All hadronic ttH(bb) analysis using the Matrix Element Method

Daniel Salerno

2015 PhD seminar

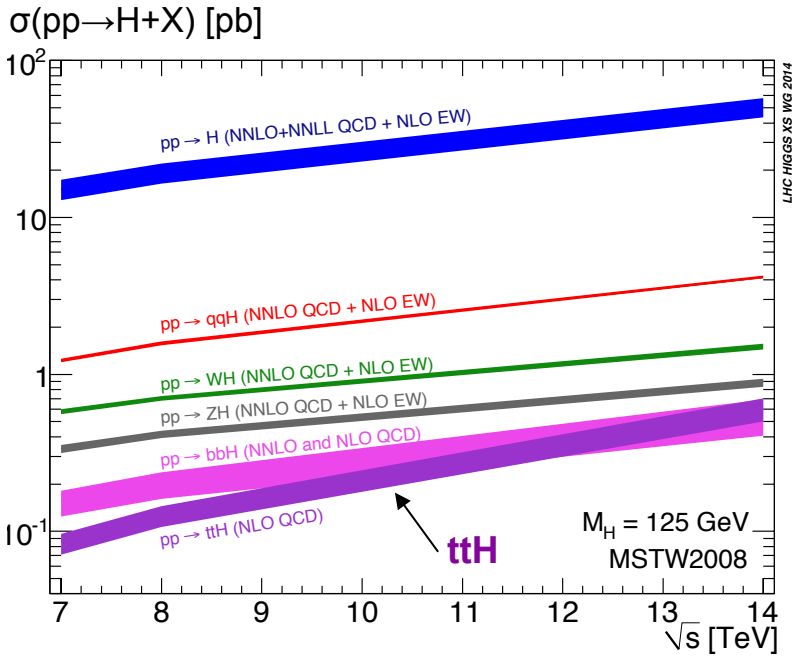
26 August 2015, PSI

Standard model ttH production

Motivation

- Higgs boson with 125 GeV mass discovered by CMS and ATLAS
  - Focus now on studying its properties
- ttH provides a direct probe of the Higgs/top Yukawa coupling y_t
  - Most important fermion coupling
  - Only one with $y_f \sim 1$
    - $y_f = m_f / v$, where $v \approx 246$ GeV
  - Provides insight into possible new physics
- This search is at CMS
  - Multipurpose detector at the LHC

Production cross section at LHC

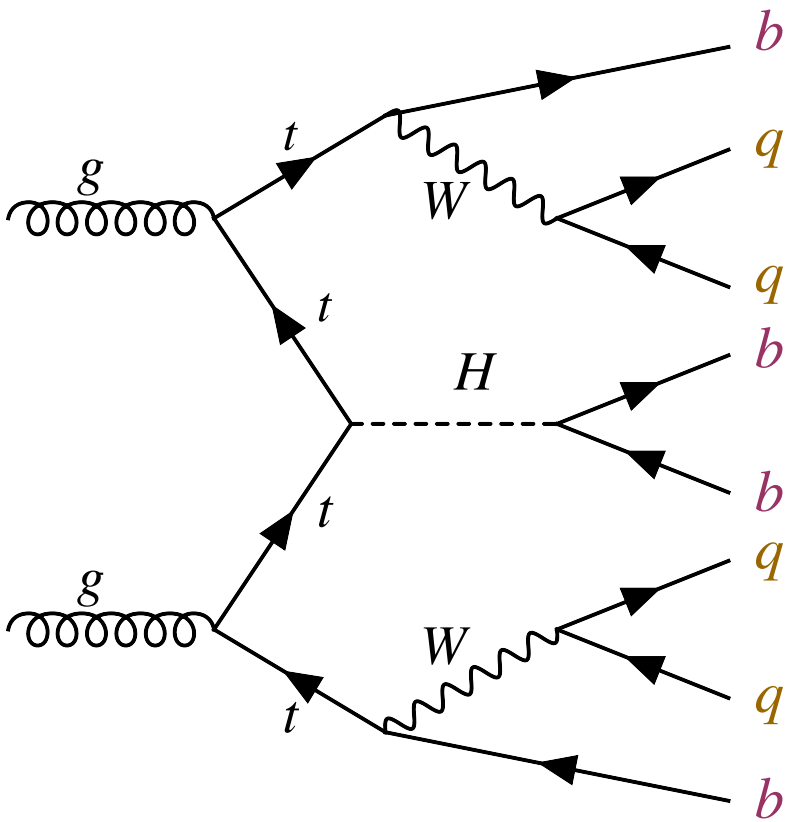


ttH has the lowest cross section of all Higgs production mechanisms

~ 0.1, 0.5, 0.6 pb @ 8, 13, 14 TeV respectively

All hadronic ttH(bb) channel

Feynman diagram

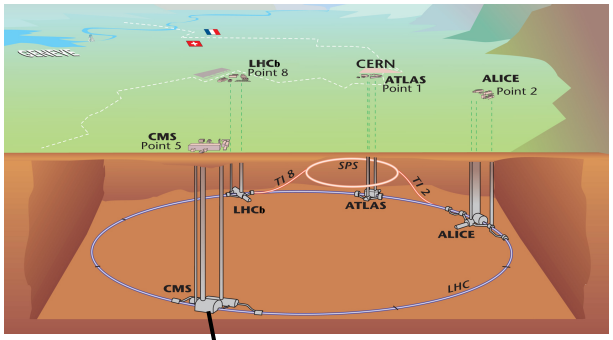


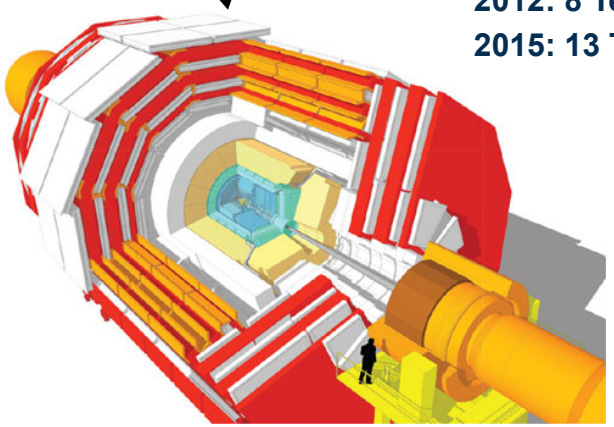
Characteristics

- ☺ H→bb has largest branching ratio of Higgs decays (~58%)
- ☺ All hadronic represents ~46% of all ttH(bb) decays
- ☺ Fully reconstructed final state
  - 8 jets: 4 b-jets and 4 light-jets
- ☹ Large QCD Multijet background
  - Cross section ~10^6 times ttH(bb)
- ☹ Large combinatoric self-background

Experimental search

LHC and the CMS detector





CoM energy
2012: 8 TeV
2015: 13 TeV

8 TeV leptonic analysis

- Completed analysis in single lepton and di-lepton channels
- Used 19.5 fb^{-1} of 8 TeV data
- Background dominated by tt+jets
  - Matrix element showed good discrimination power
- 95% CL upper limit:

$$\mu = \frac{\sigma_{\mathrm{ttH}}}{\sigma_{\mathrm{SM}}} < 4.2\ (3.3)$$

observed (expected)

- Best-fit value:

$$\hat{\mu} = 1.2^{+1.6}_{-1.5}$$

Eur. Phys. J. C, Vol. C75, No. 6, 2015, p. 251

Analysis overview

Organisation

- Part of the ttH-MEM group*
  - Semi leptonic + dilepton ttH(bb)
  - **All hadronic ttH(bb)**
  - Boosted topologies ttH(bb)
  - Leptonic ttH($\tau\tau$)

Data and Monte Carlo

- Currently collecting data from the LHC at 13 TeV
- Monte Carlo samples used to simulate signal and background
  - aMC@NLO and MadGraph interfaced with PYTHIA 8

*** The ttH-MEM group is a collaboration between UZH, ETH, NICPB (Tallinn) and LLR (Ecole Polytechnique)**

Analysis strategy

1 Trigger
- Large p-p collision rate
- Cannot save all events
- Need to select interesting events

2 Selection
- Large amount of background
- Need to reduce it by cutting on measured/calculated variables

3 Matrix Element Method
- Employed after the selection
- Provides final discriminant to further separate signal from background

High level trigger

New paths in HLT menu

- Developed dedicated HLT paths
  - Now integrated in CMS menu
- Control paths also integrated
  - Help measure the efficiency
- Successfully taking data in new runs at 13 TeV

Path	H_T* cut [GeV]	Jet cut [GeV]	b-tag cut
Signal paths			
Trigger 1	>450	≥6 j: p_T>40, |η|<2.6	≥1 b
Trigger 2	>400	≥6 j: p_T>30, |η|<2.6	≥2 b
Prescaled control paths			
Control 1	>450	≥6 j: p_T>40, |η|<2.6	–
Control 2	>400	≥6 j: p_T>30, |η|<2.6	–

