CMS Draft Analysis Note

2016/07/26
Head Id: 357992
Archive Id: 357906:358606M
Archive Date: 2016/07/21
Archive Tag: trunk

Combination of Searches for Top Quark Pair Resonances Using the All-Hadronic and Semileptonic Final States (2015 Dataset)

The B2G-15-002 and B2G-15-003 Teams

Abstract

We present a combination of searches for heavy resonances decaying to top quark-antiquark pairs, using the full set of proton-proton collision data collected at a center-of-mass energy of $\sqrt{s} = 13$ TeV in 2015, corresponding to an integrated luminosity of 2.6 fb^{-1}. The combination includes contributions from channels using the lepton+jets final state, as well as the all-hadronic decay mode. The combination results presented here further increase the mass exclusion limits for several models of new physics, including Z′ resonances and Randall-Sundrum Kaluza-Klein gluon production.

Semileptonic Channel	All-Hadronic Channel
e + 0 b-tag + 0 top-tag	0 subjet b-tag + $\|\Delta y\| < 1.0$
e + 1 b-tag + 0 top-tag	1 subjet b-tag + $\|\Delta y\| < 1.0$
e + 1 top-tag	2 subjet b-tag + $\|\Delta y\| < 1.0$
μ+ 0 b-tag + 0 top-tag	0 subjet b-tag + $\|\Delta y\| > 1.0$
μ+ 1 b-tag + 0 top-tag	1 subjet b-tag + $\|\Delta y\| > 1.0$
μ+ 1 top-tag	2 subjet b-tag + $\|\Delta y\| > 1.0$

Table 1: Event categories used in the combination, from each of the two channels.

1 Introduction

The CERN LHC physics production period of 2015 resulted in the first significant dataset to be collected and analyzed at a center-of-mass energy of $\sqrt{s} = 13$ TeV. Due to the increase in collision energy from 8 to 13 TeV, it is possible to probe further many new physics models which have not yet been constrained. This is due to the increase of cross sections of new physics processes relative to the standard model backgrounds.

The CMS experiment has produced two public results of searches for heavy resonances decaying to $t\bar{t}$ pairs using the dataset collected in 2015, which corresponds to 2.6 fb^{-1} of integrated luminosity. These searches both probe the same physics models, including generic Z' resonances having widths of 1, 10, and 30% of the Z' mass [1], as well as a Randall-Sundrum Kaluza-Klein gluon production model [2]. In each of these models, the heavy particle (Z'/g_{kk}) decays to a $t\bar{t}$ pair. Due to the high mass regime probed (0.5 - 4.0 TeV), the top quarks are expected to be highly boosted, requiring special reconstruction algorithms. The two searches to be combined in this result include (i) a search probing the lepton plus jets channel [3], using both non-isolated leptons and top-tagging algorithms, and (ii) an all-hadronic analysis [4] using top-tagging in conjunction with subjet b-tagging.

This analysis note will not describe in detail the individual channels, but instead will focus on the combination strategy and relevant differences from the preliminary results mentioned above. Further details on the individual analyses can be found in the documentation of the semileptonic [5] and all-hadronic [6] channels.

2 Combination Strategy

Events in the semileptonic channel fall into six independent categories – they are divided by the lepton flavor (electron / muon), as well as the b-tagging and top-tagging content. The 0 b-tag, 0 top-tag category is mainly used to constrain the W+jets background process. Similarly, the 1 b-tag, 0 top-tag category will be enriched in $t\bar{t}$ events and can also help to constrain that contribution. Finally, the 1 top-tag category is best for signal discrimination.

Events in the all-hadronic channel follow a similar categorization. Events are always required to have 2 top-tagged jets, but are divided based on how many subjet b-tags are identified. Because the dominant background in the all-hadronic analysis is non-top multijet production, the 0 and 1 subjet b-tag categories are useful to constrain this background. Additionally, events are divided based on the rapidity separation ($|\Delta y|$) between the two top-tagged jets, resulting in additional signal discrimination for high resonance mass hypotheses. The event categories entering the combination, for each channel, are detailed in Table 1.

The only background in common between the two channels is the $t\bar{t}$ background; this back-

ground is taken from simulation and treated as fully correlated between all the analysis categories. Because simulated events have been generated with a different distribution of activity due to additional interactions (pileup), these samples are reweighted to match the observed distribution of number of pileup vertices for the 2015 data-taking period. For the combination, the all-hadronic channel changed the minimum-bias cross section from 69.5 mb to 72 mb in order to be consistent with the semileptonic channel. This change has little impact on the kinematic distributions, and is within the systematic uncertainty due to the method used.

