R_b and b-tagging efficiency

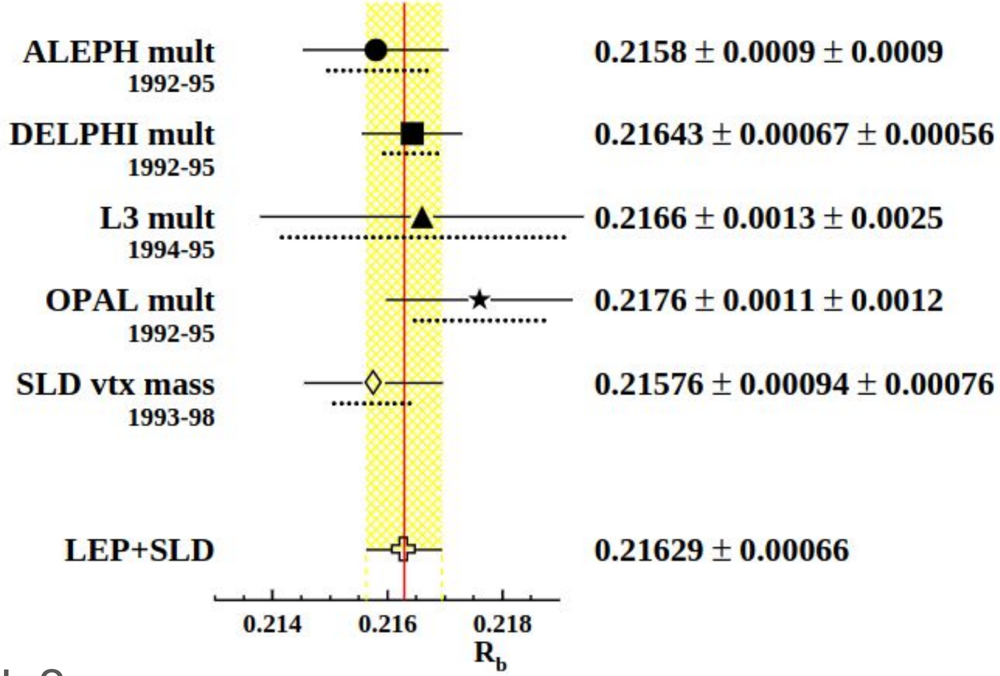
Loukas Gouskos, Emmanuel Perez, Michele Selvaggi (CERN)

R_b at LEP and SLD

$$R_b{}^0 = \Gamma_{b\bar{b}}/\Gamma_{had}$$

$\Delta R_b/R_b$ (combined) ~ 0.002

~ 50% (syst) 50% (stats)



Is 10, 100x better than LEP possible?

Double-tag method (ideal case)

$$f_S = \epsilon_b R_b$$
$$f_D = \epsilon_b^2 R_b$$

$\longrightarrow$

$$R_b = \frac{f_S^2}{f_D} \quad \text{and} \quad \epsilon_b = \frac{f_D}{f_S}$$

- Measure ratio of single and double tag hemisphere:
  - f_S, f_D
- Simultaneous extraction of: R_b, ϵ_b

<u>Advantage:</u>
Measure directly b-tagging efficiency from data

Expected stat. precision ~ $$\left(\frac{\Delta R_b}{R_b}\right)_{\mathrm{stat}} \approx \frac{\sqrt{N(Z \to b\bar{b})}}{\epsilon_b^2}$$ **~ 1e-6**

Double-tag method (with hemisphere correlations)

https://arxiv.org/pdf/hep-ex/9810002.pdf

$$f_S = \epsilon_b R_b + \epsilon_c R_c + \epsilon_{uds}(1 - R_b - R_c)$$
$$f_D = (1 + C_b)\epsilon_b^2 R_b + (1 + C_c)\epsilon_c^2 R_c + (1 + C_{uds})\epsilon_{uds}^2(1 - R_b - R_c)$$

- Measure single and double tag fractions : f_S, f_D
- Extract POIs: R_b, ε_b

Input from MC:

- Mistag rates: ε_c, ε_{uds},
- tag correlation coefficients: C_b, C_c, C_{uds}
- theory: R_c

Expected stat. precision ~ 1e-6

Advantage:
Measure directly b-tagging efficiency from data

Assumptions and selection

- 2 years at sqrt(s) = 91.188 GeV
- σ(ee→ had) = 30050 pb (at NLO QCD)
- N(ee→ had) = 1.13e12 events

- R_b = 0.2155
- R_c = 0.1720
- $R_q = 1 - R_b - R_c$

- no backgrounds (ee → τ τ) < 0.1%
- no "had" preselection (E_{vis}, N_{tracks})
  - neglecting bias introduced by preselection

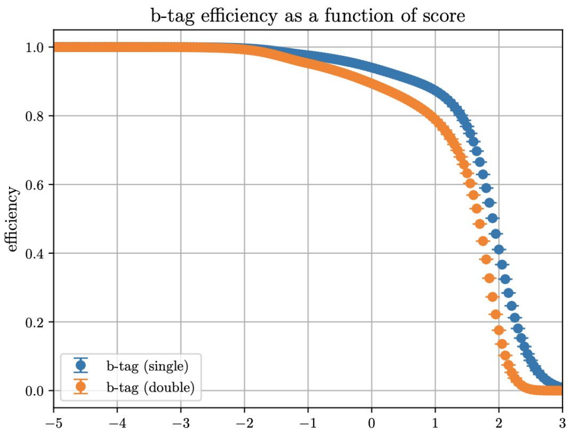
Selection

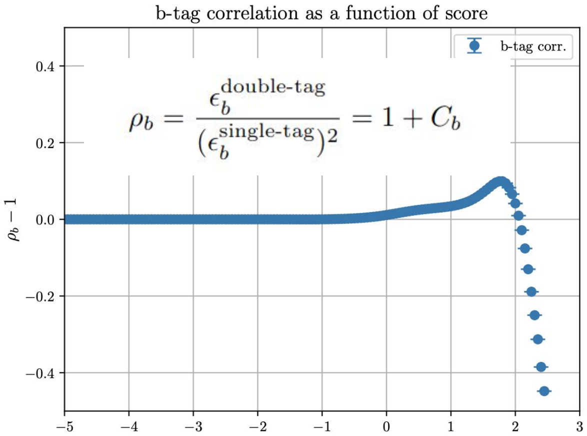
- N = 2 Durham k_T clustering
- $\cos(\theta_T) < 0.7$

