LHC Workshop Berkeley

Workshop on Physics Opportunities with the First LHC Data

6-8 May 2009, Berkeley, CA, USA

Top quark pair production in di-leptonic decay channels at CMS

- Introduction:
  - Top Quark Production
  - CMS and Commissioning
- Observing Early Dileptonic tT
- Analyses @ 14 TeV:
  - $10pb^{-1}$
  - $100pb^{-1}$
  - eτ/μτ $100pb^{-1}$
- Perspectives for 2009

Julien Caudron
(julien.caudron@cern.ch)

University of Louvain
Center for Particle Physics and Phenomenology

Top Quark

If the top quark is not observed at LHC , we cannot claim observing new processes

History:

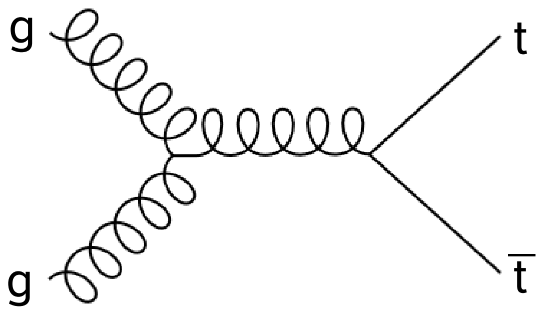
- in top anti-top pair production (tT) (cf. next slide)
  observed at Tevatron in 1995
- in single top production (t-channel, s-channel, tW-channel)
  observed at Tevatron in 2009 (arXiv:0903.0850, arXiv:0903.0885)

Motivations:

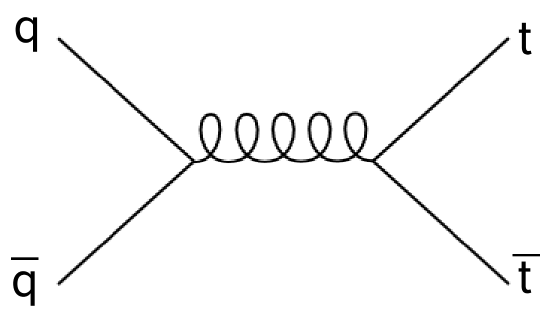
- It will be the first complex process (leptons, b-jets, MET) at LHC
  → ▪ physics commissioning: used to calibrate the detector
  ▪ performance of the detector: b-tag efficiency, JES
- Top quark physics: mass, spin, charge, V_{tb}, ...
- It will be a background for other signals: Higgs, BSM, ...

Top Quark Pair Production

Production:



87% at LHC
15% at Tevatron



13% at LHC
85% at Tevatron

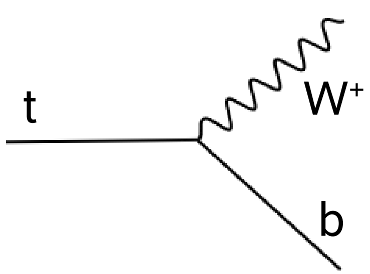
LHC 14TeV	LHC 10TeV	Tevatron
~900pb	414pb	7.6pb

Observed (Tevatron):
~20k evts at CDF ($3fb^{-1}$)
~8k evts at D0 ($1fb^{-1}$)

Predicted (LHC):
9M evts per year ($10fb^{-1}$ @ 14TeV)
80k evts in 2009-2010 run
($200pb^{-1}$ at 10TeV)

Decay:

