A Di-Higgs Resonance Search at the LHC
X → H(bb)H(bb)

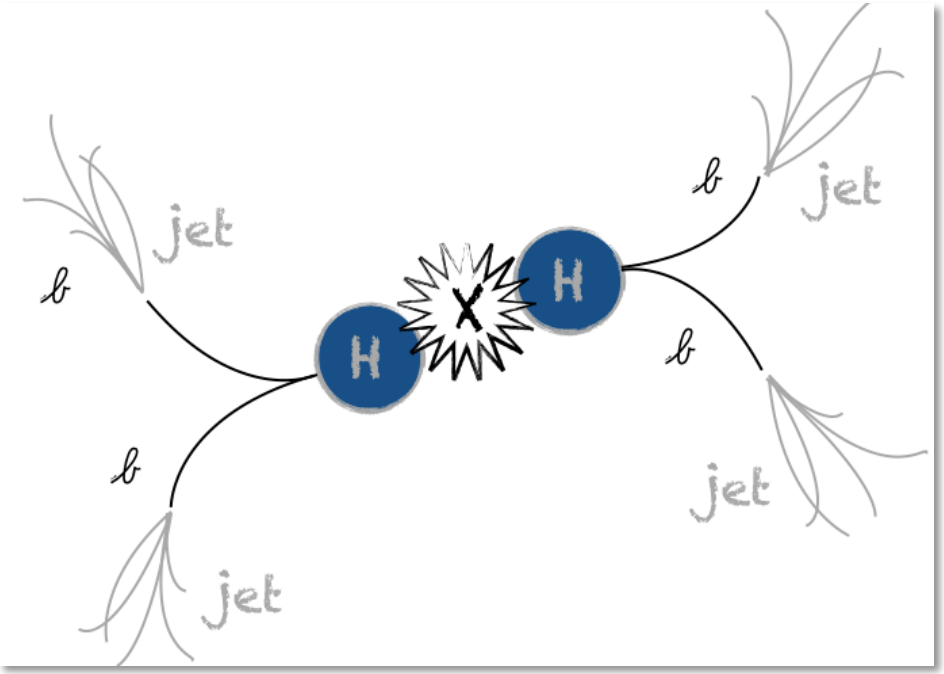
Caterina Vernieri and Andrea Rizzi
INFN, Scuola Normale Superiore, Pisa

Souvik Das, Jacobo Konigsberg and Jia Fu Low
University of Florida

December 20, 2013

Outline

- **Introduction**
  - Theoretical Motivations
  - Analysis Strategy
- **Analysis**
  - The Trigger and Data
  - Signal Samples
  - Physics Objects & Event Selection
  - Systematic Uncertainties
  - Signal Modeling
  - Background Modeling
- **Results**
  - Bias Studies
  - Expected Upper Limits
- **Conclusion**



Analysis Note

Available on the CMS information server

AN-13-227

CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

2013/08/14
Head Id: 198589
Archive Id: 200811:201202M
Archive Date: 2013/07/25
Archive Tag: trunk

Search for di-Higgs resonances decaying to 4 b-jets in pp collisions at 8 TeV

Souvik Das[1], Michele de Gruttola[1], Jacobo Konigsberg[1], Jia Fu Low[1], Caterina Vernieri[2], and Andrea Rizzi[2]

[1] University of Florida
[2] University of Pisa

DRAFT

Abstract

A search for a model-independent, narrow, di-Higgs resonance with the Higgs bosons reconstructed through their decay to bottom quarks is presented in this analysis note. The search is performed in 18.6 fb^{-1} of proton-proton collision data recorded by the CMS detector at the LHC with a center-of-mass energy of 8 TeV during 2012. Upper limits on the production cross section for such a resonance, with masses between 300 GeV and 1 TeV, are reported in this note.

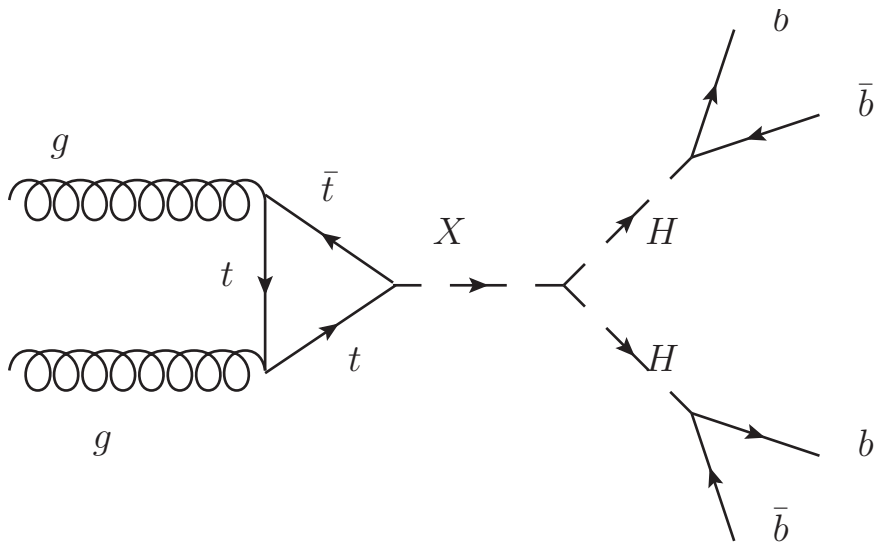
This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor: Souvik Das
PDFTitle: Search for di-Higgs resonances decaying to 4 b-jets in pp collisions at 8 TeV
PDFSubject: CMS
PDFKeywords: CMS, physics, Higgs, resonance

Please also verify that the abstract does not use any user defined symbols

Introduction

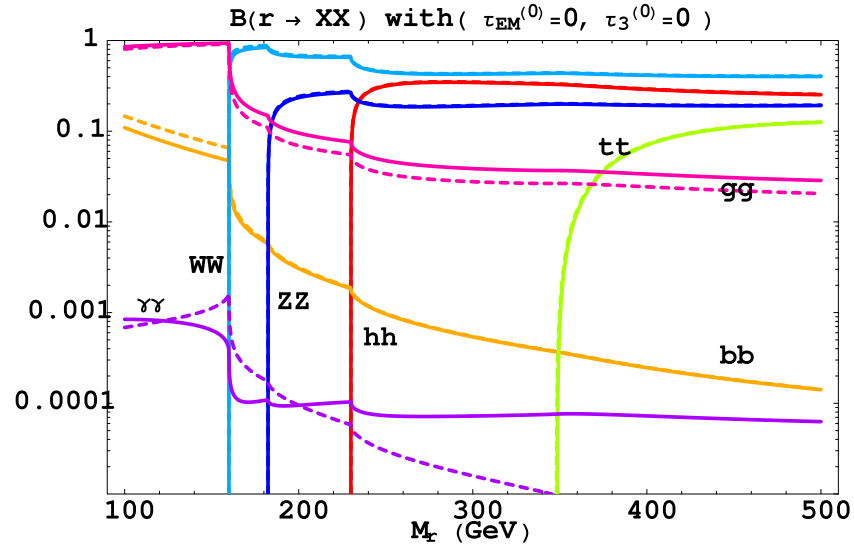
What We Are Searching For



- A model-independent resonance, *X*, decaying to two Standard Model Higgses, decaying to 4 b-jets
- Having discovered the Higgs, it is important to ask if it is pair-produced
- Signal would be buried under immense 4-jet QCD background. Challenge is to **unearth it using b-tagging** and a **data-driven estimation of the background**.

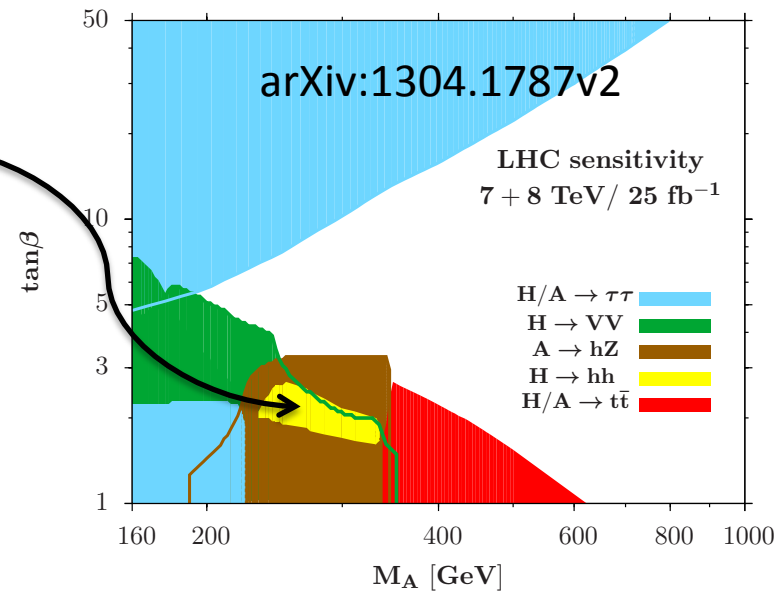
Theoretical Motivation

- A Randall Sundrum (RS1) **warped extra dimension** model with a radion of mass above 300 GeV can decay to two SM Higgs.
  - "Radion Phenomenology", Csaba Csaki et al: arXiv:0705.3844,
  - "Higgs-Radion interpretation of LHC data", Bohdan Grzadkowski et al: arXiv:1202.5017



- *X→HH(→4b)* final state can be sensitive only to X with mX < 350 GeV for low tanβ value in **MSSM** scenario.
  arXiv:1304.1787v2

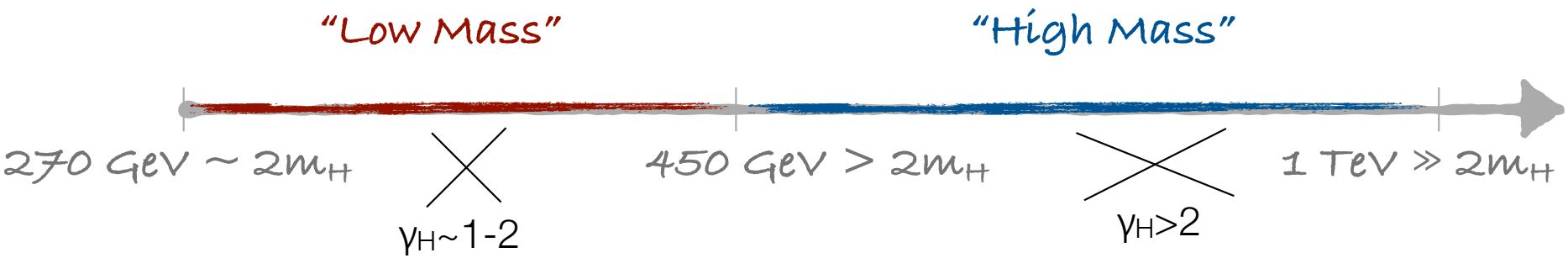
- **NMSSM** CP-even heavy Higgs H^0 decaying to h^0h^0 is interesting for m_{H0} between 300 and 500 GeV. Cross sections of ~ 60 fb assuming SM Br(H->bb).
  arxiv:1304.3670



Analysis Strategy

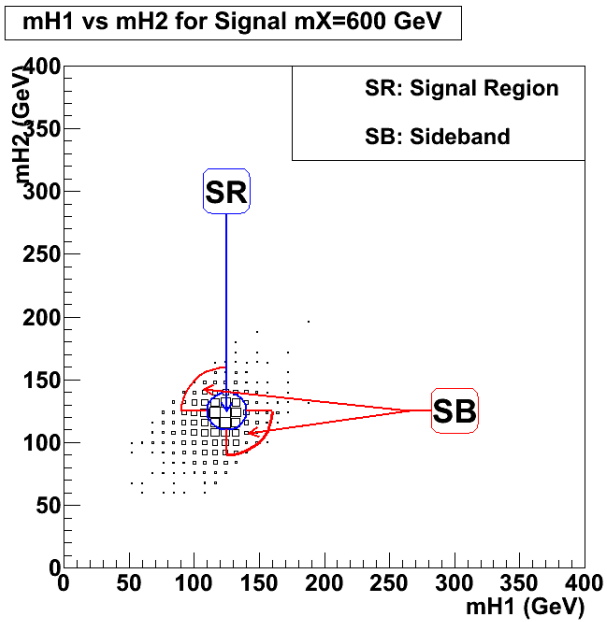
- The signal is a **narrow width resonance** X→H(bb)H(bb)

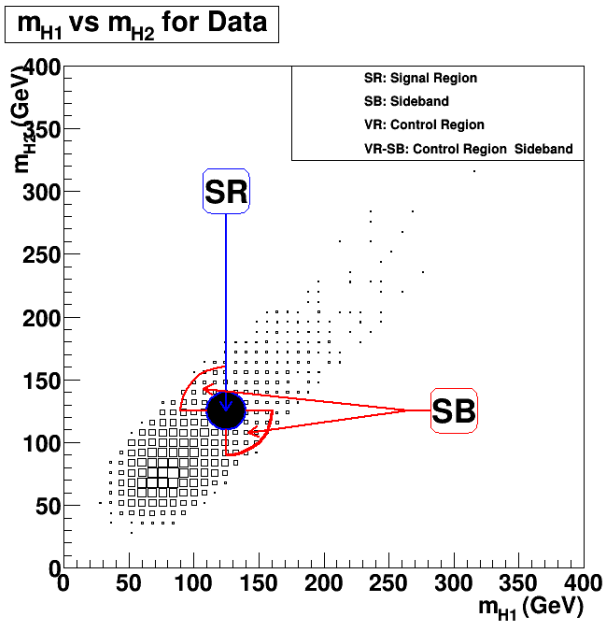
- This **blind search** covers mass hypotheses from **270 GeV to 1.1 TeV**
  - The kinematics involved change substantially over this range
  - **Two different event selections** are used in two kinematic regimes



- Main background through our selection criteria is multi-jet QCD production
  - **Data driven background estimation**

Analysis Strategy – Data Driven Background Estimation





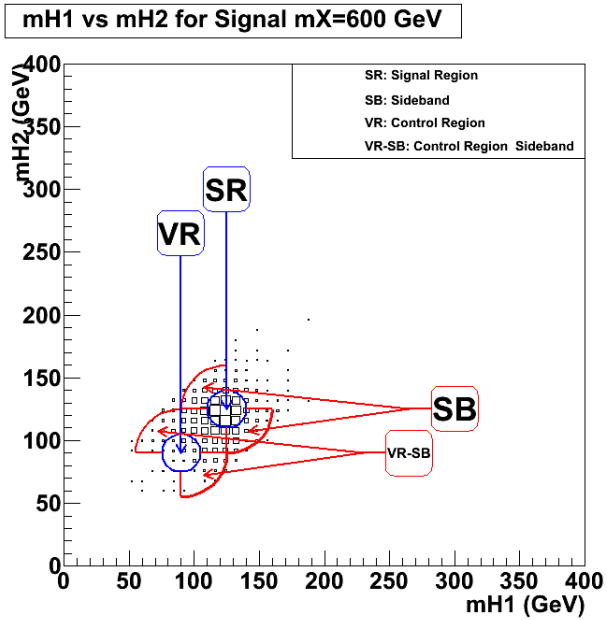
- **Select** 4 b-jets that reconstruct to two H(125) (described in event selection), randomize the selection of the two H and make a scatter plot of m_{H1} vs m_{H2}

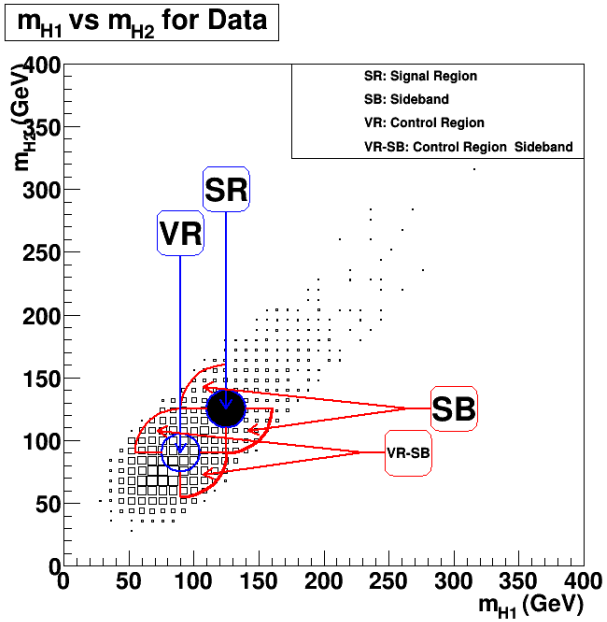
- **Define** the Signal Region ***SR*** centered around (125 GeV, 125 GeV) with a radius of 15 GeV. This is **BLINDED** for data

- **Define** a SideBand region ***SB*** around it with an inner radius of 15 GeV and an outer radius of 35 GeV, and only the top-left and bottom-right quadrants

… continued

Analysis Strategy… continued





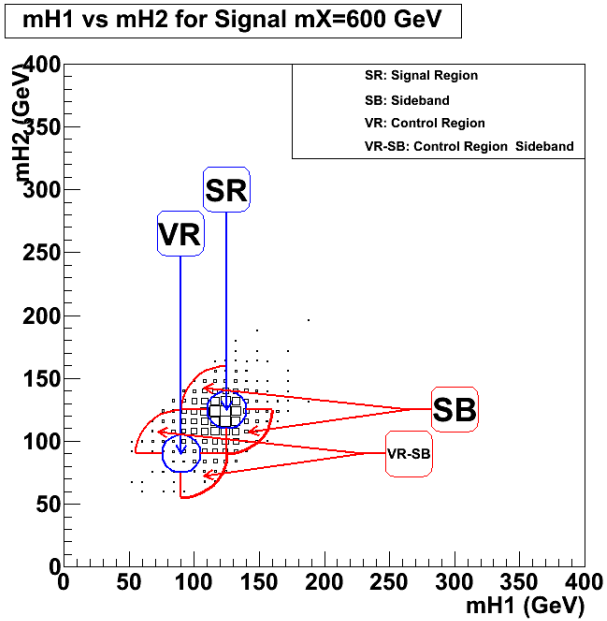
- **Empirical Claim**: The distribution of m_X within ***SR*** can be predicted, within quantifiable uncertainty, by the distribution of m_X within ***SB***

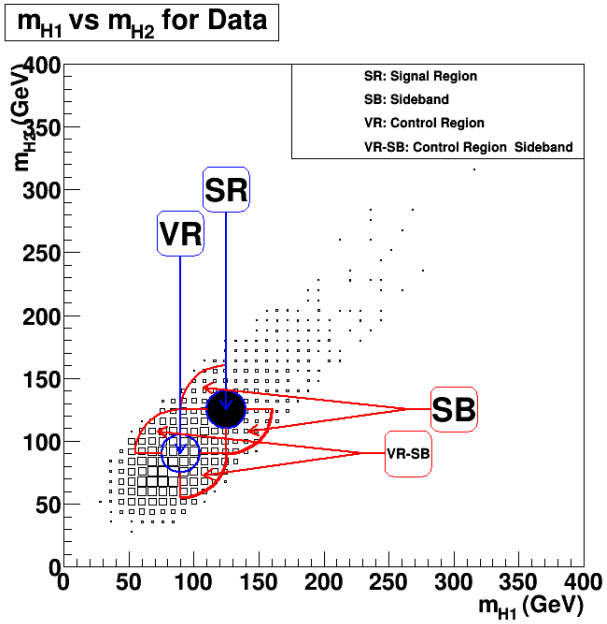
- **Check** this claim in regions ***VR*** and ***VR-SB*** centered around (90 GeV, 90 GeV)

- **Validate** this in an orthogonally tagged MMM(anti-M) Control Region

… continued

Analysis Strategy… continued





- **Propagate** JES, JER, b-tagging and trigger efficiency systematics through signal since background is data-driven

- **Expected Limit**: Use **functional fits** to the background and to the signal m_X, and mentioned systematics, to determine expected CL_S upper limits on production x-section of X using the **Higgs Combination Tool**

- **Observed Limit: Unblind** the ***SR*** and fit m_X to the predicted background and signal. Infer significance of observation if any

The Analysis

The Trigger

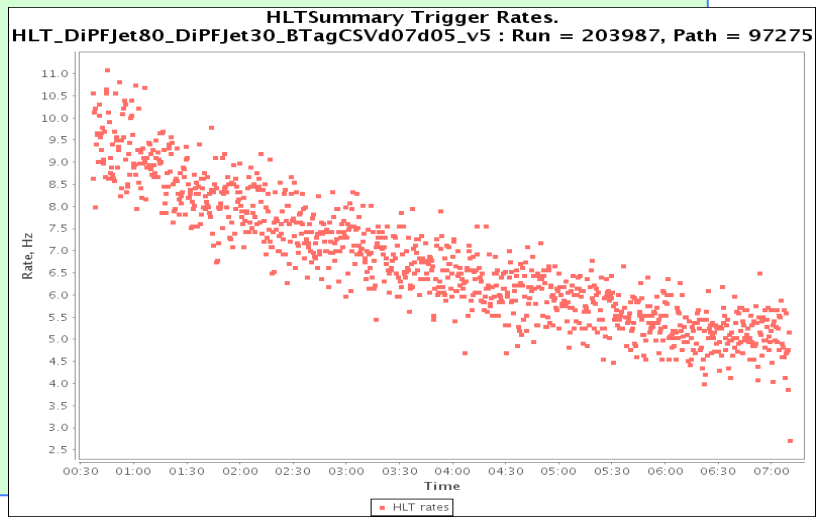
- Path used: **HLT_DiPFJet80_DiPFJet30_BTagCSVd07d05**

Been in place since /cdaq/physics/Run2012/7e33/v2.1/HLT/V9 (9 May, 2012), used for a Z(bb)H(bb) study

Structure

- L1_DoubleJetC56 OR L1_DoubleJetC64
- Reconstruct anti-KT 0.5 L1FastJetCorrected CaloJets
  - 2 Jets with |eta| < 2.6 and pT > 75 GeV
  - 4 Jets with |eta| < 2.6 and pT > 25 GeV
- Fast Primary Vertex Reconstruction
  - |z| < 25 cm, r < 2 cm
- CSV Computation
  - 1 CaloJet with pT > 20 must have CSV > 0.7
  - 2 CaloJets with pT > 20 must have CSV > 0.5
- PF Reconstruction Sequence
  - 2 PFJets with |eta| < 2.6 and pT > 80 GeV
  - 4 PFJets with |eta| < 2.6 and pT > 30 GeV

Rate ~ 6 Hz at 5e33



Datasets Used

- We use 2012 **8 TeV data** for our study.

- ***BJetPlusX*** datasets contain our trigger.

  - 2012B: (/BJetPlusX/Run2012B-13Jul2012-v1, 193834 - 196531): 4.428 /fb
  - 2012C:
    - (/BJetPlusX/Run2012C-24Aug2012-v2, 198049 - 198522): 495 /pb
    - (/BJetPlusX/Run2012C-PromptReco-v2, 198941 - 203002): 6.396 /fb
  - 2012D: (/BJetPlusX/Run2012D-PromptReco-v1, 203894 – end): 7.274 /fb

- Total Integrated Luminosity used is **18.60 /fb**

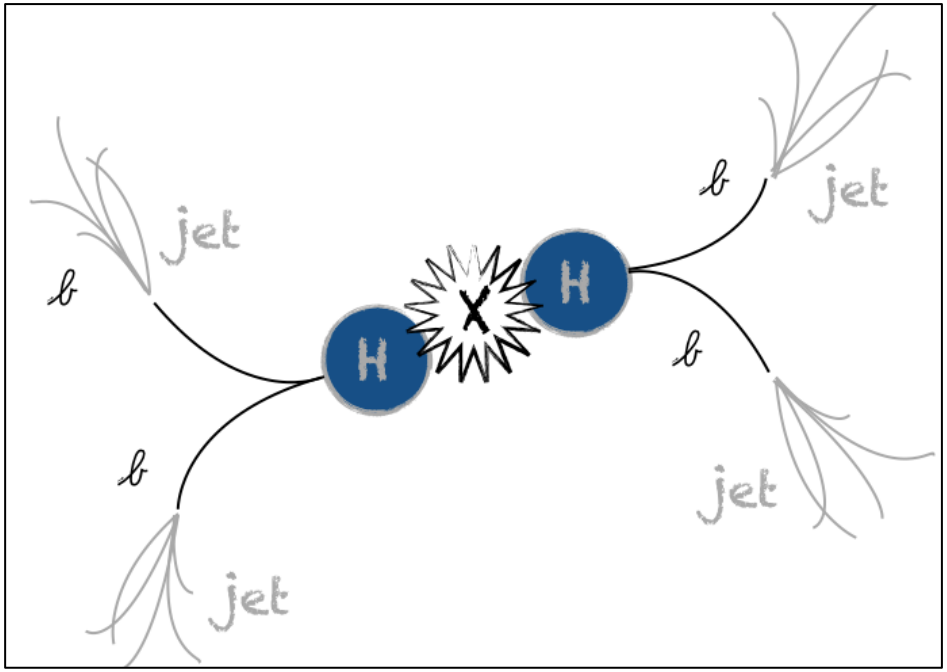
Signal Event Generation

- Events generated with Madgraph and showered with PYTHIA6
- We used the RS1 radion to stand in for *X* and generated

$$gg \rightarrow X \rightarrow HH$$
$$H \rightarrow b\bar{b}; H \rightarrow b\bar{b}$$

- for m_X = {270, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1100} TeV
- Generated 100,000 events per mass point
- Pileup Scenario: Summer 2012 In-Time
- Passed sample through HLT step containing trigger, and then reconstruction through CMSSW_5_3_3_patch2

Physics Objects



- **Jets**

  Jets are **Particle Flow** objects clustered with the **anti-k_T algorithm using R = 0.5**. Standard Jet Energy Corrections are applied (described in systematics).

  - **b-jet Identification**

    The **Combined Secondary Vertex** (CSV) algorithm is used.

