Background estimation strategies in CMS

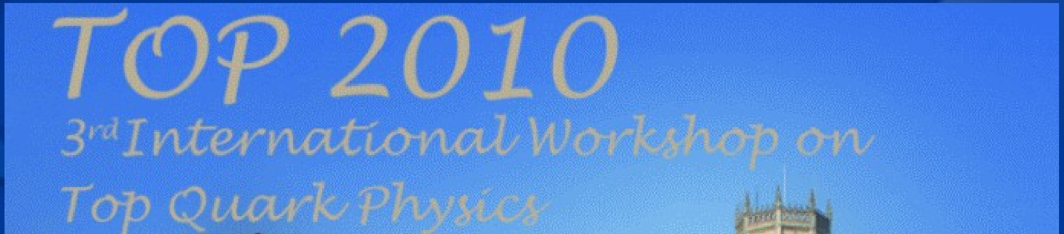
Matthew Barrett

Experimental Particle Physics Group, Centre for Sensors and Instrumentation,

Department of Electronic and Computer Engineering, School of Engineering and Design, Brunel University

On behalf of the CMS collaboration

Presented at:

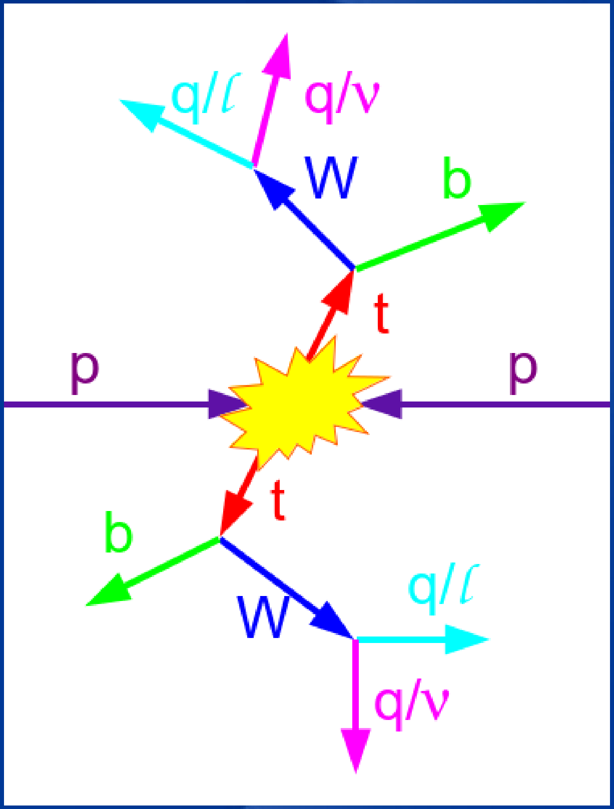


Outline

- Introduction.
- Lepton+jets mode:
  - Selecting top events and remaining backgrounds.
  - QCD background.
  - W/Z+jets background.
- Dilepton mode:
  - Selecting dileptonic top events.
  - DY+jets background, and fake leptons.
- Summary.

Introduction

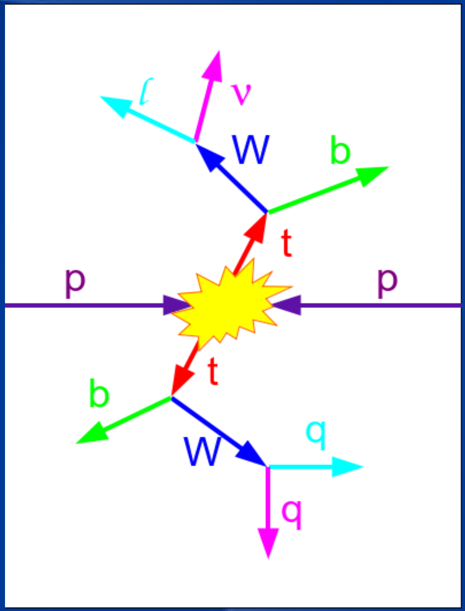
- For early data dileptonic, and lepton+jets are the top top topologies to look at.
- Methods developed with MC to study backgrounds:
  - still to be tested on data...
  - Multiple methods to cross-check.
  - robust for early data.
- Figures and tables shown are for 20pb^{-1} (lepton+jets), and 10pb^{-1} (dileptonic), except where noted.
- All studies used 10TeV MC.
- References are to CMS PAS (Physics Analysis Summary).



Lepton+jets

Selecting lepton+jets events

- Simple, robust selections for early data:
- At least 4 jets:
  - pT> 30GeV/c, $|\eta| < 2.4$, (for e+jets veto jets with $\Delta R(e, jet) < 0.3$).
- 1 lepton:
  - electron:
    single electron trigger (E_T>15GeV), E_T>30GeV,
    $|\eta|<2.5$ (exclude $1.442<|\eta|<1.560$), $d_0<200\mu m$, RelIso < 0.1
  - muon:
    single muon trigger (p_T>9GeV), p_T>20GeV/c, $|\eta|<2.1$, $d_0<200\mu m$, RelIso <0.05
- Veto events with extra isolated leptons :
  - Looser selections used (muon p_T>10GeV/c, electron E_T>15GeV, RelIso <0.2) for μ+jets.
- No MET criterion or b-tagging used.

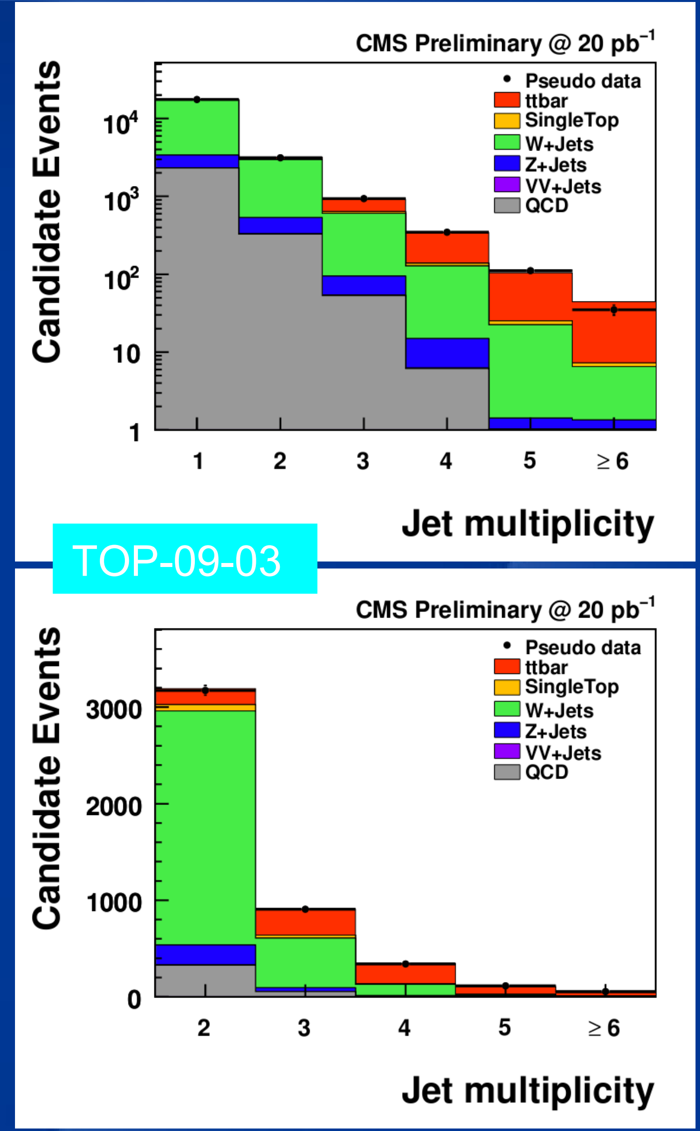


RelIso = $(I_{Tracker}+I_{Ecal}+I_{Hcal}) / E_T$ [p_T] for electron [muon].
d_0 = transverse impact parameter (with respect to beam spot).