Efficiency ~ 60%

Impact of systematics on R_b and ε_b precision

B-tagging efficiency and correlation



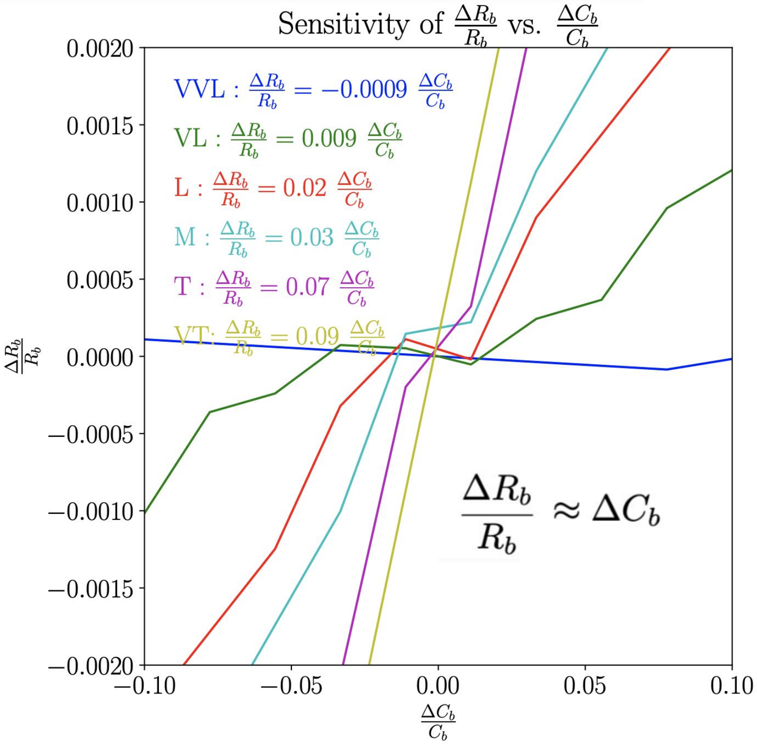


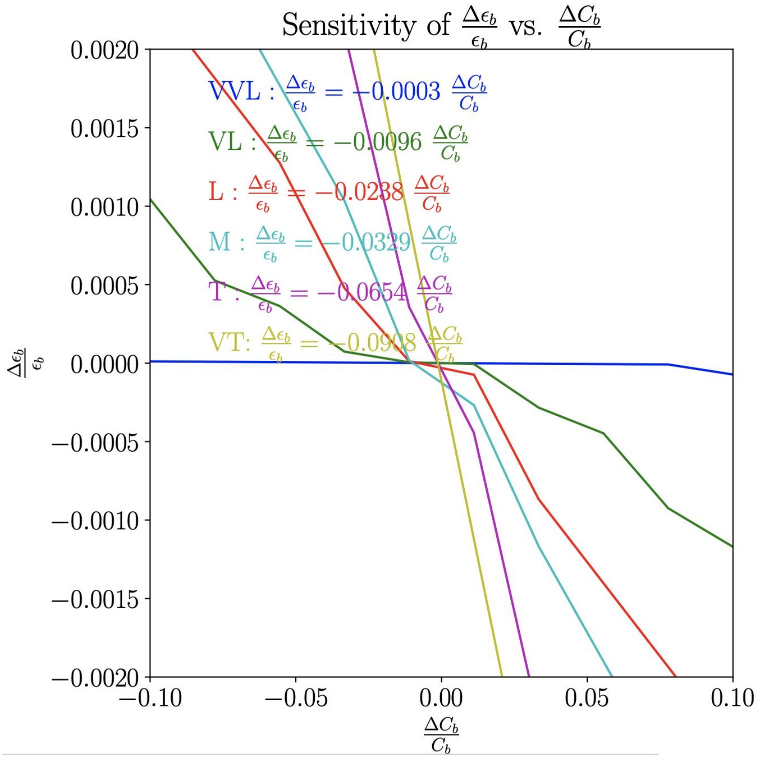
Working points:

- VVL : $\epsilon_b = 96\%,\ \epsilon_c = 3\%,\ \epsilon_{uds} = 0.6\%,\ C_b = 0.25\%$
- VL : $\epsilon_b = 94\%,\ \epsilon_c = 1\%,\ \epsilon_{uds} = 0.3\%,\ C_b = 1.1\%$
- L : $\epsilon_b = 91\%,\ \epsilon_c = 0.5\%,\ \epsilon_{uds} = 0.2\%,\ C_b = 2.5\%$
- M : $\epsilon_b = 87\%,\ \epsilon_c = 0.2\%,\ \epsilon_{uds} = 0.08\%,\ C_b = 3.4\%$
- T : $\epsilon_b = 77\%,\ \epsilon_c = 0.08\%,\ \epsilon_{uds} = 0.03\%,\ C_b = 7.0\%$
- VT: $\epsilon_b = 63\%,\ \epsilon_c = 0.04\%,\ \epsilon_{uds} = 0.02\%,\ C_b = 9.9\%$

