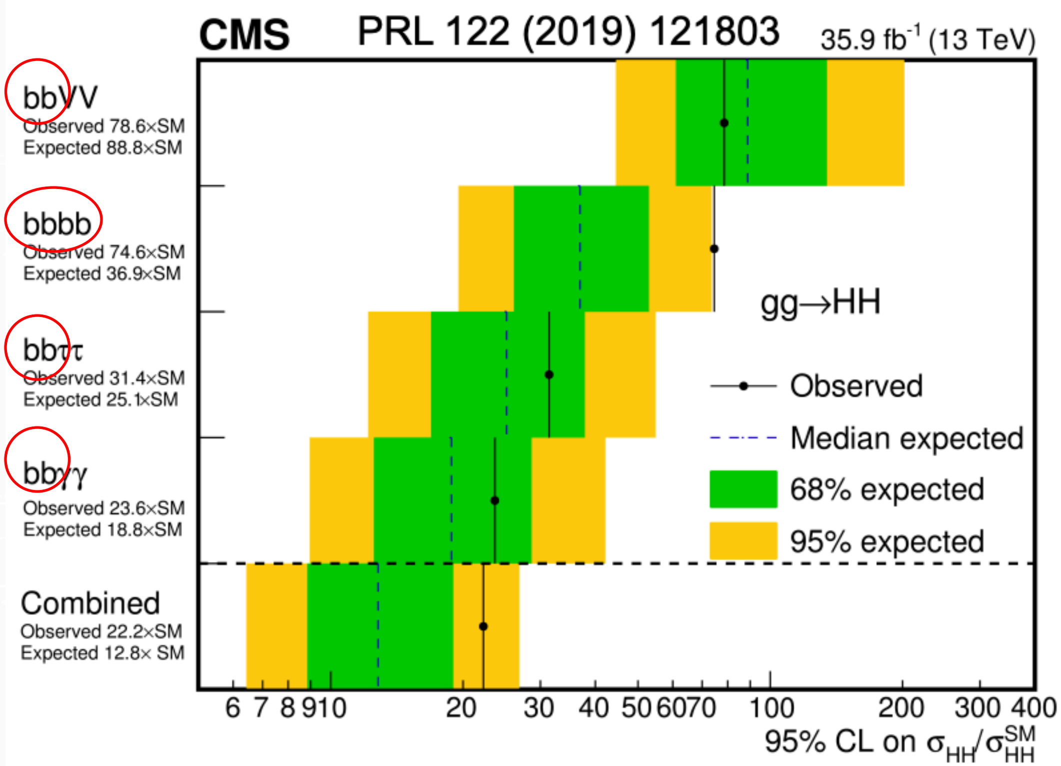
Di-Higgs searches with bottom quarks

Loukas Gouskos (CERN)
on behalf of the CMS Collaboration

Set the scene

- Searches with b quarks play an important role in the LHC physics program



Characteristic case:
Searches with Higgs boson(s)

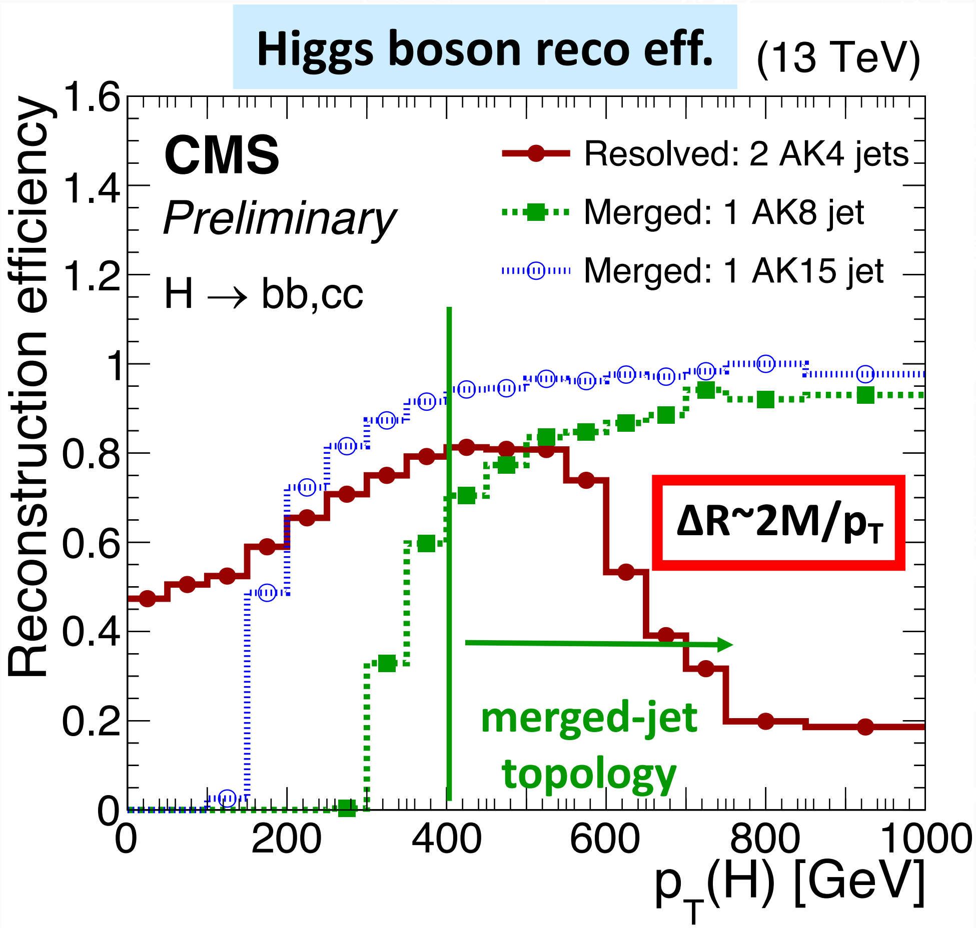
- H→bb: largest BR
- **but:** large QCD backgrounds

Key for success:
- Advanced b-jet identification algorithms
- Sophisticated analysis techniques

- Today's talk: Recent CMS results w/ **boosted** Higgs bosons decaying to b quarks
  - Particularly focus on the latest tools and techniques developed to enhance sensitivity
    - e.g., jet identification ("tagging"), jet mass regression, search design
- Full suite of results: CMS-B2G-analyses , CMS-HIG-analyses

General strategy

- Focus on scenarios that produce Higgs bosons with high-p_T ("boosted"):
  - i.e., decay products can be reconstructed as a single jet
- Target H→bb final states (largest BR)



Traditional approach

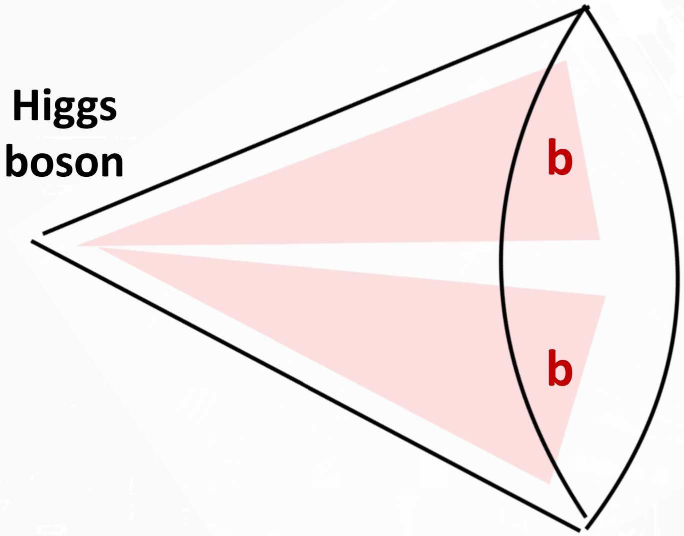
- Higgs decay products resolved in two "small-*R*" jets (R=0.4)

"Merged-jet topology"

- A single "large-R" jet to reconstruct the H→bb decay
- Better RECO efficiency at high-p_T
- Exploit the correlation between the two bottom quarks
- Reduced combinatorial BKG

The challenge: bottom quark identification

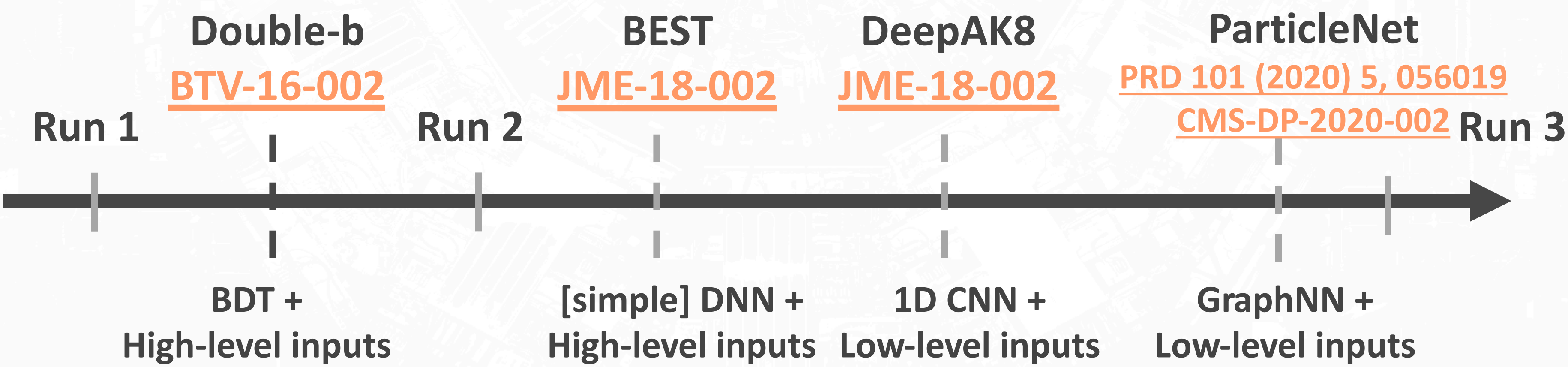
- Jet flavour identification (“tagging”): Topic of high interest in both TH and EXP



Main handles:

- **Jet mass [*]**
- **Jet substructure:** identify the 2-prong structure in a single large-R jet
- **Jet flavor:** Identify the 2 bottom quarks
- **Challenges:** pile-up, soft-radiation, etc..

- Enormous progress over the last few years:

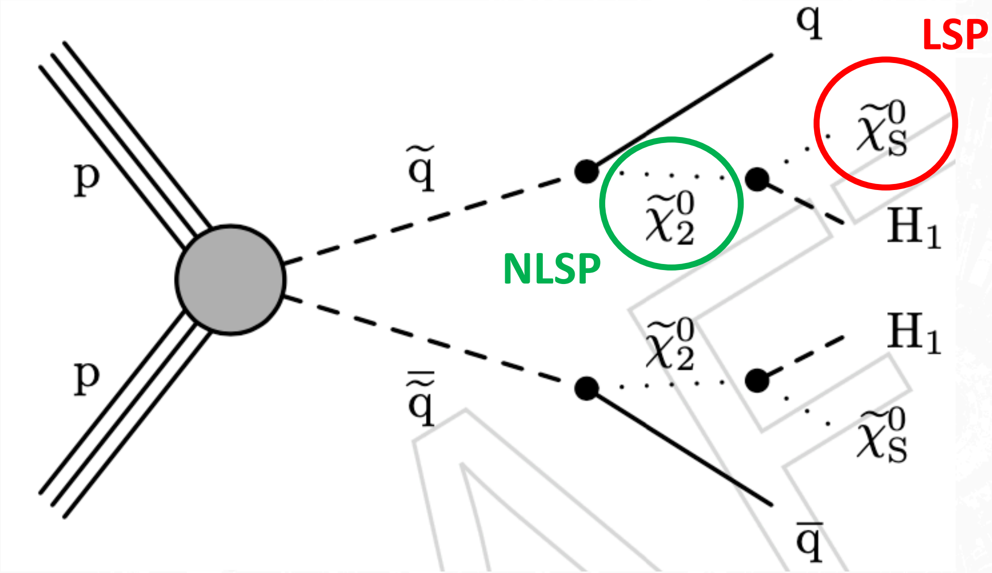


Search for light Higgs boson pairs in SUSY cascades

Search for light Higgs bosons in SUSY

Hot off the press! HIG-20-018

- Search of a pair of boosted light Higgs bosons in SUSY cascades
  - in the NMSSM SUSY extension to SM



