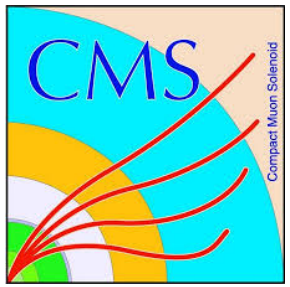
Search for ttH production in boosted all-jet final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the CMS detector

Garyfallia (Lisa) Paspalaki

National Center of Scientific Research "Demokritos" and National Technical University of Athens (NTUA)

in collaboration with
NTUA CMS group: K.Kousouris, Y. Tsipolitis

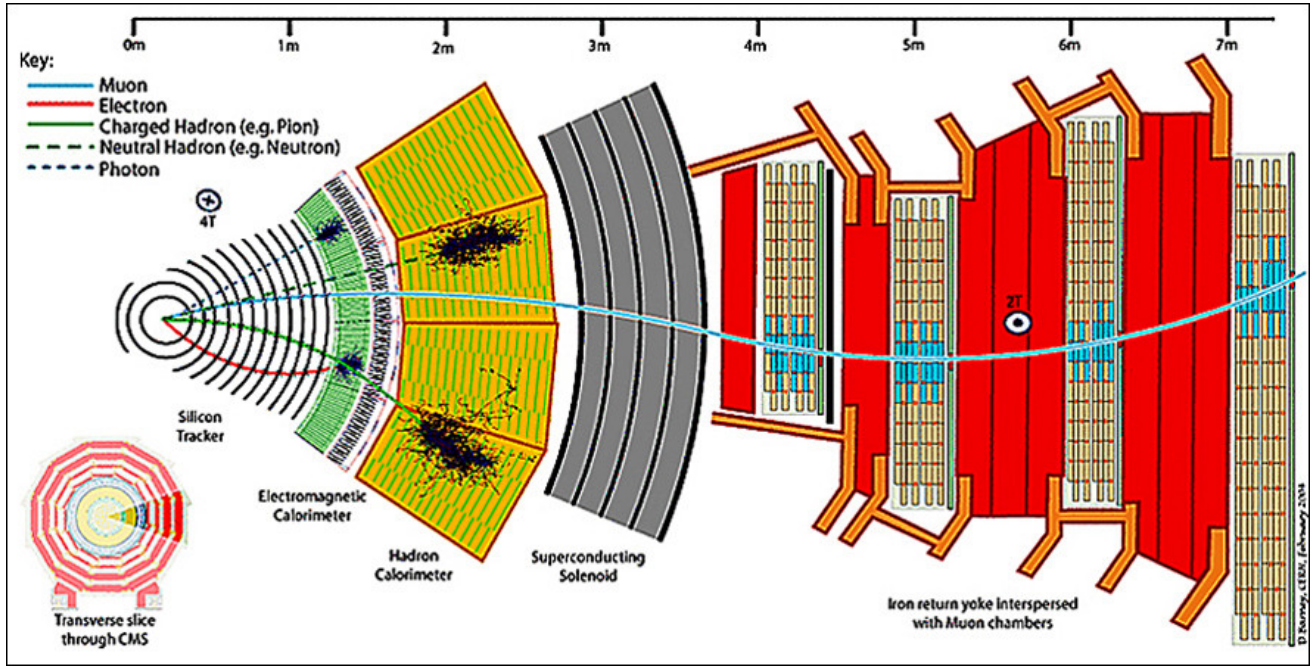
HEP 2019 Conference of recent developments in High Energy Physics and Cosmology
19 April 2019



Outline

- The CMS experiment
- ttH @ 13 TeV
- ttH in boosted topologies
- analysis strategy
- conclusions

The CMS detector



The Compact Muon Solenoid (CMS) is a general purpose detector and is designed to observe any new physics phenomena at the LHC

CMS detector has many detection layers

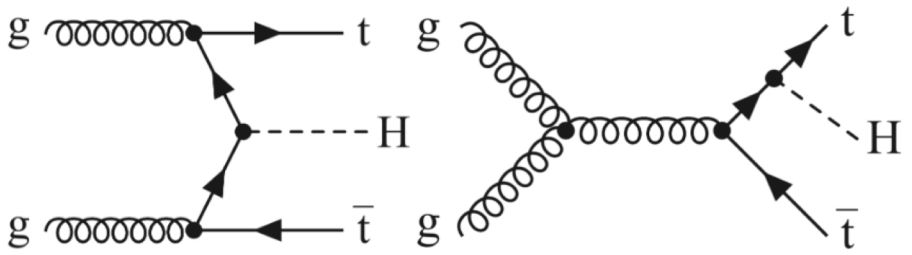
- tracker
- Electromagnetic Calorimeter (ECAL)
- Hadron Calorimeter (HCAL)
- Muon Detectors

the object reconstruction is done using sophisticated algorithms that take into account information **from all the detector layers**

Higgs boson production in association with a top quark-antiquark pair

ttH@13TeV

- ttH process is **observed** with a significance of 5.2 σ (arXiv:1804.02610v2)
- combining statistically independent searches for several Higgs bosons decaying channels
- **direct probe of the top–Higgs coupling**

