Quark- Gluon discrimination for τ lepton identification

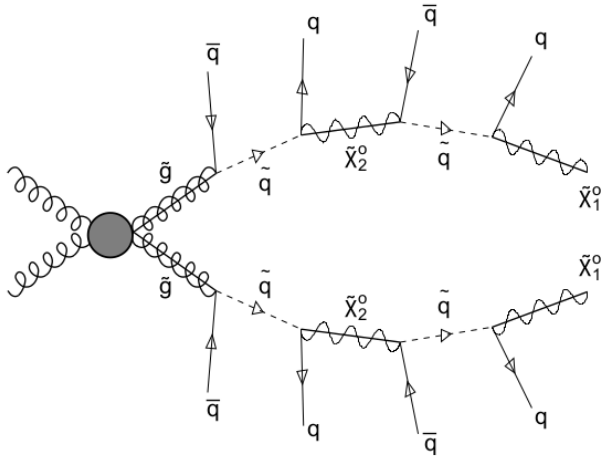
Marzieh Bahmani

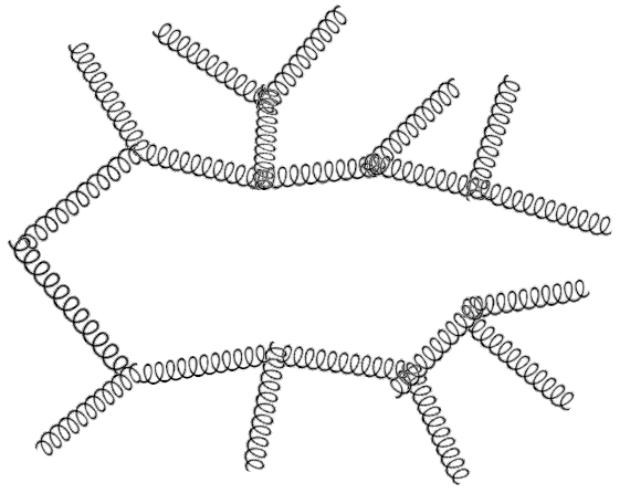
IFJ-PAN Krakow

October 2018

Introduction and Motivation

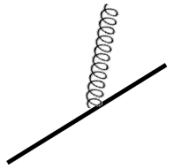
- The production of quarks and gluons via strong interactions is the dominant high-momentum-transfer process at the LHC.
- The power to discriminate between jets initiated by quarks and those initiated by gluons provide a tool in search for new physics.
- BSM New searches:
  New searches mostly quark jets Background mostly gluon jets





Quark and gluon initiated jets

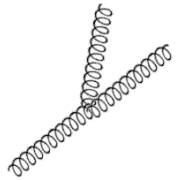
Probability of quark radiating:



$$P(q \to qg) = \frac{\alpha_s}{2\pi} C_F(\cdots)$$

$$C_F = \frac{4}{3} = 1.3$$

Probability of gluon radiating:

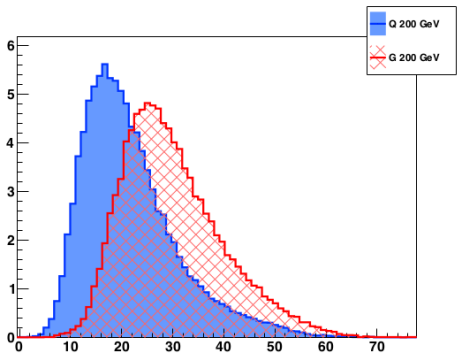


$$P(g \to gg) = \frac{\alpha_s}{2\pi} C_A(\cdots)$$

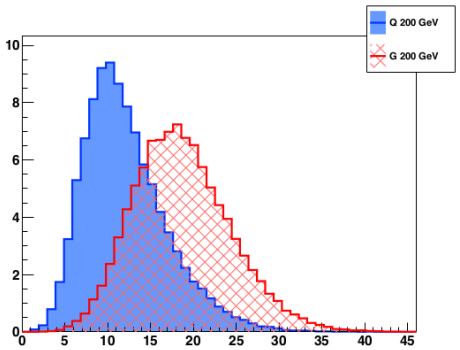
$$C_A = 3$$

- Gluons around twice as likely to radiate than quarks
- According to the charecteristic of quarks and gluons, one can find discriminant variables for distinguishing quark initiated jets from qluon initiated jet. Some example distributions :

