Heavy-flavour jet identification at the CMS experiment for Run 2

Mauro Verzetti
on behalf of the CMS collaboration

UNIVERSITY of ROCHESTER

Why flavour tagging?

Just look at the talks in this conference!

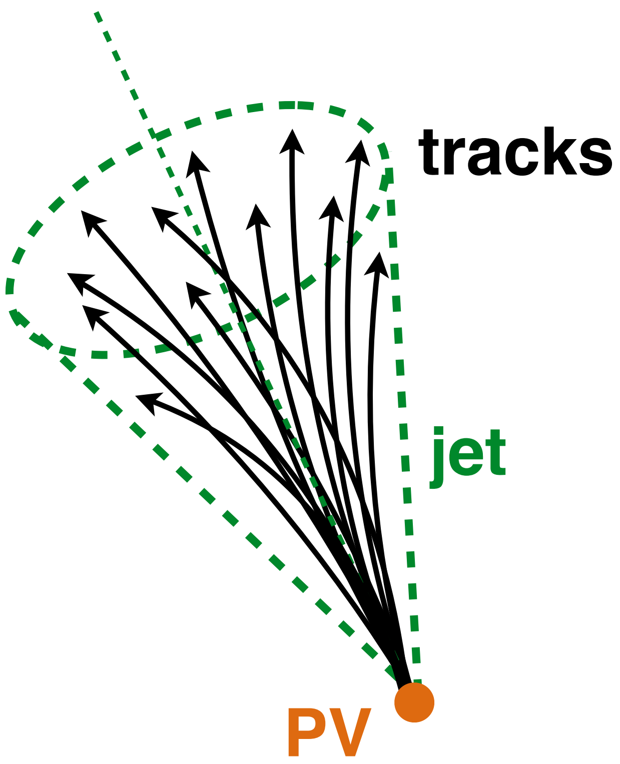
Why flavour tagging?

Just look at the talks in this conference!

Fundamental tool for:

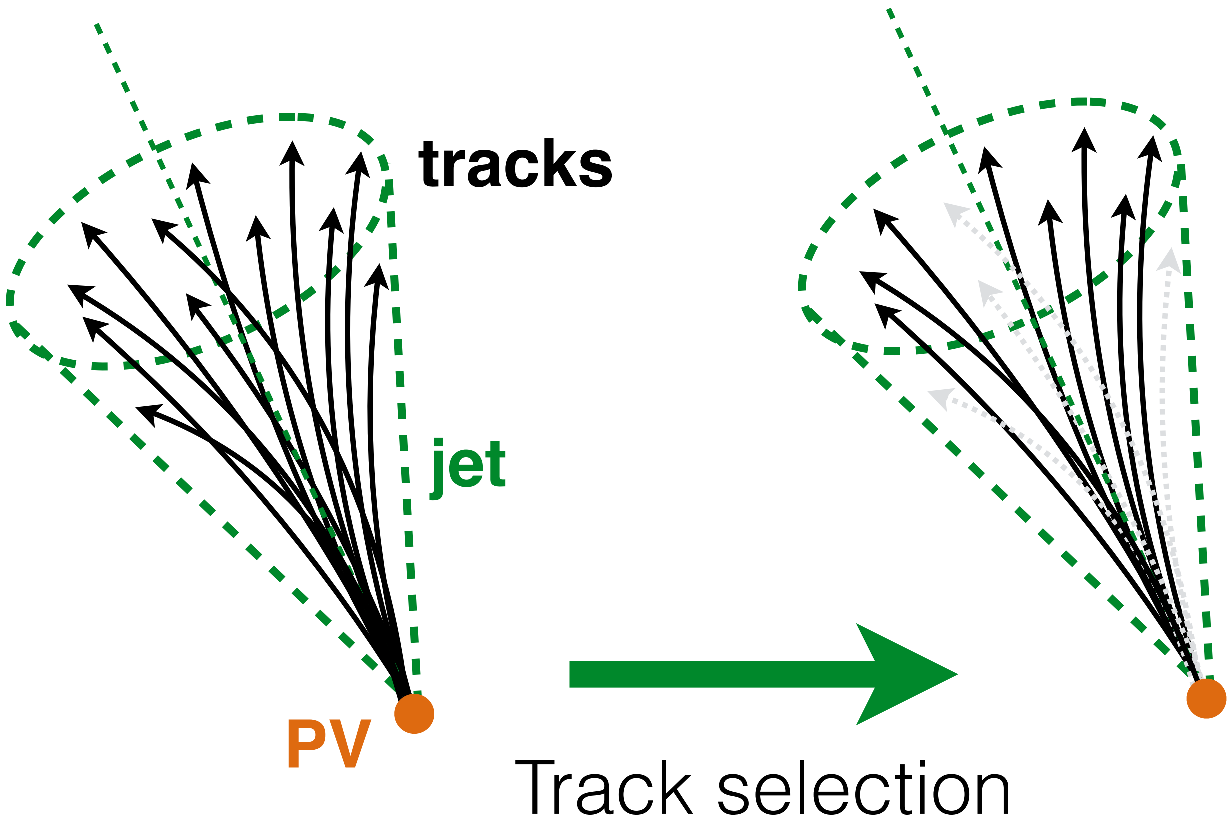
- Higgs physics ($H\rightarrow bb$, ttH)
- SUSY
- Heavy exotic resonances
- Top physics, SM

HF tagging @ CMS — BTV-15-001

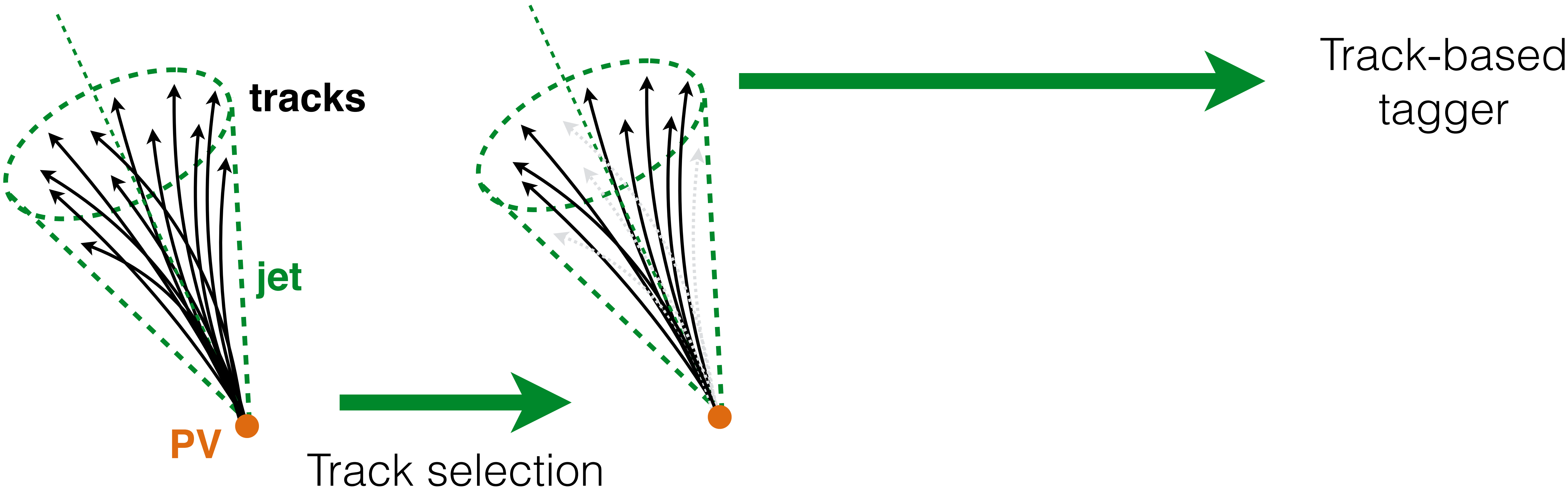


M. Verzetti, U. Rochester

HF tagging @ CMS — BTV-15-001 — J. Instrum. 8 P04013

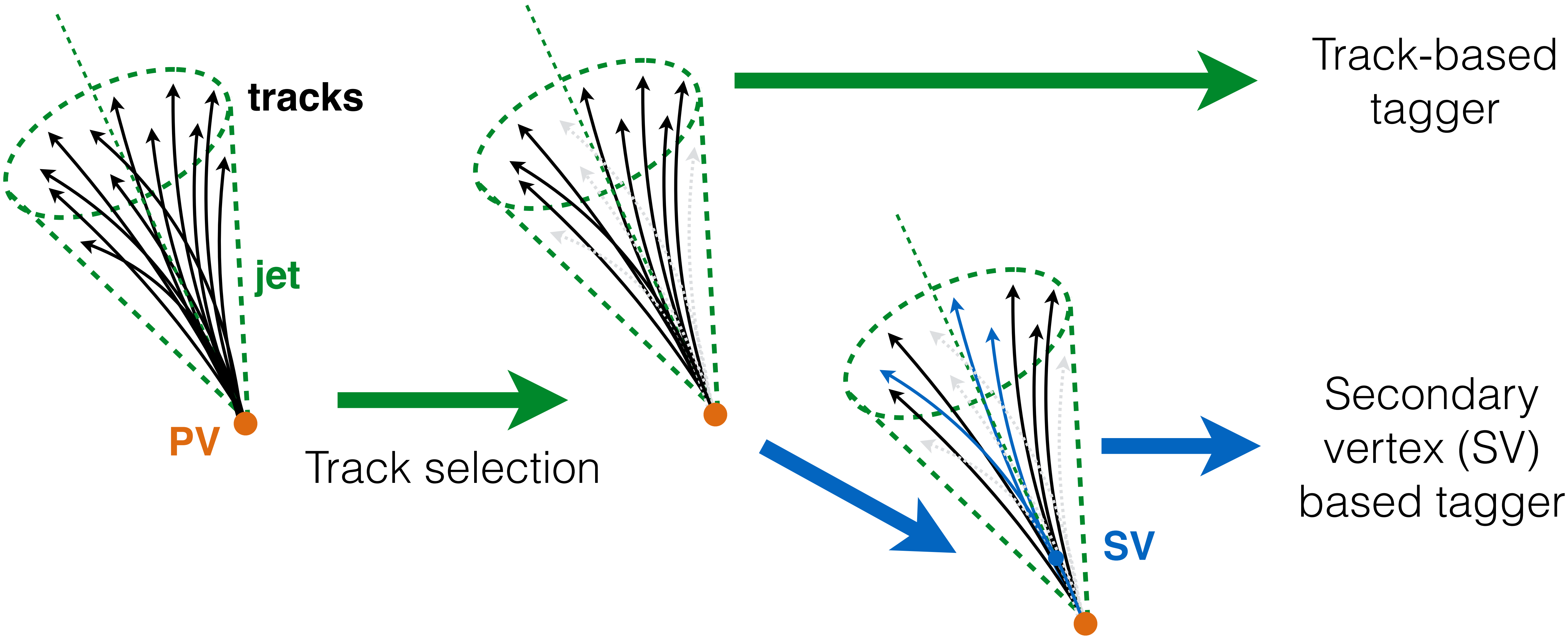


HF tagging @ CMS — BTV-15-001 — J. Instrum. 8 P04013



e.g.: **J**et **P**robability (**JP**)

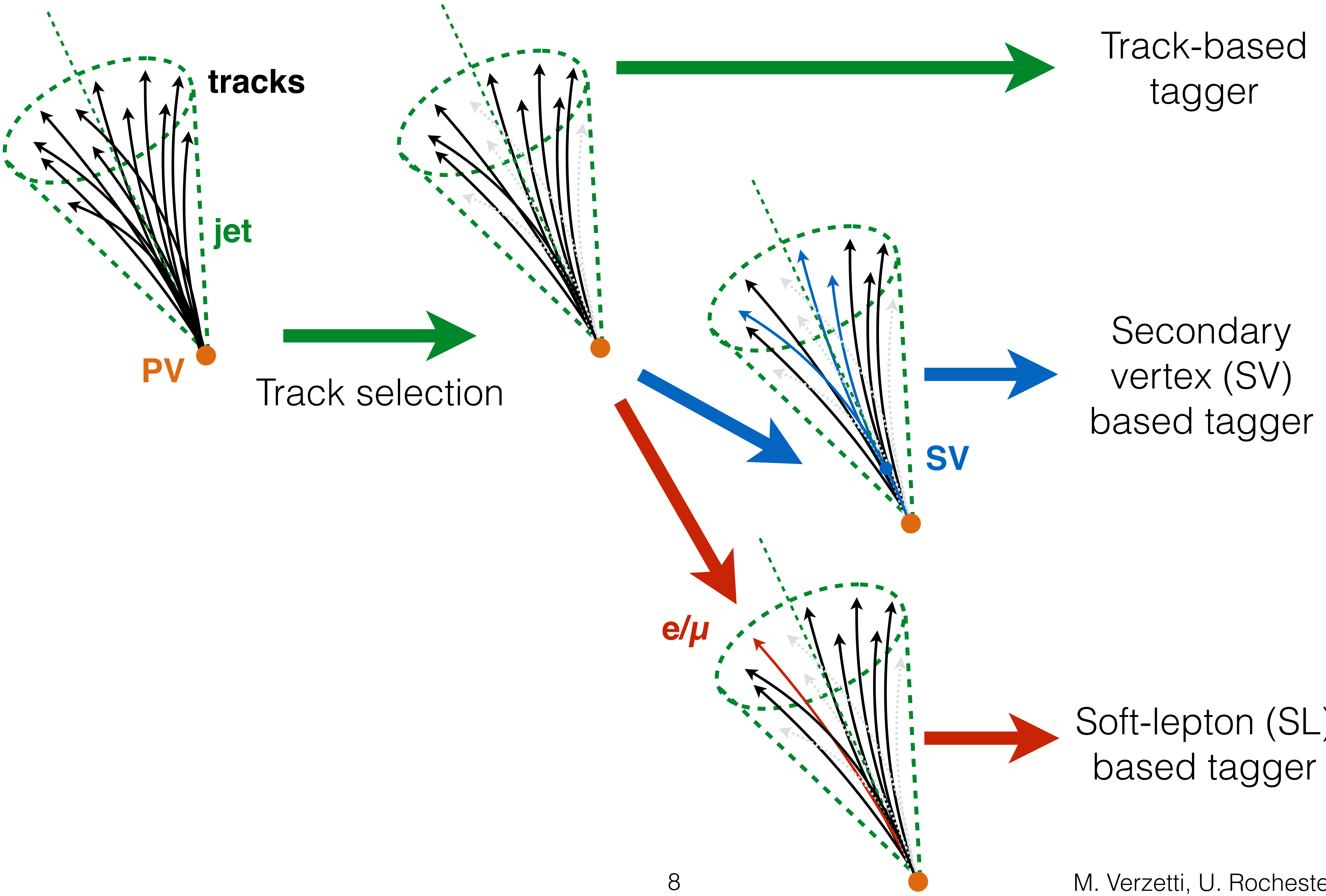
HF tagging @ CMS — BTV-15-001 — J. Instrum. 8 P04013



- **A**daptive **V**ertex **R**econstruction **(AVR)**: applied on tracks associated to the jet
- **I**nclusive **V**ertex **F**itter **(IVF)**: on the full set of tracks recorded in the event (SV ΔR-matched to jet).
  Current reconstruction default

