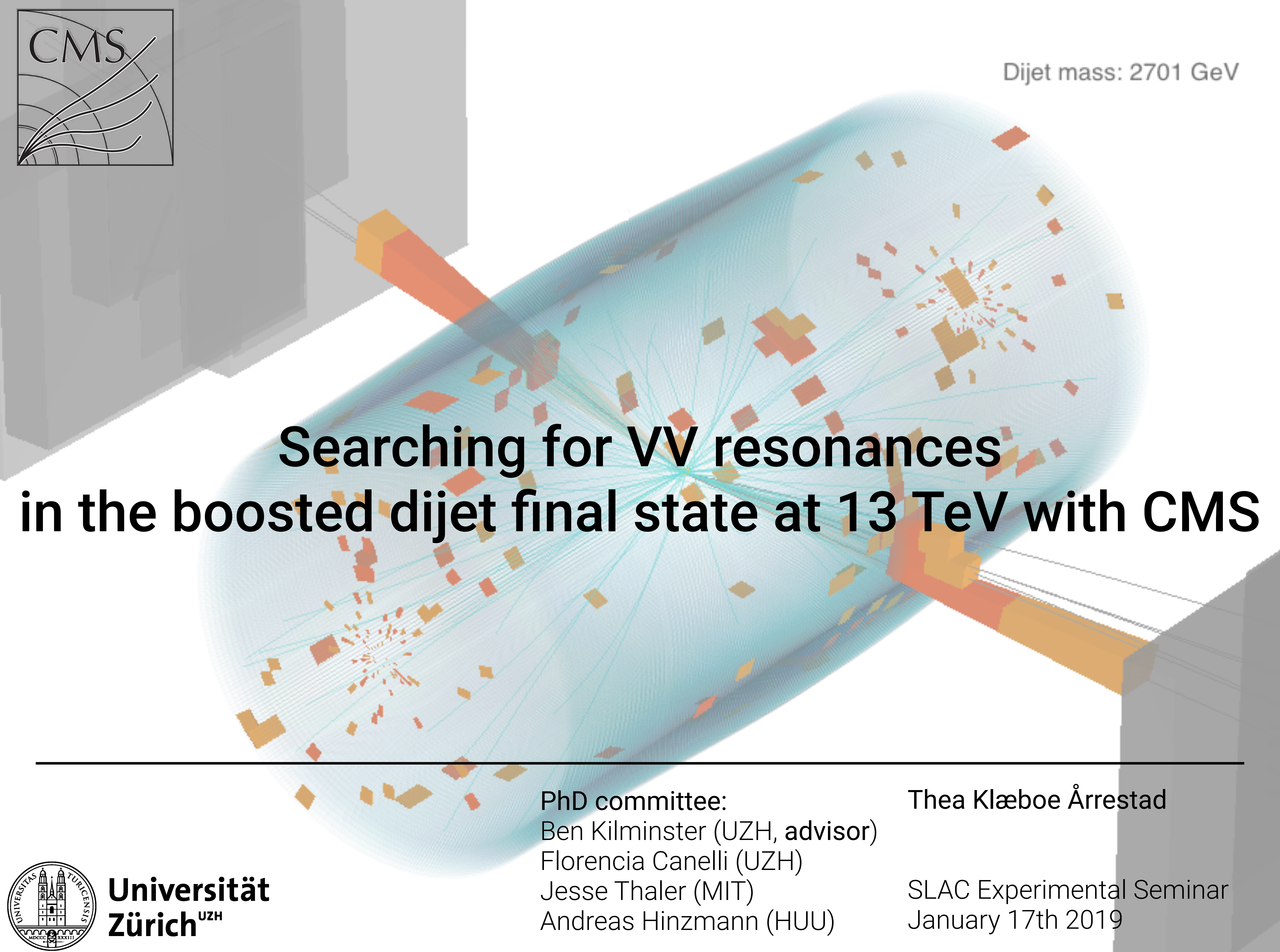


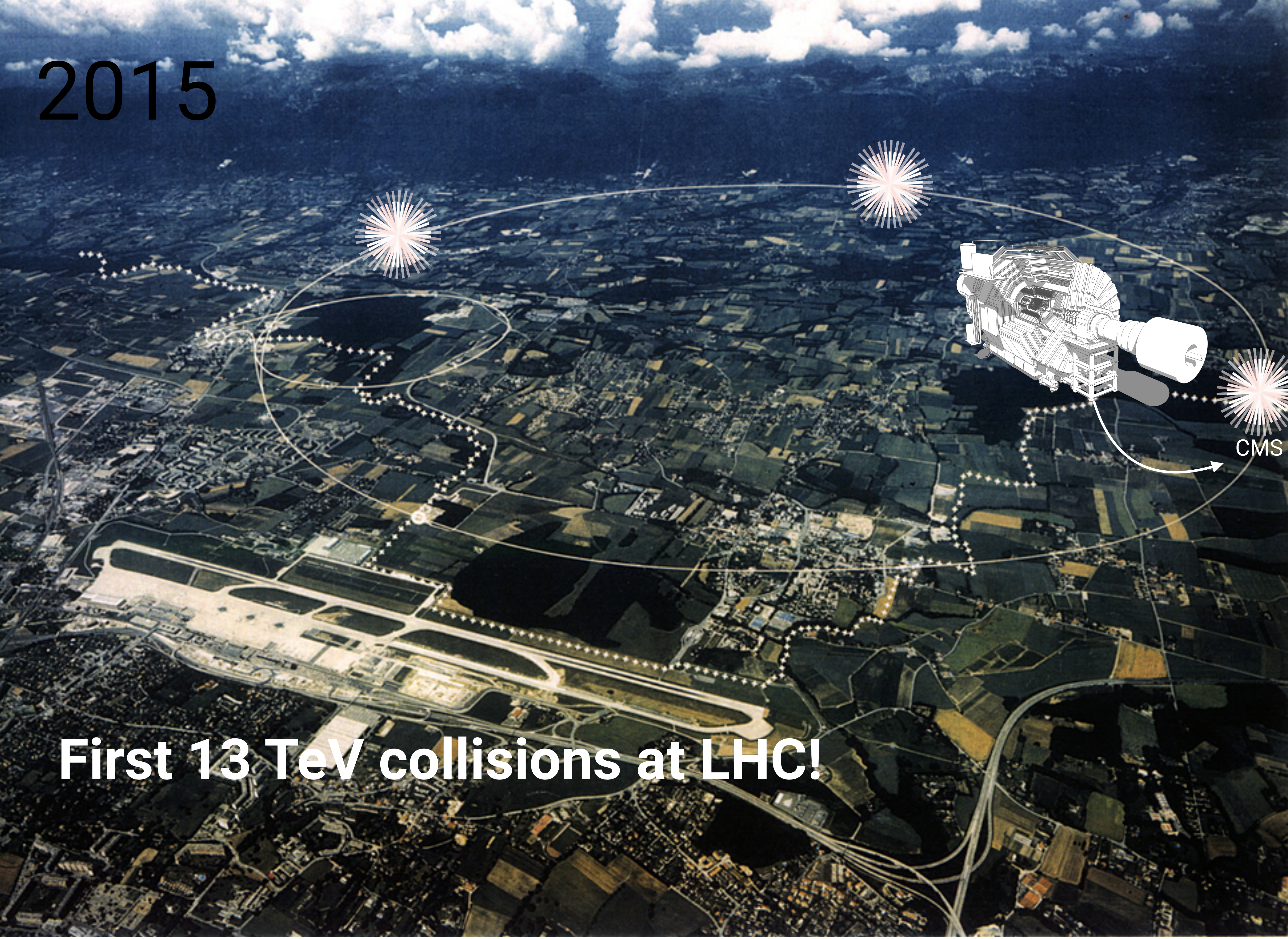
Searching for VV resonances in the boosted dijet final state at 13 TeV with CMS

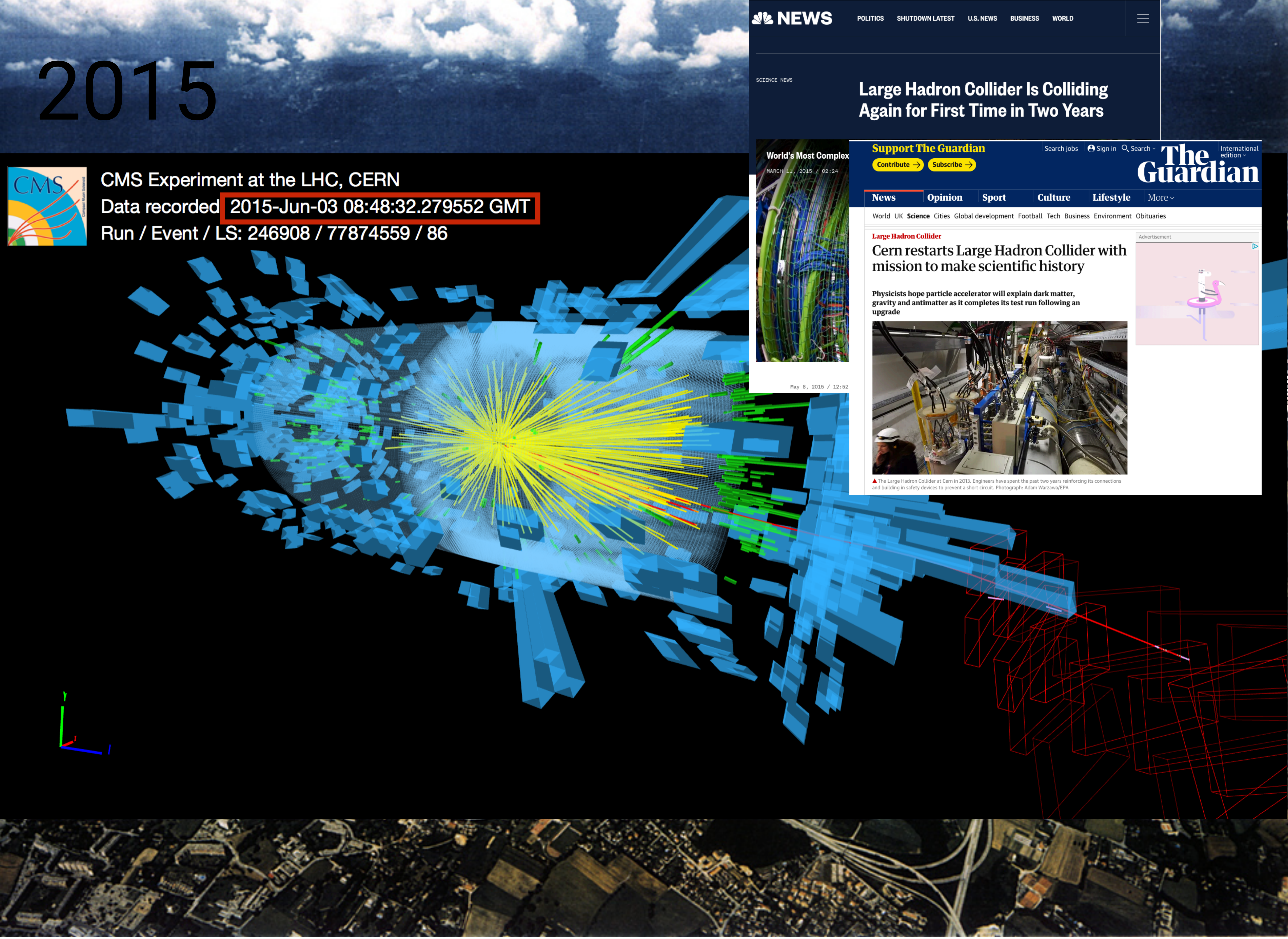
PhD committee:
Ben Kilminster (UZH, **advisor**)
Florencia Canelli (UZH)
Jesse Thaler (MIT)
Andreas Hinzmann (HUU)

Thea Klæboe Årrestad

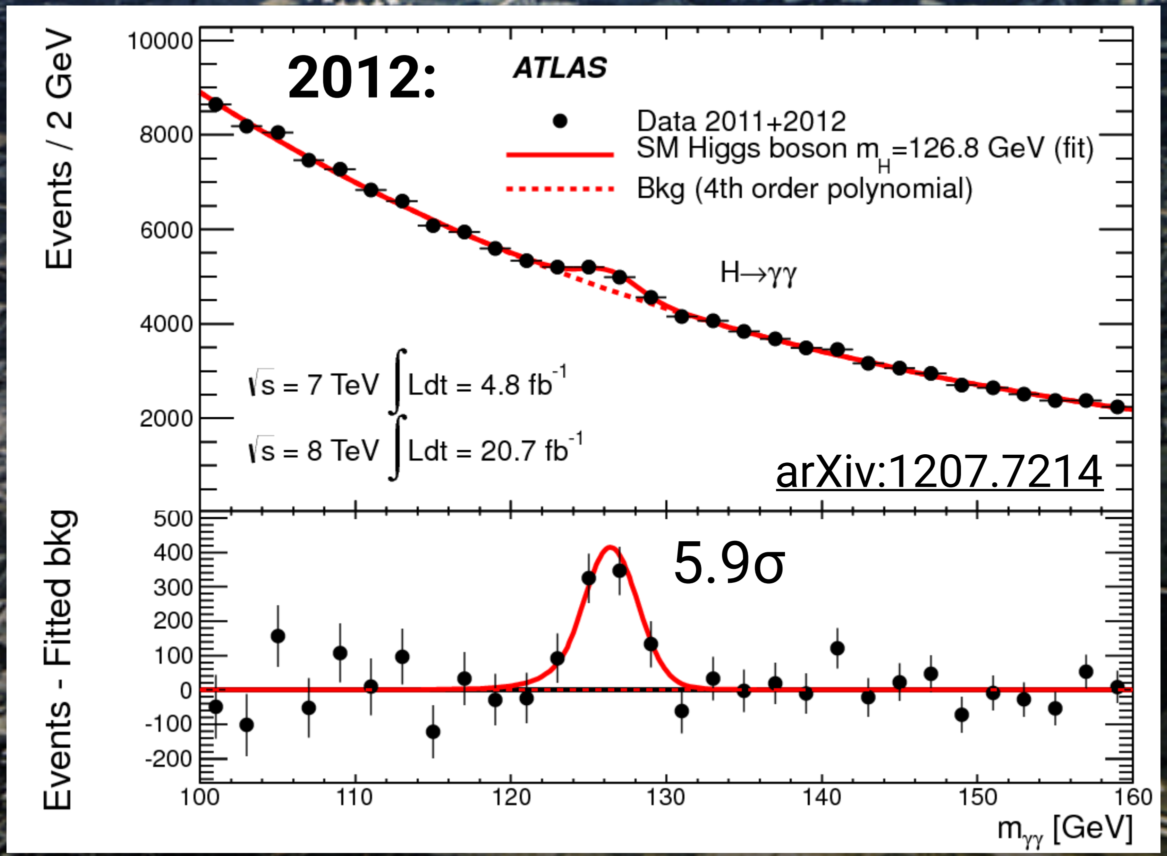
Universität Zürich[UZH]

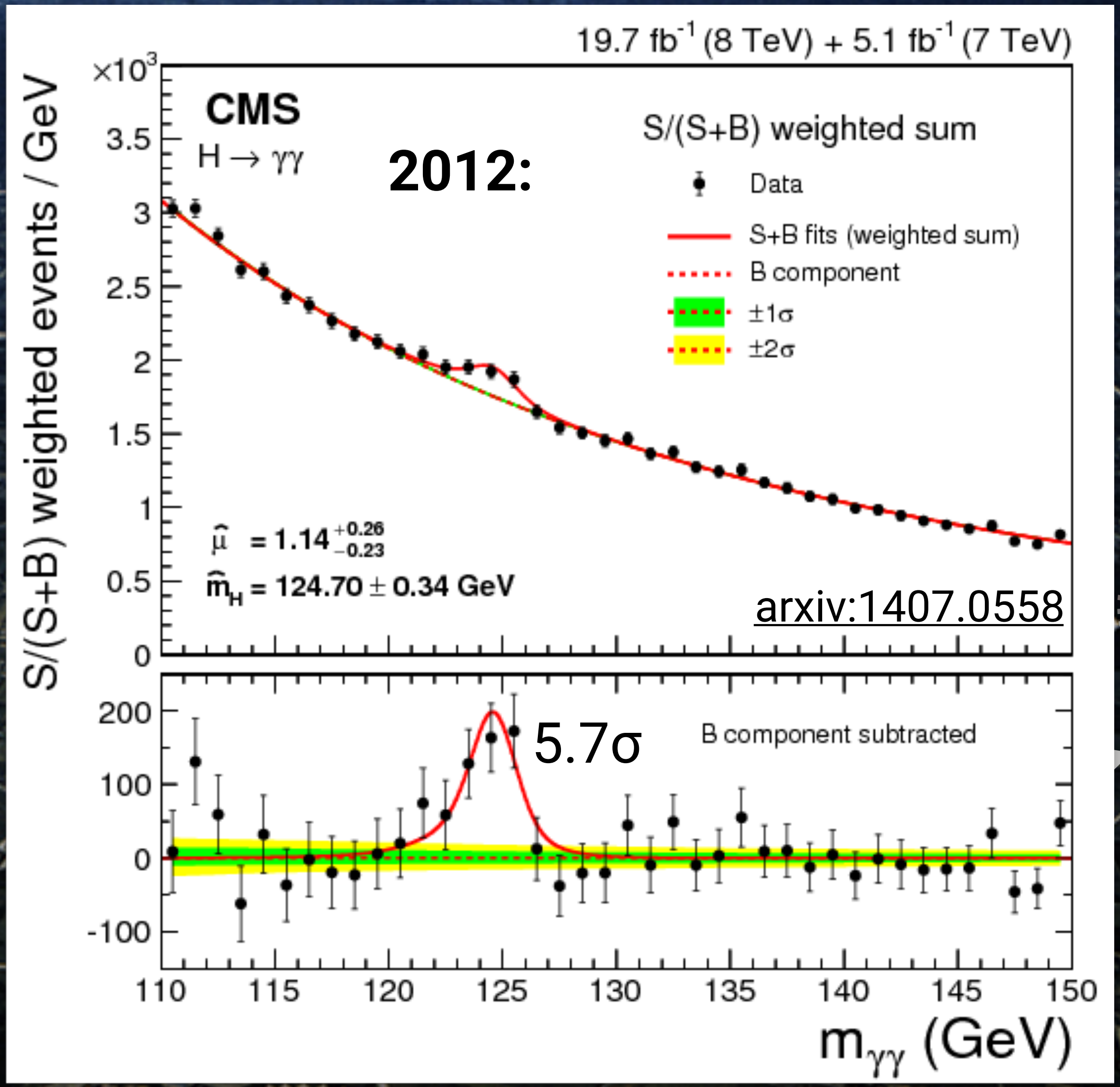
SLAC Experimental Seminar
January 17th 2019


2015
CMS
First 13 TeV collisions at LHC!


2015
CMS Experiment at the LHC, CERN
Data recorded: 2015-Jun-03 08:48:32.279552 GMT
Run / Event / LS: 246908 / 77874559 / 86
NEWS
POLITICS
SHUTDOWN LATEST
U.S. NEWS
BUSINESS
WORLD
SCIENCE NEWS
Large Hadron Collider Is Colliding Again for First Time in Two Years
World's Most Complex
MARCH 11, 2015 / 02:24
May 6, 2015 / 12:52
Support The Guardian
Contribute
Subscribe
Search jobs
Sign in
Search
International edition
The Guardian
News
Opinion
Sport
Culture
Lifestyle
More
World UK Science Cities Global development Football Tech Business Environment Obituaries
Large Hadron Collider
Cern restarts Large Hadron Collider with mission to make scientific history
Physicists hope particle accelerator will explain dark matter, gravity and antimatter as it completes its test run following an upgrade
Advertisement
The Large Hadron Collider at Cern in 2013. Engineers have spent the past two years reinforcing its connections and building in safety devices to prevent a short circuit. Photograph: Adam Warzawa/EPA

2015


2012:
ATLAS
Data 2011+2012
SM Higgs boson m_H=126.8 GeV (fit)
Bkg (4th order polynomial)
H→γγ
$\sqrt{s}$ = 7 TeV $\int$Ldt = 4.8 fb^{-1}
$\sqrt{s}$ = 8 TeV $\int$Ldt = 20.7 fb^{-1}
arXiv:1207.7214
5.9σ
Events / 2 GeV
Events - Fitted bkg
$m_{\gamma\gamma}$ [GeV]


19.7 fb^{-1} (8 TeV) + 5.1 fb^{-1} (7 TeV)
CMS
H → γγ
2012:
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
±1σ
±2σ
$\hat{\mu}$ = 1.14$^{+0.26}_{-0.23}$
$\hat{m}_H$ = 124.70 ± 0.34 GeV
arxiv:1407.0558
5.7σ
B component subtracted
S/(S+B) weighted events / GeV
$m_{\gamma\gamma}$ (GeV)

arXiv.org > hep-ex > arXiv:1506.00962

Search or Art

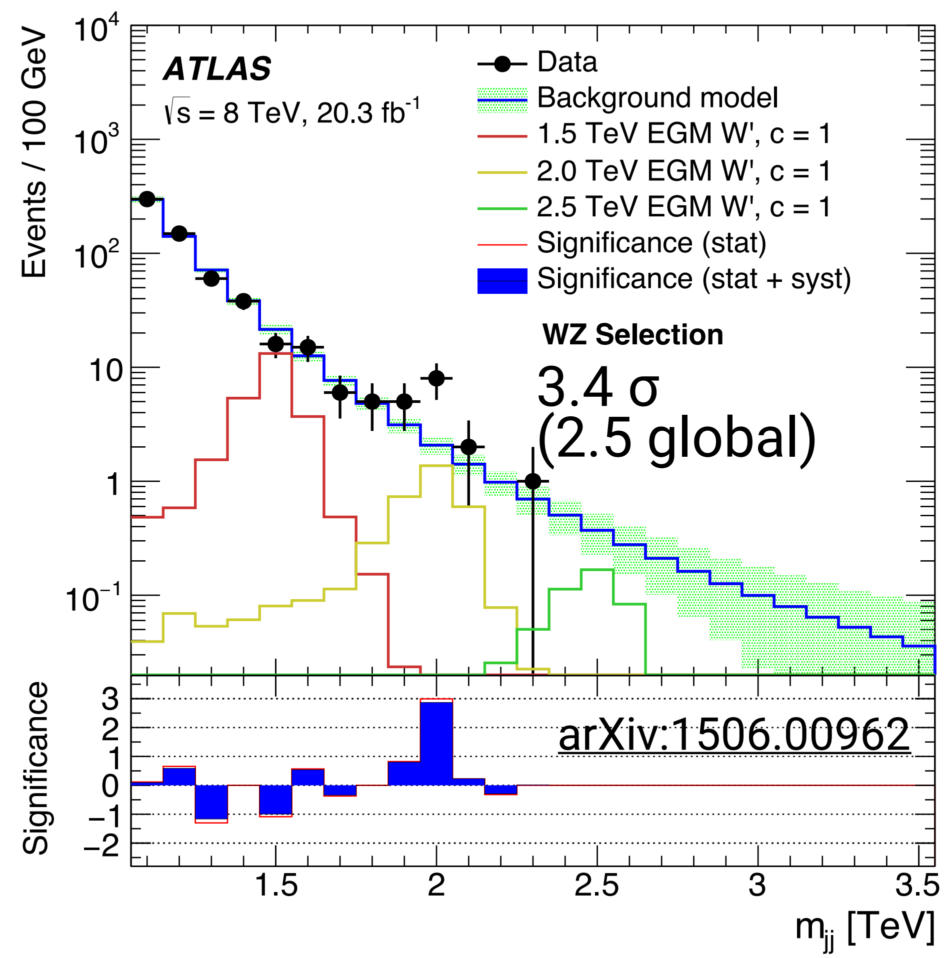
(Help | Advanced

High Energy Physics – Experiment

Search for high-mass diboson resonances with boson-tagged jets in proton–proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector

ATLAS Collaboration

(Submitted on 2 Jun 2015 (v1), last revised 22 Jan 2016 (this version, v3))



arXiv.org > hep-ex > arXiv:1506.00962

Search or Art

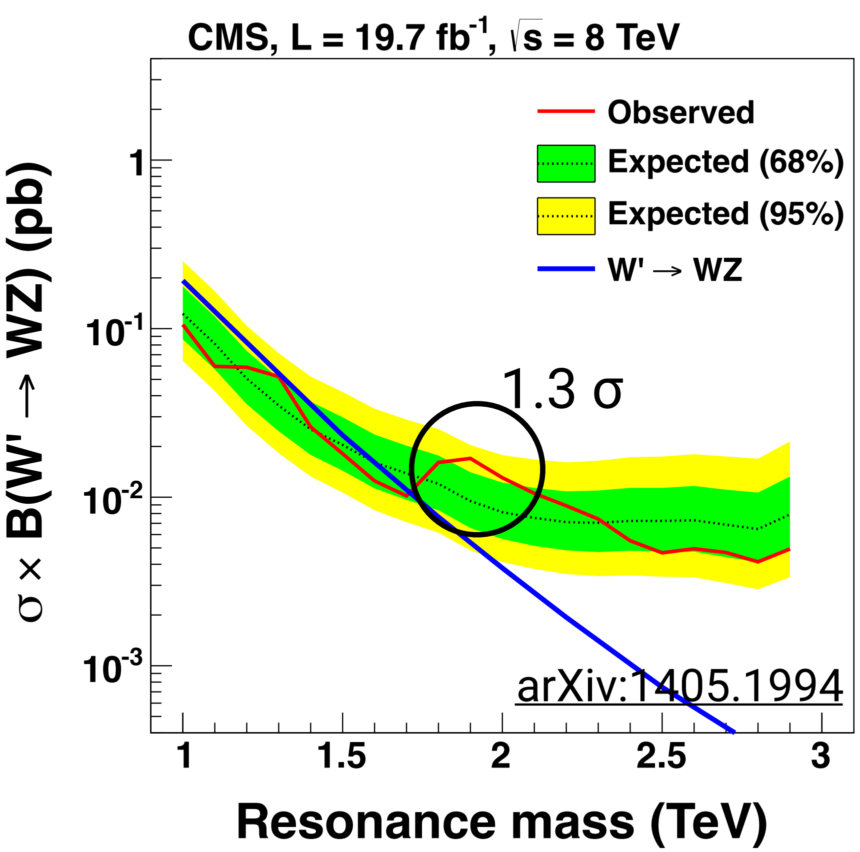
(Help | Advanced

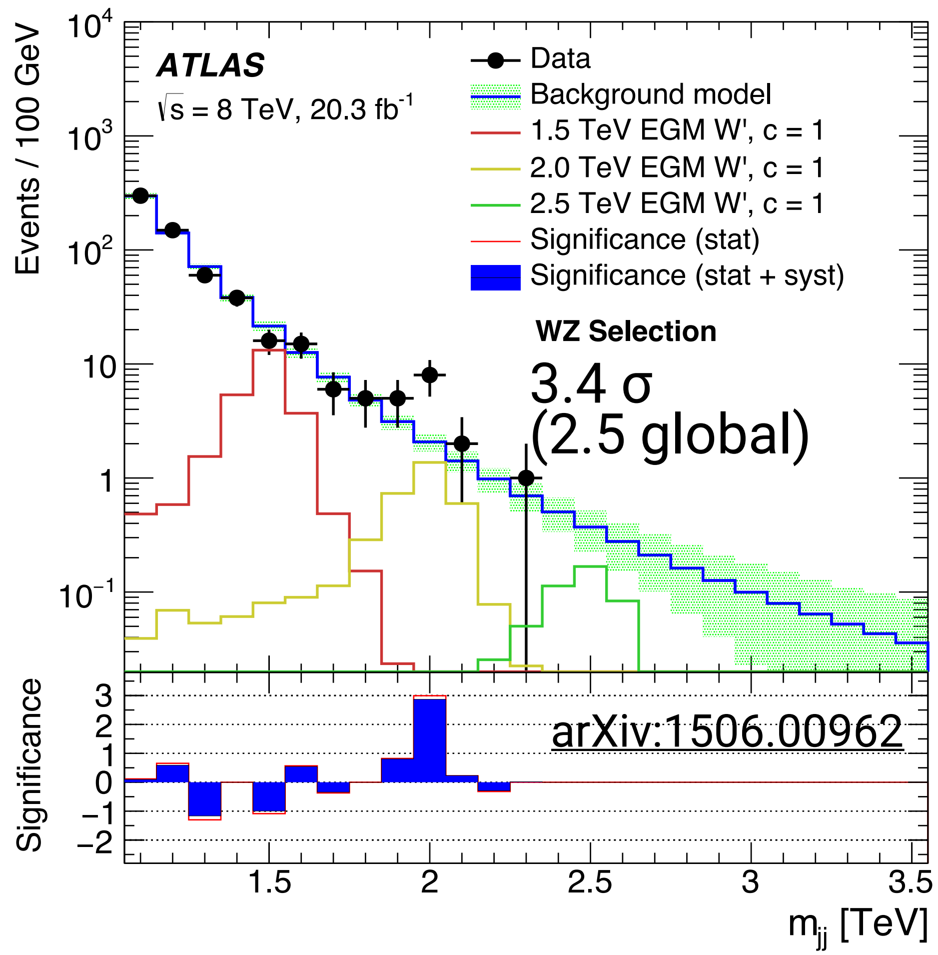
High Energy Physics – Experiment

Search for high-mass diboson resonances with boson-tagged jets in proton–proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector

ATLAS Collaboration

(Submitted on 2 Jun 2015 (v1), last revised 22 Jan 2016 (this version, v3))





arXiv.org > hep-ex > arXiv:1506.00962

Search or Art

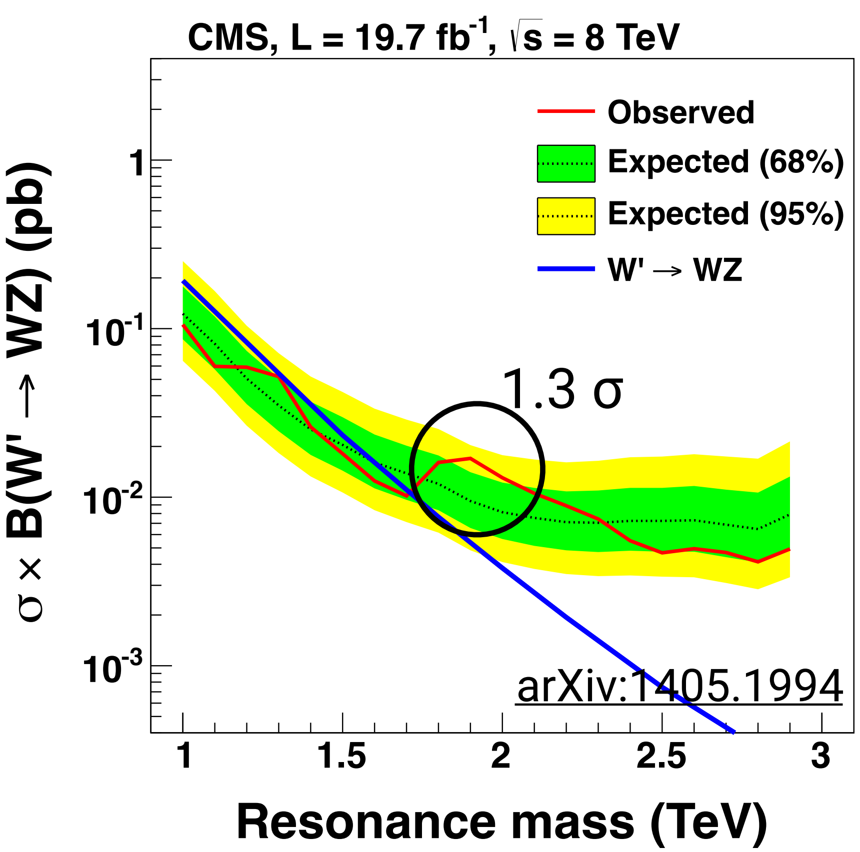
(Help | Advanced

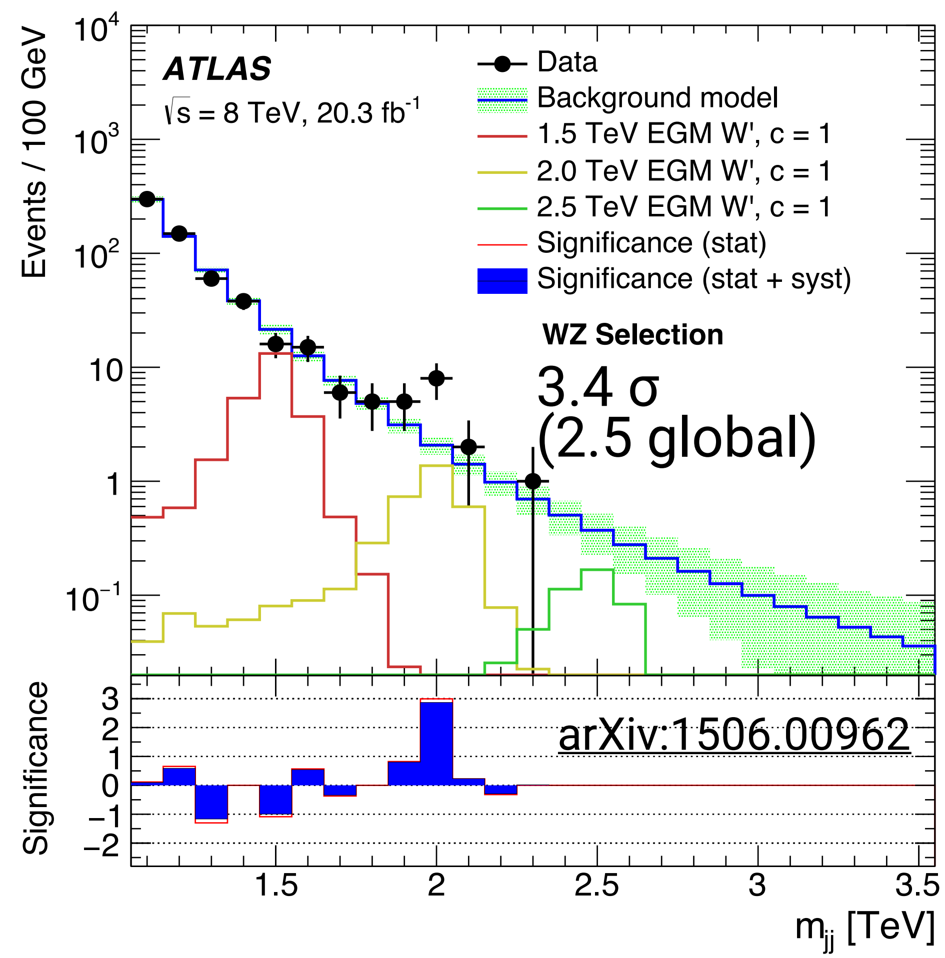
High Energy Physics – Experiment

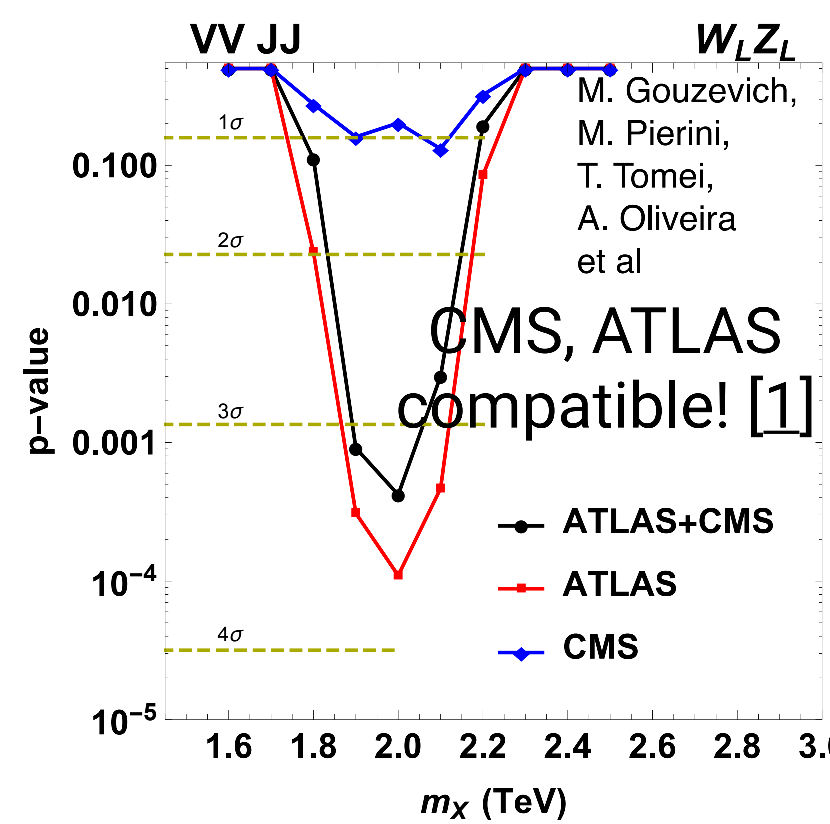
Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector

ATLAS Collaboration

(Submitted on 2 Jun 2015 (v1), last revised 22 Jan 2016 (this version, v3))







What could it be?


1. arXiv:1506.06739 [pdf, ps, other] hep-ph hep-ex
Triboson interpretations of the ATLAS diboson excess
Authors: J. A. Aguilar-Saavedra
Submitted 25 September, 2015; v1 submitted 22 June, 2015; originally announced Ju

2. arXiv:1506.07511 [pdf, other] hep-ph hep-ex doi 10.1103/Phy
G221 Interpretations of the Diboson and Wh Excesses

3. arXiv:1507.00268 [pdf, other] hep-ph doi 10.1103/PhysRevD.92.09
Simple non-Abelian extensions of the standard model ga
Authors: Qing-Hong Cao, Bin Yan, Dong-Ming Zhang
Submitted 27 November, 2015; v1 submitted 1 July, 2015; originally announced July

4. arXiv:1507.01681 [pdf, other] hep-ph hep-ex doi 10.1007/JHE
Prospects for Spin-1 Resonance Search at 13 TeV LHC an
Authors: Tomohiro Abe, Teppei Kitahara, Mihoko M. Nojiri
Submitted 22 January, 2016; v1 submitted 7 July, 2015; originally announced July 201
Comments: 38 pages, 19 figures, 1 table; minor changes, references added, version p

5. arXiv:1507.03098 [pdf, other] hep-ph hep-ex doi 10.1103/PhysRevLett.115.171802
A scalar hint from the diboson excess?
Authors: Giacomo Cacciapaglia, Aldo Deandrea, Michio Hashimoto
Submitted 21 October, 2015; v1 submitted 11 July, 2015; originally announced July 2015.
Comments: 6 pages, 2 figures, 2 tables, two column. Typos in (5)-(9) are corrected and several discussions are clarified. References added. Matches published version in PRL

6. arXiv:1507.03553 [pdf, other] hep-ph doi 10.1155/2016/3279568
On the compatibility of the diboson excess with a gg-initiated composite sector
Authors: Veronica Sanz
Submitted 6 April, 2016; v1 submitted 13 July, 2015; originally announced July 2015.

7. arXiv:1507.04431 [pdf, ps, other] hep-ph hep-ex doi 10.1016/j.physletb.2015.08.028
2 TeV Higgs bo
Authors: Chuan-H
Submitted 18 August,
Comments: 12 pages,

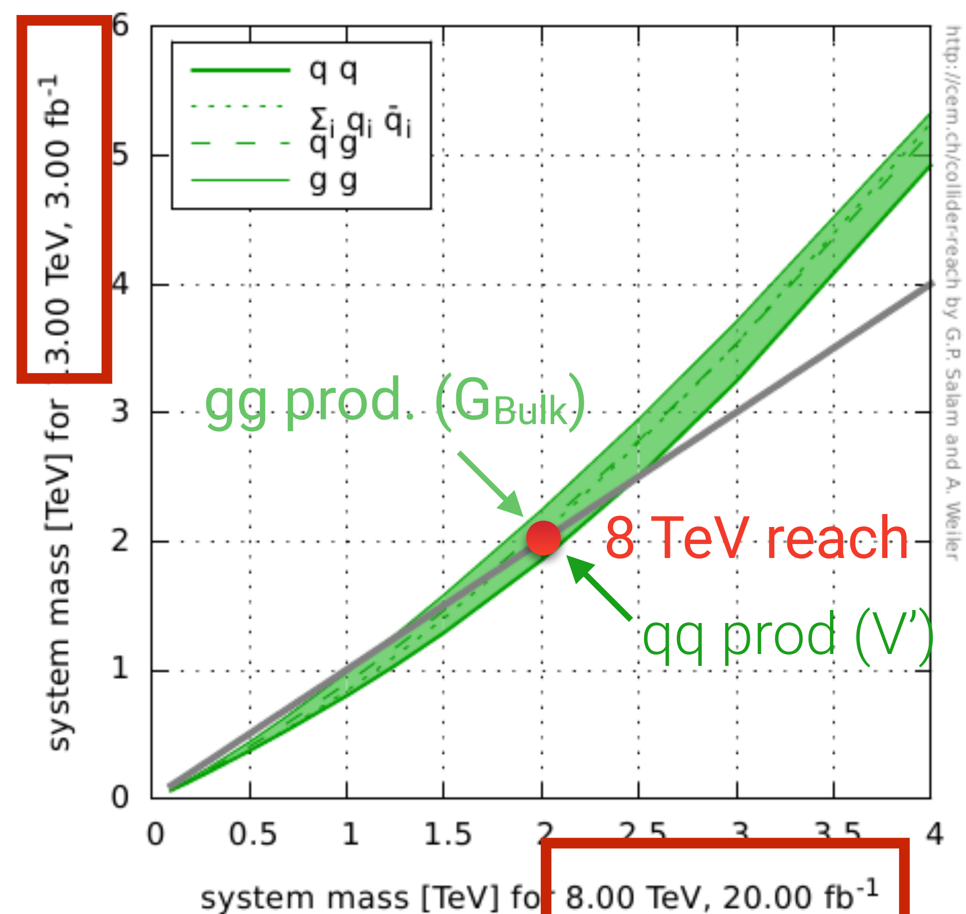
11. arXiv:1507.06018 [pdf, other] hep-ph
Low Scale Composite Higgs Model and 1.8 ∼ 2 TeV Diboson Excess
Authors: Ligong Bian, Da Liu, Jing Shu
Submitted 21 July, 2015; originally announced July 2015.

