Search for the Higgs Boson in the decay channel H → WW

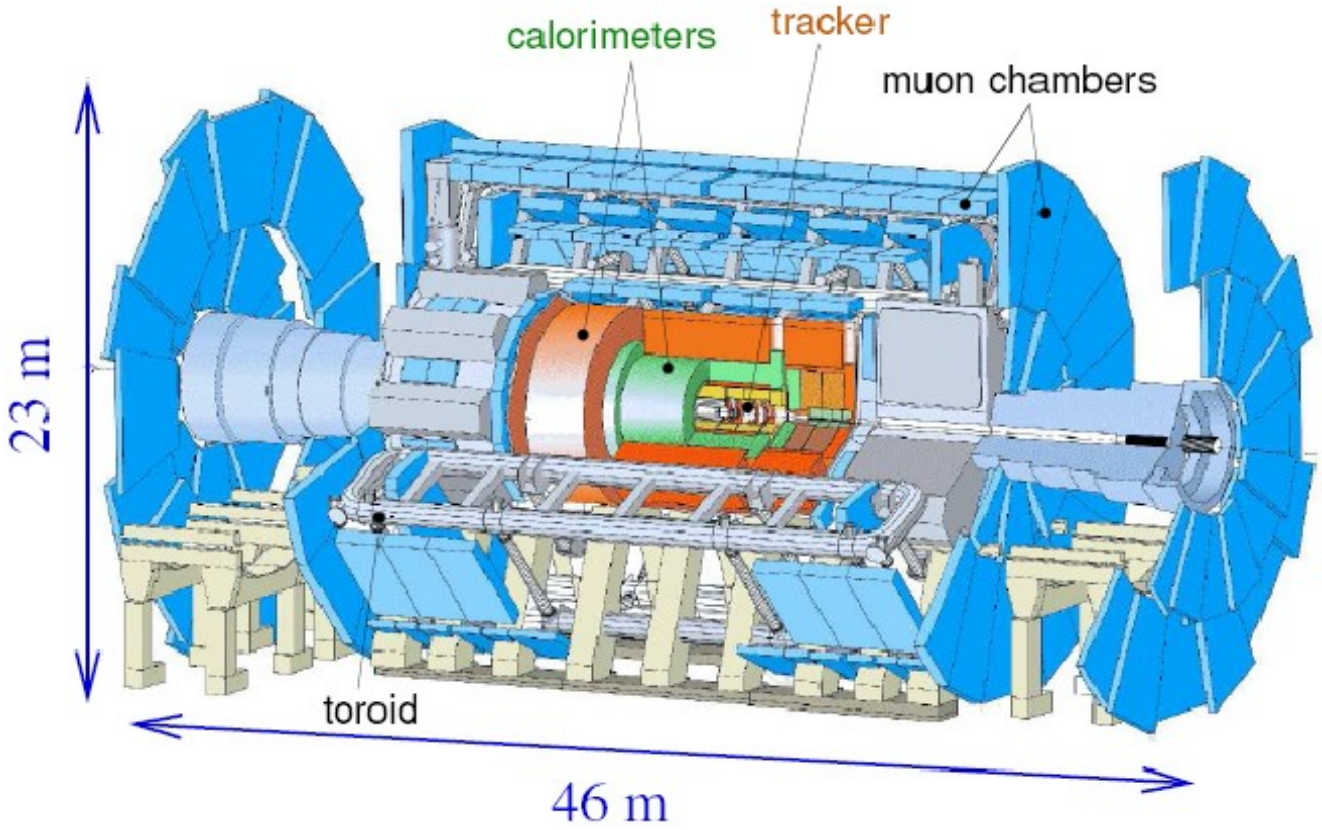
(Results of HG4 CSC note)

Steffen Kaiser

Physics at LHC Seminar
July 1, 2008

Max-Planck-Institut
für Physik
(Werner-Heisenberg-Institut)

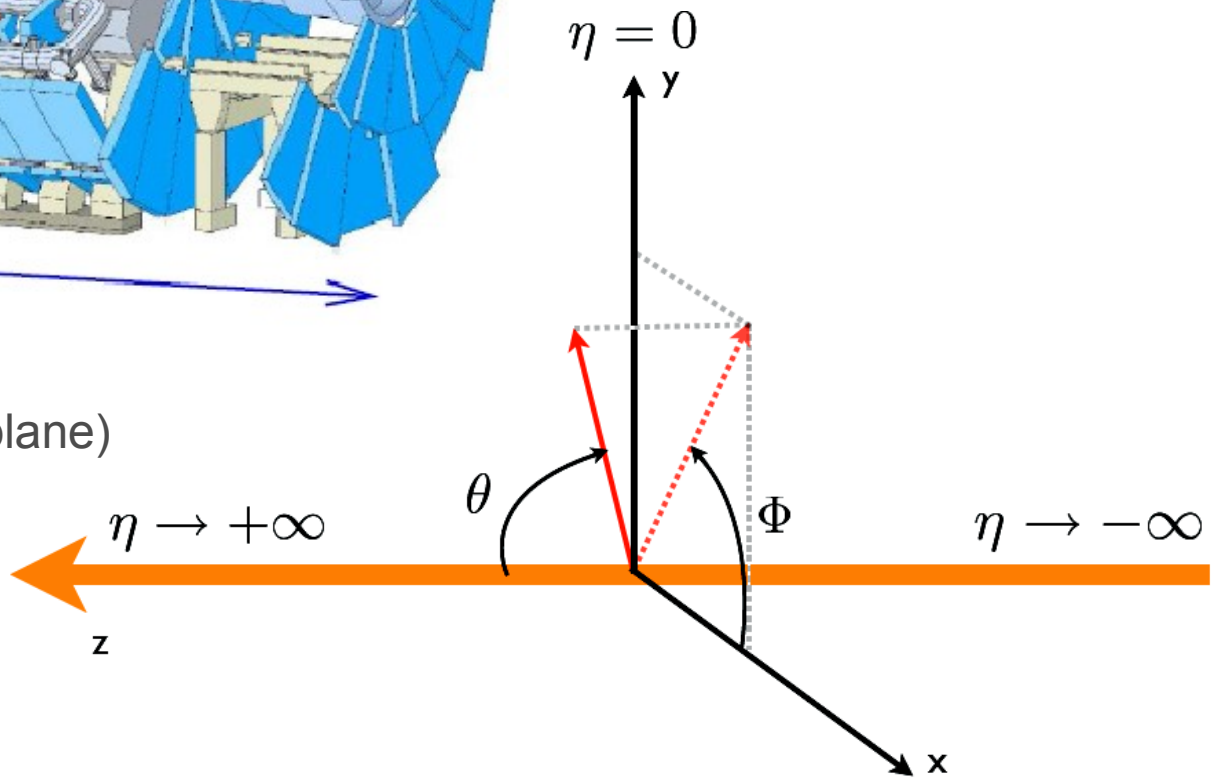
The ATLAS Detector



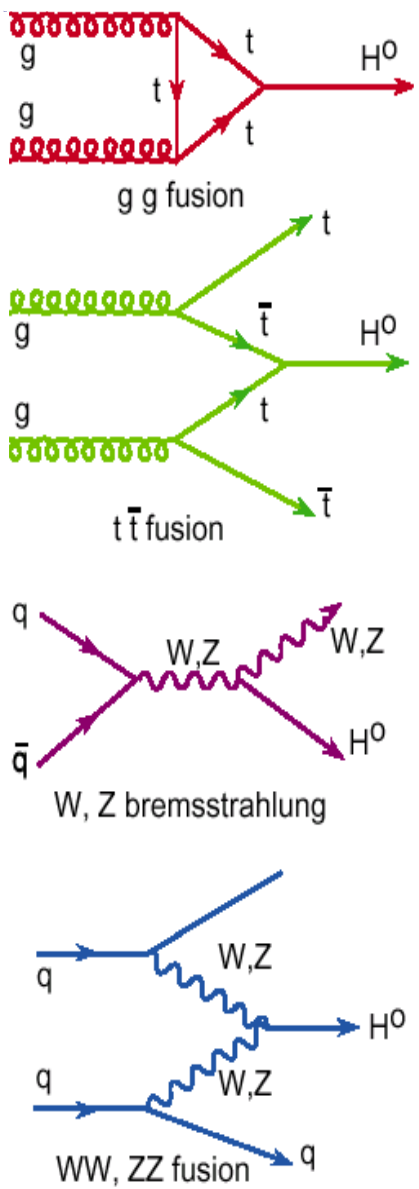
p_T = transverse momentum (x-y plane)

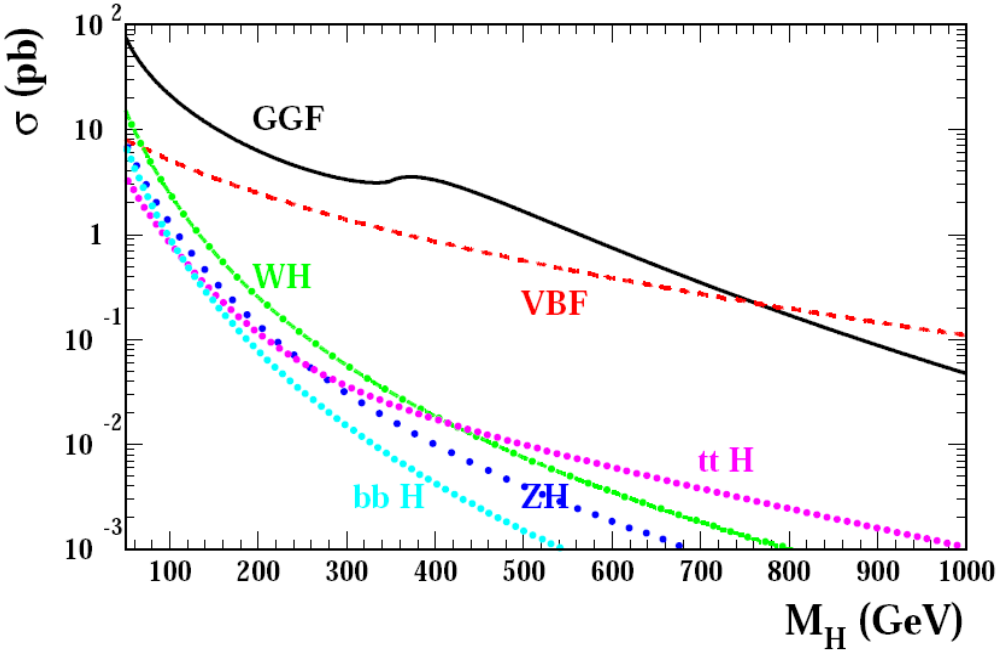
$$\eta = -\ln \tan\left(\frac{\theta}{2}\right)$$

$$\Delta R = \sqrt{\Delta^2\eta + \Delta^2\phi}$$



SM Higgs Production

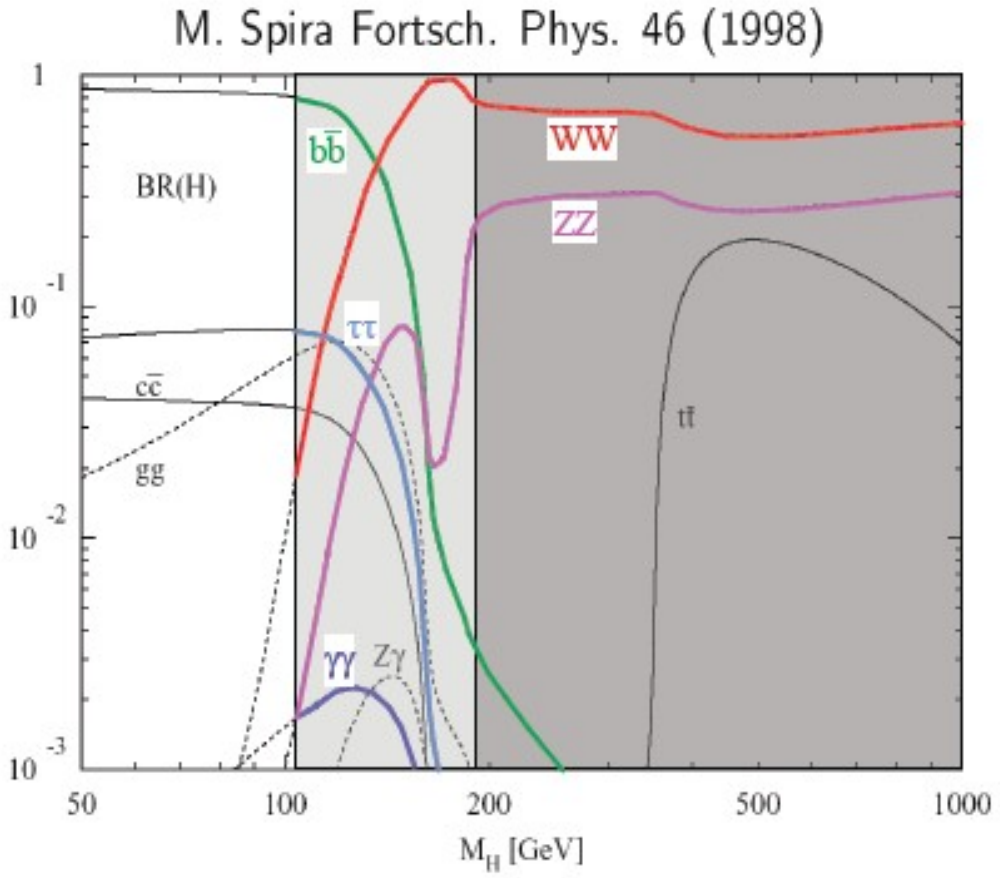




- Gluon fusion dominant up to 800 GeV
- σ(VBF) ~ 0.2 σ(gluon fusion) at low masses
  ~ σ(gluon fusion) for large M_H
- HW, HZ, Htt only relevant for small M_H