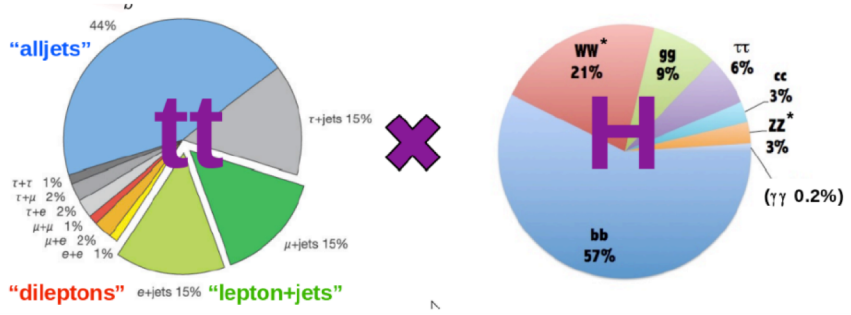


H(bb) @13 TeV

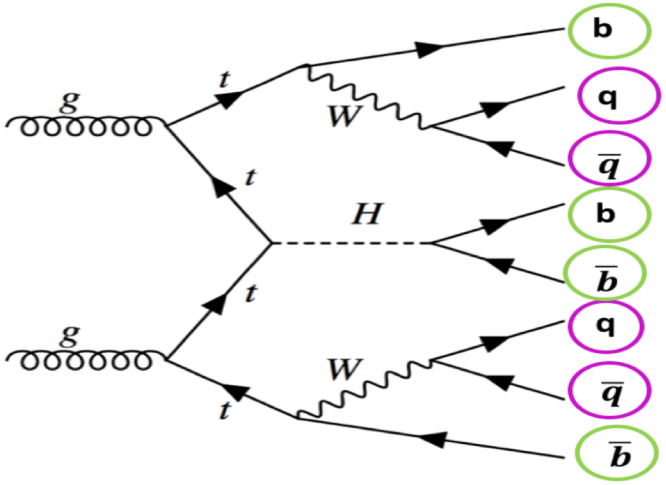
- SM Higgs Boson decay to a pair of bottom quarks is observed with significance 5.6 σ (arXiv:1808.08242v2)
- main contribution from Higgs production in association with W or Z boson
- **H → bb has the largest branching fraction compared to other decay channels**

ttH(bb) @ 13 TeV

- combined search of ttH(bb) for several decay channels
- events in the final state with lepton and jets
- **all hadronic represents ~ 46% of all ttHbb final state**
- **specific Higgs coupling space**: all couplings are fermionic and restricted to the third-generation quarks only



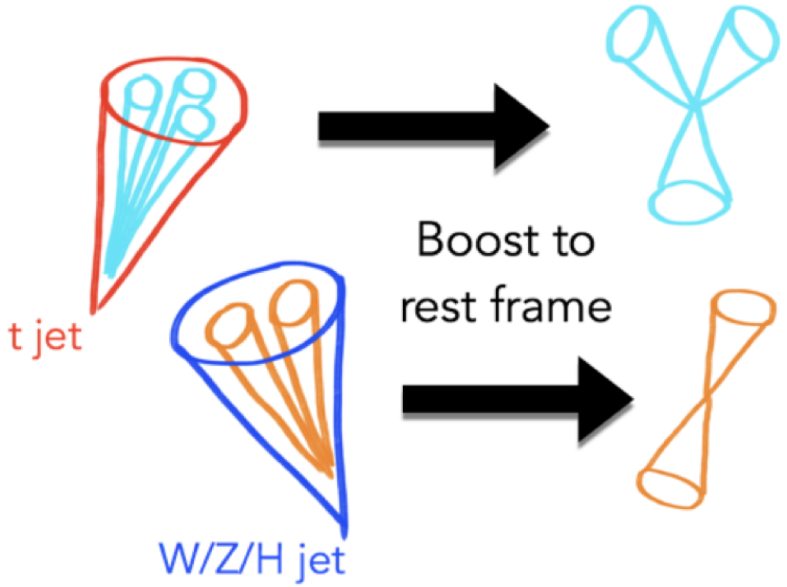
ttH(bb) in fully boosted hadronic final states



- fully hadronic ttH(bb) is challenging to target
- zero leptons
- **8 jets** in the final state at least **4 of them b tagged**
- successfully identify the jets special identification techniques are developed (arXiv:1803.06986v2)

Boosted Topologies

- high P_T ***"boosted"*** jets can be reconstructed within **a large radius jet** with all the **decay products merged**
- "*large R jets*" the contain information about the Higgs and Top decays products
- setup an analysis using the full **2016 dataset (35.86 fb^{-1})**
- use MVA techniques to tag boosted Candidates (**Higgs** and **Top**)



Trigger and Object Selection

Signal Trigger

- boosted jets reconstructed within a large radius cone of 0.8
- large hadronic activity (small R and boosted jets) in the events
- the signal efficiency is studied with respect of the **sum of P_T off all jets (S_T)** in the event and is found to be ~98% efficient

Boosted Jets

- P_T >200 GeV
- $|\eta| < 2.4$
- JetMassSoftDrop >50 GeV
- Tight ID requirements

