SEARCHES FOR BSM PARTICLES IN BOOSTED FINAL STATES AT CMS

Lauren Hay on behalf of the CMS Collaboration

BOOST Workshop (July 31st - Aug. 4th, 2023)

Introduction

Many BSM hypotheses predict heavy resonances that decay into final states with high boost in the CMS detector

- Excited quarks ($b*$)
- Spin-1 vector boson resonances (W′)
- Scalar diquark/VLQ
- RPV SUSY (squarks, higgsinos, gluinos)
- Randall-Sundrum warped extra dimensions (G_{KK}, ϕ)

This talk summarizes some of the latest CMS results on such resonance searches with Run II data

$b* \rightarrow tW$ (semileptonic top)	CMS-B2G-21-005
$W' \rightarrow tb$ (semileptonic top)	CMS-B2G-20-012
Paired dijet resonances (*)	CMS-EXO-21-010
High mass resonance —> boosted resonance + jet	CMS-EXO-20-007
Pair produced multijet resonances using data scouting	CMS-EXO-21-004

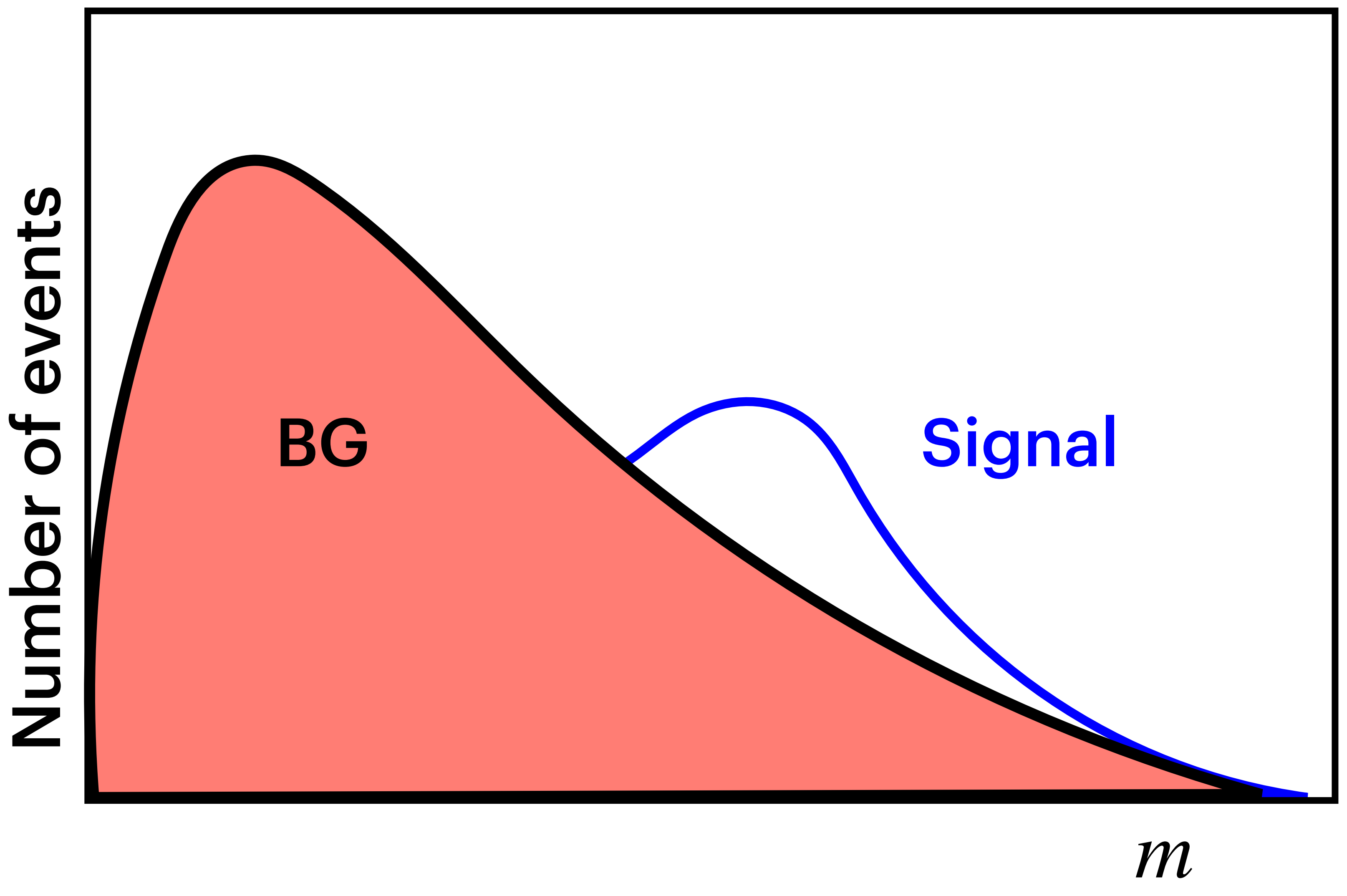
Full list of public results:

Exotica, B2G

Searches utilize a variety of techniques to distinguish signal and background

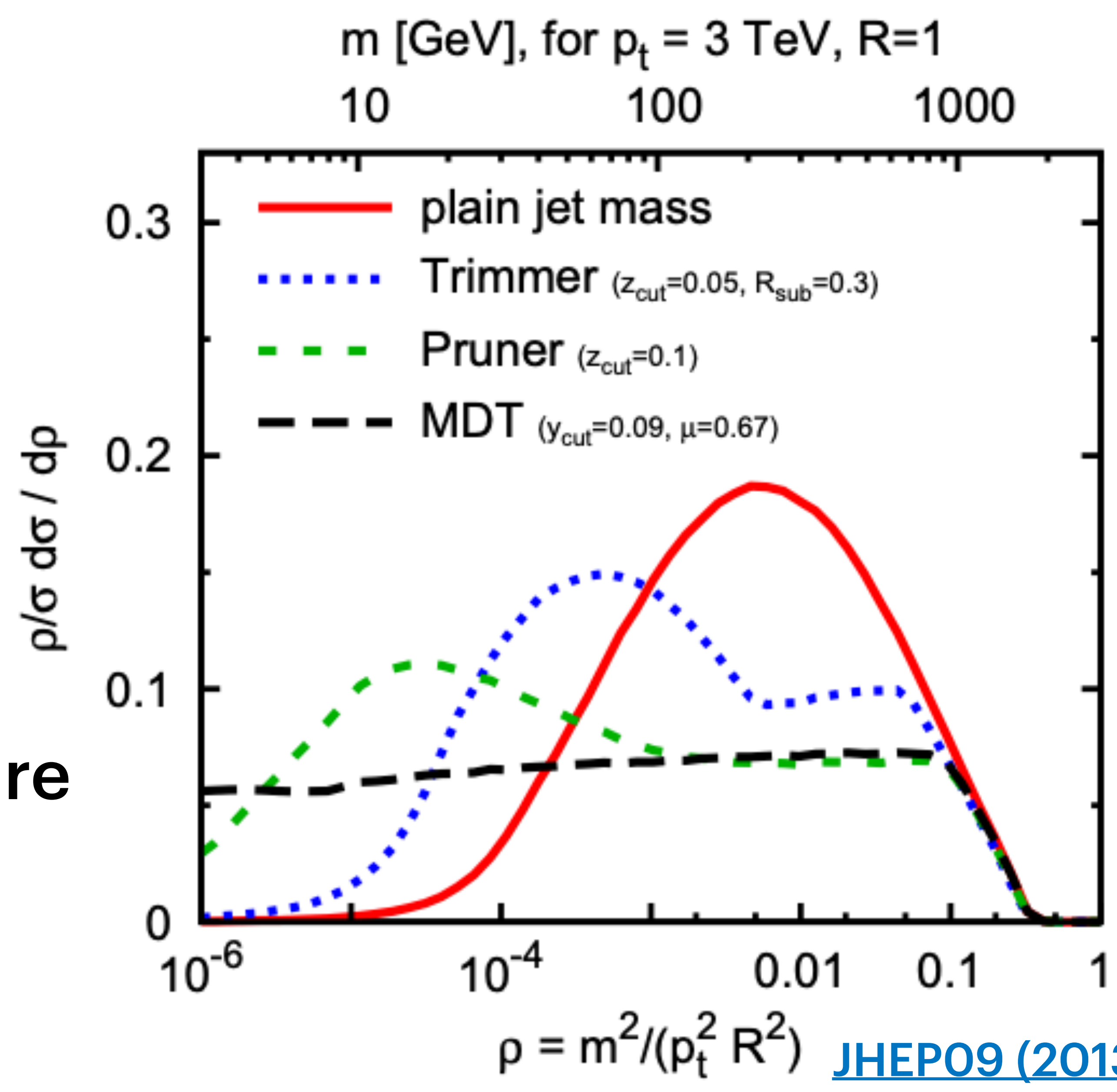
Data driven estimation of background

- Binned maximum likelihood fit
  - ABCD
- Gaussian process regression

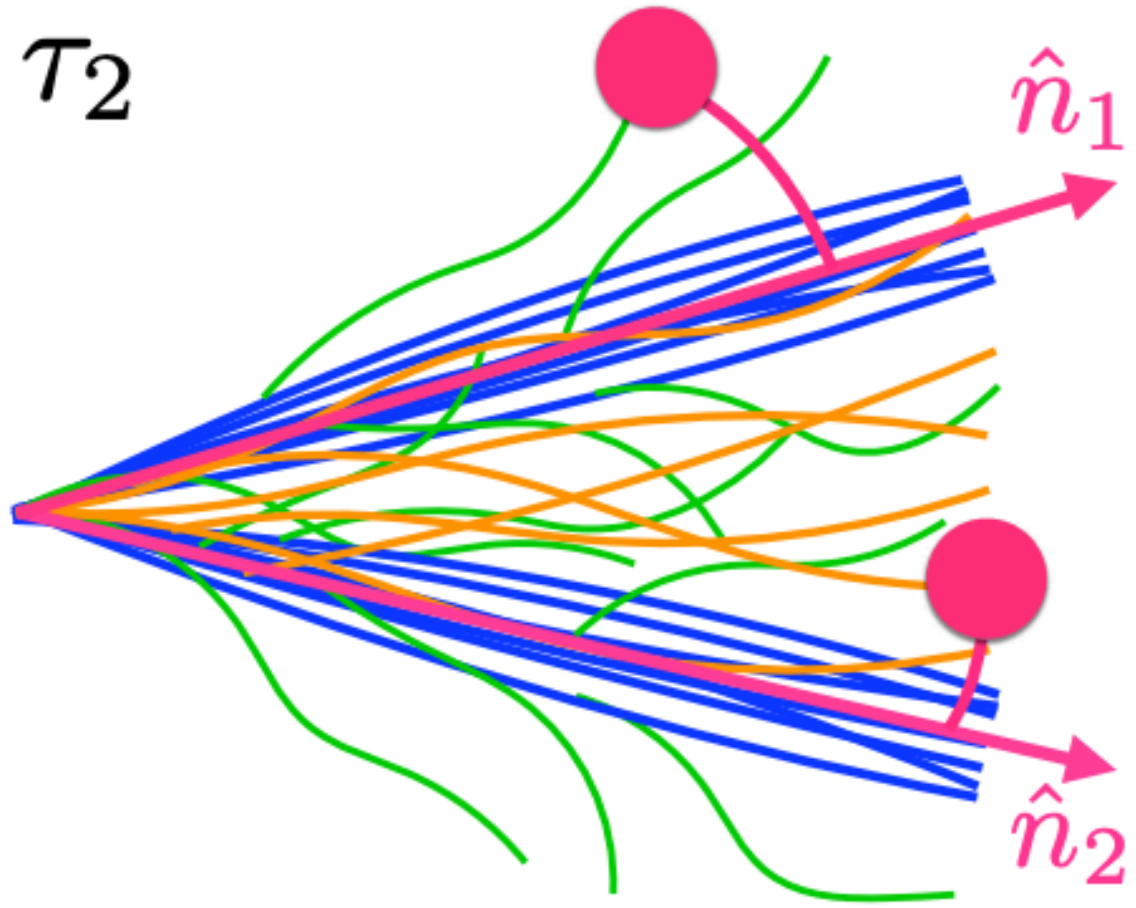


To select signal and background regions

- Removal of soft, non-FSR radiation
  - Soft drop
  - Jet Trimming
- Classification based on pronged substructure
  - N-subjettiness
    - τ_{21} — two-pronged (W, ϕ)
    - τ_{32} — three pronged $(\tilde{h}, \tilde{g})$
  - Energy correlation observable N_2
- Designed Decorrelated Tagger (DDT)
  - Reduces sculpting of jet mass
- DNN classifiers for b tagging and QGL



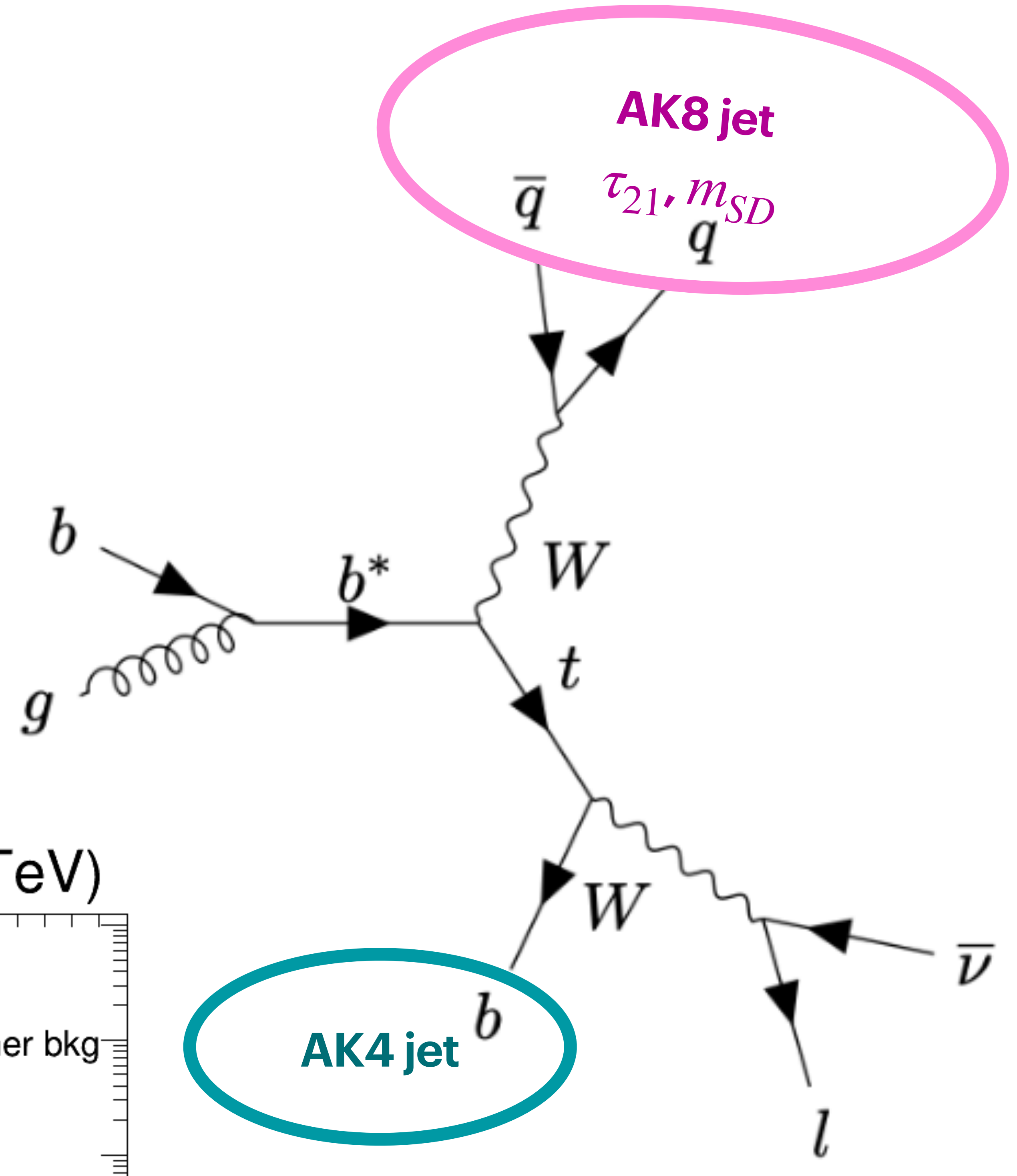
JHEP09 (2013) 029



JHEP 12 (2016) 153

Event selection

- boosted t $\rightarrow$ non-isolated ℓ, MET, and AK4 b jet
- boosted W $\rightarrow$ large radius jet w/ 2 pronged structure $\tau_{21} < 0.4, 0.45$ (2016, 2017-2018) and $65 < m_{SD} < 105 GeV$

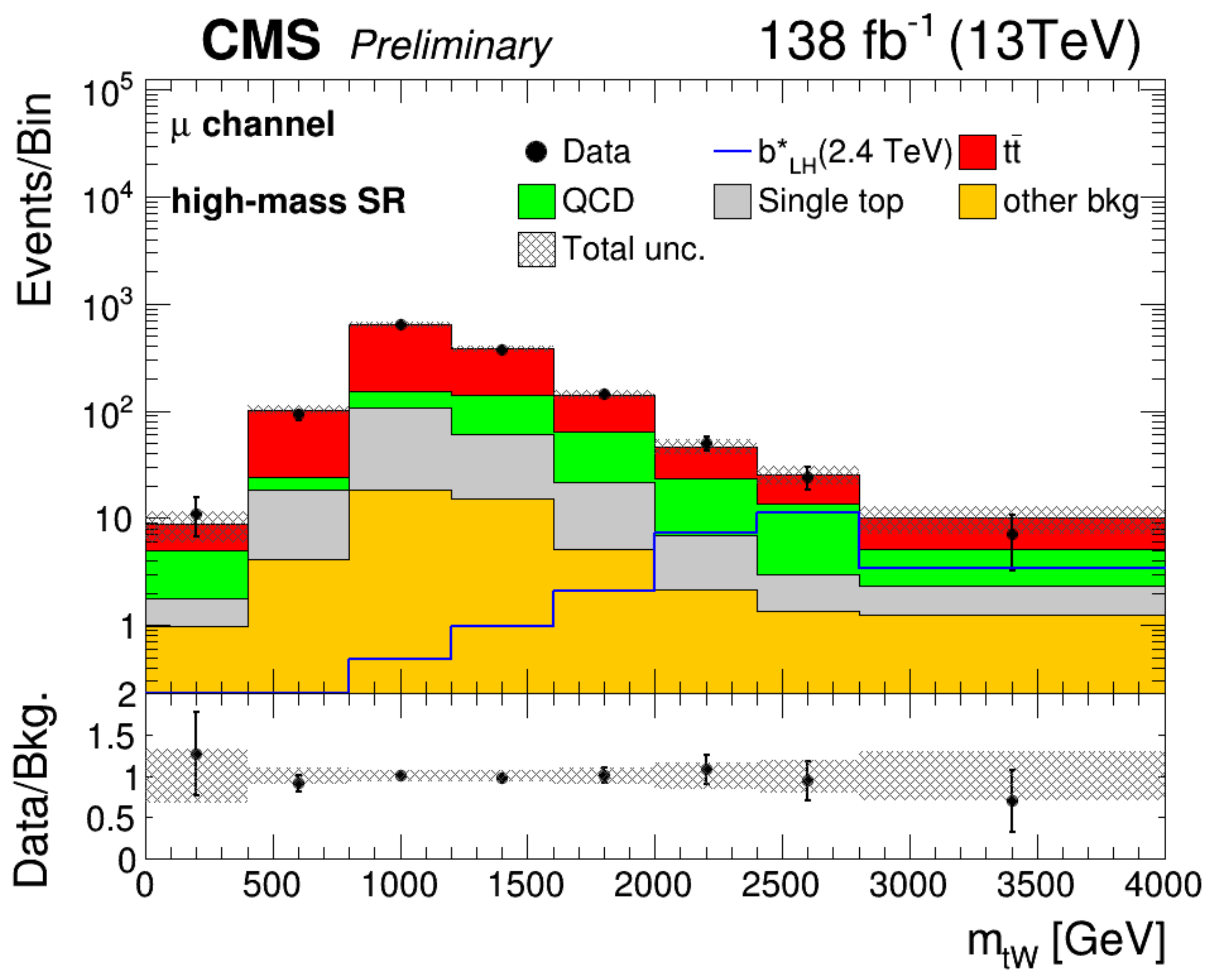


Background estimation and signal extraction

Background is estimated using sideband extrapolation and fit to m_{tW}

- TTbar CR established using algorithm similar to W jet
- Minor backgrounds derived from MC
- High mass and low mass SR

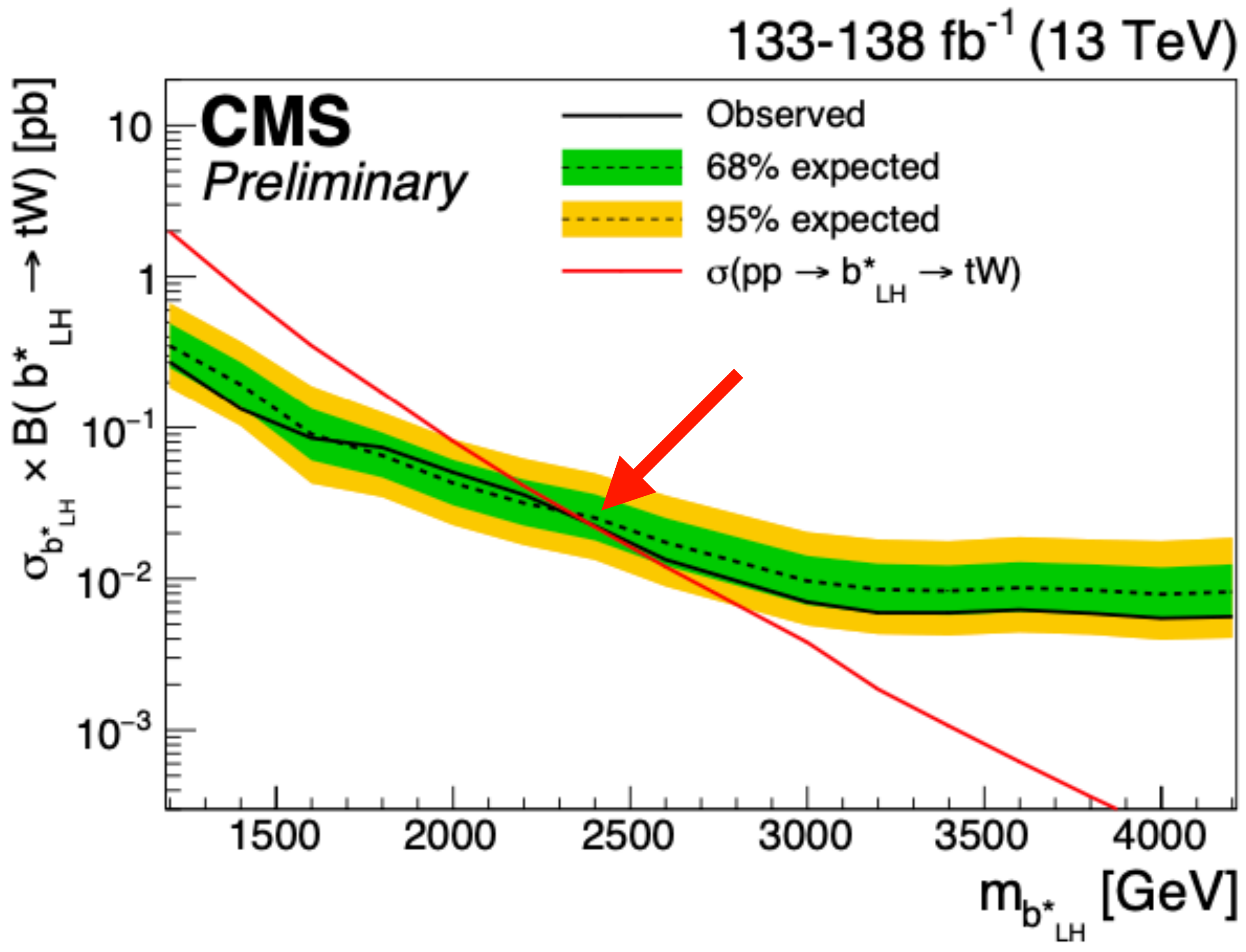
$p_T^W > 400GeV$
$p_T^{miss} > 80GeV$

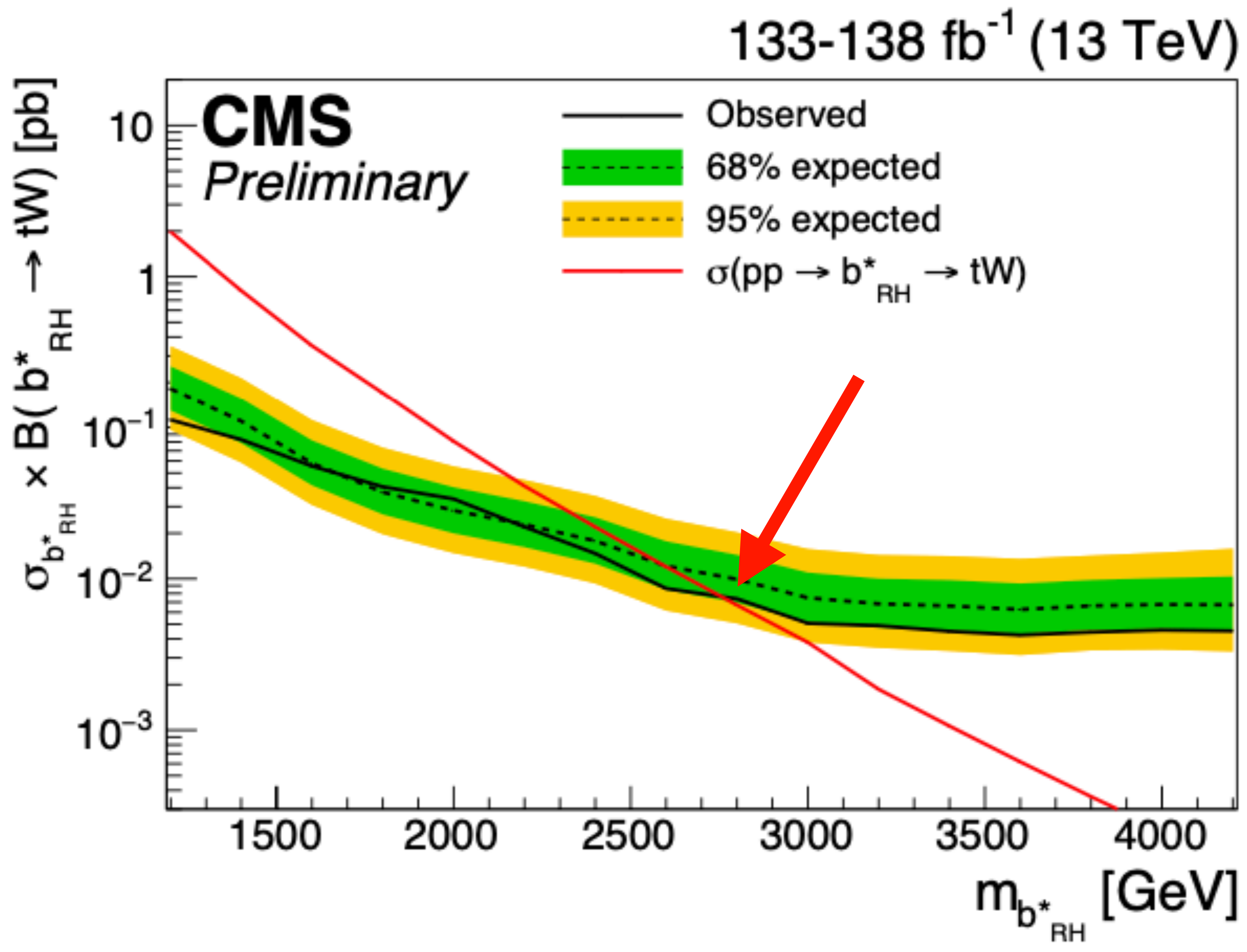


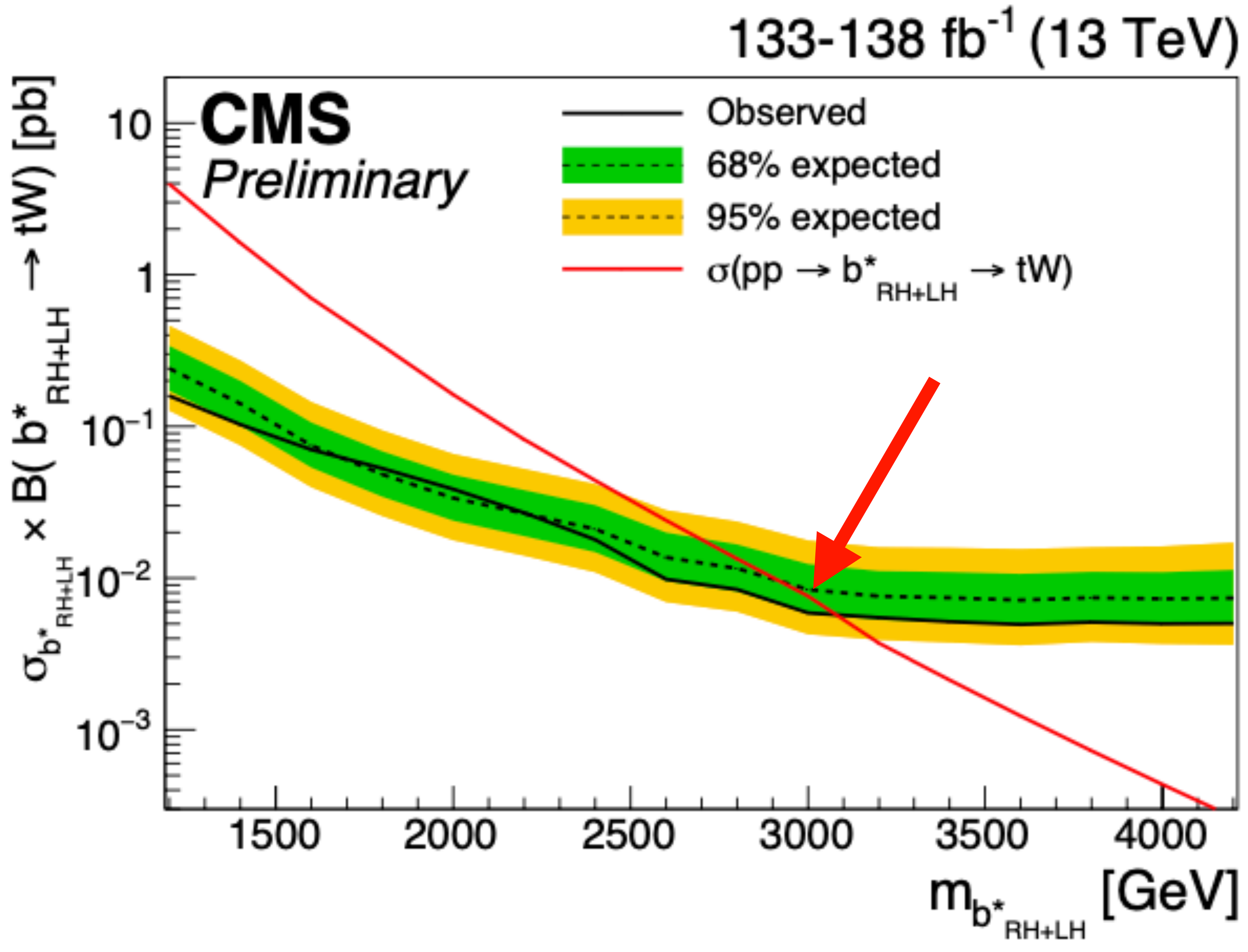
Results

- No statistically significant excess over SM background is observed.

