



## Search for di-Higgs resonances decaying to 4 b-jets on CMS at 13 TeV

F. Nechanský<sup>1</sup>, Supervisor Caterina Vernieri<sup>1</sup>;  
Consultants: Silvio Donato<sup>3</sup>, Jacobo Konigsberg<sup>4</sup>;  
Additional members of analysis team: Souvik Das<sup>4</sup>, Andrea Rizzi<sup>2</sup>

<sup>1</sup>FNAL    <sup>2</sup>U Pisa & SNS    <sup>3</sup>U Zurich    <sup>4</sup>U Florida

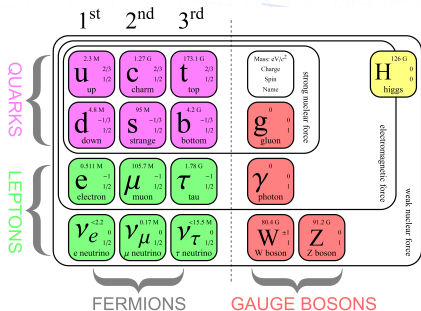
9/22/2016

Final report presentation

# Outline

- ▶ Motivation
- ▶ Trigger efficiency (data,  $t\bar{t}$ )
- ▶ Preselection
- ▶ Signal Region
- ▶ Regression
- ▶ Background
- ▶ Limits
- ▶ Summary

# Motivation



- ▶ 2012 Higgs discovery completes the Standard model  
 $m_H = 125.7 \pm 0.4 \text{ GeV}$
- ▶ SM incomplete, does not explain e.g.:
  - ▶ Neutrino mass
  - ▶ Dark matter
  - ▶ Dark energy
  - ▶ Gravity

- ▶ Necessary to go beyond standard model
- ▶ Many theories with various predictions and free parameters - e.g. SuperSymmetry



Is there anything beyond the Standard Model?

# Experimental Point of View

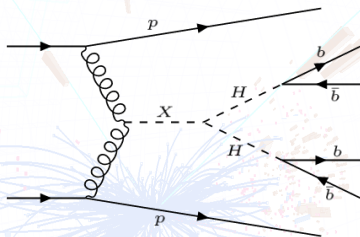
- ▶ New particle Higgs Boson
  - in which ways it can be produced?
- ▶ Many modern BSM theories
  - which one is true?
- ▶ Some theories predict particle ( $X$ ) decaying to pair of Higgs bosons (Randall-Sundrum radion, massive KK graviton[1], 2HDM[2])
- ▶ Possible channels of Higgs decay, e.g.:
  - ▶  $H \rightarrow \gamma\gamma$ , BR: 0.23%, high res. (1-2%)
  - ▶  $H \rightarrow b\bar{b}$ , BR: 57.7%, low res. ( $\approx 10\%$ )



This is why experimental scientists hate theoretical scientists.

# Current search

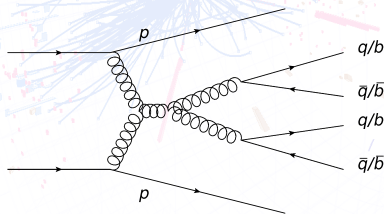
- ▶ This presentation reports on channel  $X \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ :



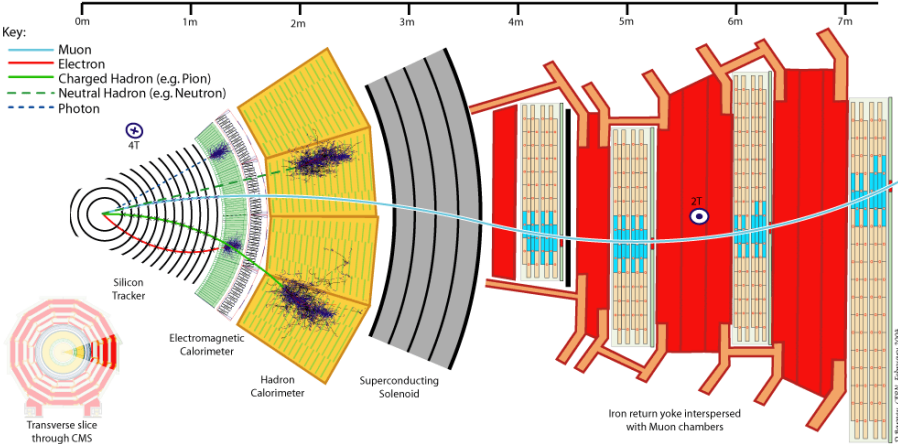
- ▶ Best sensitivity for X mass  $m_X > 400$  GeV
- ▶ Done on CMS experiment at CERN for  $\sqrt{s} = 13$  TeV using data from 2016 (currently  $9.3 \text{ fb}^{-1}$ )
- ▶ Similar study was done at 8 TeV ( arXiv:1503.04114 ) and at 13 TeV ( 2015 data, CMS-PAS-HIG-16-002 )

# Analysis work-flow

1. Four jet events dominated by e.g. multi-jet background, necessary to select only events with  $b$ -jets  $\implies$   $b$ -tagging
2. Triggers not modelled perfectly  $\implies$  study of trigger behaviour
3. After selection of four  $b$ -jets need to check if they originate from Higgs decay
4. Corrections for imperfection of detector on jet energy/momentum
5. Subtraction of  $4b$  multi jet background
6. Estimation of cross-section for new processes

