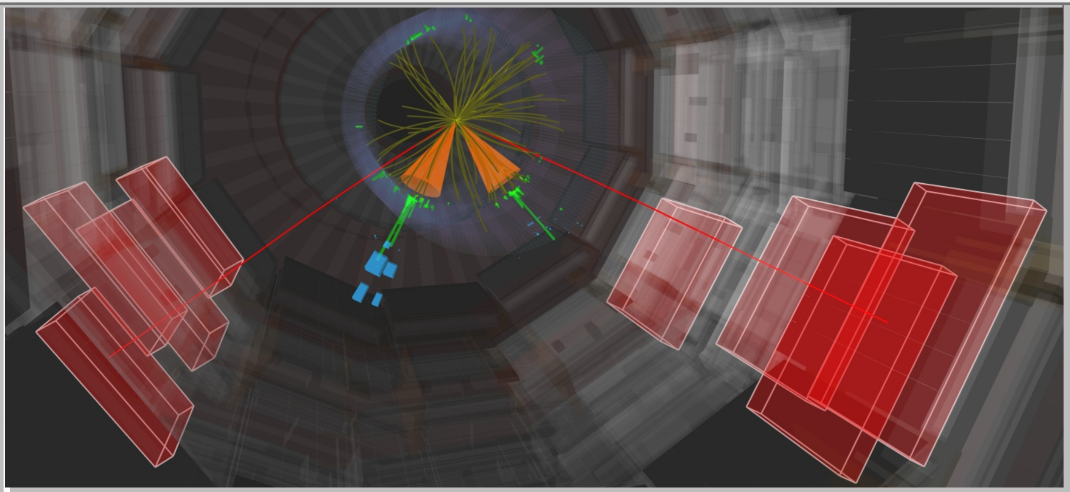
CMS early Run 3 measurement of the $t\bar{t}$ cross-section (@ DESY!)

FH Particle Physics Discussion

13.11.2023

Evan Altair Ranken

for the DESY CMS group



Flashback: LHC Run 3 begins

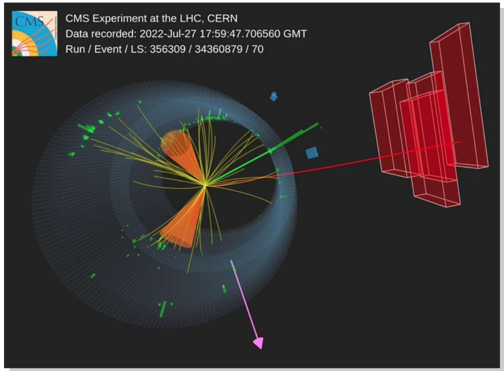
5 July 2022:

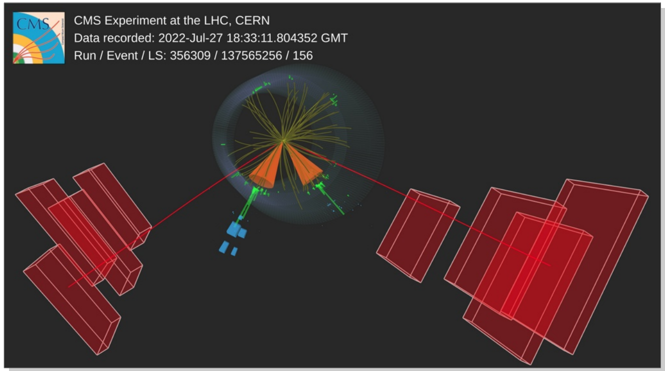
The LHC raises the bar again, pushing to $\sqrt{s}$ = 13.6 TeV

- **Late July:**
  - Ramp-up complete, ~1200 bunches circulating
  - Physics-ready!
- **Early September:**
  - First preliminary Run 3 result from CMS with heavy DESY involvement

    Now superseded by a publication in JHEP

    arXiv:2303.10680





Introduction

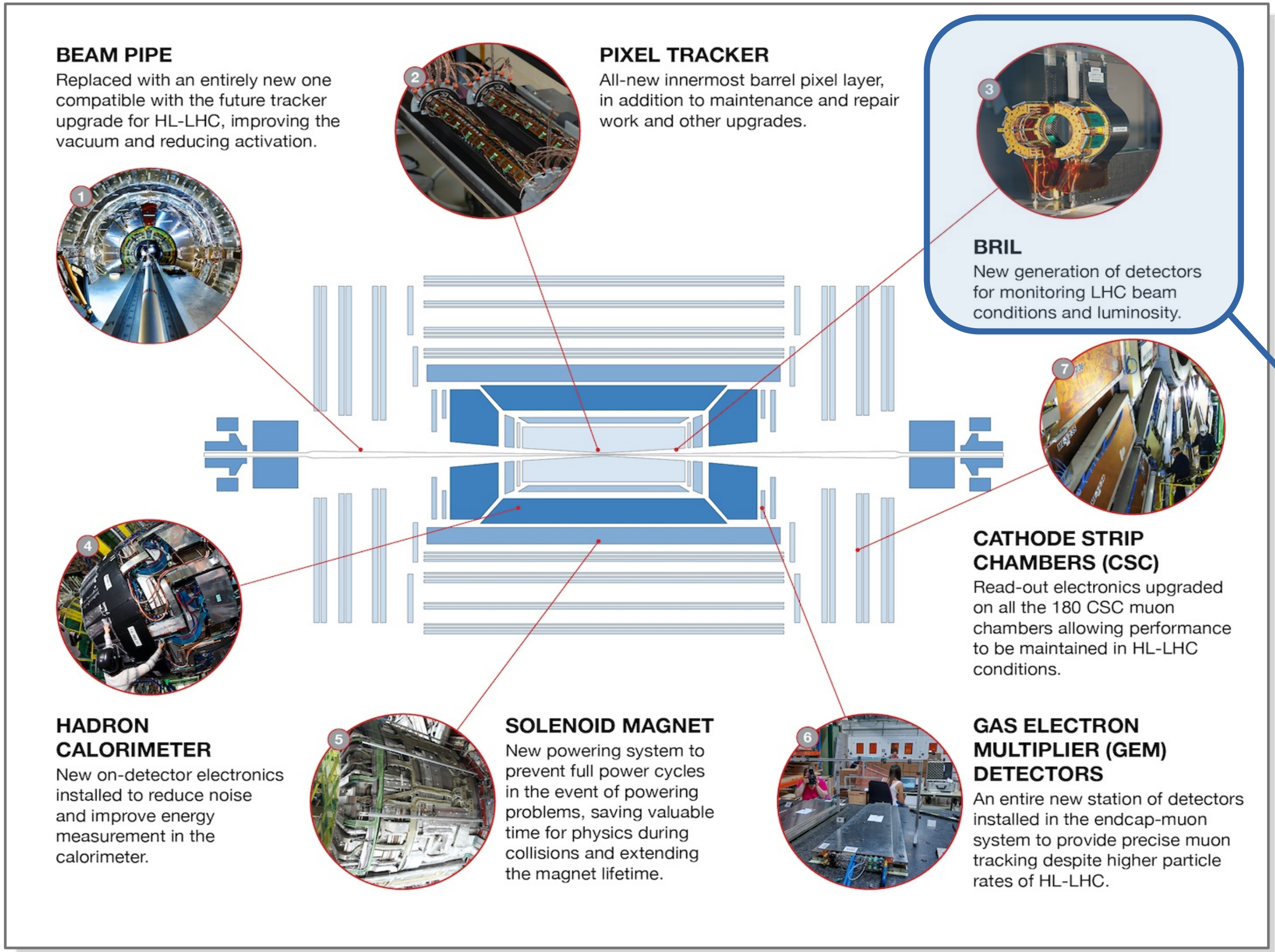
- **Questions to address in this talk:**
  - How did CMS and its DESY members achieve this timescale?
  - Are this and other “early measurements” worthwhile?
  - What did we do differently with this measurement?
  - What are the lessons from this measurement?