* H_T = sum of transverse momentum

Efficiency estimates

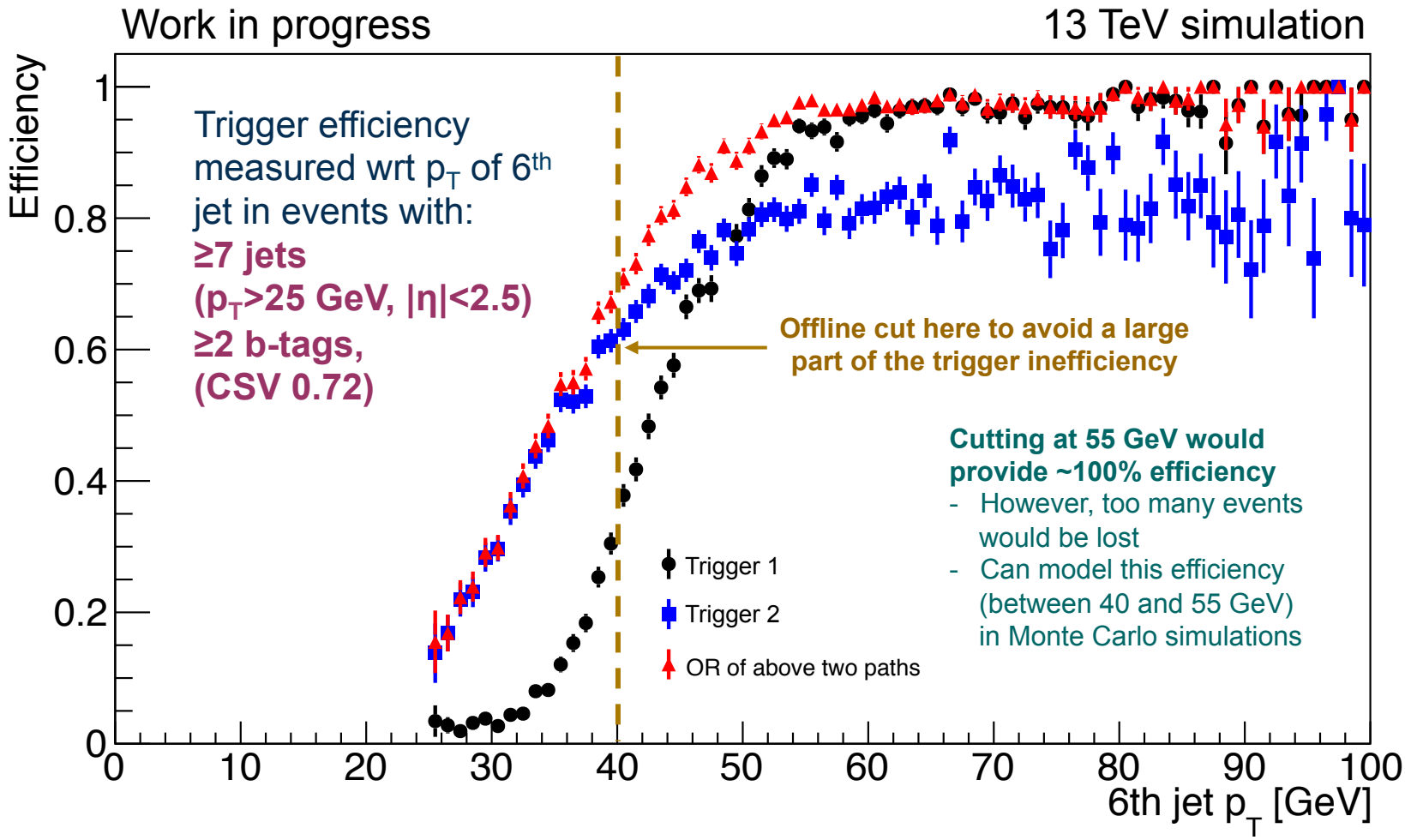
Selection	# in 20 fb^{-1}	Efficiency	
		Incr.	Total
Total ttH(bb) events	5 868		100%
All hadronic events	2 674	46%	46%
≥7 jets with p_T>25 GeV and |η|<2.5, ≥6 jets with p_T>35 GeV, ≥3 b-tags	798	30%	14%
Trigger 1 OR Trigger 2	689	86%	12%

Rate estimates (L = 1.4e34)

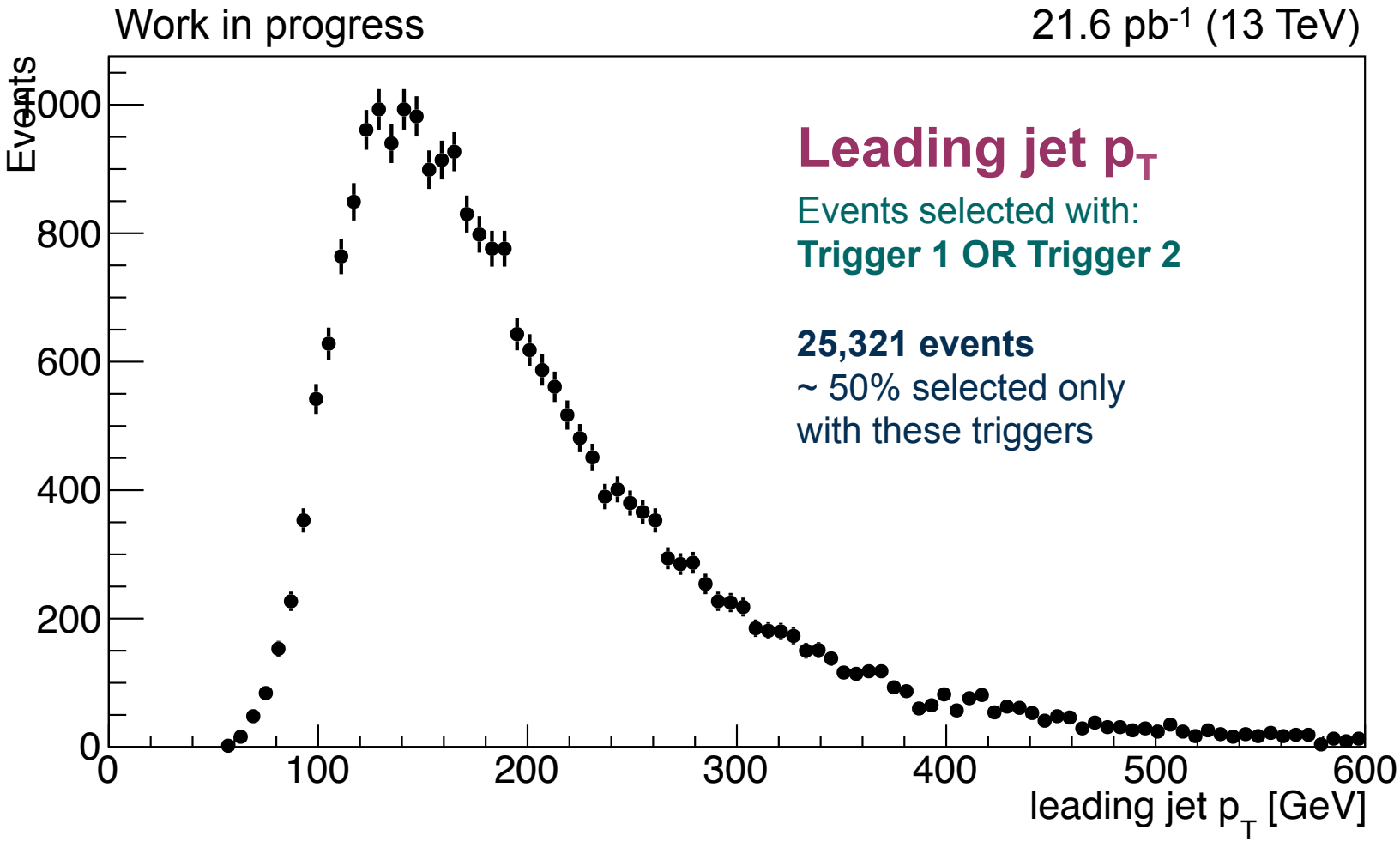
(Hz)	Total	Unique
Total p-p collision rate	40 000 000	
Max total Level 1 Trigger rate	**100 000**	
Maximum total HLT rate	**1200**	
Trigger 1 OR Trigger	17.5±1.4	8.6±0.8

Need to keep rate < 20 Hz, unique rate < 10 Hz

Trigger turn on curve



First 13 TeV data



Event selection

Preselection

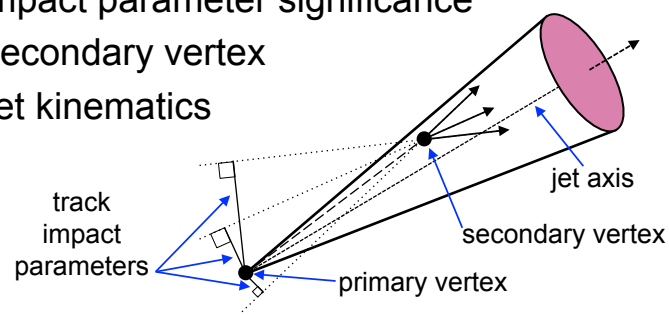
- 6 jets with p_T > 40 GeV, |η| < 2.4
- 2 b-tagged jets (CSV 0.814)
- H_T > 500 GeV (selected jets only)
- Lepton veto (none with p_T > 20 GeV)

Selection possibilities

- Simple cut-and-count
  - e.g. 8 jets p_T > 30 GeV, 4 b-tags
- B-tag likelihood ratio cut
- Kinematic fit
- Quark-gluon separation
  - See next slide
- Combination of the above

B-tag likelihood ratio

- Combined Secondary Vertex (CSV) values (ζ) of each jet calculated
  - Impact parameter significance
  - Secondary vertex
  - Jet kinematics



