ETH Institute for Particle Physics

Higgs boson production in association with a top quark pair: ttH(bb) with the matrix element method

Joosep Pata

PSI PhD seminar

26.08.2015

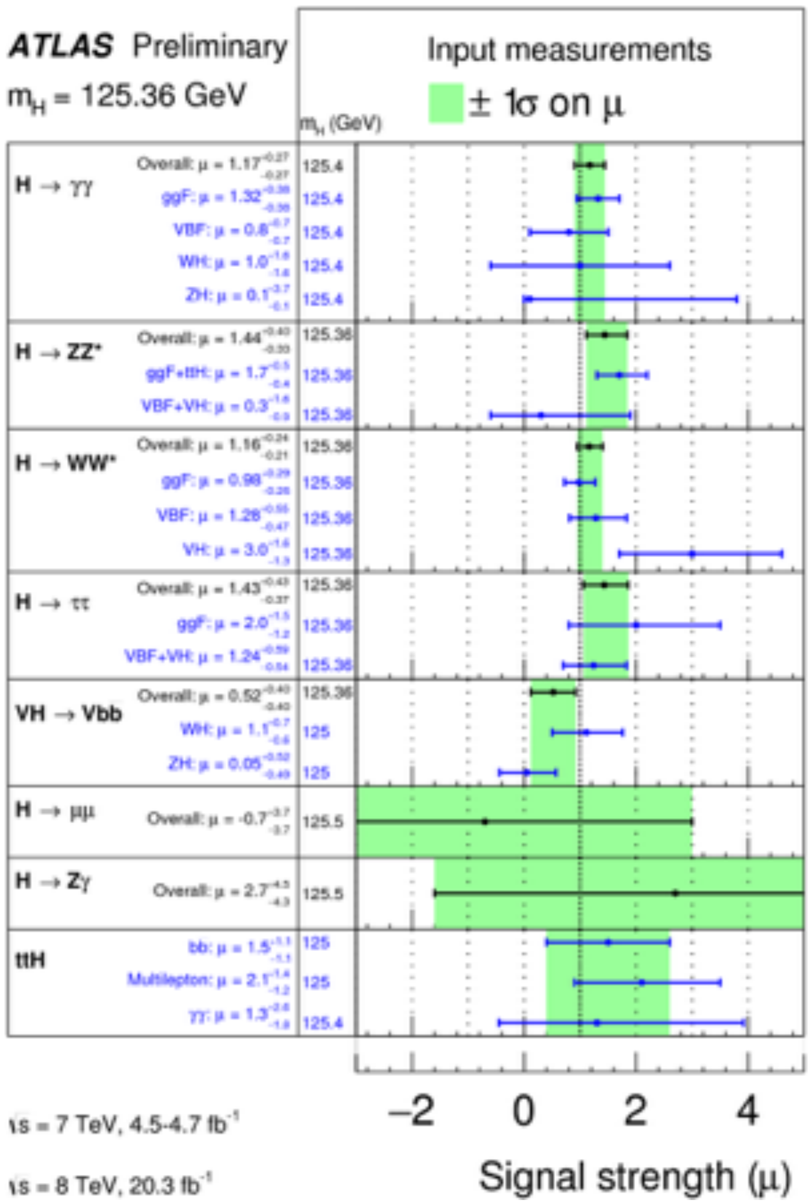
In collaboration with L. Bianchini, L. Caminada, F. Canelli, G. Dissertori, G. Kasieczka, D. Salerno

ETH Zürich

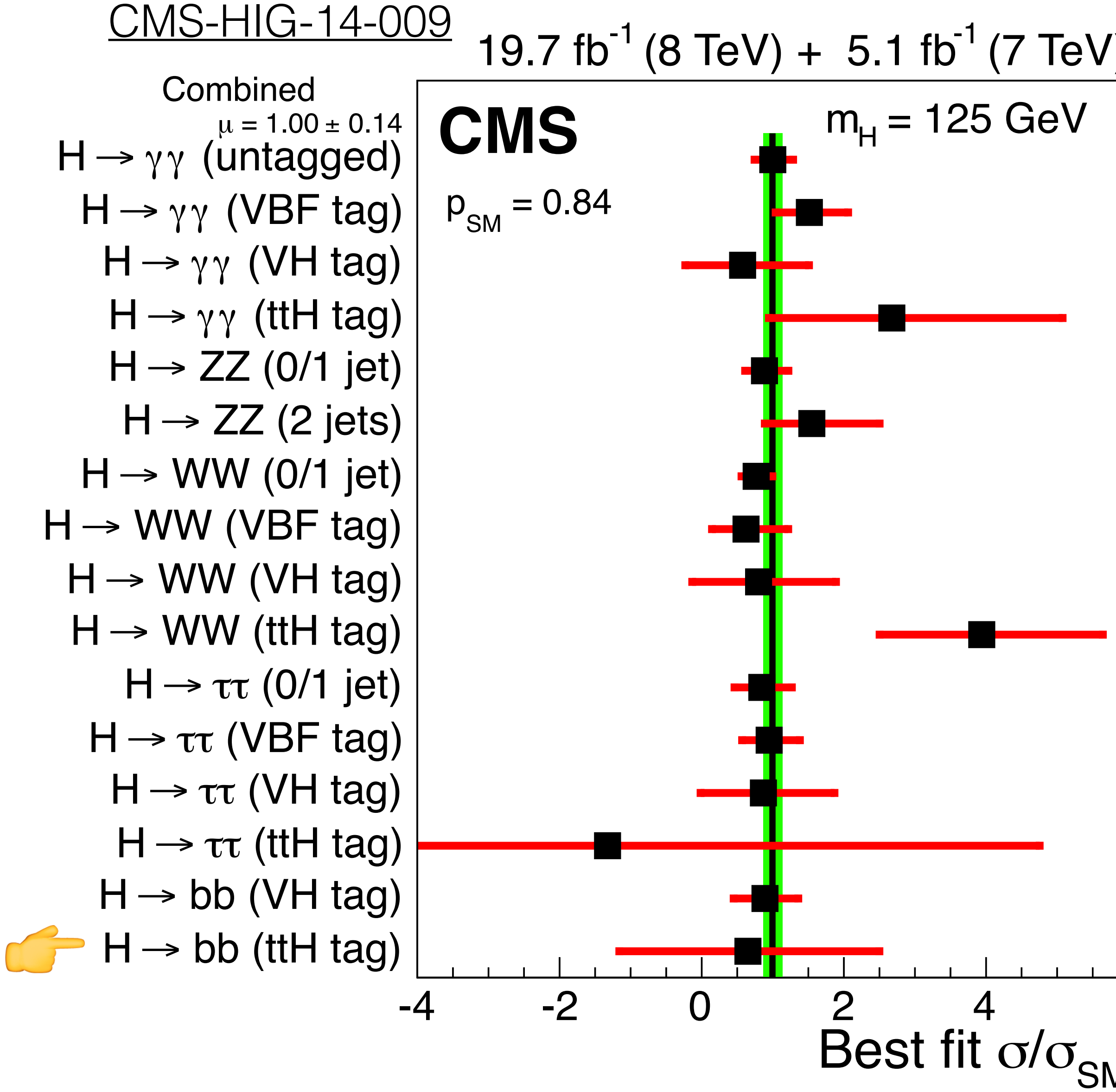
higgstools

What we know about H(125)

ATLAS-CONF-2015-007



CMS-HIG-14-009



H(125) coupling data mostly consistent with SM Higgs boson, but small differences could be very interesting!

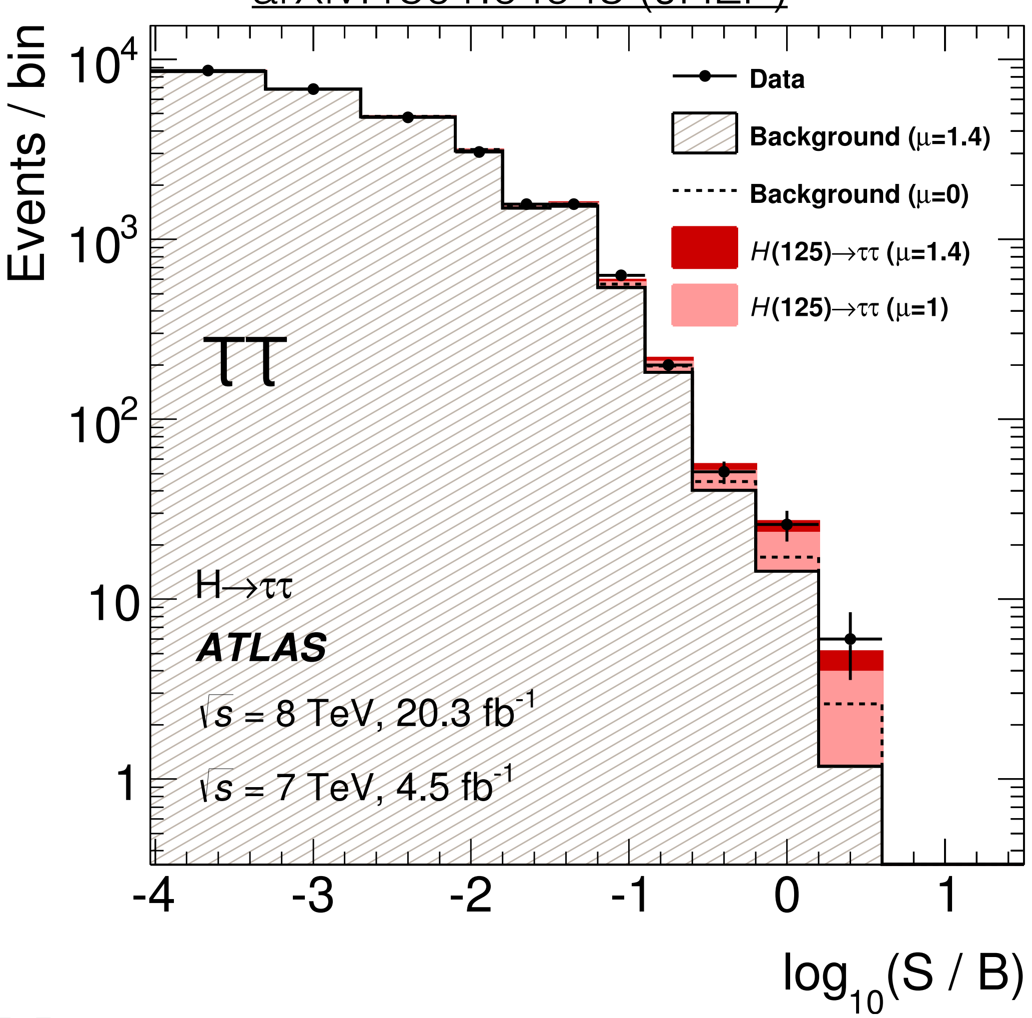
~15% uncertainty on top Yukawa with 300 fb^{-1} in Run II: **reaching the precision phase!**

