$HH \rightarrow bb\tau\tau$ search at CMS

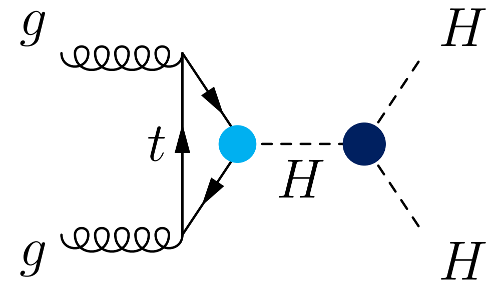
KONSTANTIN ANDROSOV[1] , MARIA TERESA GRIPPO[1,2]

[1]INFN Pisa, [2]University of Pisa

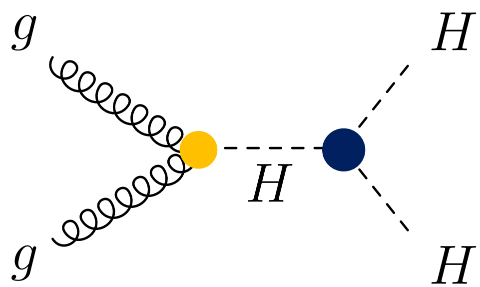
CMSDAS Pisa January 2019

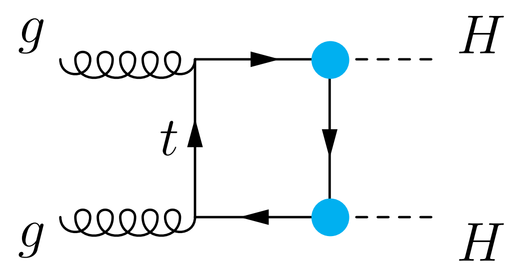
Motivation – Non Resonant production

- $\sigma_{SM}(HH) = 33.49\ fb$ for $m_H = 125\ GeV$ at $\sqrt{s} = 13\ TeV$ (LHCHXSWG Yellow Report 4)
  - SM Double Higgs production not accessible with current data
  - The Beyond Standard Model (BSM) scenarios can be still explored, defining an Effective Field Theory (EFT) Lagrangian [1]:

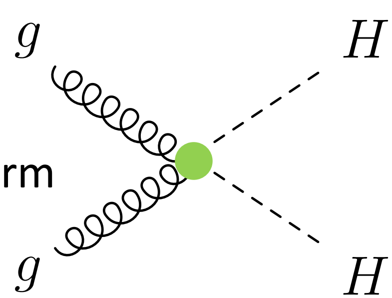


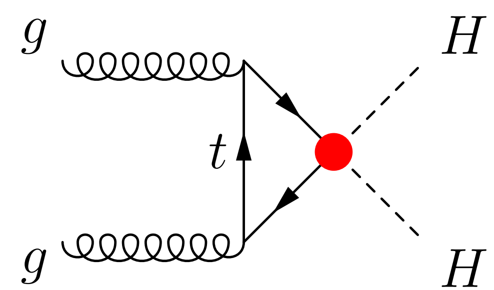
$$L_{hh} = \frac{1}{2}\partial_\mu h \partial^\mu h - \frac{m_h^2}{2}h^2 - k_\lambda \lambda_{SM} v h^3$$
$$- \frac{m_t}{v}\left(v + k_t h + \frac{c_2}{v} hh\right)(\bar{t}_L t_R + h.c.)$$
$$+ \frac{\alpha_s}{12\pi v}\left(c_{1g} h - \frac{c_{2g}}{2v} hh\right) G^A_{\mu\nu} G^{A,\mu\nu}$$





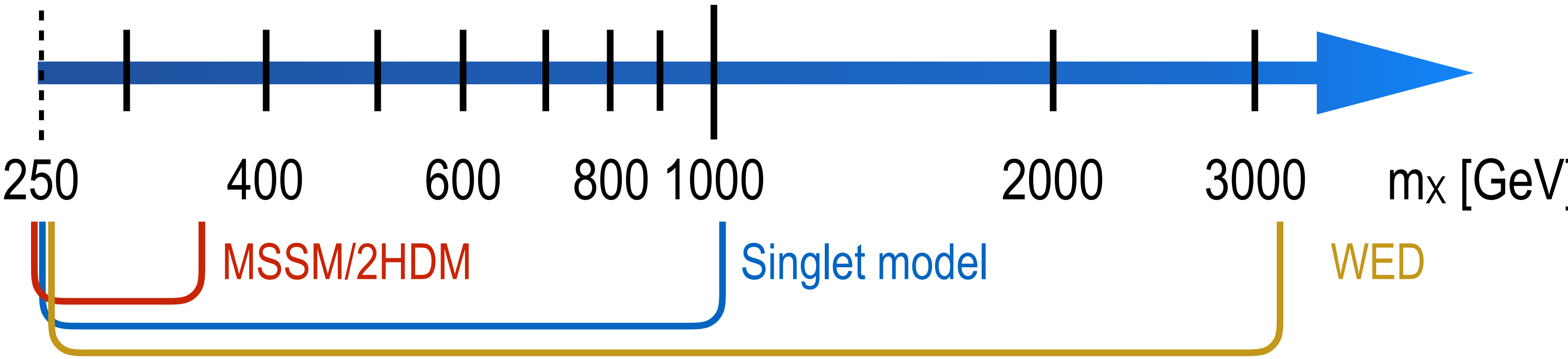
- Test non-resonant BSM effective models with anomalous couplings:
  - Define 2D-planes (e.g. $k_t k_\lambda$-plane) within the parameter space and perform a grid scan inside each plane [2]
  - 12 benchmarks are defined

