Top background to H->WW->2l2   at CMS

TOP 2006 Workshop

Coimbra - 14th January 2006

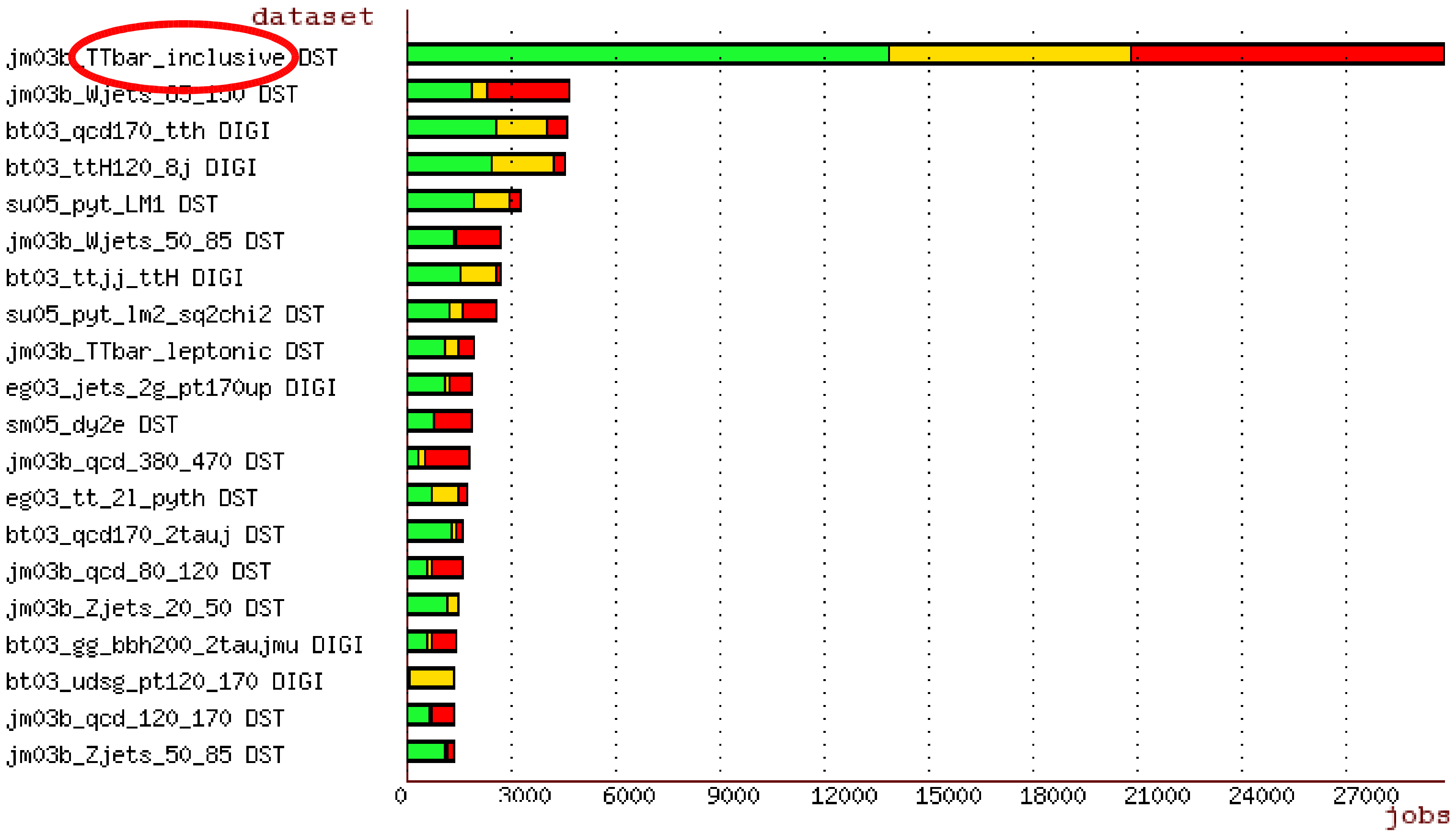
Marco Zanetti - INFN Padova

Introduction

Top as a background at LHC

- Providing with very high luminosity almost all possible final states (b-jets, jets muons, electrons, Missing Et), processes with top are backgrounds to almost every discovery channel at LHC

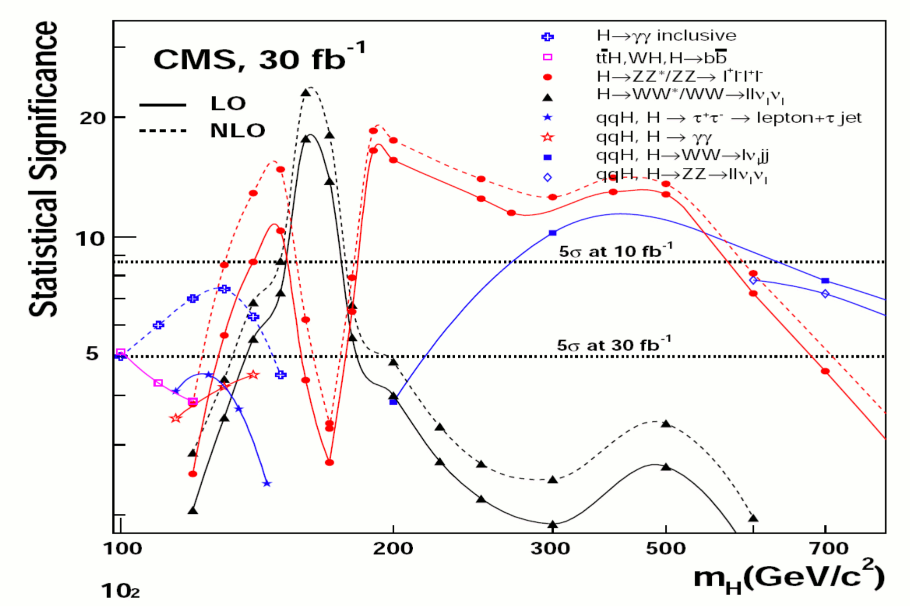
Datasets analysed at CMS

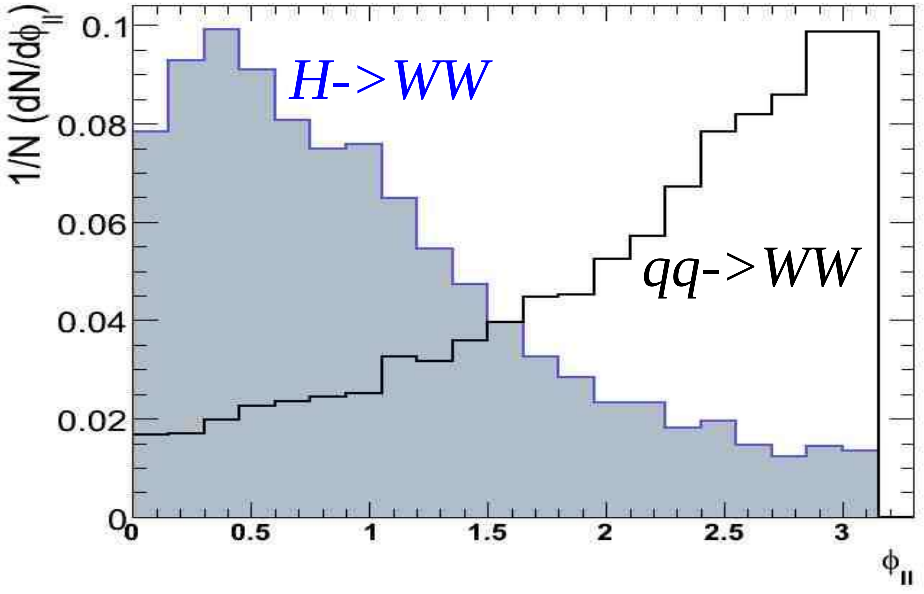


Introduction

the H->WW->2l+X channel

- Main Higgs discovery channel at LHC in the Higgs mass range 150-170 GeV
- Signal features: 2 isolated, high Pt and nearby leptons with small M_{ll}, high missing Et and no jets.
- No invariant mass peak, all the backgrounds processes must be determined from the data
- Discriminating variable is the opening angle between the leptons. Spin correlations matter!
- Main background from *WW*, *tt*, *Wtb* and Drell Yan.





