Probing the heavy flavor content in ttbar events

&

Using ttbar events as a calibration tool

at

Roberta Volpe
Università degli Studi di Perugia
and INFN Perugia
on behalf of the CMS collaboration

DPF 2009

R measurement motivation

$$R = \frac{B(t \to Wb)}{B(t \to Wq)}$$ q = b,s,d

Standard Model (SM) predicts:

$$R = \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2}$$

- If 3 SM quark generations: Unitarity 3x3 CKM matrix => $R=|V_{tb}|^2$

R measurement → **V_{tb} measurement**

- If a 4th generation of quarks (t' and b') exists →unitarity of a 4x4 matrix

  - R measurement → **constraints on V_{tb}**
  - $R_{SM} - R > O(10^{-1})$ → **clue of the existence of a 4th generation**

J. Alwall et al., "Is V(tb) = 1?," *Eur. Phys. J.* **C49** (2007) 791–801

What we know about R and V_{tb}

- **From Standard Model :** $|V_{tb}| = 0.999133^{+0.000044}_{-0.000043}$

- **Last R measurements:**

CDF: $R > 0.61$ at 95 % C.L. with L = 162 pb^{-1}

D0: $R = 0.97^{+0.09}_{-0.08}$ $R > 0.79$ at 95 % C.L.

Simultaneus measurement of R and ttbar cross section with L = 900 pb^{-1}

- **V_{tb} Measurements from single top studies:**

CDF: $|V_{tb}| = 0.91 \pm 0.11(\text{stat} + \text{syst}) \pm 0.07(\text{theory})$

$|V_{tb}| > 0.71$ at the 95% C.L. with L = 3.2 fb^{-1}

D0: $|V_{tb}| > 0.78$ at 95% C.L. with L = 2.3 fb^{-1}

Plans for R measurement with CMS experiment

Physical Observable: distribution of number of b-tagged jets
b-tagging algorithm used relies on the track impact parameter

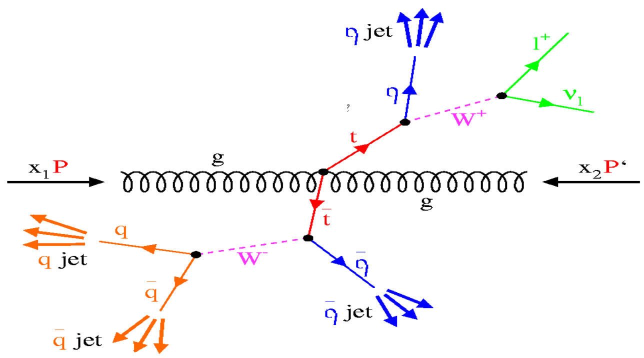
ϵ_b = b-tagging efficiency
ϵ_q = mis-tagging probability

Probability to find i b-tagged jets:

$$P_i(R;\epsilon_b,\epsilon_q) = R^2 P_i(tt \to WWbb) + 2R(1-R)P_i(tt \to WWbq) + (1-R)^2 P_i(tt \to WWqq)$$

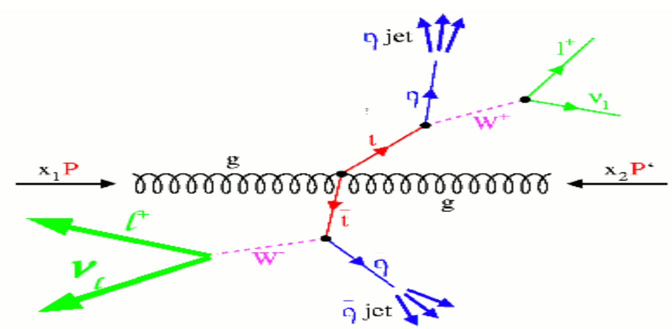
with q = s,d and i = number of b-tagged jets

- **Semileptonic channel (L= 1 fb^{-1})**
  - **R measurement** ***CMS PAS TOP-09-007***



- **Dileptonic channel e μ (L=250 pb^{-1})**
  - **R measurement**
  - b-tagging efficiency measurement (by fixing R)

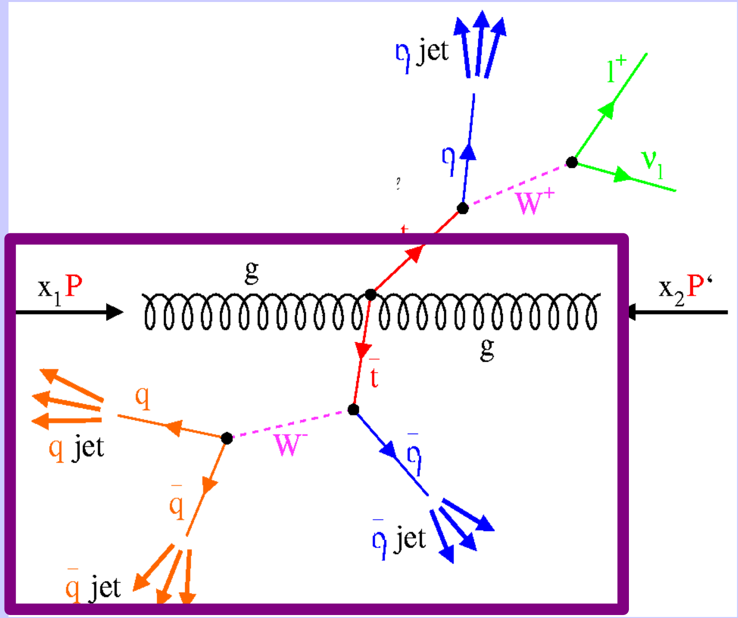
CMS PAS TOP-09-001



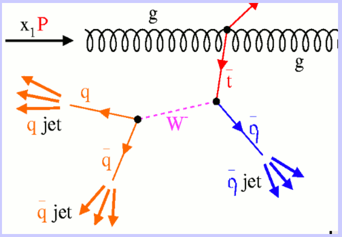
Semileptonic channel

L= 1 fb^{-1} @ 10 TeV

- R *measurement*



CMS PAS TOP-09-007



Selection and event reconstruction

1. Trigger request;
2. a single isolated and high energy lepton (electron or muon);
3. at least four selected jets;
4. $Centrality > 0.35$;
5. $|m_{ij} - m_{Whad}| < \sigma(m_{Whad})$;
6. $\chi^2 < 4$

	$\sigma \cdot BR$ (pb)	N_{ev}
$t\bar{t}$ Semilep	182	2650
other $t\bar{t}$	232	109
W + jets	$4 \cdot 10^4$	260
Z + jets	$3.7 \cdot 10^3$	52
tW	29	52
QCD	10^9	56

L=1 fb^{-1}

N(Semilept ttbar) / N(Bkg) ~ 5

- **Centrality:**

$$\frac{\sum E_T}{\sqrt{\left(\sum E\right)^2 - \left(\sum p_z\right)^2}}$$

Sums on every jet

- **Hadronic W reconstruction:**

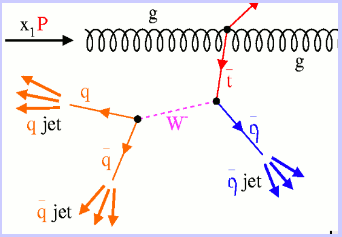
4 good jets with largest Et: Invariant mass Mij for each pair of jets, choose (i, j) in order to have minimum $\Delta M_W = |M_{ij} - M_{Whad}|$