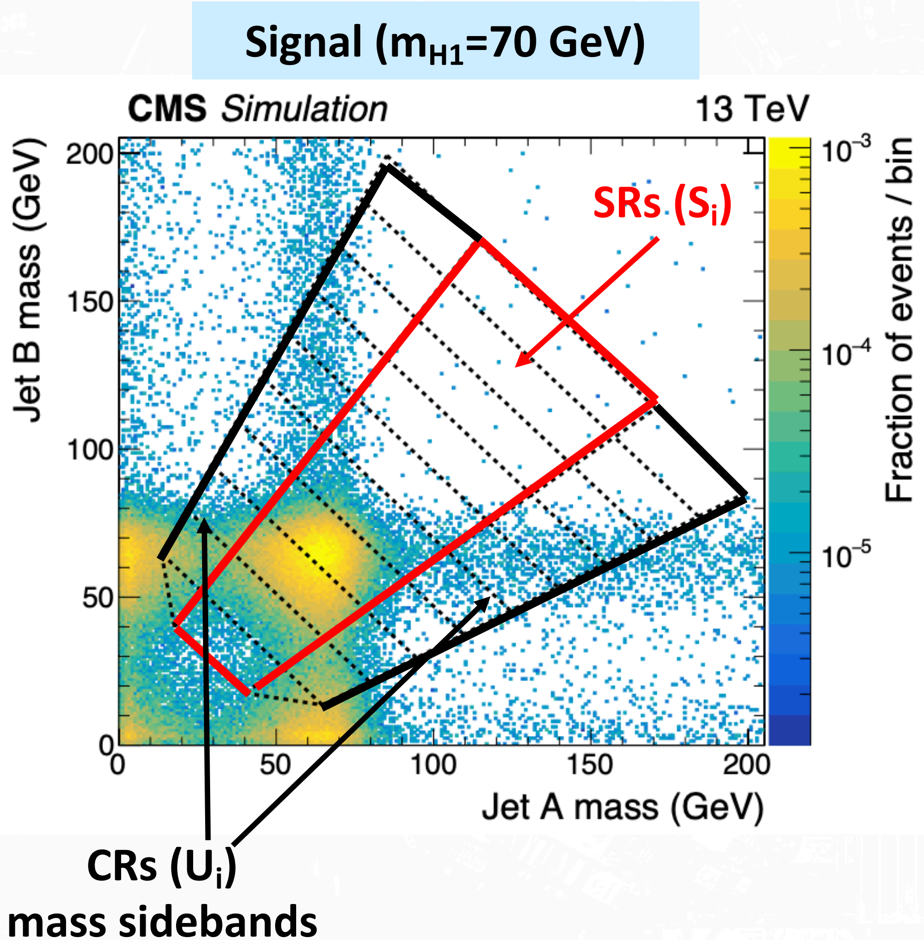
Model parameters:
- **m_{SUSY}:** mass scale of squarks/gluinos
- **$m_{H1.}$** : mass of CP-even Higgs
- $\boldsymbol{R_m} = m_{NLSP}/m_{H1} = 0.99$ [never explored at the LHC]
- **Δm** $= m_{NLSP} - m_{H1} - m_{LSP}$

- **Signature:** Multiple (b-) jets, high-p_T Higgs bosons, very little ME_T
- **Search strategy:**
  - Target $H_1 \rightarrow bb$ [largest BR] reconstructed as a single large-*R* jet: bb-tagging critical
  - No requirement on ME_T; disjoint regions in H_T to probe the SUSY scale

Search design

HIG-20-018

- $H\rightarrow bb$ cand: Identified using Double-b algorithm; jet mass RECO w/ softdrop
- **Signal extraction:** 2D fit of the mass of the two leading bb-tagged jets
- Dominant BKG: QCD multijet estimated from data [subdominant ttbar from MC]



BKG estimation

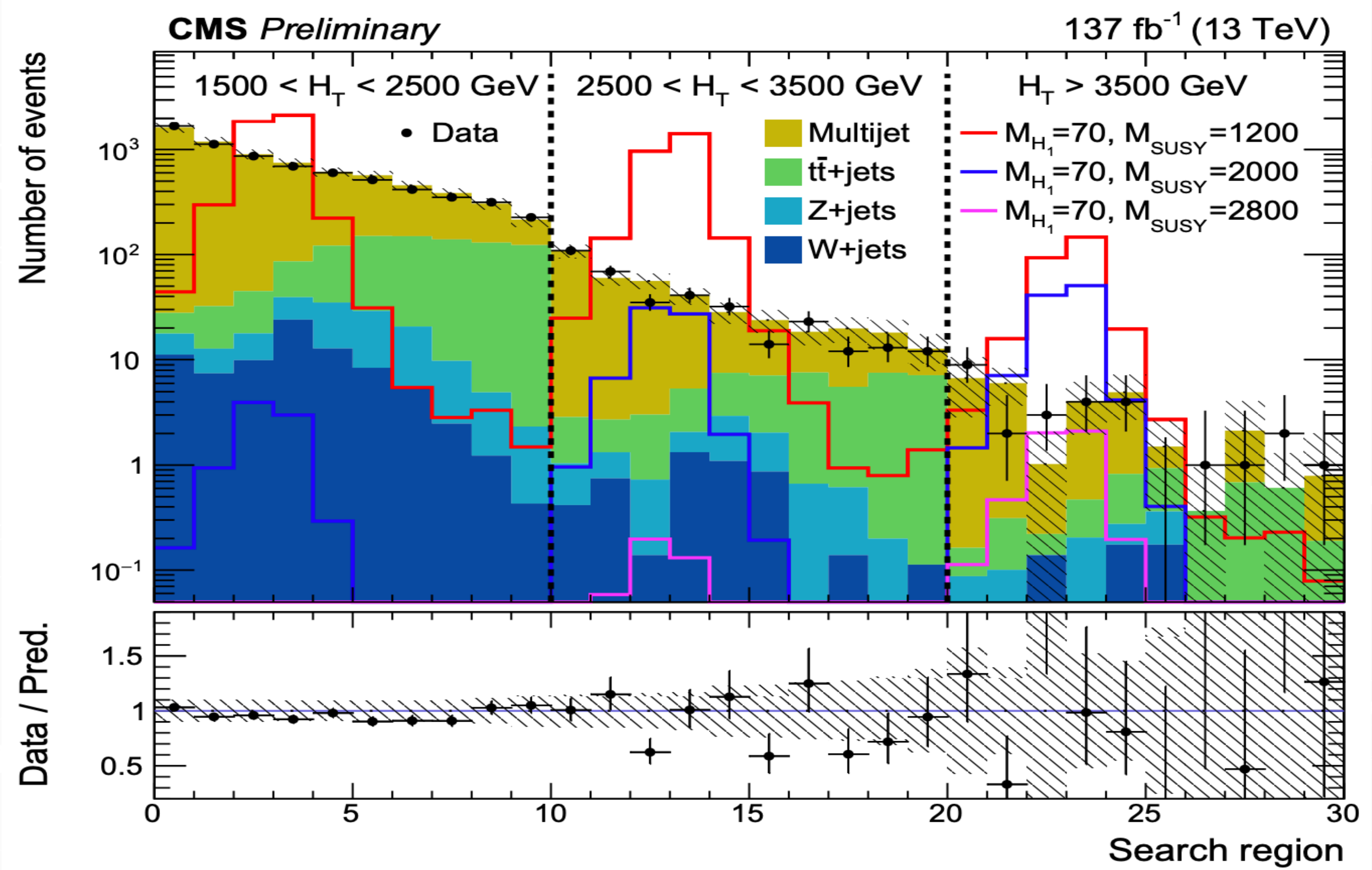
- QCD contribution in SRs from signal depleted mass sidebands (CRs)
- CRs designed to have similar yield as the corresponding SRs
- Correction factor extracted by inverting the double-b selection
  - Signal depleted region
- Prediction:

$$\hat{S}_i^{TR} = F_i \cdot \hat{U}_i^{TR},$$

F_i: True ratio of SR vs mass sideband yields obtained in the double-b inverted data region

Results

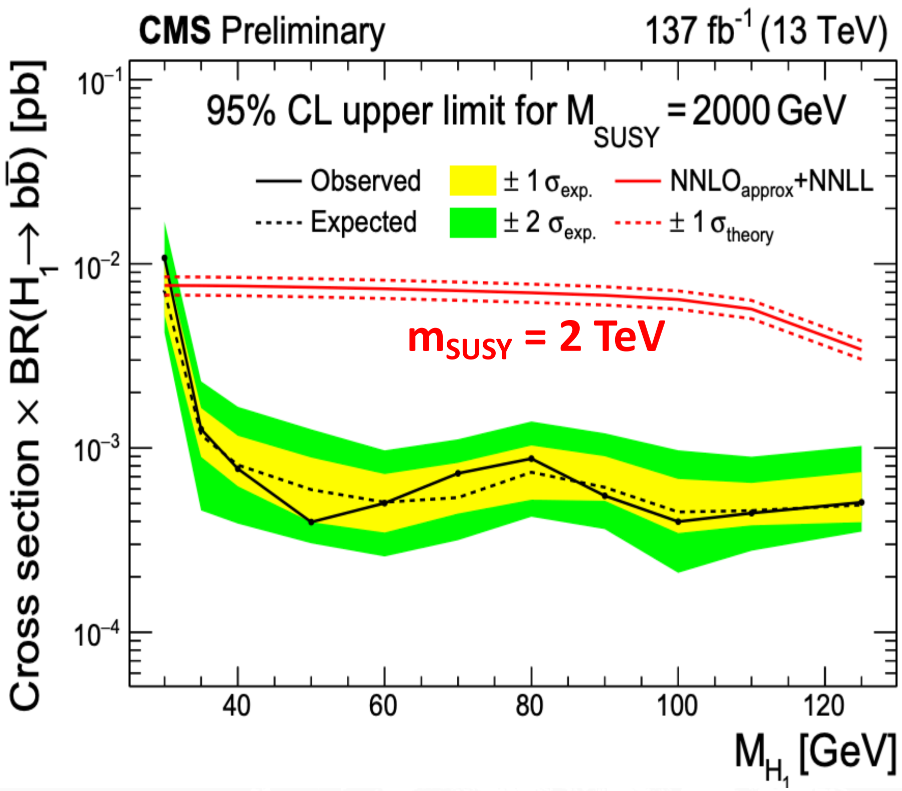
HIG-20-018



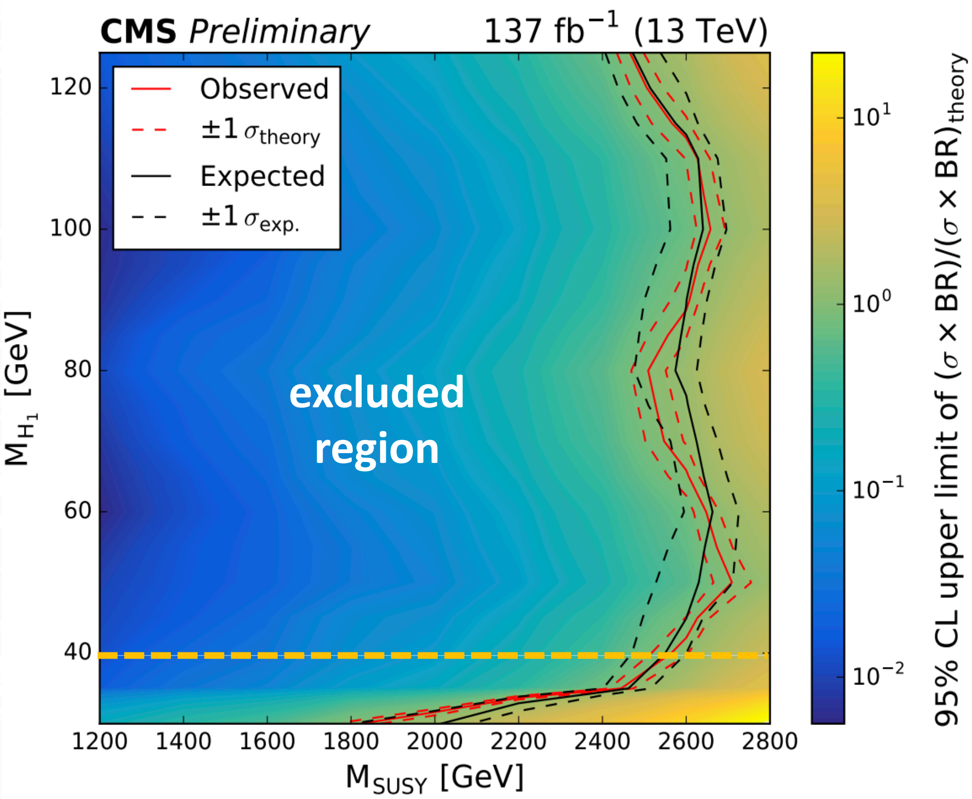
- No statistically significant excess observed

