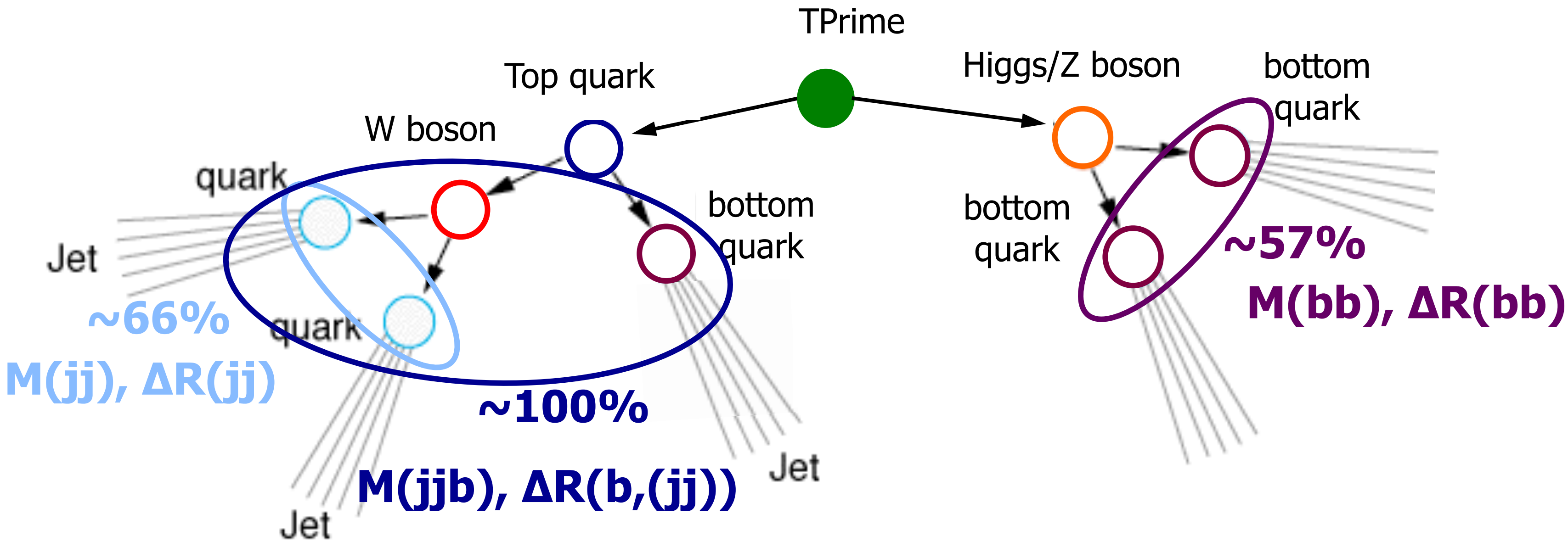
Search for Single VLQ T'→ top+H/Z in all hadronic final state in CMS data

Stéphanie Beauceron
Top LHC France 2024
LPNHE



Vector-Like Quarks

VLQ are colored, fractionally-charged fermions, nonchiral under SU(2) - why these particular particles?

- appear in many BSM models that address the naturalness issue (Little Higgs, extra dimensions, Compositness etc)

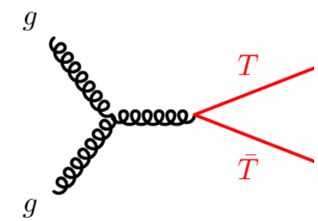
- (maybe) explain fermion mass hierarchy

Candidates:

Name	charge	Lagrangian	Candidates	SM
Y	4/3	$c_L \bar{Y}_L \not{W} b_L$	Y	Triplet
$X_{5/3}$	5/3	$c_{Rt} \bar{X}^R_{5/3} \not{W} t_R + c_{Lt} \bar{X}^L_{5/3} \not{W} t_L$	$X_{5/3}$	Doublet
T'	2/3	$c_{Rt} \bar{T}'_R \not{Z} t_R + c_{Lt} \bar{T}'_L \not{Z} t_R$	$X_{2/3}$	Doublet
		$+c_{Lb} \bar{T}'_L \not{W} b_L$	$\tilde{T}$	Singlet
		$+c_{Rh} h \bar{T}'_R t_L + c_{Lh} h \bar{T}'_L t_L$	T	Doublet
B'	-1/3	$c_{Rt} \bar{B}'_R \not{W} t_R + c_{Lt} \bar{B}'_L \not{W} t_R$	B	Doublet
		$+c_{Lb} \bar{B}'_L \not{Z} b_L$		
		$+c_{Rh} \bar{B}'_R b_L$	$\tilde{B}$	Singlet

Exotic charge partners

Production/Decay

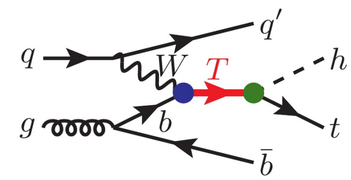


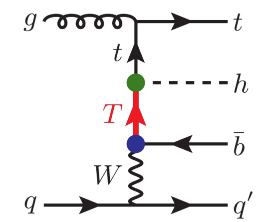
Pair-production:

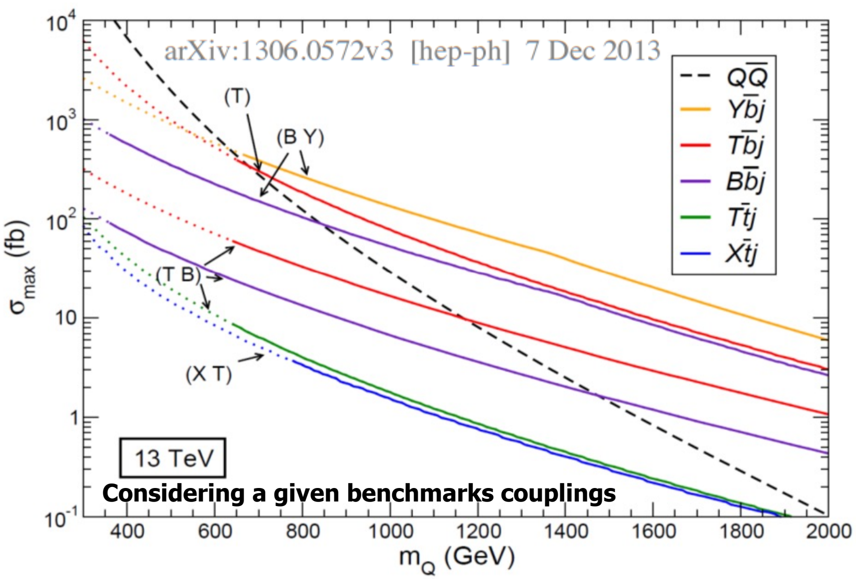
Strong mechanism, the cross section depends only on the VLQ mass

Single production:

Electroweak mechanism, the cross section depends on VLQ mass and on its couplings with SM particles

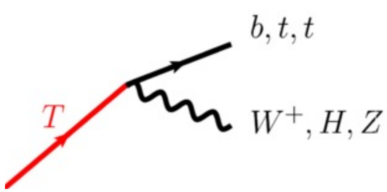


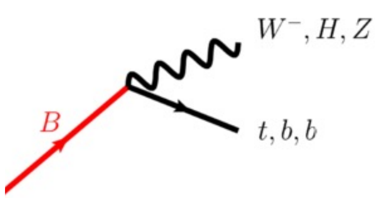




Pair production cross section falling very quickly and single production dominates as soon as 800 GeV for T and Y.

Can decay: (equivalent for Y/X)



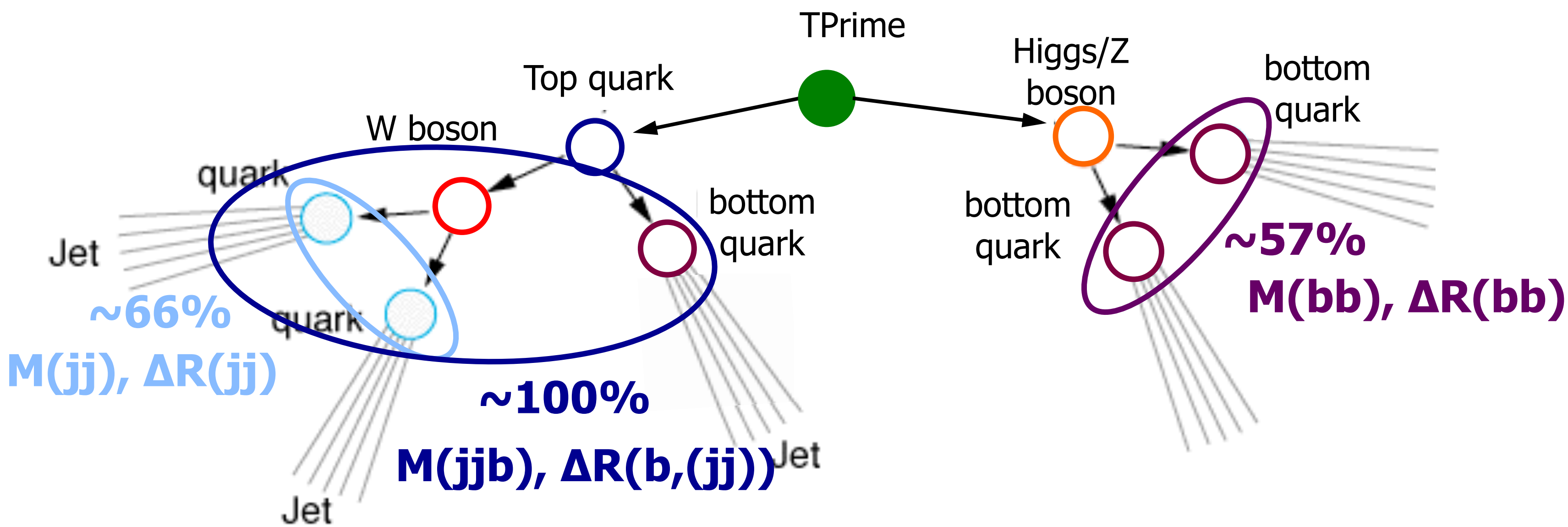


VLQ	W-decay	Z-decay	h-decay
T	Wb	Zt	ht
B	Wt	Zb	hb
$T_{5/3}$	Wt	-	-
$Y_{-4/3}$	Wb	-	-