The all-hadronic channel has changed the top-tagging selection to be consistent with that of the semileptonic channel. This required increasing the p_{T} cut on the jets from 400 to 500 GeV. While this does lower the signal efficiency, the effect is most dramatic for the lower resonance mass points (< 1500 GeV), which is already in the excluded region from the individual analysis channels. The change was made in order to measure simultaneously the top-tagging efficiency scale factor between the two channels, as detailed below.

In the all-hadronic channel, the preliminary result included a top-tagging efficiency scale factor which was measured using a lepton plus jets control region. However, when combining these results, the control region used now will overlap with the lepton plus jets signal region(s). In order to avoid this circularity, the top-tagging efficiency scale factor is instead left as a free parameter during the maximum likelihood fit used for limit setting. This allows the measurement of this scale factor *in situ* during limit setting. A placeholder scale factor value of 1.0 ± 0.25 is used for the normalization of templates before the fitting procedure. The result of this scale factor measurement can be seen further in this document.

Due to the kinematic requirements of the all-hadronic channel, there are two mass points included in the semileptonic channel – 500 and 750 GeV – that are not probed with the all-hadronic category. These could be added, but would require a re-optimization of the all-hadronic channel, where it is simply not sensitive to these signal mass points, due to the jet and event kinematic cuts applied.

Finally, one needs to ensure orthogonality between all analysis channels to avoid statistical complications and double-counting of events. We ensure that the semileptonic channel is totally independent of the all-hadronic by vetoing events containing two top-tagged jets. The top-tagging selection is identical between the two channels, so this ensures the orthogonaity of analysis categories.

3 Systematics

For the combination, there are several sources of systematic uncertainties which can be correlated between the semileptonic and all-hadronic channels. The correlation is done by using the identical nuisance parameter for the uncertainty source for both of the channels during limit setting. Table 2 details the sources of systematic uncertainty for each analysis and their correlations between the two channels.

4 Signal Discrimination

Figures **??** and **??** show the distributions of $m_{\mathrm{t\bar{t}}}$ for the all-hadronic and semileptonic channels, respectively. In these plots, the post-fit distributions are used for the backgrounds, while the signal is scaled to a cross section of 1 pb. The post-fit values are determined by a maximum-likelihood fit described in the following section, by using the preferred values for the normalization and shape nuisance parameters and their uncertainties determined from this fit.

Uncertainty	Semileptonic	All-Hadronic
Z+jets σ	✓	
W+jets σ	✓	
$t\bar{t}$ σ	✓	✓
Single t σ	✓	
VV σ	✓	
t-tagging Efficiency	✓	✓
t-mistagging Efficiency	✓	
Pileup reweighting	✓	✓
Parton Distribution Functions	✓	✓
Muon ID	✓	
Muon Trigger	✓	
W+jets Q^2 Scale	✓	
$t\bar{t}$ Q^2 Scale	✓	✓
$t\bar{t}$ Parton Shower Scale	✓	✓
NTMJ Jet Kinematics		✓
NTMJ Closure Test		✓
Luminosity Measurement	✓	✓
Jet Energy Resolution	✓	✓
Jet Energy Scale	✓	✓
Electron ID	✓	
Electron Trigger	✓	
b-tagging (HF) Efficiency	✓	
b-tagging (LF) Efficiency	✓	
Subjet b-tagging Efficiency		✓

Table 2: Sources of systematic uncertainty for the combination. If an uncertainty applies to both the semileptonic and all-hadronic channels, it is treated as fully correlated between all event categories.