TOP-09-03
&
TOP-09-04

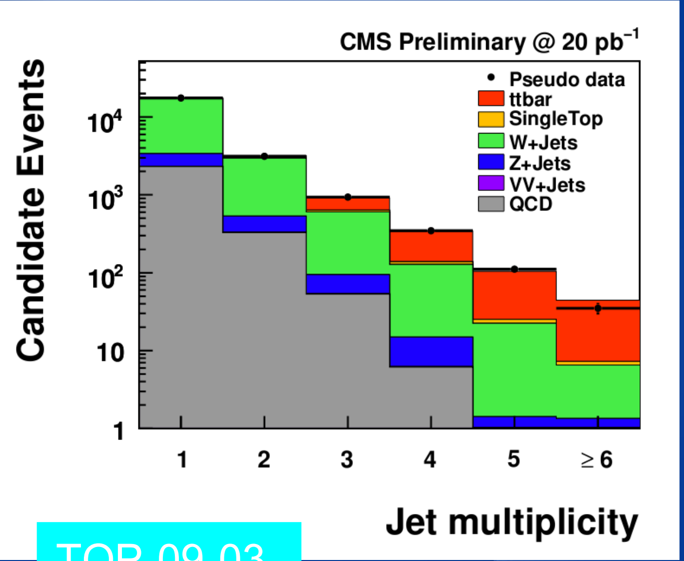
Muon+jets: Selected events

- At low jet multiplicities, dominated by W+jets.
  - Z+jets and QCD also contribute.
- ttbar events dominant for multiplicities of 4 (or higher).
- QCD contribution can be estimated from data by :
  - finding appropriate control regions
  - extrapolating from control regions to signal region.
- W(Z)+jets can be estimated by fitting discriminating variables.



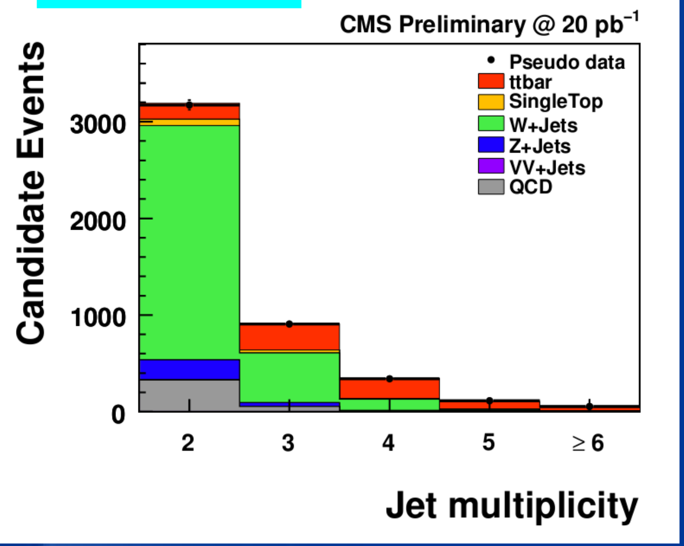
Muon+jets: Selected events

- At low jet multiplicities, dominated by W+jets.
  - Z+jets and QCD also contribute.
- ttbar events dominant for multiplicities of 4 (or higher).

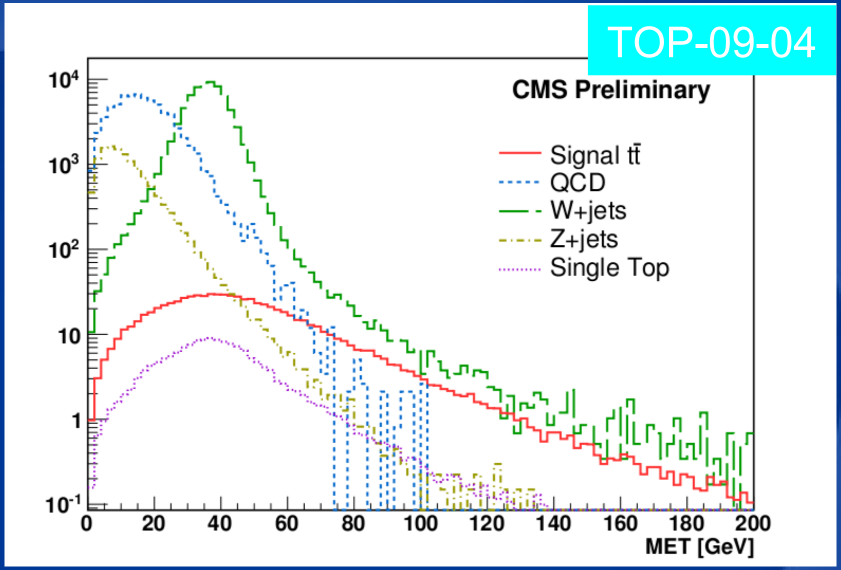


TOP-09-03

	$t\bar{t}$+jets s.l. μ	$t\bar{t}$+jets other	Single top s-Ch.	t-Ch.	tW	W+jets	Z+jets	VV+jets	QCD
AllEvents	1,220	7,060	32	832	580	912,000	76,240	236	2,546,279
Trigger	978	1,418	10	260	147	168,633	20,952	100	2,032,021
$\geq$ 1 tight μ	620	345	5	140	69	110,509	15,296	73	7,200
$<$ 2 tight μ	620	309	5	140	66	110,509	9,300	62	7,200
no tight e	620	264	5	140	62	110,508	9,292	53	7,200
veto on loose μ	618	228	5	140	60	110,503	5,492	44	7,192
veto no loose e	616	183	5	140	56	110,469	5,415	34	7,188
$\geq$ 1 jet	614	180	4	125	55	16,998	1,325	18	2,701
$\geq$ 2 jets	593	158	3	63	47	3,076	256	5	387
$\geq$ 3 jets	489	99	1	18	27	651	51	1	60
$\geq$ 4 jets	277	43	0	5	9	140	10	0	7



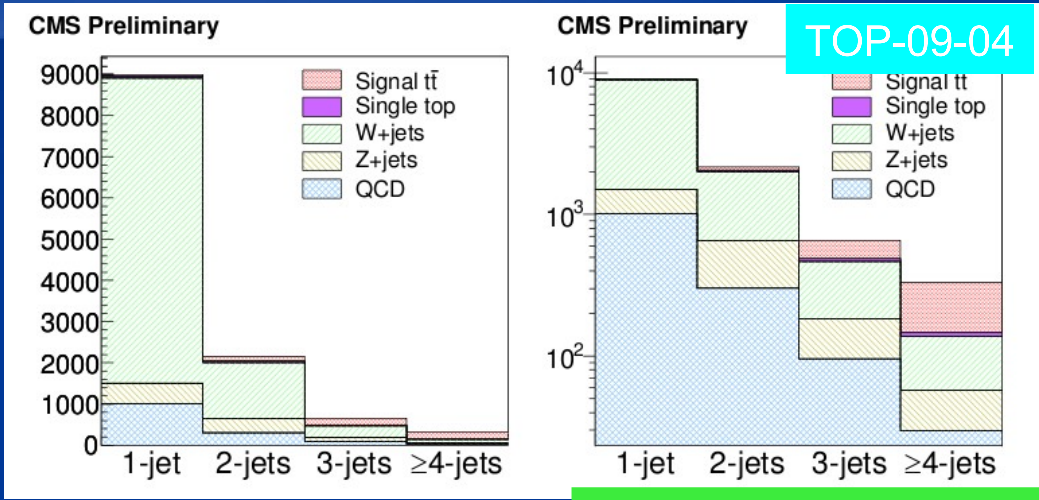
Electron+jets: Further selection



- W+jets and QCD are major backgrounds.
- Z+jets and QCD reduced by further selection.
- Strategies to reduce Z+jets:
  - Veto events with extra loose electrons (E_T>20GeV).
  - With events with extra electrons veto 76 < m_{ee}< 106 GeV/c^2.
- To reduce QCD backgrounds:
  - Restrict electrons to ECAL barrel only (|η| < 1.442) - most material before calorimeters in forward region.
  - Missing E_T >20 GeV.
  - Reject electrons consistent with coming from conversions.