Signal

Looking for:

Top+bb final state, all hadronic, resolved



→ **Challenges: high background from multijet/ttbar**

→ **Opportunities: full possibility to reconstruct each of the invariant mass and 3 b-tags can be used to constraint multijet events**

→ **M(Top+bb), main variable, and look for a bump!**

Signal Reconstruction

2-step Chi2 minimization: first Higgs/Z, then W/Top:

- **b-tag jets: select 2 → Higgs/Z candidate**
- **Remaining jets: select 2 → W-candidate**
- **Remaining b-tag jets: merge 1 with W-candidate → Top-candidate**

→ 30% gain on S/N (wrt 1 step)

$$\chi^2_{\mathrm{H/Z}} = \left(\frac{m^{\mathrm{meas}}_{\mathrm{H/Z}} - \mu^{\mathrm{MC}}_{\mathrm{H/Z}}}{\sigma^{\mathrm{MC}}_{\mathrm{H/Z}}}\right)^2,$$

$$\chi^2_{\mathrm{W}} = \left(\frac{m^{\mathrm{meas}}_{\mathrm{W}} - \mu^{\mathrm{MC}}_{\mathrm{W}}}{\sigma^{\mathrm{MC}}_{\mathrm{W}}}\right)^2$$

$$\chi^2_{\mathrm{t}} = \left(\frac{m^{\mathrm{meas}}_{\mathrm{t}} - \mu^{\mathrm{MC}}_{\mathrm{t}}}{\sigma^{\mathrm{MC}}_{\mathrm{t}}}\right)^2,$$

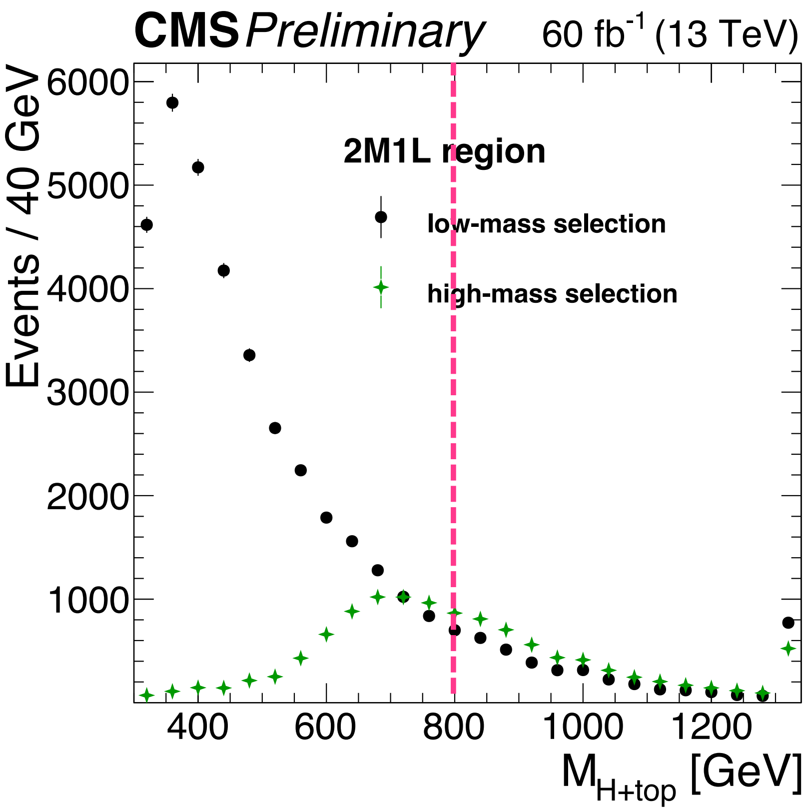
$$\chi^2 = \chi^2_{\mathrm{H/Z}} + \chi^2_{\mathrm{W}} + \chi^2_{\mathrm{t}}$$

Input chi2 values obtained in M=700 GeV samples with matching to MCTruth after trigger + kinematics selection

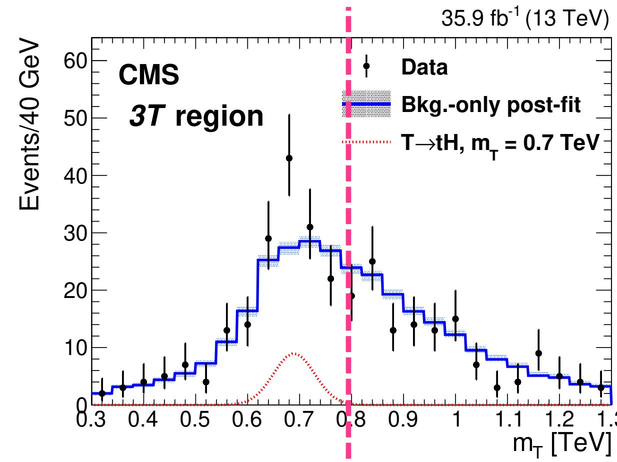
	2016		2017		2018	
Particle	μ^{MC}	σ^{MC}	μ^{MC}	σ^{MC}	μ^{MC}	σ^{MC}
H	121.9	13.5	118.9	14.7	120.2	14.3
W	83.8	10.9	82.5	12.6	83.9	10.8
t	173.8	16.0	172.8	18.9	175.9	17.2
Z	90.9	11.4	89.2	12.0	90.9	11.3

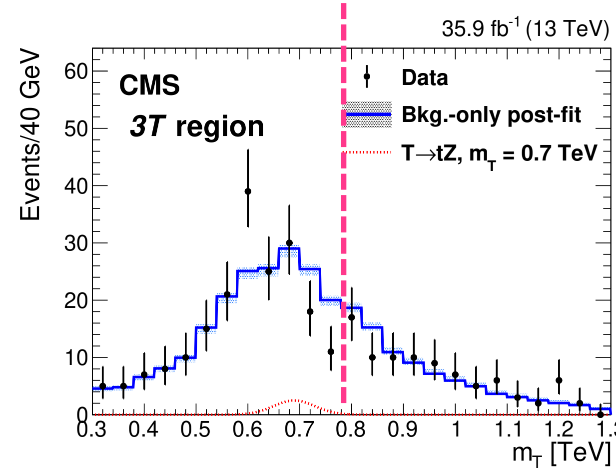
Two Selections

The published previous analysis induces a shaping of the main variable, defined a second selection to remove the shaping
➔ more robust analysis



JHEP01(2020)036





Define 2 regions of selection:

- **low mass selection (m(top+bb)<~800 GeV)**
- **high mass selection (m(top+bb)>~800 GeV)**

High Mass Selection

Label	2M1L	3M	3T
Basic selection	Trigger	Trigger	Trigger
	6 jets $p_T > 40$ GeV/c, $\lvert\eta\rvert < 4.5$	6 jets $p_T > 40$ GeV/c, $\lvert\eta\rvert < 4.5$	6 jets $p_T > 40$ GeV/c, $\lvert\eta\rvert < 4.5$
	$j^1_{p_T} > 170$ GeV/c	$j^1_{p_T} > 170$ GeV/c	$j^1_{p_T} > 170$ GeV/c
Cut 0	$j^2_{p_T} > 130$ GeV/c	$j^2_{p_T} > 130$ GeV/c	$j^2_{p_T} > 130$ GeV/c
	$j^3_{p_T} > 80$ GeV	$j^3_{p_T} > 80$ GeV	$j^3_{p_T} > 80$ GeV
	$H_T > 500$ GeV/c	$H_T > 500$ GeV/c	$H_T > 500$ GeV/c
B - tagging	**2M1L vetoing 3M**	**3 Medium vetoing 3T**	3 Tight
	$\chi^2 < 15$	$\chi^2 < 15$	$\chi^2 < 15$
	2nd Top Mass > 250 GeV/c^2	2nd Top Mass > 250 GeV/c^2	2nd Top Mass > 250 GeV/c^2
	Higgs Mass > 100 GeV/c^2	Higgs Mass > 100 GeV/c^2	Higgs Mass > 100 GeV/c^2
Cut 1	Relative $H_T > 0.4$	Relative $H_T > 0.4$	Relative $H_T > 0.4$
Cut 2	$\mathrm{Max}(\chi^2) < 3$	$\mathrm{Max}(\chi^2) < 3$	$\mathrm{Max}(\chi^2) < 3$
Cut 3	$\Delta R(b_{Higgs}, b_{Higgs}) < 1.1$	$\Delta R(b_{Higgs}, b_{Higgs}) < 1.1$	$\Delta R(b_{Higgs}, b_{Higgs}) < 1.1$
Cut 4	$\chi^2_{Higgs} < 1.5$	$\chi^2_{Higgs} < 1.5$	$\chi^2_{Higgs} < 1.5$
Cut 5	$\Delta R(j_W, j_W) < 1.75$	$\Delta R(j_W, j_W) < 1.75$	$\Delta R(j_W, j_W) < 1.75$
Cut 6	$\Delta R(b_{Top}, W) < 1.2$	$\Delta R(b_{Top}, W) < 1.2$	$\Delta R(b_{Top}, W) < 1.2$

Top+Z: in chi2, M(bb) is set at mass(Z), in the overall selection, Higgs Mass<100 GeV AND chi2_Higgs < 1.0 (instead of 1.5)

Some Top+H can be reconstructed in top+Z final state (~15%)

