Draft version x.y

ATLAS NOTE

May 22, 2013

A template for ATLAS notes

First Author[a], Second Author[a], Third Author[b]

[a]*One Institution*
[b]*Another Institution*

Abstract

This is the abstract

To be submitted to *Phys. Lett. B*

1 Introduction

The discrimination between quark-initiated and gluon-initiated jets is a topic of large interest because it could improve many physics analyses and measurements. A natural field of application of a quark-gluon tagging tool is the QCD, where the estimations of several quantum numbers and couplings still need to be achieved. However, also other processes, which involve hadronic jet signatures, could benefit from a separation between quarks and gluons content to enhance the separation between signal and background processes.

In the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ decay channel, for example, the signature is characterized by two high-p_{T} leptons with opposite signs and two high-p_{T} jets. Since the two jet arise from the hadronic decay of the Z, they are certainly produced by the fragmentation and hadronization of two quarks. The main source of background to this channel originates from Z+jets processes, i.e. the production of a Z in association with other jets, where the additional jets mostly arise from QCD interactions and they can be inititiated from both quarks and gluons. In Z+jets events, tipically, gluon contribution is dominant with respect to quark contribution and this could be handled to separate background from signal.

Several theoretical and experimental studies exploited the relations between the flavour of the originating parton and jet properties. In particular, the analyses of 3-jet events at LEP showed that gluon jets are tipically broader than quark jets according to perturbative QCD calculations, and fragmentation and hadronization models. Furthermore, ATLAS studies on the jet energy scale showed that the calorimeter response is larger for jets originated from light-quark. Recently, J. Gallicchio and M. D. Schwartz published a phenomenological study of quark and gluon jet properties that exploits, in different energy regime, a large number of (potentially) discriminating variables and the potential of gluon jet rejection with respect to quark jet acceptance.

Neverthless, the performances of the application of a quark-gluon tagger to a specific analysis depends not only on the charachteristics of the tagger itself (multivariate method, efficiency and rejection, etc.) but also on the topology of the considered processes. The chance to successfully separate the signal from the background relies on a pre-existent difference in the jet flavour composition, i.e. the percentage of total events with jets of a given flavour (quarks and gluons); for example, there is no chance of discrimination if the quark jets contribution to the total events of signal is equal to the quarks contribution to the background.

This work presents the application of the ....

2 Quark-gluon generalities

The separation between quark and gluon jets, and so the possibility to construct a tagger of quarks and gluons, relies on the existence of potentially discriminating variables, which can be used to estimate quantitatively whether a jet looks as a quark- or as a gluon-initiated jet.

Several theoretical and experimental studies had demonstrated that a difference between quark and gluon jets exists and it arises from the differences in their properties (e.g. color charge, electrical charge, spin, etc.). For example, the ratio of the average multiplicity of all particles between gluon and quark jets as well as the ratio of the respective variances are described by the semi-classical approximation

$$\frac{<N_g>}{<N_q>} = \frac{C_A}{C_F} \quad \frac{\sigma_g^2}{\sigma_q^2} = \frac{C_A}{C_F} \tag{1}$$

where the ratio between the gluon and quark color charges corresponds to $C_A/C_F = 9/4$. Applying the Sterman-Weinberg definition, the relation between the angular widths of the quark and gluon jets, instead, can be described to the leading order as

$$\delta_g = \delta_q^{\frac{C_F}{C_A}} \tag{2}$$

These results imply that quark jets have a lower average number of particles than gluon jets whereas, considering the widths, the formers are tipically broader than the latters, which can be intuitively explained considering that quark jets are dominated by the first gluon emission. Many LEP studies have investigated these properties confirming the differences and so the possibility to find the discriminating variables needed to construct a quark-gluon tagger. Furthermore, ALEPH and OPAL experiments have reported a deviation of jets originated from b-quarks, and so called b-jets, from the light-quarks behaviour described above. The average number of particles and the width of b-jets are tipically larger than light-quark jets showing distributions more similar to gluon jets.

