Single VLQ Search: T’→ top+H with H→ $\gamma\gamma$

Stéphanie Beauceron
Institut de Physique des 2 Infinis
Lyon

ICHEP Release: CMS PAS B2G-21-007

Outline:
- Motivations
- VLQ production and decays
- Analysis Strategy
- Results
- Conclusion

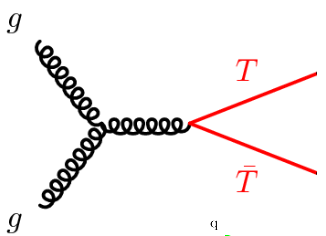
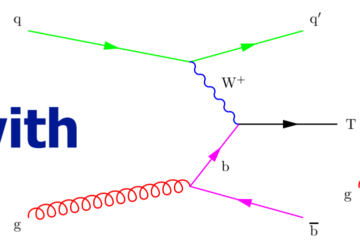
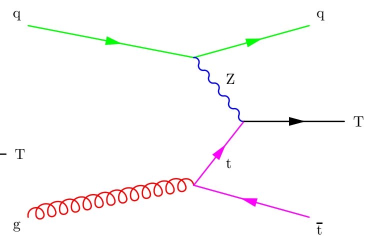
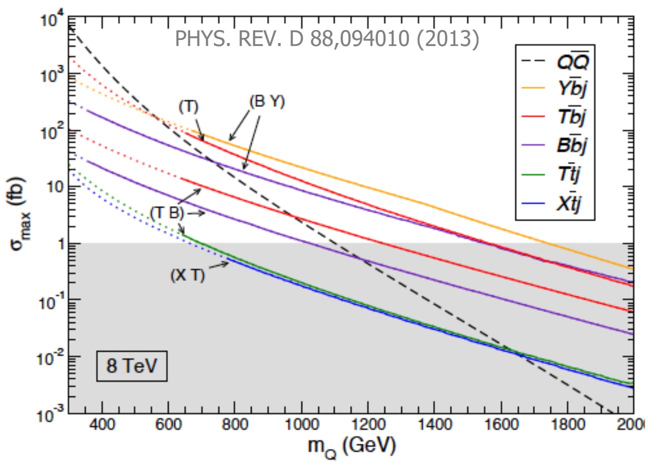
Motivation

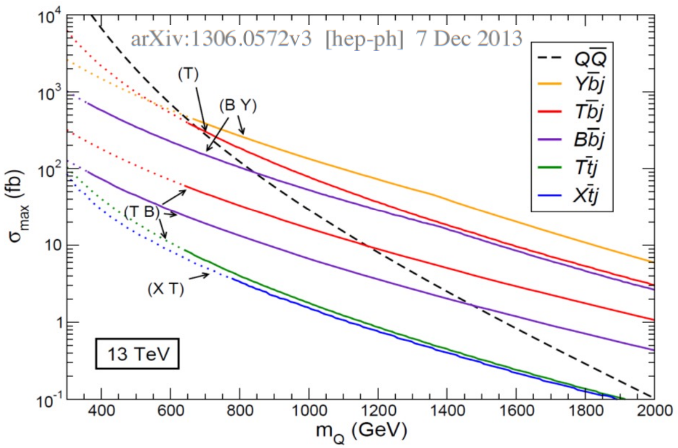
After the discovery of Higgs boson, the Standard Model (SM) is complete as a low-energy effective theory describing all known fundamental particles and their interactions
→ However, the origin of Higgs mass stability at electroweak scale is still mystery

➔ Introduction of vector like quark T′ and B′ provides a feasible solution, having electric charges of +2e/3 and −1e/3 and coupling to 3rd generation is considered here

Production

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

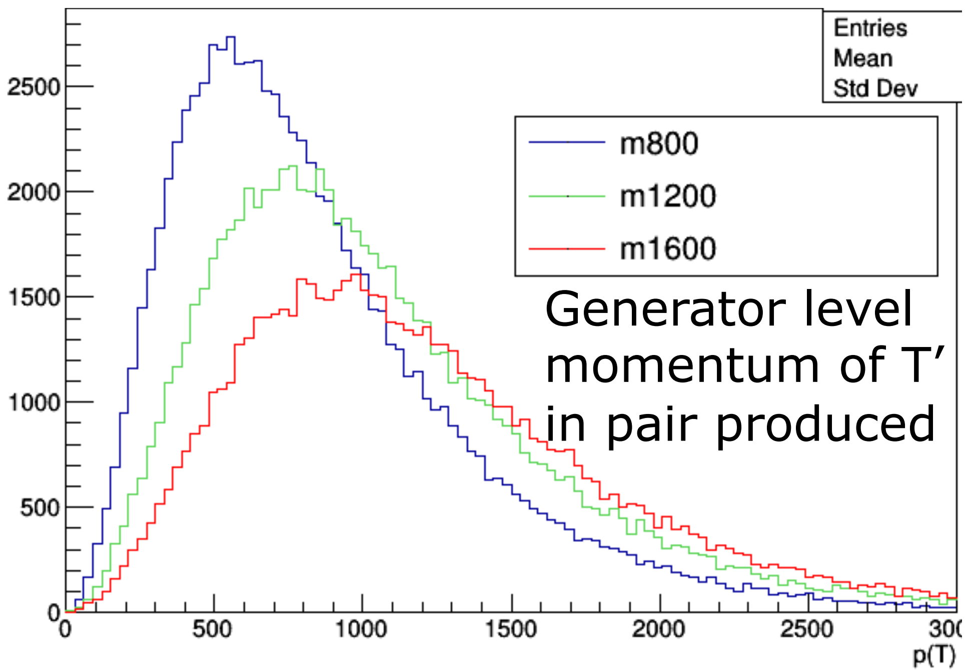


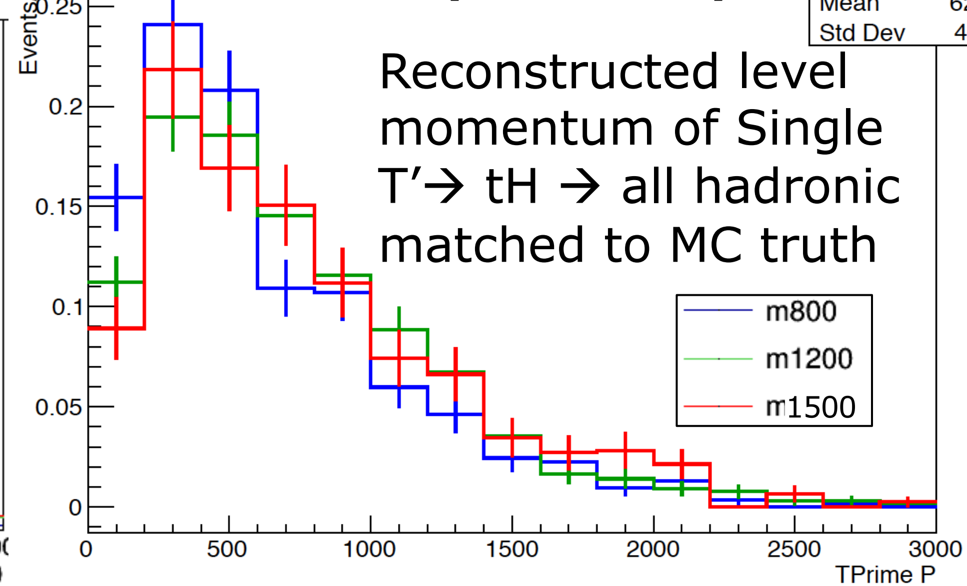
Plots above are for given benchmarks couplings but still giving an idea of what is happening…