Coupling to fermions

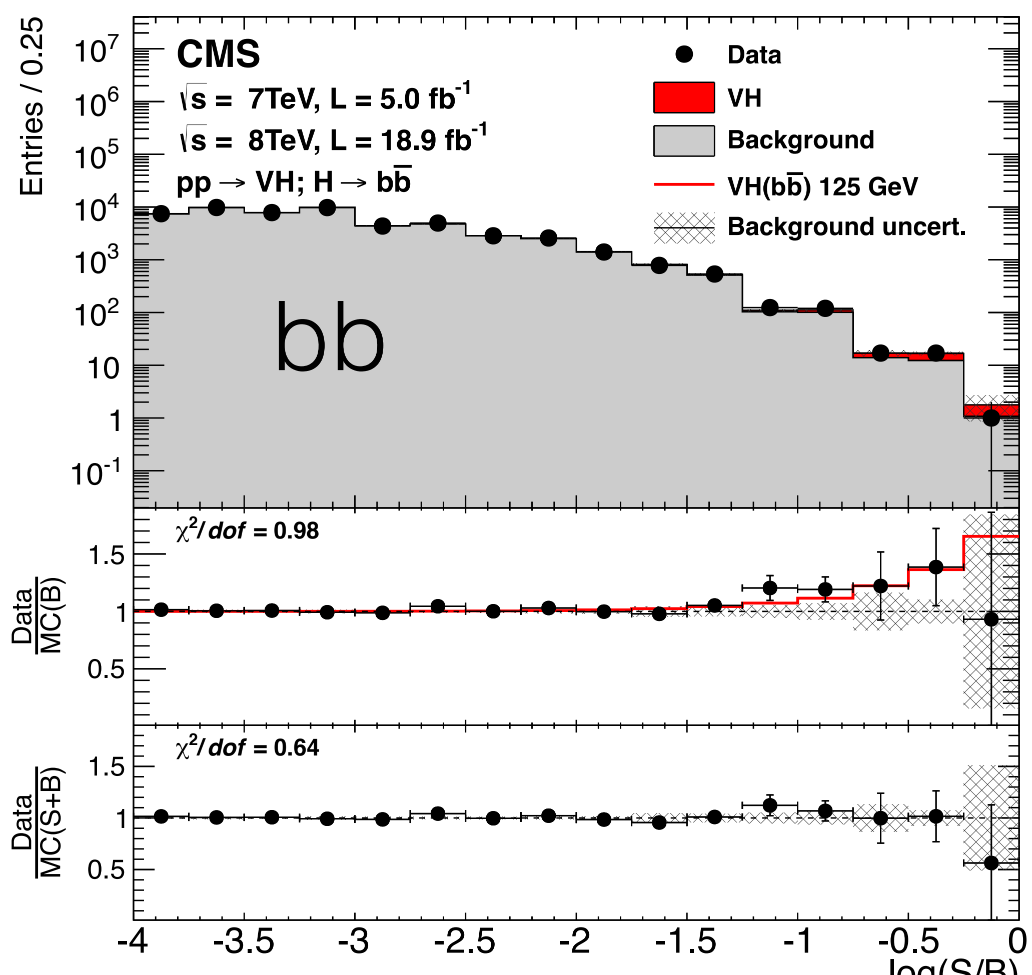
Verify EWSB mass generation by comparing **SM H decay to fermions to measurements**.

Direct evidence **down-type** (b, tau), indirect to **up-type** (top loops).

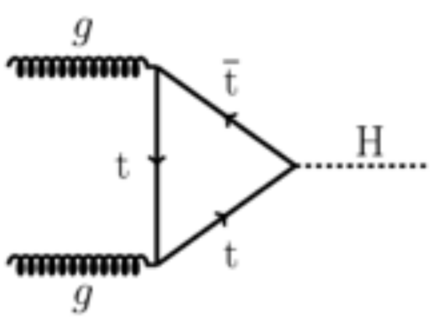
arXiv:1501.04943 (JHEP)



1310.3687 (PRD)

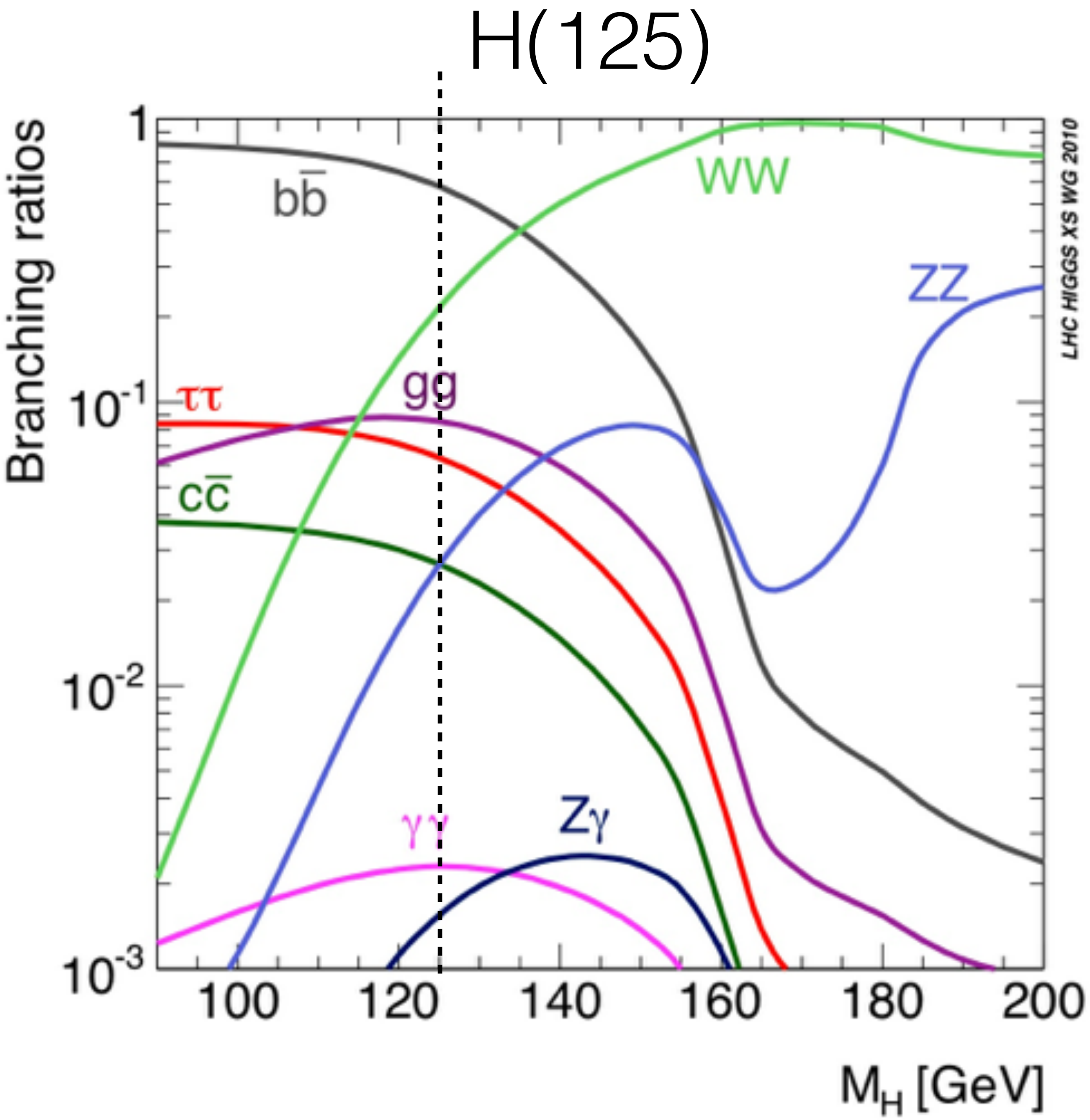


indirect evidence for coupling to tops through loop diagrams

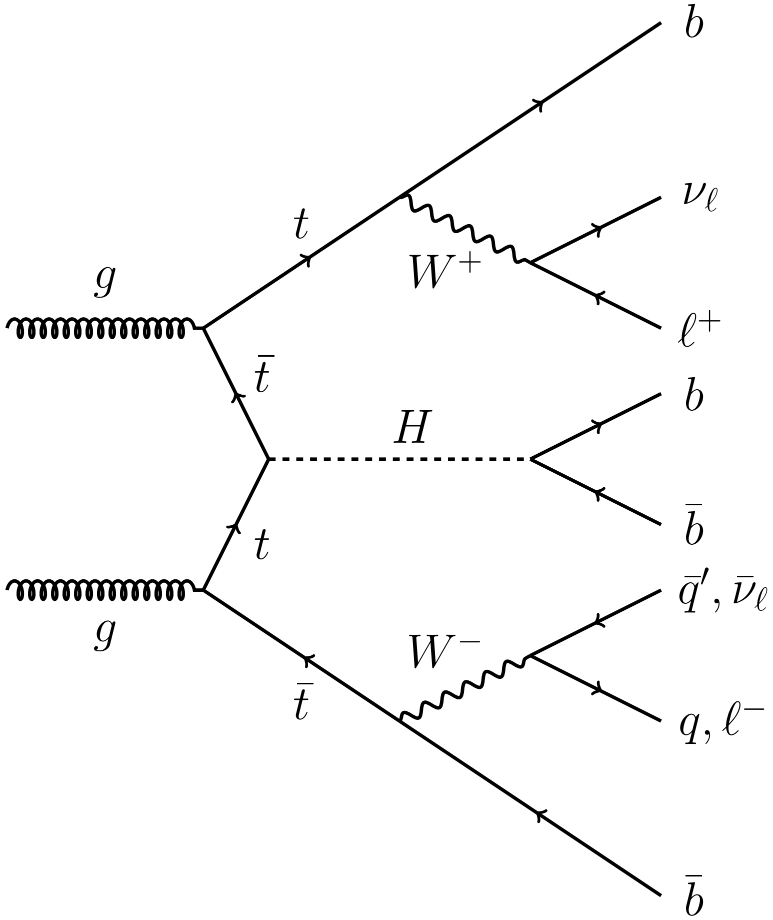


What to look for?

