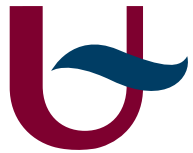
Jet vetoes and tagging in Higgs-boson studies at CMS

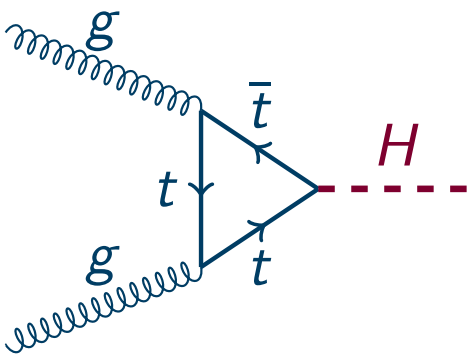
Tom Cornelis
for the CMS Collaboration



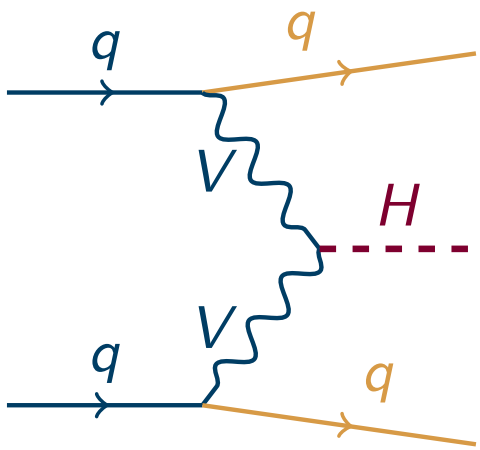
Jet Vetoes and Jet Multiplicity Observables at the LHC, Durham, July 19th 2013

