



# JVF Efficiency for b-Jets and light Jets in top events

Paola Ruggiero, University of Pisa, Italy

September 8, 2012

## Abstract

The DESY ATLAS group is doing a measurement of additional particle jets in events with top-anti-top quark pairs at the LHC. This measurement is important to understand QCD at hard scales and as background for Higgs searches and searches for new physics. With current LHC luminosities ( $10^{30} \text{cm}^{-2} \text{s}^{-1}$ ), there is a considerable effect of pile-up from several proton-proton interactions occurring in the same bunch crossing. This effect has to be accounted for in the measurement. In particular, there is a need to distinguish jets coming from the vertex of the main hard interaction from those of pile-up ones. One way to do this is to make a cut on the jet vertex fraction (JVF). The efficiencies of this cut need to be understood and corrected for. In particular, in this project we will study the efficiency of the JVF cut on jets stemming from bottom quarks (so-called *b*-jets) and for *light* jets (*u*, *d*, *s* flavour jets, including gluons).

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Theoretical overview</b>	<b>4</b>
2.1	ATLAS detector: a short description . . . . .	4
2.2	Some more definitions... . . . .	5
2.3	JVF alghoritm . . . . .	6
<b>3</b>	<b>Our data sample: some details</b>	<b>8</b>
3.1	Monte Carlo $t\bar{t}$ events . . . . .	8
3.2	Reco-level event selection . . . . .	8
<b>4</b>	<b>JVF distribution</b>	<b>10</b>
4.1	$\Delta R$ matching and JVf distribution for all jets . . . . .	10
4.2	JVF distribution for $b$ -jets and light jets . . . . .	11
<b>5</b>	<b>JVF efficiency and rejection plots</b>	<b>14</b>
5.1	Definition and results for all jets . . . . .	14
5.2	Efficiency and rejection for light jets . . . . .	15
5.3	Efficiency for $b$ -jets . . . . .	16
5.4	Efficiency with respect to other variables . . . . .	17
<b>6</b>	<b>Conclusion</b>	<b>20</b>

# 1 Introduction

The aim of the project is the JVF efficiency study, where the Jet Vertex Fraction is a new variable (we'll define in the following section), that ATLAS would like to implement in future and that can be used in order to determine the origin of a certain jet: this concerns with the general issue of having a clearer and more detailed vision of events we study at LHC.

Working with Monte Carlo simulations of  $t\bar{t}$  semileptonic events, we first checked how this variable works in general with jets: we had a look to the JVF distribution for jets, independently of the flavour of the quark from where they come from. And then we made look at the efficiency and the rejection as well, having in mind the idea of finding the best cut on JVF, optimizing both.

In particular we were interested in understanding how the JVF variable works for  $b$ -jets and *light* jets (meaning  $u$ ,  $d$  and  $s$  quarks, including gluons). So, we repeated the same study, after a selection on flavour, separately for  $b$ -Jets and *light* jets.

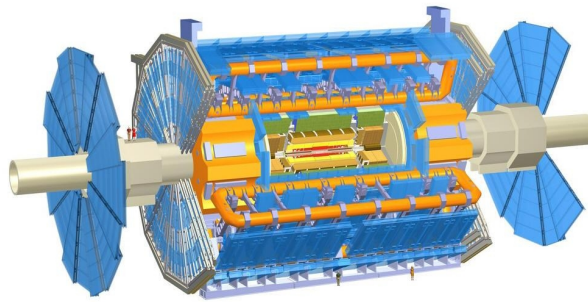
In the following sections, after a short theoretical introduction, including some definition to become familiar with this topic, we'll come to discuss some results, explaining how we got them and trying to draw a conclusion from them.

## 2 Theoretical overview

### 2.1 ATLAS detector: a short description

ATLAS is a particle physics experiment at the Large Hadron Collider at CERN. The ATLAS detector is searching for new discoveries in the head-on collisions of protons of extraordinarily high energy.

One of the most important goals of ATLAS is to investigate a missing piece of the Standard Model, the Higgs boson in addition to the asymmetry between the behavior of matter and antimatter, known as CP violation. The properties of the top quark are also object of study and further lines of investigation are those searching directly for new models of physics. A short description of the detector is needed before going into details of our project.



The ATLAS detector consists of a series of ever-larger concentric cylinders around the interaction point where the proton beams from the LHC collide. It can be divided into four major parts: the Inner Detector, the calorimeters, the Muon Spectrometer and the magnet systems. Each of these is in turn made of multiple layers. The detectors are complementary:

- the *Inner detector* measures the momentum and giving tracks of each charged particle
- the *Calorimeter* measures the energies carried by the particles
- the *Muon spectrometer*, identifying and measures the momenta of muons
- the *Magnet system* bends charged particles, allowing for momentum measurement.