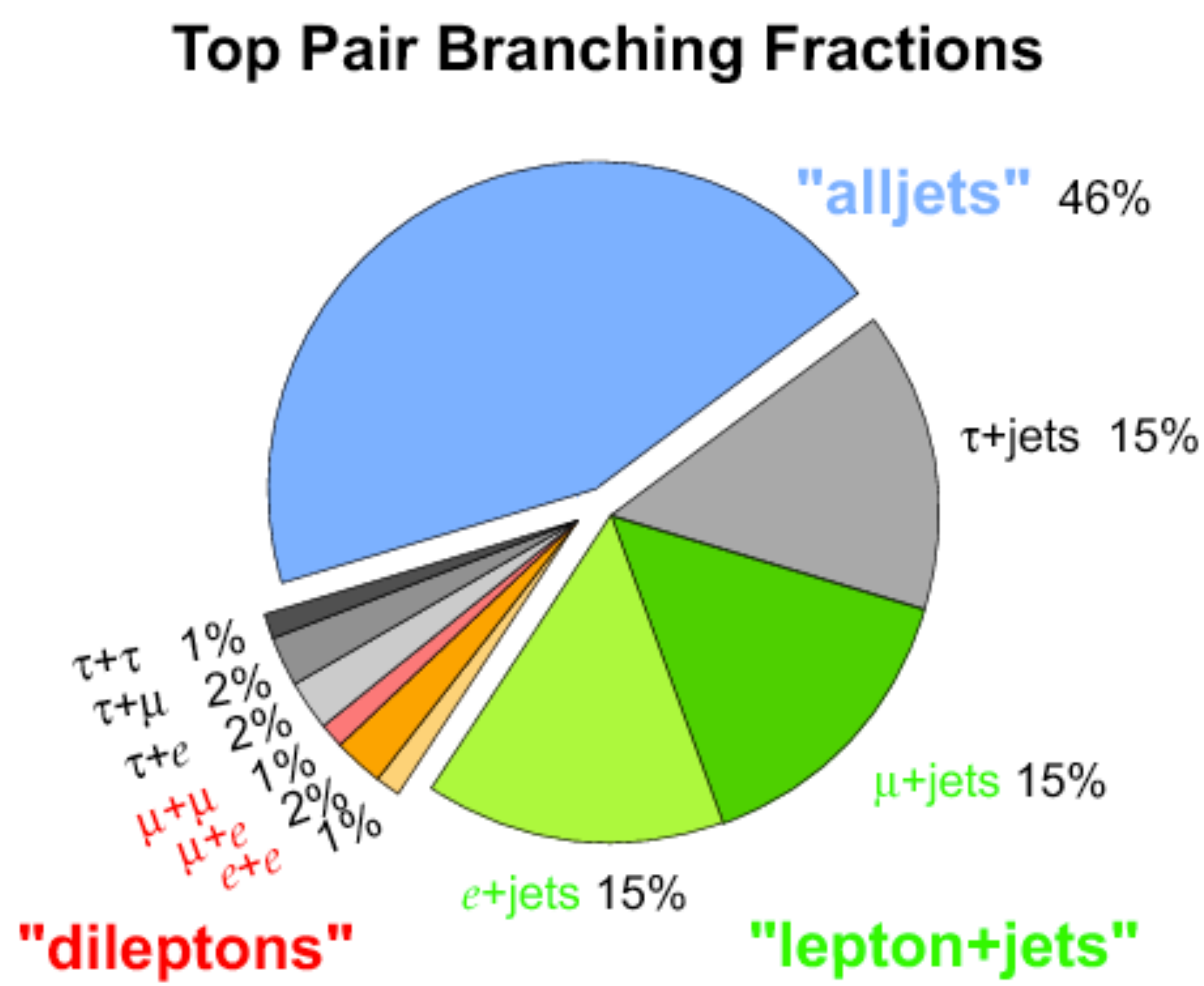
H(125) near-ideal for **bb** statistics-wise, but a very complicated final state.



Final state



1-2 leptons,
4-6 jets (4b),
MET



Reconstruction and selection

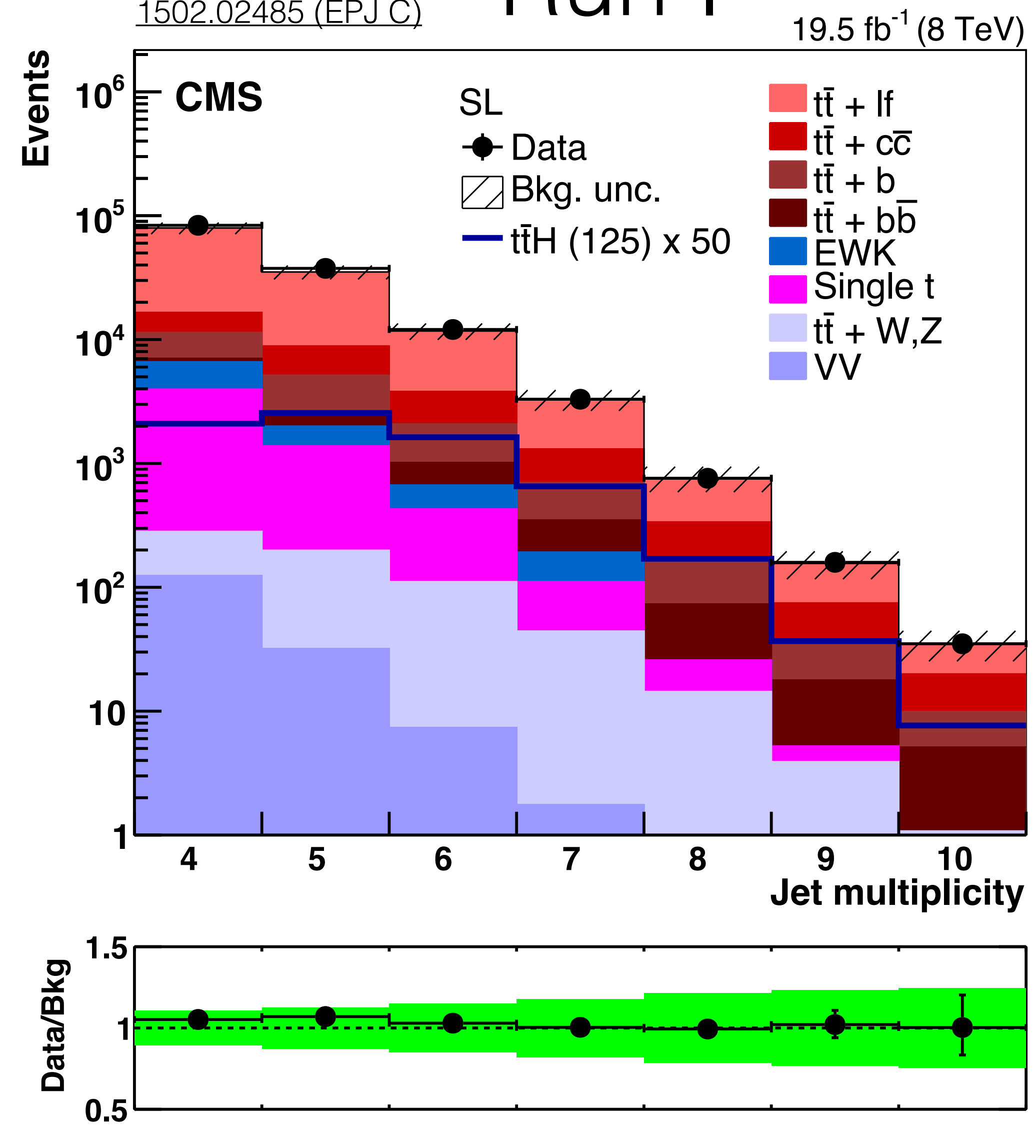
Require **1-2 isolated e/mu** → multi-jet QCD, but also fully-hadronic ttH(bb).

4 or more central jets p_t > 30(20) GeV, b-jets based on impact parameter and SV properties. → single top, EWK.

b-tagging for tt+heavy vs. light: N_b, likelihood discriminant → tt+light.

1502.02485 (EPJ C)

Run I



top quark background

Bevilacqua et al 2014

top quark pair background

$$\sigma^{\rm NLO}_{t\bar{t}b\bar{b}}({\rm LHC}_{13{\rm TeV}}, m_t = 173.5~{\rm GeV}, {\rm CT10}) = 1078.3^{+222.1(20\%)}_{-249.7(23\%)}~[{\rm fb}]\,,$$

$$\sigma^{\rm NLO}_{t\bar{t}jj}({\rm LHC}_{13{\rm TeV}}, m_t = 173.5~{\rm GeV}, {\rm CT10}) = 85.5^{-18.3(21\%)}_{-8.4(10\%)}~[{\rm pb}]\,.$$

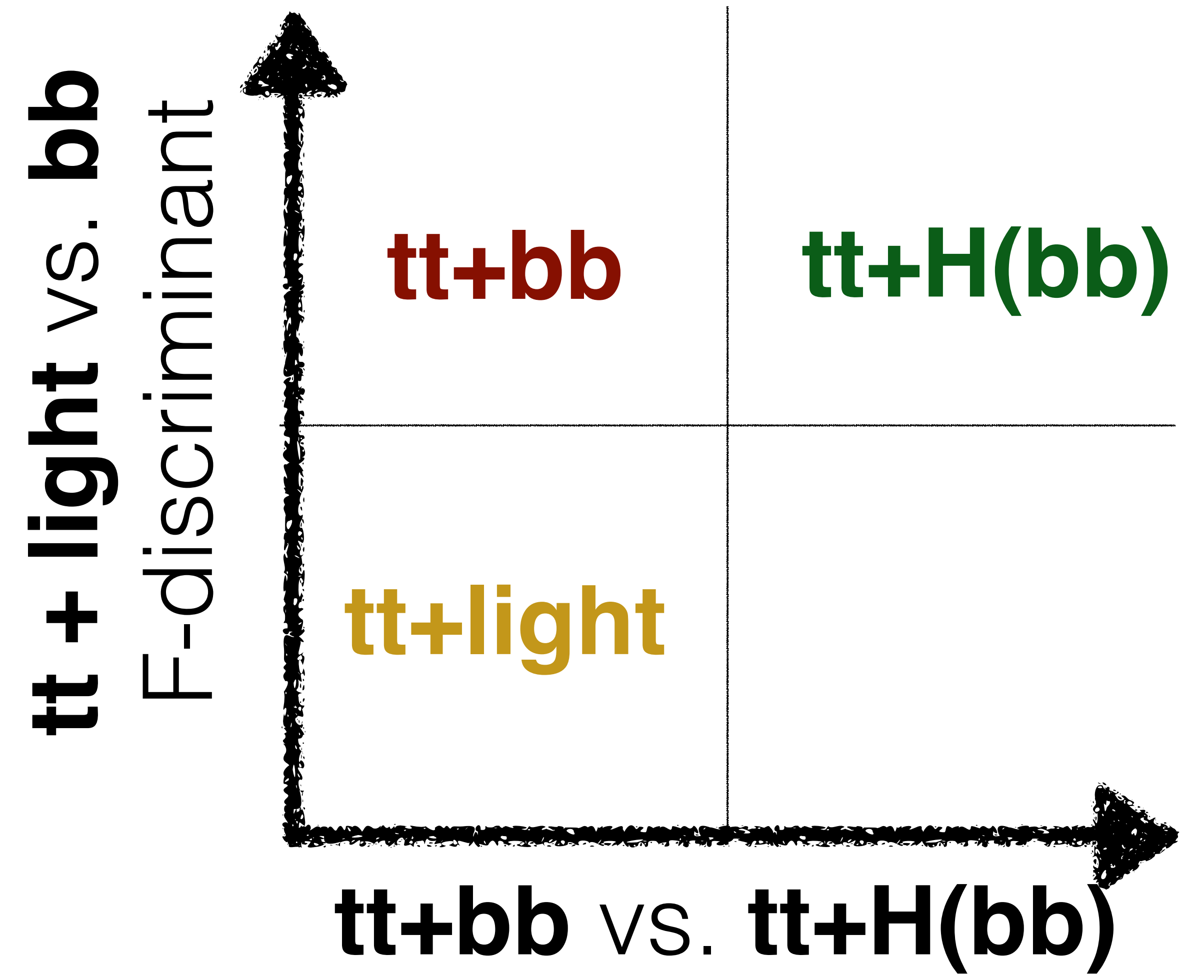
vs.

