# Data Scouting in CMS

**Dustin Anderson**

*For the CMS Collaboration*
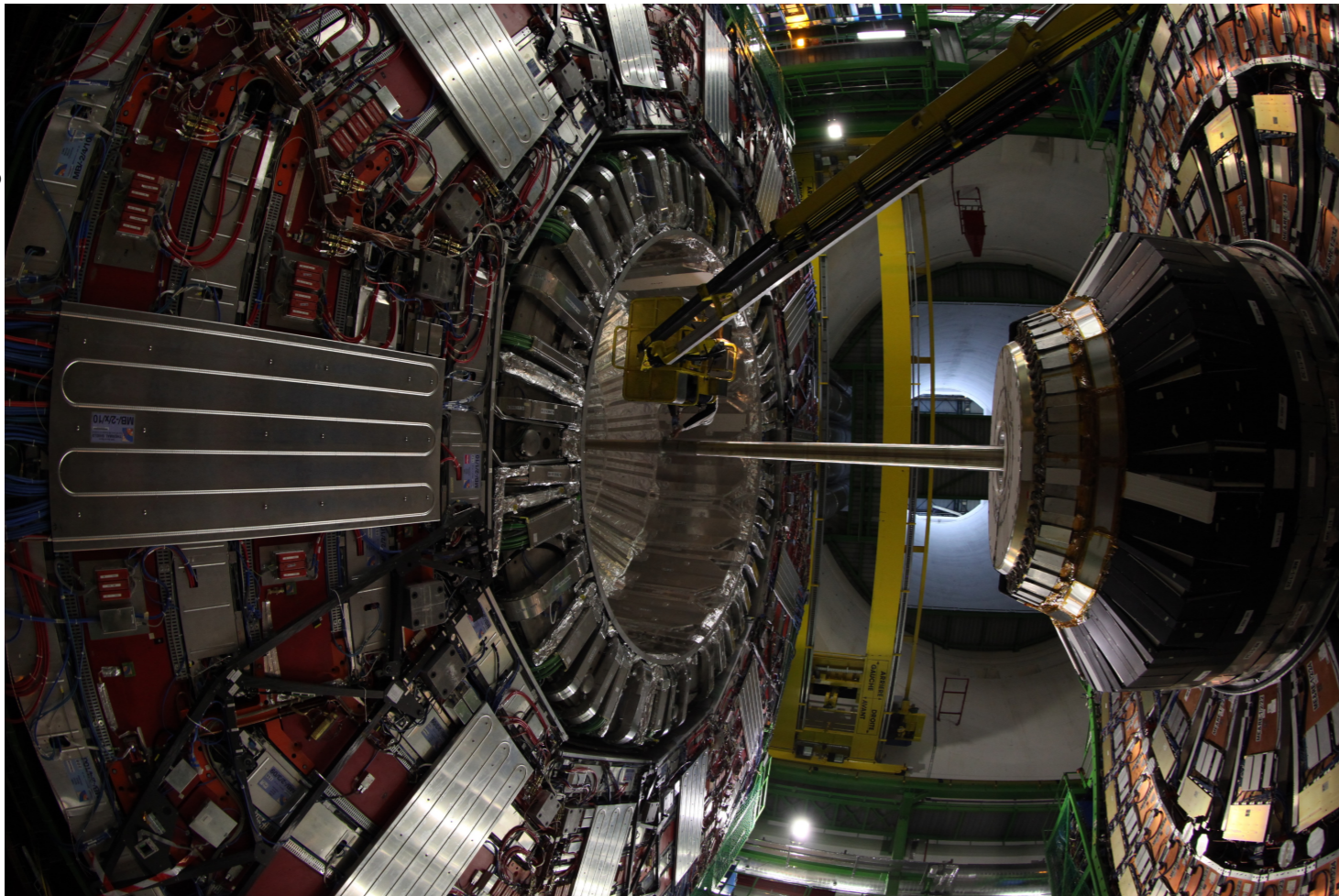
*ICHEP 2016*
*Chicago, Illinois*

