b-Jet Identification in CMS

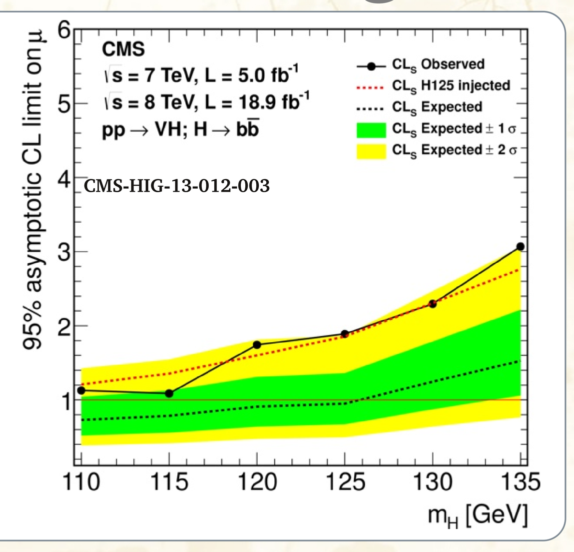
Camille Beluffi (on behalf of the CMS collaboration)

Université Catholique de Louvain-la-Neuve, Belgium - Institut Pluridisciplinaire Hubert Curien, Strasbourg

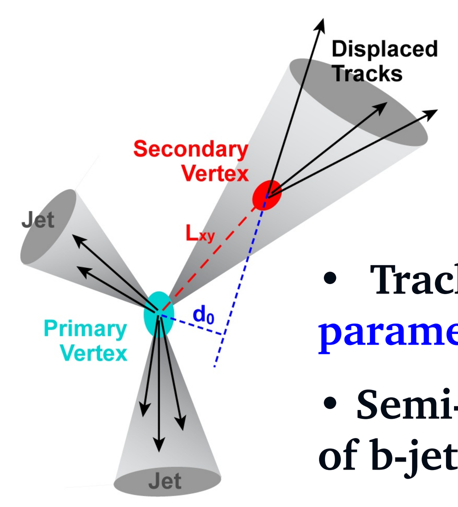
The LHC, the CMS detector, and the challenges

The CMS experiment uses proton-proton collisons from the LHC to probe the limits of the Standard Model (SM).

B-tagging is the identification of jets coming from the hadronization of a b-quark (b-jet); it is then used to suppress the main backgrounds involving jets from gluons and light-flavor quarks for precise measurements of SM processes (top quark pair cross section measurement), Higgs properties studies (measurement of $H \rightarrow b\bar{b}$ to probe the Higgs-boson coupling to down-type quarks) or search for new physics, for example with boosted topologies.



B-tagging in a nutshell

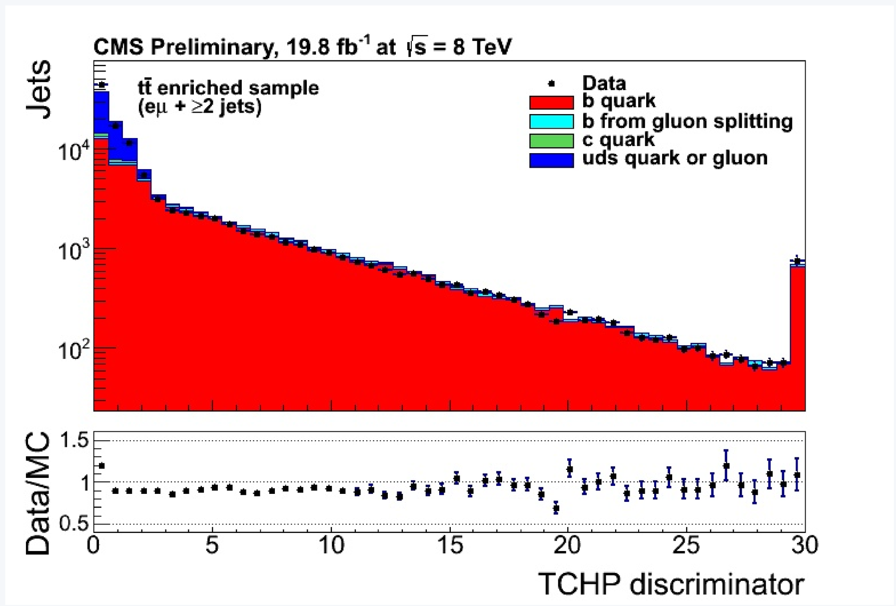


- B-hadrons propagate a measurable distance before decaying; it creates a displaced secondary vertex (SV), with a flying distance L_{xy} higher than its resolution.
- Tracks coming from SV have a large impact parameter (IP, distance track-primary vertex).
- Semi-leptonic decay of b-hadrons leads to ∼20% of b-jets containing a muon/electron.

