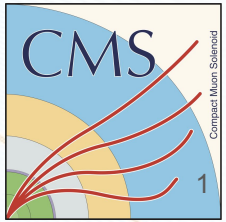
Search for HH→4b using boosted large-radius jets with the CMS detector

Irene Dutta (Caltech/Fermilab*)
On behalf of the CMS collaboration

12th Higgs Hunting workshop

13th September 2022



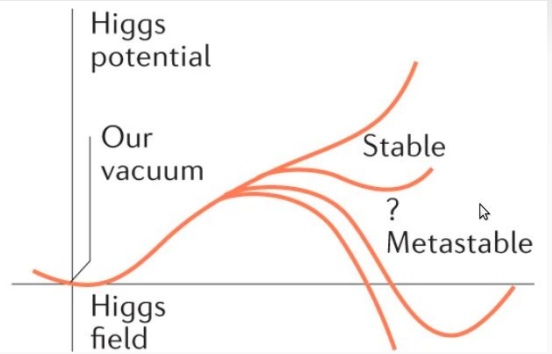
* joined Fermilab as a postdoc yesterday; was a PhD student at Caltech until recently

Higgs self coupling

- Higgs trilinear self coupling is $\lambda_{HHH} = m_h^2/2v$ (v~246 GeV)

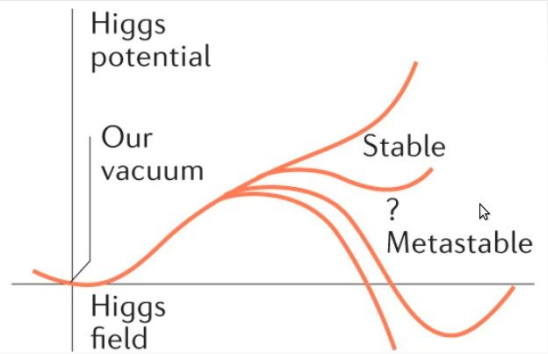
Higgs self coupling

- Higgs trilinear self coupling is $\lambda_{HHH} = m_h^2/2v$ (v~246 GeV)
- Important to study the trilinear coupling
  - probe the structure of the Higgs potential at large scales - metastability of the EW vacuum



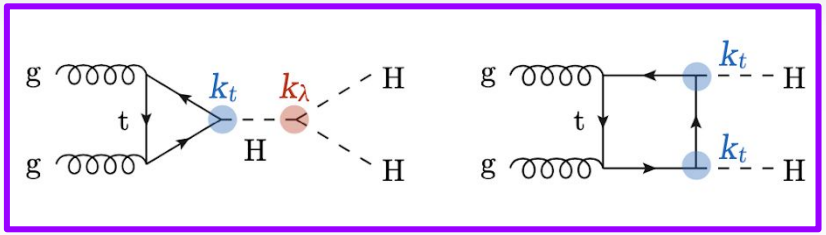
Higgs self coupling

- Higgs trilinear self coupling is $\lambda_{HHH} = m_h^2/2v$ (v~246 GeV)
- Important to study the trilinear coupling
  - probe the structure of the Higgs potential at large scales - metastability of the EW vacuum
- Study HH production to measure λ_{HHH}
  - Two major production modes at LHC - ggF and VBF



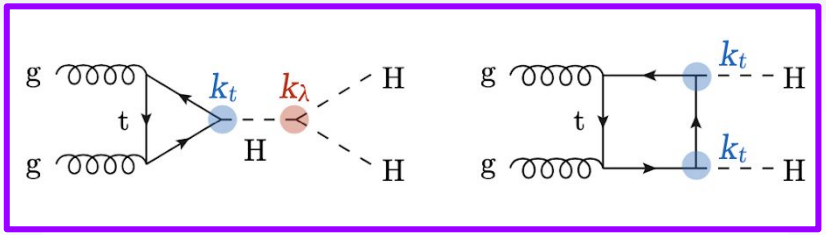
HH production modes –ggF

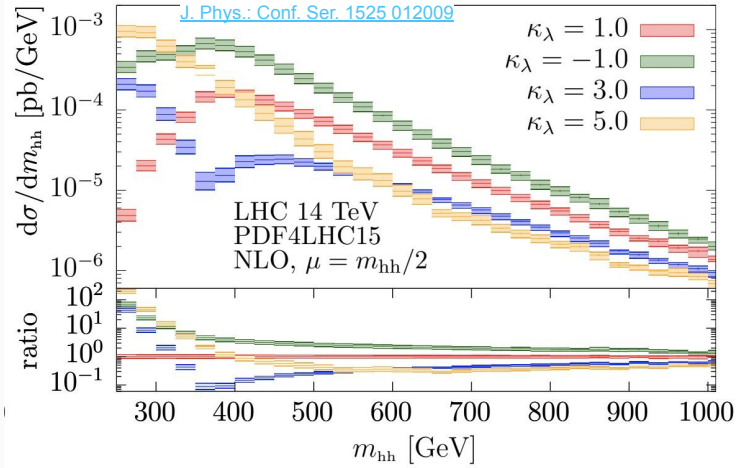
- Two main leading order **ggF** diagrams interfere destructively - 31.05 fb at 13 TeV at NNLO



HH production modes -ggF

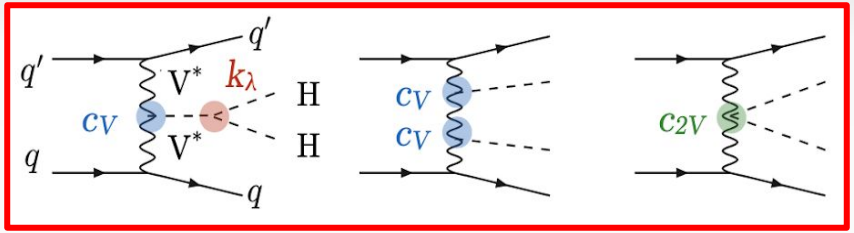
- Two main leading order **ggF** diagrams interfere destructively - 31.05 fb at 13 TeV at NNLO
- Spectrum of m_{HH} is softer for large $|\kappa_\lambda|$ and harder for intermediate $|\kappa\lambda|$
  - boosted ggF signatures sensitive to intermediate $|\kappa_\lambda|$





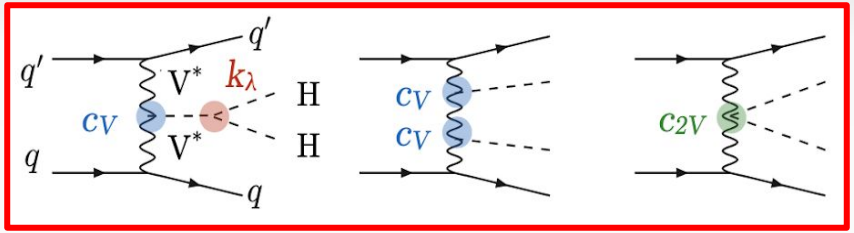
HH production modes –VBF

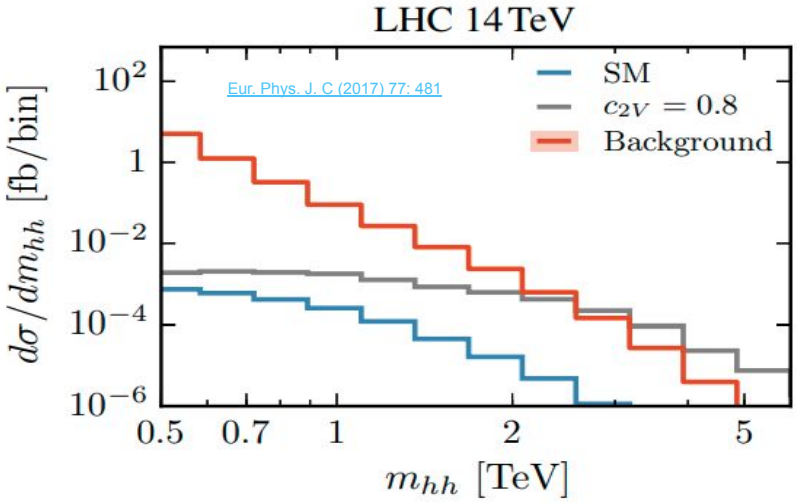
- **VBF** diagrams sensitive to C_V and C_{2V} - 1.73 fb at 13 TeV



HH production modes –VBF

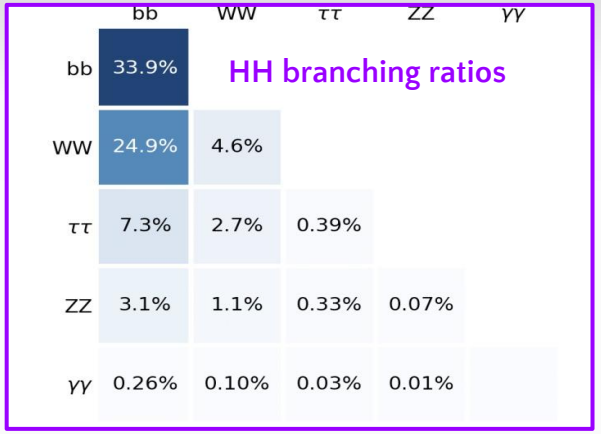
- **VBF** diagrams sensitive to C_V and C_{2V} - 1.73 fb at 13 TeV
- Smaller κ_{2V} leads to larger cross section, harder m_{HH} spectrum, and boosted VBF signatures





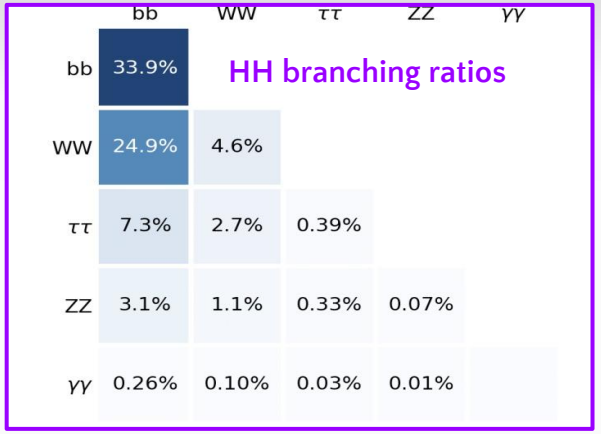
HH→bbbb at LHC

- HH→(bb)(bb) has the highest BR (33.9 %)
  - Large QCD background and poor decay channel resolution



HH→bbbb at LHC

- HH→(bb)(bb) has the highest BR (33.9 %)
  - Large QCD background and poor decay channel resolution
  - Considered *inaccessible* until a few years ago

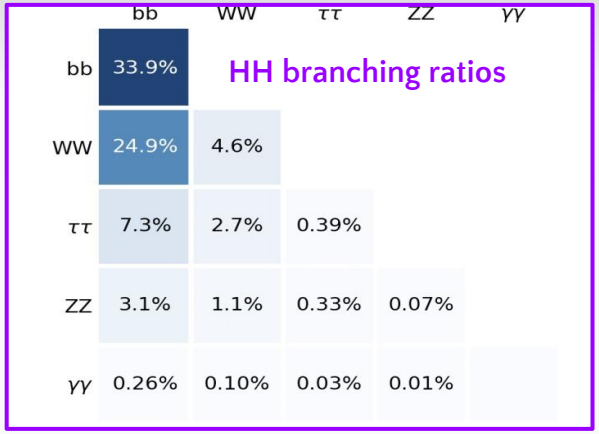


HH→bbbb at LHC

- HH→(bb)(bb) has the highest BR (33.9 %)
  - Large QCD background and poor decay channel resolution
  - Considered *inaccessible* until a few years ago

How can we make HH→4b more powerful ?

