Strategy for a $t\bar{t}$ Production Cross Section Measurement at CMS

Frankie Bostock
on behalf of the CMS collaboration

IOP HEPP & APP
30th March 2010

Aims

- Our intention is to extract a $t\bar{t}$ signal, and furthermore to make a measurement of the production cross section, at the LHC.
- The motivations behind this are:
  1) Commissioning of the detector - top signatures are characterized by high pt leptons, jets and missing energy. A successful measurement means a successful detector.
  2) "Rediscovery" of the top - we are now testing the standard model at the highest energies ever achieved.
  3) Potential signs of new physics

- Our main interests as we enter the early days of data-taking obviously lie within the first two categories, as we continue to understand and commission our detector.

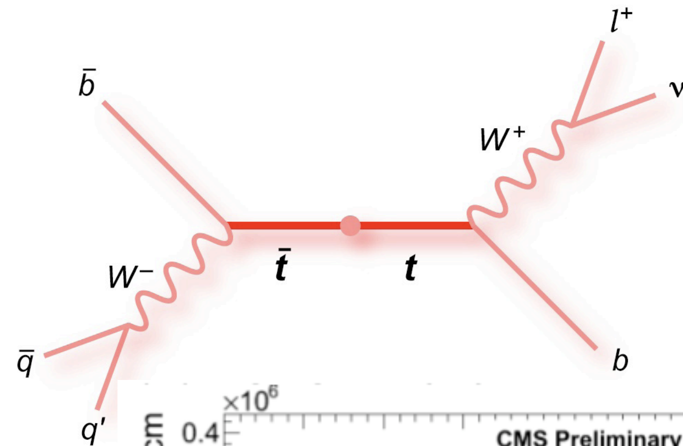
- Here I will present the results of our analysis based on $\sqrt{s}$ = 10 TeV with an integrated luminosity of $20pb^{-1}$.
- Most of what follows applies to directly to how we are approaching 7TeV, with a luminosity of ~5-10 pb^{-1}.

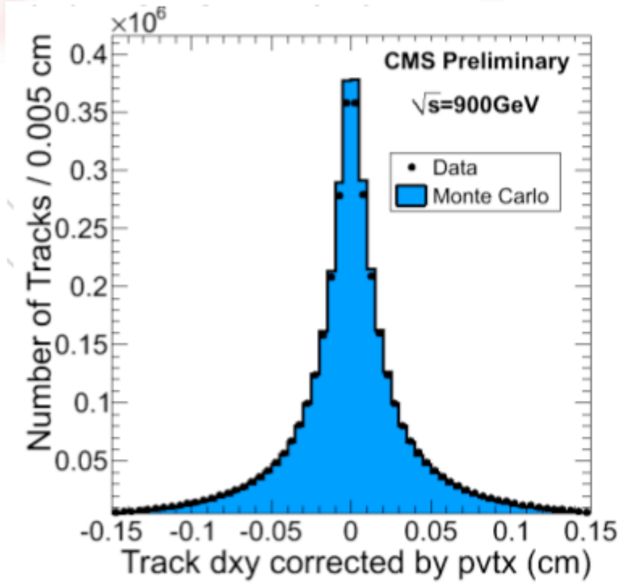
Top Topology

- We consider semileptonic decays in $t\bar{t}$, working in the "electron+ jets" channel - one top decays to an electron, the other to jets.

- Our signal is characterized by:
  - Electrons with high transverse momentum
  - Missing transverse energy from the neutrino
  - At least 4 jets

