Upper limits on the branching ratio for top decaying to a bottom and a charged Higgs boson

(ATLAS-CONF-2011-151)

Silje Hattrem Raddum

University of Oslo

5th January 2012

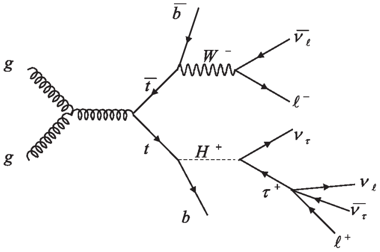
Outline

Some theory...

- Simple extension to the Higgs sector: the two Higgs-doublet model (2HDM) with five physical states of which two are charged (H^+ and H^-)
- An example of a 2HDM is the Minimal Supersymmetric Standard Model (MSSM)
- In the MSSM, a light H^+ decays primarily to $c\bar{s}$, bW^+ and $\tau^+\nu$
- Branching ratios depend on $\tan\beta$ (ratio of the two Higgs doublet expectation values) and m_{H^+}
- For $\tan\beta > 3$, the dominating decay mode is $\tau^+\nu$ (90%)
- This analysis is based on the assumption that $\tan\beta$ is large

Analysis

- 1.03 fb^{-1} of pp collision data recorded at $\sqrt{s} = 7$ TeV with the ATLAS detector
- Dominant charged Higgs production mode at the LHC is $t \rightarrow bH^+$
- Analysis of $t\bar{t}$ decays with leptonically decaying τ in the final state
- Assume $\mathcal{B}(H^+ \rightarrow \tau\nu) = 1$
- $\mathcal{B}(H^+ \rightarrow \tau\nu \rightarrow l + N\nu) \approx 35\%$, $\mathcal{B}(W \rightarrow l + N\nu) \approx 25\%$

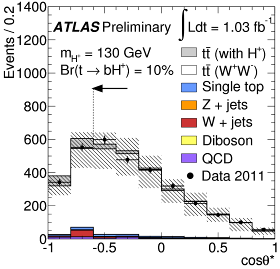


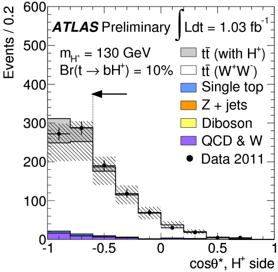
Discriminating variables: the invariant mass m_{bl}

Or more conveniently:

$$cos\theta_l^* = \frac{2m_{bl}^2}{m_{top}^2 - m_W^2} - 1 \simeq \frac{8E_bE_l(1 - \cos\theta_{bl})}{m_{top}^2 - m_W^2}$$

- If the top-quark decay is mediated through a H^+, the b-quark usually has a smaller momentum (given that $m_H > m_W$)
- A light charged lepton from a τ-decay is likely to have a smaller momentum than a lepton coming directly from a real W boson
- Signal events thus have $cos\theta_l^*$ values closer to -1.
- Use $cos\theta_l^*$ to define signal and control region (left plot: single-lepton, right plot: di-lepton)

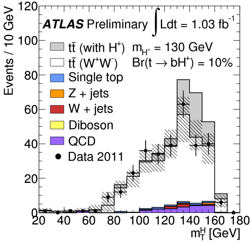




Discriminating variables: m_T^H (single-lepton)

Charged Higgs boson transverse mass:

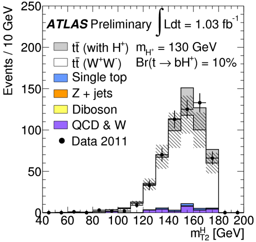
- Three final-state neutrinos $\rightarrow (p^{miss})^2 \neq 0$
- Maximize over the invariant mass $m_{H^+}^2 = (p^l + p^{miss})^2$ by differentiating with respect to $p_{\parallel}^{miss}$
- Constraint: $m_{top}^2 = (p^{miss} + p^l + p^b)^2$
- Finally, we obtain:
  $(m_T^H)^2 = \left(\sqrt{m_{top}^2 + (\vec{p}_T^{\,l} + \vec{p}_T^{\,b} + \vec{p}_T^{\,miss})^2} - p_T^b\right)^2 - (\vec{p}_T^{\,l} + \vec{p}_T^{\,miss})^2$
- The following holds: $m_{top} > m_T^H > m_{H^+}$
- Use m_T^H to search for H^+ in the single-lepton channel