Event Selection – High Mass (m_X > 450 GeV) Analysis

- Event contains at least 4 jets with $p_T > 40$ GeV, $|\eta| < 2.5$
- Loop over all di-jet pairs and converge on the two di-jet pairs with:
  - the least difference in masses between Higgs candidates.
  - ΔR between the jets within a pair $< \pi/2$
- Randomly swap the two Higgs candidates (based on parity of a digit)
- Require all 4 jets to have b-tag CSV > 0.679 (CSVM)
- (m_H, m_H) must fall within 15 GeV of (125 GeV, 125 GeV) (***SR***)

Event Selection – Low Mass (m_X < 450 GeV) Analysis

- Jets in the event are CSV-ordered
  - At least 4 jets with CSV > 0.244 (CSVL), pT > 30 GeV & |η| < 2.5

- Among selected jets, search for 2 H candidates such that:
  - Starting from the CSV-leading jet, we look for a second jet such that $m_{H1,2}$ is in the window [90, 160] GeV
  - Then, search for another pair with the same criterion

- Among the 4 jets identified as HH candidates, we require:
  - At least 2 jets with pT > 80 GeV
  - At least 2 jets with CSV > 0.898 (CSVT)

- All combinations of these 4 jets are inspected and the one with the minimal difference in masses of the Higgs candidates is chosen

- (m_H, m_H) must fall within 15 GeV of (125 GeV, 125 GeV) (***SR***)

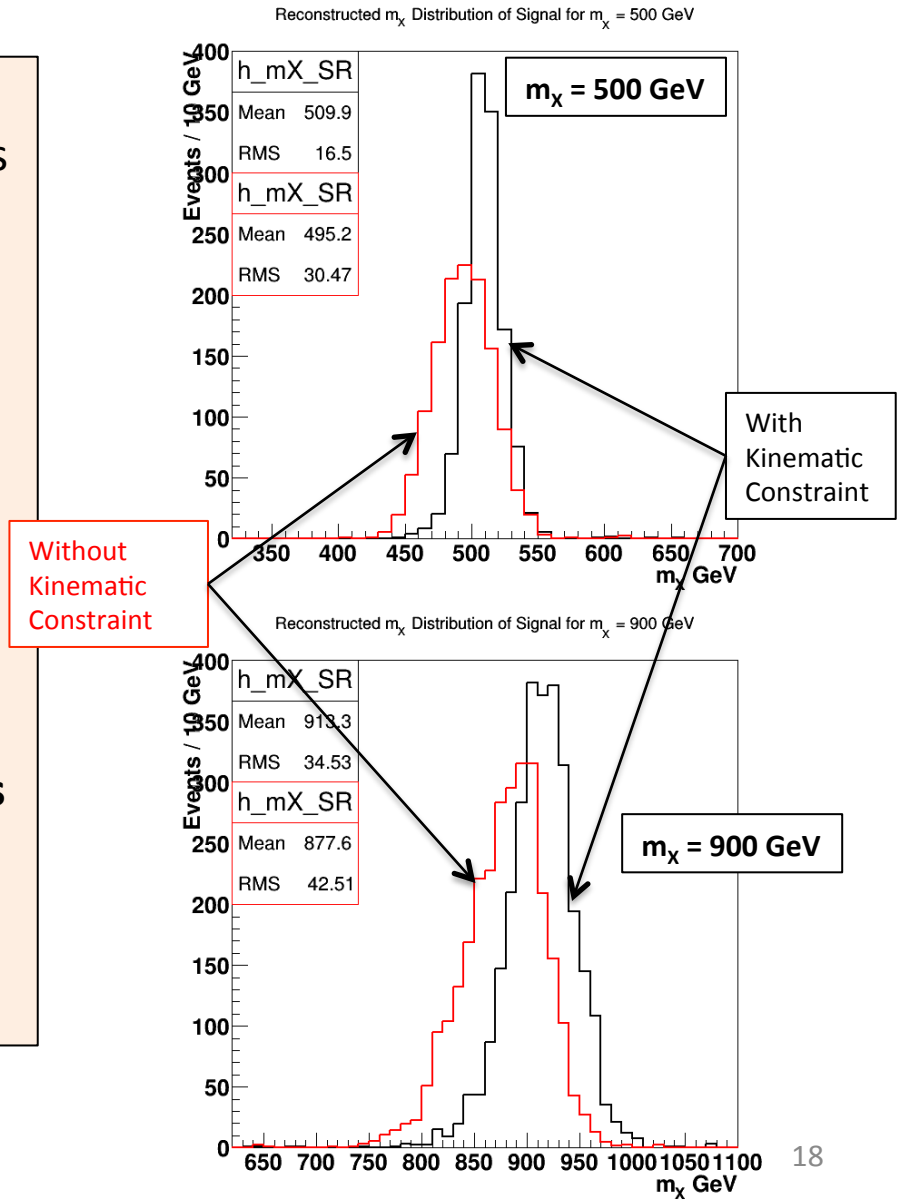
Kinematic Constraint

- **Kinematic constraint** applied on the b-jets reconstructing to each Higgs candidate so as to constrain Higgs mass to be 125 GeV. The jet 3-momenta are varied within known resolutions.

- Constraining is done using Lagrange Multipliers to converge to a least-square fit. This is implemented in CMSSW PhysicsTools/KinFitter.
  See **CMS NOTE 2006/023**

- Method improves **resolution of the di-Higgs resonance**, m_X, by factors of 20% to 40%

- Impact on expected upper limit varies between 10% to 18%



Event Selection – Efficiency Table

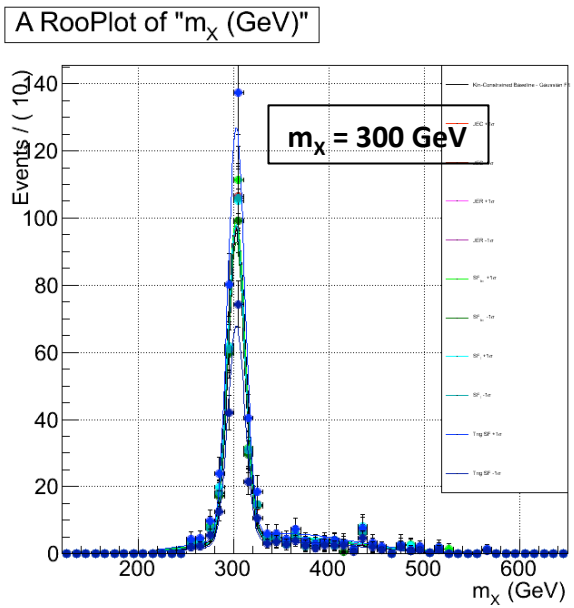
High Mass Event Selection Efficiency

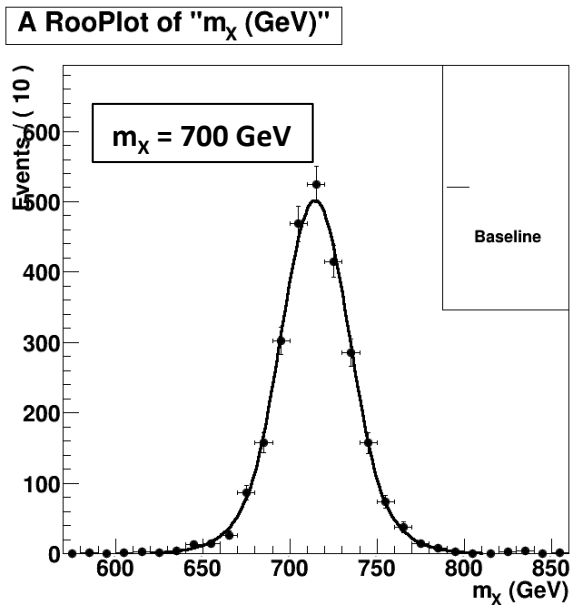
Selection Criterion	450 GeV	500 GeV	550 GeV	600 GeV	650 GeV	700 GeV	800 GeV	900 GeV	1000 GeV	1100 GeV
Trigger	43.86%	50.36%	54.65%	57.11%	59.81%	61.54%	63.70%	63.04%	58.89%	52.94%
nJets >3	74.23%	77.23%	79.86%	82.04%	83.57%	84.60%	86.68%	87.15%	85.51%	83.11%
Finding HH Candidate	51.50%	59.00%	63.37%	65.38%	68.19%	69.26%	70.02%	67.95%	60.55%	52.72%
b-tagging MMMM	13.05%	14.49%	15.41%	15.83%	16.19%	15.96%	15.56%	15.02%	13.44%	12.23%
Signal Region	37.79%	39.25%	38.94%	40.31%	42.80%	43.84%	45.04%	46.49%	44.52%	43.89%
Cumulative Eff.	0.823%	1.267%	1.624%	1.949%	2.319%	2.503%	2.686%	2.6%	1.807%	1.242%

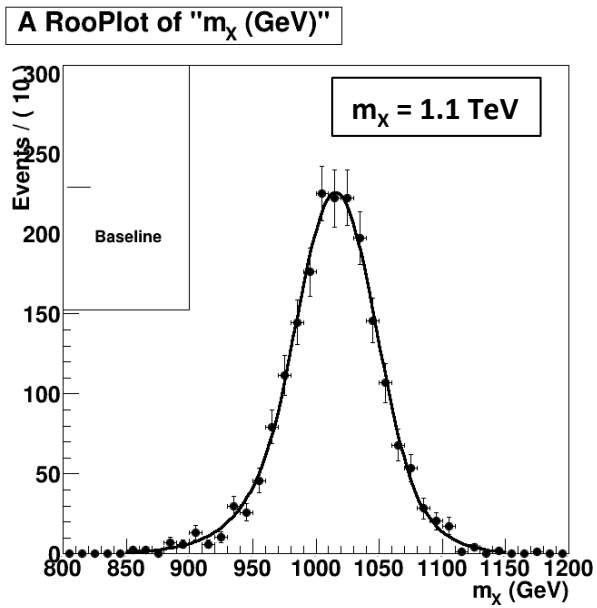
Low Mass Event Selection Efficiency

Selection Criterion	270 GeV	300 GeV	350 GeV	400 GeV	450 GeV	500 GeV	550 GeV
Trigger	7.75%	12.69%	23.21%	34.72%	43.86%	50.27%	54.65%
n *b*-jets	24,63%	26.36%	29.41%	32.05%	34.12%	35.69%	37.22 %
HH candidate	20.26%	29.81%	41.28%	48.81 %	51.13%	50.79%	51.72%
SR	36.48%	38.47%	37.40%	36.68 %	37.59%	37.67 %	38.41%
Cumulative Eff.	0.14%	0.39%	1.05%	1.99%	2.87%	3.42%	4.03%

Signal Modeling



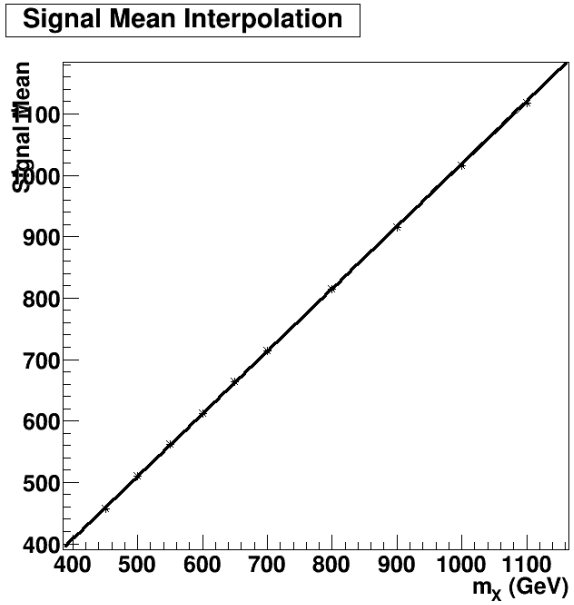


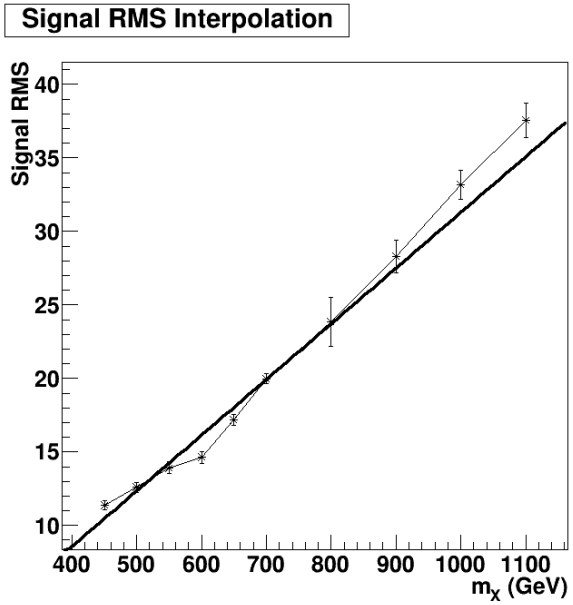


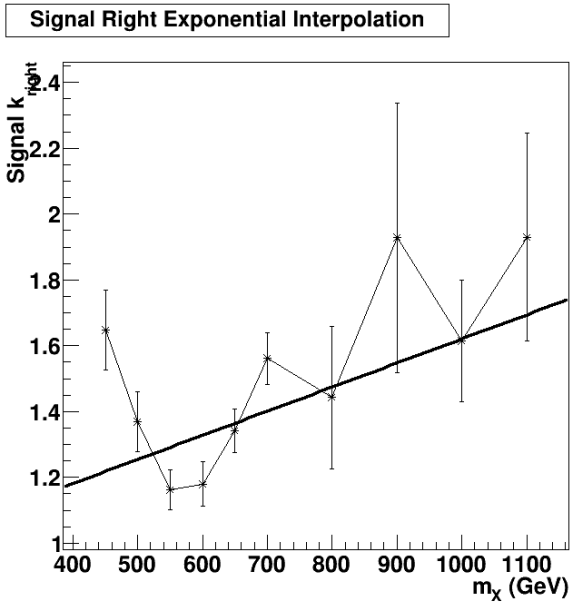
- The signal m_X distributions are modeled thus:
  - **High Mass Analysis**: An *Exp-Gauss-Exp* function invented for the High Mass Analysis, which is a Gaussian core analytically continued on both sides to exponential tails. 4 parameters.
  - **Low Mass Analysis**: Two Gaussian functions, one for the core and one for the combinatorial background. 5 parameters.

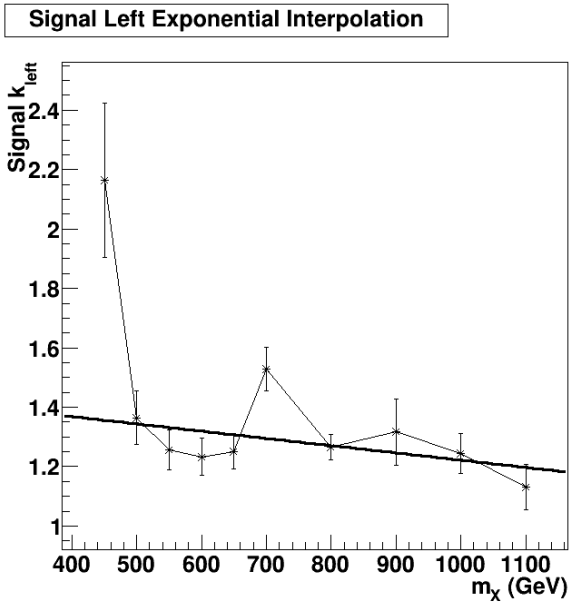
- Fit details:
  - LMA: http://www.pi.infn.it/~vernieri/TLTL8030/SignalSystematics.html
  - HMA:http://home.fnal.gov/~souvik/FullSim/focus/SignalSystematics/SignalSystematics.html

Interpolation for Fit Parameters of Intermediate Mass Points High Mass Analysis





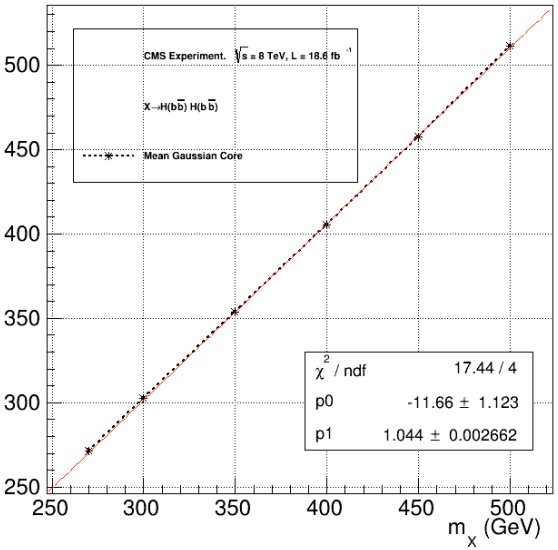




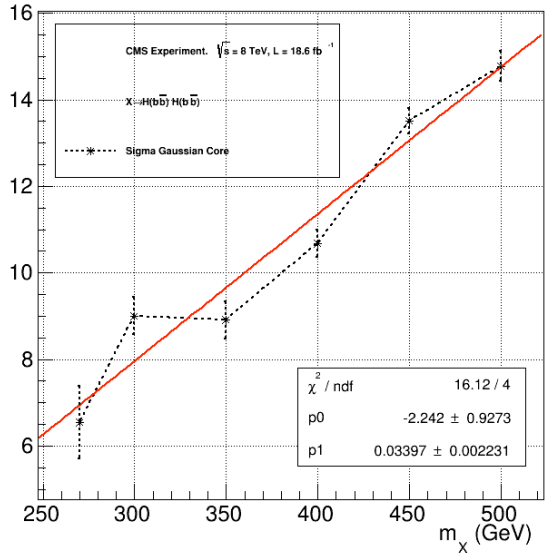
- Will be used to fit signal hypotheses for intermediary mass points when data is unblinded.

Interpolation for Fit Parameters of Intermediate Mass Points Low Mass Analysis

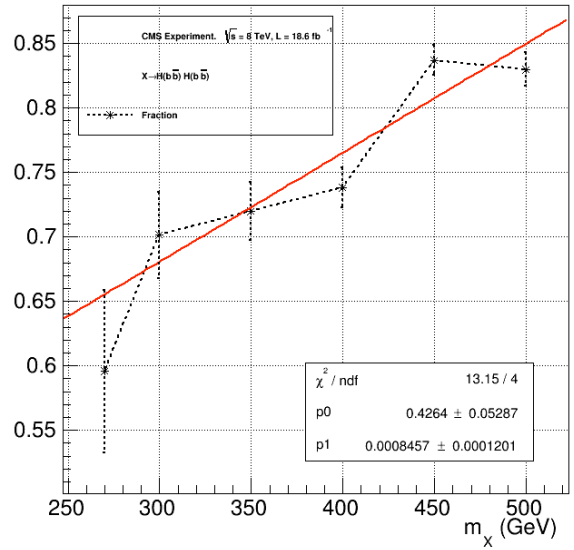
mean gaussian core



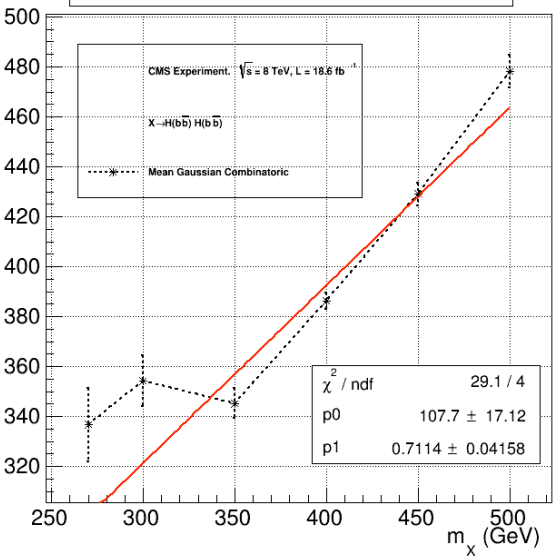
sigma gaussian core



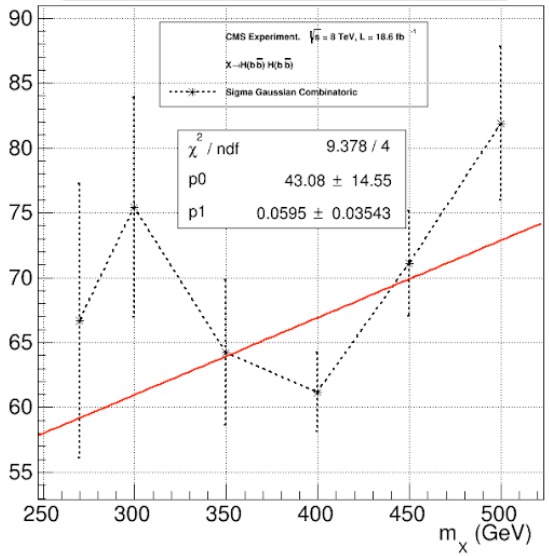
frac gaussian core



Mean Gaussian for combinatoric



Sigma Gaussian for combinatoric



- Will be used to fit signal hypotheses for intermediary mass points when data is unblinded.

Systematic Uncertainties – From JES, JER

Since the **background is modeled entirely from data**, systematic uncertainties arising from jet energy scale & resolution, b-tagging and trigger efficiency **affect only the signal modeling**. These are briefly discussed in the following slides.

Jet Energy Scale

Jet p_T moved by $\pm 1\sigma$ of total JES-uncertainty arising from sources listed here:
https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECUncertaintySources

Jet Energy Resolution

Smearing of Jet p_T w.r.t gen-jet p_T increased and decreased by $\pm 1\sigma$ of the JER uncertainties listed here:
https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution

Systematic Uncertainties – From b-tagging

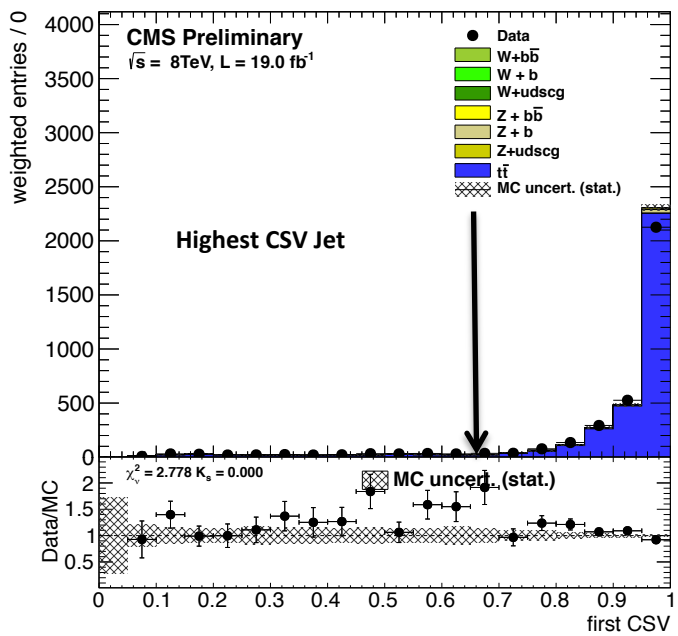
b-tagging Scale Factor

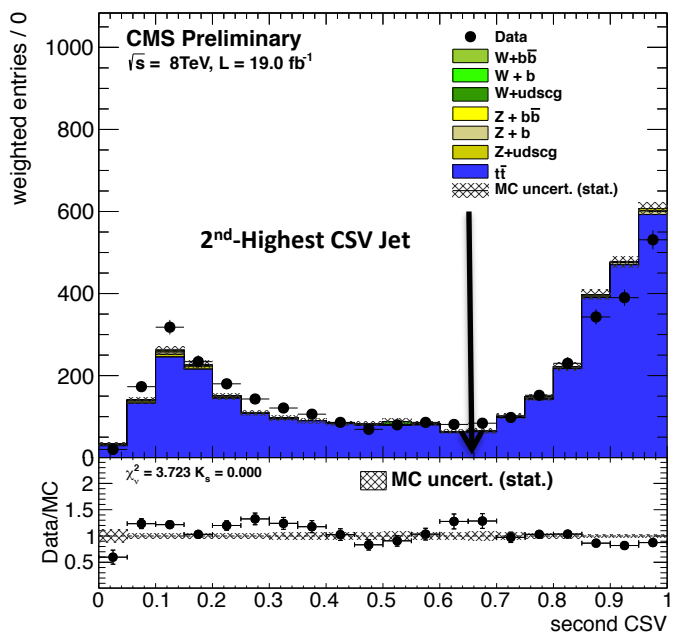
- Since we're on the CSV working points, Loose, Medium and Tight, we use **CSV-reshaping** to account for the b-tagging scale factor between data and Monte Carlo. Reshaping also returns equivalent CSV distributions for ±1σ variations of the SF. Used:
  - One **nominal** CSV_{equiv} distribution that folds in SF based on jet flavor, jet p_T and $|\eta|$
  - Two CSV_{equiv} distributions derived from pushing the $\mathbf{SF_{bc}}$ scale factors by $\pm 1\sigma$
  - Two CSV_{equiv} distributions derived from pushing the $\mathbf{SF_{light}}$ scale factors by $\pm 1\sigma$

Hard work was already done by Hbb group!

