Search for $A/H \to t\bar{t}$

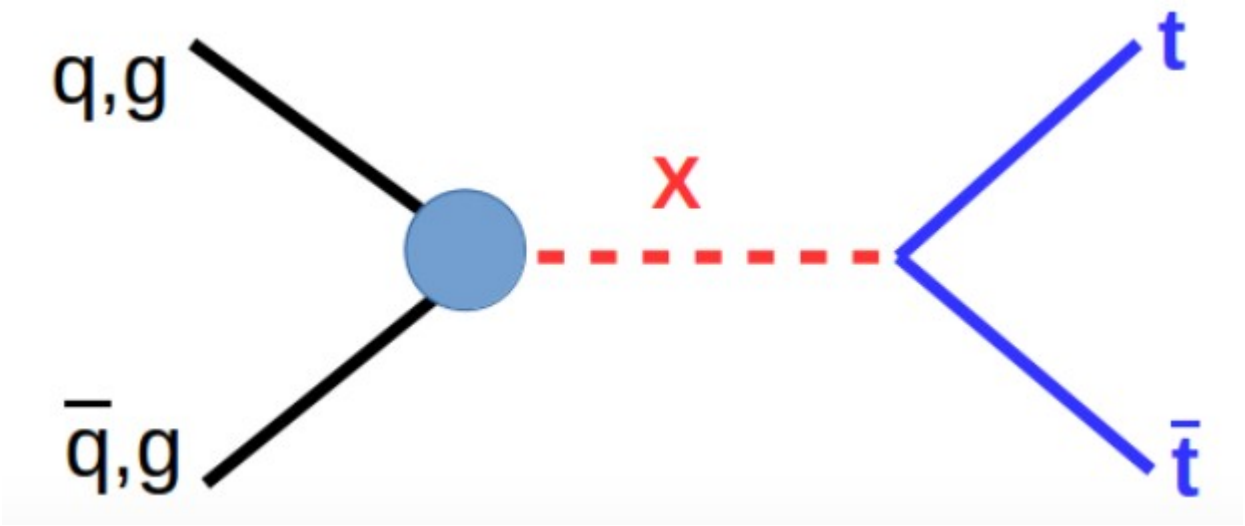
- PRL.119(2017)191803

Jike Wang (PI, Universitaet Heidelberg)

CLHCP 2017 Workshop, Nanjing

The Overall Picture

- Search for massive resonances decaying to $t\bar{t}$
- Cover various benchmark models over a large mass range



- Resonances

Z' boson (narrow, spin 1)

Kaluza-Klein gluon (wide, spin 1)

Kaluza-Klein graviton (narrow, spin 2)

} negligible interference

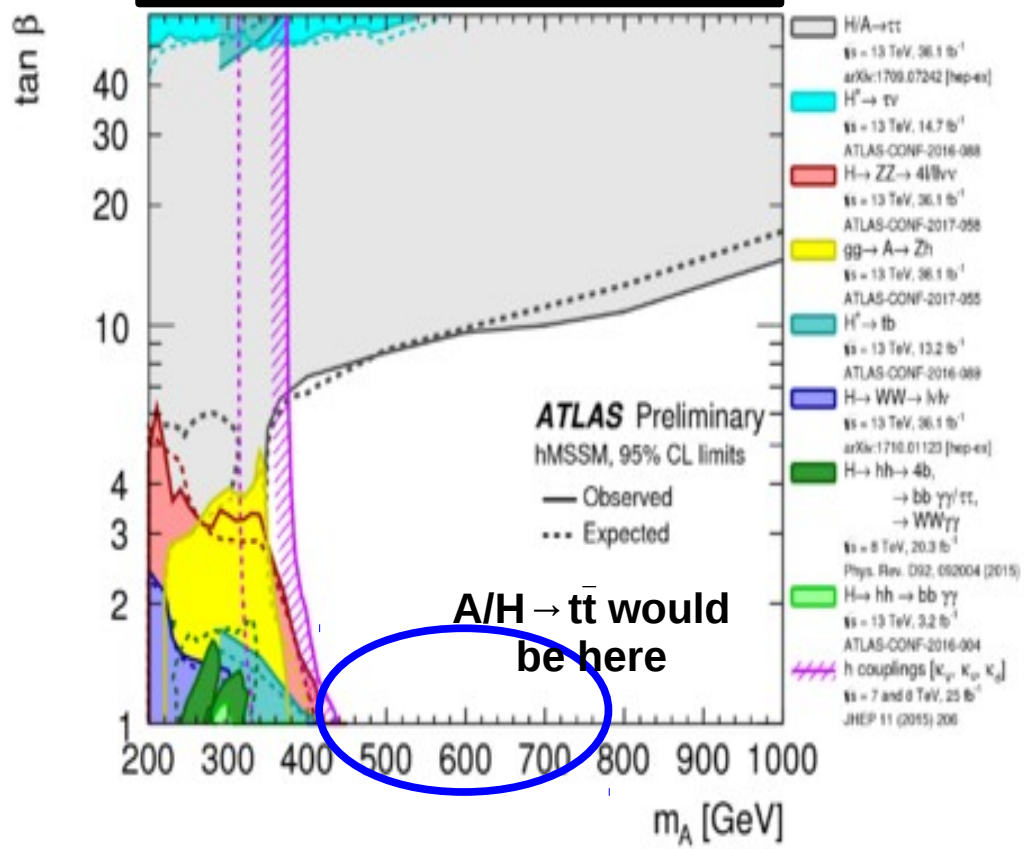
Heavy Higgs boson (wide, spin 0) ⇒ large interference