The only established stable particles that cannot be detected directly are neutrinos; their presence is inferred by measuring a momentum imbalance among detected particles.

The coordinate system and nomenclature used to describe ATLAS detector and the particles emerging from  $p - p$  collisions are briefly summarized here:

- the positive x-axis is defined as pointing from the interaction point to the center of the LHC ring

- the positive y-axis is defined as pointing upwards
- the azimuthal angle  $\varphi$  is measured as usual around the beam axis
- the pseudorapidity  $\eta$  is related to the polar angle  $\theta$ , that is the angle from the beam axis, by the relation  $\eta(\theta) = -\ln(\tan(\theta/2))$ .

## 2.2 Some more definitions...

When we talk about an "event", we mean all the information emerging from the collision point every 25 ns (actually until now ATLAS is running with 50 ns). But in reality we will have a main collision point, but also many other softer p-p interaction will occur in the same bunch crossing. For this reason, there are many vertexes associated to each event, so that, among them, we can distinguish the hardest ones (i.e. "hard scatter vertex") from the softer ones (so called "pile-up vertex").

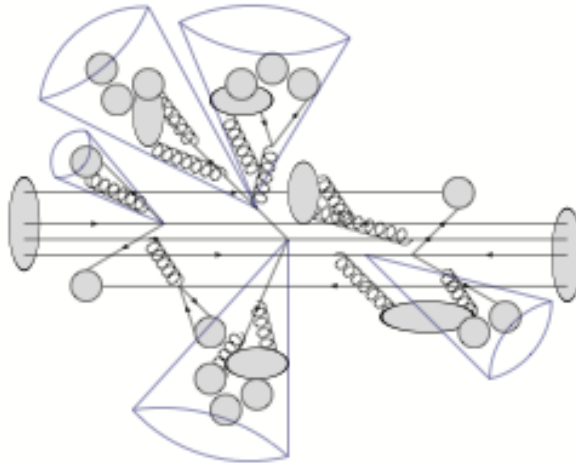


Figure 1: A typical event including HS and PU jets in the final state

And this also allow for a classification of jets into:

- *Hard Scatter Jets*: Jets coming from the main interaction (primary vertex)
- *Pile-up Jets*: Pile-up refers to the presence of a signal in the detector which originates from an interaction other than the hard-scattering event which triggers the readout. It therefore consists primarily of two components:
  - "in time" pile-up, referring to multiple proton-proton interactions occurring simultaneously within a single event
  - "out-of-time" pile-up, referring to contributions of previous bunch-crossing due to relatively large calorimeter integration time<sup>1</sup>.

<sup>1</sup>The reason for choosing quite a short integration time is because we also need to maximize the rate  $R = L \cdot \sigma$ , where  $L$  is the luminosity and  $\sigma$  the crosssection.

## 2.3 JVF algorithm

In order to take into account the effect of pile-up, there is a need to distinguish pile-up jets from the hard scatter ones. If we know which is the main vertex, this distinction can be done by associating a jet to this vertex. So what we need is an Algorithm for Jet-Vertex association.

The JVF algorithm (the one we consider in this project) makes use of tracks in order to make this association: tracks are matched to each jet and used to calculate the fraction of track momentum from each PV <sup>2</sup>.

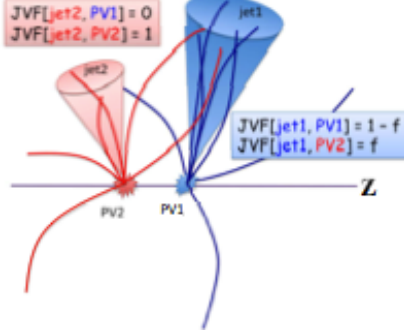


Figure 2: *Schematic representation of the jet-to-vertex association provided by the jet-vertex fraction (JVF) discriminant*

The resulting discriminant is termed the *jet-vertex fraction* ( JVF ). This variable measures the probability that a jet originated from a particular vertex and it is defined to be the fraction of each jet's constituent transverse track momentum contributed by each vertex. More specifically, it's the sum  $p_t$  of all matched-tracks from a given vertex divided by the total-jet-matched track  $p_t$  from all vertices; for a single jet  $jet_i$  the JVF with respect to the vertex  $vtx_j$  within an event it is:

$$JVF(jet_i, vrt_j) = \frac{\sum_k p_t(trk_k^{jet_i}, vtx_j)}{\sum_n \sum_l p_t(trk_l^{jet_i}, vtx_n)} \quad (1)$$

In Figure 3 a typical JVF distribution (referred to the hard-scatter vertex, as selected by the default  $max(\sum p_t^2)$  criterion) is shown: 4 distinct components are labeled:

---

<sup>2</sup>primary vertex.

