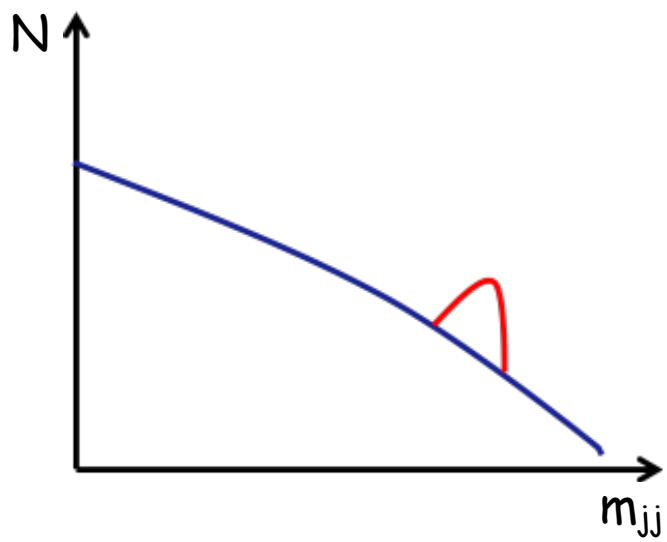
Precision searches in dijets at the HL-LHC and HE-LHC

S.Chekanov, T.Childers, D.Frizzell, G.Bodwin, J.Proudfoot, R.Wang
Argonne National Laboratory

HL/HE-LHC Meeting
April 5th, 2018

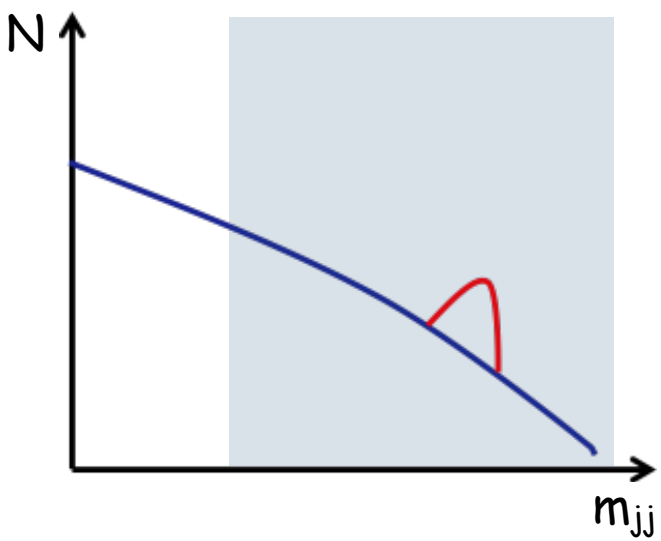
Introduction

- Di-jet final states are generic consequence of heavy particle decay in many BSM theories

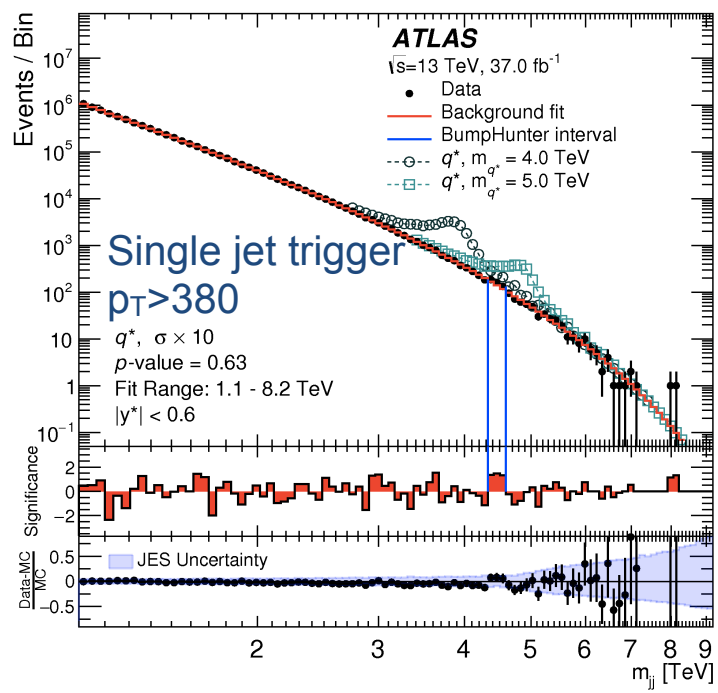


Introduction

- Di-jet final states are generic consequence of heavy particle decay in many BSM theories
- In the past, searches in dijet mass spectrum (mjj) at the LHC were explored in various mass regions according to the jet trigger thresholds

