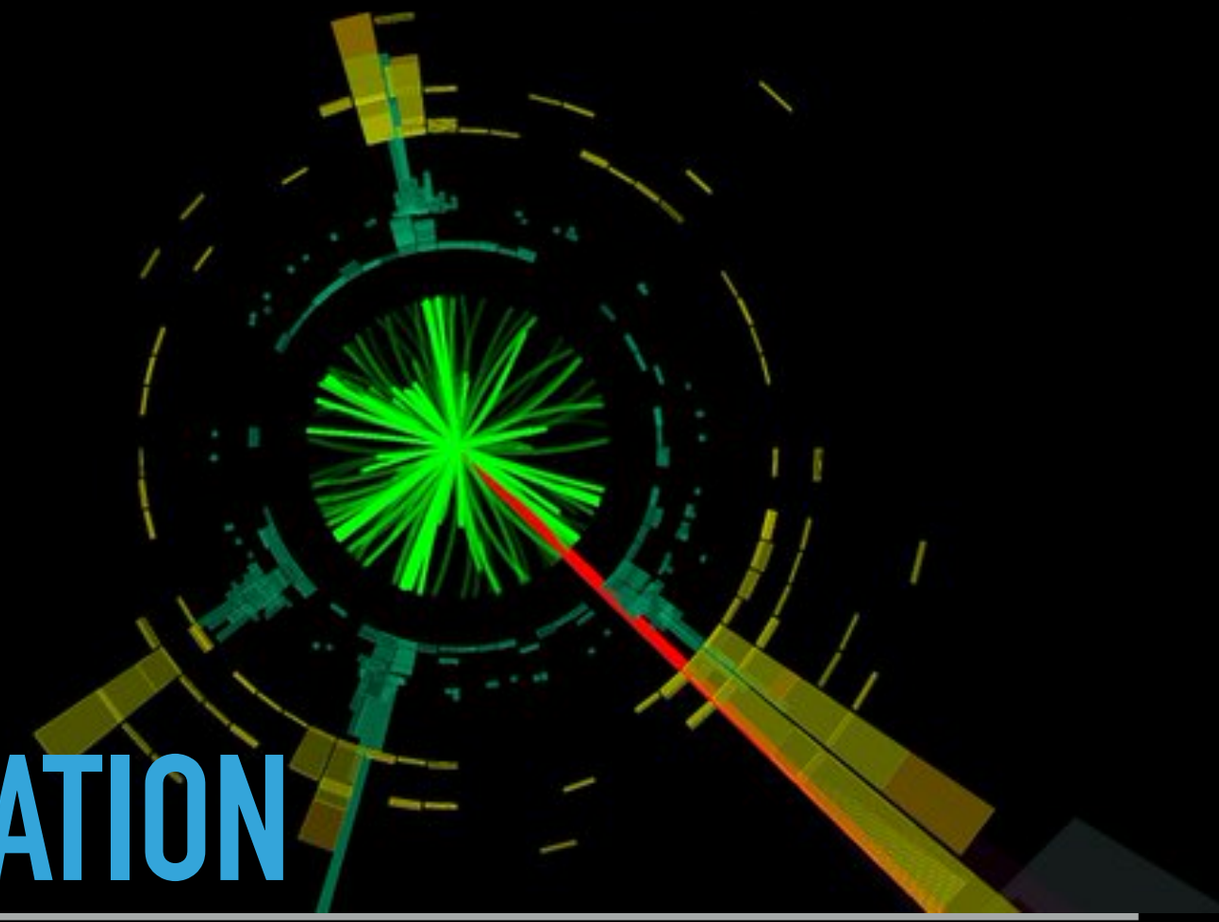


BTAGGING METHODS AND CALIBRATION

SEBASTIAN TAPIA, MAR 08, 2021, ISU

Outline

- btagging algorithms
  - Lifetime based tagging algorithms
    - Impact parameter based
    - Vertex-based
  - Lepton based tagging algorithm

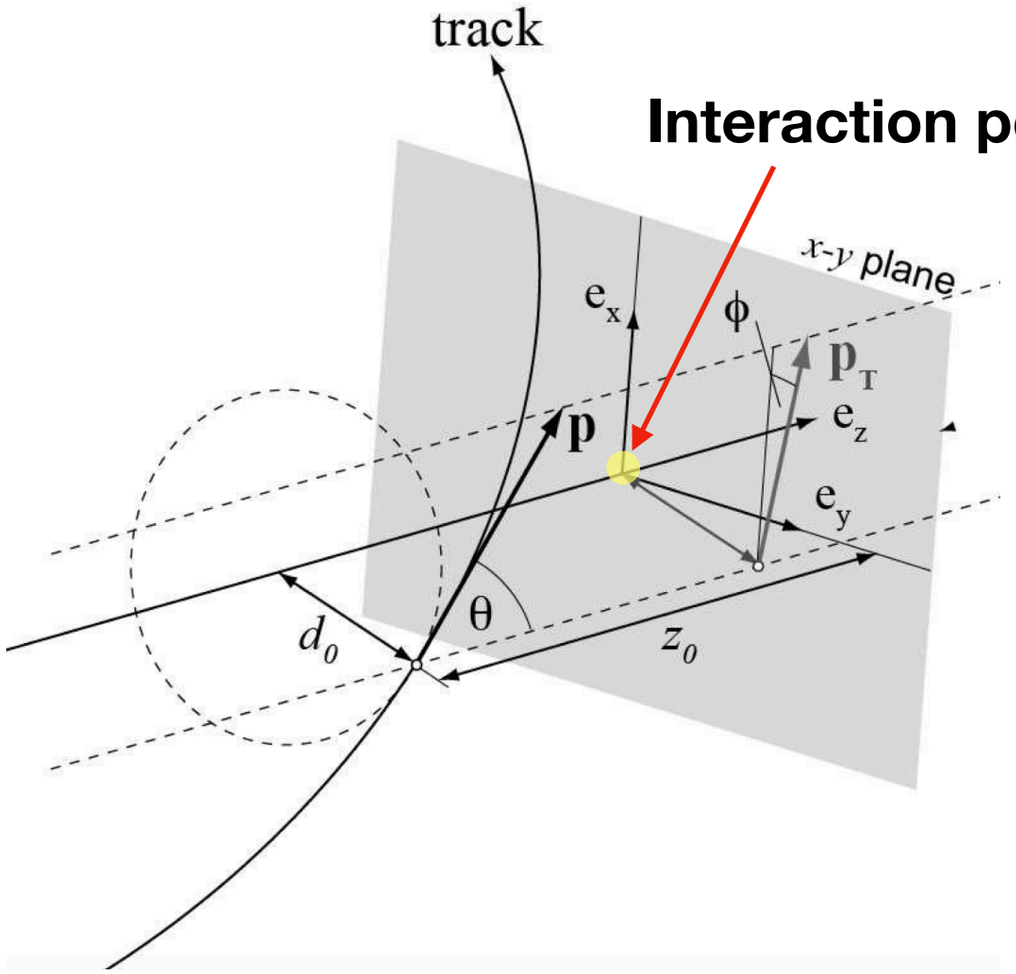
- btagging efficiency calibration
  - System8
  - p_T-rel

- Mistag rate calibration
  - Displaced vertex mass
  - Negative tagging

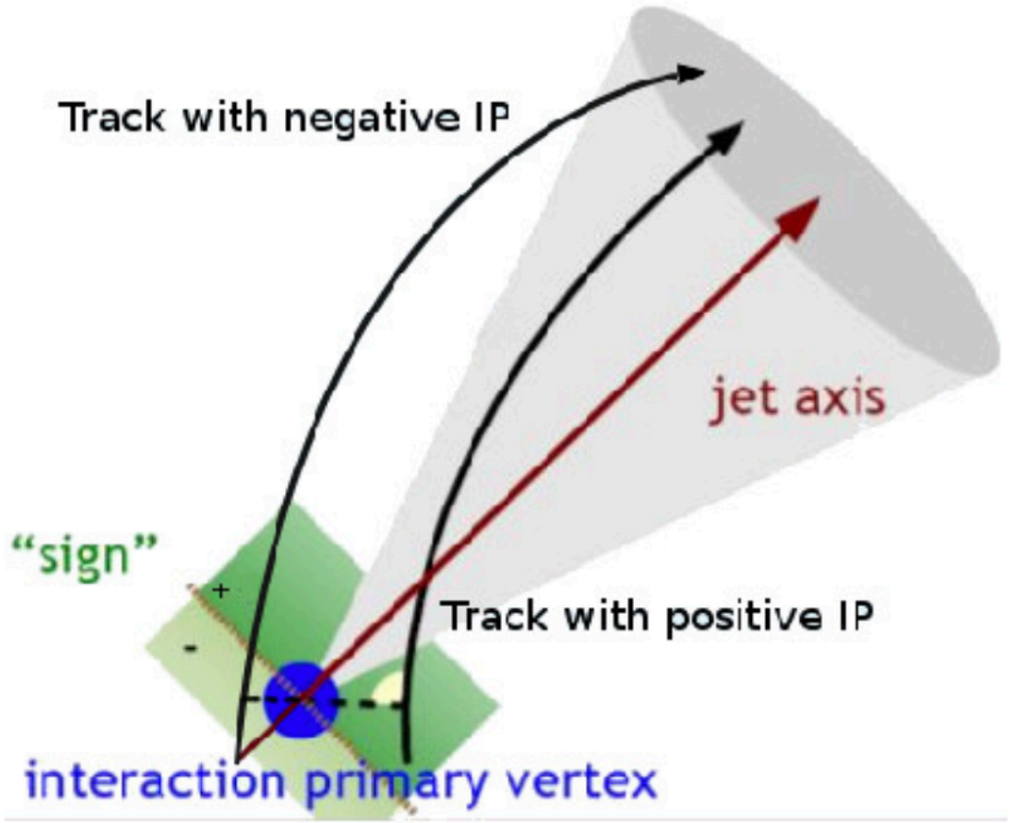
Detailed documentation:
1512.01094
ATLAS-CONF-2012-040
ATLAS-CONF-2010-099

