UPPSALA UNIVERSITET

ATLAS EXPERIMENT

Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel with the ATLAS detector

Petar Bokan

supervisors:

Arnaud Ferrari, Pedro Sales de Bruin (Uppsala University)
Stan Lai (University of Goettingen)

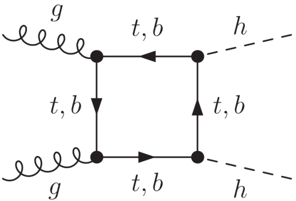
Partikeldagarna, Stockholm, 2017

Outline

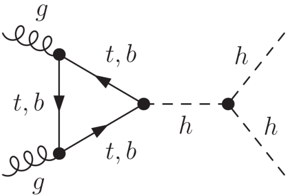
- Motivation for the di-Higgs searches and the $b\bar{b}\tau^+\tau^-$ final state.
- The Run-2 (resolved) analysis status and other efforts.
- The Run-2 analysis strategy
  - Trigger and event selection
  - Background estimation
  - Results
- High Luminosity Large Hadron Collider (HL-LHC) $\rightarrow b\bar{b}\tau^+\tau^-$ prospects.
  - Limit extrapolation: Run-2 analysis $\rightarrow$ ($\sqrt{14}$ TeV, 3000 fb^{-1})
- Summary

Motivation for the di-Higgs searches

- Standard Model (SM) Higgs boson pair production:



Higgs-fermion Yukawa coupling

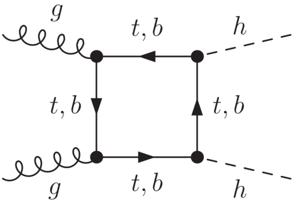


Higgs boson self-coupling

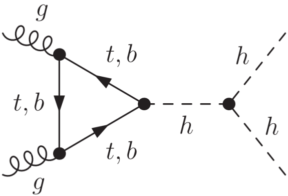
SM Higgs boson pair production is very small due to destructive interference between these two diagrams (33.41 fb @ $\sqrt{s} = 13$ TeV)

Motivation for the di-Higgs searches

- Standard Model (SM) Higgs boson pair production:



Higgs-fermion Yukawa coupling

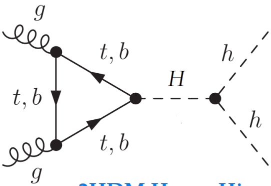


Higgs boson self-coupling

- Looking for beyond the SM resonances decaying to two Higgs bosons.

Tested BSM hypotheses:

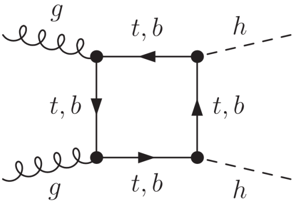
- Randall-Sundrum graviton → hh.
- Heavy Higgs H → hh.
  (Two Higgs Doublet Model - 2HDM)



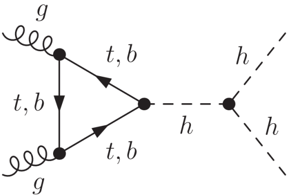
2HDM Heavy Higgs

Motivation for the di-Higgs searches

- Standard Model (SM) Higgs boson pair production:



Higgs-fermion Yukawa coupling

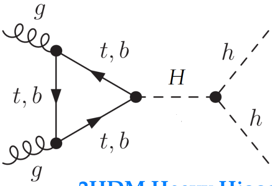


Higgs boson self-coupling

- Looking for beyond the SM resonances decaying to two Higgs bosons.

Tested BSM hypotheses:

- Randall-Sundrum graviton → hh.
- Heavy Higgs H → hh.
  (Two Higgs Doublet Model - 2HDM)



2HDM Heavy Higgs

- Non-resonant BSM enhancements.
  (due to new couplings, or modified Yukawa/self-coupling)

hh→ $b\bar{b}\tau^+\tau^-$ final state

- Decay modes of the di-Higgs system and their relative branching fractions:

	bb	WW	ττ	ZZ	γγ
bb	33%				
WW	25%	4.6%			
ττ	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
γγ	0.26%	0.10%	0.029%	0.013%	0.0053%

$b\bar{b}\tau^+\tau^-$ is a relatively clean final state and it still benefits from a relatively large branching fraction

