Search for $t\bar{t}H$ in the all-hadronic channel

Zürich PhD Seminar

08.03.18

L. Caminada, F. Canelli, S. Donato, D. Salerno, K. Schweiger

University of Zurich[UZH]

Why are we looking for $t\bar{t}H$

The Higgs boson and the Top-Higgs coupling

The Higgs Boson

Properties

- Gives mass to fermions and weak gauge bosons
  - Coupling to fermions $y_f \sim m_f$

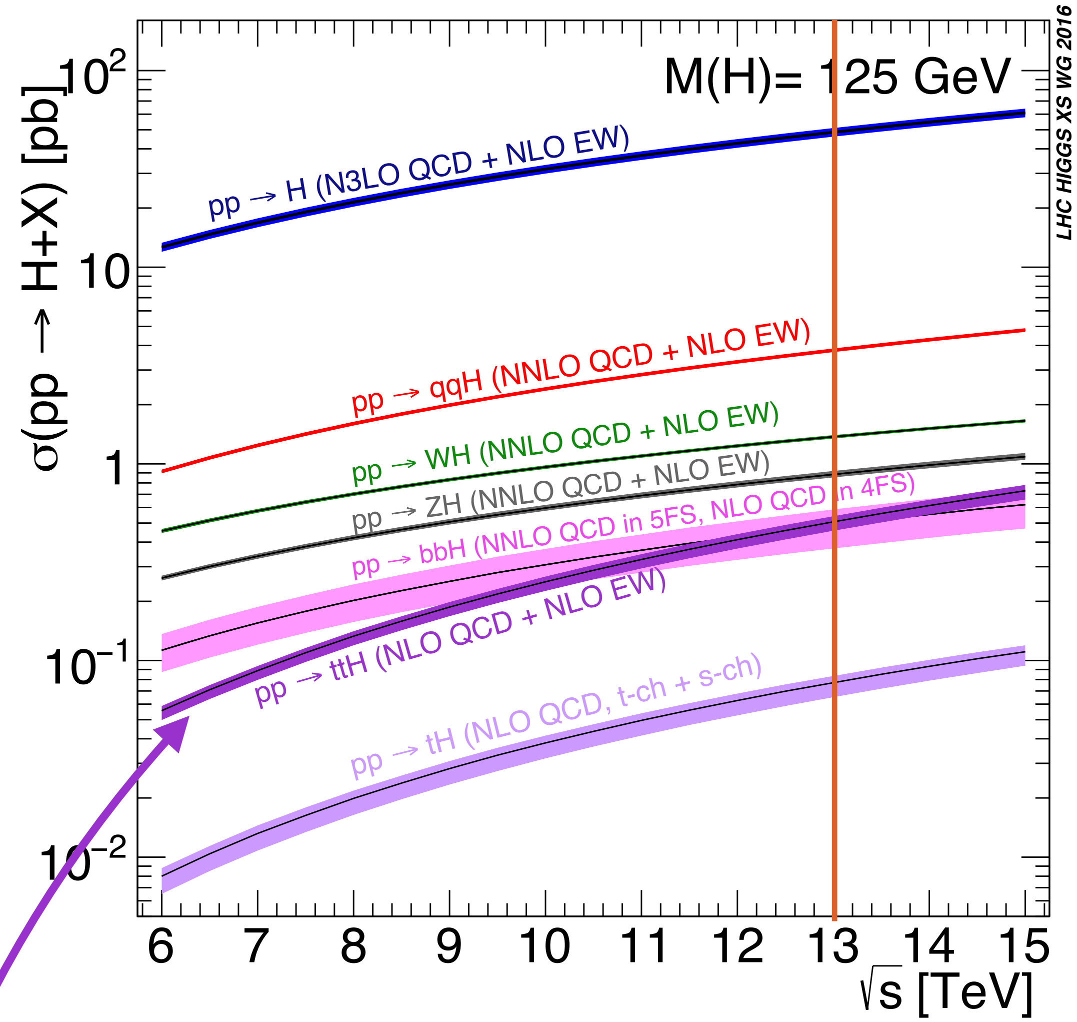
Measurements

- Mass 125,09 ± 0,24 GeV
- Spin and couplings consistent with standard Model (SM)

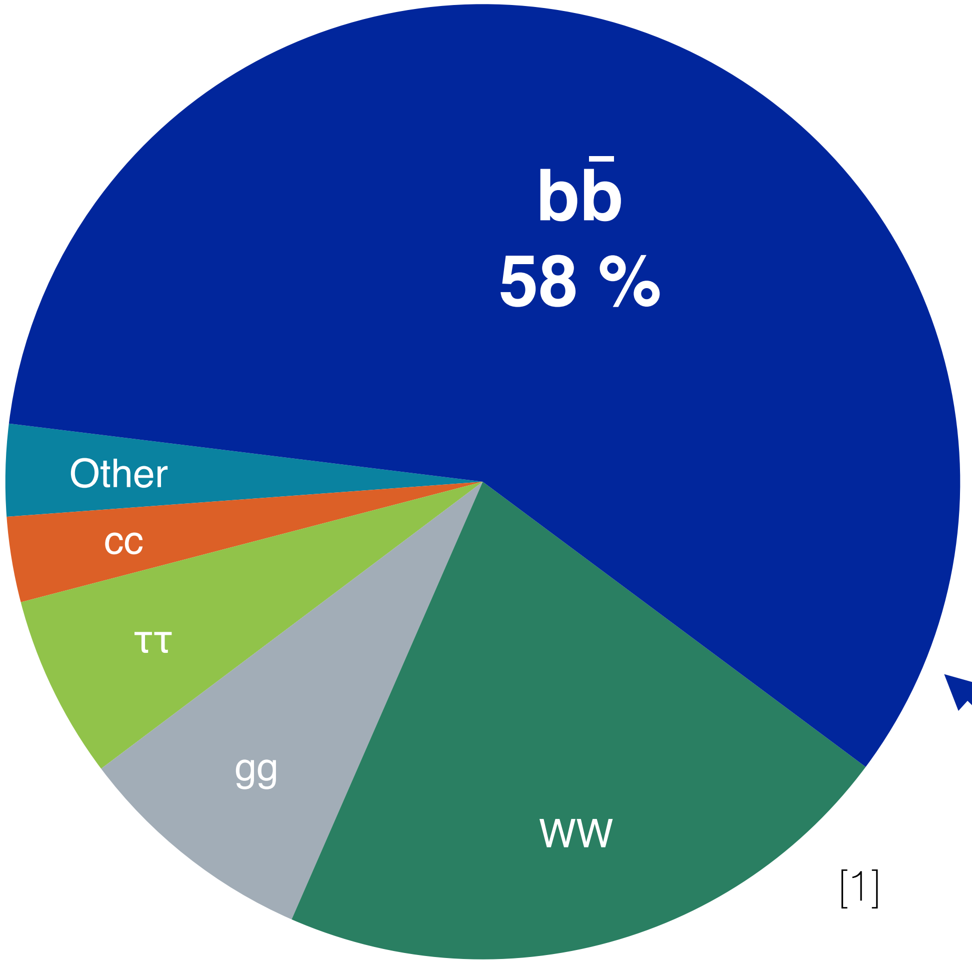
The Top-Higgs coupling

- Top quark is the heaviest SM particle
  - Strongest Yukawa coupling to Higgs
  - Indicator for BSM physics
- Measurement of top-Higgs coupling is crucial test of the SM
  - Only indirect knowledge

Higgs boson at the LHC - Choosing a channel



Higgs boson branching fractions for m_H = 125 GeV



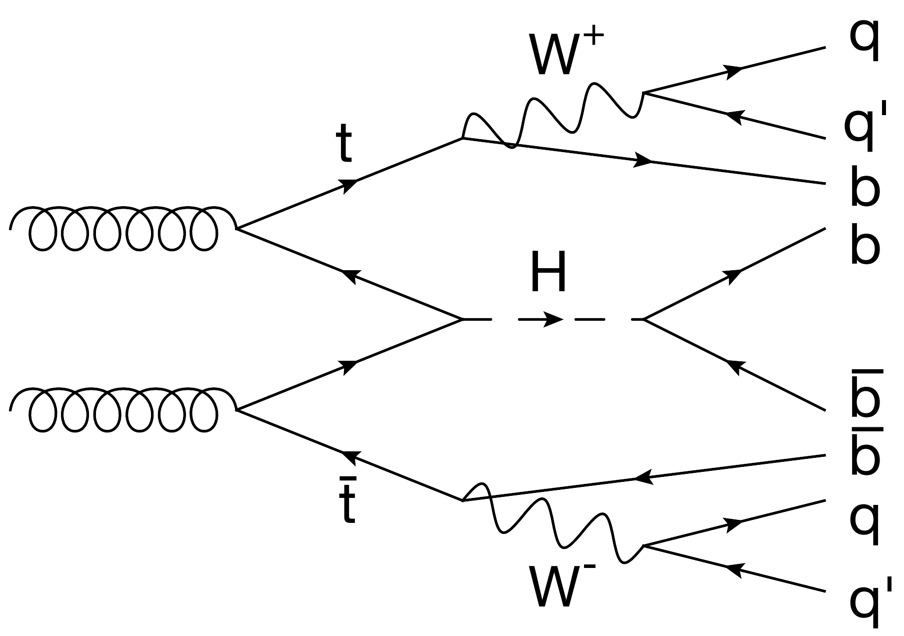
Process	σ [pb]
Glu-Glu fusion	44
Vector boson fusion	3,7
Associated production	1,4
tt̄H	0,51

tt̄H has small cross section

H→bb̄ has largest branching fraction

$t\bar{t}H$ production

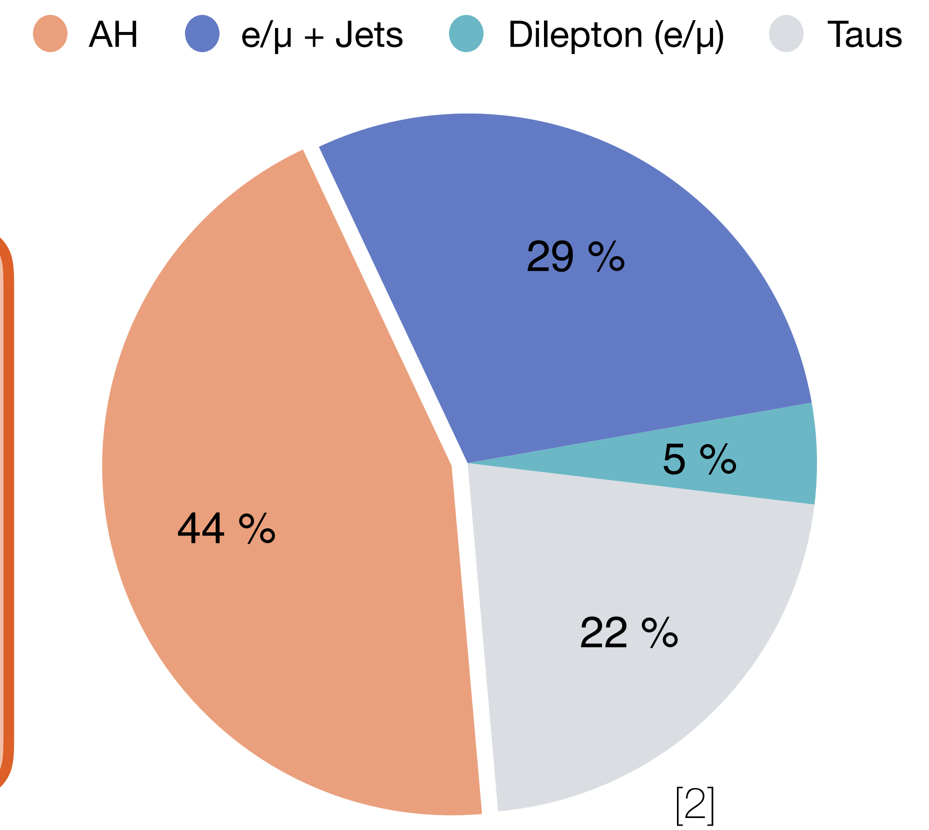
- Last remaining main production of the Higgs boson not yet experimentally observed
- Direct measurement of the Top-Higgs Yukawa coupling → crucial test of the SM
- $H \to b\bar{b}$ has largest branching fraction but is challenging
- Multiple channels depending on top-quark pair decay



