Impact of Tracker Misalignment on the CMS b-Tagging Performance

Christophe Saout

CERN, Karlsruhe Institute of Technology (KIT)

for the CMS experiment
on behalf of the b-Tag and Vertexing
Physics Objects Group

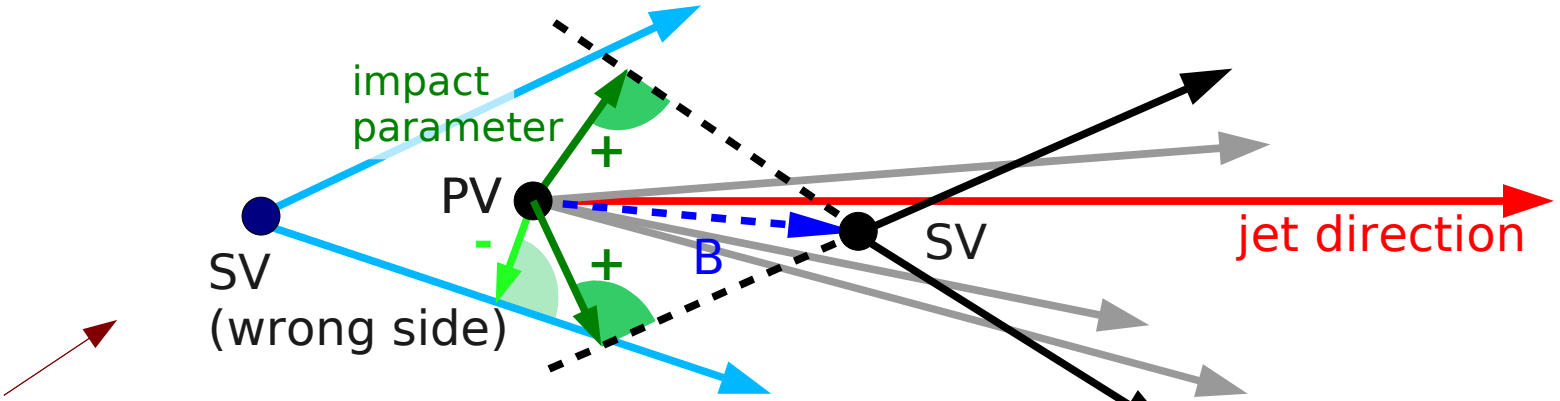
- Introduction
- CMS tracking system
- Misalignment Scenarios
- Effect on Observables
- Effect on Algorithm Performance
- Conclusions

Introduction

- b-tagging is of crucial importance for physics analyses that need to identify jets from **heavy quark** flavours *(b, top)*,
  e.g. **top, SUSY, Higgs**

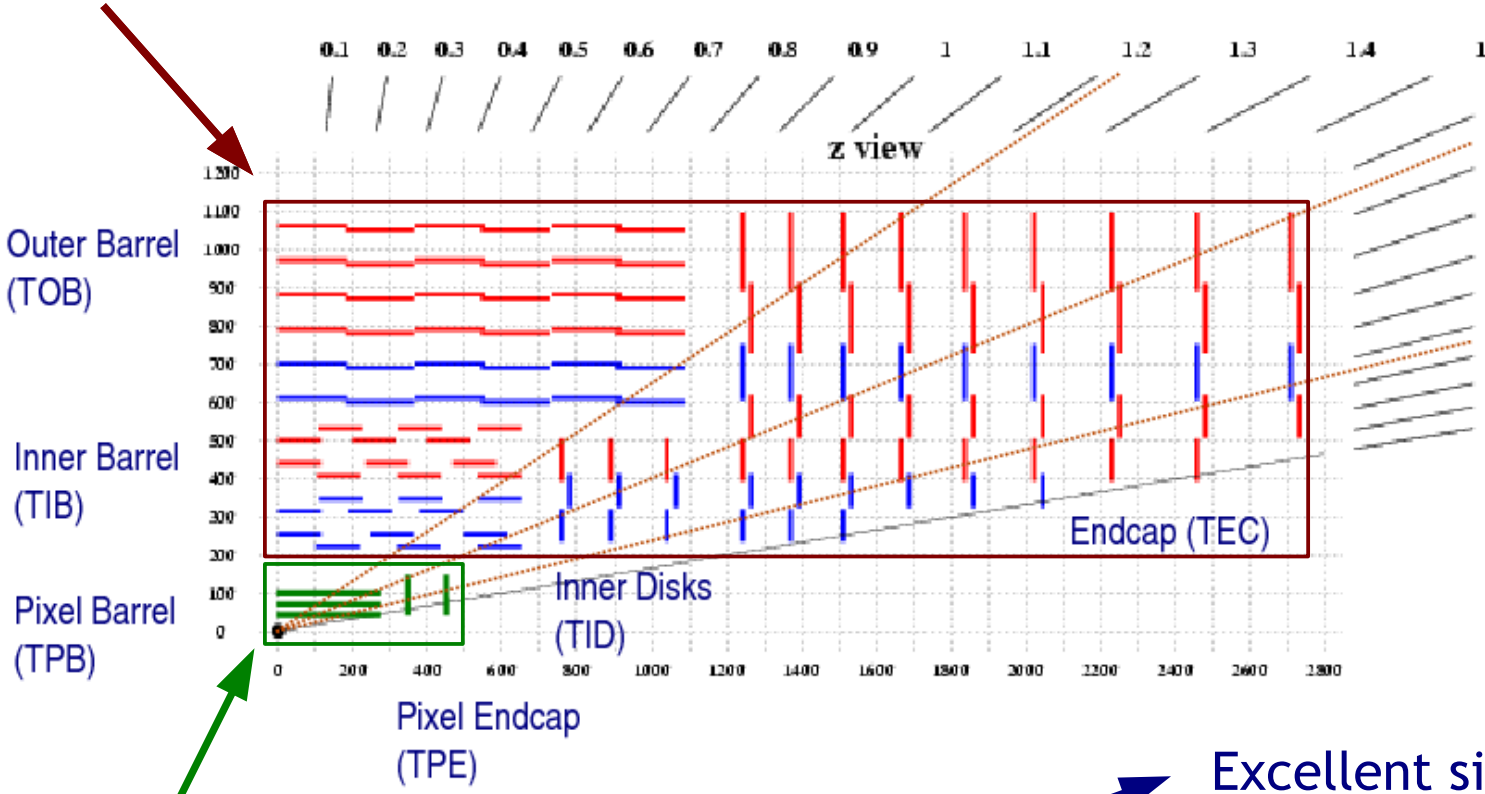
- **b**-quarks significantly differ from light flavour quarks by:
  - mass: m = 4.2 GeV
  - lifetime: τ ≈ 1.5 ps → ~1.8mm *(at 20 GeV)* before decay
  - decay: weak, mostly into c-quarks *(→ 3rd decay)* → 20% into leptons
  - tracks: high decay multiplicity, significant displacement
  - Secondary vertices *(SV)*: tracks intersecting at a common vertex