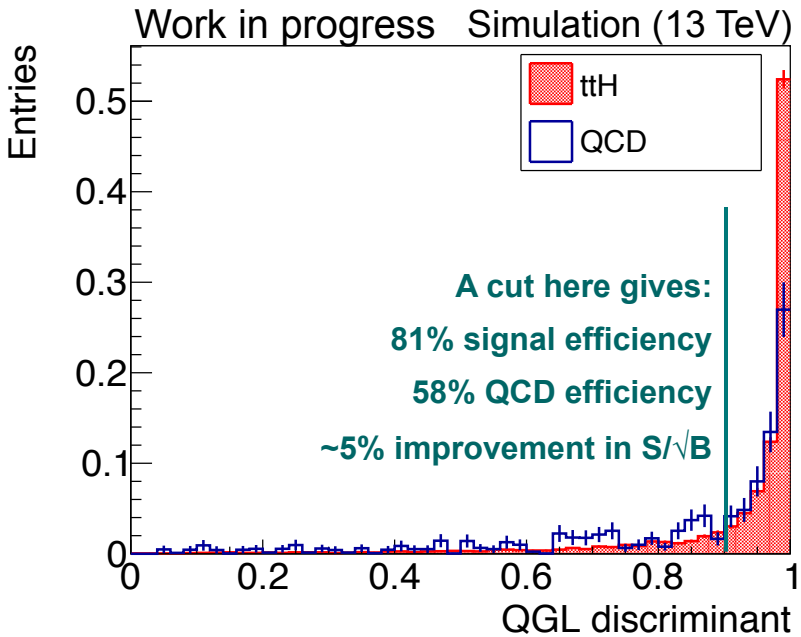
- High CSV values are more likely to be from b-jets
- CSV values used in a likelihood function for competing hypothesis
  - e.g. 4b,4q vs. 2b,6q
- $$b_{\mathrm{LR}} = \frac{\mathcal{L}_{4b4q}(\zeta_1, ..., \zeta_n)}{\mathcal{L}_{4b4b}(\zeta_1, ..., \zeta_n) + \mathcal{L}_{2b6q}(\zeta_1, ..., \zeta_n)}$$

Quark-gluon separation

Quark-gluon likelihood

- Discriminates jets from quarks and jets from gluons
  - QCD Multijet is likely to have more jets from gluons
- Calculated for each jet based on particle-flow composition
  - Energy of constituents
  - Number of constituents
  - Direction of constituents
- Optimized to discriminate light quarks (u,d,s) from gluons
- QGL values used in a likelihood function for competing hypotheses
  - e.g. 8q,0g vs. 4q,4g

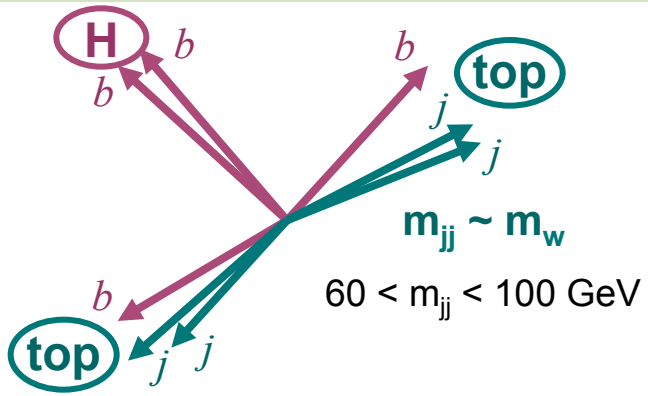
QGL for jet with highest QGL



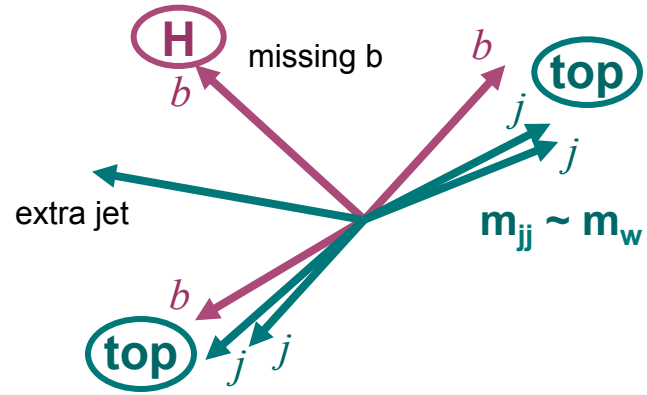
Work ongoing to develop a likelihood ratio incorporating the QGL discriminant and b-tag CSV value