→ Look at top+H, top+Z and top+H reconstructed in top+Z

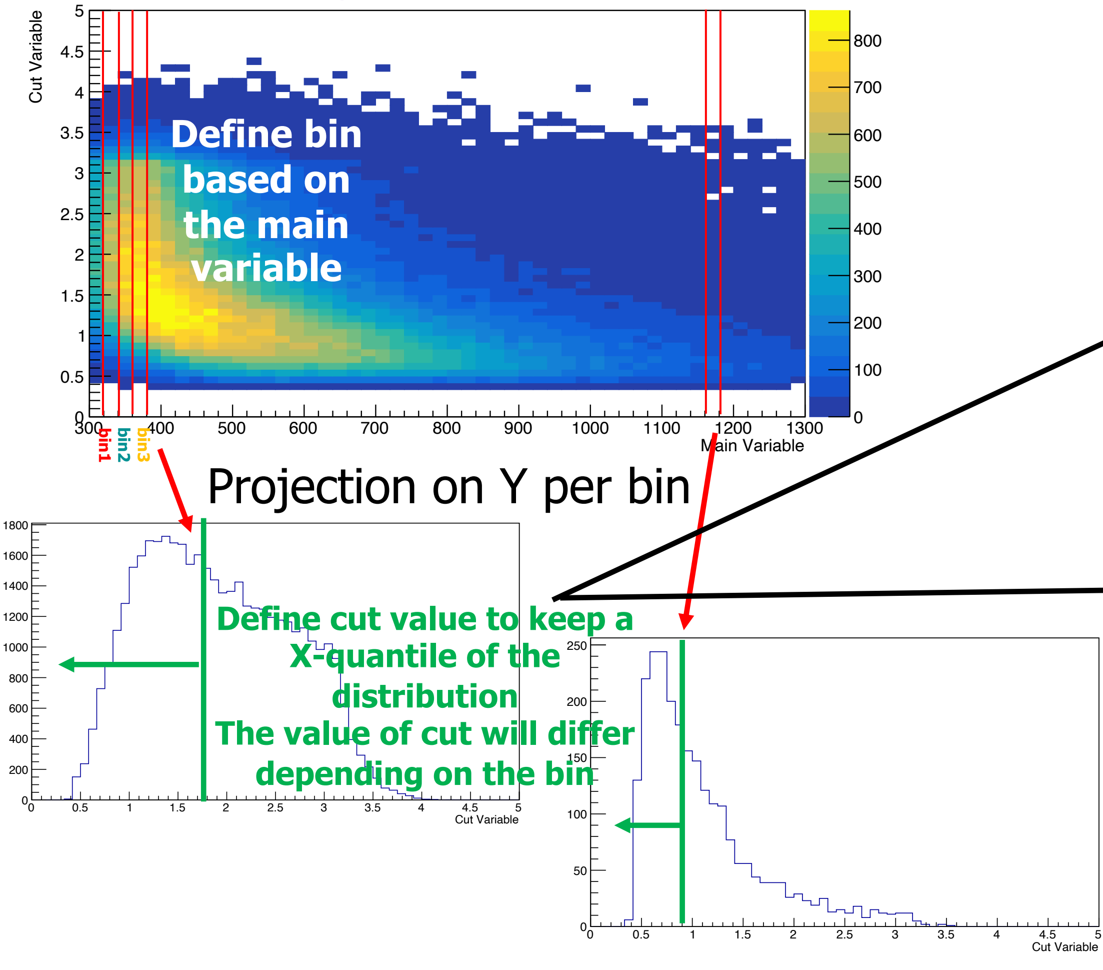
→ If bb resonance not a Higgs, more events could migrate

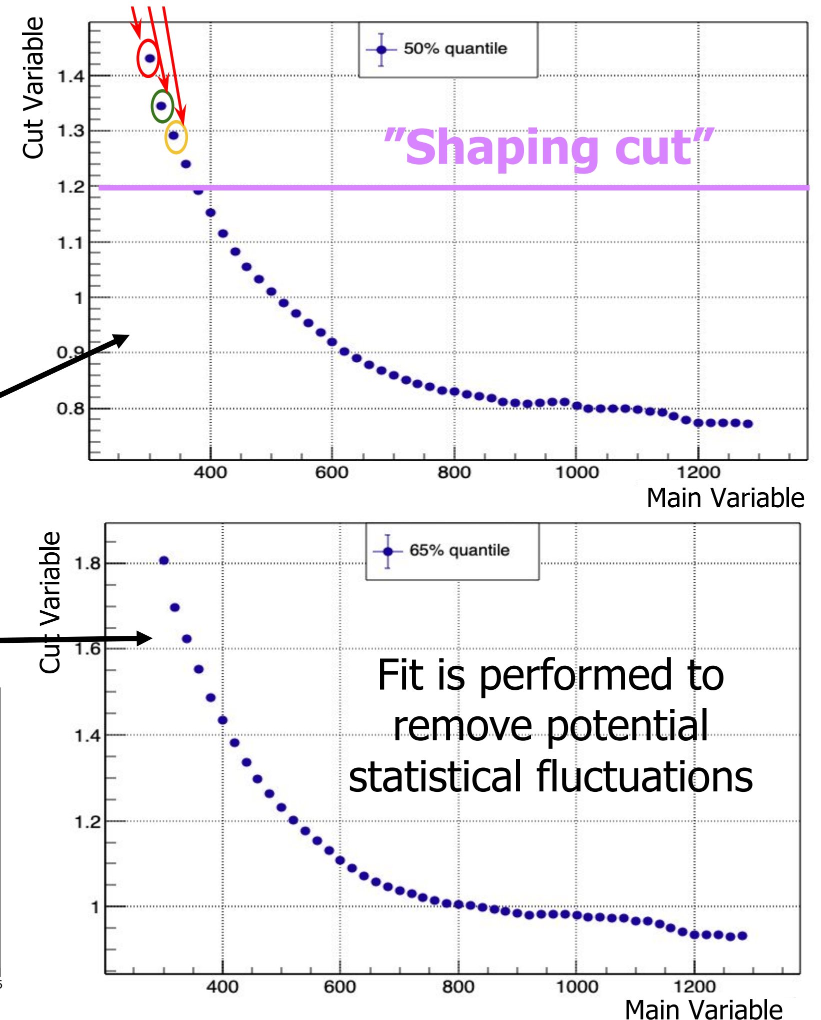
Remove Selection Bias

The same criteria than before are used, simply modify the cut value: Define a cut as function of the main variable to keep a quantile of events from the previous cut

➔ Preserving the shape as just reducing the shape by a given quantile

How to design the new cut value:





Low Mass Selection

Cut 1	Relative $H_T > 0.4$
Cut 2	$\mathrm{Max}(\chi^2) < 3$
Cut 3	$\Delta R(b_{Higgs}, b_{Higgs}) < 1.1$
Cut 4	$\chi^2_{Higgs} < 1.5$
Cut 5	$\Delta R(j_W, j_W) < 1.75$
Cut 6	$\Delta R(b_{Top}, W) < 1.2$

This two variables are kept as is, the others induce a shaping

For the other cuts, polynomials from order 5 to order 7 are used to apply the cut

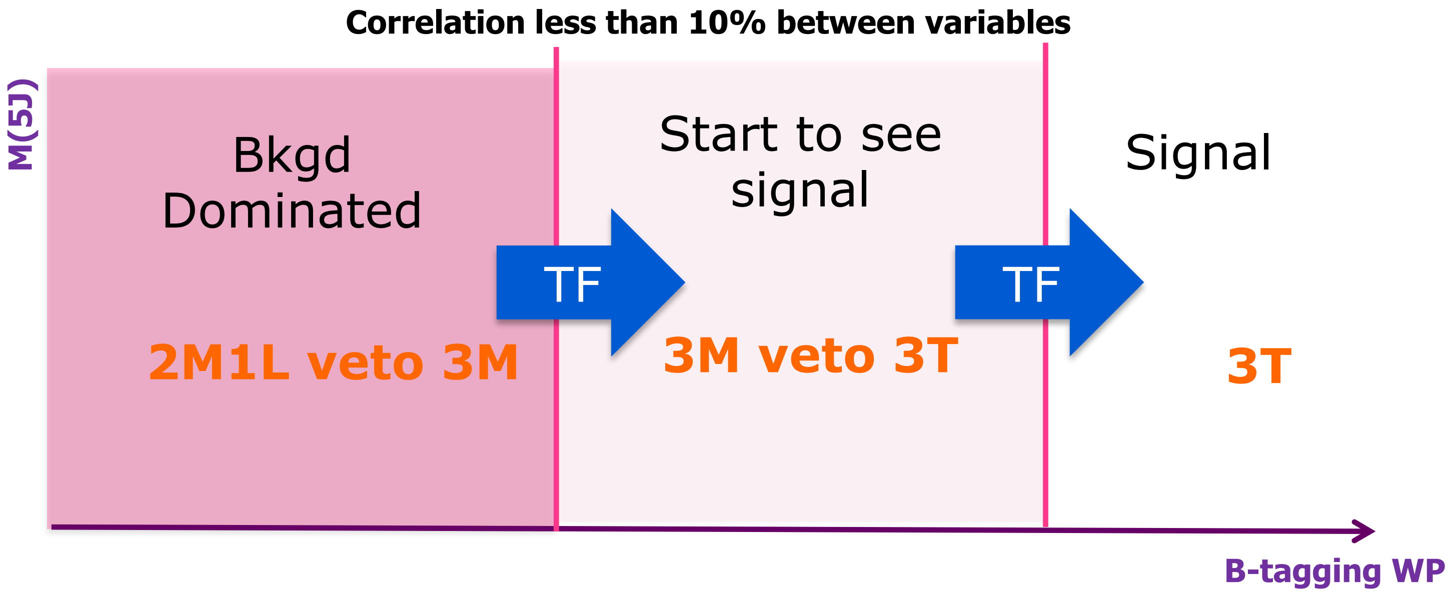
Cuts	Fraction of events kept
3: $\Delta R(b, b)$	35%
4: $\Delta R(j_W, j_W)$	20%
5: Relative H_T	99.5%
6: $\Delta R(Top, W)$	15%

→ Fraction designed to keep the same efficiency than the high mass cut at M=700 GeV (value used for cut designed)

Why keeping 2 selections?