The all hadronic (AH) channel

- Largest fraction of all top-quark decay channels
- Fully reconstructed final state
- Very challenging selection because of all jet signature

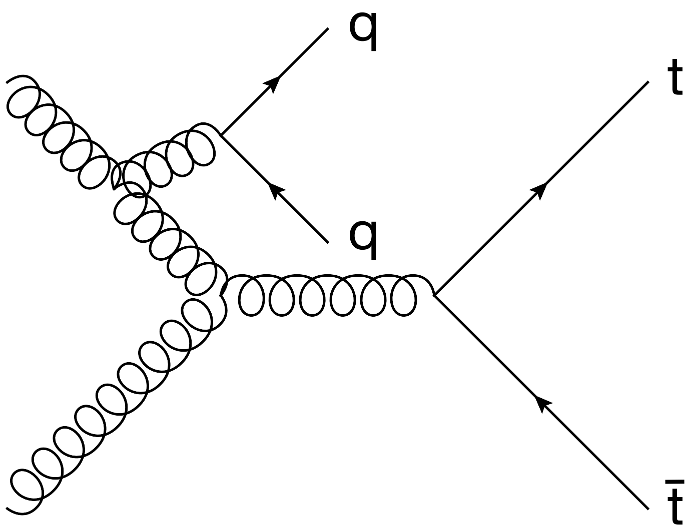
Top-quark pair decay channels



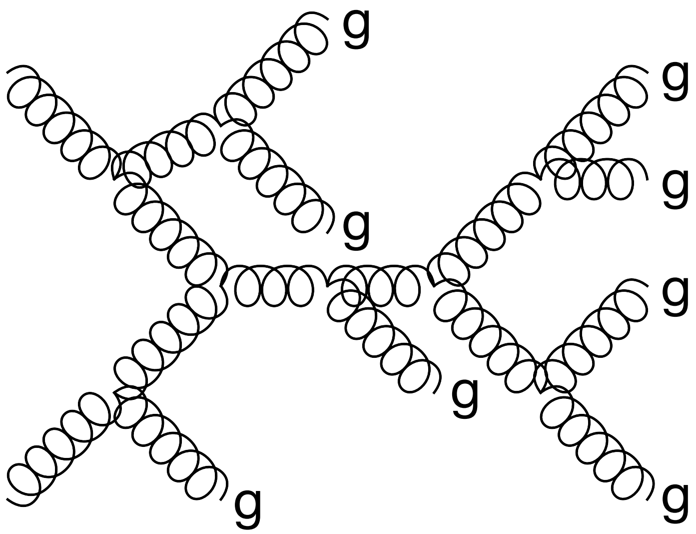
[2]

$t\bar{t}H(b\bar{b})$ backgrounds

- Top-quark pair backgrounds
  - $t\bar{t}$+jets
    - Similar final state → Less b-Jets
    - Reducible background
  - $t\bar{t}$+$b\bar{b}$
    - Same final state
    - Irreducible background

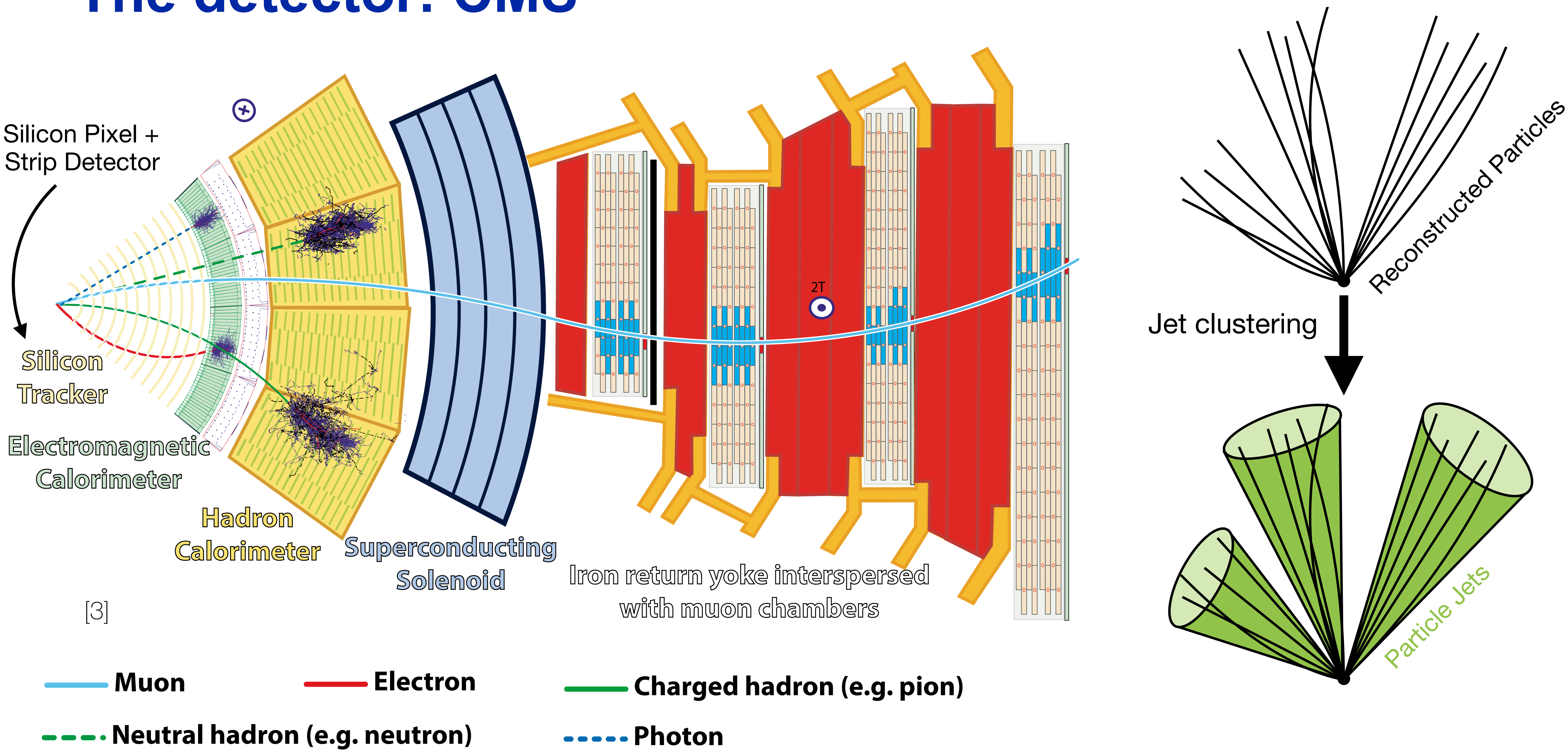


- QCD multijet background
  - Final state different
  - Much more events → some events have similar signatures



How are we searching for $t\bar{t}H$

The detector: CMS

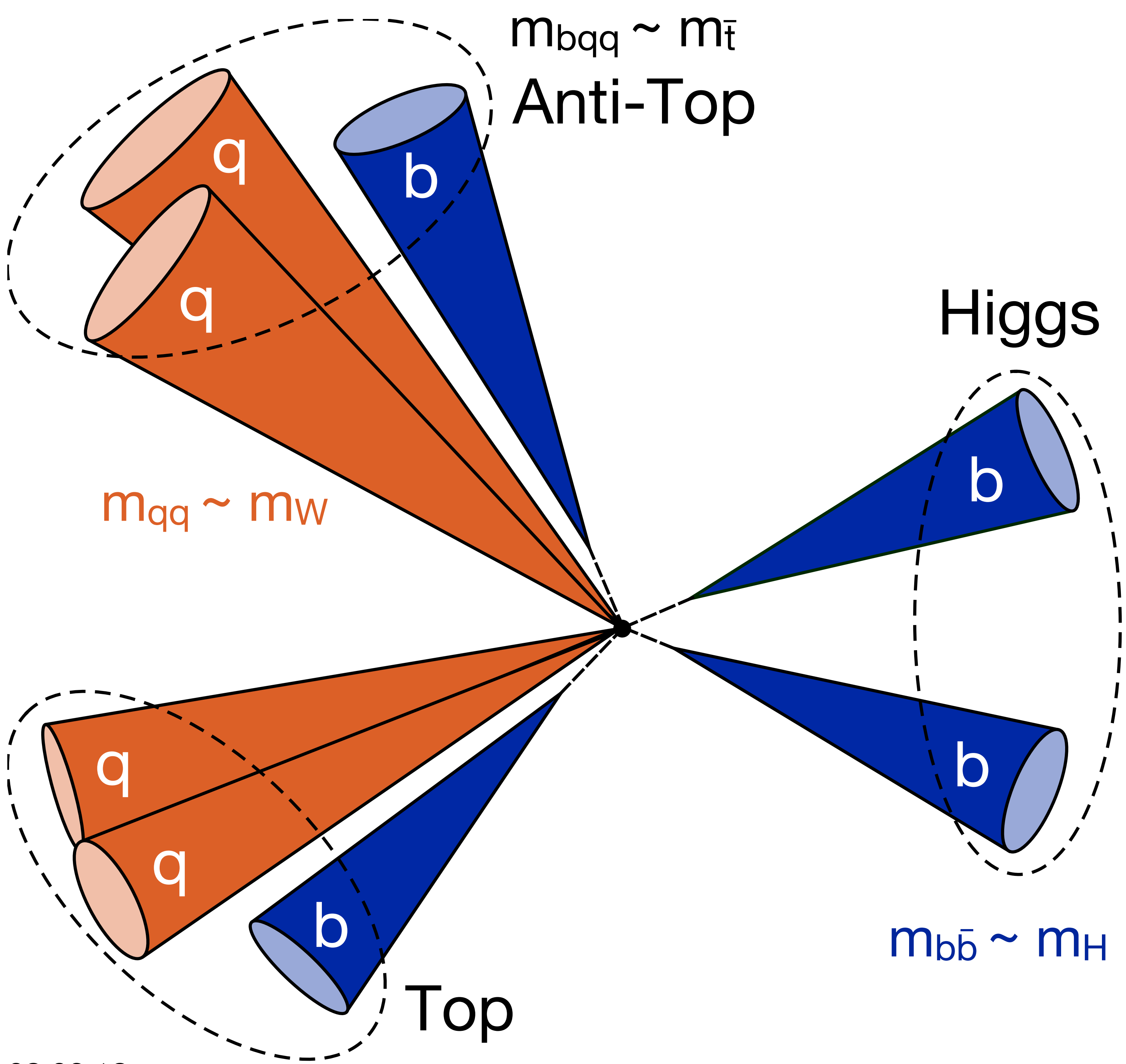


- Recorded interactions with the detector are reconstructed to particle tracks and energies
- Algorithms are used to cluster reconstructed particles to jets

What are the challenges?

