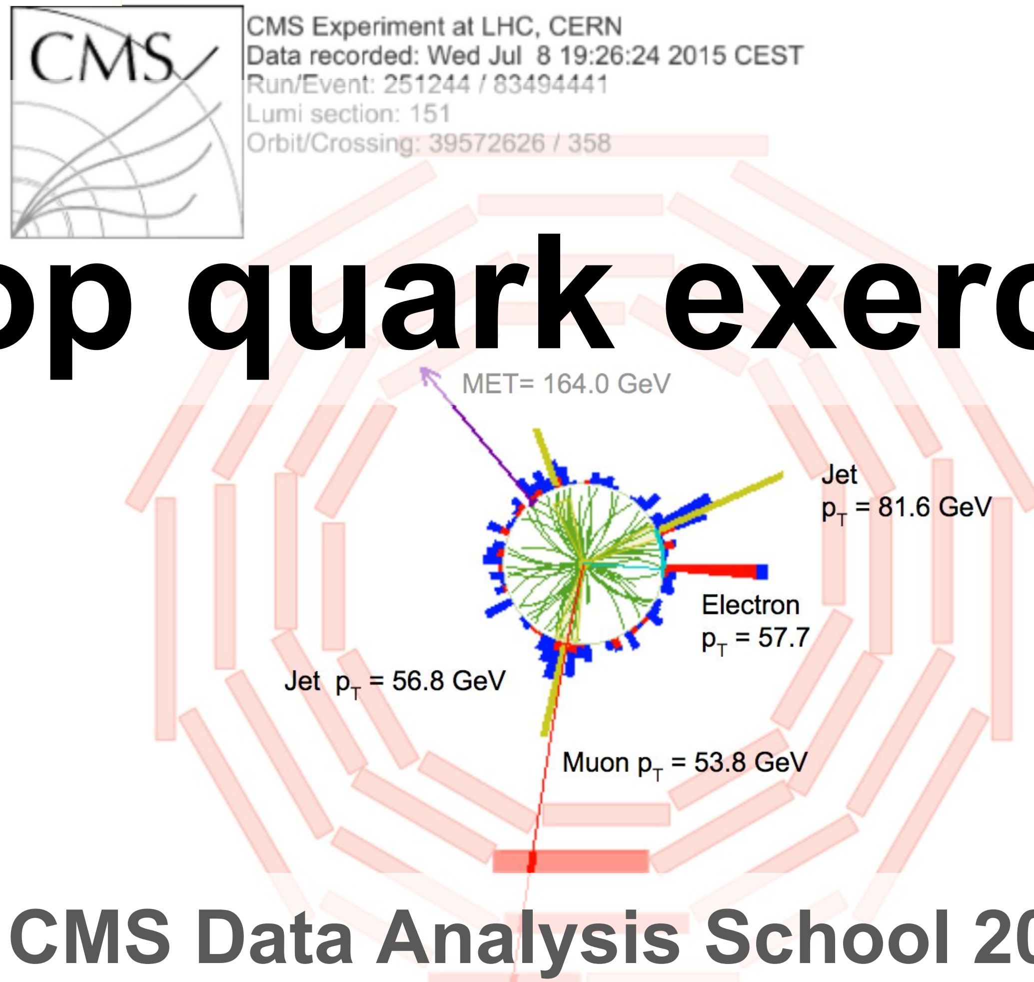
Universidad de Oviedo

UCLouvain
Institut de recherche
en mathématique et physique

Top quark exercise



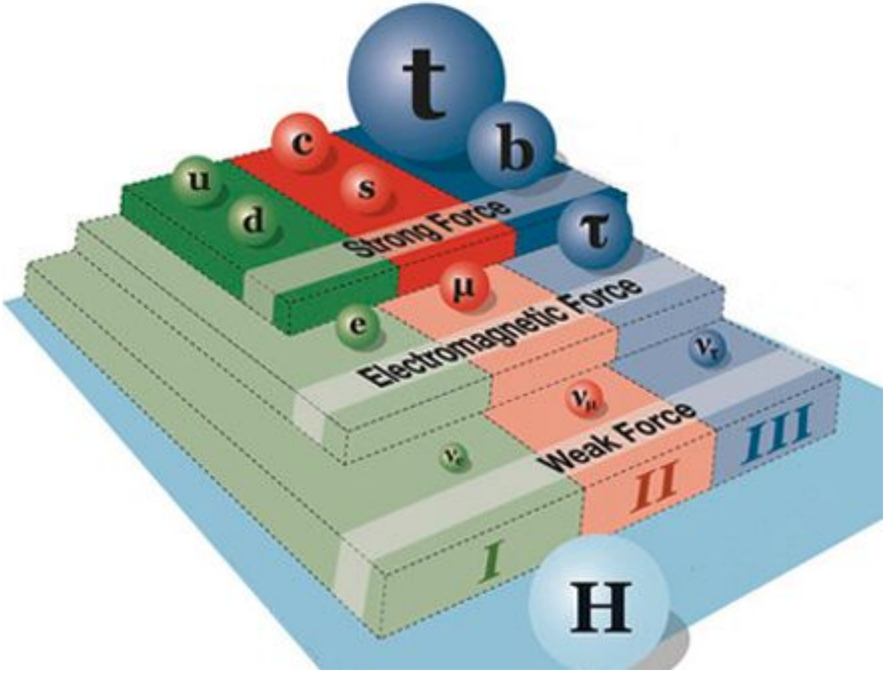
CMS Data Analysis School 2019
28 Jan - 01 Feb, Pisa

Welcome to the top quark long exercise at CMS Data Analysis School 2019 @ Pisa

Why the top quark?

