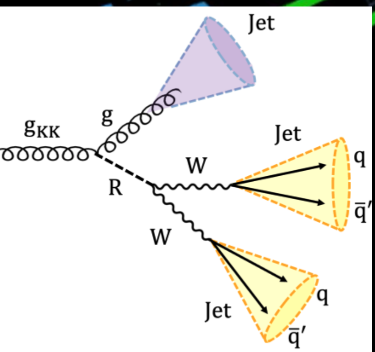
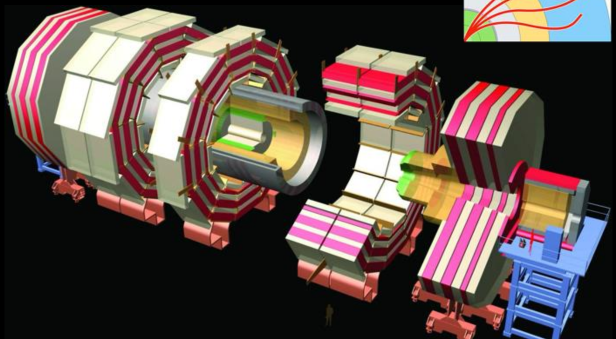
February 20th 2025,

Workshop: Polarized Perspectives: Tagging and Learning in the SM

Search for Di-resonant New Physics with Massive Jets at CMS





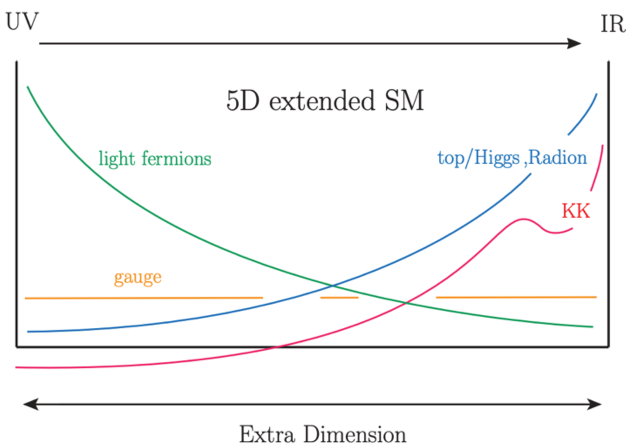
Antonis Agapitos

Motivation: BSM physics beyond minimal

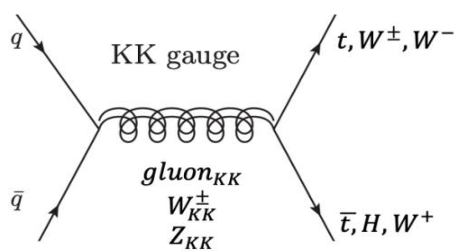
- Hierarchy: EW-M_{Pl} scale gap motivates BSM physics.
- No BSM physics yet → time to look in non-standard final states/scenarios.

Minimal Warped ED model:

- 2 Branes in bulk (in the RS framework). Everything propagates to the same bulk:

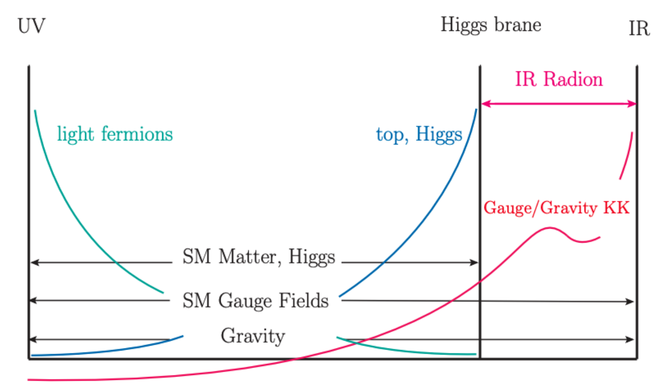


- Di-SM dominant phenomenology:
- Constrained by LHC searches.

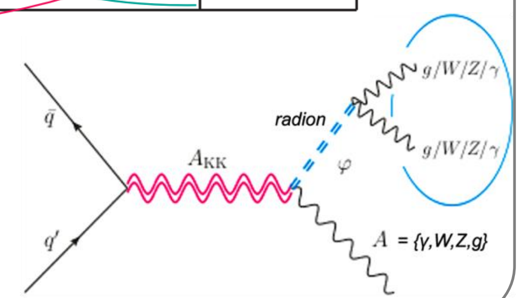


Extended Warped ED (EWED) model:

- Extra brane by splitting → extended bulk.
- 3 or more branes, 2 or more Radions.
- Various fields propagate in diff. subregions:



- Di-SM suppressed in favor of tri-SM:
- A wealth of new signatures emerges:

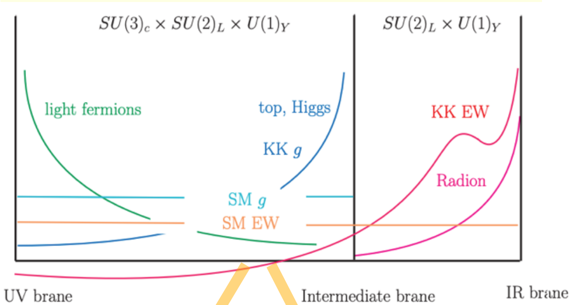


Theory sources:
Kaustubh Agashe, et al
his talk at CMS

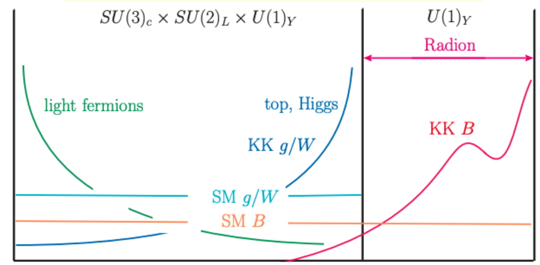
- LHC Signals from Cascade Decays of Warped Vector Resonances arXiv:1612.00047
- Dedicated Strategies for Triboson Signals from Cascade Decays of Vector Resonances arXiv:1711.09920
- Detecting a Boosted Diboson Resonance arXiv:1809.07334

EWED landscape & CMS searches

EW fields propagate at the bulk

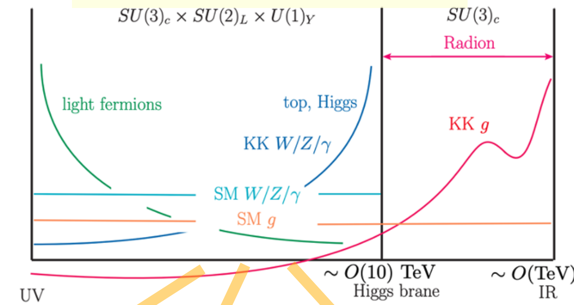


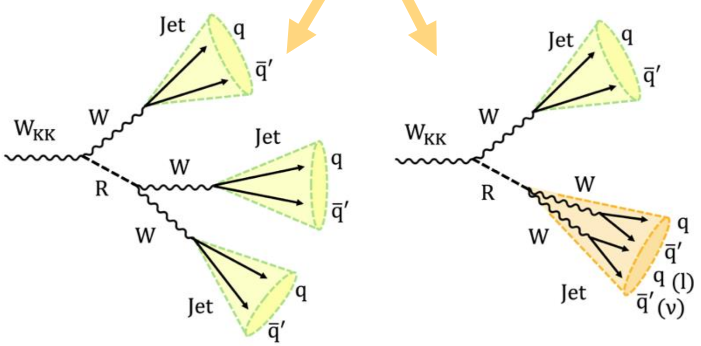
Hypercharge at the bulk



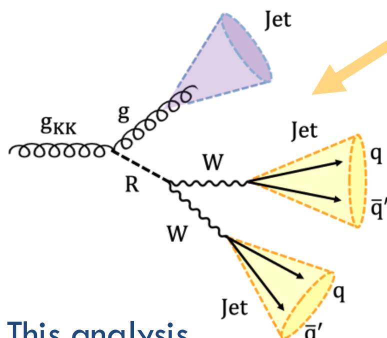
tri-γ signature - yet unexplored

QCD fields at the bulk

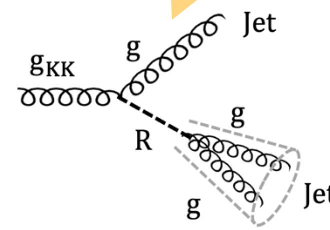




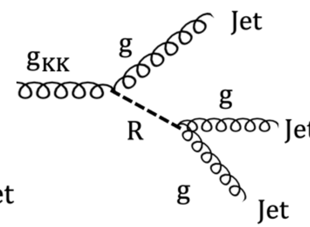
Tri-W; $W_{KK} \rightarrow WR \rightarrow WWW$
merged & resolved R; final states: 0l & 1l
2201.08476 , 2112.13090



