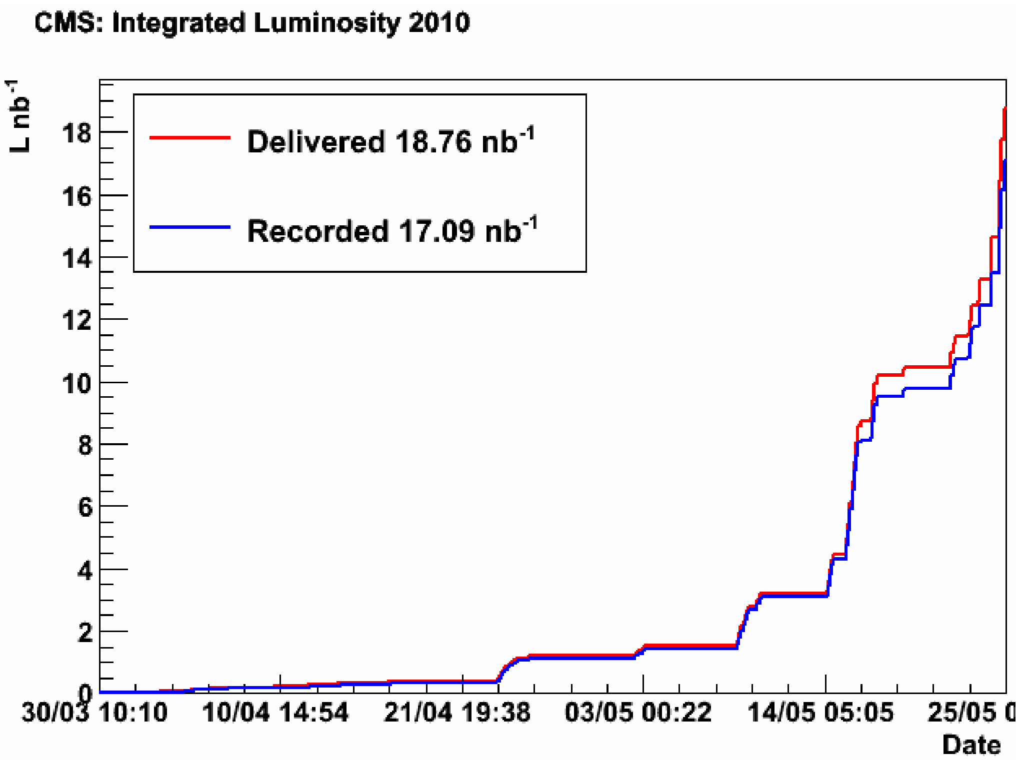
Reconstruction of Jets in Top Events with CMS

Roger Wolf, Universität Hamburg

TOP2010 Bruges 03.06.2010

The Status of LHC

- Delivered(Recorded) **L≈19(17)nb^{-1}** (@ 7TeV).
- CMS uptime: **ε_{Tape}≈92%**.
- Performance (last weekends): **L_{sp} ≈ 6 μb^{-1}min^{-1}**.
- Official target: **0.3-1.0pb^{-1}** by end of June.



Expectation (of reco'ed events):

Event Type	Cross Section	Reco Events per pb (1)
TTbar	165 (NNLO)	O(10)
W-->lv	28000 (NLO)	O(1500)
Z-->ll (2)	3110 (LO)	O(150)

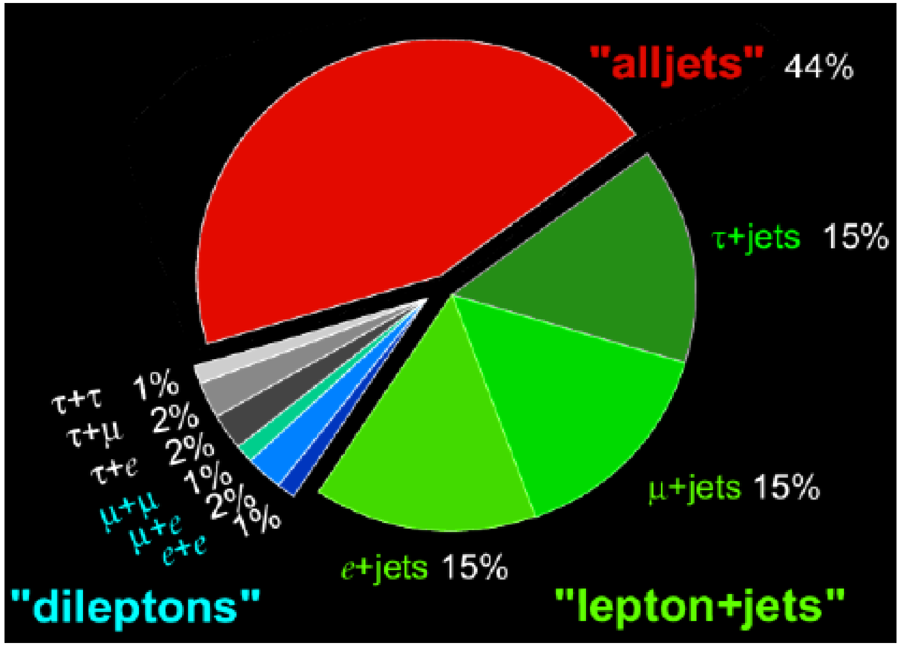
(1) for 5% selection efficiency **(2)** with m(ll)>10!GeV

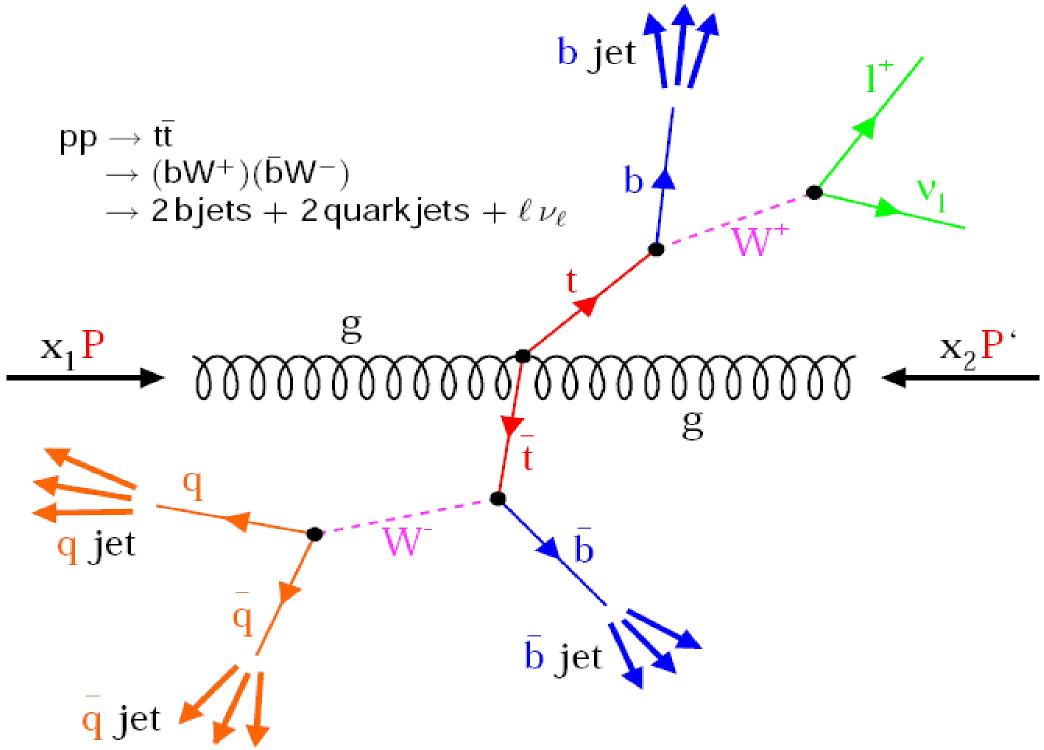
Looking forward to first Top Candidates by ICHEP.

Working hard on a re-discovery before September.

Rediscovery of the Top Quark

Relevant Decay Channels:

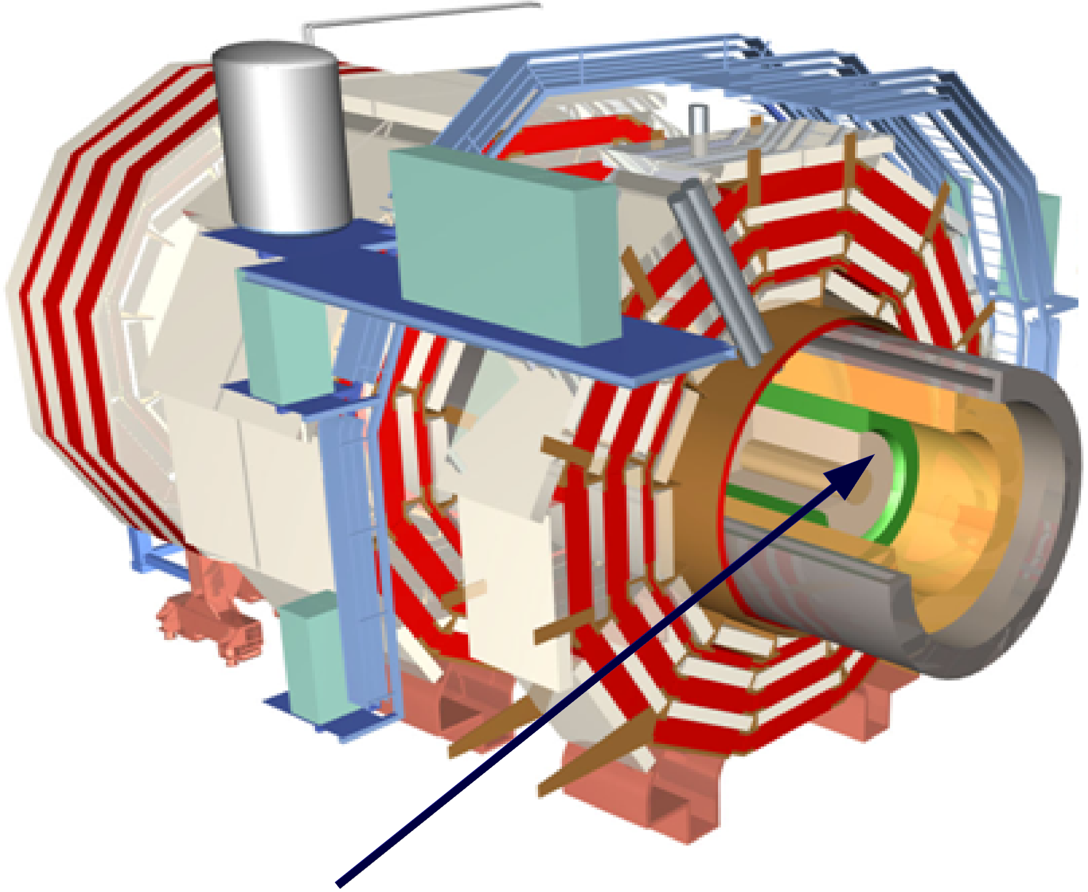




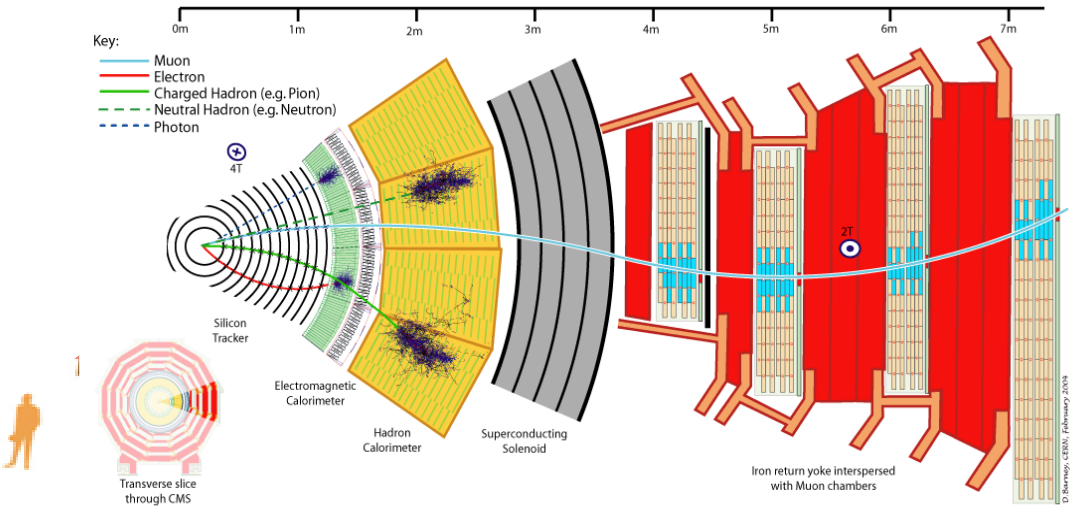
- **dileptonic (eμ, μμ, ee)** ΣBR= 4% (~ 5pb) (Study for $10pb^{-1}$)
- **lepton+jets (μ, e)** ΣBR=30% (~67pb) (Study for $20pb^{-1}$)

For both measurements jets will play an important role!

Jets in the CMS Detector

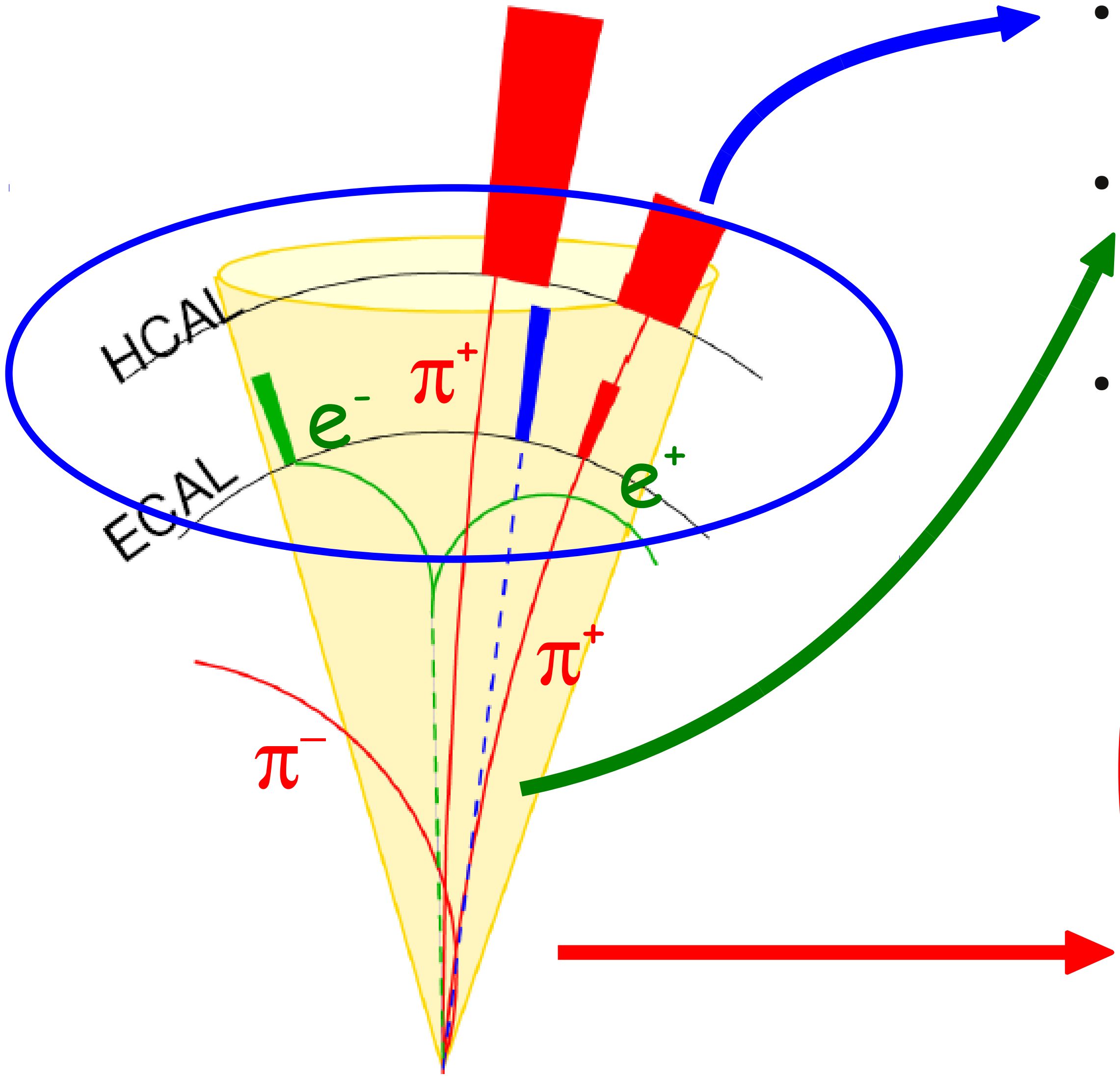


- **Pure Calorimeter Jets** (**Calo**):
  Jets clustered from ECAL and HCAL deposits.
- **Track Supported CaloJets** (**JPT**):
  Replace calorimeter information by more reliable tracker information (H1 approach).
- **Jets from Particle Flow** (**PFlow**):
  Cluster particle flow objects.