The top quark decays before hadronization



SM: $t \to Wb \approx 100\%$

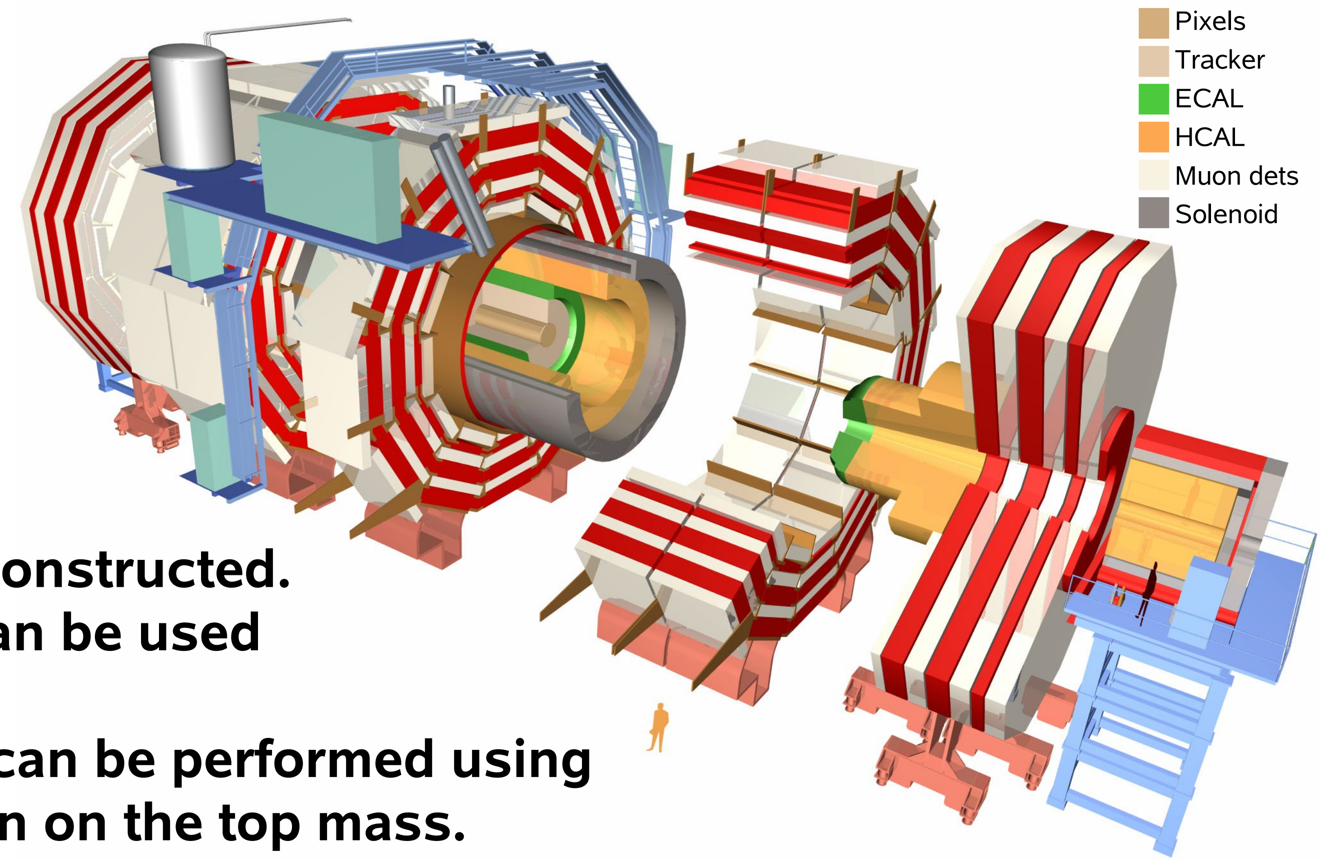
$W \to qq'$: 68%
$W \to l\nu_l$: 32%

→ hadronic: 44%
semi-leptonic: 45%
dileptonic: 11%

Dileptonic:
→ ee: 1.2%
eμ: 2.4%
μμ: 1.2%
ττ : 1%
τl : 5%

CMS and Commissioning

- **The sub-detectors have to be aligned and calibrated.**
- **Reconstruction methods have to be controlled**



<u>Jet Energy Scale:</u>
From semi-leptonic tT, the W mass can be reconstructed. Its well known value can be used to adjust the JES. Similarly, JES of b-jet can be performed using the additional constrain on the top mass.

<u>B-tagging efficiency:</u>
Pure b-jet samples can be obtained from eμ, e+jet and μ+jet channels. b-tagging efficiency can be extracted

Alignment and calibration was (partially) performed in CRAFT and CRUZET

Dileptonic $t\bar{T}$

Dileptonic events: 2 high p_T isolated opposite sign leptons, 2 b-jets and MET

- Clear signature
- Less affected by multi-jets backgrounds than semi-leptonic
  → early observation
- Two neutrinos → not used to measure m_{top} (semi-leptonic instead)
- Divided in leptonic channels:
  - eμ: clear channel, not contaminated by D.Y.
  - ee/μμ: complementary to the eμ channel
  - eτ/μτ: if τ → e/μ (35%), incorporated in ee/μμ/eμ channels
    if τ → had. (65%), tau-jet reconstructed (→ similar to semi-leptonic)

Main Backgrounds:

- Diboson, Drell-Yan ($\gamma/Z^{(*)}$+jets) (with 2 real isolated leptons)
- W+jets or semi-leptonic tT (with 1 real isolated leptons and a "fake" lepton*)
- Multi-jet QCD or fully hadronic tT (with 2 "fake" leptons*)

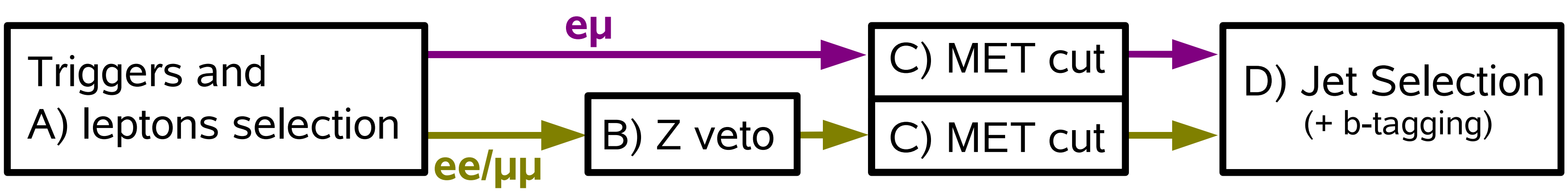
*"fake" leptons: from heavy flavor, in a jet, or not a real lepton

Early Dileptonic tT

Early Analyses:

- **Robustness and simplicity → simple counting experiment**
- **Use of Mis-alignment and Mis-calibration (MisAlCa) scenarios**
- **Data-driven methods needed to control backgrounds**
- **Some variables are not reliable at startup: MET, b-tag,... → loose/robust cut**
- **Complementary strategies (cross-check, robustness against failures, ...)**

Schematic Cut Selections Steps:



Data Driven Background Estimation:

Large uncertainties on normalization of backgrounds (QCD, D.Y., W+Jets)

- **Low jet multiplicity bins**
- **Control Region (i.e. for Z+Jets, inverting the Z veto)**
- **Fake Rate estimate (needed for QCD and W+jets)**
- **Template Fit or other methods**

Early Dileptonic tT Observation

Cross-section extraction:

simple counting method:

$$\sigma_{t\bar{T}} \times BR = \frac{N_{sel} - N_{bkg}}{\varepsilon_{t\bar{T}} \times \int \mathcal{L}}$$

σ_{tT}: tT cross-section

BR: dileptonic branching ratio

N_{sel}: events selected

N_{bkg}: events estimated for the background

$\varepsilon_{tT} = \varepsilon^{HLT}_{tT} \times \varepsilon^{MC}_{tT} \times \varepsilon^{reco/sel}_{tT}$

with ε^{HLT}_{tT} : trigger eff. estimated from data

ε^{MC}_{tT} : selection eff. estimated from MC

$\varepsilon^{reco/sel}_{tT}$: corr. factor between MC and data

$\int \mathcal{L}$: integrated luminosity

Systematics uncertainties:

- Experimental:
  - Leptons Reconstruction and Selection
  - Jet Energy Scale and Jet Energy Resolution
  - b-tag Efficiency and Mis-tag
- Theoretical:
  PDF, ISR/FSR, Jet Multiplicity Spectrum, ...
- Background Normalization
- MC Statistics
- Luminosity
- Multiple pp collision

10/pb @ 14 TeV

PAS: TOP-08-001

Cross-section after 10pb^{-1}, in ee/μμ/eμ channels

Early analysis:

- **MisAlCa startup scenario** (expected mis-alignment and mis-calibration of the detector)
- **Quality cuts on leptons are robust**
- **MET:** not calibrated, but corrected with respect to muons
  not stringent cut
  Mis-measurement in D.Y. → angle α between dilepton p_T and MET vector
- **No b-tagging**

A) Leptons selection:
- $p_T > 20$ GeV, $|\eta| < 2.4$
- Quality cuts (leptonID, χ^2/ndof, ...)
- Isolation: $p_T / (p_T + S) > 0.92$ with $S = \Sigma\, p_T(\text{tracks}) + \Sigma\, E_T(\text{calo})$ around the lepton ($\Delta R<0.3$)

B) Z-Veto: $|m_{ll} - m_Z| < 15$ GeV, for ee and μμ channel only

C) MET cut: ee/μμ channel: MET >30 GeV
MET $>0.6\times p_T^{ll}$ OR $\alpha>0.25$
eμ channel: MET >20 GeV

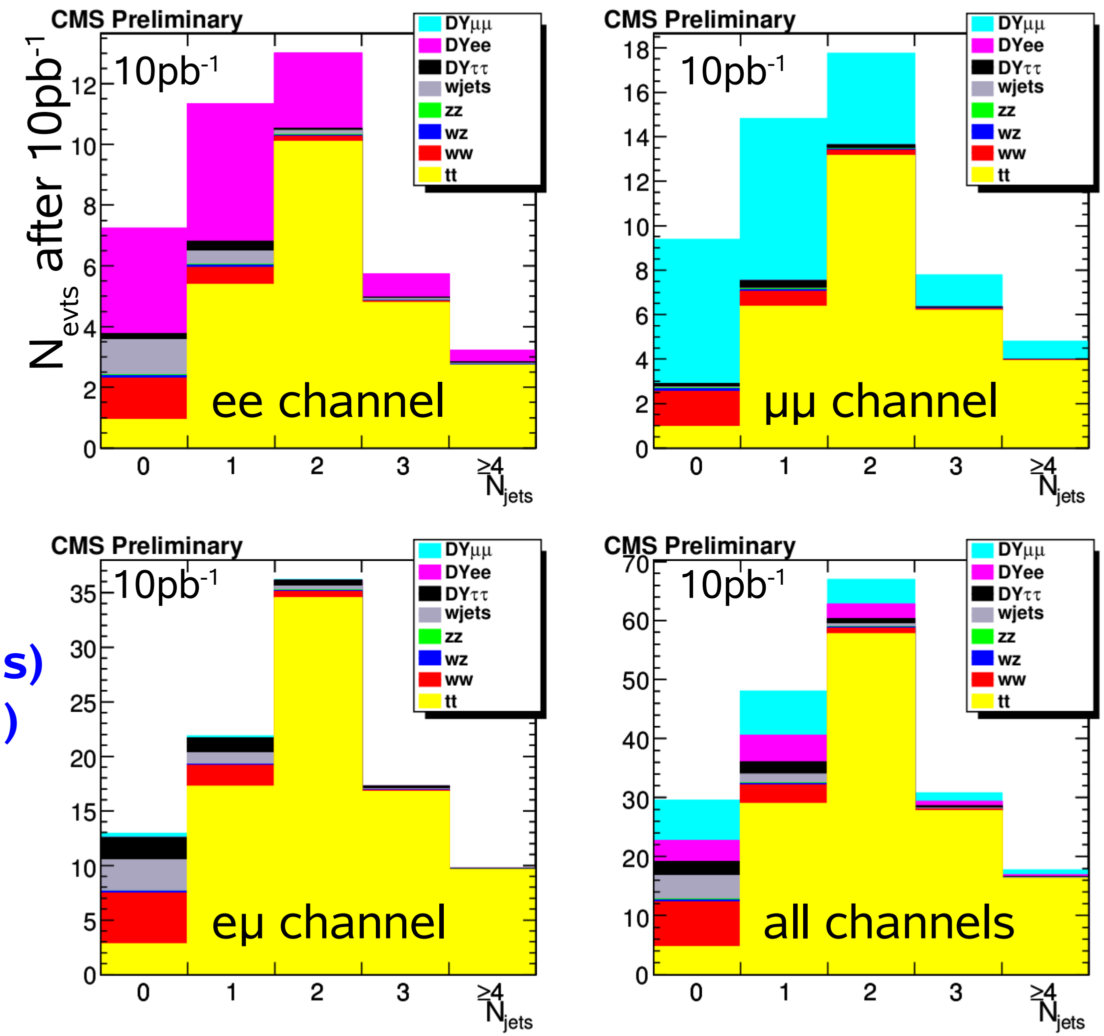
D) Jet Selection: >= 2 Jets with $E_T > 30$ GeV, Iterative Cone with radius 0.5

10/pb @ 14 TeV

PAS: TOP-08-001

Results:

- **>= 2 jets region quite pure, especially in the eμ channel**
- **Backgrounds:**
  - **D.Y.: main background in ee/μμ**
  - **W+jets: can be improved by quality cut on leptons**
  - **Diboson: small contribution**
- **Pure eμ channel, which can be used alone if backgrounds (D.Y.) or MET not correctly controlled**