“Early measurement”

We will walk through the timeline:

- Start of Run 3
- Steps to a preliminary measurement!
- Updates to the publication
- Lessons

CMS upgrades for Run 3

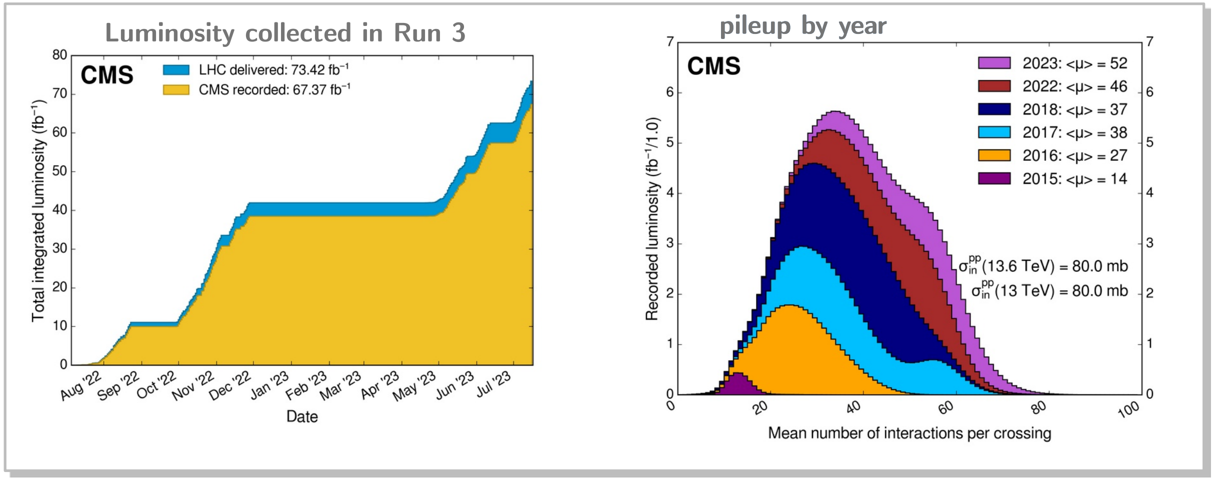


- Many upgrades done during the long shutdown
- Preparing already for HL-LHC (Run 4)

Large group @ DESY!
See FH seminar by Andreas

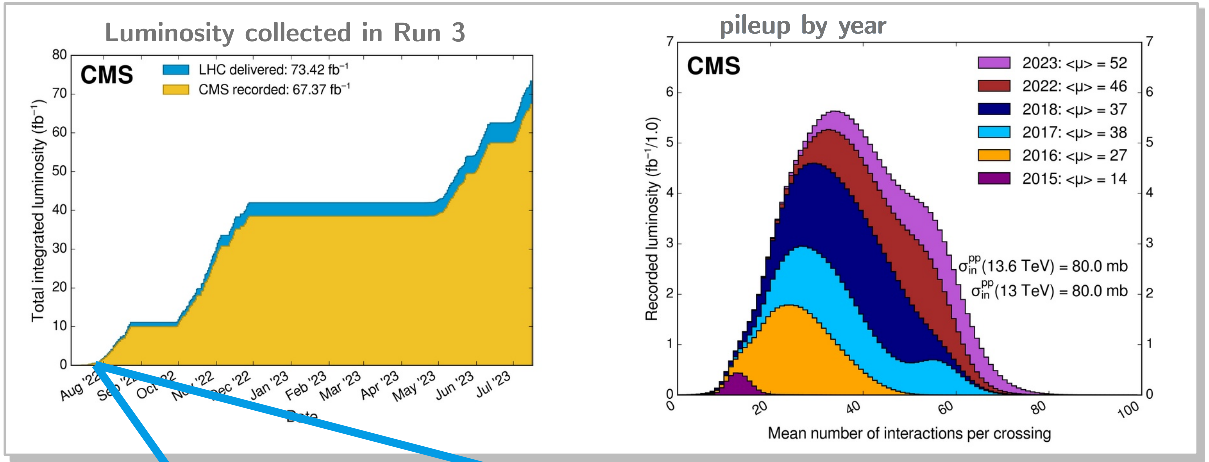
LHC @ 13.6 TeV

- To date, **CMS** has recorded over 60 fb^{-1} in Run 3
- Successful data-taking at our highest pileup values yet

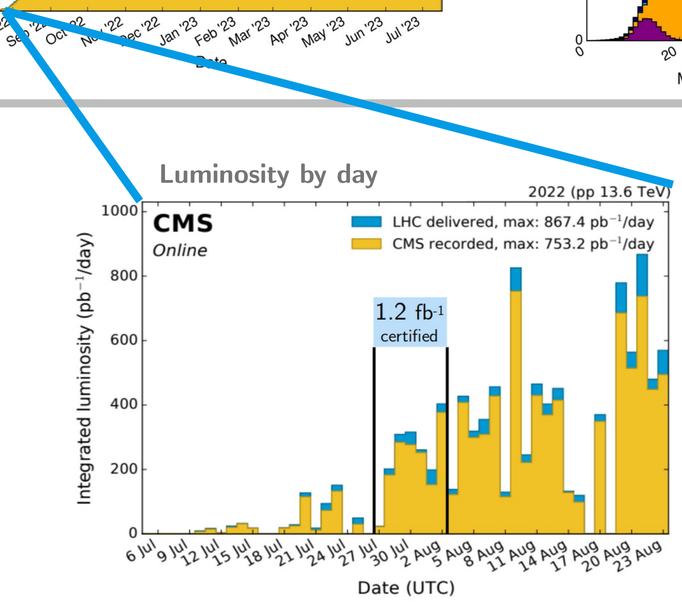


LHC @ 13.6 TeV

- To date, **CMS** has recorded over 60 fb^{-1} in Run 3
- Successful data-taking at our highest pileup values yet



This measurement targeted only
1fb^{-1} of integrated luminosity
From the first weeks of data-taking



Early analysis targets

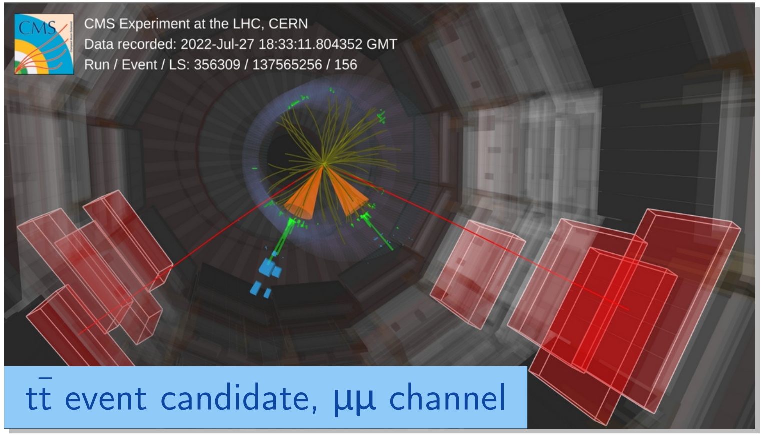
At the start of Run 3, we are eager to analyze new data:

- First look at new energy frontier, want to test our models by measuring quantities sensitive to the LHC energy increase
- Vital to verify physics-readiness after years of upgrade work & commissioning at CMS

Why top quark pairs?

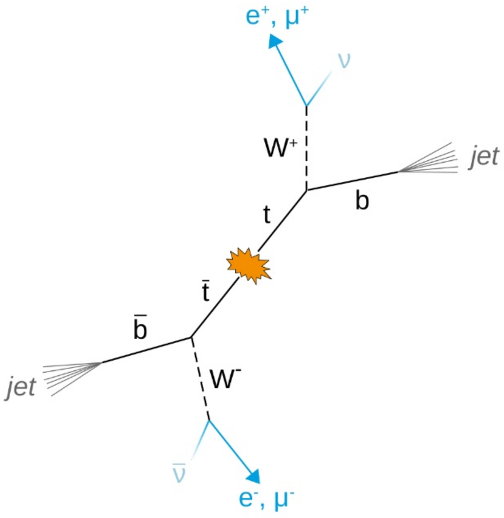
$t\bar{t}$ production

- Involves a wide variety of particles
- Uses information from all main detector components
- Great for early analysis + validation of new data