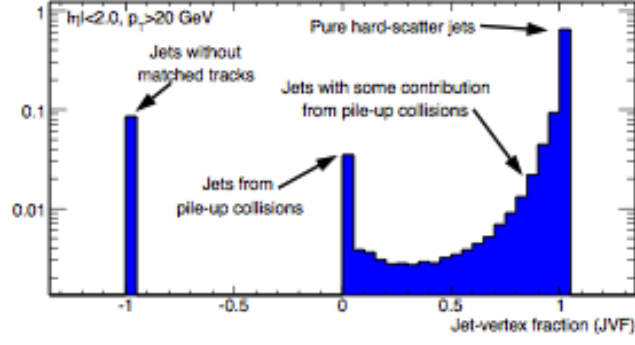


Figure 3: *Typical JVF distribution*

- The first is the peak at  $JVF = 1$ , corresponding to jets that are very likely to come from the HS vertex;
- The second component consist of jets with  $JVF = 0$ , for which matched tracks did not point back to the selected hard-scatter vertex, but to an additional minimum bias vertex;
- The third component is the distribution of jets having  $0 < JV F < 1$ , that is for jets with some contribution from pile-up collisions;
- The last component is for calorimeter jets which have not been matched to track and are therefore assigned with a value of  $JVF = -1$  .

## 3 Our data sample: some details

### 3.1 Monte Carlo $t\bar{t}$ events

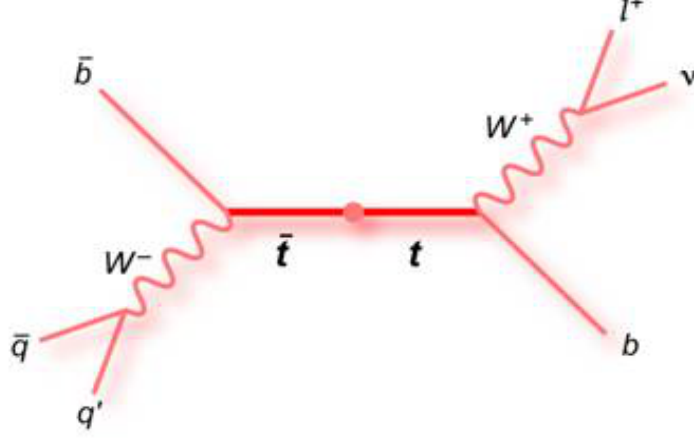


Figure 4:  $t\bar{t}$  semileptonic event

For this project we worked with simulated Monte Carlo data of semileptonic  $t\bar{t}$  events. In Figure 5 is shown what a  $t\bar{t}$  semileptonic event looks like: assuming to have a  $t\bar{t}$  quark pair after a proton-proton collision, a  $t$  quark will decay with a probability greater than 99% into a  $b$ -quark and a  $W$ -boson, so that in the final state we expect to see:

- 2  $b$ -jets, coming from the hadronization of the  $b$ -quarks
- 2 more jets from the decay of a  $W$  boson.

Our simulation comes from the multi-leg LO eventgenerator AlpGen: it simulates events only up to the leading order, whereas, for instance there are other event generators, such as MC@NLO, simulating up to next to leading order; but AlpGen is supposed to work better for describing additional jets, making use of matrix elements approach to perturbative QCD, instead of the pure parton showers one<sup>3</sup>.

### 3.2 Reco-level event selection

The Monte Carlo simulation will give a variegated spectra of events, corresponding to different decay channels for  $t\bar{t}$ : fully hadronic channel, leptonic channel, semileptonic

---

<sup>3</sup>One can use the corresponding matrix elements to simulate multi-jets, even if they are in general available at leading order in  $\alpha_S$  and don't allow for a full simulation of the final state, not taking into account hadronization. That's why any realistic model will include parton showering (even if a pure parton shower approach would on the other hand give a poor simulation of configurations). Regarding AlpGen, it makes use of both parton showers and Matrix Elements for extra jets, using a matching scheme for avoiding a potential double counting of jet configurations.



channel. So that we need a preselection in order to focus on semileptonic  $t\bar{t}$  events only. In this preselection, also several other standard requirements are included; just the most important ones are listed below <sup>4</sup>. We actually required for each event:

- *Transverse momentum*: trigger on high- $p_t$  lepton
- *Vertex*: at least 4 tracks for the first primary vertex
- *Leptons*: exactly 1 lepton with energy over 25 GeV and not other leptons over 15 GeV
- *Jets*: at least 3 jets
- *Missing tranverse energy*: over 30 GeV
- *W transverse mass*: over 35 GeV
- *b-Jets*: at least 1 jet b-tagged <sup>5</sup>.

---

<sup>4</sup>Notice that there are no JVF requirements, of course, and this is because JVF is the variable we want to study the efficiency of.

<sup>5</sup>Even if from semileptonic channel, we expect to see 2 b-jets.