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

Pair Produced=Momentum

Illustrative plots for T'→ tH (all hadronic)





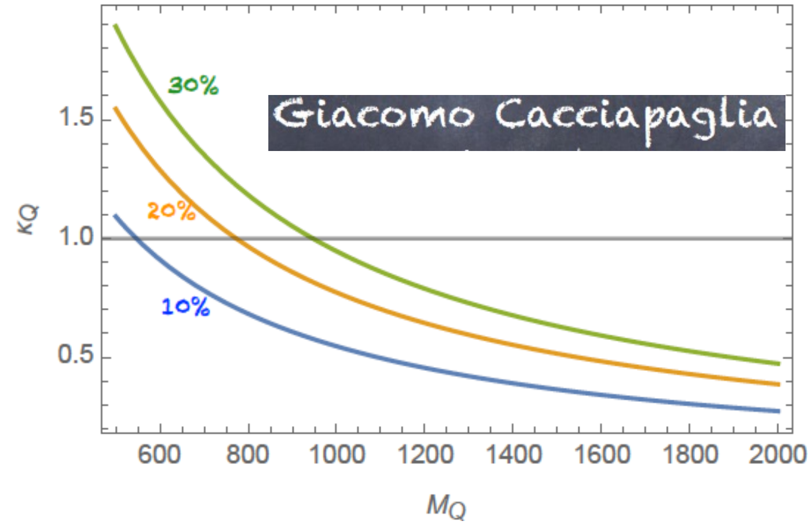
In pair production, the momentum of the produced VLQ is ~M/2, so it is increasing with mass while cross section is going down

Coming from threshold artifact linked from spin ½ (low cross section but more energy available for production)

In single production mode, the momentum does not change much with mass.

Question of Width?

Single VLQ only EW contributions and sensitive to both the VLQ mass and its mixing parameters
→Mixing parameters entering the width of VLQ
→Model dependent



Currently all pair analysis only doing narrow width while acceptance/analysis selection could be not optimal for large width (as 30%)

→ Single VLQ and pair VLQ search are complementary
→ New theory development: for width >10%, interference with SM top+H final state → Not yet considered here ([PLB Volume 793 (2019)](#))

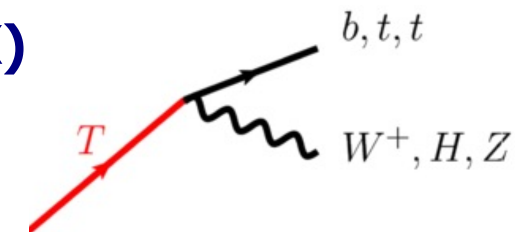
VLQ Decay

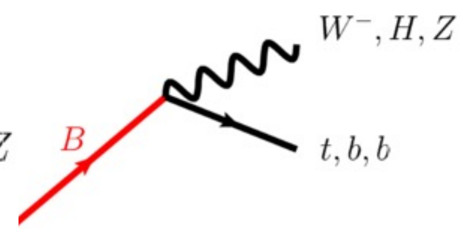
Heavy Vector like Quark

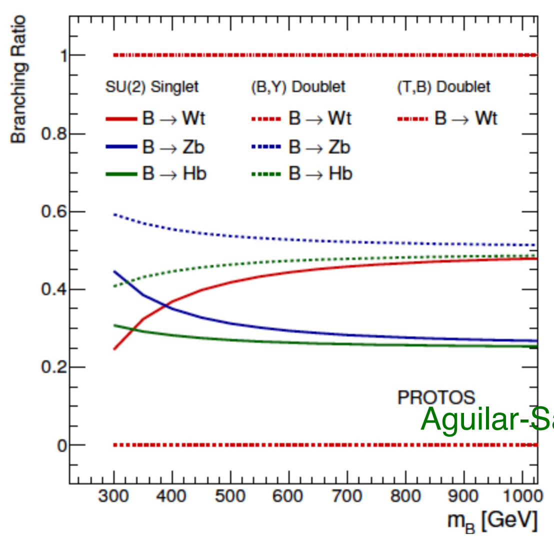
	$T \quad B$	$\binom{X}{T}\ \binom{T}{B}\ \binom{B}{Y}$	$\begin{pmatrix}X\\T\\B\end{pmatrix}\ \begin{pmatrix}T\\B\\Y\end{pmatrix}$
$SU(2)_L$ multiplet	1	2	3
Charge	2/3 -1/3	$\binom{5/3}{2/3}\ \binom{2/3}{-1/3}\ \binom{-1/3}{-4/3}$	$\begin{pmatrix}5/3\\2/3\\-1/3\end{pmatrix}\ \begin{pmatrix}2/3\\-1/3\\-4/3\end{pmatrix}$

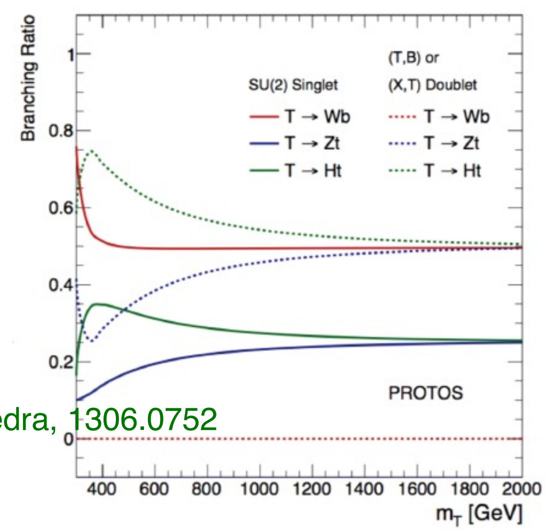
VLQ	W-decay	Z-decay	h-decay
T	Wb	Zt	ht
B	Wt	Zb	hb
$T_{5/3}$	Wt	-	-
$Y_{-4/3}$	Wb	-	-

(equivalent for Y/X)