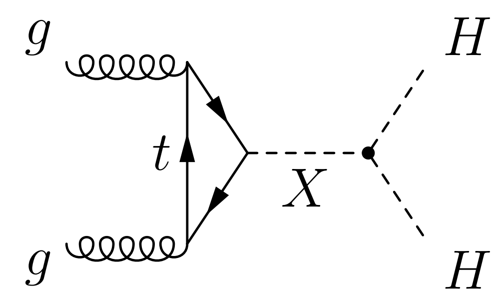




[1] doi:10.1103/PhysRevD.91.115008
[2] doi:10.1007/JHEP04(2016)126

Motivation – Resonant production

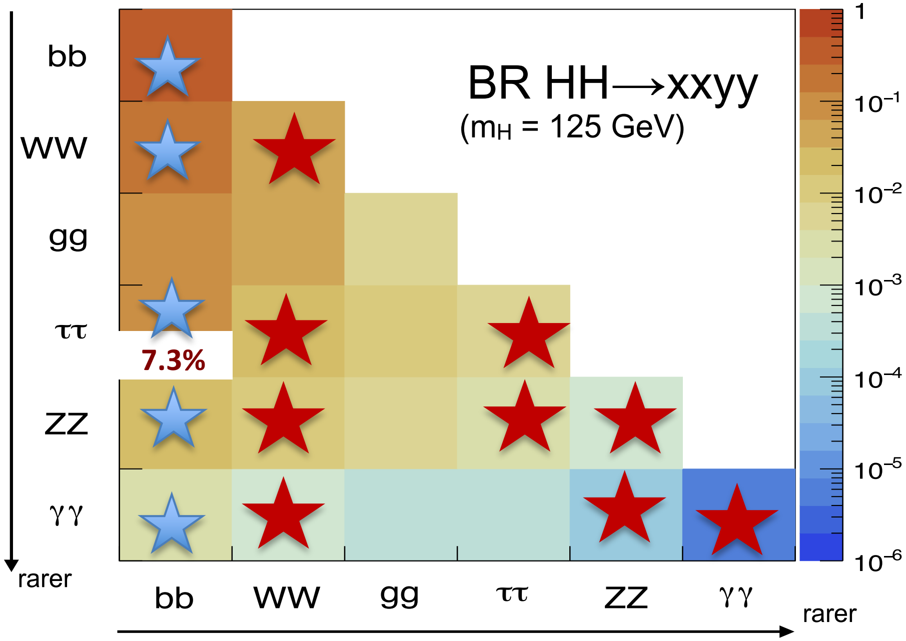


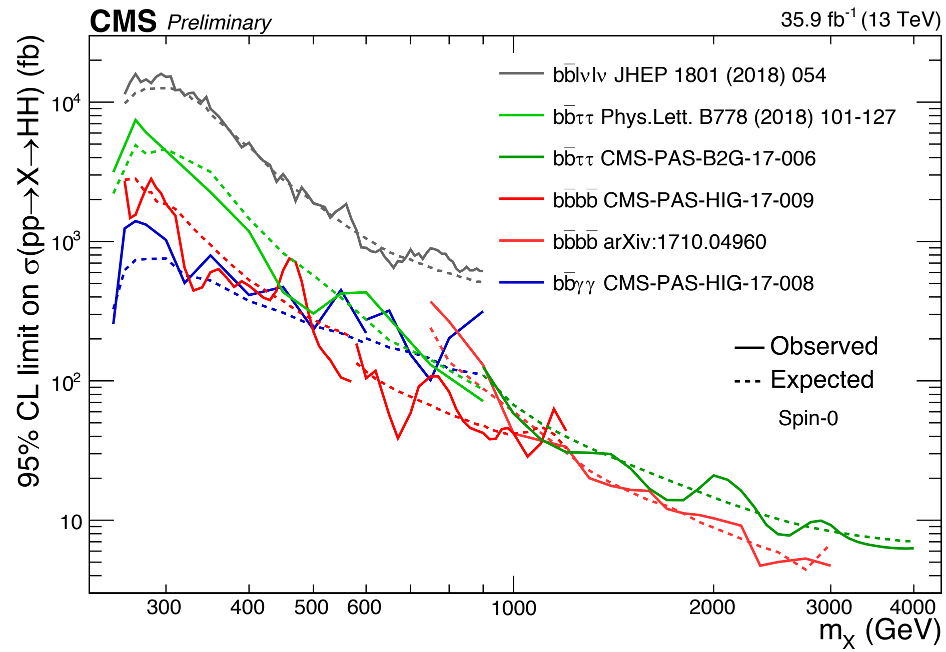


- Model independent search of narrow width resonance not predicted by the SM
- Different possible scenarios for a wide mass range:
  - MSSM low tanβ high [3], hMSSM [4] and Two Higgs Doublet Model (2HDM)[5]: Additional Higgs doublet→CP-even scalar H
  - Singlet model [6]: additional Higgs singlet with an extra scalar H; not negligible width at high m_H
  - Warped Extra Dimensions (WED): spin-2 (KK-graviton) [7] and spin-0 (radion) [8] resonances

[3] doi:10.1007/JHEP10(2013)028
[4] doi:10.1140/epjc/s10052-013-2650-0
[5] doi:10.1016/j.physrep.2012.02.002
[6] doi:10.1007/s002880050442
[7] doi:10.1103/PhysRevD.76.125015
[8] doi:10.1103/PhysRevD.76.036006