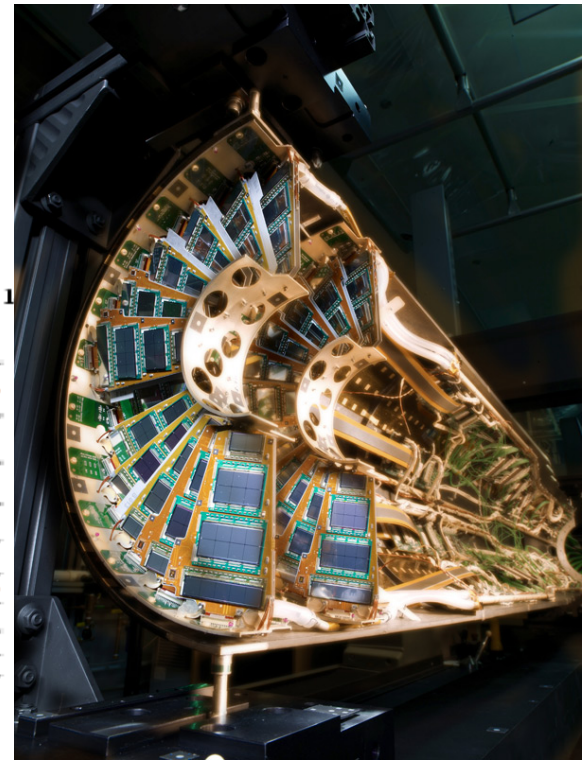
  → need good tracking resolution at impact point! *(→ alignment!)*



The CMS Tracking System

10(*) layers of silicon strip detectors
- r-φ strip pitch of 80μm-180μm
- stereo layers: angle of 5.7°





Three(*) layers of pixel detectors:
- 768 modules
- Inner ring at r = 4.4cm
- 100 μm × 150 μm pixel size

Excellent single-point resolution:
- 10μm in r-φ, 20μm in z

→ *good for b-tagging*

(*) in the central detector

Misalignment Scenarios

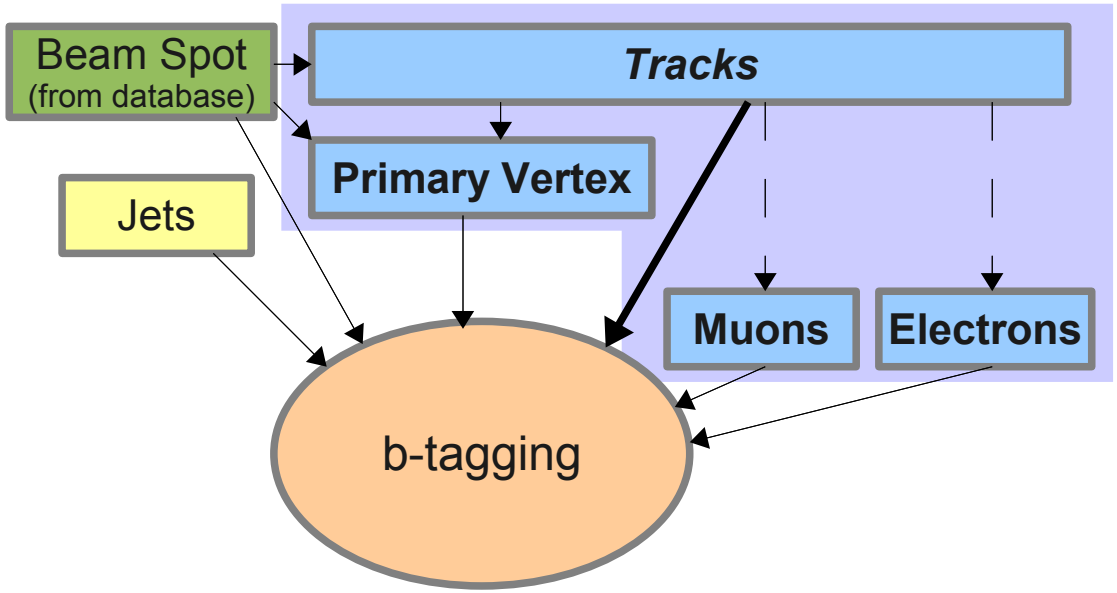
- **Startup**: *survey, laser alignment, cosmics*
- **10 pb^{-1}**: cosmics, min. bias, J/Psi, Ypsilon
- **100 pb^{-1}**: high p_T μ, W/Z
- **10 pb^{-1} and pixel barrel layer 1 disabled** *(academic interest / early beam safety)*
- **Ideal** *(no misalignment)*

"realistic misalignment scenarios" for the tracker as defined by the alignment group in ***2007***
(before CRAFT data-taking)

- b-tagging uses the impact parameter, which is dominated by the *pixel detector*.
- The "startup" scenario is being continuously updated as alignment with cosmics proceeds. It currently is somewhere between the old 10 pb^{-1} and 100 pb^{-1}. *(around 10 pb^{-1} for the central pixel detector)*
- The point of this study is not to give b-tagging performance numbers to be used for early analysis, but to study the sensitivity of the algorithms. ***The actual efficiencies will be measured on data!***

