CMSDAS @ CERN 2024

Long exercise: tt cross section

Evan Altair Ranken
Dominic Stafford
Otto Hindrichs
Laurids Jeppe

19.06.2024

Long exercise basics

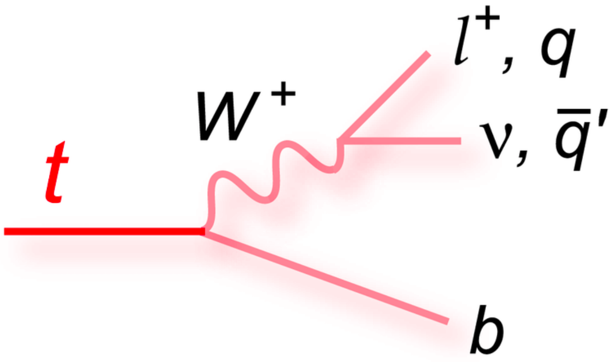
- A full CMS analysis would be hard to do in 3 days!

- So that we're all on the same page, let's review some basics and introduce the exercise in three steps:

  - **Top quarks**
  - **CMS Analysis**
  - **Measuring the $t\bar{t}$ cross section with Run 3 data**

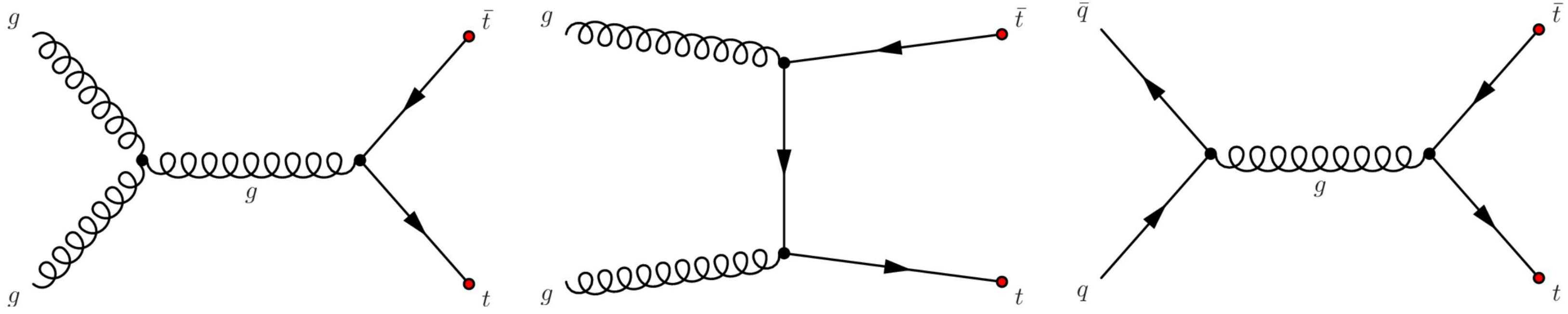
Part 1: **Top quarks**

- Top quarks are the heaviest known fundamental particle with a mass of 172.5 GeV
- They are the only quark heavy enough that decay before hadronizing, decaying almost exclusively to bW:



Top quarks

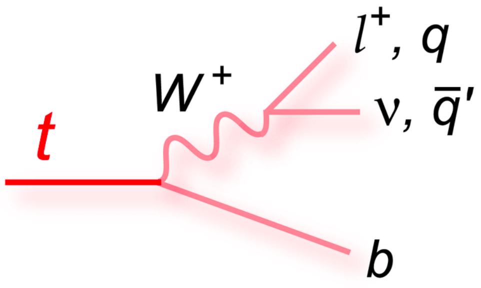
- They are produced in large numbers at the LHC, especially in pairs, and leave distinct experimental signatures



- Each top quark ultimately decays

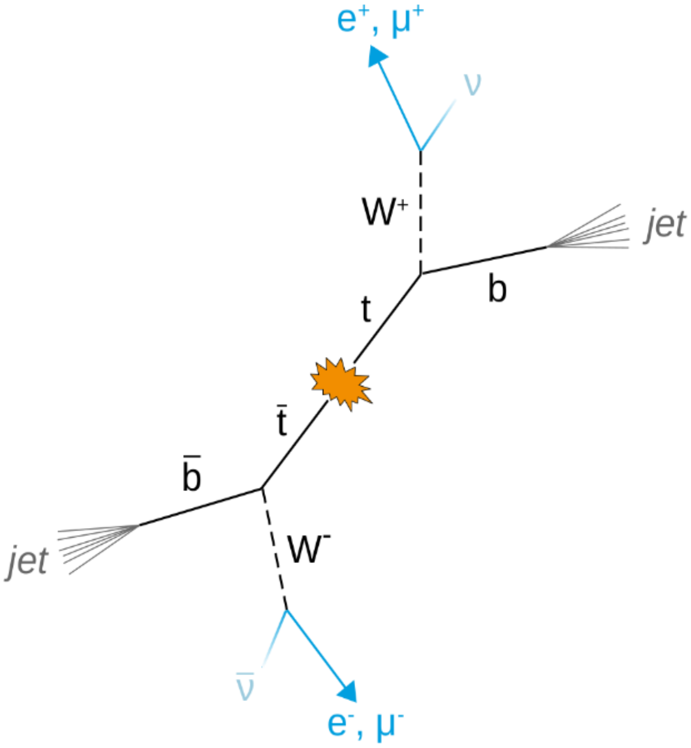
  hadronically (t → $q\bar{q}$)

  or leptonically (t → lν) = much cleaner signal!



