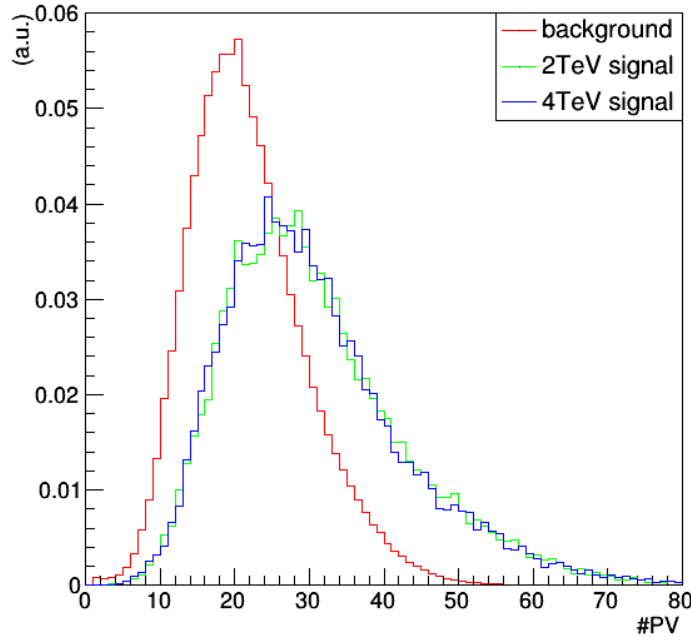


Figure 21: Normalized histograms showing the number of primary vertices per event in all three samples.

An evident property of the PU profiles shown in Fig. 21 is that both signal samples have the same distribution but they differ notably from the background. Thus, as in the case of the different  $p_T$  distributions, special care has to be taken such that a meaningful comparison can be made. Analogously to Section 4.1.2, the data samples are divided into the same separate  $p_T$  bins such that a  $p_T$  dependent cut value corresponding to a 10% mistag rate can be identified. A first trial on the stability can thus be made by drawing the efficiency as a function of  $p_T$ , as before in Fig. 17a. This is shown in Fig. 22a, next to Fig. 22b where again the mistag rate as a function of  $p_T$  is shown to be flattened to an approximate value of 10%.

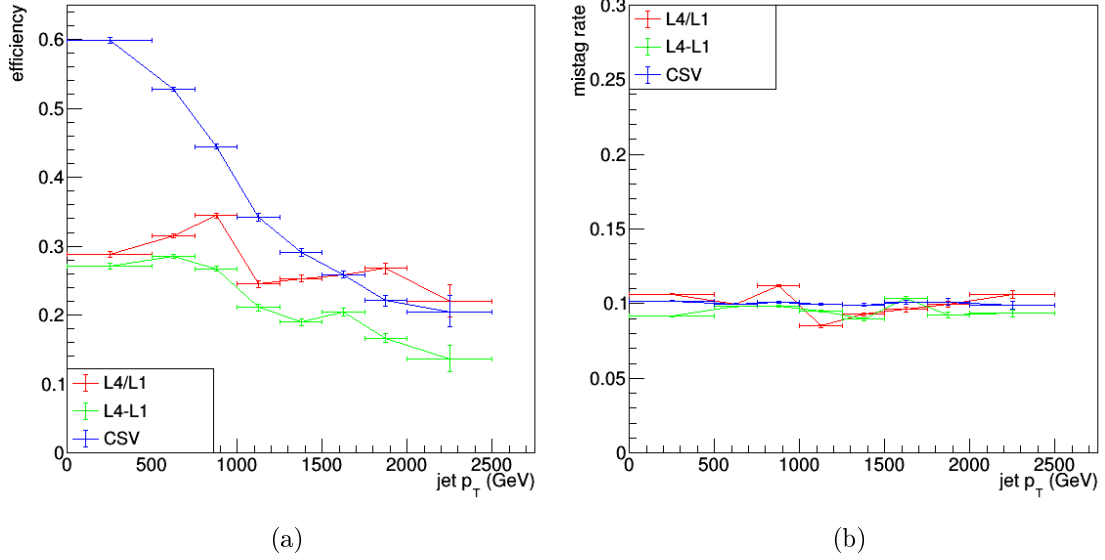


Figure 22: Efficiency (a) and mistag rate (b) of each tagger as a function of jet  $p_T$  evaluated with  $p_T$  dependent cut values on the PU samples.

All three taggers yield an equivalent behavior to Section 4.1.2. CSV and  $L4/L1$  even seem to have a slightly higher efficiency compared to before, particularly  $L4/L1$ . In order to see the effect of PU on each tagger, the efficiency and mistag rate were also plotted as a function of the number of primary vertices in Figs. 23a and 23b.

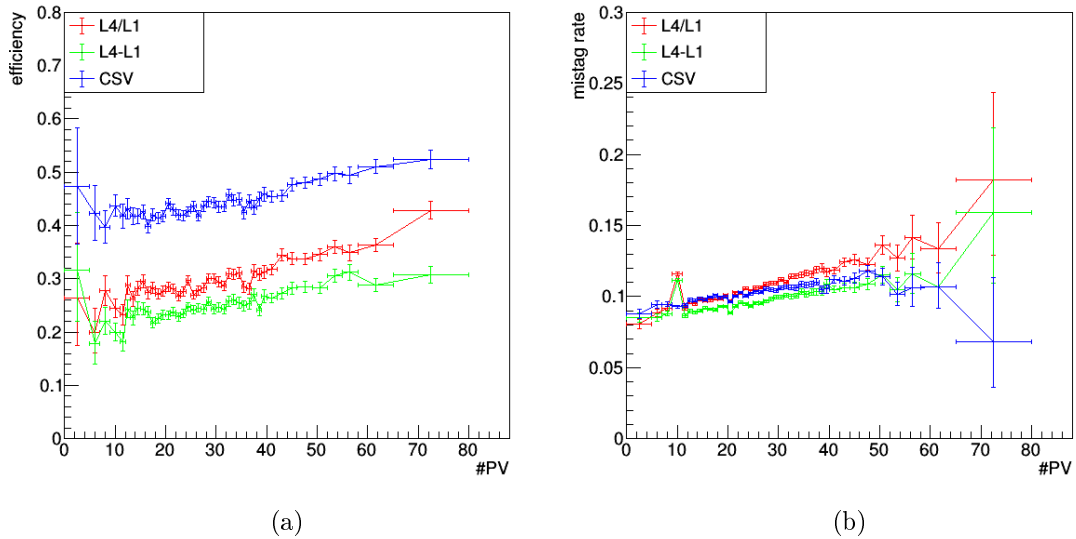


Figure 23: Efficiency (a) and mistag rate (b) of each tagger as a function of the number of primary vertices.

From the analysis above, it can be seen that both the efficiency and mistag rate of the hit-based taggers increase as a function of the number of primary vertices. However, as the efficiency as well as the mistag rate increase similarly as a function of primary vertices, the impact on the absolute performance is reduced. The behavior with respect to PU is similar to that of CSV. The conclusion of Section 4.1.2 that a cut-based  $L4/L1$  tagger yields similar and at very high  $p_T$  even slightly better performance than the currently used CSV algorithm is verified here, as well as the lower efficiency of the  $L4 - L1$  tagger. Because the analysis of hit-based taggers yielded no instance where using the difference  $L4 - L1$  was superior to using the ratio  $L4/L1$ , only the latter is used for comparison in the next section.