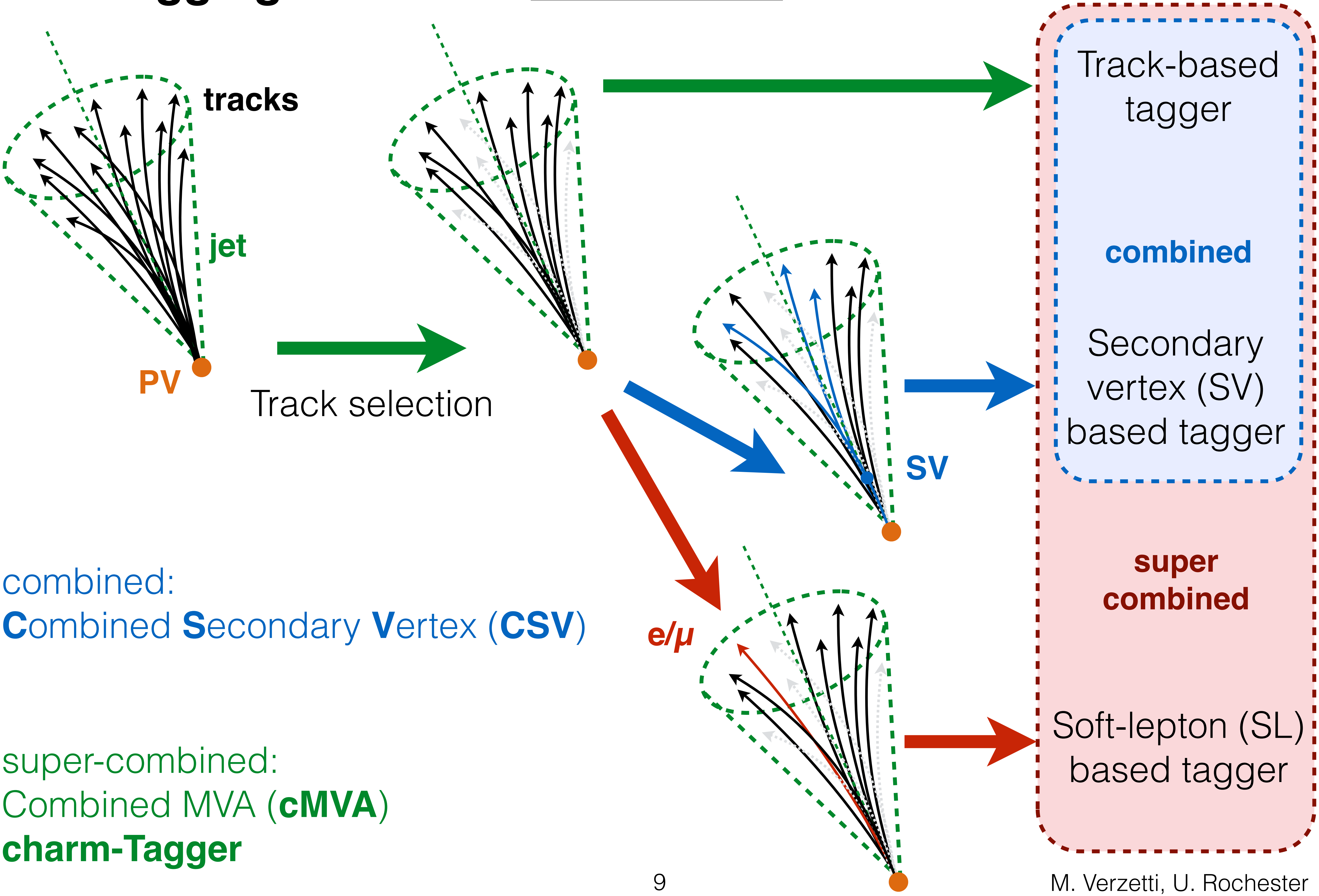
M. Verzetti, U. Rochester

HF tagging @ CMS — BTV-15-001



M. Verzetti, U. Rochester

HF tagging @ CMS — BTV-15-001

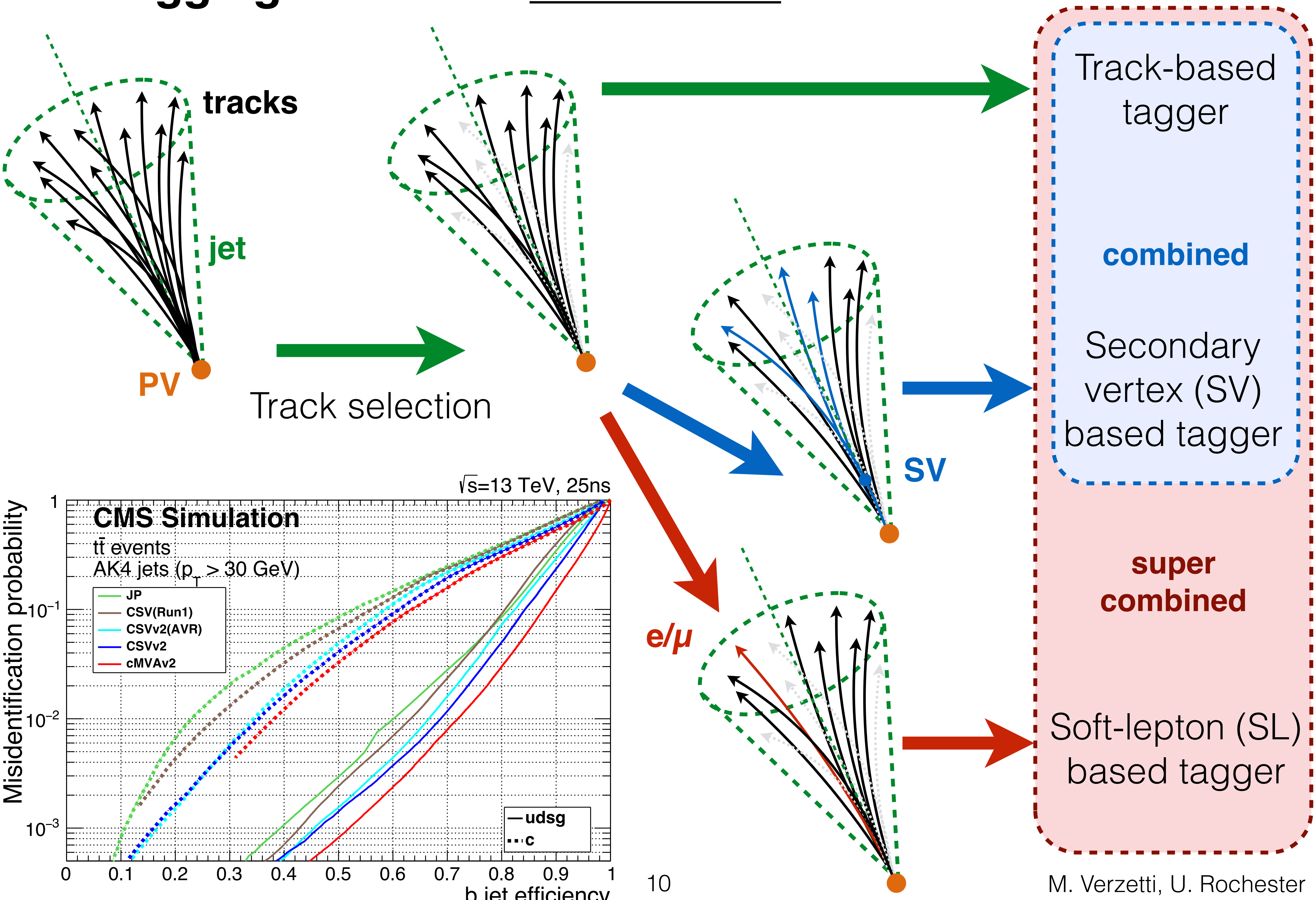


combined:
Combined **S**econdary **V**ertex (**CSV**)

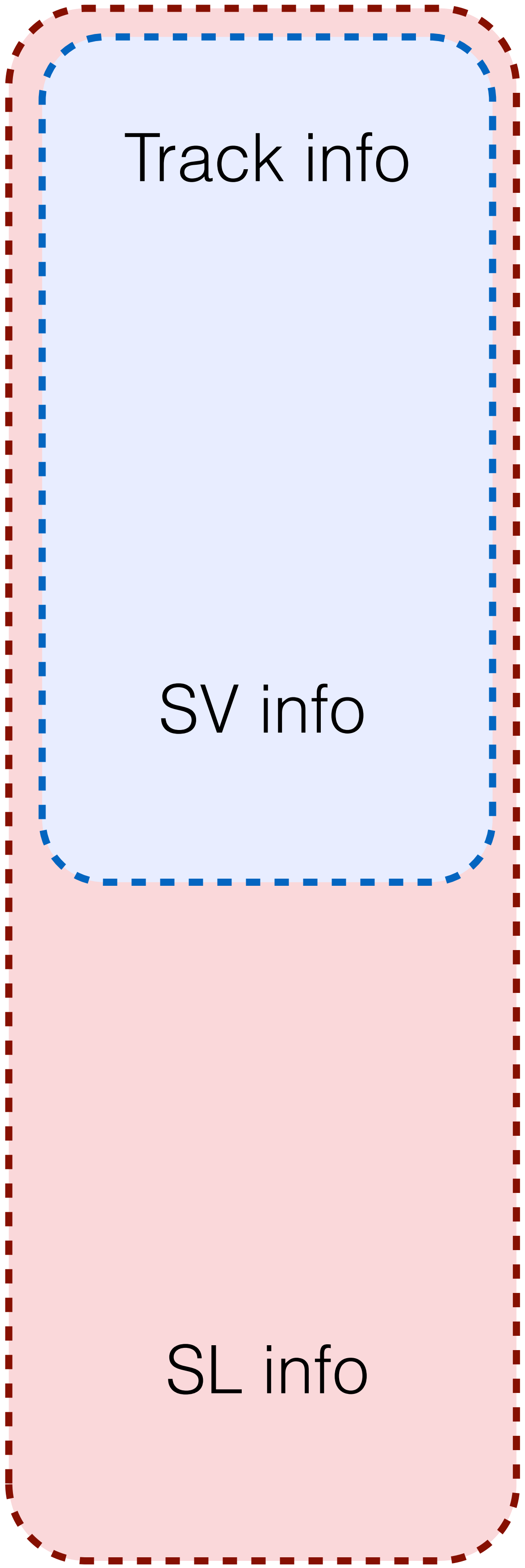
super-combined:
Combined MVA (**cMVA**)
charm-Tagger

M. Verzetti, U. Rochester

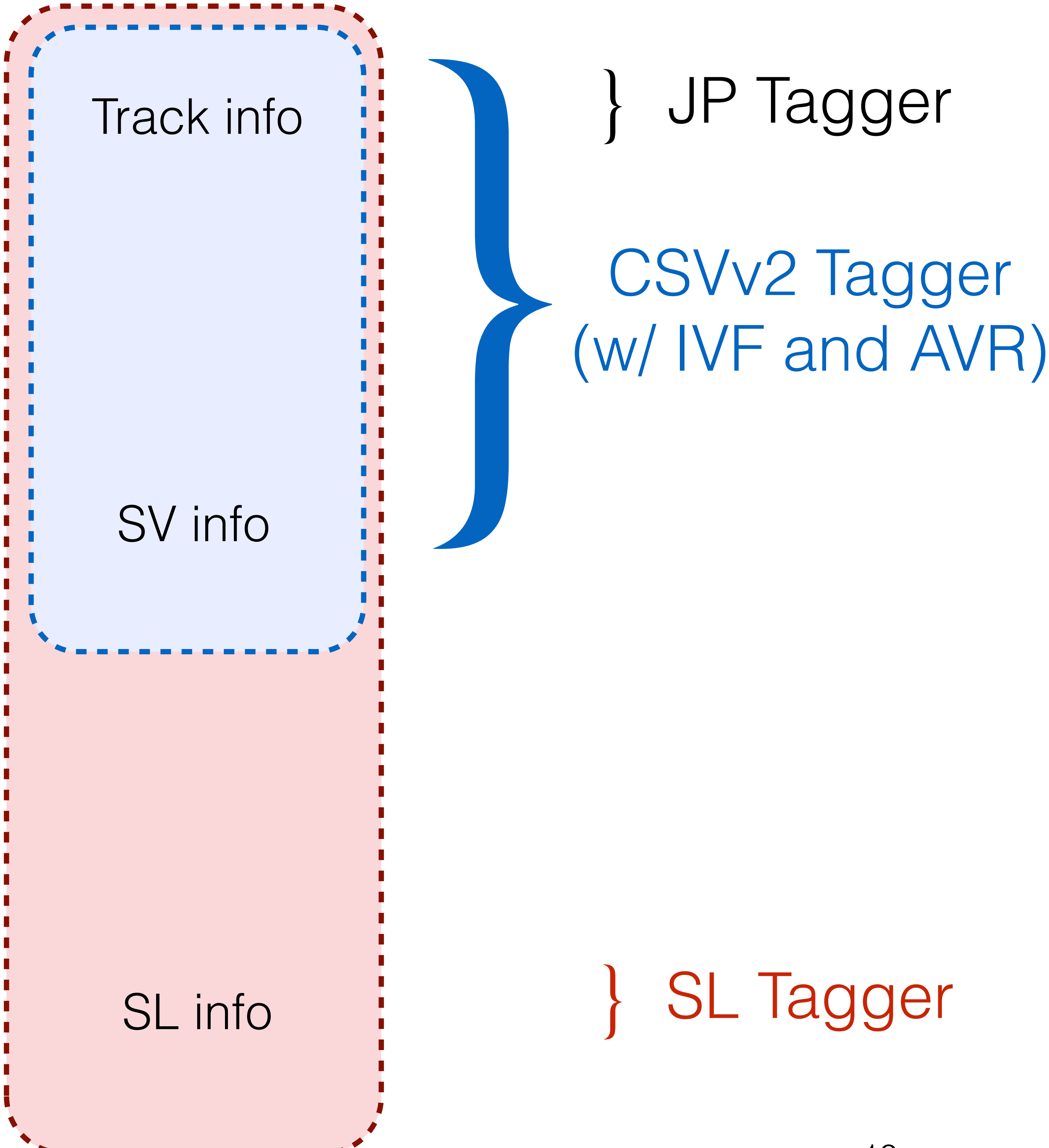
HF tagging @ CMS — BTV-15-001



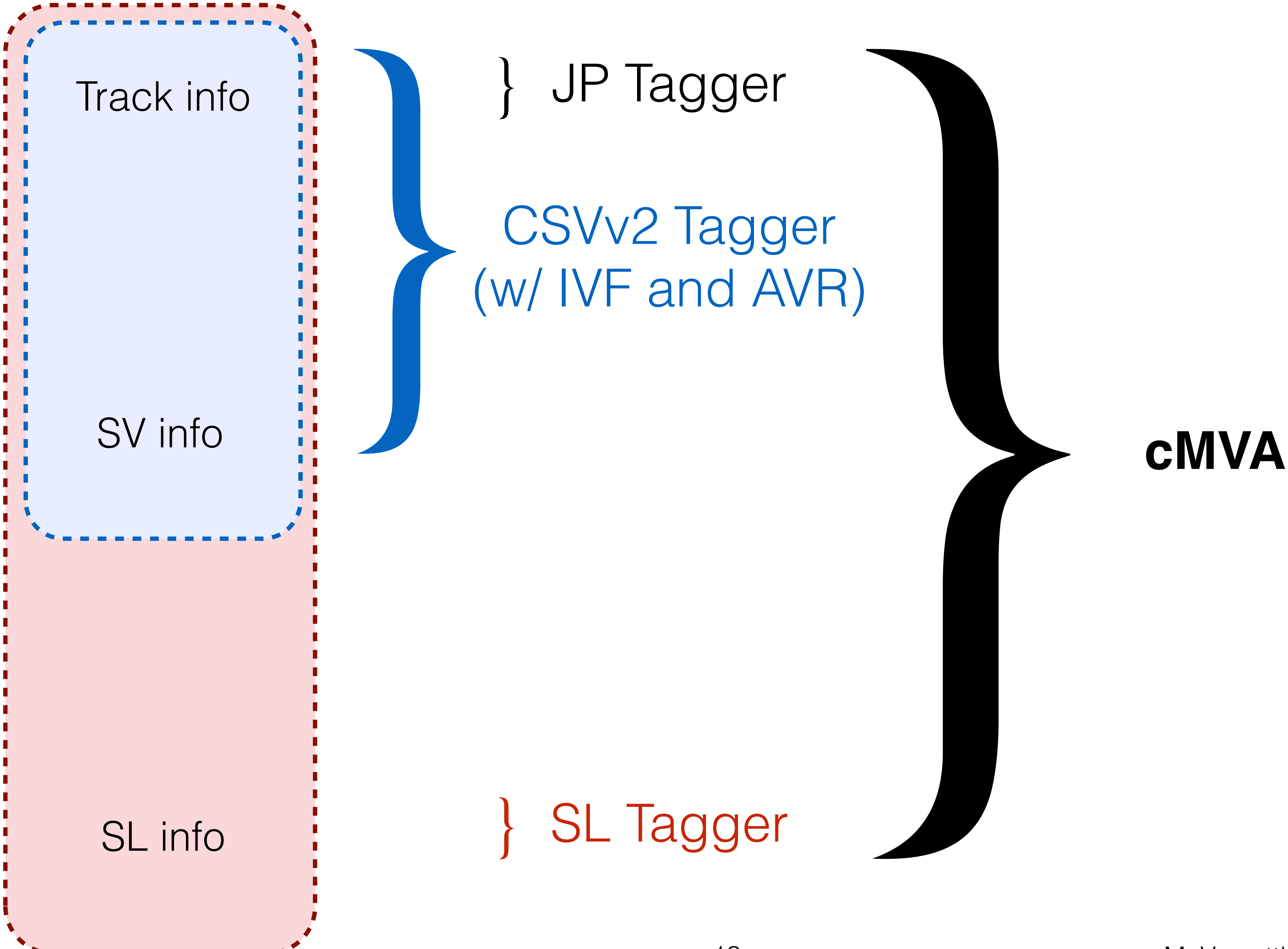
cMVA — BTV-15-001



cMVA — BTV-15-001



cMVA — BTV-15-001



c-tagger — BTV-16-001

Track info

SV info

SL info

c-tagger — BTV-16-001

Track info

SV info

SL info

c-Tagger

All the input features directly fed into the discriminator training.