Systematic Uncertainties – From the Trigger

- **Simulation of the HLT in signal MC** is taken as a baseline, and corrections to it by studying data/MC in a *tt-bar* control region are considered systematic uncertainties
- The SingleMu dataset was used to characterize the efficiency of our trigger, **HLT_DiPFJet80_DiPFJet30_BTagCSVd07d05** in order to use *tt-bar* with at least two b-jets
- To get this control region, the *HH* candidates were found using our selection and at least one muon with $p_T > 40$ GeV was required





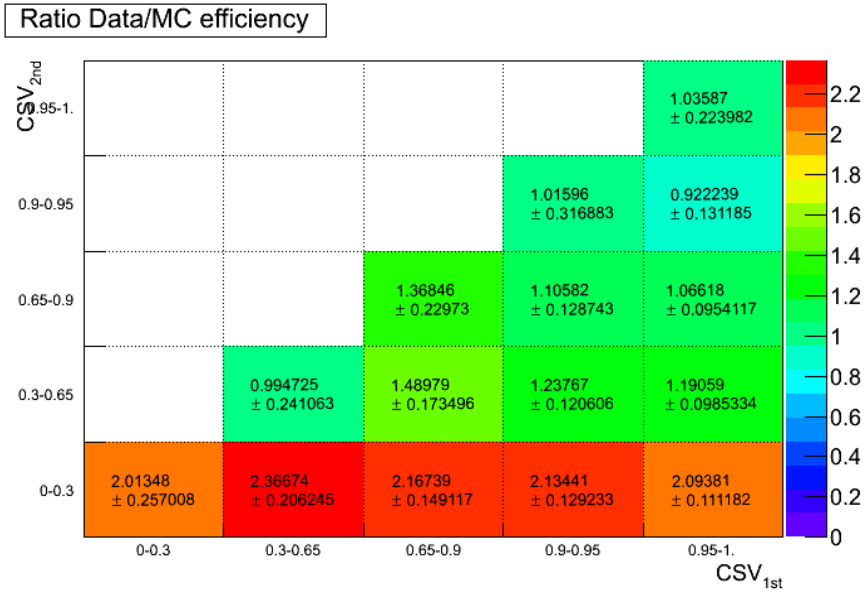
Systematic Uncertainties – From the Trigger

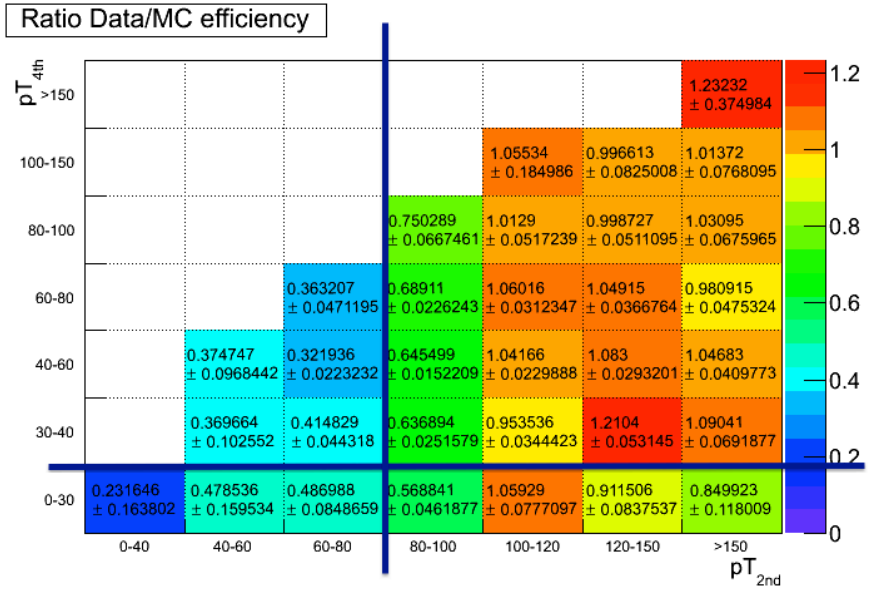
- Uncertainty on the following measure of efficiency was computed:

  $$\varepsilon = P(HLT \mid offlineCuts\ AND\ SingleMu)$$

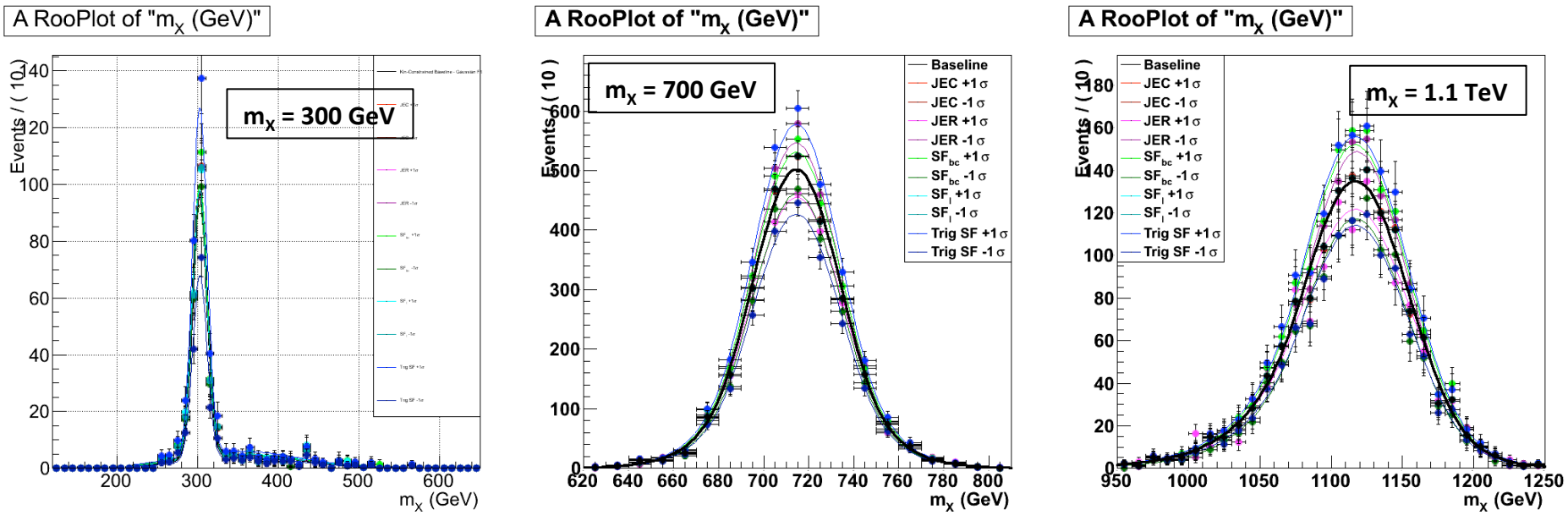
  and plotted as a function of:
  - the highest and second highest CSV for both MC and data
  - the 2nd and 4th highest jet p_T

- ±1σ variations of each these characterizations were folded into the signal line-shape in order to compute the systematic uncertainties from the trigger





Signal Modeling with Systematics



- ±1σ variations of each source of systematics are also fitted to the same functions.
- The variations in **normalization** and **fit parameters** are used to compute limits
- Fit details:
  - LMA: http://www.pi.infn.it/~vernieri/TLTL8030/SignalSystematics.html
  - HMA: http://home.fnal.gov/~souvik/FullSim/SignalSystematics/SignalSystematics.html

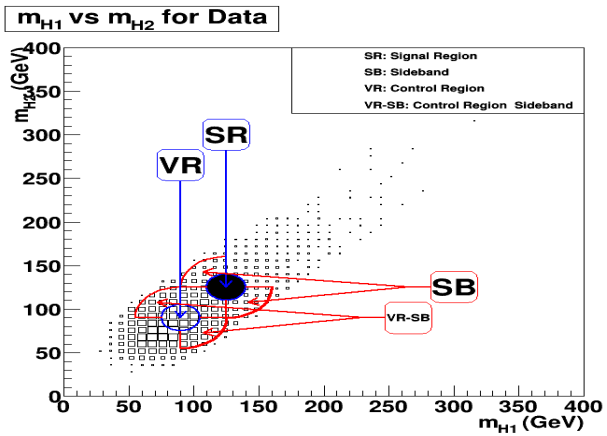
Signal Modeling with Systematics

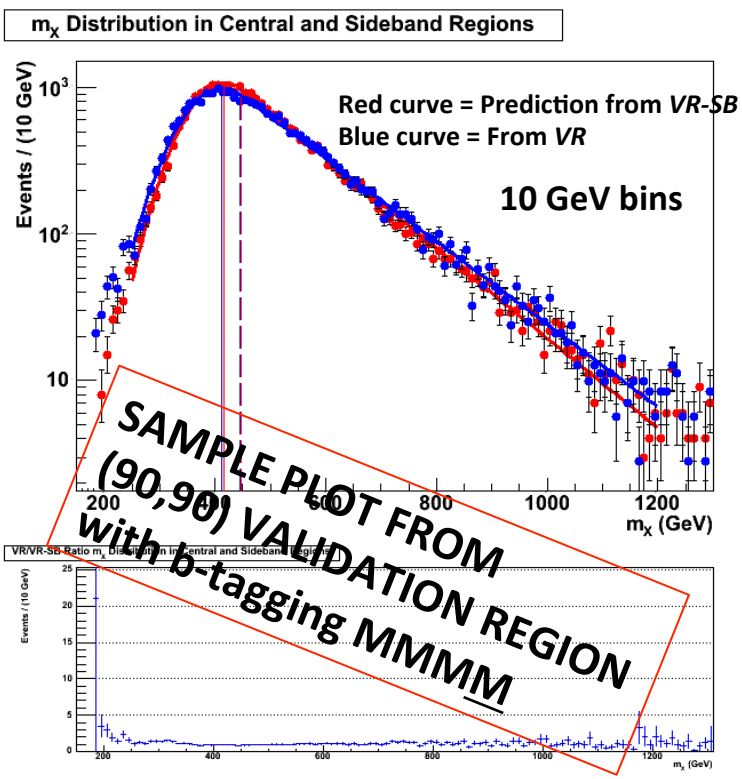
- Systematic uncertainties affect both the normalization of the signal shape and the fit parameters. Effects on the normalization are tabulated below

Source of Systematic Uncertainty	Range of Signal Efficiency Normalization Uncertainty High Mass Analysis	Range of Signal Efficiency Normalization Uncertainty Low Mass Analysis
Luminosity	4.4%	4.4%
Jet Energy Scale	0 – 0.2%	0.1% – 0.5%
Jet Energy Resolution	5.5% – 6.7%	2.4% – 7.0%
b-tagging SF for *bc* flavors	5.2% – 13.0%	1.8% – 5.7%
b-tagging SF for light flavors	0 – 0.2%	0.8% – 2.3%
Trigger SF	14.7% – 17.0 %	21.5% – 30.3%

- Details of how systematics affect signal normalization and fit parameters here:
  Low Mass Analysis: http://www.pi.infn.it/~vernieri/TLTL8030/SignalSystematics.html
  High Mass Analysis: http://home.fnal.gov/~souvik/FullSim/SignalSystematics/SignalSystematics.html

Modeling the Background Distribution – High Mass Analysis





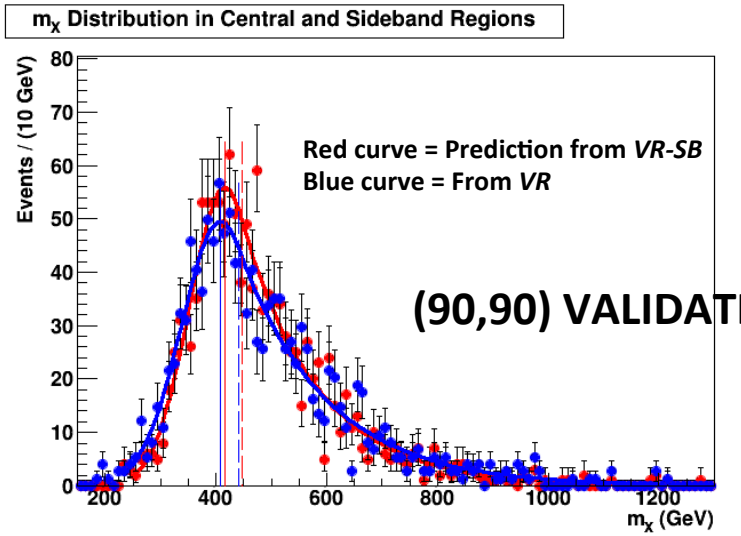
- Crystal Ball function was too unstable in the exponent. CDF 2009 function for di-jet resonances was also unstable.

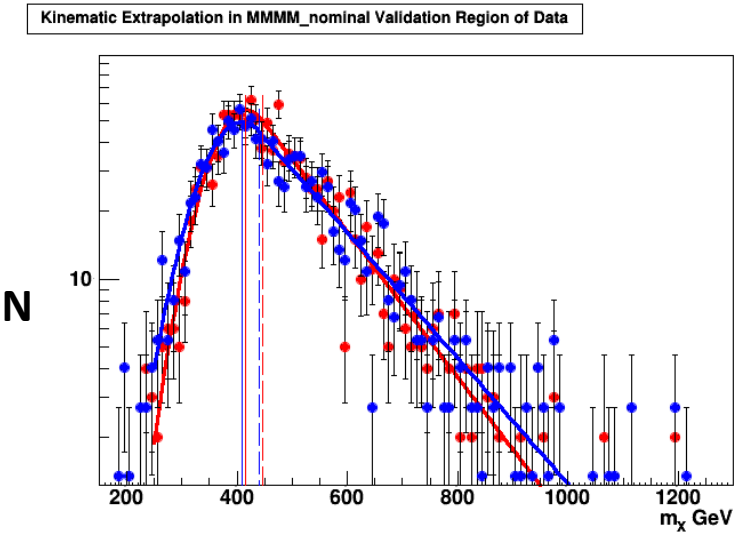
- Our function has Gaussian Core & Exponential Tail

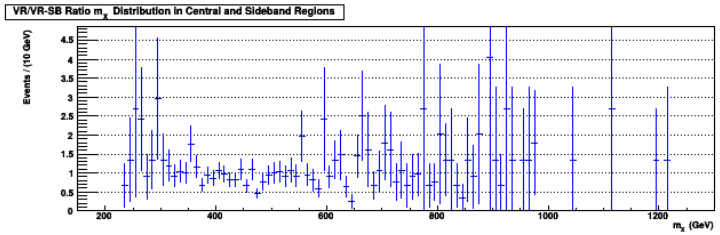
$$y(x;\bar{x},\sigma,k) = A\left(e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}, \quad \text{for} \quad \frac{x-\bar{x}}{\sigma} < k\right)$$
$$\text{and} \quad A\left(e^{\frac{k^2}{2}-k\left(\frac{x-\bar{x}}{\sigma}\right)}, \quad \text{for} \quad \frac{x-\bar{x}}{\sigma} \geq k\right)$$

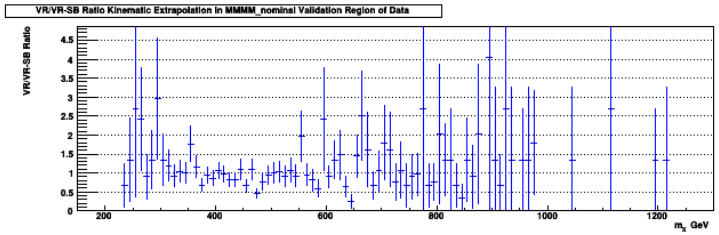
- Just three parameters and we know what they do:
  - P_0: Mean of the Gaussian core
  - P_1: Width of the Gaussian core
  - P_2: Exponent of the high tail, also the number of std dev at which Gaussian turns into exponential

Data in the Validation Region









Validation Region Sideband Parameters

P_0^{VRSB} = 416.2 GeV; δP_0^{VRSB} = 1.2 GeV

P_1^{VRSB} = 66.7 GeV; δP_1^{VRSB} = 1.0 GeV

P_2^{VRSB} = 0.471; δP_2^{VRSB} = 0.009

chi^2/ndof = 0.78

Validation Region Parameters

P_0^{VR} = 412.1 GeV; δP_0^{VR} = 1.2 GeV

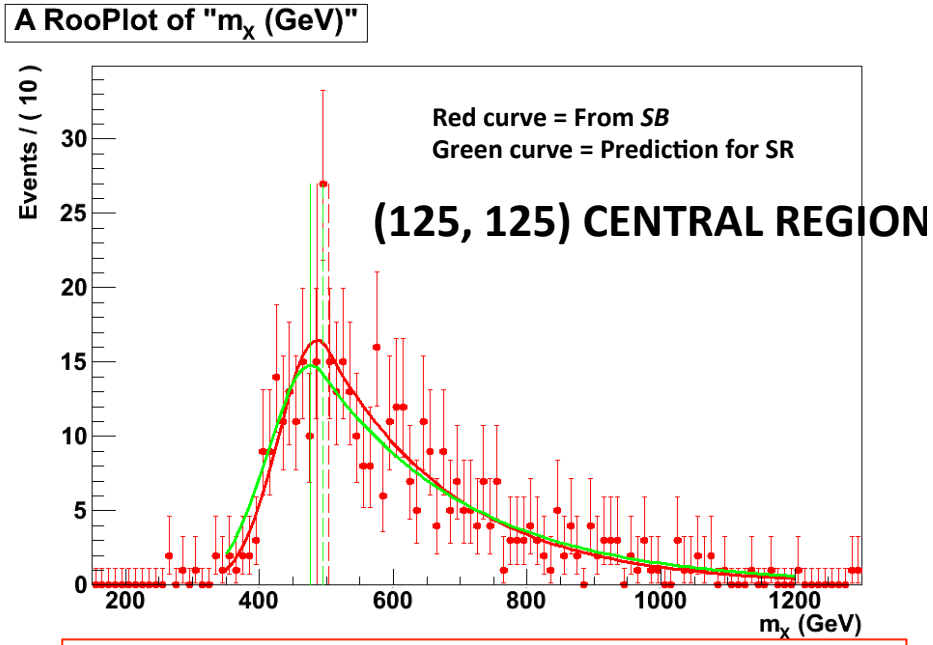
P_1^{VR} = 72.1 GeV; δP_1^{VR} = 1.0 GeV

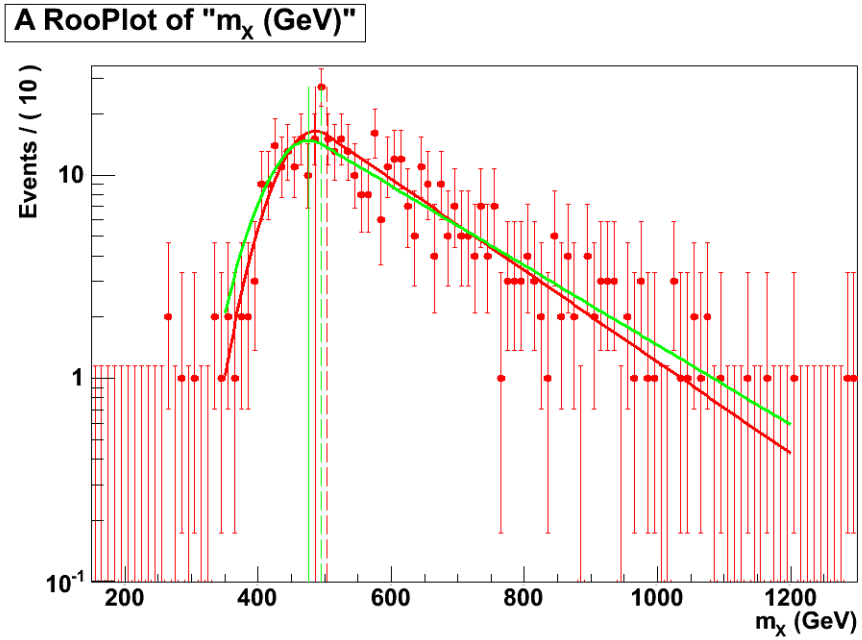
P_2^{VR} = 0.468; δP_2^{VR} = 0.009

chi^2/ndof = 0.98

The Gauss-Exp function that works for the VR-SB region also fits data in the VR region.

Data in the Signal-Sideband Region





Signal Region Sideband Parameters

P_0^{SB} = 486.0 GeV; δP_0^{SB} = 12.0 GeV
P_1^{SB} = 57.8 GeV; δP_1^{SB} = 9.0 GeV
P_2^{SB} = 0.299; δP_2^{SB} = 0.057
chi^2/ndof = 0.44

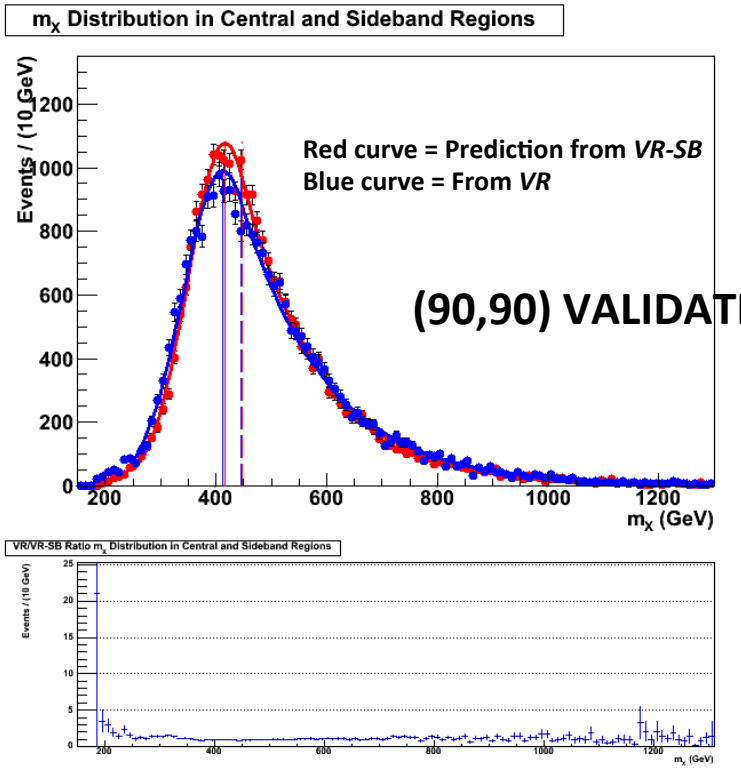
Predicted Signal Region Parameters

P_0^{SR} = 476.6 GeV; δP_0^{SR} = 14.9 GeV
P_1^{SR} = 64.0 GeV; δP_1^{SR} = 11.6 GeV
P_2^{SR} = 0.288; δP_2^{SR} = 0.066

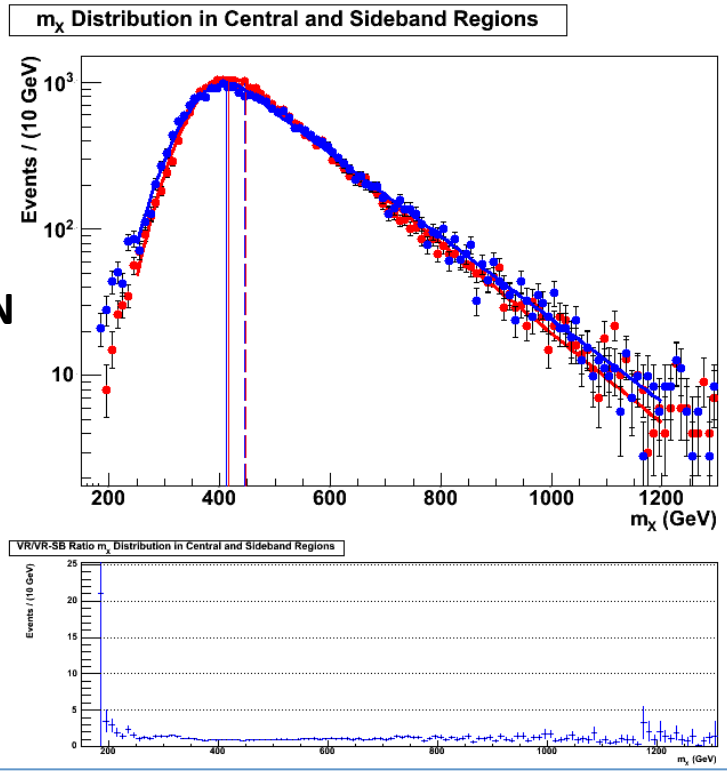
- The **Gauss-Exp form fits** the m_X distribution in the ***SB*** region.
- We posit it will also fit the distribution in the ***SR*** region.

- We **test this position in an orthogonally b-tagged control region** with one b-tag reversed

Data in Orthogonally b-tagged MMMM Control Region



(90,90) VALIDATION REGION



Validation Region Sideband Parameters

P_0^{VRSB} = 416.2 GeV; δP_0^{VRSB} = 1.2 GeV

P_1^{VRSB} = 66.7 GeV; δP_1^{VRSB} = 1.0 GeV

P_2^{VRSB} = 0.471; δP_2^{VRSB} = 0.009

chi^2/ndof = 1.43

Validation Region Parameters

P_0^{VR} = 412.1 GeV; δP_0^{VR} = 1.2 GeV

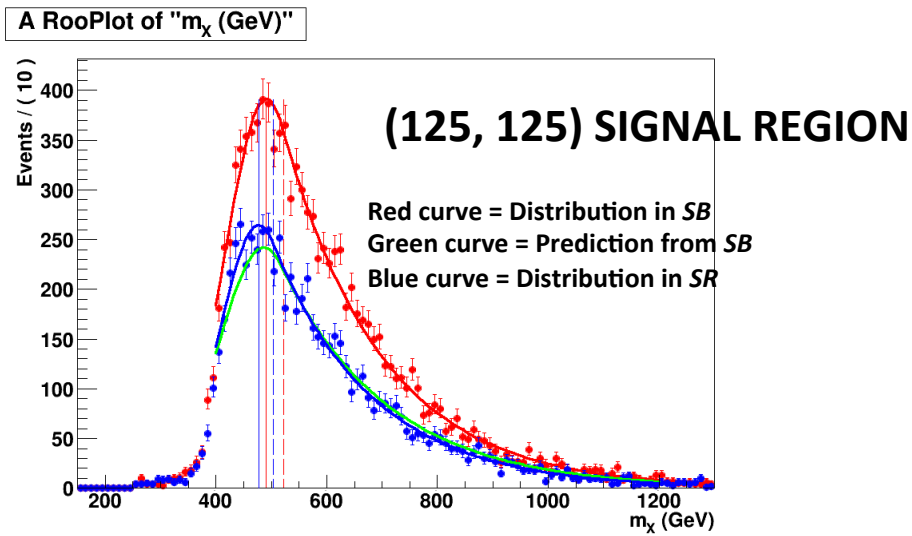
P_1^{VR} = 72.1 GeV; δP_1^{VR} = 1.0 GeV

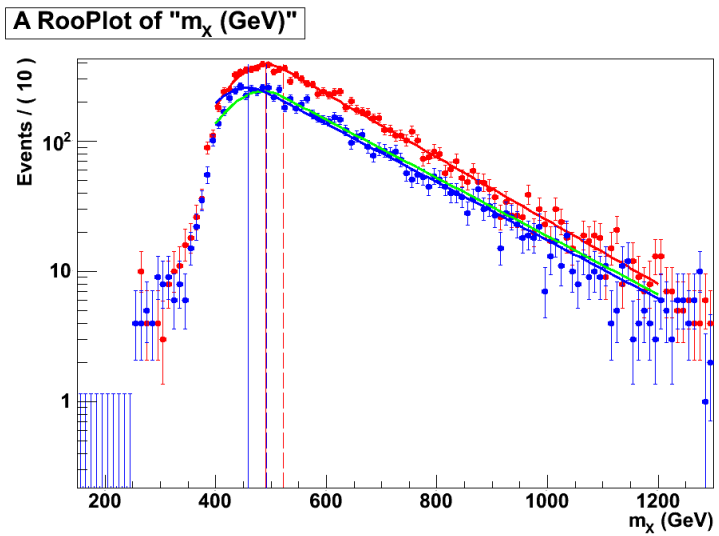
P_2^{VR} = 0.468; δP_2^{VR} = 0.009

chi^2/ndof = 1.59

The Gauss-Exp function that works for the VR-SB region again fits data in the VR region.
The tail is an exponential. Polynomial forms & Landau curves don't do it justice.

