Searches for heavy BSM particles resulting in boosted final states at CMS

Xudong Lyu (Peking University)

On behalf of the CMS Collaboration

BOOST 2021

August 2-5, 2021

Introduction

New CMS results related with heavy BSM particles using full Run-2 data:

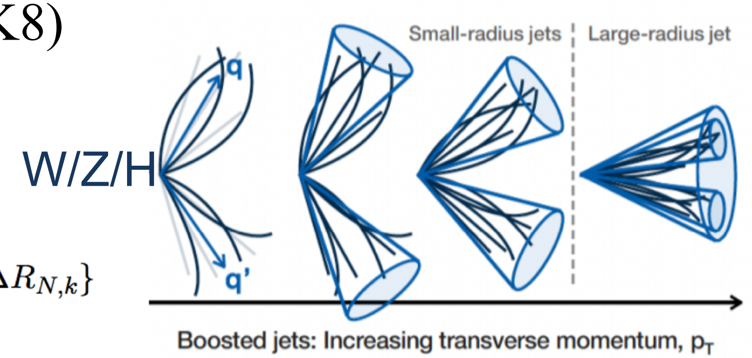
- Part 1: searches for resonances with enhanced couplings to third generation quarks
- Part 2: search for a heavy resonance that decays in cascade though a radion scalar into three bosons

Part 1	**$b^* \to tW \to$ jets** **$b^* \to tW \to$ lepton+jets**	B2G-19-003 CMS-PAS-B2G-20-010
	$W' \to tb$	B2G-20-005
	$W' \to VLQ + t/b$	CMS-PAS-B2G-20-002
	VLQ $T' \to tZ(\nu\nu)$	CMS-PAS-B2G-19-004
	VLQ $B\bar{B}$, $B \to H/Z+b$	B2G-19-005
Part 2	**$W_{KK} \to WR \to WWW \to l\nu$+jets**	CMS-PAS-B2G-20-001

Jet tagging

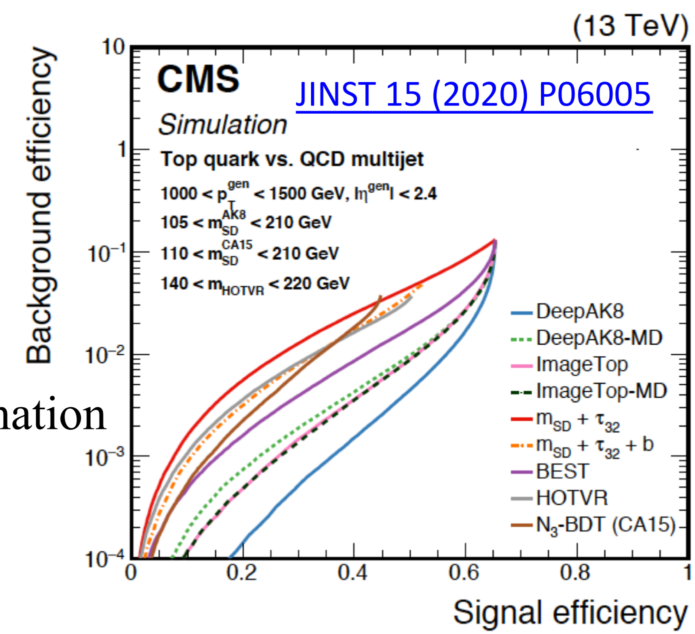
Boosted objects → small angular separation → merged jets (W/Z → qq; H → bb)

- Jets: anti-kt with R = 0.4 (AK4) or R = 0.8 (AK8)
- Jet grooming, soft drop mass



Techniques for a successful boosted analysis:

- N-subjettiness $\tau_N = \frac{1}{d_0}\sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \cdots, \Delta R_{N,k}\}$
  - Designing decorrelated taggers (DDT)
- HOTVR
  - new jet clustering technique with a variable R
- Double-b tagger
  - discriminate H → bb decays
- ImageTop
  - CNN training based on pixelized images using jet information
- DeepAK8 tagger
  - multi-class tagger for t/W/Z/H tagging
  - use PF candidates and secondary vertices

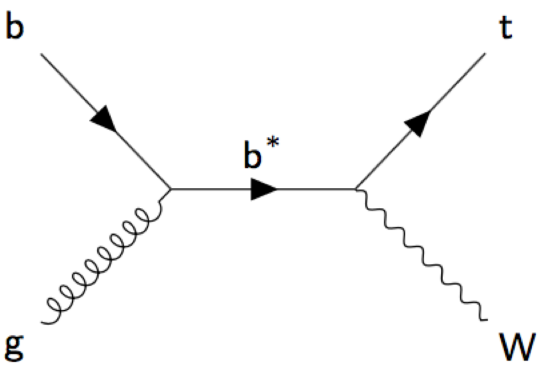


b*→ tW

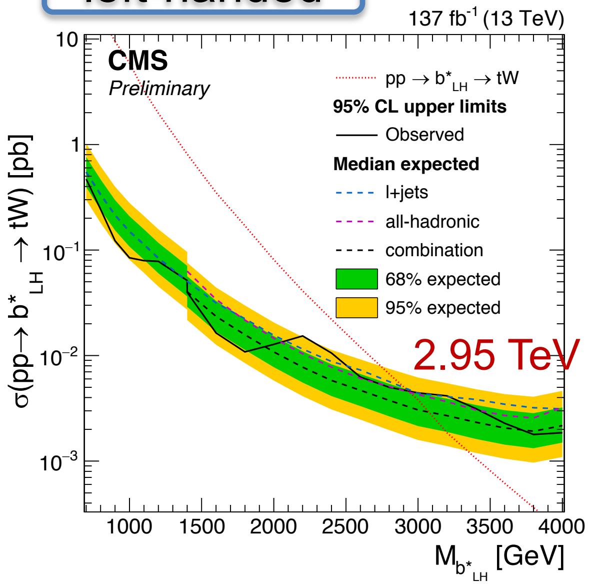
CMS B2G-19-003

CMS-PAS-B2G-20-010

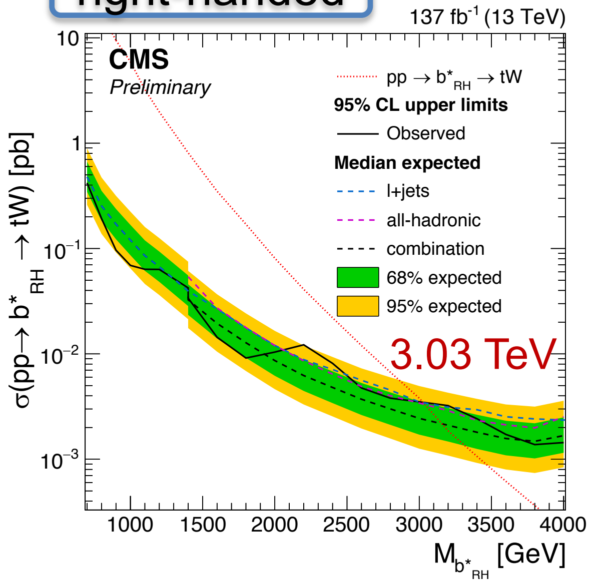
- Combination of all-hadronic and single-lepton final states
- Reconstruction of the tW system
  - Hadronic top: N-subjettiness and HOTVR tagging
  - W boson:
    - N-subjettiness, SD mass (all-hadronic)
    - reconstruct from lepton and MET (single-lepton)
- Most stringent limit on excited b quarks



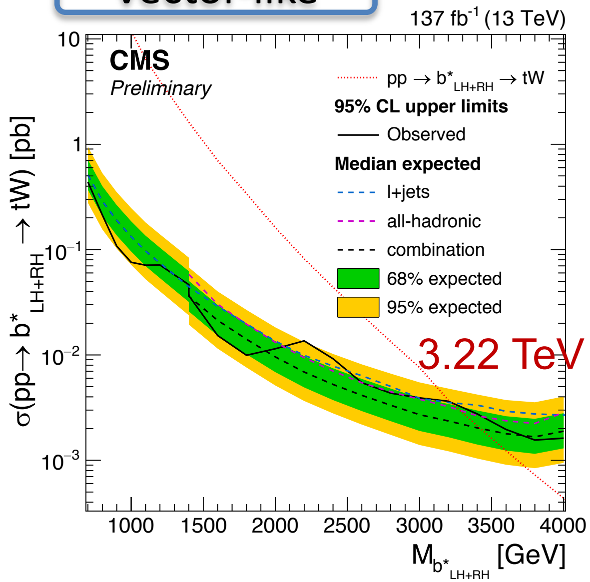
left-handed



right-handed



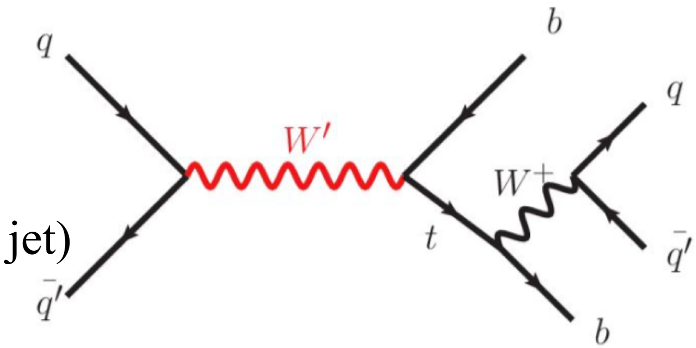
Vector-like



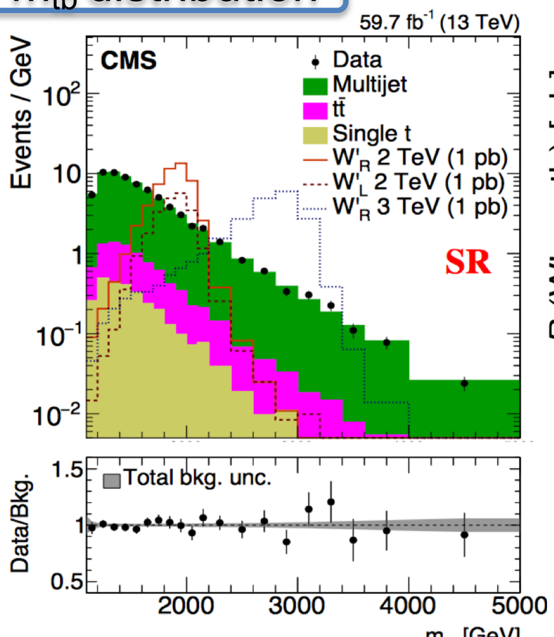
W'→ tb in full-had mode

CMS B2G-20-005

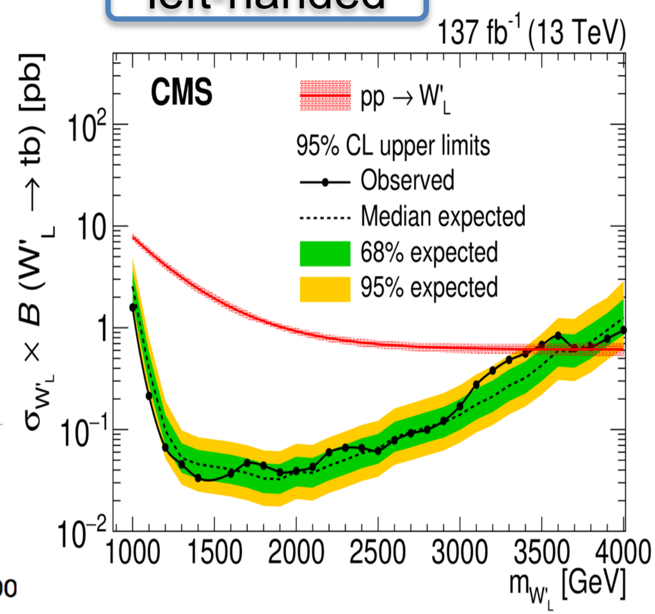
- Hunt for a bump in m_{tb} spectrum
- Reconstruction of the tb system
  - b candidate: DeepJet tagger (AK4 jet)
  - t candidate: MD DeepAK8 tagger, SD mass (AK8 jet)
- Dominant background: QCD multi-jet production, estimated from data in jet mass sideband using b tagging pass/fail ratio
- The most stringent limit to date: W' bosons with masses below 3.4 TeV are excluded at 95% CL