Minimal information loss

c-tagger — BTV-16-001

Track info

SV info

SL info

c-Tagger

light jets

c jets

b jets

c-tagger — BTV-16-001

Track info

SV info

SL info

Fast
Reliable
Many implementations
Binary*

c-Tagger
- **light jets**
- **c jets**
- **b jets**

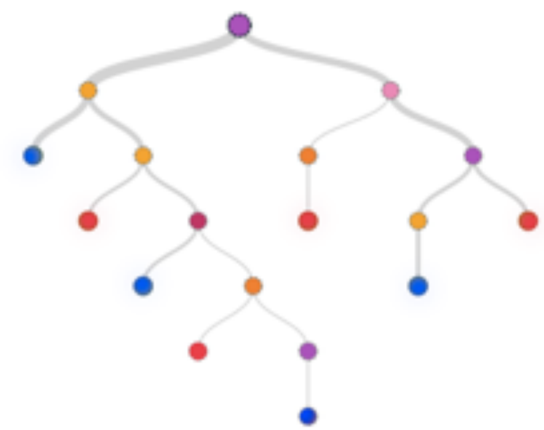
We need separation against two kind of backgrounds!

c-tagger — BTV-16-001

Track info

SV info

SL info



Fast
Reliable
Many implementations
Binary*

c-Tagger

- **light jets**
- **c jets**
- **b jets**

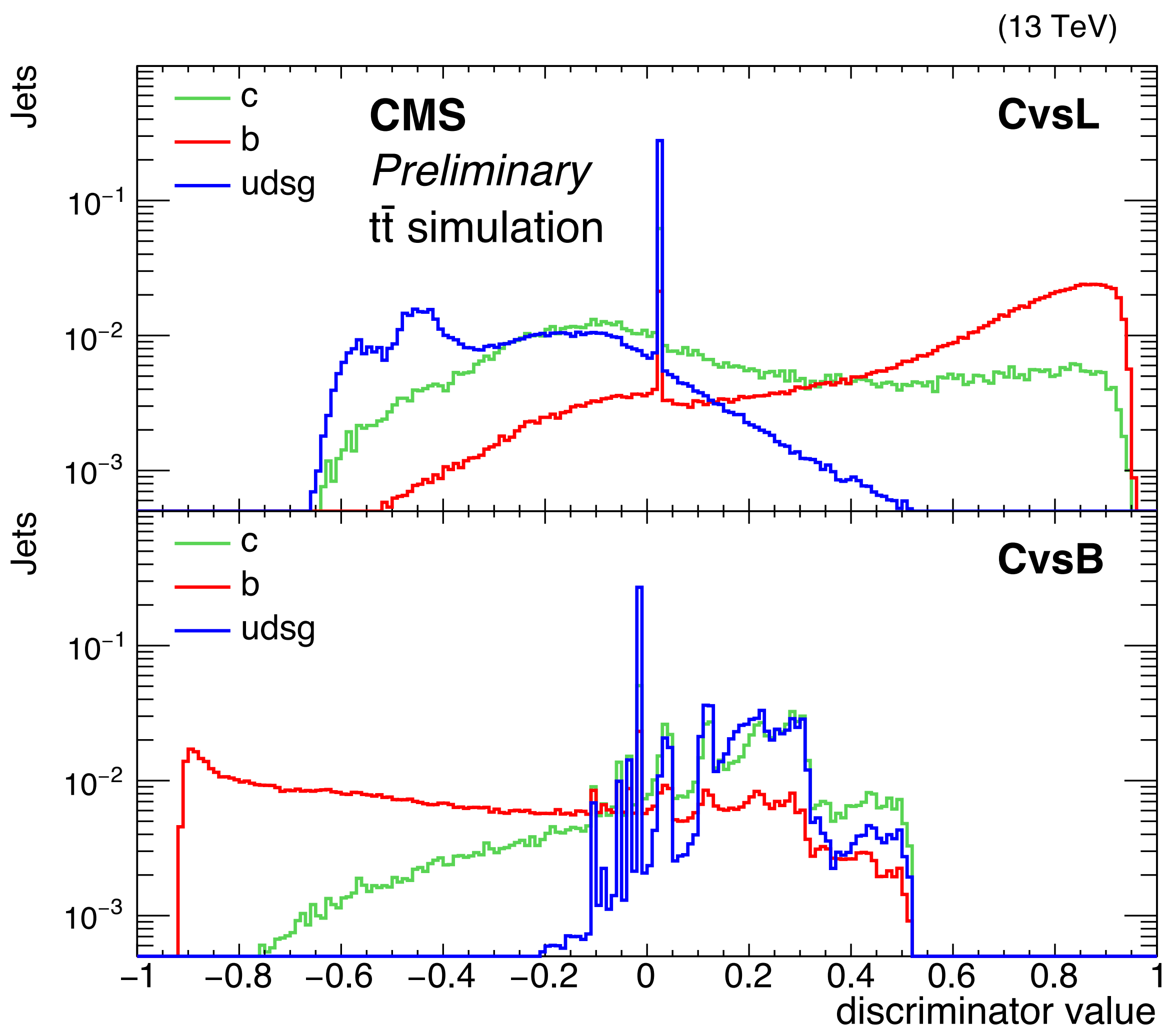
CvsL (light jets vs c jets)

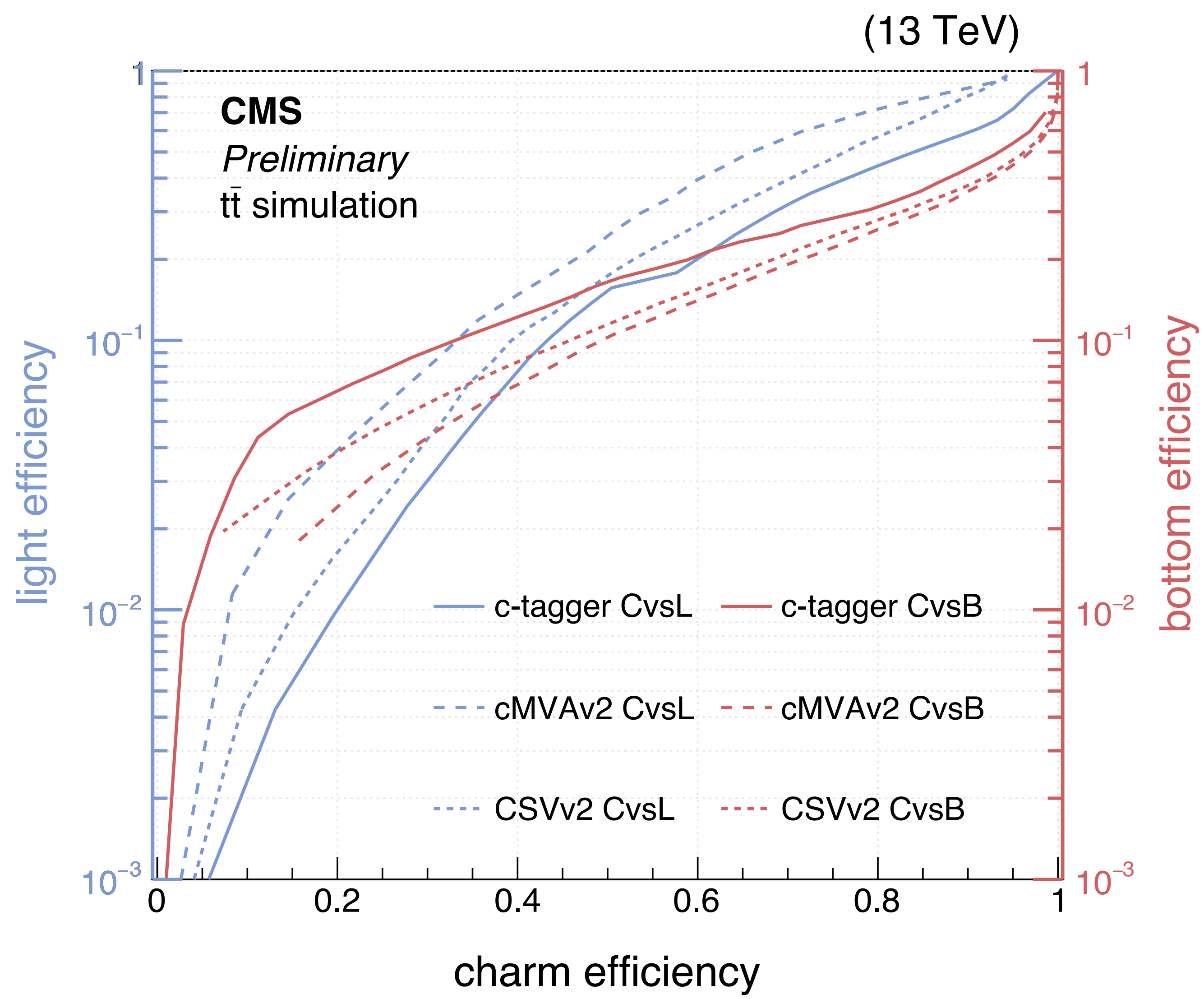
CvsB (c jets vs b jets)

We need separation against two kind of backgrounds!

