Search for 2HDM neutral Higgs bosons through the H->ZA->llbb process with the full run2 data

Khawla Jaffel
Supervisor : Christophe Delaere

EOS PHD DAYS

27-Nov-2020

2HDM

“Whether you can observe a thing or not depends on the theory which you use. It is the theory which decides what can be observed.”
Albert Einstein (1879 - 1955)

Model parameters:

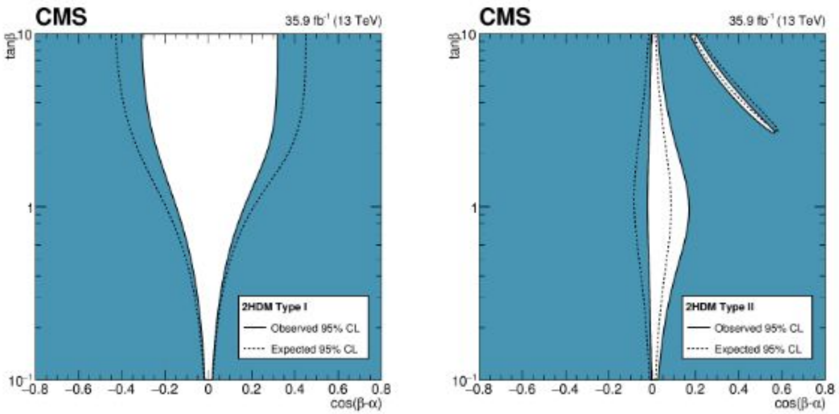
- tanβ, cos(β-α), mA, mH, mH± (when imposing symmetries to forbid tree-level FCNC and CP violation)

Four different types (I, II, flipped, lepton-specific) determined by Yukawa couplings to up-/down-type quarks and leptons

New states: A, H, H±

Signatures of CP-conserving and FCNC

- Modified h125 couplings;
- H decays: as in singlet + to fermions, ZA
- A decays: to fermions, Zh125, ZH
- H± decays: tb, cs, τν; Wh125, WH, WA



cos(β-α) =0: alignment
Decays A→Zh125, H→h125 h125 vanish, but not A→ZH/H→ZA etc

2HDM searches

Conventional channels: Most of the existing direct searches for BSM Higgs bosons focus on their conventional decays into a pair of quarks, leptons or gauge bosons.
(A/H → μμ, bb , ττ, γγ, tt) and **(H→ ZZ, WW)**

- **Exotic decays into the SM Higgs:** **A→Z (h125 → bb,ττ) , H→ hh** (away from the alignement)
- **Exotic decays of the SM Higgs:** If the BSM scalars are sufficiently light , exotic decays of the SM-like Higgs h→ AA/HH opens up.
  h→ AA (bbbb, bbττ, bbμμ, ττττ, ττμμ, μμμμ) searches focusing on masses of mA > 4GeV
- **Exotic decays in BSM sector:** For |mA- mH| > mZ, the exotic decay channel **A/H-> ZH/ZA → (llbb, llττ)** when allowed, typically dominant decay and hence strong exclusion.
- **Searches for charged Higgs : H± → cs, τν and tb.** At large masses of **mH± > mt**, constraints on these charged Higgs searches are typically weaker than their neutral counterparts, due to the suppressed **tbH±** associated production cross section as well as large backgrounds.

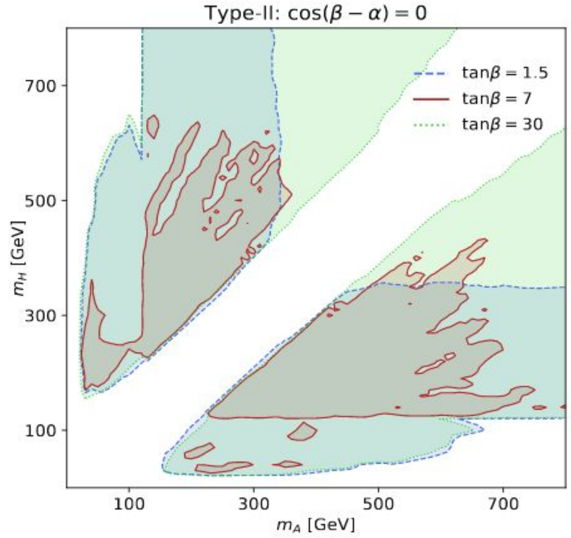
mA vs.mH plane of the 2HDM parameter space

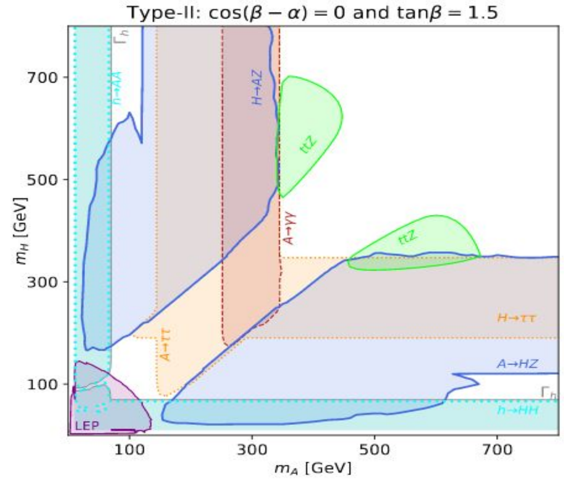
Parameter space excluded at 95% C.L. by the H/A-->ZA/ZH search in the alignment limit:

- ❑ Away from the alignment limit, these constraints are weakened given the suppressed coupling H→ ZA ∝ sin(β-**α**) (BACKUP)
- ❑ In the gap region along mA~mH, the exotic decay modes are kinematically inaccessible.

constraints at 95%C.L. for cos(β-α)= 0 and tanβ= 1.5:

- ❑ THe combination of all channels cover the majority of the region in which one of the Higgs masses is below the di-top threshold mA, mH < 2mt
- ❑ In the gap region: **A/H→ ττ , A/H→ γγ** cover nevertheless these channels become inefficient for Higgs mass above 2mt, where A/H→ tt opens up. This not only decreases the branching fraction into the clean signatures but it can also increase the heavy Higgs widths significantly which imposes a general problem for resonant searches.(BACKUP)





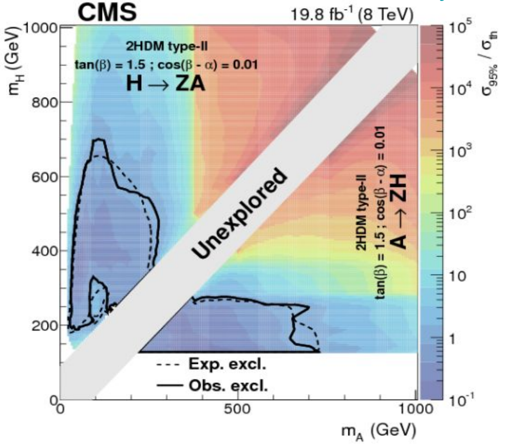
What was the landscape at the start of Run2 ?

- Before exploring Full run2 data, a small stop to explore the view from previous searches:
    - Many extra scalar searches performed
    - No hints of new physics
    - But there is still a lot of uncovered phase--space -- specially at high mass!