Leptons

- electrons and muons
- P_T > 20 GeV
- $|\eta| < 2.4$

Small Radius Jets

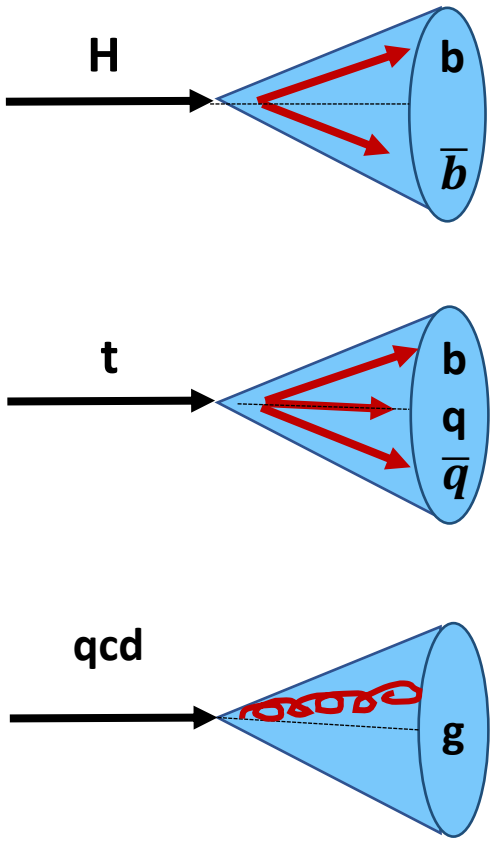
- P_T >30 GeV
- $|\eta| < 2.4$
- Tight ID requirements
- cleaned from boosted jets

Baseline Selection

- at least one boosted jet
- leading Jet P_T >300 GeV
- S_T >900 GeV
- lepton veto

MVA Training

Idea: Train MVA Boosted Decision Trees (BDTs) to identify **boosted candidates (Higgs, Top)**



boosted training:

- boosted jets that pass the basic selection

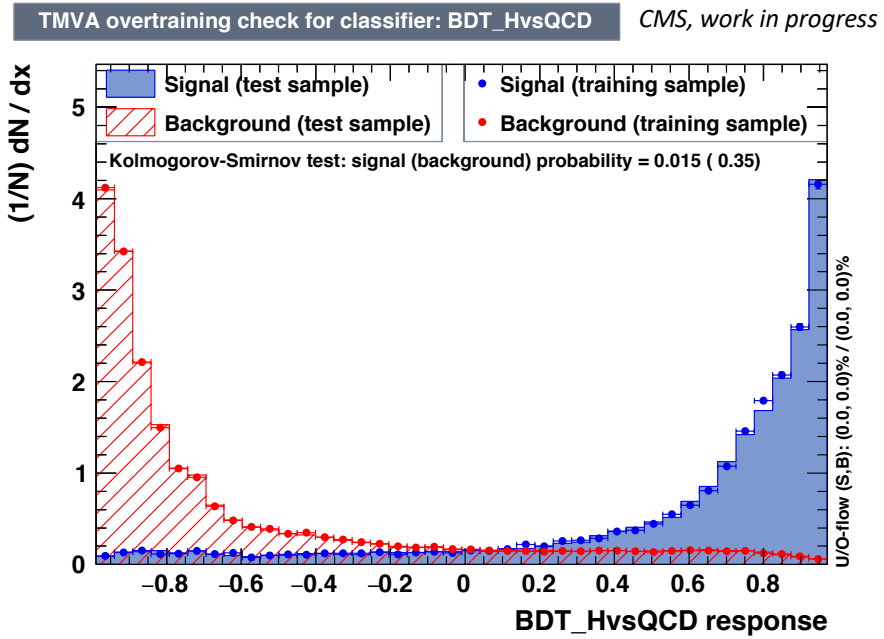
Perform 3 individual trainings

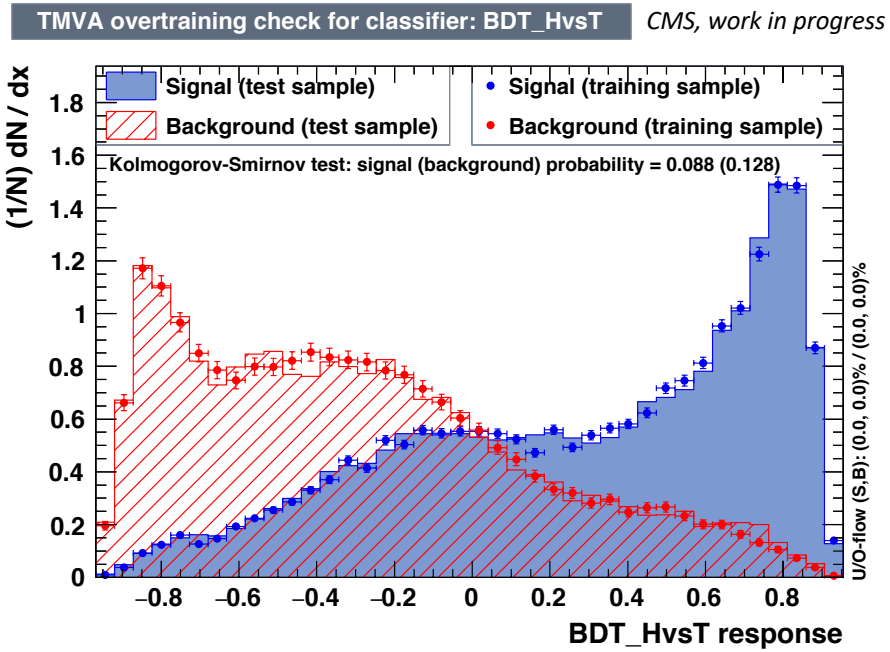
- use the same variables with good separating power
- train **boosted Higgs** Candidates against **QCD Jets** (**HvsQCD**)
- train **boosted Higgs** Candidates against **Top Jets** (**HvsT**)
- train **boosted Top** Candidates against **QCD Jets** (**TvsQ**)

