Search for a SM Higgs Boson in the H→ZZ→2l2q Decay Channel

Francesco Pandolfi
Università di Roma 'Sapienza'
INFN Sezione di Roma

on behalf of the H→ZZ→2l2q Group

H→ZZ Subgroup Meeting
03.06.11

The Group

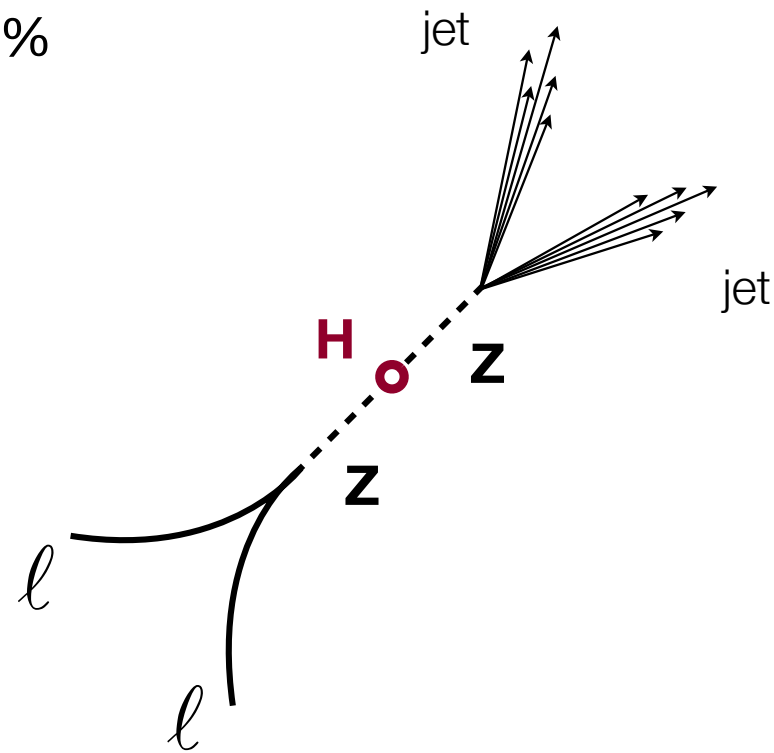
- Johns Hopkins University
  - A. Bonato, A. Gritsan, N. Tran, A. Whitbeck
- Purdue University
  - L. Borrello, D. Bortoletto, M. Kress, M. Vidal
- CERN
  - S. Bolognesi, M. Mannelli, M. Mozer
- UAM
  - G. Codispoti, J. F. de Troconiz
- University of Torino
  - R. Castello
- CIEMAT
  - J.P. Fernandez, P. Garcia-Abia, O. Gonzalez, J. Hernandez, M. Soares
- University of Rome
  - D. del Re, F. Pandolfi
- Brown University
  - M. Narain, M. Segala
- University of Naples
  - A. De Cosa, F. Fabozzi, L. Lista

Literature

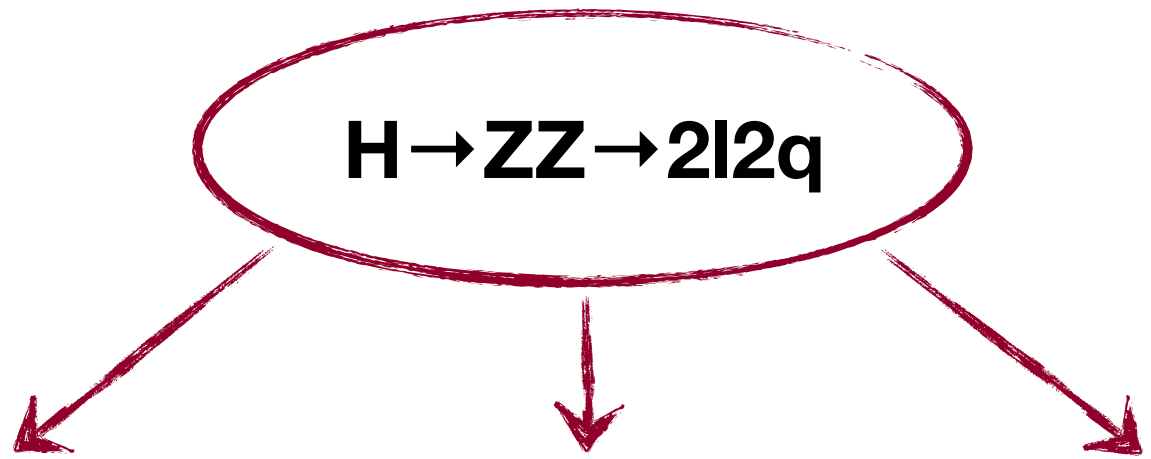
- Analysis:
  - ‘Search for a SM Higgs or BSM Boson H→ZZ→2l2q’, CMS AN-2011/100
  - ‘Search for the Standard Model Higgs Boson in the decay channel H→ZZ→2l2b’, CMS AN-2011/125
  - ‘Search for a SM Higgs Boson H→ZZ→2l2q at CMS’, CMS PAS HIG-11-006
- Angular discriminant:
  - ‘Angular Analysis of Resonances pp→X→ZZ’, CMS AN-2010/351
  - ‘Spin-determination of single produced resonances at hadron colliders’, http://arxiv.org/abs/1001.3396v2
- Quark-Gluon discriminant:
  - ‘Quark-Gluon Jet Discrimination through Particle Flow Jet Structure’, CMS AN-2011/215

The H→ZZ→2l2q Channel

- Large yields: exploit very large BR(Z→qq) = 70%
  - BR(ZZ→2l2q) = **20 ×** BR(ZZ→4l)
  - BR(ZZ→2l2q) = **3.5 ×** BR(ZZ→2l2ν)
- Drawbacks:
  - Bad resolution coming from jets
  - Large backgrounds coming from Z+jets
- But full decay is reconstructed, closed kinematics
  - Good for exclusion and discovery



Maximize Significance: Three Categories



2 b-tags

- One jet medium tagged, one jet loose tagged
- Purest category
- Lowest signal yield

1 b-tag

- !(2 b-tags) and one jet loose tagged
- Worst category
- Large background

0 b-tags

- !(2 b-tags) and !(1 b-tag)
- Highest signal yield
- Large background

Multiple Candidates?

- Many jets in event → multiple Higgs candidates
- **But** each event has to be counted in one category only
- Logic:
  - Look at jet pairs which pass selection
  - Choose jet pair with highest number of b-tags
  - If multiple, choose best Z mass pair
- Exclusive classification ensured

Monte Carlo Samples

- Signal:

  JHU (main): /SMHiggsToZZTo2L2Q_M-*_7TeV-jhu-pythia6/Spring11-PU_S1_START311_V1G1-v1

  Powheg: /GluGluToHToZZTo2L2Q_M-*_7TeV-powheg-pythia6/Spring11-PU_S1_START311_V1G1-v1

- BG:

  Z+jets: /Z*Jets_ptZ-*_TuneZ2_7TeV-alpgen-tauola/Spring11-PU_S1_START311_V1G1-v1

  Z+bb: /ZBB*JetsToLNu_TuneZ2_7TeV-alpgen-tauola/Spring11-PU_S1_START311_V1G1-v1

  Z+cc: /ZCC*JetsToLNu_TuneZ2_7TeV-alpgen-tauola/Spring11-PU_S1_START311_V1G1-v1

  ZZ: /ZZtoAnything_TuneZ2_7TeV-pythia6-tauola/Spring11-PU_S1_START311_V1G1-v1

  WZ: /WZtoAnything_TuneZ2_7TeV-pythia6-tauola/Spring11-PU_S1_START311_V1G1-v1

  WW: /WWtoAnything_TuneZ2_7TeV-pythia6-tauola/Spring11-PU_S1_START311_V1G1-v1

  tt: /TTTo2L2Nu2B_7TeV-powheg-pythia6/Spring11-PU_S1_START311_V1G1-v1

  tW: /TToBLNu_TuneZ2_tW-channel_7TeV-madgraph/Spring11-PU_S1_START311_V1G1-v1

Physics Objects

- Electrons:
  - VBTF80/95 working points
- Muons:
  - VBTF selection
  - Relative combined iso < 15%
  - $d_{xy} < 0.02$, $d_z < 1$
- Jets:
  - Anti-k_T 0.5 PFJets
  - Pile-up subtraction with FastJet
  - Minimal jet ID (CHF > 0%)

Studied possibility of correcting isolation with FastJet
Not a big effect, not in this talk

CiC Electrons?

will study the possibility
expected small effect:
leptons not main concern

Preselection and Triggers

- Preselection:
  - Two opposite-charge leptons
  - Lepton $p_T > 20$ GeV and fiducial η
  - At least two jets with $p_T > 30$ GeV and $|\eta| < 2.4$

- Triggers:
  - Will run on HZZ skim
  - Could stick to only single lepton triggers

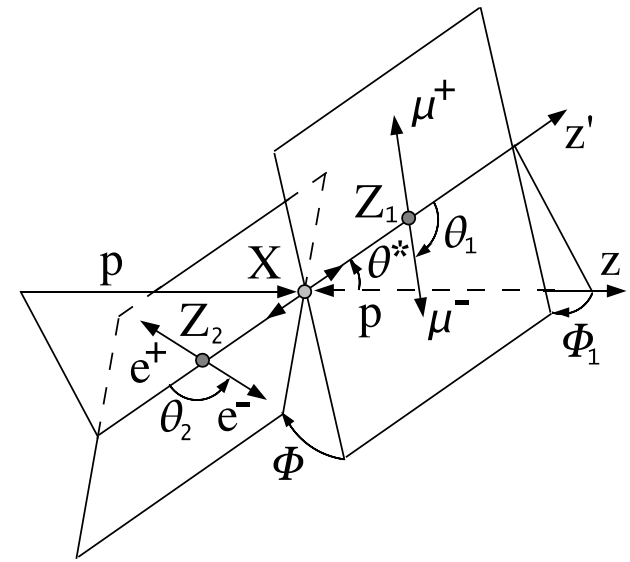
	MC-truth expected efficiency		
HLT Path	**m_H = 300**	**400**	**500**
HLT_IsoMu17	94.9±0.2%	94.9±0.2%	94.3±0.2%
HLT_Mu30	97.5±0.2%	98.1±0.1%	98.2±0.1%
HLT_DoubleMu7	90.7±0.3%	91.2±0.3%	90.7±0.3%
HLT_Ele27_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_v2	95.9±0.2%	96.7±0.2%	96.9±0.2%
HLT_Ele45_CaloIdVT_TrkIdT_v2	91.4±0.3%	94.8±0.1%	95.9±0.2%
HLT_Ele17_CaloIdL_CaloIsoIVL_Ele8_CaloIdL_CaloIsoIV	86.9±0.4%	87.7±0.3%	88.0±0.3%

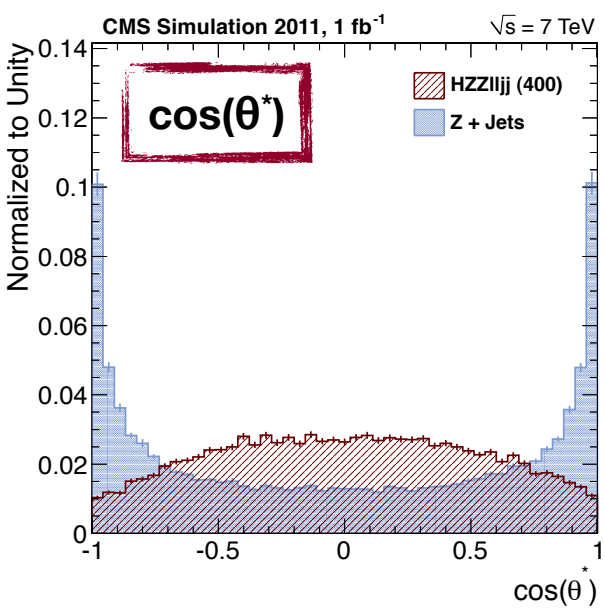
The Analysis Strategy At A Glance

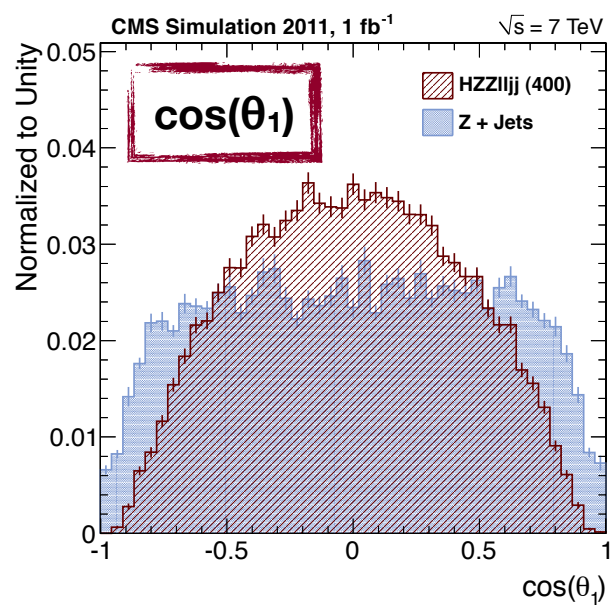
- **Main background:** Z+jets
  - Random jets emulate Z→qq by having invariant mass compatible with m_Z
  - Cross section more than **~10^5** times larger than signal
- Handles:
  - Higgs sets the scale: signal events are harder
  - Higgs decay (spin 0) follows established PDF's: angular analysis
  - Jet flavor: Z→qq is democratic, background enriched in gluons and u,d
- **Other backgrounds:**
  - top: important in 2-btag category. Further handle: missing E_T
  - diboson: (almost) irreducible, but small

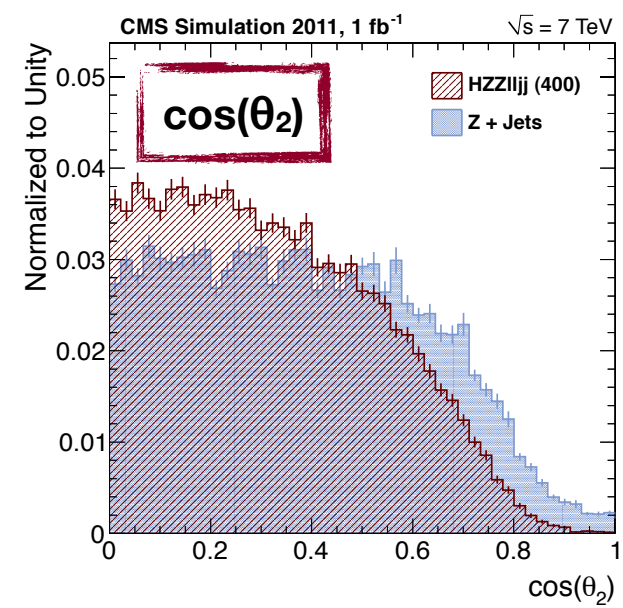
Helicity (and Production) Angles

- Final state kinematics univocally determined by 5 angles
- Different distributions between signal and non-resonant backgrounds