Electron+jets: Further selection

- W+jets and QCD are major backgrounds.
- Z+jets and QCD reduced by further selection.

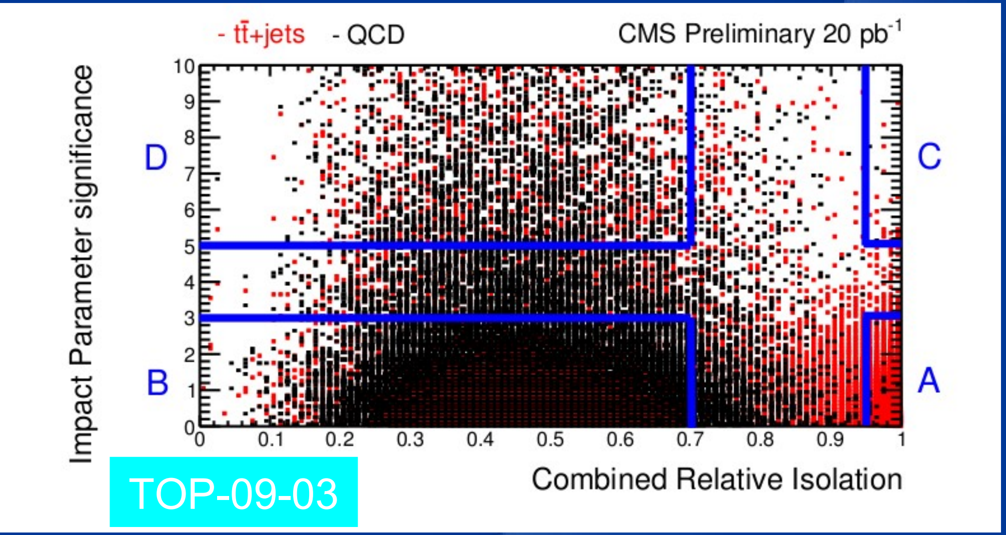


Events after baseline + m_{ee} veto, MET cut and conversion rejection.

	Cuts	$t\bar{t}$	W+jets	Z+jets	QCD	Single Top
	-	8280 ± 6	$9.1E5 \pm 265$	$8.4E{+}04 \pm 60$	$1.7E8 \pm 3.4E4$	1455 ± 2
	Trigger	4727 ± 5	$2.0E5 \pm 175$	$2.7E{+}04 \pm 42$	$3.4E7 \pm 1.9E4$	669 ± 2
	$\geq$1 Iso e	654 ± 2	$6.4E4 \pm 102$	$1.2E{+}04 \pm 29$	9030 ± 318	148 ± 1
	=1 Iso e	640 ± 2	$6.4E4 \pm 102$	8672 ± 25	9030 ± 318	146 ± 1
	Muon Veto	590 ± 2	$6.4E4 \pm 102$	8664 ± 25	9030 ± 318	143 ± 1
	$\geq$ 4 jet	215 ± 1	95 ± 4	46 ± 2	76 ± 20	10 ± 0
Option 1	Loose e Veto	208 ± 1	95 ± 3	20 ± 1	76 ± 13	10 ± 0
	$\|\eta\| < 1.442$	172 ± 1	57 ± 2	12 ± 1	31 ± 10	8 ± 0
Option 2	$\not{E}_T > 20$ GeV	188 ± 1	83 ± 4	34 ± 2	48 ± 15	9 ± 0
	Z Veto	186 ± 1	83 ± 4	29 ± 2	48 ± 15	9 ± 0
	Conv. Veto	183 ± 1	80 ± 4	28 ± 1	30 ± 14	9 ± 0

QCD - ABCD method

- Used in muons+jets analysis.
- Assign a phase space with two (uncorrelated) variables.
- Define four regions (A, B, C, D):
  - Each dominated by signal or (QCD) background.
  - Region A is the signal region.
  - Assume ratio of number of events in regions is: $N_A/N_B = N_C/N_D$
- Therefore $N_A = N_B \times N_C / N_D$



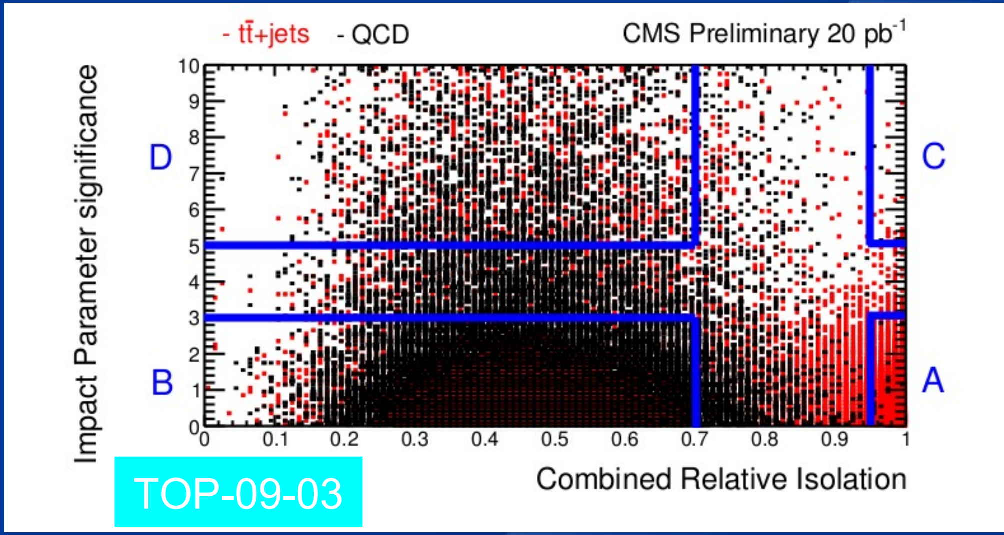
QCD - ABCD method

- Variables used:
  - Combined relative isolation of muon:
    RelIso' = 1/(1+RelIso)
  - muon impact parameter significance:
    significance(d_0) = $d_0/\sigma(d_0)$
    with respect to beam spot
- Produces results in good agreement.
- Boundary changes used to test stability and estimate uncertainty:
- Estimated conservatively as 50%.