Solution: use two BDTs

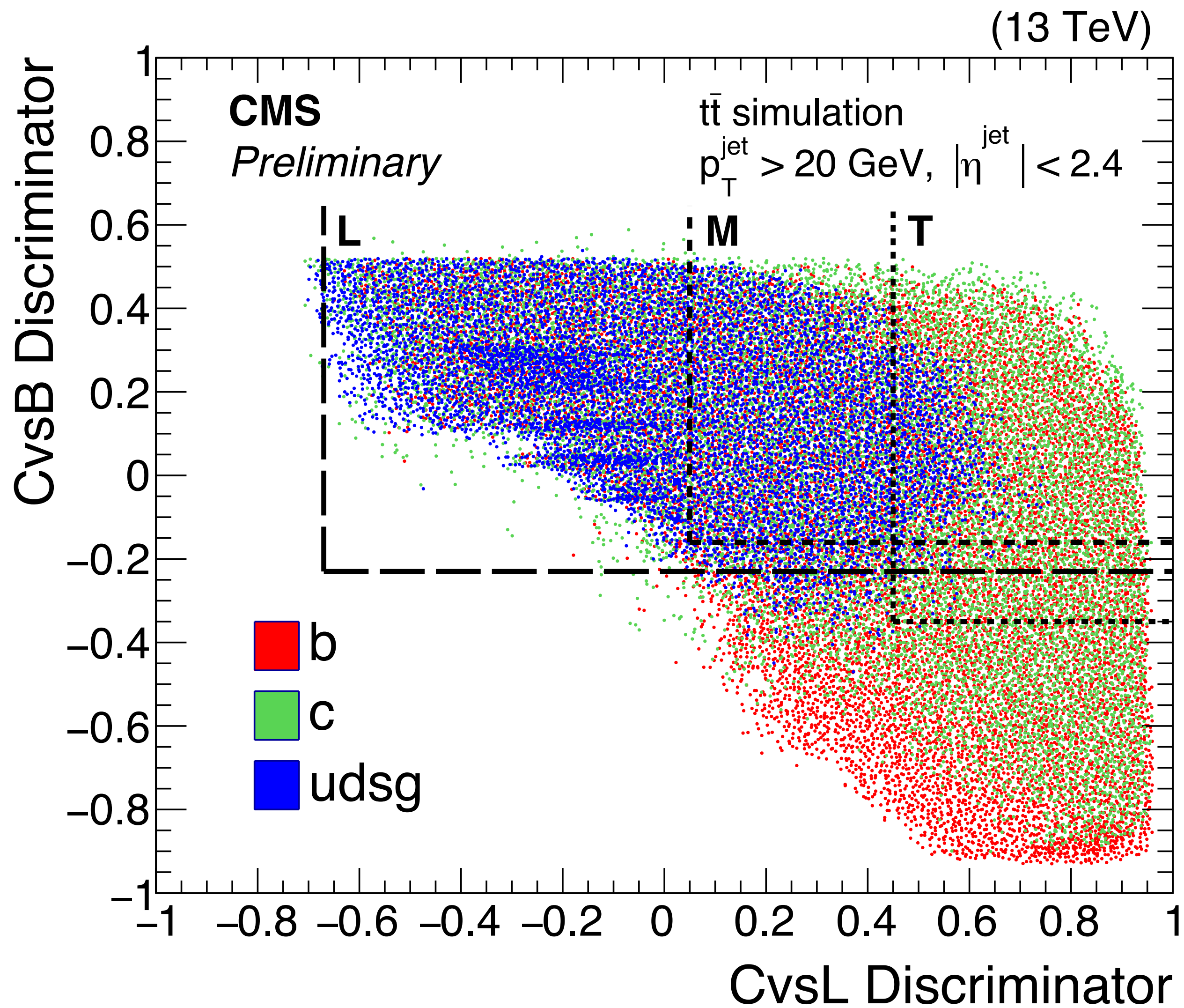
Use TWO BDTs — BTV-16-001

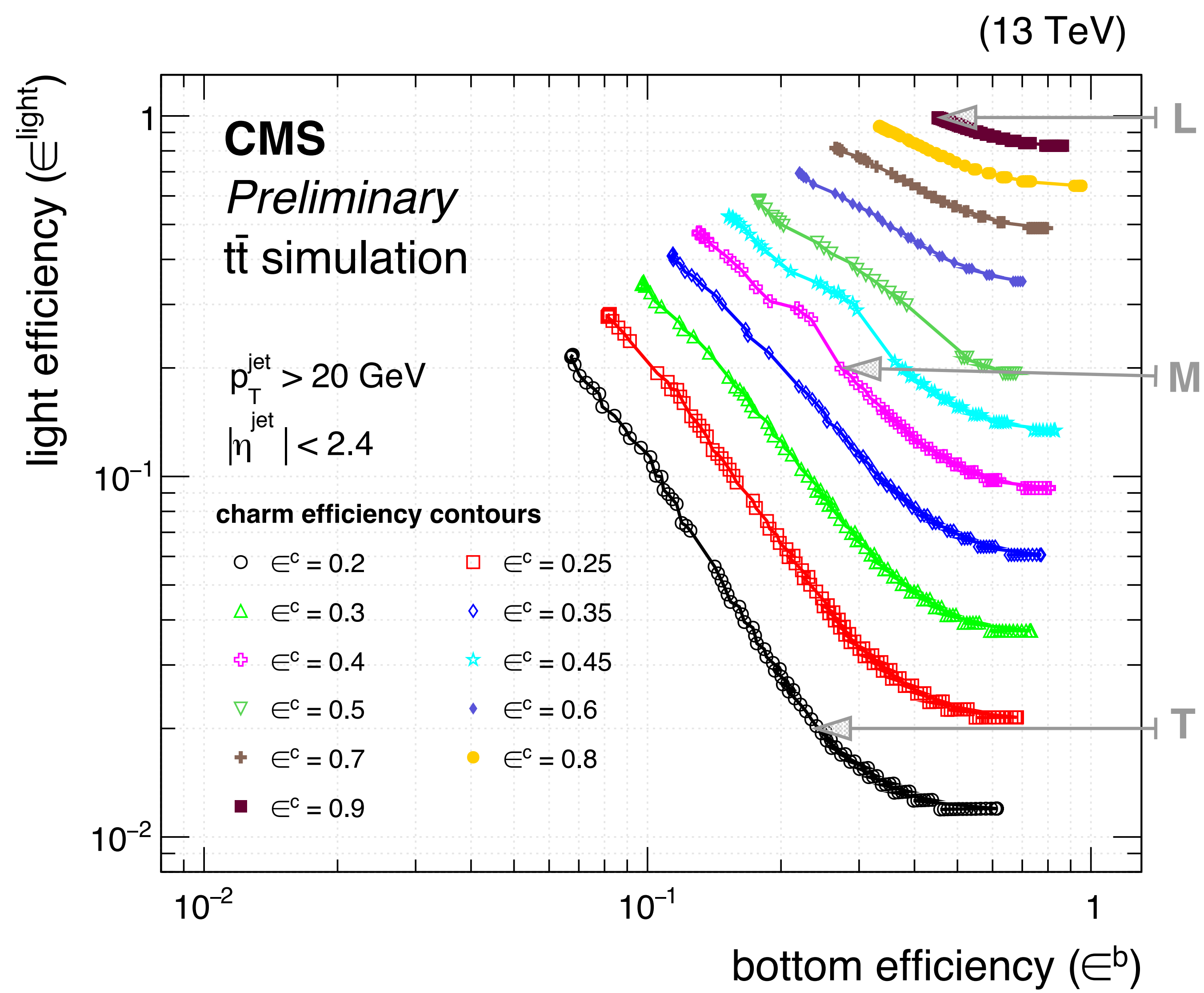




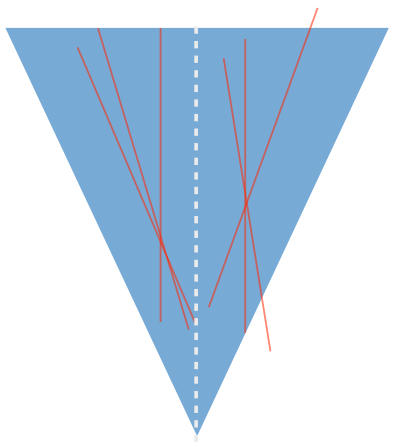
M. Verzetti, U. Rochester

Use TWO BDTs — BTV-16-001



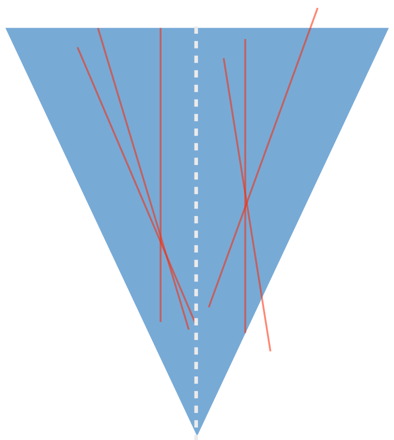


Boosted tagging @ CMS — BTV-13-001

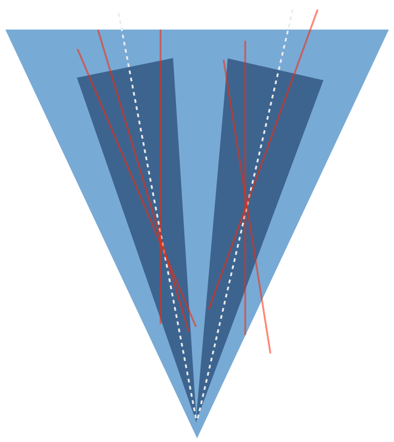


FatJet: CSVv2 w/o retraining. Custom (relaxed) track and SV association directly on anti-k_T 0.8

Boosted tagging @ CMS — BTV-13-001



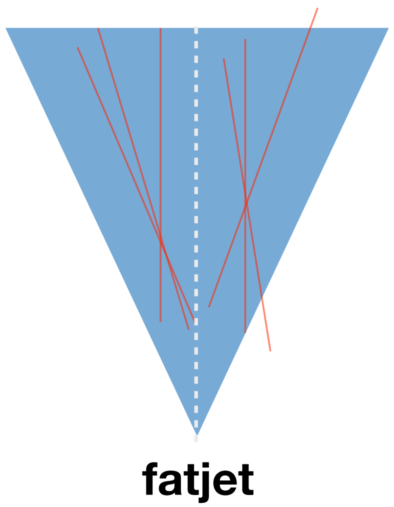
FatJet: CSVv2 w/o retraining. Custom (relaxed) track and SV association directly on anti-k_T 0.8



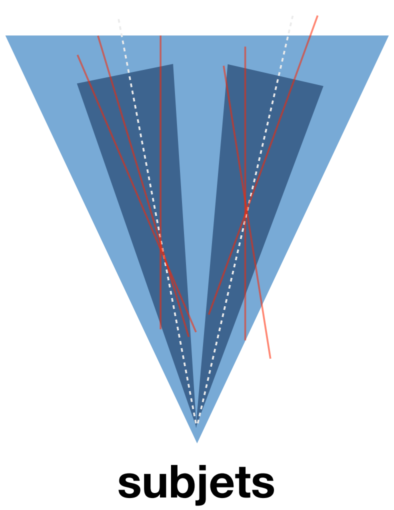
Sub-jet: CSVv2 w/o retraining applied to sub-jets (soft drop, pruned, etc…)

M. Verzetti, U. Rochester

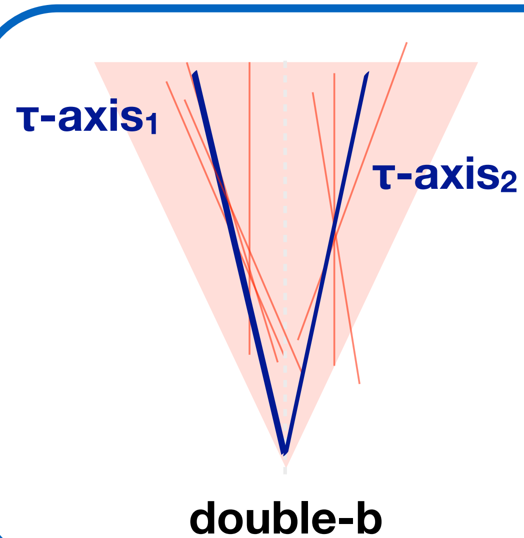
Boosted tagging @ CMS — BTV-15-002



FatJet: CSVv2 w/o retraining. Custom (relaxed) track and SV association directly on anti-k_T 0.8



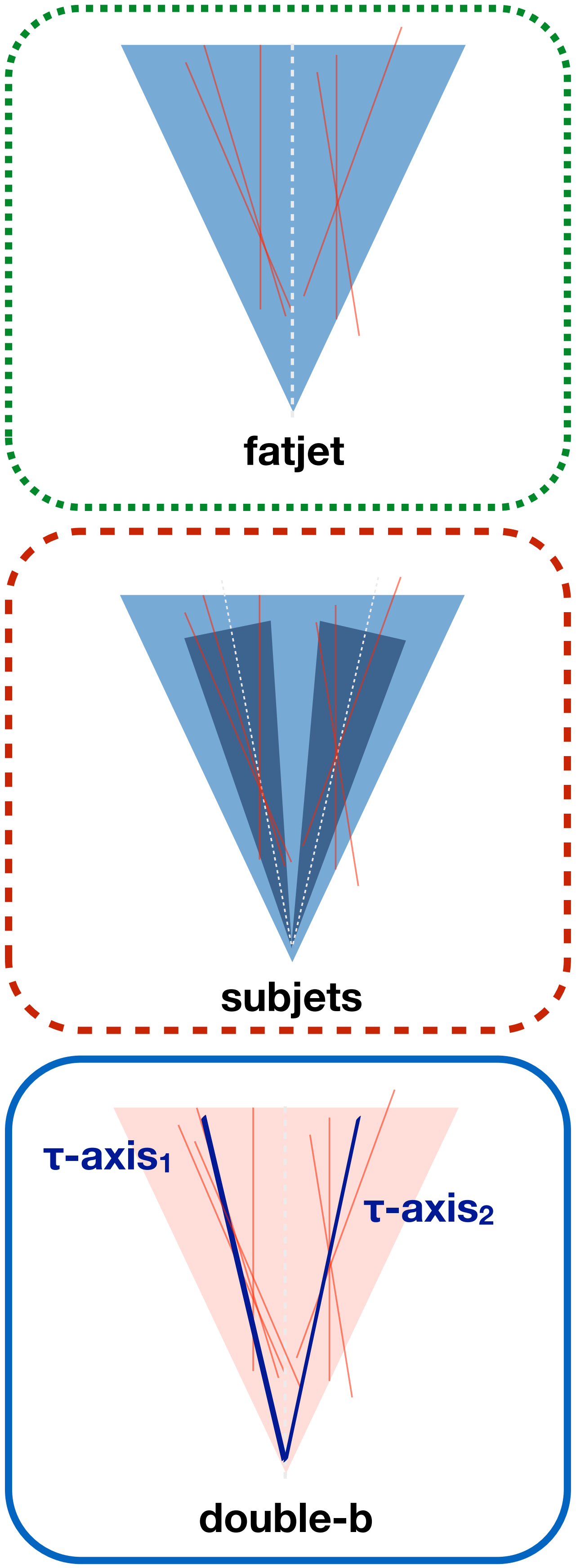
Sub-jet: CSVv2 w/o retraining applied to sub-jets (soft drop, pruned, etc…). Used for boosted top

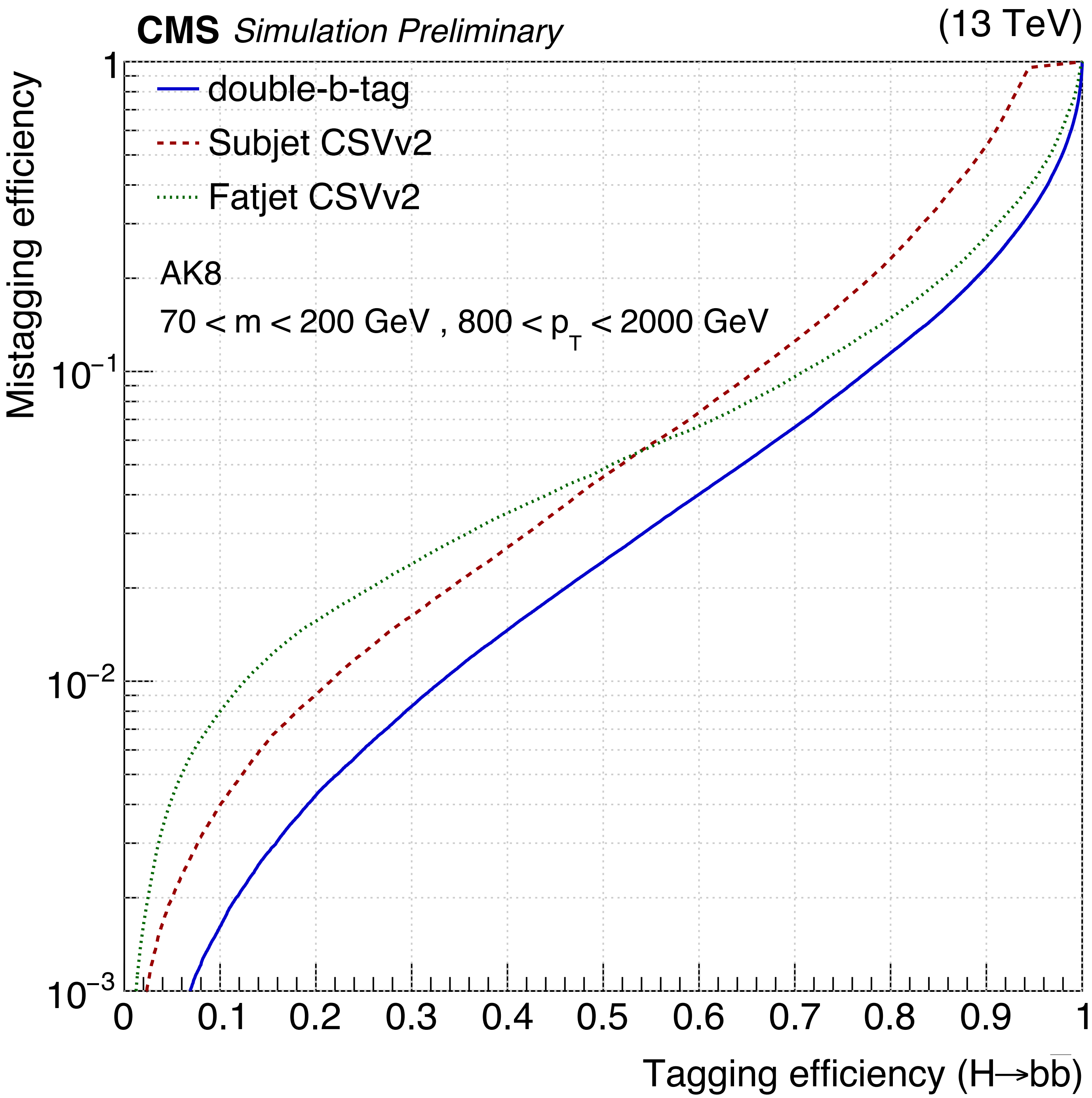


Double b: dedicated training targeting boosted resonances X→bb

M. Verzetti, U. Rochester

Boosted tagging @ CMS — BTV-15-002





M. Verzetti, U. Rochester

Why to measure performance on data?

Simulation is not perfect!

Why to measure performance on data?

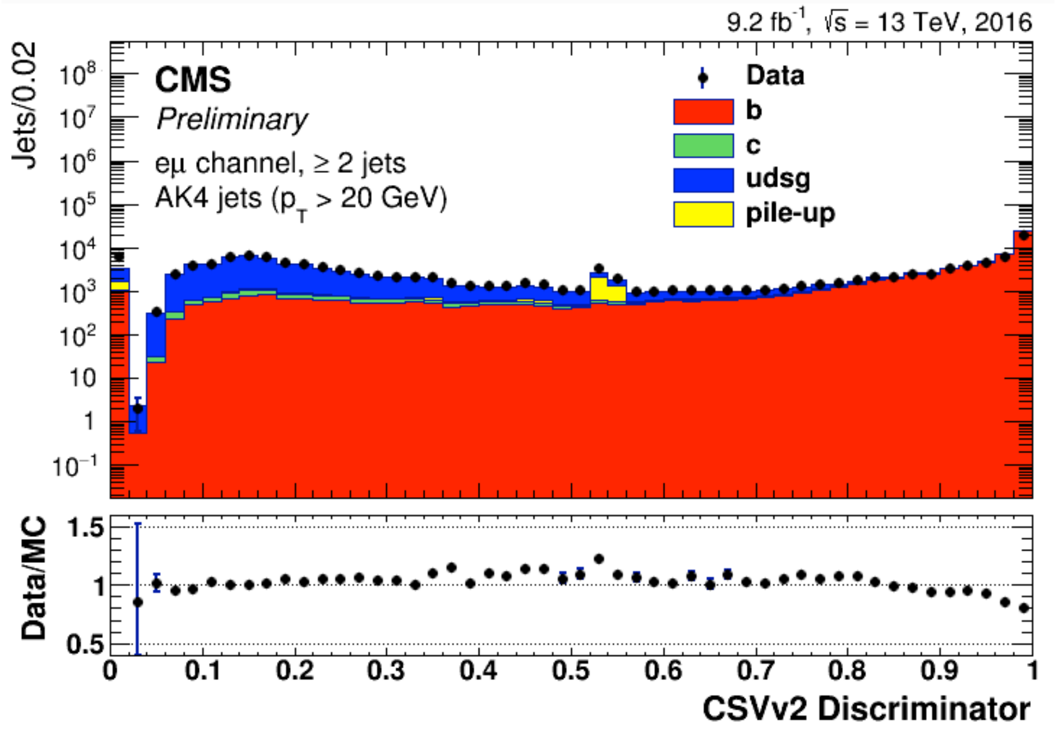
Simulation is not perfect!