This analysis
$g_{KK} \rightarrow gR \rightarrow gWW$
This search 2410.17303



Di-jet (merged)
$g_{KK} \rightarrow gR \rightarrow ggg$
2201.02140



Tri-jet (resolved)
$g_{KK} \rightarrow gR \rightarrow ggg$
2310.14023

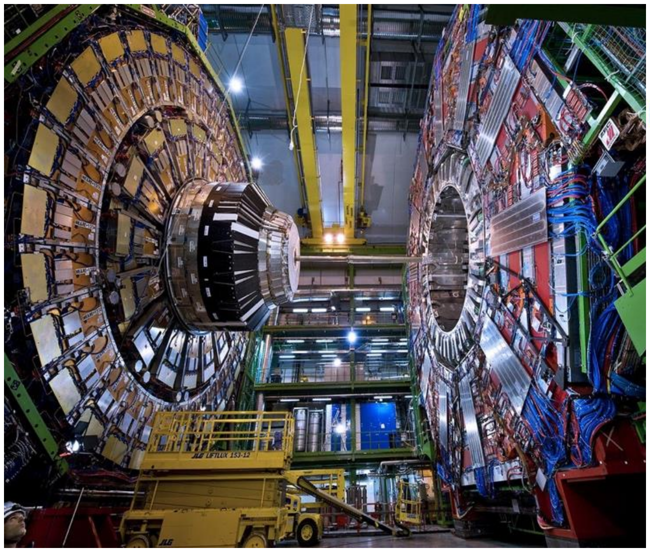
Benchmark points & associated dominant process →
(from 1612.00047)

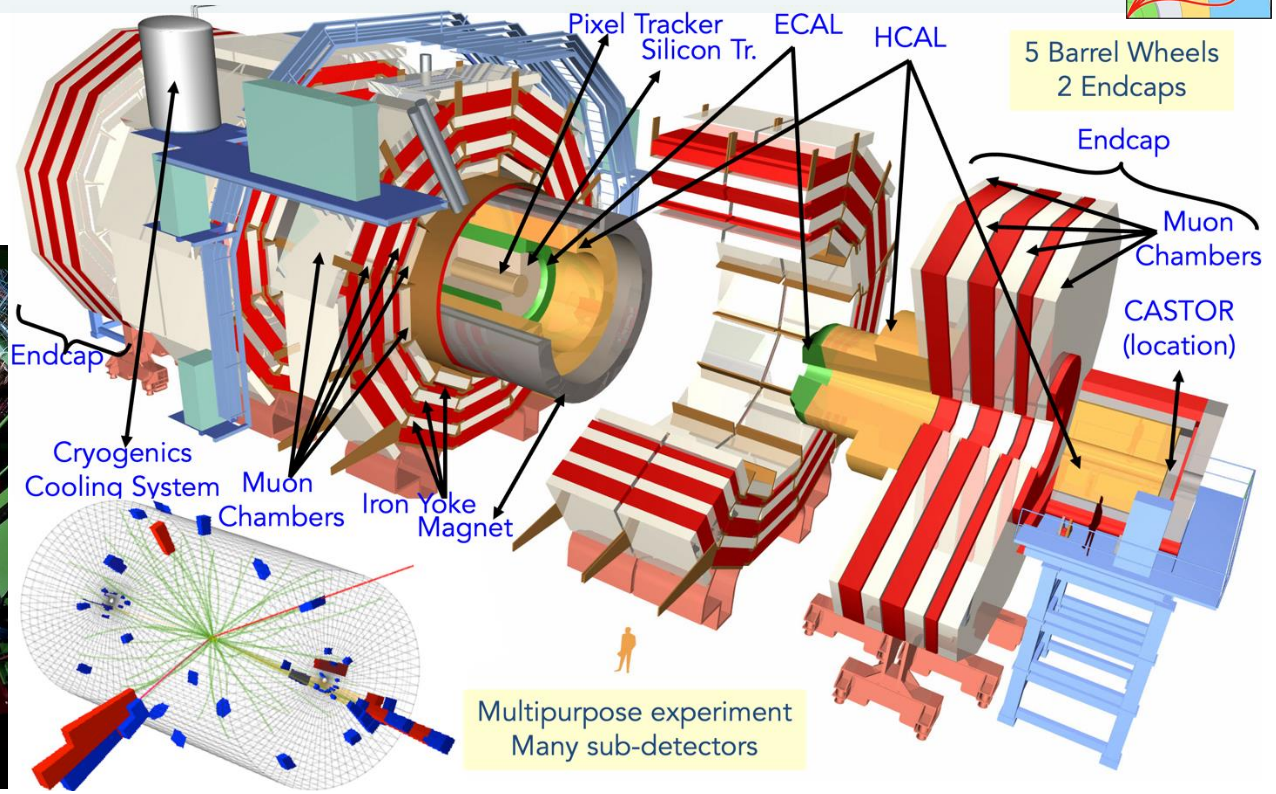
	Process	Name	m_{KK}	m_φ	$g_{\gamma_{KK}}$	$g_{W_{KK}}$	$g_{g_{KK}}$	g_{grav}
γ_{KK}	$\gamma_{KK} \to \gamma\varphi \to \gamma gg$ (5.1)	γ-γgg-BP1	3	1	3	6	3	3
		γ-γgg-BP2	3	1.5	2.7	6	3	4.1
g_{KK}	$g_{KK} \to g\varphi \to g\gamma\gamma$ (5.2.2)	g-$g\gamma\gamma$-BP1	3	1	2.7	6	6	2.25
		g-$g\gamma\gamma$-BP2	3	1.5	2.7	6	6	3
	$g_{KK} \to g\varphi \to ggg$ (5.2.1)	g-ggg-BP1	3	1	2.7	6	3	2.45
		g-ggg-BP2	3	1.5	2.7	6	3	4
	$g_{KK} \to g\varphi \to gV_hV_h$ (5.2.3)	g-gVV-BP1	3	1	2.65	3	6	3
		g-gVV-BP2	3	1.5	2.65	3	6	5
W/Z_{KK}	$W_{KK} \to W_l\varphi \to W_lgg$ (5.3)	W-Wgg-BP1	2.5	1	3.5	4.4	3	3.5
		W-Wgg-BP2	3	1.5	3	3.5	3	5.1

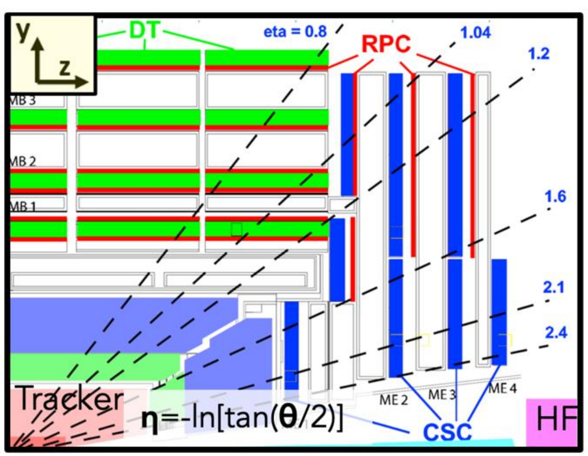
← 2201.02140 & 2310.14023
← This search 2410.17303
← 2201.08476 & 2112.13090

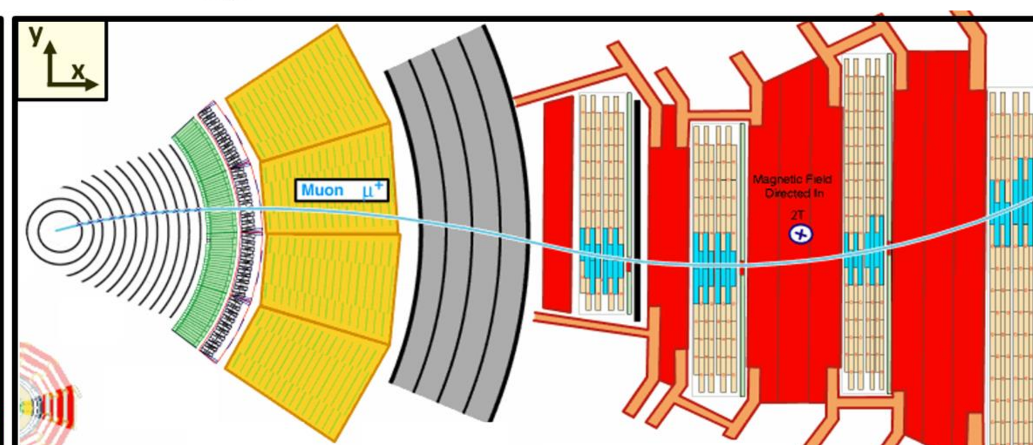
The CMS detector at the LHC

Compact Muon Solenoid
Mass: ~12500 Tones
Size: ~15m x 22m
Magnetic field: 4 T (3.8 T)
CMS collaboration is 30 y.o.
~6100 collaborators
~250 Institutes ~57 countries here for more



