# ELPHI 6024 / 1

b-tagging algorithms

Sebastian Fleischmann, Univ. Wuppertal

- Repetition: b-quark properties
- Why b-tagging?
- Signatures of b-jets & tagging variables
- Combining tagging variables
- Efficiency and rejection ratio
- Summary





Un	+2/3	Charm + 2/3	Top/ +2/3	Particle	Symbol	Quark Content	Spin	Mass $(GeV/c^2)$
op			Truth	charged pion	$\pi^+$	nd	0	0.140
11	3	<u> </u>	I <b>∔</b> 3	neutral pion	$\pi^0$	$\mathbf{n}\overline{\mathbf{n}},\mathbf{d}\overline{\mathbf{d}}$	0	0.135
u				$J/\psi$	$\psi$	$\overline{0}$	1	3.097
	~ 4	~1300	>131,000	Upsilon	Υ	bb	1	9.460
D	1/2	Stuance 1/2	Bottom/ 1/2	D plus	D <sup>+</sup>	$c\overline{d}$	0	1.869
Down	-1/3	Strange -113	Beauty	D zero	D <sup>0</sup>	cīī	0	1.865
1	з	<b>~</b> <sup>3</sup>	<b>h</b> 3	D sub s	$D_{s}^{+}$	cs	0	1.969
<b>d</b>		$\mathbf{s}$ —	m - m	B plus	B+	ub	0	5.279
M.	~7	~150	~4,800	B zero	B <sup>0</sup>	$d\overline{b}$	0	5.279
				B sub s	$B_s^0$	$s\overline{b}$	0	5.375
				Particle	Symbo	ol   Quark Content	Spin	Mass (GeV/c
				Proton	р	und	1/2	0.938
				Neutron	n	udd	1/2	0.940
				Lambda c	$\Lambda_{c}^{+}$	ude	1/2	2.285
				Lambda b	$\Lambda_b^0$	udb	1/2	5.641

#### • b-hadron properties

 lifetimes ~ 1,5 ps, corresponds to flight distances of ~4 mm for 50 GeV particles (but keep in mind: exponentially distributed!)



•  $\sim$  20% probability to decay into muon or electron

# **Repetition:**

#### Cabibbo-Kobayashi-Maskawa-Matrix

 CKM-Matrix describes mixing of electro-weak quark eigenstates in the QCD flavour (mass) eigenstates

$$\begin{pmatrix} d'\\ s'\\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub}\\ V_{cd} & V_{cs} & V_{cb}\\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d\\ s\\ b \end{pmatrix}$$

with its values (S. Eidelman et al., Phys. Lett. B 592, 1 (2004))

- $\begin{pmatrix} 0.9739 \text{ to } 0.9751 & 0.221 & \text{to } 0.227 & 0.0029 \text{ to } 0.0045 \\ 0.221 & \text{to } 0.227 & 0.9730 \text{ to } 0.9744 & 0.039 & \text{to } 0.044 \\ 0.0048 & \text{to } 0.014 & 0.037 & \text{to } 0.043 & 0.9990 & \text{to } 0.9992 \end{pmatrix}$
- therefore is the decay of b-quarks into lighter quarks suppressed by the CKM values (decay into top obviously not possible because of higher mass)



# Why b-tagging?

 Many important channels in collider experiments at LEP and Tevatron / LHC energies contain b-quarks

- Higgs-decay:  $H \rightarrow bb$  predominant for  $M_{H} < 130 \text{ GeV}$
- top-decay:  $t \rightarrow bW$
- SUSY particles
- B-Hadrons important for CP-Violation measurements
- b-jets are signatures for these processes; the tagging of a b-quark gives direct hints about the underlying (primary) process before hadronization occurred
- other jet-types (lighter quarks / cquarks /gluons) are much harder or not to distinguish in the detector





WW

# Signatures of b-jets

- Most tagging variables are based on the longer lifetime of B-Mesons and the related separation of primary and secondary vertex
  - Impact parameter (IP) and variables derived from the IP
  - measurement of a secondary vertex and the distance of primary and secondary vertex
- other variables tag on the kinematic properties of the Bmeson decay
  - invariant mass at the secondary vertex
  - transverse momentum
  - track rapidities
- or on other signatures related to the production of the bquark
  - soft lepton tag