Introduction

Top as a background to H->WW->2l+X

- Huge rejection needed, $\sim 10^{-4}$. How well do our Monte Carlos describe such a narrow phase space area?
- The most powerful cut is a jet veto. A challenge from the theoretical and experimental point of view.
- Once the jet veto is applied, *Wt* is enhanced with respect to *tt*. A reliable double-counting free NLO description of *Wtb* is needed.
- Strategies and related systematics for normalization of *tt* from the data

Selections on signal and top backgrounds

		$\mathrm{H} \rightarrow \mathrm{WW}$ ($\mathrm{m_H} = 165$ GeV)	$\mathrm{t\bar{t}}$	tWb
	$\sigma \times \mathrm{BR}(\mathrm{e}, \mu, \tau)$ [fb]	2360	86200	3400
1)	Trigger	1390 (59%)	57380 (67%)	2320 (68%)
2)	lepton ID	393 (28%)	15700 (27%)	676 (29%)
3)	$\mathrm{E_t^{miss}} > 50$ GeV	274 (70%)	9332 (59%)	391 (58%)
4)	$\phi_{\ell\ell} < 45$	158 (58%)	1649 (18%)	65 (17%)
5)	12 GeV $< m_{\ell\ell} <$ 40 GeV	119 (75%)	661 (40%)	28 (43%)
6)	30 GeV$<$ $\mathrm{p_t^{\ell\,max}}$ $<$55 GeV	88 (74%)	304 (46%)	13 (46%)
7)	$\mathrm{p_t^{\ell\,min}}$ $>$25 GeV	75 (85%)	220 (73%)	9.2 (71%)
8)	Jet veto	46 (61%)	9.8 (4.5%)	1.4 (15%)

Overview

- Monte Carlo studies on *tt*:
  - LO-NLO comparison
  - Shower model effect
  - Spin correlation effect
- *Wtb* simulation:
  - NLO description

Parton level analysis

- Normalization of *tt* from the data:
  - Strategies
  - Experimental systematics

Full CMS detector simulation analysis

Monte Carlo Studies

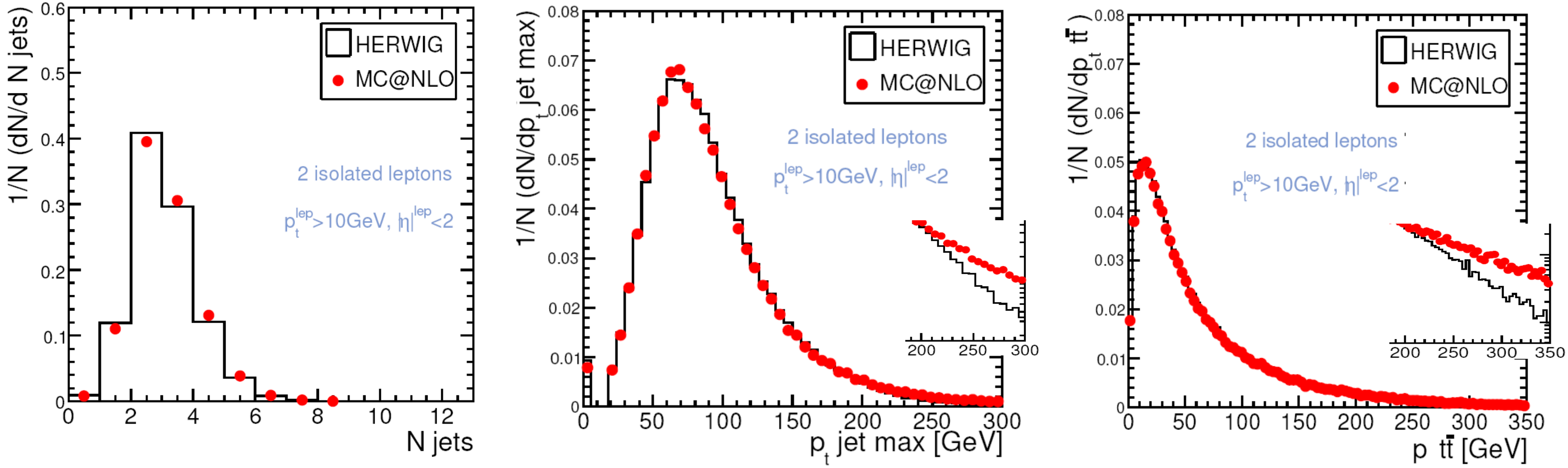
Monte Carlos programs

- PYTHIA and HERWIG are multiprocesses Leading Order generators. They provide additional jet activity through the parton shower
- TopREX is 2->n (n up to 6) generator based on exact matrix element computation. Spin correlations are propagated throughout the whole process. It is linked to PYTHIA for the showering
- MC@NLO combines exact NLO computation with parton shower. It is based on HERWIG for the hadronization step

	PYTHIA 6.227	TopREX 4.11	HERWIG 6.508	MC@NLO 2.31
hadronization model	LUND	LUND	Cluster	Cluster
shower model	Q^2 ordered	Q^2 ordered	angular ordered	angular ordered
$t\bar{t}$ spin correlations	no	yes	yes	no

NLO effects on *tt*

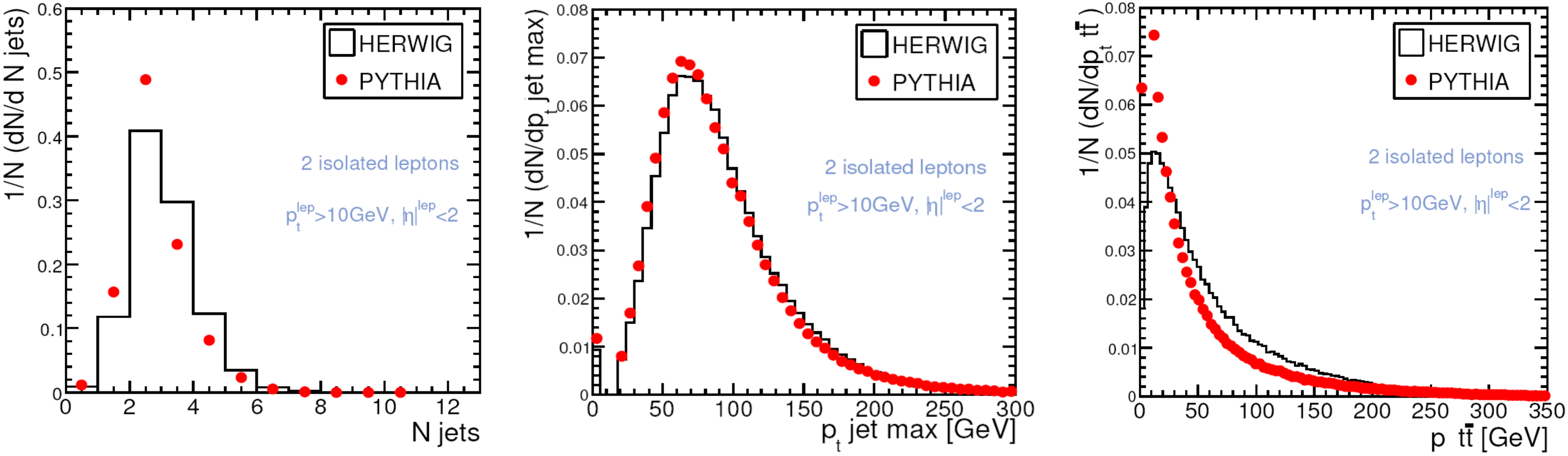
- Comparison of MC@NLO and HERWIG (w/o spin correlation)



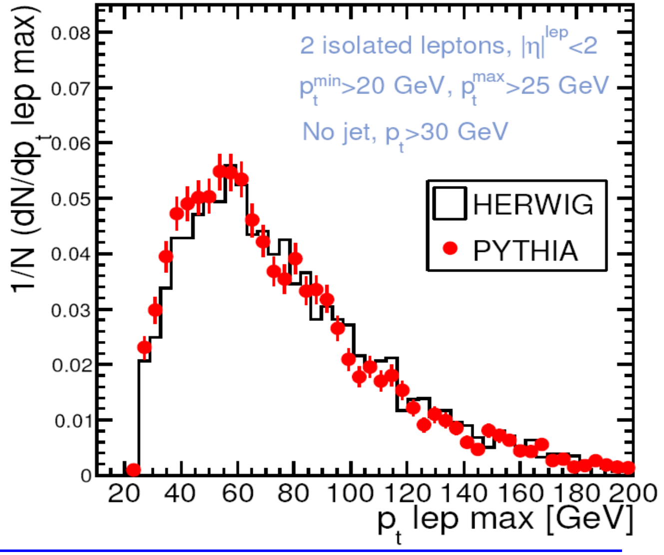
- The bulk of the events presents no major differences neither in jets spectra nor in jets multiplicity
- Typical NLO-LO discrepancy in the high Pt part of the spectra
- Relatively small difference of the jet veto efficiency (<10%)
- Leptons' kinematics very similar. Safe to simply rescale the cross section to the NLO one (0.4 -> 0.8 nb).