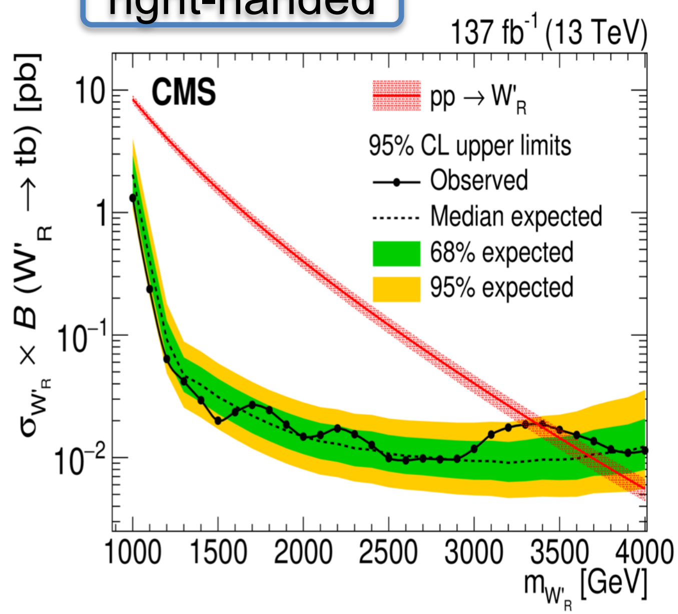
m_{tb} distribution



left-handed



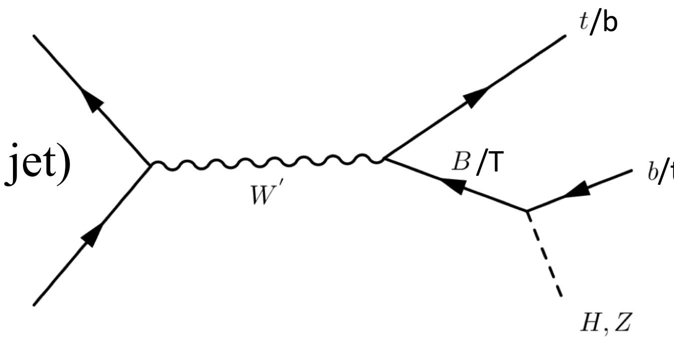
right-handed



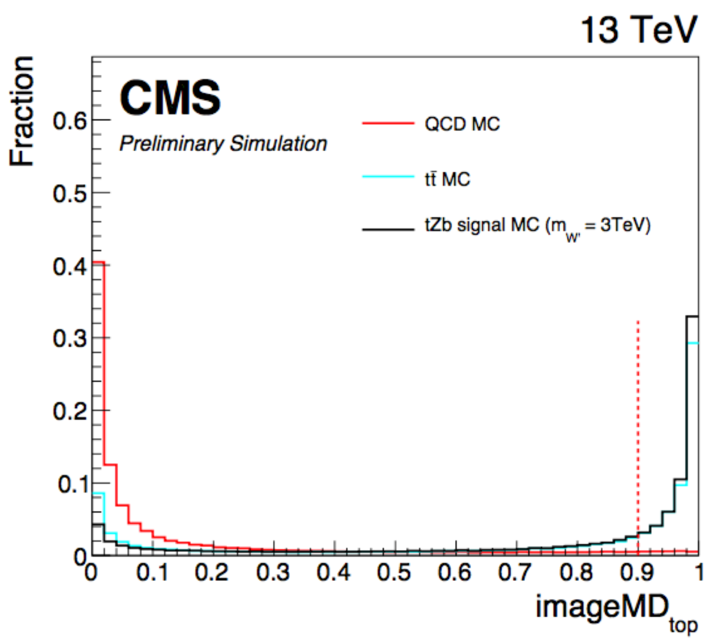
W'→VLQ + t/b

CMS-PAS-B2G-20-002

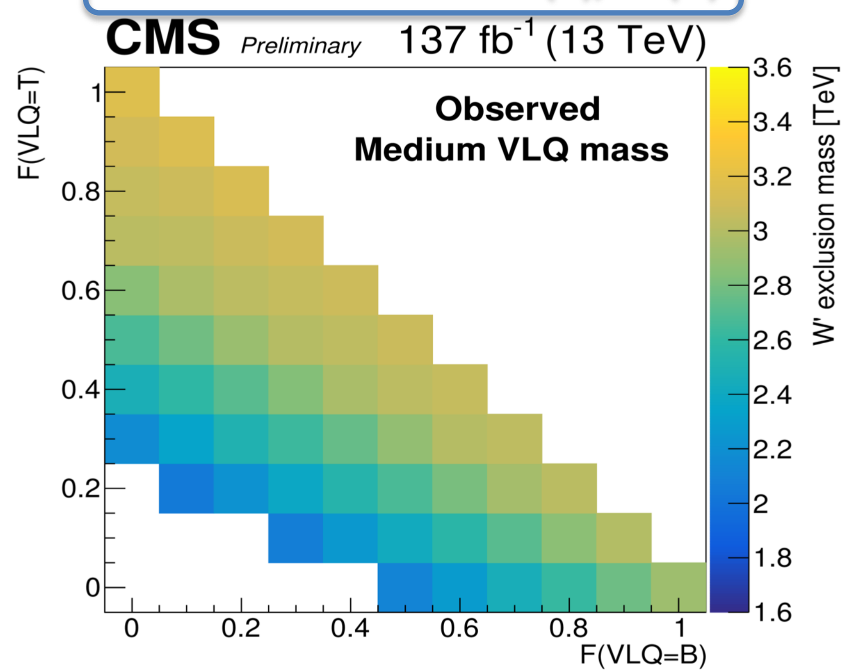
- Signal: W' → Tb/Bt → tHb/tZb, and full hadronic channel
- Trijet topology:
  - Top: imageMD$_{TOP}$ (first application), SD mass (AK8 jet)
  - Z/H: τ_{21}/Double-b tagger, SD mass (AK8 jet)
  - b quark: DeepFlavour b tag (AK4 jet)



- Dominant background: QCD multi-jet production, estimated from data using alphabet method



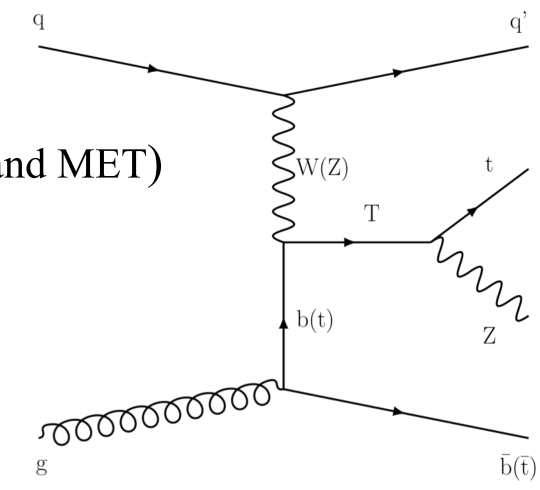
Limit as function of BR(T)/BR(B)



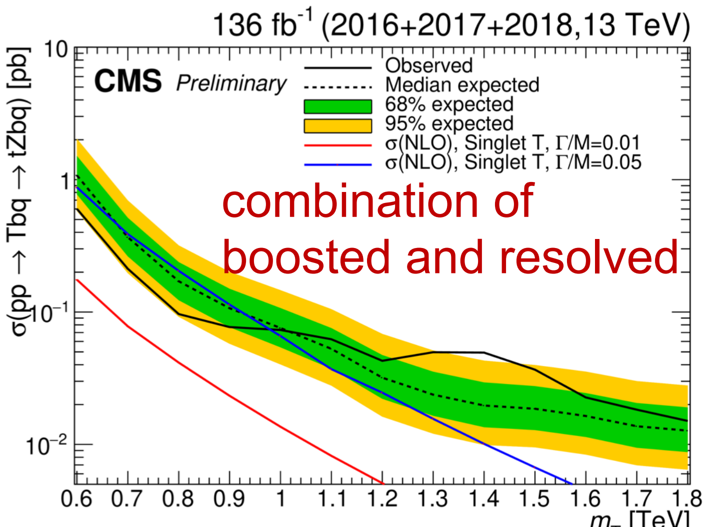
VLQ T'→ tZ →(bqq)(νν)

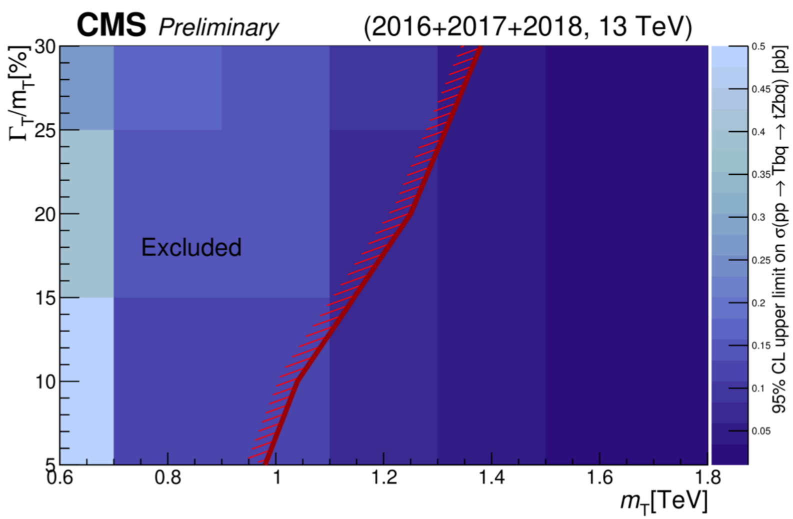
CMS-PAS-B2G-19-004

- First result of single production of VLQ in tZ (top hadronic, Z to neutrinos) with the full Run-II
- Reconstruction of the tZ system (Observable: M_T between top and MET)
  - Z to neutrinos : MET
  - hadronic top: (N-subjettiness, τ_{21}^{DDT}, τ_{32})
    - fully-resolved: three AK4 jets
    - partially-merged: a W-jet and a b-jet
    - fully-merged: one top-jet



- Background (W/Z+jets, ttbar): taken from MC with corrections estimated from data
- Limits for masses and widths of T' and excluded region in the width-mass plane

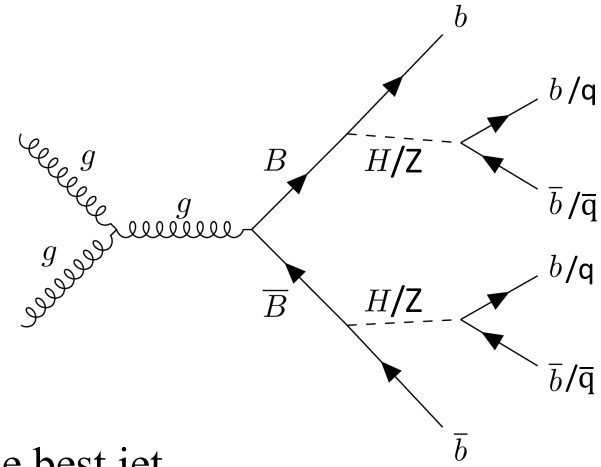




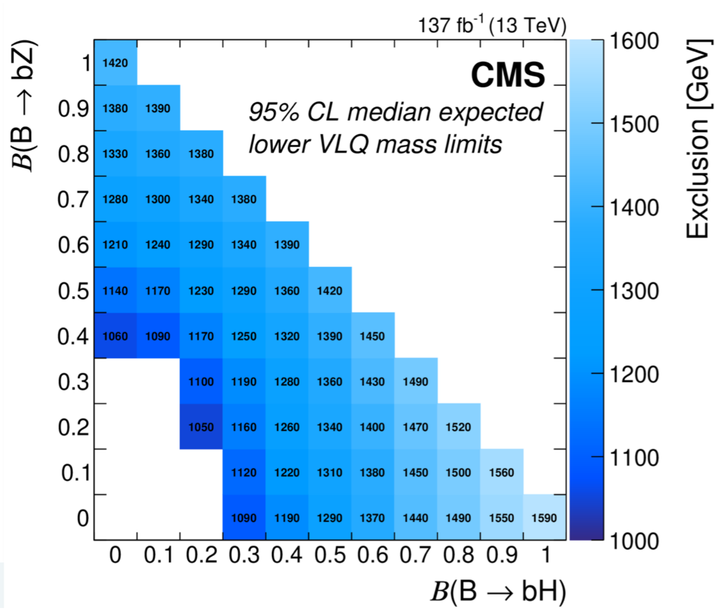
VLQ $B\bar{B}$ production (full-had)