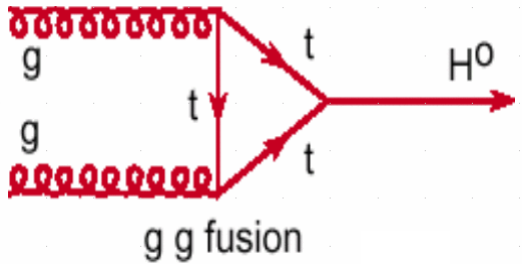
VBF ... Vector Boson Fusion

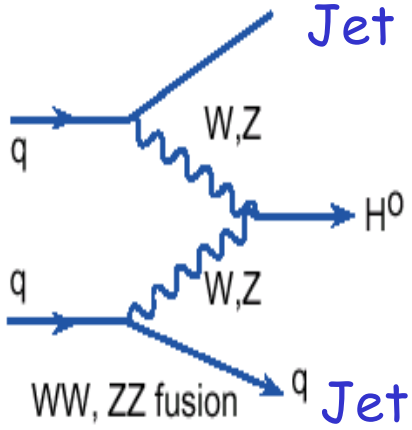
SM Higgs Decays



- M_H < 114.4 GeV (at 95% CL) excluded by LEP
- For $M_H < 2M_W$:
  - H → bb, → ττ dominant
  - H → γγ small but also relevant (high precission γ reconstruction)
- For $M_H > 2M_W$:
  - H → WW, → ZZ dominant

Higgs Decay: H → WW → lνlν



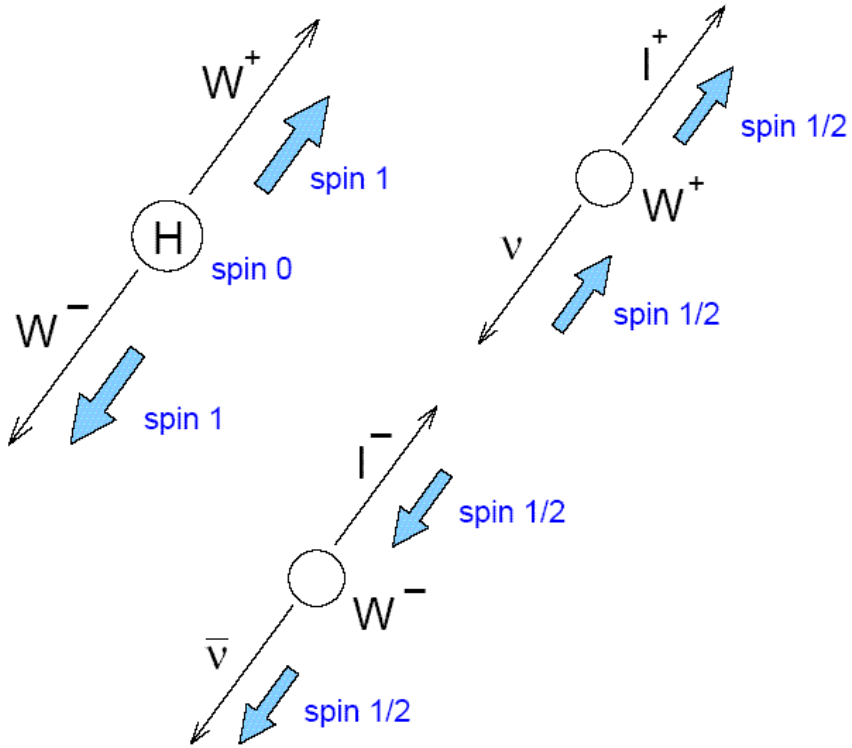


Signature:

- 2 high p_T leptons + large missing E_T
- Lepton spin correlation (spin 0↔1)

→ W^+ and W^- have opposite spins

→ Leptons $l^\pm$ tend to be emitted in the same direction

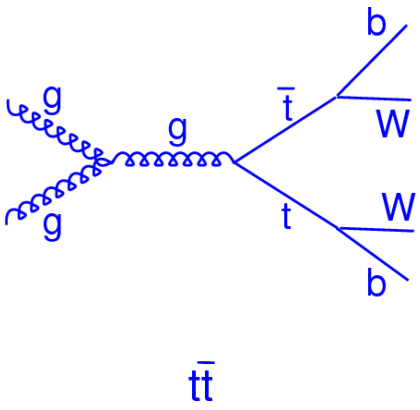


- No mass peak → transverse mass

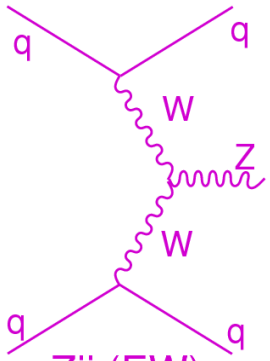
$$M_T = \sqrt{(E_T^{\ell\ell} + E_T^{\nu\nu})^2 - (\vec{p}_T^{\,\ell\ell} + \vec{\not{P}}_T)^2}$$

Signal and Background Samples

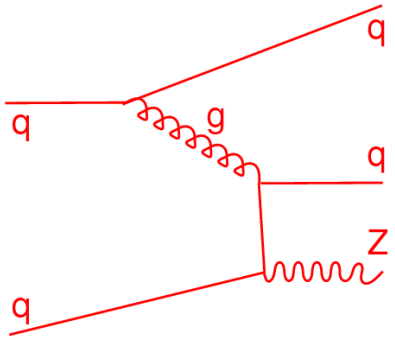
Backgrounds:



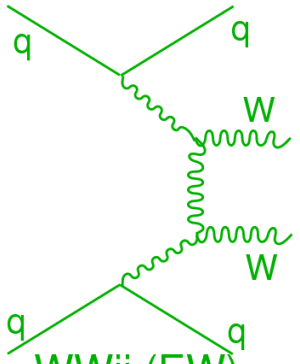
$t\bar{t}$



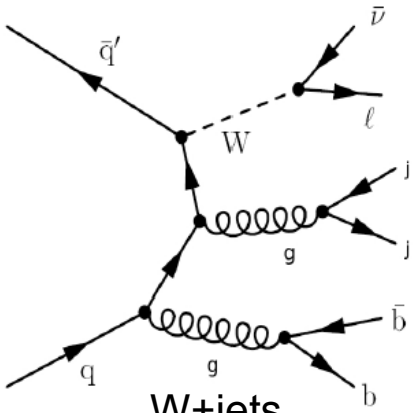
Zjj (EW)



Zjj (QCD)



WWjj (EW)



W+jets

MC Samples:

Process	Generator	σ (pb)
$gg \to H \to WW$ $(M_H = 170$ GeV)	MC@NLO	19.418
VBF $H \to WW$ $(M_H = 170$ GeV)	PYTHIA/Sherpa	2.853
VBF $H \to WW$ $(M_H = 300$ GeV)	HERWIG	0.936
$qq/qg \to WW$	MC@NLO/Alpgen	111.6
$gg \to WW$	GG2WW	5.26
$pp \to t\bar{t}$	MC@NLO	833
$Z \to \tau\tau$+jets	PYTHIA/ALPGEN	2015
W+jets	ALPGEN	20510

Channels and Analysis Methods

- H + 0j ($H \rightarrow WW \rightarrow e\nu\mu\nu$)
  - 2D-Fit (M_T,p_T^{WW})
  - Cut & Count
- H + 2j ($H \rightarrow WW \rightarrow e\nu\mu\nu$)
  - 2D-Fit (NN,M_T)
  - 5D-Fit (M_T, $\Delta\varphi_{ll}$, $\Delta\eta_{ll}$, $\Delta\eta_{jj}$,M_{jj})
- H + 2j ($H \rightarrow WW \rightarrow l\nu qq$)
  - 1D-Fit (M_H)

Lepton/Jet Reconstruction

Electrons/Muons:

- MediumElectrons: shower shape in Calo + ID track quality
- ThightElectrons: + B-layer and TRT hits + cluster isolation cuts
- Track match ($d_0/\sigma_{d0}<10$)
- Calorimeter isolation: E_T in cone($\Delta R < 0.2$) < 5 GeV, 2.5 GeV
- Track isolation: Σp_T(tracks) in cone($\Delta R < 0.4$) < 5 GeV, 3 GeV
- $p_T > 15$ GeV, $|\eta| < 2.5$
- Efficiency in $H \rightarrow WW$: (50.0 ± 0.5)% (77.1 ± 0.2)%
- Fakerate in $W \rightarrow \mu/ev$ + jets: $(6.7 \pm 1.5) \cdot 10^{-5}$ $(1.7 \pm 0.5) \cdot 10^{-5}$

Jets:

- TopoCluster, Cone: $\Delta R < 0.4$, $|\eta| < 4.8$
- Efficiency ~ 95%

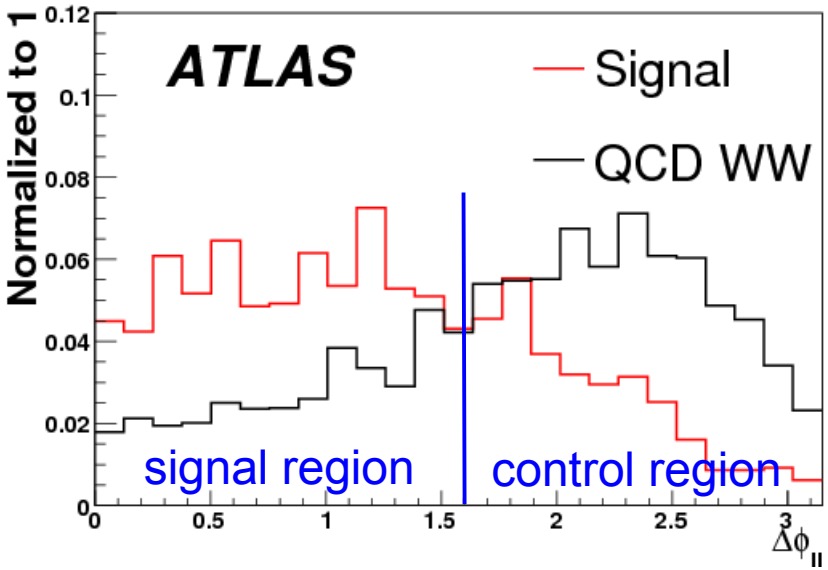
$H + 0j\ (H \rightarrow WW \rightarrow e\nu\mu\nu)$

- 2D-Fit (M_T, p_T^{WW})
- Cross Check: Cut & Count
- Outlook: Multivariate Techniques

