Search for light vector resonances decaying to a quark pair produced in association with a jet in proton-proton collisions at $\sqrt{s}$ = 13 TeV

CMS-PAS-EXO-17-001

... or in other words BOOSTED DIJET + ISR search

Outline

- Motivation
- Event Selection
- Background Estimate
- Results

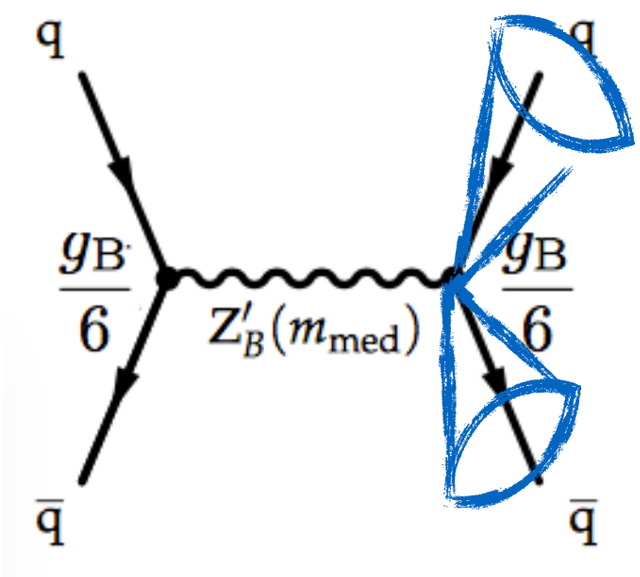
Cristina Mantilla Suarez (Johns Hopkins University)
on behalf of the CMS Collaboration

June 6th, 2017
New Perspectives Conference

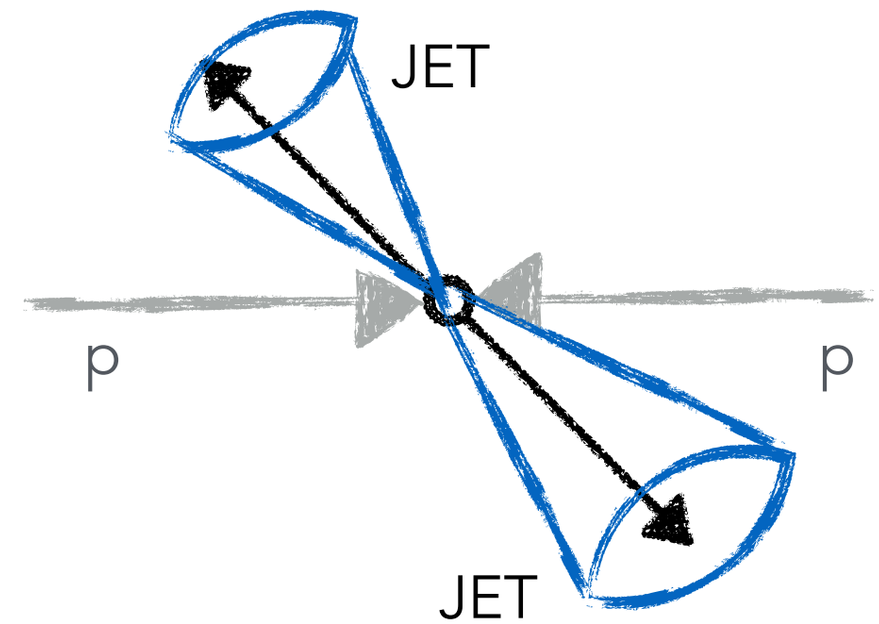
Motivation

CLASSIC Dijet Resonances Searches

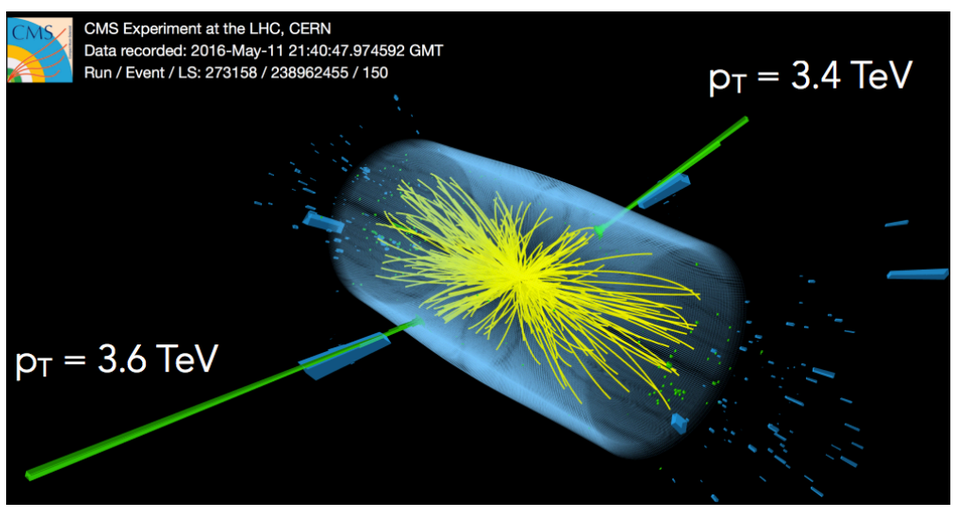
The model: $\mathcal{L} \propto \frac{g_B}{6} Z'_{B\mu} \bar{q} \gamma^\mu q$



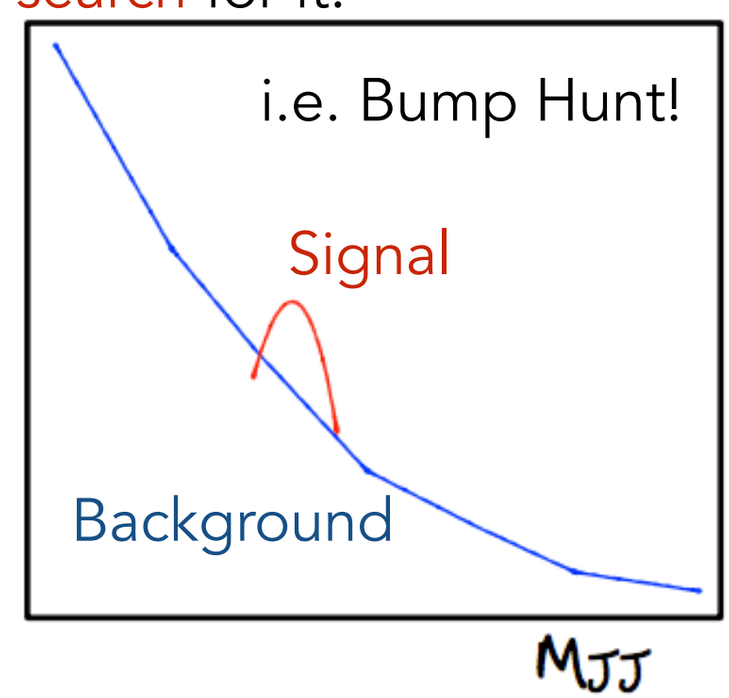
What happens in the collision:



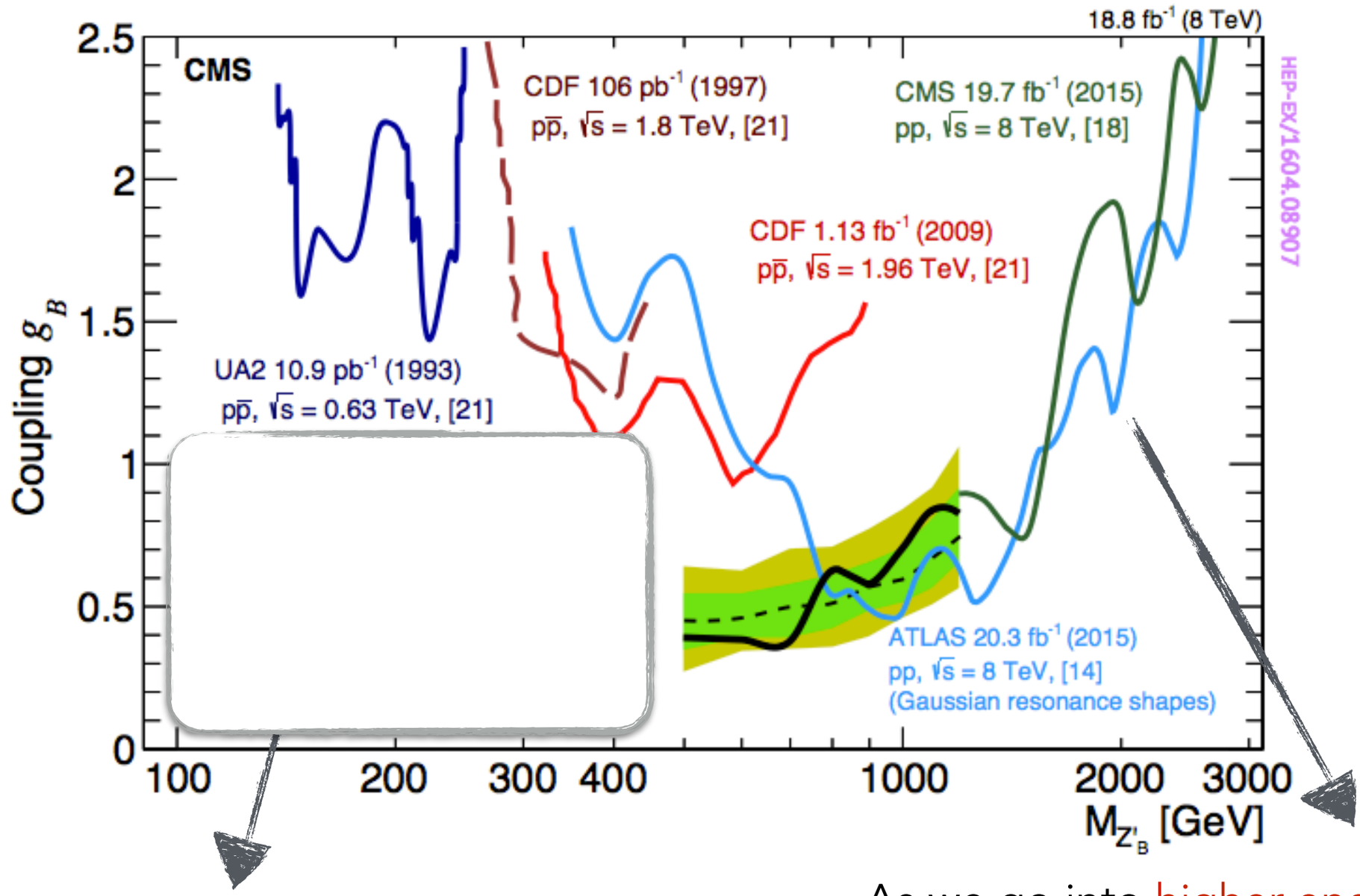
What we see in the detector:



How we search for it:



CLASSIC Dijet Resonances Searches



How can we go lower in mass?

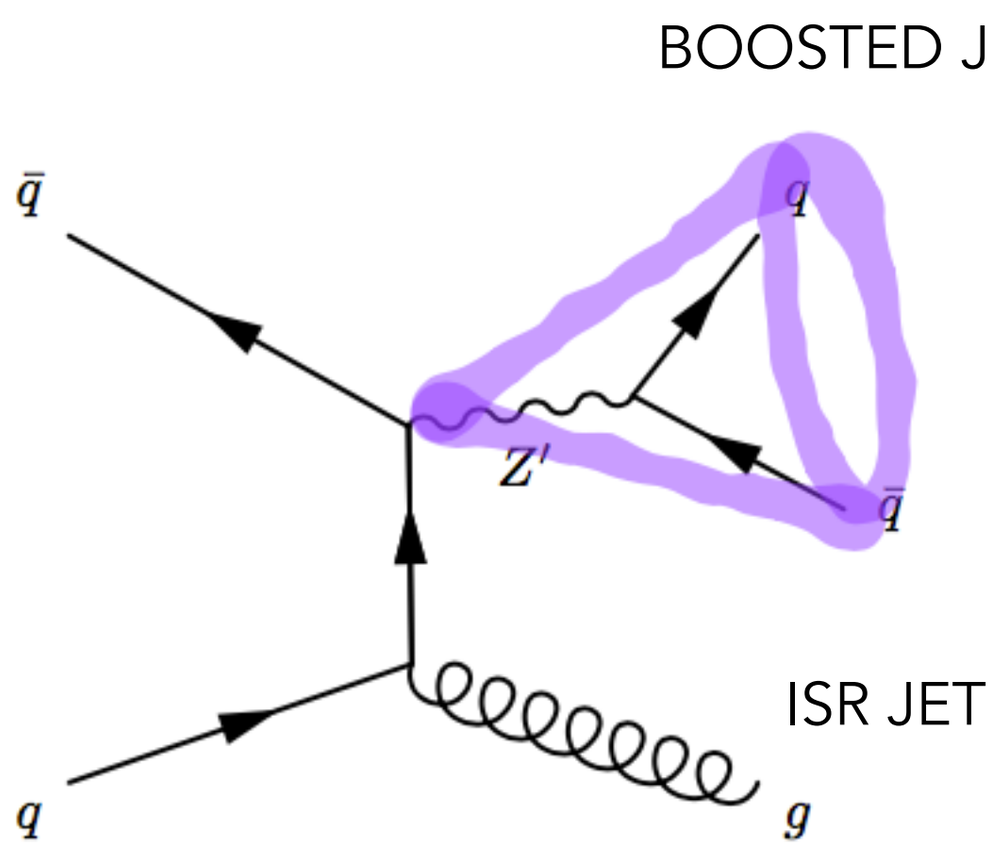
As we go into higher energies
(and more 13 TeV data)
we try to push to higher masses