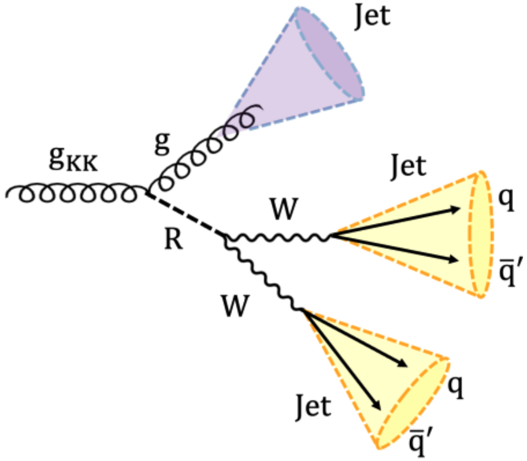


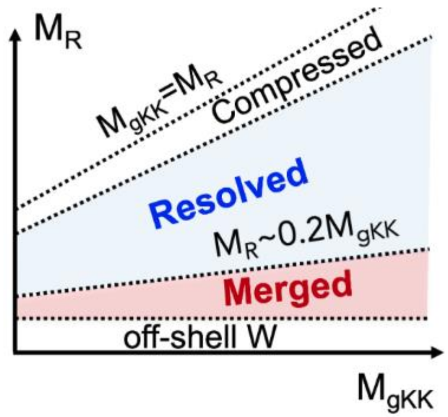


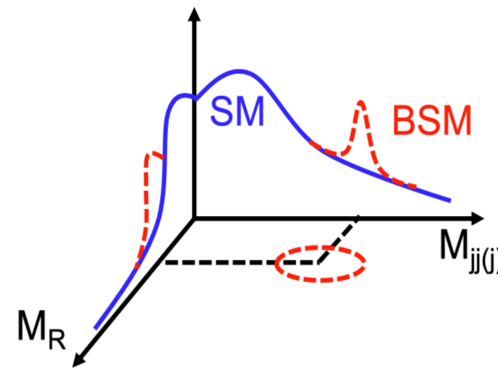


Signal topology & Preselection

- We use benchmark point at which the dominant process is: $g_{KK} \rightarrow gR \rightarrow gWW$
- Big advantage of the W-tagging & narrow mass-window to suppress QCD BKG.



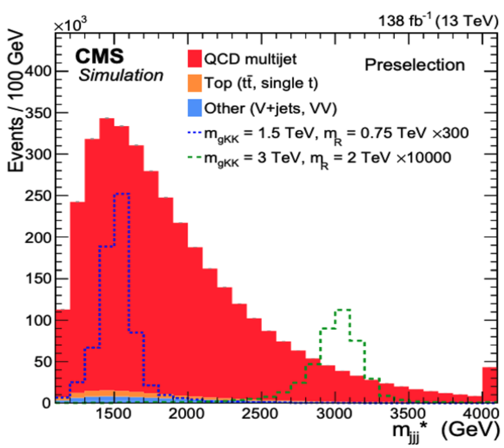




- g_{KK} is spin-1, R is spin-0
- We focus on the 0l channel: $g_{KK} \rightarrow gR \rightarrow gWW \rightarrow$ jets (BR~56%)
- We cover only the resolved R case: $0.2<m_R/m_{gKK}<0.9 \rightarrow$ 3 jets

Strategy:

1. Tri-jet selection,
2. Identify (tag) 2 jets as W-candidates with PNet,
3. form m_{jj} (R) and m_{jjj} (g_{KK}),
4. bin over m_{jj}, fit m_{jjj} →



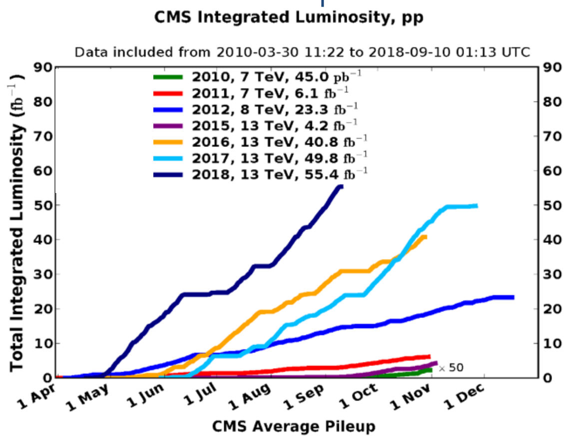
Preselection cuts:

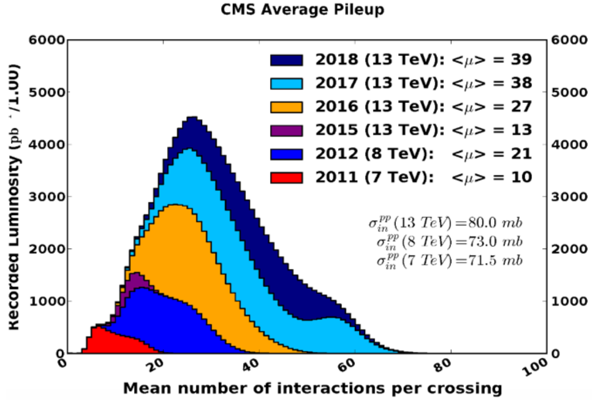
1. $N_{j\text{-}AK8}=3$, $N_{lep}=0$,
2. $p_{Tj1(j2,j3)} > 400\ (200)$ GeV, $|\eta_j| < 2.4$, $\eta = \ln[\tan(\theta/2)]$
3. $m_{ja,jb} > 50$ GeV,
4. $H_T \equiv \sum_i p_T(jet[i]) > 1.1$ TeV

Datasets, Trigger, & MC samples

DATA: pp collision at 13 TeV

- Full Run 2 (JetHT) dataset used.
- Trigger paths:
  H_T ($H_T \equiv \sum_i p_T(jet[i])$) & m_{jAK8}–based
- L = 138 fb^{-1}
- Triggers OR combination found to be eff. >~99% for H_T>1.1 TeV.





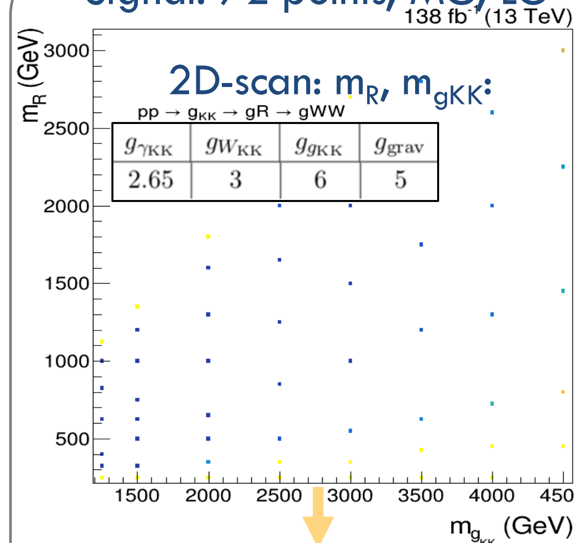
Simulation (MC) Madgraph, Pythia ...

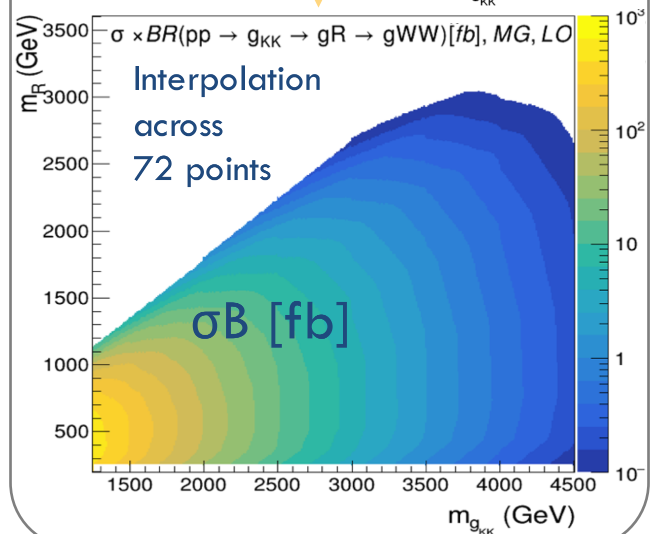
BKG samples

QCD_HT500to700_TuneCP5_1
QCD_HT700to1000_TuneCP5
QCD_HT1000to1500_TuneCP
QCD_HT1500to2000_TuneCP
QCD_HT2000toInf_TuneCP5_
TTToHadronic_TuneCP5_13T
TTToSemiLeptonic_TuneCP5
WJetsToQQ_HT-400to600_Tu
WJetsToQQ_HT-600to800_Tu
WJetsToQQ_HT-800toInf_Tun
ZJetsToQQ_HT-400to600_Tun
ZJetsToQQ_HT-800toInf_Tune
ZJetsToQQ_HT-600to800_Tun
ST_tW_antitop_5f_inclusiveD
ST_tW_top_5f_inclusiveDecay
ST_t-channel_antitop_4f_Inclu
ST_t-channel_top_4f_Inclusive
ST_s-channel_4f_hadronicDec
WW_TuneCP5_13TeV-pythia8
ZZ_TuneCP5_13TeV-pythia8
WZ_TuneCP5_13TeV-pythia8