Showering Models

- Comparison between PYTHIA (Lund model) and HERWIG (cluster model)

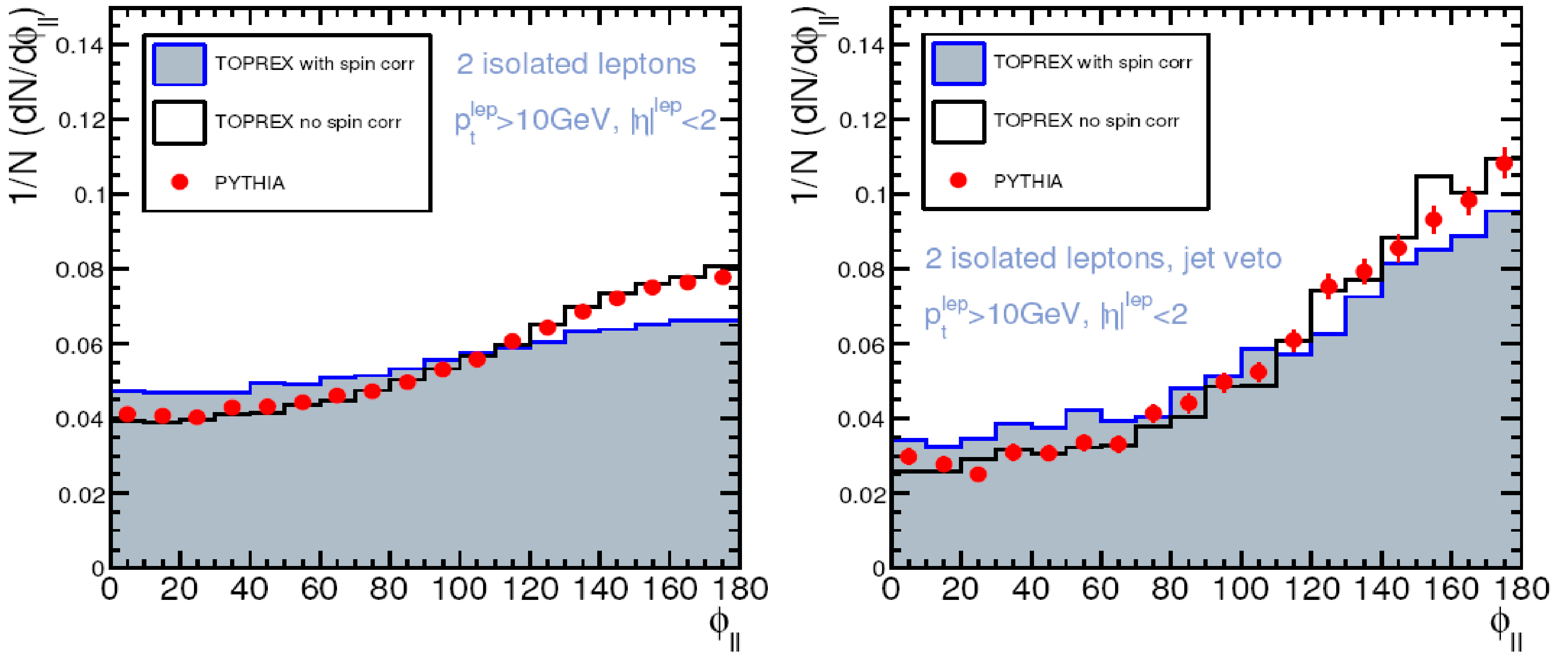


- PYTHIA produces less and softer jets.
- Sizeable difference in jet veto efficiency (>15%)
- Leptons' kinematics very similar down to the signal region



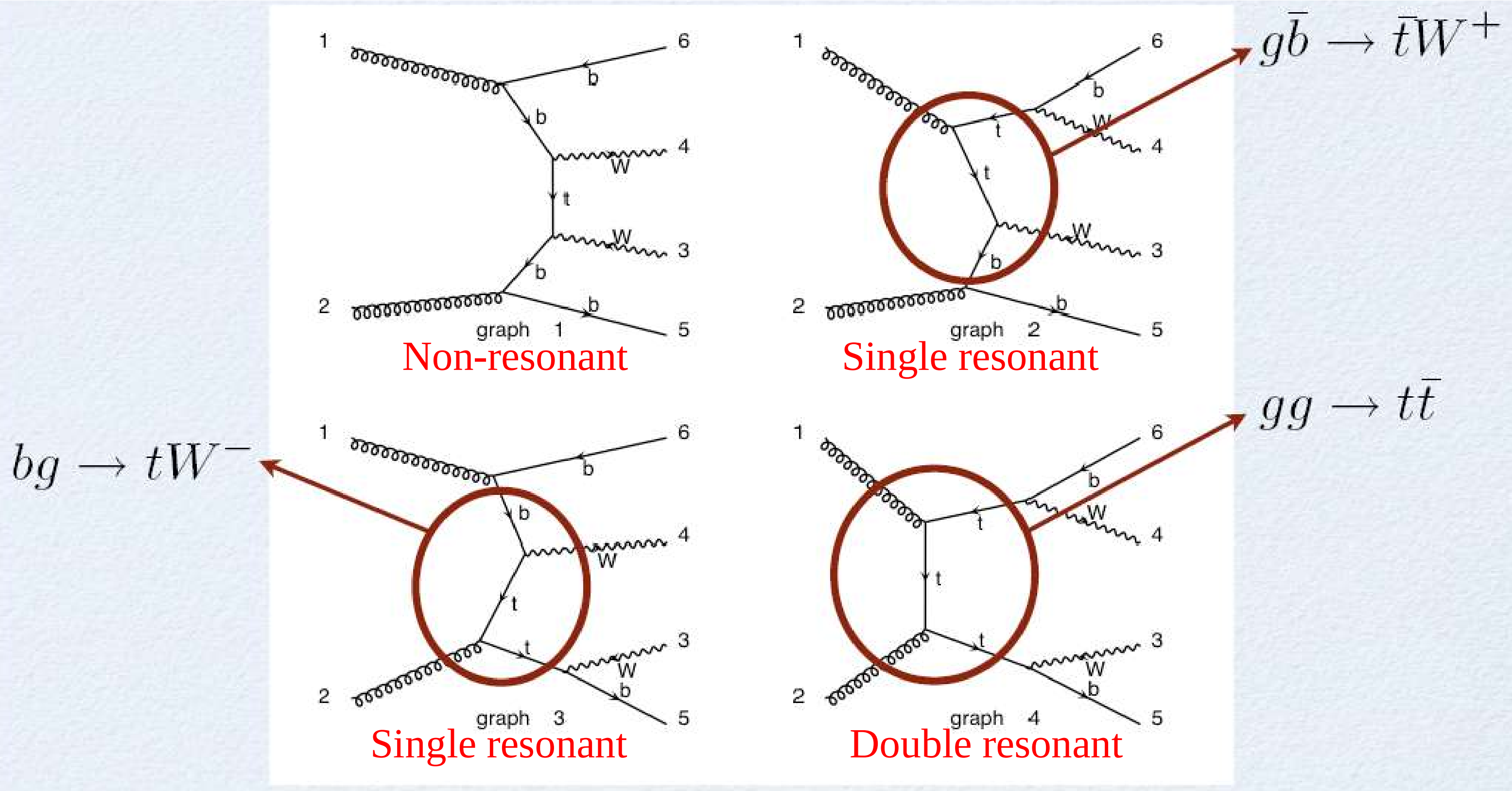
Spin Correlation

- Comparison between PYTHIA (w/o spin correlation) and TopRex (w/ spin corr.)



- *tt* spin correlation tends to flat the $\Delta\phi_{ll}$ distribution.
- $\Delta\phi_{ll}$ distribution more similar after the jet veto
- *tt* spin correlation accounts for a difference of ~10% in selection efficiency

Wtb Simulation



- NLO description of singly resonant top production includes the LO doubly resonant processes
- A solution that allows a Monte Carlo implementation is the “b-PDF” approach: a veto is applied on the p_t of the spectator *b* corresponding to the factorization scale (J. Campell, F. Tramontano)
- Naturally extendible to H->WW signal region (jet veto)

Wtb Simulation

- Comparison between MCFM (LO and NLO with $p_t^{b\ veto} = \mu_f = 40$ GeV) and TopREX (LO)
- To match the MCFM partonic veto, for TopREX events (with the shower included) the jet veto has been shifted down to 30 GeV

	MCFM				TopREX
	LO		NLO		LO
Selection cuts	$\sigma \times$ BR (fb)	rel. eff	$\sigma \times$ BR (fb)	rel. eff	rel. eff
No cuts	271		377		
2 lep, $\lvert\eta\rvert < 2$, $p_t > 20$ GeV	204	0.75±0.002	277	0.73±0.002	
$E_t^{miss} > 40$	148	0.73±0.002	209	0.75±0.003	0.75±0.001
$\phi_{\ell\ell} < 45$	20.8	0.14±0.002	34.4	0.16±0.002	0.17±0.001
5 GeV $< m_{\ell\ell} <$ 40 GeV	10.6	0.51±0.01	15.6	0.45±0.008	0.50±0.005
Partonic jet veto, 40 GeV	1.55	0.15±0.01	1.12	0.07±0.05	0.16±0.005
30 GeV$< p_t^{\ell\ max} <$55 GeV	1.08	0.70±0.03	0.73	0.65±0.05	0.63±0.002
$p_t^{\ell\ min} >$25 GeV	0.73	0.68±0.04	0.49	0.67±0.05	0.67±0.002