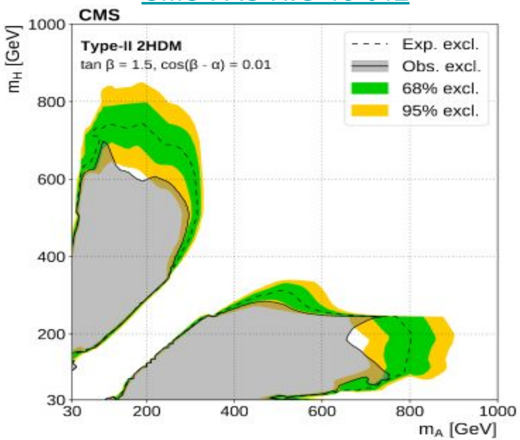
<u>The questions we ask is:</u> **Which parts of the 2HDM parameter space are favoured after imposing the latest experimental data from the LHC?** Full run 2 CMS update should address weak spots (b-associated production) and retain its strengths (H→ZA, lower + higher masses)

2.6 sigma excess was observed at(mH , mA)= (285, 95) GeV → **statistical fluctuation**

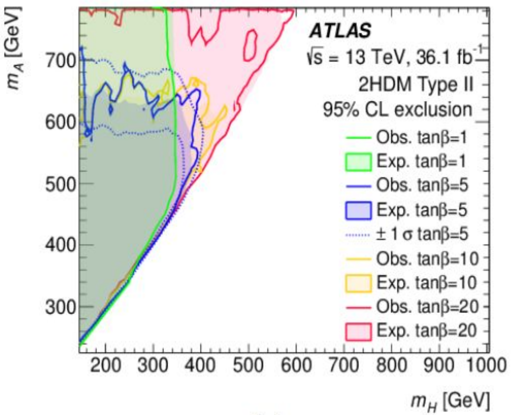
CERN-THESIS-2017-474 8TeV Analysis



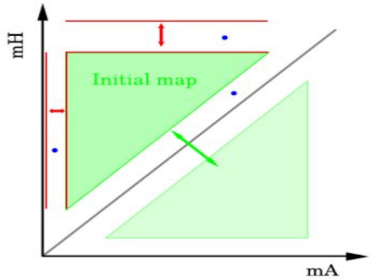
CMS-PAS-HIG-18-012



ATLAS PAPER - HDBS-2018-13/



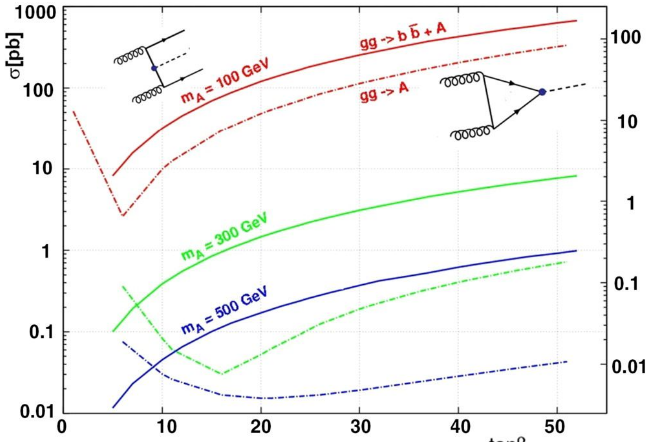
Direct searche: Analysis strategy

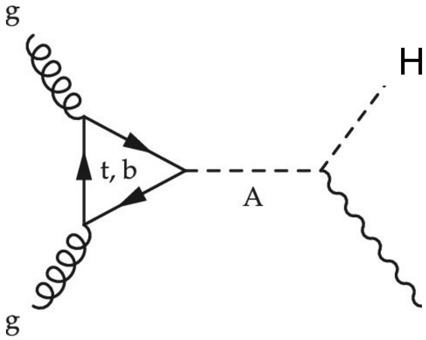


Signal :

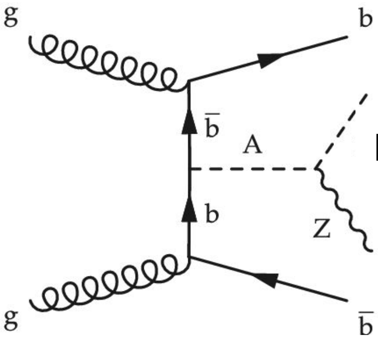
H->ZA->OS Leptons(ee /μμ) + 2 b-tagged jets(2 AK4 resolved **OR 1 AK8 boosted**)

- In the alignment limit the coupling to vector bosons is negligible, only two production modes will be studied: the gluon-gluon-fusion with a top loop and the **bb associated mechanism.**
- The cross-sections will depend on the different parameters of the model.
- Conduct the search to high tanβ values.





Production pseudoscalar A boson via gluon–gluon fusion

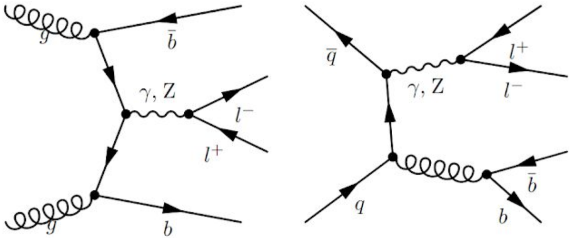


Production pseudoscalar A boson in association with b quarks*

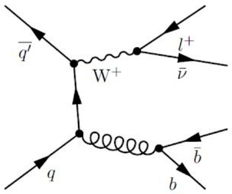
Backgrounds:

- **MC samples used are NanoAODv7 . [(Full list in backup)]**

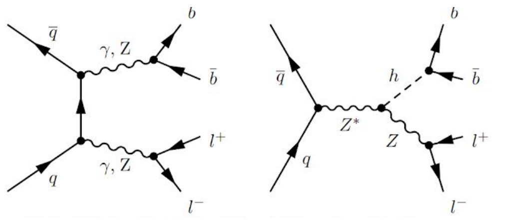
main background



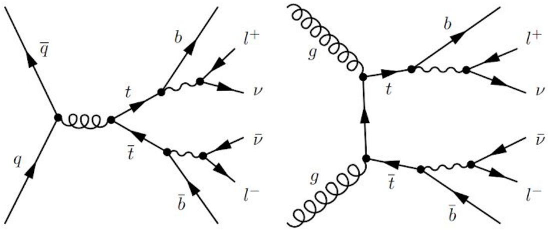
DY+ Jets : gg→Z(l+l−)jj has the biggest cross-section thanks to the high gluon pdf, while the q q̄→Z(l+l−)jj is reduced by the parton density function of the antiquark.