Discriminating variables: m_{T2}^{H} (di-lepton)

Generalized charged Higgs boson transverse mass:

- Di-lepton case: two leptons and missing energy on both sides of event - more complicated
- Set of constraints: m^2_{top} on each side of the event, m^2_W, $\vec{p}_T^{miss}$, invariant mass of a neutrino = 0
- Two free variables: p^{H^+} and $p^{\nu W}$
- Maximize w.r.t. to the free variables (nasty math):
  $m_{T2}^{H} = \max_{\{constraints\}} [m_T^H(\vec{p}_T^{H^+})] \quad , \quad \vec{p}_T^{H^+} = \vec{p}_T^{l} + \vec{p}_T^{miss}$
- The following holds: $m_{top} > m_{T2}^{H} > m_{H^+}$
- Use m_{T2}^{H} to search for H^+ in the di-lepton channel



Top cross section and branching ratios

- Can not rely on predicted cross-section for $t\bar{t} \rightarrow b\bar{b}W^+W^-$ in the presence of H^+
- Use control region enriched with SM-like $t\bar{t}$ events to measure the fiducial cross section σ_{bbWW}
- Control region:
  - $-0.2 < cos\theta_l^* < 1$ in single-lepton analysis
  - $-0.4 < cos\theta_l^* < 1$ in di-lepton analysis

Branching ratios in the presence of H^+:

$$B \equiv \mathcal{B}(t \rightarrow bH^+) \quad \rightarrow \quad \mathcal{B}(t \rightarrow bW^+) = 1 - B$$

$$N_{t\bar{t}} = \Big[\underbrace{(1-B)^2}_{bbWW} + \underbrace{2B(1-B)}_{bbWH} + \underbrace{B^2}_{bbHH}\Big] N_{t\bar{t}}$$

The profile likelihood method

- The profile likelihood is the likelihood maximized over the nuisance parameters (denoted θ) given a value of the parameter of interest (POI), here denoted μ:

$$\max_\theta \mathcal{L}(\mu, \theta)$$

(It's often more practical to work with the negative log-likelihood (NLL). Minimizing the NLL is equivalent to maximizing the likelihood)

- True nuisance parameters can vary freely during minimization
- Additional systematic uncertainties are introduced through pseudo-measurement terms in the likelihood (the external constraints introduce penalizing terms to the NLL)

Example: systematic uncertainty on luminosity measurement $\tilde{L}$

$$\mathcal{L}(\mu) = \mathcal{P}(n_i^{obs}|n_i^{exp})\mathcal{G}(L|\tilde{L}, \sigma_L) \quad \rightarrow \quad -\ln \mathcal{L}(\mu) = -\ln \mathcal{P}(n_i^{obs}|n_i^{exp}) + \frac{(L-\tilde{L})^2}{2\sigma_L^2}$$

- $\tilde{L}$ is the measured luminosity, σ_L is the uncertainty
- The penalizing Gaussian term increases as $|L - \tilde{L}|$ increases
- Note that n^{exp} depends on L and μ

Implementation

Earlier searches for H^+ suggest that $\mathcal{B}(t \to bH^+) < 10\%$. Hence, the contribution from $t\bar{t} \to b\bar{b}H^+H^-$ is small. Due to low statistics, which made minimization unstable, it was therefore decided to take out the HH-signal.