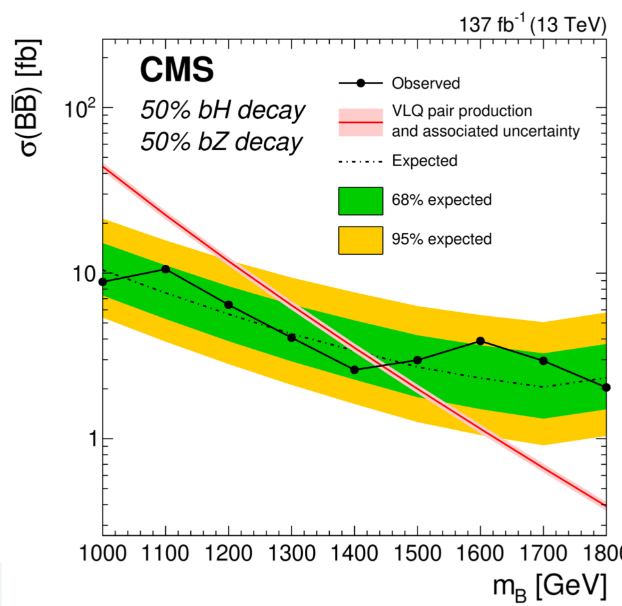
CMS B2G-19-005

- Bottom-like VLQ pair production in the full hadronic mode
- B →bH/bZ
- 4, 5, or 6 jets depending on the number of merged H/Z
- Object reconstruction:
  - Individual jets: DeepJet b tagger applied to AK4 jets
  - Merged jets from bb pairs: Double-b tagger (AK8 jets)
  - Apply a jet multiplicity dependent $\chi 2$ metric to determine the best jet combination and reconstruct the event



- Determine final exclusion limits as a function of B → bH and B → bZ branching fractions

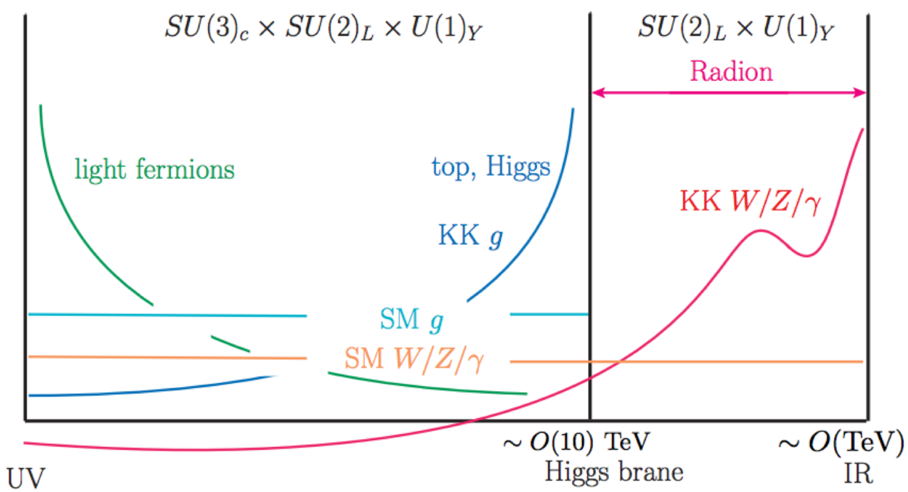


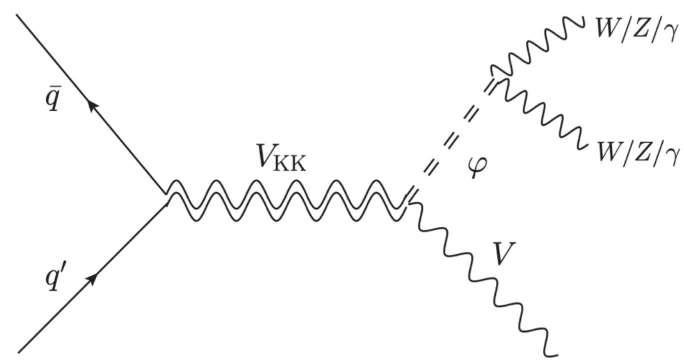


$W_{KK} \to WR \to WWW \to l\nu$+jets

CMS-PAS-B2G-20-001

- First search for resonances decaying in cascade to final state with 3 W bosons
- Extended Warped ED model :
  - Extra brane by splitting the Bulk (arXiv:1612.00047 , arXiv:1711.09920 , arXiv:1809.07334);
  - Various fields propagate in different regions
  - "di-SM" suppressed in favor of "tri-SM"
  - Only EW bosons in the extended bulk → dominant: $V_{KK} \to R\,V \to VVV$
    - – WWW having the largest contribution





$W_{KK}\to$WWW: Selection

CMS-PAS-B2G-20-001

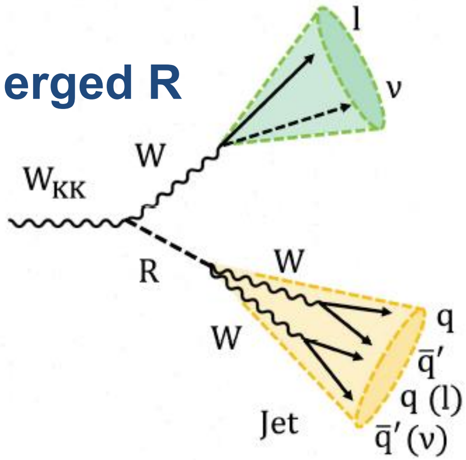
- 1-lepton channel with BR of 42% is explored: $W_{KK} \to WWW \to l + \nu + jets$

- Split to 6 signal regions based on:
  - Merged: (SR1-3)
    - single massive large-radius jet
    - Bin over M_R: 60-100, 100-200, >200 GeV
    - For $60 < M_j < 100$ GeV, use deep-W
    - For Mj > 100 GeV, use deep-WH to tag radion

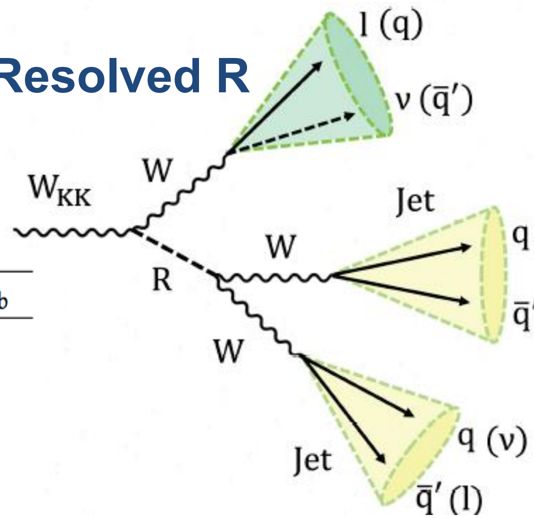
$$\text{deep-WH} = \frac{W_{qq,qc} + H_{4q}}{QCD_{g,q,b,...} + W_{qq,qc} + H_{4q}}$$

  - Resolved: (SR4-6)
    - 2 jets, ordered due to mass
      M_j^{max}: 60-100 GeV
      M_j^{min}:0-60-100 GeV binning

Merged R



Resolved R

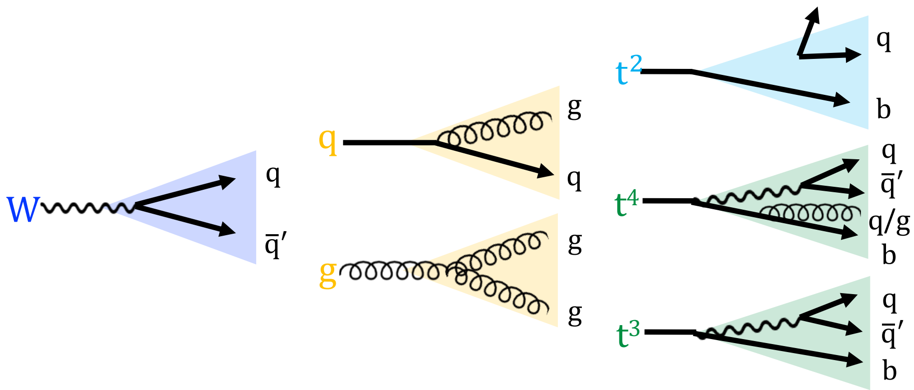


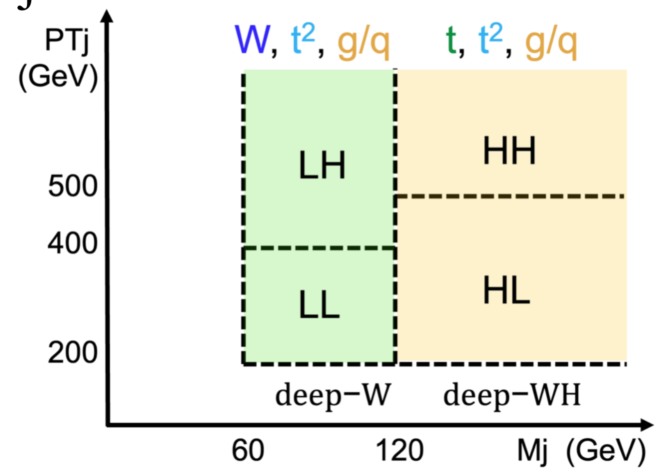
Region	m_j^{max} [GeV]	taggers	m_j^{min} [GeV]	tagger	N_j^{AK8}	N_j^{AK4}	N_b
SR1	60–100	deep-W > 0.7	—	—	1	≤ 2	0
SR2	100–200	deep-WH > 0.7	—	—	1	≤ 2	0
SR3	≥ 200	deep-WH > 0.7	—	—	1	≤ 2	0
SR4	60–100	deep-W > 0.5	60–100	deep-W > 0.5	2	≤ 2	0
SR5	60–100	deep-W >(<) 0.5	60–100	deep-W <(>) 0.5	2	≤ 2	0
SR6	60–100	deep-W > 0.7	0–60	—	2	≤ 2	0

DeepAK8 tagger calibration

CMS-PAS-B2G-20-001

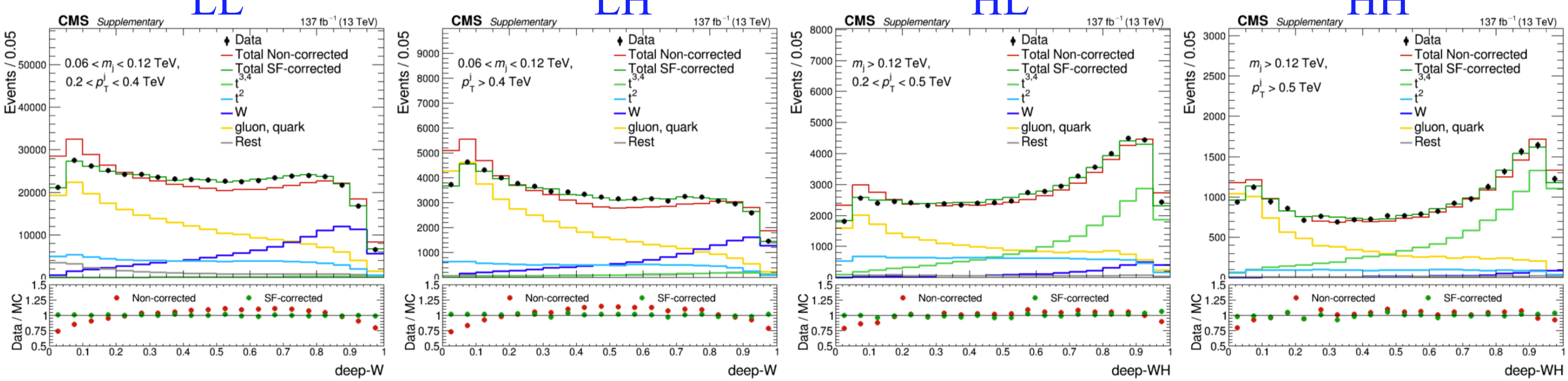
- Calibrate deep tagger discriminant shape using SM objects:





- Apply parton-level matching and break down into 4 samples (LL, LH, HL, HH)

LL LH HL HH

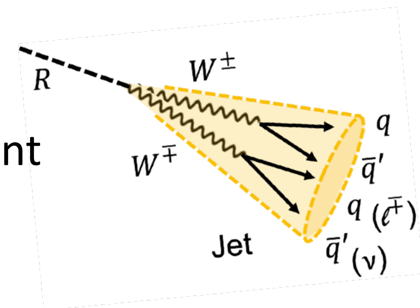


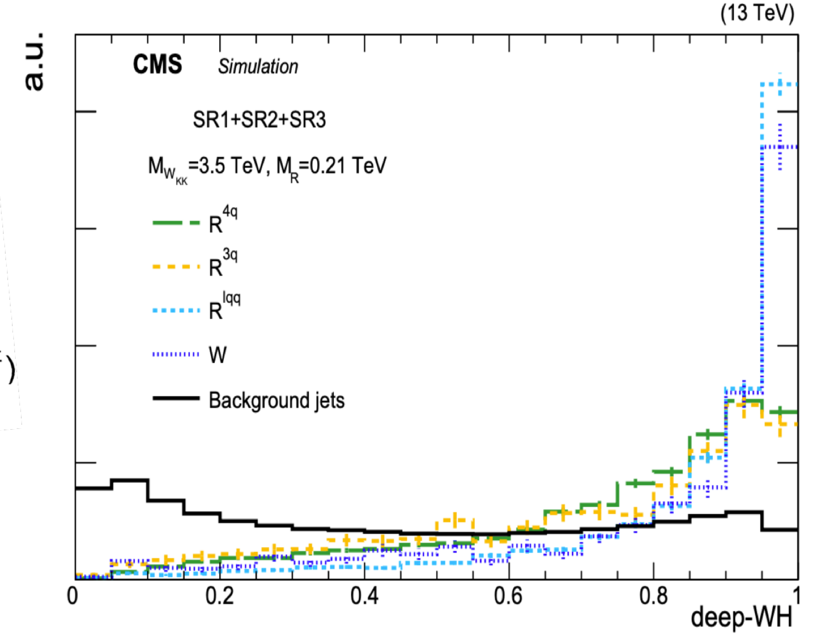
- Split each sample into 3 pure subsets and correct MC shapes bin by bin to derive scale factors for each SM object (details in backup)

Calibration for signal

CMS-PAS-B2G-20-001

- Merged Radion jet≈ $R^{4q} + R^{3q} + R^{lqq}$
- no standard candle in SM
  - special calibration treatment





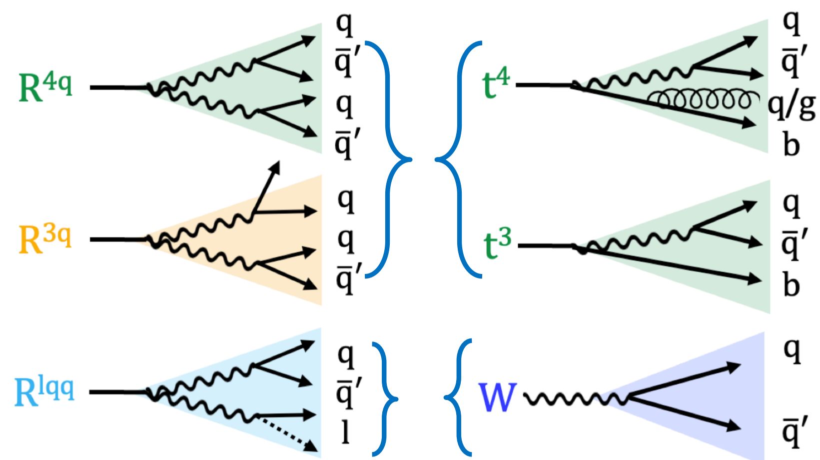
1. Observe similarity between $R^{4q} \leftrightarrow R^{3q}$ jets with merged top: $t^{3,4}$
   → we apply SF($t^{3,4}$) on R^{4q}, R^{3q}

2. Observe similarity between $W \leftrightarrow R^{lqq}$ jets
   → we apply scale factors for W, SF(W), on R^{lqq}

3. The difference between the performances of the SM candle and signal is taken into account as the systematic uncertainty.

R components

SM proxy jet

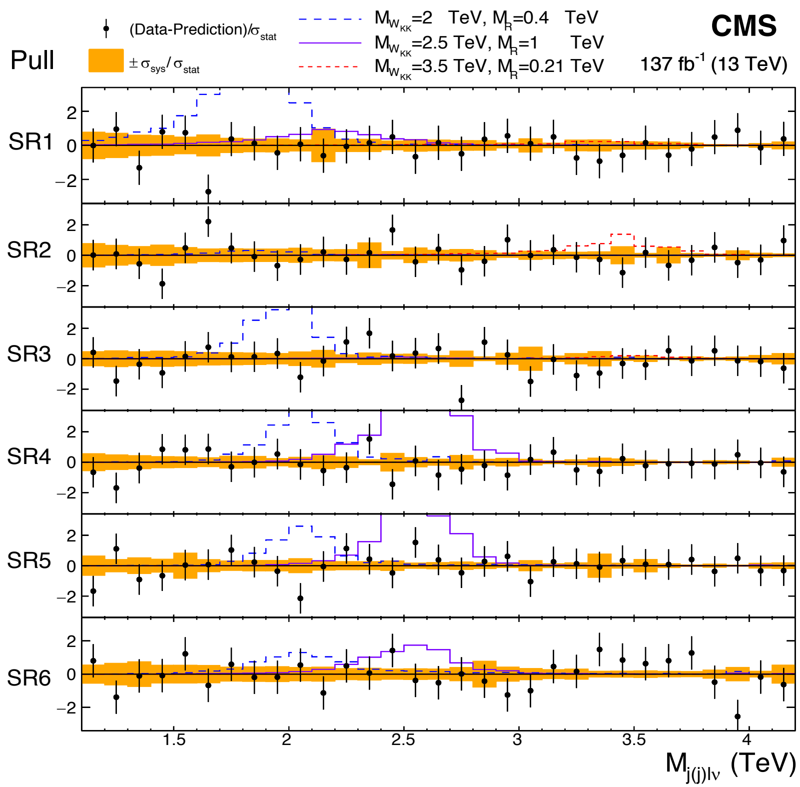


Results and limit

CMS-PAS-B2G-20-001

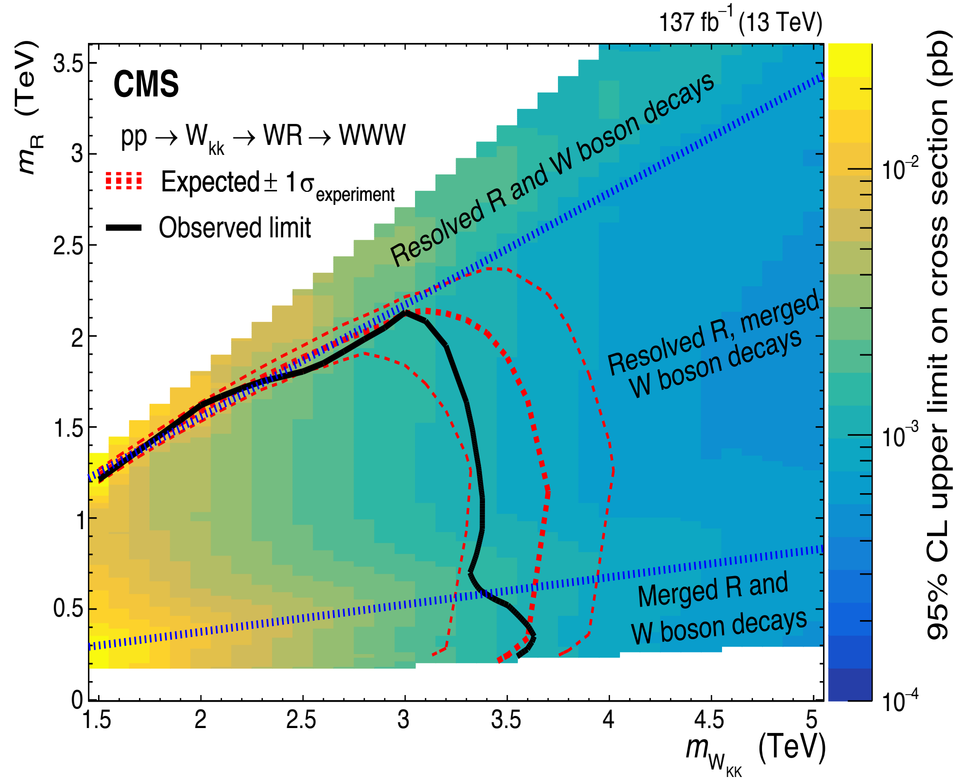
- Combined fit of six signal regions.

(No excess over the background estimation is observed.)



- Limits in 2D W_{KK} vs. R mass plane.

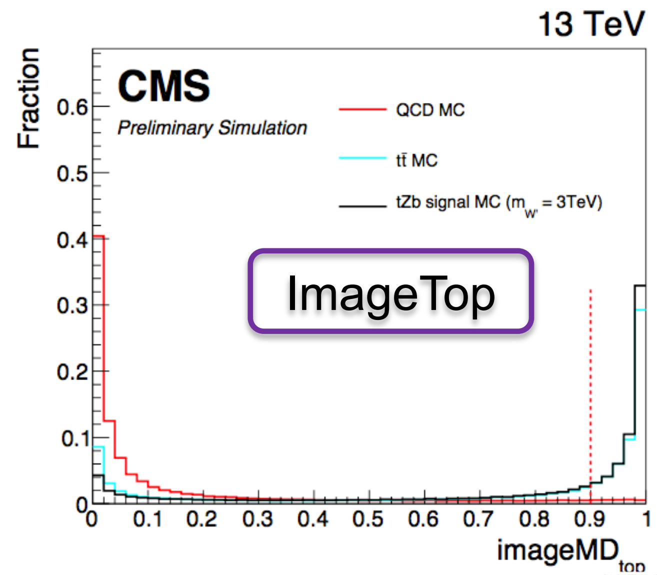
The first of their kind!

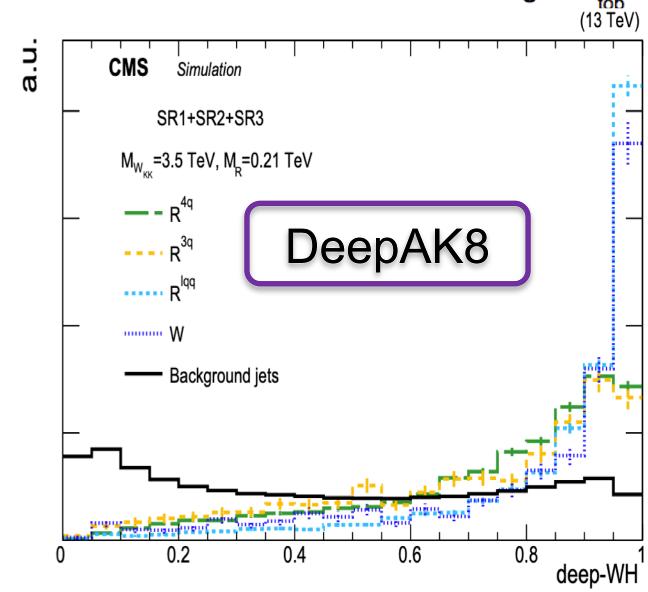


The triboson resonances are excluded up to m_{WKK} = 3.4 (3.6) TeV for m_R = 1 (0.35) TeV.