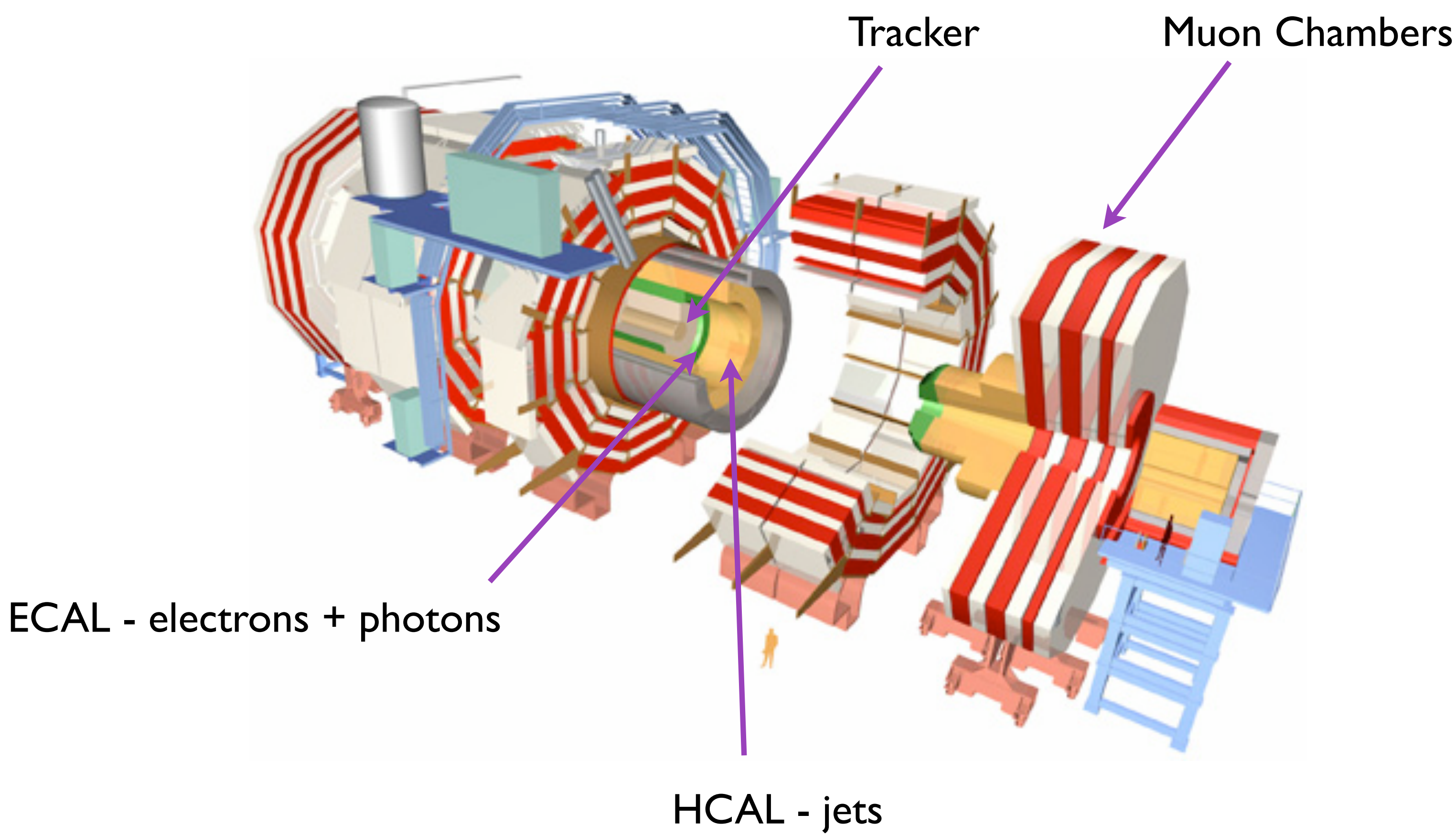
- In addition, two of these jets are b-jets.
- We could make use of b-taggers to gain a purer sample, but as a first approach we are aiming at an un-tagged analysis.
- With the current performance we have seen at 900GeV, the use of b-taggers might be possible a lot earlier than anticipated.





Transverse Impact Parameter

The CMS detector



A successful cross section measurement, which requires all of these components, would be a great step towards commissioning of the detector.

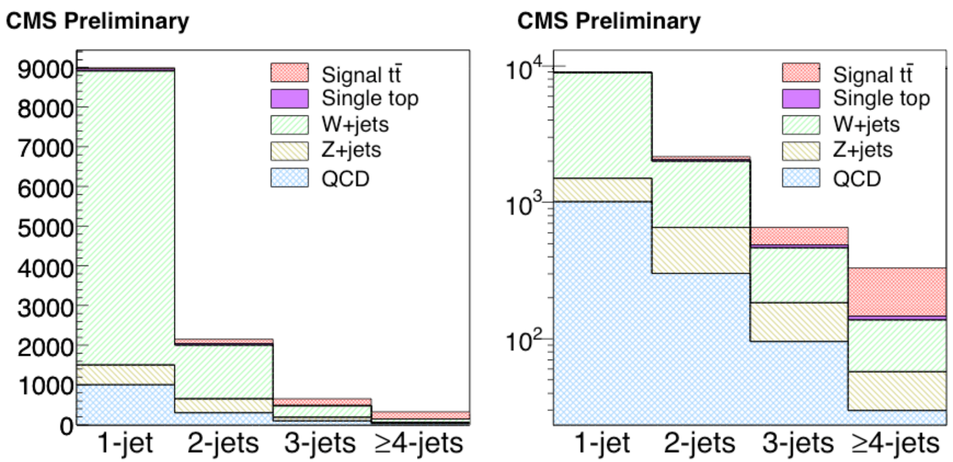
Strategy Breakdown

Firstly, we start with an event selection that discriminates against the majority of the background.
This mostly comes from fake electrons in multi-jet events (QCD).