- POI: $B \equiv \mathcal{B}(t \to bH^+)$
- *HW*-signal scale factor $\sigma_{bbHW} = \sigma_{bbWW} \times \frac{2B}{1-B}$
  (*HH*-signal scale factor $\sigma_{bbHH} = \sigma_{bbWW} \times \frac{B^2}{(1-B)^2}$)
- True nuisance parameter: σ_{bbWW}
- Psuedo-nuisance parameters: systematic uncertainties from theory, energy scales, triggering, reconstruction, pile-up etc...
- The single- and di-lepton channels are orthogonal, so that they can easily be combined
- Assume all nuisance parameters to be 100% correlated across the two channels

Implementation

In ROOFIT/HISTFACTORY each channel was implemented as two "channels"; one for the signal region and one for the control region

Signal region (SR):

We use the m_T^H and m_{T2}^H distributions, where each bin represents a counting experiment. The product defines the likelihood function

$$\mathcal{L}_{SR}(B) = \prod_i^{bins} \mathcal{P}(n_i^{obs,SR}|n_i^{exp,SR})$$

Control region (CR):

We are not interested in the shape of the distribution, but the fiducial cross section σ_{bbWW}. The control region "channel" thus represents a pure counting experiment

$$\mathcal{L}_{CR}(B) = \mathcal{P}(n^{obs,CR}|n^{exp,CR})$$

Introducing the systematic uncertainties (constraints), the resulting likelihood function becomes

$$\mathcal{L}(B) = \max_{\sigma_{bbWW},\theta} \mathcal{L}(B, \sigma_{bbWW}, \theta) = \mathcal{P}(n^{obs,CR}|n^{exp,CR}) \prod_i^{bins} \mathcal{P}(n_i^{obs,SR}|n_i^{exp,SR}) \prod_j^{syst} p(\tilde{\theta}_j|\theta_j)$$

Likelihood-ratio test statistics

Wilks theorem:

The log likelihood-ratio (LLR) for a nested model will be asymptotically χ^2 distributed when the sample size approaches ∞

- The test-statistic t_μ for a hypothesized value of μ is defined as

$$t_\mu = -2\ln\lambda(\mu) = -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu},\hat{\theta})}$$

- $\hat{\hat{\theta}}$ in the numerator denotes the value of θ that maximizes $\mathcal{L}$ for the given test value μ
  - conditional estimator of θ
- $\hat{\theta}$ and $\hat{\mu}$ in the denominator are the values of θ and μ that maximizes $\mathcal{L}$
  - unconditional estimators ("best fit" values) of μ and θ
- Thus, $M(\mu,\hat{\hat{\theta}})$ is a nested model of $M(\hat{\mu},\hat{\theta})$ with one free parameter less
- For high statistics, the p-value for a given test value μ is given by

$$p_\mu = 2[1-\Phi(\sqrt{t_\mu})]$$

The Asimov dataset

- The Asimov dataset is an artificial dataset constructed from a model - a pseudo-experiment with no statistical fluctuations
- b-only Asimov: profile nusiance parameters for $\mu = 0$
- s+b Asimov: profile nusiance parameters for $\mu = 1$
- Profiling the nuisance parameters means to perform a likelihood fit to data (maximize) for the given value of μ
- We then obtain a model $M(\mu, \hat{\hat{\theta}}(\mu))$ from which we can construct an Asimov dataset
- We use the b-only Asimov for the extraction of expected upper limit and error bands
- The s+b Asimov is used for the extraction of expected p_0 values (to test the b-only hypothesis)

More on test-statistics and the Asimov dataset:

"Asymptotic formulae for likelihood-based tests of new physics"
Glen Cowan, Kyle Cranmer, Eilam Gross, Ofer Vitells
http://arxiv.org/abs/1007.1727

Profiling is conservative: expected limit and p_0

- In presence of signal, the b-only Asimov dataset may be an over-estimate of the background
- In such a case, the apparent sensitivity becomes worse (conservative)
- Without signal, the s+b Asimov might under-estimate the signal strength (μ)
- In such a case, the expected agreement with the b-only hypothesis (p_0) becomes higher (conservative)
- It might be useful to compare with results obtained with a non-profiled Asimov dataset (i.e. all nuisance parameters set to their nominal value)