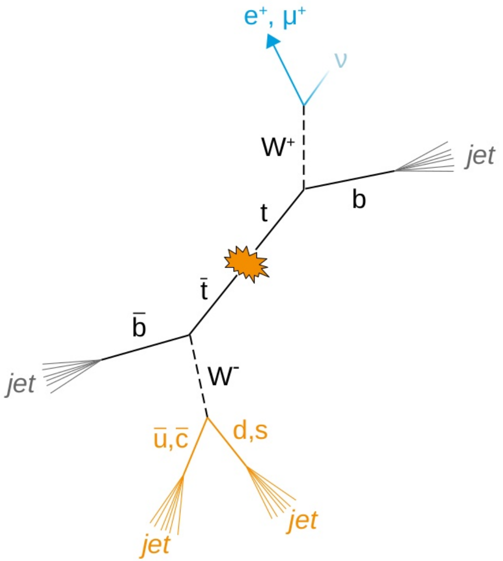


$t\bar{t}$ event candidate, $\mu\mu$ channel

dilepton

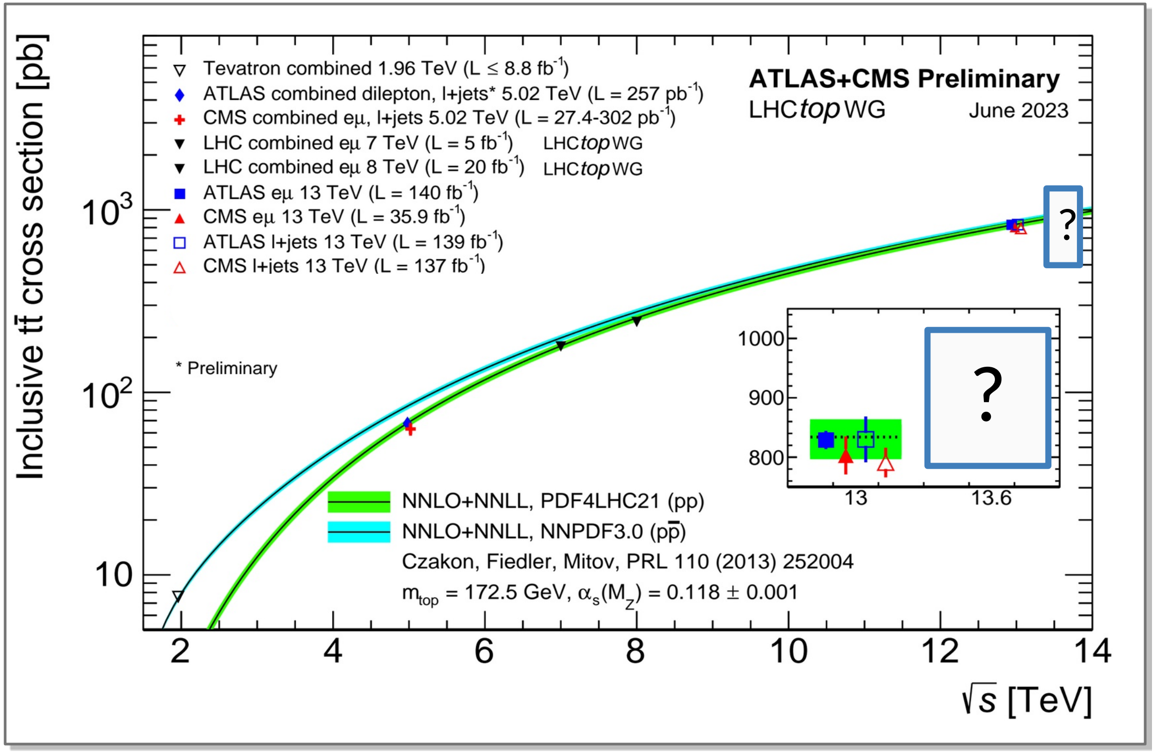


lepton + jets



Why the $t\bar{t}$ cross section?

- As of 2022, the top quark pair production cross section ($\sigma_{t\bar{t}}$) had been measured at 5 energies, of which 4 measured at LHC experiments



- $p\bar{p} \to t\bar{t}$ measured first by CDF & D0 at the **Tevatron**

  @ $\sqrt{s}$ = 1.96 TeV

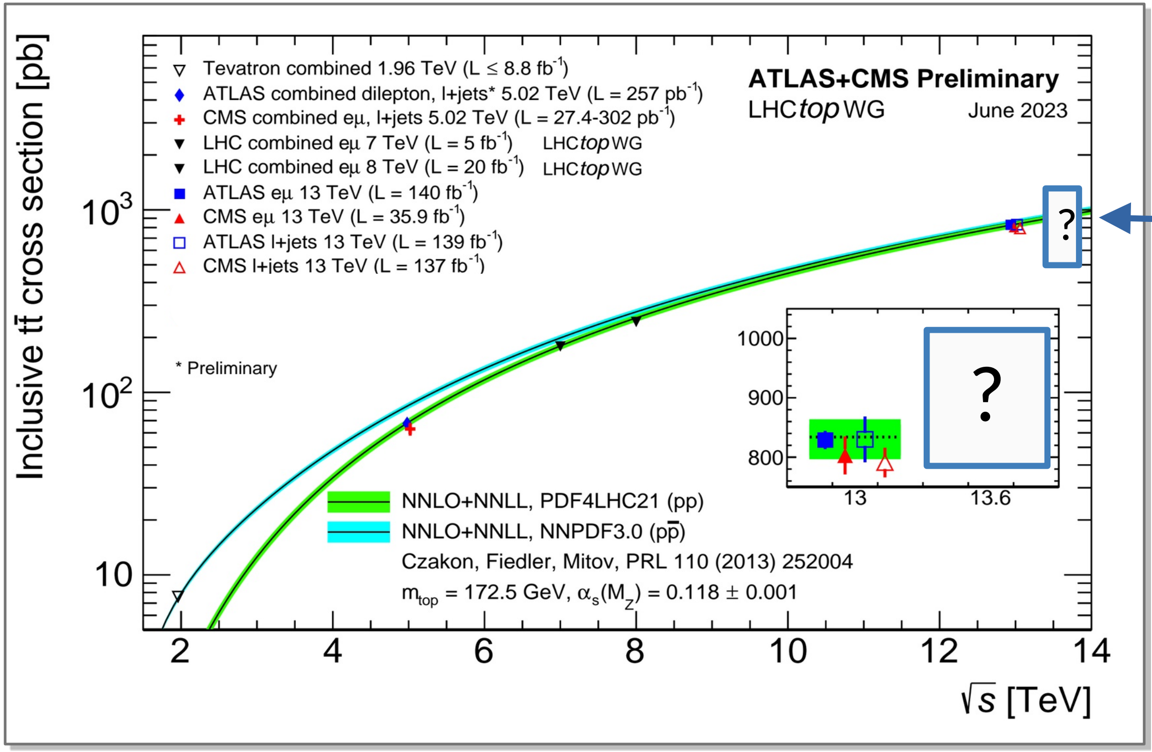
- **$pp \to t\bar{t}$**
  measured at the **LHC** by:

  CMS and ATLAS
  $\sqrt{s}$ = 5.02, 7, 8, 13 TeV

  +LHCb in the forward region
  $\sqrt{s}$ = 7, 8, 13 TeV

Why the $t\bar{t}$ cross section?

- As of 2022, the top quark pair production cross section, $\sigma_{t\bar{t}}$, had been measured at 5 energies, of which 4 measured at LHC experiments



$\sqrt{s}$: 13 TeV → 13.6 TeV

$\sigma_{t\bar{t}}$: 834 pb → 924 pb

over 10% increase in LHC Run 3!

Limiting factors

- Cross section measurements boil down to counting events. However, there are many limiting factors
  - Luminosity
  - Lepton identification
  - Jet energy response
  - b-flavor jet tagging

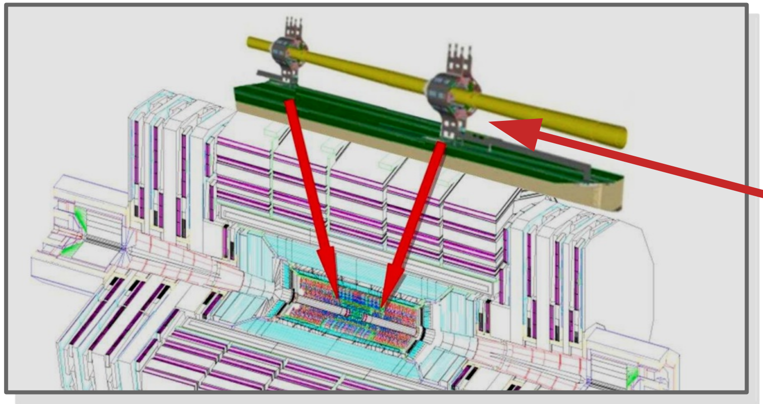
Luminosity

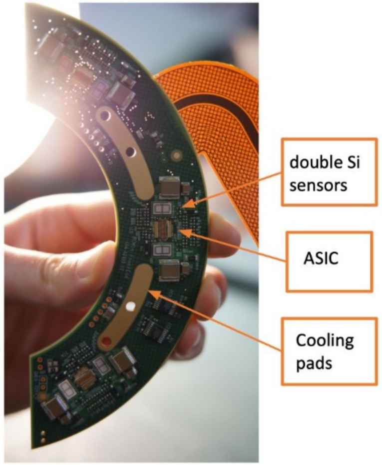
Early luminosity estimates come from emittance scans at start/end of LHC fills