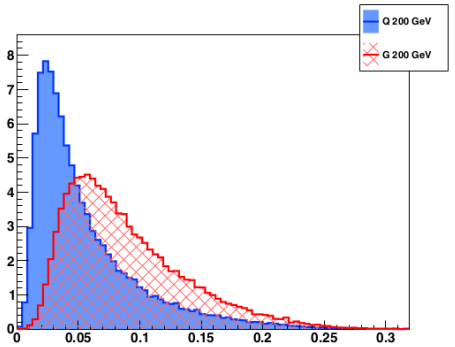
Jet mass



Charged particle count



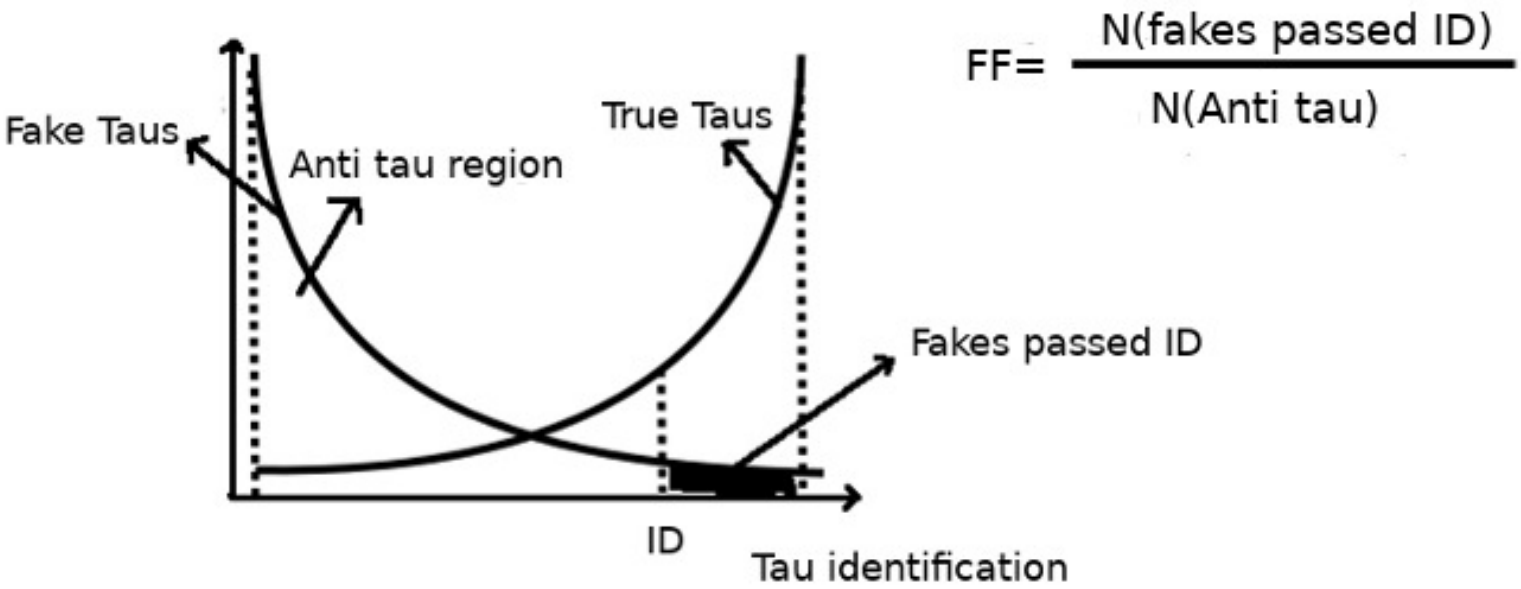
Jet broadening



τ Identification

- Jets genuinely fake hadronic τ and all the τ-related analyses suffer from such backgrounds.
- The probability for a hadronic jet to fake a τ depends on its origin.($\tau_{had-vis}$ jet width and Charged track multiplicity)
- There are different approaches for estimation of jet-to-tau misidentified hadronic τ decays: the fake factor method is using the help of discriminant for the origin of jets.
- The fake tau background is not well modeled by MC therefore, we developed data driven techniques.