- **χ^2 based on tops masses:**

2 jets not associated to Whad: k,p

$$\chi^2 = \left(\frac{m_{ijk} - m_{tHad}}{\sigma(m_{tHad})}\right)^2 + \left(\frac{m_{l\nu p} - m_{tLep}}{\sigma(m_{tLep})}\right)^2$$

choose k and p in order to have the minimum χ^2

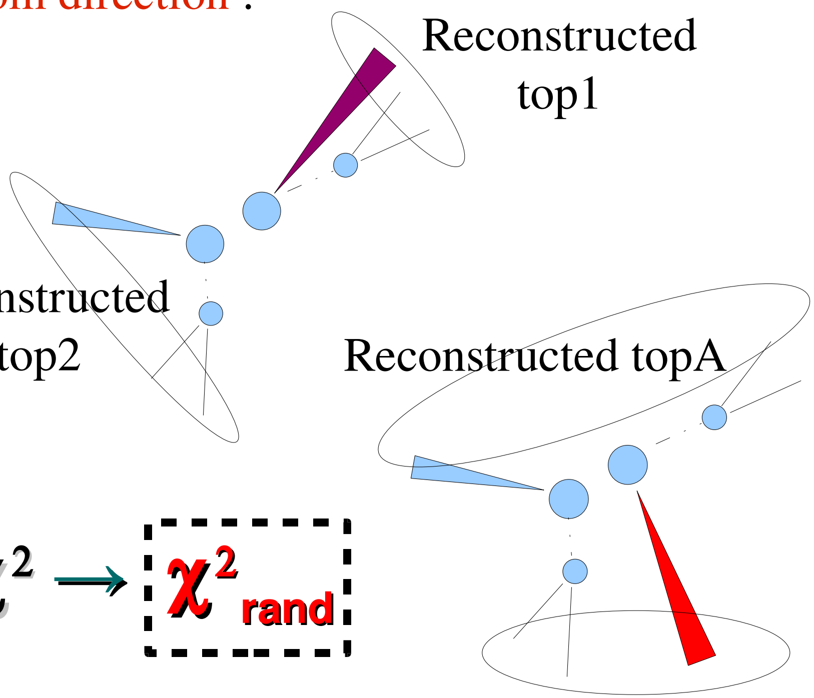


Background Subtraction

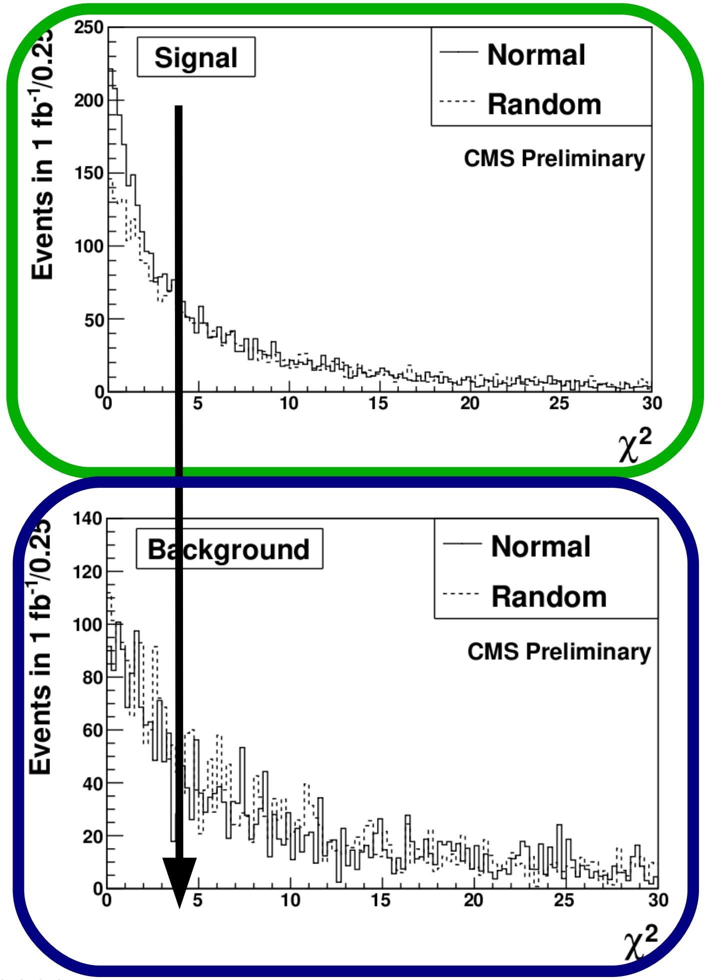
Signal: semilept ttbar events with 2 jets well matched to partons coming from the direct decays of top

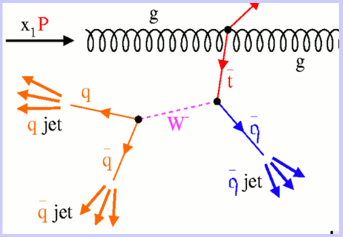
Background: semilep ttbar other than Signal ; dilep and hadr ttbar, W+jets, Z+jets, single top, QCD.

Change the momentum direction of the jet with larger Et coming from the top decay to a random direction :



$\chi^2 \rightarrow \chi^2_{rand}$

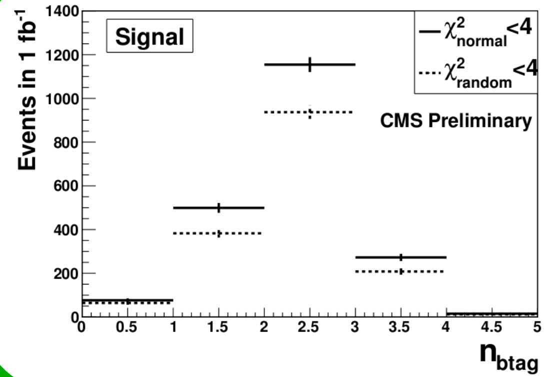


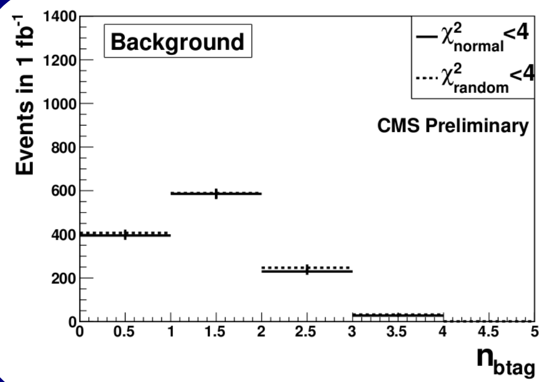


Background Subtraction

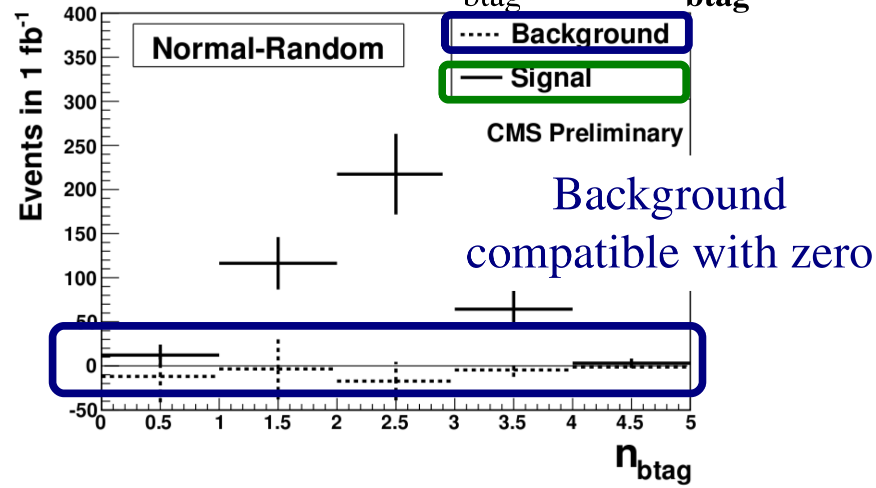
$\chi^2_{norm} < 4$
$\rightarrow n_{btag}^{norm}$