TOP-09-03

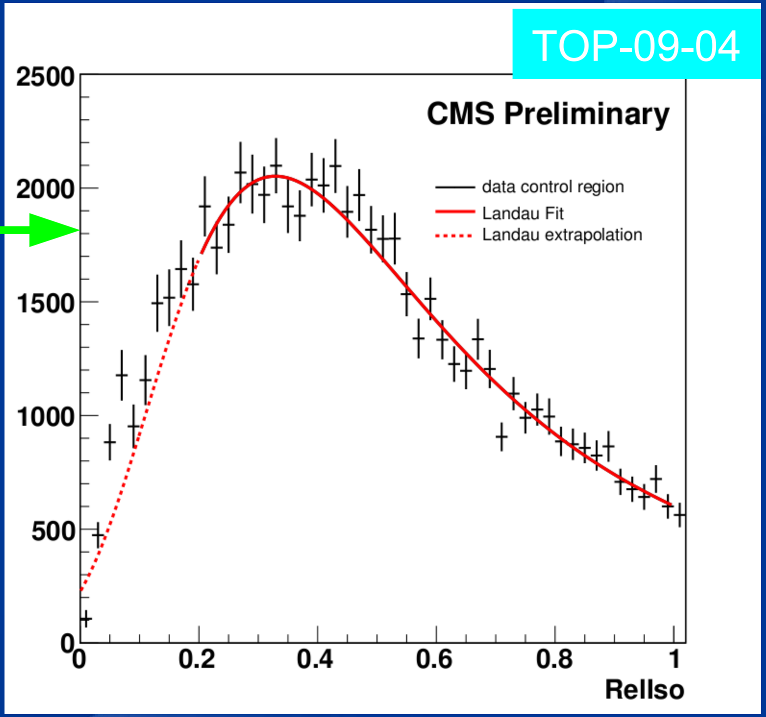
N_A → N(QCD) Estimated

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



QCD - RelIso extrapolation

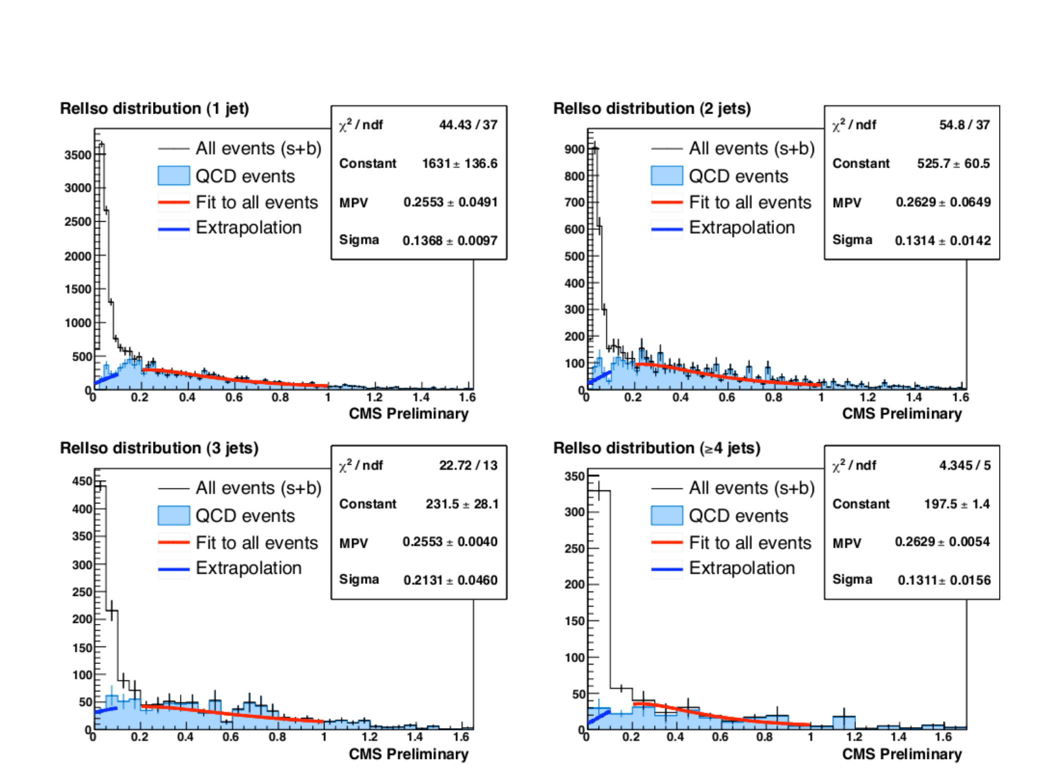
- Used by both electron+jets and muon+jets channels.
- For electron+jets channel:
  - RelIso = $(I_{tracker} + I_{ecal} + I_{hcal})/E_T$
- Signal (and non-QCD backgrounds) region: RelIso < 0.1
- Use higher values as a control region.
  - Extrapolate from background control region into signal region.
- Test various functional forms on pure QCD sample:
  - invert Missing E_T selection.
  - loosen electron selection.
  - tighten Z veto.
  - Apply H_T selection



H_T = scalar sum jet E_T, electron E_T, missing E_T.

QCD - RelIso extrapolation

- Landau function gave best fit.
  - This is used to fit main RelIso dist^n.
  - No physical motivation.
- Very few events with 3 or 4 selected jets:
  - Fix mean peak value from 1 and 2 jet fits.
  - Improves stability.
- A 50% total uncertainty is applied to result.
- Results are systematically slightly low:
  - remaining conversions are the major factor.



TOP-09-04

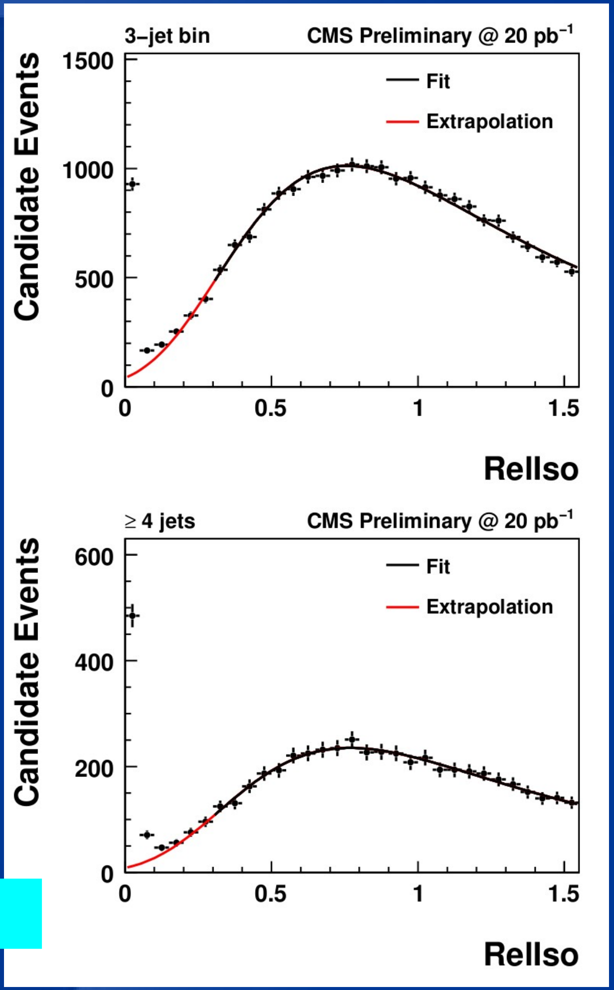
	Signal region	
	True QCD $20\ pb^{-1}$	Estimate $20\ pb^{-1}$
1j	1007 ± 102	815
2j	301 ± 47	227
3j	96 ± 28	71
$\geq$4j	30 ± 14	17

QCD - RelIso extrapolation

- RelIso extrapolation in mu+jets channel:
- Similar to e+jets channel:
  - Narrower signal region (RelIso <0.05).
  - Wider control region.
  - Landau again gives best fit.
- RelIso' also studied using a Gaussian fit.
- Stability studied by varying signal and region ranges.
- From studies a conservative uncertainty of 50% is placed on the extrapolation.