BOOSTED Dijet Resonances Program

At low masses:

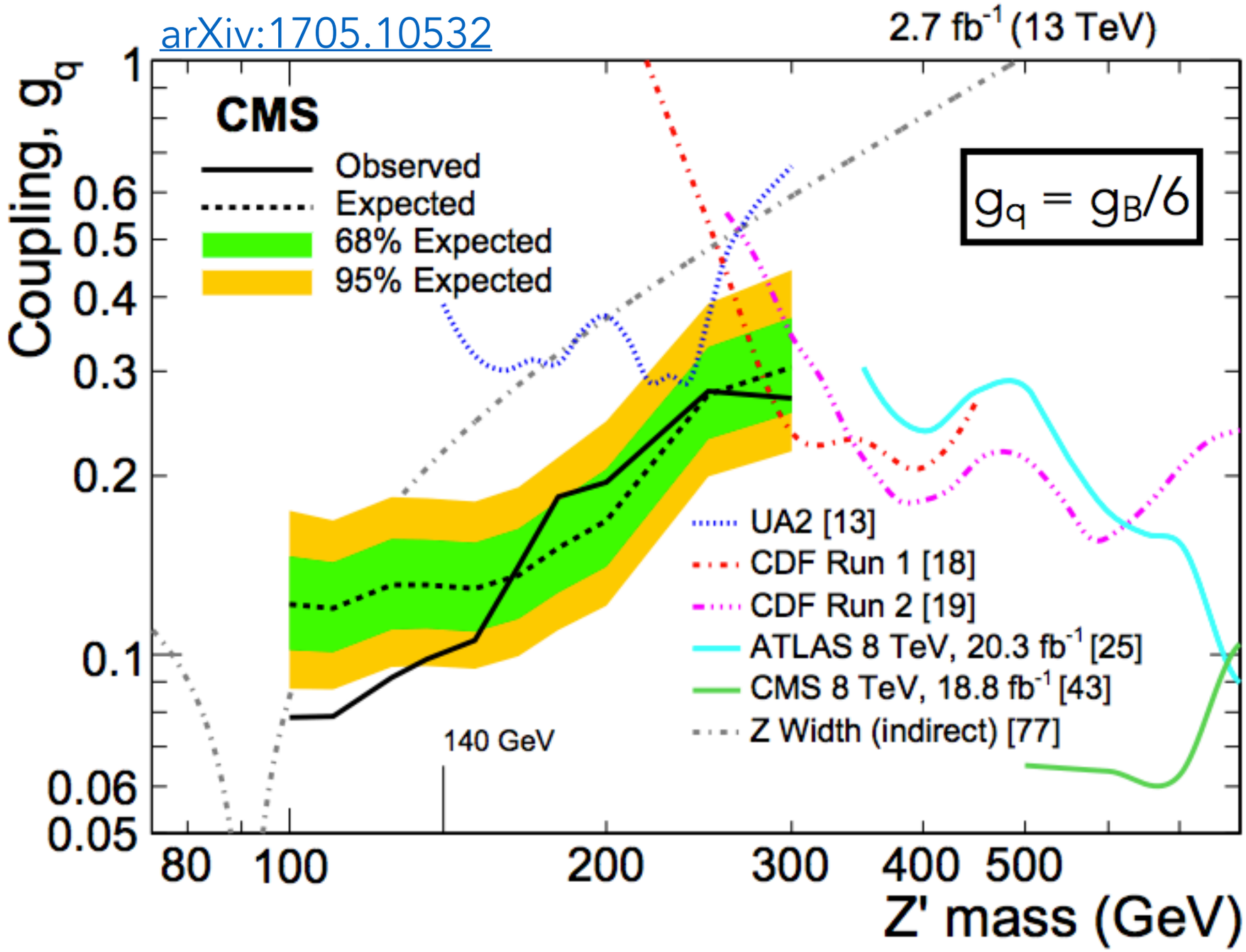
- large cross section of QCD multi jet processes
  - this is WAY TOO MUCH DATA to store
  - we don't store it
  - set trigger thresholds

Solution: Use ISR jet



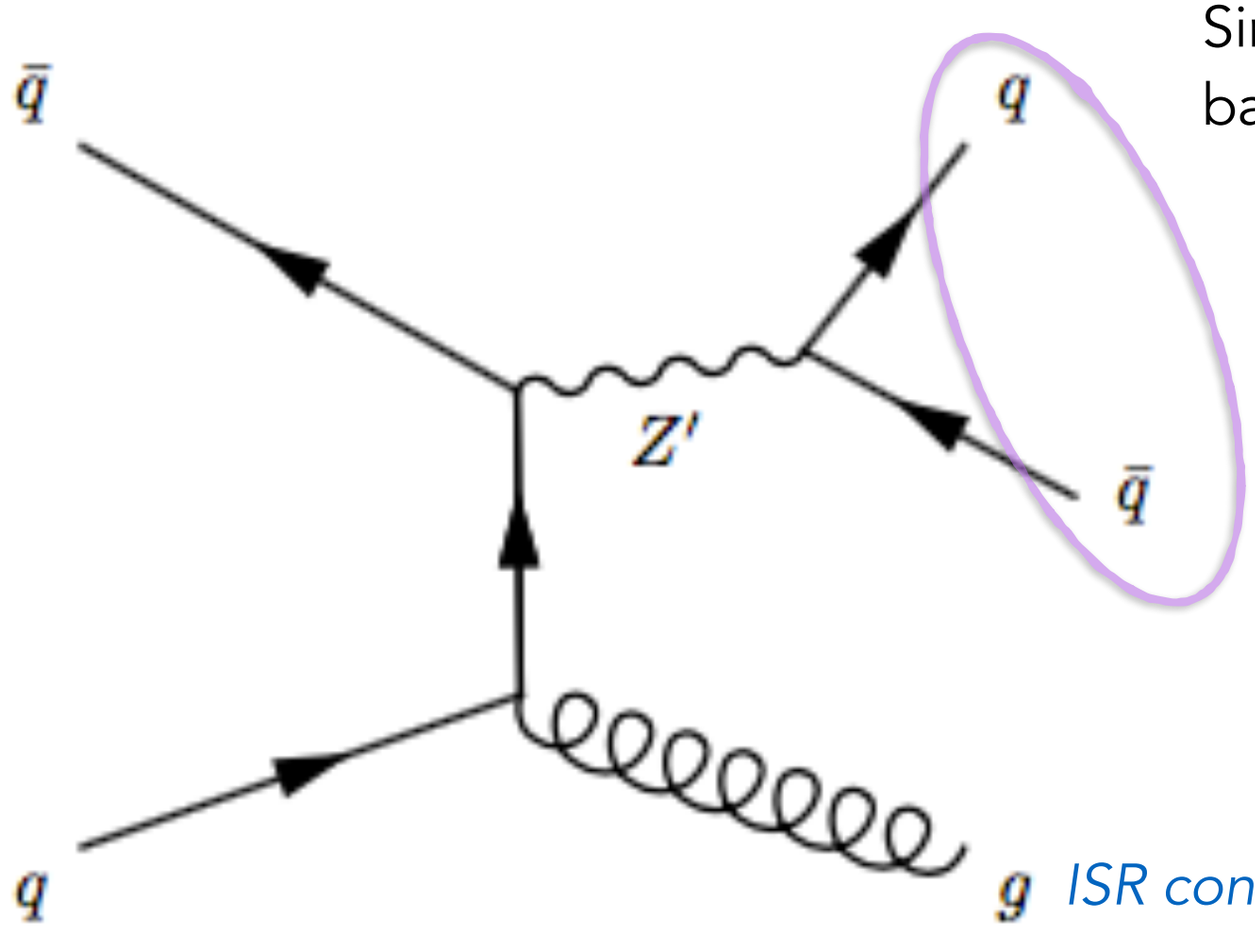
BOOSTED Dijet Resonances Program

CMS: FIRST constraint from the LHC below 250 GeV - employed 2015 data (EXO-16-030)



The present search uses 2016 dataset and improved techniques (see next sls.)

Analysis Strategy



Single *boosted jet* to reduce background combinatorics

SIGNAL Z': Identified by *substructure* variable: N_2^{Deco}

ISR constraint provides enough energy to event to get above trigger

Backgrounds :

- QCD (overwhelming and difficult to simulate) ← *data-driven estimate*
- SM W/Z → qq,
- top

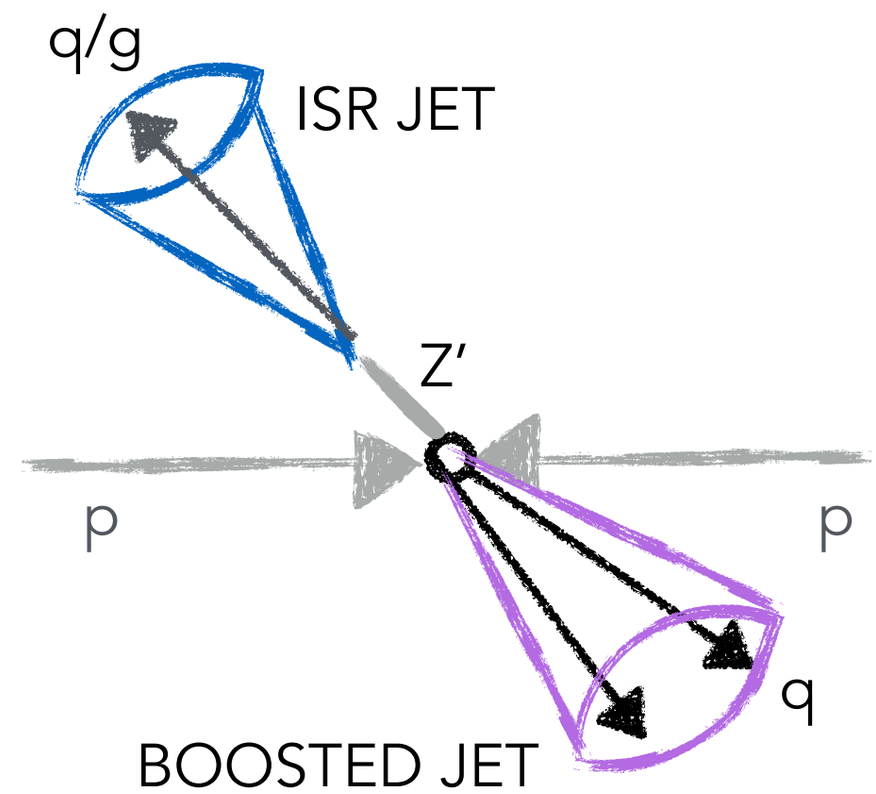
Event Selection

Event Selection

We aim to identify a boosted 2-prong Z' signal:

- Veto on leptons
- Jet kinematic selection:
  - Jet cone: $\Delta R < 0.8$
  - At least one HIGH ENERGY: **p_T > 500 GeV**
- Jet tagging
  - Selection on jet mass and p_T (jet rho)
  - Use jet substructure variable $N_2^{DDT} < 0$ decorrelated from jet mass and p_T

This is KEY



Jet tagging observables

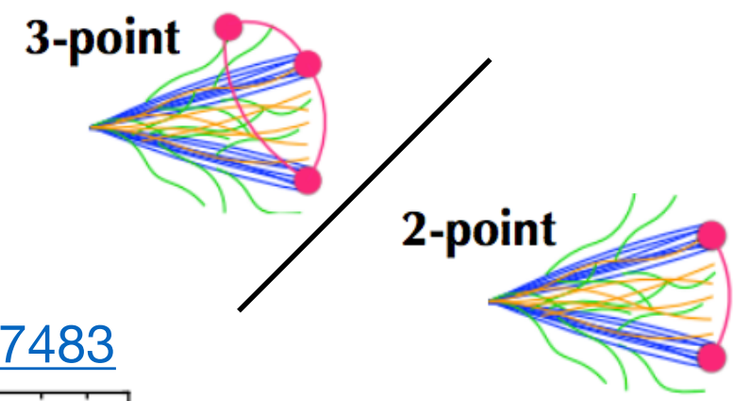
We use **JET SUBSTRUCTURE** variables, e.g.:

1-prong

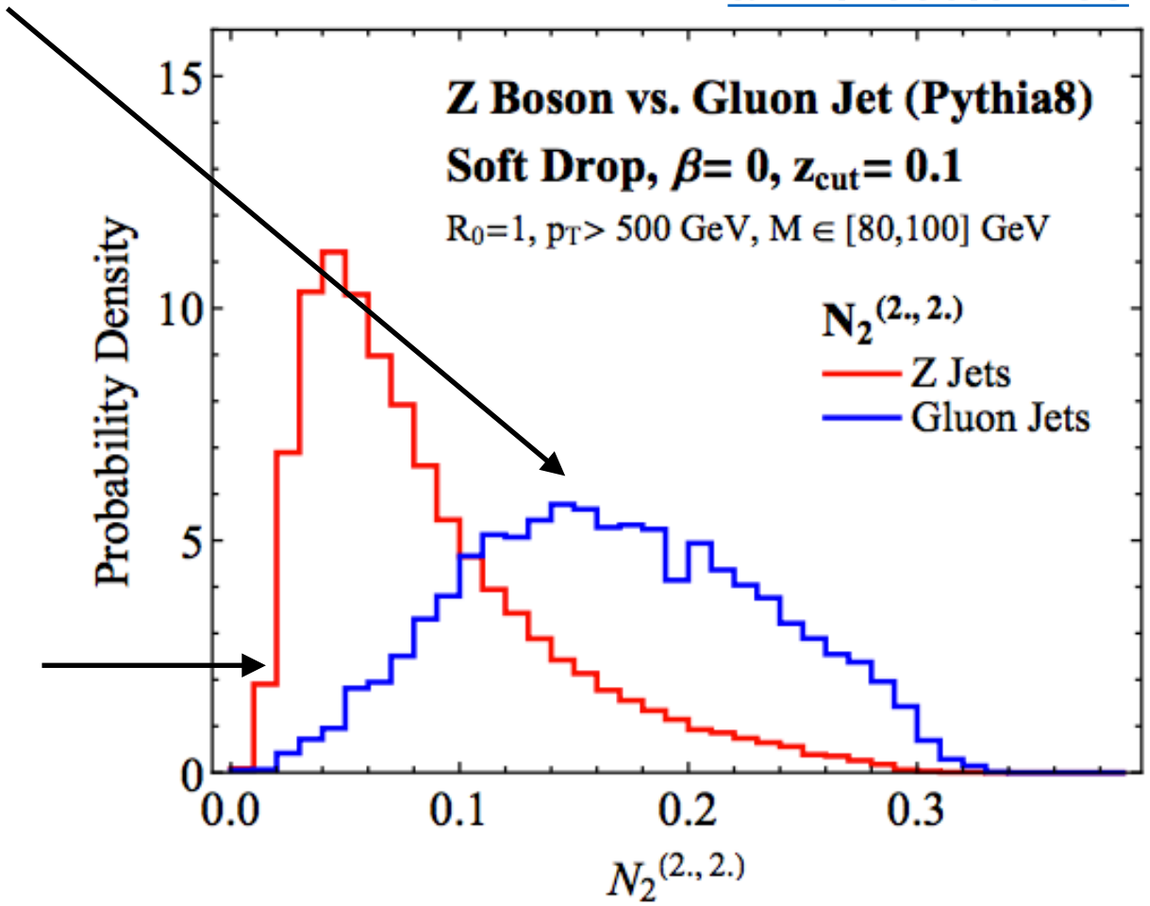
2-prong

Energy Correlation Functions: correlate particles inside jet

$$N_2(\beta) = \frac{{}_2e_3^{\beta}}{(e_2^{\beta})^2}$$



arxiv.1609.07483

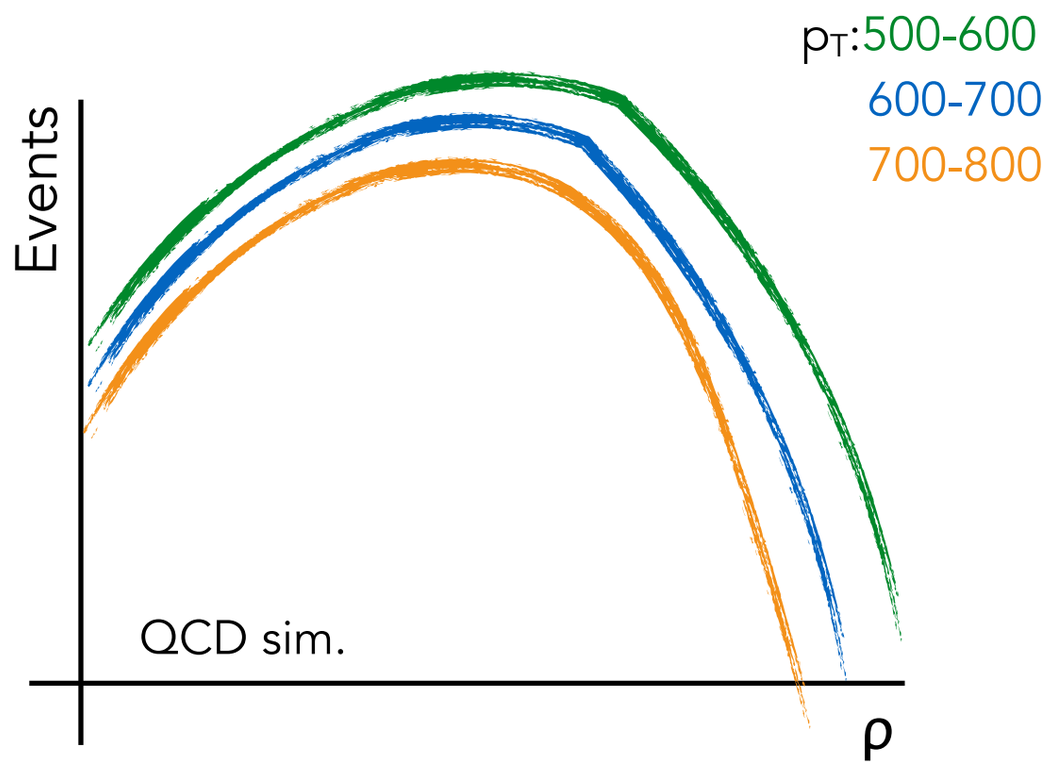


Tagging observables

We use the QCD scaling variable : **JET RHO** $\rho = \ln(m_{SD}^2/p_T^2)$

i.e. for every pT bin the distribution is the same

Is a function of the jet mass m_{SD} and jet pT!



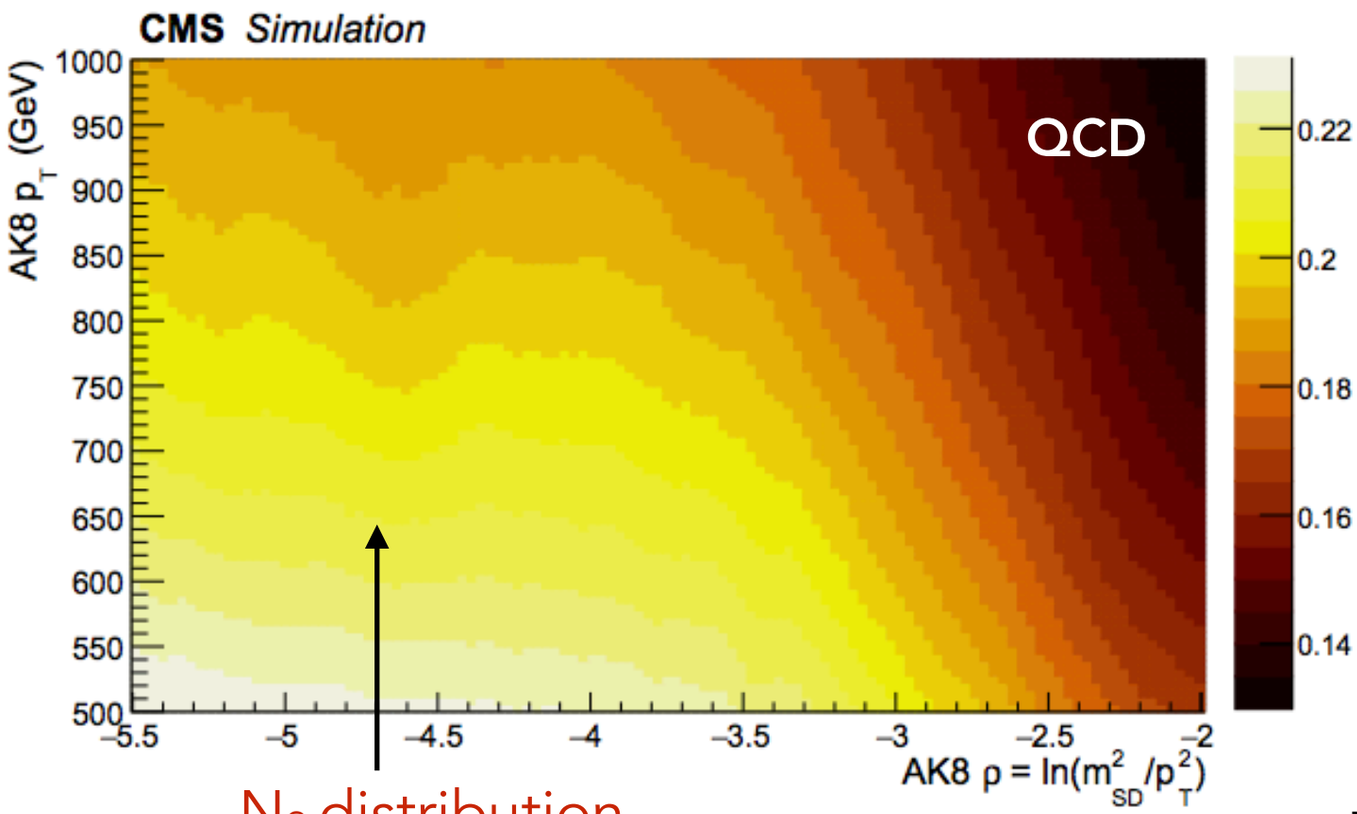
We select: $-5.5 < \rho < -2$

e.g. for a jet of pT: 500 GeV, we select on $30 < m_{SD} < 184$ GeV

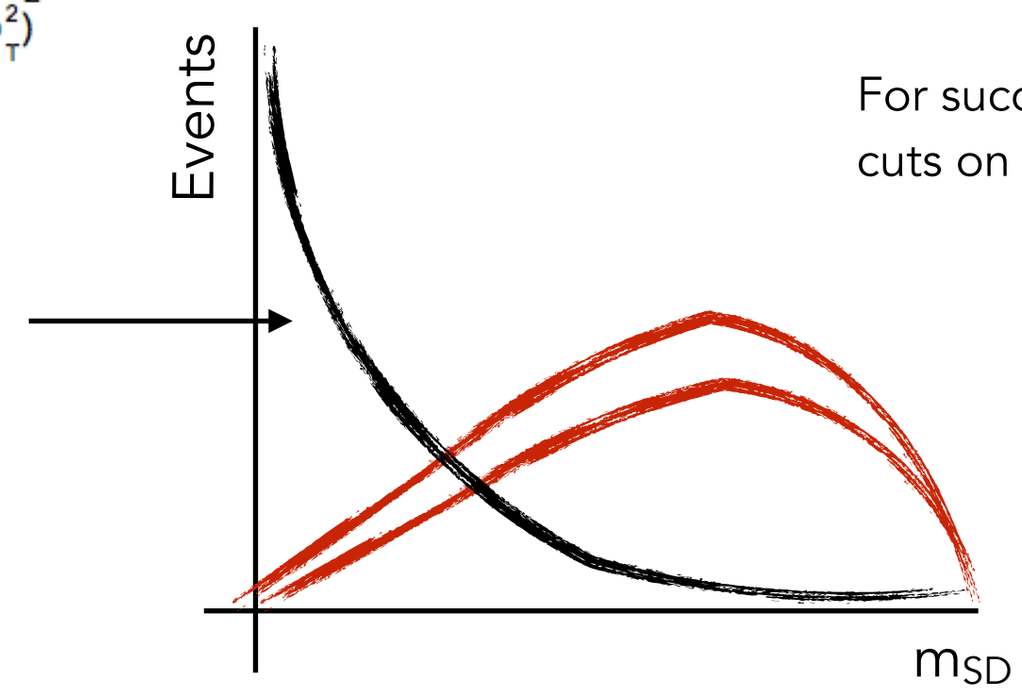
And use it to study the **dependence** of our **selection variables** with the **jet mass and pt**

Jet tagging observables

- Challenge: for QCD events the **N_2 variable is correlated with jet mass and pT:**



N_2 distribution



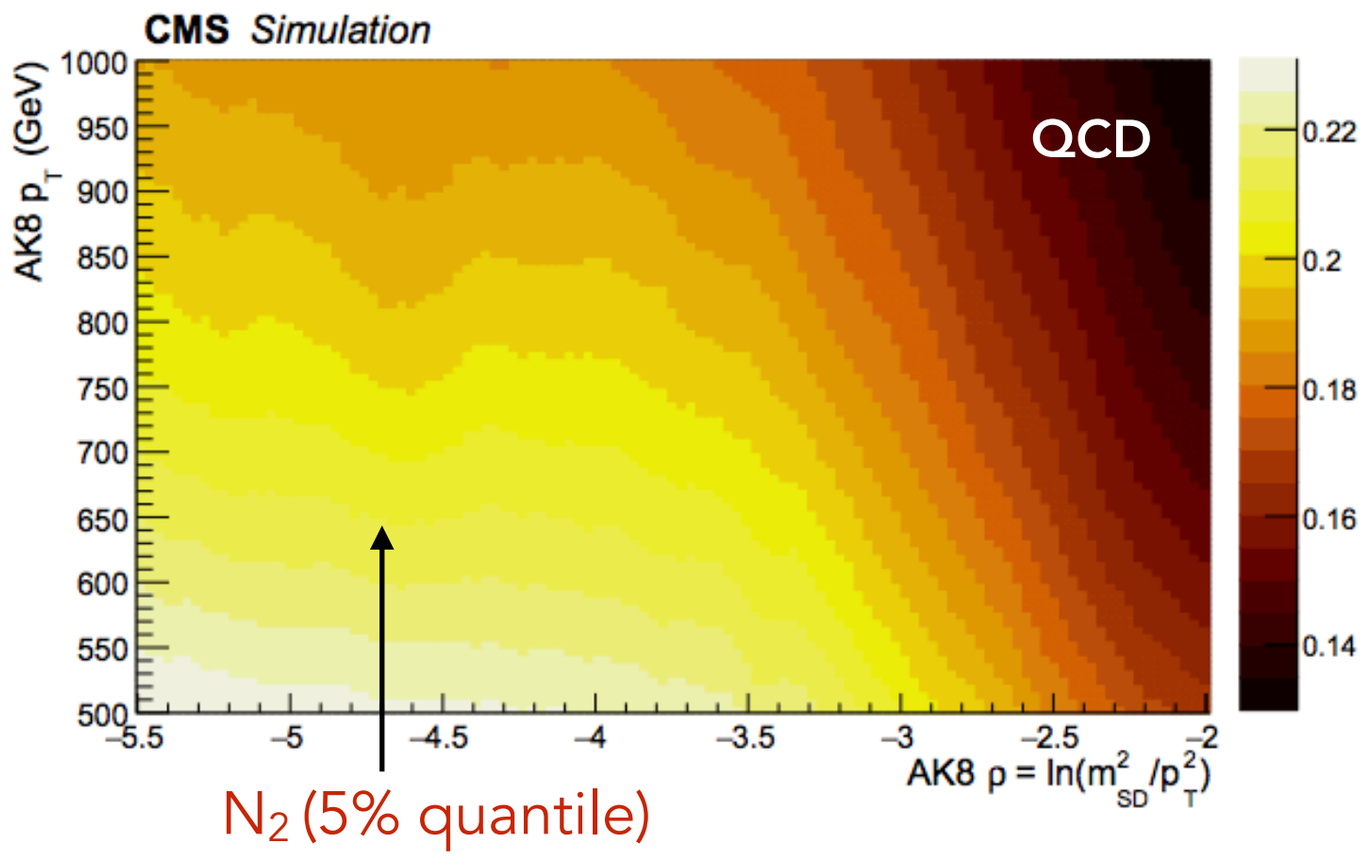
For successive cuts on N_2

This should be **smoothly falling** but instead it will start to **peak depending on jet pT**