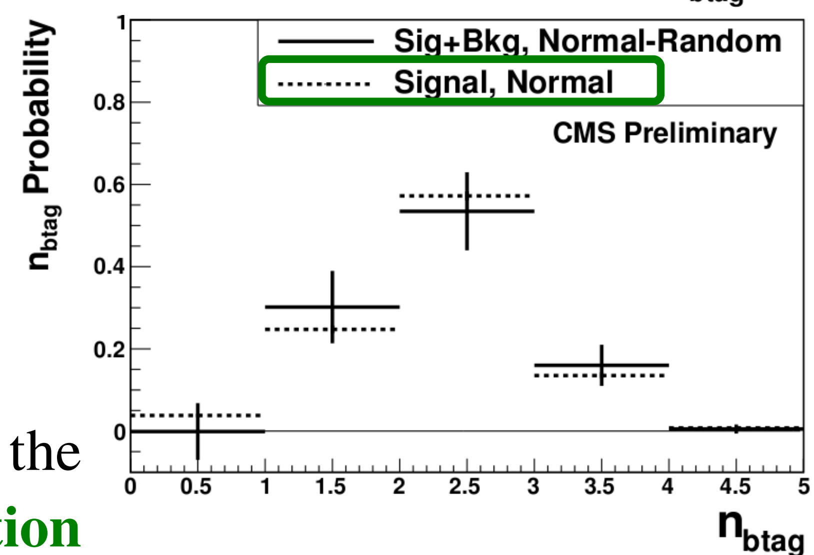
$\chi^2_{rand} < 4$
$\rightarrow n_{btag}^{rand}$



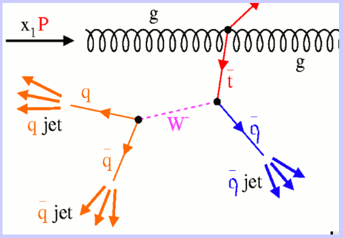


Bin by bin subtraction $n_{btag}^{norm} - n_{btag}^{rand}$





The resulting distribution is proportional to the **signal n_{btag} distribution**



R measurement method

$n_{btag}^{norm} - n_{btag}^{rand}$ (all data) is to be fitted with:

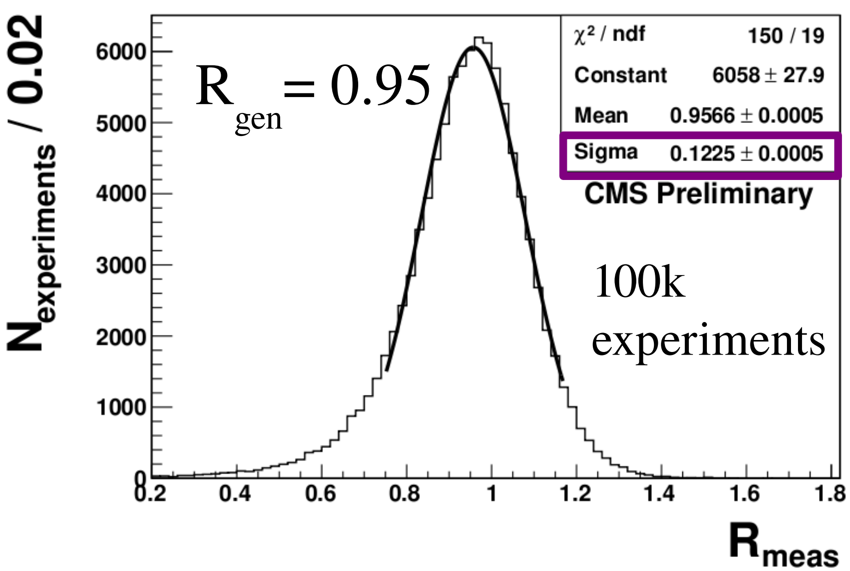
$$P_i(R;\epsilon_b,\epsilon_q) = R^2 P_i(bb) + 2R(1-R)P_i(bq) + (1-R)^2 P_i(qq)$$

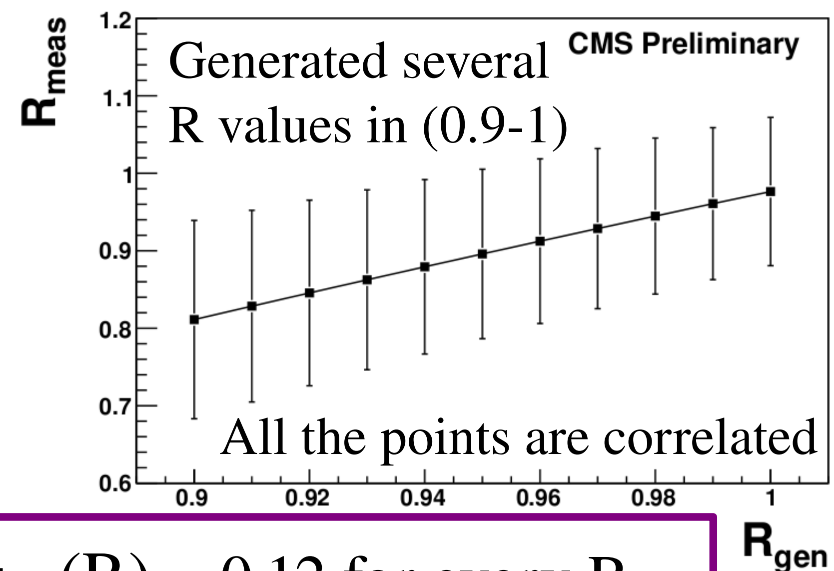
$P_i(\ldots)=f(\epsilon_b,\epsilon_q)$ R and ϵ_b are *fully correlated* $\rightarrow$ ϵ_b **is fixed**

Free parameters: R and ϵ_q

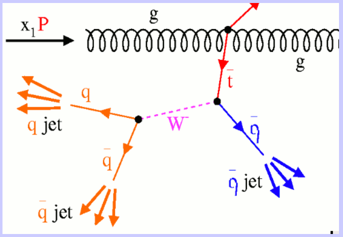
CMS will use methods to evaluate ϵ_b from data with an independent sample

Fit Validation with a Toy Monte Carlo experiment:





$\sigma_{stat}(R) \sim 0.12$ for every R_{gen}



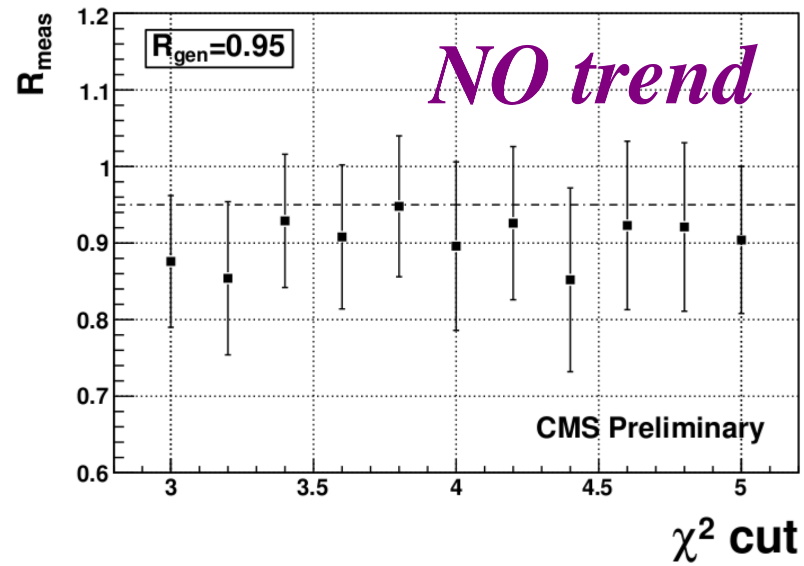
Expected uncertainties

Systematics evaluation:

systematics	σ_{sys}
b tagging efficiency	0.04
b tagging efficiency bias	0.04
Jet Energy Scale	0.09
χ^2 cut	0.02
Selection efficiency	0.006
total	0.11