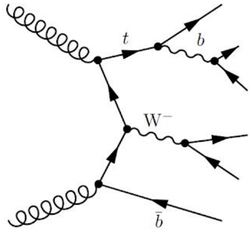
W + heavy flavour jets



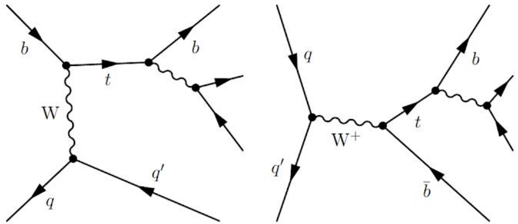
Leading order **ZZ** and **Zh** process diagram



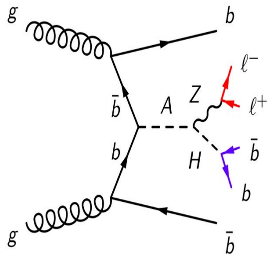
ttbar process, leading diagrams



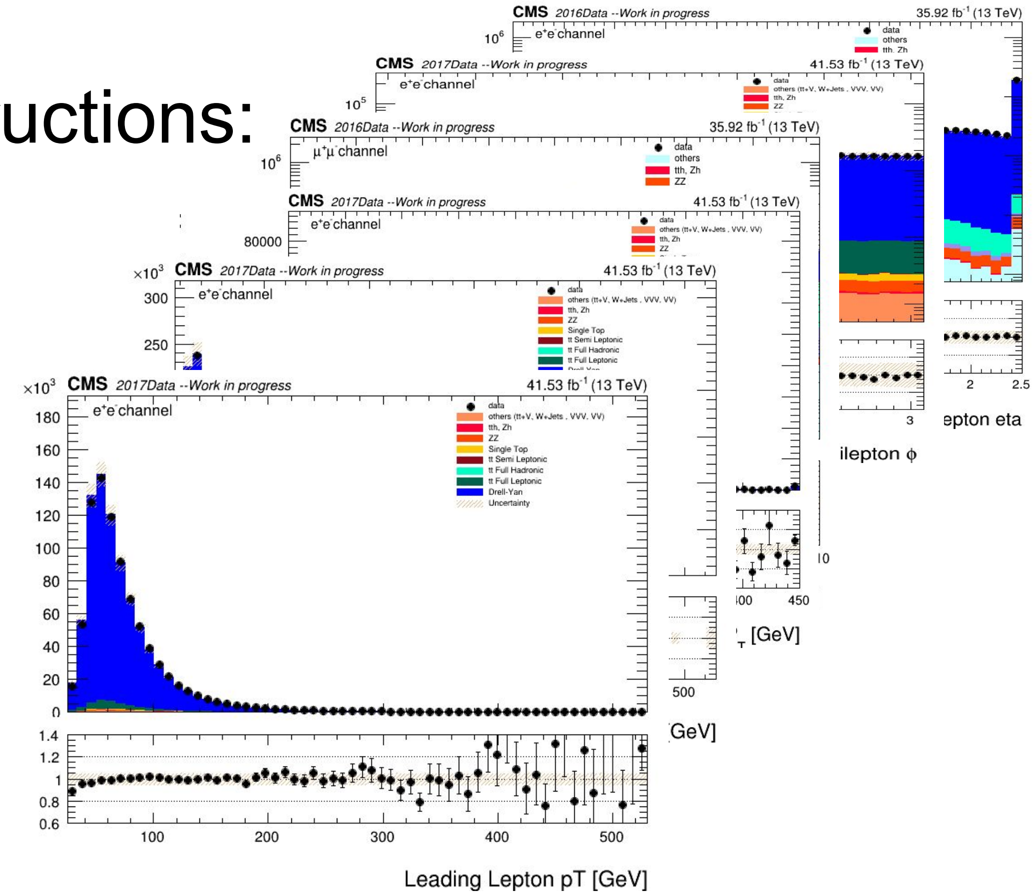
Single top production in addition to a W boson.



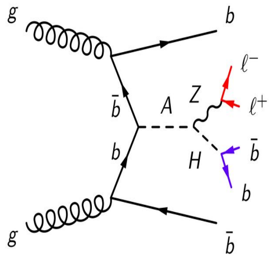
Single top with a W boson in the **t-channel** on the left and **s-channel** on the right.



Object reconstructions:



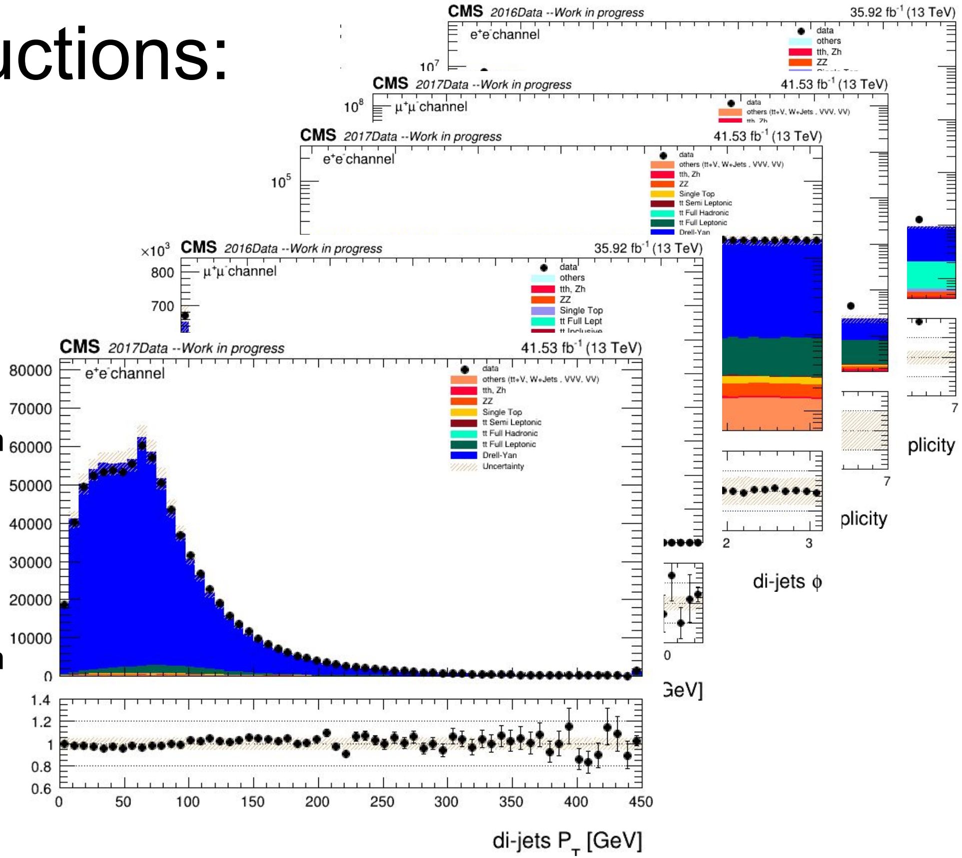
- **Muons:**
  - medium ID cut
  - PF ISO (Δß-corr) < 0.15
  - Leading (sub-leading) pT>20(10) GeV
  - |eta| < 2.4
- **Electrons:**
  - medium ID cut
  - **Barrel:** dz < 1mm , dxy < 0.5 mm
  - **Endcap:** dz < 2mm, dxy < 1mm
  - Leading (sub-leading) pT > 20(15) GeV
  - |eta| < 2.5
- 70 < mll < 120 GeV to enhance the presence



Object reconstructions:

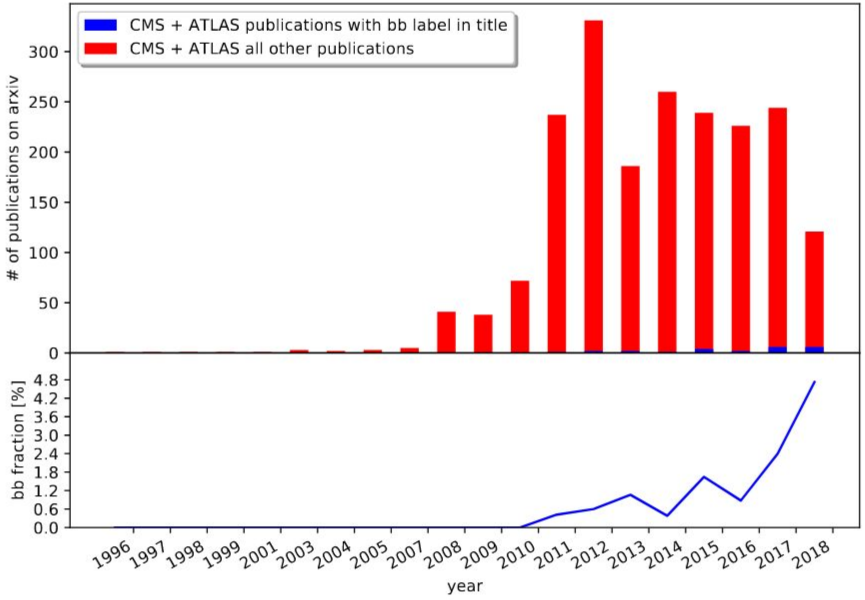
- **Jets:**
  Resolved:
  - |eta| < 2.4
  - AK4 CHS PF Jet pT > 20 GeV (2016) ;
    AK4 CHS PF Jet pT > 30 GeV (2017/2018)
  - 3D impact parameter w.r.t first PV |sip3d| < 4cm

  Boosted :
  - N-subjettiness parameters tau2/tau1 < 0.75
  - |eta| < 2.4
  - 3D impact parameter w.r.t first PV |sip3d| < 4cm
  - AK8 CHS PF Jet pT >200 GeV



Publications on Arxiv with a b-quark pair in the finalstate

- With increasing luminosity analyses with a b-quark pair in the finalstate become increasingly popular.
- mbb resolution is key for the sensitivity of most analyses.
- Identification of jets originating from the hadronization of b-quarks is a powerful method to reduce backgrounds in analysis involving Higgs bosons or top quarks since their decays involve b in almost 60% and 100% of cases respectively.



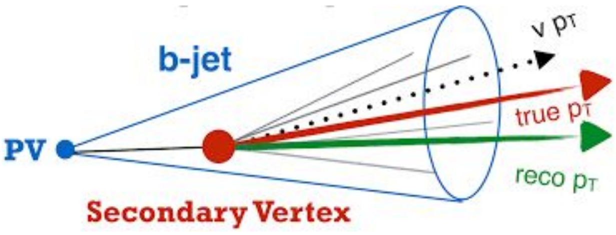
b-jet Energy Regression

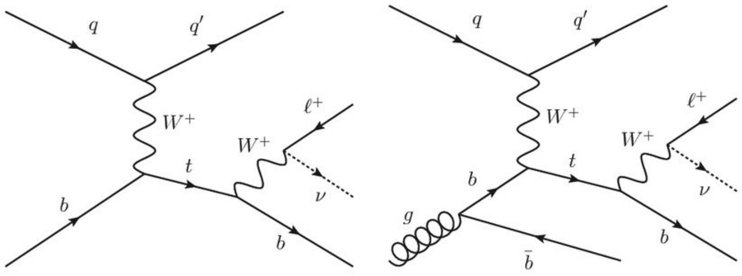
https://twiki.cern.ch/twiki/bin/view/CMS/HiggsWG/BJetRegression

- **Semileptonic decays lead to mismeasured energy due to escaping neutrino worsening the b-jets resolution.**
- The goal of b-jet energy regression is to correct b-jet energy scale using a multidimensional correction and information at generator level.
- This correction improves detector response and corrects for semi leptonic b-decays.

<u>What's New?</u>

- Switch from BDT to Deep Neural Net ; additional inputs added to the training
- Training is done on 100 M b-jets from TTbar
- Train on b-jets from jet pT > 15 GeV and |η|<2.5
- Target : ratio of gen jet pT with included neutrinos over reco pT

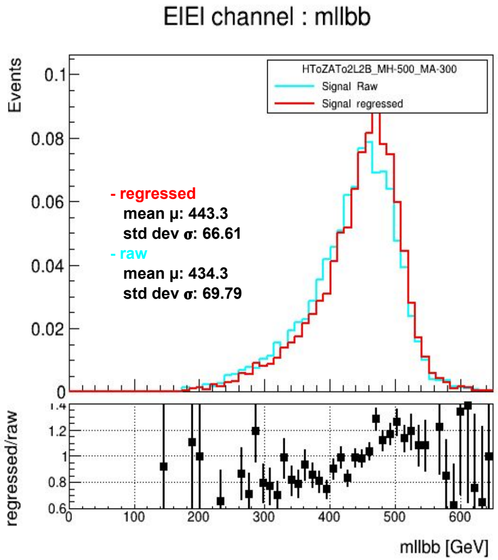


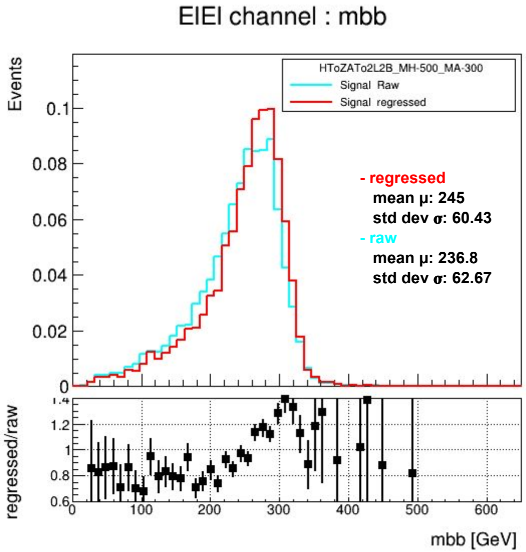


t-channel single top-quark production and decay.