Btagging algorithms

Btagging: The identification of jets containing b-hadrons.
Several dedicated algorithms exploiting specific features like long-lifetime, high mass and decay multiplicity of b-hadrons and the hard b-quark fragmentation.



d_0 and z_0 transverse, longitudinal DCA



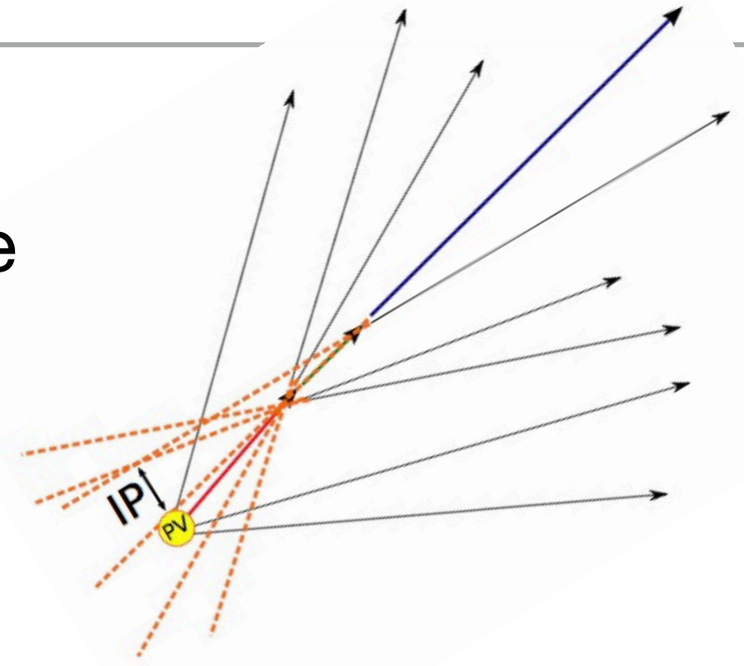
Exploit jet direction, and define "signed" impact parameter and secondary vertex displacement.

Impact parameter

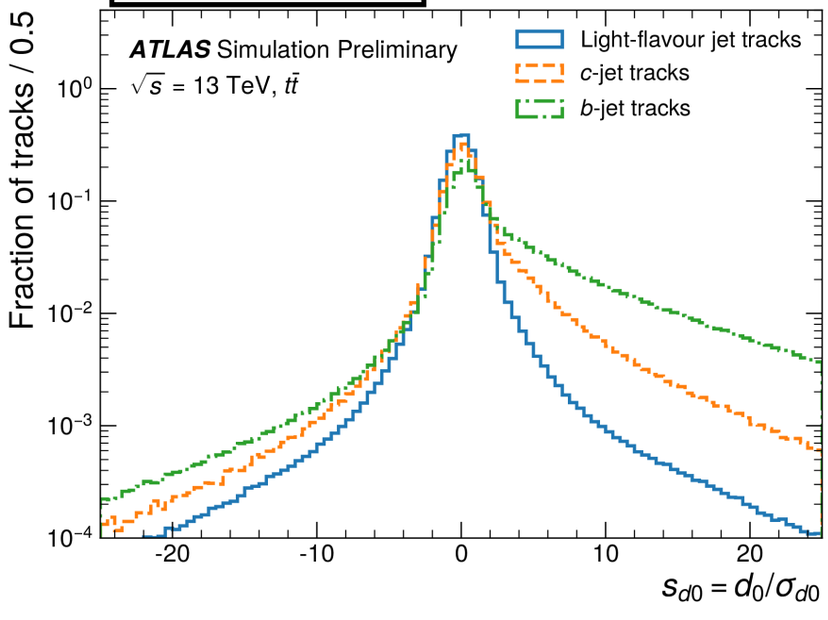
IP3D: log-likelihood base algorithm
Use transverse and longitudinal IP significance as the PDF's

$$w_{track} = p_b / p_l$$

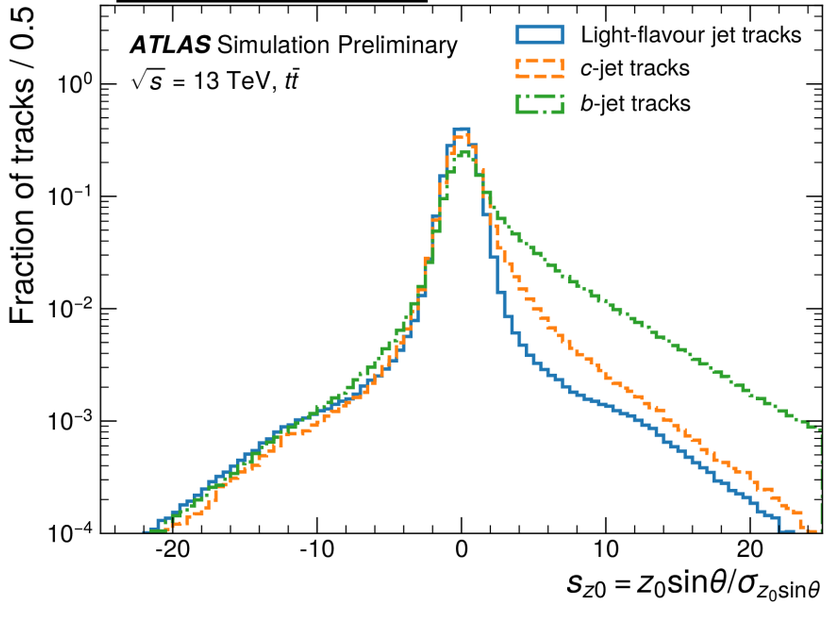
$$w_{jet} = \sum log(w_{track})$$



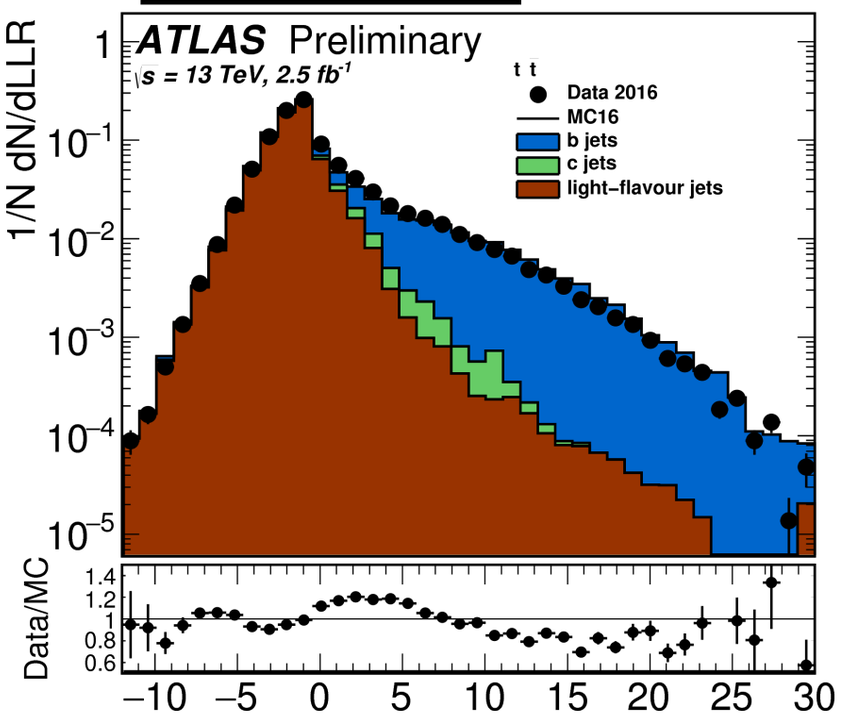
S_{d0} prob



S_{z0} prob



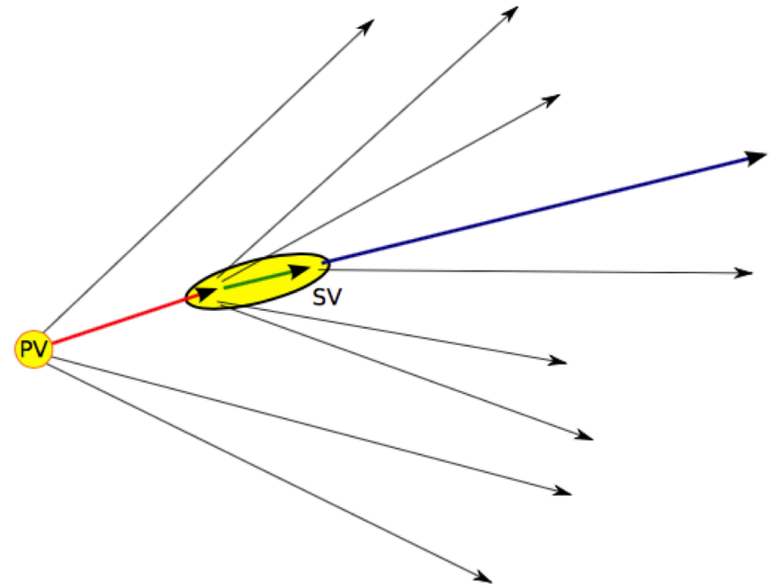
Jet weight

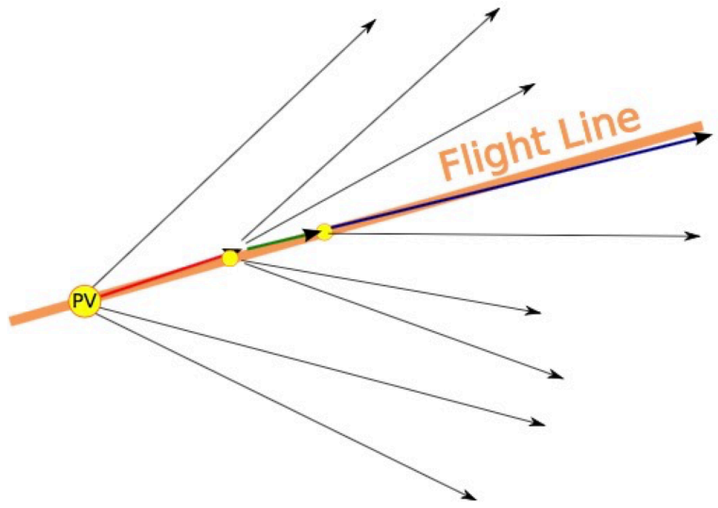


Displaced vertex

Two approaches:

- **Inclusive secondary vertex:**
  - All selected tracks used to form all possible two-track vertices.
  - Selection perform on these two-track vertex candidates.
  - All tracks corresponding to the remaining accepted two-track vertices are used to determine a single secondary vertex.
- **Decay chain multi-vertex:**
  - Exploits the topological structure of weak b- and c- hadron decays inside the jet.
  - Tries to reconstruct the full PV→ b → c-hadron decay chain with the Kalman filter.

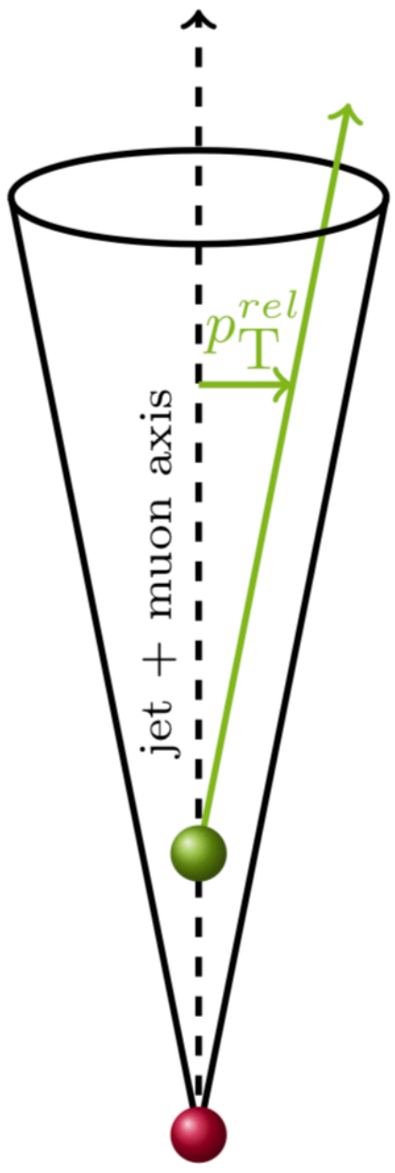




SV	$m(SV)$	Invariant mass of tracks at the SV assuming π masses
	$f_E(SV)$	Fraction of the charged jet energy in the SV
	$N_{TrkAtVtx}(SV)$	Number of tracks used in the SV
	$N_{2TrkVtx}(SV)$	Number of two track vertex candidates
	$L_{xy}(SV)$	Transverse distance between the PV and the SVs
	$L_{xyz}(SV)$	Distance between the PV and the SVs
	$S_{xyz}(SV)$	Distance between the PV and SVs divided by its uncertainty
	$\Delta R(jet, SV)$	ΔR between the jet axis and the direction of the SV relative to the PV

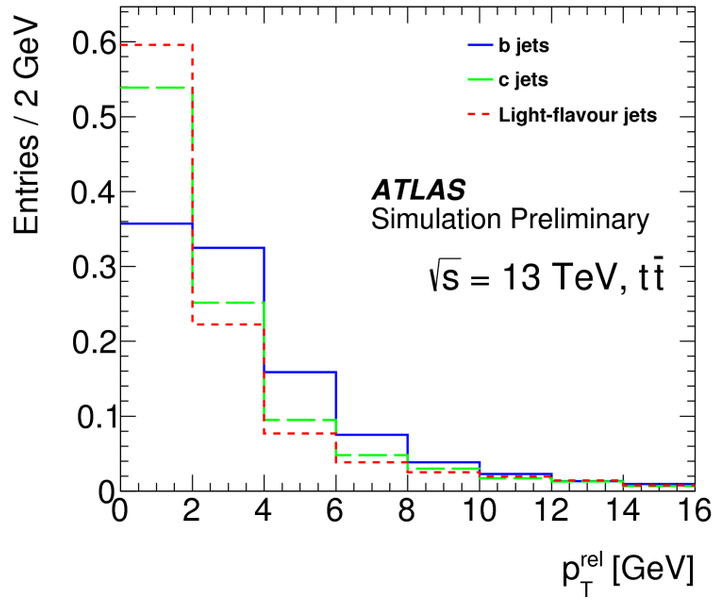
Semi-leptonic decay

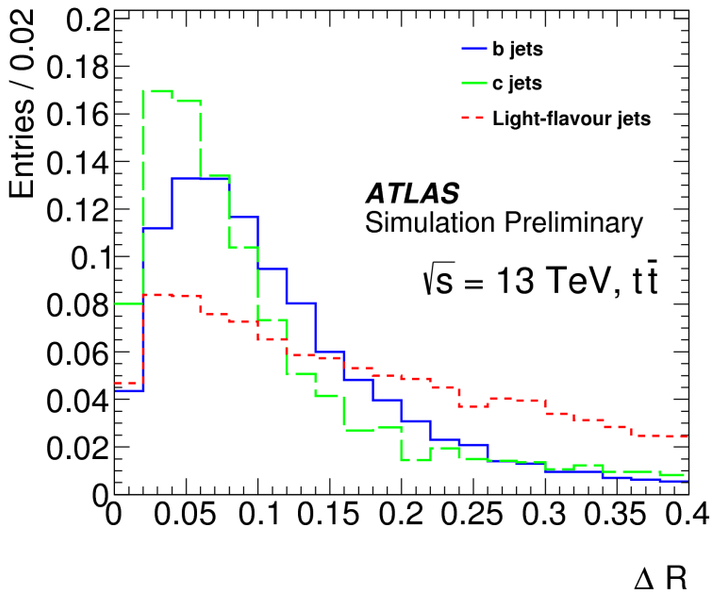
Branching fraction of b-hadrons to muons ~ 20%
Inputs: ΔR, p_T-rel, and DCA

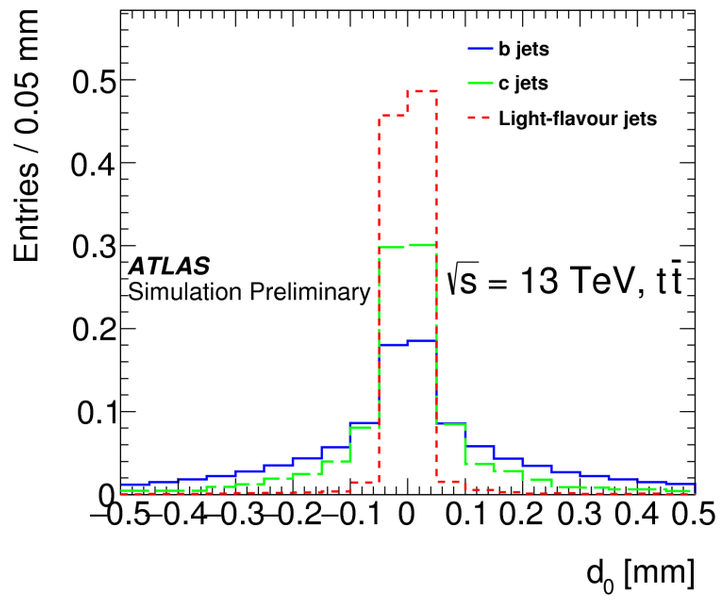


$$p_T^{rel} = \left| \vec{p}_{\mu} \times \frac{\vec{p}_{jet+\mu}}{|\vec{p}_{jet+\mu}|} \right|$$