In the recent article of Schwarz and Gallicchio, which also reviews the actual status of quark-gluon discrimination studies, a variety of jet observables (and combinations of those) has been exploited, using a multivariate approach and estimating the respective separation powers. The authors show that just few variables, tipically two, describe almost all differences between quark and gluon jets. For high-p_{T} jets the most powerfull variable is the number of tracks, i.e. the number of charged particles, inside a cone of given radius around the jet axis whereas for low-p_{T} jets geometric moments, which measure the spread of the jet, give greater separations.

2.1 Discriminating variables: N_{trk} and $Width$ definitions

Since the best separation is obtained combining a discrete jet variable togheter with a continous one, respectively particles multiplicity and geometric moments of jets, in this study two observables belonging to these category are chosen: the tracks multiplicity and the first geometric moment, the width, defined respectively as

$$N_{\mathrm{trk}} = \sum_{i:\Delta R_i<0.4} |n_i| \text{ and } Width = \frac{\sum_{i:\Delta R_i<0.4} p_{\mathrm{T}}^i \Delta R_i}{\sum_{i:\Delta R_i<0.4} p_{\mathrm{T}}^i} \tag{3}$$

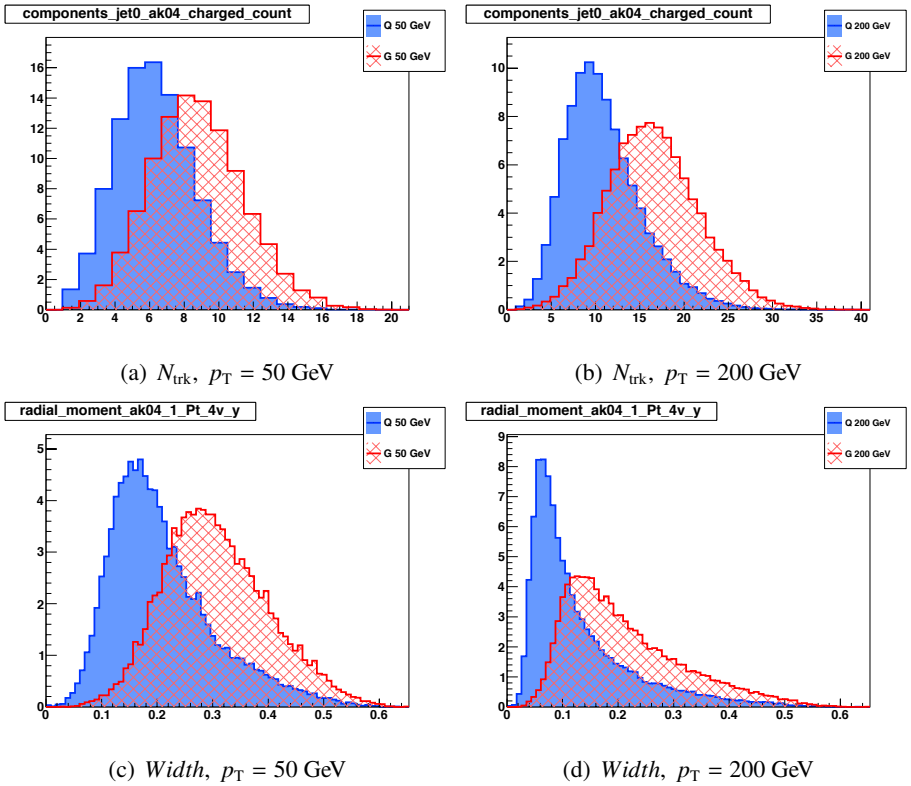
where all sums run over all charged particles within a cone of radius $\Delta R = 0.4$ around the jet axis, n_i is just a counter of charged tracks, p_{T}^i represents the transverse momentum of the i-th track and ΔR_i the relative distance from jet axis. Since jet properties are strongly p_{T}-dependent, the analysis is splitted in different bins of p_{T} trying to disentangle and reduce momentum-dependent effects from variables distributions.