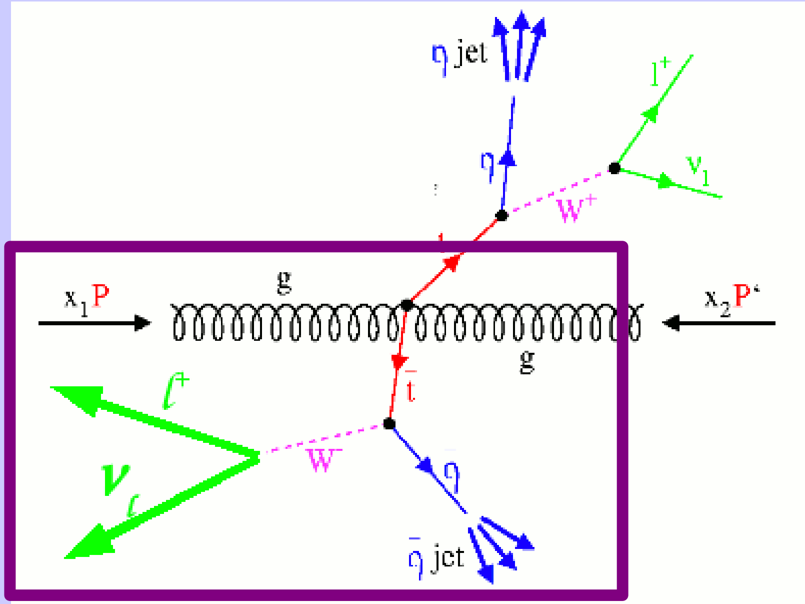
The main systematics sources (JES and ϵ_b) will be reduced once the detector is well understood

Stability under changes of χ^2 cut:



- σ_R (stat) = 0.12
- σ_R (sys) = 0.11

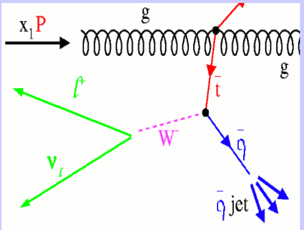
Dileptonic channel



L= 250 pb^{-1} @ 10 TeV

- R *measurement*
- ϵ_b *measurement*

CMS PAS TOP-09-001



Analysis strategy

The nbtag distribution depends mainly on:

- ϵ_b and ϵ_q
- R
- The probability to reconstruct and select jets from top decays: ***α***

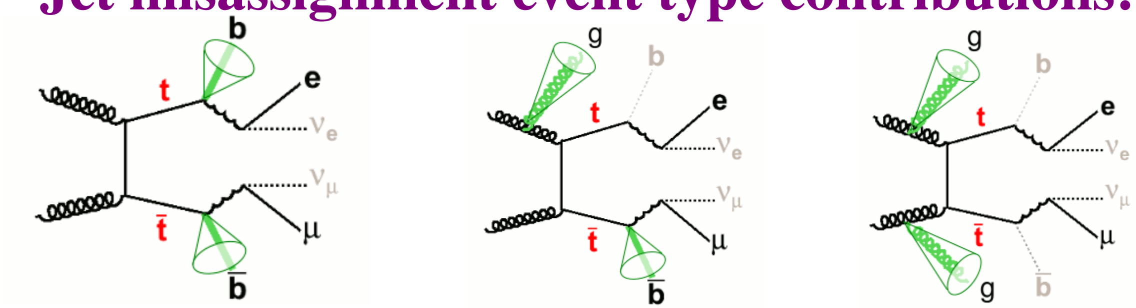
Selection
Triggered
$\geq$ 2 leptons (>20 GeV/c)
1 e and 1 μ
$\geq$ 2 jets (>30 GeV)
$\not\!\!E_T \geq$ 30 GeV
Opp. sign leptons

	$\sigma \cdot BR\ (pb)$	N_{ev}
$t\bar{t}$ Dilep	41	787
Single top	144	29
other $t\bar{t}$	372	14
Di-boson	40	11
$W/Z+$ jets	$4.4 \cdot 10^4$	27
$V+QQ$	290	1
QCD	10^9	0

L=250 pb^{-1}

S/B~10

Jet misassignment event type contributions:



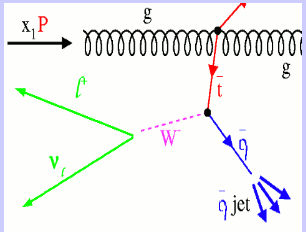
$\alpha_2 = \alpha^2$ $\alpha_1 = 2(1-\alpha)\alpha^2$ $\alpha_0 = (1-\alpha)^2$

Distribution considered in order to evaluate α : $\mathbf{M_{lb}}$

l and b lepton and jet coming from the same top

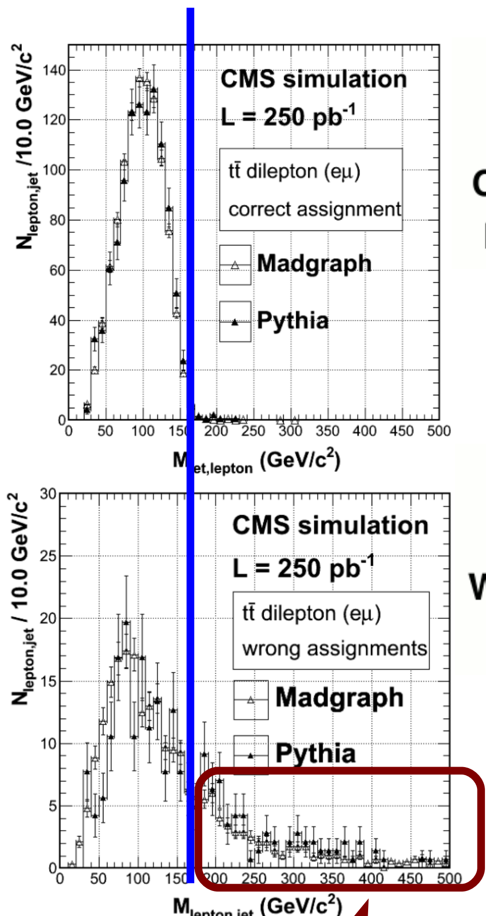
2 methods to estimate from data each event type contribution:

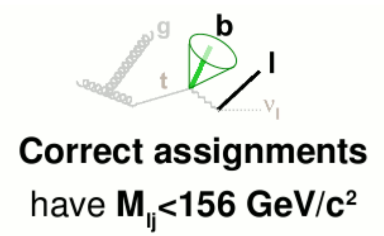
- Swap: pair all leptons from one event with all jets from a different event
- Random rotation: randomize the direction of all leptons before pairing with all jets

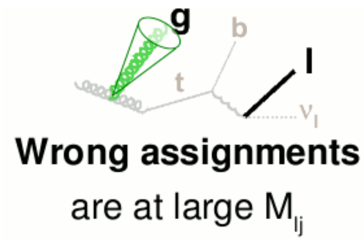


Jet misassignment estimation from data

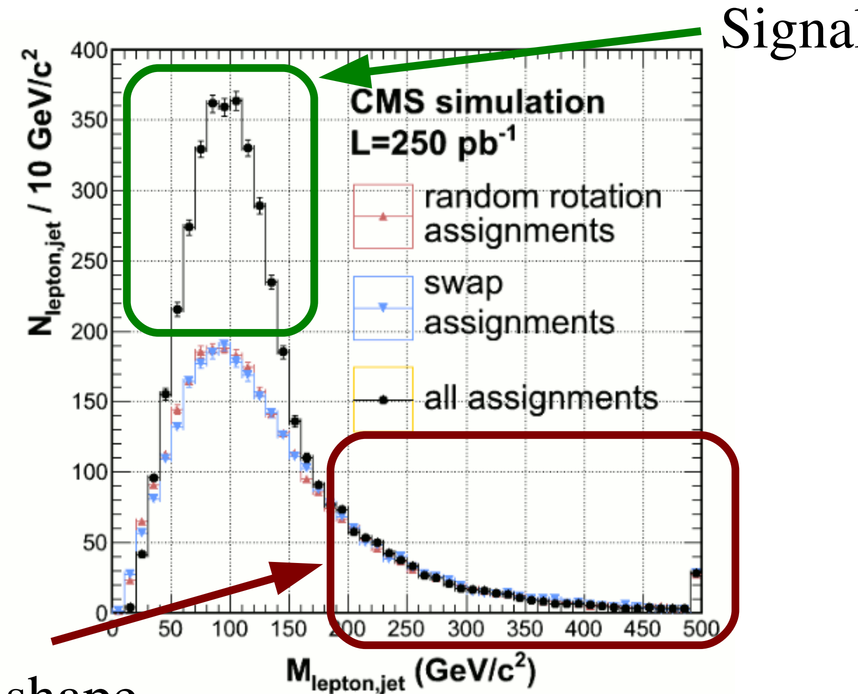
$$M_{l,b}^{max} \equiv \sqrt{m_t^2 - m_W^2} = 156\ \mathrm{GeV/c^2}$$