Sensitivity to C_b

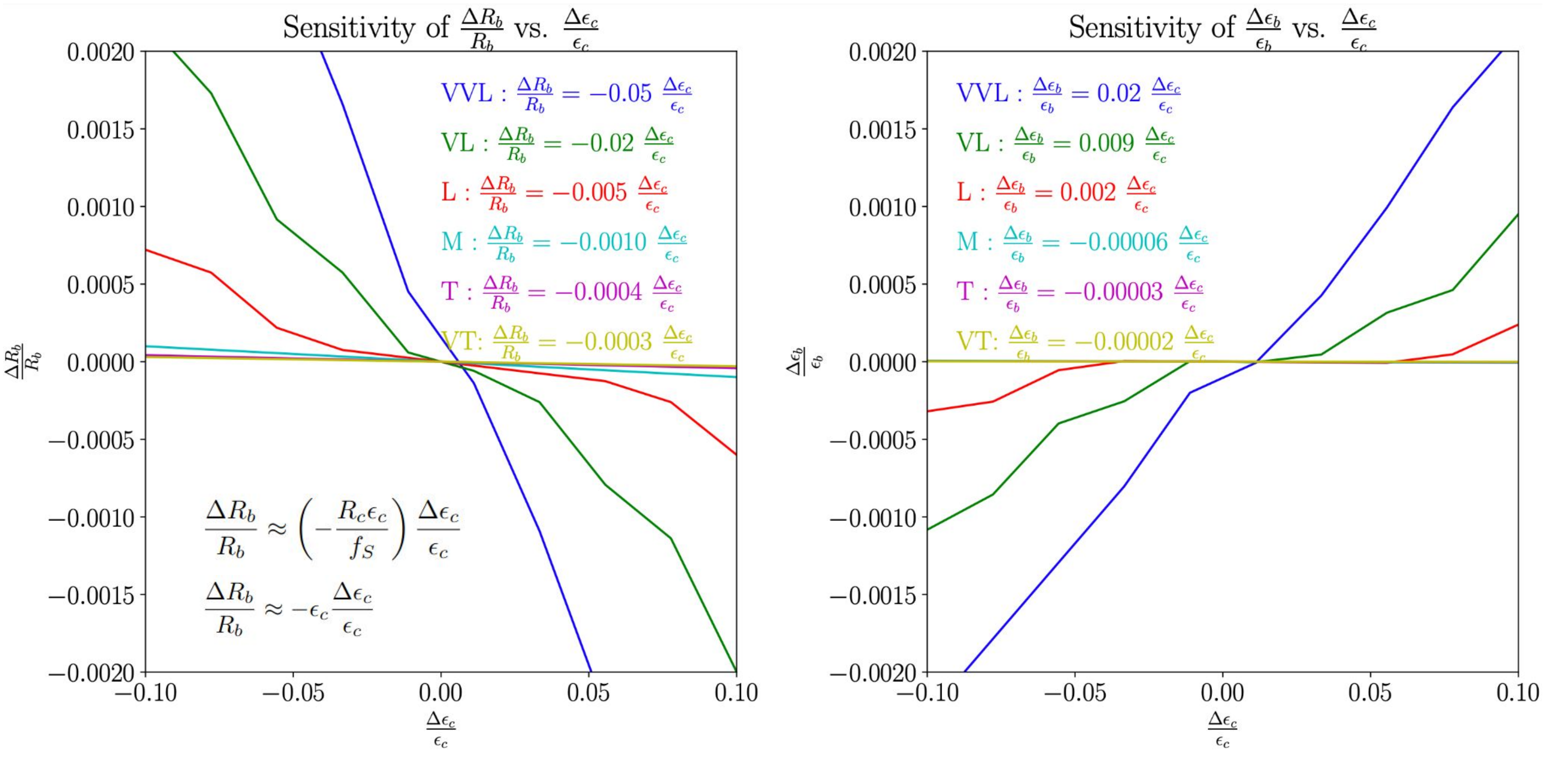
$$\frac{\Delta R_b}{R_b} = \frac{\Delta C_b}{1+C_b} = \frac{C_b}{1+C_b}\frac{\Delta C_b}{C_b} \approx C_b \frac{\Delta C_b}{C_b} = \Delta C_b$$





- **Loose** tagging WPs are preferred to minimise syst. of C_b to R_b

Sensitivity to charm mistag rate



- **Tight** tagging WPs are preferred to minimise impact of charm mistag

Comments

- Assuming all systematics of the same size (and independent of the tagging purity), largest sensitivity
  - $C_b > \varepsilon_c > \varepsilon_{uds} > C_c > C_l$
- To minimise impact of systematic uncertainties, trade-off between possible b-tagging working points of various purities
  - C_b prefers loose tag WP
  - mistag rates prefer tight WP

Systematics at LEP (OPAL)

https://arxiv.org/pdf/hep-ex/9810002.pdf

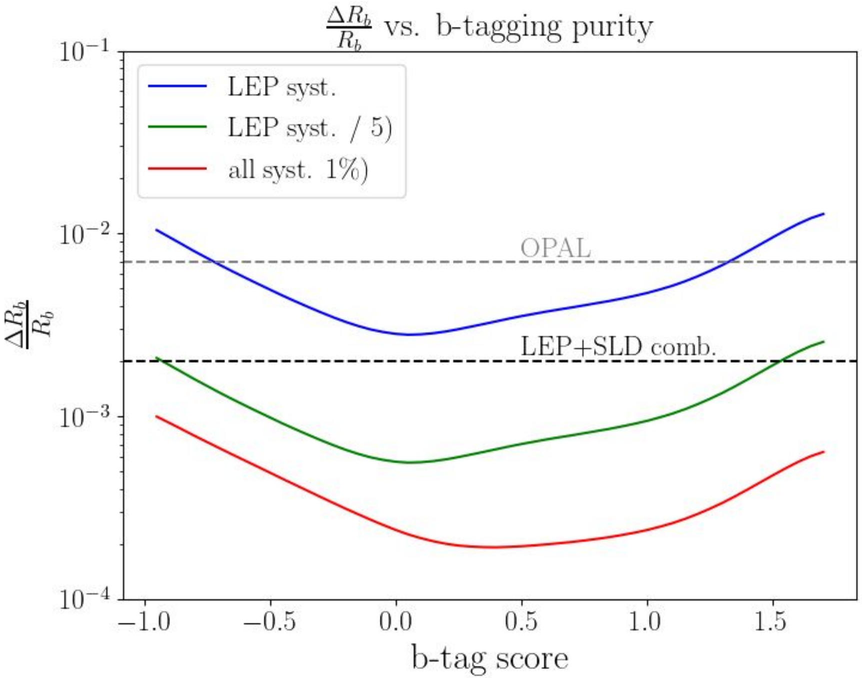
Source	$\Delta\epsilon^{c}/\epsilon^{c}$ (%)	$\Delta\epsilon^{uds}/\epsilon^{uds}$ (%)	ΔR_{b}
Tracking resolution	1.24	4.0	0.00017
Tracking efficiency	0.80	4.0	0.00014
Silicon hit matching efficiency	0.82	2.8	0.00009
Silicon alignment	0.58	2.1	0.00008
Electron identification efficiency	1.11	0.5	0.00015
Muon identification efficiency	0.64	0.2	0.00009
c quark fragmentation	2.26	-	0.00028
c hadron production fractions	3.66	-	0.00046
c hadron lifetimes	0.55	-	0.00007
c charged decay multiplicity	1.09	-	0.00014
c neutral decay multiplicity	2.39	-	0.00030
Branching fraction $B(\mathrm{D} \to \mathrm{K}^{0})$	1.20	-	0.00015
c semileptonic branching fraction	2.44	-	0.00031
c semileptonic decay modelling	2.34	-	0.00029
Gluon splitting to $\mathrm{c\bar{c}}$	0.34	6.3	0.00018
Gluon splitting to $\mathrm{b\bar{b}}$	0.50	9.3	0.00027
K^{0} and hyperon production	-	0.3	0.00001
Monte Carlo statistics (c, uds)	0.66	2.5	0.00010
Subtotal $\Delta\epsilon^{c}$ and $\Delta\epsilon^{uds}$	6.65	13.3	0.00090
Electron identification background			0.00039
Muon identification background			0.00041
Efficiency correlation ΔC^{b}			0.00066
Event selection bias			0.00033
Total			0.00129