Interpretation

σ x BR UL for fixed m_{SUSY}



m_{SUSY} vs. m_{H1}



- 1^{st} limits on this signature [i.e., low ME_T] at the LHC

Resonant HH production
to 4b or bb+L(L) final states

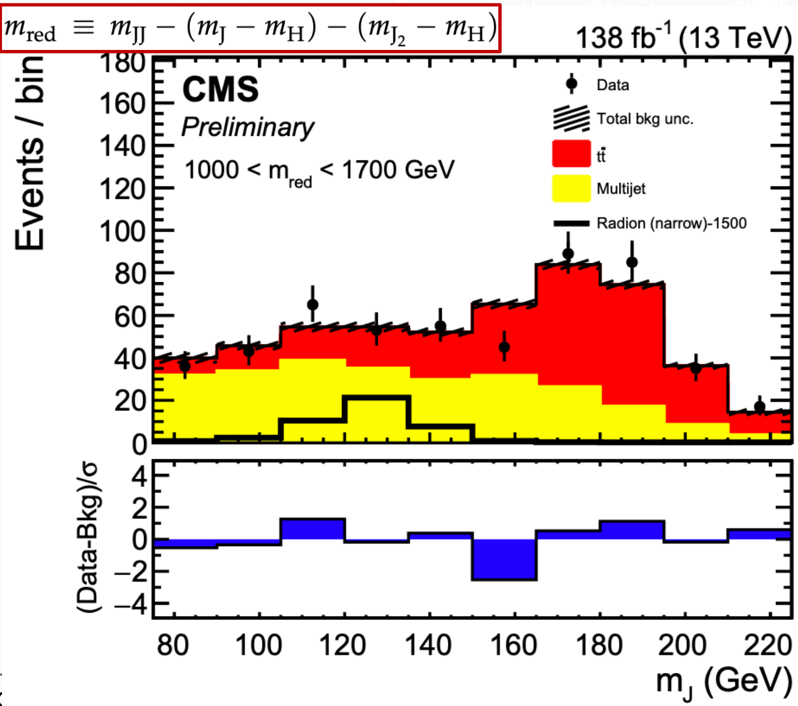
[Disclaimer: Only flash main points of the two searches and provide pointers to further info]

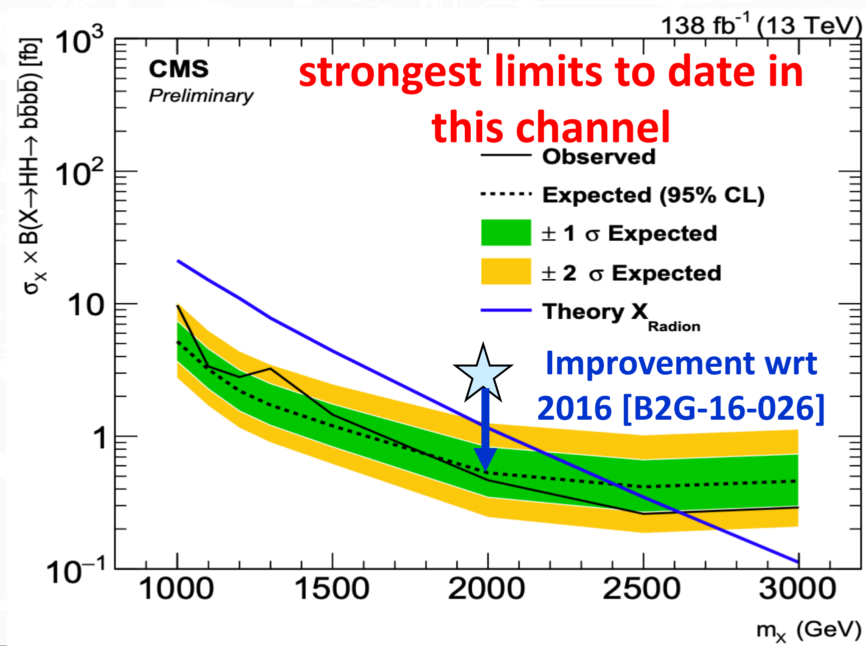
X → HH → 4b

More details: B2G-20-004

- **Search strategy:** Split in two main analysis categories
  - Boosted: H→bb candidates reconstructed as two large-*R* jets
    - identified using DeepAK8 algorithm [2.5x improved sensitivity vs. Double-b]
  - Semi-resolved: One large-*R* and two small-*R* jets to reconstruct Higgs candidates
    - small-R jet b-tagging using DeepJet [ref]
- **BKG estimation:** Main QCD bkg from data [based on DeepAK8 & m(j)]
  - ttbar: templates from MC; corrections extracted from data CRs
- **Signal extraction:** 2D fit of reduced HH mass (m_{red}) and leading jet mass (m_J)

$$m_{red} \equiv m_{JJ} - (m_J - m_H) - (m_{J_2} - m_H)$$



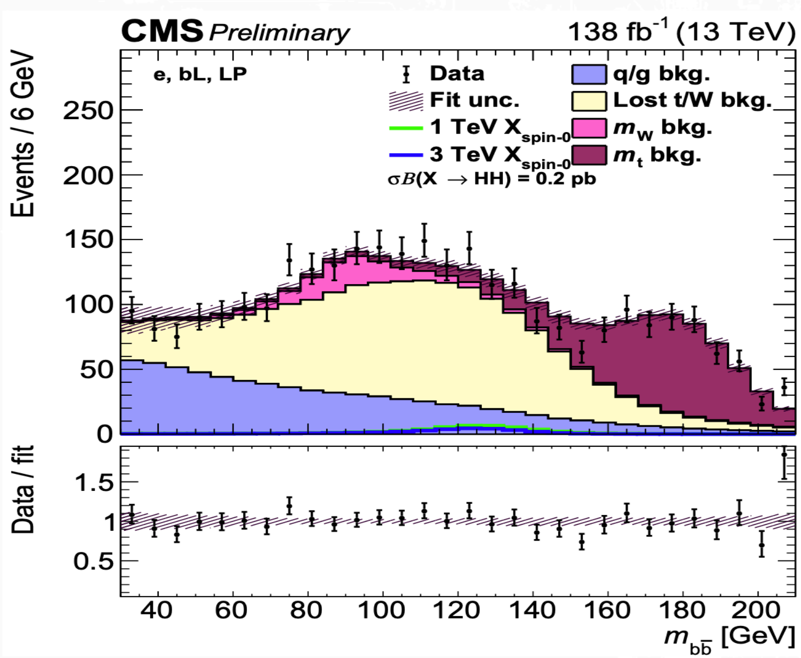


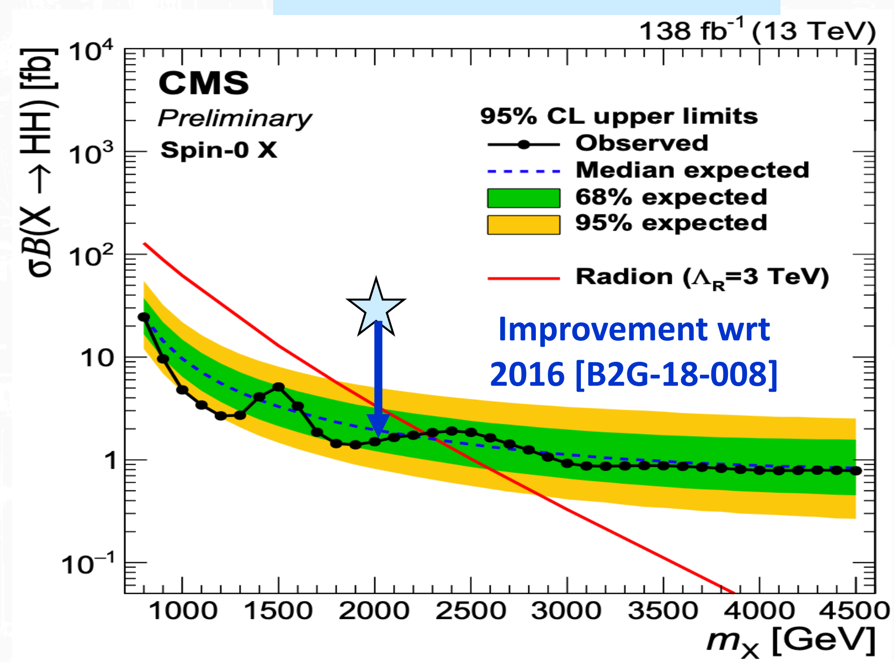
$X \to HH \to 2bL(L)$

More details: B2G-20-007

- **Search strategy:**
  - $H \to bb$: A large-R jet identified using DeepAK8
  - Leptonic leg: $H \to WW^*$ or $H \to \tau\tau$
    - **1L** and **2L** subcategories; in 1L $W \to qq$ reconstructed as a large-*R* jet [using τ_{21}]
- **BKG estimation** from data: Templates from MC w/ generous pre-fit unc.
  - Sophisticated approach: templates morph to account data-mc differences
  - Post-fit uncertainties constrained
- **Signal extraction:** 2D fit of m_{HH} and m_{bb} mass

Strongest limits to date in this channel

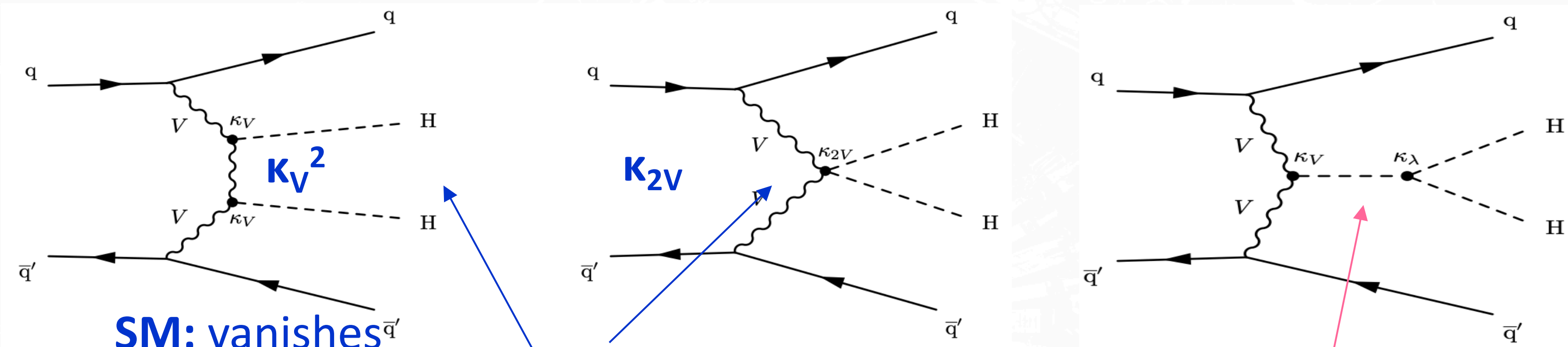




Nonresonant VBF HH → 4b production

Nonresonant VBF HH→4b production

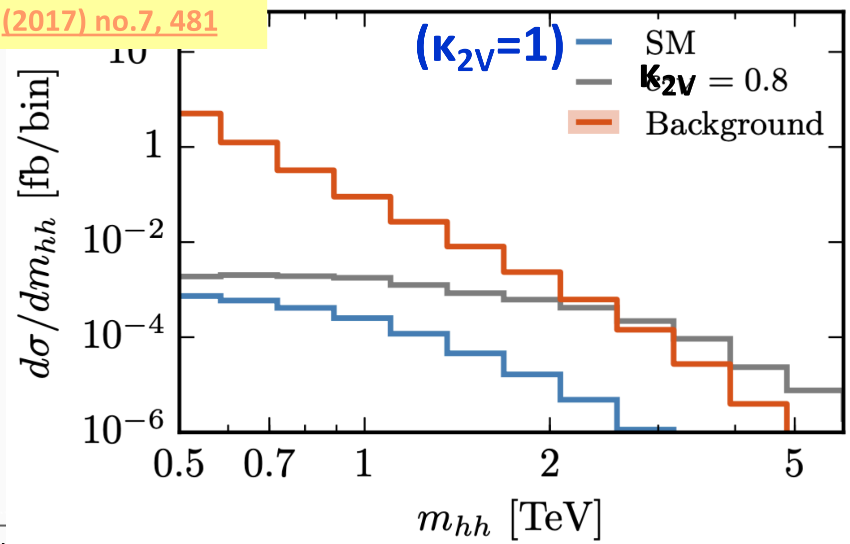
- Higgs pair production: Usually searched for in the ggF channel (σ~31 fb)
- VBF production: powerful probe of BSM physics [but very rare: σ~1.7 fb]
  - provides direct sensitivity to the VVHH (κ_{2V}) coupling



