# Measurement of the top quark pair production cross-section at 5.02 TeV

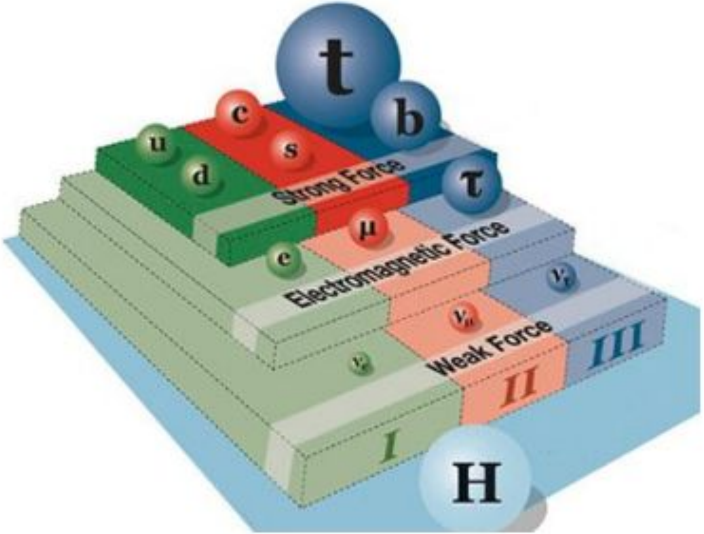
CMS DAS Pisa

1 February 2019

Why the top quark?

- The **most massive particle** in the SM - the largest Yukawa coupling!

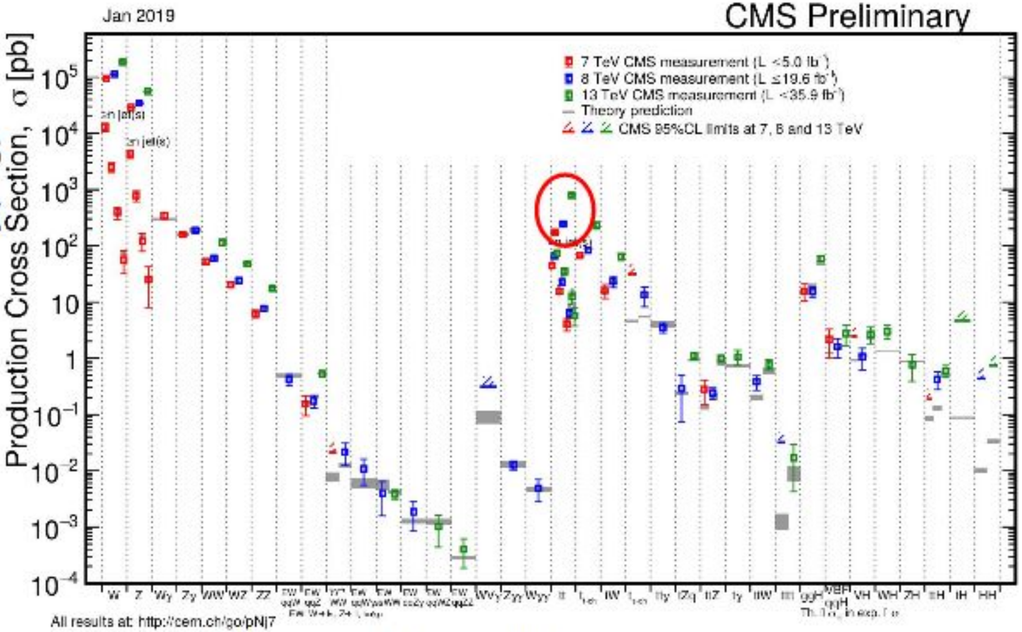
  Used to probe consistency between $\mathbf{M_H}$, $\mathbf{M_W}$, $\mathbf{M_t}$.



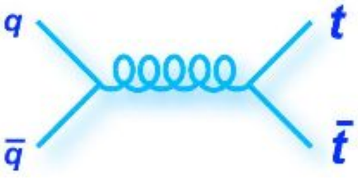

- The only quark that **decays before hadronizing**: best candidate to study QCD predictions!

# Top quarks in proton-proton collisions

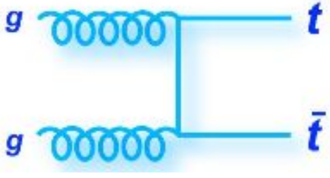
The LHC is a top quark factory!