Measurement basics

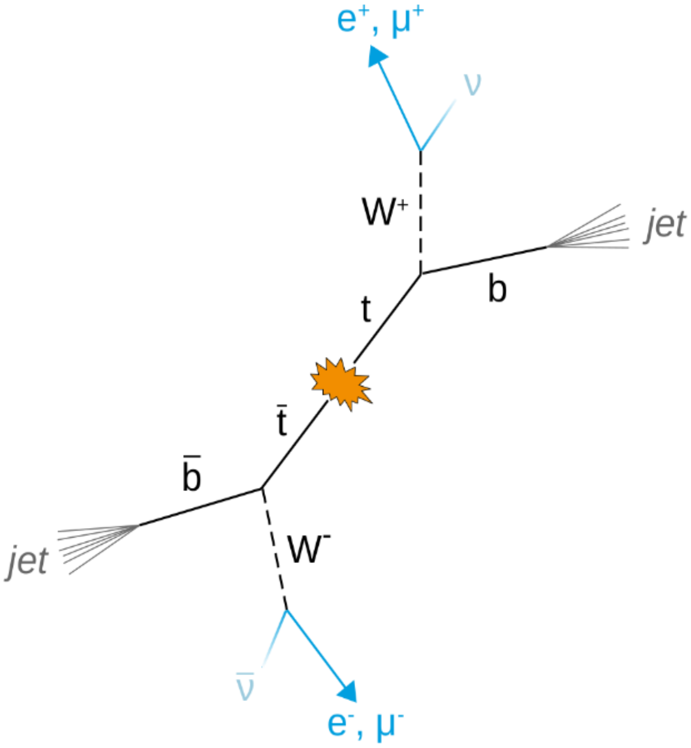
- Physics target: $t\bar{t}$ dilepton decays
  - Cleanest signal of top pair production
  - Characterized by two opposite sign leptons w/ high p_T and 2 b-jets
  - Significant MET due to neutrinos
  - ee, eμ, μμ final states



Measurement basics

- Inclusive cross section: a conceptually simple measurement

$$N_{\text{events}} = \mathcal{L}_{\text{int.}} \cdot \sigma_{\text{t}\bar{\text{t}}}$$



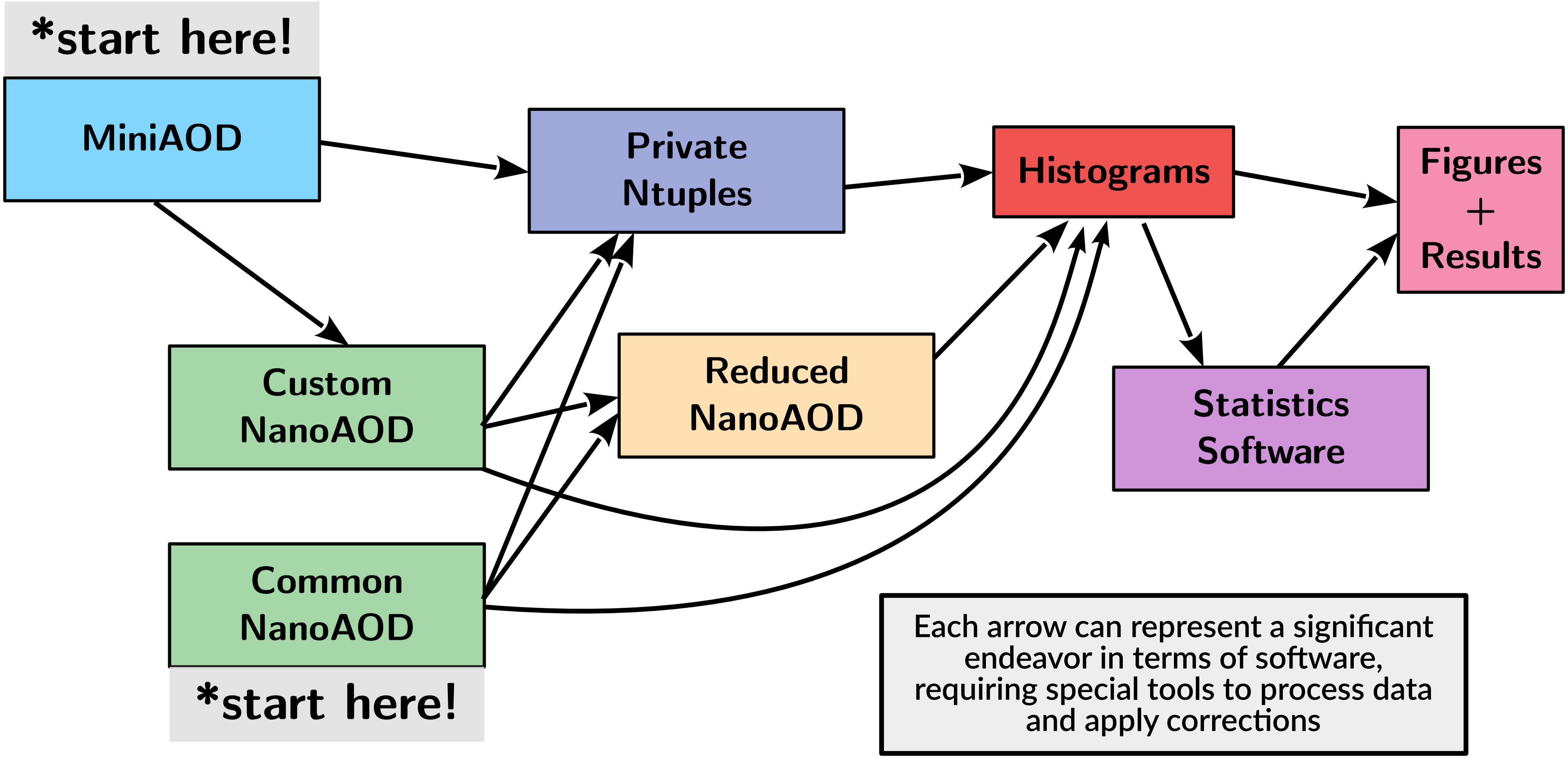
What does a CMS analysis look like?

Part 2: **CMS Analysis**

I will intersperse some results from a recent TOP PAG survey

To give a picture of the CMS analysis landscape

What does a CMS analysis look like?

