Triple-differential measurement of the dijet cross section at CMS

7th KSETA Plenary Workshop

21 February, 2020

Daniel Savoiu, Klaus Rabbertz

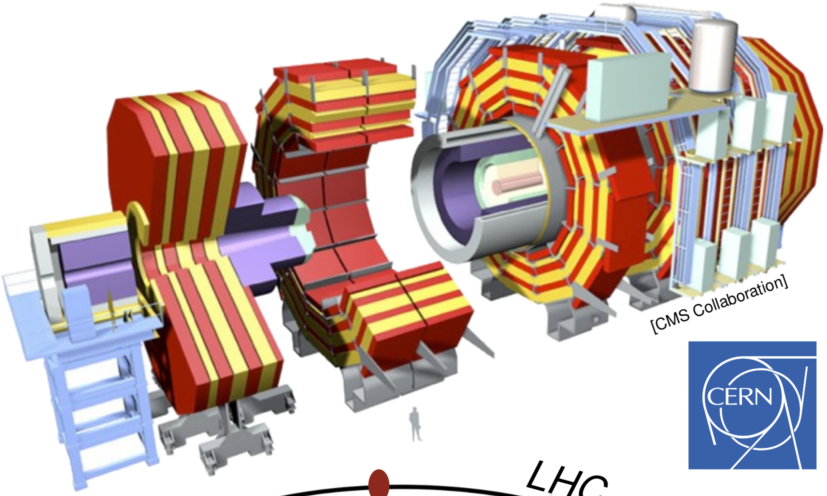
Institute of Experimental Particle Physics (ETP)

The CMS experiment

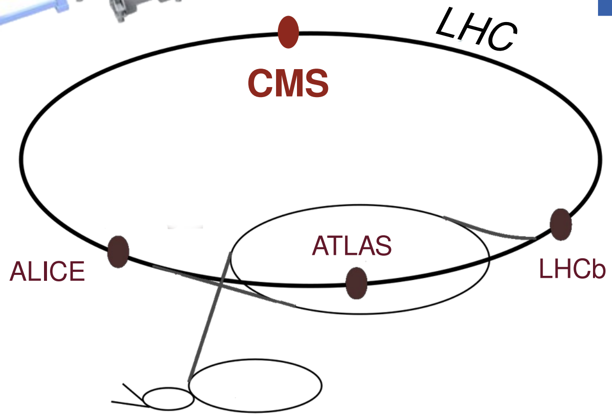
- "**Compact Muon Solenoid**" – large general-purpose particle detector at the Large Hadron Collider (LHC) at CERN

- Run 2 (2016–2018): ~**140 fb**$^{-1}$ of data collected
  - **systematic uncertainties** are becoming the limiting factor in many analyses
  - improving **precision** becomes more important

- hadron collider: composite initial state
  - **proton structure** remains a significant source of uncertainty



[CMS Collaboration]

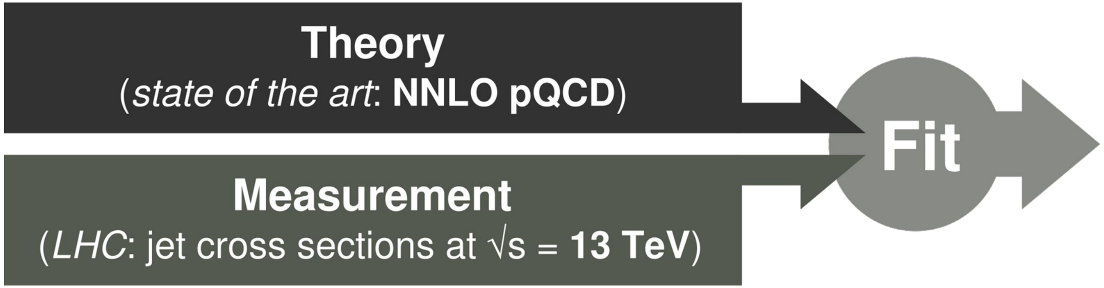


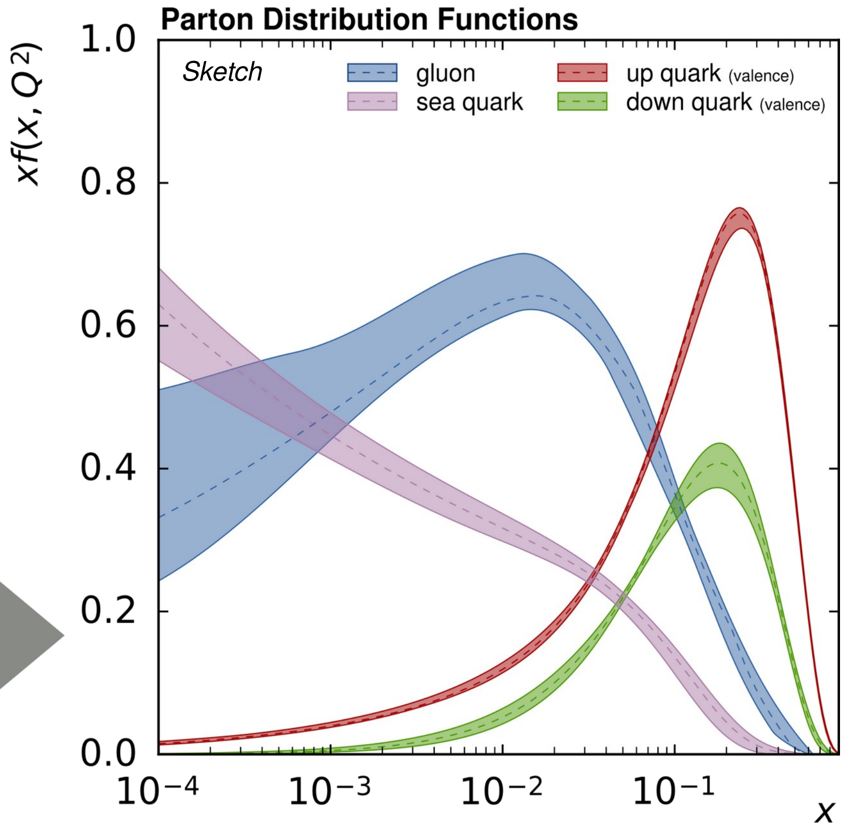
[https://en.wikipedia.org/wiki/File:LHC.svg]

PDFs & jet cross sections

- **parton distribution functions** (PDFs) vital for physics understanding at proton colliders
  - parameterized probability of finding a constituent *parton* (quark or gluon)
- not given by theory

  → **fit** of theory predictions to measurements
- **jet cross sections** particularly suited for this