Most popular candidates

- Bulk Graviton models: Spin 2
  Main production: gluon-fusion
  $G_{Bulk}\rightarrow$WW and $G_{Bulk}\rightarrow$ZZ

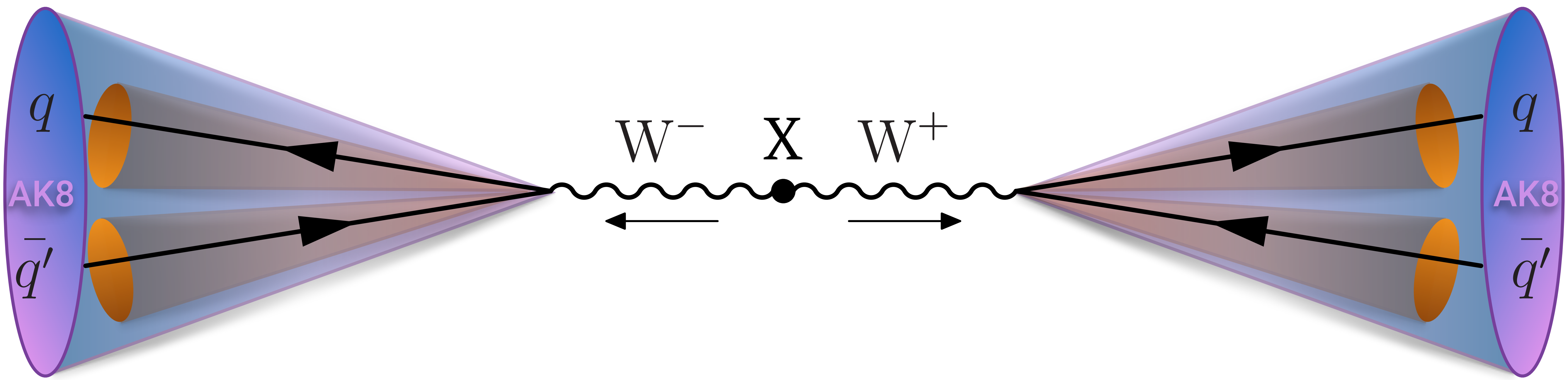
- Composite Higgs models: Spin-1
  Main production: qq-annihilation
  Z'→WW and W'→WZ

With only 3 fb^{-1} of 13 TeV data, same discovery potential as 8 TeV dataset of 20 fb^{-1}



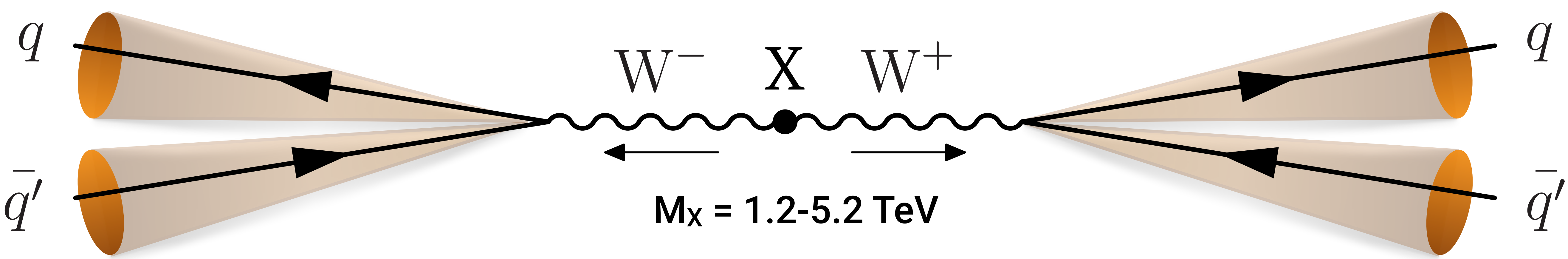
Search I:
First search for $X \rightarrow VV \rightarrow q\bar{q}q\bar{q}$ at 13 TeV with CMS

Published in Journal of High Energy Physics (2017), DOI: 10.1007/JHEP03(2017)162

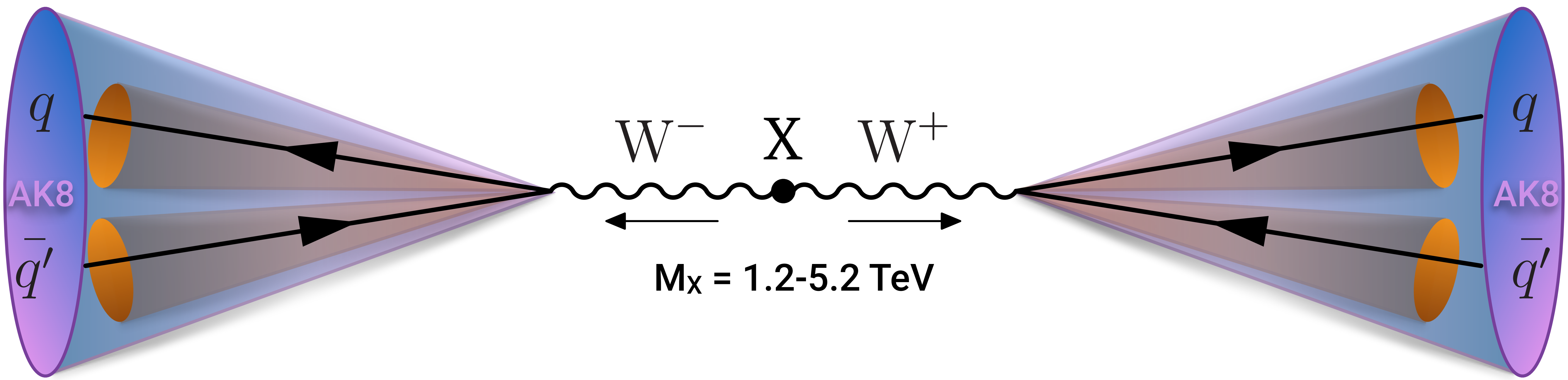


~ The first 'boosted' search with 13 TeV data and the first to use dedicated jet substructure triggers. Published with the full 2015 dataset, 2.7 fb^{-1}

$X \rightarrow VV \rightarrow q\bar{q}q\bar{q}$



$X \to VV \to q\bar{q}q\bar{q}$

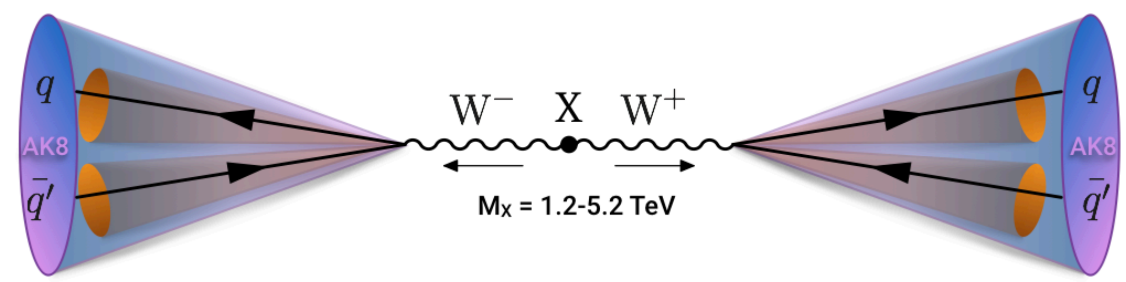


$$R \sim \frac{2m_V}{p_{T,V}}$$ → Fully contained in AK R=0.8 jet for V p_T > 200 GeV

$X \rightarrow VV \rightarrow q\bar{q}q\bar{q}$

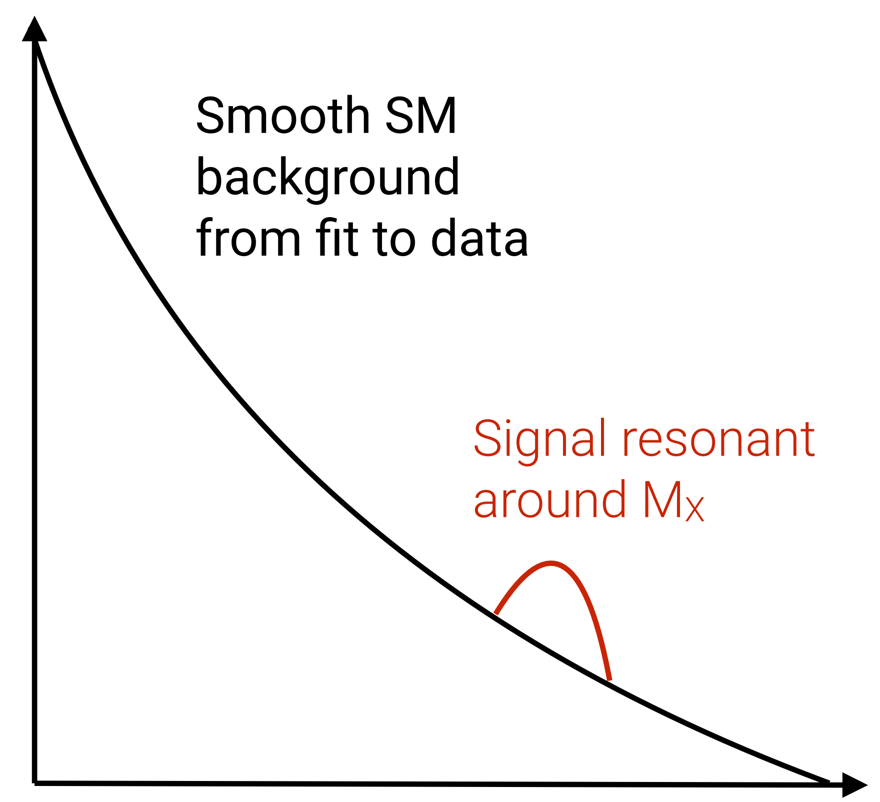
Reconstruct two hadronic W/Z bosons

- require two high-p_T jets (>200 GeV), tag using dedicated jet substructure methods



Bump hunt in dijet invariant mass spectrum

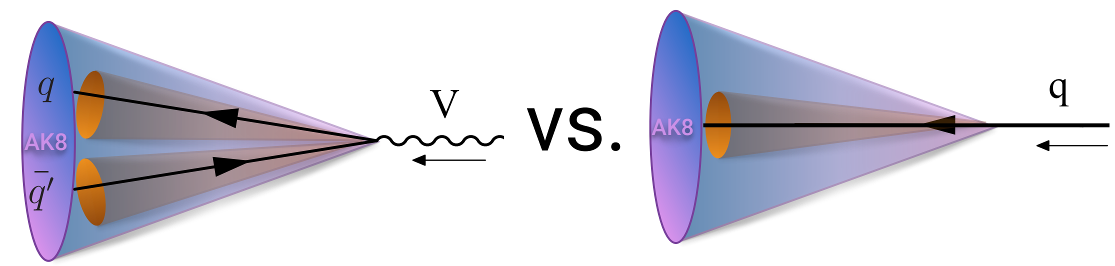
- QCD (dominant) background estimated from fit to data in signal region
- smoothness test of observed data (no MC)

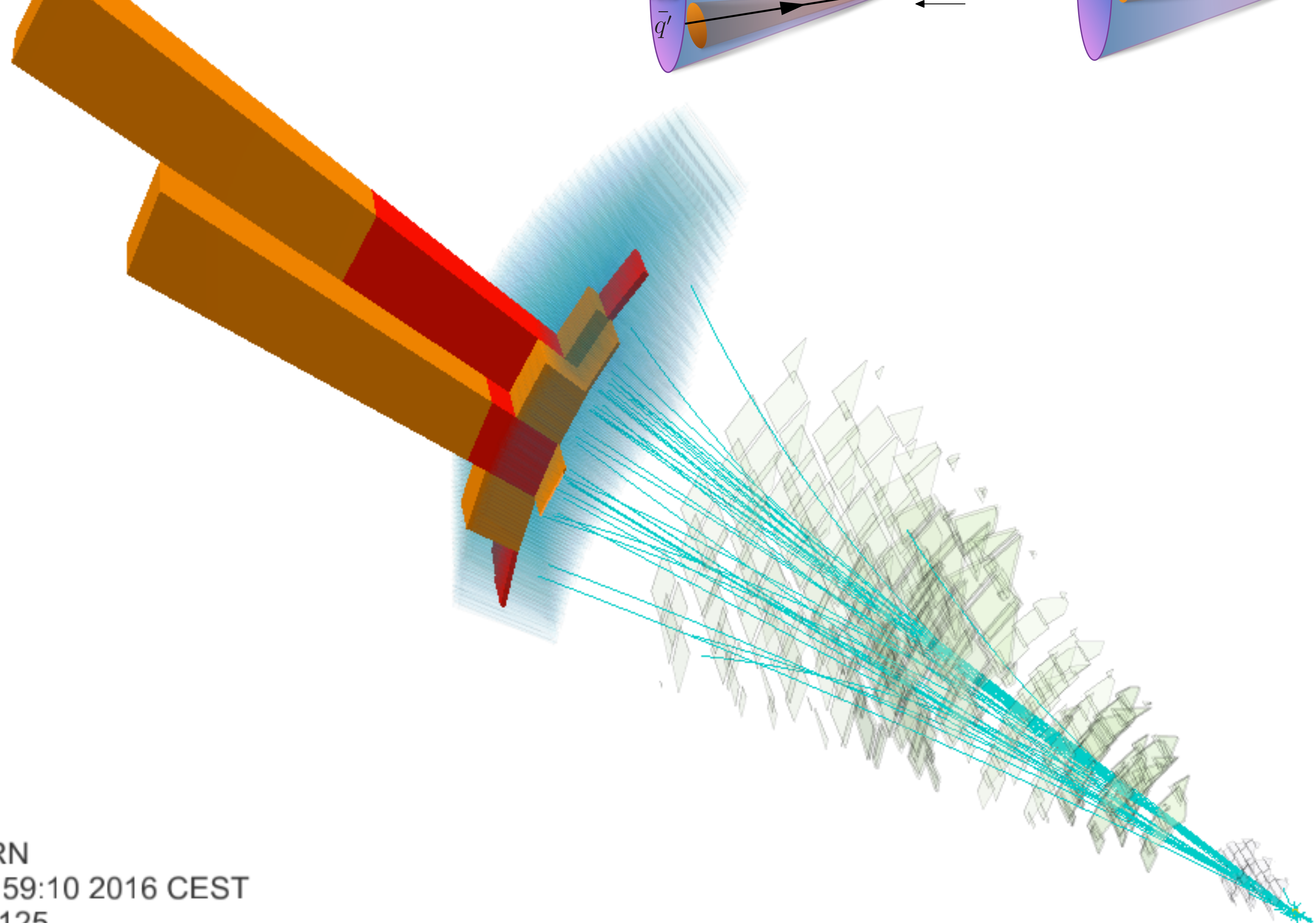


Simple and robust one-background analysis, some caveats:

1. Overwhelming QCD multijet background
2. Can only model smoothly falling m_{jj} (trigger limited)

Getting rid of QCD



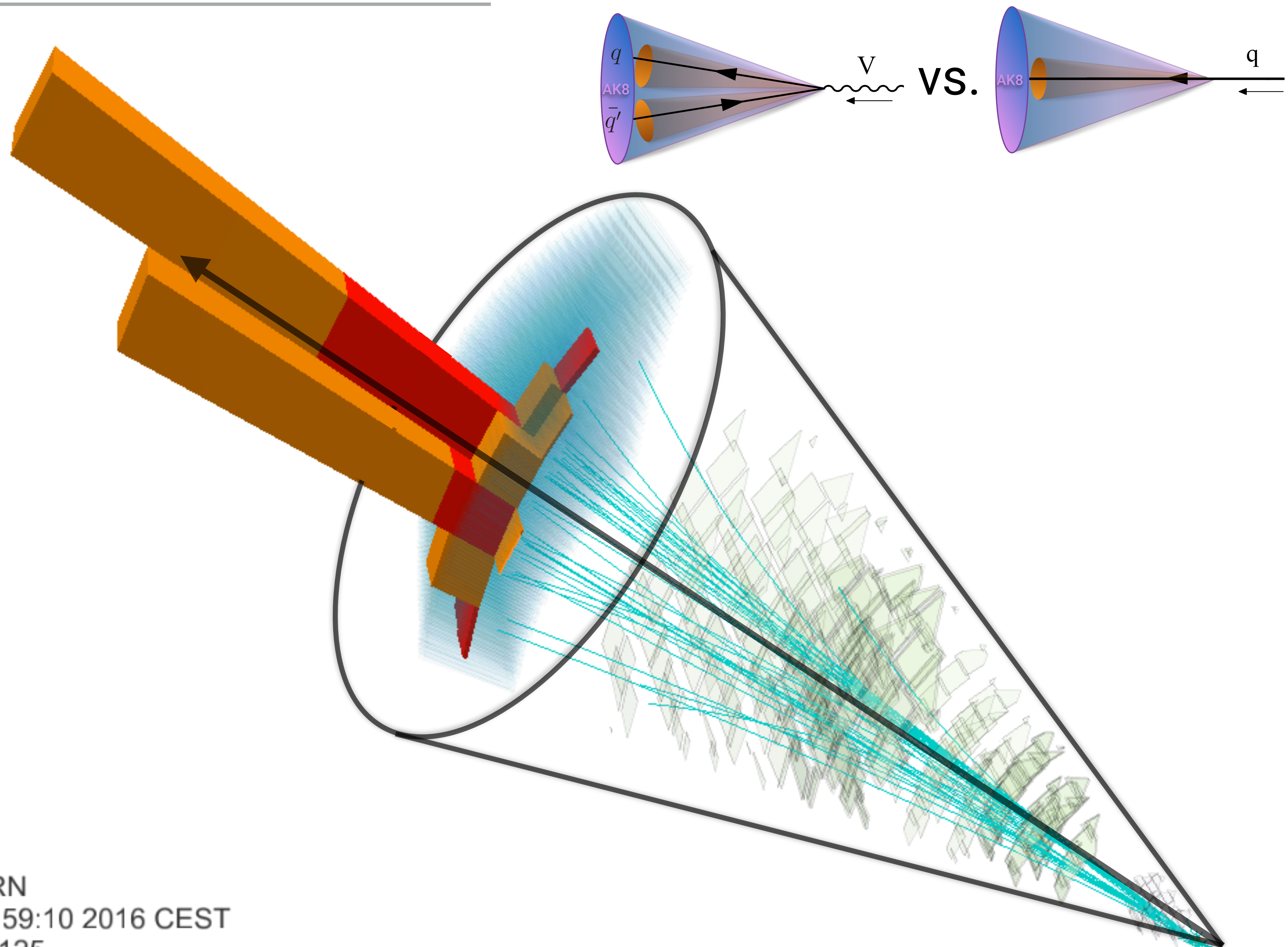


CMS-PAS-B2G-17-001

CMS Experiment at LHC, CERN
Data recorded: Mon Jul 18 19:59:10 2016 CEST
Run/Event: 276950 / 1080730125
Lumi section: 573

$V \rightarrow q\bar{q}$?

Getting rid of QCD

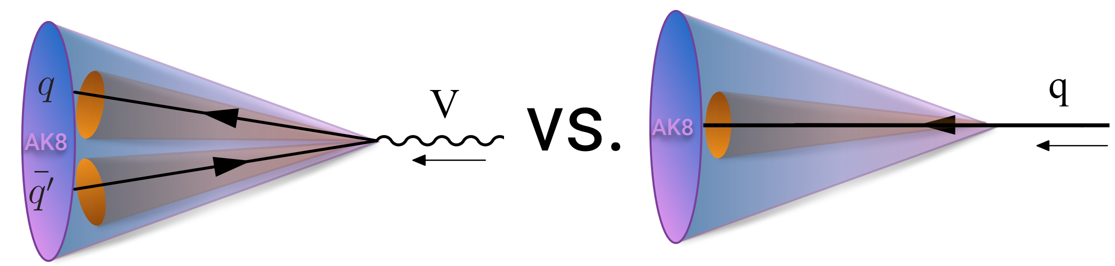


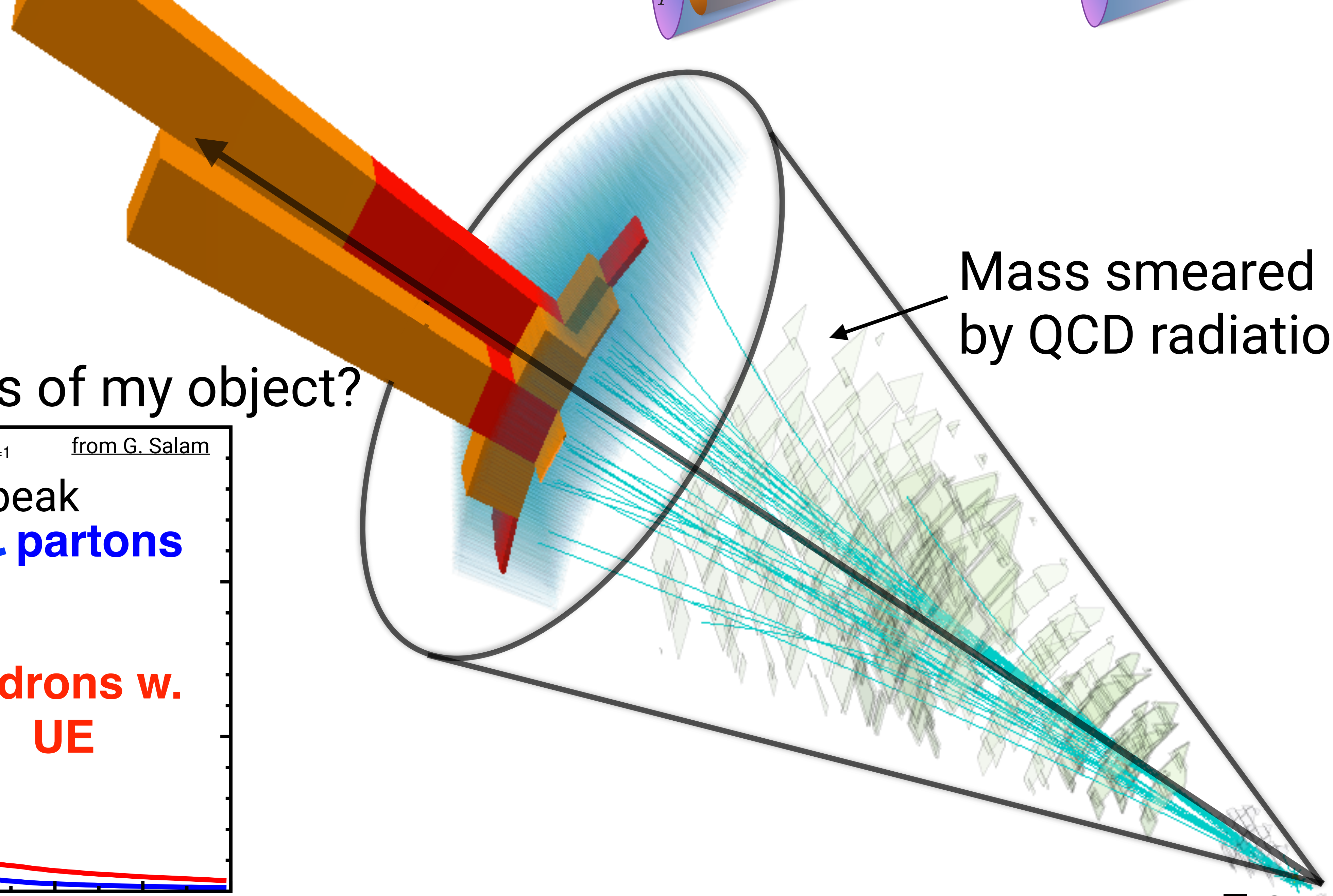
CMS-PAS-B2G-17-001

CMS Experiment at LHC, CERN
Data recorded: Mon Jul 18 19:59:10 2016 CEST
Run/Event: 276950 / 1080730125
Lumi section: 573

$V \rightarrow q\bar{q}$?

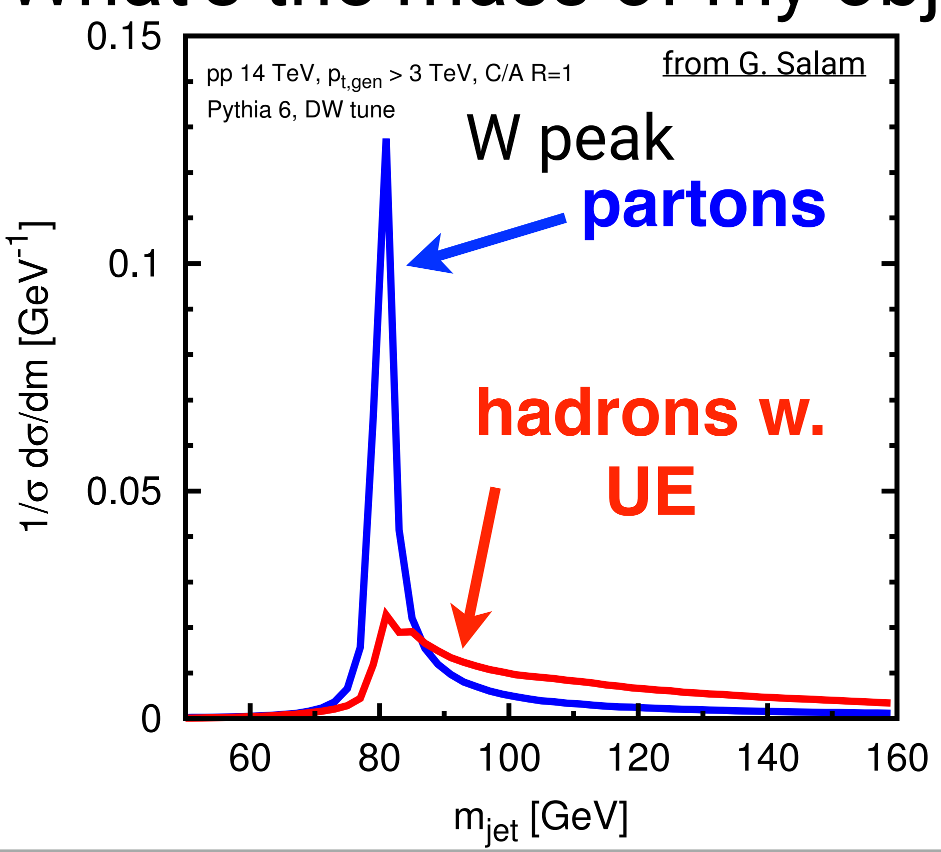
Getting rid of QCD





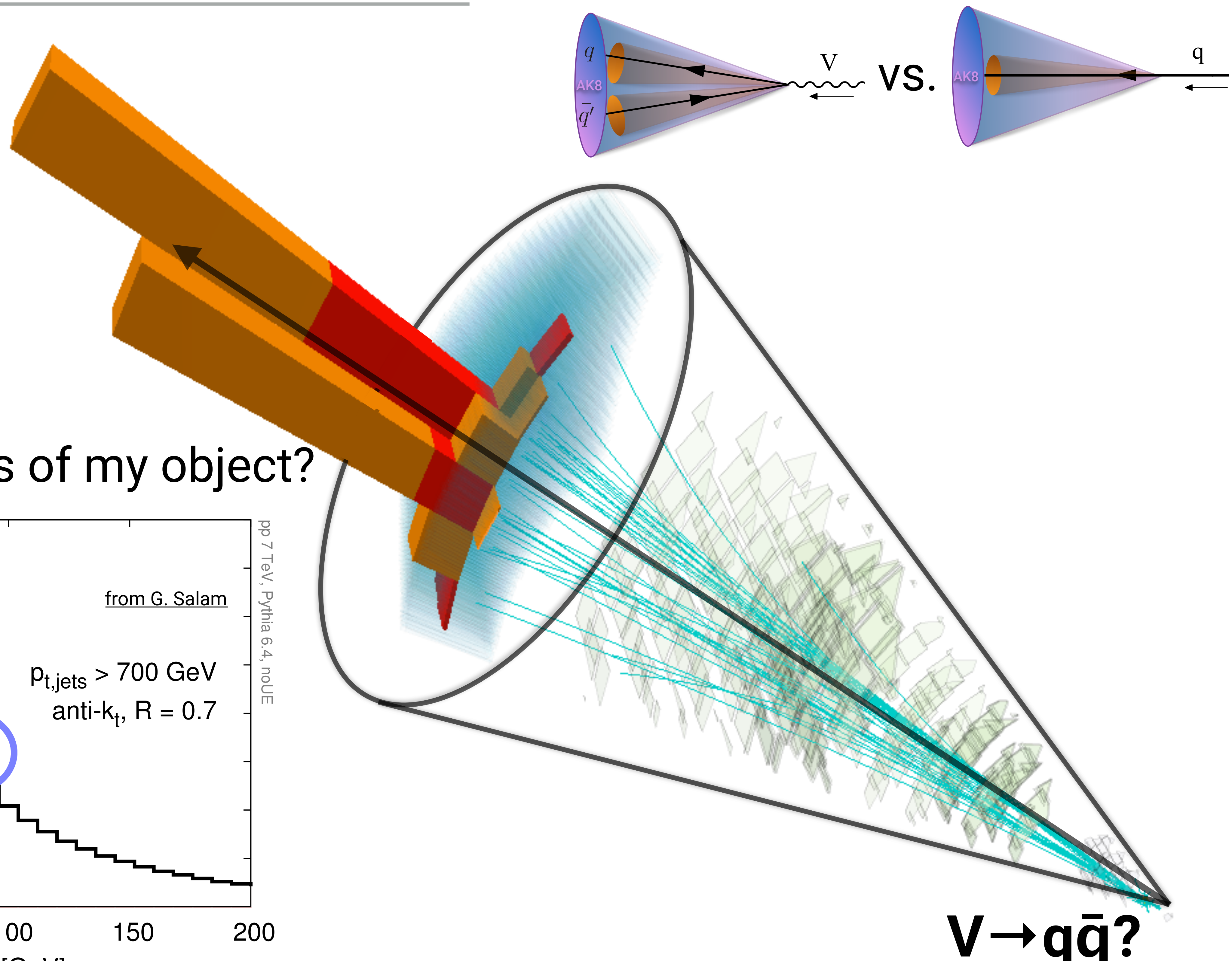
Mass smeared by QCD radiation

1) What's the mass of my object?

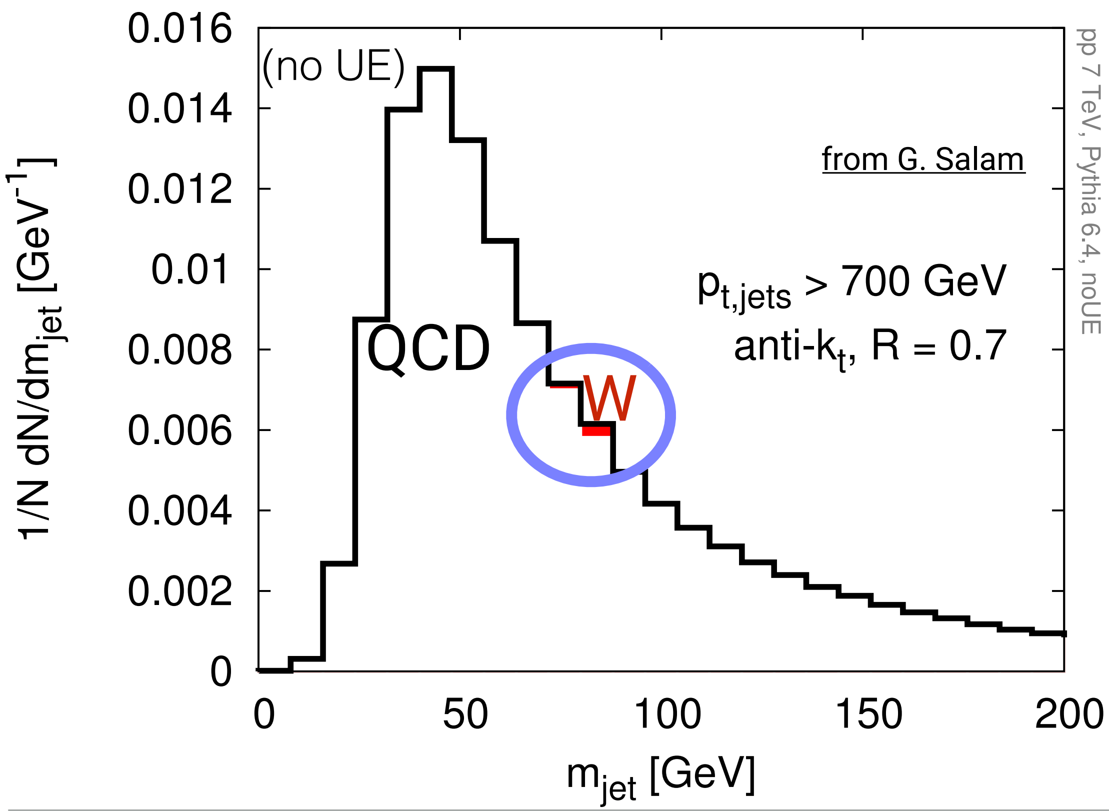


$V \rightarrow q\bar{q}$?

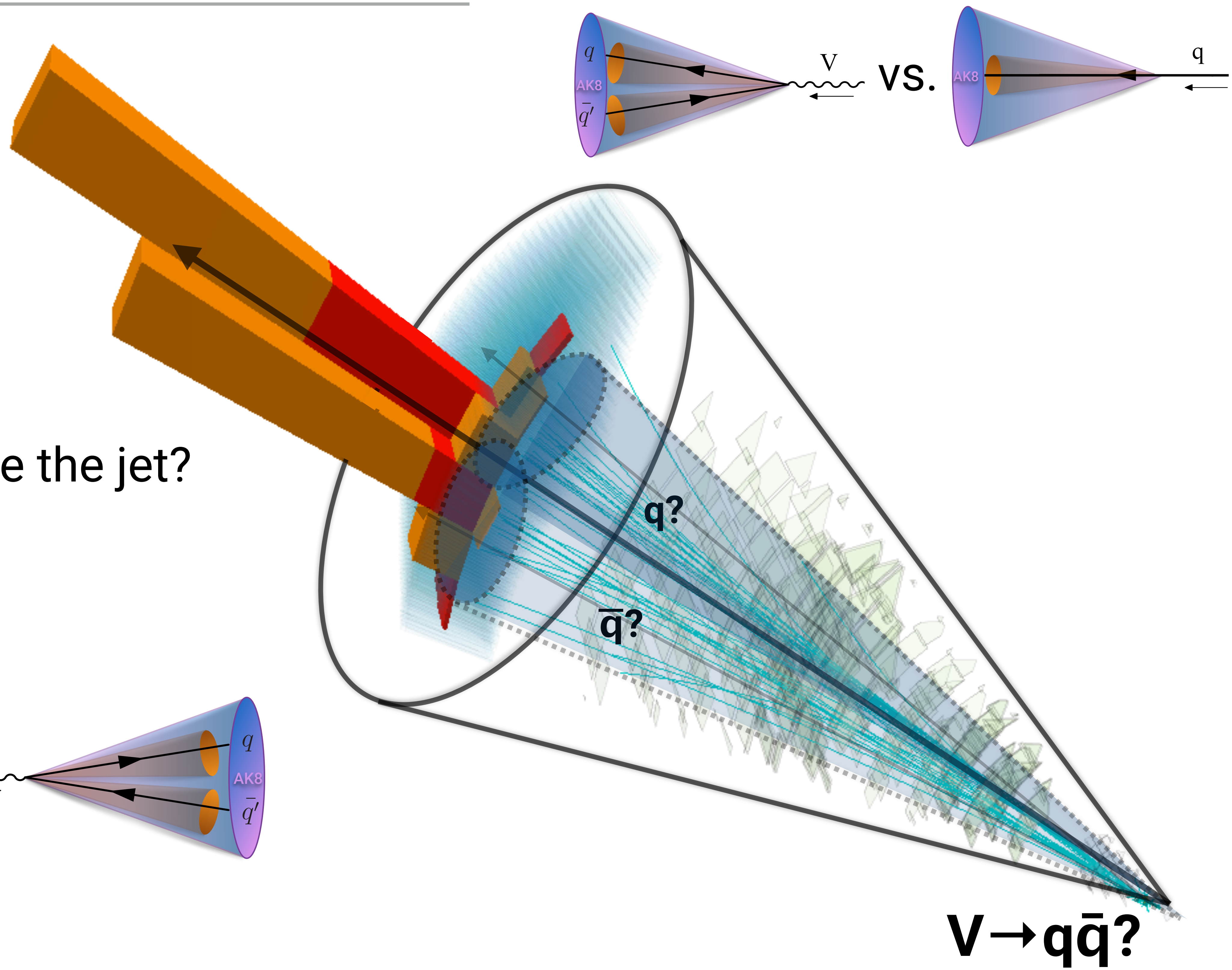
Getting rid of QCD



1) What's the mass of my object?

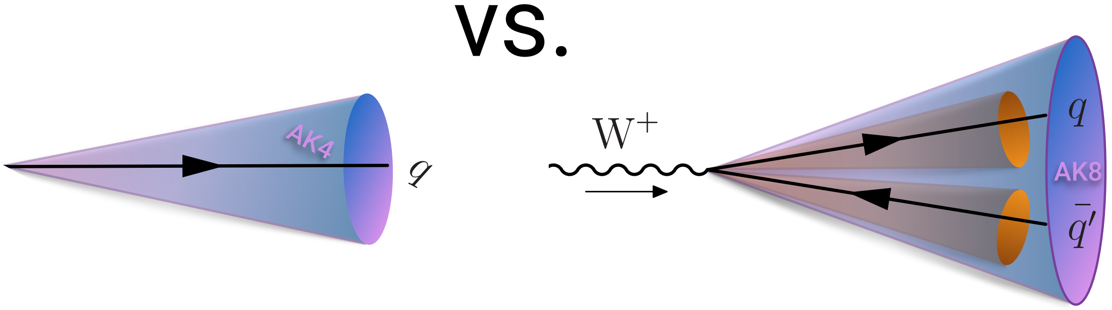


Getting rid of QCD

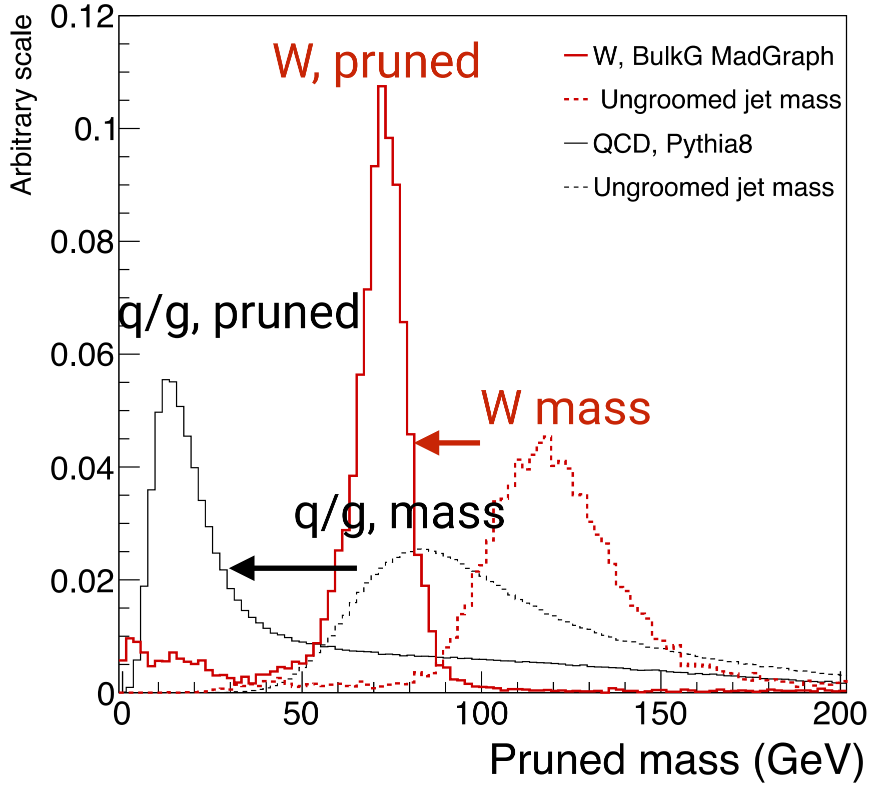


2) Can I peak inside the jet?

vs.