- Sub-channels considered:
  - $\tau_{\text{lep}}\tau_{\text{had}}$ (branching ratio = 45.8%)
  - $\tau_{\text{had}}\tau_{\text{had}}$ (branching ratio = 41.9%)

hh$\rightarrow$ $b\bar{b}\tau^+\tau^-$ analysis

- Analysing data recorded during the 2015 and 2016 data taking periods. Corresponding to $\int$ Ldt $= 36.1$ fb^{-1}.
- Publication is expected before Christmas.
- Signals hypotheses tested:
  - Non-resonant production: Standard Model di-Higgs signal.
  - Resonant production: RS Kaluza-Klein graviton and 2HDM Heavy Higgs: $m = 260, 275, 300, 325, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000$ GeV.
  - Variations of the trilinear Higgs self-coupling (λ_{hhh}) strength will be tested as well.

hh$\rightarrow$ $b\bar{b}\tau^+\tau^-$ analysis

- Analysing data recorded during the 2015 and 2016 data taking periods. Corresponding to $\int$ Ldt $= 36.1$ fb^{-1}.
- Publication is expected before Christmas.
- Signals hypotheses tested:
  - Non-resonant production: Standard Model di-Higgs signal.
  - Resonant production: RS Kaluza-Klein graviton and 2HDM Heavy Higgs: $m = 260, 275, 300, 325, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000$ GeV.
  - Variations of the trilinear Higgs self-coupling (λ_{hhh}) strength will be tested as well.

Other efforts and plans:

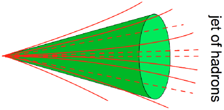
- 3rd generation of leptoquarks signal re-interpretation of the analysis.
- High Luminosity LHC Standard Model hh$\rightarrow$ $b\bar{b}\tau^+\tau^-$ prospects.
- Boosted regime analysis.
- Include new 2017 data.
- Combination result with the other di-Higgs channels.

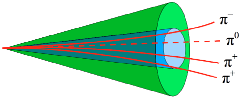
Event selection

$\tau_{\text{lep}}\tau_{\text{had}}$		$\tau_{\text{had}}\tau_{\text{had}}$	
Single lepton trigger SLT	Lepton tau trigger LTT	Single tau trigger STT	Di-tau trigger DTT
1 e/μ and 1 medium τ of oppostite sign		2 medium τs of opposite sign	
$p_T^{e/\mu} > 25, 27$ GeV (for 24, 26 GeV triggers) $p_T^{\tau} > 20$ GeV	18 GeV $< p_T^{e} <$ SLT threshold 15 GeV $< p_T^{\mu} <$ SLT threshold $p_T^{\tau} > 30$ GeV	$p_T^{\text{lead}\,\tau} > 100, 140, 180$ GeV (for 80, 125, 160 GeV triggers) $p_T^{\text{subl}\,\tau} > 20$ GeV	$p_T^{\text{lead}\,\tau} > 40$ GeV $p_T^{\text{subl}\,\tau} > 30$ GeV
≥ 2 central jets		≥ 2 central jets	
$p_T > 45, 20$ GeV	$p_T > 80, 20$ GeV	$p_T > 45, 20$ GeV	$p_T > 80, 20$ GeV $p_T > 45, 20$ GeV (2015 data)
$m_{\tau\tau}^{\text{MMC}} > 60$ GeV		$m_{\tau\tau}^{\text{MMC}} > 60$ GeV	

- **Signal region (SR): events with exactly 2 b-tagged jets.**
- A **Boosted Decision Tree** classification is applied to the SR.
- Events with 0, 1 b-jets, high transverse mass and "same-sign" events are used to define the control/validation regions.

Background estimation

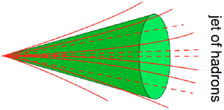


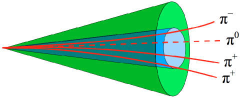


- Hadronically decaying τs are seen as narrow jets with small track multiplicity (1 or 3 tracks).
- Other jets can be misidentified as τs.
- Data-driven methods used to estimate jet $\rightarrow$ fake τ background for different processes.

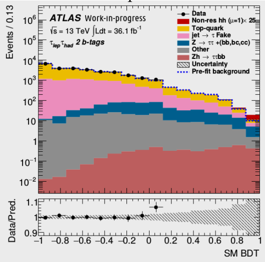
Background estimation

- Hadronically decaying τs are seen as narrow jets with small track multiplicity (1 or 3 tracks).
- Other jets can be misidentified as τs.
- Data-driven methods used to estimate jet $\rightarrow$ fake τ background for different processes.