- QCD multijet
- Top ($t\bar{t}$, single t)
- Other (V+jet, VV)

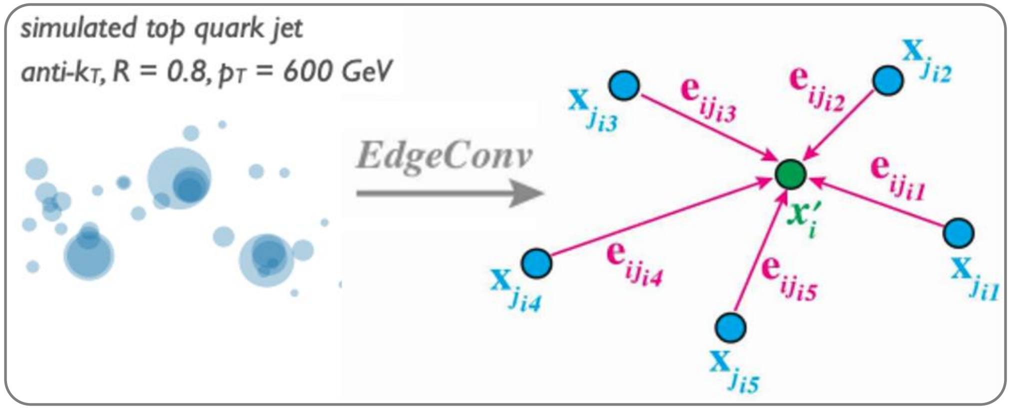
Signal: 72 points, MG, LO





W-tagging with PNet & SR binning

- Use Particle Net (PNet) tagger (1902.08570) to identify W→qq merged jets.
  → Graph NN, treat jets as particle cloud
  → Convolution on point clouds (EdgeConv 1801.07829)

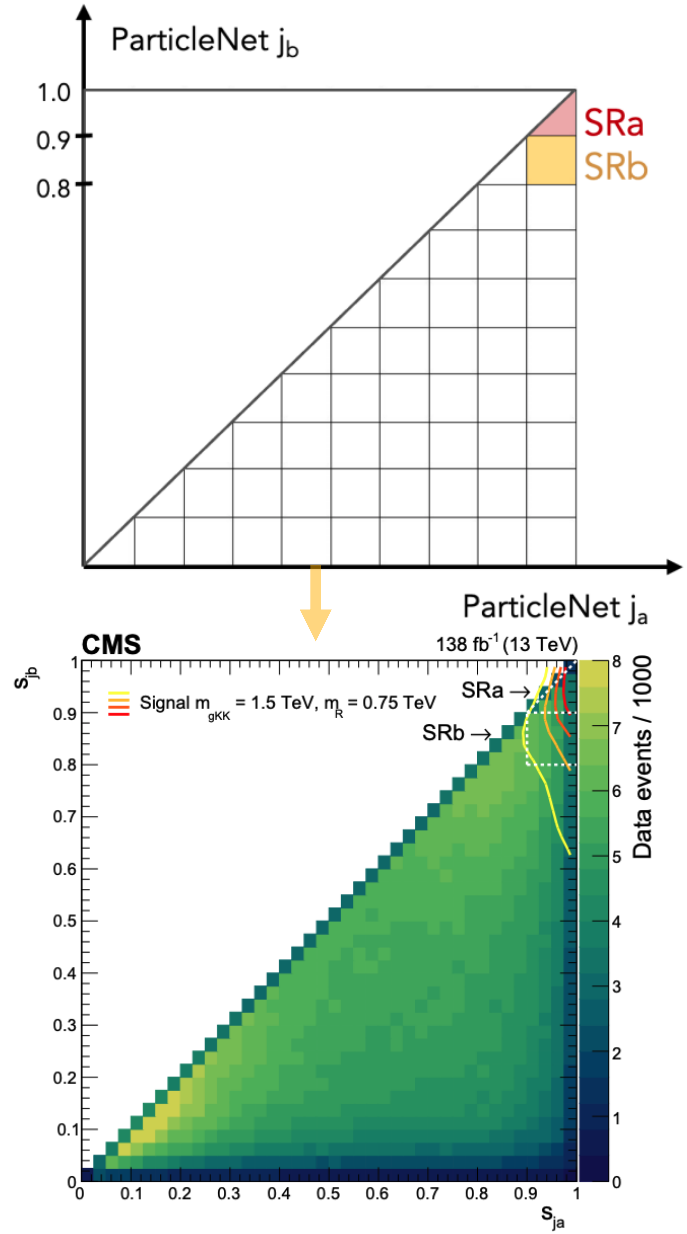


- Form ratio of "W_{qq}/QCD" classes score.

- The 2 highest PNet score jets j_a, j_b are assigned as W-candidates, is j_c. is the gluon.

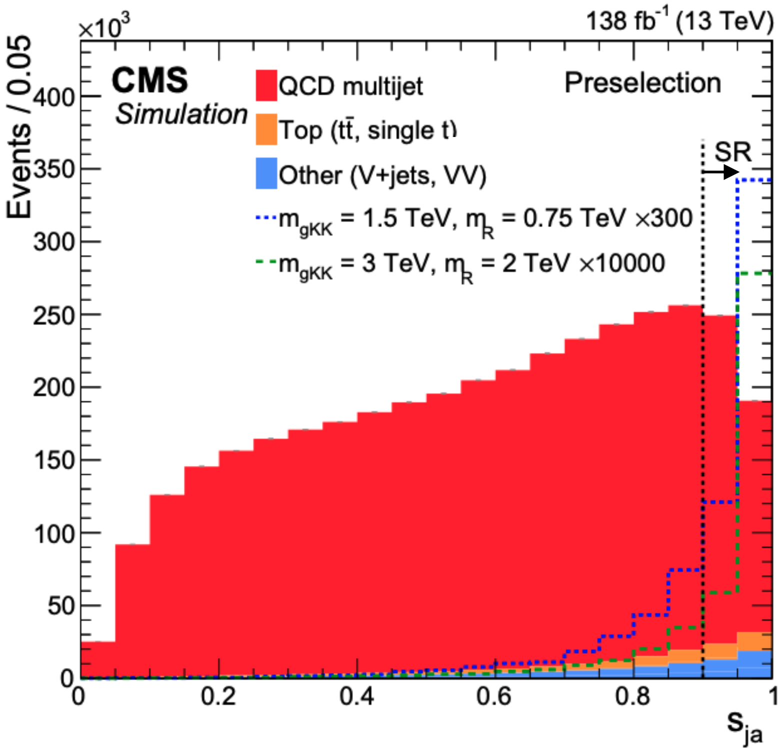
- Use PNet (MD) scores of j_a & j_b to select as:

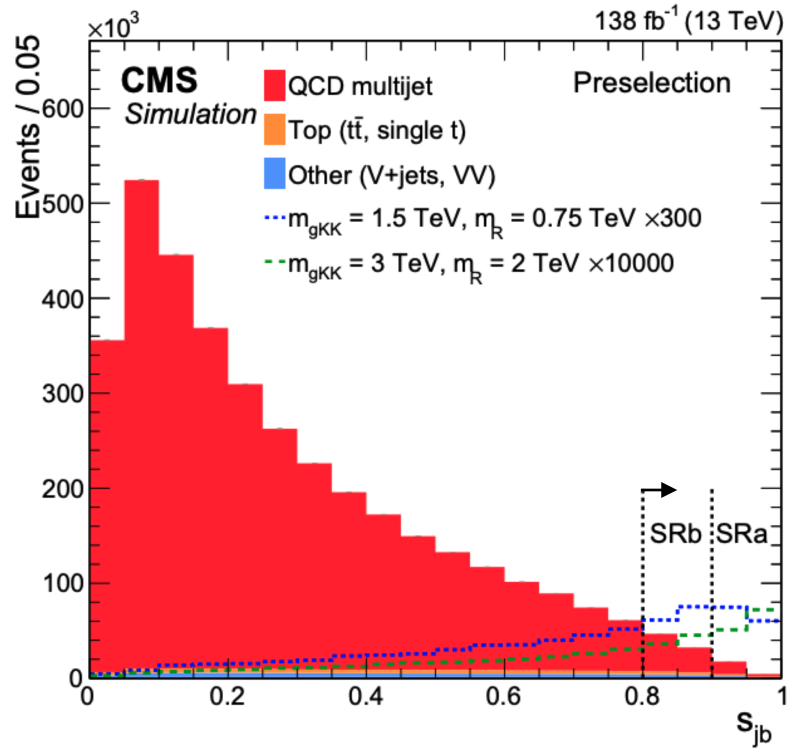
- SRa → both jets with $PNet_{ja,jb}$ >0.9
- SRb → $PNet_{ja}$>0.9 & 0.8<$PNet_{jb}$<0.9



W-tagging with PNet & SR binning

- The PNet (MD) scores of j_a & j_b according to simulation
- SRa → both jets with $PNet_{ja,jb} > 0.9$
- SRb → $PNet_{ja} > 0.9$ & $0.8 < PNet_{jb} < 0.9$





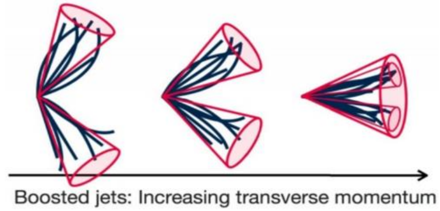
- No demands for 3rd jet, as the gluon candidate: m_{jc} or $PNet_{jc}$.
  → This maintains generality and provides sensitive to signals like: X→AWW, or X→WW + $j^{ISR/FSR}$.