We want to **AVOID** this sculpting

Jet tagging observables

- Challenge: for QCD events the **N_2 variable is correlated with jet mass and pT:**



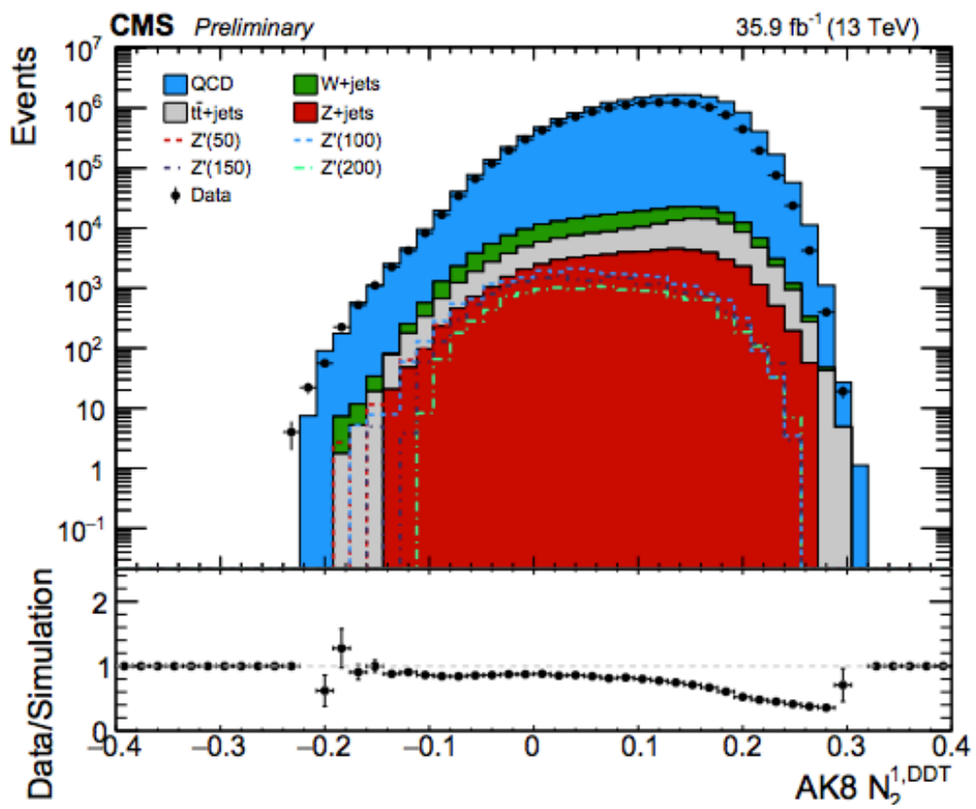
New variable defined from kNN interpolated transformation as:

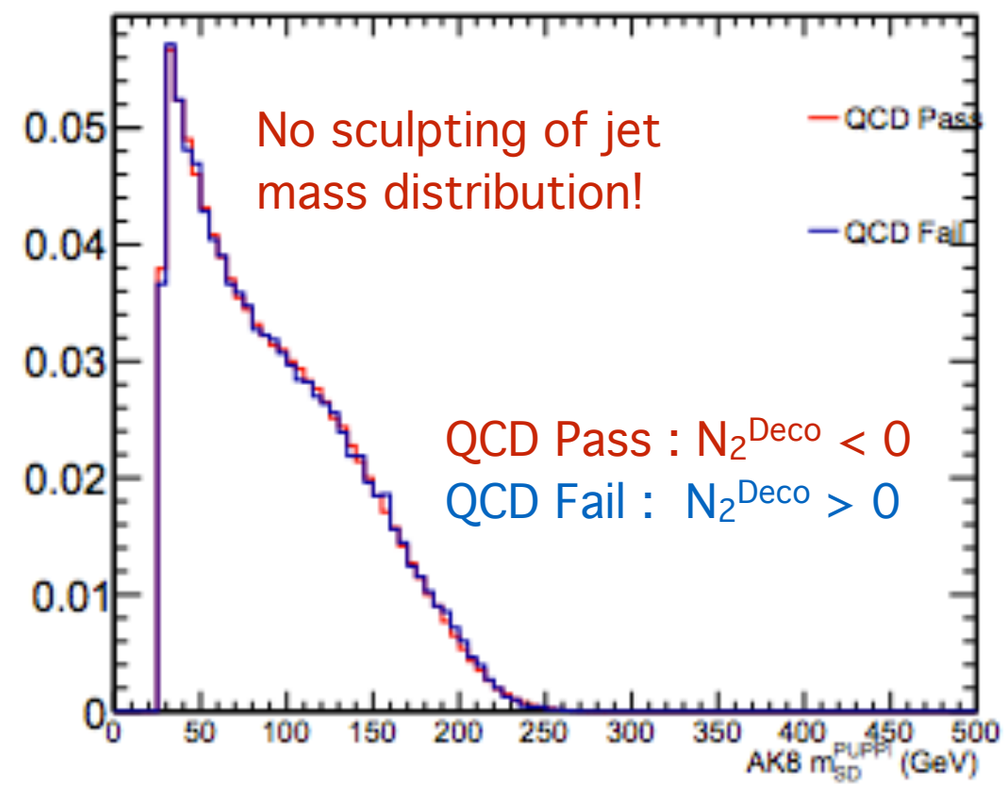
$$N_2^{DDT} = N_2 - N_2(\varepsilon_{QCD})$$

Jets Substructure Selection:

- $N_2^{Deco} < 0$:
  - N_2^{Deco} derived with ε_{QCD} =5% Bkg Efficiency.

Jet tagging observables





New variable defined from kNN interpolated transformation as:

$$N_2^{DDT} = N_2 - N_2(\varepsilon_{QCD})$$

Jets Substructure Selection:

- $N_2^{DDT} < 0$:
  - N_2^{DDT} derived with ε_{QCD} =5% Bkg Efficiency.

Background Estimate

Background Processes

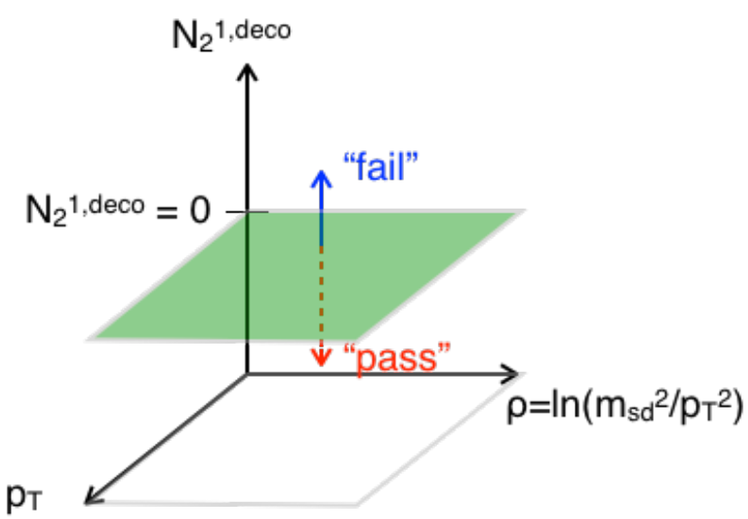
- Top
- W/Z SM candles

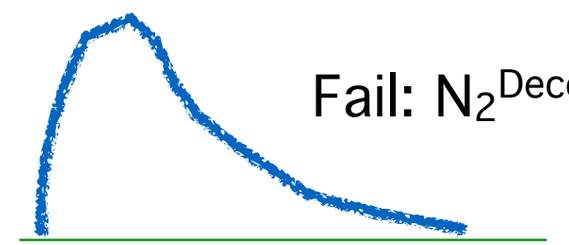
→ Estimated from simulation

- QCD contribution

→ Estimated from **data**

  - Fitting jet mass distribution directly would require very high order polynomial
  - Main idea: Try sideband prediction from failing to passing region:

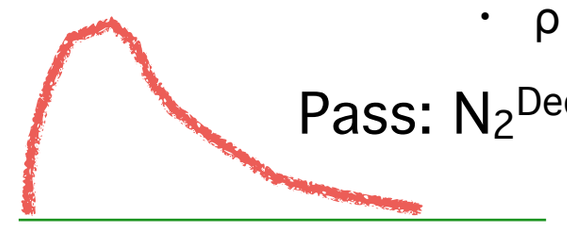




Fail: $N_2^{Deco} > 0$

same shape

- $\rho = \ln(m_{SD}^2/pT^2)$

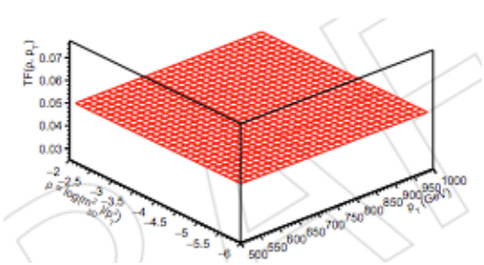


Pass: $N_2^{Deco} < 0$

- $\rho = \ln(m_{SD}^2/pT^2)$

$$p_{\text{pass}}^{\text{QCD}}(m_{SD}, p_T) = \mathcal{F}(\rho(m_{SD}, p_T), p_T) \times p_{\text{fail}}^{\text{QCD}}(m_{SD}, p_T)$$

← **QCD pass = pass/fail * QCD fail.**



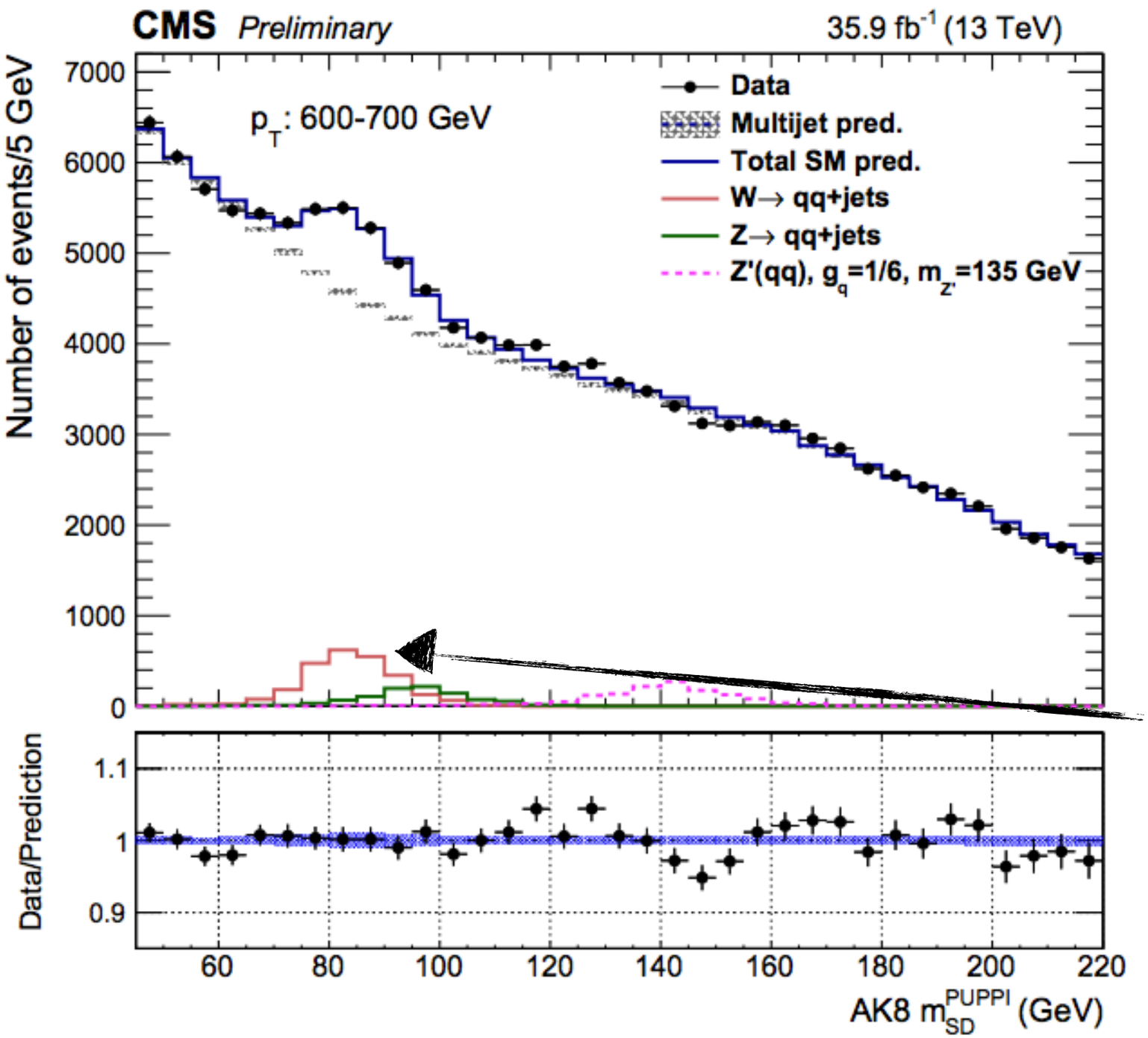
i.e. pass/fail ratio models discrepancies between QCD MC and data