HH final States





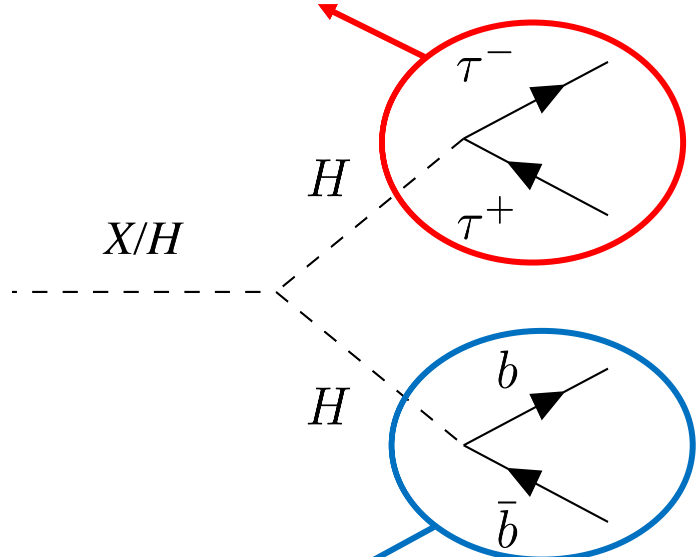
- Four channels are published at 13 TeV:
  - Blue star analyses entered in the combination
  - Red star analyses on going
- $bb\tau\tau$ final state:
  - robust analysis since Run1
  - trade off between BR and purity

$HH \rightarrow bb\tau\tau$

The double Higgs production can be detected through the reconstruction of its decay products.

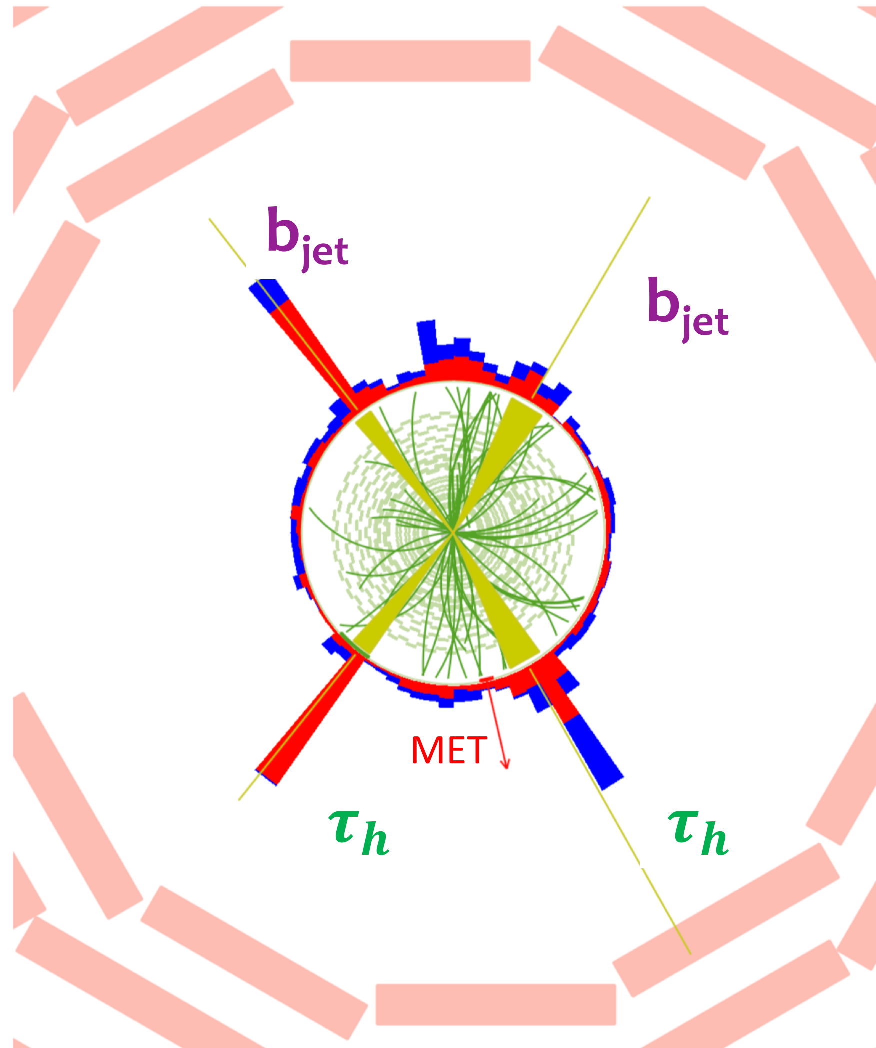
$$H \rightarrow \tau^{\pm}\tau^{\mp}\ (\mathcal{BR} \approx 6\%)$$

Good performance of CMS in reconstruction of τ leptons



$$H \rightarrow b\bar{b}\ (\mathcal{BR} \approx 57\%)$$

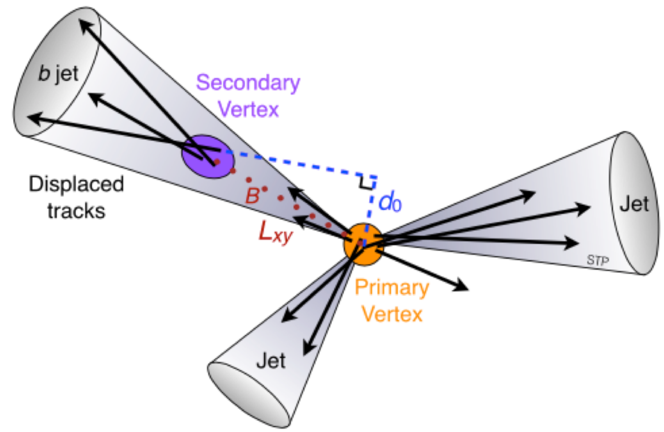
High branching ratio and good identification of b quarks



Reconstruction of the objects: b, τ_e, τ_μ, τ_h

b jet

In order to separate the b jets from the jets coming from gluons and light quarks (b-tag), the Combined Secondary Vertex (CSVv2) algorithm is used.
CSVv2 is based on the information from secondary vertices, on impact parameter and on distances of the tracks wrt the jet axis.



τ_h

The τ_h are reconstructed using the Hadron Plus Strips (HPS) algorithm.
HPS uses the information coming from the calorimeter e from the tracker to reconstruct the topology of the hadronic tau decays.