Event categories

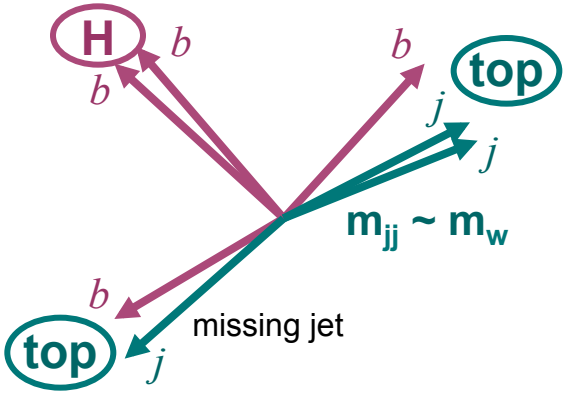
8 jets, ≥4 b



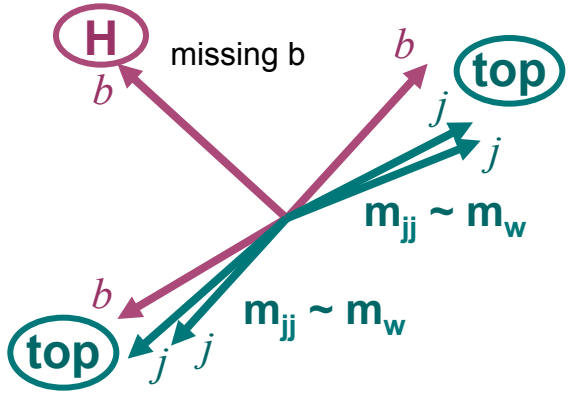
8 jets, 3 b



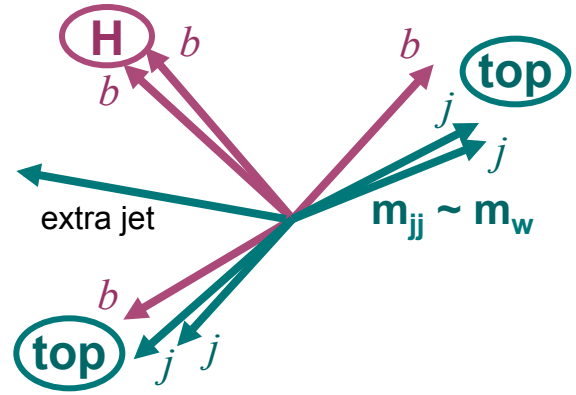
7 jets, ≥4 b



7 jets, 3 b

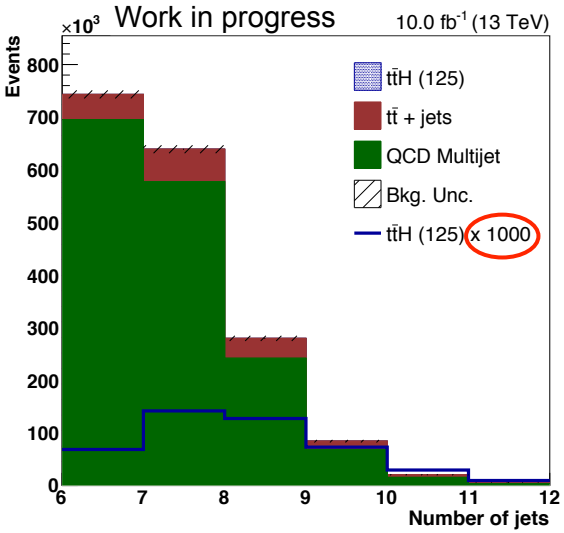


9 jets, ≥4 b

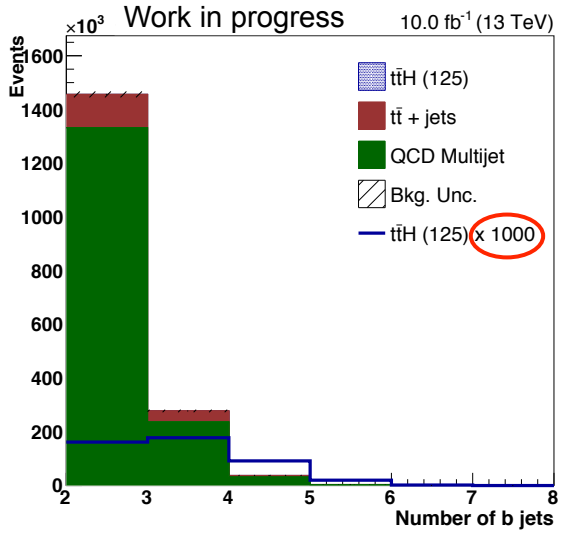


Preselection distributions

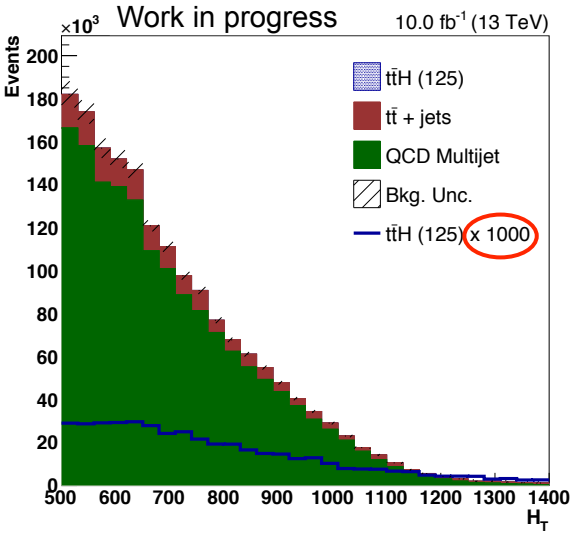
Number of jets



Number of b jets

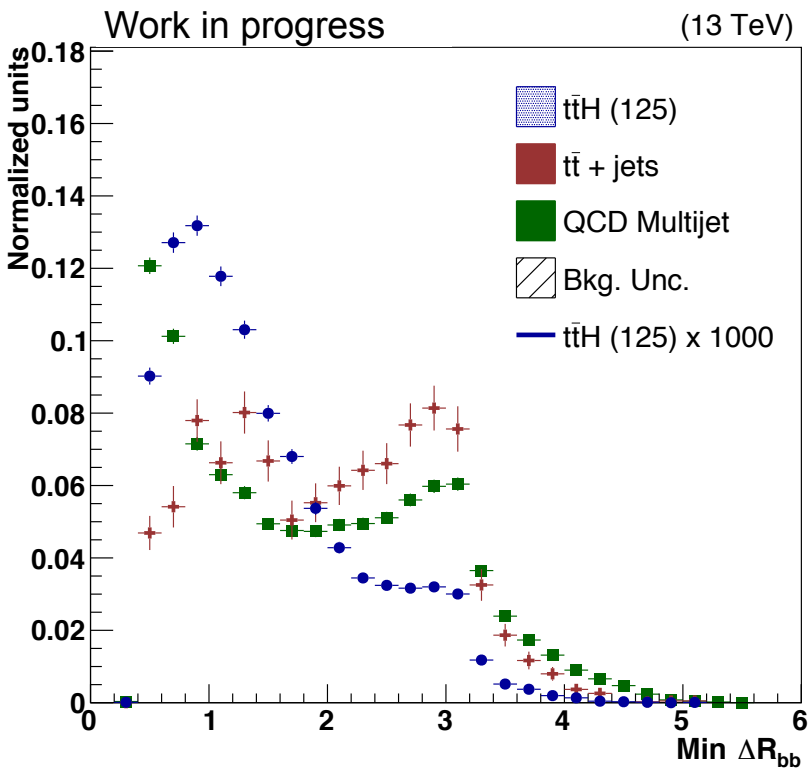


Jet H_T (GeV)

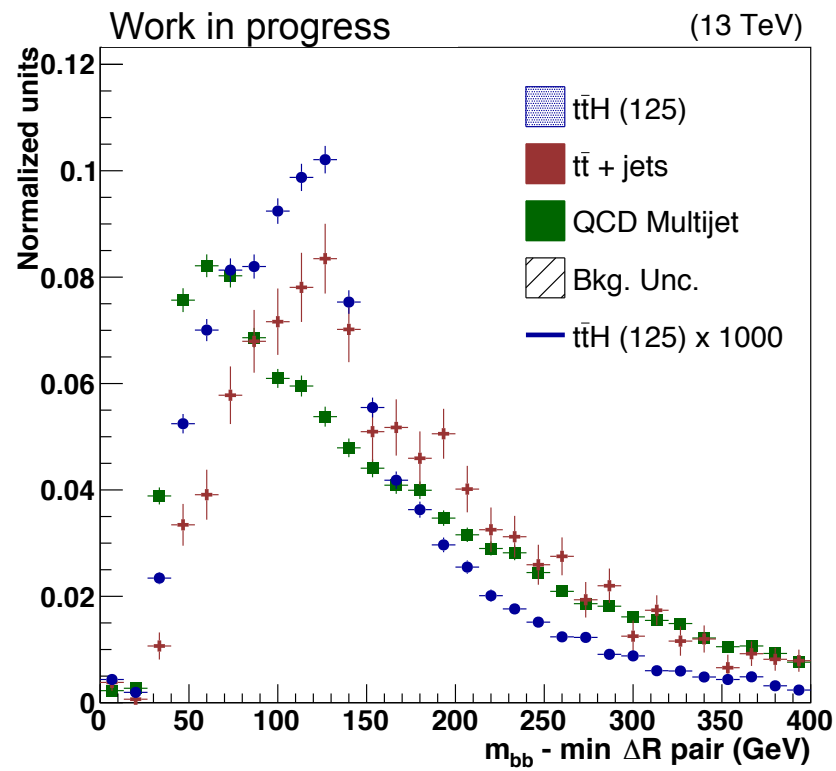


Preselection distributions

Closest bb pair – ΔR



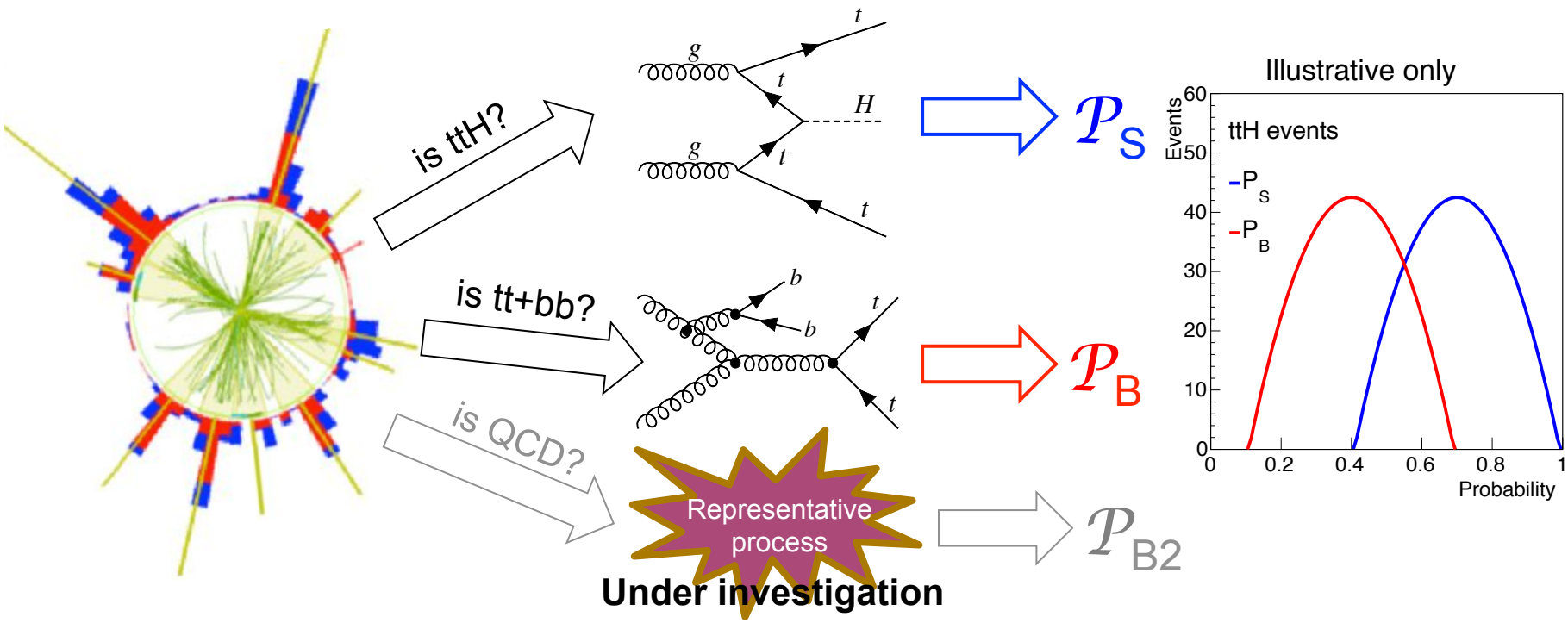
Closest bb pair – mass



The Matrix Element Method

Overview

- Provides optimal separation of signal and background
- Reduces combinatorial self-background (sums over all combinations)
- Calculates the probability of an event being signal/background



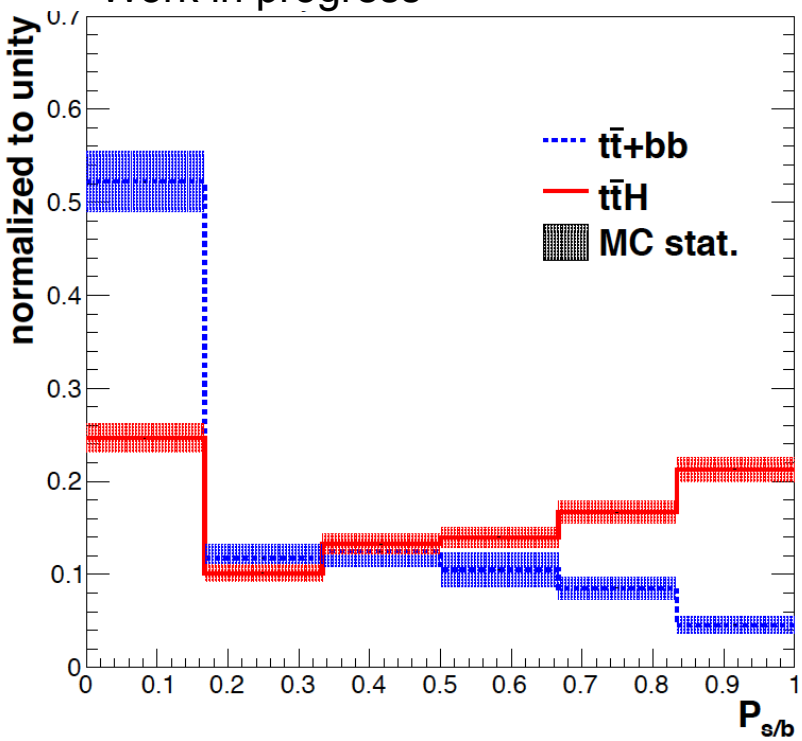
The ME discriminant

Calculation

- For each event $\mathcal{P}_S$ and $\mathcal{P}_B$ are calculated
- Final discriminant is built
  - $P_{s/b} = \dfrac{\mathcal{P}_S}{\mathcal{P}_S + \mathcal{P}_B}$
    - Lies between 0 and 1 by definition
- May eventually add QCD probability, $\mathcal{P}_{B2}$
  - $P_{s/b} = \dfrac{\mathcal{P}_S}{\mathcal{P}_S + \mathcal{P}_B + \mathcal{P}_{B2}}$