Physics motivation

- $A/H \to t\bar{t}$ is very important for high mass regions (above $2m_t$). The only channel could be sensitive to low tanβ (at high mass)

- gg production ⇒ strong interference between signal and the dominated background (SM $t\bar{t}$). Generate very complicated shapes

Combined summary plots

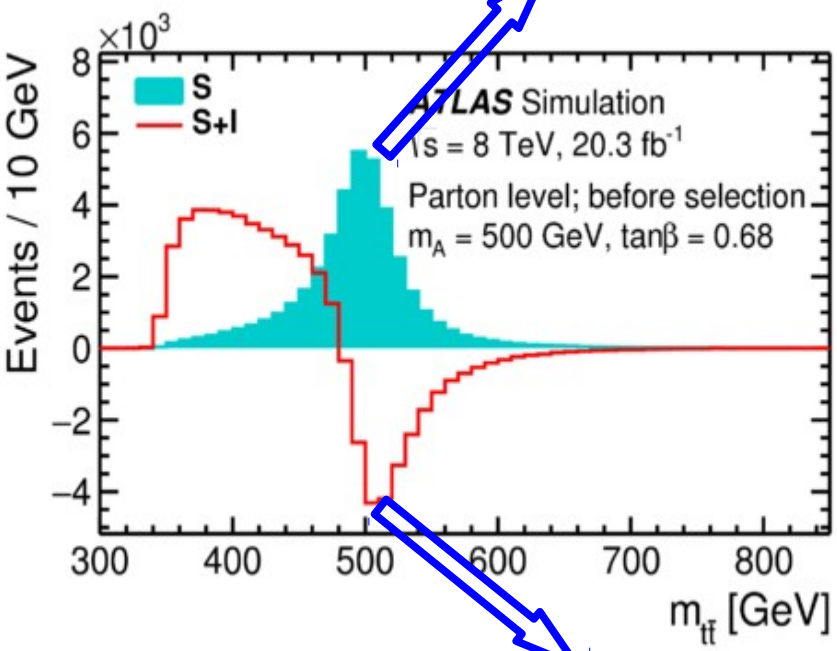


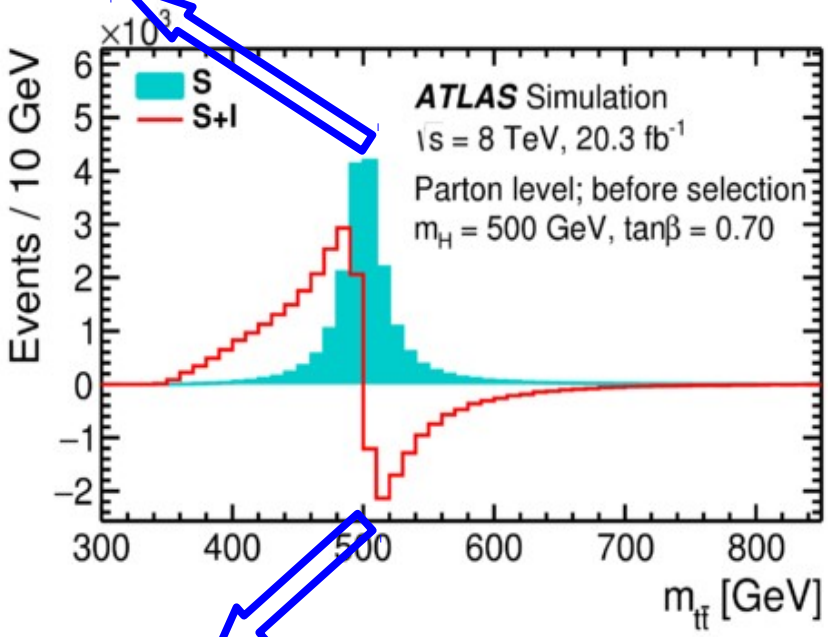
- This analysis is the first LHC search of $A/H \to t\bar{t}$ with the interference effects properly considered and treated

Interference

- The strong interference between the signal and the SM $t\bar{t}$, makes this search difficult
  - bump + dip shape (or even pure dip), quite different from normal searches

Normal searches (bump-hunter)





We are looking for

Signal Generation: S+I

- Not realistic to generate/produce <u>A/H → $t\bar{t}$ + SM $t\bar{t}$ + interference</u>