LHC XSWG

ttH: 508.5 [fb] ± ~10%

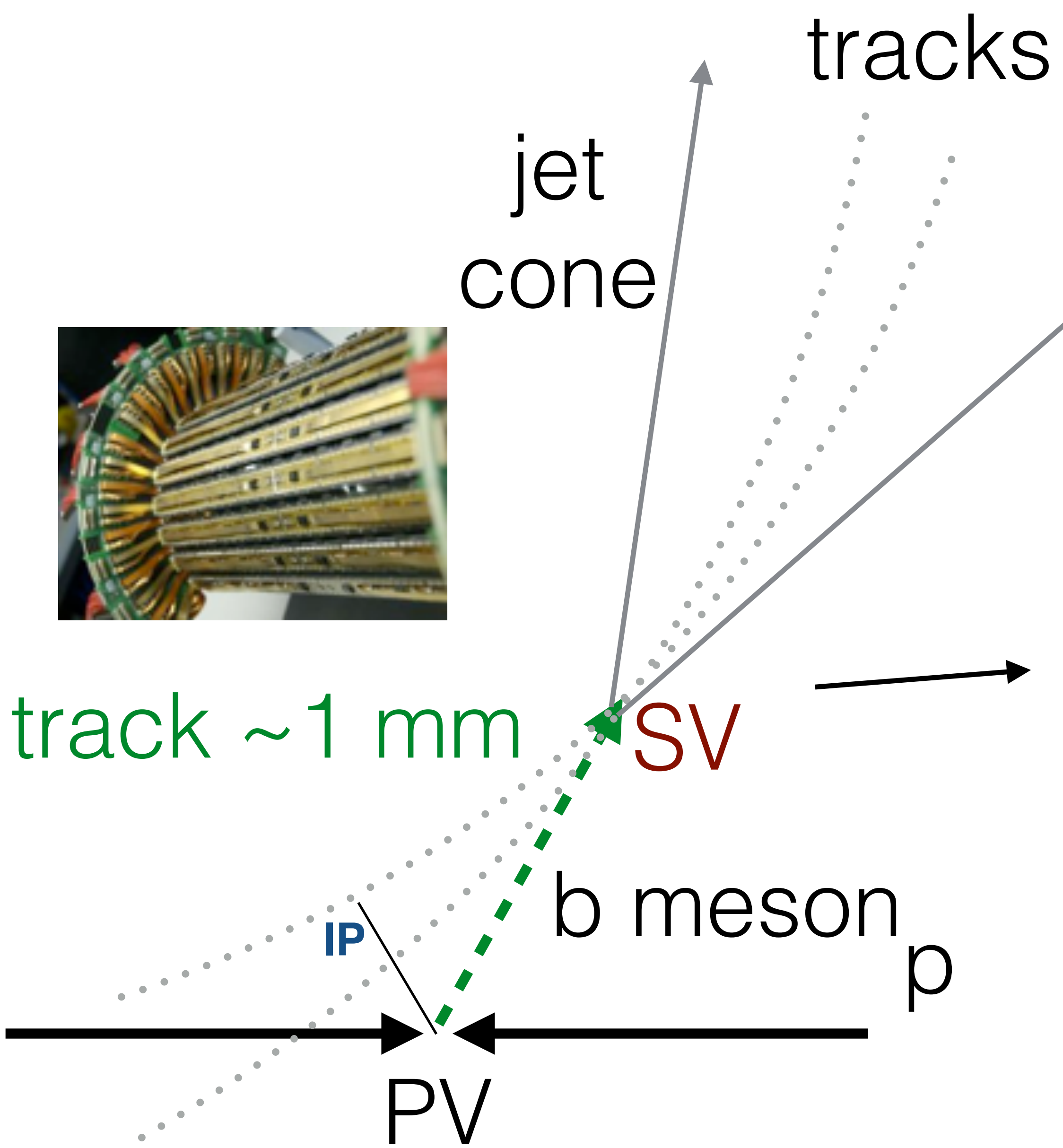
ttbb uncertainties cover expected signal!

Modelling uncertainties of tt+bb (gluon splitting).



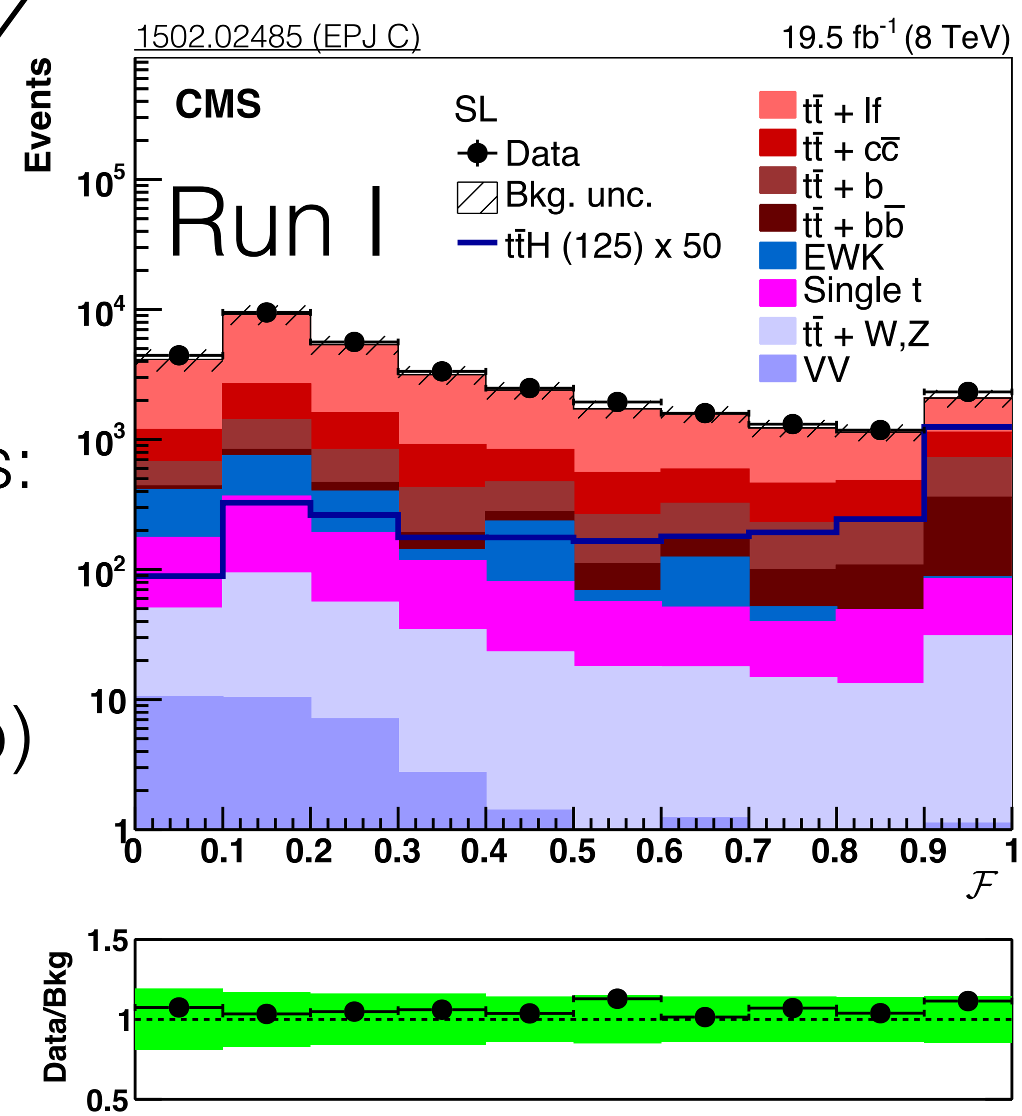
combined fit in σ(ttH) vs. σ(ttbb)/σ(ttjj) space!

b-tagging: tt+heavy vs. tt+light



in multi-jet, combine probabilities: p(jet | b) & p(jet | non-b)

$$\mathcal{F}(\vec{\xi}) = \frac{f(\vec{\xi}|t\bar{t}+\mathrm{hf})}{f(\vec{\xi}|t\bar{t}+\mathrm{hf}) + f(\vec{\xi}|t\bar{t}+\mathrm{lf})}.$$

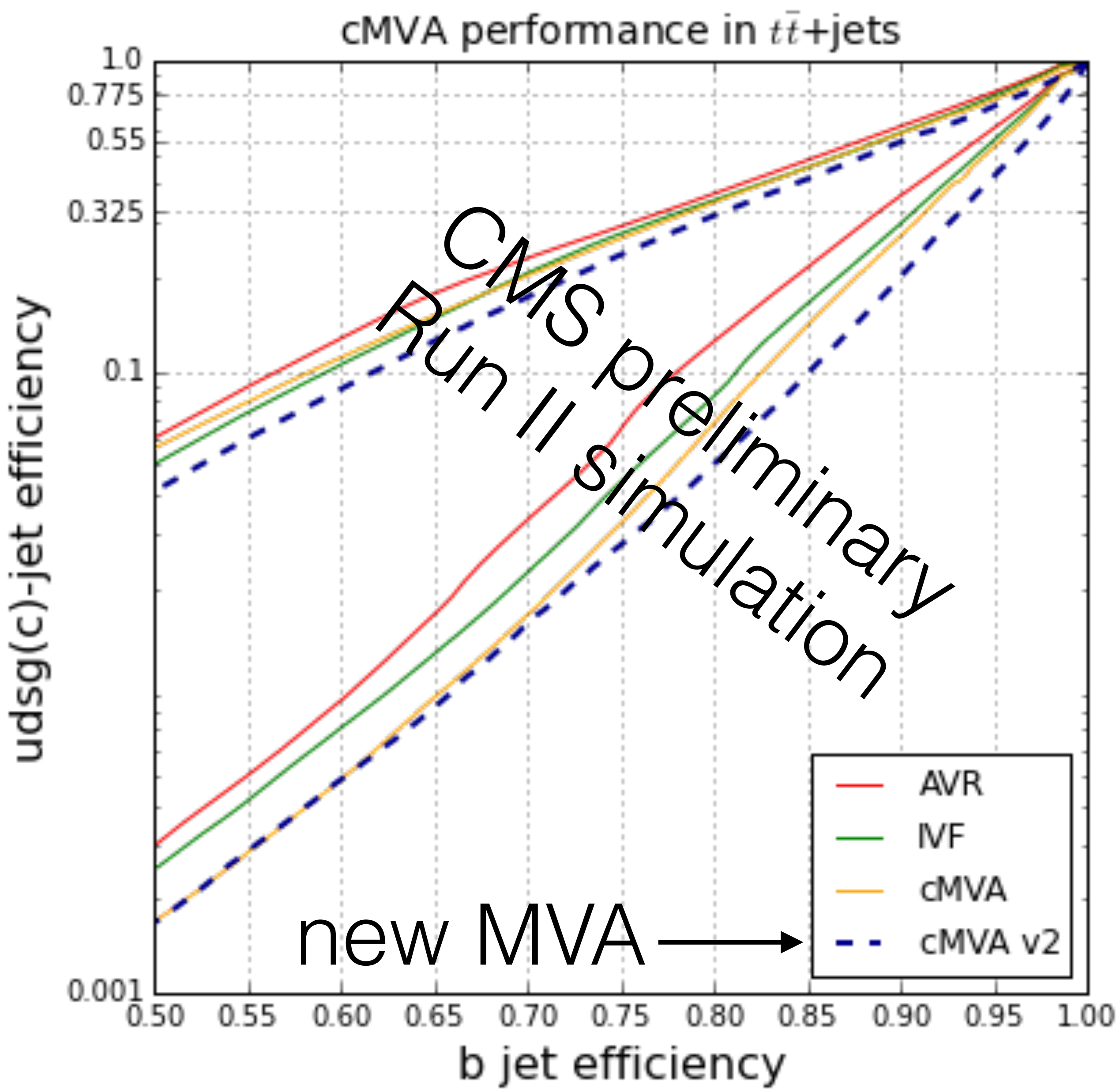


b-tagging in Run II

Studying a new “super-MVA” for b vs. light.

Combines different vertexing algorithms, likelihood-based jet probability, soft lepton taggers.

c mistag, high-eff b-tagging improved by **~20-30% (rel.)**.