Final state	LH, RH, VL excluded at 95% CL below
all hadronic	2.6, 2.8, 3.1 TeV
lepton + jets (hadronic top)	3.0, 3.0, 3.2 TeV
combined	2.0, 3.0, 3.2 TeV
lepton +jets (leptonic top)	**2.35, 2.77, 3.1 TeV**

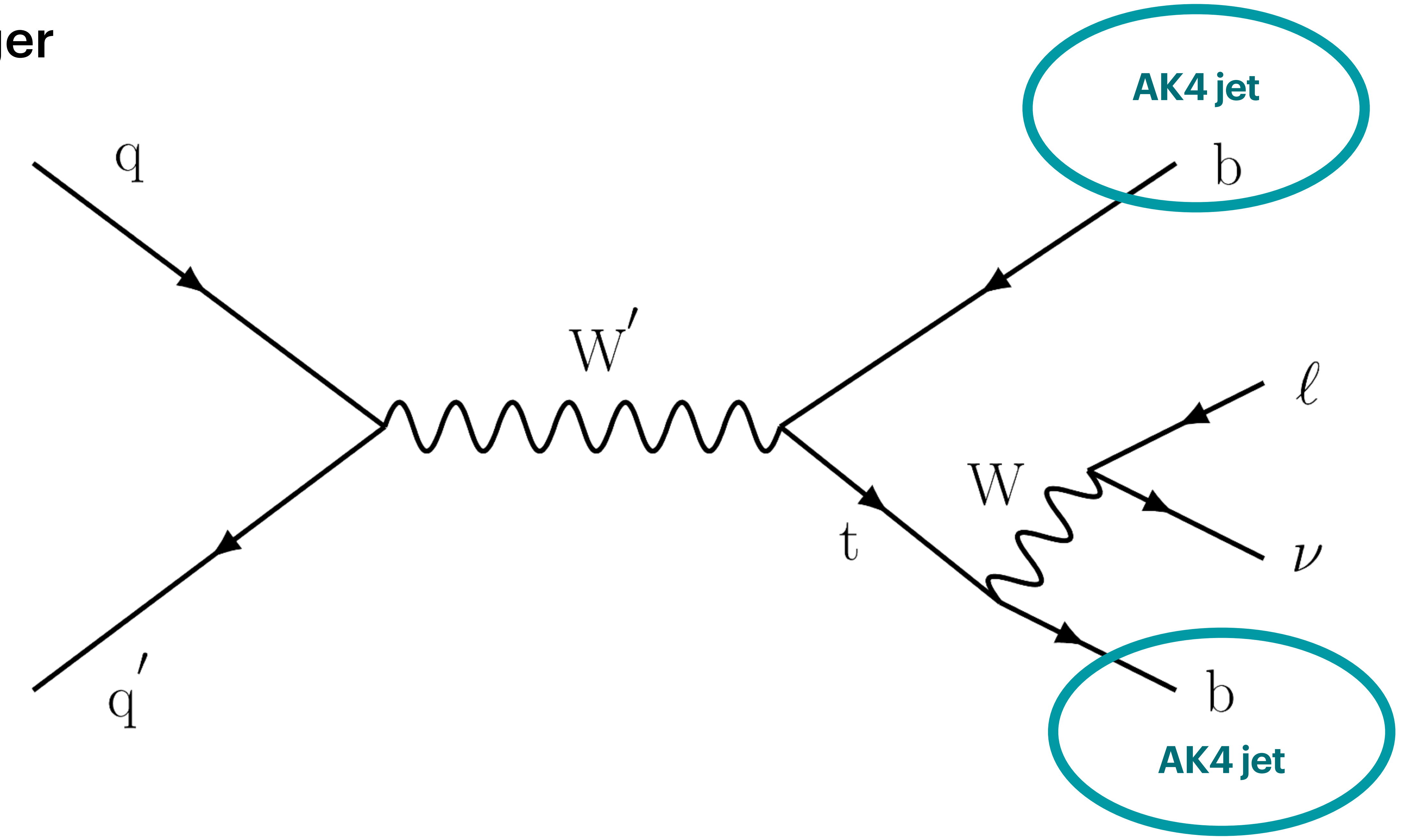






Search for W' decaying to a top and a bottom quark in leptonic final states

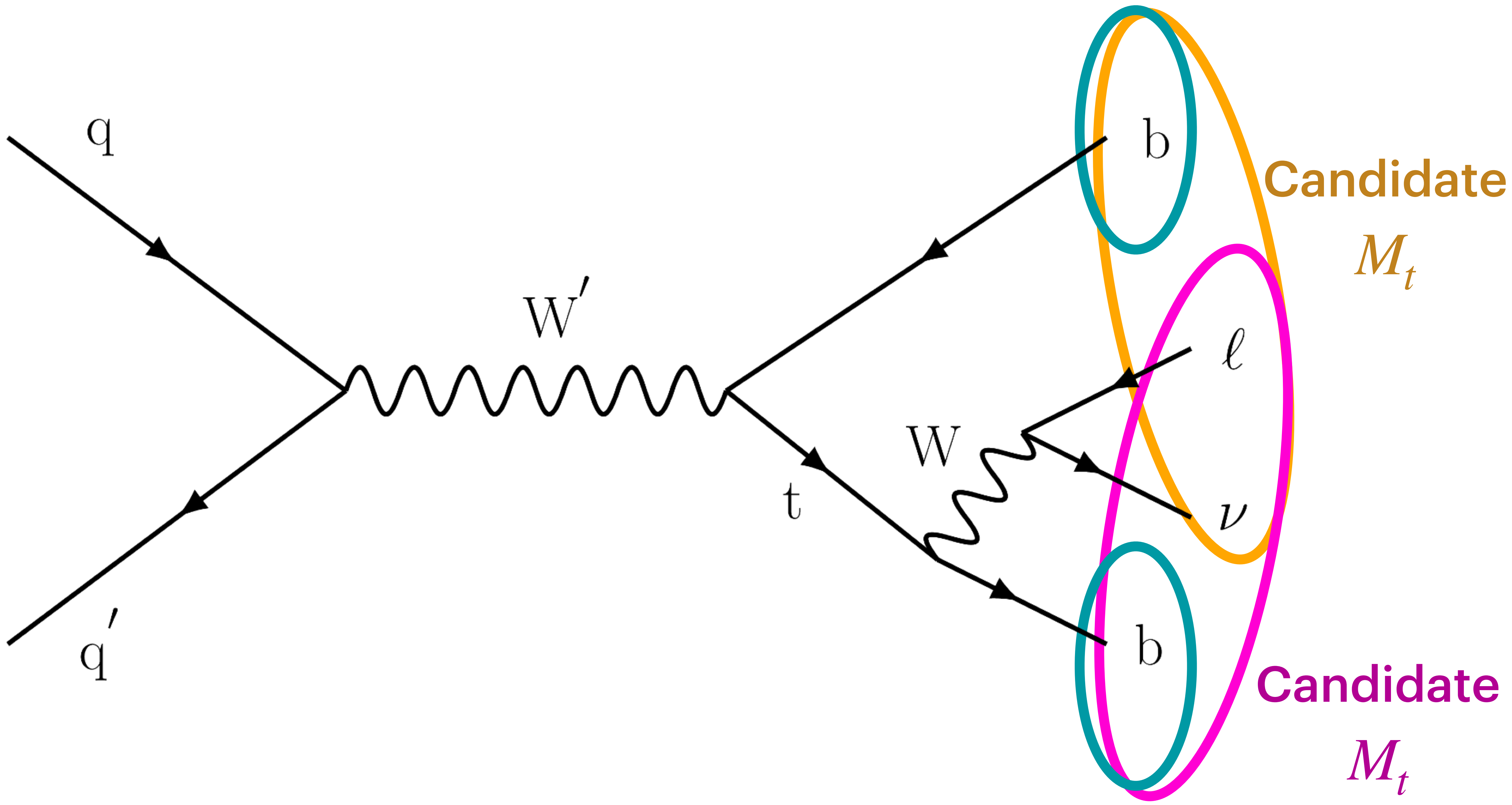
- This analysis probes $\Gamma/m_{W'}$ of 1, 10, 20, and 30%, for the first time
- 1% corresponds to the case where W’ only decays to tb, and larger widths have additional decays



Event selection

- t $\rightarrow \ell$ w/ $I_{mini} < 0.1$, MET, AK4 jet
- W’ jet $\rightarrow$ AK4 jet
- Require at least two AK4 jets and at least two AK8 Jets
- Use AK8 jets to veto hadronic top decays w/ soft drop mass

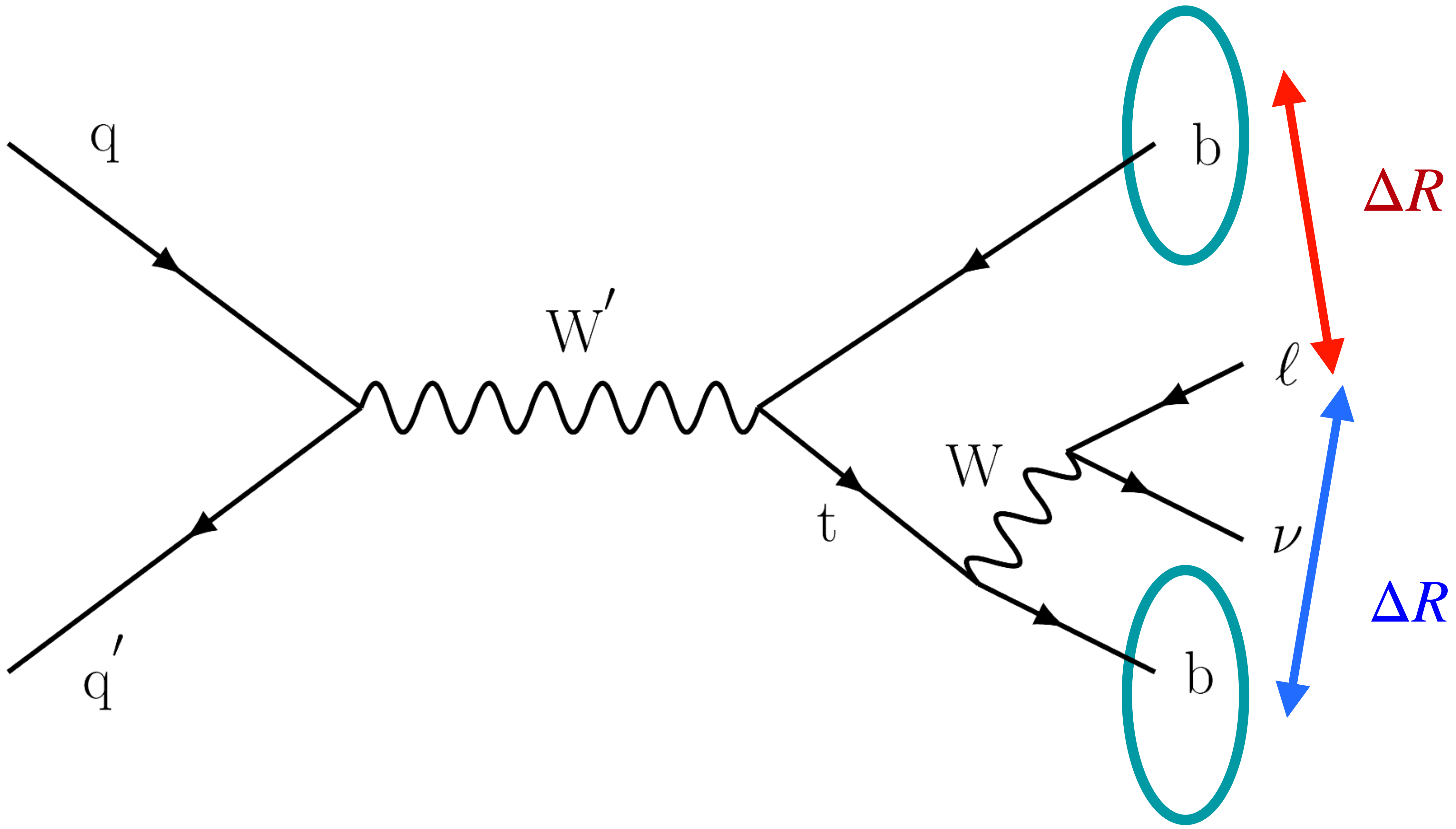
Kinematic reconstruction



j_t and $j_{W'}$ assignment strategy

- Mass criterion
  - choose jet whose invariant mass sum w/ candidate $\ell\nu$ minimizes $|M_t - M_t^{nom}|$ for j_t
- Closest criterion
  - choose the jet closest to the lepton as j_t
- Subleading criterion
  - expect leading jet to come directly from the W' boson $\rightarrow$ choose subleading jet as j_t
- If jet passes 2 of these criteria $\rightarrow j_t$; otherwise, preference given to the mass criteria.
- $j_{W'} \rightarrow$ highest pT jet not identified as top jet
- If there are < 2 b-tagged jets, criteria are applied to the entire jet collection; otherwise, only b-tagged jets are considered

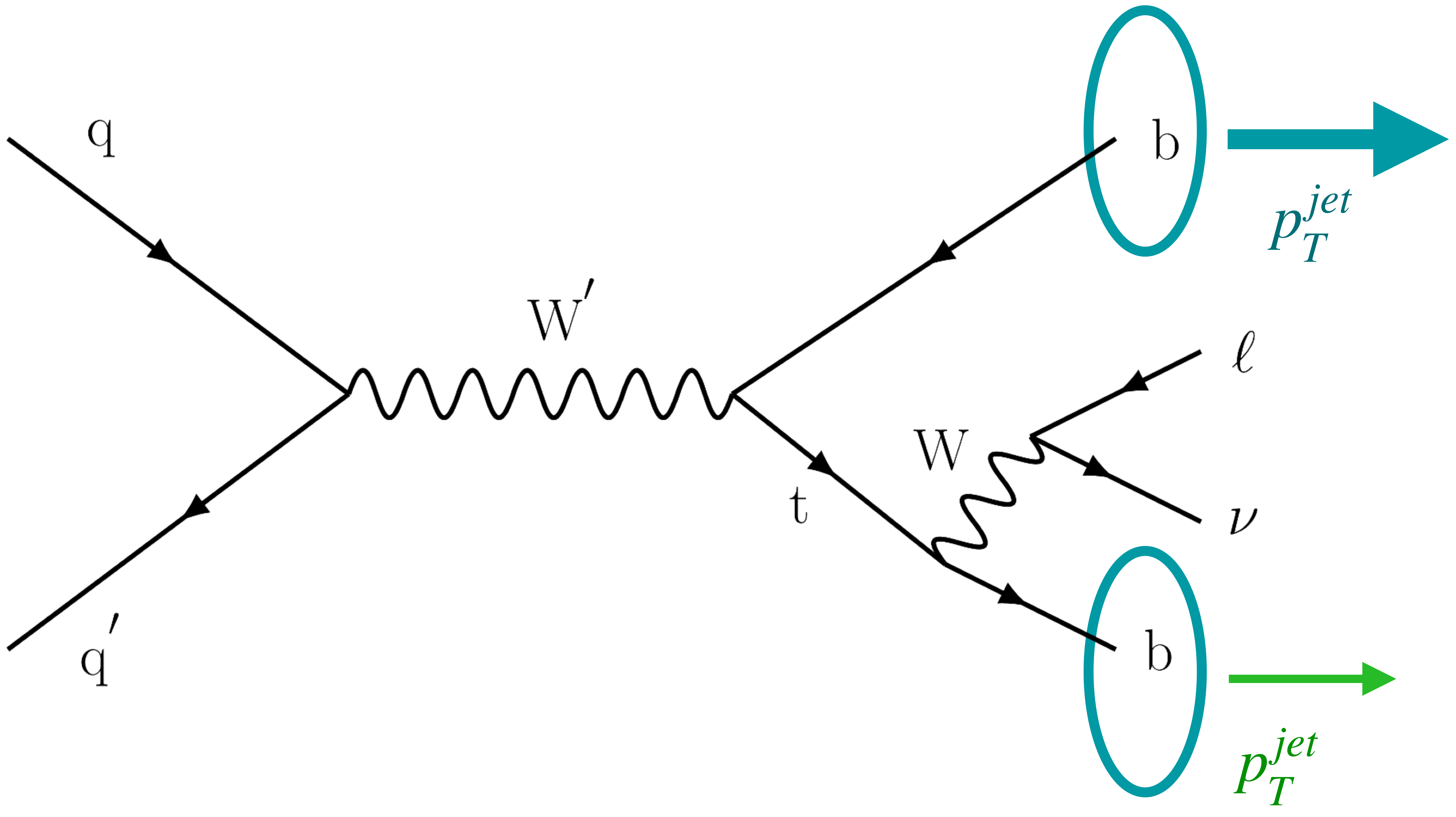
Kinematic reconstruction



j_t and $j_{W'}$ assignment strategy

- Mass criterion
  - choose jet whose invariant mass sum w/ candidate $\ell\nu$ minimizes $|M_t - M_t^{nom}|$ for j_t
- Closest criterion
  - choose the jet closest to the lepton as j_t
- Subleading criterion
  - expect leading jet to come directly from the W' boson → choose subleading jet as j_t
- If jet passes 2 of these criteria $\rightarrow j_t$; otherwise, preference given to the mass criteria.
- $j_{W'} \rightarrow$ highest pT jet not identified as top jet
- If there are < 2 b-tagged jets, criteria are applied to the entire jet collection; otherwise, only b-tagged jets are considered

Kinematic reconstruction

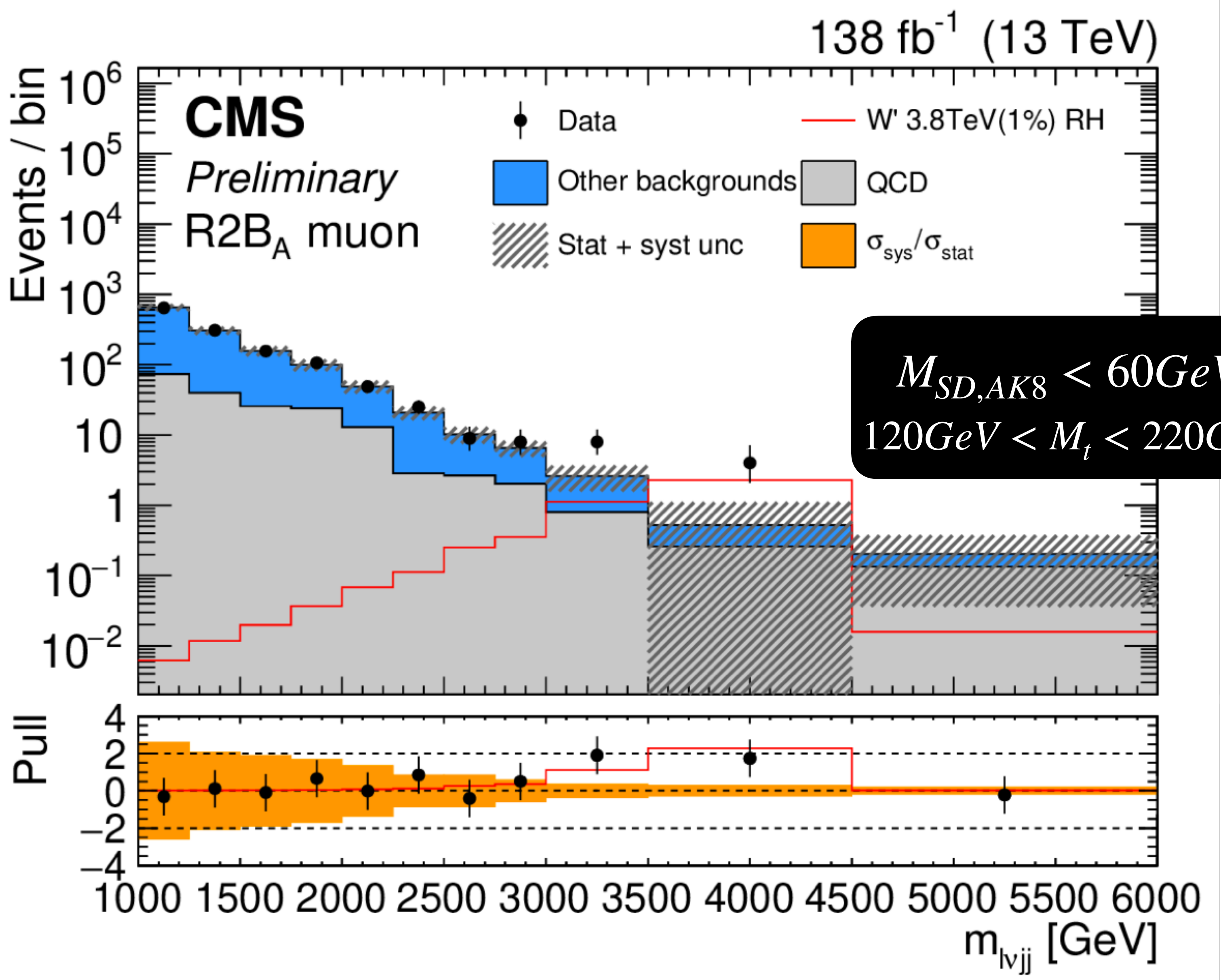


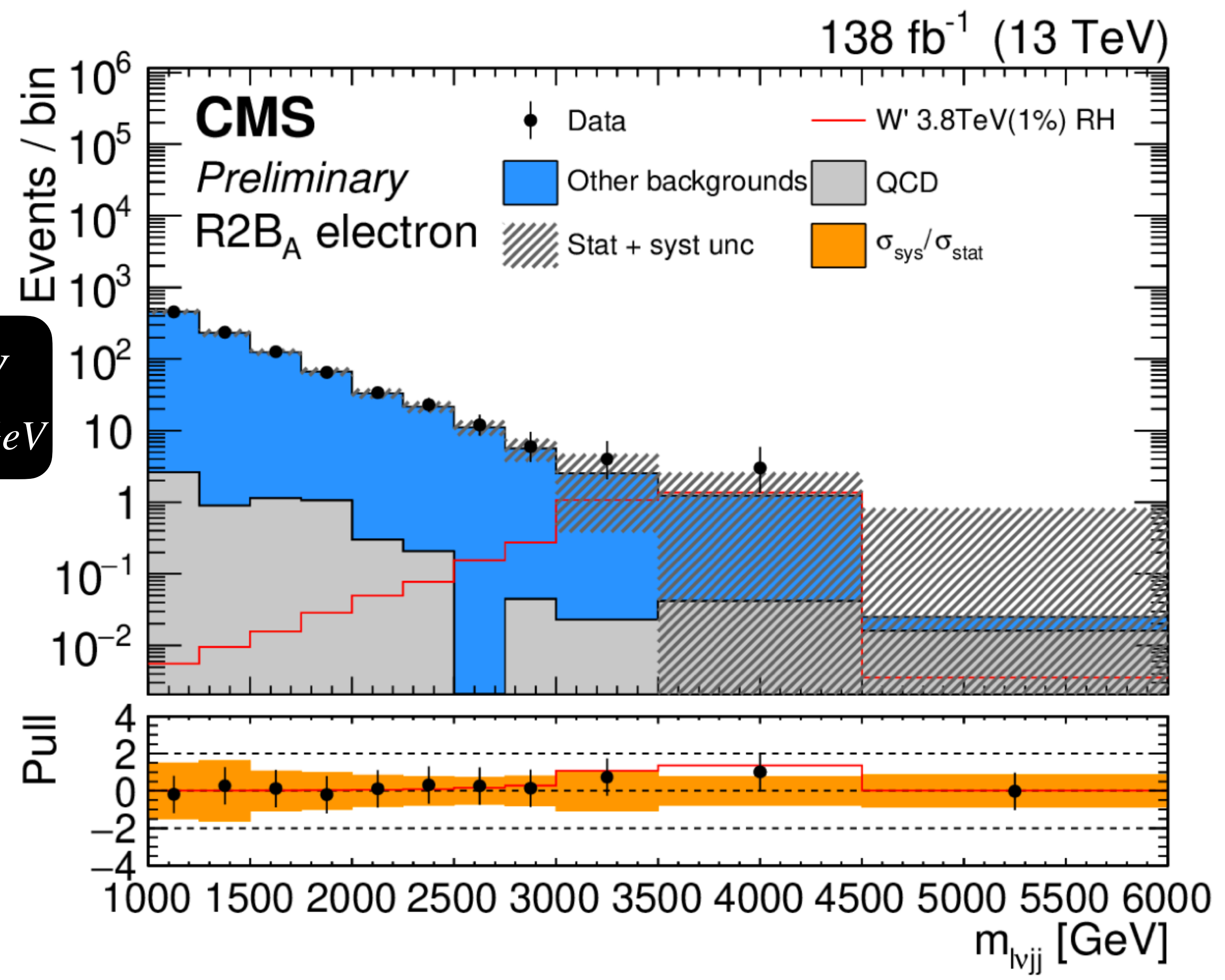
j_t and $j_{W'}$ assignment strategy

- Mass criterion
  - choose jet whose invariant mass sum w/ candidate $\ell\nu$ minimizes $|M_t - M_t^{nom}|$ for j_t
- Closest criterion
  - choose the jet closest to the lepton as j_t
- Subleading criterion
  - expect leading jet to come directly from the W' boson $\rightarrow$ choose subleading jet as j_t
- If jet passes 2 of these criteria $\rightarrow j_t$; otherwise, preference given to the mass criteria.
- $j_{W'} \rightarrow$ highest pT jet not identified as top jet
- If there are < 2 b-tagged jets, criteria are applied to the entire jet collection; otherwise, only b-tagged jets are considered

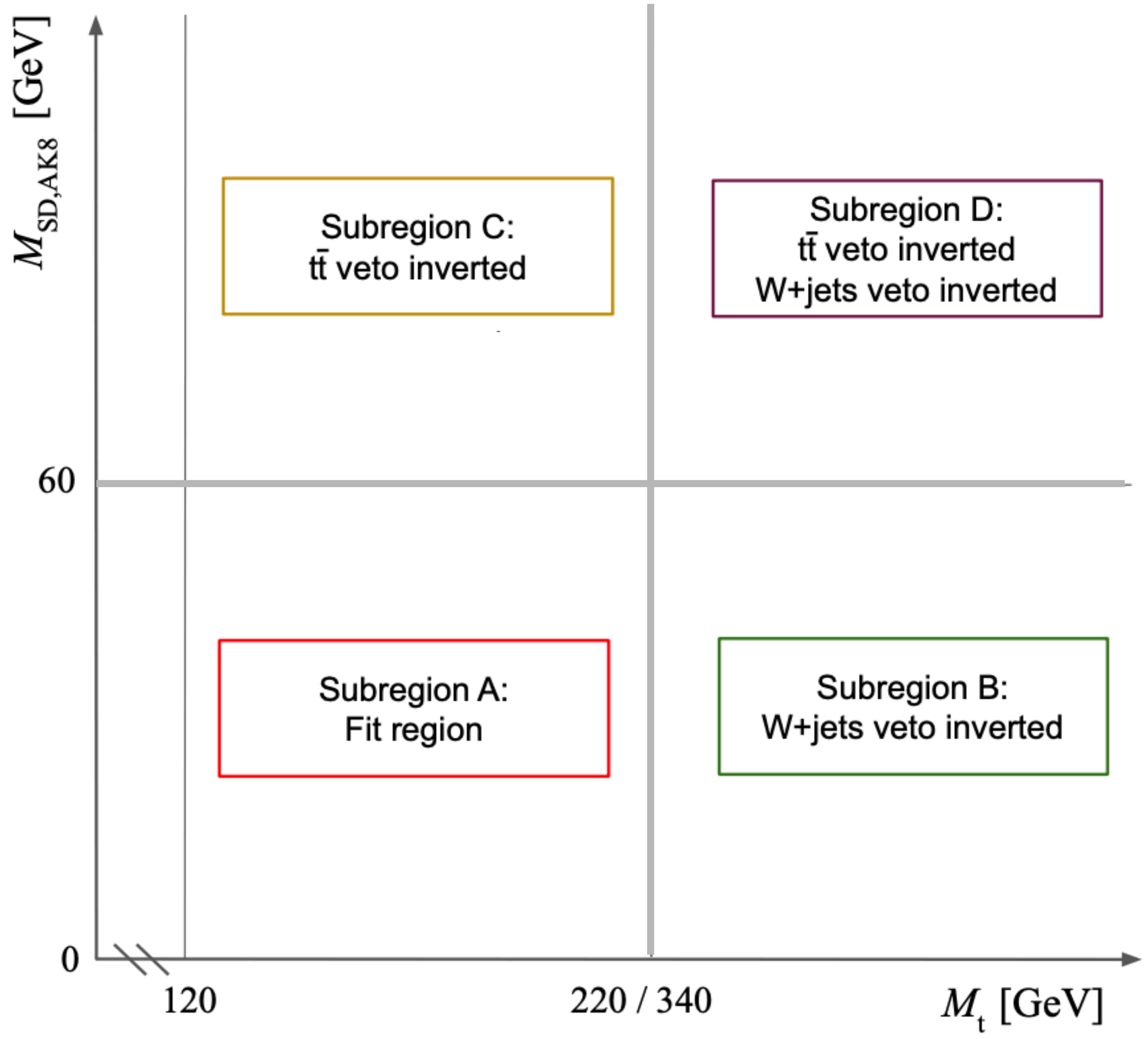
Background estimation and signal extraction