  - imagine we would need to generate (+ simulate and reconstruct) a whole SM $t\bar{t}$ dataset for each signal point

- The quick way: change the MG code to remove the SM ttbar matrix-element; then we just do <u>A/H + interference</u> [I] **(D. B. Franzosi)**

  - there is another reweighting way being studied for Run-II

- Another is subtraction between the follow two samples [II]:

  (a) SM $t\bar{t}$ + A/H + interference A/H-SM

  (b) SM $t\bar{t}$

- compare distributions between [I] and [II], as validation of approach-[I] (very nicely agree)

Reweighting

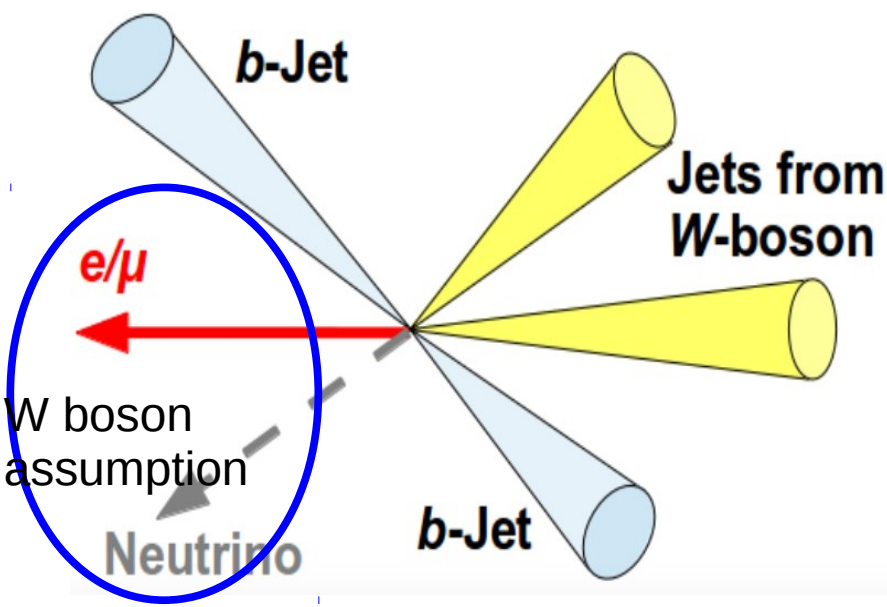
- Reweight: S ⇒S and S ⇒ S+I
  - e.g. [500GeV, tanb=0.5] S ⇒ [550GeV, tanb=0.5] S and [550GeV, tanb=0.5] S+I
- Event by event reweighting
  - based on the ratio of two matrix-elements

$$\mathcal{W}_{\mathrm{new}}(\{s_i\}_{\mathrm{new}}, \{p_j\}) = \frac{|\mathrm{ME}(\{s_i\}_{\mathrm{new}}, \{p_j\})|^2}{|\mathrm{ME}(\{s_i\}_{\mathrm{old}}, \{p_j\})|^2}$$