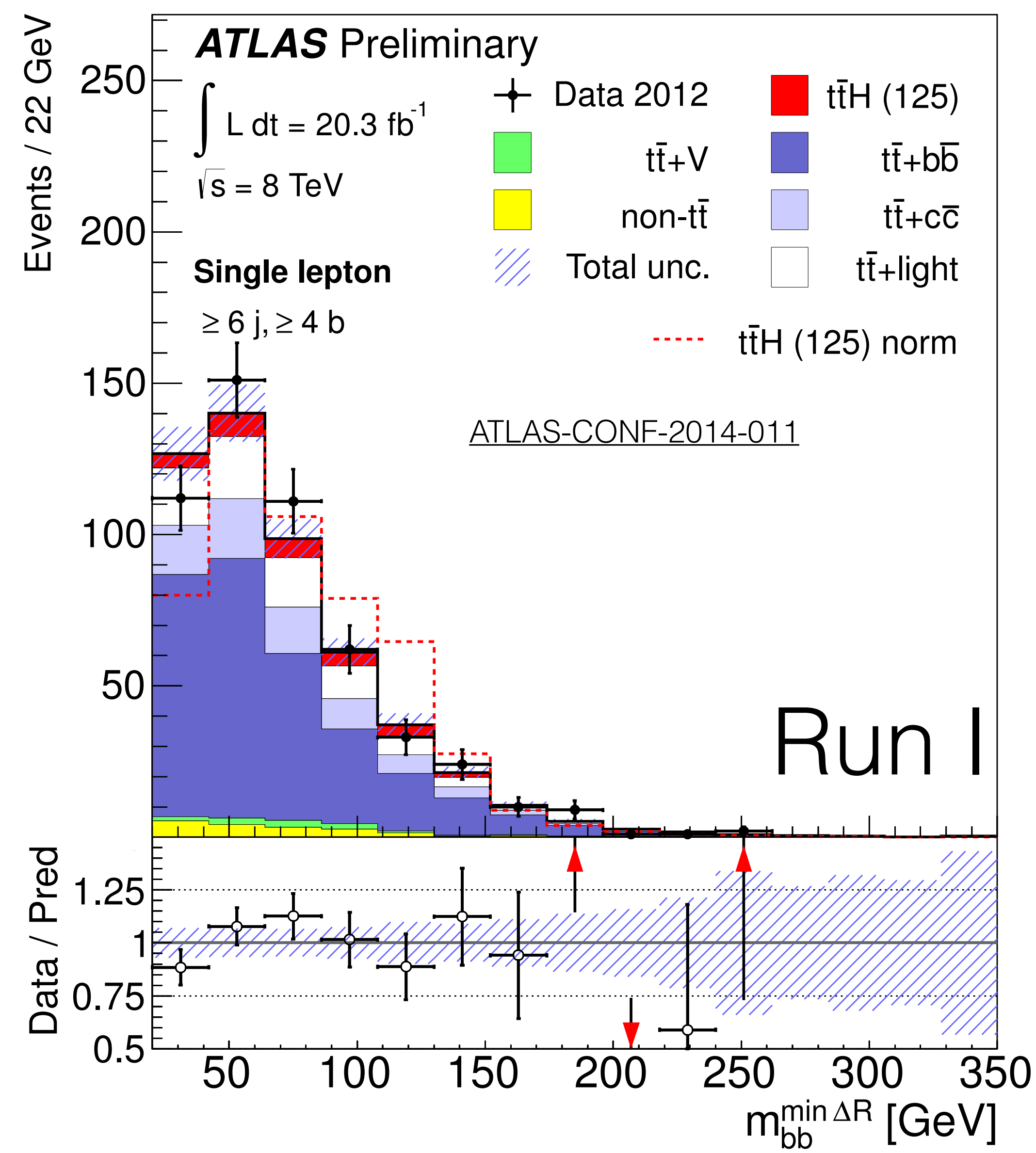


tt+bb vs. tt+H(bb)

We could reconstruct H(→bb) system invariant mass, but how to choose the right b-jet candidates?

H(→bb) smeared out due to combinatorics, cannot observe "Higgs peak".

Need to exploit H resonance properties experimentally.



Main problems

tt+jets without H dwarfs the signal by several orders of magnitude.

complicated final state: multi-jet, missing energy

Extracting the signal

Reconstructed event → probability density value from theoretical models: Matrix Element Method (MEM), **full use of kinematics** in LO.

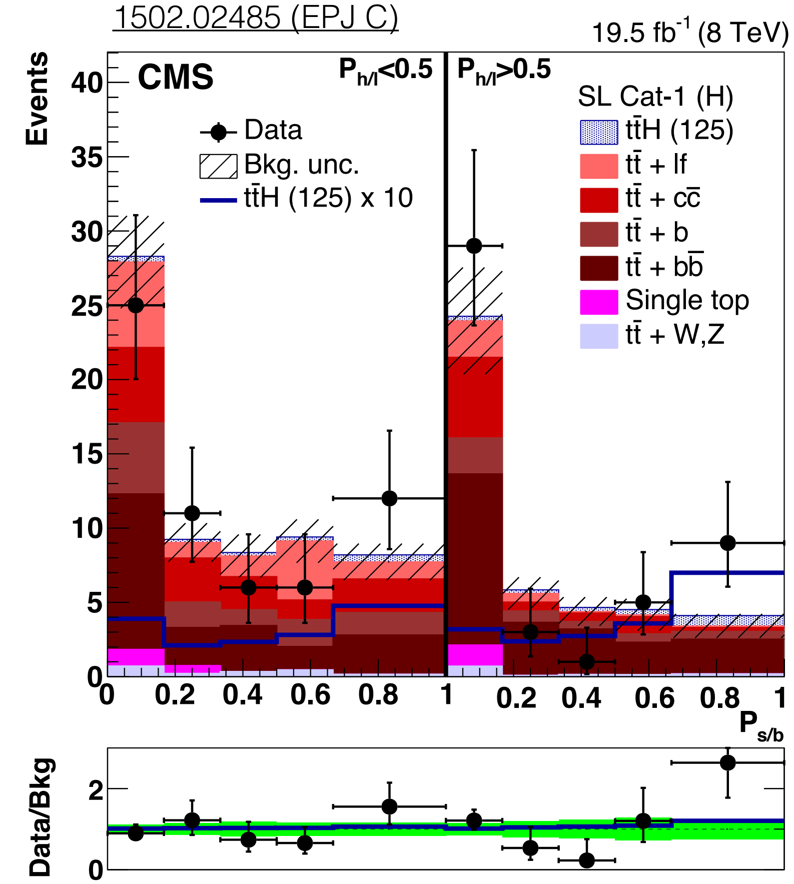
Unknown, poorly measured quantities → integrated out directly.

combinatorics for associating the multi-particle final state to the theoretical model.

High-level features: **neural networks** (NN) or **boosted decision tree** (BDT) -> exploit best simulation (NLO), but little physical insight, need high statistics.

Run I

1502.02485 (EPJ C)

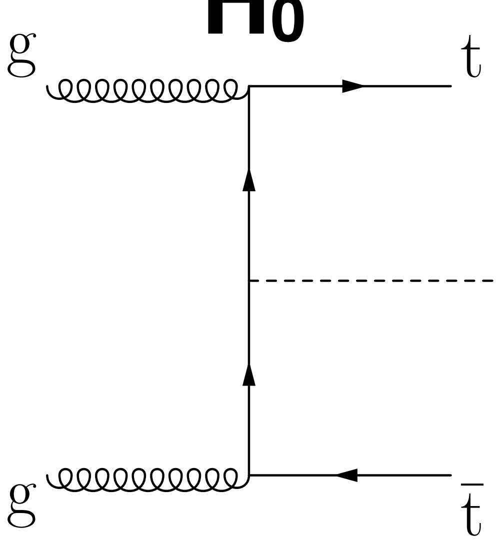


Matrix Element Method

Y - measured event properties

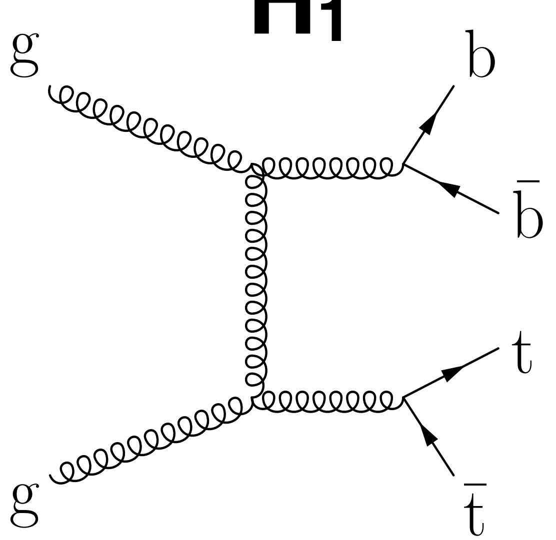
X - "true" parton-level quantities

H_0



or

H_1



likelihood ratio

$$P_{s/b} = \frac{w(\vec{y}|t\bar{t}H)}{w(\vec{y}|t\bar{t}H) + k_{s/b} w(\vec{y}|t\bar{t}{+}b\bar{b})}$$

parton level → reconstruction level

$$f(Y \mid H_0, \lambda) = 1/\sigma(H_0, \lambda) \int dX \int dx_a dx_b \; \Phi(x_a, x_b) \rightarrow \textbf{PDF}$$