PNet Tagger is calibrated with SFs formed on tt data sample:

Jet p_T [GeV]	200–300	300–400	> 400
$s_{ja} > 0.9$	0.83 ±0.03	0.84 ±0.04	0.82 ±0.05
$0.8 < s_{jb} < 0.9$	1.08 ±0.03	1.01 ±0.04	1.02 ±0.05

W-candidate selection on m_{jet}

- $W \to qq$ are boosted: using the anti-KT algo form single AK8 jets

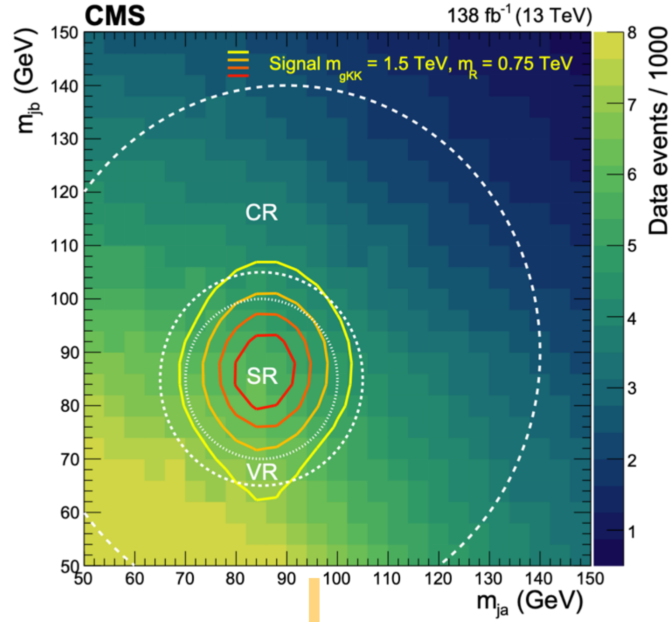


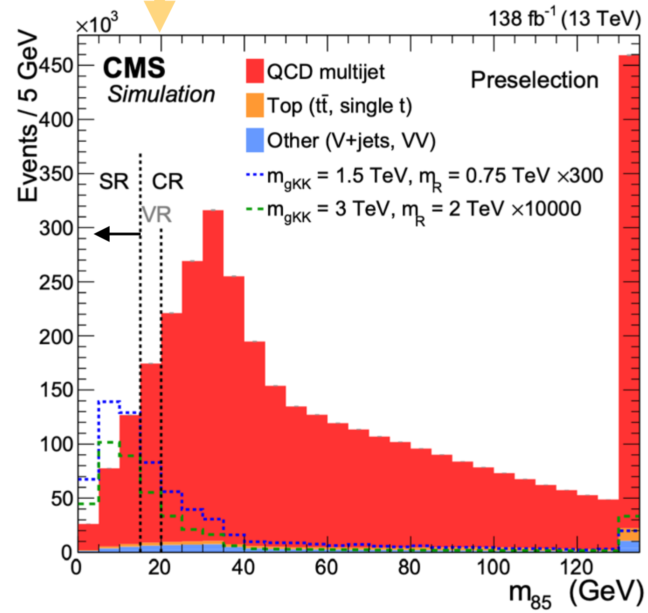
- The 2 highest PNet score jets j_a, j_b are assigned to be the W-candid., gluon is j_c.

- We demand the jets Soft Drop masses $m_{ja,jb}$, to be on W-peak with the condition of m_{85} variable:

$$m_{85} \equiv \sqrt{(m_{ja}+85)^2 + (m_{jb}+85)^2} < 15\ \text{GeV}$$

- We define 3 regions based on m_{85}:
  - Signal Regions (SRs)have: $m_{85} < 15$ GeV.
  - Control Regions (CRs) are: $m_{85} > 15$ GeV & $m_{90} < 50$ GeV
  - Validation Regions (VRs): $15 < m_{85} < 20$ GeV.



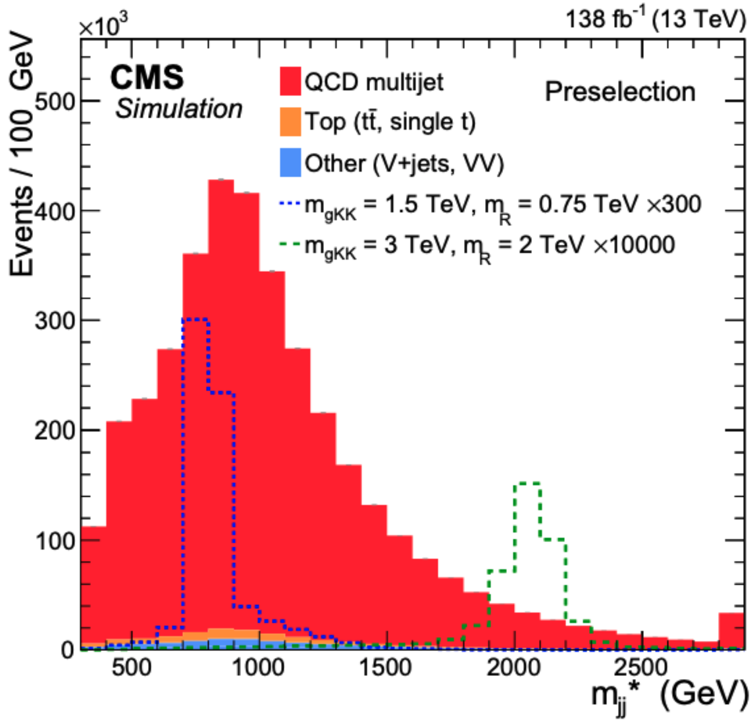


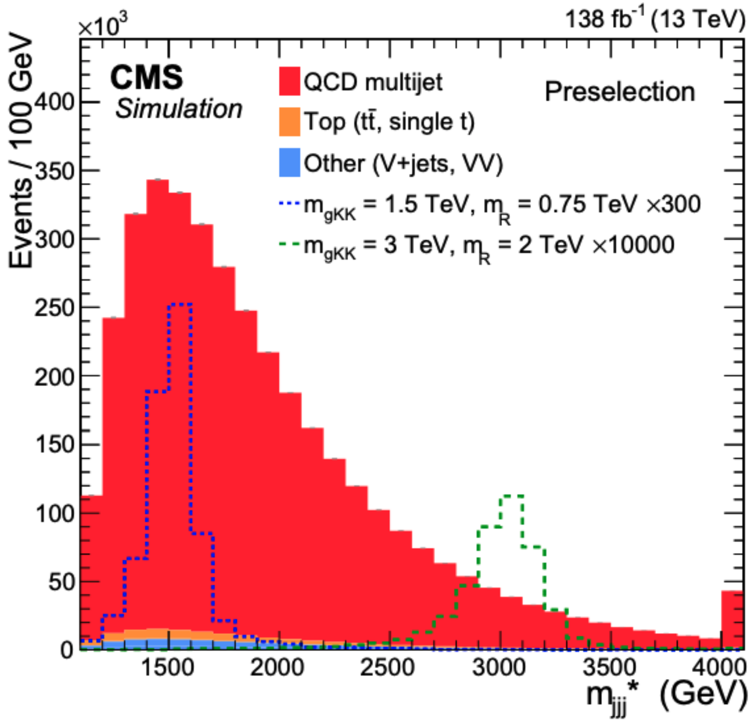
The Soft-Drop is an algorithm which removes soft & wide-angle radiation from within the jet, improving mass scale & resolution:

We use the anti-kT algo to cluster individual particles (PF candidates) into jets (using clustering param. R).