- Various electron kinematic and quality cuts
- Veto Muons
- Veto Z events (reject events with two electrons)
- Missing ET - Mainly for rejecting QCD and Z+jets events
- Photon Conversion veto

Signal	Background	S/B
183	147	1.25

Final numbers after selection

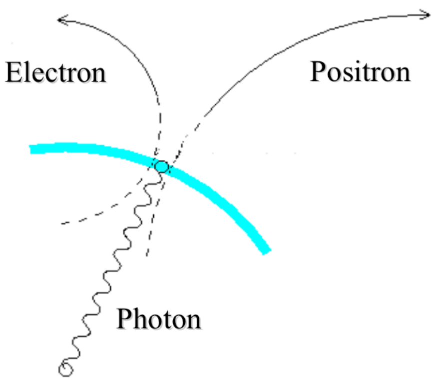


N(jet) multiplicity after selection

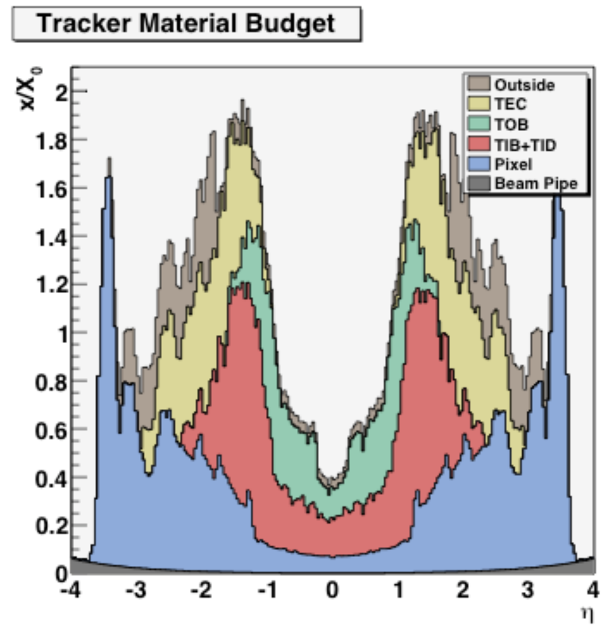
Once we've taken this as far as we can go, we apply data-driven techniques to estimate the two remaining dominant backgrounds

- QCD
- W+jets

Conversion Background

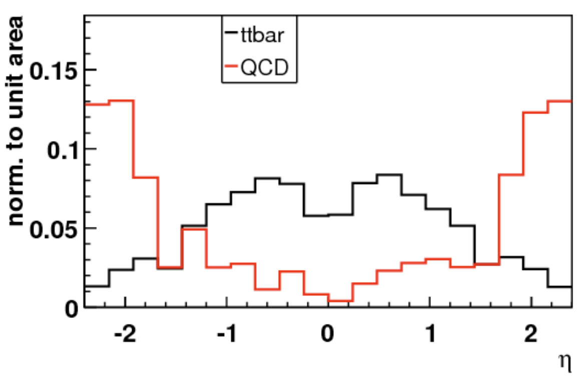


The dominant background coming from QCD is converting photons. Large material budget in CMS means we get a lot of conversions.



A few ways to tackle these:

- eta cut - remove the forward region - brute force approach

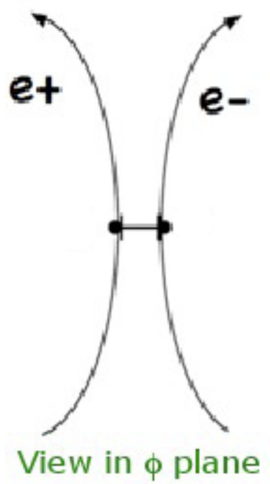


- Impact parameter cut on the track - removes the more symmetrical conversions
- Geometry cuts
- Missing tracker hits

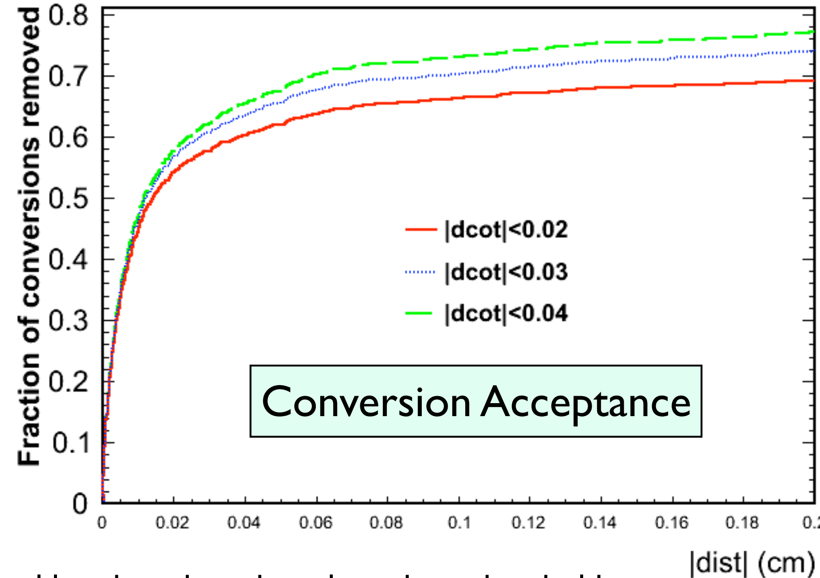
The analysis presented only makes use of the impact parameter cut and geometry cuts