- Backgrounds: WW, $t\bar{t}$, $Z \rightarrow \tau\tau$, W +jets

Event Selection

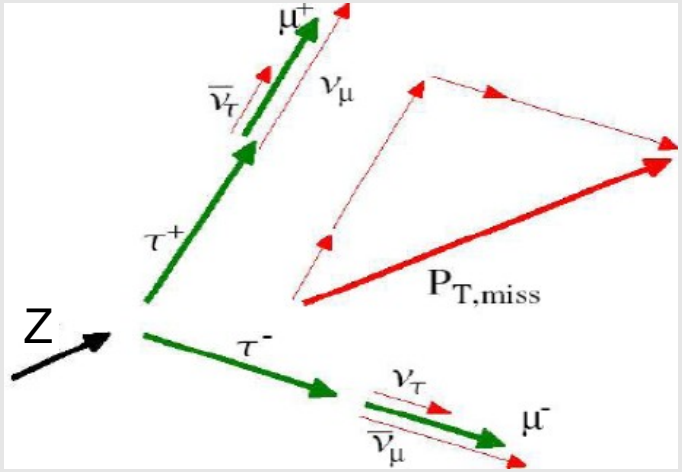
- 2 isolated leptons, opposite charge, p_T > 15 GeV (TightElectrons)
- 12 GeV < m_{ll} < 300 GeV
- Missing E_T > 30 GeV
- $Z \rightarrow \tau\tau$ veto
- Jet veto (p_T > 20 GeV, |η| < 4.8)
- b-jet veto (p_T > 15 GeV, b-weight > 4)
- σ (fb):



Selection	Selection cuts	$gg \rightarrow H$	$t\bar{t}$	WW	$Z \rightarrow \tau\tau$	$W + jets$
pre-selection	Lepton Selection+M_{ll}	169.0	6501	718.12	4171	209.1
	$p_T^{miss} > 30$ GeV	133.2	5617	505.25	526.3	181.6
	$Z \rightarrow \tau\tau$ Rej.	129.8	5215	485.12	164.2	150.4
	Jet Veto	52.85	14.84	238.35	31.91	76.12
	b-veto	52.62	6.85	237.87	30.76	76.12
signal region	$\Delta\phi_{ll} < 1.575$, $M_T < 600$ GeV	35.7±1.1	2.3±1.6	85.4±2.7	<1.7	38±38
control region	$\Delta\phi_{ll} > 1.575$, $M_T < 600$ GeV	16.9±0.7	4.6±2.3	151.9±3.6	30.8±4.2	38±38

$Z \rightarrow \tau\tau$ Veto

Collinear Approximation

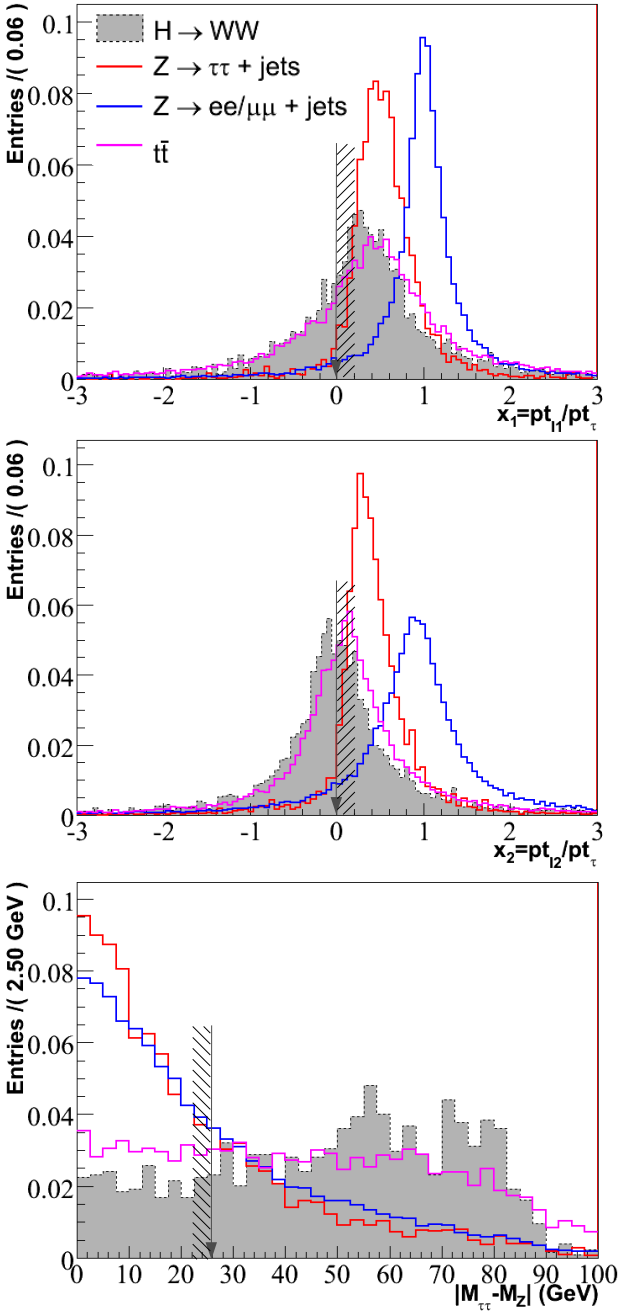


Large boost of tau

→ Same direction: tau, lepton, neutrino

$$x = \frac{p_T(Lepton)}{p_T(\tau)}$$

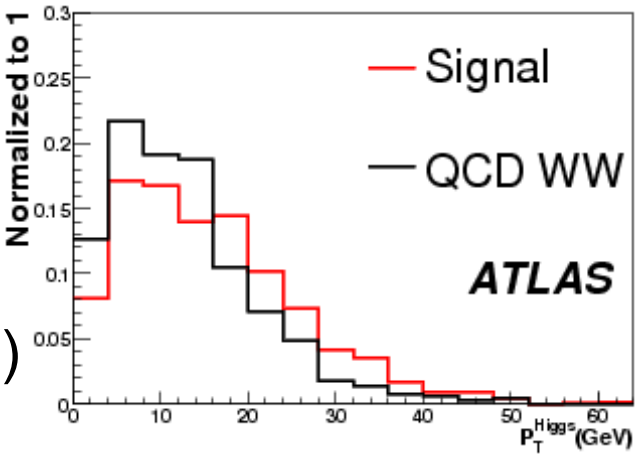
→ Reject event if $x_1 > 0$, $x_2 > 0$ and $|M_Z - M_{\tau\tau}| < 25$ GeV



Fit Strategy

Top Background:

- Shape determined on b-tagged control sample
- Cross section and shape extrapolated from b-tagged to b-vetoed region based on MC ($t\bar{t}$ - MC@NLO)

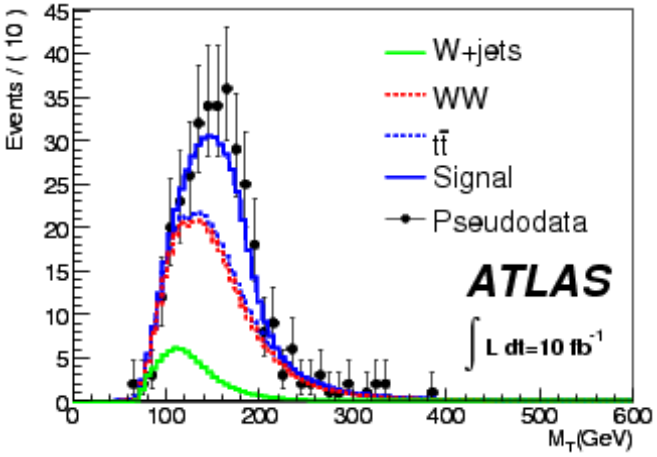


$Z \rightarrow \tau\tau$:

- Normalization and shape from $Z \rightarrow \mu\mu$ events with replacing μ's with simulated τ's ($82 < m_{\mu\mu} < 98$ GeV)

Fake Backgrounds (W+jets):

- Shape and normalization from MC (loose isolation)



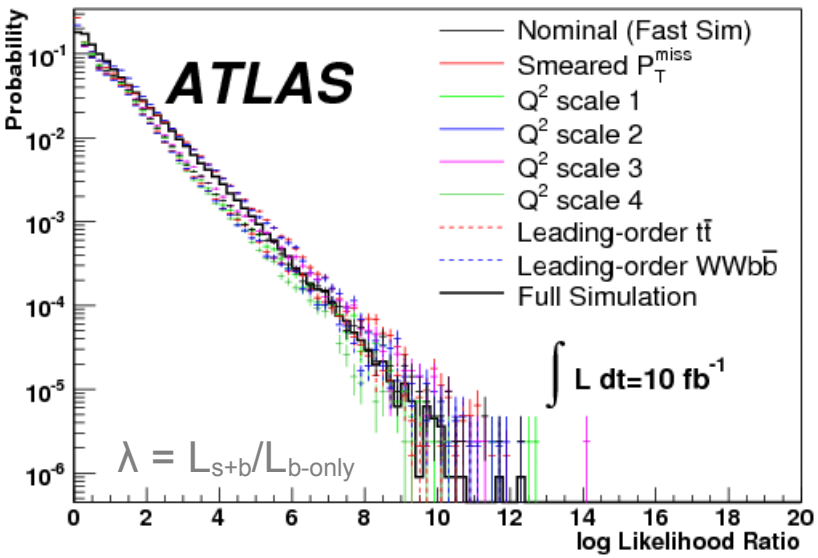
Combined fit:

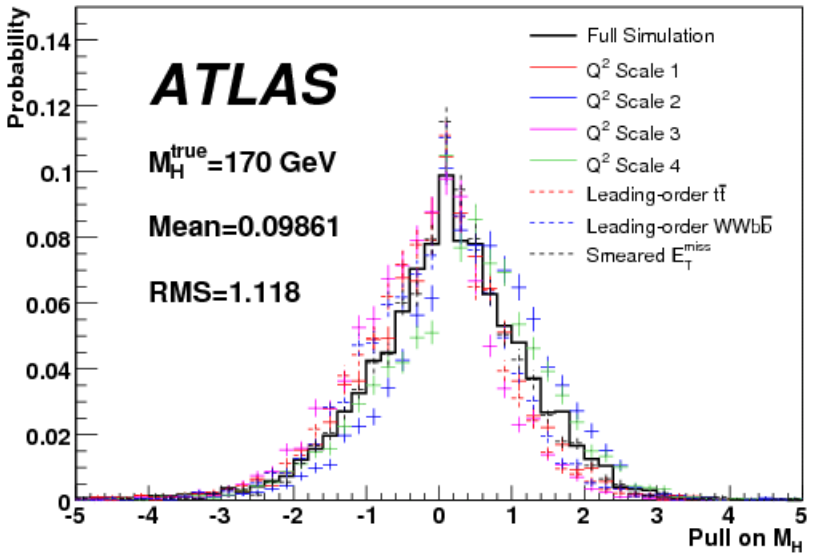
- Simultaneous fit of both $\Delta\varphi_{ll}$ bins in (M_T, p_T^{WW})
- Parameters for WW bkg allowed to float in the fit (same shape for signal and control region except for correction factor)

Systematics Studies

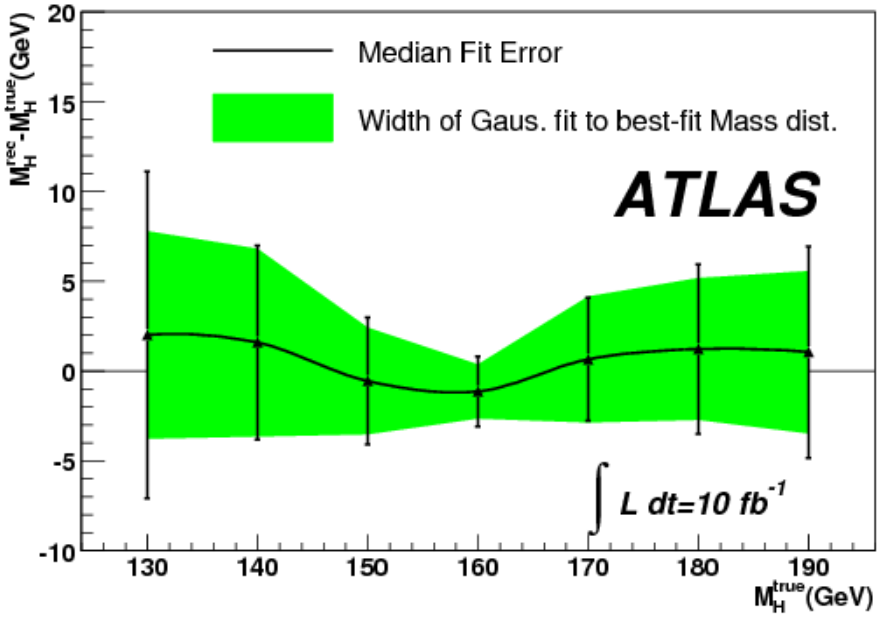
Toy MC with different distorted scenarios (derived on ATLFAST):