SM: vanishes
BSM: can be significantly modified

SM: negligible at large m_{HH}
highly off-shell

Eur.Phys.J. C77 (2017) no.7, 481



- Even "small" modifications on κ_{2V} have striking effect on m_{HH}
  - leading to high-p_T ("boosted") Higgs bosons

VBF HH→4b: search strategy in a nutshell

B2G-21-001

- Target the boosted regime and the 4b final state [largest possible BR]
  - Each Higgs boson reconstructed using large-*R* jets [*R*=0.8]
- Cornerstone of the search:
  - Identification of the H→bb candidates
  - Also: as precise as possible reconstruction of the large-*R* jet mass

Significant developments on these fronts

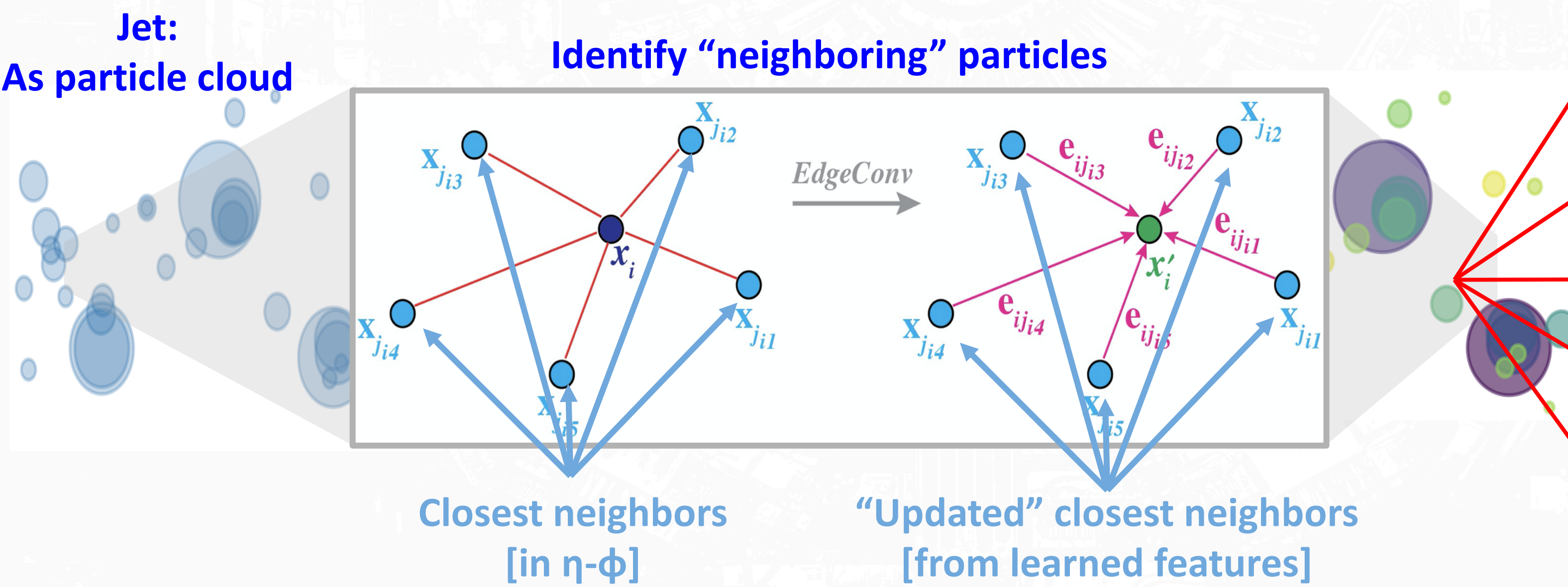
- A simple and robust VBF selection:
  - Two highest p_T small-*R* jets (R=0.4)
  - Large separation in η: $\Delta\eta(jj) > 4.0$ and large dijet mass (m_{jj}>500 GeV)
- Fit m_{HH} for signal extraction
  - Dominant BKGs estimated from data

Pushing the limits in jet tagging

- **ParticleNet**: Novel algorithm w/ improved **jet representation** & **network arch.**
  - Jet represented as a "particle cloud"
  - Architecture: Graph Neural Networks [i.e., DGCNN – add ref]
  - **Input:** PFcands & SV, **Output:** W/Z/H/top/QCD + decays; [same as DeepAK15]

PRD 101 (2020) 5, 056019
CMS-DP-2020-002

- Follow a hierarchical learning approach
  - **First:** Learn "local" structures; **Then:** move to more "global" features
  - Treat the particle cloud as a graph
    - **Particles** are the **vertices** of the graph
      Relationships between the particles are the **edges** of the graph



Category	Label
Top	bcq bqq bq cq
Higgs	bb cc VV*→qqqq
Z	bb cc qq
W	cq qq
QCD	g→bb g→cc b c others

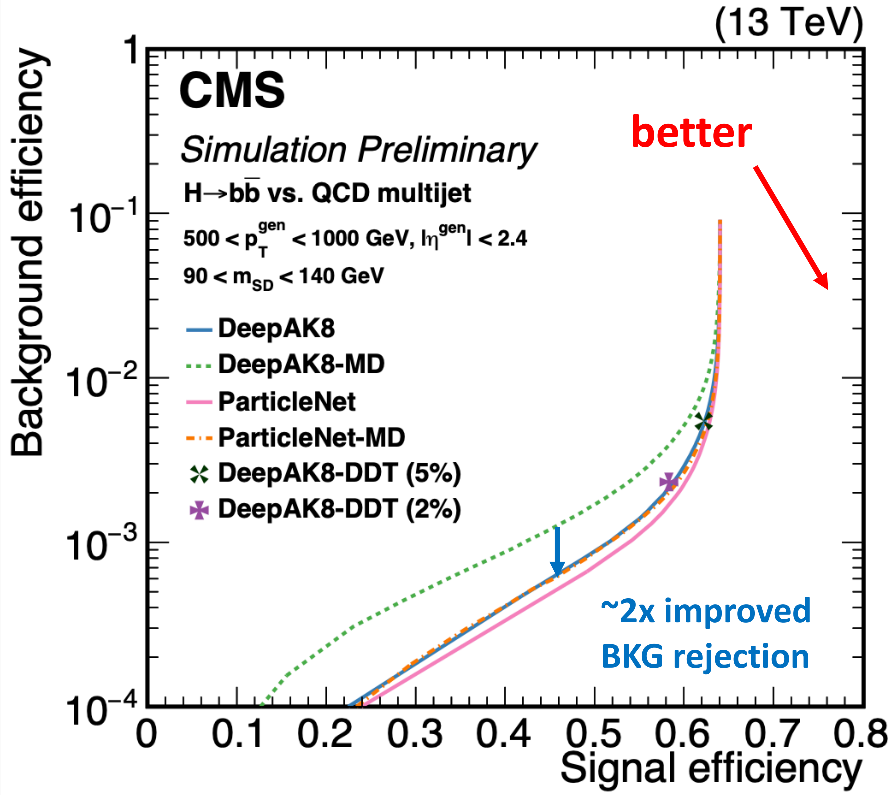
Pushing the limits in jet tagging (II)

- bb-tagging discriminant:

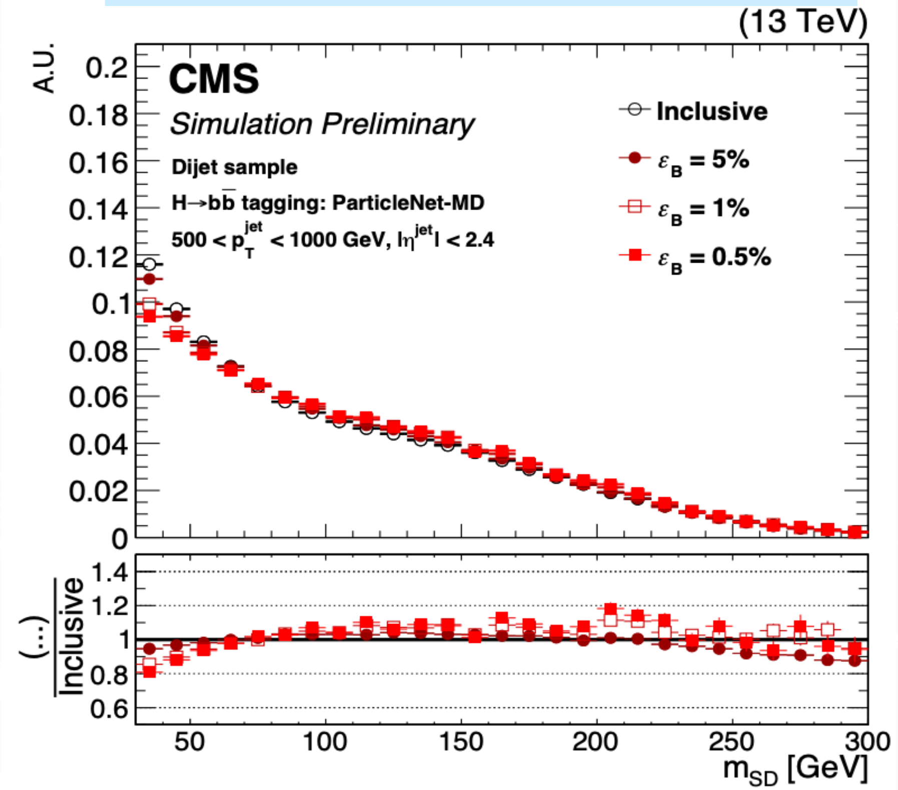
$$D_{bb} = \frac{\text{score}(X \to b\bar{b})}{\text{score}(X \to b\bar{b}) + \text{score}(QCD)}$$

PRD 101 (2020) 5, 056019
CMS-DP-2020-002

Performance in MC



New jet mass decorrelation method
No signs of mass sculpting



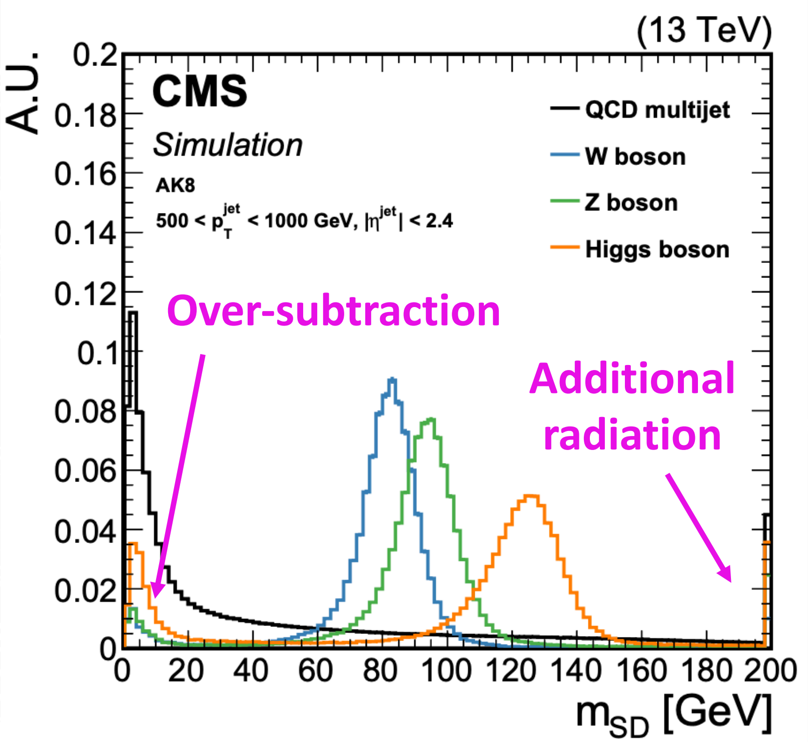
- Calibration in data using proxy jets from gluon→bb
  - Data-MC correction factors typically ~1 with ~20% uncertainty

Large-*R* jet mass regression

- Jet mass: powerful observable to discriminate signal (e.g., H→bb jets) from BKGs [e.g., QCD jets]
  - but very sensitive to soft radiation, pileup, …

Large-*R* jet

- Grooming techniques [e.g., SoftDrop] have been developed to mitigate this effect:
  - Iteratively decluster the jet and remove constituents that are:
    - soft and/or wide angle
  - Pros: simple and well tested in data
  - Cons: some inefficiency
    - e.g., some two prong jet identified as 1-prong



- Decays to bb/cc:
  - additional energy loss via the (undetected) neutrinos from semileptonic decays

Large-*R* jet mass regression (II)