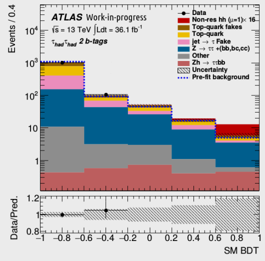




$\tau_{\text{lep}}\tau_{\text{had}}$



$\tau_{\text{had}}\tau_{\text{had}}$



$\tau_{\text{lep}}\tau_{\text{had}}$

- true ttbar (Monte Carlo)
- Jet $\rightarrow$ fake τ_{had}
  mostly from ttbar, but also from multi-jet and W+jets processes (Combined Fake Factor Method)

$\tau_{\text{had}}\tau_{\text{had}}$

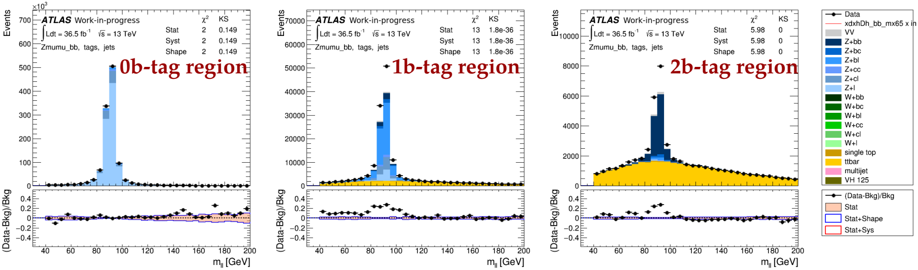
- true ttbar (Monte Carlo)
- Jet $\rightarrow$ fake τ_{had}
  from multi-jet processes (ABCD Fake Factor method)
- Jet $\rightarrow$ fake τ_{had}
  from ttbar processes (Fake Rate method)
- $Z \rightarrow \tau\tau$ + bb/bc/cc (MC)

- ttbar normalization freely floated in the fit. Constrained by the low BDT region.
- $Z \rightarrow \mu\mu$ + bb/bc/cc **control region used to correct the** $Z \rightarrow \tau\tau$ + bb/bc/cc **normalization.**

Z + Heavy Flavour normalization

- The $Z + HF$ cross-section is not well described by the Monte Carlo (Sherpa).
- A $Z \rightarrow \mu\mu$ + bb control region is defined to correct the Z + bb/bc/cc normalization in the bb$\tau\tau$ signal region.
- The event selection is very similar to the one used for the signal region.

  In addition: $81 < m_{\mu\mu} < 101$ GeV and ($m_{bb} < 80$ GeV or $m_{bb} > 140$ GeV), one-bin region.



- For the $\tau_{\rm lep}\tau_{\rm had}$ Standard Model signal fit (background only hypothesis):

$$SF(Z + HF) = 1.44 \pm 0.17.$$

- Similar results for other signals.

Results, non-resonant signal

- A BDT score in the 2b-tag signal region is used as the final discriminant for the fitting and limit setting.
- Work-in-progress results:

channel	-2σ	-1σ	expected $\sigma/\sigma_{\mathrm{SM}}$ 95% C.L. limit	$+1\sigma$	$+2\sigma$
$\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ SLT	14.02	18.82	26.12	36.35	48.73
$\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ LTT	38.16	51.28	71.10	98.95	132.65
$\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	13.67	18.35	25.47	35.44	47.51
$\tau_{\mathrm{had}}\tau_{\mathrm{had}}$	8.40	11.28	15.66	21.79	29.21
combined	7.01	9.14	13.06	18.18	24.37

- Observed (expected) 95% C.L. limit on σ/σ_{SM}:

channel	bbbb	bbWW	bb$\tau\tau$	bb$\gamma\gamma$	WW$\gamma\gamma$
ATLAS	29 (38)	-	-	117 (161)	747 (386)
CMS	342 (308)	410 (227)	28 (25)	91 (90)	-

2.3-3.2 fb^{-1} 13.3 fb^{-1} 35.9 fb^{-1}

HL-LHC SM hh$\rightarrow b\bar{b}\tau^+\tau^-$ prospects

Run-2 analysis:

- Large fraction of the background for both sub-channels is estimated in a data-driven way.
- Huge improvement after switching from a cut-based to the BDT based analysis.

- For those reasons, extrapolation of the Run-2 results is a prefered way of studying the SM di-Higgs HL-LHC prospects in the $b\bar{b}\tau^+\tau^-$ channel.
- Similar strategy used for the SM hh $\rightarrow$ 4b prospects study.