1

# Introduction: CMS Data Rates



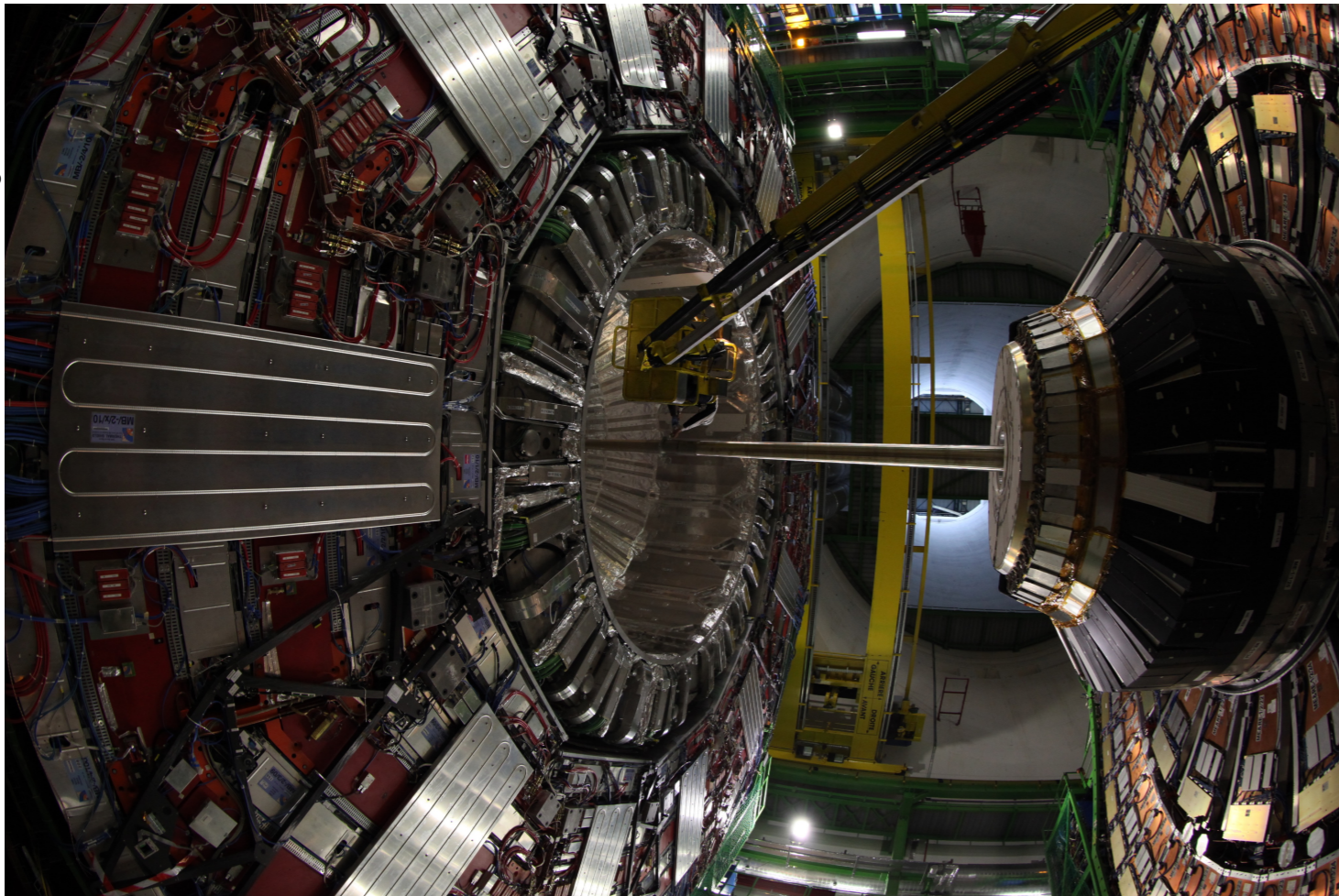Image credit: CERN

The CMS Experiment at CERN

- The LHC provides p-p bunch crossings at a rate of up to **40 MHz**

- The CMS experiment uses a two-level trigger system to reduce the data volume

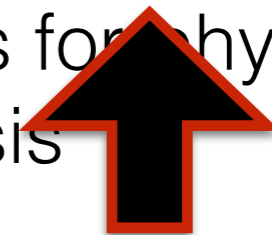  ➡ Selects ~**1 kHz** of events for physics analysis

2

# Introduction: CMS Data Rates



Image credit: CERN

The CMS Experiment at CERN

- The LHC provides p-p bunch crossings at a rate of up to **40 MHz**

- The CMS experiment uses a two-level trigger system to reduce the data volume

  ➡ Selects ~**1 kHz** of events for physics analysis

**This talk is about how (and why) to increase this number**

# Outline of Talk

*A New Trigger Paradigm*

*History of Scouting in CMS*

*Scouting in LHC Run II*

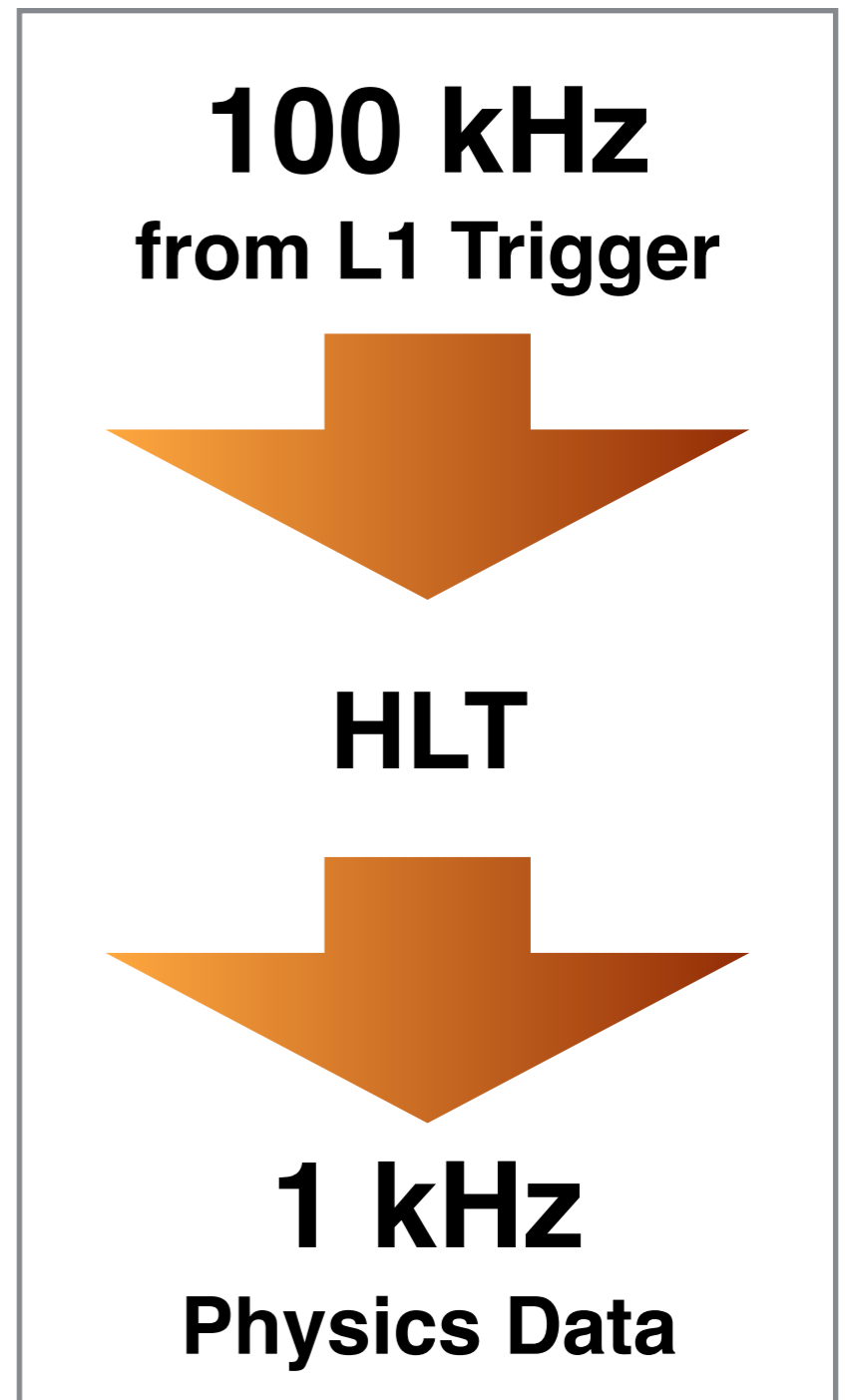*Scouting Search Results in LHC Run II*
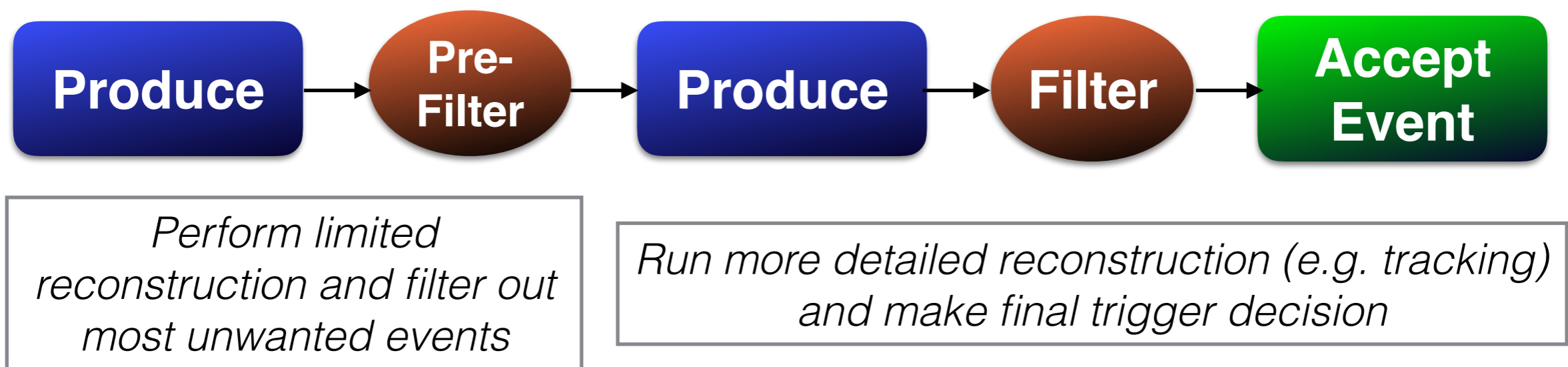
# *A New Trigger Paradigm*