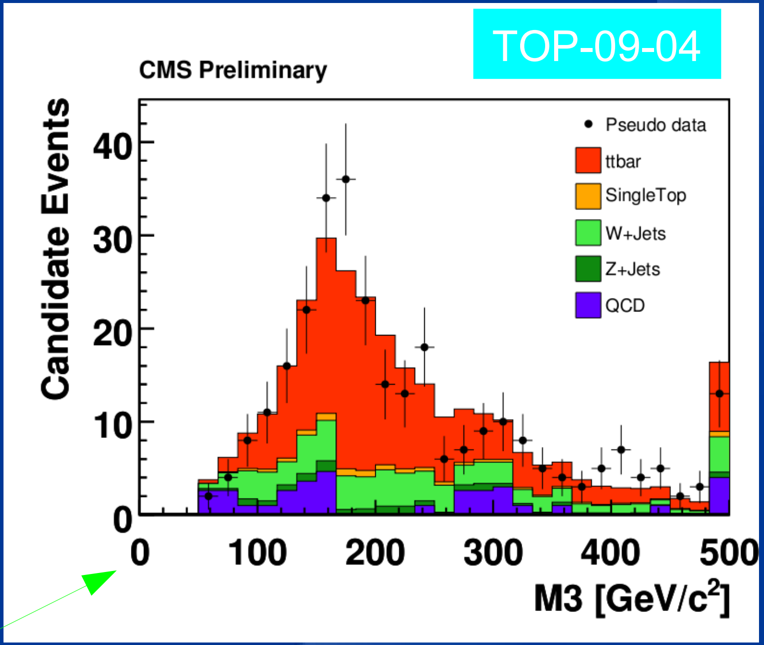
Jets	N(QCD) Predicted	N(QCD) Estimated
2	327	378 ± 82
3	53	47 ± 24
≥ 4	7	13 ± 7

TOP-09-03



W+jets - M3

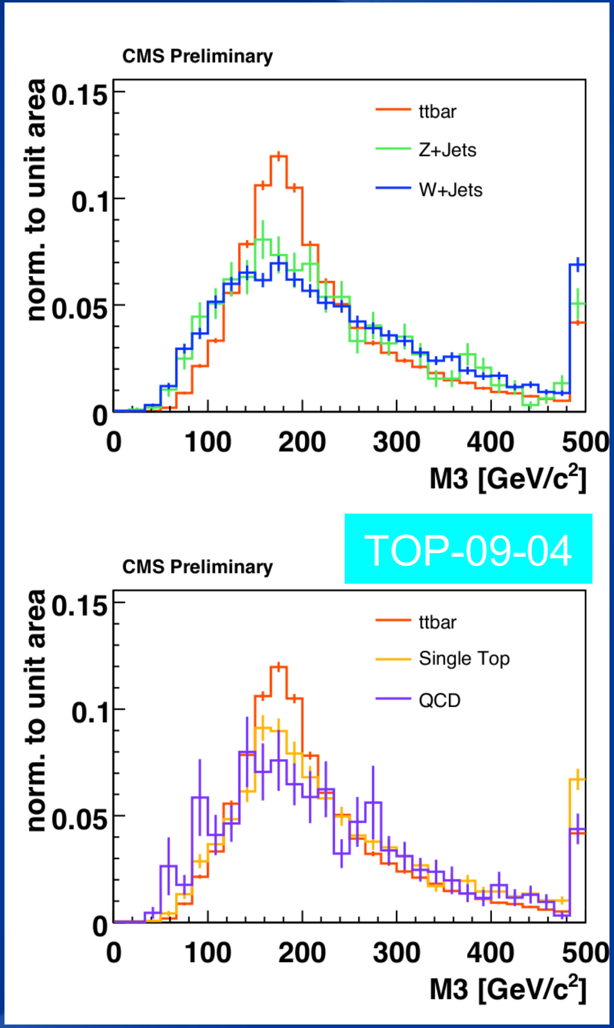
- W+jets is the dominant remaining background.
  - It can be separated from ttbar by performing a fit to a discriminating variable distribution.
  - Remaining Z+jets background may be included with W+jets.
- One discriminating variable is “M3”:
  - invariant mass of the 3 jet combination, with highest p_T (vector sum).
- Approximates top mass
  - should peak near m_{top} for $t\bar{t}$.



e+jets M3 distribution

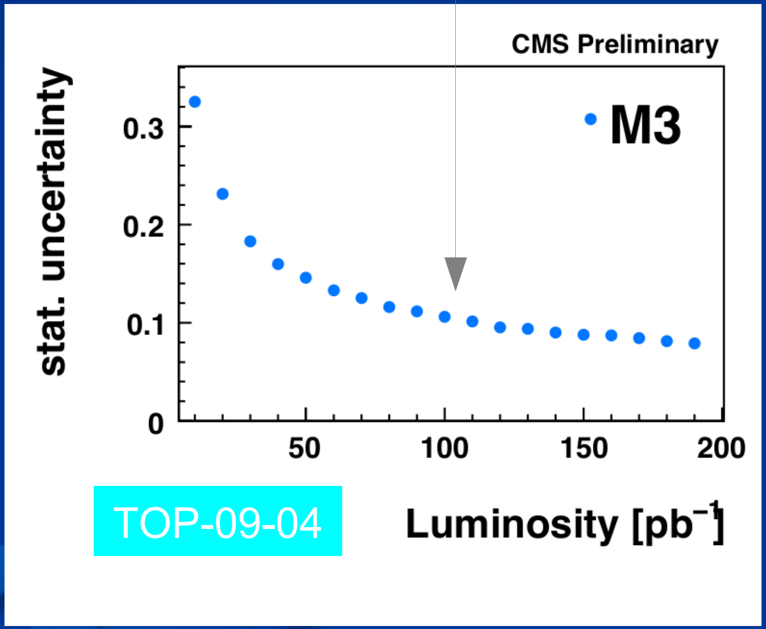
W+jets - M3 fit - electron channel

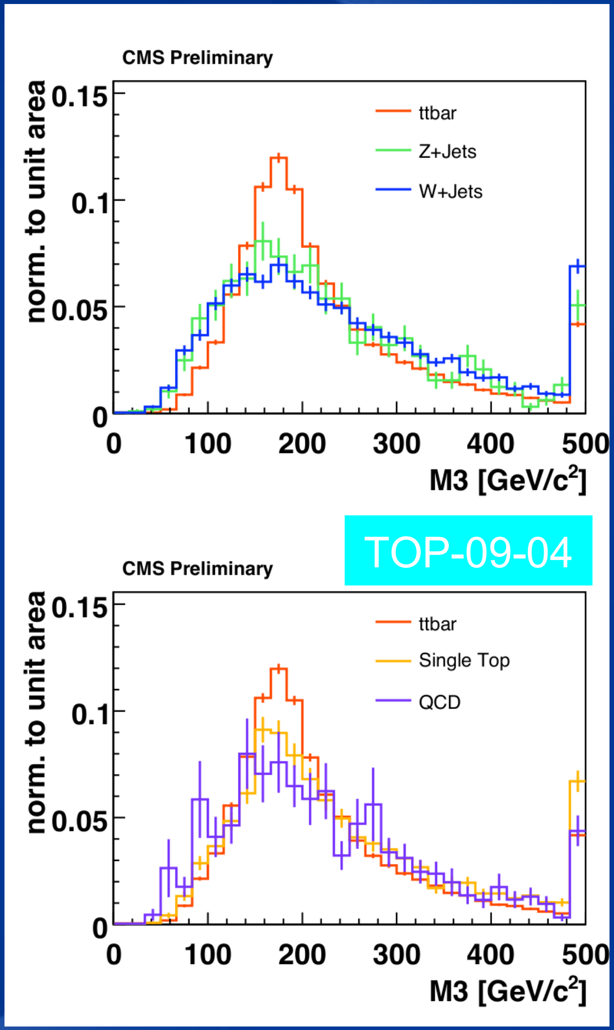
- Perform a fit to 4 components:
  - ttbar, W(Z)+jets, QCD, and single top.
  - QCD template can be taken from data (control region).
  - Others taken from simulation.
  - W(Z)+jets shape could be taken from Z+jets data, but not enough early on.
- QCD and single top contributions are constrained in the fit.
- N_{ttbar} and $N_{W/Z+jets}$ are extracted.
- pseudo-experiments are preformed to check for bias and estimate error.
  - Error estimated at 23% for $20pb^{-1}$.