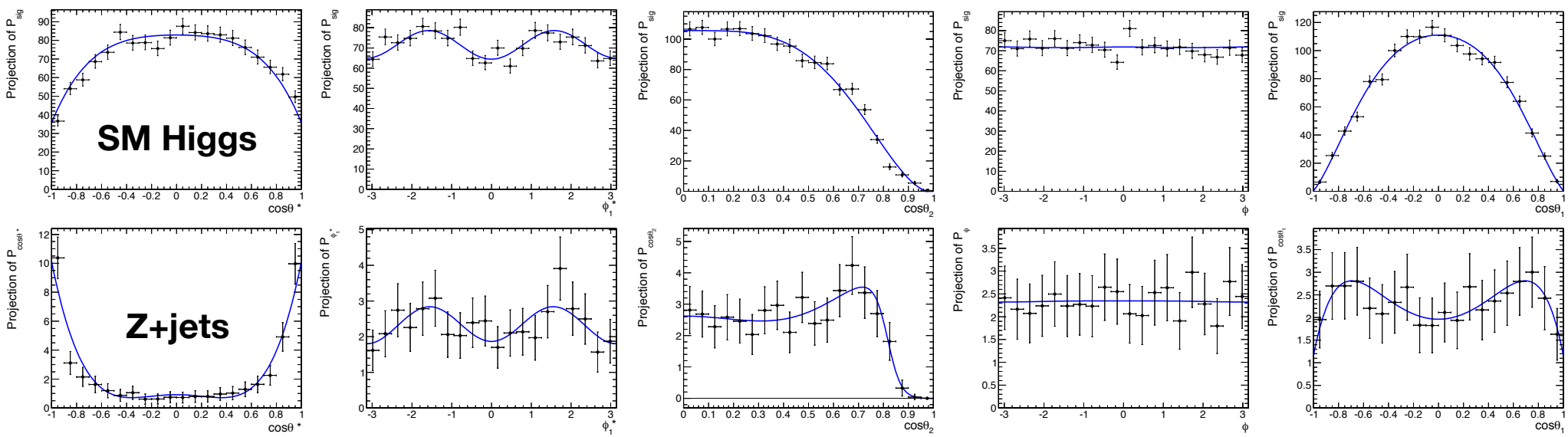


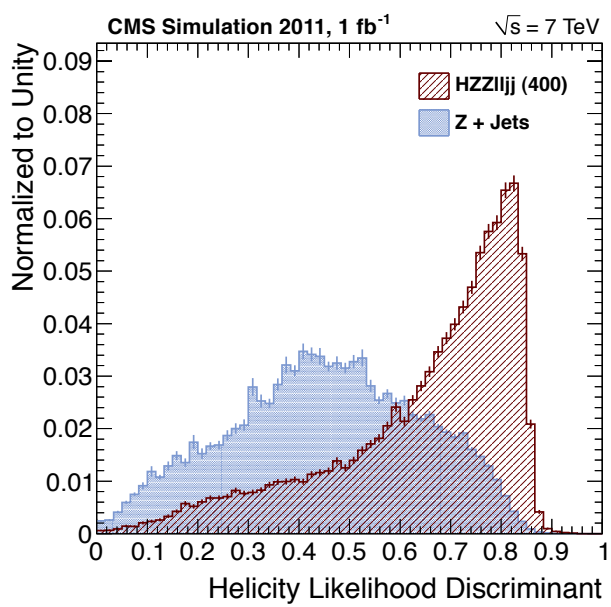






Helicity Likelihood Discriminant



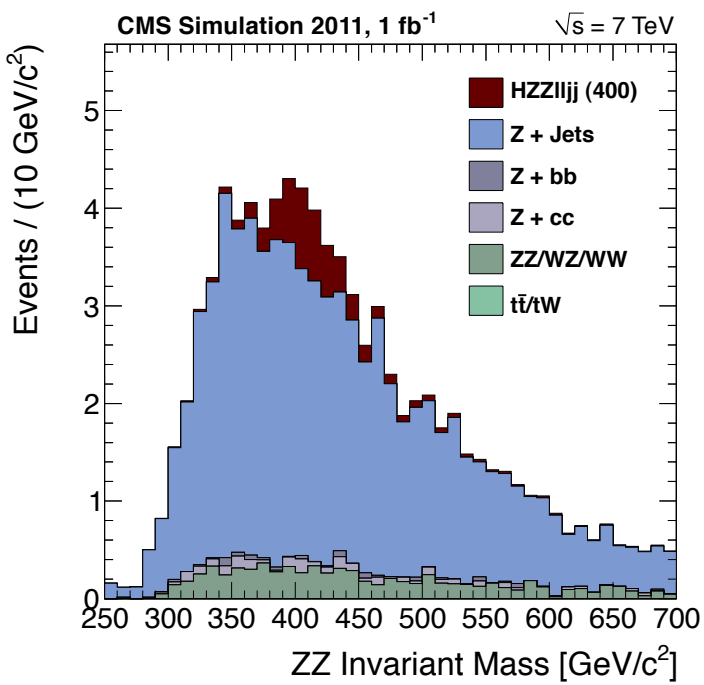


5D PDF for signal and BG used to compute LD

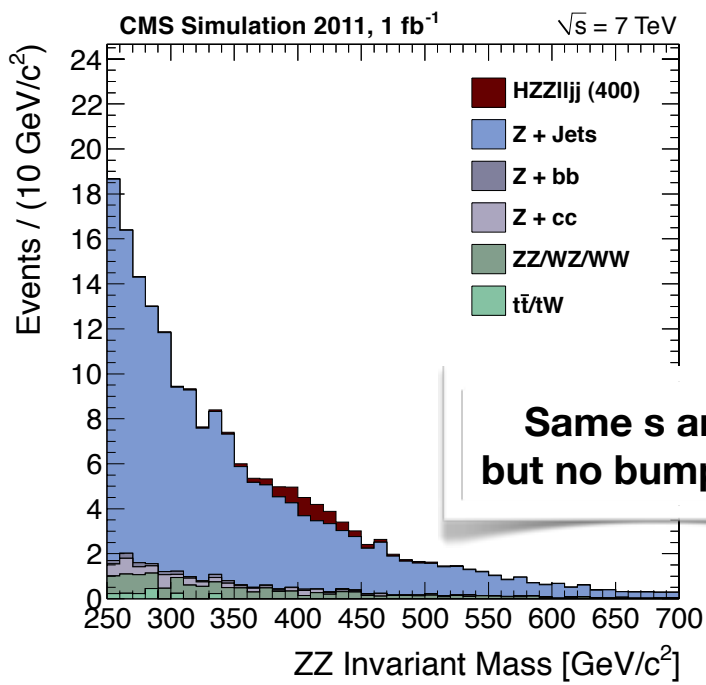
Angular vs. Kinematic Selection

- **Two** big advantages of using angular selection:
  - Works also at low mass, where not many kinematic handles
  - BG shape preserved: sidebands!

Kinematic Cuts



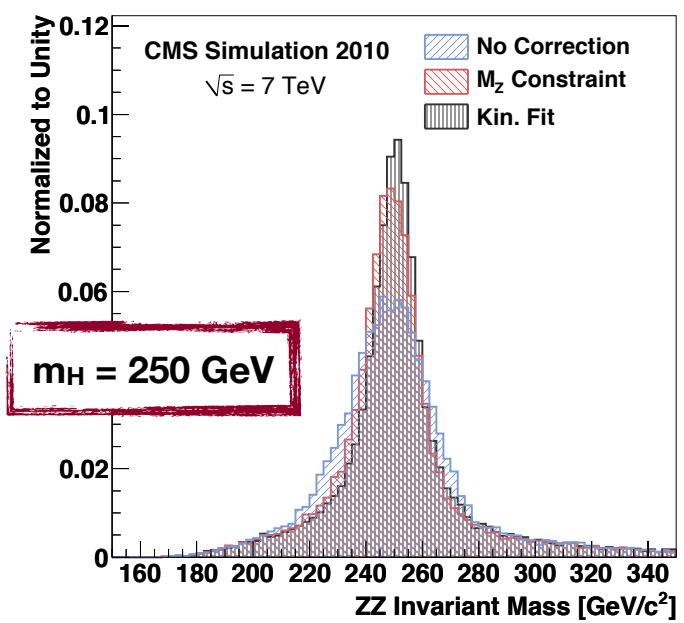
HelicityLD Cut

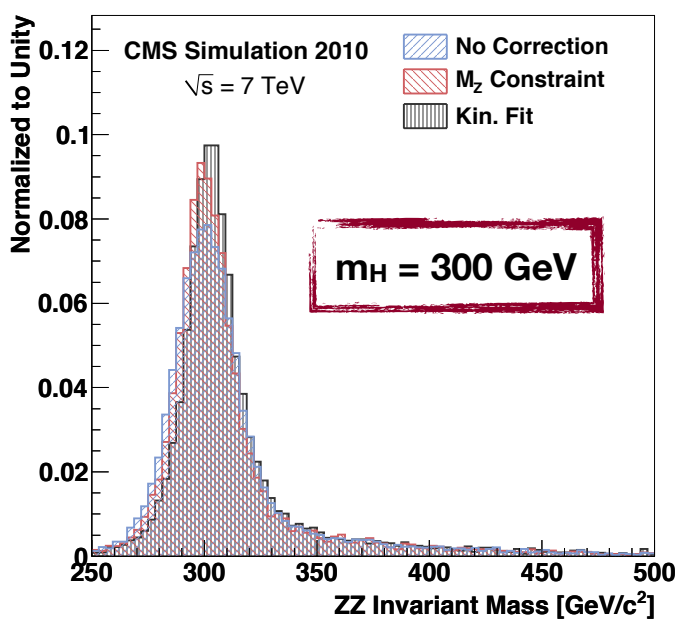


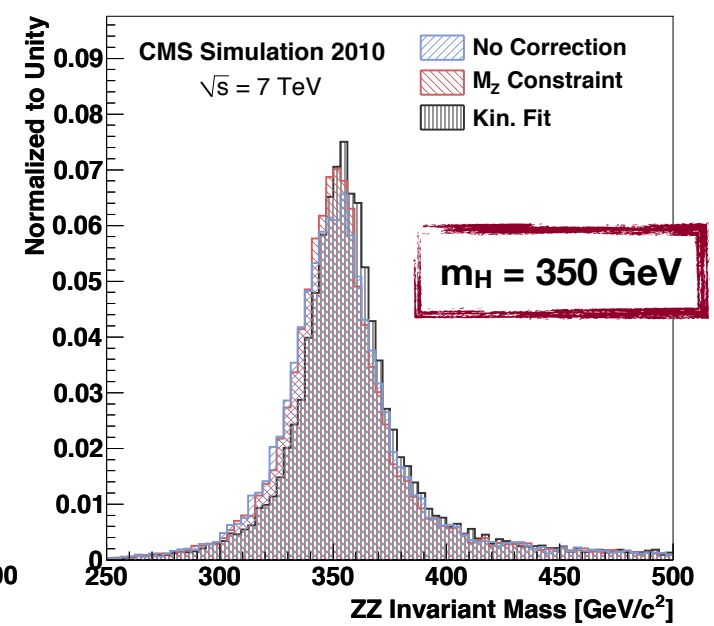
Same s and b, but no bump in BG

Dealing with Jet Resolution: Kinematic Fit

- Hadronic Z mass width ~15 GeV due to jet resolutions
- A kinematic fit can force m_Z weighting jets with expected resolutions
  - Improved m_H resolution in signal
  - Reshuffling of events in backgrounds
- Cannot expect much for $m_H > 350$ GeV: Higgs intrinsic width too large



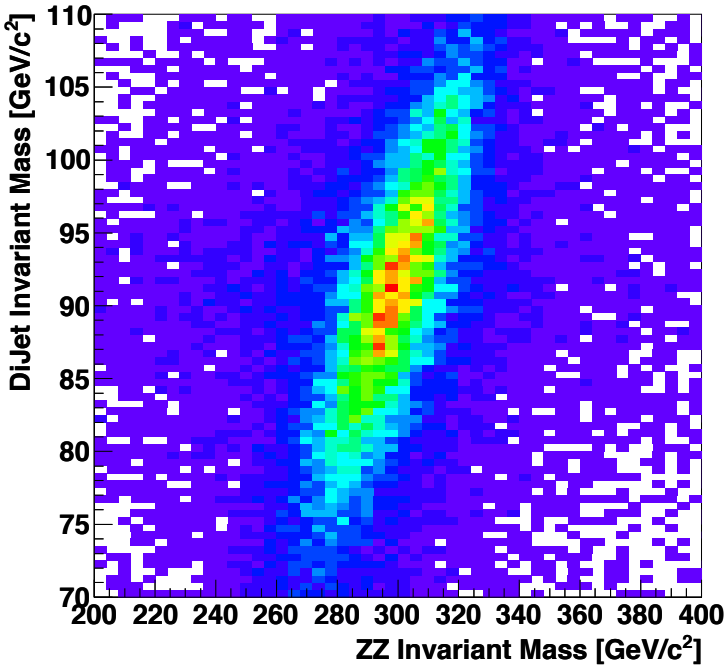




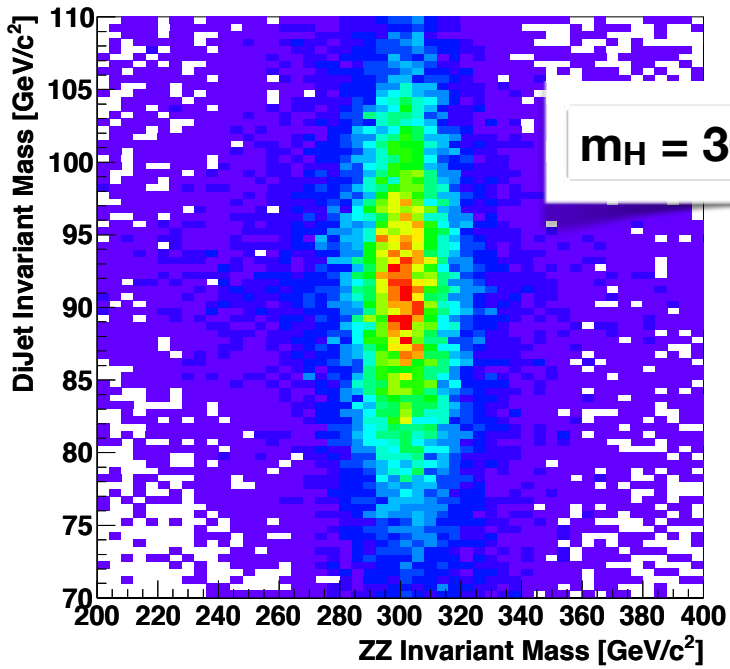
KinFit Removes m_{jj}-m_H Correlation

- Removal of correlation between dijet and ZZ reconstructed masses
- Simple definition of signal box and sidebands

Before Kin Fit



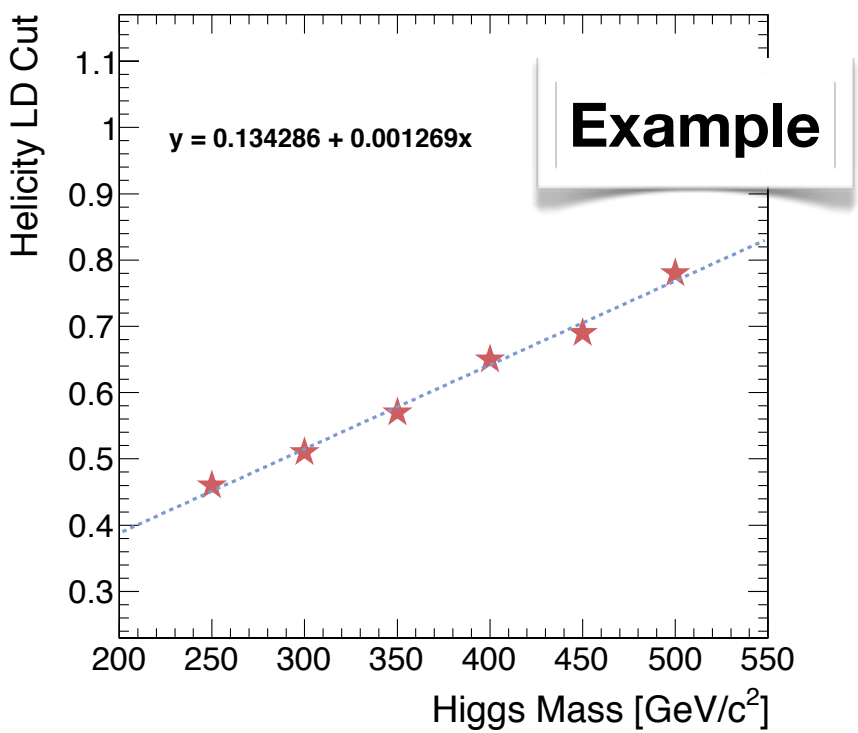
After Kin Fit



m_H = 300 GeV/c^2

Optimization in Categories

- Used TMVA toolkit to find optimal thresholds on helicityLD
  - Separately optimized 3 categories at all mass points
  - Chose threshold which minimizes UL
- Found smooth variation of threshold vs m_H