- Background is estimated using the ABCD method
  - ABCD subregions based on cuts to the $M_{SD,AK8}$ matched to $j_{W'}$ and M_t
- A simultaneous fit to $M_{\ell\nu jj}$ in region A is performed
- 3 signal regions: j_t and/or j'_W are b-tagged





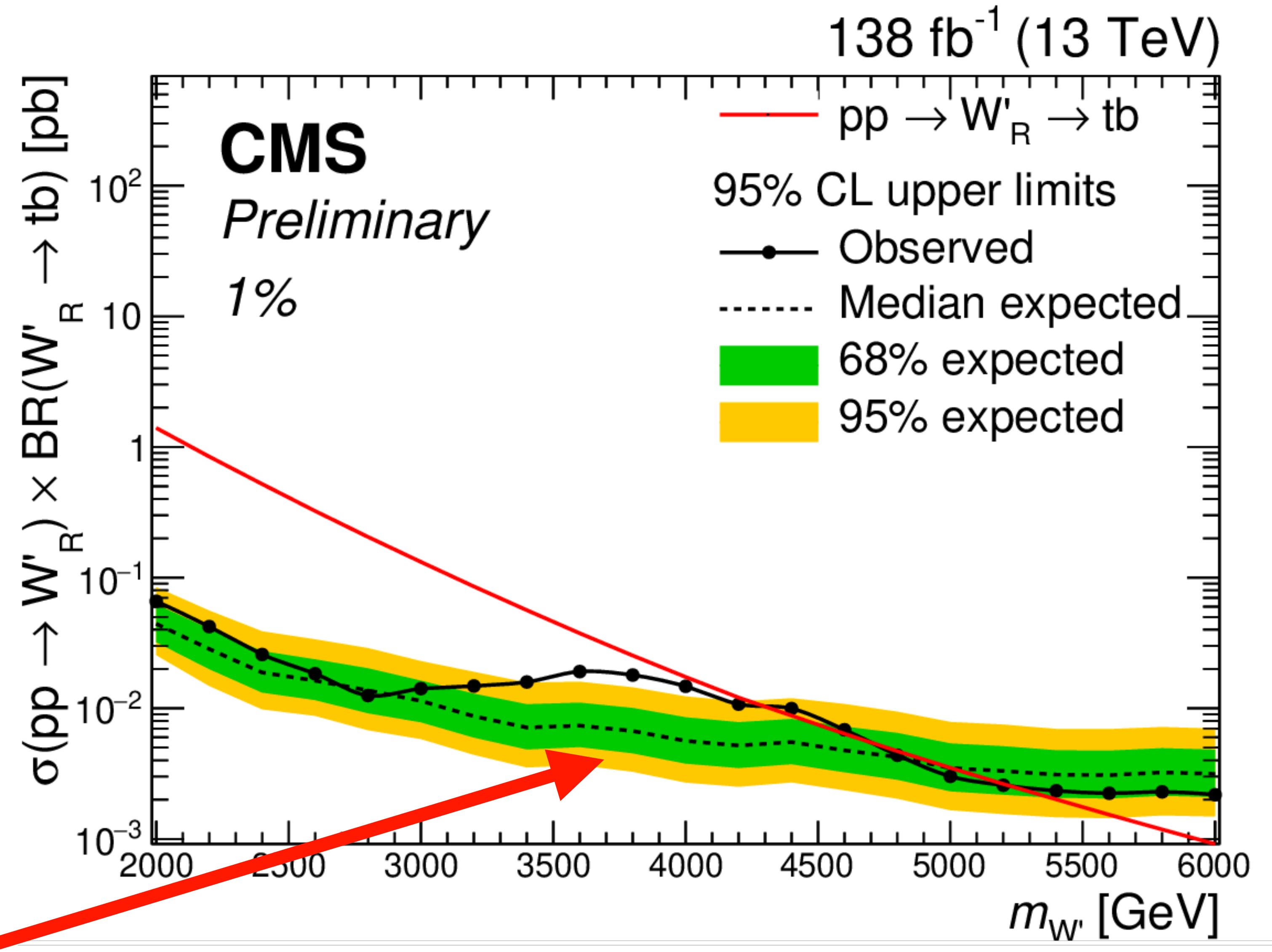
$M_{SD,AK8} < 60GeV$
$120GeV < M_t < 220GeV$

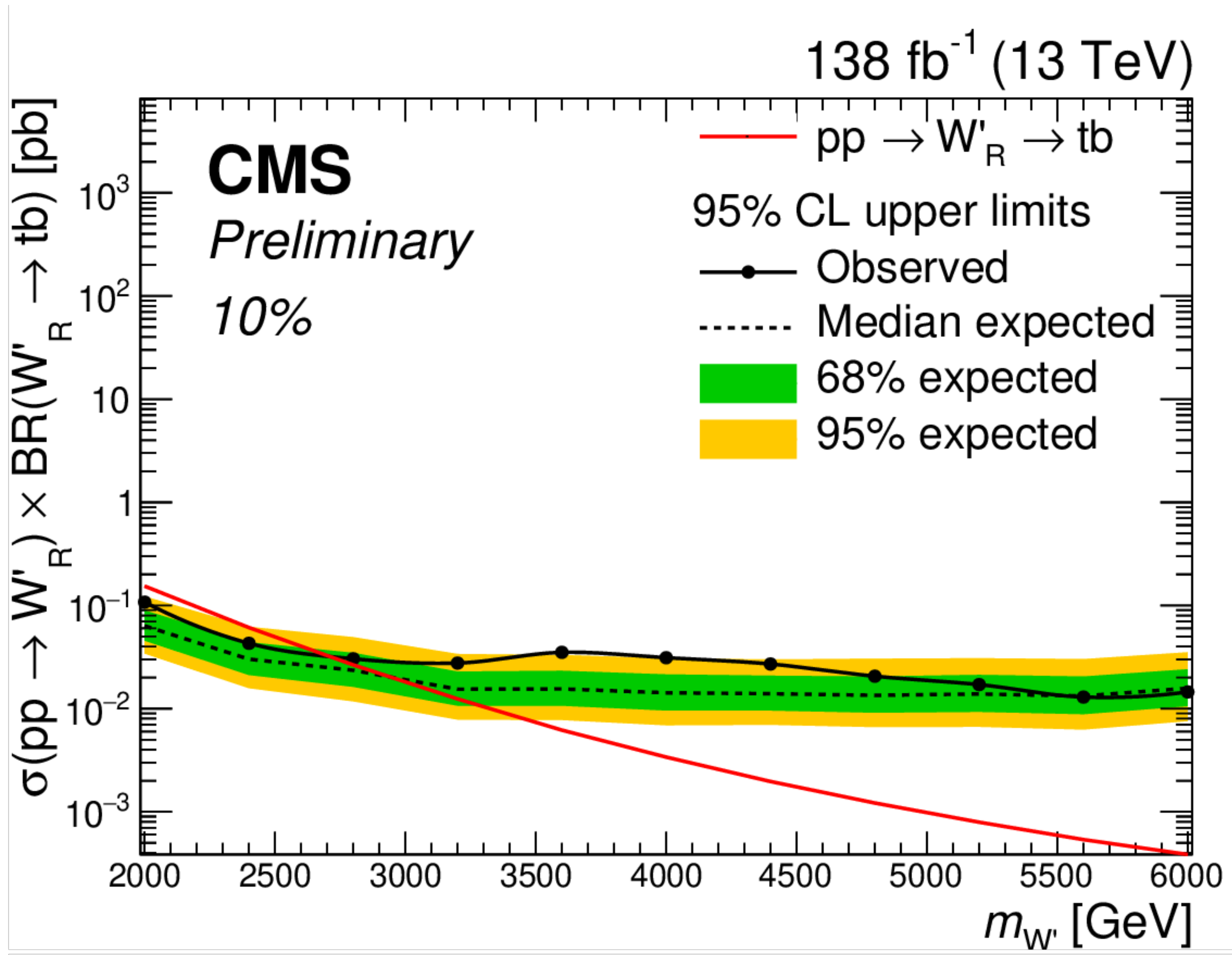


$R2B_A$ SR for 1% width (left: e, right: μ)

Results

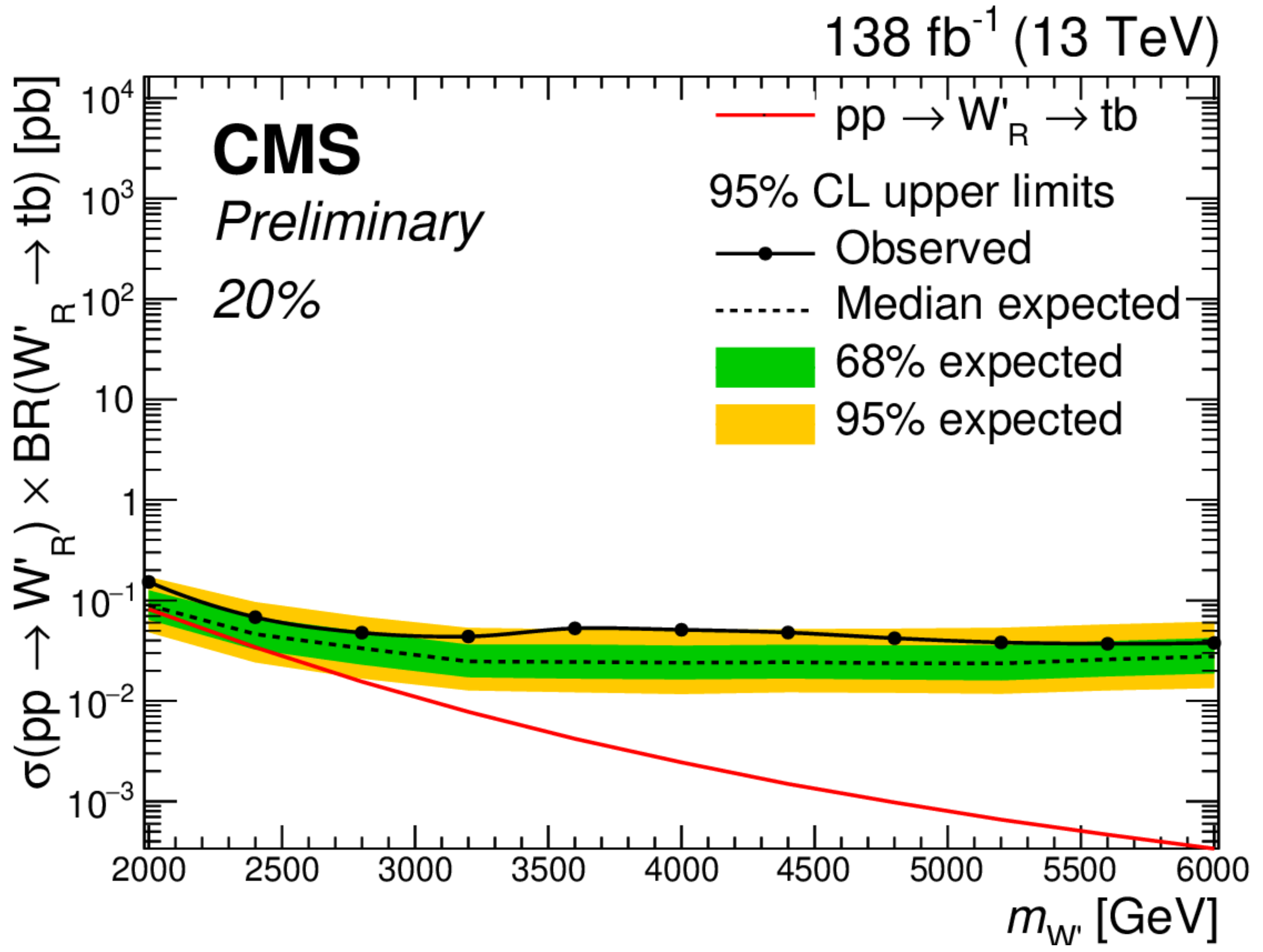
Agreement between data and SM background

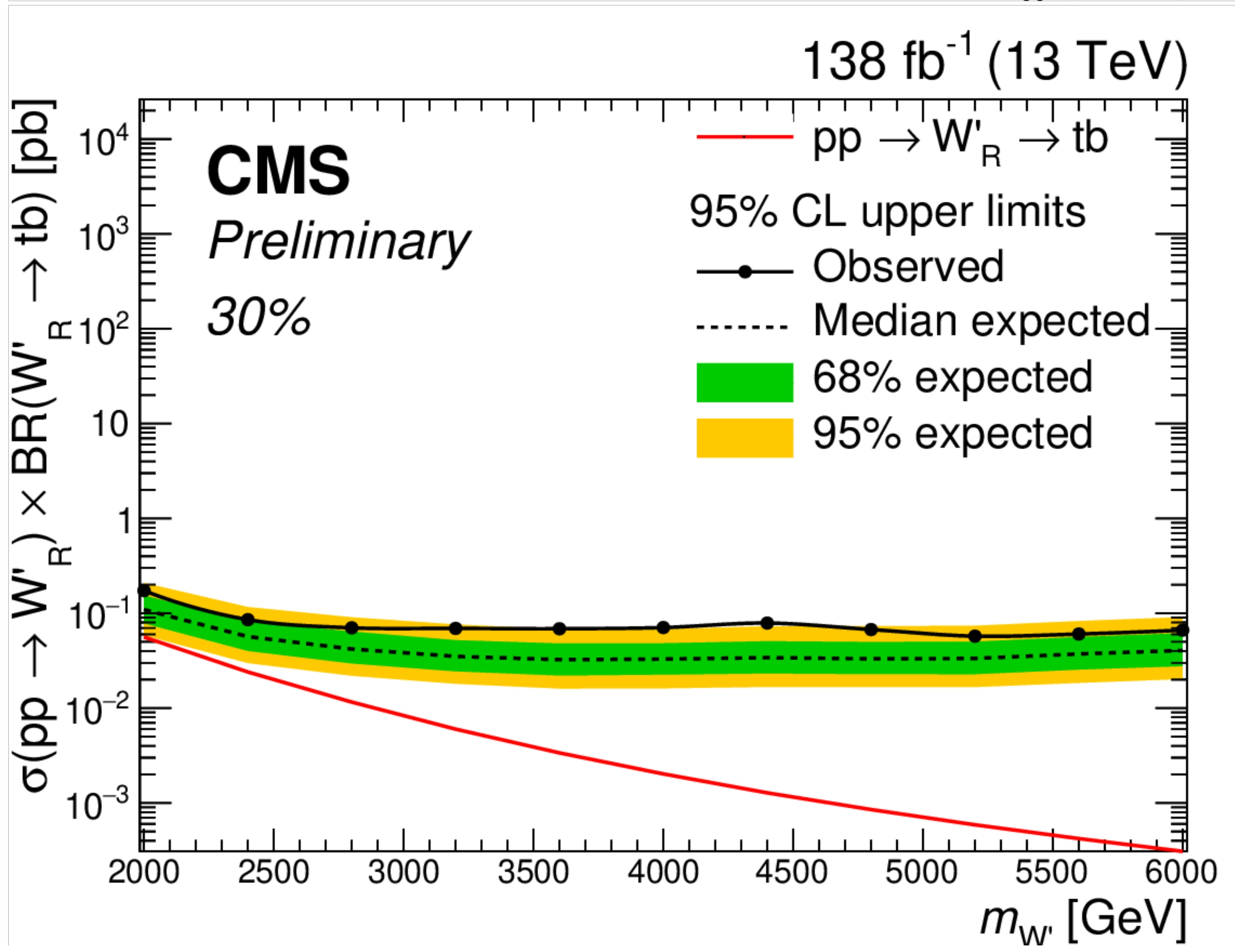




Limits also set for different couplings of $\alpha_{R/L}$ in 1% width (backup)

2.0 sigma global excess

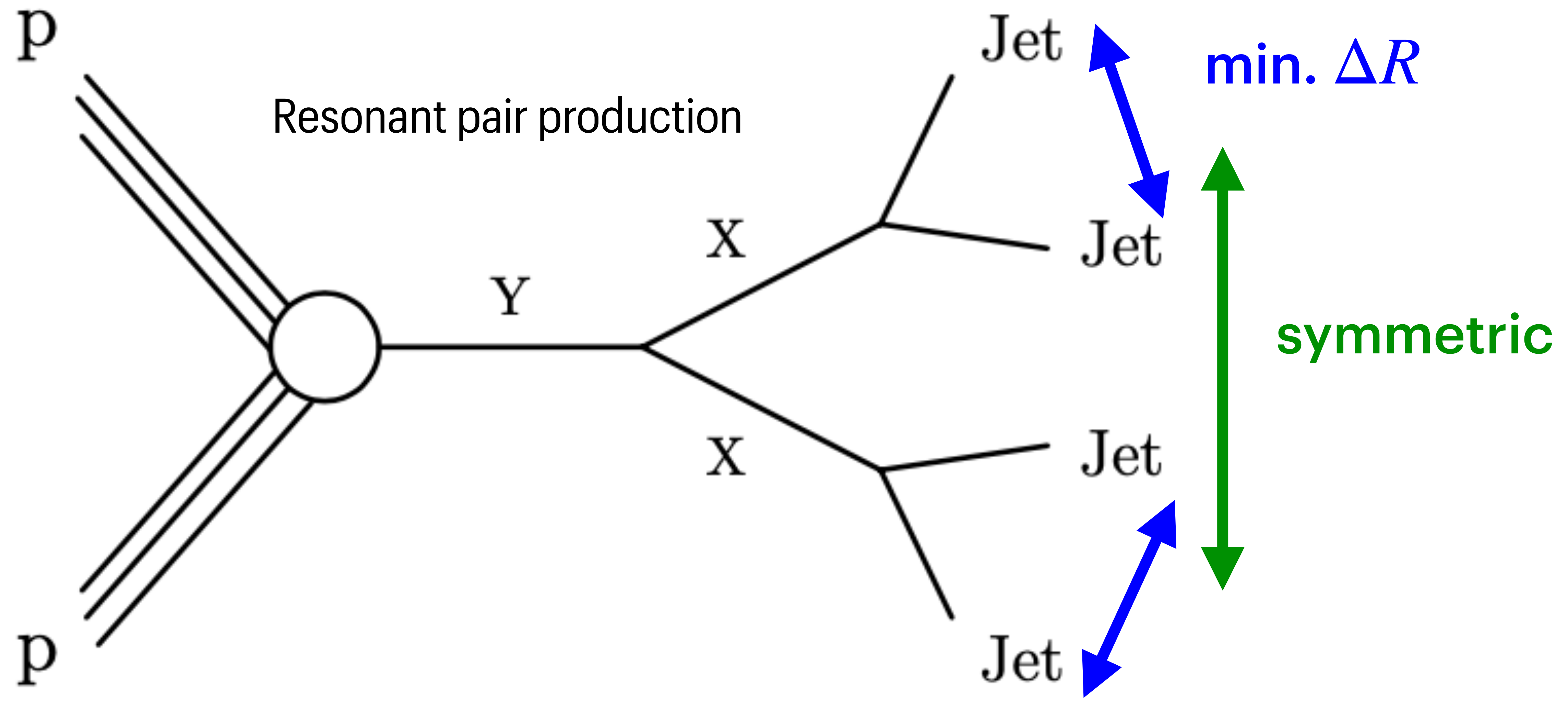


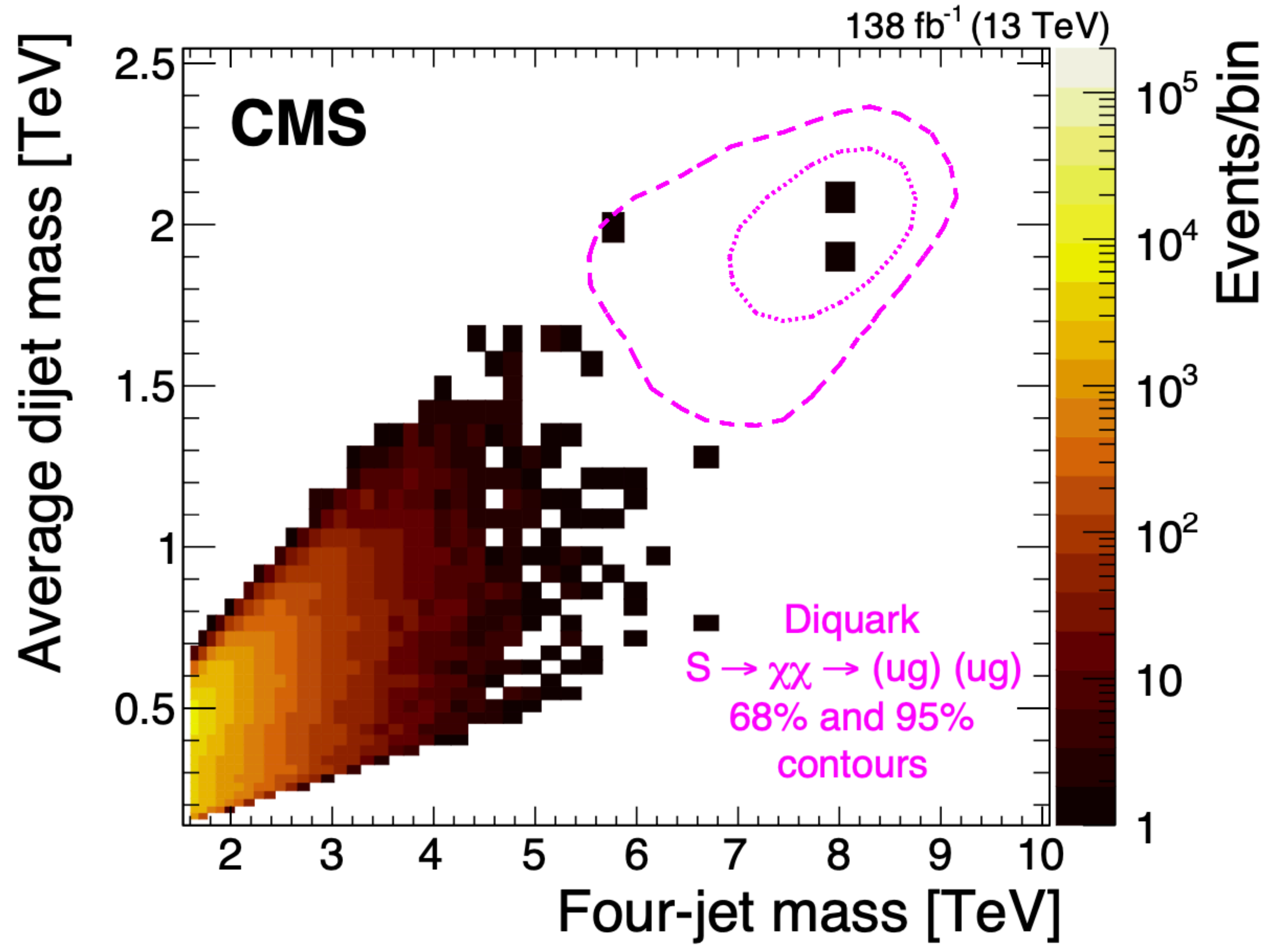


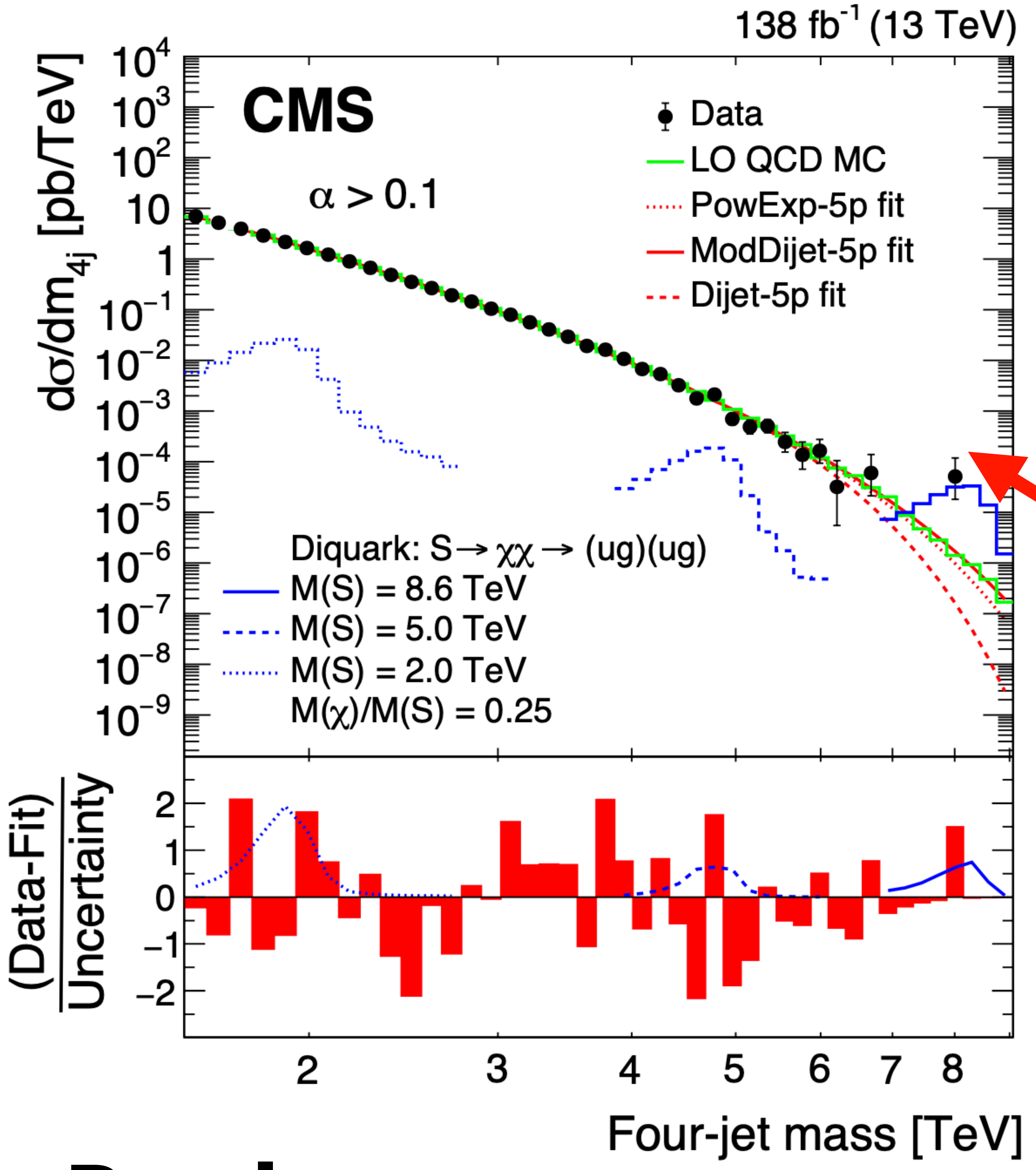
For the first time, limits on production cross section are set for a W’ boson w/ relative decay widths of 20% and 30%

Event selection for resonant and nonresonant production of pairs of dijet resonances

- Resonant benchmark model is scalar diquark to VLQs $uu \rightarrow S \rightarrow \chi\chi \rightarrow (ug)(ug)$
- Nonresonant benchmark is two top squarks $pp \rightarrow \tilde{t}\tilde{t}* -> (\bar{d}\bar{s})(ds)$
- Resonant search — reconstruct the four-jet mass m_{4j}
- Nonresonant search — reconstruct the average dijet mass $\bar{m}_{jj}$
- Near $m_{4j} \sim 8TeV$ and $m_{jj} \sim 2TeV$ observe two candidates for a resonant signal
- Bin observables by $\alpha = \bar{m}_{jj}/m_{4j}$ to remove sculpting