Jet substructure techniques



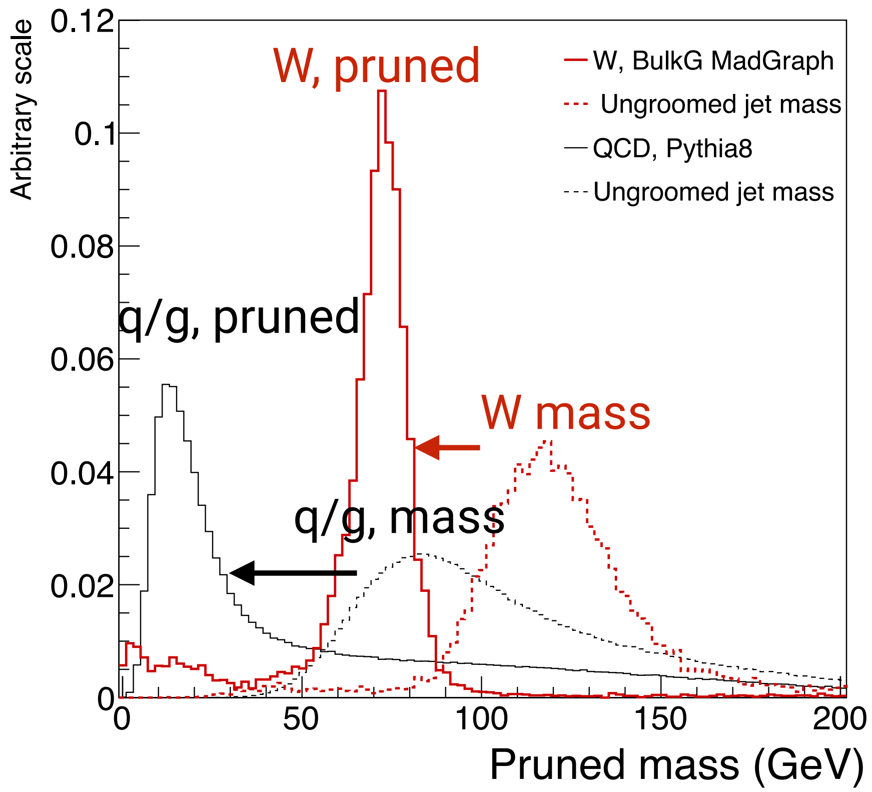
What is the mass of my jet → Pruning

arxiv:0912.0033

Improve mass resolution by removing radiation

- recluster jet, veto soft+wide angle constituents

Jet substructure techniques

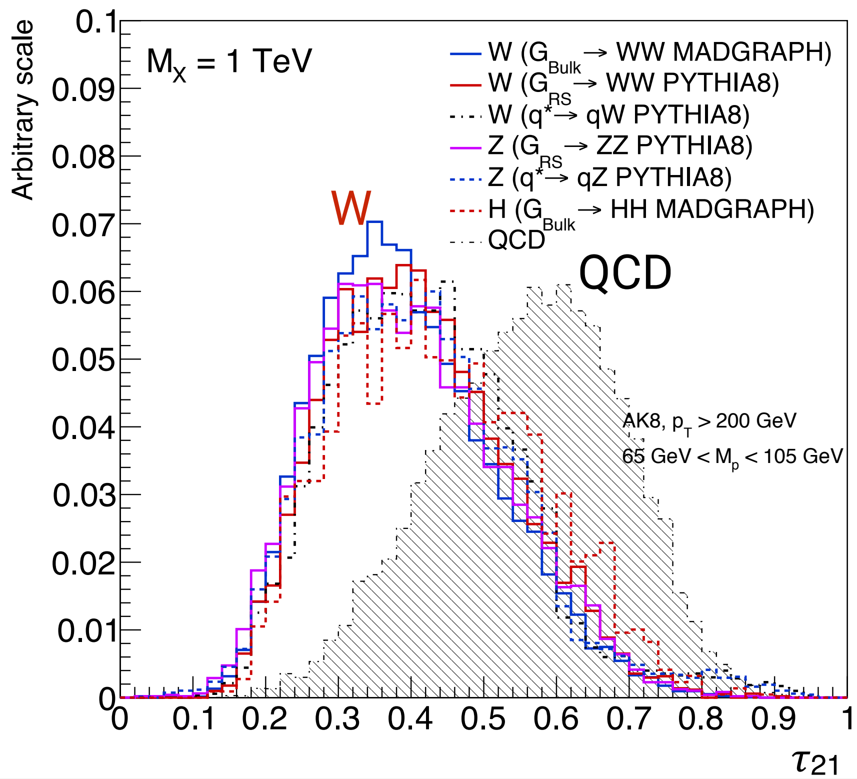


What is the mass of my jet → Pruning

arxiv:0912.0033

Improve mass resolution by removing radiation

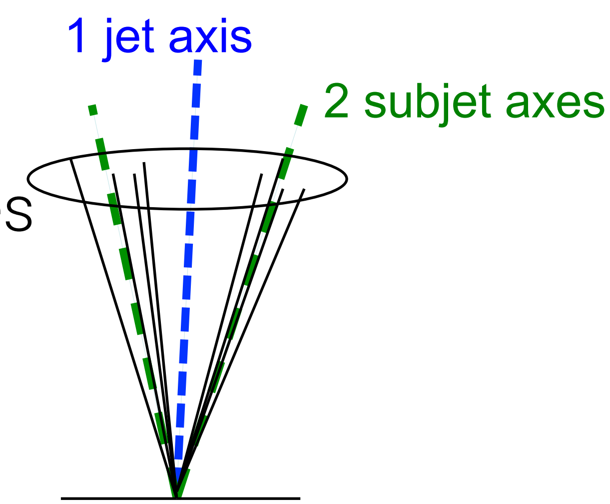
- recluster jet, veto soft+wide angle constituents



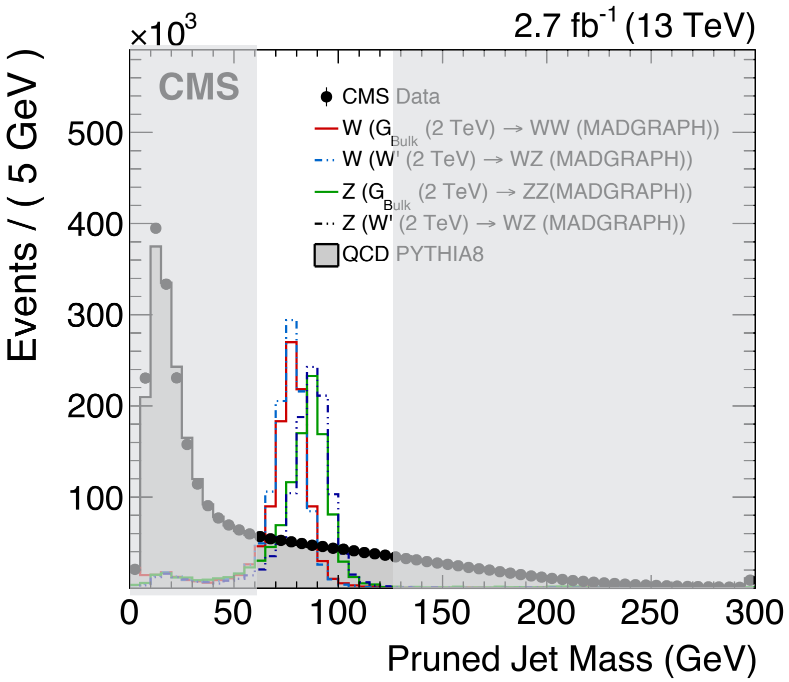
Can I peak inside the jet? → n-subjettiness

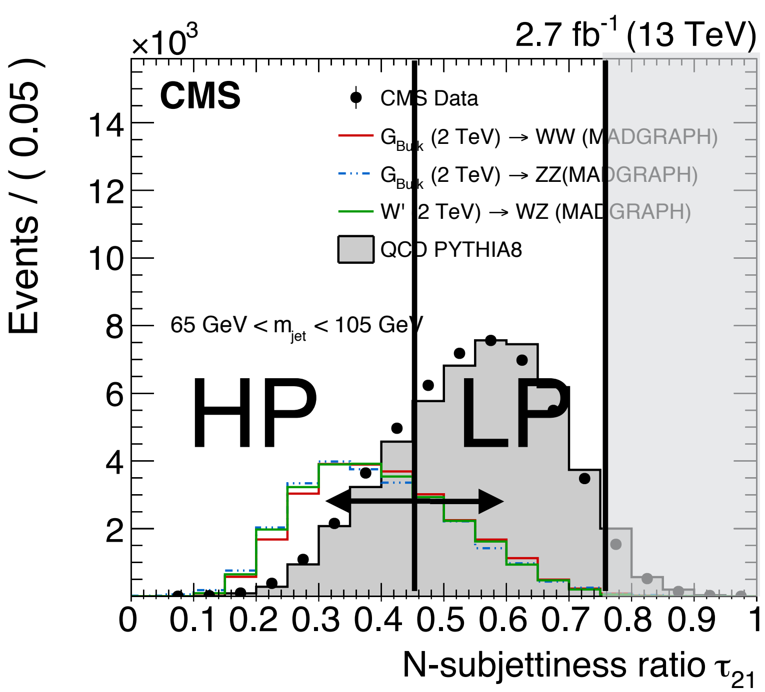
arxiv:1011.2268

- How compatible jet is with having N subjets
  - undo clustering N times, calculate sum of distance between all constituents and axes (small τ_{N+1} →N+1 prong)
  - Rather τ_2/τ_1: q/g with large τ_1, usually large τ_2



Tagging vector bosons





W/Z-tagging: Pruning + τ_{21}

arxiv:0912.0033

Pruned mass window optimised for best S/B and non-overlap with the Higgs boson

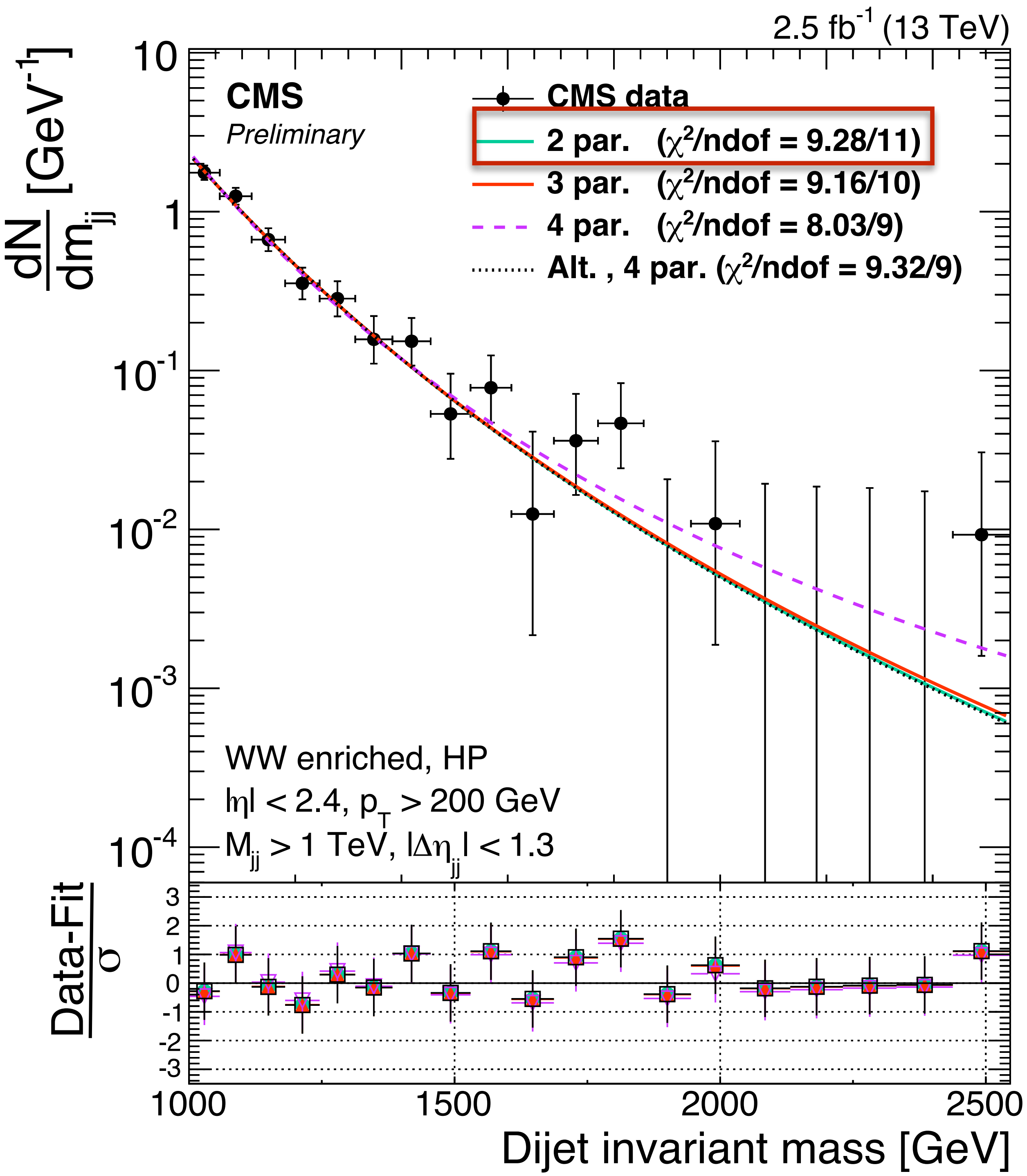
- 65 GeV < M_{Pruned} < 105 GeV

W/Z-tagger (pruning+ $\boldsymbol{\tau}_{21}$):
~55% efficiency
~1-2% mistag rate

Two τ_{21} analysis categories:

- High-purity: $\boldsymbol{\tau}_{21} \leq 0.45$
  (best possible S/B)
- Low-purity: $0.45 < \boldsymbol{\tau}_{21} \leq 0.75$
  (enhance sensitivity at high M_X where bkg is low)

Background modelling



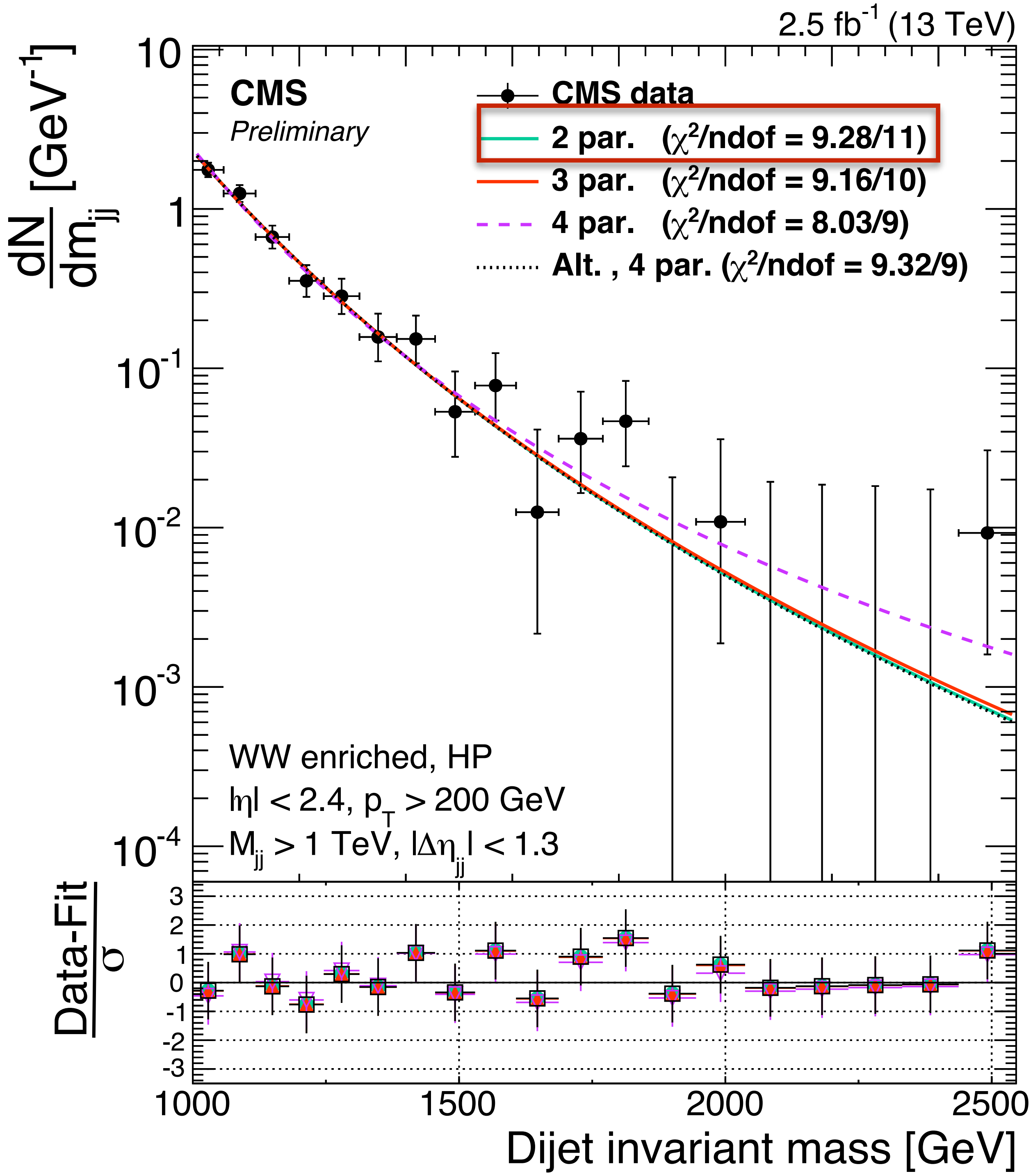
Background assumed to be described by smoothly falling function

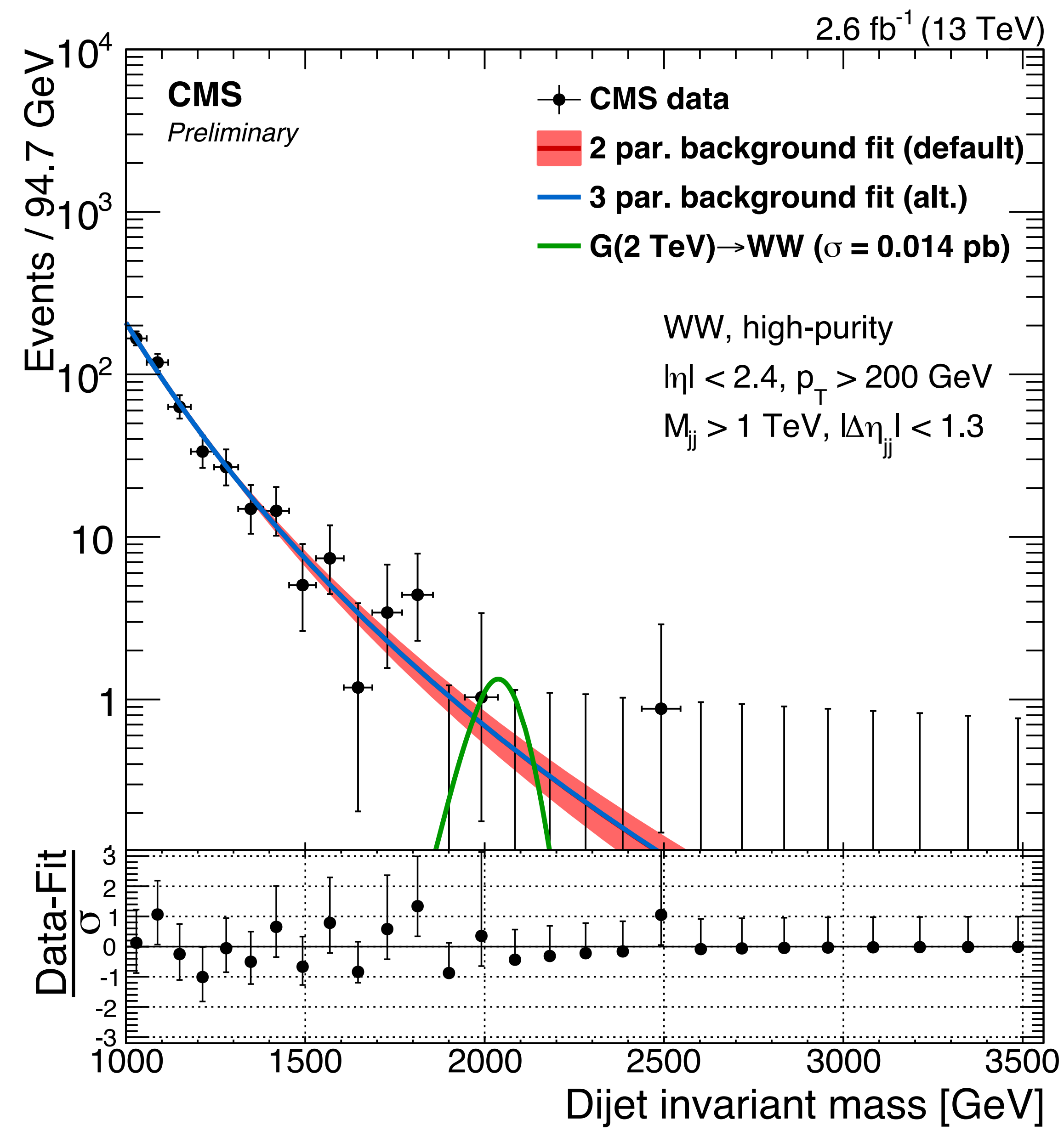
$$\frac{d\sigma}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2}}$$

Function sufficient to describe background?

- F-test: Increase N parameters until no significant improvement (in data!)

Background modelling



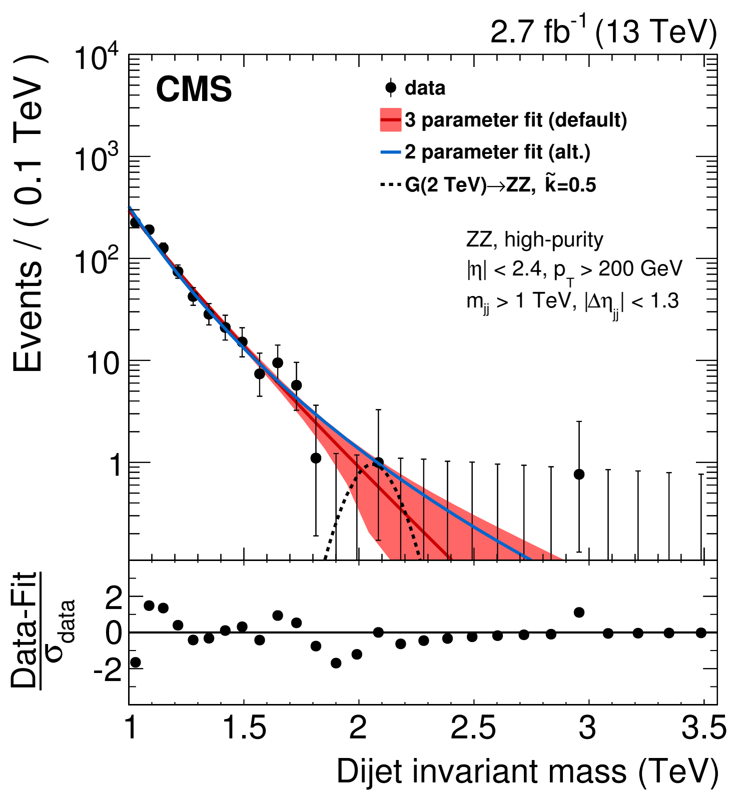


Check that alternate fit functions are within fit uncertainty of nominal fit

- taken as background PDF uncertainty

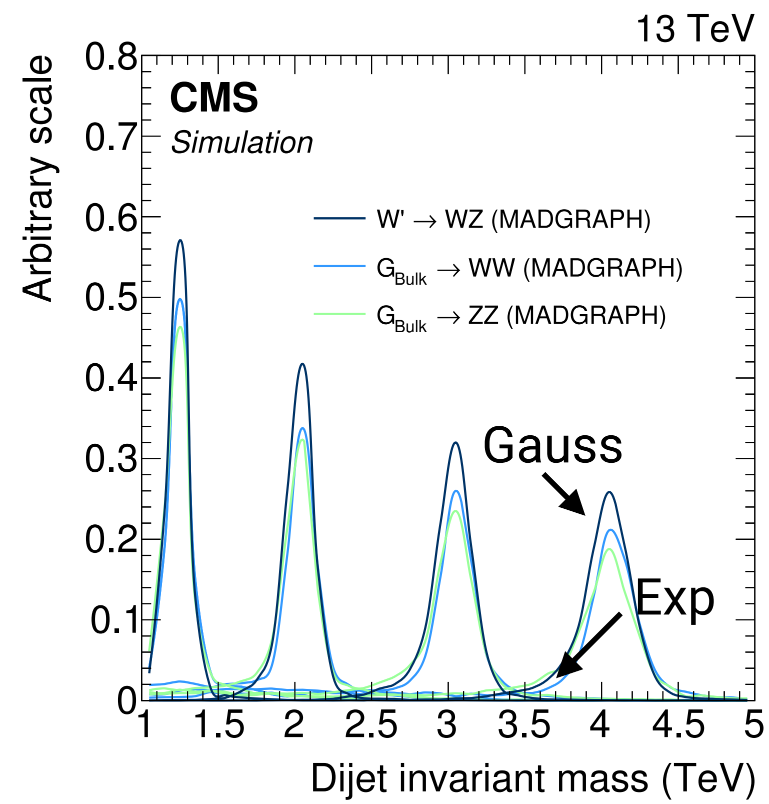
Statistical interpretation

- Background assumed smoothly falling



Smoothness test of observed data

- Signal PDF extracted from MC

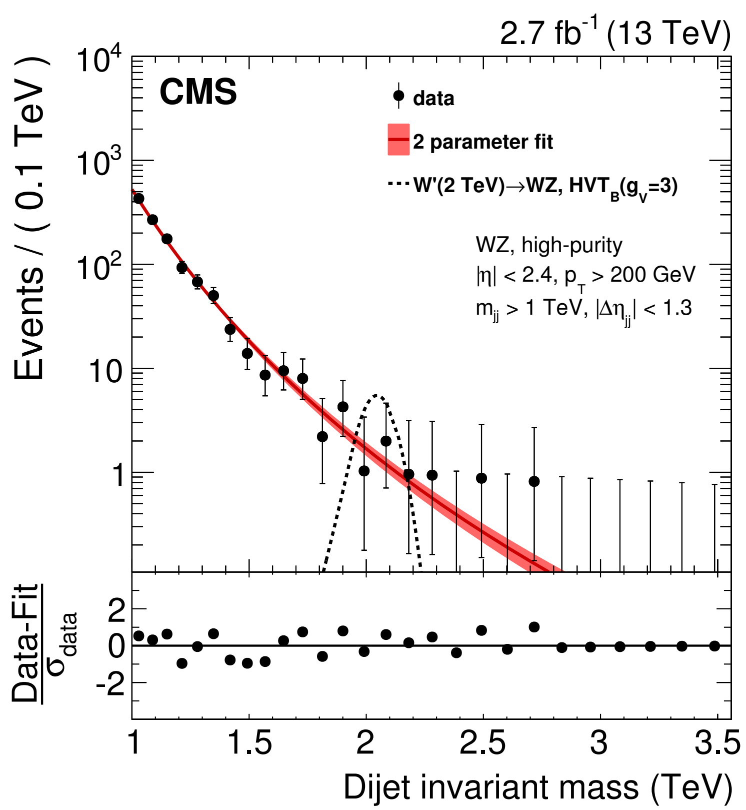


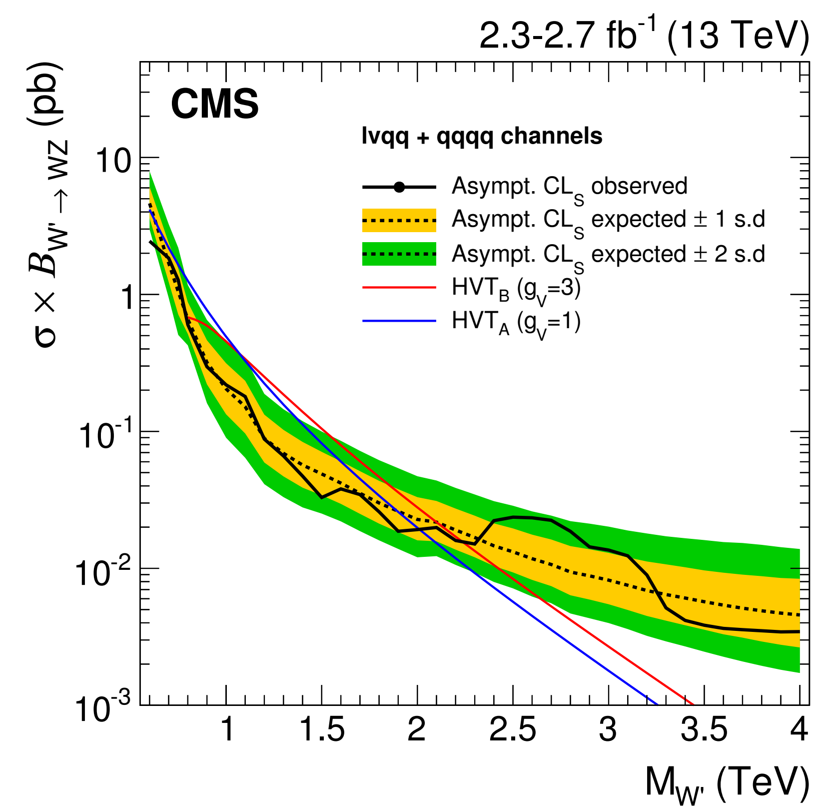
Composite models with Gaussian core and an exponential tail.

Hypothesis test by comparing fits of observed data with "background-only" and "background + signal" function.

- signal strength and background function parameters left floating

Results





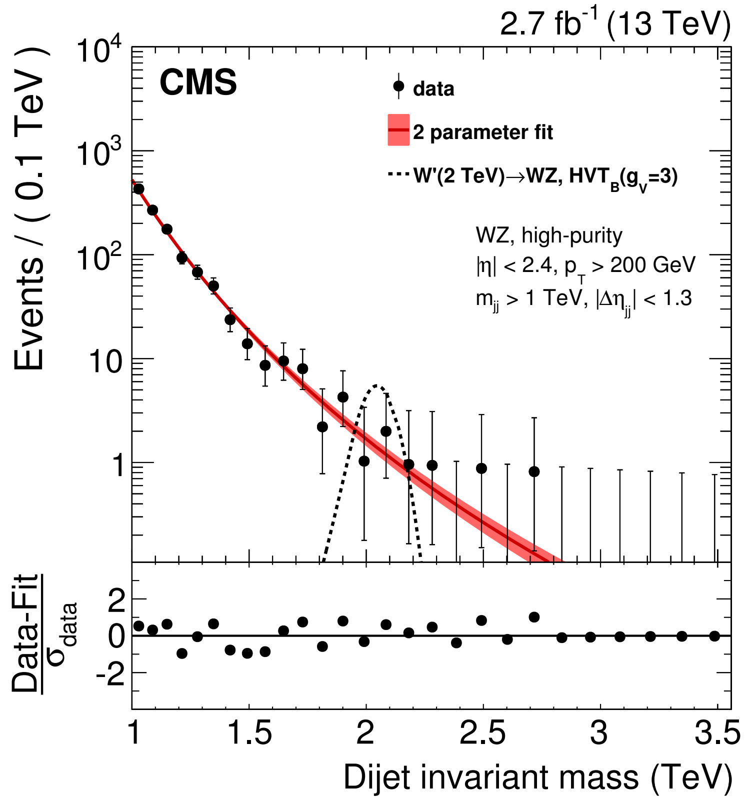
For maximum sensitivity, combined results with semi-leptonic search (VV→$\ell\nu$qq)

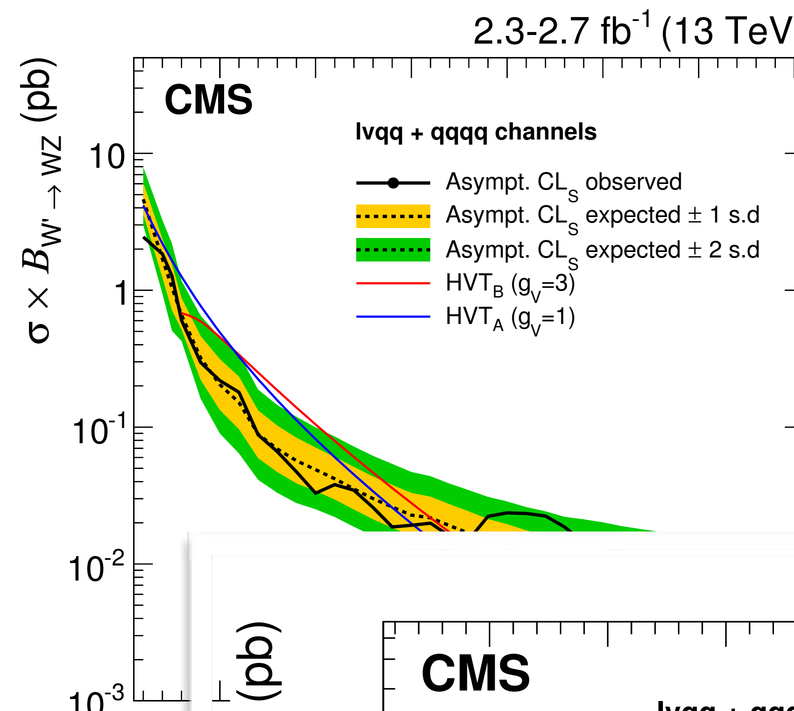
- <1.5 TeV: $\ell\nu$qq dominant, less background
- > 1.5 TeV: dijet dominant, higher branching ratio

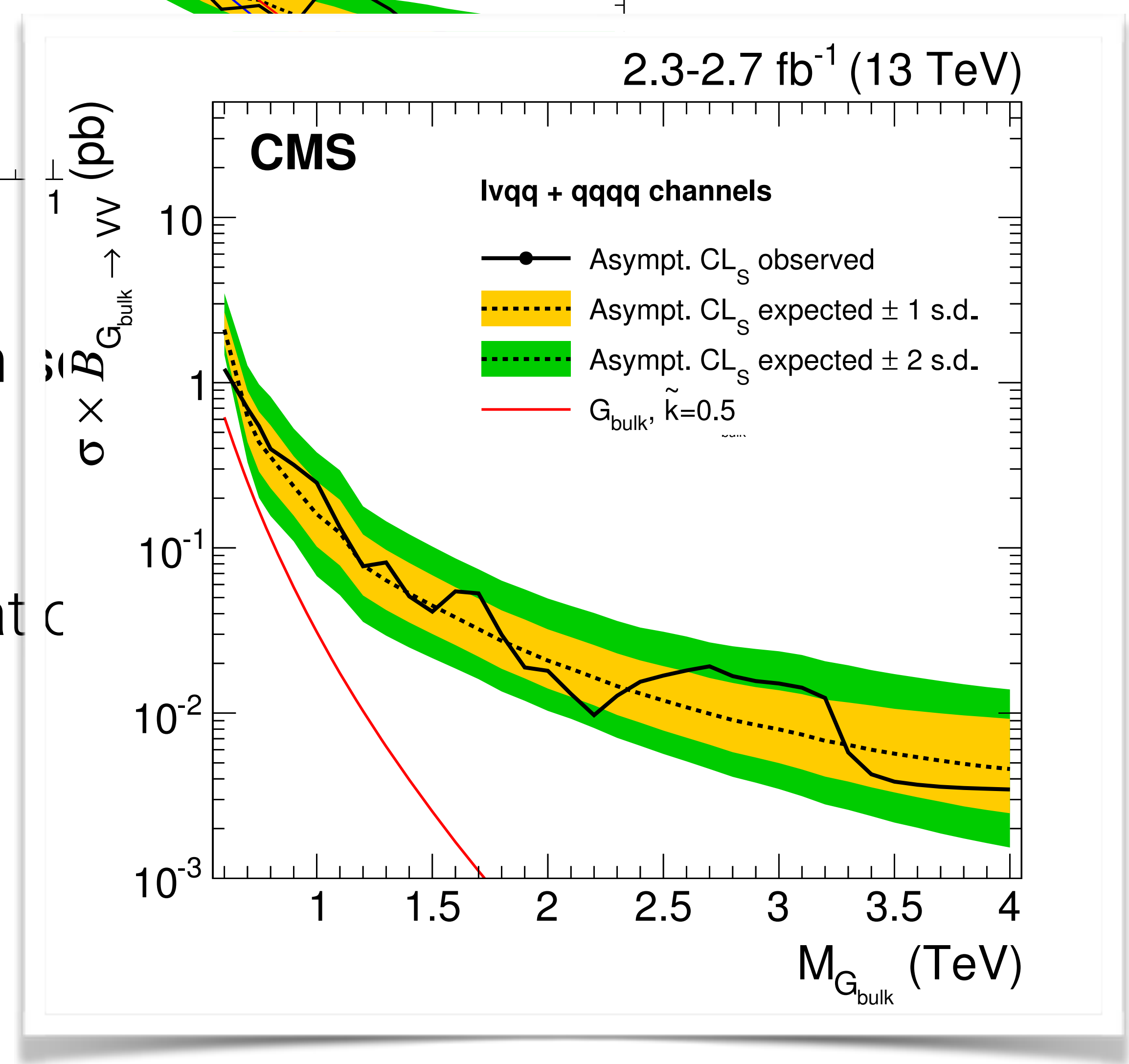
Just exclude 2 TeV excess for W'→WZ!

However, other signals far from excluded!

Results







For maximum sensitivity, combined results with s

- <1.5 TeV: $\ell\nu$qq dominant, less background
- > 1.5 TeV: dijet dominant, higher branching rat

Just exclude 2 TeV excess for W'→WZ!

However, other signals far from excluded!

Interlude: Softdrop

Developments on the theory front:

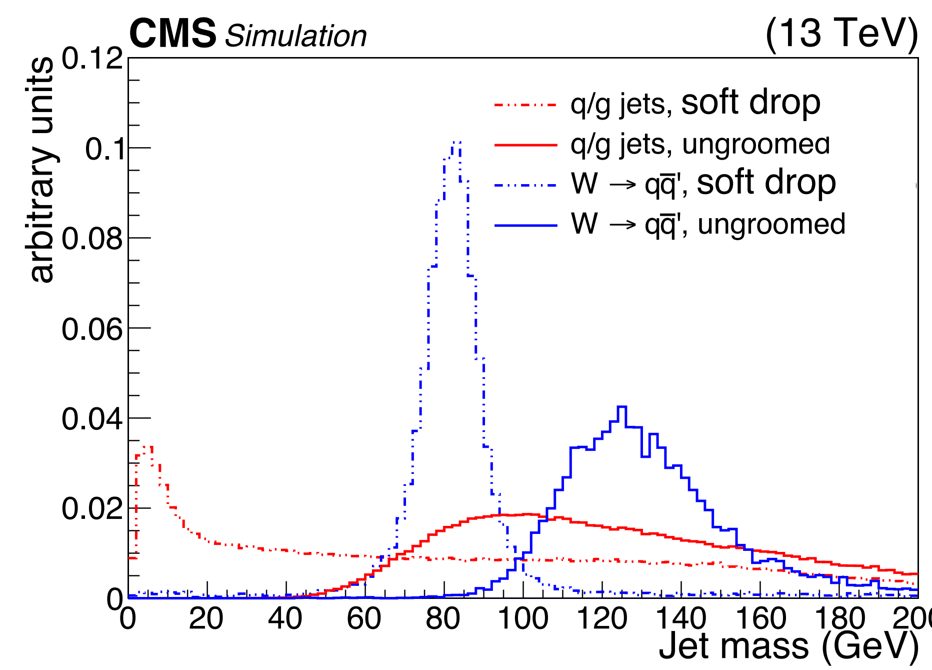
Softdrop (β=0)

arxiv:1402.2657
arXiv:1307.0007

- Remove all soft emission
  - decluster with C-A, remove subjet if

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} < \boxed{0.1}$$

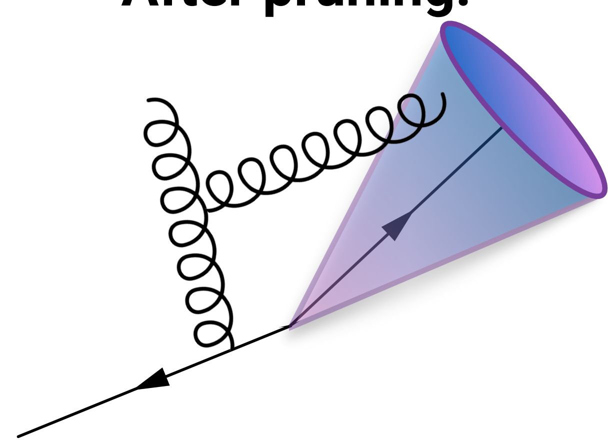
z_{cut}



Pruning NOT completely soft radiation free. Non global logarithmic terms (NGLs) in mass → not "perturbatively robust"

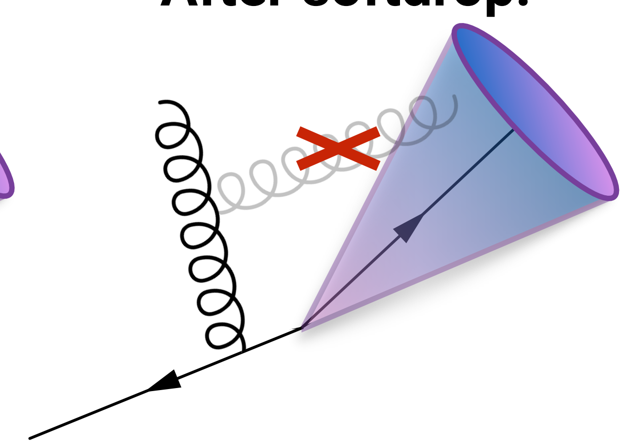
- Softdrop removes all sensitivity to soft divergences, only groomer which is theoretically calculable!

After pruning:



Soft gluon radiating into jet not removed

After softdrop:

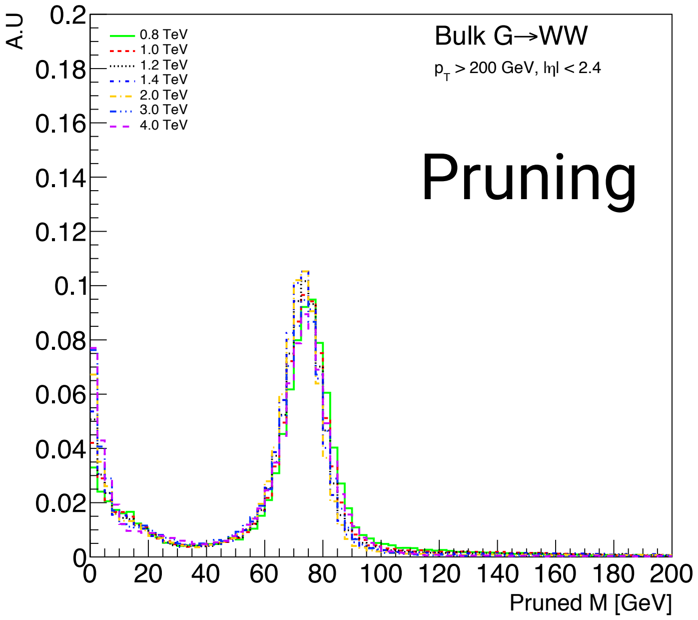


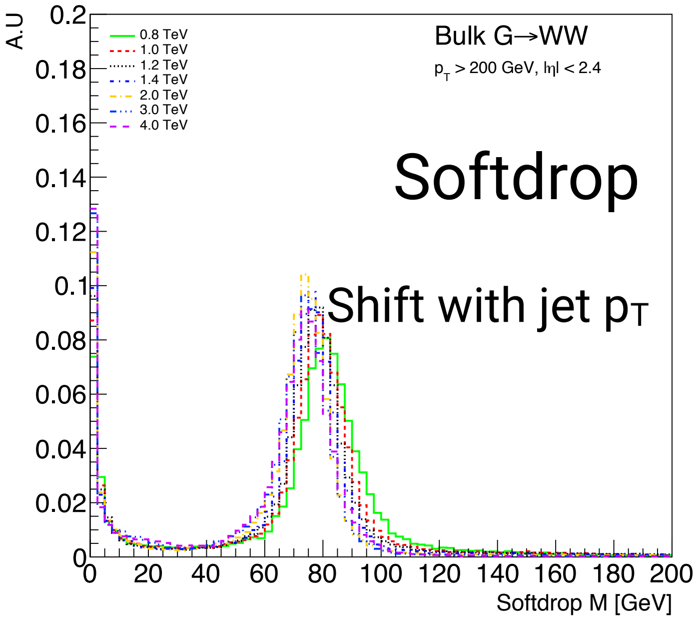
All soft radiation removed → no non-global logs!

Interlude: Softdrop

However,

- found softdrop mass for signal jets highly p_T dependent!

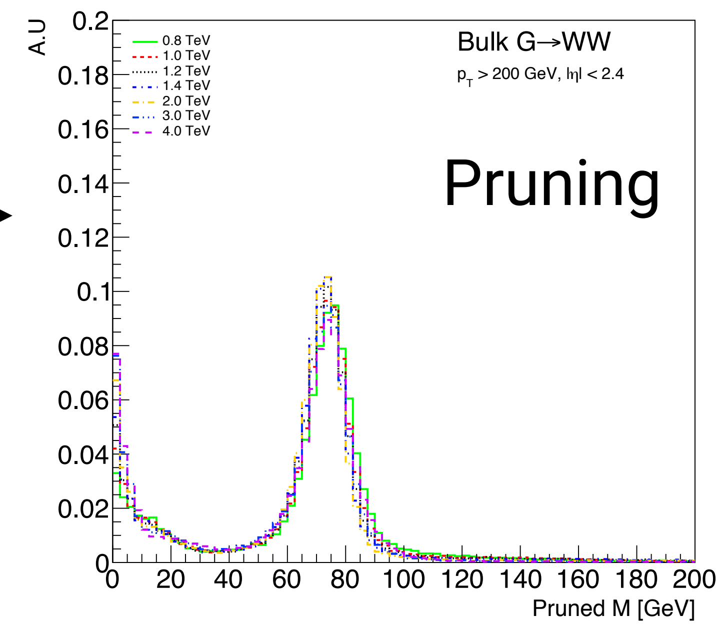


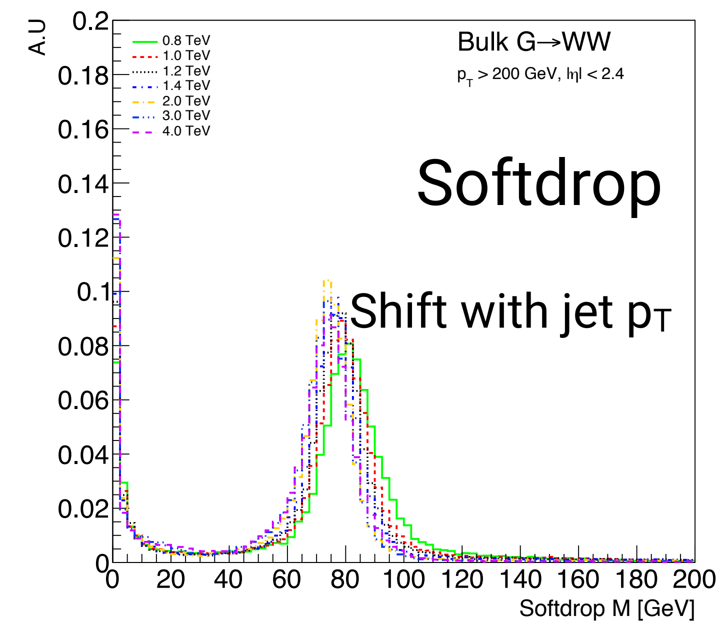


Interlude: Softdrop

However,

- found softdrop mass for signal jets highly p_T dependent!





