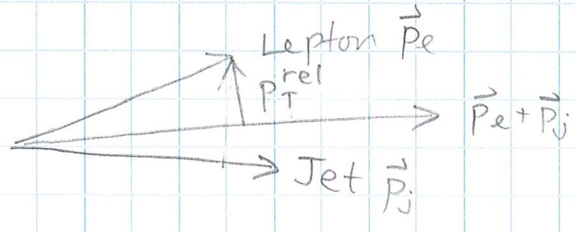
B-Jet Tagging

B-hadrons have relatively large semileptonic branching fractions. (10% each to μ and e)

Large mass (compared to m_e or m_μ) means that the lepton $|\vec{p}_T^{\,rel}|$ is large.

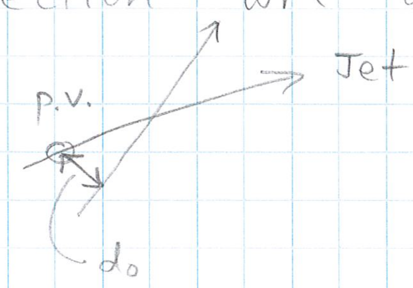


The shape of the p_T^{rel} distribution is different for b-jets and udsg+c jets.

Finding a lepton enhances the b-fraction. Fitting the shape of the p_T^{rel} distribution to templates gives the b-purity.

Track impact parameter tagging.

In two-dimensions the impact parameter can be given a sign based on the track direction wrt a jet axis:



The track parameter d_0 carries a sign such that

$$x_0 = -d_0 \sin\phi_0$$
$$y_0 = d_0 \cos\phi_0$$

Then, if $\hat{u}$ is the jet axis, $\hat{u} = \dfrac{\vec{p}_{xy}}{|\vec{p}_{xy}|}$

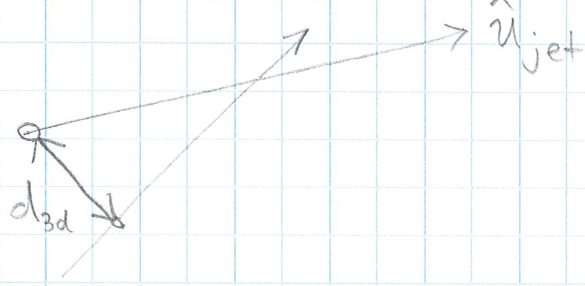
then $d_{xy} = x_0 u_x + y_0 u_y$

$$= d_0 \left(u_y \cos\phi_0 - u_x \sin\phi_0 \right) = d_0\, \hat{z} \cdot (\hat{t} \times \hat{u})$$

d_{xy} will generally be positive if the track was from a displaced secondary vertex and equally probable to be positive or negative if it came from the origin.

Generalization to 3-d:

The point of closest approach in 3-d is unique.

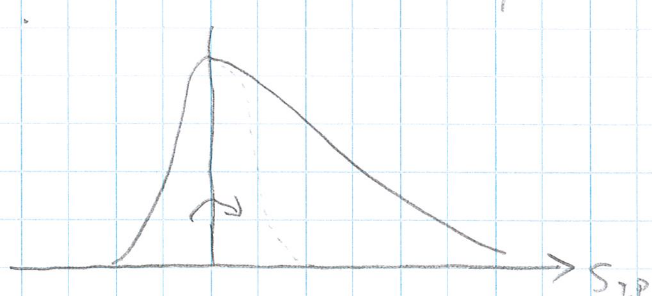


$$d_{3D} = x_0 u_x + y_0 u_y + z_0 u_z$$

In practice, consider $\frac{d_{3D}}{\sigma_{d_{3D}}}$ which should be distributed as a gaussian with mean 0 and width 1 for prompt tracks.

$d_{3D}/\sigma_{d_{3D}} = S_{IP}$ (Impact parameter significance)

Even if S_{IP} is not a unit gaussian for prompt tracks, it is still expected to be symmetric. The negative half should describe the shape of the prompt component.

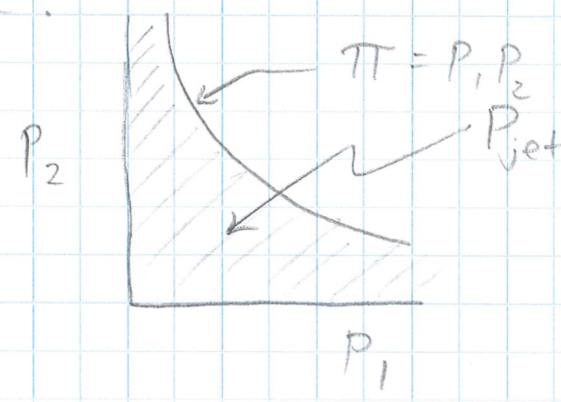


Consider a jet with two tracks. Calculate

$$P(s) = \int_{-\infty}^{-|s|} R(t)\,dt$$

where R is the resolution function for prompt tracks.