Algorithms Overview

"Ingredients"
from other
"Physics Objects Groups"

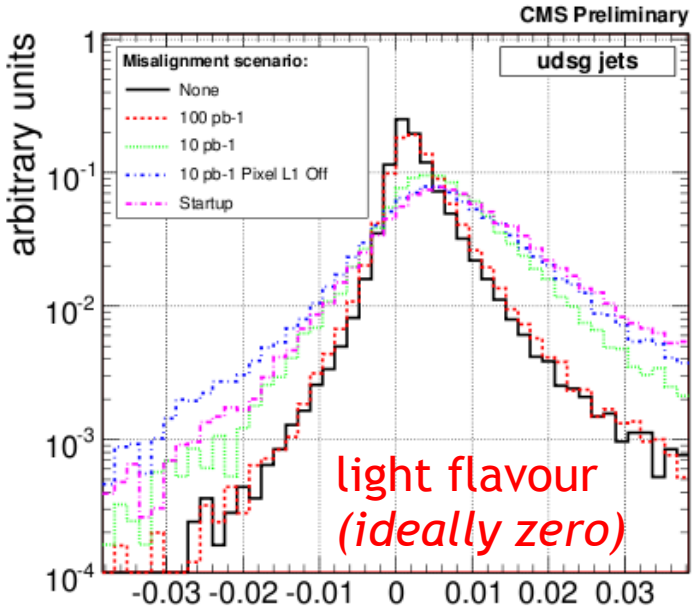


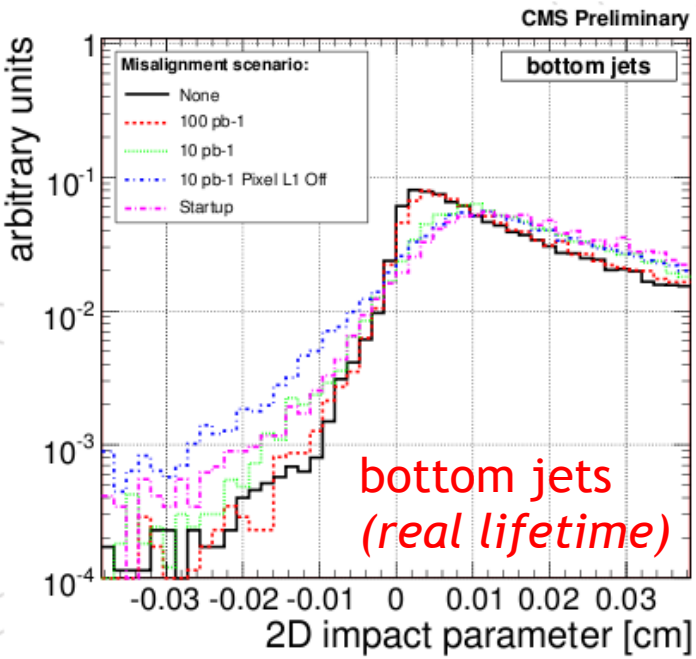
- The CMS offline software contains a variety of algorithms
  - The "simple" ones *(exploiting one characteristic at a time)*:
    - Track Counting — 1 IP
    - Simple Secondary Vertex *(new)* — 1 SV
    - Soft Lepton — 1 μ/e

    } *Need the least knowledge about the detector → suitably for early data*

  - Combined algorithms *(using MVA techniques → training/PDFs)*
    - Jet Probability — all IP
    - Combined Secondary Vertex — all SV + all IP + more

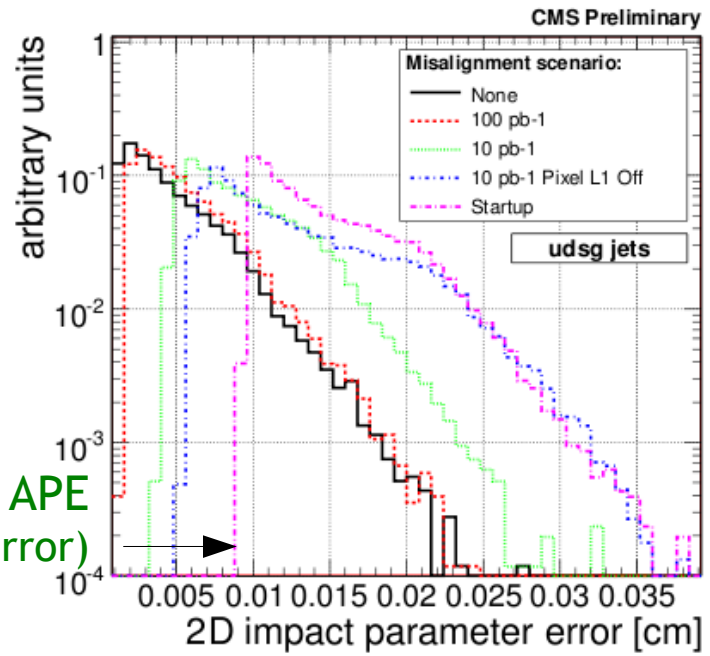
Impact Parameters





Distributions for the 2nd-highest signed impact parameter (IP) in jets:
(→ "track counting" b-tag algorithm)

error on the IP measurement
dominated by pixel hit resolution, extrapolation from innermost hits