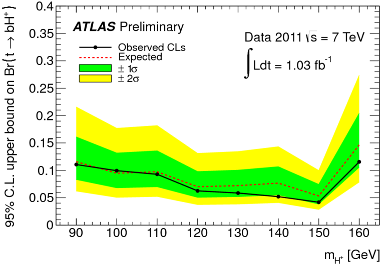
Profiling is conservative: observed limit and p_0

- Similar effects applies to the observed limit and p_0
- For upper limits, we use an iterative method for increasing μ until CLs is at the 5% level
- Without signal in data, nuisance parameters will attempt to pull down the modeled signal, weakening the signal strength (conservative)
- Need larger μ to reach the 5% level
- For p_0, we simply test the unconditional (best) fit against the conditional fit for $\mu = 0$
- If there is signal in data, the conditional fit will attempt to pull the nuisance parameters in order for background to compensate for the observed signal
- The result is less difference between the unconditional and conditional fits (conservative)

This somewhat conservative procedure is chosen since it gives approximate coverage of the POIs for any values of the nuisance parameters

Single-lepton limits

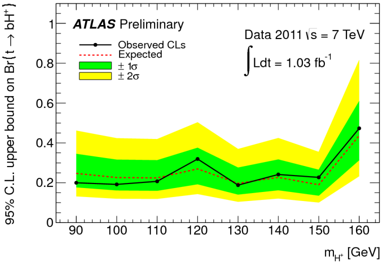
- Fitted values of σ_{bbWW} lie between 0.99 and 1.03 times the SM prediction, with uncertainties in the range 2-3%
- Note that for $m_{H^+} = 160$ GeV, the b-jets are usually too soft to pass the p_T cut at 20 GeV - low sensitivity



m_{H^+} (GeV)	90	100	110	120	130	140	150	160
observed expected	11.1% (11.6%)	9.9% (9.5%)	9.3% (9.7%)	6.3% (7.0%)	5.8% (7.2%)	5.2% (7.7%)	4.2% (5.3%)	11.6% (14.6%)

Di-lepton limits

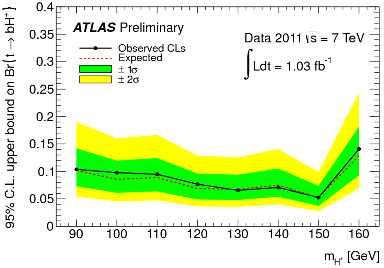
- A downward fluctuation in the CR yields fitted values of σ_{bbWW} between 0.78 and 1.06 times the SM prediction, with uncertainties in the range 5-25%



m_{H^+} (GeV)	90	100	110	120	130	140	150	160
observed expected	20.0% (24.7%)	19.2% (22.6%)	20.7% (22.4%)	32.0% (26.9%)	18.8% (19.8%)	24.2% (22.6%)	22.7% (19.0%)	47.3% (43.7%)

Combined limits

- p_0-values range between 26% and 50%.
- No indication of an H^+-like signal is found
- Tevatron experiments have placed upper limits on B in the 15-20% range
- Values of tan β larger than 30-56 are excluded for 90 GeV $< m_{H^+} <$ 140 GeV in the context of the m_h^{max} scenario of the MSSM



m_{H^+} (GeV)	90	100	110	120	130	140	150	160
observed expected	10.4% (10.2%)	9.8% (8.5%)	9.5% (8.9%)	7.7% (6.9%)	6.6% (6.7%)	7.1% (7.5%)	5.2% (5.2%)	14.1% (12.9%)

Thank you!

Backup material

Alternative test-statistic

- To take into account that $B \geq 0$, we use the alternative test-statistic $\tilde{t}_\mu$
- For data yielding $\hat{\mu} < 0$, the best *physical* fit value of μ is 0

$$\tilde{t}_\mu = \begin{cases} -2\ln \frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\theta})} & \hat{\mu} \geq 0 \\ -2\ln \frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \end{cases}$$

- When setting an upper limit, it is only meaningful to test values of μ greater than the value of μ most compatible with the data obtained ($\hat{\mu}$)
- The upper limit test-statistic is thus defined as

$$\tilde{q}_\mu = \begin{cases} \tilde{t}_\mu & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$