- Similar efficiencies for lepton selections between MCFM LO and NLO
- Similar efficiencies for lepton selections between MCFM LO and TopREX
- NLO descriptions ends to a K-factor smaller than 1 (~0.7)

tt Normalization from data

tt normalization from data

the method

- Identify a phase space region where *tt* is enhanced. From the measured number of events there, estimate the *tt* background in the signal region exploiting the relation:

$$N^{estimated}_{signal\ \ reg.} = \frac{N^{Monte\,Carlo}_{signal\ \ reg.}}{N^{Monte\,Carlo}_{control\ \ reg.}} N^{measured}_{control\ \ reg.} = \frac{\sigma_{signal\ \ reg.}}{\sigma_{control\ \ reg.}} \frac{\epsilon_{signal\ \ reg.}}{\epsilon_{control\ \ reg.}} N^{measured}_{control\ \ reg.}$$

- The ratio $(N_{\text{signal reg}}/N_{\text{control reg}})$ carries smaller theoretical and experimental systematics uncertainties
- A candidate control region:
  - should be dominated by *tt* events
  - should be as close as possible to the signal region
- Two approaches are proposed:
  - Drop the jet veto + requiring 2 b-tagged jets (optimal choice)
  - Drop the jet veto + requiring 2 high Et jets (to avoid b-tag uncertainties)
- For both approaches the same cuts on leptons and missing Et are used

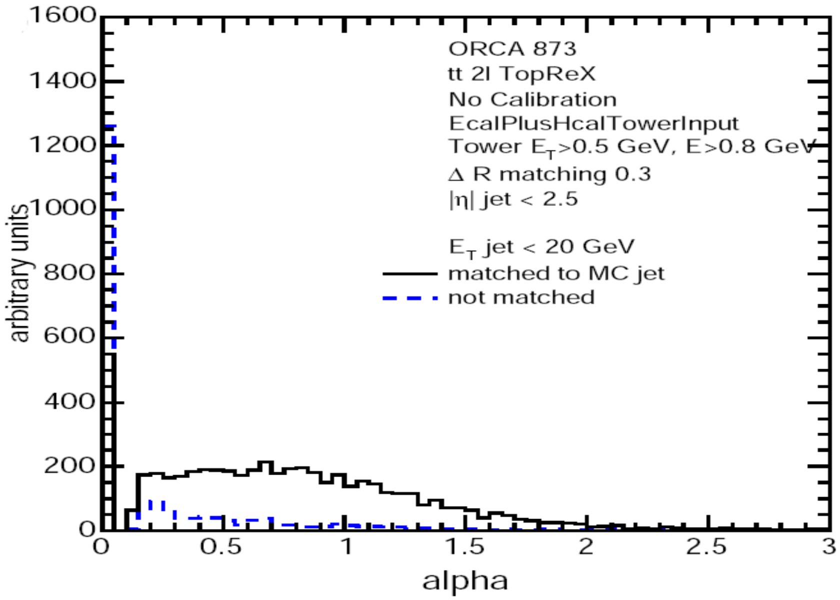
Signal region

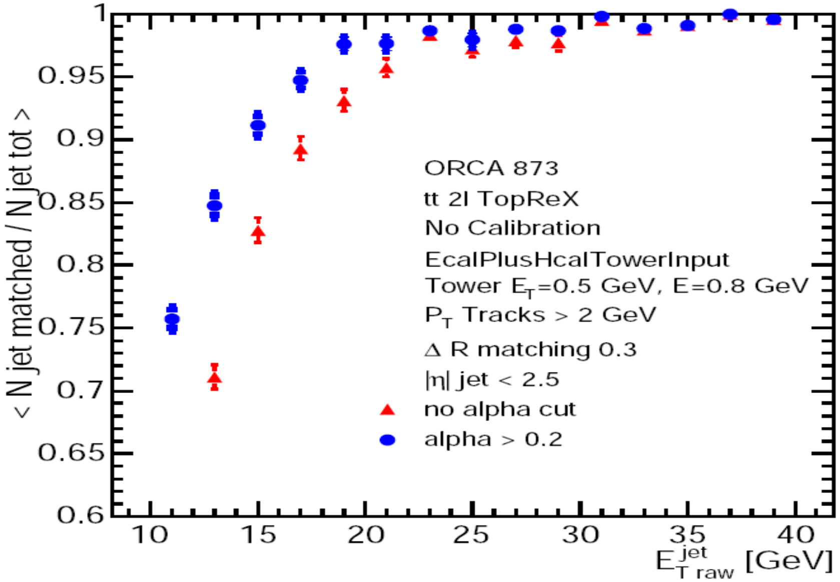
Experimental Jet Veto

- The LHC will provide events with multiple interactions and thousands of charged and neutral tracks. Considering also the 4 Tesla CMS magnetic field, this is a very hostile environment for jet reconstruction at low E_t.
- Below measured $E_t = 20$ GeV, the tracks within the jet cone belonging to the leptons vertex are used to discriminate between real and fake jets via the α parameter:

$$\alpha=\frac{\Sigma P_t(tracks)}{E_t(jet)}$$

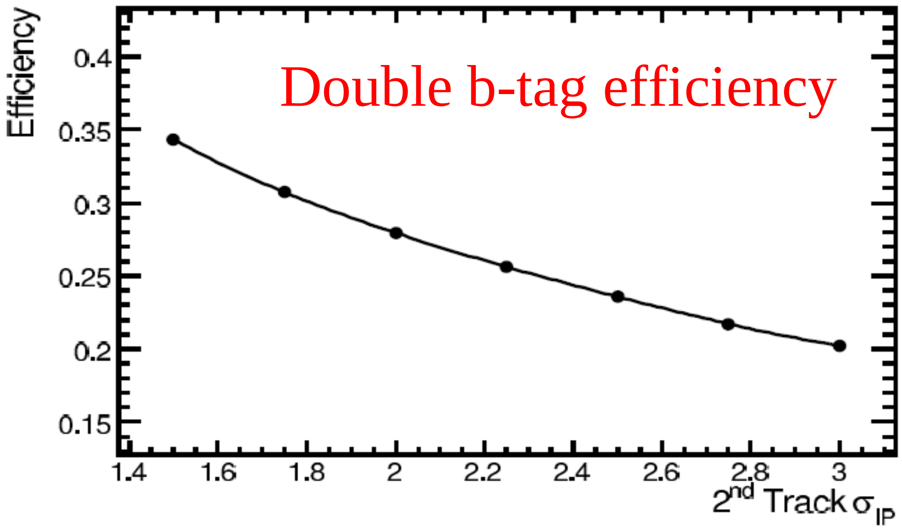
- Jets with 15 < Et < 20 GeV are required to have $\alpha > 0.2$

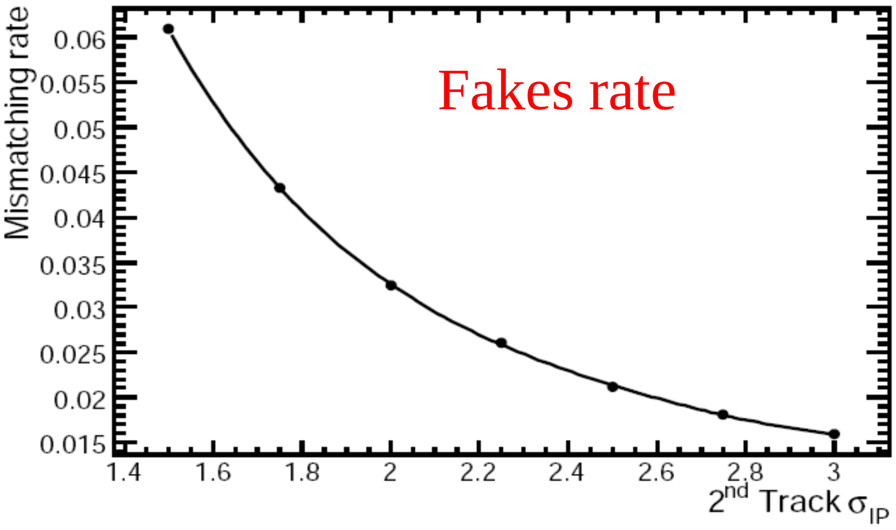




Double b-tag control region

- “Track counting” b-tagging algorithm: σ_{IP} of the second most energetic (p_t) track associated to a jet used as discriminating variable





- Requiring jets with E_t>20 GeV and σ_{IP} > 2, leads to efficiency ~ 30% and fakes rate ~ 3%.