The same criteria are used for top+H and top+Z channels, as low mass selection is always preserving a given fraction of the input variable → it is cutting tighter at higher mass
➔ lower efficiency than the high mass selection, for example, at M=900 GeV, high-mass selection is 46% efficient while the low-mass is 36% for top+H channel

Background Estimation



- Use shape of 2M1L/3M to estimate background, transition between regions done via Transfer Function (TF [b-tag weights])
- Normalisation of the shape done in combine
- Hypothesis: the shape of background in 2M1L==3M==3T
- Define validation regions to validate the hypothesis
→ Produce pre-Fit plots
- Use Combine for background shape and final results

➔ Independent of the selection (identical to previous publication)

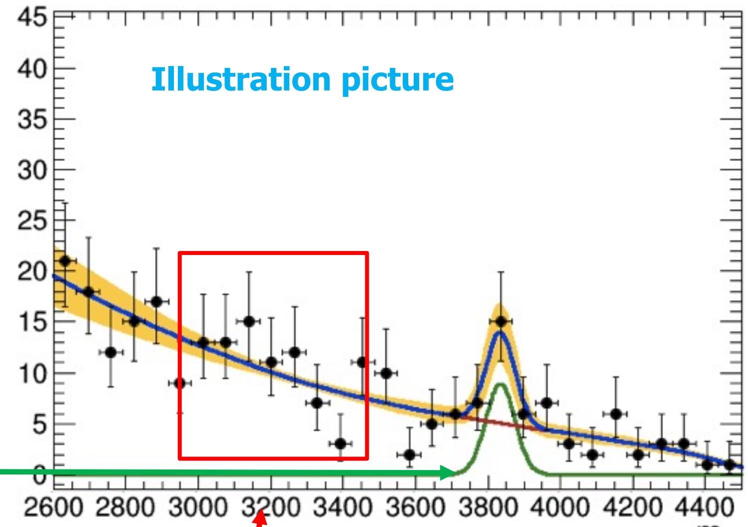
Background Shape in the Fit

Like doing a matrix inversion:

Signal: clustered in a few bins

In each b-tagging region, the bin content is either S+B or B only: (r=μ, signal strength)

$$n_{3T}^{data}(bin) = B_{3T}(bin) + \mu \times S_{3T}(bin)$$
$$n_{3M}^{data}(bin) = B_{3M}(bin) + \mu \times S_{3M}(bin)$$
$$n_{2M1L}^{data}(bin) = B_{2M1L}(bin) + \mu \times S_{2M1L}(bin)$$



Signal → known (gaussian from MC)

Background links between regions by b-tag efficiency:

$$B_{3T}(bin) = Norma_{3M}^{3T} \times TF_{3Mto3T}(bin) \times B_{3M}(bin)$$
$$B_{3M}(bin) = Norma_{2M1L}^{3M} \times TF_{2M1Lto3M}(bin) \times B_{2M1L}(bin)$$

→Transfer Function (TF): change of b-tag eff. as function of M(H/Z+top) [later]

→Norma$_X^Y$: overall b-tag efficiency from region X to Y

$$n_{3T}^{data}(bin) = Norma_{2M1L}^{3M} \times Norma_{3M}^{3T} \times TF_{2M1Lto3M}(bin) \times TF_{3Mto3T}(bin) \times B_{2M1L}(bin) + \mu \times S_{3T}(bin)$$
$$n_{3M}^{data}(bin) = Norma_{2M1L}^{3M} \times TF_{2M1Lto3M}(bin) \times B_{2M1L}(bin) + \mu \times S_{3M}(bin)$$
$$n_{2M1L}^{data}(bin) = B_{2M1L}(bin) + \mu \times S_{2M1L}(bin)$$

→Norma$_X^Y$ → determined by the fit method in bins without signal

→Solvable system, fitting [$B_{2M1L}(bin)$] by taking into account all the parameters' errors (bin statistics + systematics)

Transfer Function

Physics concerns:

- B-tagging efficiency is not flat with respect to η x pt
- Looser b-tagging prefer jets at high η (wrt to medium b-tag)

→ Selected 4-vector jets → slightly different in b-tag regions

➔ Calculated b-tagging weights to correct each of the b-tag region

➔ = Transfer Function

Per jets, ratio of loose/medium selected b-tag as function of momentum/ η
→fitted function for each histograms

The weight for 3M → 3T is:

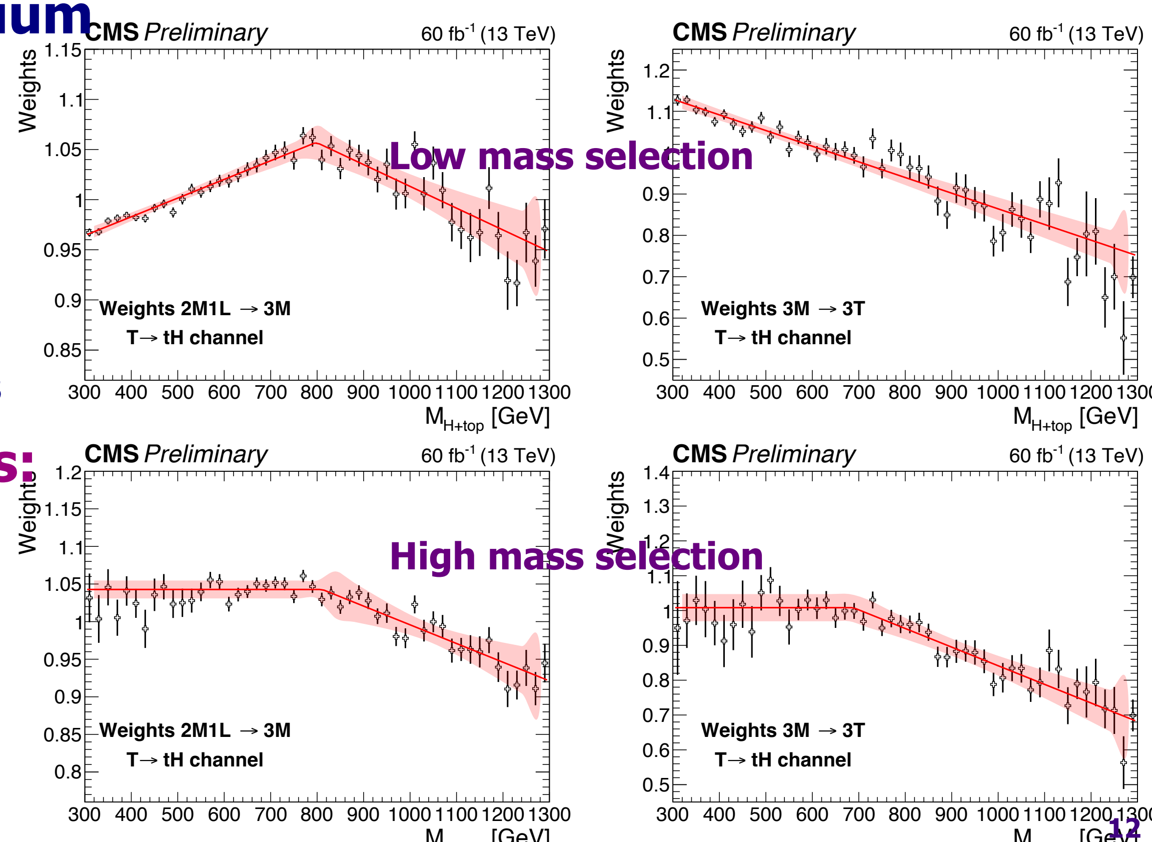
$(W^{momentum}_{Medium} * W^{\eta}_{Medium})^3$

The weight for 2M1L→ 3M is:

$W^{momemtum}_{Loose} * W^{\eta}_{Loose}$

Low mass selection

High mass selection



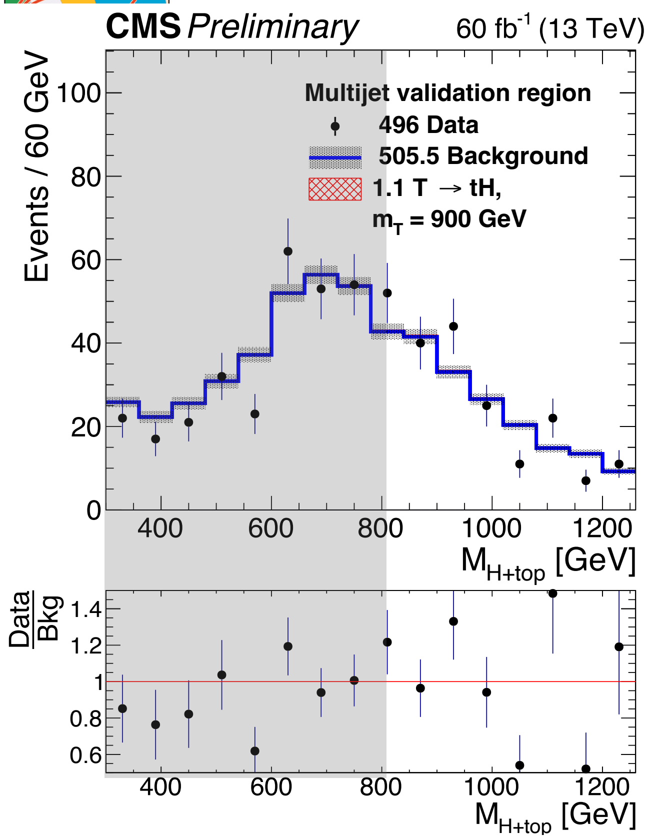
Validation Regions

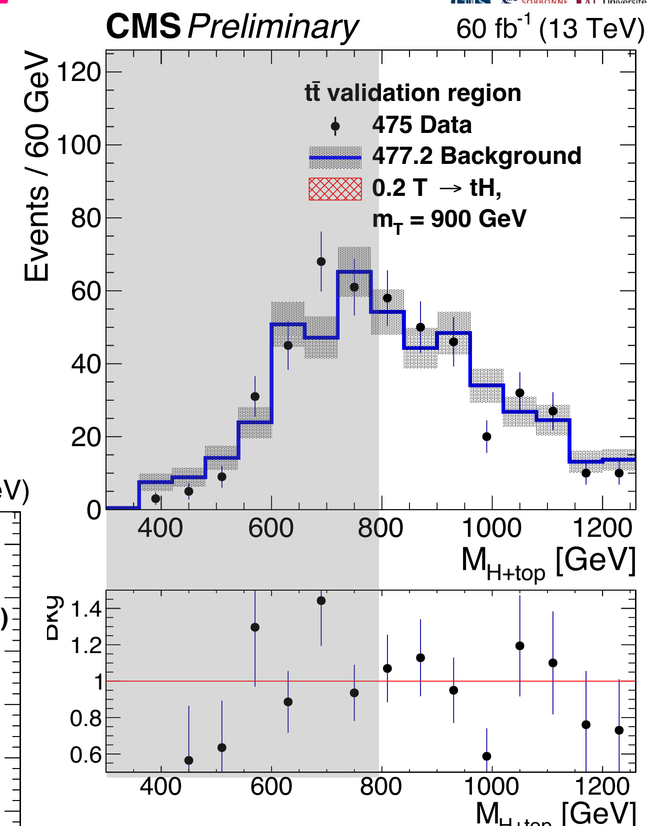
Label	QCD 2M1L CR	QCD 3T CR	$t\bar{t}$ 2M1L CR	$t\bar{t}$ 2T1L CR
Basic selection	Trigger	Trigger	Trigger	Trigger
Cut 0	6 jets $p_T > 40$ GeV/c, $\|\eta\| < 4.5$ $j^1_{p_T} > 170$ GeV/c $j^2_{p_T} > 130$ GeV/c $j^3_{p_T} > 80$ GeV $H_T > 500$ GeV/c	6 jets $p_T > 40$ GeV/c, $\|\eta\| < 4.5$ $j^1_{p_T} > 170$ GeV/c $j^2_{p_T} > 130$ GeV/c $j^3_{p_T} > 80$ GeV $H_T > 500$ GeV/c	6 jets $p_T > 40$ GeV/c, $\|\eta\| < 4.5$ $j^1_{p_T} > 170$ GeV/c $j^2_{p_T} > 130$ GeV/c $j^3_{p_T} > 80$ GeV $H_T > 500$ GeV/c	6 jets $p_T > 40$ GeV/c, $\|\eta\| < 4.5$ $j^1_{p_T} > 170$ GeV/c $j^2_{p_T} > 130$ GeV/c $j^3_{p_T} > 80$ GeV $H_T > 500$ GeV/c
B - tagging	**2M1L vetoing 3M** $\boldsymbol{\chi^2 < 50}$ 2nd Top Mass> 250 GeV/c^2 Higgs Mass> 100 GeV/c^2	3 Tight $\boldsymbol{\chi^2 < 50}$ 2nd Top Mass> 250 GeV/c^2 Higgs Mass> 100 GeV/c^2	**2M1L vetoing 2T1L** $\boldsymbol{\chi^2 < 50}$ 2nd Top Mass> 250 GeV/c^2 Higgs Mass> 100 GeV/c^2 **Top b-tag Medium**	**2 Tight +1 Loose** $\boldsymbol{\chi^2 < 50}$ 2nd Top Mass> 250 GeV/c^2 Higgs Mass> 100 GeV/c^2 **Top b-tag Tight**
Cut 1	Relative $H_T > 0.4$	Relative $H_T > 0.4$	Relative $H_T > 0.4$	Relative $H_T > 0.4$
Cut 2	**5<Max(χ^2) < 20 and $\chi^2_{Top} > 1.0$**	**5<Max(χ^2) < 20 and $\chi^2_{Top} > 1.0$**	**3<Max(χ^2) < 5**	**3<Max(χ^2) < 5**
Cut 3	$\Delta R(b_{Higgs}, b_{Higgs}) < 1.1$	$\Delta R(b_{Higgs}, b_{Higgs}) < 1.1$	$\boldsymbol{\Delta R(b_{Higgs}, b_{Higgs}) < 1.5}$	$\boldsymbol{\Delta R(b_{Higgs}, b_{Higgs}) < 1.5}$
Cut 4	$\chi^2_{Higgs} < 1.5$	$\chi^2_{Higgs} < 1.5$	**$\chi^2_{Top} < 1.5$ and $\chi^2_{Higgs} > 3$**	**$\chi^2_{Top} < 1.5$ and $\chi^2_{Higgs} > 3$**
Cut 5	$\Delta R(j_W, j_W) < 1.75$	$\Delta R(j_W, j_W) < 1.75$	$\Delta R(j_W, j_W) < 1.75$	$\Delta R(j_W, j_W) < 1.75$
Cut 6	$\Delta R(b_{Top}, W) < 1.2$	$\Delta R(b_{Top}, W) < 1.2$	$\Delta R(b_{Top}, W) < 1.2$	$\Delta R(b_{Top}, W) < 1.2$

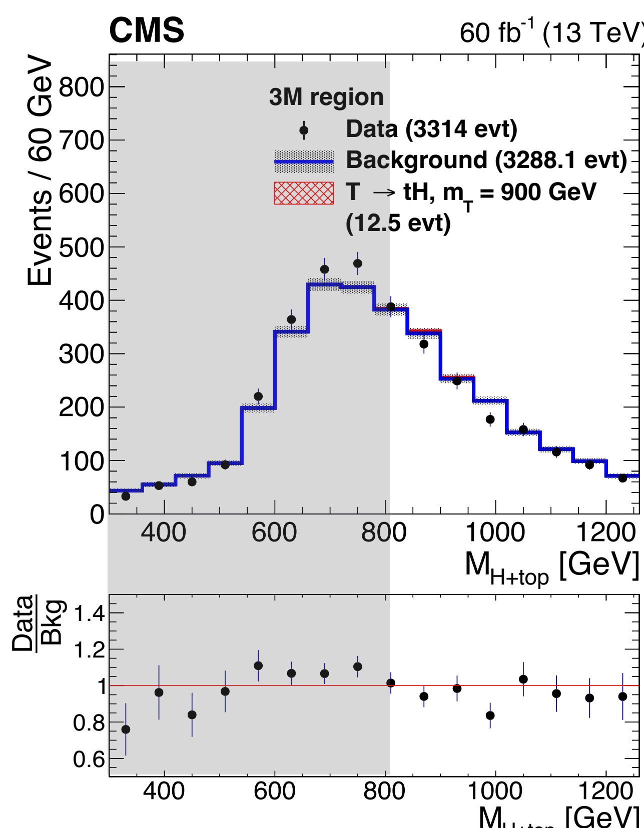
Each region is dominated (>70%) by the expected background
Up to Cut2, QCD 2M1L and tt2M1L are almost identical

+Mtop=140 GeV:
In the chi2, instead of Mtop from MC, fix it to 140 GeV
Produce VR SR/QCD/ttbar with this new chi2
+Mtop=250 GeV (lower stat)
➔ 45 validation regions exist… Performed in high mass selection are more complex shape

Validation Regions







Signal Parametrisation and Systematics

Thanks to full hadronic final state, signal shape is gaussian!
Parametrize signal shape (mean and sigma of the gaussian) for each of the b-tag regions/each selection as function of Mgen
→ Linear dependance, simple parametrisation

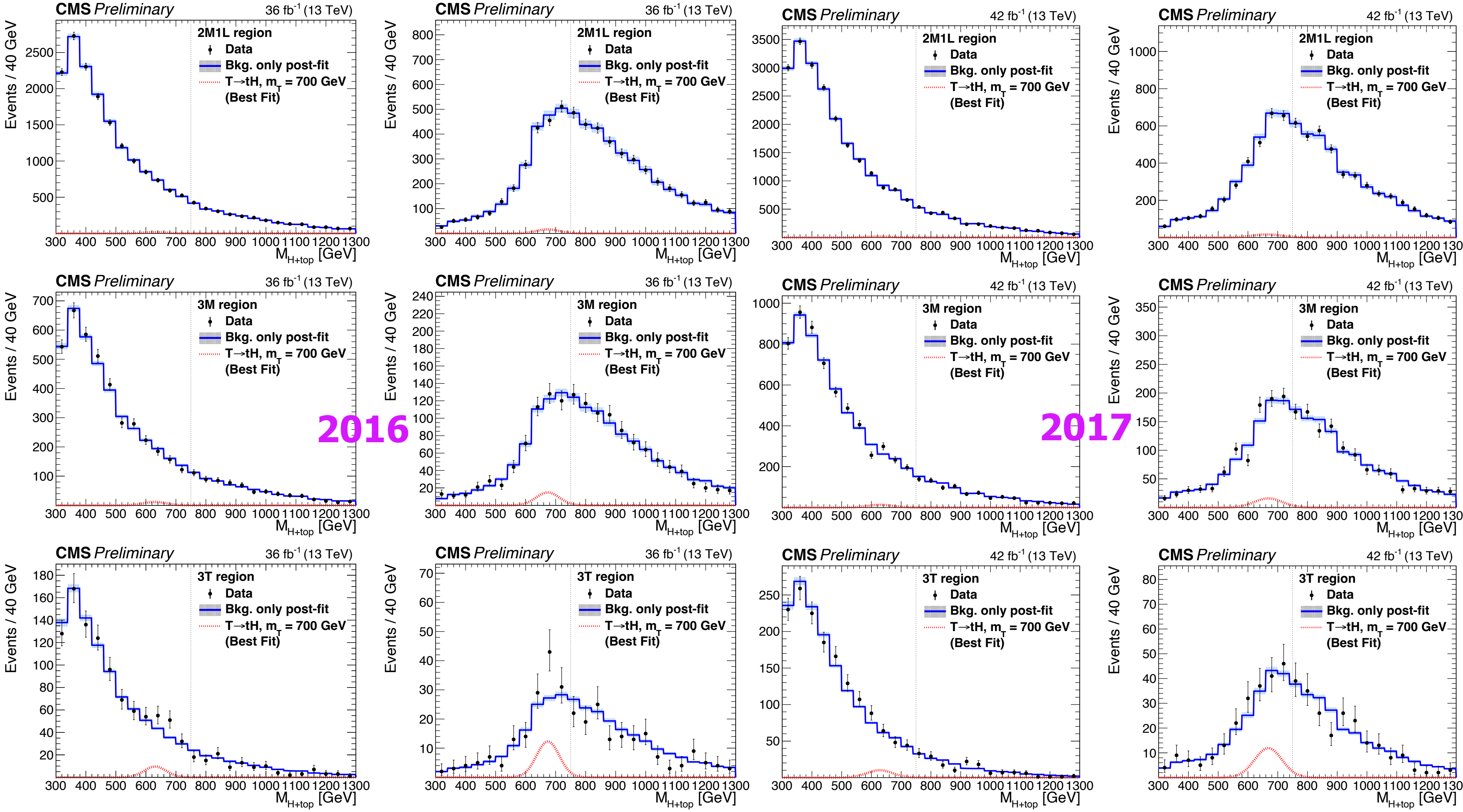
Systematics:
Background is coming from data → no systematics
Systematics error on the transfer function (from the fit parameter and also flavor composition in the sample used for the determination)