## 4.2 Artificial neural networks

As a next step to examine whether a hit-based approach would improve the  $b$  tagging performance at high  $p_T$ , the number of fired clusters per layer were used as inputs to an artificial neural network (ANN). For this aim, different combinations of input variables as well as various kinds of network architectures were examined and compared. As before, the performance of each tagger was first evaluated on the pileup free data samples. Later, the behavior with respect to an increased number of primary vertices was investigated on the pileup samples.

### 4.2.1 Choice of model

First, a simple network architecture consisting of five densely connected layers with 150 nodes and dropout of 0.5 [17] was used to test different combinations of input variables. The activation functions were all chosen to be rectified linear units [18] and the cost function binary cross entropy [19] that returns a value between 0 and 1 corresponding to the probability of a jet being a  $b$  jet, as the CSV tagger does.

An intuitive first approach was simply inputting the number of clusters in each layer at a fixed cone size. Inspired by the evaluation in Section 6.3,  $\Delta R < 0.04$  and  $\Delta R < 0.1$  were both tried separately as cone sizes. The comparison of the corresponding ROC curves in Fig. 24 between these approaches reveals that the former cone size performs significantly better than the latter. The next approach attempted was to include input variables from both cone sizes into the same ANN. This further increased the performance (see Figure. 24) and thus, all the five cone sizes that were analyzed in Section 6.3 (0.04, 0.06, 0.08, 0.1, 0.16) were combined and used as input variables of another ANN model, also compared in Fig. 24. This resulted in a slightly higher performance compared to the previous approaches. The latter two models have a comparable efficiency to the CSV tagger at a 10% mistag rate and a higher one at smaller mistag rates.



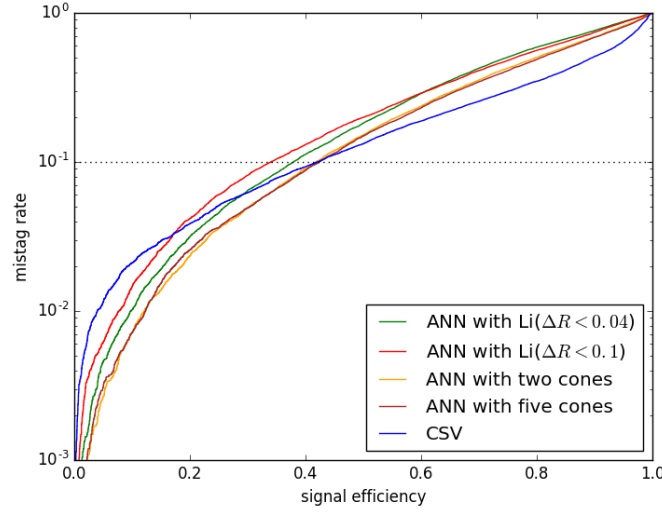


Figure 24: Comparison of different input variables: the number of clusters in each layer using two separate cone sizes ( $\Delta R < 0.04$ ,  $\Delta R < 0.1$ ), both cone sizes combined and five cone sizes combined (0.04, 0.06, 0.08, 0.1, 0.16.)

Concluding the combination of the number of hits in each layer at five different cone sizes to be the most promising input variables, different architectures were then compared to exploit these variables the most. Since a spatial structure lies behind these input variables, replacing the first densely connected layer with a convolutional layer [20] seemed to be a plausible improvement. Thus, the variables were arranged in a two dimensional matrix such that the rows correspond to the cone size and the columns to the layer number. A convolutional kernel that extends over all rows with a width of two columns thus sweeps over the input matrix, resulting in 32 output filters. The performance of such a convolutional network can be seen in Fig. 25.

To see whether the discriminants ( $L4/L1$  and  $L4 - L1$ ) elaborated in Section 4.1 might add any non redundant discrimination power, they were used together with the ratios of the number of clusters in  $\Delta R < 0.1$  in subsequent layers,  $L2/L1$ ,  $L3/L2$  and  $L4/L3$ , as input variables to a separate branch in the ANN starting with a densely connected layer of 150 nodes and dropout of 0.5. This additional branch then merges the output of the convolutional branch, after which the network continues as before.

A last approach was made by additionally plugging the number of clusters per layer in all five cone sizes into a densely connected layer with 150 nodes and dropout of 0.5, parallel to the convolutional layer and then merging the outputs of the other two branches. The idea was that the convolutional layer alone might also have disadvantages compared to the densely connected layer and thus they would complement each other. The number of nodes in the layer after the merging was increased from 150 to 200 to take into account the larger input.

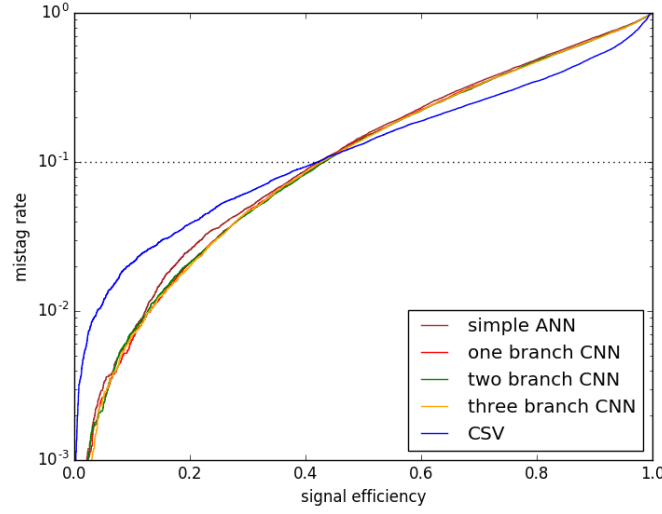


Figure 25: Comparison of different ANN architectures: a simple densely connected network, a convolutional network (one branch CNN), a convolutional network with an additional branch of discriminants (two branch CNN) and one with an additional densely connected input branch (three branch CNN).

The ROC curves for these three modifications are plotted in comparison to the simple densely connected five layer network in Fig. 25. There it can be seen that the additional branches add only a small if any increase in efficiency. The two branch approach was chosen, since it seems to outperform the one branch convolutional network slightly at higher mistag rates but is not as complex as the three branch model yielding the same performance.

Since there is in general a correlation expected between  $p_T$  and the extent of the jet cone, it appears also promising to add  $p_T$  to the input variables. The issue one has to face here is that the signal samples have a significantly different  $p_T$  profile than the background (see Fig. 5a). Thus, if  $p_T$  is plugged into an ANN for training without reweighting the training samples, the network will simply overfit, i.e., it will categorize high  $p_T$  events as signal and low  $p_T$  events as background. To avoid such overfitting, the  $p_T$  distributions were represented in histograms with a bin size of 5 GeV. The background histogram was divided by the signal distribution, resulting in a weight function for each signal event on each  $p_T$  bin. The result of this reweighted training with an additional  $p_T$  variable can be seen in Fig. 26 in comparison to the two branch approach discussed above.

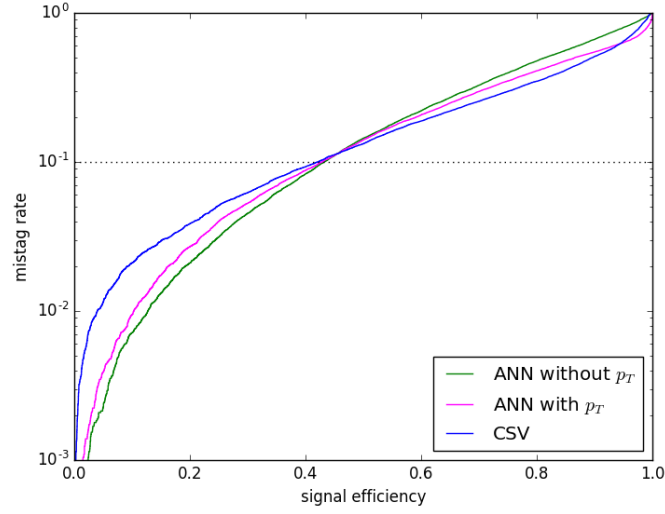


Figure 26: Comparison of the same ANN (two branch CNN) with and without  $p_T$  as input variable.