$$|\mathscr{M}(X \mid H_0, \lambda)|^2 \; W(Y,X)$$

ME amplitude @ LO

exp. resolution

ETH Institute for Particle Physics

Reconstruction categories

Reconstructed event

Interpretation

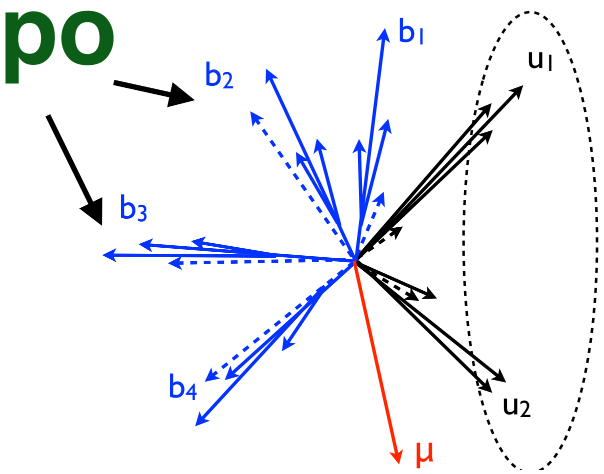
Integration

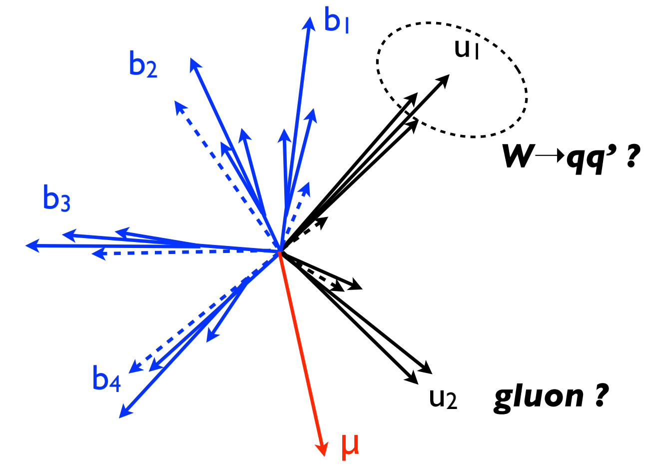
Discriminant **P**

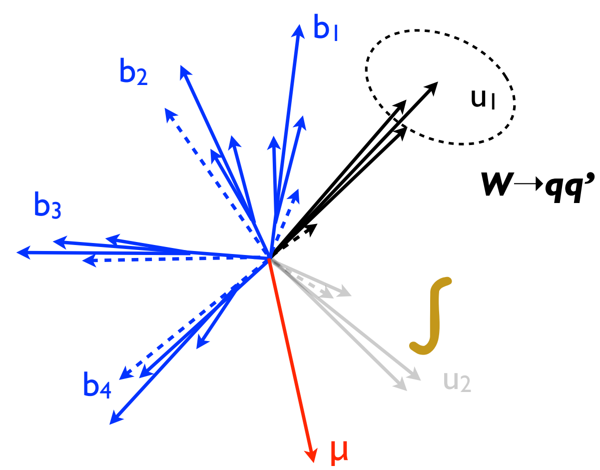
W→qq' in acceptance

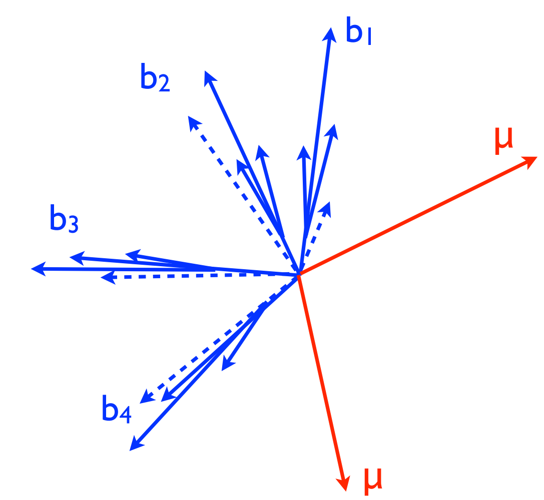
one light quark not reconstructed

hypo







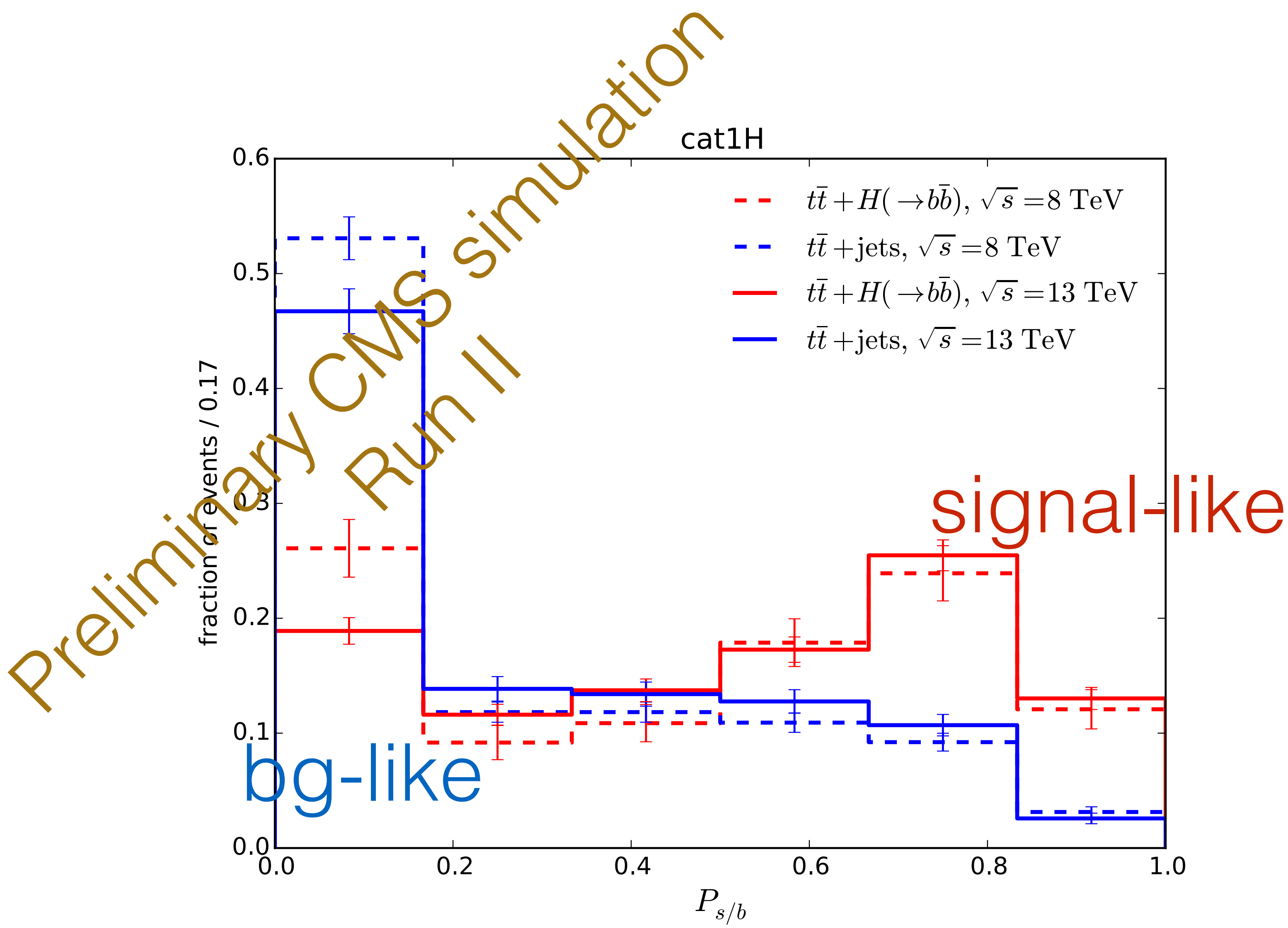


SL

DL

ME is just a discriminator, no need to be fully theoretically sound.

An approximation already does a very good job.



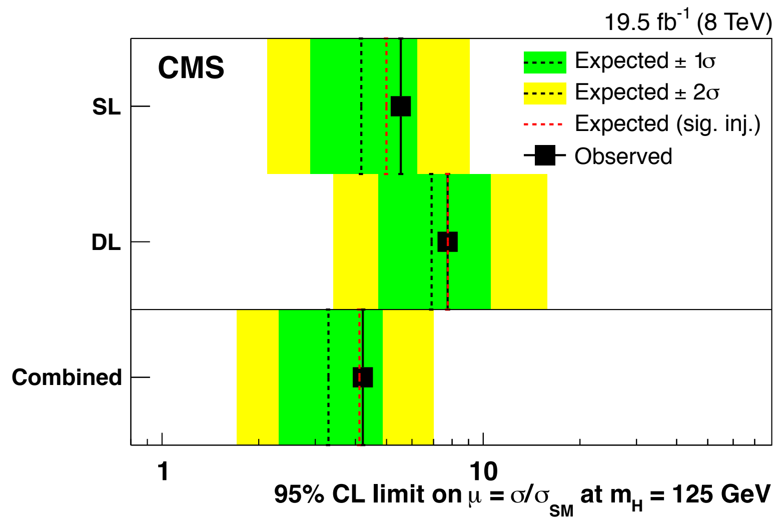
Results from LHC run I

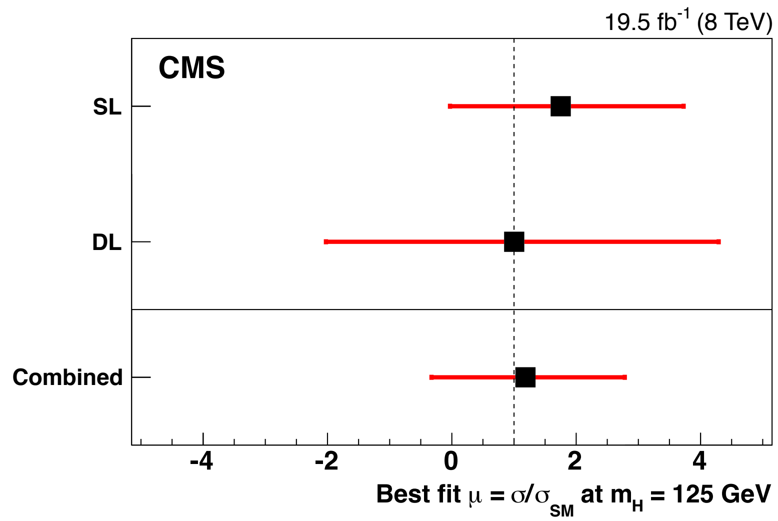
👉**CMS** ttH(bb) with **MEM** 20/fb: μ < 4.2 (3.3) obs (exp) @ 95%CL

CMS with **NN** 5+5/fb: μ < 5.8 (5.2)

ATLAS MEM(SL) + **NN** 20/fb: μ < 3.4 (2.2)

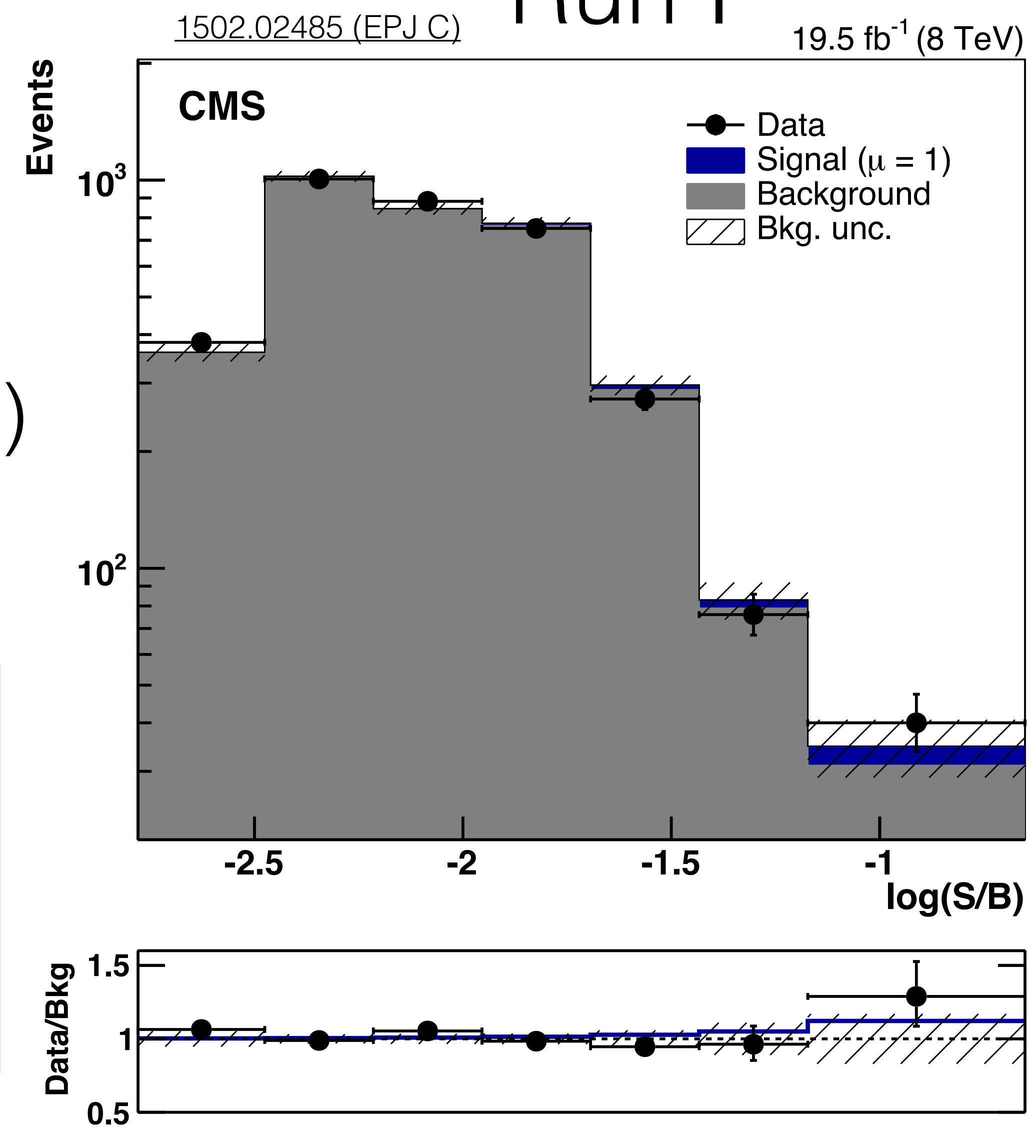
in Run II, first limits with ~10/fb.