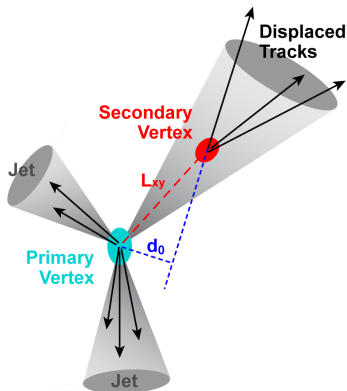


# Compact Muon Solenoid (CMS)



# Identification of $b$ -jets ( $b$ -tagging)

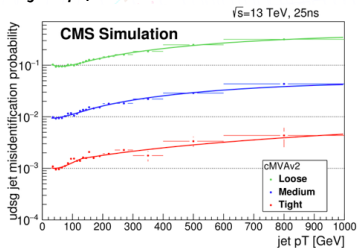
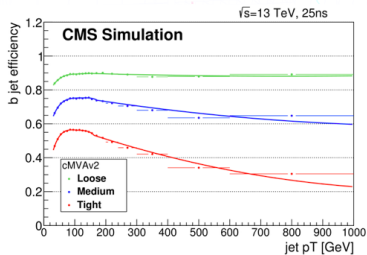
- ▶  $b$ -quark hadronize to  $b$ -hadrons with relatively large lifetime
- ▶ Identification using secondary vertex (few mm from PV)
- ▶ Tracks often have non-zero and positive impact parameter
- ▶ Multivariate discriminant exploit this information to identify  $b$ -jets: e.g. CSV (online) and CMVA (offline)





# CMVA performance

- ▶ Performance of discriminants characterized by:
  - ▶  $b$ -jet efficiency - fraction of  $b$ -jets correctly identified
  - ▶ misidentification prob. - probability of identifying non- $b$ -jet as  $b$ -jet
- ▶ Performance of CMVA as function of jet  $p_T$ :



- ▶ Medium working point used



# Trigger efficiency

# Triggers

- ▶ LHC produces large amount of collisions ( $\approx$ MHz), impossible to record all
- ▶ Preference for interesting events  $\implies$  implementation of triggers
- ▶ Triggers decide based on portion event information to keep the event or not
- ▶ Search for  $4b$  resonance  $\implies$  necessary reduction of background (multi jet)  
 $\implies$  online  $b$ -tagging

## Quad Jet trigger (QJ):

HLT\_BIT\_HLT\_QuadJet45\_TripleBTagCSV\_p087\_v

- ▶ L1 jet activity
- ▶ 4 jets  $|\eta| < 2.6, p_T > 45$  GeV  
(Calorimeter and Particle flow level)
- ▶ three  $b$ -tagged jets

## Double Jet trigger (DJ):

HLT\_BIT\_HLT\_DoubleJet90\_Double30\_TripleBTagCSV\_p087\_v

- ▶ L1 jet activity
- ▶ 4 jets  $|\eta| < 2.6, p_T > 30$  GeV
- ▶ 2 jets  $|\eta| < 2.6, p_T > 90$  GeV  
(Calorimeter and Particle flow level)
- ▶ three  $b$ -tagged jets

# Trigger efficiency

- ▶ Not all events detected by the trigger - necessary to derive correction
- ▶ Complicated triggers - usage of data driven technique
- ▶ Consider trigger, that requires event to pass specific selections  $A, B, C, D, \dots$ , then it can be shown that efficiency can be rewritten:

$$P(A \& B \& C \& \dots) = P(A) \cdot P(B|A) \cdot P(C|A \& B) \cdot \dots$$

- ▶ Trigger divided in stages, each studied as function of some relevant variable
- ▶ e.g. Quad Jet trigger:
  - ▶ L1 as function of sum of  $p_T$  of four leading jets ( $\sum^4 p_T$ )
  - ▶ Four Calorimeter-jet selection as function of  $p_T$  of the fourth jet ( $p_{T,4}$ )
  - ▶ Three  $b$ -tagged jets as function of discriminant of the third jet ( $CSV_3$ )
  - ▶ Four Particle-flow-jets selection as function of  $p_T$  of the fourth jet ( $p_{T,4}$ )

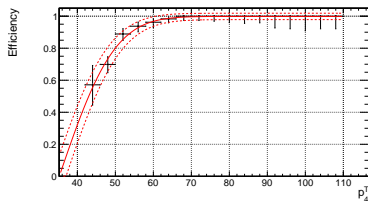
- ▶ Efficiency of each step estimated from data (turn-on function)
- ▶ Trigger efficiency can be written as:

$$Eff(\sum p_T, p_{T,4}, CSV_3) = Turn\_on\_L1(\sum p_T) \cdot Turn\_on\_Calo\_pt4(p_{T,4}) \cdot Turn\_on\_btag(CSV_3) \cdot Turn\_on\_PF\_pt4(p_{T,4})$$

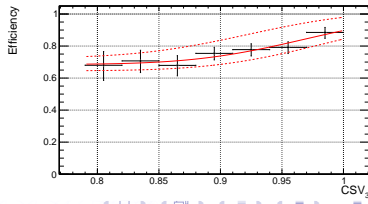
- ▶ Preselection designed to resemble final selection
  - ▶ Preselection (orthogonal) trigger HLT\_BIT\_HLT\_IsoMu24\_v
  - ▶ diHiggs cut - two pairs of b-jets compatible with Higgs mass
  - ▶ Only considering jets with  $|\eta| < 2.6$ ,  $CMVA > 0.185$  and  $PUID \geq 4$

Turn-on:

Calot pt4:

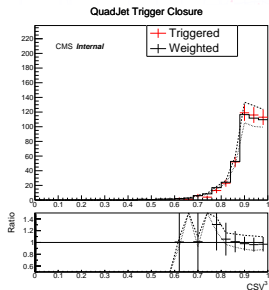


CSV3:

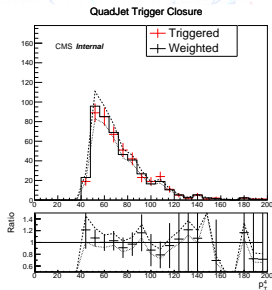


# Validation

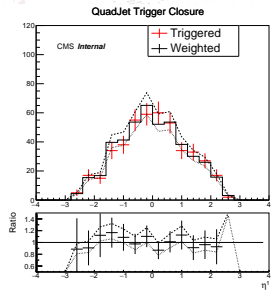
- ▶ It is necessary to validate derived efficiencies (closure tests)
- ▶ Done for events after preselection
- ▶ Compare distributions of events weighted by the efficiency and of events which pass the studied trigger



CSV<sub>3</sub>



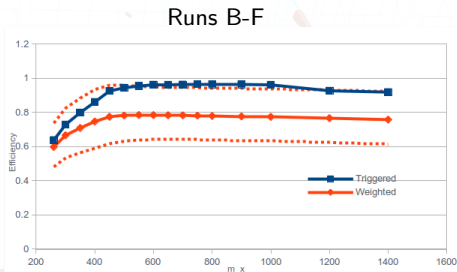
p<sub>T,4</sub>



η<sub>1</sub>

# Comparison between data and signal MC efficiency

- ▶ Done for events after analysis preselection and diHiggs selection
- ▶ Comparison between events weighted by data driven efficiency and event passing the triggers



- ▶ Problem during data-taking for Runs B-F - tracking inefficiency not present in our simulation
- ▶ Results consistent nevertheless (within stat. uncertainty)

CMS Experiment at the LHC, CERN

Data recorded: 2012-May-27 23:35:47.271030 GMT

Run/Event: 195099 / 137440354

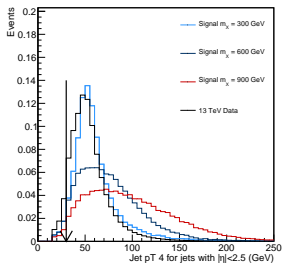
# HbbHbb analysis of 2016 data

A 3D visualization of a particle detector, likely the CMS detector, showing a central collision point with numerous tracks radiating outwards. The tracks are colored in green, orange, and blue. The detector structure is shown in a light blue, semi-transparent style. The background is dark grey with some faint red wireframe structures.

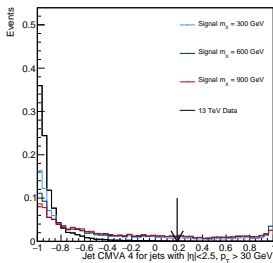


# Event selection

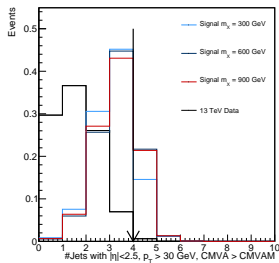
- ▶ Trigger: Quad Jet OR Double Jet Trigger
- ▶ At least 4 jets with:
  - ▶  $p_T > 30$  GeV,  $|\eta| < 2.5$ ,
  - ▶  $CMVAV2 > 0.185$



$p_{T,4}$



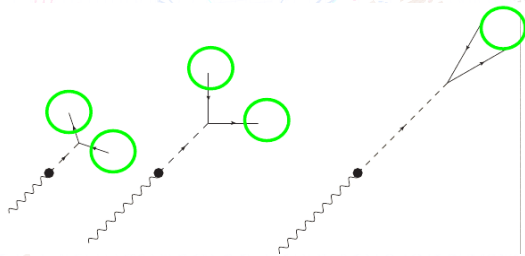
B-tagging discriminant



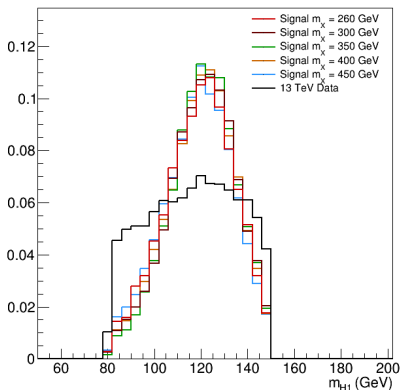
Number of selected jets

# Higgs selection

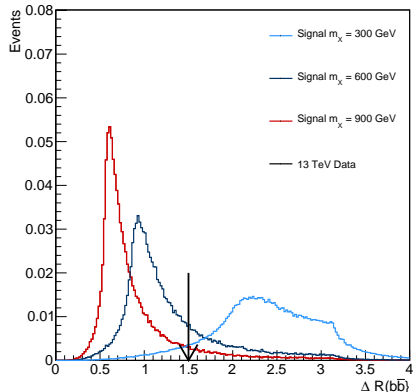
- ▶ Four  $b$ -jets from two Higgs bosons - looking for two Higgs candidates (H1,H2)
- ▶ Several regions (based on invariant mass of the original particle  $X$ ) due to different event topology:
  - ▶ Low mass region (LMR): two dijets with mass compatible with Higgs boson
  - ▶ Medium mass region (MMR):  $b$ -jets are more boosted  $\rightarrow$  requirement on distance between the  $b$ -jets ( $\Delta R < 1.5, \Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ )



LMR:



MMR (Gen):