Due to increased sensitive to UE

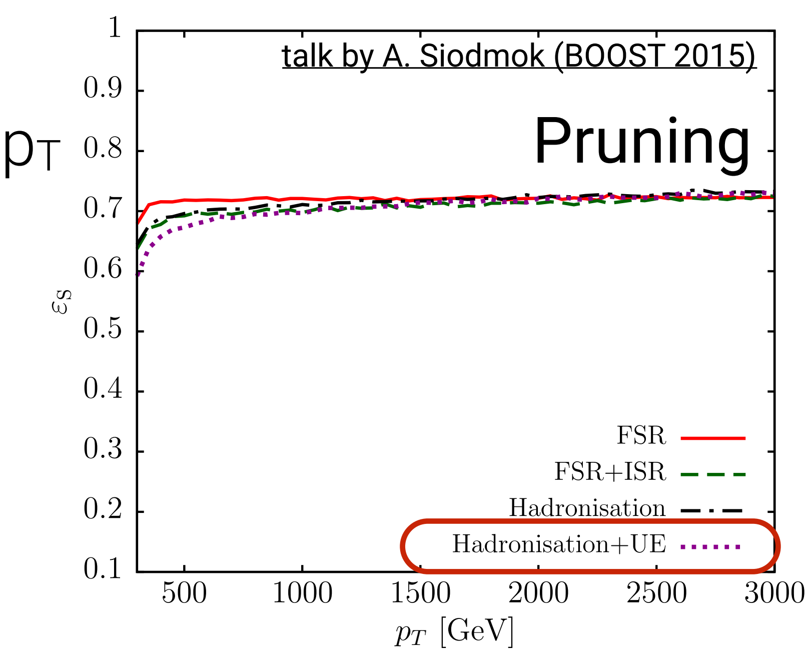
- softdrop effective radius increases as jet p_T decreases

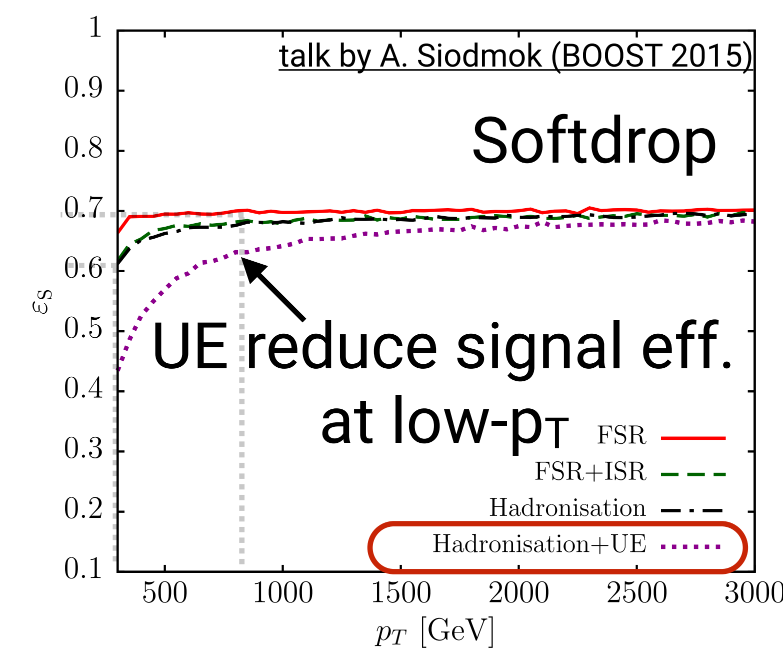
  $\propto m_V/p_T\sqrt{z_{cut}(1-z_{cut})}$

  (for pruning, $\propto m_V/p_T$)

- Absorb more radiation at low-p_T

Vector boson tagging efficiency vs. p_T (Herwig++)





Interlude: Softdrop

However,

- found softdrop mass for signal jets highly p_T dependent! →

Due to increased sensitive to UE

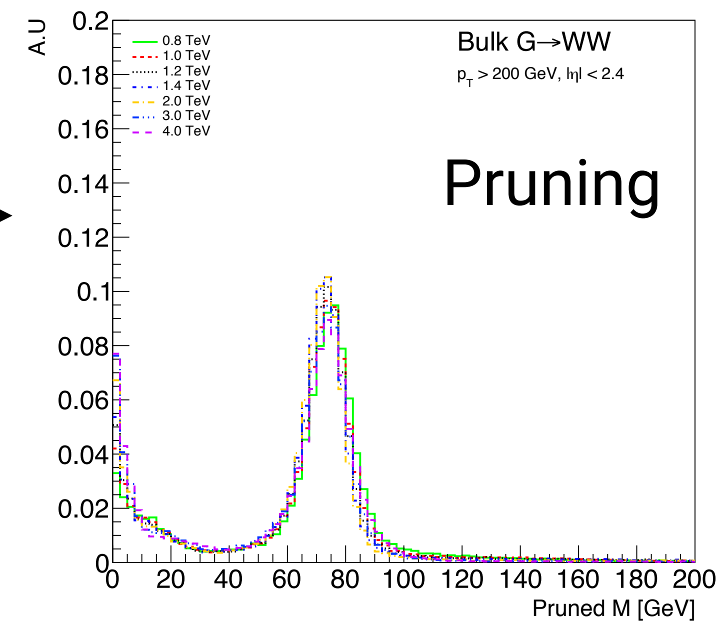
- softdrop effective radius increases as jet p_T decreases

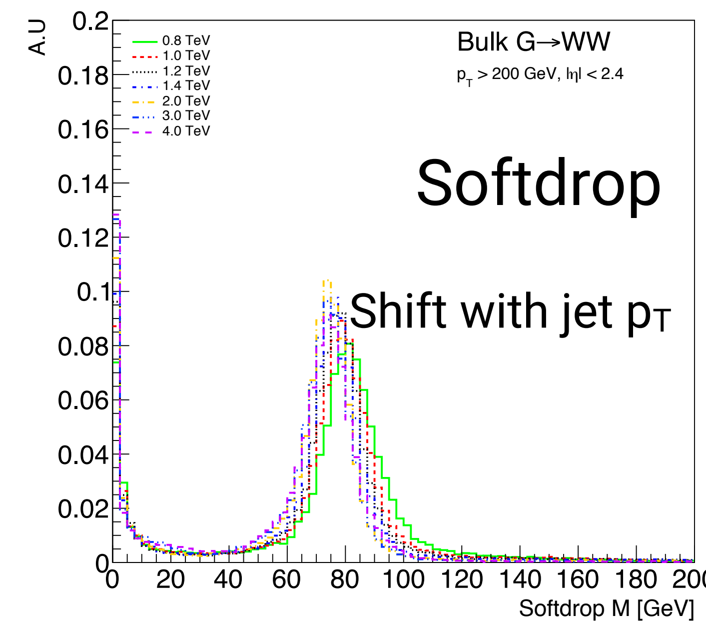
  $\propto m_V/p_T\sqrt{z_{cut}(1-z_{cut})}$

  (for pruning, $\propto m_V/p_T$)

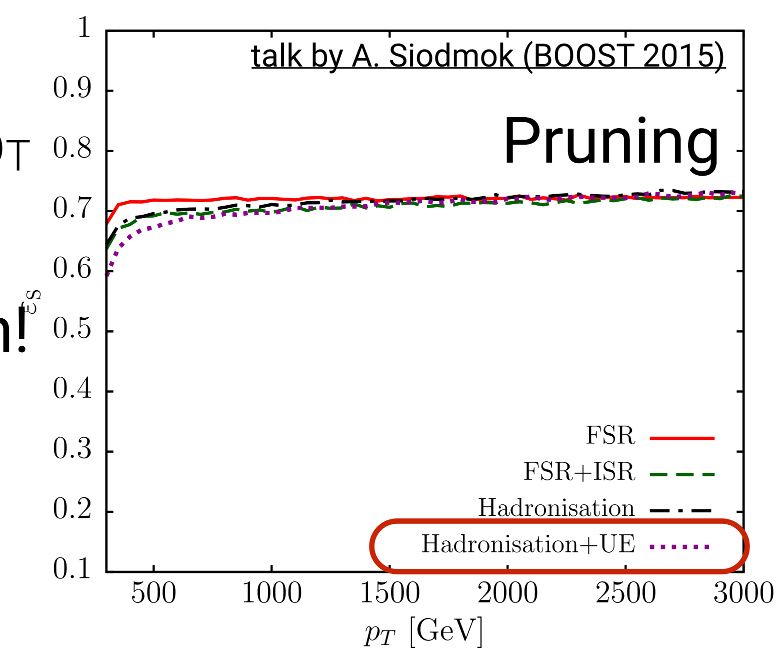
- Absorb more radiation at low-p_T

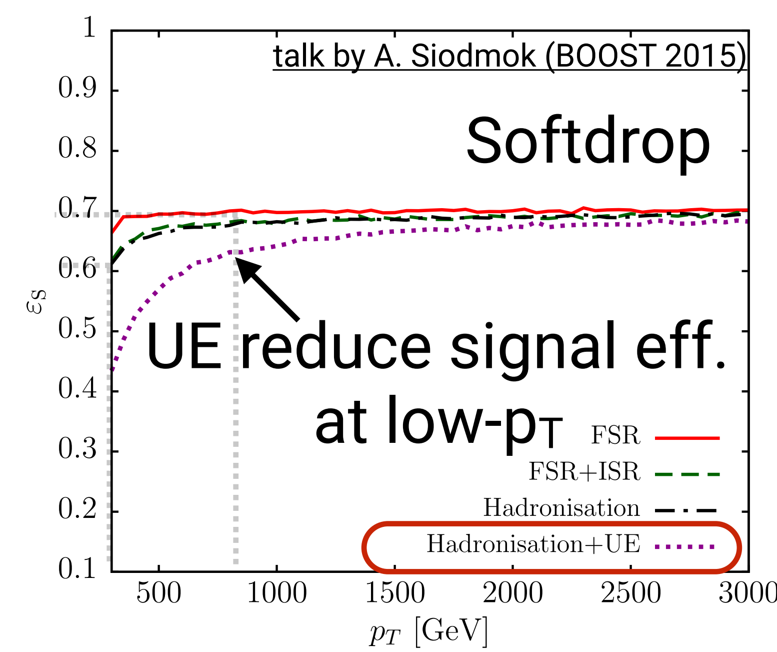
Need better pileup/UE subtraction!





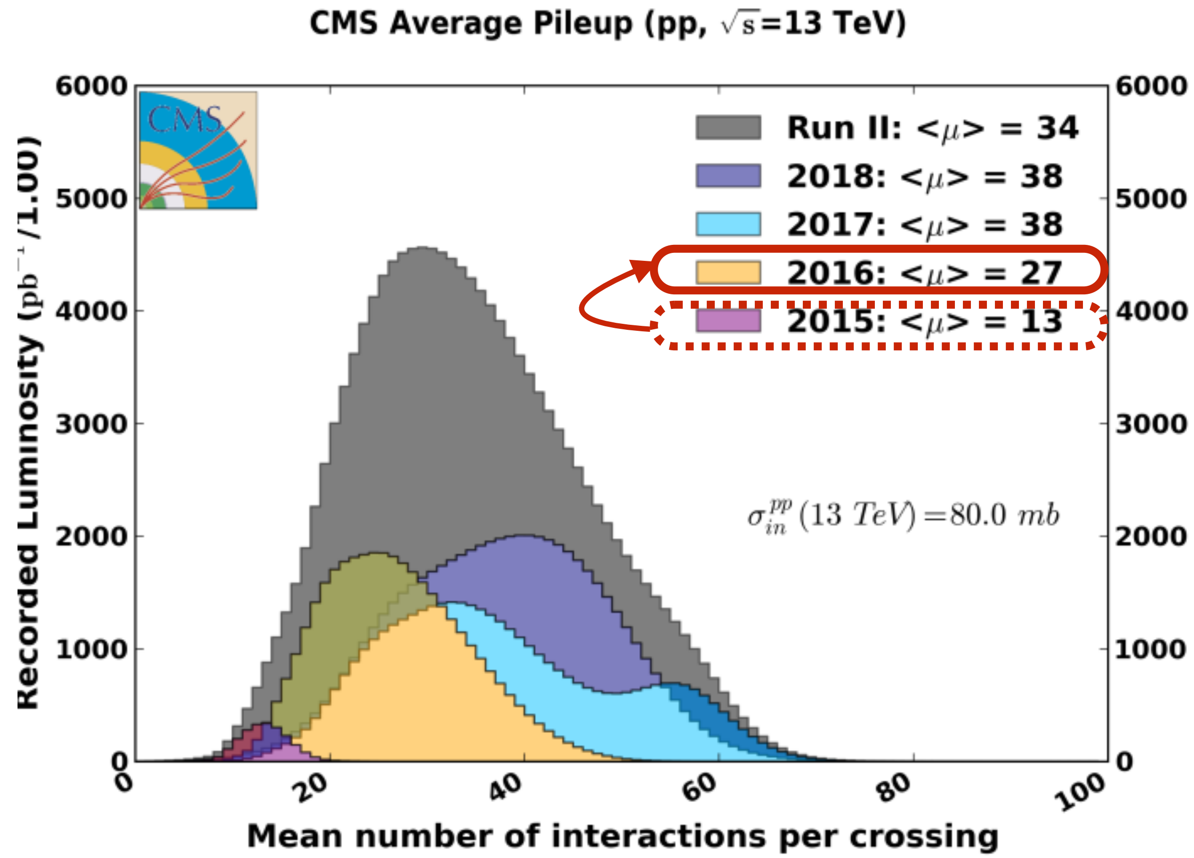
Vector boson tagging efficiency vs. p_T (Herwig++)





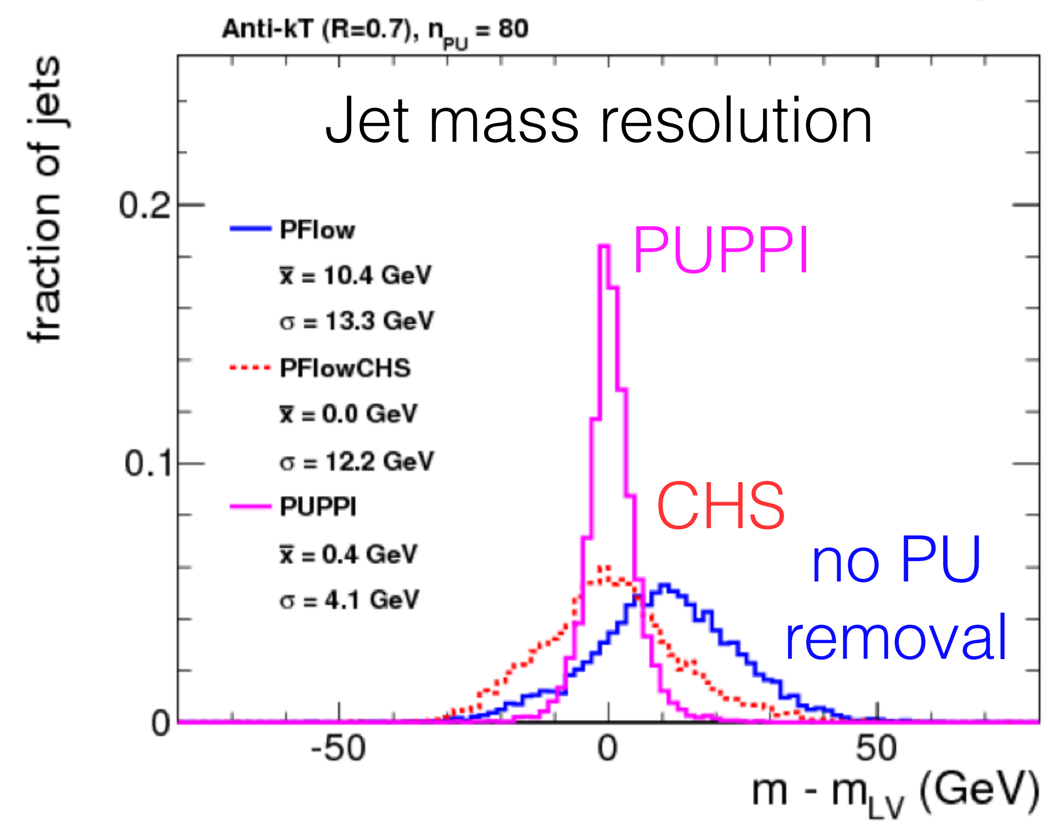
Interlude: PUPPI

Unfortunately, pileup in 2016 expected to be double that of 2015!



Fortunately,
PileUp Per Particle Identification (PUPPI)

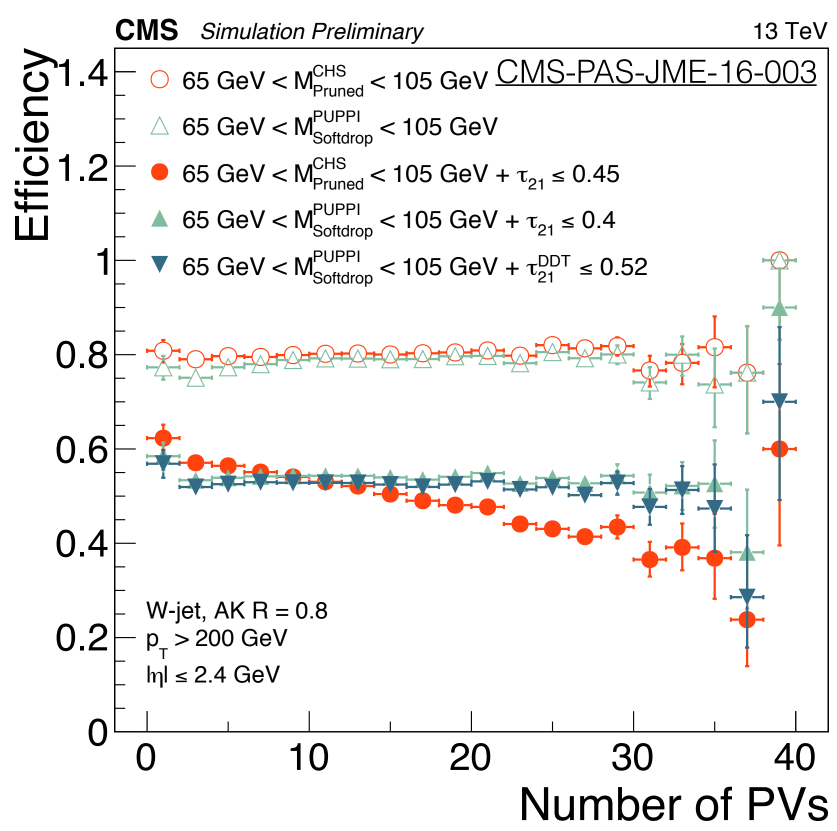
- CHS (old): remove charged hadrons not associated with primary vertex
- PUPPI (new): probability for ANY particle (neutral+charged) to be from pileup, reweights each accordingly



Huge resolution improvement for jet observables in large-cone jets

Search II:
A novel pileup resistant, perturbative safe tagger

Published in PRD, DOI: 10.1103/PhysRevD.97.072006; CMS-PAS-B2G-16-021; CMS-PAS-JME-16-003



~ First analysis to use the PUPPI+softdrop algorithm; optimizing and commissioning new tagger in the process (now default for W-tagging in CMS). Adding new analysis never before explored at 13 TeV: q→ qV. Published with the full 2016 dataset, 35.9 fb^{-1}*

Developing a new V-tagger: Softdrop mass corrections

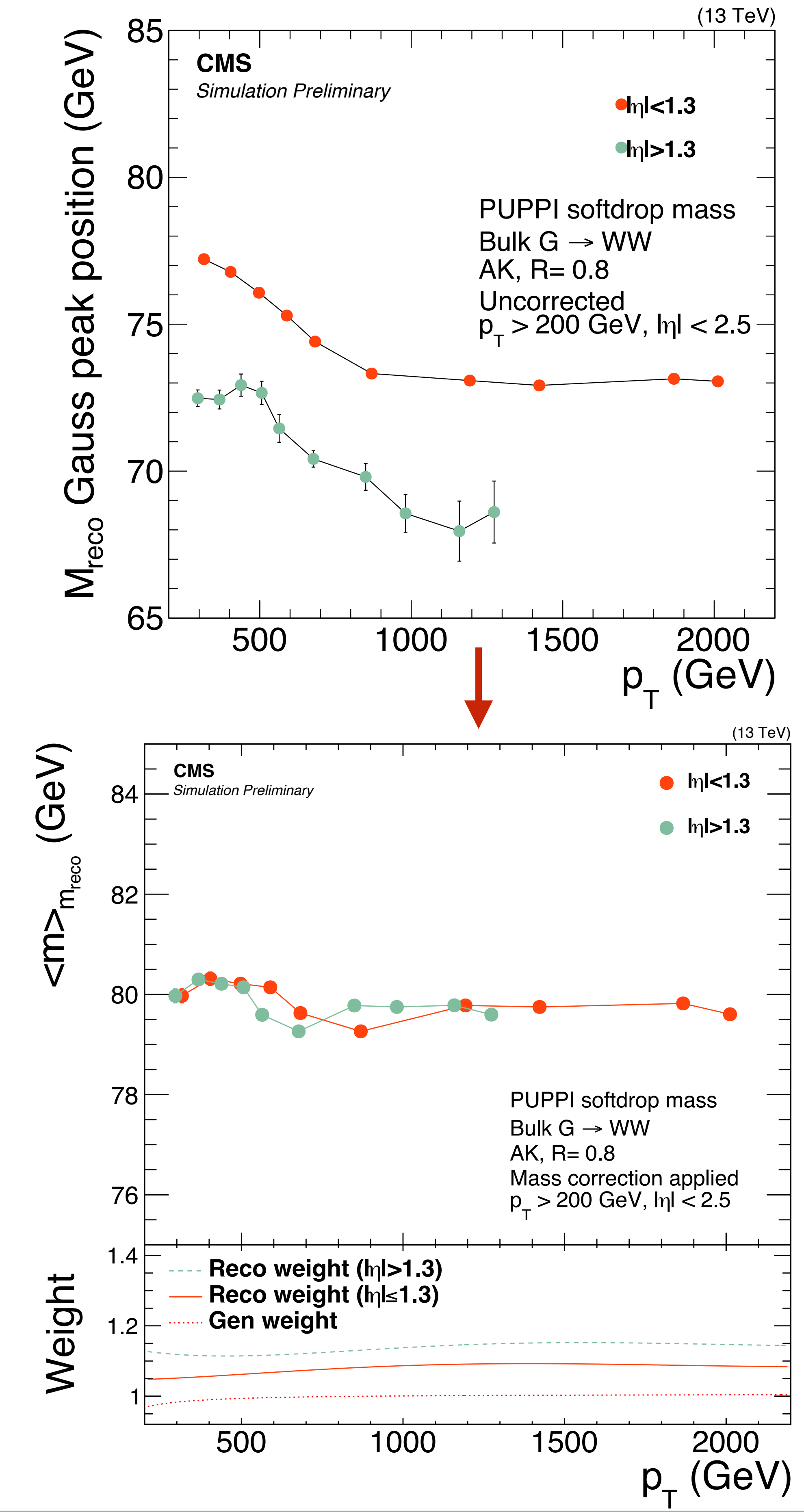
With PUPPI, was softdrop saved?

- Better, but still residual p_T dependence
- Enhanced when applying standard CMS jet energy corrections

Solution: Compute dedicated PUPPI softdrop jet mass scale corrections

- remove p_T/η-dependence, shift mass to 80 GeV

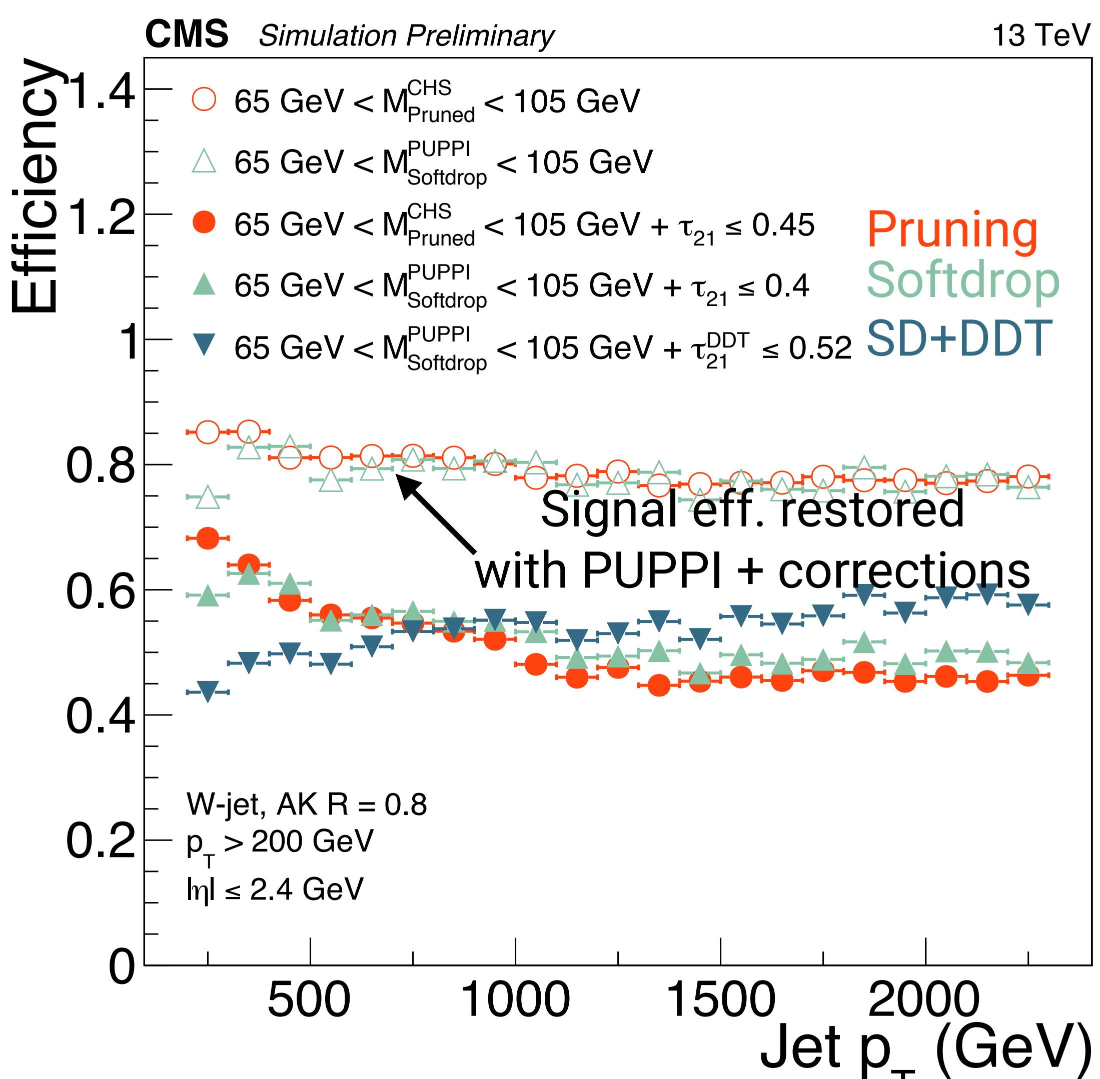
Finally stabile softdrop mass peak

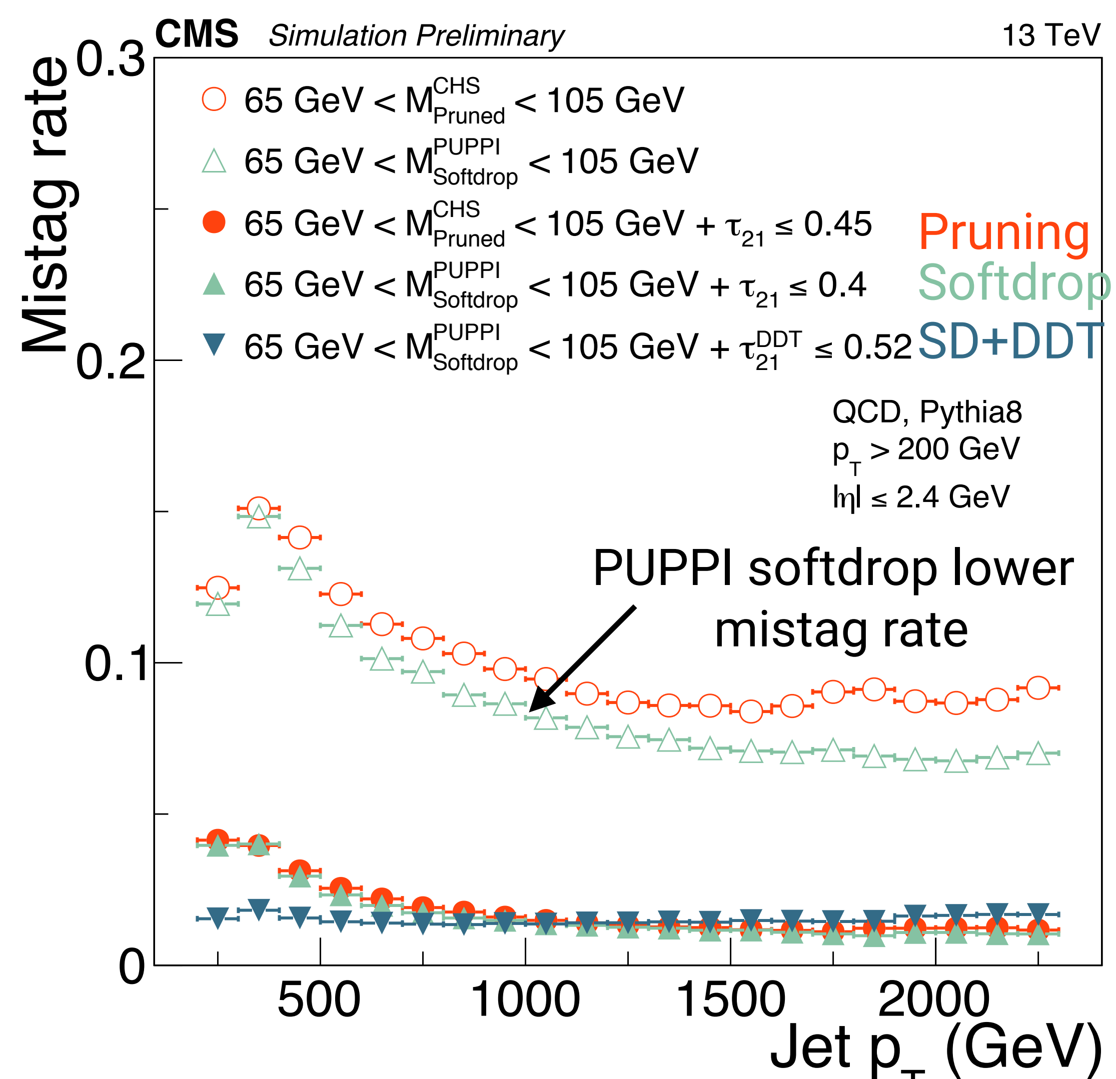


Developing a new V-tagger: Performance

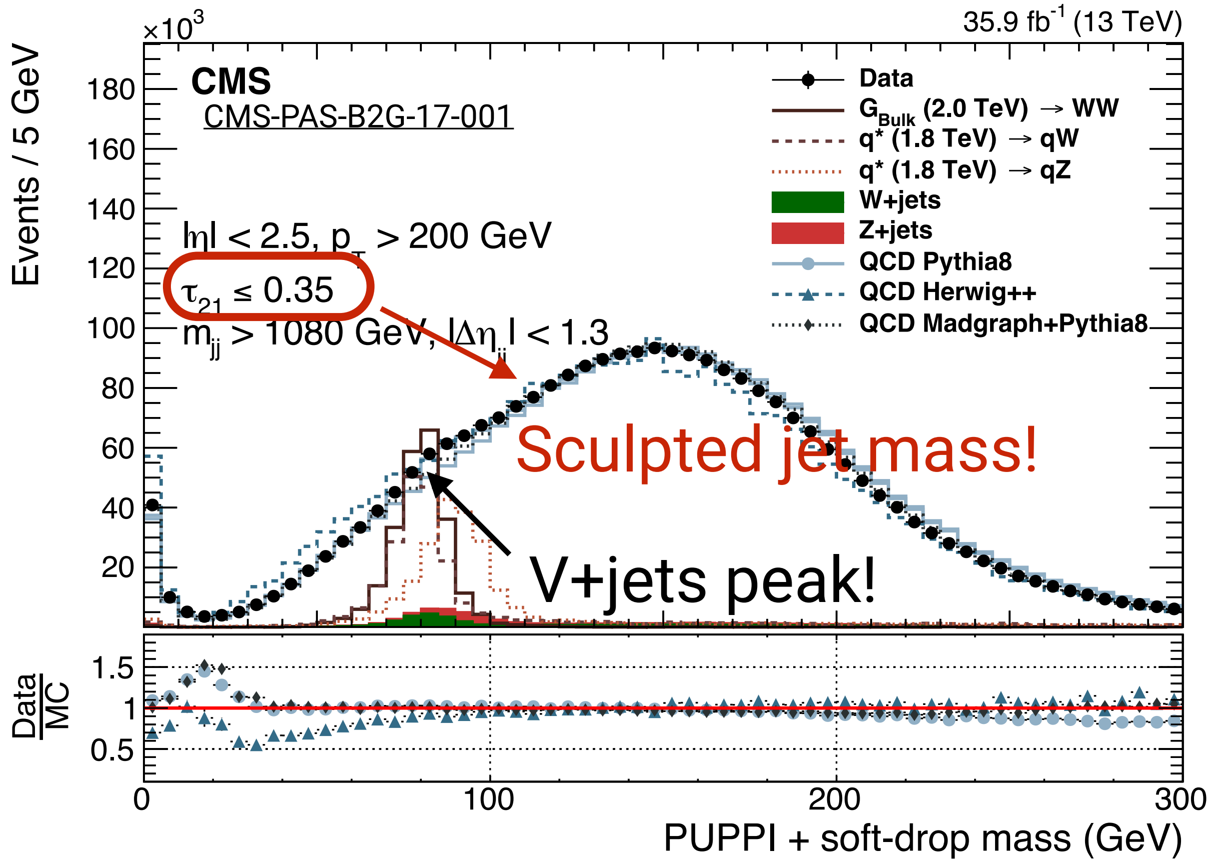
Compare 3 taggers

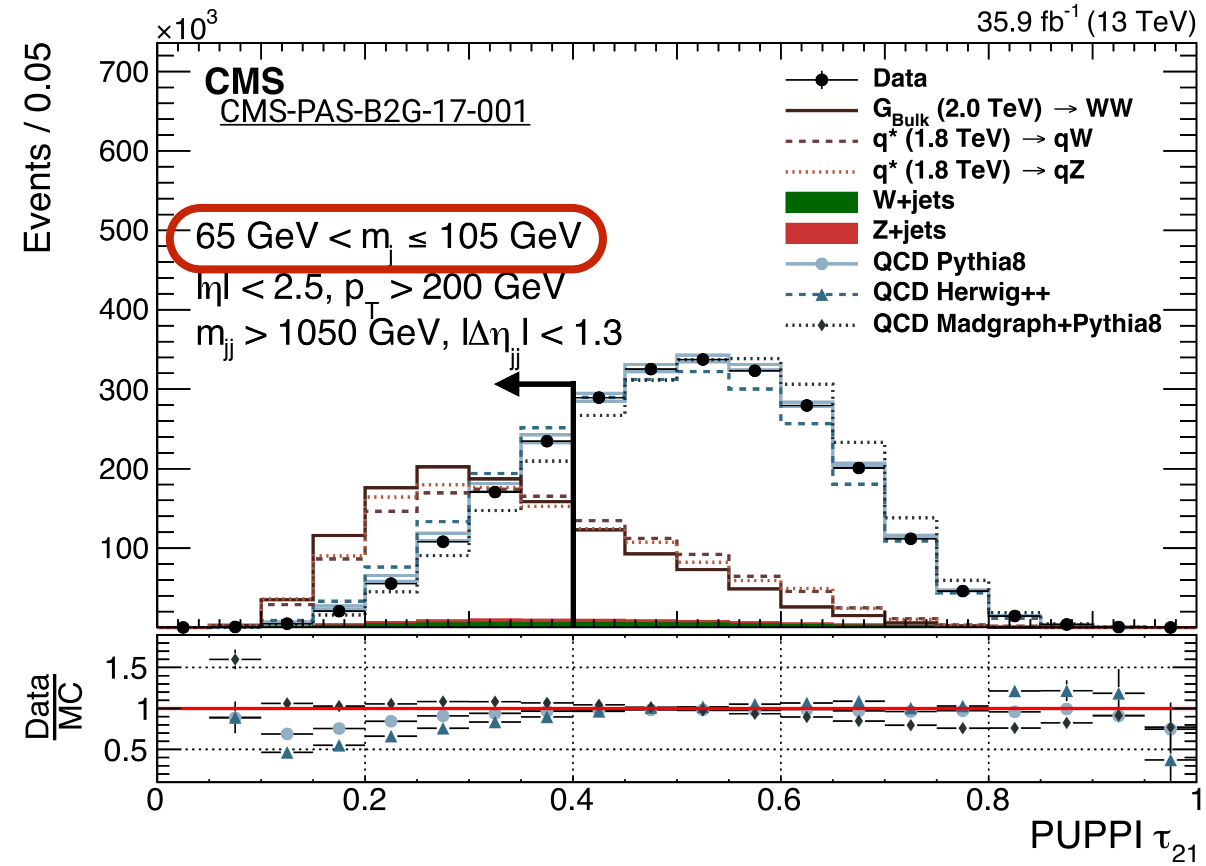
- Pruning + τ_{21}
- PUPPI Softdrop + τ_{21}
- PUPPI Softdrop + τ_{21}^{DDT} → linear transformation of τ_{21} decorrelated from m/p_T





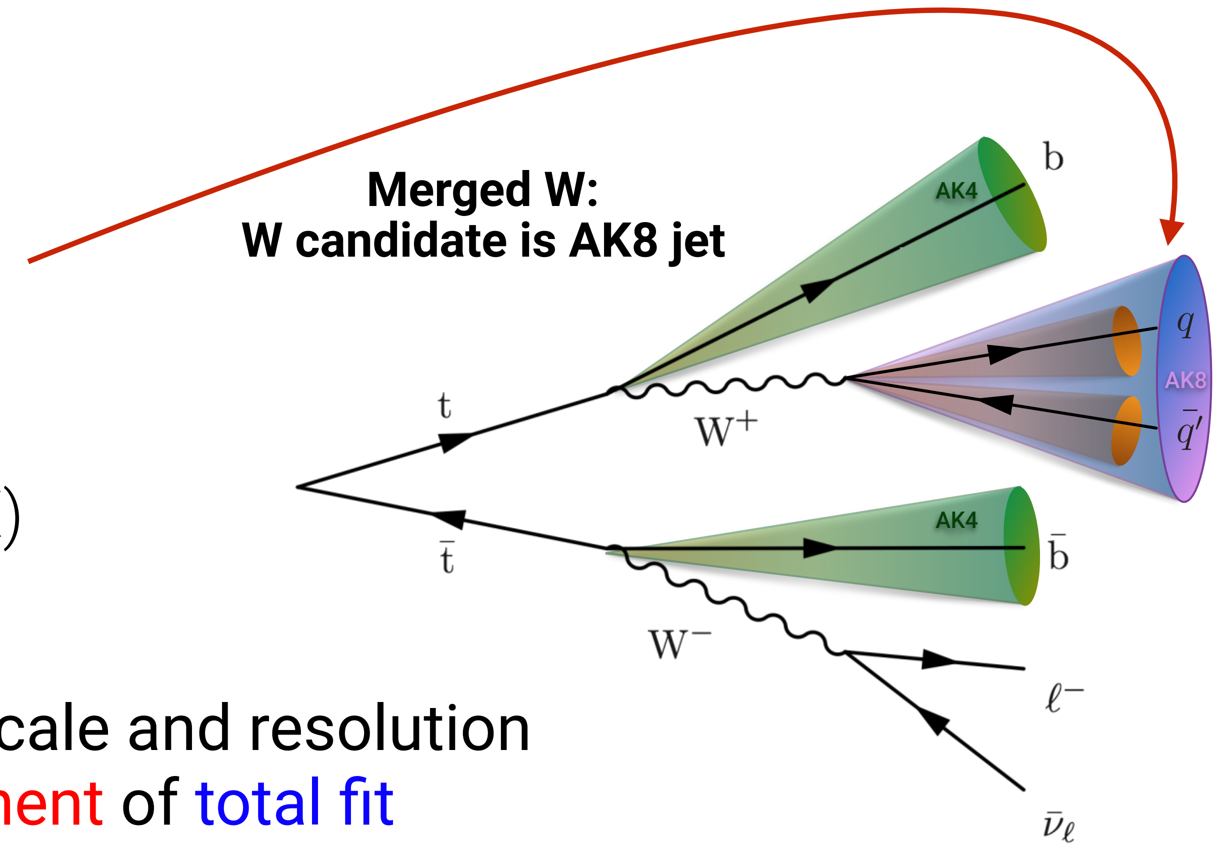
Developing a new V-tagger: Performance in data





- Substructure variables strong dependence on shower generator
- Need to ensure we know real signal eff. in data from region well described by MC!