$$p_T^{rel} \approx p_T^{\mu} * sin(\Delta R(Jet, \mu))$$



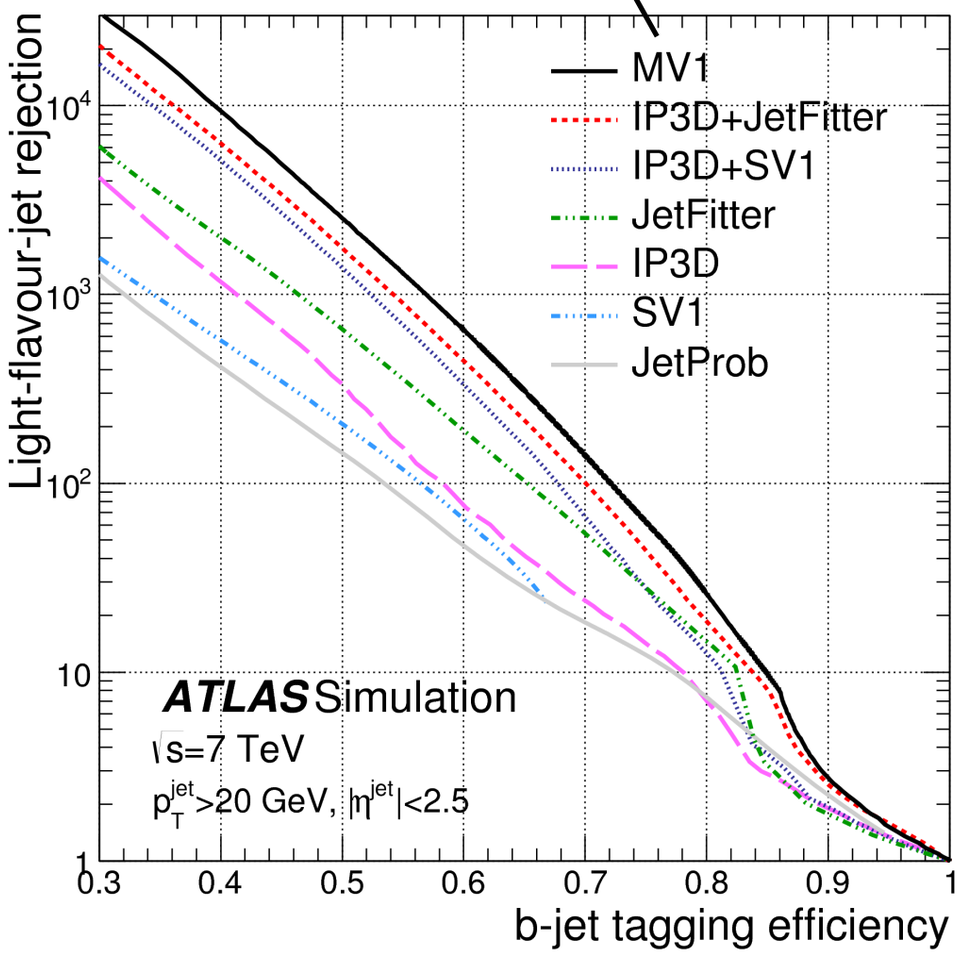




Multi-variable Btagging

MV1 used a BTD

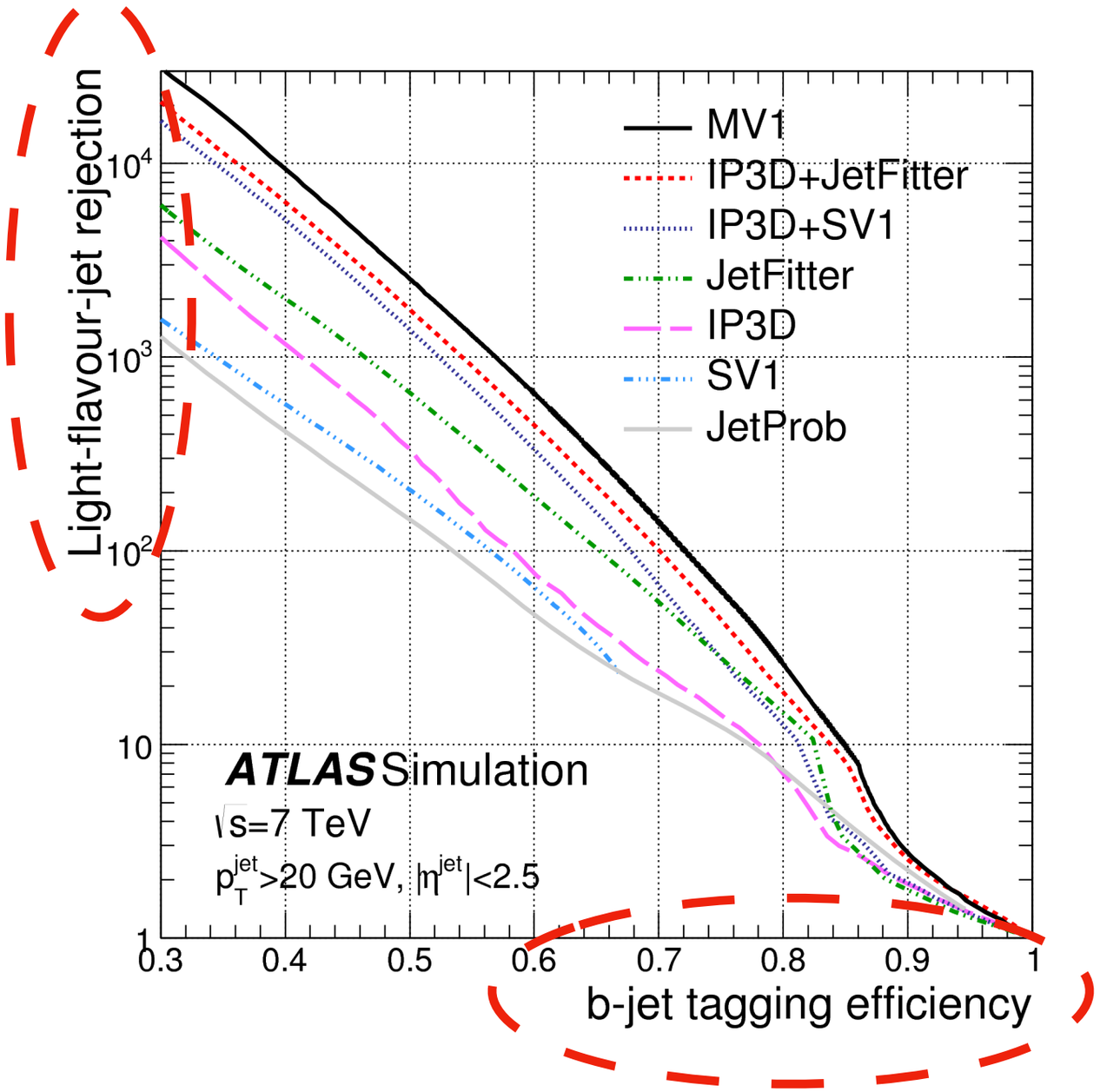
Combining the several taggers into one usually done through ML algorithms such as boost decision threes (BDT), or Neural-nets (NN)



Input	Variable	Description
Kinematics	$p_T(jet)$	Jet transverse momentum
	$\eta(jet)$	Jet pseudo-rapidity
IP2D, IP3D	$\log(P_b/P_{\text{light}})$	Likelihood ratio between the b- and light jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio between the b- and c-jet hypotheses
	$\log(P_c/P_{\text{light}})$	Likelihood ratio between the c- and light jet hypotheses
SV	$m(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion masses
	$f_E(\text{SV})$	Fraction of the charged jet energy in the secondary vertex
	$N_{\text{TrkAtVtx}}(\text{SV})$	Number of tracks used in the secondary vertex
	$N_{\text{2TrkVtx}}(\text{SV})$	Number of two track vertex candidates
	$L_{xy}(\text{SV})$	Transverse distance between the primary and secondary vertices
	$L_{xyz}(\text{SV})$	Distance between the primary and secondary vertices
	$S_{xyz}(\text{SV})$	Distance between the primary and secondary vertices divided by its uncertainty
	$\Delta R(\text{jet}, \text{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex
Jet Fitter	$N_{\text{2TrkVtx}}(\text{JF})$	Number of 2-track vertex candidates (prior to decay chain fit)
	$m(\text{JF})$	Invariant mass of tracks from displaced vertices assuming pion masses
	$S_{\text{xyz}}(\text{JF})$	Significance of the average distance between the primary and displaced vertices
	$f_E(\text{JF})$	Fraction of the charged jet energy in the secondary vertices
	$N_{\text{1-trk vertices}}(\text{JF})$	Number of displaced vertices with one track
	$N_{\geq\text{2-trk vertices}}(\text{JF})$	Number of displaced vertices with more than one track
	$N_{\text{TrkAtVtx}}(\text{JF})$	Number of tracks from displaced vertices with at least two tracks
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})$	ΔR between the jet axis and the vectorial sum of the momenta of all tracks attached to displaced vertices

Btagging Calibration

How do we know that our btagging **efficiency** and **rejection** extracted from simulation is right?



Calibration factor:

$$\kappa_b^{data/sim} = \frac{\epsilon_{b,l}^{data}}{\epsilon_{b,l}^{sim}}$$

where:

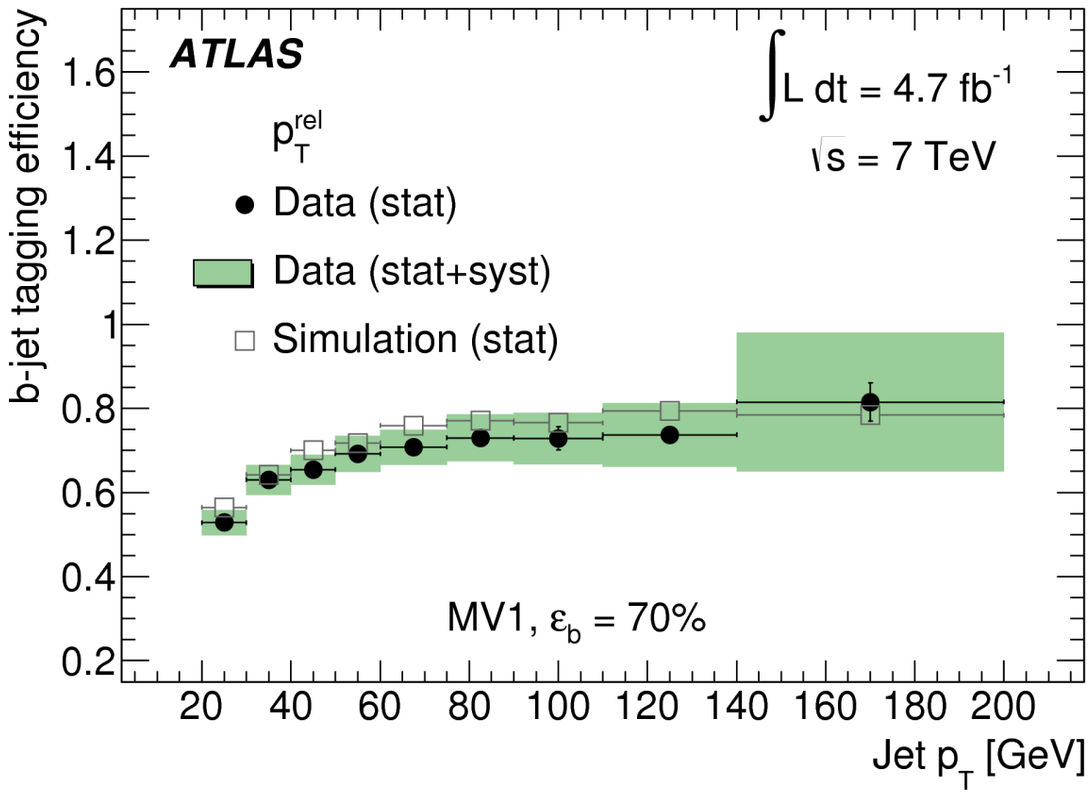
$$\epsilon_{b,l}^{data,sim} = \left. \frac{N_{b,l}^{tag}}{N_{b,l}} \right|^{data,sim}$$