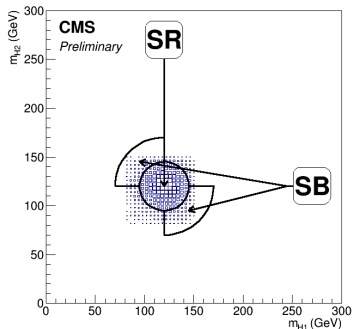


left Comparison of dijet mass after Higgs selection for LMR

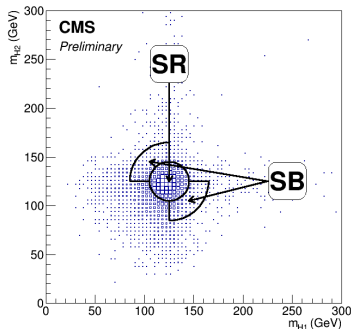
right Distance between two jets from Higgs decay (generated level)

# Signal selection

- ▶  $\chi^2 = (m_{H1} - \bar{m}_H)^2 / \bar{\sigma}_H^2 + (m_{H2} - \bar{m}_H)^2 / \bar{\sigma}_H^2$
- ▶ Reconstruct two Higgs Boson candidates with lowest value of  $\chi^2$
- ▶ Signal region (SR):  $\chi < 1$
- ▶ Sideband region (SB):  $1 < \chi < 2$ ,  $(m_{H1} - \bar{m}_H) \cdot (m_{H2} - \bar{m}_H) < 0$



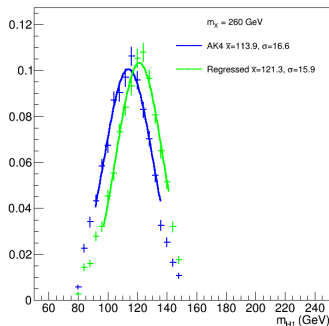
MC  $m_\chi = 350$  GeV



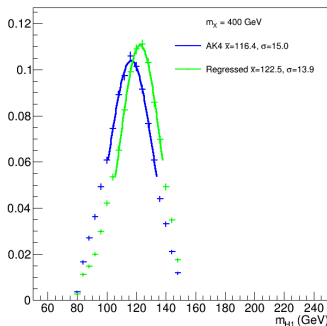
MC  $m_\chi = 650$  GeV

# Regression

- ▶ Presence of neutrinos + imperfection of the detector  
⇒ lower recorded energy of jets
- ▶  $b$ -jet  $p_T$  correction using regression: uses multivariate algorithm trained on signal Monte Carlo
- ▶ e.g. LMR:



Improvement 10.06%



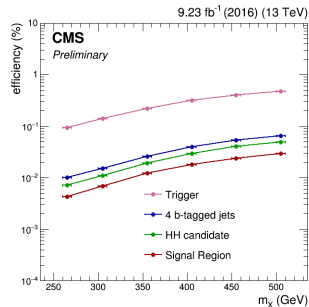
Improvement 11.94%

# Signal region optimization

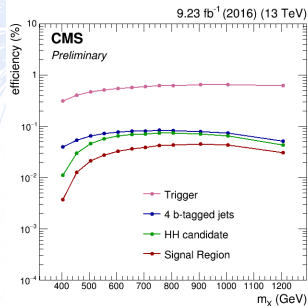
- ▶ Signal region is optimized for both before and after the regression
- ▶ Regression reduces radius of the SR and leads to better signal significance

	LMR		MMR	
	center	radius	center	radius
baseline	115	25	120	20
after regression	120	20	125	20

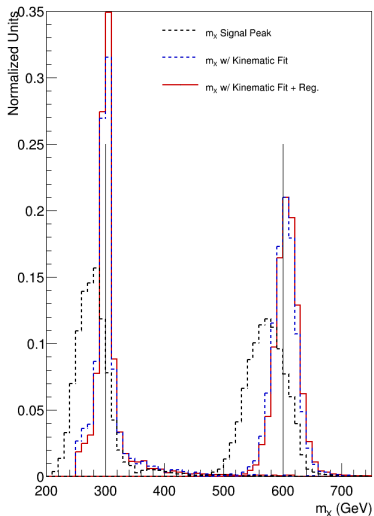
LMR:



MMR:



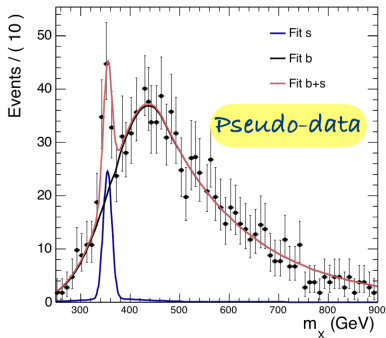
# Kinematic fit



- ▶ We expect  $m_{bb} = 125$  GeV (Higgs mass)
- ▶ Correction of momentum to achieve the invariant mass
- ▶ Modification based on resolution of jet variables ( $p_T, \eta, \phi$ ) to achieve the lowest possible  $\chi^2$
- ▶ Resolution measured in Monte Carlo

	$m_x$	300	600
baseline	$\mu$	276.3	572.8
	$\sigma$	23	32.2
	$\sigma / \mu$	8.32	5.62
kinematic fit	$\mu$	300.6	604.3
	$\sigma$	7.6	17.7
	$\sigma / \mu$	2.53	2.93
kinematic fit + regression	$\mu$	301.1	606.9
	$\sigma$	7.7	17.5
	$\sigma / \mu$	2.56	2.88

# Signal extraction



- ▶ Signal as a peak on smooth background
- ▶ Fit done for LMR/MMR, components:
  - ▶ Signal - fit shape determined from MC
  - ▶ Multi-jet background - determined from data (poor modelling)
  - ▶  $t\bar{t}$  - negligible