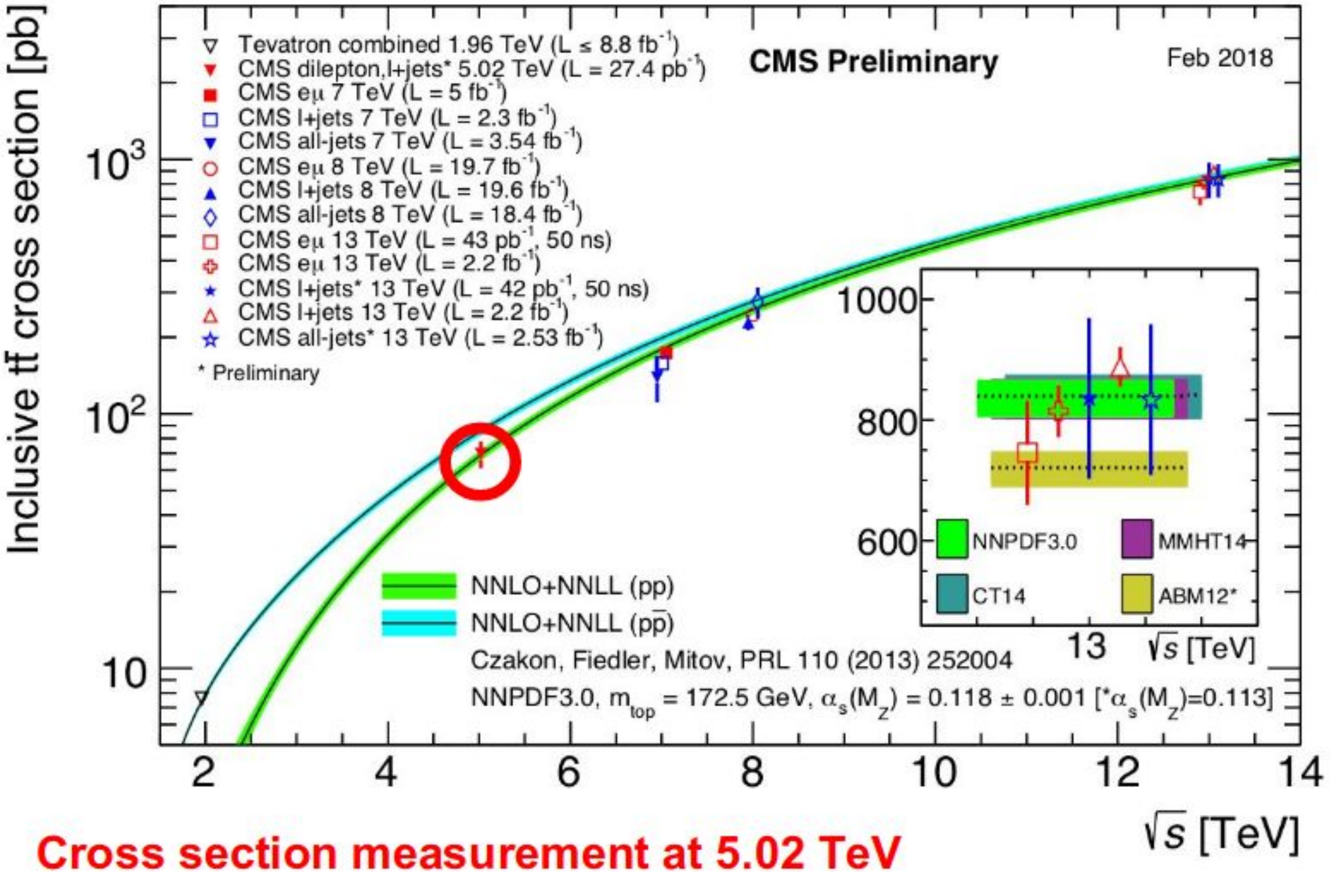

Top quarks are mainly produced in pairs:





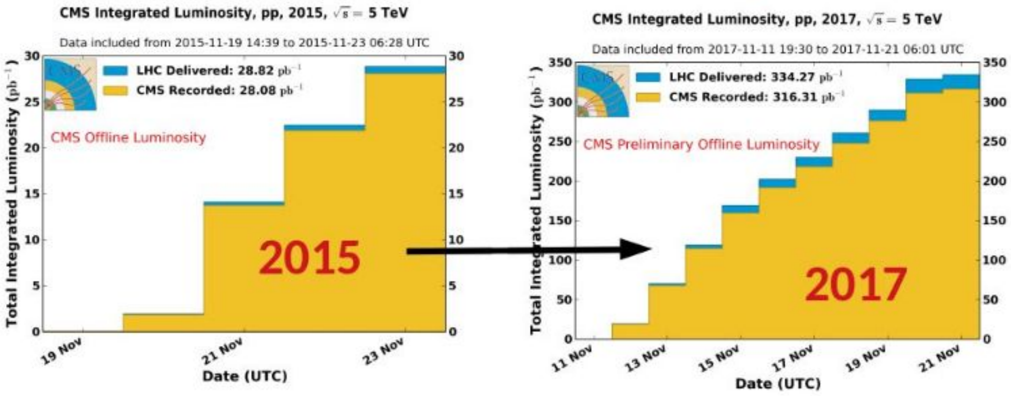


# Top pair production at CMS

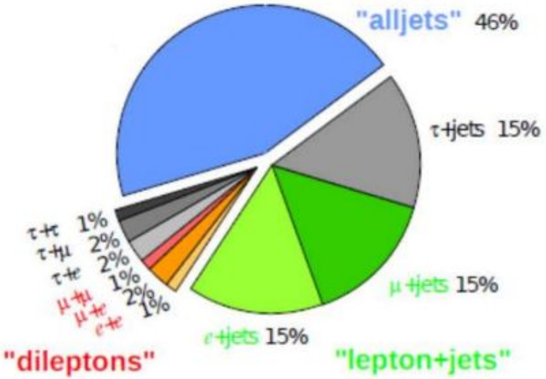



**Cross section measurement at 5.02 TeV**

## A factor ~10 on luminosity




## Where to look for top quarks?




Dileptons: small but very clean!

e/μ+jets: not so clean but higher statistics.

# tt - dilepton selection

At least 2 leptons, with at pT > 12 GeV, Eta < 2.4, dxy < 0.05, dz < 0.1

Muons: Tight ID, tight ISO, RelIso04 < 0.15

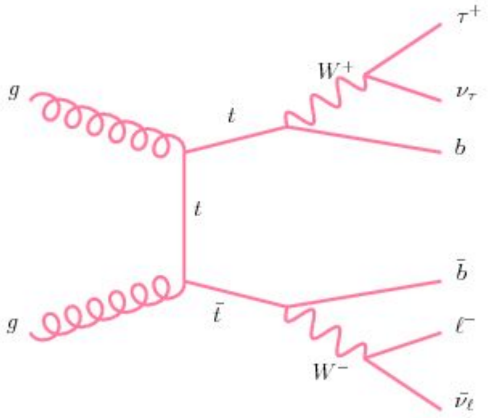
Electron: Tight cut-based Id, convVeto, RelIso03 tight