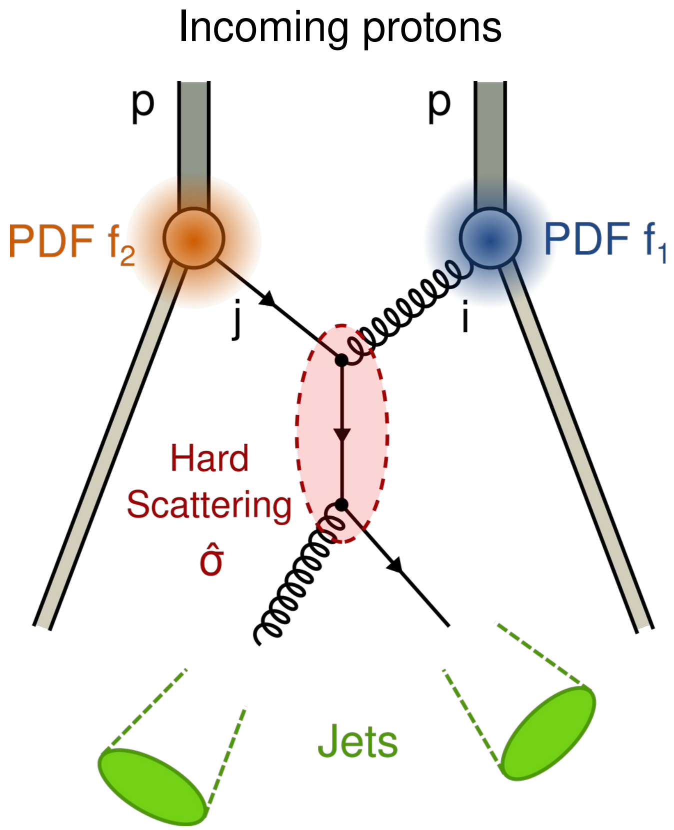




Dijet production

- events with **two jets** (or more) in the final state
- differential cross section sensitive to $\boldsymbol{\alpha_s}$ and parton distribution functions **PDFs**
- one of the **highest cross sections** at the LHC
  → high event rates
  → low background

- **triple-differential** measurement
  - event counts sufficiently high for **three-dimensional** division of phase space
  - beneficial for constraining PDFs by exploiting dependence on **dijet topology**



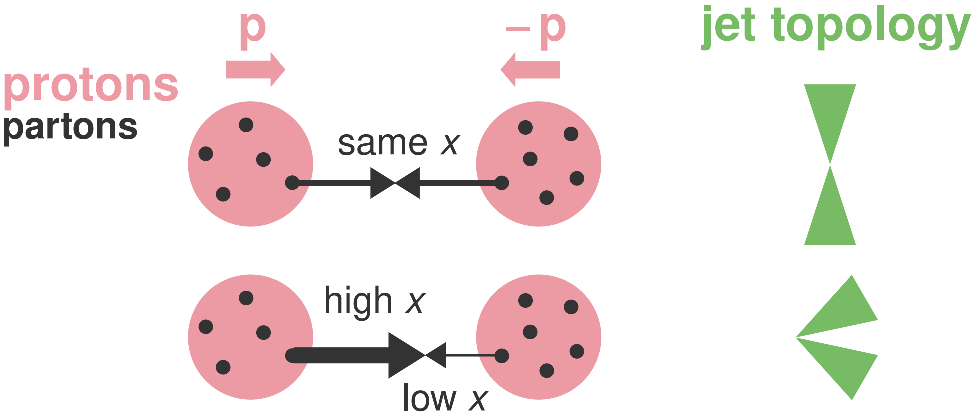
PDF sensitivity

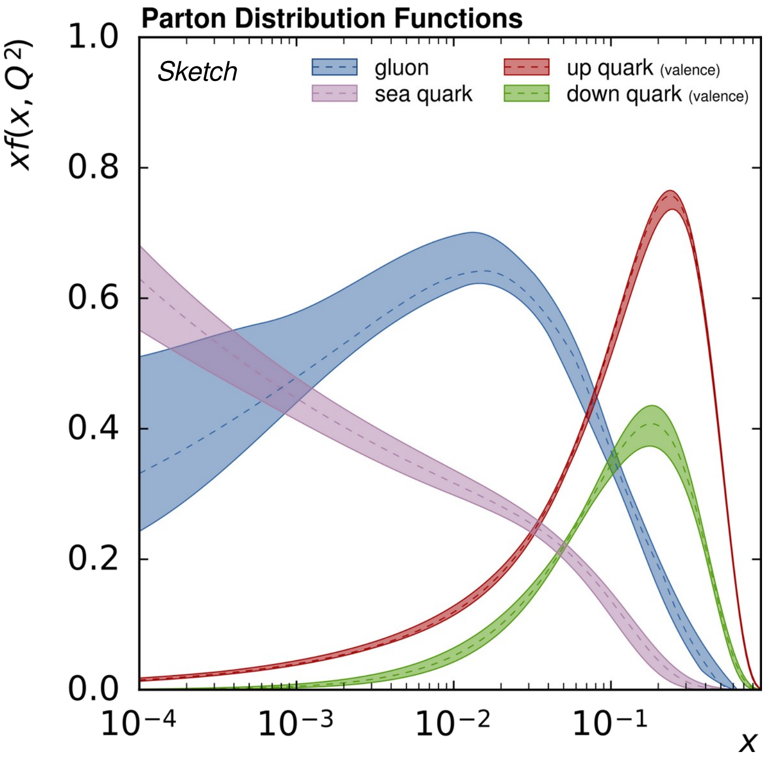
- **cross section** for processes at pp colliders:

$$\int_0^1 \mathrm{d}x_1 \int_0^1 \mathrm{d}x_2\ \underbrace{f_1(x_1, Q^2)\ f_2(x_2, Q^2)}_{\text{PDFs}}\ \hat{\sigma}(x_1, x_2)$$

momentum fractions carried by partons

partonic cross section





- **jet topology** can help disentangle PDFs from unrelated effects across a wider x region

Triple-differential measurement

- *objective*: **triple-differential** measurement of the **dijet production** cross section

$$\frac{d^3\sigma}{dy^*\, dy_b\, d\langle p_T\rangle_{1,2}} = \frac{1}{\varepsilon \cdot \mathcal{L}} \frac{N_{\text{events}}}{\Delta y^*\, \Delta y_b\, \Delta\langle p_T\rangle_{1,2}}$$

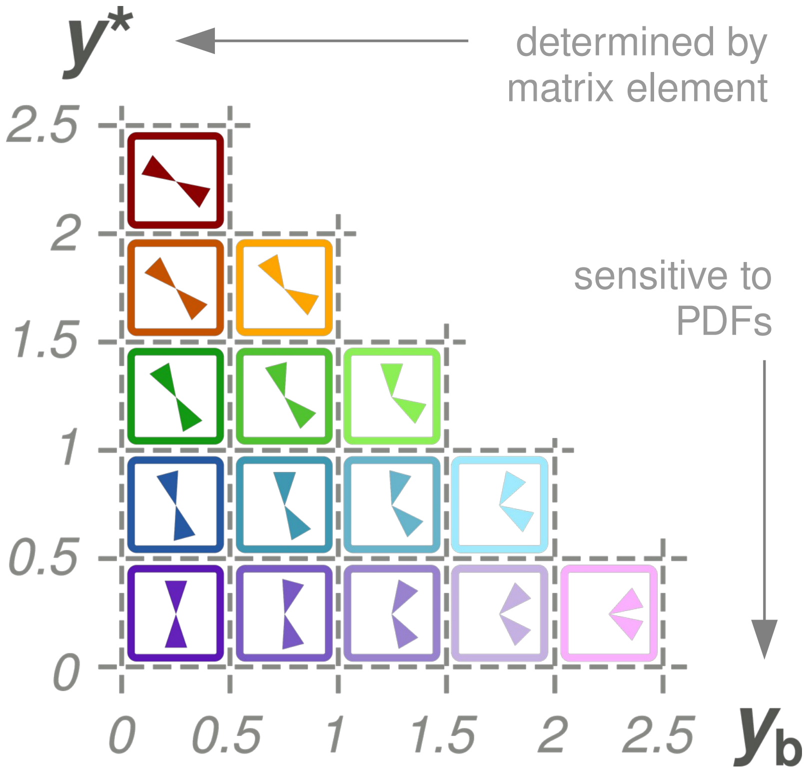
- variables chosen in such a way as to exploit jet topology and increase PDF sensitivity:

Variables

$\langle p_T\rangle = ½ \, (p_T^{\text{jet1}} + p_T^{\text{jet2}})$ "p_T-average"