## 4 JVF distribution

### 4.1 $\Delta R$ matching and JVF distribution for all jets

The efficiency of JVF variable, that we want to study, will depend for sure on how we define a jet to be an HS one. There are different way to do this, for example looking into the content of a jets, checking the tranverse momentum value for each track and so on. But the simlest one (that is also the one we used in this project) is the  $\Delta R$  matching: essentially, we calculate the  $\Delta R$ , defined as :

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

between the jets coming from our montecarlo simulations (that here we call *TruthJets*) and the ones reconstructed in the detector (*RecoJets*). According to our definition, a *RecoJet* will be selected as an HS one if it has a *TruthJet* closeby (which means with  $\Delta R < 0.3$ ). In Figure 5 the distribution of  $\Delta R$  between a *RecoJets* and the closest true one is shown: the ones selected as HS are about the 98% of the total.

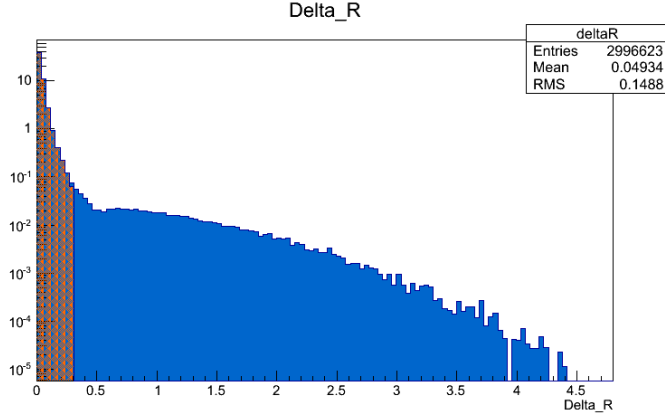


Figure 5:  $\Delta R(TruthJets, RecoJets)$

So now, after isolating contributions from pile-up, we plotted the JVF distribution, separately for HS and PU jets. The plot shows the expected behaviour since actually the most of HS jets are associated to high JVF value, whereas the most of PU jets have low JVF. Anyway, it is also to be noticed a peak at  $JVF = 1$  for pile-up. We had not enough time to really understand the reason, but, just making some guess, it could be because we actually didn't select the right HS vertex, so that we have the value of JVF referred to the wrong one. Or maybe the matching method implemented here ( $\Delta R$ ) is too simple to work properly or the cut on  $\Delta R$  itself too strict (and in fact raising the cut to the value 0.4, the peak decreases). But another good explanation could be the following one: also for *TruthJets* there threshold on several variables are set (for instance an upper limit on  $p_t$  or a lower limit on  $|\eta|$  and so on), so that it can happen that a *TruthJet* (i.e. coming from the HS vertex that means with  $JVF \approx 1$ ) having

some variables value under threshold is reconstructed over threshold, and in this case it will have no matching.

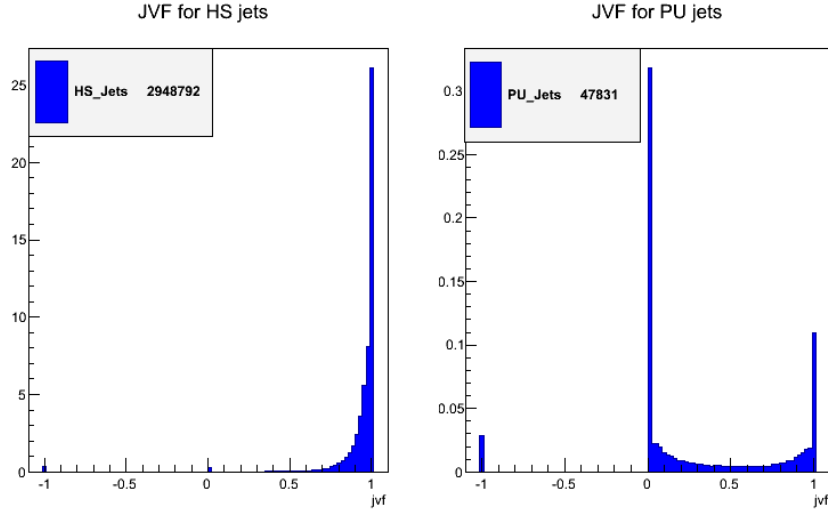


Figure 6: *JVF distribution for  $b$ -jets*

## 4.2 JVF distribution for $b$ -jets and light jets

In order to better understand the behaviour of JVF for different flavour jets, we plotted the JVF distribution after the same  $\Delta R$  matching separately for  $b$ -jets and light jets.