- BCM1F upgraded for Run 3: more channels, silicon diodes

  (joint project @ DESY + CERN)

- CMS has multiple luminometers, which showed good agreement





BCM1F C-shape

(CMS-DP-2022-033)

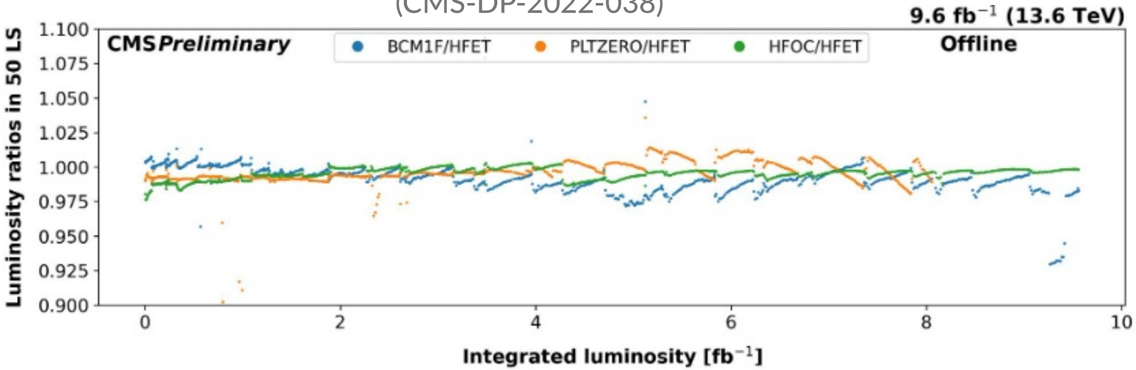
Hard to obtain absolute calibration

- LHC must run in a special configuration once per year: Van der Meers scans

  (October 2022 + analysis time)

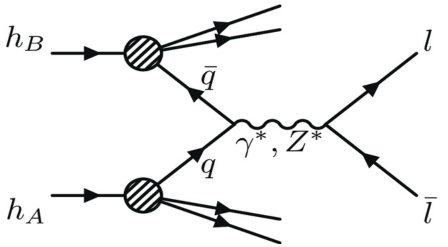
- Emittance scans are short scans much more vulnerable to systematic uncertainty

(CMS-DP-2022-038)

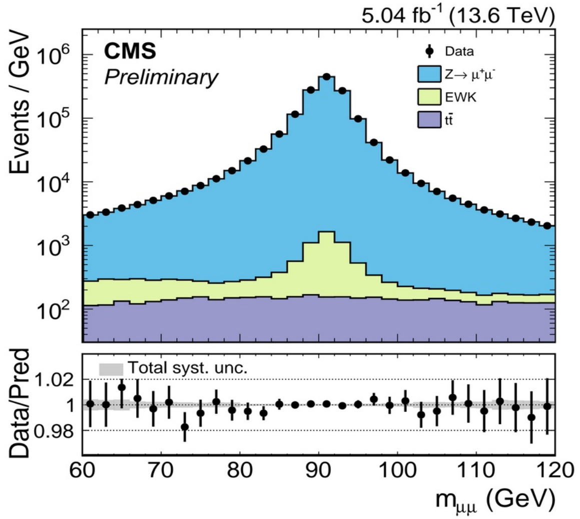


Luminosity / Z-counting

- **Drell-Yan** events involving a **Z boson** are particularly useful for early calibration



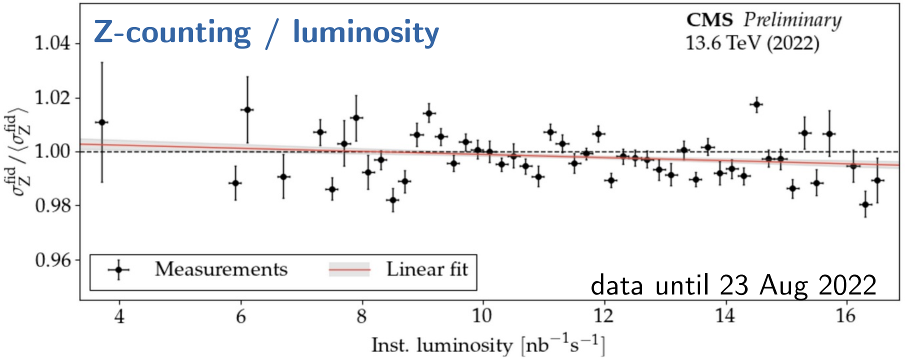
CMS-PAS-SMP-22-017



- Counting $Z \rightarrow \mu\mu$ events as an independent cross-check on luminosity:

  → combine with emittance scan results

  → obtain best possible lumi estimate before VdM scans



CMS-DP-2023-003

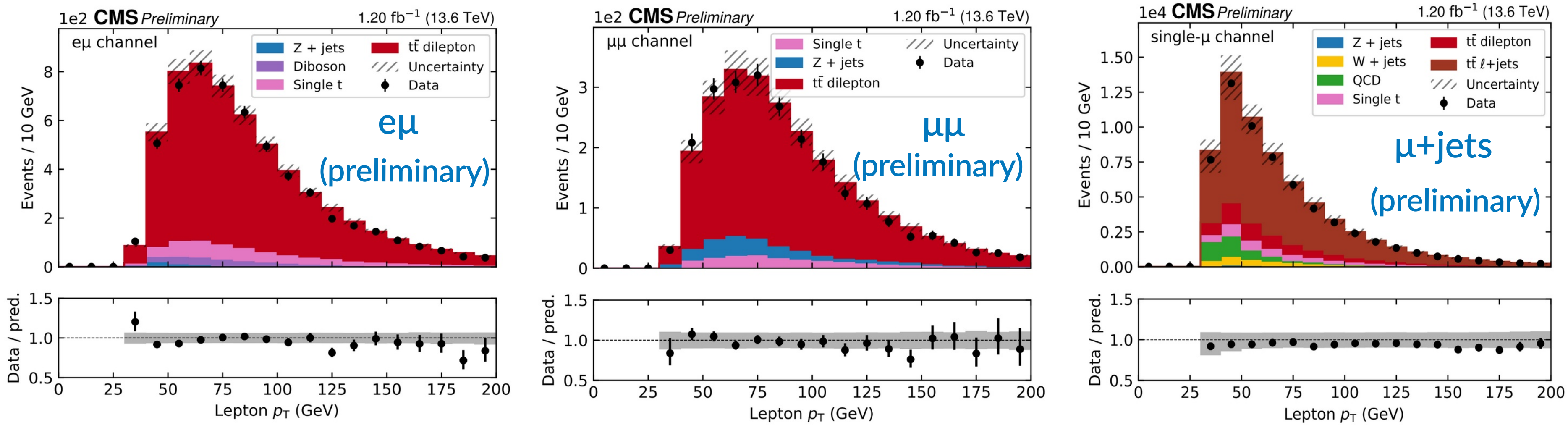
CMS measurement strategy

- **Channel combination:** eμ, ee, μμ, e+jets, μ+jets ~1 fb^{-1} of data

  → use information from multiple channels to constrain nuisance parameters *in situ*

- Lepton selection: uniform between e/μ and across channels
  - p_T > 35 GeV for all leptons → constraint of lepton ID efficiency performance

leading lepton p_T



CMS measurement strategy

- **Channel combination:** eμ, ee, μμ, e+jets, μ+jets   ~1 fb^{-1} of data

  → use information from multiple channels to constrain nuisance parameters *in-situ*

- Lepton selection: uniform between e/μ and across channels

- ID efficiencies depend on lepton kinematics, but these variables are not needed for a simple cross section measurement IF you don't bin into categories with different kinematics

  → selection efficiencies ϵ_μ and ϵ_e factorize!

ee yield $\sim \epsilon_e^2$	e+jets yield $\sim \epsilon_e$
eμ yield $\sim \epsilon_e \epsilon_\mu$	μ+jets yield $\sim \epsilon_\mu$
μμ yield $\sim \epsilon_\mu^2$	

Efficiency correction = scale factor (SF)
Different effect on different channels:

→ distinguish ID SF from σ_{tt}

→ float+constrain ID SF in situ

Object selection + corrections

Lepton selection:

- p_T > 35 GeV
- cut-based IDs (70% signal efficiency) ported from Run-2

Jet selection: AK4 jets

- p_T > 30 GeV
- b tagging: Deepjet algorithm

Designed for early analysis:

b tagging efficiency

- Bin content follows binomial distribution in b-tagging efficiency ϵ_b

$$\epsilon_b^{N_{\mathrm{b-tag}}} (1 - \epsilon_b)^{N_{\mathrm{b-jet}} - N_{\mathrm{b-tag}}}$$

- Floating efficiency parameter in the fit

Jet energy calibration (JEC)

- Take time to derive on new data
- Define hadronic W in lepton+jets channel with 2 highest p_T non-b-tagged jets

→ derive coarse JEC, sensitive to large discrepancies in data/MC

→ include additional uncertainty

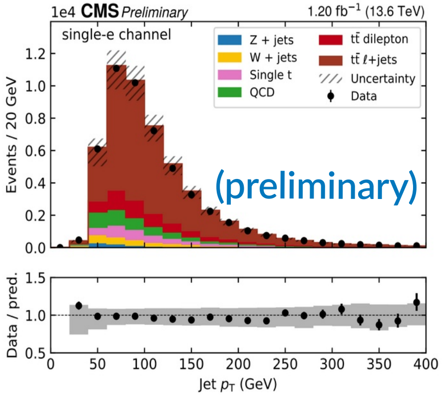
Jet control plots

Jet energy calibration (JEC)

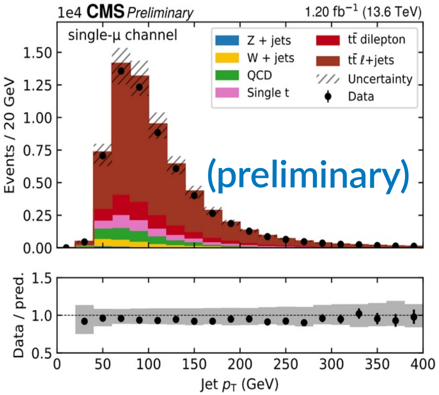
- Only preliminary calibrations were available
- Compared these with Run 2 JEC to define additional uncertainty
- Still, overall jet behavior looked good

Leading jet p_T

e+jets

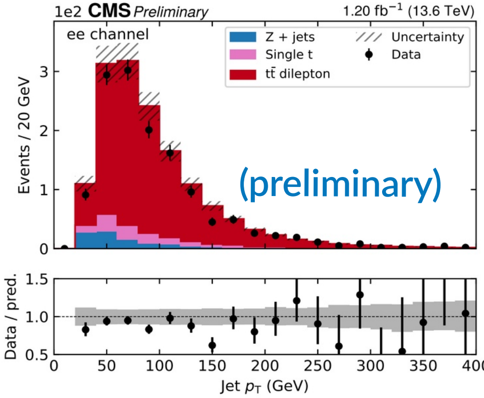


μ+jets

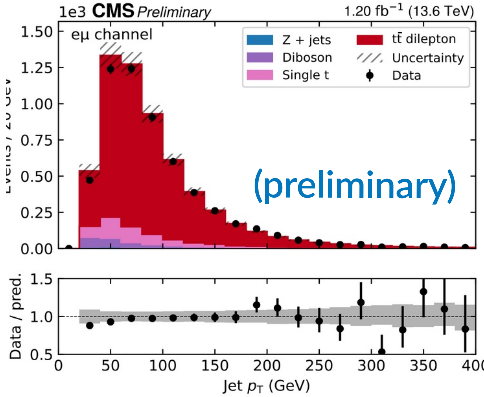


Leading jet p_T

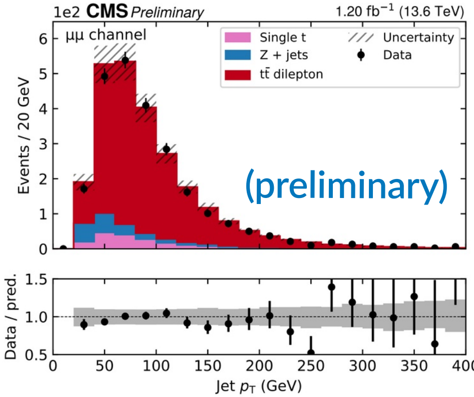
ee



eμ



μμ



Fit outcome (preliminary result)

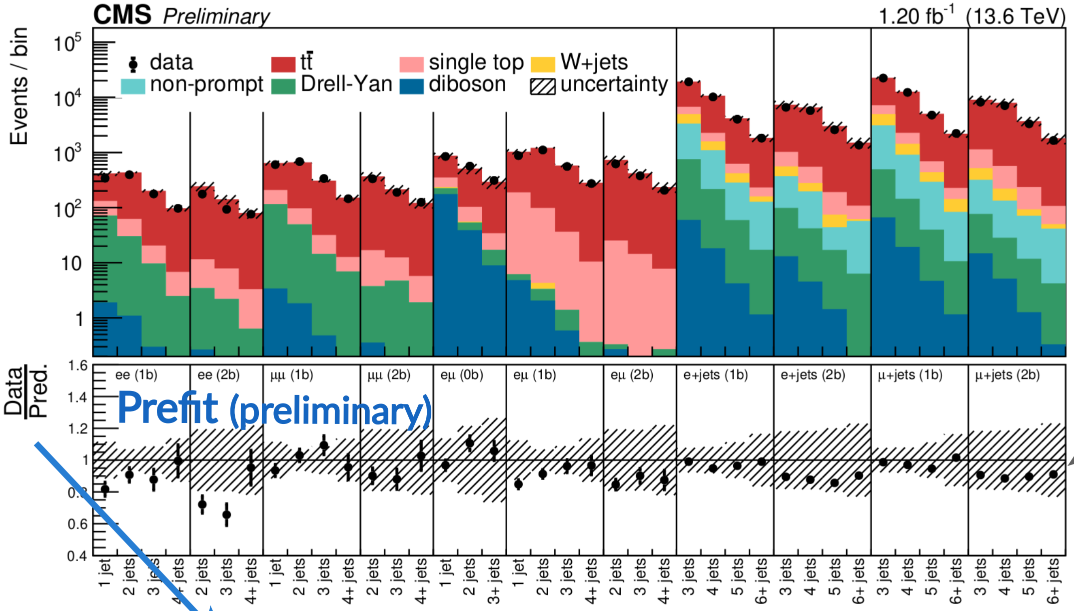
- Bins:
  - ➤ Lepton flavor
    - ➤ b jet multiplicity
      - ➤ Jet multiplicity
- **b-tag efficiency** scale factors determined by fit
- **6 % Luminosity** unc. externalized, not drawn

$$\sigma_{t\bar{t}}^{\text{theory}} = 924^{+32}_{-40}(\text{scale} + \text{PDF} + \alpha_S)\,\text{pb}$$

Result: CMS-PAS-TOP-22-012

$$\sigma_{t\bar{t}} = 887\,^{+43}_{-41}\,(\text{stat} + \text{syst}) \pm 53\,(\text{lumi})\,\text{pb}$$

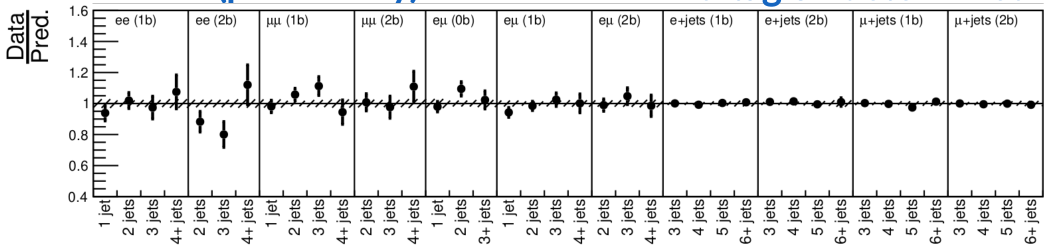
~ 0.5% statistical uncertainty



no b-tag SF

Postfit (preliminary)