# The CMS High-Level Trigger

- **Dedicated CPU farm** processes the output of the Level-1 trigger in real time

- The High-Level Trigger (HLT) has more than **500 trigger paths**, each selecting for a particular physics signature

- CPU time available per event: up to **200 ms**

**100 kHz**
**from L1 Trigger**

**HLT**

**1 kHz**
**Physics Data**

# Trigger Object Reconstruction

- The HLT reconstructs some physics objects for each event in order to make accept/reject decision

- A trigger path usually works like this:

**Produce** → **Pre-Filter** → **Produce** → **Filter** → **Accept Event**

*Perform limited reconstruction and filter out most unwanted events*

*Run more detailed reconstruction (e.g. tracking) and make final trigger decision*

- **Object performance is as similar as possible to offline, to maximize trigger efficiency**

7

# Restrictions on Trigger Rates

- **Offline reconstruction of events** must be performed in under 48 hours to avoid delays and backlogs

- **Total DAQ data volume** and disk/tape space are limited

- **HLT CPU resources** limit the physics objects that can be reconstructed online

# A New Paradigm: Data Scouting

- **Data Scouting** is a technique for saving events that CMS cannot normally record due to trigger restrictions

- How it works:

  · **Produce physics objects online at the HLT** (as for a normal trigger path)

  · **Save the HLT physics objects to disk**

  · **Perform your physics analysis on those objects** (rather than on offline reconstructed objects)

- No raw data is saved and no further reconstruction is performed on these events

**100 kHz**
**from L1 Trigger**

**HLT**

**1 kHz**
**Physics Data**

**1-5 kHz**
**Scouting Data**

9
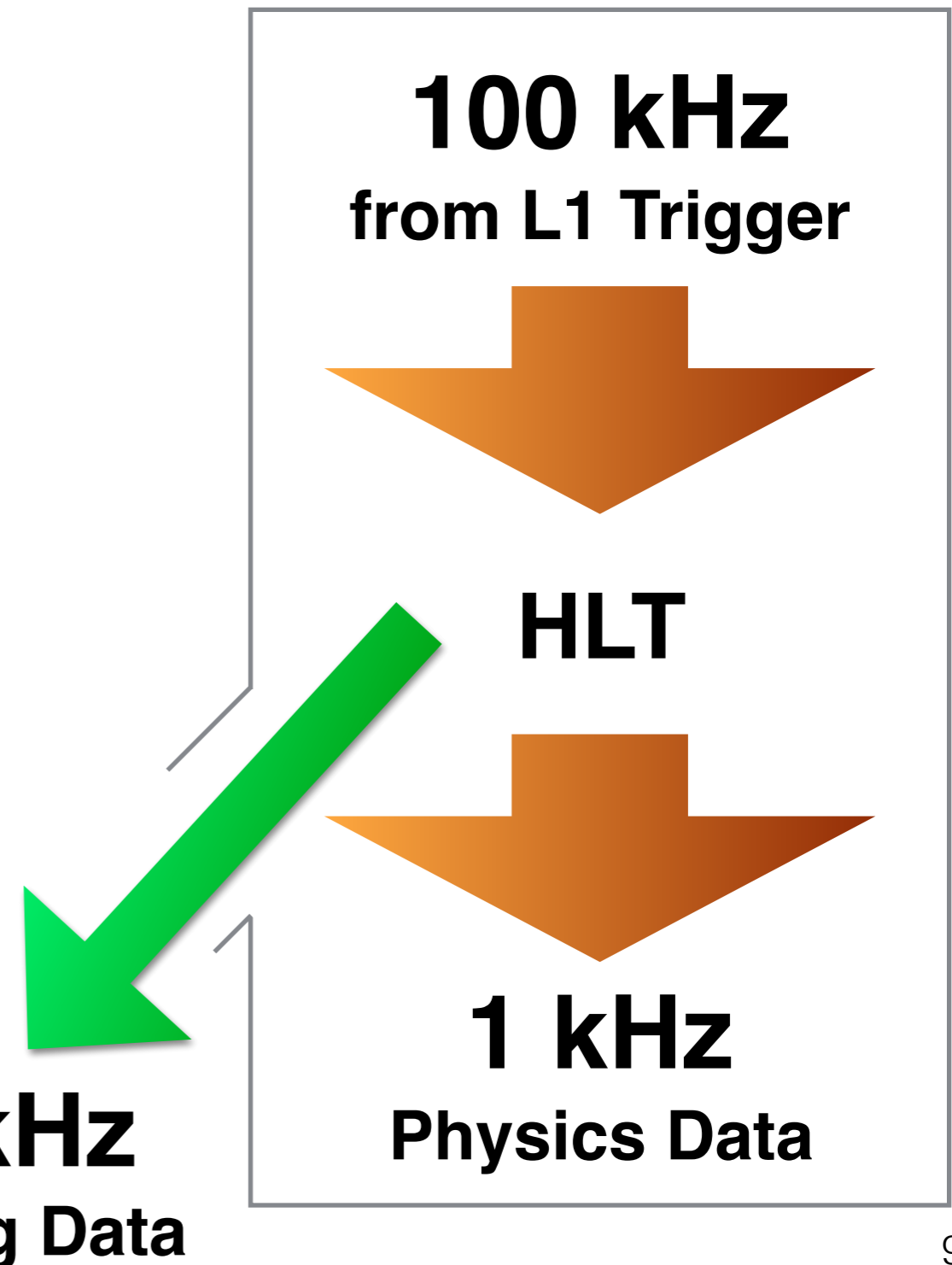
# Scouting Enables Very High Trigger Rates

- **No offline reconstruction** needed!