Data in Orthogonally b-tagged MMMM Fake Signal Region





Fake Signal Region Sideband Parameters

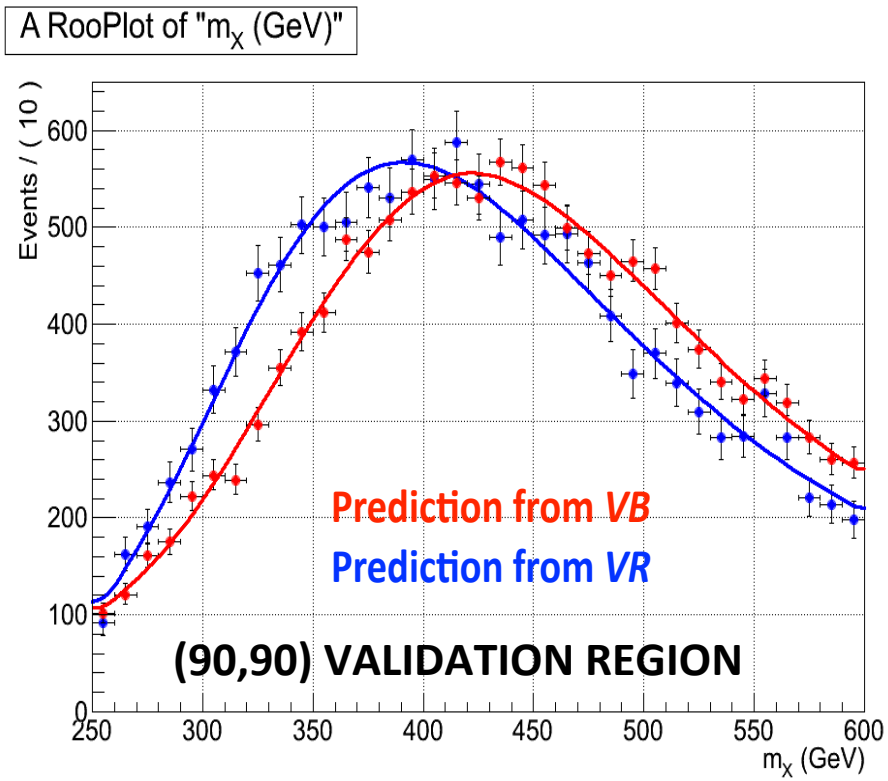
P_0^{SB} = 490.9 GeV; δP_0^{SB} = 1.2 GeV
P_1^{SB} = 74.1 GeV; δP_1^{SB} = 3.6 GeV
P_2^{SB} = 0.415; δP_2^{SB} = 0.021
chi^2/ndof = 1.12

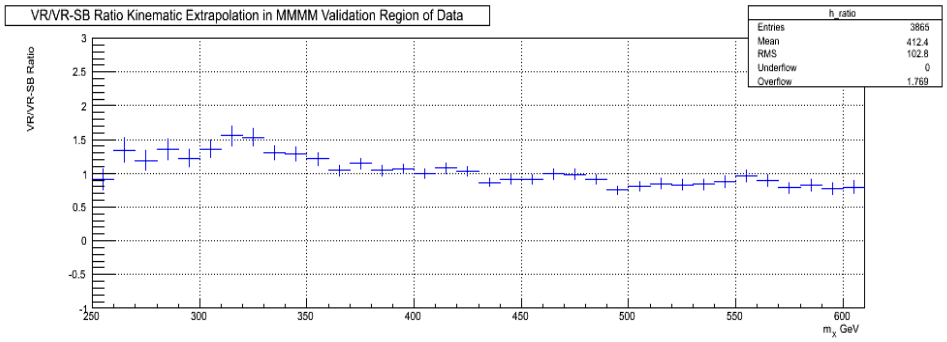
Fake Signal Region Fit Parameters

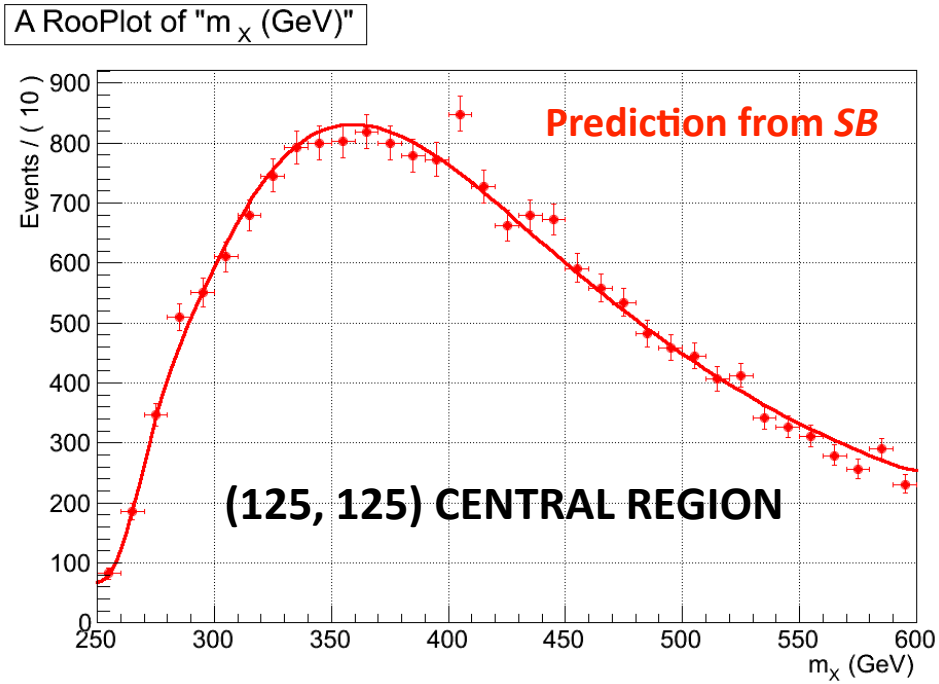
P_0^{VR} = 477.0 GeV; δP_0^{VR} = 6.5 GeV
P_1^{VR} = 69.3 GeV; δP_1^{VR} = 12.5 GeV
P_2^{VR} = 0.384; δP_2^{VR} = 0.075
chi^2/ndof = 1.07

- The **Gauss-Exp function fits the Fake Signal Region** of the orthogonally b-tagged region, starting from fit parameters obtained from the SB region
- **This function can be used to model the background on unblinding**

Modeling the Background Distribution – Low Mass Analysis

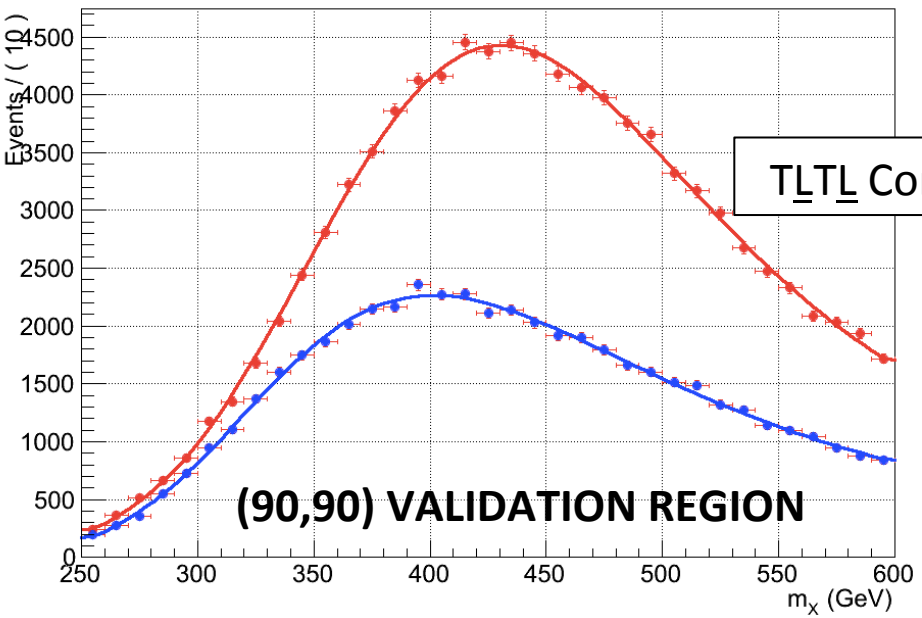


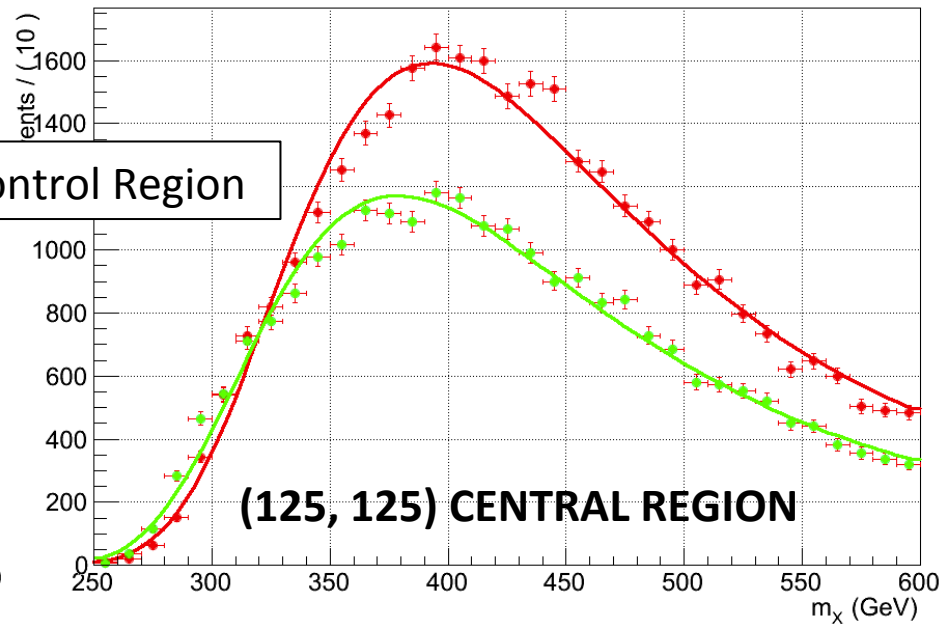




- A Landau function convoluted with a Gaussian is used to model the background. 4 parameters
- We test this not only in the VR, VR-SB and SB regions, but also in an orthogonally b-tagged $T\bar{L}T\bar{L}$ region

Modeling the Background Distribution – Low Mass Analysis





Prediction from *VB, X/ndof = 0.99*
Prediction from *VR, X/ndof = 0.89*

Prediction from *SB, X/ndof= 3.23*
Prediction from *SR, X/ndof= 2.19*

- We reverse the requirement on the CSV for the 2 jets which are not tagged as CSV-tight. Jets are ordered by p_T for this control region
- The Landau-Gaussian is seen to **fit the fake signal region distribution** well. Can be used when data is unblinded.

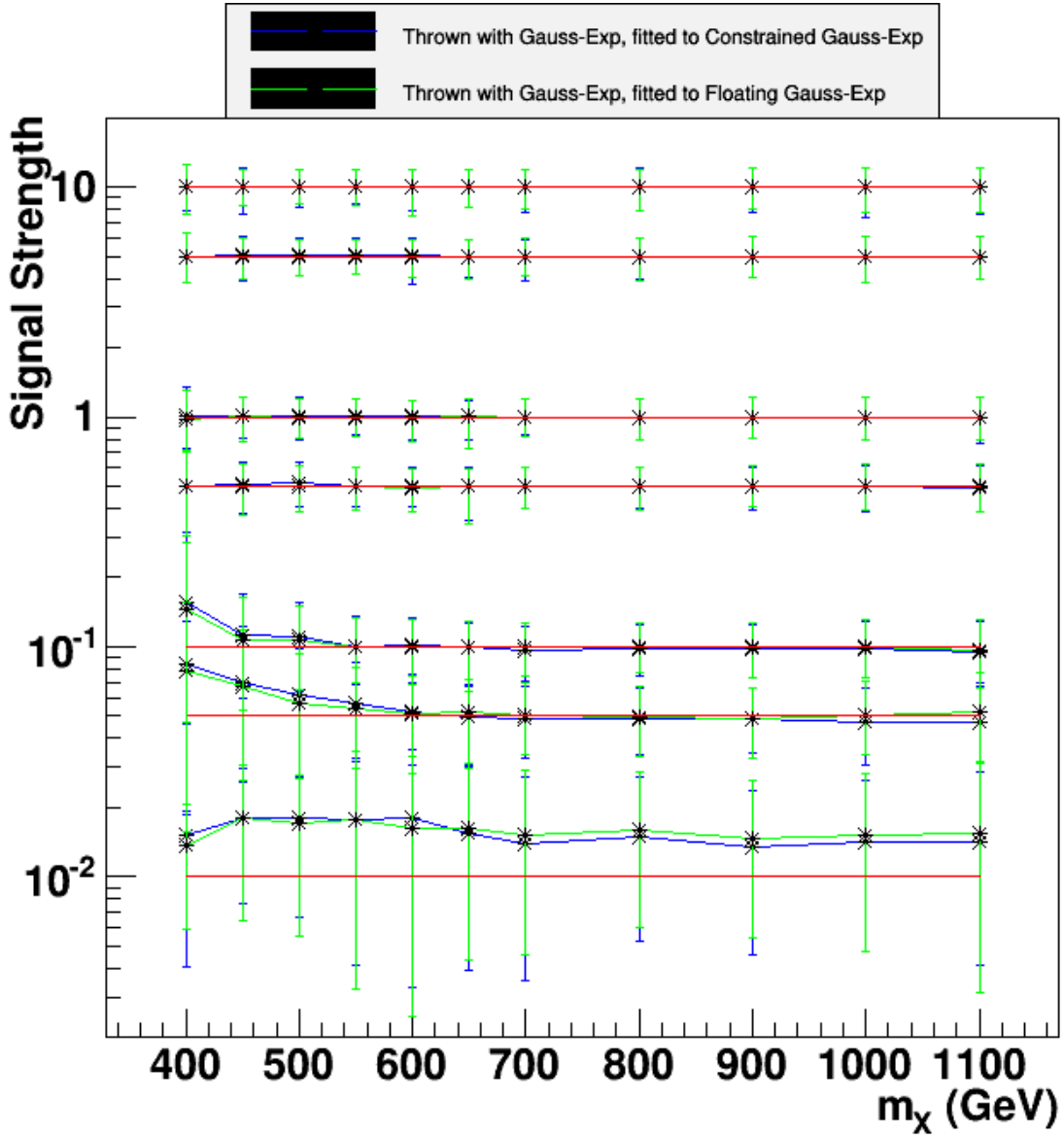
Results

Signal Strength Bias Studies

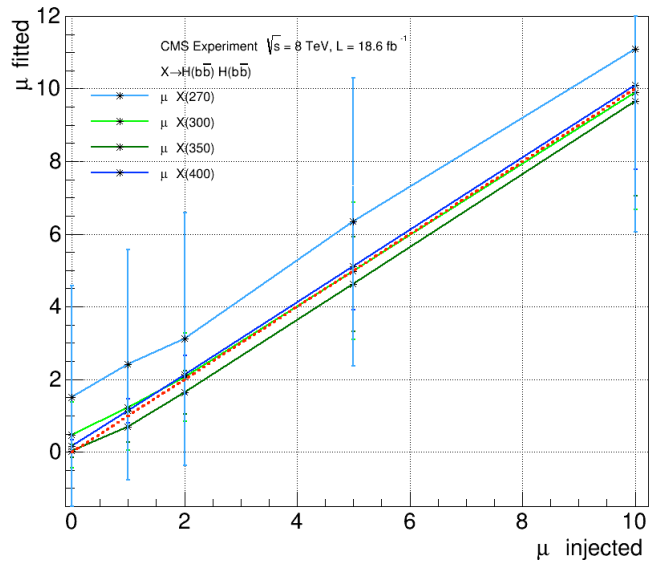
- Given our signal and background model, we test the systematic bias in the signal strength extracted with the Higgs Combination Tool for various cross sections of injected signal
- 200 – 500 toys are thrown from a background & signal model, for different values of signal strength, for each mass point studied.
- The toys in the m_X distribution are fitted with the same or a different functional form for the background and a signal whose mean is left to float
- The reconstructed signal strength is a measure of the bias inherent in this signal extraction procedure

Signal Strength Bias for Various Functional Forms High Mass Analysis

- 200 toys were thrown for each point, with the Gauss-Exp fit to the data in the SB region, normalized to number of events in the SR region.
- Signal strengths of 0.01 pb, 0.05 pb, 0.1 pb, 0.5 pb, 1 pb, 5 pb and 10 pb were injected
- The toys were fit to a Gauss-Exp for background with floating parameters, and also to a Gauss-Exp whose parameter ranges are predicted by extrapolation. Both show similar bias.
- The **background + signal fit is robust**.

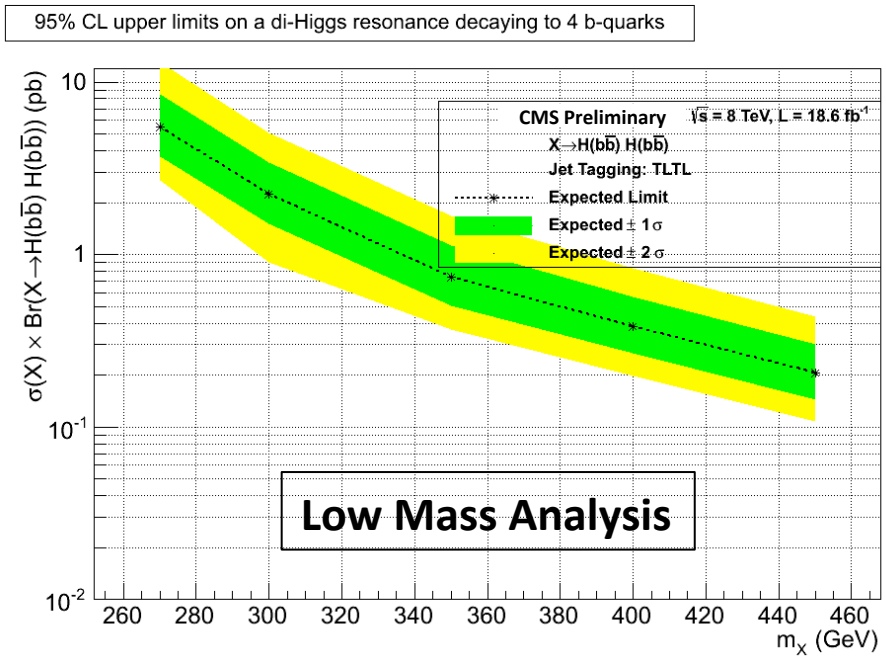


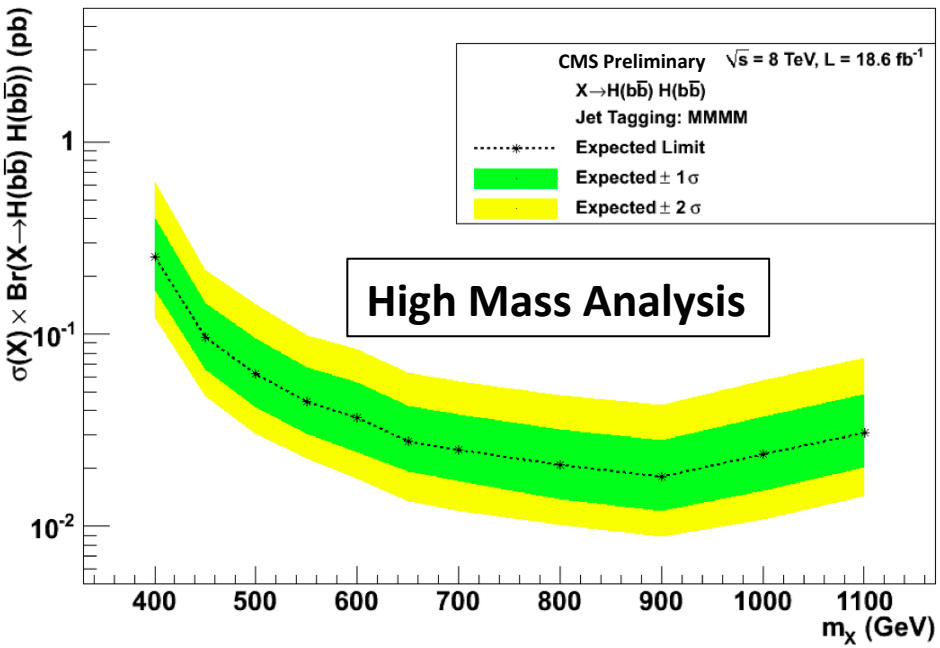
Signal Strength Bias for Various Functional Forms
Low Mass Analysis



- 500 toys were thrown for each point. Signal strengths of 0 pb, 1 pb, 2 pb, 5 pb and 10 pb were injected
- The toys were generated with a 6th order Bernstein polynomial as the background, and fitted to the Landau-Gauss.
- The **background + signal fit is robust**.

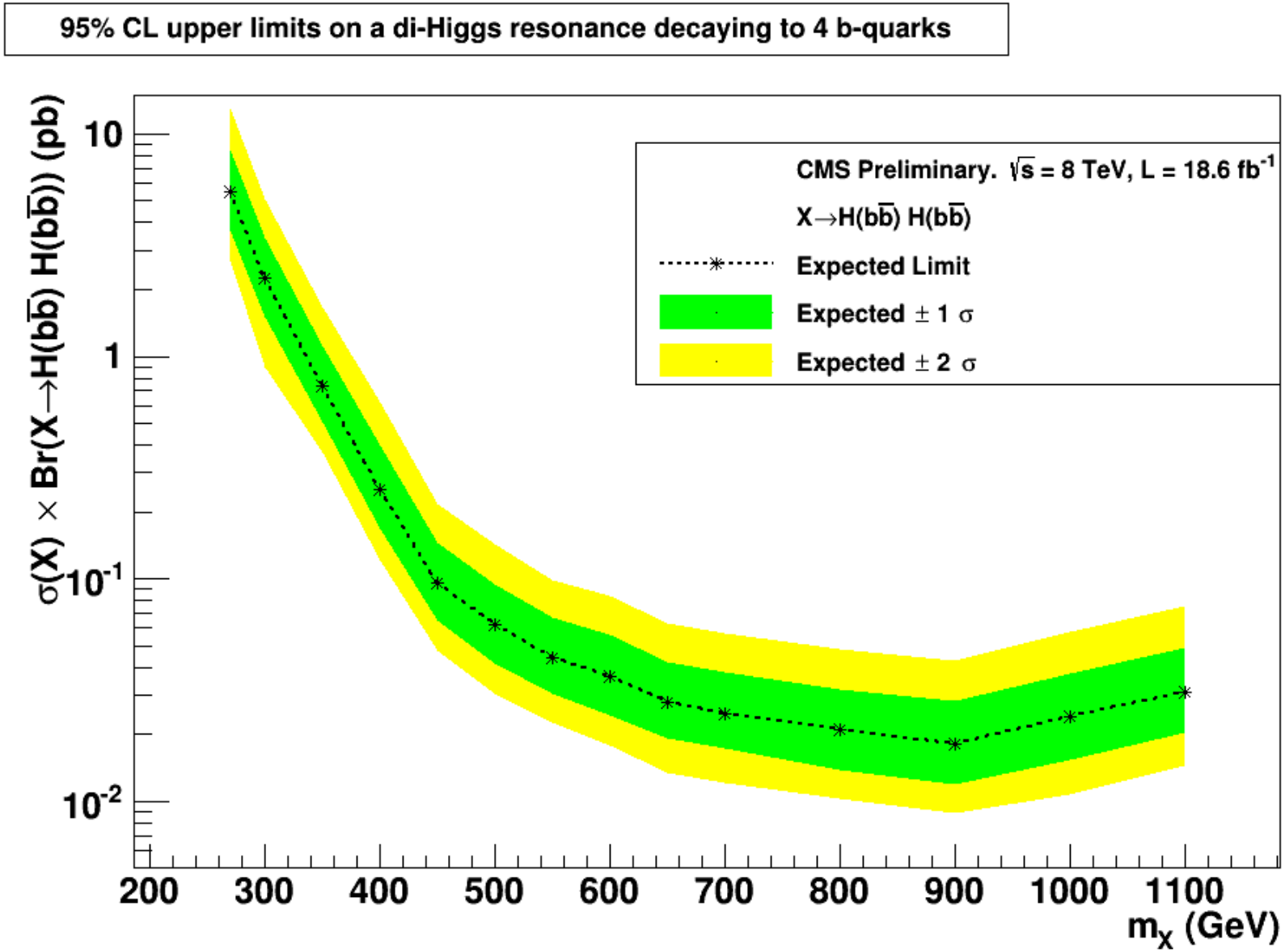
Expected Upper Limit with 18.6 /fb



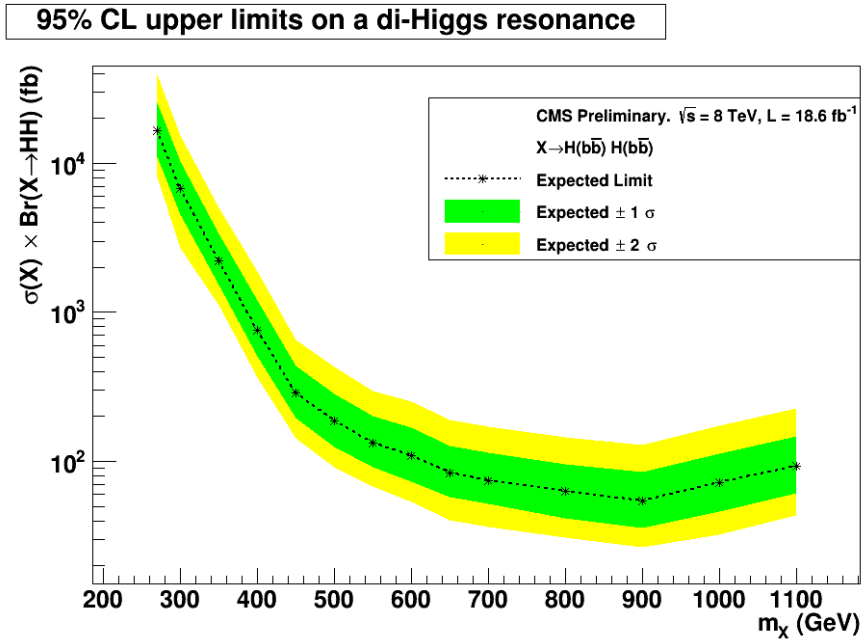


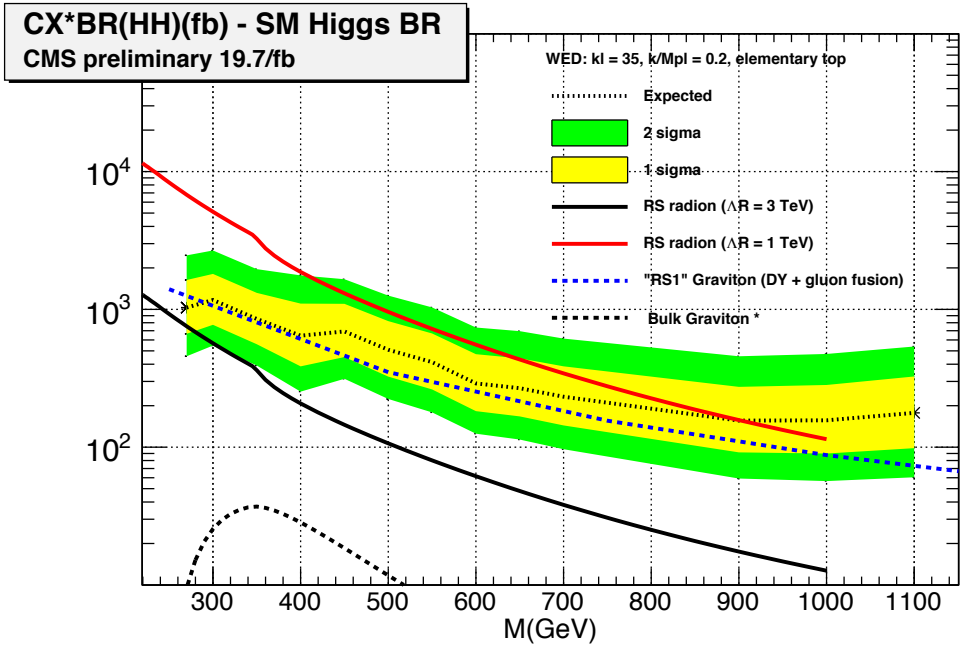
- Using the signal and background models described, we obtain the expected upper limits on the signal cross section at 95% CL using the **Asymptotic CL_S** method
- The Low Mass Analysis and High Mass Analysis are each powerful in their domains with the **turn-over at 400 GeV**
- The High Mass Analysis begins to lose sensitivity beyond 900 GeV, signaling the **relevance for a jet substructure based analysis beyond 1 TeV**