Events for 10 fb^{-1} (at NLO)

	2μ	$2e$	$e\mu$
$t\bar{t}$	194	107	245
Wt	1	< 1	2
$Signal(m_H = 165)$	< 1	< 1	1

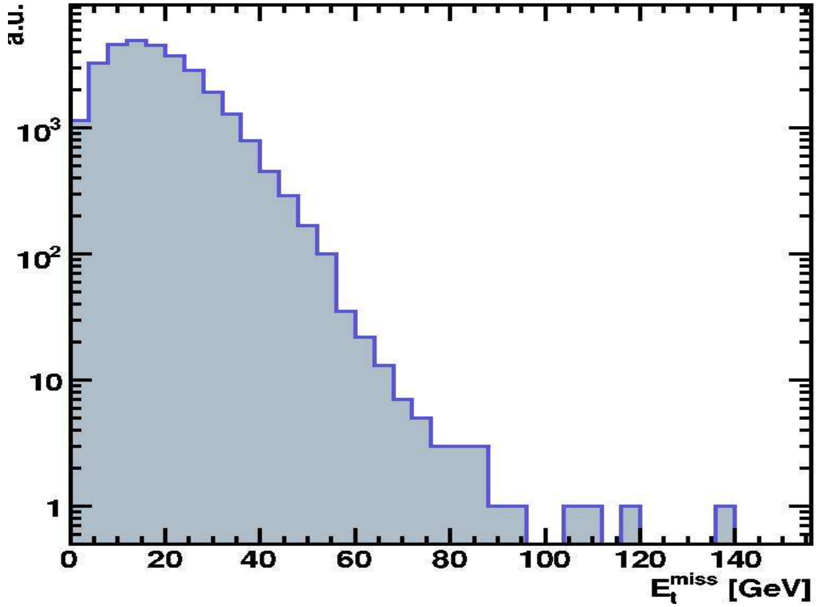
Double b-tag control region

Possible backgrounds

Analysis at parton level on MadGraph samples

- WWbb; suppressed with respect to *tt->WbWb* (σ < 1 pb). Efficiency for signal selections ~10^{-3}. Including b-tag efficiency => negligible.
- DYbb; (for ee or μμ final state); ~200 events foreseen for 10 fb^{-1} but negligible due to missing Et cut (eff. ~ 10^{-2}) and b-tag efficiency.

Missing Et distribution for fully simulated DY+2b events



Two high Et jets control region

- Alternative method that allows to avoid experimental systematics related to b-tagging.
- Jets are reconstructed by an iterative algorithm with cone size = 0.5 and tower thresholds set to $E_t = 0.5$ and $E = 0.8$.
- DY+2j has much higher cross section than DY+2b. Although missing E_t cut, its contribution is relevant => only eμ final state are considered.
- The jets Et thresholds are set to 50 and 30 GeV respectively (optimized with respect to *H->WW* and *Wt*)
- The number of events expected for 10 fb^{-1} for *tt*, the signal and *Wt* are respectively 411, 11 and 6.

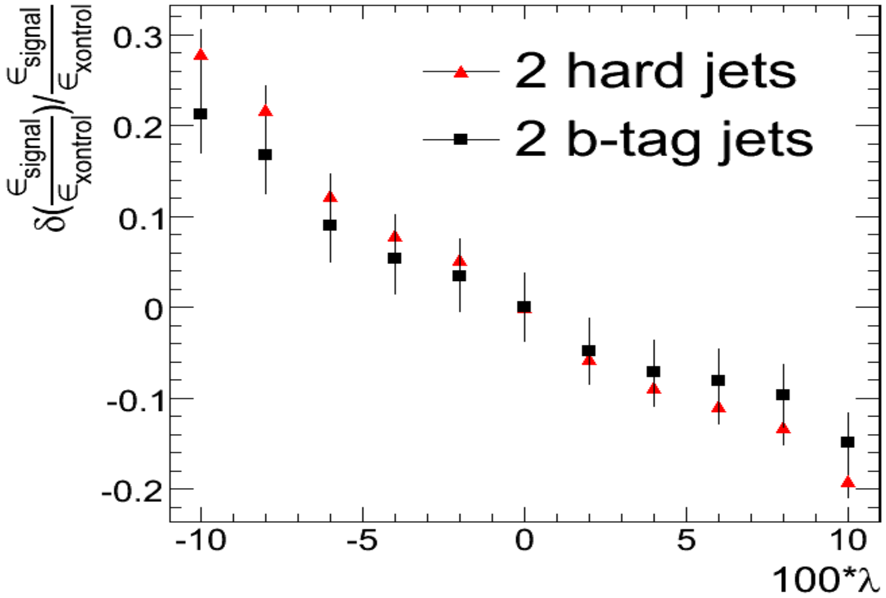
Possible backgrounds

- WWjj; σ~ 0.4 pb after geometrical acceptance cuts. Efficiency for signal selections $\sim 5 \cdot 10^{-4}$. => negligible
- Wjjj; (when a jet is misidentified as electron); σ~ 200 pb. Negligible due to cuts eff. $\sim 10^{-4}$ and low misidentification rate ($\sim 10^{-4}$)

Systematic uncertainties

Experimental uncertainties

- Jet Energy Scale (JES)
  - Estimated by rescaling the jet momentum, $P_\mu = (1 + \lambda) P_\mu$. λ represents the precision with which the energy scale is known.
  - The errors on ε_{signal} (related to the jet veto) and $\varepsilon_{control}$ (related to the request of 2 jets) are strongly anticorrelated. To evaluate the uncertainty, consider the error on their ratio.



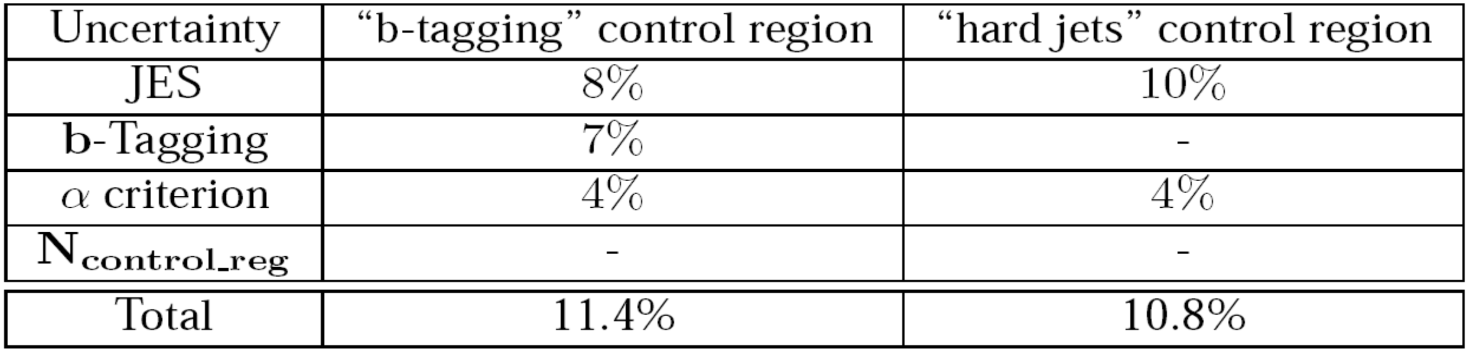
  - For a JES uncertainty ~5% (reasonable at CMS with 10 fb^{-1}) , the error on the ratio ~ 10%

Systematic uncertainties

Experimental uncertainties

- α parameter
  - varying the cut value from 0.15 to 0.25 and rescaling the track Pt threshold from 2 to 3 GeV, the ε_{signal} variation is very small, ~4%
- b-tagging
  - With 10 fb^{-1} at CMS the b-tagging efficiency is foreseen to be known with 7% precision thanks to the calibration on *tt* data
- uncertainty on $N_{control}$:
  - Negligible for both 2 b-tagged jets and 2 high Et jets Control Regions

Experimental systematics summary



Uncertainty	"b-tagging" control region	"hard jets" control region
JES	8%	10%
b-Tagging	7%	-
α criterion	4%	4%
$N_{control\_reg}$	-	-
Total	11.4%	10.8%

Conclusions

- The searches for the Higgs via the decay H->WW->2l+2ν offers the chance to study *tt* and *Wt* in narrow and peculiar phase spaces. This is a major challenge from the theoretical and experimental point of view

- The comparison between a set of Monte Carlo programs on *tt* shows that:
  - NLO effects can be safely included as a global K-factor
  - Showering models differs in predicting jets multiplicity and spectra
  - *tt* spin correlations accounts for an effect on selection efficiency ~ 10%