W+jets - M3 fit - electron channel

~10% for $100pb^{-1}$

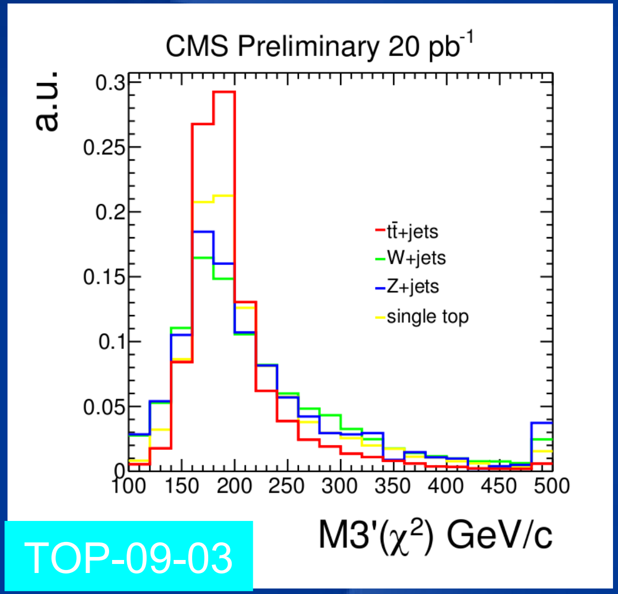




- pseudo-experiments are preformed to check for bias and estimate error.
  - Error estimated at 23% for $20pb^{-1}$.

W+jets - muon channel

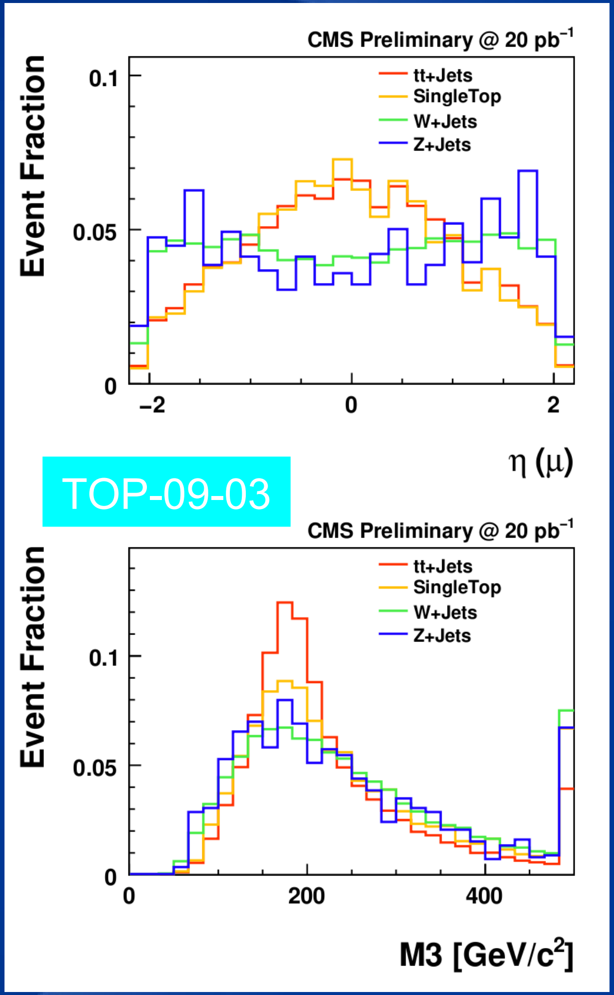
- Three discriminating variables tested:
  - $\eta(\mu)$, M3, and M3'.
- M3' is calculated using a χ^2 distribution:
- $\chi^2 = (m_{j1j2} - m_W)^2/\sigma^2_{jj} + (m_{j1j2j3} - m_t)^2/\sigma^2_{jjj} + (m_{\mu\nu j4} - m_t)^2/\sigma^2_{\mu\nu j}$
- σ_x are resolutions of each jet combination.
- Requires MET to calculate $m_{\mu\nu j4}$.



- χ^2 for Combinations of up to 7 jets calculated.
- M3' is the invariant mass of the three jets forming the hadronic component of the χ^2 for the lowest calculated χ^2.
- Fit carried out using M3' distributions as templates.
- Pseudo experiments used to evaluate an uncertainty of ±12%.

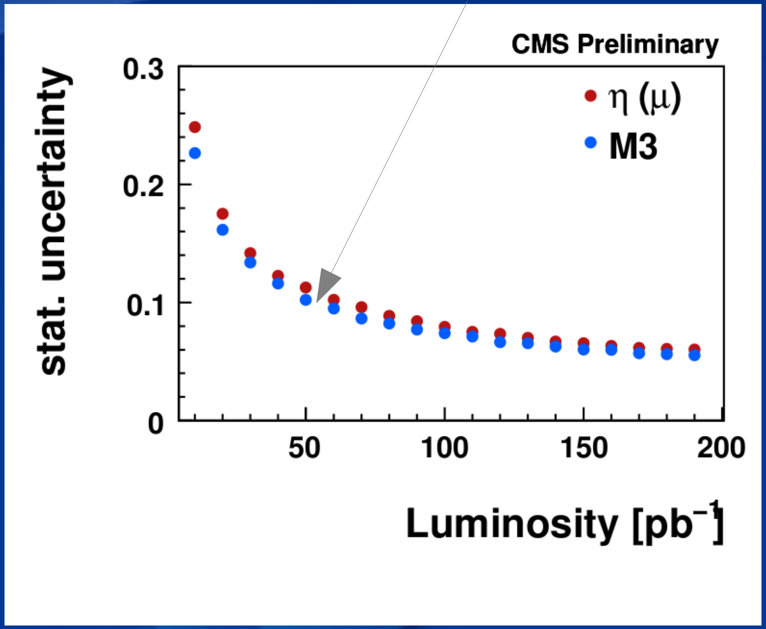
W+jets - muon channel

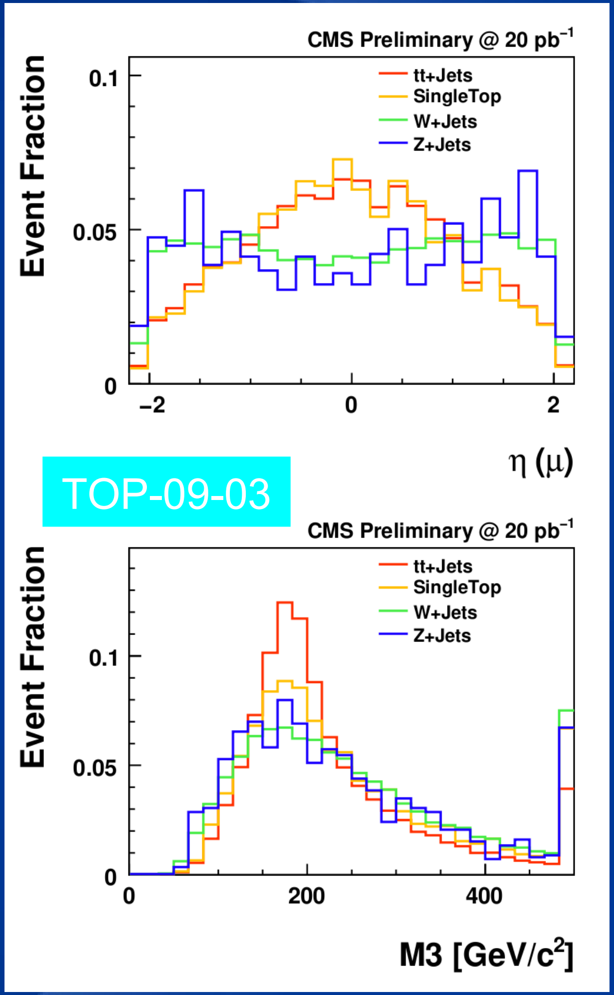
- Three discriminating variables tested:
  - $\eta(\mu)$, M3, and M3'.
- Three components fitted:
  - ttbar, single top, and W+jets (including Z+jets and QCD).
  - Taken from simulation.
  - Could be taken from lower jet multiplicities with enough data.
  - Single top constrained.
- Pseudo experiments used to check method, and estimate uncertainty.
  - Uncertainty estimated to be ±16% for M3, and ±18% for $\eta(\mu)$.