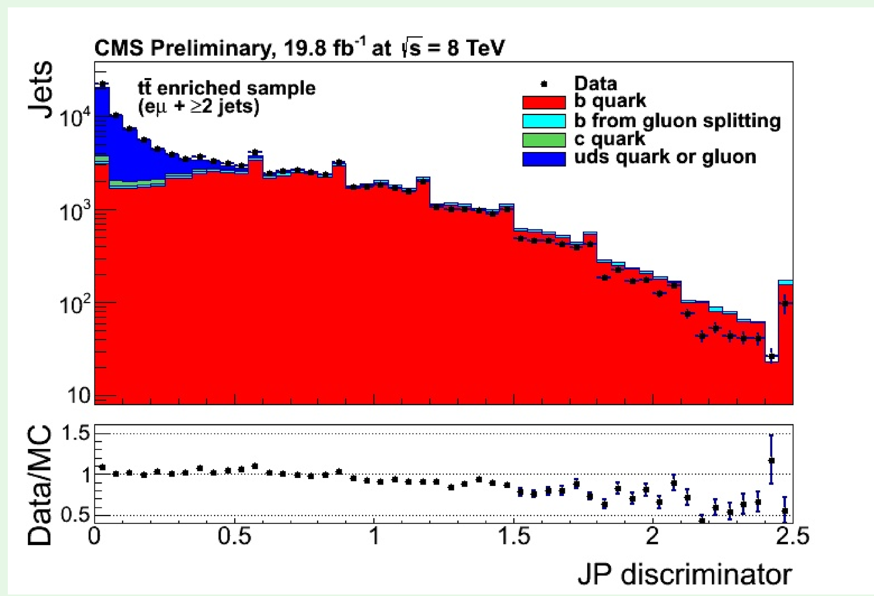
For 2012 8 TeV dataset, mainly three taggers were used, based on b-jet properties. Each one yields a single discriminator value for a jet.
Three operation points, corresponding to a misidentification probability (P) for light partons, are defined: "Loose" (P=10%), "Medium" (P=1%), Tight (P=0.1%).

→ → → Increasing performance and complexity → → →

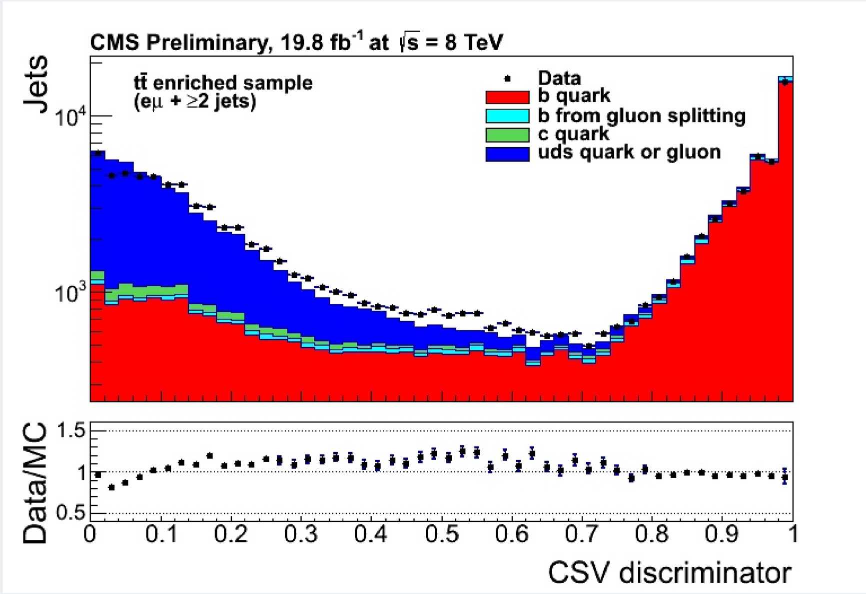
- **Track Counting High Purity** (TCHP): based on the third track with the highest impact parameter significance .



- **Jet Probability** (JP): the jet is assigned a likelihood estimation that all associated tracks come from the primary vertex (PV).