$y_b = ½ \, | y^{\text{jet1}} + y^{\text{jet2}} |$ "y-boost"

$y^* = ½ \, | y^{\text{jet1}} - y^{\text{jet2}} |$ "y-star"



Data & event selection

- data set: **2016** data (35.9 fb^{-1} @ 13 TeV)

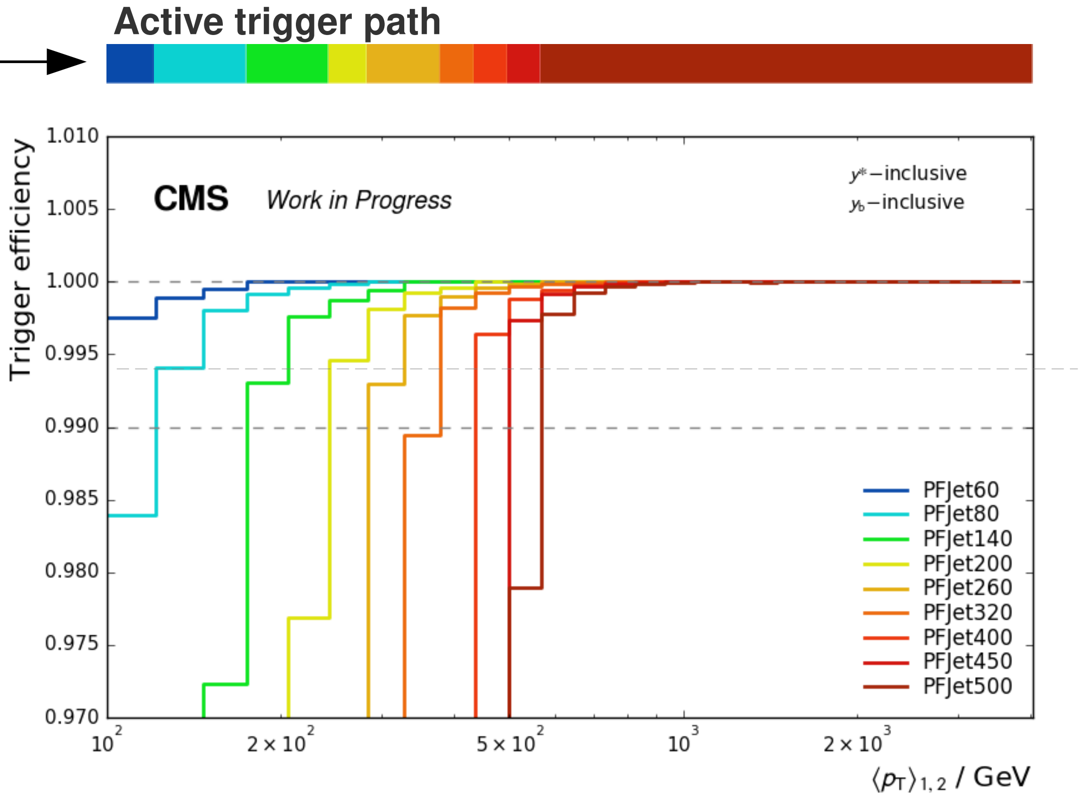
- select events with at least **two "good" jets**
  - jets reconstructed with two jet radii: R = **0.4** and R = **0.8**
  - calibrated **jet energy scale**
  - calibrated **jet energy resolution** (in Monte Carlo simulation)
  - jets required to pass additional **jet identification** criteria (prevent spurious jets due to noise)

- kinematic cuts for final selection
  - $p_T^{jet1} > 100$ GeV
  - $p_T^{jet2} > 50$ GeV
  - $|y^{jet1}| < 3.0$
  - $|y^{jet2}| < 3.0$
  - $E_T^{miss} \Big/ \sum_i E_{T,i} < 0.3$

background rejection

Trigger strategy

- cross section is **steeply falling** in $\langle p_T \rangle$
  - to cover entire phase space, multiple prescaled triggers must be used
- no single trigger fully efficient across entire phase space
  - divide $\langle p_T \rangle$ into regions
  - only one trigger active per region
- switch to next trigger once fully efficient



Trigger combination

- measure efficiency using **bootstrapping** method:
  - pre-select events triggered by T_{n-1}
  - emulate T_n trigger decision:

$$\varepsilon(T_n) = N(T_n \cap T_{n-1}) / N(T_{n-1})$$

- **combine** triggers:
  - only one trigger **active** in each measurement bin → T_a
  - choose lowest-prescale trigger with an **efficiency ε > 99.5%**

- event yield maximized while ensuring full efficiency

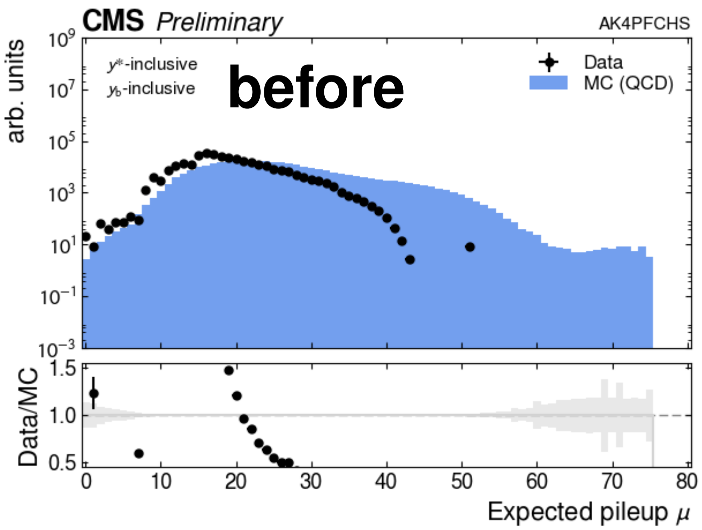
Trigger path	Eff. lumi. 2016
HLT_DiPFJetAve40	0.1 pb^{-1}
HLT_DiPFJetAve60	1.7 pb^{-1}
HLT_DiPFJetAve80	4.2 pb^{-1}
HLT_DiPFJetAve140	27.6 pb^{-1}
HLT_DiPFJetAve200	138.7 pb^{-1}
HLT_DiPFJetAve260	522.7 pb^{-1}
HLT_DiPFJetAve320	2968.7 pb^{-1}
HLT_DiPFJetAve400	9026.4 pb^{-1}
HLT_DiPFJetAve500	29309.3 pb^{-1}

→ event weight in data:

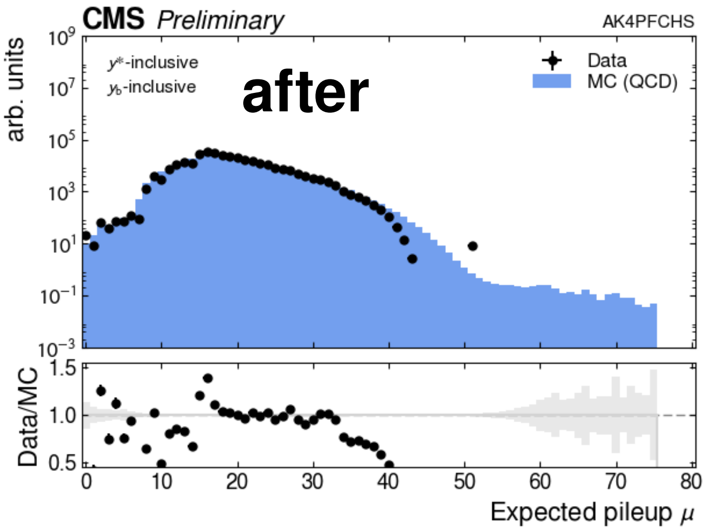
$$w_{Data} = \begin{cases} 1 / L_{eff}(T_a) & \text{if trigger } T_a \text{ fired,} \\ 0 & \text{otherwise} \end{cases}$$