W+jets - muon channel

~10% at $50pb^{-1}$





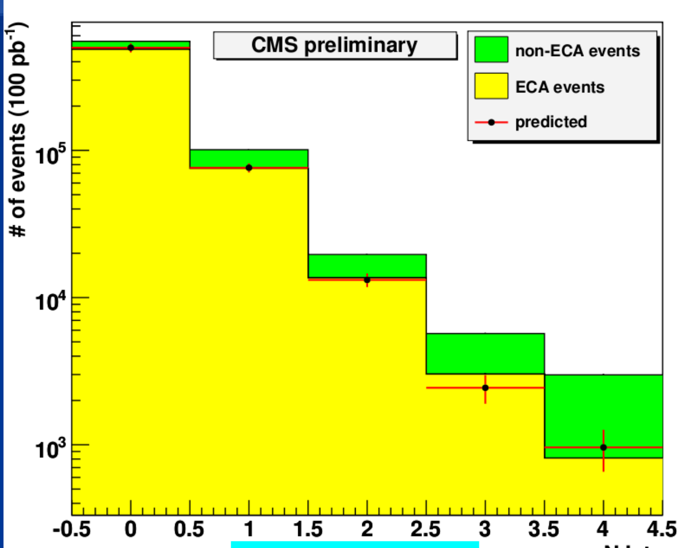
- Pseudo experiments used to check method, and estimate uncertainty.
  - Uncertainty estimated to be ±16% for M3, and ±18% for η(μ).

W+jets - Charge Asymmetry

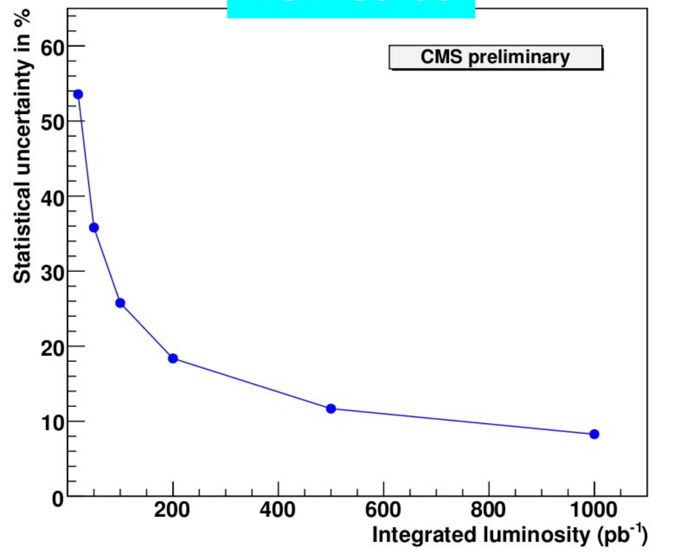
- Signal $t\bar{t}$ is charge symmetric:
  - Consider W+jets with other asymmetric backgrounds (Vbb, single top (s and t-channels) as “Events leading to Charge Asymmetry”.
- Total ECAs:
  - $(N_+ + N_-)_{data} = R_\pm(W) \times (N_+ - N_-)_{data}$
  - $R_\pm(W)$ corresponds to inverse of W charge asymmetry:
    - $$R_\pm(W) = \frac{N_{W+} + N_{W-}}{N_{W+} + N_{W-}} = \frac{A_+\sigma_{W+} + A_-\sigma_{W-}}{A_+\sigma_{W+} + A_-\sigma_{W-}}$$
  - $\sigma_{+(-)}$: $W^{+(-)}$ cross section, $A_{+(-)}$: geometrical acceptance.
- Assume ECAs dominated by W+jets, and $R_\pm$ is same for all ECAs and W+jets.

W+jets - Charge Asymmetry

- $(N_+ - N_-)_{data}$ will have a large statistical uncertainty.
  - Method studied for $100pb^{-1}$.
- $(N_+ - N_-)_{data}$ estimated from pseudo data, counting muon and anti-muon events.
- $R_\pm$ estimated from statistically independent W+jets sample.
- In data an independent sample may be found (low jet multiplicities).
- Statistically dominated:
  - systematic (PDF and ECA components) ~11%.
- Total uncertainty ~30% for $100pb^{-1}$.



TOP-09-03

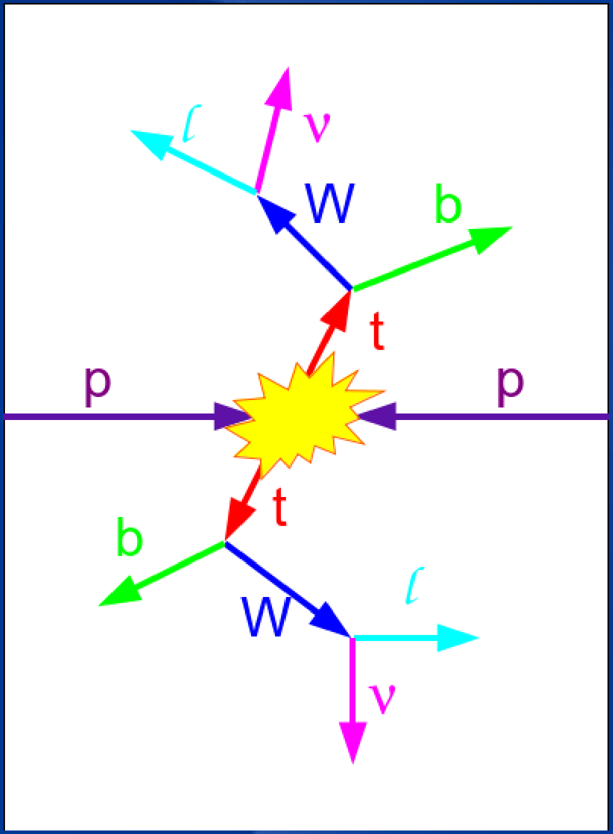


Dileptonic

Selecting Di-leptonic events

- 2 leptons (μ or e):
  - single electron (E_T>15GeV) or single muon (p_T>9GeV/c) trigger,
  - p_T >20GeV, $|\eta|$<2.4, opposite signs.
  - Individual relative isolations*: I_{trk} > 0.9, I_{cal} > 0.9 (0.8) for μ (e).
- 2 (or more) jets:
  - p_T>30, $|\eta|$<2.4
  - no b-jet identification.
- Missing E_T:
  - ME_T > 20/30 GeV (decay channel dependent).

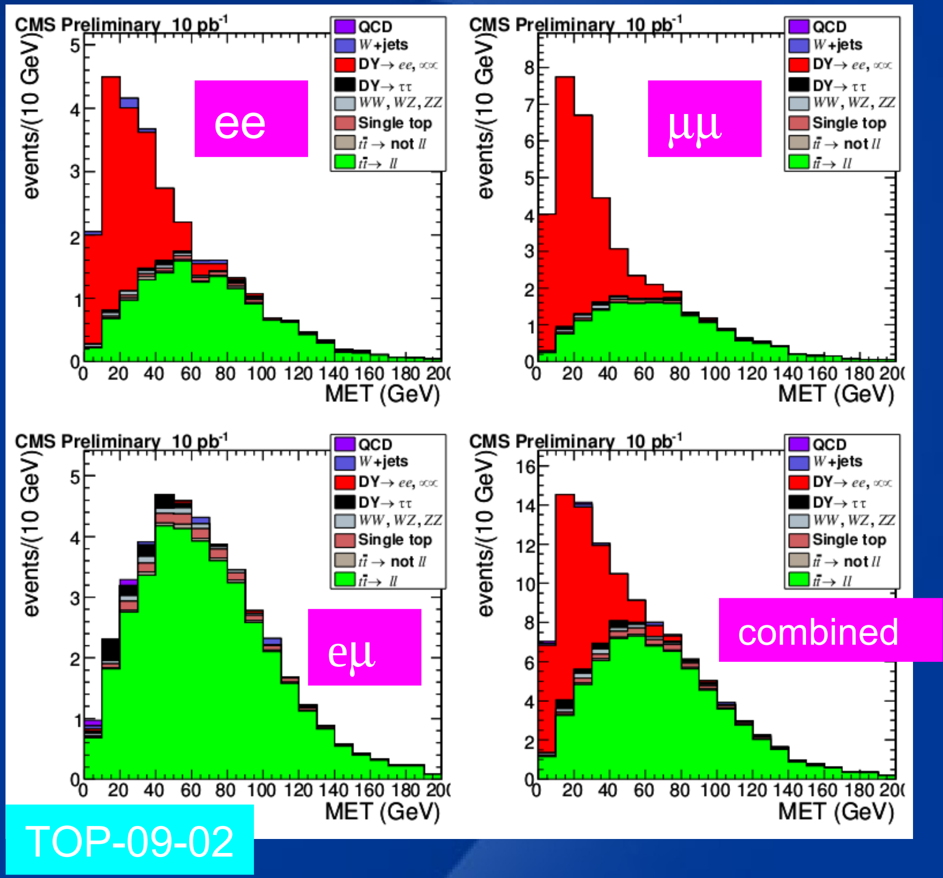
* Individual relative isolation
  = $p_T/(p_T$ + absolute_isolation)