τ_e

The τ_e are reconstructed considering the energy deposits from ECAL, which have a link in the tracker.

τ_μ

The τ_μ are reconstructed using the combined information coming from the tracker and from the muon system.

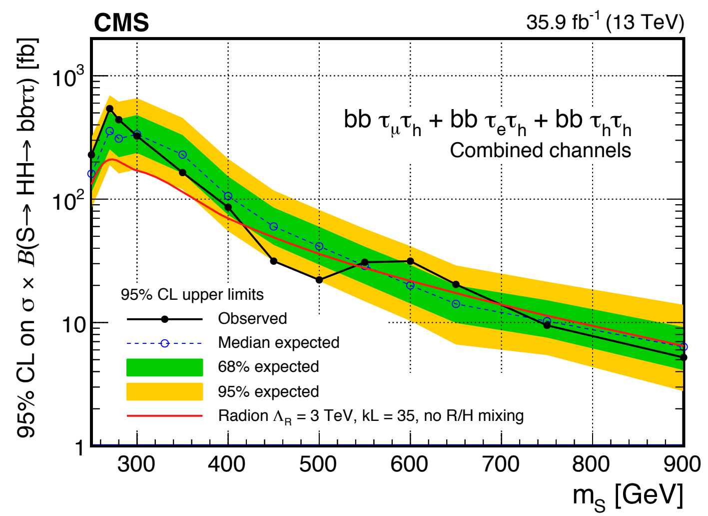
$HH \rightarrow bb\tau\tau$ analysis

Last published results

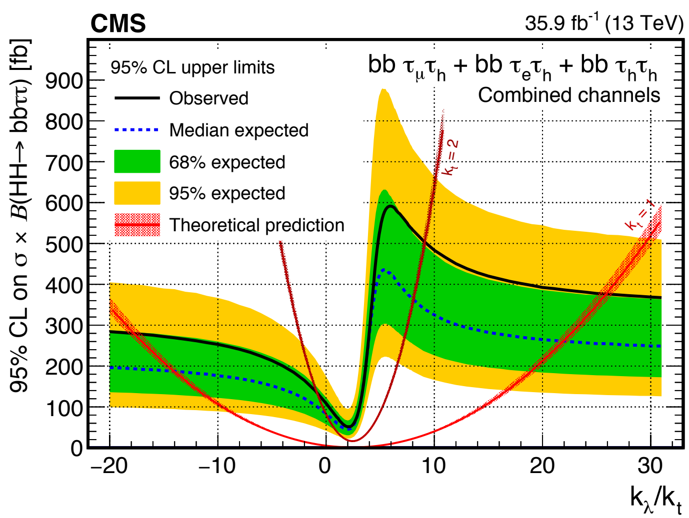
Analysis **HIG-17-002**: published in PLB (doi:[10.1016/j.physletb.2018.01.001](https://doi.org/10.1016/j.physletb.2018.01.001))

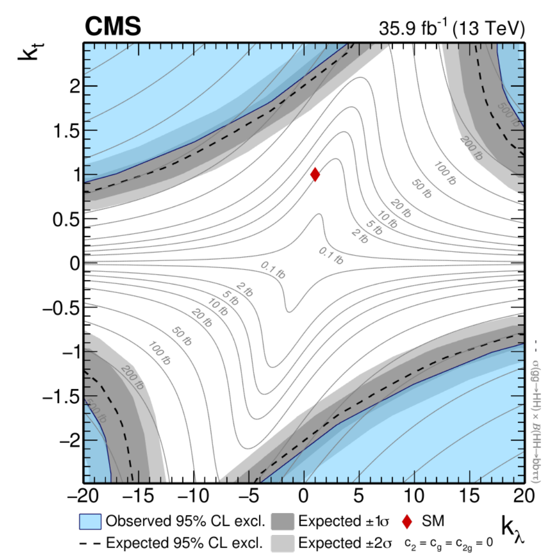
- No evidence for a signal is observed
- Non resonant search: set an obs (exp) 95% CL upper limit on the $\sigma(HH) \approx 30\ (25) \times \sigma_{SM}(HH)$

RESONANT



NON – RESONANT





Cross-section of the main backgrounds

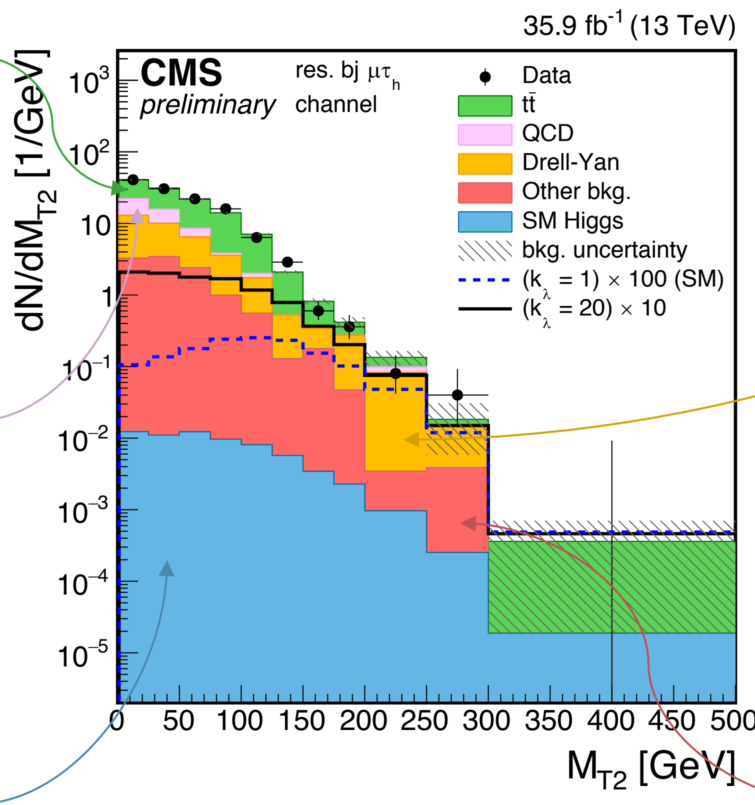
Process	Cross-section $\sqrt{s} = 13$ TeV [pb]
QCD multijets	$O(2 \cdot 10^5)$
$W \to l\nu_l + jets$	$O(6 \cdot 10^4)$
$Z_0/\gamma^* \to ll + jets$ (DY)	5765
$t\bar{t} + jets$	832
VV + jets	180
Single top	71
SM Higgs (ZH)	0.46