Conversion Background 2

Geometrical Cuts



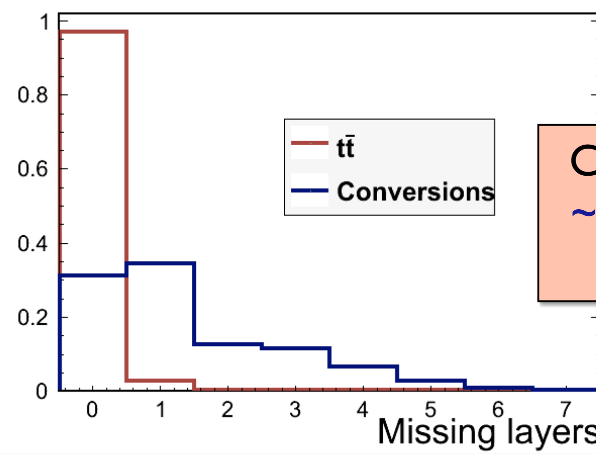
- We look for tracks near the electron track.
- After simple requirements (e.g. opposite sign) we then apply two further cuts:
  - "dist": distance of closest approach of both tracks in phi plane
  - "dcot" minimum requirement based on the theta angles of each track.
- If all these are met, the electron is flagged as coming from a conversion.



This alone can reject ~50% of conversions with around ~4% signal loss based on the selected cut thresholds.

Missing tracker hits

Prompt electrons will have tracker hits starting from the primary vertex.
In comparison, electrons from conversions will have tracks starting later in the tracker.
Can reject electrons which don't have initial hits close to the PV.

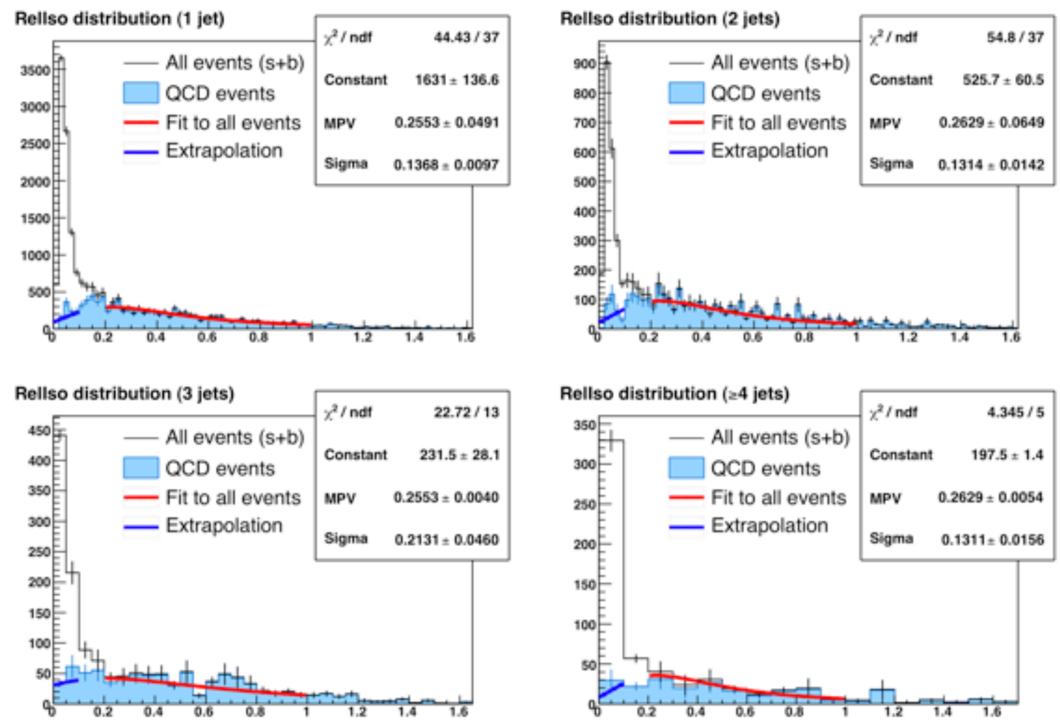


Can potentially remove ~70% conversions with loss of ~3% signal

QCD background Estimation

- To estimate QCD, we consider a variable which contains the signal region and a QCD dominated region.
- We fit a function to the QCD region.
- Extrapolate this into the signal region.
- For this we use "RelIso" - ratio of the electron isolation to the electron Et.