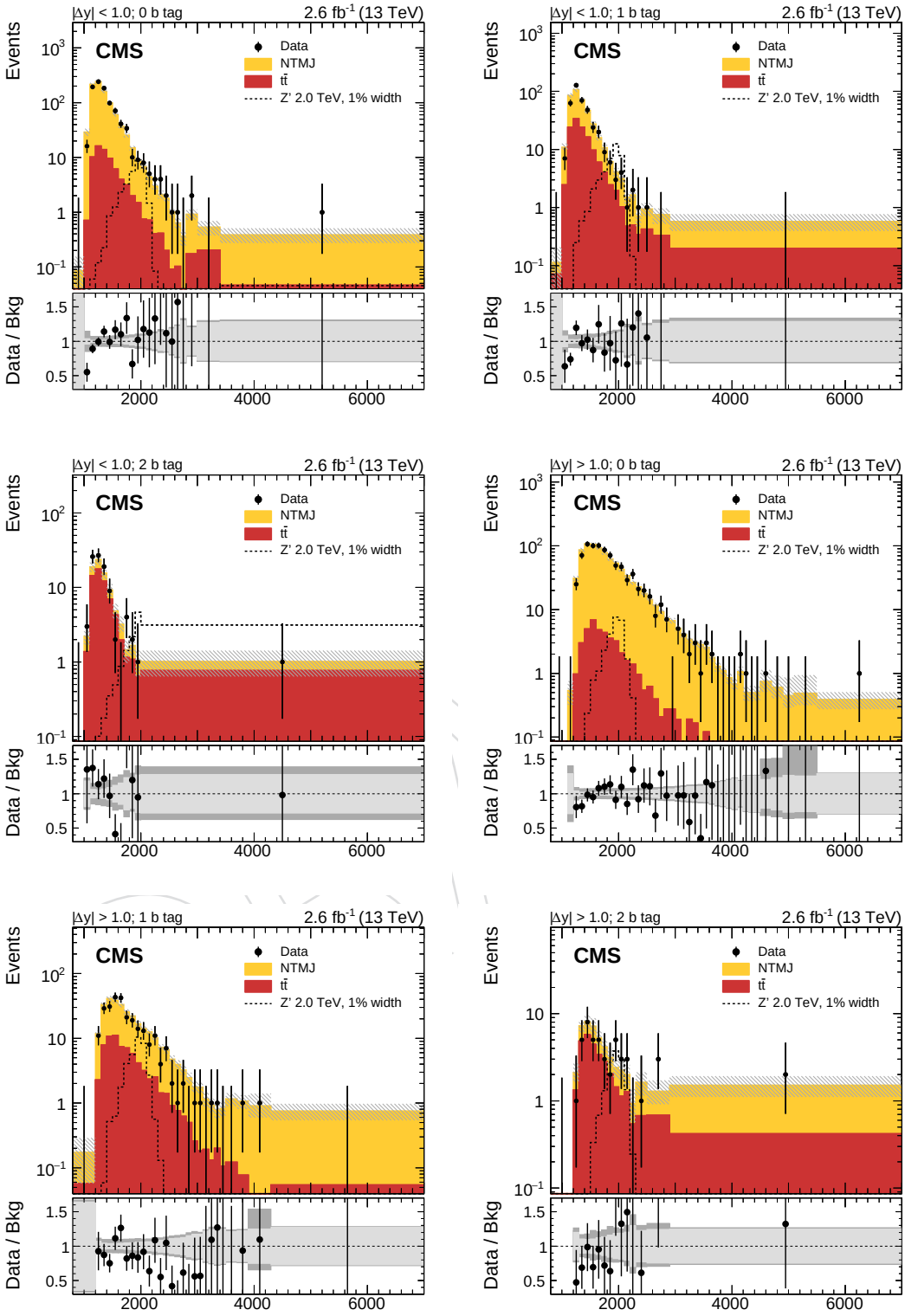

Figure 1: Post-fit distributions of $m_{\mathrm{t\bar{t}}}$, used for signal discrimination, for the six categories of the all-hadronic channel entering the combination. Example signal shapes are scaled to an arbitrary cross section of 1 pb for visualization.