Cut: HelicityLD > a · m_H + b

	a [GeV^{-1}]	b
0-btag	0.00025	0.55
1-btag	0.000656	0.302
2-btag	0	0.5

looser cuts in purer categories

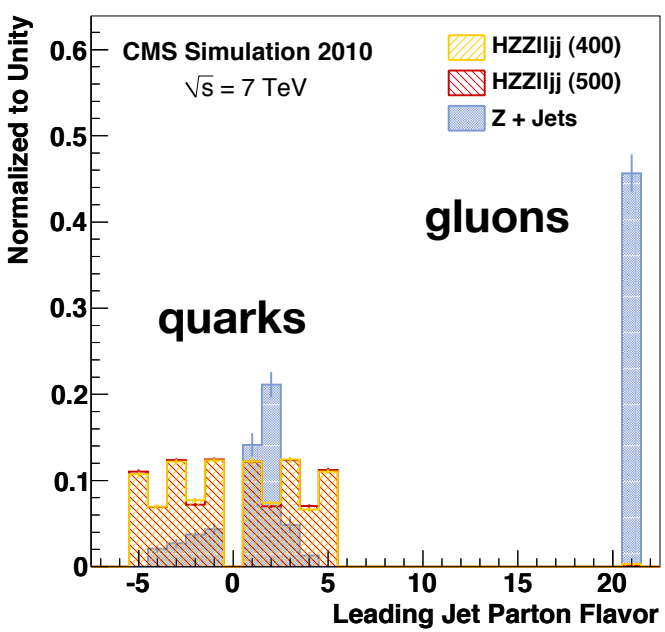
Jet Flavor as Means of Discrimination

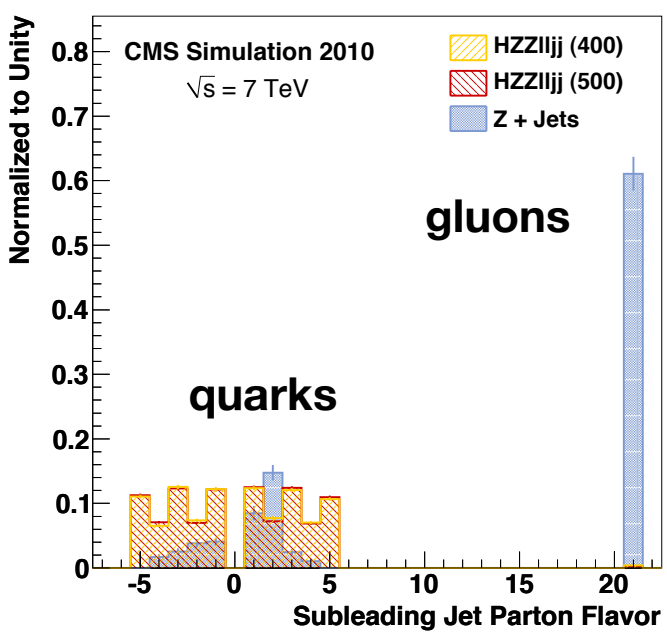
- Jets in **signal** democratic about quark flavour
- 45% (60%) of leading (subleading) jets in **background** are gluon jets

Two handles:

b-tagging
only ~1/5 of signal, but better s/b

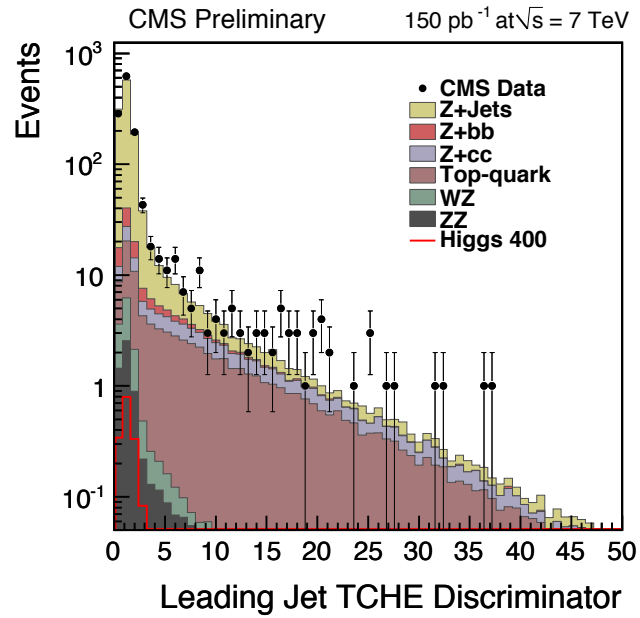
Quark-Gluon discrimination
remaining BG has large gluon component. can we isolate it?

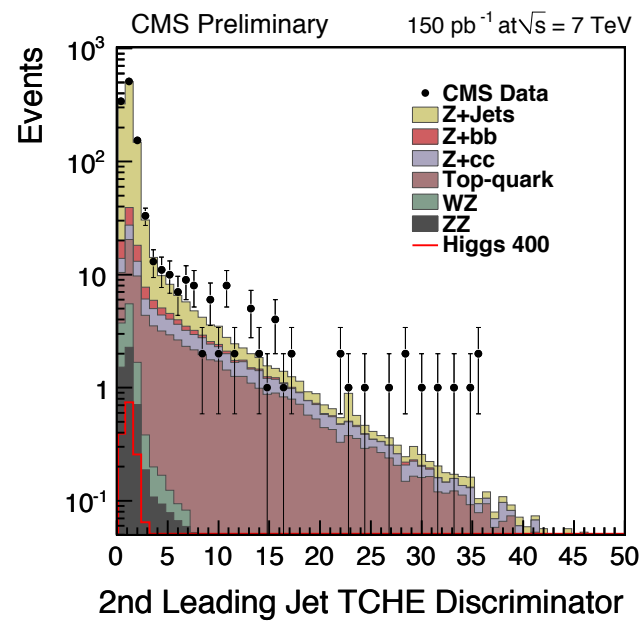




B-Tagging Details

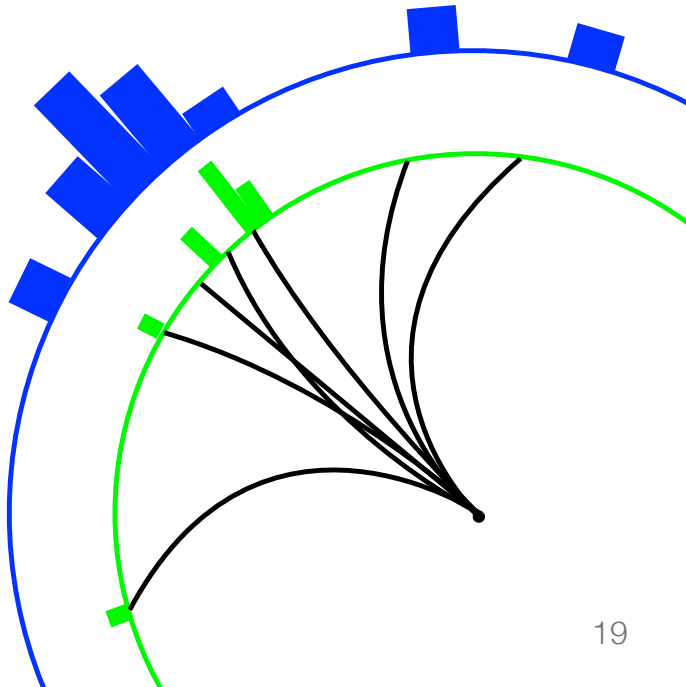
- TrackCountingHighEfficiency (TCHE) tagger
- BTV recommended thresholds:
  - Loose: TCHE > 1.7
  - Medium: TCHE > 3.3
- Applying scale factors (SF) to take into account loss in efficiency in data
  - SF(real b) = 90%
  - SF(mistag) ~ 1.1 (p_T and η dependence)
  - On a jet-by-jet basis: no events lost, just migrating from one category to other



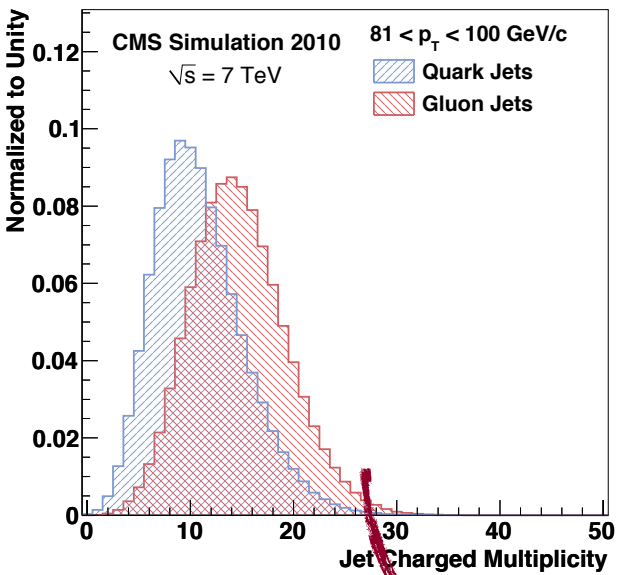


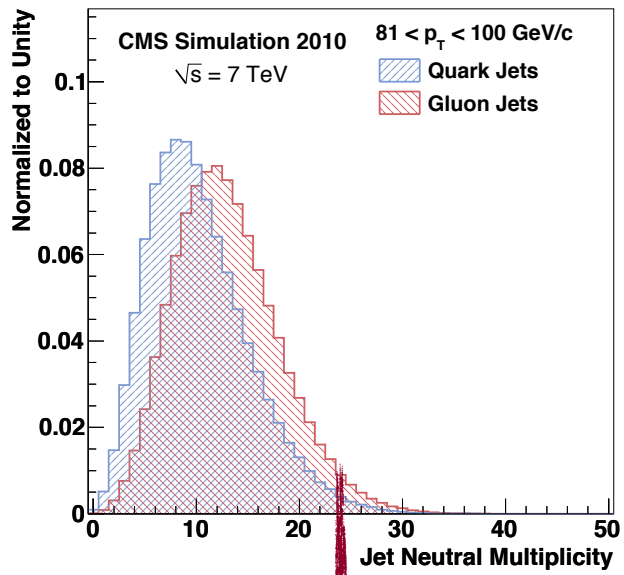
Quark-Gluon Discrimination: The Idea

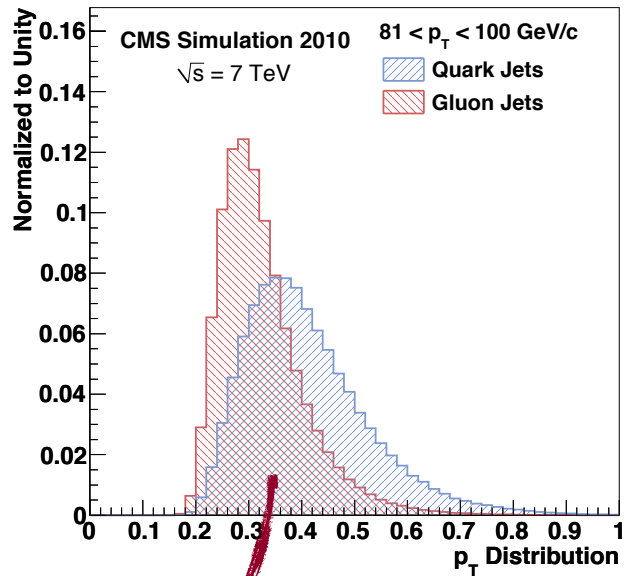
- Gluons are more coupled to the strong field wrt quarks
  - Gluon hadronization produces more particles (and at wider angles)
- Plus: gluon splitting can yield jets with marked substructure (2 quark subjets)
- Should mirror in multiplicity and composition variables
- Considered three variables:
  - PF charged multiplicity
  - PF neutral multiplicity
  - p_T distribution among PFCandidates in jet

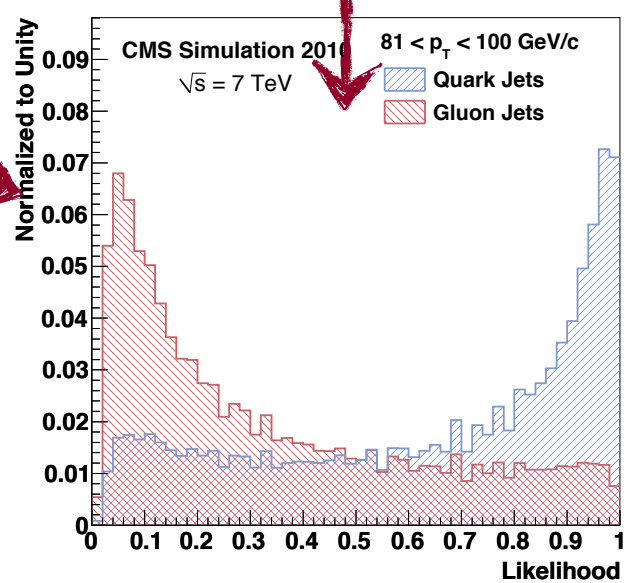


Building the QGLikelihood Discriminator









PDFs obtained on QCD

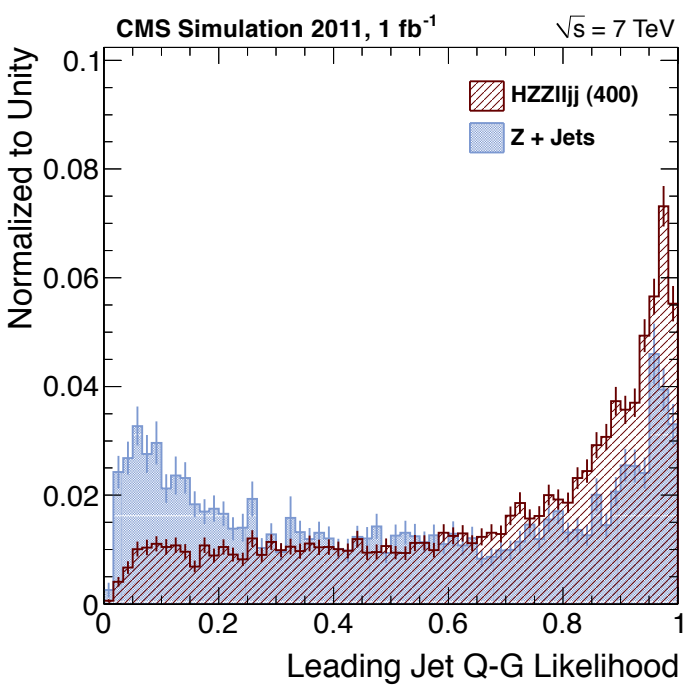
Three variables combined in an independent likelihood