- **One simultaneous fit derives QCD prediction and performs signal extraction!**

Results

Jet mass shape

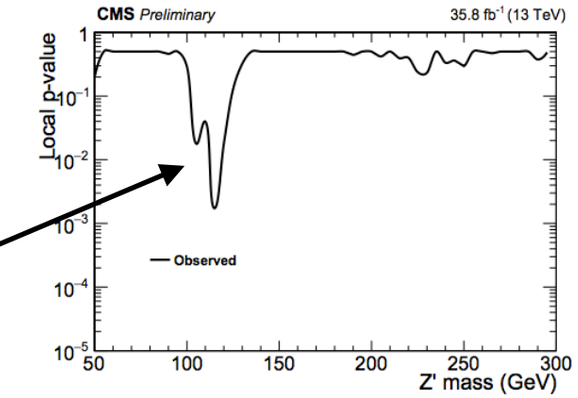
- Jet mass distribution fit down to ~ 50 GeV.
- Interpretation for resonance masses in terms of coupling!

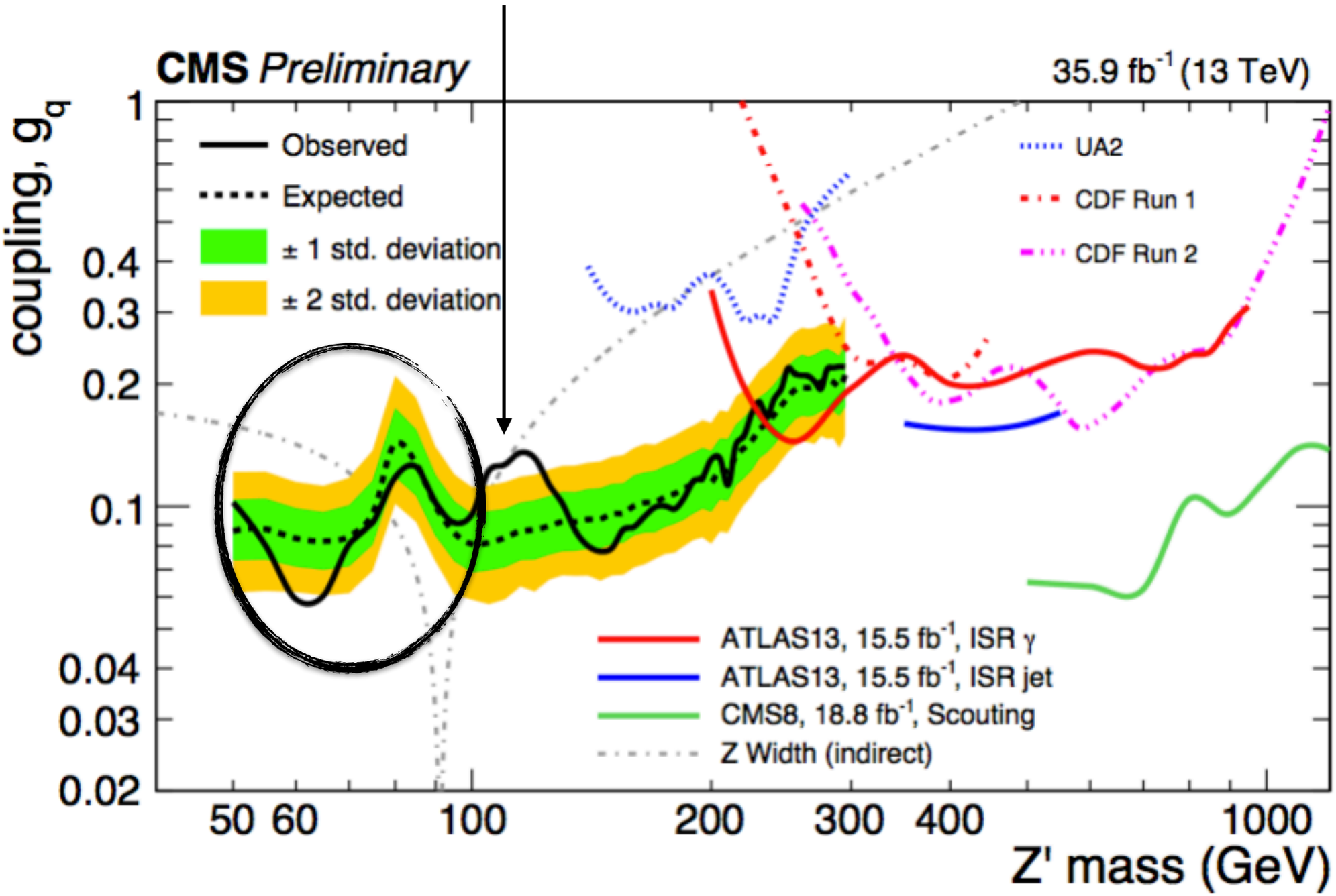


SM candles: in-situ constraint of Z' signal systematics

Boosted Z' limit interpretation

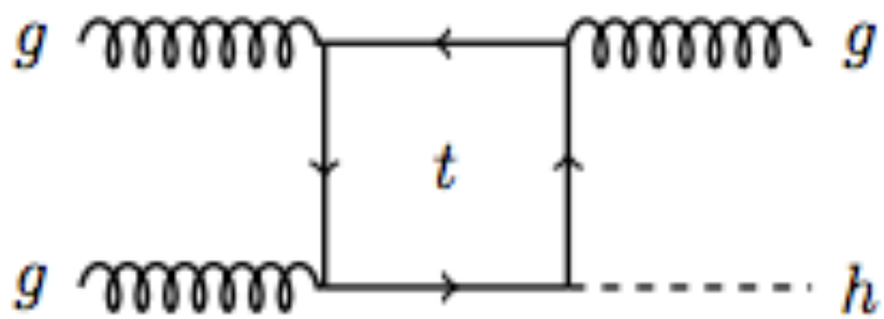
- First limits in the 50-100 GeV region!
- Modest excess observed in 115-125 GeV region corresponding to 2.2 σ deviation (global).





Summary

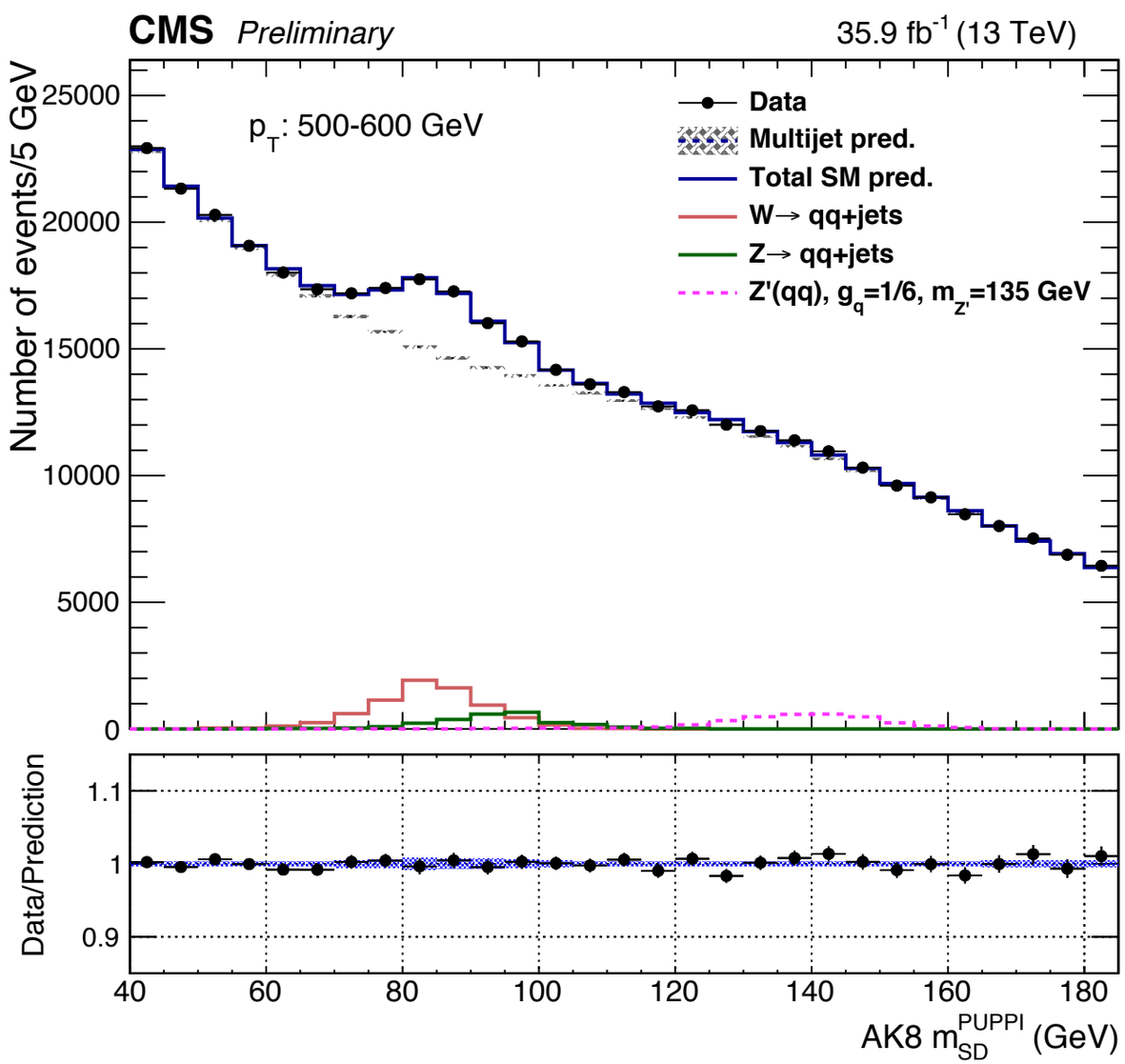
- Search for light mass resonances using full 2016 dataset
  - Extends search to lower masses (below 100 GeV!)
- Includes new substructure tagging techniques (ECFs)
- Novel background estimation technique and fitting strategy
- Technique applied to ggH analysis - check it out! CMS-HIG-17-010

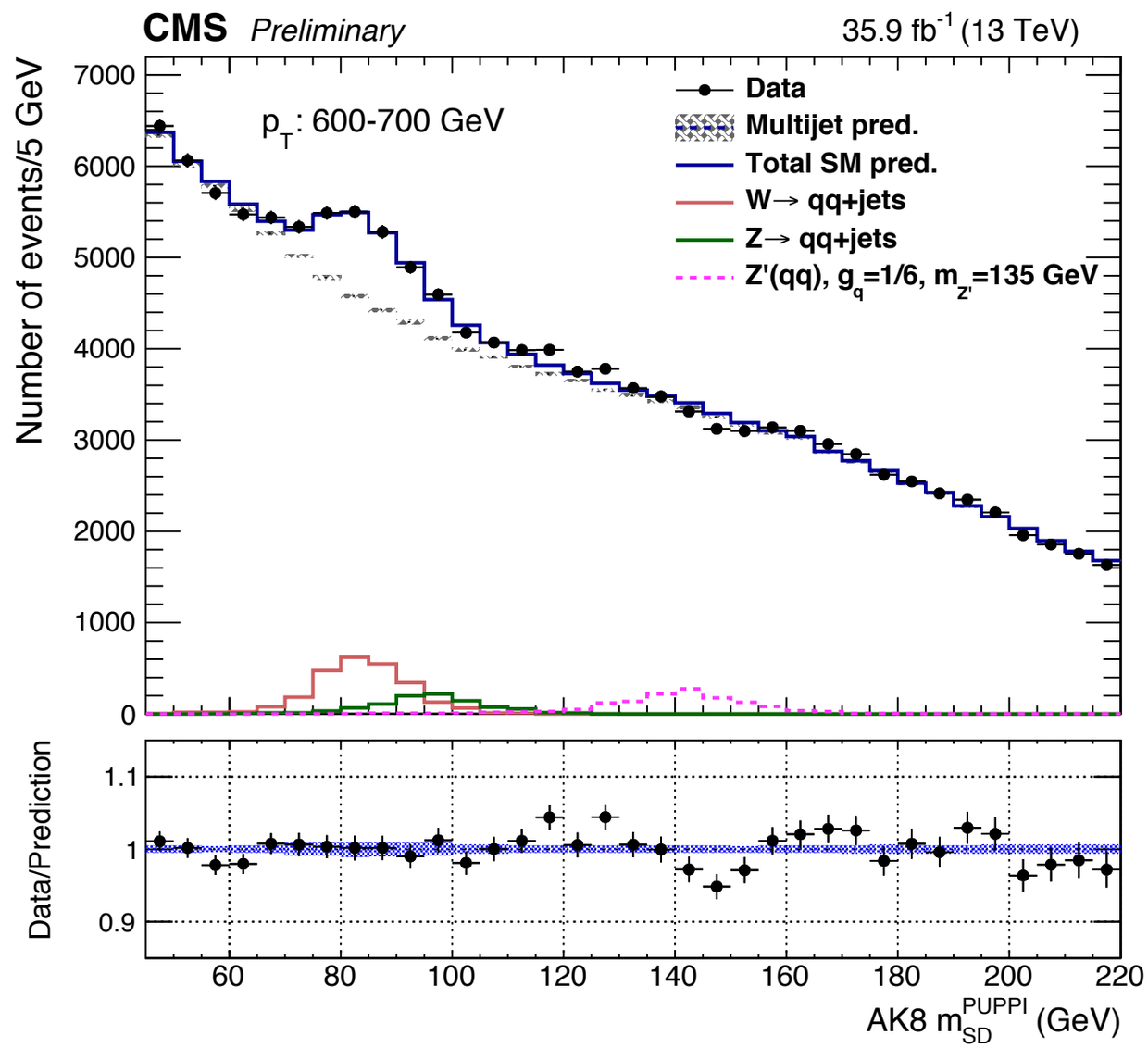


- Simultaneous fit: background estimation and signal extraction
- Setting limits at < gB = 1 + Added DM interpretation
- Expect updates with 2017 data!