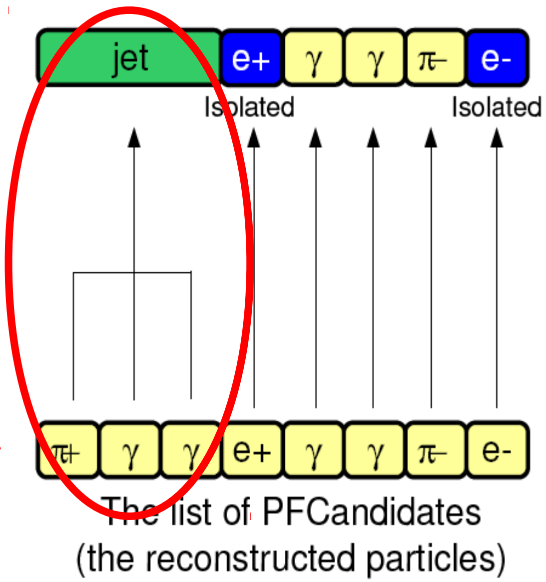


- Silicon-Strip (Particle Flow)
- ECAL ($PbWO_4$, $26X_0$),
- HCAL (Brass,Scintillator, $10\lambda_i$)

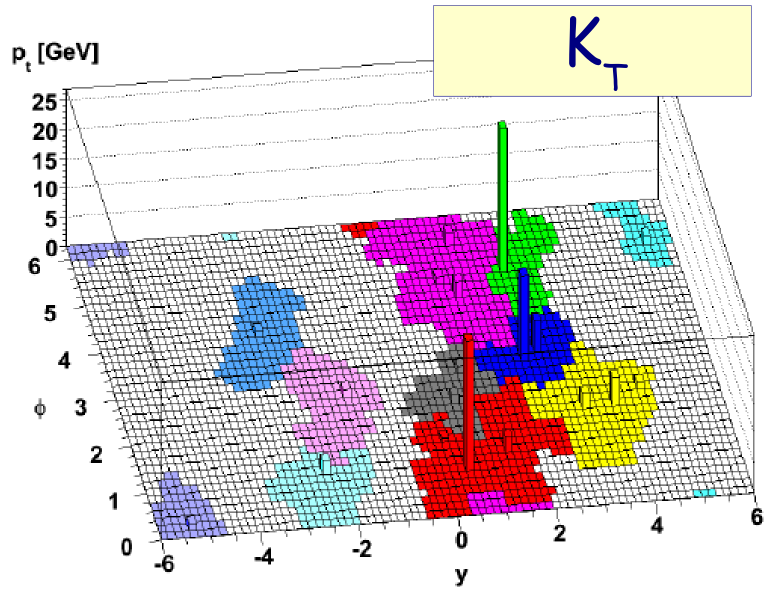
JPT Jets and Particle Flow @ CMS



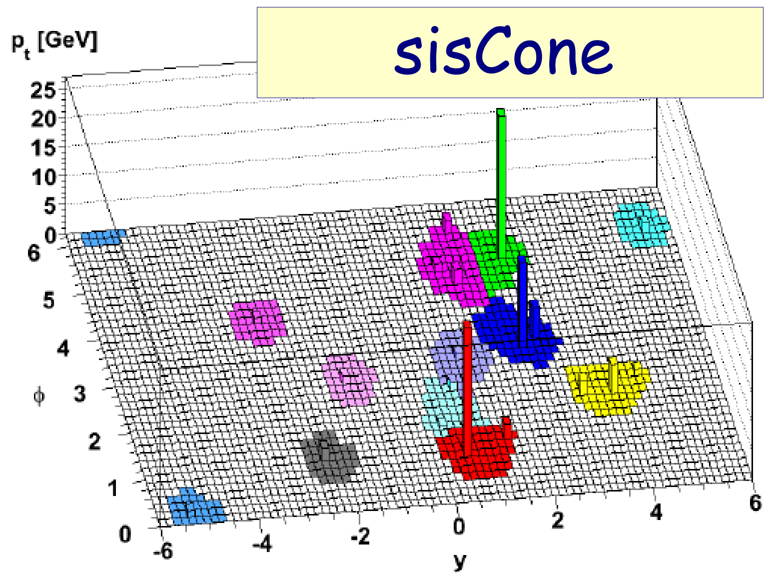
- **Pure Calorimeter Jets (Calo):**
  Jets clustered from ECAL and HCAL deposits.
- **Track Supported CaloJets (JPT):**
  Replace calorimeter information by more reliable tracker information (H1 approach).
- **Jets from Particle Flow (PFlow):**
  Cluster particle flow objects.

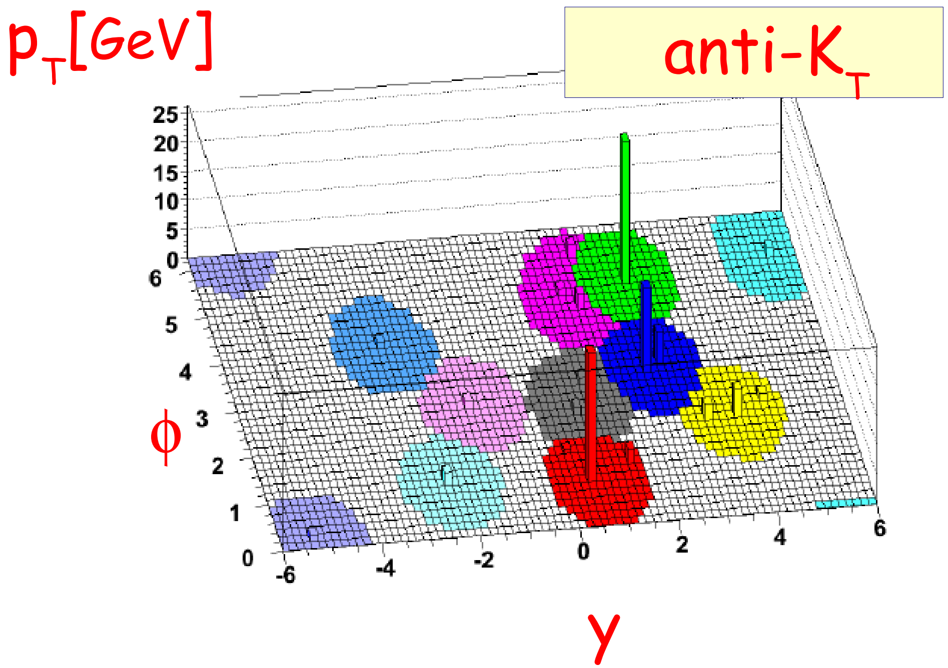


Jet Algorithms in Use

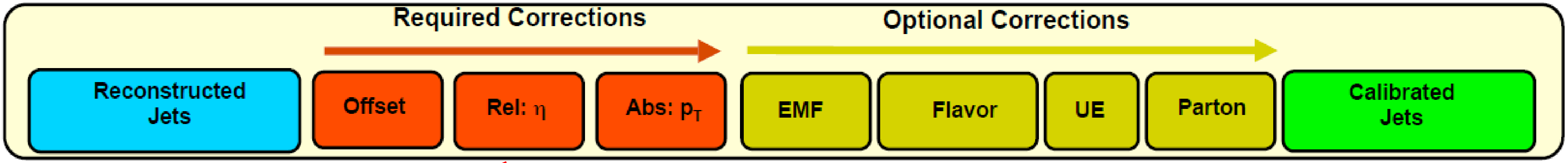


Algorithm	R	GenJet-Label	CaloJet-Label	PFJet-Label
AntiKt	0.5	ak5GenJets	ak5CaloJets	ak5PFJets
AntiKt	0.7	ak7GenJets	ak7CaloJets	ak7PFJets
SISCone	0.5	sisCone5GenJets	sisCone5CaloJets	sisCone5PFJets
SISCone	0.7	sisCone7GenJets	sisCone7CaloJets	sisCone7PFJets
kT	0.4	kt4GenJets	kt4CaloJets	kt4PFJets
kT	0.6	kt6GenJets	kt6CaloJets	kt6PFJets

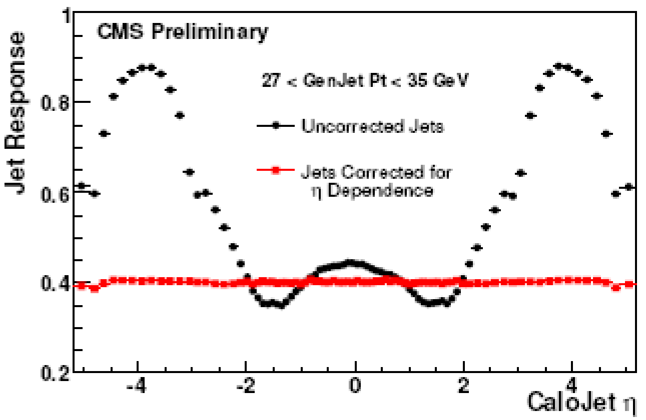




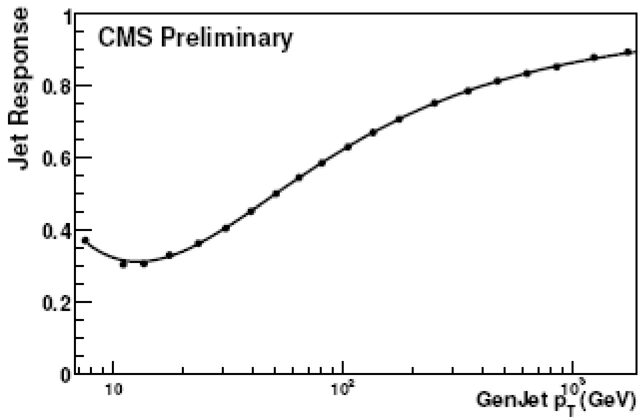
Jet Corrections



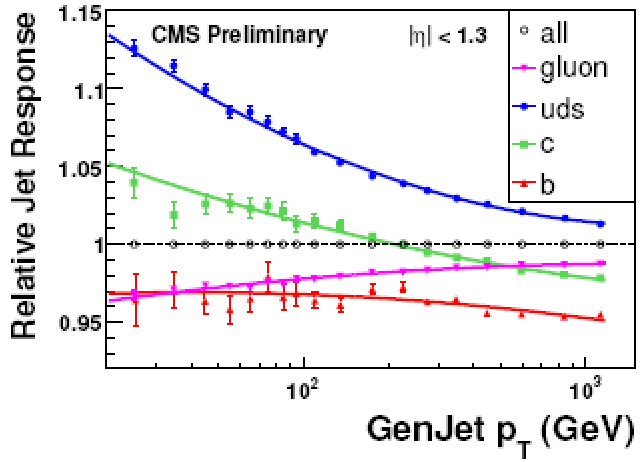
Relative:
correct to uniform calorimeter response in η.



Absolute:
correct absolute energy scale.



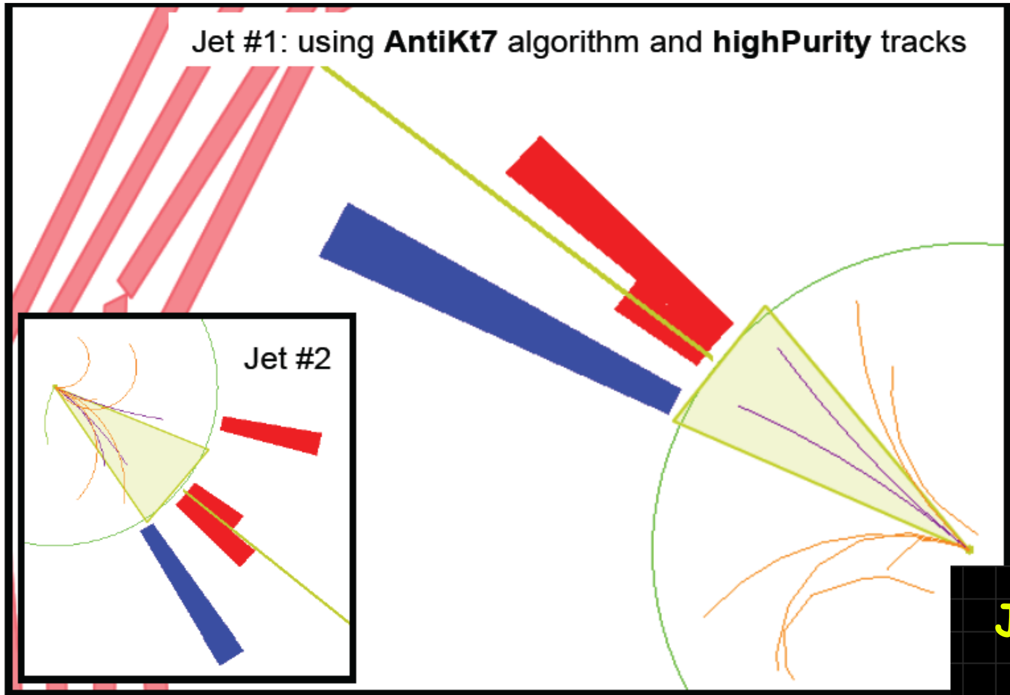
Flavor:
correct for different calorimeter response of different flavors.



- Initial correction up to **Absolute** energy scale only.
- Initial assumption on **JEC uncertainty: ±10%**.

Typical Dijet Event (all representations)

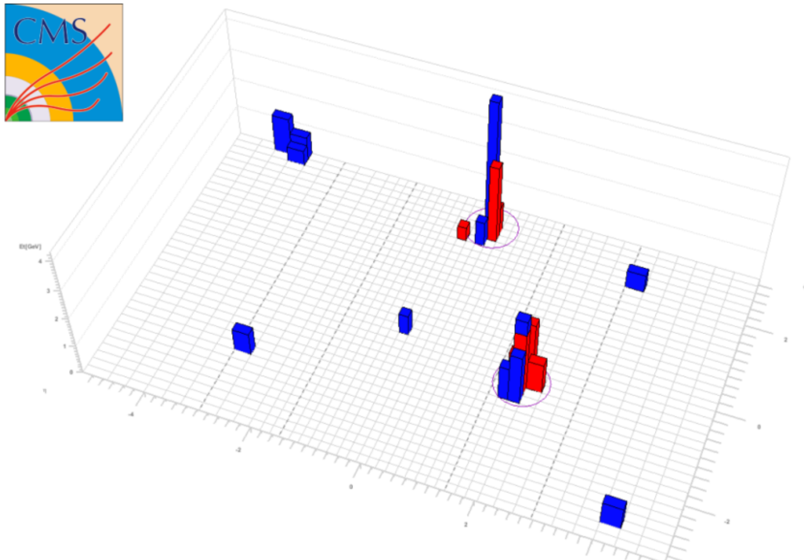
anti-K_T (R=0.7)

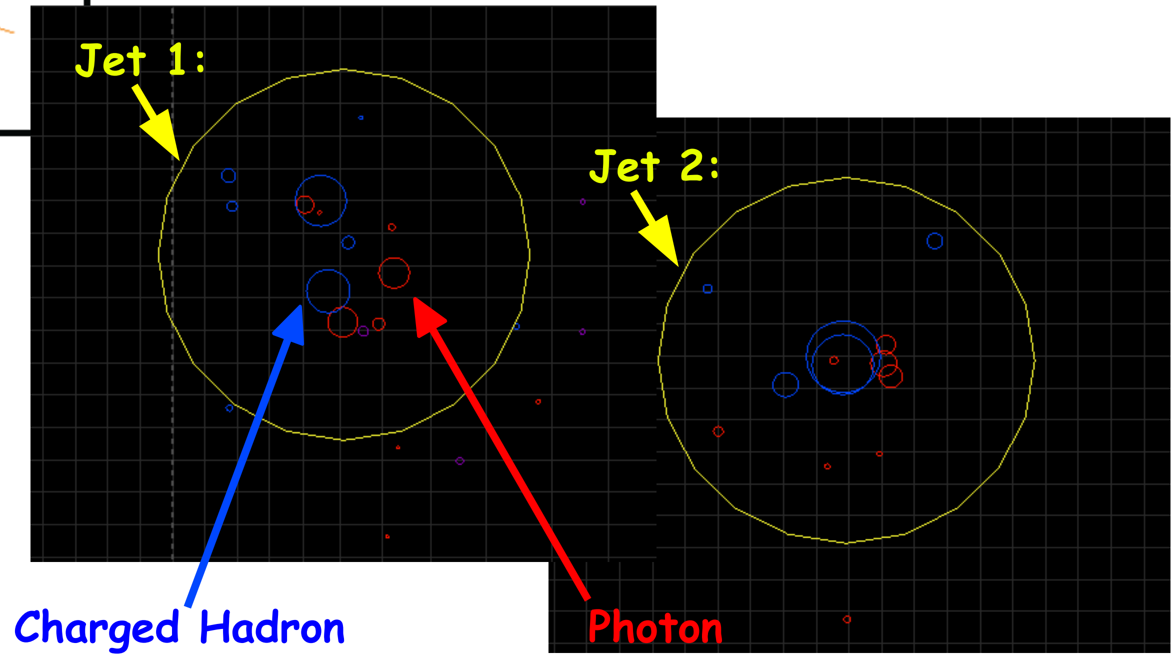


CaloJet:
- p_T(Jet1, Abs) = 26 GeV
- p_T(Jet2, Abs) = 24 GeV

Track Supported Jet (JPT):
- p_T(Jet1, JPT) = 25 GeV
- p_T(Jet2, JPT) = 24 GeV

Particle Flow Jet:

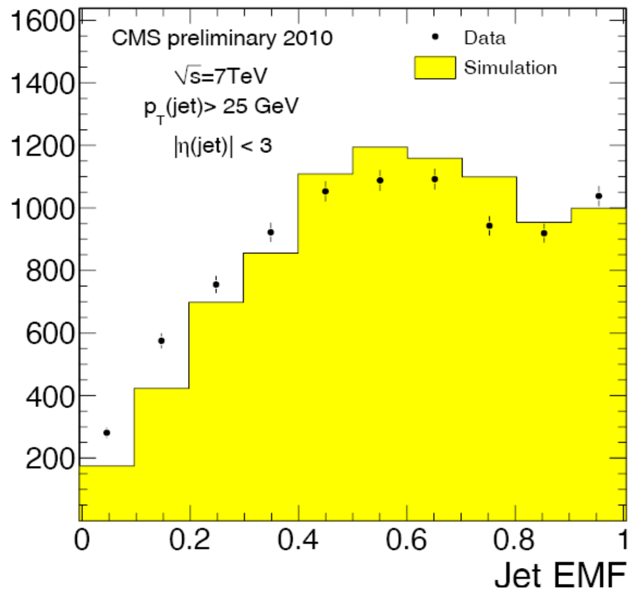


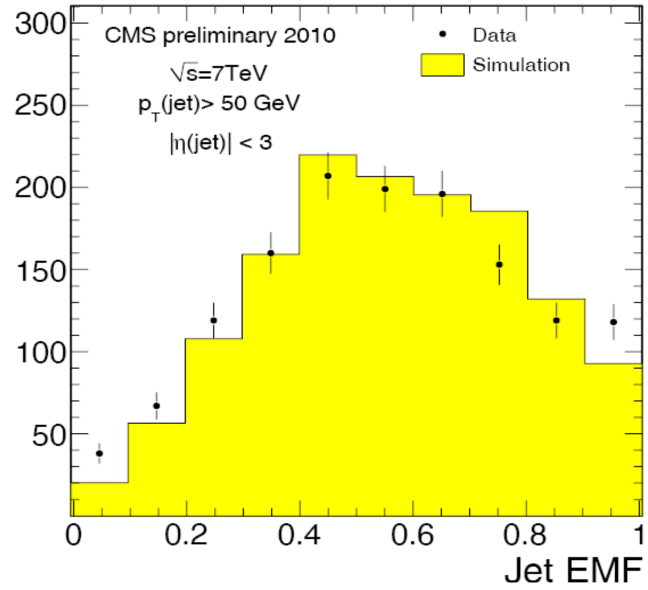


Data vs. Monte Carlo (Dijet Events)

CaloJets

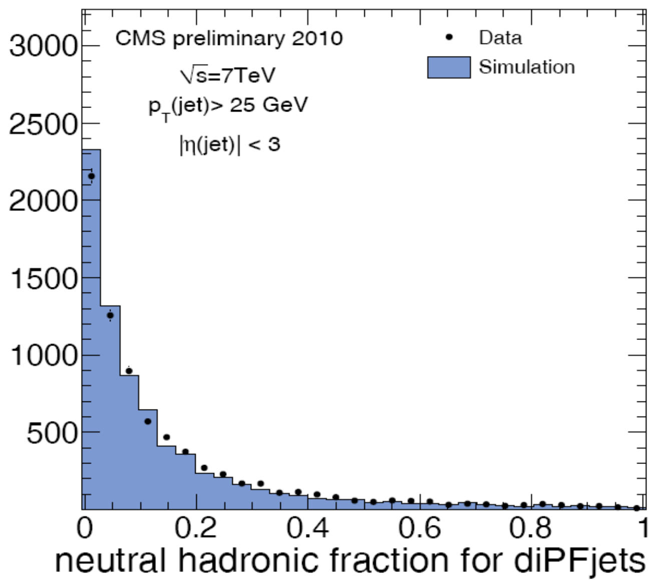
Jet electromagnetic fraction (Jet EMF); central region, different p_T(Jet) thresholds.

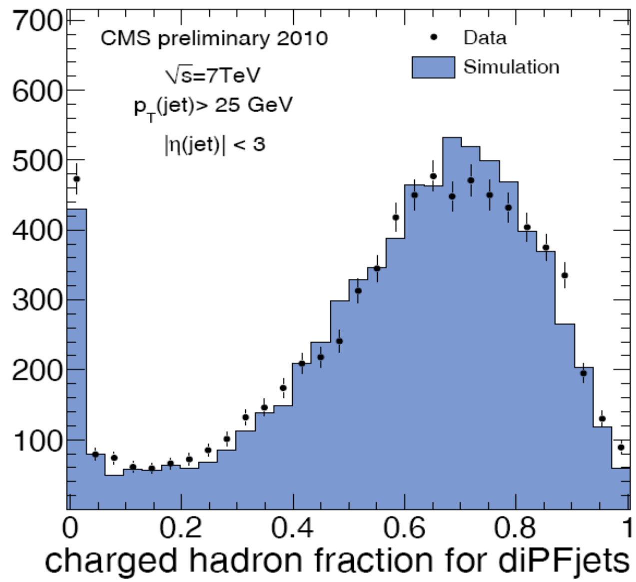




PFlow Jets

Jet charged and neutral hadronic fraction.



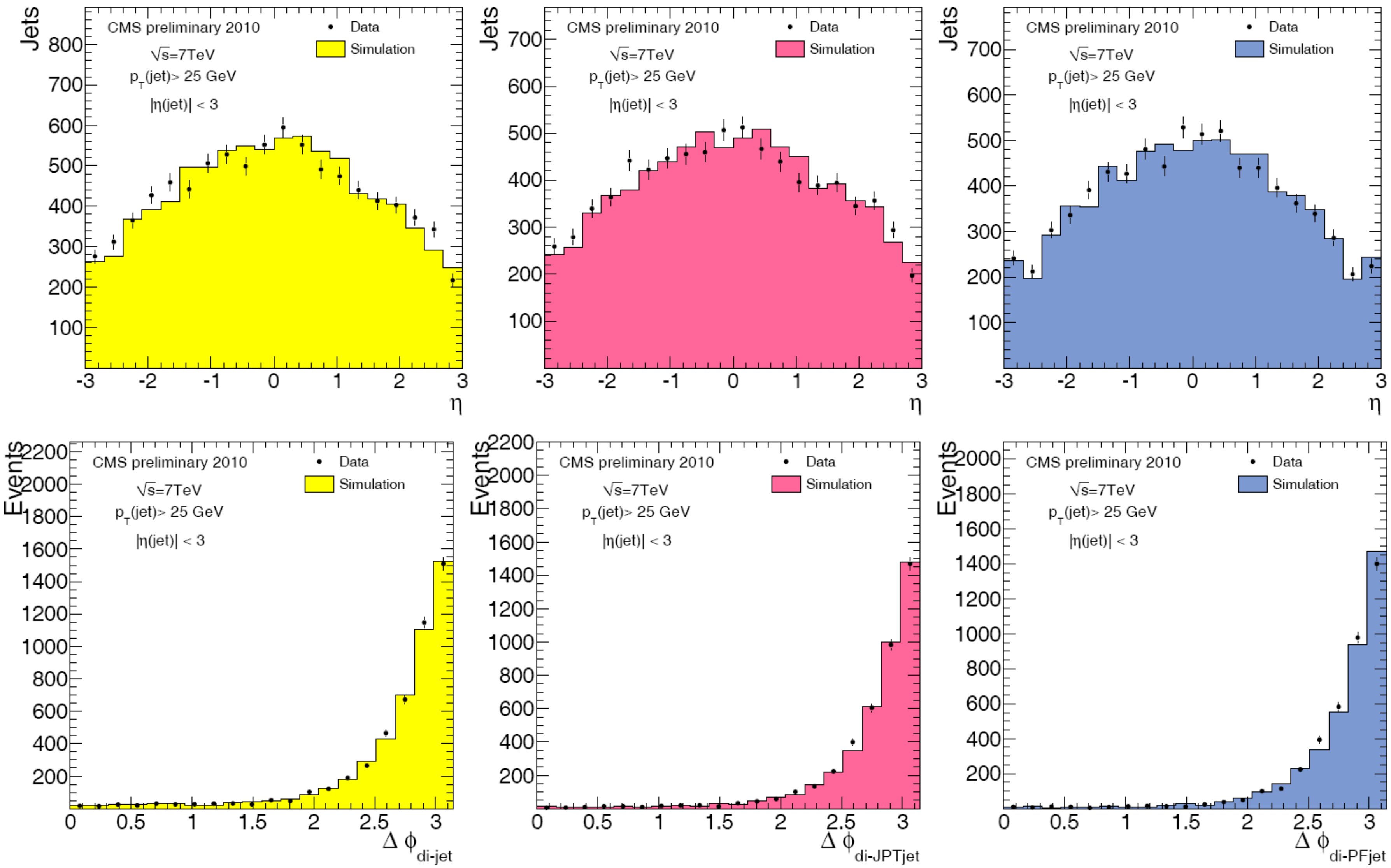


Data vs. Monte Carlo (Dijet Events)

CaloJets | **JPT Jets** | **PFlow Jets**

η of Jets

Δφ(Jet1, Jet2)

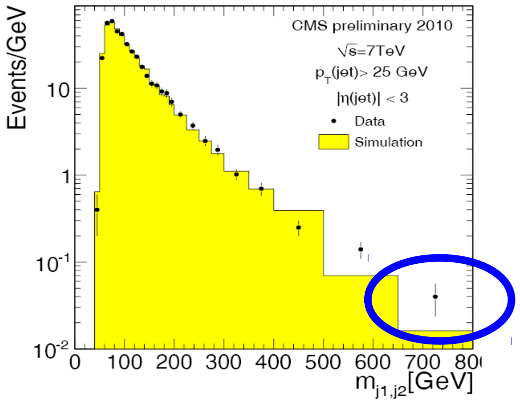


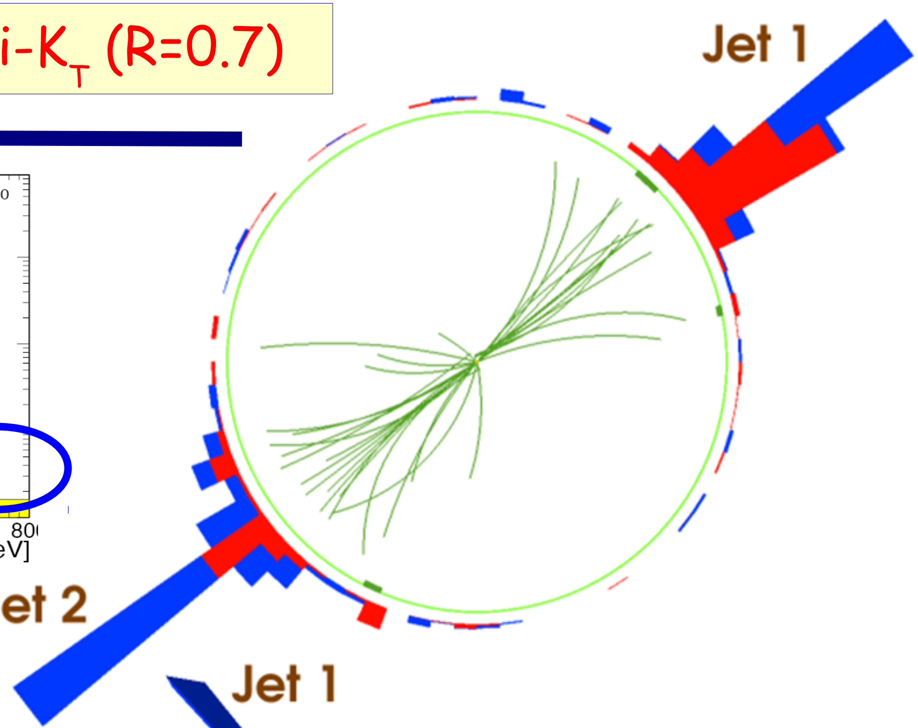
High Mass Dijet Event

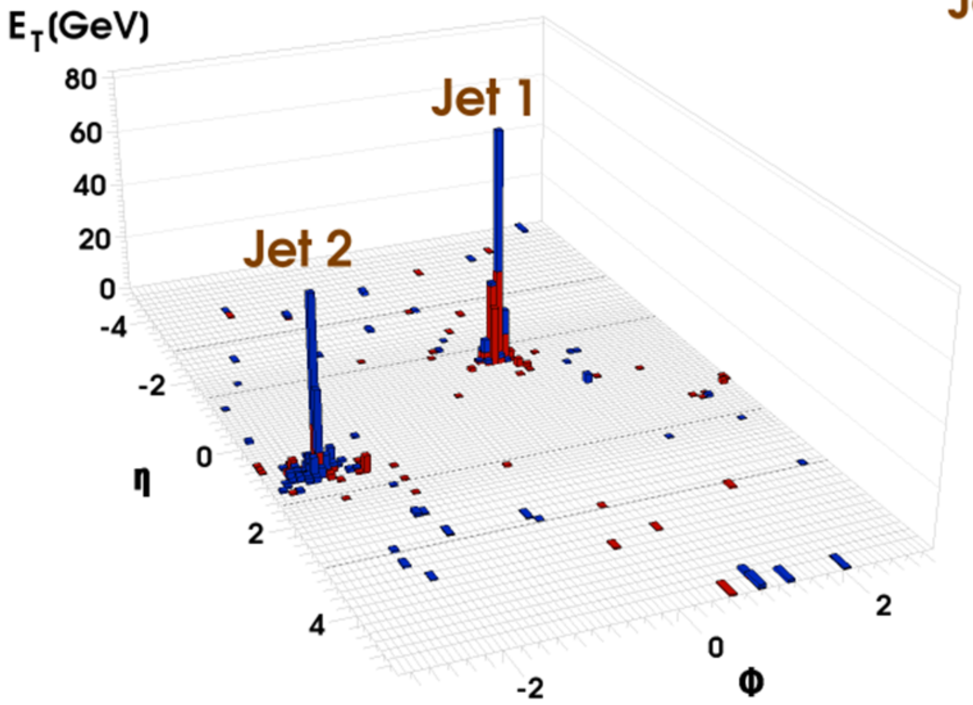
anti-K_T (R=0.7)

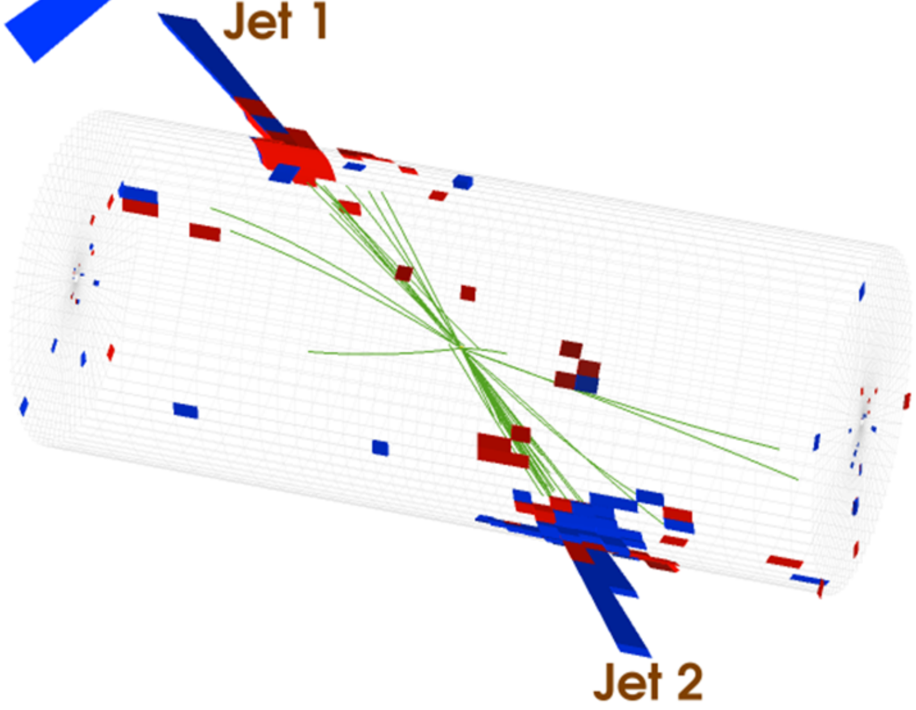
CaloJet:

- p_T(Jet1, Abs) = 253 GeV
- p_T(Jet2, Abs) = 244 GeV
- **Dijet Mass = 764 GeV**

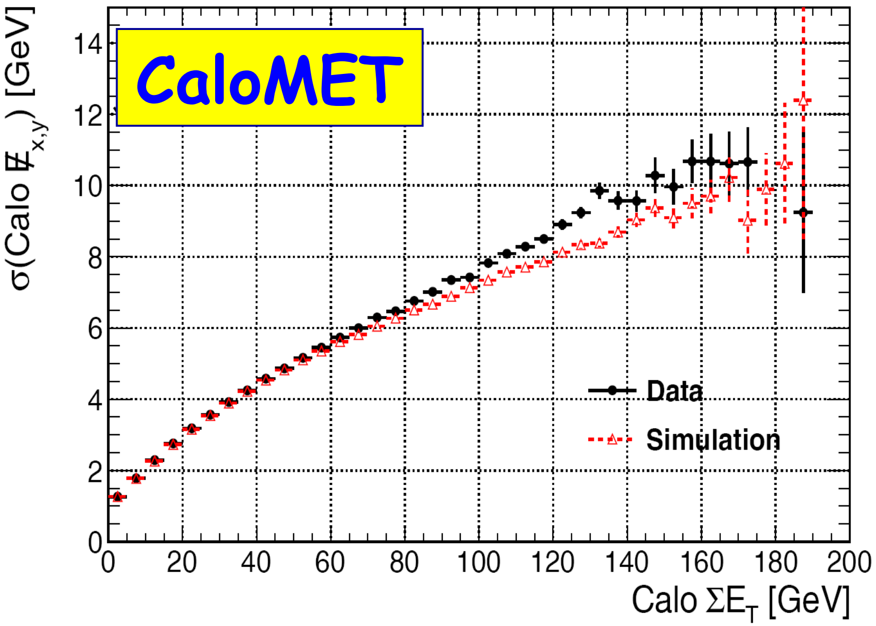


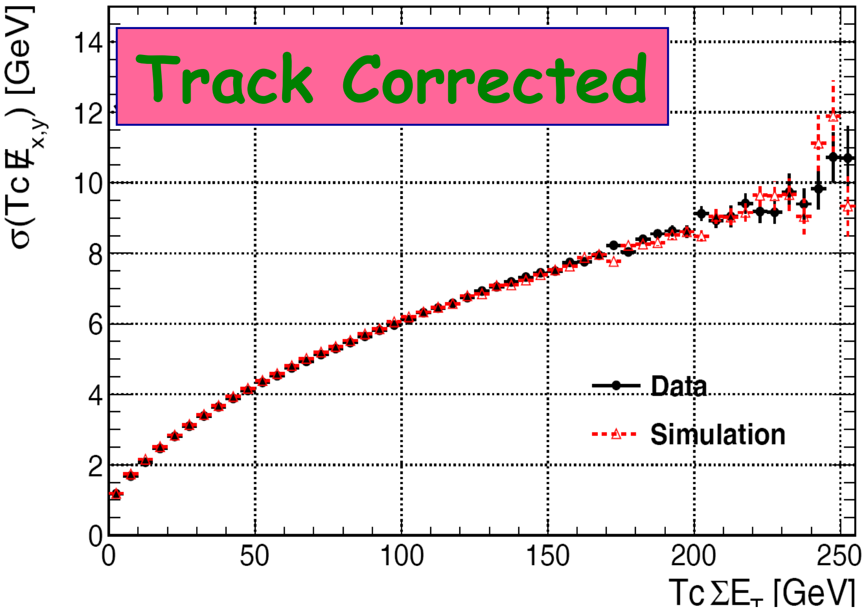






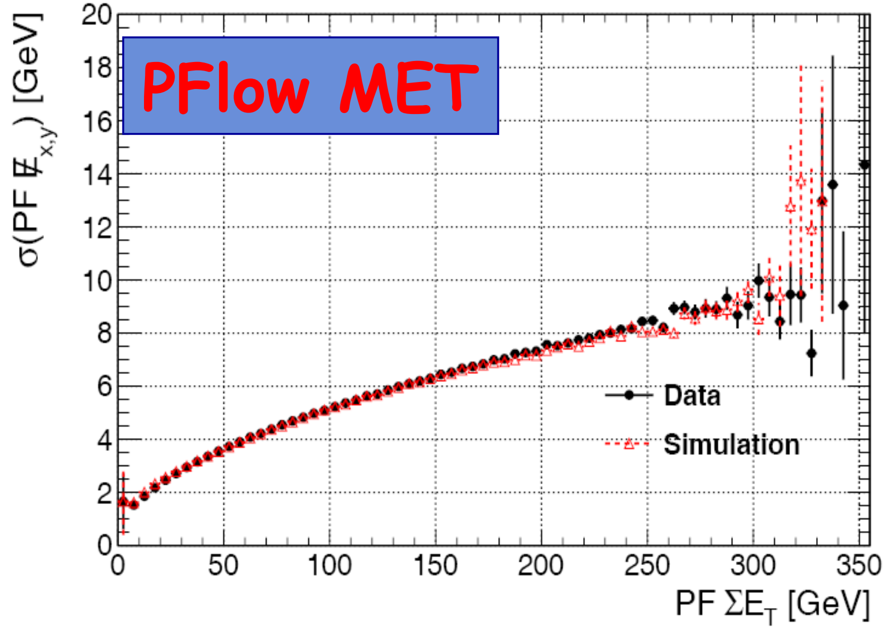
Data vs. Monte Carlo (MET performance)





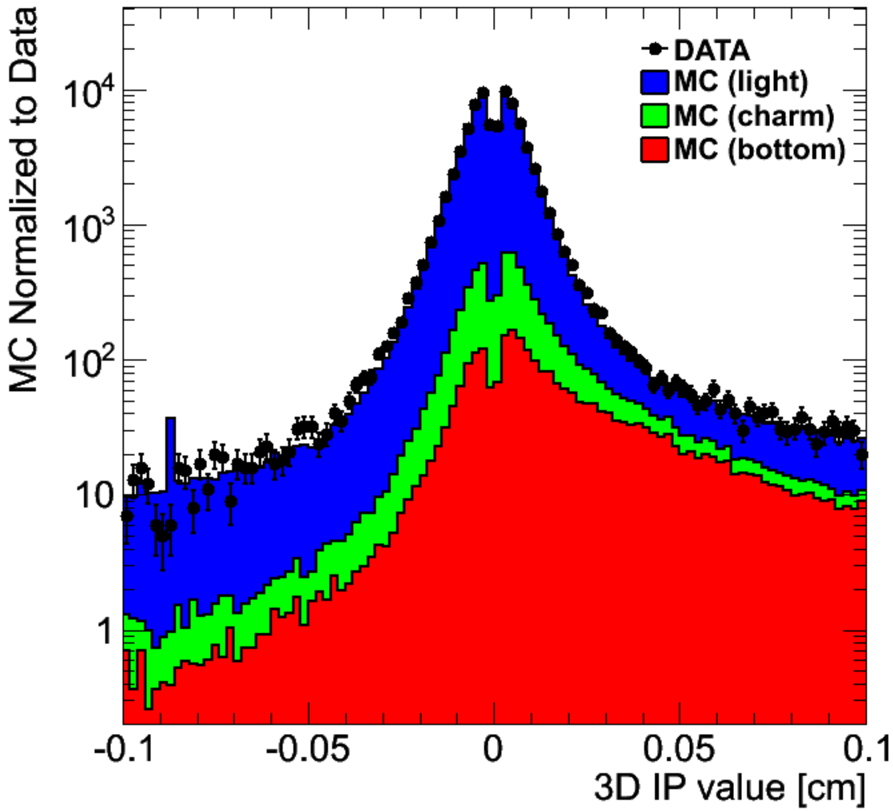
MET Significance (Dijet events):

- Minor discrepancies in the high ΣE_T region for CaloMET.
- **Astonishingly well described** track corrected and particle flow MET.

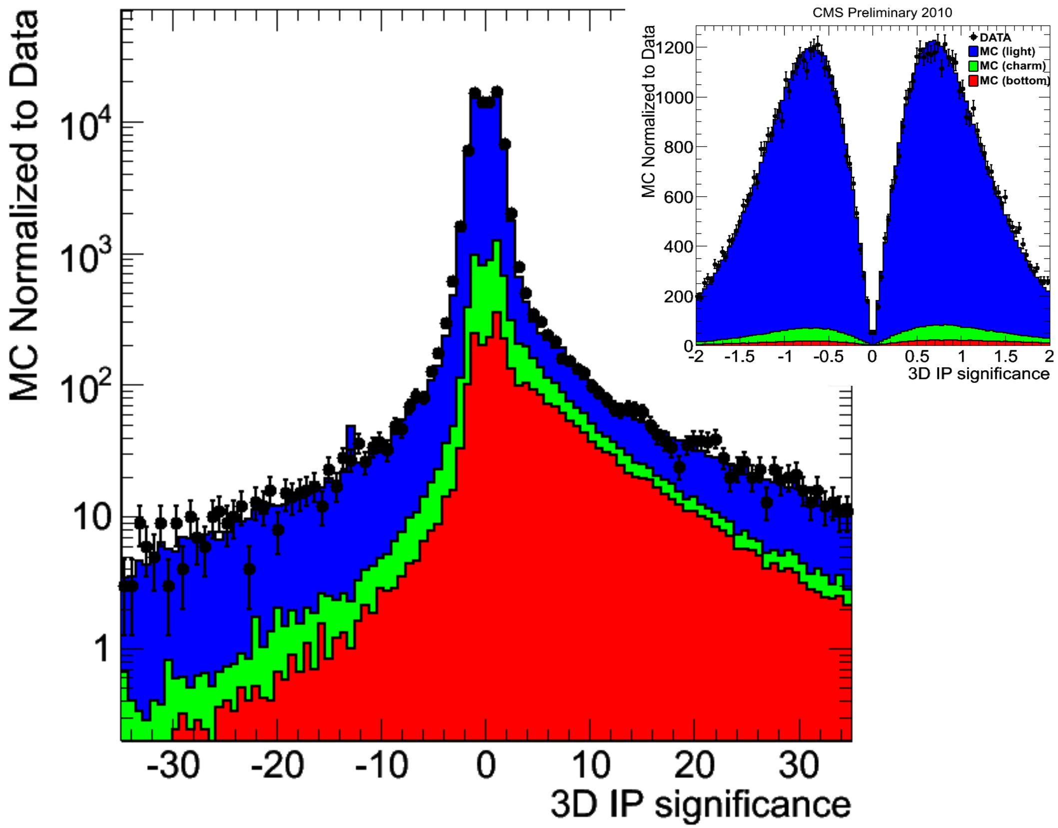


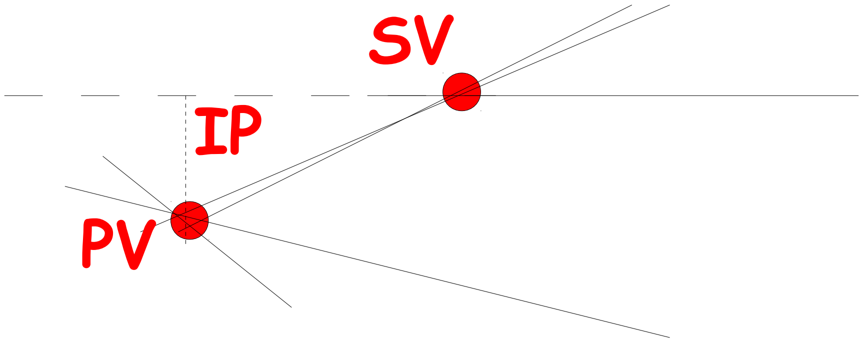
Data vs. Monte Carlo (B-Tag performance)

Impactparameter 3dim (signed)



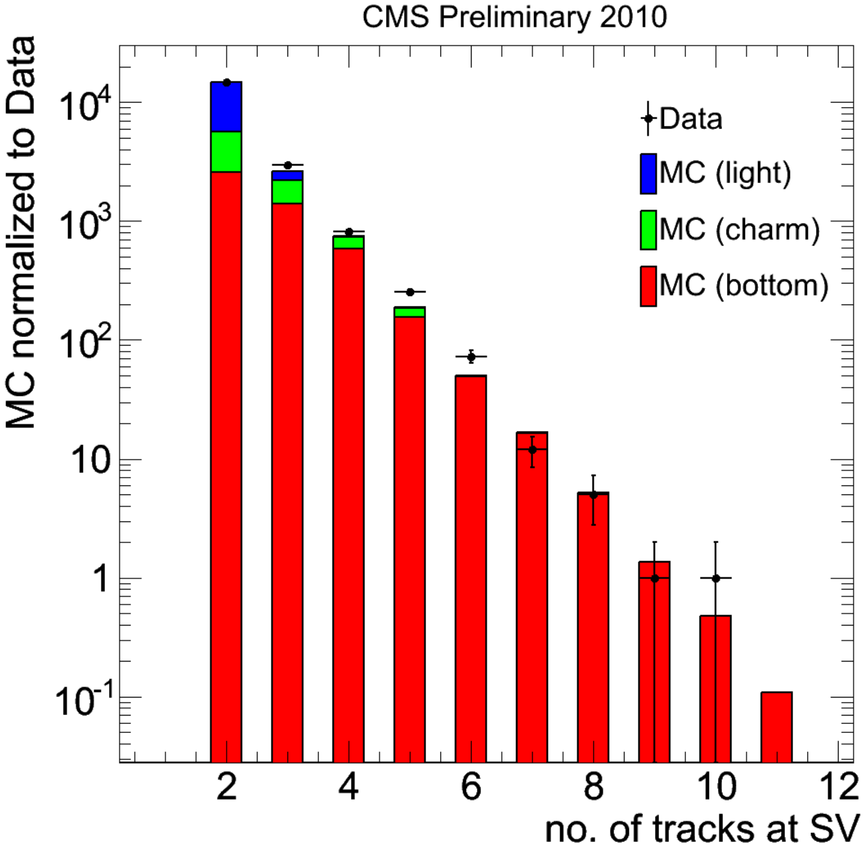
Impactparameter significance

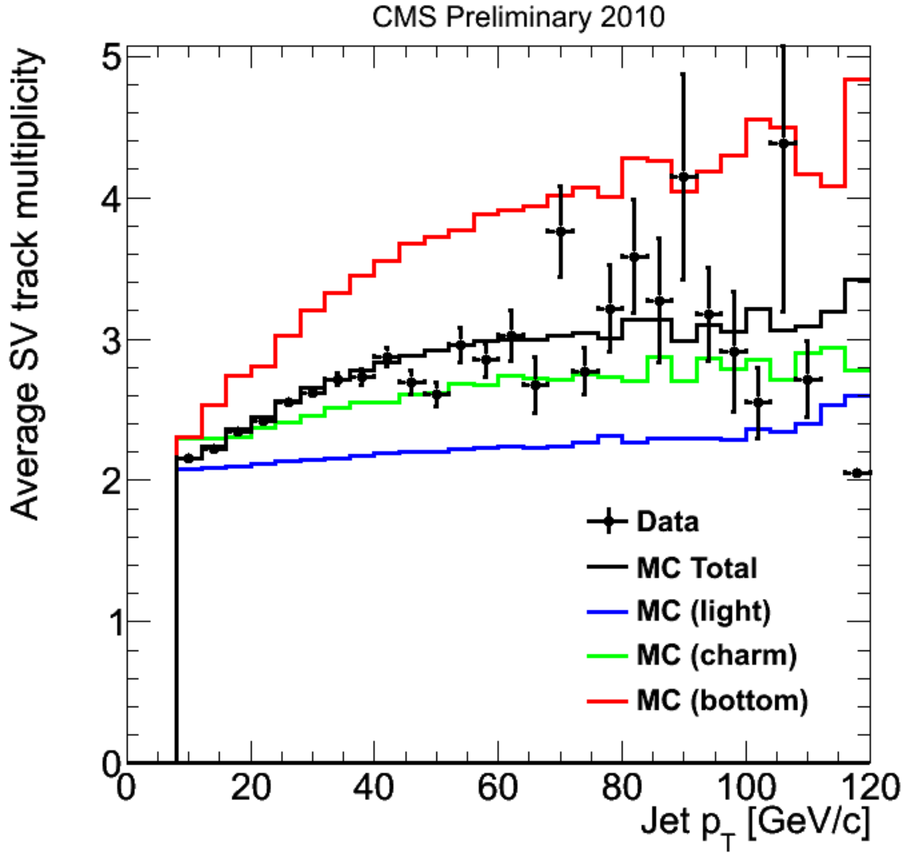




- PFlow Jets (anti-k_T (R=0.5)).
- p_T>40GeV $|\eta|$<1.5.
- Overall very good description.