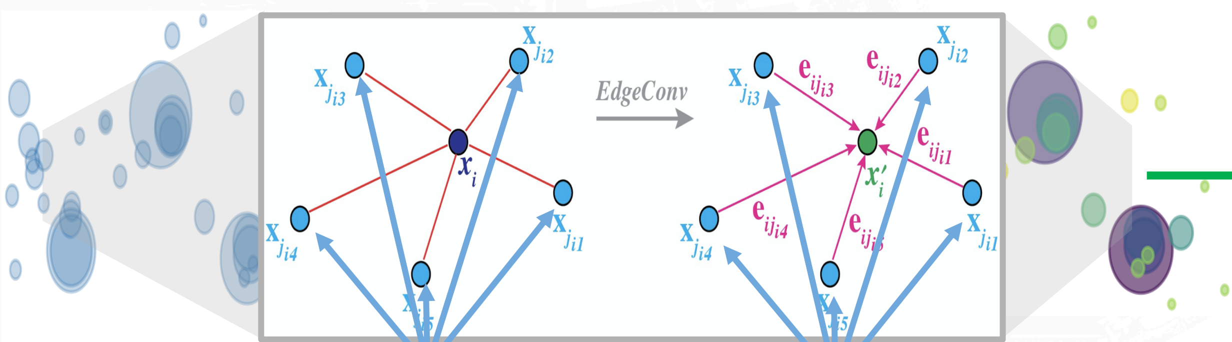
CMS-DP-2021-017

- Develop algorithm to reconstruct jet mass with best possible scale & resolution
  - Meanwhile: avoid "sculpting" of the QCD jet mass distribution
- Exploit **ParticleNet** architecture to predict m(jet) directly from jet constituents
  - Same inputs (PF candidates + SV) and same training configuration as for jet tagging

~~**ParticleNet for Jet Classification**~~ → **ParticleNet for Jet mass regression**

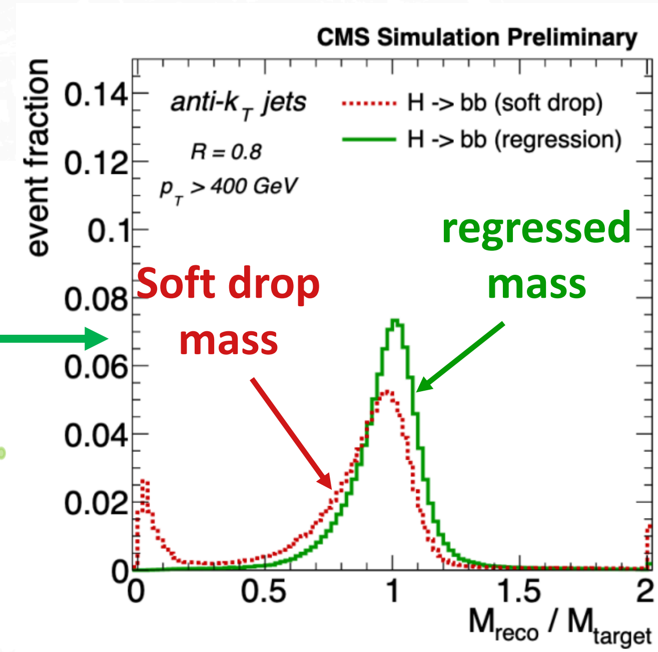
Jet:
As particle cloud

Identify "neighboring" particles



Closest neighbors [in η-ϕ]

"Updated" closest neighbors [from learned features]

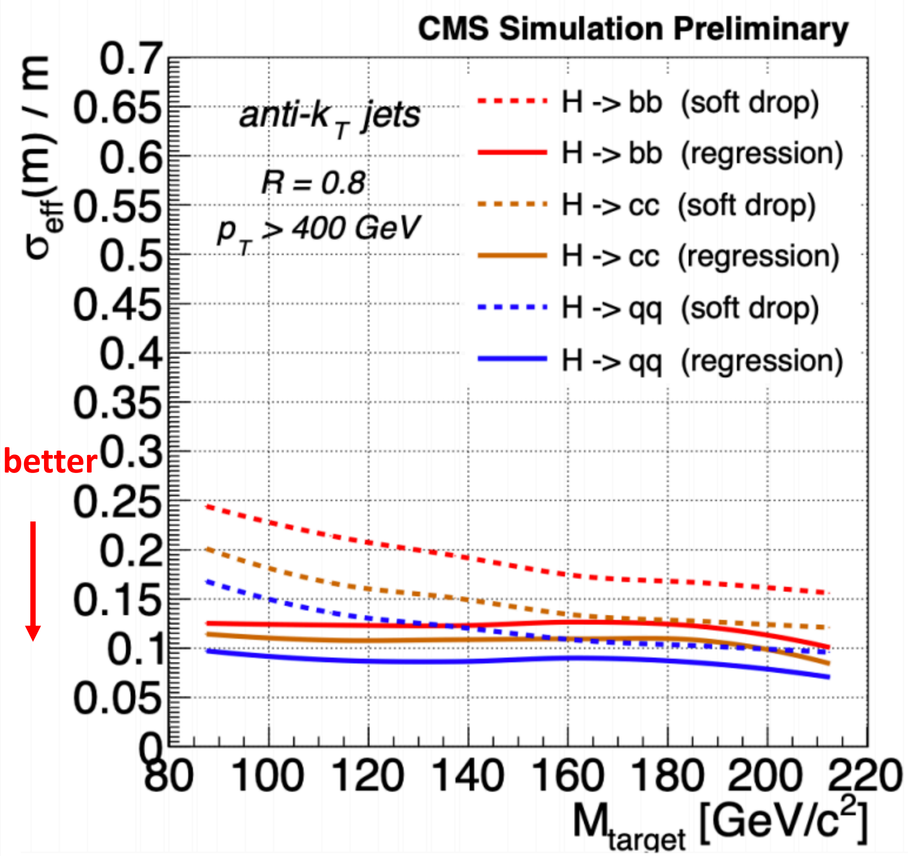


- Substantial improvement in both mass scale & mass resolution
- Tails in m(SD) significantly reduced

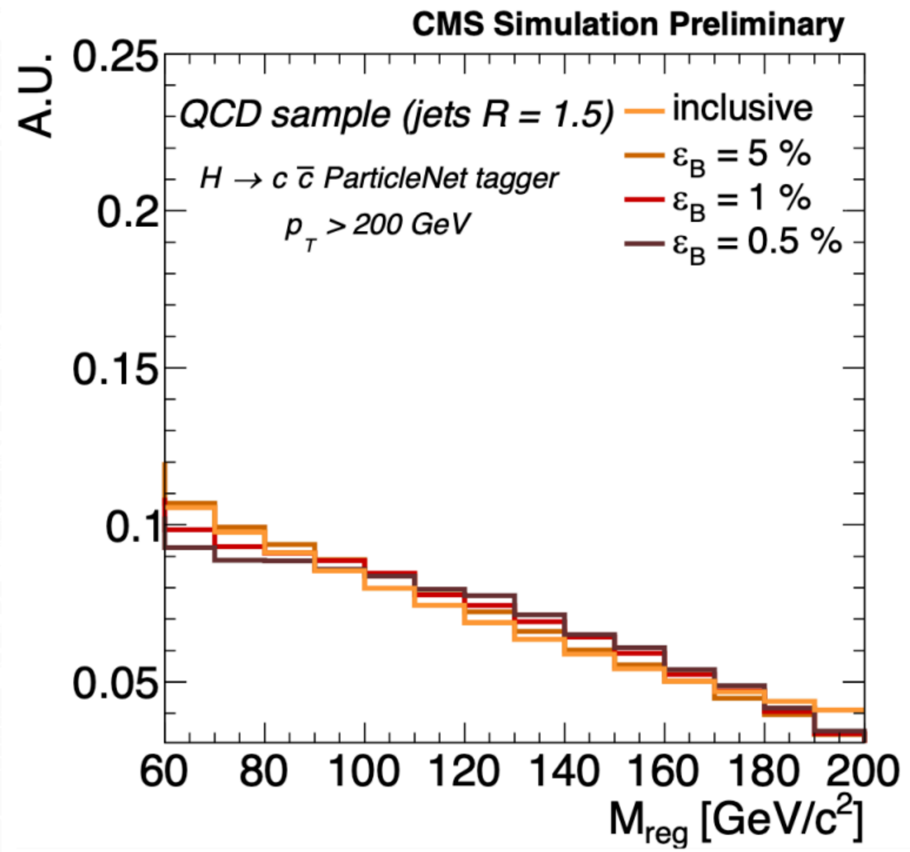
Large-*R* jet mass regression: Performance

CMS-DP-2021-017

Mass resolution vs. m(X)



Regressed Mass vs. Tagger WP



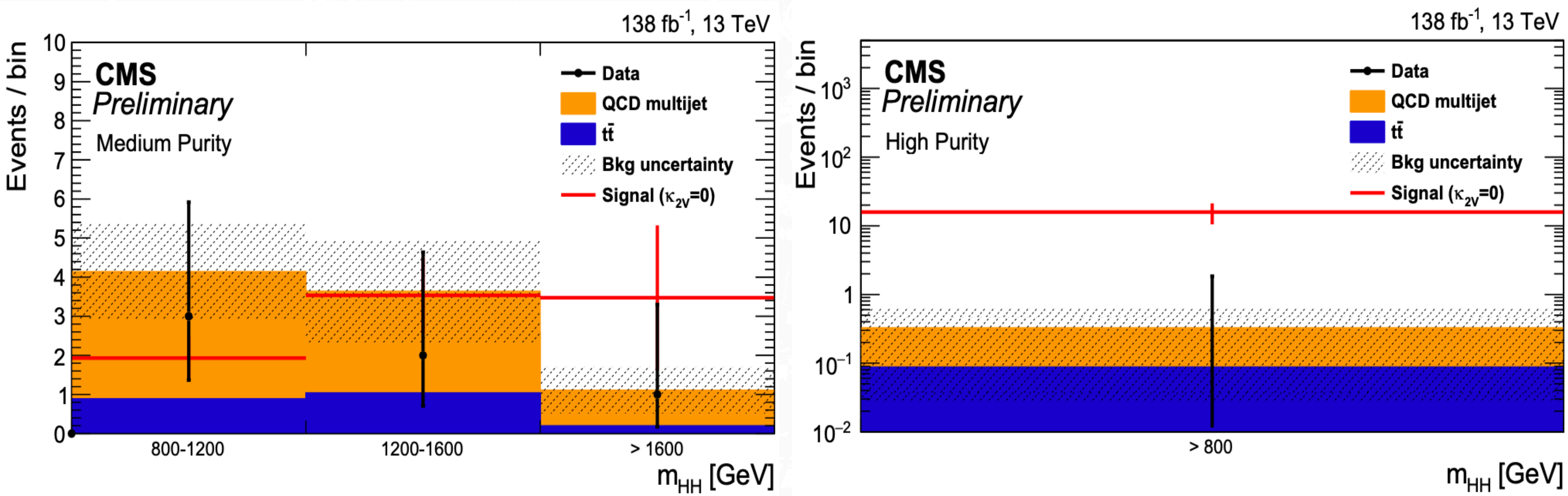
- Mass resolution stable across m(X)
- No indication of mass sculpting – even for very tight WPs
- Up to ~20-25% improvement in analysis sensitivity with H→bb/cc
- Calibration using W jets: scale (resolution) correction < 1% (3%)

Signal extraction

B2G-21-001

- Analysis carried out in three disjoint categories based on the bb-discriminant of the Higgs candidates
- Fit m_{HH} **QCD** and **ttbar** templates, in each of the three analysis categories
  - QCD estimated from data using an ABCD method [back-up]