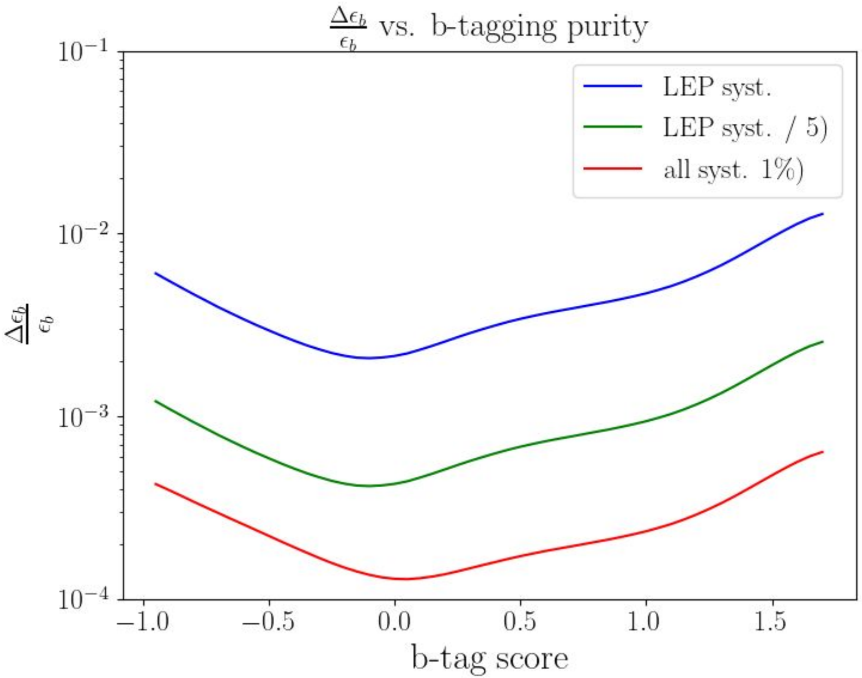
Correlation $C^{b}-1$ (%)	Vertex	Combined
Same hemisphere events	0.02 ± 0.02	-0.03 ± 0.02
Momentum correlation	0.04 ± 0.05	0.06 ± 0.03
Geometrical correlation	0.88 ± 0.02	0.71 ± 0.02
Component sum	0.94 ± 0.06	0.74 ± 0.04
Overall correlation	0.83 ± 0.20	0.93 ± 0.17

- $\Delta C_{b}/C_{b} \approx 20\%$
- $\Delta\varepsilon_{c}/\varepsilon_{c} \approx 7\%$
- $\Delta\varepsilon_{uds}/\varepsilon_{uds} \approx 13\%$

Results vs purity

<u>Assumption:</u> systematics constant over tagging score





- Optimal working point is ~ Loose depending on syst assumptions
- 1% syst. scenarios prefer slightly higher purity because larger relative reduction of error on C_b

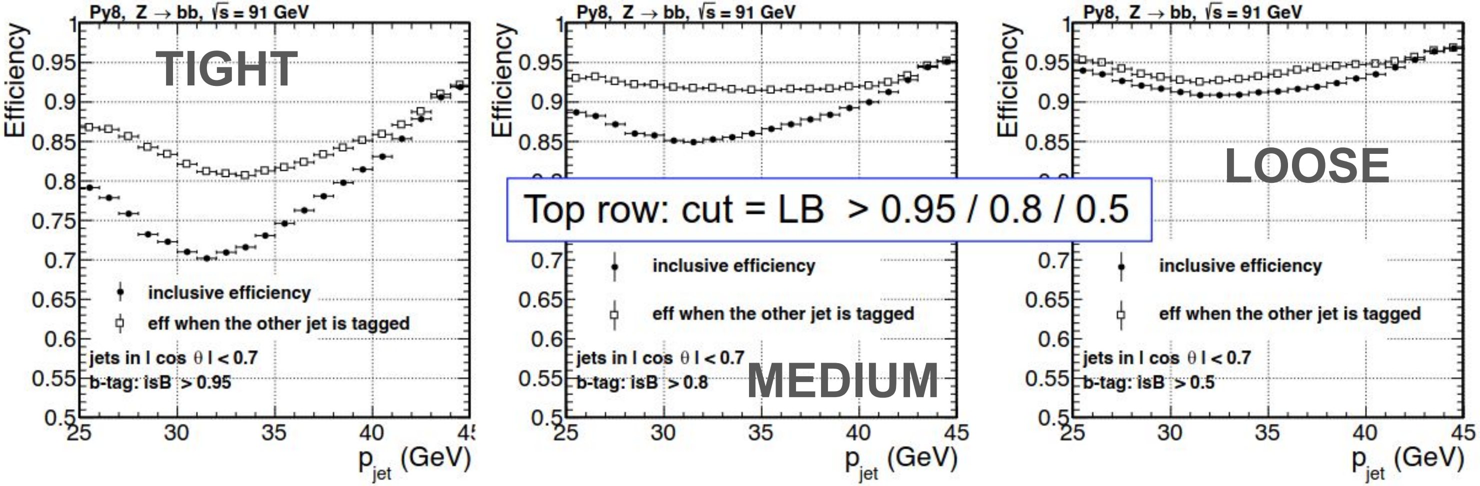
Prel. conclusion and next

- Start assessing how systematics **scale vs tagging purity**
- **Most likely this method will require to trust MC to better that 1% to achieve goal $\Delta\varepsilon_b/\varepsilon_b \sim \Delta R_b/R_b \sim$ 2e-4**
- Multi tag method to fit all flavor mis-tagging efficiencies and $R_{b,c,s,ud}$ from data will rely less on MC modelling
- Scale factors to extrapolate for Z ($\sqrt{s}$=90 GeV) to ZH ($\sqrt{s}$=240 GeV) ?