Overview of the backgrounds

$t\bar{t}$
Shape and normalization from MC

QCD
Shape and normalization from data

SM Higgs
Shape and normalization from MC



Drell-Yan
Shape from MC, normalization from data

Other backgrounds
Shape and normalization from MC

Summary of HIG-17-002 analysis

- This analysis covers three final states: $\tau_e\tau_h$, $\tau_\mu\tau_h$, $\tau_h\tau_h$ $BR(\approx 88\%)$
- Analysis flow:
  - **$\mathbf{H \rightarrow \tau\tau}$ candidate:**
    - **$\mathbf{H \rightarrow \tau\tau}$** baseline selection with few modifications tuned for bbττ final state
  - **$\mathbf{H \rightarrow}$ bb candidate:**
    - select two jets with the highest CSVv2 score in the event
  - **Events categorization:**
    - Splits the events in 3 categories: resolved 1 btag & 2 btag, and boosted
  - **$\boldsymbol{HH}$ tag:**
    - elliptical mass cut, based on m($\tau\tau$) and m(bb) resolution
    - BDT discriminant against $t\bar{t}$ background in $e\tau_h$ and $\mu\tau_h$ channels
  - **Limit extraction performed on:**
    - HH mass after a kinematic fit (resonant)
    - MT2 (non-resonant)
- Main backgrounds modelling:
  - $t\bar{t}$: using MC simulation
  - QCD: data driven ABCD method
  - DY+Jets: shape from MC simulation, normalization from Z → μμ sideband data sample