Challenges of the $t\bar{t}H(b\bar{b})$ AH final state

Signature: 8 jets: 4 b-jets and 4 light jets

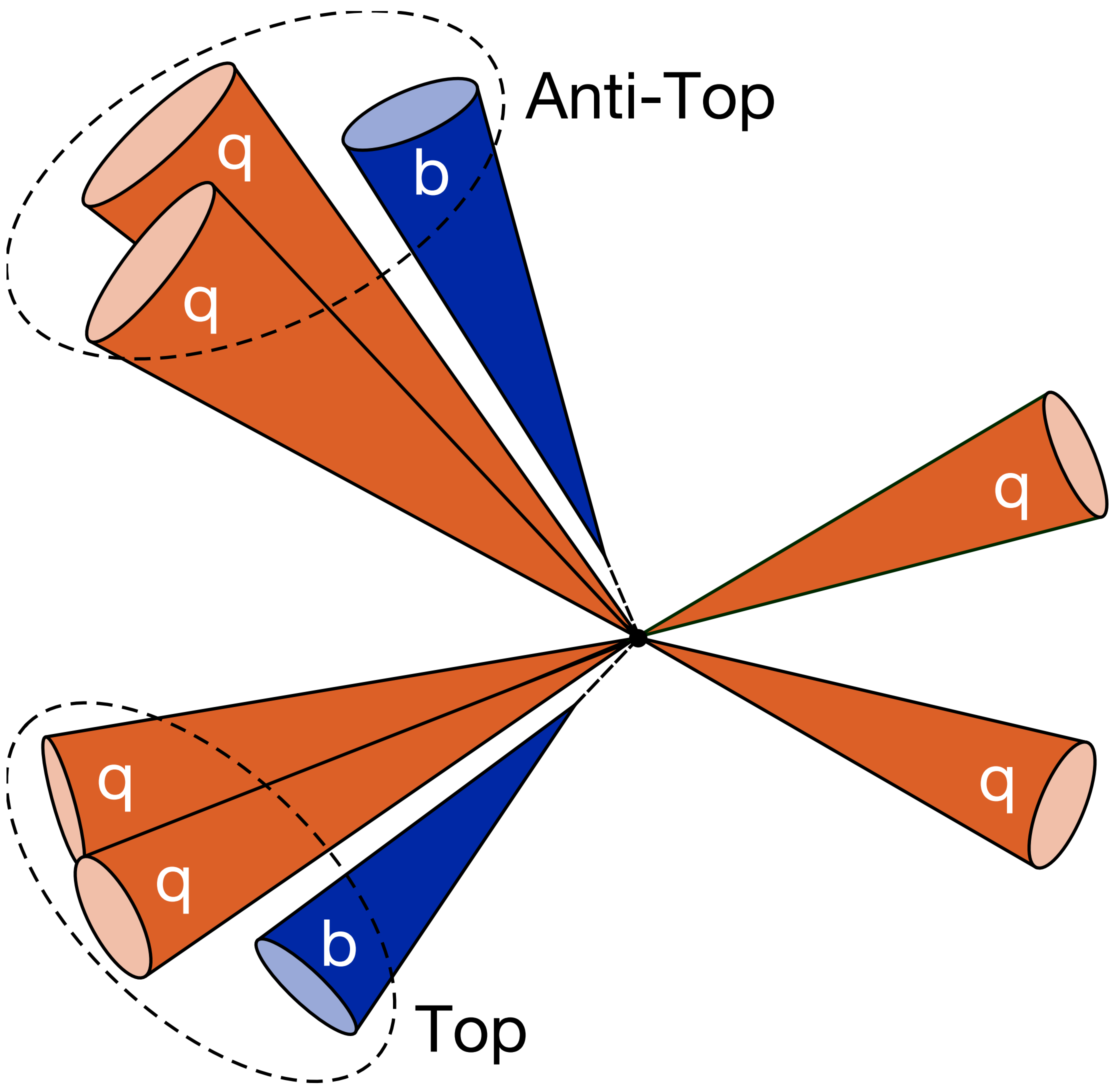


- Large QCD multijet background
- Tough selection at data taking
- Irreducible $t\bar{t} + b\bar{b}$ background
- Additional jets from radiation
- Matching jets to quarks is not easy

Identifying event properties

$t\bar{t}$ + Jets

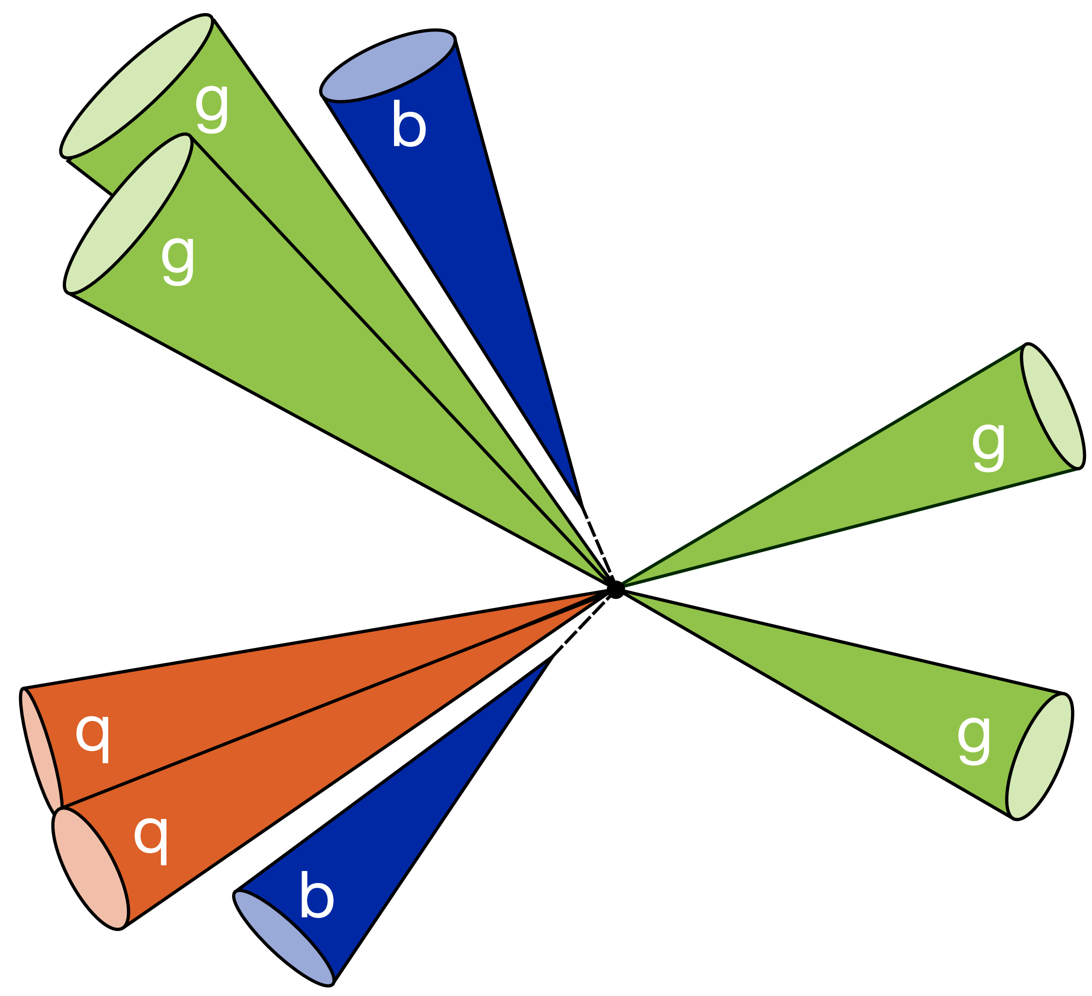
Signature: 8 jets: 2 b-jets and 6 light jets



Less b-jets → Can we identify them?

Multijet final state

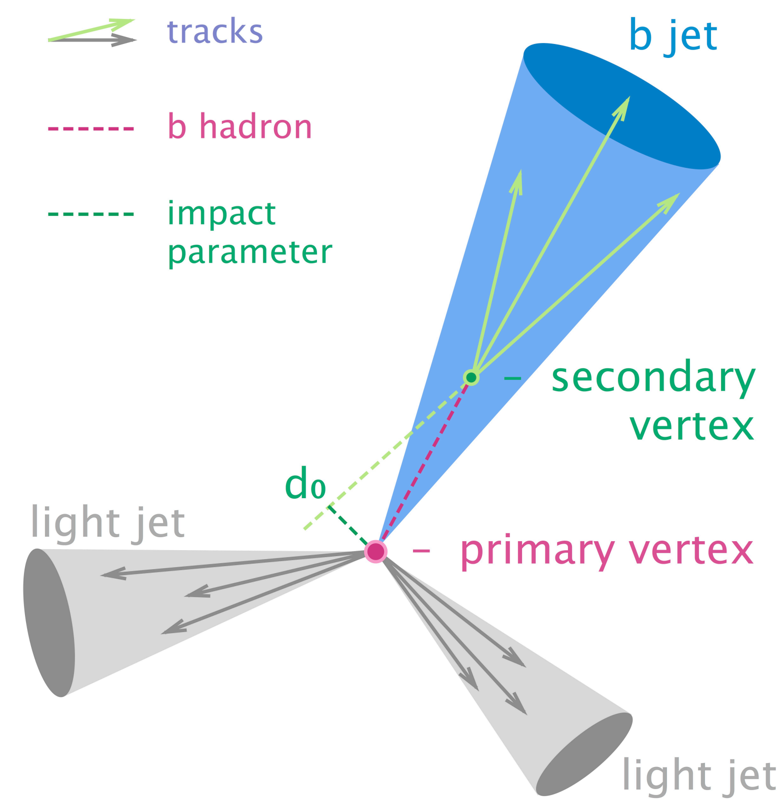
Signature: Variable number of jets: mostly gluon jets

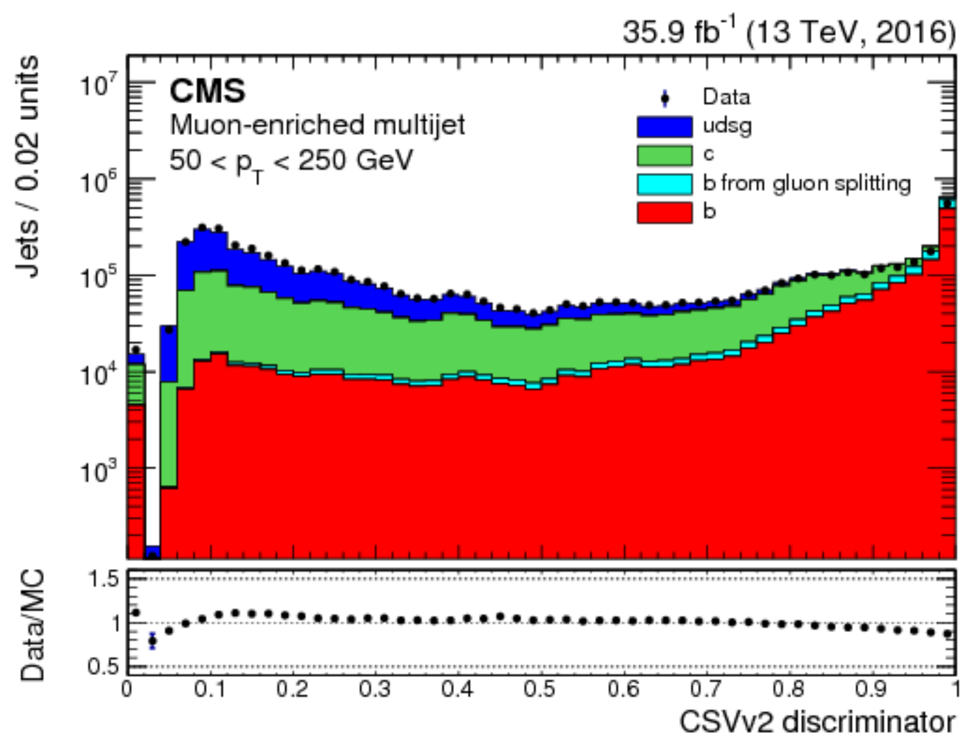


Many gluon jets → Can we identify them?