Effect on H->ZA signal samples

2Lep, 2b-tagged jets (DeepCSV M WP) selection, year 2016





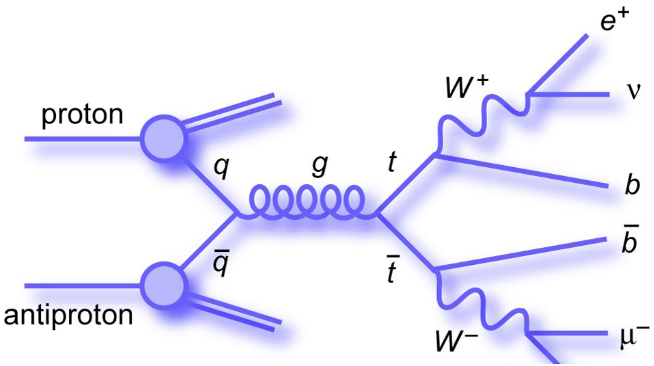
16.38 % improvement in FWHM/peak of regressed M(bb) wrt to JEC corrections

Application to the ZH Associated Production Analysis:
https://arxiv.org/abs/1709.07497

EPR:

B-tagging Efficiencies Measurements in ttbar dilepton events at 13 TeV

- Identifying jets originating from the hadronisation of b-quarks is a powerful method to reduce background in analysis involving Higgs bosons or top quarks .
- 2TagCount measures the b jet identification efficiency by counting the fraction of events with 2 b-tagged jets within a sample of events requiring exactly two selected jets and exactly one isolated electron and one isolated muon.
- The b tagging efficiency can be obtained from :



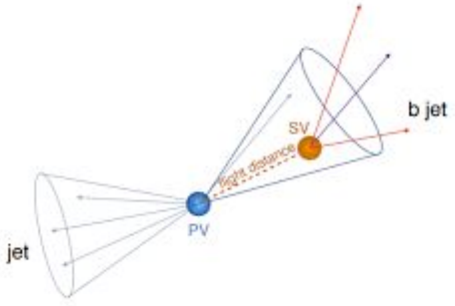
$$N_{2btag} - N_{2btag}^{non-b-jet} = \varepsilon_b^2 n_{2bjet},$$

$$\varepsilon_b = \sqrt{\frac{F_{2tag} - F_{2btag}^{non-b-jet}}{f_{2b}}} \quad *$$

number of events with 2 btagged jets (taken from data)

the number of events with 2 b-tagged jets with at least one of the selected jet being a light or charm jet (taken from MC)

the number of events in which both jets are b-jets (taken from MC).



Event reweighting using scale factors and MC b-tagging efficiencies:

- The goal of this method is to predict the correct event yield in data by only changing the weight of the selected MC events [1]
- **Plots shown:** pt vs eta efficiencies map for 2017 data, resolved region, e+e- channel.

$$\omega = \frac{P(Data)}{P(MC)}$$

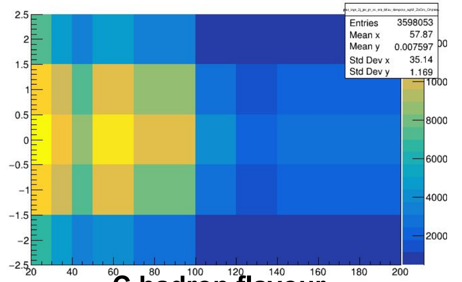
where:

$$P(MC) = \prod_{\text{tagged}} \epsilon_i \times \prod_{\text{not tagged}} (1 - \epsilon_j)$$

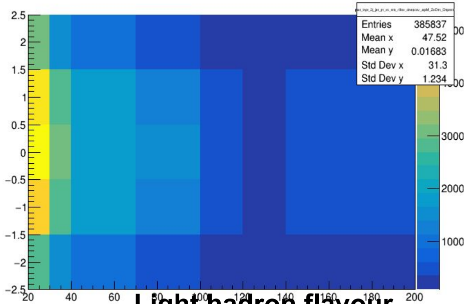
$$P(Data) = \prod_{\text{tagged}} \mathrm{SF}_i \epsilon_i \times \prod_{\text{not tagged}} (1 - \mathrm{SF}_j \epsilon_j)$$

DeepFlavourM :	Tot.MC	Data	Data/MC	
No bTagEventWeight	31779.9 ± 1726.4	29453 ± 171.6	0.93 ± 0.05	
with bTagEventWeight	22087.4 ± 1205.8	22197 ± 149.0	1.00 ± 0.06	7% improvment
DeepCSVM :				
No bTagEventWeight	23942.4 ± 1303.0	21259 ± 145.8	0.89 ± 0.05	
with bTagEventWeight	16075.3 ± 879.3	15808 ± 125.7	0.98 ± 0.05	8% improvment

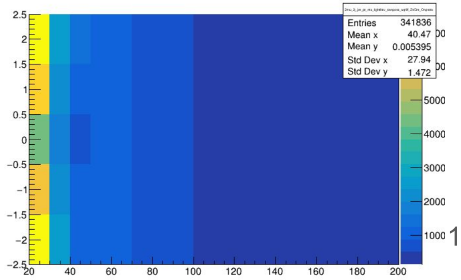
B hadron flavour



C hadron flavour



Light hadron flavour

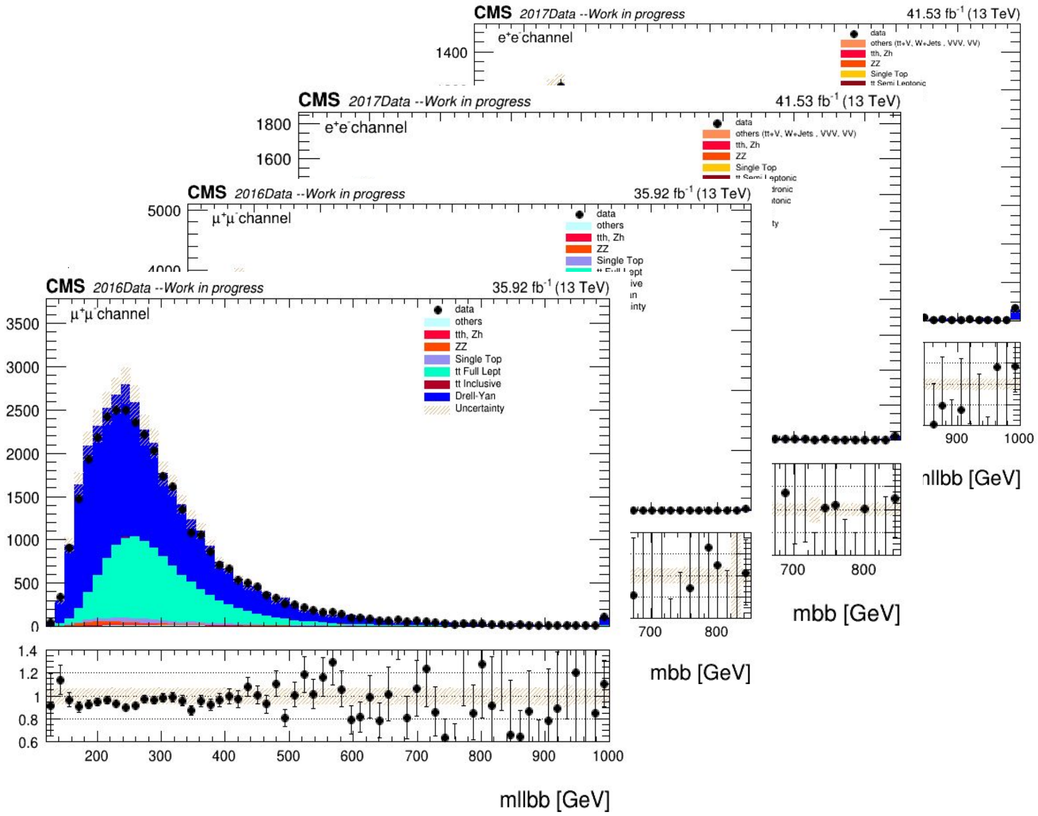


[1] https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods

Final selection:

Medium b-tag discriminator cut:

- ❑ **Resolved :** both the AK4 CHS PF jets to pass the medium b-tagging wp.
- ❑ **BoOsted:** b tagging can be applied either on the AK8 jet, on jets matched to AK8 jets within ΔR(AK4, AK8)<0.4, or on its sub-jets. Since the latest perform better both sub-jets are required to be b-tagged.
- ❑ MET pT < 80 GeV + [MET Filter Recommendations for Run II]



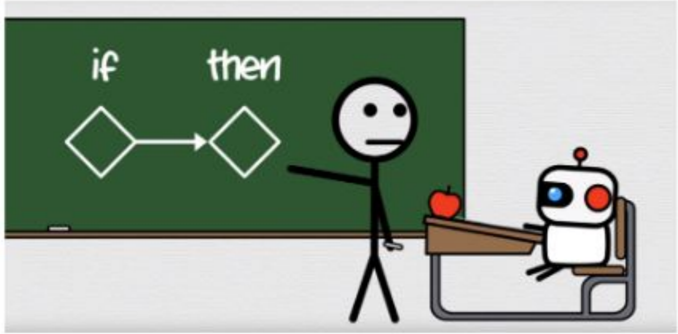
ML techniques for enhanced sensitivity:

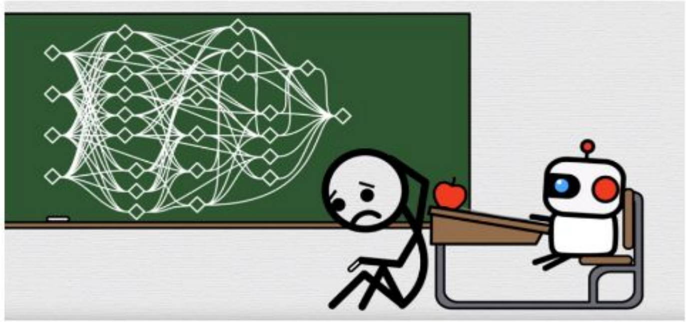
It may be much easier than you think !

- Use all the available signal samples in the training phase , the classifier is then able to guess the dependence of the signal behavior on these parameters.

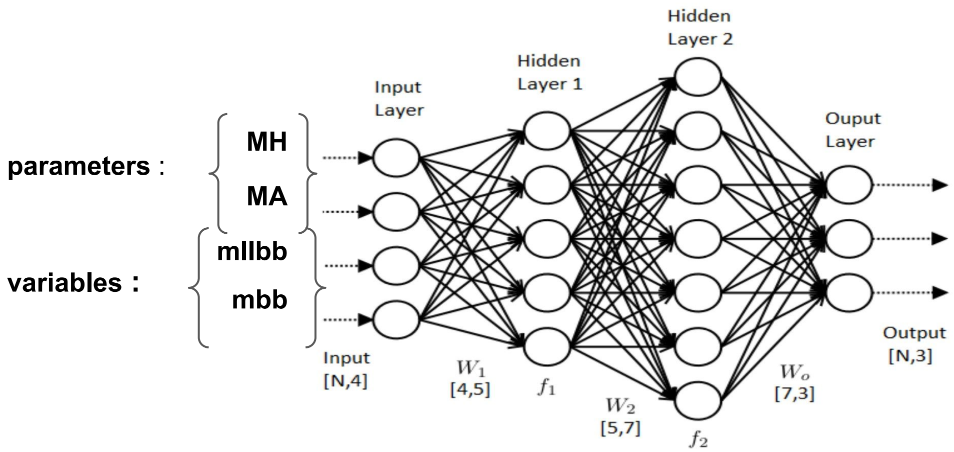
- When evaluating the classifier on data and simulation to derive the distributions needed for the signal search, the signal parameters are frozen to a specific value and only the signal sample corresponding to that value is considered.

- This procedure is repeated for every parameter value for which a signal sample is available





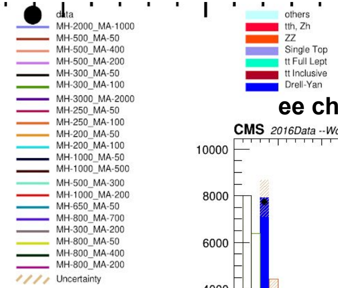
Signal extraction using Deep Neural Network:



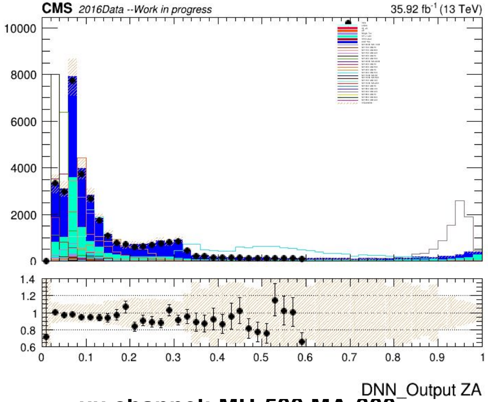
The result model should be able to interpolate the behaviour of the signal as function of the signal parameters and also perform well on samples not seen during the training phase.

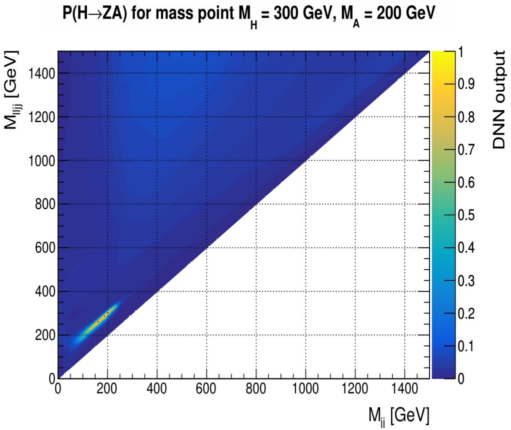
- **Variables:** Reconstructed masses
  mbb : Invariant mass of at least 2 AK4 b-tagged jets (resolved) or at least 1 AK8 b-tagged jet (boosted)
  mllbb : Invariant mass of 2 leptons + at least 2 AK4 b-tagged jets (resolved) or 2 leptons + at least 1 AK8 b-tagged jet (boosted)
- **Parameters:**
  (MA, MH) : Signal mass hypothesis.