Efficiency calibration

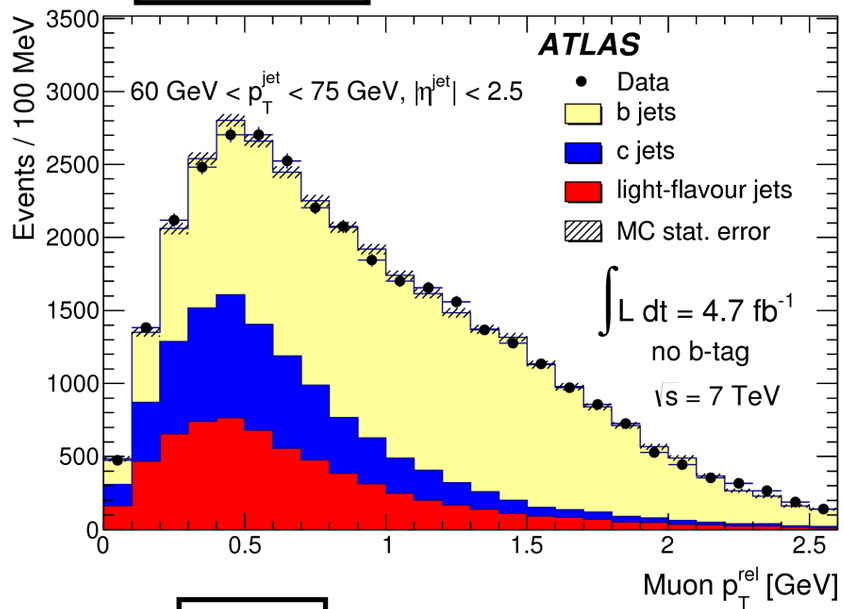
$$p_T^{rel} = \left| \vec{p}_{\mu} \times \frac{\vec{p}_{jet+\mu}}{|\vec{p}_{jet+\mu}|} \right| \approx p_T^{\mu} * sin(\Delta R(Jet, \mu))$$

p_T^{rel} method: Template fit of muon p_T respect to jet axis (p_T^{rel}) to get flavor fraction before and after b-tagging

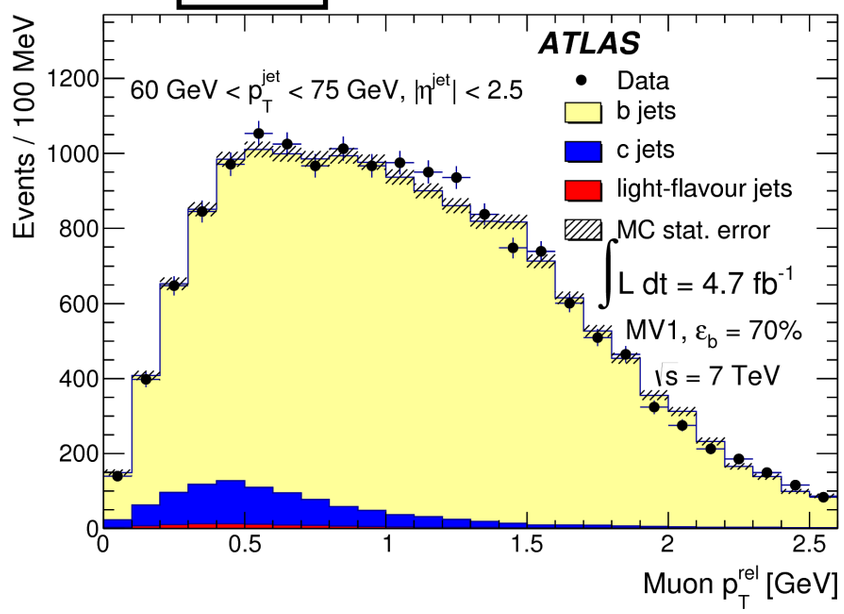
$$\epsilon_b^{data} = \frac{f_b^{tag} \cdot N^{tag}}{f_b \cdot N} = \frac{f_b^{tag} \cdot N^{tag}}{f_b^{untag} \cdot N^{untag} + f_b^{tag} \cdot N^{tag}}$$



Untag



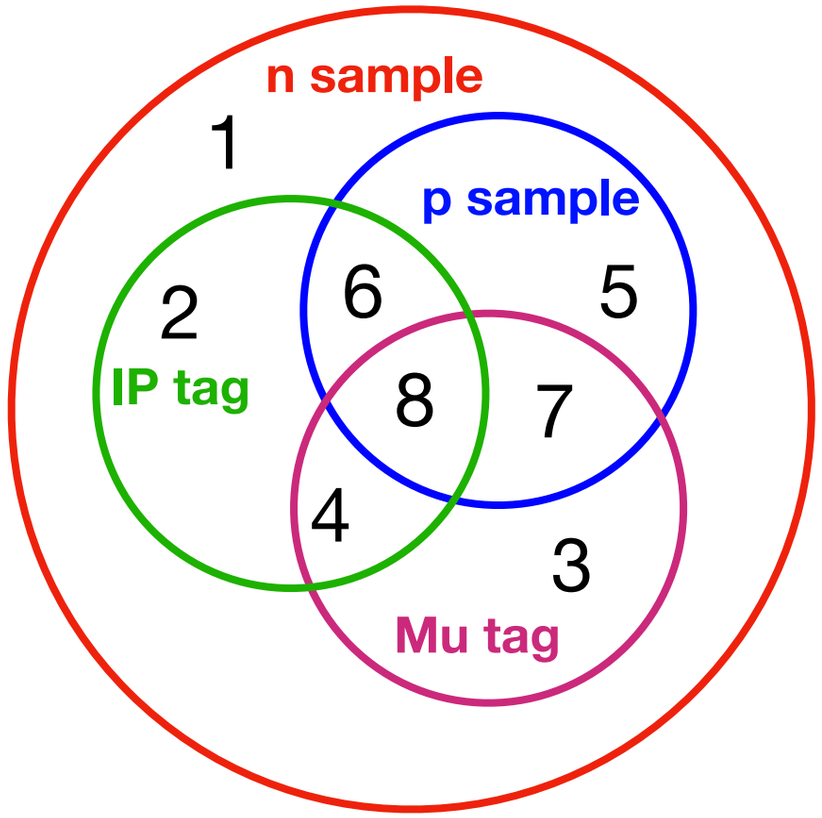
Tag



Efficiency calibration

System8 : Define 3 independent jet selection, construct 8 samples.
System8 solved by χ^2 minimization.

$$
\begin{aligned}
n &= n_b + n_{cl} \\
p &= p_b + p_{cl} \\
n^{LT} &= \epsilon_b^{LT} n_b + \epsilon_{cl}^{LT} n_{cl} \\
p^{LT} &= \alpha_6 \epsilon_b^{LT} p_b + \alpha_4 \epsilon_{cl}^{LT} p_{cl} \\
n^{MT} &= \epsilon_b^{MT} n_b + \epsilon_{cl}^{MT} n_{cl} \\
p^{MT} &= \alpha_5 \epsilon_b^{MT} p_b + \alpha_3 \epsilon_{cl}^{MT} p_{cl} \\
n^{LT,MT} &= \alpha_1 \epsilon_b^{LT} \epsilon_b^{MT} n_b + \alpha_2 \epsilon_{cl}^{LT} \epsilon_{cl}^{MT} n_{cl} \\
p^{LT,MT} &= \alpha_7 \alpha_6 \alpha_5 \epsilon_b^{LT} \epsilon_b^{MT} p_b + \alpha_8 \alpha_4 \alpha_3 \epsilon_{cl}^{LT} \epsilon_{cl}^{MT} p_{cl}
\end{aligned}
$$