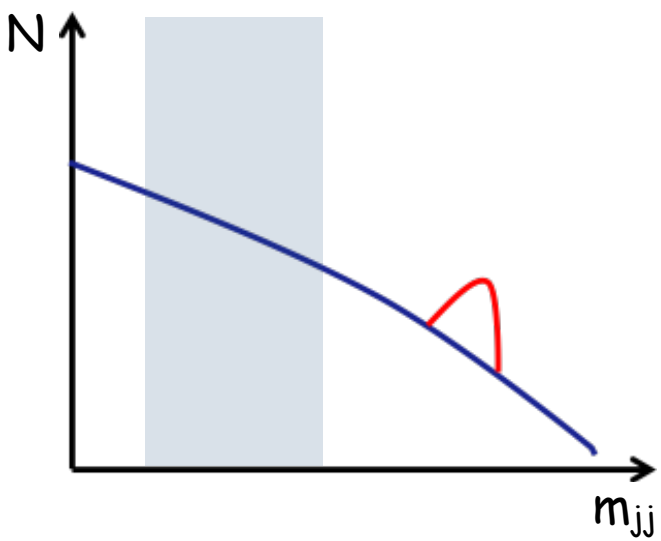


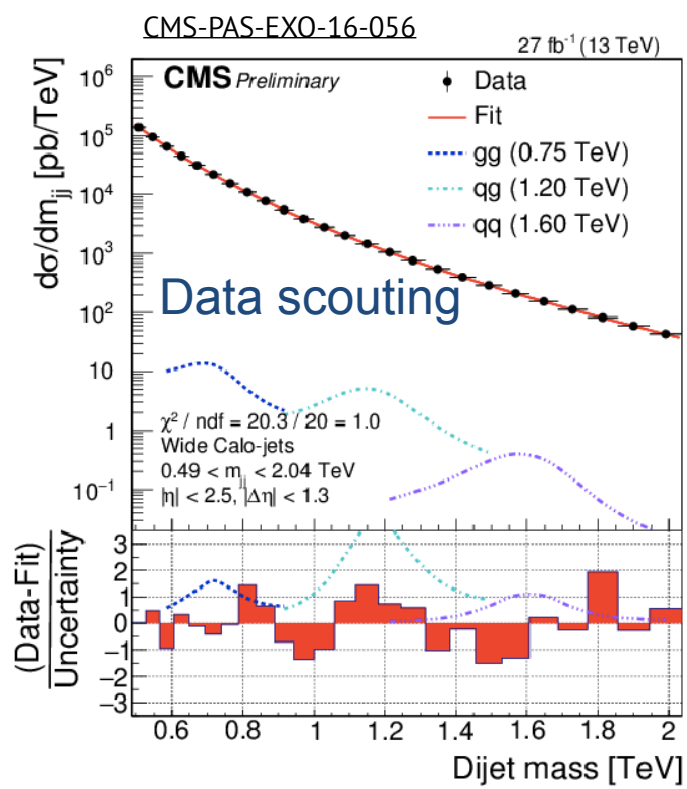
Phys. Rev. D 96 (2017) 052004



Introduction

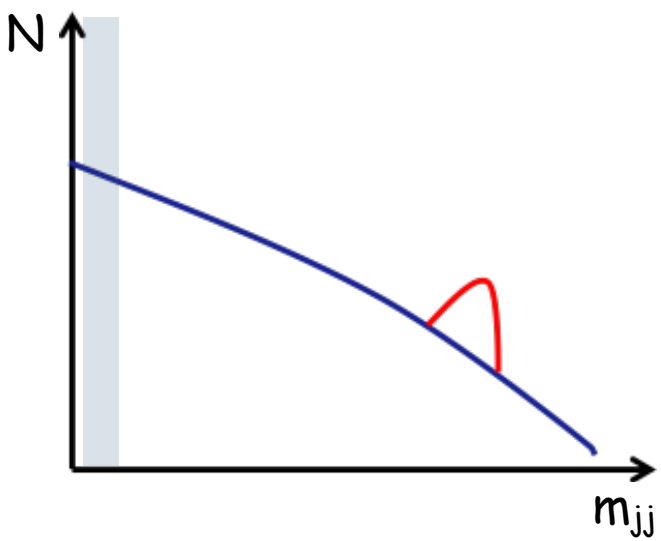
- Di-jet final states are generic consequence of heavy particle decay in many BSM theories
- In the past, searches in dijet mass spectrum (mjj) at the LHC were explored in various mass regions according to the jet trigger thresholds



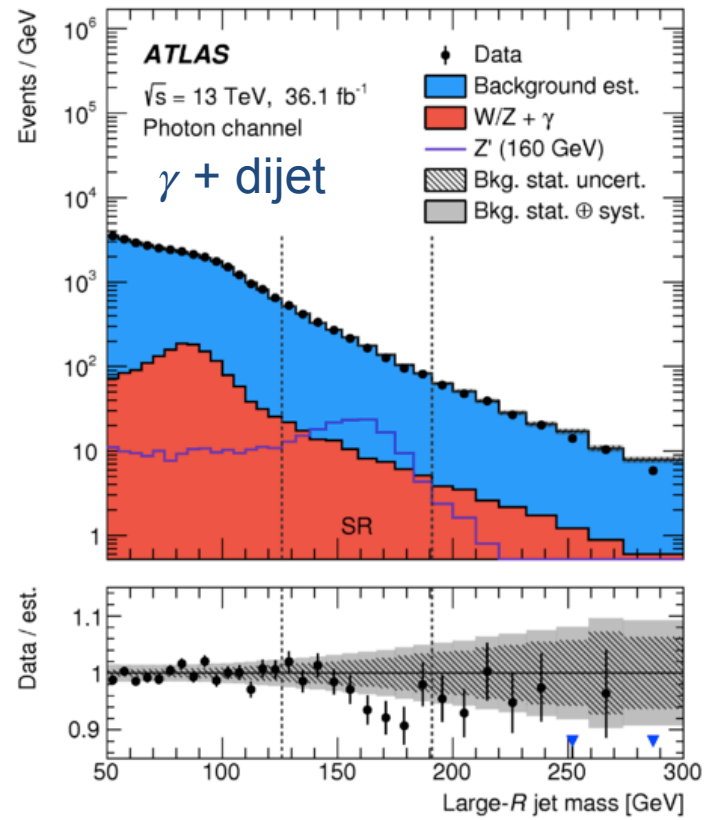


Introduction

- Di-jet final states are generic consequence of heavy particle decay in many BSM theories
- In the past, searches in dijet mass spectrum (mjj) at the LHC were explored in various mass regions according to the jet trigger thresholds
- Requiring a lepton, **photon** and or other identified objects produced in association with a dijet pair, helps to collect data with low mjj that can not be triggered by the jet triggers



arXiv:1801.08769



Introduction

arXiv:1710.09484

- Di-jet final states are generic consequence of heavy particle decay in many BSM theories
- In the past, searches in dijet mass spectrum (mjj) at the LHC were explored in various mass regions according to the jet trigger thresholds
- Requiring a lepton, photon and or other identified objects produced in association with a dijet pair, helps to collect data with low mjj that can not be triggered by the jet triggers

- The inclusive, b-tagged and muon associated dijet production will be discussed in this presentation
- We assume with further developments, a very wide range over the mjj spectrum (~100 GeV - several TeV) will be covered by a single analysis

MC sample

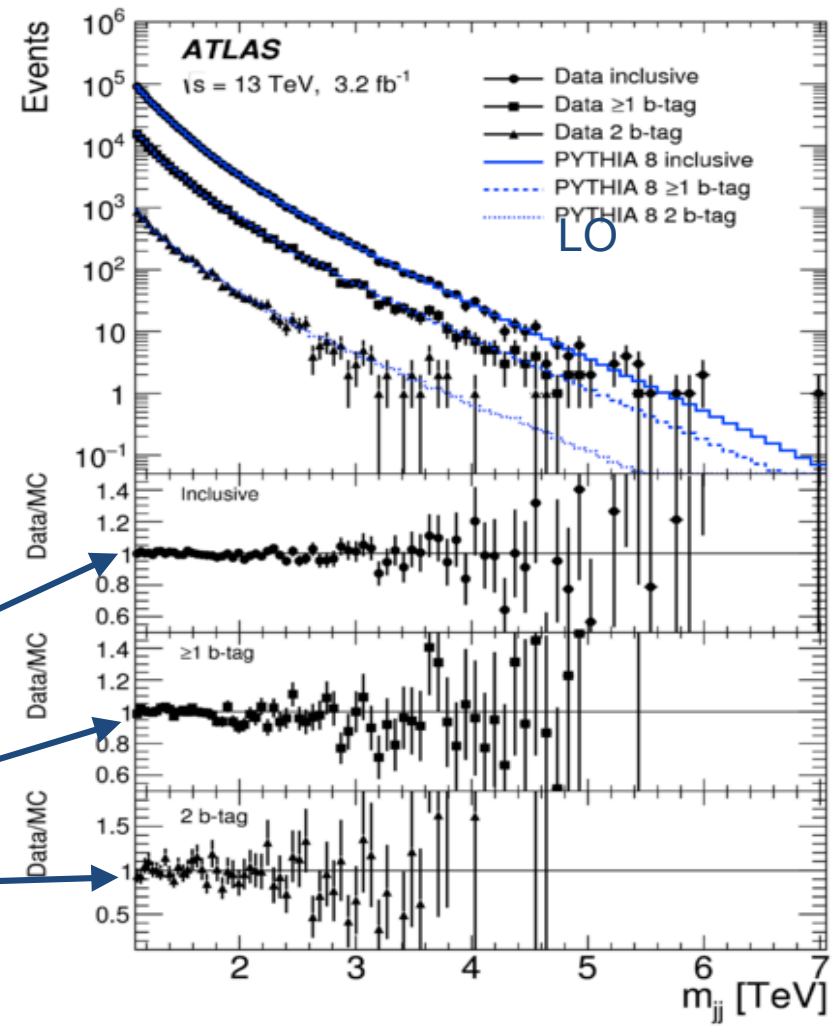
arXiv:1710.09484

- In this presentation, we use the LO MC produced on the super computer to
  - emulate the data at HL/HE-LHC
  - study the signal sensitivities
  - and explore the background modeling