Background estimation and signal extraction

Background is derived exclusively from data

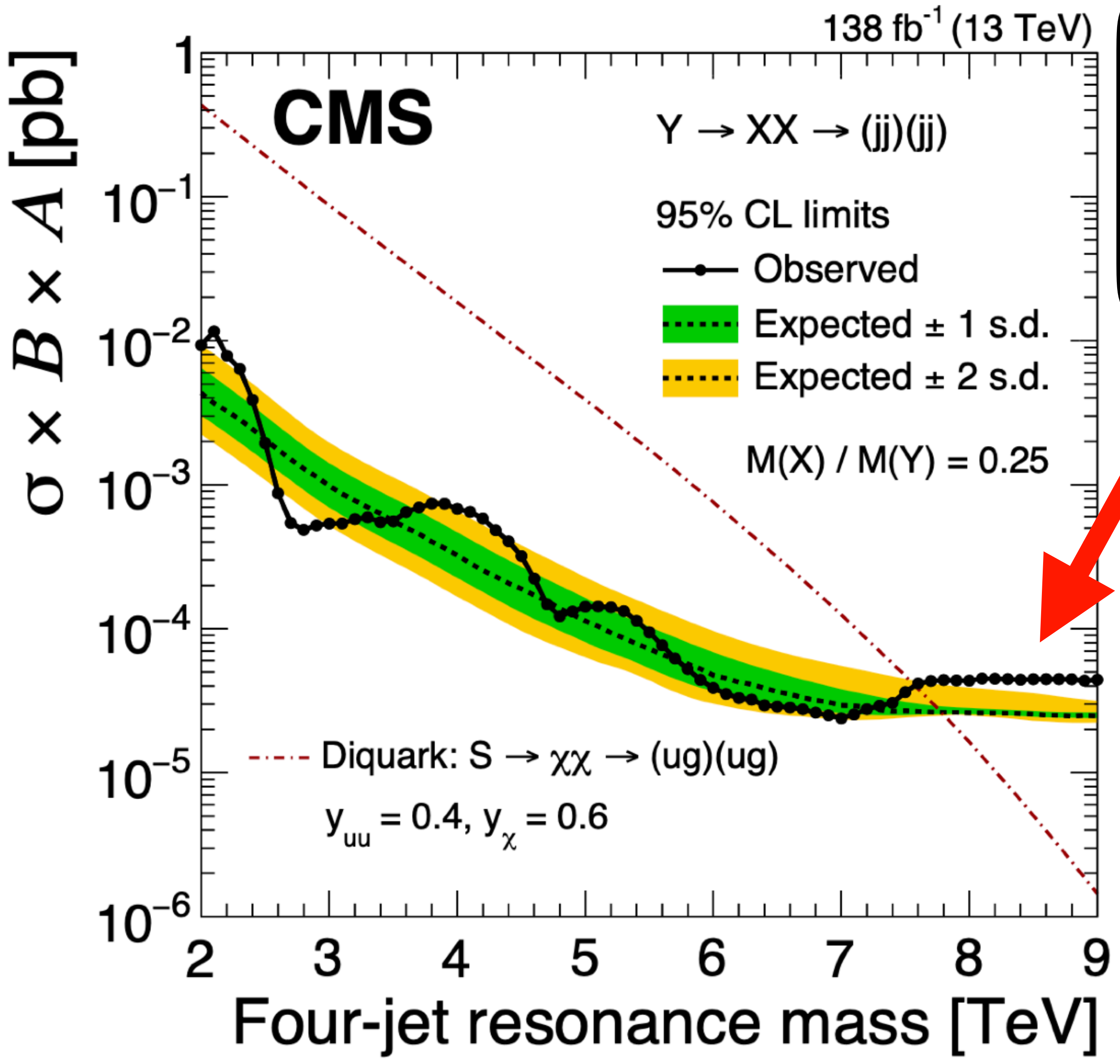
3 parameter smoothly falling spectrum for background:

- 13 bins of α for resonant search
- 3 bins of α in nonresonant search

To include the two events of interest together, a 5 parameter fit is performed to the sum of the alpha bins

Results

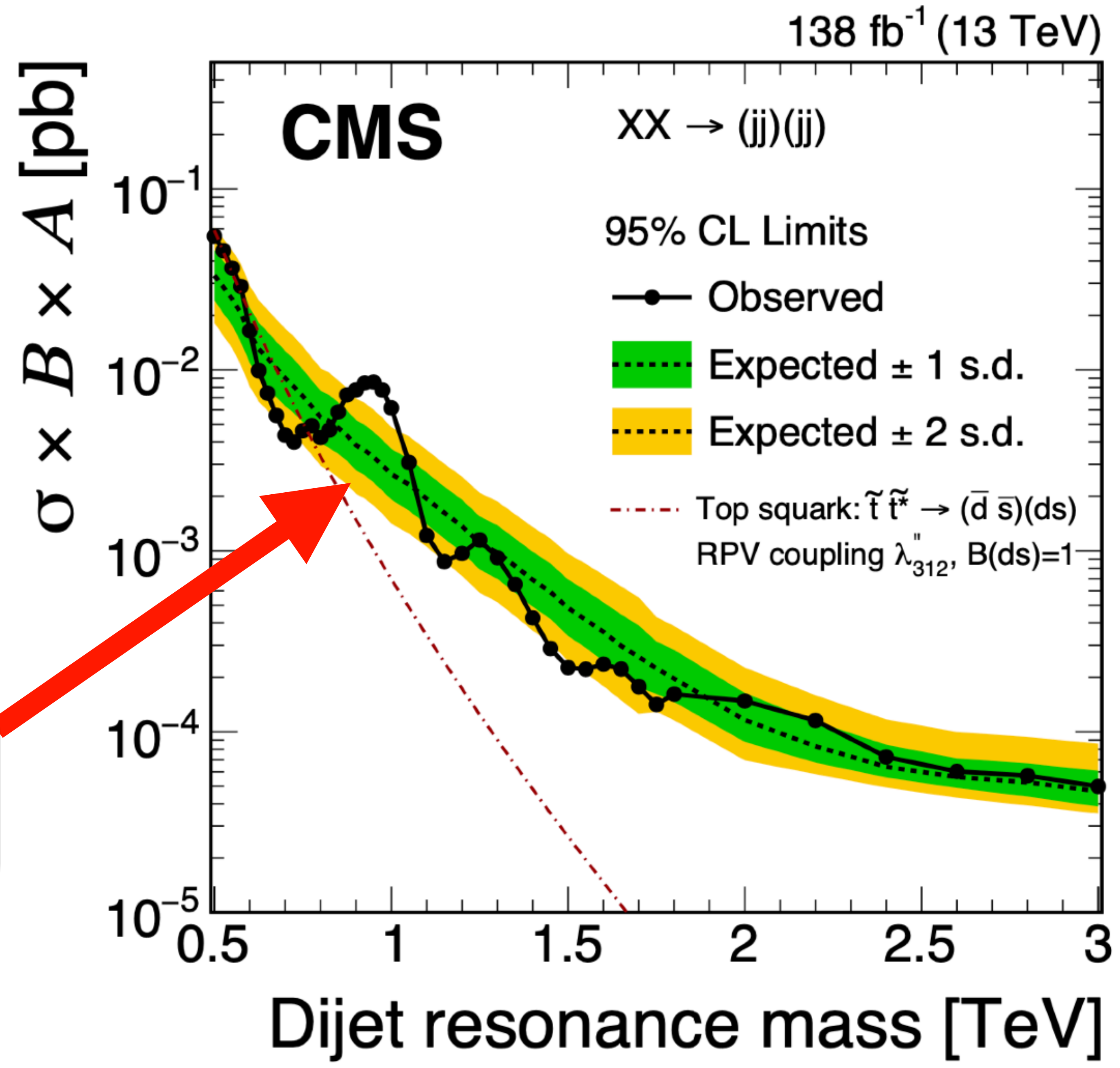
All functions used for the background estimation give a good fit to the data



$\sigma_{loc} = 3.9$

$\sigma_{glob} = 1.6$

$m_{jj} = 2.1 TeV$

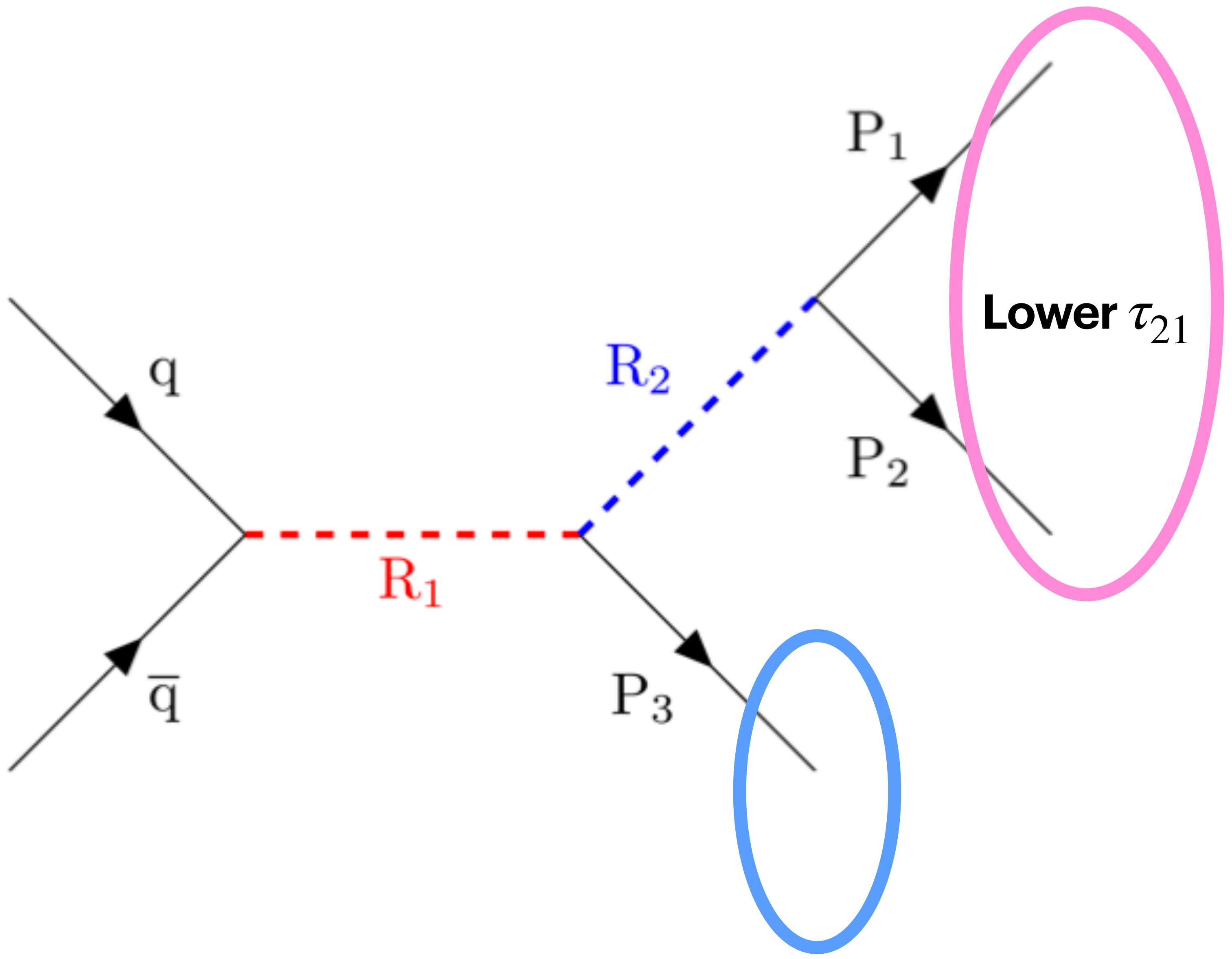


$\sigma_{loc} = 3.6$

$\sigma_{glob} = 2.5$

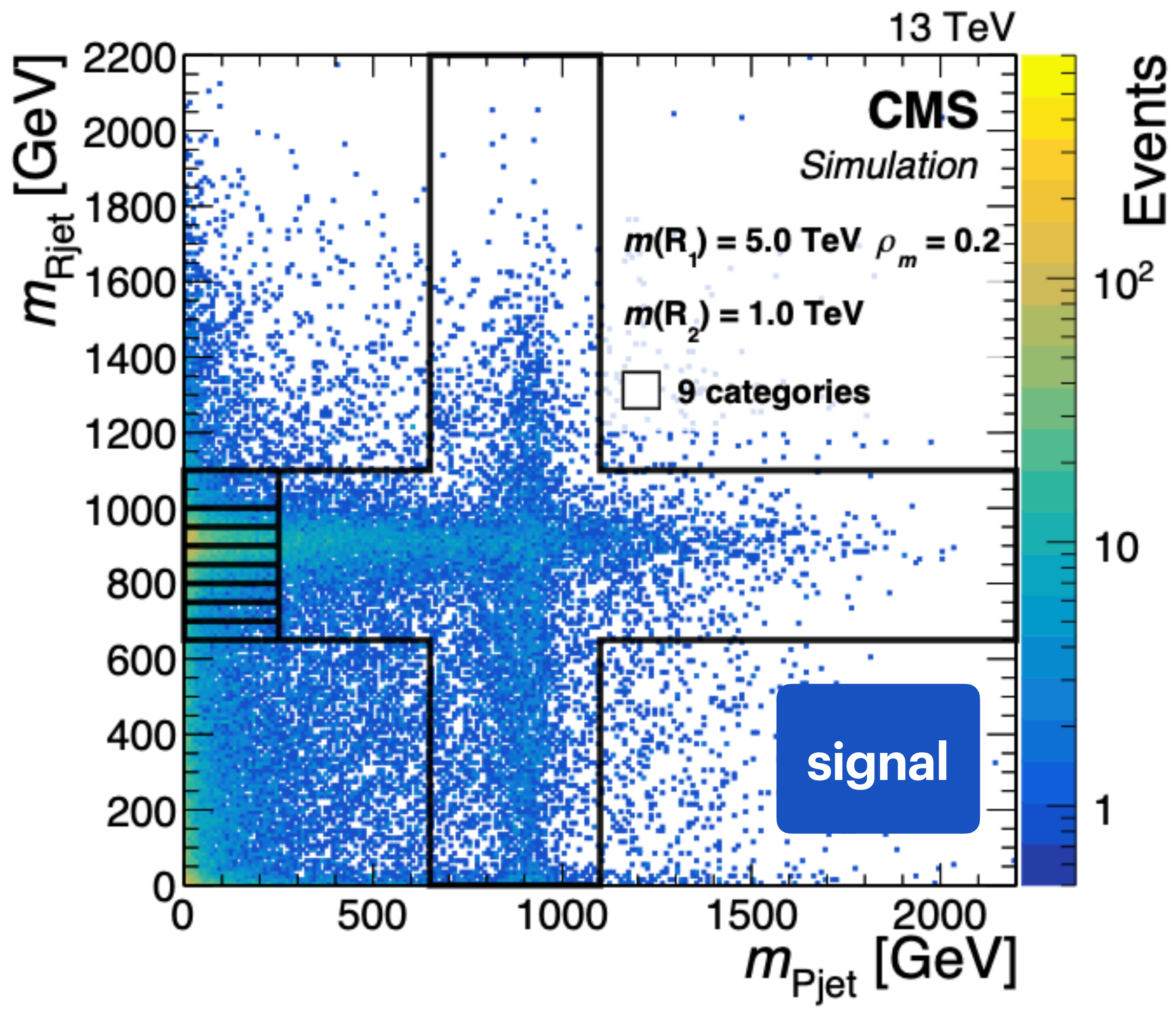
Event selection

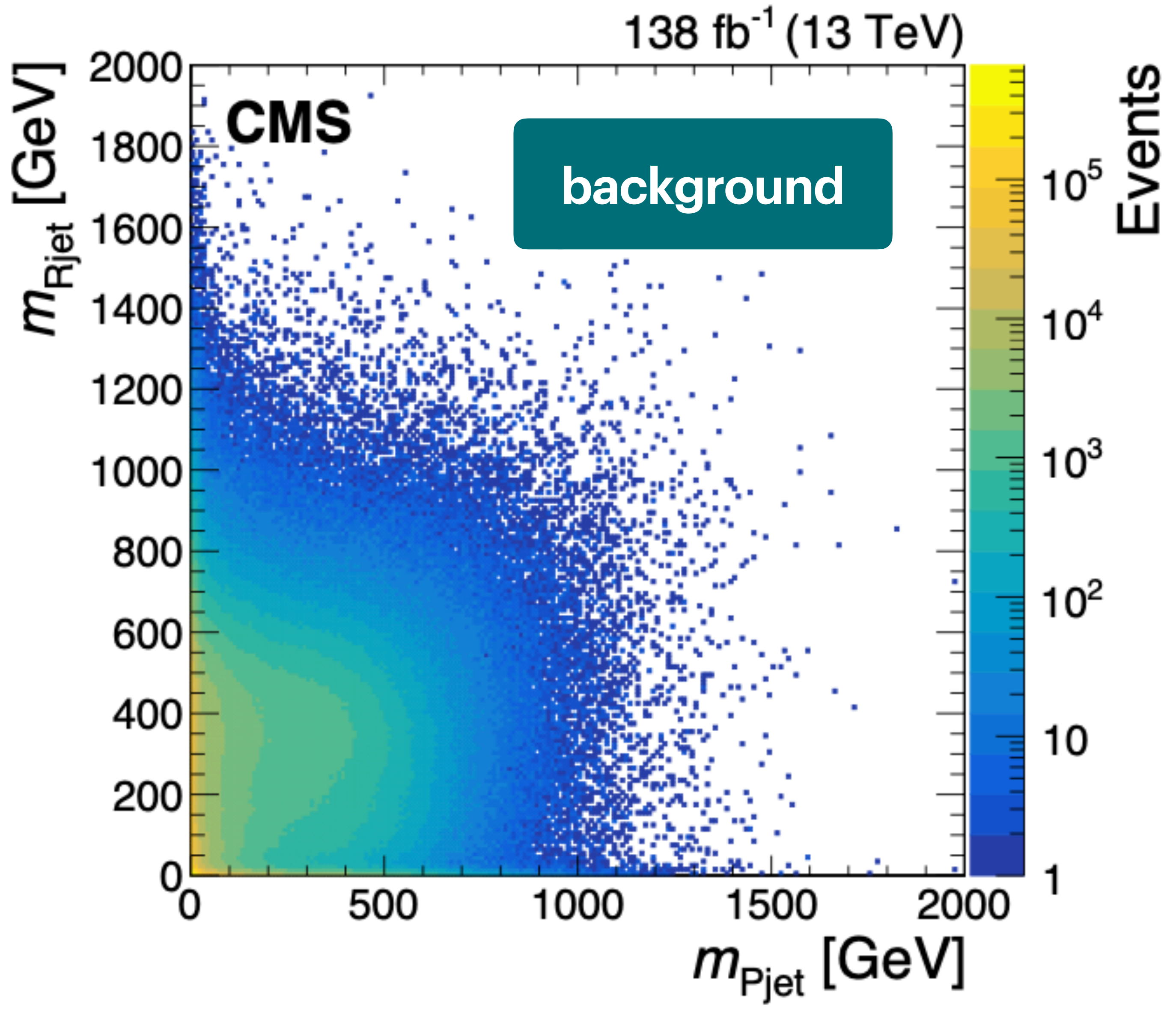
- First dedicated search for resonances decaying into 3 final state partons
- Consider new process: $q\bar{q} \rightarrow R_1 \rightarrow R_2 + P3 \rightarrow (P_1 + P_2) + P_3$
- Benchmark: $q\bar{q} \rightarrow G_{KK} \rightarrow \phi + g \rightarrow (g + g) + g$
- Select two highest pT, wide AK15 jets
- Consider only $\rho = m(R_2)/m(R_1) < 0.2$ so that R_2 is boosted enough that its decay products resolve as one jet, and $m(R_1) > 2TeV$



Analysis Strategy

- Exploit cross pattern in experimental signature to divide the data into signal enriched regions
- Binned fit to m_{jj}
  - SM background: smoothing falling spectrum
  - R_1 signal shape: double sided crystal ball w/ parameters derived from MC fit



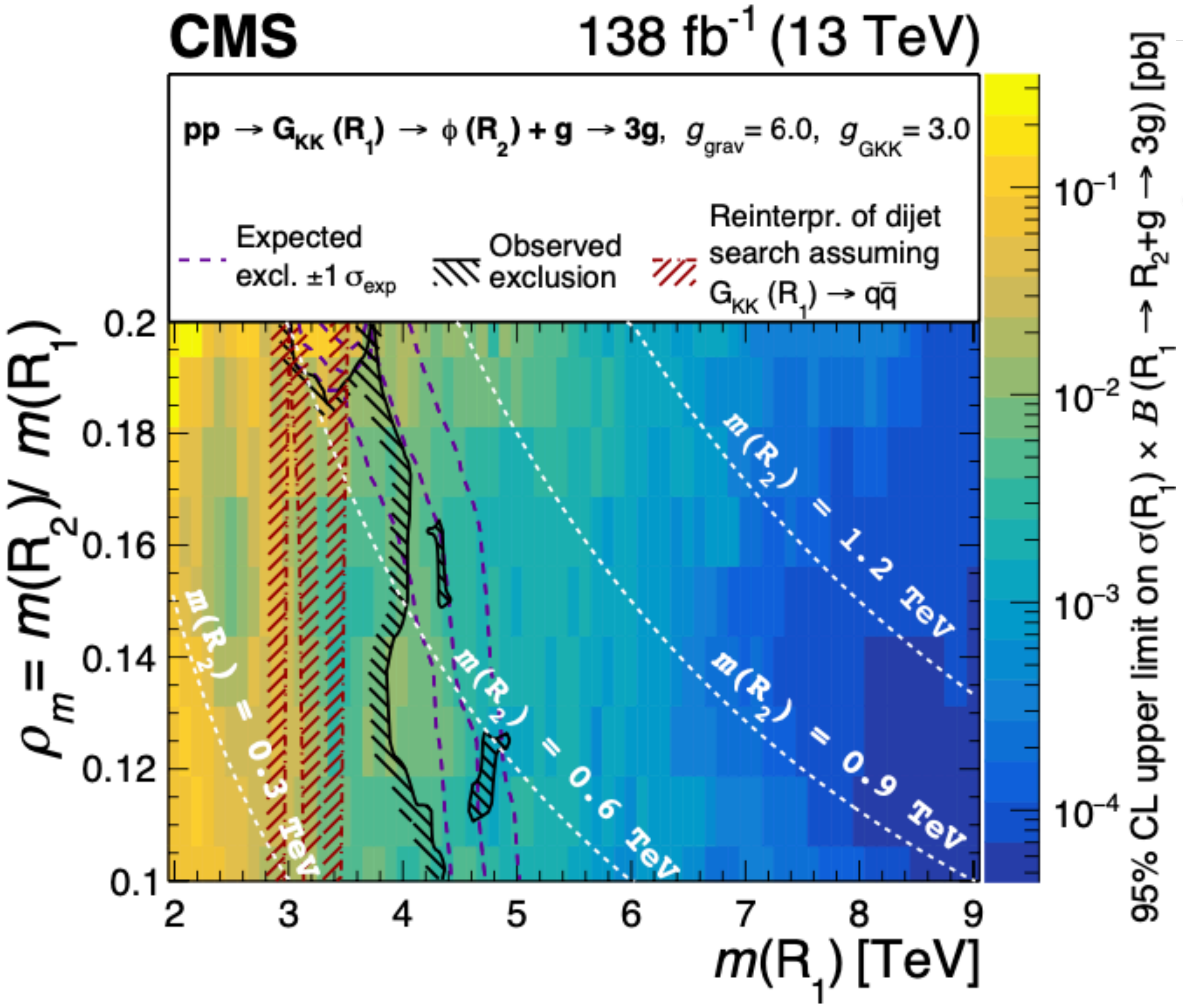


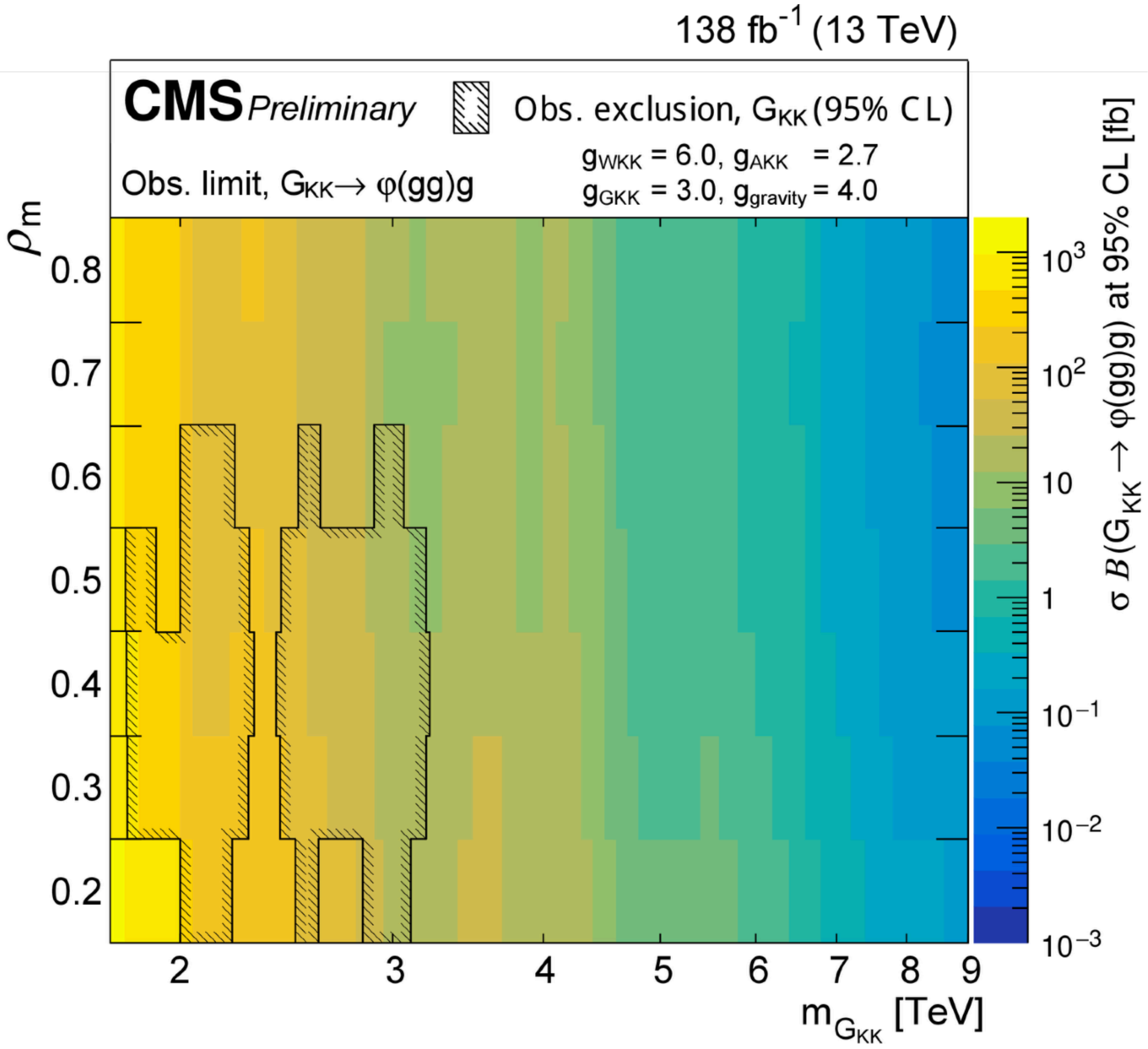
CMS-EXO-20-007

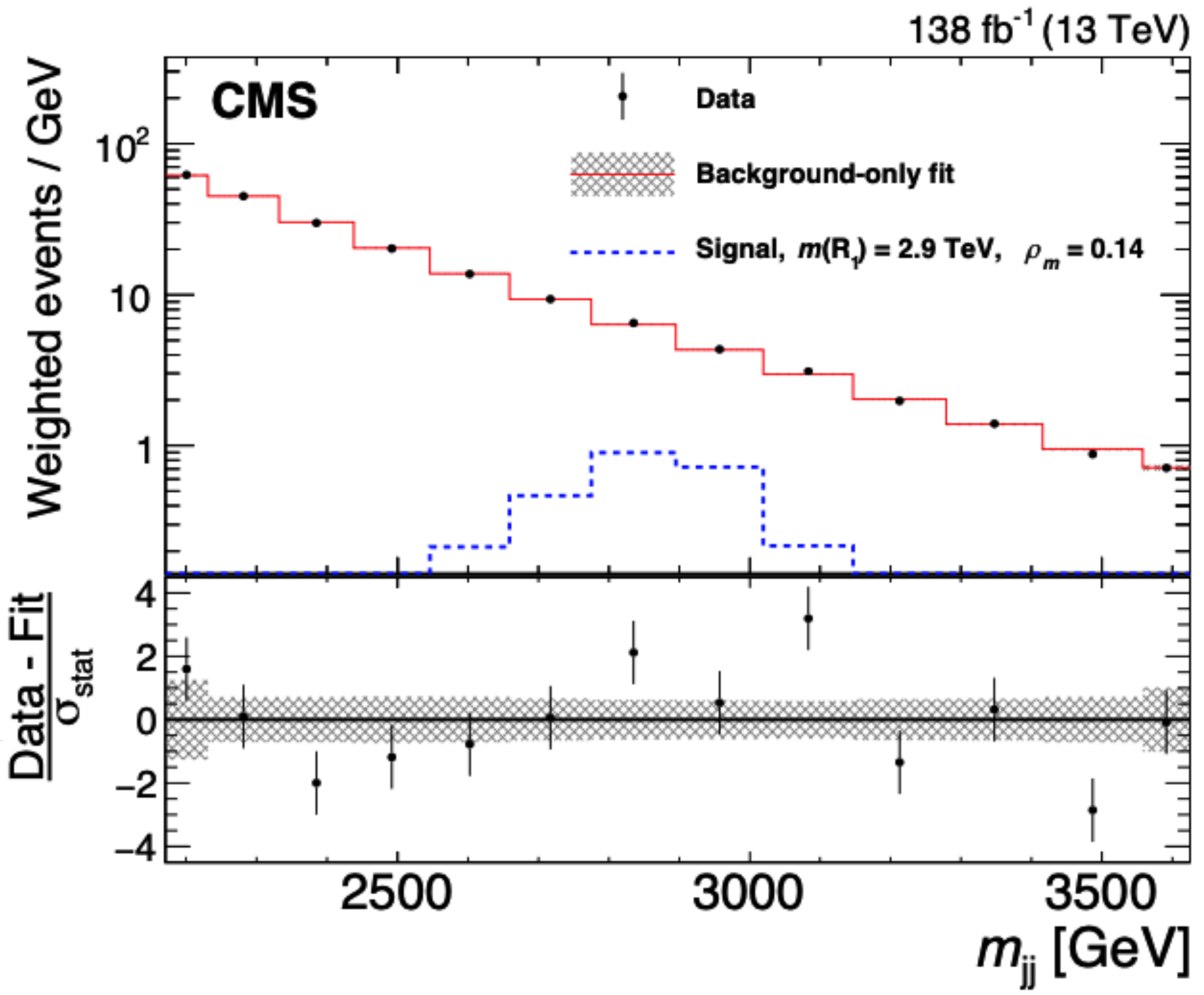
Results

- Compatible with background only predictions
- Observed limits on $m(G_{KK})$ are extended $\sim 1TeV$ compared to previous searches and resolved search