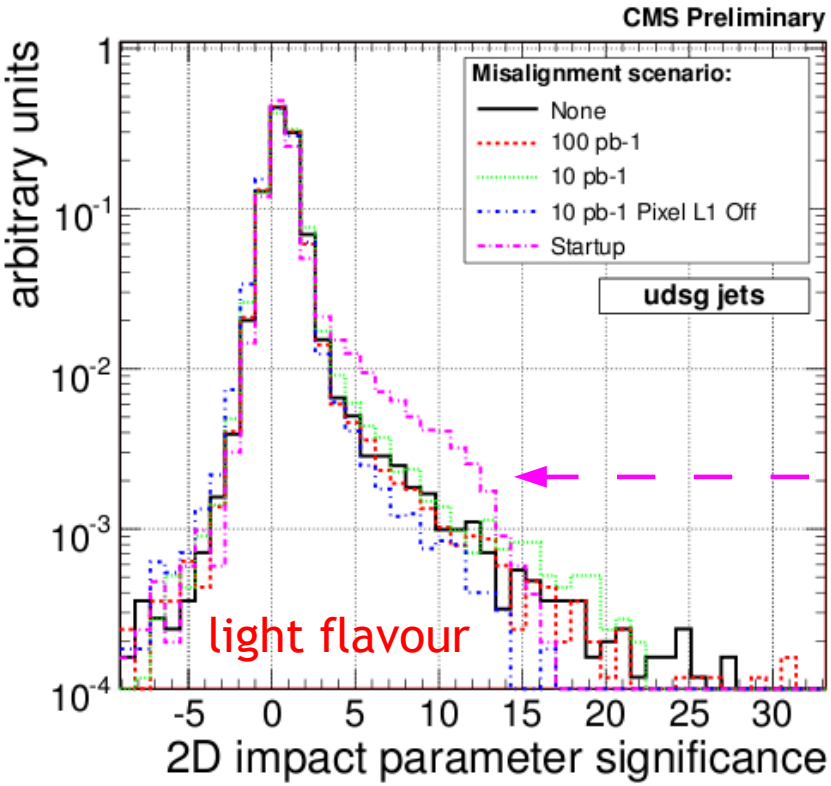
Effect of the APE
(alignment position error)

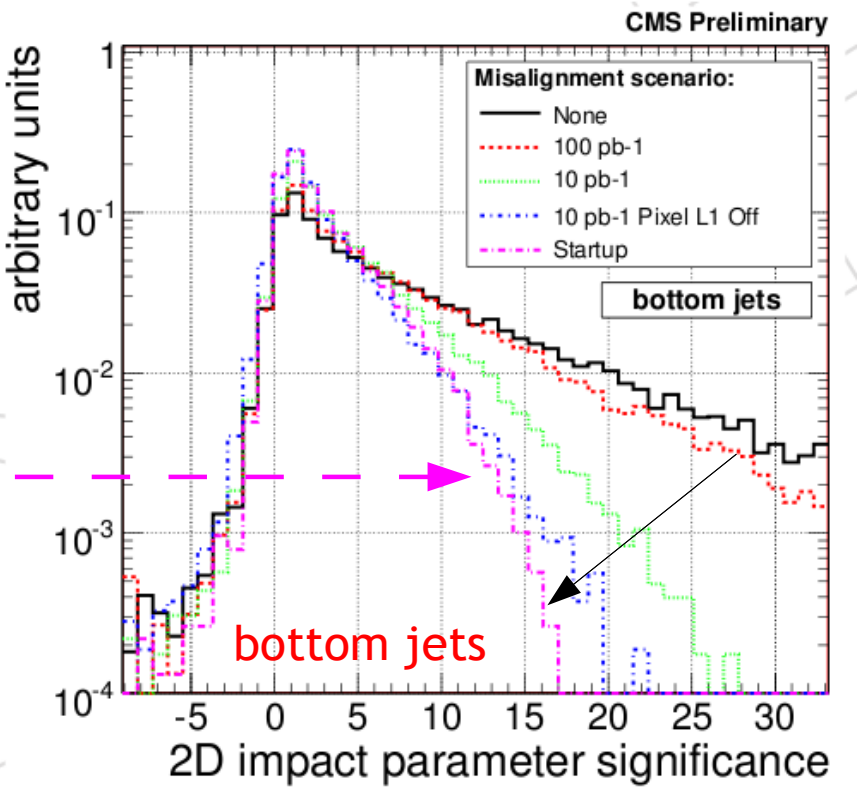
Impact Parameter Significances

b-Tagging algorithms are more robust if invariant against per-track errors.
Therefore, the significance of a track being displaced wrt. the PV is taken:

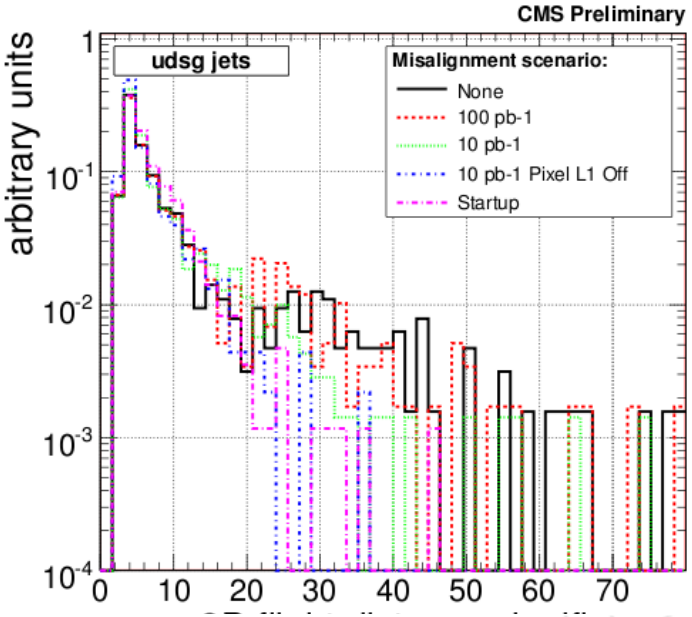
$$IP_{sig} = IP / \sigma_{IP}$$

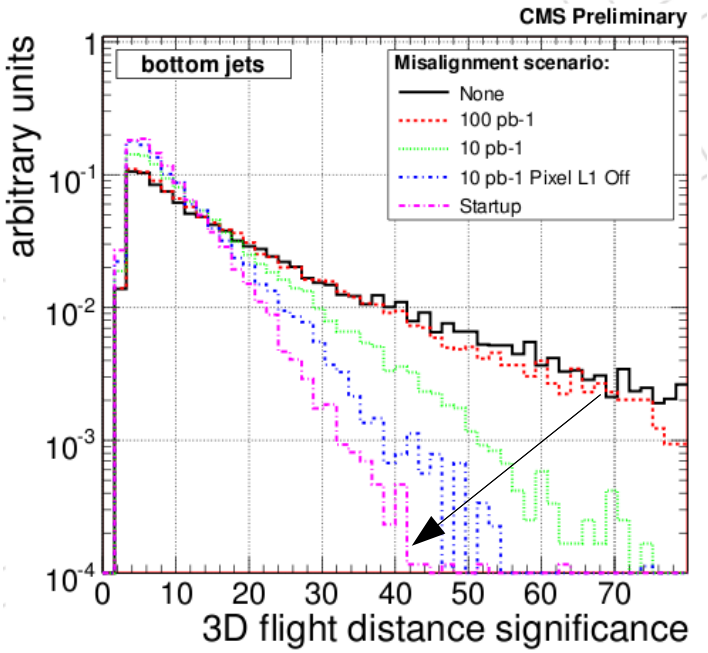
the 2nd-highest signed IP significance:





Secondary Vertices





Given a reconstructed displaced secondary vertex (SV), one can compute the "flight distance significance" as

$$D^{3D} = |\overrightarrow{SV} - \overrightarrow{PV}|$$

with the significance computed analoguosly *(using errors from both vertex fits)*

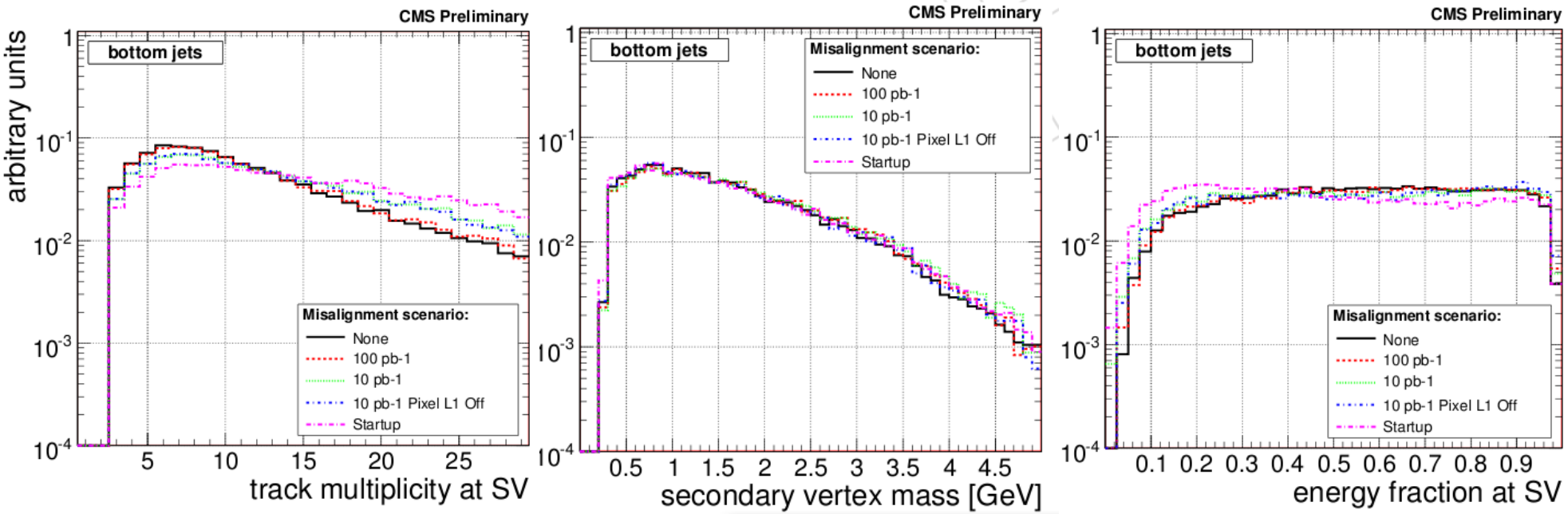
→ *discr. For the "Simply Secondary Vertex" algorithm*

Since essentially, candidate tracks for SV fit are those incompatible with the PV *(related to significance)*, the SV finding efficiency also decreases with the tracker misalignment:

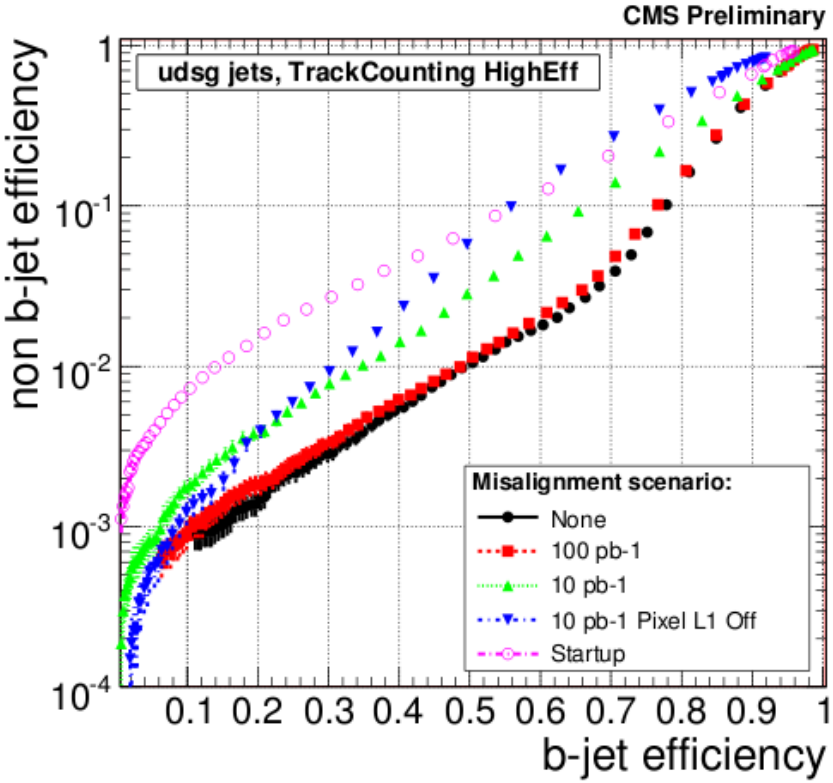
Misalignment scenario	Secondary vertex fraction [%] b-jets	c-jets	udsg-jets
No Misalignment	62.6	22.0	2.7
100 pb^{-1} Misalignment	62.1	19.6	2.4
10 pb^{-1} Misalignment	53.0	12.7	2.9
10 pb^{-1} Pixel L1 Off Misal.	39.2	7.6	1.9
Startup Misalignment	37.8	7.7	3.5