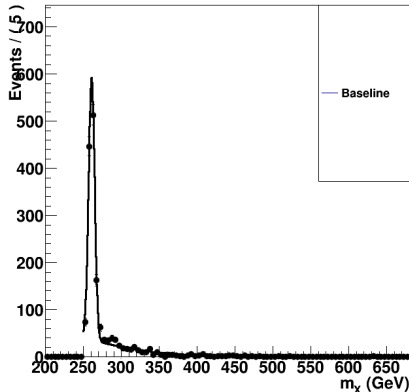


# Signal fits

- ▶ Low mass: Gaussian signal + Gaussian combinatoric background
- ▶ High mass: ExpGaussExp function

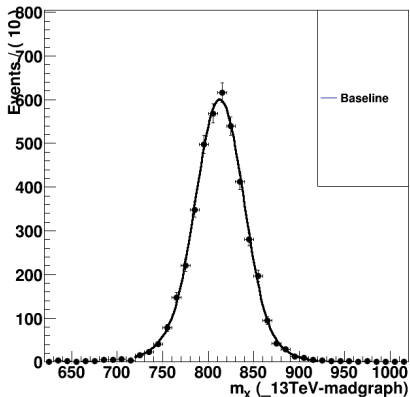
Signal 260 GeV (LMR):

A RooPlot of " $m_x$  (GeV)"



Signal 800 GeV (MMR):

A RooPlot of " $m_x$  (\_13TeV-madgraph)"

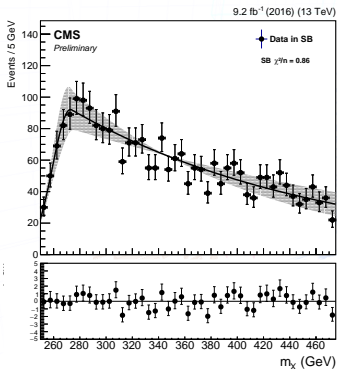


# Background modeling

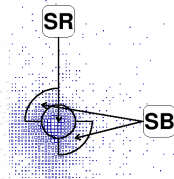
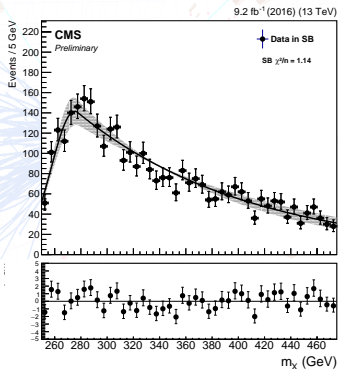
- ▶ Signal region blinded
  - fit in Sideband (SB), using Gauss-Exp function

SB definition:  
(H1 vs H2)

Example: SB+LMR  
Without regression:



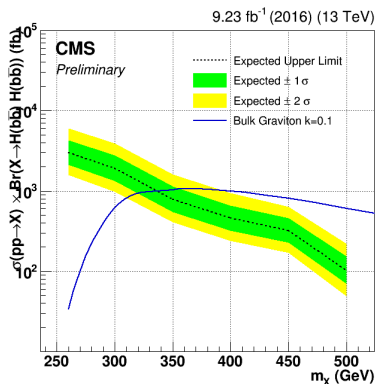
With regression:



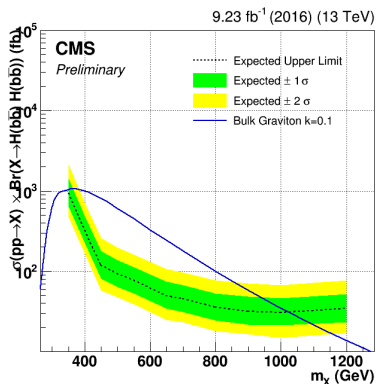
- ▶ Regression does not sculpt the background

# Expected upper limits

LMR:



MMR:



- ▶ Expected upper limits on the signal cross sections at 95% confidence level (computed using Asymptotic CL<sub>S</sub> method)
- ▶ Already better exclusion with only 9 fb<sup>-1</sup> than at 8 TeV
- ▶ With regression improvement 3.3-4.1% for LMR and 9.4-19.0% for MMR ( $m_X > 400$  GeV)

# Summary

- ▶ This talk summarizes current status of the 2016 search for two Higgs boson resonance decaying into four  $b$  quarks
- ▶ My contribution over the summer:
  - ▶ Bulk of work on trigger efficiency
  - ▶ Optimization of Signal region
  - ▶ Study of effects of regression
- ▶ Corrections already applied: regression of jet  $p_T$ , kinematic fit on Higgs mass; improvement of mass resolution
- ▶ Optimization of signal region, background estimation and expected upper limits finished

Thank you for your attention!



L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension," Phys. Rev. Lett. **83** (1999) 3370  
doi:10.1103/PhysRevLett.83.3370  
[hep-ph/9905221].

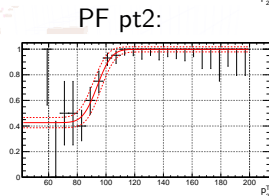
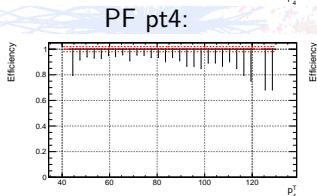
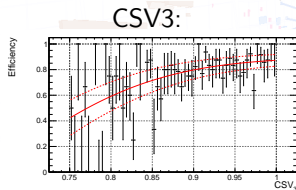
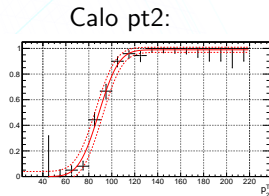
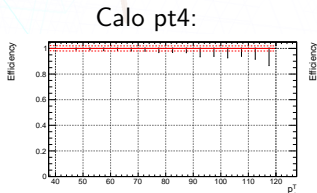
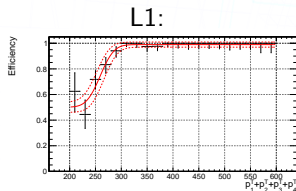


R. Barbieri, D. Buttazzo, K. Kannike, F. Sala and A. Tesi, "Exploring the Higgs sector of a most natural NMSSM," Phys. Rev. D **87** (2013) no.11, 115018 doi:10.1103/PhysRevD.87.115018  
[arXiv:1304.3670 [hep-ph]].