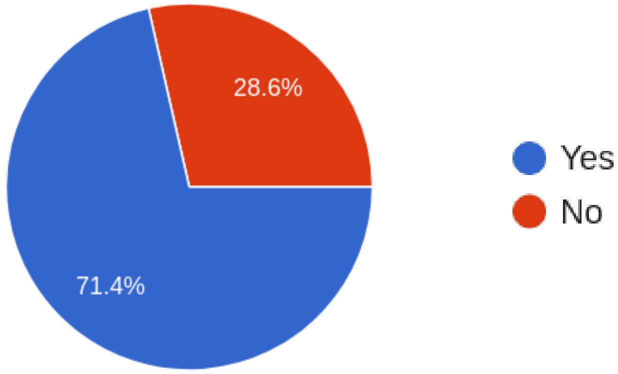


Each arrow can represent a significant endeavor in terms of software, requiring special tools to process data and apply corrections

Analysis trends

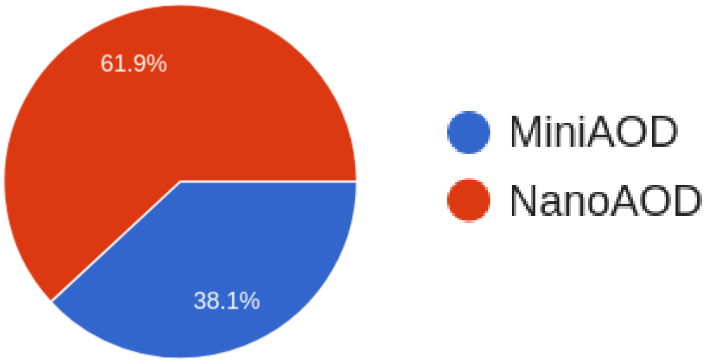
Recent analyses tending towards shared frameworks, NanoAOD

Does the analysis use a "shared framework" that is used by multiple other analyses measuring different quantities?

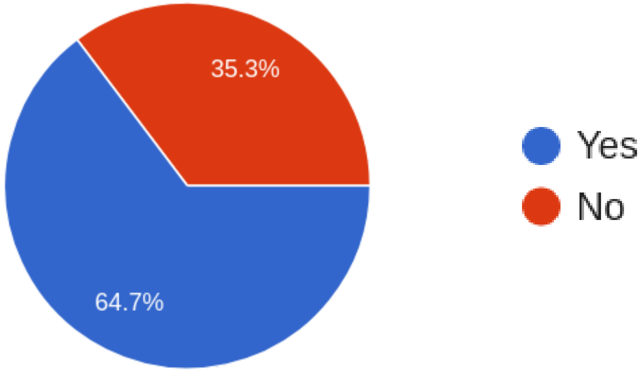


*See also: arrival of the [CAT group](#)

Your analysis framework uses



If using a shared framework, is the framework used by others outside of your institute?



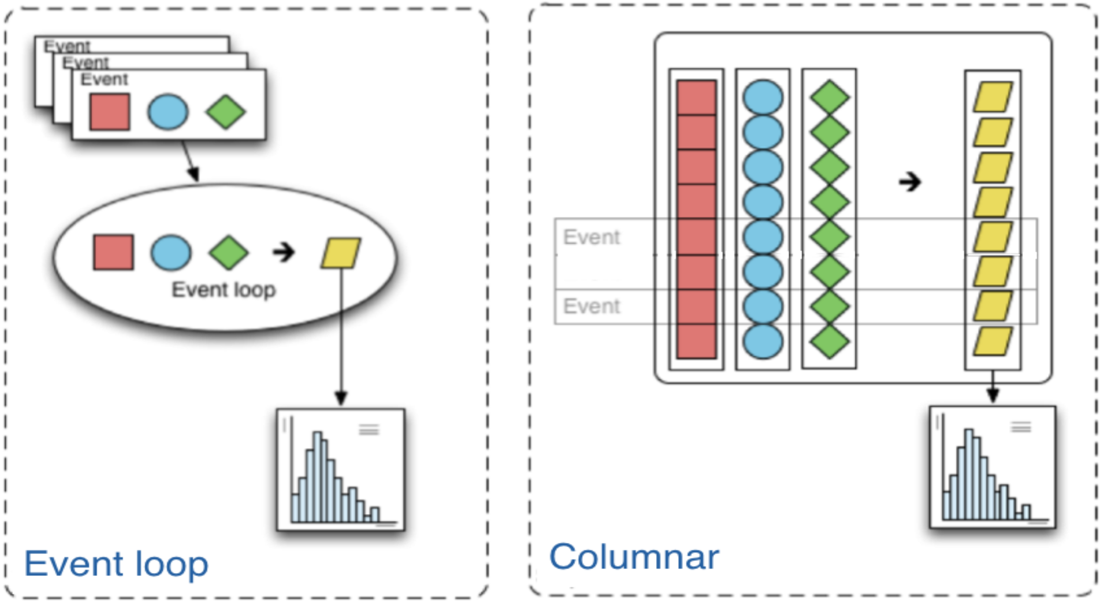
Data-processing approaches

Event loop vs. RDataFrame vs. Awkard arrays

Event-by-event thinking
Impossible in Python

Requires columnar thinking,
ROOT familiarity

Requires columnar thinking,
python/numpy familiarity



Processing data → histograms

Event loop vs. RDataFrame vs. uproot + awkward arrays

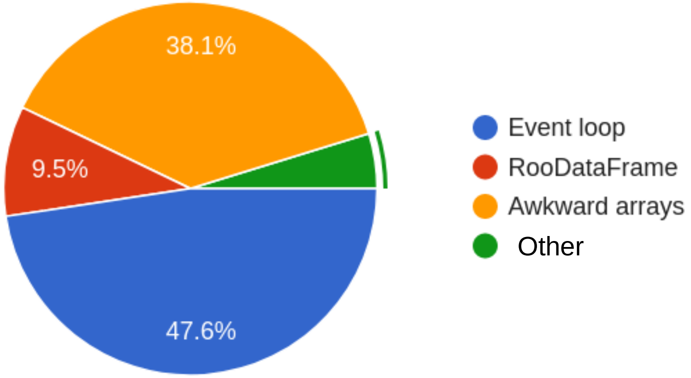
- Conceptually simple
- Reasonable in C++
- Slow in python

(never recommended
in PyRoot)

- High-level interface to TTrees
- Relies on “smart” event loop
- Usable + fast in ”python” via PyRoot

- Fully pythonic way of of handling TTree data
- Part of scikit-HEP ecosystem (uproot, hist, vector...)
- Numpy-like syntax

Which does your analysis framework rely on when processing your final Ntuples?