- *Wtb* can be reliably calculated at NLO in the signal region. The global K factor that can be applied is 0.7
- Two strategies for the *tt* normalization from data have been studied, one based on a double b-tagging and the other on two high Et jets.
- Each approach provides a reliable control region dominated by *tt* events and leads to an experimental systematics ~11%

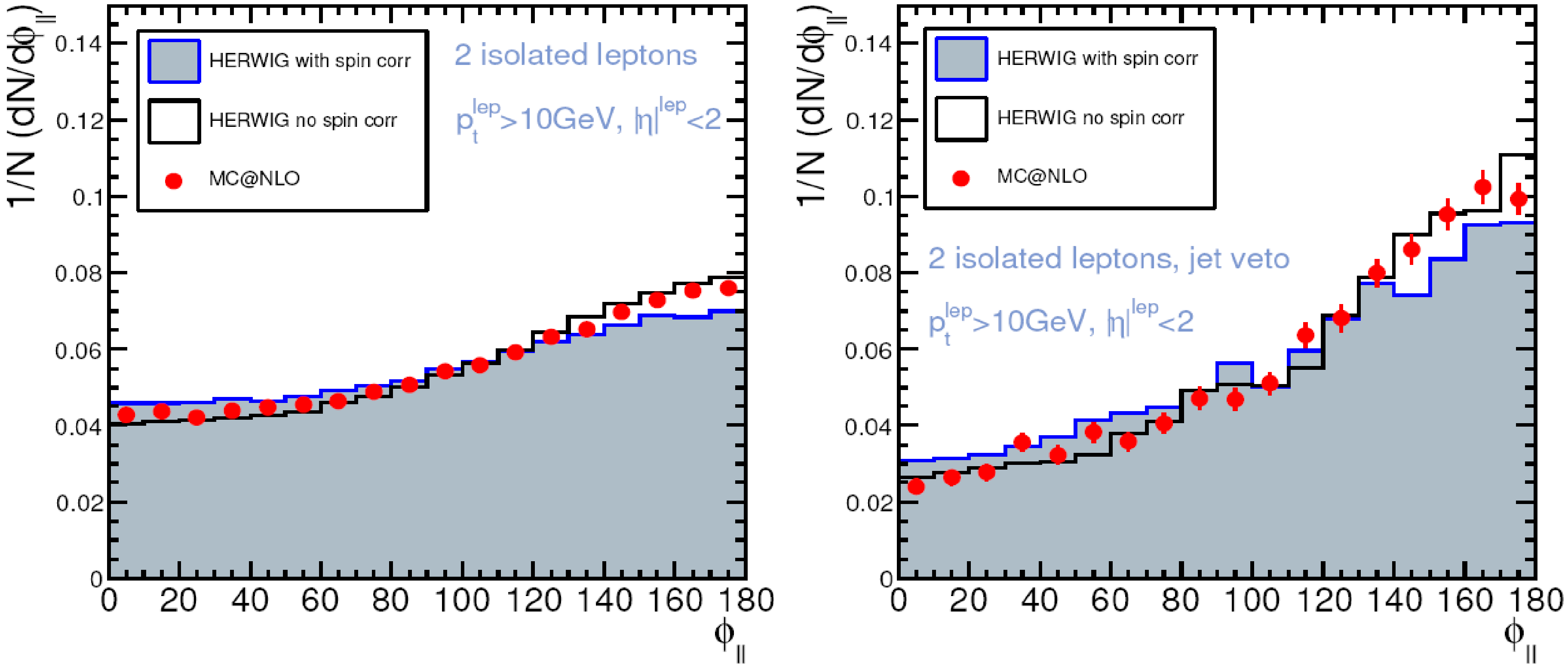
BAKUPS

tt Monte Carlos summary tables

	MC@NLO 2.31		HERWIG 6.508			
	without spin correlations		without spin correlations		with spin correlations	
	nr of events	rel. eff.	nr of events	rel. eff.	nr of evts	rel. eff.
2 isol. leptons	280656	0.2807 ± 0.0004	284876	0.2849 ±0.0004	288015	0.2880±0.0004
$\lvert\eta^{\ell ep}\rvert < 2$	197614	0.7041± 0.0009	193553	0.6795±0.0009	196034	0.6806±0.0009
jet veto	5764	0.0292± 0.0004	6159	0.0318±0.0004	6046	0.0308±0.0004
$E_t^{miss} > 40$	4027	0.699±0.006	4414	0.717±0.006	4489	0.743±0.006
$\phi_{\ell\ell} < 45^\circ$	608	0.151± 0.006	632	0.143±0.005	724	0.161±0.006
5 GeV < $m_{\ell\ell}$ < 40 GeV	354	0.58±0.02	379	0.60±0.02	416	0.57± 0.02
30 GeV< $p_t^{\ell\,max}$ <55 GeV	164	0.46±0.02	194	0.51±0.03	191	0.46±0.02
$p_t^{\ell\,min}$ >25 GeV	71	0.43± 0.04	76	0.39±0.04	77	0.40±0.04

	PYTHIA 6.227		TopReX			
	without spin correlations		without spin correlations		with spin correlations	
	nr of events	rel. eff.	nr of events	rel. eff.	nr of evts	rel. eff.
2 isol. leptons	281624	0.2816 ± 0.0004	293670	0.2937 ± 0.0005	295707	0.2957 ± 0.0005
$\lvert\eta^{\ell ep}\rvert < 2$	195343	0.6936±0.0009	203689	0.6936 ± 0.0009	205605	0.6953 ±0.0009
jet veto	7128	0.0365±0.0004	7804	0.0383±0.0004	7834	0.0381±0.0004
$E_t^{miss} > 40$	4976	0.698±0.005	5442	0.697±0.005	5586	0.713±0.005
$\phi_{\ell\ell} < 45^\circ$	731	0.147±0.005	801	0.147±0.005	962	0.172±0.005
5 GeV < $m_{\ell\ell}$ < 40 GeV	434	0.59±0.02	499	0.62±0.02	594	0.62±0.02
30 GeV< $p_t^{\ell\,max}$ <55 GeV	214	0.49±0.02	258	0.52±0.02	296	0.50±0.02
$p_t^{\ell\,min}$ >25 GeV	85	0.40±0.03	113	0.44±0.03	125	0.42±0.03

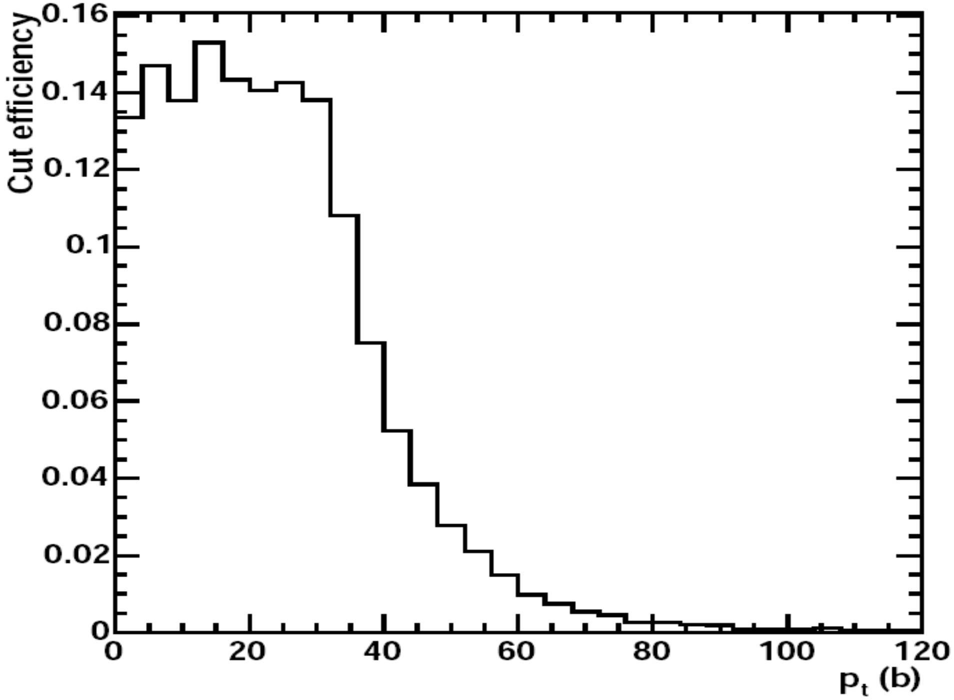
HERWIG-MC@NLO spin correlation effect



- The difference between TopREX w/ and w/o spin correlation is slightly bigger than in the HERWIG MC@NLO comparison. This may be due to the fact that TopREX does not provide gluon radiation from the top