R & gKK masses reconstruction

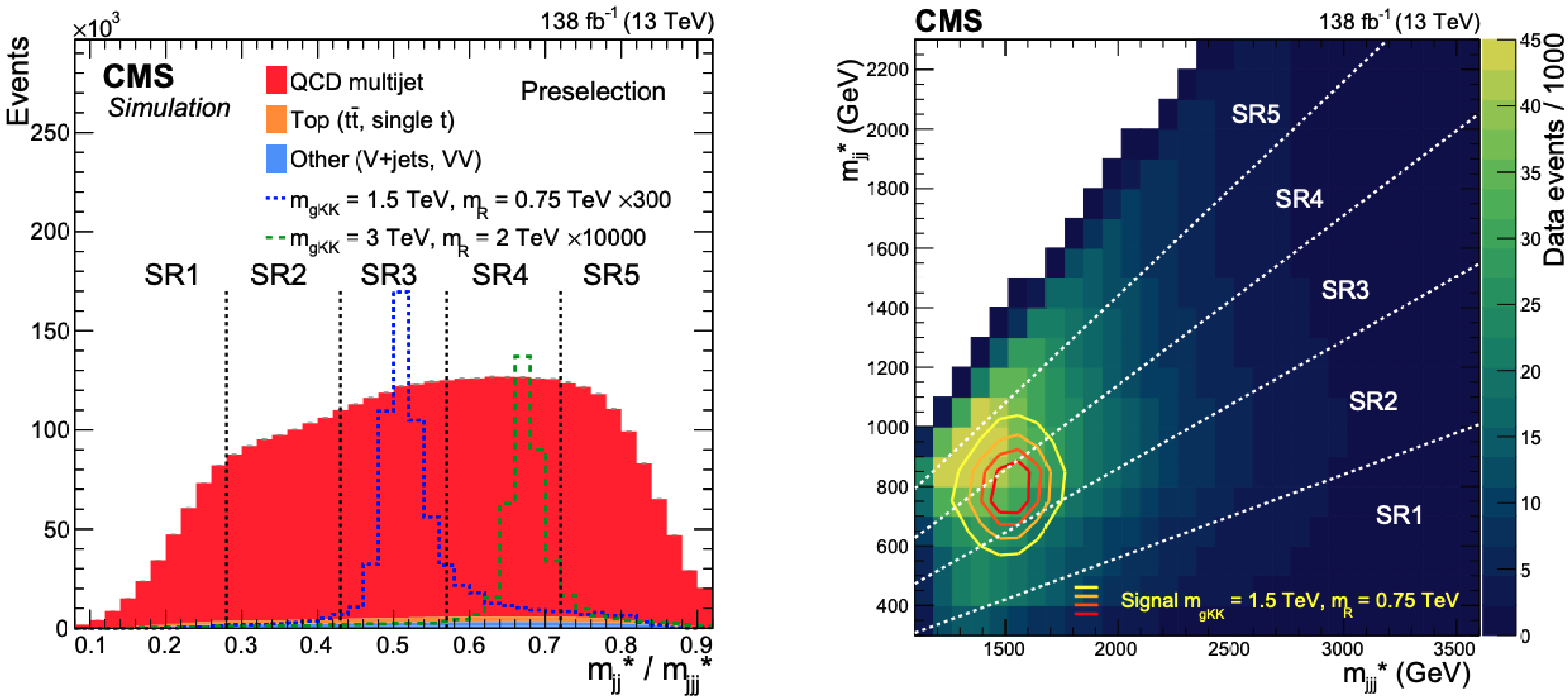
- M_R reco. from j_a, j_b: $$m^*_{jj} \equiv m_{jj} - m_{ja} - m_{jb} + 2 \times 85\,\text{GeV}$$
- M_{gKK} reco. from j_a, j_b, j_c: $$m^*_{jjj} \equiv m_{jjj} - m_{ja} - m_{jb} + 2 \times 85\,\text{GeV}$$
- → i.e. we correct invariant masses to mitigate resolution effect from jet SD masses.
  → sharper peaks (see Fig.4).
  → ~3% significance gain.





(The variable which we fit)

- From ratio m_{jj}^*/m_{jjj}^* and define 5 bins SR1—5. → Effectively binning over m_R.



- In each of these 5 SR we have 2 SRs (SRa, SRb) based on PNet scores.
  → Thus, we have 10 SRs in total.

- We fit the m_{jjj}^* spectra.

BKG prediction in 10 SRs

SR full selection summary

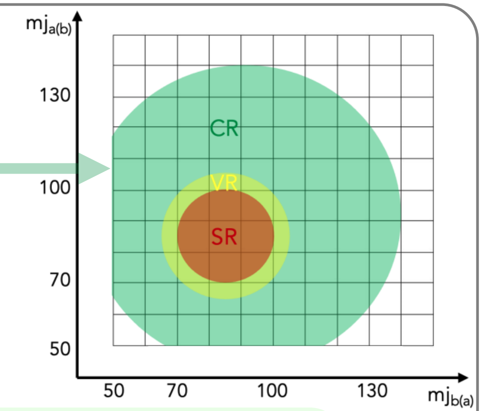
1. $N_{j\text{-}AK8}=3$, $N_{lep}=0$,
2. $p_{Tj1\,(j2,j3)}>400(200)$GeV $|\eta_j| < 2.4$,
3. $m_{ja,jb} > 50$ GeV,
4. $H_T > 1100$ GeV,
5. $m_{85} < 15$ GeV,
6. PNet > 0.8, & binning
7. $|\Delta\eta_{jj}|^{max} < 3$
8. $N_b=0$ (CHS, tight, deepflavor)

10 SRs categories:

Region	m^*_{jj}/m^*_{jjj}	s_{jb}
SR1a	< 0.28	> 0.9
SR1b		0.8–0.9
SR2a	0.28–0.43	> 0.9
SR2b		0.8–0.9
SR3a	0.43–0.57	> 0.9
SR3b		0.8–0.9
SR4a	0.57–0.72	> 0.9
SR4b		0.8–0.9
SR5a	> 0.72	> 0.9
SR5b		0.8–0.9

QCD multijet 80-90%

- Dominant → data-driven prediction
- Form Control Regions (CRs) defined in $m_{ja,jb}$ sideband as: $m_{85}>15$ & $m_{90}<50$ GeV keeping the rest conditions as in SRs.
- Form 10 CRs: CR1–5a & CR1–5b
- Similar kinem/cs to SRs; high QCD purity.
- Predict QCD with → $$\mathrm{Pred}^{\mathrm{QCD}}_{\mathrm{SRxy}} \equiv [\mathrm{Data}-\mathrm{Rest}]_{\mathrm{CRxy}}\frac{\mathrm{QCD}_{\mathrm{SRxy}}}{\mathrm{QCD}_{\mathrm{CRxy}}}$$
- We validate QCD pred. in 10 VRs (defined by $15<m_{85}<20$ GeV).



Top ($t\bar{t}$, single t) 3–8%

Other (V+jet, VV) 8–16%

- Subdominant BKGs → use MC for prediction.
- We correct the MC applying SFs for PNet selection eff. per matched W→qq jets.
- We validate Top MC (shape & rate) in dedicated samples (bRs) like the SRs but with $N_b\geq1$.
- We assign conservative (large) rate unc. for these 3 BKGs.

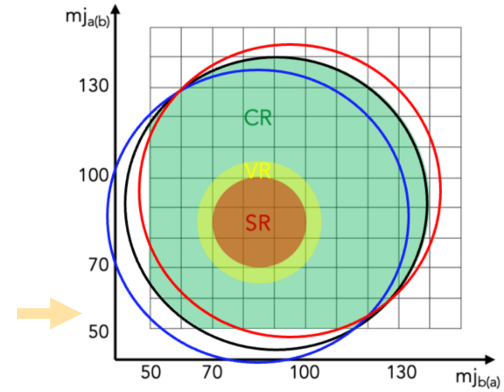
Systematic Uncertainties

BKG

Uncertainty source	Effect on	Magnitude	Number of NPs & correlations
Normalization QCD	Rate	20% ← Dominant	10, uncorr. across SRs
Normalization Top	Rate	50%	10, uncorr. across SRs
Normalization Other	Rate	30%	10, uncorr. across SRs
QCD bkg. shape due to m_{90} usage	Shape	$\pm1\sigma$ templates	10, uncorr. across SRs
QCD bkg. shape due to other processes	Shape	$\pm1\sigma$ templates	10, uncorr. across SRs

RATE
- QCD 20% based on validation prefit disclosure & MC low stat.
- Top 50% based on data in bRs, Other 30% based on similar search.
  All uncorrelated across 10 SRs → 30 nuisances.

SHAPE
- Vary "rest" in QCD BKGs prediction by x2 down, x0 up.
- Shift CR circle center: m_{90}<50 (central) → m_{85}<50 (down), m_{95}<50 (up).