Processing data: TOP PAG analyzers

	Event loop	RDataFrame	uproot+awkard
From a recent survey:	NanoAOD-tools TopAnalysis HeavyNeutrino + ewkino (Ghent) URAnalysis (Rochester)	Latinos CMG tools	Pepper (mostly DESY) Coffea
Other frameworks worth a look! **From CAT general repo**		Bamboo Crown	Columnflow pocket-coffea

Relevant analysis frameworks

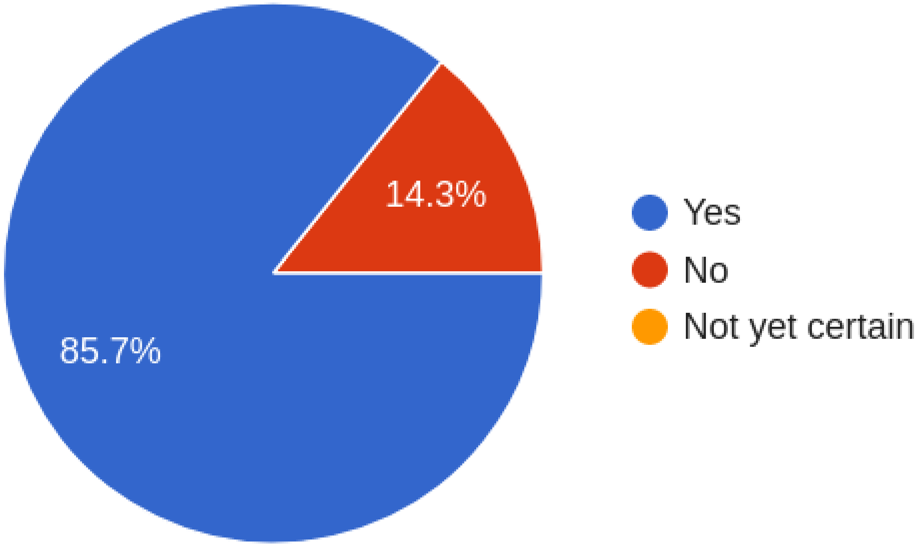
- **Coffea**
  - Early and well-known “pure-python” awkward-array-based framework
  - Now has offshoots like “pocket-coffea”
- **Pepper**
  - Framework expanding from DESY, built using some coffea classes for job splitting + execution, jet energy correction loading, and some utilities
  - Has it’s own processor and config classes, does most things independently

Statistics

- Combine is by far the most widely used statistics software in CMS.
- Originally developed by the Higgs group, it became the go-to for binned likelihood fits, parametric model fits, and limit setting (+ sometimes unfolding)

Recent TOP Survey:

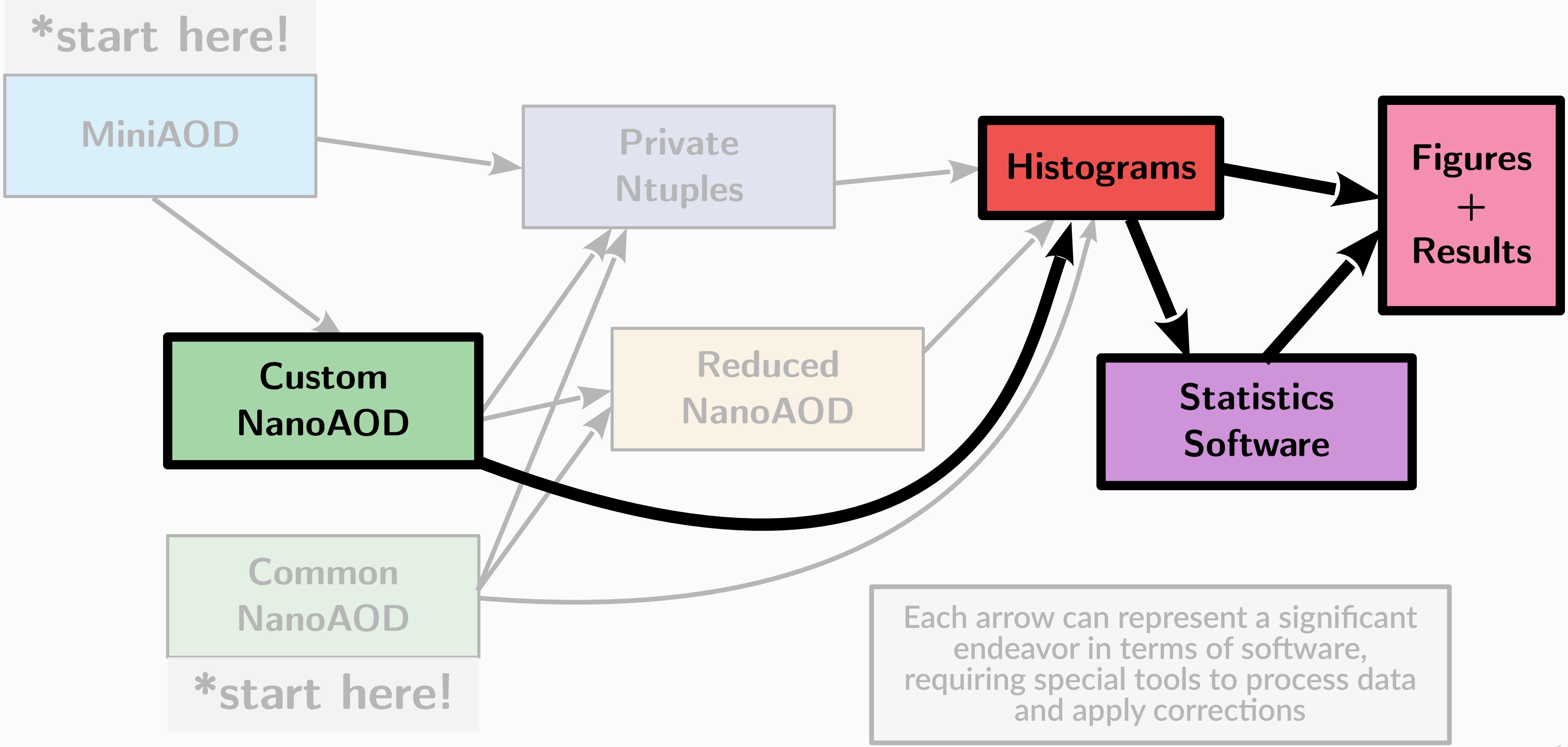
Does your analysis use combine?



*Traditionally a module of CMSSW, a standalone installation is also available

*Recently had to run with centos7 containers, but we will be using a cutting edge new version (for el9)

What does a CMS analysis look like?