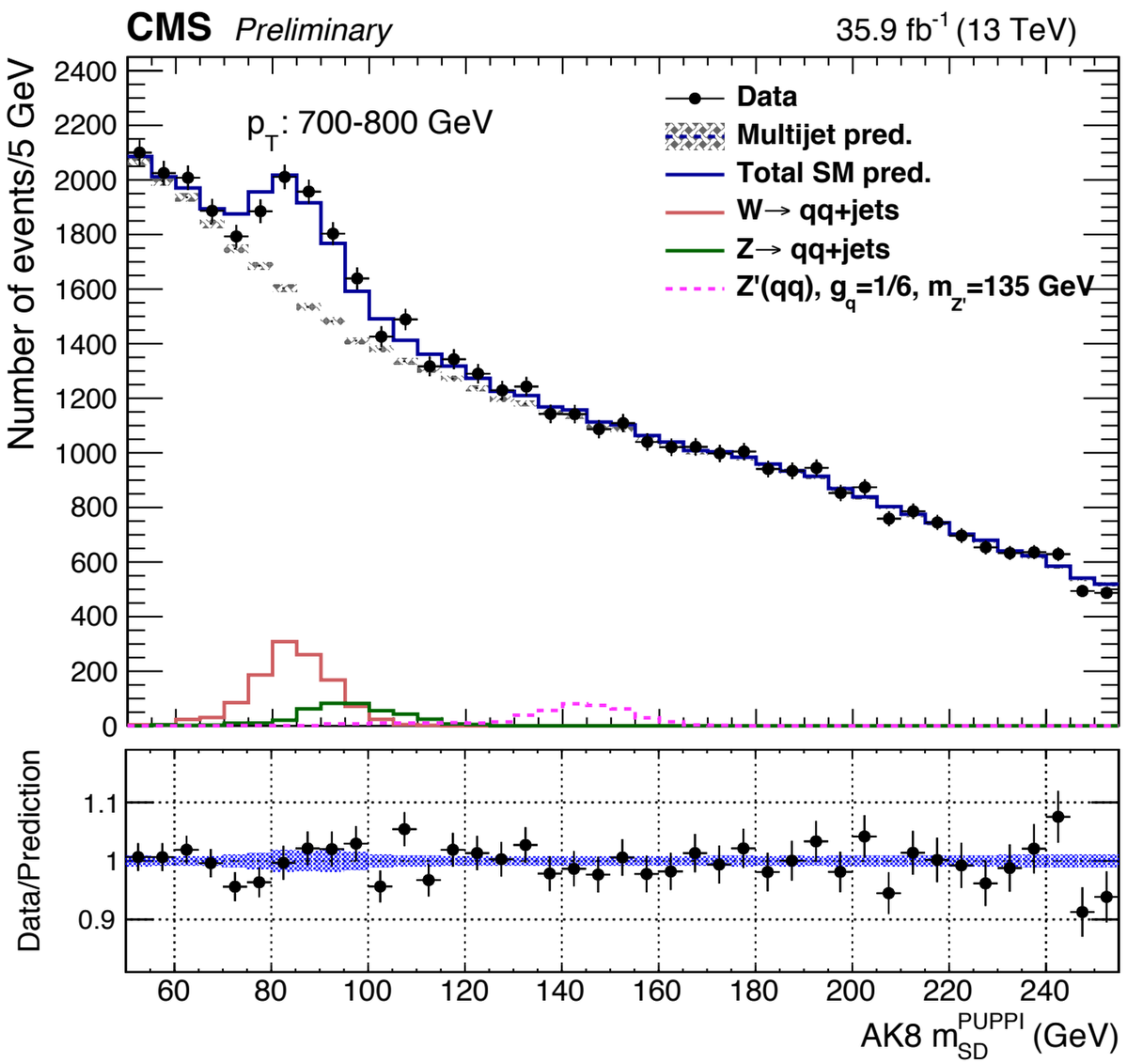
Additional Material

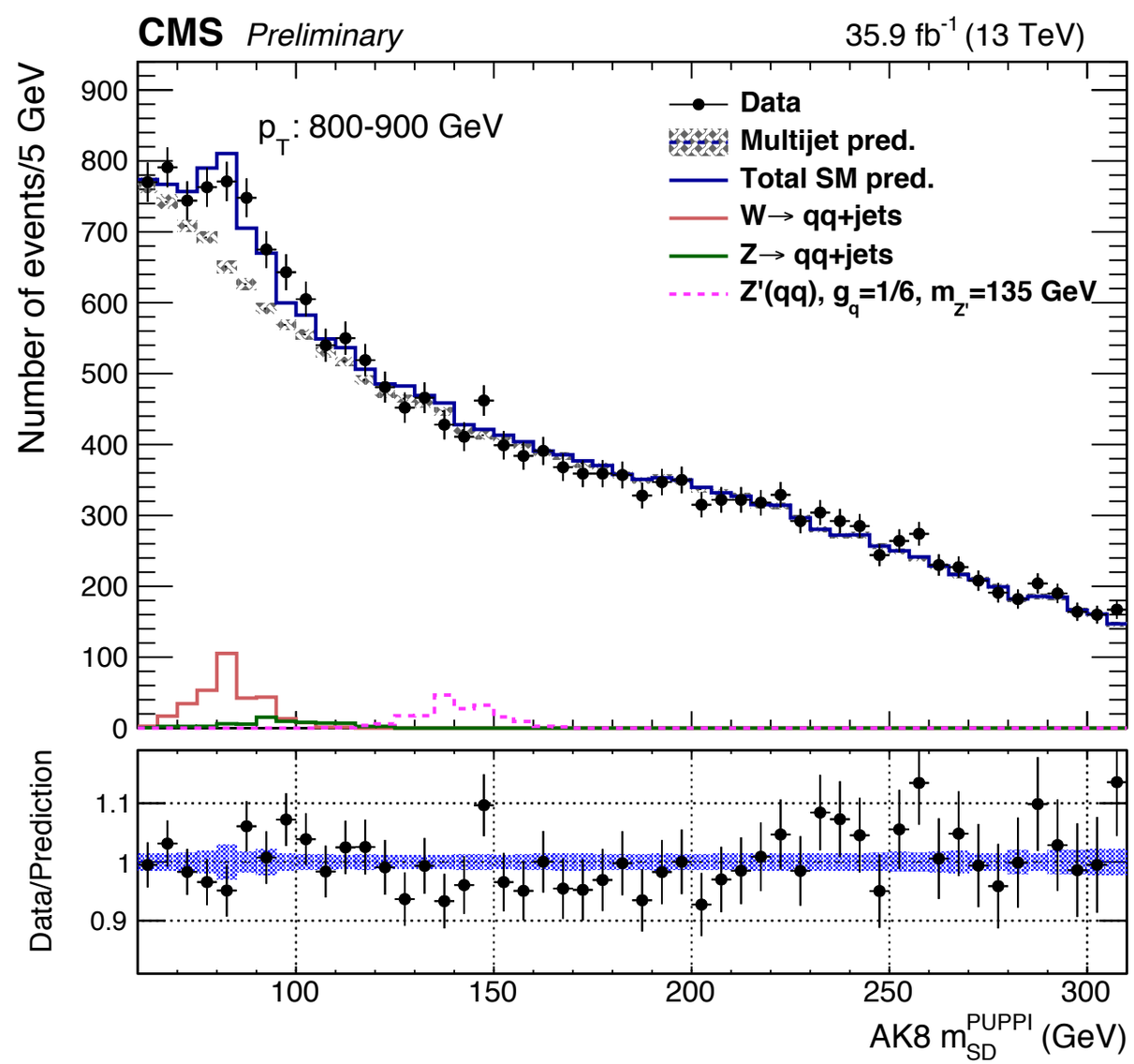
Jet mass by pT categories



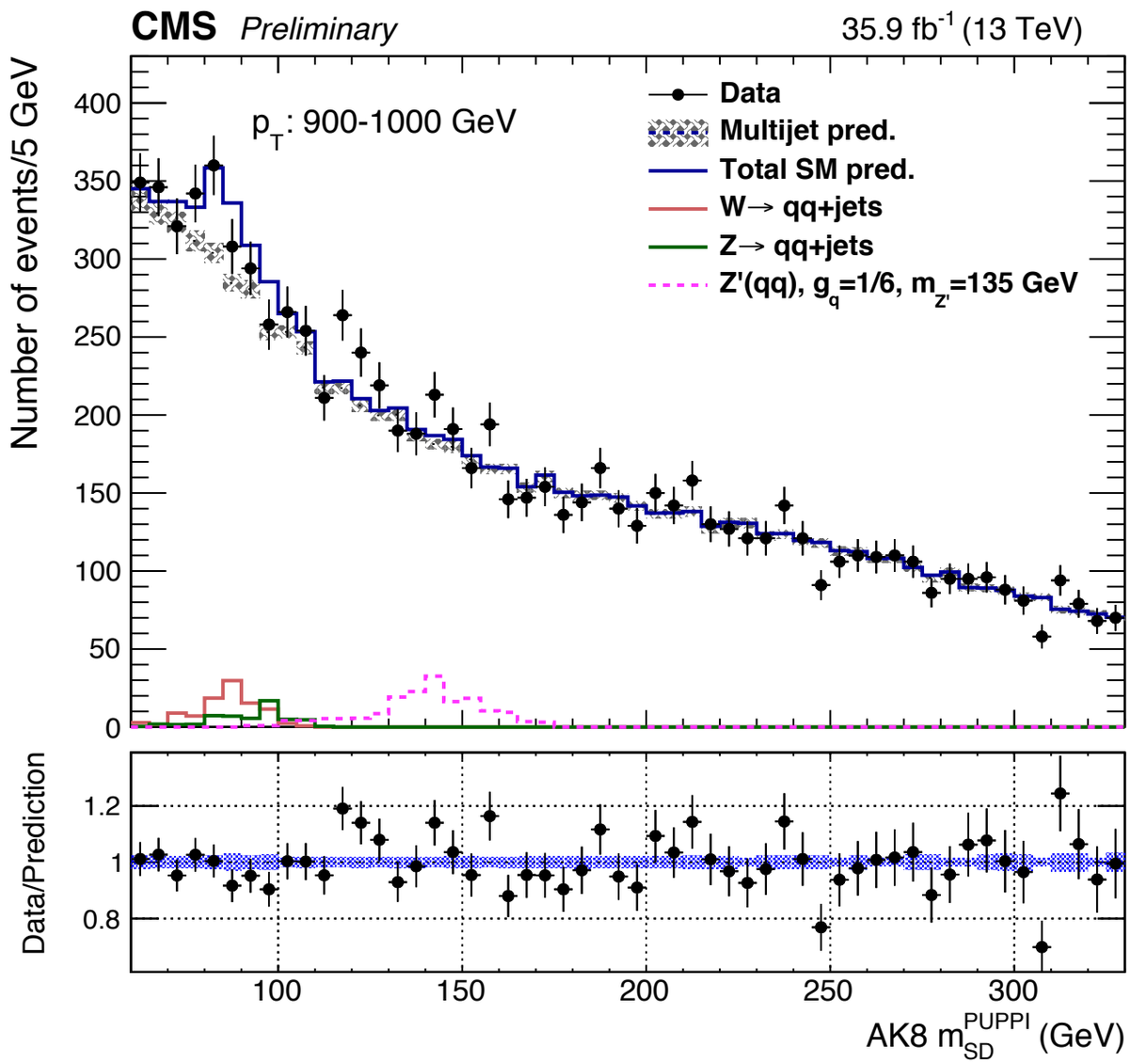


Jet mass by pT categories

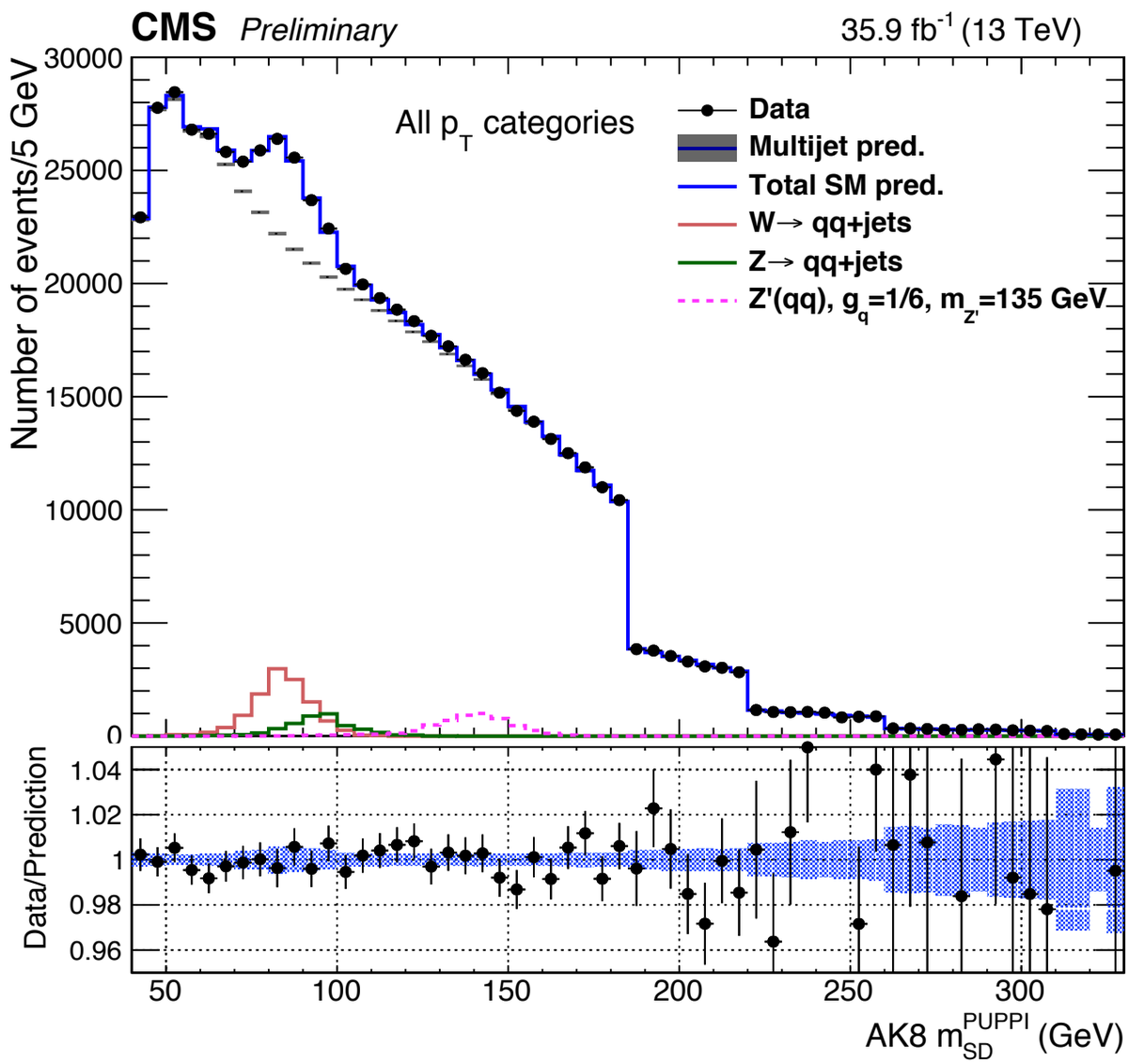




Jet mass by pT categories



Jet mass by pT categories

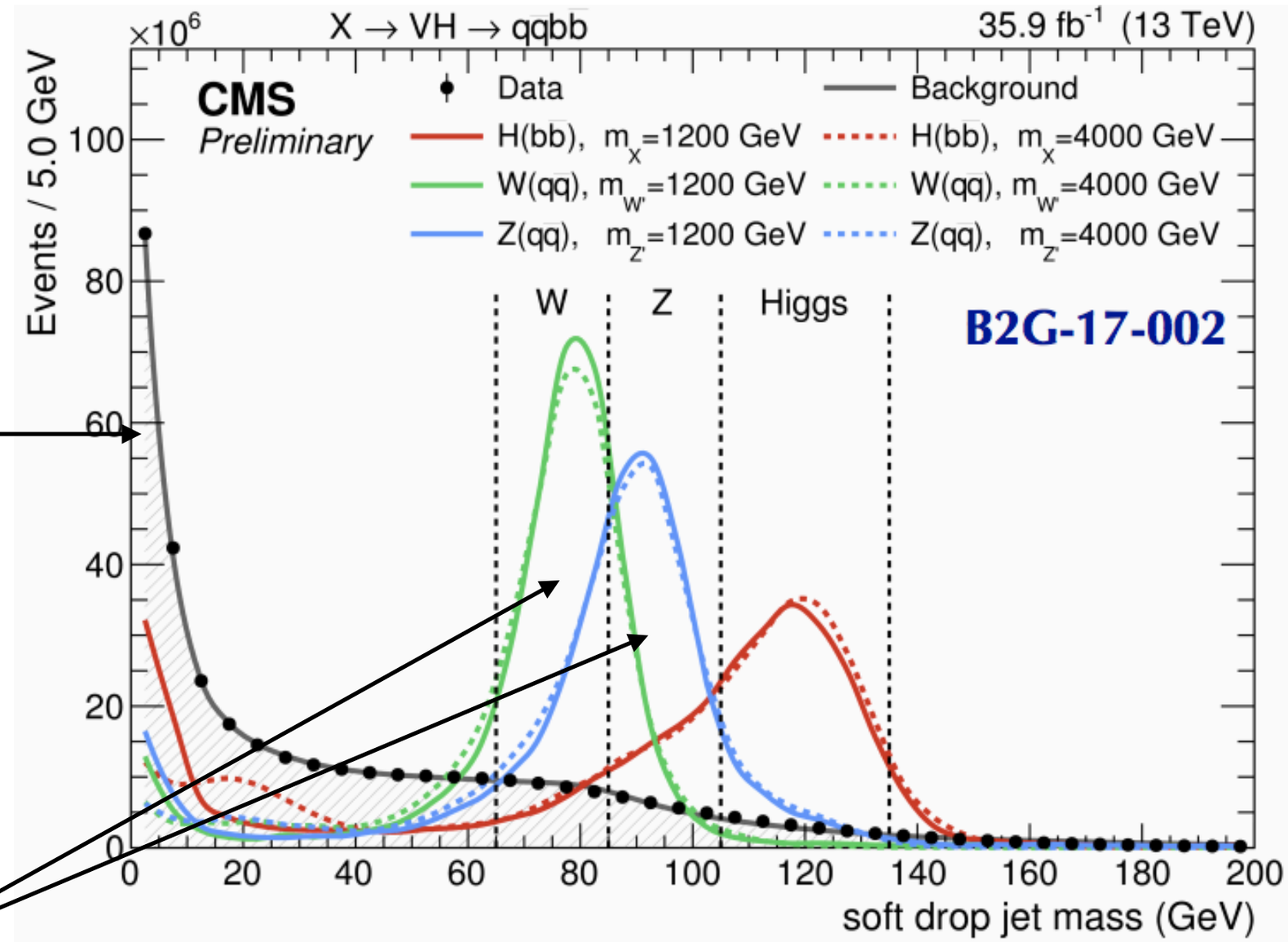


Jet tagging observables

We look for a bump in the **JET MASS**:

smoothly-falling

peaks at signal



Soft drop algorithm: gets rid of linear and soft radiation in the jet

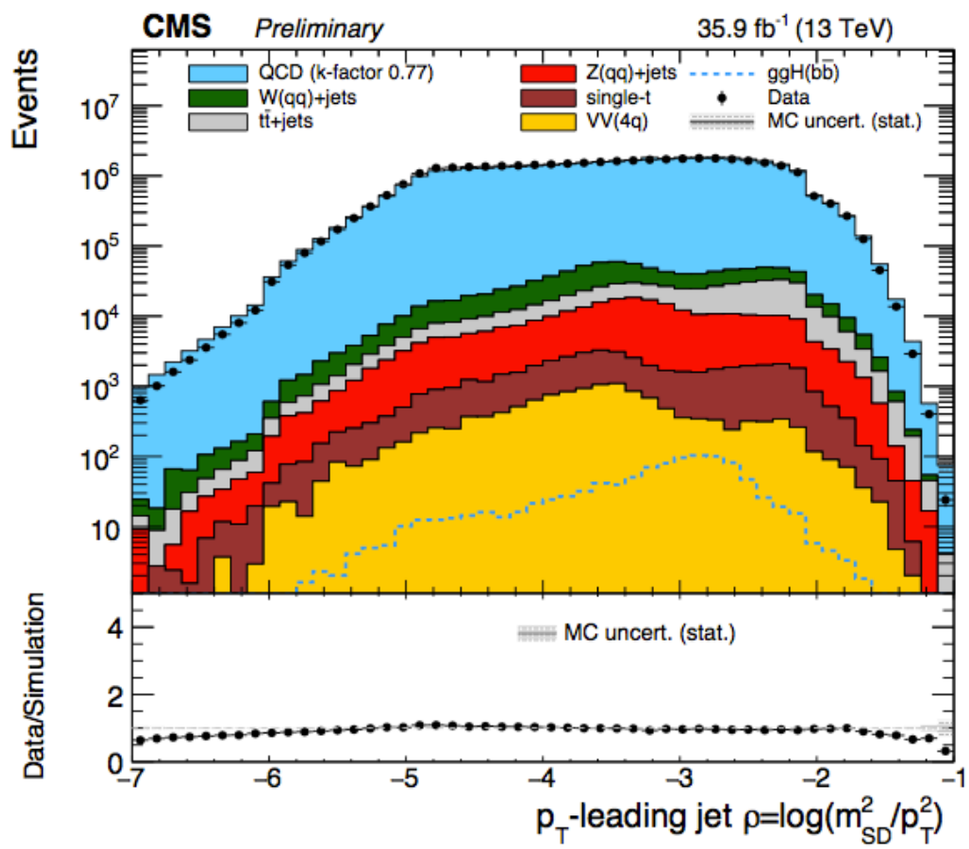
Tagging observables

We use the QCD scaling variable : **JET RHO** $\rho = \ln(m_{SD}^2/p_T^2)$