Simulations & Pileup

- Monte Carlo simulations available in CMS:
  - Madgraph + Pythia (LO pQCD)
  - full detector and pileup simulation
- pileup distribution in **Monte Carlo** adapted to actual **data** taking conditions
  - events are **reweighted** based on expected number of pileup interactions μ

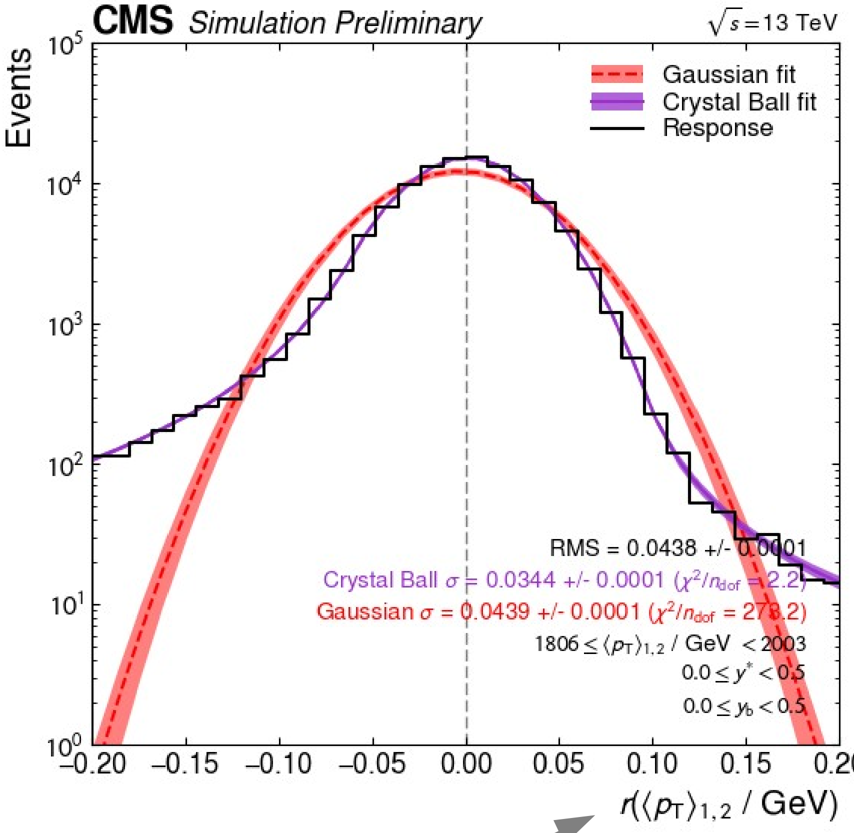


$w_{PU} = \mu_{data} / \mu_{MC}$



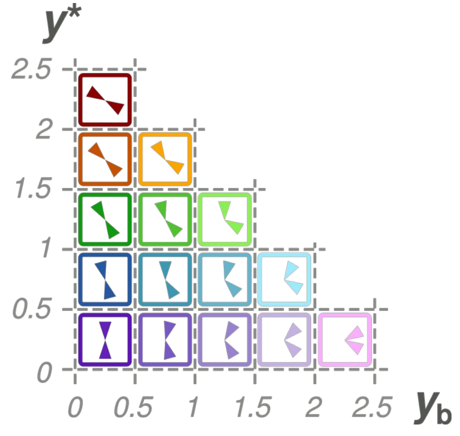
Unfolding

- measurements affected by finite **detector resolution**
  - distribution of reconstructed quantity is **smeared** compared to the true distribution
- resolution is **specific to each detector**
  - direct **comparison** with other data / with theory not directly possible
  - limited usefulness or theory fits
- **unfolding** → “reversing” detector smearing effects

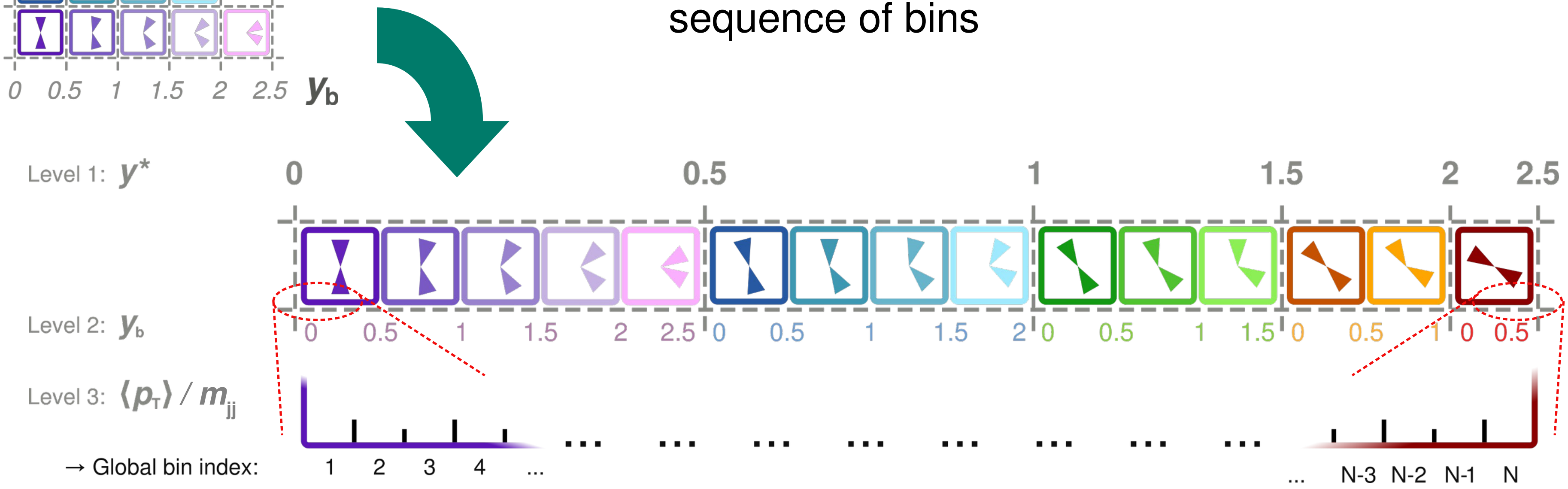


$r(x) = (x_{detector} - x_{true}) / x_{true}$

Insert: phase space unraveling



- finite detector resolution causes bin-to-bin **migrations** of events
- analysis has **three-dimensional** bin structure
  - **unravel** phase space (y^*, y_b, $\langle p_T \rangle$) into a "flat" sequence of bins