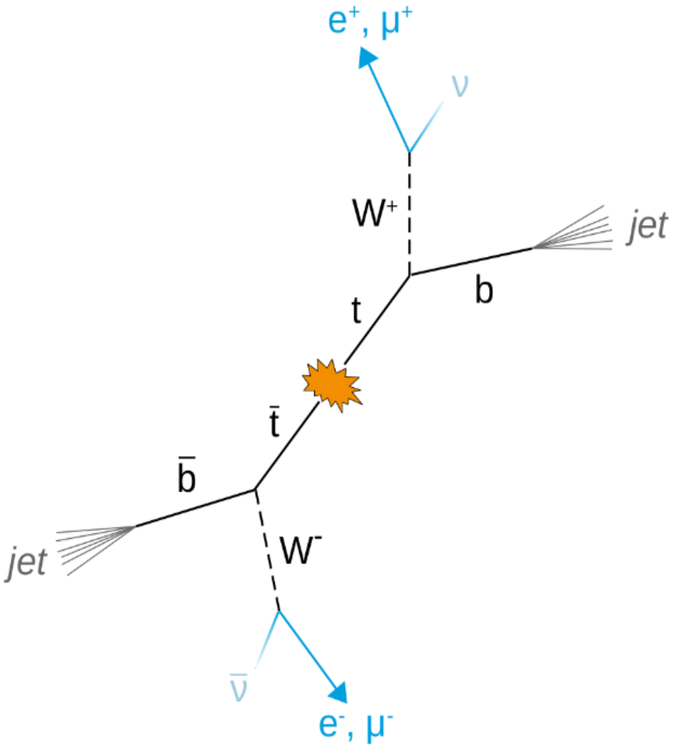


Each arrow can represent a significant endeavor in terms of software, requiring special tools to process data and apply corrections

Part 3: $\mathbf{t\bar{t}}$ cross section with Run 3 data

- Inclusive cross section: a conceptually simple measurement

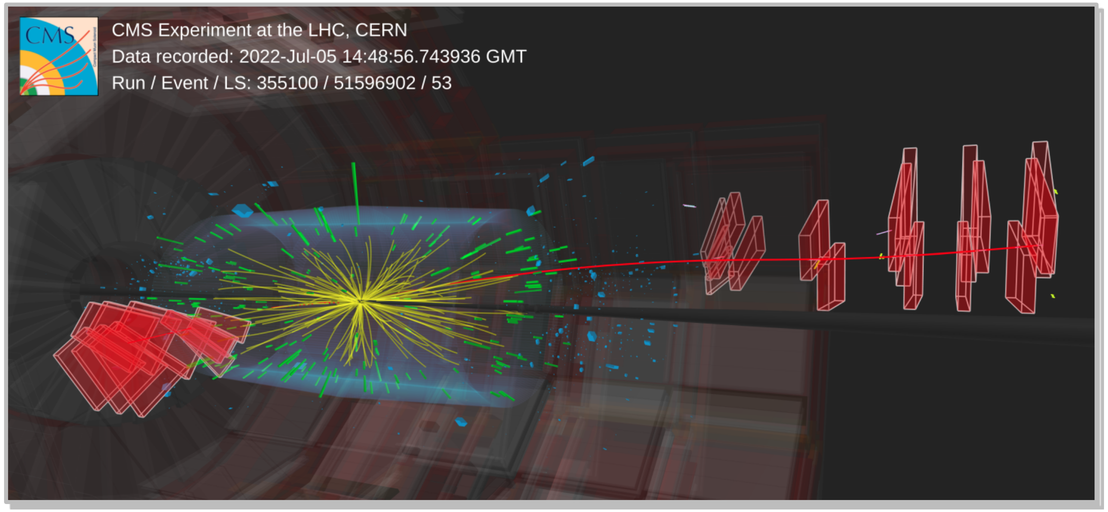
$$N_{\text{events}} = \mathcal{L}_{\text{int.}} \cdot \sigma_{t\bar{t}}$$



Run 3

In July 2022, the LHC pushed HEP into a new energy frontier

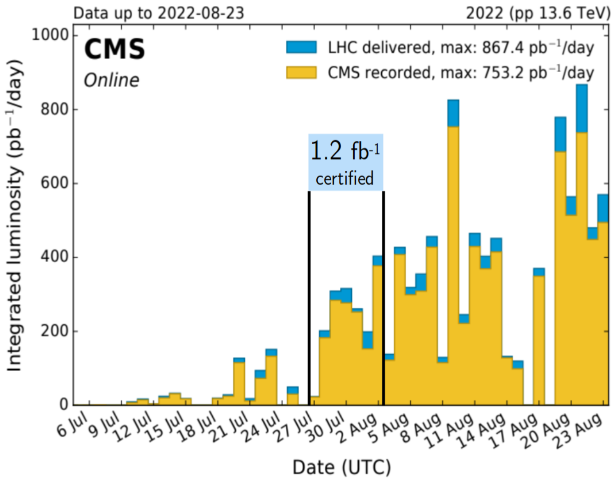
CMS control room


CMS Experiment at the LHC, CERN
Data recorded: 2022-Jul-05 14:48:56.743936 GMT
Run / Event / LS: 355100 / 51596902 / 53