α terms are correlation factors

Application to lepton sample

n = sample of Jets with a lepton within

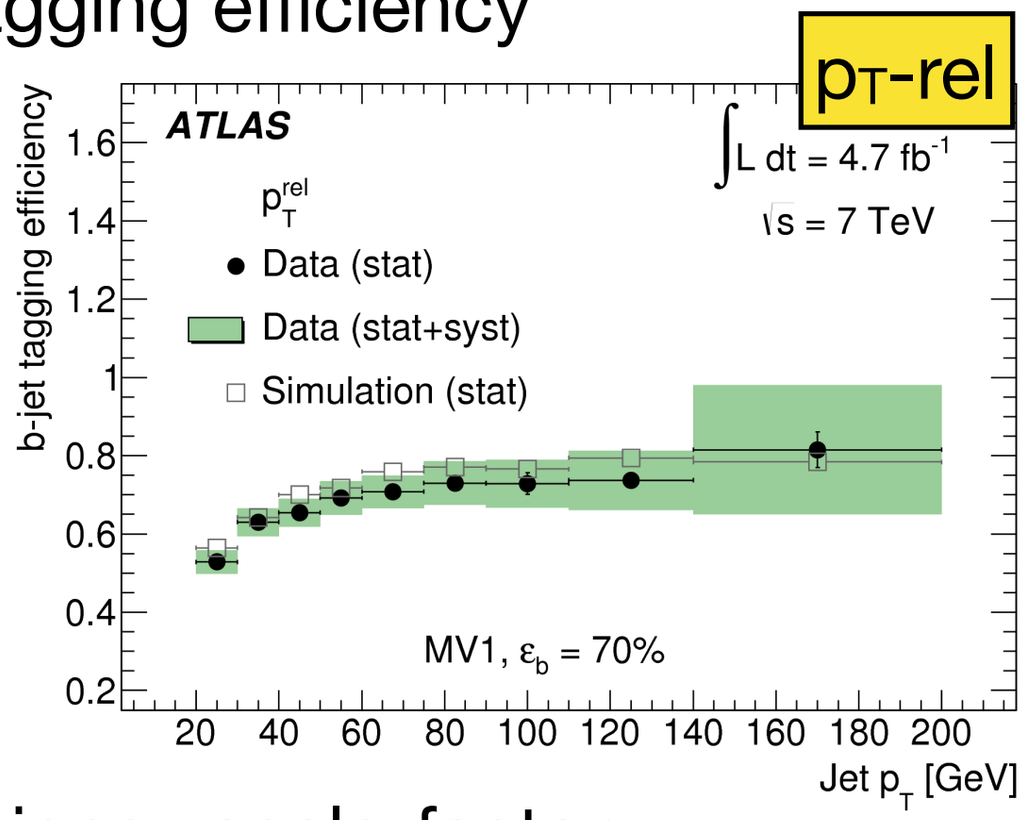
p = sample of Jets with a lepton within + a b-tagged opposite jet

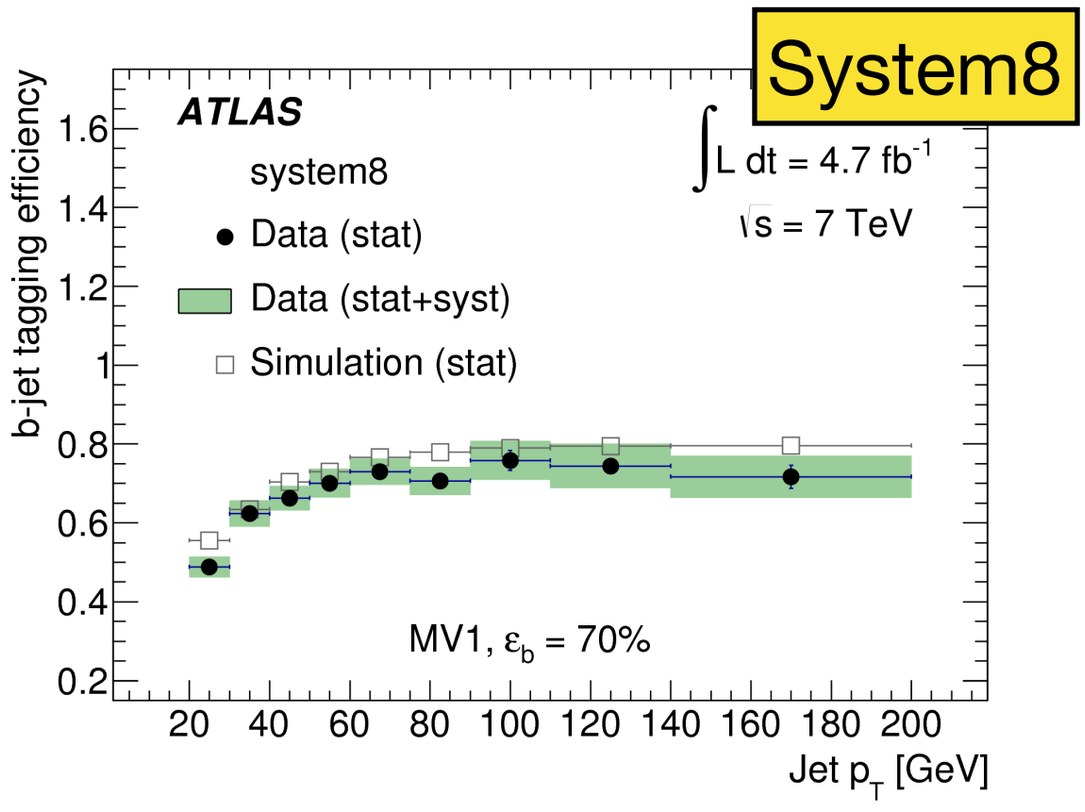
n^{LT}/p^{LT} = respective sample + “IP tagging”