- Explore the regime where both Higgs bosons are boosted (p_T > 300 GeV)

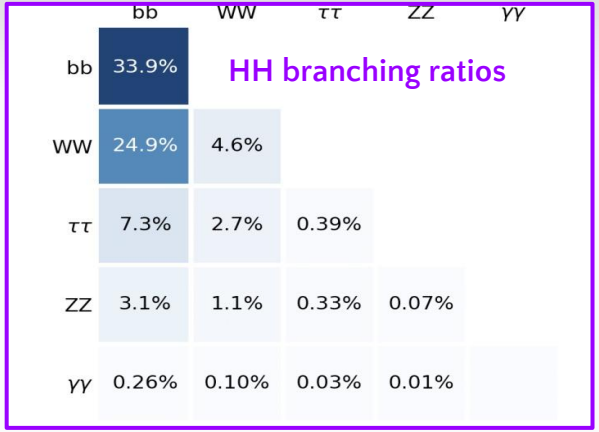


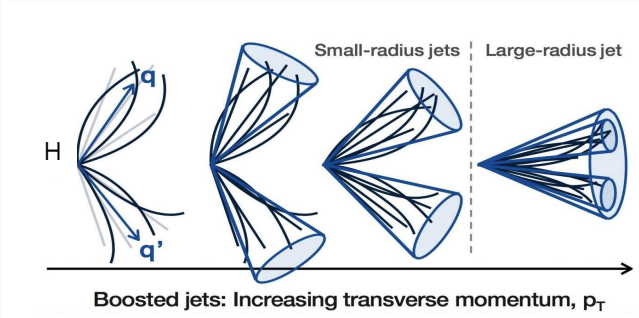
HH→bbbb at LHC

- HH→(bb)(bb) has the highest BR (33.9 %)
  - Large QCD background and poor decay channel resolution
  - Considered *inaccessible* until a few years ago

How can we make HH→4b more powerful ?

- Explore the regime where both Higgs bosons are boosted (p_T > 300 GeV)
  - The two b quark jets from the Higgs decay merge into one fat-jet



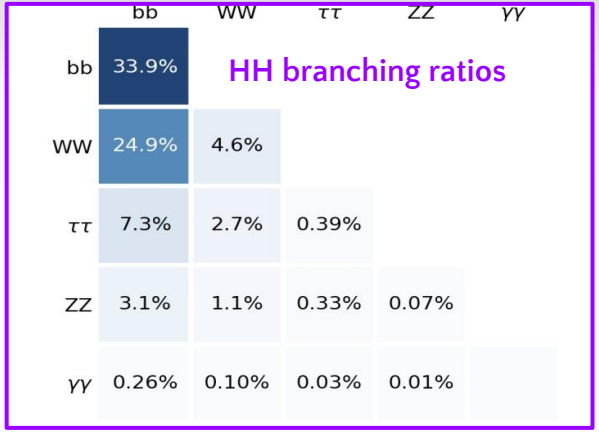


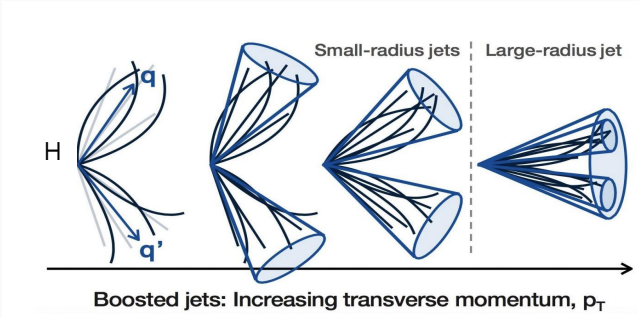
HH→bbbb at LHC

- HH→(bb)(bb) has the highest BR (33.9 %)
  - Large QCD background and poor decay channel resolution
  - Considered *inaccessible* until a few years ago

How can we make HH→4b more powerful ?

- Explore the regime where both Higgs bosons are boosted (p_T > 300 GeV)
  - The two b quark jets from the Higgs decay merge into one fat-jet
- Exploit fat-jet sub-structures for better S/B
  - Use machine learning techniques to identify b quarks inside fat jets





Boosted H→(bb) jet identification

ParticleNet Jet Tagger

Graph Neural Network (GNN) based classifier for large radius jets (Phys. Rev. D 101, 056019)

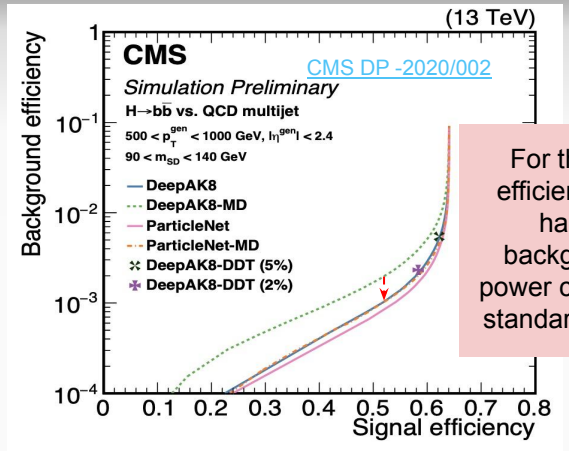
- low-level jet information (PF candidates, secondary vertices from b-quark decays etc.) as inputs

Boosted H→(bb) jet identification

ParticleNet Jet Tagger

Graph Neural Network (GNN) based classifier for large radius jets (Phys. Rev. D 101, 056019)

- low-level jet information (PF candidates, secondary vertices from b-quark decays etc.) as inputs
- output scores: X→bb, X→cc, X→light quarks, QCD



For the same signal efficiency, Particle Net has a 2x better background rejection power compared to other standard Jet taggers

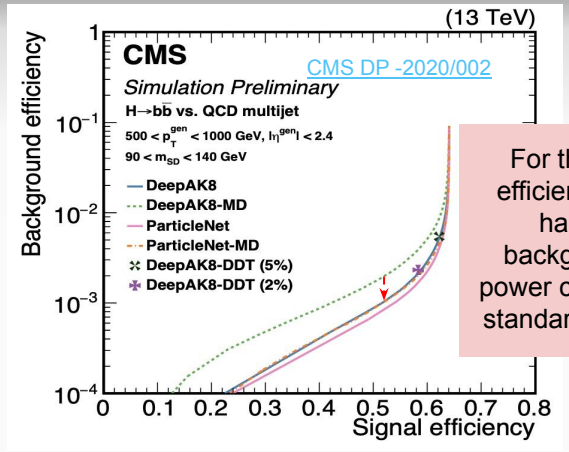
Boosted H→(bb) jet identification

ParticleNet Jet Tagger

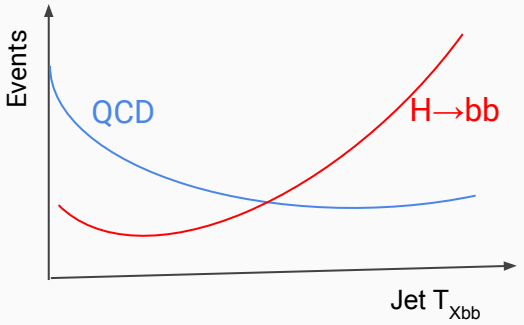
Graph Neural Network (GNN) based classifier for large radius jets (Phys. Rev. D 101, 056019)

- low-level jet information (PF candidates, secondary vertices from b-quark decays etc.) as inputs
- output scores: X→bb, X→cc, X→light quarks, QCD
  - For our purposes, discriminate X→bb vs QCD jets:

$$T_{Xbb} = \frac{P_{Xbb}}{P_{Xbb} + P_{QCD}}$$



For the same signal efficiency, Particle Net has a 2x better background rejection power compared to other standard Jet taggers



Boosted HH → 4b search from CMS

CMS-B2G-22-003 ; CERN-EP-2022-090

Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks

arXiv:2205.06667 (accepted by PRL)

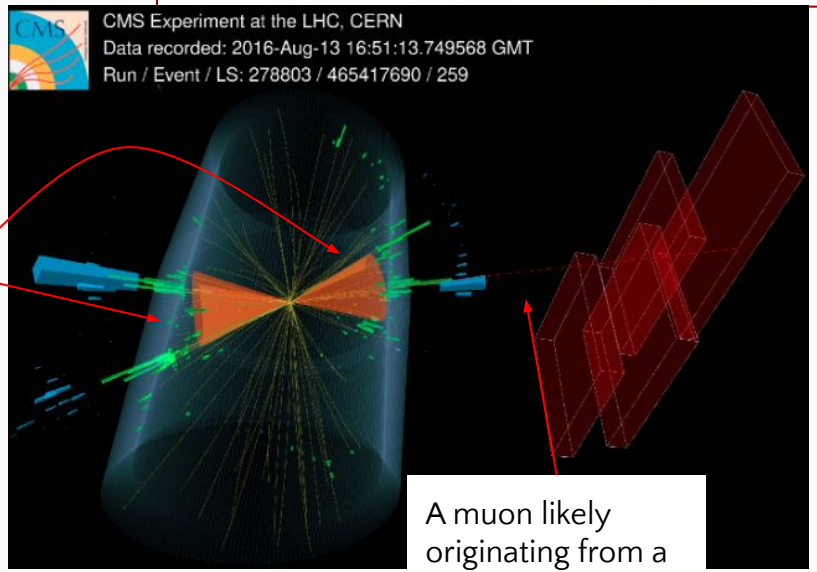
CMS Collaboration

13 May 2022

Accepted for publication in Phys. Rev. Lett.