- Inputs: the four-vectors of the incoming gg and outgoing $t\bar{t}$ particles (built from their 6 decaying partons)
- Validation:
  - compare MC generated and reweighted
  - generally nice agreement; still conservatively Put on 10% reweighting uncertainty for S+I, 5% for S

Analysis

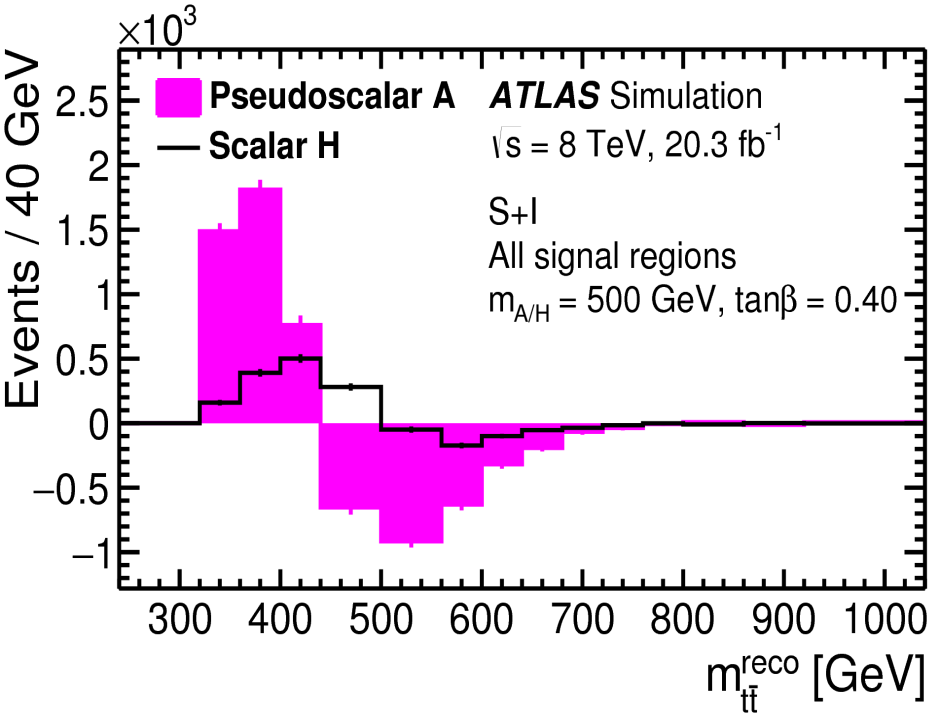
- $== 1$ electron or muon
  - $p_T > 25$ GeV, $|\eta| < 2.5$
  - tight, mini-isolated
- $E_T^{\text{miss}} > 20$ GeV
- $E_T^{\text{miss}} + m_T^W > 60$ GeV
- ≥ 4 anti-k_t $R = 0.4$ jets
  - $p_T > 25$ GeV, $|\eta| < 2.5$
- $\geq 1 b$-tagged jets
  - MV1 70% operating point