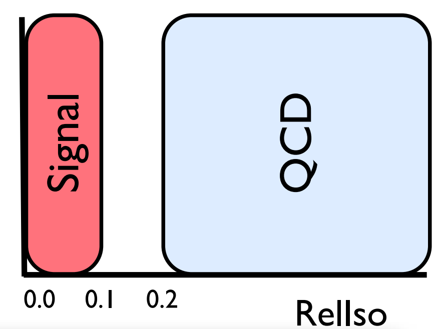
Isolation - sum of ECAL and HCAL energy deposits, and Pt of tracks in a cone around the electron



Example fit results

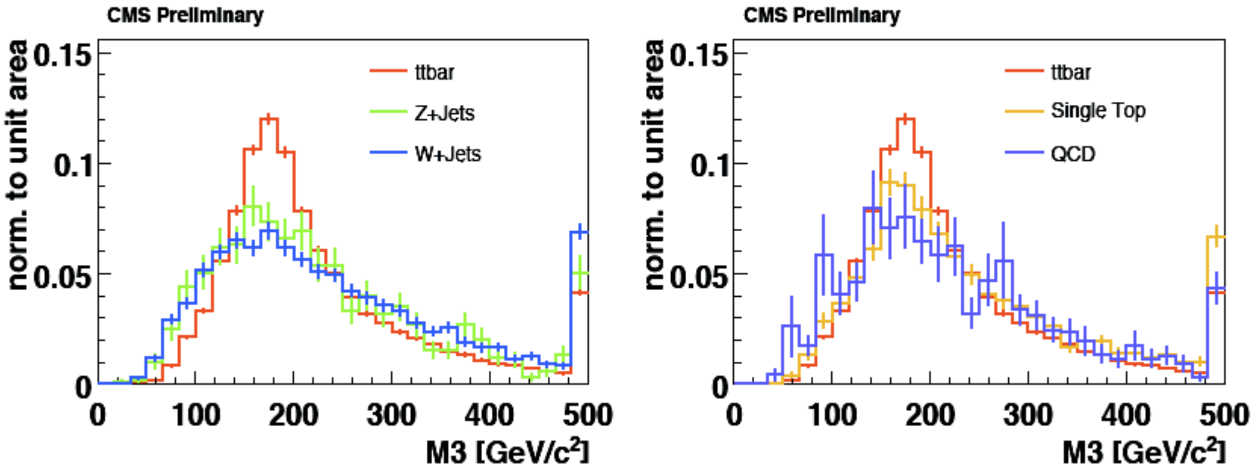
	Signal Region	
	True (QCD) 20 pb-1	Estimate 20 pb-1
1 jet	1006.6	814.5
2 jets	301.4	227.4
3 jets	95.9	71.0
≥4 jets	29.9	17.0

Method has been tested for robustness by using various fit regions, and typically accurate to within 40%.
QCD makes up ~10% of all events so, this is manageable.



W+Jets background Estimation

•To estimate the W+jets background, we find a discriminating variable between this and our signal.
•We use M3:
•Reconstructed mass of the three jets which give the highest vectorially summed Pt.
➡ Simple “reconstruction” of the hadronically decaying top quark.
•We take templates for each data type, and then fit the normalisations based on data.
•Templates come from Monte Carlo, except for QCD would can be attained from data.



By testing with pseudo-experiments, we can ascertain the effect of systematics on the fit, and ergo the final cross section.

Final Result Expectations

After applying our event selection, we get the following typical expected yields:

20 pb^{-1}

ttbar	W+jets	Z+jets	Single top	QCD	Total
183	80	28	9	30	330

$$\sigma_{t\bar{t}} = \frac{N^{fit}_{t\bar{t}}}{A \cdot \varepsilon_{t\bar{t}} \times \int \mathcal{L} dt}$$

We take the number of ttbar events taken from the M3 fit. Measurement of signal efficiency will come from data-driven methods and MC.

Systematic Uncertainties

	Relative Systematic Uncertainty
Jet Energy Scale	15%
MC Generator	10%
$t\bar{t}$ ISR/FSR uncertainty	3%
W+jets MC Factorization Scale	1%
W+jets MC Matching threshold	5%
Shape uncertainty of Single Top	1%
Shape uncertainty of QCD	5%
PDF uncertainty	5%
Total	20%

Together with this systematic uncertainty, we expect a statistical uncertainty of 23% at 20 pb^{-1}

Summary and Outlook

- I have covered the analysis that we developed for $\sqrt{s}$ = 10 TeV, based on 20 pb^{-1} of data.

- We have moved the analysis to 7 TeV which mostly follows the same lines.
- It has various improvements, in particular the conversion rejection.

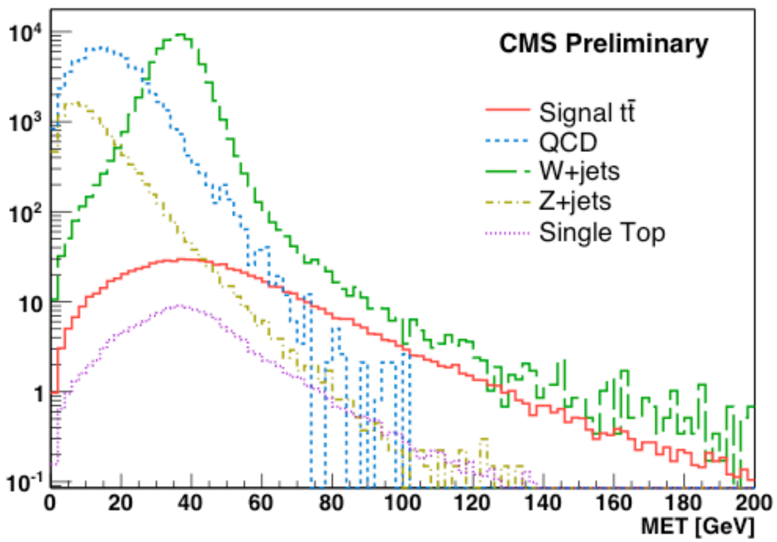
- Although the cross section drops to less than half that at 10 TeV, we can expect to see top signals from ~1 pb^{-1} onwards at 7 TeV.
- Towards 10-20 pb^{-1} we will be able to make a quantifiable cross section measurement.