Identifying jets originating from B hadrons

- B hadrons have relatively long lifetime → secondary vertices
- Identify b-jets based on kinematic information and vertex information from the pixel detector
  - Implemented as neural network (Combined Secondary Vertex - CSV)
- Trade off between efficiency and purity
  - Medium Working point:
    - 70% b jet efficiency

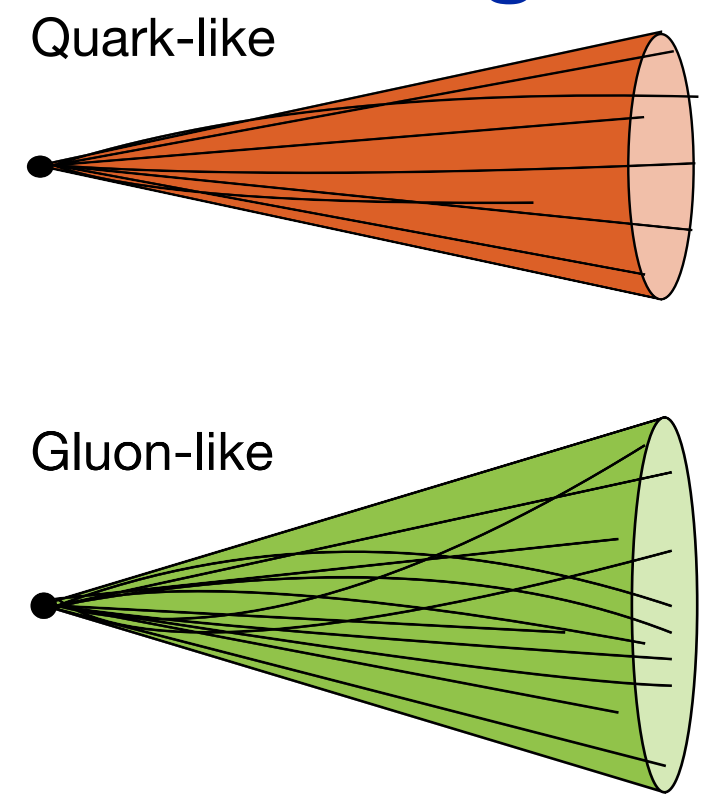


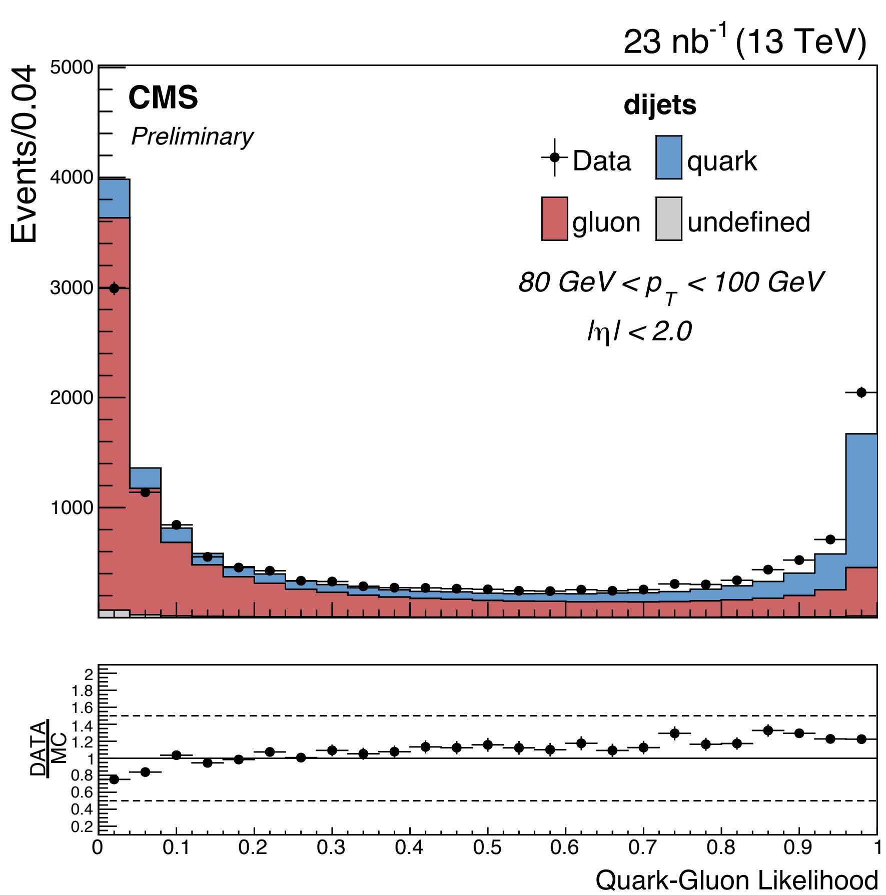


[4]

Distinguish jets originating from quarks and gluons

- Discriminate between jets from quarks and gluons by jet composition, e.g.
  - Number of constituents
  - Spatial collimation
- Quark-Gluon likelihood
  - Optimized to discriminate light quark (u,d,s) jets from gluon jets
- Use as event based discriminator by calculating a likelihood ratio for certain hypotheses (QGLR)
  - e.g. 4q,0g vs 0q,4g





[5]

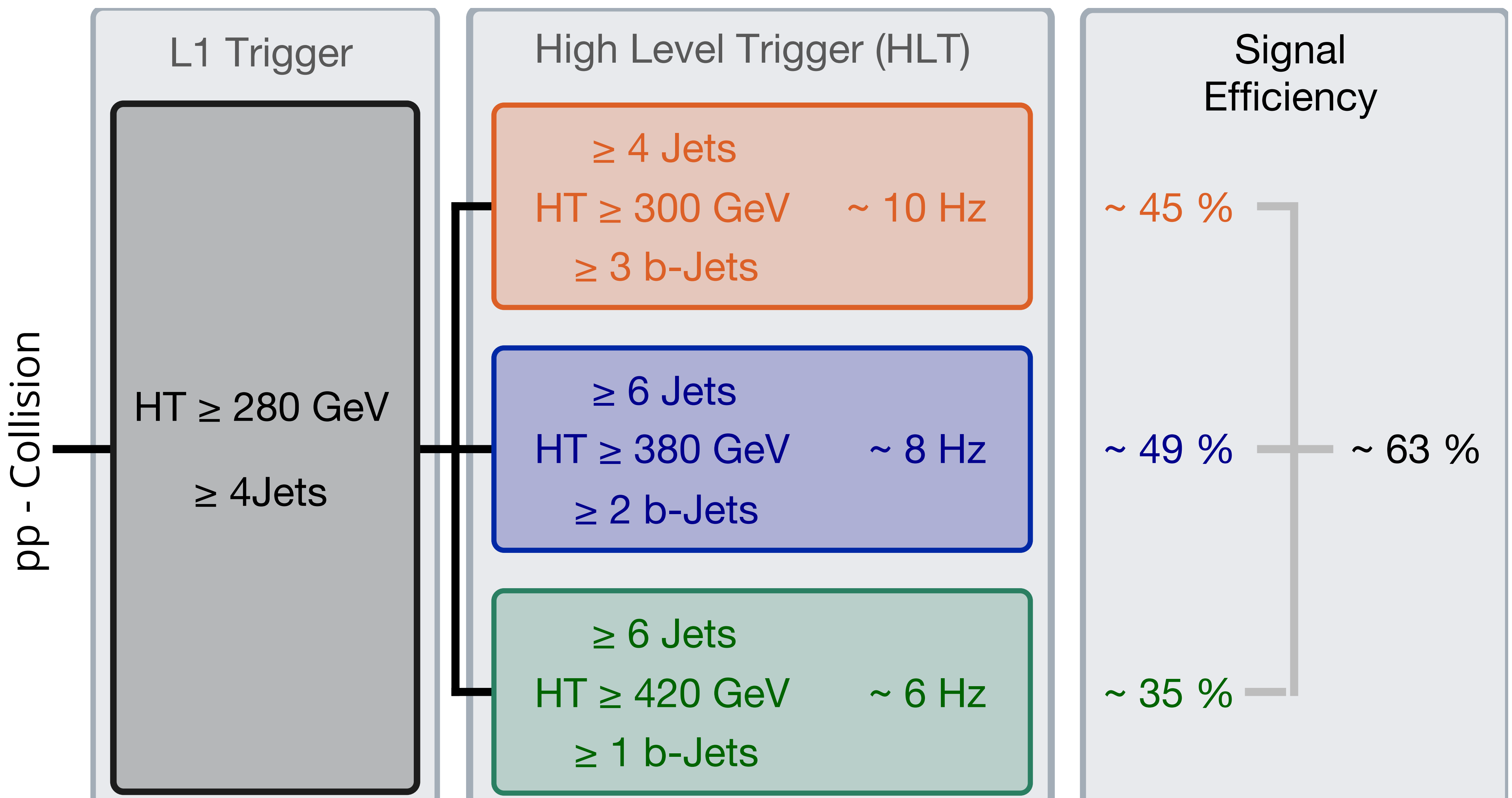
Selecting signal events

Collecting data

- Because of small $\sigma_{t\bar{t}H}$, selecting events with high efficiency at data taking is very important
  - In 2016 ~ $2x10^{15}$ pp collisions (→ ~ 18 000 $t\bar{t}H$ events)
  - Not all of these events can be recorded!
- In the LHC collisions occur every 25 ns → 40 MHz
- CMS is technically limited to record events at 1 kHz
- This reduction is achieved with a 2-stage trigger system (online)
  - Level 1-Trigger (40 MHz → 100 kHz)
  - High-Level-Trigger (100 kHz → 1 kHz)

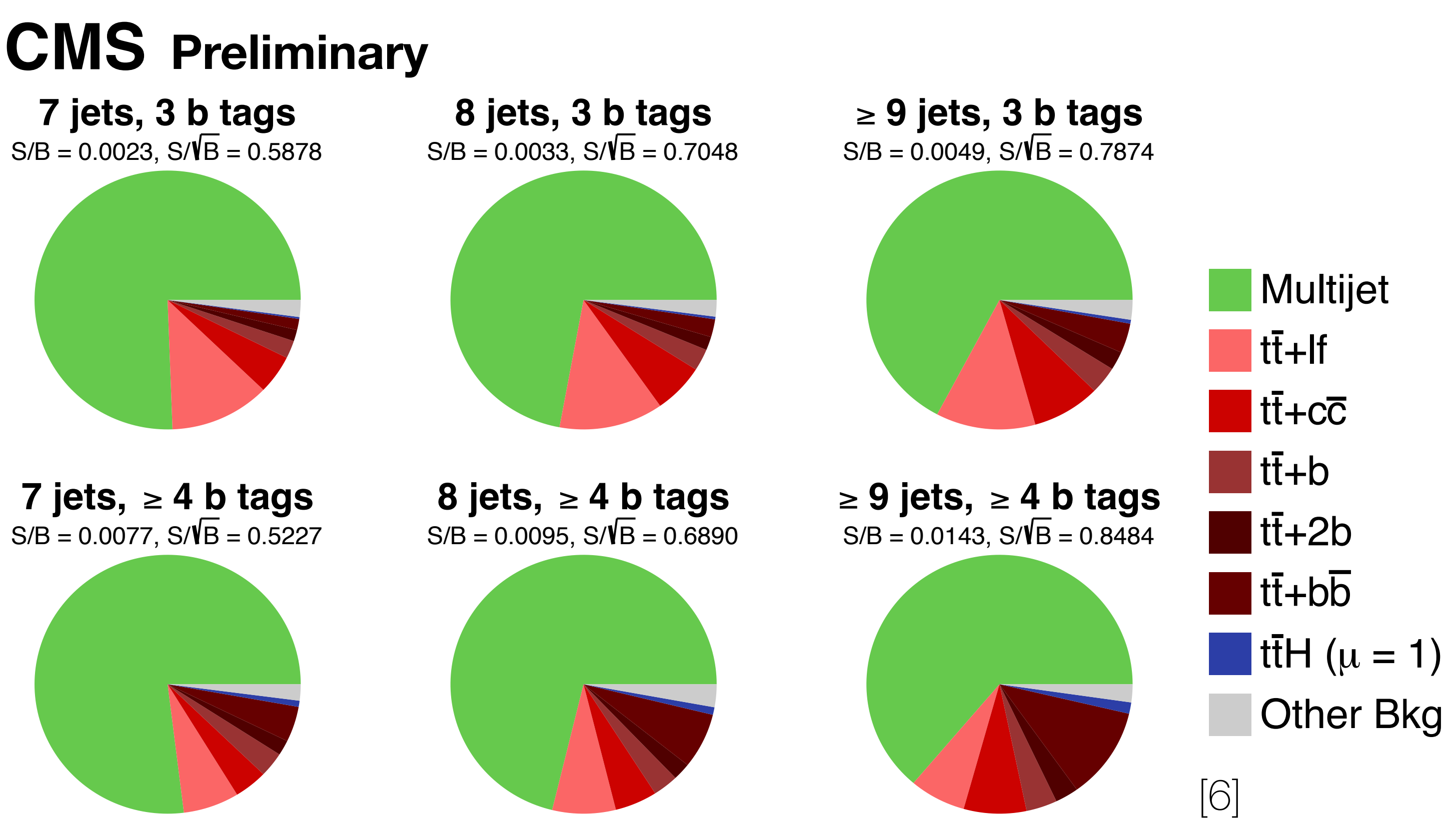
Trigger strategy for $t\bar{t}H(b\bar{b})$ AH