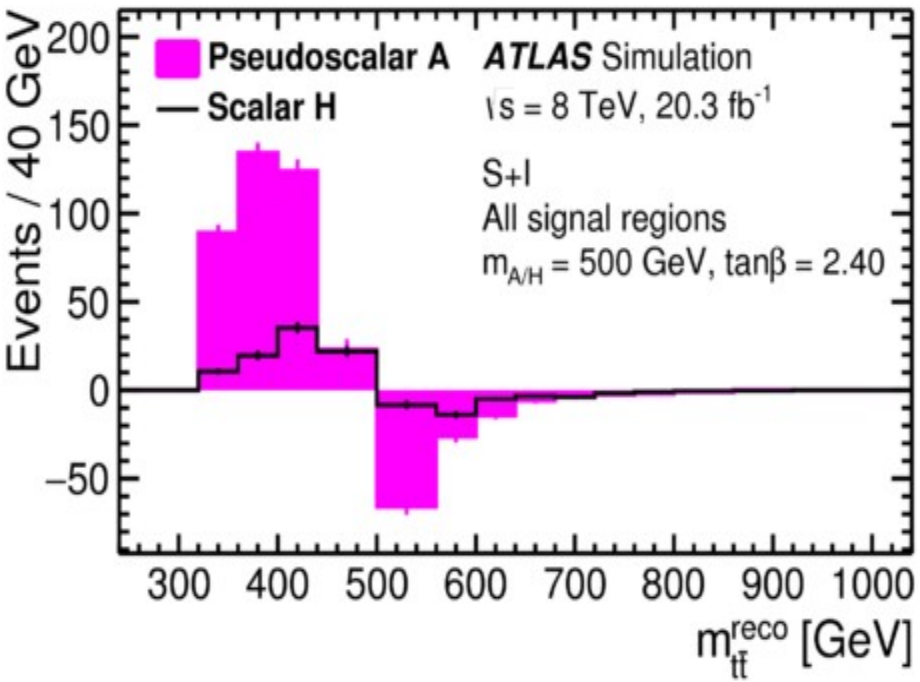


- Semi-leptonic, solved analysis
  - little sensitive when go to very high mass due to very huge interferences
- Event selections:
  - At least 4 small-R jets
  - at least one of these jets must be b-tagged
- Binning:
  - Optimization to avoid positive and negative contents cancel out
- Categorization:
  - Cat1: both leptonic and hadronic top have b-jets matched
  - Cat2: only hadronic top has b-jets matched
  - Cat3: only leptonic top has b-jets matched

Signal shape at Detector Level

- Event Reconstruction
  - Neutrino pZ is from W-boson mass constraint
  - Kinematic χ2 fit to reconstruct $t\bar{t}$ system





K-factor on interference

- Wouldn't have any calculation for interference at NLO from theory for this moment

- Geometry mean:
  - Discussed and agreed with several theorists including Zhen Liu, Marcela Carena, Peter Galler, etc.
  - Also explicitly used in several phenomenology papers

- k-factor for interference term:
  $k_I = \sqrt{k_B \cdot k_S}$
- $(S + I) = [(S + I) - S] \cdot k_I + S \cdot k_S$

Statistics

- Binned profile likelihood fit to data. S and B floated simultaneously. Uncertainties incorporated as NPs. Fitting model is shown as like: $\mu \cdot S + \sqrt{\mu} \cdot I + B$ ~ data
  - rewrite as: $k \cdot (S+I) + (k^2 - k) \cdot S + B, \quad where\, k = \sqrt{\mu}$
  - k is used as PoI, allowed to be negative in the fitting

- Comes from the amplitudes defined in our MadGraph model:

$$\frac{d\sigma}{ds} = \frac{\alpha_s^2 G_F^2 c_t^2 m_t^2 s^2}{1536\pi^3}\beta^3 \left| \frac{\sum_f c_f A_S(\tau_f)}{s - m_H^2 + i m_H \Gamma_H(s)} \right|^2 - \frac{\alpha_s^2 G_F c_t m^2}{66\pi\sqrt{2}}\beta^2 \log\left(\frac{1+\beta}{1-\beta}\right) \mathrm{Re}\left[\frac{\sum_f c_f A_S(\tau_f)}{s - m_H^2 + i m_H \Gamma_H(s)}\right] + \frac{d\sigma_{QCD}}{ds}$$