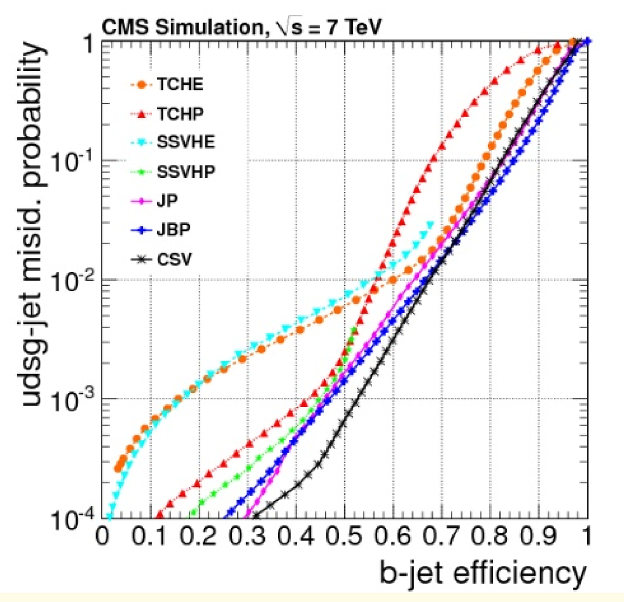


- **Combined Secondary Vertex** (CSV): secondary vertices and track-based lifetime information are used to build a likelihood-based discriminator.



Performance measurements

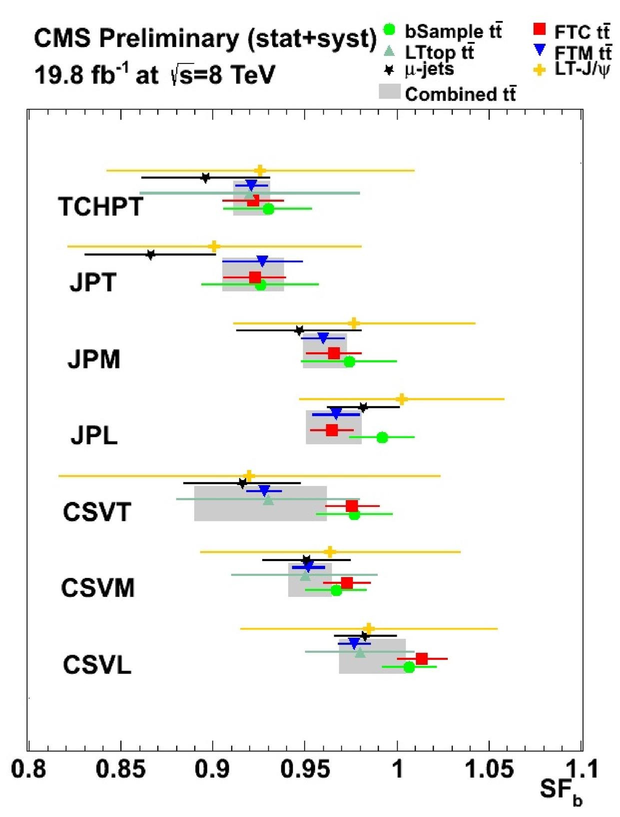
Performances are computed in simulated events, measured in data. Scale factors (SF) quantify the difference. They are measured for b-tagging efficiency but also for mis-identification probability, in different jet p_T ranges (and η for the misidentification rate).



b-tagging efficiency

- using multijets events:
- System8 method: uses two weakly correlated taggers, one of which is the one to be probed. They are tested in two samples with different b-quark enrichement.
- PtRel, IP3D and Life-Time (LT) methods based on a fit in data of a discriminant variable distribution. LT method also used with top quark pairs events.

- using top quark pairs events:
- bSample method defines two regions, one enriched in b-jets, the other depleted. Efficiencies are derived from the difference between the discriminator distributions.



Different measurements are combined taking into account their correlations.

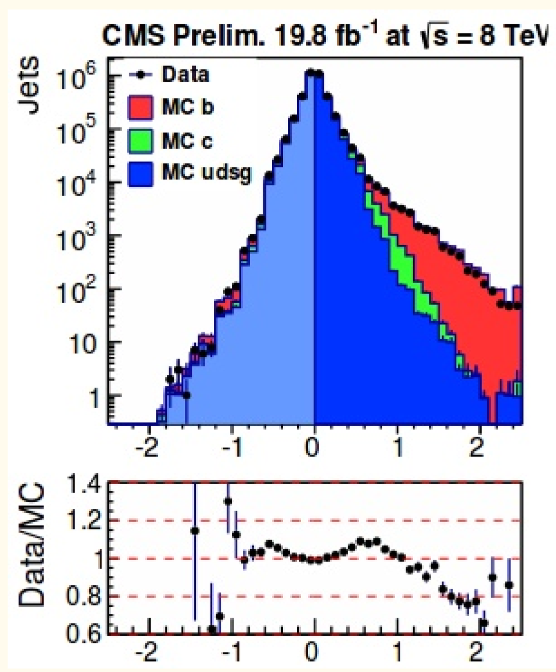
- Flavor Tag Matching/Consistancy (FTM/C) methods requires consistency between the observed and expected number of tags in the lepton+jets/dilepton events.