Only wrong assignment

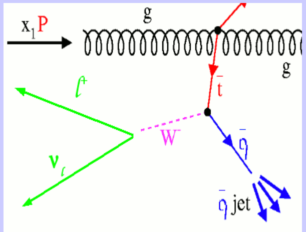


Scale the shape to match the data for $M_{lep,jet} > 190\ \mathrm{GeV/c^2}$

From the normalization factor **find** α

The everage of the swap and random results is used

Method	$N_{mis}^{M>190} / N_{mis}$	α
$t\bar{t}$ events from MADGRAPH		
average	0.21 ± 0.01	0.82 ± 0.04
MC truth	0.20 ± 0.01	0.80 ± 0.01



Fit strategy

$$P_i(R;\epsilon_b,\epsilon_q) = R^2 P_i(bb) + 2R(1-R)P_i(bq) + (1-R)^2 P_i(qq)$$

α from data → α_2

$P_i(\ldots)=f(\epsilon_b,\epsilon_q,\alpha_0,\alpha_1,\alpha_2)$

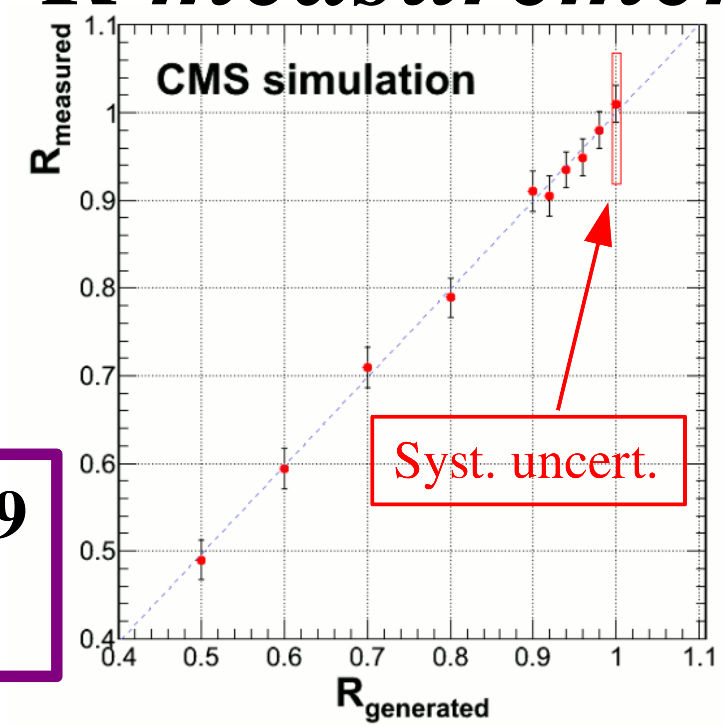
α_1 from normalization

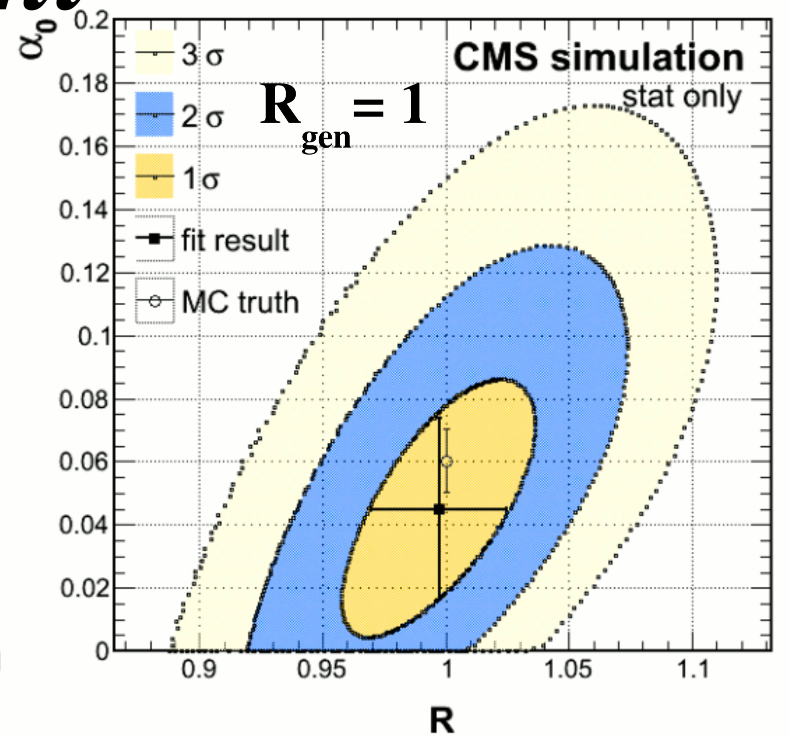
B-tagging working point slightly different from semileptonic analysis

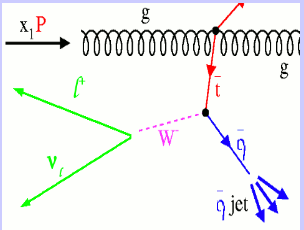
R measurement

ϵ_b,ϵ_q fixed
(from other measurements)
free parameters:
R and **α_0**

$\sigma_{stat+sys}$(R) = 0.09
for L = 250 pb^{-1}

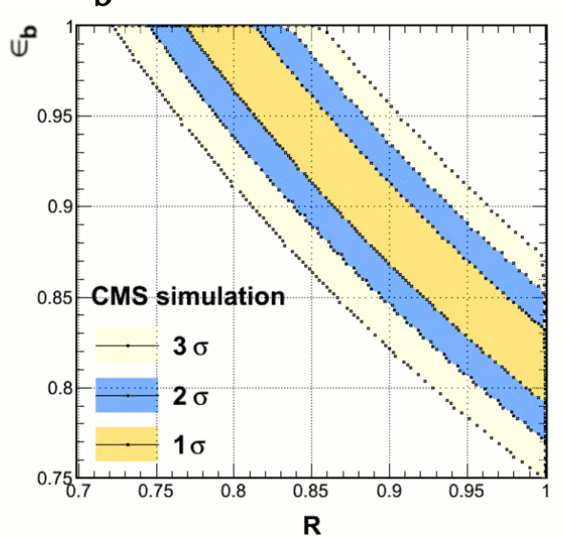






Fit strategy

ϵ_b and **R** are fully correlated:

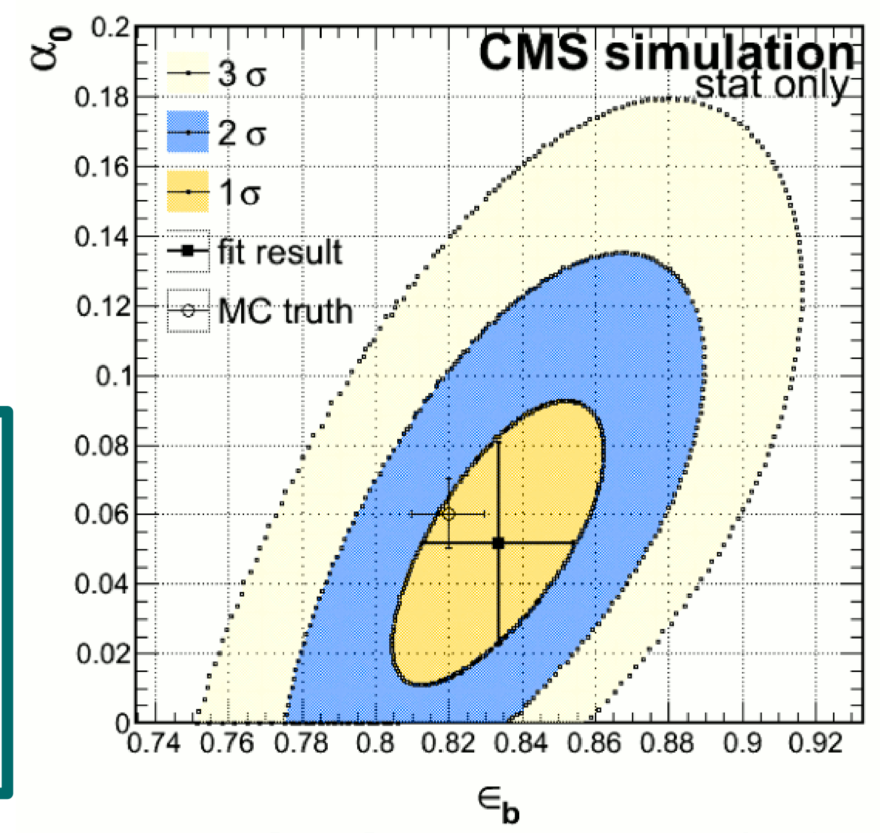


R fixed from Standard Model

$0.79 < \epsilon_b < 0.83$

in agreement with MC

ϵ_b measurement

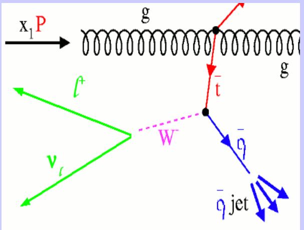


R fixed from SM

ϵ_q fixed from other source

free parameters: α_0 and ϵ_b

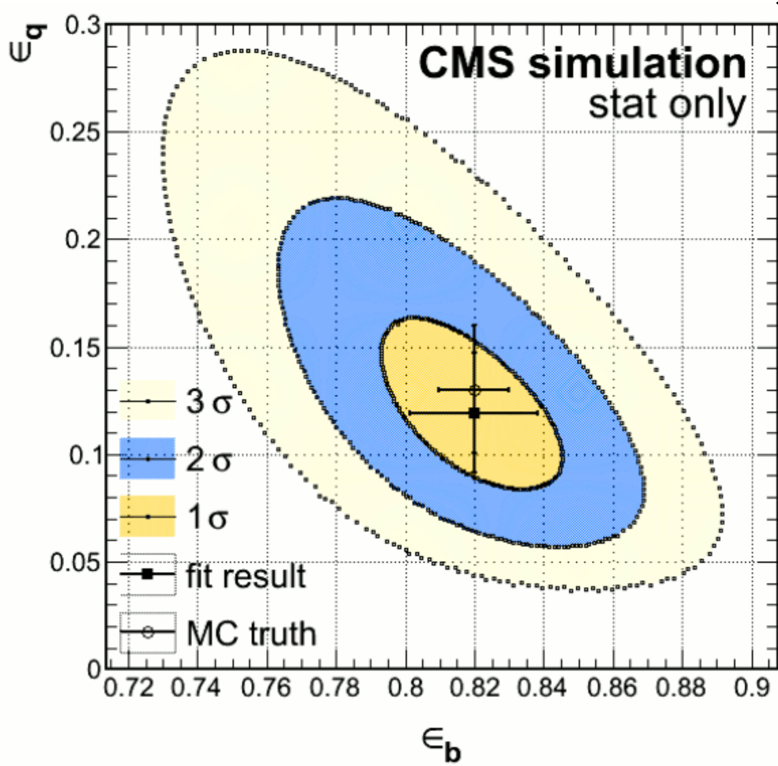
(constraint $0 < \epsilon_b < 1$)



Fit strategy: ϵ_b *measurement*

2 different b-tagging algorithms, each for 3 different working points

algorithm	working point	ε_b (MC truth)	ε_b
Jet Probability	loose	0.82 ± 0.01	0.81 ± 0.02
	medium	0.63 ± 0.01	0.63 ± 0.02
	tight	0.41 ± 0.01	0.41 ± 0.02
Track Counting	loose	0.80 ± 0.01	0.82 ± 0.02
	medium	0.65 ± 0.01	0.65 ± 0.02
	tight	0.40 ± 0.01	0.41 ± 0.02



Closure test:
fitting ϵ_b, ϵ_q

$\epsilon_b = 0.82 \pm 0.02$

$\epsilon_q = 0.12 \pm 0.03$

Syst. Uncert. ~ 4 %
for $L = 250\ pb^{-1}$

Conclusions

- **R measurement strategies (data-driven background evaluation) have been developed in both semileptonic and dileptonic ttbar events**

Semilep (L = 1 fb^{-1}) $\sigma_{stat}(R) = 0.12$ $\sigma_{sys}(R) = 0.11$

Dilep (L = 250 pb^{-1}) $\sigma_{stat+sys}(R) = 0.09$

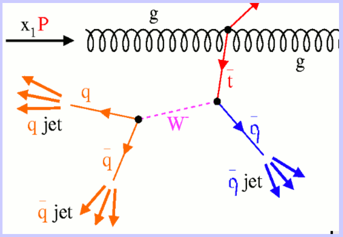
- **B-tagging efficiency measurement from data using dileptonic ttbar events is possible already with early LHC data**

$\sigma_{stat}(\epsilon_b) = 0.02$

$\sigma_{sys}(\epsilon_b) \sim 0.04$

Dilep (L = 250 pb^{-1})

Back-up



Selection (semil)

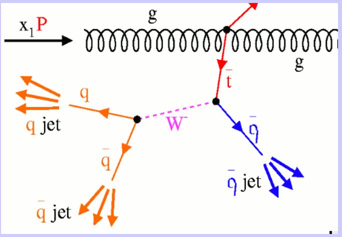
Physics Objects Reconstruction

- **HLT:** single lepton trigger (single electron || single muon)
- **Good Jets:**
  - $|\eta| < 2.4,\ E_T^{corr} > 40$ GeV
  - $\Delta R(Jet, Lep) > 0.5,\ EmFr < 1$
- **Missing Transverse Energy:**
  ·Vectorial sum of calorimetric energy
  ·deposits (JES and muon energy
  · deposits considered)
- **Good Lepton (e or μ) :**
  lepton ID, Isolation cut, p_t(lep) > 30 GeV,$|\eta| < 2.4$

Preliminary selection:
- ✓ HLT
- ✓ one single lepton
- ✓ at least 4 jets

	$\varepsilon(HLT)$ %	$\varepsilon(Lep)$ %	$\varepsilon(Jets)$ %	$\varepsilon(Centr)$ %	$\varepsilon(\Delta M_W)$ %	$\varepsilon(\chi^2)$ %	N_{evts} with 1 fb^{-1}
$t\bar{t}$ semil	54	27.5	7.23	5.62	3.79	1.48	2650 ←
$t\bar{t}$ others	22	7.9	$6.3\cdot10^{-1}$	$4.9\cdot10^{-1}$	$2.4\cdot10^{-1}$	$4.7\cdot10^{-2}$	109
$W + jets$	34	16	$9.5\cdot10^{-3}$	$6.8\cdot10^{-3}$	$3.7\cdot10^{-3}$	$6.5\cdot10^{-4}$	260
$Z + jets$	50	19	$1.9\cdot10^{-2}$	$1.4\cdot10^{-2}$	$7.4\cdot10^{-3}$	$1.4\cdot10^{-3}$	52
tW	35	18	1.2	$8.4\cdot10^{-1}$	$6.4\cdot10^{-1}$	$1.8\cdot10^{-1}$	52
QCD	2.7	$7.8\cdot10^{-3}$	$5.2\cdot10^{-6}$	$4.3\cdot10^{-6}$	$2.4\cdot10^{-6}$	$1.1\cdot10^{-6}$	56