S I B

Scalar

Pseudo-scalar

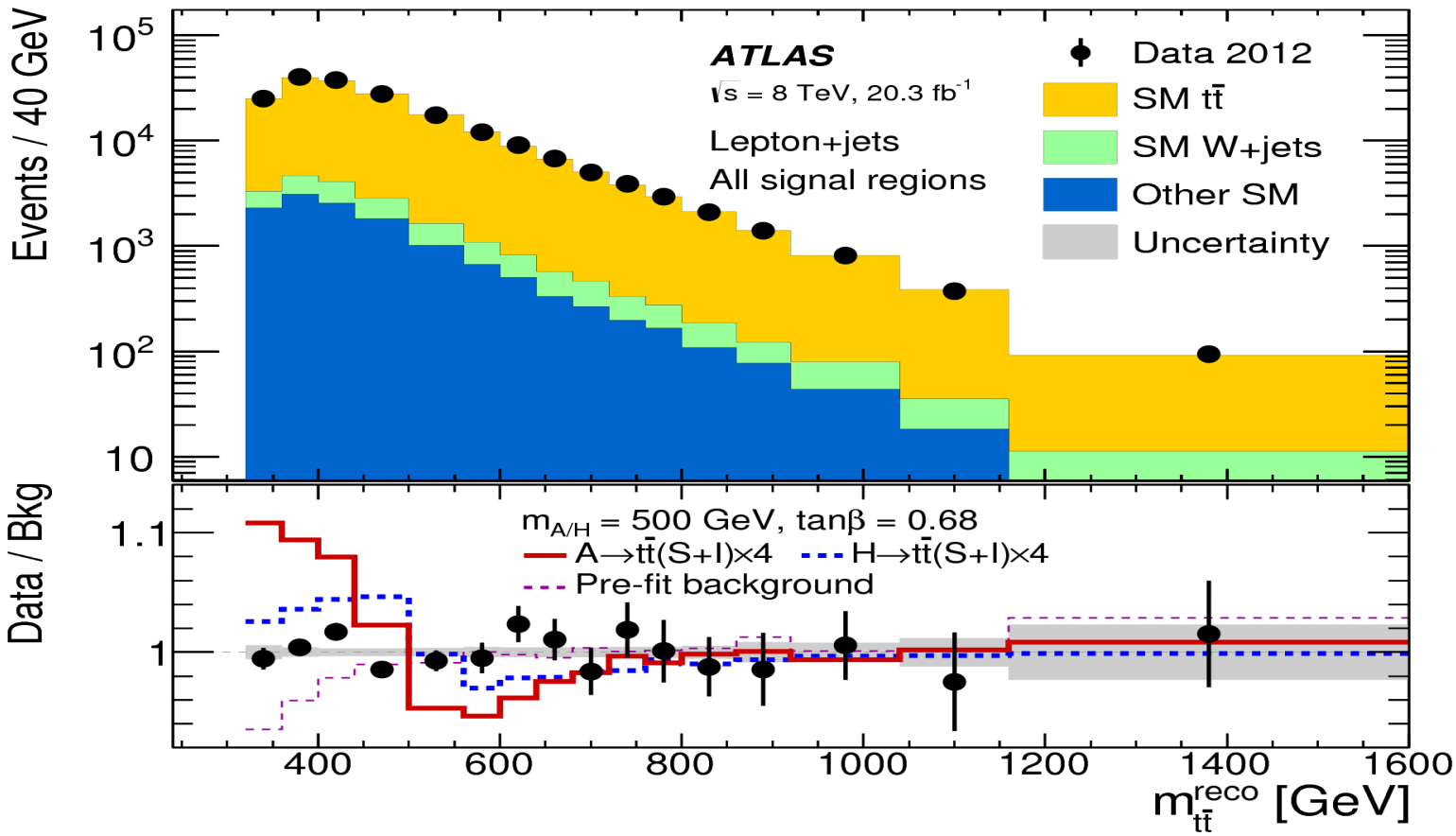
$$\frac{d\sigma}{ds} = \frac{3\alpha_s^2 c_t^2 G_F^2 m^2 s^2}{2048\pi^3}\beta \left| \frac{\sum_f c_f f(\tau_f)/\tau_f}{s - m_A^2 + i m_A \Gamma_A(s)} \right|^2 - \frac{\alpha_s^2 c_t G_F m^2}{64\pi\sqrt{2}} \log\left(\frac{1+\beta}{1-\beta}\right) \mathrm{Re}\left[\frac{\sum_f c_f f(\tau_f)/\tau_f}{s - m_A^2 + i m_A \Gamma_A(s)}\right] + \frac{d\sigma_{QCD}}{ds}.$$

S I B

- By default RooFit cut away all the negative bins when transform histogram to pdf; solved by changing the HistFactory code