Many thanks to my colleagues at Bristol and the CMS Top PAG

Thank you for your attention!

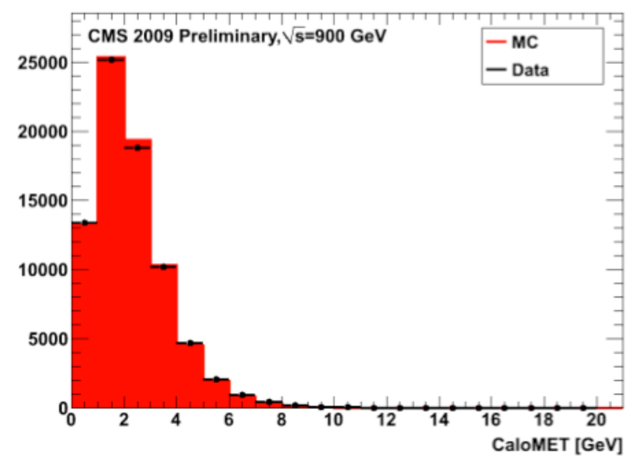
Backup Slides

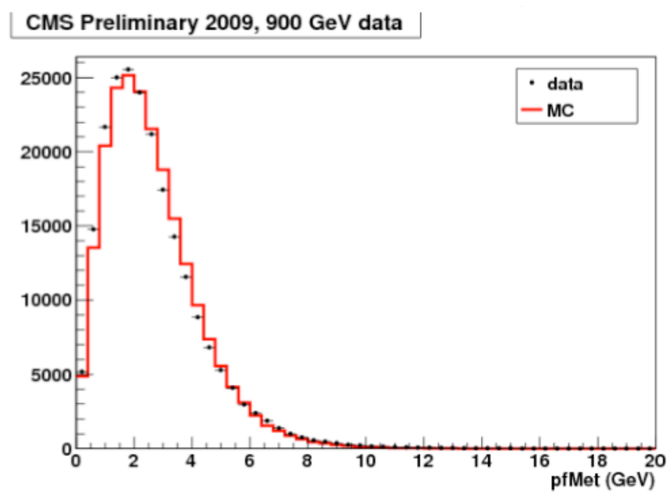
MET



Plot shows MET rejection power at 10 TeV. Potentially a problematic variable to cut on, though initial performances seen on data are very promising.

MET performance on data - approved plots comparing MET from MC with 900 GeV data.