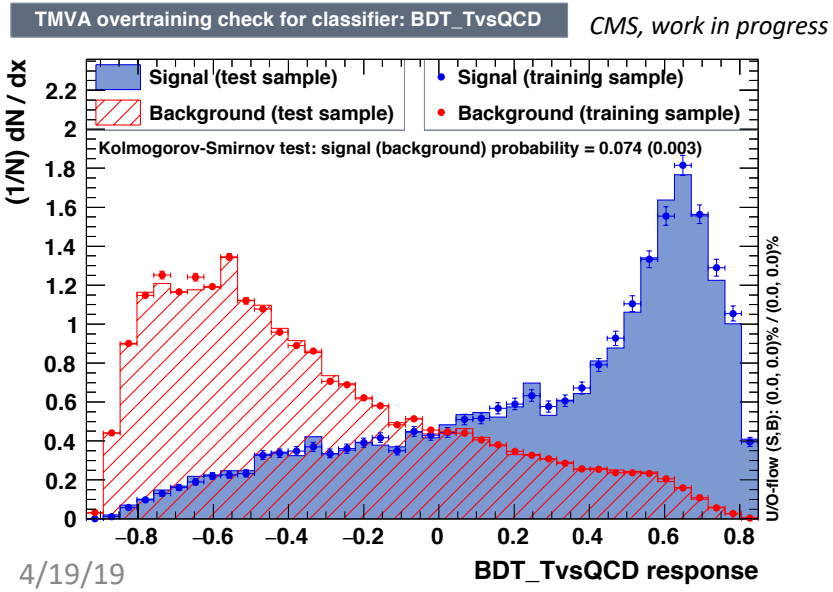
Variables use for training for boosted candidates:

- Jet substructure variables
- score of the dedicated btagging discriminator (CSVv2) of the leading and subleading jet
- Mass of the leading and subleading subjet

BDT Responses







- Good discriminating power for HvsQ and TvsQCD
- able to distinguish between Higgs and Top Candidates
- cuts on BDT are optimized based on the expected limit
- use a combination of the scores to identify Higgs and Top candidates

Boosted Candidate Tagging

Higgs Candidate

- jet with the highest BDT_HvsQ + BDT_HvsT
- BDT_HvsQ>0.8 , BDT_HvsT>0.1
- jetPt>300 GeV
- MSD >70 GeV

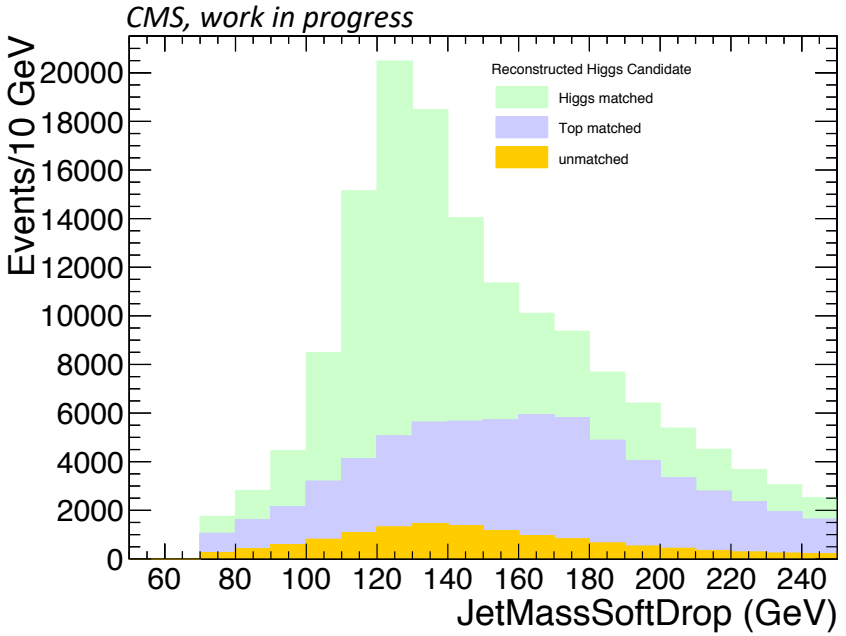
Higgs Tagging: 54% (matched_Higgs/Reco Higgs)