- Event size on disk is **100-1000 times smaller** than standard raw data

- Trigger rates of **several kHz** are possible

**Provides a major increase in the amount of physics data available for analysis!**

(a factor of 2-6, with rates limited by L1 trigger selection and HLT CPU resources)

## 100 kHz
**from L1 Trigger**

## HLT

## 1 kHz
**Physics Data**
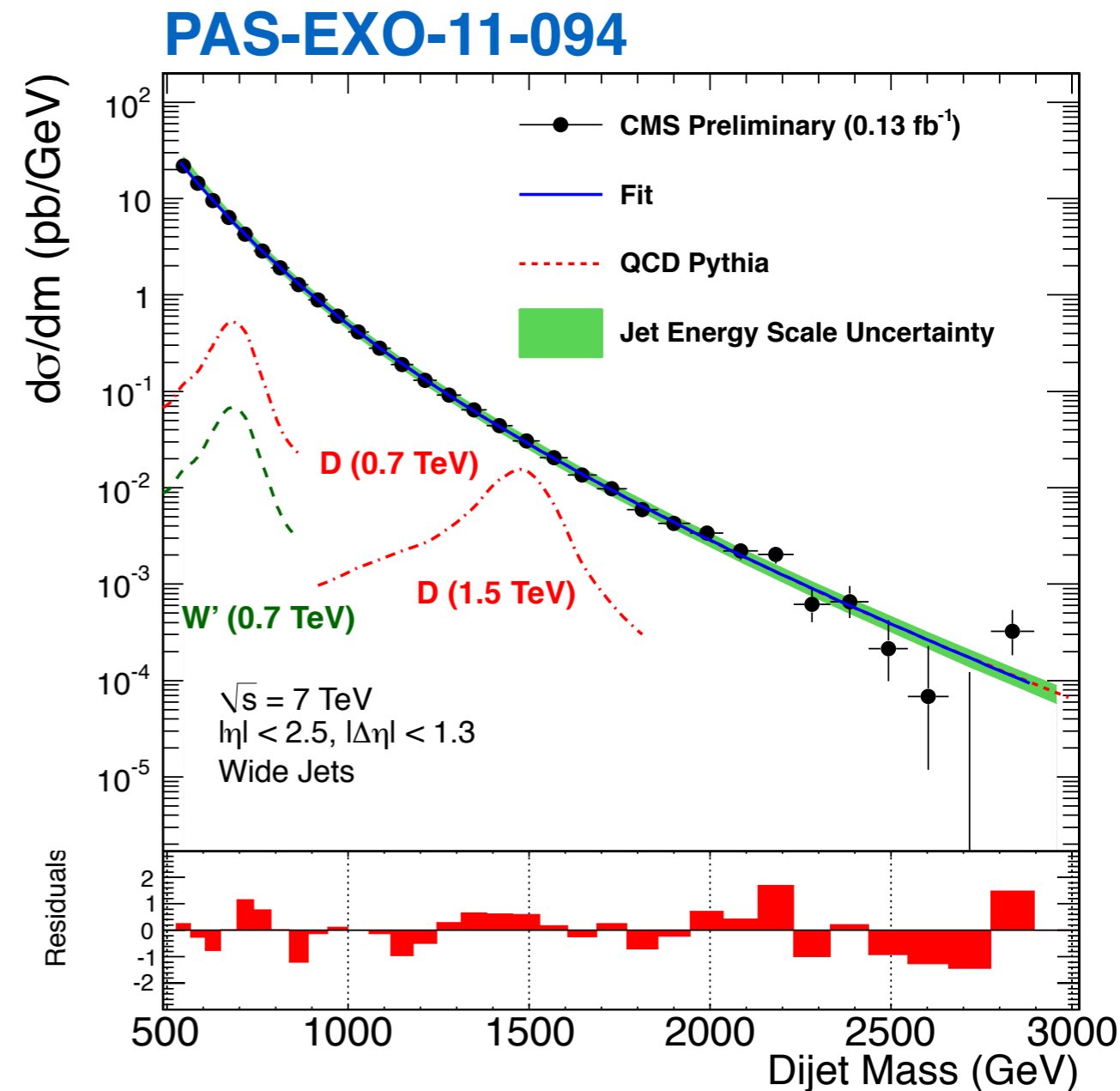
## 1-5 kHz
**Scouting Data**
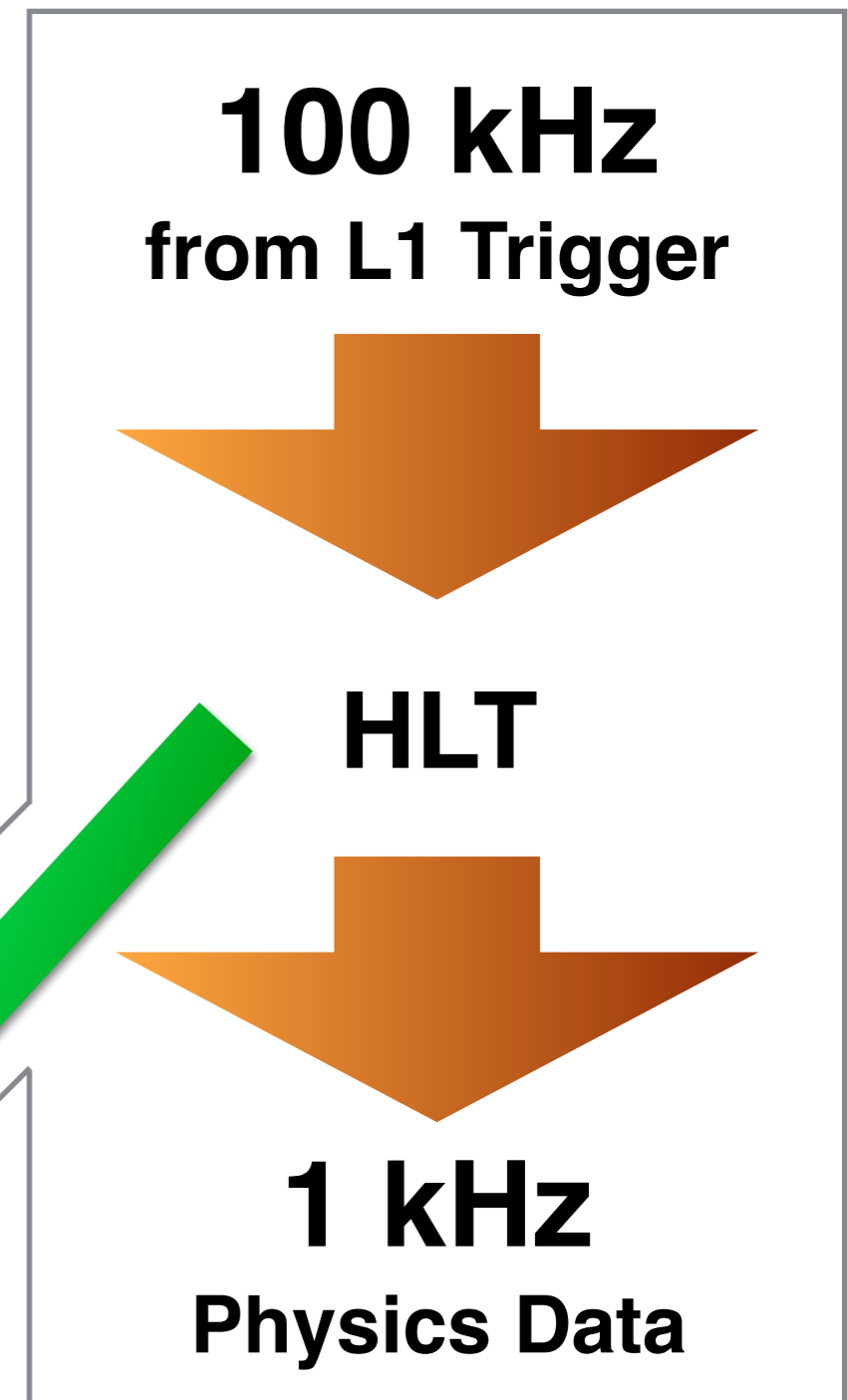
# *History of Scouting in CMS*
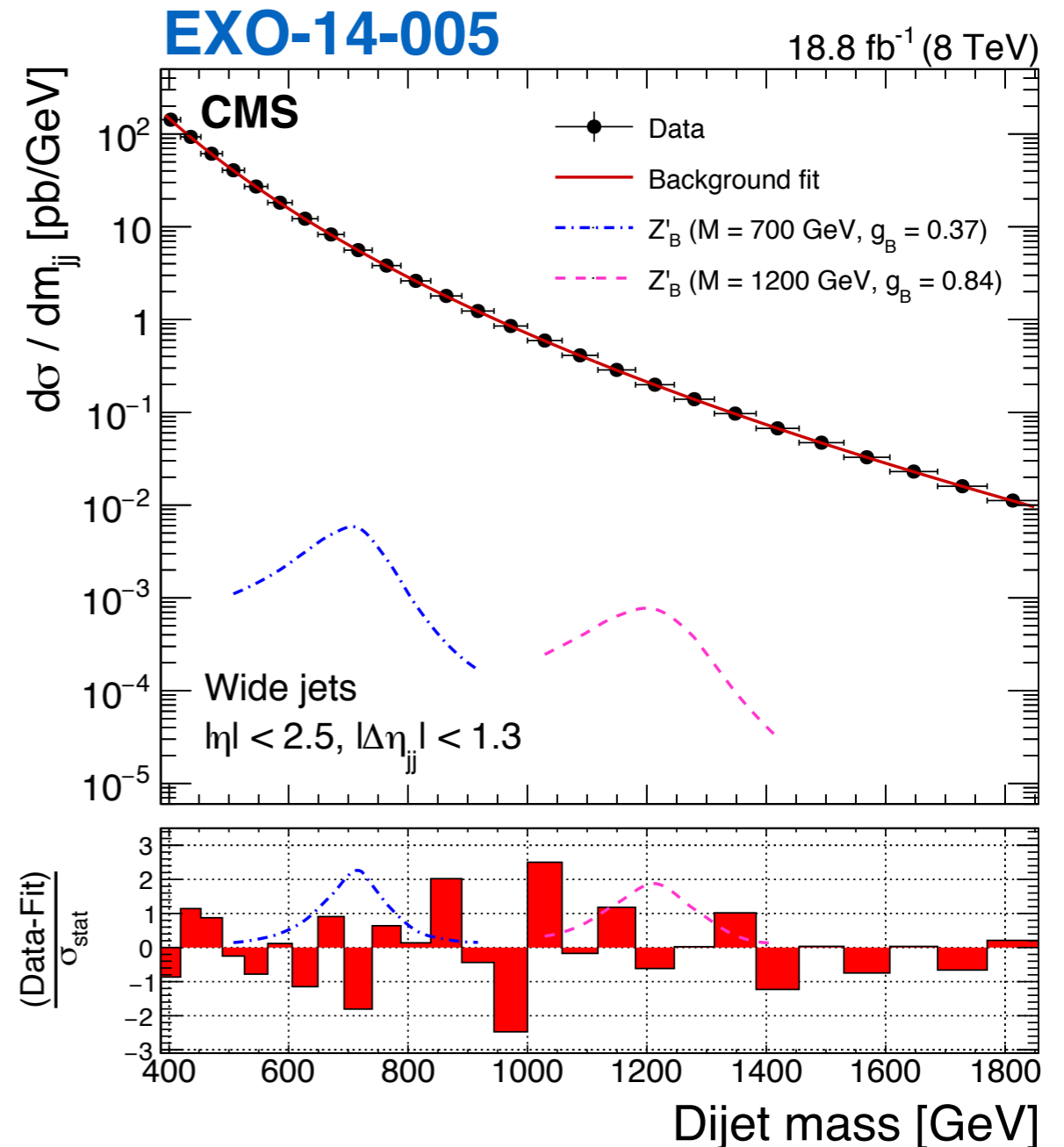
# Scouting in 2011