CMS-EXO-22-008







Trijet paper extends limits to higher ρ and finds 0.36σ deviation for $\rho = 0.3, m_{G_{KK}} = 4.2TeV$

Search for pair of resolved trijets, merged trijets, and merged dijets

- Search for RPV Higgsinos and gluinos and stop quarks below the previous lowest masses explored

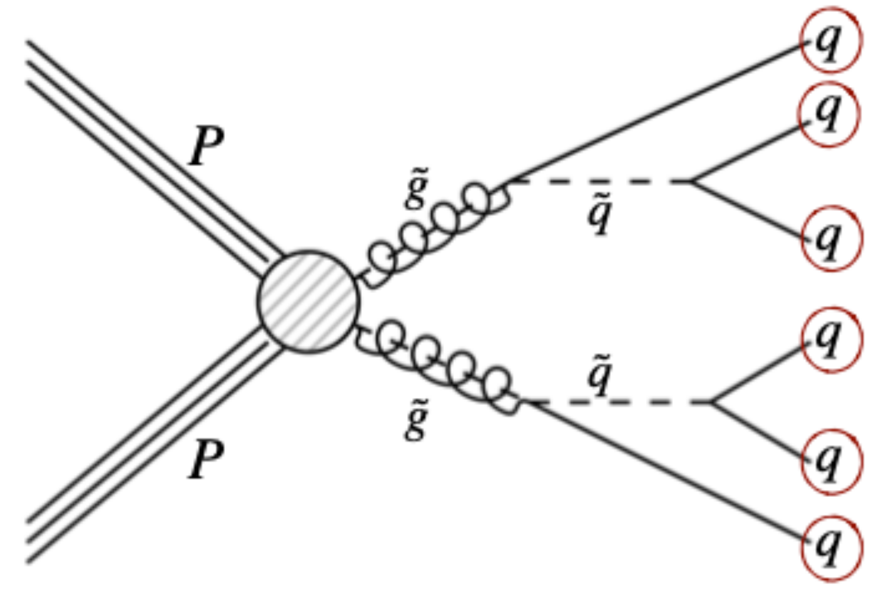
Data scouting

H_T trigger threshold was $800GeV(1.5TeV)$ in 2016(2017-2018) → implement data scouting: partially reconstructed events at the HLT are saved for further analysis w/ a lowered trigger threshold ($H_T > 410GeV$)

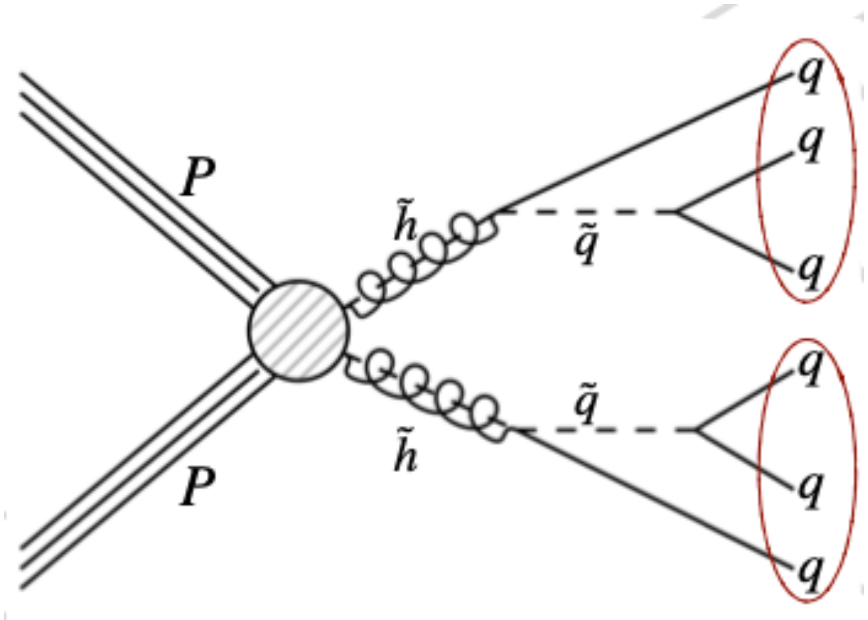
Event reconstruction

- AK4 jets for resolved trijet search and AK8 jets for the merged resonances
- Apply jet trimming to AK8 jets
- $\tau_{32}^{DDT} < 0$ and $N_2^{1,DDT} < 0$ for boosted jet searches
- For the resolved trijet resonances implement a QGL NN and Dalitz variables

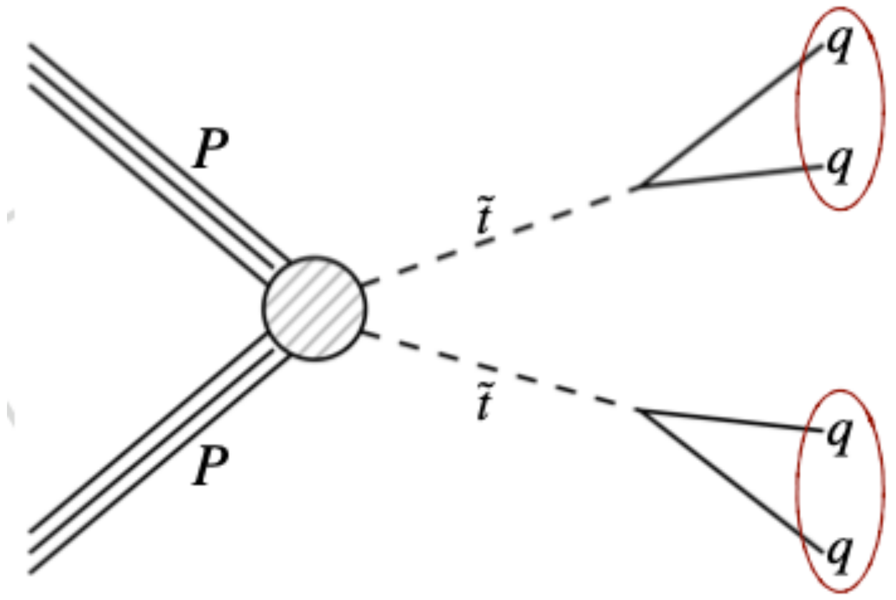
resolved trijets ($\tilde{h}, \tilde{g}$)



boosted trijets ($\tilde{h}, \tilde{g}$)

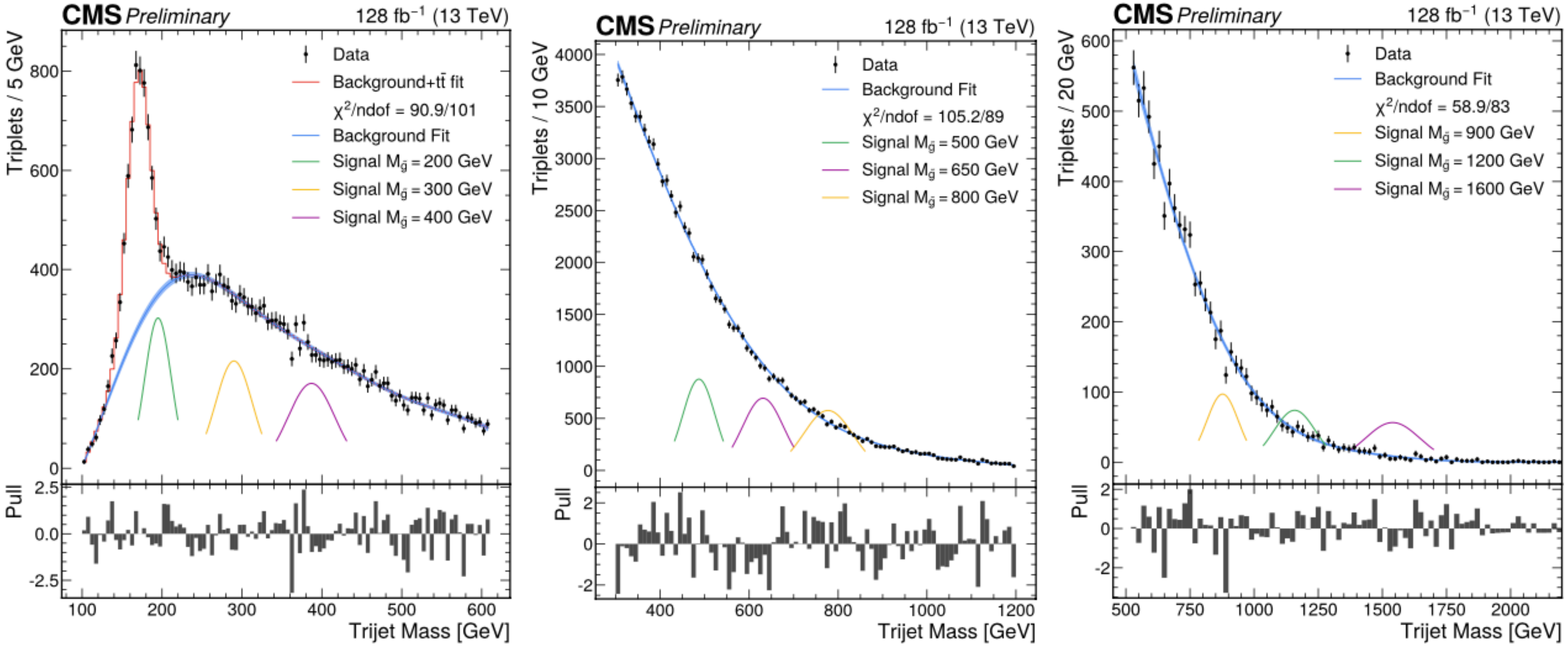


boosted dijets ($\tilde{t}$)



Signal extraction

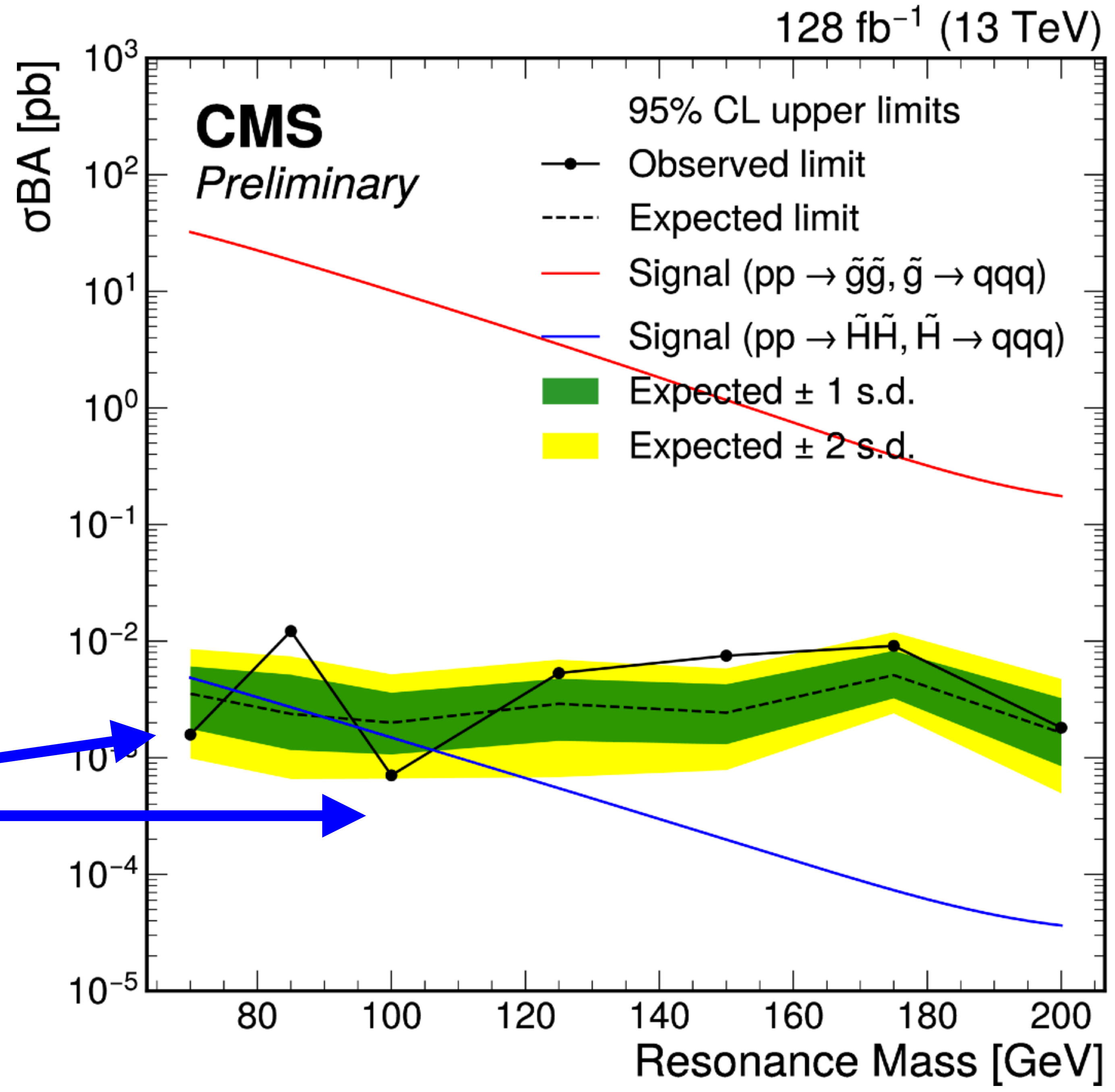
- QCD background is estimated using a Gaussian process regression JHEP 02 (2023) 230
- $t\bar{t}$ and other minor backgrounds estimated using simulation
- Signal is modeled with Gaussian distributions
- Fit the signal templates and backgrounds simultaneously to data



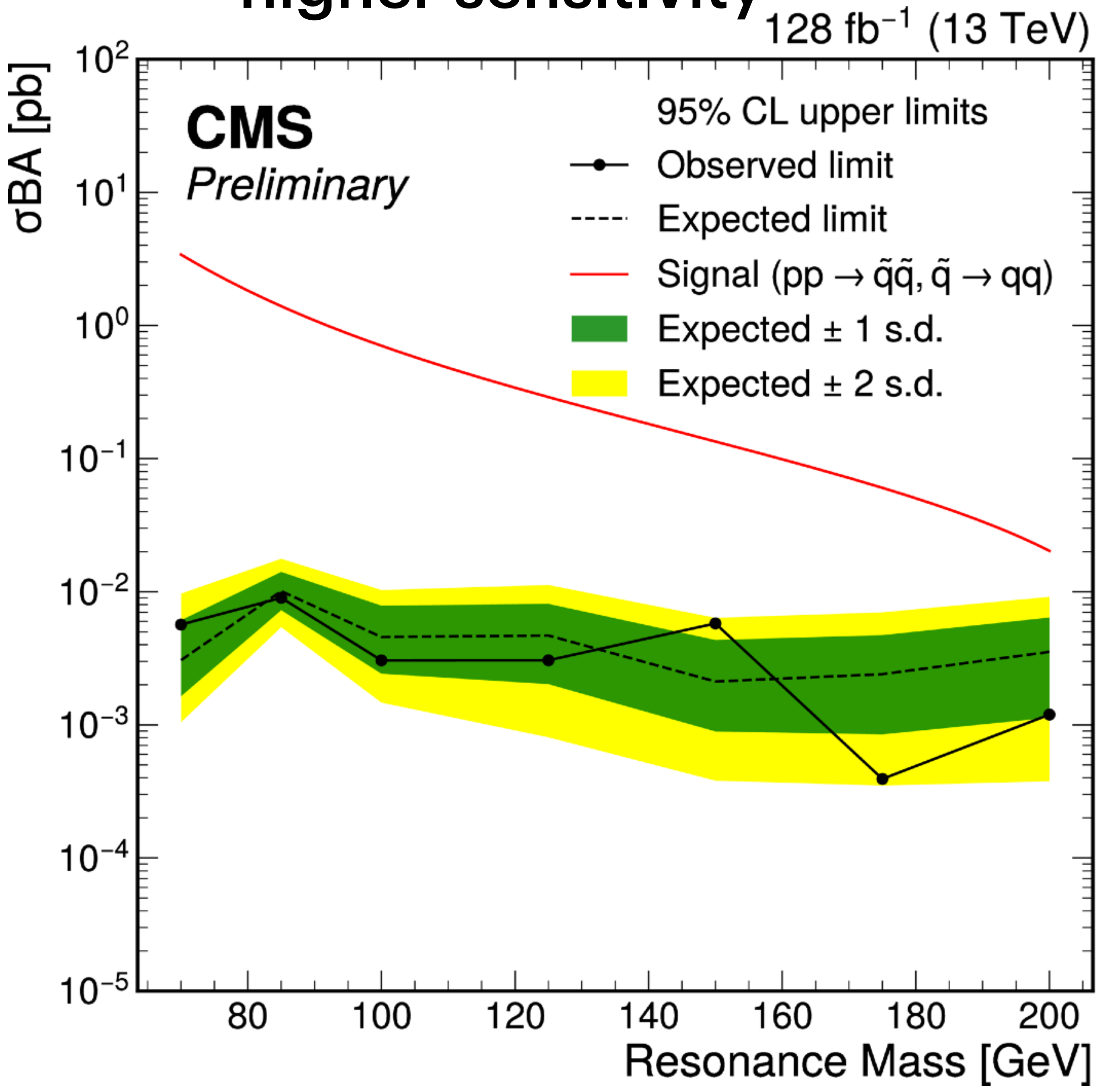
Results

Find no significant deviations from the SM

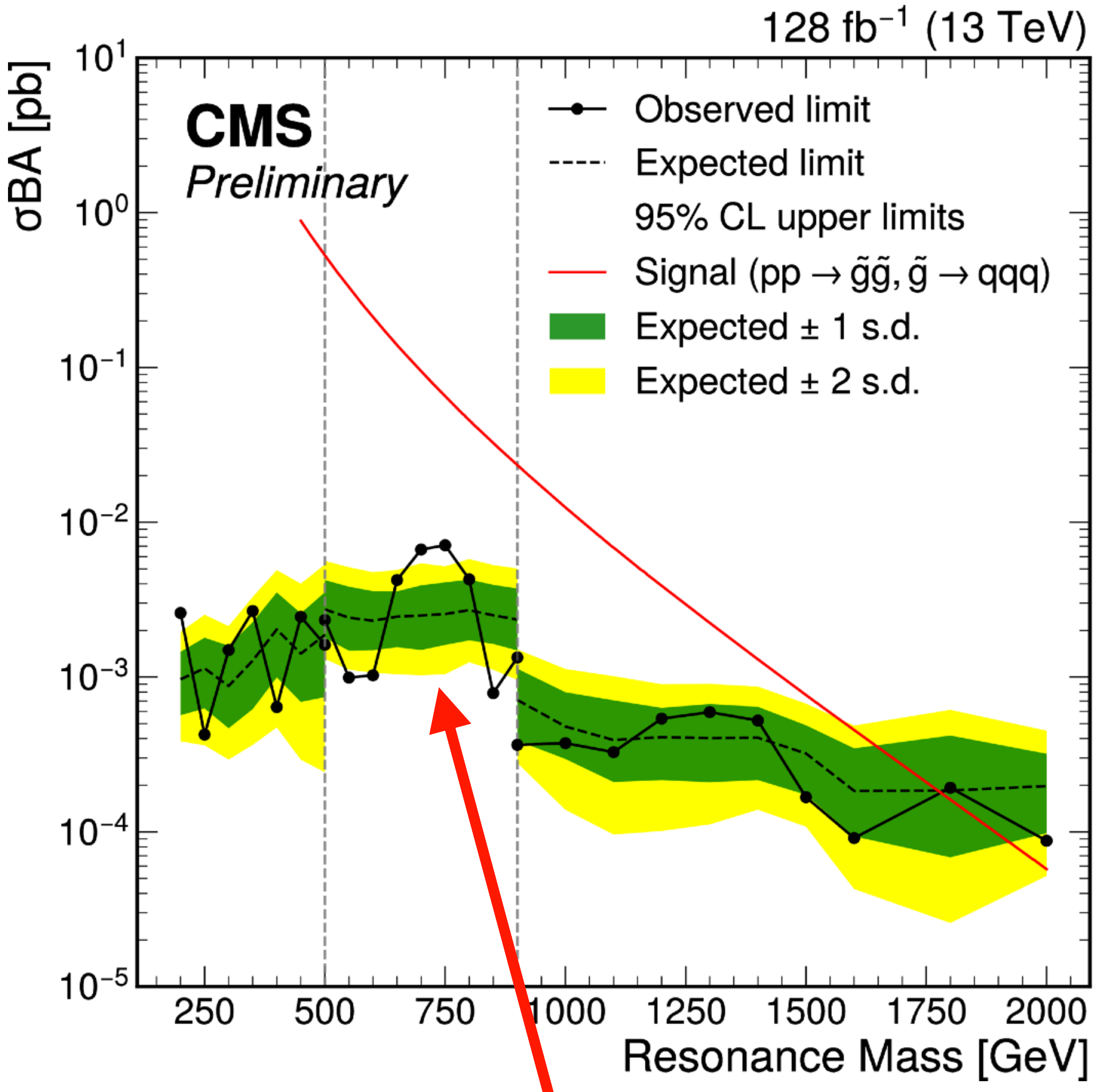
First limits on prompt RPV Higgsino



Extends $m(\tilde{t})$ limit to lower mass and higher sensitivity



Extends prior limit to lower and higher mass



2.8σ local deviation

Summary

- Latest searches for heavy resonances decaying to boosted final states at CMS with Run II data have been presented
- New unexplored channels and signatures
- No evidence for BSM physics observed
- Extended limits
- More results to come with evolving techniques and Run 3 data

Thank you!

BACKUP

Analysis	Fitting	Discr.	
$b* \to tW$ (semileptonic top)	**ABCD, ML fit to Poisson dist.**	τ_2^1 m_{SD}^W **N(b jets)**	
$W' \to tb$ (semileptonic top)	**ABCD w/ TF, ML fit to poisson**	$m_{SD}^{W'}$ m_t	
Paired dijet resonances (*)	Ratio method binned ML fit	τ_{21}	
High mass resonance —> boosted resonance + jet	Binned ML fit		
Pair produced multijet resonances using data scouting	Gaussian Prior fit		

Kinematic reconstruction

If more than one b jet is identified, the one which gives a reconstucted top mass closest to 172.5GeV is chosen

	High Mass SR	Low mass SR	TTbar CR
p_T^{miss}	> 80 GeV	< 80GeV	NA
$m_{SD,AK8}$	> 400 GeV	< 400 GeV	> 400 GeV
$\tau_{3,2}$	NA	NA	< 0.65

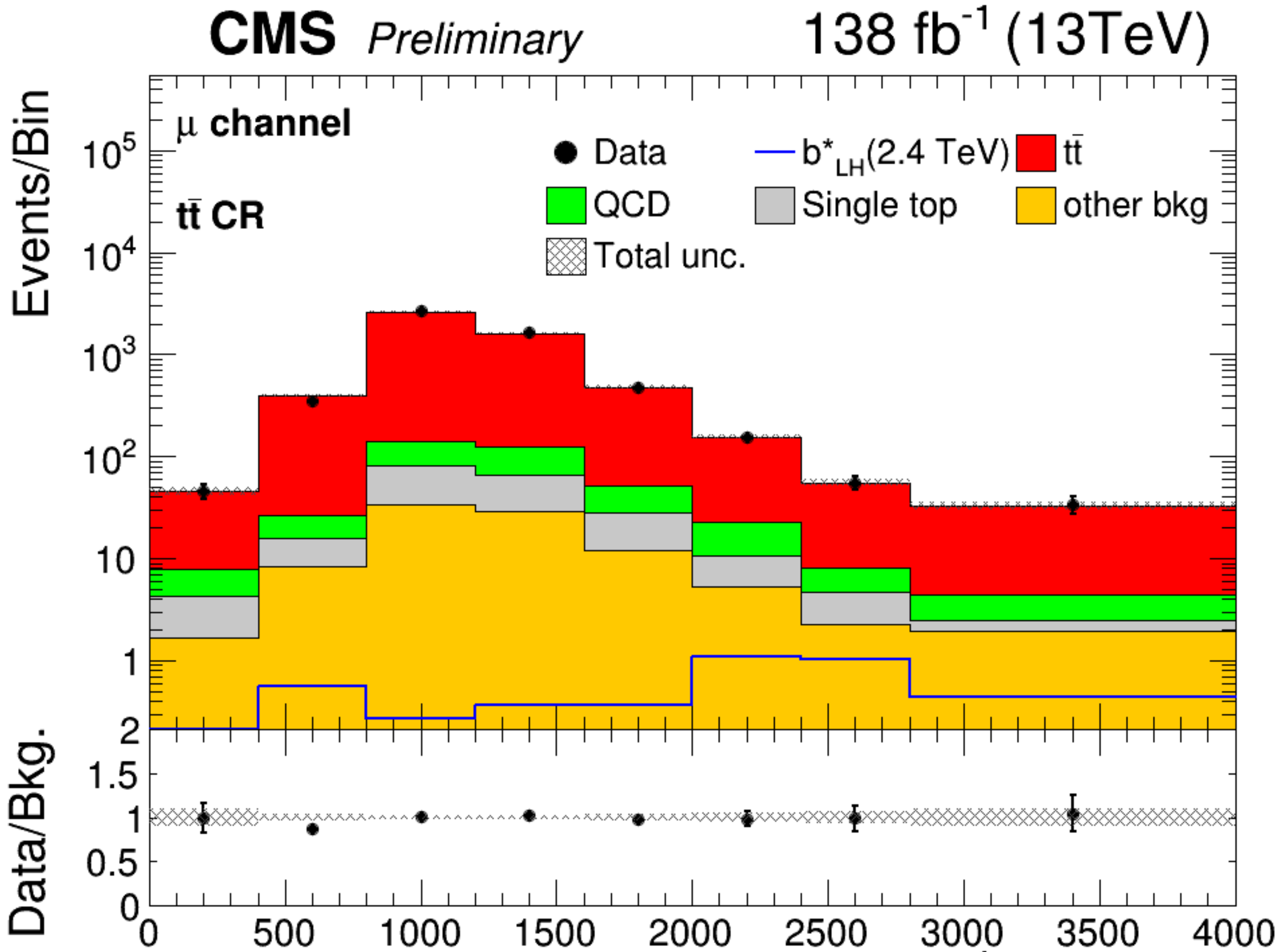
	A	B	C	D	E
2D isolation	< 60 GeV	< 60 GeV	> 60 GeV	> 60 GeV	$p_{rel}(lep,jet)$>30 GeV
τ_{21}	> 120 GeV < 220 GeV	> 220 GeV	> 120 GeV < 220 GeV	> 220 GeV	ΔR(b-jet, AK8 jet) < 0.8
N(b-jet)	0	>=1	0	>=1	$p_{rel}(lep,jet)$>30 GeV