Extrapolation steps:

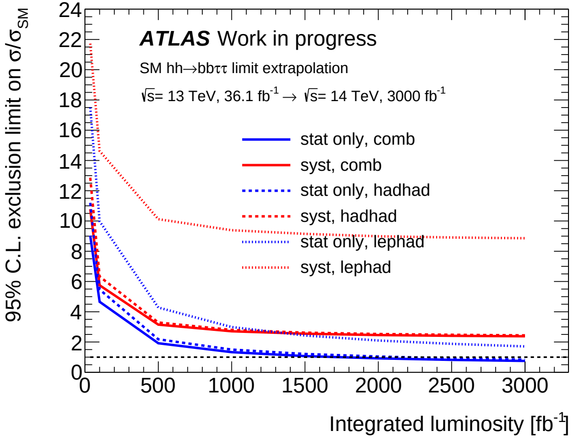
- All the distributions (including the statistical uncertainties) are scaled by

$$f = (\int Ldt)_{\text{target}} / (\int Ldt)_{2015+2016}$$

- All the distributions are scaled by 1.18 to account for an increase in cross-sections that comes from a change in gluon luminosity.
- ttbar and $Z + HF$ normalizations fixed to the best Run-2 analysis fit results.

HL-LHC SM hh$\rightarrow b\bar{b}\tau^+\tau^-$ prospects

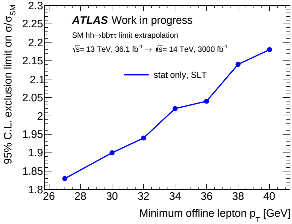
- Assuming the same triggers and no improvements for any of the systematic uncertainties.



Extrapolation results:

channel	stat (syst)
4b	1.50 (5.20)
$\tau_{lep}\tau_{had}$	1.71 (8.86)
$\tau_{had}\tau_{had}$	0.84 (2.44)
$b\bar{b}\tau^+\tau^-$	0.75 (2.38)

Single lepton trigger threshold study (stat only limit)



Summary

- Many improvements since the last iteration of the analysis.
  - Lepton + τ trigger category included.
  - Jet $\rightarrow$ fake τ background modeling improvements.
  - $Z \rightarrow \tau\tau$ + bb/bc/cc modeling improved.
  - Moved from cut-based to the BDT based analysis.
- For the non-resonant search, $b\bar{b}\tau^+\tau^-$ is currently the most sensitive channel.
- Measuring the λ_{hhh} is one of the primary physics goals for HL-LHC and beyond. Promising HL-LHC prospects of the $b\bar{b}\tau^+\tau^-$ channel.

backup

Boosted Decision Tree training

Variable	$\tau_{lep}\tau_{had}$ SLT-resonant	$\tau_{lep}\tau_{had}$ non-resonant LTT-resonant	$\tau_{had}\tau_{had}$
m_{hh}	■	■	■
$m_{\tau\tau}^{MMC}$	■	■	■
m_{bb}	■	■	■
$\Delta R(\tau,\tau)$	■	■	■
$\Delta R(b,b)$	■	■	■
E_T^{miss}	■		■
$E_T^{miss}\ \phi$ centrality	■		■
$m_T^{l\nu}$	■	■	
$\Delta\phi(h,h)$	■		
$\Delta p_T(l,\tau)$	■		
Sub-lead b-jet	■		

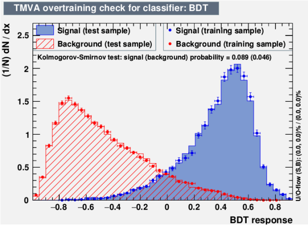
■ used for the BDT

BDT trainings done on 2b-tag signal region events.

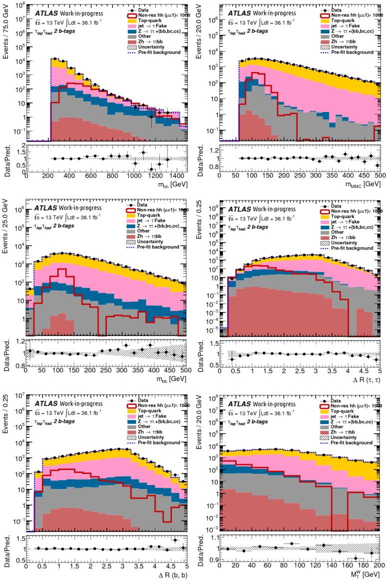
- BDT clasifier trained against all dominant backgrounds.
- One training for each signal type (and mass) hypothesis.