closer look at the correlation

B-tagging efficiency and correlation

cf. Emmanuel



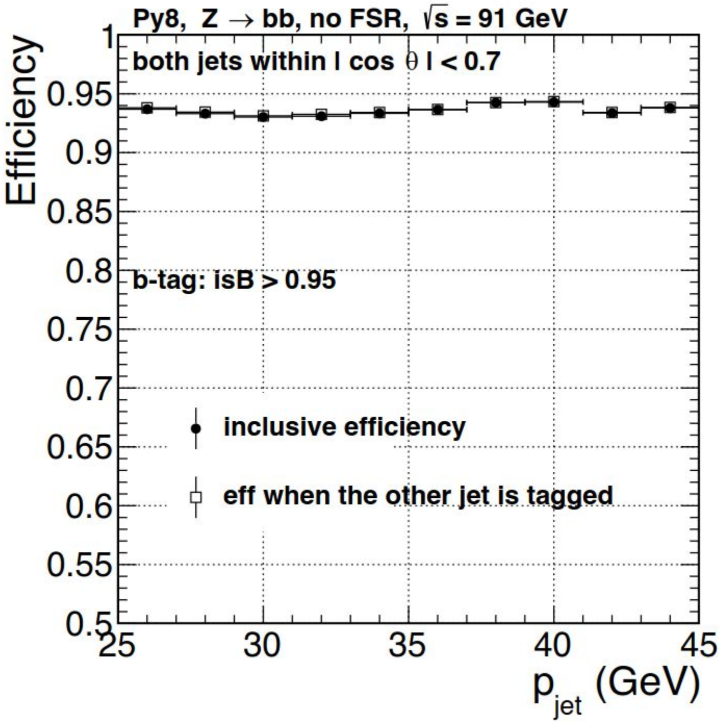
- Efficiency dependence with jet momentum
  - max eff at high p (OK)
  - non monotonic vs p (?)
- **Correlation dependence with the jet momentum**
  - max correlation at p_{jet} ~ 30 GeV
  - correlation vanishes at p_{jet} ~ E_{beam}
- **Correlation dependence with the tagger**
  - max correlation at high purities

B-tagging efficiency and correlation

cf. Emmanuel

When gluon radiation is switched off in Pythia:

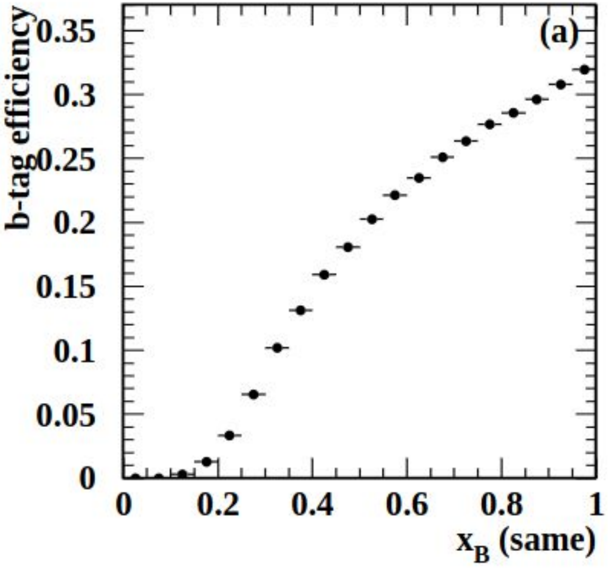
- Flat efficiency vs p
- Tagging one leg does not bias the efficiency of the other leg



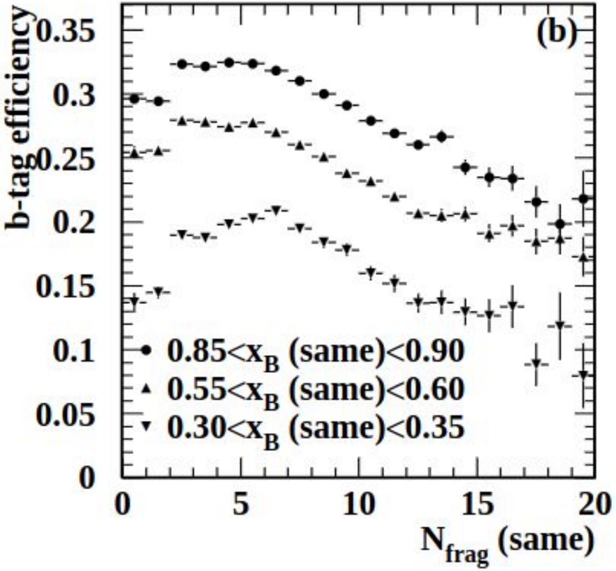
gluon radiation responsible for correlation