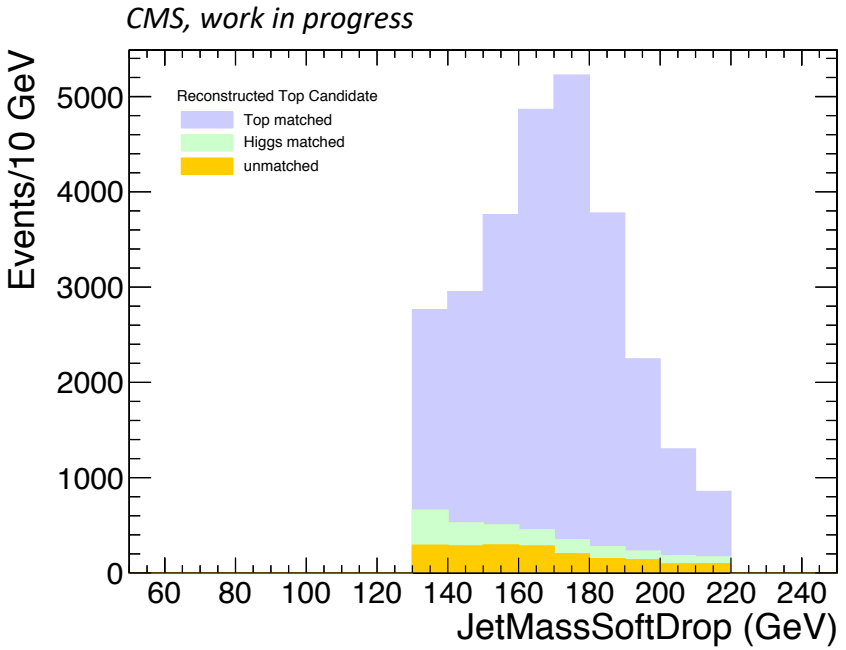
Top Candidate

- not the higgs Candidate
- jet with the highest TvsQ score
- jetPt>300 GeV
- BDT_TvsQ>0.5
- 130<MSD<220 GeV

Top Tagging : 88% (matched_Top/Reco Top)

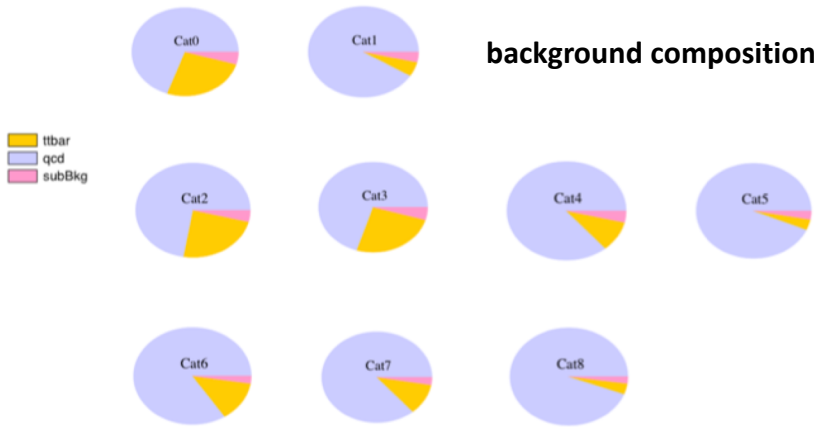
(*) Parton Matching: $\Delta R_{minimum}$(jet, parton) < 0.3 and PartonId=25 (Higgs) PartonId = ±6 (Top)





Analysis Categories- Boosted Higgs Candidate Channel

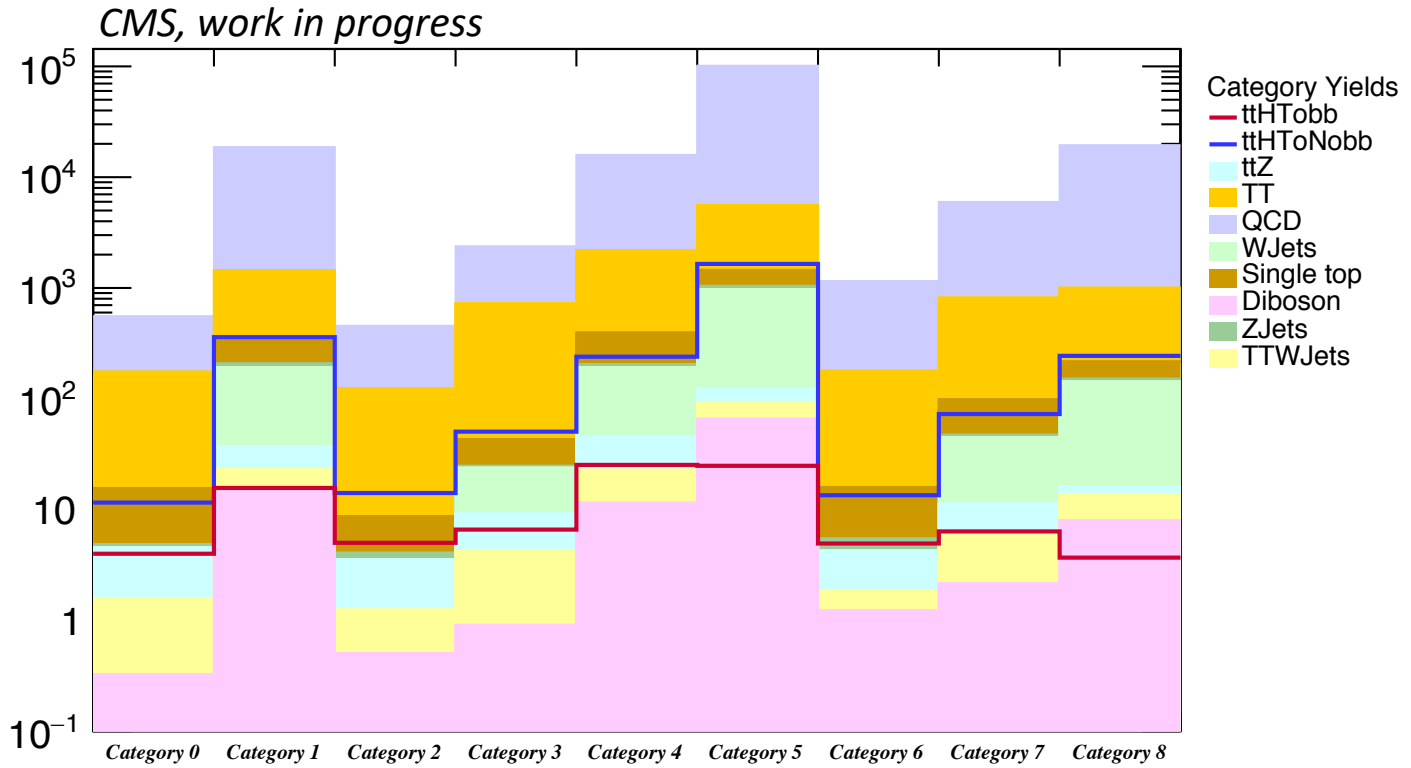
Categories	Nak8	Higgs Tagged	Top Tagged	Nak4	NAk4Bjets
0	3	✓	✓	–	–
1		✓	✗	–	–
2	2	✓	✓	>1	1
3		✓	✓	>1	0
4		✓	✗	>1	1
5		✓	✗	>1	0
6	1	✓	–	>4	≥2
7		✓	–	>4	≥ 1
8		✓	–	>4	=0