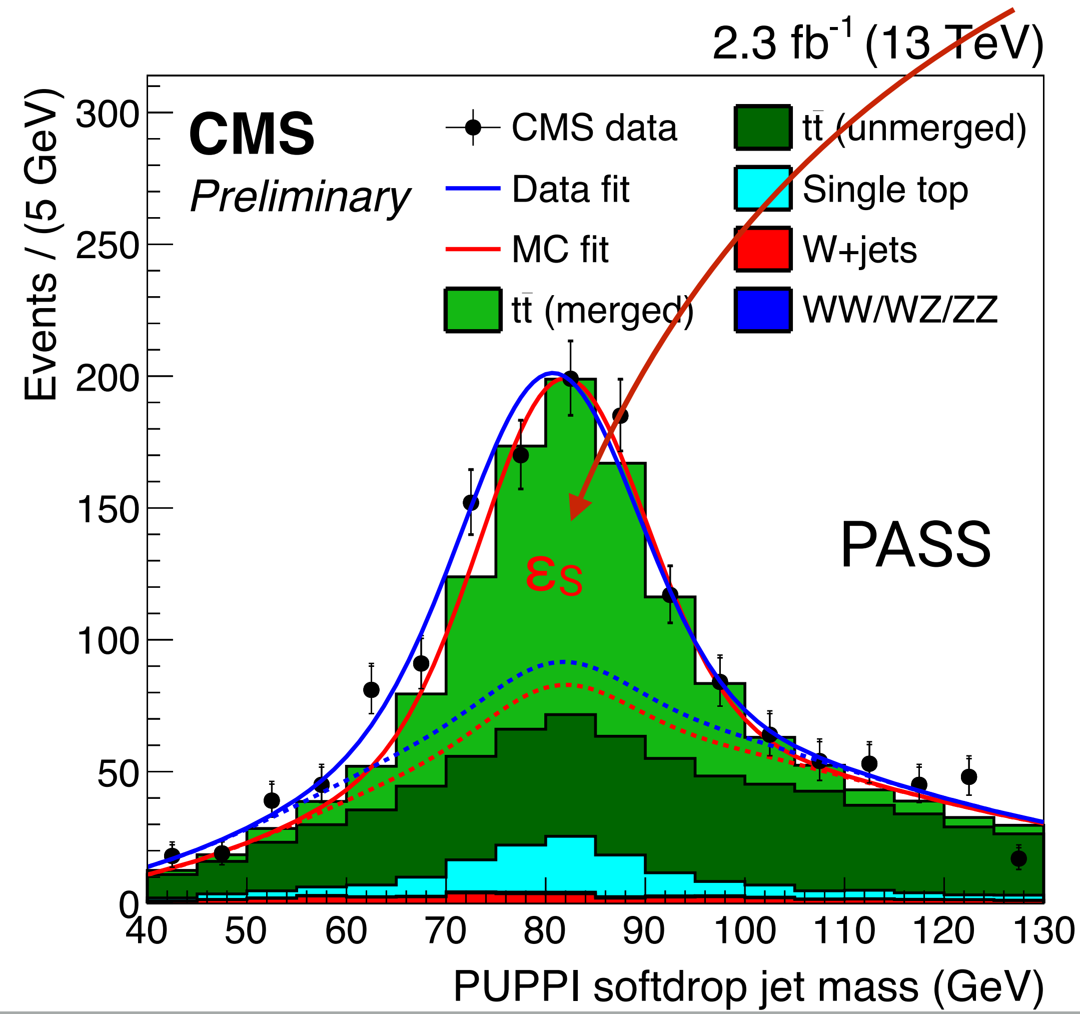
Developing a new V-tagger: Efficiency scalefactors

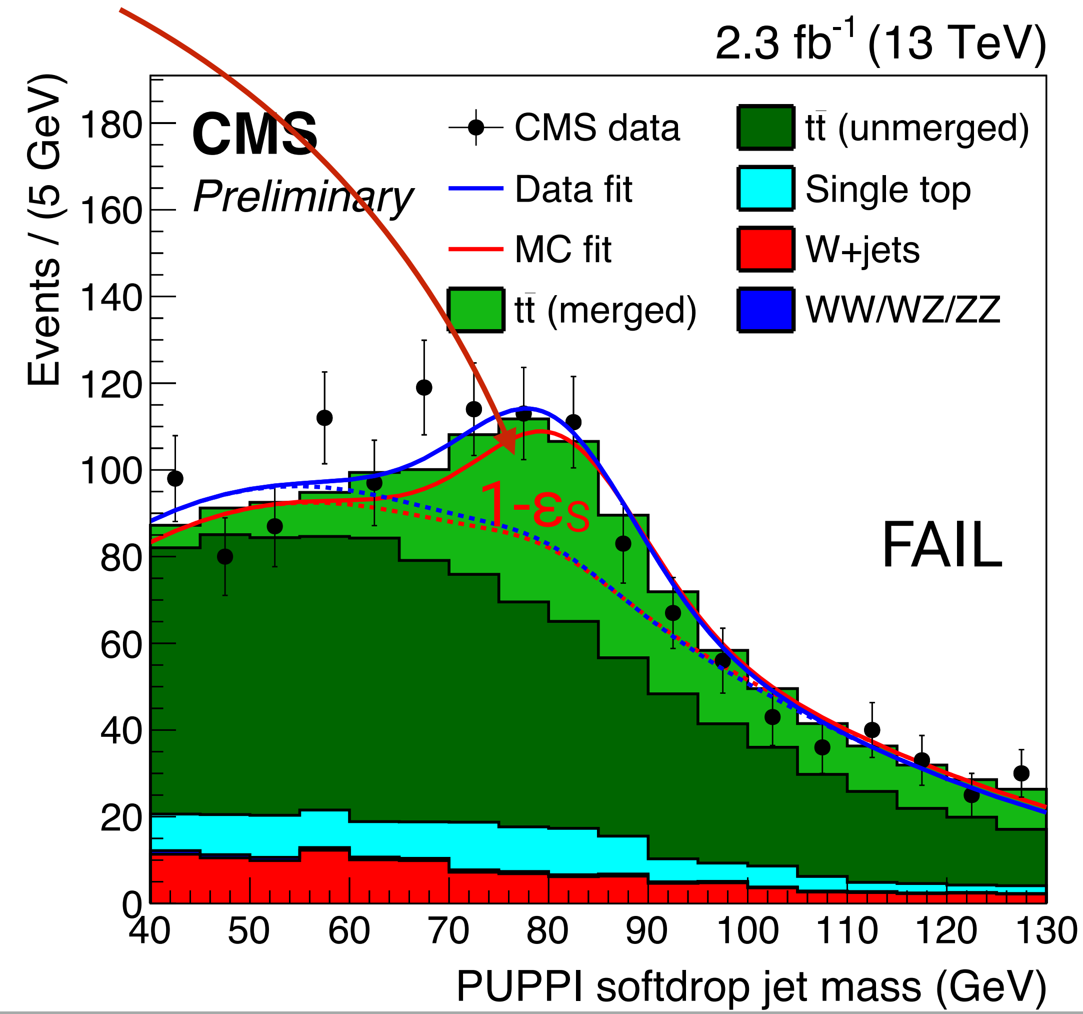


Efficiency (ε_S), jet mass scale and resolution estimated in semileptonic $t\bar{t}$

- Simultaneous fit of τ_{21} pass(<X) and fail (>X)

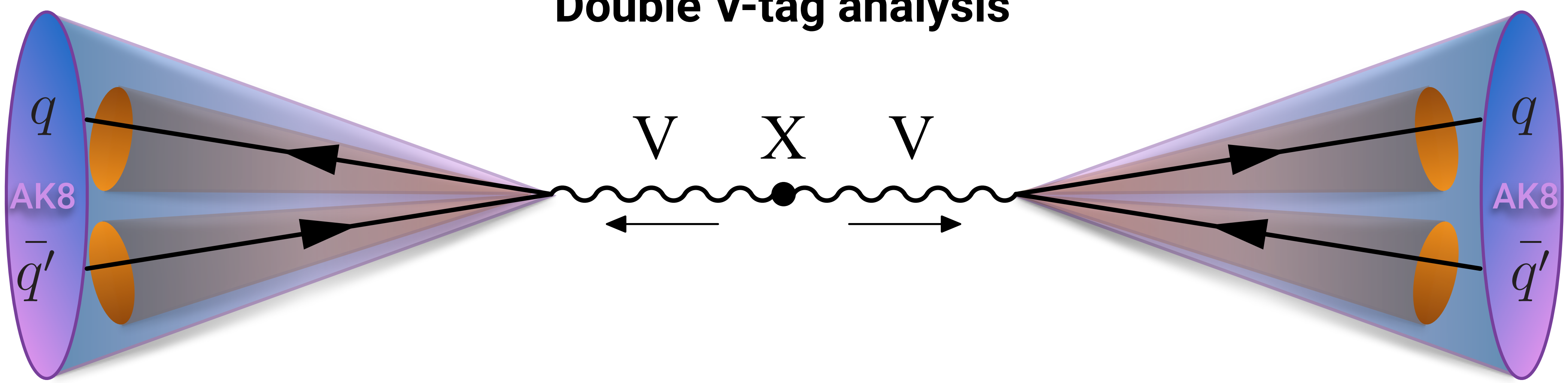
Signal efficiency, jet mass scale and resolution from Gaussian component of total fit



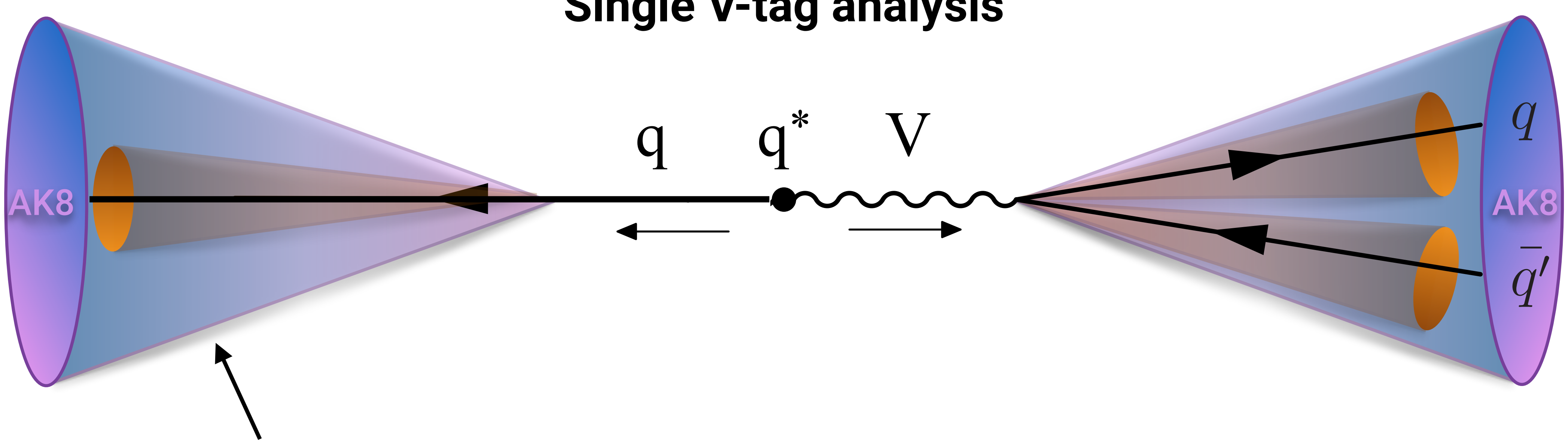


Analysis strategy

Double V-tag analysis

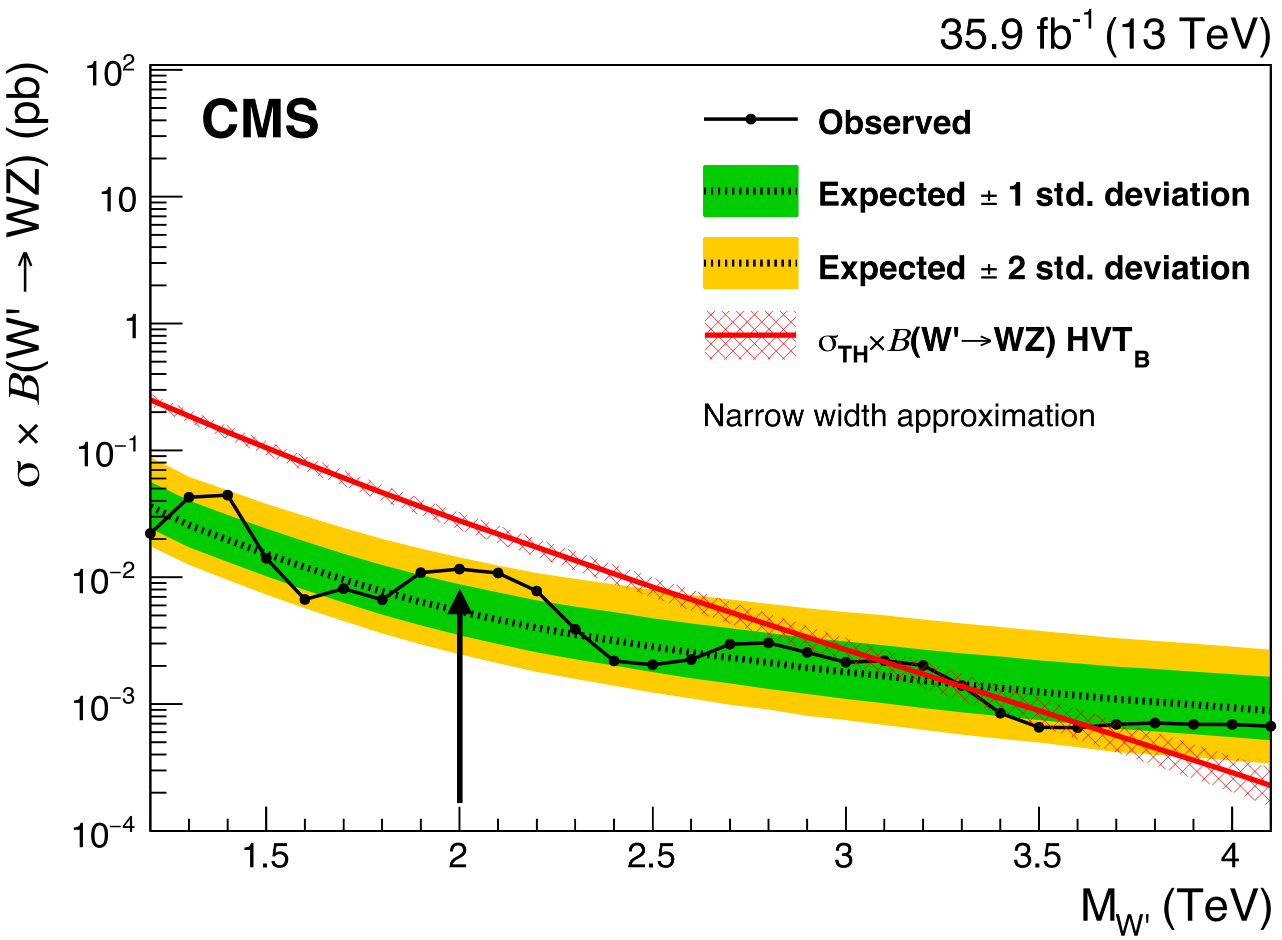


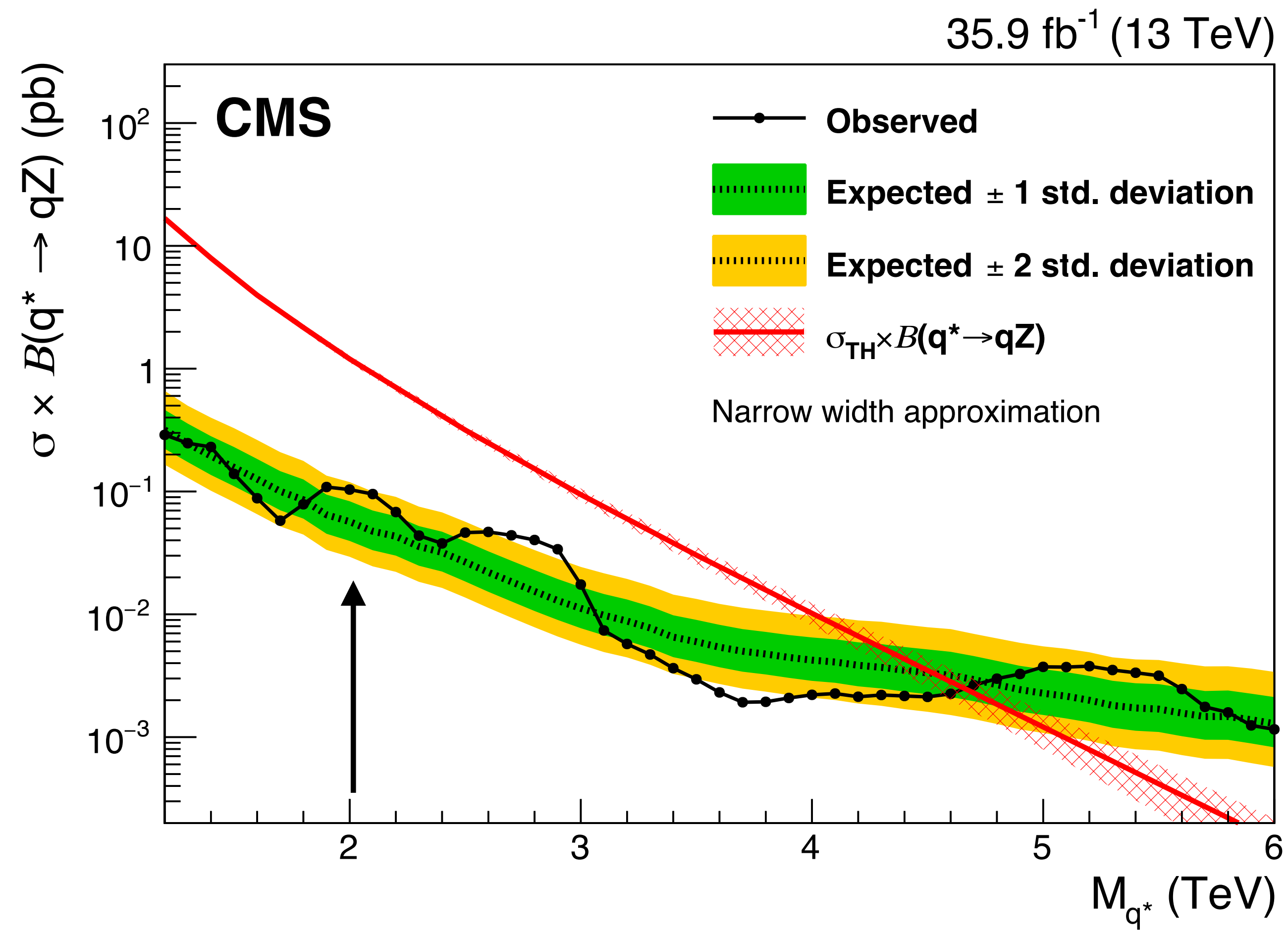
Single V-tag analysis



Adding search for excited quarks decaying to qV by removing W-tag.
Never before analysed channel at 13 TeV!

Results



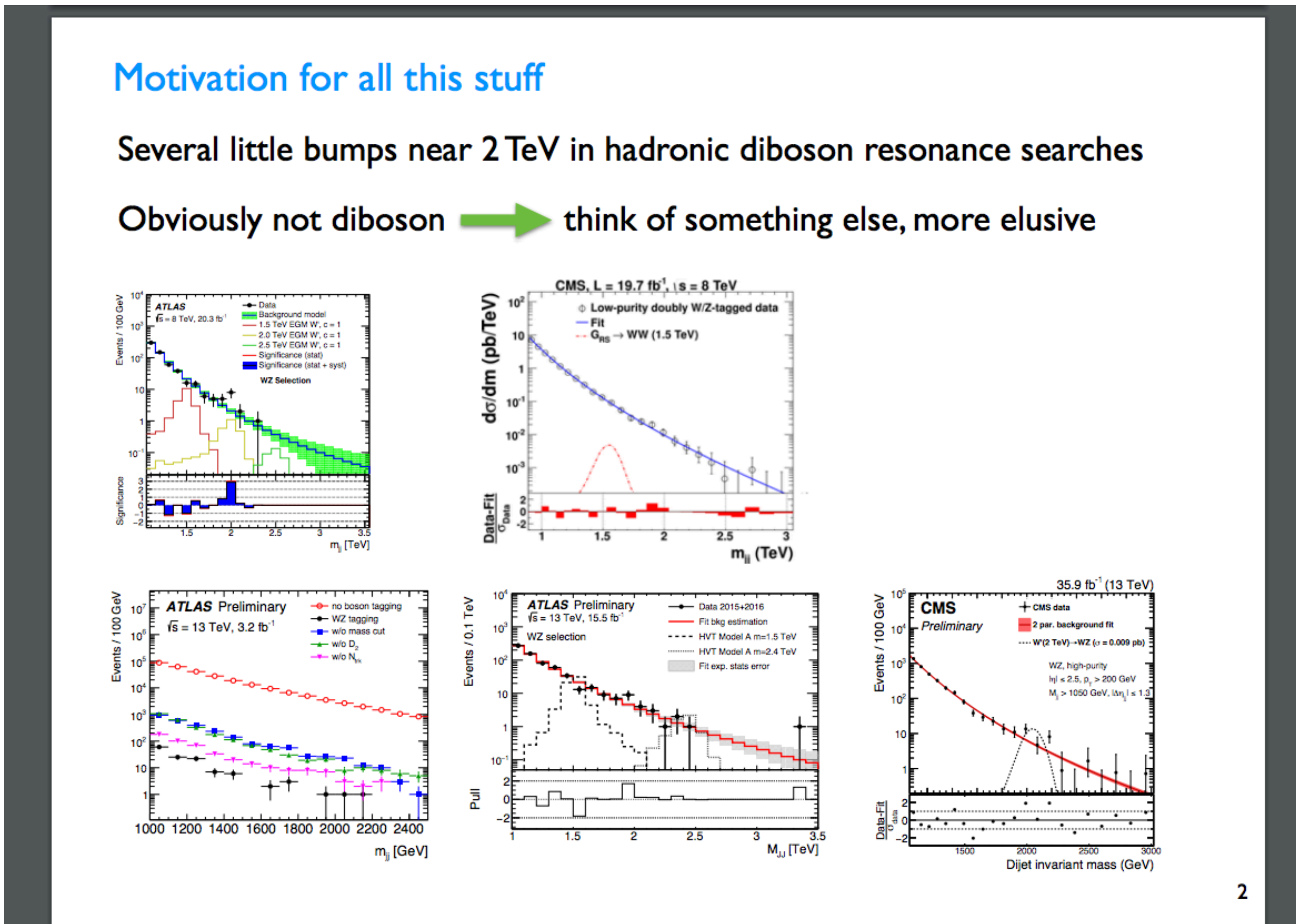


Excluding vanilla signal models (BulkG, V', q*), but still see (statistically insignificant) enhancements around 2-3 TeV in qV and VV.

What now?

Interlude: Stealth bosons

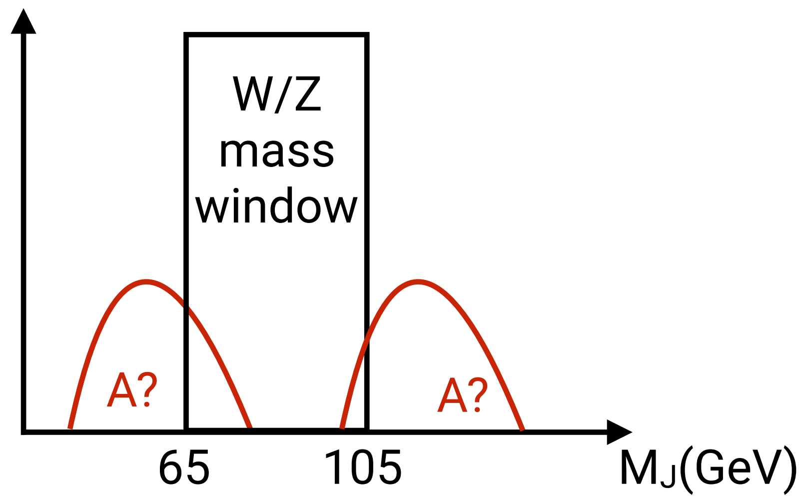
Slide from J.A Aguilar-Saavedra:
“Stealth bosons and where to find them”
(BOOST 2018)

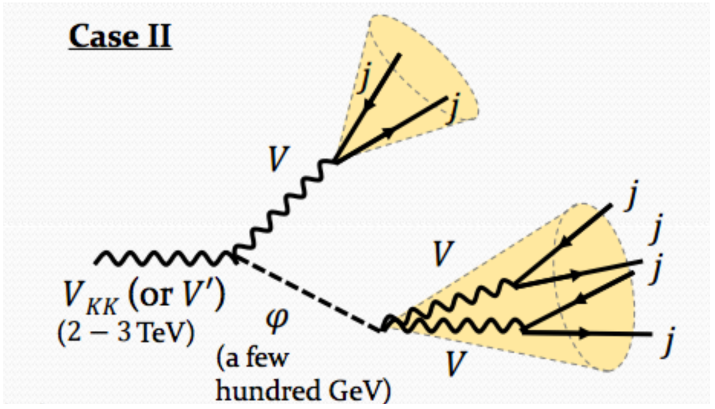


Interlude: Stealth bosons

What if tiny excesses were not due to diboson resonances, but something else

- catching tail of other non-SM boson?
- not necessarily 2-, but N-pronged?

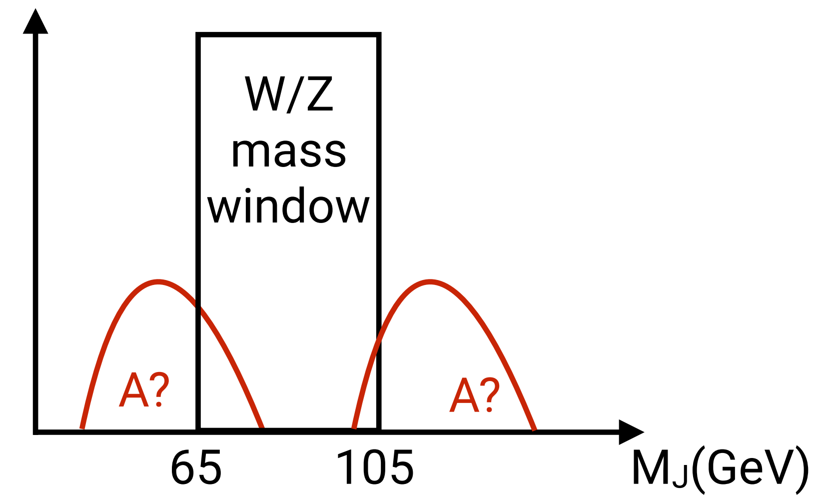


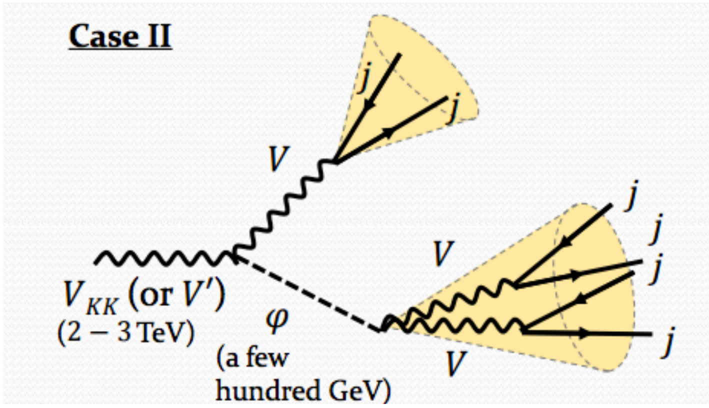


Interlude: Stealth bosons

What if tiny excesses were not due to diboson resonances, but something else

- catching tail of other non-SM boson?
- not necessarily 2-, but N-pronged?



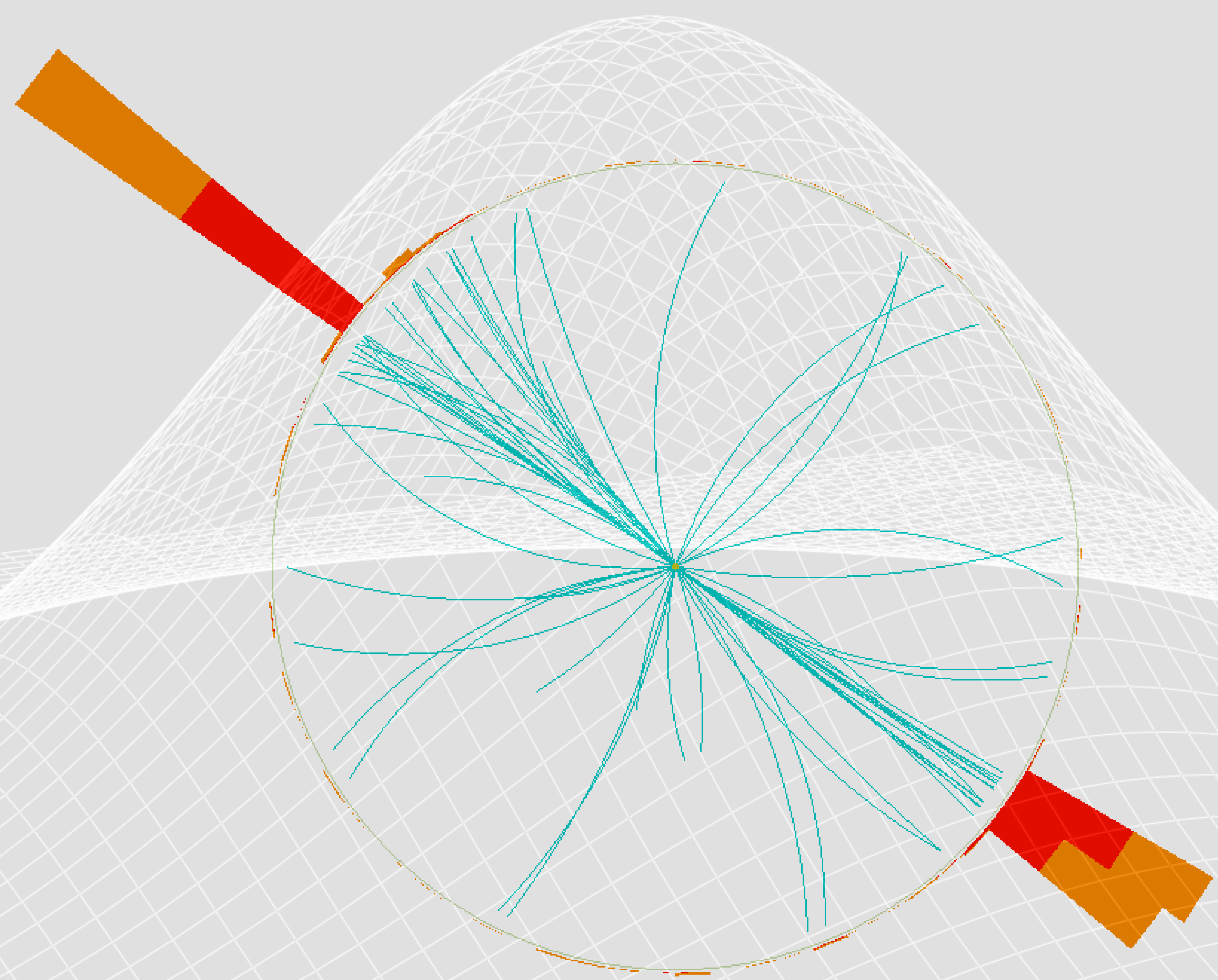


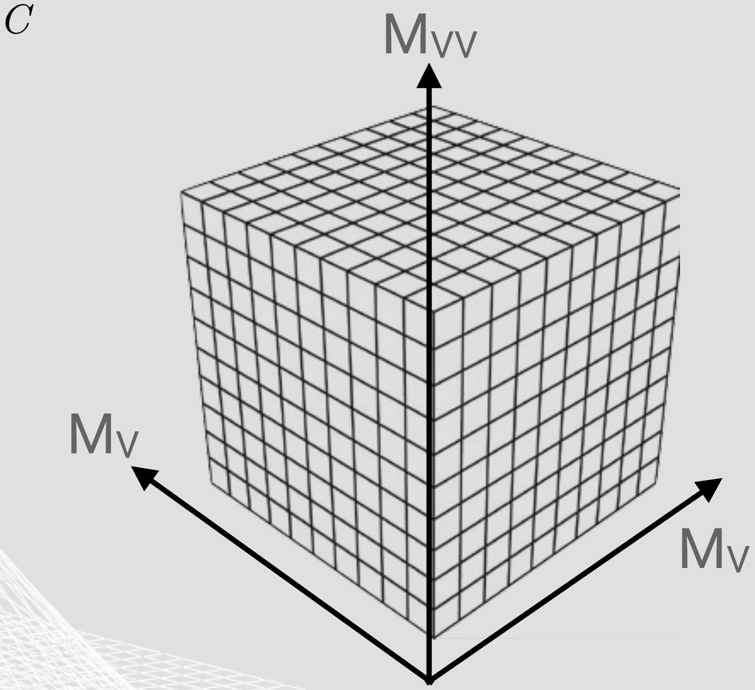
→ Make framework for easily scanning full jet groomed mass spectrum which would

- yields gain in sensitivity for VV analysis
- allow to search for VV/VH/HH and non-SM bosons anywhere in softdrop mass spectrum

Search III:
A novel framework for multi-dimensional searches

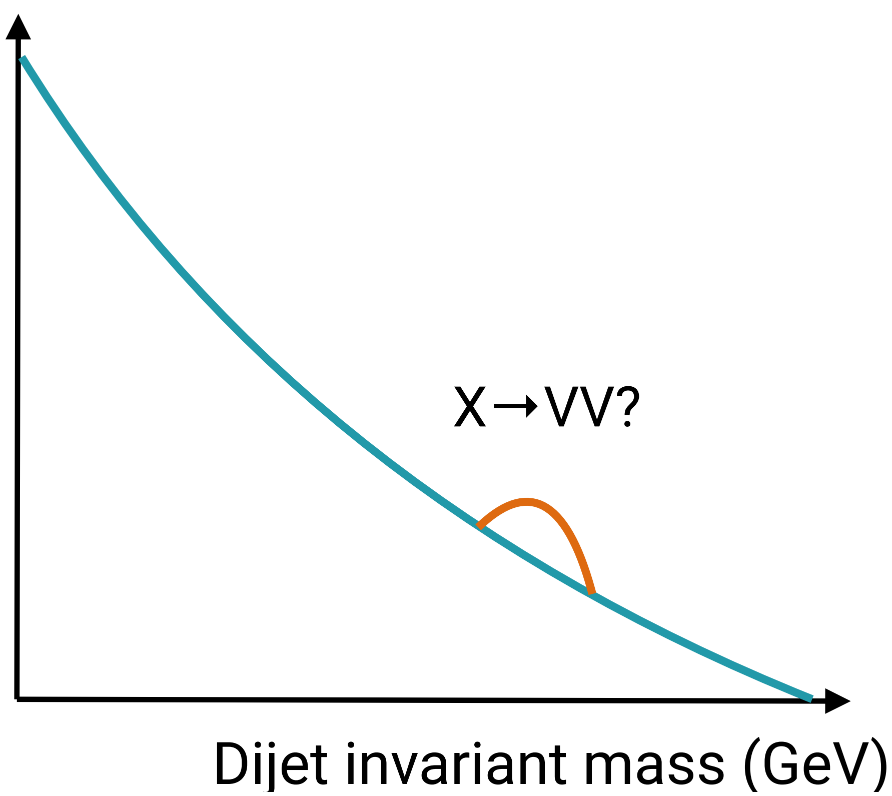
In progress. To be submitted to The European Physical Journal C





~ Paper introducing a novel three-dimensional search method allowing for simultaneously searching for W/Z/H peaks, and eventually non-SM bosons, in the softdrop jet mass spectrum. To be published with full 2016+2017 dataset, ~80 fb^{-1}

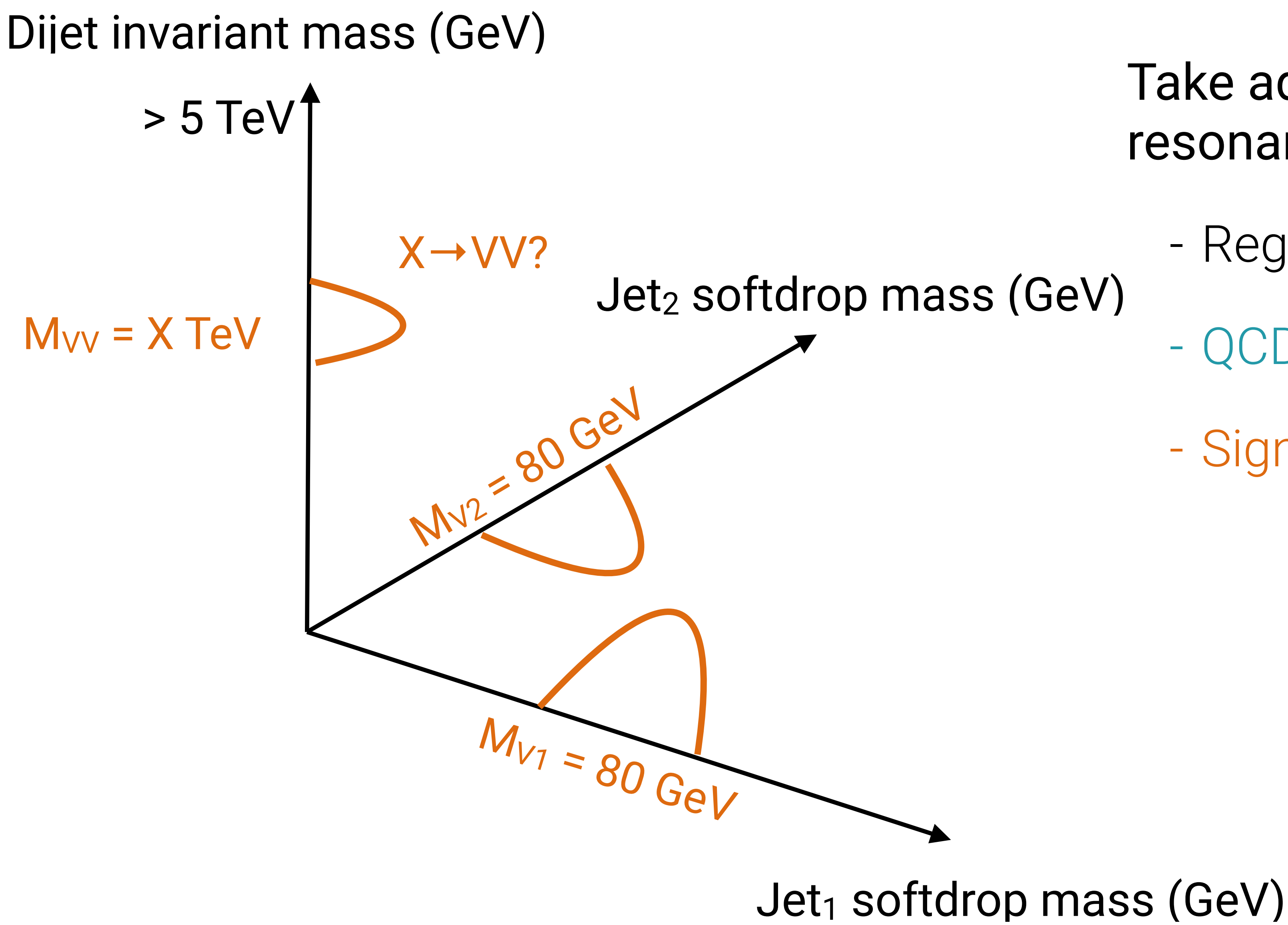
Three-dimensional VV



Until now: two AK8 jets with groomed mass between 65-105 GeV and $\tau_{21} < X$

- Region of interest is dijet invariant mass
- QCD background estimated from smooth fit to data signal region using “dijet fit”
- Signal parametrised with double CB

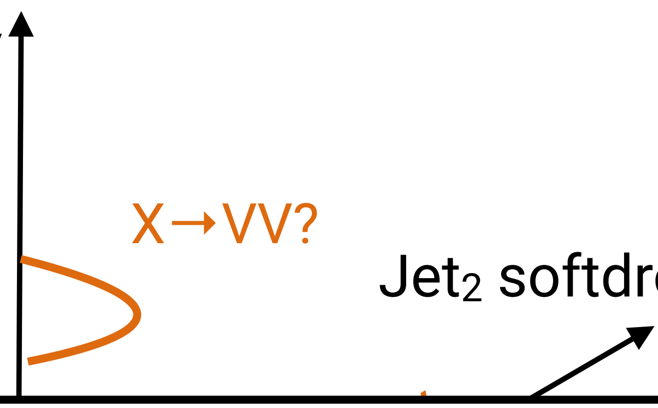
Three-dimensional VV



Take advantage of the fact that signal is resonant in 3D: M_V, M_V and M_{VV}

- Region of interest is M_{V1}-M_{V2} - M_{VV} plane
- QCD background starting from simulation
- Signal parametrised with 3D PDF

Three-dimensional VV



Take advantage of the fact that signal is resonant in 3D: M_V, M_V and M_{VV}

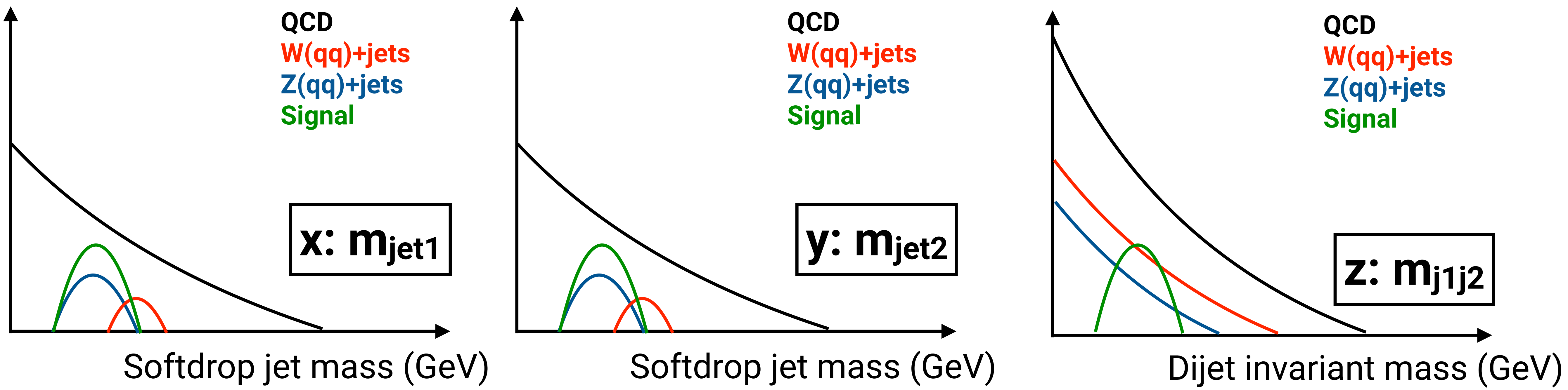
- Region of interest is M_{V1}-M_{V2} - M_{VV} plane
- QCD background starting from simulation
- ed with 3D PDF

The pros of this procedure:

1. Can model turn-ons in m_{jj}/m_{jet}
2. Can take jet mass and dijet mass correlations fully into account
3. Larger signal acceptance without mass cuts
4. Opens door to scan full groomed mass spectrum in one analysis

Building PDFs

4 steps to full model:

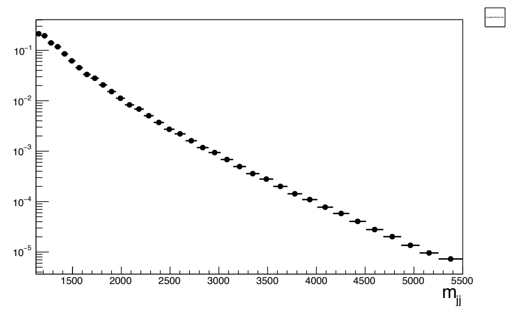


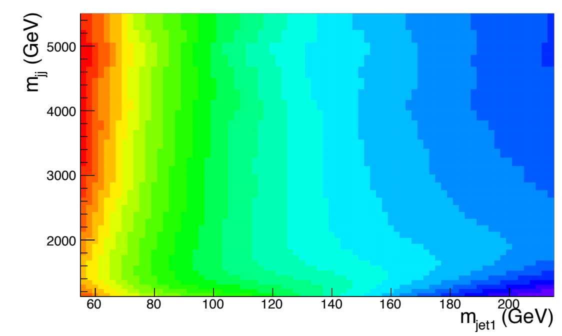
1. Signal 3D PDF
   - Resonant in x, y and z
2. Background, non-resonant
   - Non-resonant in x, y and z
   - Dominant background
3. Background, resonant
   - W/Z+jets, resonant in x+y
4. Alternate PDFs
   - 5 additional shape uncertainties