B-tagging efficiency and correlation

https://arxiv.org/pdf/hep-ex/9810002.pdf



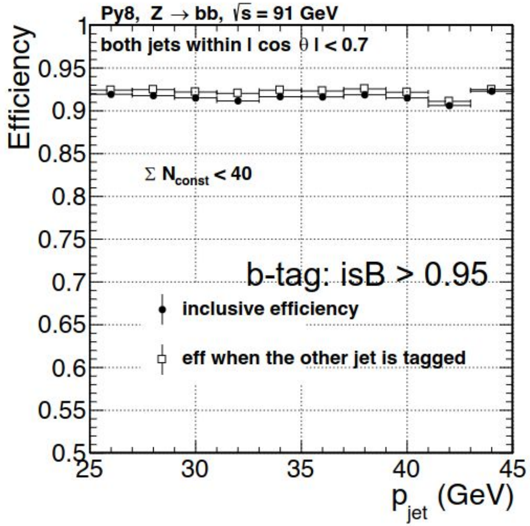
= p_B/E_{beam}

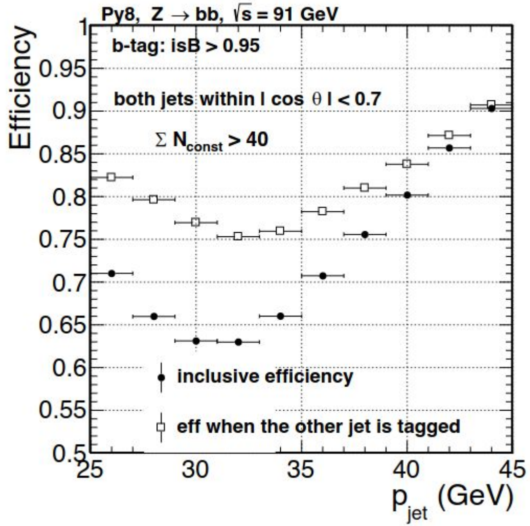


- B-tagging efficiency increases with b-hadron momentum
- B-tagging efficiency decreases if gluon emission in the same hemisphere
  - →N_{frag} (number of fragmentation tracks) increases
  - → SV more easily mistaken for PV

B-tagging efficiency and correlation

https://arxiv.org/pdf/hep-ex/9810002.pdf





- gluon emission increases number of N_{frag} tracks
  - decreases available momentum for bhadron p_B and overall p_{jet} (increase the jet mass)
  - momentum balance →opp. hemisphere also softer

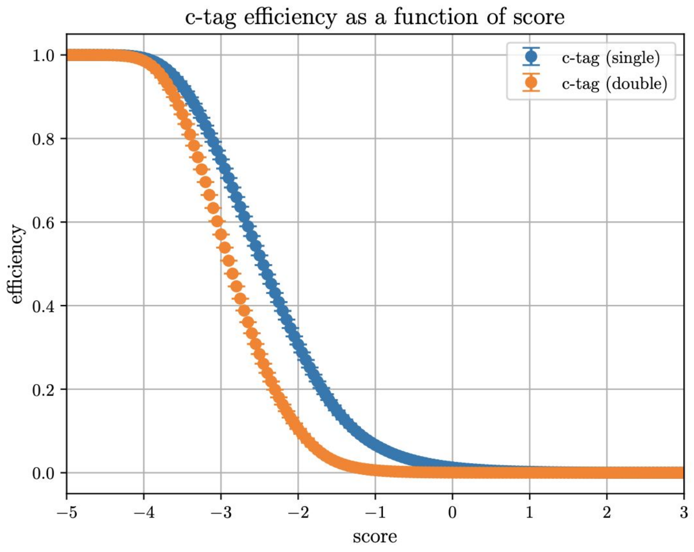
Open questions

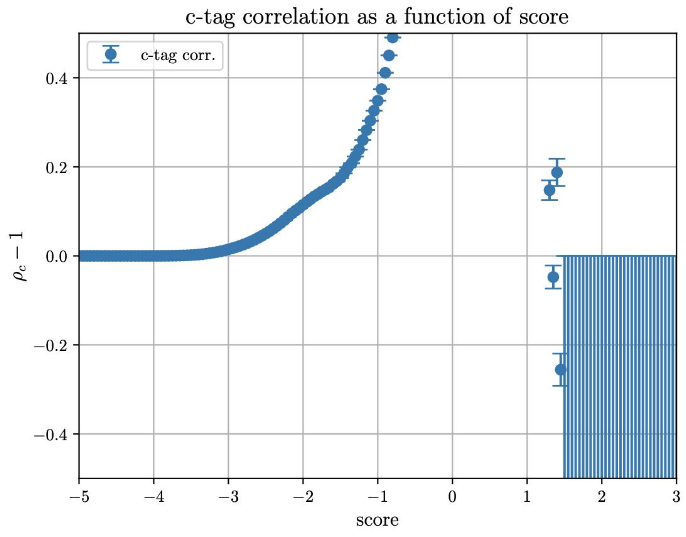
- Tagging correlation depends:
    - on the tagger purity
    - on the jet momentum
        - correlated with amount of gluon radiation in the event, and relative momentum carried by the bhadron
          → including such information is crucial to understand this effects (IN PROGRESS ..)