Wtb Simulation

- To match the selections applied on MCFM and TopREX samples the parton level veto on spectator b set at 40 GeV (in the MCFM case) is shifted to 30 GeV after the showering (for the TopREX case)
- When 2 leptons within $|\eta|=2.5$ are required and a jet veto with a threshold at 30 GeV is applied, 85% of the events have $p_t^b < 40$ GeV and 94% have $p_t^b < 60$ GeV

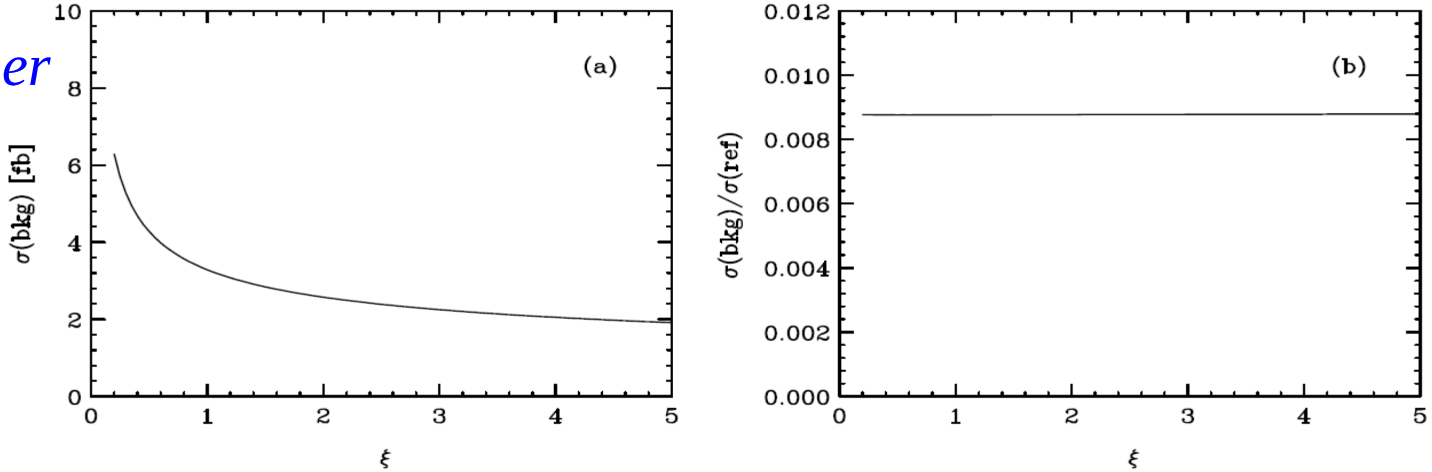


Comments on normalization relation

$$N^{estimated}_{signal\ \ reg.} = \frac{N^{Monte\,Carlo}_{signal\ \ reg.}}{N^{Monte\,Carlo}_{control\ \ reg.}} N^{measured}_{control\ \ reg.} = \frac{\sigma_{signal\ \ reg.}}{\sigma_{control\ \ reg.}} \frac{\epsilon_{signal\ \ reg.}}{\epsilon_{control\ \ reg.}} N^{measured}_{control\ \ reg.}$$

- $N_{signal}/N_{control}$ is calculated from the full simulation and it is what can be directly used.
- The second term is a rephrasing of the relation that points out the possible contribution of systematic uncertainties
- At parton level, the x-secs ratio is well under control from theoretical point of view

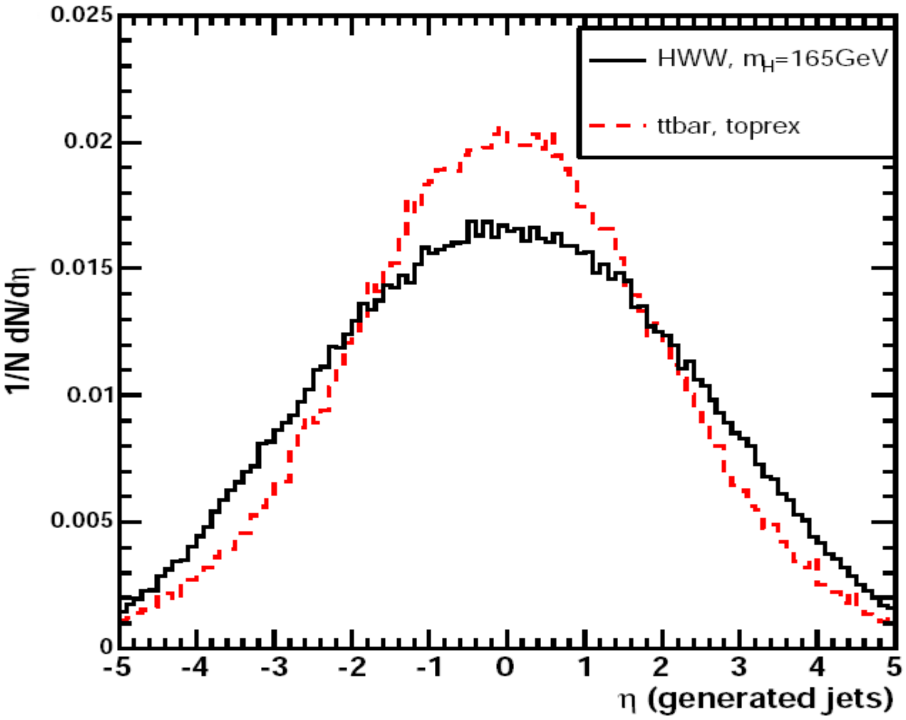
N. Kauer

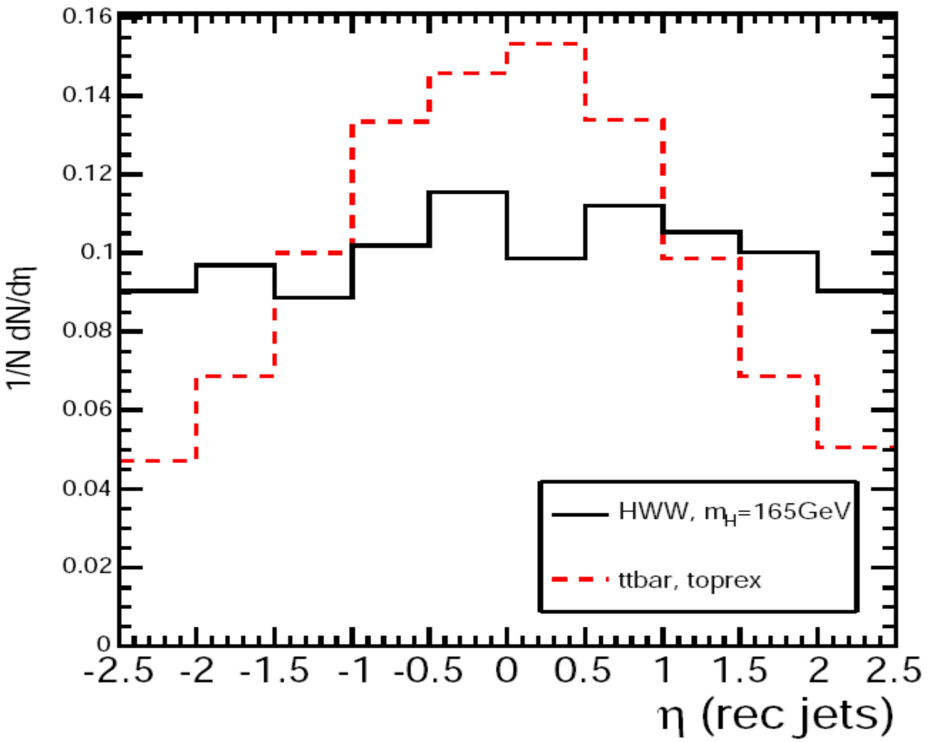


- The “ ” terms represent the experimental efficiencies in finding events in the signal/control regions.
- In this notation the “ ” terms accounts also for the smearing of the phase space regions bounds

Jet Veto

- Jets are reconstructed with an iterative cone algorithm with DR=0.5, considering towers formed by deposits in the Electromagnetic and Hadronic calorimeter. The threshold for the seed tower is Et = 1 GeV whereas for the constituent towers thresholds are set to E=0.8 and Et=0.5.
- A pseudorapidity cut is set at |η| =2.5 due to different distribution of jets in the signal and *tt*