In Fig. 26, it appears that adding  $p_T$  to the input variables and reweighting the training data does not result in a significant improvement in efficiency at a 10% mistag rate. It only performs better at larger mistag rates and performs worse at lower ones. In order to obtain more insight into the  $p_T$  dependent behavior of the neural networks, the ROC curves that were produced for the entire  $p_T$  spectrum in Fig. 26 were also plotted on two separate  $p_T$  bins: one for the events below (Fig. 27a) and one above (Fig. 27b) a  $p_T$  threshold of 1200 GeV.

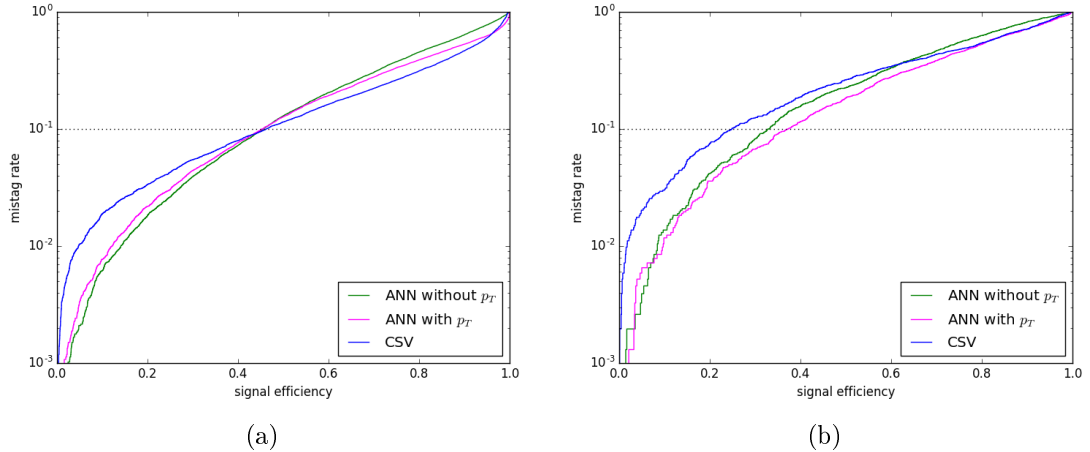


Figure 27: Comparison of the same ANN (two branch CNN) with and without  $p_T$  as input variable below (a) and above (b) a  $p_T$  value of 1200 GeV.

While the ROC curves in Fig. 27a have a very similar behavior as before in Fig. 26, a different one can be witnessed in Fig. 27b. Above this  $p_T$  threshold of 1200 GeV, the ANN that included  $p_T$  as a variable has become the most efficient one throughout the whole spectrum. This plot also gives a first hint that the hit-based ANNs outperform CSV notably at very high  $p_T$ . More on this will be elaborated in the next section.

#### 4.2.2 Comparison to cut-based approach and CSV

Analogous to Section 4.1.2, the behavior of the hit-based ANNs at different values of  $p_T$  is evaluated at a 10% mistag rate by determining each corresponding cut in the ANN  $b$  tag values on separate  $p_T$  bins. The same bins as in Section 4.1.2 were used. Thus, the efficiency of both ANN models (with and without  $p_T$  as input variable) could be plotted, as before, against  $p_T$  in comparison to CSV and the cut-based  $L4/L1$  tagger from Section 4.1.2 in Fig. 28a. Again, the mistag rate as a function of  $p_T$  is shown in Fig. 28b.

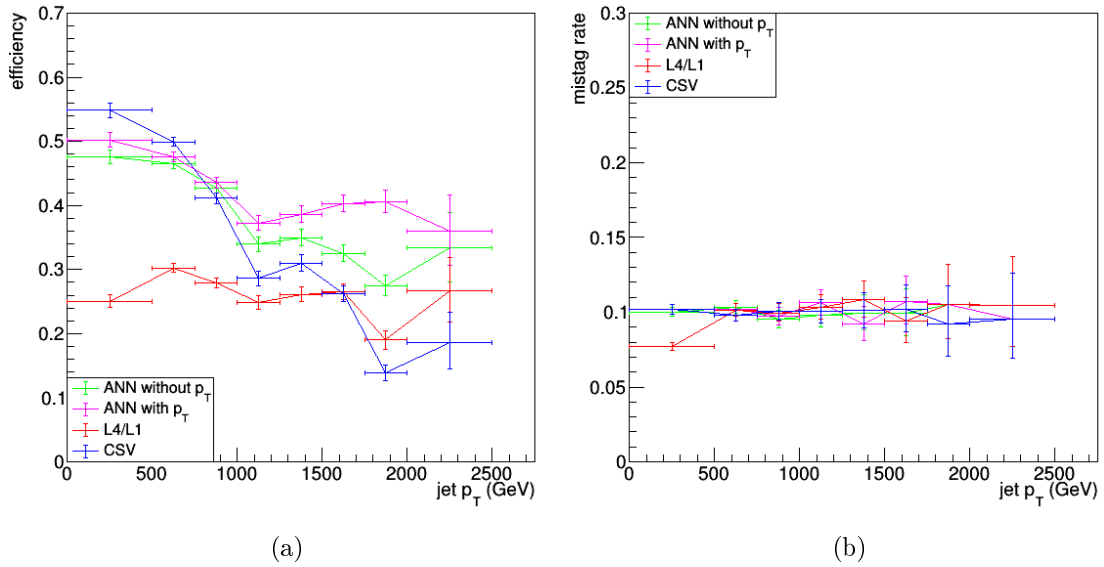


Figure 28: Efficiency (a) and mistag rate (b) of the ANN with and without  $p_T$  as input variable, the  $p_T$  dependent cut-based  $L4/L1$  tagger and CSV as a function of jet  $p_T$ .

The ANN efficiencies in Fig. 28a have a similar behavior as the cut-based  $L4/L1$  tagger from before: at the lower end of the  $p_T$  spectrum, the efficiencies are lower than the one of CSV, then they become increasingly comparable at higher  $p_T$  until the ANN efficiencies become notably higher than the one of CSV. The difference now to  $L4/L1$  is that both ANN models have a higher performance throughout the entire  $p_T$  spectrum, particularly at the lower end. Thus, the region where the ANN models have a higher efficiency than CSV already begins between 800 and 900 GeV. The reason why the graph of  $L4/L1$  looks

slightly different than before in Fig. 17b is that the ANN models (and thus also  $L4/L1$  in this section) were only evaluated on a distinct subset (20%) of the data sample on which the models were not already trained. For the difference between the two ANN models, it can be noted that their efficiency is comparable below 1000 GeV, while at higher  $p_T$  values, the  $p_T$  dependent model outperforms the other model significantly. This indicates that a useful correlation between  $p_T$  and the other input variables could be made, particularly since the reweighting process penalizes the training samples at higher  $p_T$  to ensure that  $p_T$  itself is not learned as a discriminant, but rather correlations between  $p_T$  and other input variables.