$$\sqrt{s} = 13.6\ \text{TeV}$$

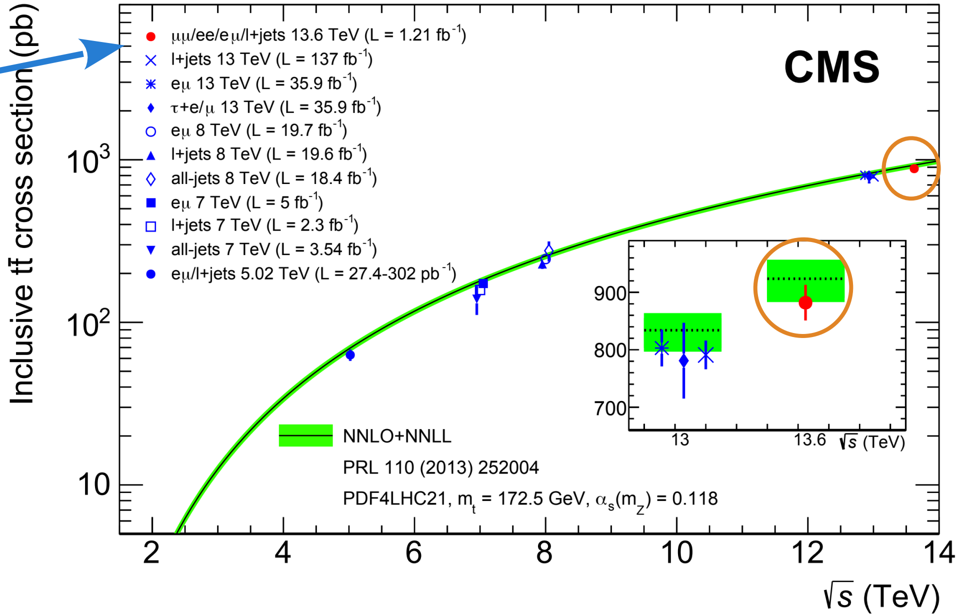
Historical details: TOP-22-012

- **First Run 3 physics result:**



Top quark pair production cross section at $\sqrt{s} = 13.6$ TeV

arXiv:2303.10680 (accepted by JHEP!)

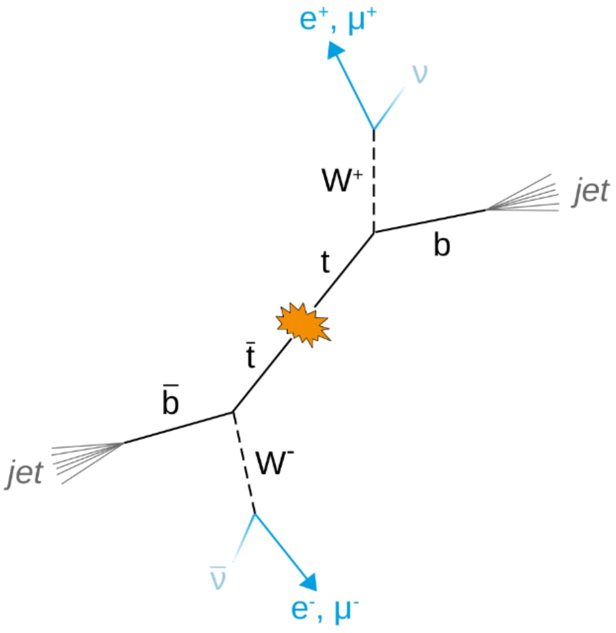


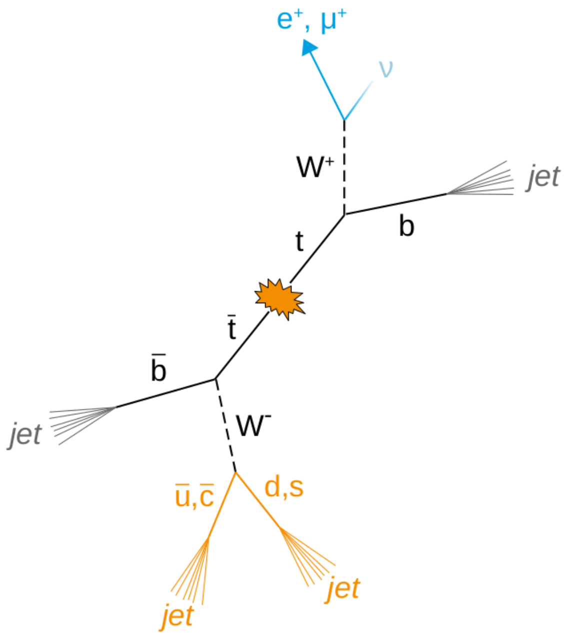
$$\boxed{\sigma_{t\bar{t}} = 881 \pm 23(\text{stat.} + \text{syst.}) \pm 20(\text{lumi.})\ \text{pb}}$$

Why $\sigma_{t\bar{t}}$?