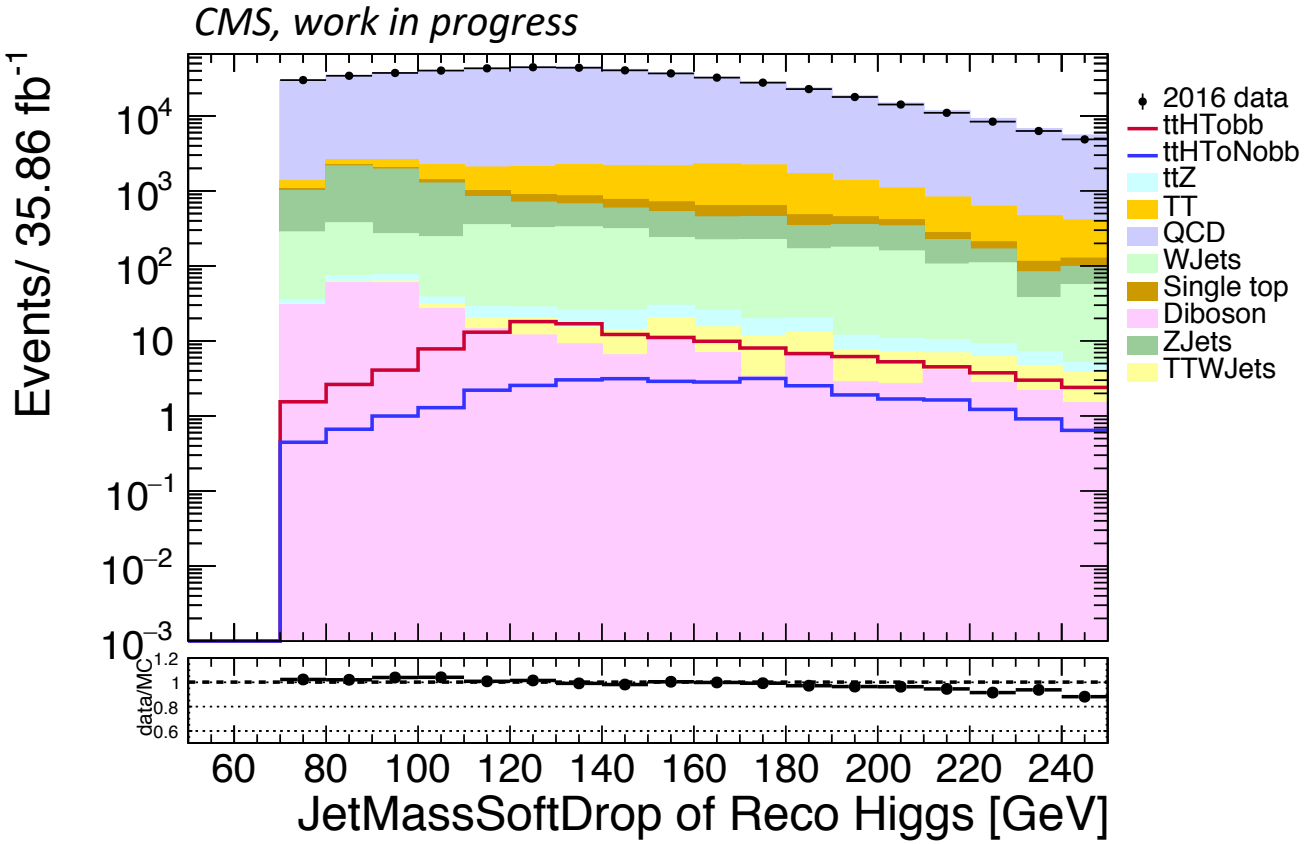
background composition

- 9 orthogonal categories based on the number of boosted jets
- Higgs candidate **always** is reconstructed as a boosted Jet
- **Observable**: Jet Mass Soft drop of the reconstructed Higgs Candidate