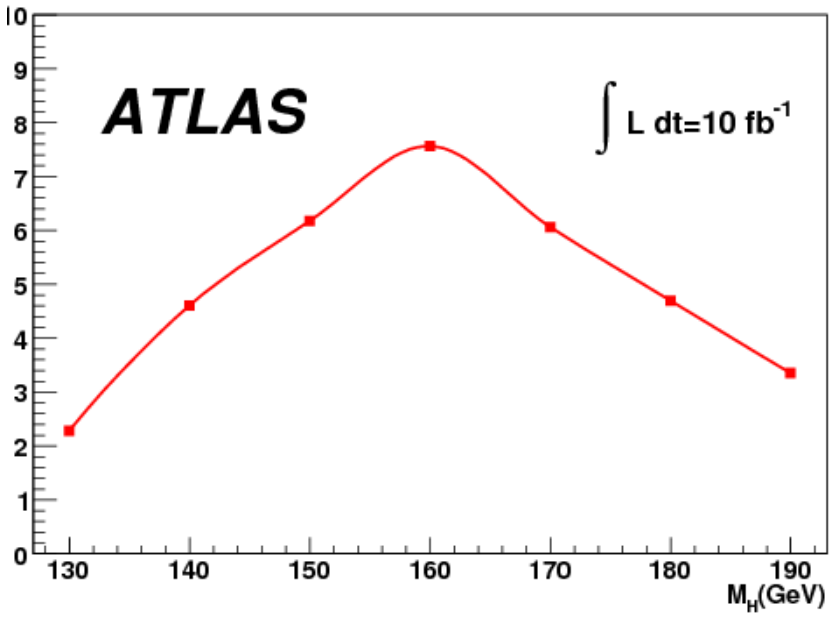
- Q^2 scale (factorization and renormalization raised, lowered by factor 8)
- 2 alternative top bkg models
  (leading order pp $\rightarrow$ WWbb, pp $\rightarrow t\bar{t} \rightarrow$ WWbb)
- For bkg: smear x and y components of missing E_T independently by 5 GeV





Fit Results





- Most promising for M_H near the WW threshold
- Maximum significance: 7.8σ (M_H=160GeV)
- At other masses:
  Larger systematic uncertainties on bkg predictions

Cross Check and Alternatives

Cut & Count:

Region	Signal, $M_H = 170$ GeV (fb)	$t\bar{t}$	WW	$Z \to \tau\tau$
Signal-like	19.61±0.80	1.14±1.14	29.35±1.59	<1.74
Control	2.27±0.27	5.71±2.55	61.13±2.33	4.06±1.53

- Signal region: $p_T^{WW} > 10$GeV, $M_{ll} < 64$GeV, $\Delta\varphi_{ll} < 1.5$, $50 < M_T < 180$GeV
- Control region: $p_T^{WW} > 10$ GeV, $80 < M_{ll} < 300$ GeV, $\Delta\varphi_{ll} > 1.5$
- Top background from b-tagged control region
- W+jets bkg neglected since only 1 event passes cuts
- Ratios of cross sections in the different regions taken from MC
- Systematic uncertainties on cross sections: WW (5%), top (9%)
- Significance (10 fb^{-1}): 7.1σ (fit: 6.7σ, ignoring fake backgrounds)

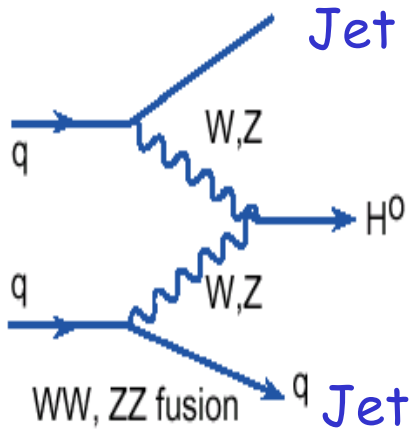
Multivariate Techniques:

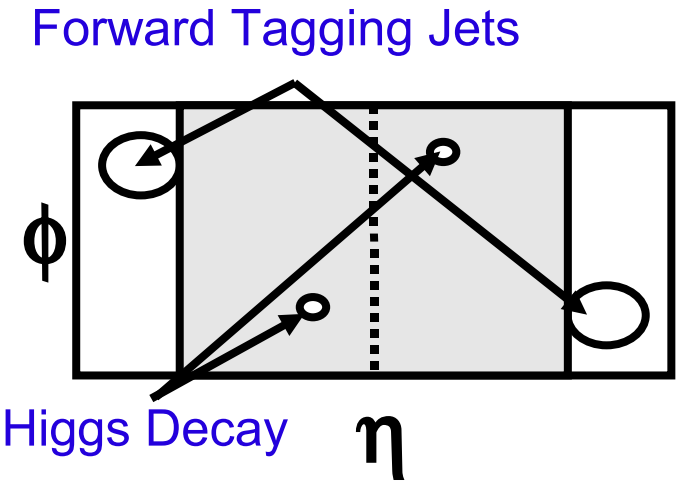
- NN (5 variables), BDT (21 variables) under study

H + 2j ($H \rightarrow WW \rightarrow e\nu\mu\nu$)

- 2D-Fit (NN, M_T)
- 5D-Fit (M_T, $\Delta\varphi_{ll}$, $\Delta\eta_{ll}$, $\Delta\eta_{jj}$, M_{jj})
- Backgrounds: $t\bar{t}$, WW+jets, W +jets

Higgs Production via VBF: qq → qqH

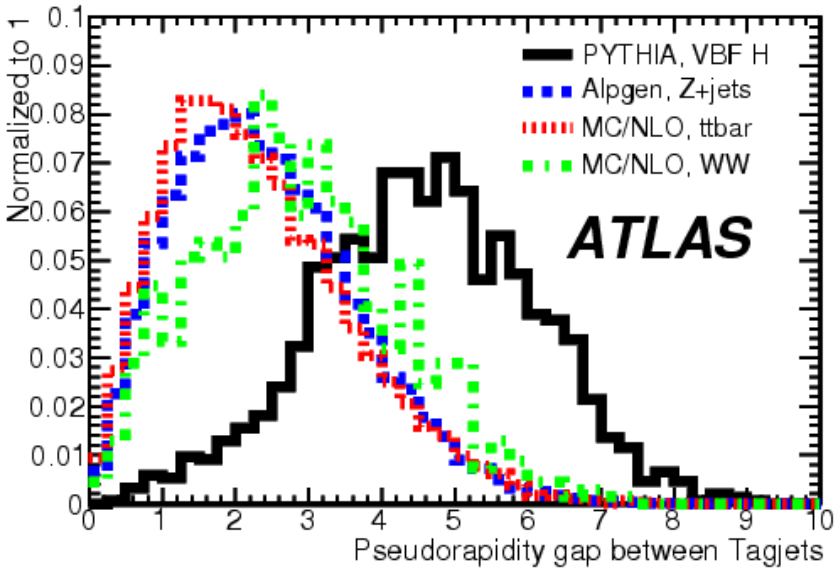




Signature:

- 2 forward jets with large rapidity gap
- Suppressed central jet activity

→ Only Higgs decay products in central part of the detector

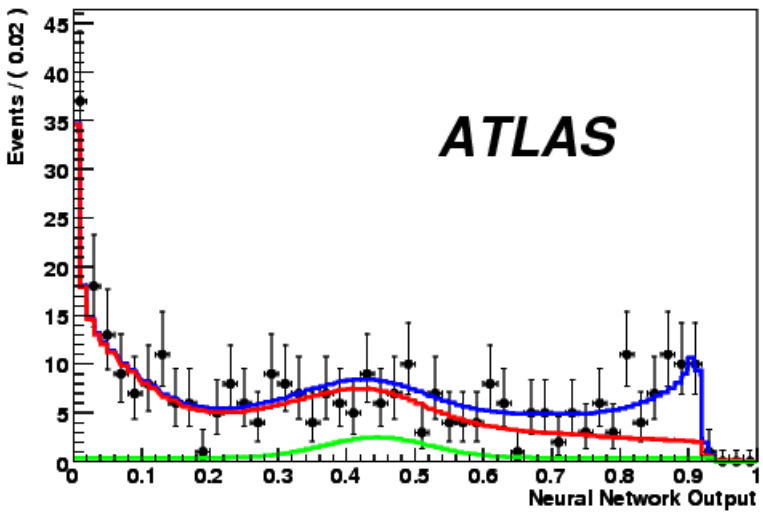


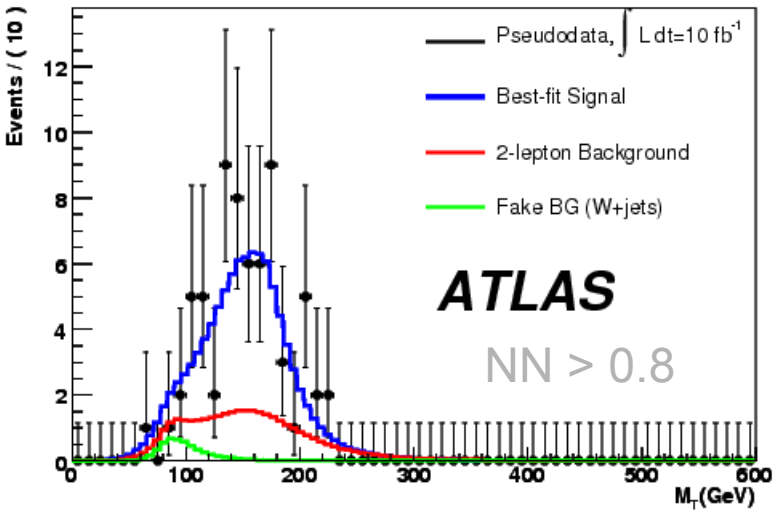
2D Fit - Event Selection

- 2 isolated leptons, opposite charge, p_T > 15 GeV (MediumElectrons)
- Missing E_T > 30 GeV
- At least 2 jets with p_T > 20 GeV and $|\eta|$ < 4.8
- $\eta_{j1} \cdot \eta_{j2} < 0$, $|\Delta\eta_{jj}| > 3$, leptons between jets
- $Z \rightarrow \tau\tau$ veto, b-jet veto
- 50 < M_T < 600 GeV
- Signal box: $\Delta\varphi_{ll} < 1.5$ and $\Delta\eta_{ll} < 1.4$, control region: the rest

Cut [σ(fb)]	Signal (170 GeV)	$t\bar{t}$	WW+jets	$Z \rightarrow \tau\tau$	W+jets
Lepton Selection	30.20	8317	838.96	2096	1323
Forward Jet Tagging	17.27	946.6	32.77	79.30	31.83
Leptons Between Jets	16.47	617.8	22.92	55.13	27.91
$Z \rightarrow \tau\tau$ Rejection	15.68	561.8	21.20	39.03	27.91
p_T^{miss}, M_T, $m_T^{ll\nu}$	12.78	425.9	15.28	0	13.96
b-veto	12.67	206.72	-	-	-
signal box	9.28±0.27	28.5±5.7	4.75±0.30	-	4.3±4.3
control region	3.02±0.15	89±10	9.78±0.43	-	7.9±5.0

2D Fit - Setup





- Neural Net: $\Delta\eta_{jj}$, M_{jj}, p_T(3rd jet) in $|\eta| < 3.2$, $\eta^* = \eta_3 - (\eta_1 + \eta_2)/2$
- 2D fit: NN (parameters floating), M_T (parameters fixed)
- Uncorrelated product probability density functions (PDFs)
- 2-leptons backgrounds combined in one PDF ($t\bar{t}$, WW+jets)
- Fake backgrounds:
  - PDF determined on MC sample with loose cuts
  - Shape and normalization fixed in final fit
- Same bkg NN distribution in signal and control region except for slope of a linear extrapolation factor