#### Tagging variables: Impact Parameter (IP)

- Impact parameter can be defined for each track individually by the distance between reconstructed track and primary vertex
  - this distance is often split into two independent components because of unequal detector resolutions in those directions:
    - Rz component along z-direction (beam direction)
    - RΦ component transverse to z (higher resolution)





#### Tagging variables: Impact Parameter (IP) sign definition

- Sign of the impact parameter can be defined
  - using only the track directions ("geometrical sign")
  - including the estimated flight path of the B-meson by calculating the point of closest approach between track and B flight path and separating between upstream and downstream points ("lifetime sign")





## <u>Tagging variables: Impact Parameter (IP)</u> general properties

- IPs are correlated with each other for tracks which are included into primary vertex fit used for the IP calculation
- can be calculated for each track even without reconstructed secondary vertex
- impact parameter depends 10 on the resolution of the detector and precision of each measurement: create a new dimensionless variable called 10 "track significance" by

$$S = \frac{IP}{\sigma_{IP}}$$





# Tagging variables: Impact Parameter (IP)

track significance



- tracks with negative significance mainly come from the primary vertex; large negative (and positive) significance can result from scatters, wrong hit association etc.
  - negative significance distribution (*resolution function*) is expected to be Gaussian with tails

excess of tracks with positive significance due to long-lived particles

Tagging variables: Impact Parameter (IP) track probability

By integrating the known track significance distribution f(S) for tracks from the primary vertex (i.e. using the negative distribution resp. resolution function) one can get the PDF of the so-called track probability

$$P(S) = \frac{1}{\int_0^\infty f(\tilde{S}) d\tilde{S}} \int_S^\infty f(\tilde{S}) d\tilde{S}$$

- using this new PDF one can calculate the probability of an individual track to origin from the primary vertex
- the track probability distribution is expected to be flat for PV tracks and shows an excess at zero for SV tracks





#### Tagging variables: Impact Parameter (IP) track probability

By integrating the known track significance distribution f(S) for tracks from the primary vertex (i.e. using the negative distribution resp. resolution function) one can get the PDF of the so-called track probability





#### Tagging variables: Impact Parameter (IP) lifetime probability

• the track probabilities  $P(S_i)$  for a given group of tracks can be combined into the so-called *lifetime probability* 

$$P_{N} = \Pi \sum_{i=0}^{N-1} \frac{(-\log(\Pi))^{i}}{i!}$$
 where  $\Pi = \prod_{j=1}^{N} P(S_{j})$ 

- possibility to calculate this variable for any group of tracks (e.g. jet, hemisphere) makes it very flexible and allows for adoptions to specific physics processes
- the lifetime probability distribution is flat between 0 and 1 for uncorrelated tracks from the PV and peaks at 0 for track groups from secondary vertices



#### Tagging variables: Secondary vertex

- In some b-jets secondary vertices can be reconstructed
  - different methods for vertex fitting are available
    - Billoir method (global fit)
    - Kalman fitter based method
    - adaptive methods which improve the selection of tracks for a given vertex candidate
  - selection of tracks is crucial for this task  $_{10^4}$
- due to cascade decays ( $B \rightarrow D$ ) more than one secondary vertex may be reconstructed in a jet
  - cascade decays can also be used to improve the rejection of c-quarks





Sebastian Fleischmann





#### Tagging variables: Fraction of total energy at secondary vertex

- the fraction of charged energy reconstructed at the secondary vertex with respect to the whole jet energy (reconstructed calorimeter energy) allows to discriminate between b-jets and c-jets
- fragmentation properties yield to differences in the distributions





#### Tagging variables: Transverse momentum at secondary vertex

- the transverse momentum with respect to the estimated b-hadron flight direction takes into account missing particles not included in the secondary vertex
- as the sum of all track momenta is calculated one gets one variable per secondary vertex
- missing particles can be neutrinos from semileptonic decays, neutral particles or non-reconstructed charged particles





#### **Tagging variables:** Track rapidity at secondary vertex

- rapidities of tracks included in the secondary vertex differ for b- and c-jets due to the higher mass of B-hadrons
- the variable can be defined for each track in the vertex fit which improves the whole btagging significantly due to the higher statistic





## Tagging variables: Soft leptons

 Soft lepton tag is based on semileptonic B decays, therefore limited use due to branching ratio (BR~0.10)