- **The first scouting trigger** was deployed during the last few pp fills of 2011 data taking

- **0.13/fb** of data were collected

- **The scouting dataset** consisted of all events with $H_T$ > 350 GeV*

  - Jets reconstructed from particle flow candidates were saved for each selected event

- **A search for dijet resonances** was performed, demonstrating sensitivity in the range 0.6-0.9 TeV

**PAS-EXO-11-094**

*Figure legend:*
- CMS Preliminary (0.13 fb$^{-1}$)
- Fit
- QCD Pythia
- Jet Energy Scale Uncertainty

D (0.7 TeV)
D (1.5 TeV)
W' (0.7 TeV)

$\sqrt{s}$ = 7 TeV
$|\eta|$ < 2.5, $|\Delta\eta|$ < 1.3
Wide Jets

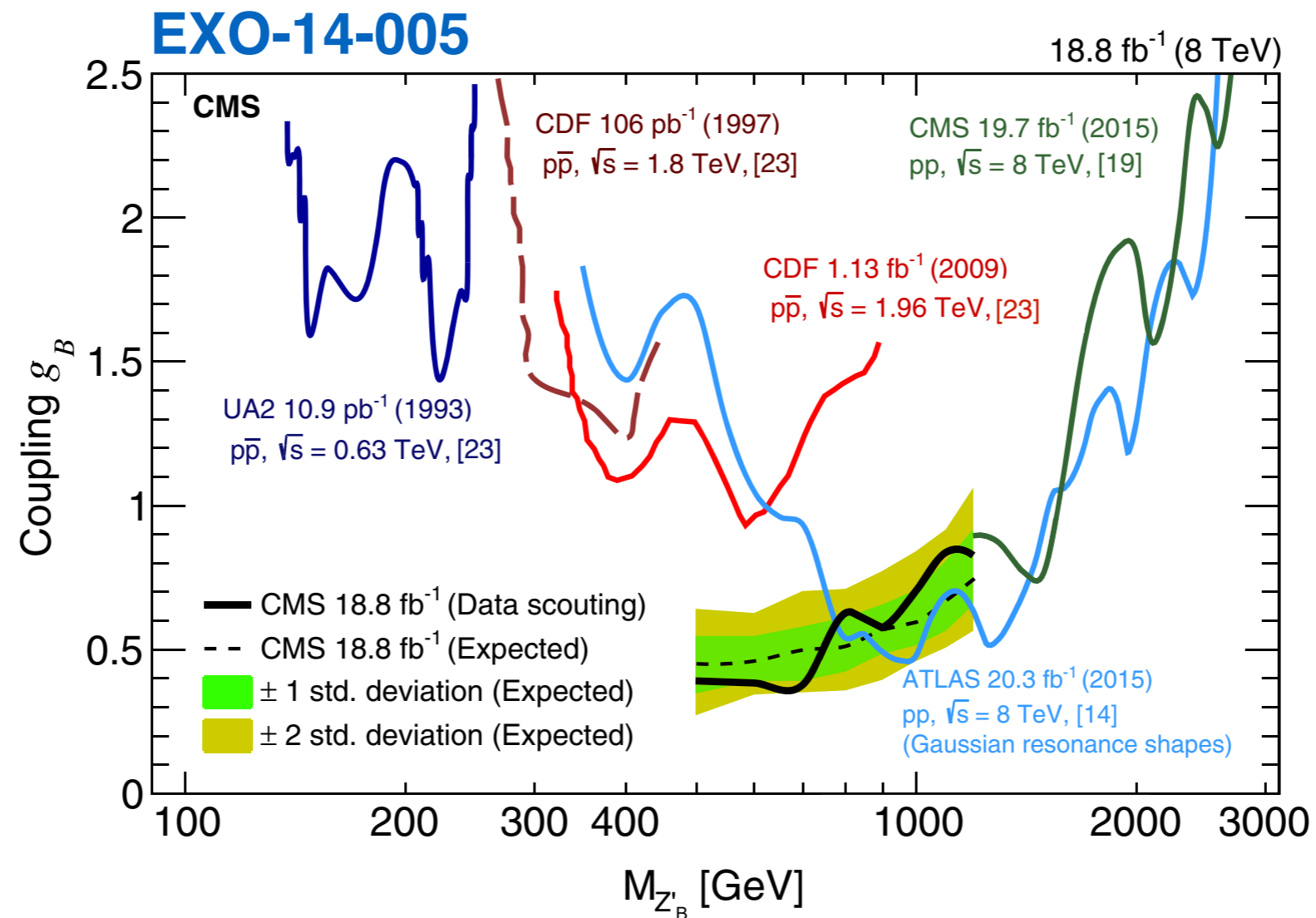Axes: d$\sigma$/dm (pb/GeV) vs Dijet Mass (GeV); Residuals

$*H_T = \Sigma \, | \, \text{jet } p_T \, |$

12

# Scouting in 2012

- **The strategy was repeated** in 2012, this time for the entire data taking period

- **Events with $H_T$ > 250 GeV** were recorded

  - Calorimeter-level jets were saved for each event

- **A dijet search was performed** on 18.8/fb of data

- **Limits were placed** on resonances between 0.5 and 1.6 TeV



13

# Scouting in 2012



- Limits were placed on the coupling of a hypothetical leptophobic resonance decaying to quarks

- **Provides the most stringent limits to date between 500 and 800 GeV**
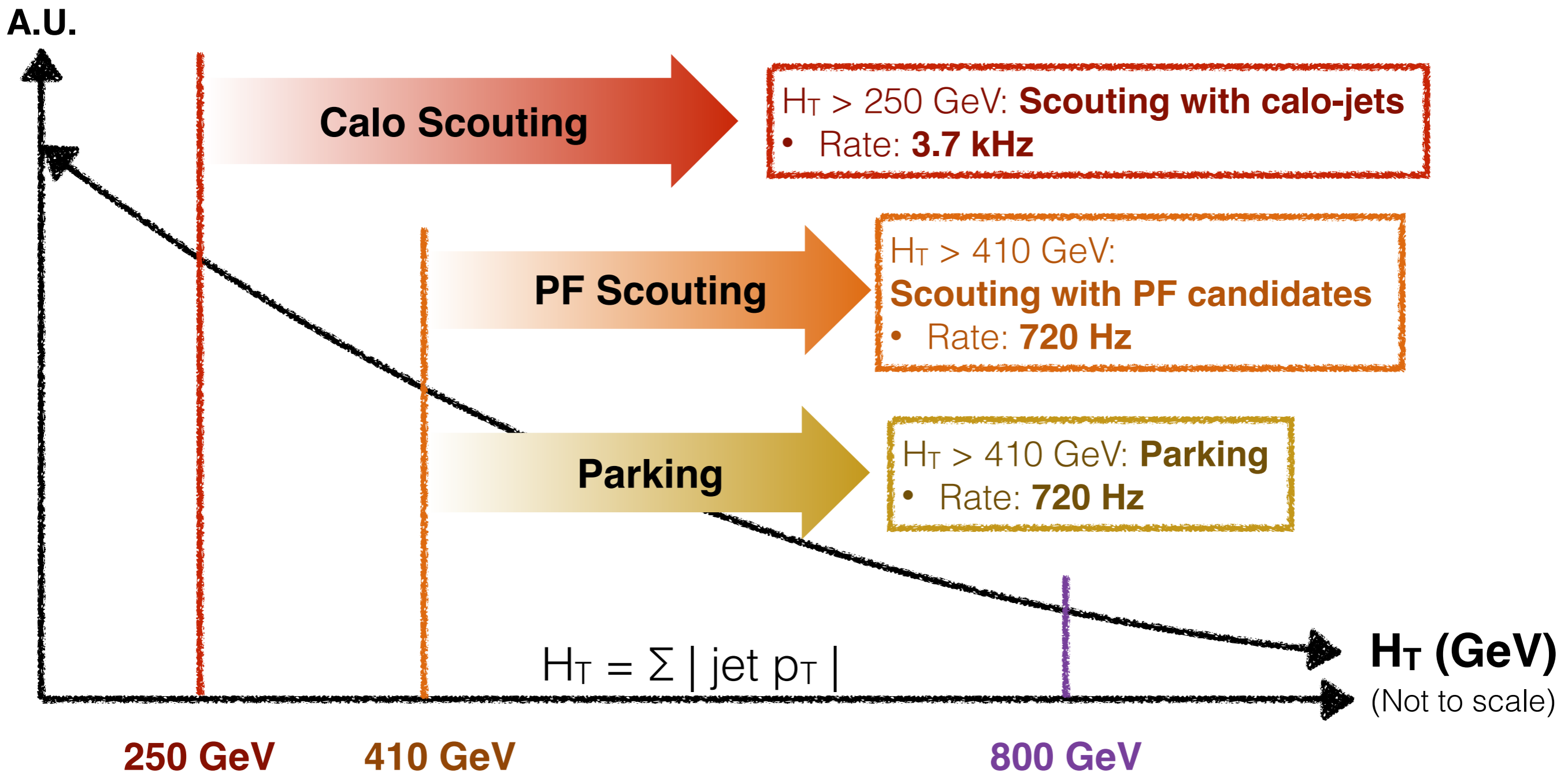
14

# *Scouting in LHC Run II*

# Scouting in LHC Run II

- Following the success of Run I scouting, the strategy was expanded for Run II:

  - "**Calo Scouting**": reconstruct and save jets reconstructed from calorimeter deposits (as in 2012)

    ➡ B-tagging and basic pileup jet ID information obtained from local pixel track reconstruction (added in 2016)