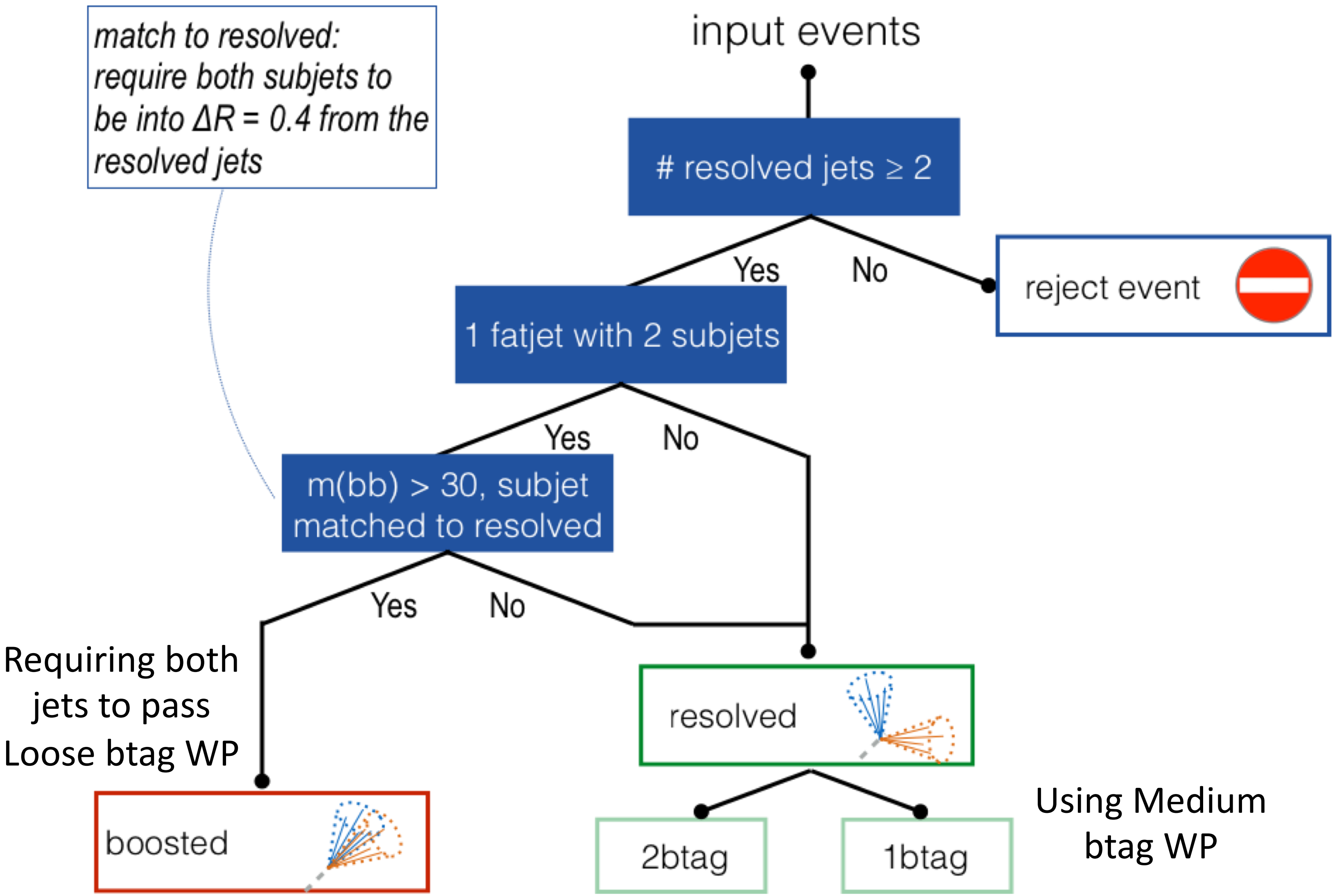
Baseline Selection

- Electrons
  - τ_e: $p_T > 27\ GeV\ \wedge\ \eta < 2.1$
  - MVA ID: 80% WP - signal, 90% WP - veto
  - PF relative $\Delta\beta$ isolation < 0.1
- Muons
  - τ_μ: $p_T > 23\ GeV\ \wedge\ \eta < 2.1$
  - ID: Tight WP - signal, Loose WP - veto
  - PF relative isolation < 0.15
- Taus
  - τ_h: $p_T > 20\ (45)\ GeV\ \wedge\ \eta < 2.1$
  - MVA isolation: Medium WP (*after pair selection*)
- Tau pair for $H \rightarrow \tau\tau$ candidate
  - ΔR between τ candidates > 0.1
  - Opposite sign (*after pair selection*)
  - $m_{\tau\tau}$ reconstructed using SVfit algorithm
- AK4 jets
  - $p_T > 20\ GeV\ \wedge\ \eta < 2.4$
  - PF loose ID
  - CSVv2 is used for b jet identification
  - ΔR with signal objects > 0.5
- AK8 jets
  - Soft drop mass $> 30\ GeV$
- PF MET

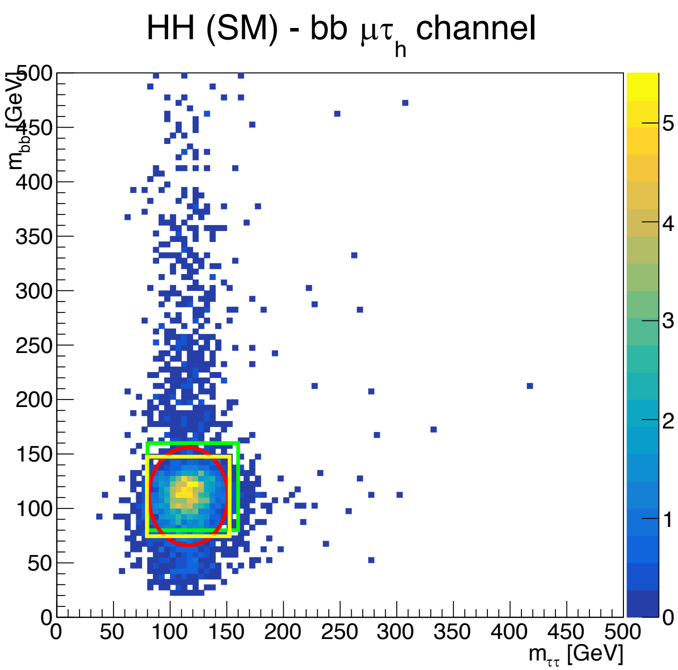
MC/Data correction factors are applied:

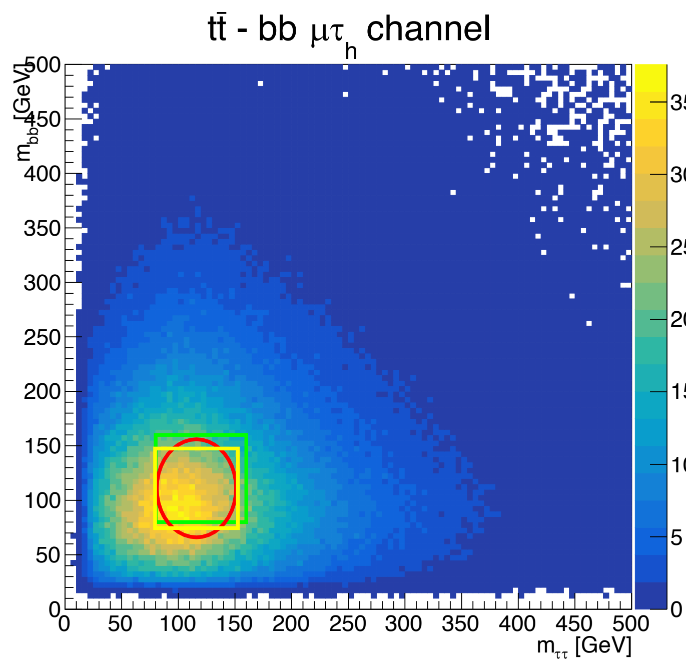
- ***e***: the ID/iso SFs are provided by HTauTau group
- ***μ***: the ID/iso SFs are provided the muon POG
- Weights for **btag efficiency** SF are provided by b POG