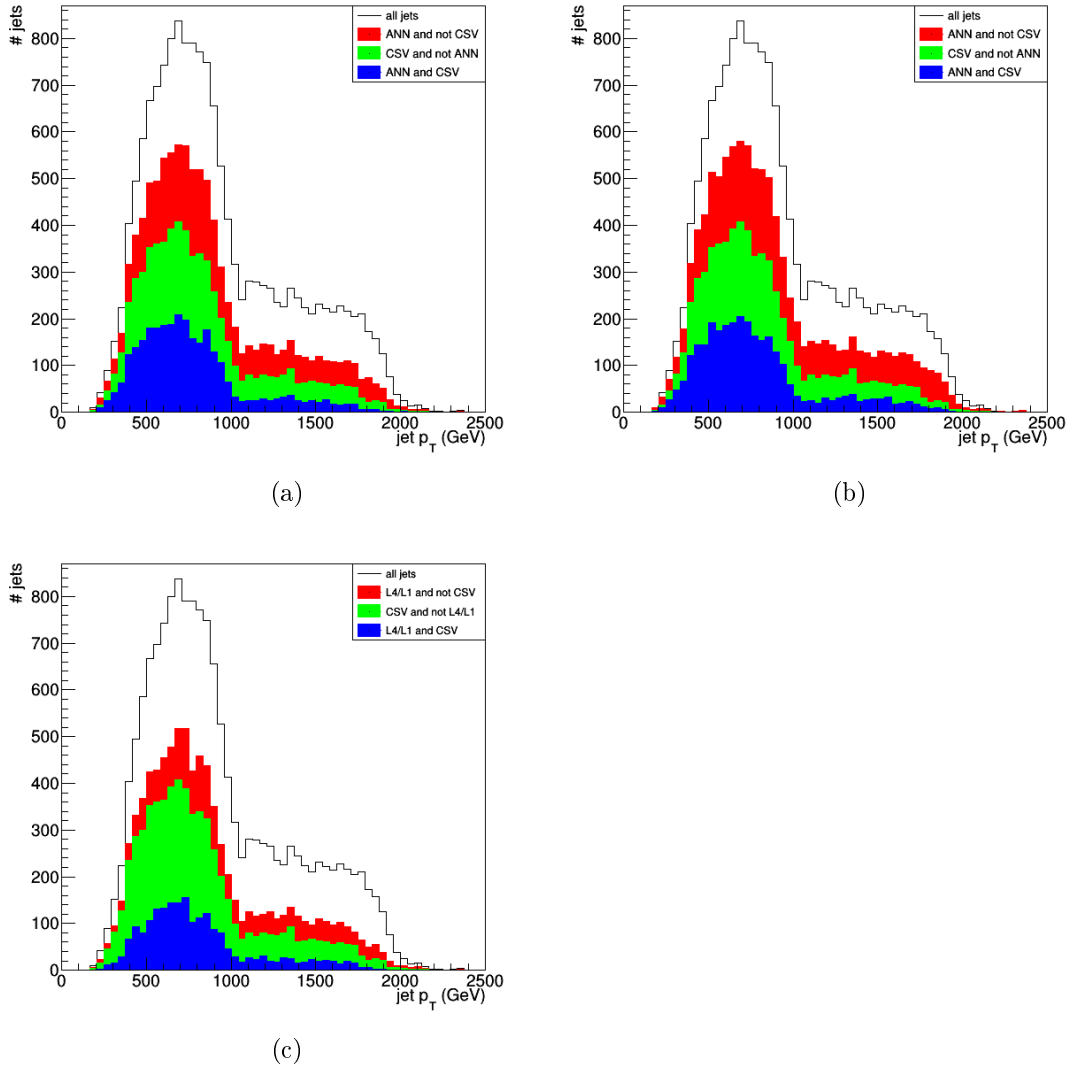


Figure 29: Number of  $b$  jets from the 2 TeV and 4 TeV sample (black line) correctly tagged exclusively by CSV and those tagged only by one of the hit-based taggers (a: ANN without  $p_T$ , b: ANN with  $p_T$ , c:  $L4/L1$ ) as a function of jet  $p_T$  at a mistag rate of approximately 10%.

To see the  $p_T$  dependent gain in using an ANN model or  $L4/L1$ , the number of correctly tagged  $b$  jets exclusively by the tagger in question is compared to the number of those tagged exclusively by CSV in Figs. 29a to 29c. Figure 29a corresponds to the ANN model without  $p_T$  and Fig. 29b to the one with  $p_T$ .

Analogous to Section 4.1.2, the gain in efficiency from additionally using the hit-based taggers is quantified by comparing the number of correctly tagged  $b$  jets exclusively by the corresponding tagger to the number of jets tagged by CSV above different  $p_T$  thresholds. The result is plotted in Fig. 30. Evaluated on the entire  $p_T$  spectrum, a gain of approximately 34% can be observed for  $L4/L1$ , 54% for the ANN without the  $p_T$  variable and 60% for the ANN with  $p_T$  input. When a threshold of  $p_T > 1200$  GeV is chosen, this gain increases to approximately 68% for  $L4/L1$ , 92% for the ANN without  $p_T$  input and 112% for the ANN with  $p_T$  input.

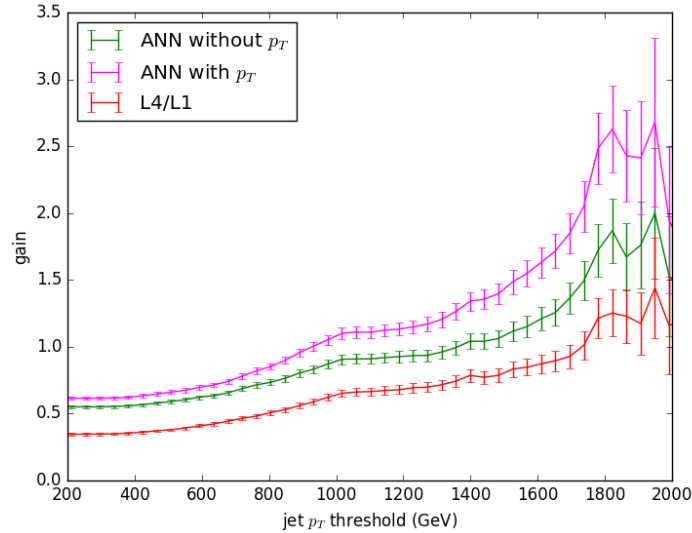


Figure 30: Efficiency gain when additionally using a hit-based tagger with respect to only using CSV, evaluated at a mistag rate of 10%.

### 4.2.3 Stability with respect to pileup

The ANN approach is also tested on the pileup samples. In this case, not only  $p_T$  can be used as an input variable, but also the number of primary vertices. The same issue as in the case of  $p_T$  applies here: the signal and the background have different pileup profiles. Thus, the number of PV can only be used as input variable when the samples are reweighted according to the sample profiles. This is done analogously to before with  $p_T$ . Figure 31 shows the ROC curves of the models where either  $p_T$ , PV or none of these two variables were used as additional inputs. The models with and without  $p_T$  as input variable behave analogously to before. Using PV as variable, on the other hand, has a

lower efficiency than without throughout the greater part of the spectrum, at most being of comparable efficiency at higher mistag rates.