S/B= 7, S= 120 events (all channels)
S/B= 25, S= 61 events (eμ channel)

Expected stat. error:
9% all channels
13% eμ channel

Syst. Error expected to be similar

100/pb @ 14 TeV

PAS: TOP-08-002

Cross-section after $100pb^{-1}$, in ee/μμ/eμ channels

Early analysis:

- MisAlCa $10pb^{-1}$ scenario (based on cosmic muons, minimum bias and low mass resonances)
- Quality cuts on leptons are robust
- MET calibrated, corrected with respect to muons
- b-tagging: track counting algorithm, loose working point

A) Leptons selection:

- p_T > 20 GeV, |η| < 2.4
- Quality cuts (leptonID, χ^2/ndof, ...)
- Isolation: TrackIso(lept) < 3 GeV/c, CaloIso(e) < 6 GeV, CaloIso(μ) < 1 GeV — sum in tracker/calo around the lepton (ΔR<0.3)

B) Z-Veto: m_{ll} < 75 GeV OR m_{ll} > 105 GeV, for ee and μμ channel only

C) MET cut: ee/μμ/eμ channel: MET >50 GeV

D) Jet Selection: >= 2 Jets with E_T > 30 GeV, Iterative Cone with radius 0.5
both jets are required to be b-jets

100/pb @ 14 TeV

PAS: TOP-08-002

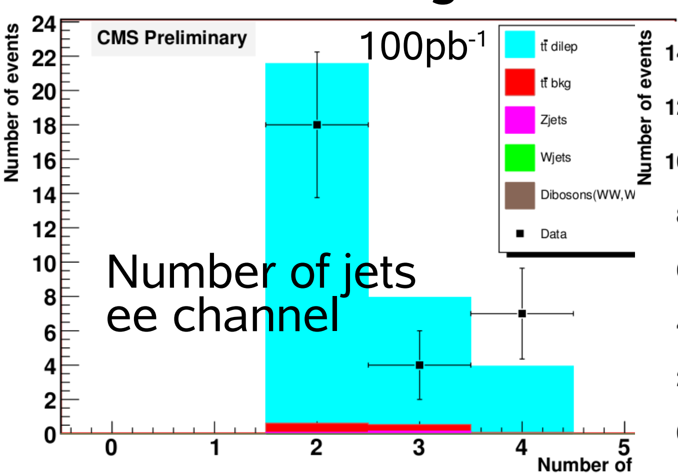
Results:

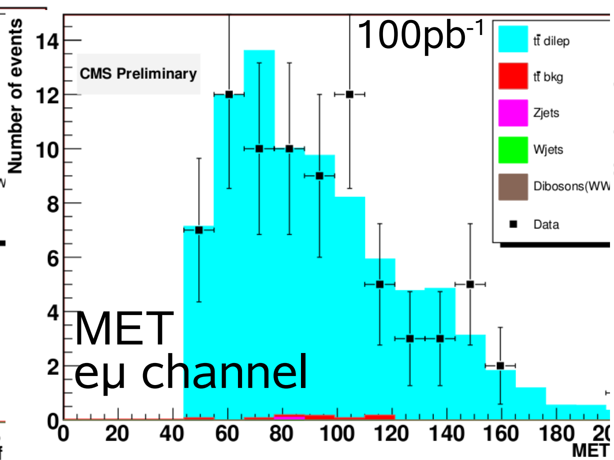
- More stringent cuts than 10pb^{-1} analysis → better S/B
- Backgrounds:
  - ee channel mainly contaminated by fake leptons
  - μμ channel contaminated by Z+jets
  - eμ channel quite pure
- Heavy flavour sensitivity tested with enriched MC samples
- Matrix Method and Factorization used to evaluate backgrounds

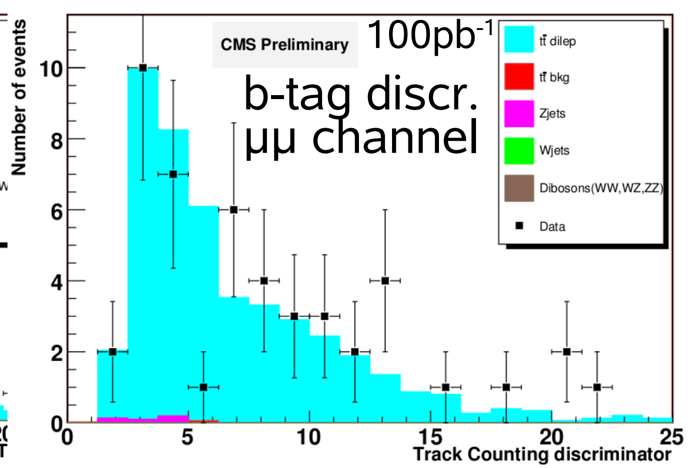
S/B= 26, S= 32 events (ee channel)
S/B> 90, S= 84 events (eμ channel)
S/B> 90, S= 45 events (μμ channel)

S/B without b-tag: ee: 4, eμ: >20, μμ: 6

Expected stat. error:
ee: 15%, μμ: 18%, eμ: 11%
all channels: 8%







eτ/μτ 100/pb @ 14 TeV

PAS: TOP-08-004

Cross-section ratio $\sigma[t\bar{t} \rightarrow (l\nu)(\tau\nu_\tau)b\bar{b}] / \sigma[t\bar{t} \rightarrow (l\nu)(l\nu)b\bar{b}]$ after 100pb^{-1} (l=e,μ)

Early analysis:

- MisAlCa 10pb^{-1} scenario
- Quality cuts on leptons are robust
- MET calibrated, corrected with respect to muons
- 10% variation on the tau reconstruction algorithm parameters to evaluate syst.
- different b-tagging scenarios considered

A) Leptons selection:

- p_T > 20 GeV, |η| < 2.4
- Quality cuts for tau-jet
- Isolation: TrackIso(lept) < 3 GeV/c, CaloIso(e) < 6 GeV, CaloIso(μ) < 5 GeV — sum in tracker/calo around the lepton (ΔR<0.3)
- other leptons veto

C) MET cut: eτ/μτ channel: MET >60 GeV

D) Jet Selection: >= 2 Jets with E_T > 30 GeV, Iterative Cone with radius 0.5

eτ/μτ 100/pb @ 14 TeV

PAS: TOP-08-004

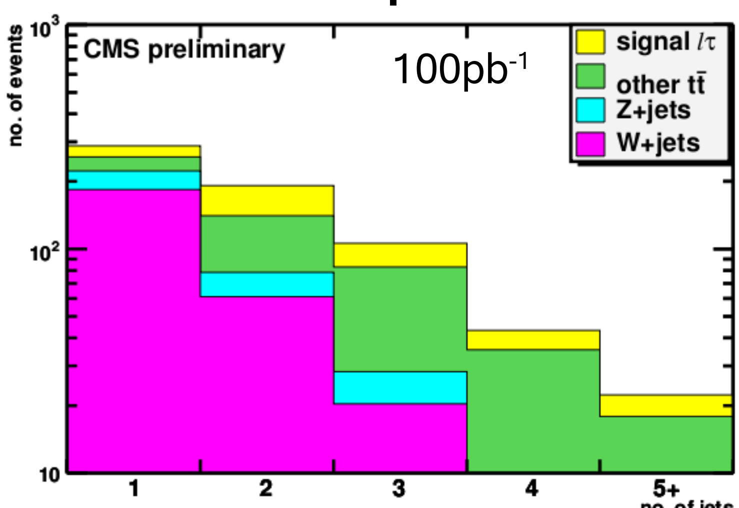
Results:

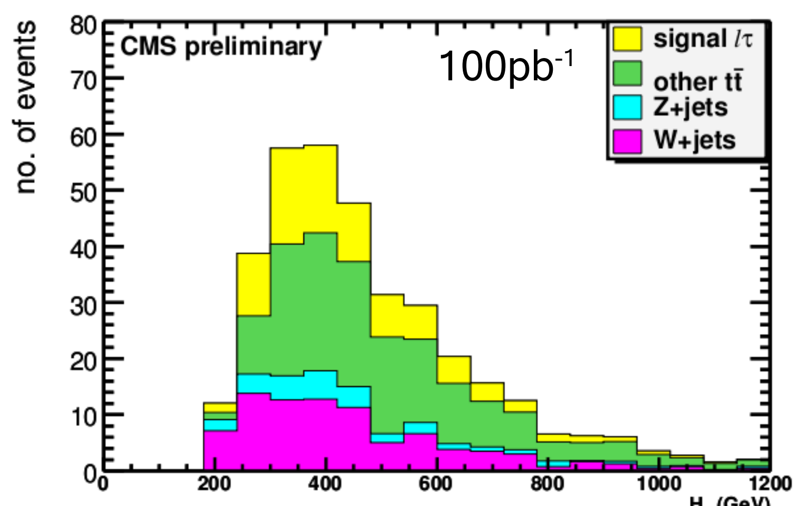
- R = $\sigma^{l\tau} / \sigma^{e\mu}$ can be computed, in order to evaluate deviation w.r.t. SM. The syst. uncertainties cancel in this ratio. The pure eμ channel is used as normalization
- Backgrounds:
  - W+jets: main non-tT background, but can be derived from data
  - Other tT: mainly semi-leptonic
- Fake Rate for tau-jet can be evaluated in jet dominated samples

without b-tag:
S/B= 0.4, S= 86 events (1-prong)

b-tag increase S/B by 2

Expected stat. error:
1-prong without b-tag: ~10%





Perspectives for 2009

The results presented here were released 1 year ago

Possible scenarios:

- **10 TeV instead of 14 TeV**
- **One dileptonic note for three dileptonic tT studies:**
  - **standard analysis after 10/pb @ 10 TeV**
  - **track-based analysis after 10/pb @ 10 TeV**
  - **standard analysis after 100/pb @ 10 TeV**
- **MisAlCa scenario = Ideal scenario (thanks to cosmic data)**
- **Improved data-driven methods:**
  - **Z+Jets: "In/Out Z-peak" at 10/pb, Template Fit at 100/pb**
  - **Fake Rate estimation**
  - **Matrix Method for QCD at 100/pb**

next: theoretical uncert., Top Commissioning, ...

Conclusions

The di-leptonic top pair production will be observable at CMS after few tens pb^{-1} of integrated luminosity

Early observations can be performed in complementary and robust way:

- **after $10pb^{-1}$, without b-tagging**
  backup scenario (eμ channel only) **without MET**
- **after $100pb^{-1}$, with loose working point b-tagging**
- **after $100pb^{-1}$, in the eτ and μτ channel**

Early tT observation at LHC:

- **First complex SM process at LHC energies**
  → understanding and calibration of the detector
- **Observation of a background in many signals**
- **First step to the Top Quark Physics program at LHC**
  (top physics, BSM, ...)