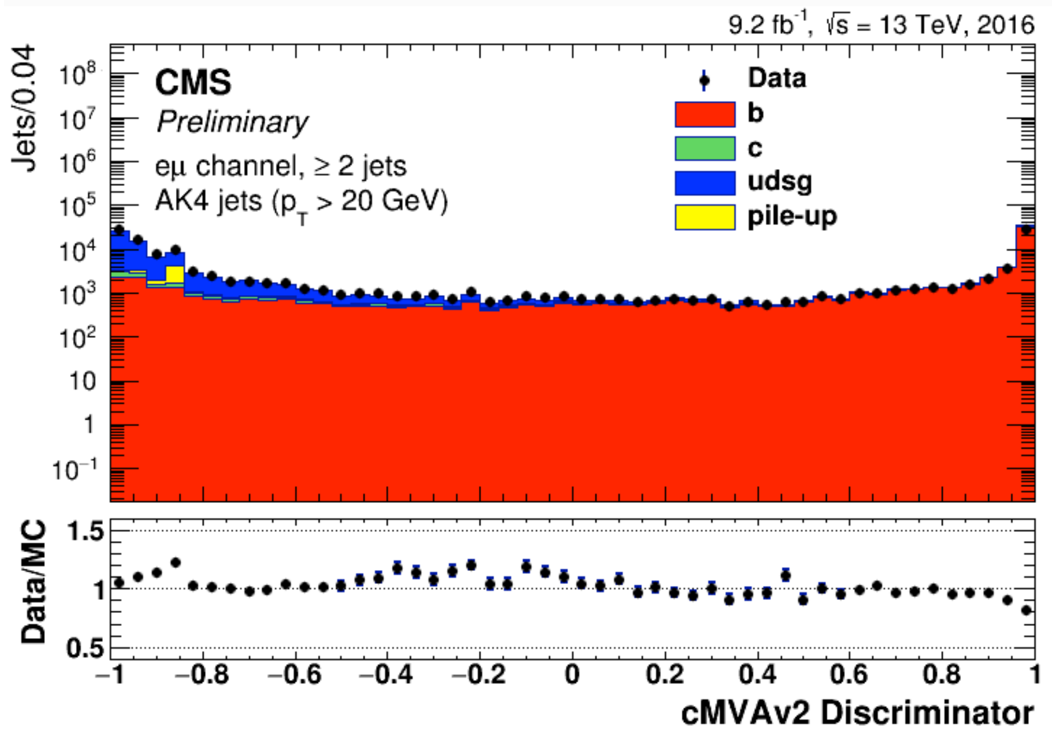
We need to correct simulation by introducing scale factors

$$SF_f = \frac{\varepsilon_{DATA}}{\varepsilon_{MC}}$$

performance on data — CMS-DP-2016-018

Extensive performance programme based on QCD multi-jet and tt events.



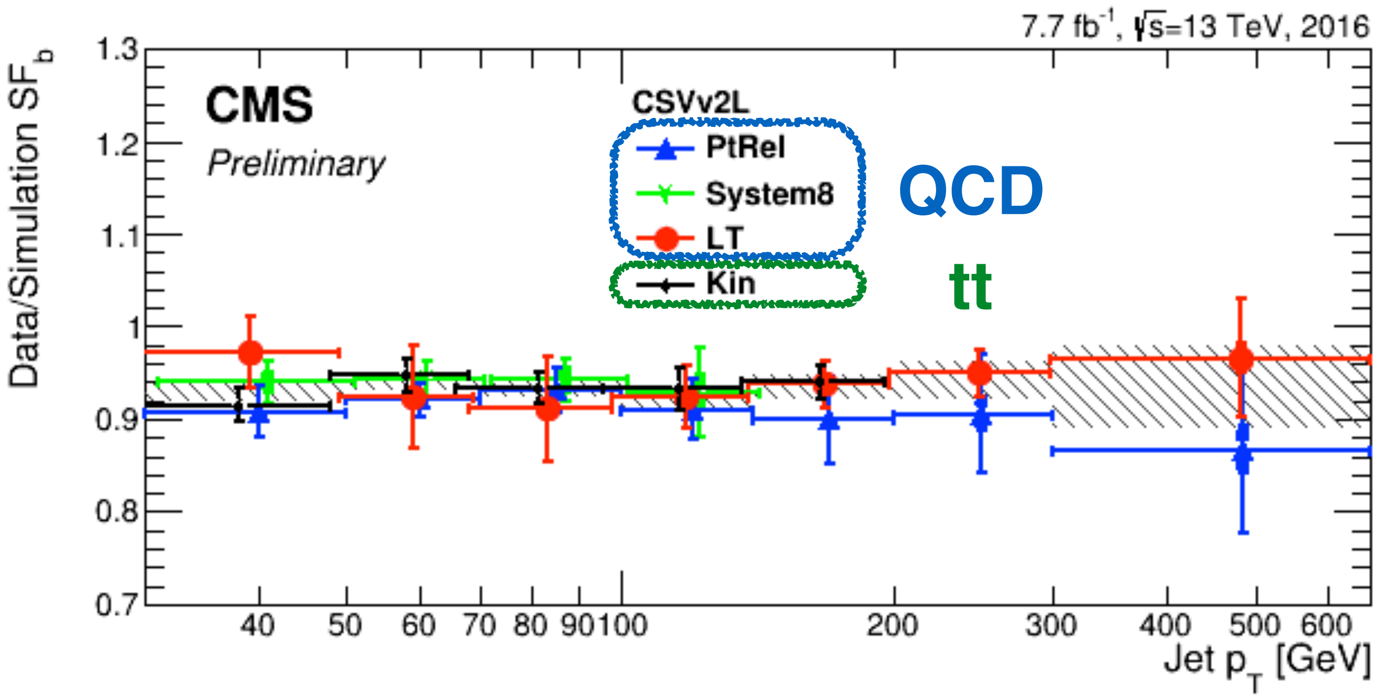


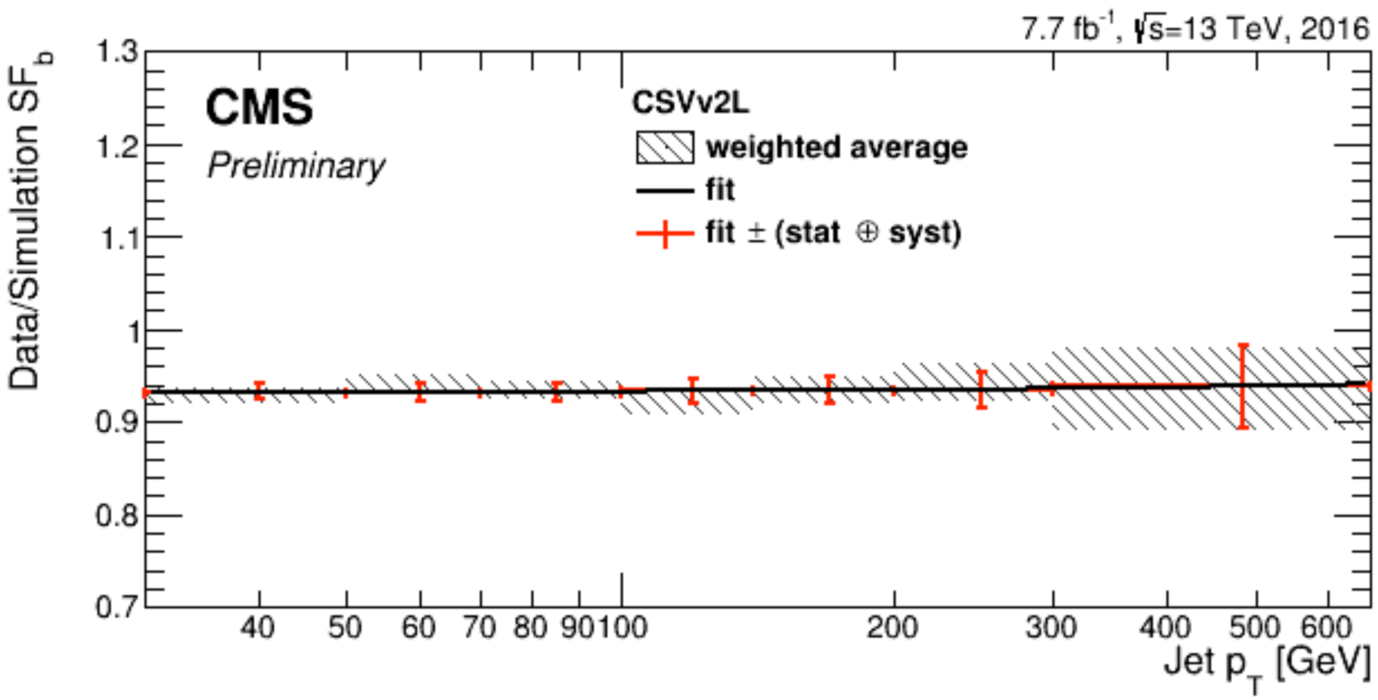
performance on data — CMS-DP-2016-018

Extensive performance programme based on:

— QCD multi-jet

— tt events.



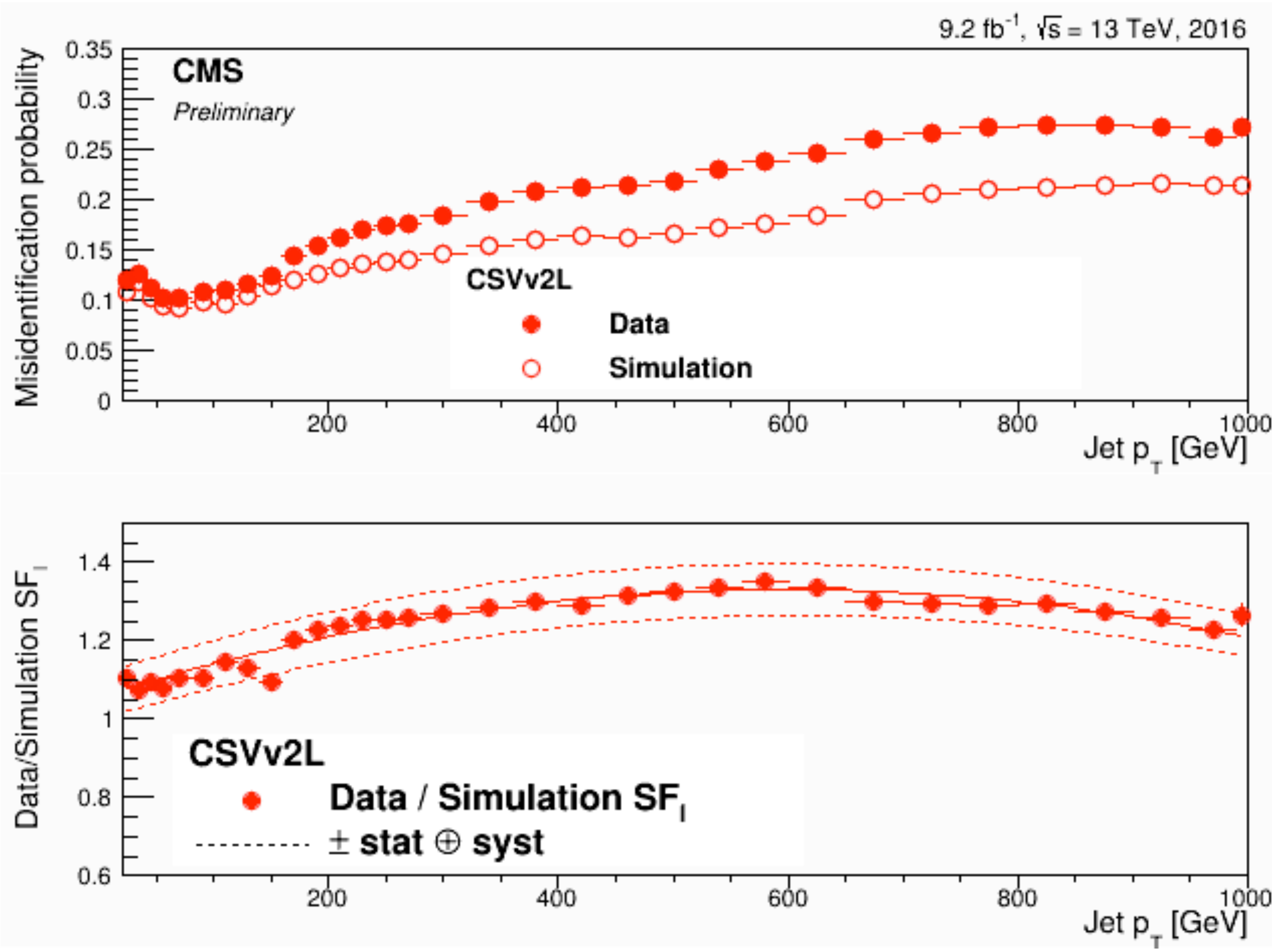


performance on data — CMS-DP-2016-018

Extensive performance programme based on:

— QCD multi-jet

— tt events.



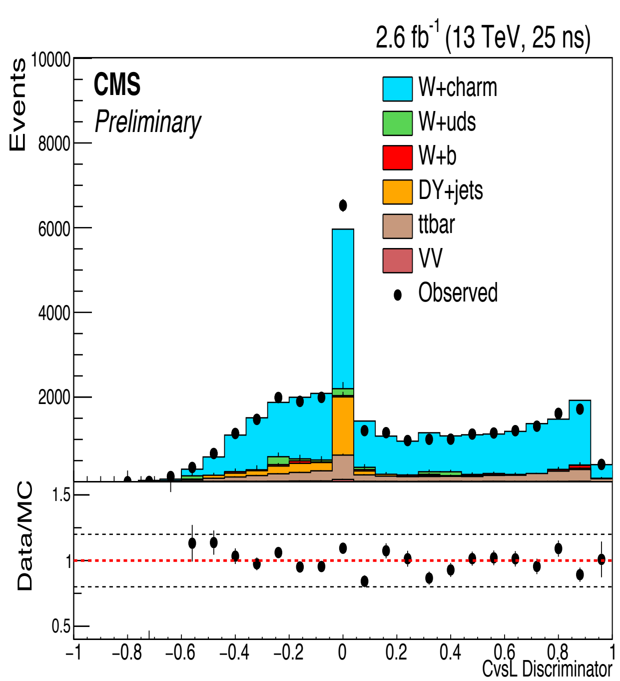
M. Verzetti, U. Rochester

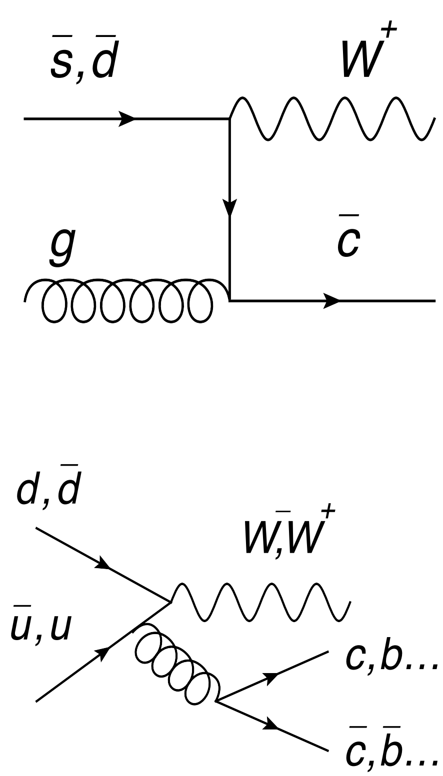
A harder challenge: charm tagging performance — <u>BTV-16-001</u>

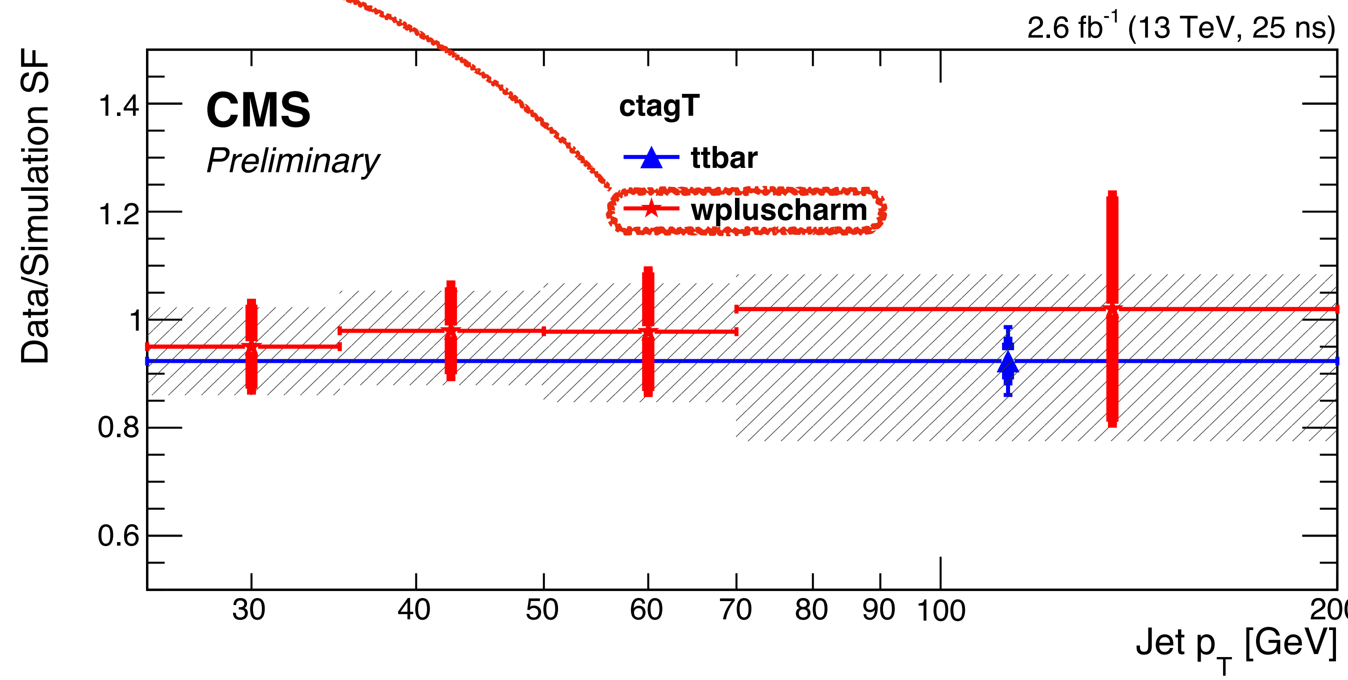
W+c events require a μ in the jet.