Combined Expected Upper Limit with 18.6 /fb



Compared with H(γγ)H(bb)





Conclusion

- This analysis is sensitive to resonant di-Higgs production in the 4 b-jet final state over a wide range of resonance masses and down to

  $\sigma(X \rightarrow HH; H \rightarrow bb; H \rightarrow bb)$ = 18 fb

  with 18.6 /fb of 8 TeV data

- Awaiting pre-approval if the *H->bb* group agrees

- Documentation is at:

  Analysis Note AN-13-227

  https://twiki.cern.ch/twiki/bin/viewauth/CMS/HHto4bUnboosted

BACKUP

Experimental Motivation

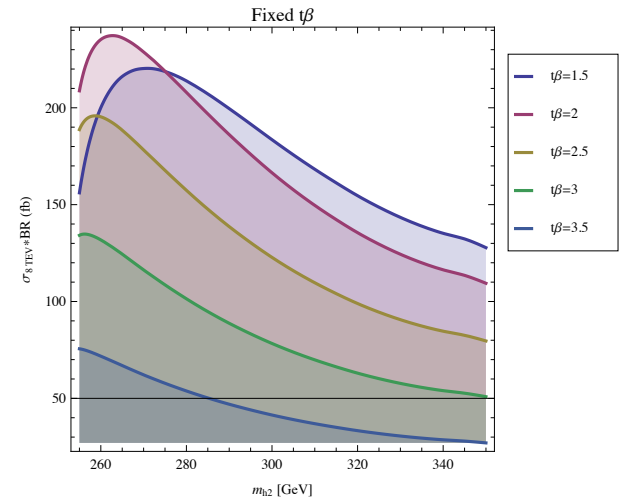
- The analysis uses data collected by a trigger with b-jet identification
- The granularity of tracking with the CMS Pixel Detector allows online secondary-vertexing and b-tagging. Demonstration of capability with a physics result
- Fidelity of b-jet identification and reconstruction reflected in the reach of this analysis through a 4 b-jet final state

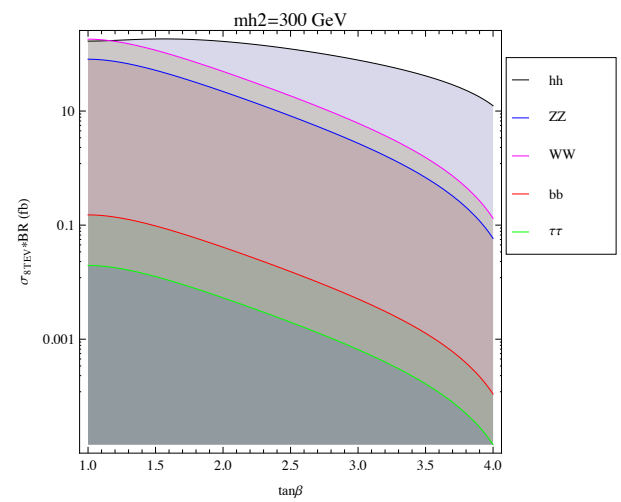
NMSSM, σ(pp→h2)·BR (h2→HH)

- Prediction on the σ(pp→h2)·BR (h2→HH) varying tan β and m(h2)

Table: σ_{8TeV}*BR$(h_2 \to hh)$ [fb]

mh2[GeV]	$t_\beta = 1.5$	t_β=2	t_β=2.5	t_β=3	t_β=3.5	t_β =4
255	156.101	208.753	188.721	134.168	75.5825	23.5914
260	200.057	235.238	195.341	131.795	71.7849	21.8852
265	216.157	236.133	188.447	124.457	66.9385	20.2386
270	220.269	228.789	178.208	116.291	62.1283	18.7044
275	218.742	218.886	167.691	108.592	57.7793	17.3518
280	213.89	207.953	157.41	101.402	53.812	16.1356
285	207.168	196.859	147.666	94.7731	50.2072	15.0404
290	199.573	186.154	138.661	88.7571	46.9668	14.0618
295	191.613	176.017	130.391	83.3039	44.0499	13.1848
300	183.525	166.454	122.771	78.3291	41.4031	12.3915
305	175.695	157.654	115.878	73.8631	39.0366	11.6842
310	168.262	149.62	109.671	69.8652	36.9253	11.0545
315	161.261	142.29	104.071	66.2774	35.036	10.4921
320	154.603	135.516	98.9476	63.0093	33.3194	9.98193
325	148.677	129.583	94.4919	60.1782	31.8362	9.54201
330	143.422	124.393	90.6195	57.7274	30.5558	9.16305
335	138.898	119.963	87.3322	55.6557	29.4772	8.84471
340	135.187	116.334	84.6529	53.9765	28.6073	8.58906
345	132.31	113.502	82.5741	52.684	27.943	8.39523
350	127.789	109.406	79.6334	50.863	27.0057	8.12092





- For m_{h2} ~ 300 GeV we expect h2→4b to contribute with a σ·BR < 60 fb assuming SM BR(H→bb)
- HH decay mode is, in this particular phase space, 2 odm greater than bb and ττ
- For tan β> 1.5, HH decay mode is much more dominant with respect to ZZ, WW.

Physics Objects

Analysis uses standard CMS physics objects provided by the **Physics Analysis Tools** (PAT) framework and approved by relevant **Physics Object Groups** (POG).

- **Primary Vertex Selection and Pile-up**

  Primary vertices identified using tracks clustered by **Deterministic Annealing**. Required to have $|z\text{-}z_0| < 24$ cm, $|r| < 2$ cm, and fit n.d.o.f. > 4. 8 TeV data contained 10 – 30 primary vertices, and this affects jet reconstruction unless corrected for.

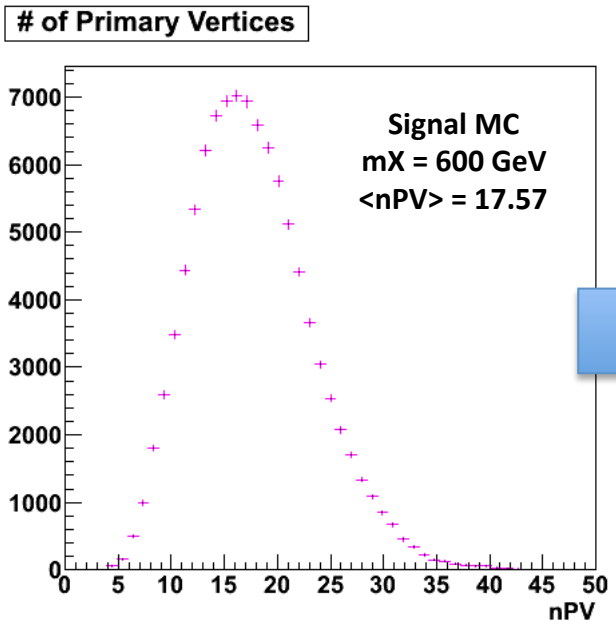
  - **PFnoPU**

    Subtracts tracks from jets that do not originate from the primary interaction. Works in the η region covered by the tracker.

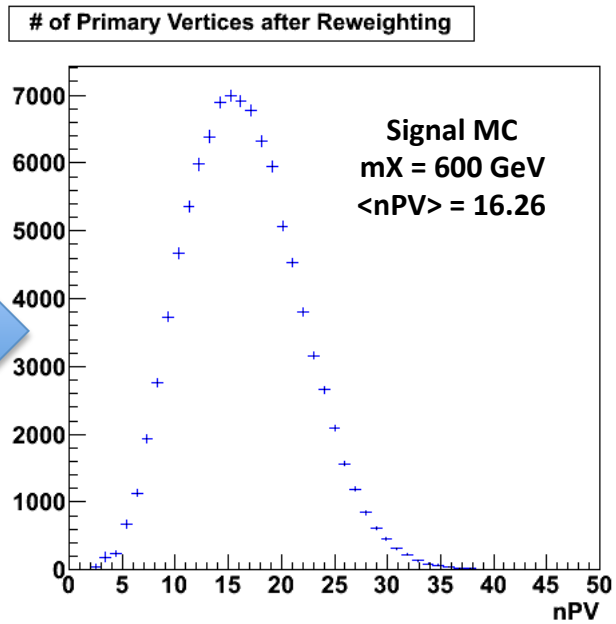
  - **Fastjet**

    Subtracts a momentum density per unit area due to PU for each event from each jet.

Pileup Reweighting

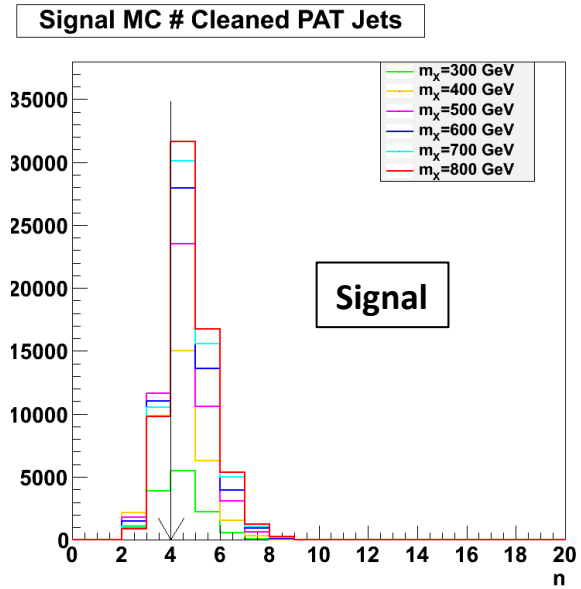


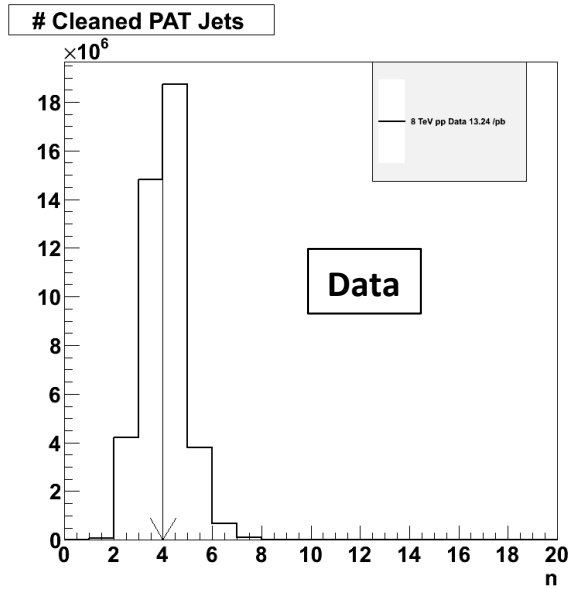
Pileup Reweighting



- Standard pile-up reweighting technique used in H(bb) analyses used here.
- **nPV reweighted from 17.57 to 16.26**, thus mimicking data

Number of Central Jets in the Events



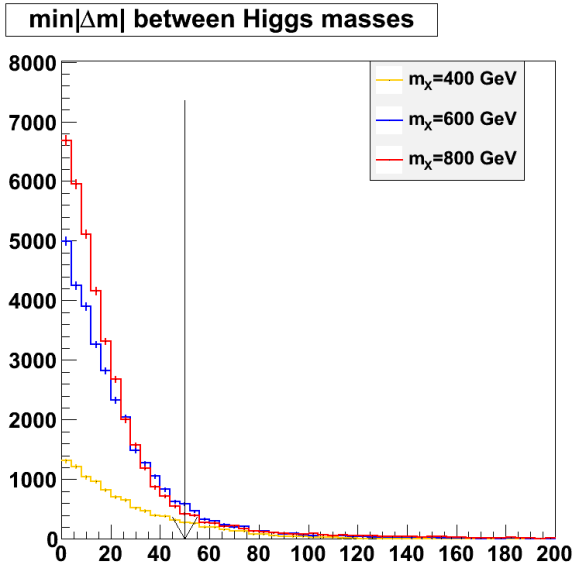


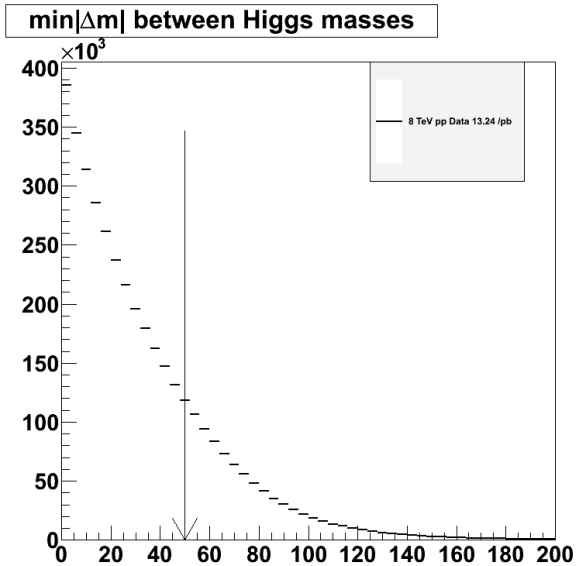
In following slides, the distributions will be used to motivate selection criteria for the High Mass Analysis. Corresponding distributions for the Low Mass Analysis may be found in the Note.

- Jets have $p_T > 40$ GeV, $|\eta| < 2.5$ (no b-tagging applied yet)
- Require at least 4 such jets
- Representative selection efficiencies:

	300 GeV	400 GeV	500 GeV	600 GeV	700 GeV	800 GeV
Trigger	13%	35%	51%	59%	63%	65%
not VH(bb)	99%	99%	97%	95%	93%	90%
nCJets> 3	63%	65%	73%	78%	82%	83%

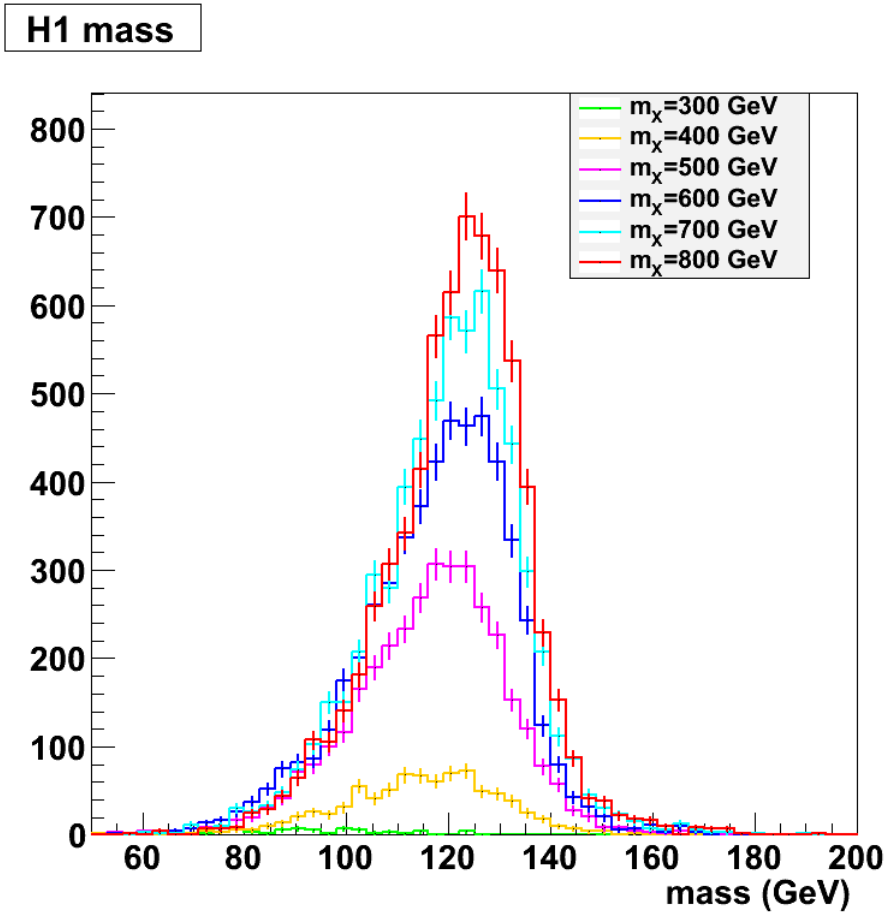
Choice of Four Jets. min(|mH1-mH2|)





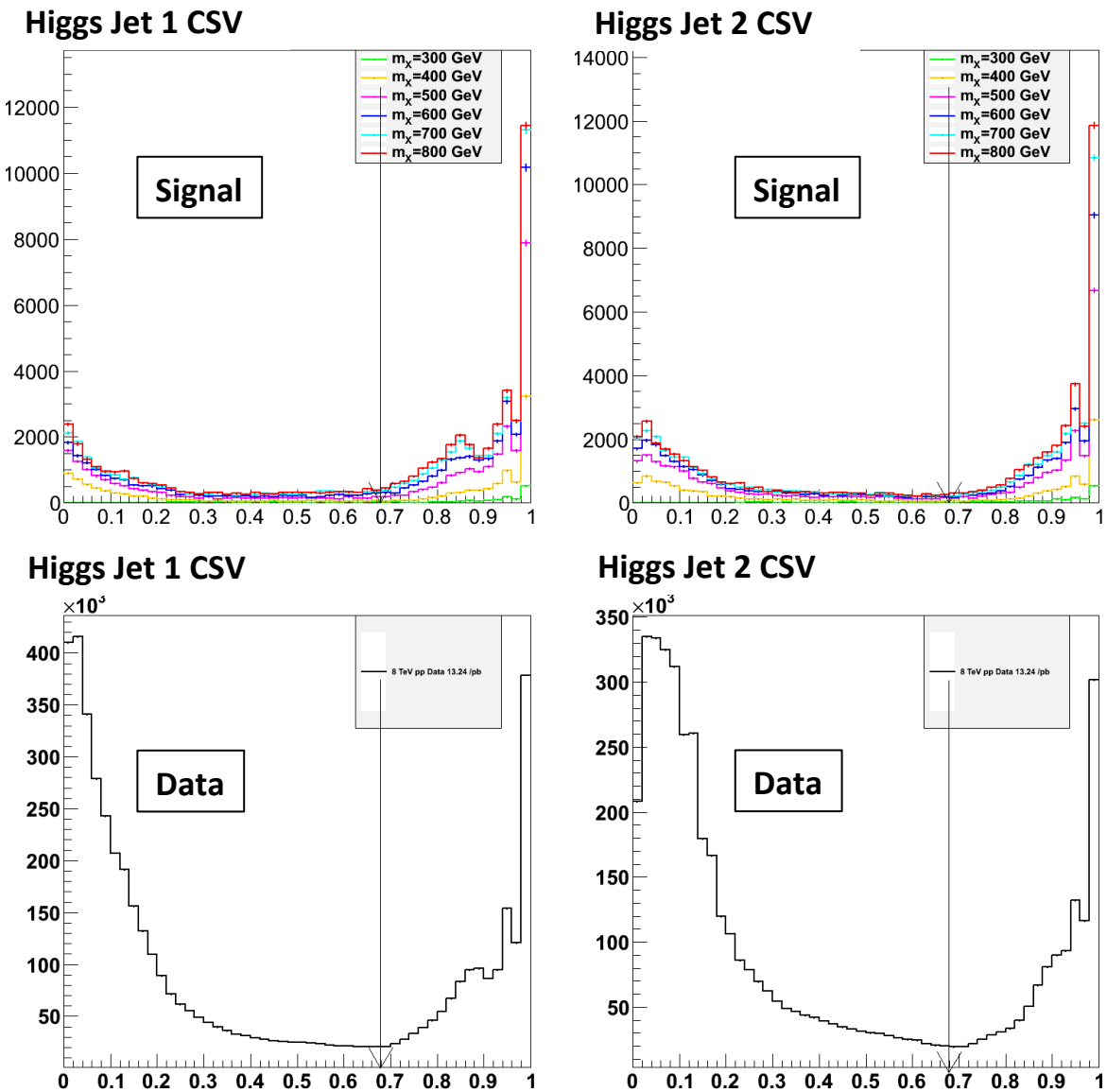
- Loop over all di-jet pairs and converge on the two di-jet pairs with the least difference in masses. The ΔR(jet, jet) must be less than π/2
- Now you have **4 kinematically chosen jets** to reconstruct HH

The Reconstructed Higgs Candidate Mass. $\Delta R < \pi/2$



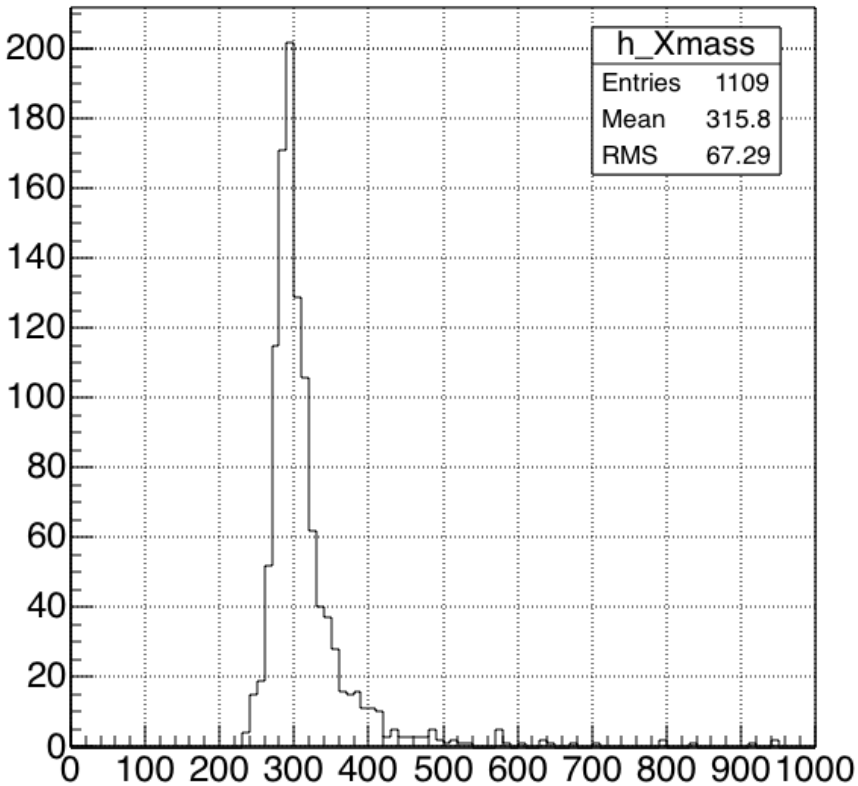
Efficiency Table here

CSV Distributions of the 4 Chosen Jets



- Within a Higgs, Jets are ordered by p_T
- Will **cut symmetrically** on all four jets in order to make kinematic extrapolation easy.
- CSV-medium (0.679) was chosen through a n-1 S/√B optimization study