  - "**PF Scouting**": reconstruct and save particle-flow (PF) candidates

    ➡ Leptons, photons, and jets with b-tag information saved

  - **Parking** of the full RAW data in case of a discovery in the scouting data

- ATLAS has also adopted the scouting technique for its Run II dijet search (see ATLAS-CONF-2016-030)

16

# Scouting With Hadronic Events

**A.U.**

**Calo Scouting**

$H_T > 250$ GeV: **Scouting with calo-jets**
- Rate: **3.7 kHz**

**PF Scouting**

$H_T > 410$ GeV:
**Scouting with PF candidates**
- Rate: **720 Hz**

**Parking**

$H_T > 410$ GeV: **Parking**
- Rate: **720 Hz**

$H_T = \Sigma \mid jet\ p_T \mid$

**$H_T$ (GeV)**
(Not to scale)

**250 GeV**        **410 GeV**        **800 GeV**

**Rates from 2016 data.**
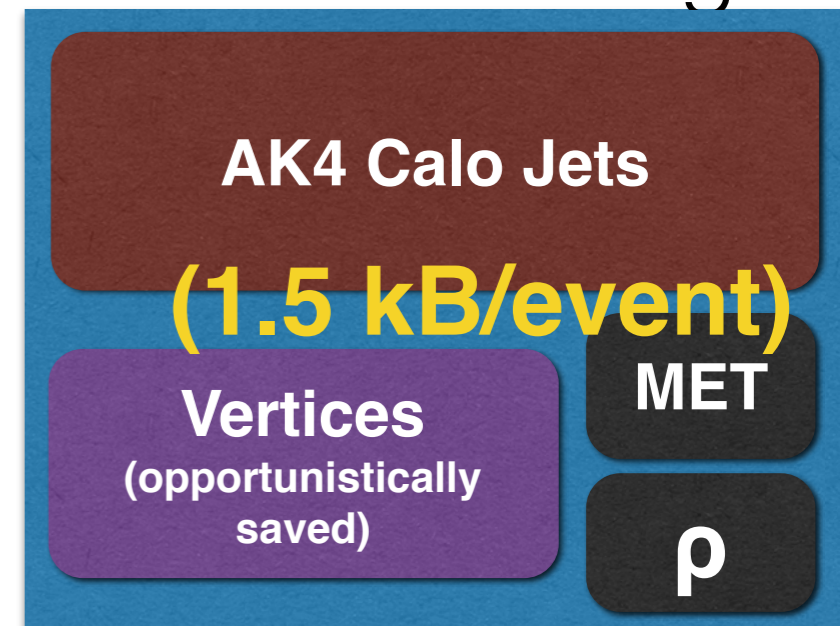**Luminosity = $10^{34}$ cm$^{-2}$ s$^{-1}$**        17        **Loosest $H_T$ trigger in standard HLT menu**

# Scouting Data

**Calo Scouting**

- Dedicated compact event formats are used for Calo Scouting and PF Scouting

- In 2016 Calo and PF Scouting each took **1%** of the disk resources used by CMS physics data

**AK4 Calo Jets**

**(1.5 kB/event)**

**Vertices (opportunistically saved)**

**MET**

**ρ**

**PF Scouting**

**PF Candidates**

**AK4 PF Jets**

**(10 kB/event)**

**Vertices**

**MET**

**ρ**

**e**

**μ**

**γ**

PF Scouting Event

Calo Scouting Event

Ordinary CMS Event (~1MB)

ρ: average energy density in event
AK4 jets: clustered using Anti-$k_T$ with R = 0.4

18

# Scouting Dataset Overview

- Other scouting data:

  - **Dimuon scouting dataset** records 2-muon events with $m_{\mu\mu} > 10$ GeV

  - **Commissioning datasets** contain very loose prescaled scouting triggers, useful for measuring trigger performance

  - **Monitoring dataset** contains events reconstructed both at the HLT and offline, useful for online-offline object comparisons

| Data Stream | Rate at $10^{34}$ cm$^{-2}$ s$^{-1}$ (Hz) | Bandwidth (MB/s) |
|---|---|---|
| Calo Scouting H$_T$ | 3700 | 11 |
| PF Scouting H$_T$ | 720 | 9 |
| PF Scouting Dimuon | 480 | 6 |
| Commissioning | 30 | < 1 |
| Monitor | 26 | 23 |

# Data Parking

- Advantageous to record the full raw data in case a signal is seen in scouting data

- **Parking stream at HLT sends events directly to tape without reconstruction**

- Rate limited by DAQ and disk storage

  ➡ Park only the $H_T > 410$ GeV data from PF Scouting

| Range in $H_T$ (GeV) | Rate at $10^{34}$ cm$^{-2}$ s$^{-1}$ (Hz) |
|---|---|
| 410-430* | 130 |
| 430-450 | 100 |
| 450-470 | 83 |
| 470-500 | 95 |
| 500-550 | 110 |
| 550-650 | 110 |
| 650-∞ | 92 |

$H_T$ ranges can be enabled/disabled individually as needed

\* Due to rate constraints, the 410-430 GeV range will be disabled for the post-ICHEP running period

20
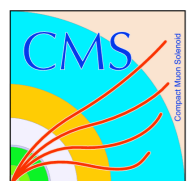
# *Scouting Search Results in LHC Run II*

# Run II Scouting Dijet Search

- **Search for resonances** between 0.6 TeV and 1.6 TeV, using 12.9 fb$^{-1}$ of scouting data collected in 2016

- **Calo Scouting trigger** selecting events with $H_T > 250$ GeV is used

- **Dedicated scouting monitor dataset** is used to make detailed online-offline comparisons for jets

  - The HLT calorimeter jets are then re-calibrated to have the same response as offline PF jets