Summary

- Boosted jet physics can be probed efficiently due to the developments of various ML taggers.
- A wealth of new physics signatures with massive particles can be probed with boosted jets.
- No evidence for new physics observed yet, 95% CL limits are set.
- More results to come out, and follow also here to keep track!
    - [CMS publications]()



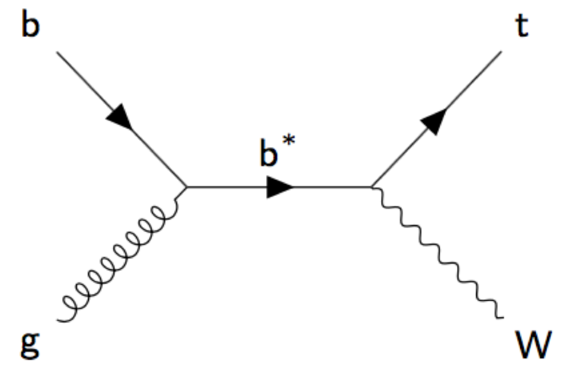


BACKUP

b*→ tW

CMS B2G-19-003

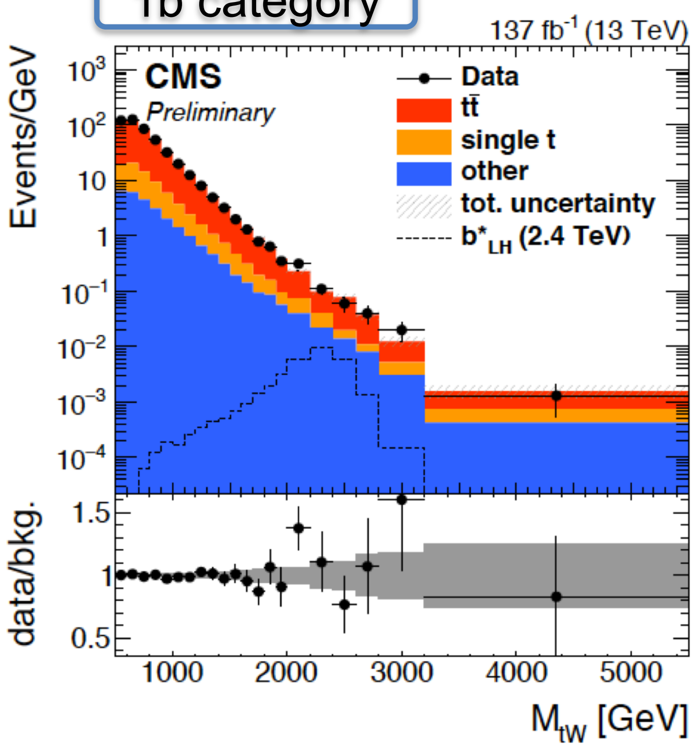
CMS-PAS-B2G-20-010



Uncertainties affecting both the l+jets and all-hadronic channels are marked by an asterisk (∗)

Source	Uncertainty	Samples	Impact (up/down)
$t\bar{t}$ cross section*	± 20%	$t\bar{t}$	+5.1 / -5.0%
Single tcross section*	± 30%	single t	-4.2 / +4.8%
Luminosity*	± 1.8%	$t\bar{t}$, single t, signal	-1.9 / +1.8%
Top p_T reweighting*	Shape	$t\bar{t}$	-5.0 / +5.1%
PDF*	Shape	$t\bar{t}$, single t, signal	-4.1 / +4.6%
Background estimation (1b)	Shape	non-top (from data)	-5.3 / +6.9%
Background estimation (2b)	Shape	non-top (from data)	-0.3 / -0.4%
Pileup*	Shape	$t\bar{t}$, single t, signal	-0.4 / +0.5%
JEC*	Shape	$t\bar{t}$, single t, signal	-1.3 / +2.3%
JER*	Shape	$t\bar{t}$, single t, signal	+0.0 / +0.4%
Trigger prefiring	Shape	$t\bar{t}$, single t, signal	+0.1 / -0.0%
Muon identification	Shape	$t\bar{t}$, single t, signal	+0.4 / -0.1%
Muon isolation	Shape	$t\bar{t}$, single t, signal	+0.3 / -0.1%
Muon trigger	Shape	$t\bar{t}$, single t, signal	-0.1 / +0.4%
Electron identification	Shape	$t\bar{t}$, single t, signal	+0.4 / -0.4%
Electron reconstruction	Shape	$t\bar{t}$, single t, signal	+0.2 / -0.0%
Electron trigger	Shape	$t\bar{t}$, single t, signal	+0.3 / -0.0%
ttagging (fully merged)	Shape	$t\bar{t}$, single t, signal	-1.2 / +1.5%
ttagging (partially merged)	Shape	$t\bar{t}$, single t, signal	-0.7 / +0.8%
ttagging (non-merged)	Shape	$t\bar{t}$, single t, signal	-0.0 / +0.2%
btagging (b,c)	Shape	$t\bar{t}$, single t, signal	-3.6 / +4.0%
btagging (u,d,s,g)	Shape	$t\bar{t}$, single t, signal	+0.7 / -0.6%

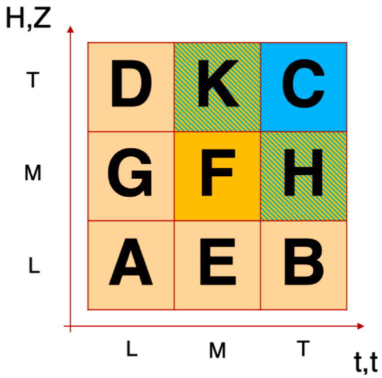
1b category



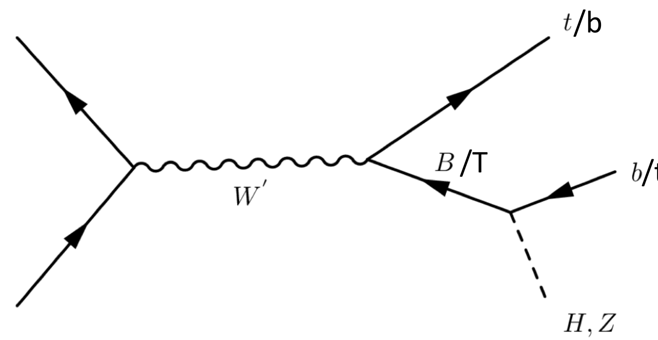
W'→VLQ + t/b

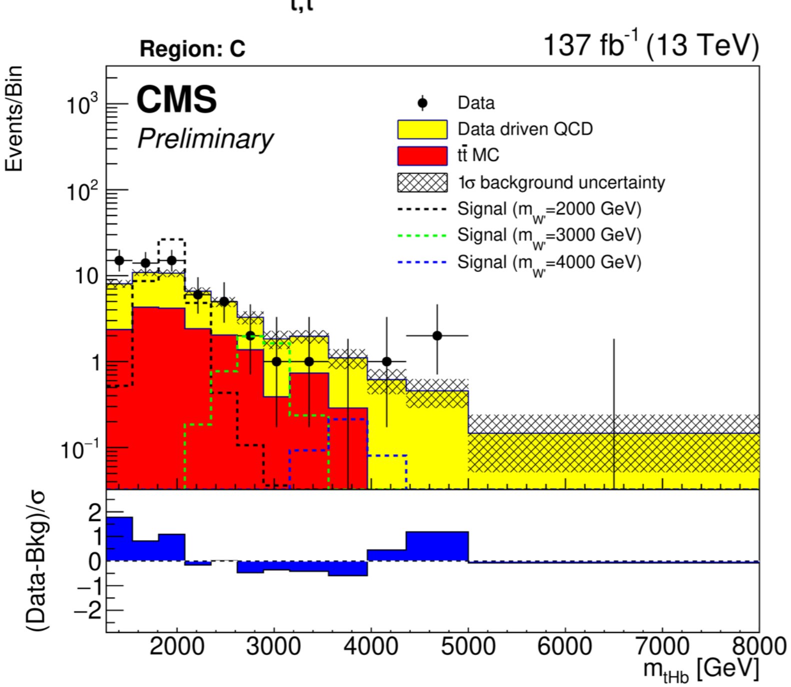
CMS-PAS-B2G-20-002

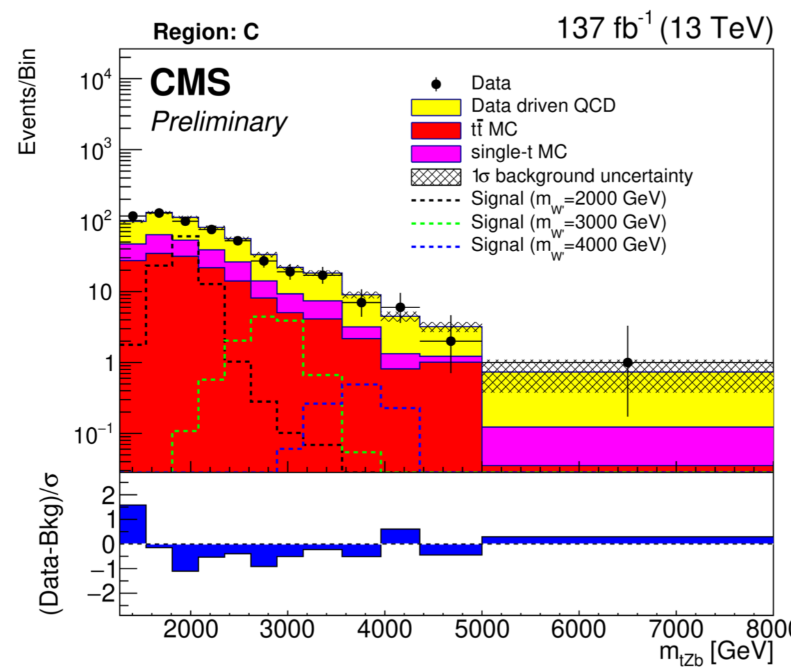
- Dominant background: QCD multi-jet production, estimated from data using alphabet method



$$TF(p_T,\eta) \equiv (B_{data} - B_{t\bar{t}})/(A_{data} - A_{t\bar{t}}),$$
$$TF_v(p_T,\eta) \equiv (E_{data} - E_{t\bar{t}})/(A_{data} - A_{t\bar{t}}),$$
$$C_{qcd} \simeq (D_{data} - D_{t\bar{t}}) \times TF(p_T,\eta),$$
$$H_{qcd} \simeq (G_{data} - G_{t\bar{t}}) \times TF(p_T,\eta),$$
$$K_{qcd} \simeq (D_{data} - D_{t\bar{t}}) \times TF_v(p_T,\eta),$$
$$F_{qcd} \simeq (G_{data} - G_{t\bar{t}}) \times TF_v(p_T,\eta).$$