Illustrative $P_{s/b}$ distribution

Work in progress



Expected performance

- The matrix element discriminant provides ~15% improvement on a simple yield analysis after selection
- Estimate of expected performance based on yields after simple cut-and-count selection in 5 event categories
  - Worst case scenario, as b-tag and QGL likelihood ratios as well as kinematic fit are expected to boost performance
  - However, **systematic uncertainties are not considered** in this estimate

	10 fb^{-1}	20 fb^{-1}	300 fb^{-1}
$S / \sqrt{B}$	0.36	0.51	2.0
95% CL limit on $\sigma_{ttH} / \sigma_{SM}$	4.9	3.4	–

~15-20% contribution to the total ttH(bb) sensitivity

Semi leptonic
Di leptonic
All hadronic

Summary and next steps

Summary

- The all hadronic ttH(bb) analysis faces many challenges
  - Large QCD multijet background given the lack of leptons
  - Large self background due to the ambiguity of assigning jets to partons
- Many tools are available to reduce the amount of background
  - Kinematic fit, b-tag likelihood ratio and quark-gluon likelihood ratio
- The matrix element method offers a good possibility to overcome both of the major challenges

Next steps

- Optimise selection strategy
  - Implement W mass requirements, b-tag likelihood ratio, quark-gluon likelihood and kinematic fit
- Incorporate boosted topologies
- Optimise the matrix element discriminant

Backup

MC Samples – Spring15

Spring15 production, PU2015_25ns, MINIAODSIM

ttHJetTobb_M125_13TeV_amcatnloFXFX_madspin_pythia8	4.2M events
ttHTobb_M125_13TeV_powheg_pythia8	3.9M events
TTJets_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	43M events
TTJets_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	11M events
QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	20M events
QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	19M events
QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	15M events
QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4.5M events
QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4.0M events
QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.0M events

Low HT QCD samples have insufficient statistics (given the large cross section)

May need to derive a data driven estimate similar to that done with 8 TeV data

The Matrix Element Method

Method

- Measured kinematical variables (**y**) used as input
  - Integration over poorly measured variables (E_{jet}, p_ν)
- Sum over all possible permutations of jet–quark matching

$$w_i(\mathbf{y}) = \frac{1}{\sigma_i} \sum_{\text{perm}} \int_\Omega d\mathbf{x} \int dx_a dx_b \Phi(x_a, x_b) \delta^4\{(x_a P_a + x_b P_b) - \sum p(\mathbf{x})\} |\mathcal{M}_i(\mathbf{x})|^2 W(\mathbf{y}|\mathbf{x})$$

  - Ω = phase space volume of final particles x, $x_{a,b}$ = parton momentum fraction
  - Φ = parton flux factor, $\mathcal{M}_i$ = scattering amplitude of process i (i = ttH, tt+bb)
  - W = transfer function: probability of measuring **y** given **x**

Final discriminants

Two discriminants defined:

- Matrix element $P_{sb} = \dfrac{w_S}{w_S + k_{sb} w_B}$
- B-tag likelihood $P_{bj} = \dfrac{\mathcal{L}_{bbbb}}{\mathcal{L}_{bbbb} + k_{bj} \mathcal{L}_{bbjj}}$
  - where $\mathcal{L}_{bbjj} = \sum_i P(\zeta_1, ..., \zeta_6 | \{bbjjjj\}_i)$ and $\zeta_1, \ldots, \zeta_6$ are the jet CSV values

In 8 TeV leptonic analysis a 2D analysis was performed: 6 bins in P_{sb} × 2 bins in P_{bj}

Matrix Element details

Inputs

t – W
- Light jet: θ, ϕ, E → q
- Light jet: θ, ϕ, E → q'
- b-tag jet: θ, ϕ, E → b

t – W
- Light jet: θ, ϕ, E → q
- Light jet: θ, ϕ, E → q'
- b-tag jet: θ, ϕ, E → b

H
- b-tag jet: θ, ϕ, E → b_1
- b-tag jet: θ, ϕ, E → b_2

Precisely measured
Integrated over resolution (±4σ)
Calculated from other variables
Integrated over full range (?)

Kinematic reconstruction – top

$$E_{\mathrm{q}'} = \frac{m_{\mathrm{W}}^2}{4E_{\mathrm{q}}\sin^2(\frac{\theta_{\mathrm{qq}'}}{2})}$$

$$E_{\mathrm{b}} = \frac{a\Delta_{m_{\mathrm{t}}} \pm |b|\sqrt{\Delta_{m_{\mathrm{t}}}^2 - (a^2-b^2)m_{\mathrm{b}}^2}}{a^2-b^2}$$

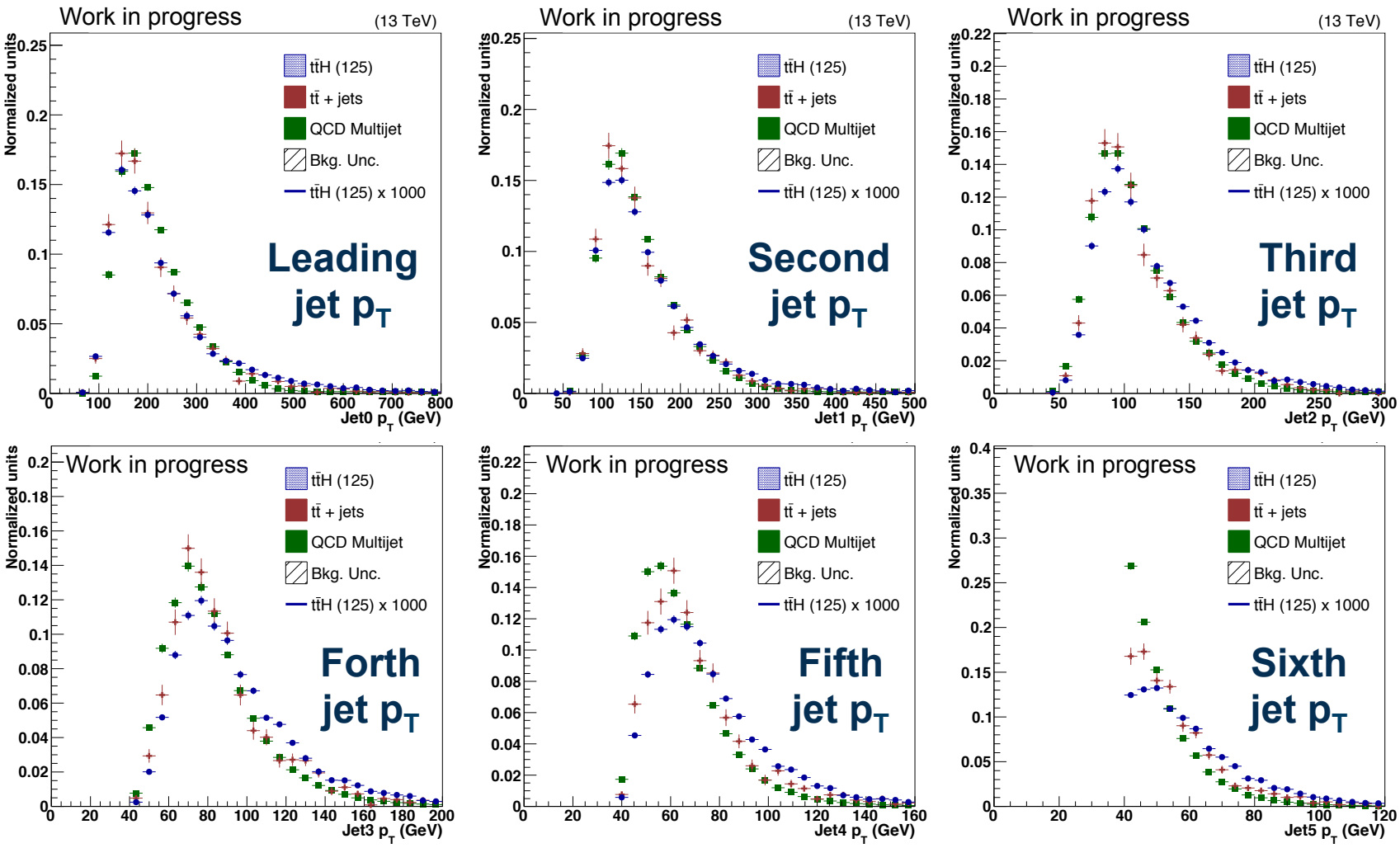
where
$$a \equiv E_{\mathrm{q}} + E_{\mathrm{q}'}$$
$$b \equiv E_{\mathrm{q}}(\vec{e}_{\mathrm{q}}\cdot\vec{e}_{\mathrm{b}}) + E_{\mathrm{q}'}(\vec{e}_{\mathrm{q}'}\cdot\vec{e}_{\mathrm{b}})$$
$$\Delta_{m_{\mathrm{t}}} \equiv \tfrac{1}{2}(m_{\mathrm{t}}^2 - m_{\mathrm{b}}^2 - m_{\mathrm{W}}^2)$$