# Backup slides

A 3D visualization of a particle detector, likely a CMS detector, showing a central interaction point with a dense spray of blue tracks. Surrounding this are various colored tracks (red, pink, brown) and clusters of particles. The detector structure is represented by a light blue grid and wireframe.

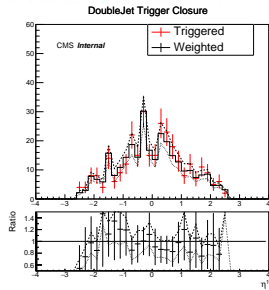
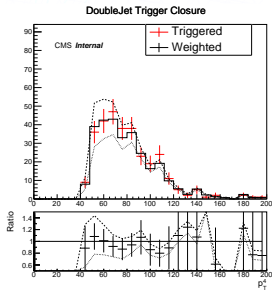
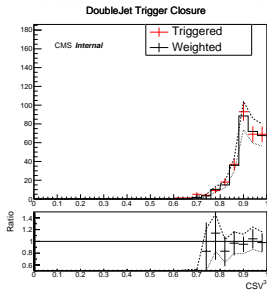
# Turn-ons (e.g. Double Jet, data driven)



Bands visualize uncertainty of the fit

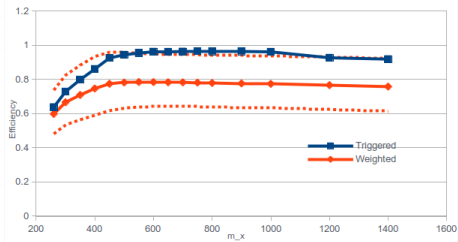
# Closure test

Double Jet trigger (DJ):

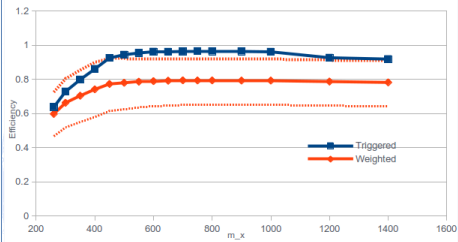




### Runs B-F

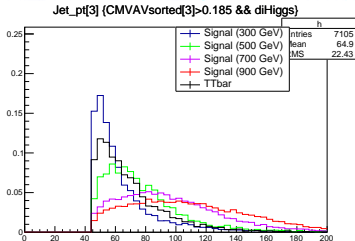
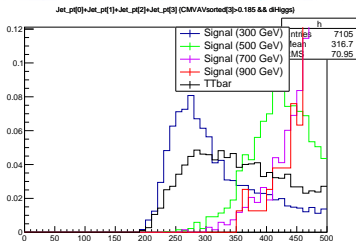


### Run G



# Kinematic compatibility

- Comparison of  $t\bar{t}$  sample and several signal masses:



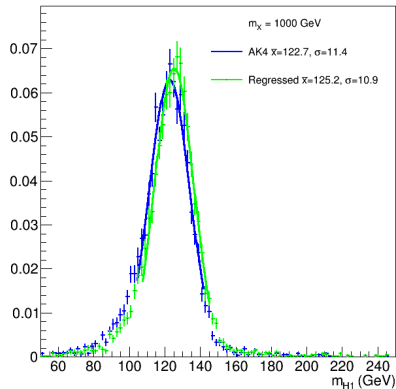
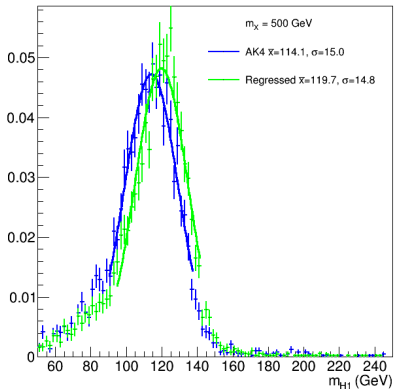
$$\sum^4 Jet_{pT}$$

$$Jet_{pT,4}$$

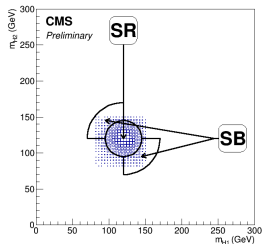
- Low mass kinematically more similar than high mass  
= same behavior as in efficiency comparison