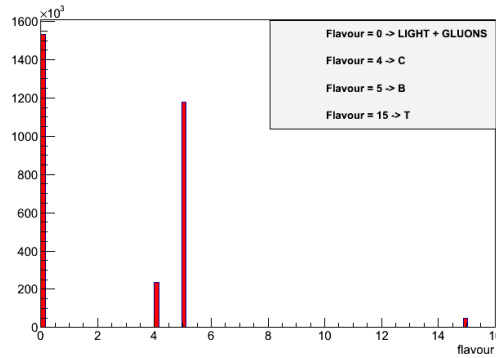


Figure 7: *Flavour distribution*

So, how to select them? We used information taken from Monte Carlo truth information<sup>6</sup>, stored in a variable (Figure 7).

---

<sup>6</sup>coming from Athena.

The result is shown in Figure 8 and Figure 9.

As we can see the shape of JVF distribution for light jets look very similar to the one we had for all jet, independently of the flavour, especially if we look to the distribution for pile-up and to the number of entries for them: this is what we expect, since, of course, the most of pile-up are supposed to be light.

Whereas, for  $b$ -jets, the JVF distribution for HS jets seems to be consistent, but not the one for PU jets, because actually we don't expect to have  $b$ -jets at all in pile-up (the probability of producing a  $t$ -quark in minimum bias is several orders of magnitude smaller than in the HS). The explanation for this sample of PU  $b$ -jets we got could be the same we mentioned for all jets, but here it can really be considered a small effect (smaller than 0.2 % since the ratio (HS  $b$ -Jets / PU  $b$ -Jets) is very small as well).

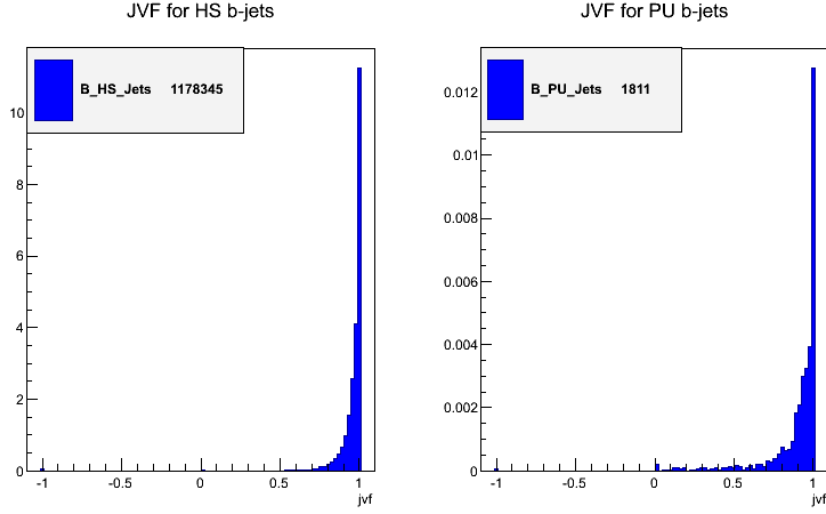


Figure 8: *JVF distribution for  $b$ -jets*

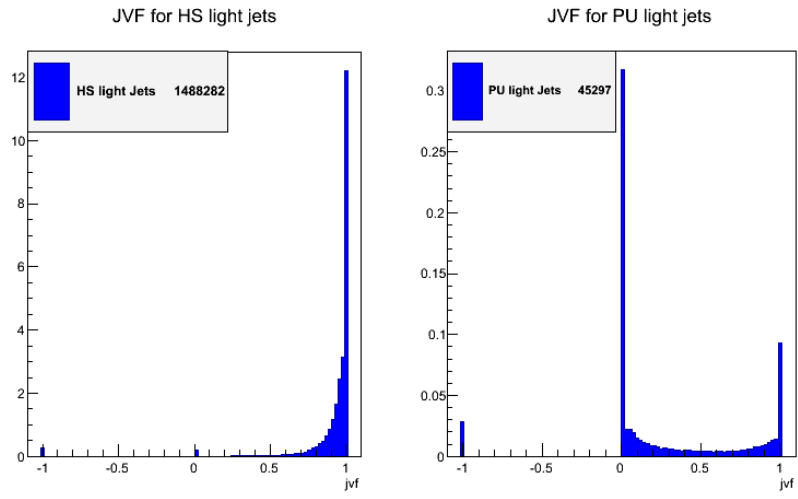


Figure 9:  $JVF$  distribution for light jets

## 5 JVF efficiency and rejection plots

### 5.1 Definition and results for all jets

To have a clearer drawing of how our results change by applying different cuts on JVF, we would like to have a look into the efficiency (and rejection as well) as a function of the JVF cut.

So, we can define the JVF efficiency to be the number of HS jets passing the cut on JVF, over all HS jets:

$$E_{JVF} = \frac{HSJets(JVF > cut)}{HSJets} \quad (3)$$

Whereas another important information is contained into the JVF rejection, that is basically related to how many "bad" jets are still among the "good" ones after our cut on JVF. It's defined as:

$$R_{JVF} = 1 - \frac{PUJets(JVF > cut)}{PUJets} \quad (4)$$

JVF efficiency and rejection are defined to be kind of complementary: when the efficiency decreases, the rejection will rise, so that, by comparison, we should be able to estimate the best compromise between them, aiming to maximize both .