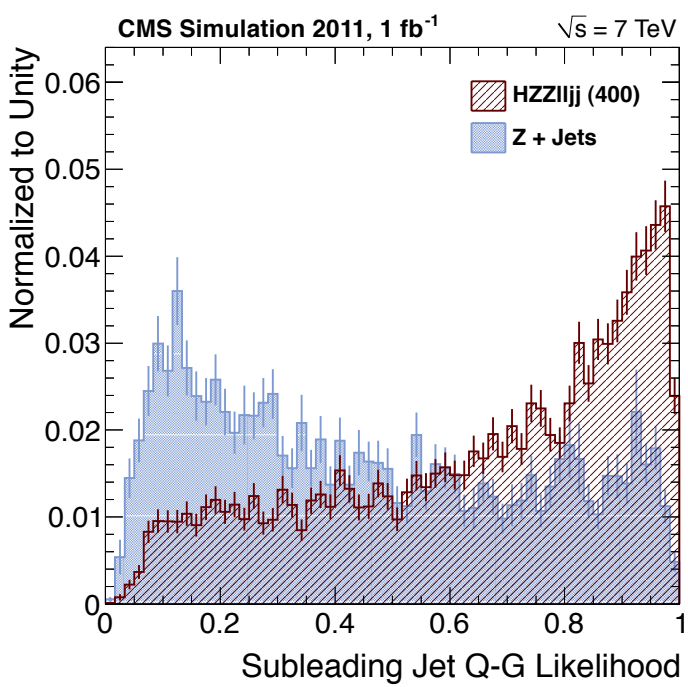
Complementary to B-Tagging

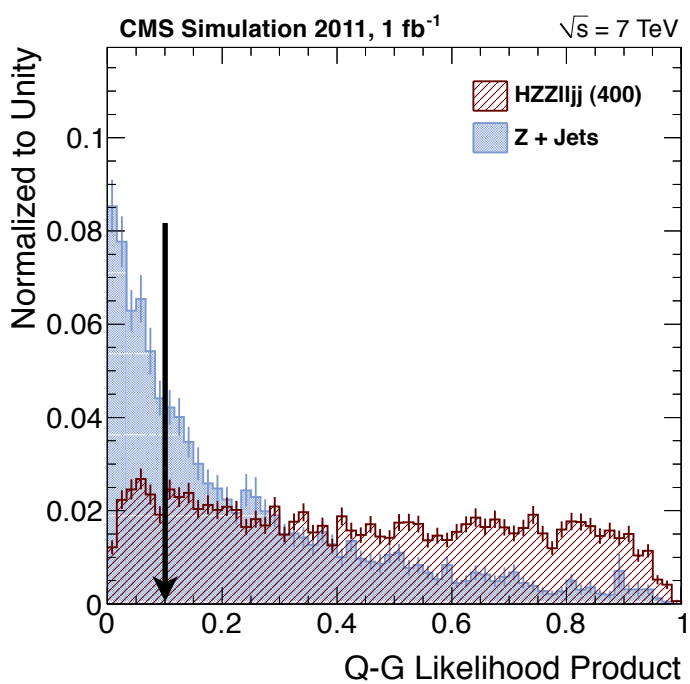
- b quarks hadronize “like gluons” (already seen by OPAL)
  - Would be cut away by Q-G discriminator
- **Cut:** (QGJet1*QGJet2) > 0.1 in 0-btag category only
  - Cuts away ~half of the BG, small effect on signal

Cut Efficiency

Mass	Signal	Z+Jets
300	86%	57%
400	85%	50%
500	87%	44%

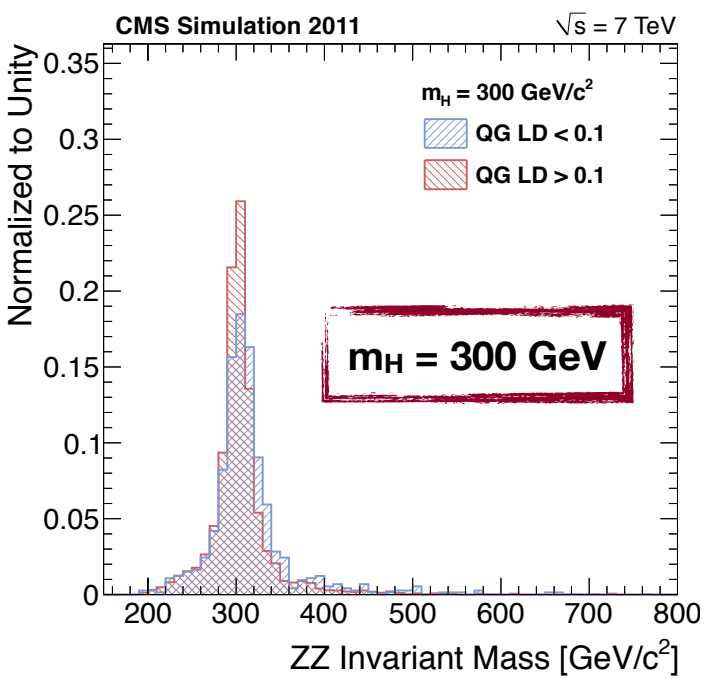


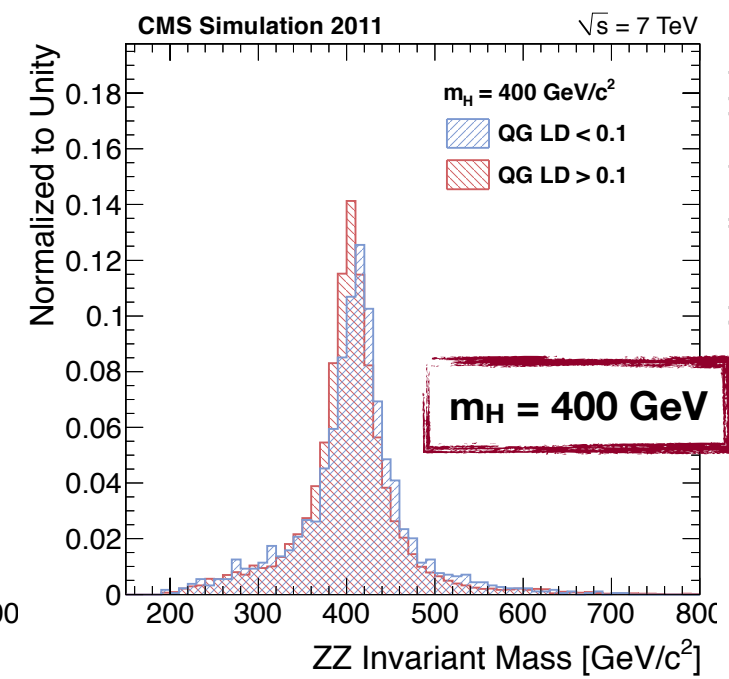


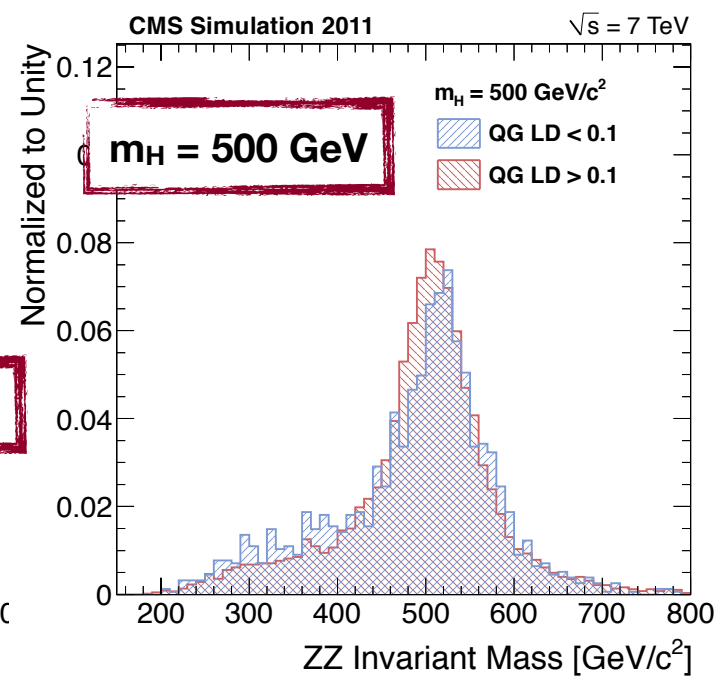


Not Only Background Discrimination

- QGLikelihood cut also cuts away the tails in signal
  - Misreconstructed jets
  - (Small) residual self-combinatorics
- Cut **improves** ZZ mass resolution

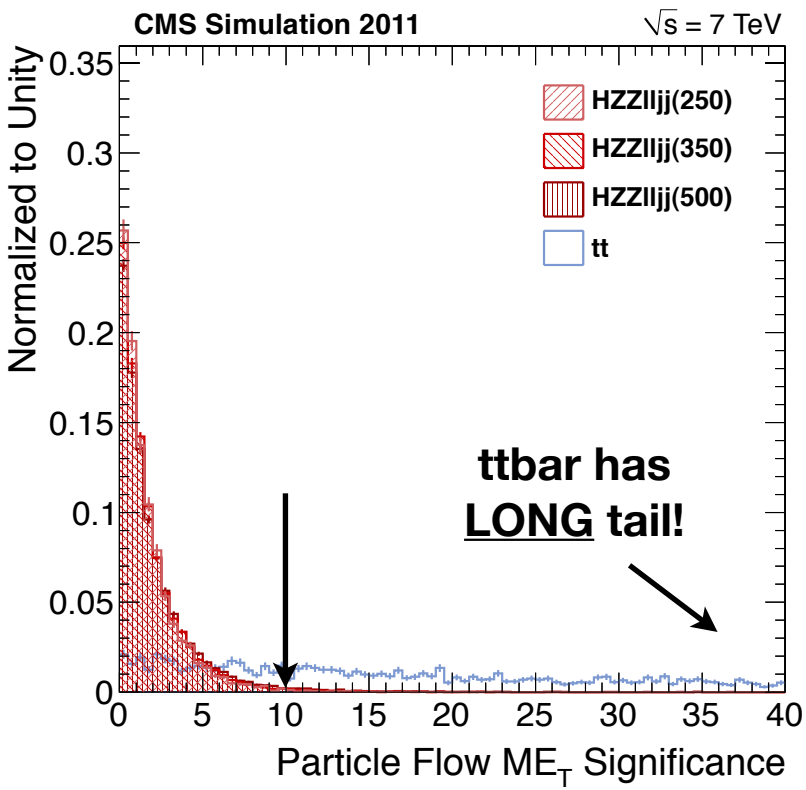






Dealing with Top in the 2-tag Category

- Large infiltration of top background in 2-tag category
  - Can reduce it by looking at missing E_T
- **But** cutting on pfME_T introduces inefficiencies at large signal mass



- Cut on pfME_T significance (require < 10)
  - Computed from PFJet resolutions
- Losing small fraction of signal (2-3%)
- Gain factor 2-3 rejection power in top BG

Final Analysis Selection

0-btag	1-btag	2-btag
Lepton ID + lepton p_T > 40/20 GeV		
70 < Dilepton invariant mass < 110 GeV		
75 < Dijet invariant mass < 105 GeV		
Kinematic fit applied to jets		
HelicityLD > $0.00025 \cdot m_H + 0.55$	HelicityLD > $0.000656 \cdot m_H + 0.302$	HelicityLD > 0.5
QGLikelihood Product > 0.1		$pfME_T$ Significance < 10

Expected Yields @ 1 fb^{-1}

0-btag:

Mass [GeV]	Signal	Z+Jets	Z+cc	Z+bb	Diboson	tt/tW	Total Background
250	4.5 (4.2%)	1e+02	2.7	1.1	5.6	1.4	1.2e+02
300	4.3 (5.4%)	77	1.3	0.63	3.9	0.55	83
350	4.8 (6.6%)	48	0.57	0.37	2.7	0.054	52
400	3.5 (6.3%)	27	0.27	0.23	1.5	0.036	29
450	2.1 (5.7%)	15	0.18	0.13	1.1	0.018	16
500	1.2 (5.2%)	8.5	0.076	0.05	0.68	0.018	9.3

1-btag:

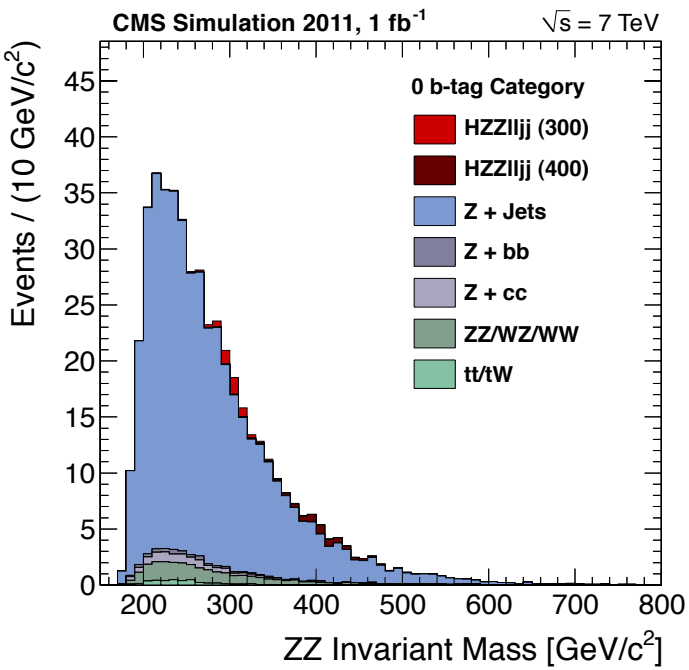
Mass [GeV]	Signal	Z+Jets	Z+cc	Z+bb	Diboson	tt/tW	Total Background
250	3.4 (3.2%)	80	9.2	9.1	5.2	15	1.2e+02
300	3.1 (3.9%)	58	2.9	4.6	3.3	6.3	75
350	3.7 (5.1%)	42	1.4	2.2	2.1	1.8	50
400	2.9 (5.2%)	27	2	0.9	1.4	0.56	32
450	1.8 (5%)	18	0.36	0.49	0.88	0.25	20
500	1 (4.4%)	11	0.27	0.16	0.81	0.091	12

2-btag:

Mass [GeV]	Signal	Z+Jets	Z+cc	Z+bb	Diboson	tt/tW	Total Background
250	1.6 (1.5%)	1.9	0.48	3.4	0.83	2.5	9.1
300	1.7 (2.1%)	1.8	0.45	1.9	0.64	1.3	6.2
350	2 (2.8%)	1.5	0.14	1.3	0.43	0.38	3.7
400	1.6 (2.9%)	1.1	0.16	0.75	0.29	0.054	2.4
450	1.1 (2.9%)	0.92	0.11	0.49	0.26	0.036	1.8
500	0.65 (2.8%)	0.67	0.076	0.31	0.18	0.036	1.3

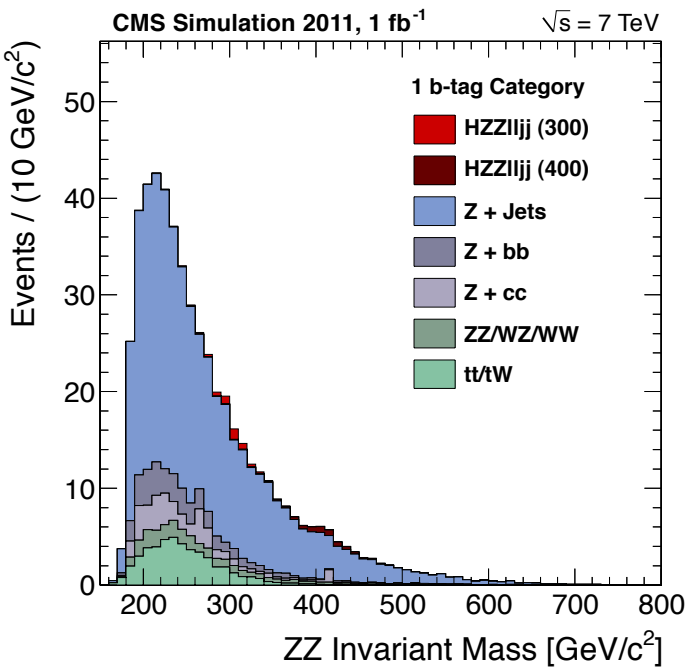
Expected Invariant Mass Spectra @ 1 fb^{-1}