Abstract: A search for nonresonant Higgs boson (H) pair production via gluon and vector boson (V) fusion is performed in the four-bottom-quark final state, using proton-proton collision data at 13 TeV corresponding to 138 fb^{-1} collected by the CMS experiment at the LHC. The analysis targets Lorentz-boosted H pairs identified using a graph neural network. It constrains the strengths relative to the standard model of the H self-coupling and the quartic VVHH couplings, κ_{2V}, excluding $\kappa_{2V} = 0$ for the first time, with a significance of 6.3 standard deviations when other H couplings are fixed to their standard model values.

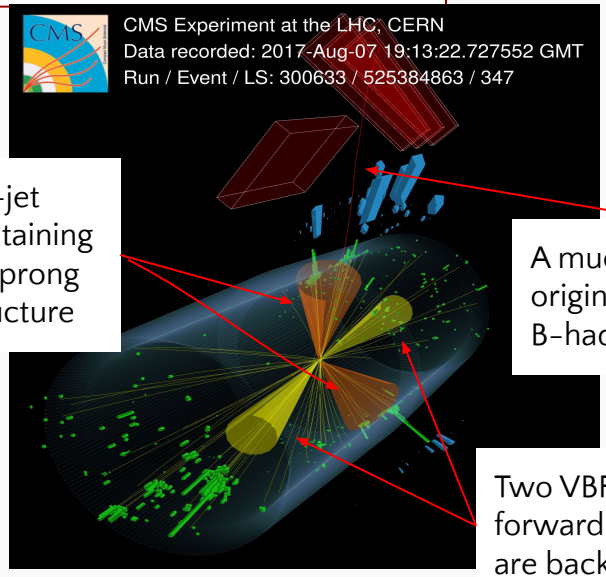
Links: e-print arXiv:2205.06667 [hep-ex] (PDF) ; CDS record ; inSPIRE record ; HepData record ; CADI line (restricted) ;



Fat-jet containing a 2 prong structure

A muon likely originating from a B-hadron decay

A ggF like event ..



Fat-jet containing a 2 prong structure

A muon likely originating from a B-hadron decay

Two VBF like forward jets that are back-to-back

A VBF like event ..

Analysis strategy

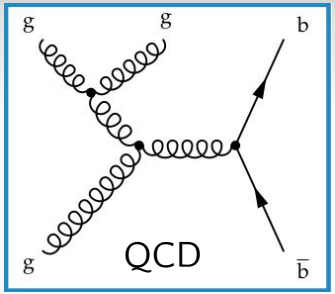
- Identify two high p_T Higgs candidate large radius jets using
  - Particle Net T_{Xbb} score
  - Jet mass (m_j)

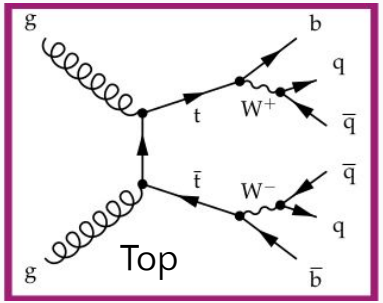
Analysis strategy

- Identify two high p_T Higgs candidate large radius jets using
  - Particle Net T_{Xbb} score
  - Jet mass (m_j)
- Based on additional kinematic features, separate into ggF or VBF category

Analysis strategy

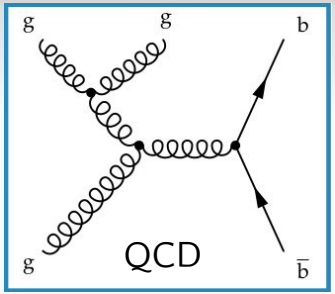
- Identify two high p_T Higgs candidate large radius jets using
  - Particle Net T_{Xbb} score
  - Jet mass (m_j)
- Based on additional kinematic features, separate into ggF or VBF category
- Use signal depleted control regions to perform data-driven estimates of backgrounds
  - QCD
  - Top

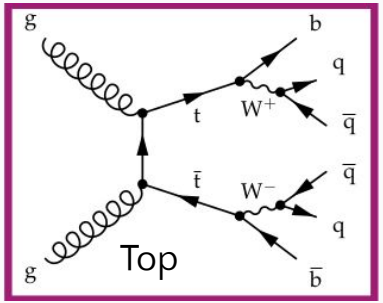


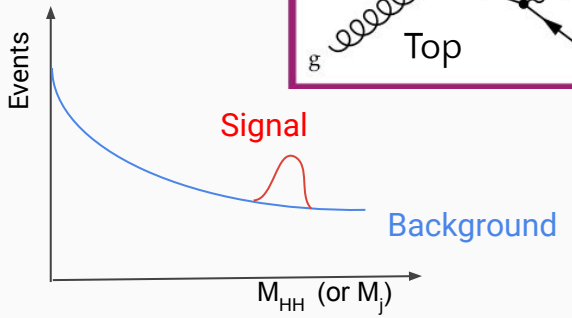


Analysis strategy

- Identify two high p_T Higgs candidate large radius jets using
  - Particle Net T_{Xbb} score
  - Jet mass (m_j)
- Based on additional kinematic features, separate into ggF or VBF category
- Use signal depleted control regions to perform data-driven estimates of backgrounds
  - QCD
  - Top
- Perform final signal extraction fit to m_{HH} or Jet m_j

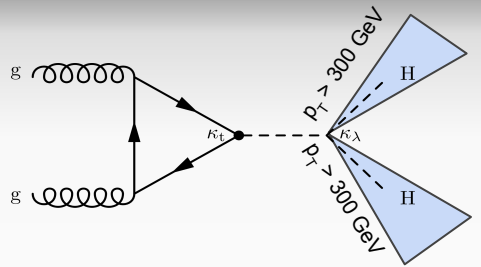






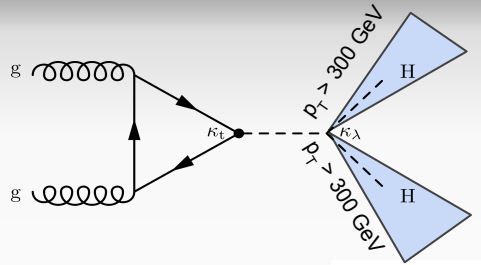
ggF strategy

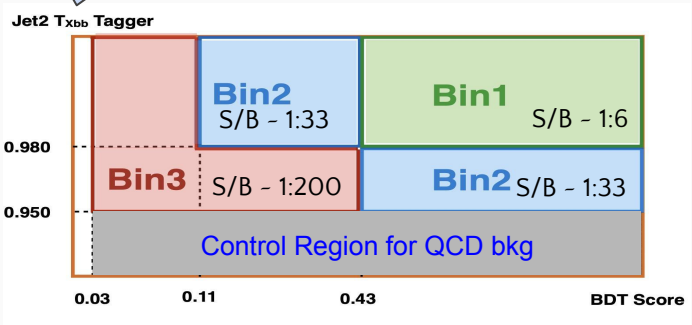
- Boosted Decision Tree (BDT) to discriminate HH signal events from QCD and top



ggF strategy

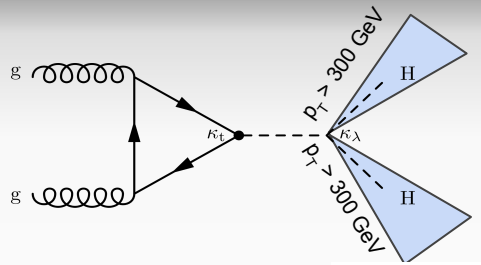
- Boosted Decision Tree (BDT) to discriminate HH signal events from QCD and top
- Three event categories optimized based on BDT score and Jet 2 T_{Xbb} score

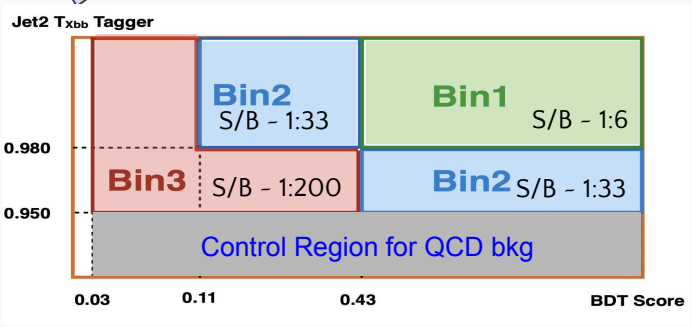




ggF strategy

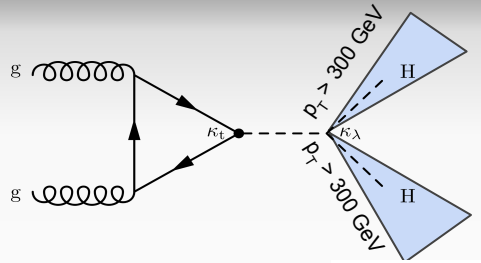
- Boosted Decision Tree (BDT) to discriminate HH signal events from QCD and top
- Three event categories optimized based on BDT score and Jet 2 T_{Xbb} score
- Observable for HH signal extraction : Jet 2 Mass

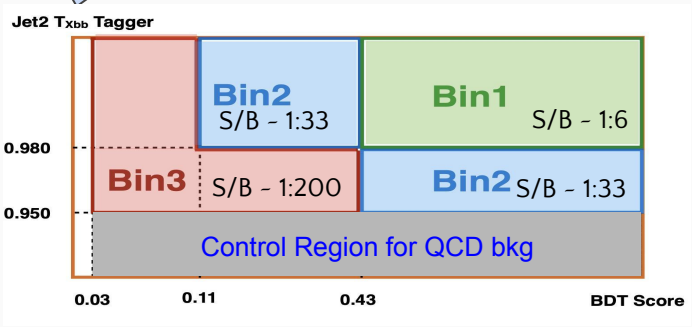




ggF strategy

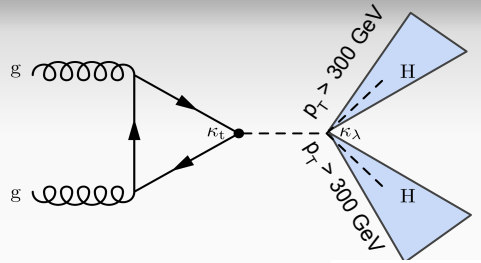
- Boosted Decision Tree (BDT) to discriminate HH signal events from QCD and top
- Three event categories optimized based on BDT score and Jet 2 T_{Xbb} score
- Observable for HH signal extraction : Jet 2 Mass
- Background estimation
  - **Top** : data driven correction factors applied to simulation