Below, we did it first for all jets:

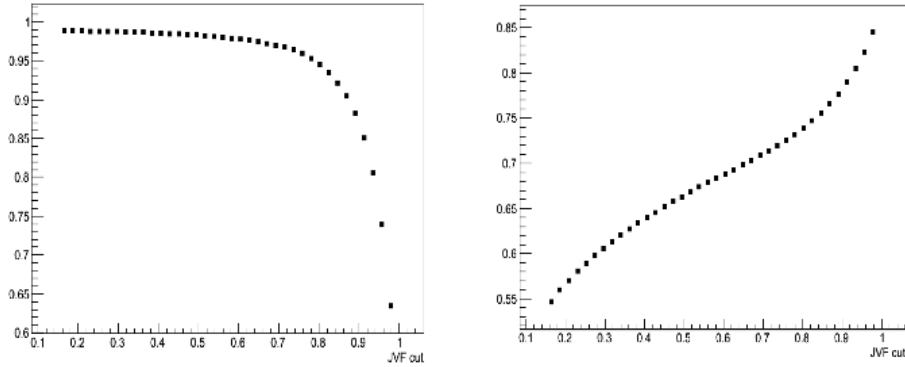


Figure 10: *Efficiency (a), Rejection (b)*

So this comparison can be performed by plotting rejection versus efficiency, as shown in Figure 11.

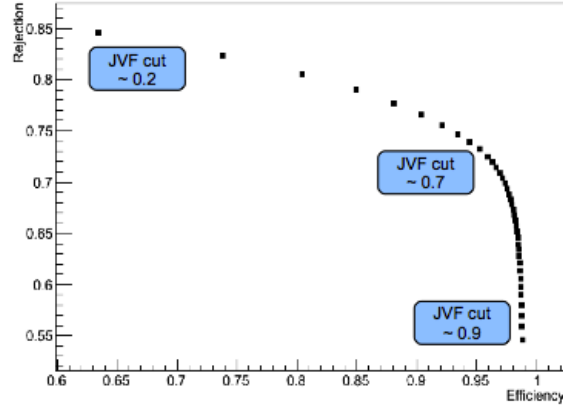


Figure 11: *Efficiency versus rejection*

The extimation for the best compromise between them we extrapolate maximizing the product (Efficiency\*Rejection), is shown in the table below for all jets:

JVF cut	Efficiency	Rejection
0.803	0.94	0.70

## 5.2 Efficiency and rejection for light jets

We also performed the same plots separately for light jets, as shown in Figure 12 and Figure 13, coming to similar results for the best JVF cut.

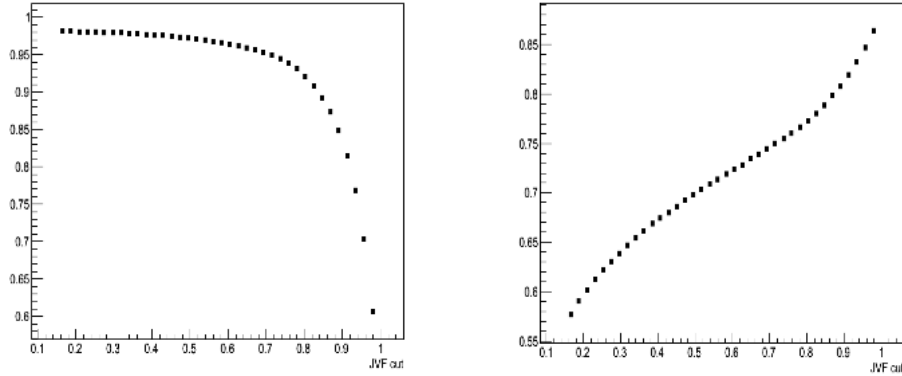


Figure 12: *Efficiency (a), Rejection (b) (light jets)*

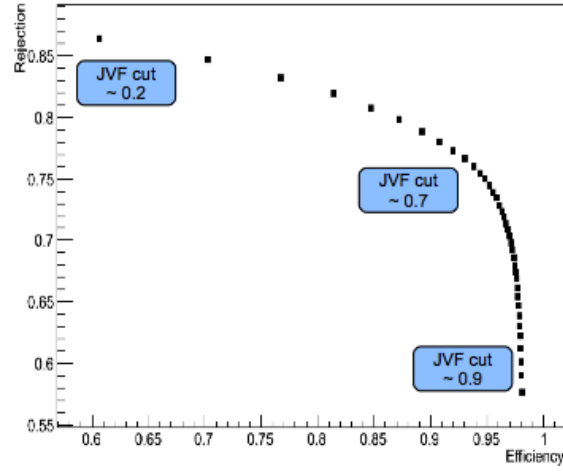


Figure 13: *Efficiency versus rejection (light jets)*

JVF cut	Efficiency	Rejection
0.781	0.93	0.76

### 5.3 Efficiency for $b$ -jets

For  $b$ -jets, instead, we just plotted the efficiency as a function of the JVF cut, since, as said before, it makes no sense to define rejection for PU  $b$ -jets, since we don't expect to have them at all.