Total Charge of the most energetic leptons: -1

Pt of leading lepton > 20 GeV/c

InvMass(ll) > 20 GeV

At least 2 Jet with Pt > 25 GeV, |eta| < 2.4




**ee - µµ selection:**

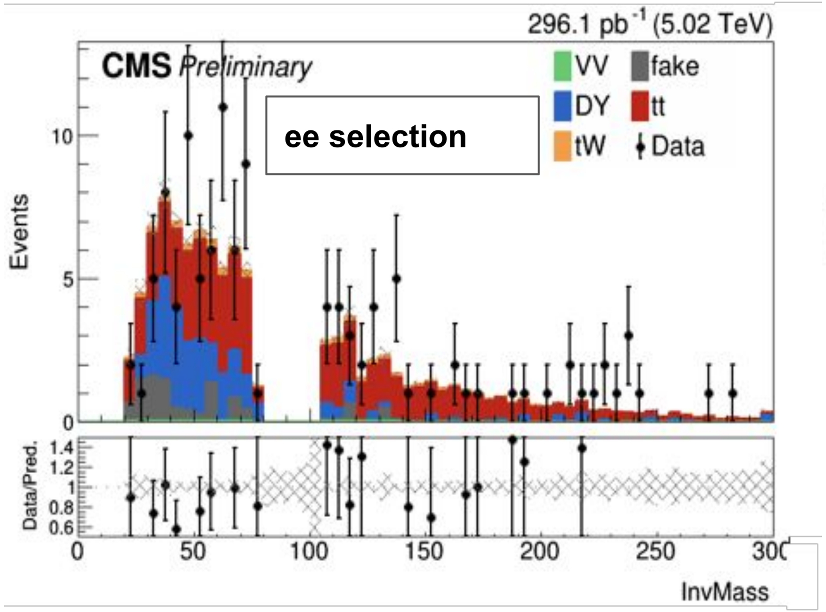
|InvMass - MassZ| > 15

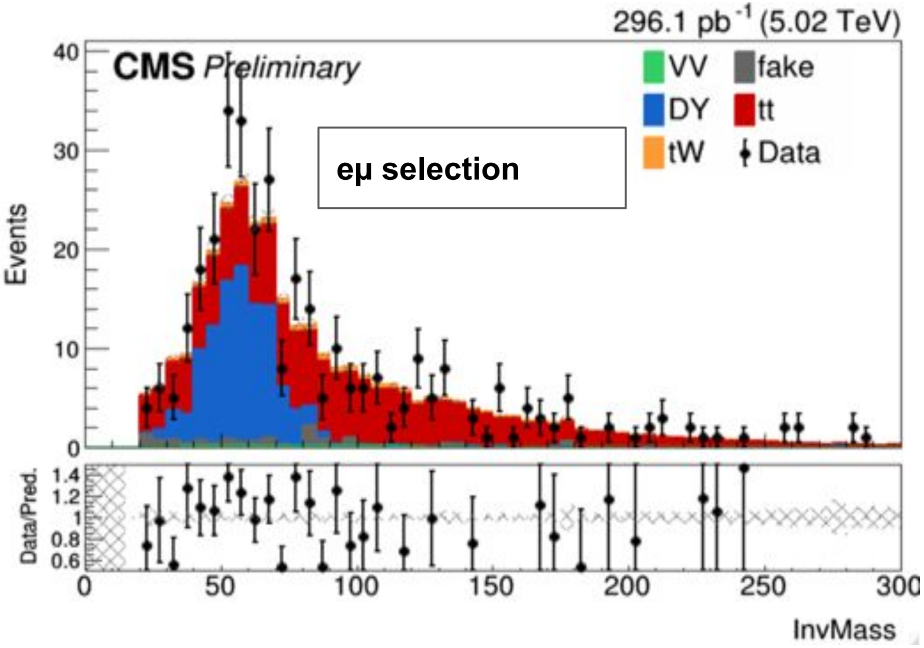
Missing ET 35 GeV

**Triggers:**

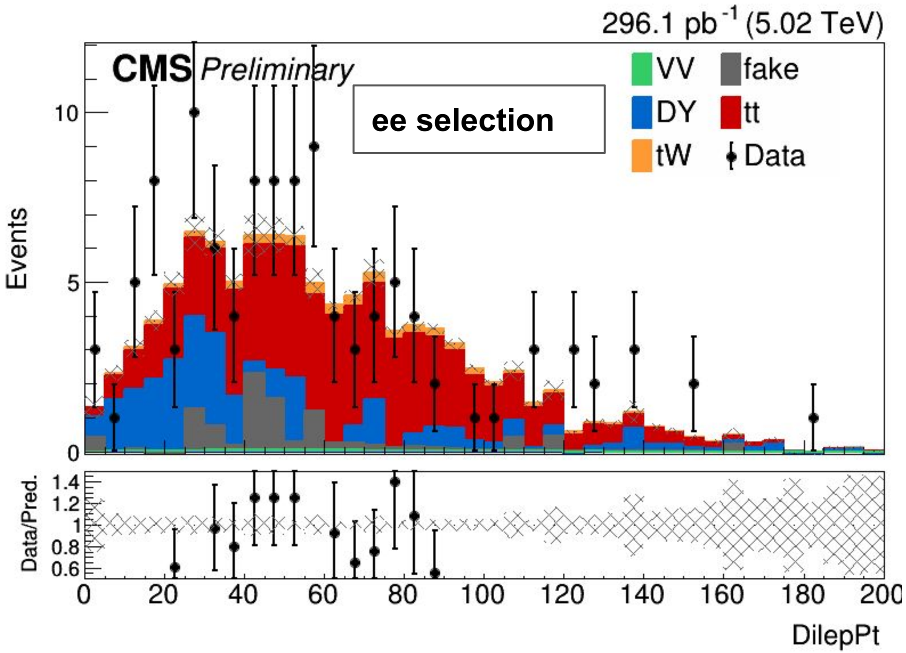
SingleMuon, DoubleMuon and HighEGJet

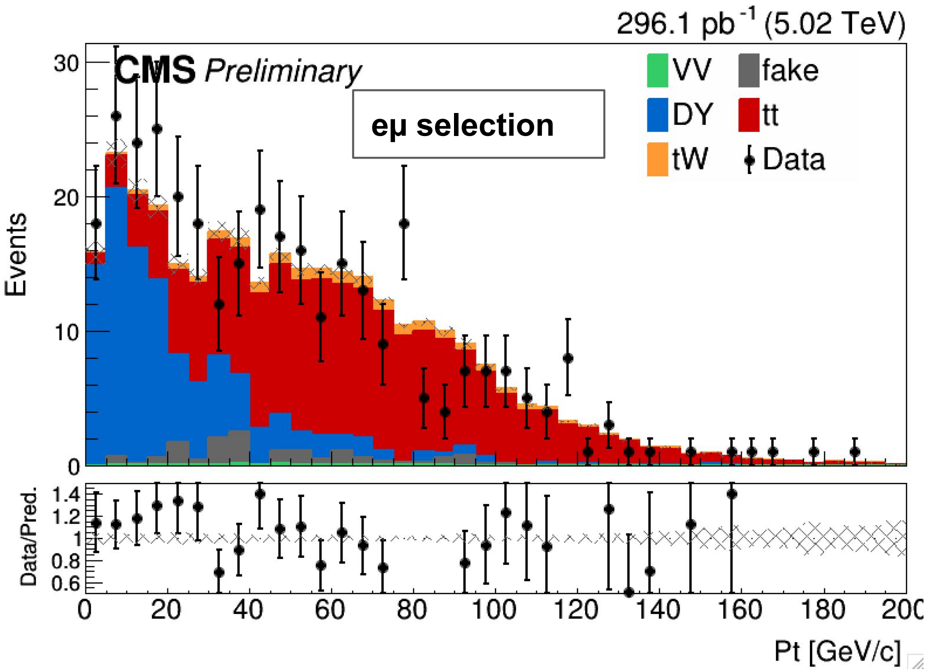
# Dilepton invariant mass distribution



# Dilepton Pt


296.1 pb$^{-1}$ (5.02 TeV)
CMS Preliminary
ee selection
VV
fake
DY
tt
tW
Data
Events
Data/Pred.
DilepPt


296.1 pb$^{-1}$ (5.02 TeV)
CMS Preliminary
eμ selection
VV
fake
DY
tt
tW
Data
Events
Data/Pred.
Pt [GeV/c]