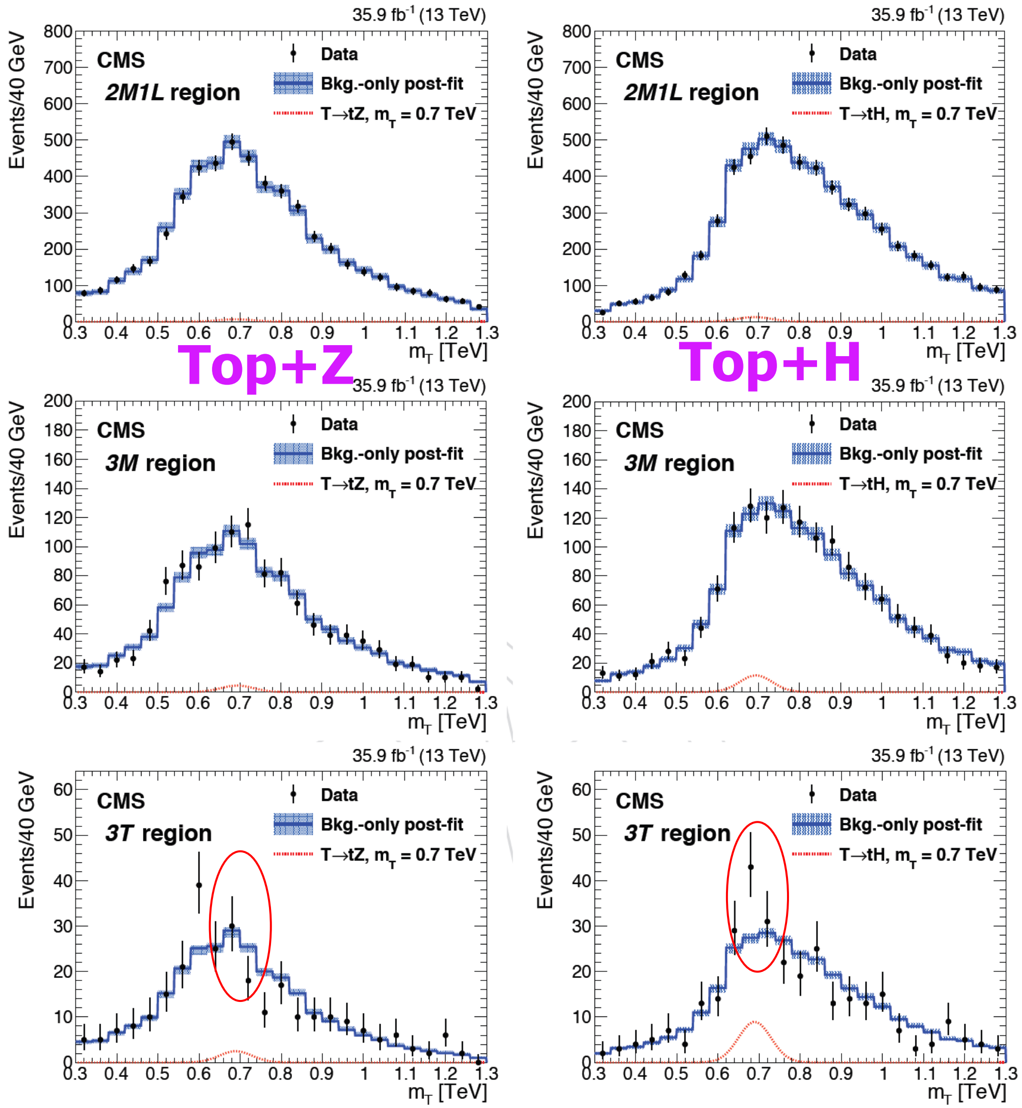
Aguilar-Saavedra, 1306.0752

no mixing between $B^{-1/3}$ and b for $(T,B)_R$ doublet in these plots ($\theta_R^d = 0$)

Top+H

[JHEP 01 (2020) 036](https://doi.org/10.1007/JHEP01(2020)036)

Post-Fit background only (low mass)



Interest in top+H as an excess observed in all hadronic final state but mainly in top+H and not really in top+Z… Theory interpretation in top+a with a→ bb or gluon-gluon (dominant) [JHEP 06 (2018) 065]

Work on other decay channel from Higgs

Semi-leptonic analysis on going as full run2 analysis for all hadronic decay

Overview of Analysis Strategy

Use all tools developed for H→γγ analysis are used
Analysis is performed via Boosted Decision Tree

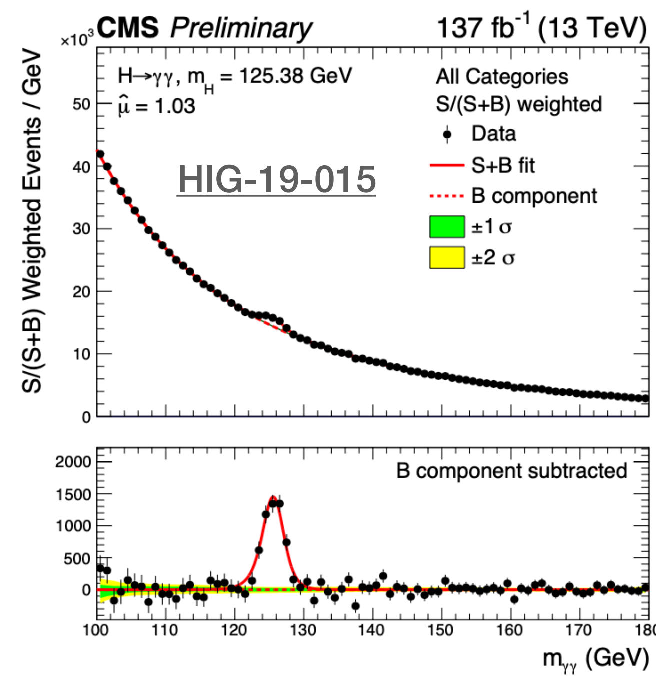
T'→tH (H→γγ)

Hadronic (t→bqq̄) | Leptonic (t→blν̄)

	Hadronic (t→bqq̄)	Leptonic (t→blν̄)
Top Reconstruction	Minimum χ2 Method	Quadratic Equation
MVA Training* / SR Optimization	BDT (NRB) BDT (SMH)	BDT (SMH)
Modeling / Limit extraction	Model Signal & Background ($m_{\gamma\gamma}$ distribution)	Model Signal & Background ($m_{\gamma\gamma}$ distribution)

Upper limit

*Separate trainings are performed for three T' mass categories
[600, 625, 650, 675, 700] [800, 900, 1000] [1100, 1200]



Analysis uses H→γγ as a probe to tag T'

Define signal window: $m_{\gamma\gamma} \in [115, 135]$ GeV

Channels Selection

First selection to have channel orthogonal to each other, train BDT after the selection:
Trigger: Diphoton
== 2 photons: p_T(1st)> $M_{\gamma\gamma}/3$, p_T(2nd) > $M_{\gamma\gamma}/4$
Jets: p_T>25 GeV and $|\eta|<4.5$
Leptons: p_T>10 GeV and $|\eta|<2.4$

Leptonic Preselection

- $N_{leptons} \geq 1$
- $N_{jets} \geq 1$, $N_{bjets\ (loose)} \geq 1$
- photon ID MVA ≥ -0.7
- $m_{\gamma\gamma} \in$ [100, 180] GeV

Hadronic Preselection

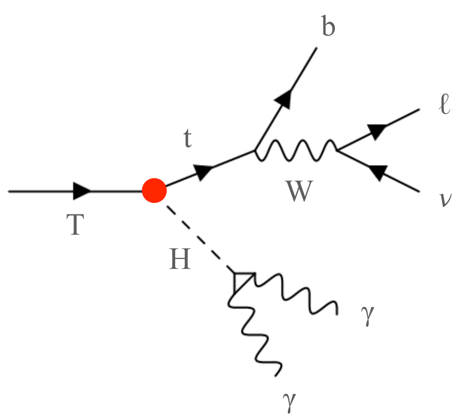
- $N_{leptons} = 0$
- $N_{jets} \geq 3$, $N_{bjets\ (loose)} \geq 1$
- photon ID MVA ≥ -0.7
- $m_{\gamma\gamma} \in$ [100, 180] GeV

Top Reconstruction

Depending on the channel:

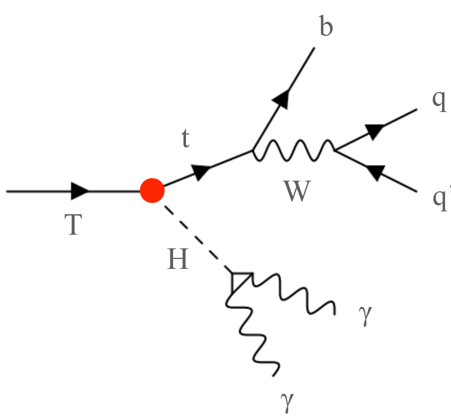
Leptonic channel

- Task: evaluate Pz component of neutrino
- Method: quadratic equation
- top candidate = blν



Hadronic channel

- Task: identify jets coming from top decay
- Method: minimum χ2 method
- top candidate = bjj