**PAS-EXO-16-032**

# Results

**PAS-EXO-16-032**
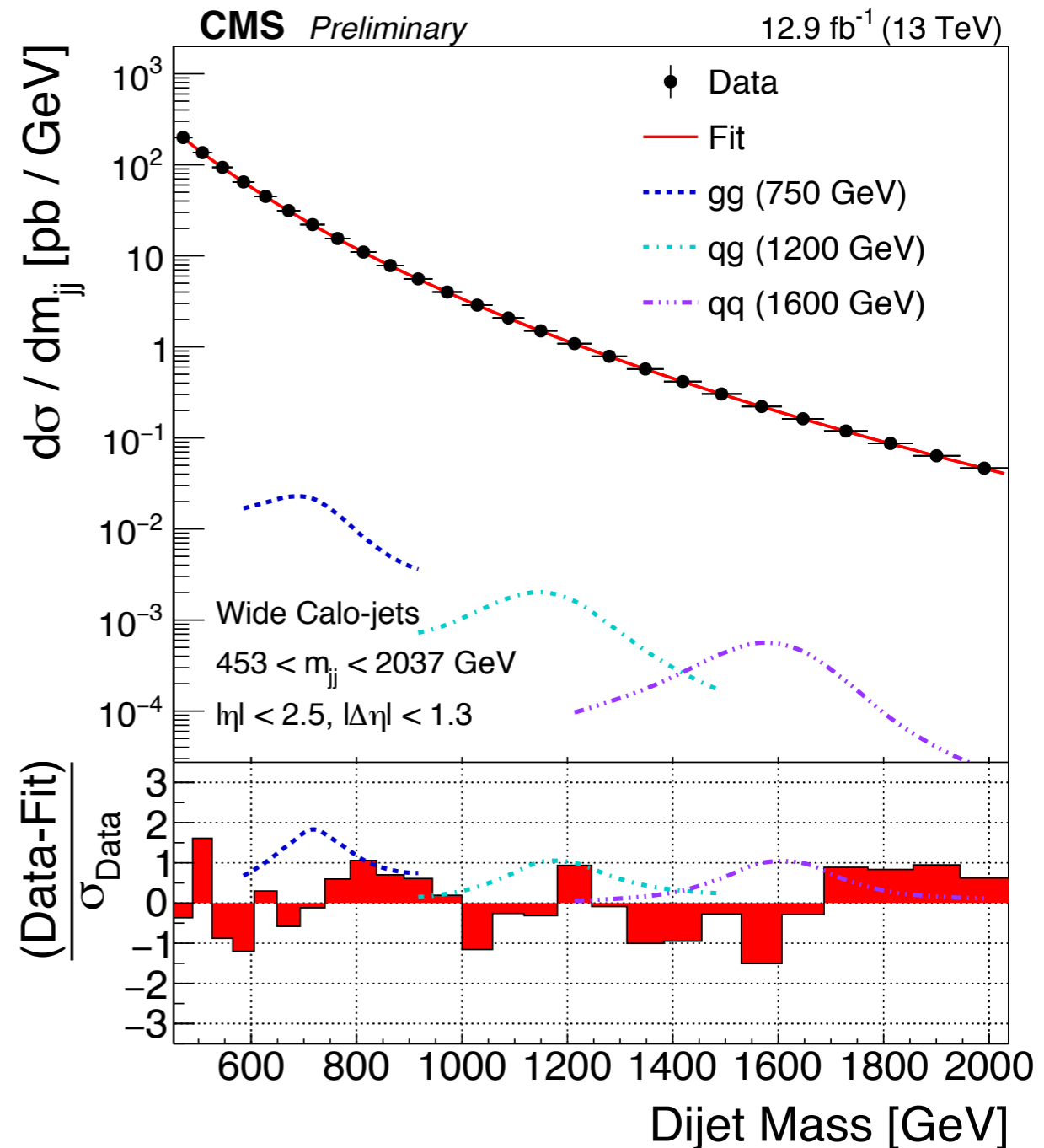


- **Limits are placed** on gg, qg, and qq resonances

- **Largest local significance** is 2.6σ, for a gg resonance of mass 850 GeV
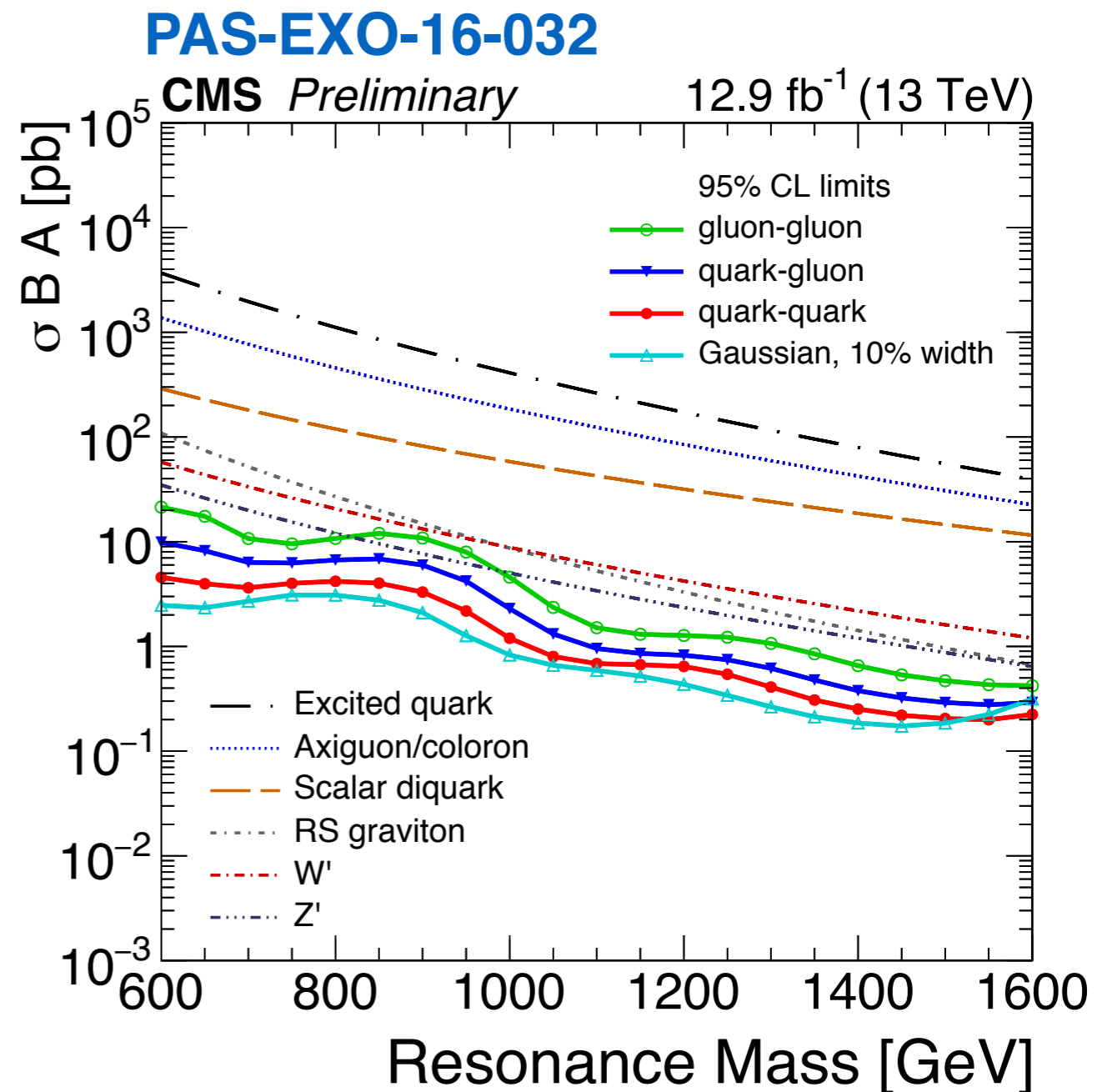
**For more details see yesterday's talk by Saptaparna Bhattacharya**

# Conclusion

- **Scouting significantly increases the number of events recorded by CMS, with negligible impact on data volume**

- **Successful scouting searches in LHC Run I motivate a more comprehensive strategy for Run II:**

  - Scouting with calorimeter jets

  - Scouting with Particle Flow candidates

  - Raw data parked in case of a discovery

- **First physics result with the Run II scouting framework has been released.  More to come!**