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

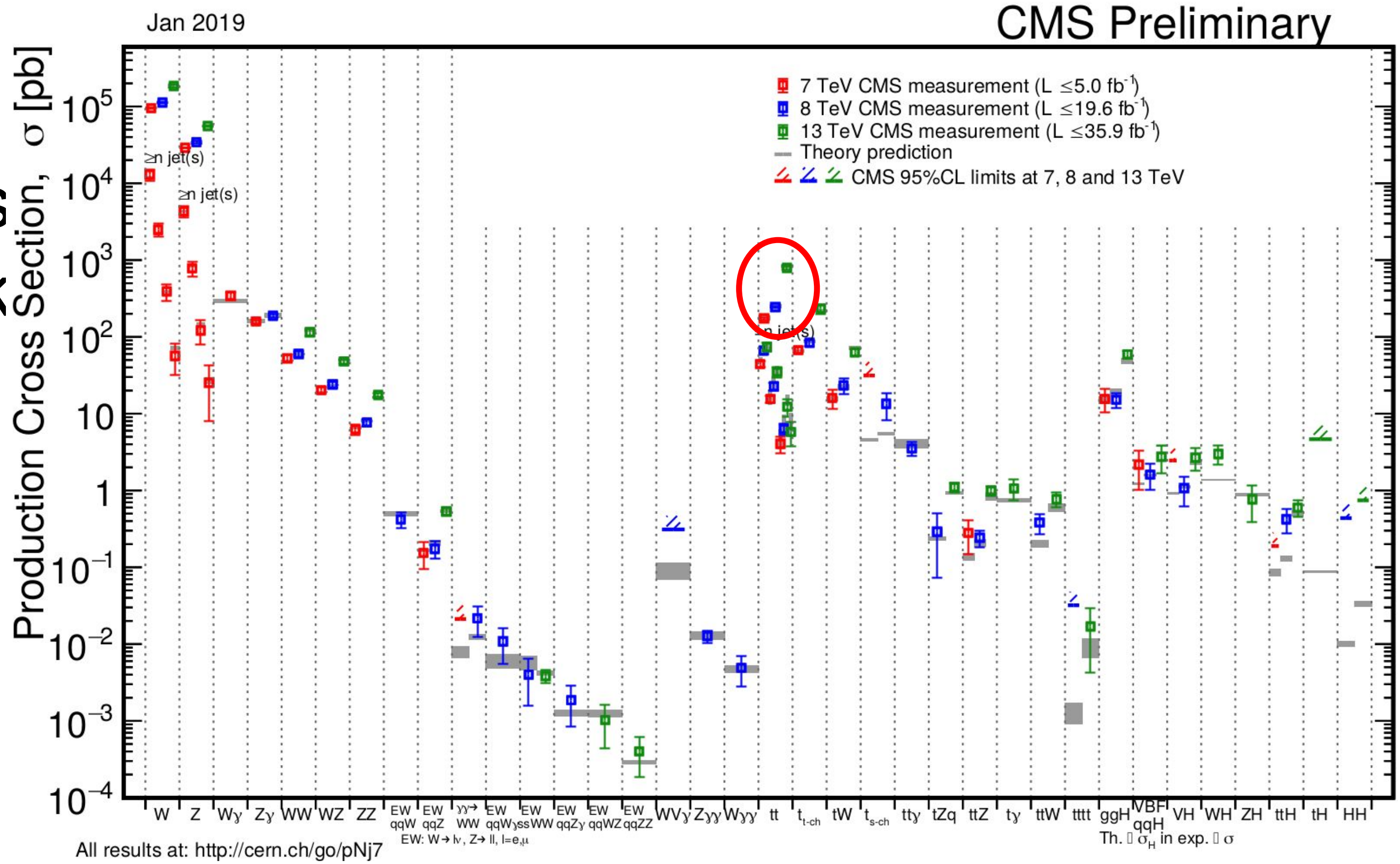
  Used to probe consistency between $\mathbf{M_H}$, $\mathbf{M_W}$, $\mathbf{M_t}$.



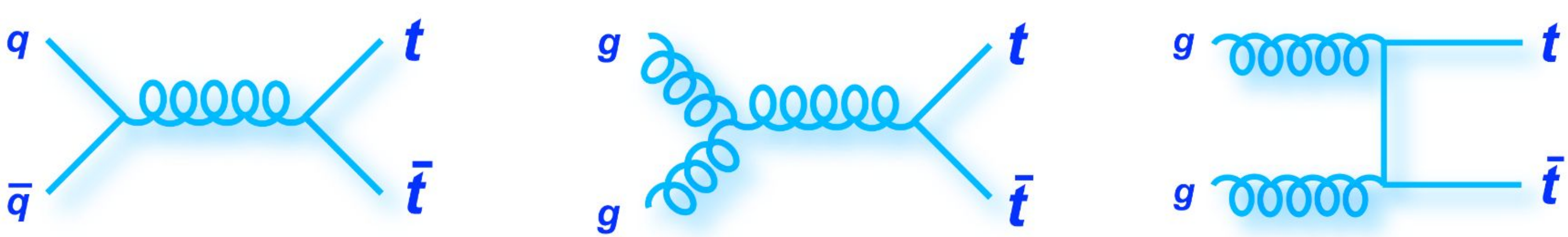
- The only quark that **decays before hadronizing**: best candidate to study QCD predictions!

Top quarks in proton-proton collisions

The LHC is a top quark factory!

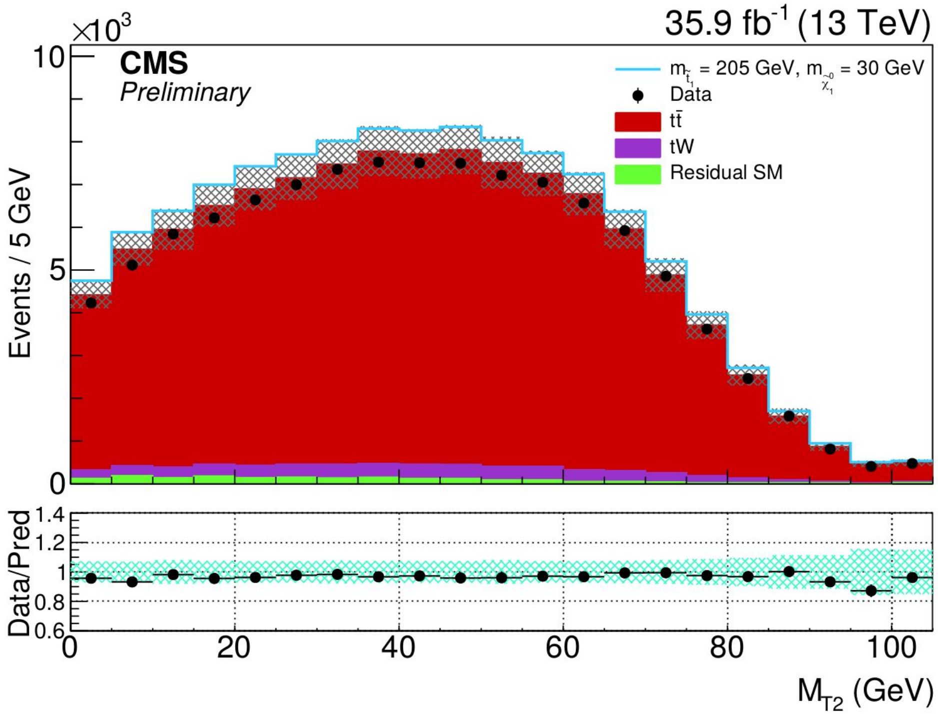


Top quarks are mainly produced in pairs:

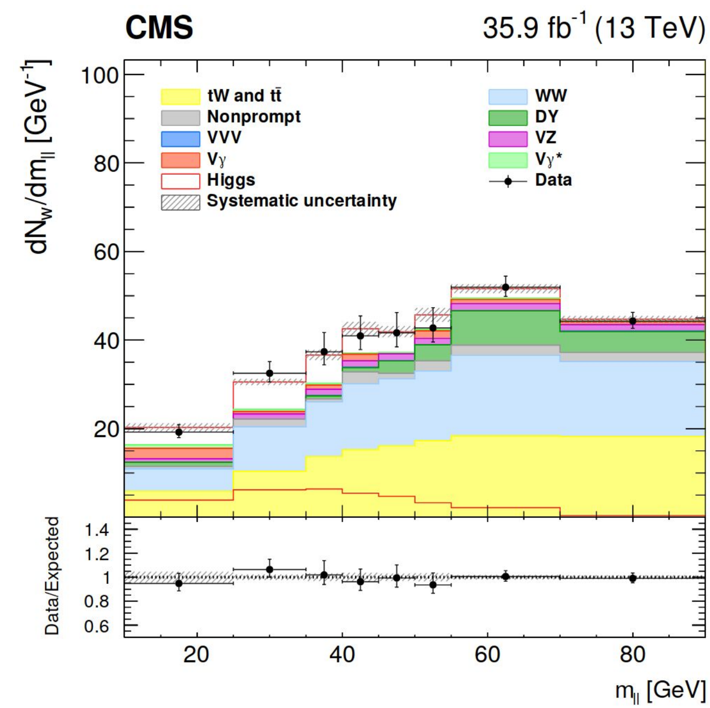


Top quarks seen from outside

Top quark pair production is normally the largest background in most BSM searches… and in other SM measurements!