Run I

1502.02485 (EPJ C)



Ongoing work

Improve MEM categorization: less constrained hypotheses - more separation at the cost of CPU time.

Combination of MEM discrimination with **NLO-based machine learning**.

Jet substructure can improve already existing limits: top tagging [1503.05921], higgs tagging of fat jets [1402.2657], direct integration with MEM/MVA.

Deploy MEM in **additional topologies**: H(ττ), fully-hadronic ttH(bb).

Fully exploit and improve **b-tagging to constrain tt+bb/cc/light**, quark-gluon discrimination.

Moving to LHC Run II

gg-dominated

ttH ~0.5085 pb (3.9x) favoured over

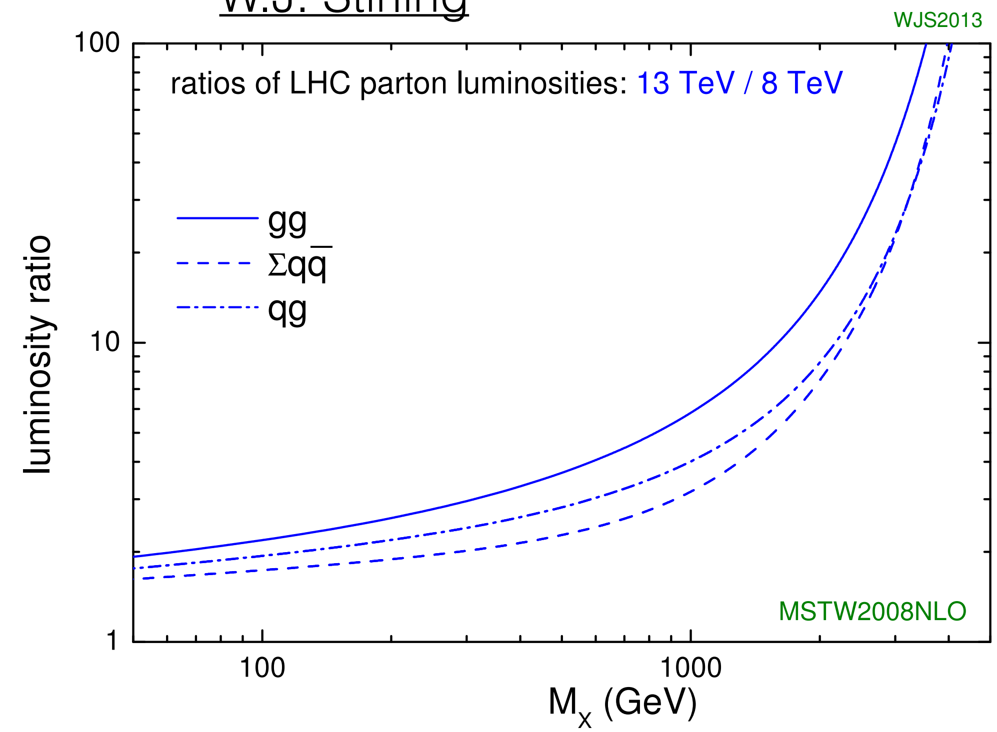
tt+jets ~832 pb (3.3x)

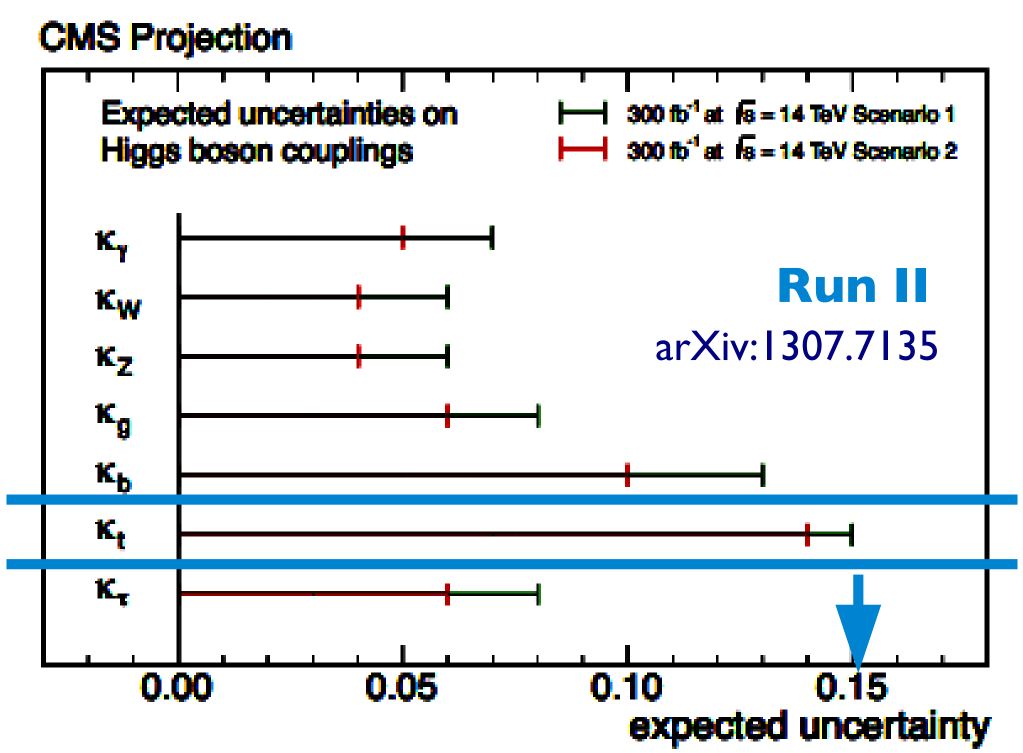
With 300/fb, coupling scale factor κ_t uncertainty 0.15.

Jet, lepton spectra harder in pt, higher multiplicity, increased signal acceptance.

But crucial to keep b-tagging performance with high pile-up, exploit jet substructure.

W.J. Stirling





Summary & outlook

ttH(bb) MEM proof of concept at 8 TeV: **μ < 4.2 (3.3) @ 95%**.

Need the **best possible interpretation of data** : matrix element method natural for high jet multiplicities, complements machine learning.

tt+bb NLO crucial, MEM naturally suited for systematically dominated (LHC 300/fb).

Extend to additional reco. hypotheses, use of jet substructure information.

LHC has only started to speak up about the Higgs, Run II will be crucial!

Backup

ME details

OpenLoops-based.

Verified against MadWeight at discriminator-level.

10-100 CPU-sec / event.

Integration via VEGAS, PDF using LHAPDF.

Detector effects encoded in transfer functions.

assumptions

No spin correlations.

Leading order only - no p_t etc.

Narrow-width t, W, H.

Only ggH.

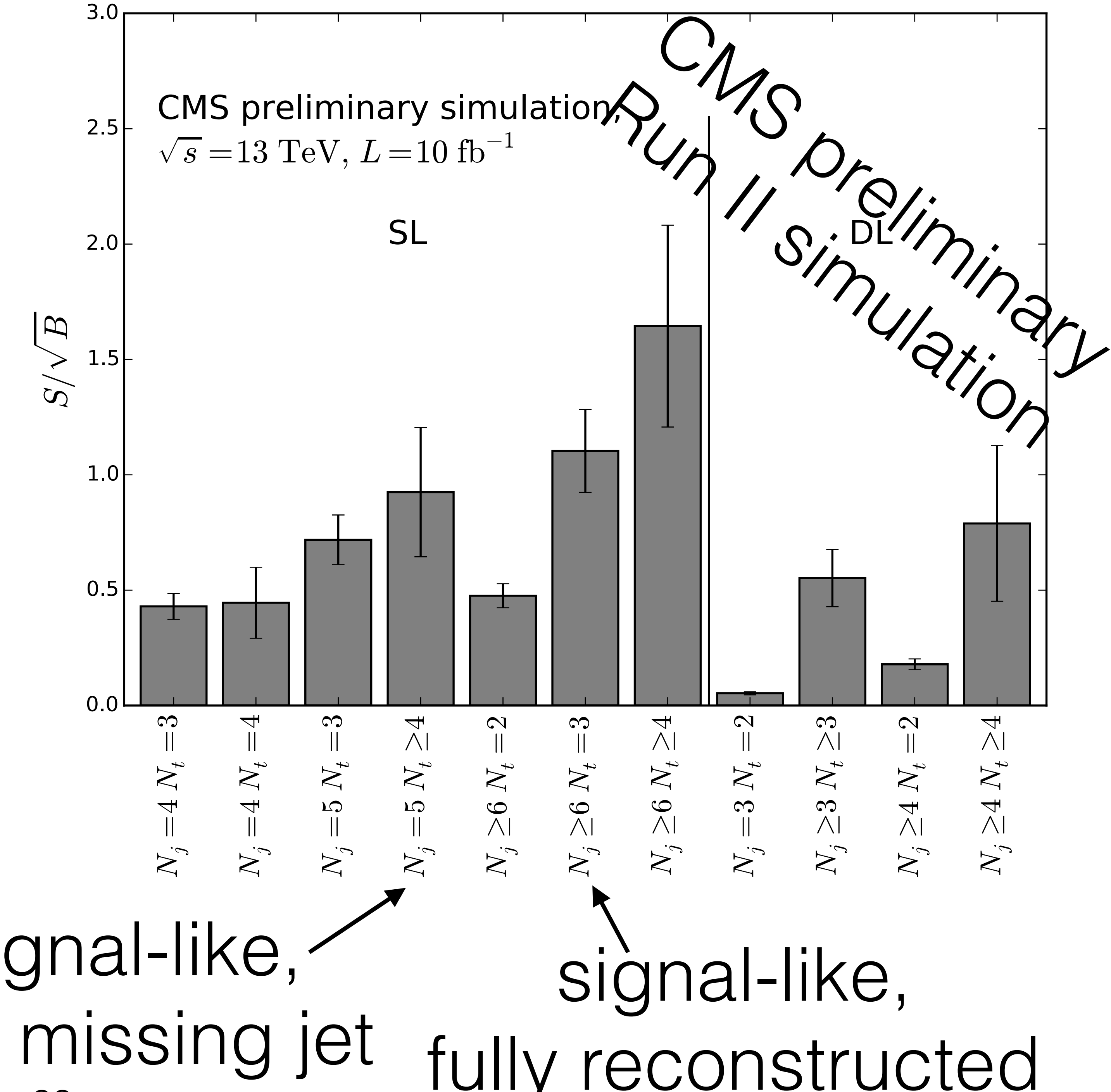
Require b-tagging of b-quarks (F-discriminant).