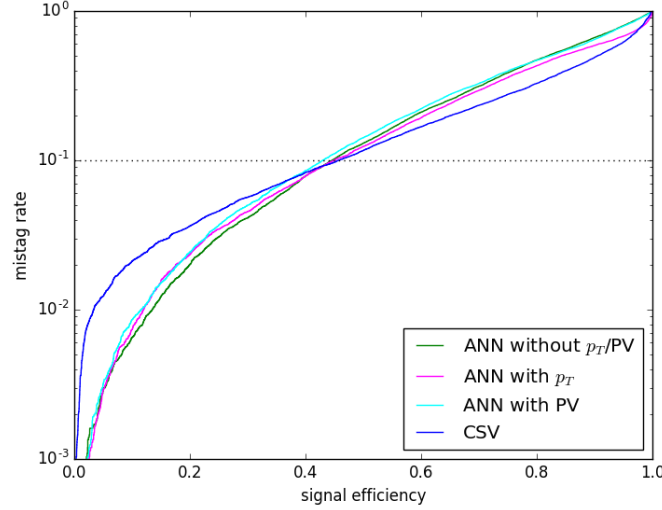


Figure 31: Comparison of the same ANN (two branch CNN) with  $p_T$  or PV as input variable and without any additional variables.

When evaluated on two separate  $p_T$  bins, above and below 1200 GeV, the model with and without  $p_T$  as an input variable have the same behavior as before, which can be seen in Figs. 32a and 32b. The model using a PV variable suffers the same relative loss in efficiency as the model without additional input variables on the high- $p_T$  bin.

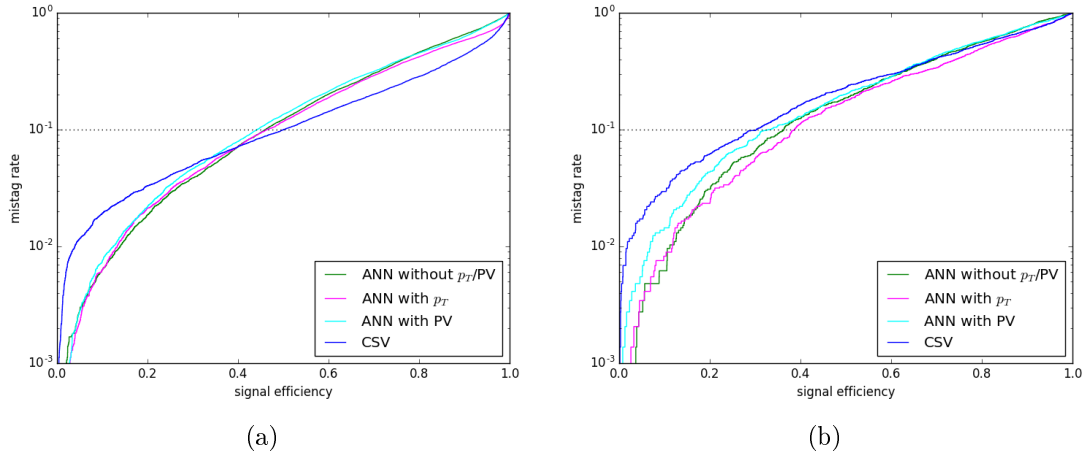


Figure 32: Comparison of the same ANN (two branch CNN) with  $p_T$  or PV as input variable and without any additional variables at below (a) and above (b) a  $p_T$  value of 1200 GeV.

The efficiency of the three ANN models at different values of jet  $p_T$  is very similar when evaluated on the PU sample compared to the analysis in Section 4.2.2. The corresponding graphs are shown in Fig. 33a. The newly introduced model using PV as input variable is again shown to have a lower efficiency at higher values of  $p_T$  compared to the other models but still outperforms CSV in the same region. The mistag rate in Fig. 33b is sufficiently flat in the case of the model without additional input variables and the one with  $p_T$  for a reasonable comparison to CSV. The  $L4/L1$  tagger, on the other hand, fluctuates with some statistical uncertainty making an exact comparison to the other taggers more difficult. Similarly, the model using PV as input variable tends to have a falling mistag rate above 1000 GeV, although there are statistical fluctuations. Thus, the ROC curves in Figs. 32a and 32b yield a preciser comparison between the models.

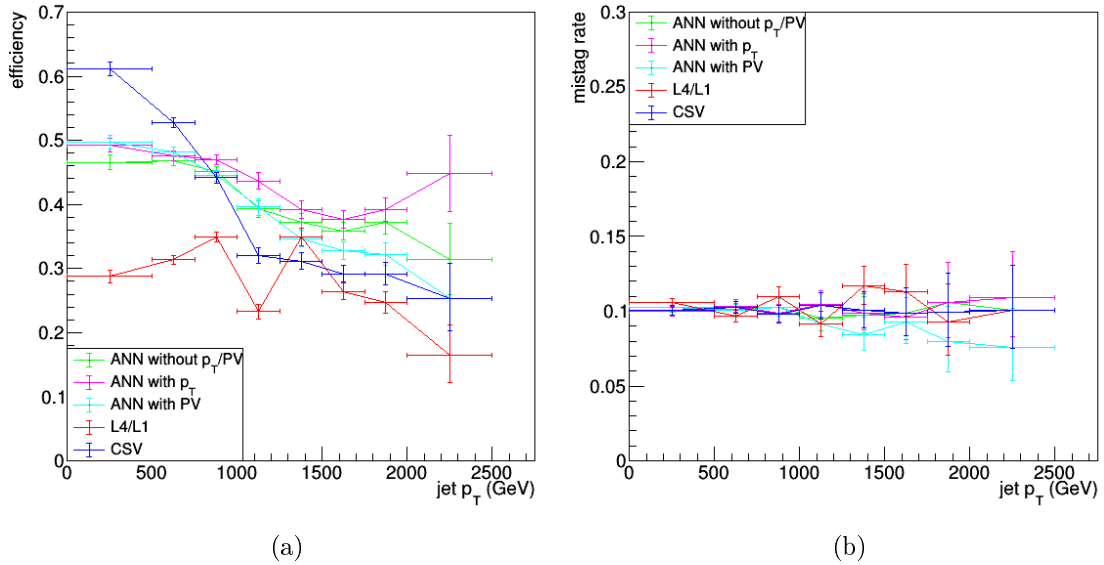


Figure 33: Efficiency (a) and mistag rate (b) of the three ANN models, the  $p_T$  dependent cut-based  $L4/L1$  tagger and CSV as a function of jet  $p_T$ , evaluated on the PU samples.

Finally, in order to best see the behavior of the ANNs at different numbers of primary vertices, their efficiency and mistag rate are plotted as a function of PV in Figs. 34a and 34b. From these plots, it can be seen that a non-negligible sensitivity to pileup is also observable for the ANNs. Again, the efficiency is increasing as well as the mistag rate at higher numbers of primary vertices with a slope comparable to the one corresponding to CSV. The ANN model that was reweighted with respect to the pileup profiles has a flatter behavior as a function of PV, which matches the expectation that such training should decrease the sensitivity with respect to pileup.