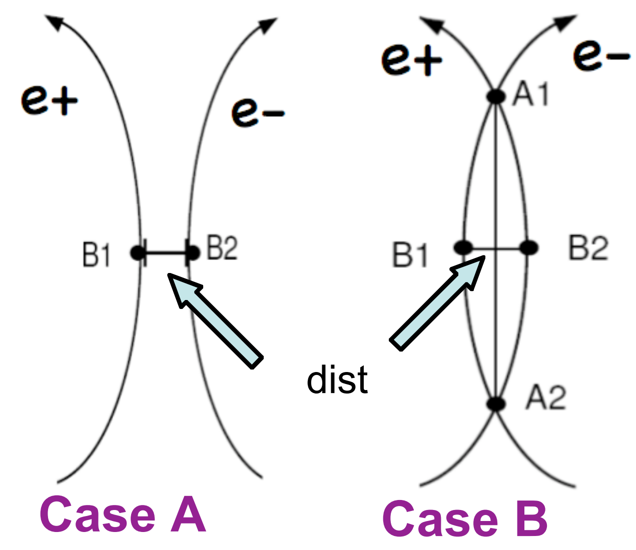


Conversion Algorithm

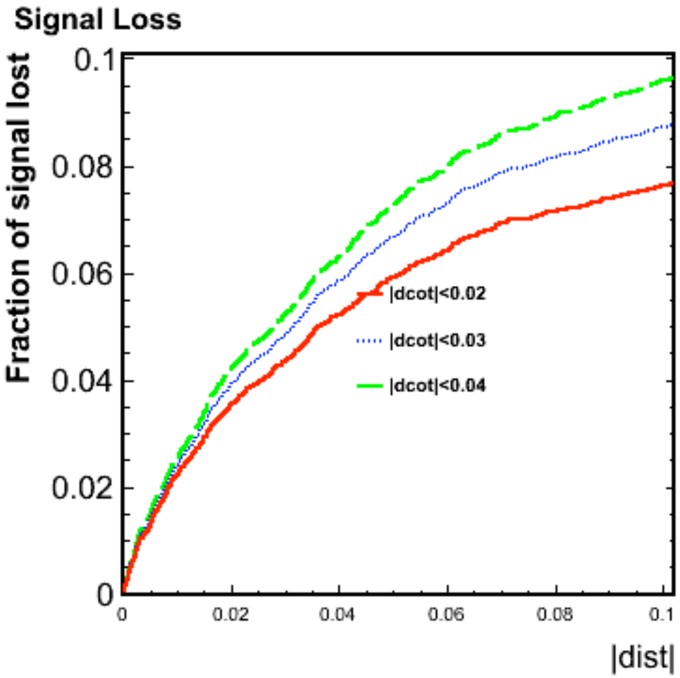
- In order to reject conversions, we employ an algorithm that takes advantage of the geometry of a conversion event - picking out the two tracks from the e+ e- pair.
- We loop over the reconstructed track collection looking for a potential conversion partner to the track of our reconstructed electron.

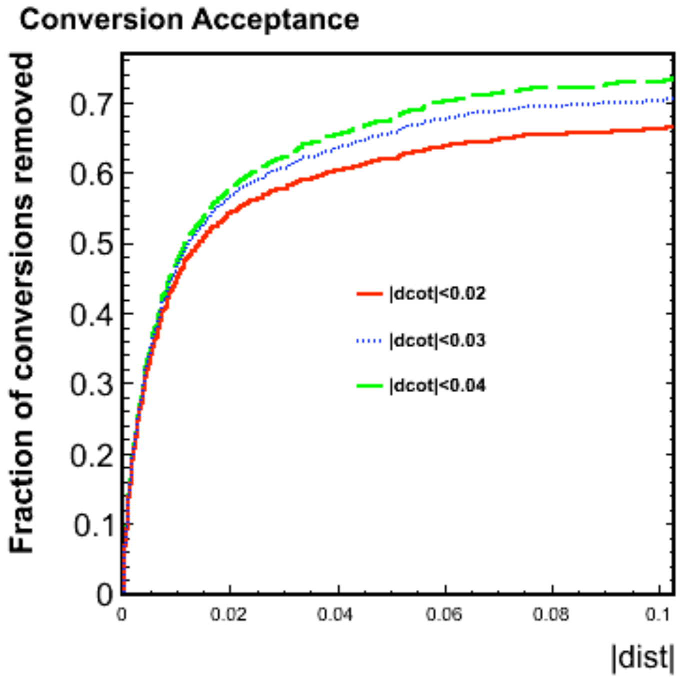
- We require that the two tracks have opposite curvature.
- Both tracks are required to be within a small cone which help decrease signal rejection with little effect on conversion rejection.
- For a conversion we expect the variables **"dist"** and **ΔCot(ϑ)** to be very small where:

- "**dist**" is the 2D distance at the point where the tracks are parallel
- **ΔCot(ϑ)** is the difference in the cotangents of theta of each track.