Universiteit Antwerpen

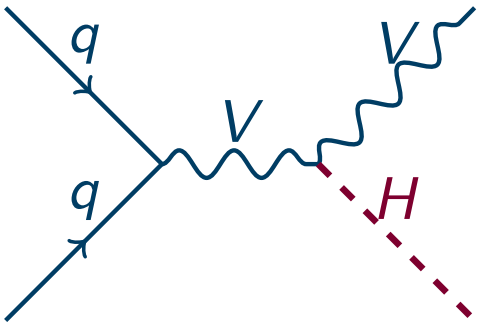
Higgs production modes



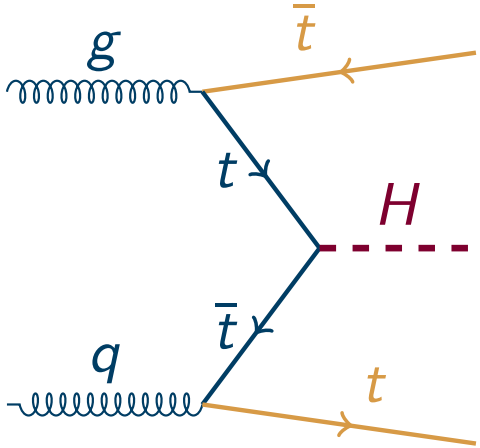
gluon-gluon fusion



vector boson fusion



Higgs strahlung

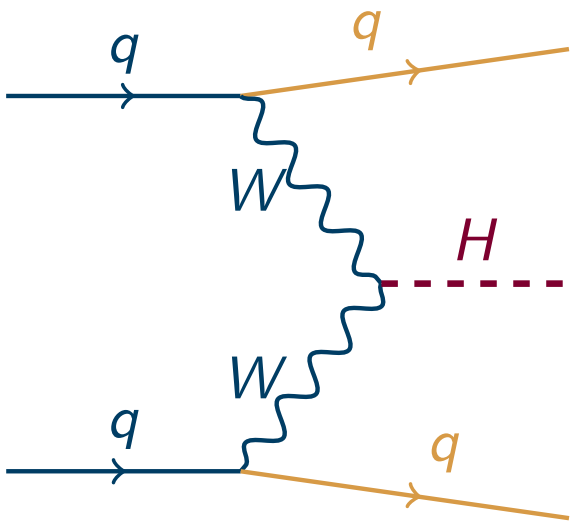


$t\bar{t}$ fusion

Higgs analyses with jets

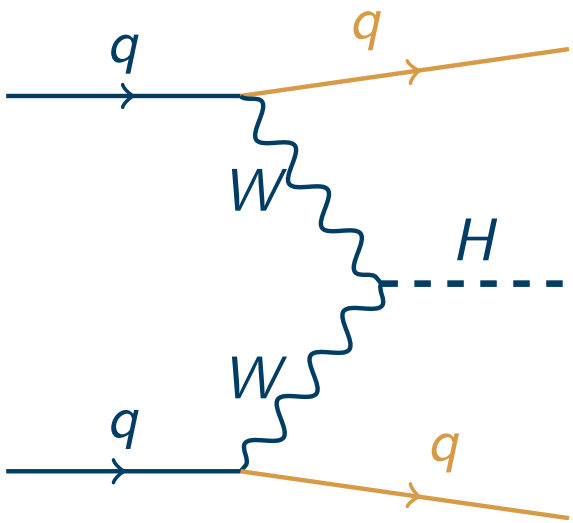
- Higgs analyses are divided in sub-categories, which are mostly based on jet multiplicities
- This improves sensitivity and helps to identify production modes, useful to measure couplings
- The 2-jet category is specific to the vector boson fusion (VBF) process
- VBF topologies could be used in this categorie to discriminate between signal and background

Vector boson fusion



- **Higgs** boson is produced in the central rapidity region
- Two energetic **quark jets** at large rapidities
- Jets are color connected with the proton remnant
- Expect no (or small) hadronic activity between the two tagging jets

Tagging jets

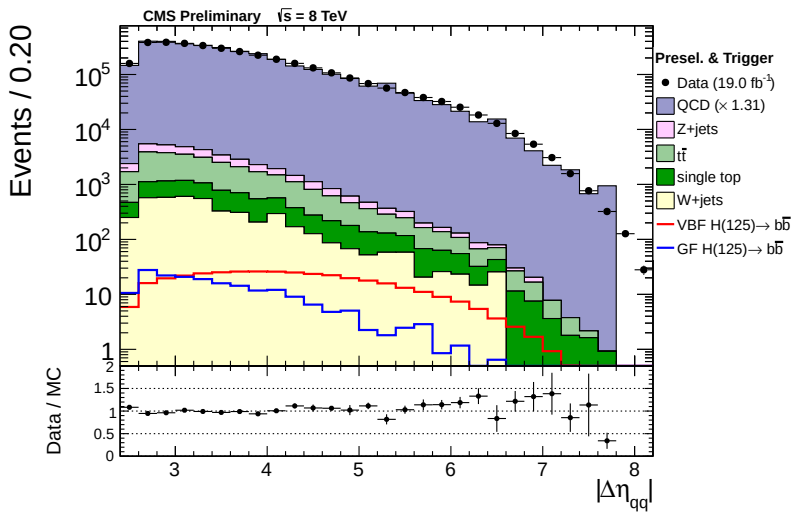


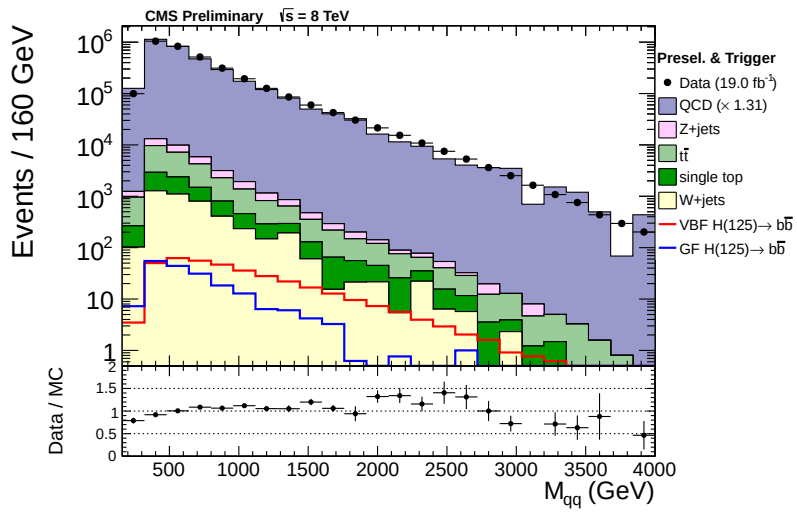
- The **VBF tagging jets** are found at large rapidities
- These jets originate from the hadronization of light quarks while the backgrounds will also contain jets originating from gluons

Tagging jets

The forward - backward jet signature is mostly exploited by

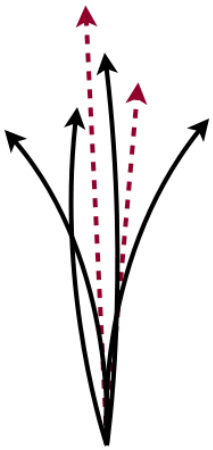
- a large rapidity separation
- a large invariant dijet mass