- Designing a trigger strategy is important to collect as many signal event as possible
  - All analyses share the same Trigger system → Trade off between efficiency and trigger rate
- Multiple dedicated Triggers for $t\bar{t}+b\bar{b}$ and $t\bar{t}H(\bar{b}b)$ are used together to improve signal efficiency



Event selection and categorization

- Select events to increase signal purity and define control regions
- Events are categorized by jet and b-tag multiplicity to increase overall sensitivity → Categories differ by Signal to Background ratio (S/B)
  - High S/B → Drive signal sensitivity
  - Low S/B → Help to unterstand background



[6]

Trigger

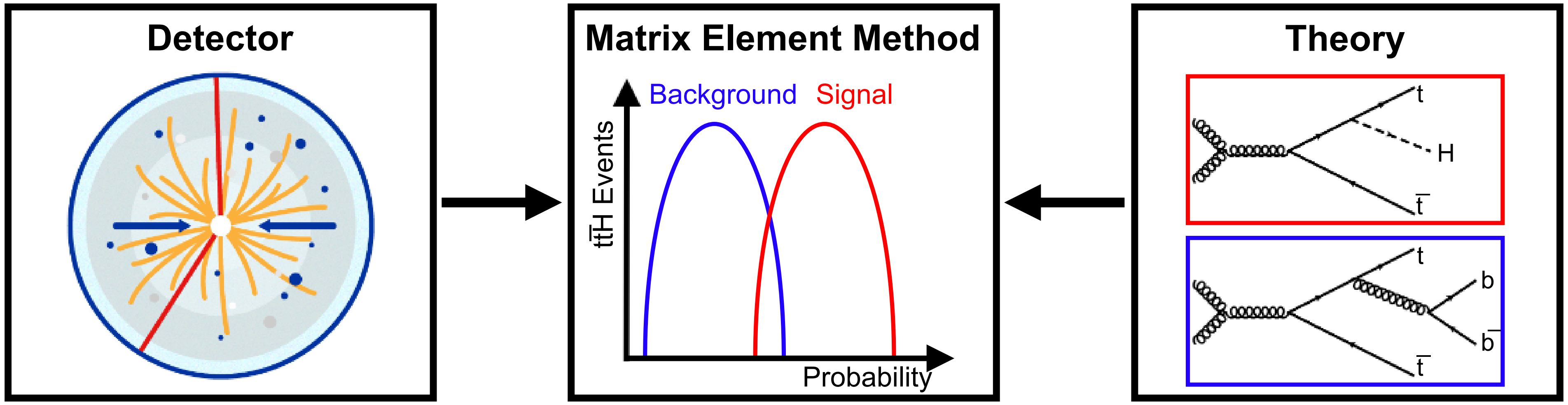
↓

Selection
- 6 Jets, pT > 40 GeV
- HT > 500 GeV
- 2 b-tagged jets
- No leptons (e/μ)
- QGLR > 0,5
- W mass cut

↕ ↕

Process	Events	Efficiency
tt̄H AH (Signal)	$5x10^2$	10 %
tt̄ + jets	$4x10^4$	0,1 %
tt̄ + bb̄	$5x10^3$	3,3 %
QCD multijet	$1x10^5$	0,001 %

Matrix Element Method



- Powerful tool to separate signal and background
  - Combining theory knowledge from the standard model and detector information
- Sum over all combinations of quark-jet matching
- Considers $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ hypotheses (works also well against QCD)
- Calculation of the probability density ω for each event to originate from $t\bar{t}H$ (S) or $t\bar{t}+b\bar{b}$ (B)

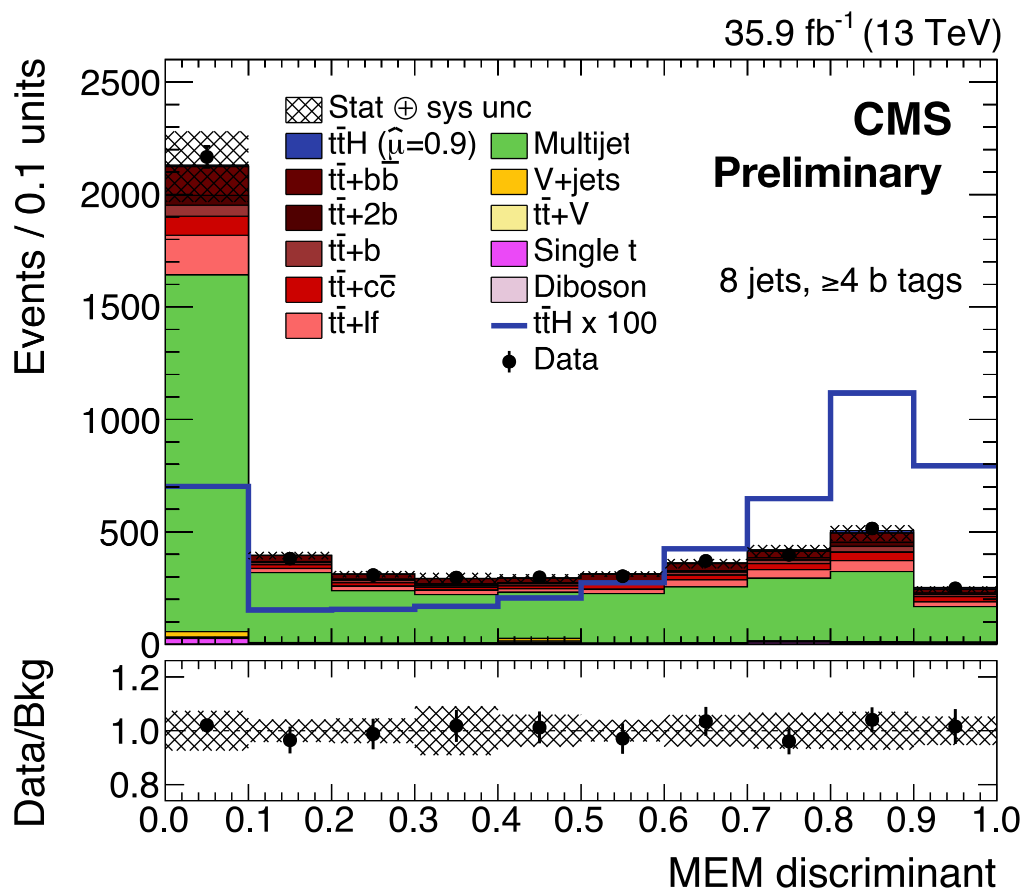
Measuring $t\bar{t}H$

Final discriminant

- The MEM is evaluated under the signal and background hypothesis for each event → A discriminant is formed:

$$\text{Final Discriminant: } P_{sb} = \frac{\omega_S}{\omega_S + \kappa_{sb} \cdot \omega_B}$$

- Distribution between 0 and 1 → different shapes for tt̄H and background events!



[6]

Statistical treatment and uncertainties

Statistical treatment

- Signal strength modifier: $\mu = \sigma/\sigma SM$

- Simultaneous fit in all 6 analysis categories under signal+background and background-only hypothesis (Binned Maximum Likelihood Fit)

    - Low sensitivity→ calculation of upper limit on μ

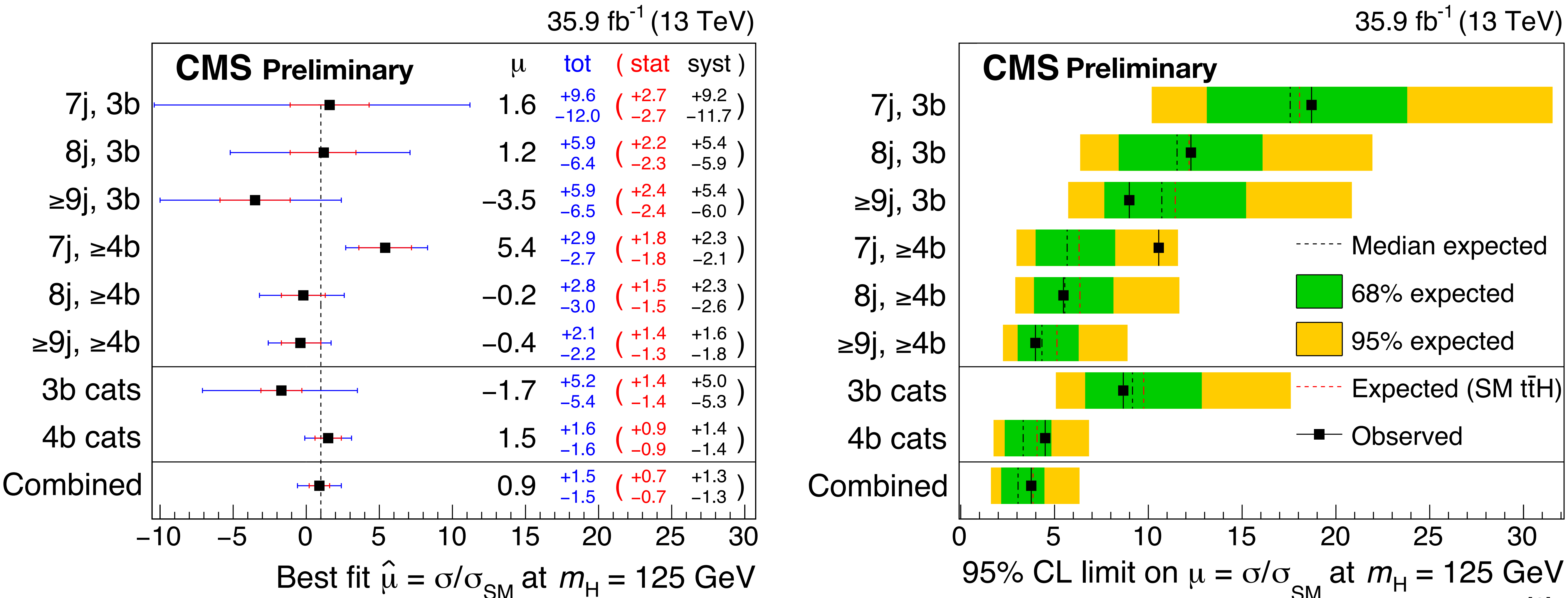
Uncertainties

- Signal and background predictions are not perfect! → Given as $\pm 1\sigma$ range

- Experimental uncertainties, e.g.
    - Corrections dealing with difference between data and simulation
    - Trigger efficiencies

- Theoretical uncertainties, e.g.
    - Cross sections

Latest results with data recorded in 2016



[6]

- Results consistent with SM expectations (μ = 1)
- Also consistent with background-only hypothesis → Observed exclusion limit μ < 3.8
- Analysis limited by systematic uncertainties → Only adding data will not be enough to reduce them

Outlook

Combination

- Currently all CMS $t\bar{t}H(b\bar{b})$ searches are being combined
- All $t\bar{t}H(b\bar{b})$ also part of the overall $t\bar{t}H$ measurements

Background rejection

- Improvement necessary → QCD events dominating in all categories
  - DNN/BDT for early rejection
  - Representative process in MEM

Systematic uncertainties

- dominating uncertainties in current analysis → Improvements e.g. for:
  - $t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$ cross section are poorly known

Adding boosted region

- Special treatment of high p_T events
  - Multiple jets can merge to one single „FatJet“
  - Algorithm to investigate such jets can improve sensitivity

Conclusion

- $t\bar{t}H$ is a **direct** probe of the Top-Higgs Yukawa coupling
- The all hadronic channel is an important contribution to the $t\bar{t}H$ sensitivity
- Challenging final state exploring many different techniques
  - Dedicated Triggers
  - Selection utilizing methods like b-tagging or Quark-Gluon likelihood
  - Matrix Element Method used for calculating a final discriminant
- This analysis was conducted and published for the first time at CMS and first time ever with 13 TeV at UZH
  - ▸ This opened a new channel for future development