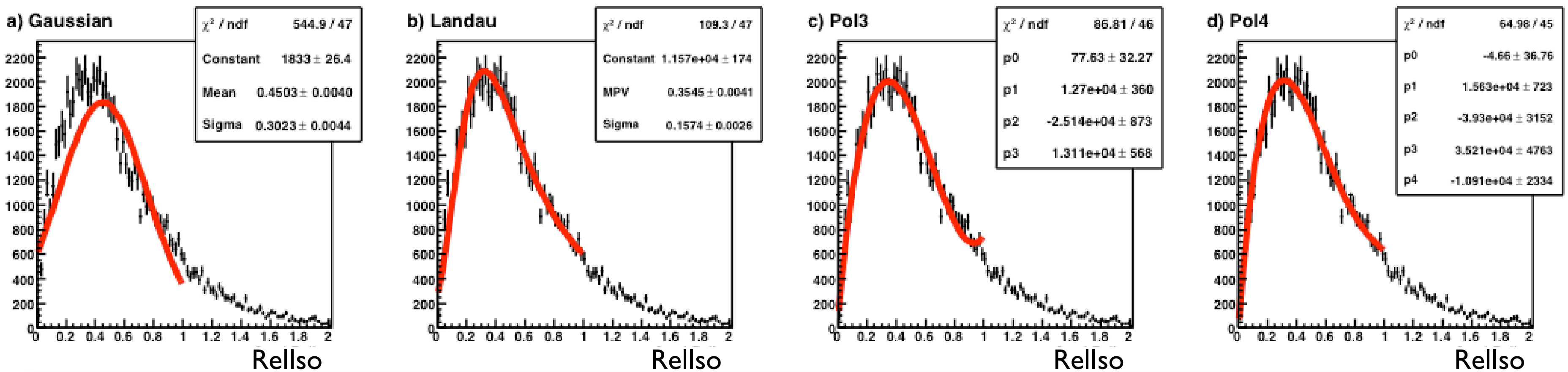
Conversion Algorithm





QCD determination

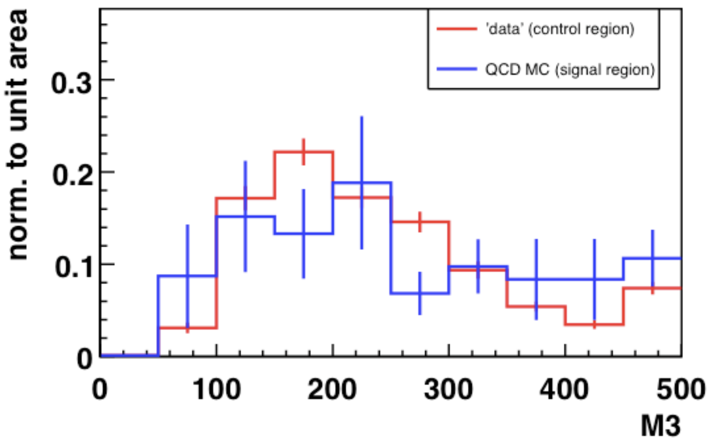
- To order to put some control on the QCD extrapolation, we have explored a data-driven approach to find the shape to use.
- We have investigated event selection criteria which reject signal and increase the QCD level of dominance in the RelIso distribution.
- This allows to see the behaviour of QCD at low isolation, and to find the most appropriate functional form to use in the extrapolation method.
- We can look at various functional forms, and then apply this to the data sample.



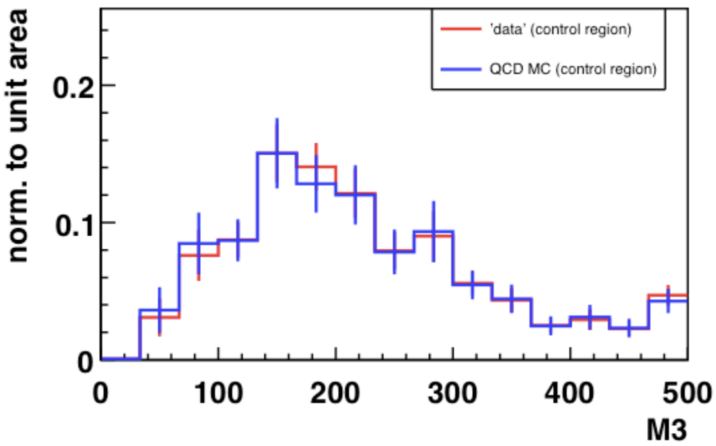
QCD M3 Template

We take the M3 shape for QCD in the signal based on data in the control (inverted RelIso) region.

3-jet bin: Comparison of QCD in the signal region to all the samples combined in the control region.



4-jet bin: Comparison of QCD alone in the control region to all samples combined in the control



M3 Method

- We use a maximum likelihood method with **four parameters**:
- $N_{t\bar{t}}$, $N_{W/Z}$, $N_{singletop}$, N_{QCD}.
- The templates for $t\bar{t}$, W+Jets (used for W+Jets and Z+Jets) and single top are derived from Monte Carlo.
- The QCD template is taken from data from the control region (inverted RelIso cut).
- $N_{singletop}$ is contrained with a Gaussian with the Monte Carlo expectation as mean and a width of (taking 30% uncertainty on the theoretical cross section into account).
- N_{QCD} is constrained with a Gaussian with the QCD-estimation as mean and a width of 50% (based on the level of uncertainty we apply to the QCD method).

Systematics

- The following sources of systematic uncertainties have been studied:
- **Jet energy scale:** Multiply the Four-vectors of all jets by 1.1 (+10% JES) or 0.9 (-10% JES)
- **MC generator:** Different MC generators are used to compare the effects of different modeling.
- **ISR/FSR:** Used $t\bar{t}$ samples with either less or more gluon radiation
- **W+Jets MC factorization scale:** Used W+Jets samples with varied Q^2 scale (varied by factor 0.5 or 2.0)
- **W+Jets MC matching threshold:** Used W+Jets samples with different matching threshold (5 or 20 GeV/c, default is 10 GeV/c)
- **Singletop Shape uncertainty:** Alter M3 shape by enhancing either the fraction of the tW-channel or the fraction of the t-channel.
- **QCD Shape uncertainty:** Take only QCD events from the control region for the pseudodata, and not the data-driven template from the control region.
- **PDF uncertainty:** Using the CTEQ6.6 PDF set and the LHAPDF package using a re-weighting technique. The errors for the set of 22 eigenvectors are added in quadrature.

JES Systematics

Jet energy scale: Multiply the Four-vectors of all jets by 1.1 (+10% JES) or 0.9 (-10% JES)

Two effects:

1) The acceptance is altered for all processes
2) The shape of the M3 distribution is altered for all processes.

Both effects have to be taken into account:

- ➡ Change the mean value for the pseudo-data for each process according to the shifted numbers due to the JES variation.
- ➡ Use the altered M3 distributions for all processes to smaple the pseudo data from, while we fit with the default templates.

Effect on the $t\bar{t}$ cross section ~ 15%