 # selected well reconstructed Signal events = 2650 * 0.74 (well rec. fr.) = 1960



Fitting function (semil)

The nbtag ^{norm} - nbtag ^{rand} (sig+bkg) distribution is to be fitted with:

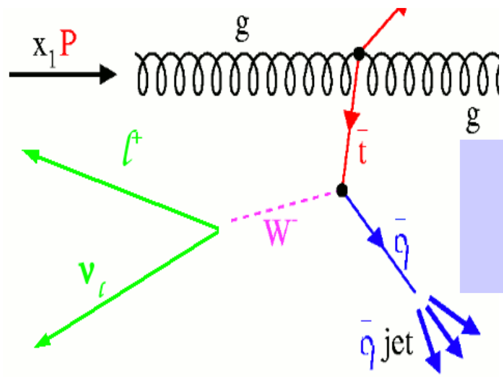
$B=\epsilon_b$, $M=\epsilon_q$

$$\epsilon_i(R;B,M) = R^2 P_i(t\bar{t} \to bWbW) + 2R(1-R)P_i(t\bar{t} \to bWqW) + (1-R)^2 P_i(t\bar{t} \to qWqW)$$

With Pi( ) from:

	P_0	P_1	P_2	P_3	P_4
bb	$(1-B)^2(1-M)^2$	$2B(1-B)(1-M)^2+$ $2M(1-B)^2(1-M)$	$B^2(1-M)^2+$ $4BM(1-M)(1-B)+$ $M^2(1-B)^2$	$2B^2M(1-M)+$ $2M^2B(1-B)$	B^2M^2
bq	$(1-B)(1-M)^3$	$B(1-M)^3$ $(1-B)3M(1-M)^2$	$3BM(1-M)^2+$ $3M^2(1-B)(1-M)$	$M^3(1-B)+$ $3M^2B(1-M)$	BM^3
qq	$(1-M)^4$	$4M(1-M)^3$	$6M^2(1-M)^2$	$4M^3(1-M)$	M^4

Fixed parameter: B Free parameters: R and M



dilep.

Trigger, Reconstruction, Selection

- The **dilepton channel** (eμ) **event selection** is used (cf. TOP-09-002)
  - **2 hard-p_T**, prompt, **isolated leptons**

Trigger

Single electron or muon

Lepton Selection

p_T > 20 GeV/c |η| < 2.4
| d_0 | < 400 mm I_{trk}>0.9 I_{calo}>0.9 (0.8 for e)
Δ(e,μ) > 0.1
Lepton id.

Jet Selection

Calorimeter Jets (SIS cone ΔR<0.5)
2 constituents with E_T>2GeV+1 track
p_T^{corr}>30 GeV/c |η|<2.4
emf<0.98   **ΔR(jet, e/μ)>0.3**

Missing Et >30 GeV (corrected for μ)

nalysis Approval 5

Event selection: results

Event yields for L = 250 pb^{-1}

Table 1: Expected event yield for an integrated luminosity of L=250 pb^{-1} using a MADGRAPH sample. Only statistical uncertainties from the MC samples are shown.

Selection	Total	$t\bar{t}$ dileptons
Triggered	$(426 \pm 1) \cdot 10^6$	6251 ± 25
$\geq$ 2 leptons (>20 GeV/c)	$(204.7 \pm 0.5) \cdot 10^3$	2595 ± 16
1 e and 1 μ	2531 ± 32	1344 ± 12
$\geq$ 2 jets (>30 GeV/c)	1041 ± 12	914 ± 10
$\not{E}_T \geq 30$ GeV	884 ± 10	789 ± 9
Opp. sign leptons	867 ± 10	787 ± 9

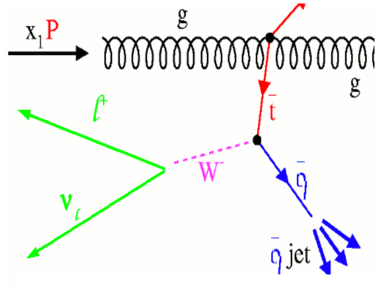
Background processes only

Table 2: Expected event yield for background processes for an integrated luminosity of L=250 pb^{-1}. Only statistical uncertainties are shown. VQQ corresponds to the background from $W/Zb\bar{b}$ and $W/Zc\bar{c}$. All samples were generated with MADGRAPH except for QCD which was generated using PYTHIA.

Selection	Single top	other $t\bar{t}$	Di-boson	W/Z+jets	VQQ	QCD
Triggered	27563 ± 422	42303 ± 77	3400 ± 13	$(4338 \pm 2) \cdot 10^3$	38046 ± 64	$(422 \pm 1) \cdot 10^6$
$\geq$ 2 leptons (>20 GeV/c)	225 ± 10	66 ± 3	620 ± 5	$(193 \pm 0.4) \cdot 10^3$	8170 ± 26	401 ± 183
1 e and 1 μ	122 ± 8	34 ± 2	207 ± 2	749 ± 24	49 ± 2	26 ± 15
$\geq$ 2 jets (>30 GeV/c)	38 ± 4	27 ± 2	14 ± 1	45 ± 6	2.0 ± 0.4	-
$\not{E}_T \geq 30$ GeV	29 ± 1	23 ± 2	11.6 ± 0.5	30 ± 5	0.7 ± 0.2	-
Opp. sign leptons	29 ± 1	14 ± 1	10.5 ± 0.5	26 ± 5	0.7 ± 0.2	-

dilep.

Misassignment models in background samples



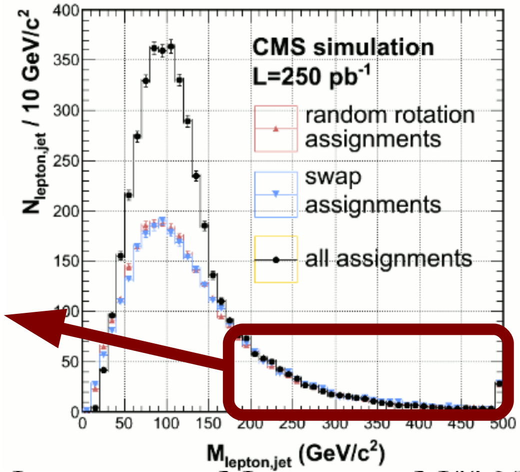
$$1-\alpha = \frac{N_{mis}}{N} = 1 - \frac{N - N_{mis}}{2N_{evts}}$$

$N_s = N_{swap/random}$

$$N^{M>z}_{mis} = N^{M>z}$$

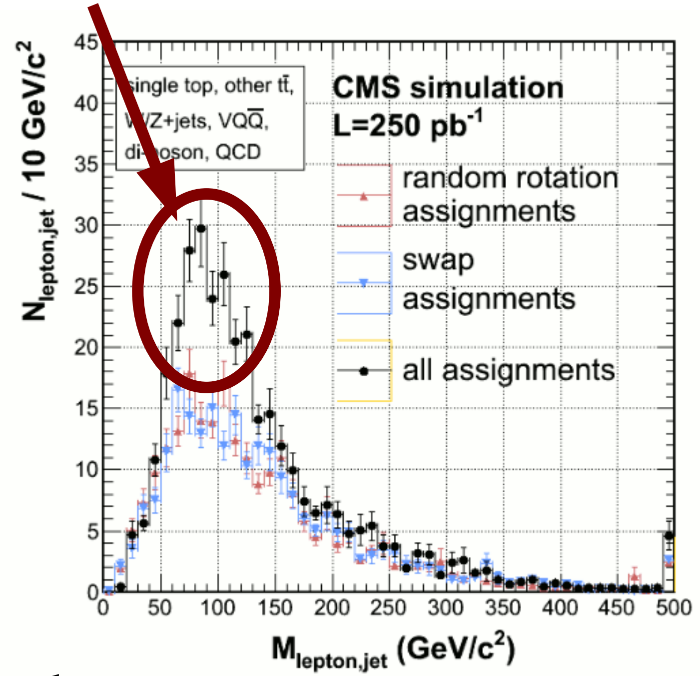
$$\frac{N^{M>z}_{mis}}{N_{mis}} = \frac{N^{M>z}_{s}}{N}$$