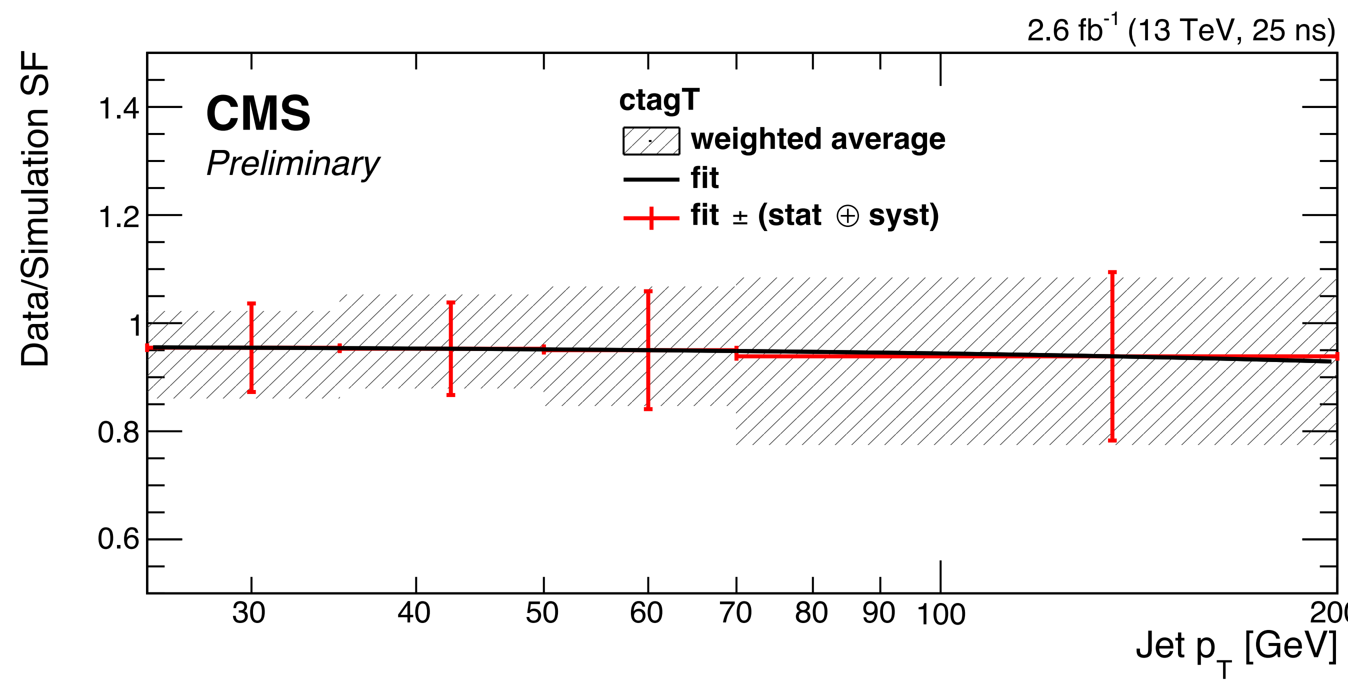
Background removed by OS - SS

High purity sample. Possible p_T(jet) binning









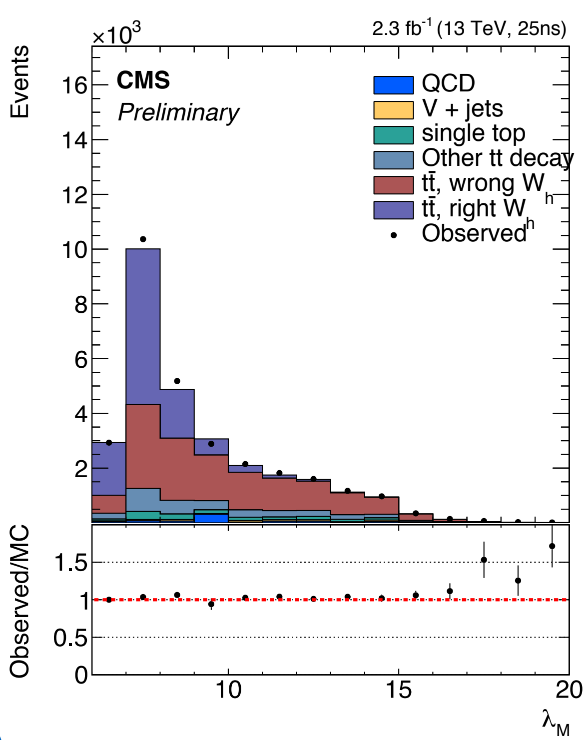
M. Verzetti, U. Rochester

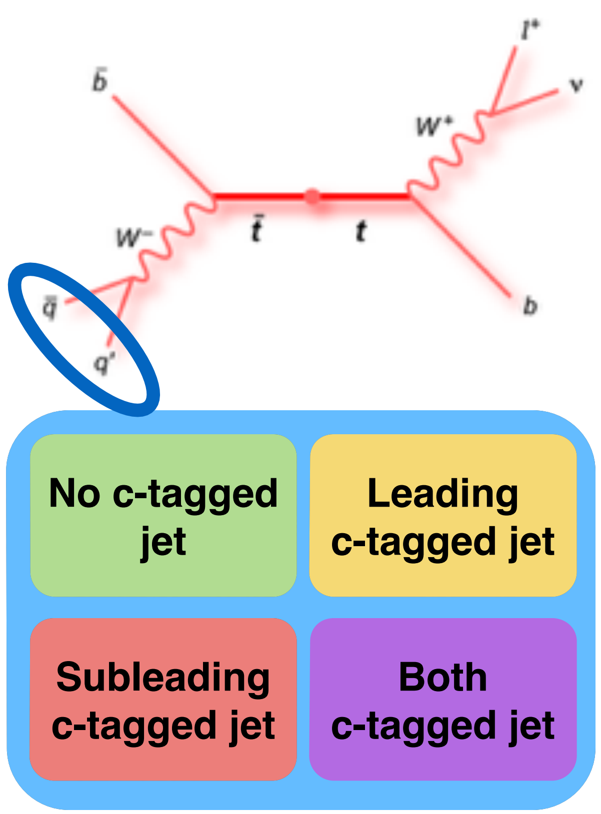
A harder challenge: charm tagging performance — BTV-16-001

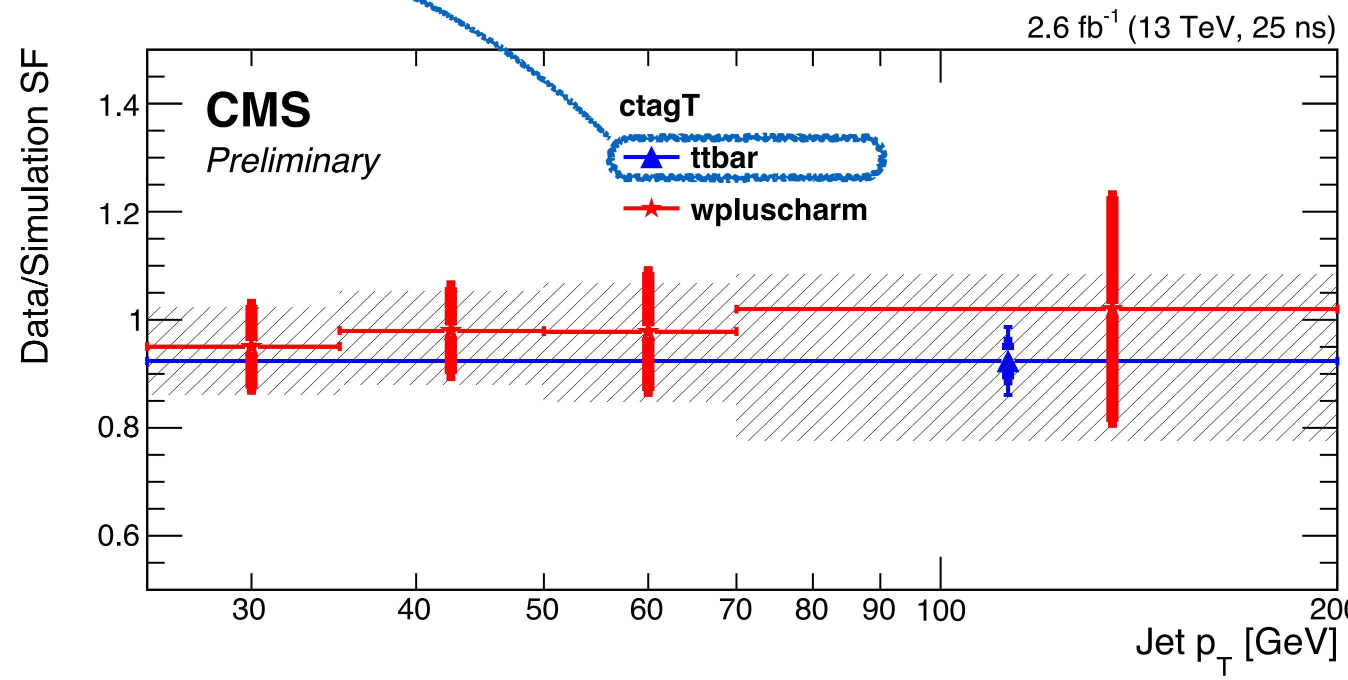
Semi-leptonic tt events contain 25% of c-jets in the W decay.

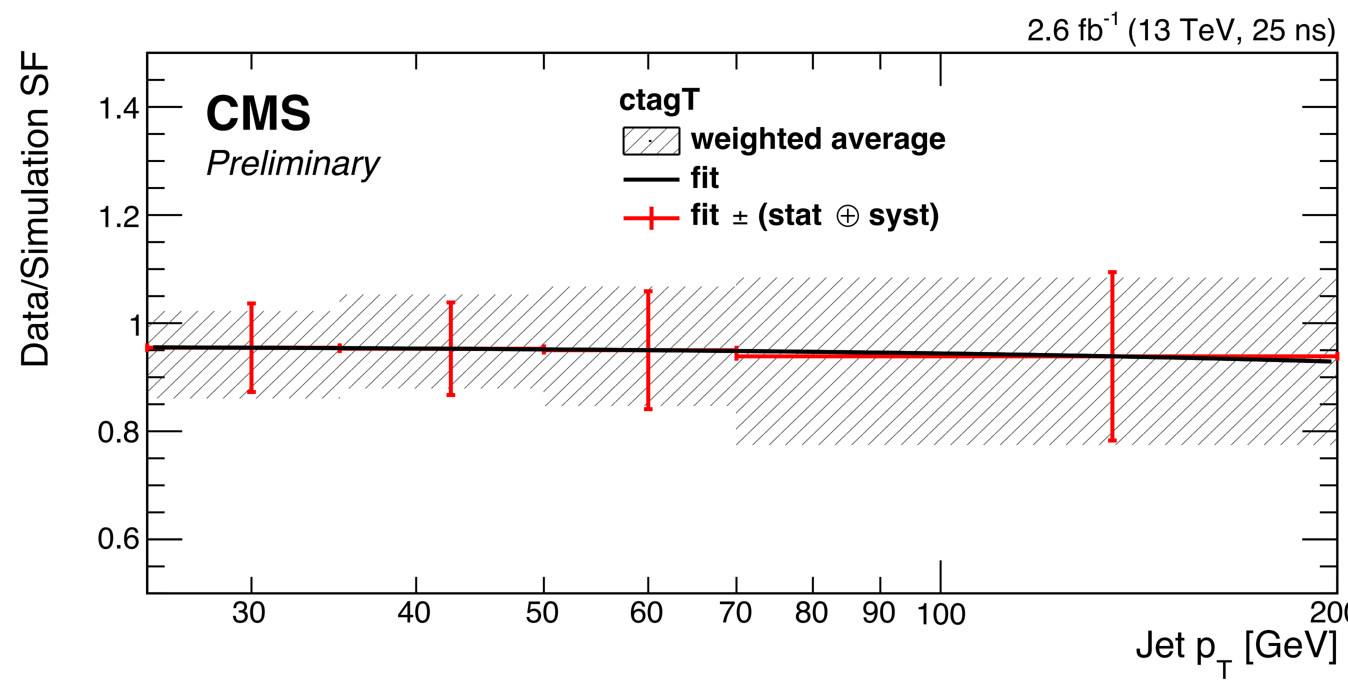
Two jet samples (leading, subleading) enriched/depleted with c-jets by double weak decay

Simultaneous template fit in four categories to extract the SF value





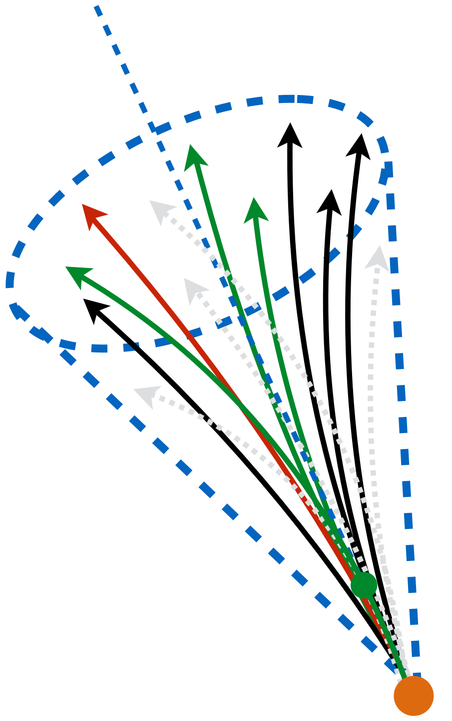




M. Verzetti, U. Rochester

Take home messages

- CMS has a wide range of heavy flavour taggers
- Calibration is performed on data with multiple, complementary methods
- New charm tagger tool
  - New methods to calibrate on c jets