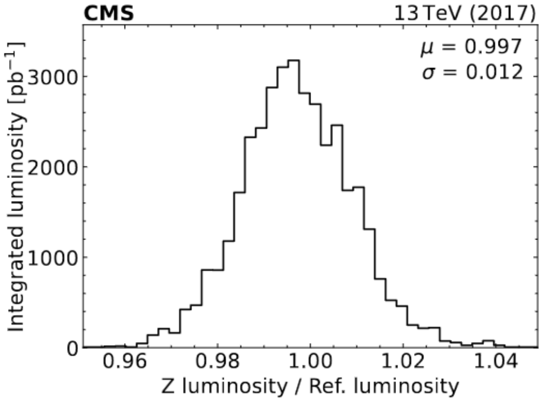
b-tag SF determined



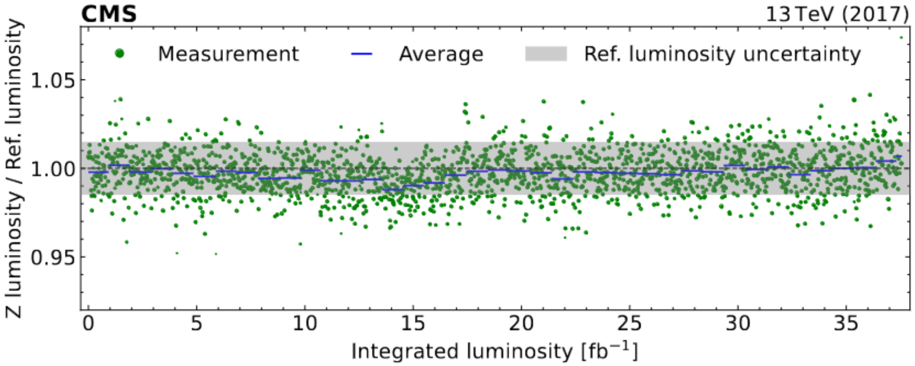
Lessons from the preliminary result

- Many useful techniques, some unusual and some brand new, were used in the search for an early result
  - In situ estimation of...
    - **lepton ID SF**
    - b-tag SF with binning in N_{jet}
  - New/novel methods
    - **Z-boson counting for luminosity**

      → now a preprint using 2017 data!
    - Jet energy calibration
      (coarse cross-check only)



arxiv:2309.01008



Lessons from the preliminary result

- Many useful techniques, some unusual and some brand new, were used in the search for an early result
  - In situ estimation of...
    - **lepton ID SF**
    - b-tag SF with binning in N_{jet}
  - New/novel methods
    - **Z-boson counting for luminosity**

      → now a preprint using 2017 data!
    - Jet energy calibration
      (coarse cross-check only)

Note: these were influenced by the drive to obtain the result quickly, **but** were not dependent on its completion!

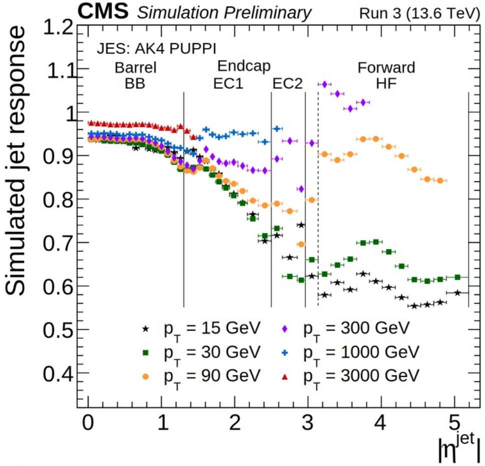
Updates for publication

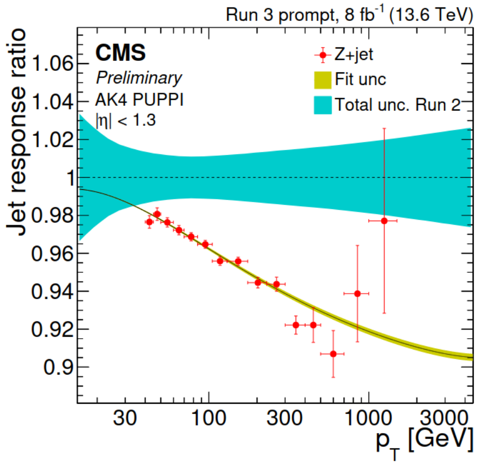
- Wanted to push the precision as far as possible while keeping the basic measurement strategy
  - **Luminosity**
    - VdM scans delayed, ready in early 2023
    - 6% luminosity uncertainty → 2.3%
  - **Lepton ID scale factors**
    - Use tag-and-probe method in Z-enriched sideband
    - Analysis-specific ID efficiencies using same dataset
    - 2% uncertainty → ~1% (per lepton flavor)
  - **Cross check**
    - Event counting method in eμ channel, no b-tagging needed

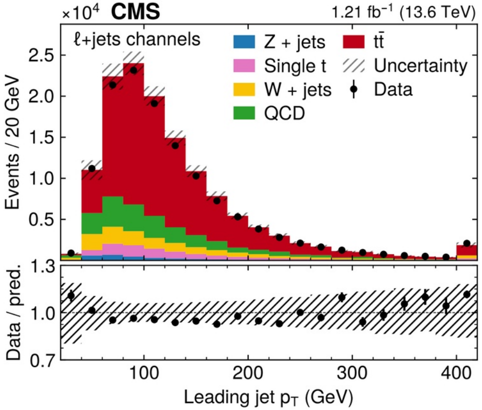
Updates: jet energy

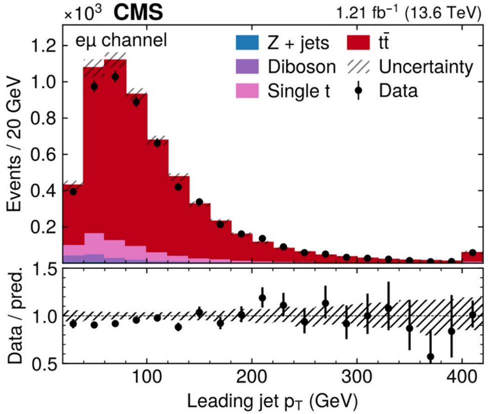
- **New jet calibrations** for 2022 data,
- reconstruction algorithm including subtraction of neutral hadrons **(PUPPI)**

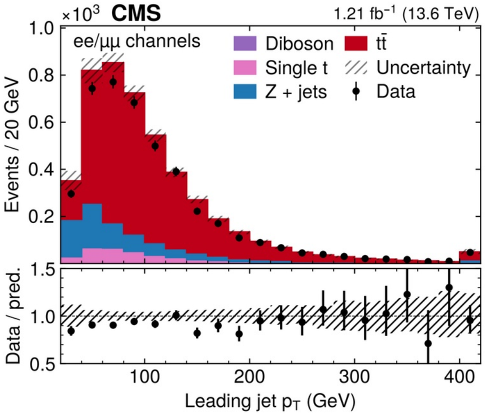
(CMS-DP-2022-054)











Fit outcome (published result)

Source	Uncertainty (%)
Statistical uncertainty	0.5
Combined uncertainty	2.5
Integrated luminosity	2.3
Total uncertainty w/ lumi	3.4

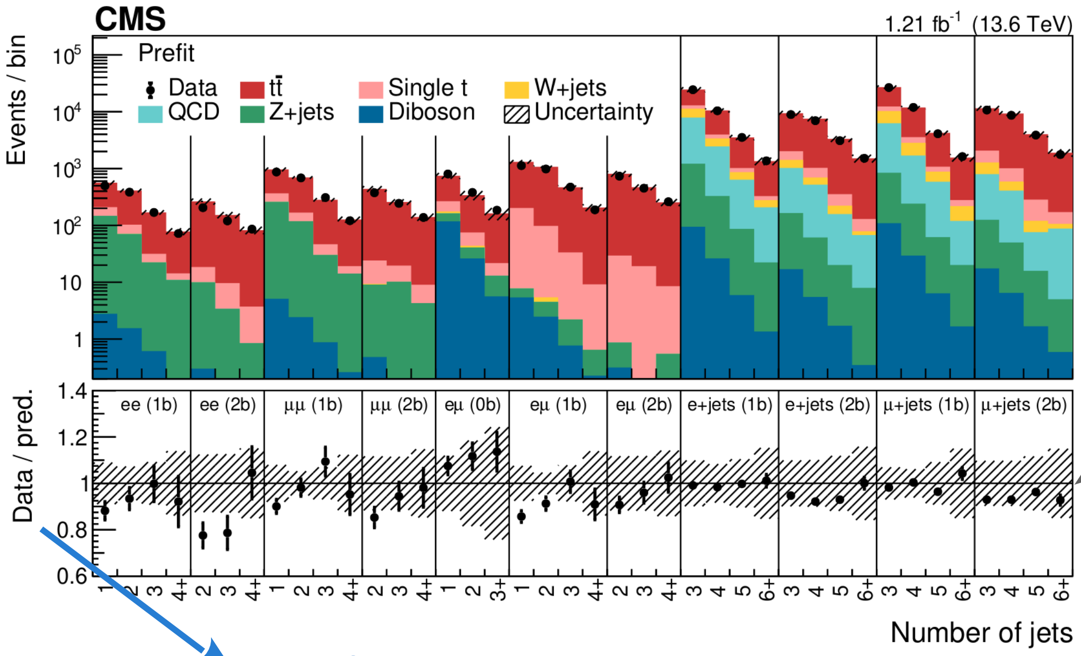
$$\sigma_{t\bar{t}}^{\text{theory}} = 924^{+32}_{-40}(\text{scale} + \text{PDF} + \alpha_S)\ \text{pb}$$