“CSV-selection”, Xmass in the SR



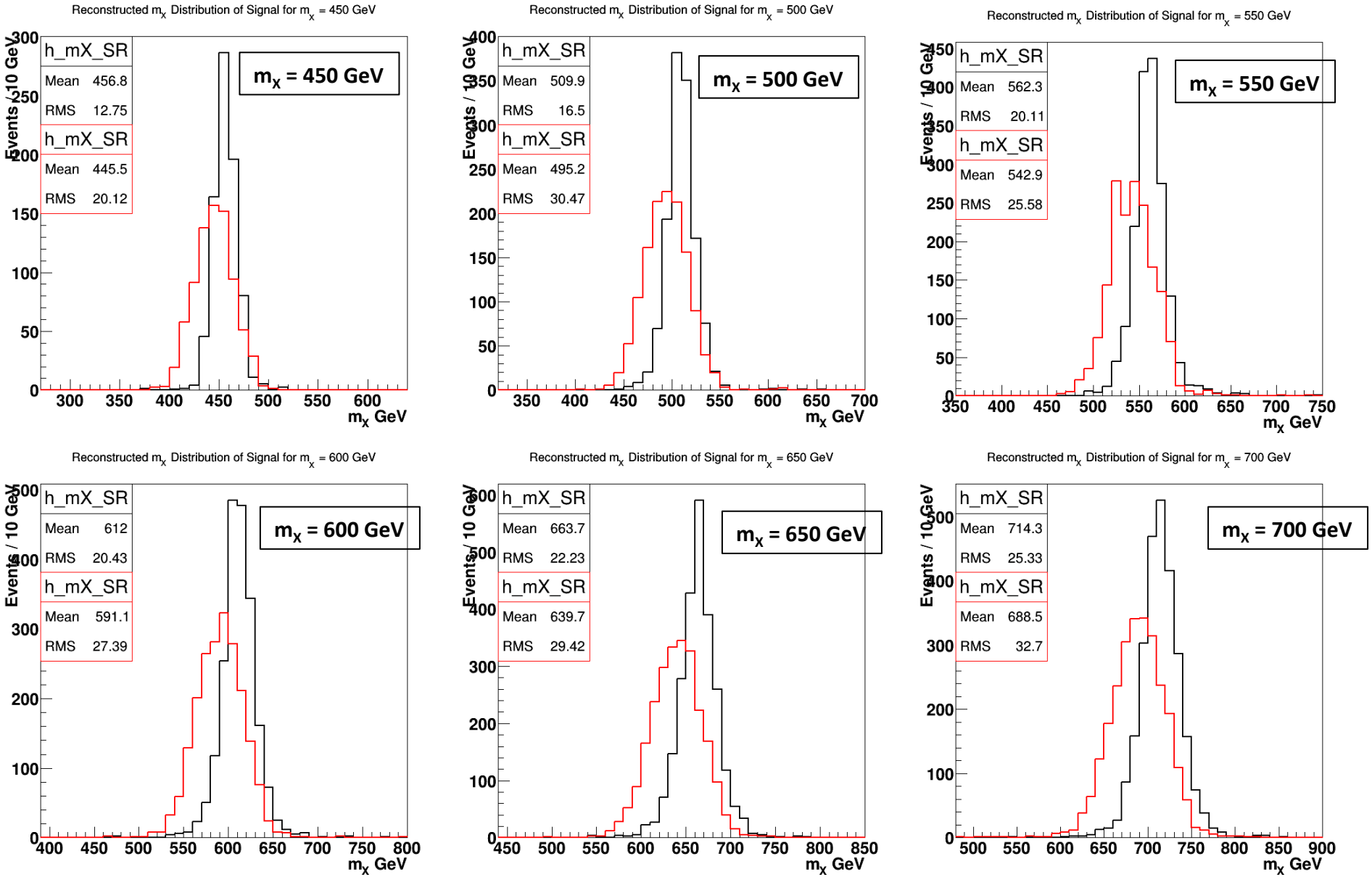
	angular	box	minDm
csv-trigger	13,42	13,42	13,42
nJets>3&b-tag	2,05	12,77	17,88
HH candidate	23,04	45,33	83,77
Signal Region	12,69	43,91	40,07
Cumulative	0,008	0,343	0,8
Purity*	55	60,1	62,2

Purity = (#evt matched | HH cand & btag)

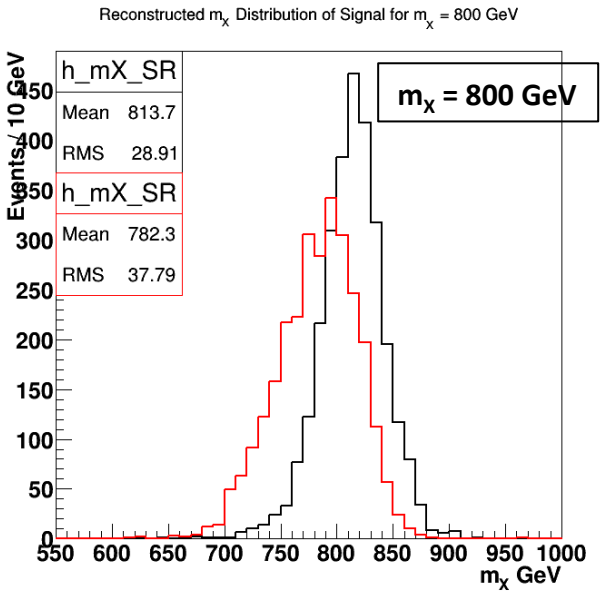
/(# evt HH cand & btag)

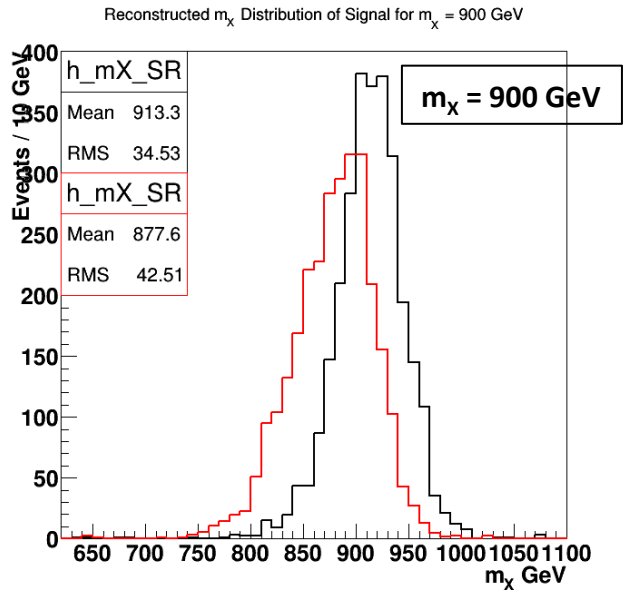
(*) ‘evt matched’ has to be intended as an event in which the 4 jets selected as HH candidates are matched (dR<0.4) to the jets at the generator level originated by the 2 H from the X

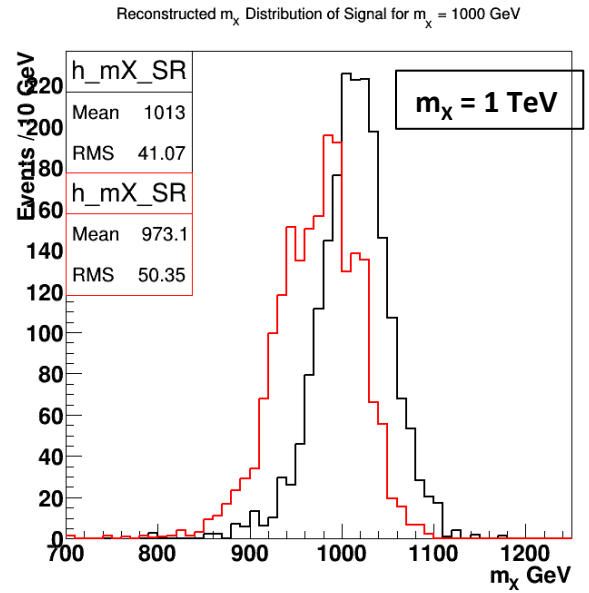
Signal m_X Distributions – High Mass Analysis w& w/o Kinematic Constraints

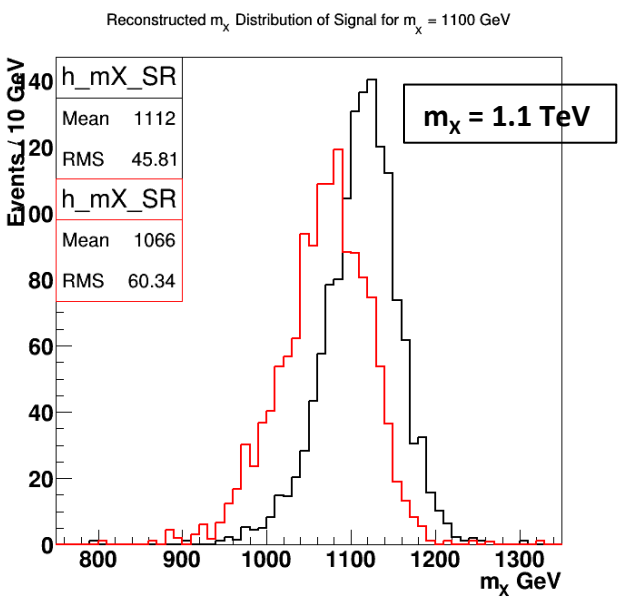


Signal m_X Distributions – High Mass Analysis w& w/o Kinematic Constraints



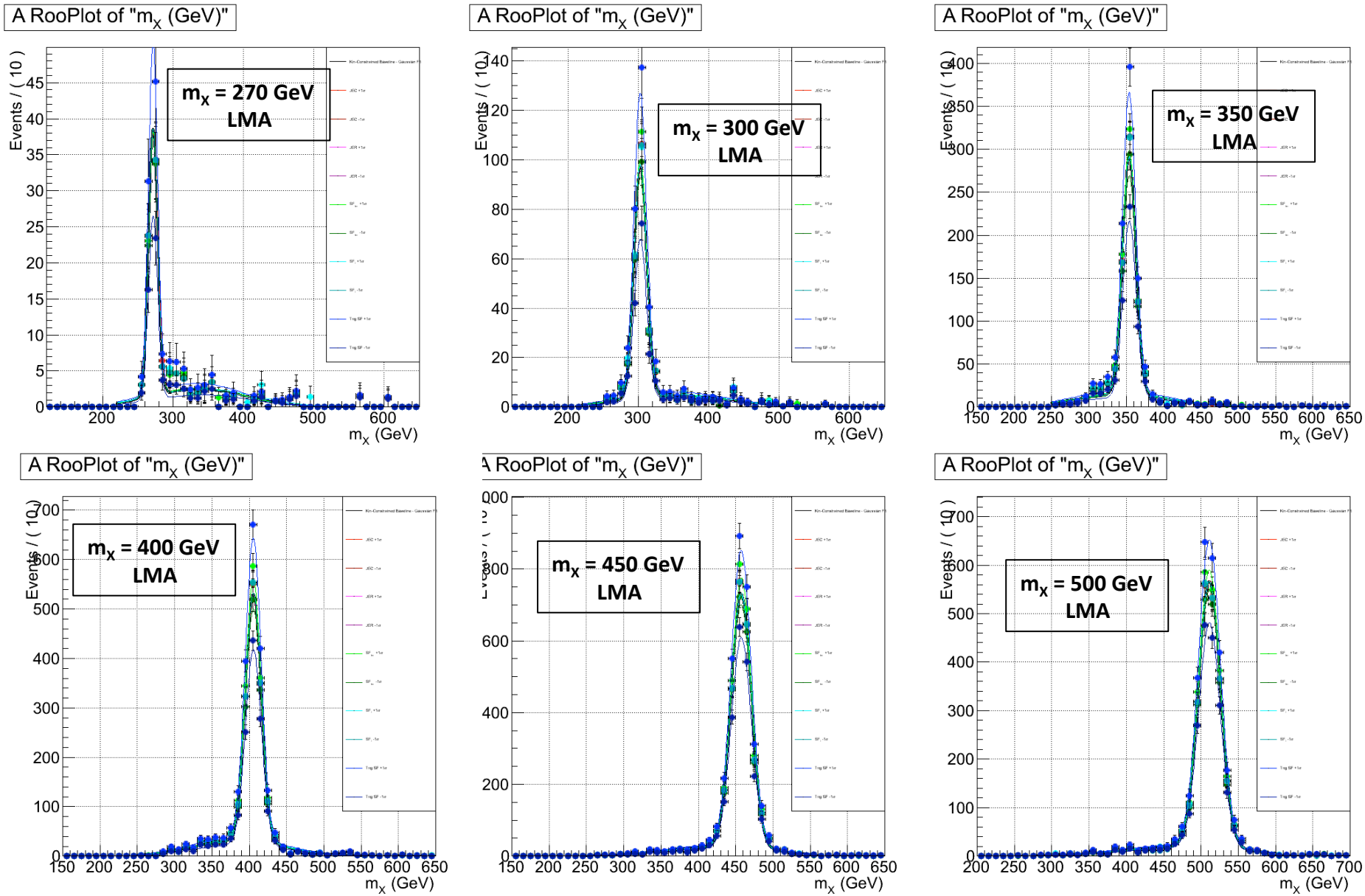




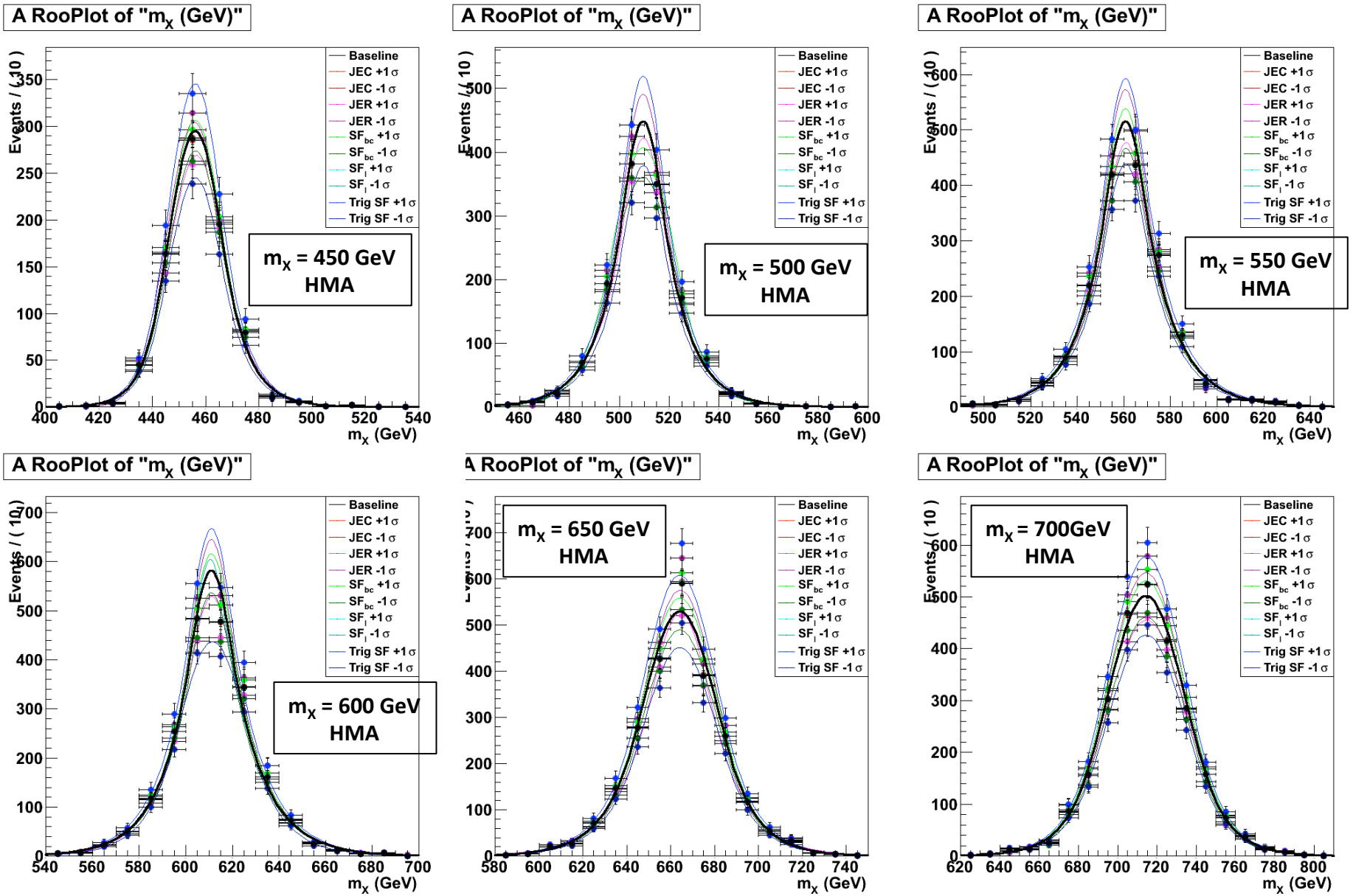


- Similar distributions for the m_X = 270, 300, 350 and 400 GeV are shown in the Note.
- Method improves **resolution of the di-Higgs resonance**, m_X, by factors of 20% to 40%

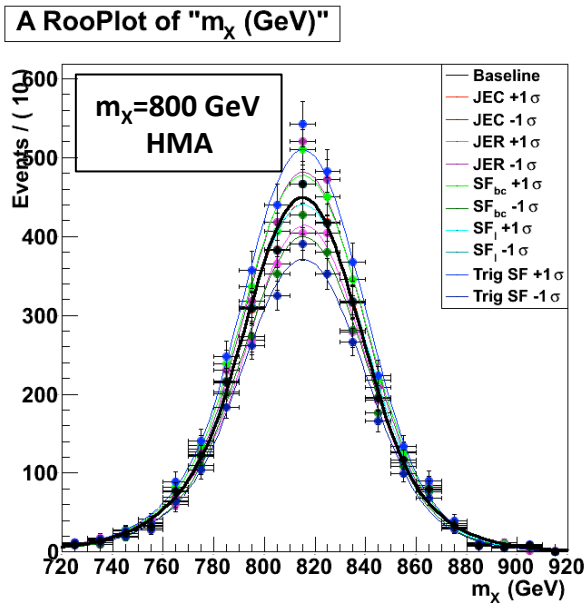
Signal Modeling with Systematics – Low Mass Analysis

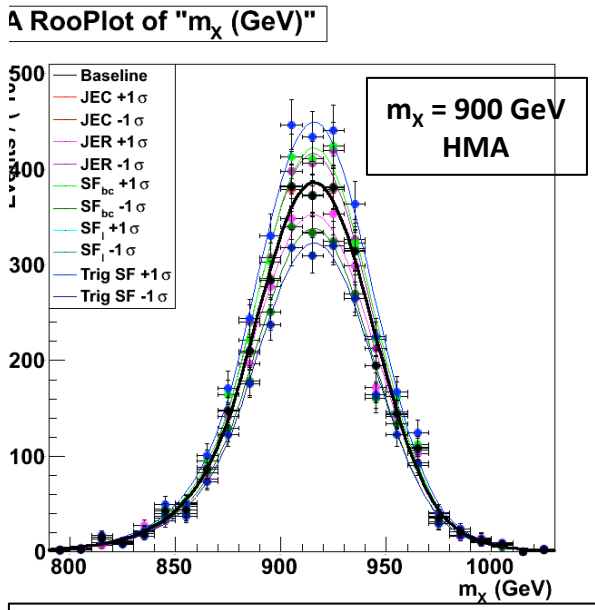


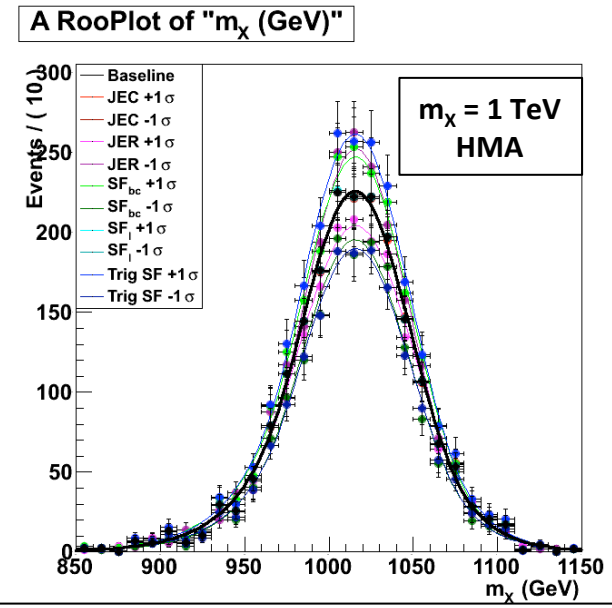
Signal Modeling with Systematics – High Mass Analysis

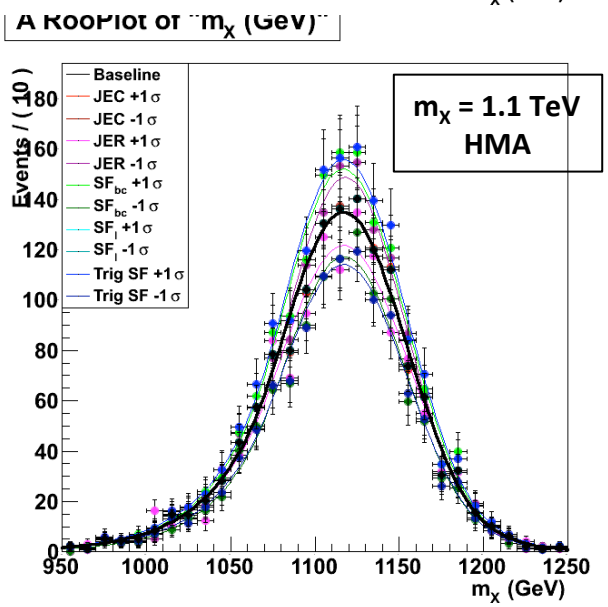


Signal Modeling with Systematics – High Mass Analysis









Typical Contributions to Systematic Uncertainties

- JEC: ~ 0.1%
- JER: ~ 5%
- bTag heavy flavor: ~ 10%
- bTag light flavor: ~ 0.01%
- Trigger: ~ 15%

Detailed tabulation of fits and uncertainties in:

LMA: http://www.pi.infn.it/~vernieri/TLTL8030/SignalSystematics.html

HMA: http://home.fnal.gov/~souvik/FullSim/SignalSystematics/SignalSystematics.html

300	S	b	s/sqrt(b)
TLTL	63.3502477935514	3603	1.05539780571638
TLbarTLbar	12.1191778387664	18826	0.088327132450769
400			
TLTL	347.383372452629	3798	5.63678599561288
TLbarTLbar	42.1581762685229	21823	0.285380693516967

Low mass, $T\bar{L}T\bar{L}$ - p_T thresholds choice

- If we use as thresholds for the jet p_T the default ones (80,30) while reverting the CSV requirements for 2 of the 4 jets, we exploit a phase space where the trigger bias is stronger and shapes the m_X distribution

- A study with on a top sample has shown that this feature is seen in that sample too

  - This can be settled by looking at the mX distribution in ttbar with and without the trigger.

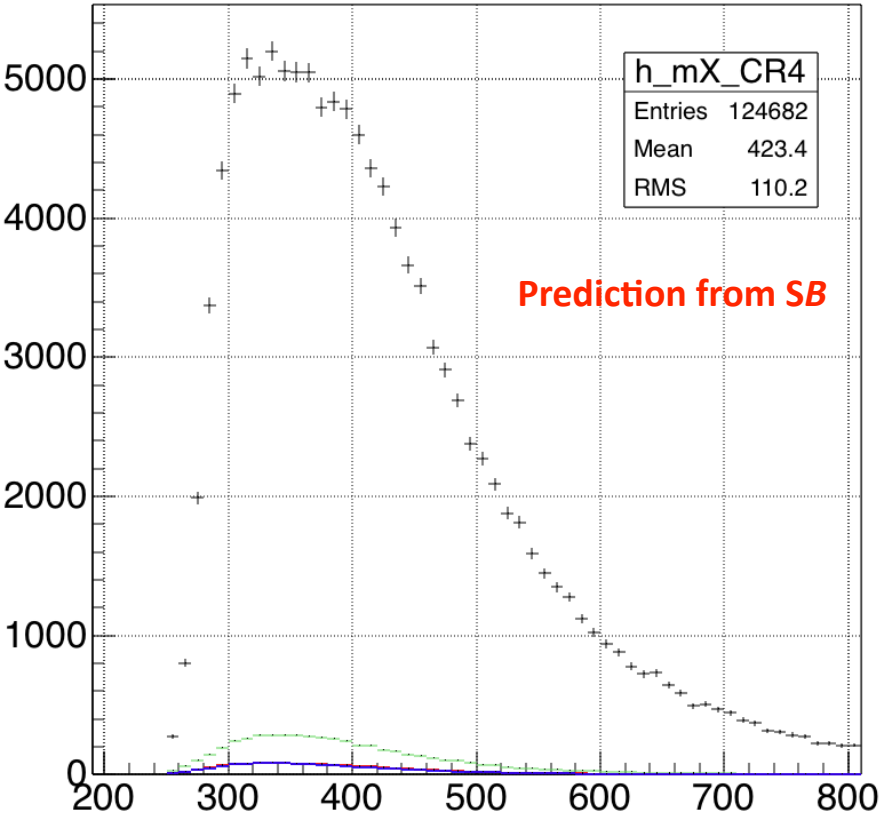
  - Also for a QCD sample with and without the trigger.