Signal

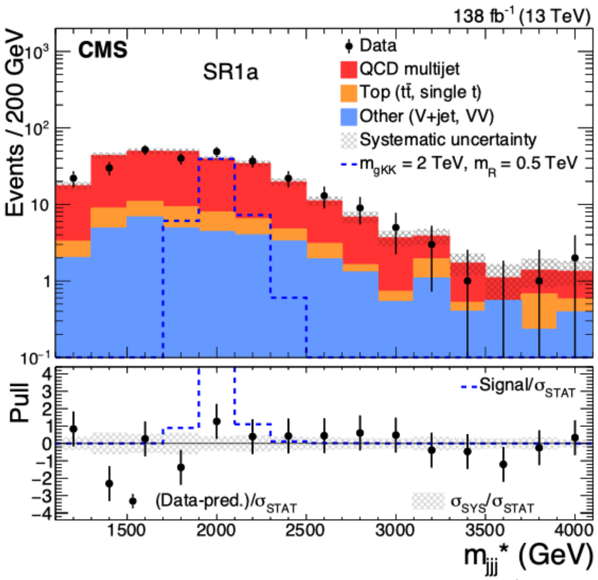
Uncertainty source	Effect on	Magnitude	Number of NPs & correlations
PU reweighting & int. luminosity	Rate	1.7%	1, correlated across all SRs
PDFs	Rate	$\leq 10\%$	1, correlated across all SRs *
μ_R/μ_F scales	Rate	$< 0.8\%$	1, correlated across all SRs *
PNet$_W$ selection eff. per jet (event)	Rate	6% (12%) ← Dominant	1, correlated across all SRs
JEC	Shape	$\pm1\sigma$ templates	1, correlated across all SRs *
JER	Shape	$\pm1\sigma$ templates	1, correlated across all SRs *

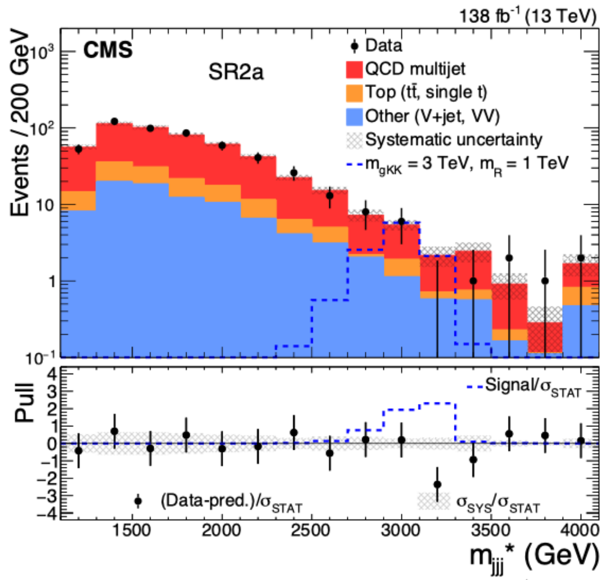
RATE
- Lumi, PU, PDFs, QCD scales μ_F, μ_R : 1—10%
- PNet SFs unc. → 6% [12%] per jet [event] (we have 2 W→qq jets/event)

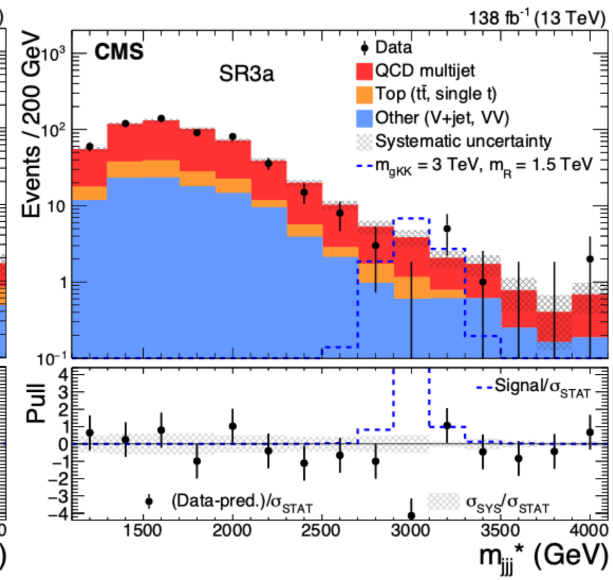
SHAPE
- JEC & JER: +σ/-σ variations → forming templates per point, per SRs.

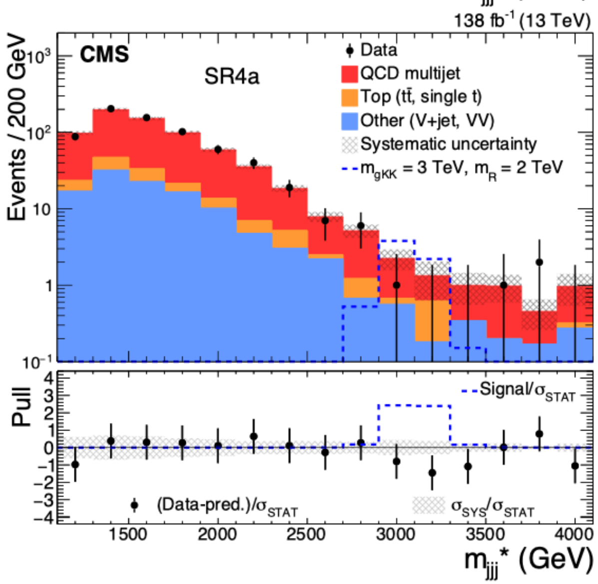
Results: SR1a—SR5a

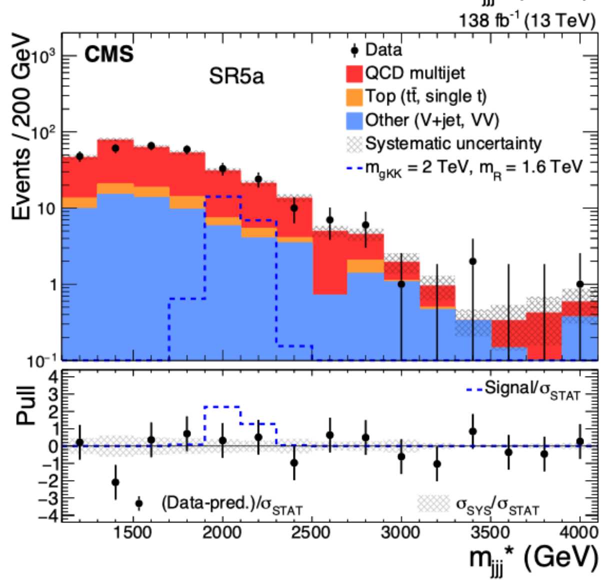
- We fit simultaneously the m_{jjj}^* spectra in the 10 SRs, using Combine tool:







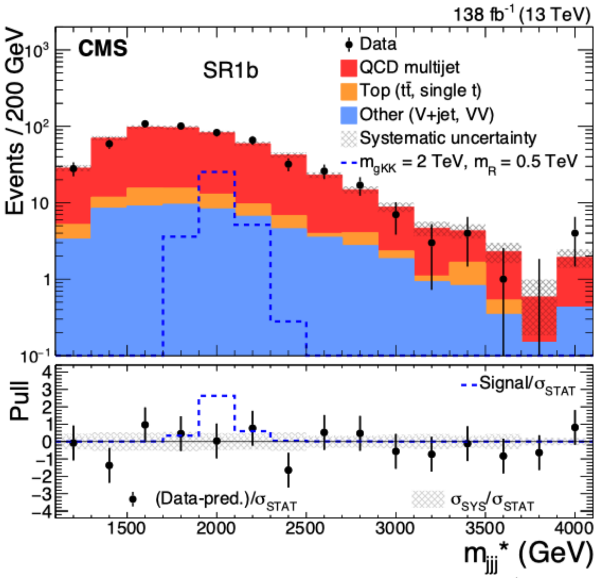


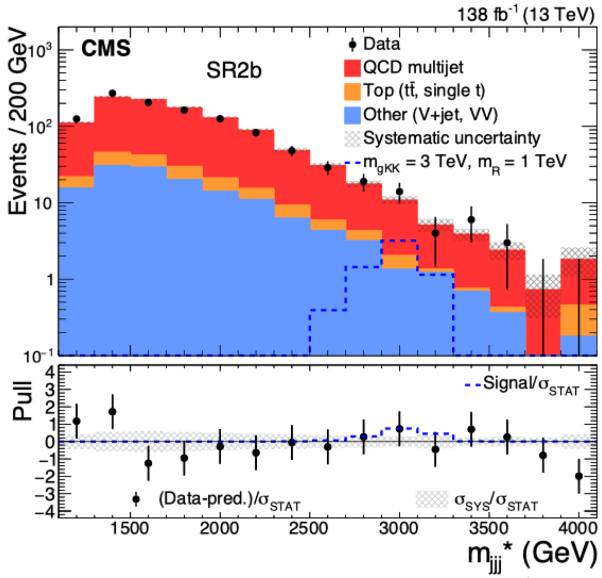


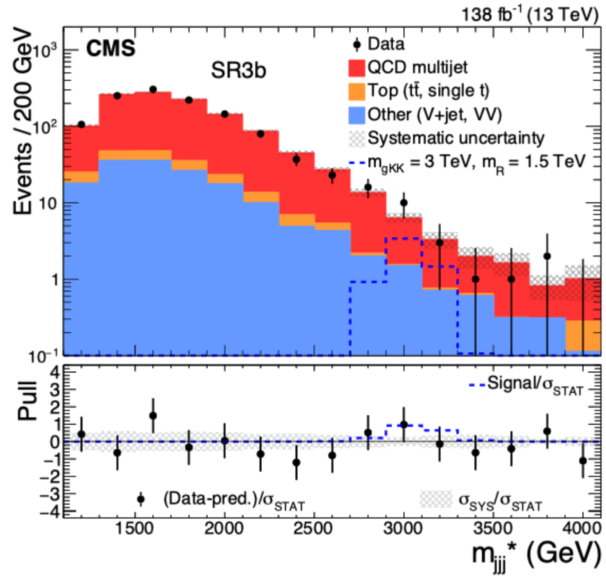
- No indication of signal.
- Excellent post-fit data/pred. agreement.