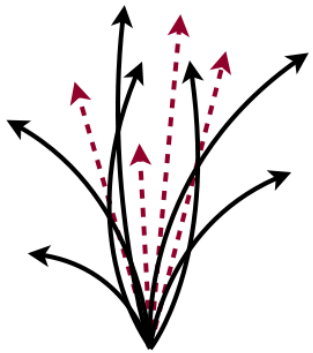


Quark - gluon jet discrimination

- Quarks and gluons have different color interaction
- This will mirror in different hadronization
- At a given energy a gluon jet will, on average:
  - have a higher multiplicity
  - be angularly wider
  - have a more uniform energy fragmentation



quark jet

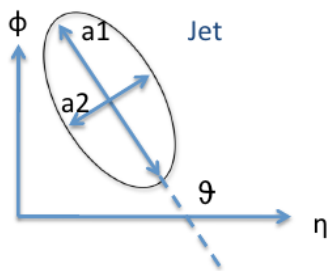


gluon jet

Quark - gluon jet discrimination

Likelihood based discrimination on properties of the jet constituents:

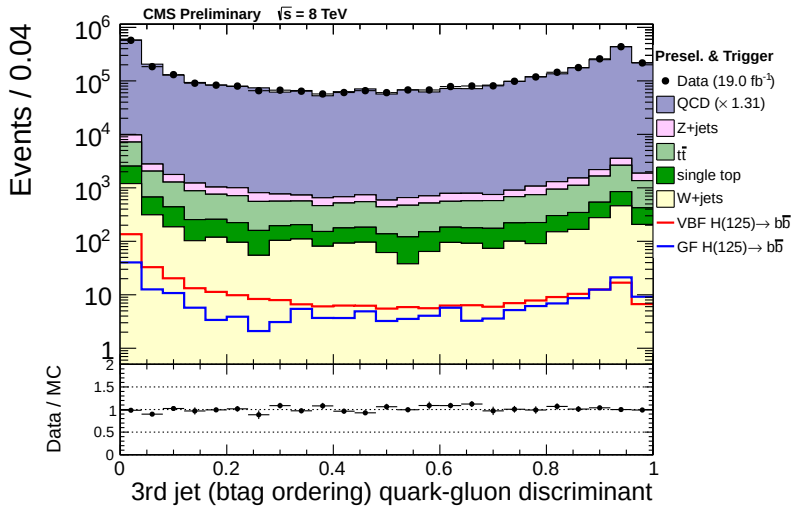
- Angular spread in the (η,ϕ) plane
  - major axis
  - minor axis



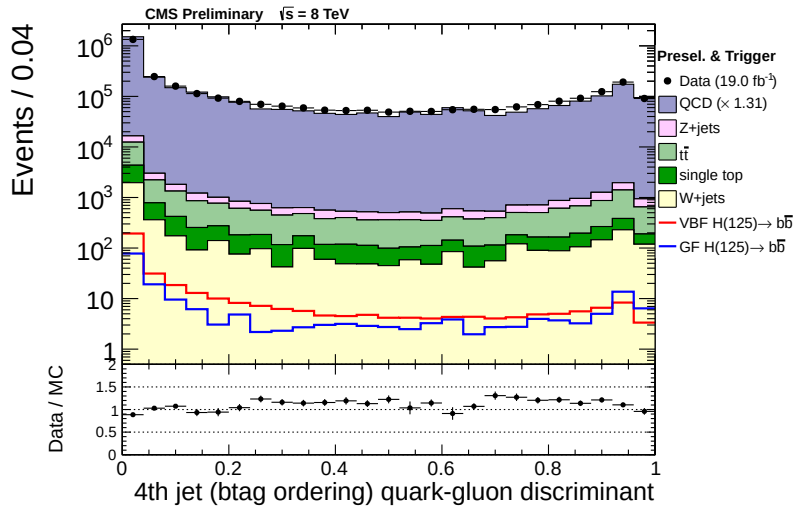
- Multiplicities
  - central region: charged multiplicity with PV association
  - transition and forward region: charged + neutral multiplicity ($p_T > 1$ GeV)
- Asymmetry
  - Jet Pull: $$t = \left| \sum_{i \in jet} \frac{{p_T^i}^2 \mid \vec{r}_i \mid}{p_T^{jet}} \vec{r}_i \right|$$
- Jet R: Energy fraction carried by the leading constituent $\frac{max(p_T^i)}{sum(p_T^i)}$ (only for charged in central region)