# **Combining tagging variables**

- the different tagging variables need to be combined into a single variable which allows to decide whether or not a jet is tagged as a b-jet
- several methods exits for this purpose:
  - likelihood ratio
  - neural networks
  - "specialized methods" for correlated variables



# Combining tagging variables: Likelihood ratio,

• define the combined tagging variable *y* by:

$$y = \frac{f^{bgd}(x_1, \dots, x_n)}{f^{sig}(x_1, \dots, x_n)}$$

where  $f^{bgd}(x_1,...,x_n)$ ,  $f^{sig}(x_1,...,x_n)$  are the probability density functions of the discriminating variables  $x_1,...,x_n$ 

- problem: *n*-dimensional PDFs difficult to determine
- for independent variables this expression gets

$$y = \frac{f_1^{bgd}(x_1)}{f_1^{sig}(x_1)} \cdots \frac{f_n^{bgd}(x_n)}{f_n^{sig}(x_n)} \equiv \prod_{i=1}^n y_i$$

i.e. for each variable  $x_i$  an independent  $y_i$  can be computed (the so-called *weight*)



# Combining tagging variables: Likelihood ratio.

• define the combined tagging variable y by:

where  $f^{bgd}(x_1,\ldots,x_n)$ ,  $f^{sig}(x_1,\ldots,x_n)$  are t functions of the discriminating variables

- problem: *n*-dimensional PDFs difficult to
- for independent variables this expression

$$y = \frac{f_1^{bgd}(x_1)}{f_1^{sig}(x_1)} \cdots \frac{f_n^{bgd}(x_n)}{f_n^{sig}(x_n)} \equiv \prod$$

i.e. for each variable  $x_i$  an independent (the so-called *weight*)





#### Combining tagging variables: Jet classes in the likelihood ratio method

- jets with reconstructed secondary vertex allow for more tagging variables to be used, therefore different classes of jets with different sets of variables have to be formed, e.g.
  - jets with reconstructed secondary vertices
  - jets with x tracks with low track probability respectively high positive significance
  - other jets
- the tagging variable  $y_{\alpha}$  for the jet class  $\alpha$  is then normalized by the according production rates  $n_{\alpha}^{b}$ ,  $n_{\alpha}^{c}$ ,  $n_{\alpha}^{q}$  for b-, c- and light quark jets of the class  $\alpha$

$$y_{\alpha} = \frac{n_{\alpha}^{c}}{n_{\alpha}^{b}} \prod_{i=1}^{n} y_{i,\alpha}^{c} + \frac{n_{\alpha}^{q}}{n_{\alpha}^{b}} \prod_{i=1}^{n} y_{i,\alpha}^{q}$$



# Combining tagging variables: Properties of the likelihood ratio method

• Pro

- easy to implement
- number of used tagging variables can be easily varied

Contra

variables have to be (at least approximately) uncorrelated



# Combining tagging variables:

Neural Networks

- Neural networks are widely used in multivariate analysis
- connection strength between nodes is set dynamically in a training phase



'Probability' of being b-jet

Hidden layer

 $f = \frac{1}{1 + e^{-x}}$ 

 $\mathbf{h}_{j} = \mathbf{f}(\Sigma \mathbf{x}_{i} \mathbf{w}_{ij} + \theta_{j})$ 

Input layer

Output layer O =  $\Sigma h_j w_j + \theta_o$ 

$$\mathsf{E} = \frac{1}{\mathsf{N}} \sum (\mathsf{O} - \mathsf{T})^2$$



# Combining tagging variables:

**Properties of Neural Networks** 

- pro
  - correlations between variables are automatically taken into account
  - no PDFs need to be known; they are intrinsic "knowledge" of the net
- contra
  - "obscure" how the net came to its decision
    - but enhancements of neural nets take a priori knowledge into account
  - set of variables has to be the same for training and use phase



# **b-tagging efficiency and rejection rate**

 to quantify the performance of b-tagging algorithms it is common to define the b-tagging *efficiency*



top-Seminar - 2006-01-27: b-tagging algorithms

Sebastian Fleischmann

#### Summary

- b-tagging is very important for collider experiments
- though it has high demands on
  - the tracking performance of the detector
  - Monte-Carlo simulations
- many different tagging variables are used, but the Impact Parameter (IP) of tracks is the most important one
- variables have to be combined to a single discriminating value for each jet or the whole event by the use of likelihood ratios, probability functions or neural networks



#### Backup