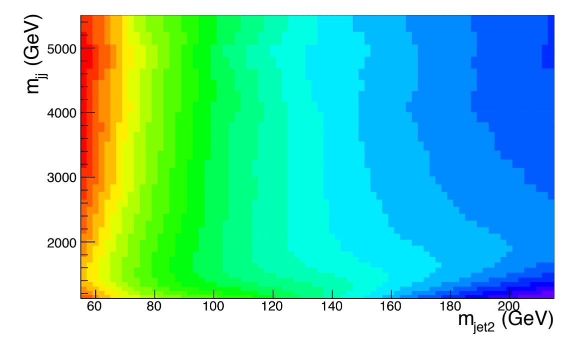
Replacing the dijet fit

To account for correlations m_{jet}/m_{jj}, non-resonant background modelled conditionally

- $P_{non\text{-}res}(m_{jj}\,, m_{jet1}, m_{jet2}) = \quad P_{jj}\,(\,m_{jj} \mid \theta_1\,) \quad \times \quad P_j\,(m_{jet1} \mid m_{jj}\,, \theta_2\,) \quad \times \quad P_j\,(m_{jet2} \mid m_{jj}\,, \theta_2\,)$







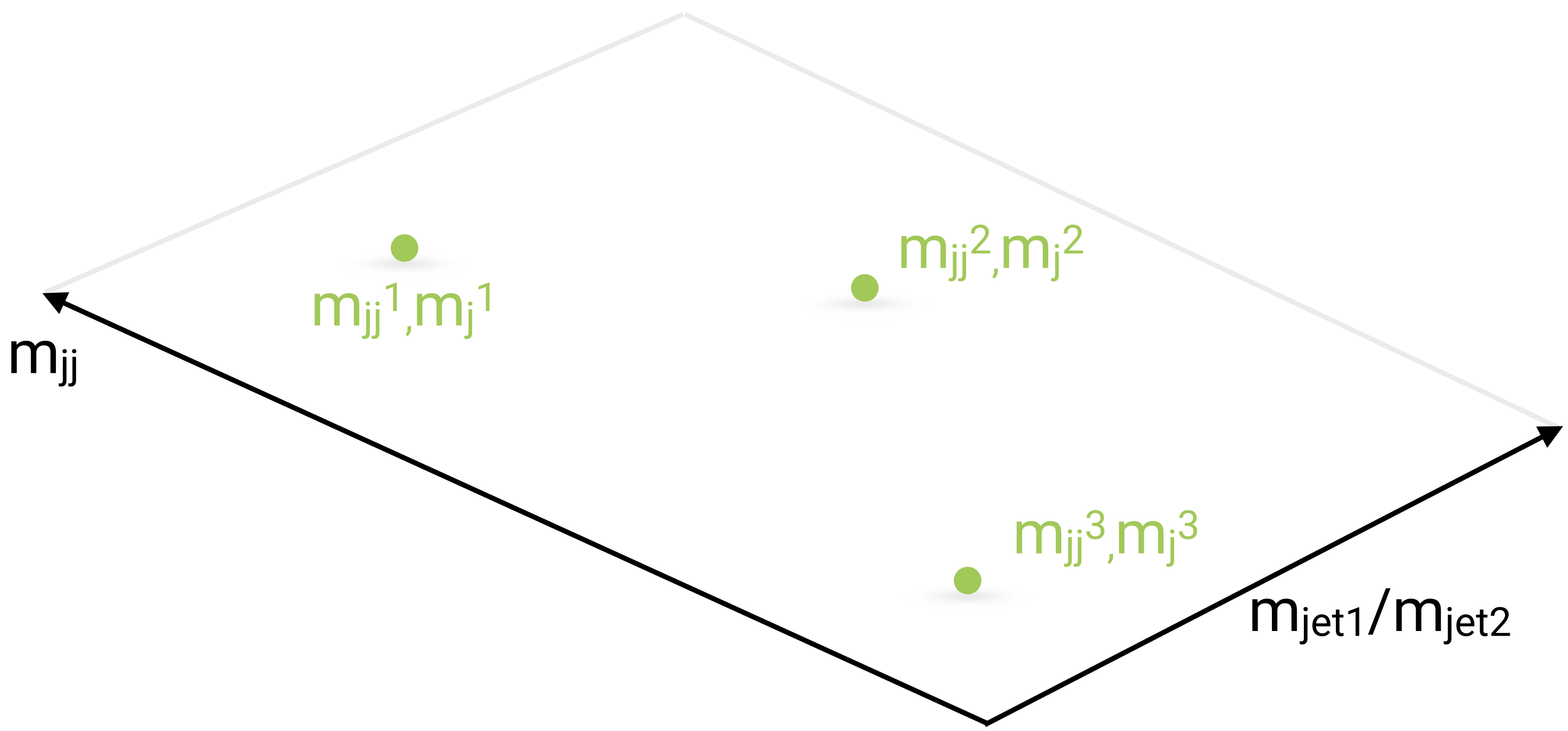
Replacing the dijet fit

To account for correlations m_{jet}/m_{jj}, non-resonant background modelled conditionally

- $P_{non\text{-}res}(m_{jj}, m_{jet1}, m_{jet2}) = P_{jj}(m_{jj} \mid \theta_1) \times P_j(m_{jet1} \mid m_{jj}, \theta_2) \times P_j(m_{jet2} \mid m_{jj}, \theta_2)$

250k bins, need to ensure smooth and full shape → kernel approach

- rather than filling 1D/2D histogram with m_{jet}, m_{jet}/m_{jj} (sparse), let each event contribute 1D/2D gaussian kernel defined through generator level quantities



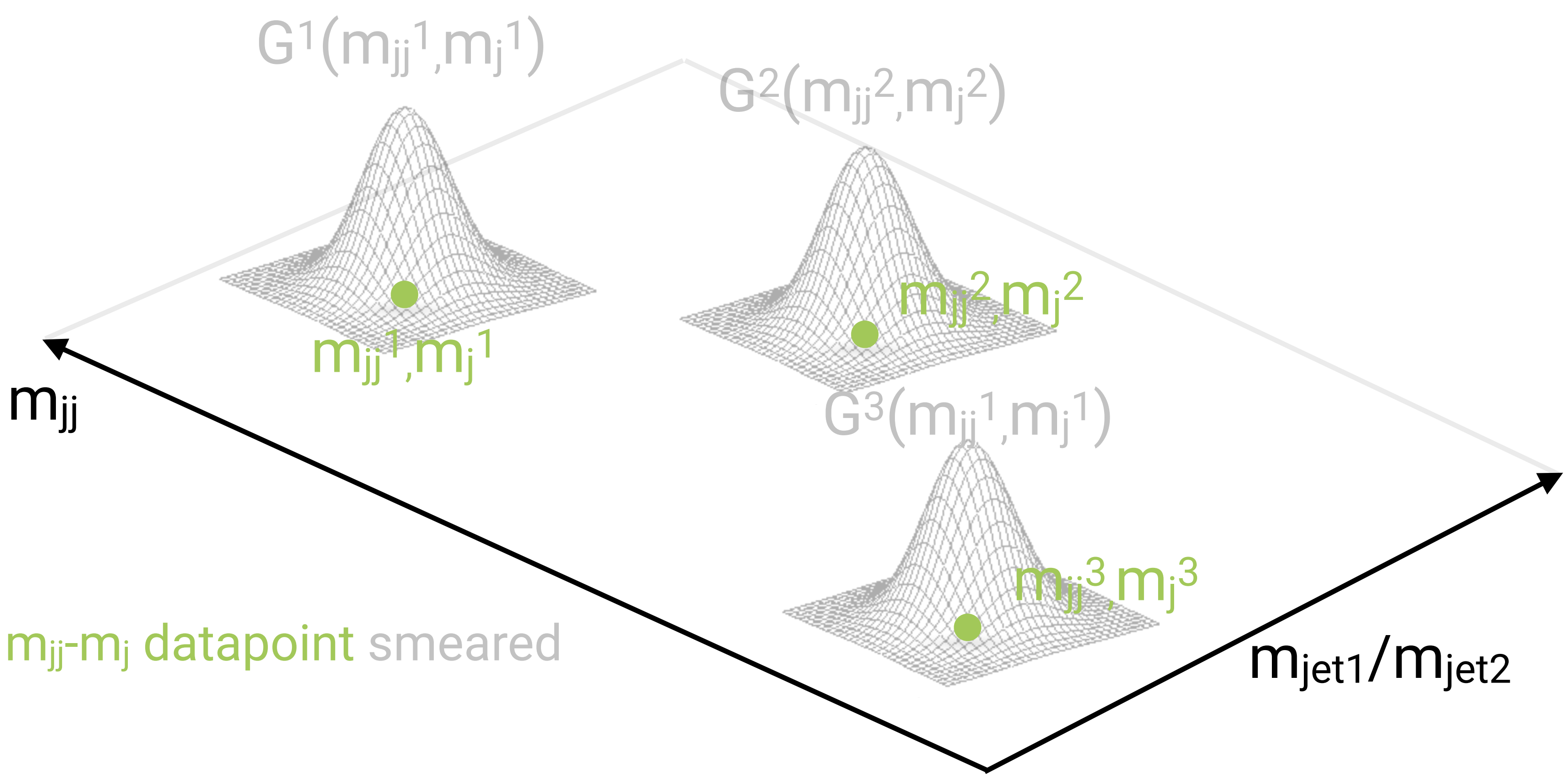
Replacing the dijet fit

To account for correlations m_{jet}/m_{jj}, non-resonant background modelled conditionally

- $P_{non\text{-}res}(m_{jj}, m_{jet1}, m_{jet2}) = P_{jj}(m_{jj} \mid \theta_1) \times P_j(m_{jet1} \mid m_{jj}, \theta_2) \times P_j(m_{jet2} \mid m_{jj}, \theta_2)$

250k bins, need to ensure smooth and full shape → kernel approach

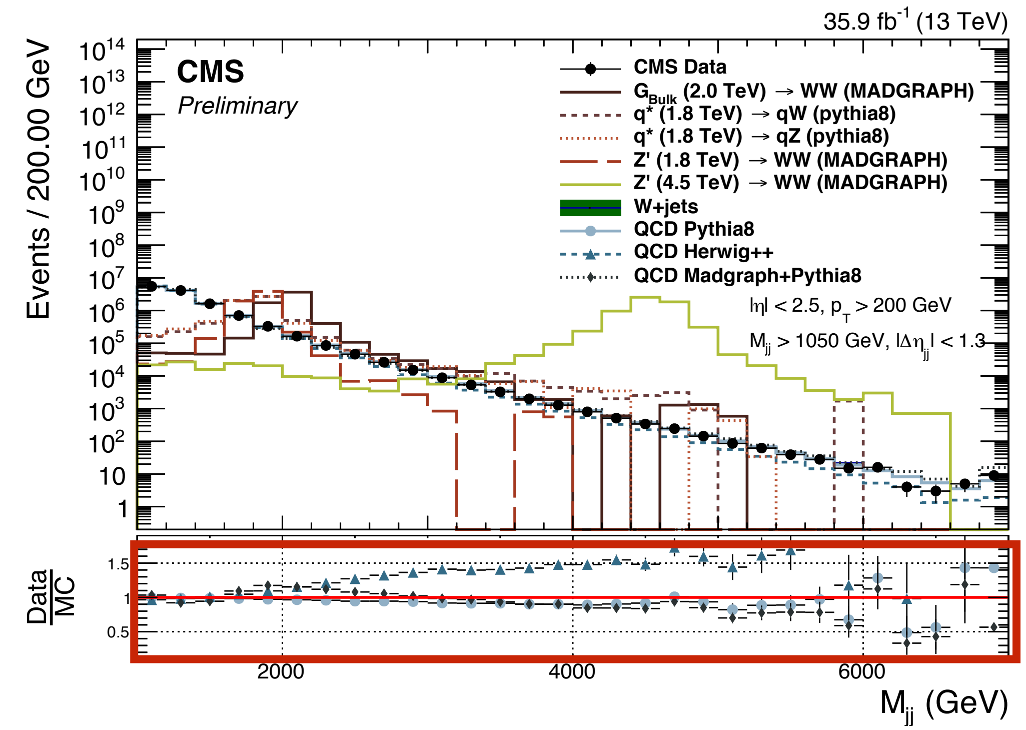
- rather than filling 1D/2D histogram with m_{jet}, m_{jet}/m_{jj} (sparse), let each event contribute 1D/2D gaussian kernel defined through generator level quantities



Replacing the dijet fit

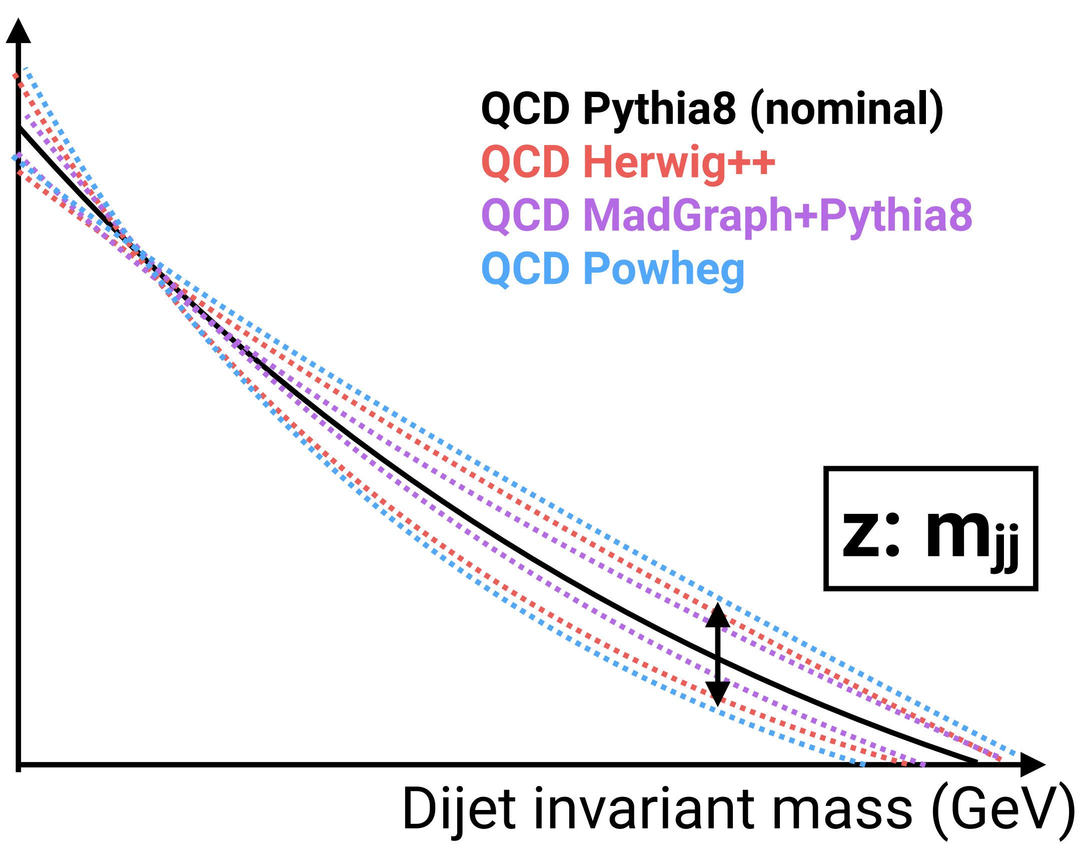
Is Nature Herwig++, MadGraph or Pythia?
LO(Pythia) or NLO (Powheg)?

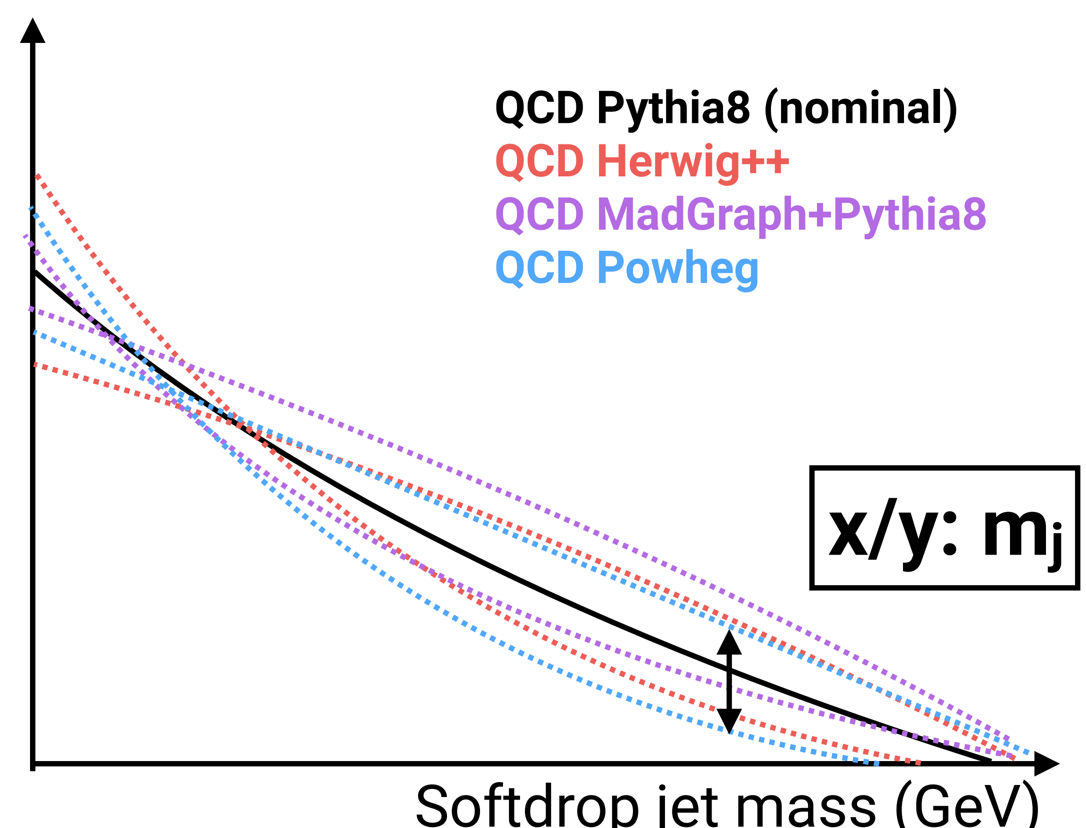
- predictions disagree, let's allow it to be all!



Add alternate shapes based on different QCD MC, simultaneously affecting m_{j1}, m_{j2}, m_{jj}

- PDF can take any shape to match data!



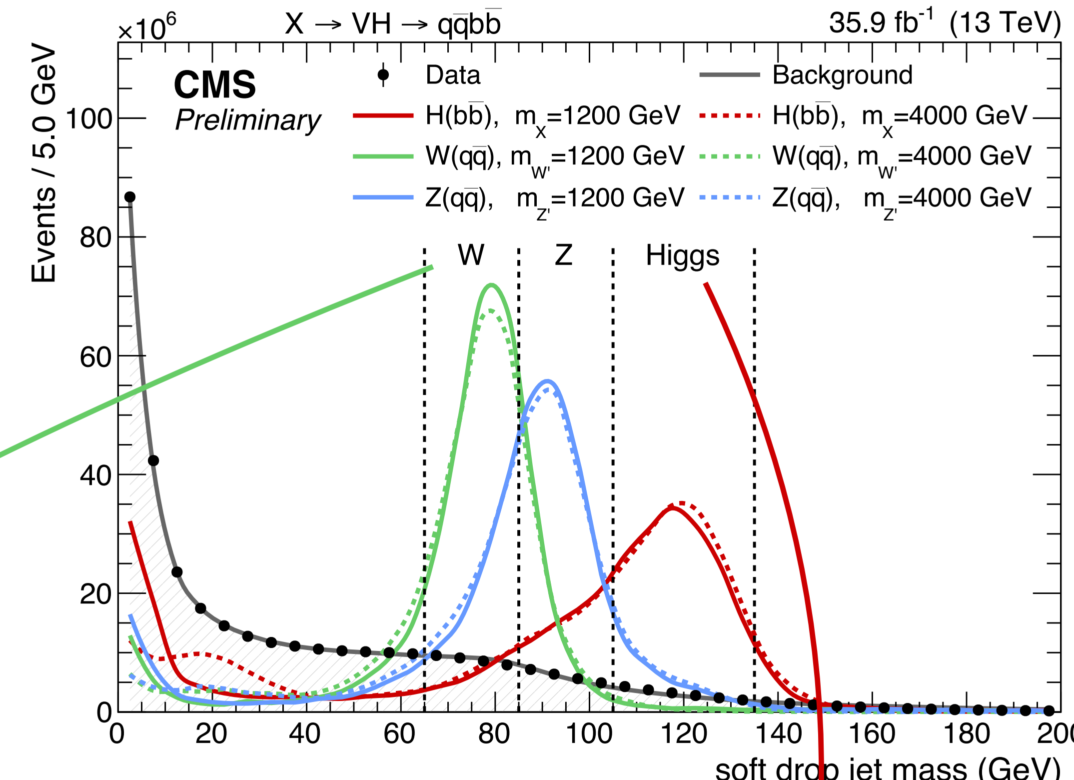


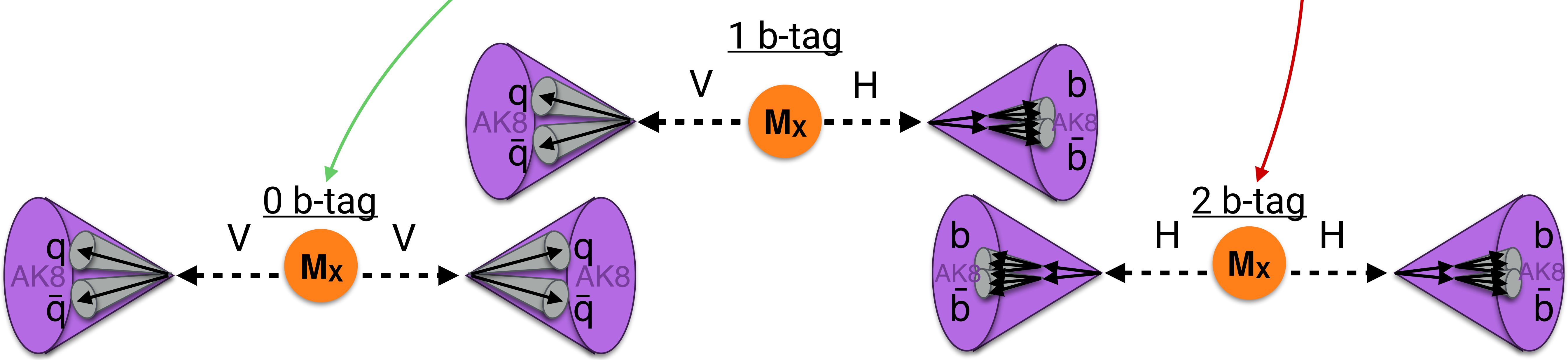
Plan forward

3D method improve 1D sensitivity by 20-30%

For full 13 TeV dataset (16+17+18)

- VV, VH(bb) + HH
  in one analysis!





And tribosons?

Do we still need to "scan" τ_N for N-prong signals?

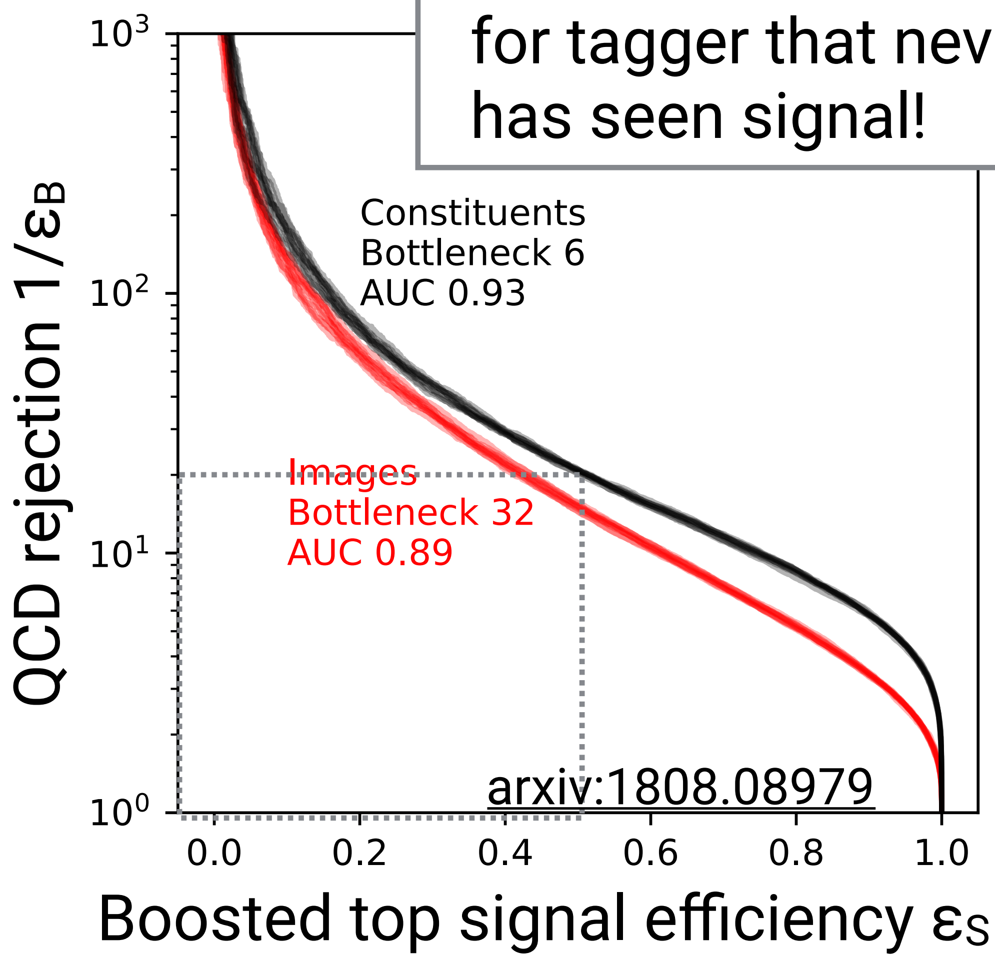
No! Switch from τ_{21} to generic anti-QCD tagger

- Deep Neural Networks trained to learn how "QCD-like" event is [1] [2]
- identify signal without having seen it, ideal for model independent searches

Combined with 3D fit, one background model for any signal peaking in softdrop+dijet mass

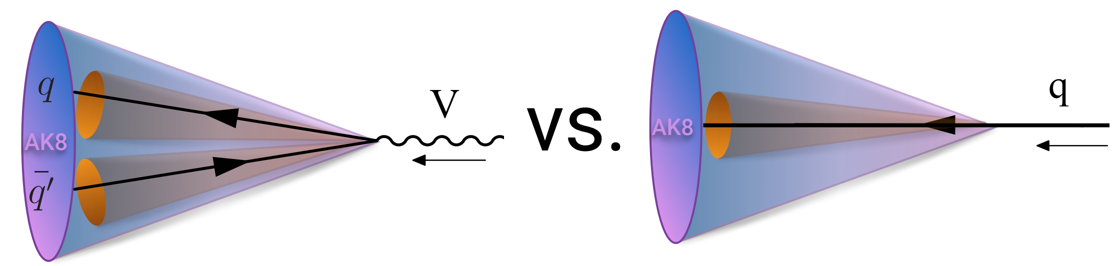
- truly scanning the full m_{j1}-m_{j2}-$m_{j1,j2}$ plane!

50% ε_S at 5% ε_B for tagger that never has seen signal!



LoLa: DNN for W-tagging

See more here

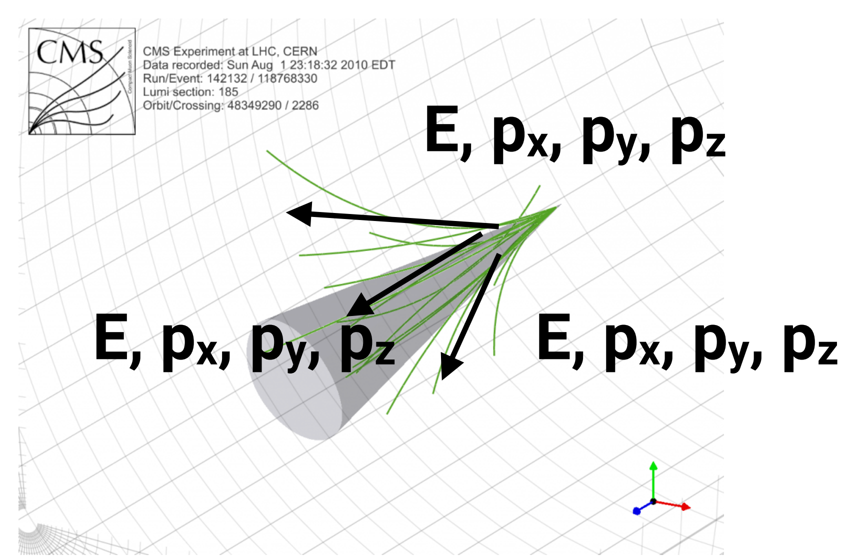


Physics based deep neural network
(first introduced for top tagging)

- look at jet constituent 4-vectors only and teach network Minkowski space and jet clustering

Input:
4-vectors of N = 20
highest p_T jet
constituents
of AK8 jets

$$\begin{array}{l} E_i \ldots E_N \\ p_x \\ p_y \\ p_z \end{array} \quad (k_{\mu,i}) = \begin{pmatrix} k_{0,1} & k_{0,2} & \cdots & k_{0,N} \\ k_{1,1} & k_{1,2} & \cdots & k_{1,N} \\ k_{2,1} & k_{2,2} & \cdots & k_{2,N} \\ k_{3,1} & k_{3,2} & \cdots & k_{3,N} \end{pmatrix}$$

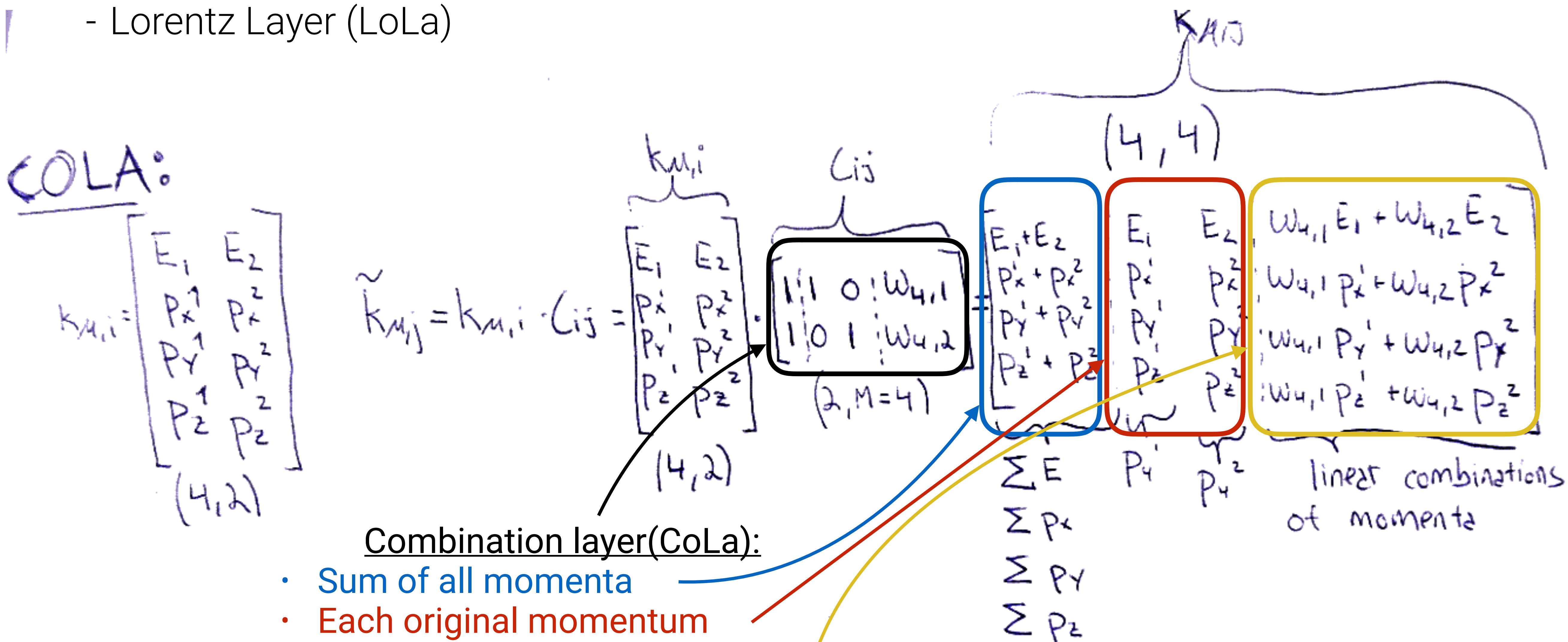


LoLa: DNN for W-tagging

See more here

4 layer deep neural network, 2 custom layers:

- Combination Layer (CoLa)
- Lorentz Layer (LoLa)



Combination layer(CoLa):

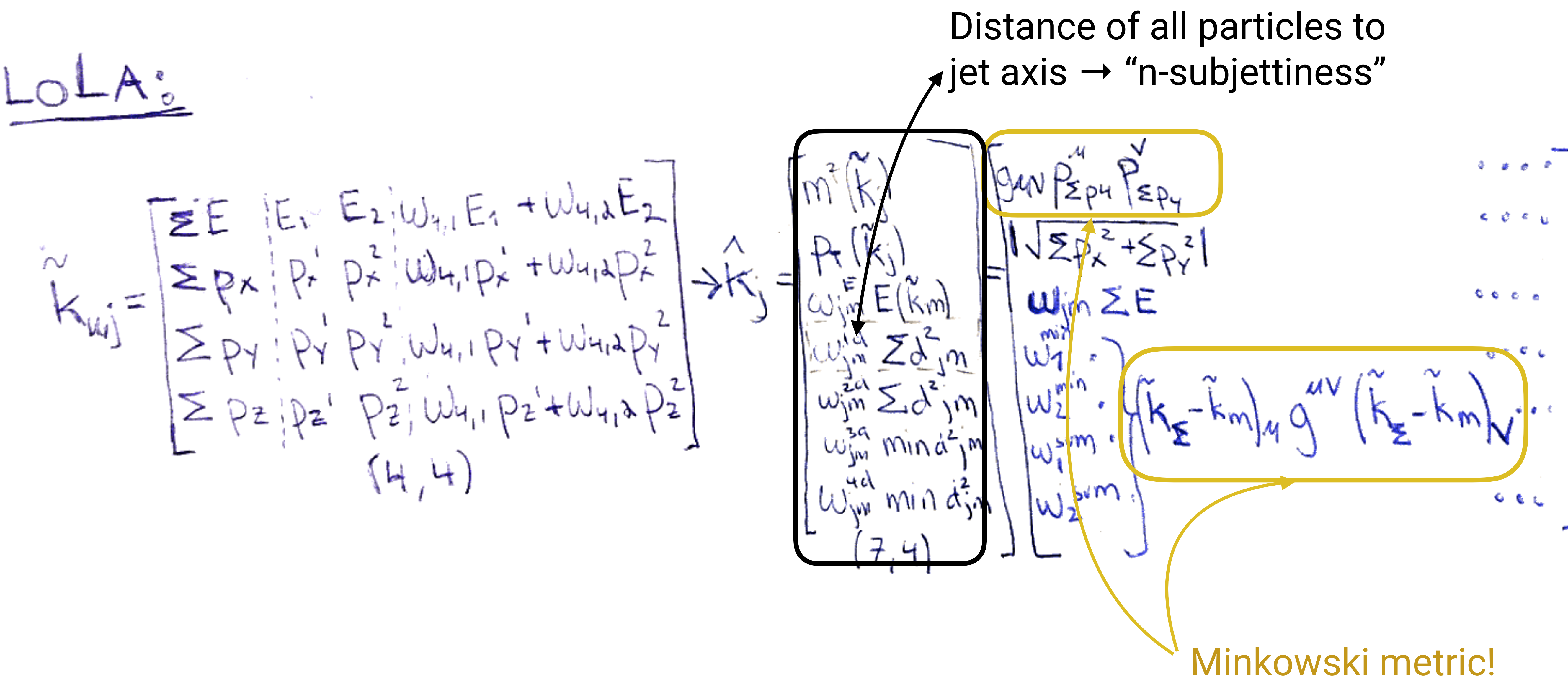
- Sum of all momenta
- Each original momentum
- Linear combination of particles with trainable weights

LoLa: DNN for W-tagging

See more here

4 layer deep neural network, 2 custom layers:

- Combination Layer (CoLa)
- Lorentz Layer (LoLa)



Encoding jet substructure

LoLa output is Prob(QCD) and Prob(W), trained with QCD and W signal

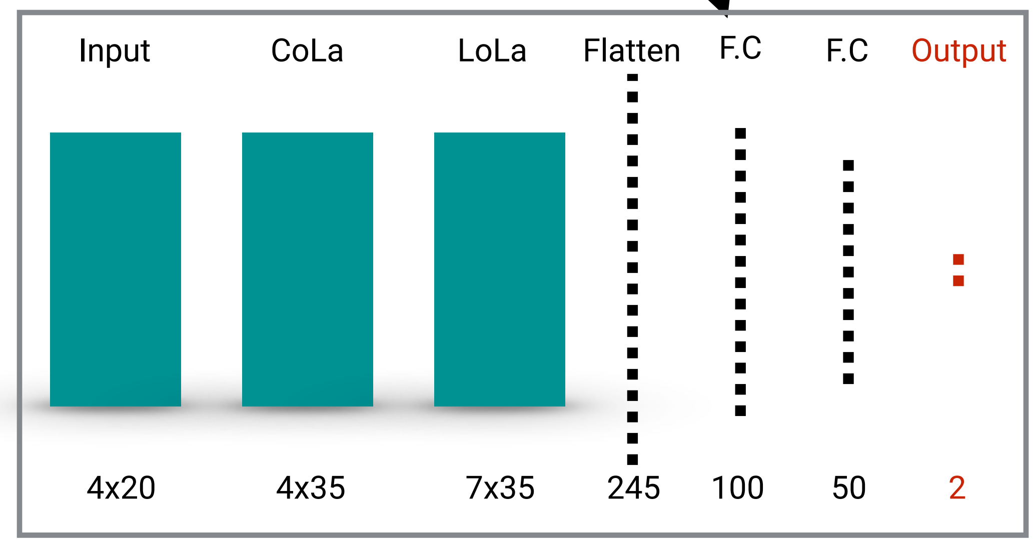
Instead, train DNN to reconstruct QCD jet constituent 4-vectors by itself

- compress LoLa output to smaller dimensional space, “encode QCD”, then make DNN blow up to 4-vectors again
- Novel anti-QCD tagger based on LoLa demonstrated <u>here</u>: Auto-encoding jet substructure!

Hope to see auto-encoder LoLa for boosted generic searches in near future!

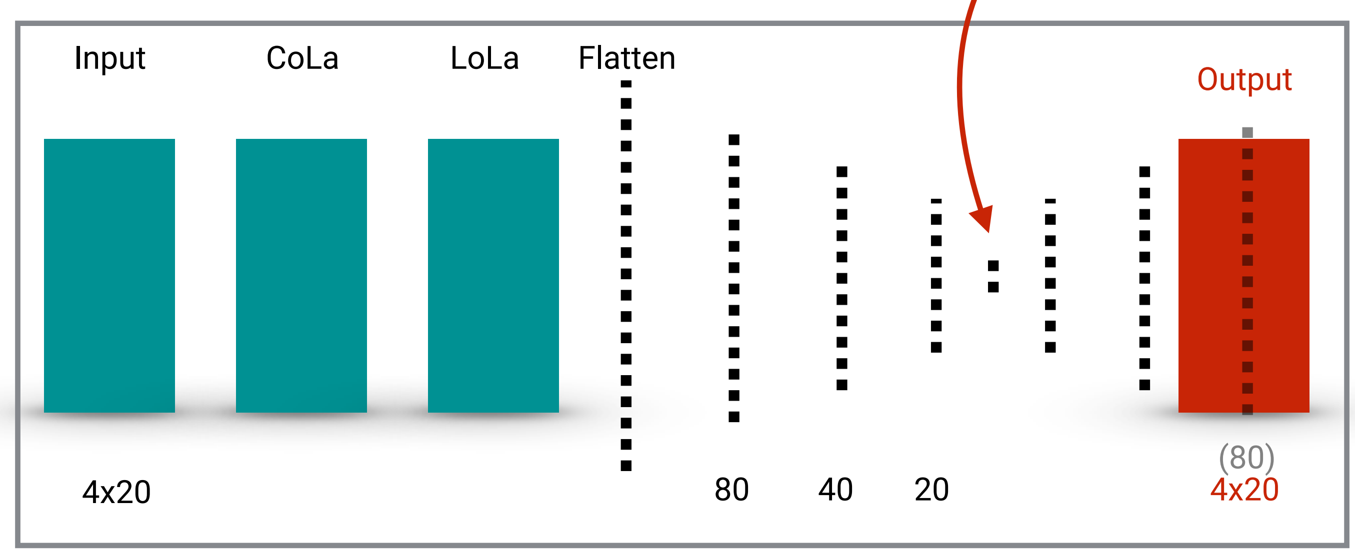
$$\sigma\left(\underbrace{\begin{bmatrix} W_{1,1} & W_{2,1}\\ W_{1,2} & W_{2,2}\end{bmatrix}}_{\text{trainable weights}} \begin{bmatrix}x_1\\ x_2\end{bmatrix} + \underbrace{\begin{bmatrix}b_1\\ b_2\end{bmatrix}}_{\text{trainable biases}}\right)$$

Vanilla LoLa:



Bottleneck:
Encode QCD jets in smaller dim. space!