0-btag



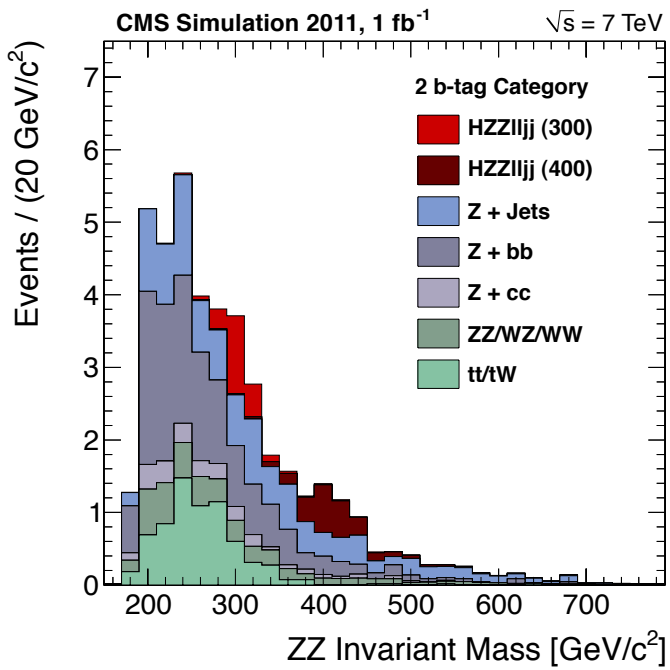
- High signal yield
- Large background

1-btag



- Worse performance

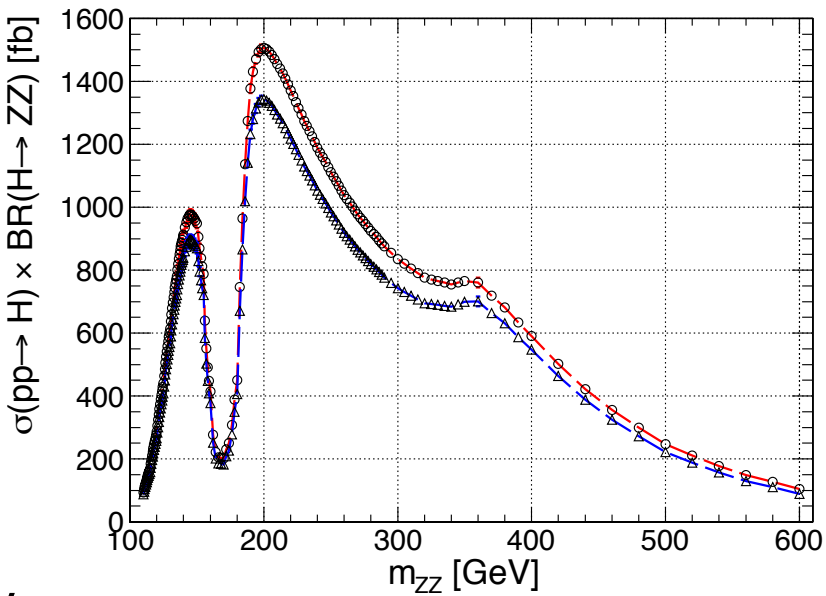
2-btag



- Purest category
- Low signal yield

Statistical Treatment of Results

- We have a total of 6 channels
  - 3 btag categories × 2 lepton flavours
- Combination: HiggsCombination tool
  - Bayesian (flat prior), MCMC
- Sensitive in high mass region: 200-600 GeV
  - Will provide results for 73 mass points
- Two independent approaches: shape analysis (main) and cut&count (check)



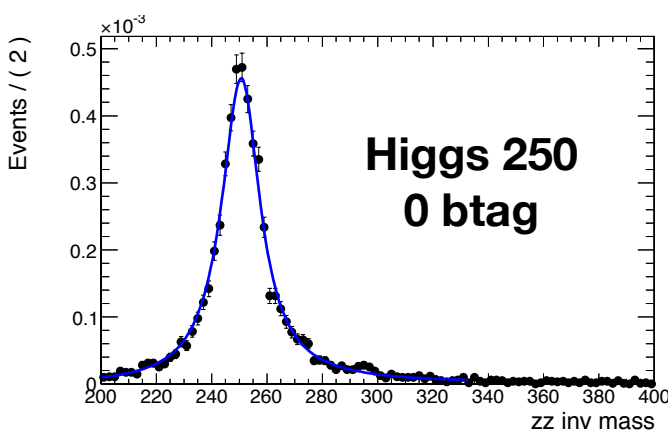
Cut&Count vs. Shape Analysis

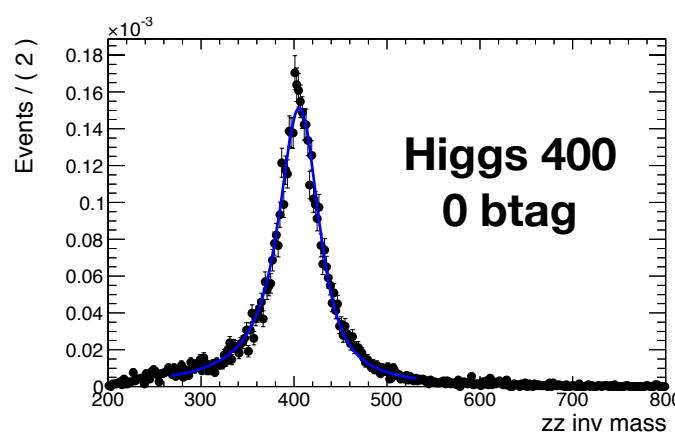
Shape Analysis

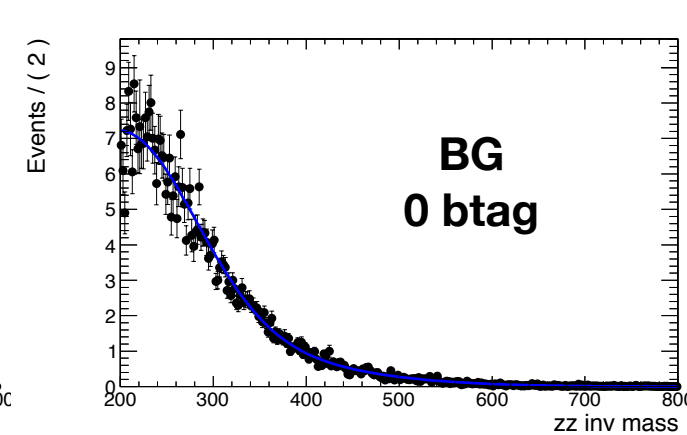
- Used for all 73 mass points
- Likelihood fit to data
  - Signal: BW ⊗ CB
  - BG: empirical
- Smooth parametrization vs. mass

Cut and Count

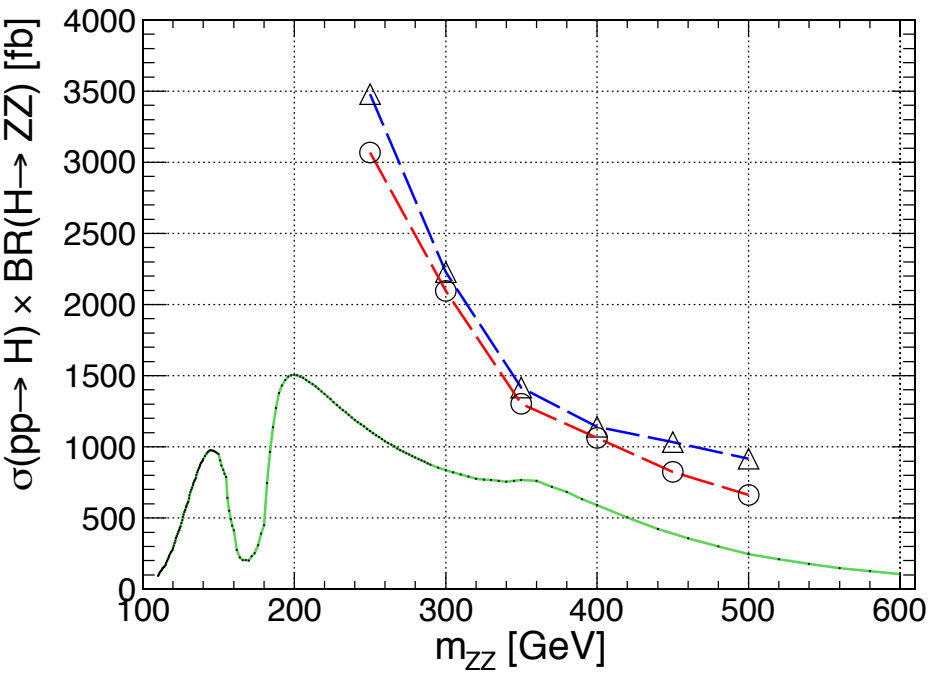
- Check at 6 pivotal mass points
- Count in m_H -6/+10%
  - Maximizes significance

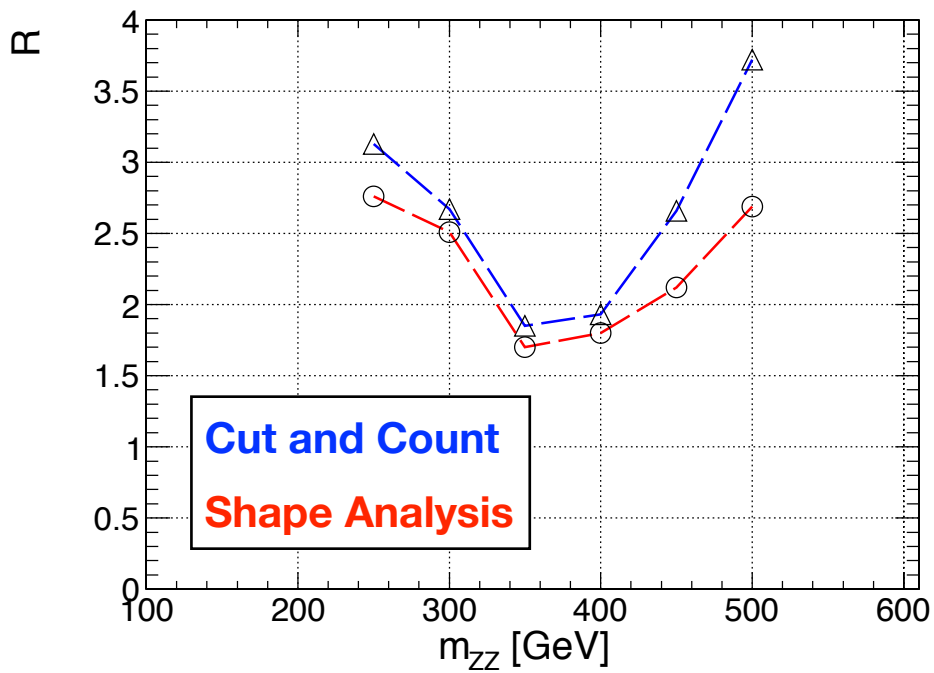






Shape Analysis Gives 15% Improvement





- Here statistical only uncertainties
- Shape analysis gives better results
- Analysis 10% better than Higgs Review (and lower expected syst)

UL/SM with no syst.

m_H	Higgs Review	This Analysis
300	2.8	2.51±0.03
400	2.1	1.80±0.02
500	3.1	2.69±0.03

Systematic Uncertainties

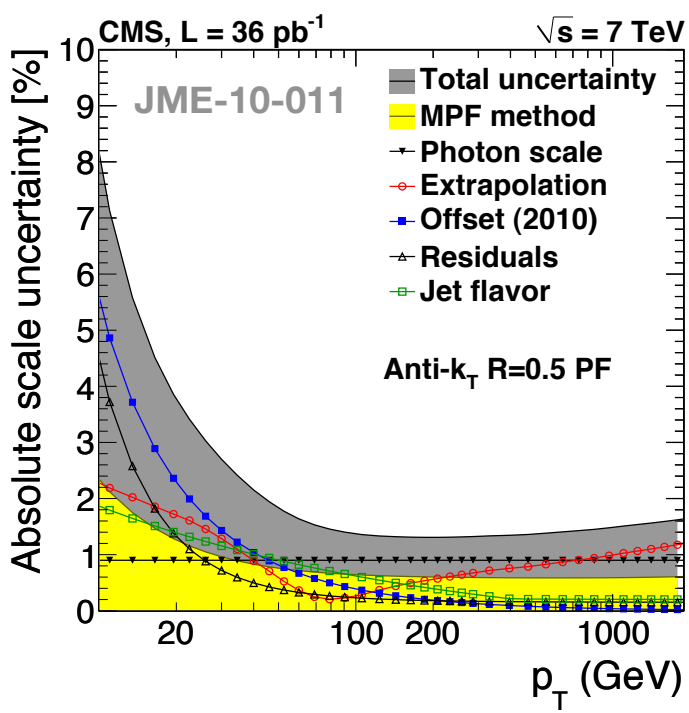
Systematic Uncertainties: Summary

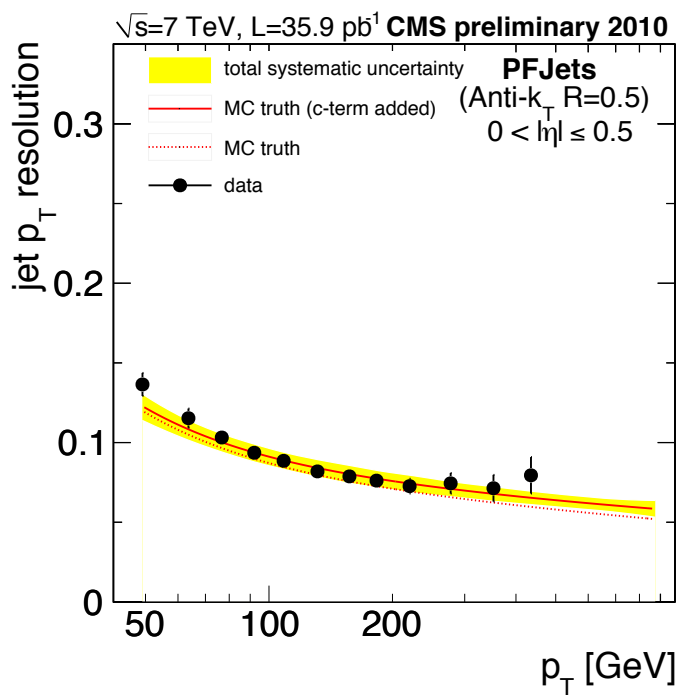
red: syst on signal efficiency **blue:** syst on background

	0-btag	1-btag	2-btag
jets	1.5%	1.5%	1.5%
background	(5%)	(5%)	30%
leptons and HLT	< 3%	< 3%	< 3%
pile up	8%	8%	8%
btag	~3% ~1.5%	~2% ~5-10%	~25% ~20%
luminosity	4%	4%	4%
signal cross section	17%	17%	17%
production mechanism	2%	2%	2%
ME_T	--	--	discussion on hn-met
QG	work in progress	--	--

Uncertainties on Jets

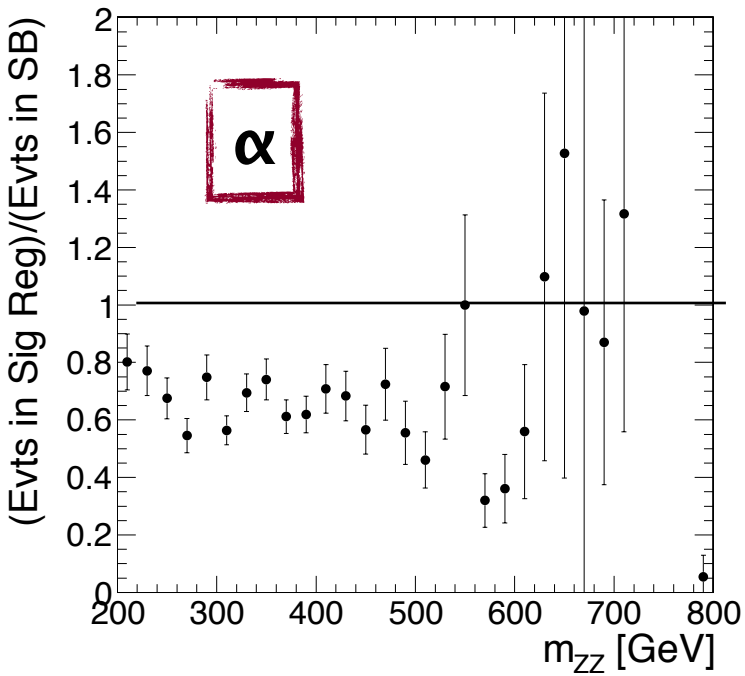
- Jet energy scale and resolution:
  - Re-ran analysis by varying jet p_T within uncertainties (±1 σ)
  - Re-ran analysis by smearing jet resolution by data/MC
- Total effect: 1.5% on signal efficiency (kinematic fit minimizes effect)