Tprime mass = top+H mass

Background Rejection

- Multivariate analysis technique
  - Distinguish VLQ signal from background events
  - Gradient boosted decision trees (BDT)
  - Separate trainings are performed for three T' mass categories [600, 625, 650, 675, 700] [800, 900, 1000] [1100, 1200]
- Leptonic channel
  - One BDT is trained for each T' mass category
  - Signal: VLQ
  - Background: $t\bar{t}H$, ggH, VH, VBF, tHq
- Hadronic channel
  - Two BDTs are trained for each T' mass category
  - Signal: VLQ
  - Non-resonant background (NRB): $\gamma\gamma$+jets, data-driven QCD, $t\bar{t}\gamma\gamma$, $t\bar{t}\gamma$+jets, $t\gamma$+jets, $t\bar{t}$+jets, V+γ
  - SM Higgs background (SMH): $t\bar{t}H$, ggH, VH, VBF, tHq

Training configuration
- Algorithm: Gradient BDT
- Decision trees: 1000
- Tree depth: 2
- Training samples: 50%
- Testing samples: 50%

In all hadronic channel the QCD is derived from data via inversion of MVA cut

Input Variables to BDTs

- Mainly kinematics, ID MVA, and b-tagging scores
  - Expect that there exist some difference between VLQ signal and SM background
  - Prevent BDTs from learning information of $m_{\gamma\gamma}$ and T' mass

Leptonic Channel

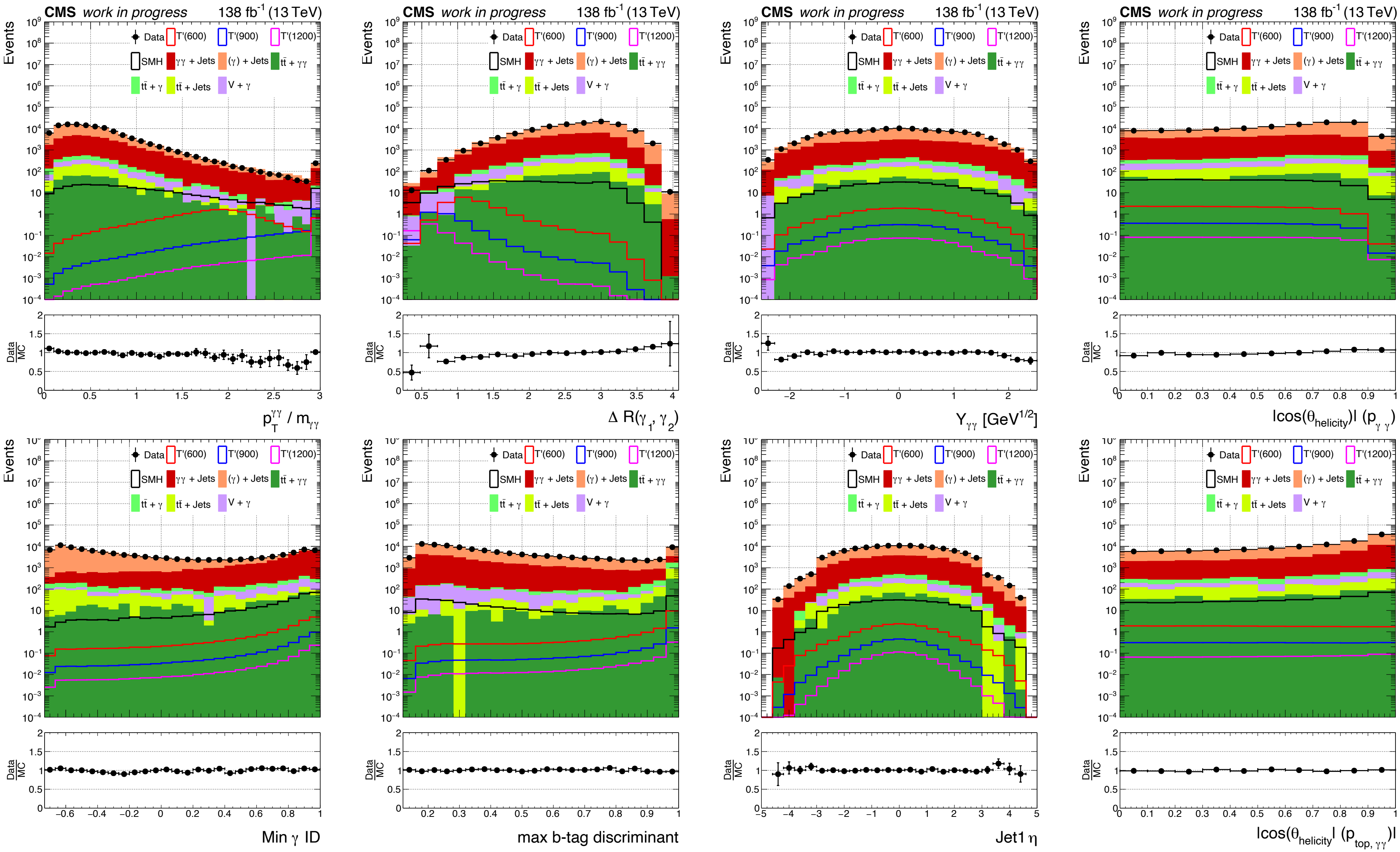
	Objects	Variables		
PHOTON RELATED	leading photon	pT/mγγ	pixel seed veto	eta
	subleading photon	pT/mγγ	pixel seed veto	eta
	photon relevant	γ1 IDMVA	γ2 IDMVA	
JETS / LEPTON RELATED	lepton relevant	pT	eta	charge
	jet multiplicity	n_jets	n_bjets	n_central_jets
	leading jet 1	pT	eta	
	leading jet 2	pT	eta	
	leading b-jet	pT	eta	b-tag score
	"forward" jet	pT	b-tag score	
OTHERS	open angle	dR(γ1, forward jet)	dR(γ2, forward jet)	dR(b-jet, forward jet)
		dR(tH, forward jet)	dR(lepton, forward jet)	dR(lepton, b-jet jet)
	T' quark & MET	(pT(blν)+pT(γγ)) / all pT	MET	

Note: "forward jet" stands for the jet with the highest |η| among the reconstructed jets

Hadronic Channel

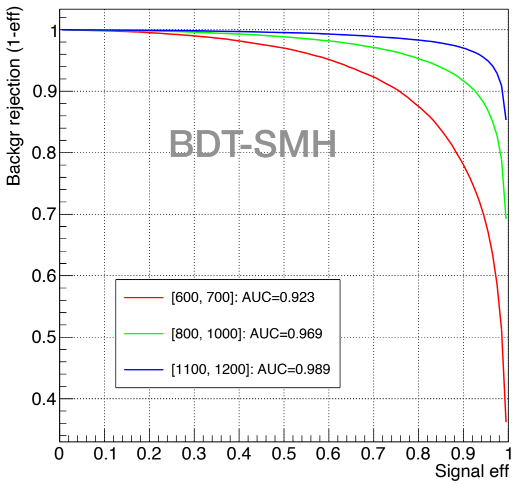
	Objects	Variables		
PHOTON RELATED	leading photon	pT/mγγ	pixel seed veto	eta
	subleading photon	pT/mγγ	pixel seed veto	eta
	photon relevant	max IDMVA	min IDMVA	dR(photon, photon)
	di-photon	pT/mγγ	cos ϕ	| cos (helicity angle) |
		Rapidity		
JETS RELATED	jet relevant	n_jets	n_bjets	
		max b-tag score	min b-tag score	HT
	leading jet 1	pT / Mbjj	eta	b-tag score
	leading jet 2	pT / Mbjj	eta	b-tag score
	leading jet 3	pT / Mbjj	eta	b-tag score
	leading jet 4	pT / Mbjj	eta	b-tag score
	b-jet	pT / Mbjj	eta	b-tag score
	w-jet 1	pT / Mjj	eta	b-tag score
	w-jet 2	pT / Mjj	eta	b-tag score
	W boson	pT / Mjj	eta	Mjj
		dR(W boson, b-jet)	dR(w-jet1, w-jet2)	Minimum χ2 value
	top quark	pT / Mbjj	eta	Mbjj
OTHERS	T' quark	pT / Mbjjγγ	eta	(pT(bjj) + pT(γγ)) / HT
		| cos (helicity angle) |		
	MET	MET		

Before training a few variables...