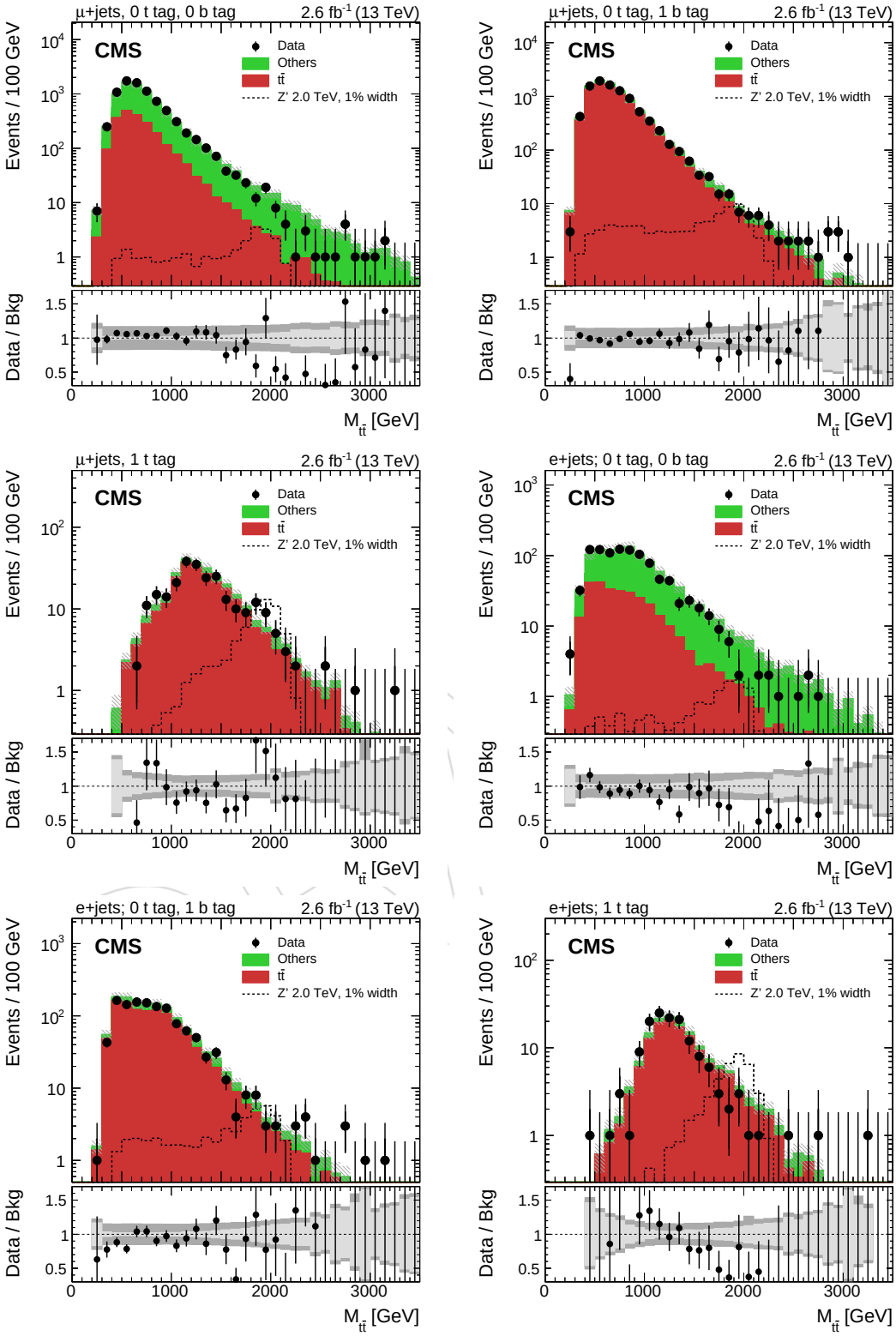

Figure 2: Distributions of $m_{\mathrm{t\bar{t}}}$, used for signal discrimination, for the six categories of the semileptonic channel entering the combination. Example signal shapes are scaled to an arbitrary cross section of 1 pb for visualization.

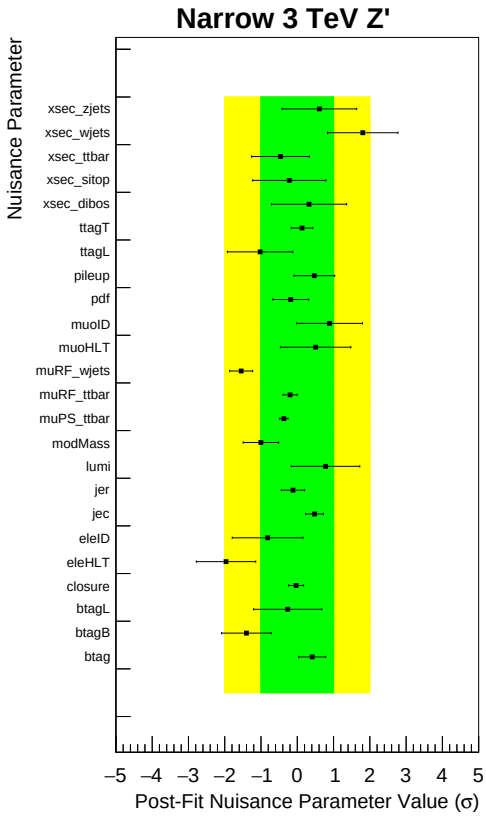

Figure 3: Post-fit values of all nuisance parameters entering the fitting procedure for the combination.