All data



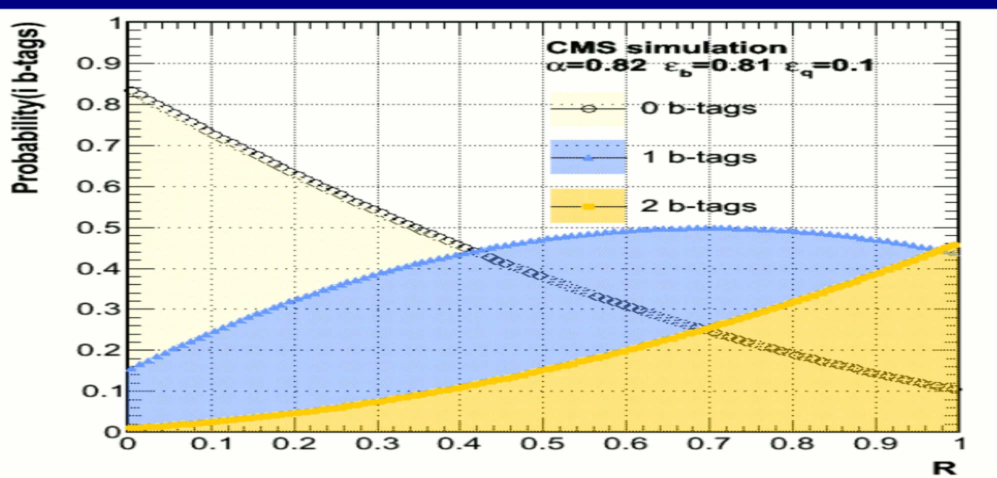
Contributions with one top

Background



$$\alpha = \frac{N - N_{mis}}{2\ N_{evts}} = \frac{N}{2\ N_{evts}}\left(1 - \frac{N_{mis}}{N}\right) = \frac{N}{2\ N_{evts}}\left(1 - \frac{N^{M>z}}{N^{M>z}_{s}}\right)$$

Probabilistic model for the b-tagging multiplicity



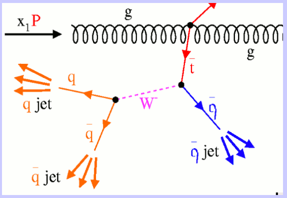
example:

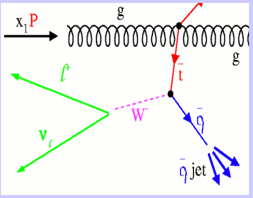
Expect 2 *b*-tags in $t\ \bar{t}$ events from:

P(t→Wb; t→Wb)	∝	$R^2\ \varepsilon_b^{\ 2}$
+ P(t→Wb; t→Wq)	∝	$2\ R\ (1\text{-}R)\ \varepsilon_b\,\varepsilon_q$
+ P(t→Wq ;t→Wq)	∝	$(1\text{-}R)^2\ \varepsilon_q^{\ 2}$

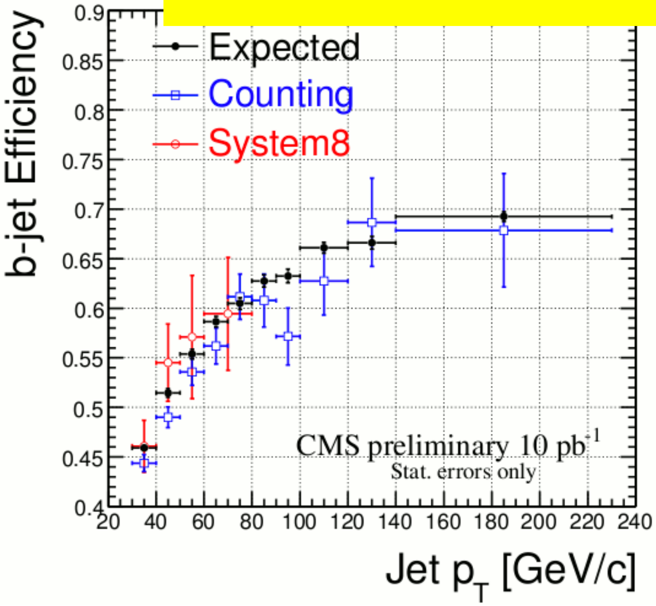
(where q= d,s)

For R measurement, ϵ_b will be fixed to the value extracted from an independent sample: di-jet b bbar events





CMS PAS BTV-07-001



System8:

$$
\begin{aligned}
n &= n_b + n_{cl} \\
p &= p_b + p_{cl} \\
n^{tag} &= \epsilon_b^{tag} n_b + \epsilon_{cl}^{tag} n_{cl} \\
p^{tag} &= \beta \epsilon_b^{tag} p_b + \alpha \epsilon_{cl}^{tag} p_{cl} \\
n^{\mu} &= \epsilon_b^{\mu} n_b + \epsilon_{cl}^{\mu} n_{cl} \\
p^{\mu} &= \epsilon_b^{\mu} p_b + \epsilon_{cl}^{\mu} p_{cl} \\
n^{tag,\mu} &= \kappa_b \epsilon_b^{tag} \epsilon_b^{\mu} n_b + \kappa_{cl} \epsilon_{cl}^{tag} \epsilon_{cl}^{\mu} n_{cl} \\
p^{tag,\mu} &= \kappa_b \beta \epsilon_b^{tag} \epsilon_b^{\mu} p_b + \kappa_{cl} \alpha \epsilon_{cl}^{tag} \epsilon_{cl}^{\mu} p_{cl}
\end{aligned}
$$

From data

From MC

$$\kappa_b = \frac{\epsilon_b^{tag,\mu}}{\epsilon_b^{tag} \epsilon_b^{\mu}} \qquad \kappa_{cl} = \frac{\epsilon_{cl}^{tag,\mu}}{\epsilon_{cl}^{tag} \epsilon_{cl}^{\mu}}$$

$$\beta = \frac{\epsilon_{b,p}^{tag}}{\epsilon_{b,n}^{tag}} \qquad \alpha = \frac{\epsilon_{cl,p}^{tag}}{\epsilon_{cl,n}^{tag}}$$

8 unknowns:

n_b, n_{cl}, p_b, p_{cl}

$\varepsilon_b^{tag}, \varepsilon_b^{mu}, \varepsilon_{cl}^{tag}, \varepsilon_{cl}^{mu}$

operating point	Loose		
Luminosity (pb^{-1})	10	100	1000
Systematics (%)			
β	5.8	5.8	2.9
α	0.4	0.4	0.2
κ_b	3.4	3.4	1.7
κ_{cl}	0.2	0.2	0.1
p_{Trel}	2.8	2.8	2.8
statistics MC (%)	2.3	2.3	2.3
statistics data (%)	7.2	2.3	0.7
Total error (%)	10.5	8.0	6.4

Counting:

Method relying on fit to the pTrel distribution

operating point	Loose		
Luminosity (pb^{-1})	10	100	1000
Counting			
statistics data (%)	2.5	0.9	0.3
Template (%)	15	10	5
Mistag (%)	3.1	3.1	3.1
Total error (%)	16	11	6