Data vs. Monte Carlo (B-Tag performance)



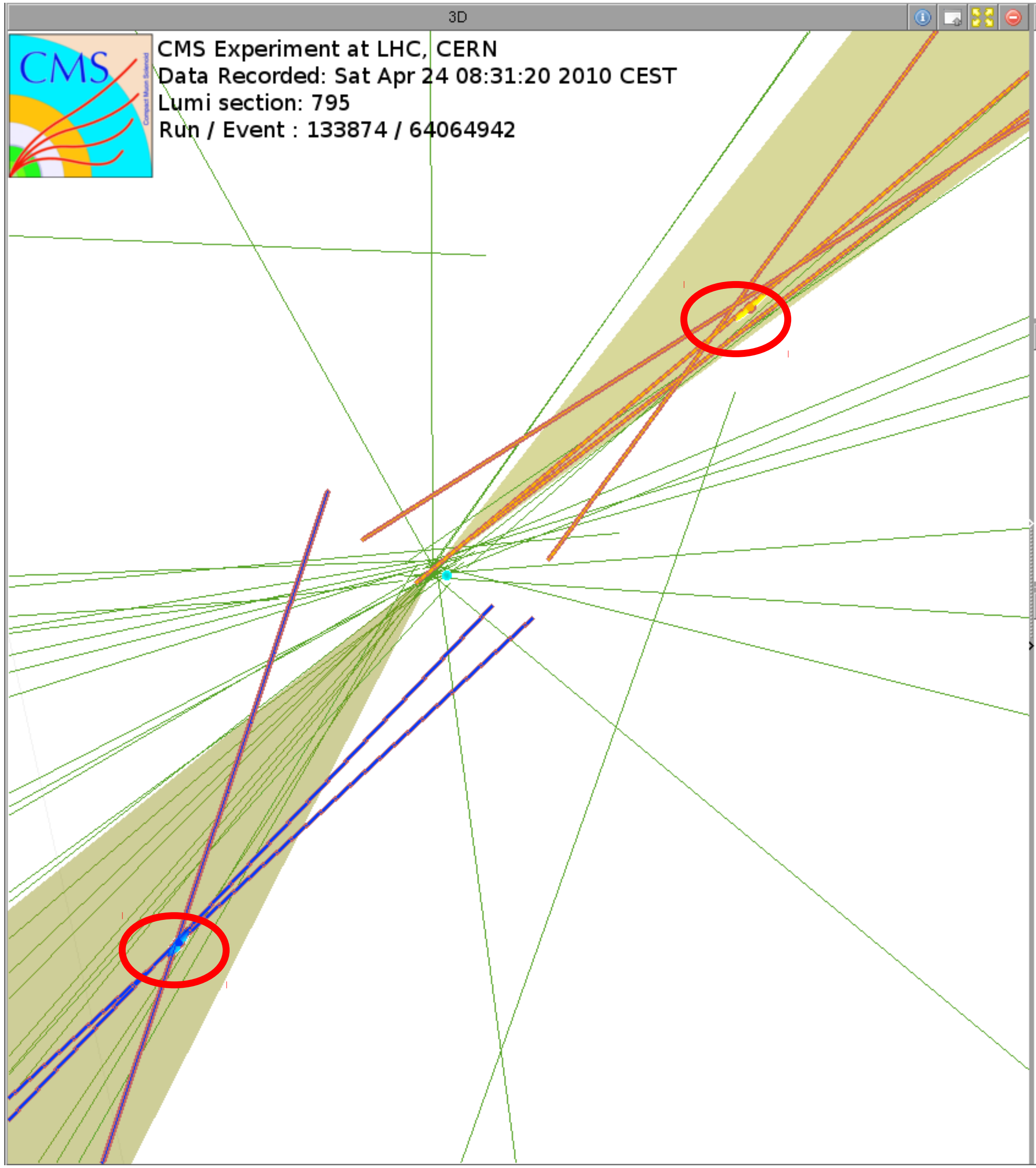


Number of tracks at secondary vertex.

Number of tracks at secondary vertex as function of jet p_T.

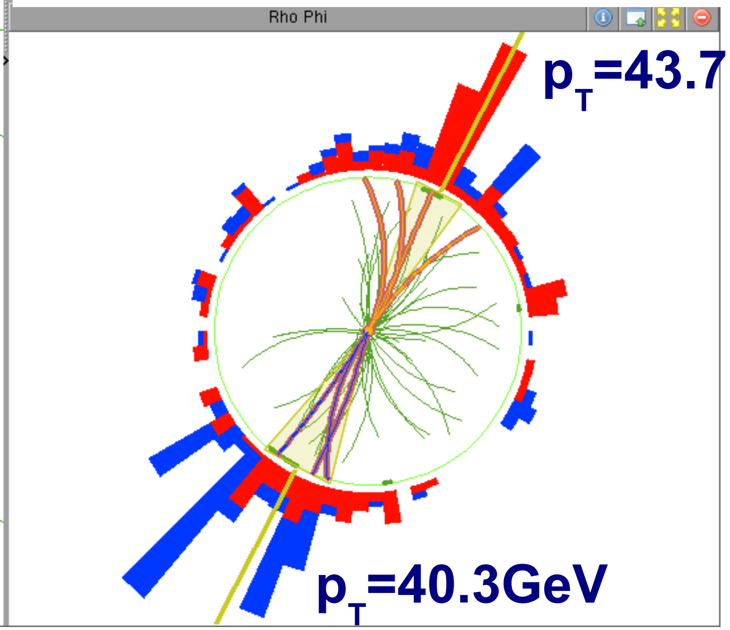
Overall very good description.

Event Display of 2 B-Tagged Jets



- PFlow Jets (anti-k_T (R=0.5)).

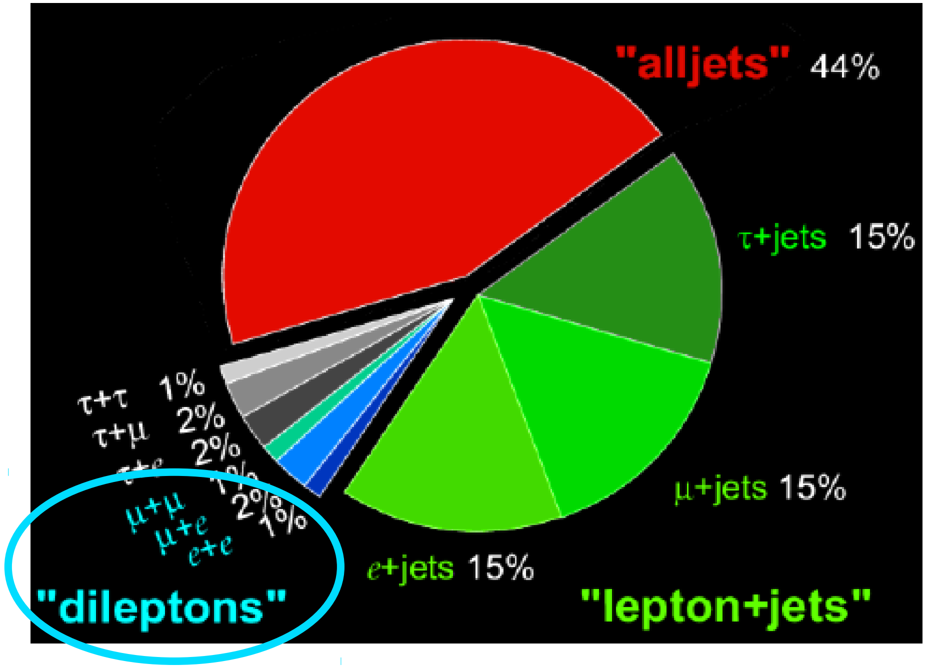
	Jet1	Jet2
dist(3dim)	6.2mm	8.6mm
σ(dist)	43	55
m_{SV}	2.9GeV	3.1GeV
p_T	25.7 GeV	17.2 GeV
χ^2/ndof	6.3/5	15.9/3



Rediscovery (Dileptonic first $10pb^{-1}$)

Clearest signature:

- 2 isolated leptons.
- Missing transverse energy.
- 2 Jets (with B-Tags).



Selection:

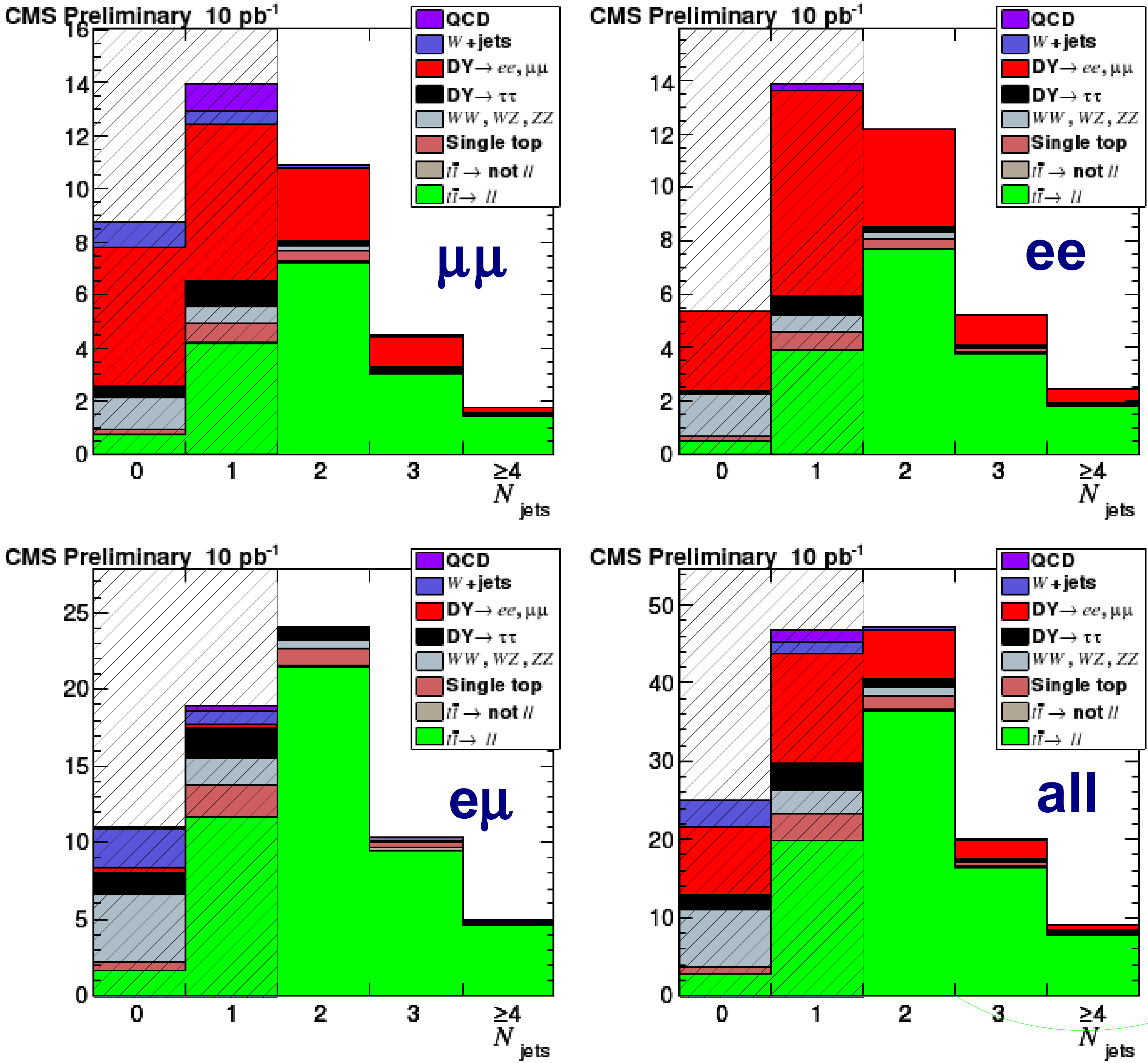
- pt(lepton)>20GeV, |η|(lepton)|<2.4.
- I_{trk}: pt/(pt + T) >0.9 ; I_{cal}: pt/(pt + C) >0.9(0.8).
- T=Σpt (all tracks in ΔR<0.3).
- C=Σpt (all calo clusters in ΔR<0.3(0.4)).
- Exclude same flavor events in M(Z)±15GeV.

- Select for events with Missing ET>30(20)GeV
- Reconstruct jets with sisCone R<0.5 (Calo/Tracks)
- Check for jet multiplicity (potentially with B-Tag)

→ Rejects 70% of all Drell-Yan events

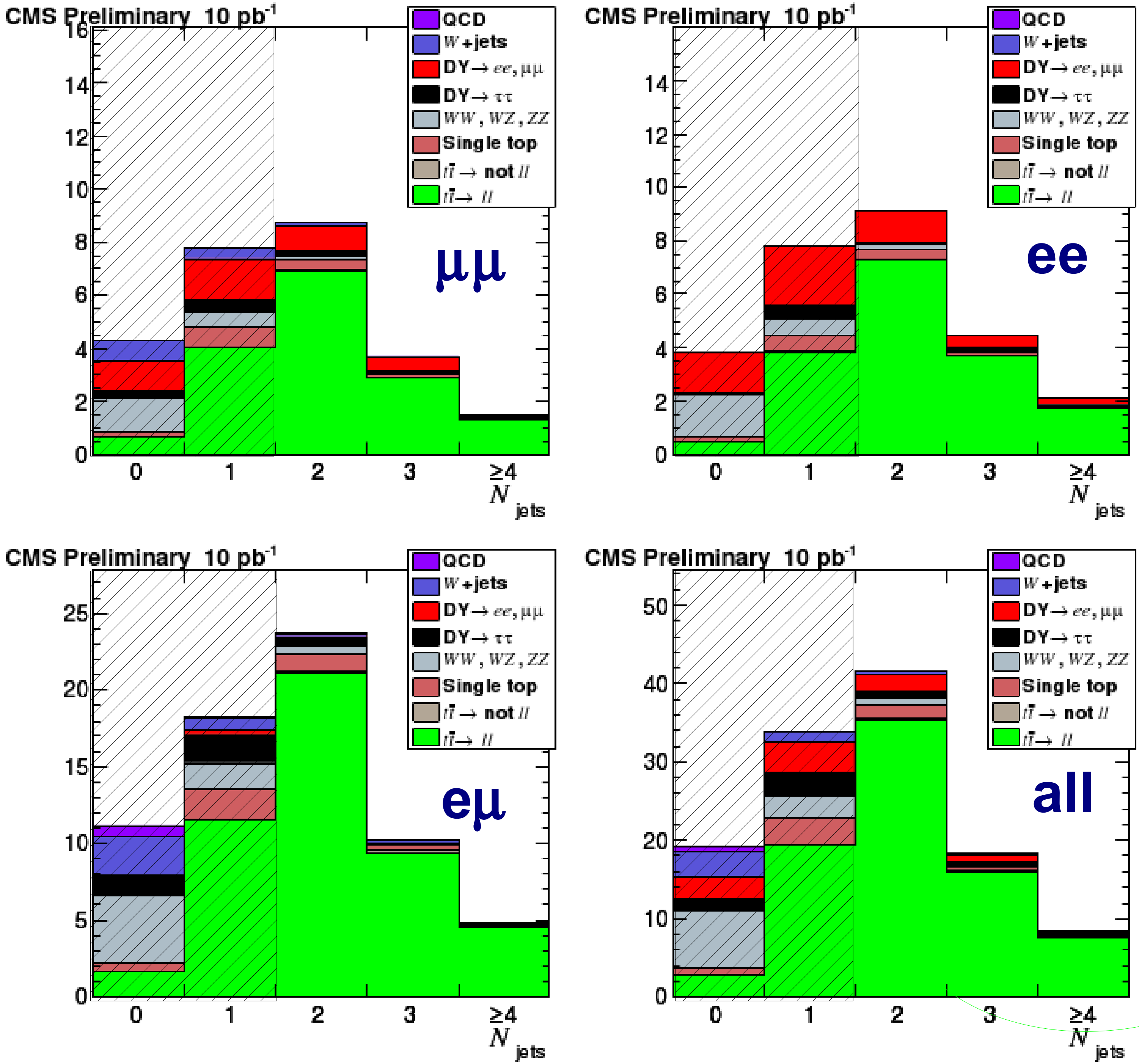
→ Find top in the 2 Jet bin

N(CaloJets)