Quark - gluon jet discrimination

Already applied in the VBF $H \to b\bar{b}$ analysis:



quarks ⟷ gluons



quarks ⟷ gluons

These QG discrimination values were used as inputs to the neural network

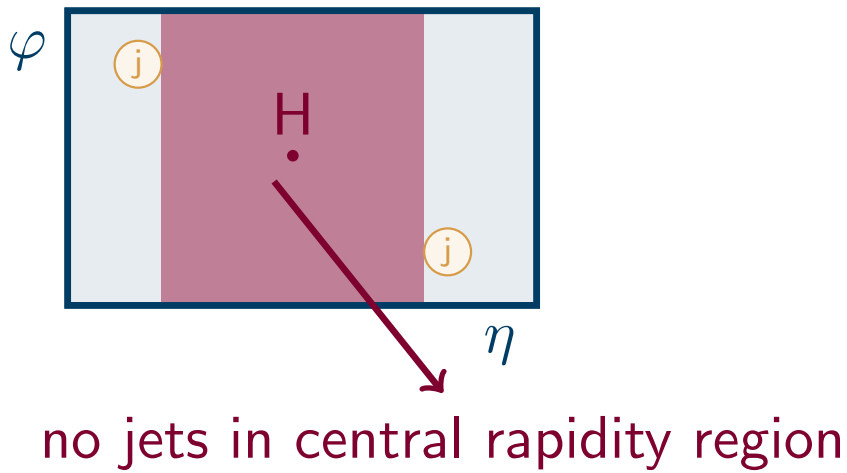
Slightly modified QG discriminator under construction for future analyses

Pile-up jet discrimination

- PU jets could cause events to migrate between the different jet categories
- The forward backward jets in the VBF selections suffer most from pile-up jet backgrounds
- Therefore we try to identify these pile-up jets
- Tradionally done by cuts on the most sensitive variables:
  - $\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R^2 p_T^2(i)}{\sum_i p_T^2(i)}$
  - $\beta^* = \frac{\sum \text{charged candidates} \in \text{other } PV}{\sum \text{charged candidates}}$
- More recently, a boosted decision tree is constructed using
  - Track/vertex variables
  - Shape variables
  - Multiplicities

Central jet veto

No (or small) hadronic activity in the rapidity region between the **tagging jets** $\Rightarrow$ Apply central jet veto



CMS uses the central jet veto of 30 GeV in:

- $H \to WW$
- $H \to \tau\tau$

Zeppenfeld variable

The topology of VBF events is also reflected in the Zeppenfeld variable, which places an observed pseudorapidity with respect to the tagging jets:

$$Z = \eta\,|_{\text{obs}} - \frac{\eta(j_1) + \eta(j_2)}{2} \tag{1}$$

The Zeppenfeld variable can be measured for

- the Higgs candidate, e.g. $\eta\,|_{\text{obs}} = \eta(H)$, the VBF signal will be more strongly peaked around 0 with respect to the backgrounds This is exploited in the VBF $H \to \gamma\gamma$ analysis with a cut on $Z < 2.6$
- a 3rd jet in the event, e.g. $\eta\,|_{\text{obs}} = \eta(3rd\ \text{jet})$, the VBF signal will show a double peak, contrasting with the peak around 0 in the backgrounds

$\Delta\phi_{H-jj}$

- The VBF $H \to \gamma\gamma$ search uses a cut on the azimutal angle between the dijet and diphoton systems:

$$\Delta\phi_{\gamma\gamma-jj} > 2.6$$

- If the event only contains the diphoton system and two jets, momentum conservation leads to $\Delta\phi_{\gamma\gamma-jj} \approx \pi$
- Thus, this variable is sensitive to remove events with additional radiation or PU jets