Fake-Factor method: (FF determination)



$$FF = \frac{N(\text{fakes passed ID})}{N(\text{Anti tau})}$$

- The *tau* candidate matching a true hadronic *tau* dacay, an electron or a muon at generator level must be subtracted.
- Fake-factors are usually measured in bins.(e.g. p_T, number of tracks) They can also be measured in opposite- or same-sign regions, with or without b-jets, depending of the topology of interest in the analysis.

Considering q/g jet composition

- There can be one or several CR(s) where FFs are measured, for one CR, one must ensure that the fake τ composition is close to the one in the signal region (SR). Otherwise, one should measure FFs in several CRs that have different fake τ compositions and then combine them.
- Usually, FFs are measured in CRs enriched in either gluon-initiated or quark-initiated jets
- In the case where two (or more arXiv:1808.00336) CRs are used, and if one is enriched in gluon-initiated jets, FF for each bin:

$$FF = \alpha_g \times FF(g) + [1 - \alpha_g] \times FF(other(s))$$

In that case, one only needs to compute the fraction of gluon-initiated jet events in the SR-like anti-τ region

Application of FFs

- Define an anti-τ region, which is similar to the signal region but where a τ candidate fails the ID-requirement, instead of fulfilling it.
- In a bin i, the number of events with a j$\rightarrow\tau$ fake is

$$N^{\tau}_{fakes}(i) = N^{anti-\tau}_{fakes}(i) \times FF(i),$$

$$N^{anti-\tau}_{fakes}(i) = N^{anti-\tau}_{fakes}(data, i) - N^{anti-\tau}_{fakes}(MC, \tau \neq j, i)$$

Example of $H^{\pm} \rightarrow \tau\nu$ analysis

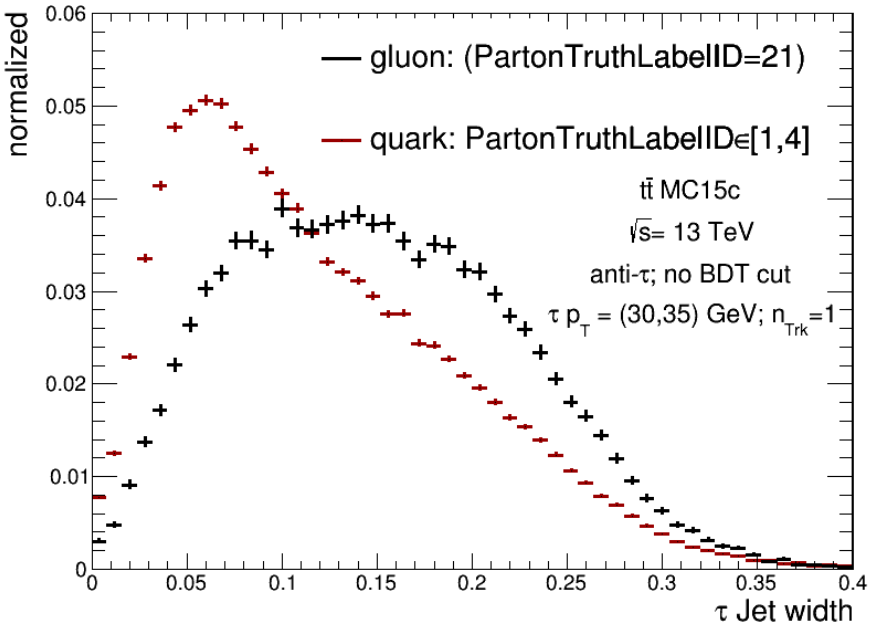
- Two control region with different jet compositions are used in order to determine the rate of the fake $\tau_{had-vis}$ objects.
  1. Multi-jet CR (dominated by gluon-initiated jets)
  2. W+jet CR (dominated by quark-initiated jets)
- In the anti-$\tau_{had-vis}$ regions, the fractions of quark- and gluon-initiated jets misidentified as $\tau_{had-vis}$ candidates are measured using a template-fit approach, based on variables that are sensitive to the difference in quark- and gluon-fractions between these two types of jets

Combined Fake Factor in $H^{\pm} \rightarrow \tau\nu$ analysis

- Chosen variables : $\tau_{had-vis}$ identification BDT output score for 3-track and the $\tau_{had-vis}$ jet width for 1-track
  τ jet width definition

$$w_{\tau} = \frac{\Sigma[p_T^{track} \times \Delta R(\tau_{had-vis}, track)]}{\Sigma p_T^{track}}$$