- Cut on MET(Calo)
- DY dominates low multiplicity bin in ee or μμ
- Low multiplicity bins can serve as cross check
- eμ is already now nearly BG free
- p_T(jet, calib)>30GeV

Ttbar dilepton signal

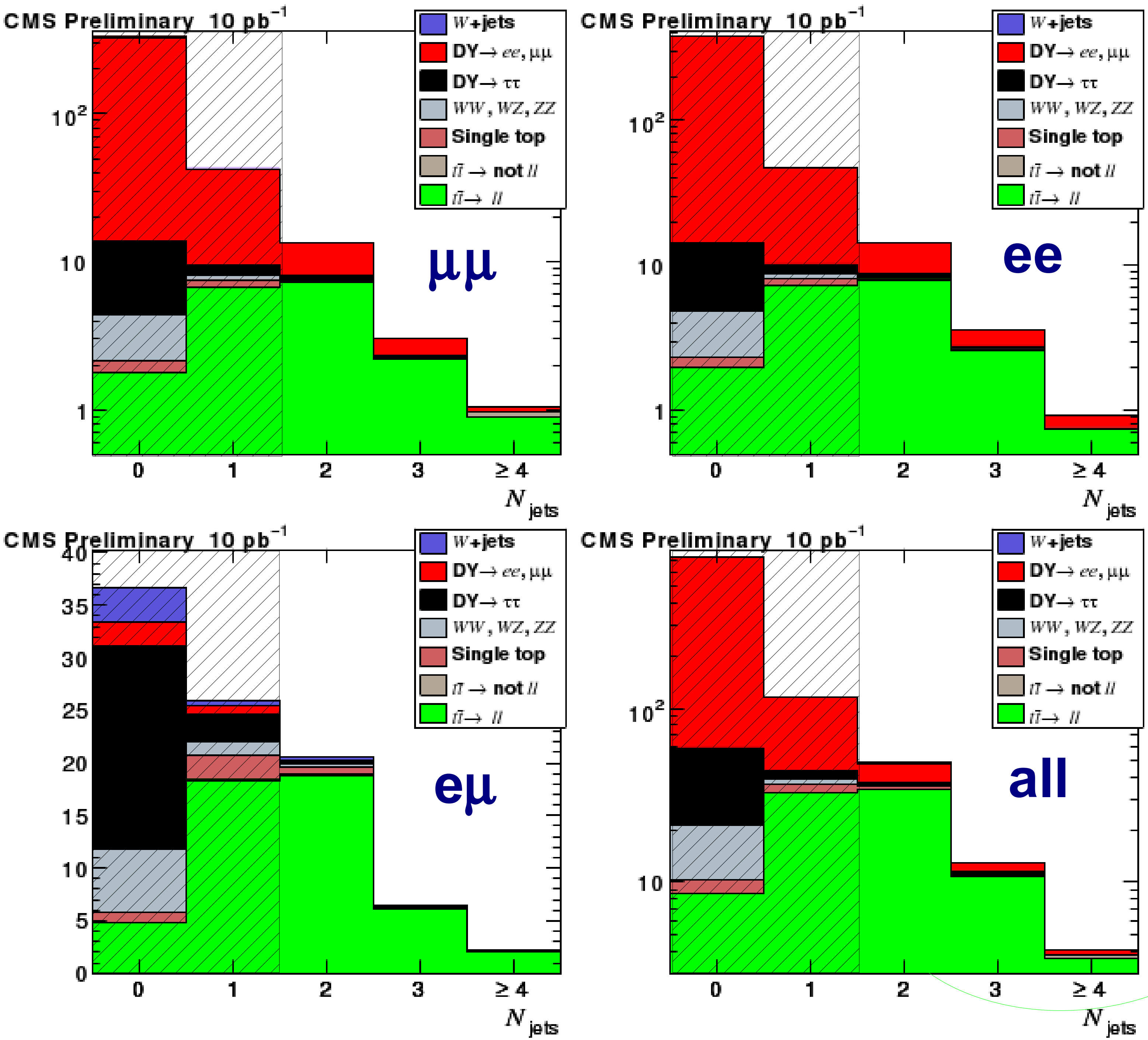


N(CaloJets)

- Cut on MET(Track)
- DY background is largely reduced in ee or μμ
- Low multiplicity bins can serve as cross check
- eμ is already now nearly BG free
- p_T(jet, calib)>30GeV

Ttbar dilepton signal

N(TrackJets)



- w/o MET cut
- Using track jets instead of calo jets
- Only calo depen-dency from lepton isoaltion!
- eμ is already now nearly BG free
- p_T(jet, calib)>30GeV

Ttbar dilepton signal

Cross Section Estimates (Dileptonic first 10pb^{-1})

Data sample	Main selection			Track-jet selection		
	e^+e^-	$\mu^+\mu^-$	$e^\pm\mu^\mp$	e^+e^-	$\mu^+\mu^-$	$e^\pm\mu^\mp$
$t\bar{t} \to \ell\ell$	11.6 ± 0.2	13.2 ± 0.2	35.6 ± 0.4	10.4 ± 0.2	11.3 ± 0.2	26.7 ± 0.4
other $t\bar{t}$	0.21 ± 0.03	0.04 ± 0.01	0.46 ± 0.04	0.4 ± 0.04	0.05 ± 0.02	0.43 ± 0.04
Single top	0.46 ± 0.03	0.56 ± 0.03	1.40 ± 0.06	0.32 ± 0.03	0.35 ± 0.03	0.85 ± 0.04
$WW/WZ/ZZ$	0.26 ± 0.02	0.33 ± 0.03	0.71 ± 0.05	0.15 ± 0.02	0.2 ± 0.02	0.28 ± 0.02
DY$\to \tau\tau$ + jets	0.3 ± 0.1	0.3 ± 0.1	0.7 ± 0.2	0.1 ± 0.06	0.15 ± 0.07	0.22 ± 0.08
DY$\to ee/\mu\mu$ + jets	4.1 ± 0.4	5.3 ± 0.4	0.08 ± 0.05	6.1 ± 0.5	6.7 ± 0.6	0.19 ± 0.07
W + jets	0.2 ± 0.1	< 0.1	0.3 ± 0.1	0.2 ± 0.1	< 0.1	0.3 ± 0.1
QCD	< 1	< 0.4	< 0.4	< 1	< 0.4	< 0.4
Total backgrounds	5.5 ± 0.4	6.6 ± 0.4	3.7 ± 0.2	6.9 ± 0.5	6.9 ± 0.6	1.6 ± 0.2
Data driven fakes	1.1 ± 0.6	0.8 ± 0.4	2.5 ± 1.2	1.3 ± 0.6	0.5 ± 0.2	1.9 ± 1.0
Data driven DY	4.0 ± 1.3	5.1 ± 1.6		6.6 ± 2.2	7.2 ± 2.4	

Rel . uncertainties 10pb^{-1}

Source	e^+e^- and $\mu^+\mu^-$	$e^\pm\mu^\mp$
Statistical	25	18
Lepton ID	5	5
Lepton isolation	3	3
Jet energy scale	8	5
Theory	4	4
DY$\to ee,\mu\mu$ method	10	
Fake leptons method	4	4
Residual background	5	4
Integrated luminosity	10	10

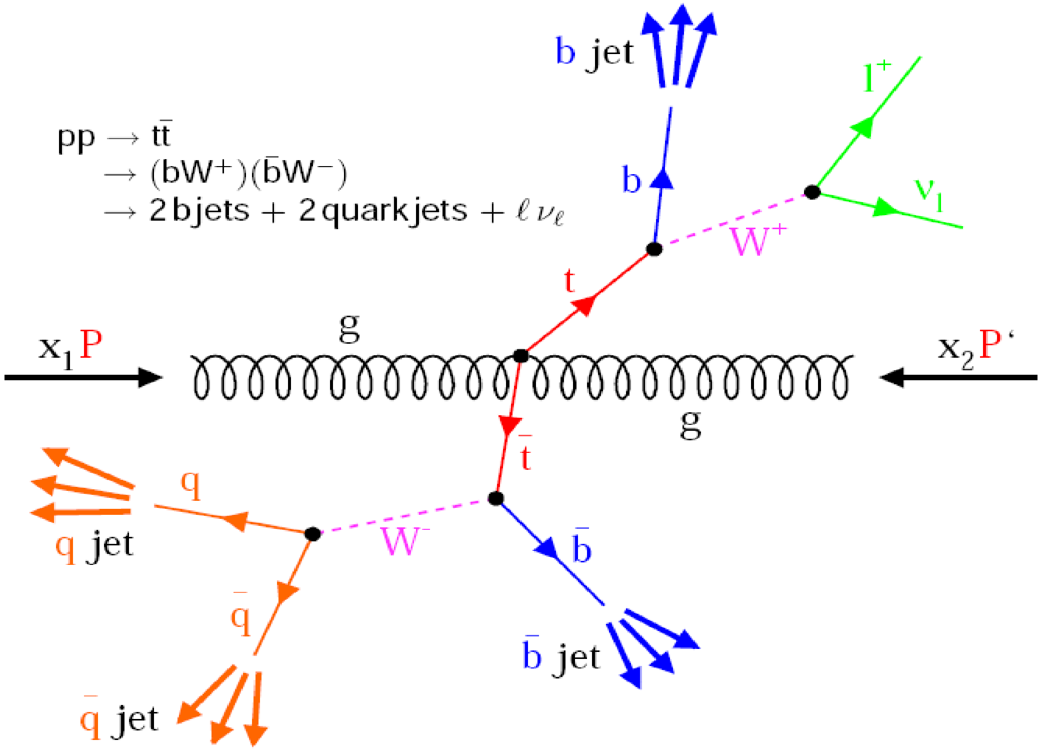
in %

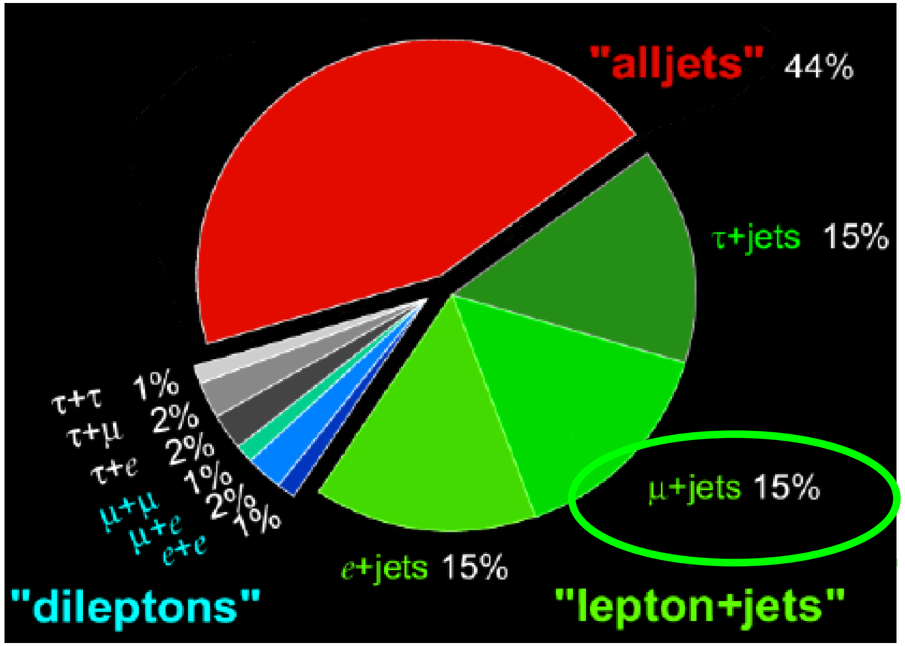
With 10pb^{-1} (ee, μμ, eμ): S/B~4, **15%(stat)+10%(sys)+10%(lumi).**

Rediscovery (μ+Jets first 20pb^{-1})

Signature:

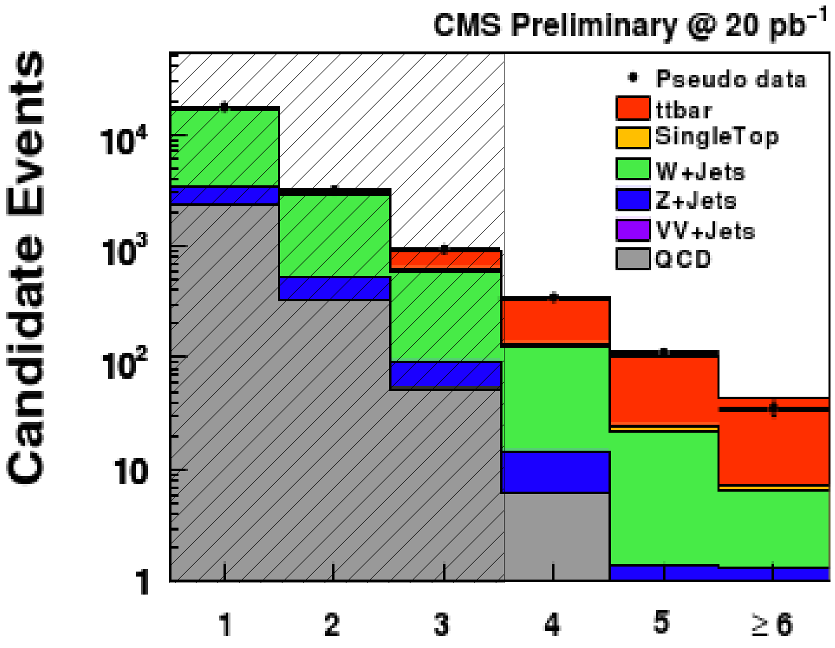
- 1 isolated lepton
- Missing transverse energy
- 4 Jets (among those 2 jets from b-Quarks)

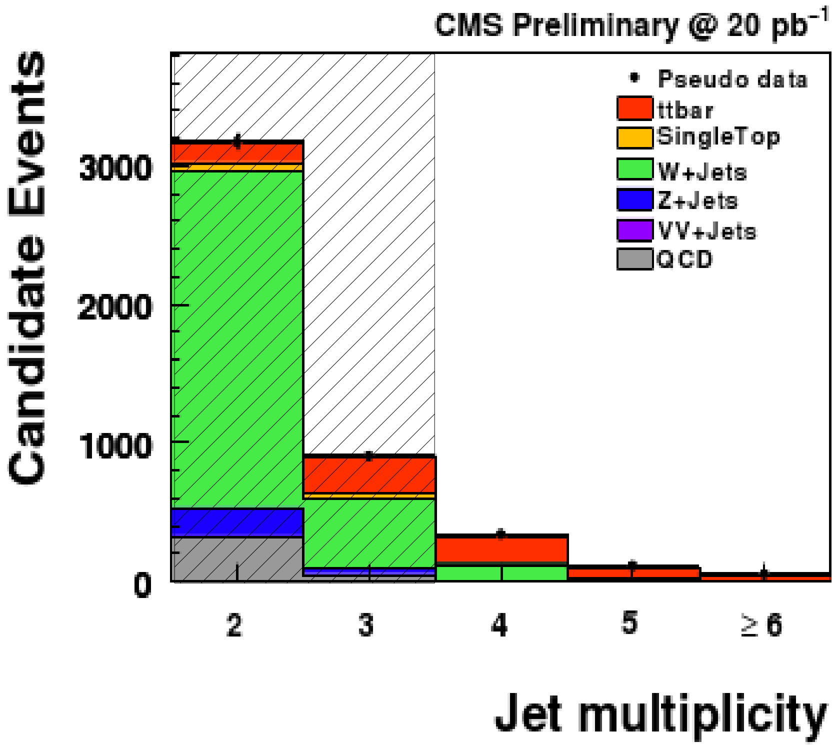




Event Selection

- pt(μ)>20GeV, |η|(μ [e])|<2.1[2.4]
- I_{trel}: (C + T)/pt >0.05
- T=Σpt (all tracks in ΔR<0.3)
- C=Σpt (all calo clusters in ΔR<0.3[0.4]) [1)]
- Missing transverse energy MET>30GeV
- Veto on a second μ with looser selection (pt>10GeV, I_{re}l<0.2)
- Veto on an electron with looser selection (pt>10GeV, I_{re}l<0.2)
- Jets are reconstructed using sisCone with ΔR<0.5
- Require at least 4 Jets with pt(calib)>30GeV (in tracker acc.)