$t\bar{t}$ production

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





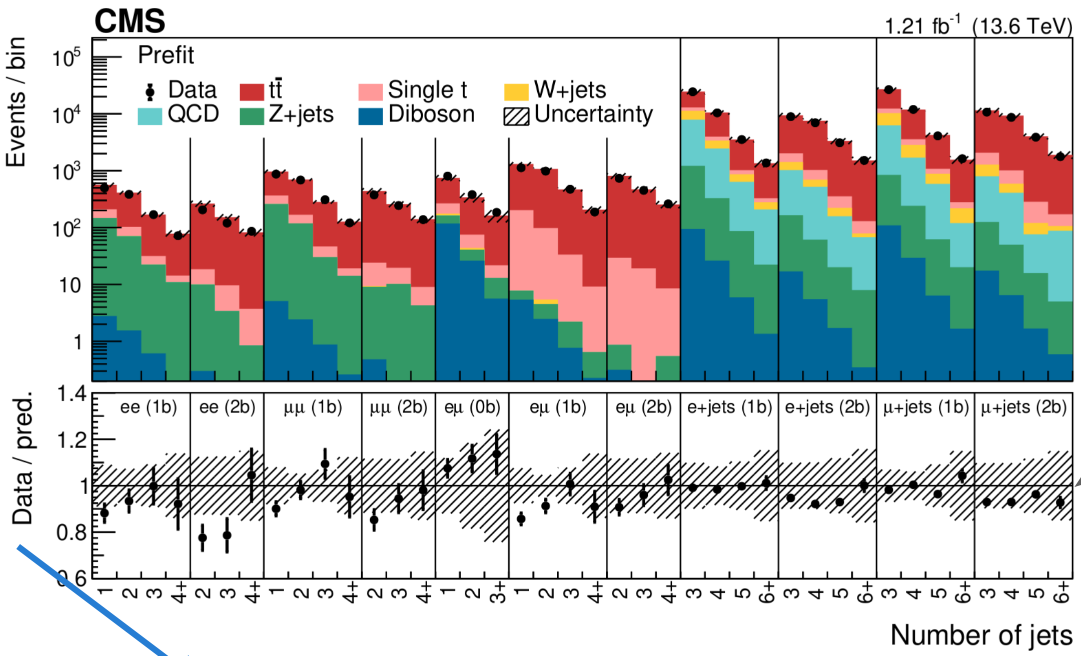
- ~10% increase in LHC Run 3

$\sqrt{s}$: 13 TeV → **13.6** TeV

$\sigma_{t\bar{t}}$: 834 pb → **924** pb

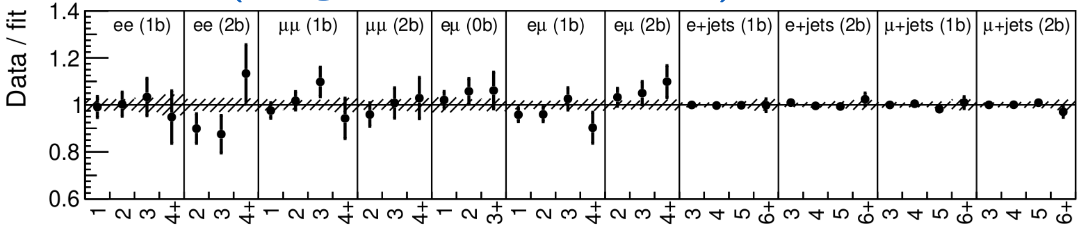
Measurement setup: original

- Bins:
  - ➤ Lepton flavor
    - ↳ b jet multiplicity
      - ↳ Jet multiplicity



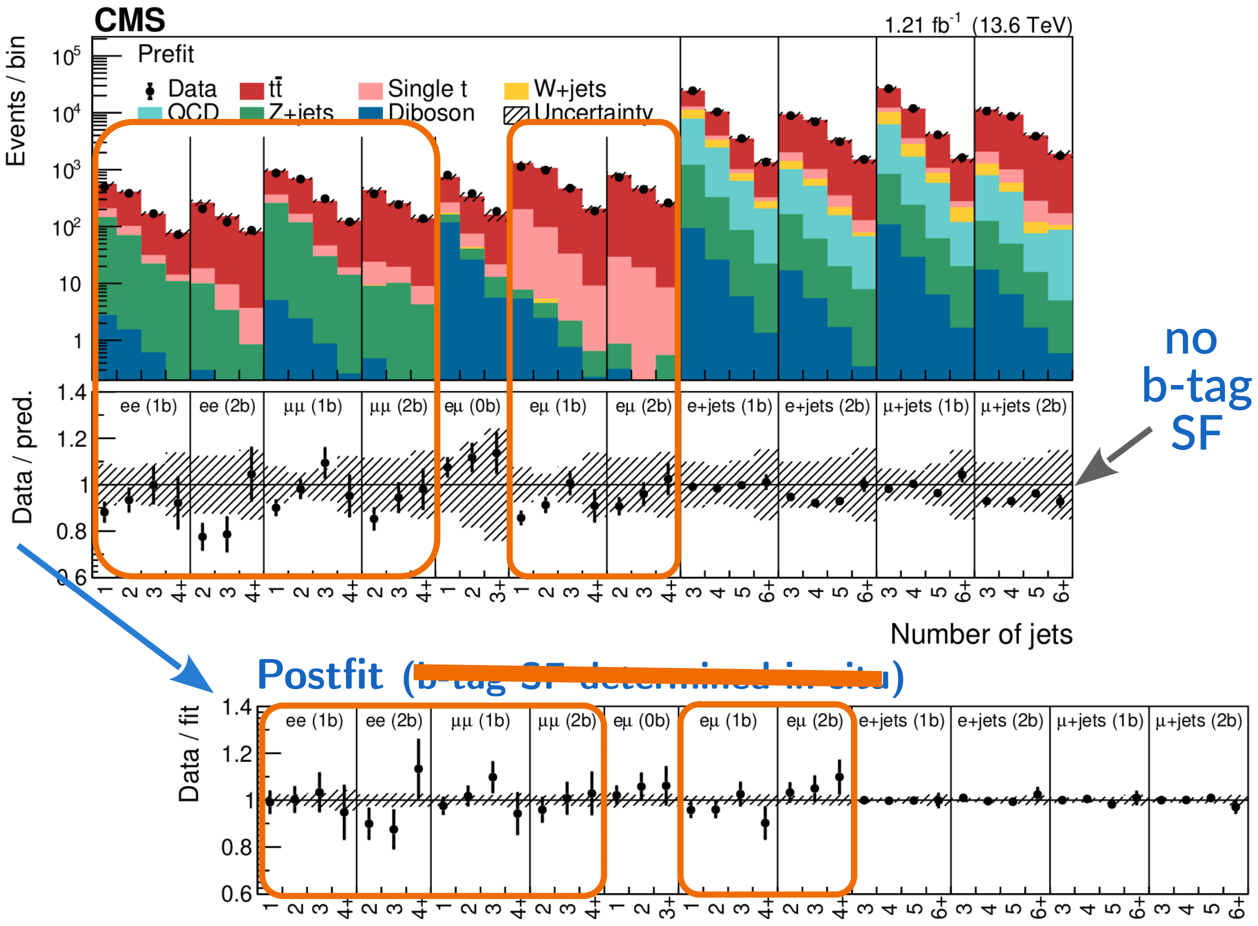
no b-tag SF

Postfit (b-tag SF determined in situ)



Measurement setup: DAS

- Bins:
  - ➤ Lepton flavor
    - ↳ b jet multiplicity
      - ↳ Jet multiplicity



no b-tag SF

Postfit (~~b-tag SF determined in situ~~)

- **DAS**
  - **→ Dilepton channel only**
  - **→ Reduced set of uncertainties**

Analysis basics

- **Pepper**, a fast python-based analysis framework
  - Fast to implement changes, re-run framework

    → fast and flexible analysis strategy played a major role in speeding up TOP-22-012,
- **Samples**
  - We provide custom nanoAOD files from the Winter22 early run 3 MC campaign—newer samples are used now elsewhere!
- **Measurement**
  - The measurement is performed via multi-channel profile likelihood fit in combine
- **Teamwork**
  - The documentation is designed for everyone to work through *most* steps, but teamwork will speed things up greatly!
  - The best analysis result will come from combining work from different participants