5 Statistical Analysis

We use the theta software package [7] to set limits on the production cross sections of the new physics hypotheses, using the statistical combination of the 12 event categories listed in Table 1. We use a Bayesian methodology to set limits using a likelihood composed of the product of likelihoods for each bin of events entering the combination. Statistical uncertainties due to the finite number of simulated events are included as an additional nuisance parameter through the Barlow-Beeston 'lite' method. Nuisance parameters for the systematic uncertainty contributions are implemented using lognormal prior distributions, with the exception of the signal normalization, which uses a flat prior distribution. Additionally, to measure the top-tagging efficiency scale factor *in situ* we also use a flat prior distribution for the corresponding nuisance parameter. Correlated systematic uncertainties between the all-hadronic and semileptonic channels are implemented by using identical nuisance parameters for the two channels.

Figure 3 shows the resulting post-fit values of the nuisance parameters and their uncertainties. These values are used to scale the bacgkround contributions in the post-fit $m_{\mathrm{t\bar{t}}}$ distributions shown in the previous section of this document.

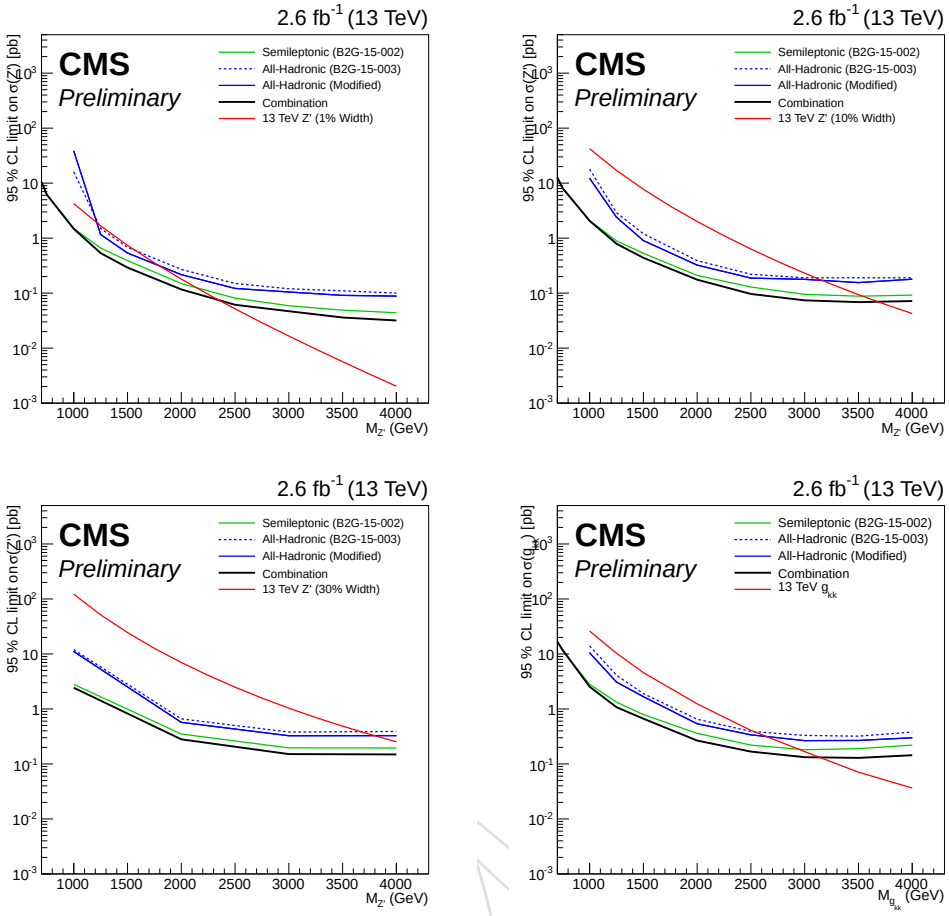

Figure 4: Comparison of expected limits from the sub-channels entering the combination, for each new physics hypothesis.

6 Results

Figure 4 shows a comparison of the expected limits for the two individual channels (approved results), along with the resulting expected limit for the combination. This is shown for each of the four physics models tested. The semileptonic channel sensitivity is better overall, but a gain in sensitivity is still obtained by including the all-hadronic channel event categories in the combination. The resulting improvement in sensitivity ranges from 100 to 300 GeV depending on the physics model.