- Assuming the LO MC can reflect the shape of dijet background
- Jet reconstruction and selection do not introduce signal-like structures

- ATLAS full simulations confirm such assumptions: Good agreement in shape between MC (full simulation) and data
  - $|y^*|<0.6$ ($y^*=\Delta y_{j1,j2}/2$)
    - Suppress the QCD processes
  - For both inclusive
  - and b-tagged selections

Search for resonances in the mass distribution of jet pairs with one or two jets identified as b-jets in proton-proton collisions at 13 TeV with the ATLAS detector



arXiv:1710.09484

MC sample

- 10 M CPU hours (10 days) on supercomputers at NERSC

- 100 billion events are generated in each of these three categories:
  - Light jet QCD including bb production (exclude $t\bar{t}$)
  - Vector and scalar boson production that includes the W, Z and H^0 boson processes
  - tt and single top quark production
- The combined Monte Carlo dijet mass spectrum provides a better precision than will be achievable for data at 3 ab^{-1}

- LO order Pythia8 generation with the
- default parameter settings and the ATLAS A14 tune for minimum-bias events
- 14 TeV and 27 TeV for the HL-LHC and HE-LHC

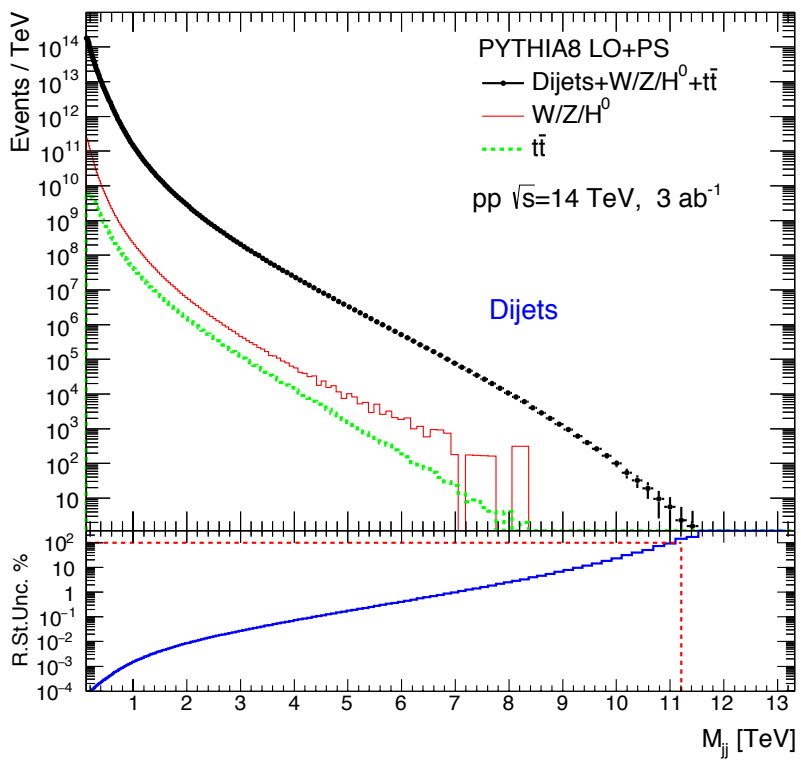
http://www.nersc.gov/

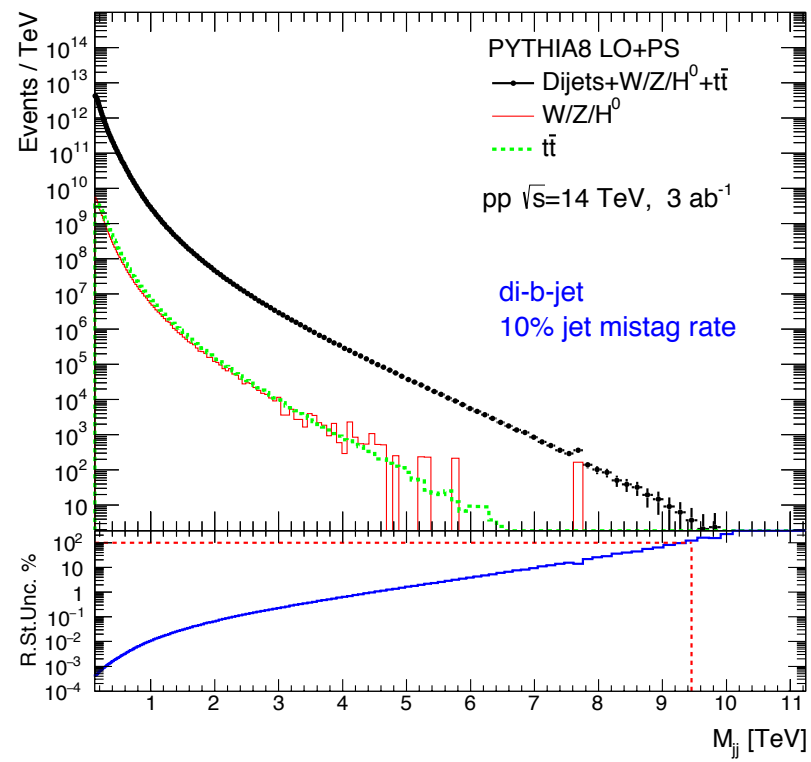
arXiv:1710.09484

MC sample

- Jets are reconstructed using FastJet
  - pt > 40 GeV to reduce pile-up contributions, |η| < 2.4
- The b-jets are selected using truth matched jets to the b quarks (ΔR=0.4 cone), and then mixed with 10% of the light jets to model the mistag rate
  - The b-jet tagging efficiency is set to 100% because the b-jet fraction is << than the light jet fraction in the background