n^{MT}/p^{MT} = respective sample + $p_T^{rel} > 0.7$

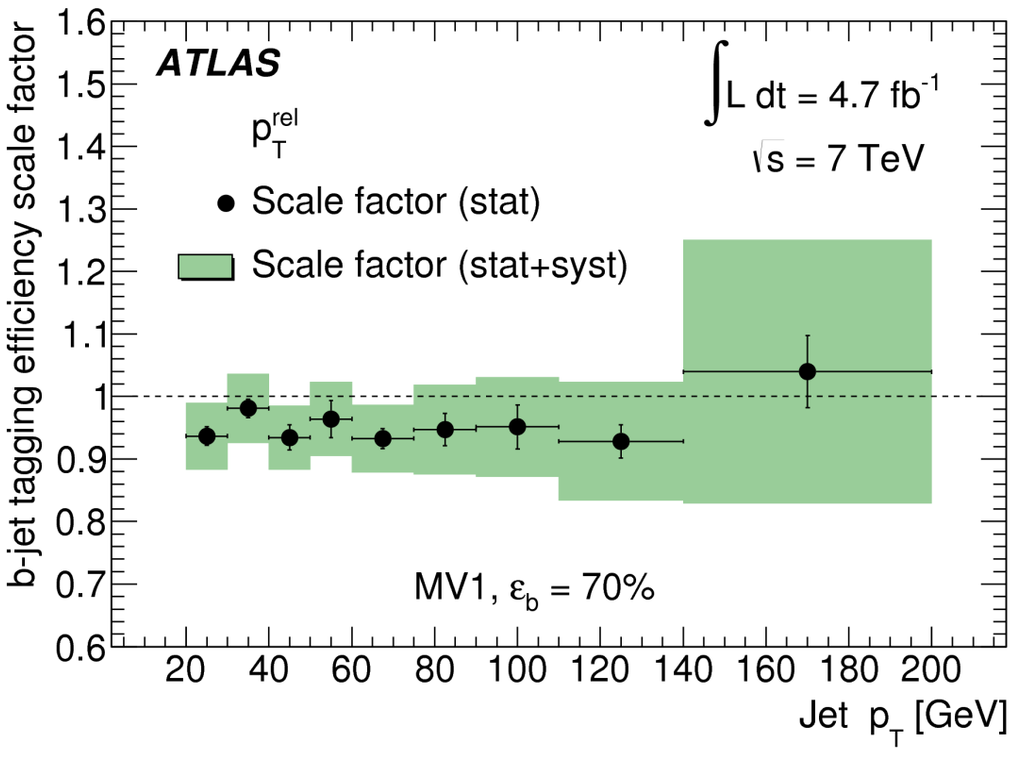
Efficiency scale factor

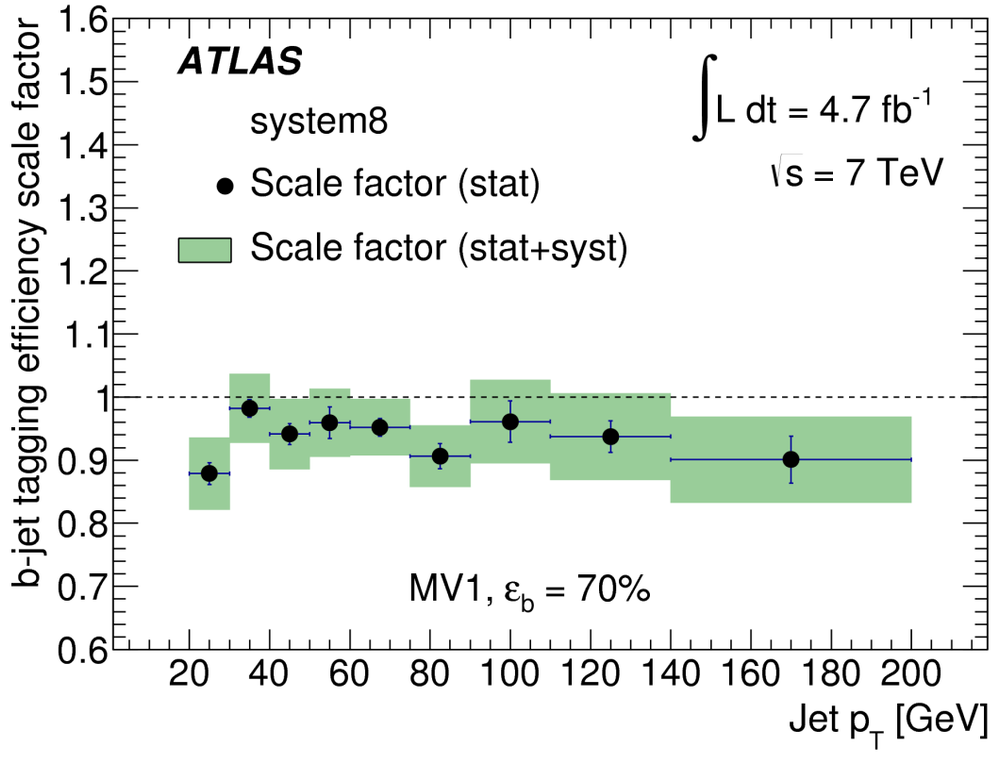
Btagging efficiency





Efficiency scale-factor

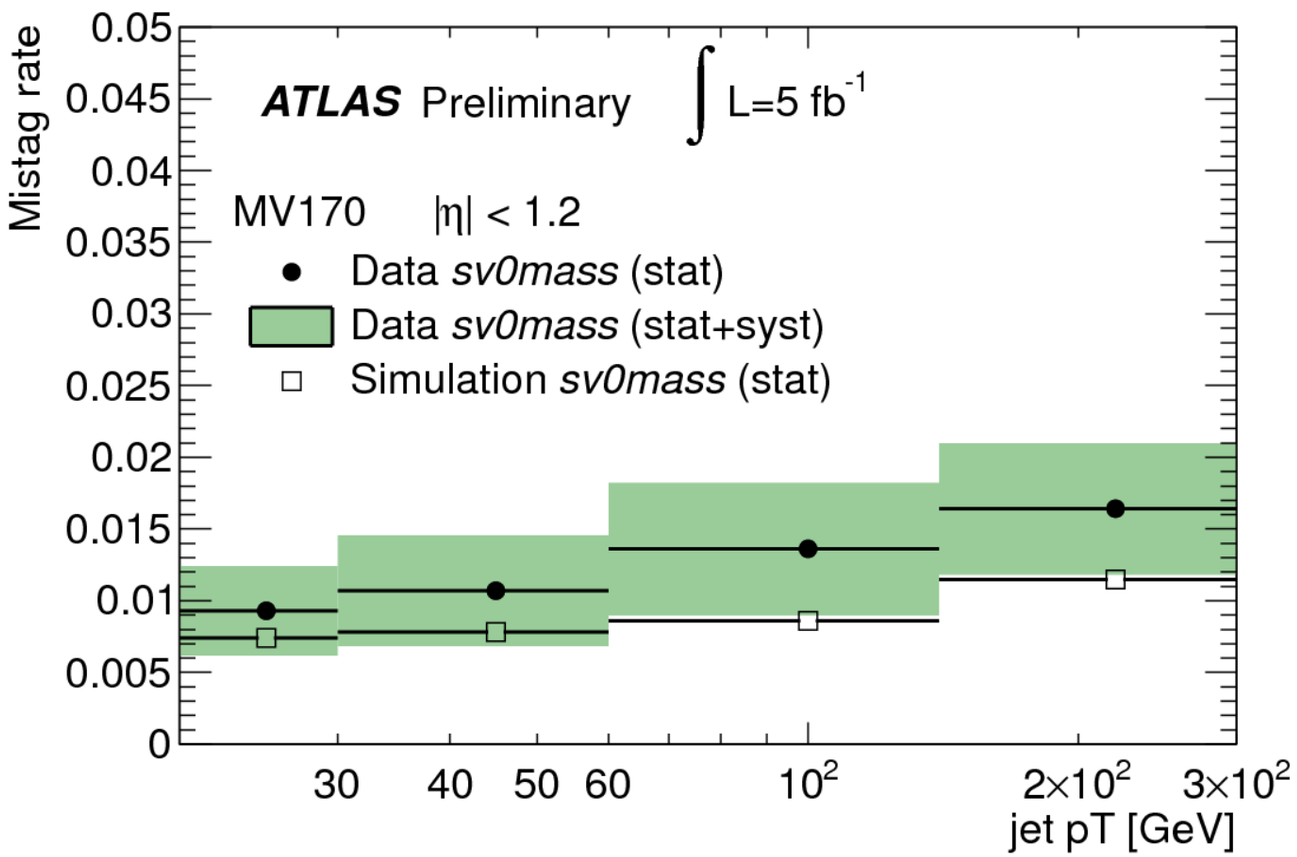




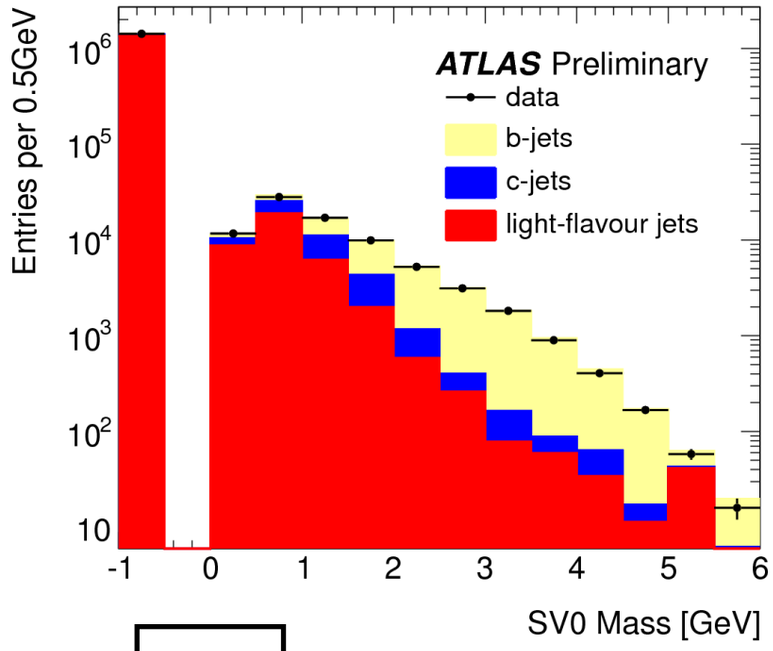
Mistag rate calibration

Displaced vertex mass: Same idea than pt-rel method by focus in light-jet efficiency

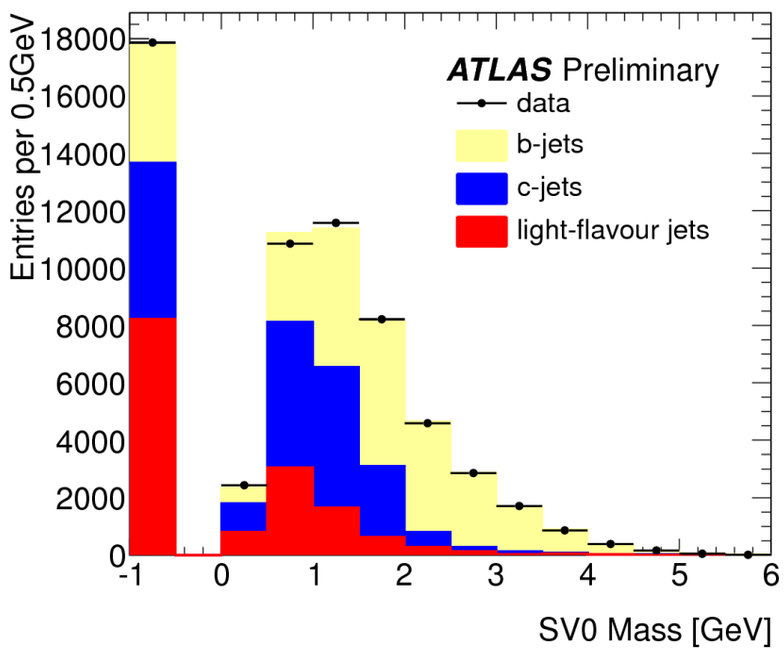
$$\epsilon_l^{data} = \frac{f_l^{tag} \cdot N^{tag}}{f_l \cdot N} = \frac{f_l^{tag} \cdot N^{tag}}{f_l^{untag} \cdot N^{untag} + f_l^{tag} \cdot N^{tag}}$$