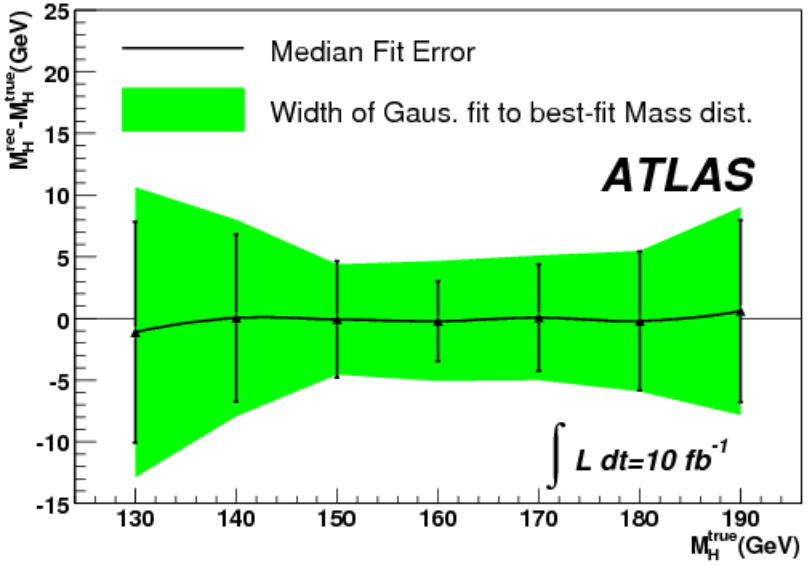
Signal box: $\Delta\varphi_{ll} < 1.5$, $\Delta\eta_{ll} < 1.4$

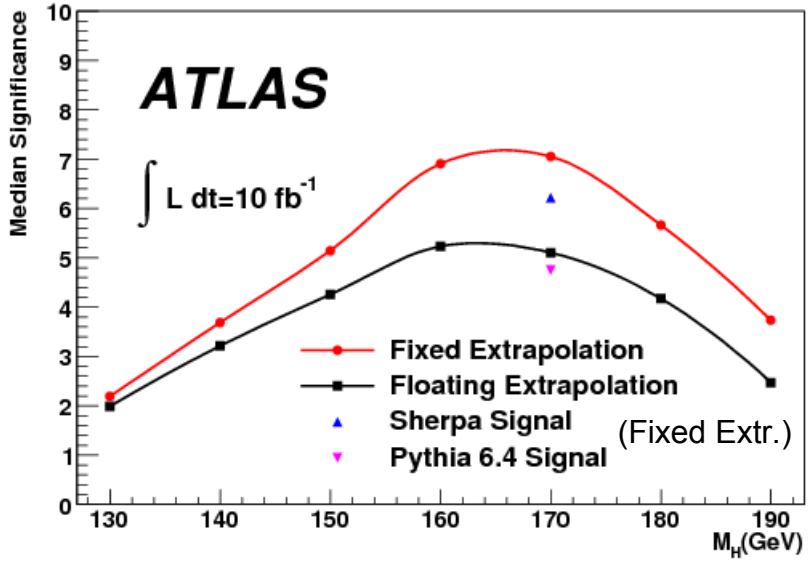
2D Fit - Results

- NN shape independent against:
  - Jet energy scale changed ±5% ($|\eta|<2.5$) and ±10% ($|\eta|>2.5$)
  - Jet p_T thresholds: 20, 30, 40 GeV
  - Q^2 scale uncertainty
- If bkg uncertainties in ratio NN(signal region)/NN(control region) small

  → Extrapolation parameter can be fixed (atm determined on MC)

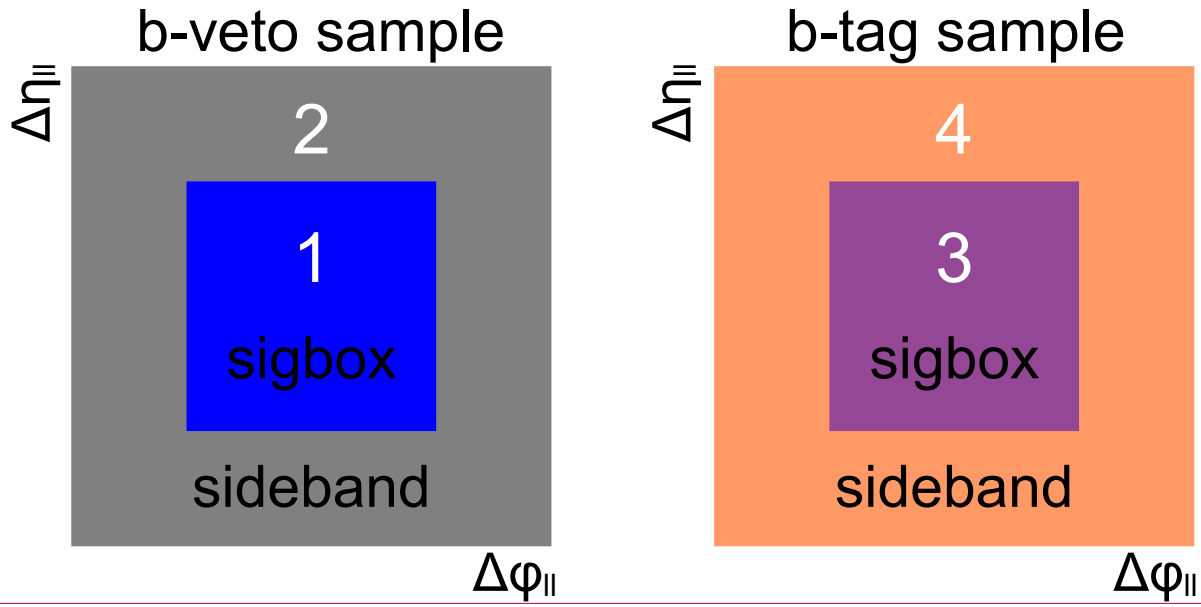
  → 2 scenarios: fixed and floating NN extrapolation parameter





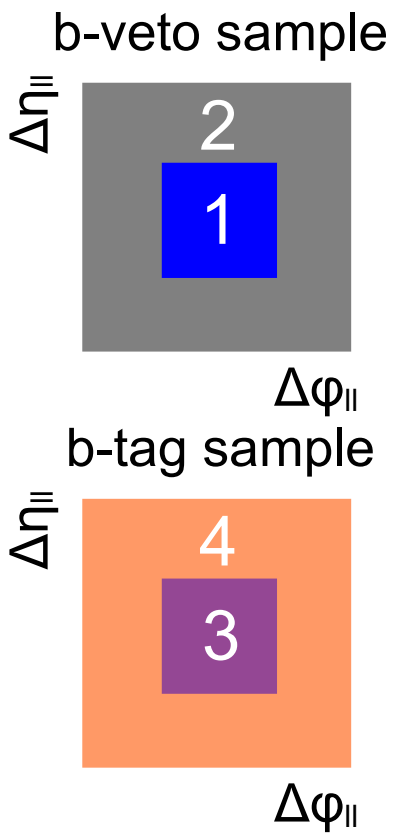
5D Fit – Event Selection

- Event selection similar to 2D fit, changes:
  - 2 jets with $p_T > 15$ GeV and $|\eta| < 4.9$
  - Missing $E_T > 20$ GeV
  - $|\Delta\eta_{jj}| > 2.5$, $m_{jj} = [600, 3000]$ GeV
  - b-jet veto: displaced vertex significance $d_0/\sigma_{d0} > 4.5$
- Unbinned maximum likelihood fit in: M_T, $\Delta\varphi_{ll}$, $\Delta\eta_{ll}$, $\Delta\eta_{jj}$, M_{jj}
- Multidimensional kernel estimation technique [K. Cranmer, hep-ex/0011057]
- Signal box: $|\Delta\varphi_{ll}| < 1.5$ and $|\Delta\eta_{ll}| < 1.4$
- Fit categories:
  - btag, bveto
  - signal box, sideband
  - ee, eμ, μμ

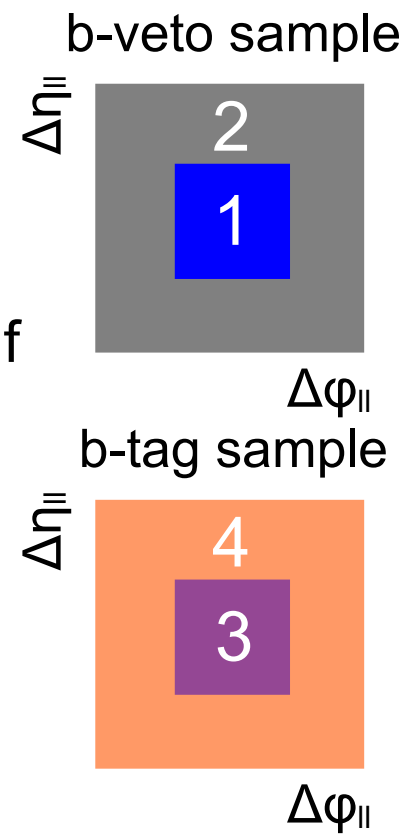


5D Fit - Setup

- Fit for Higgs candidates only in b-vetoed signal box
- Extrapolate bkg from regions 2,3,4 $\rightarrow$ 1
- WW+jets (1,2), $t\bar{t}$ (3,4)
- f_{ee}, $f_{e\mu}$, $f_{\mu\mu}$: relative fractions of events
  - Same for signal and bkg (dominant bkg's have 2 W's)
  - Same for all sample categories
- f_{bveto}: ratio #bveto/#btag for bkg
  - Same for signal box and sideband
- f_{sigbox}: ratio #sigbox/#sideband for bkg
  - Same for bveto and btag categories



5D Fit – PDF Shapes



Signal PDF:

- M_T: 2 sided exponential ⊗ Gauss
- $\Delta\varphi_{ll}$, $\Delta\eta_{ll}$: Simple Gauss
- ($\Delta\eta_{jj}$, M_{jj}): Strongly correlated → 2D kernel estimation pdf

- Only mean of M_T distribution (m_H) free in combined fit
- Independent of lepton flavor
- Largest unmodelled correlation on MC: 14%

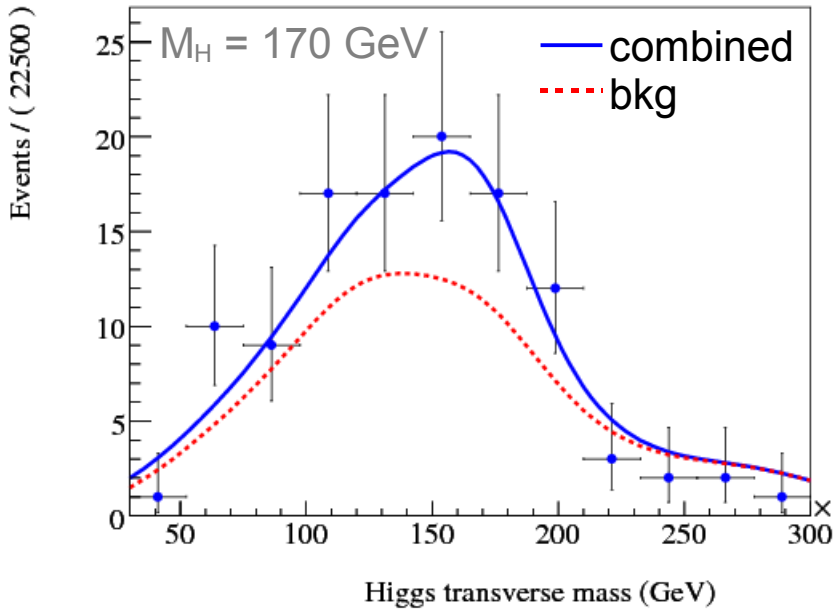
Background PDF:

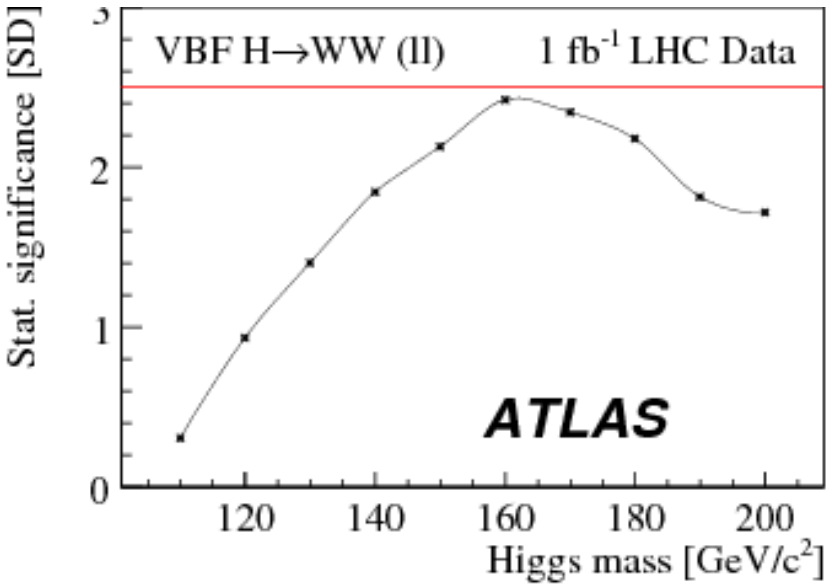
- (M_T, $\Delta\varphi_{ll}$, $\Delta\eta_{ll}$): 3D kernel estimation pdf in region 3
- ($\Delta\eta_{jj}$, M_{jj}): 2D kernel estimation pdf in region 2

- Largest unmodelled correlation on $t\bar{t}$ and WW MC < 10%

5D Fit - Results

- 15k toy MC fits: 1 fb^{-1}
- Bias on N_{Signal}: < 0.8 events, compatible with 0 at $M_H = 160$ GeV
- When not fixing N_{Bkg} in fit → error on N_{Signal} about 25% larger
- Maximum significance ($1fb^{-1}$): 2.5σ at $M_H = 160$ GeV





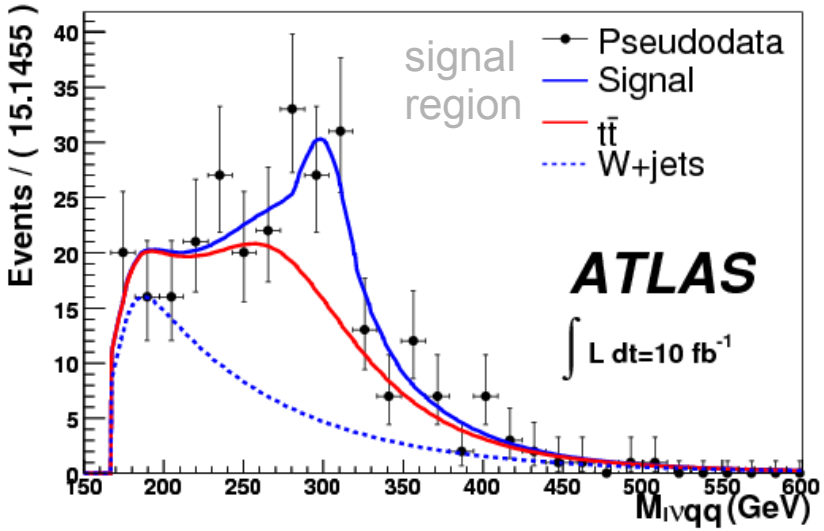
H + 2j ($H \rightarrow WW \rightarrow l\nu qq$)

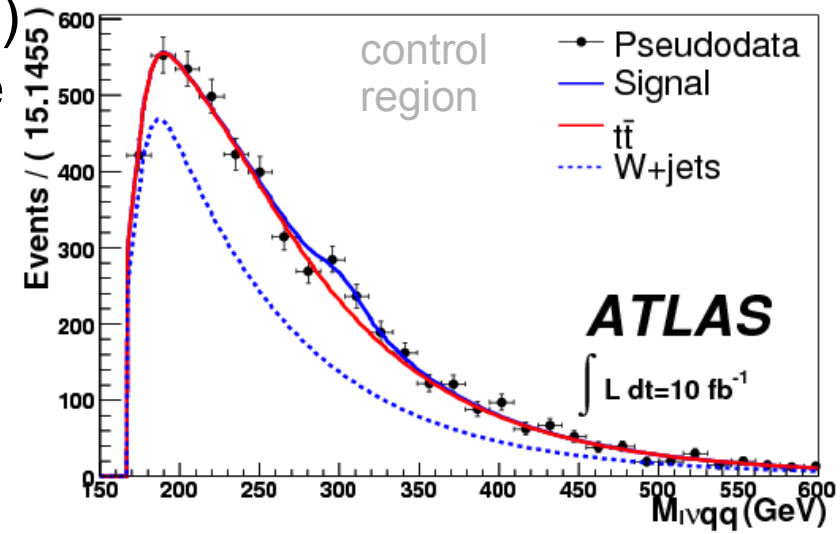
- 1D-Fit (M_H)
- Backgrounds: $t\bar{t}$, W +jets, QCD multijets

1D Fit

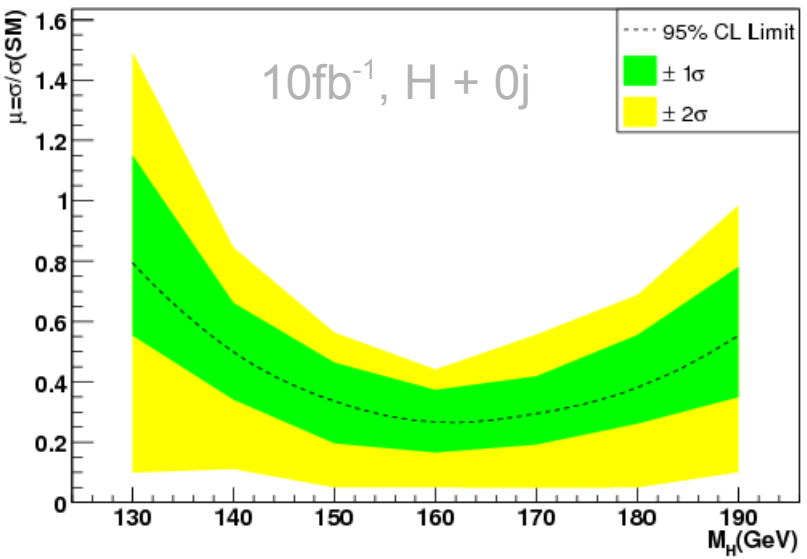
- Reconstruct M_H using W mass constraint to estimate p_z of the neutrino
- Consider out-of-cone corrections to jet energies
- Signal region: $|\Delta\eta_{jj}| > 4.4$
- Control region: $|\Delta\eta_{jj}| < 4$
- S/B ~ 1/5
- Simultaneous binned fit of M_H in signal and control region
- Signal and Bkg shapes:
  - Signal: shape from MC (M_H=300GeV)
  - $t\bar{t}$: from b-tagged control sample
  - W+jets: shape free in fit

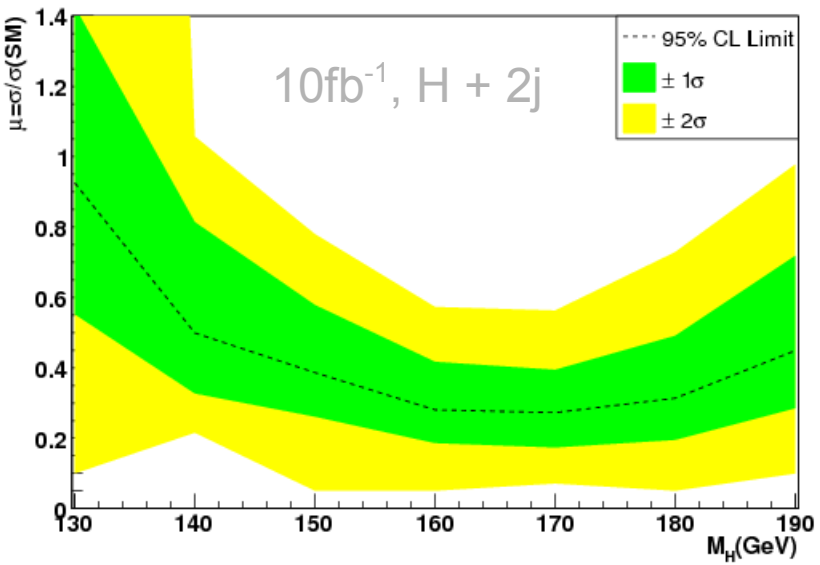
→ No significance calculated:
  - Large uncertainties of W+jets
  - No estimate of QCD multijet bkg

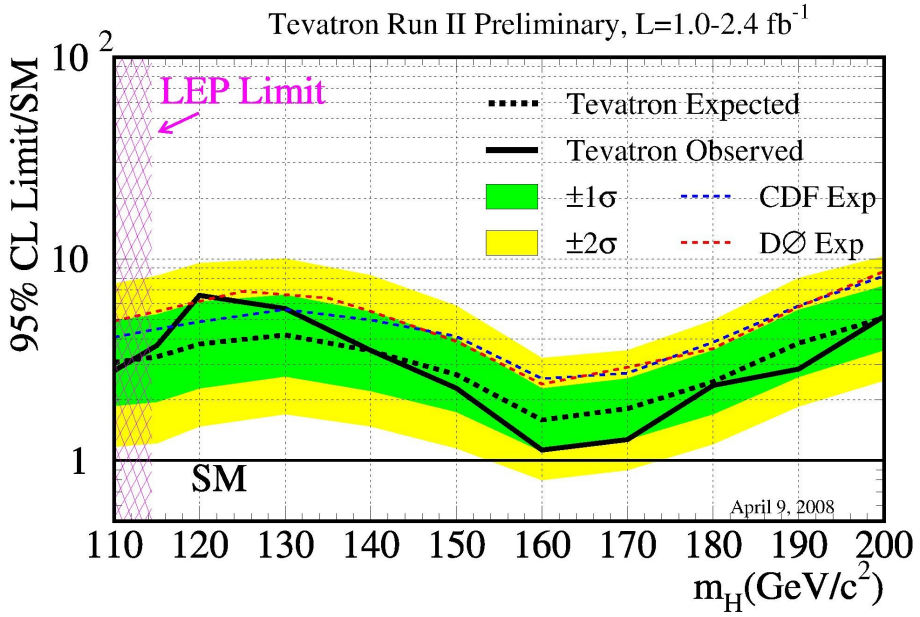




Exclusion Limits

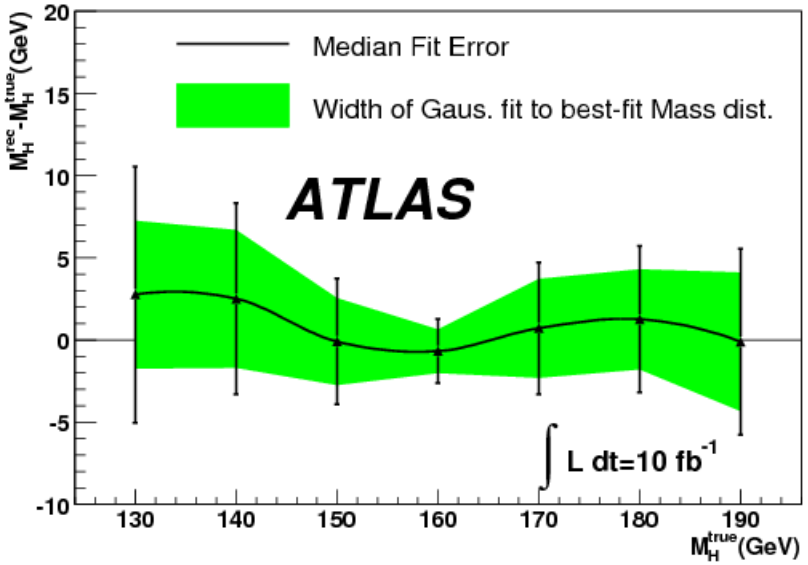


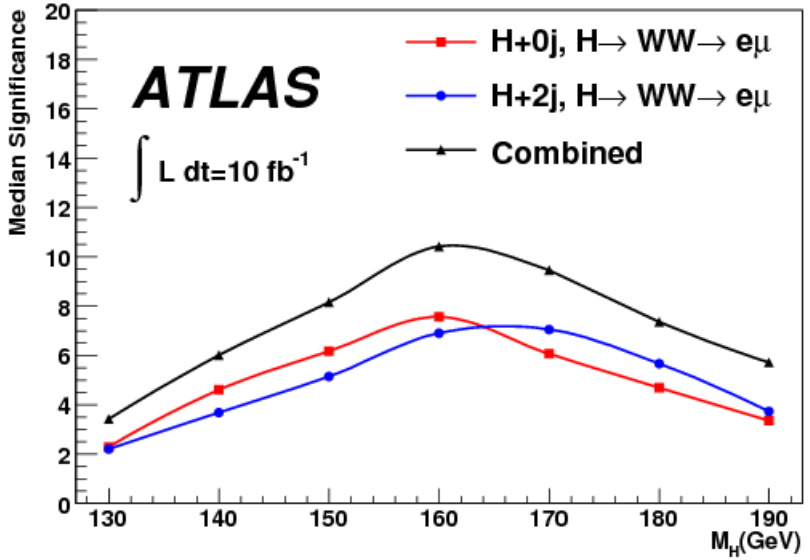




Conclusions

- Most important backgrounds: $t\bar{t}$, WW(+jets), W+jets, $Z \rightarrow \tau\tau$
  $\rightarrow$ Estimation from data seems promising
- $H \rightarrow WW \rightarrow e\nu\mu\nu$ most promising around the WW threshold (160 GeV)
- With 0jet and 2jet channel alone a 5σ discovery can be reached with 10 fb^{-1} for $150 < M_H < 180$ GeV
- Combined fit with shared mass parameter, independent normalization:
  $\rightarrow$ Significance > 5σ for $M_H > 140$ GeV