Stacked MCs, m_{jj} > 125 GeV, bin error = $\sqrt{(N_{event})}$



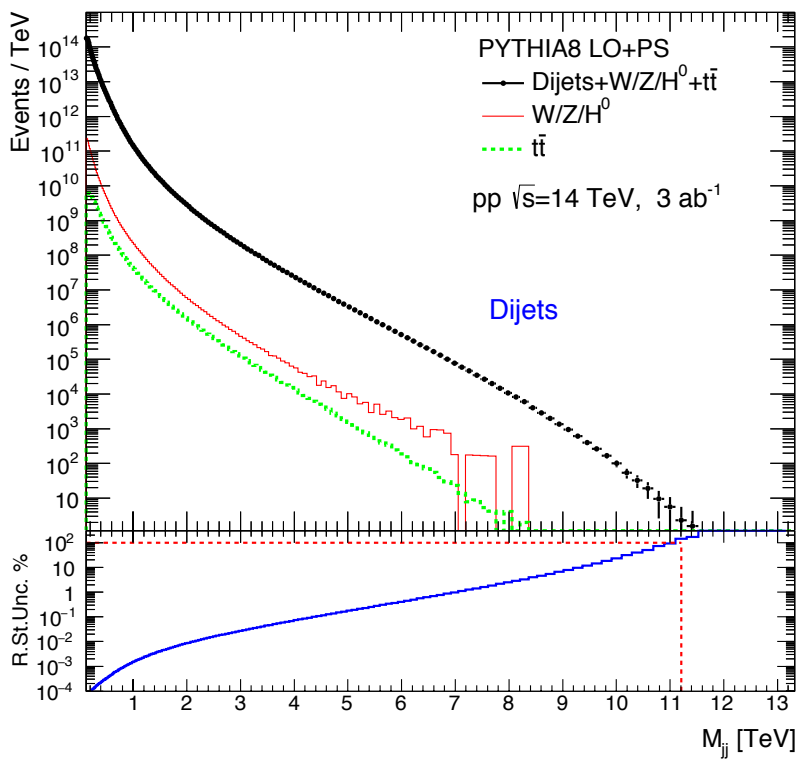


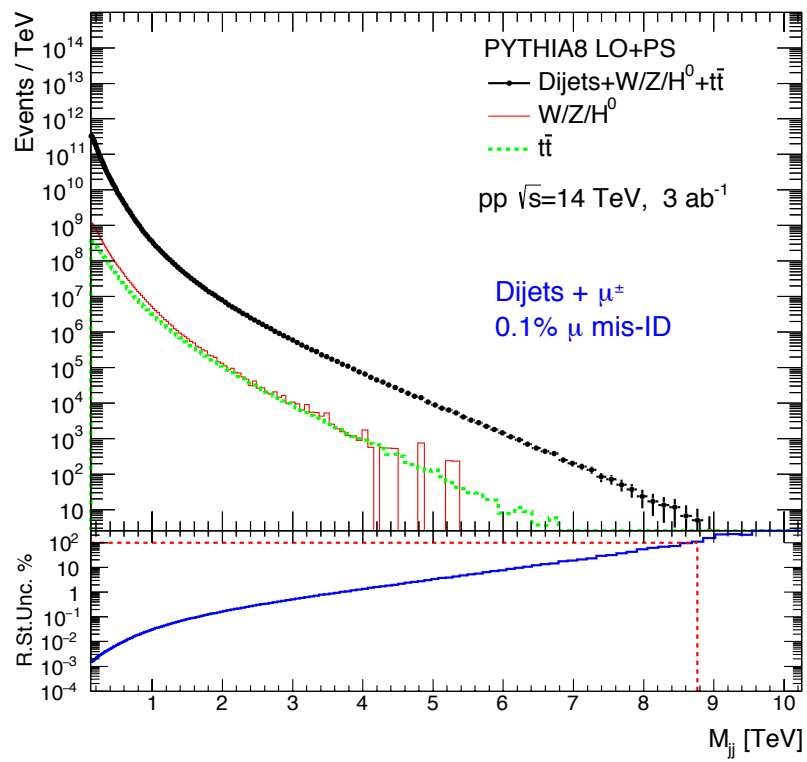
arXiv:1710.09484

MC sample

- Jets are reconstructed using FastJet
  - pt > 40 GeV to reduce pile-up contributions, |η| < 2.4
- The muons are selected with pt > 60 GeV and isolated within ΔR=0.2 cone, and then mixed with 0.1% of the light jets to model the misidentification rate
  - The fake muon pt is assigned to be the jet pt in 0.1% of these light jets selected randomly

Stacked MCs, m_{jj} > 125 GeV, bin error = $\sqrt{(N_{event})}$

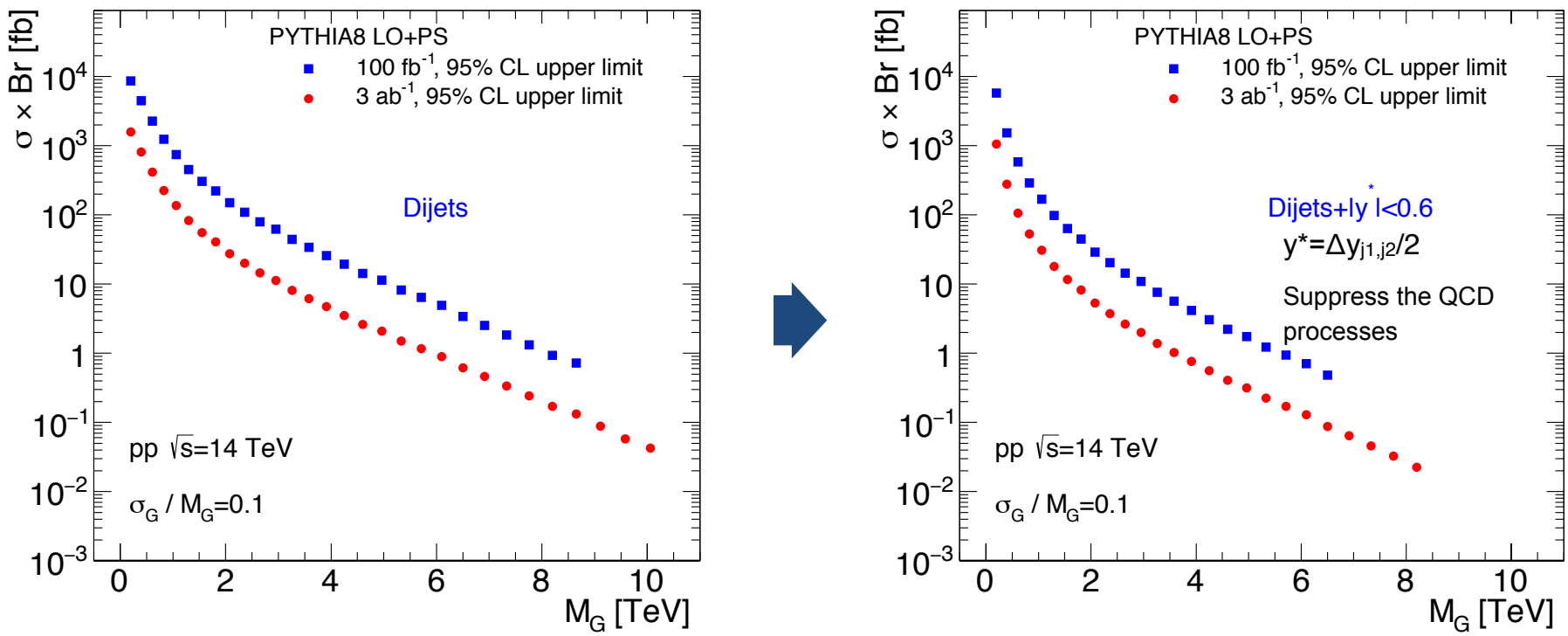




Exclusion limit

arXiv:1710.09484

- Resonance peak after reconstruction can be approximated by a Gaussian
- Gaussians with different widths can provide us guidance on the exclusion limit of possible signals
  - We study the exclusion limits of Gaussians with width $\sigma_G / M_G = 10\%$