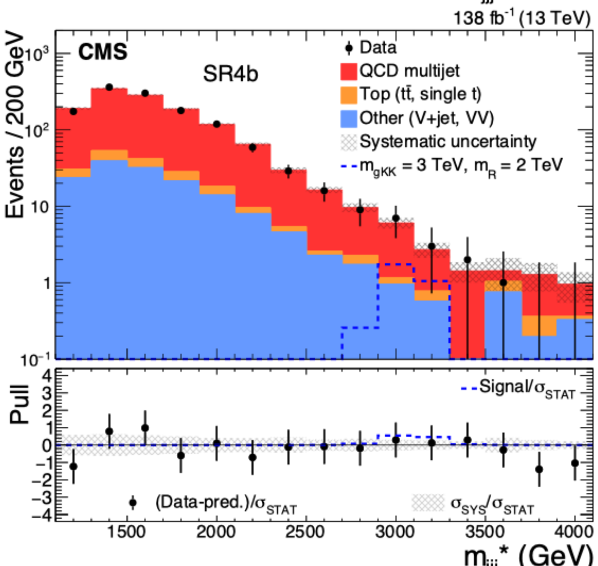
Results: SR1b—SR5b

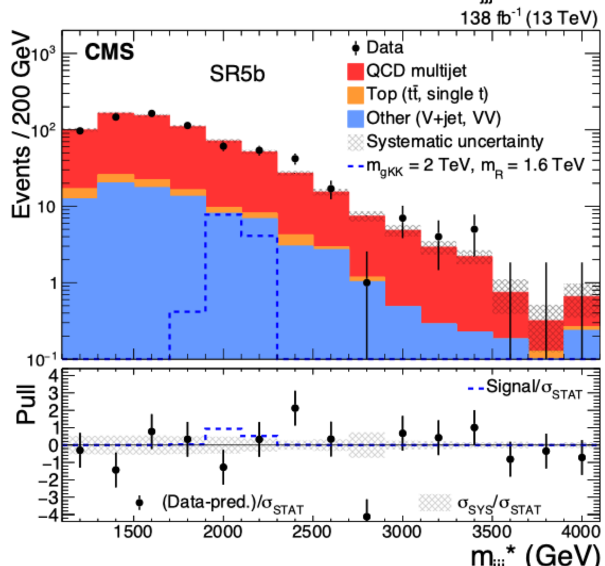
- We fit simultaneously the m_{jjj}* spectra in the 10 SRs, using Combine tool:







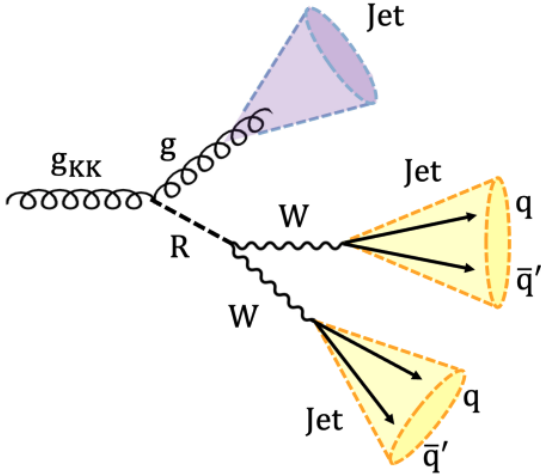


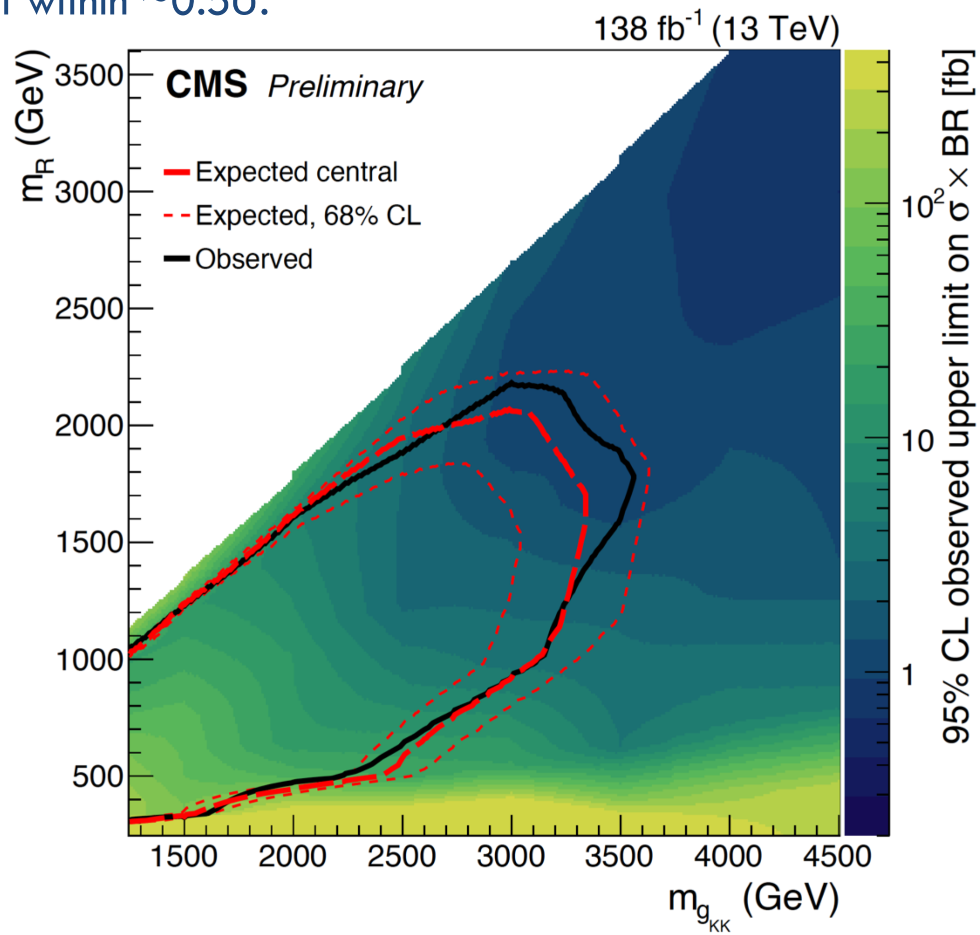


- No indication of signal.
- Excellent post-fit data/pred. agreement.

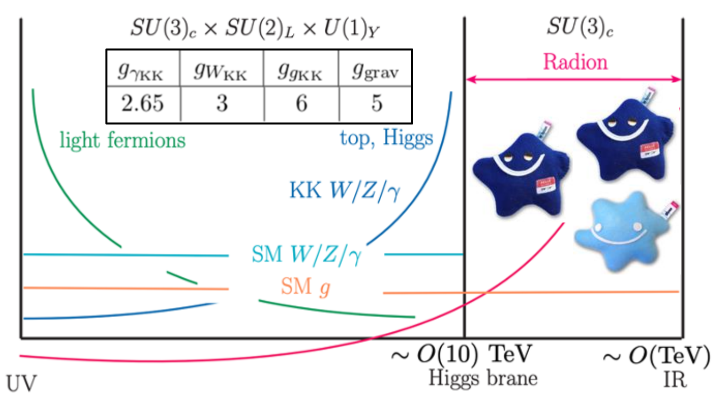
Interpretation: σB & m_{gKK}-m_R limits

- We set upper limits, at 95% CL, on σB, and lower limits on m_{gKK}-m_R masses plane:
- Expected and observed in agreement within ~0.5σ.





- First ever limits of this kind: EWED with QCD in bulk & gWW resonant...



- Read our full paper here for more.
- Visit the CMS B2G public results page and see our summary plots for more.