Distributions of the classifier scores in data and in mc :

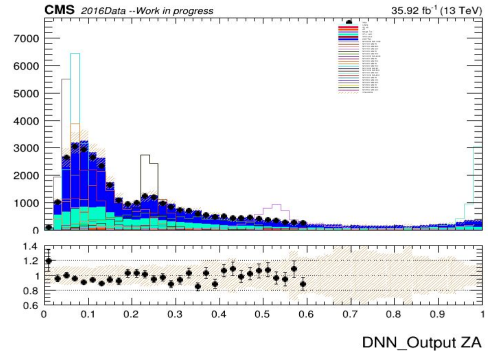


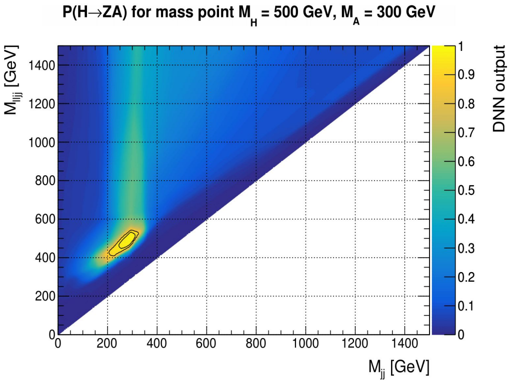
ee channel: MH-300 MA-200





μμ channel: MH-500 MA-300

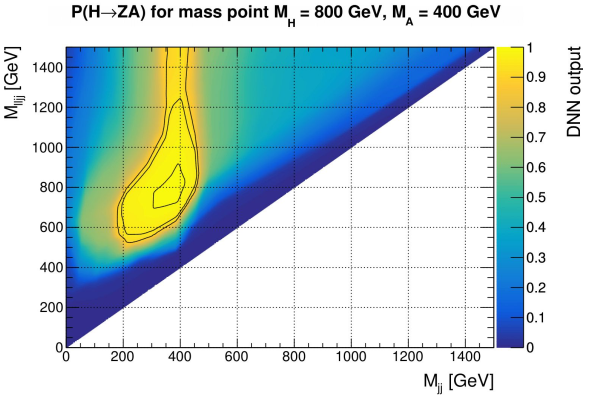


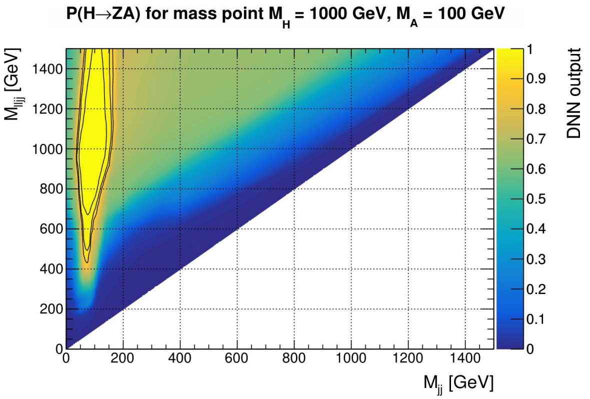


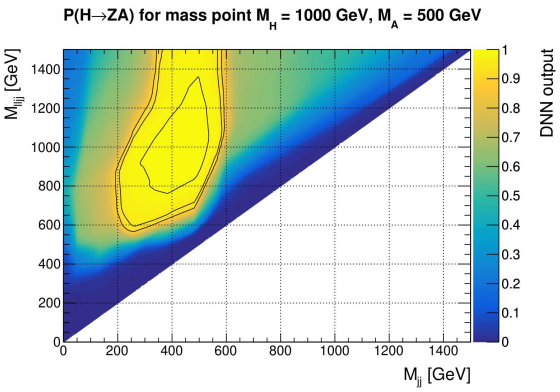
Note:
In the training we consider only the main background process in the training (**DY, TTbar+Single Top**) after requiring all the selection criteria described before.

BoOsted regions

- For important mass differences between H and A, the daughter particle is produced with an important boost resulting in close b-jets;
- The curves enclosing the bulk of the signal in the mass plane are drawn only to emphasize the elliptical shape in the resolved region as the previous CMS study showed.





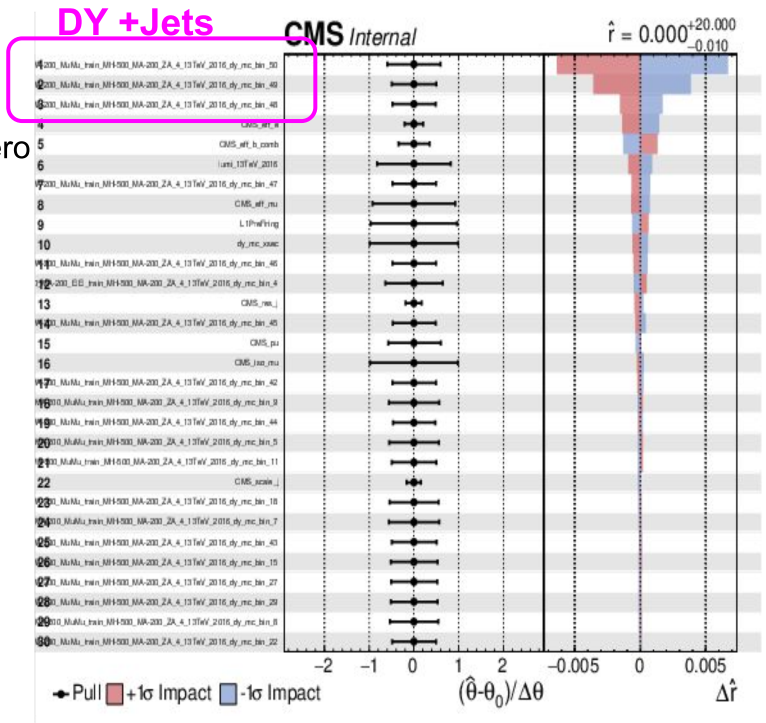


Pull and Impacts parameters

- Impacts represent a useful tool for determining which nuisance parameters have the largest effect on the POI (signal strength r) uncertainty.
  - ❖ **If central value != 0 :** means the data absorbed by non zero value => Investigate large pull
  - ❖ **If uncertainty <1 :** systematic constrained by data => Needs checking if this legitimate or a modeling issue

We have to judge if our analysis should really be able to provide more information about this parameter than the external measurement that gave us the input uncertainty.

MH-500_MA-200



Plans ...

- There's a lots on my ToDo List …
  - Validate the behavior of the Parameterized DNN for signal extraction
  - Optimize the best cut and rebining for our DNN output for the best trade-off between stat and syst uncertainties
  - Fun part: Looking to the Goodness of Fit; Does data agree with our theory and Hypothesis Testing; which theory our data prefer BSM or SM ?
  - Finalizing mc production is our PRIORITY !