systematics ultimately will depend on TH parton shower and fragmentation models

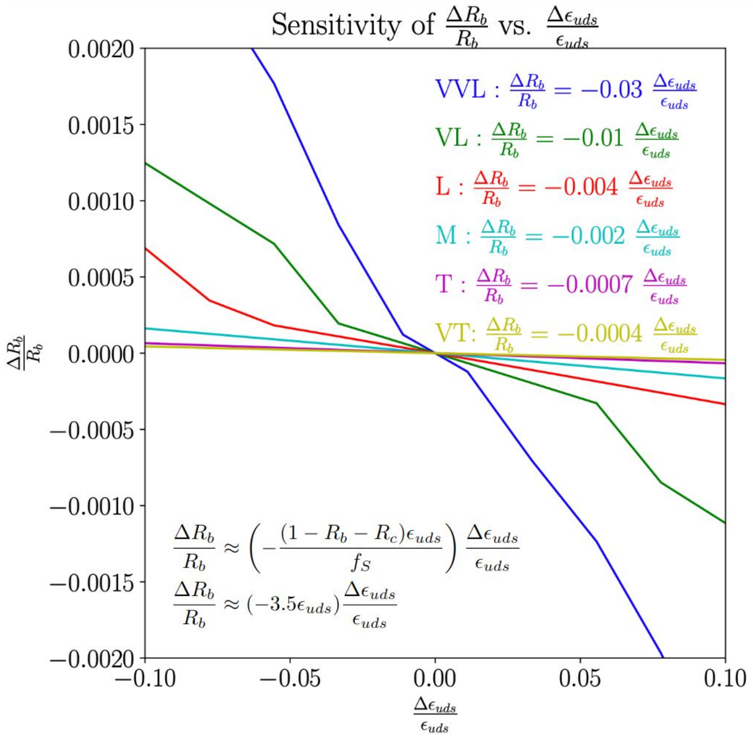
Backup

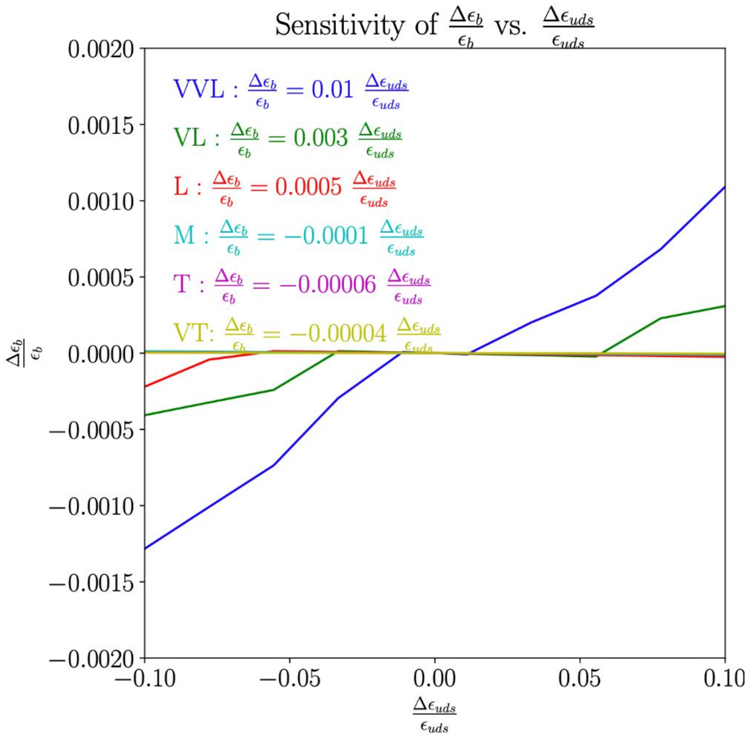
C-mistag rate efficiency and correlation





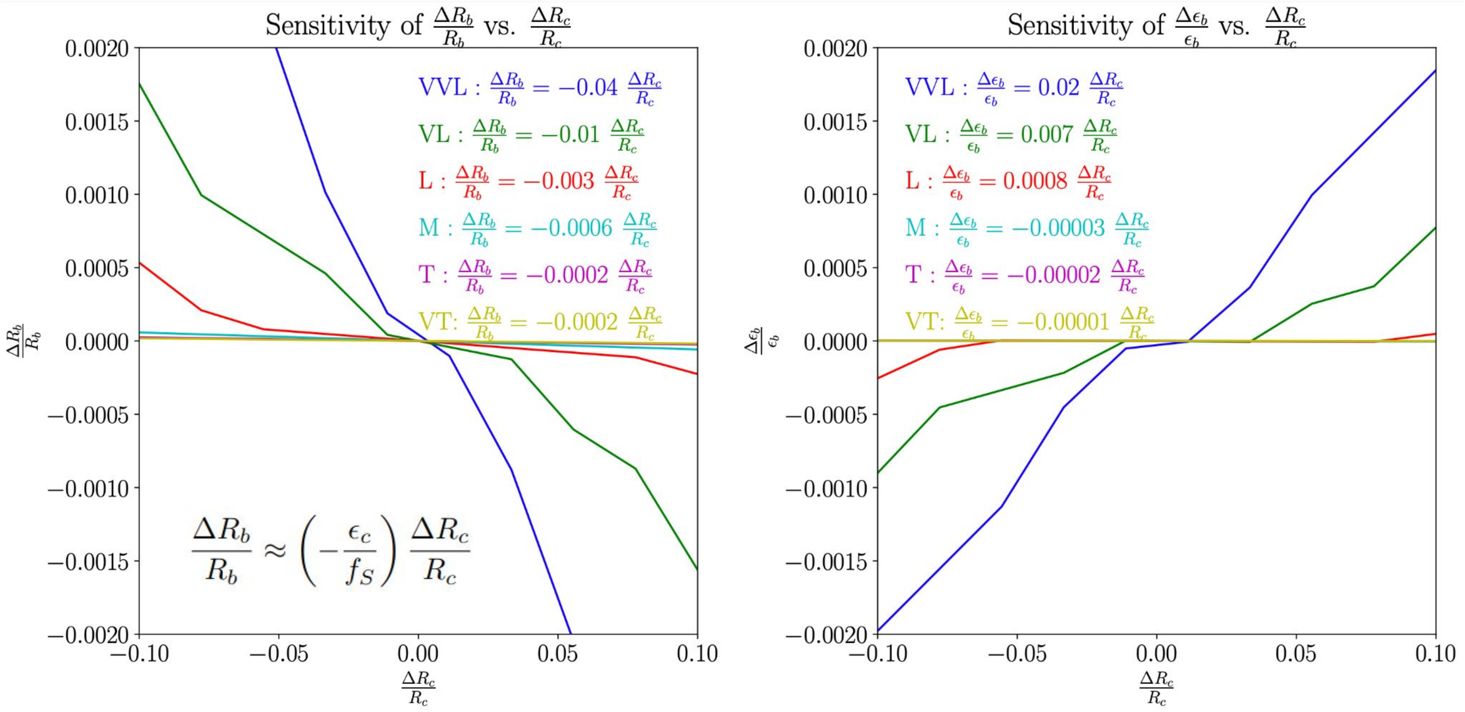
Sensitivity to light mistag rate





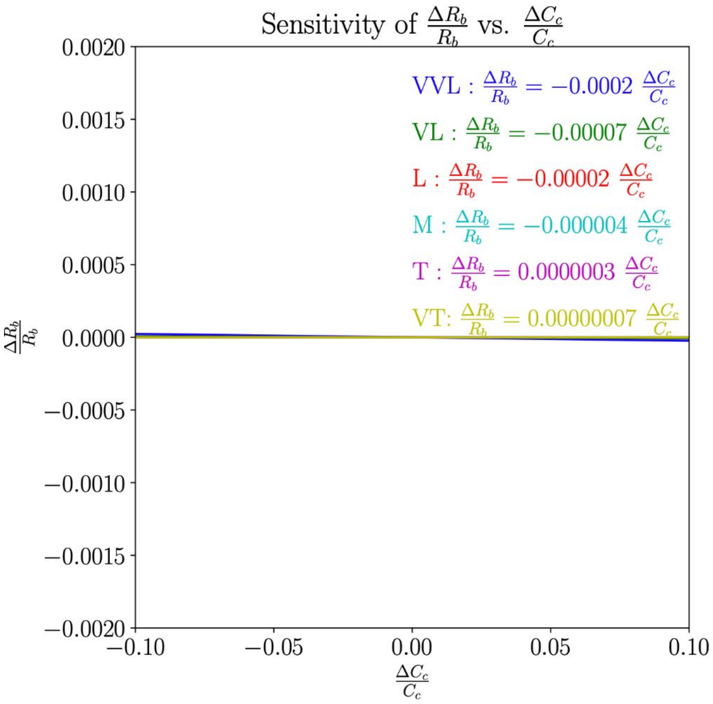
- **Tight** tagging WPs are preferred to minimise impact of light quark mistag systematics
- Similar sensitivity than ε_c, ($\varepsilon_{uds} < \varepsilon_c$ but $R_{uds} \sim 3\ R_c$)