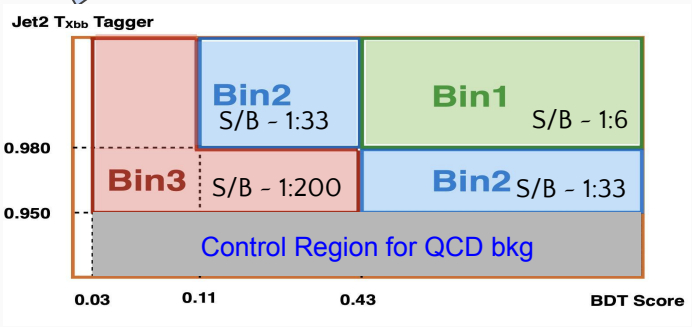




ggF strategy

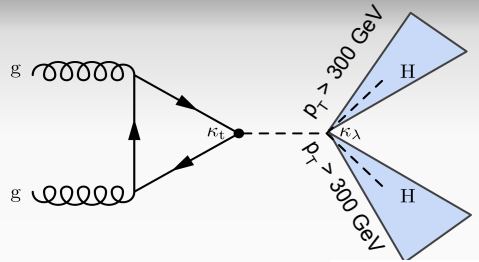
- Boosted Decision Tree (BDT) to discriminate HH signal events from QCD and top
- Three event categories optimized based on BDT score and Jet 2 T_{Xbb} score
- Observable for HH signal extraction : Jet 2 Mass
- Background estimation
  - **Top** : data driven correction factors applied to simulation
  - **QCD**: parametric fit to data in control region

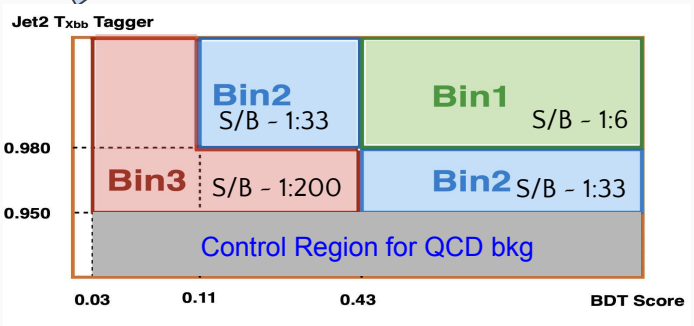


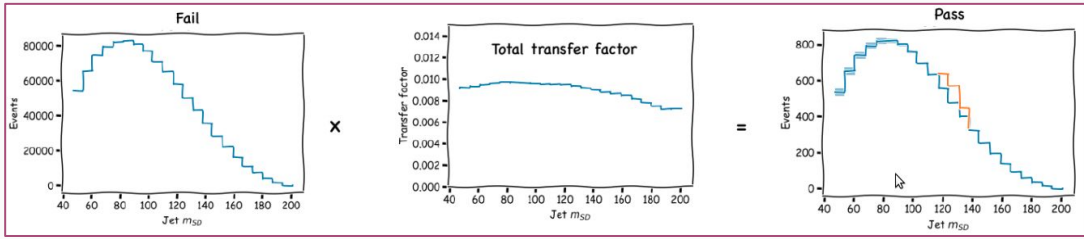


ggF strategy

- Boosted Decision Tree (BDT) to discriminate HH signal events from QCD and top
- Three event categories optimized based on BDT score and Jet 2 T_{Xbb} score
- Observable for HH signal extraction : Jet 2 Mass
- Background estimation
  - **Top** : data driven correction factors applied to simulation
  - **QCD**: parametric fit to data in control region
    - Signal region QCD shape = QCD fail shape ⊗TF (j_2 Mass)

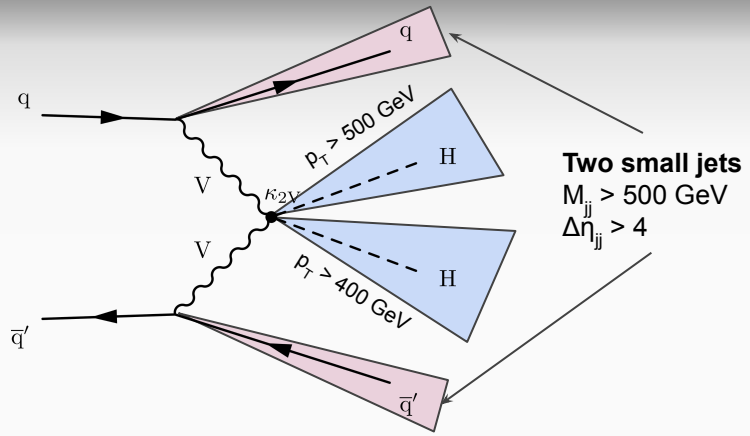






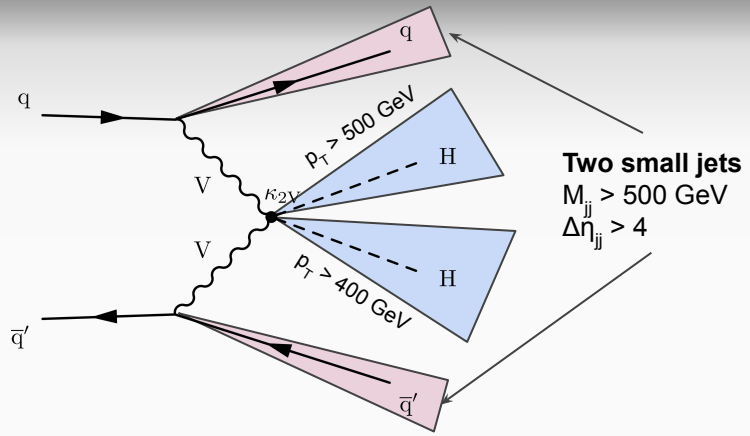
VBF strategy

- Three exclusive signal regions based on both jets satisfying:
  - **High Purity (HP)**: $T_{Xbb} > 0.98$
  - **Medium Purity (MP):** $0.94 < T_{Xbb} < 0.98$
  - **Low Purity (LP)**: $0.90 < T_{Xbb} < 0.94$



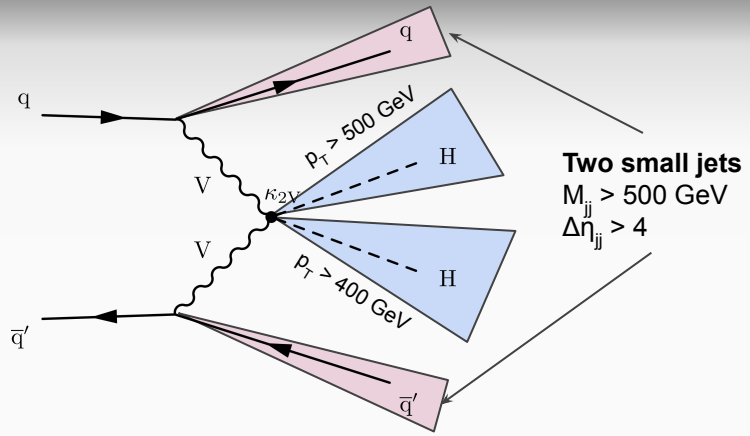
VBF strategy

- Three exclusive signal regions based on both jets satisfying:
  - **High Purity (HP)**: $T_{Xbb} > 0.98$
  - **Medium Purity (MP):** $0.94 < T_{Xbb} < 0.98$
  - **Low Purity (LP)**: $0.90 < T_{Xbb} < 0.94$
- Background estimation
  - **Top**: estimated from simulation with data driven correction factors

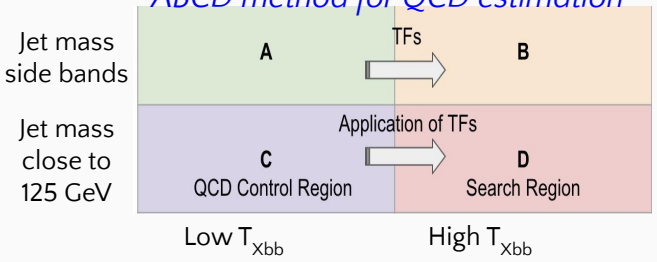


VBF strategy

- Three exclusive signal regions based on both jets satisfying:
  - **High Purity (HP)**: $T_{Xbb} > 0.98$
  - **Medium Purity (MP):** $0.94 < T_{Xbb} < 0.98$
  - **Low Purity (LP)**: $0.90 < T_{Xbb} < 0.94$
- Background estimation
  - **Top**: estimated from simulation with data driven correction factors
  - **QCD** - ABCD method



ABCD method for QCD estimation

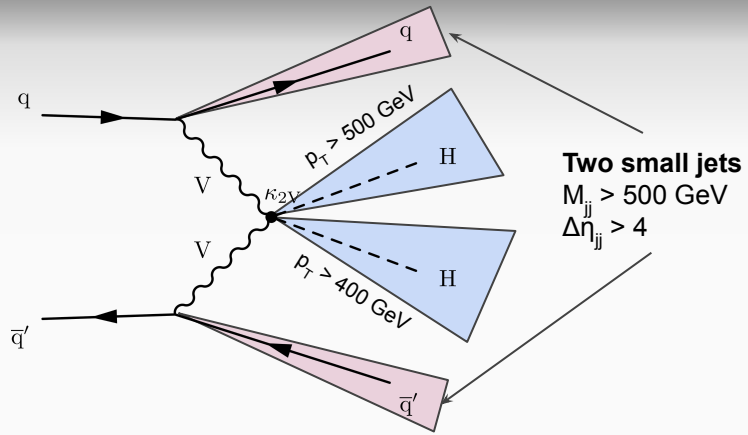


$TF = N_B / N_A$

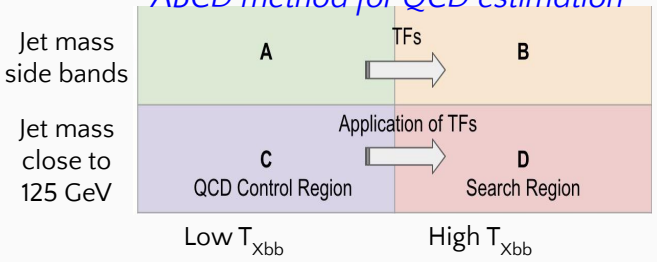
$N_D = TF * N_c$

VBF strategy

- Three exclusive signal regions based on both jets satisfying:
  - **High Purity (HP)**: $T_{Xbb} > 0.98$
  - **Medium Purity (MP):** $0.94 < T_{Xbb} < 0.98$
  - **Low Purity (LP)**: $0.90 < T_{Xbb} < 0.94$
- Background estimation
  - **Top**: estimated from simulation with data driven correction factors
  - **QCD** - ABCD method
- Perform final fit to m_{HH} in different WP regions