Secondary Vertex Observables

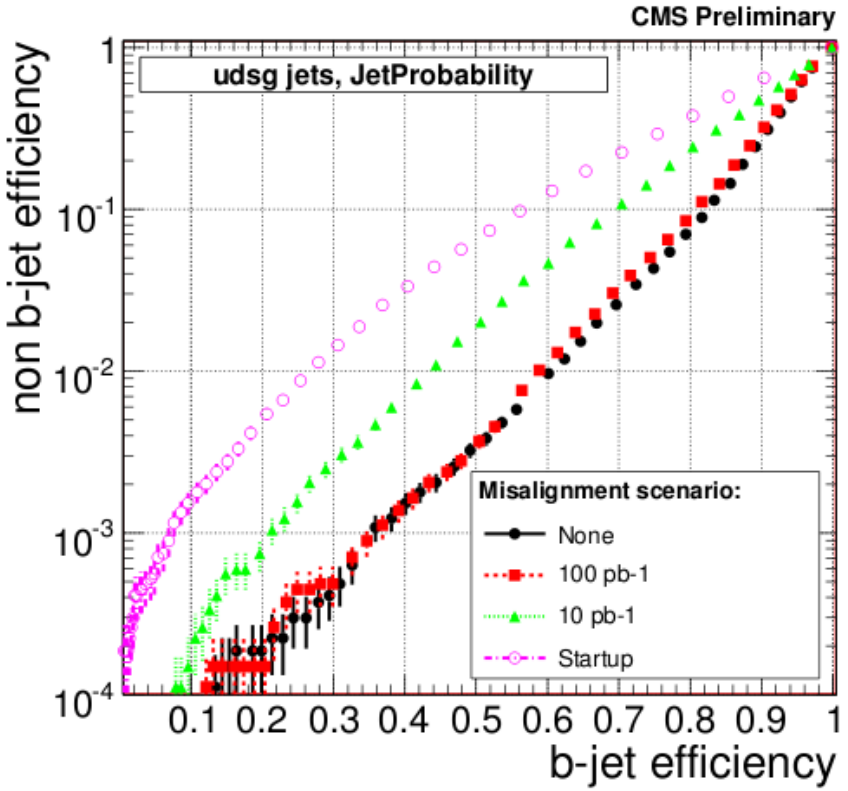
Given a reconstructed SV, observables derived that do not directly depend on errors, are in fact more robust against alignment quality.



Algorithm Performance



"Track Counting":
2nd -highest IP significance



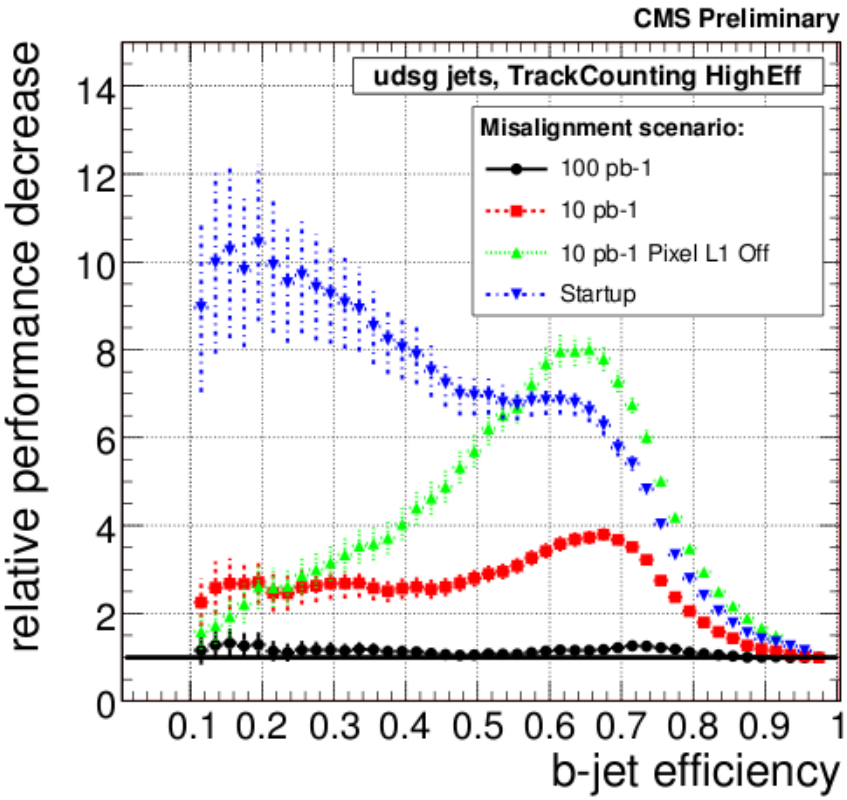
"Jet Probability":
Combination of all "Track Probabilities"
(per-track IP sig. pdf's)

data points represent light flavour mistag vs. b-tagging efficiency
for different working points *(discriminator cuts)*