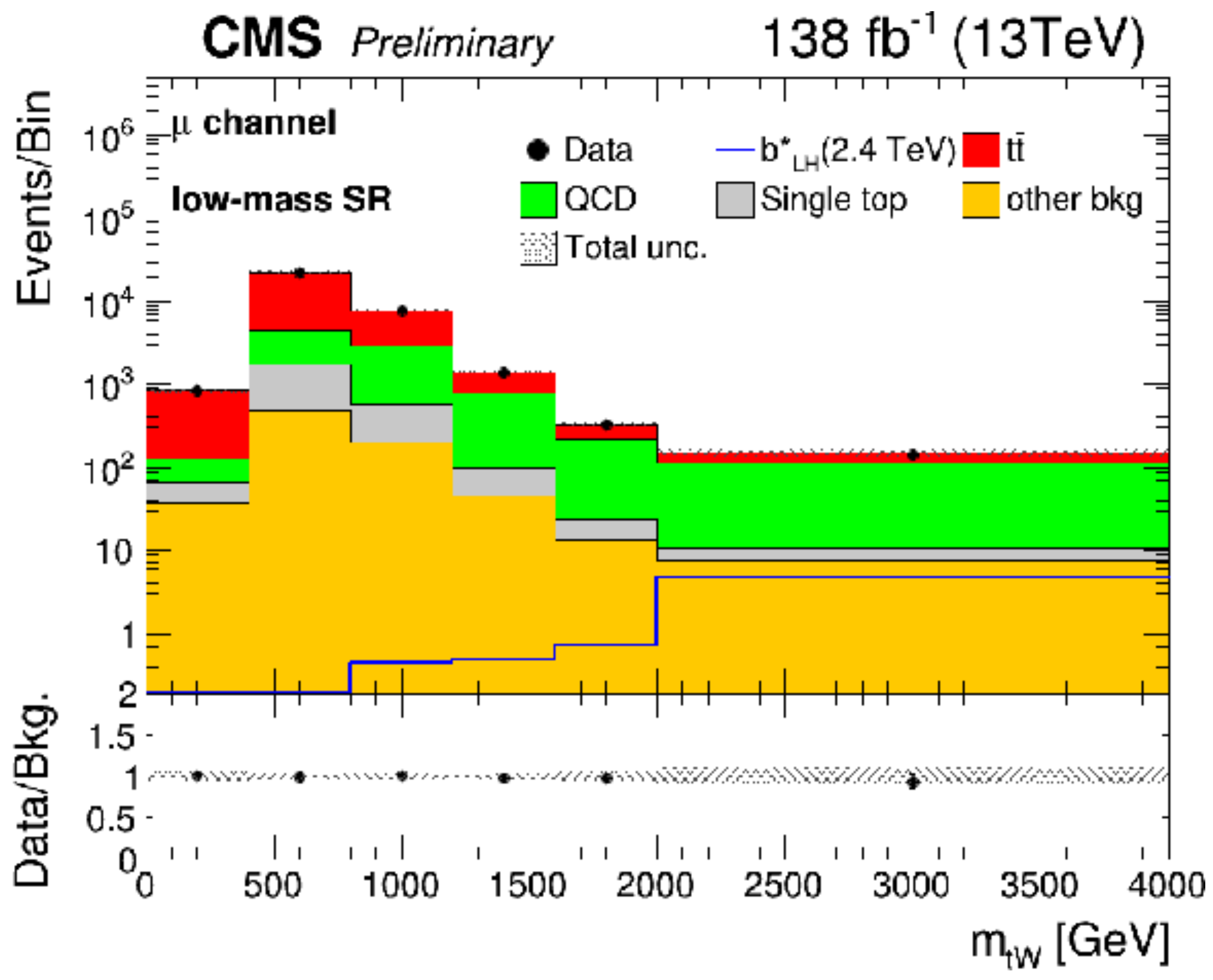
tW signal and control regions

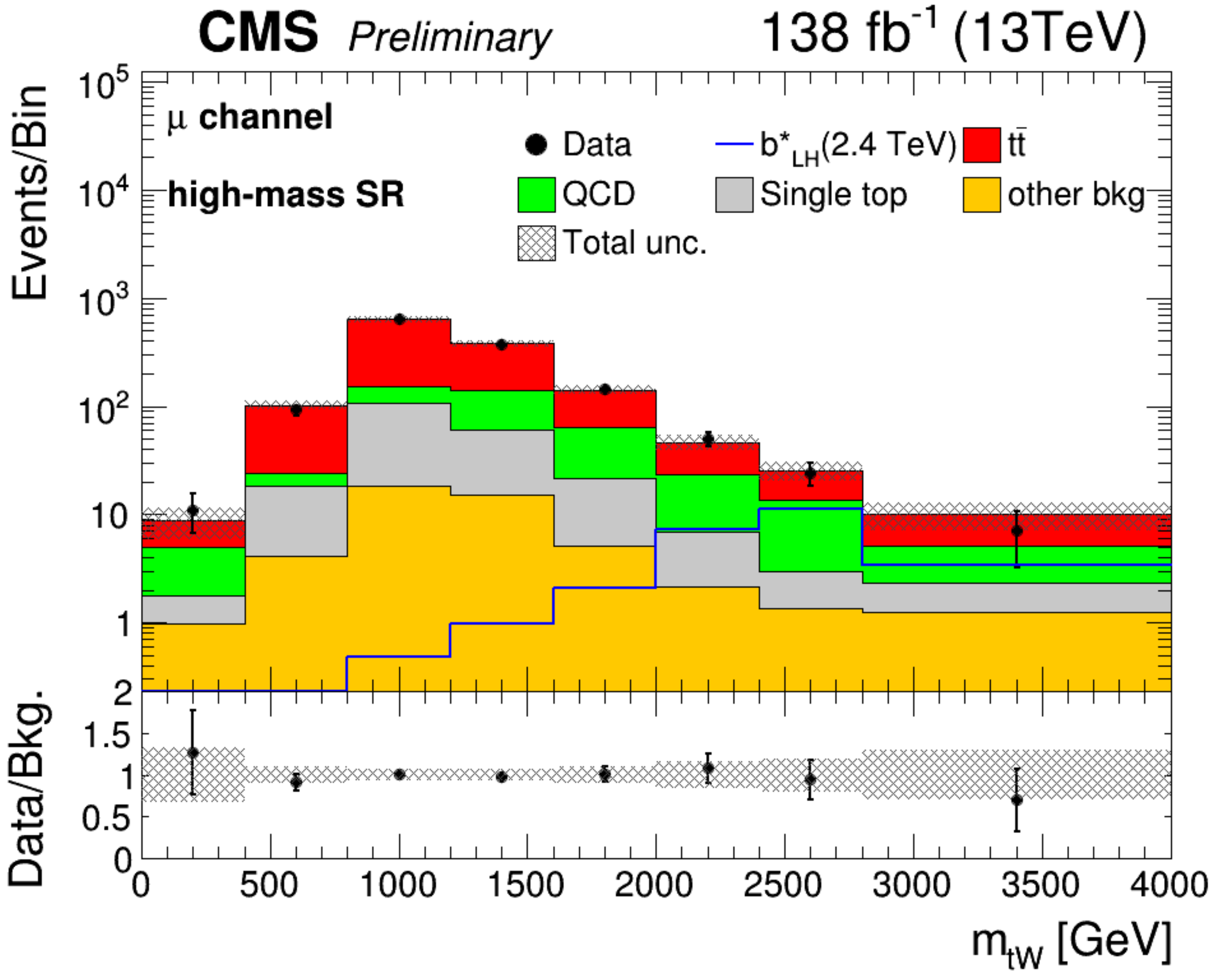
TTbar CR:

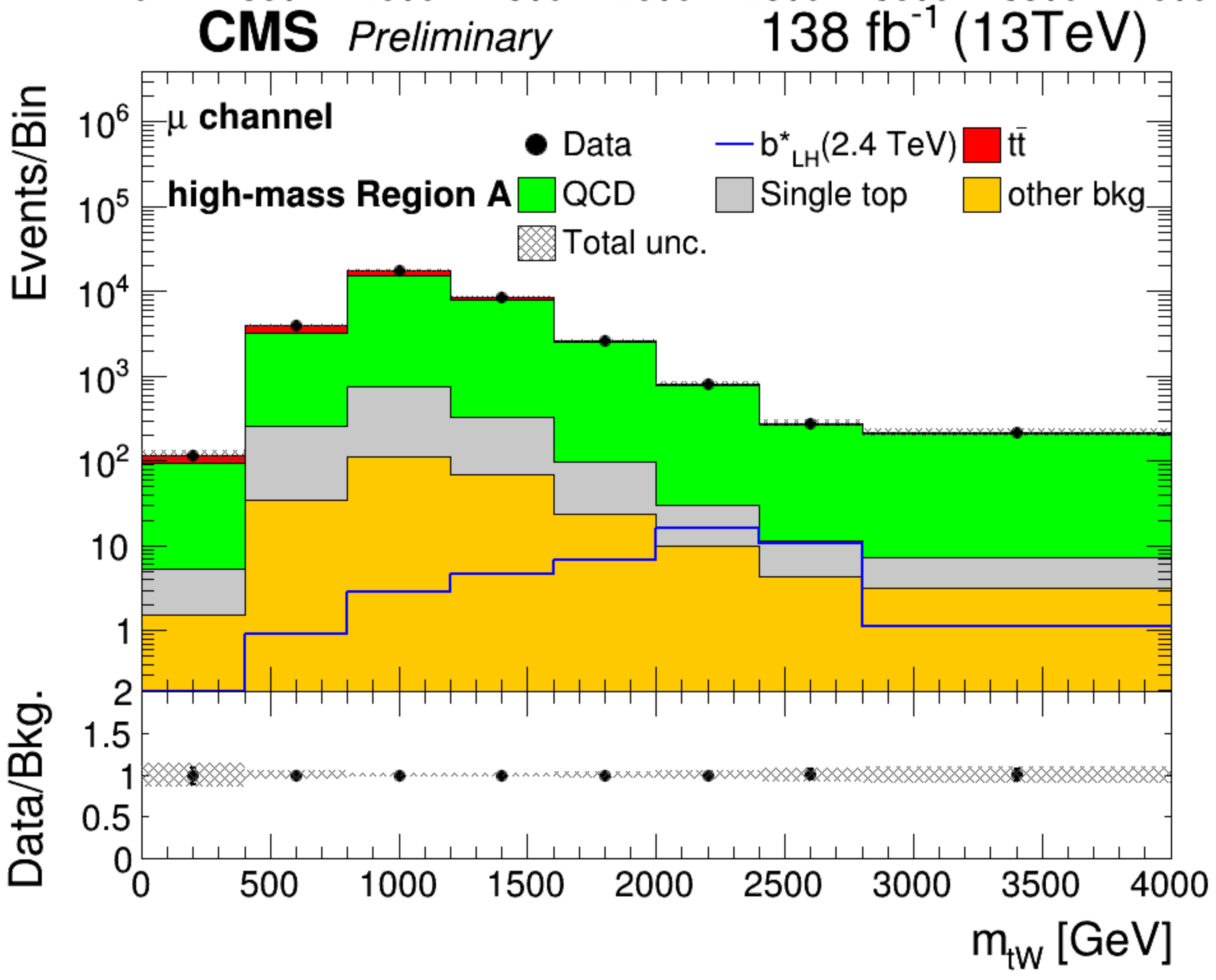
$$p_T^{rel}(lep, jet) > 30Gev$$

$$\Delta R(bjet, AK8jet) < 0.8$$

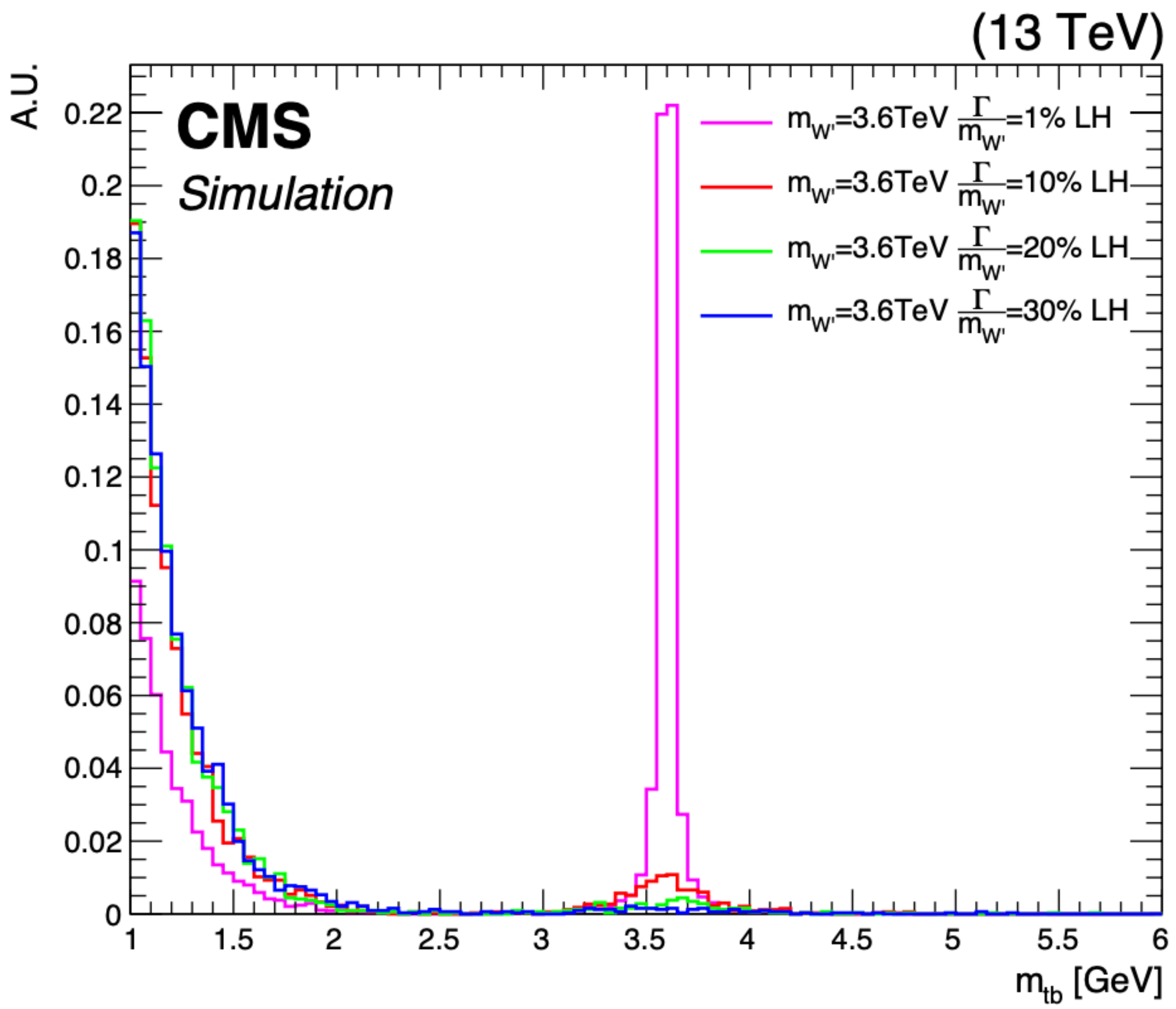


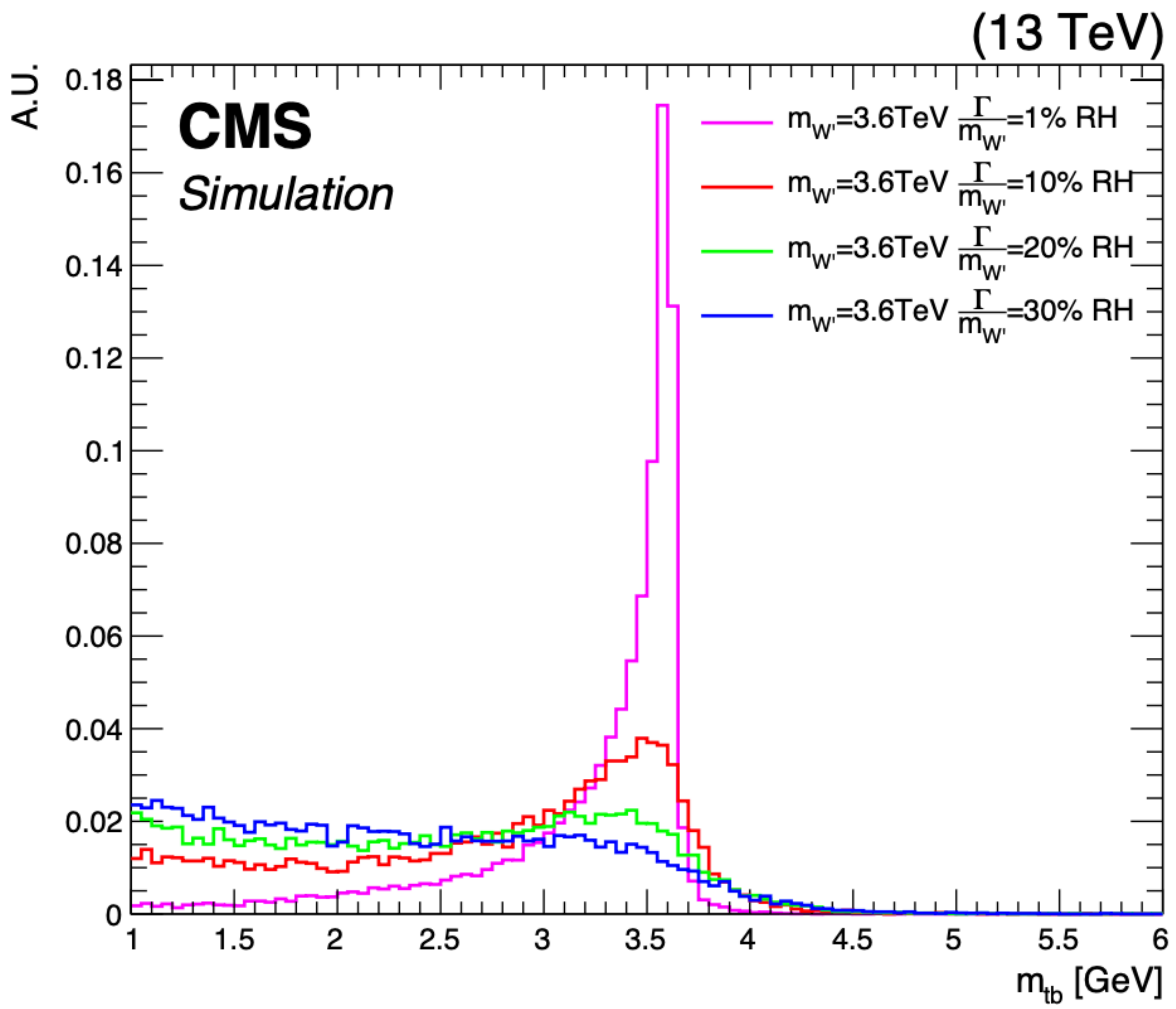




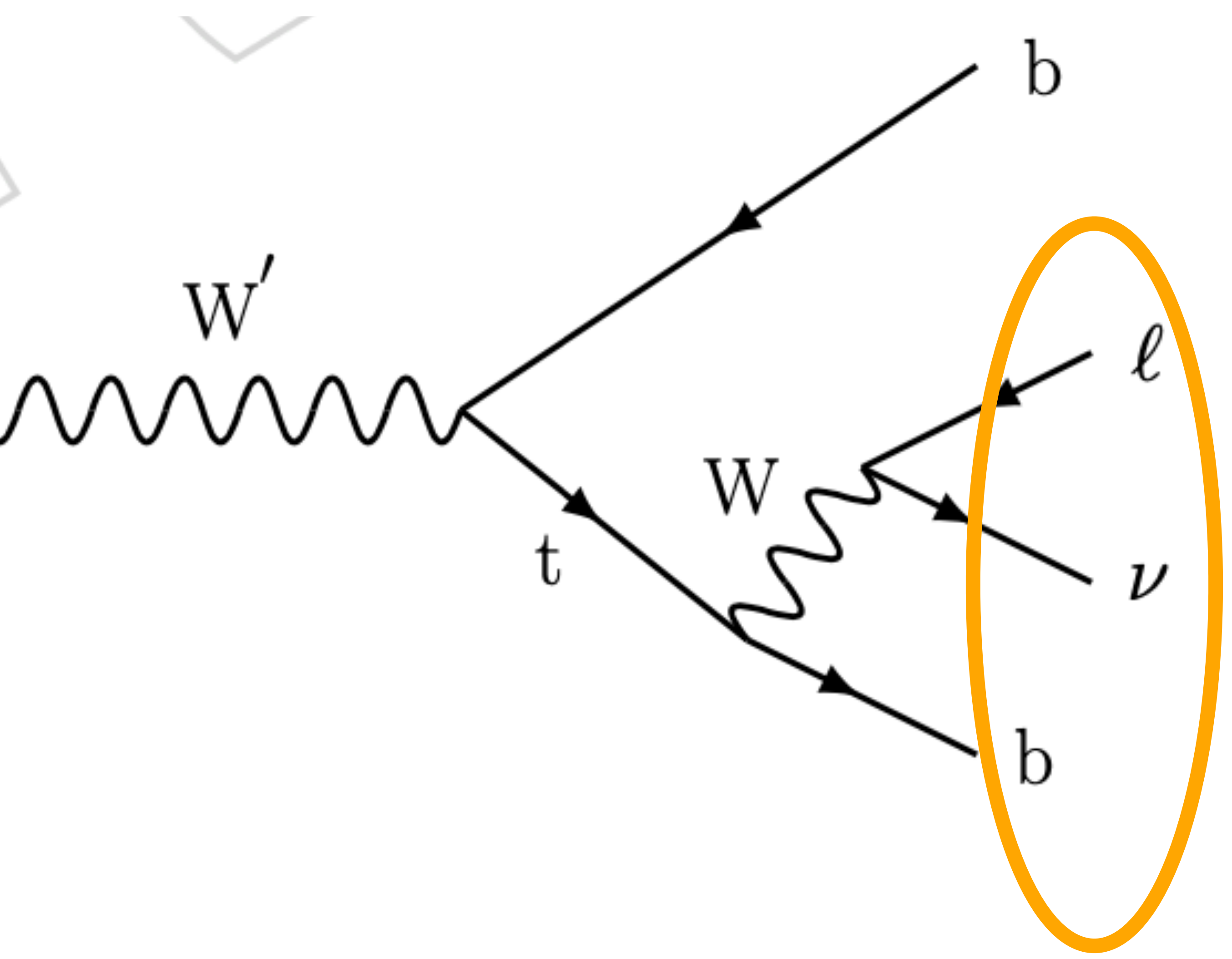


For the LH hypothesis, the region of the mW' spectrum below 2 TeV is dominated by the SM s-channel single top quark production.





Kinematic reconstruction



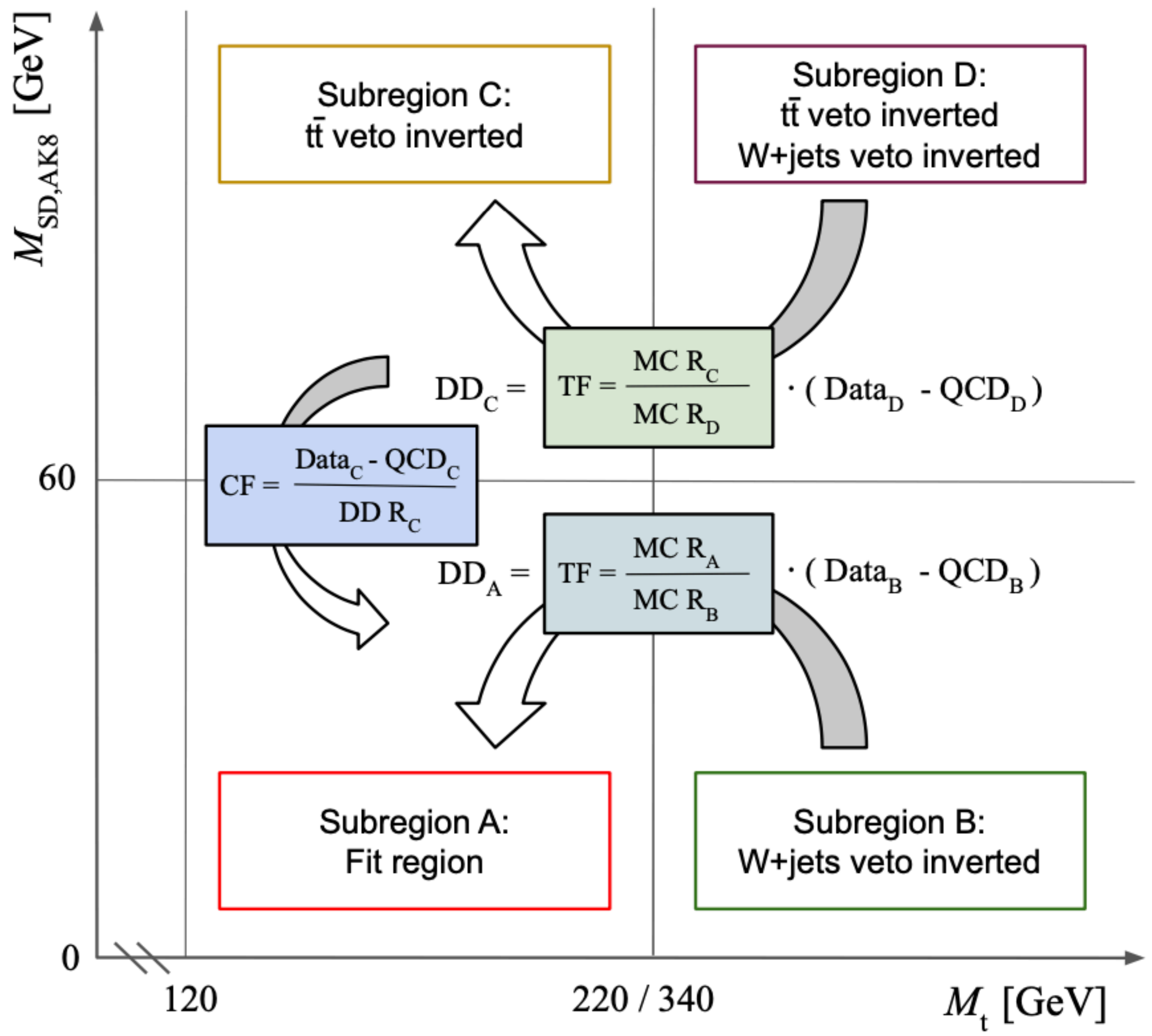
Mini isolation

$$I_{\mathrm{mini}} = \frac{S_I(R)}{p_T^{\ell}}, \text{with } R = \frac{10\,\mathrm{GeV}}{\min(\max(p_T^{\ell}, 50\,\mathrm{GeV}), 200\,\mathrm{GeV})},$$

Top quark momentum reconstruction

- $\overrightarrow{p_T}^{miss}$ is taken as $\overrightarrow{p_T}^{\nu}$
- p_z^{ν} is calculated by setting the mass of $\ell\nu$ to the mass of the W boson, and the masses of the lepton and neutrino to zero, forming a second order equation for p_z^{ν}
- if the equations results in two real or two imaginary solutions
  - two real: the one that results in a value of the W boson mass closest to the value from the world average is chosen
  - two imaginary: remove constraint $\overrightarrow{p_T}^{miss} = \overrightarrow{p_T}^{\nu}$, allowing discriminant of the second order equation to be set to zero, finding a single solution for p_z^{ν} and setting a constraint on $\overrightarrow{p_T}^{\nu}$. $\overrightarrow{p_T}^{\nu}$ is then found by minimizing the vectorial distance between $\overrightarrow{p_T}^{\nu}$ and $\overrightarrow{p_T}^{miss}$

Background estimation - sideband extrapolation



- Fit transfer function $TF = ae^{bM_{\ell\nu jj}} + cM_{\ell\nu jj} + d,$ for region B.
- Estimate number of events, N from this fit, subtract the QCD multijet contribution estimated from simulation, and multiply by the TF to get the number of nonQCD events in region A.
- $N^{\mathrm{non-QCD,R_A}} = TF(N^{\mathrm{Data\,R_B}} - N^{\mathrm{QCD,MC\,R_B}}),$
- Repeat this process for regions C and D.
- $N^{\mathrm{non-QCD,R_C}} = TF_{\mathrm{CD}}(N^{\mathrm{Data\,R_D}} - N^{\mathrm{QCD,MC\,R_D}}).$
- A bin by bin correction factor is found from regions C and D and applied to region A.
- $$CF = \frac{N^{\mathrm{non-QCD,Data\,R_C}}}{N^{\mathrm{non-QCD,R_C}}},$$
- $N^{\mathrm{non-QCD,corr\,R_A}} = CFN^{\mathrm{non-QCD,R_A}}.$

Background estimation and signal extraction

# b jets	j_t is b-tagged	$j_{W'}$ is b-tagged	Region
0	no	no	R0 (CR)
1	yes	no	RT (SR)
1	no	yes	RW (SR)
≥2	yes	yes	R2B (SR)

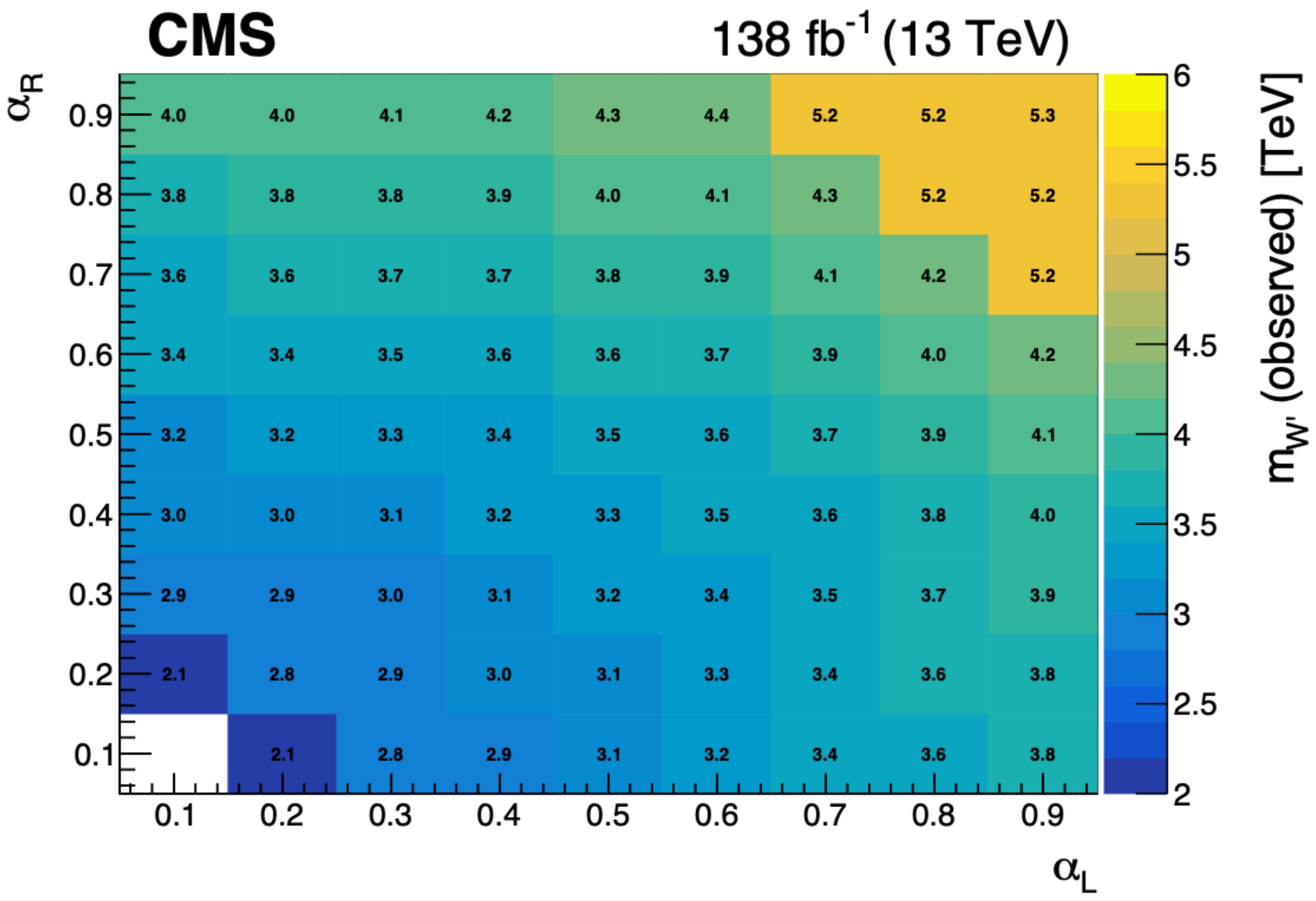
	A	B	C	D (R0, RW’, RT)	D (R2B)
$M_{SD,AK8}$	< 60 GeV	< 60 GeV	> 60 GeV	> 60 GeV	< 30 GeV
M_t	> 120 GeV < 220 GeV	> 220 GeV	> 120 GeV < 220 GeV	> 220 GeV	> 340 GeV