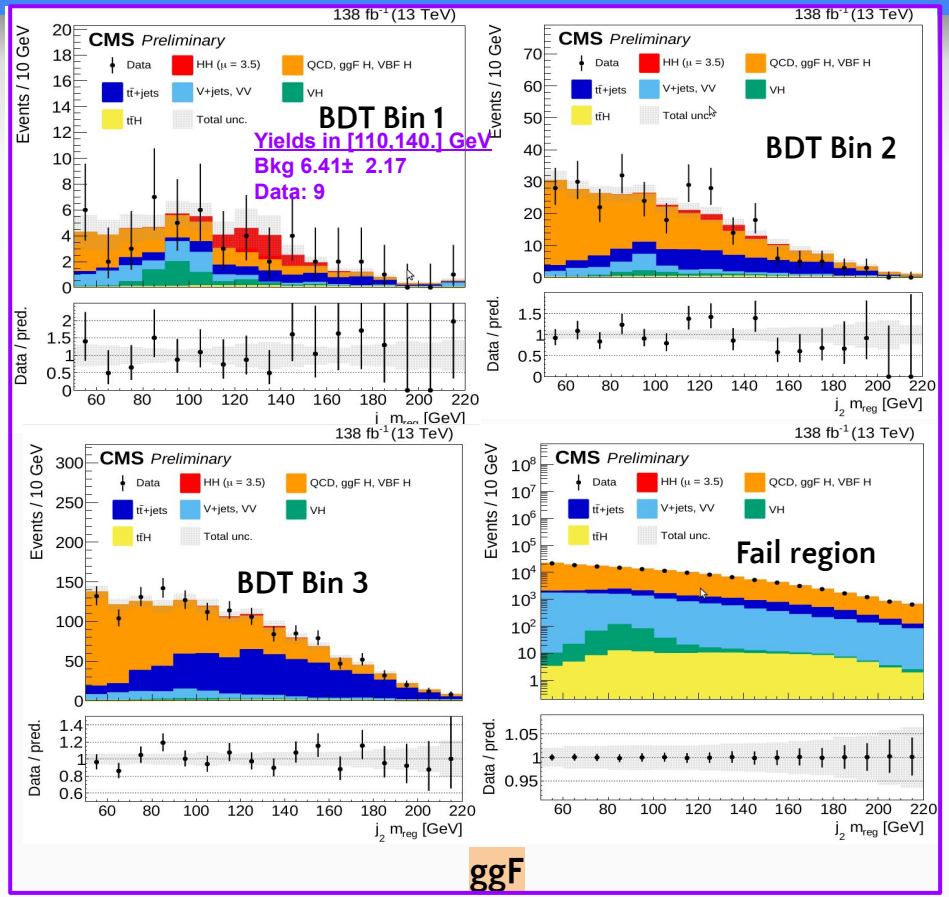
ABCD method for QCD estimation



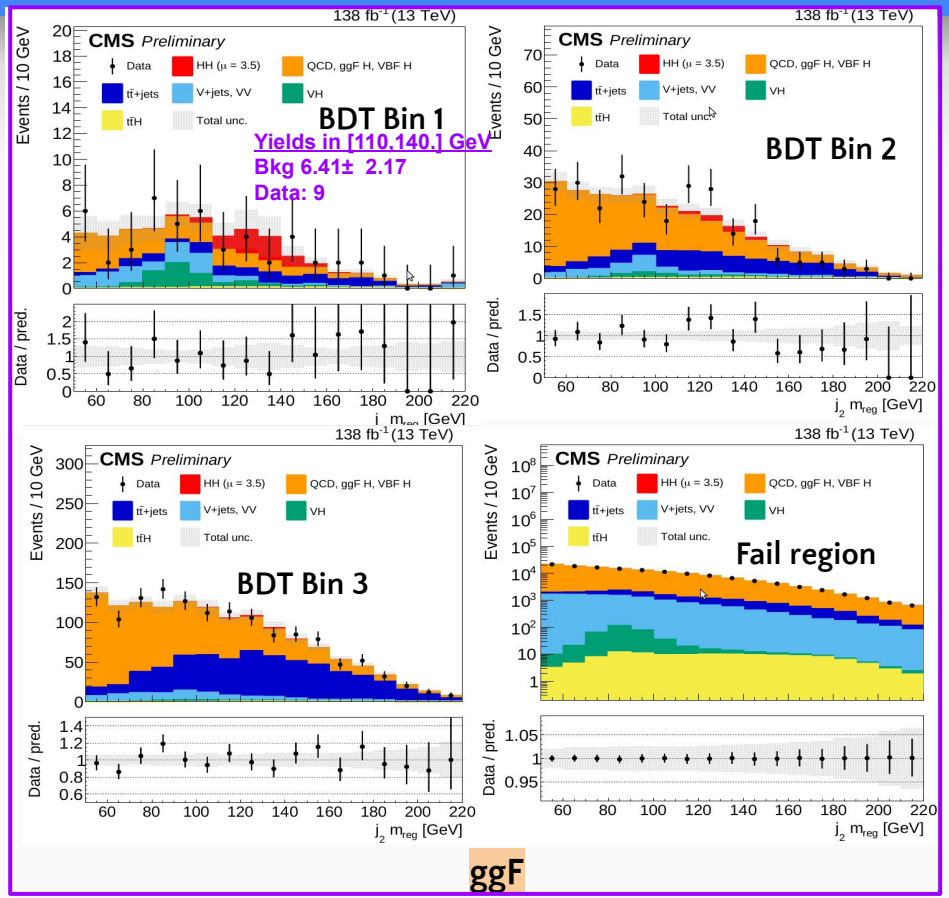
$TF = N_B / N_A$

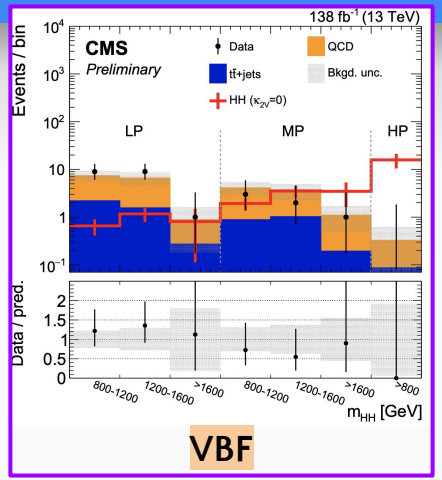
$N_D = TF * N_c$

Results

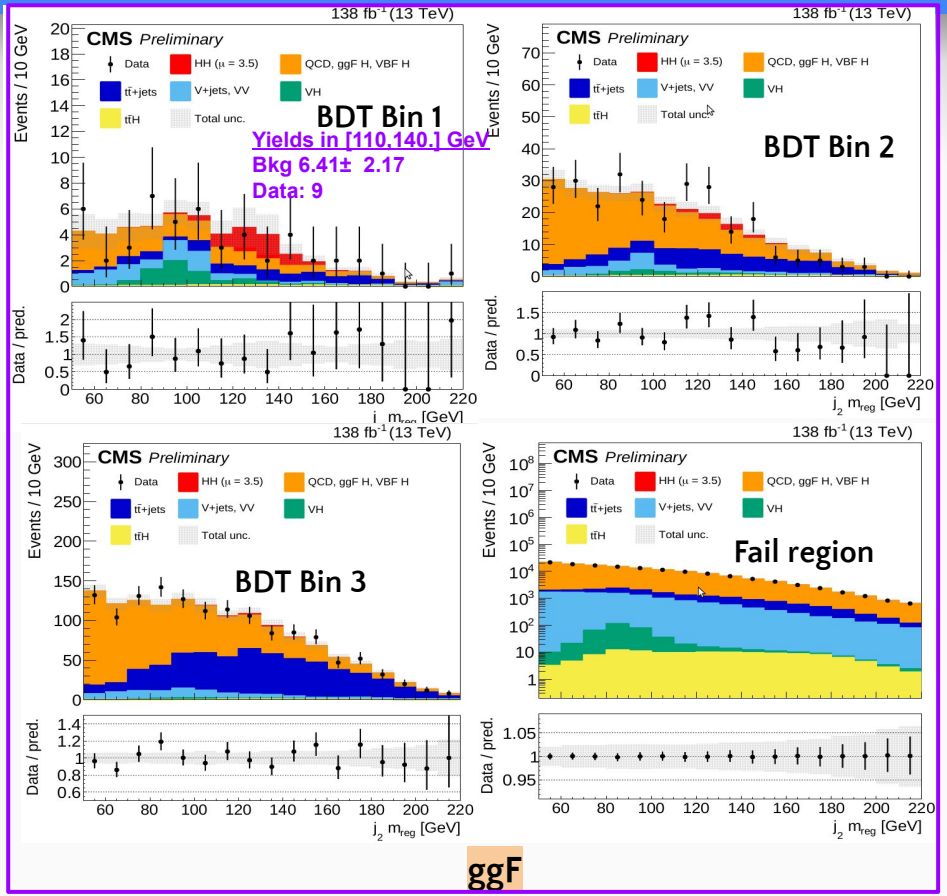

138 fb-1 (13 TeV)
CMS Preliminary
Data
HH (μ = 3.5)
QCD, ggF H, VBF H
tt̄+jets
V+jets, VV
VH
tt̄H
Total unc.
BDT Bin 1
Yields in [110,140.] GeV
Bkg 6.41± 2.17
Data: 9
BDT Bin 2
BDT Bin 3
Fail region
Events / 10 GeV
Data / pred.
j2 mreg [GeV]
ggF

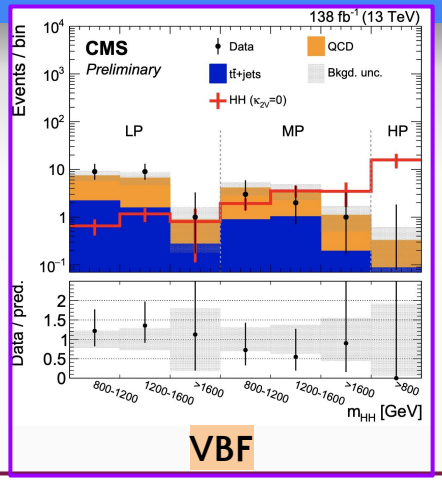
Results


138 fb-1 (13 TeV)
CMS Preliminary
Data
HH (μ = 3.5)
QCD, ggF H, VBF H
tt̄+jets
V+jets, VV
VH
tt̄H
Total unc.
BDT Bin 1
Yields in [110,140.] GeV
Bkg 6.41± 2.17
Data: 9
BDT Bin 2
BDT Bin 3
Fail region
Events / 10 GeV
Data / pred.
j2 mreg [GeV]
ggF