Untag



Tag

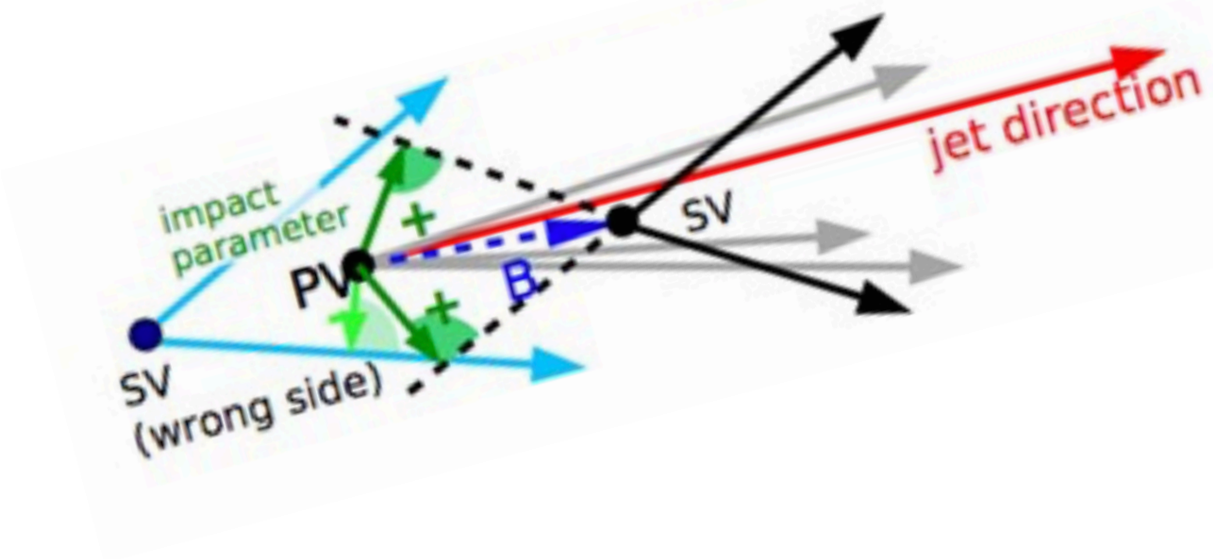


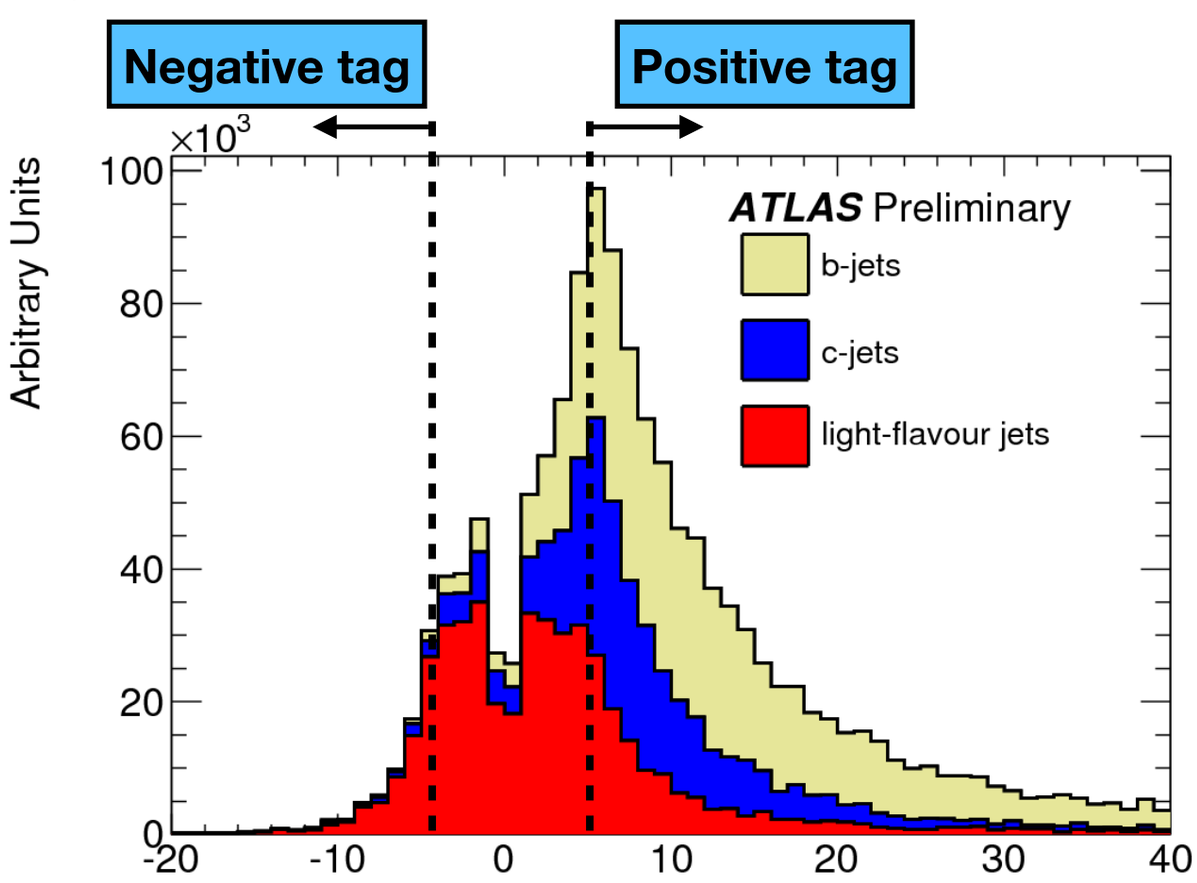
Mistag rate calibration

Negative tagging: inverting tagging based on impact parameter or displayed length.

- The mistag rate ε_l is then approximated by the negative tag rate of the inclusive jet sample, ε_{inc}, plus corrections from:

- Negative tag distribution is not the same for all flavors
  - b-, c-jet more sensitive to finite jet direction resolution which can flip the discriminant sign
- Negative tag distribution is not perfectly symmetric
  - Symmetric distribution from tracking resolution effects
  - Asymmetric distribution from ling-lived particles(K,L) decays and material interaction.

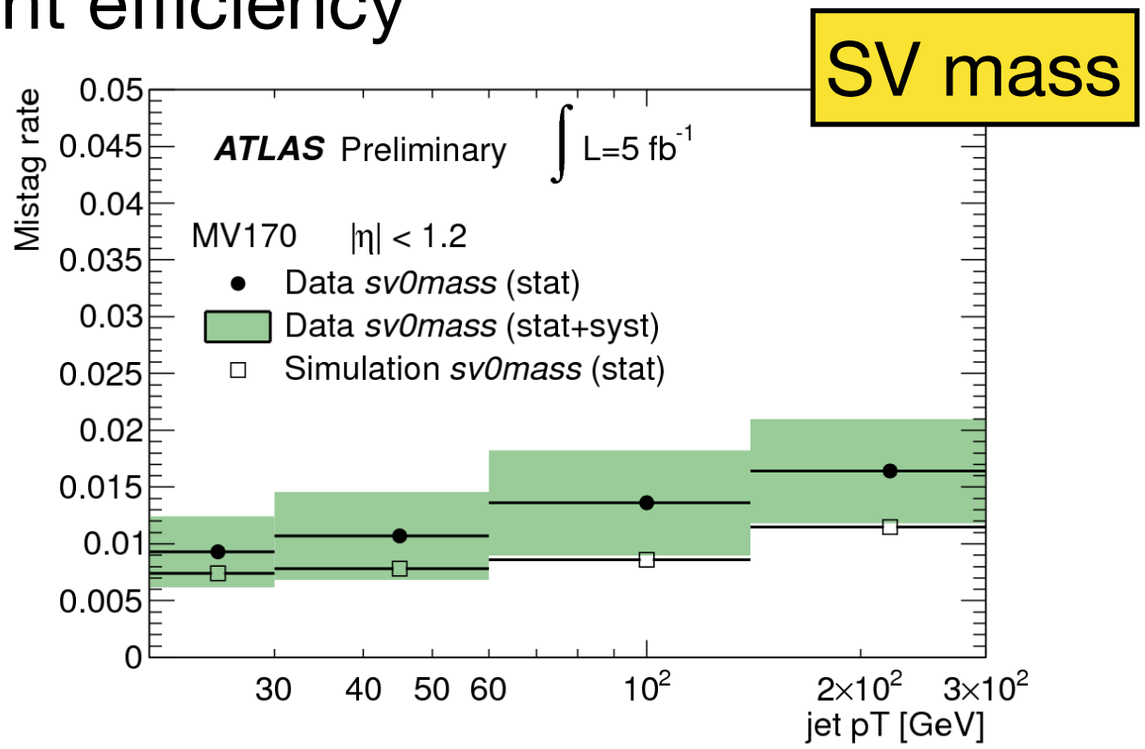


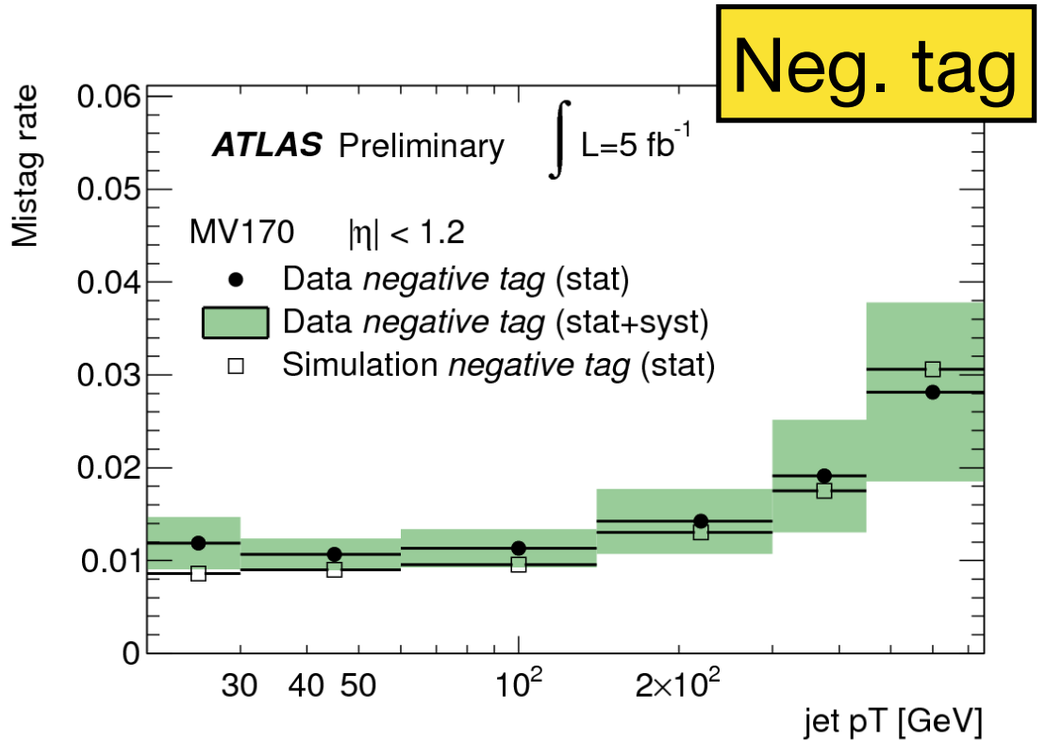


Signed decay length

Efficiency scale factor

Light efficiency





Efficiency scale-factor