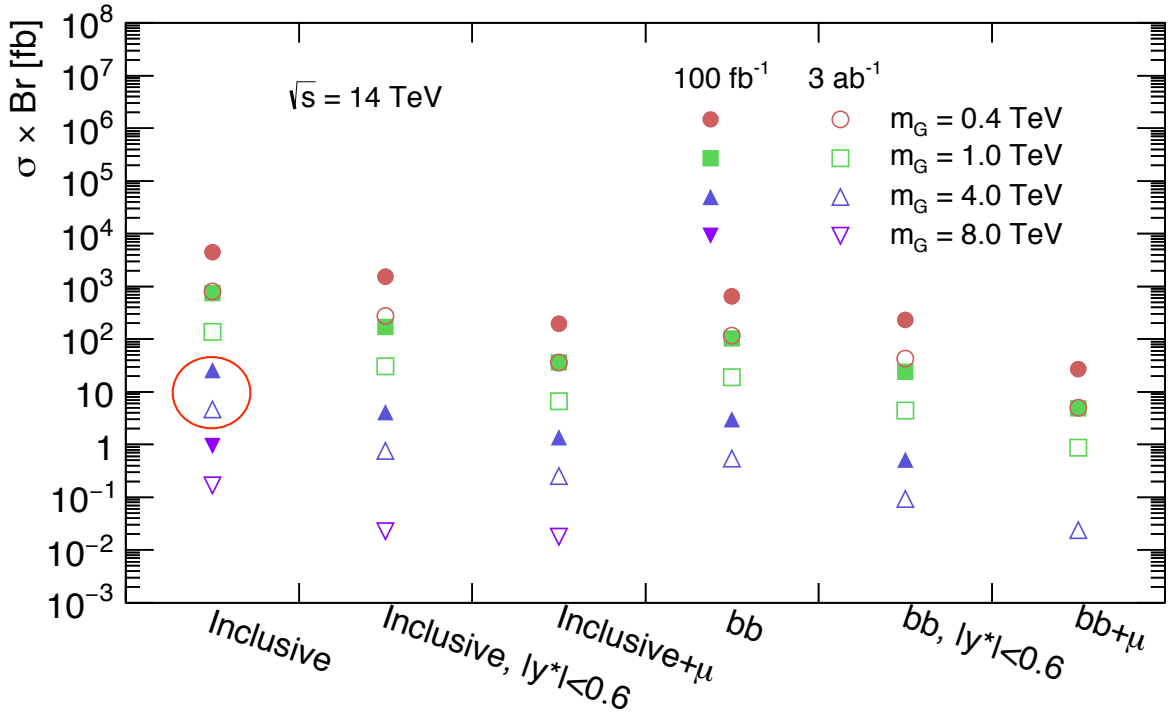
Gaussian limit

arXiv:1710.09484

- Resonance peak after reconstruction can be approximated by a Gaussian
- Gaussians with different width can provide us guidance on the exclusion limit of possible signals
  - We study the exclusion limits of Gaussians with width $\sigma_G / M_G = 10\%$
  - Compared the exclusion limit with different selections
    - Inclusive, |y*|<0.6, inclusive+muon, 2 b-tag, 2 b-tag+|y*|<0.6 and 2 b-tag+muon

When luminosity Increases from 100 fb^{-1} to 3 ab^{-1}, it gives ~10 times better exclusive limit on the cross section * Br

(eg. blue filled triangle vs blue open triangle)



Background modeling

arXiv:1710.09484

- For searches, how to model the background critical!
  - Due to computational resources, large scale MC production is limited
  - Data driven methods are becoming more and more important

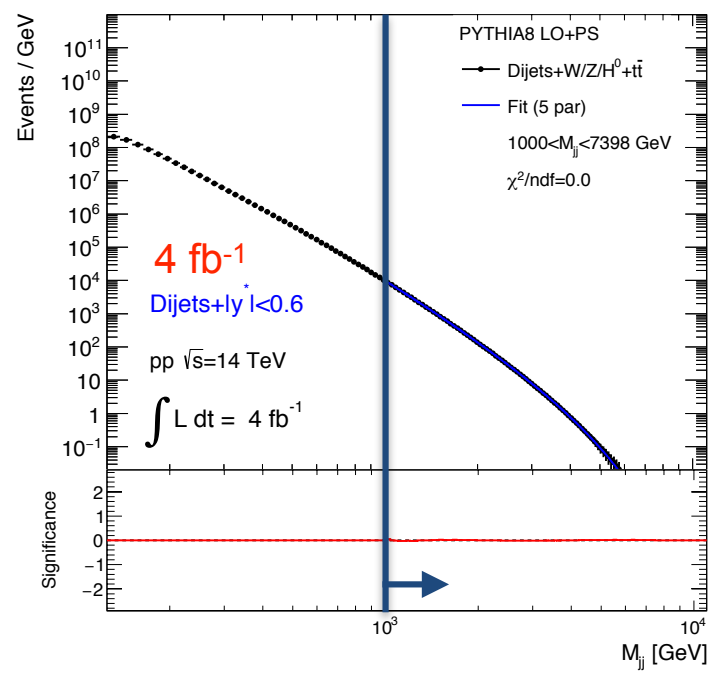
- From past experience, analytical functions have been used to describe the smoothly falling shape of the QCD background in limited mass range
  - eg. dijet resonance search (2015 data) with mjj > 1.1TeV, |y*|<0.6

➡ Test on the MC sample in the same mjj region

  - Same luminosity of 4 fb^{-1}
  - Fit function: $f(x) = p_1(1-x)^{p_2}(x)^{p_3+p_4\ln x+p_5(\ln x)^2}$

✦ The fit is perfect

Fit 1 TeV - 7 TeV



Background modeling

arXiv:1710.09484

- For searches, how to model the background critical!
  - Due to computational resources, large scale MC production is limited
  - Data driven methods are becoming more and more important

- However when move to much larger luminosity (event statistics), the fit function is not capable any more
  - eg. dijet resonance search (2015 data) with mjj > 1.1TeV, |y*|<0.6

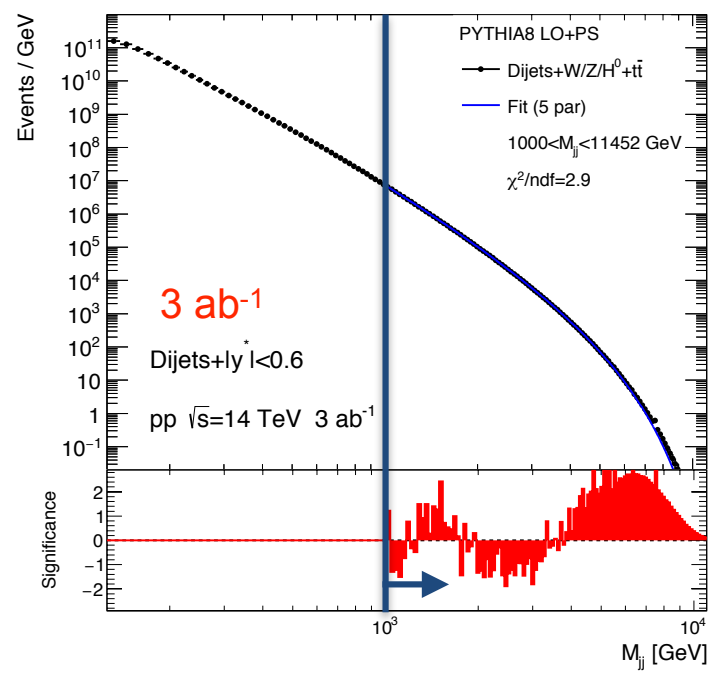
➡ Test on the MC sample in the same mjj region