An example of BDT response

($\tau_{lep}\tau_{had}$ SLT, Heavy Higgs 2HDM $m_H = 300$ GeV, signal normalized to the background)



Boosted Decision Tree output

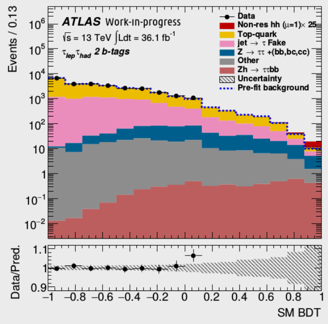


$\tau_{had}\tau_{had}$ SLT

BDT input variables used for the training on the SM signal.

- 2b-tag signal region
- Signal yield multiplied by 1000.

BDT output distribution:



- Signal yield multiplied by 25.
- Region that contains 90% of the signal is blinded.

Combined Fake Factor Method

- **Jet → τ fake background** contribution from all processes with jets (QCD, W+jets, ttbar). Estimated using a combined **Fake Factor (FF) method .**
- $\boldsymbol{FF = \frac{N(\tau)}{N(anti-\tau)}}$ as function of τ p_T and prongness, separately for SLT and LTT category, for each process in **separate CRs**:

  (**anti-τ:** τ-ID requirement inverted: **!BDTMedium +** τ-ID BDT > 0.35)

QCD CR	inverted lepton isolation (e/μ fail loose isolation working point)
W+jets CR	$m_T(l, MET) > 40$ GeV, 0b-tag region
ttbar CR	$m_T(l, MET) > 40$ GeV, 2b-tag region

- These are used to calculate:

$$FF_{comb} = FF_{QCD} \times r_{QCD} + FF_{ttbar/W+jets} \times \left(1 - r_{QCD}\right)$$

where r_{QCD} is a fraction of multi-jet events in the anti-τ signal region.

$$r_{QCD} = \mathrm{N}_{\mathrm{multijets}}^{\mathrm{data-driven}} / (\mathrm{data} - \mathrm{MC}_{\mathrm{true}})$$

ABCD Matrix Fake Factor Method

- A data-driven **"matrix" method** is used to estimate the multi-jets background!

A OS τ	B SS τ
C OS anti-τ	D SS anti-τ

$$\frac{\mathrm{N(OS,ID)}}{\mathrm{N(OS,anti-ID)}} = \frac{\mathrm{N(SS,ID)}}{\mathrm{N(SS,anti-ID)}}$$

$$\mathrm{FF} = \frac{\mathrm{N(SS,ID)}}{\mathrm{N(SS,anti-ID)}}$$

2D fake factors, as a function of leading and sub-leading τ p_T

(**anti-τ:** τ-ID requirement inverted: **!BDTMedium +** τ-ID BDT > 0.35)

- Separate fake-factors are calculated for STT and DTT events, and for 1-prong and 3-prong hadronic taus.
- 1b-tag FFs used for 2b-tag region due to the lack of statistics. (transfer factor applied)

$$(\text{multi-jet})|_{\mathrm{OS,ID}} = \mathrm{FF} * (\text{data} - \mathrm{MC})|_{\mathrm{OS,anti-ID}}$$

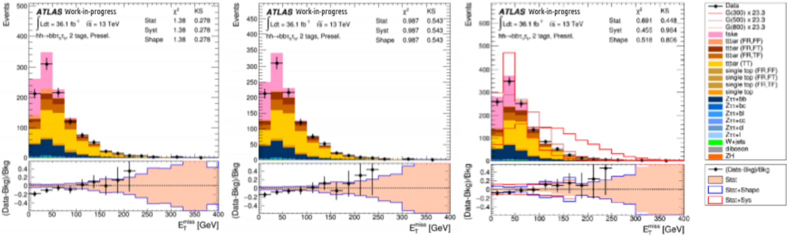
Fake Rate Method

- Use $\tau_{lep}\tau_{had}$ ttbar CR ($m_T^W > 80$ GeV) to derive fake rates (FR)

$$\mathrm{FR} = \left.\frac{\mathrm{N}^{\mathrm{pass}\ \tau-\mathrm{ID}}}{\mathrm{N}^{\mathrm{total}}}\right|_{\mathrm{fake}\ \tau s}$$

Separately for 1-prong and 3-prong as a function of τ p_T eta

- Applied to all non-truth-matched ttbar MC events according to which tau is fake.



ttbar fakes from MC | fake rate method (stat only) | fake rate method (syst)