THANKS :)

BACKUP

In the alignment limit :

- **cos(β−α)= 0 :** the decays of the heavy neutral Higgs H→ hh, AA are absent and the SM Higgs self coupling obtains its SM value, while a decay of h→ AA is possible if kinematically open.

$$g_{Hhh} = -\frac{c_{\beta-\alpha}}{v}\frac{s_{2\alpha}}{s_{2\beta}}(m_H^2 - m_h^2) + \frac{c_{\beta-\alpha}}{2v}m_H^2,$$

$$g_{HAA} = -\frac{c_{\beta-\alpha}}{2v}(m_H^2 - 2m_A^2),$$

$$g_{hhh} = -\frac{c_{\beta-\alpha}^2}{v}\left[\frac{c_{\beta-\alpha}}{t_{2\beta}} - s_{\beta-\alpha}\right](m_H^2 - m_h^2) + \frac{s_{\beta-\alpha}}{2v}m_h^2,$$

$$g_{hAA} = -\frac{s_{\beta-\alpha}}{2v}[2(m_H^2 - m_A^2) - m_h^2] - \frac{c_{\beta-\alpha}}{t_{2\beta}v}(m_H^2 - m_h^2).$$

- The couplings of the neutral CP-even Higgses to a pair of vector bosons are:

$$g_{hVV} = \frac{m_V^2}{v} s_{\beta-\alpha} \quad \text{and} \quad g_{HVV} = \frac{m_V^2}{v} c_{\beta-\alpha},$$

- Additionally, the neutral CP-even Higgses can couple to the CP-odd Higgs A and a Z-boson with couplings:

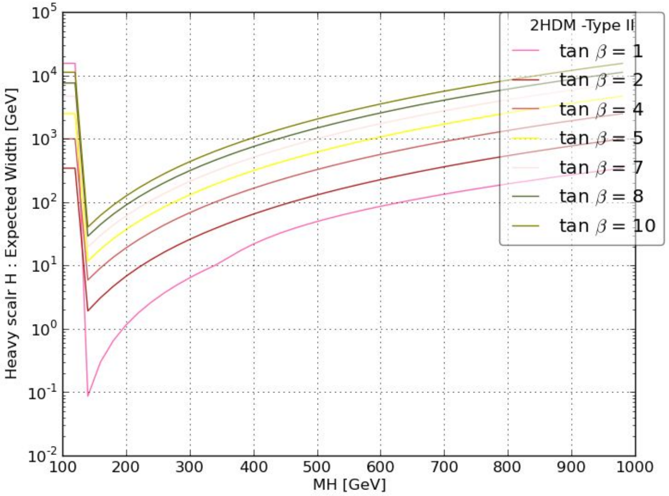
$$g_{hAZ} = \frac{g c_{\beta-\alpha}}{2 c_{\theta_W}} (p_h - p_A)_\mu$$

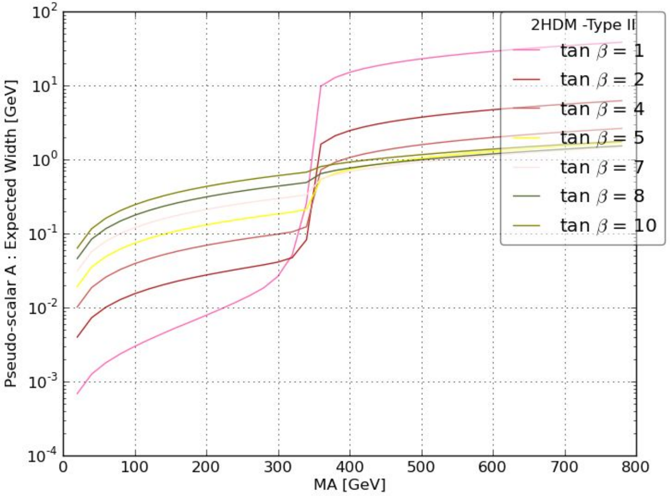
$$g_{HAZ} = -\frac{g s_{\beta-\alpha}}{2 c_{\theta_W}} (p_H - p_A)_\mu ,$$

- LHC Higgs coupling measurements favor the alignment limit, $s_{\beta-\alpha} \approx 1$ in which the couplings of the 125 GeV Higgs to fermions and gauge bosons are consistent with those predicted by the SM **=> non-observation of such a state in the H -> V V channel**
- In the alignment limit, the CP-odd Higgs will couple more strongly to the BSM Higgs H than its SM-like counterpart h. In particular, this implies that A is more likely to decay to HZ than hZ, if kinematically possible.
⇒ This motivates the exotic Higgs searches for A-> HZ and H-> AZ as complementary probe in the alignment limit.

Heavy Higgs H , Pseudo-scalar A Width

- We use the `SusHi` package to calculate the production cross-section at NNLO level, and the `2HDMC` code for Higgs decay branching fractions at tree level.





EPR Tasks :

- **BTV-POG:** B-tagging Efficiencies Measurements using 2TagCount method in ttbar dilepton events at 13 TeV :
  - 2018 and 2017 using MiniAOD Ntuples.
  - We moved to NanoAOD v5 then UL: works is done for 2017UL , WIP for 2018UL
  - My talks:
    - https://indico.cern.ch/event/853831/#6-twotagcount-report
    - https://indico.cern.ch/event/916806/#69-2tagcount-ul17-sf-report
    - https://indico.cern.ch/event/862144/#77-updates-on-2tagcount-method
    - https://indico.cern.ch/event/828712/#9-2tagcount-sf-results-for-201
  - My NanoAOD -git repo: https://github.com/kjaffel/b-taggingEfficienciesMeasurment ,
  - MiniAOD repo: https://github.com/kjaffel/TTbarCalib
- **Tracker-DPG:**
  - Cluster Parameter Estimator: Test the compatibility between the predicted width based on track direction and the reconstructed one.
  - My Talks:
    - https://indico.cern.ch/event/879450/#45-cpe-error-studies
    - https://indico.cern.ch/event/934813/#60-cpe-reparameterization

MC simulated samples:

DY:

- /DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/
- /DYJetsToLL_{0,1,2}J_TuneCP5_13TeV-amcatnloFXFX-pythia8

TTbar-Full hadronic

- /TTToHadronic_TuneCP5_13TeV-powheg-pythia8

TTbar-Full leptonic

- /TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8

TTbar-Semi leptonic

- /TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/

Single Top

- /ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/
- /ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/
- /ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8/
- /ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8/
- /ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/

Standrd model higgs : Zh & tth

- /ZZTo2L2Nu_13TeV_powheg_pythia8/
- /ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/
- /ZZTo4L_13TeV_powheg_pythia8/
- /HZJ_HToWW_M125_13TeV_powheg_jhugen714_pythia8_TuneCP5/
- /ZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/
- /ggZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/
- /ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8/
- /ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8/

Others= WZ/ WW / VVV / W+Jets / ttbar+V

- /WWToLNuQQ_NNPDF31_TuneCP5_13TeV-powheg-pythia8/
- /WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8/
- /WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/
- /WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8_v2/
- /WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/
- /WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/
- /WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/
- /WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/
- /WZZ_TuneCP5_13TeV-amcatnlo-pythia8/
- /ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/
- /WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/
- /TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/
- /TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/
- /TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8/
- /TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/

Data:

- /DoubleMuon/Run2017{B, C, D, E, F}-02Apr2020-v1/NANOAOD
- /MuonEG/Run2017{B, C, D, E, F}-02Apr2020-v1/NANOAOD
- /DoubleMuon/Run2017{B, C, D, E, F}-02Apr2020-v1/NANOAOD

The performance of b tagging for boosted events:

- POG provide # taggers; some clearly better than others, but many taggers behave similarly & details depend on analysis (+ MC choice) !
- The performance of AK8 and subjet b tagging is compared. When b tagging is applied to the subjets of boosted top quark jets, at least one of the subjets is required to be tagged. In addition, the performance of b tagging applied to AK4 jets matched to AK8 jets within$\Delta R(AK4,AK8)<0.4$ is also shown. When b tagging is applied to AK4 jets matched to the AK8 jet, at least one of the AK4 jets is required to be tagged.