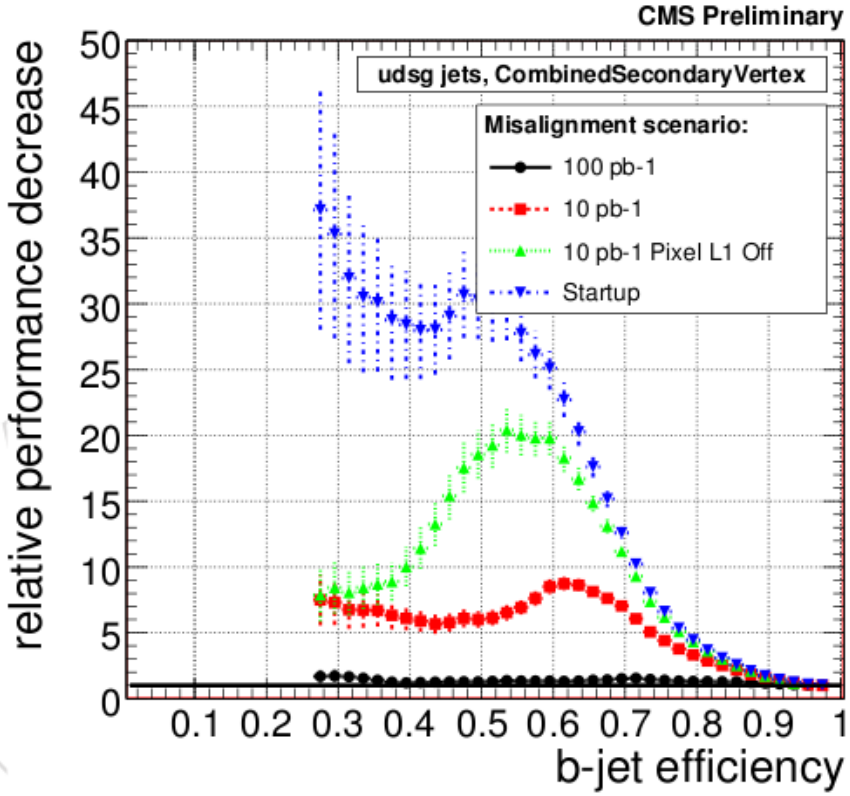
Relative Performance

One can depict the sensitivity of an algorithm by the relative increase in mistag rate at a given b-tagging efficiency

(which is not a fixed discriminator cut across scenarios!)



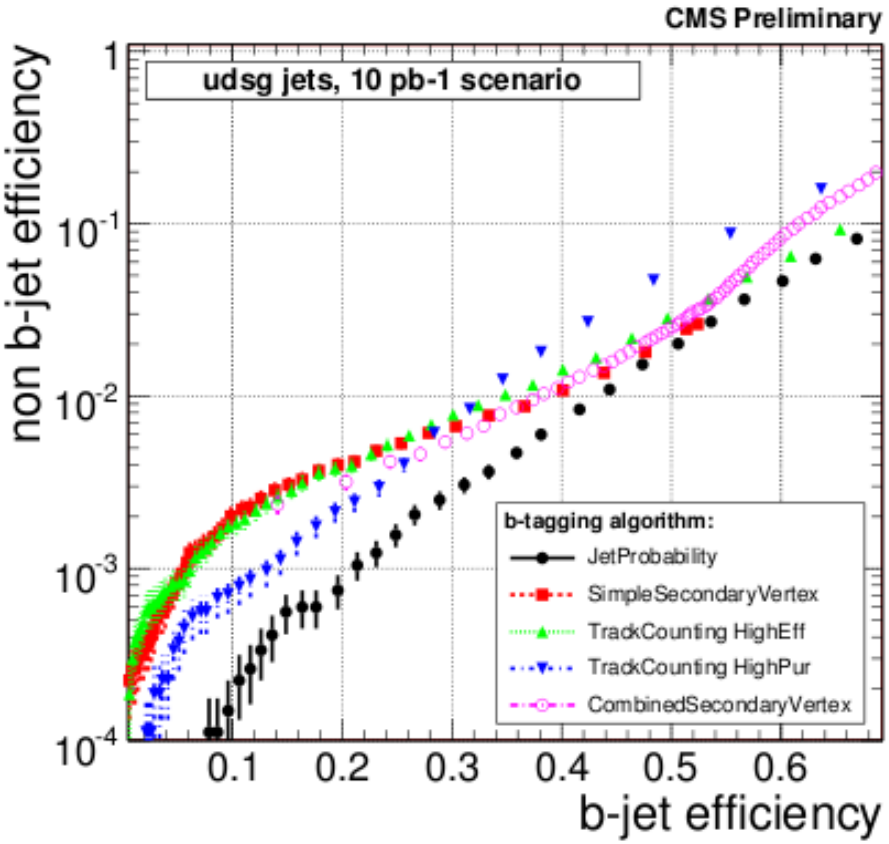
"Track Counting" 2nd track

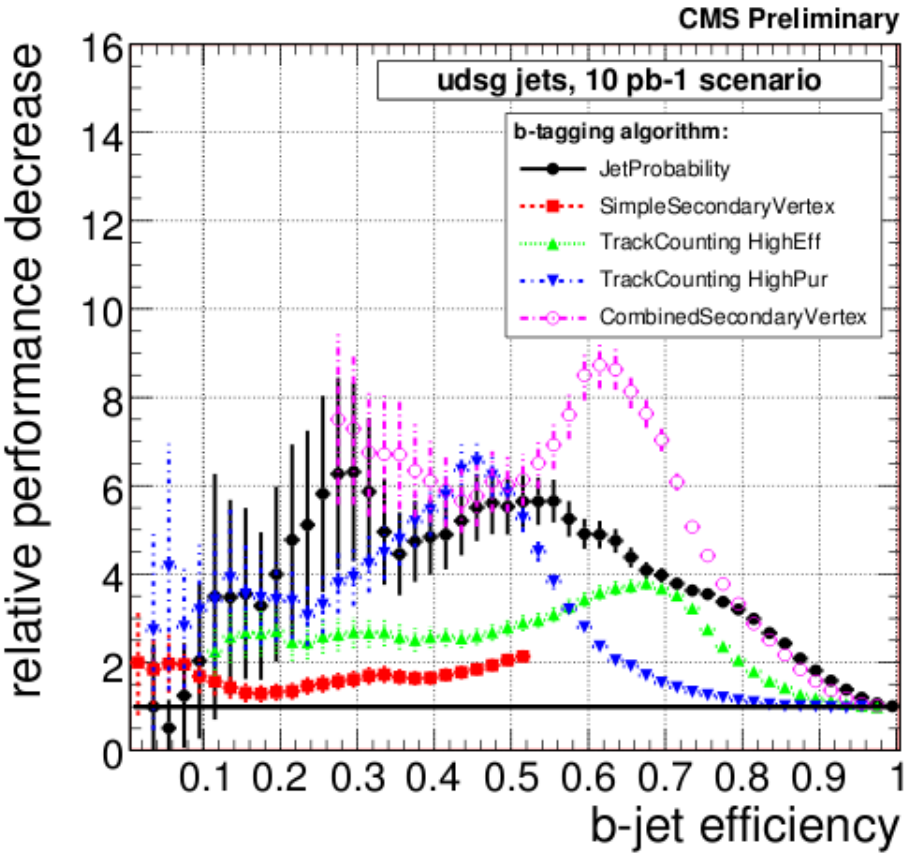


"Combined Secondary Vertex"
(IP + SV observables via MVA)