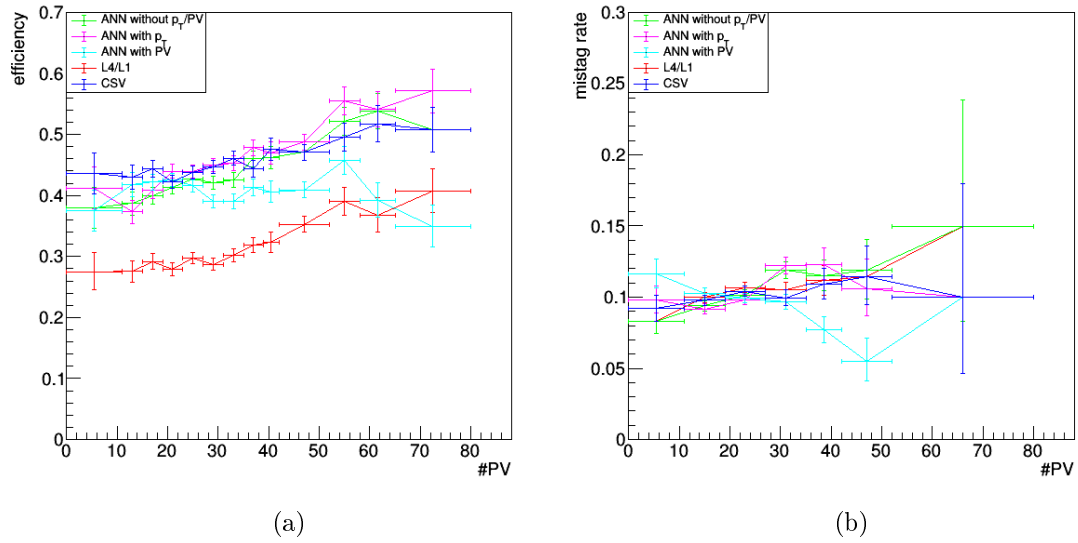


Figure 34: Efficiency (a) and mistag rate (b) of the three ANN models, the cut-based  $L4/L1$  tagger and CSV as a function of the number of primary vertices.

## 5 Summary

Above a jet  $p_T$  of 1 TeV, a significant loss in efficiency is observed for the track-based CSV algorithm. Using a hit-based tagging approach, a similar or improved performance can be achieved in this region of high  $p_T$ .

A first approach towards a hit-based tagger consisted in using the number of hits in the fourth pixel detector layer divided by the number of hits in the first one inside a cone of size  $\Delta R < 0.1$  around the jet axis ( $L4/L1$ ) and the difference between the same quantities inside a cone of size  $\Delta R < 0.04$  around the jet axis ( $L4 - L1$ ) as discriminants. Applying a single cut corresponding to a mistag rate of 10% yielded promising results with both discriminants exceeding the efficiency of CSV above 1.2 TeV. But the mistag rate also increased notably as a function of  $p_T$ . Thus, a  $p_T$  dependent cut was applied to obtain a flat mistag rate of 10% throughout the entire examined  $p_T$  spectrum. While the two hit-based discriminants do not exceed the efficiency of CSV at this working point, the efficiency of  $L4/L1$  remains comparable to the one of CSV. Additionally using  $L4/L1$  under these constraints yields a gain of approximately 32% compared to using merely CSV on the entire sample. Above a  $p_T$  threshold of 1.2 TeV, this gain increases to 63%. For  $L4 - L1$  these gains become 29% and 54% respectively. The absolute performance with respect to pileup was investigated for each tagger and found to be relatively stable and similar to CSV.

As a second approach of hit-based taggers, an artificial neural network (ANN) was implemented. The most optimal choice of model turned out to be one with a convolutional input layer processing the number of hits for each detector layer in five different choices of cone sizes with an additional branch using the previously analyzed quantities  $L4/L1$  and  $L4 - L1$  as well as  $L2/L1$ ,  $L3/L2$  and  $L4/L3$ . A variation was also introduced that takes  $p_T$  as input as well where the training samples were reweighted according to their  $p_T$  profiles. Evaluating these models on a constant mistag rate of 10% throughout the entire  $p_T$  spectrum, they achieved a comparable efficiency as CSV below 1.2 TeV and notably exceeded the efficiency of CSV above the same threshold, particularly the model using  $p_T$  as input variable. The gain from using the ANN model without a  $p_T$  input variable at this working point relative to merely using CSV on the entire sample is approximately 54%, increasing to 92% at  $p_T > 1.2$  TeV. In the case of the model using  $p_T$ , these gains become 60% and 112 %. Also the neural networks yield a relatively stable absolute performance with respect to pileup, comparable to CSV.

## 6 Appendix

### 6.1 General concepts

The coordinate system commonly used in CMS is shown in Fig. 35, where the  $z$  axis corresponds to the beam axis.

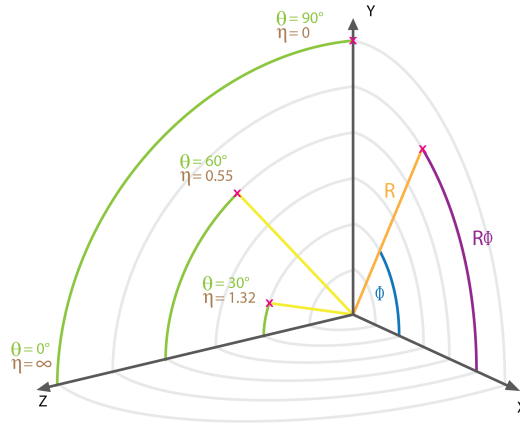


Figure 35: Illustration of the coordinate system convention used in CMS [21].

$\Delta R$  is a quantity used to describe a difference in direction between vectors.

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

where  $\eta$  is the pseudorapidity defined by

$$\eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right]$$

### 6.2 Hadrons vs quarks

The discussion in Section 2.2 on whether to match the signal jets to the hard process  $b$  quarks or to the outgoing  $B$  hadrons was also looked at by comparing their features. Namely  $p_T$ ,  $\eta$  and  $\phi$  of the corresponding quark and hadron were compared in Figs. 36a to 36c where the modulus of the difference between the hadron and the quark observable divided by the quark observable were plotted in normalized histograms for both signal samples.

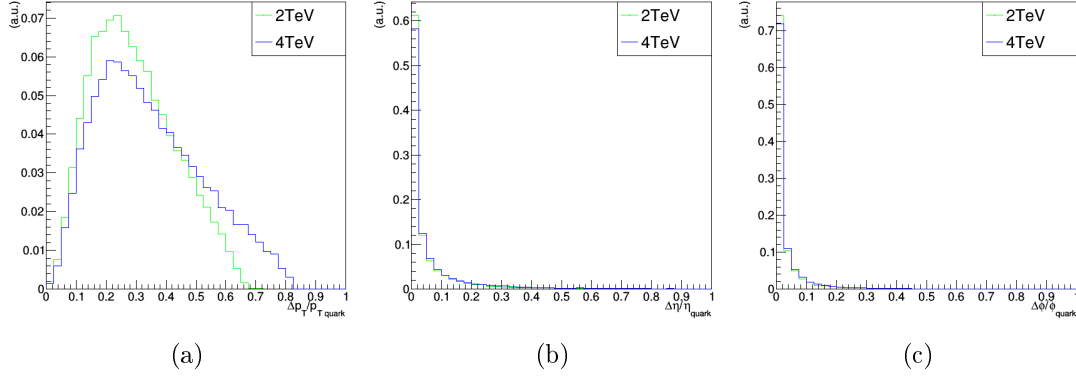


Figure 36: Distribution of the normalized difference in  $p_T$  (a),  $\eta$  (b) and  $\phi$  (c) between corresponding  $B$  hadrons and  $b$  quarks.

From these plots it can be clearly seen that the direction of travel hardly differs, which makes the matching of jets indifferent with respect to the two options since the matching is done by comparing the direction of the two momentum 4-vectors. There is, on the other hand, a significant decrease in  $p_T$  visible due to the hadronization process. Since the  $B$  hadron is the particle that possibly traverses the first detector layer before decaying, it was then decided that this should be the one a jet is matched to in order to be counted as a  $b$  jet.

### 6.3 Determination of optimum cone sizes

In order to determine what value of  $\Delta R$  should be used when matching clusters to a jet axis, two approaches were followed. The first one consisted of plotting the number of matched clusters in each layer as a function of  $\Delta R$ . This procedure yields a first insight on which cone size looks a priori reasonable and can be seen in Figs. 37a to 37c.