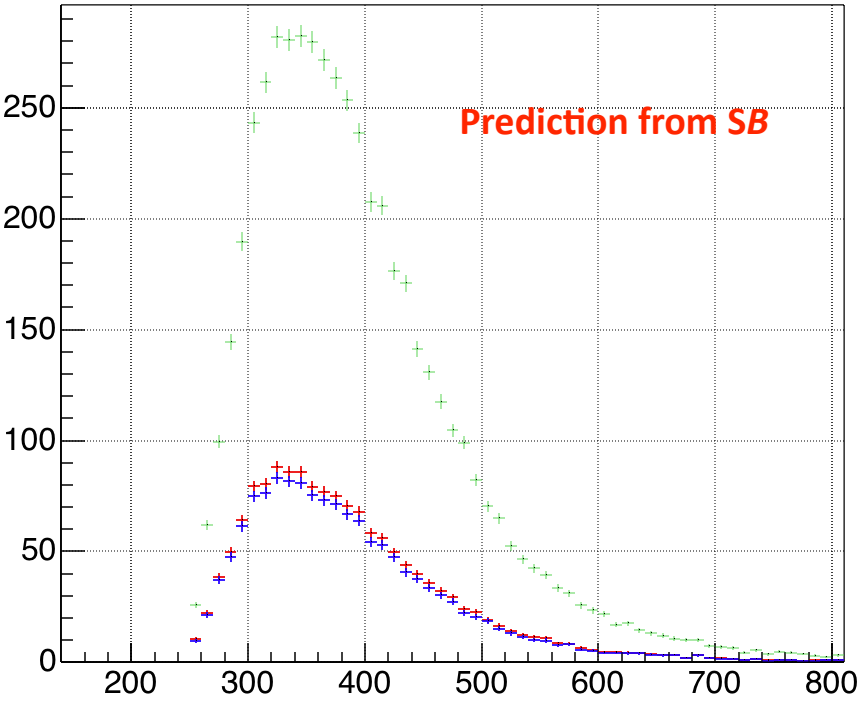
Top contribution

TLbarTLbar

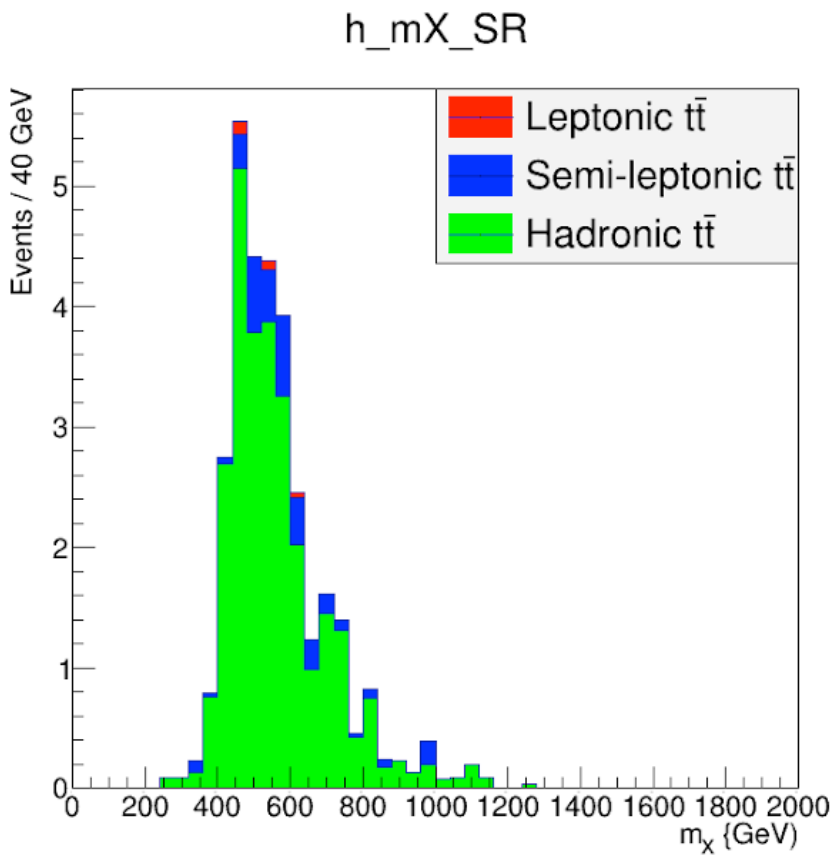
- ttbar has similar shape to the selected data
- we checked several distributions, not just mX

TTjets Semilept
TTjets FullLept
TTjets Hadronic
Data





High Mass Analysis ttbar Contribution



xsec_ttbar_fulllept = 24.56 pb
xsec_ttbar_semilept = 103.12 pb
xsec_ttbar_hadronic = 106.32 pb

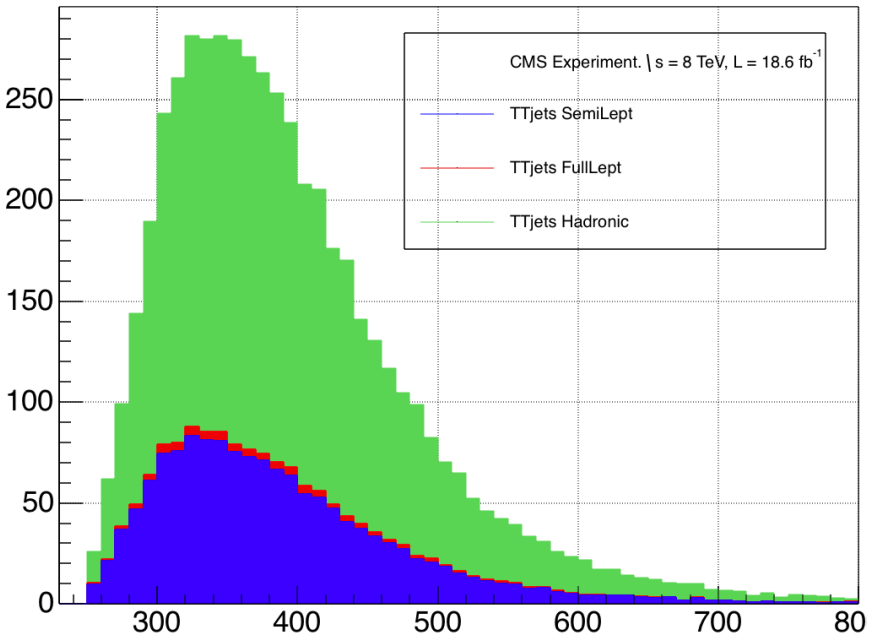
Contribution of fully leptonic ttbar is 0.217462
Contribution of semi leptonic ttbar is 3.46977
Contribution of hadronic ttbar is 28.0169
Total contribution of ttbar is 31.7041

Number of data events in SR = 298

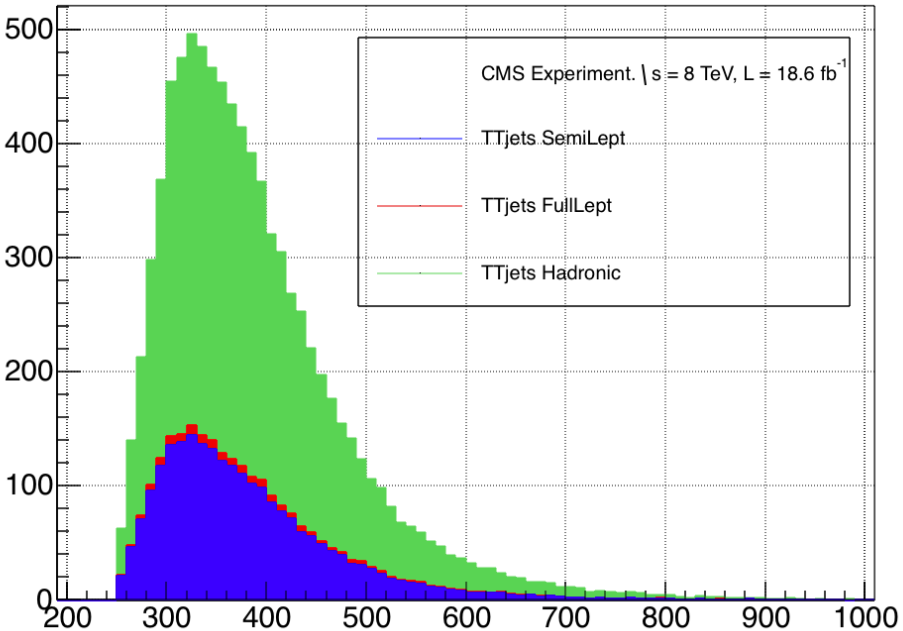
Trigger Bias? TTbar

SB, TLbarTLbar

Trigger on



Trigger off

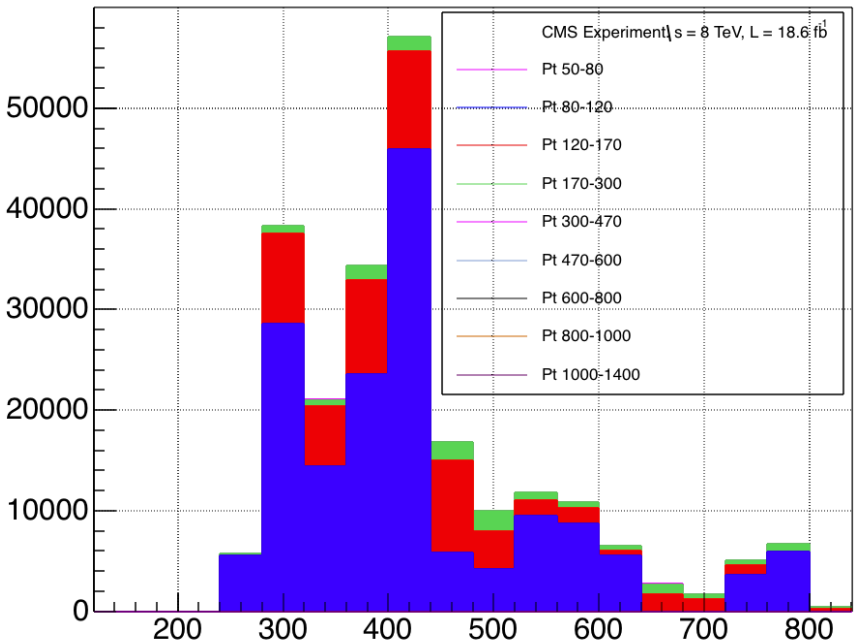


- The shaping due to the trigger is clear, since the analogies with the selected data we are confident the same effect is affecting our selected data sample

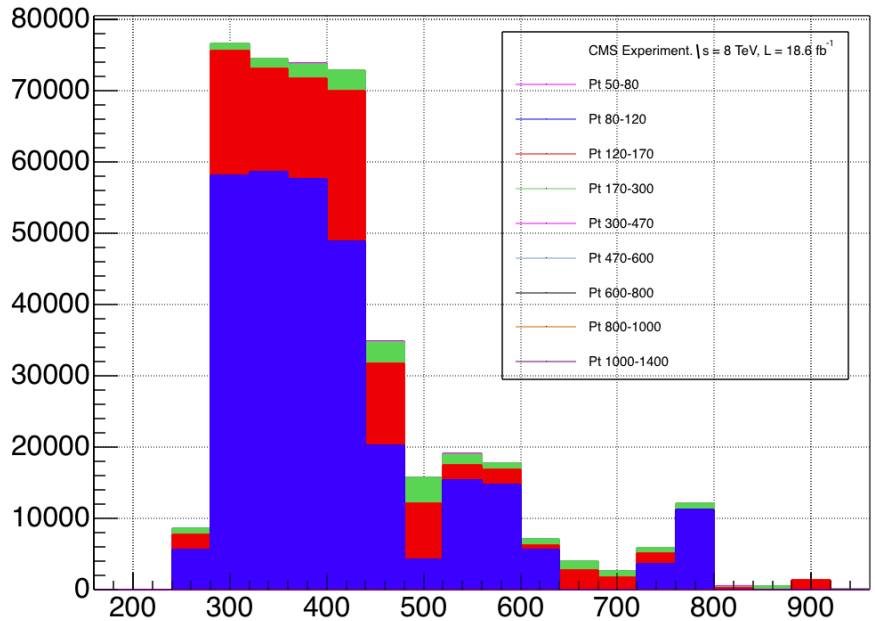
Trigger Bias? QCD

SB, TLbarTLbar

Trigger on



Trigger off

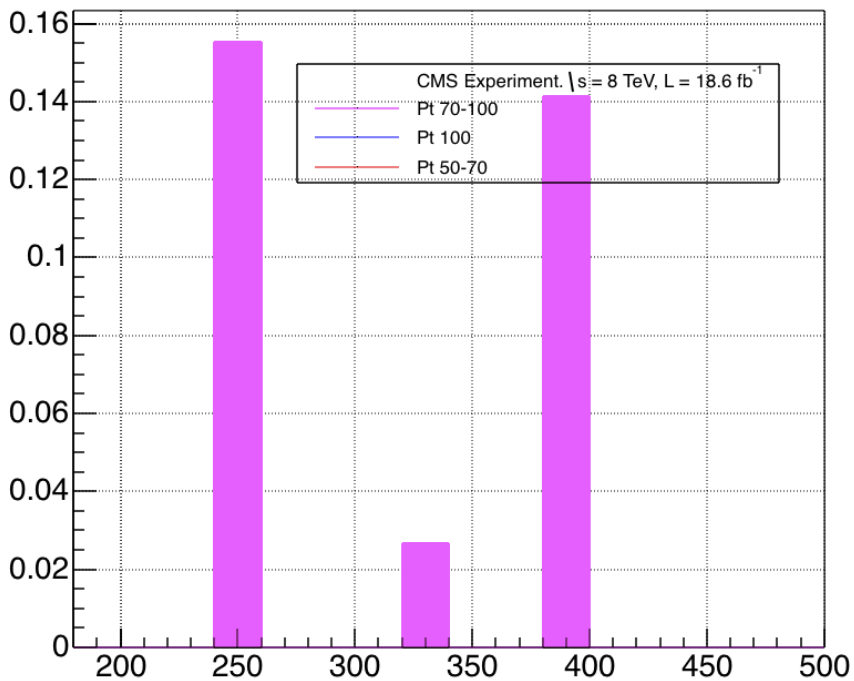


Trigger Bias? Z+jets

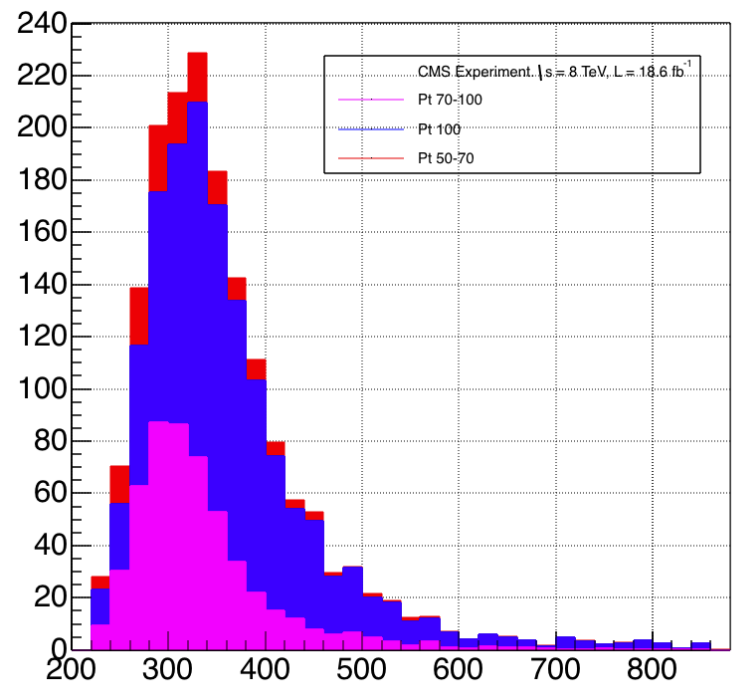
SB, TLbarTLbar

We use DY samples in which the Z is forced to decay into ll
We assume the leptons are b-quarks passing the 2CSVT requirement
Though we over-estimate this bkg we found its contribution is negligible

Trigger on



Trigger off

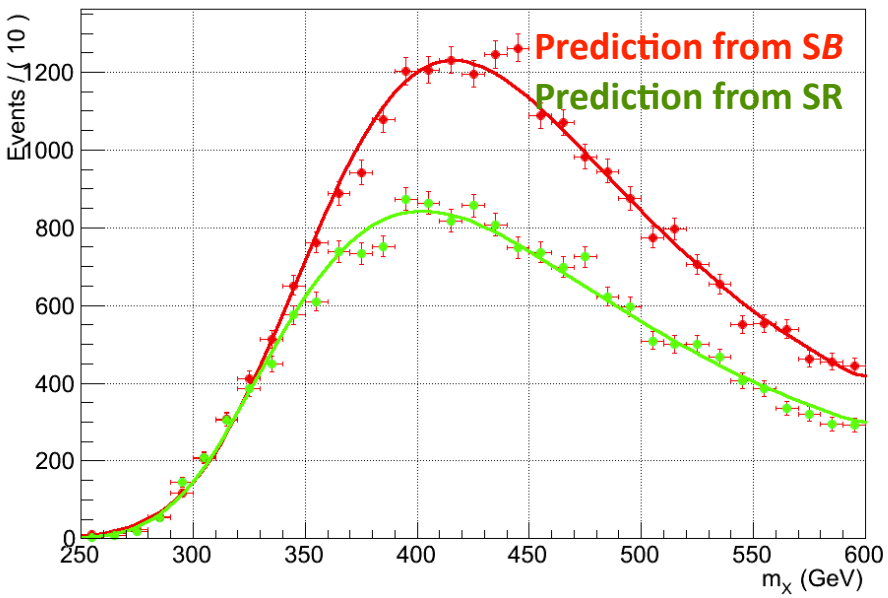


Solution I (90,40)

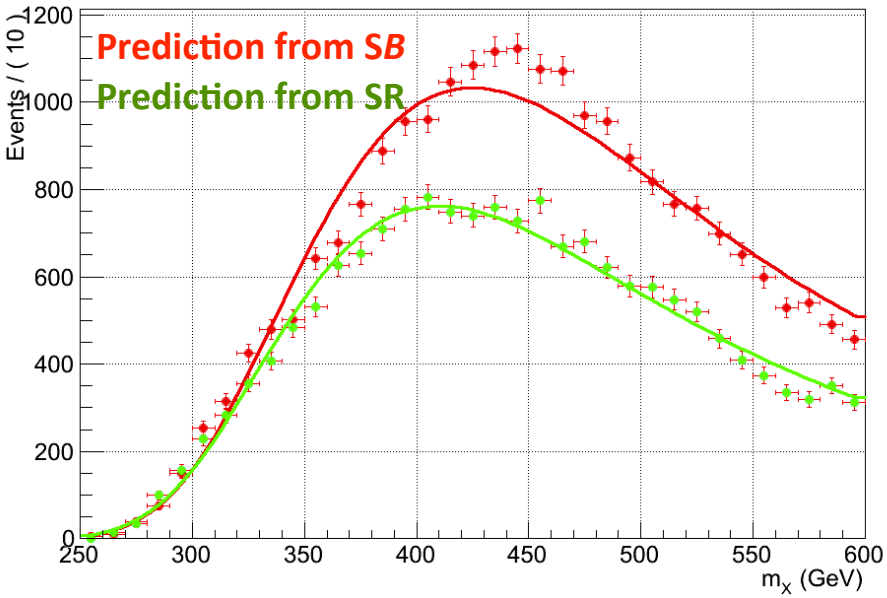
TLbarTLbar

- The solution we tried is to use higher thresholds for the jet pT and to order jet according to their pT value since we require to anti-b-tagged jets
    - No further splitting in categories is then needed

No Kin Fit



Kin Fit



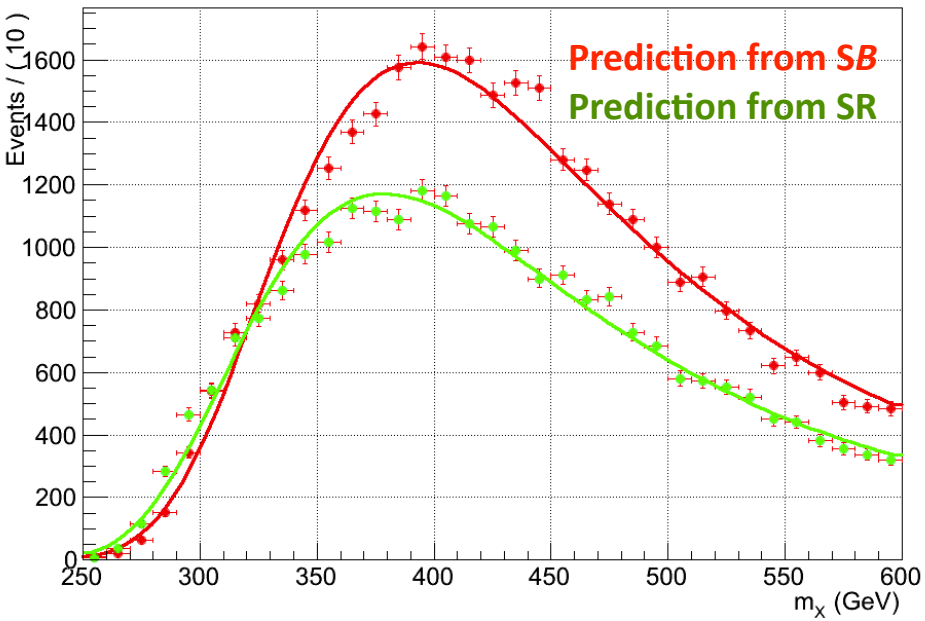
Fit to be fixed, but no double peak feature

Solution II (80,40)

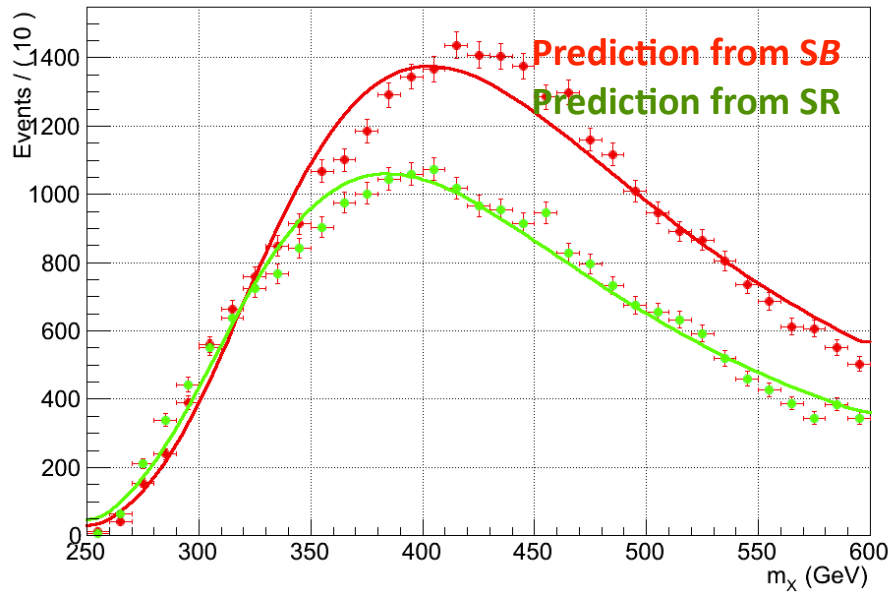
TLbarTLbar

- The solution we tried is to use higher thresholds for the jet pT and to order jet according to their pT value since we require to anti-b-tagged jets
    - No further splitting in categories is then needed

No Kin Fit

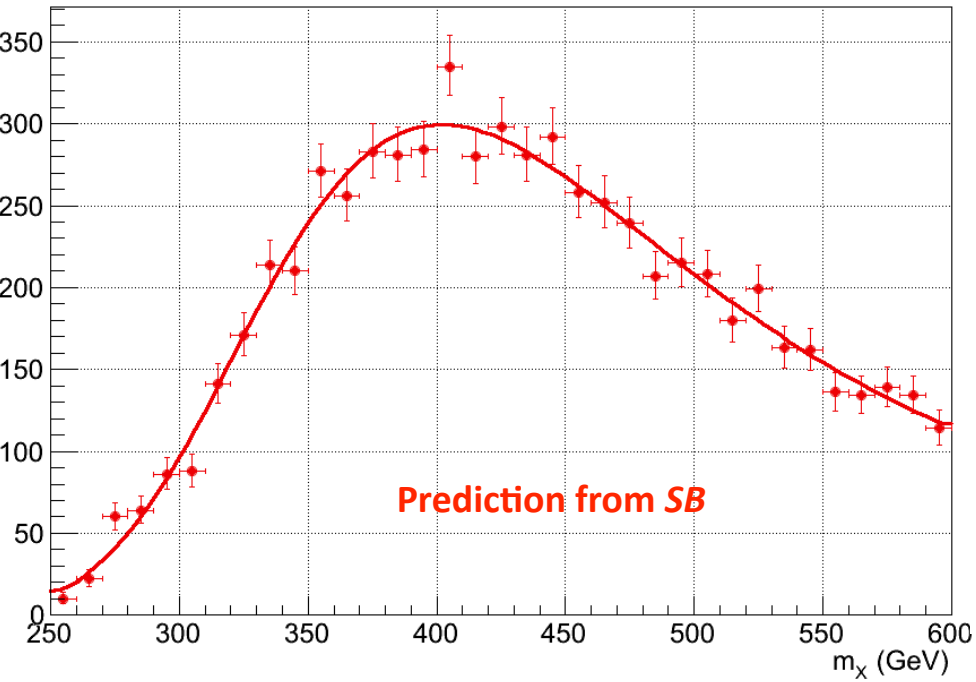


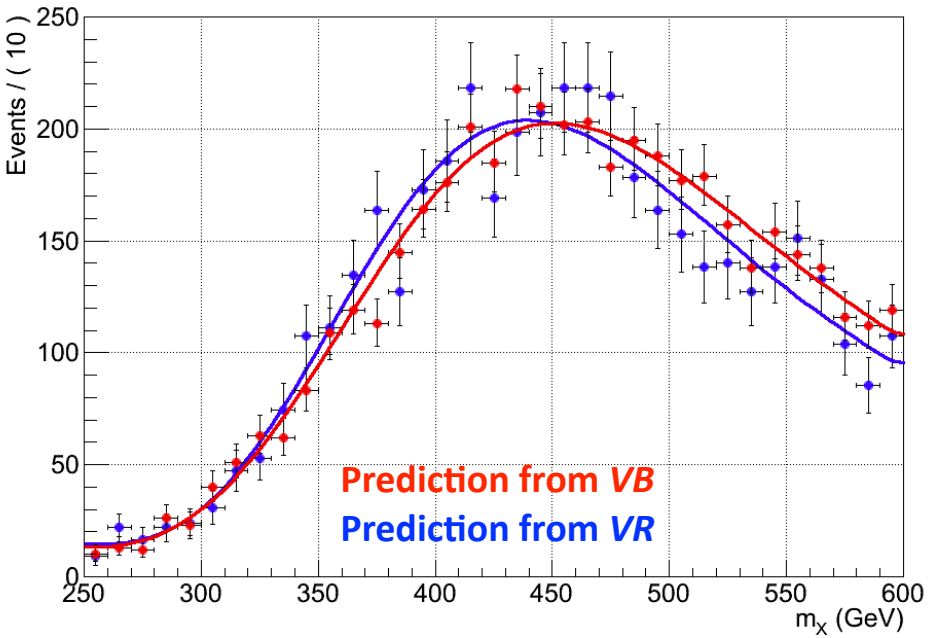
Kin Fit



Fit to be fixed, but no double peak feature

80/40, TLTL





chi^2/ndof = 0.959972

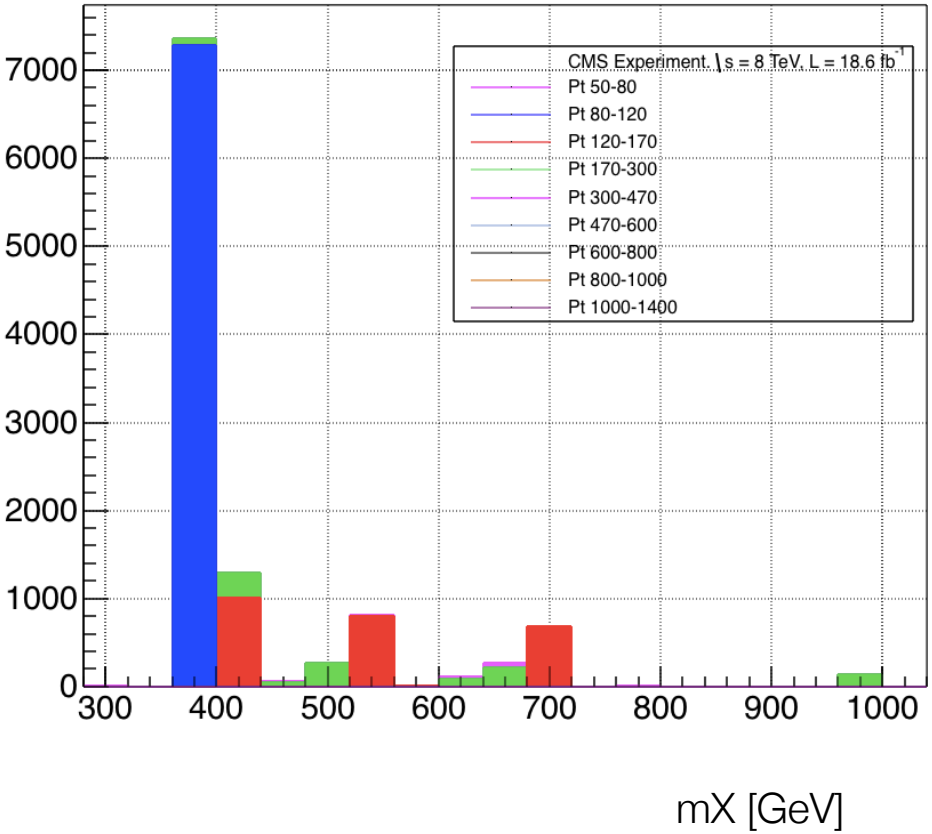
chi^2/ndof of bC = 0.582375
chi^2/ndof of bS = 0.868697