  - A different luminosity of 3 ab^{-1}
  - Fit function: $f(x) = p_1(1-x)^{p_2}(x)^{p_3+p_4\ln x+p_5(\ln x)^2}$

✦ Significant deviations

$$\frac{d\sigma}{dm_{jj}} \sim \sigma(x1, x2)$$

Fit 1 TeV - 7 TeV



Background modeling

arXiv:1710.09484

- For searches, how to model the background critical!
  - Due to computational resources, large scale MC production is limited
  - Data driven methods are becoming more and more important

- Similarly, a fit function may not be able to describe the data in a different mass region with much higher statistics
  - eg. dijet resonance search (2015 data) with mjj > 1.1TeV, |y*|<0.6

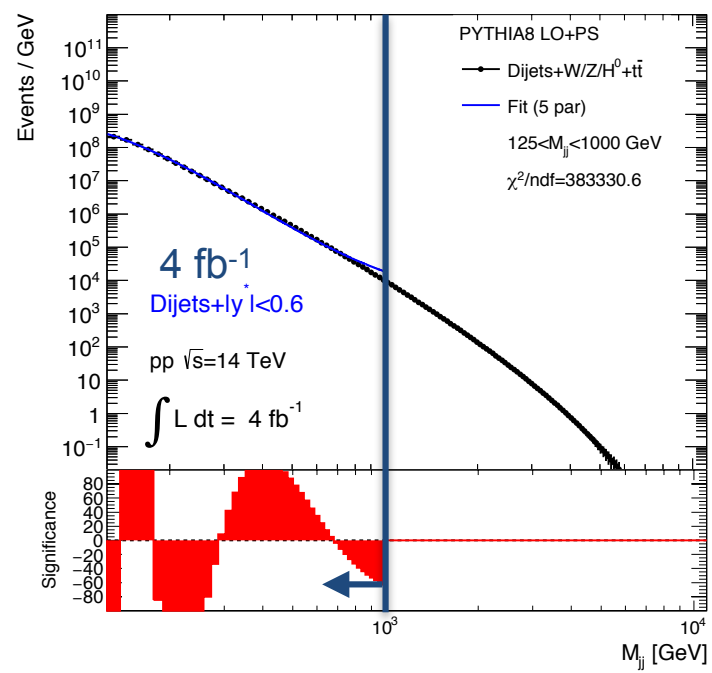
➡ Test on the MC sample in a different mjj region (125 GeV - 1 TeV)

  - Same luminosity of 4 fb^{-1}
  - Fit function: $f(x) = p_1(1-x)^{p_2}(x)^{p_3+p_4\ln x+p_5(\ln x)^2}$

✦ Significant deviations

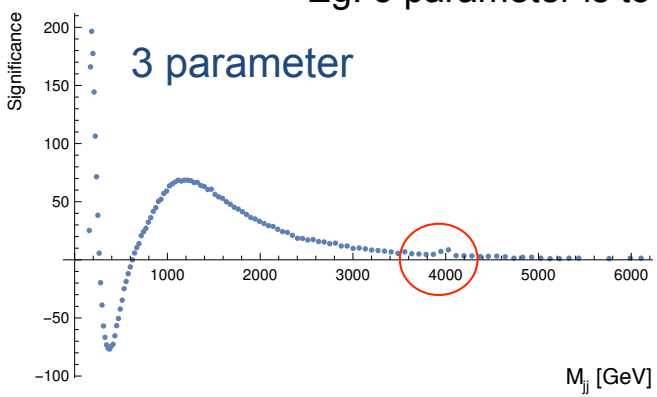
$$\frac{d\sigma}{dm_{jj}} \sim \sigma(x1, x2)$$

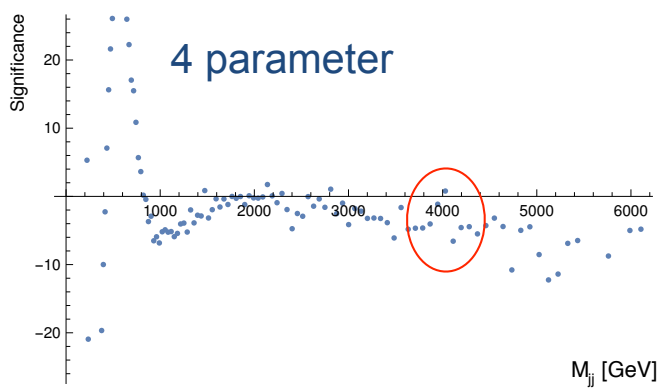
Fit 125 GeV - 1 TeV

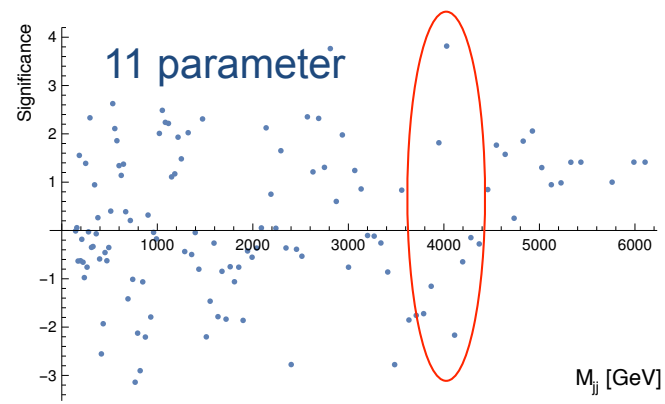


Data driven fit

- First we keep exploring the analytic fit over the mass region of mjj > 125GeV
  - We scale the b-tagged MC sample to 16 fb^{-1} then inject random poison fluctuation to make it data-like
  - Inject a 5σ Gaussian with M_G = 4TeV and σ_G = 50GeV
  - Extend the previous function using Tchebyshev polynomial $f_n(x) = e^{p1+p2*\log(1-x)+\sum_{i=3}^{n} p_i T_{i-2}[\log(x)]}$, where T_k is the kth Tchebyshev polynomial
  - Fit with different number of parameters
    - Eg. 3 parameter is to set n = 3



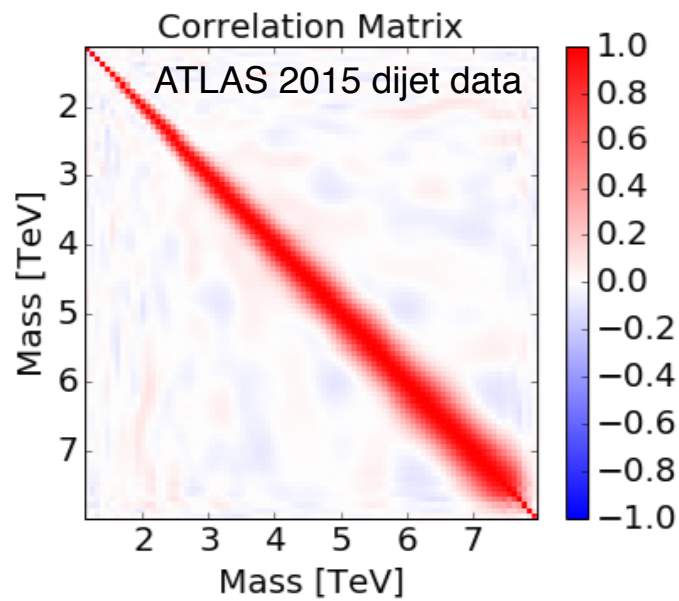




- As the number of parameters increases, the fit function reproduces the background more closely
- If too many parameters are used, the fit function reproduces the signal, as well as the background: There is no signal in the residuals
- The number of parameters in the fit function has to be tuned according to the signal width