# Drell-Yan background estimation

The Drell-Yan background can be estimated using the R*out/in* method.

Events inside the Z peak are counted in ee/μμ channels.

The non DY contribution inside the peak is estimated as half the contribution from the eμ channel.

The DY background estimate outside the peak:

$$N_{out}^{\mu\mu} = R_{out/in}^{\mu\mu}(N_{in}^{\mu\mu} - 0.5N_{in}^{e\mu}k_{\mu\mu})$$

$$R_{out/in} = \frac{N_{out}^{\mu\mu,DY}}{N_{in}^{\mu\mu,DY}} \qquad k_{\mu\mu} = \sqrt{\frac{N_{in}^{\mu\mu}}{N_{in}^{ee}}}$$

**Work (was) in progress...**

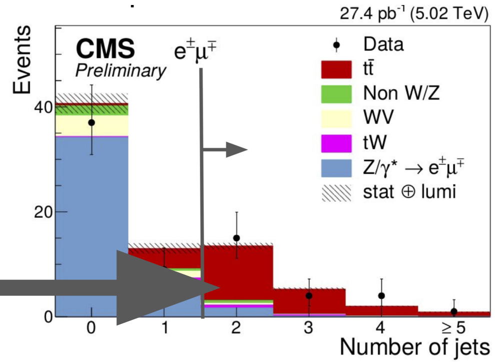
# Dilepton channel: systematic uncertainties

**Counting experiment**: count the number of observed events over the background expectation and extrapolate to the full phase space.