Event Categorization



Mass cut for HH candidate selection



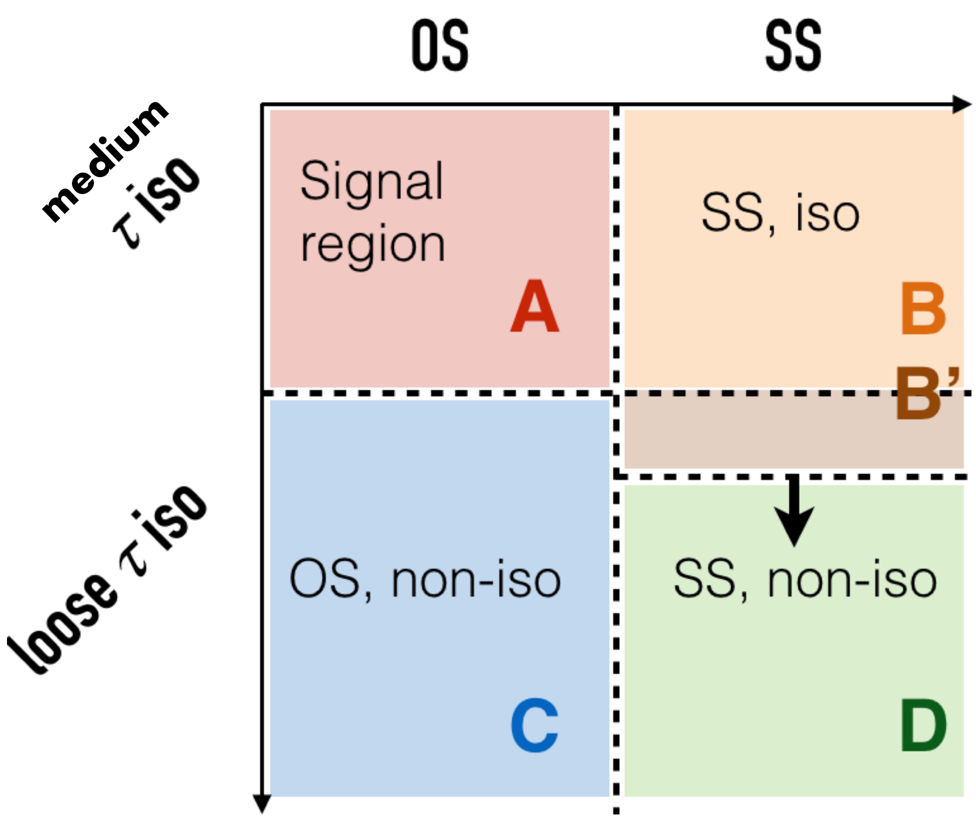


Mass window is chosen accordingly to the resolution and mean value of $m_{\tau\tau}$ and m_{bb} distributions:

$$\left(\frac{m_{\tau\tau}\ [GeV] - 116}{35}\right)^2 + \left(\frac{m_{bb}\ [GeV] - 111}{45}\right)^2 < 1$$

There will be an exercise to optimize these parameters!

QCD estimation

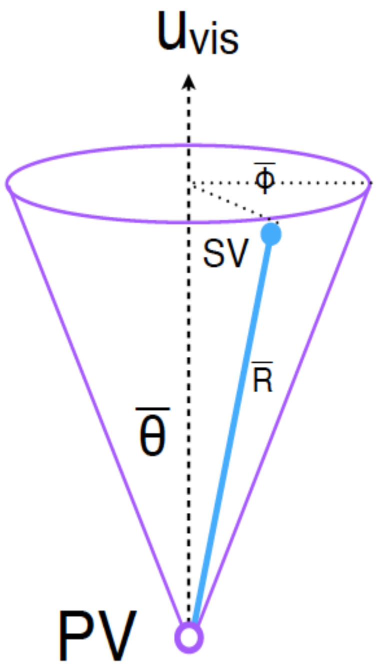


- QCD is estimated entirely from data and the method used is based on the assumption that the contribution from the taus with the same sign and with the opposite sign for QCD is roughly the same. This is not completely true so an opposite sign / same sign extrapolation factor is calculated using data inverting the tau isolation
- Yield in region A is calculated from this formula: $B \cdot C/D$
  - From Data subtracting bkg contribution from MC

There will be an exercise to estimate the QCD contribution!

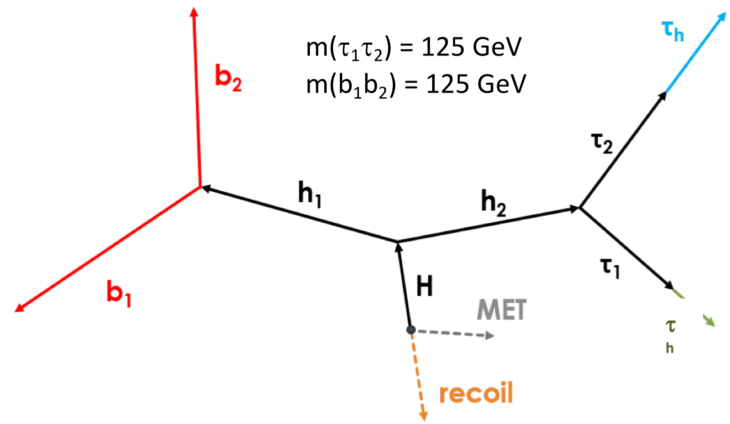
Di-tau invariant mass reconstruction (SVFit algorithm)

- SVfit is a likelihood based algorithm for the reconstruction of h boson decaying to τ leptons.
- The kinematics of τ decays can be parameterized by following variables:
  - θ – the angle between the boost direction of the τ lepton and the momentum of the visible decay products in the rest frame of the τ.
  - ϕ – the azimuthal angle of the τ in the CMS detector frame.
  - $m_{\nu\nu}$ – invariant mass of the invisible momentum system for leptonic τ decays
- The kinematics of the τ pair decays depends upon 4-6 parameters, which are constrained only by 2 observables from MET
- Using Dynamical Likelihood Methods, SVfit reconstruct kinematic quantities on an event-by-event basis.

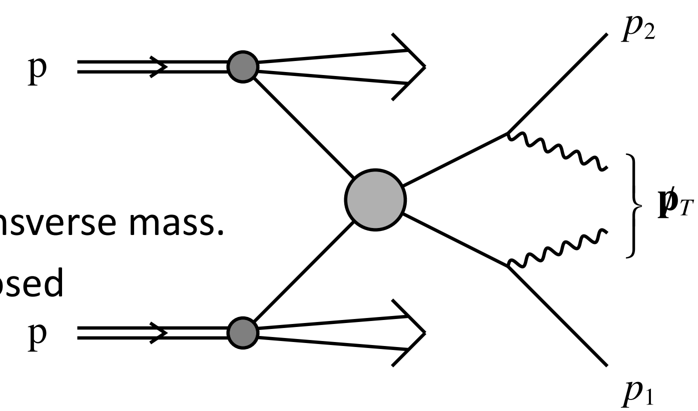


Variables for limit extraction

- In the resonant analysis, **fitted invariant mass of HH-candidate** is used for the signal extraction, M_H^{KinFit}
  - Kinematic constraints of $m_H = 125\ GeV$ is applied for $H_{\tau\tau}$ and H_{bb} candidates.
  - Collinear approximation is considered for τ decays.
- The fit improves the mass resolution for signal events, while for the background the M_H^{KinFit} distribution is still wide and quite unchanged.