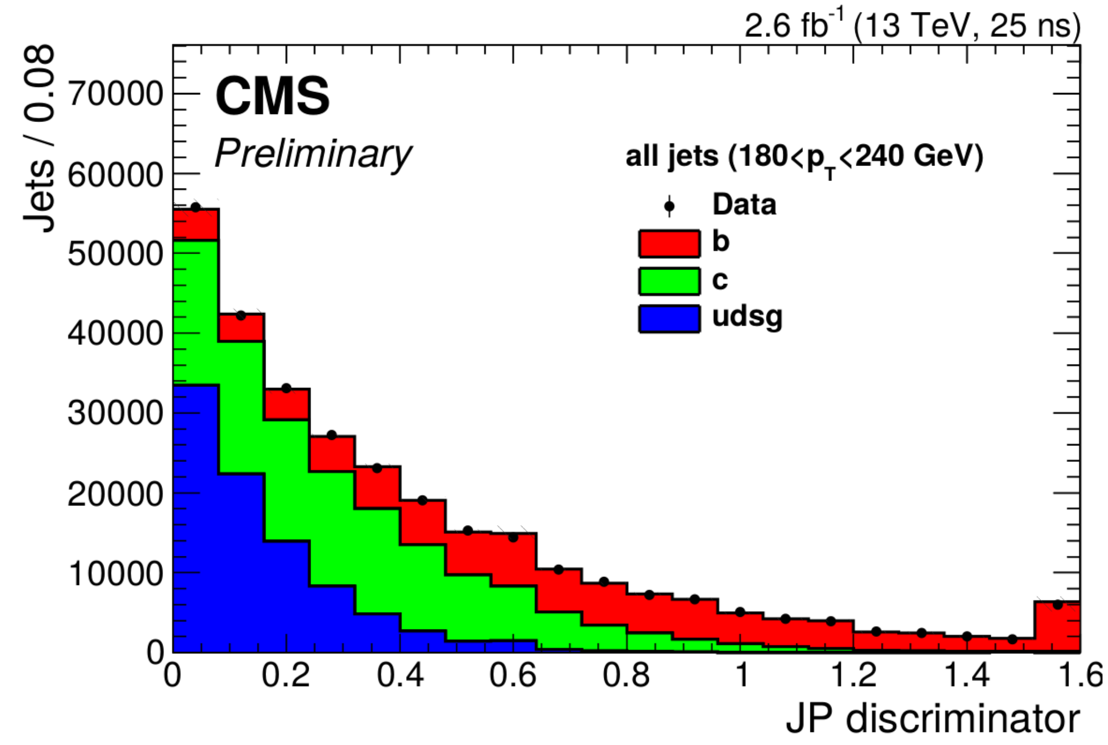


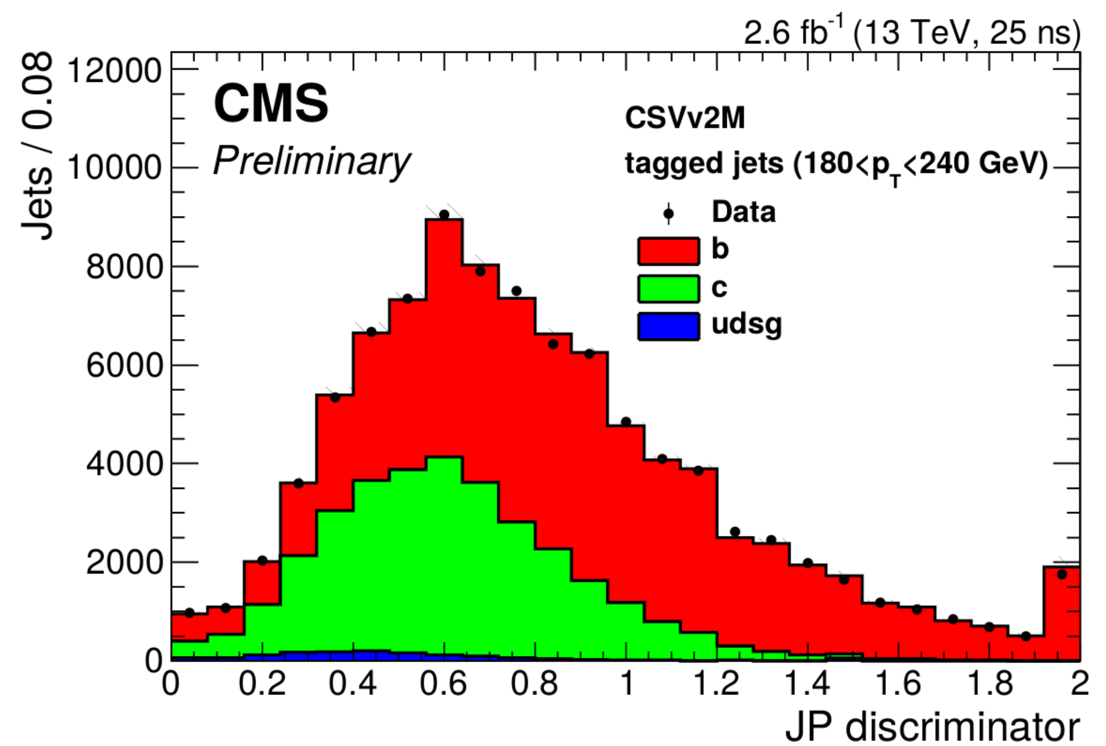
Back up

LT Method

- Requires muon in the jet to increase purity
- Template fit of the lifetime tagger (JP discriminator) distribution before and after the tag

$$\varepsilon_b = \frac{C_b f_b^{tag} N_{data}^{tag}}{f_b^{before\ tag} N_{data}^{before\ tag}}$$