Expected rates

S/√B ~ 2% after lepton and jet selection.

Validate analysis in different jet / b-tag categories, combined fit.

MEM hypotheses optimized per category.



Uncertainties

Differential (shape) and **inclusive** (normalization) cross-sections affected. Fit **μ** within uncertainties. CL based on profile likelihood q(μ) properties

$$q(\mu) = -2\ln\left[\mathcal{L}(\mu,\hat{\theta}_\mu)/\mathcal{L}(\hat{\mu},\hat{\theta})\right]$$

tt+jets modelling: tt+heavy flavour cross-section (data → 15-20%), renormalization, factorization and resummation scale and functional form, shower recoil, PDF choice, MPI, FSR, top quark and top pair kinematics.

Jet energy **scale** and **resolution**:

B-tagging: discriminator distributions, not only total efficiencies need to be modelled well -> differential corrections.

Theory uncertainties on tt+H and tt+jets discriminant shape: crucial to describe gluon splitting.

Data & models

ATLAS and **CMS**: 5+20 fb^{-1} of 7 and 8 TeV collision data with lepton triggers -> 3.9 x 10^3 ttH events [CERN-2013-004], 55 (65) % gg->H at (N)LO in 8 TeV, more in 13 TeV

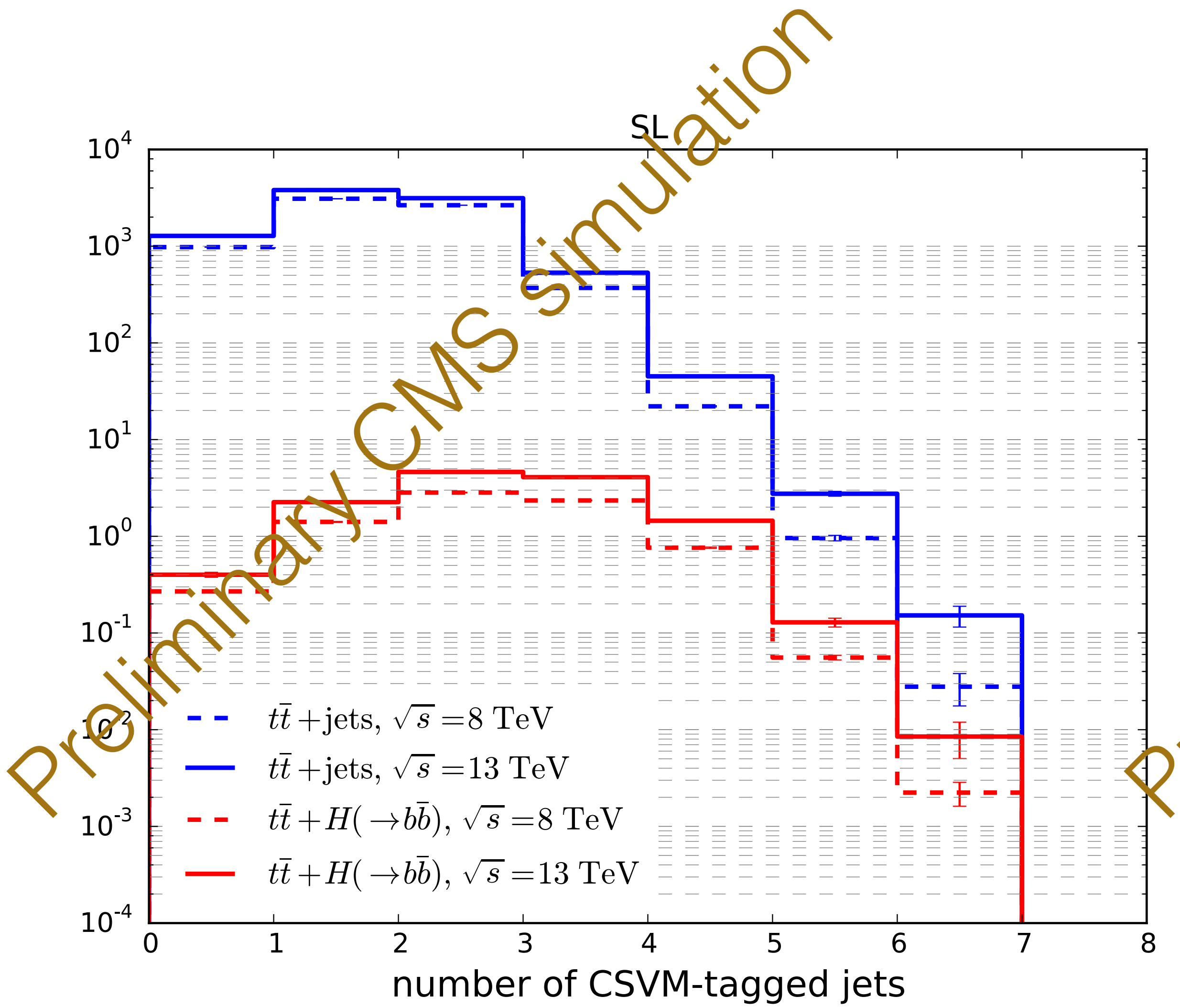
signal: LO->NLO simulation for tt+H, corrected to NNLO + leading log, Pythia (aMC@NLO at 13 TeV)

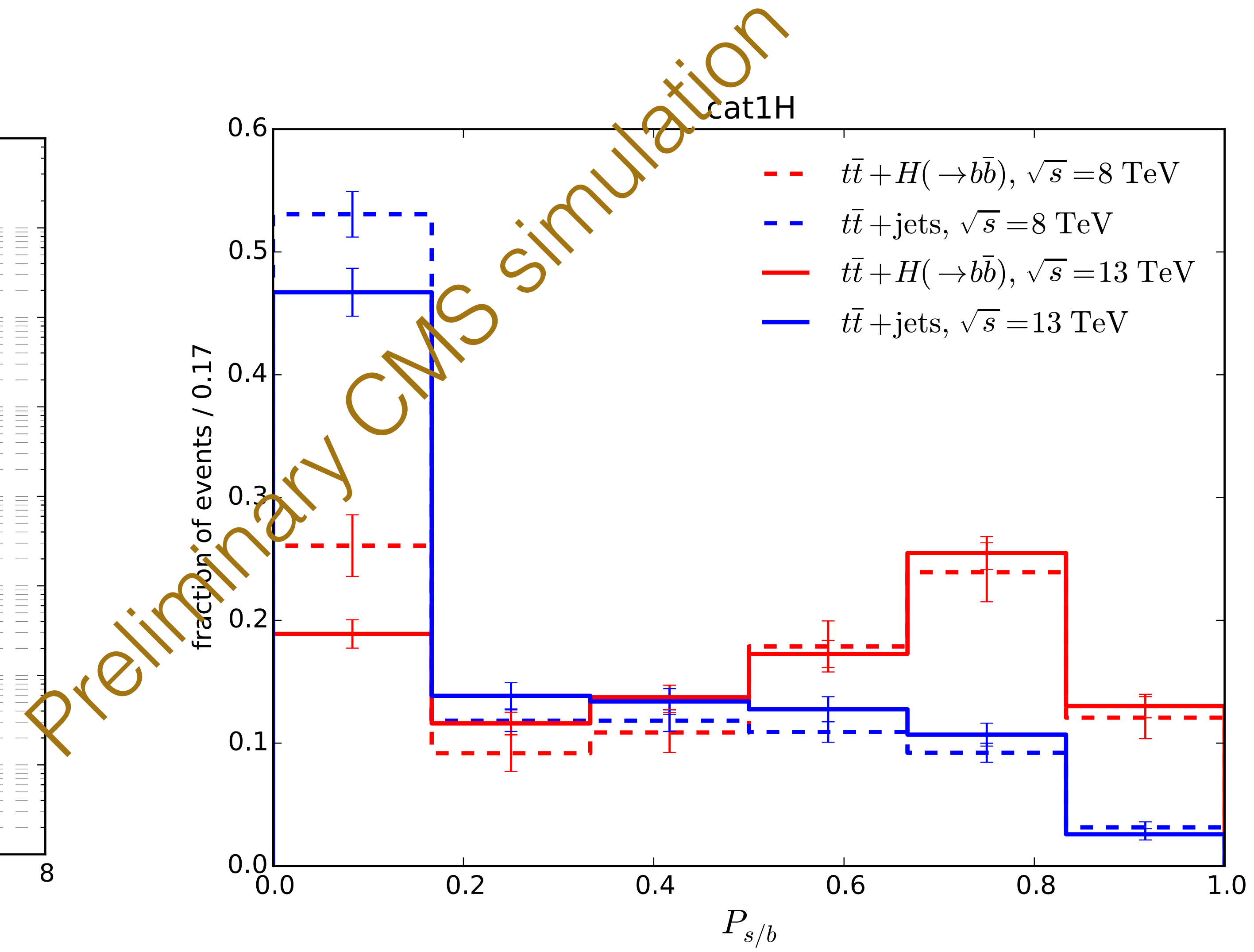
tt+bb: LO -> NLO, significant theoretical uncertainties, combine predictions from multiple generators, NLO recently available with Sherpa +OL

tt+cc, tt+light: LO (MG5), empirical corrections to top-antitop system and top quark kinematics.

minor: single-top: NLO (aMC@NLO), W/Z+jets, diboson

First look at 13 TeV simulation



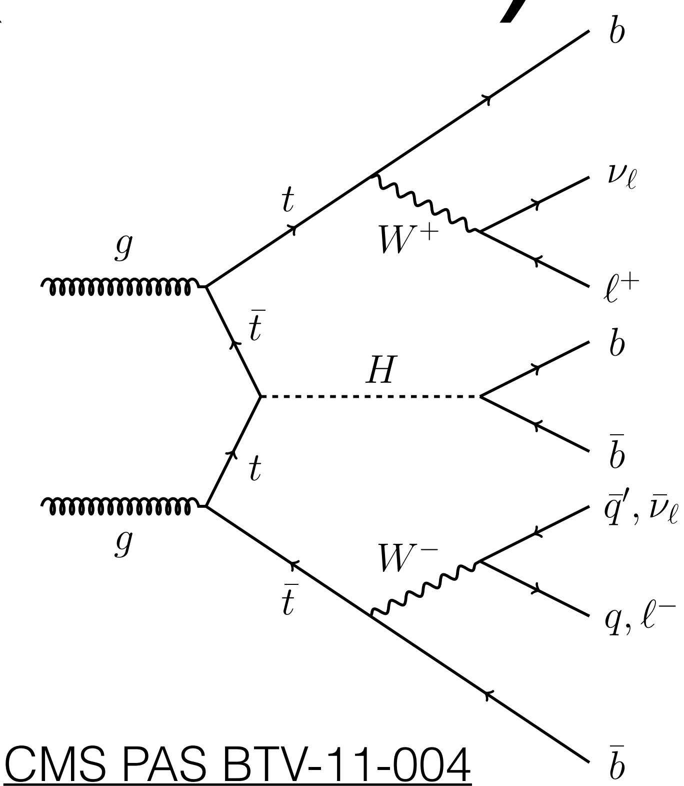


Case study: ttH(→bb)

H(125) → decays to b-quarks most abundant.

Multi-jet events with high b-quark multiplicity are a prolific source of ttH.

The experiments are good at detecting jets from b-quarks: **1-2% udsg** fake rate with **70% b-quark** efficiency.



CMS PAS BTV-11-004