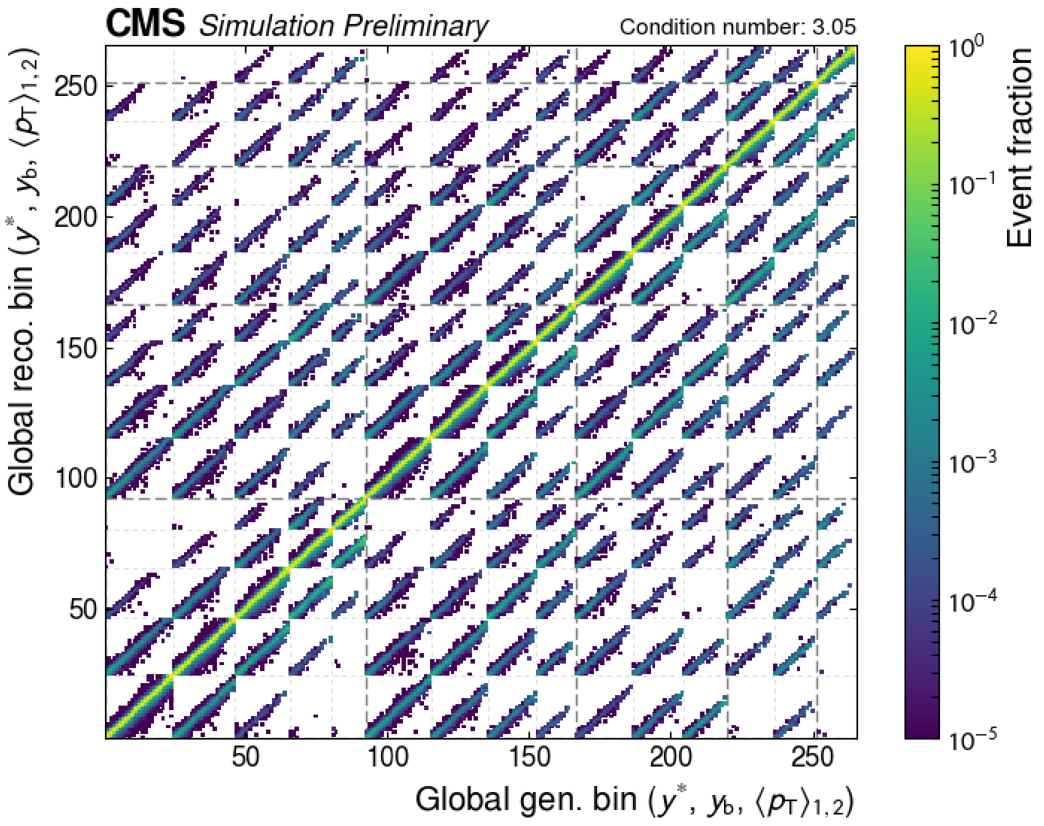


Detector response matrix

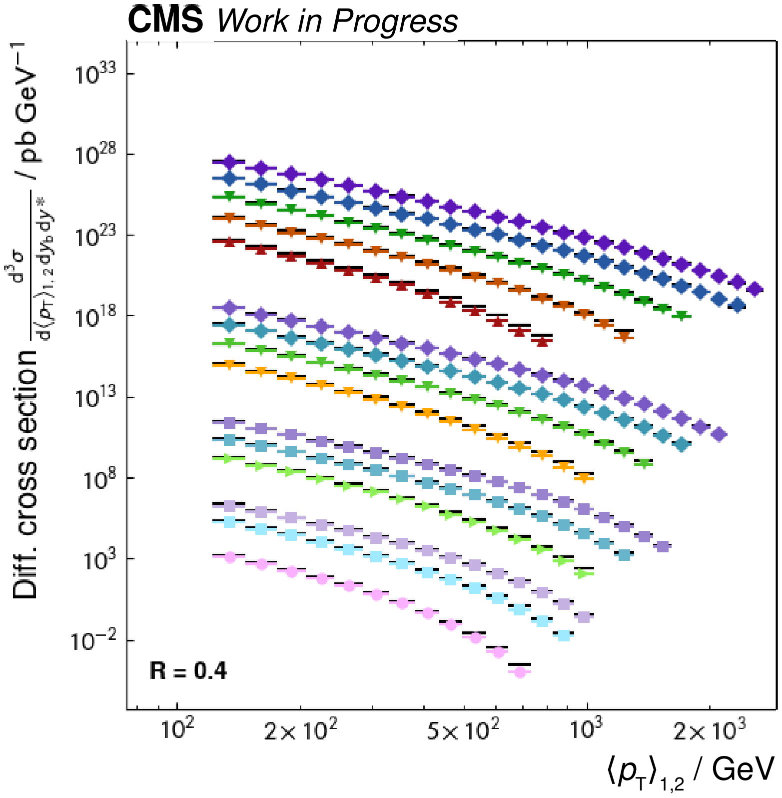
- effect of finite detector resolution can be described by a **matrix**
  - entry A_{ij} → probability of an event **generated** in bin j to be **reconstructed** in bin i
  - **estimated** from Monte Carlo simulation where generator-level information is known ("Monte Carlo truth")

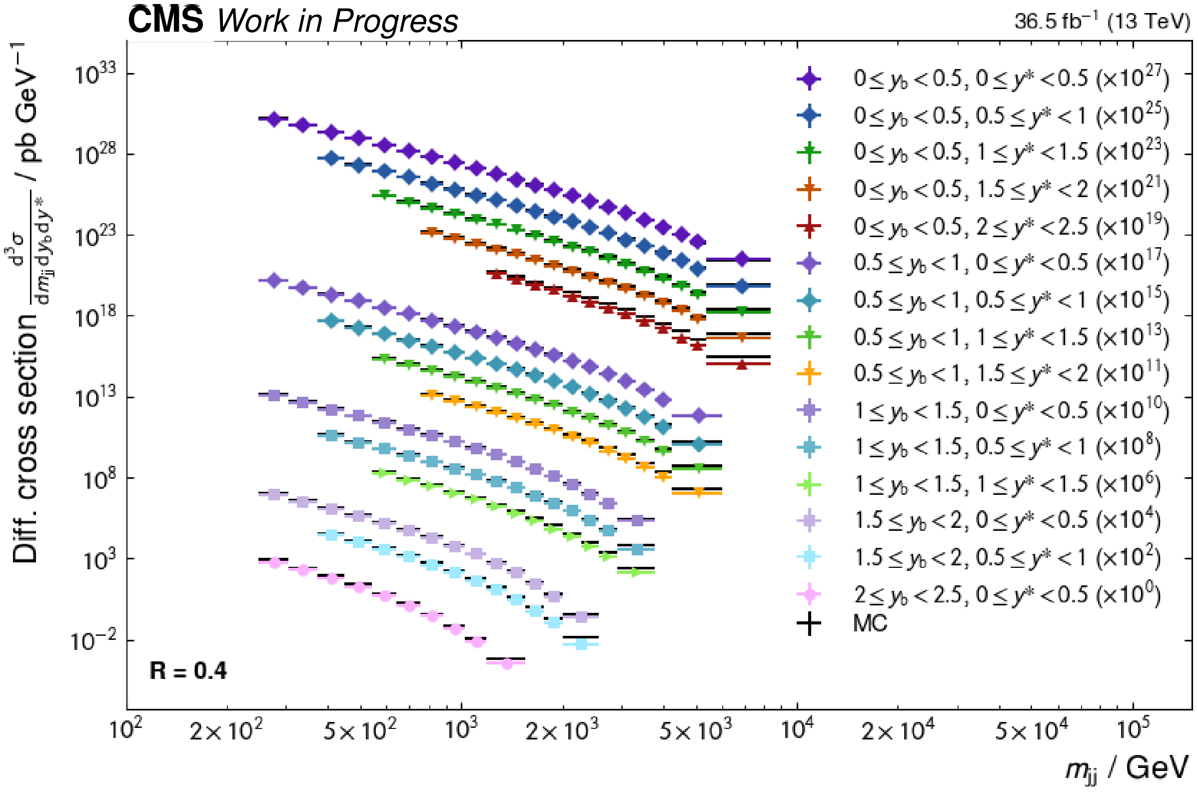
- unfolded distributions are determined by **inverting** this matrix and multiplying it to the measured distribution