Medium purity **High purity**



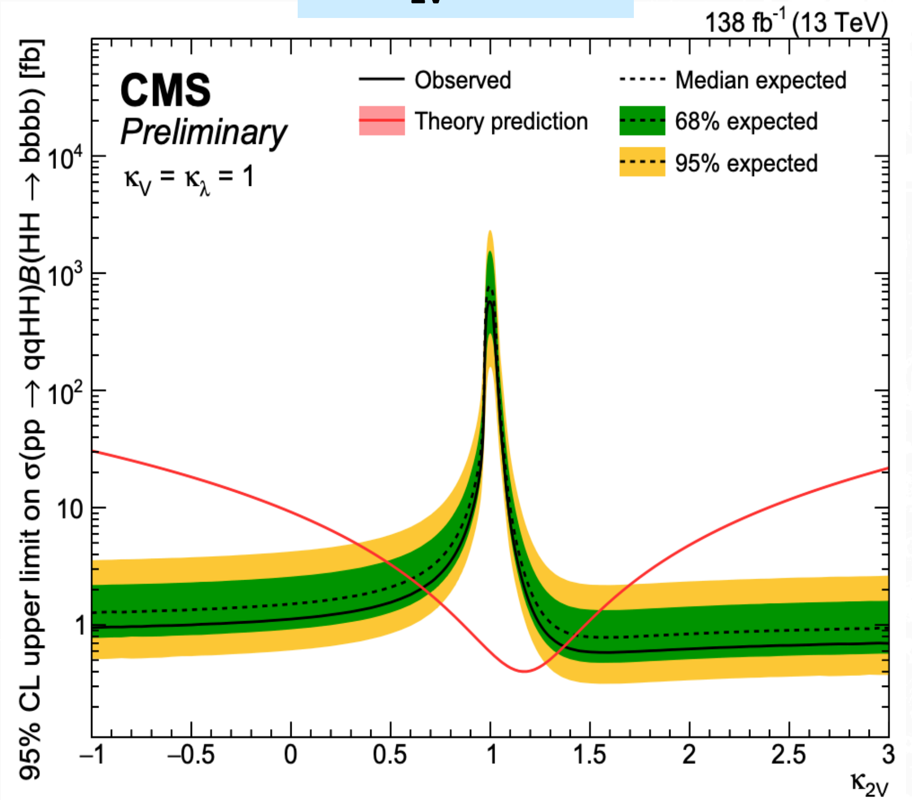
- Prediction agrees well with observed data
- Expected contribution for κ_{2V}=0 shown for illustration
  - all other couplings fixed to SM values

1st analysis to use ParticleNet [more analyses in the pipeline]

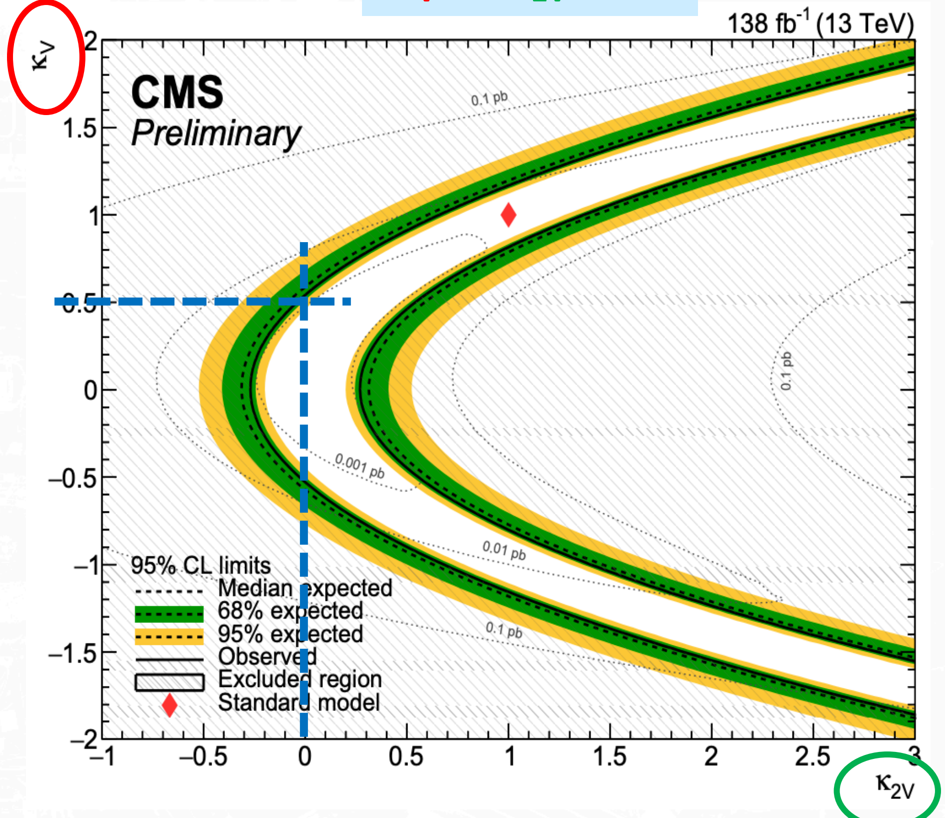
Interpretation

B2G-21-001

κ_{2V} scan



κ_V vs. κ_{2V} scan

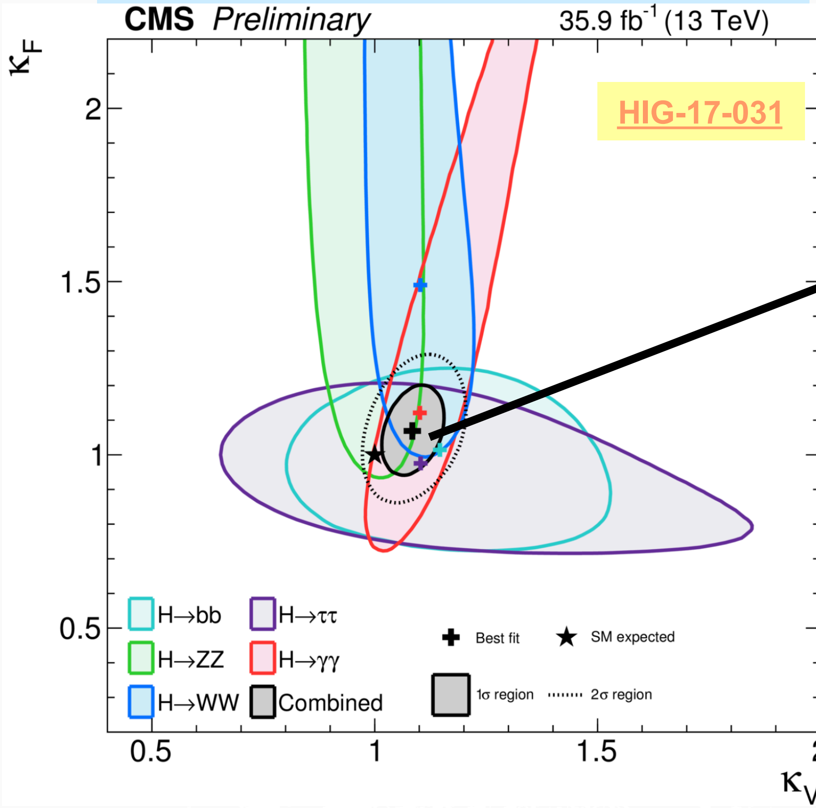


- Allowed values: **0.6 < κ_{2V} < 1.4**
- strongest constraints on κ_{2V} to date:
  - ATLAS: -0.6<κ_{2V}<3.1 [JHEP07(2020)108]
  - CMS [resolved]: -0.4<κ_{2V}<2.5 [HIG-19-018]
- 1[st] time to exclude κ_{2V} = 0 hypothesis

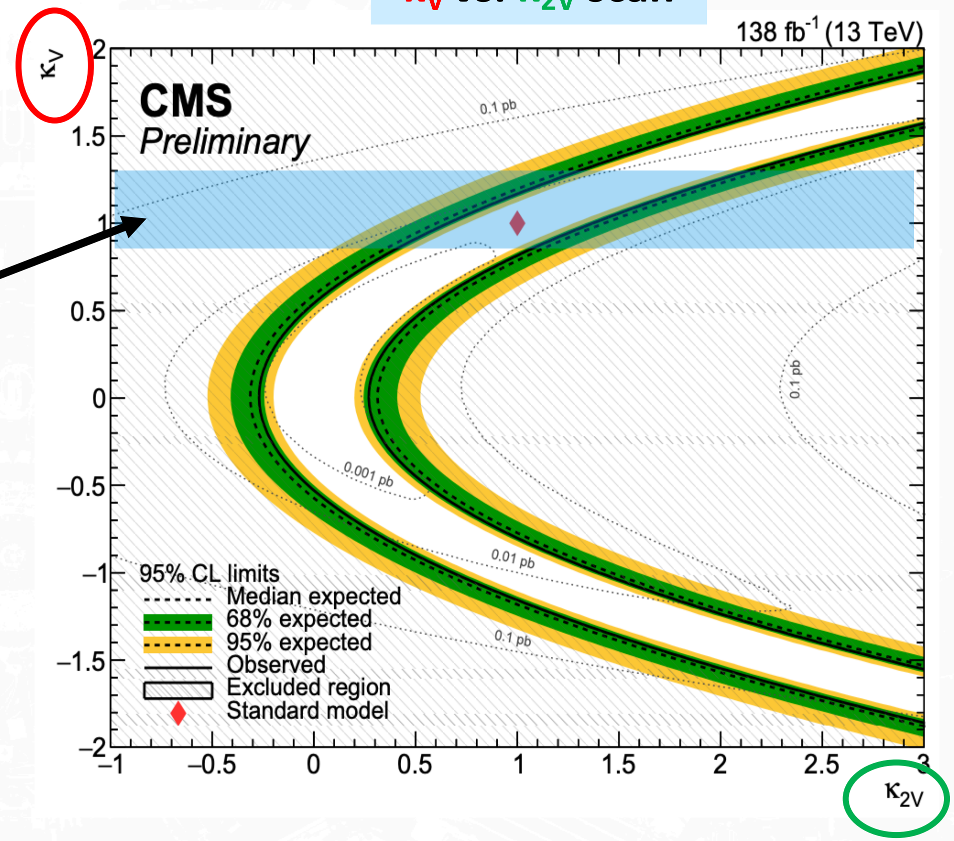
- Understanding interplay between κ_V and κ_{2V}
- For all values with κ_V > 0.5, "confirms" existence of the HHVV coupling

Interpretation (II)

κ_V constraints from single-Higgs



κ_V vs. κ_{2V} scan



- Single-Higgs measurements provide the tightest constraints on κ_V
  - Combined measurement: κ_V ~1.1 w/ O(20%) at 2σ
- κ_{2V} = 0 hypothesis highly disfavored when constraints on κ_V are considered

Summary

- Searches with boosted Higgs bosons and multiple bottom quarks in the final state are of particular importance for the success of the LHC physics program
  - a small subset shown today

- Enormous effort in both the Theory and Experiment communities to improve existing jet tools
  - Developments in jet tagging traditionally led the way
  - Now extended to other areas: e.g., jet mass regression

- More sophisticated techniques and/or analyses targeting different topologies [e.g., VBF HH in boosted regime] have been exploited

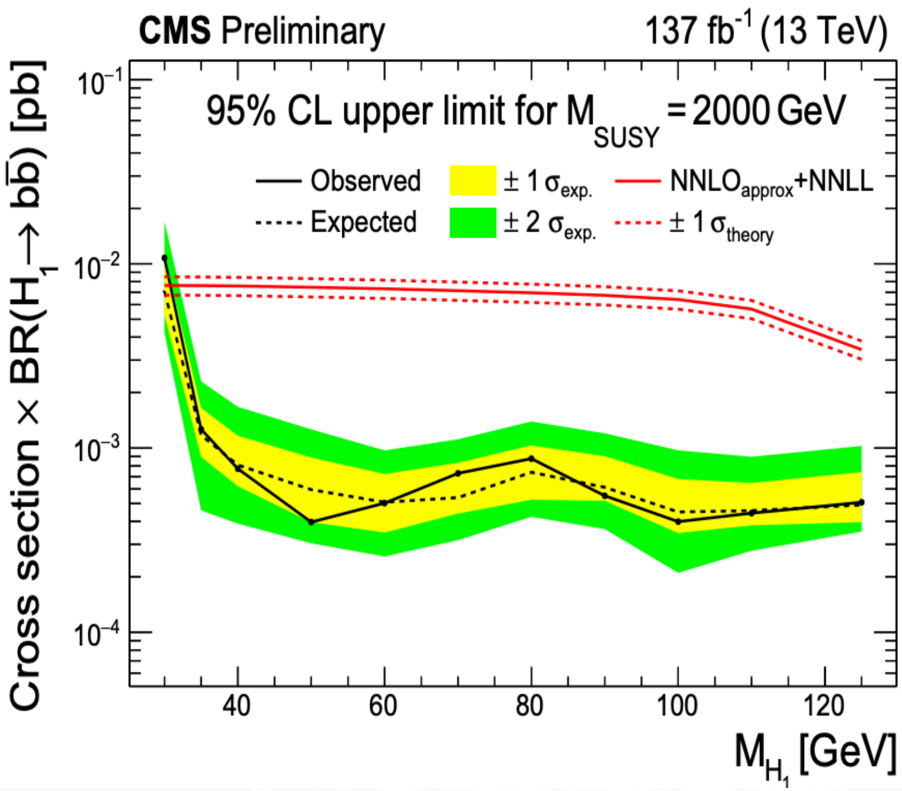
- All these efforts pay off yielding substantial improvements in the sensitivity of the physics analyses;
  - reaching already sensitivity expected with [much] much more data