As described above, we use the fitting procedure during limit setting to also extract a measurement of the top-tagging efficiency scale factor. The nuisance parameter is left free to float during the fit, and we use the post-fit value of the nuisance parameter to extract the value of the scale factor. Using the method outlined in the previous section, we arrive at an *in situ* determination of the top-tagging efficiency of 0.996 ± 0.065.

The full results of the combination, including the observed limits on the resonance production cross section, are shown in Figure 5. The results are tabulated numerically in Tables 4–7. The combination of the semileptonic and all-hadronic channels allows the exclusion limits to increase relative to the individual results. We exclude masses up to 4 TeV for the extra-wide (30% width) Z′ samples, and well above 3 TeV for the wide (10% width) Z′ and RS KK gluon hy-

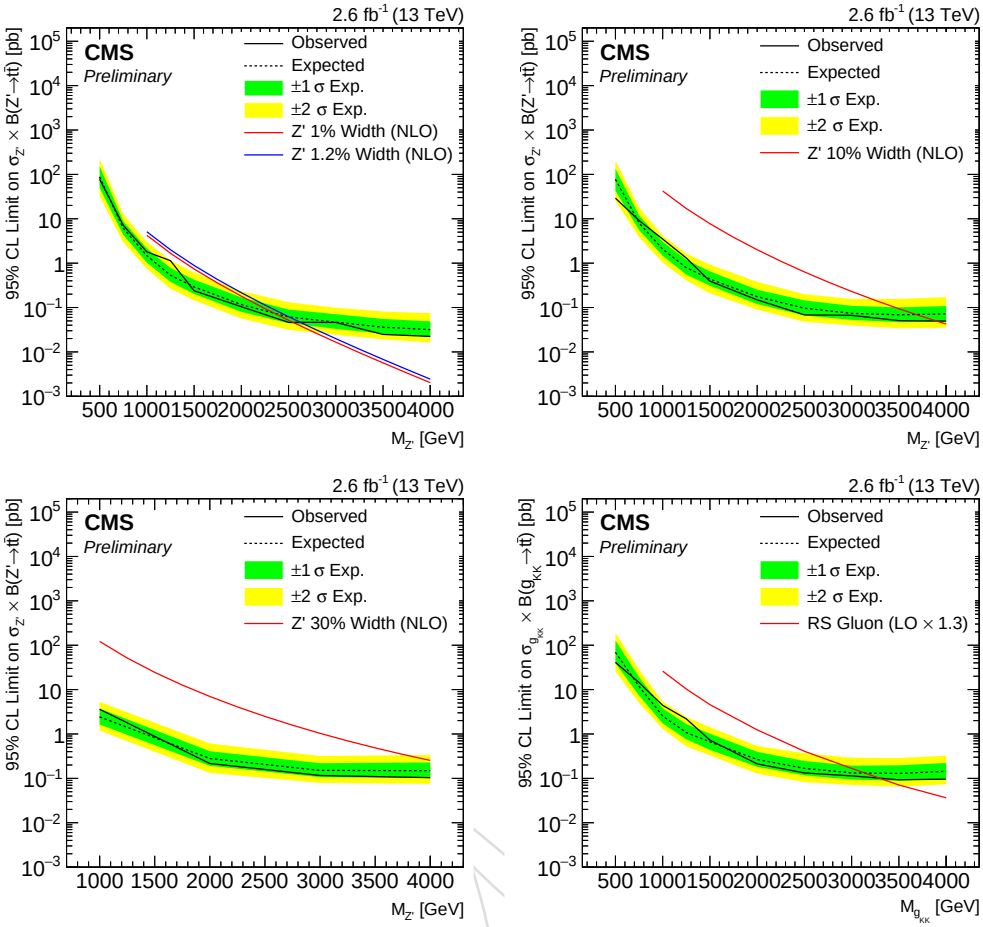

Figure 5: Observed and expected limits for the full combination of analysis results, for each new physics hypothesis.

potheses. Table 3 compares the exclusion limits of the two individual preliminary results with the results of the combination detailed in this note. Figure 6 presents the Z′ limits as a function of width instead of mass.