Figure 1 shows the comparison of N_{trk}, Fig. 1(a) - 1(b), and $Width$ distributions, Fig. 1(c) - 1(d), between quark and gluon jets for two different p_{T} of the jets, namely 50 GeV and 200 GeV. It is possible to observe that the separation between quark and gluon jets distributions of charged particle multiplicity increases with the p_{T} of the jets while width distributions separation gets worst. According to the behavior described in Sec. 1, gluon jets tipically fragment in a number of charged particle greater but softer (with a lower average p_{T}) than quark jets, leading to a lower and broader gluon width distribution that is observable in the bottom plots of Fig. 1.

2.2 Multivariate methods and Self-Organizing Map

A quark-gluon tagging is a method based on a set of discriminating variables that can be used to construct a classificator able to quantitatively establish if a jet looks more as a quark or a gluon jet, by means of multivariate methods.

The choice of the method, and so of the classificator, characterizes the tagger response and the respective performances, therefore the test of several approaches is crucial once a comparison parameter has been chosen. For each selected method, tagger abilities can be described by the quark efficiency - gluon rejection curves, the so called ROC curves, produced applying a sliding cut on the discriminant that defines the tagger working point.



(a) N_{trk}, $p_{\mathrm{T}} = 50$ GeV (b) N_{trk}, $p_{\mathrm{T}} = 200$ GeV

(c) *Width*, $p_{\mathrm{T}} = 50$ GeV (d) *Width*, $p_{\mathrm{T}} = 200$ GeV

Figure 1: Distributions of N_{trk} (top) and *Width* (bottom) for two different p_{T} of the jet. Both quark (blue) and gluon (red) jet distributions are shown.

Tagger performances, however, will depend on the definition and the purity of the sample to which the tagger is applied. Indeed, defining in a jet-by-jet analysis the quark efficiency (ϵ_q) and the gluon rejection ($r_g = 1 - \epsilon_g$) as the fraction, respectively, of quark jets selected and gluon jets rejected for a given working point, the efficiency of the cut on a sample with a certain initial composition of quarks, f_q, and gluons, f_g, will be $\epsilon_{\mathrm{cut}} = \epsilon_q f_q + (1 - r_g) f_g$ with a new quarks fraction equal to $f'_q = \epsilon_q f_q / \epsilon_{\mathrm{cut}}$. If the searched signal in the analysis is composed of just quark jets, it is straightforward to choose the tagger working point that maximizes the quark purity f_q of the sample since this implies also the maximization of the signal with respect to the background number of events. Otherwise, if the signal itself is composed of a mixture of quarks and gluons, the maximization of the ratio ϵ_q/ϵ_g is no longer usefull but the signal significance $S/\sqrt{B}$ has to be used in turn, maximizing the relative improvement defined as the ratio between the signal significance after and before tagger application, which correspond to the ratio between the signal and background cutting efficiency, i.e. $\epsilon_S^{\mathrm{cut}} / \sqrt{\epsilon_B^{\mathrm{cut}}}$.

3 Multivariate method and Self-Organizing Map

Machine learning and data mining had, in the last decades, a large diffusion thanks to many theoretical contributions and a wide range of practical applications in many fields, e.g. cognitive neuroscience, medical imaging as well as high energy physics. Classification and clustering, as in general pattern recognition, are well known and well studied problems and several algorithms, based on multidimensional models of complex systems, was proposed to extract unknown properties or to reproduce knowledge in high dimensional data. Artificial Neural Networks (ANNs) are an example of nonlinear multivariate models used to describe complex relationships and/or to find patterns in input data and they are widely

employed in high energy physics, in particular for offline data analysis.

A Self-Organizing Map (SOM) is a particular type of ANN based on an unsupervised learning model proposed by Kohonen to reduce multidimensional data distributions onto a typically bidimensional representation of training input, called map, trying to preserve the topological properties of input data space in the $\mathbb{R}^n \to \mathbb{R}^2$ mapping. A reference weight vector $w_i \in \mathbb{R}^n$ is therefore associated to each node of the bidimendional map and given an input vector $x \in \mathbb{R}^n$ is possible to define as response of the SOM the unit with the nearest reference vector, called Best-Matching Unit (BMU), in the chosen metrics.

4 Dijet events study

5 The standard *llqq* analysis

6 SOM discriminant construction

7 Results

8 Conclusions