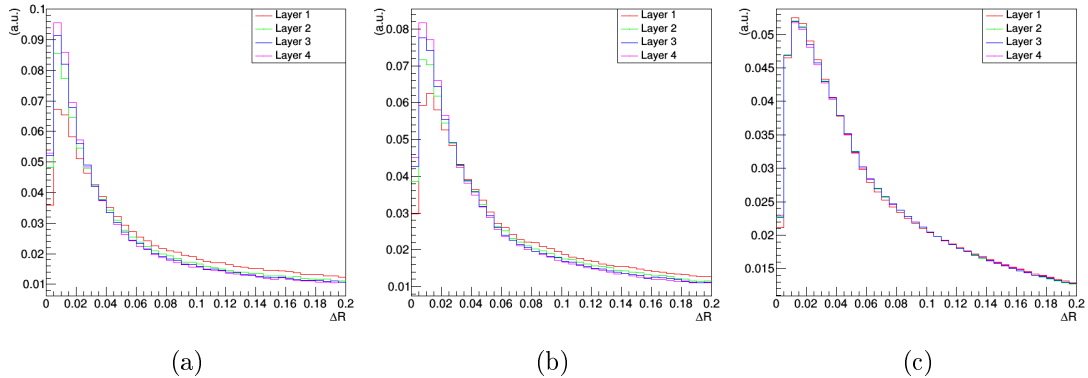


Figure 37: Normalized number of matched clusters at each value of  $\Delta R$  for each layer in the 2 TeV sample (a) the 4 TeV sample (b) and the background sample (c).



These figures show that a large part of the matched clusters would be cut off if a  $\Delta R$  value smaller than 0.04 was chosen. Also, for  $\Delta R$  values larger than 0.16, the functions for layer 3 and layer 4 become very small and increasingly flat in the case of the signal samples. Thus, the search for an optimal value of  $\Delta R$  is confined to a range from 0.04 to 0.16. In this range, the values 0.04, 0.06, 0.08, 0.1 and 0.16 were used for drawing ROC-curves of all possible discriminants consisting of ratios  $L_i/L_j$  (Figs. 38a to 38e) and differences  $L_i - L_j$  (Figs. 39a to 39e) in hits between later and earlier detector layers. Comparing these curves, the value of  $\Delta R < 0.1$  was found to yield the best performance for the  $L_i/L_j$  taggers and  $\Delta R < 0.04$  for the  $L_i - L_j$  taggers.

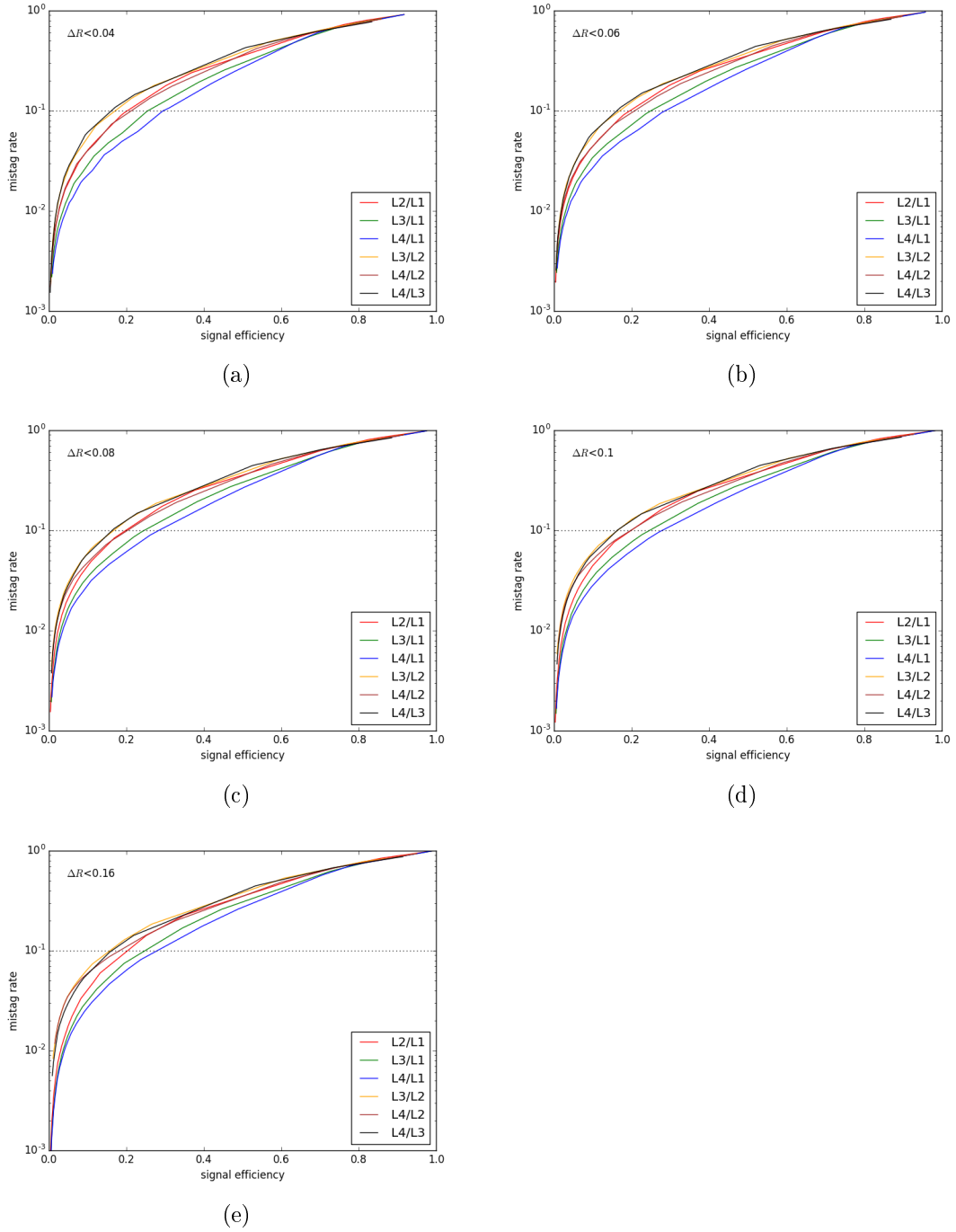


Figure 38: ROC curves comparing every combination  $L_i/L_j$  using different values of  $\Delta R$  for the cone size.