Kinematic reconstruction – Higgs

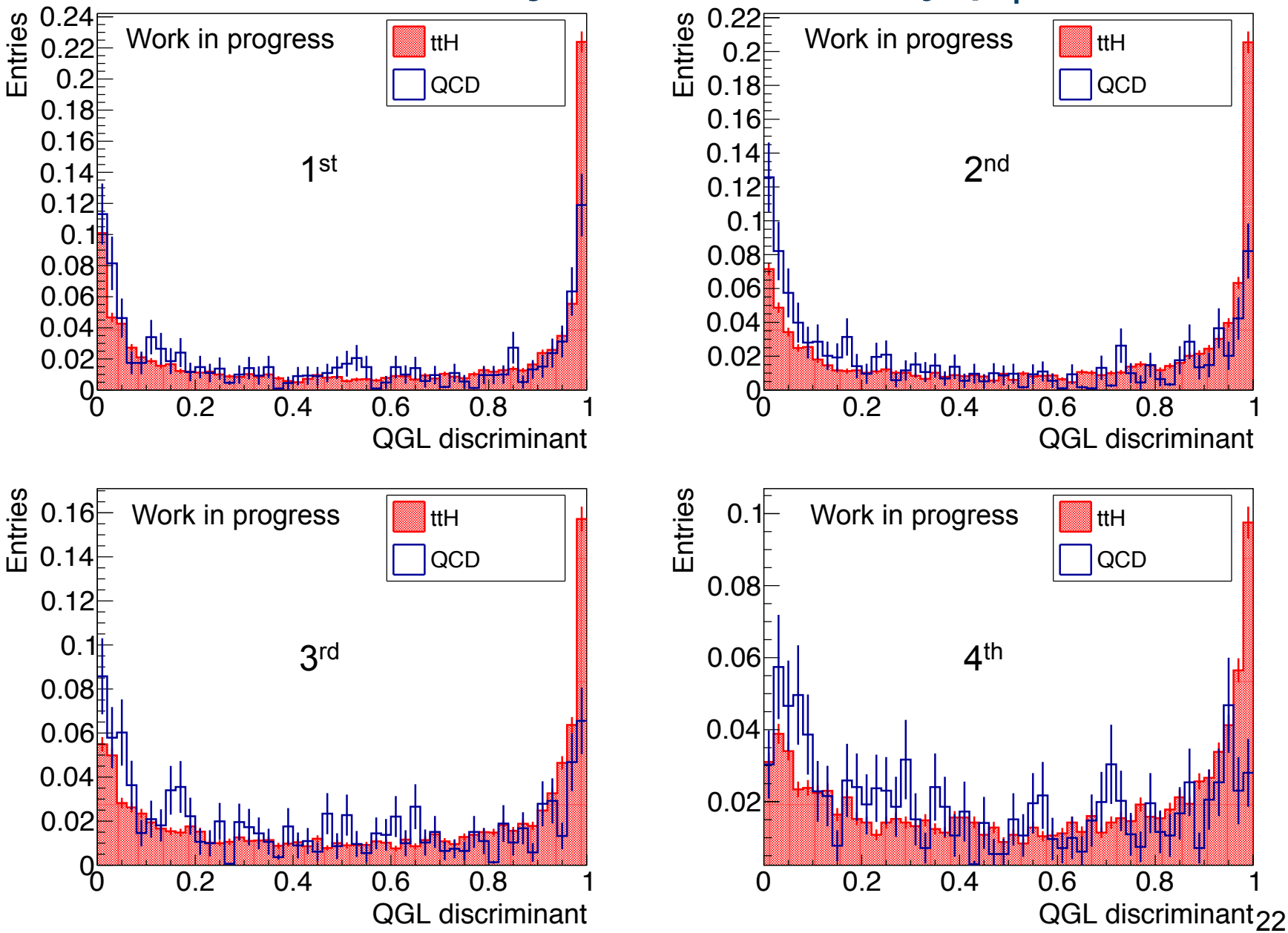
$$E_{\mathrm{b}_2} = \frac{a\Delta_{m_{\mathrm{H}}} \pm |b|\sqrt{\Delta_{m_{\mathrm{H}}}^2 - (a^2-b^2)m_{\mathrm{b}}^2}}{a^2-b^2}$$

where
$$a \equiv E_{\mathrm{b}_1}$$
$$b \equiv \sqrt{E_{\mathrm{b}_1}^2 - m_{\mathrm{b}}^2}(\vec{e}_{\mathrm{b}_1}\cdot\vec{e}_{\mathrm{b}_2})$$
$$\Delta_{m_{\mathrm{H}}} \equiv \tfrac{1}{2}(m_{\mathrm{H}}^2 - m_{\mathrm{b}}^2)$$