References

[1] CERN Yellow Report

[2] **C. Patrignani et al. (Particle Data Group)**, Chin. Phys. C, 40, 100001 (2016) and 2017 update

[3] CMS-PHO-GEN-2016-001

[4] arXiv:1712.07158

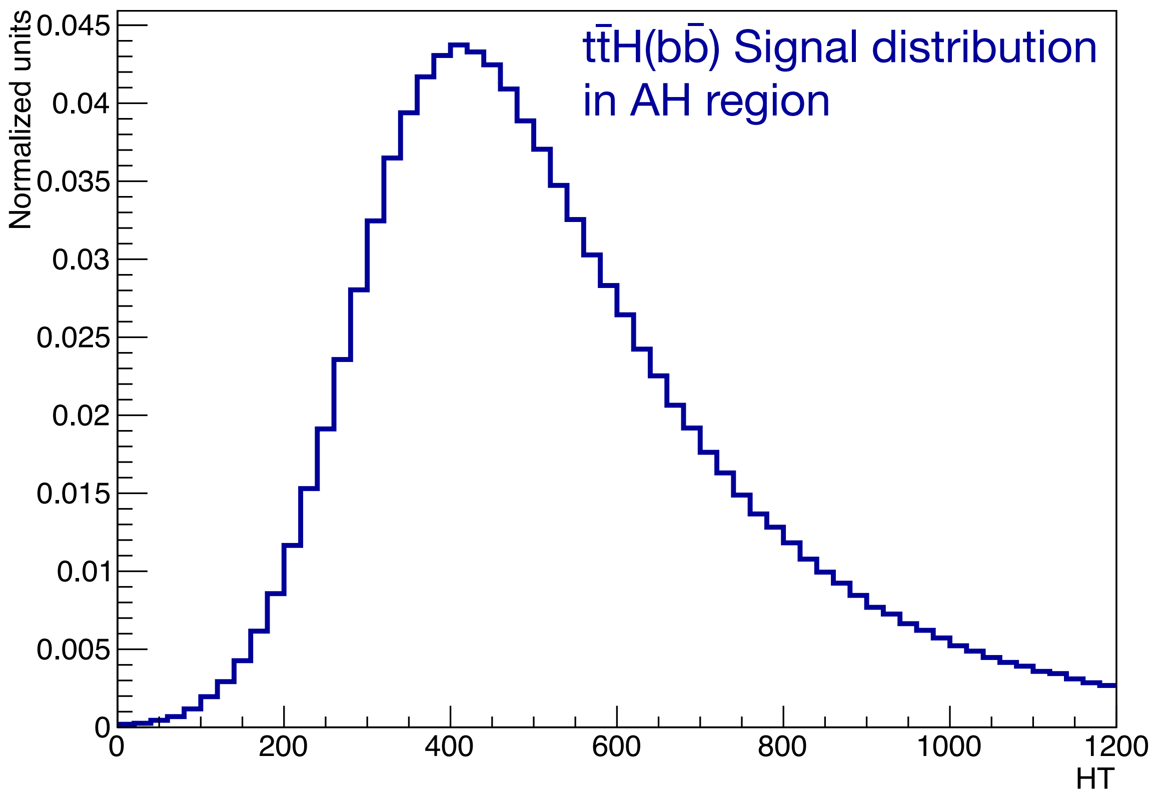
[5] CMS-DP-2016-070

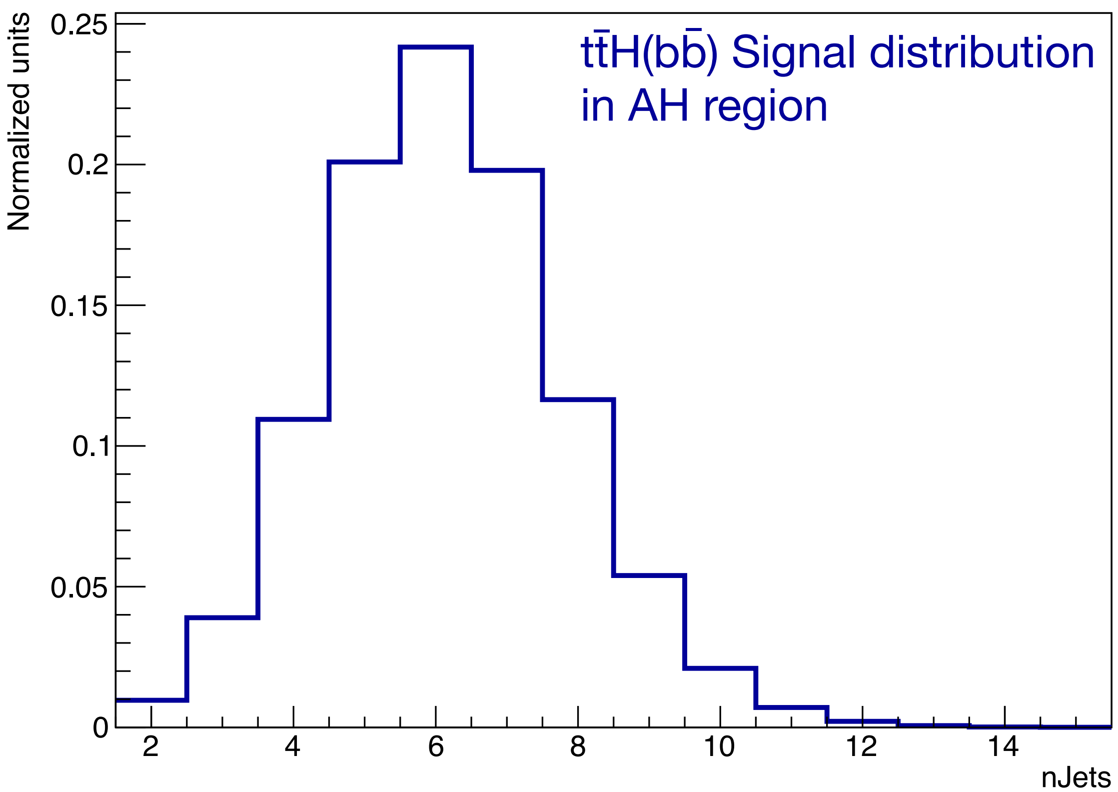
[6] CMS HIG-17-022

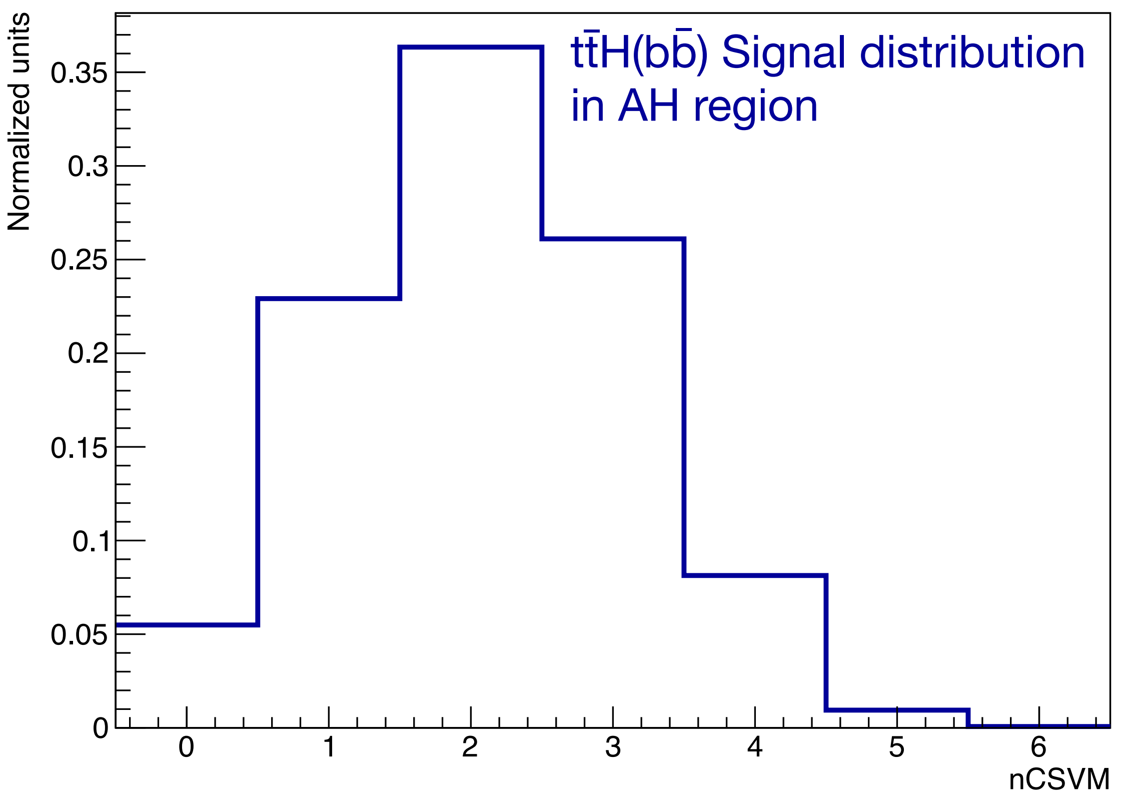
Backup

High Level Trigger for $t\bar{t}H(b\bar{b})$

- Designing a trigger strategy is important to collect as many signal event as possible
  - All analyses share the same Trigger system → must limit rate
  - Trade off between efficiency and trigger rate
- All-jet signature makes it difficult to use generic triggers → Usage of special triggers/ combinations