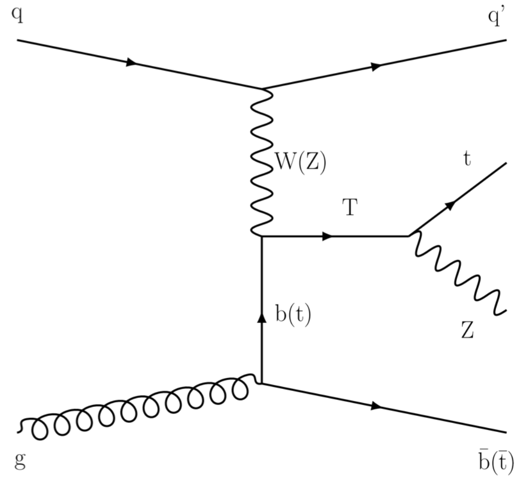




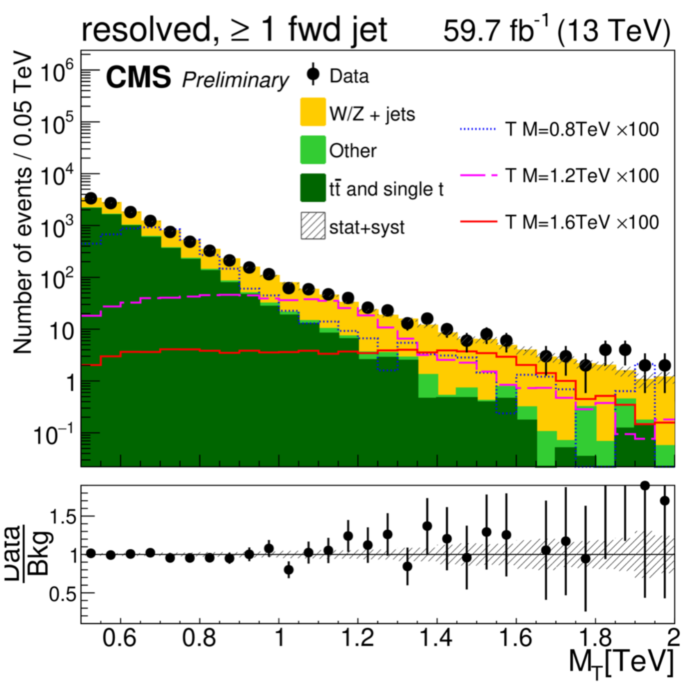


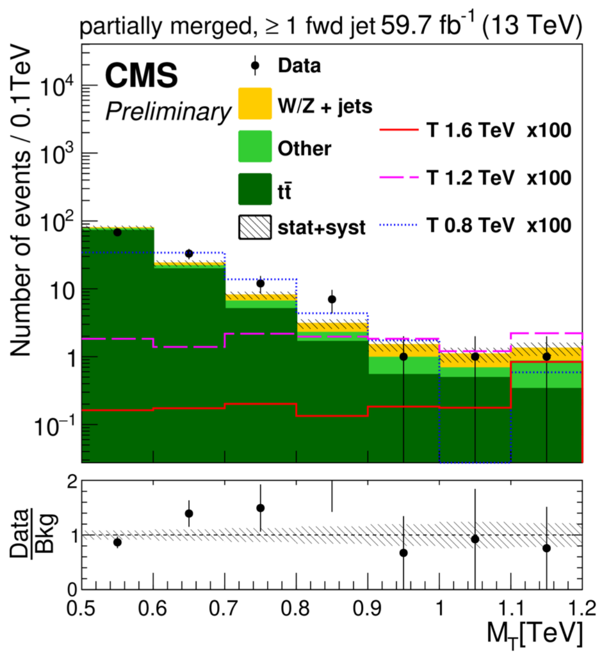
VLQ T'→ tZ →(bqq)(νν)

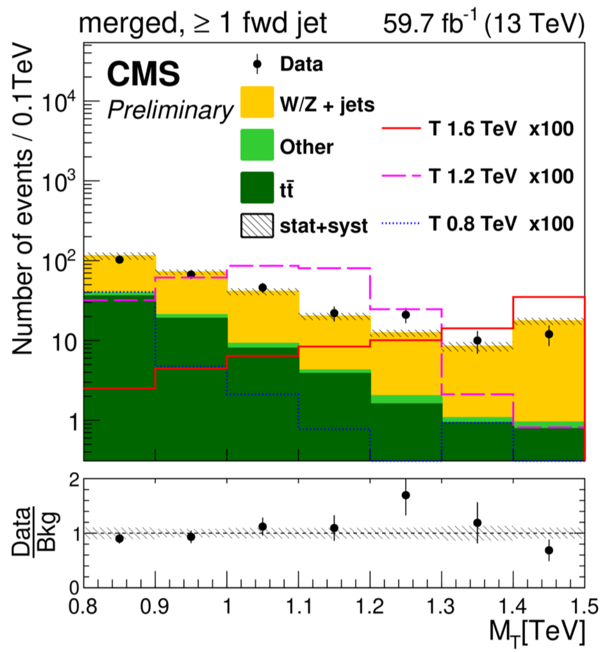
CMS-PAS-B2G-19-004



- First result of single production of VLQ in tZ
- hadronic top:
  - fully-resolved: three AK4 jets
  - partially-merged: a W-jet and a b-jet
  - fully-merged: one top-jet
- Background (W/Z+jets, ttbar): taken from MC with corrections estimated from data







VLQ $B\bar{B}$ production (full-had)

CMS B2G-19-005

- In the 4 & 5 multiplicity analyses, we have merged H/Z jets
  - Match potential merged boson candidate AK4 jets with AK8 jets within $\Delta R < 0.3$.
  - To avoid overlap, discard the AK8 candidate if there is an additional AK4 match within $\Delta R < 0.6$.
  - Additionally disregard low mass AK8 jets (< 50 GeV) to avoid low mass matches.
  - In the $\chi 2$ calculation, the mass of the merged candidate is taken to be the AK8 softdrop mass.
  - Disregard events that do not have a minimum number of AK8 matches (2 for 4 jets, 1 for 5 jets, 0 for 6 jets)

- Use a $\chi 2$ metric to select the correct jet combination and event mode.

$$\text{6-jet events}: \chi^2_{\text{mod}} = \frac{(m_{\text{dijet}_1} - \overline{m}_{\text{dijet}})^2}{\sigma^2_{m_{\text{dijet}}}} + \frac{(m_{\text{dijet}_2} - \overline{m}_{\text{dijet}})^2}{\sigma^2_{m_{\text{dijet}}}} + \frac{(\Delta m_{\text{VLQ}} - \overline{\Delta m}_{\text{VLQ}})^2}{\sigma^2_{\Delta m_{\text{VLQ}}}},$$

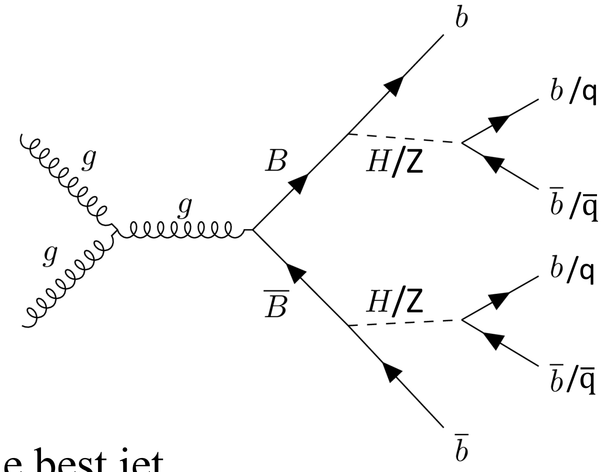
$$\text{5-jet events}: \chi^2_{\text{mod}} = \frac{(m_{\text{dijet}} - \overline{m}_{\text{dijet}})^2}{\sigma^2_{m_{\text{dijet}}}} + \frac{(m_{\text{merged}} - \overline{m}_{\text{merged}})^2}{\sigma^2_{m_{\text{merged}}}} + \frac{(\Delta m_{\text{VLQ}} - \overline{\Delta m}_{\text{VLQ}})^2}{\sigma^2_{\Delta m_{\text{VLQ}}}},$$

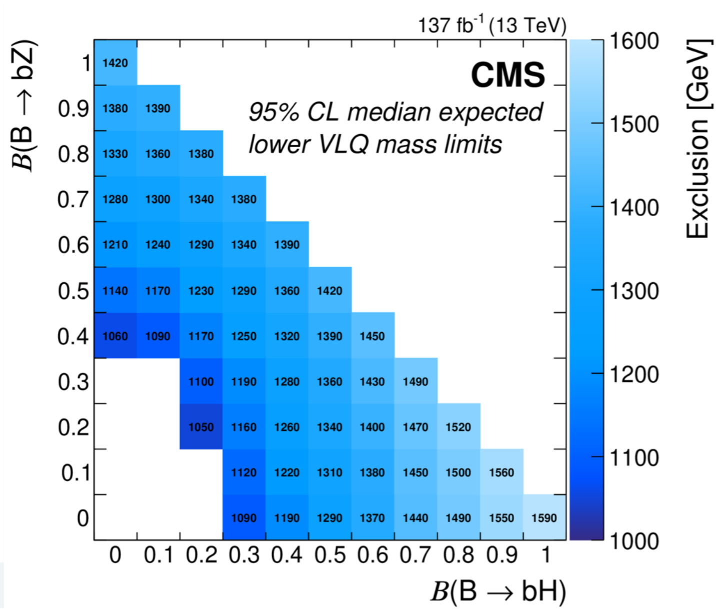
$$\text{4-jet events}: \chi^2_{\text{mod}} = \frac{(m_{\text{merged}_1} - \overline{m}_{\text{merged}})^2}{\sigma^2_{m_{\text{merged}}}} + \frac{(m_{\text{merged}_2} - \overline{m}_{\text{merged}})^2}{\sigma^2_{m_{\text{merged}}}} + \frac{(\Delta m_{\text{VLQ}} - \overline{\Delta m}_{\text{VLQ}})^2}{\sigma^2_{\Delta m_{\text{VLQ}}}}.$$

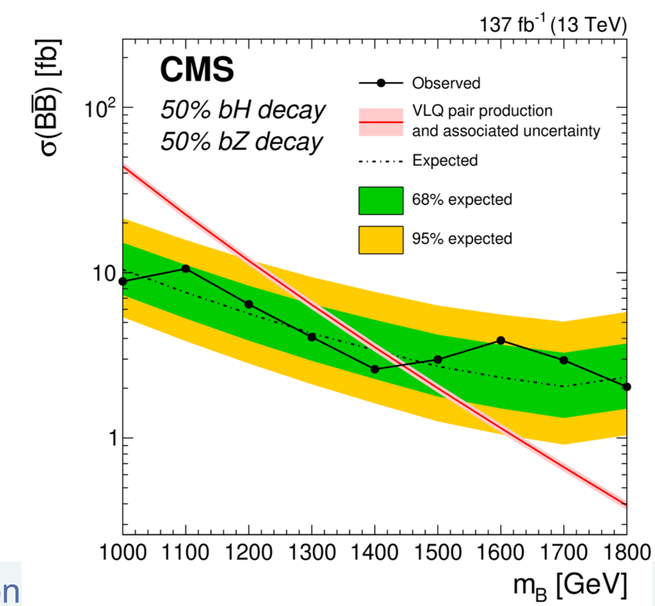
VLQ $B\bar{B}$ production (full-had)

CMS B2G-19-005

- Bottom-like VLQ pair production in the full hadronic mode
- B →bH/bZ
- 4, 5, or 6 jets depending on the number of merged H/Z
- Object reconstruction:
  - Individual jets: DeepJet b tagger applied to AK4 jets
  - Merged jets from bb pairs: Double-b tagger (AK8 jets)
  - Apply a jet multiplicity dependent $\chi 2$ metric to determine the best jet combination and reconstruct the event
- Determine final exclusion limits as a function of B → bH and B → bZ branching fractions