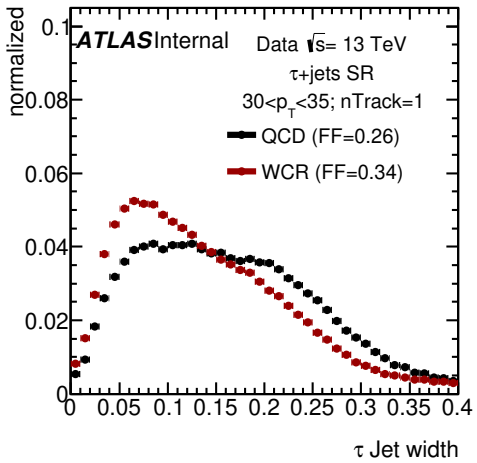
where $\Delta R < 0.4$

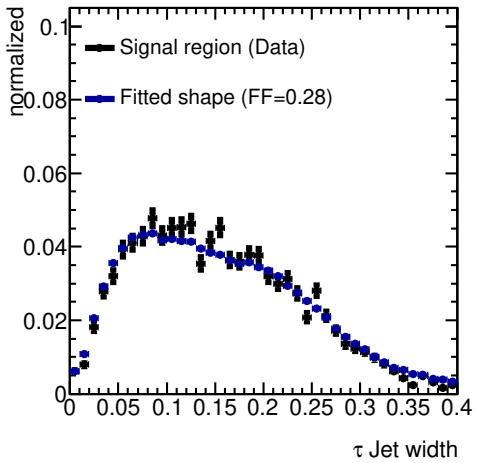


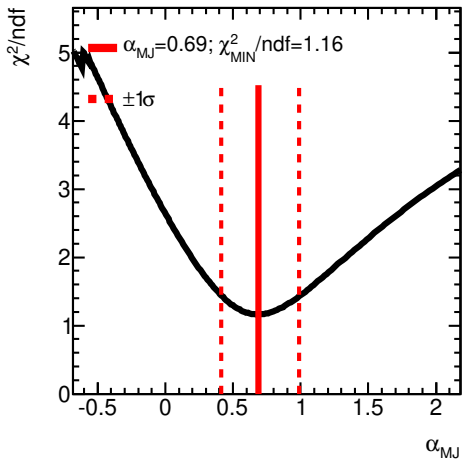
- For each bin, two binned templates, denoted f_{MJ} (Multijet CR) and f_W (W+jet CR) , are obtained in their corresponding CRs.
- Their fractional contribution in the SR is determined using a template fit to the respective distributions in the anti tau SR:

$$f(x|\alpha_{MJ}) = \alpha_{MJ} \times f_{MJ} + (1 - \alpha_{MJ}) \times f_W$$

- α_{MJ} is a free parameter.