Data driven fit

- One the other hand, many smoothing techniques are available on the market
- They can be adopted to create data driven background fit
  - Fit small window
  - Using FFT
  - Gaussian process (arXiv:1709.0568)
    - Captures correlations of data points using a covariance kernel
    - We are exploring this with the MC
  - …



Correlation between pairs of mass bins from the GP fit

Summary

- Production of Large scale MC sample for data emulation at the HL-LHC and HE-LHC
  - Larger statistics than expected for the planned HL-LHC integrated luminosity ($3ab^{-1}$)
  - Generated with 10 M CPU hours (10 days) on supercomputers at NERSC

- Studied the potential of precision searches in dijet invariant masses at the HL-LHC and HE-LHC
  - Exclusion limit at 100 fb-1 and 3 ab-1 using Gaussian
  - Compared the exclusion limit with different selections

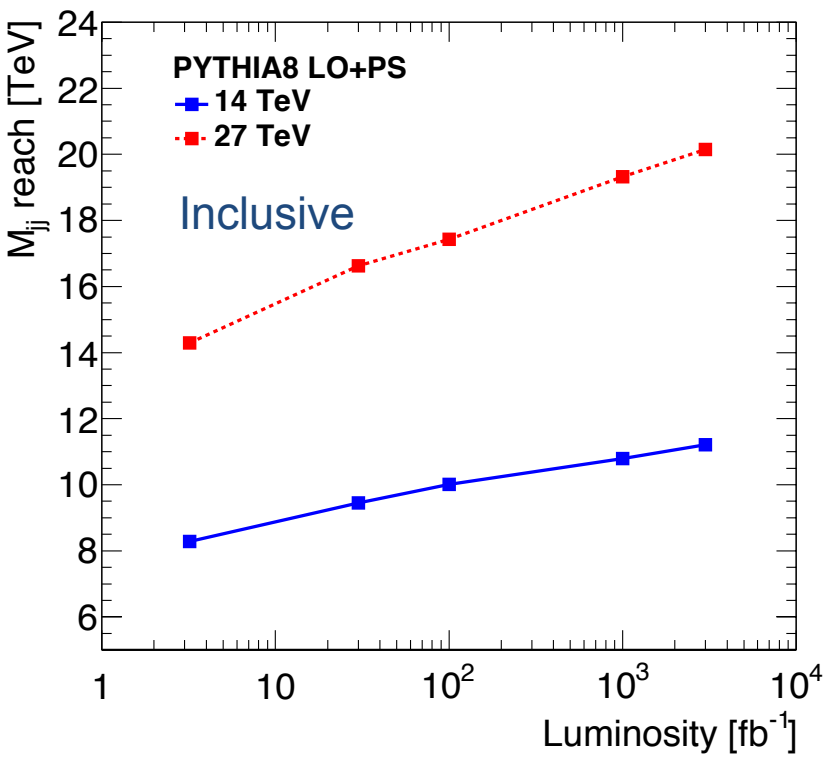
- Exploring the background modeling using data driven fit
  - Data complexity increases with statistics
  - Analytic function / new smoothing method

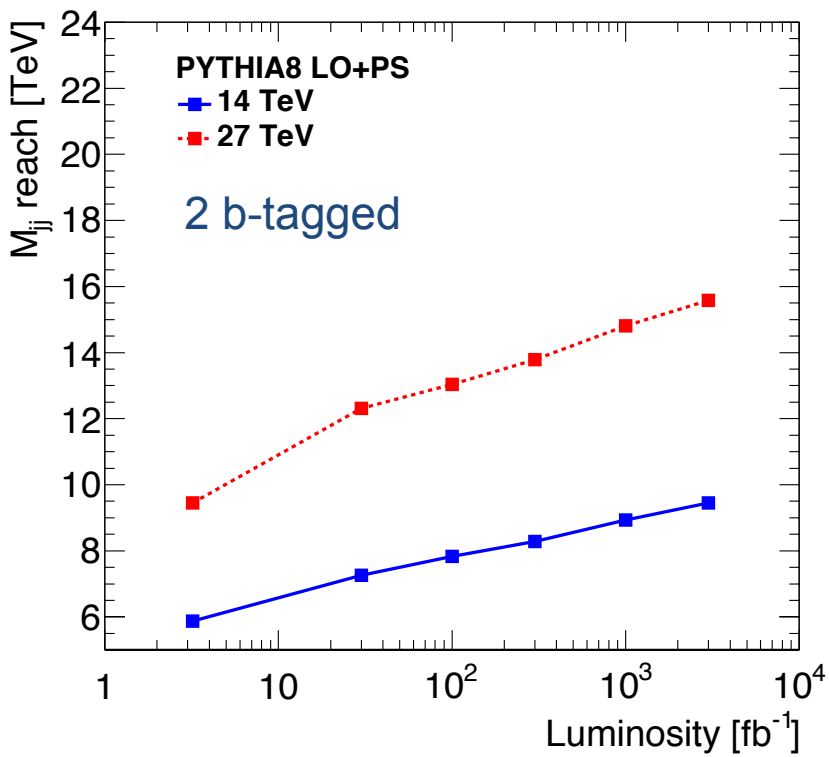
Backup

Mass reach

arXiv:1710.09484

- LO prediction of the mass reach of the dijet spectrum
  - Defined as the first bin with error >= 100%
  - NLO correction should be considered for realistic prediction