Misidentification probability

The probability that a light jets is misidentified as a b-jet is derived in data from taggers built as the original ones but filled with only tracks with negative IP or with SV with negative decay length, the negative part being dominated by detector resolution effects. It is corrected by a mistag/negative rate factor derived from simulations.

b tagger	misidentification probability	SF_{light}
JPL	0.0944 ± 0.0004	$1.03 \pm 0.01 \pm 0.07$
CSVL	0.0990 ± 0.0004	$1.10 \pm 0.01 \pm 0.05$
JPM	0.0105 ± 0.0002	$1.10 \pm 0.02 \pm 0.20$
CSVM	0.0142 ± 0.0002	$1.17 \pm 0.02 \pm 0.15$
TCHPT	0.0026 ± 0.0001	$1.27 \pm 0.06 \pm 0.27$
JPT	0.0013 ± 0.0001	$1.11 \pm 0.07 \pm 0.31$
CSVT	0.0016 ± 0.0001	$1.26 \pm 0.07 \pm 0.28$

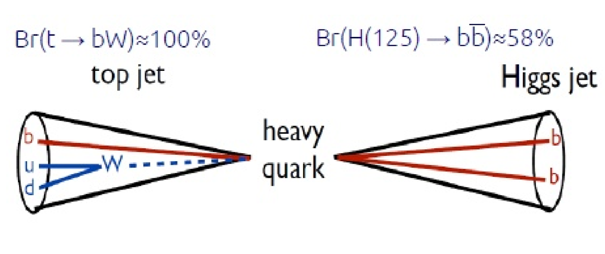
Misidentification probability for jet p_T in range [80-120] GeV/c.

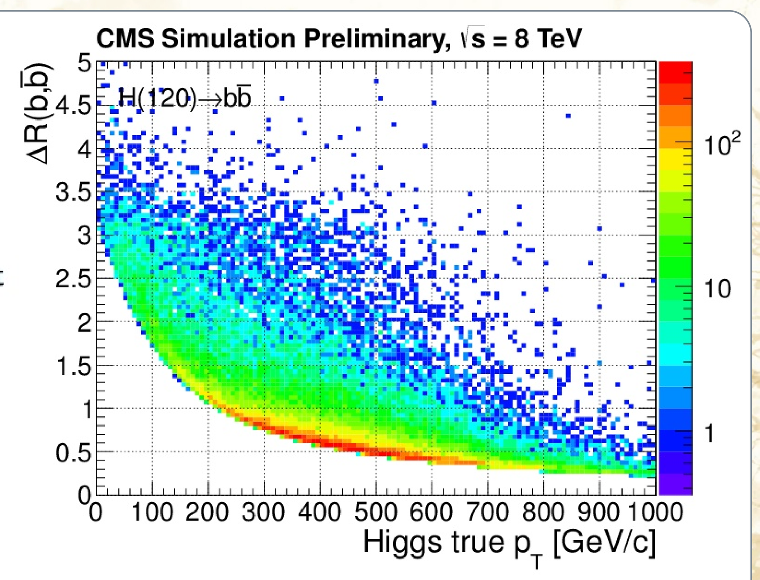


JP discriminator for tracks with negative IP.

B-tagging in boosted topologies

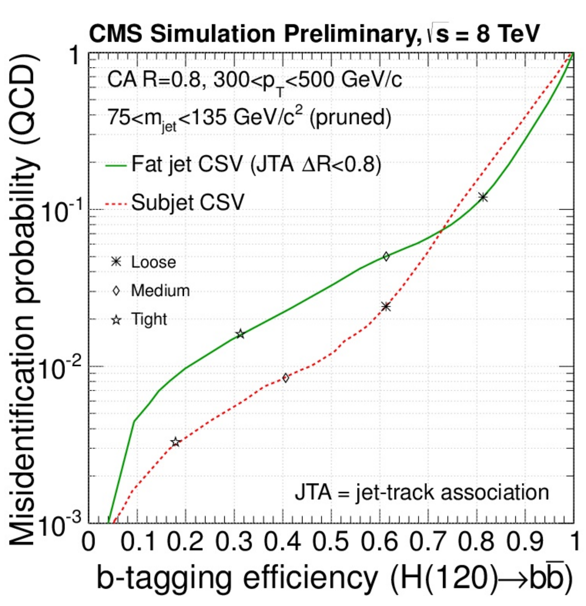
High mass resonances with a final state containing b quarks are predicted by various models of new physics. Some may decay to top quark pairs, some into Higgs bosons and if they have a large enough momentum, their decay products are very collimated (small angular distance ΔR between them), ending up clustered in a single fat jet (boosted regime).