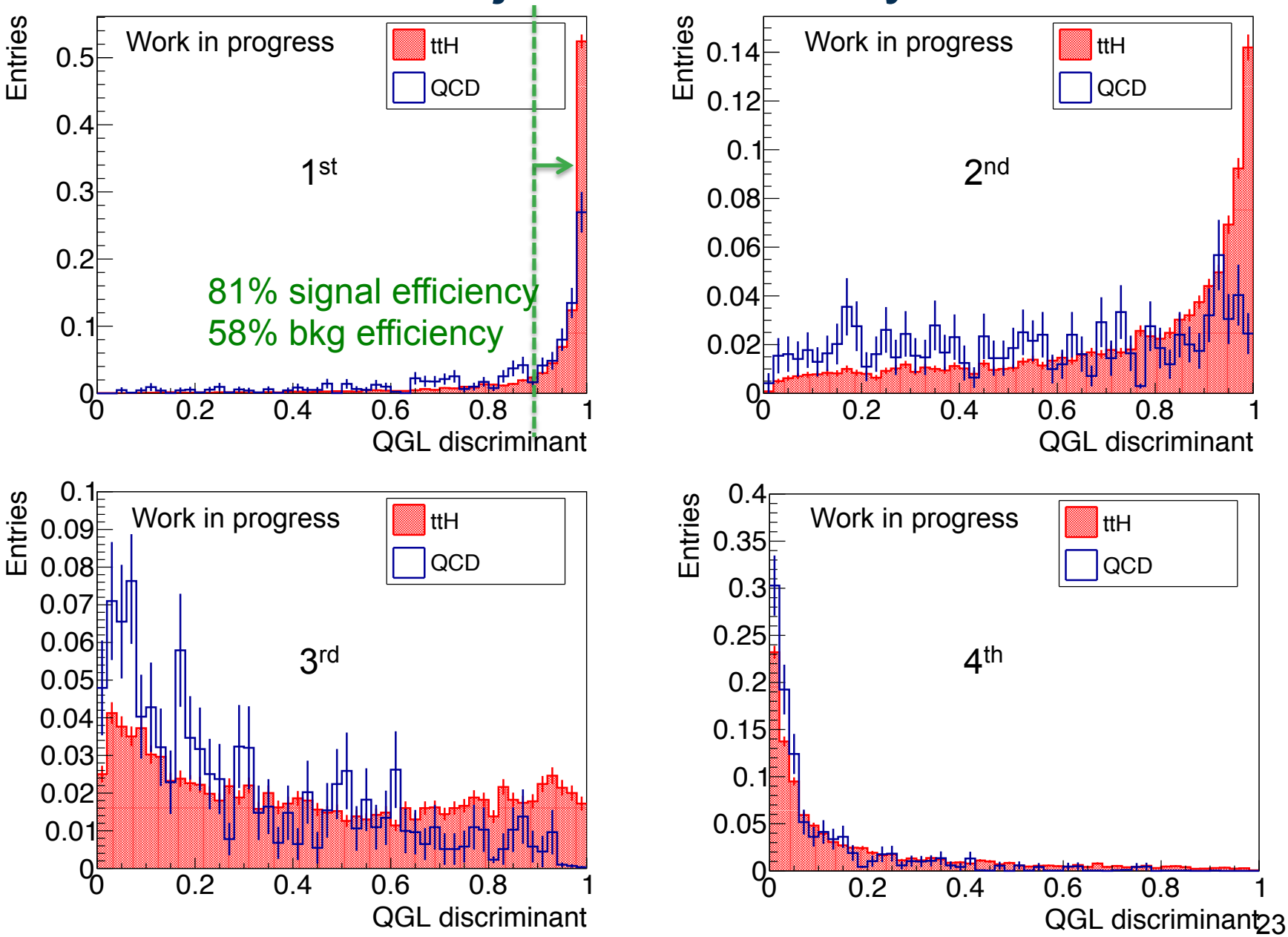
Preselection distributions: jet p_T



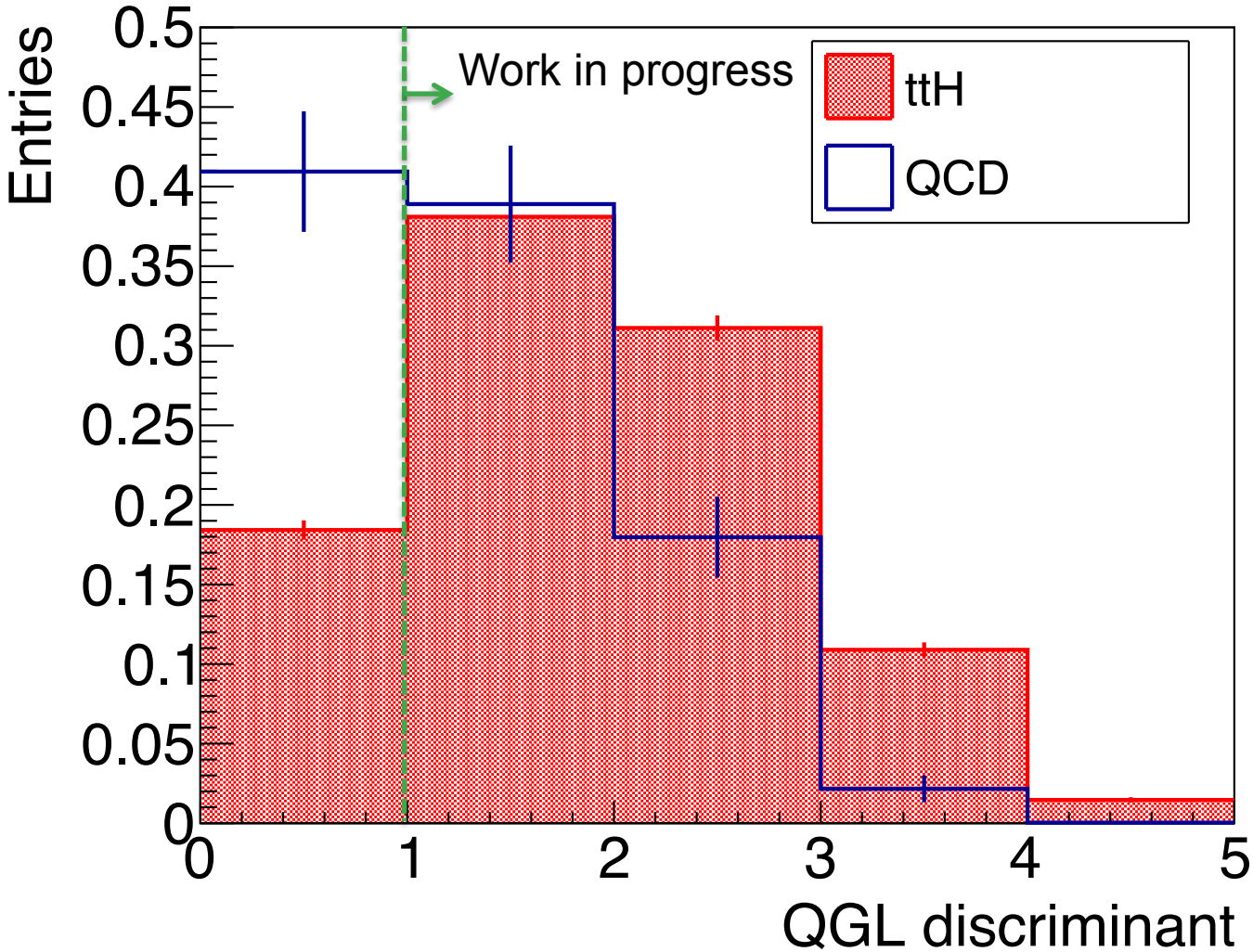
QGL for jets ordered by p_T



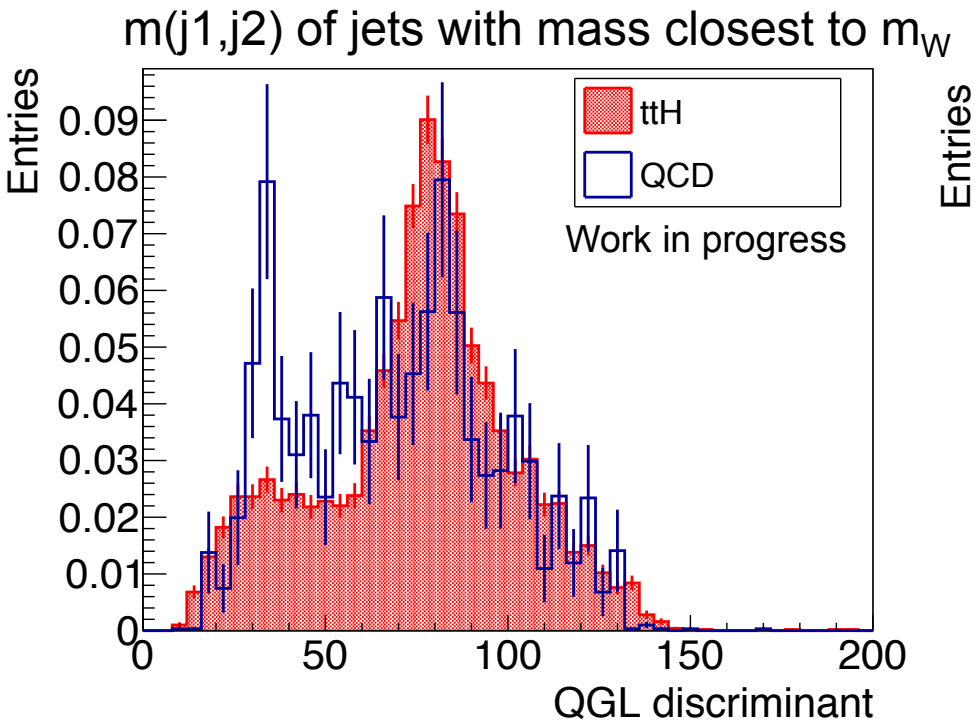
QGL for jets ordered by QGL

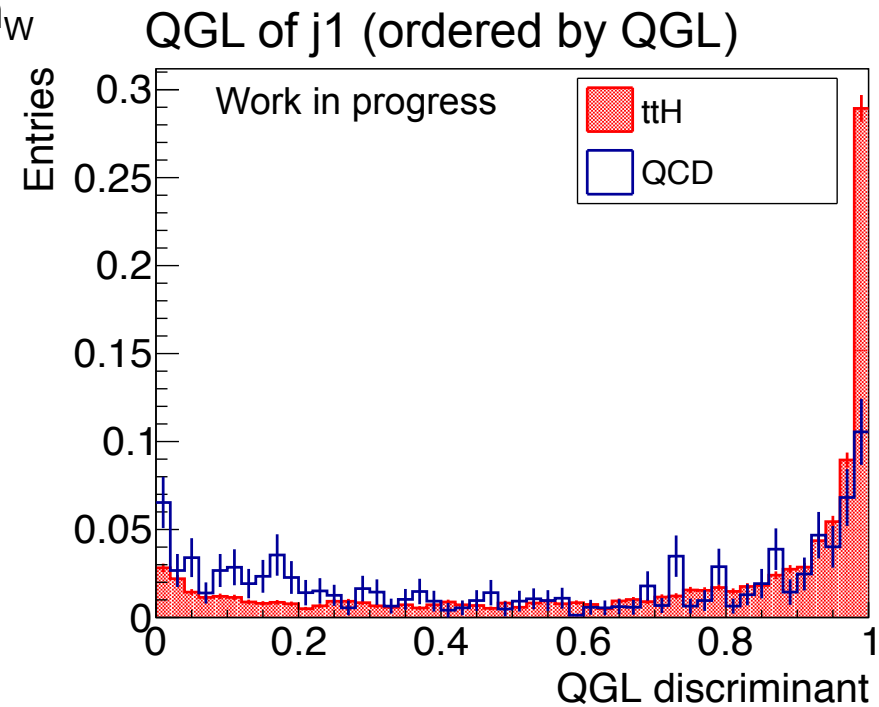


Number of non-b-jets with QGL>0.9



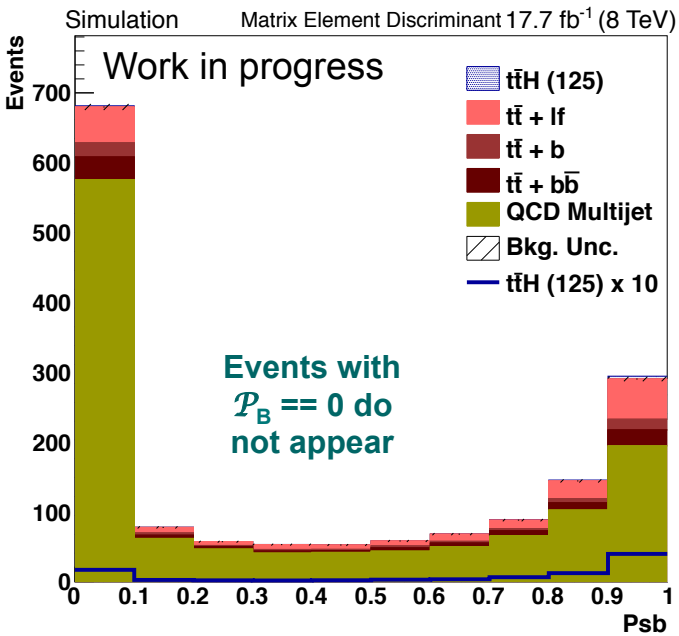
QGL of jets with best W mass





8 TeV performance

$P_{s/b}$ distribution: 8j, ≥4b



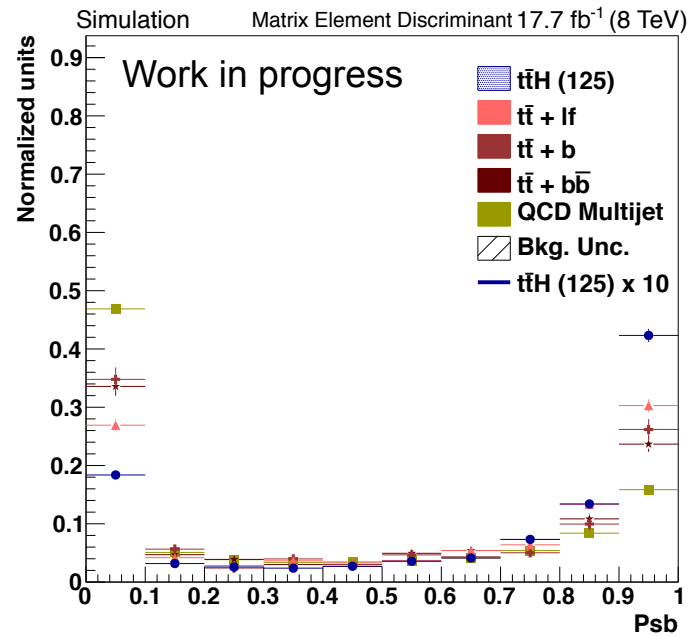
Yields

QCD	1228
tt+jets	345
Background	**1574**
Signal	9.5
S/√B	0.240

Cut on $P_{s/b}$ to maximise sensitivity

→ **$P_{s/b}$ > 0.50**

Normalised



Yields

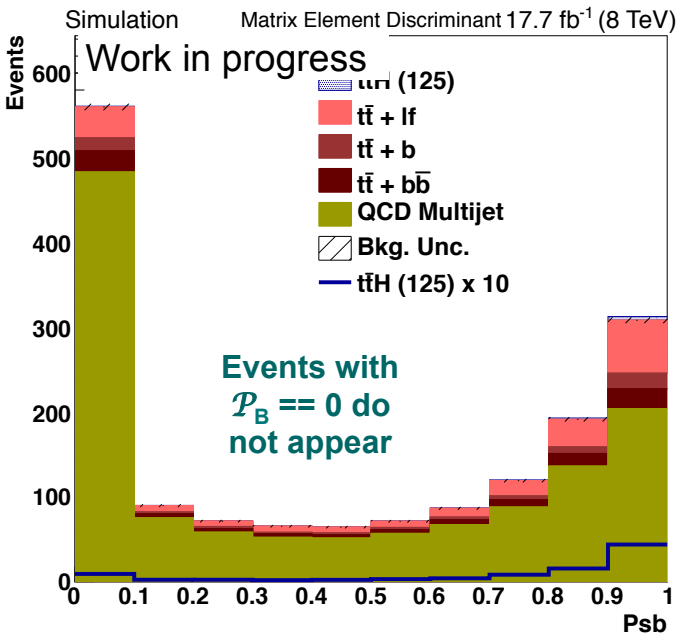
QCD	459
tt+jets	192
Background	**651**
Signal	6.7
S/√B	0.264

→ **95% CL upper limit on σ_{ttH}/σ_{SM} ~ 7.5**

8 TeV performance

Integrating over one "missing" parton

$P_{s/b}$ distribution: 8j, ≥4b



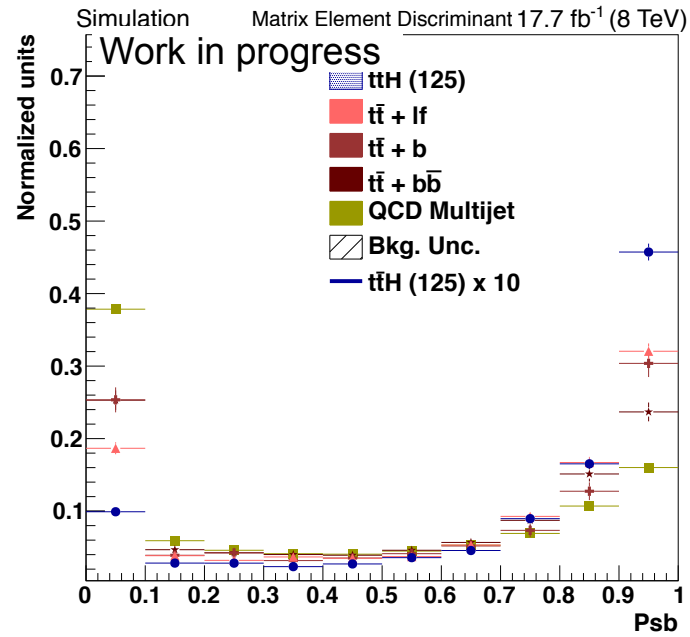
Yields

QCD	1277
tt+jets	354
Background	**1631**
Signal	9.7
S/√B	0.240

Cut on $P_{s/b}$ to maximise sensitivity

→ $P_{s/b} > 0.68$

Normalised



Yields

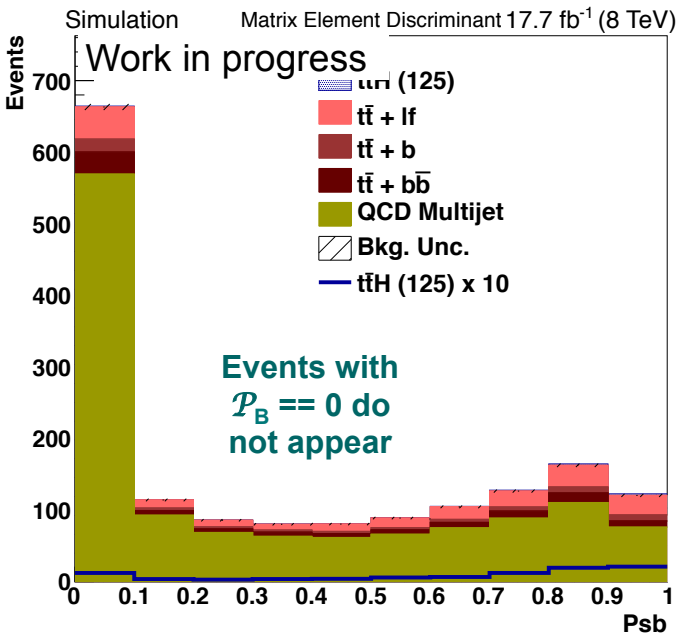
QCD	445
tt+jets	194
Background	**639**