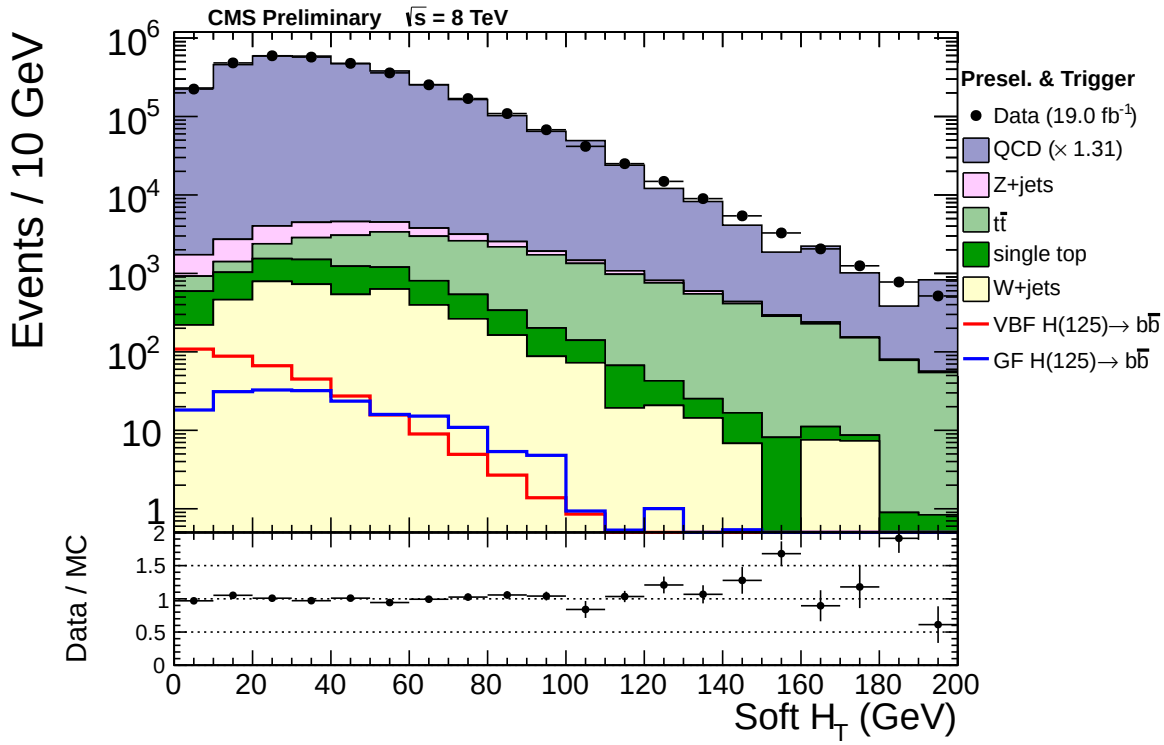
Soft hadronic activity

A way to measure central hadronic activity with track jets:

- Select charged high-purity tracks associated with the primary vertex
- Exclude tracks associated with the tagging jets and Higgs decay products
- Cluster these tracks into soft track jets with the anti-k_T algorithm
- Only keep track jets between the tagging jets

Soft hadronic activity

The central soft hadronic is measured in the VBF H $\rightarrow$ bb analysis:



The soft hadronic activity was also used as one of the inputs of the neural network to achieve maximal separation between signal and background

Summary

- CMS is using VBF topologies in analyses with the 2-jet category
- A quark-gluon jet discrimination is successfully applied
- Pile-up jets could be tagged and rejected from the event
- The low hadronic activity in VBF events is exploited by:
  - The central jet veto in $H \to WW$ and $H \to \tau\tau$
  - The $\Delta\phi_{\gamma\gamma-jj}$ variable in $H \to \gamma\gamma$
  - The soft hadronic activity in $H \to b\bar{b}$

Back-up

References

- CMS Collaboration, "Search for the standard model Higgs boson produced in vector boson fusion, and decaying to bottom quarks", CMS PAS HIG-13-011
- CMS Collaboration, "Updated measurements of the Higgs boson at 125 GeV in the two photon decay channel", CMS PAS HIG-13-001
- CMS Collaboration, "Search for the Standard-Model Higgs boson decaying to tau pairs in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV", CMS PAS HIG-13-004

PU jet ID variables

A boosted decision tree is constructed using

- Track/vertex variables:
  - $\beta = \frac{\sum \text{charged candidates} \in PV}{\sum \text{charged candidates}}$
  - $\beta^* = \frac{\sum \text{charged candidates} \in \text{other } PV}{\sum \text{charged candidates}}$
  - d_Z: distance in the z-axis of the highest p_T charged candidate wrt the PV
  - n_{vertices}
- Shape variables:
  - $\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R^2 p_T^2(i)}{\sum_i p_T^2(i)}$
  - $p_T^A(\Delta R) = \frac{1}{p_T} \sum_{i \in A < \Delta R < A+0.1} p_T^i$
  - $p_T^D = \frac{\sum_i p_T^2(i)}{p_T}$
- Multiplicities:
  - N_{charged}
  - N_{neutral}

CMS σ/σ_{SM} results