Result: arXiv:2303.10680
published in JHEP

$$\sigma_{t\bar{t}} = 881 \pm 23(\text{stat.} + \text{syst.}) \pm 20(\text{lumi.})\ \text{pb}$$

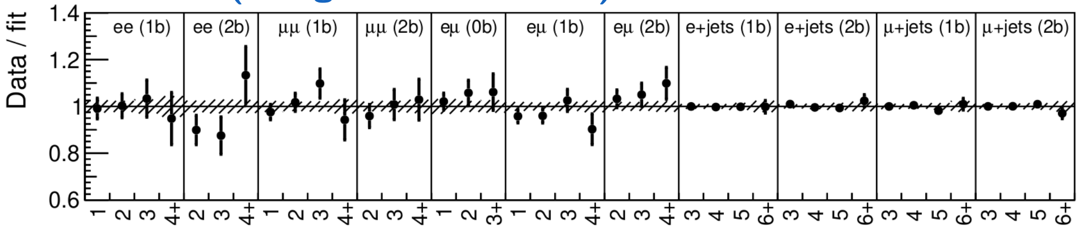
eμ cut and count cross-check:

$$\sigma_{t\bar{t}}^{e\mu-\text{only}} = 888 \pm 34(\text{stat.} + \text{syst.}) \pm 20(\text{lumi.})\ \text{pb}$$



no b-tag SF

Postfit (b-tag SF determined)



Uncertainties

$t\bar{t}$ cross section

2 clear leading uncertainties: lepton, b-tag eff.

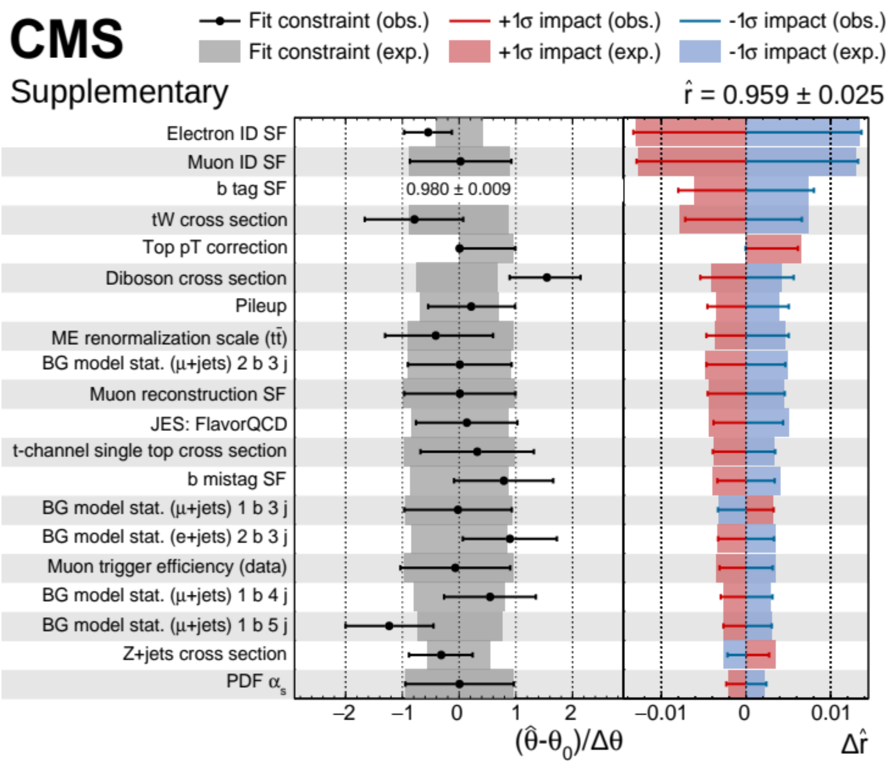
Many "3rd place" uncertainties

Not statistics limited

Lumi comparable to likelihood stat+syst unc.

Source	Uncertainty (%)
Lepton ID efficiencies	1.6
Trigger efficiency	0.3
JES	0.6
b tagging efficiency	1.1
Pileup reweighting	0.5
ME scale, $t\bar{t}$	0.5
ME scale, backgrounds	0.2
ME/PS matching	0.1
PS scales	0.3
PDF and α_S	0.3
Top quark p_T	0.5
tW background	0.7
t-channel single-t background	0.4
Z+jets background	0.3
W+jets background	<0.1
Diboson background	0.6
QCD multijet background	0.3
Statistical uncertainty	0.5
Combined uncertainty	2.5
Integrated luminosity	2.3
Total unc. (with lumi.)	3.4

= Resulting uncertainty in $\sigma_{t\bar{t}}$



Results

- The top quark pair production cross section was measured at a 5th energy by CMS

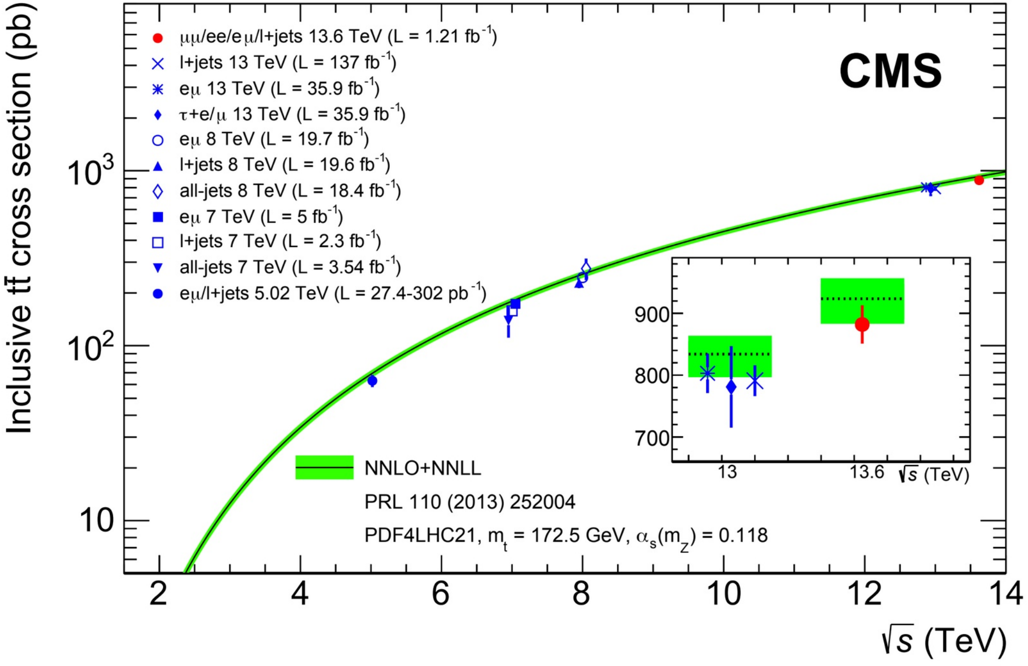
  First physics publication of Run 3!

- The DESY CMS group was heavily involved in every stage of the analysis

- **"Early analysis"** requires a different sort of problem solving, has risks and rewards

  ... but the journey is also important!

arXiv:2303.10680 (now in JHEP)



$$\sigma_{t\bar{t}} = 881 \pm 23(\text{stat.} + \text{syst.}) \pm 20(\text{lumi.})\ \text{pb}$$

$$\sigma_{t\bar{t}}^{\text{theory}} = 924^{+32}_{-40}(\text{scale} + \text{PDF} + \alpha_S)\ \text{pb}$$

Are "early measurements" worth it?

- From an **outside** perspective, they give a first look at the data

  → need an enticing physics target!