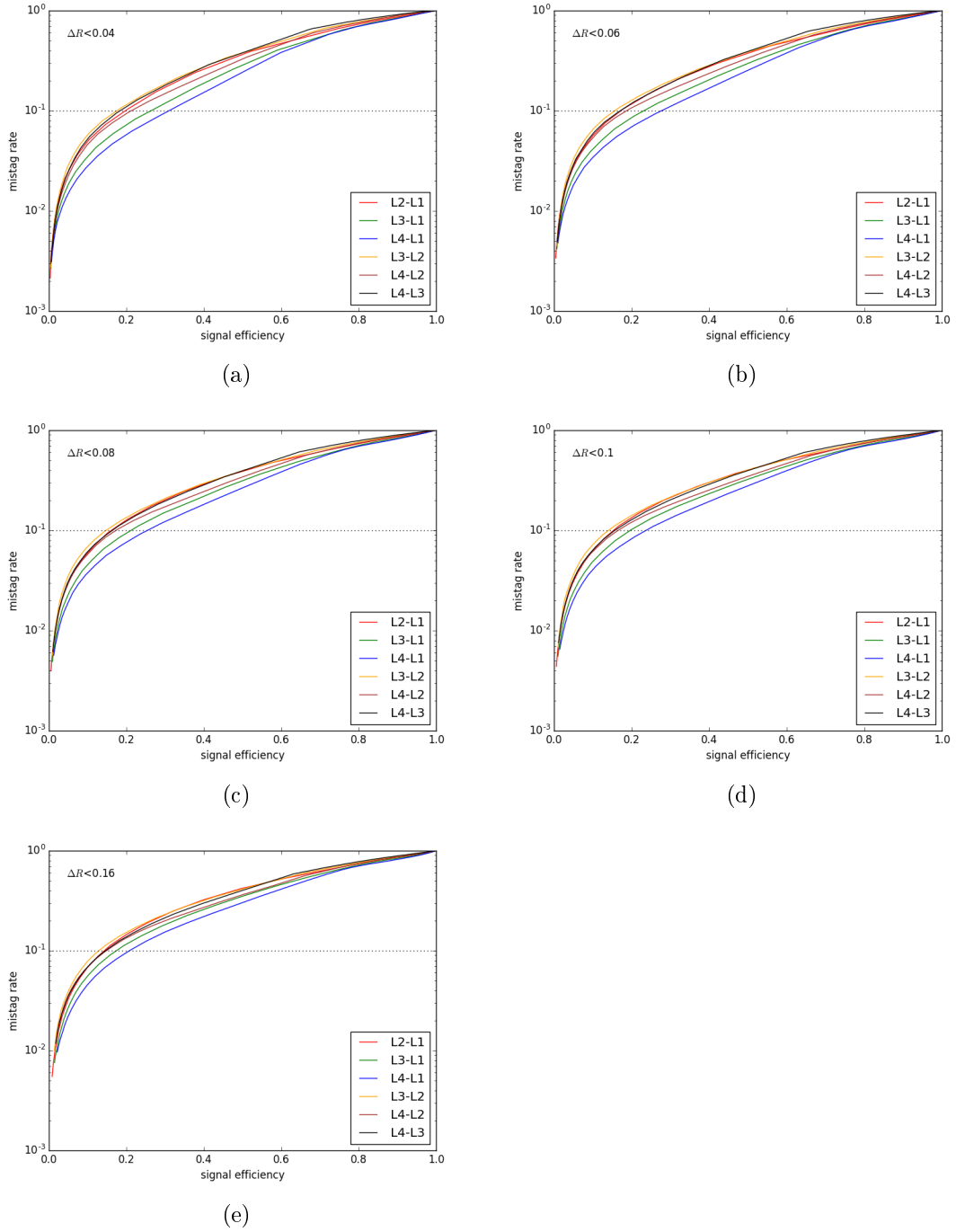


Figure 39: ROC curves comparing every combination  $L_i - L_j$  using different values of  $\Delta R$  for the cone size.

## References

- [1] J. Katharina Behr et al. “Boosting Higgs pair production in the  $b\bar{b}b\bar{b}$  final state with multivariate techniques”. In: (2015). DOI: 10.1140/epjc/s10052-016-4215-5. arXiv: 1512.08928. URL: <http://arxiv.org/abs/1512.08928><http://dx.doi.org/10.1140/epjc/s10052-016-4215-5>.
- [2] Maxime Gouzevitch et al. “Scale-invariant resonance tagging in multijet events and new physics in Higgs pair production”. In: (2013). DOI: 10.1007/JHEP07(2013)148. arXiv: 1303.6636. URL: <http://arxiv.org/abs/1303.6636>[http://dx.doi.org/10.1007/JHEP07\(2013\)148](http://dx.doi.org/10.1007/JHEP07(2013)148).
- [3] Johan Alwall, Philip Schuster, and Natalia Toro. “Simplified Models for a First Characterization of New Physics at the LHC”. In: (2008). DOI: 10.1103/PhysRevD.79.075020. arXiv: 0810.3921. URL: <http://arxiv.org/abs/0810.3921><http://dx.doi.org/10.1103/PhysRevD.79.075020>.
- [4] CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: (2017). DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158. URL: <http://arxiv.org/abs/1712.07158>.
- [5] B. Todd Huffman, Charles Jackson, and Jeff Tseng. “Tagging  $b$  quarks at extreme energies without tracks”. In: (2016). DOI: 10.1088/0954-3899/43/8/085001. arXiv: 1604.05036. URL: <http://arxiv.org/abs/1604.05036><http://dx.doi.org/10.1088/0954-3899/43/8/085001>.
- [6] Cian O’Luanaigh. *Proton beams are back in the LHC / CERN*. 2015. URL: <https://home.cern/about/updates/2015/04/proton-beams-are-back-lhc> (visited on 06/11/2018).
- [7] CERN. *Detector / CMS Experiment*. URL: <https://cms.cern/detector> (visited on 06/11/2018).
- [8] David Barney. *CMS Detector Slice*. 2016. URL: <https://cds.cern.ch/record/2120661>.
- [9] V. Veszpremi. “Performance verification of the CMS Phase-1 Upgrade Pixel detector”. In: *Journal of Instrumentation* 12.12 (2017), pp. C12010–C12010. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/12/C12010. URL: <http://stacks.iop.org/1748-0221/12/i=12/a=C12010?key=crossref.301b6490592416ad70aad4464829998a>.
- [10] C. Patrignani et al. (Particle Data Group). *Particle Data Group - 2017 Summary Tables*. 2016. URL: [http://pdg.lbl.gov/2017/tables/contents\\_tables.html](http://pdg.lbl.gov/2017/tables/contents_tables.html) (visited on 06/12/2018).
- [11] E. Norrbin and T. Sjöstrand. “Production and Hadronization of Heavy Quarks”. In: (2000). DOI: 10.1007/s100520000460. arXiv: 0005110 [hep-ph]. URL: <http://arxiv.org/abs/hep-ph/0005110><http://dx.doi.org/10.1007/s100520000460>.





- [12] M. Kreps and Y. Kwon. “Production and decay of b-flavoured hadrons”. In: *Particle Data Group* 0.March (2014), pp. 1–32. URL: <http://pdg.lbl.gov>.
- [13] Thomas G. Rizzo. “Z’ Phenomenology and the LHC”. In: (2006). arXiv: 0610104 [hep-ph]. URL: <http://arxiv.org/abs/hep-ph/0610104>.
- [14] ATLAS Collaboration. “Search for resonances in the mass distribution of jet pairs with one or two jets identified as *b*-jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: (2018). arXiv: 1805.09299. URL: <http://arxiv.org/abs/1805.09299>.
- [15] Torbjörn Sjöstrand. *PYTHIA 8 Documentation*. 2014. URL: <http://home.thep.lu.se/~simstorbjorn/pythia81html/Welcome.html> (visited on 08/17/2018).
- [16] V. Chiochia. “Experience with CMS pixel software commissioning”. In: August (2008), pp. 1–8. arXiv: 0812.0681. URL: <http://arxiv.org/abs/0812.0681>.
- [17] *Core Layers - Keras Documentation*. URL: <https://keras.io/layers/core/> (visited on 07/23/2018).
- [18] *Activations - Keras Documentation*. URL: <https://keras.io/activations/> (visited on 07/23/2018).
- [19] *Losses - Keras Documentation*. URL: <https://keras.io/losses/> (visited on 07/23/2018).
- [20] *Convolutional Layers - Keras Documentation*. URL: <https://keras.io/layers/convolutional/#conv2d> (visited on 07/23/2018).
- [21] Thomas Lenzi and Gilles De Lentdecker. “Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors”. In: (2013). arXiv: 1306.0858. URL: <http://arxiv.org/abs/1306.0858>.