Signal	7.0
S/√B	0.277

→ **95% CL upper limit on σ_{ttH}/σ_{SM} ~ 7.2**

8 TeV performance

"Optimising" discriminant scale factor

$P_{s/b}$ distribution: 8j, ≥4b



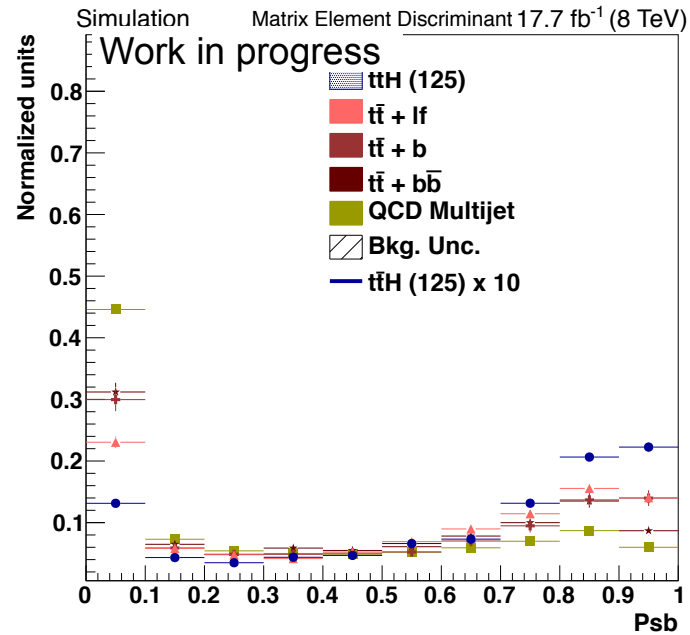
Yields

QCD	1277
tt+jets	354
Background	**1631**
Signal	9.7
S/√B	0.240

Cut on $P_{s/b}$ to maximise sensitivity

→ $P_{s/b} > 0.46$

Normalised



Yields

QCD	444
tt+jets	194
Background	**638**
Signal	7.0
S/√B	0.277

→ **95% CL upper limit on σ_{ttH}/σ_{SM} ~ 7.1**

Predictions for Run II

Pre discriminant yields

- Extrapolated from 8 TeV yields
- So far only 8j,≥4b and ≥9j,≥4b categories provide meaningful sensitivity
  - Preselection on W mass not yet implemented
- Yields for 20 fb^{-1} at 13 TeV
  - ttH 75
  - tt+jets 2 100
  - QCD 5 500
  - Total background 7 600

Early analysis of 13 TeV MC samples suggest more favourable yields

Estimated post-fit sensitivity

- Assuming 10% post-fit error on Signal and Background
- Assuming 15% improvement from the Matrix Element Discriminant
  - (Based on 8 TeV leptonic analysis)

	5 fb^{-1}	20 fb^{-1}	300 fb^{-1}
$S / \sqrt{B}$	0.37	0.74	2.9
95% CL limit on $\sigma_{ttH} / \sigma_{SM}$	5.4	2.7	–