# Regression (MMR)

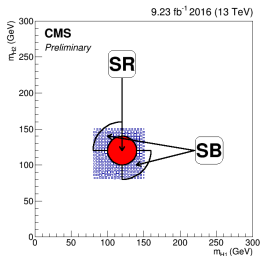
▶ = bjet  $p_T$  correction



# Signal region

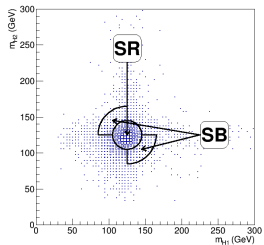


MC  $m_\chi = 300$  GeV

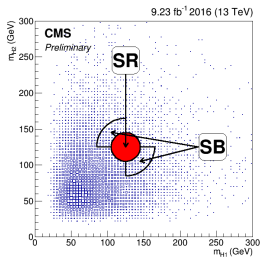


Data

< LMR



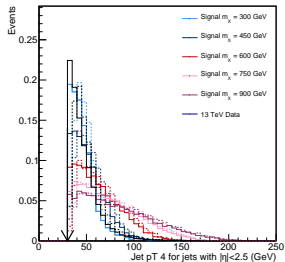
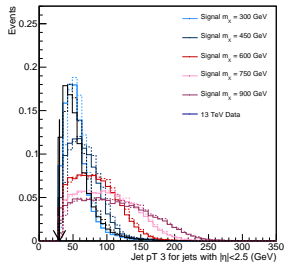
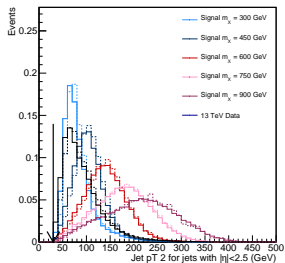
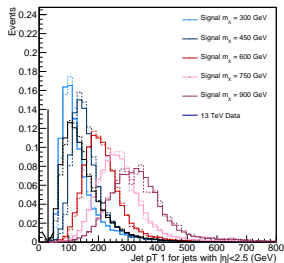
MC  $m_\chi = 900$  GeV



Data

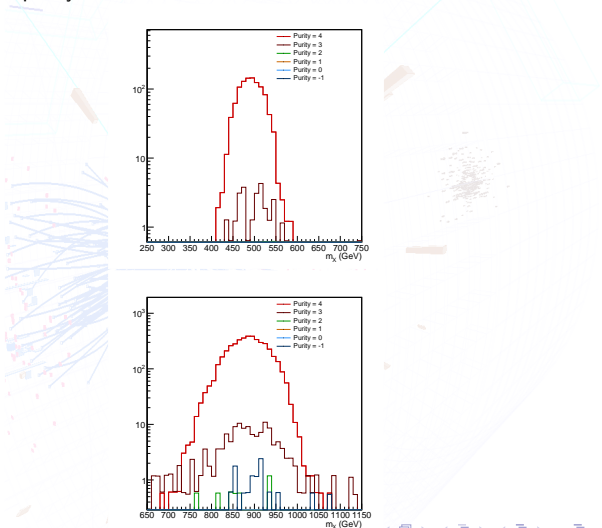
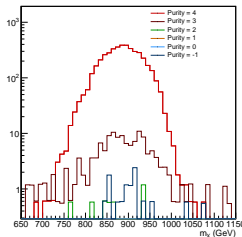
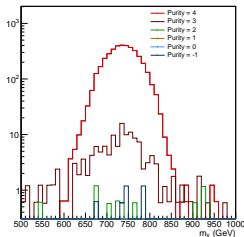
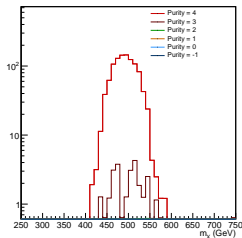
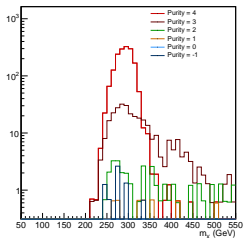
< MMR

# pt(1,2,3,4) before/after regression



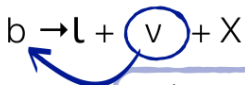
# Purity

Purity is equal to number of b-jets matched to generated b-quark. When two jets are matched to one quark, the purity is -1.



# b-jet energy MVA regression

B.R. 35%



Multidimensional calibration  
targeting the gen jet  $p_T$

- ✓ Basic kinematic and jet structure
- ✓ SecVtx and soft lepton information
- ✓ MET related

- ✓ Final resolution is 8-10%
- ✓ The sensitivity increases by **10-20%**
- ✓ The VZ/VH separation power improves

Jet  $\eta$   
Jet  $p_T$   
Jet Mass  
Neutral hadron energy fraction  
Photon energy fraction

SecVtx Decay Length  
SecVtx Mass  
Soft Lepton  $p_T$   
Soft Lepton  $p_T^{\text{Rel}}$

$N_{\text{tot}}$ , total number of jet constituents  
 $p_T$  of the Leading Track

MET  
 $d\Phi(\text{MET}, \text{jet})$   
 $\rho$

# Signal significance

- ▶ Study of signal significance to determine best signal region
- ▶ e.g. LMR: (TODO update or remove)

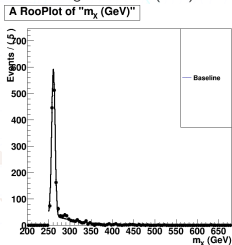
Before Reg.						
$\tilde{m}_H$ [GeV]		115	110	120	115	115
$\sigma_H$ [GeV]		30	30	30	25	35
$m_\chi$	260	46.82	48.40	42.91	45.17	48.14
	300	59.35	58.49	57.33	58.41	59.49
	350	97.18	93.48	96.47	97.02	93.20
	400	144.53	137.74	144.25	143.30	140.28
	450	74.28	73.94	72.57	76.10	70.49
	500	52.61	52.30	50.85	55.16	48.98
	550	147.56	144.60	142.90	155.50	137.96
After Reg.						
$\tilde{m}_H$ [GeV]		120	125	120	120	120
$\sigma_H$ [GeV]		25	25	20	30	35
$m_\chi$	260	49.89	47.30	46.44	51.50	53.43
	300	64.87	63.31	61.33	65.54	65.91
	350	112.73	112.73	107.80	113.06	109.79
	400	190.34	188.10	186.58	187.89	176.66
	450	91.47	87.40	91.64	88.06	82.30
	500	63.74	61.11	65.67	60.40	56.87
	550	178.99	170.77	184.70	168.21	152.47