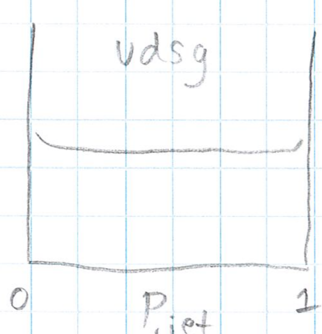
If s is large and positive then $P(s)$ is small.



P_{jet} is the set of two track combinations with probability less than or equal to Π.

In general
$$P_{jet} = \Pi \sum_{k=0}^{N-1} \frac{(-\log \Pi)^k}{k!}$$

$$\Pi = P_1 P_2 \cdots P_N .$$



Secondary Vertex Algorithms.

Reconstruct a secondary vertex within a jet. Discriminating variable is the decay length significance.

Require that

1. Most tracks are not also in the primary vertex
2. Reject vertices that are too far from the primary (K_s^0) or are too massive (more than 6.5 GeV).
3. Flight direction is within a cone around the jet axis.

Multivariate Methods.

Inputs:

- vertex category
- L_{xy}
- Mass
- # of tracks in the vertex
- E_{vertex} / E_{jet}
- η of tracks in the jet wrt jet axis
- # of tracks in the jet
- S_{IP} of tracks in the jet

Combined to form a likelihood discriminant.

Form C
Approved for use in
Purdue University

Calibrating b-jet tagging algorithms.

Usually we need to know the efficiency as precisely as possible.

In Monte Carlo we can cheat because we know whether a jet was a b-jet or not.

$$\varepsilon_b^{mc} = \frac{n_{tag}}{n_b}$$

This does not measure purity, which depends on the sample. But the algorithms can be applied to light flavor jets to estimate the mistag rate under various assumptions.

Typical operating point: 10% mis-id probability
80% b-tag efficiency

Tight selection 0.1% misid probability
50% b-tag efficiency.

Calibration in data.

To measure the efficiency in data we need a pure source of b-jets.

Example: $t\bar{t}$ decay

$$t \to W^+ b$$
$$\hookrightarrow \mu^+ \nu_\mu$$

$$\bar{t} \to W^- \bar{b}$$
$$\hookrightarrow e^- \bar{\nu}_e$$

Signature is two high p_T isolated leptons (ie, not in a jet) missing E_T (from the neutrinos) and two jets which are pure $b\bar{b}$.

But there are some backgrounds:

tW (with an additional jet)
WW (with additional jets)
WZ
ZZ

$$\varepsilon_b = \sqrt{\frac{F_{2\text{-tag}} - F^{MC}_{non\,2b}}{f_{2b}}}$$

$F_{2\text{-tag}}$ fraction of the 2-jet sample with double b-tag

$F^{MC}_{non\,2b}$ expected fraction from non $b\bar{b}$ estimated from MC

f_{2b} fraction of sample with 2 b-jets.

Efficiencies cancel.

But, $2.5\ fb^{-1}$ at $\sqrt{s} = 13$ TeV $\Rightarrow$ 2000 double tags.

Efficiency from di-jet samples.

In QCD, $g \to b\bar{b}$ so if one jet is tagged as a b, the other jet has a significantly enhanced b-jet purity.

Apply a tight tag to one jet, apply the test tag to the other jet and measure the b-purity by fitting p_T^{rel} or impact parameter of leptons in the jet.

Restriction: lepton p_T^{rel} and IP is used to measure the fraction before and after the b-tag. The b-tag must not change the shape of these distributions.

System8 method: Requires a muon jet and at least one other away-jet.

Require $p_T^{rel} > 0.8$ GeV.

Efficiency ε_b^{ptRel}

Apply a lifetime tagger with efficiency ε_b^{tag}.

Full muon-jet sample $n = n_b + n_{cl}$
away-jet sample $p = n_b + p_{cl}$

$n^{tag} = \varepsilon_b^{tag} n_b + \varepsilon_{cl}^{tag} n_{cl}$
$p^{tag} = \beta^{tag} \varepsilon_b^{tag} p_b + \alpha^{tag} \varepsilon_{cl}^{tag} p_{cl}$

$n^{ptRel} =$
$p^{ptRel} =$

$n^{tag,ptRel} =$
$p^{tag,ptrel} =$

There are then 8 equations and 8 unknowns :

$$n_b \,, \quad n_{cl}$$
$$p_b \,, \quad p_{cl}$$
$$\varepsilon_b^{tag} \,, \quad \varepsilon_{cl}^{tag}$$
$$\varepsilon_b^{ptRel} \,, \quad \varepsilon_{cl}^{ptRel}$$

where α, β are scale factors estimated from MC.

Systematic Uncertainties.

- pileup
- gluon splitting (more $g \to b\bar{b}$ in data?)
- muon p_T^{rel} template shapes

- different away jet tagger
- differences between muon jets and inclusive jets

Currently these systematic effects are at about the 1-2 % level.