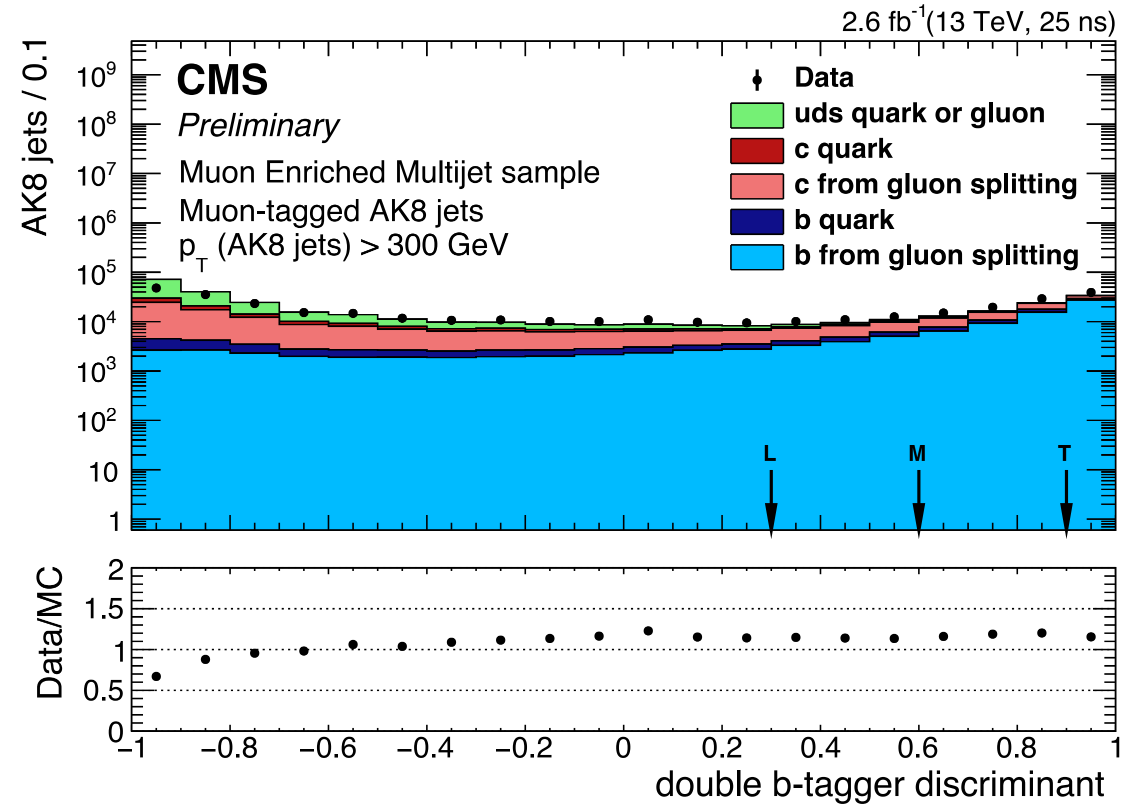


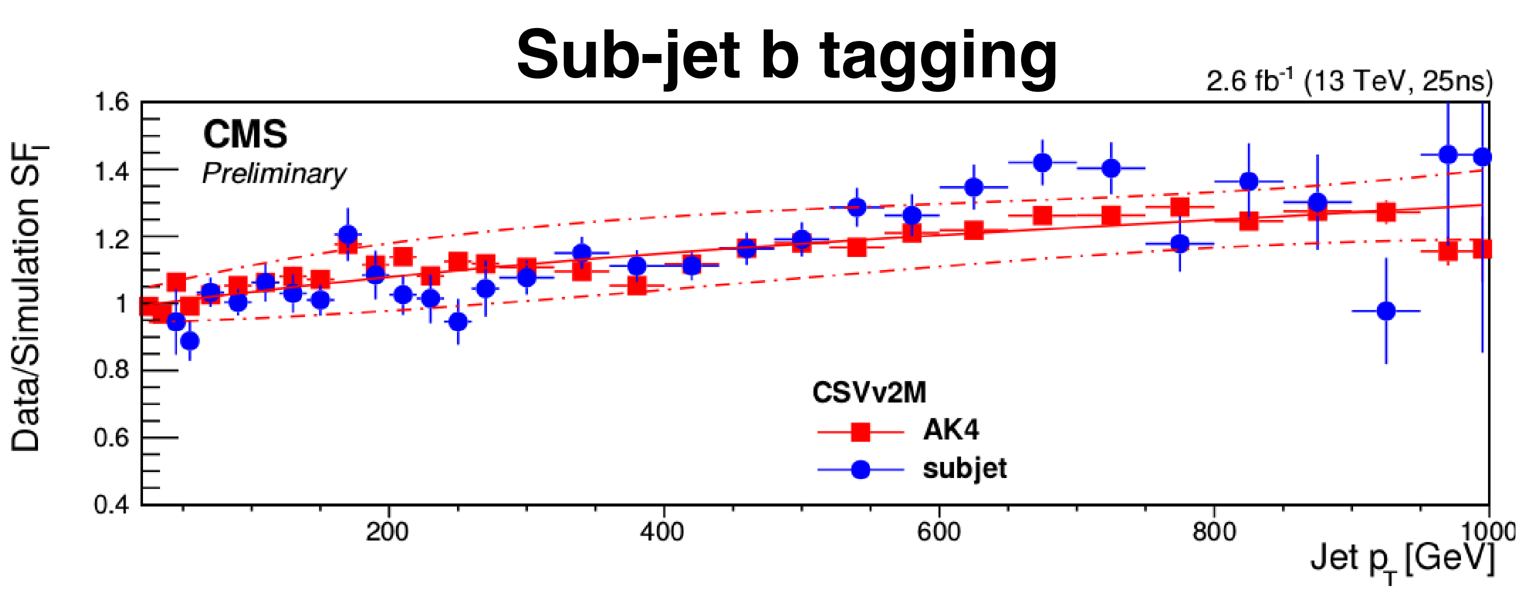


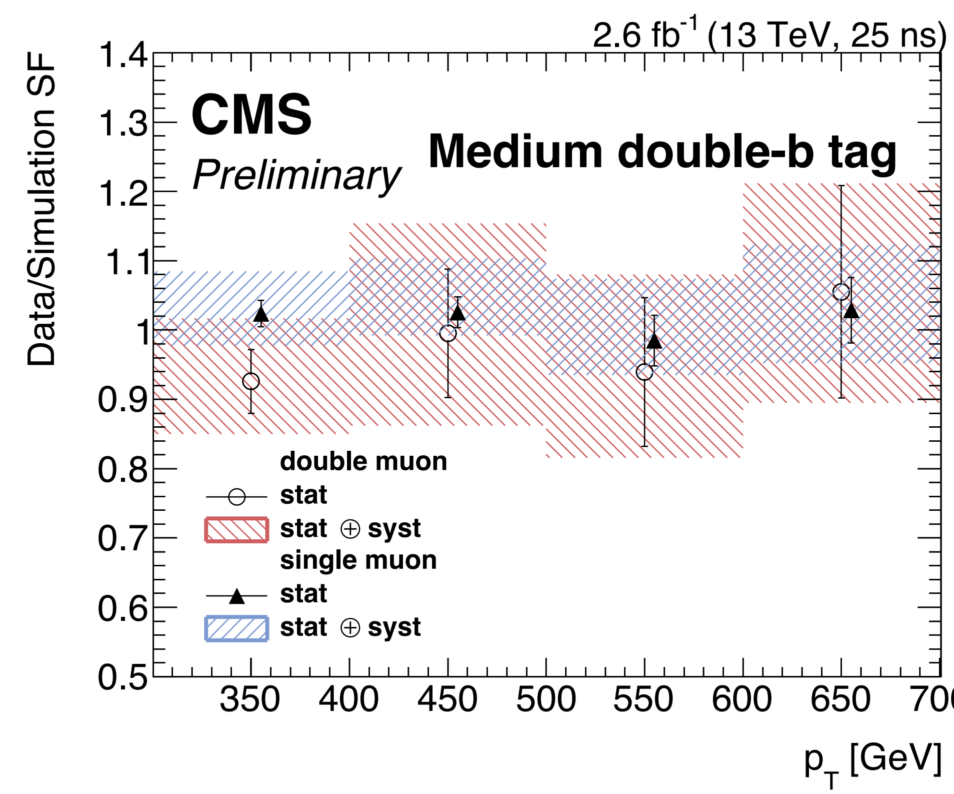
Boosted b tagging performances

- Efficiency measured with the LT method

- Mis-tag rate measured with the Negative tags method



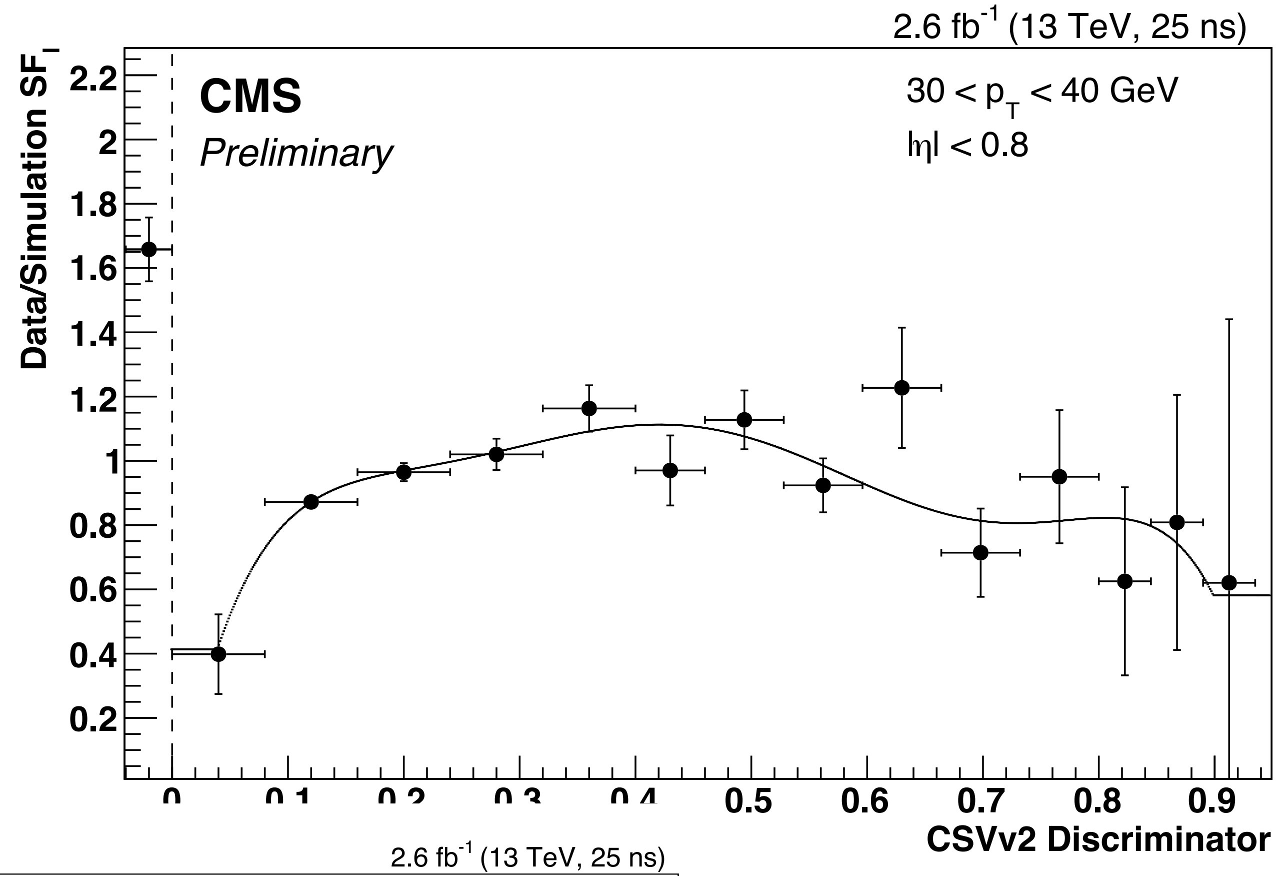


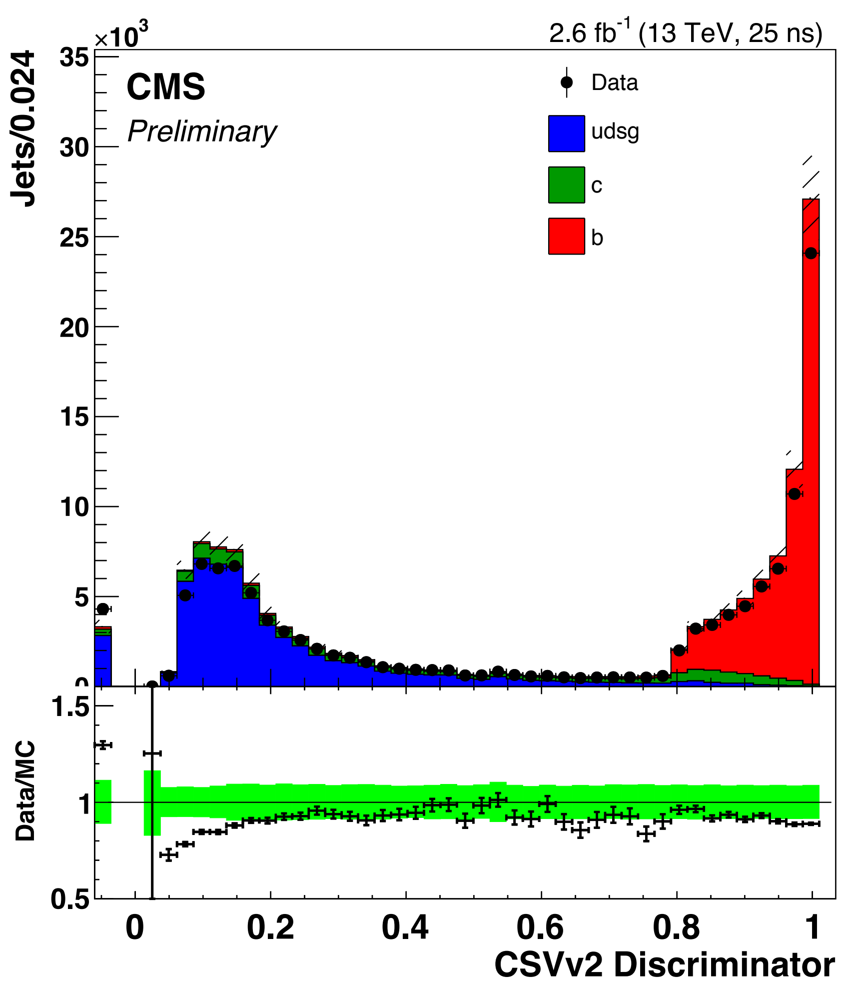


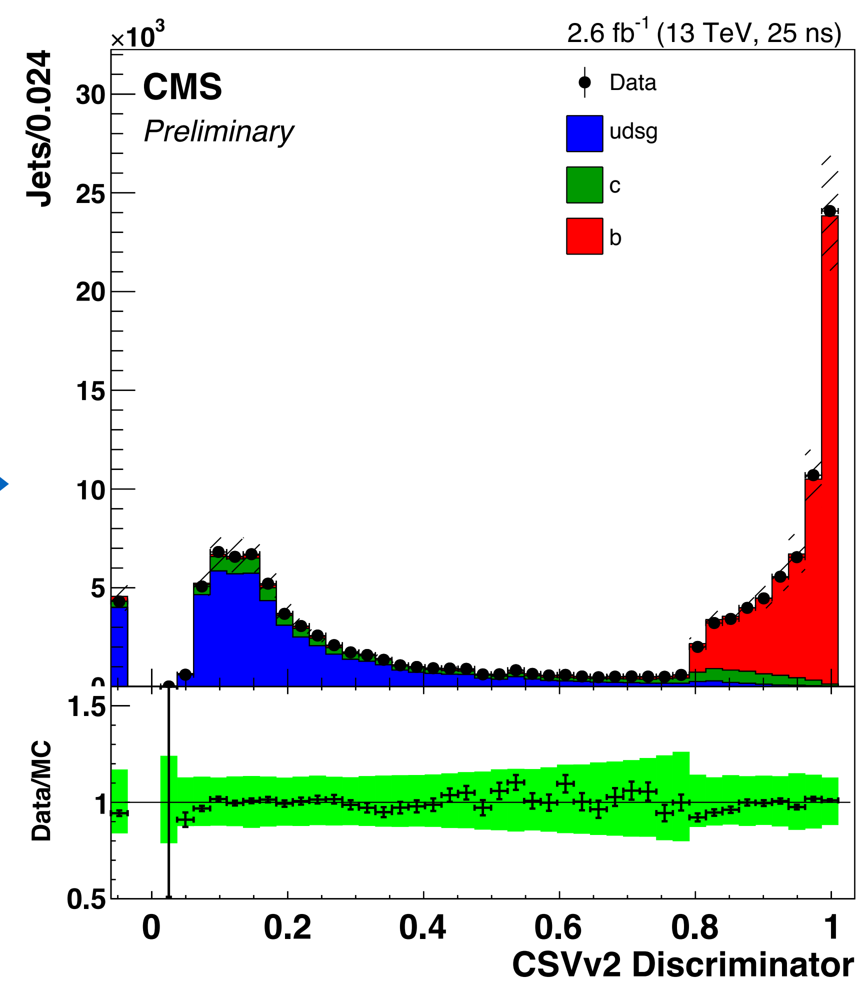
M. Verzetti, U. Rochester

Discriminator reshaping

- Di-leptonic tt events + Z+jets events
- Iterative method to extract SF_b and SF_L





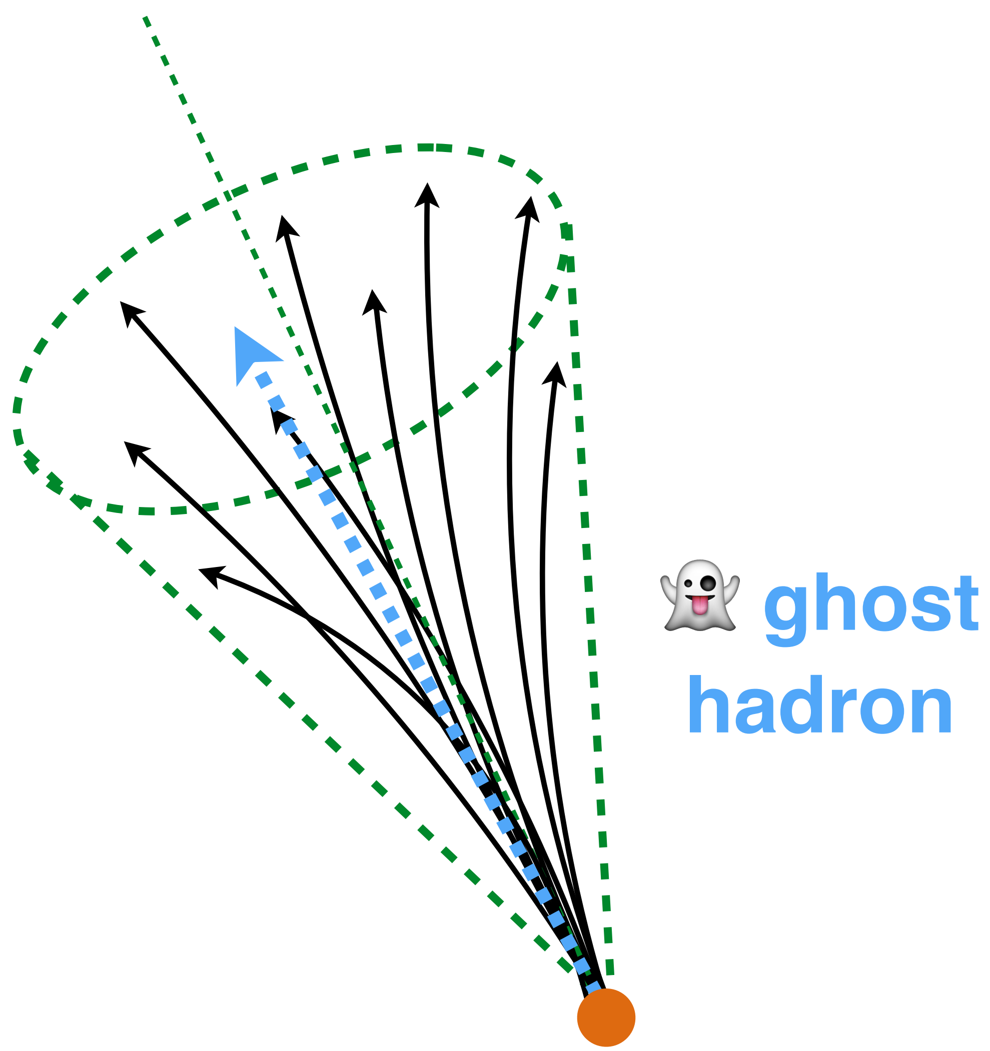


M. Verzetti, U. Rochester

Flavour Matching

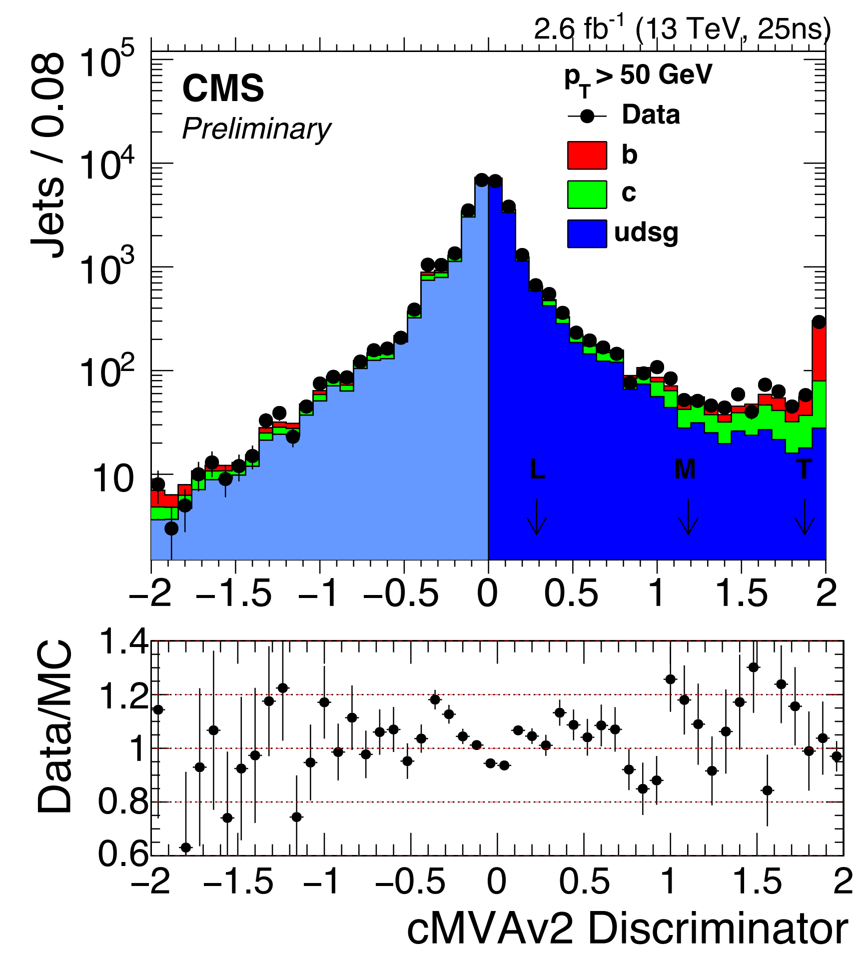
Similar definition to particle-level b-quarks in top physics:

- Generated B/C hadron w/ scaled-down momentum
- Add the hadrons to the stable particle collections
- Re-run jet clustering
- Jets containing the hadron are assigned the hadron flavour (priority to Bs over Cs)



M. Verzetti, U. Rochester

Negative tags method (l → b mistag)



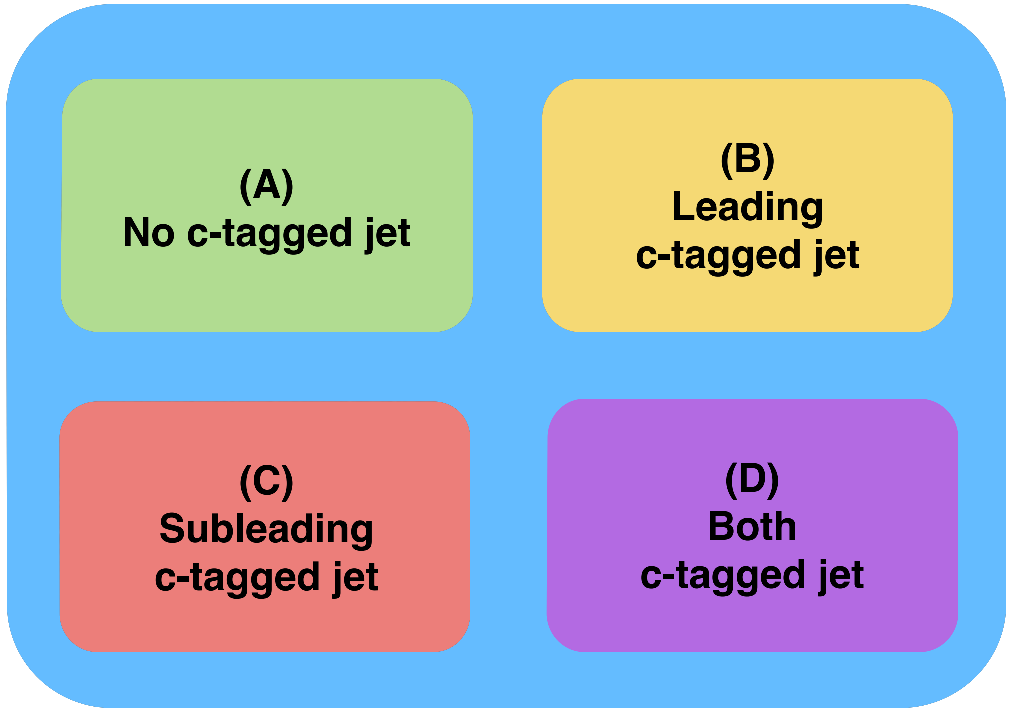
- Builds tagger using only information from tracks with positive/negative IP
- Assumption: IP for light jets is a resolution effect / due to fake tracks
- Measure SF with the negative tagger only
- Used for:
  - b-tagging
  - c-tagging
  - boosted tagging

$$\varepsilon_{\text{data}}^{\text{misid}} = \varepsilon_{\text{data}}^{-} \cdot R_{\text{light}}$$

$$R_{\text{light}} = \frac{\varepsilon_{\text{MC}}^{\text{misid}}}{\varepsilon_{\text{MC}}^{-}}$$

c tagging with tt events

- **Histogram-based template fit.**
- **1 parameter of interest**: ***SF_c***
- No SF_c binning: impossible with this statistics
- Systematic effects added and profiled in the fit
- **4 categories** according to hadronic W jets that pass B/C-tagging WP



$$
\begin{aligned}
N_A = \;& N_T((1-\varepsilon^l_c)(1-\varepsilon^s_L)f^l && +(1-\varepsilon^l_L)(1-\varepsilon^s_c)f^s && +(1-\varepsilon^l_L)(1-\varepsilon^s_L)(1-f^l-f^s))\\
N_B = \;& N_T(\varepsilon^l_c(1-\varepsilon^s_L)f^l && +\varepsilon^l_L(1-\varepsilon^s_c)f^s && +\varepsilon^l_L(1-\varepsilon^s_L)(1-f^l-f^s))\\
N_C = \;& N_T((1-\varepsilon^l_c)\varepsilon^s_L f^l && +(1-\varepsilon^l_L)\varepsilon^s_c f^s && +(1-\varepsilon^l_L)\varepsilon^s_L(1-f^l-f^s))\\
N_D = \;& N_T(\varepsilon^l_c\varepsilon^s_L f^l && +\varepsilon^l_L\varepsilon^s_c f^s && +\varepsilon^l_L\varepsilon^s_L(1-f^l-f^s))
\end{aligned}
$$

(c, L) (L, c) (L, L)

$$
\varepsilon^i_j = \mathrm{SF}_j\,\varepsilon^i_j(MC), \text{ with } i=(l,s) \text{ and } j=(c,L)
$$

M. Verzetti, U. Rochester