BACKUP SLIDES

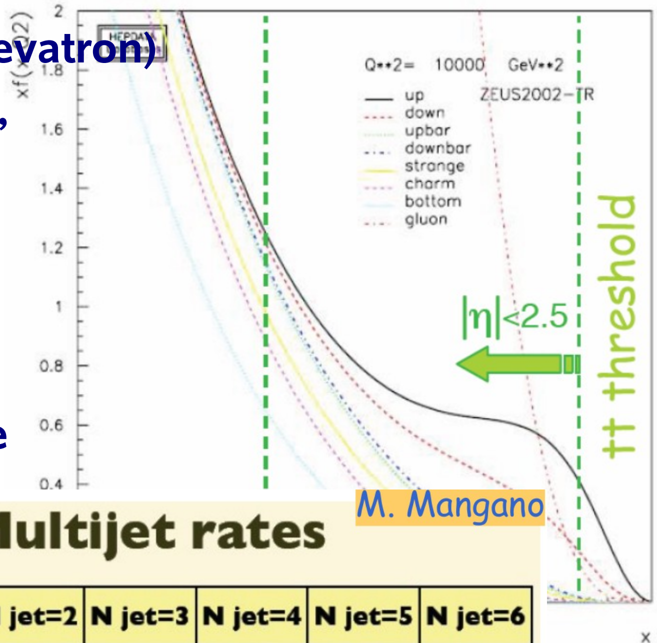
Tevatron → LHC

$\sigma_{tT}(LHC)=100\times\sigma_{tT}(Tevatron)$

σ_{tT}/σ_W is also larger, but not $\sigma_{tT}/\sigma_{W+>=4jets}$

Others difficulties:

- pile-up
- FR and HF difficult to simulate
- ...

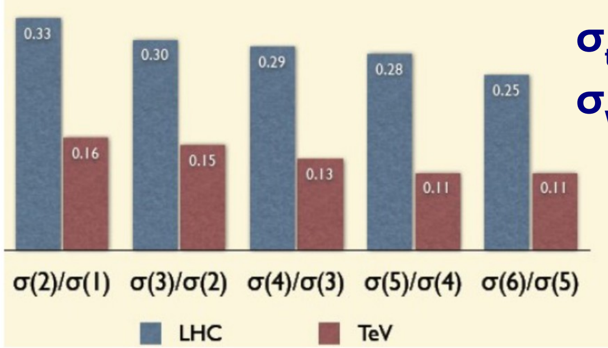


M. Mangano

W+Multijet rates

σxB(W→eν)[pb]	N jet=1	N jet=2	N jet=3	N jet=4	N jet=5	N jet=6
LHC	3400	1130	340	100	28	7
Tevatron	230	37	5.7	0.75	0.08	0.009

E_T(jets) > 20 GeV , |η|<2.5 , ΔR>0.7



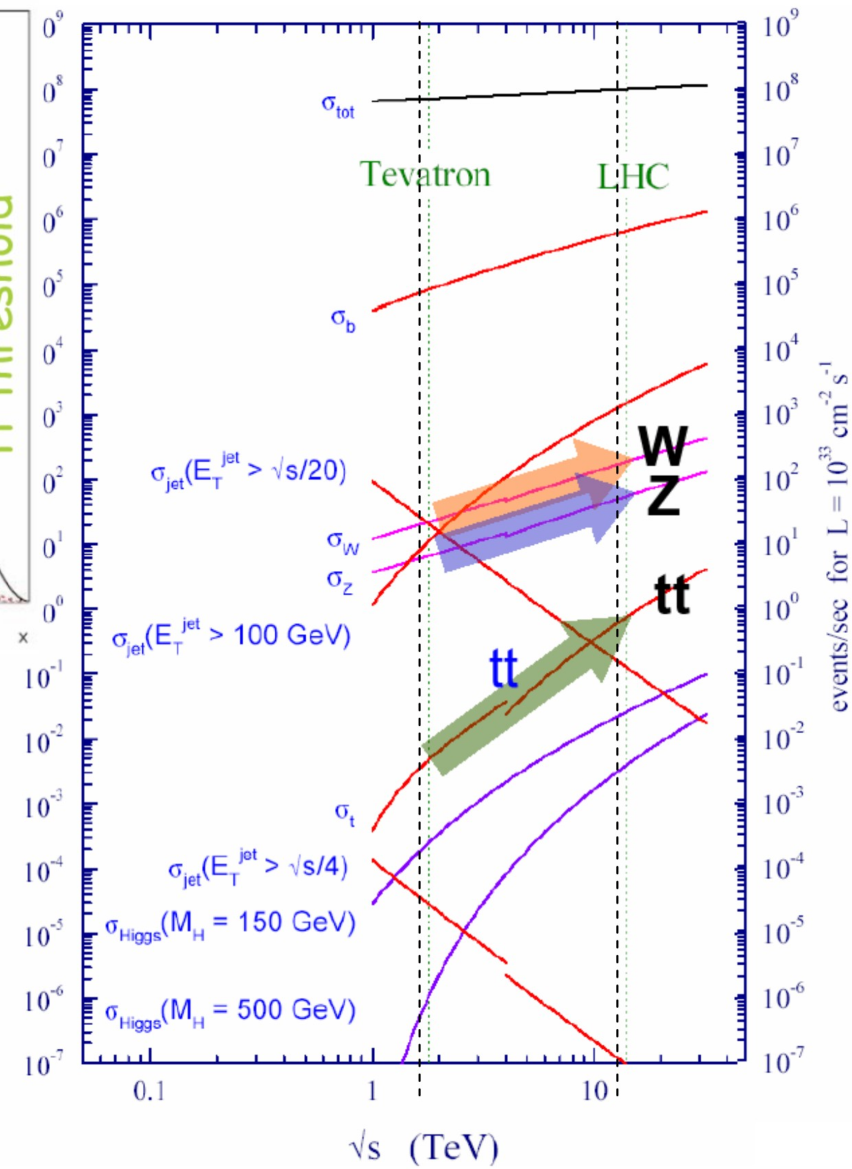
- Ratios almost constant over a large range of multiplicities
- O(α_s) at Tevatron, but much bigger at LHC