$$\rightarrow g_j = \sum_i (A^{-1})_{ij} (r_i - f_i)$$



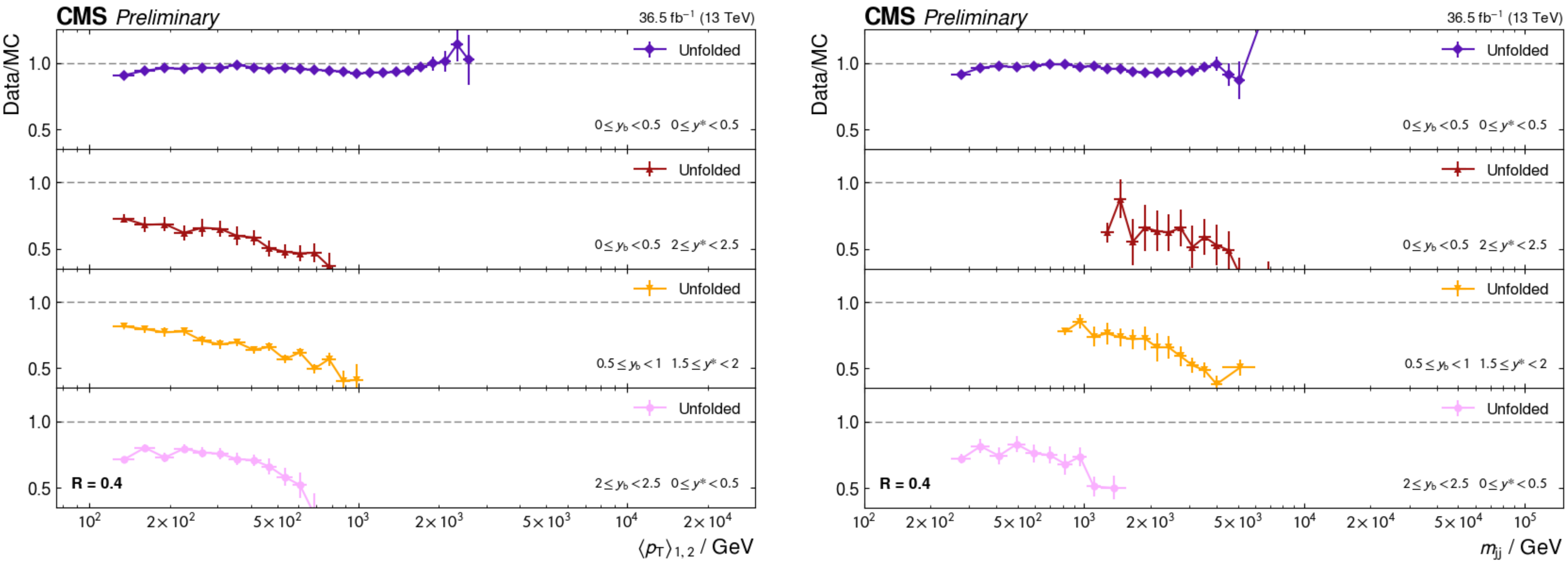
Unfolded cross section





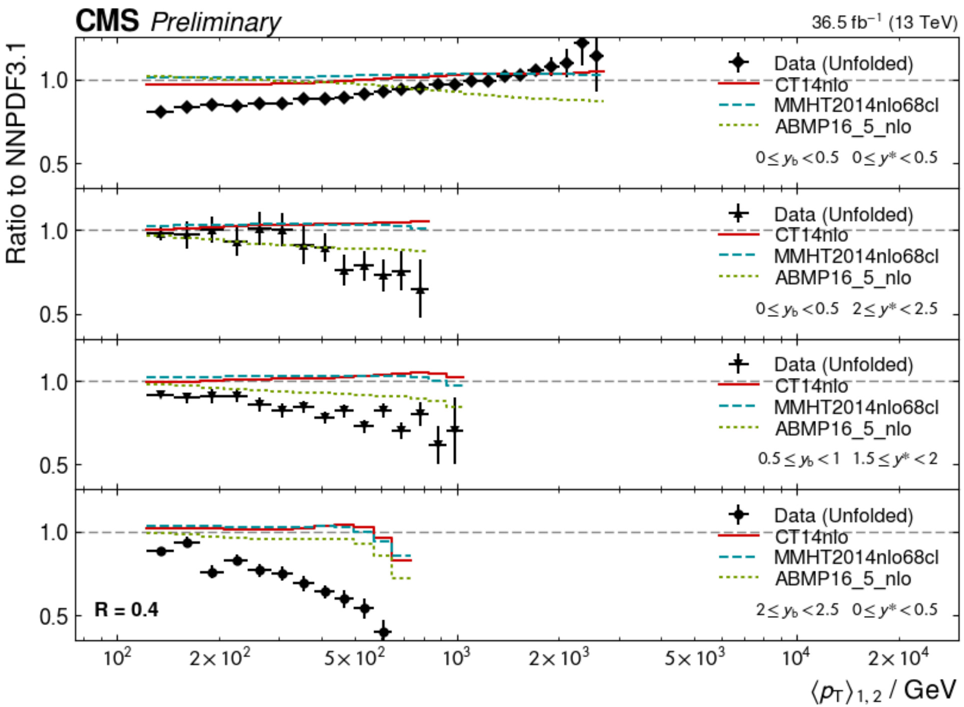
- shown: $R = 0.4$ ($R = 0.8$ similar)
- *right*: complementary measurement in dijet invariant mass (m_{jj}) instead of $\langle p_T\rangle$

Comparison to LO simulation (Madgraph + Pythia)



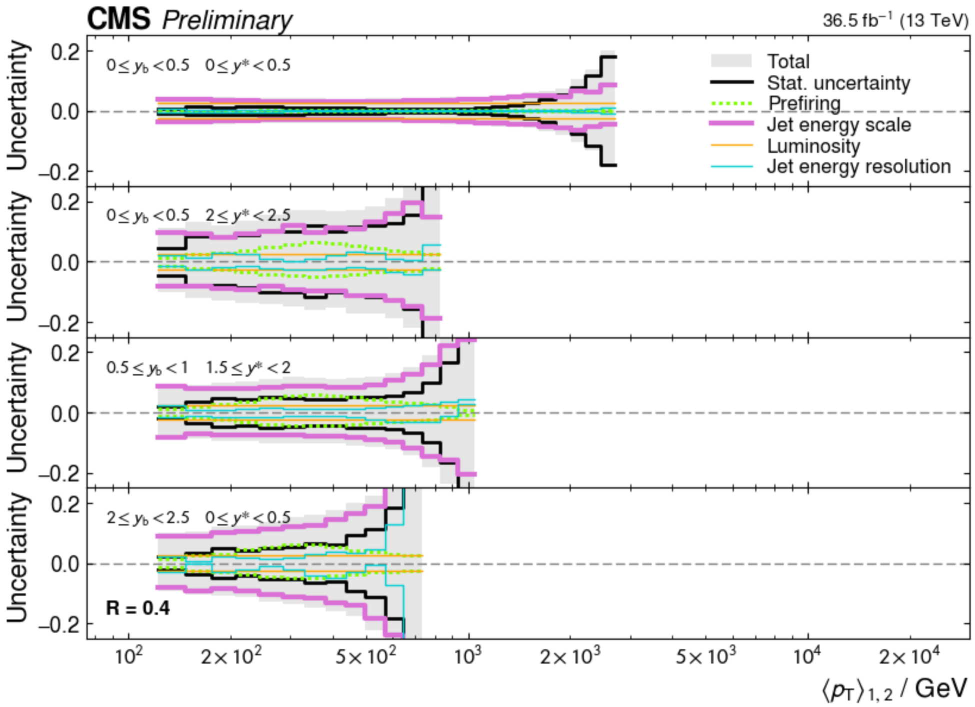
- *shown*: ratio of unfolded data to generator-level distribution in Monte Carlo (LO)
- data deviates from prediction at high values of y^* and y_b