Expected Yields for each category



Reconstructed Higgs Candidate



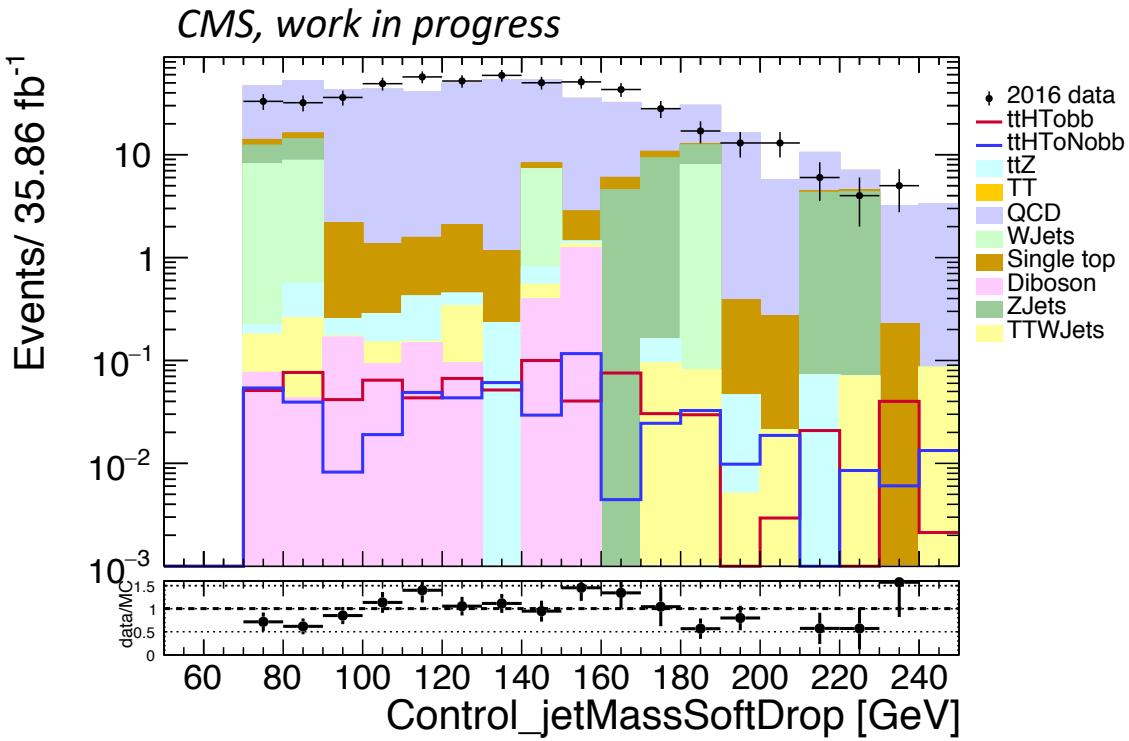
- reconstructed mass *"JetMassSoftDrop"* of the boosted Higgs Candidate before entering in the analysis categories
- 2016 data are compared with simulation
- The simulated backgrounds are scaled to the luminosity of the data and then the QCD background is further scaled to match the data yield
- analysis is blinded in Higgs Mass Window for all the categories

Data-Driven QCD estimate for Boosted Higgs analysis

- most significant background
- model the QCD shape from the data
- use this shape to set a systematic uncertainty on the overall QCD shape