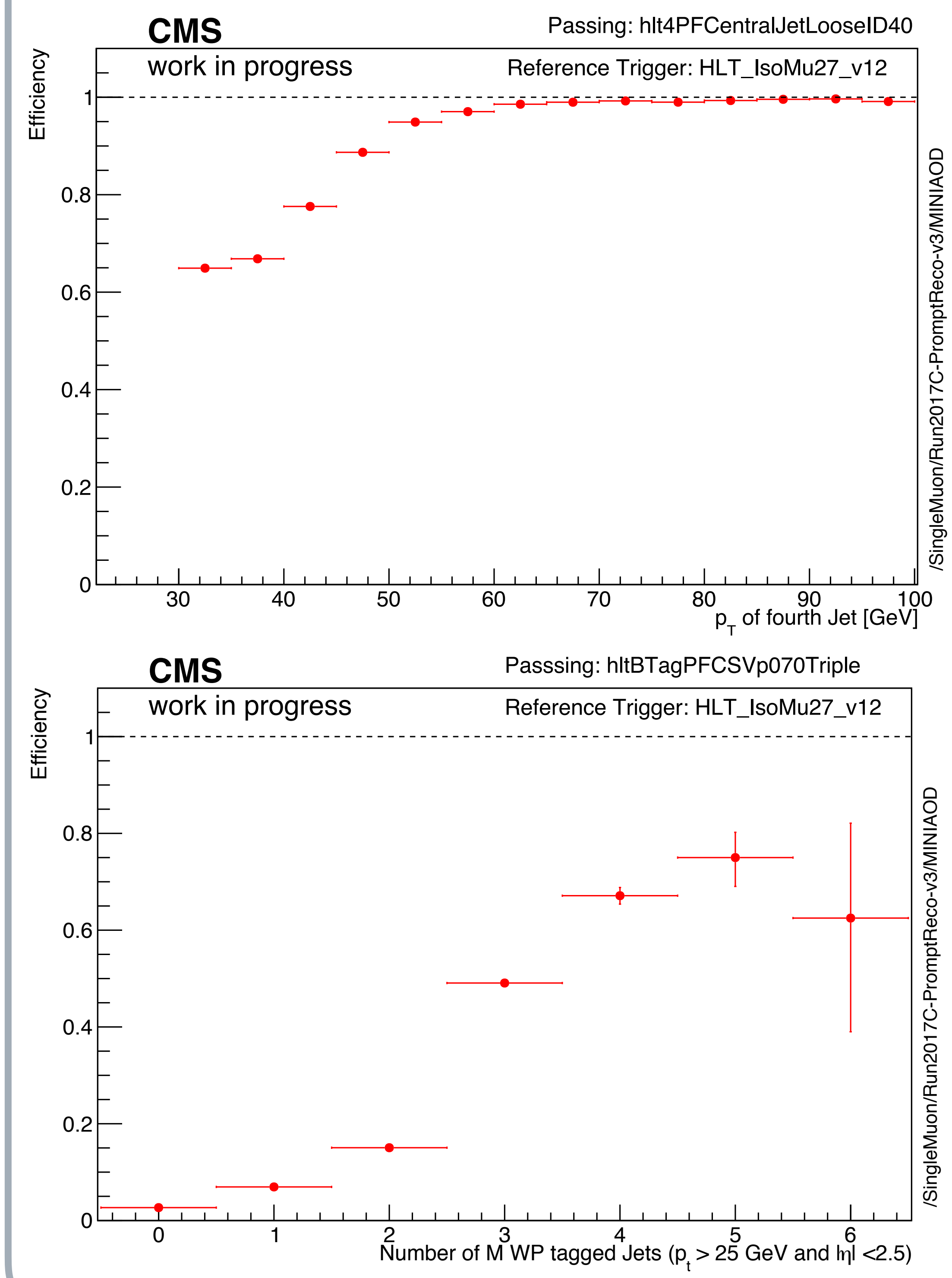




Trigger Performance

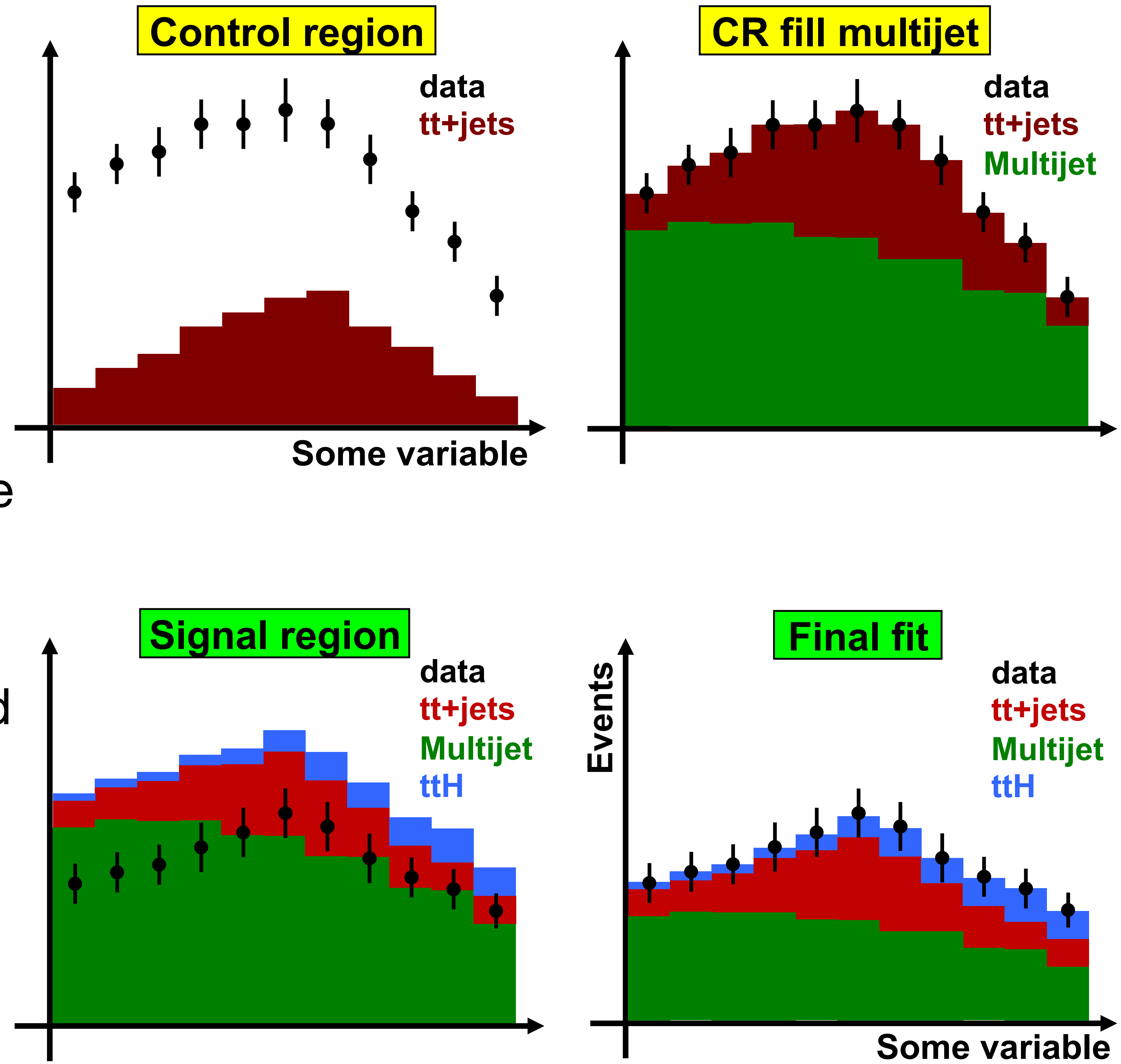
- High level triggers are build as sequence of modules cutting on single variables
  - Performance between modules important
- Analysis (offline) selection cuts dependent on trigger efficiency
- Online and offline algorithms differ
  - monitoring trigger performance with respect to offline variables

Trigger:
≥ 4 Jets, ≥ 3 b Jets, HT > 300



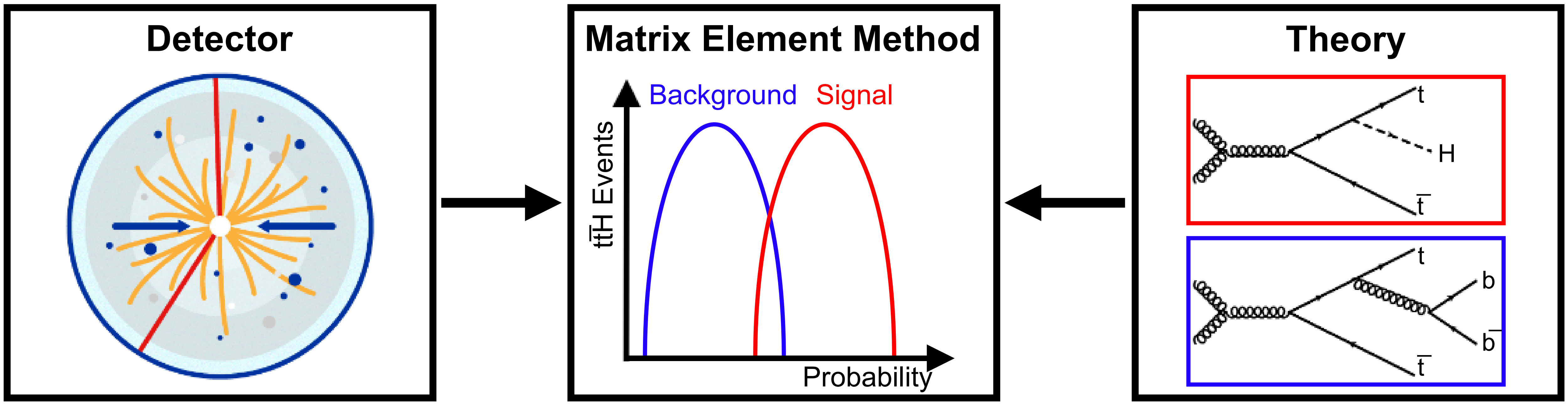
Data driven background estimation

- QCD background is very large, complex and **poorly understood**
  - MC simulation too little and imperfect
  - Estimation from data
- Control regions: Estimate the shape of the QCD multijet background
- Normalization of background in Signal region determined by the final fit
- Has to be done separately for each category and distribution



The Matrix Element Method - Introduction

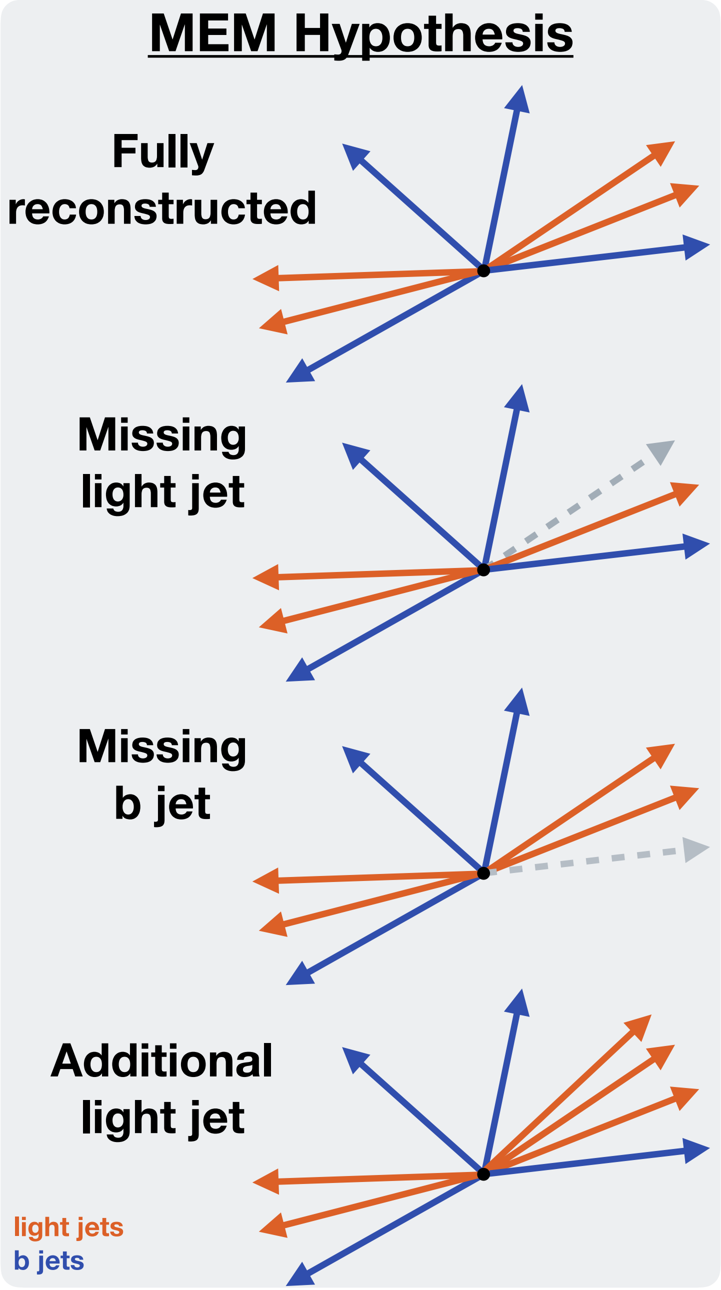
- Powerful tool to separate signal and background
  - Combining theory knowledge from the standard model and detector information
- Sum over all combinations of quark-jet matching
- Considers $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ hypotheses (works also well against QCD)