138 fb-1 (13 TeV)
CMS
Preliminary
Data
QCD
tt̄+jets
Bkgd. unc.
HH (κ2V=0)
LP
MP
HP
Events / bin
Data / pred.
800-1200
1200-1600
>1600
mHH [GeV]
VBF

Results





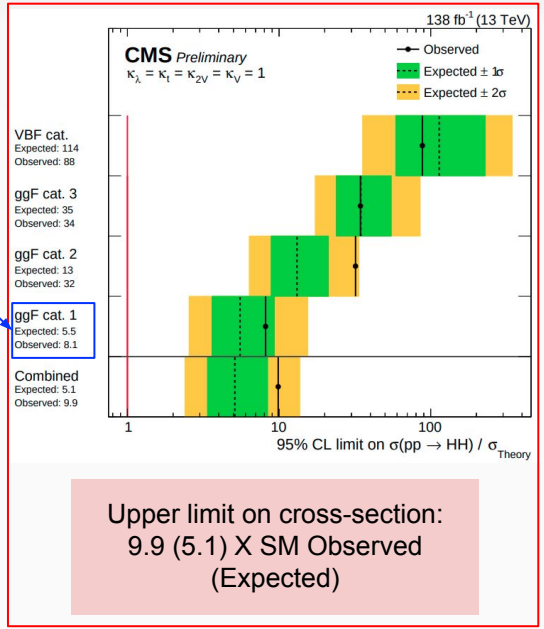
Uncertainty source	$\Delta\mu$	
Statistical	+2.55	−2.30
Signal extraction	+2.32	−2.06
QCD multijet modeling	+1.12	−1.01
$t\bar{t}$ modeling	+0.28	−0.19
Systematic	+2.09	−0.89
Simulated sample size	+0.55	−0.55
T_{Xbb} selection	+0.72	−0.32
Jet energy and mass scale and resolution	+0.54	−0.39
Trigger selection	+0.26	−0.03
Luminosity measurement	+0.13	−0.04
Pileup modeling	+0.05	−0.06
Other experimental uncertainties	+0.05	−0.03
Theoretical	+0.63	−0.63
Total	+3.30	−2.47

ggF + VBF boosted HH→4b combination

- Upper limit on σ_{HH}: 9.9 (5.1) X SM Obs(Exp) (~1.4σ excess over SM)

ggF + VBF boosted HH→4b combination

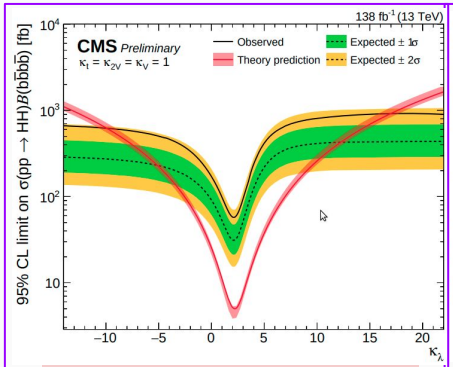
- Upper limit on σ_{HH}: 9.9 (5.1) X SM Obs(Exp) (~1.4σ excess over SM)
- ggF category dominates the upper limit on HH cross section



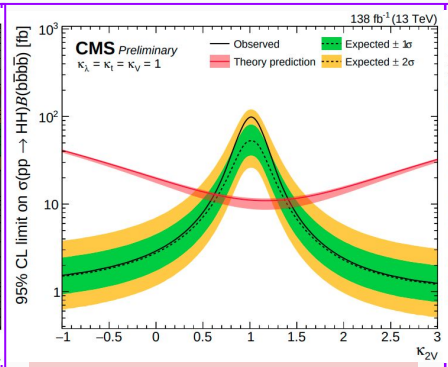
Upper limit on cross-section:
9.9 (5.1) X SM Observed
(Expected)

ggF + VBF boosted HH→4b combination

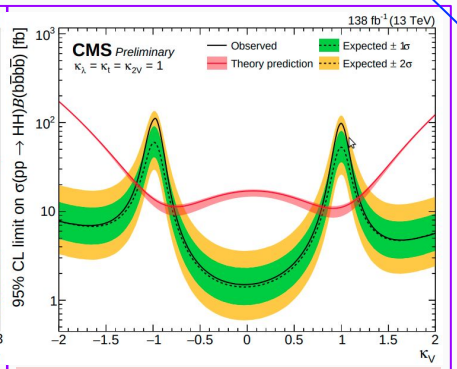
- Upper limit on σ_{HH}: 9.9 (5.1) X SM Obs(Exp) (~1.4σ excess over SM)
- ggF category dominates the upper limit on HH cross section
- VBF category sensitive to BSM values of κ_{2V}
  - κ_{2V}=0 excluded at 6.3σ for the first time when other couplings at SM values



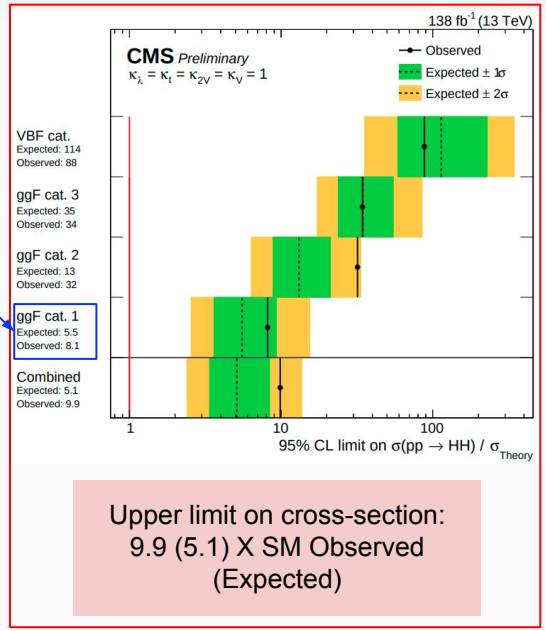
$k_\lambda \in$ [-9.9, 16.9] when $k_V = k_{2V} = 1$ at 95 %CL



$k_{2V} \in$ [0.62, 1.41] when $k_\lambda = k_V = 1$ at 95 %CL

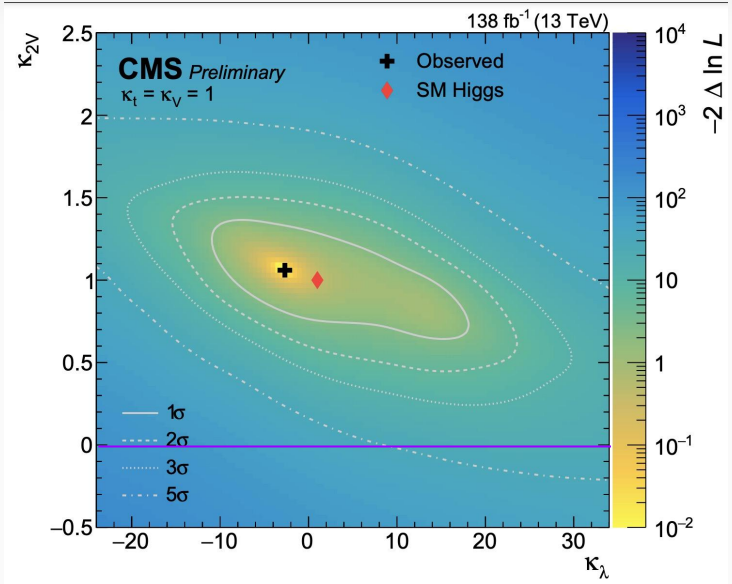


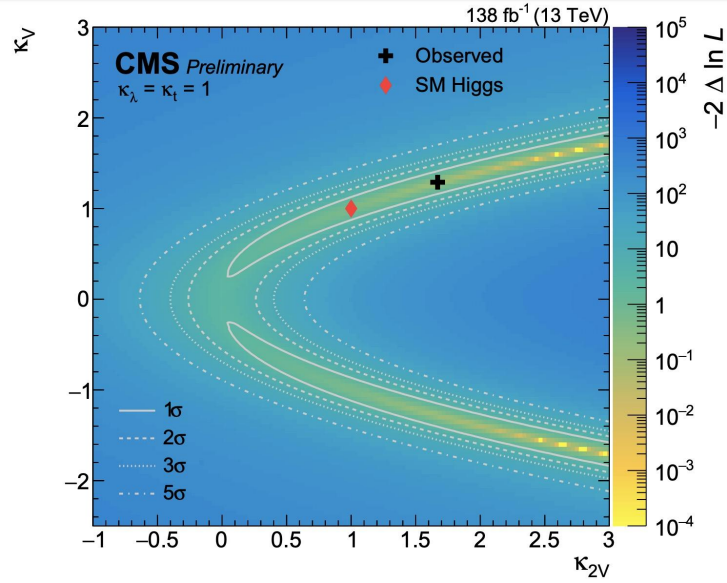
$k_V \in [-1.17,-0.79] \cup [0.81,1.18]$ when $k_\lambda = k_{2V} = 1$ at 95 %CL



Upper limit on cross-section: 9.9 (5.1) X SM Observed (Expected)