Signal fits are available here:
http://www.pi.infn.it/~vernieri/test/SignalSystematics.html

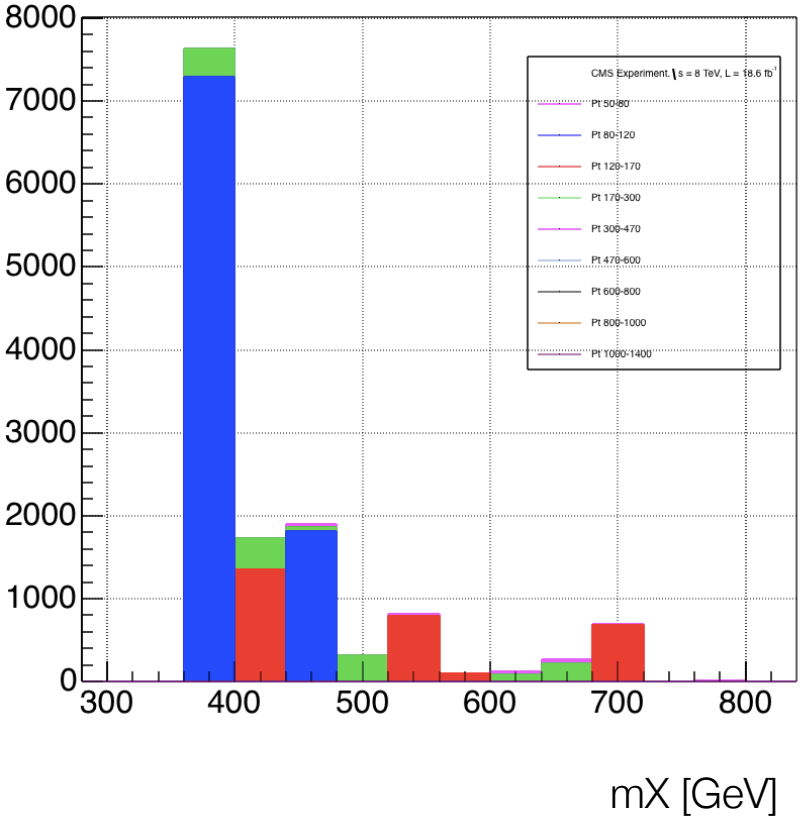
QCD Contribution

SB, TLTL

Trigger on



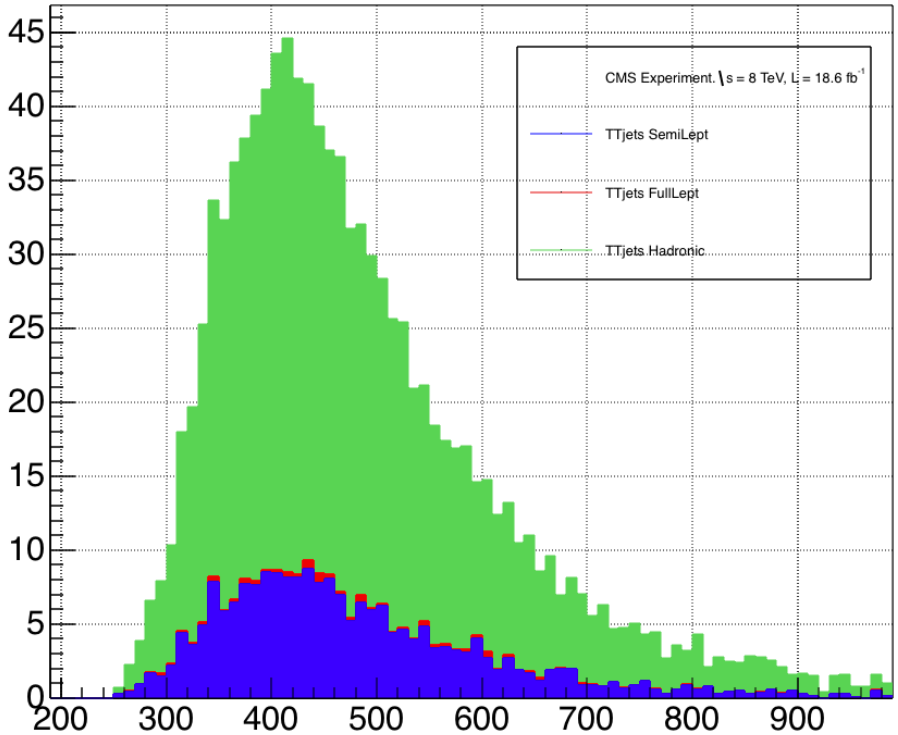
Trigger off



TTbar Contribution

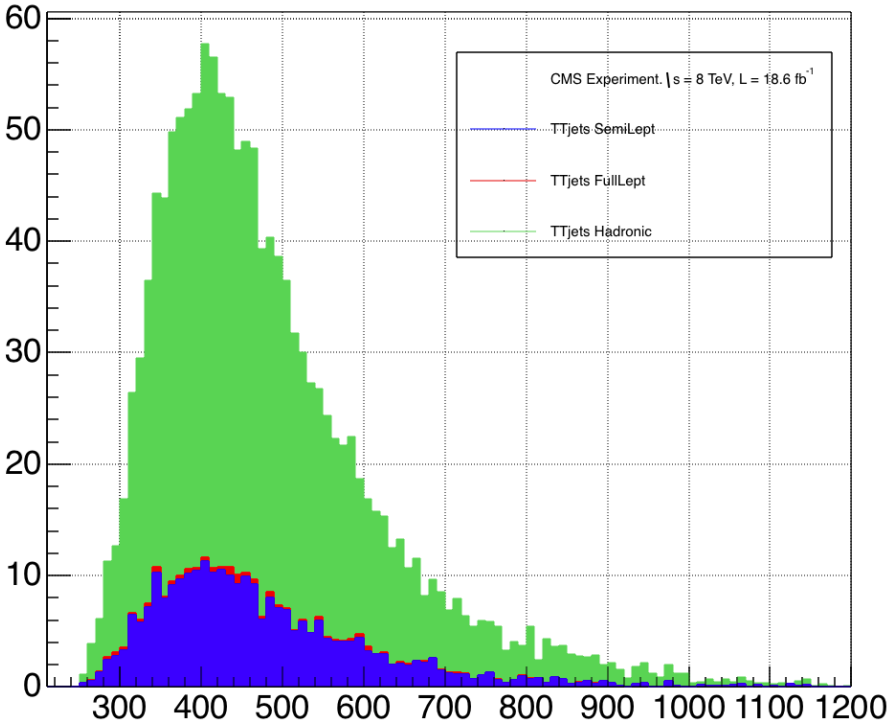
SB, TLTL

Trigger on



mX [GeV]

Trigger off



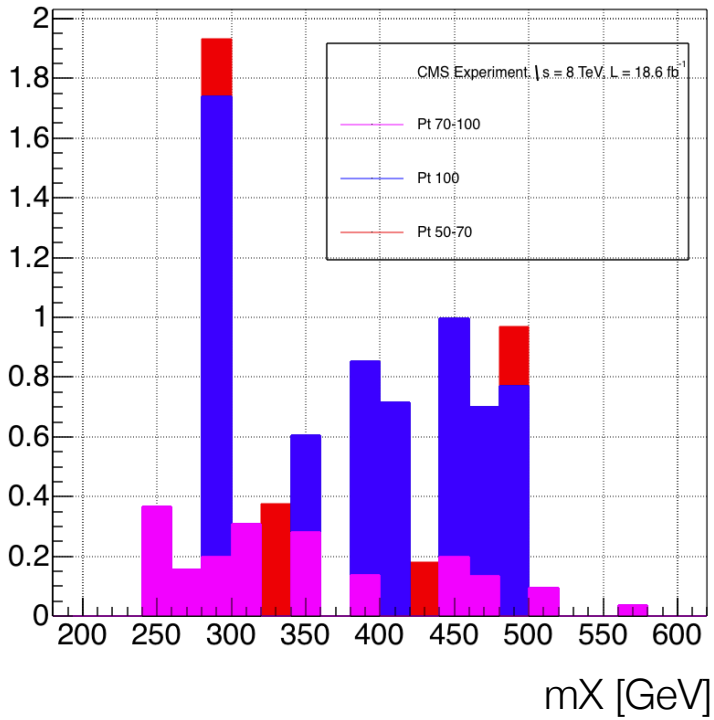
mX [GeV]

Z+jets Contribution

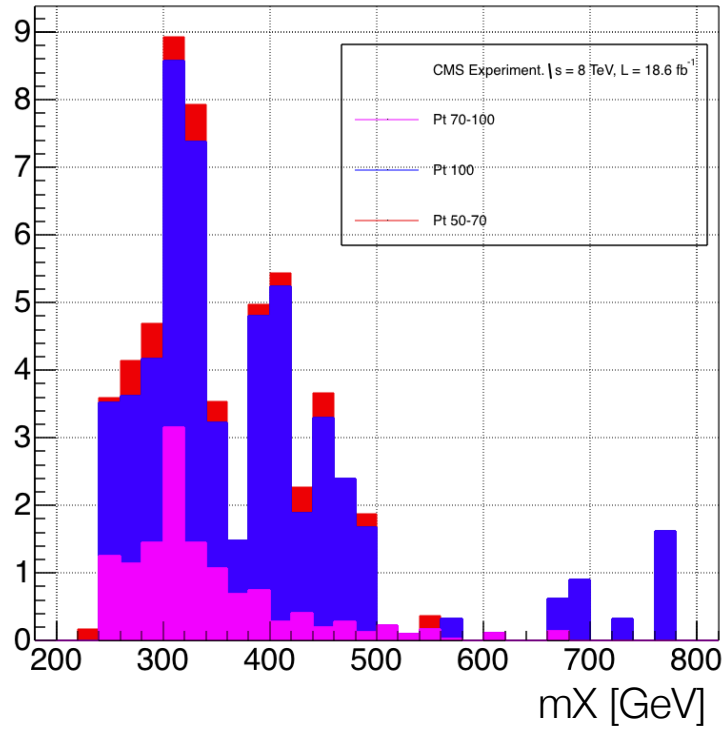
SB, TLTL

We use DY samples in which the Z is forced to decay into ll
We assume the leptons are b-quarks passing the 2CSVT requirement
Though we over-estimate this bkg we found its contribution is negligible

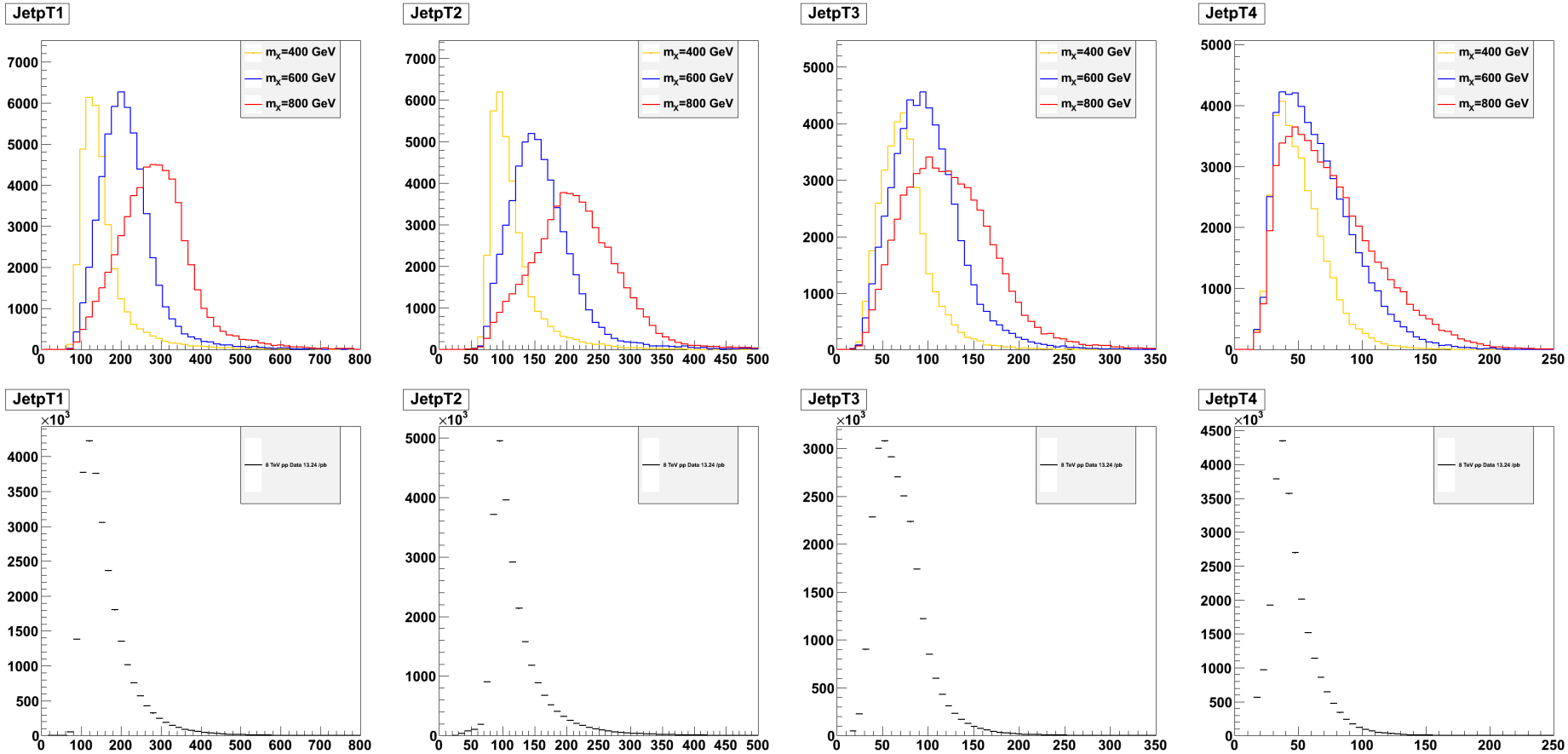
Trigger on



Trigger off

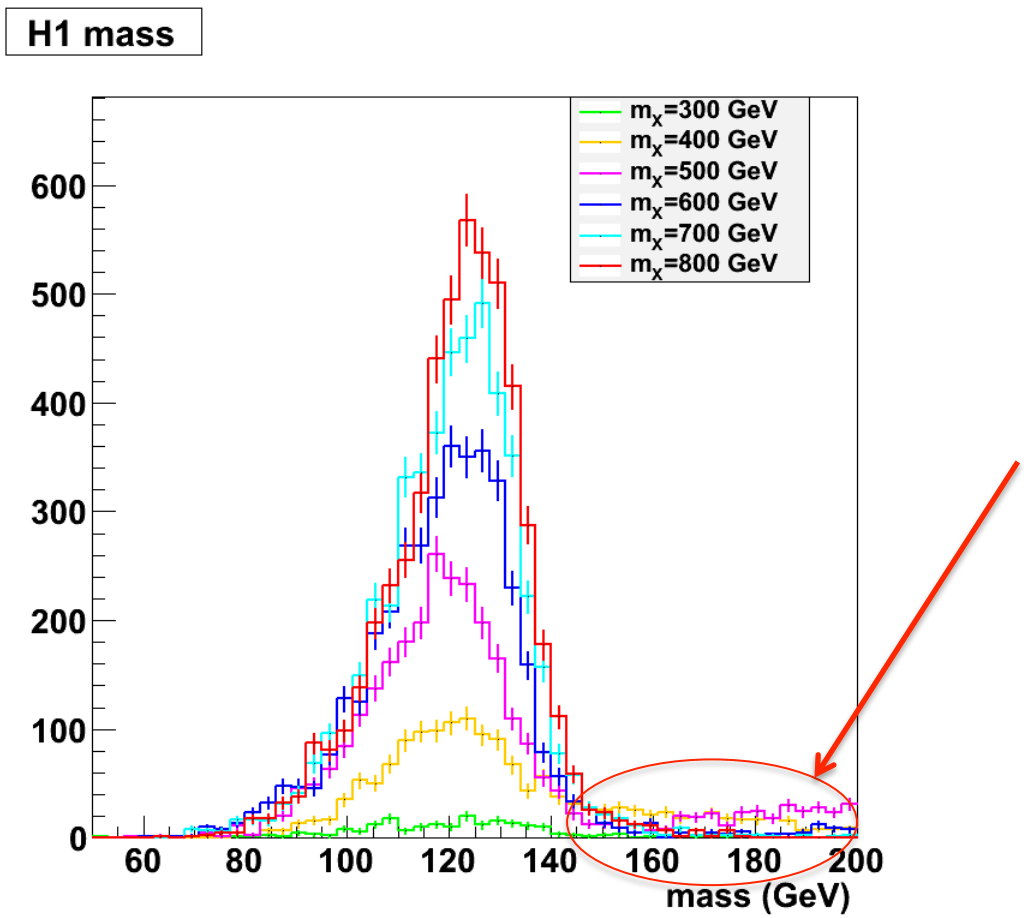


First 4 Hardest Central Jet pT Distributions



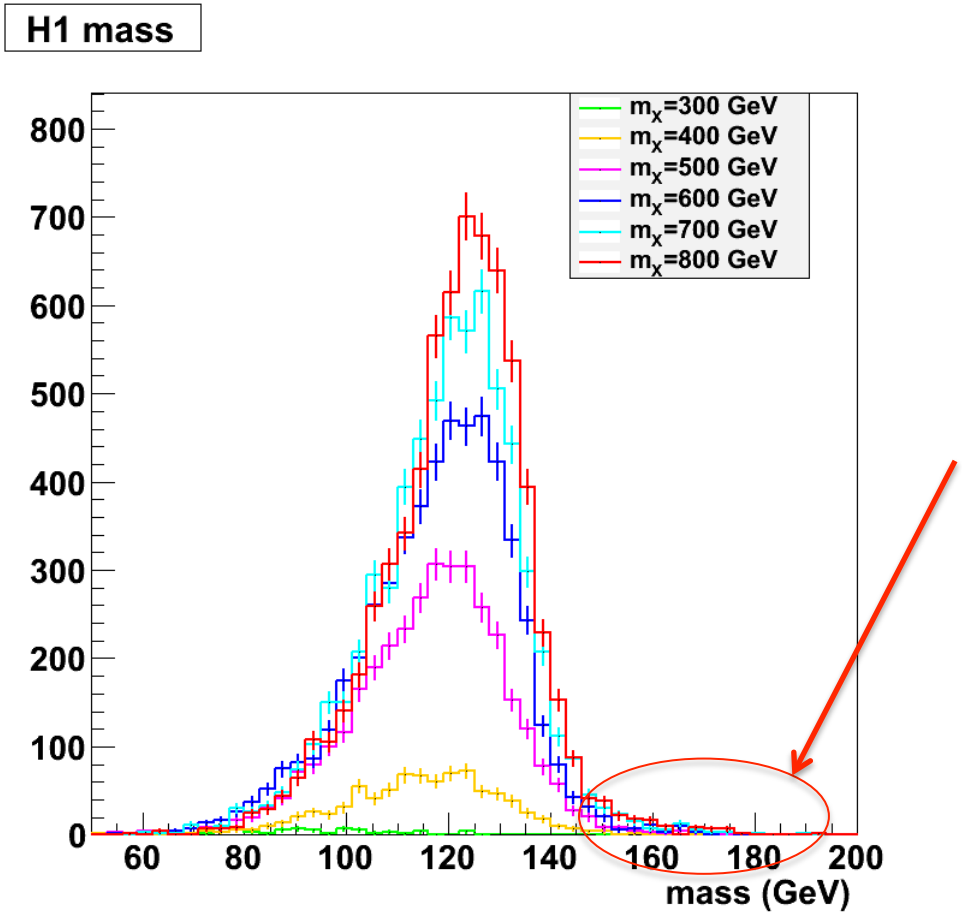
- First four highest pT central jets plotted after trigger, !(VH(bb)) and nCJets>3 criteria
- The **effect of the trigger** requiring 2 jets > 80 GeV and 4 jets > 30 GeV is apparent
- We do not use the hardest jets. We **use different combinatorial method**. Next slide.

Reconstructed Higgs Candidate Mass



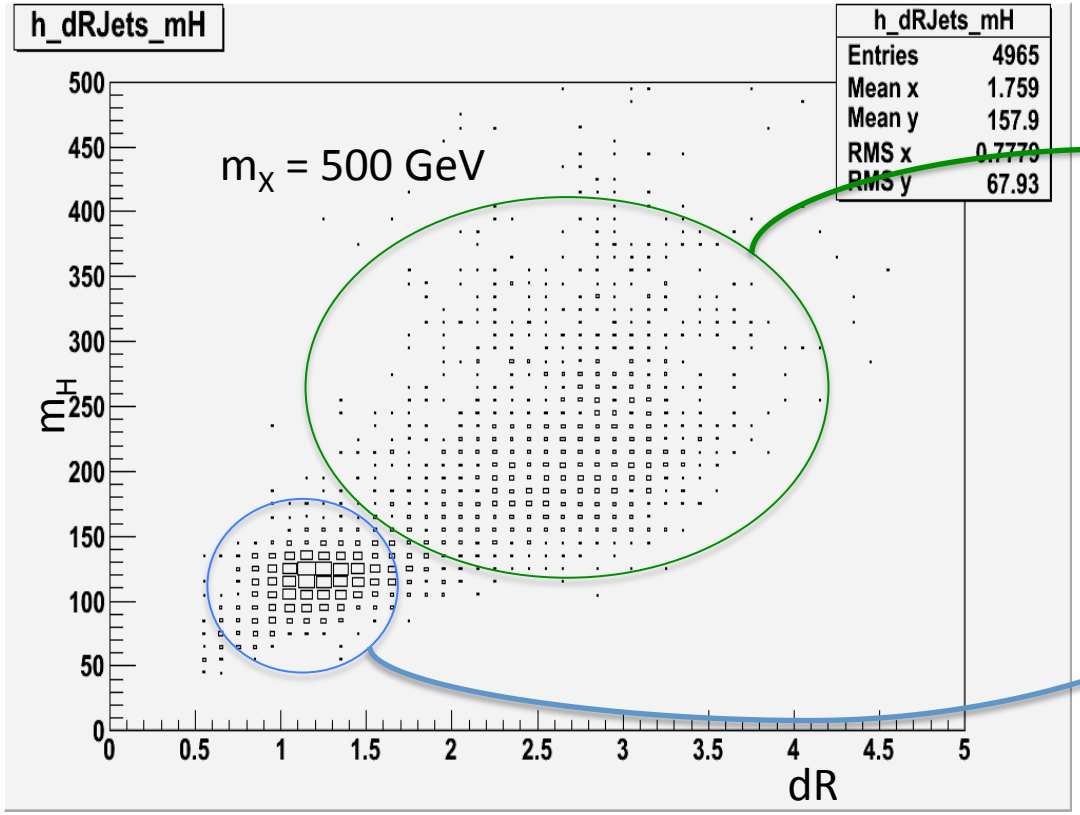
- The reconstructed m_H distribution exhibits this shoulder on the higher side.
- It does not diminish with tighter b-tagging of the jets.
- Prominent for mX < 600 GeV
- Where does it come from?

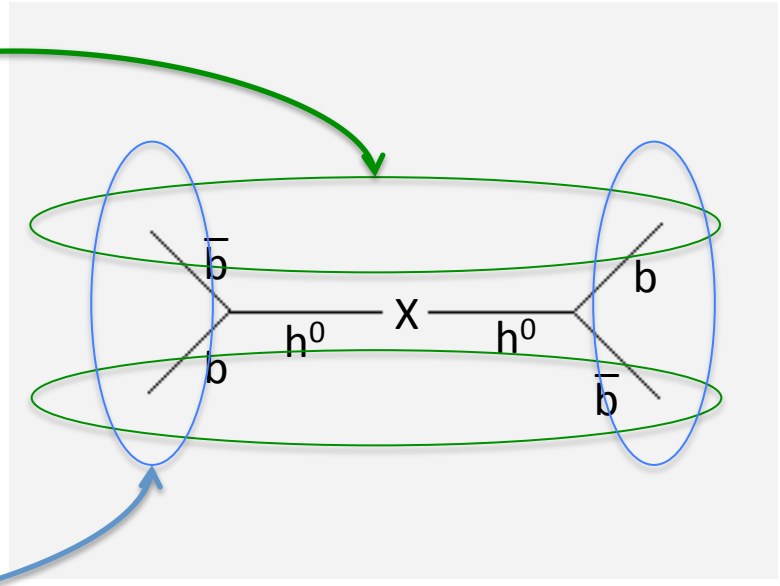
The Reconstructed Higgs Candidate Mass. dR < 1.7



- Once we require dR(jet, jet) < pi/2, the shoulder is gone!
- Now, we are ready to optimize the b-tagging working point.

The Reconstructed Higgs Candidate Mass vs dR(jets)





- To diagnose this shoulder, we plot a scatter between the reconstructed m_H and dR(jet, jet) within the Higgs candidate.
- We see two distinct populations of signal candidates. The ones with higher dR give us the shoulder.
- In order to **recover the event**, we only compare diJet pairs for mass-difference IF the dR(jet, jet) < pi/2. This doesn't work for mX=300 GeV.