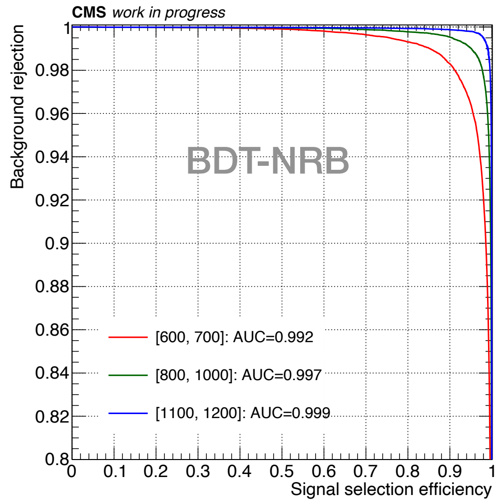


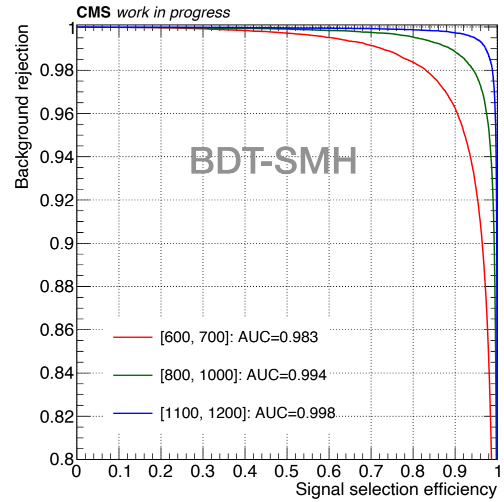
Performances

Leptonic Channel



Hadronic Channel





Performance Metrics

$M_{T'}(GeV)$	MVAs	Metric (%)			
		AUC	Signal efficiency at fixed background efficiency		
			$\epsilon_{\rm sig}$ ($\epsilon_{\rm bkg}$ = 1%)	$\epsilon_{\rm sig}$ ($\epsilon_{\rm bkg}$ = 10%)	$\epsilon_{\rm sig}$ ($\epsilon_{\rm bkg}$ = 30%)
[600, 700]	BDT-SMH	0.923	0.299	0.757	0.945
[800, 1000]	BDT-SMH	0.969	0.470	0.926	0.994
[1100, 1200]	BDT-SMH	0.989	0.715	0.990	1.000

Performance Metrics

$M_{T'}(GeV)$	MVAs	Metric (%)			
		AUC	Signal efficiency at fixed background efficiency		
			$\epsilon_{\rm sig}$ ($\epsilon_{\rm bkg}$ = 1%)	$\epsilon_{\rm sig}$ ($\epsilon_{\rm bkg}$ = 10%)	$\epsilon_{\rm sig}$ ($\epsilon_{\rm bkg}$ = 30%)
[600, 700]	BDT-NRB	0.992	0.853	0.986	0.998
[800, 1000]	BDT-NRB	0.997	0.951	0.997	1.000
[1100, 1200]	BDT-NRB	0.999	0.990	1.000	1.000
[600, 700]	BDT-SMH	0.983	0.725	0.965	0.994
[800, 1000]	BDT-SMH	0.994	0.889	0.993	0.999
[1100, 1200]	BDT-SMH	0.998	0.973	0.999	1.000

Very good separation between signal and background

Optimization

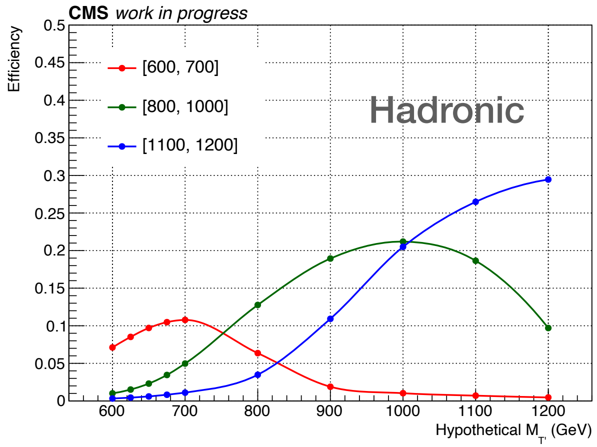
IP2I

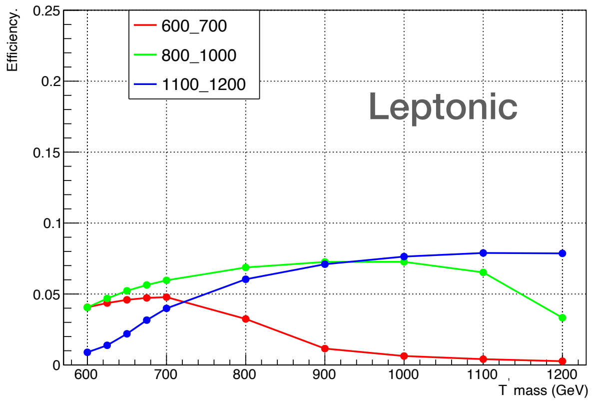
- The analysis is based on H→γγ which has good mass resolution (< 2 GeV) and is indep. of the T' mass
- Design BDT and large T' mass window to improve efficiency for all T' mass points.
- An estimator is used to choose the high significance point

$$\text{Estimator} = -\frac{S}{\sqrt{B}} + \text{Pen},\ \text{Pen} = 0.1 \times \min(0, \text{B}-\text{B}_{\text{min}}) \times (0, \text{B}-\text{B}_{\text{min}}) \times \frac{S}{\sqrt{B}}$$