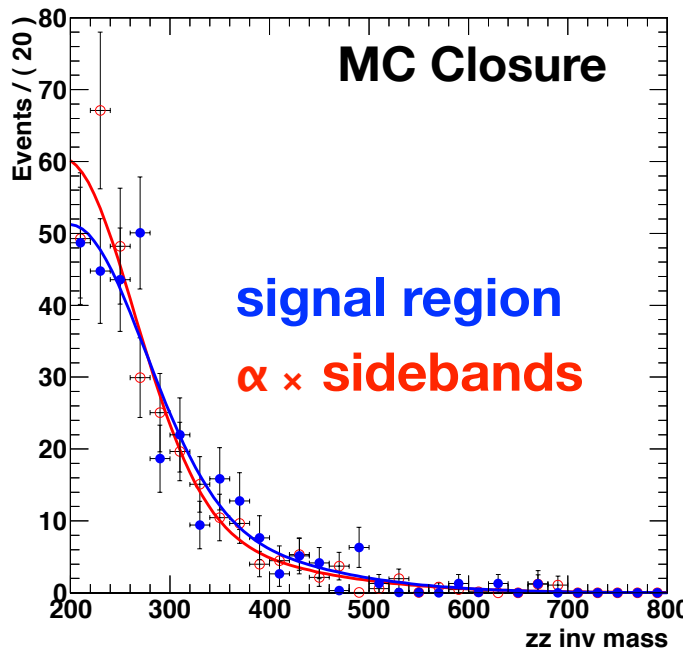




Background Control: Sidebands

- Will use sidebands in 0- and 1-tag category (high stat):
  - Invert m_{jj} cut and look at m_{ZZ} distribution
  - From MC: **α** = bin-by-bin ratio of (m_{ZZ} sideband)/(m_{ZZ} signal region)
  - BG estimate in signal region = **α** × sideband m_{ZZ} distribution





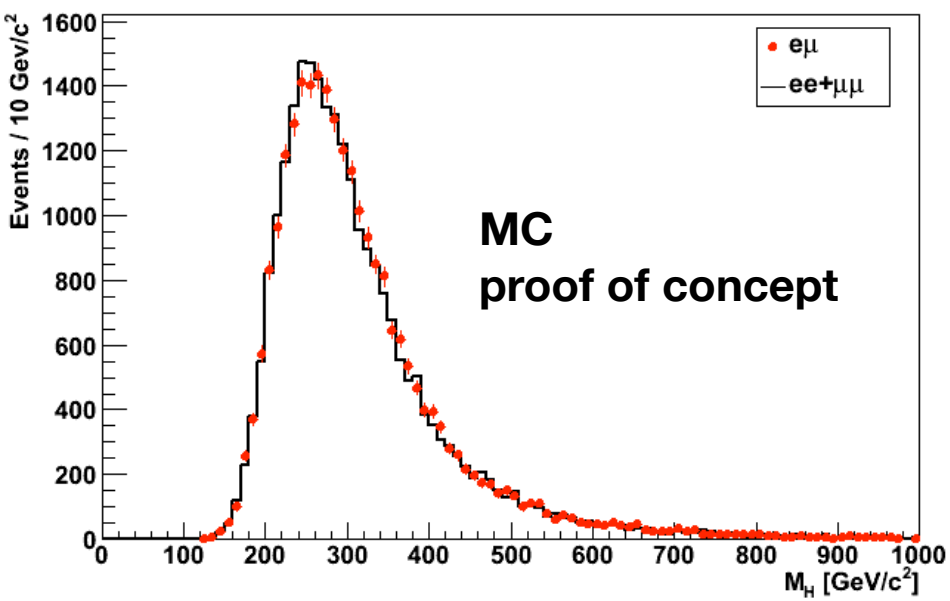
Expected uncertainty:

- 5% for cut&count (stat only)
- shape uncertainty for shape

Background Control in 2-btag

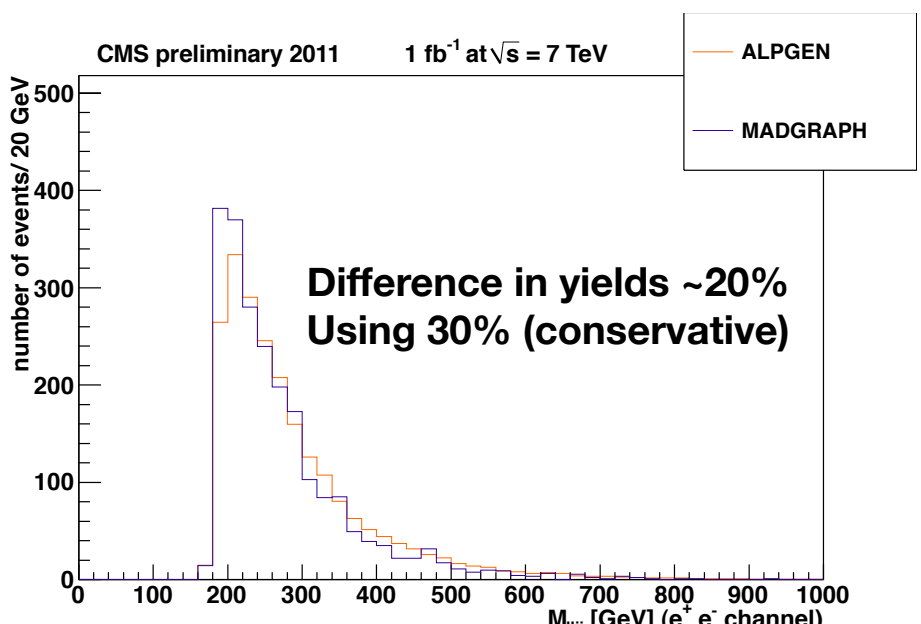
ttbar

- Control region: eμ events
- N(eμ) = N(ee) + N(μμ)



Z+jets

- Not enough stat in 1fb^{-1} for sidebands
- Will have to rely on MC
- Uncertainty: Alpgen-Madgraph



Uncertainties on Leptons and HLT

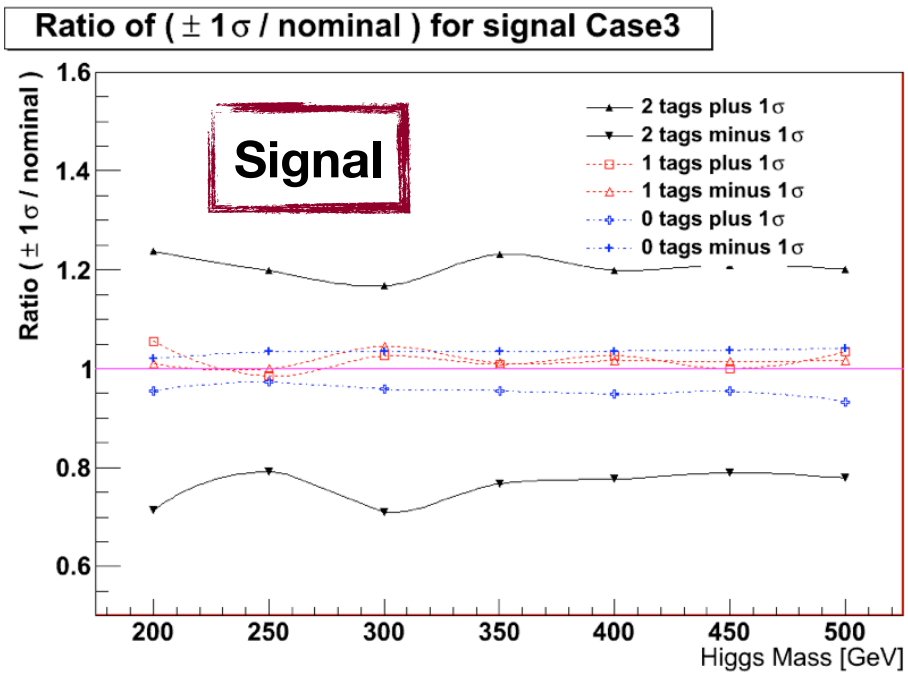
- Three aspects:
  - ID
  - Isolation
  - Trigger

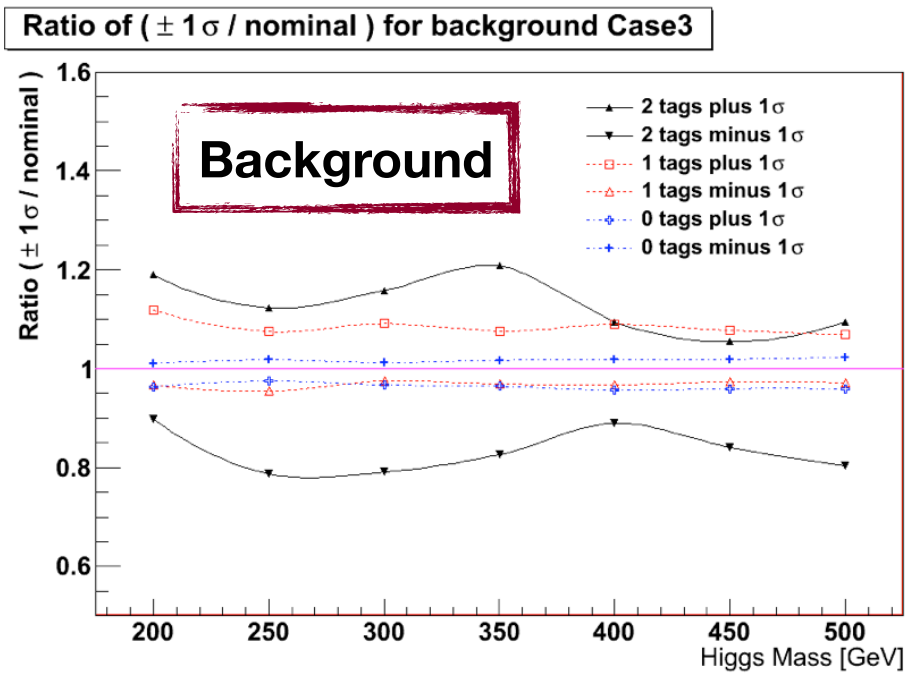
	Electrons	Muons
ID	3.3%	0.5%
Isolation	0.8%	0.2%
Trigger	TBD	0.5%

(independent of mass point)

- Data/MC ratio measured with Z Tag&Probe
  - Muons: 2011 data; Electrons: still 2010 results, will update on 2011
- Uncertainty on ratio used to modify lepton efficiencies
- Systematic: difference in signal yield when varying lepton efficiency $\pm\sigma$

Uncertainty on B-Tagging





- Using BTV recommended scale factors (SF) to emulate b-tag efficiency in MC
  - SF = 0.9 for real b-jets
  - On a jet-by-jet basis: no reweighting, events migrate between categories
- Uncertainty: see how yields change while changing SF by one sigma (0.9±0.15)

Uncertainty on Production Mechanism

- Done by comparing two generators:
  - **JHU:** LO generator with correct angular correlations (not only for spin-0 Higgs but also exotic particles)
  - **Powheg:** NLO generator with correct angles for spin-0 Higgs **only**
- Uncertainty: max difference in signal efficiency while applying selection

EXAMPLE: 300 GeV

	powheg	jhu	difference
mjj 75-107 GeV	95.96%	96.57%	0.4%
mll 70-112 GeV	97.63%	97.42%	0.2%
LD > 0.58	49.3%	50.2%	1%
mH 280-320 GeV	88.17%	90.22%	2%

shape analysis

cut & count

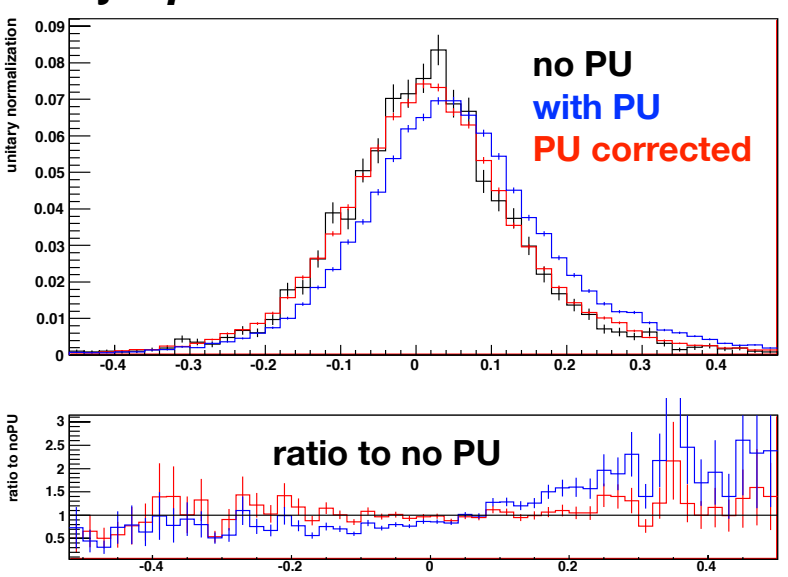
Uncertainty on Pile-Up

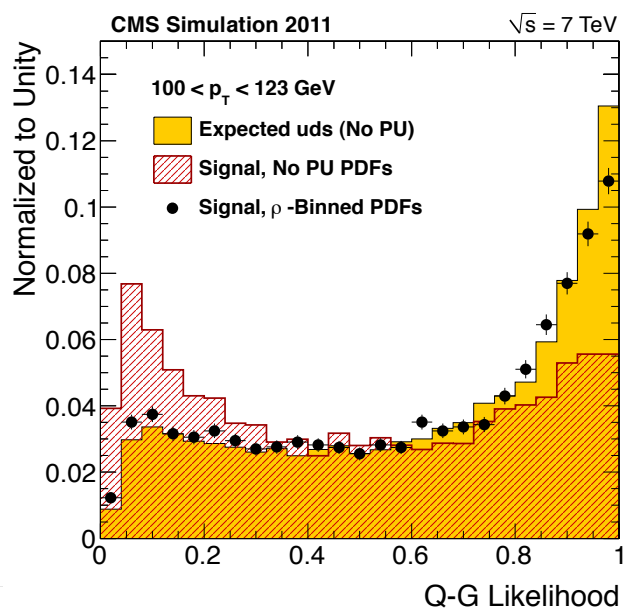
- Pile-up affects:
  - Jet p_T and dijet mass cuts
  - Helicity and QG LD's
  - Lepton isolation

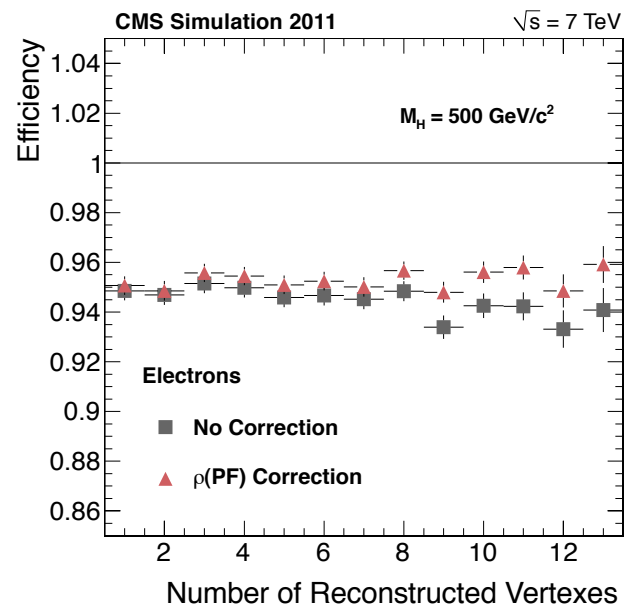
Maximal difference between No-PU MC and FastJet-corrected PU MC