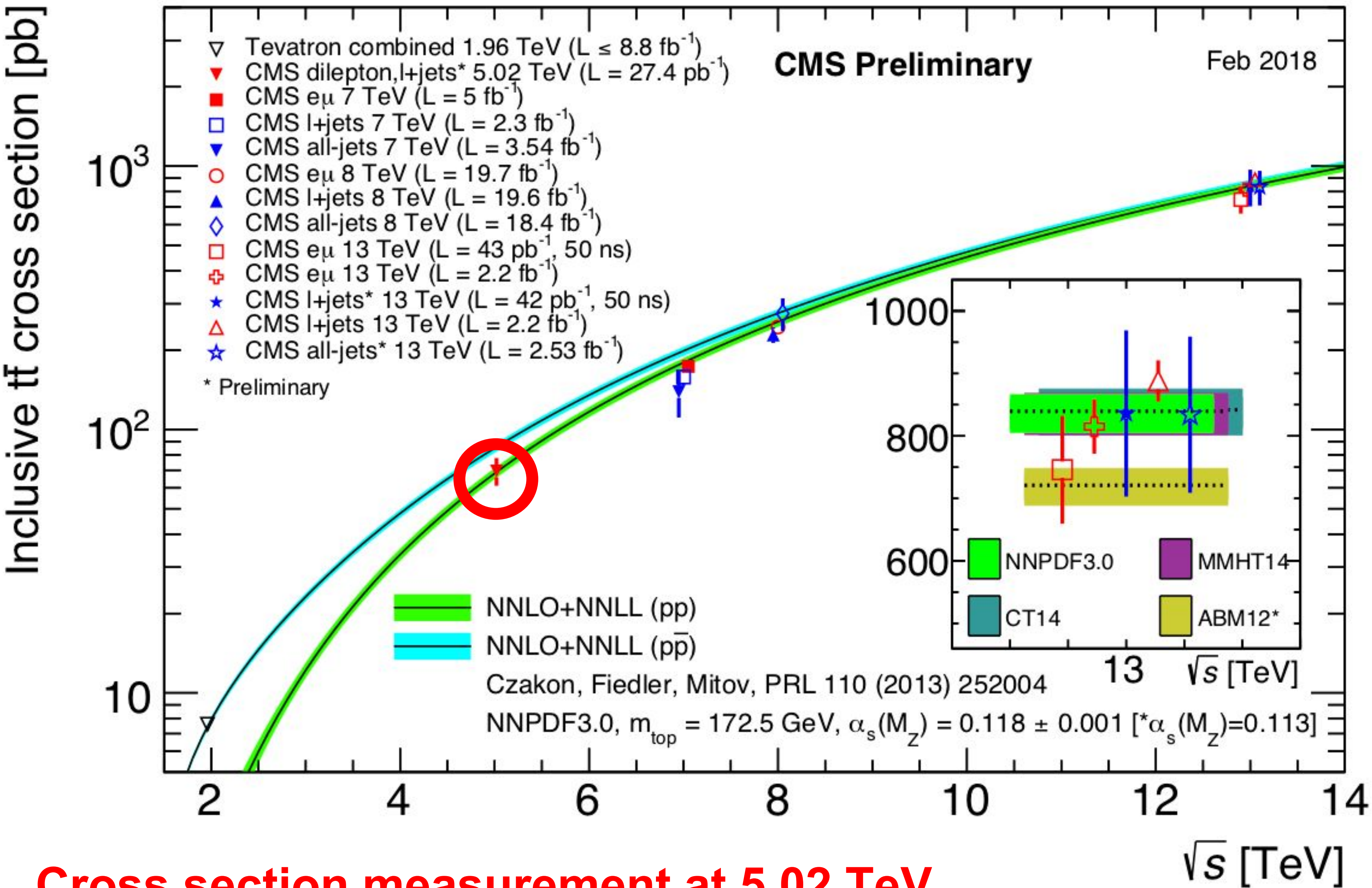


Top quarks and SUSY
SUS-18-003



Top quarks and higgs
HIG-16-042

Top pair production at CMS



Cross section measurement at 5.02 TeV

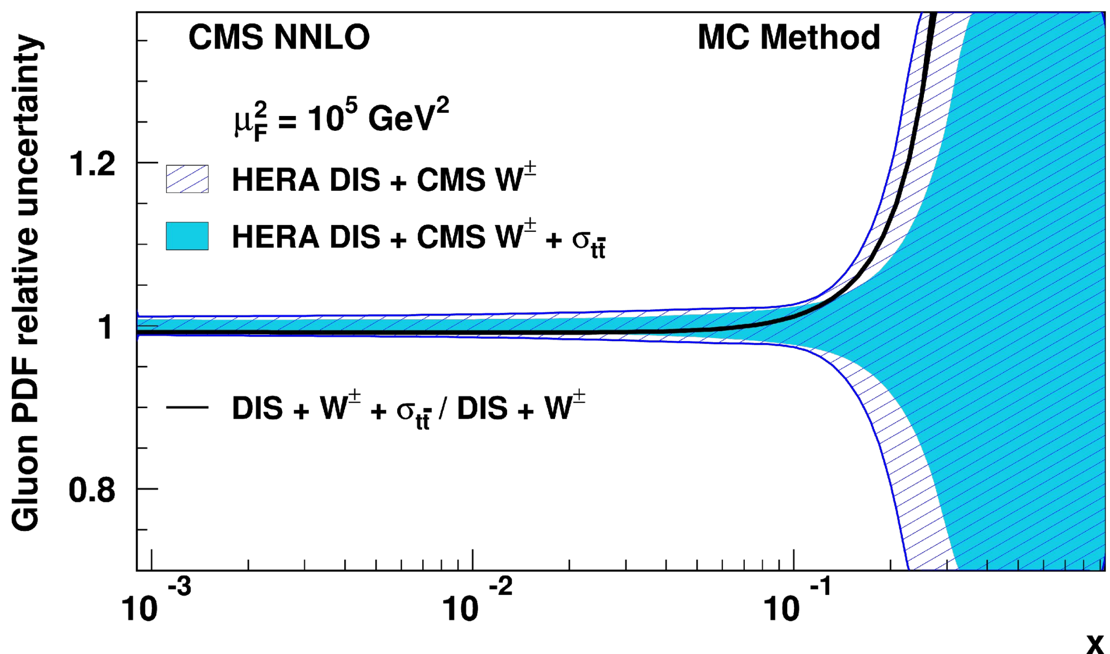
Top pair production cross section at 5.02 TeV already measured with 27.4 pb-1! JHEP 03 (2018) 115

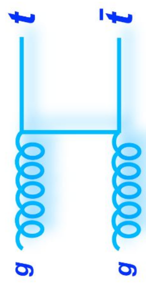
Why 5.02 TeV?

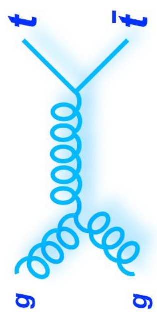
Reference pp run for HI collisions!

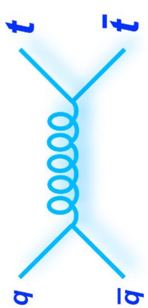
Very interesting energy to test QCD prediction: never observed in pp collisions at lower energies.

Probing high-x gluon PDF:









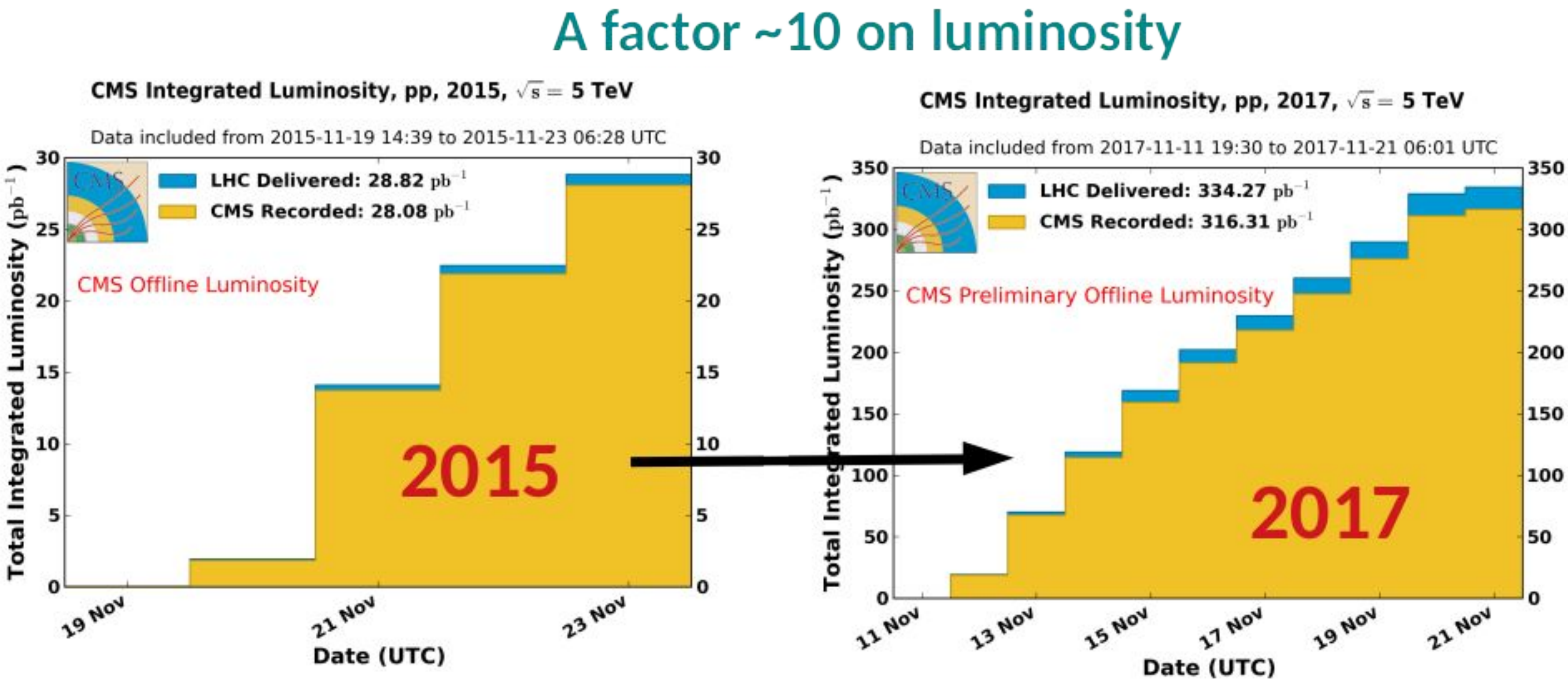
Top pair production cross section at 5.02 TeV already measured with 27.4 pb-1!

JHEP 03 (2018) 115

Summary table

				Total unc.
eμ	Cross section (pb) 76.5	Stat(pb) 18.7	Syst(pb) 4.4	25 %
μμ	Cross section 59.2	Stat 28.7	Syst 10.7	52 %
Lep+jets	Cross section 68.9	Stat 6.5	Syst 6.1	13 %
	Per Channel 76.5 59.2 68.9 Overall	Weight 0.13 0.05 0.82 **12 %**		

The Top Quark Exercise:
Measuring the tt production cross section at 5.02 TeV with the 2017 dataset.