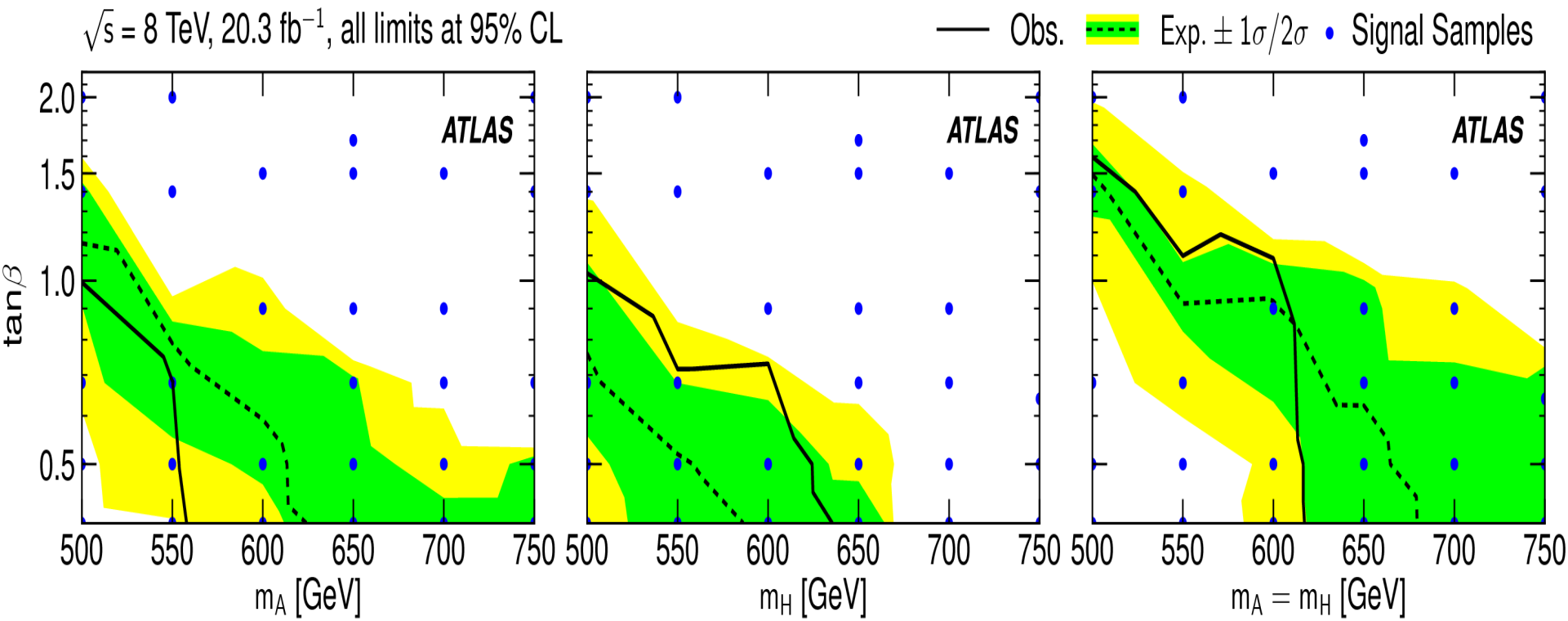
Final Mass Spectrum

All the categories merged



- Shows the individual S+I distributions for A/H relative to the total background

Interpretation Plots



- A ⇒ H ⇒ A/H degeneracy
- Upper limits at intermediate points are obtained from a linear triangular interpolation

Summary

- $A/H \rightarrow t\bar{t}$ search with proper interference treatment has been carried on for this first time
  - No clear BSM signal observed
  - Give out different interpretations (A,H, A+H) in the tanβ ~mass plane

- For interference technically we would need to be carefully about:
  - Signal generation
  - Statistical treatment
  - Binning optimization

Backup Slides

QCD Background Estimation

- Matrix method based on estimate of efficiencies (f) and fake rates (ε), which are derived from dedicated control-regions

- A looser lepton selection would contain prompt and fake leptons:

$$N_{\mathrm{L}} = N_{\mathrm{prompt}} + N_{\mathrm{multi\text{-}jet}}$$

- A tighter leptons can be estimated from those using the eefficiencies and fake rates:

$$N_{\mathrm{T}} = \epsilon \times N_{\mathrm{prompt}} + f \times N_{\mathrm{multi\text{-}jet}}.$$

- Solving these two equations for N_prompt and N_multi-jet, makes it possible to estimate the QCD contribution to the signal region

W+jets Normalization

- Using W charge asymmetric method

  - No b-tagging requirement; large-R jet fail the top-tagging. To increase statistics

$$N_{W^+} + N_{W^-} = \left(\frac{r_{\mathrm{MC}} + 1}{r_{\mathrm{MC}} - 1}\right)(D_{\mathrm{corr}+} - D_{\mathrm{corr}-}),$$

  - SF: calculated as the ratio of (N_{W+} + N_{W-}) evaluated from data to that predicted from MC simulation