Efficiency loss between 0 and 13 PU

jet p_T resolution

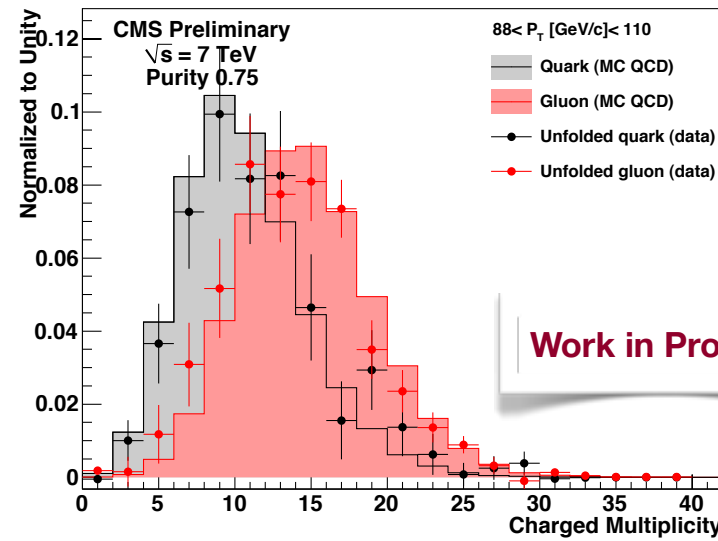




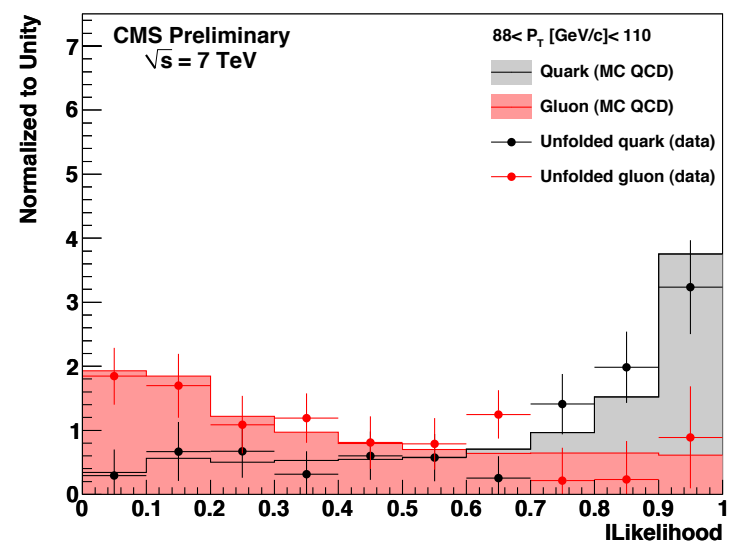


Uncertainties on QG Discriminant

- Will extract PDFs on data
- Control samples:
  - γ+jets: quark jet enriched
  - QCD dijets: gluon jet enriched
- Bin-by-bin unfolding
  - Reconstruct quark and gluon PDFs
- Work in progress
  - Will have results at preapproval



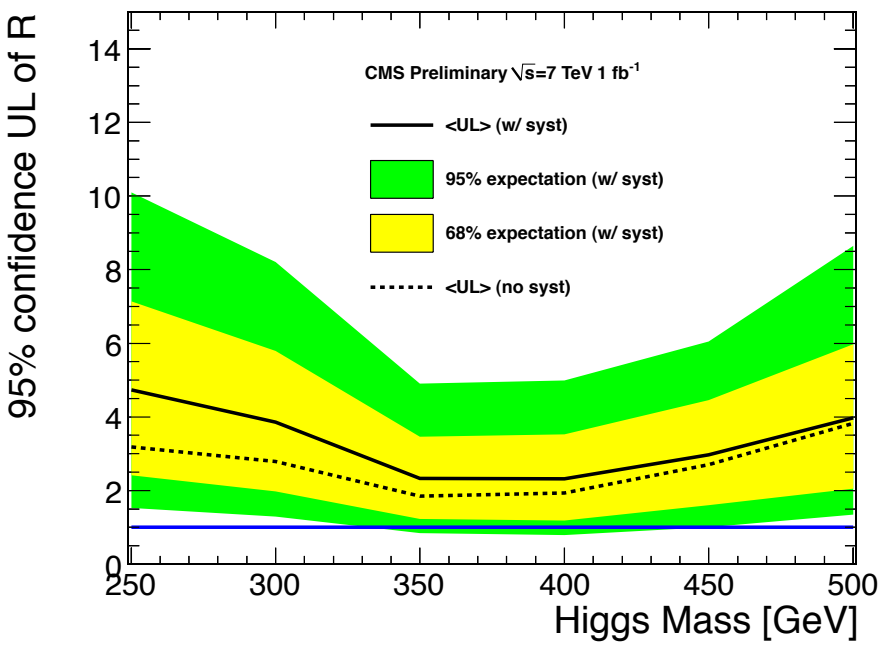
Work in Progess



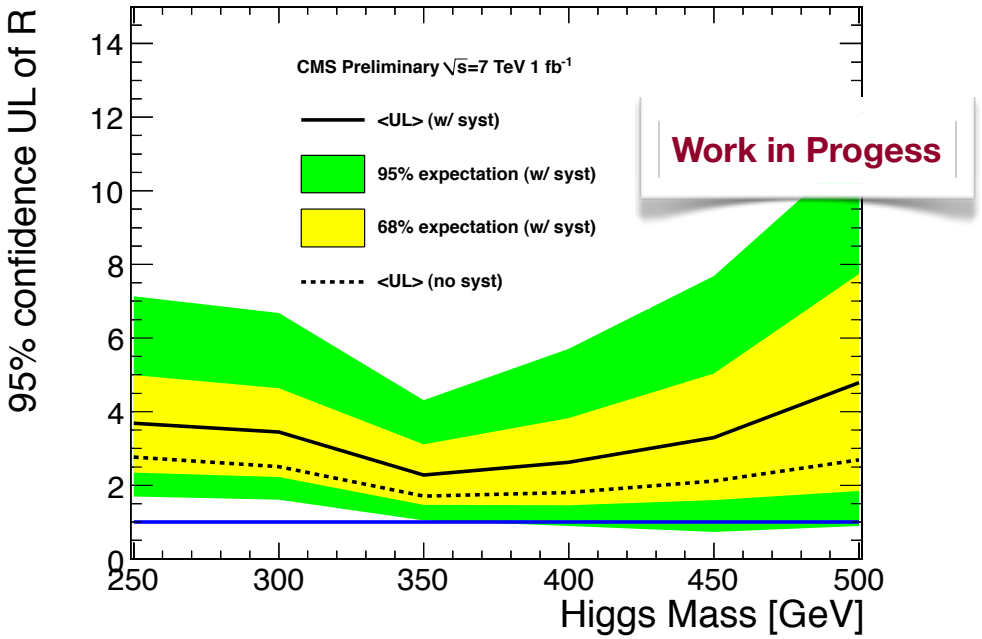
Wrapping It Up Together

Expected Exclusion @ 1 fb^{-1}

Cut and Count



Shape Analysis



- Expected to exclude ~2×SM in 350-400 GeV range with 1 fb^{-1}
- Large effect of systematics at low mass, where large BG

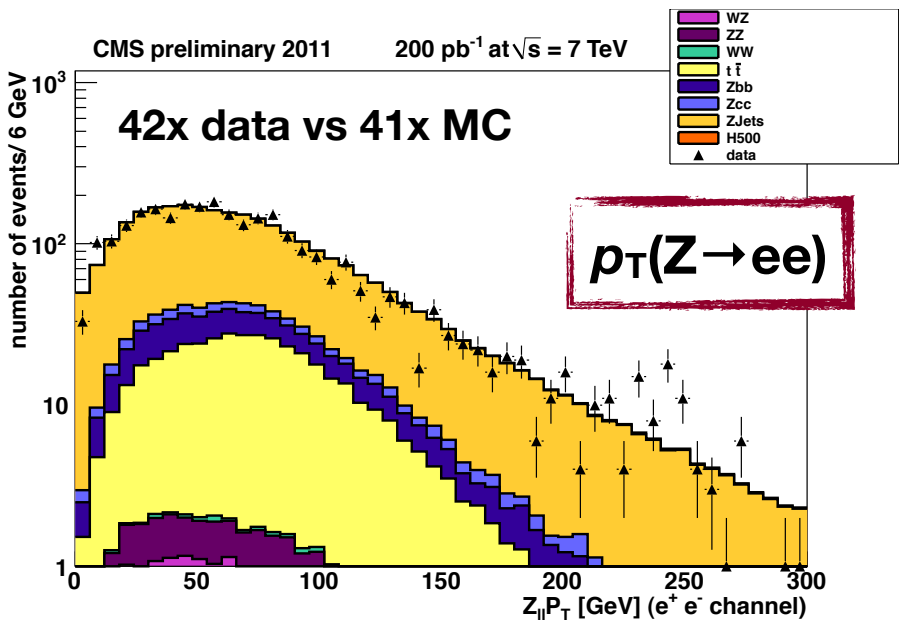
A Look at Data

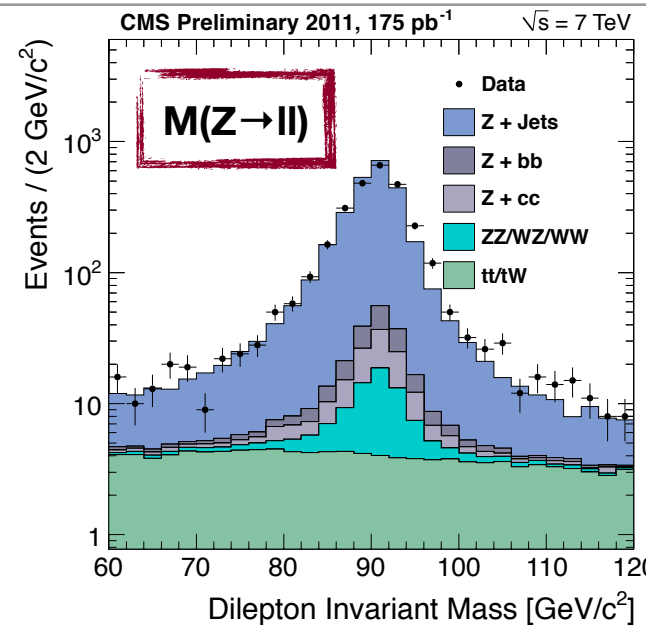
2011 Run A: DoubleElectron and Double Muon PD's

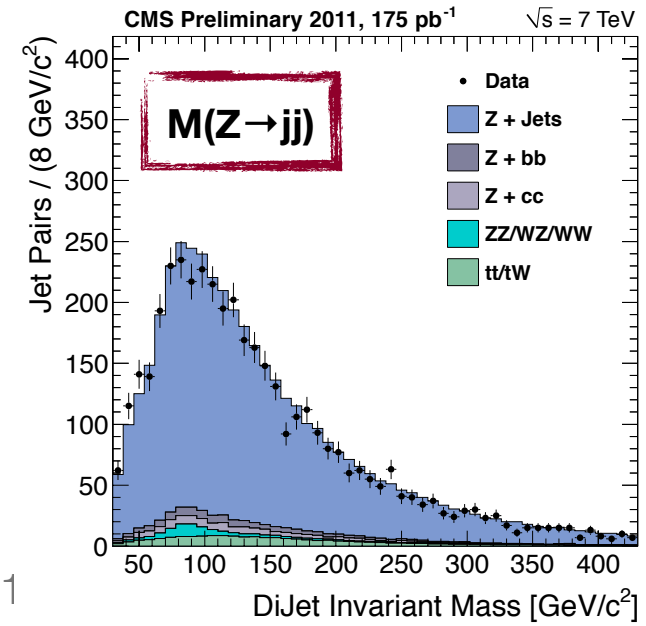
No trigger requirment (other than PD)