TOP-09-02

Selecting Di-leptonic events

- Three channels:
  - e^+e^-, $e^\pm\mu^\mp$, $\mu^+\mu^-$
- In ee and μμ,
  - Z veto: $|m_{\ell\ell} - m_Z| < 15\text{GeV}/c^2$
  - Missing E^T > 30GeV.
- eμ is far cleaner:
  - looser missing E_T cut.
- Dominant background is from Drell-Yan(DY) + jets.



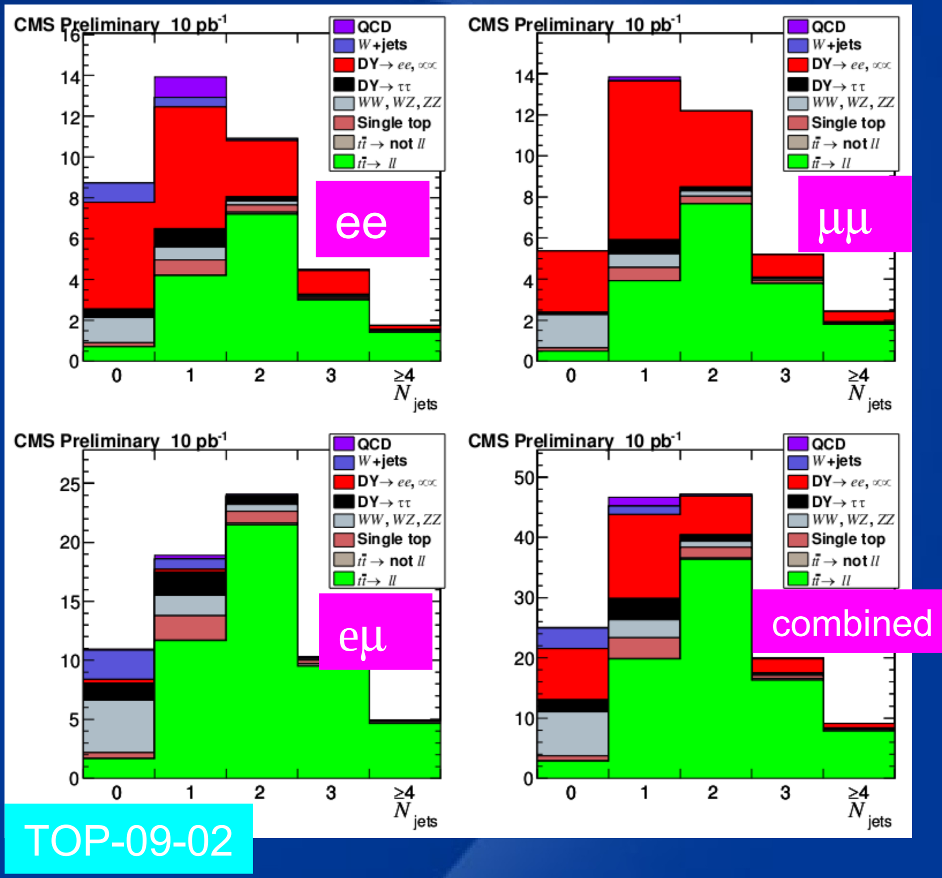
Selecting Di-leptonic events

- Three channels:
  - e^+e^-, $e^{\pm}\mu^{\mp}$, $\mu^+\mu^-$

TOP-09-02

	Main selection		
	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
other $t\bar{t}$	0.21 ± 0.03	0.04 ± 0.01	0.46 ± 0.04
Single top	0.46 ± 0.03	0.56 ± 0.03	1.40 ± 0.06
$WW/WZ/ZZ$	0.26 ± 0.02	0.33 ± 0.03	0.71 ± 0.05
DY$\to \tau\tau$ + jets	0.3 ± 0.1	0.3 ± 0.1	0.7 ± 0.2
DY$\to ee/\mu\mu$ + jets	4.1 ± 0.4	5.3 ± 0.4	0.08 ± 0.05
W + jets	0.2 ± 0.1	< 0.1	0.3 ± 0.1
QCD	< 1	< 0.4	< 0.4
Total backgrounds	5.5 ± 0.4	6.6 ± 0.4	3.7 ± 0.2
Data driven fakes	1.1 ± 0.6	0.8 ± 0.4	2.5 ± 1.2
Data driven DY	4.0 ± 1.3	5.1 ± 1.6	

- Dominant background is from Drell-Yan(DY) + jets.



Drell-Yan + jets

- DY events selected due to mis-measurement of missing E_T.
- To estimate contribution, use dileptonic events $76 < m_{\ell\ell} < 106$ GeV/c^2.
  - Estimate $R_{out/in} = N^{in}_{DY\ MC}/N^{out}_{DY\ MC}$ fraction of DY events inside the range, relative to outside, directly from simulation.
  - $N^{out}_{DY}{}^{(est)} = R_{out/in} \times N^{in}_{DYdata}$ then estimates the number of DY expected.
    - Non-DY contribution inside corrected for with eμ events.
    - Relies on simulated events, but not jet and MET properties.
- Modified selection, and different simulated samples used to estimate systematic uncertainty of 30%

Fake leptons

- Use events that pass a looser selection:
  - loosen lepton isolation (calorimeter and tracker) requirements.
- Fake Ratio (FR) is number of these events that pass main selection.
- Scale number of events passing loose selection, and failing main selection by FR/(1 – FR):
  - obtain estimate of number of fake leptons.
  - Small biases due to double counting and trigger differences.
- Uncertainty estimated from statistics of samples used, and variation in different jet multiplicities:
  - gives overall systematic uncertainty estimate of ±50%.
  - From 0.8±0.4 events in mm, to 2.5±1.2 events in em.

Summary

- Early measurements of $t\bar{t}$ cross-section are possible using some data driven methods to estimate backgrounds.
- e+jets and μ+jets decay channels, dominant backgrounds:
  – QCD - estimated from data driven methods, uncertainty ~50%.
  – W+jets - estimated from template fits, uncertainty 15-25%.
- For dilepton decay channel:
  – Drell-Yan major background for ee and μμ; estimated uncertainty ~30%.
  – Fake lepton background, uncertainty ~50%.
- Work is ongoing - these methods will soon meet data for the first time...

Backup

Conversion Algorithm

- When a photon converts the electron (positron) will bend in opposite directions in the magnetic field, in the ϕ plane.
  - Algorithm looks for pairs of tracks with opposite charges:
  - Calculate closest 2D distance between them in ϕ plane.
  - Veto events where distance < 0.04 cm and $|\Delta \cot \theta| < 0.03$
- Efficiency is ~50%, with a signal loss of 1.7% from ttbar and QCD MC samples.