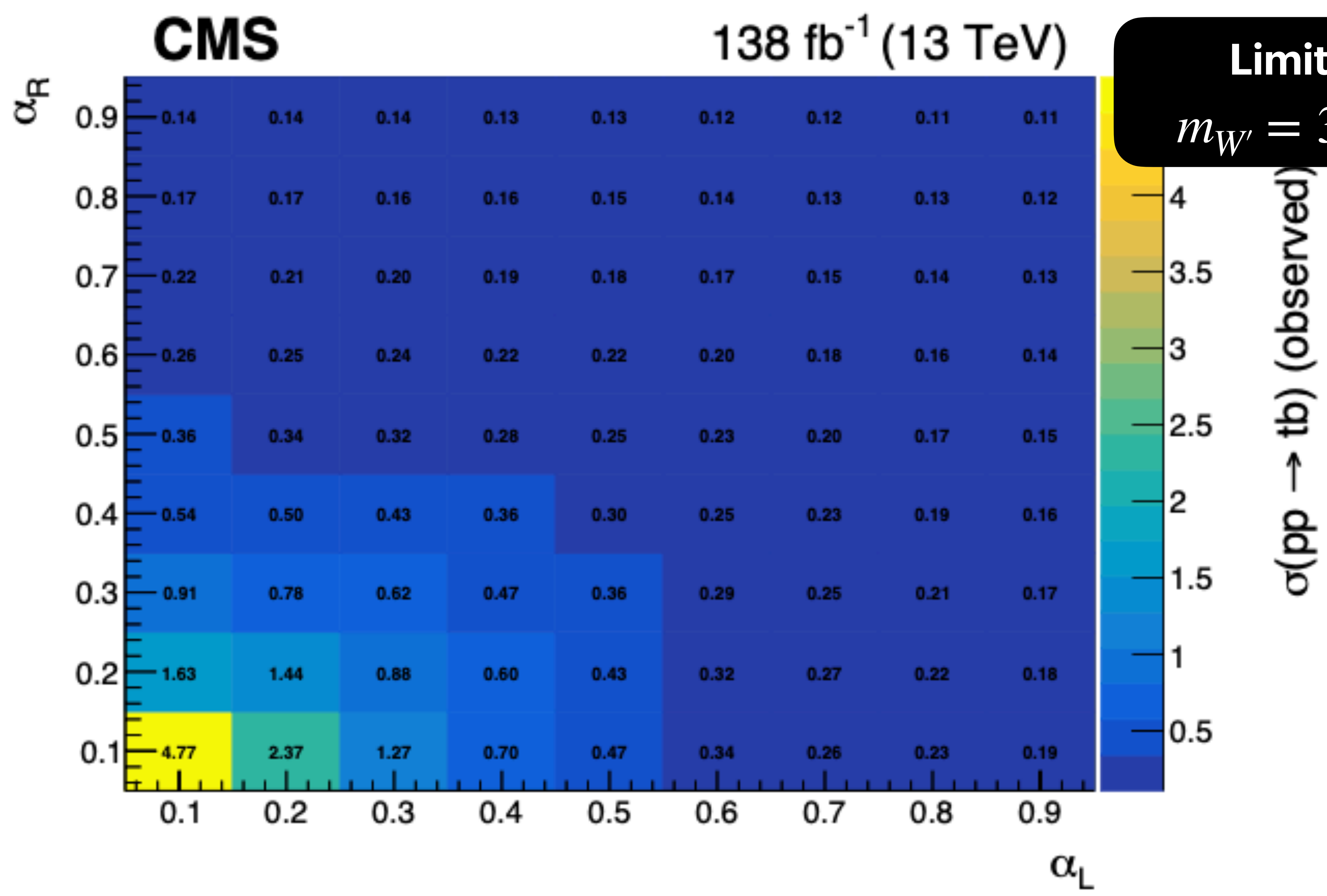
- Subregions R2B$_A$, RT$_A$, and RW'$_A$ are used for signal extraction

$\alpha_{R/L}$ coupling

- $$\mathcal{L}^{\mathrm{eff}} = \frac{V_{f_i f_j}}{2\sqrt{2}} g_{W'} \bar{f}_i \gamma_\mu [\alpha_{\mathrm{R}}^{f_i f_j}(1+\gamma^5) + \alpha_{\mathrm{L}}^{f_i f_j}(1-\gamma^5)] W'^{\mu} f_j + \mathrm{h.c.},$$
- Models with values of α_R and α_L (coupling to chiral fermions) ranging in the interval (0.1, 0.9) in steps of 0.1 were also tested in the hypothesis of $\Gamma / m_{W'} = 1\ \%$
- For this narrow width can write cross section as

$$\sigma = (1-\alpha_{\mathrm{L}}^2)\sigma_{\mathrm{SM}} + \frac{1}{\alpha_{\mathrm{L}}^2 + \alpha_{\mathrm{R}}^2}\left[\alpha_{\mathrm{L}}^2(\alpha_{\mathrm{L}}^2 - \alpha_{\mathrm{R}}^2)\sigma_{\mathrm{L}} + \alpha_{\mathrm{R}}^2(\alpha_{\mathrm{R}}^2 - \alpha_{\mathrm{L}}^2)\sigma_{\mathrm{R}} + 4\alpha_{\mathrm{L}}^2\alpha_{\mathrm{R}}^2\sigma_{\mathrm{LR}} - 2\alpha_{\mathrm{L}}^2\alpha_{\mathrm{R}}^2\sigma_{\mathrm{SM}}\right]$$

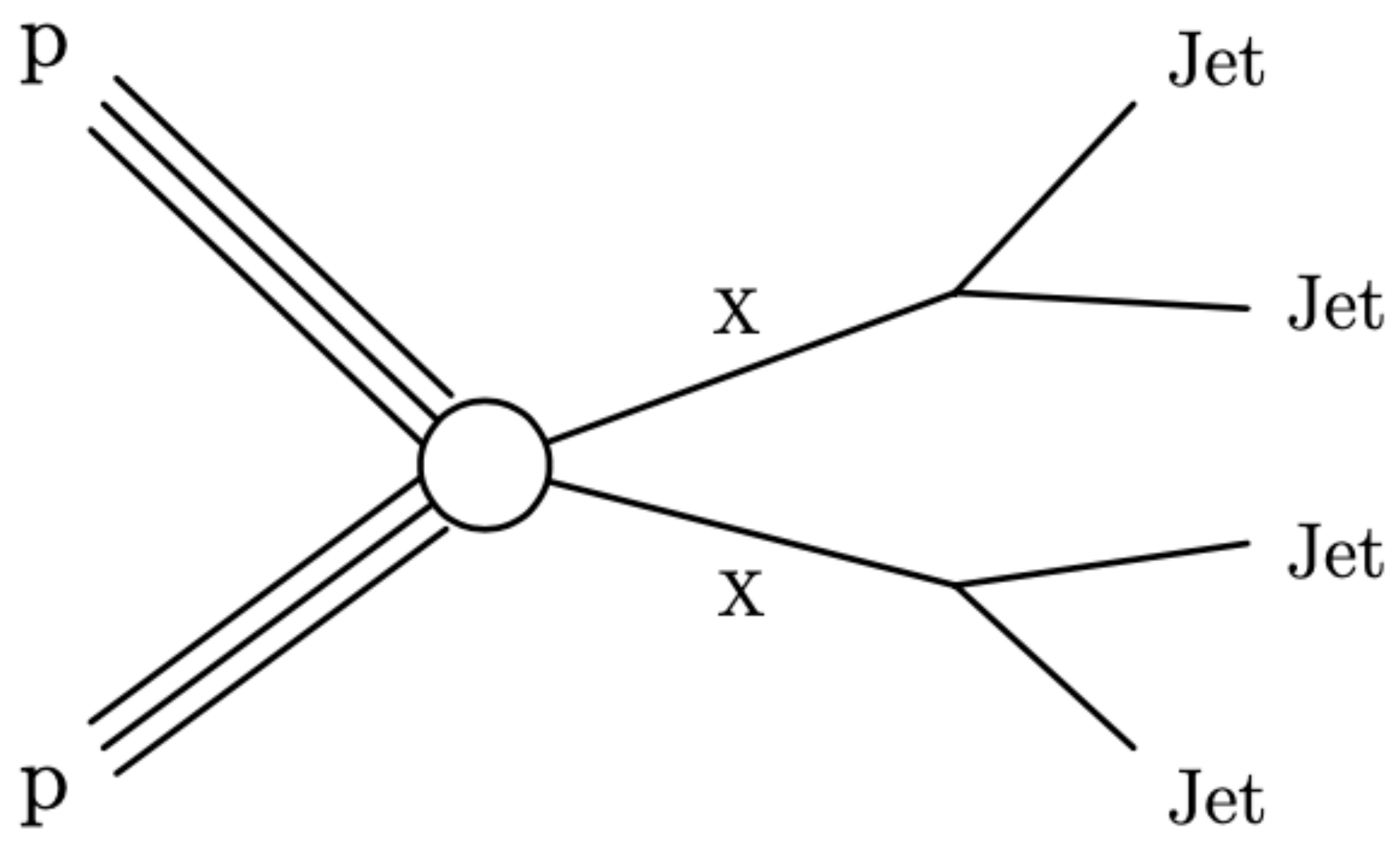




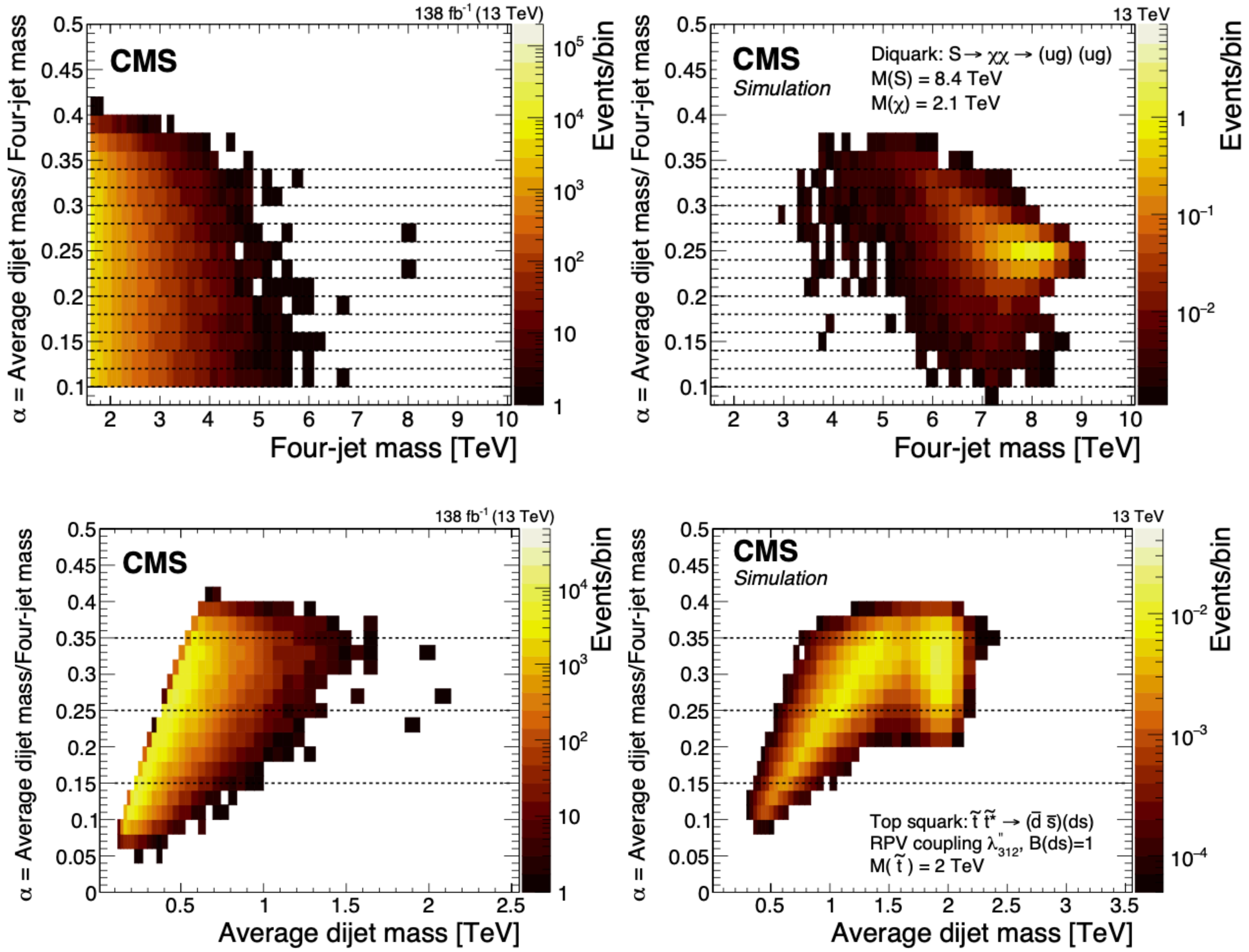
Limits for
$m_{W'} = 3.6 TeV$

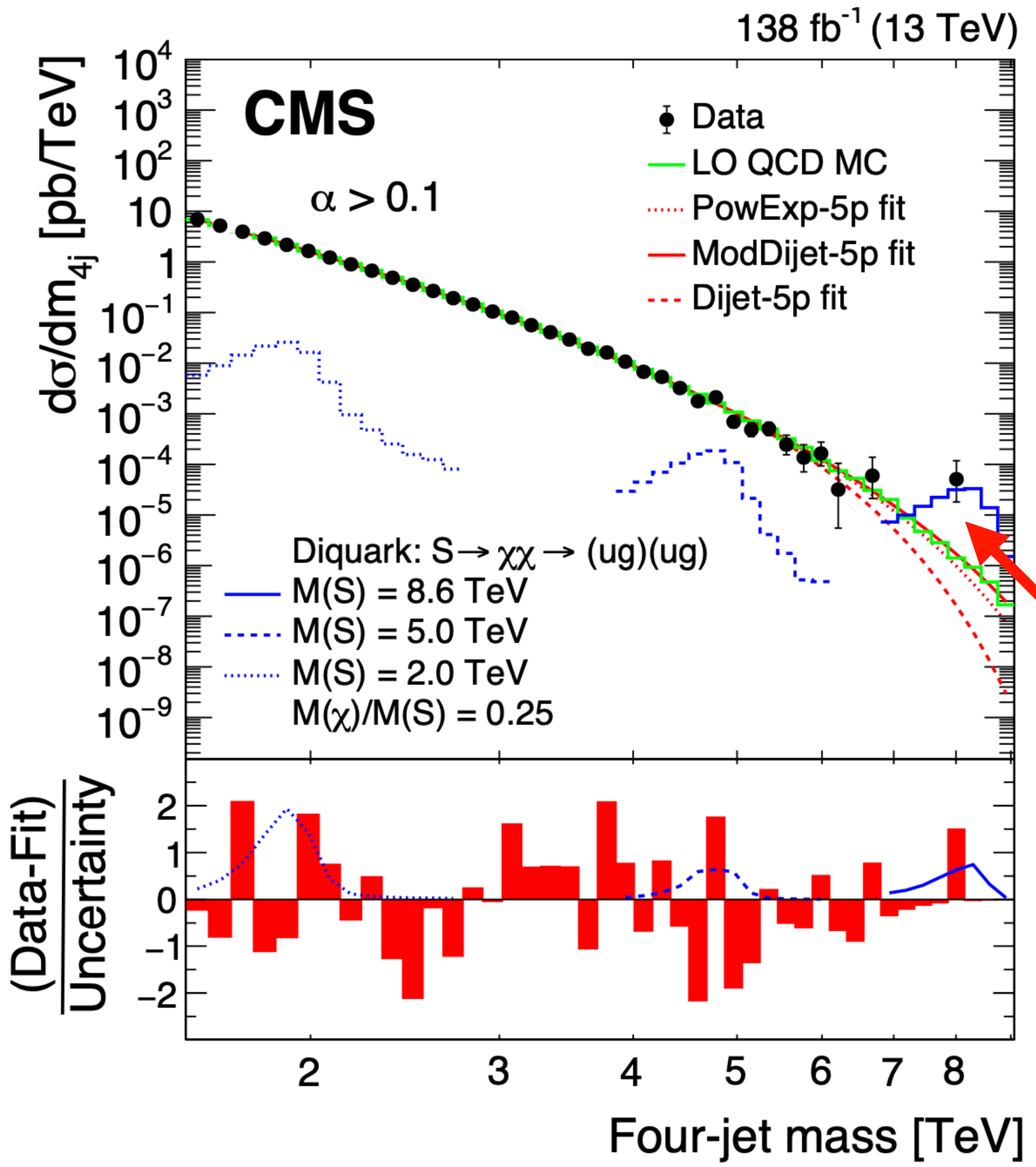
Search for resonant and nonresonant production of pairs of dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV

- After jet pairing choose events w/ $\Delta R_{12} < 2.0$, $\Delta\eta_{1,2} < 1.1$, mass asymmetry < 0.1



Nonresonant dijet pair production





Background estimation and signal extraction

Background is derived exclusively from data

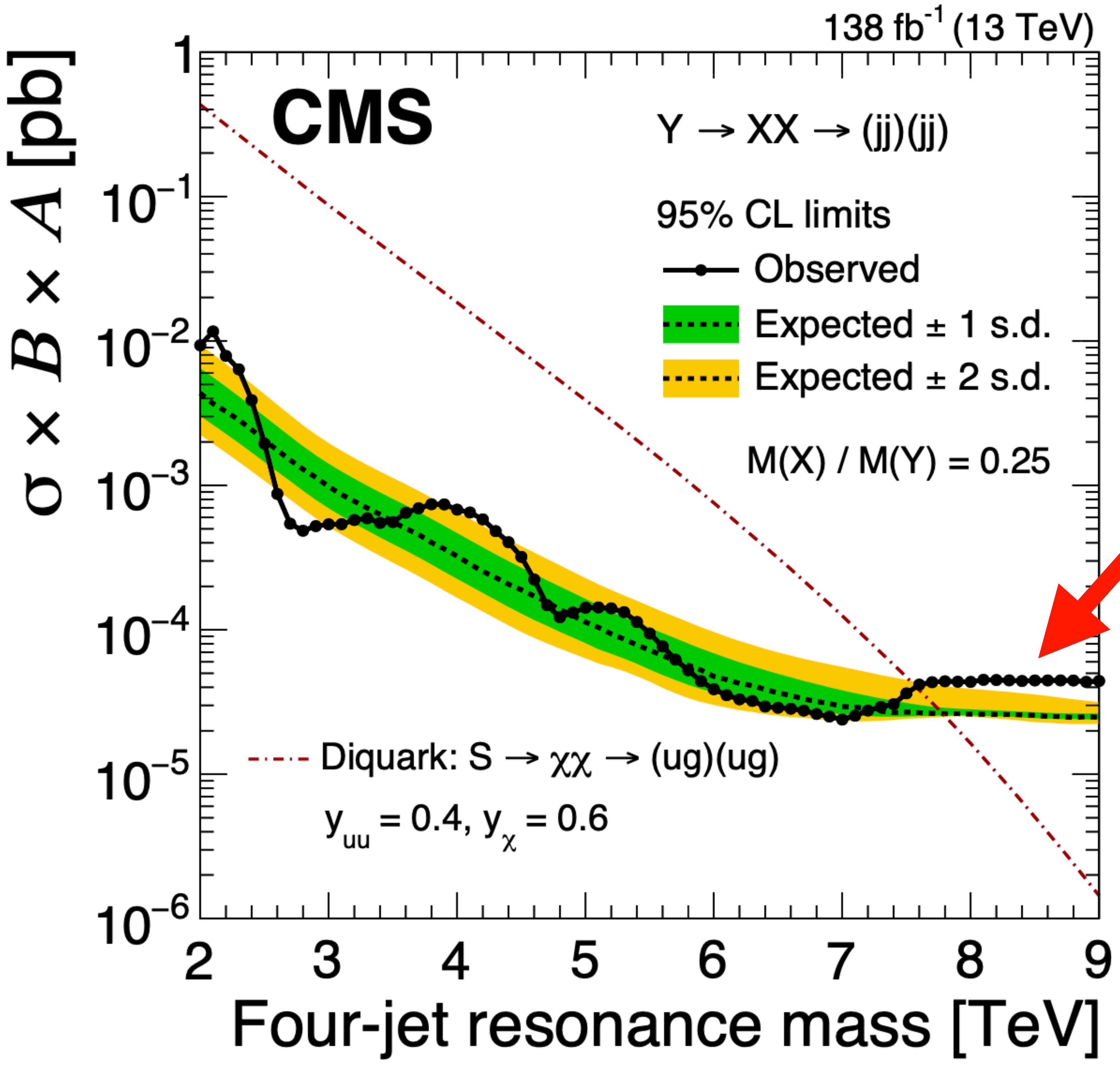
3 parameter smoothly falling spectrum for background:

- 13 bins of α for resonant search
- 3 bins of α in nonresonant search

To include the two events of interest together, a 5 parameter fit is performed to the sum of the alpha bins

Results

Background estimates do not account for the two events observed near $m_{4j} \sim 8TeV$



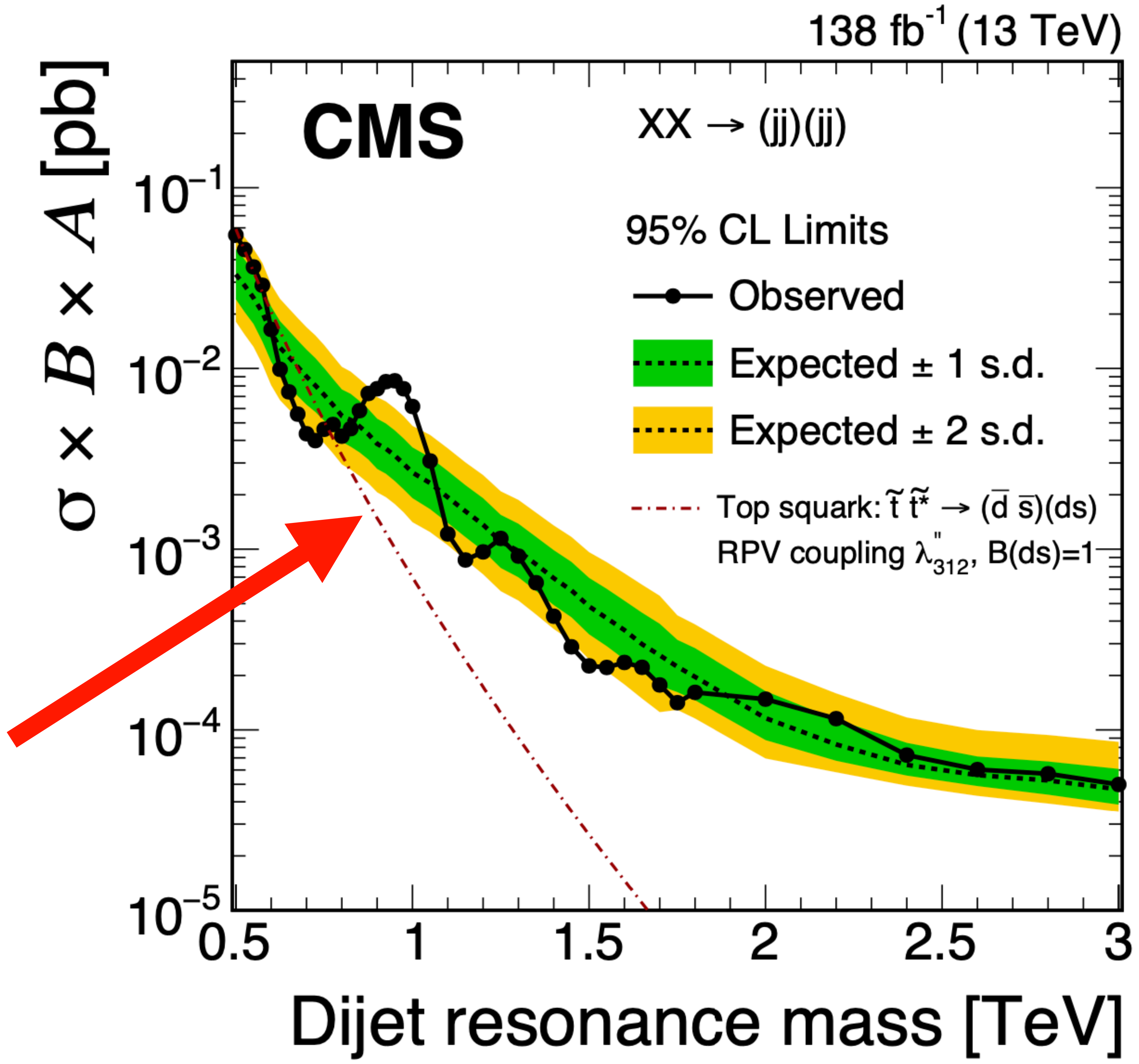
$\sigma_{loc} = 3.9$

$\sigma_{glob} = 1.6$

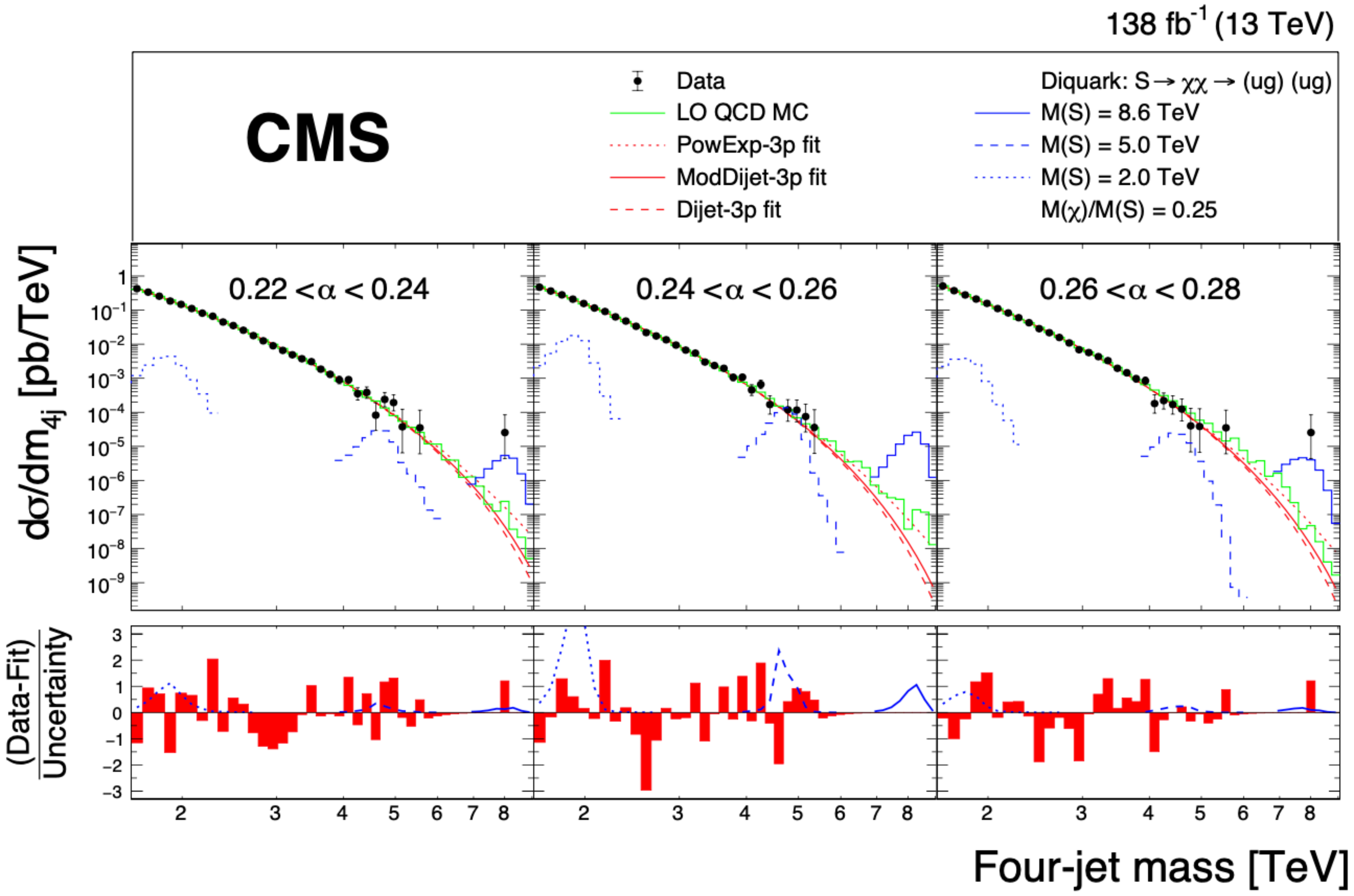
$m_{jj} = 2.1TeV$

$\sigma_{loc} = 3.6$

$\sigma_{glob} = 2.5$



Background estimation and signal extraction



Modified dijet function (ModDijet-3p):

$$\frac{d\sigma}{dm} = \frac{p_0(1-x^{1/3})^{p_1}}{x^{p_2}} \qquad x = \frac{m}{\sqrt{s}}$$