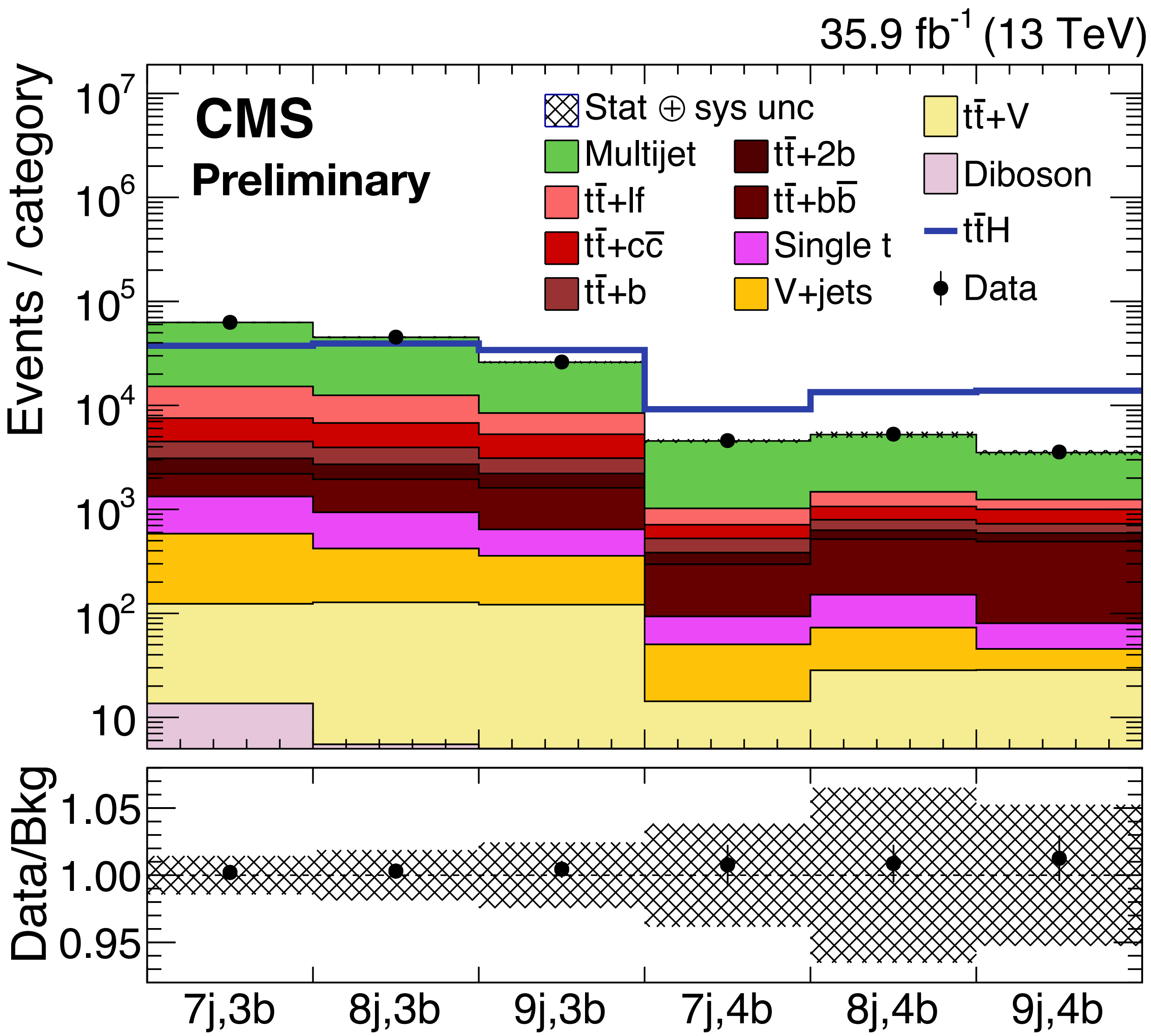
The Matrix Element Method - Probability

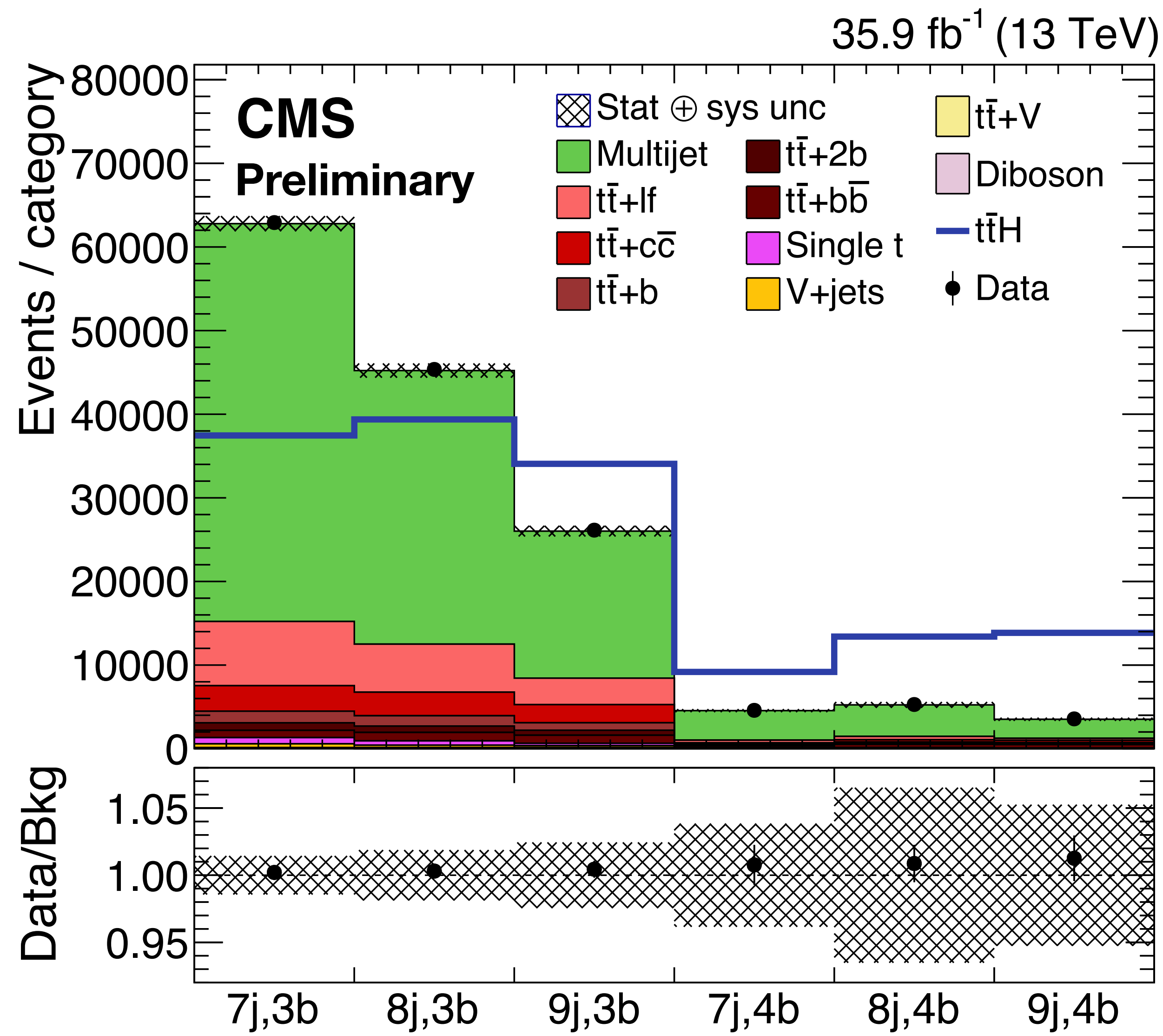
$$\omega_{s,b} = \sum_{\text{perm}} \int d\Phi_8 \, \text{PDF} \otimes |\mathcal{M}_{s,b}|^2 \otimes W$$

- Calculation of the probability density ω for each event to originate from $t\bar{t}H$ (S) or $t\bar{t}+b\bar{b}$ (B)
- Different hypotheses are considered depending on number of reconstructed light- and b jets
- Requires the gluon PDF, scattering and decay amplitude for signal an background and transfer function
- Sum over all jet assignments
- Integration over phase space of the final-state particles

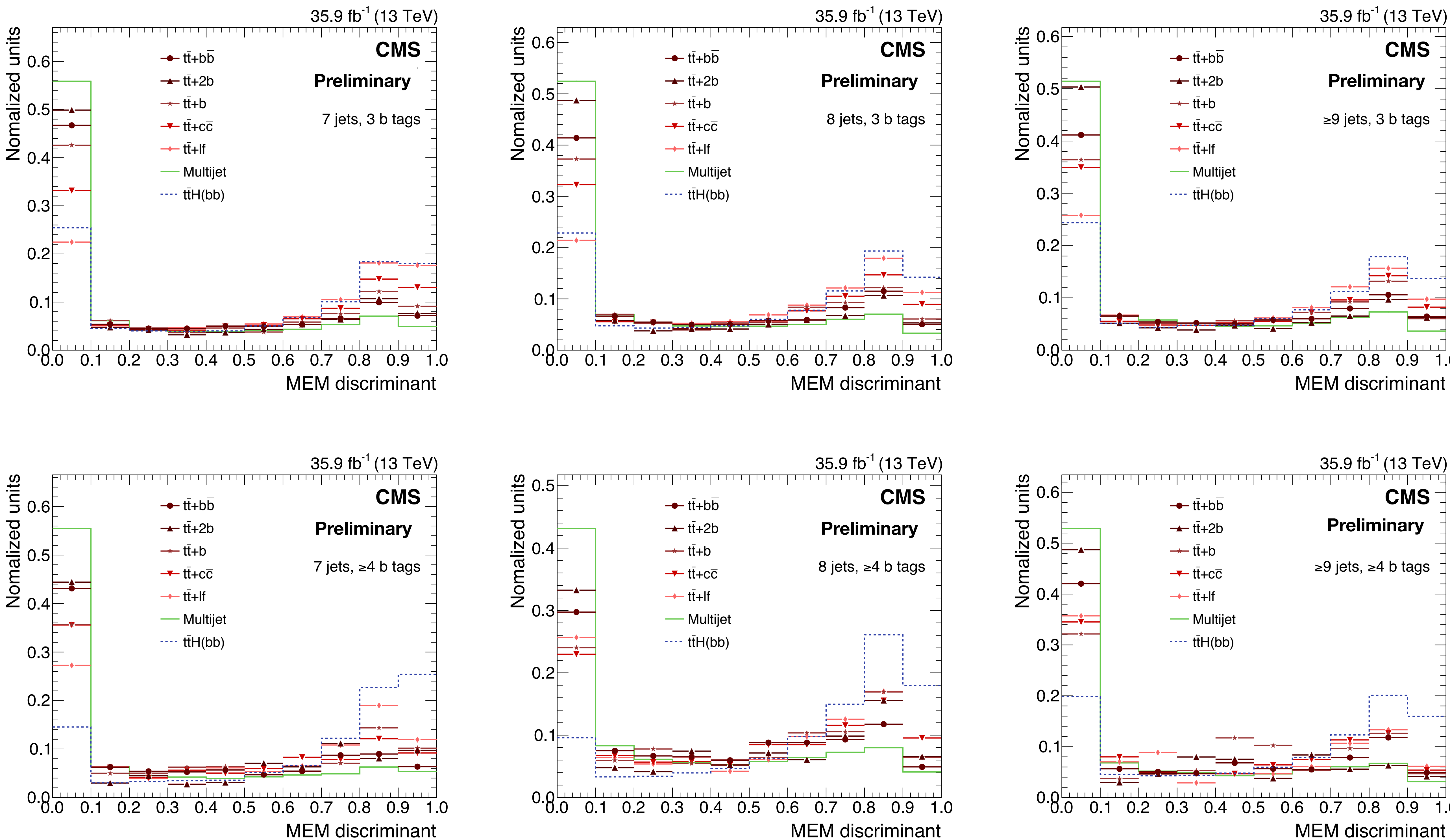


Yields


35.9 fb^{-1} (13 TeV)
CMS
Preliminary
Events / category
Stat ⊕ sys unc
Multijet
tt̄+lf
tt̄+cc̄
tt̄+b
tt̄+2b
tt̄+bb̄
Single t
V+jets
tt̄+V
Diboson
tt̄H
Data
10^7
10^6
10^5
10^4
10^3
10^2
10
Data/Bkg
1.05
1.00
0.95
7j,3b
8j,3b
9j,3b
7j,4b
8j,4b
9j,4b


35.9 fb^{-1} (13 TeV)
CMS
Preliminary
Events / category
Stat ⊕ sys unc
Multijet
tt̄+lf
tt̄+cc̄
tt̄+b
tt̄+2b
tt̄+bb̄
Single t
V+jets
tt̄+V
Diboson
tt̄H
Data
80000
70000
60000
50000
40000
30000
20000
10000
0
Data/Bkg
1.05
1.00
0.95
7j,3b
8j,3b
9j,3b
7j,4b
8j,4b
9j,4b

MEM discriminant



Final discriminant