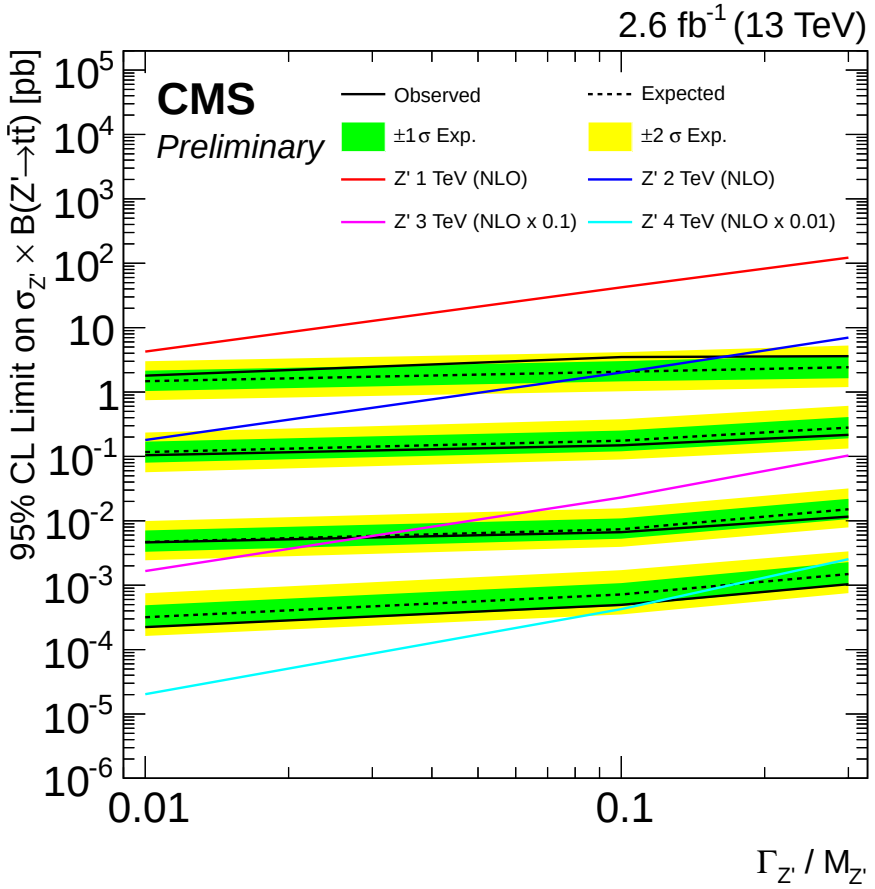

Figure 6: Expected and observed limits presented as a function of width, for various Z′ masses. Each band represents results for a different Z′ mass, going from 1 TeV on the top-most band to 4 TeV for the bottom-most band. The 3 TeV results are scaled by a factor of 0.1, and the 4 TeV results are scaled by a factor of 0.01, for visibility. The corresponding theory cross sections are shown via the colored lines.

Result	Narrow Z′		Wide Z′		Extra Wide Z′		RS KK Gluon	
	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.
Semileptonic (B2G-15-002)	0.6 – 2.1	0.6 – 2.3	0.5 – 3.5	0.5 – 3.4	0.5 – 4.0	0.5 – 4.0	0.5 – 2.9	0.5 – 2.9
All-Hadronic (B2G-15-003)	1.2 – 1.6	1.4 – 1.6	1.0 – 3.1	1.0 – 3.3	1.0 – 3.7	1.0 – 3.8	1.0 – 2.5	1.0 – 2.4
Combination (B2G-16-015)	0.6 – 2.4	0.6 – 2.5	0.5 – 3.7	0.5 – 3.9	0.5 – 4.0	0.5 – 4.0	0.5 – 3.1	0.5 – 3.3

Table 3: Comparison of mass exclusion results for the two individual preliminary results, compared with the combination presented here. The results shown for the individual channels are the same as for the preliminary approved results, not taking into account the changes made to the all-hadronic channel for this combination. The combination results include all modifications.

Mass (GeV)	Observed 95% CL Limit (pb)	Expected 95% CL Limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
500	**77.7**	32.1	50	**88.2**	153	229
750	**7.14**	2.93	4.33	**6.14**	8.81	12.7
1e+03	**1.8**	0.746	1.04	**1.47**	2.15	3.01
1.25e+03	**1.14**	0.264	0.377	**0.534**	0.778	1.16
1.5e+03	**0.239**	0.145	0.202	**0.291**	0.425	0.617
2e+03	**0.104**	0.0568	0.08	**0.117**	0.17	0.235
2.5e+03	**0.0464**	0.0314	0.0443	**0.0614**	0.09	0.132
3e+03	**0.0462**	0.0244	0.033	**0.0469**	0.0708	0.0992
3.5e+03	**0.0248**	0.0192	0.0257	**0.036**	0.0554	0.0813
4e+03	**0.0224**	0.0163	0.022	**0.0318**	0.0488	0.0749

Table 4: Expected and observed cross section limit results for the narrow (1% width) Z' resonance hypothesis.