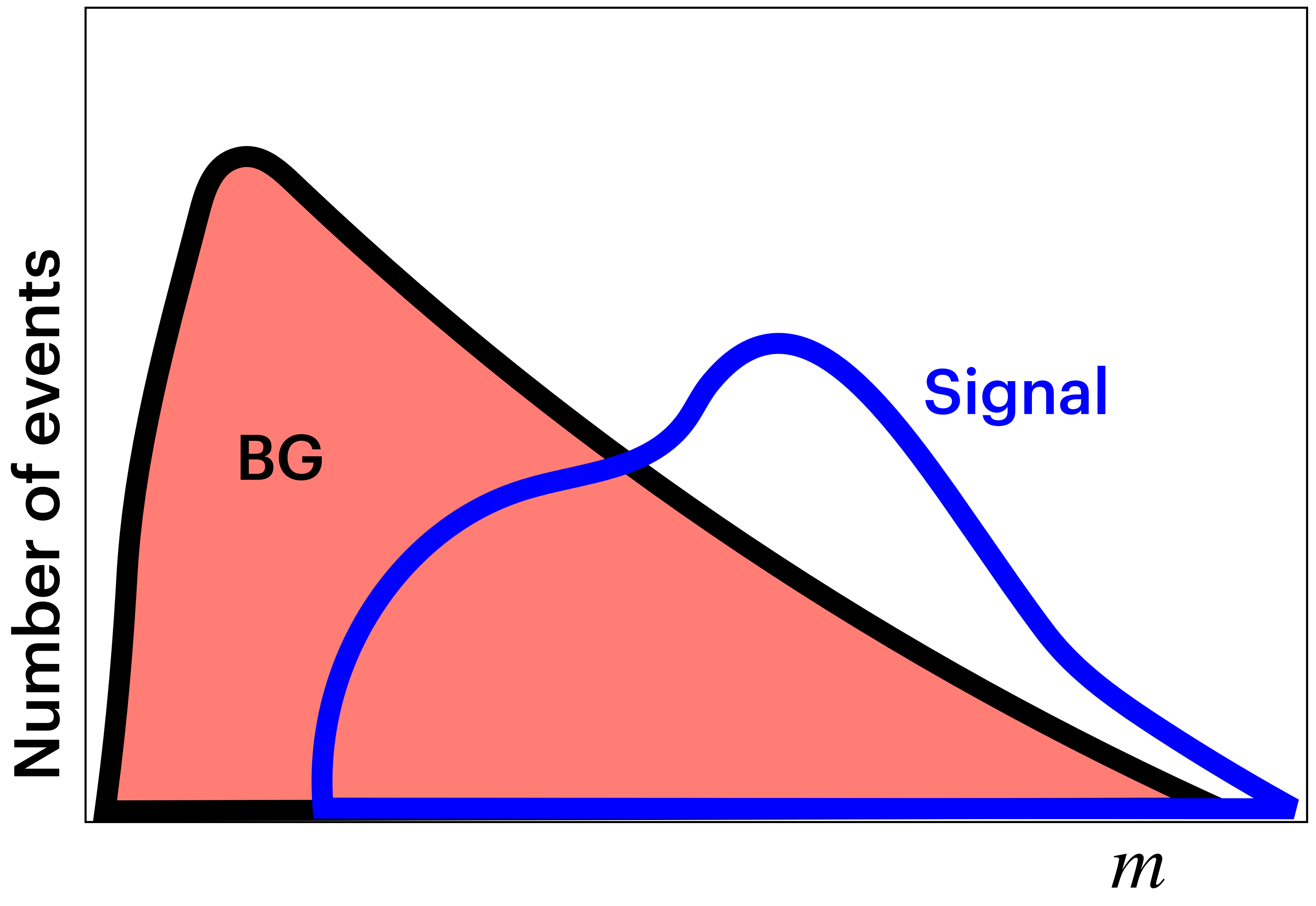
- 13 bins of α for resonant search
- 3 bins of α in nonresonant search

To include the two events of interest together, a ModDijet-5p fit is performed to the sum of the alpha bins

$$\frac{d\sigma}{dm} = \frac{p_0(1-x^{1/3})^{p_1}}{x^{p_2+p_3\log x+p_4\log^2 x}}$$

Background Extraction

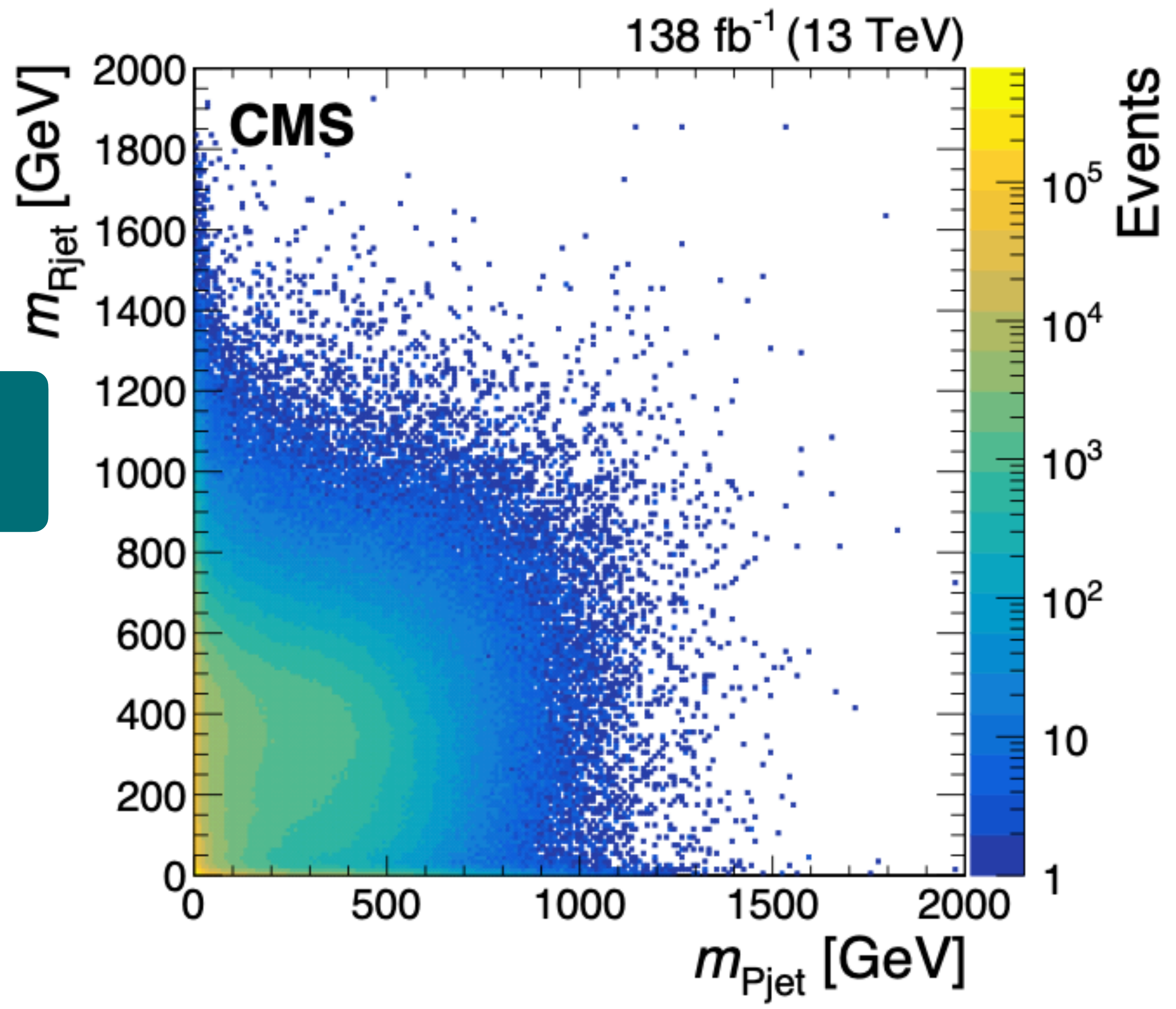
- The bin size is a function of the dijet mass and approximately equal to the dijet mass resolution (5% m(R2))
- At high masses m_{jj} is modified by a turn-on effect — apply threshold $m_{jj}^{thr} = 1.15m_{jj}$
- Simultaneous binned ML fit to m_{jj} for all categories in range $m_{jj}^{min} < m_{jj} < 1.25m(R_1)$ where m_{jj}^{min} is the greater of m_{jj}^{thr} and $0.65m(R_1)$
- If m_{jj}^{thr} > 0.9$m(R_1)$, the signal peak is truncated and the corresponding category is removed from the analysis to avoid signal biases in the fit.
- The total signal efficiency for events to pass the kinematic selection and be included in the fit range is usually between 20 and 30%. In the region with $\rho_m \approx$ 0.2 and $m(R_1) \lesssim 4TeV$, the signal efficiency is approximately 10%, because the signal peak is truncated as described above.
- QCD background simulation is used to optimize analysis strategy, but final bg estimation is done through a fit to data



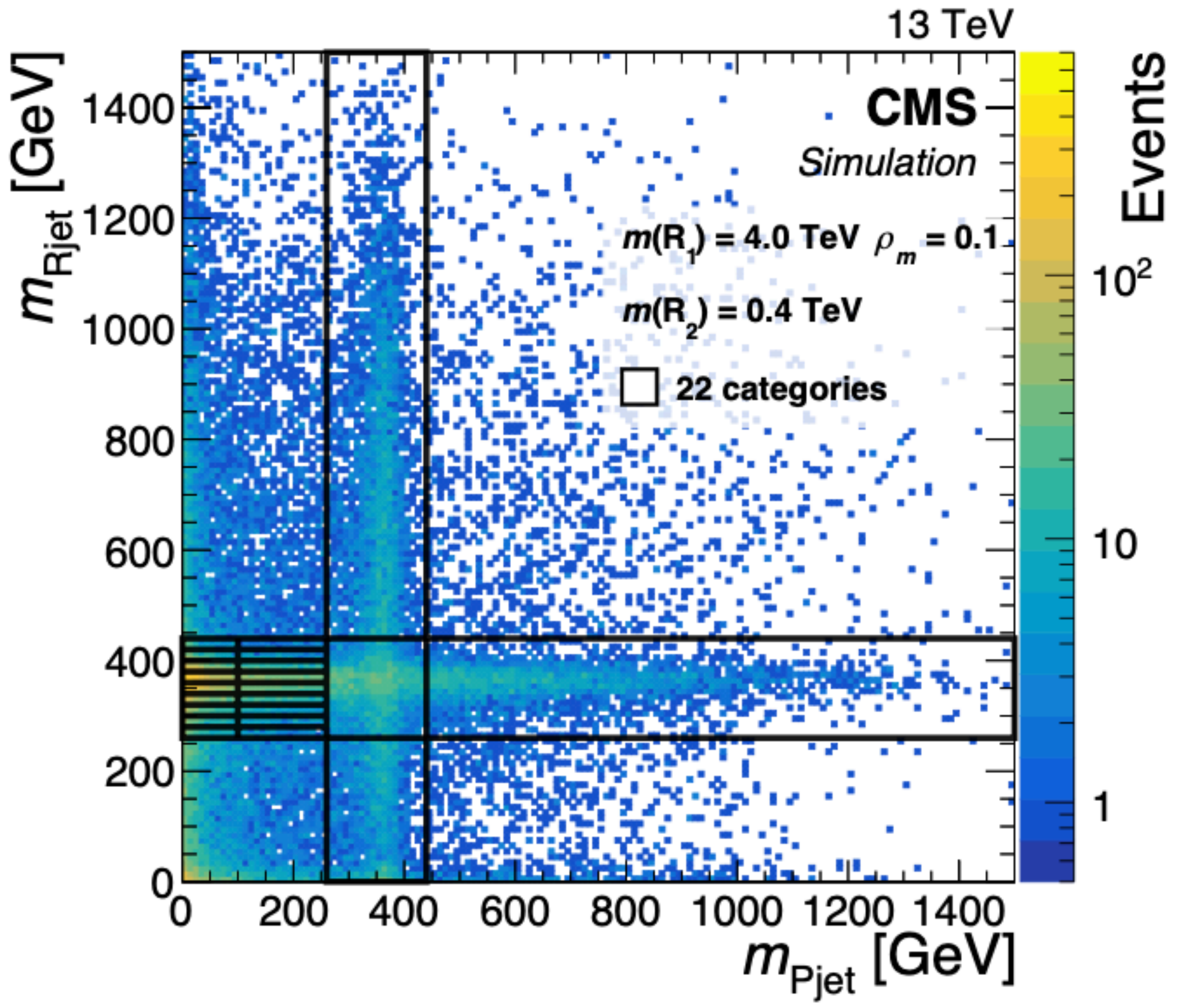
Analysis Strategy

- Binned ML fit to m_{jj}
  - SM background fit: $f(x) = p_0(1-x)^{p_1}/x^{p_2} \quad x = m_{jj}/\sqrt{s}$
  - R_1 signal shape: double sided crystal ball w/ parameters derived from MC fit

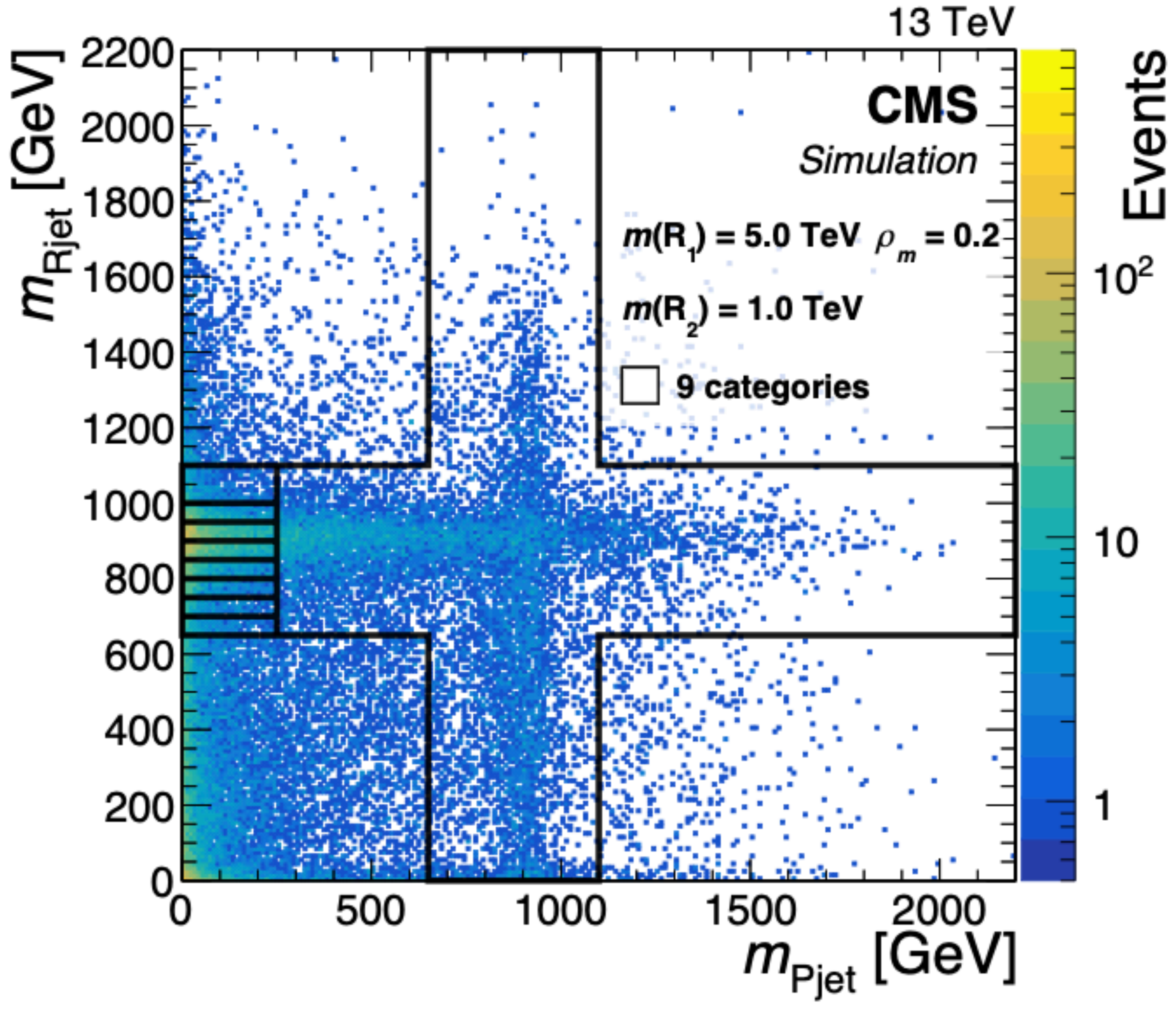
background



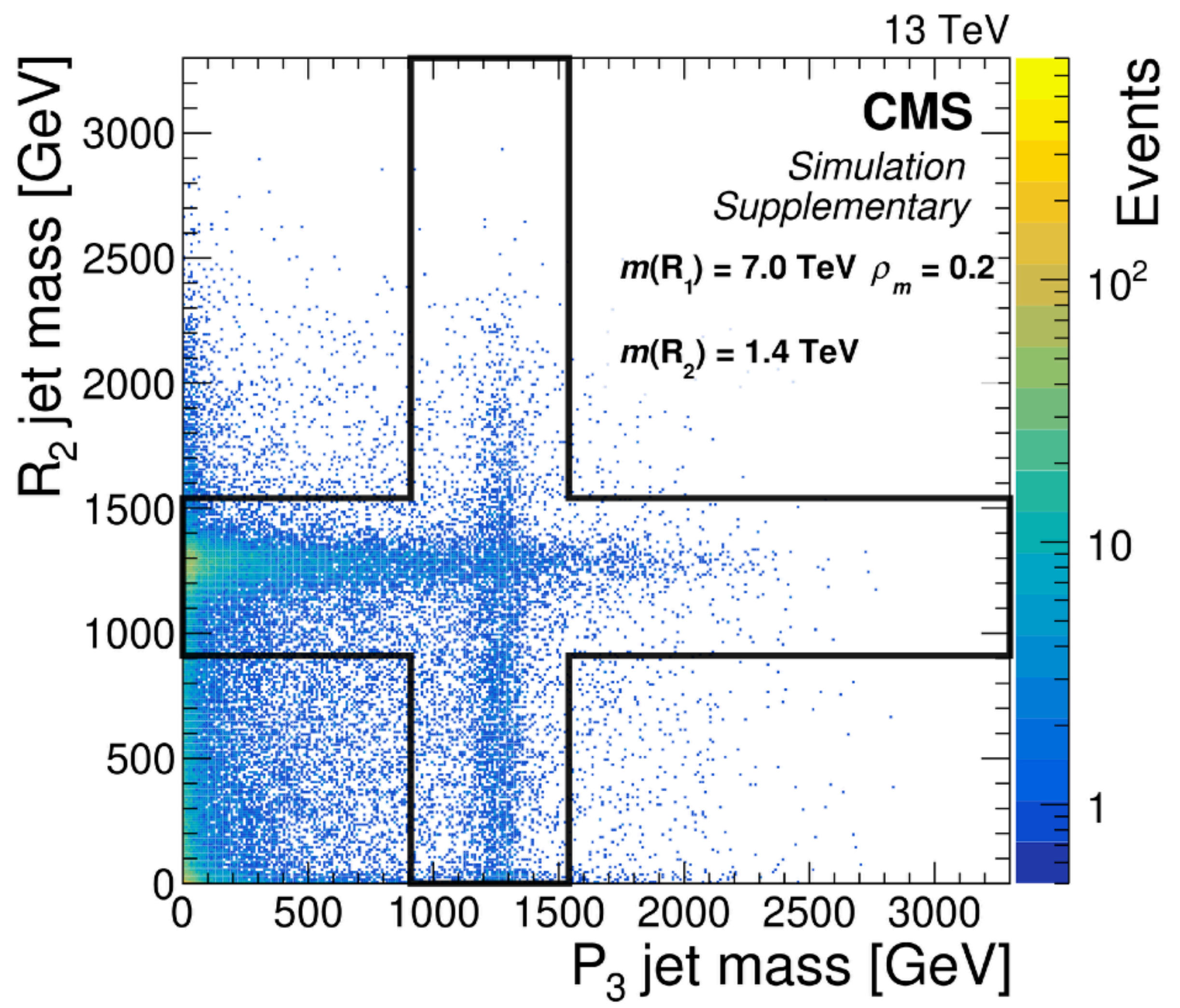
$m(R_2) < 0.6TeV$
22 categorgies



$0.6TeV \leq m(R_2) \leq 1.2$
9 categories

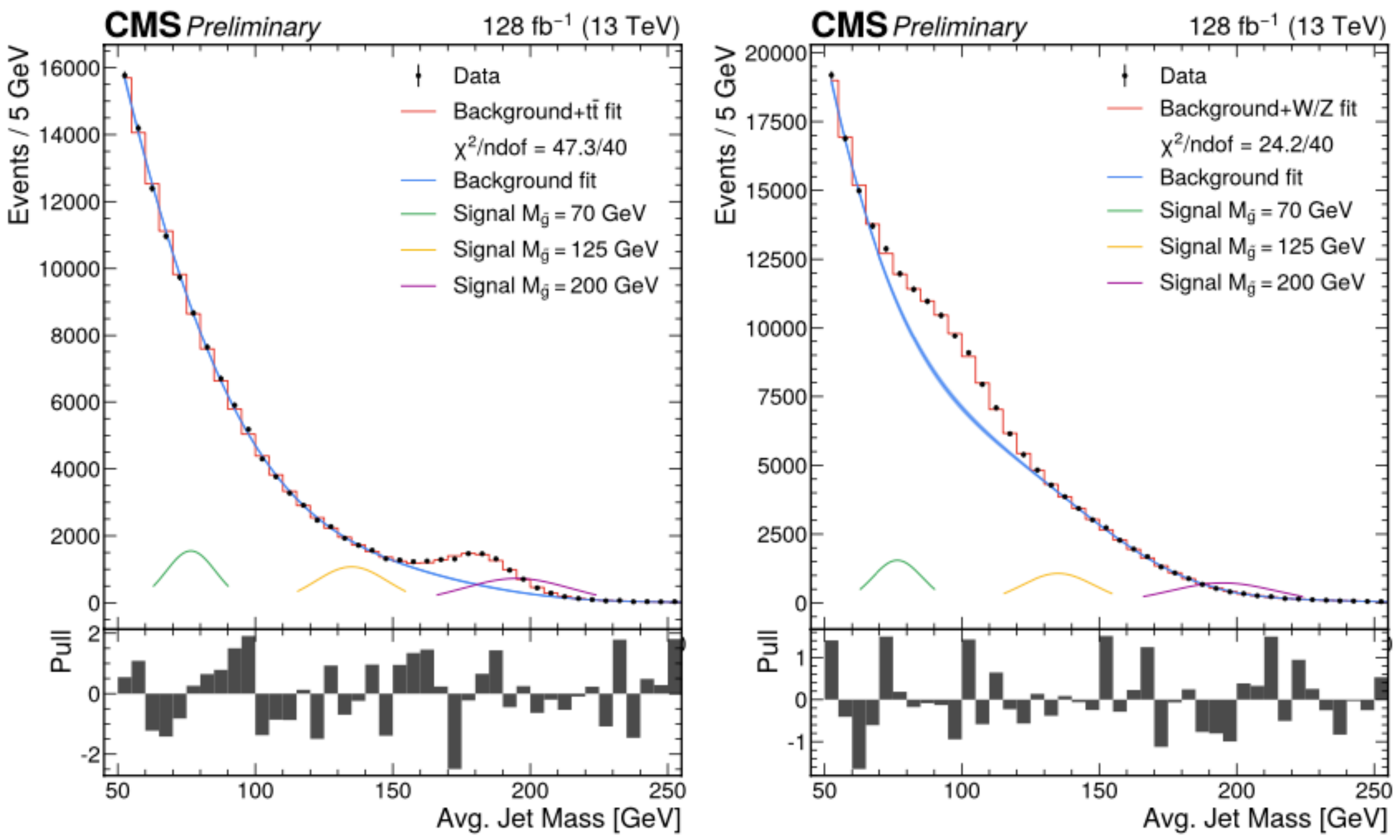


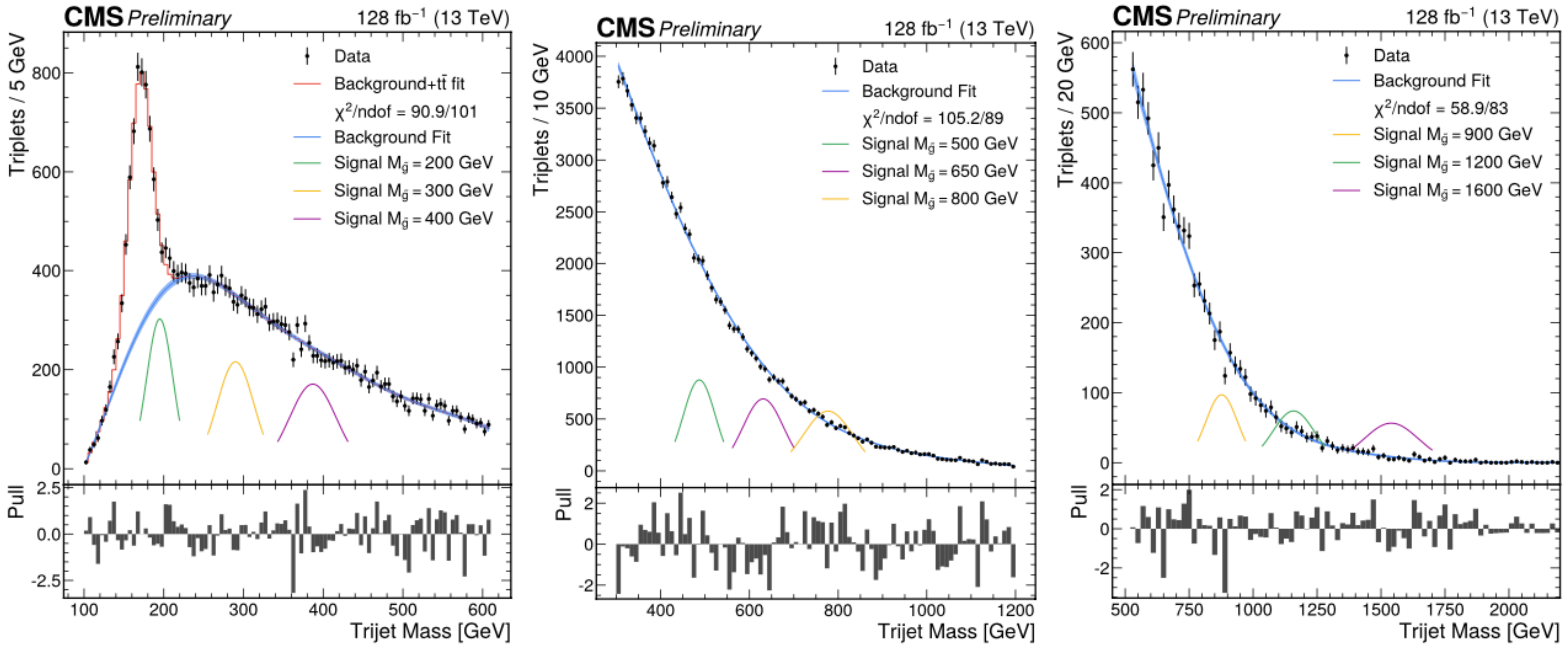
$m(R_2) > 1.2TeV$
1 category



Event selection

- Require at least two AK8 jets w/ $p_T > 300 GeV$ and $|\eta| < 2.4$
- For the resolved trijet resonances implement
  - QGL NN
    - Deep sets PFNet [arxiv:1810.05165]
  - jet ensemble technique
    - Consider an ensemble of 20 possible jet triplets.
  - Dalitz variable [arxiv:1105.2815]
    - measures the geometric spread in the jet topology (expect gluino to be more spread out than QCD background)
    - $$\hat{m}(3,2)^2_{ij} = \frac{m^2_{ij}}{m^2_{ijk} + m^2_i + m^2_j + m^2_k}.$$
    - $$D^2_{[3,2]} = \sum_{i>j} \left(\hat{m}(3,2)_{ij} - \frac{1}{\sqrt{3}} \right)^2.$$





Gaussian Process Regression

[JHEP 02 (2023) 230]

- Traditional method uses an ad-hoc model for background fit — GP's allow for the description of a broader set of functions
- Kernel function defines degree of correlation between any two datapoints
- Gaussian mean function represents the signal
- First mask signal region and fit kernel by maximizing the marginal log likelihood
- Fix kernel hyperparameters and fit to signal + bg w/ kernel + mean function