For the signal, most of them are simply rate errors (luminosity/b-tagging/PU/PDF etc), some of them are slightly changing the parametrisation (JES/JER)

Single T→ tH/Z

B2G-19-001

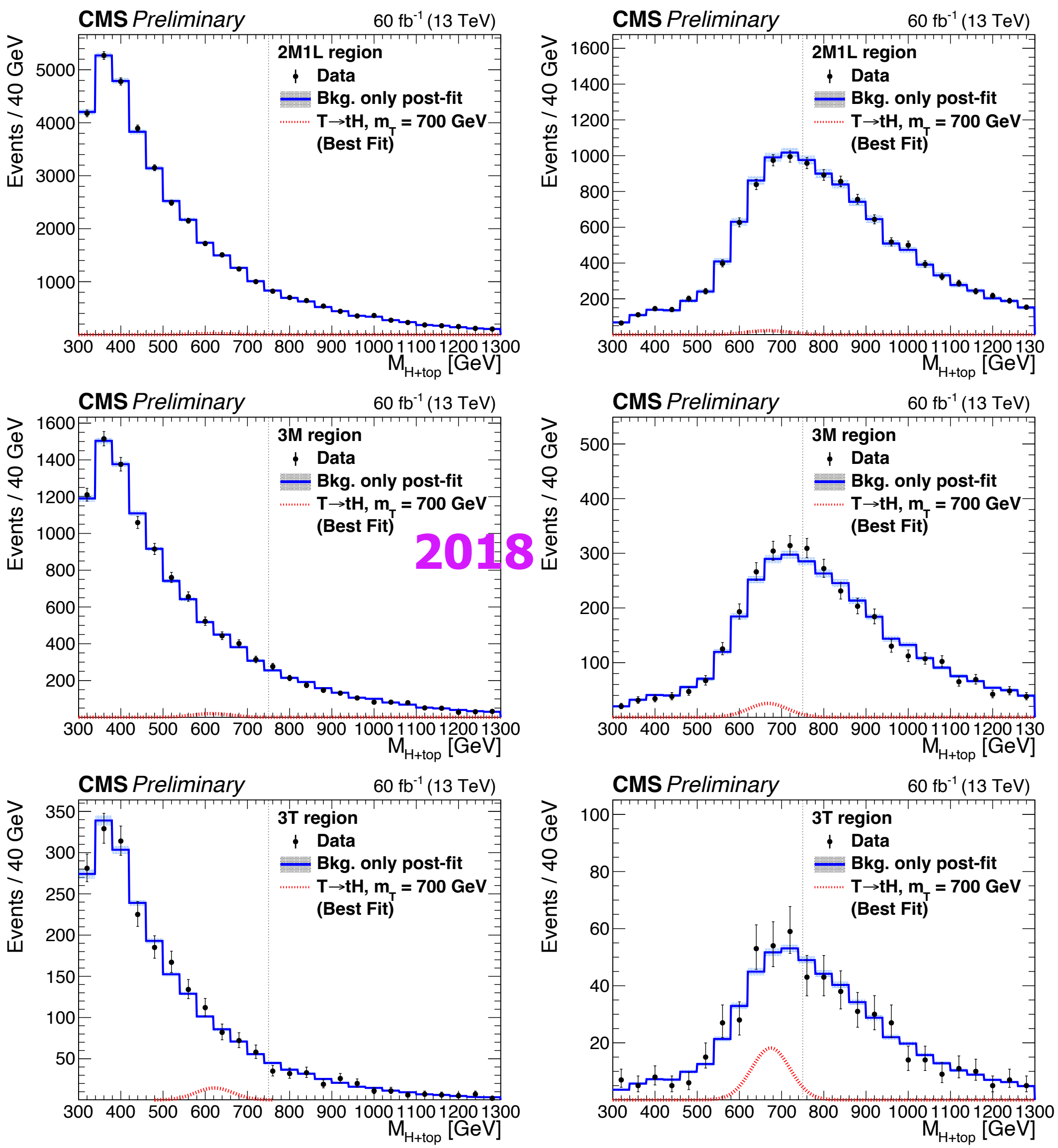
low mass selection	high mass selection	low mass selection	high mass selection