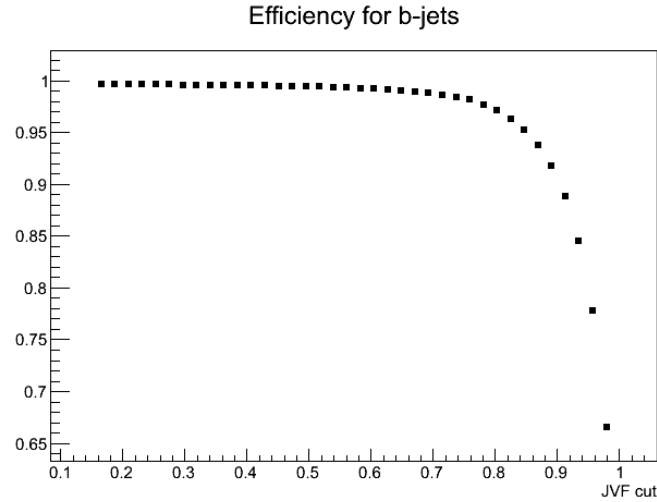


Figure 14: *JVF efficiency for  $b$ -jets*



## 5.4 Efficiency with respect to other variables

Finally, in order to better understand the meaning of different cuts on JVF, we also studied the JVF efficiency with respect to other variables, such as the tranverse momentum  $p_t$  and the pseudorapidity  $\eta$ .

In figures below the results for 3 different cuts on JVF are shown, respectively for  $b$ -jets and light jets.

We found, as expected, that, for example, for the  $p_t$ -efficiency: it decreases, raising the value of the JVF cut, and we also could guess why this effect is more visible at low  $p_t$  values; in fact, by definition, JVF is lower for less-energetic jets.

And also about the  $\eta$ -efficiency: it decreases, cutting more on JVF and this time the effect is harder for high value of  $\eta$  and this is because  $|\eta|$  covarage is different for Tracking (where the information about  $\eta$  comes from) and Calorimetry (where the information about JVF comes from).