Estimated signal strength only in eμ channel:

(N_obs-N_bkg)/N_sim = 0.4648 +- 0.0761

Bkg estimated from MC to be approx 10%




**Basic idea:** recompute the yield after varying the scale factors $\pm 1\sigma$

- **experimental uncertainties** muon & electron SFs, jes&jer corrections, PU
- **modeling uncertainties** hdamp, underlying event tune, matrix elements matching to PS

**Yield variations** (wrt. nominal) were calculated in the ee, eμ and μμ channels.

# Experimental uncertainties

Several inputs:

- **Jet energy scale and resolution**: since the jet energy cannot be reconstructed perfectly, corrections on both scale and resolution are applied. To compute the systematics we variate the corrections by the corresponding uncertainties.
- **Electrons and muons selection**: in order to match MC to data, SF are applied. To compute the systematics we variate the SF according to the corresponding uncertainties.
- **PU reweighting**: MC should be corrected with the pileup observed in data applying PUSF. The systematic uncertainties are evaluated as above.

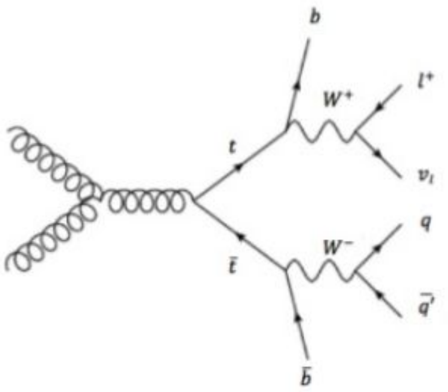
No b-tagging efficiency was computed since no b selection was applied to data (in order to increase the yield, pure tt sample for nJets > 2 anyway).

# Modeling uncertainties

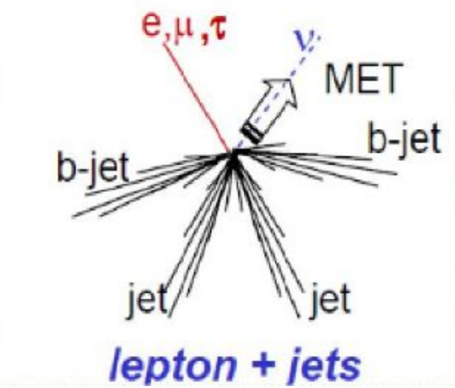
- **ME scales**: MC was computed certain values for the renormalization and factorization scales. Systematics were computed by changing the renormalization and factorization scales by factors of 0.5 and 2, while avoiding unphysical variations. Uncertainty is given by the maximum variation.
- **Underlying event tune and ME to PS matching**: since MC were calculated with an estimation of the Minimum Bias events (tune) and a model for the ME to PS matching. Uncertainties were estimated by running the very same selection on different MCs with different tunes and matchings.
- **PDFs and alpha strong**: different sets of proton pdfs were combined. Systematics computed according to different variations of the pdfs. Uncertainty is given by the maximum variation on the final yield.

| Systematic | yield / nominal yield (ee) | yield / nominal yield (eμ) | yield / nominal yield (μμ) |
|---|---|---|---|
| Electron ID | | 1.0328 0.9672 | |
| hdamp | 0.6% (up) 0.8% (down) | 1.3% (up) 0.3% (down) | 1.6% (up) 0.9% (down) |
| Jet energy scale | 0.6% (up) 0.1% (down) | 0.6% (up) 0.2% (down) | 0.4%(up) 1.7% (down) |
| Jet energy resolution | 0.3% | 0.2% | 0.2% |
| Muon ID | | 1.01% (up) 0.99 (down) | |
| Pile Up | | 1.00% (up) 0.99% (down) | |
| TuneCP5 | 1.8% (up) 0.6% (down) | 0.9% (up) 1.3% (down) | 0.5% (up) 0.2% (down) |
| PDF | | 1.07% (up) 1.03% (down) | |
| alpha_s | | 1.02% (up) 1.07% (down) | |
| ME | | 1.14% (up) 0.86% (down) | |
| Lumi | | 1.04% (up) 0.97% (down) | |

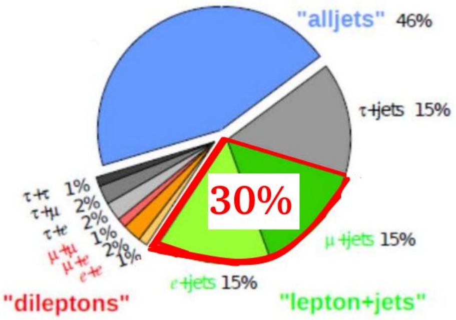
# Single Electron / Muon + jets channel




Theorist's perspective




Experimentalist's perspective




**Pros**

- Large branching ratio
  -> High statistics

**Cons**

- Large background : QCD and W+jets
- Requires a good b-tagging

**Main uncertainties**

- b-tagging efficiency
- Jet energy scale

# Events selection

## Muon

- Tight ID
- Tight isolation : PF relative isolation (RelIso04) < 0.15
- Pile-up removal : $d_{xy}$ < 0.05 and $d_z$ < 0.1
- $P_T$ > 30 GeV and $\eta$ < 2.1