i.e. for every pT bin the distribution is the same

Is a function of the jet mass m_{SD} and jet pT!

HIG-17-010



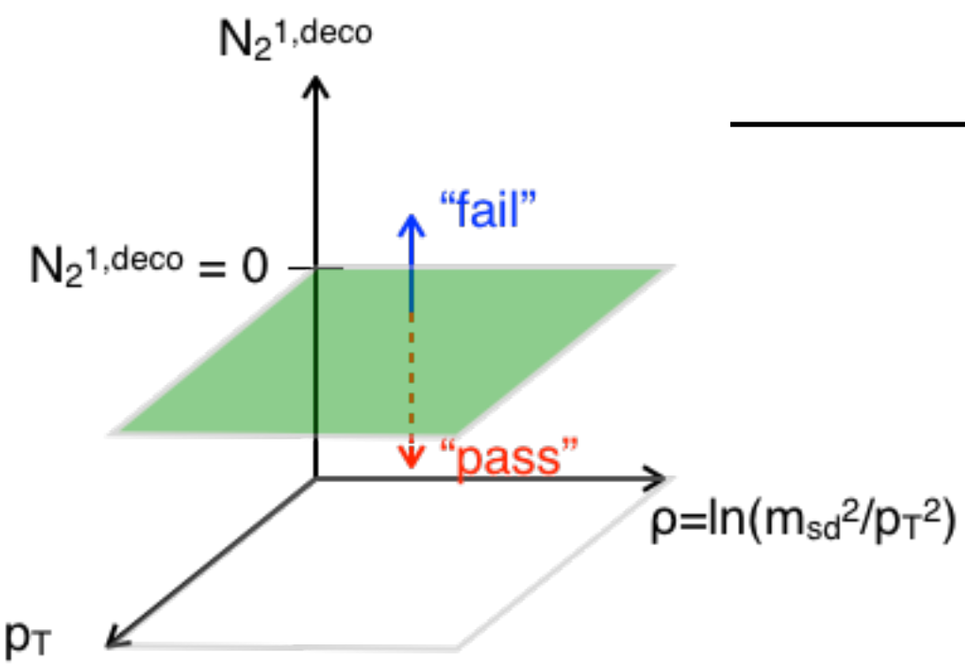
We select: $-5.5 < \rho < -2$

e.g. for a jet of pT: 500 GeV, we select on $30 < m_{SD} < 184$ GeV

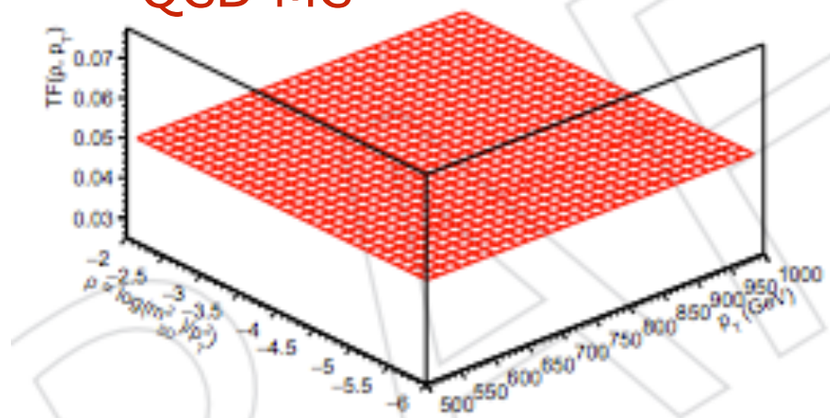
And use it to study the **dependence** of our **selection variables** with the **jet mass and pt**

QCD Background Estimate

Estimated from sideband or "failing" region:



QCD MC



If MC is absolutely correct your TF(ρ,pT) should be flat
But in data there are residual correlations -> Derive TF!