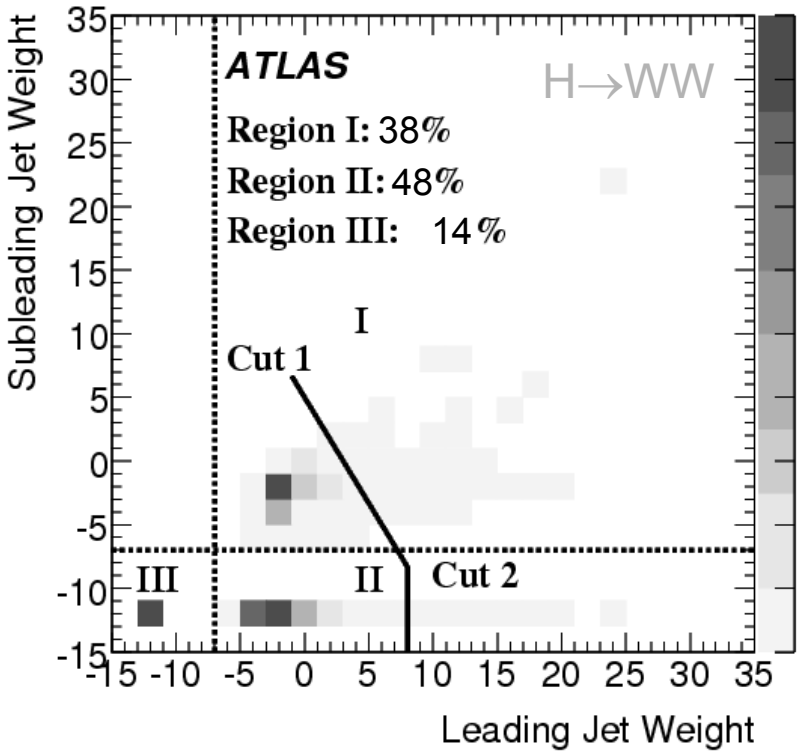


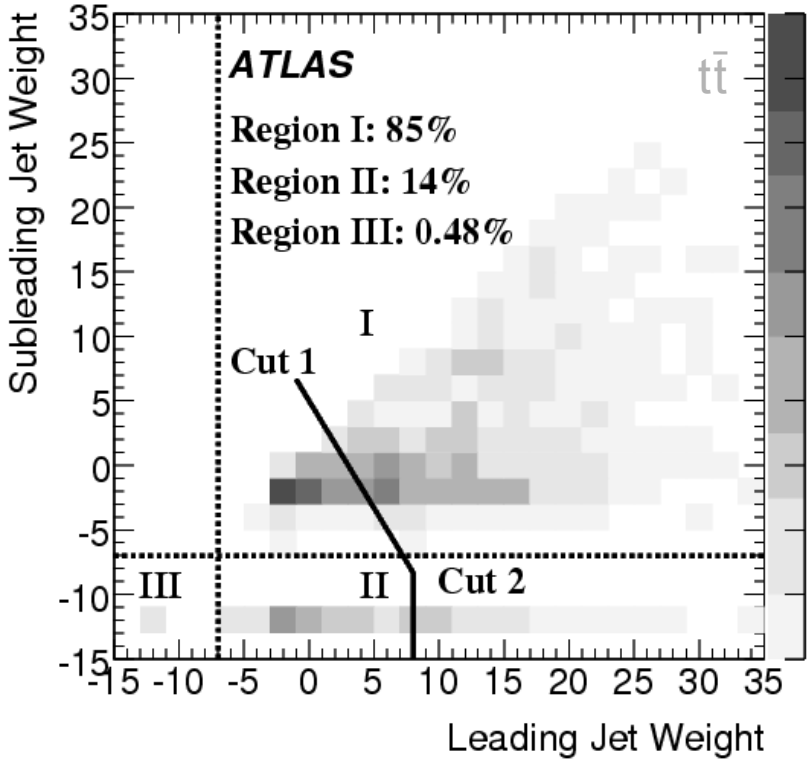


Backup Slides

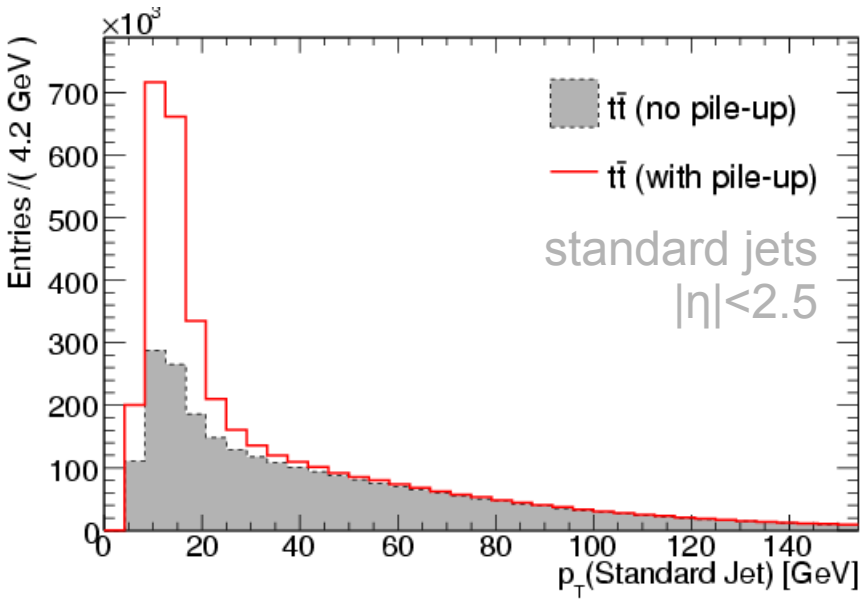
B-tagging

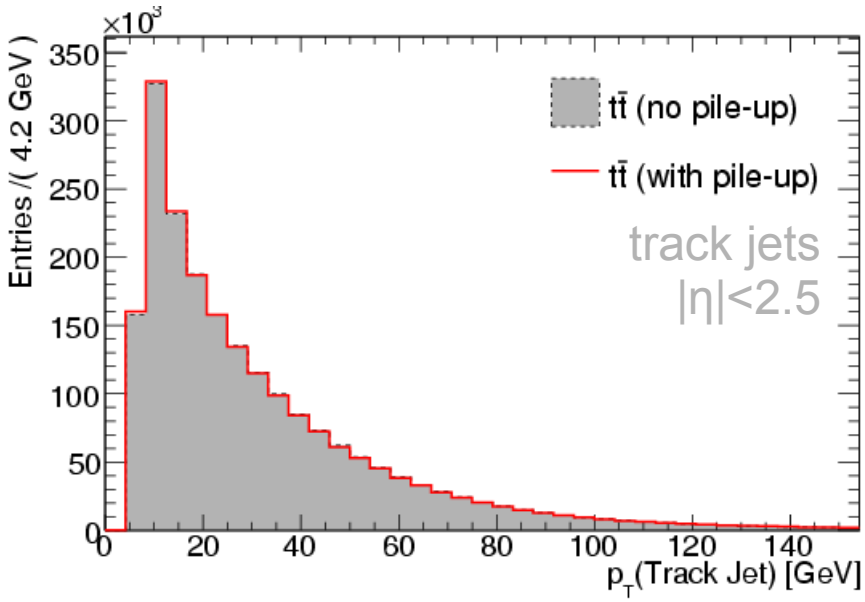
- $t\bar{t}$ most important background → contains 2 jets from b-quarks
- b-weight: IP3D+SV1
- Cut in 2D plane optimized with respect to significance
  - region I: weight(jet_1) + 0.6 · weight(jet_2) < 3
  - region II: weight(jet_1) < 8





Pile-Up and Track Jets



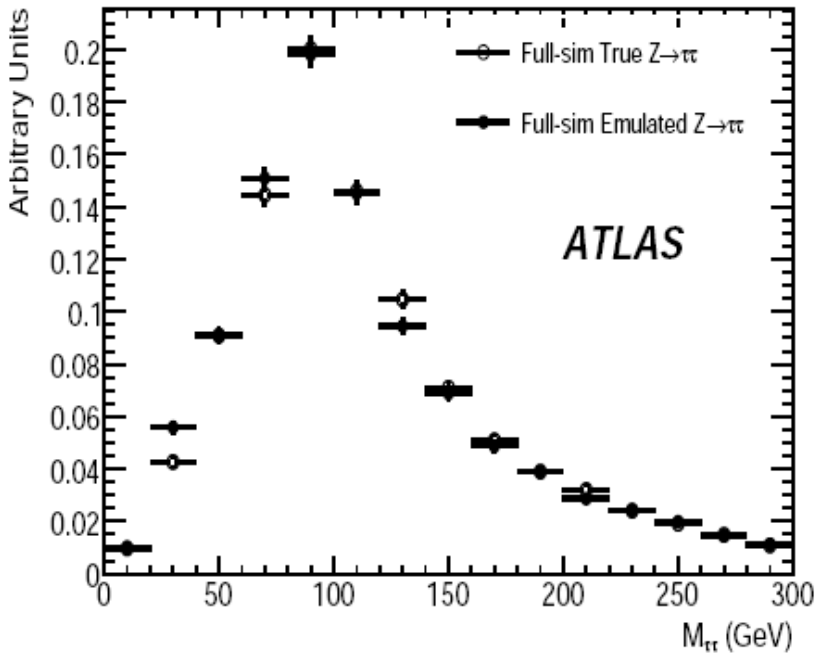


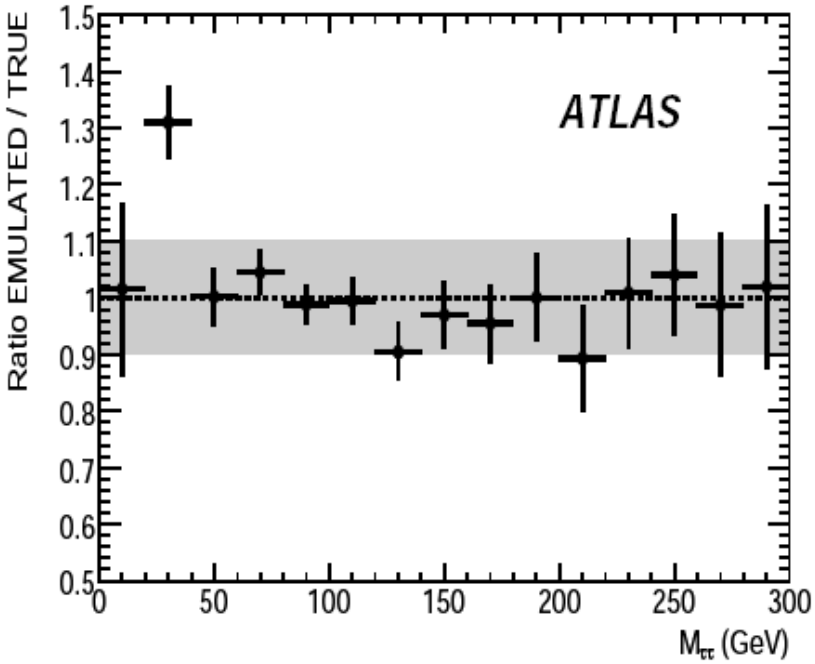
- Recontruct jets based on tracks originating in primary vertex
- Fraction of events passing the central jet veto:

	$H \to WW$		$t\bar{t}$	
	no pile-up	with pile-up	no pile-up	with pile-up
std jets ($\|\eta\| < 2.5$)	72.0 ± 1.0	63.0 ± 1.2	28.6 ± 3.4	19.7 ± 3.3
track jets	72.0 ± 1.0	73.5 ± 1.1	28.6 ± 3.4	25.9 ± 3.6
std jets ($\|\eta\| < 3.2$)	65.4 ± 1.0	57.0 ± 1.2	24.0 ± 3.2	16.3 ± 3.0
combination	65.8 ± 1.0	65.9 ± 1.1	24.0 ± 3.2	23.1 ± 3.5

$Z \rightarrow \tau\tau$ Shape from $Z \rightarrow \mu\mu$ Data

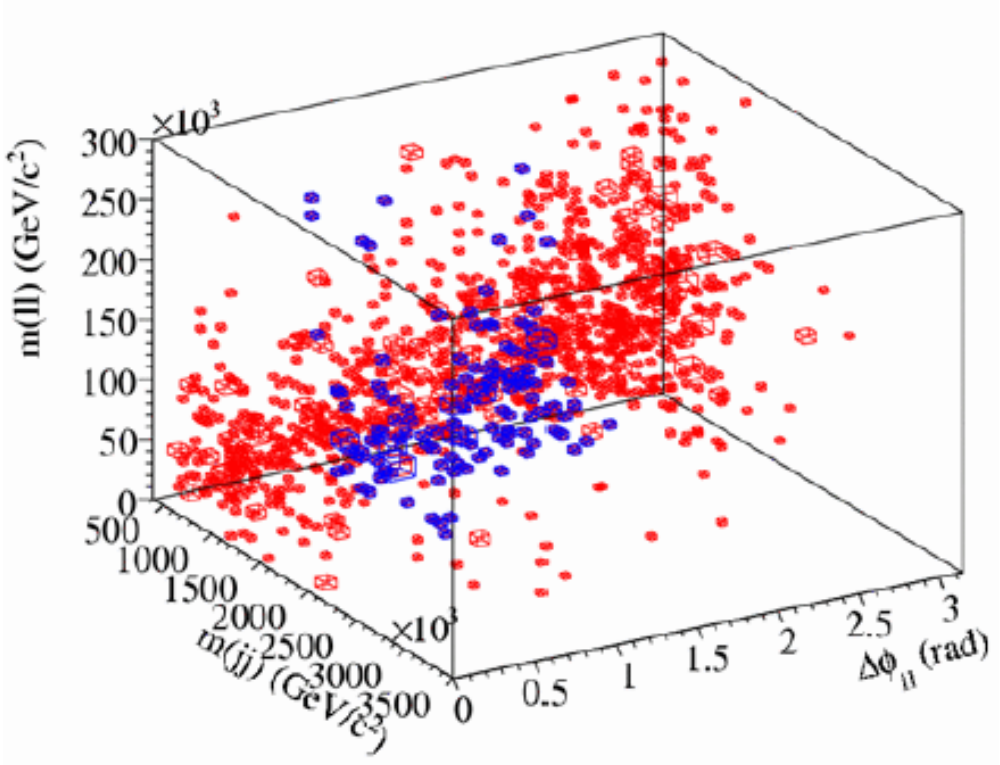
- Use only μ's since the effect on the missing E_T is smaller
- Select clean $Z \rightarrow \mu\mu$ sample on data
- Replace μ's with simulated τ's which decay trough TAUOLA
- Determine shape of $Z \rightarrow \tau\tau$ distribution
- Normalization is the same due to lepton universality

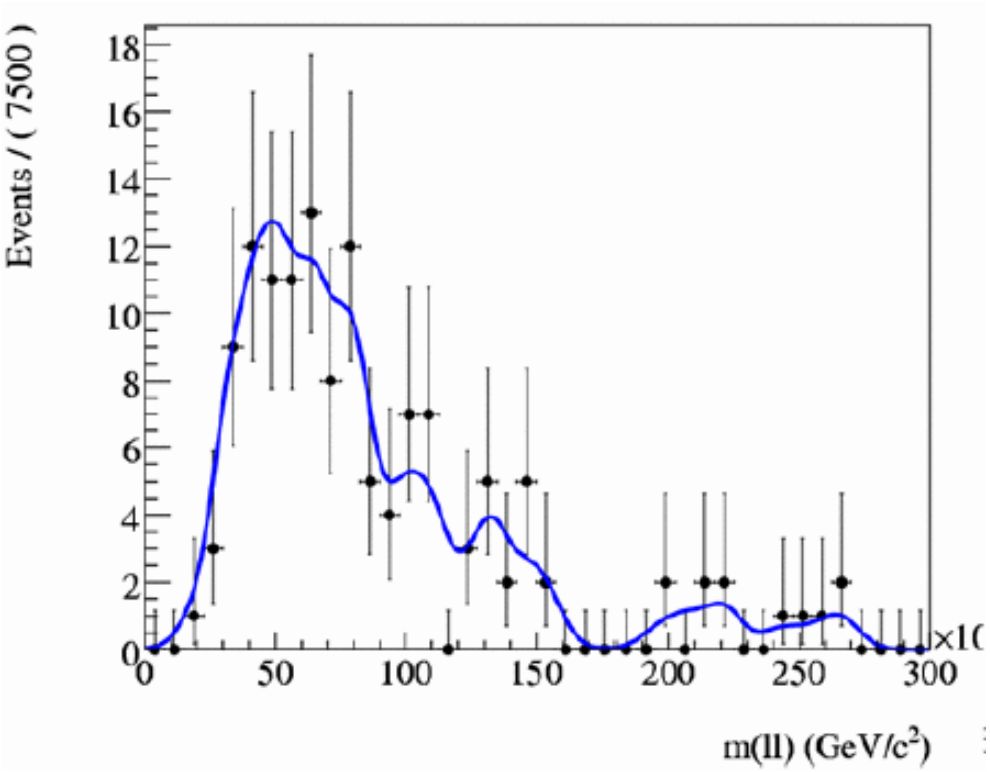




Kernel Estimation Technique

- Kernel estimation pdf: provides unbinned, unbiased estimate pdf for arbitrary set of data [K. Cranmer, hep-ex/0011057]
- 1-dim keys pdf heavily used in Babar
- Multidimensional keys pdf automaticaly includes correct correlations between observables





Transverse Mass M_T

- M_{WW} not reconstructible due to neutrinos
- If $M_H < 2M_W$

  $\rightarrow$ (virtual) W's at rest in higgs system

  $\rightarrow$ $m_{ll} \cong m_{\nu\nu}$

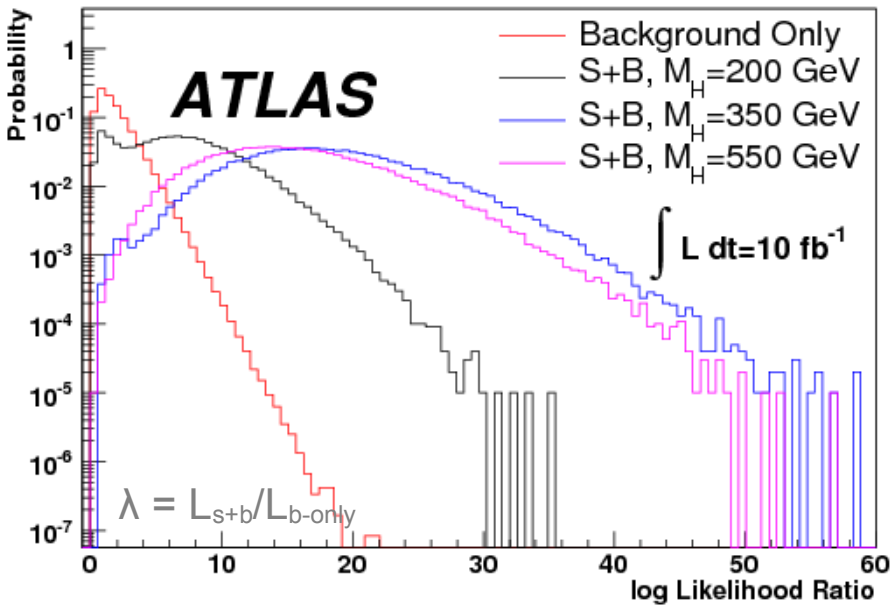
  $\rightarrow$ $E_T^{\ell\ell} = \sqrt{(P_T^{\ell\ell})^2 + m_{\ell\ell}^2}$ $\qquad$ $E_T^{\nu\nu} = \sqrt{(\not{P}_T)^2 + m_{\ell\ell}^2}$

- Transverse mass:

$$M_T = \sqrt{(E_T^{\ell\ell} + E_T^{\nu\nu})^2 - (\vec{p}_T^{\,\ell\ell} + \vec{\not{P}}_T)^2}$$

- Good approxiation upto M_H = 160 GeV
- Can also be useful above 160 GeV

Significance Determination



- Determine likelihood ratio: $\lambda = L_{s+b}/L_{b\text{-}only}$
  - L_{s+b} = Likelihood of full fit
  - $L_{b\text{-}only}$ = Likelihood of fit with N_{sig} fixed to 0
- Obtain the p-value by integration from λ_{Fit} to infinity
- Significance $= \sqrt{2}\mathrm{erfc}^{-1}(2p)$

Significance Determination

- Take some histogram (e.g. higgs mass)
- Model number of entries in each bin as Poisson variable with mean:

$$E[n_i] = \mu L \epsilon_i \sigma_i B_i + b_i \equiv \mu s_i + b_i$$
$$s_i = s_{tot} \int_{bins} f_s(x;\theta_s)dx$$
$$b_i = b_{tot} \int_{bins} f_b(x;\theta_b)dx$$

- μ (signal strength) is the only parameter of interest
- μ=0 $\rightarrow$ no Higgs, μ=1 $\rightarrow$ signal rate as expected in the SM
- The pdf's fs and fb determined from MC or control samples
- Systematic uncertainties included through θ parameters
- Calculate Likelihood function for each chanel i:

$$L_i = (\mu, \theta) = \prod_j \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_k \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

Significance Determination

- Combine likelihood for all channels: $L(\mu, \theta) = \prod L_i(\mu, \theta_i)$
- Construct profile likelihood ratio: $\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L_{max}(\hat{\mu}, \hat{\theta})}$
  - $\hat{\hat{\theta}}$: maximizes L for given μ, $\hat{\hat{\theta}}=\hat{\hat{\theta}}(\mu)$
  - denomiator: maximizes L of full phase space
  - $0 \leq \lambda \leq 1$
- Significance $\approx \sqrt{-2 \ln \lambda(\mu)}$

H + 2j ($H \rightarrow WW \rightarrow l\nu qq$)

Cut	W+jets	$t\bar{t}$	Signal ($M_H = 300$ GeV)
Leptonic W Selection	2353291	128654	174.27
Hadronic W Selection	134483	70872	73.26
Forward Jet Tagging	1076.8	1929	23.16
Lepton Between Jets	867.0	1679	22.93
M_{jj}	131.0	367.7	9.16
Central Jet Veto	57.98	58.24	8.43
$\Delta\eta_{j1,l}$	16.07	47.96	6.93
b-jet Veto	16.07	14.84	6.06
Trigger Selection	13.06	12.40	5.08
$167 < M_{l\nu qq} < 1000$ GeV	13.1±4.7	12.4±3.4	5.08±0.29