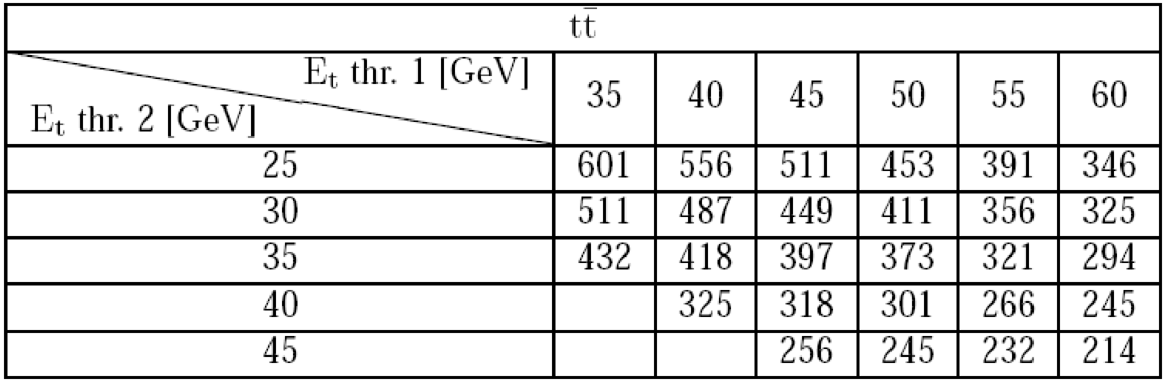
Double b-tag control region

Events for 10 fb^{-1} ($\sigma(tt)$, $\sigma(hWW, M_h=165)$ $\sigma(Wtb)$ at NLO)

Discriminator	2μ			$2e$			$e\mu$		
	$t\bar{t}$	Signal	Wt	$t\bar{t}$	Signal	Wt	$t\bar{t}$	Signal	Wt
1.5	218	1	2	128	< 1	< 1	294	2	4
1.75	211	1	2	118	< 1	< 1	266	1	3
2	194	< 1	1	107	< 1	< 1	245	1	2
2.25	183	< 1	1	86	< 1	< 1	232	1	2
2.5	173	< 1	1	69	< 1	< 1	218	< 1	1
2.75	166	< 1	1	62	< 1	< 1	211	< 1	1
3	152	< 1	< 1	59	< 1	< 1	194	< 1	1

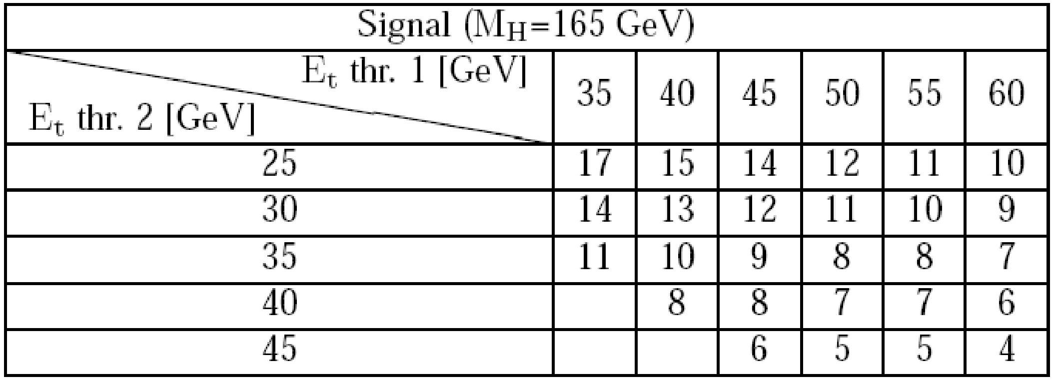
Two high Et jets control region

Events for 10 fb^{-1} ($\sigma(tt)$, $\sigma(hWW, M_h=165)$ $\sigma(Wtb)$ at NLO)



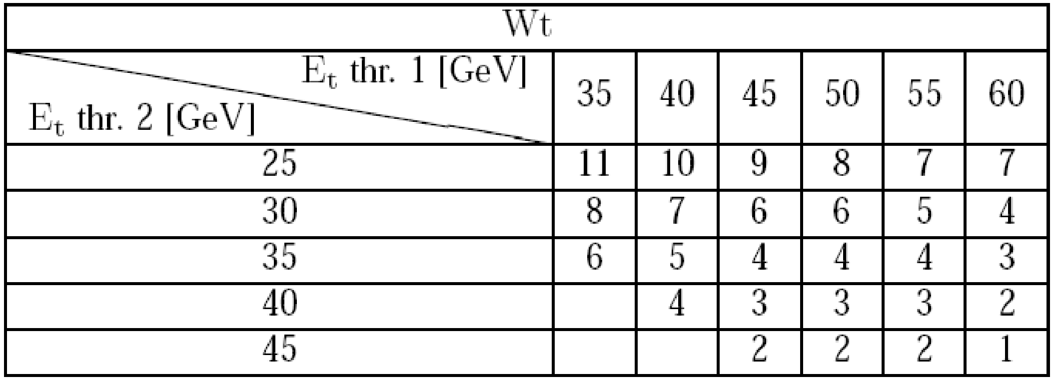
$t\bar{t}$

E_t thr. 1 [GeV] E_t thr. 2 [GeV]	35	40	45	50	55	60
25	601	556	511	453	391	346
30	511	487	449	411	356	325
35	432	418	397	373	321	294
40		325	318	301	266	245
45			256	245	232	214



Signal (M_H=165 GeV)

E_t thr. 1 [GeV] E_t thr. 2 [GeV]	35	40	45	50	55	60
25	17	15	14	12	11	10
30	14	13	12	11	10	9
35	11	10	9	8	8	7
40		8	8	7	7	6
45			6	5	5	4



Wt

E_t thr. 1 [GeV] E_t thr. 2 [GeV]	35	40	45	50	55	60
25	11	10	9	8	7	7
30	8	7	6	6	5	4
35	6	5	4	4	4	3
40		4	3	3	3	2
45			2	2	2	1

tt Normalization Procedures Summary

tt events for 10 fb^{-1}

	2μ	2e	eμ
Signal region	33 (±29%)	22 (±35%)	44 (±25%)
"b-tagging" control region	194 (±13%)	107 (±18%)	245 (±12%)
"hard jets" control region	-	-	411 (±9%)

tt, *Wt* and signal events for 10 fb^{-1}

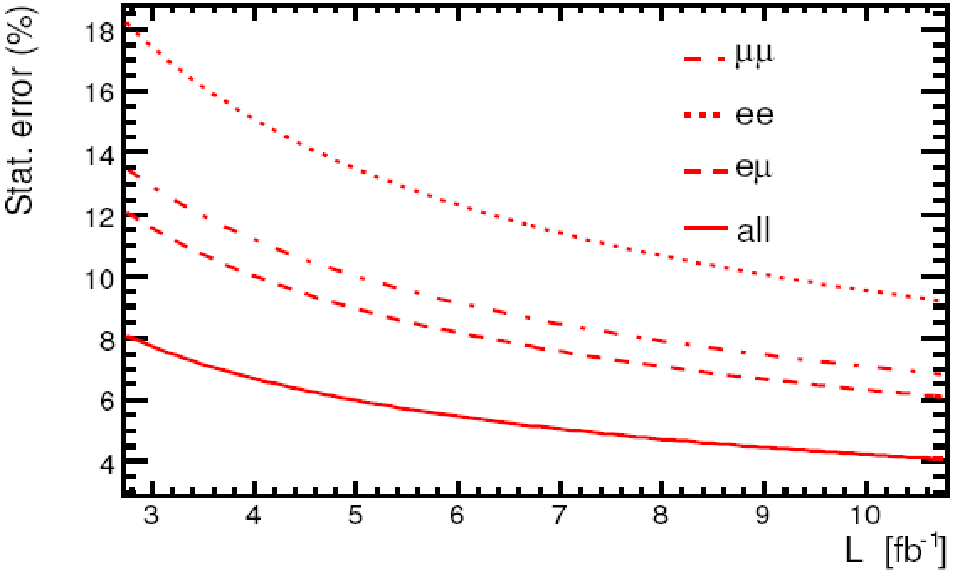
	"b-tagging" control region			"hard jets" control region			Signal region		
	2μ	$2e$	$e\mu$	2μ	$2e$	$e\mu$	2μ	$2e$	$e\mu$
$t\bar{t}$	194	107	245	-	-	411	33	22	44
Wt	1	< 1	2	-	-	6	5	3	6
$Signal\ (m_H = 165)$	< 1	< 1	1	-	-	11	156	89	214

Theoretical uncertainty on the ratio $\sigma_{\text{signal}}/\sigma_{\text{control}}$

- The dependency of the ratio $\sigma_{\text{control reg}}/\sigma_{\text{signal reg}}$ on the factorization and renormalization scale has been studied at LO at parton level by N. Kauer and it has been shown to be very small (~1%)
- In that work, the PDF choice is foreseen to accounts for an error up to ~10%
- NLO does not affect much the variables involved in the normalization procedure
- Different showering models however differ either in the jets multiplicity and in the jets spectra. How much the ratio $\sigma_{\text{control reg}}/\sigma_{\text{signal reg}}$ is affected has not been studied

tt Normalization Procedures Statistical Uncertainty

2 b-tagged jets Control Region



2 hard jets Control Region