## Electron

- Tight ID
- Tight isolation : PF relative isolation (RelIso03) < 0.0361
- Pile-up removal : $d_{xy}$ < 0.05 and $d_z$ < 0.1
- $P_T$ > 30 GeV and $\eta$ < 1.479

## Jets

- Tight ID
- $P_T$ > 30 GeV and $\eta$ < 2.1
- DeepCSV for b-tagging : medium working point (0.4941)

# Background estimation

## W + Jets

Estimated from MC

## QCD

- Data driven estimation
- Control region with inverted isolation
- Use the MET to extract the signal region

$$N_{\mathrm{SR}}(QCD) = [N_{\mathrm{CR}}(obs) - N_{\mathrm{CR}}(non-QCD)] \cdot \frac{N_{\mathrm{SR}}^{E_T^{miss}<20}(obs) - N_{\mathrm{SR}}^{E_T^{miss}<20}(non-QCD)}{N_{\mathrm{CR}}^{E_T^{miss}<20}(obs) - N_{\mathrm{CR}}^{E_T^{miss}<20}(non-QCD)}$$

# Results … Well … almost there

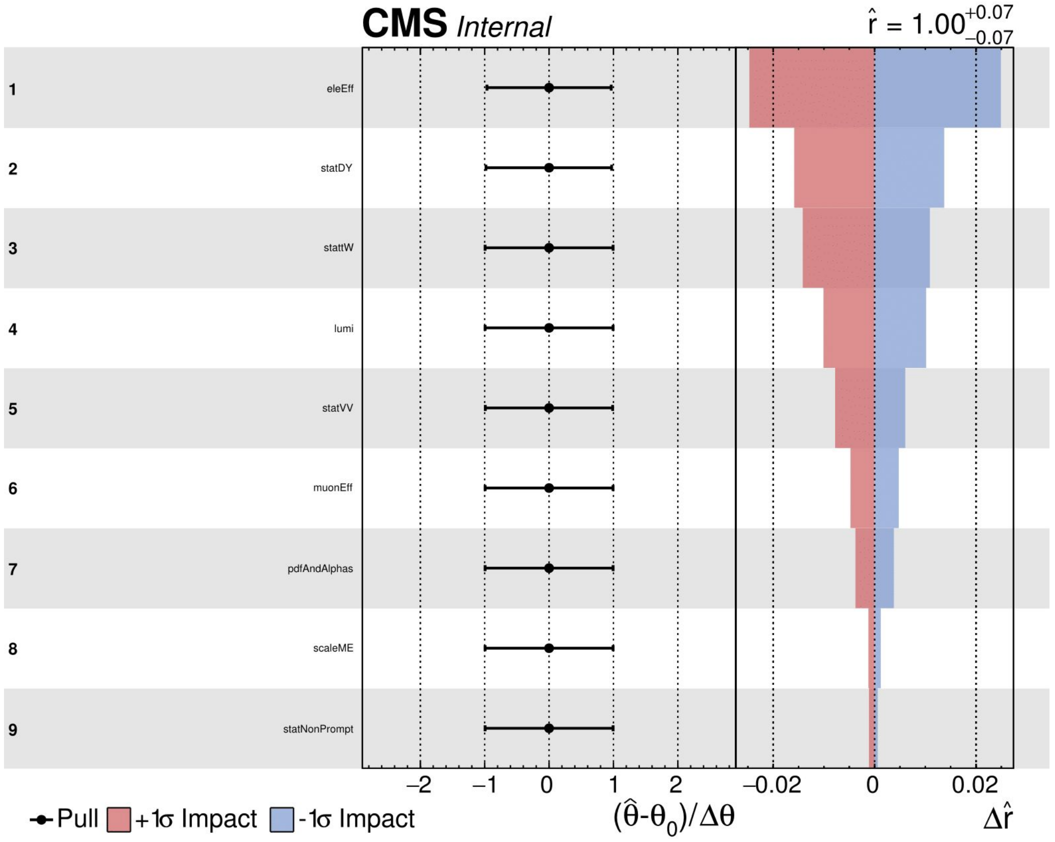
## What we did : almost from scratch

- Precuts
- Events selection
- Processing inputs
- QCD background estimation

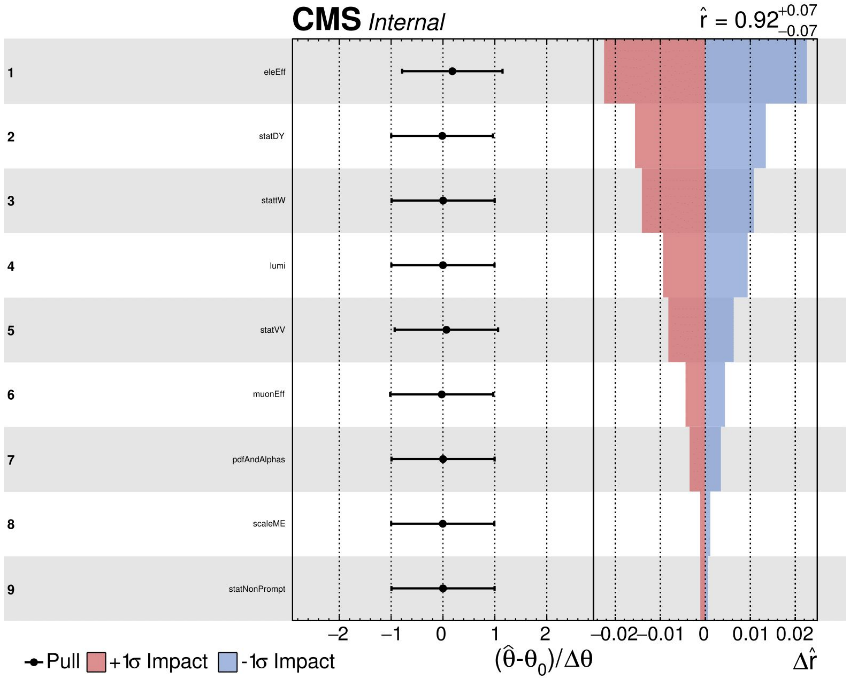
## What we still have to do :

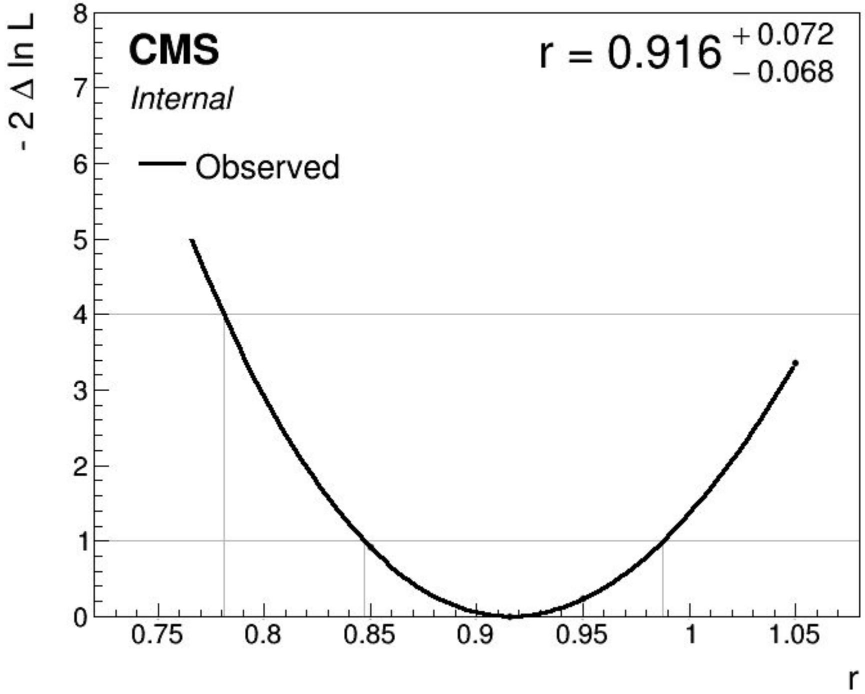
- Produce histograms
- Design Signal Region
- Cut and count analysis
- Obtain specific single lepton systematics :
    - QCD scale factors
    - Single lepton efficiency

# Checking the fitting procedure: Asimov dataset fit



# Maximum likelihood fitting to all channels



# Splitting uncertainties