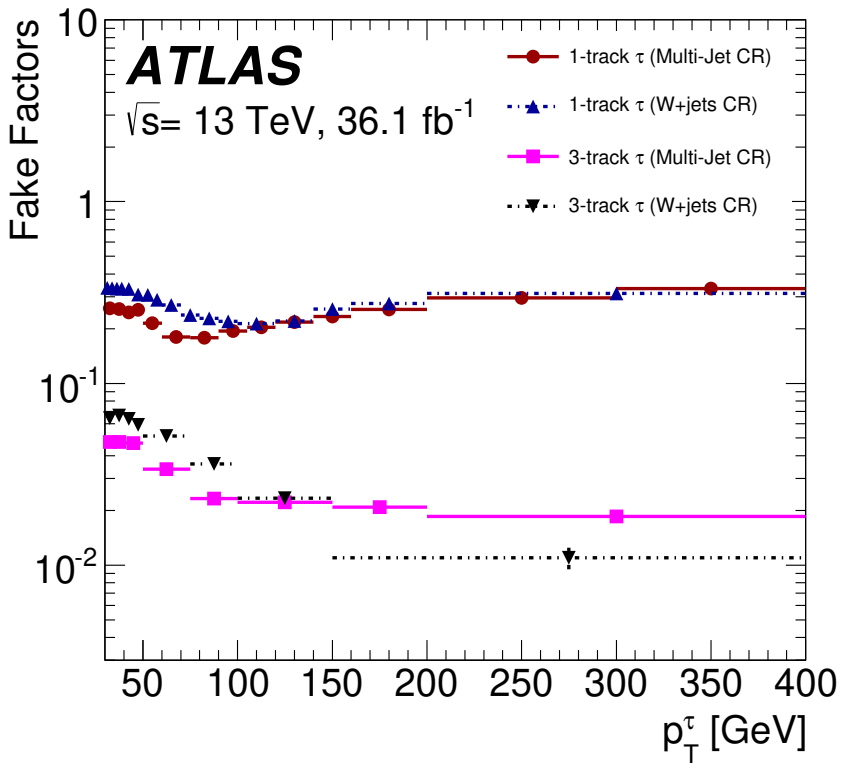


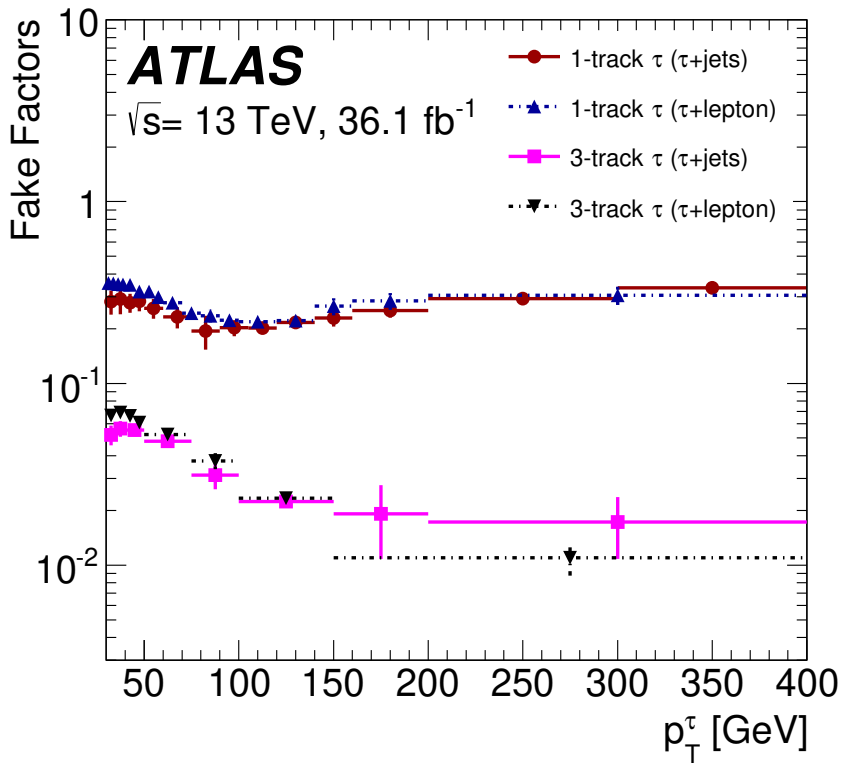
- From the best fit values of α_{MJ} , combined FF are given by :

$$FF^{comb}(i) = \alpha_{MJ}(i) \times FF^{MJ} + (1 - \alpha_{MJ}(i)) \times FF^{W}$$