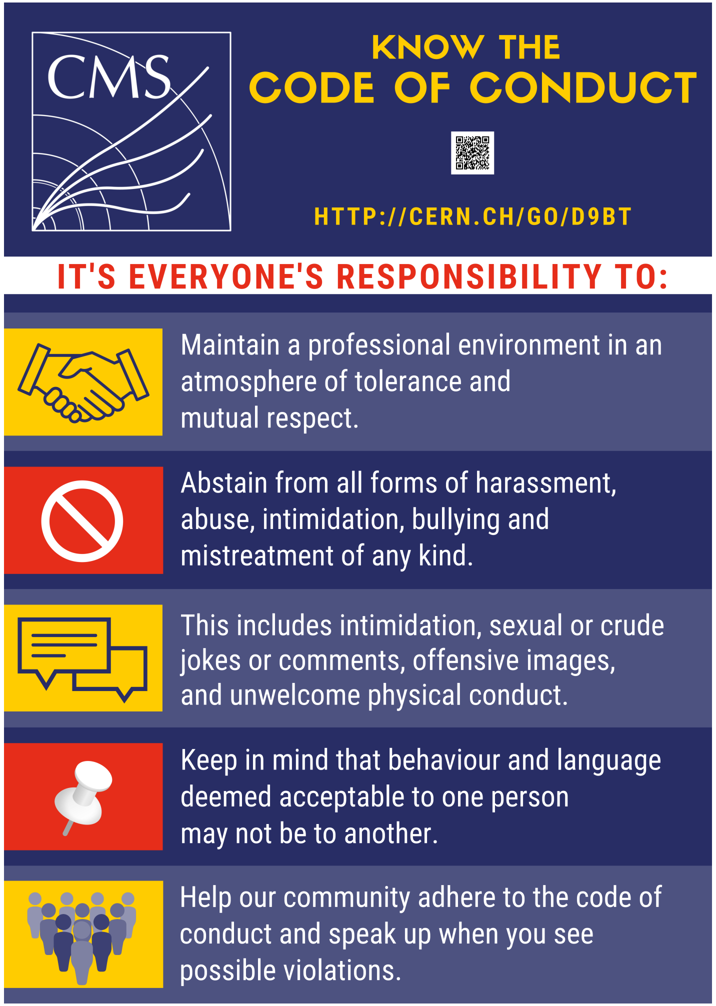
Which leads us to:

- **GOAL:** Perform the most realistic measurement possible on 1 fb^{-1} data, in 2-3 days!

Which leads us to:

- **GOAL:** Perform the most realistic measurement possible on 1 fb^{-1} data in 2-3 days!

- ...While following the <u>CMS code of conduct!</u>



KNOW THE
CODE OF CONDUCT

HTTP://CERN.CH/GO/D9BT

IT'S EVERYONE'S RESPONSIBILITY TO:

Maintain a professional environment in an atmosphere of tolerance and mutual respect.

Abstain from all forms of harassment, abuse, intimidation, bullying and mistreatment of any kind.

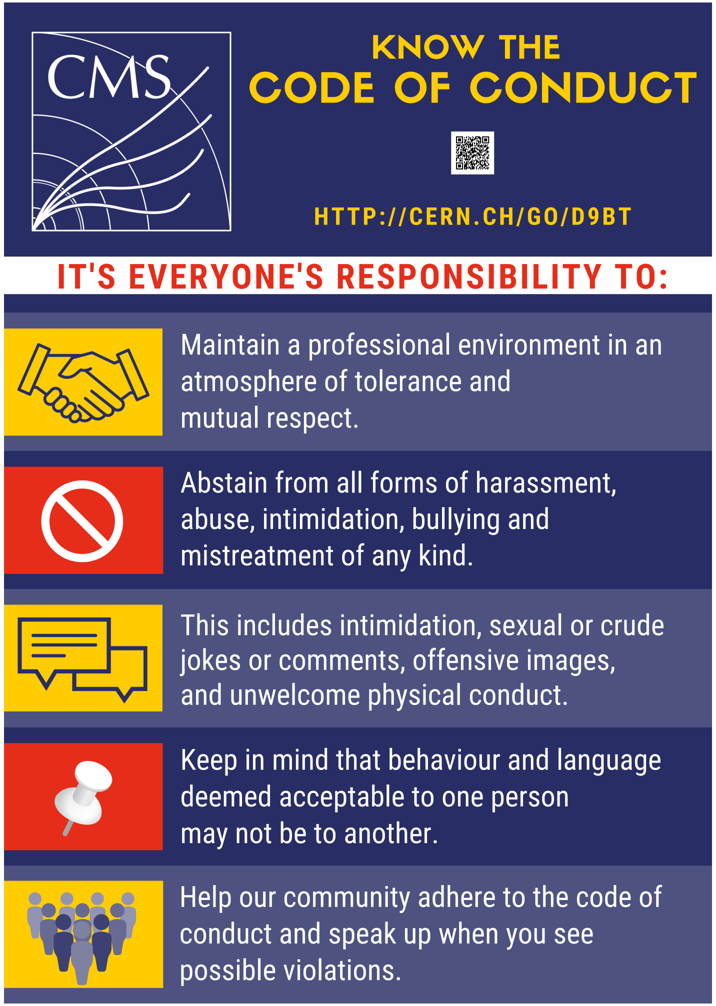
This includes intimidation, sexual or crude jokes or comments, offensive images, and unwelcome physical conduct.

Keep in mind that behaviour and language deemed acceptable to one person may not be to another.

Help our community adhere to the code of conduct and speak up when you see possible violations.

Which leads us to:

- **GOAL:** Perform the most realistic measurement possible on 1 fb^{-1} data in 2-3 days!
- …While following the CMS code of conduct!
- …And not staying up too late working on Friday!!!



KNOW THE
CODE OF CONDUCT

HTTP://CERN.CH/GO/D9BT

IT'S EVERYONE'S RESPONSIBILITY TO:

Maintain a professional environment in an atmosphere of tolerance and mutual respect.

Abstain from all forms of harassment, abuse, intimidation, bullying and mistreatment of any kind.

This includes intimidation, sexual or crude jokes or comments, offensive images, and unwelcome physical conduct.

Keep in mind that behaviour and language deemed acceptable to one person may not be to another.

Help our community adhere to the code of conduct and speak up when you see possible violations.

Backup

Relevant tools + docs

- Pepper (DAS build)
- scikit-HEP packages:
  - **Hist:** fast multi-dim histograms (front-end for boost histograms) *python-based:* easy to check things in jupyter notebook, interface with matplotlib
  - **Awkward**: "Jagged" arrays in python
- Combine