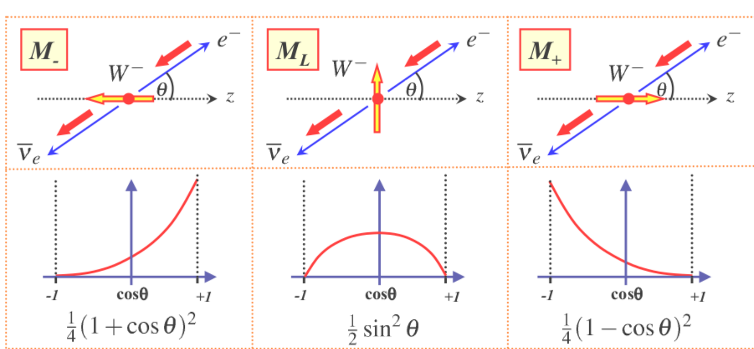
Auto-encoder LoLa:

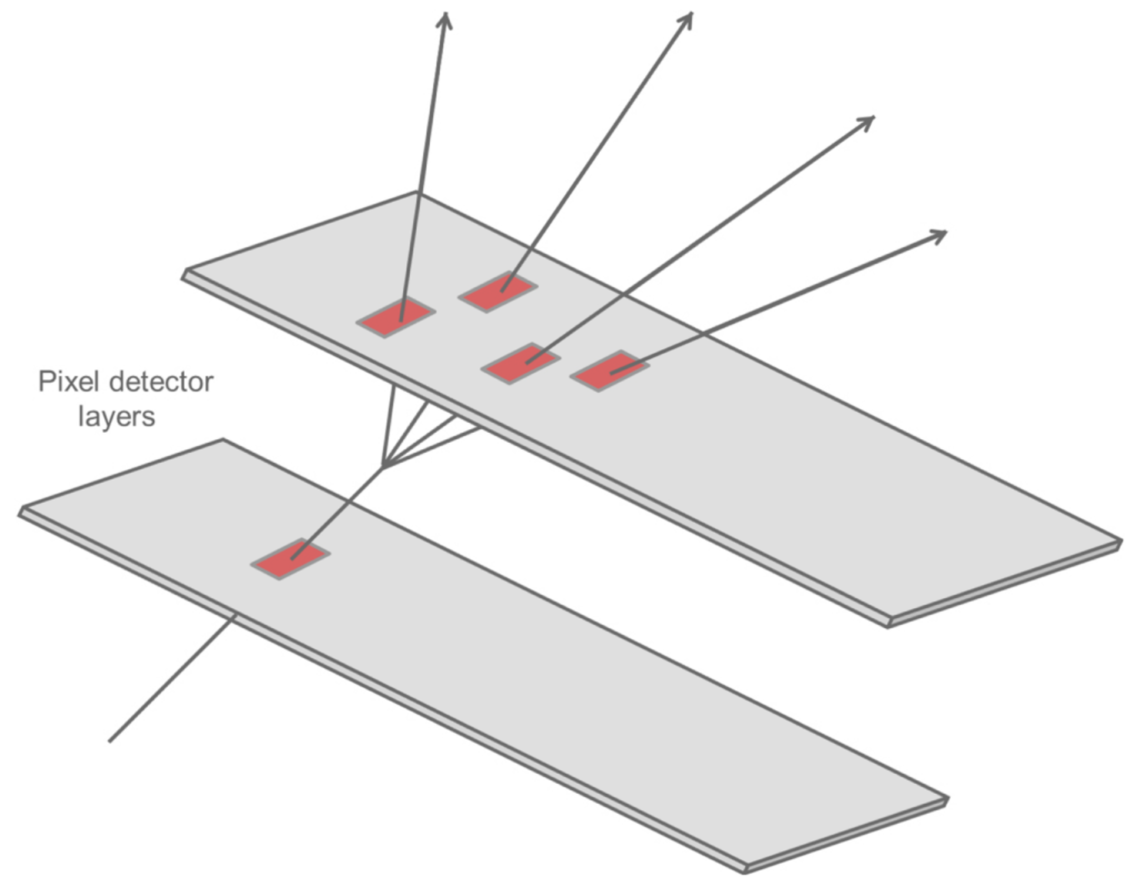


Outlook and ideas:
Ultra-high boosts and precision measurements

Energy enhanced new-physics effects in **longitudinal channel**

$$\frac{\mathcal{A}_{LL}^{\text{SM + BSM}}(q\bar{q} \to WZ)}{\mathcal{A}_{LL}^{\text{SM}}(q\bar{q} \to WZ)} \sim 1 + a_q^{(3)} E^2$$





~As we push limits on BSM to higher and higher resonance masses, need to think of new methods and analyses: How do we deal with b-tagging at extreme p_T, and how can we access BSM signals with increasingly small cross sections and/or high masses?

b-tagging with hits

High-p_T b-quarks can traverse pixel L1 before decaying (in CMS, $p_{T,B}$ >330 GeV)

- tracking fails, drop in b-tagging efficiency

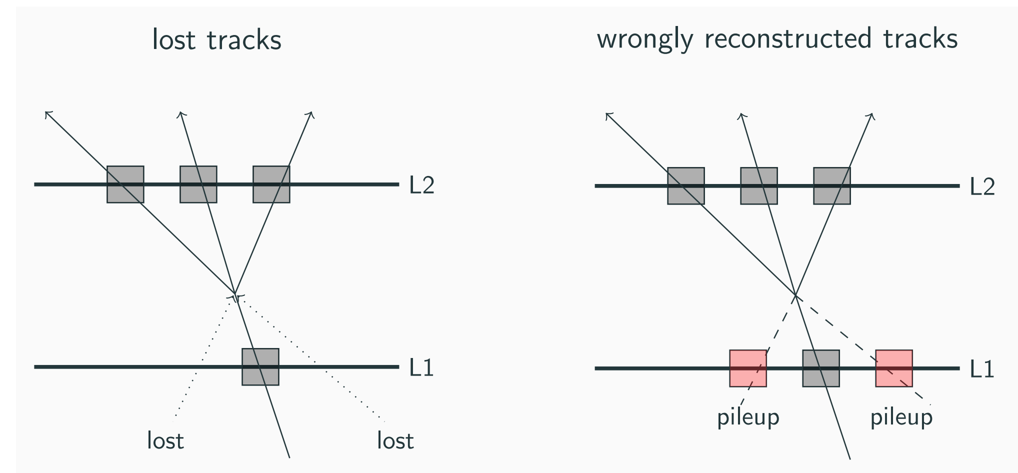
To ensure high b-tag efficiency at HL-LHC

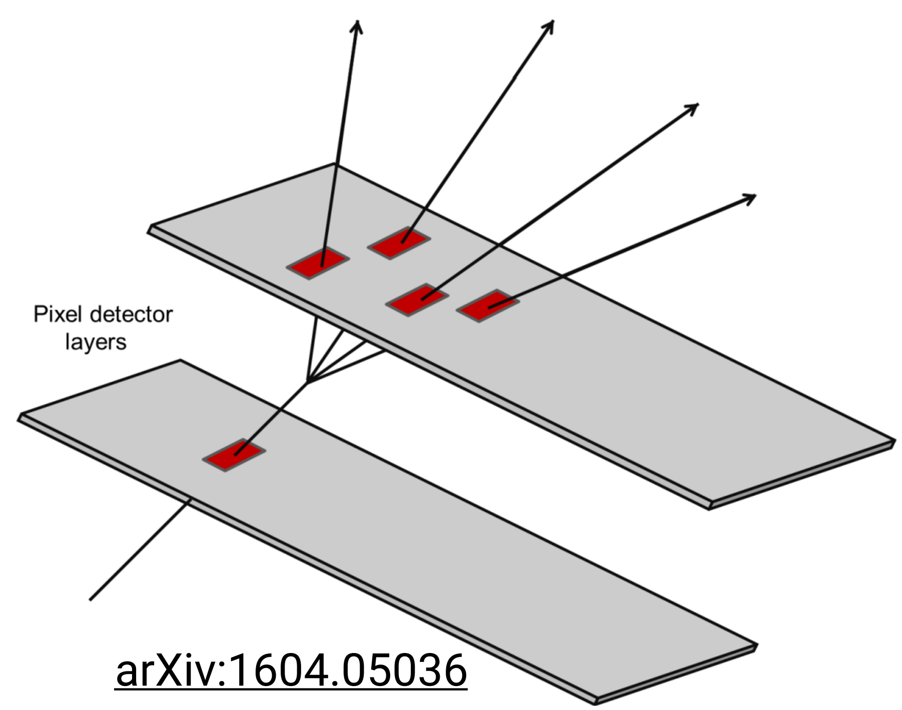
- tag high-p_T B hadrons based on increase in hit multiplicity in pixel layers using DNN

~60% gain in efficiency (112% > 1.2 TeV) (with M. Sommerhalder, Bachelor Student)

Simple! Could be used on hardware at trigger level (eg DNN of FPGAs)?

Efficiency loss for track reco due to missing inner hits!





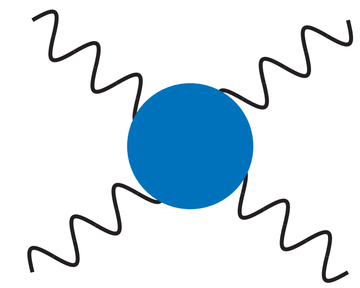
arXiv:1604.05036

WW scattering: W_T vs. W_L

What if we cant directly produce resonances and/or σ_{BSM} small, cannot directly detect?

→ Precision measurements!
BSM interference in 2→2 VV scattering!

$$VV \to VV$$



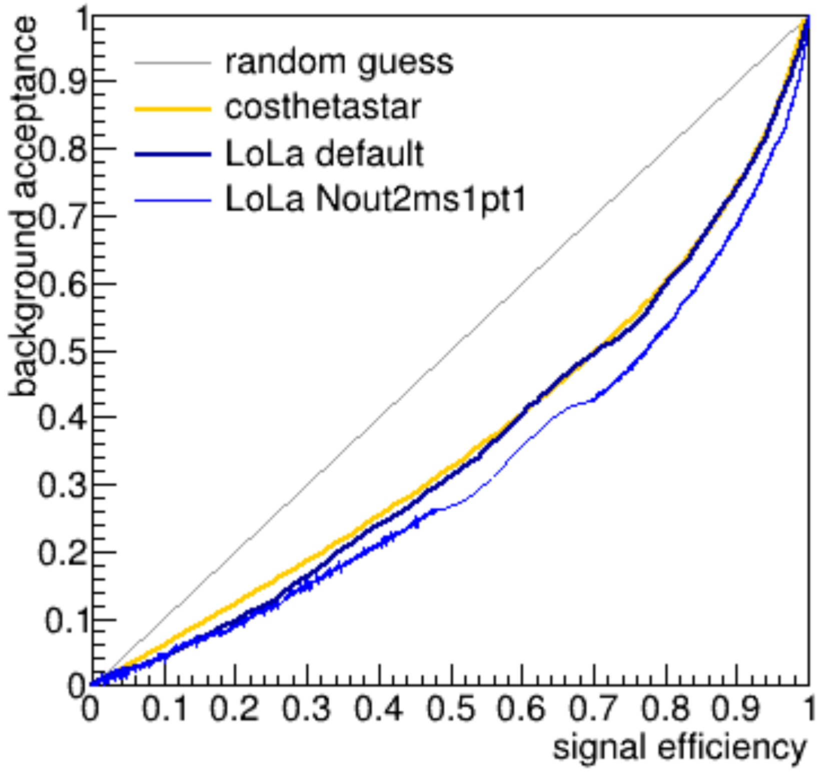
At E>>m_V, New Physics mainly couples to longitudinally polarised W_L

- 90% of SM is W_T, irreducible background!
- important to discriminate W_T and W_L at HL LHC (see G.Panicos talk, Riva et. Al)

Energy enhanced new-physics effects in **longitudinal channel**

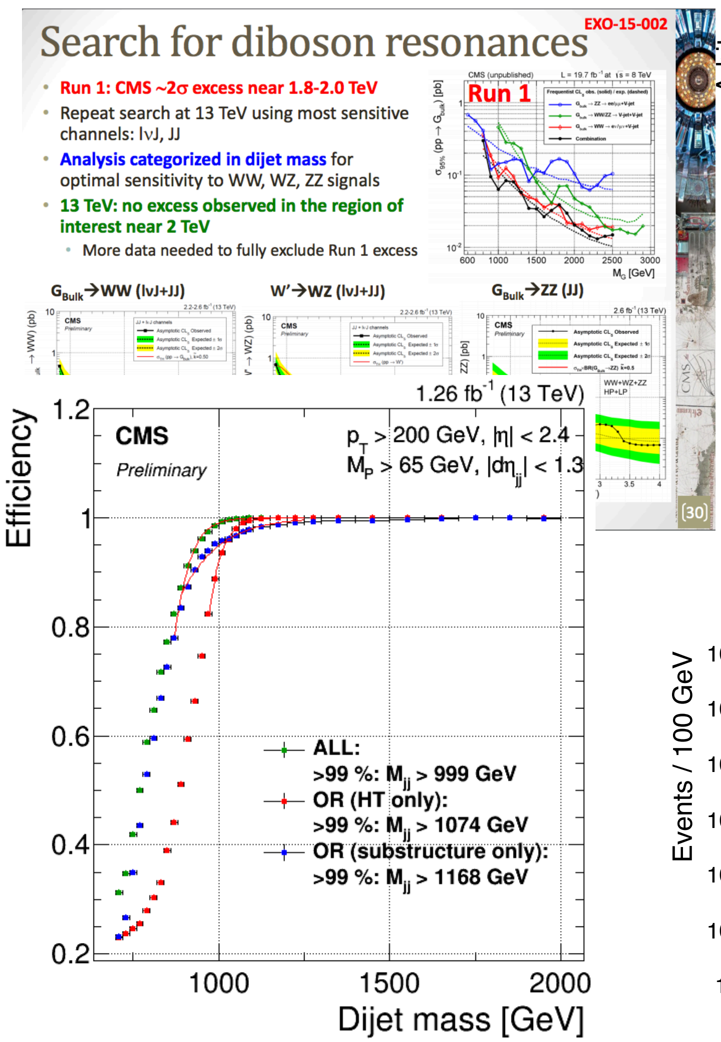
$$\frac{\mathcal{A}_{LL}^{\mathrm{SM\,+\,BSM}}(q\bar{q} \to WZ)}{\mathcal{A}_{LL}^{\mathrm{SM}}(q\bar{q} \to WZ)} \sim 1 + a_q^{(3)} E^2$$

Train LoLa to discriminate between W_T and W_L jets (w. J. Boer, CERN Summer Student)

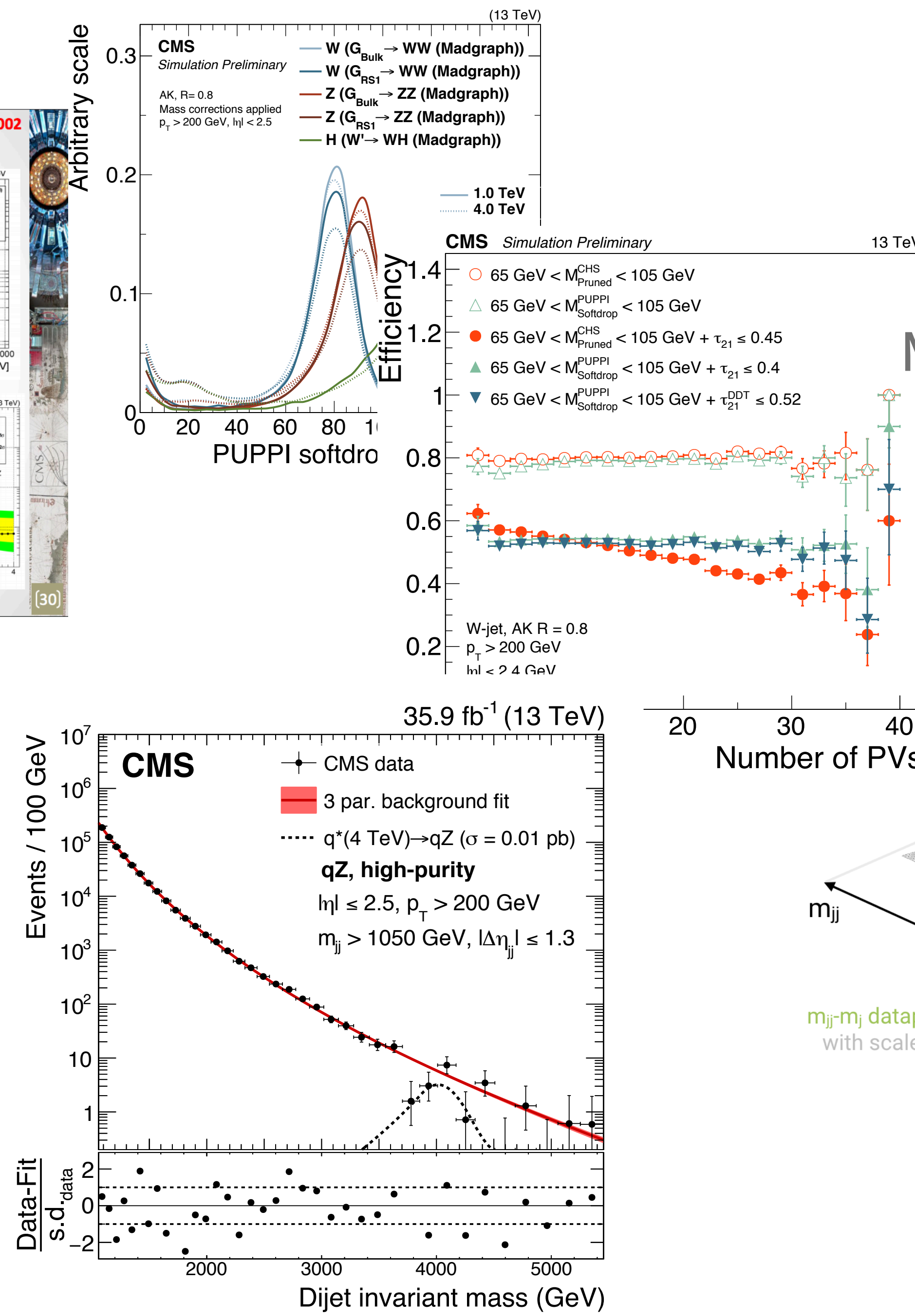


Summary and outlook

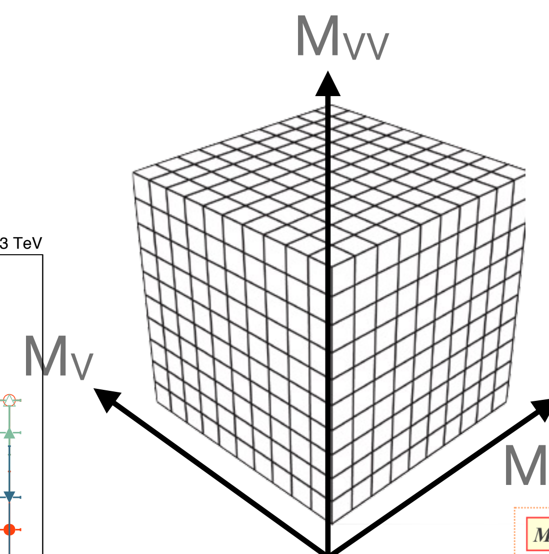
2015

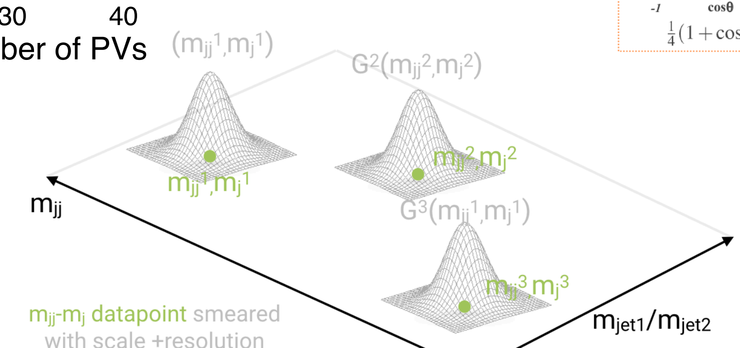


2016

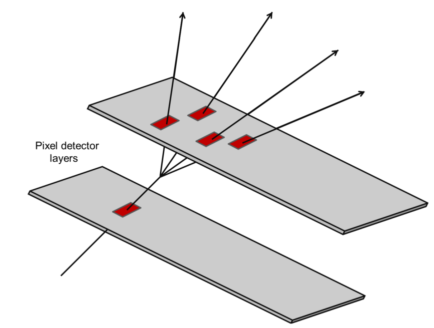


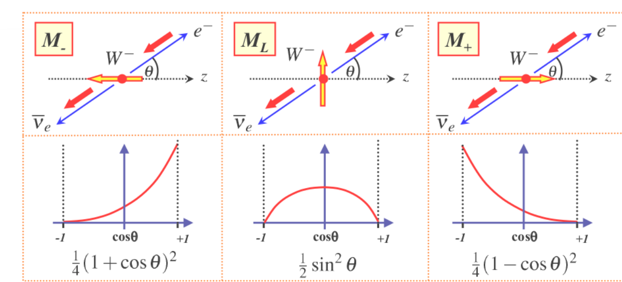
2017+2018

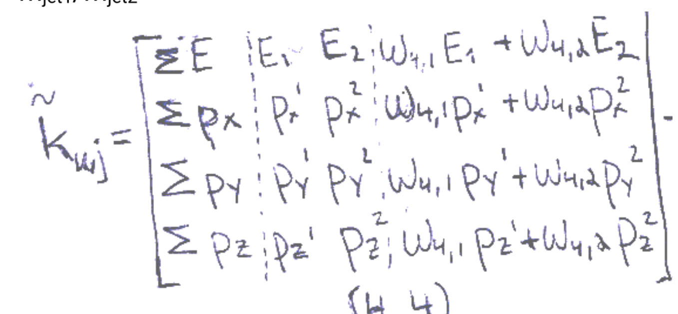




And beyond...







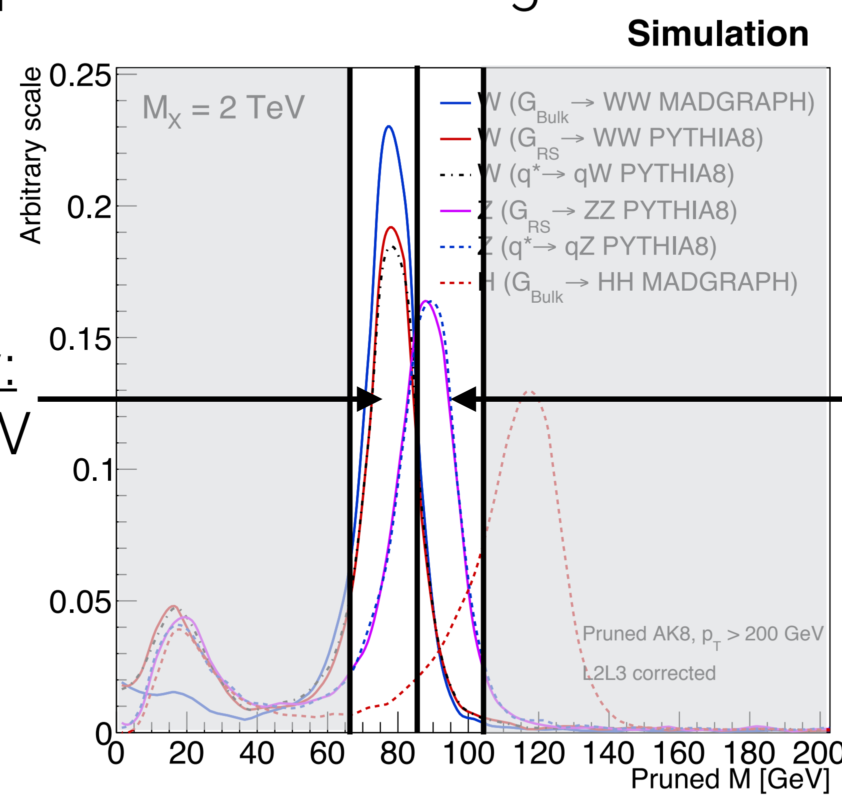
Backup

Enhancing sensitivity

Take signal serious: X→WW, WZ, ZZ?

- split into mass categories

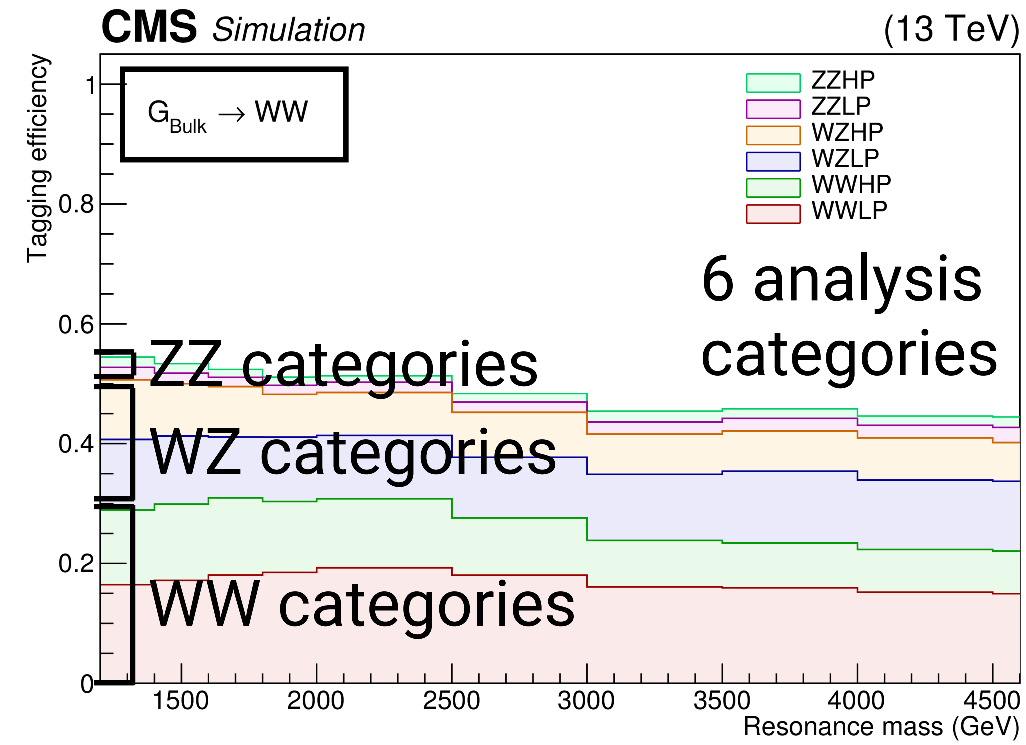
W window:
65 - 85 GeV

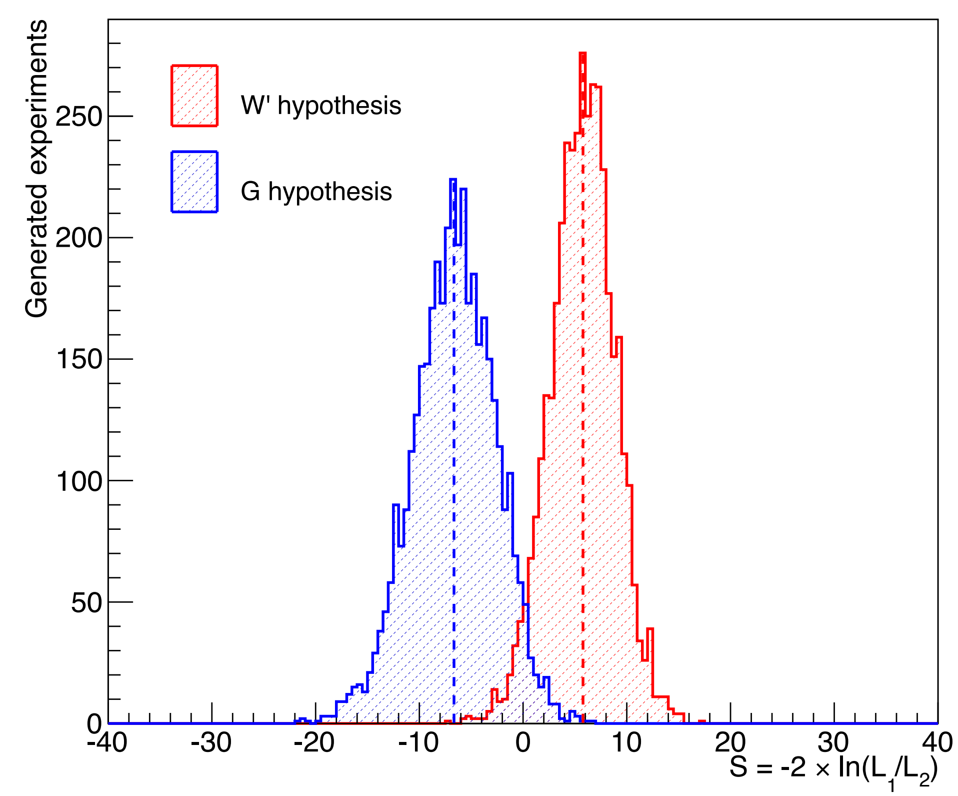


Z window:
85 -105 GeV

3 mass categories, one VV limit
~same or gain in sensitivity

Likelihood from event counting (more events in WZ for W'(WZ) than G_{Bulk}(WW))

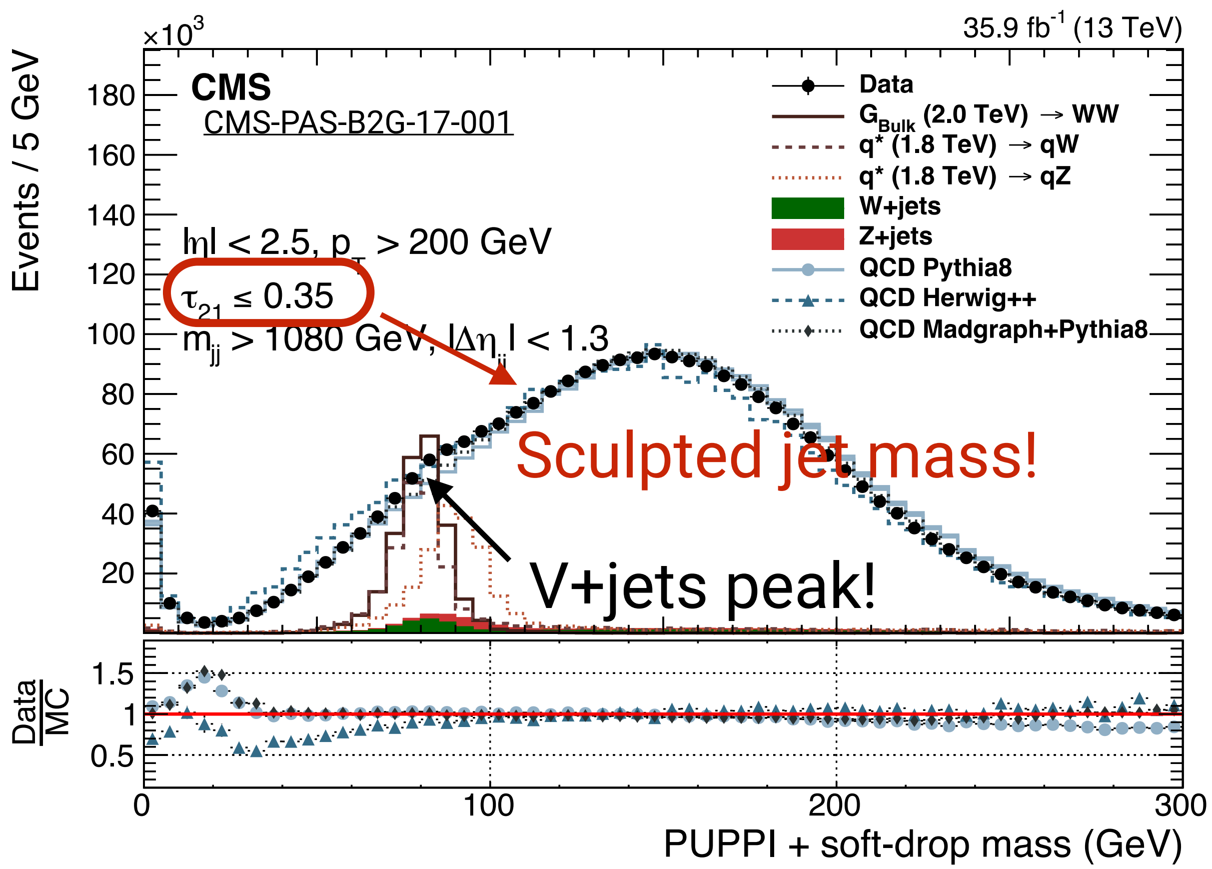


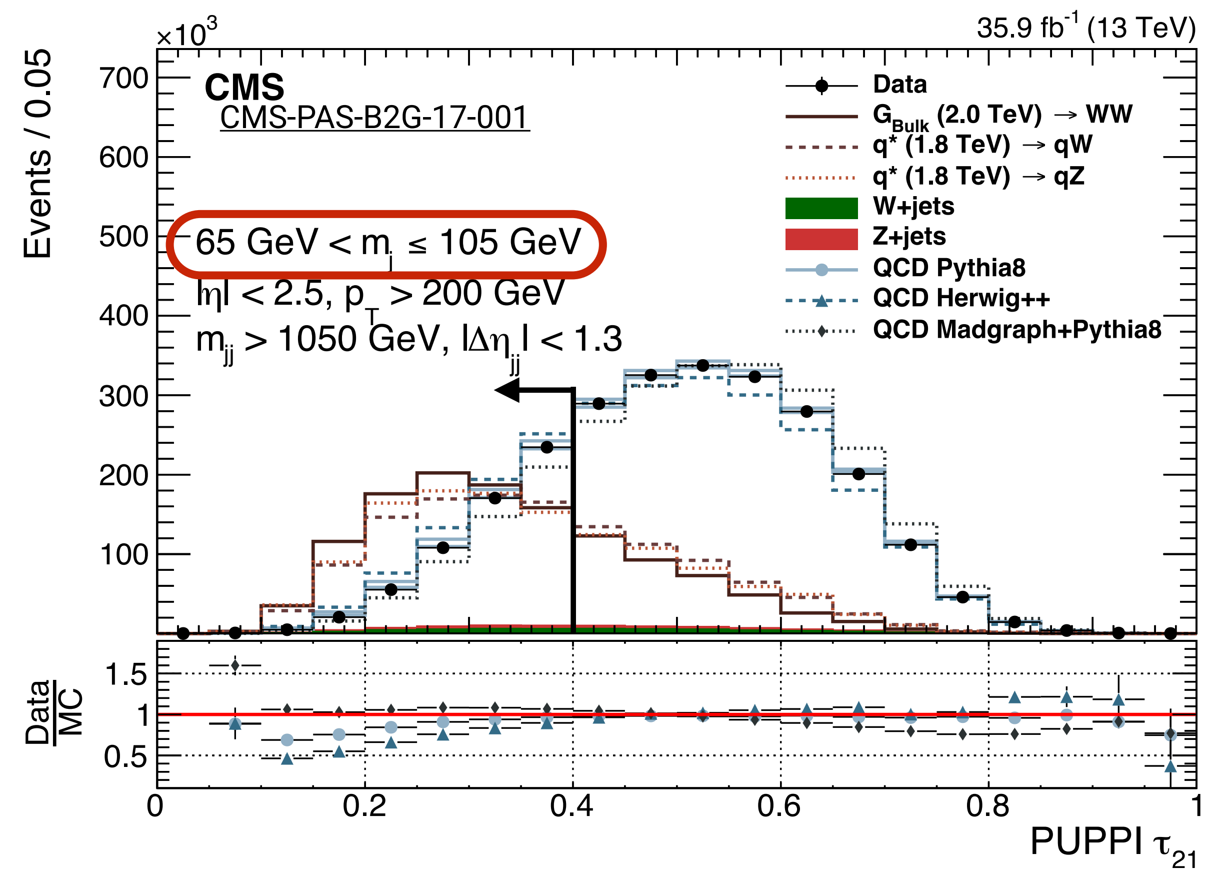


Systematics 2015

Source	Relevant quantity	HP uncertainty (%)	LP uncertainty (%)
Jet energy scale	Resonance shape	2	2
Jet energy resolution	Resonance shape	10	10
Jet energy and m_{jet} scale	Signal yield	0.1–4	
Jet energy and m_{jet} resolution	Signal yield	0.1–1.4	
Pileup	Signal yield	2	
Integrated luminosity	Signal yield	2	
PDFs (W')	Signal yield	4–19	
PDFs (Z')	Signal yield	4–13	
PDFs ($\mathrm{G}_{\mathrm{bulk}}$)	Signal yield	9–77	
Scales (W')	Signal yield	1–14	
Scales (Z')	Signal yield	1–13	
Scales ($\mathrm{G}_{\mathrm{bulk}}$)	Signal yield	8–22	
Jet energy and m_{jet} scale	Migration	1–50	
V tagging τ_{21}	Migration	14	21
V tagging p_{T}-dependence	Migration	7–14	5–11

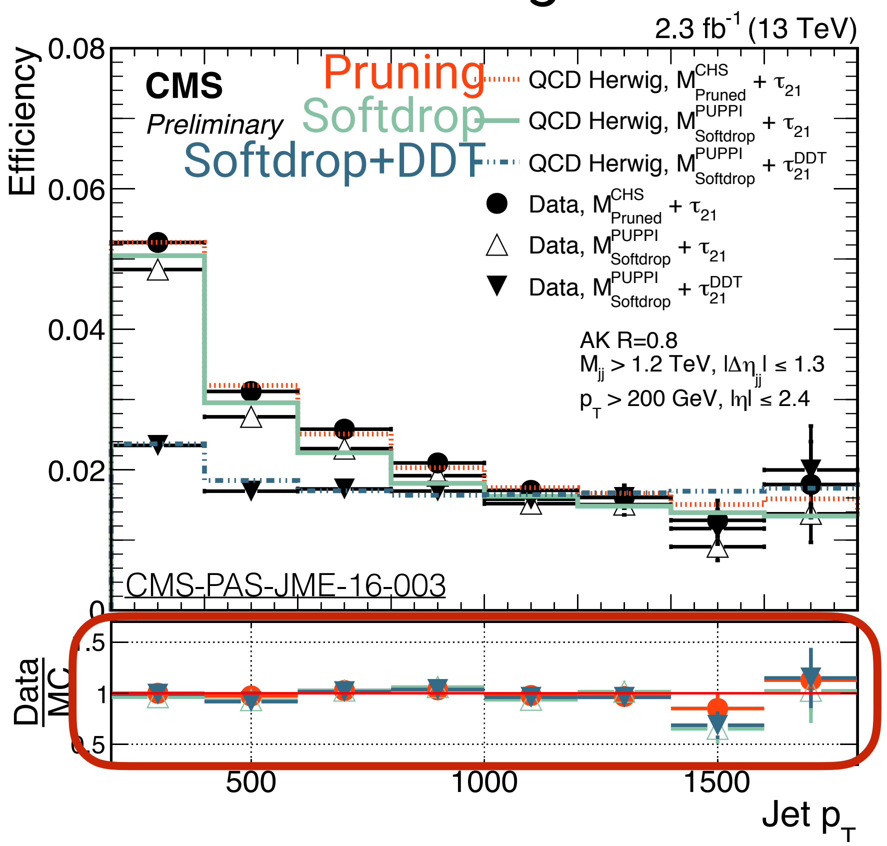
Developing a new V-tagger: Performance in data