[1)] 0.4 for the hadronic calorimeter

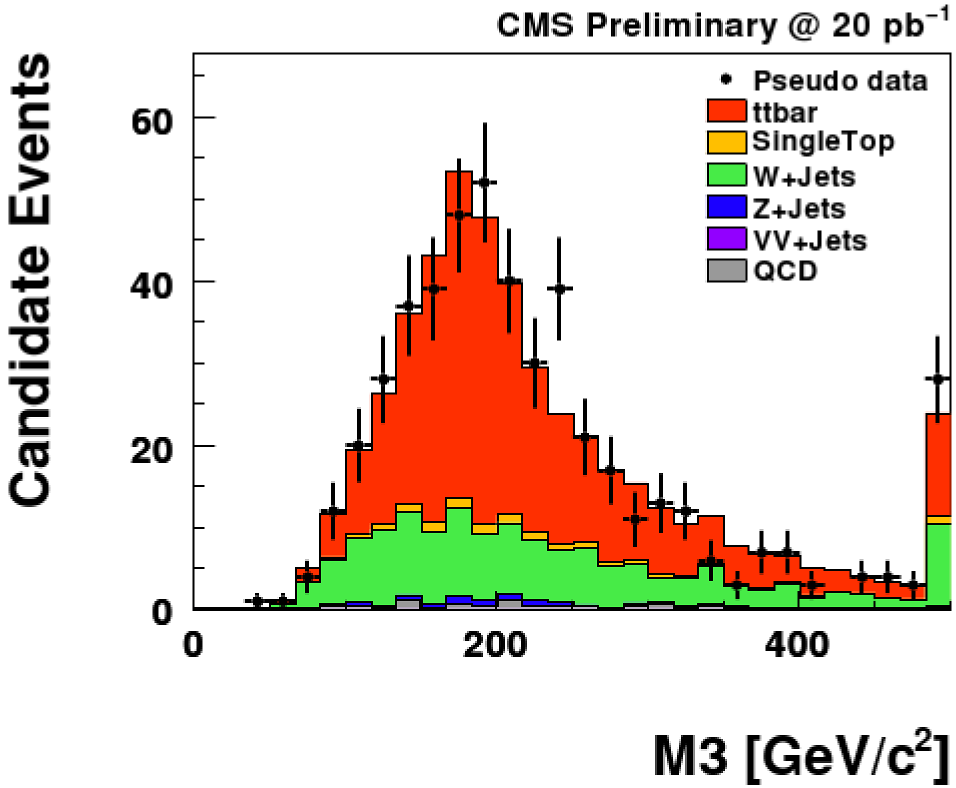
Signal Estimate

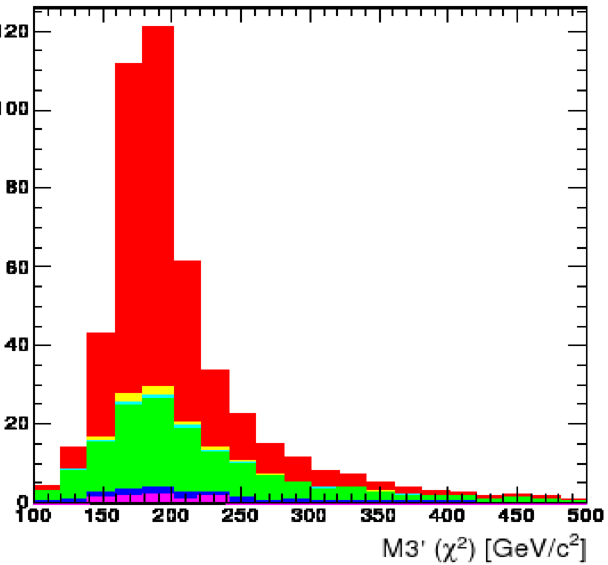
Make use of distribution(s) where signal and BG differ:

M3: inv. mass of 3 jets with largest vectorial sum pt

M3': inv. mass of 3 jets which have best χ^2 on mass hypotheses (χ^2 counting method)

$\eta(\mu)$: η distribution of isolated muon (more central for signal events)





$$\chi^2 = \frac{(m_{j_1 j_2} - m_W)^2}{\sigma^2_{jj}} + \frac{(m_{j_1 j_2 j_3} - m_t)^2}{\sigma^2_{jjj}} + \frac{(m_{\mu\nu j_4} - m_t)^2}{\sigma^2_{\mu\nu j}}$$

$\sigma(jj)\sim 10$GeV $\sigma(jjj)\sim 20$GeV $\sigma(jl\nu)\sim 24$GeV

Cross Section Estimates (μ+Jets first 20pb^{-1})

Source	Uncertainty [%]		
	Fit to $\eta(\mu)$	Fit to M3	Fit to M3′
Statistical Uncertainty (20 pb^{-1})	17.7	16.3	11.5
Jet Energy Scale	16.7	15.1	19
$t\bar{t}$ MC Generator	1.9	14.9	14
$t\bar{t}$ ISR/FSR	3.3	7.7	2
W+jets Factorization scale	4.4	4.7	4
W+jets Matching threshold	5.5	2.8	4
Single Top Shape	0.1	0.8	1
PDF Uncertainty	5.0	5.0	5.0
Total Systematic Error	19.2	23.8	25.0
Luminosity Error	10.0	10.0	10.0

Rediscovery & First Measurements

- Solid planning for first measurements for ICHEP (or "near in time").
- Unfortunately these results are too fresh to be allowed to be shown.
- We have strong development in Monitoring incoming Top Events (quasi online).
- We are sitting in our blocks to hunt for the first top candidates!
- With (10-20)pb^{-1} we expect a rediscovery or first measurements with:
  ± (10-17)% (stat) ± (15-20)% (sys) ± 10% (lumi)

Summary

- LHC will be a factory of top quarks (comp. to BarBar and Belle for b Quarks).
- The rediscovery and measurements with Top Quarks rely on a good understanding of jets.
- We have a highly redundant set of jets that will help us to proof a good & a very good understanding of our calorimeters & detector.
- TOP2010 is unfortunately too early for first Top Quarks at the LHC.
- Nevertheless the comparisons between Data and Monte Carlo show an astonishingly good agreement!
- These results show that CMS is very well prepared for high lumi data!

Summary

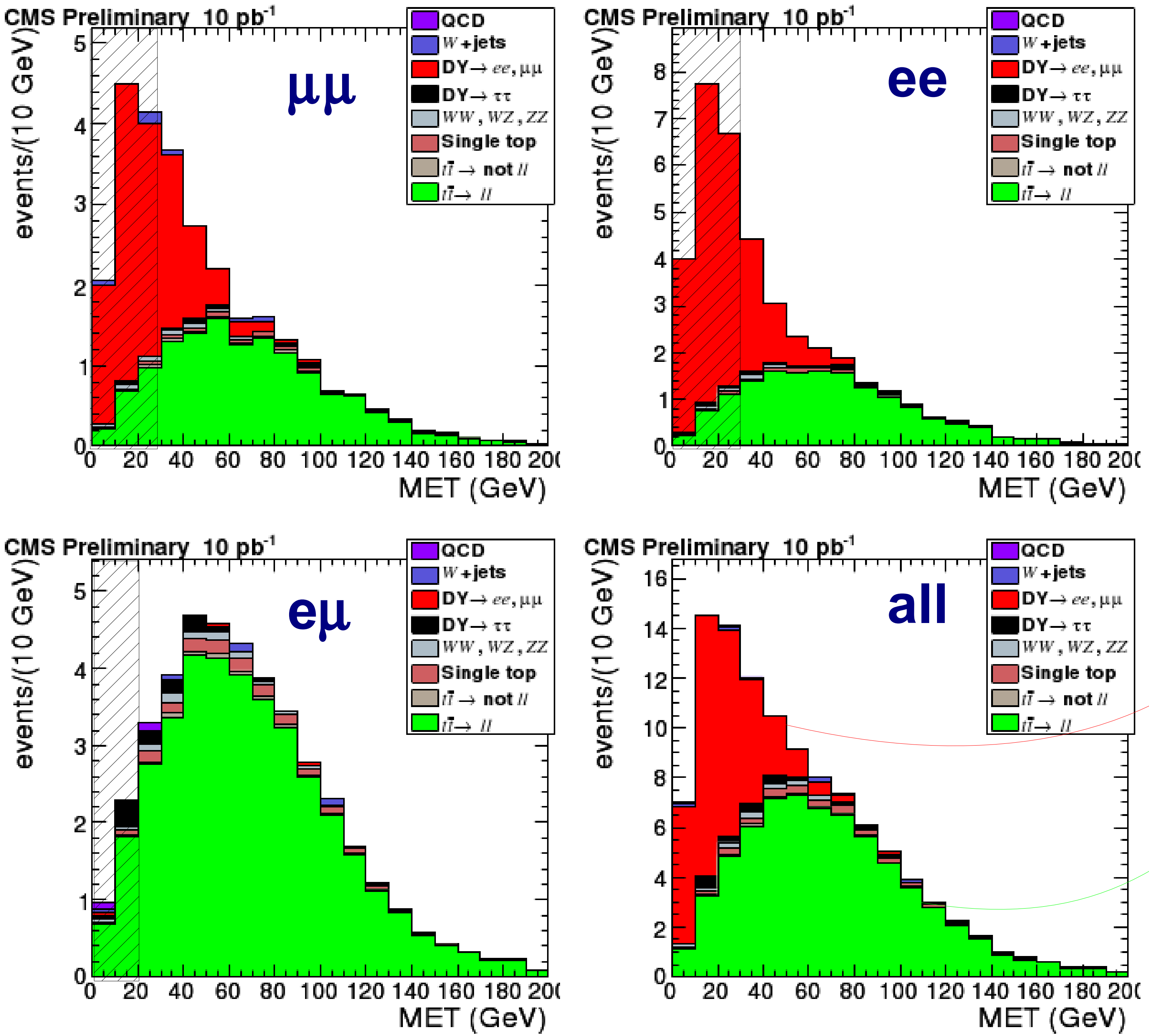
- We had already nice and very interesting events in our data!
- The number of events we see is in very good agreement with what we expect!
- We are very much looking forward to the first top events & measurements!

Many thanks for your attendance

Backup

MET

- 30 GeV for same flavor (to reject DY)
- 20 GeV for unlike sign flavor (to reject QCD)
- Corrected for μ's & for jet corrections

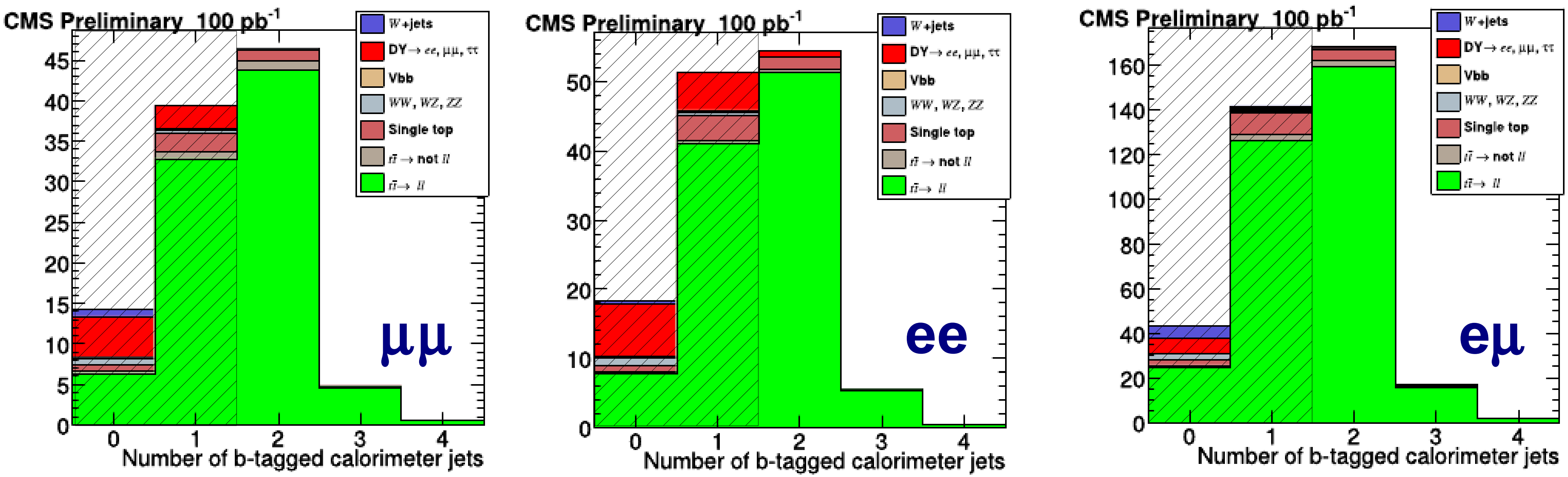


Drell-Yan

Ttbar dilepton signal

From 10pb^{-1} to 100pb^{-1}

- Exploit simple B-Tag algorithm based on lifetime of b-Quarks
  - Choose track with second largest significance in displacement from primary vertex
  - Working point here: mistag rate of 10%
  - MET > 50(30) GeV for ee/μμ (eμ)



Data Driven Methods

- Estimation of DY contribution:
  - Fix normalization outside of Z-Mass peak and extrapolate using the extrapolation factor from MC
  - **30% uncertainties assigned**

- Estimation of isolated lepton fake rate:
  - Fake isolated leptons: hadronic jets, which are misidentified as isolated leptons
  - First of all: expected to be smaller than 10%
  - Define looser isolation selection [ref] (just to define the kind of objects we are talking about)
  - Estimate FR = sel/ref

    sel := passed isolation cuts

    fail := failed isolation cuts

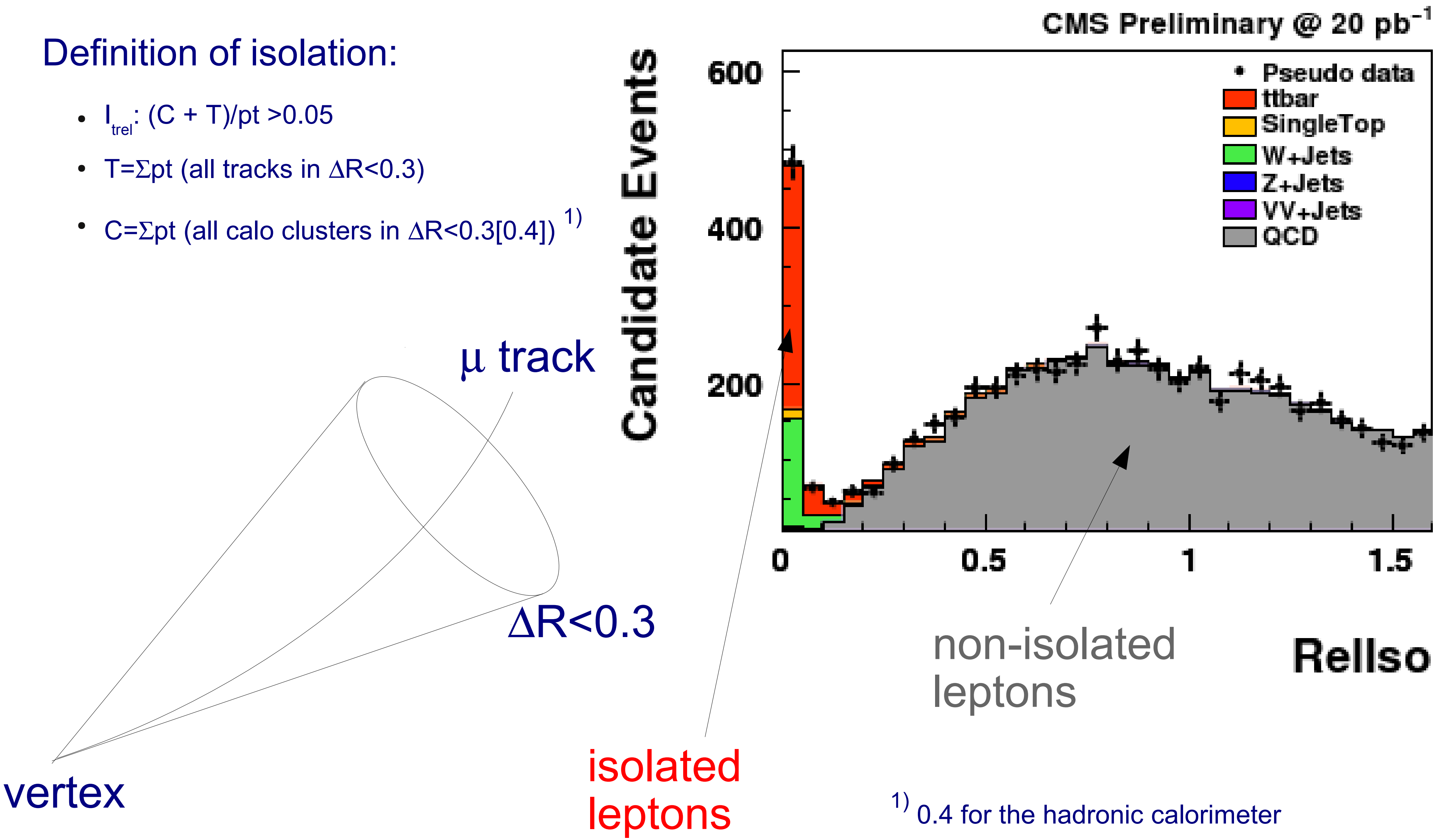
    ref := reference sample (ref=sel+fail) fail/ref = 1-FR