- In the non-resonant analysis, the signal is extracted using **stransverse mass**.
- The **stransverse mass, m_{T2}**, is a generalized version of the transverse mass.
  - it is originally designed for SUSY searches, and later proposed for $HH \to bb\tau\tau$ (doi:10.1016/j.physletb.2013.12.011)



- m_{T2} is defined as

$$m_{T2} \equiv \min_{\boldsymbol{p}_{T1}+\boldsymbol{p}_{T2}=\boldsymbol{p}_T^{\tau\tau}} \{\max[m_T(m_{b1}, \boldsymbol{p}_T^{b1}, m_{vis}^{\tau1}, \boldsymbol{p}_{T1}), m_T(m_{b2}, \boldsymbol{p}_T^{b2}, m_{vis}^{\tau2}, \boldsymbol{p}_{T2})]\}$$

- m_{T2} provides bigger discrimination comparing to $m(HH)$, because, by construction, it is bounded by m_{top} for $t\bar{t}$ background, but not for the signal

Analysis framework

Framework

- Git repositories: https://github.com/hh-italian-group/
- We will use **"cmsdas_2019" branch**
- Contributing groups:
  - *Currently active:* Pisa, Kolkata, Siena
- All code, except the tuple production step, is CMSSW-independent and can be run on SL, OSX or Ubuntu.
- Languages: C++ (>90%), Python
- Build system: CMake

Framework

Three packages

- **AnalysisTools**: https://github.com/hh-italian-group/AnalysisTools/tree/cmsdas_2019
  - General analysis tools
  - Various classes and functions that extend ROOT functionalities

- **h-tautau**: https://github.com/hh-italian-group/h-tautau/tree/cmsdas_2019
  - Definition of the EventTuple (using SmartTree class, which is a wrapper around TTree)
  - Definition of the classes that represent reconstructed e/mu/tau/Higgs candidates
  - Code for the weights: PU, bTag, lepton scale factor
  - NTuple Producer: BaseProducer that inherit from EDMAnalyzer
    - Three different producers for muTau, eTau and tauTau

- **hh-bbtautau**: https://github.com/hh-italian-group/hh-bbtautau/tree/cmsdas_2019
  - Base Analysis Class, that contains the common part for the three channels
  - Event Categorization
  - Data-driven background estimation
  - Code that produces stacked plots and root file with template shapes for the limit extraction

Analysis Flow

MiniAOD → Full tuple production

baseline selection is applied

total tuples size $\approx 250\ GB$

↓

skimming

signal + sideband regions

total tuples size $\approx 5..20\ GB$

↓

anaTuples production

Final ntuples in signal and sideband regions

total tuples size $\approx 2.5\ GB$

1. Full ntuple production

- The first step of this analysis is production of ntuples from miniAOD.
- The code is structured in a modular way to select candidates and the objects which are useful for the final selection.
- BaseTupleProducer (https://github.com/hh-italian-group/h-tautau/blob/cmsdas_2019/Production/src/BaseTupleProducer.cc) is the class for producing full ntuples. For each channel there is its own producer.
- It is part of cmssw so it should be run inside the cmssw environment.
- The configuration file that should be run to have full ntuples for all channels is Production.py (https://github.com/hh-italian-group/h-tautau/blob/cmsdas_2019/Production/python/Production.py)

2. Skimming = eventTuples

- In order to produce the final ntuples, called "anaTuples", that we will be used in the following exercises, there are two steps that have to be run.
- The first step is called TupleSkimmer (https://github.com/hh-italian-group/hh-bbtautau/blob/cmsdas_2019/Instruments/source/TupleSkimmer.cxx) , where we skim the FullTuples, applying a preliminary selection, and we weight events using different weights and the corresponding cross-section.
- For this exercise, ntuples are skimmed requiring the central energy scale, no elliptical mass cut and a cut on VLoose isolation working point for the tau identification

3. Final ntuples = anaTuples

- The final step is to produce "anaTuples" using the Analyzer classes.
- The BaseEventAnalyzer (https://github.com/hh-italian-group/hh-bbtautau/blob/master/Analysis/include/BaseEventAnalyzer.h)
  is the common class where all cuts are defined and all samples are processed. In the analyser for each channel we apply the trigger match and we identify the Event Region for each event.
- The definition of our anaTuples which are SmartTree is here https://github.com/hh-italian-group/hh-bbtautau/blob/master/Analysis/include/AnaTuple.h
- The next step is ProcessAnaTuple (https://github.com/hh-italian-group/hh-bbtautau/blob/master/Analysis/source/ProcessAnaTuple.cxx), where you can plot the main distribution observing the contribution of each bkg and signal and where data-driven bkg estimations are applied
- The last step is the limit extraction
  - The code base is defined here: https://github.com/cms-hh/HHStatAnalysis
  - It is not covered in this long exercise

Conclusion

- $HH \rightarrow bb\tau\tau$ analysis and framework have been presented
- You will find similar description and run instructions in the twiki for each step
- We have structured the long exercises in 4 sequential exercises that cover different aspects of the analysis:
  - Ntuple production and baseline selection
  - Background composition and its properties
  - Optimisation of the selection in the signal region
  - Machine learning techniques to improve signal sensitivity

Good work!