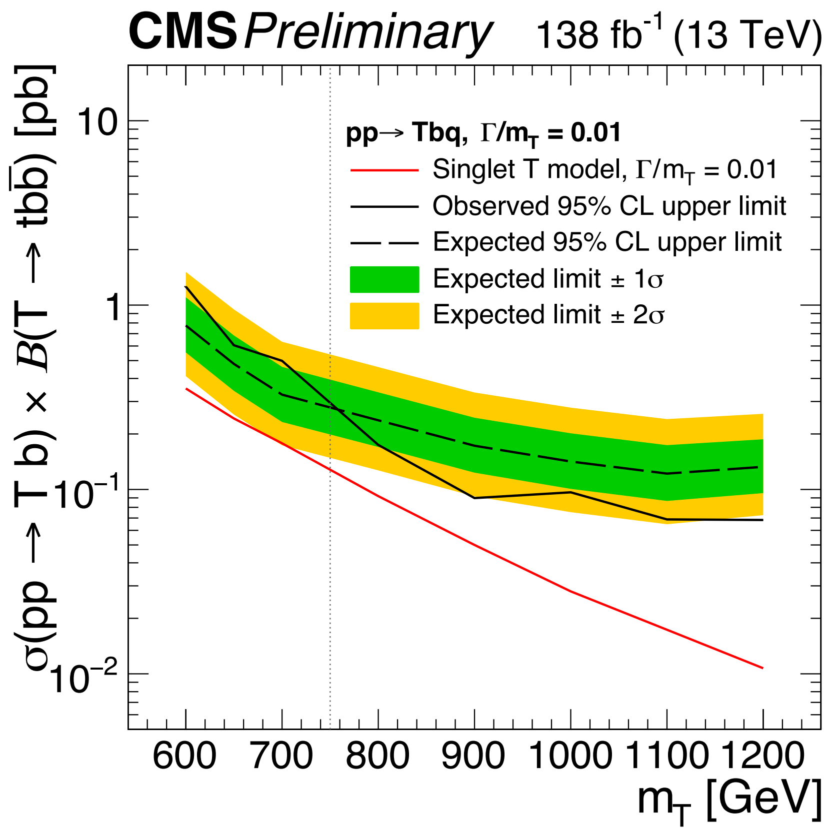


Single T→ tH/Z

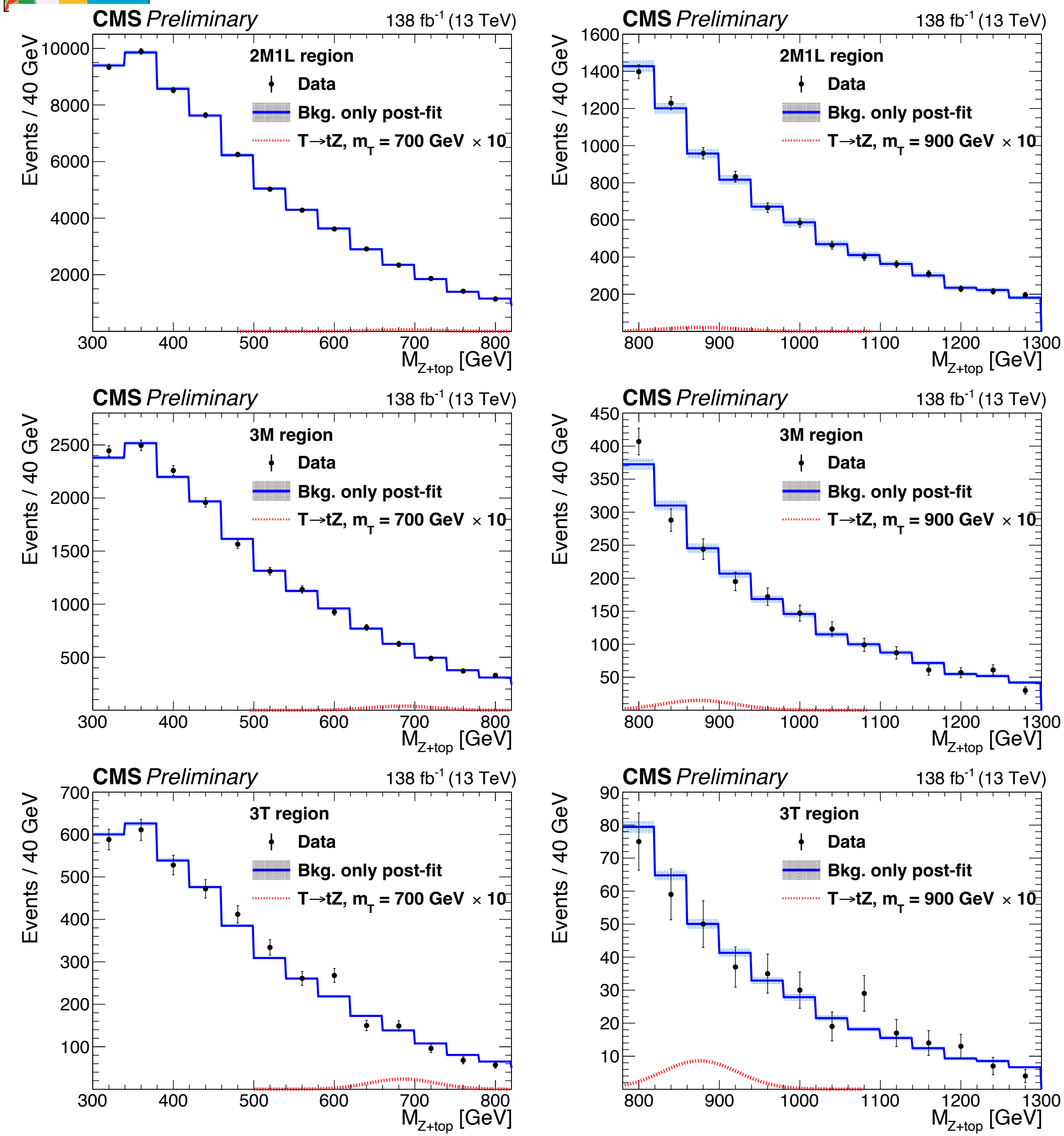
B2G-19-001

low mass selection **high mass selection**



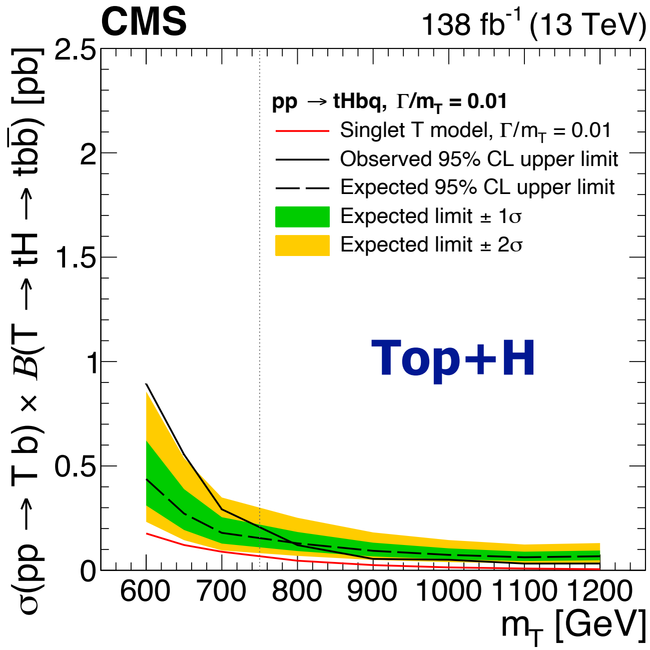


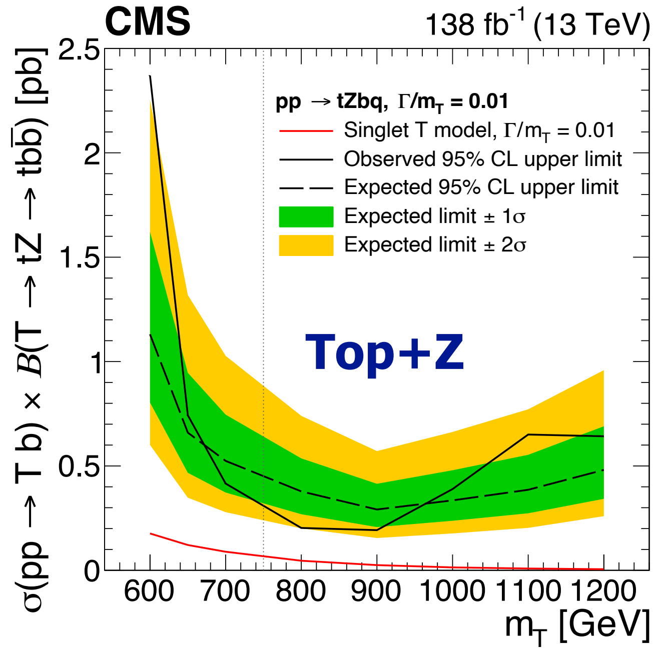
Top+Z

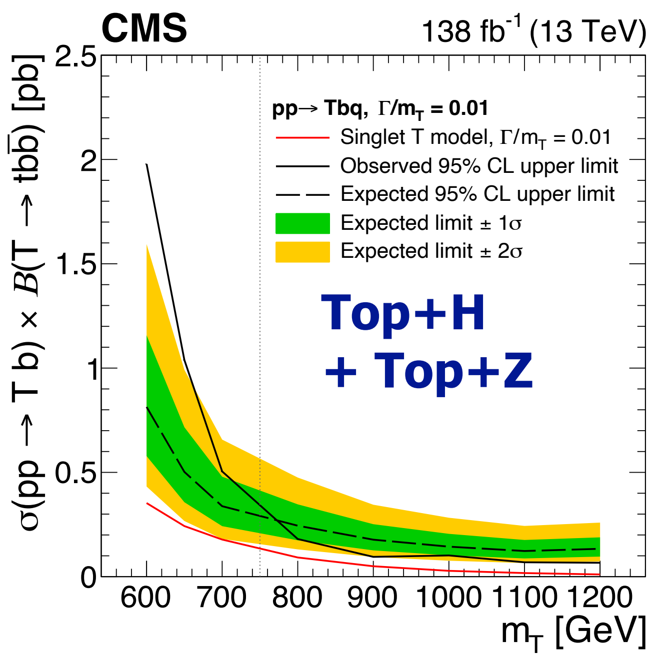


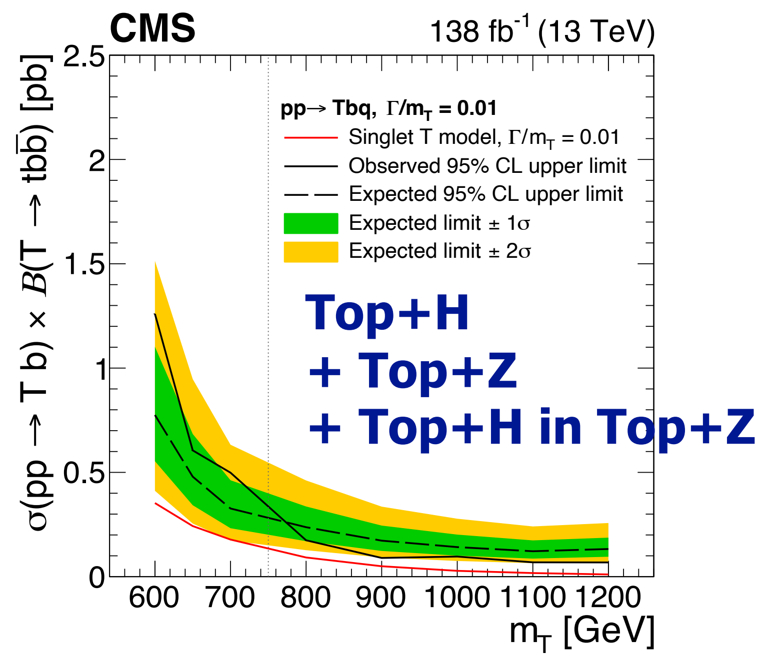
A tiny visible excess at lower mass than top+Z, but compatible with top+H linking in the top+Z channel, nevertheless combination of All top+H channel excess is 3 sigma local

Limits

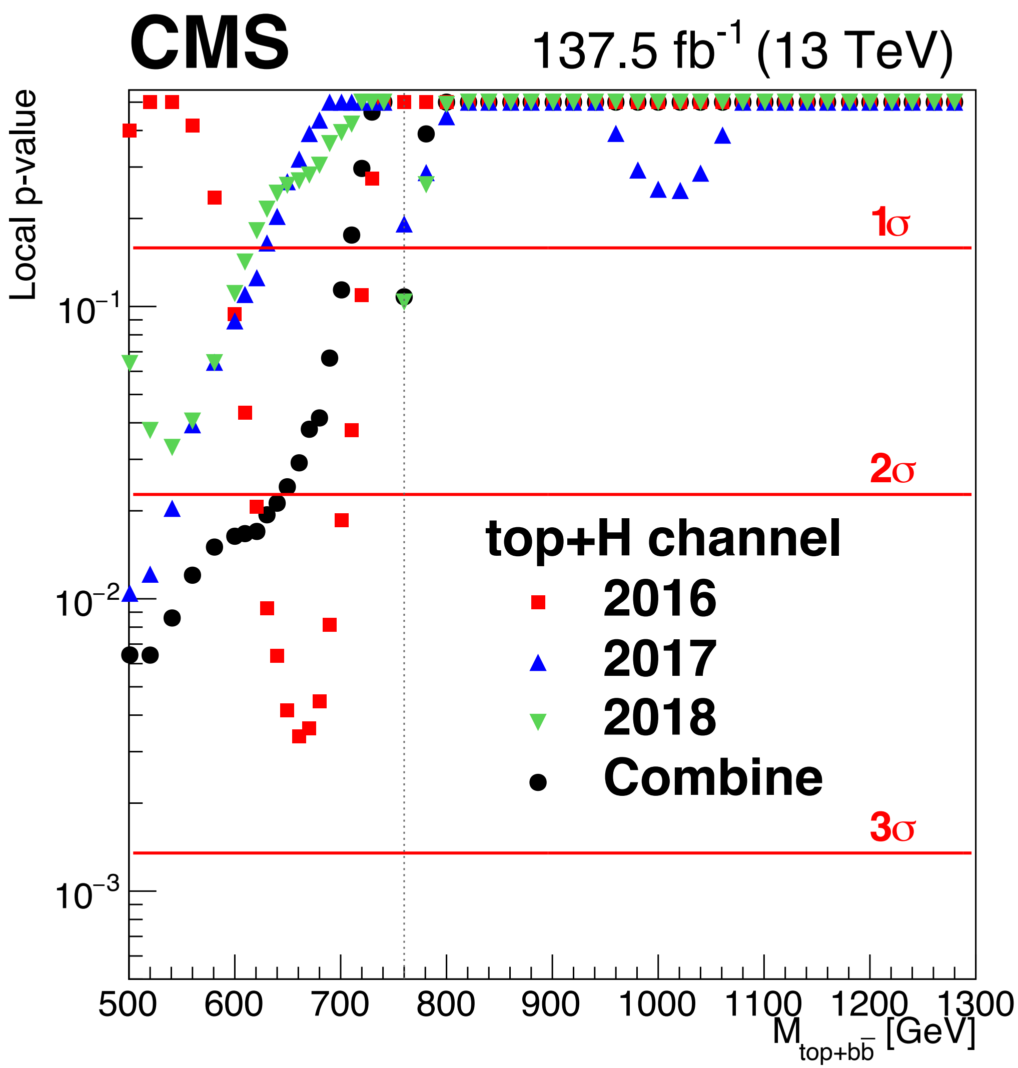

CMS
138 fb-1 (13 TeV)
pp → tHbq, Γ/mT = 0.01
Singlet T model, Γ/mT = 0.01
Observed 95% CL upper limit
Expected 95% CL upper limit
Expected limit ± 1σ
Expected limit ± 2σ
Top+H
σ(pp → T b) × B(T → tH → tbb̄) [pb]
mT [GeV]