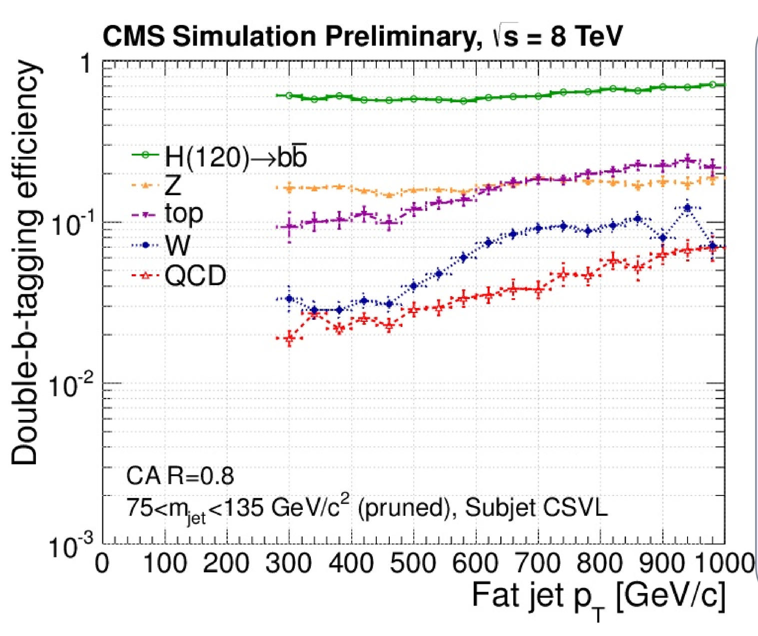




B-tagging in jet substructure

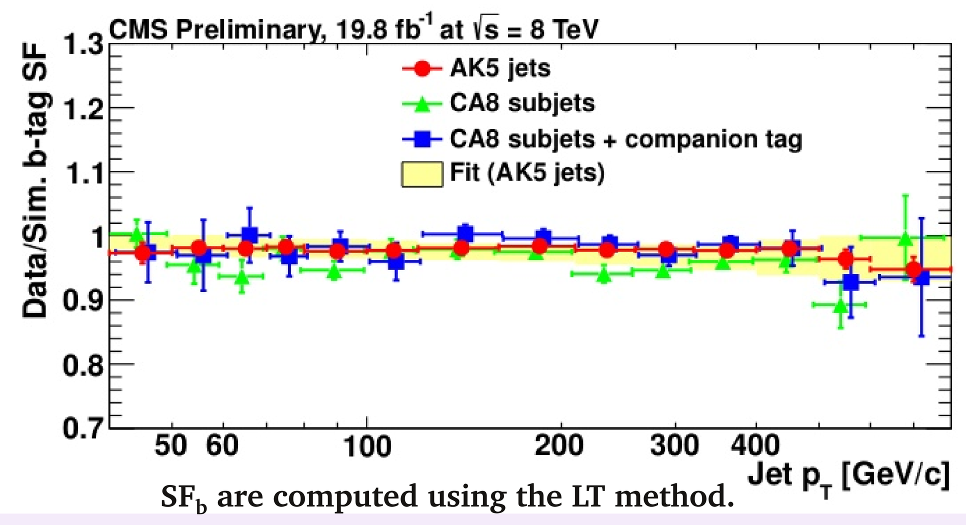
One important reconstruction parameter is the size of the jet, to include all decay products and it depends on the jet p_T. Some possibilities have been studied in detail such as Cambridge/Aachen (CA) jets of size R=1.5 for top-tagging. Fat jet substructure is identified by undoing the CA algorithm clustering. For Higgs-tagging, focus is on CA jets of size R=0.8. Pruned jets describe the jet sub-structure. B-tagging can then be applied on the fat jet or on its substructure components, the second option being the most performant.





Performances

Measurement of b-tagging efficiency in boosted topologies is challenging: standard methods use non-boosted objects so their results are not necessarily applicable. For Higgs-tagging, efficiency is measured using LT method on different control samples to study the performance of b-tagging both on fat jets and subjets. The agreement found between data and simulation is compatible with what is observed in non-boosted topologies.



SF_b are computed using the LT method.

For more information CMS PAS BTV-13-001 and 2013 JINST 8 P04013

ICHEP 2014 poster session, 2-9 July 2014 - Valencia, Spain

camille.beluffi@cern.ch