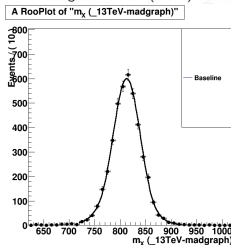
# $m_x$ fits and resolution

- ▶ Low mass: Gaussian signal + Gaussian combinatoric background
- ▶ High mass: ExpGaussExp function

Signal 260 GeV (LMR):



Signal 800 GeV (MMR):



	$m_x$	260	300	350	450	600	700	800	900
baseline	$\mu$	243.4	276.3	318.5	425.3	572.8	673.5	771.9	870.6
	$\sigma$	18.4	23	28.7	25.2	32.2	35.5	38.9	42.1
	$\sigma / \mu$	7.56	8.32	9.01	5.93	5.62	5.27	5.04	4.84
kinematic fit	$\mu$	260.7	300.6	350.4	452.4	604.3	706.6	807.9	908.3
	$\sigma$	4.1	7.6	10.4	12	17.7	21.5	27.3	31.1
	$\sigma / \mu$	1.57	2.53	2.97	2.65	2.93	3.04	3.38	3.42
kinematic fit + regression	$\mu$	260.9	301.1	351.6	453.4	606.9	710	812.2	913.9
	$\sigma$	3.9	7.7	10	12.3	17.5	21.3	27.1	31.4
	$\sigma / \mu$	1.49	2.56	2.84	2.71	2.88	3.00	3.34	3.44

## $m_x$ fits and resolution - ExpGaussExp function

- ▶ High mass: ExpGaussExp function:

$$\begin{aligned} f(x; \bar{x}, \sigma, k_L, k_H) &= \exp\left(\frac{k_H^2}{2} - \frac{k_H(x - \bar{x})}{\sigma}\right), \quad \text{for } \frac{x - \bar{x}}{\sigma} > k_H \\ &= \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right), \quad \text{for } k_L \leq \frac{x - \bar{x}}{\sigma} \leq k_H \\ &= \exp\left(\frac{k_L^2}{2} + \frac{k_L(x - \bar{x})}{\sigma}\right), \quad \text{for } \frac{x - \bar{x}}{\sigma} < k_L \end{aligned} \quad (1)$$

- ▶  $\bar{x}$ : The mean of the Gaussian core,
- ▶  $\sigma$ : The standard deviation of the Gaussian core,
- ▶  $k_L$ : The decay-coefficient of the lower exponential tail. This is also the number of standard deviations, on the low side, beyond which the Gaussian inflects into the exponential.
- ▶  $k_H$ : The decay-coefficient of the higher exponential tail. This is also the number of standard deviations, on the high side, beyond which the Gaussian inflects into the exponential.

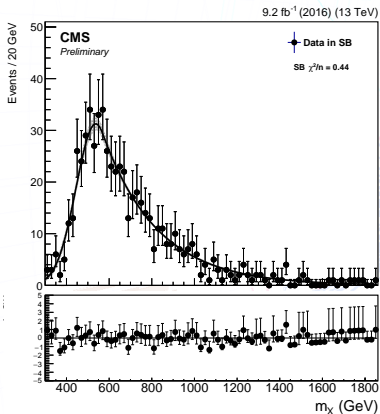
# Background modeling - ExpGaus function description

- ▶ Signal region blinded - fit in Sideband, using Gauss-Exp function:
  - ▶  $\bar{x}$ : The mean of the Gaussian core,
  - ▶  $\sigma$ : The standard deviation of the Gaussian core,
  - ▶  $k$ : The decay-coefficient of the exponential tail. This is also the number of standard deviations beyond which the Gaussian inflects into the exponential on the high side.

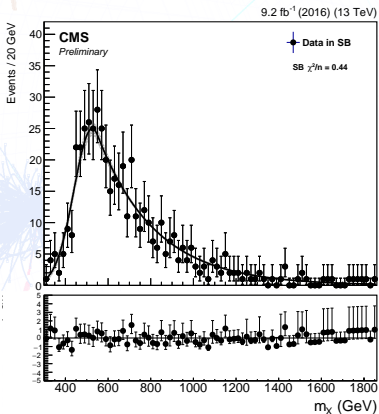
$$\begin{aligned} f(m_X; \bar{x}, \sigma, k) &= \exp\left(-\frac{1}{2}\left(\frac{x - \bar{x}}{\sigma}\right)^2\right), \quad \text{for } \frac{x - \bar{x}}{\sigma} \leq k \\ &= \exp\left(\frac{k^2}{2} - k\frac{x - \bar{x}}{\sigma}\right), \quad \text{for } \frac{x - \bar{x}}{\sigma} > k \end{aligned} \quad (2)$$

# Background modeling: SB+MMR

Without regression:



With regression:



# Regression: improvement of limits

	Kin. Fit	KF+regr.	Improv.(%)
<b>LMR</b>			
260	3085	2976	3.7
300	1976	1898	4.1
350	808	781	3.5
400	470	455	3.3
<b>MMR</b>			
350	-	925.8	-
400	315.9	321.8	-1.8
450	140.6	118.2	19.0
500	108.9	92.8	17.3
550	89.4	76.7	16.6
650	59.1	49.3	19.9
700	52.2	45.4	15.0
800	40.5	35.6	13.8
900	35.6	31.7	12.3
1000	33.7	30.8	9.4
1200	38.6	34.7	11.2

- ▶ With regression improvement  
3.3-4.1% for LMR and  
9.4-19.0% for MMR  
( $m_X > 400$  GeV)