Mass (GeV)	Observed 95% CL Limit (pb)	Expected 95% CL Limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
500	**29.1**	25.3	42.8	**77.4**	134	201
750	**9.12**	3.86	5.57	**8.06**	11.9	18
1e+03	**3.49**	1.03	1.46	**2.06**	3.02	4.15
1.25e+03	**1.31**	0.407	0.551	**0.789**	1.2	1.64
1.5e+03	**0.393**	0.209	0.31	**0.439**	0.651	0.929
2e+03	**0.149**	0.0896	0.12	**0.176**	0.252	0.378
2.5e+03	**0.0684**	0.0478	0.0663	**0.0965**	0.145	0.2
3e+03	**0.0667**	0.0394	0.0527	**0.0739**	0.109	0.156
3.5e+03	**0.0507**	0.0341	0.0467	**0.0686**	0.101	0.156
4e+03	**0.0495**	0.0351	0.0476	**0.0718**	0.108	0.171

Table 5: Expected and observed cross section limit results for the wide (10% width) Z' resonance hypothesis.

Mass (GeV)	Observed 95% CL Limit (pb)	Expected 95% CL Limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
1e+03	**3.61**	1.19	1.64	**2.43**	3.73	5.27
2e+03	**0.216**	0.133	0.191	**0.28**	0.409	0.611
3e+03	**0.116**	0.079	0.106	**0.151**	0.22	0.318
4e+03	**0.104**	0.0754	0.102	**0.149**	0.228	0.335

Table 6: Expected and observed cross section limit results for the extra-wide (30% width) Z' resonance hypothesis.

Mass (GeV)	Observed 95% CL Limit (pb)	Expected 95% CL Limits (pb)				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
500	**41**	25.6	39.5	**69.3**	128	190
750	**14.4**	5	7.34	**11.7**	19.1	28.5
1e+03	**4.41**	1.3	1.73	**2.54**	3.77	5.32
1.25e+03	**2.18**	0.527	0.757	**1.08**	1.67	2.42
1.5e+03	**0.727**	0.329	0.438	**0.665**	1.02	1.42
2e+03	**0.212**	0.127	0.182	**0.266**	0.396	0.536
2.5e+03	**0.132**	0.0824	0.117	**0.168**	0.247	0.368
3e+03	**0.114**	0.0709	0.094	**0.133**	0.193	0.291
3.5e+03	**0.093**	0.0651	0.0883	**0.129**	0.197	0.288
4e+03	**0.096**	0.0732	0.0989	**0.144**	0.222	0.323

Table 7: Expected and observed cross section limit results for the RS KK gluon hypothesis.

7 Summary

We have produced a combination of the two previously-approved searches for $t\bar{t}$ resonances using 2015 data collected at a center-of-mass energy $\sqrt{s} = 13$ TeV. While the sensitivity is slightly higher for the semileptonic channel, the combination of the all-hadronic and semileptonic channels significantly increases the mass exclusion limits relative to previously published results. This represents the first combination of results using the LHC Run 2 dataset.

References

[1] R. M. Harris, C. T. Hill, and S. J. Parke, "Cross-section for topcolor Z-prime(t) decaying to t anti-t: Version 2.6", arXiv:hep-ph/9911288.

[2] K. Agashe et al., "CERN LHC signals from warped extra dimensions", *Phys. Rev. D* **77** (2008) 015003, doi:10.1103/PhysRevD.77.015003.

[3] CMS Collaboration, "Search for $t\bar{t}$ resonances in boosted semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV", *CMS Physics Analysis Summary* **CMS-PAS-B2G-15-002** (2016).

[4] CMS Collaboration, "Search for top quark-antiquark resonances in the all-hadronic final state at $\sqrt{s} = 13$ TeV", *CMS Physics Analysis Summary* **CMS-PAS-B2G-15-003** (2016).

[5] B2G-15-002 Authors Collaboration, "Search for $t\bar{t}$ resonances in boosted semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV", *CMS AN-2015/107* (2016).

[6] B2G-15-003 Authors Collaboration, "Search for top quark-antiquark resonances in the all-hadronic final state at $\sqrt{s} = 13$ TeV", *CMS AN-2015/122* (2016).

[7] T. Muller, J. Ott, and J. Wagner-Kuhr, "theta - a framework for template-based modeling and inference", *CMS Internal Note* **IN-2010/017** (2010).