29

14 TeV → 10 TeV:

σ_{tT} ↘ factor 2.3

σ_W ↘ factor 1.4



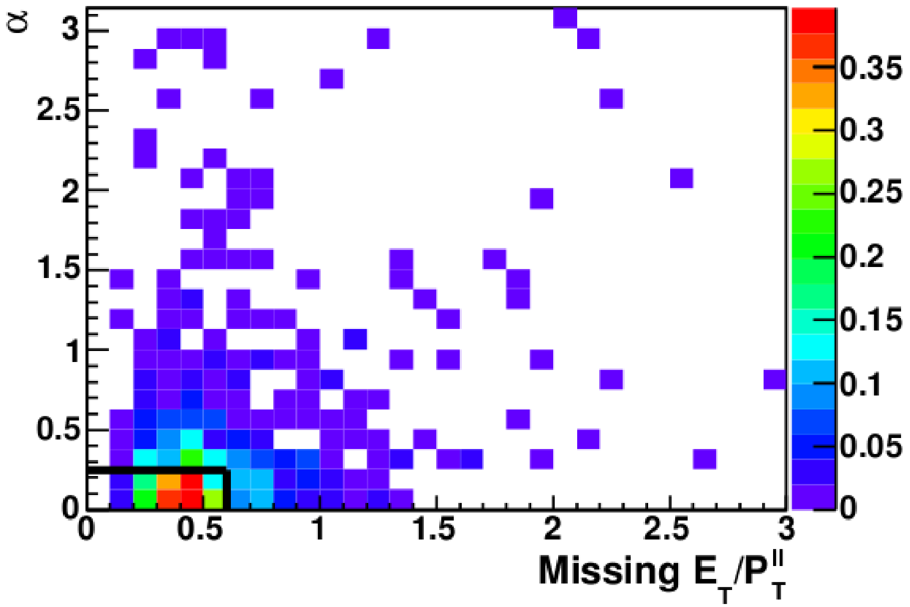
14TeV:

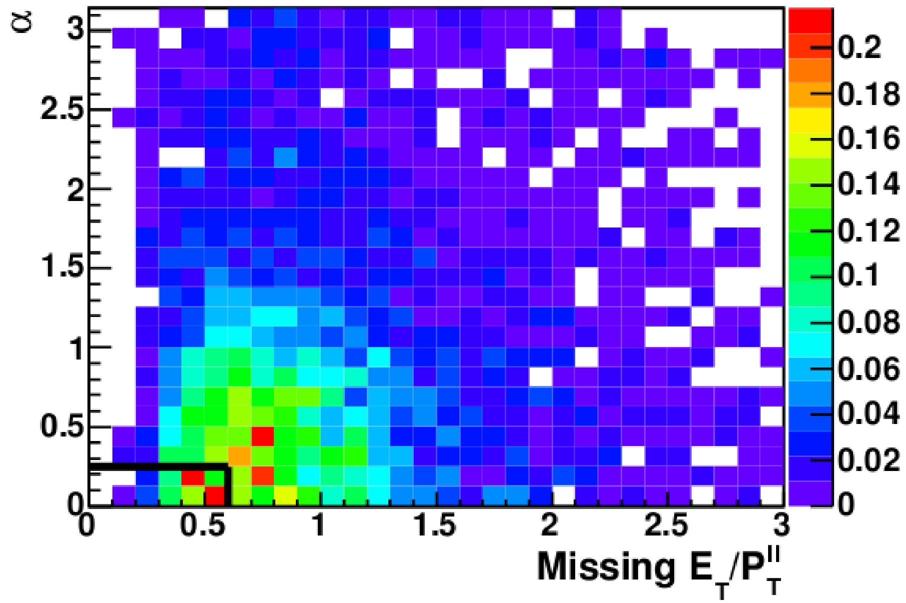
process	MC Generator	Assumed Cross Section
$t\bar{t}$	Alpgen	837 pb
W+jets	Alpgen	65.2 nb
DY+jets	Alpgen	6.46 nb
WW	Pythia	114.3 pb
WZ	Pythia	49.9 pb
ZZ	Pythia	16.1 pb

10TeV:

process	MC Generator	Cross Section
$t\bar{t}$	Madgraph	414 pb
W + jets	Madgraph	46 nb
DY+jets	Madgraph+Pythia*	6.6 nb
$WW \rightarrow 2\ell 2\nu$	Pythia	7.2 pb
WW	Pythia	74 pb
WZ	Pythia	29 pb
$ZZ \rightarrow 2\ell 2\nu$	Pythia	0.41 pb
$ZZ \rightarrow 4\ell$	Pythia	0.135 pb
ZZ	Pythia	10.4 pb
tW	Pythia	38.2 pb
$pp \rightarrow \mu + X$	Pythia	122 nb
$pp \rightarrow (b, c \rightarrow e) + X$	Pythia	455 nb
$pp \rightarrow em + X$	Pythia	8185 nb

10/pb @ 14 TeV





ee/μμ channel:

- **MET >30 GeV**
- **MET >0.6×p_T^{ll} OR α>0.25**

eμ channel: MET >20 GeV

Data-driven Method: Fake Rate

A: Fakes not selected | B: Fakes selected | C: Not fakes ← selected

Fake Rate = Selected Objects / Fakeable Objects = B+C / A+B+C

with FO is all object passing loose isolation conditions

FR will be computed from data (triggered in an other object)

N_{FR} = Events with 1 FO passing the standard conditions
1 FO failing the standard conditions

$$N_{estimated} = N_{FR} \times FR/(1-FR)$$

But the "not fakes" are supposed to have 100% efficiency under standard conditions → overestimation
MC tests show that this overestimation is small (~ 1% of $N_{estimated}$)