Comparison to fixed-order theory (NNLOJET + fastNLO)



- *shown*: ratios to fixed-order **NLO theory**
  - no non-perturbative corrections yet
  - NNLO available soon
- *points*: measured **unfolded cross sections** with **statistical** uncertainties
- *lines*: theory with alternative **PDFs**
- improved description at high y^*, deviation at high y_b remains
- indication that PDFs may benefit from differential measurements in y_b

Uncertainties



- uncertainties between 5% and 20%
- largest contribution from **jet energy scale**

- **statistical** uncertainty
- **prefiring** uncertainty
  - uncertainty on correction of inefficiency due to *trigger prefiring*
- **jet energy scale** uncertainty
  - estimated by applying systematically shifted *jet energy corrections*
- **luminosity** uncertainty
  - official recommendation: 2.5%
- **jet energy resolution** uncertainty
  - estimated by systematically varying the *jet energy resolution*

Summary

- **triple-differential** dijet cross section measurement at 13 TeV with **2016** data (35.9 fb^{-1})
  - as a function of (y^*, y_b, $\langle p_T \rangle$) and (y^*, y_b, m_{jj})
  - **2016** data set (35.9 fb^{-1})
- **combination** of multiple trigger paths to maximize accessible phase space
- **3D unfolding** via phase space unraveling and matrix inversion
- **uncertainties** dominated by jet energy scale systematics: 5% – 25%
- comparison to fixed-order theory **NLO calculations** (NNLO in progress)

Ba

ck

up

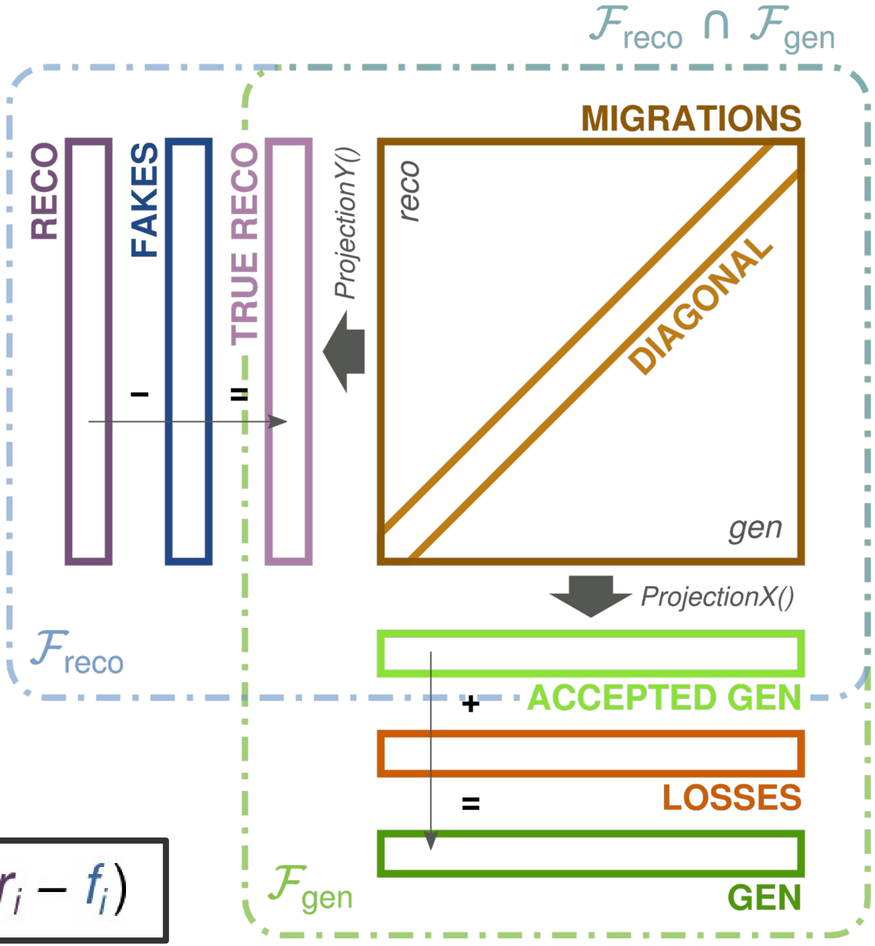
follows

Unfolding (technical details)

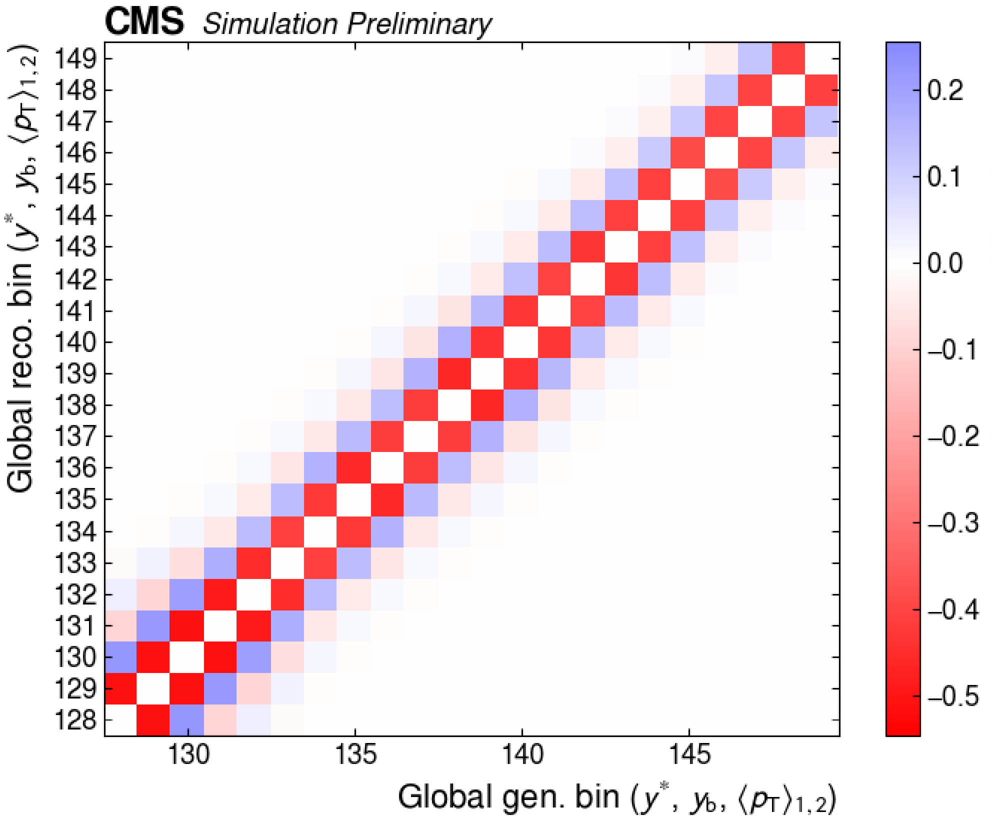
- 2 event selections:
  - $\mathcal{F}_{reco}$ → cuts applied on reco-level objects
  - $\mathcal{F}_{gen}$ → cuts applied on gen-level objects
- fill histograms with event counts per gen bin (j) and/or reco bin (i):
  - M_{ij} → migrations
  - r_i → reco-level events
  - g_j → gen-level events
- obtain fakes and response matrix:

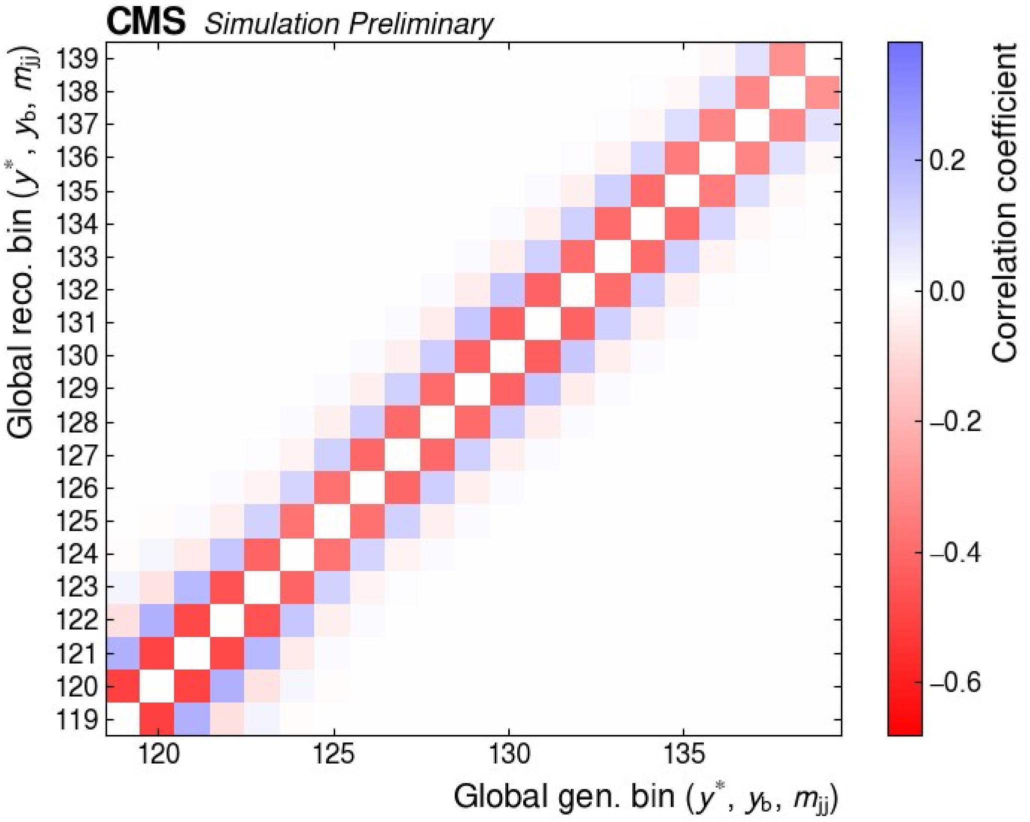
$$f_i = \underbrace{r_i - \sum_j M_{ij}}_{t_i} \qquad A_{ij} = \frac{M_{ij}}{g_j}$$

- unfold via matrix inversion: $\boxed{\rightarrow g_j = \sum_i (A^{-1})_{ij}\,(r_i - f_i)}$



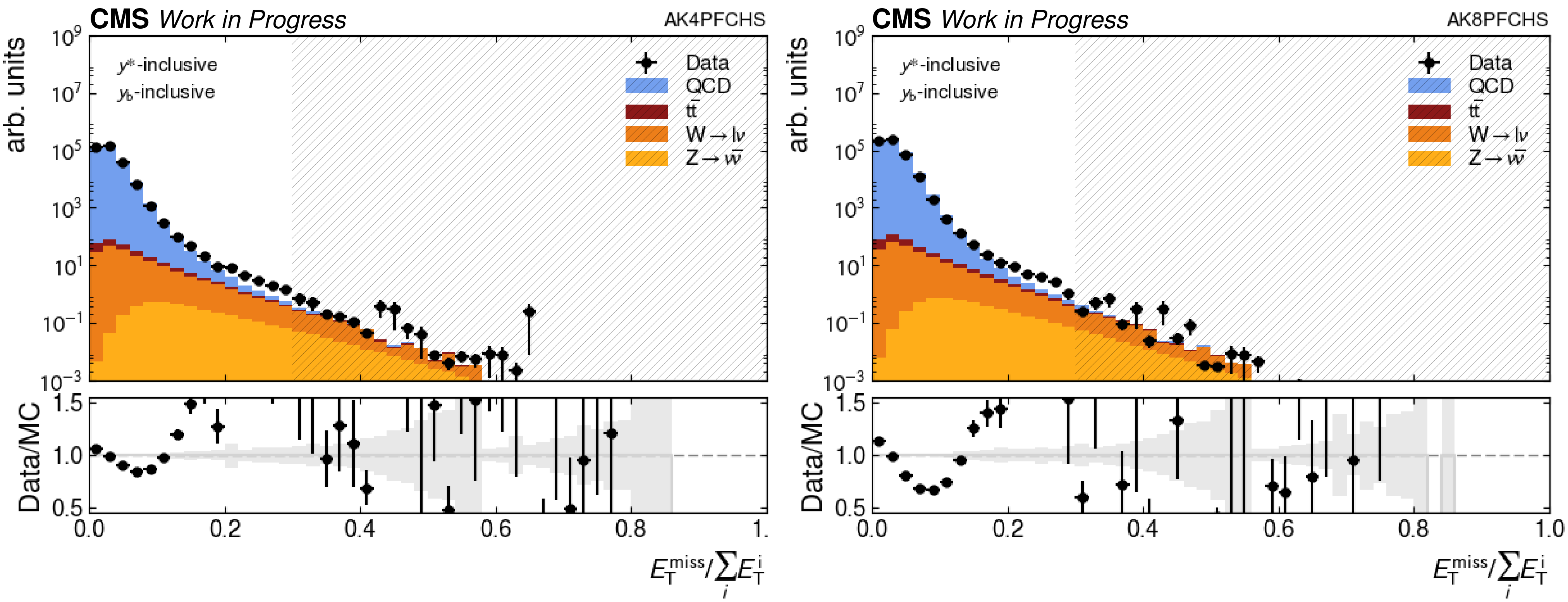
Response matrix (correlation coefficients)





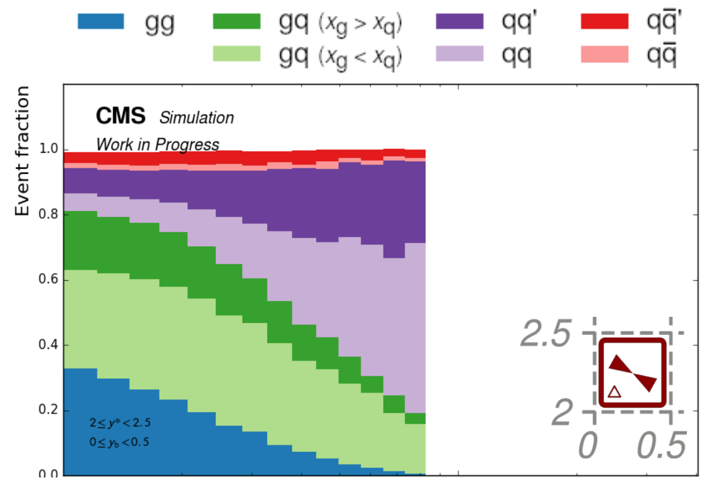
- unfolding yields familiar correlation pattern:
  - anti-correlation for nearest neighbor bins, positive correlation for second-nearest, etc.

Background rejection



- requirement $E_T^{miss} \Big/ \sum_i E_{T,i} < 0.3$ filters out background from W/Z + jets events

Subprocess composition



- initial state partons determine PDF contribution
- composition varies across ($\langle p_T \rangle$, y^*, y_b) phase space

Boosted region

- dominated by quark-gluon scattering
- gluon typically softer than quark
  → decorrelate g and q PDF contributions