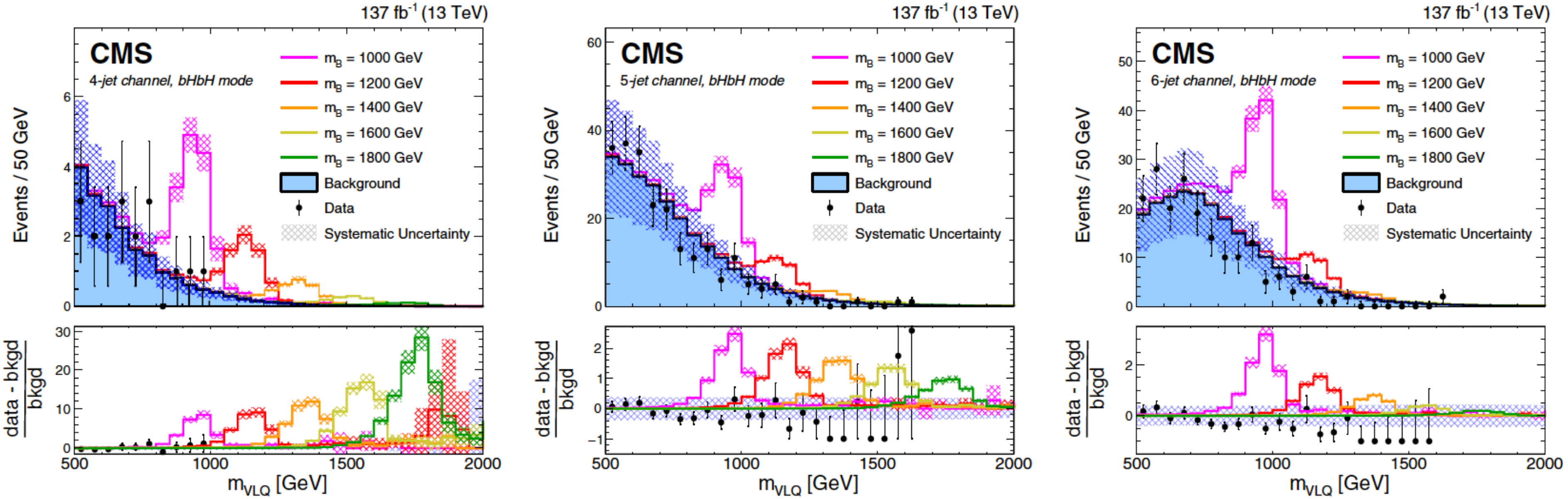




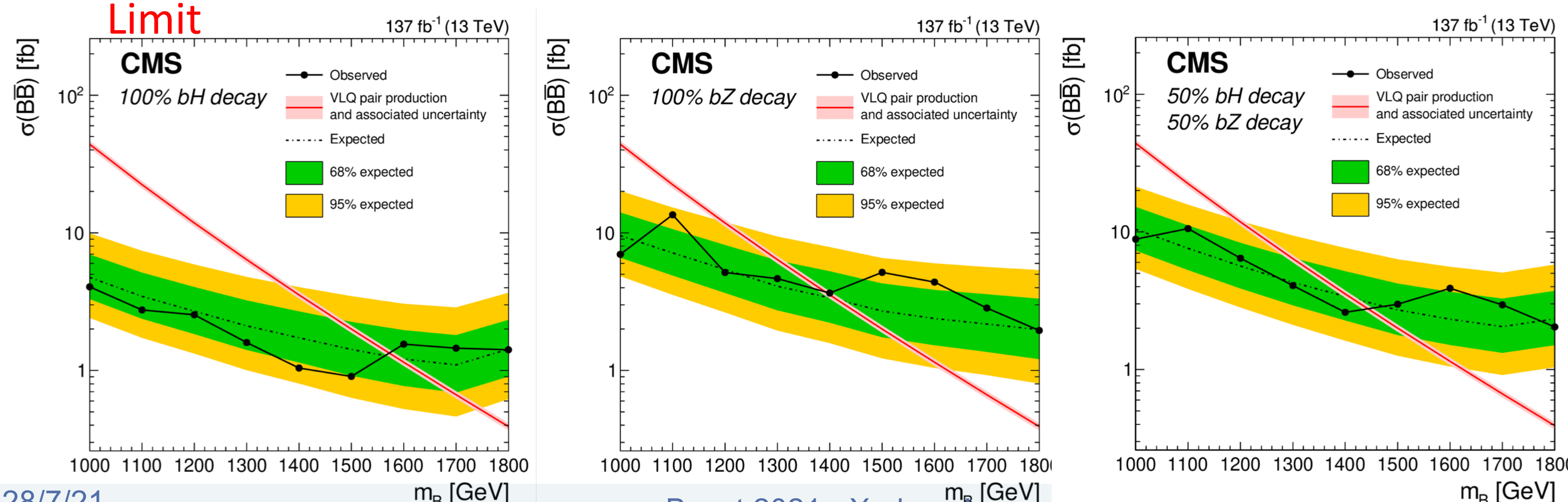
VLQ $B\bar{B}$ production (full-had)

CMS B2G-19-005

Postfit in bHbH mode



Limit



W_{KK}→WR→WWW→lv+jets

CMS-PAS-B2G-20-001

1. Focus at LL sample with W, t^2, g/q (left plot of last slides)

2. **Split the samples into 3 pure subsets** (applying cuts on τ_{ij}, deep-x/y, N_b, m_j) **in a way where each subset is dominated by a single type of jets** → mismodeling revealed

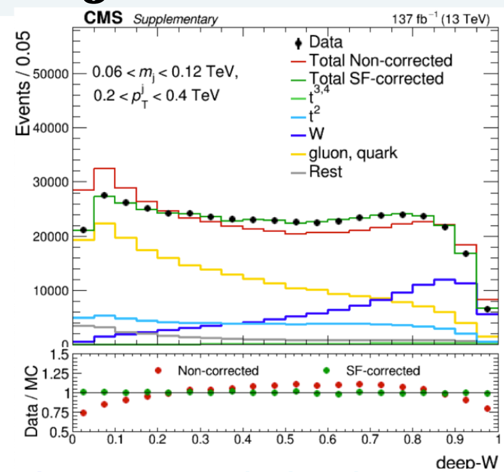
3. Demand: **Data = scaled sum of yields**

$$D_{i,k} = (g_{i,k})SF_k^g + (w_{i,k})SF_k^W + (t_{i,k})SF_k^t + d_{i,k}$$

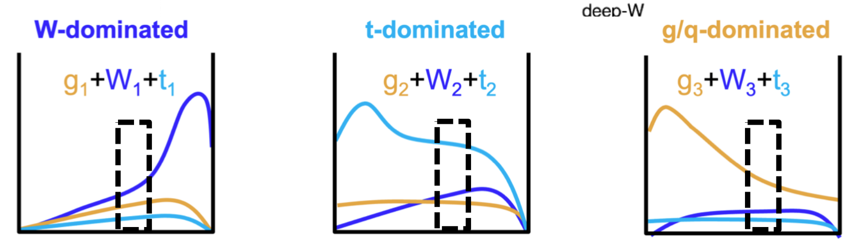
Define system of 3 equations, 1 per each subset "i", and per tagger score bin "k"

4. **Solve a 3x3 system** for SFs per each tagger score bin and **get SFs**→
   - **Known** yields: D, W, t, g/q, d
   - **Unknown** SFs

(1)



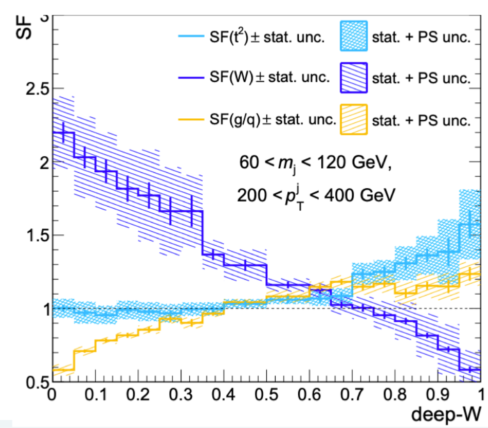
(2)



(3)

$$\begin{pmatrix} D_{1,k} - d_{1,k} \\ D_{2,k} - d_{2,k} \\ D_{3,k} - d_{3,k} \end{pmatrix} = \begin{bmatrix} g_{1,k} & w_{1,k} & t_{1,k} \\ g_{2,k} & w_{2,k} & t_{2,k} \\ g_{3,k} & w_{3,k} & t_{3,k} \end{bmatrix} \begin{pmatrix} SF_k^g \\ SF_k^W \\ SF_k^t \end{pmatrix}$$

(4)

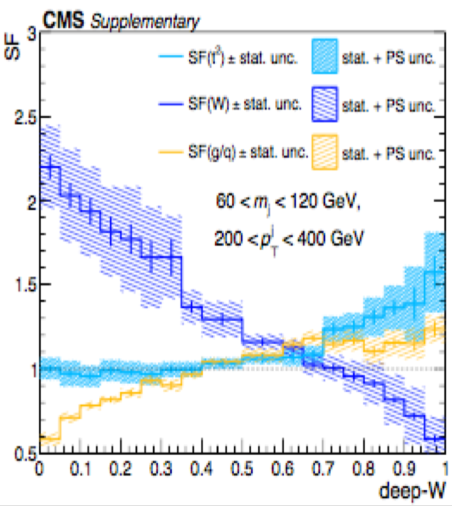


$W_{KK} \rightarrow WR \rightarrow WWW \rightarrow l\nu$+jets

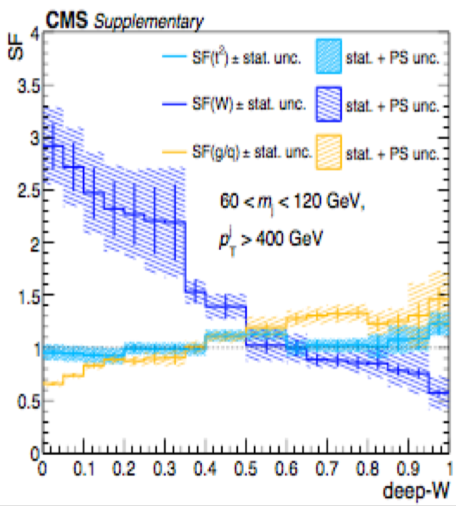
CMS-PAS-B2G-20-001

SFs for deep-W(WH)