- From the perspective of someone **inside** a large collaboration like CMS, early analysis can be like driving off-road

  - Need to think outside of the box
  - Incentive to find new, creative ways to solve problems
  - More motivation to develop measurement-specific techniques

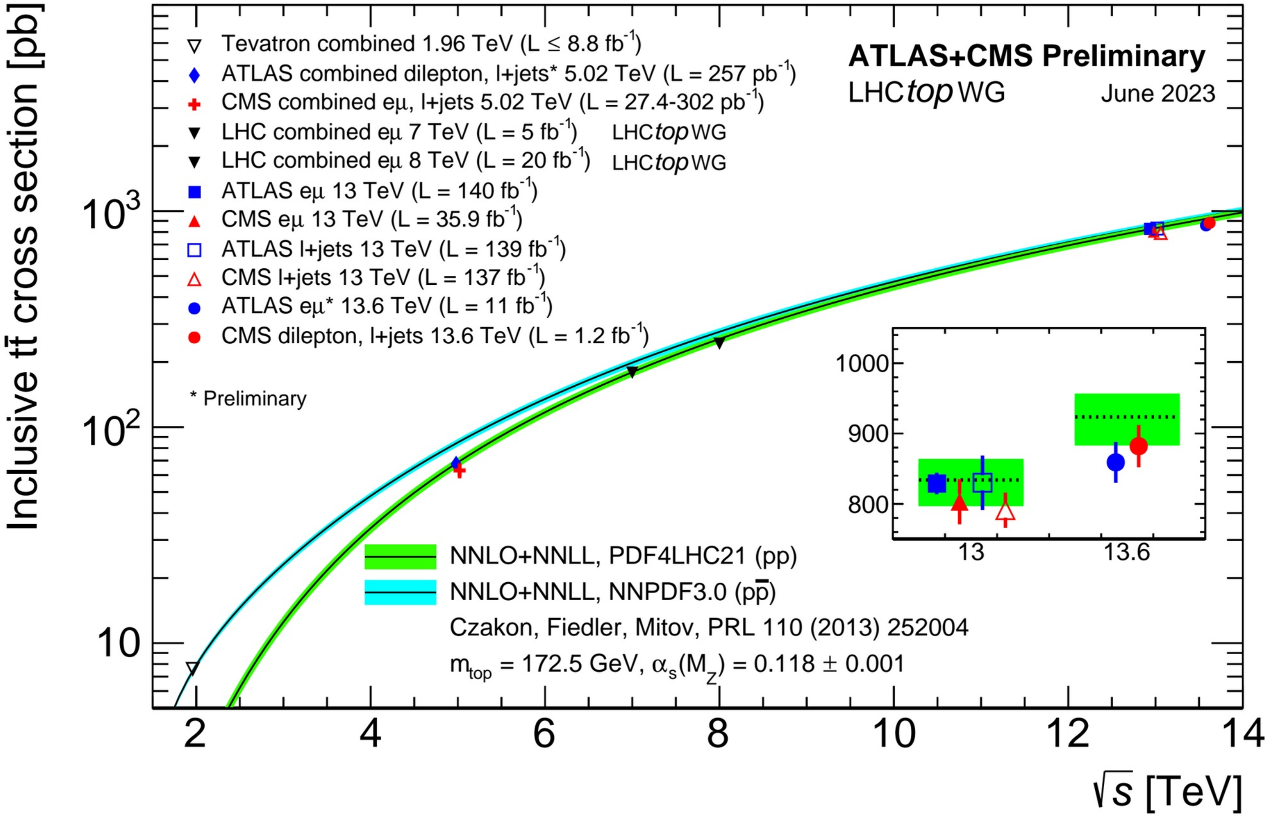
→ These things benefit **everyone**!

Thank you!

- Backup

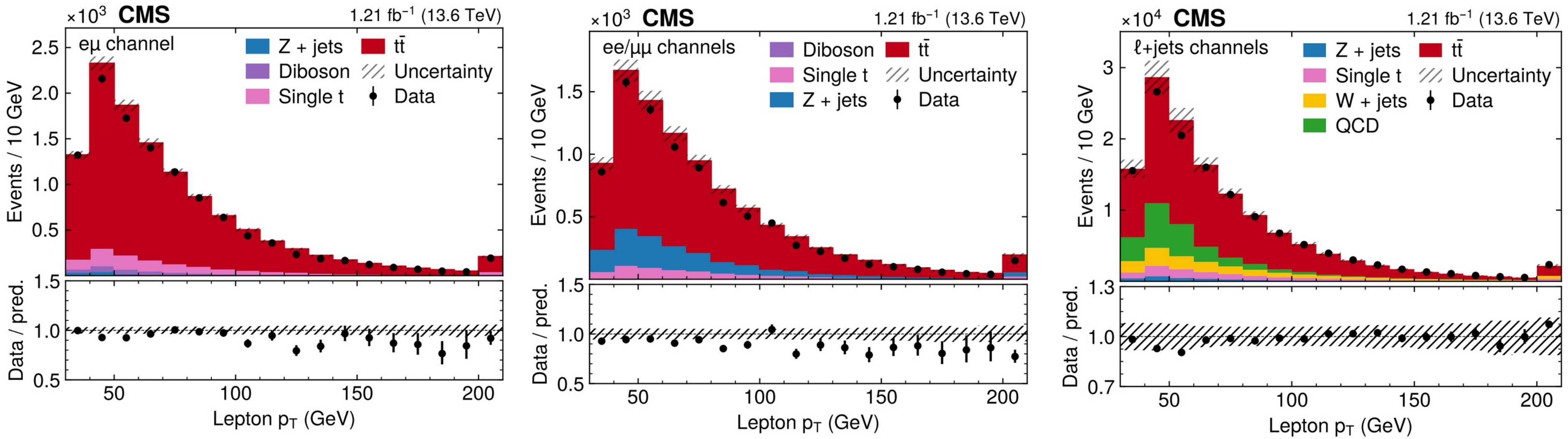
Full sqrt(s) plot

*ATLAS result not yet updated



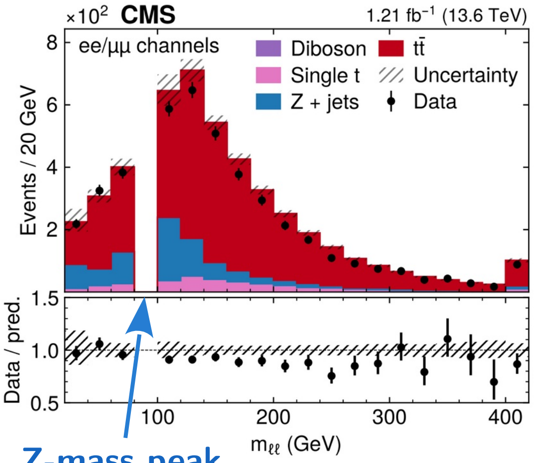
Control plots

Final lepton pT distributions



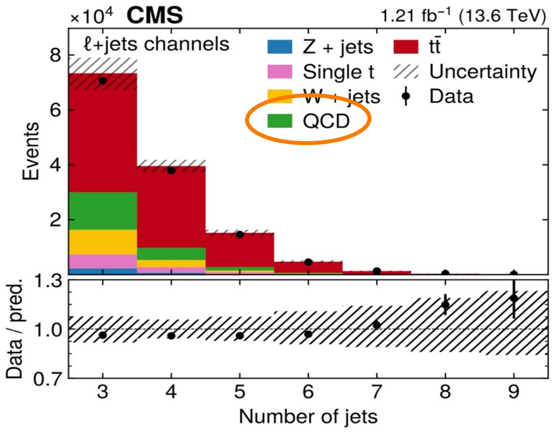
Analysis setup

- **Data-driven estimation:**
  - Z+jets events in ee, μμ channels
  - QCD multijet events in l+jets channels

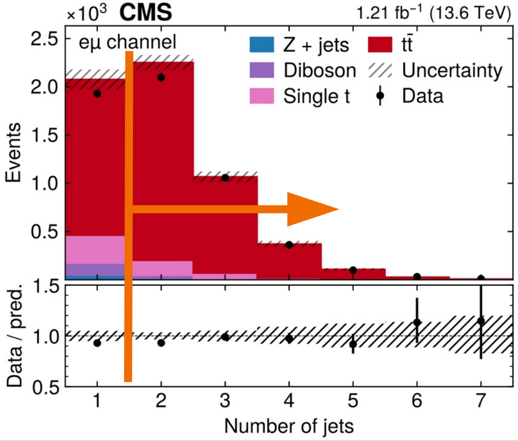


Z-mass peak sideband studied

QCD control regions:
inverted lepton isolation, $N_{jet}=1$



- **Cut and count cross-check** in eμ channel
  - Separate measurement performed via event counting in highest-purity channel
  - No b-tagging selection, only ≥ 2 Jets
    → check on consistency of result



High purity,
$N_{jet} \geq 2$
(no b-tag selection)