CMS
138 fb-1 (13 TeV)
pp → tZbq, Γ/mT = 0.01
Singlet T model, Γ/mT = 0.01
Observed 95% CL upper limit
Expected 95% CL upper limit
Expected limit ± 1σ
Expected limit ± 2σ
Top+Z
σ(pp → T b) × B(T → tZ → tbb̄) [pb]
mT [GeV]


CMS
138 fb-1 (13 TeV)
pp→ Tbq, Γ/mT = 0.01
Singlet T model, Γ/mT = 0.01
Observed 95% CL upper limit
Expected 95% CL upper limit
Expected limit ± 1σ
Expected limit ± 2σ
Top+H
+ Top+Z
σ(pp → T b) × B(T → tbb̄) [pb]
mT [GeV]


CMS
138 fb-1 (13 TeV)
pp→ Tbq, Γ/mT = 0.01
Singlet T model, Γ/mT = 0.01
Observed 95% CL upper limit
Expected 95% CL upper limit
Expected limit ± 1σ
Expected limit ± 2σ
Top+H
+ Top+Z
+ Top+H in Top+Z
σ(pp → T b) × B(T → tbb̄) [pb]
mT [GeV]

P-Value



P-Value plot can be included in the paper instead of the text (commenting on the 2016 excess not observed in the other years)

Conclusion

The excess observed in the previous publication was checked:

A new low mass selection have been made in order to avoid potential statistics fluctuation in the raising edge part of the spectrum

The excess remain clearly visible in 2016 data but it is not confirmed in the other year

→ No overall excess is observed

→ Limits are set on the VLQ mode reaching less than 2 the Singlet model cross section and improving by the 3 the previous ones

Back Up

History: 2012 Data

Analysis performed over 2012 data for my habilation thesis

→ Arrived too late for publication so ONLY thesis endorsed, analysis cuts tuned at M=500 GeV mass point

- **Basic Selection:**

Trigger: QuadJet50, H_T > 500 GeV

≥ 3 b-tag Medium (CSV>0.679)

At least 6 PFLoose ID AK5 Jets within |η|<4.5

Jets pt ordered: pt > 120/100/60/60/50/20 GeV

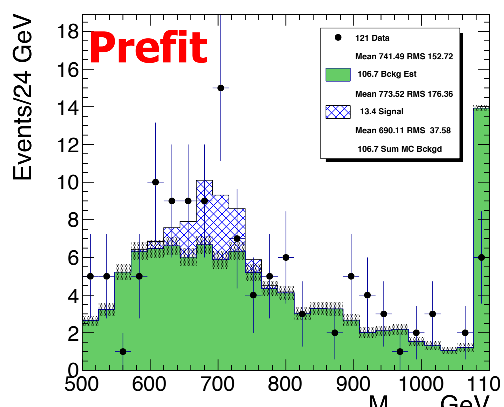
$$\chi^2 = \frac{(M_H - M_{bb})^2}{\sigma_H^2} + \frac{(M_W - M_{jj})^2}{\sigma_W^2} + \frac{(M_t - M_{bjj})^2}{\sigma_t^2}$$

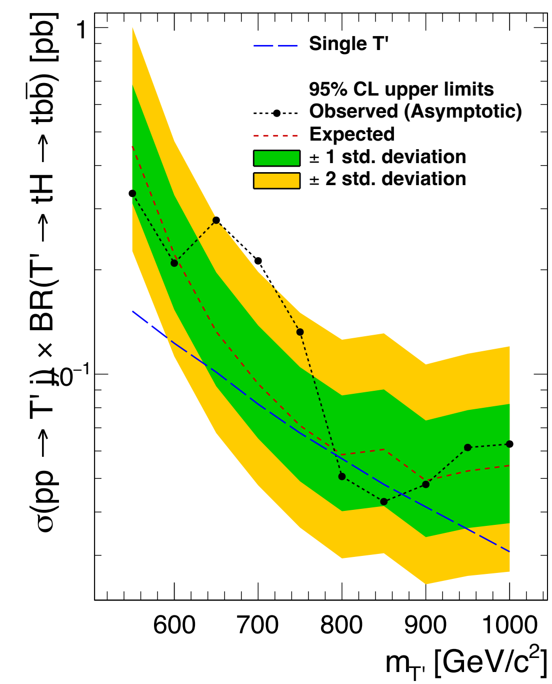
ChiSquared<30 (Not done with Higgs priority at that time)

Top_From_Higgs_Chi2->M()>250

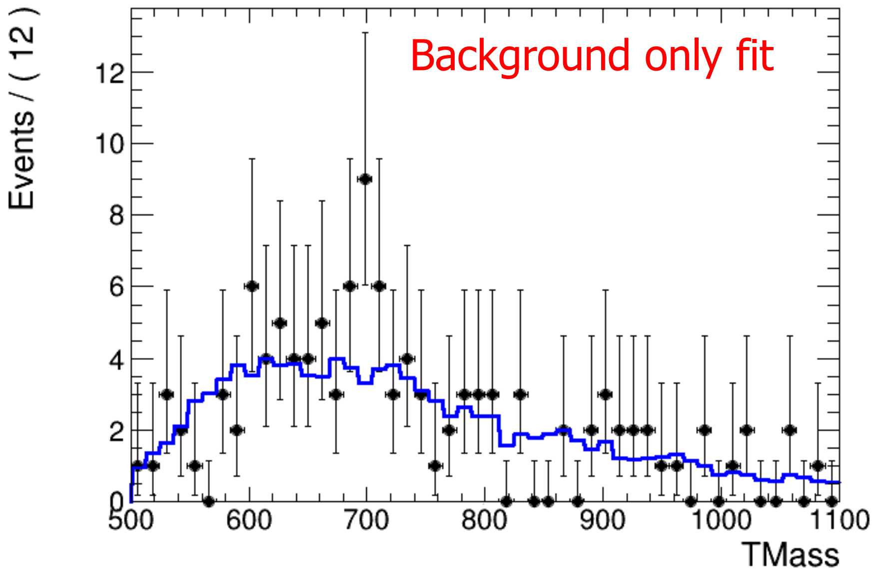
- **Full selection:**

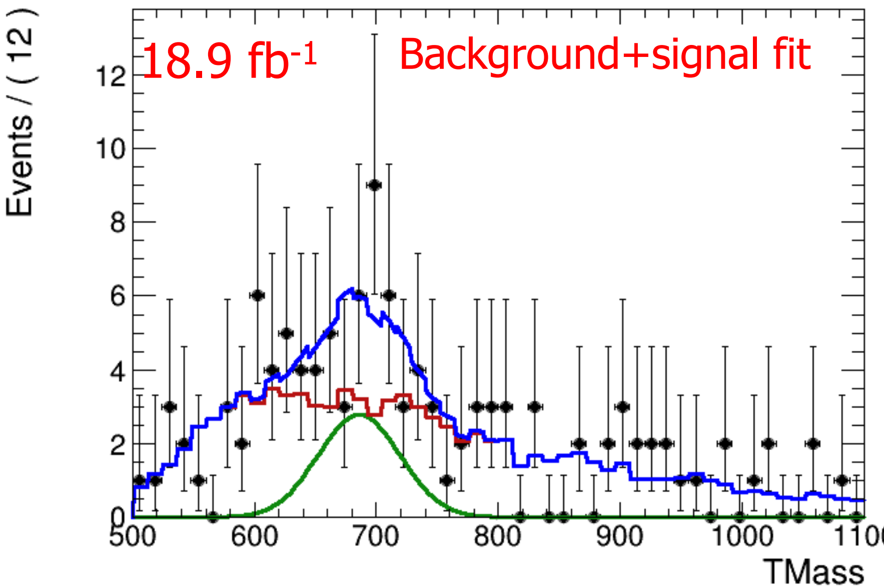
	Cuts	Signal (M=700 GeV/c^2)	Multijet	$t\bar{t}$ + single top	Diboson
Cut 0	Basic selection (number of events)	80.9 [100%]	50975.1 [100%]	11893.3 [100%]	12.6 [100%]
Cut 1	$\frac{M_{top}^{2nd}+M_W^{2nd}}{M_{Higgs}} > 5$	83.3%	63.8%	49.1%	62.7%
Cut 2	$\Delta R(b_{Top}, W) < 1.2$	54.%5	13.1%	13.8%	15.2%
Cut 3	Relative $H_T > 0.75$	32.4%	1.72%	1.90%	4.22%
Cut 4	$\chi^2_{Higgs} < 1.5$	23.4%	0.82%	0.73%	0.84%
Cut 5	$\Delta R(b_{Higgs}, b_{Higgs}) < 1.2$	23.0%	0.80%	0.68%	0.84%
Cut 6	$\mathrm{Max}(\chi^2) < 5$	18.9%	0.37%	0.52%	<0.84%
Cut 7	$\frac{M(W+H)}{M(Top+H+6^{th}Jet)} < 0.55$	16.6%	0.040%	0.365%	<0.84%





History: 2012 Data





Constraint on sigma of gaussian (GeV/c^2)	Fitted Mean (GeV/c^2)	Fitted Sigma (GeV/c^2)	r	Significance
[0, 100]	694.3±4.4	7.5±4.4	0.79±0.41	3.15
[10, 100]	694.9±5.0	10.0±5.5	0.89±0.36	3.11
[15, 100]	694.8±7.5	15.0±6.2	1.01±0.44	2.96
[20, 100]	694.3±10.0	20.0±11.6	1.13±0.49	2.85
[25, 100]	692.2±12.3	25.0±15.2	1.25±0.52	2.78
[30, 100]	688.7±14.6	30.0±47.1	1.36±0.57	2.74
[35, 100]	685.5±16.6	35.0±45.6	1.48±0.62	2.71

Fit preferred width below resolution (~5% = 35 GeV)
r is computed for cross section = 290fb (no BR)

Region enriched in ttbar

- **The analysis selection is removing ttbar events by requesting that M(Higgs + remaining leading jet) = 2nd top mass > 250 GeV**

➔ Inverse that criteria and event tighten it to [150,200] GeV

➔ Look at event after Full selection

➔ Compare Data for 3M and 2M1L B-tag WP