PU vertex reweighting

A Look at Data: Kinematic Variables

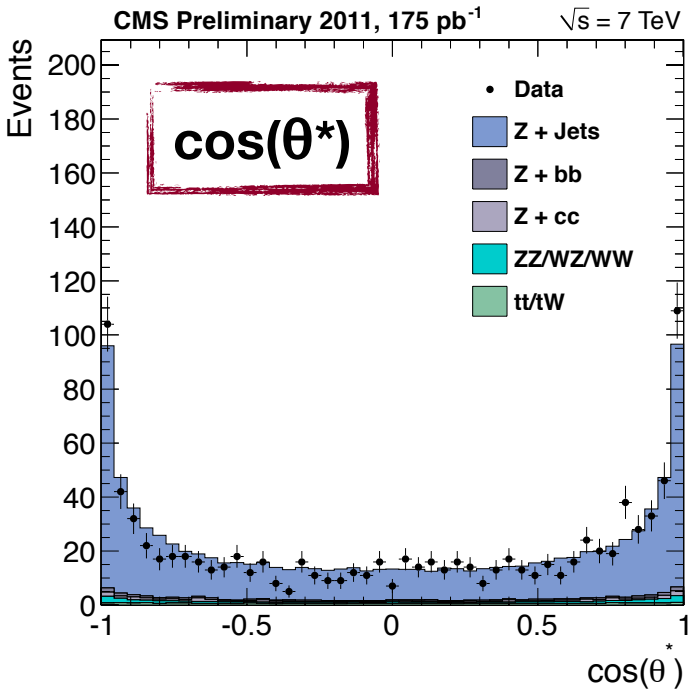


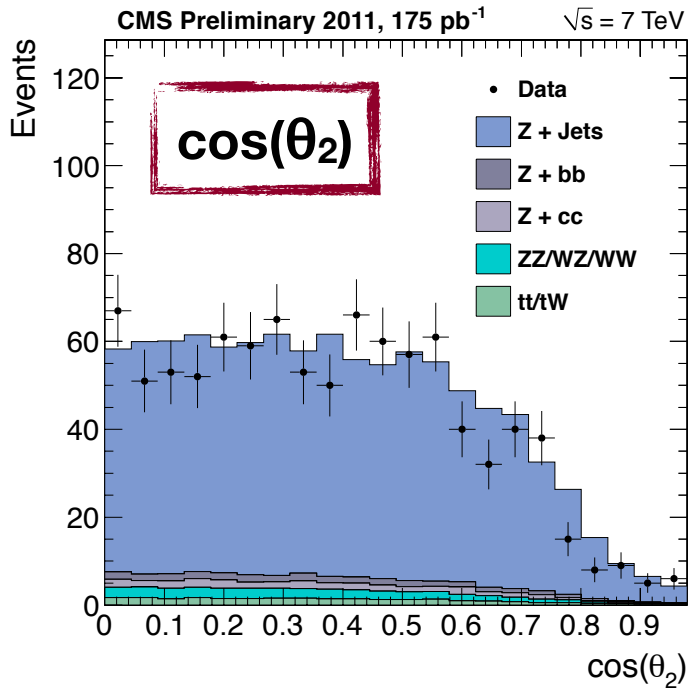


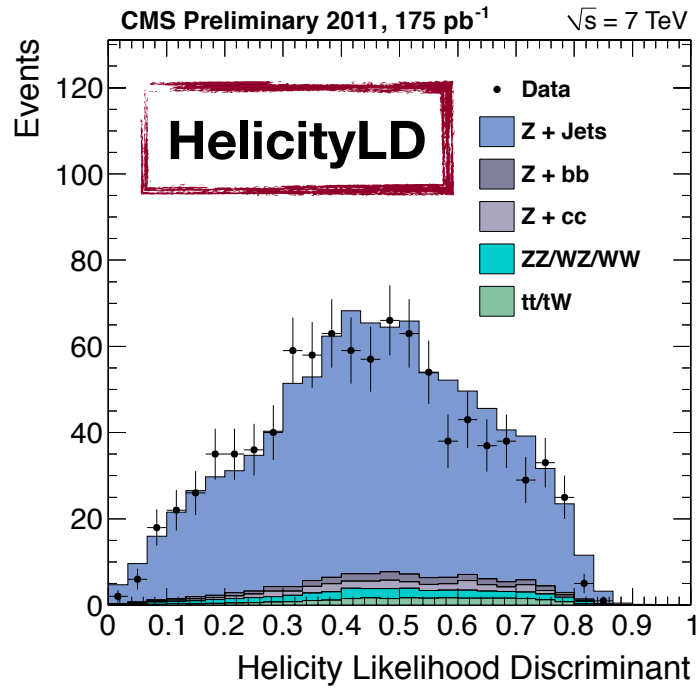


- Events passing preselection only
- Good agreement with MC

A Look at Data: Angular Variables

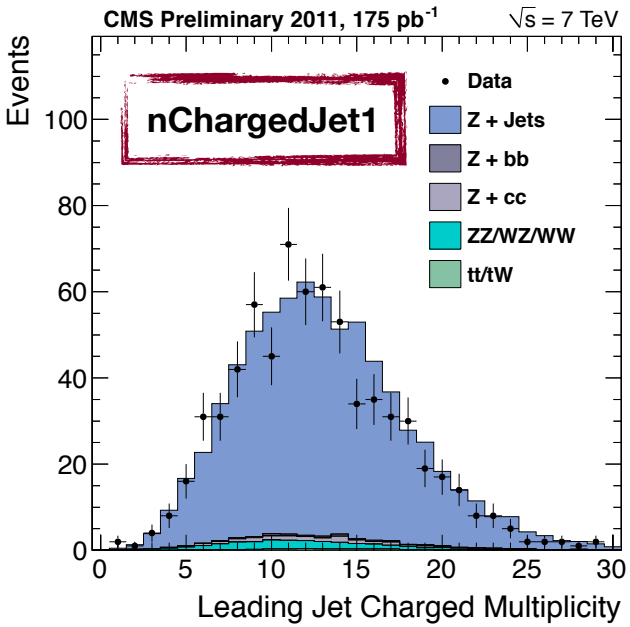


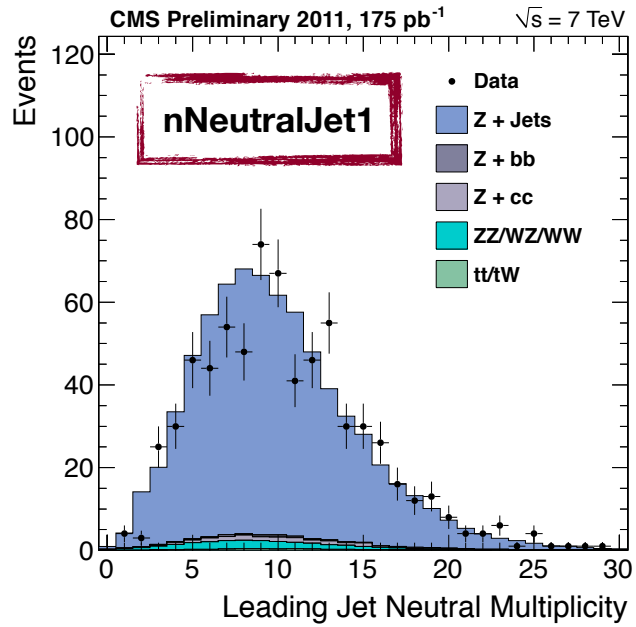


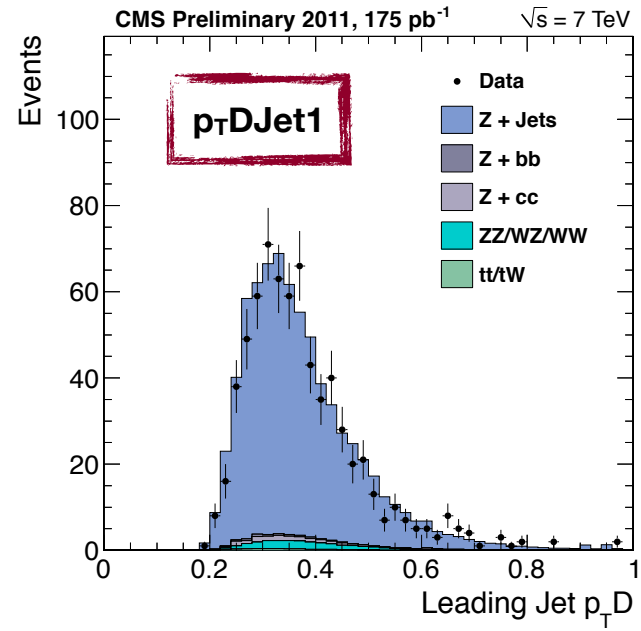


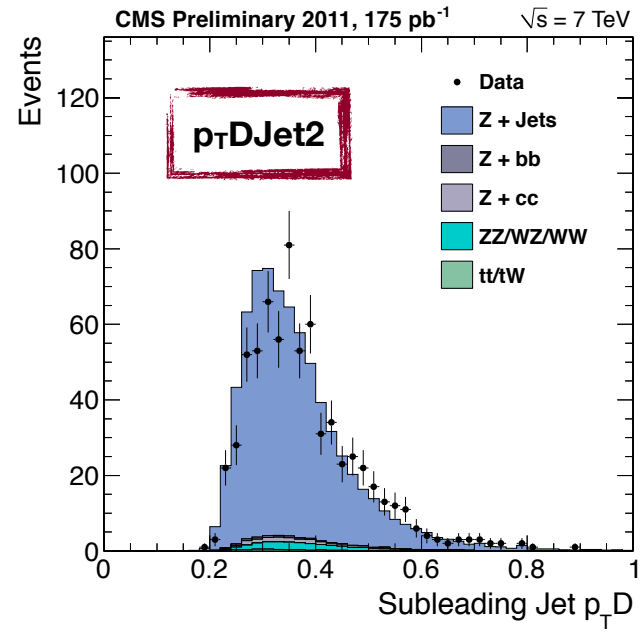
- Good agreement, both on angular variables and LD (so PDFs)
- No reason to think that helicityLD is performing differently than expected

A Look at Data: QG Variables







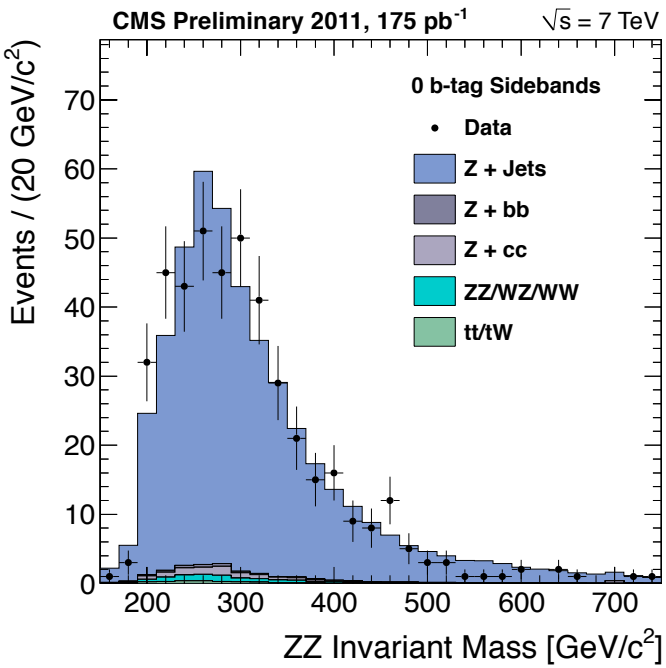


- Good agreement overall
- Hint of discrepancy in subleading jet p_TD?
  - Investigating

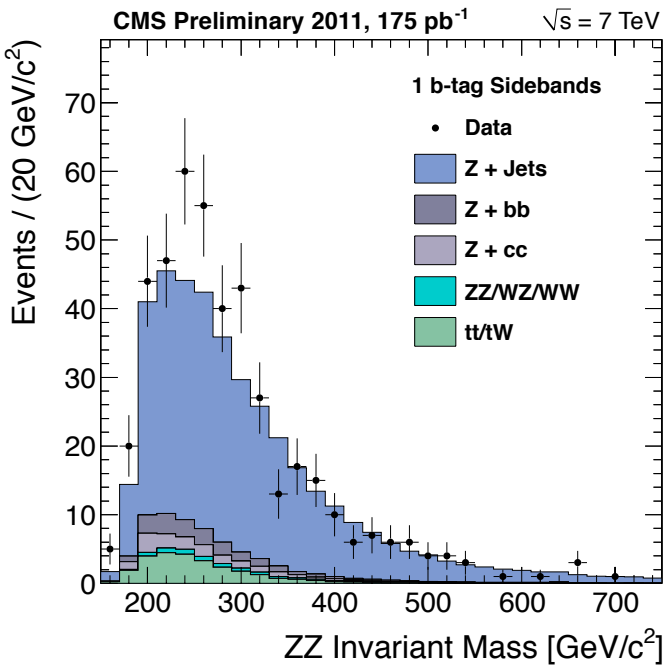
And Now Applying Full Analysis Chain

Invariant Mass Spectra: Sidebands

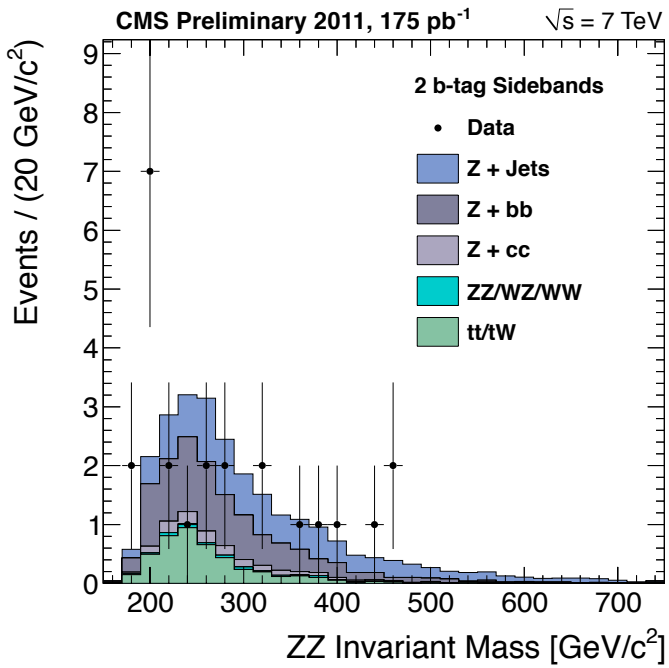
0-btag



1-btag



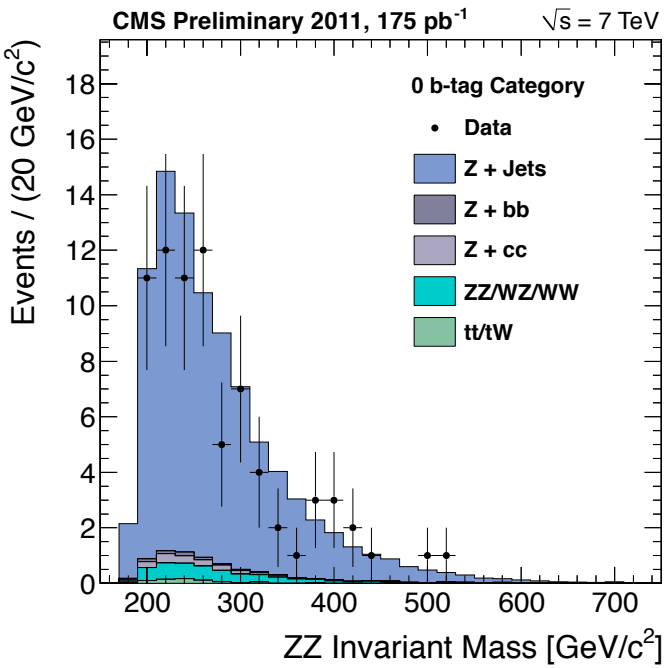
2-btag



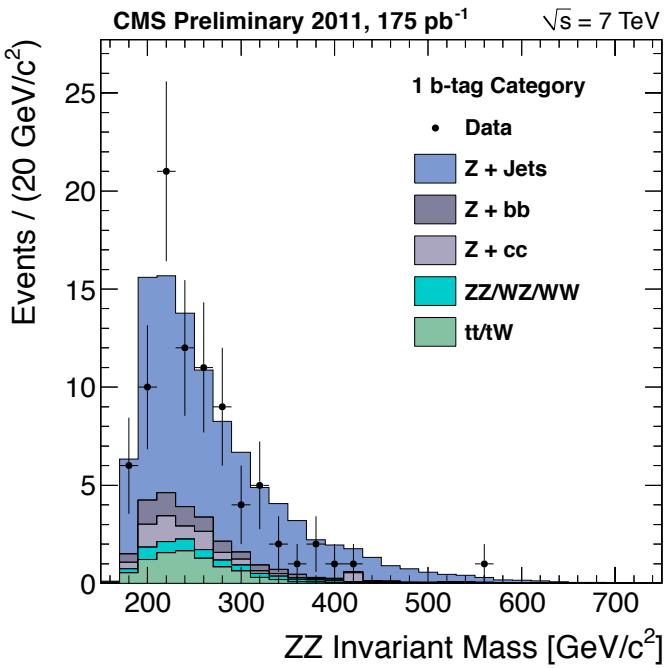
- Defined by inverting the dijet invariant mass cut
- No visible discrepancies wrt expectations

Invariant Mass Spectra: Signal Region

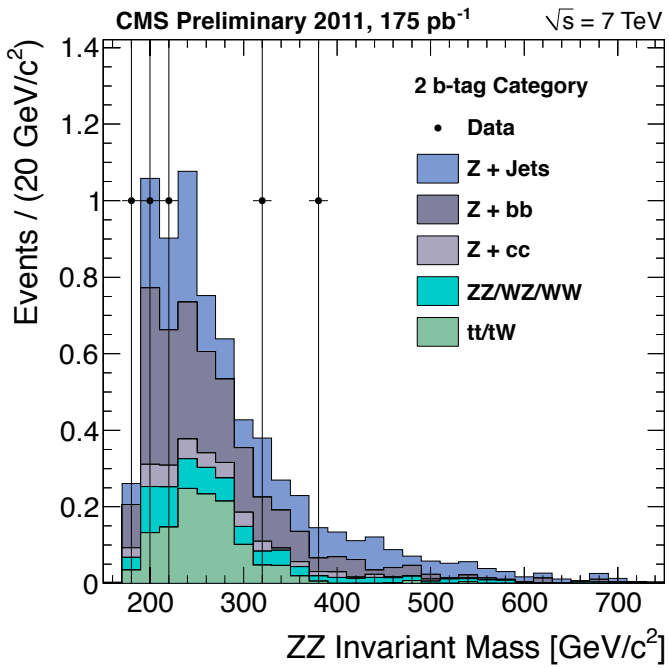
0-btag



1-btag



2-btag



- Full analysis selection
- No visible discrepancies wrt expectations

Conclusions

- The H→ZZ→2l2q analysis frozen and getting ready for EPS
  - Angular analysis to discriminate signal from BG
  - Kinematic fit to dijet system to minimize effect of jet resolutions
  - Flavour tagging (b-tags and quark/gluon) to maximize performance
- Competitive for exclusion or discovery of a heavy Higgs boson (200-600 GeV)
  - Peak performance in 350-400 GeV region: excluding 2×SM with 1 fb^{-1}
- Looking forward to the measurement