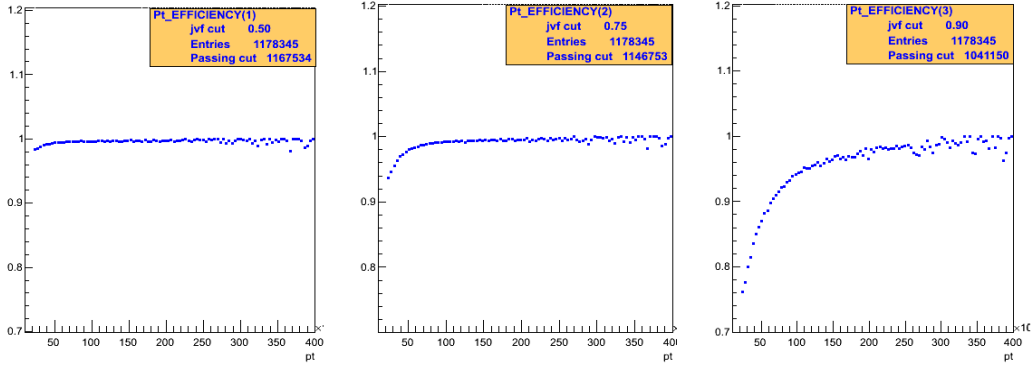


Figure 15:  $p_t$ -efficiency for  $b$ -jets with 3 different JVF cuts

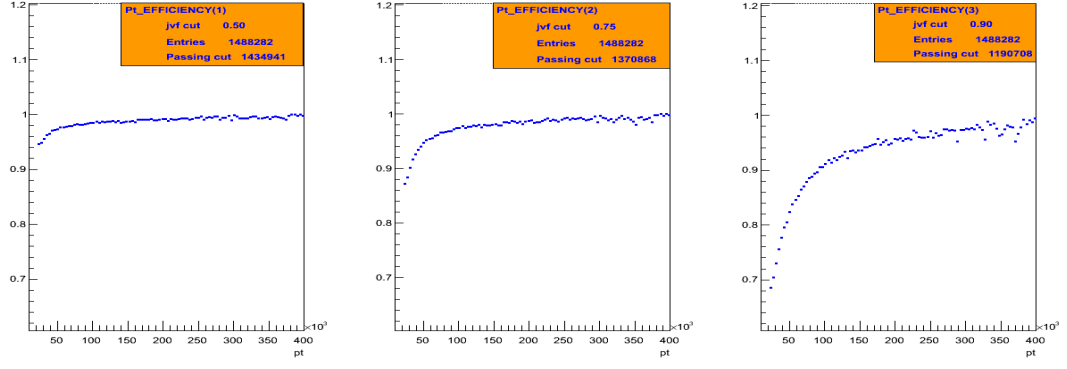


Figure 16:  $p_t$ -efficiency for light jets with 3 different JVF cuts

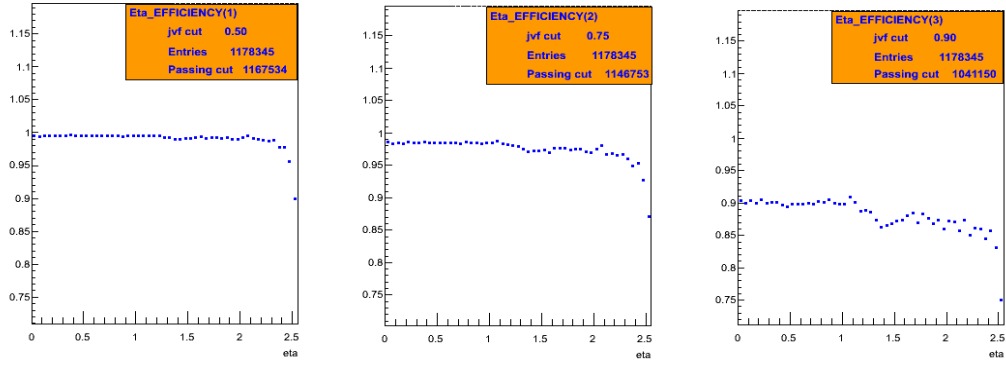


Figure 17:  $\eta$ -efficiency for  $b$ -jets with 3 different JVF cuts

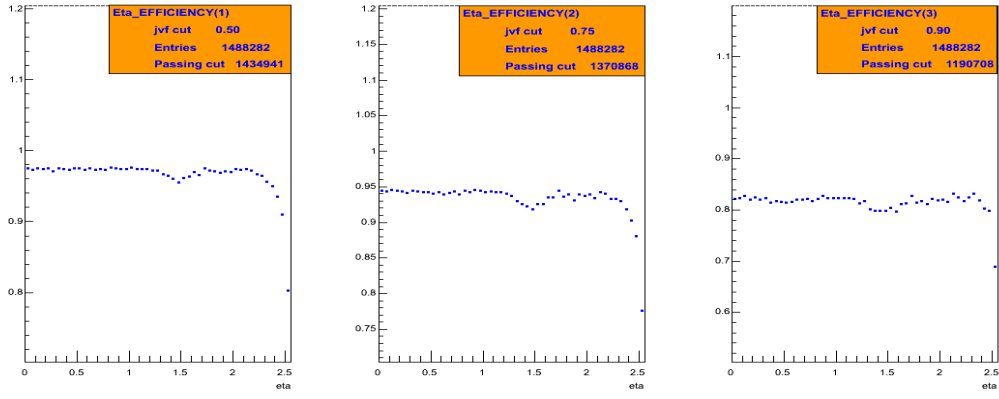


Figure 18:  $\eta$ -efficiency for light jets with 3 different JVF cuts

Having more time, it would have been also interesting to study also the rejection referred to these other variables, and read also from here some more information about the best JVF cut.

## 6 Conclusion

We have a new method for jet-identification, needing to be investigated more.

So far, it has been checked that without selection on the flavour and for light flavour jets, we obtained the expected behaviour of the JVF distribution; whereas for b-jets, we got a small sample of b-pu-jets, we didn't expected at all; if we want to take into account effects smaller than 0.2% we should investigate more about this result.

One of the main results we found is an estimation for the best compromise between efficiency and rejection: JVF cut  $\approx 0.79$  (avarage between the results obtained for light jets and all jets), even if we should have been able to find a more exhaustive result, including in the avarage the best extimation for JVF cut referred to  $p_t$ -efficiency and  $\eta$ -efficiency.

Finally, these results strongly depend on the method used for the matching: the same analysis could be done with different matching and compare them.

## References

- [1] The ATLAS Experiment at the CERN Large Hadron Collider, *Institute of physics publishing and SISSA, August 14, 2008*
- [2] Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data, *ATLAS NOTE, ATLAS-CONF-2012-020 March 2, 2012*
- [3] Jet-Vertex Association Algorithm, *ATLAS NOTE, January 29, 2008*
- [4] Pile-up jet energy scale corrections using the jet-vertex fraction method, *ATLAS NOTE, September 11, 2009*
- [5] <http://root.cern.ch/>
- [6] <http://pdg.lbl.gov/>