Fit for N_2^{Deco} residual dependence i.e. derive pass-to-fail ratio or TF(ρ,pT):

$$p_{\text{pass}}^{\text{QCD}}(m_{SD}, p_T) = \mathcal{F}(\rho(m_{SD}, p_T), p_T) \times p_{\text{fail}}^{\text{QCD}}(m_{SD}, p_T)$$

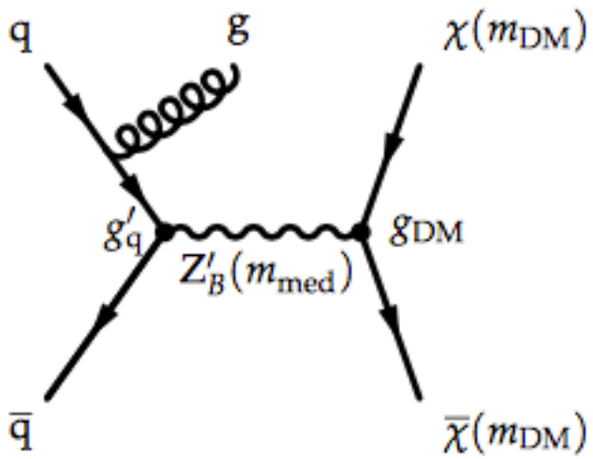
QCD pdfs.

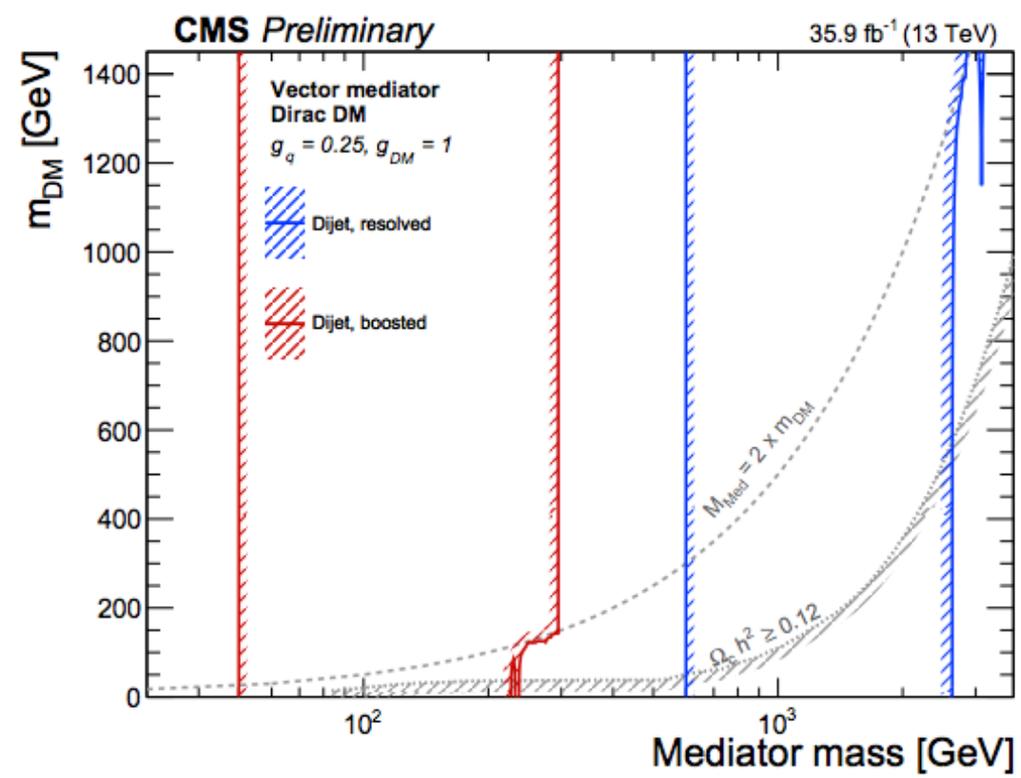
Residual correlations in data modeled as polynomial functions:

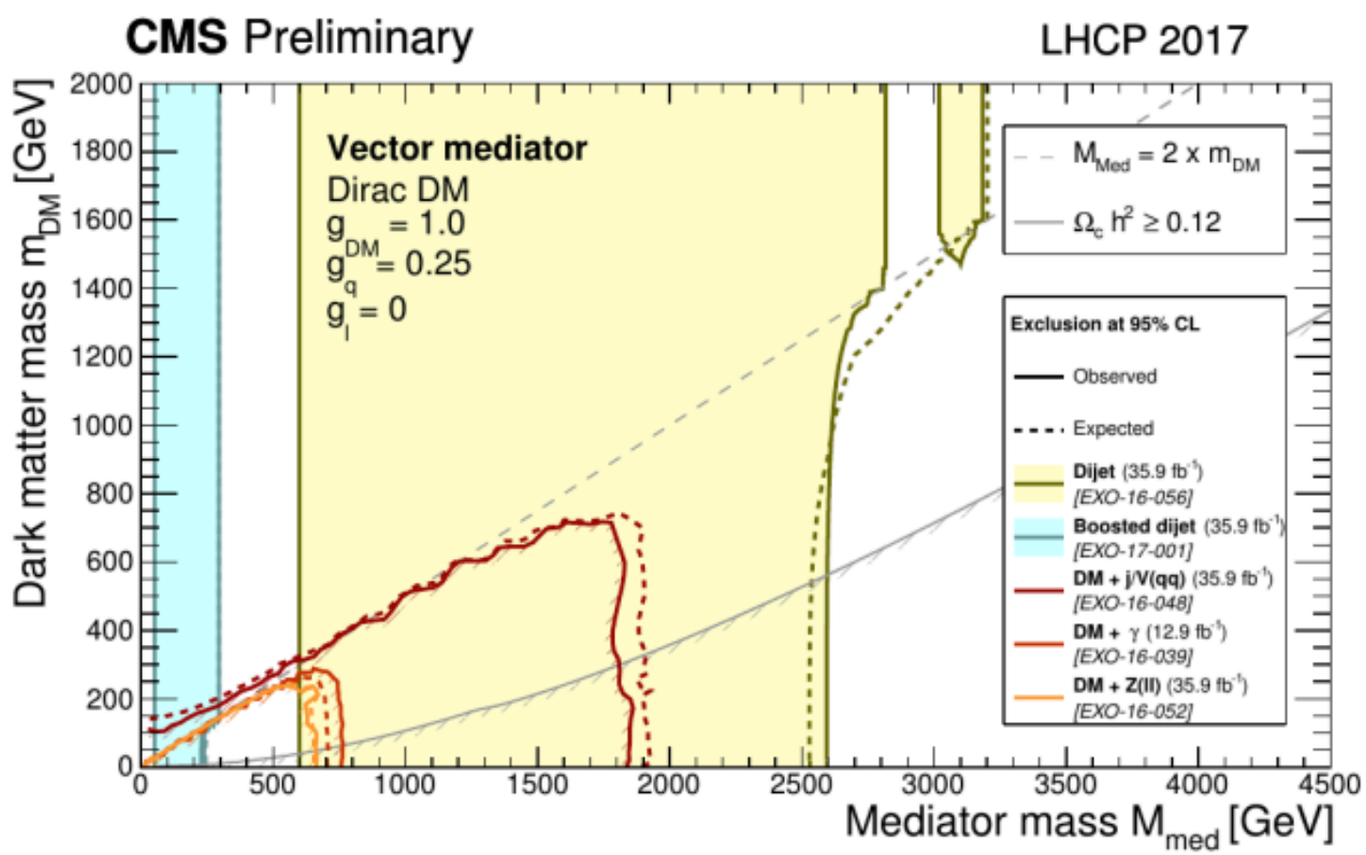
$$\begin{aligned}\mathcal{F}(\rho, p_T) = \epsilon_{\text{QCD}}(1 &+ a_{01}p_T + a_{02}p_T^2 + \cdots \\ &+ (a_{10} + a_{11}p_T + a_{12}p_T^2 + \cdots)\rho \\ &+ (a_{20} + a_{21}p_T + a_{22}p_T^2 + \cdots)\rho^2 + \cdots,\end{aligned}$$

i.e. TF models discrepancies between QCD MC and data

Connection with DM







Systematic Uncertainties

- Background estimation uncertainties:
  - From the parametric uncertainties of the fit

- Systematics for W/Z+Jets, Top and signal

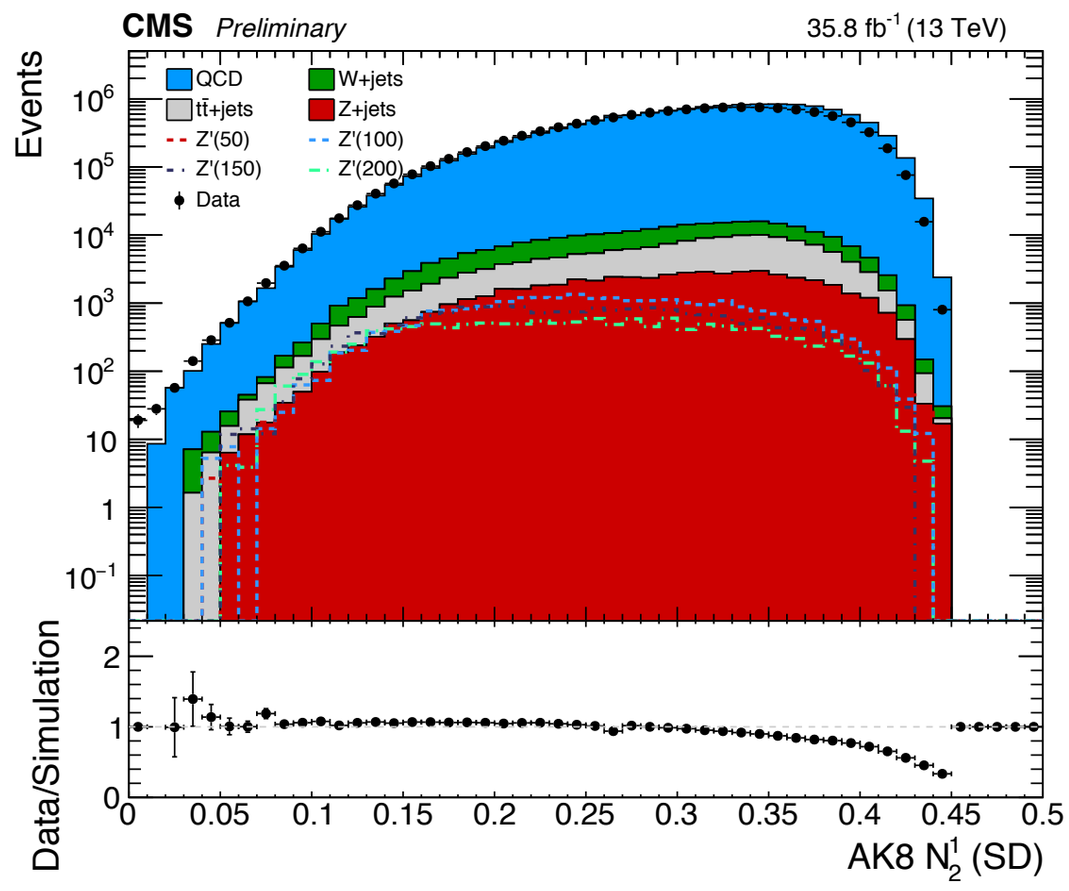
Systematic Effect	SM W/Z and Z′	$t\bar{t}$
W tag ($N_2^{1,deco}$) scale factor	9%	-
NLO QCD p_T corrections	10%	-
NLO EWK p_T corrections$^{\triangle}$	(15%,25%,35%,35%,35%)	-
W/Z ratio NLO p_T corrections$^{\triangle}$	5%	-
jet energy/mass resolution†	10%	-
jet energy/mass scale†	0.5%	-
jet energy/mass scale p_T dependent†	(0.5%,1%,1.5%,2%,2%)	-
luminosity	2.2%	-
trigger	2%	-
lepton veto	0.5%	-
top normalization	-	10%
top mistag rate	-	2%

Jet substructure selection

- Generalized ECFs (https://arxiv.org/pdf/1609.07483v1.pdf) are an alternative measure of jet substructure:

$$ {}_1e_2^\beta = e_2^\beta = \sum_{1\le i<j\le n_J} z_i z_j \Delta R_{ij}^\beta \qquad z_i \equiv \frac{p_{Ti}}{\sum_{j\in \text{jet}} p_{Tj}} $$

$$ {}_2e_3^\beta = \sum_{1\le i<j<k\le n_J} z_i z_j z_k \min\{\Delta R_{ij}^\beta \Delta R_{ik}^\beta, \Delta R_{ij}^\beta \Delta R_{jk}^\beta, \Delta R_{ik}^\beta \Delta R_{jk}^\beta\} $$

- Use power counting to relate 2 and 3 point correlation functions.
- N_2 (β=1) (N-subjetiness like variable) is of our interest:
- ECFs better probes of soft and collinear radiation

$$ N_2(\beta) = \frac{{}_2e_3^\beta}{(e_2^\beta)^2} $$