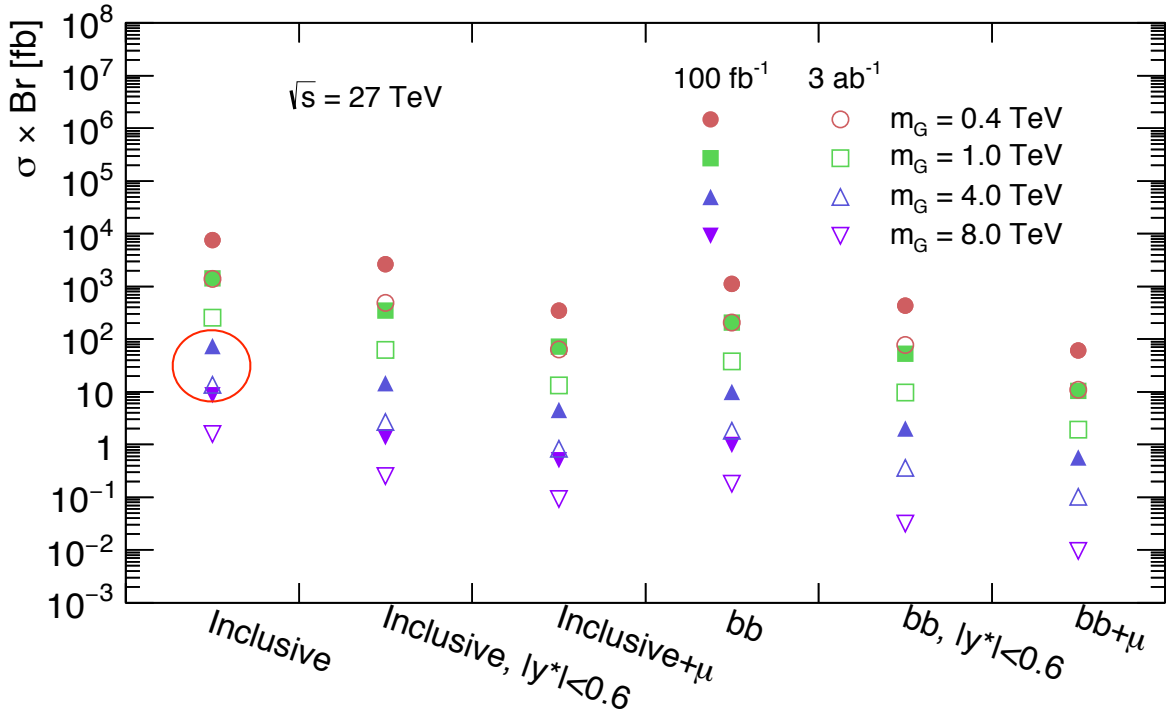
Gaussian limit

arXiv:1710.09484

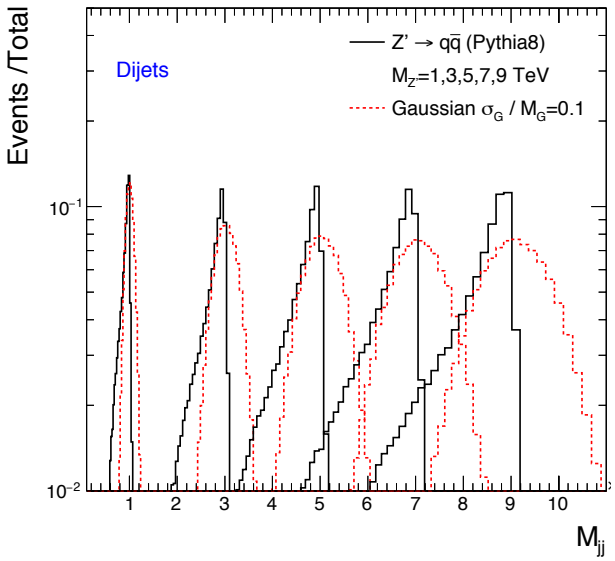
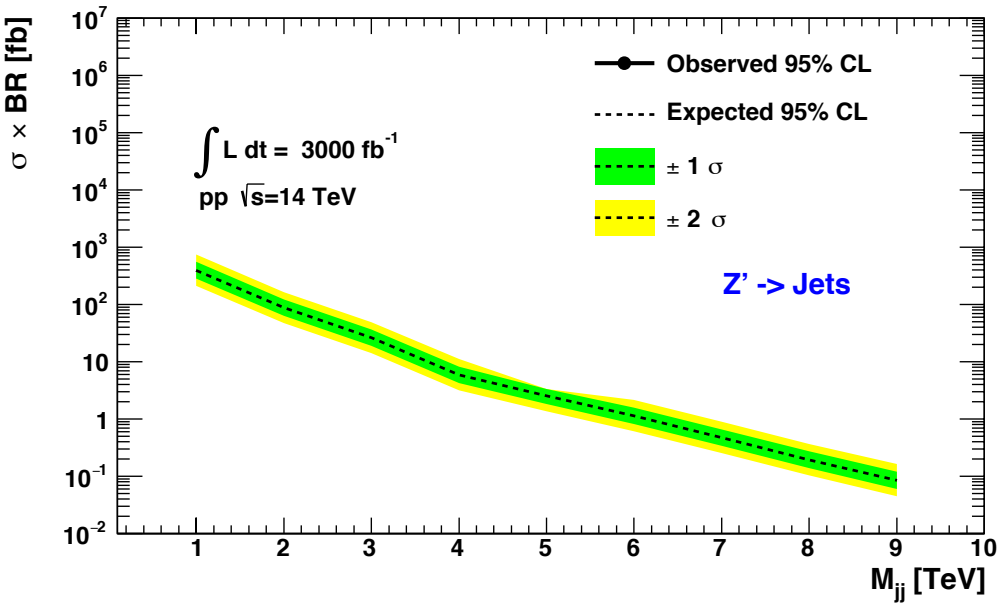
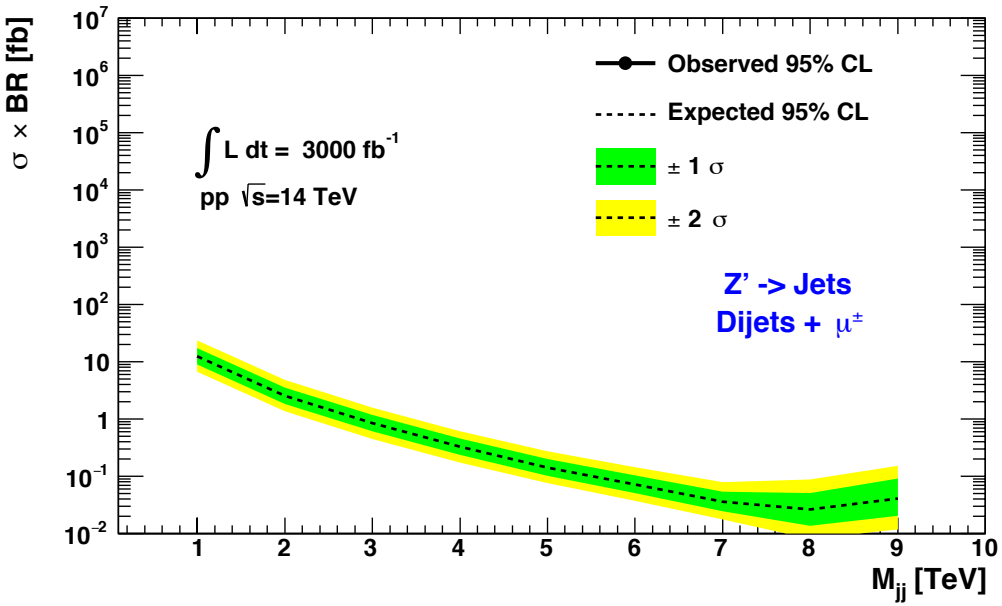
- Most of the resonance peak after reconstruction can be modeled by a Gaussian
- Gaussians with different width can provide us a guide on the exclusion limit of possible signals
  - We use Gaussians with width σ_G / M_G = 10% as an example
  - Compared the exclusion limit with different selections
    - Inclusive, |y*|<0.6, inclusive+muon, 2 b-tag, 2 b-tag+|y*|<0.6 and 2 b-tag+muon

When luminosity Increases from 100 fb^{-1} to 3 ab^{-1}, it gives ~10 times better exclusive limit on the cross section * Br

(eg. blue filled triangle vs blue open triangle)



MC limit

arXiv:1710.09484

- Taken Z' shape which have a large tail on the left