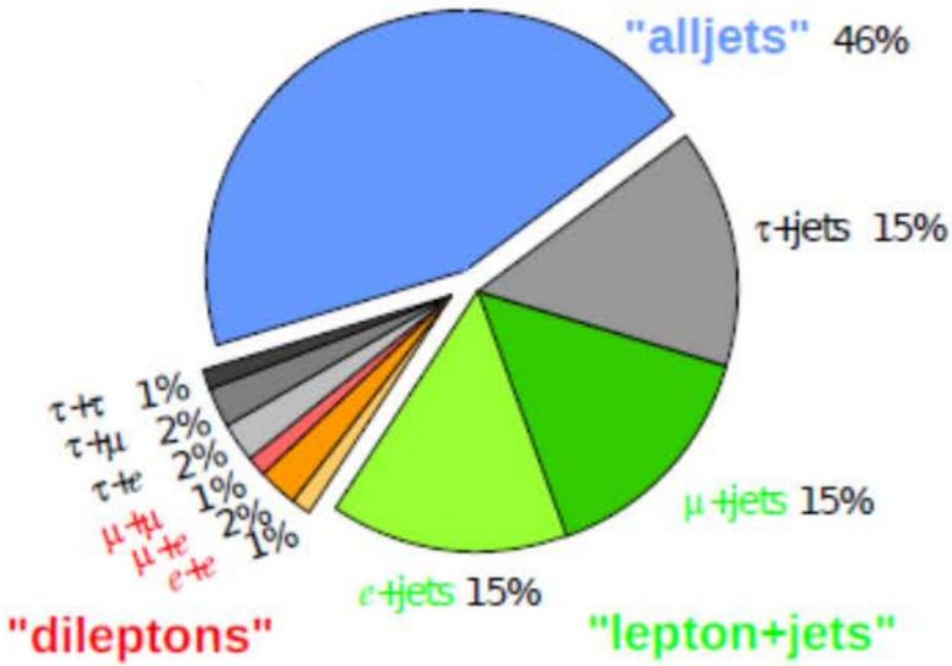


Dataset: Run2017G – ReReco 17 Nov 2017

Golden JSON: 296.08 pb-1, uncertainty of 3.5%.

https://hypernews.cern.ch/HyperNews/CMS/get/luminosity/794.html

Where to look for top quarks?



Dileptons: small but very clean!

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

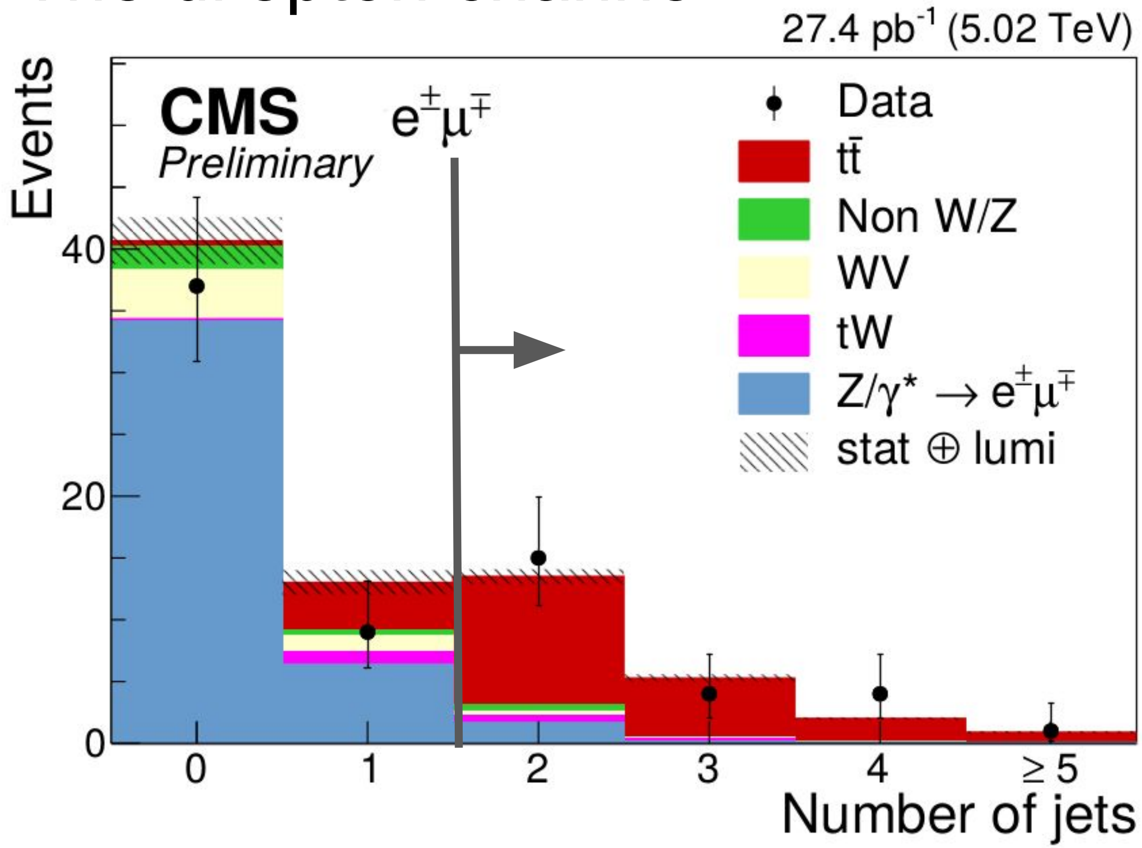
b-tagging

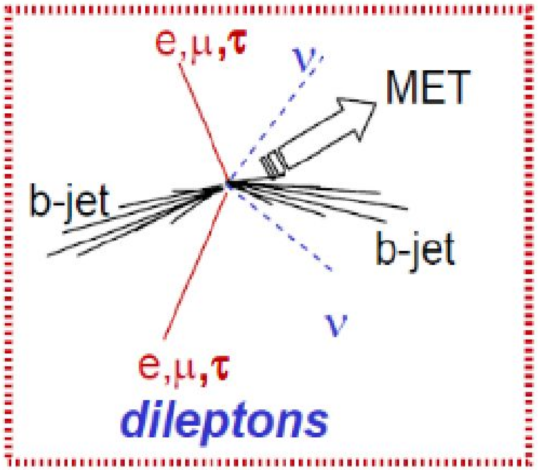
Two b quarks from the decay of the top quarks:
~high p_T b jets can be tagged (CSVv2? DeepCSV?).

➔ e/μ+jets: b-tagging is crucial!