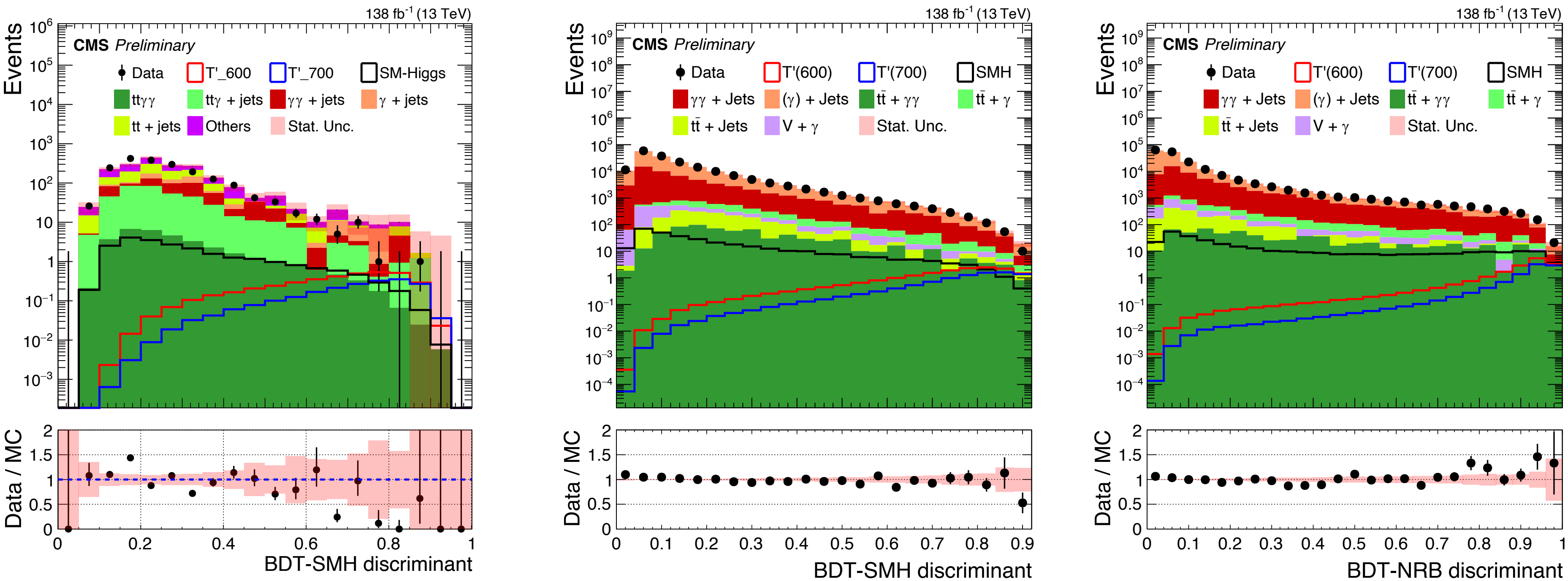
- S: signal yields in [100, 180] GeV
- B: Yields of NRB is estimated from data in the sideband [100, 115]∪[135, 180] GeV
  Yields of SMH is estimated from MC in [100, 180] GeV

MT' (GeV)	[600, 700]	[800, 1000]	[1100, 1200]
Hadronic channel			
Cut value on BDT(NRB)	0.943	0.96	0.95
Cut value on BDT(SMH)	0.80	0.80	0.80
T' mass window	[480, 800]	[550, 1150]	[650, 1600]
N events in sideband	8	17	15
Leptonic channel			
Cut value on BDT(SMH)	0.60	0.40	0.40
T' mass window	[480, 800]	[550, 1150]	[650, 1600]
N events in sideband	10	15	14





BDT output



Reasonable description by the MC of the BDT output discriminant:
→ Recall, the background will be determined from the data not using the MC

Main Systematics Uncertainties

Theoretical uncertainty

10% QCD renormalization & factorization scale uncertainty on ttH

3% Uncertainty on parton density function

2% Uncertainty on the $H \rightarrow \gamma\gamma$ branching fraction

Experimental uncertainty

Nuisance parameters associated with background modeling

5% Jet energy correction

2% Luminosity, shape of photon ID MVA, prefiring of L1 trigger

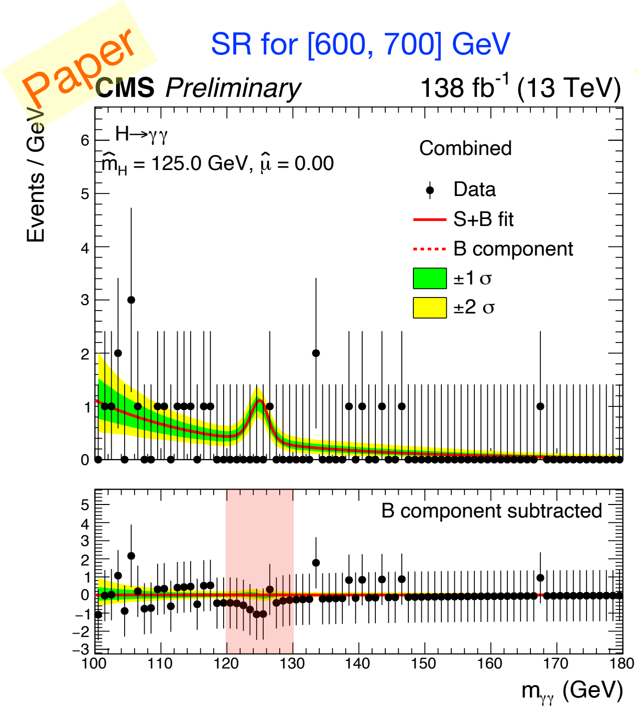
1% MET uncertainty, JER uncertainty, effect of HEM 15/16 failure

$<$ 1% Preselection SFs, b-tag reshape SFs, trigger SFs, etc.

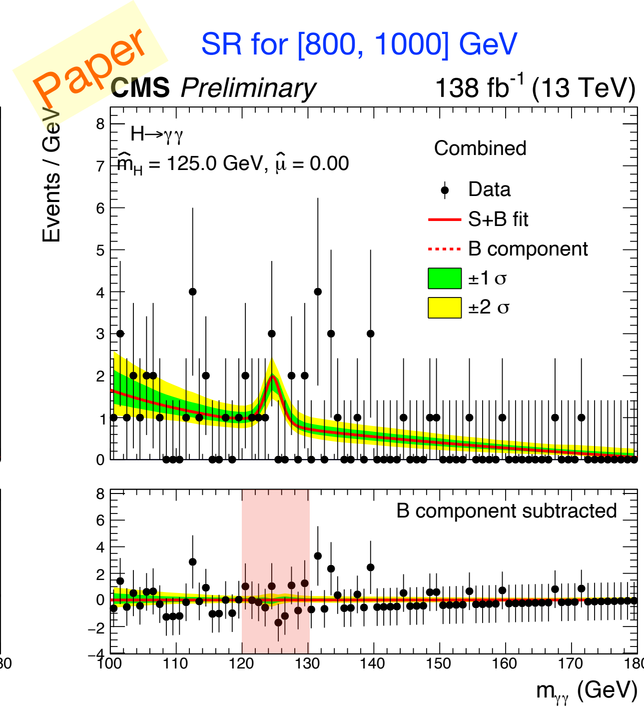
Results

- Signal model is VLQ with a hypothesized T' mass
- Background model is composed of the SM Higgs and the non-resonant backgrounds
- Higgs mass is fixed at 125 GeV in S+B fits

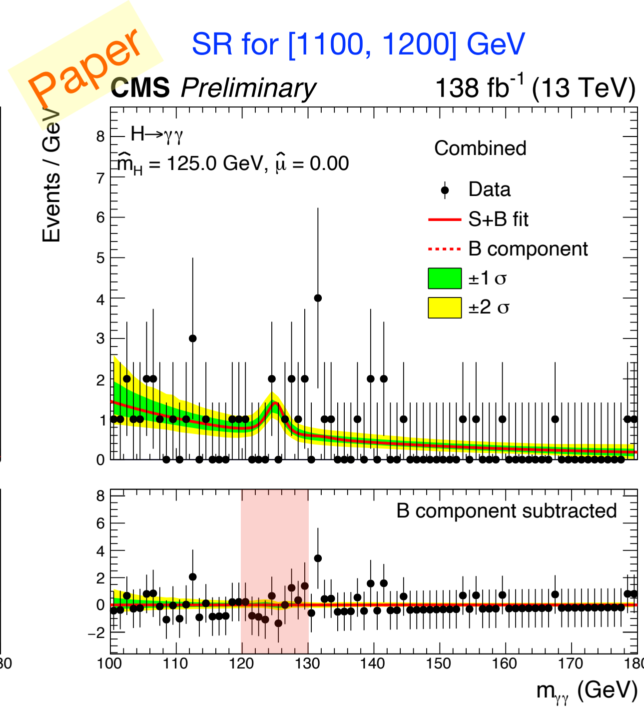
SR for [600, 700] GeV



SR for [800, 1000] GeV



SR for [1100, 1200] GeV



In Numbers

Limits extracted from fit — Composition studied from MC samples with $m_{\gamma\gamma} \in [115, 135]$ GeV