  - Determine RF on BG dominated sample (e.g. passing jet triggers) and extrapolate to signal like events
  - Scale those events in the signal region that fail the isolation by FR / (1-FR) to estimate the number of fakes in the signal region
  - Assumption: FR is the same for signal and background like events (**50% uncertainties assigned**)

μ+Isolation

Definition of isolation:

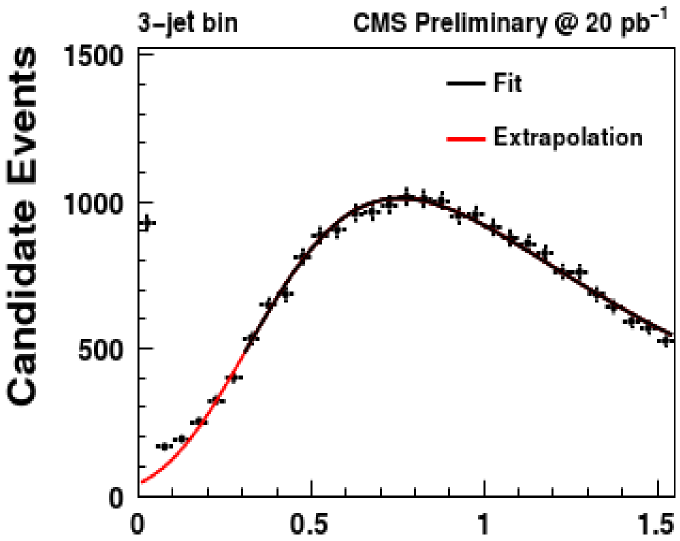
- I_{trel}: (C + T)/pt >0.05
- T=Σpt (all tracks in ΔR<0.3)
- C=Σpt (all calo clusters in ΔR<0.3[0.4]) [1)]

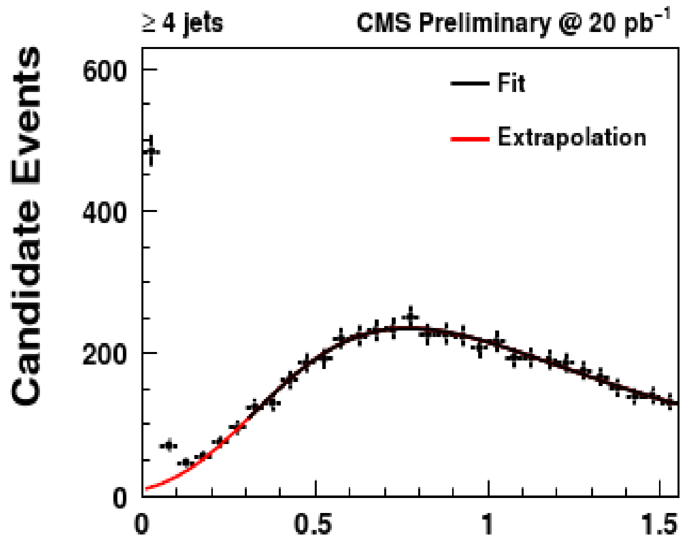


1) 0.4 for the hadronic calorimeter

Data Driven QCD Background Estimates

Extrapolation Method:



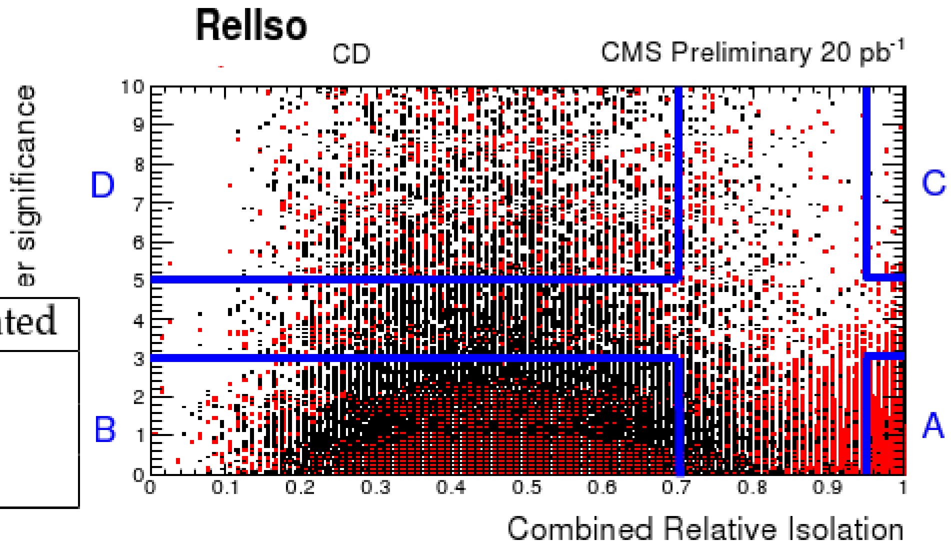


associated uncertainties <40%

Jets	N(QCD) Predicted	N(QCD) Estimated
2	327	378 ± 82
3	53	47 ± 24
≥ 4	7	13 ± 7

Classic ABCD method:

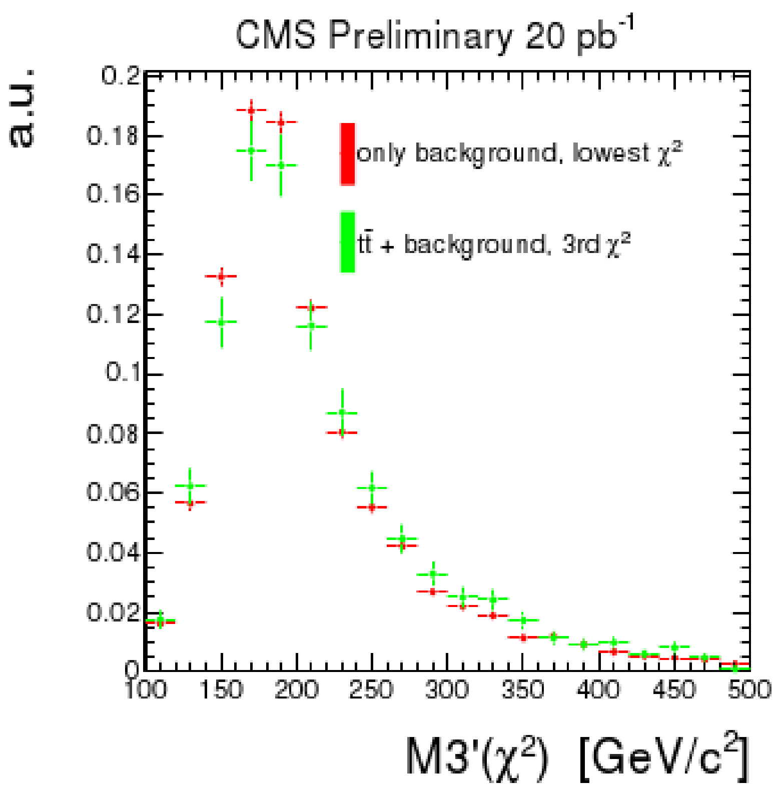
Jets	N(QCD) Predicted	N_B	N_C	N_D	N(QCD) Estimated
2	327	86625	61	16240	325 ± 26
3	53	24216	10	5058	48 ± 9
≥ 4	7	5345	3	1148	12 ± 5



Data Driven Combinatorics & BG Estimates

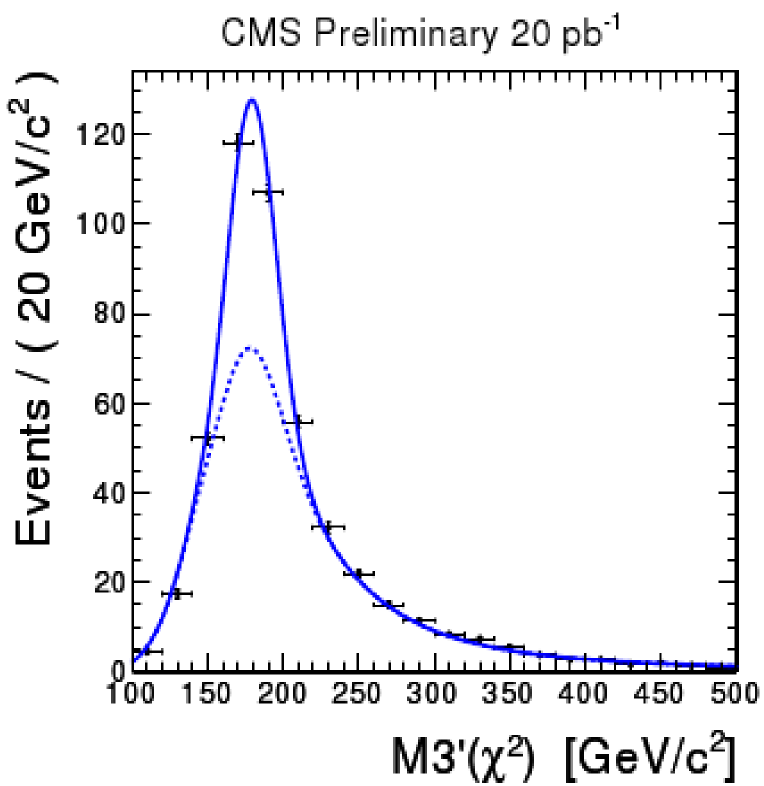
χ^2 Method:

use combinations with larger values of χ^2 to estimate combinatorics and remaining physics BG



Template Method:

use fit of BG and wrong combinations template from MC plus a Gaussian for the signal



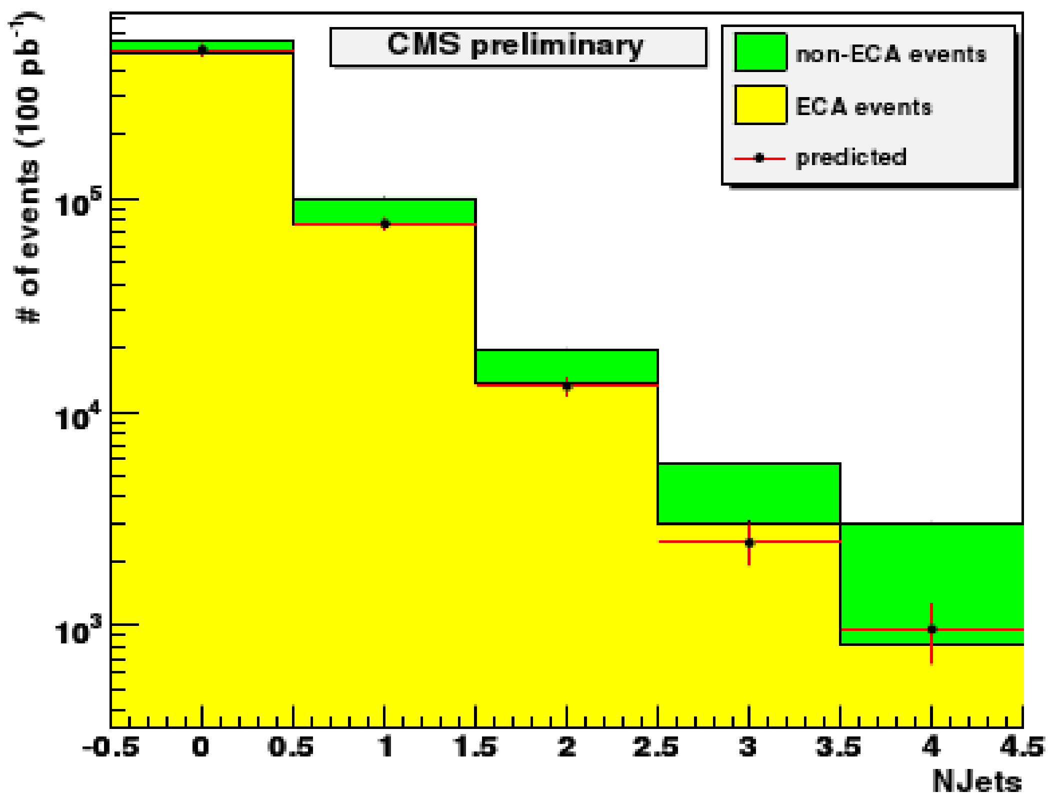
Data Driven W+Jets Background Estimate

Charge Assymmetry Method:

Make use of differences in cross section and acceptance of W^+/W^- (efficiencies should be the same to first order)

$$R_{\pm}(W) = \frac{N_{W^+} + N_{W^-}}{N_{W^+} - N_{W^-}} = \frac{A_+\sigma_{W^+} + A_-\sigma_{W^-}}{A_+\sigma_{W+} - A_-\sigma_{W-}}$$

$$N_+ + N_- = R_{\pm}(W)\,(N_+ - N_{--})$$



Data vs. Monte Carlo (MET performace)

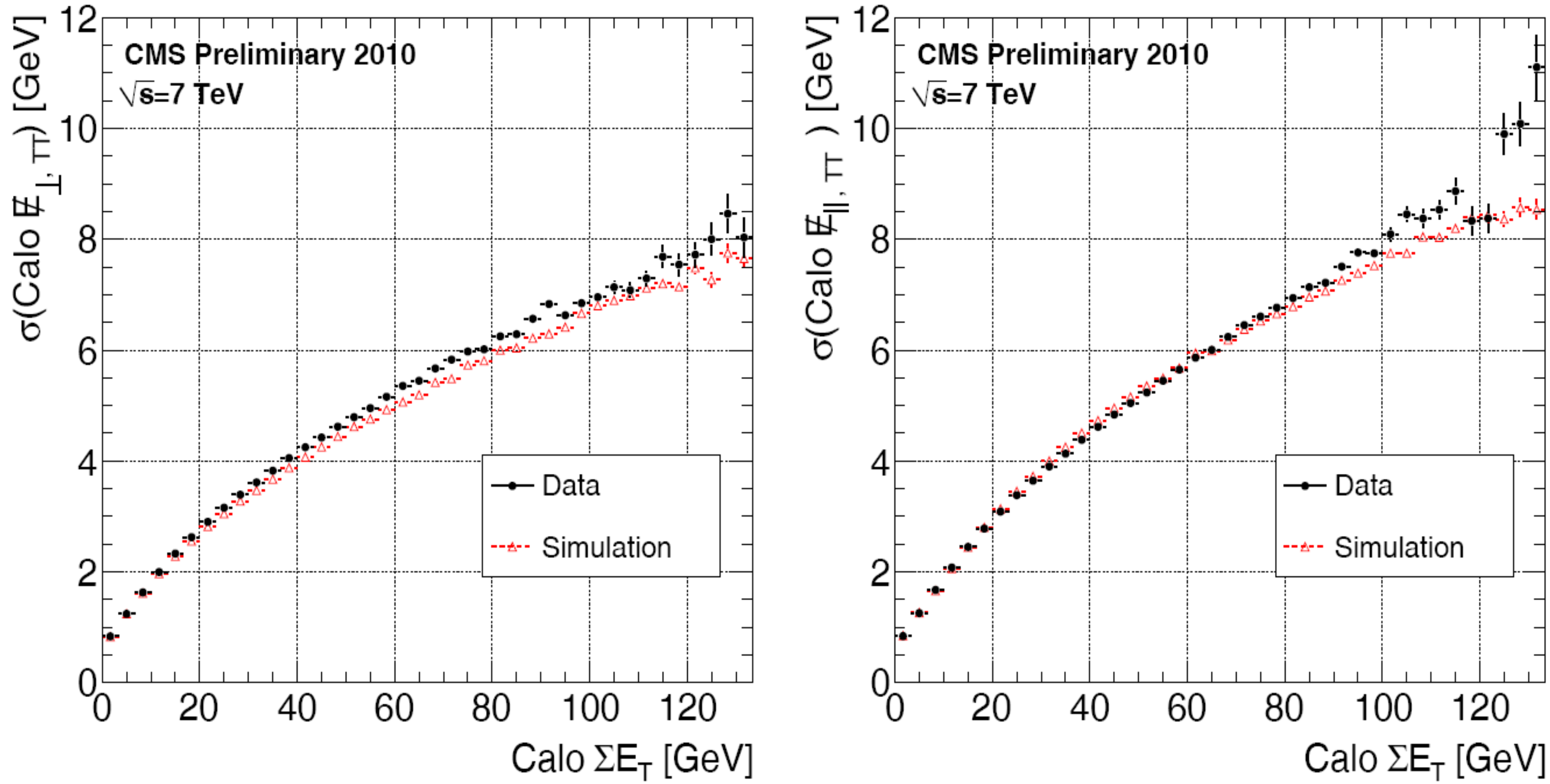


Figure: Data vs MC: MET Decomposition $\not{E}_{\perp}$ and $\not{E}_{\parallel}$ distributions