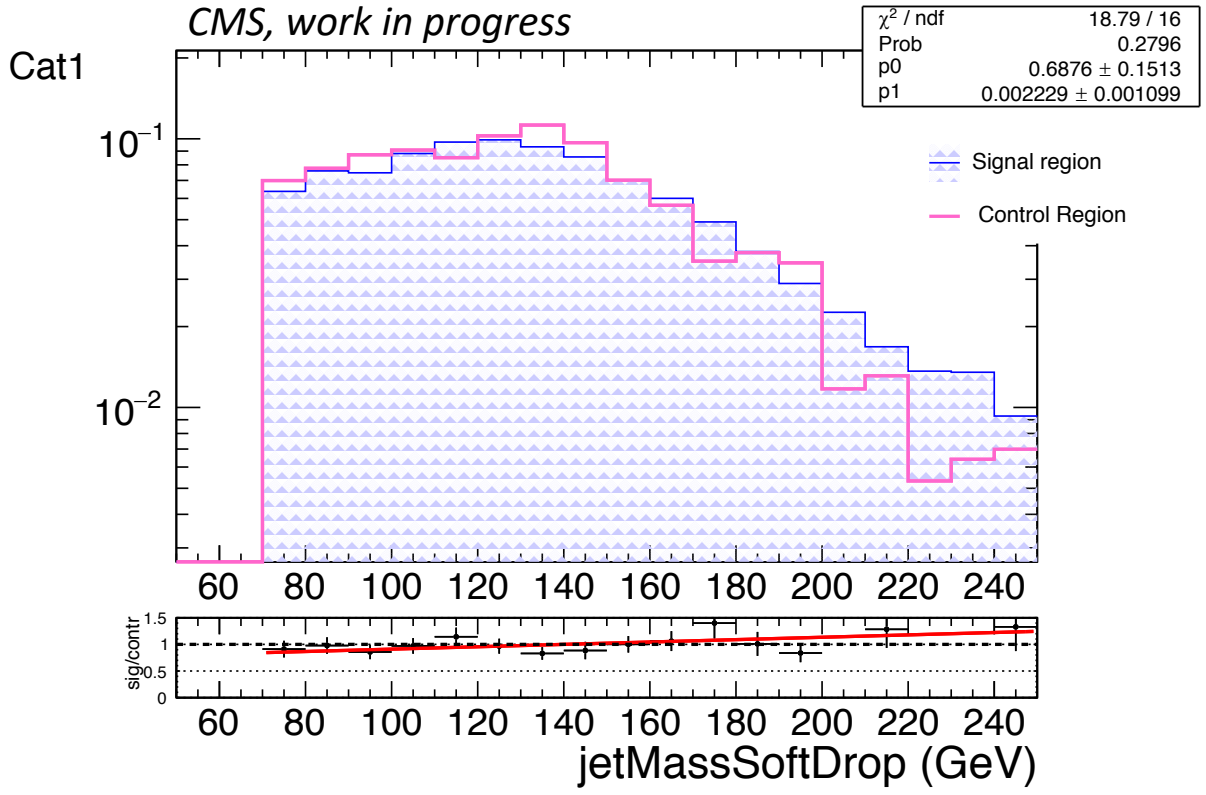
Boosted QCD Control Region Selection

- high P_T boosted jets
- BDT scores reversed



MC closure test

`Category 1:( 3 Ak8Jets, boosted Higgs, T=0 )`



Analysis Sensitivity - Systematic Uncertainties

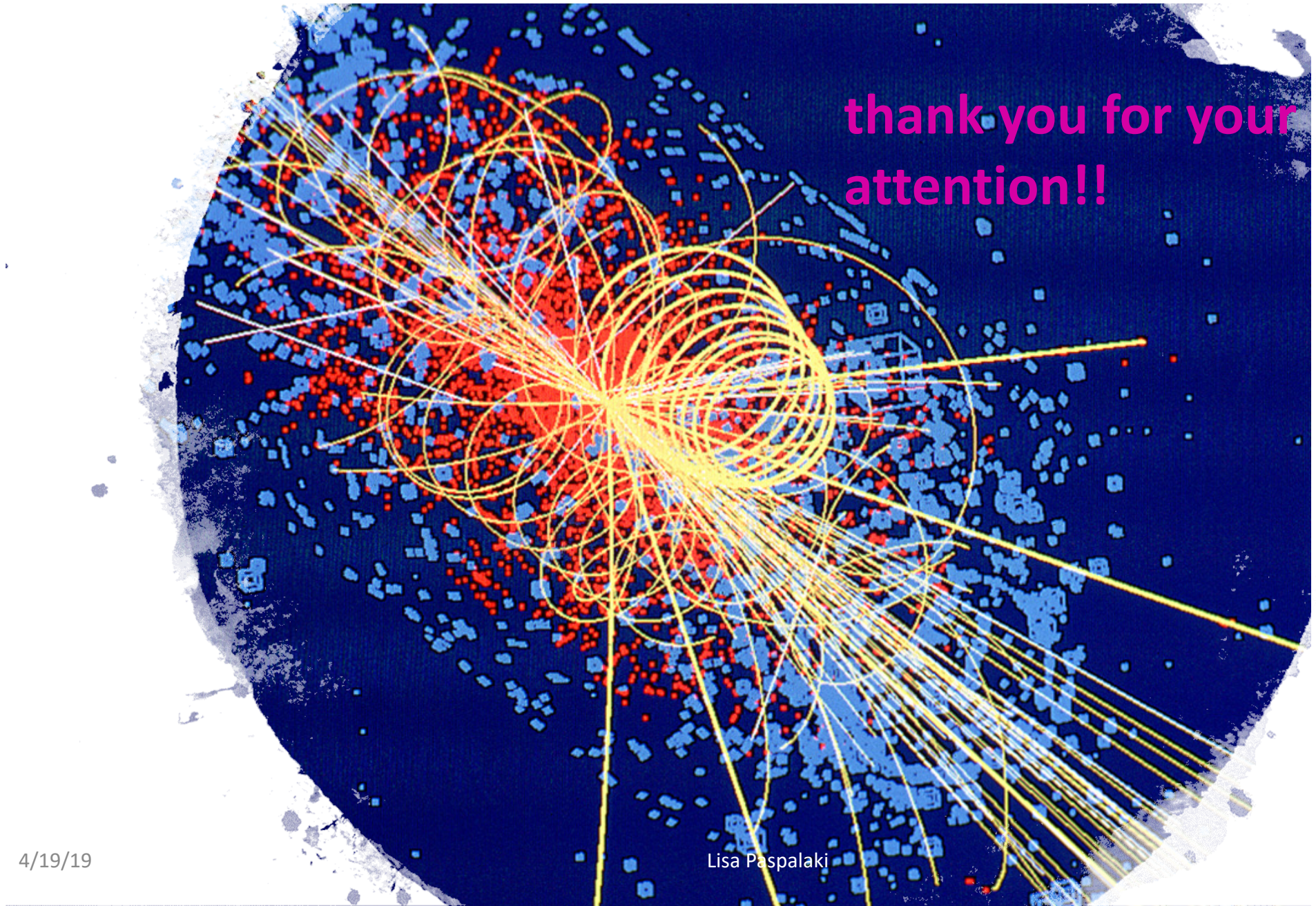
- simultaneous fit in all the categories
- take into account not only the expected events but also **the shape of the Higgs mass distribution**
- expected limit (analysis sensitivity) ~10

Systematic Uncertainties

- normalization of the background (affect the expected events)
- shape of the background (for QCD estimated from the data)
- b tagging (ability to identify successfully b quarks)
- other uncertainties need to taken into account

Conclusions-Future steps

- Analysis set up for 2016 data set (35.6 fb^{-1})
- we are still blinded in Higgs Mass window
- maximize the analysis sensitivity in simulation
- plan to use the full Run2 data set (140 fb^{-1})
- this will lead to better expected limit


thank you for your
attention!!