Backups

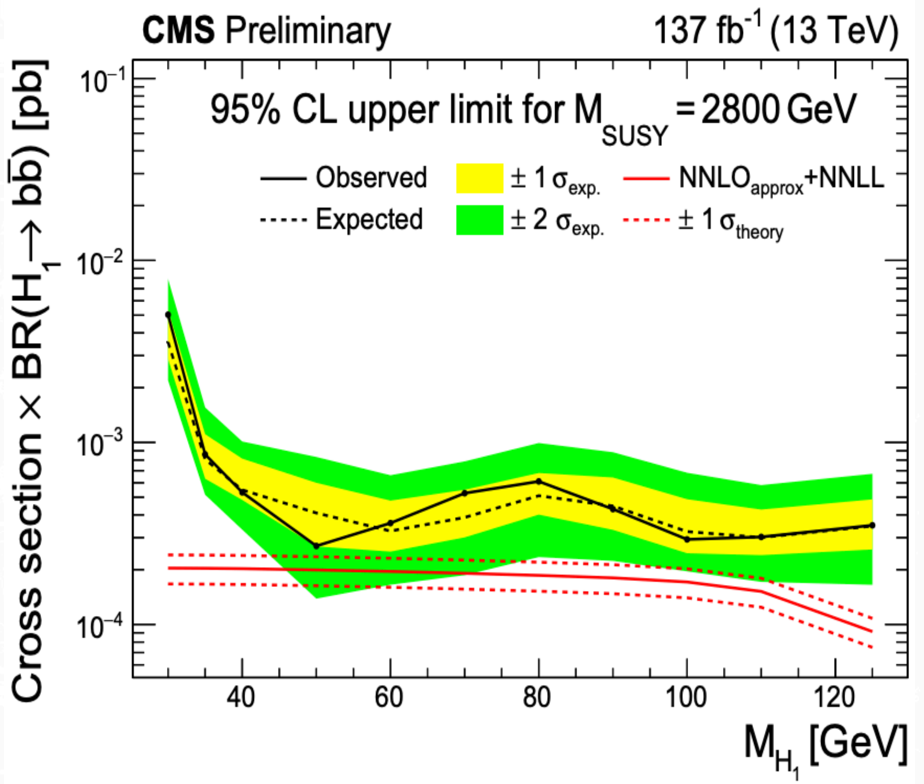
diHiggs in SUSY

HIG-20-018

m_{SUSY} = 2 TeV



m_{SUSY} = 2.6 TeV

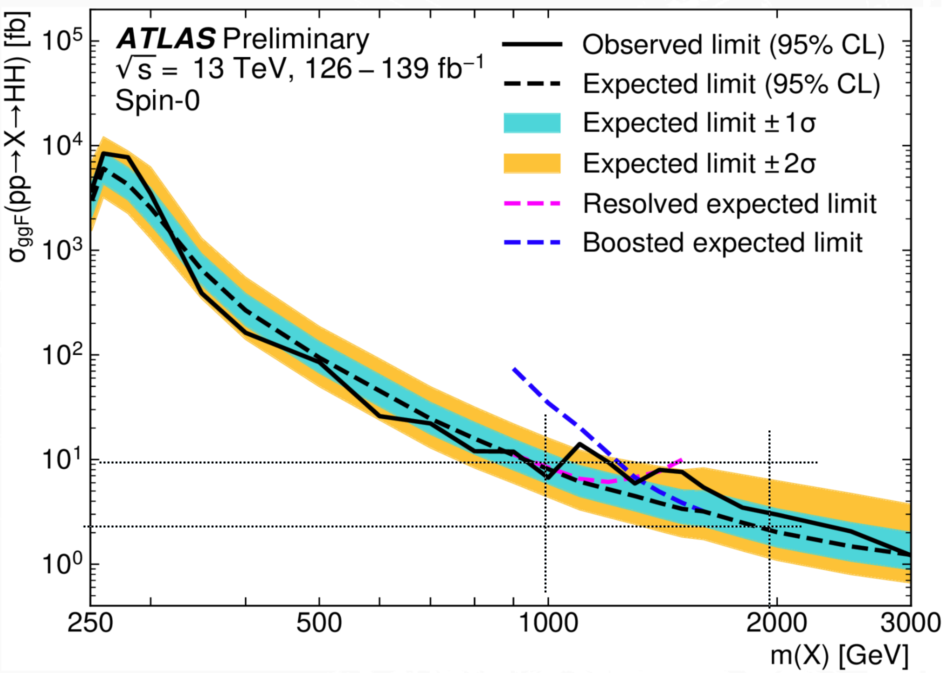


- Signal acceptance ~constant for $40 < m_{H1} < 125$ GeV
  - σ x BR limits ~constant [for given m_{SUSY}]
- Strong drop in signal acceptance for $m_{H1} < 40$GeV

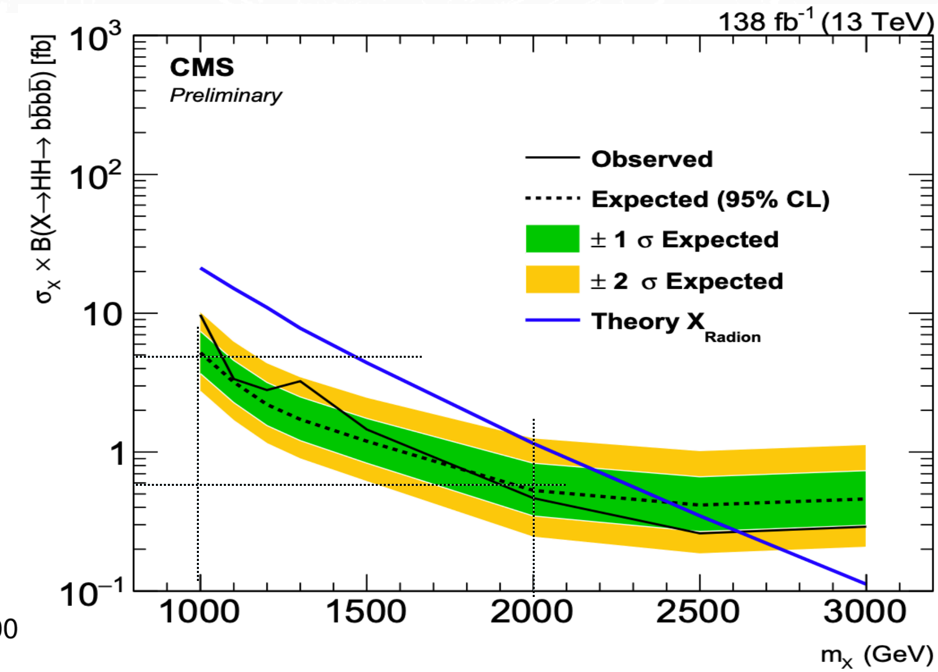
X → HH → 4b: ATLAS vs. CMS

More details:
ATLAS-CONF-2021-035
CMS-B2G-20-004

ATLAS



CMS



$X \to HH \to 2bL(L)$

- **Selection:** ME_T, angular variables b/w ME_T and L(L), m_{LL}

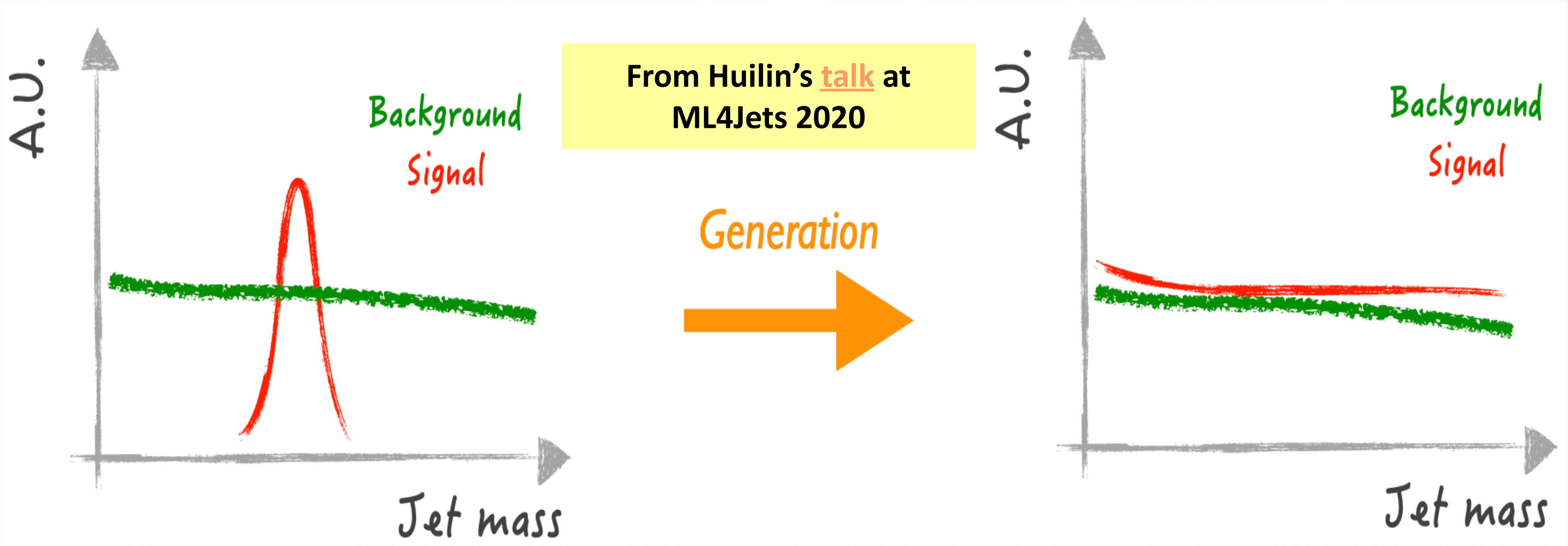
- **BKG estimation:** dominant tt BKG split in 4 categories (top, W, lost t/W, q/g)
  - Build inclusive templates: relax selection to increase stats
  - Modeling methods:
    - **non-resonant:** KDE
    - **resonant:** double-sided CB
  - Fit model accounts for correlations b/w m_{HH} and m_{bb}:

$$P_{\mathrm{bkg}}(m_{\mathrm{b\bar{b}}}, m_{\mathrm{HH}}) = P_{\mathrm{b\bar{b}}}(m_{\mathrm{b\bar{b}}} | m_{\mathrm{HH}}, \theta_1) P_{\mathrm{HH}}(m_{\mathrm{HH}} | \theta_2),$$

- **Validation regions:** invert AK4 jet veto (ttbar CR), low DeepAK8 score (non-bb)

Training details

- **Samples:** Dedicated samples to populate the full mass range
  - equal amount of QCD, X->bb, X->cc, X->qq jets [X: scalar w/ different masses]

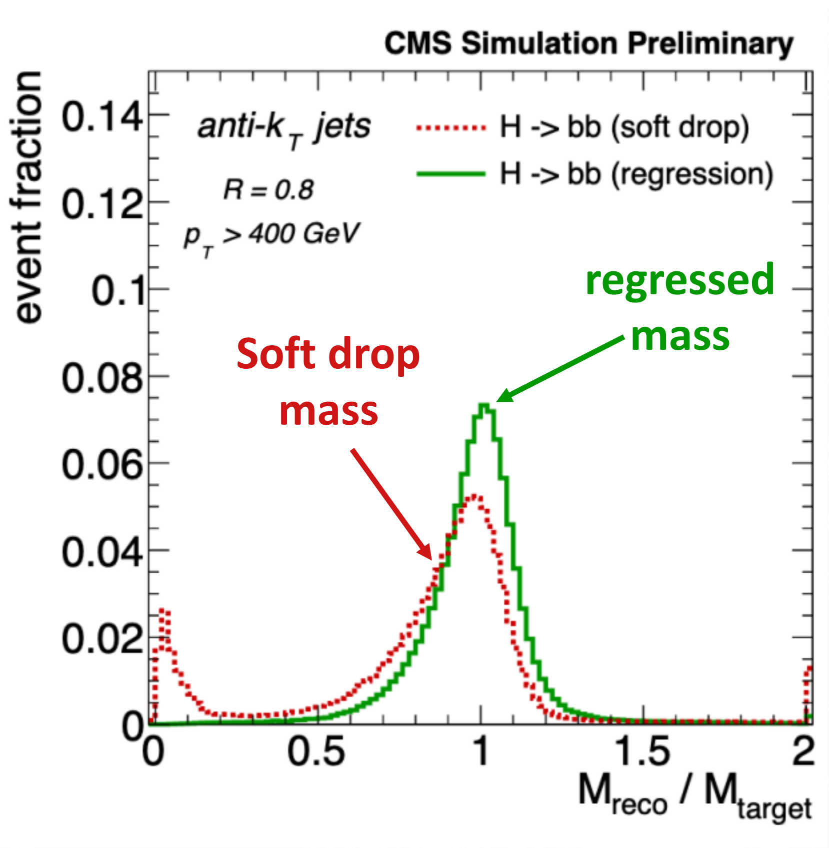


- Target mass:
  - Signal: X pole mass [15-250 GeV]
  - Background: Generated softdrop mass