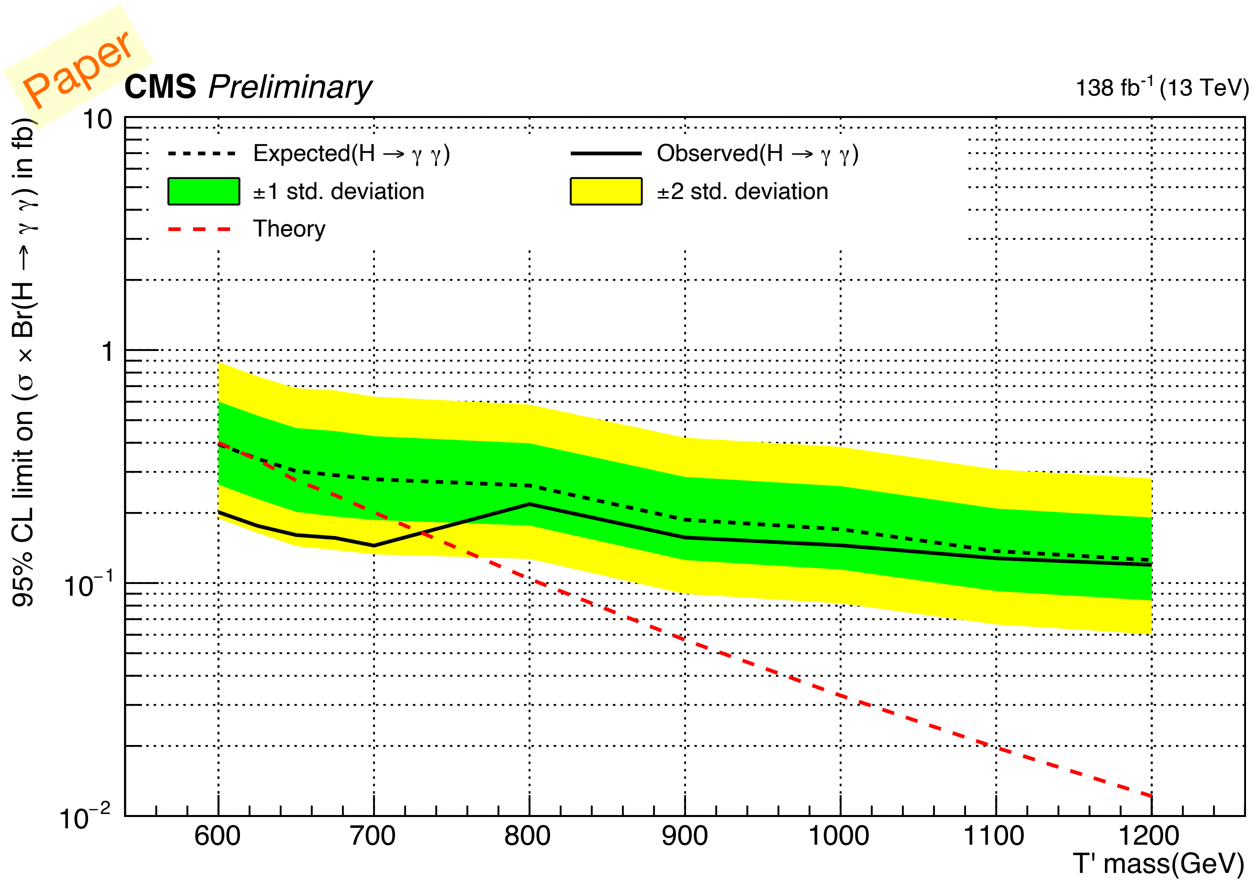
	Limits		Yield		Bkg Composition		SM Higgs Composition (%)				
	observed	expected	VLQ	Tot. Bkg	Non-Res. Bkg	SM Higgs	tth	ggh	thq	vh	vbf
Leptonic T′(600)	1.14	2.06	2.21								
Leptonic T′(625)	1.26	2.28	2.00								
Leptonic T′(625)	1.46	2.63	1.72	12.23 ± 9.08	10.94 ± 8.99	1.29 ± 0.09	61.2	3.8	18.6	14.7	1.6
Leptonic T′(675)	1.65	3.06	1.53								
Leptonic T′(700)	1.92	3.45	1.31								
Leptonic T′(800)	5.24	6.38	0.97								
Leptonic T′(900)	8.99	11.00	0.56	21.34 ± 14.51	19.04 ± 14.39	2.30 ± 0.12	76.0	0.4	12.17	10.0	1.3
Leptonic T′(1000)	15.61	19.12	0.33								
Leptonic T′(1100)	17.61	25.00	0.19	15.83 ± 13.78	14.40 ± 13.68	1.43 ± 0.10	74.8	0.69	12.58	11.8	0.69
Leptonic T′(1200)	28.48	40.75	0.11								

	Limits		Yield		Bkg Composition		SM Higgs Composition (%)				
	observed	expected	VLQ	Tot. Bkg	Non-Res. Bkg	SM Higgs	tth	ggh	thq	vh	vbf
Hadronic T′(600)	0.80	1.24	3.88								
Hadronic T′(625)	0.79	1.24	3.92								
Hadronic T′(625)	0.86	1.32	3.64	3.35 ± 0.95	1.60 ± 0.94*	1.75 ± 0.10	50.3	24.9	19.1	5.2	0.6
Hadronic T′(675)	0.96	1.44	3.40								
Hadronic T′(700)	1.04	1.64	2.96								
Hadronic T′(800)	2.62	2.91	1.81								
Hadronic T′(900)	3.20	3.56	1.47	9.29 ± 4.00	7.30 ± 4.00	1.99 ± 0.12	50.8	29.7	14.9	3.6	1.0
Hadronic T′(1000)	5.04	5.56	0.95								
Hadronic T′(1100)	7.40	7.28	0.71	11.34 ± 5.31	8.98 ± 5.31	2.36 ± 0.15	31.6	50.6	9.5	4.8	3.5
Hadronic T′(1200)	10.84	10.62	0.49								