Ways to estimate this bias from data are also proposed, in measuring FR on clean "not fake" leptons region, in general Z→ll

Data-driven Method: Matrix Method

Matrix Method:

Three kinds of backgrounds: Signal-like, 2 good leptons (S)
W+jets-like, 1 fake lepton (W)
QCD-like, 2 fakes (QCD)

Three kinds of selections: Tight (t), standard isolation for both leptons
Medium (m), at least one lepton pass the standard iso.
Loose (l), relaxed isolation

$$N^t = N^t_S + N^t_W + N^t_{QCD}$$
$$N^m = N^m_S + N^m_W + N^m_{QCD}$$
$$N^l = N^l_S + N^l_W + N^l_{QCD}$$

but $N^t_i = \varepsilon^{l\to t}_i N^l_i$

$N^m_i = \varepsilon^{l\to m}_i N^l_i$

and $\varepsilon^{l\to\{m,t\}}_i$ **can be evaluated by factorization approach**

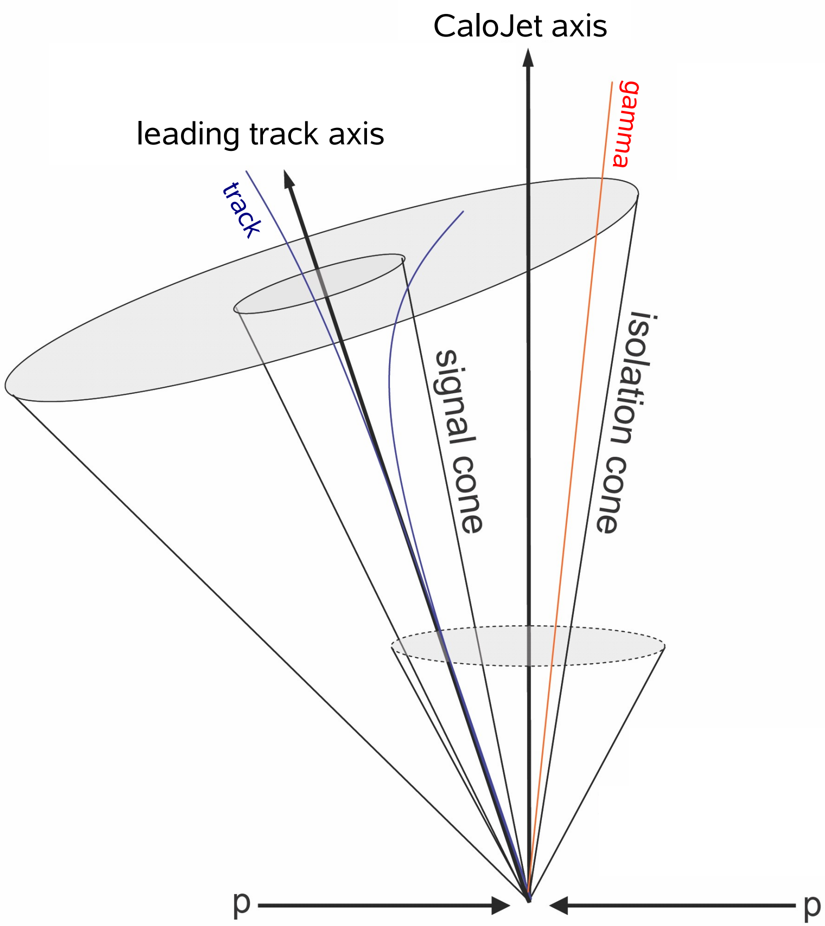
$$\begin{aligned}
\varepsilon_S^{l\to t} &= \varepsilon_s^2, \\
\varepsilon_W^{l\to t} &= \varepsilon_s.\varepsilon_{fake}, \\
\varepsilon_{QCD}^{l\to t} &= \varepsilon_{fake}^2. \\
\varepsilon_S^{l\to m} &= 2\varepsilon_s - \varepsilon_s^2 \\
\varepsilon_W^{l\to m} &= \varepsilon_s + \varepsilon_{fake} - \varepsilon_s.\varepsilon_{fake} \\
\varepsilon_{QCD}^{l\to m} &= 2\varepsilon_{fake} - \varepsilon_{fake}^2
\end{aligned}$$

$$\begin{aligned}
N_S^l &= \frac{N^t - \varepsilon_{fake}(N^m + N^t - \varepsilon_{fake}N^l)}{(\varepsilon_s - \varepsilon_{fake})^2}, \\
N_W^l &= \frac{(\varepsilon_s + \varepsilon_{fake})(N^m + N^t) - 2(N^t + \varepsilon_s\varepsilon_{fake}N^l)}{(\varepsilon_s - \varepsilon_{fake})^2}, \\
N_{QCD}^l &= \frac{N^t - \varepsilon_s(N^m + N^t - \varepsilon_s N^l)}{(\varepsilon_s - \varepsilon_{fake})^2}.
\end{aligned}$$

Tau

PAS: TOP-08-004

Tau from CaloTau Collection:



- leading track p_T > 5 around the jet axis (ΔR=0.1)
- 0 tracks in an isolation annulus (0.1< ΔR <0.5)

- Electron veto: $E_{tot}/p_{leading\ track}$ < 0.9
- Muon veto: $E_{tot}/p_{leading\ track}$ > 0.5
- "crack" region removal (1.46 < |η| < 1.56)
- E_{ecal} < 3 GeV in ΔR<0.5

-Veto tested by reversing the electron rejection cut (pure Z→ee sample)
The dilepton invariant mass is reconstructed for "all" and "e reversed". Similar method for muons

-Fake probability evaluated in sub-samples of QCD: "leading", "next-to-leading" and "back-to-back" jets
Can be evaluated from Data in OS and SS samples