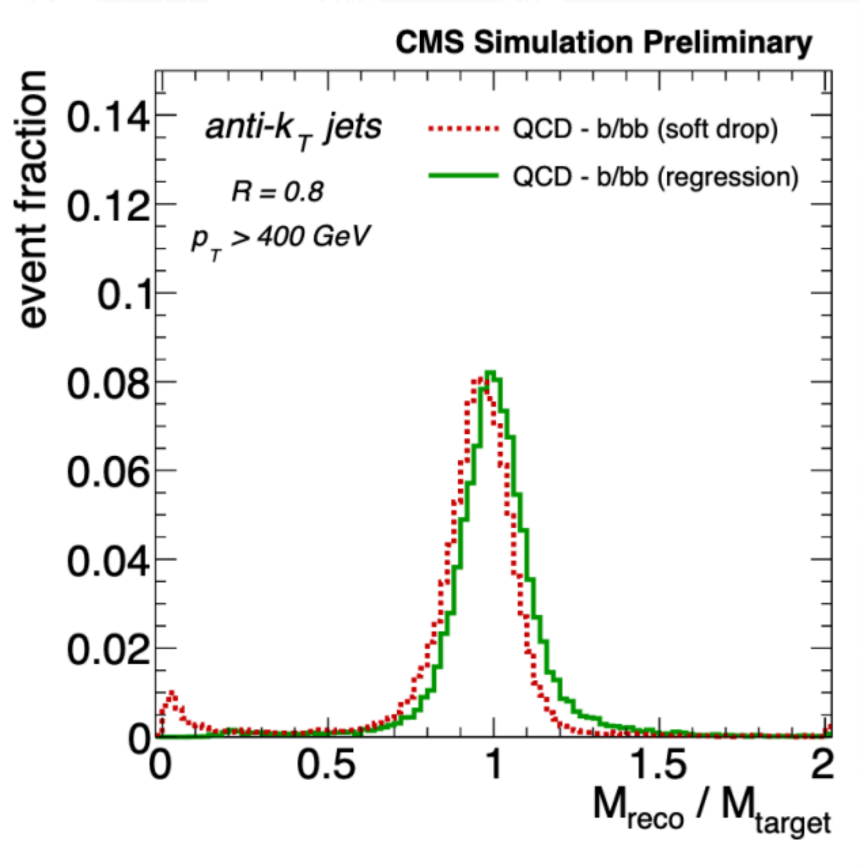
- Loss function: $$L(y, y^p) = \sum_{i=1}^{n} \log(\cosh(y_i^p - y_i))$$

Performance

Signal jets: H->bb



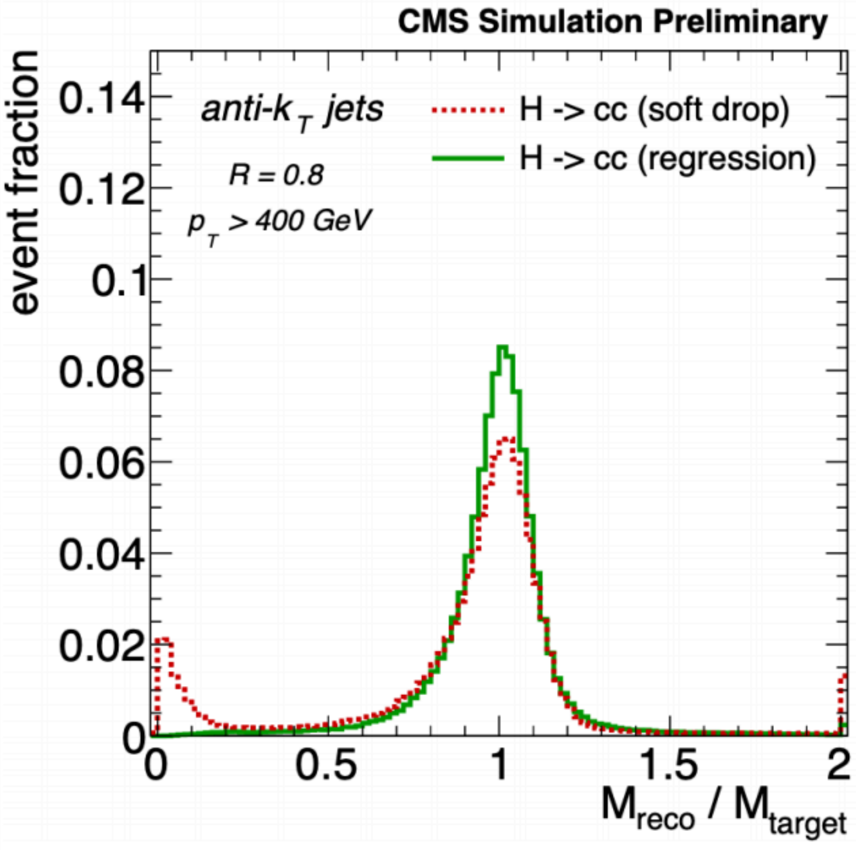
Background jets: QCD



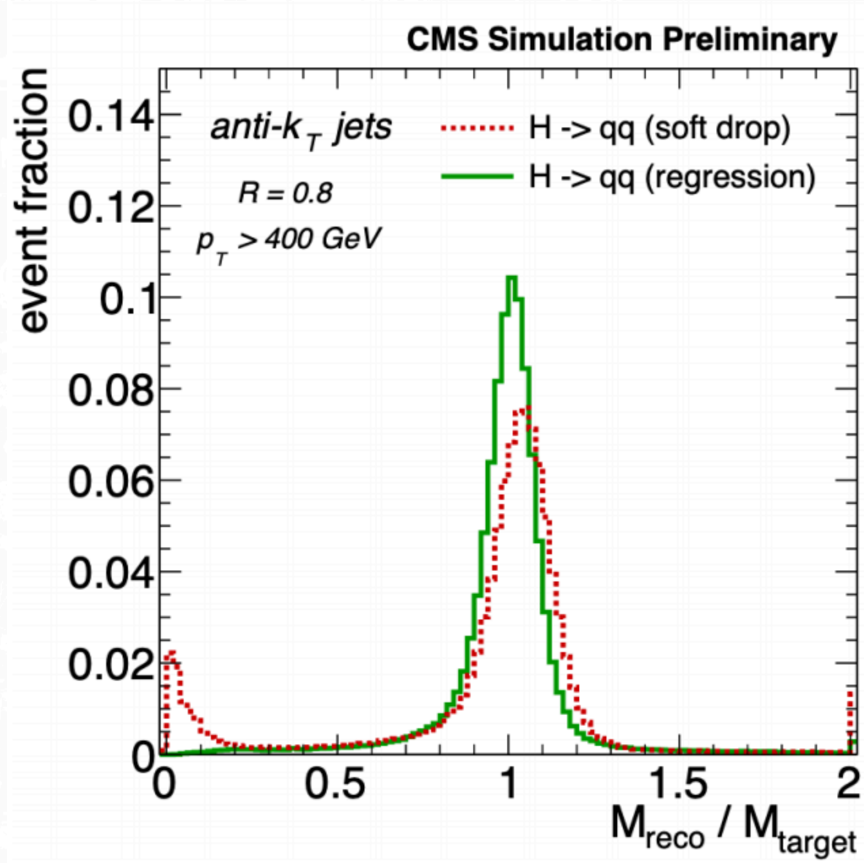
- **Substantial improvement in both mass scale & mass resolution**
- **Tails in m(SD) significantly reduced**

Performance (II)

Signal jets: H->cc



Signal jets: H->qq



- Improvement for all jet flavours

Putting pieces together

- H->bb candidates: two jets with highest bb-discriminant

SRs

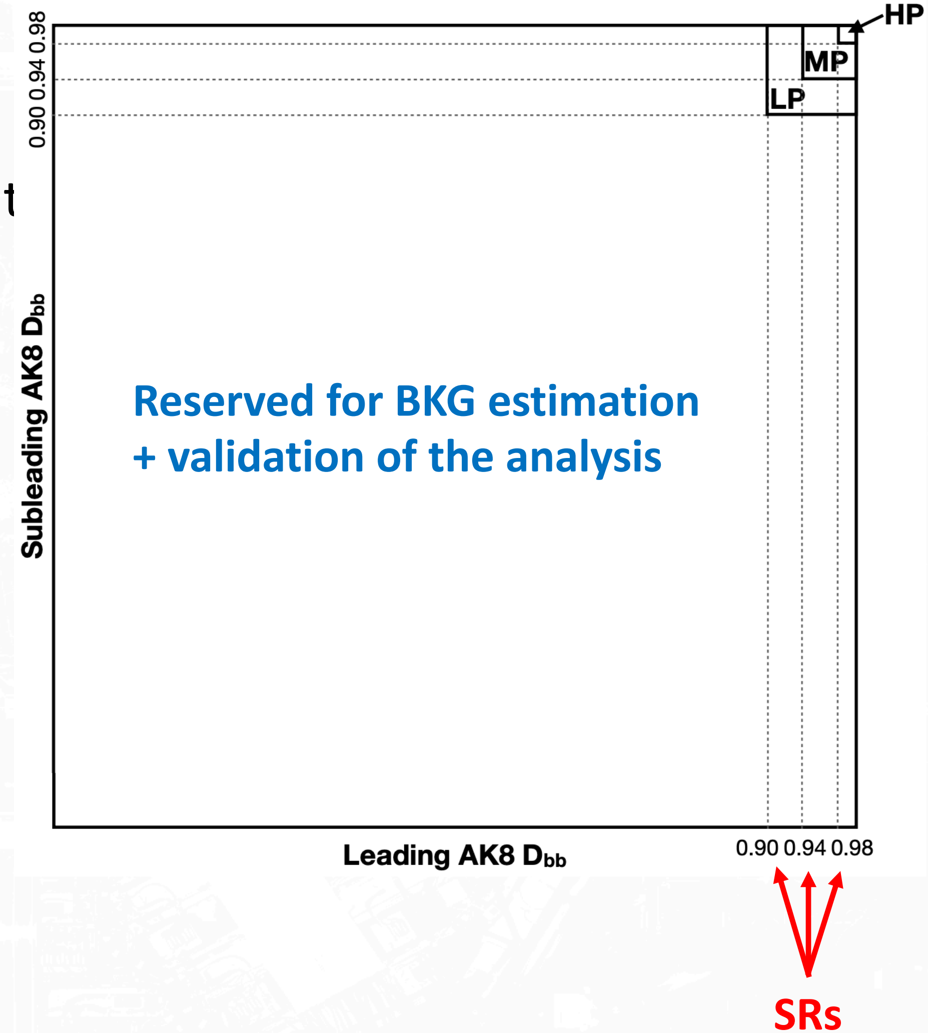
- Analysis carried out in three disjoint categories based on the bb-discriminant of the Higgs candidates
  - Low, medium and high purity

bb-discriminant	>0.90	>0.94	>0.98
ε (H→bb)	90%	80%	60%
ε (QCD)	2%	1%	0.3%

* efficiencies after analysis selection

- Add selection on regressed mass

1st analysis to use ParticleNet
[more analyses in the pipeline]



Background estimation

- Main background from QCD events; estimated directly from data
  - Smaller contribution from ttbar estimated using MC w/ necessary corrections from ttbar dominated control regions (CRs)

- Data-driven QCD estimation relies on an ABCD method
  - Define QCD-enhanced CRs (A, B, C) by inverting ParticleNet score and/or jet mass selection of the subkeading jet
  - Key: jet mass decorrelation of ParticleNet
    - i.e., tagger's response is independent of the jet mass

- Prediction in SR:

$$N^{\mathrm{D}}_{\mathrm{QCD},i} = w_i \times N^{\mathrm{C}}_{\mathrm{QCD},i}$$

  - where TF = $w_i = \frac{N^{\mathrm{B}}_{\mathrm{QCD},i}}{N^{\mathrm{A}}_{\mathrm{QCD},i}}.$

- Full analysis validated in data samples orthogonal to the SRs