2D likelihood scans

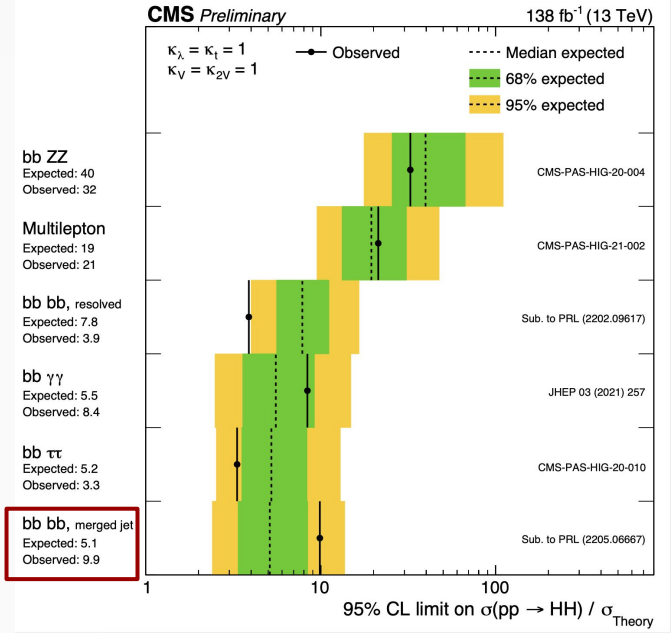


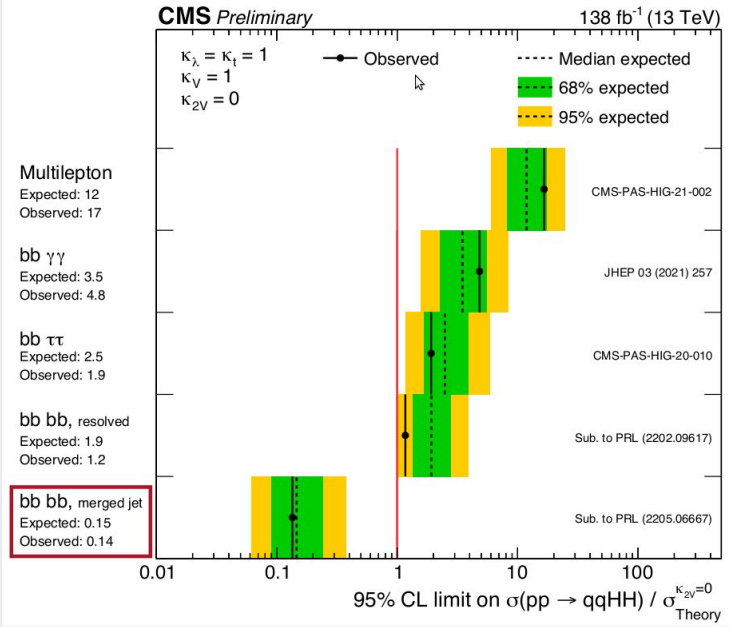


κ_{2V}=0 excluded at >3 σ for the first time for any value of κ_λ

Signal strength results in context

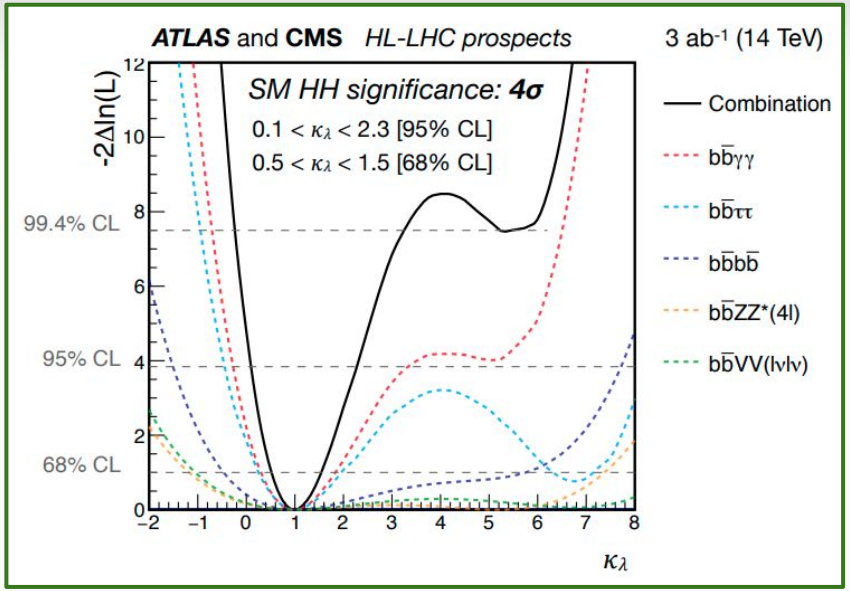
Best sensitivity to SM ggF HH production and BSM VBF HH production compared to other CMS analyses





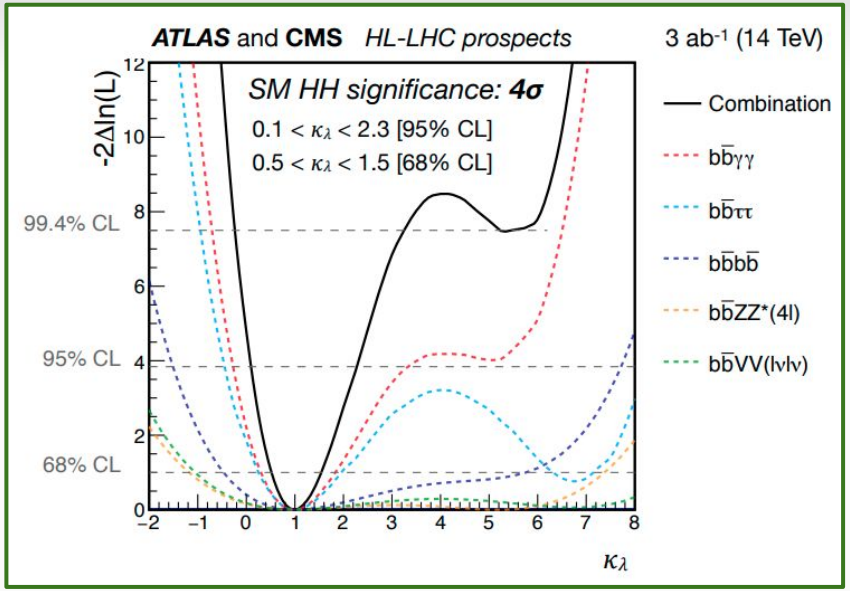
Future of HH

- HL-LHC Projected combined ATLAS +CMS ~ 4σ Exp. (CERN-LPCC-2018-04)



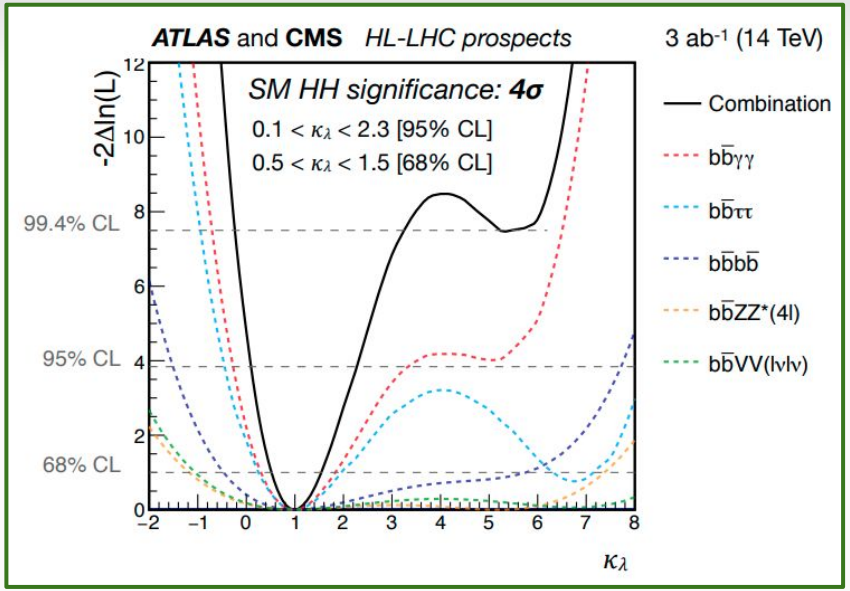
Future of HH

- HL-LHC Projected combined ATLAS +CMS ~ 4σ Exp. (CERN-LPCC-2018-04)
  - Only using the 2016/2017 analyses and does NOT include boosted analyses
  - Potential of 5σ → new boosted analysis methods/ constraining systematics



Future of HH

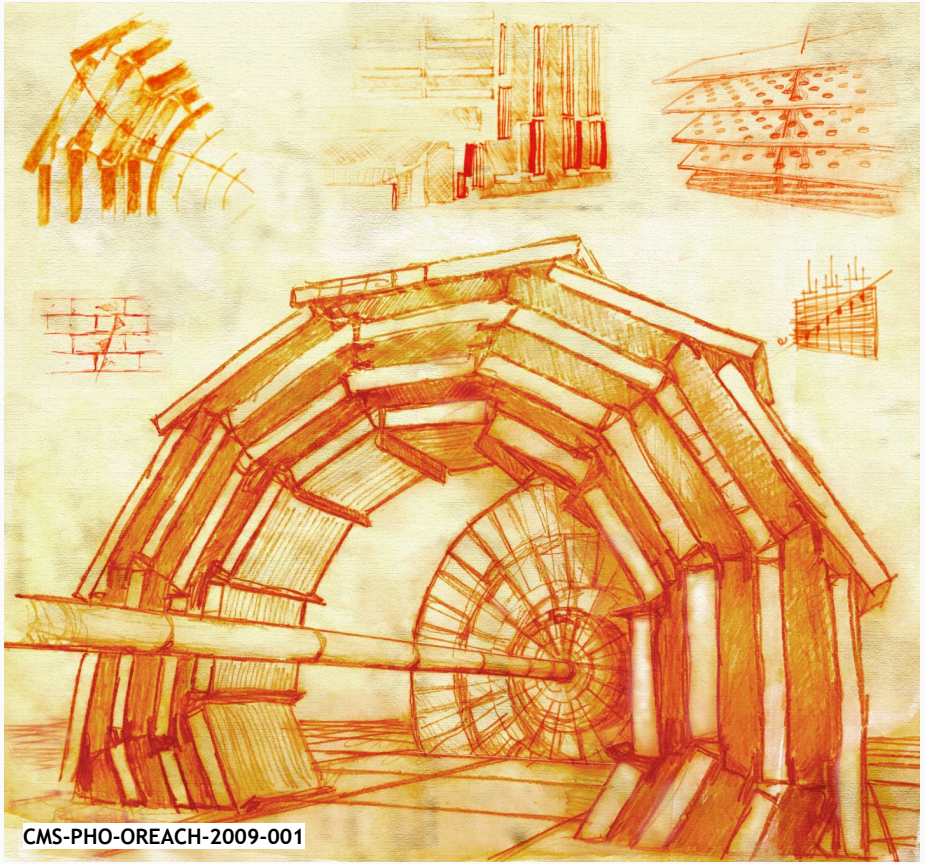
- HL-LHC Projected combined ATLAS +CMS ~ 4σ Exp. (CERN-LPCC-2018-04)
  - Only using the 2016/2017 analyses and does NOT include boosted analyses
  - Potential of 5σ → new boosted analysis methods/ constraining systematics



More boosted searches targeting bbVV etc. are underway .. stay tuned!