For low mass part, the SM Higgs is equivalent to VLQ signal

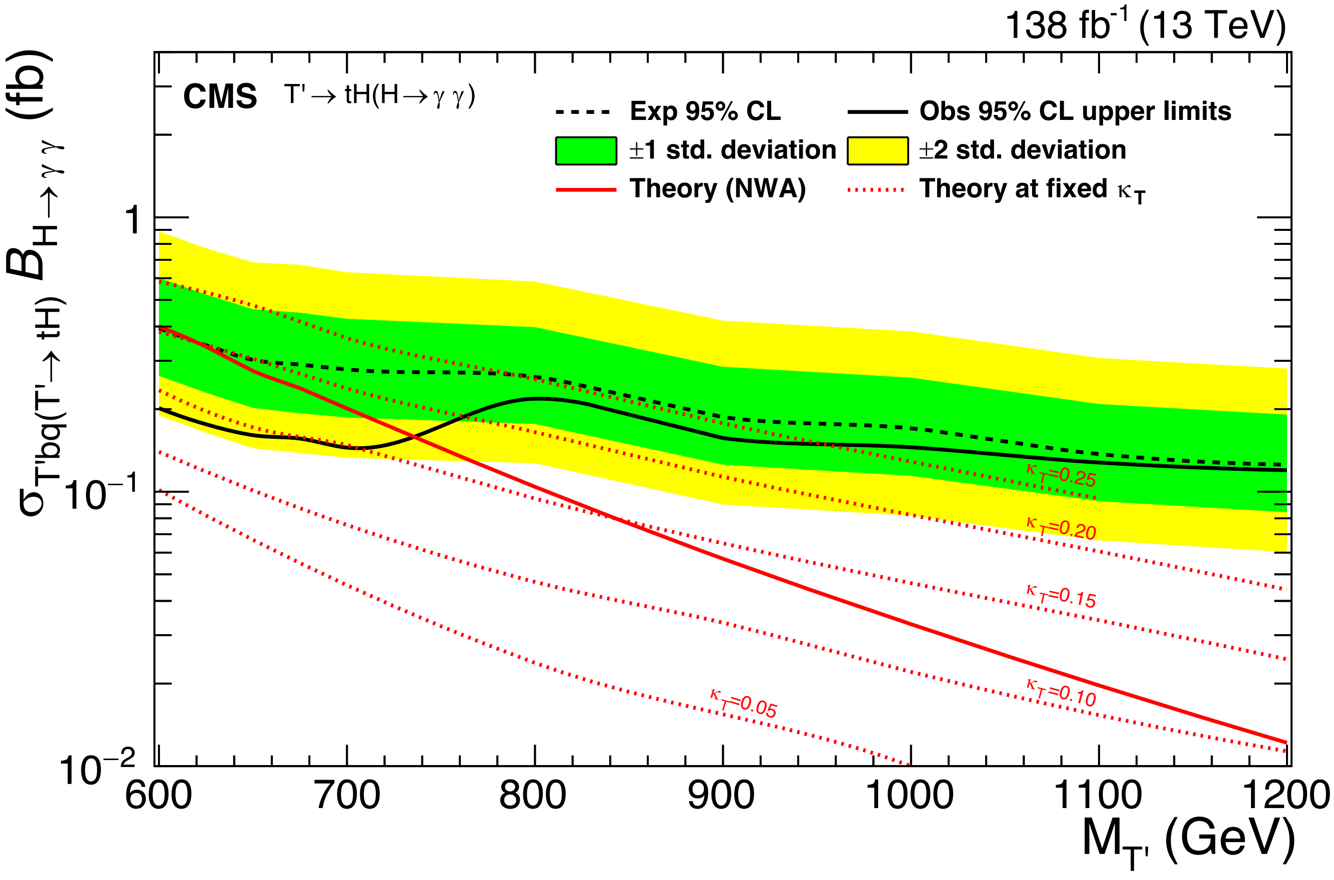
Limits

- There is an exclusion potential between 600 to 730 GeV
- Comparable with results from the search for VLQ in H→b$\bar{b}$ channel

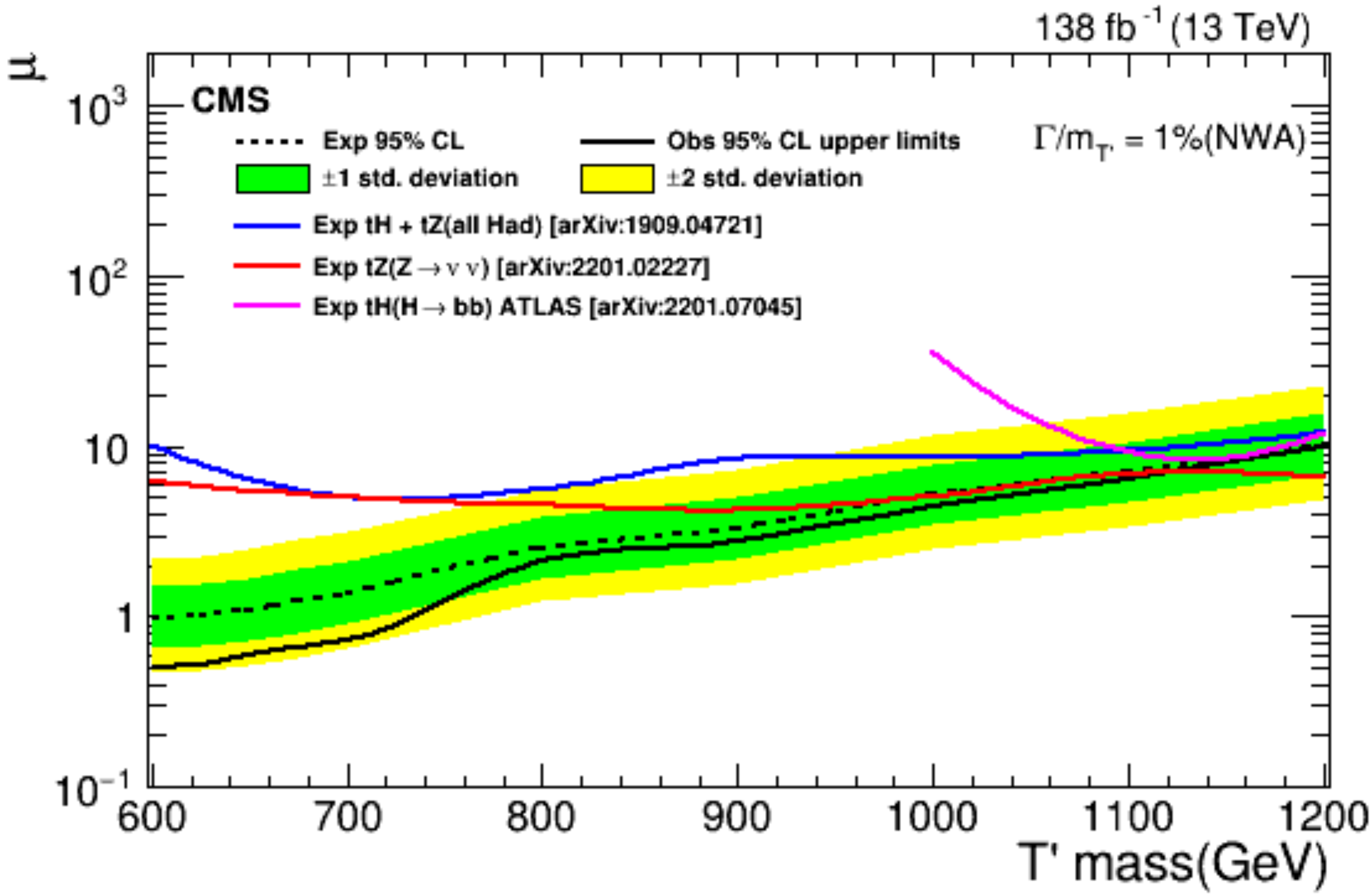


Futher Interpretation

κ_T is a parameter of VLQ models linking to VLQ width, here analysis is performed with narrow width approximation, but still able to exclude some values until the mass resolution



Other Comparison



Despite the low BR of H→γγ, competitive results at 1 TeV

Summary

VLQ searches are performed tackling low BR for Higgs decay but quite pure channel

H→$\gamma\gamma$ rules out the all hadronic excess observed

Competitive results with all hadronic boosted

Limits interpretated as function κ_T parameter, for larger width new signal MC are in the process of being produced

→ Update of the analysis to narrow down κ_T and adding other channels

BackUp