Algorithm Comparison

General observation:
The more complicated an algorithm (and the more efficient),
the more sensitive it is to misalignment





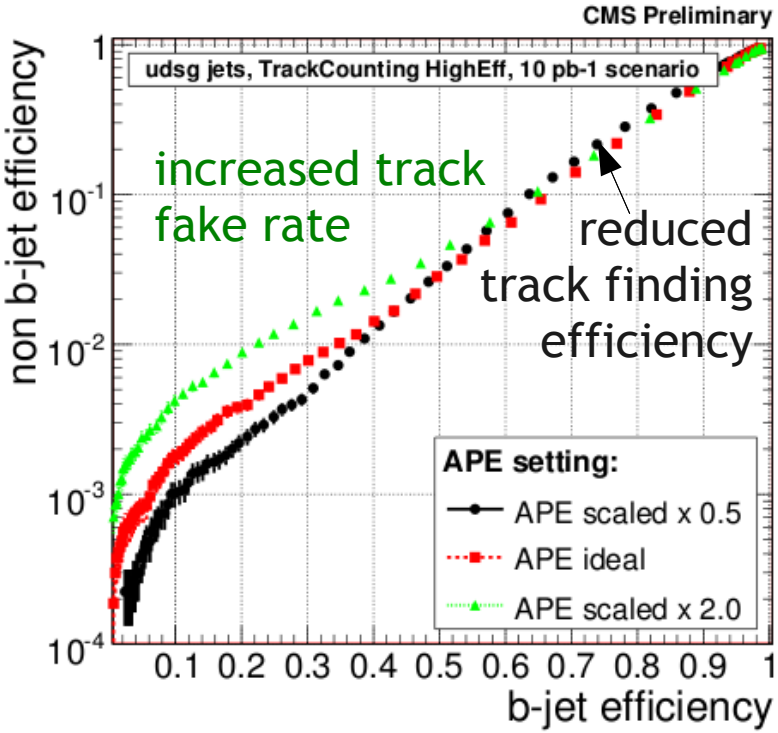
The simple "Track Counting" and "Simple SV" algorithm are presumably easiest
To get under control with early data. They do not need any "training" on MC.
→ with 10pb^{-1} of data, b-tagging will already be usable

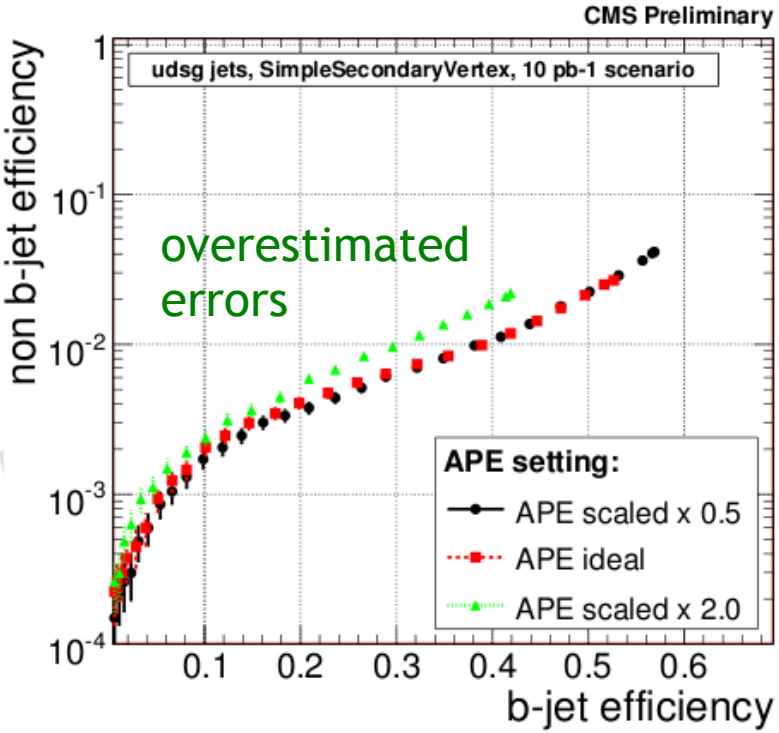
Alignment Position Error

As seen earlier, the significances strongly depend on the choice for the APE, that Is added in quadrature to the hit resolution.

For the simple (*pdf*-less) b-tagging algorithms, a different selection of the APE has no direct effect on the performance. However, a significant effect comes from differences in track and vertex reconstruction.

The effect on full track reconstruction with varied APE:





10 pb^{-1} scenario

"Track Counting" *(left)*

and

"Simple SV" *(right)*

algorithms

A factor of 2, as shown here, is very pessimistic!

Conclusions

- With **100pb^{-1}** of data: ***close to optimal alignment for b-tagging***
- With the **10pb^{-1}** scenario, **b-tagging should already be usable!**
  (i.e. a light flavour mistag of < 3% at a b-tagging efficiency of 35%)

  → which has already partly been **reached** using last year's **cosmic data**!
- simple algorithms are closely tied to understanding of tracking
- **simple algorithms** also most "**robust**" against alignment effects

  → early focus on "*Track Counting*" and "*Simple Secondary Vertex*" algorithms
- b-tagging is **highly sensitivity** to **tails of distributions**
- sensitive to effects caused by choice of APE
  → details in "CMS Tracker Alignment Results with Cosmic Muons" (E.Migliore) presentation