Thank you!

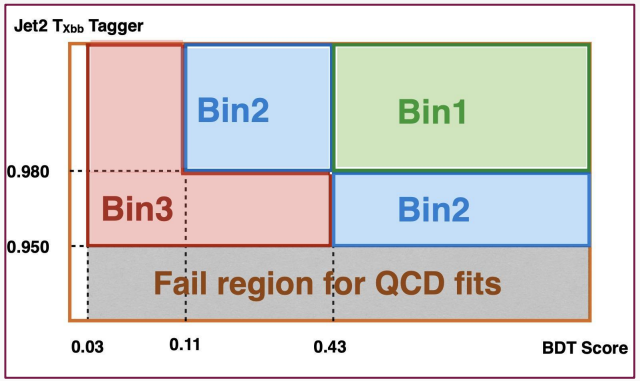
Questions ?



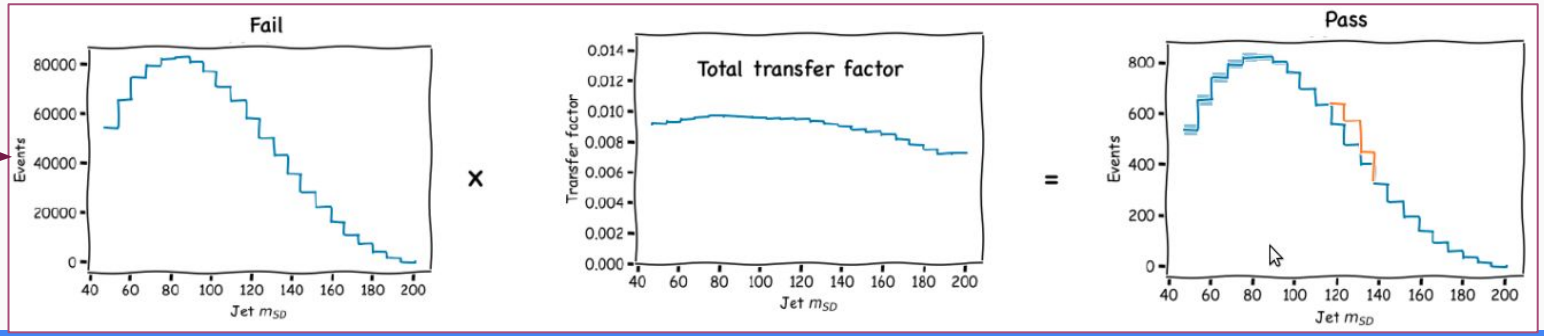
Backup

QCD background estimation in ggF category

- Fail region j_2 mass shape of QCD = Data - (non-QCD backgrounds)
- QCD distribution in the signal regions = QCD fail shape ⊗TF (j_2 Mass)
- The order of the transfer factor for each category is determined by performing a F-test and goodness-of-fit (GOF) test
  - measured to be polynomial of degree (0,0,0) for BDT Bin 1, Bin 2 and Bin 3 respectively



QCD shape estimated from data in control regions



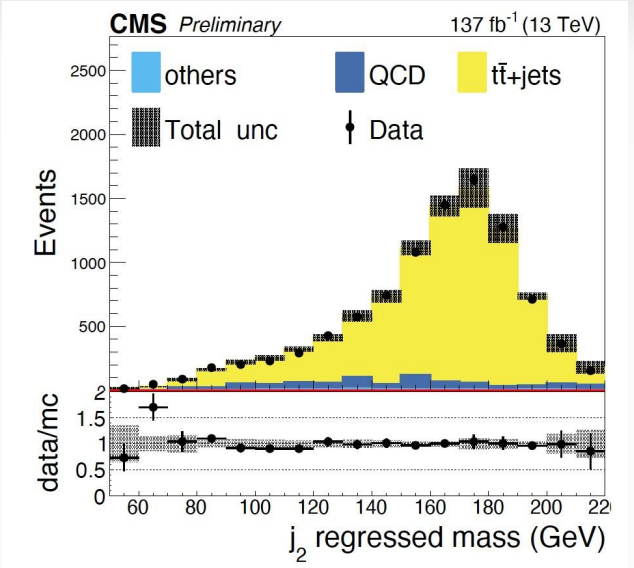
Corrections to top background in ggF category

Mis-modeling of the T_{Xbb} distribution

- Corrected in a semi-leptonic tt Control Region with one electron/muon and one fat-jet with p_T > 300 GeV
- Derive correction factors as data to simulation ratios in different jet T_{Xbb} bins.

Mis-modeling of the recoil of the tt system

- Corrected in a hadronic tt Control Region with two top-like fat-jets with p_T > 450 GeV and each fat-jet containing a sub-jet that is b-tagged
- Obtain correction factors with linear fits to data to simulation ratios in bins of p_T^{jj}



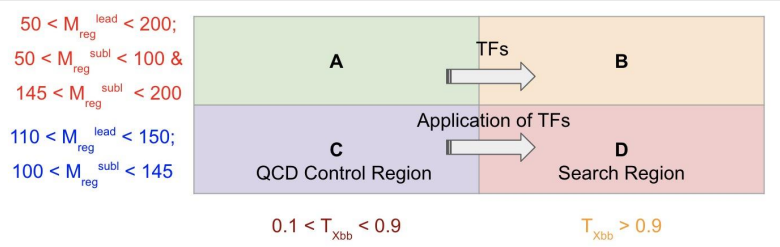
Jet 2 mass distribution in a hadronic tt Control Region post all corrections to the top background

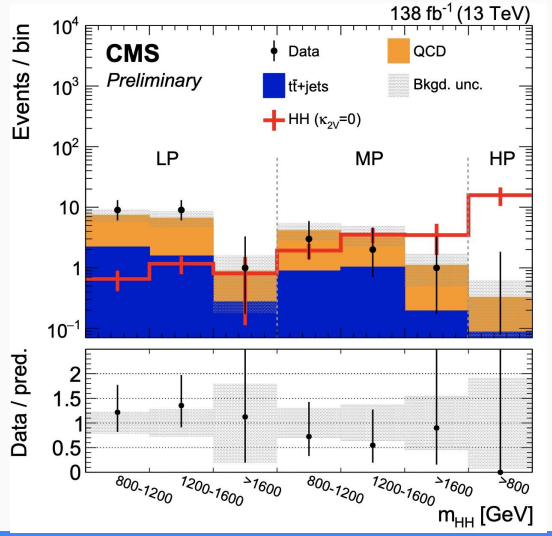
VBF analysis

- Leading fat-jet jet p_T > 500 GeV; sub-leading fat-jet jet p_T > 400 GeV
  - $\Delta\phi_{j1j2}$ > 2.6, $\Delta\eta_{j1j2}$ > 2.0
  - Jet 1 $m_{reg} \in$ [110, 150] GeV and Jet 2 $m_{reg} \in$ [100, 145] GeV
- Two small radius VBF jets
  - M_{jj} > 500 GeV, $\Delta\eta_{jj}$ > 4
  - $\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)} > 1.2$ between Higgs jets and VBF jets
- Three exclusive regions based on WPs:
  - High Purity (HP): Both Higgs candidate jets pass tight WP ($T_{Xbb} > 0.98$)
  - Medium Purity (MP): Both pass medium WP, but not tight WP ($0.94 < T_{Xbb} < 0.98$)
  - Low Purity (LP): Both pass loose WP, but not medium WP ($0.90 < T_{Xbb} < 0.94$)

VBF analysis

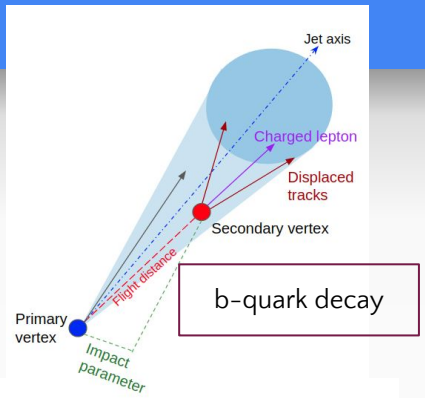
- ABCD method for QCD bkg estimation
  - Region D - search region
  - Define region C enriched in QCD multijet events with both fat jets having $0.1 < T_{Xbb} < 0.9$.
  - Define control region A and B with same T_{Xbb} selections as region C and D respectively, but in the **subleading jet mass sidebands**
  - Define TF = N_B / N_A
  - A separate TF is derived for each m_{HH} bin, in each search category (HP, MP and LP), and for each year.
  - The QCD background in the search region D: N_D = TF * N_c
- Just like ggF category, estimate Top background from simulation with data driven corrections
- Perform final fit to m_{HH} in different bins of T_{Xbb} score

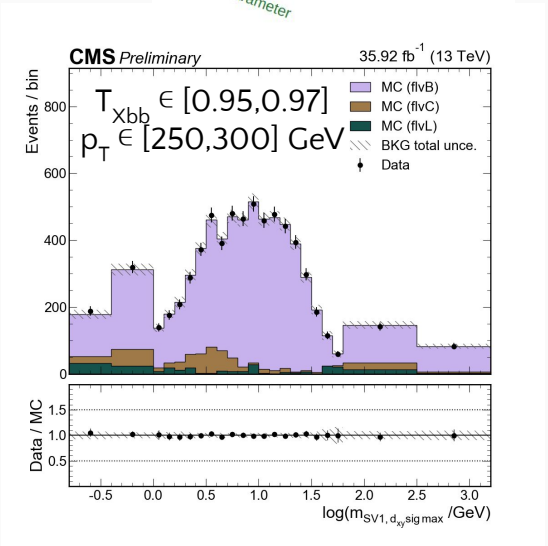




T_{Xbb} shape correction in HH signal

- T_{Xbb} gives us a boost in sensitivity, so it's important to ensure good Data/MC agreement in its shape for the signal jets
- Corrected using fat proxy jets originating from g→bb (benefits from the large statistics of QCD events)
- Each of the two sub-jets has at least one matched secondary vertex (SV)
  - Fit variable: log(M_{SV}), for SV with max d_{xy} (track impact parameter)
  - "b" template tends to peak at the B meson mass at ~ 5 GeV
  - "c" template peaks at the D meson mass at ~ 2 GeV
- Derive data to simulation ratios as correction factors in 2D grid of (p_T, Jet T_{Xbb})





Pulls and impacts

Dominant uncertainties

- shape uncertainty on the QCD multi-jet background.
- Jet energy scale, mass scale, and resolution uncertainties.
- T_{Xbb} shape
- Theory uncertainties for the gluon fusion HH signal modelling