➔ ee/μμ/eμ: dispensable

The dilepton channel





Cleanest channel:
eμ selection

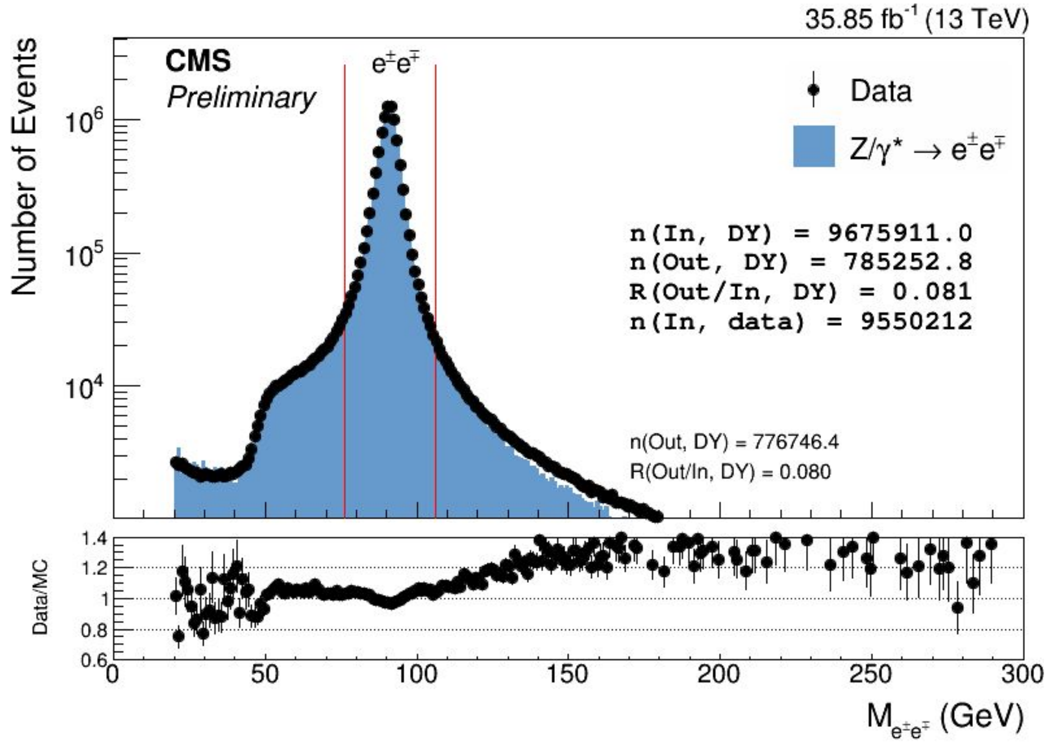
Main background processes:

★ **tW, WW/WZ/ZZ** (dibosons) → From MC simulation!

★ **DY**: from on-Z control region in data

★ **Nonprompt leptons**: from same-sign leptons in data

Drell-Yan estimate: the Rout/in method



$$N_{out}^{l^+l^-,obs} = R_{out/in}^{l^+l-}(N_{in}^{l^+l^-} - 0.5N_{in}^{e\mu}k_{ll})$$

- $\mathbf{R_{out/in}}$: the ratio between events outside and inside the Z peak (MC).
- $\mathbf{N_{in}}$: The total events in the peak (from data).
- Non-DY bkg contamination is subtracted using the eμ channel in data.
- $\mathbf{k_{ll}}$: takes into account differences between *e* and *μ* efficiencies.

Nonprompt lepton estimate

- Fake leptons background when a jet is identified as a lepton or a lepton from the decay of a hadron is taken as prompt.

Estimate from data using a same-sign control region

- Defined with the same selection but with a same-sign dilepton pair.
- Fakes mainly from W+Jets and tt with semileptonic decays.
- Prompt background subtraction from MC.

$$N_{\mathrm{data}}^{\mathrm{OS\ fakes}} = \left(N_{\mathrm{data}}^{\mathrm{SS}} - N_{\mathrm{real-pp}}^{\mathrm{SS}}\right)\frac{N_{\mathrm{MC}}^{\mathrm{OS\ fakes}}}{N_{\mathrm{MC}}^{\mathrm{SS\ fakes}}}$$

Cross section measurement

$$\sigma_{t\bar{t}} = \frac{N - N_B}{\mathcal{BR} \cdot \varepsilon \cdot \mathcal{A} \cdot \mathcal{L}},$$

Counting experiment

Count the number of observed events over the background expectation and extrapolate to the full phase space.

Fiducial cross section

Make the acceptance equal to 1: measure the cross section only with the events passing your selection (the ones you can observe!). Also called "visible" cross section → Modeling uncertainties are reduced!

Uncertainties on background estimate:

- DY and nonprompt leptons: uncertainties from the data-driven methods!: statistical (from data statistics) and systematics (MC SS prompt substraction, etc).
- tW, WW: normalization uncertainty (from the uncertainty on the best precise theoretical cross section).

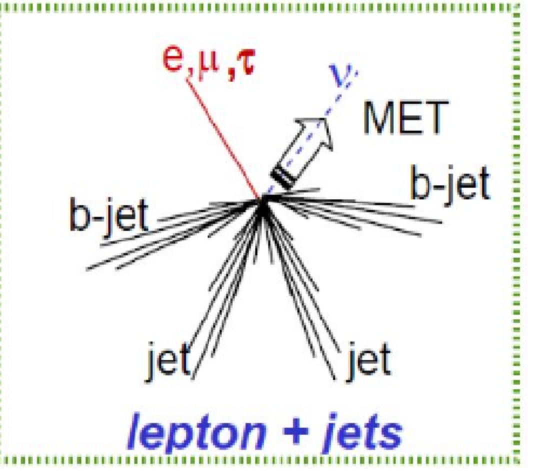
Experimental uncertainties (mainly on the efficiency):

- Jet energy scale and resolution: variating jet energy corrections by uncertainties...
- Lepton and trigger efficiencies: variating MC-to-data scale factors.
- PU reweighting? b-tagging efficiencies?

Modeling uncertainties (mainly on the acceptance)

- PDF+alpha_s:
  - Using LHE weights and following: https://arxiv.org/abs/1510.03865
  - 33 weights: nominal + 30 PDF variations + 2 alpha_s variations
- Matrix-element scales:
  - Using 9 weights (nominal + 8 variations) of μR and μF by factors 0.5 and 2.0.
  - The uncertainty is given by the maximum variation on the predicted acceptance. Variations where μR and μF go in different directions are considered unphysical.
- Underlying Event tune, hdamp:
  - Using dedicated samples with proper variations.

The e/μ+jets channel



Large backgrounds, difficult to keep under control:

- QCD with a nonprompt lepton
- W+Jets

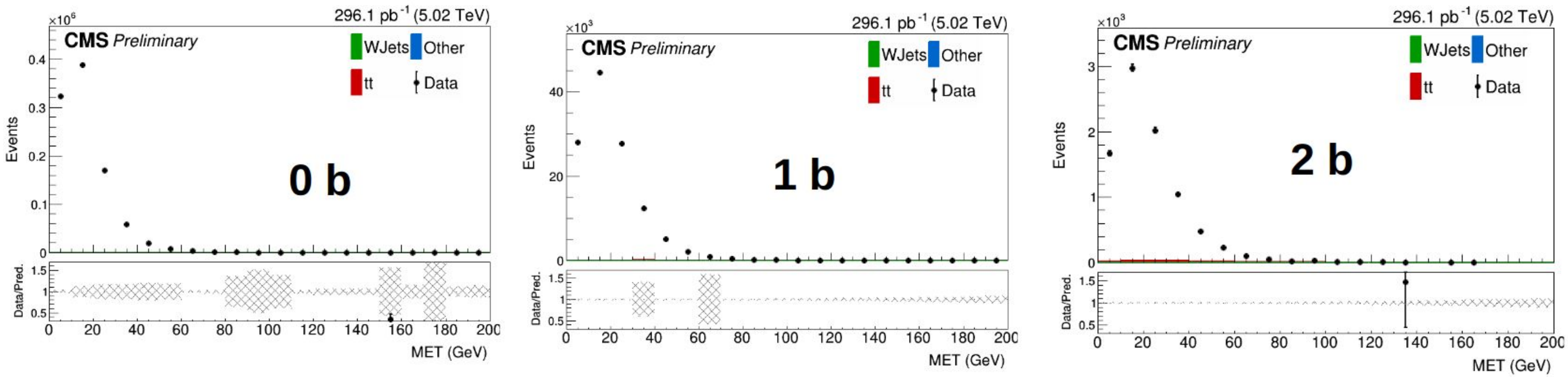
The QCD background is usually estimated from data using QCD events with non-isolated leptons (see next slide).

The W+Jets background can be estimated from MC and constrain the uncertainties using events with lower (b) jet multiplicity.

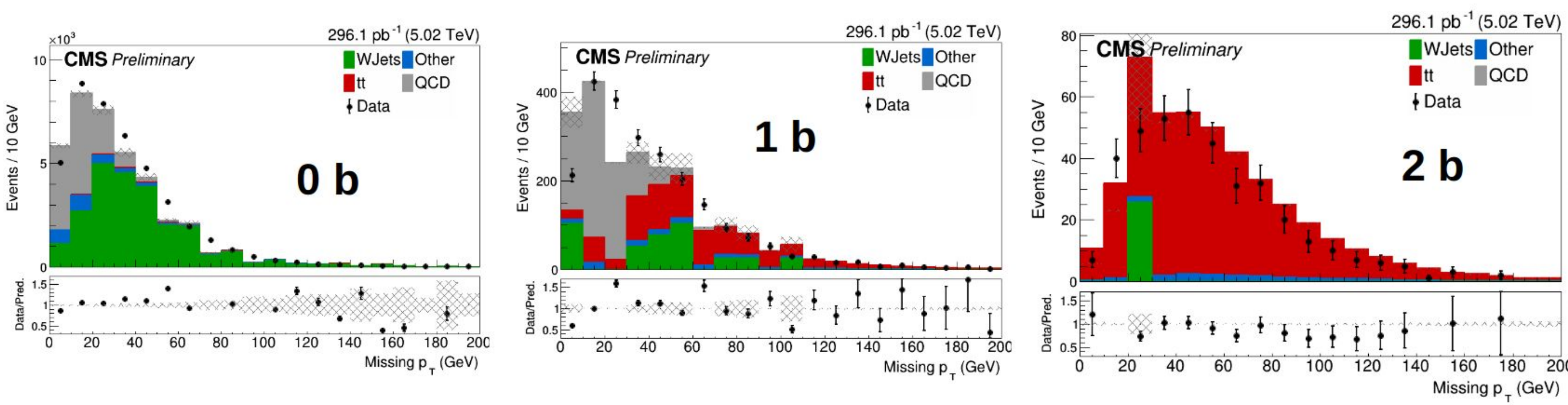
<u>QCD estimate in data</u> → Extrapolating from MET < 20 GeV.