Fake Factors from $H^{\pm} \rightarrow \tau\nu$ analysis

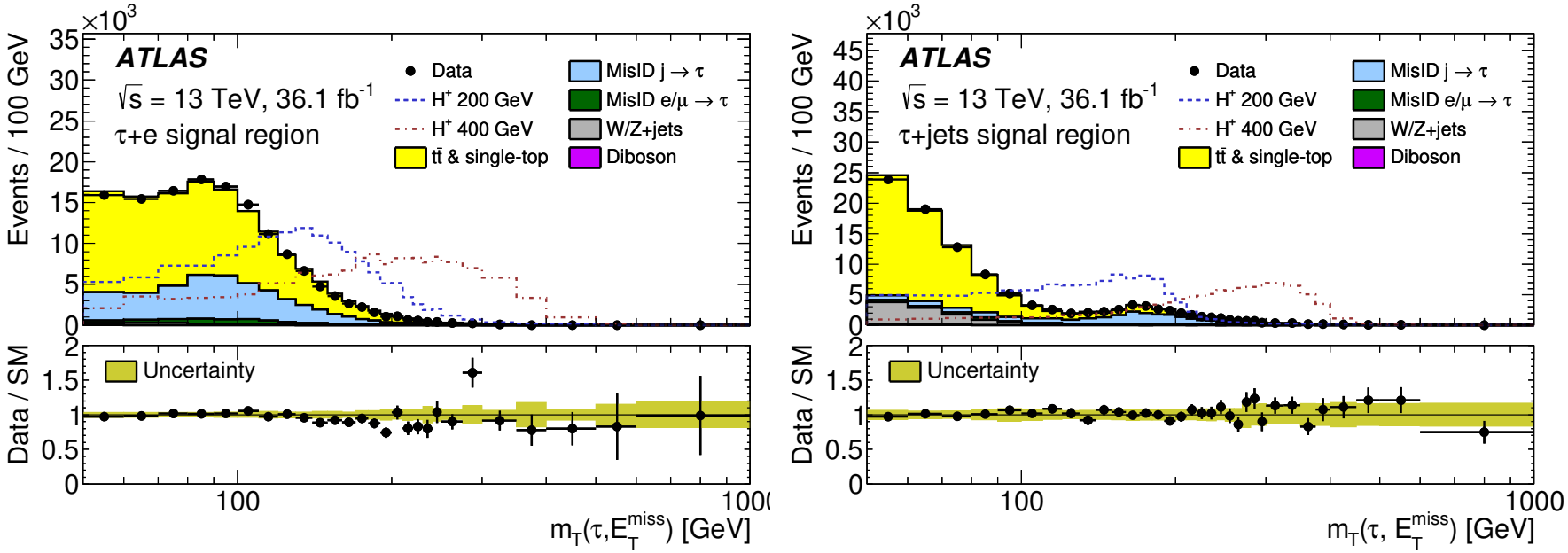
Fake factors parameterized as a function of p_T^{τ} and number of tracks, in the left plot in the multi-jet and w+jet CRs and errors represent the statistical uncertainties, in the right plot after reweighting by α_{MJ} in the $\tau_{had-vis}$+jets and $\tau_{had-vis}$+lepton channel, and it is with additional systematic uncertainties obtained from the combination in a given p_T^{τ} bin.





Validation of the background modelling in $H^{\pm} \to \tau\nu$ analysis

Distribution of m_T ($\tau_{had-vis}$,E_T^{miss}) in the two signal regions, (a) $\tau_{had-vis}$+electron, (b) $\tau_{had-vis}$+jets



Systematic Uncertainties $H^{\pm} \to \tau\nu$

Source of systematic uncertainty	Impact on the expected limit (stat. only) in % $m_{H^+} = 170\ GeV$	$m_{H^+} = 1000\ GeV$
Experimental		
luminosity	2.9	0.2
trigger	1.3	<0.1
$\tau_{\text{had-vis}}$	14.6	0.3
jet	16.9	0.2
electron	10.1	0.1
muon	1.1	<0.1
$E_{\text{T}}^{\text{miss}}$	9.9	<0.1
Fake-factor method	20.3	2.7
Υ modelling	0.8	–
Signal and background models		
$t\bar{t}$ modelling	6.3	0.1
W/Z+jets modelling	1.1	<0.1
cross-sections ($W/Z/VV/t$)	9.6	0.4
H^+ signal modelling	2.5	6.4
All	52.1	13.8

The dominant sources of systematic uncertainty of Fake factor method:

- The requirement $\tau_{had-vis}$ BDT output score in the anti-$\tau_{had-vis}$ definition.
- The contamination of true $\tau_{had-vis}$ candidates fulfilling the anti-$\tau_{had-vis}$ selection (varied by 50%).
- The statistical uncertainty of the control sample.
- The statistical error on the best-fit value of α_{MJ}

Summary

- The power to discriminate between jets initiated by quarks and those initiated by gluons provide a tool in search for new physics.
- The fake factor method is using the help of discriminant for the origin of jets, for identifying the jet misidentified as τ in tau related analysis.
- The example of Fake Factor in $H^{\pm} \rightarrow \tau\nu$ were presented.

Additional material

τ jet width for high P_T