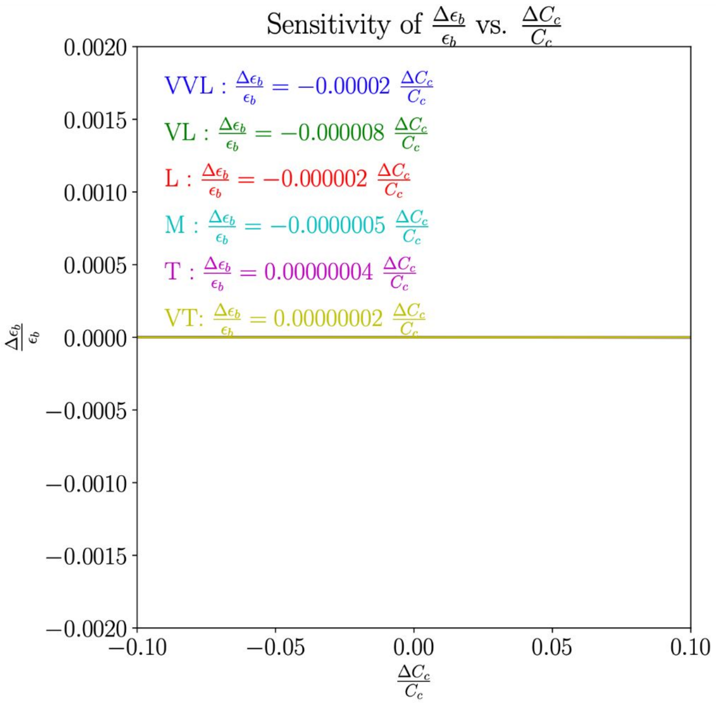
Sensitivity to R_c



- **Tight** tagging WPs are preferred to minimise impact of R_c parametric

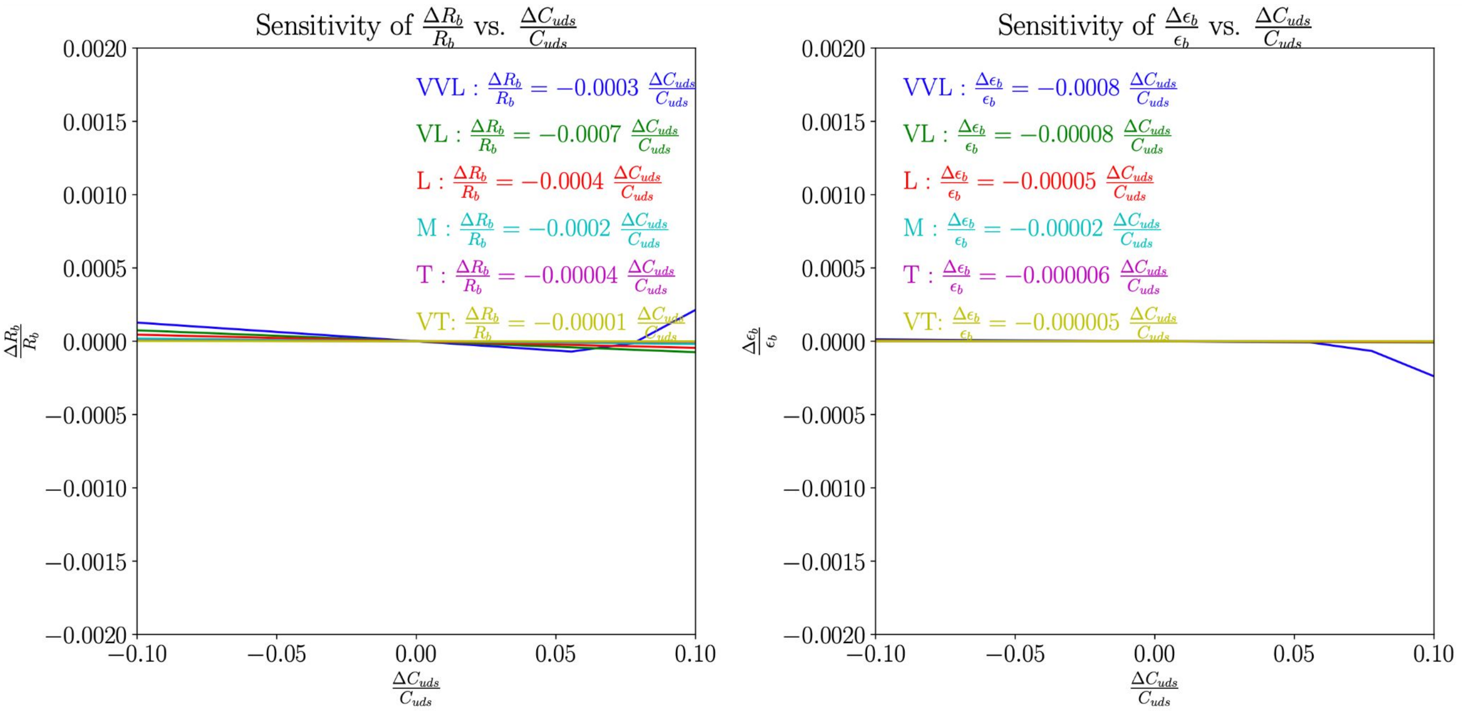
Sensitivity to C_c





- **Tight** tagging WPs are preferred to minimise impact of C_c , but almost indifferent

Sensitivity to C_{uds}



- **Tight** tagging WPs are preferred to minimise impact of C_{uds} , but almost indifferent