Take a look to AN-2016/230

Non isolated muons

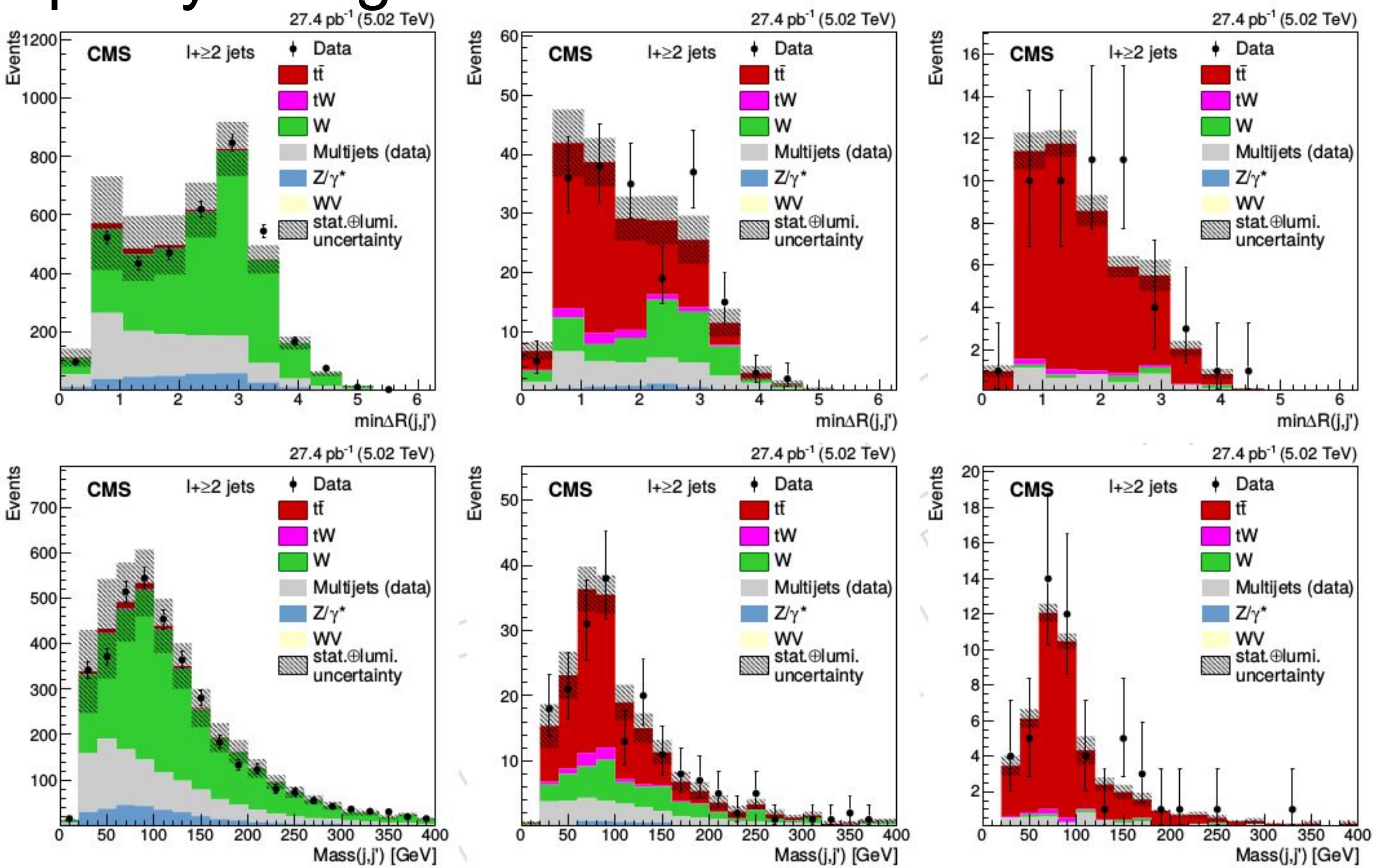


Isolated muons



Extracting the signal: a PRL fit

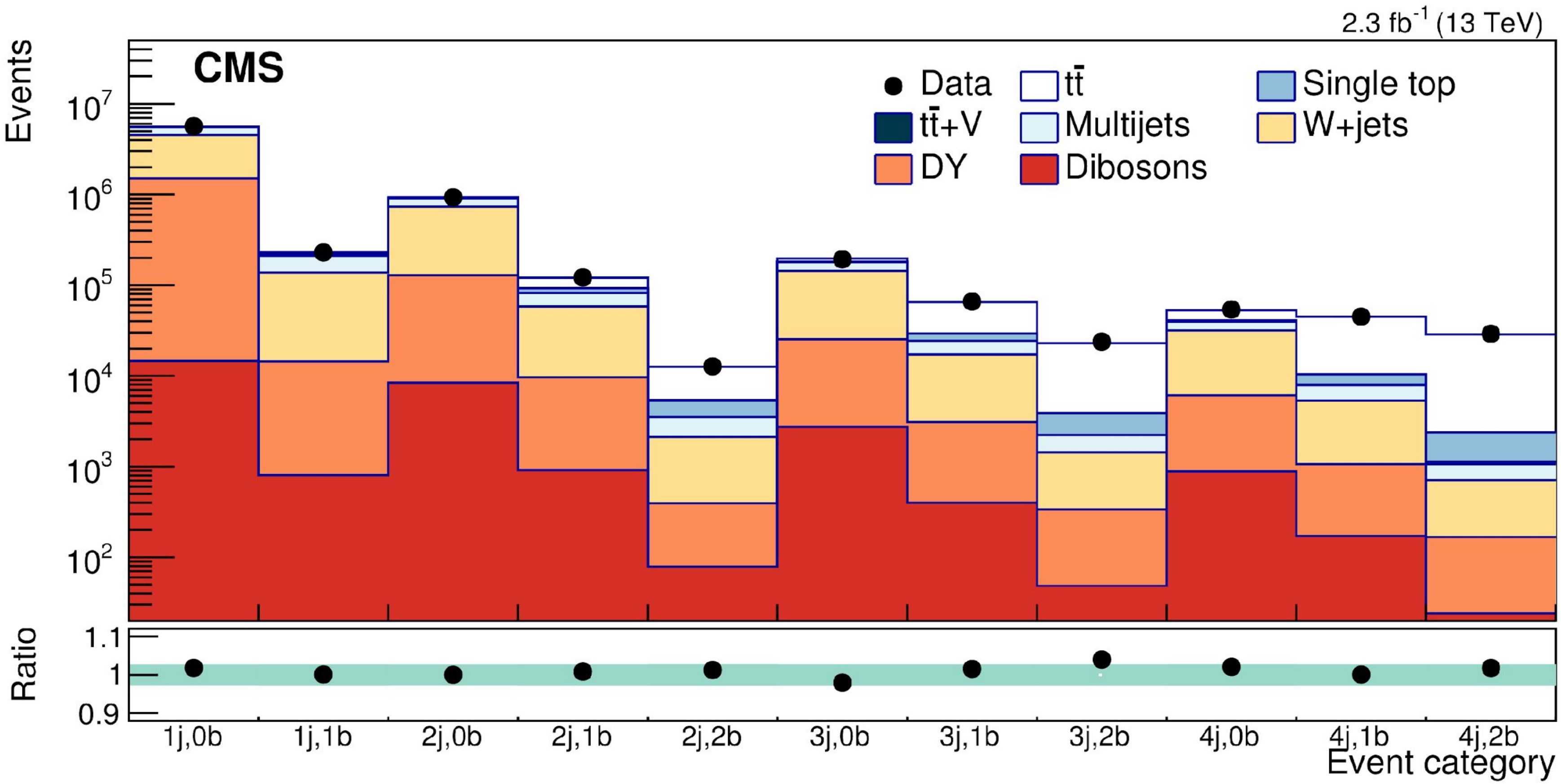
Using W-mass related variables in different b-tag multiplicity categories.



Following [JHEP 03 (2018) 115]

Extracting the signal: a PRL fit

Or using the nJet-nbtag distribution:



Following JHEP 09 (2017) 051

Combining the measurements

- You measured σ(ttbar) in different channels
  - In principle, this corresponds to having measured four observables
  - But the underlying physical quantity is the same!
  - How to go from $\sigma_{channel}$(ttbar) to a single σ(ttbar)?

Two eminent ways of combining measurements

- Obtain individual measurements (central value and uncertainty), and combine the end results
  - Simplified assumptions (e.g. Gaussian uncertainties)
  - A measurement is not a sufficient statistic; loose information along the way
  - Not trivial to encode the desired correlations among uncertainties
- Perform a combined result using all the data at once
  - No need for simplications (e.g. uncertainties not necessarily symmetric or Gaussian)
  - The combined measurement is performed using all the information (encoded in the likelihood) from the original data
  - Correlations can be taken into account in a very detailed way

The top quark exercise @CMSDAS2019-PISA:

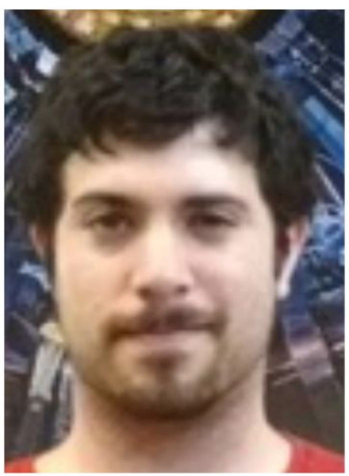
- A: dilepton channel -- event selection and background estimate: **4 people**.
- B: dilepton channel -- systematic uncertainties and cross section calculation (including modeling uncertainties at fiducial level): **4 people**.
- C: e/μ + jets channel -- QCD estimate and cross section extraction: **4 people**.
- D: Combining the measurements (ee, μμ, eμ, e+jets, μ+jets): **3 people**.

Technical details, code, and how to proceed:
https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideCMSDataAnalysisSchoolPisa2019TopPairCrossSection

And the most important thing...

Enjoy the course and learn a lot!

...and ask us when you get too stuck!


Xuan


Andrea

Pietro


Pieter