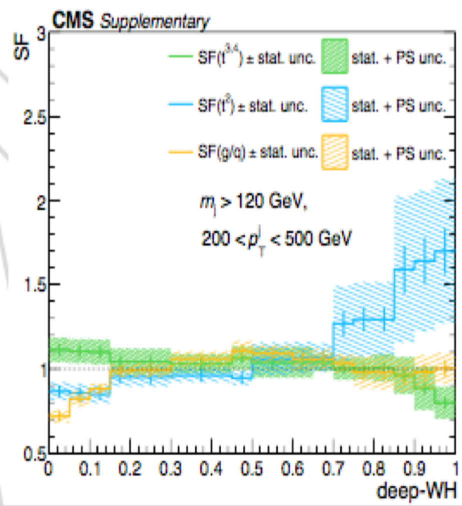
LL



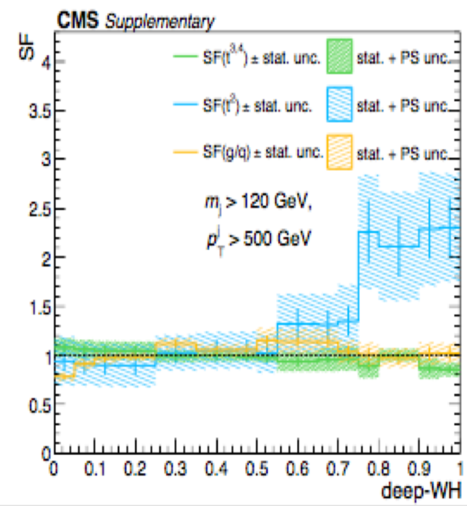
LH



HL



HH



$W_{KK} \rightarrow WR \rightarrow WWW \rightarrow l\nu$+jets

CMS-PAS-B2G-20-001

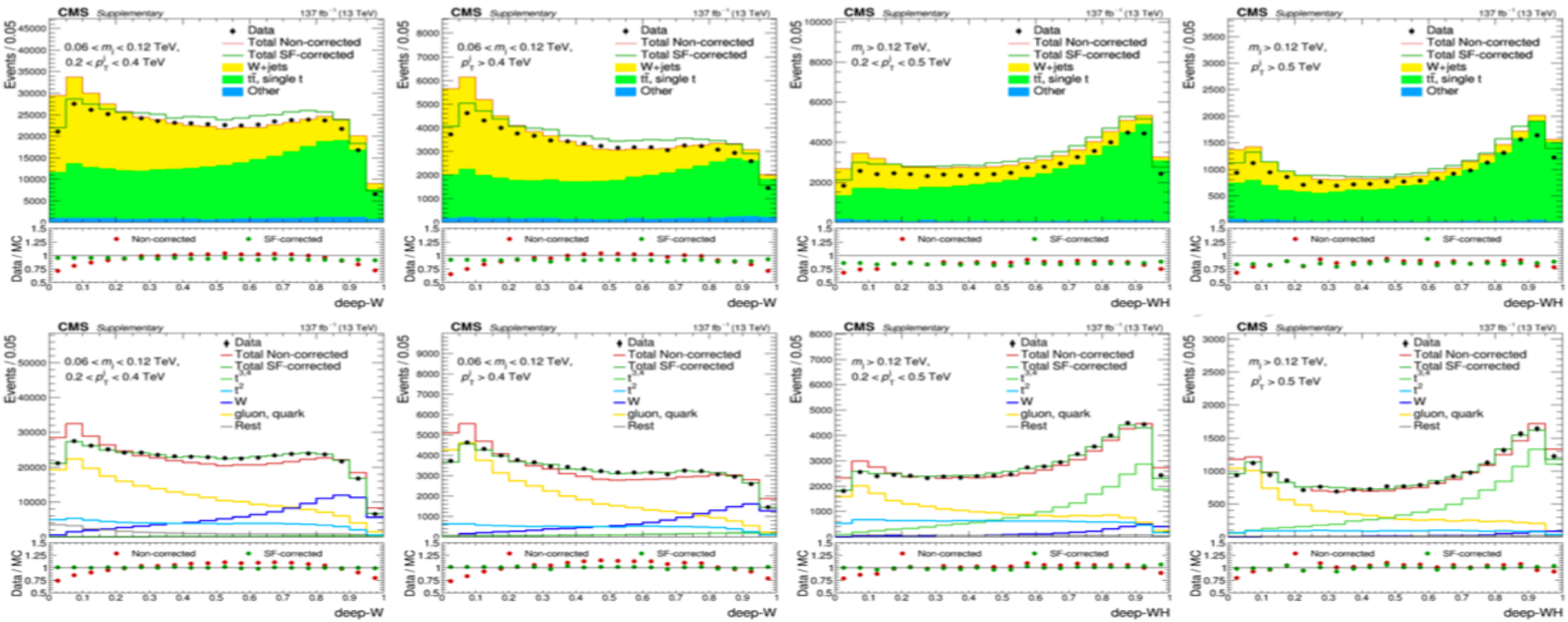


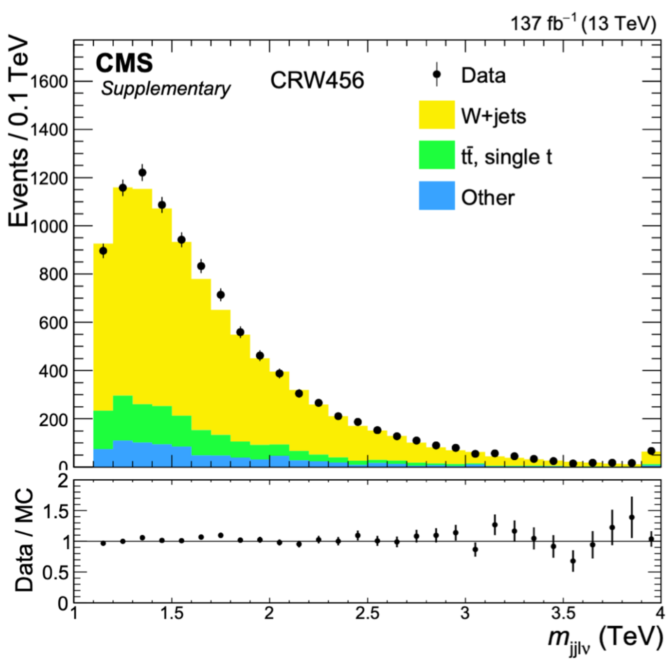
Figure 4: The deep-W (left two columns) and deep-WH (right two columns) discriminants of the jet with highest mass for different jet mass, m_j, and jet p_T ranges as indicated in the plots after preselection and vetoing the signal regions. The upper row shows the background processes (histograms) compared to the data (black markers). In the lower row, the background processes, which are normalized to the data, are split into categories that are defined by matching the reconstructed jet to processes at parton level: t^3 indicates a fully merged top-quark decay, t^4 contains an extra quark or gluon with $p_T > 50$ GeV inside the jet cone (both light green), and t^2 (light blue) contains only two quarks from the top-quark decay, but not both from the W boson decay. A merged W boson decay is indicated by W (dark blue), analoguously single quarks and gluons (yellow), and "Rest" indicates events not matching any of the categories. Before corrections (red), significant discrepancies between the prediction and the data can be observed, in particular at low and high discriminant values. The corrected distributions after scale factor (SF) application are shown using dark green. The lower panels show the Data/MC ratios before and after corrections.

$W_{KK} \rightarrow WR \rightarrow WWW \rightarrow l\nu$+jets

CMS-PAS-B2G-20-001

- Use SR-selection → Maintain kinematics as in SR
- **Invert deep-W(WH)** tagger cuts → Signal free samples with large statistics
- Reject tops: **deep-t<0.4** → Enhance W+jets purity (rejecting top)

Region	m_j^{max} [GeV]	taggers	m_j^{min} [GeV]	tagger	N_j^{AK8}	N_j^{AK4}	N_b
CRW1	60–100	deep-W(t) < 0.7(0.4)	—	—	1	≤ 2	0
CRW2	100–200	deep-WH(t)< 0.7(0.4)	—	—	1	≤ 2	0
CRW3	≥ 200	deep-WH(t)< 0.7(0.4)	—	—	1	≤ 2	0
CRW456	60–100	deep-W(t) < 0.5/0.7(0.4)	60–100/0–60	deep-W < 0.5/—	2	≤ 2	0



We have 4 such CRWs (in accordance with SR1-6); we illustrate only the CRW456 here.

- SR^{MC}, CR^{MC}, $CR^{DATA\text{-}Rest}$ have consistent $M_{j(j)l\nu}$ shapes
- Use the CR to deliver rate ($N^W_{CR_i}$) and shape (TF^W_i) correction to SR as:

$$PRED^W_{SR_i} = N^W_{CR_i}\, MC^W_{SR_i} \frac{[DATA-rest]_{CRW_i}}{N^W_{CR_i}\, MC^W_{CRt_i}} = N^W_{CR_i} MC^W_{SR_i} TF^W_i$$

0.96 to 1.03

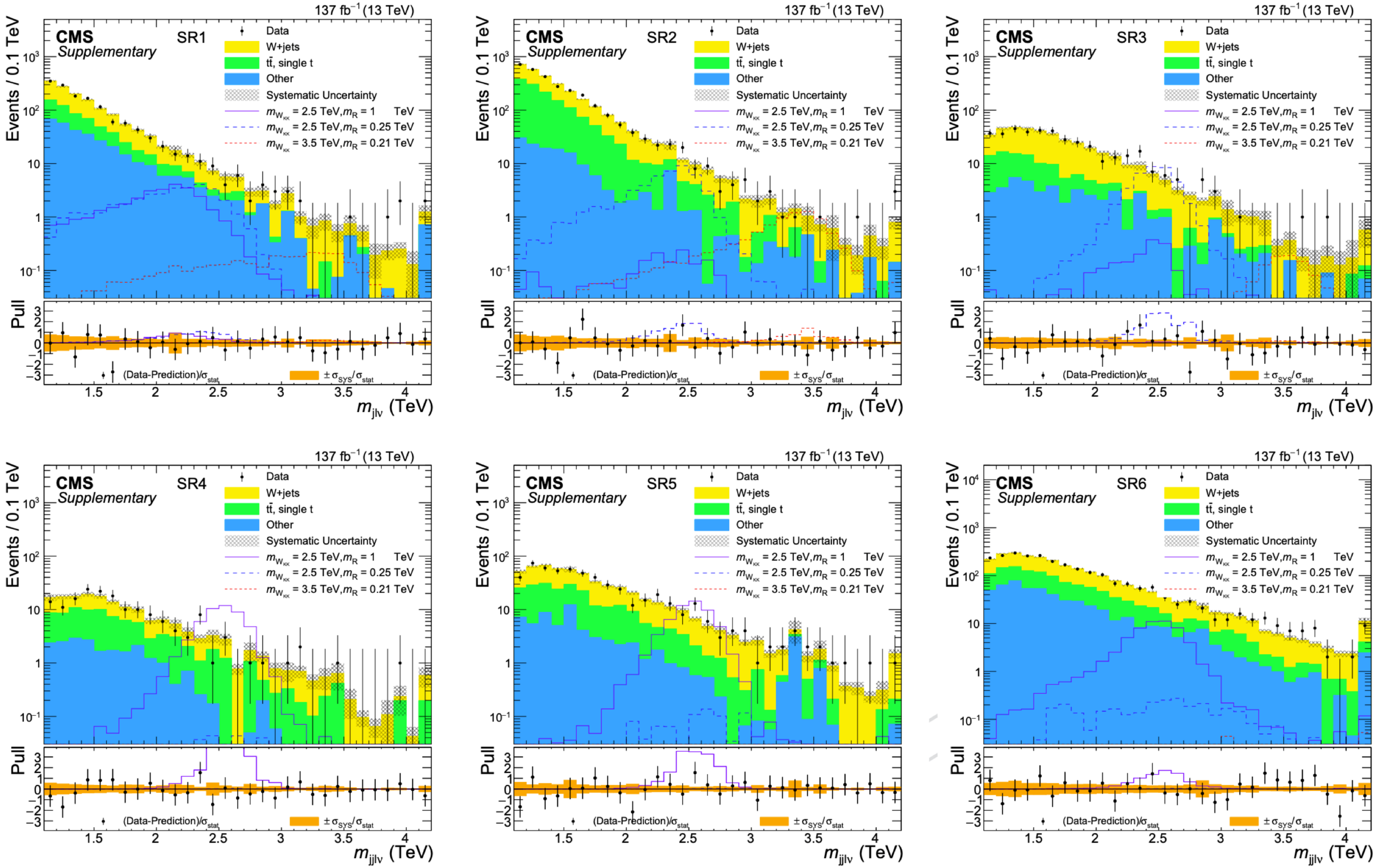
(Where MC is SF-corrected)

We **validate** prediction in low-ST samples

$W_{KK} \rightarrow WR \rightarrow WWW \rightarrow l\nu$+jets

CMS-PAS-B2G-20-001

SR1-6 Postfit Results

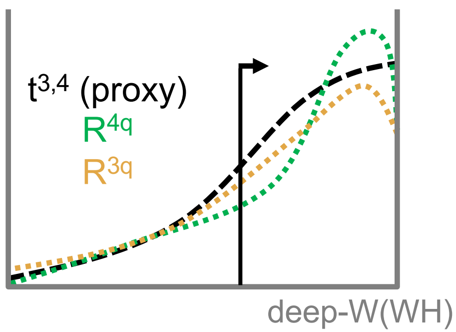


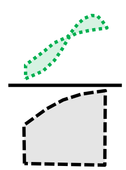
W_{KK}→WR→WWW→lv+jets

CMS-PAS-B2G-20-001

Unc. on SFs

1. **Parton Shower** ~10-20%
   Extract SFs with 3 alternative tt samples (powheg+p8, powheg+herwig7, MG+p8), maximum difference is used as unc.
2. **Bias** 10%
   (due to Matrix method selection cuts)
3. **Proxy-unc.**
   Accounts for differences between $R^{4q/3q}$, R^{lqq} and SM proxy jets: $t^{3,4}$, W.
   Compare normalized deep-W(WH) spectra to **evaluate % diff.** above the cut with metric:



Proxy unc.= $\sqrt{(\frac{\sum_i |t_i^3 - t_i^4|}{\sum_i t_i^4})^2 + (\frac{\sum_i |R_i^{3q|4q} - t_i^{3,4}|}{\sum_i t_i^{3,4}})^2} \propto$ 

4. **High-pT extrapolation**
   Signal jets much more boosted wrt SM. Generate herwig++ signal, use % diff. wrt pythia8 as unc.