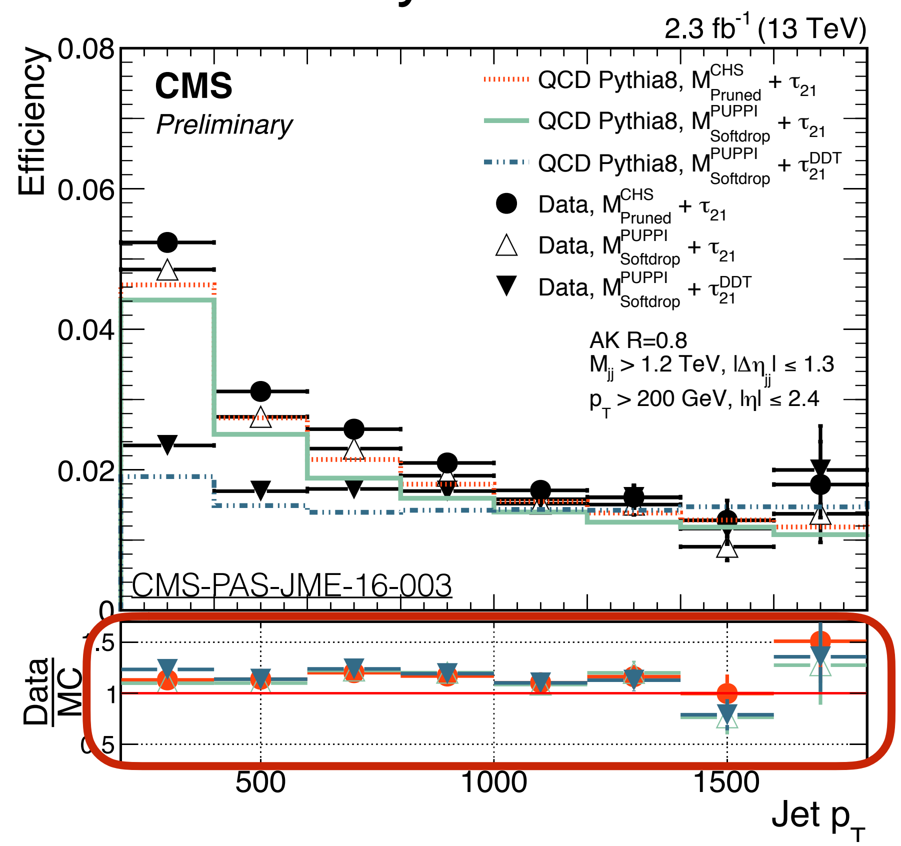


Developing a new V-tagger: Performance in data

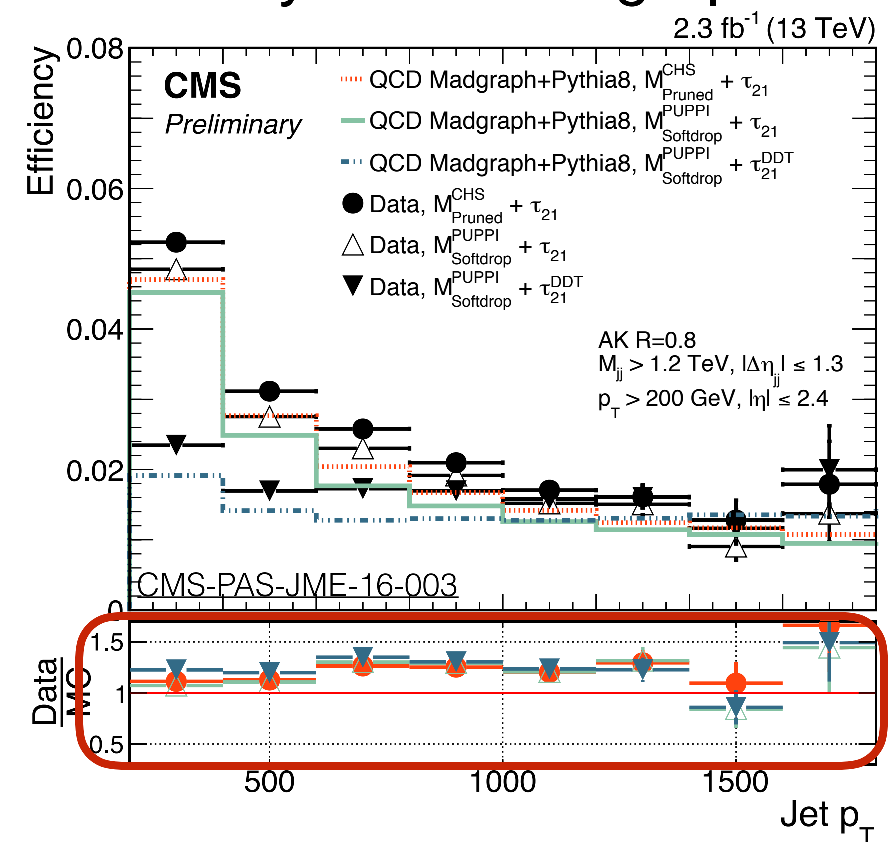
Herwig++



Pythia8

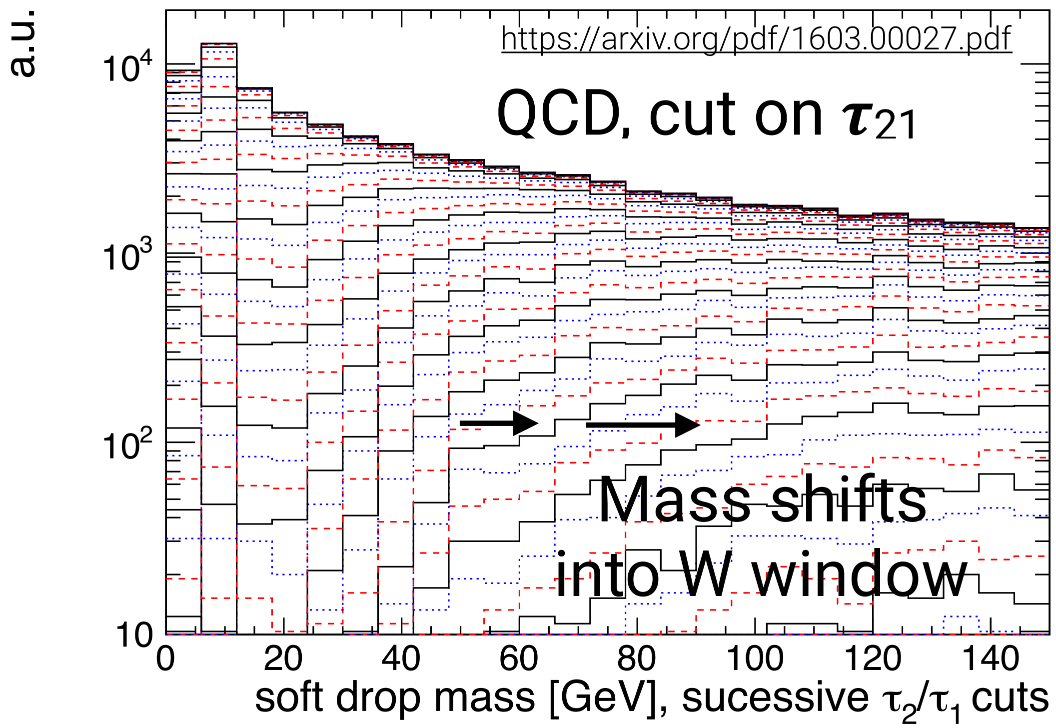


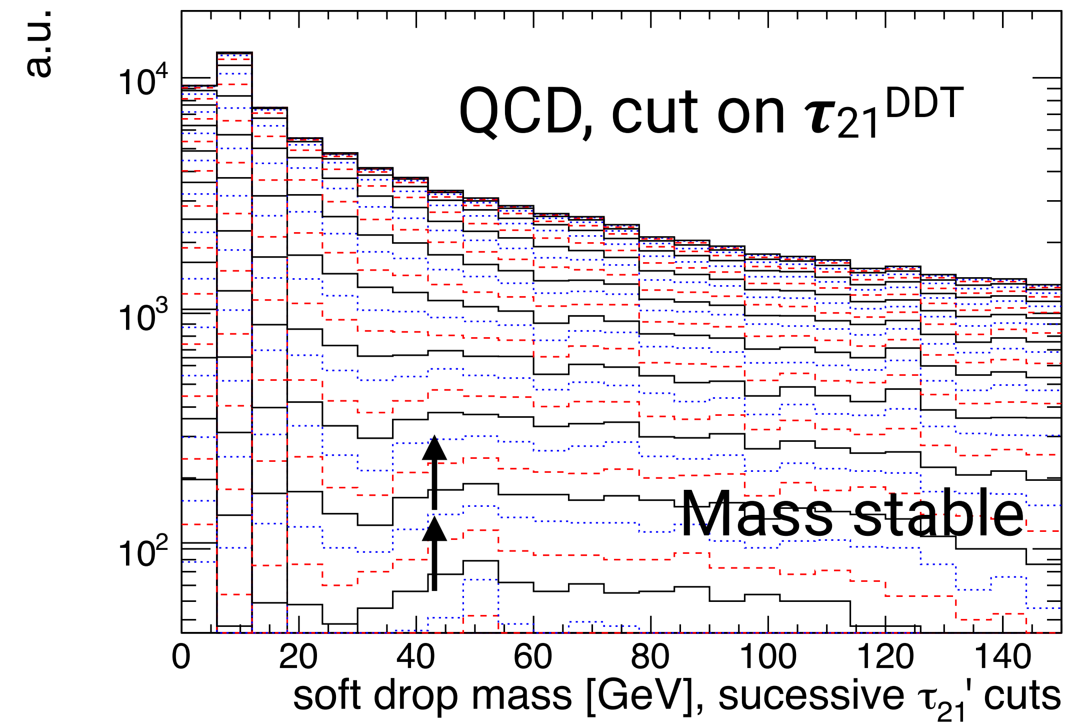
Pythia8+Madgraph

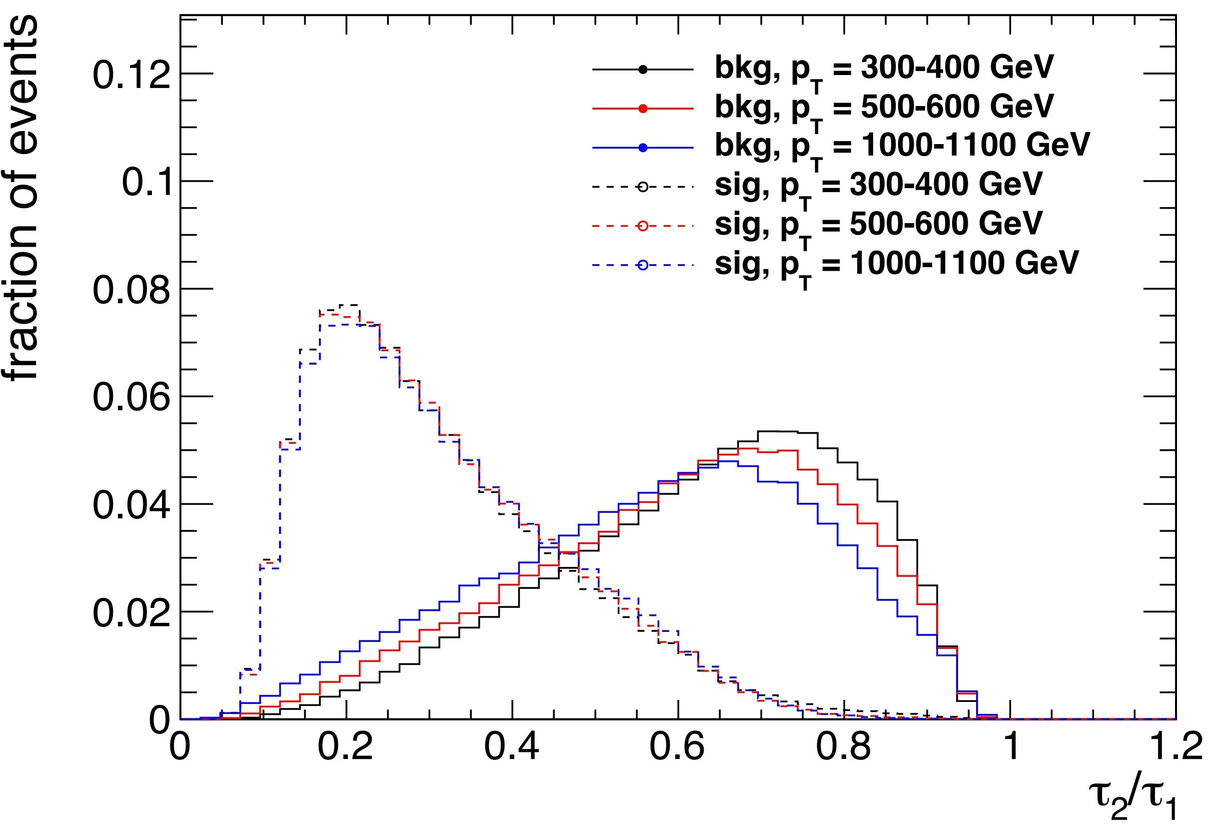


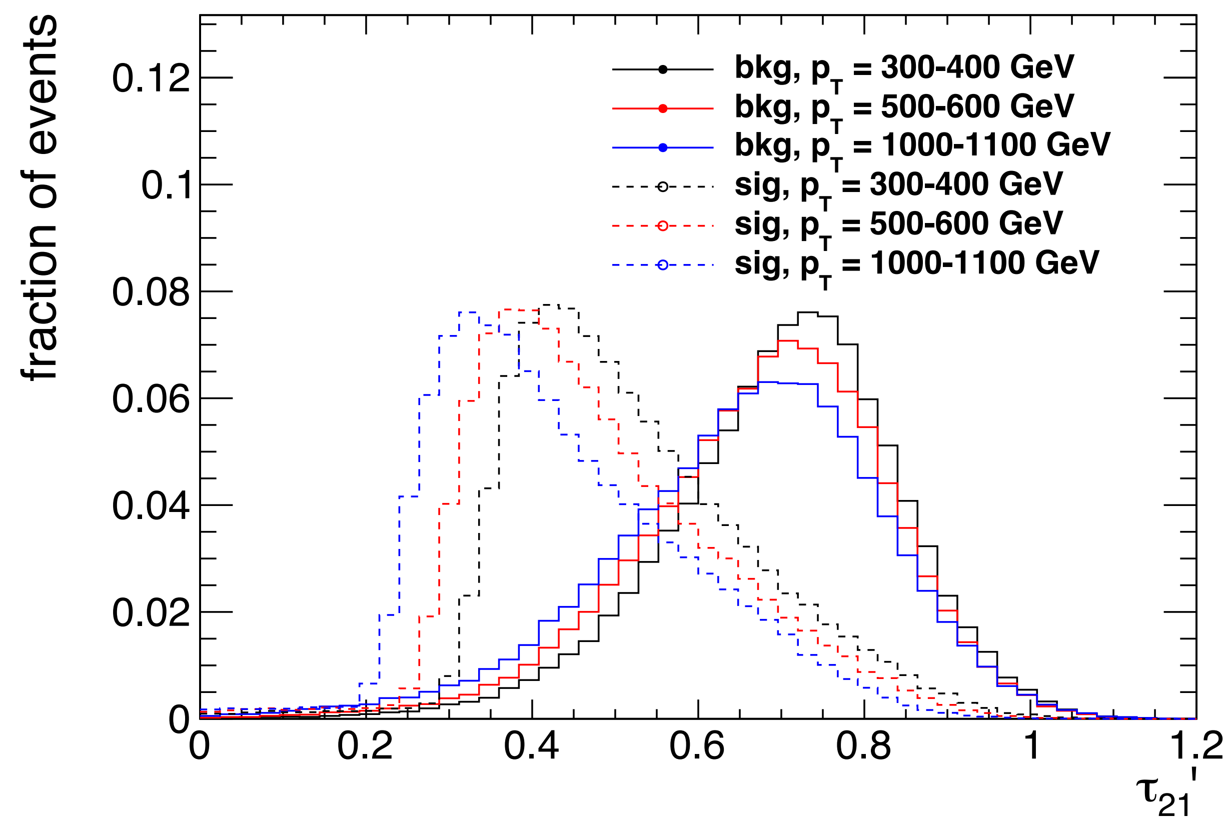
- Substructure variables strong dependence on shower generator
  - different description of gluon radiation
- Best description with Herwig++ , p_T dependence well described by all generators
- Need to ensure we know real signal eff. in data from region well described by MC!

Designing Decorrelated Taggers (DDT)



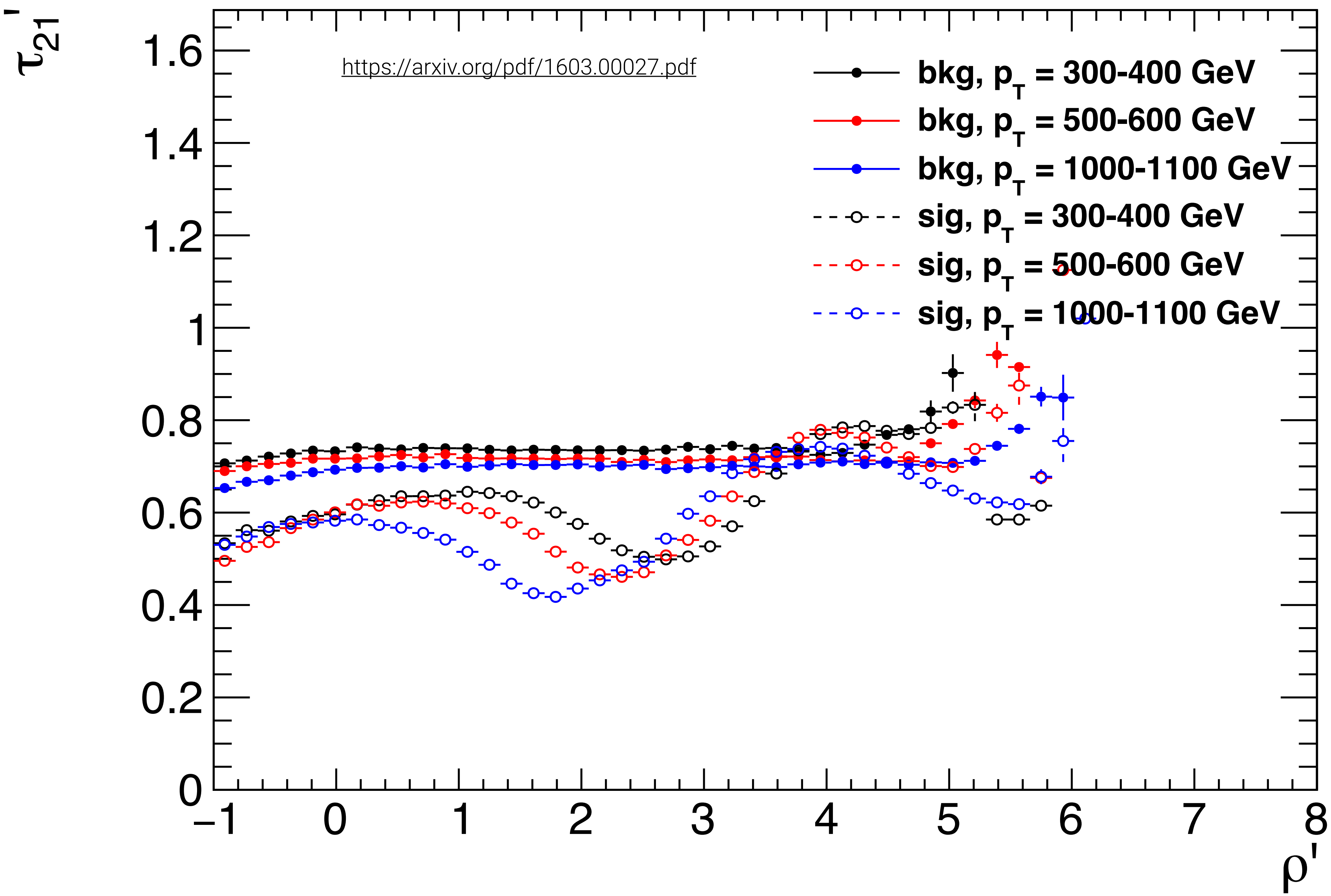






N-subjettiness DDT

$$\tau_{21}^{DDT} = \tau_{21} - 0.63 \times log(\frac{m^2}{p_T \cdot 1\,\mathrm{GeV}})$$



W-tag SF method

Scale factors for W tagging scale factors (SF) are measured for five different working points 3 for Puppi+Soft Drop algorithm (τ_{21}≳0.35; τ_{21}≳0.4; τ_{21}≳0.55) and 2 for the CHS+Pruning algorithm (τ_{21}≳0.45; τ_{21}≳0.6).
To extract the Scale Factors for the Scale (μ in the following), Resolution (σ in the following) and the τ_{21} Efficiency (ε in the following) a two step fit to a pure tt sample is used:

1. fit to W-enriched category (High Purity: τ_{21} <X) to extract μ and σ of the distribution
2. simultaneous fit to both High Purity and Low Purity (τ_{21} >X) categories, using the information for μ and σ from step 1 and extracting ε

The functions used to describe the two categories as a function of the ak08 groomed mass are:

<u>High Purity</u>

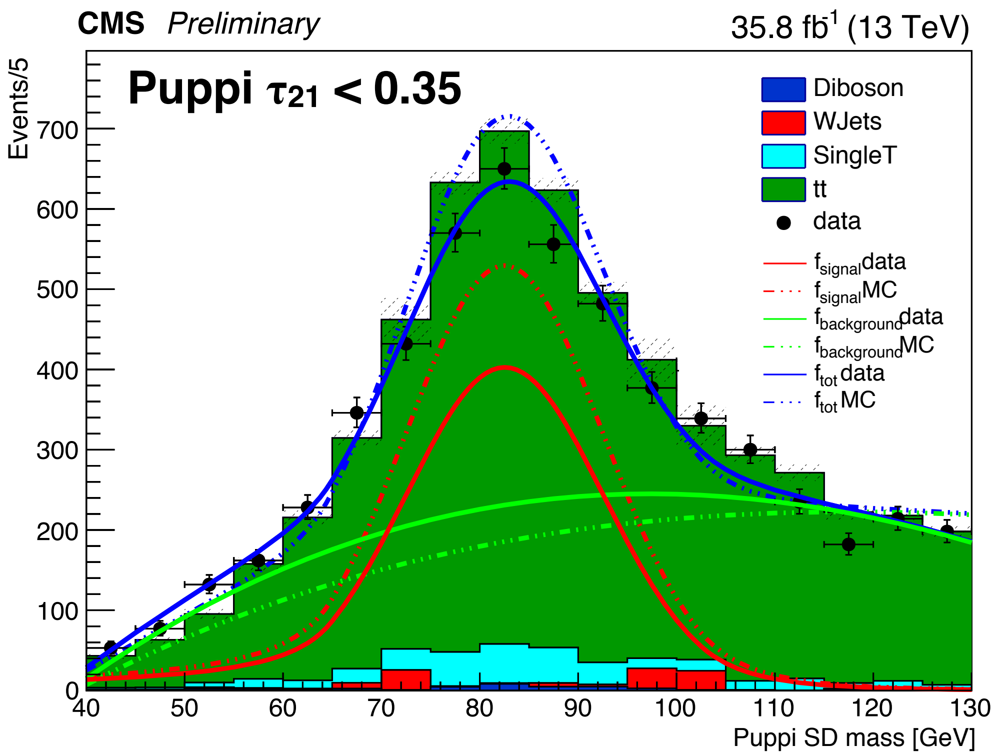
$$L = \prod_{i}^{N\,pass} \left[N_W \cdot \varepsilon_{HP} \cdot f_{passed}(m_j) + N_2 \cdot f_{comb}(m_j)\right]$$

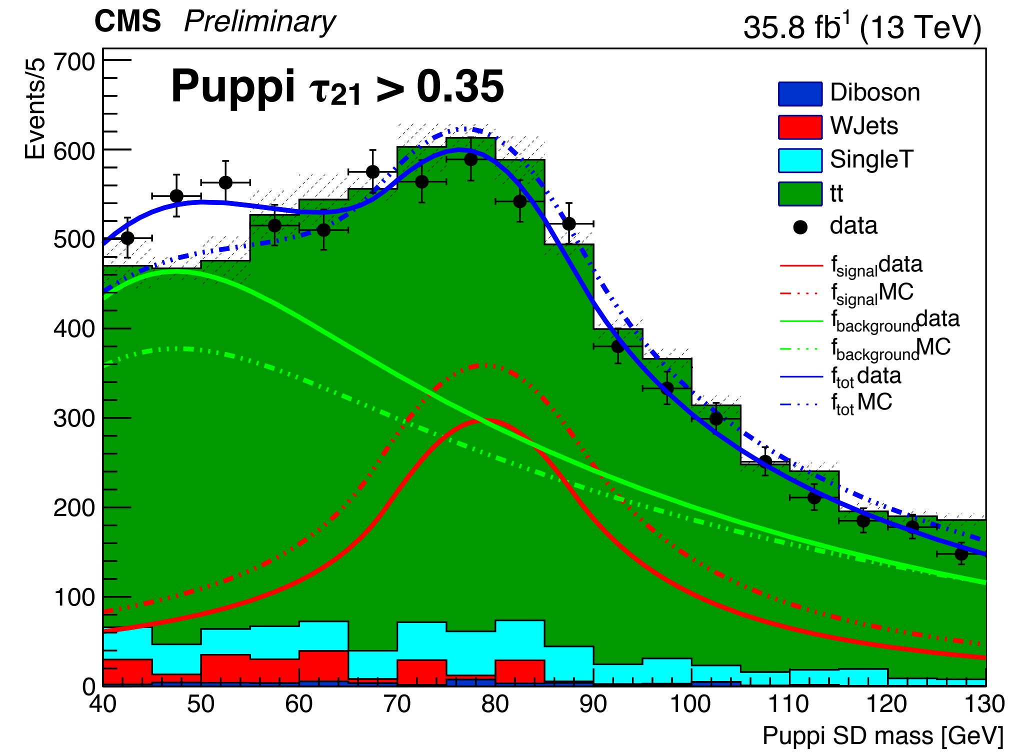
<u>Low Purity</u>

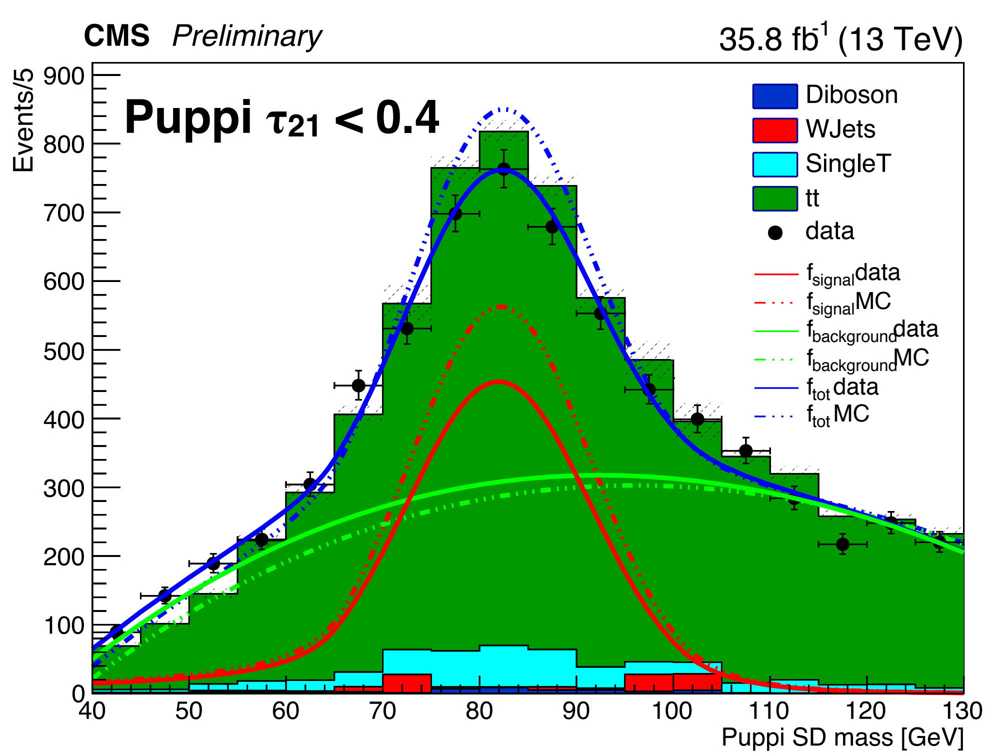
$$L = \prod_{i}^{N\,fail} \left[N_W \cdot (1 - \varepsilon_{HP}) \cdot f_{fail}(m_j) + N_3 \cdot f'_{comb}(m_j)\right]$$

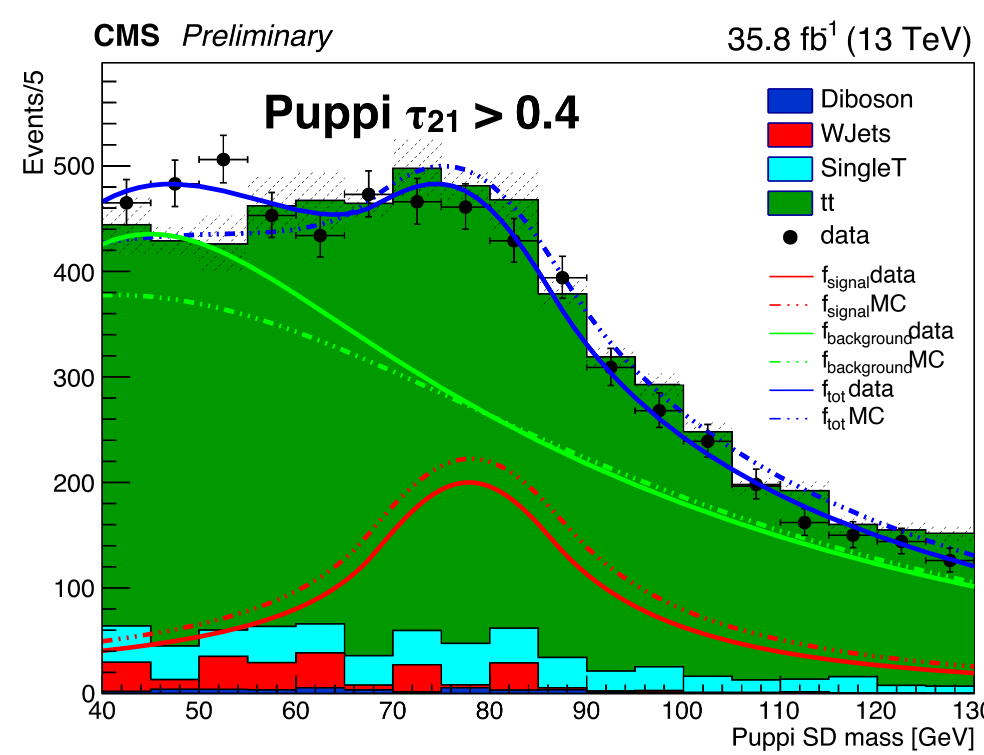
Where f_{passed} and f_{fail} (red curves in the following plots) describe the W peak in the two categories, while f_{comb} (green curves in the following plots) describe the combinatorial (e.g. from events with a b-jet merged in the AK08 jet) and general background of the tt events.
The events used for the SF extraction require in the final state 1 ak08 (p_t>200) + 1 b-tagged ak04 + 1 lepton + 1 semileptonic W (p_t>200 GeV) following the selection reported in JME-16-003.

W-tag SF - Softdrop

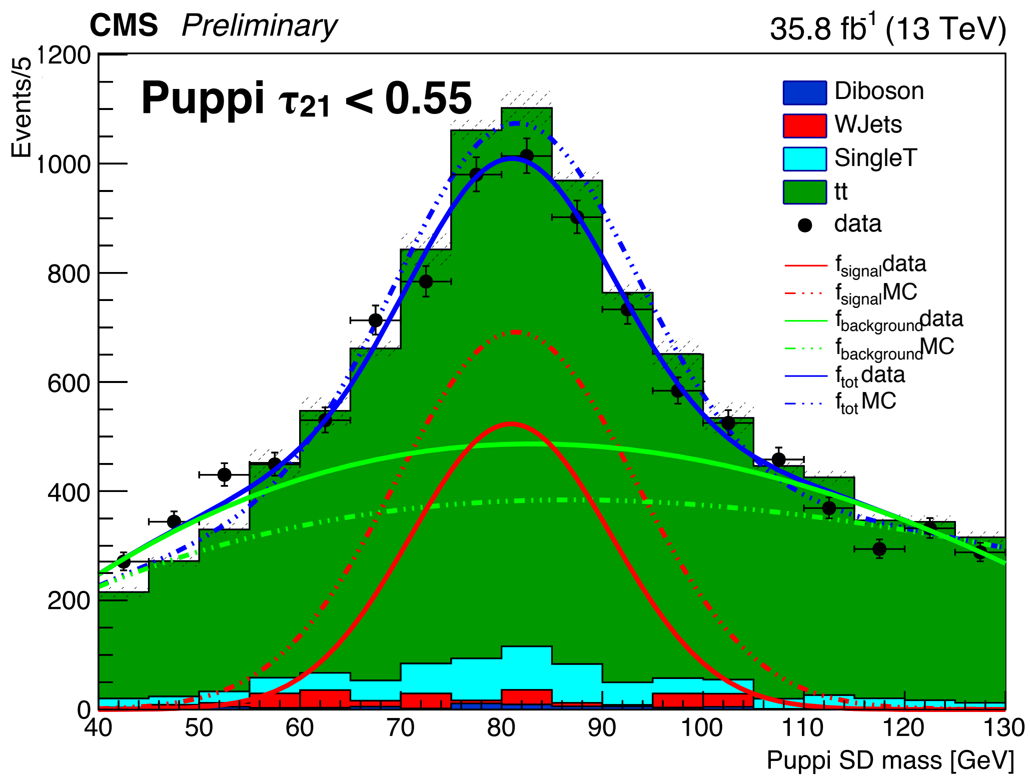


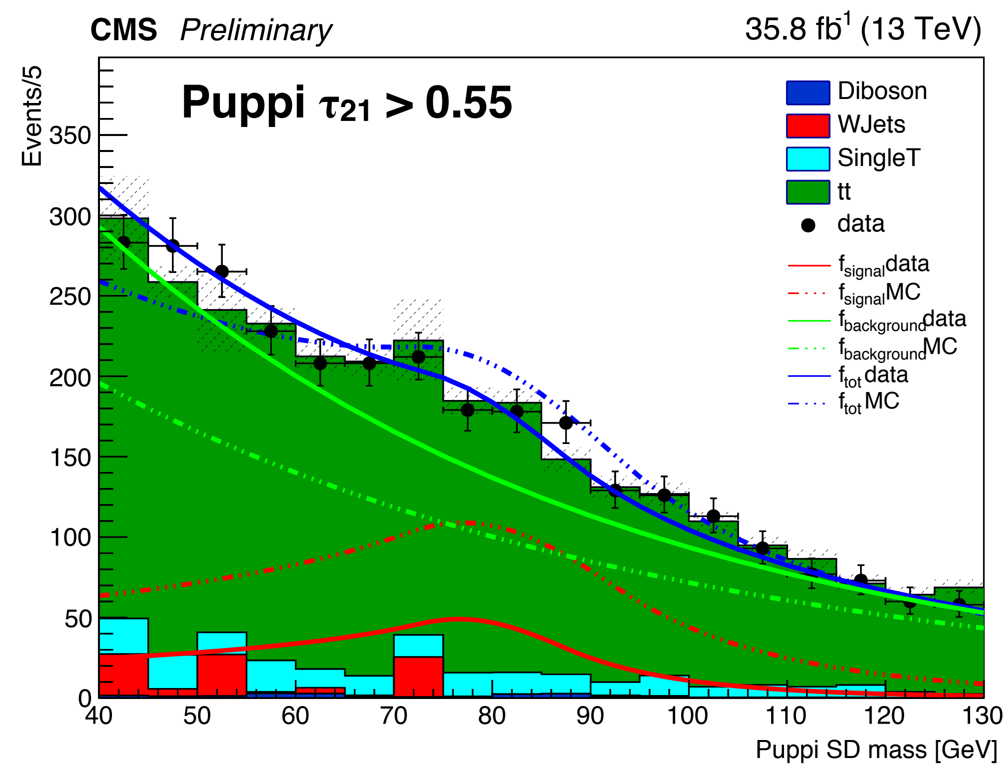






W-tag SF summary

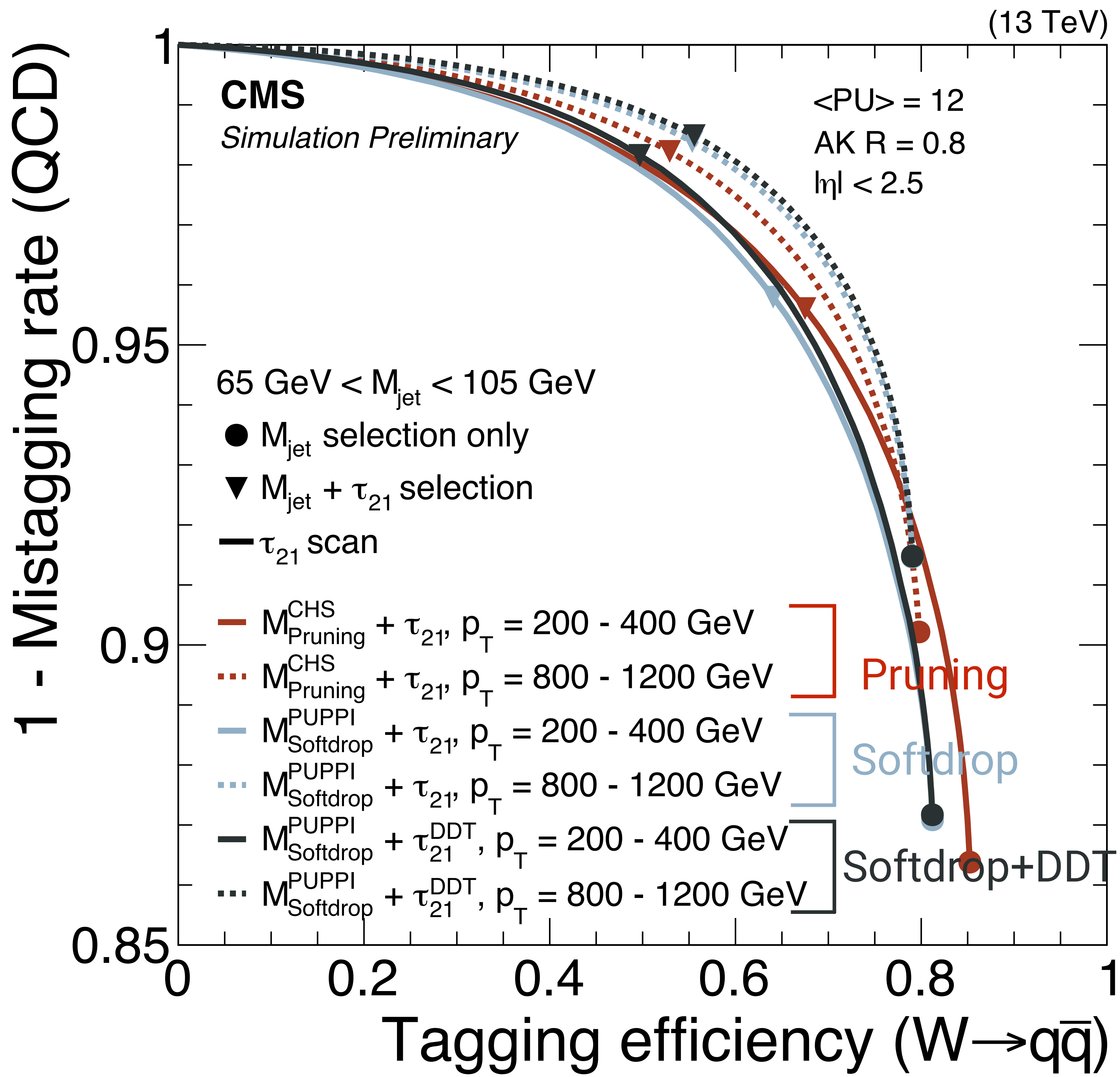




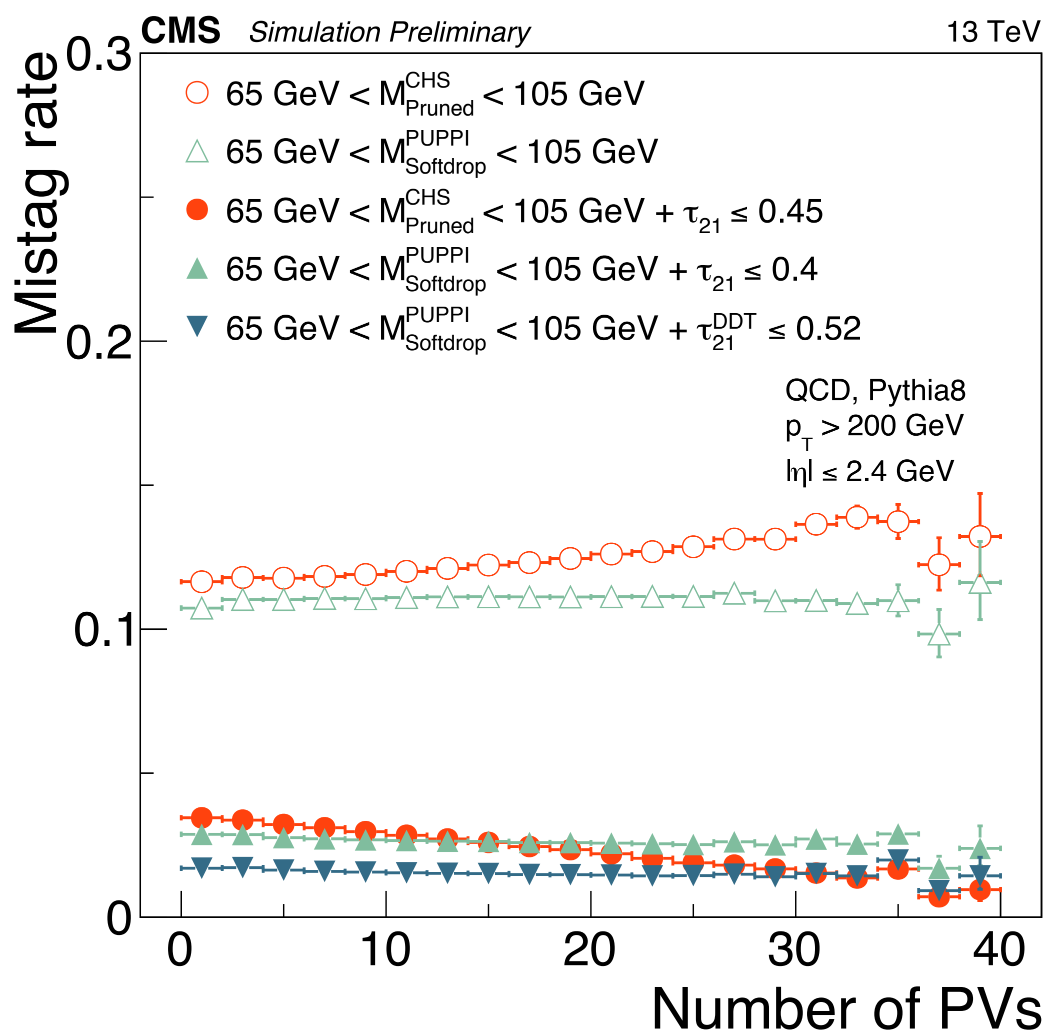
Variable	Scale Factor	Variable	Scale Factor
	τ_{21} WP $= 0.45$		Puppi τ_{21} WP $= 0.35$
ε	1.00 ± 0.06 (stat) $\pm$ 0.06 (syst) $\pm$ 0.03 (syst)	ε	0.99 ± 0.05 (stat) $\pm$ 0.03 (syst) $\pm$ 0.04 (syst)
μ	1.007 ± 0.009 (stat) $\pm$ 0.005 (syst) $\pm$ 0.002 (syst)	μ	0.999 ± 0.009 (stat) $\pm$ 0.03 (syst) $\pm$ 0.02 (syst)
σ	1.15 ± 0.04 (stat) $\pm$ 0.03 (syst) $\pm$ 0.02 (syst)	σ	1.06 ± 0.03 (stat) $\pm$ 0.09 (syst) $\pm$ 0.08 (syst)
	τ_{21} WP $= 0.6$		Puppi τ_{21} WP $= 0.4$
ε	1.08 ± 0.06 (stat) $\pm$ 0.05 (syst) $\pm$ 0.05 (syst)	ε	1.01 ± 0.06 (stat) $\pm$ 0.02 (syst) $\pm$ 0.04 (syst)
μ	1.005 ± 0.005 (stat) $\pm$ 0.006 (syst) $\pm$ 0.005 (syst)	μ	0.998 ± 0.007 (stat) $\pm$ 0.006 (syst) $\pm$ 0.001 (syst)
σ	1.12 ± 0.03 (stat) $\pm$ 0.02 (syst) $\pm$ 0.04 (syst)	σ	1.08 ± 0.02 (stat) $\pm$ 0.03 (syst) $\pm$ 0.08 (syst)
			Puppi τ_{21} WP $= 0.55$
		ε	1.04 ± 0.02 (stat) $\pm$ 0.03 (syst) $\pm$ 0.02 (syst)
		μ	0.996 ± 0.004 (stat) $\pm$ 0.009 (syst) $\pm$ 0.002 (syst)
		σ	1.08 ± 0.02 (stat) $\pm$ 0.03 (syst) $\pm$ 0.08 (syst)

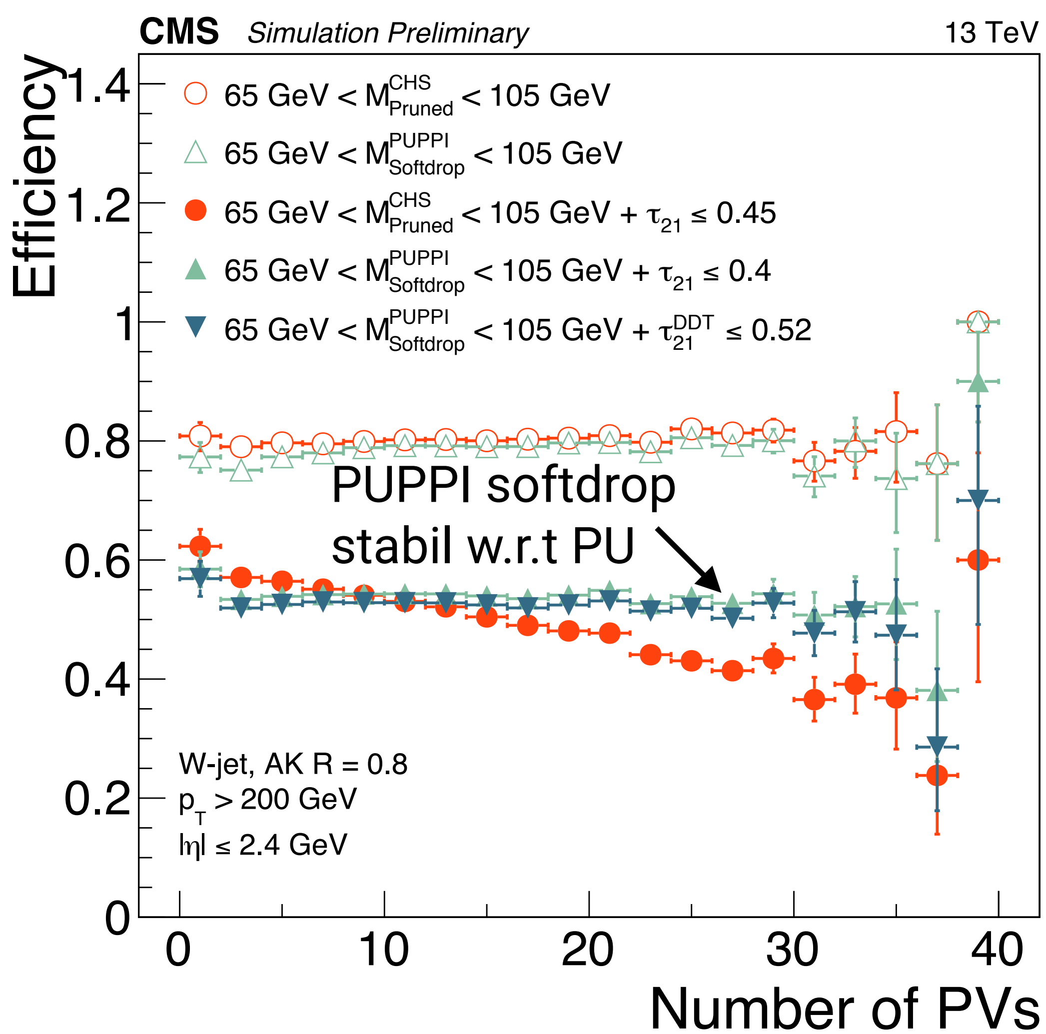
V-tagging performance

- Similar performance for
  - Low-p_T (200-400 GeV): ~65% signal efficiency at ~4% mistag rate
  - High-p_T (800-1200 GeV): 50-55% signal efficiency at 1-2% mistag rate

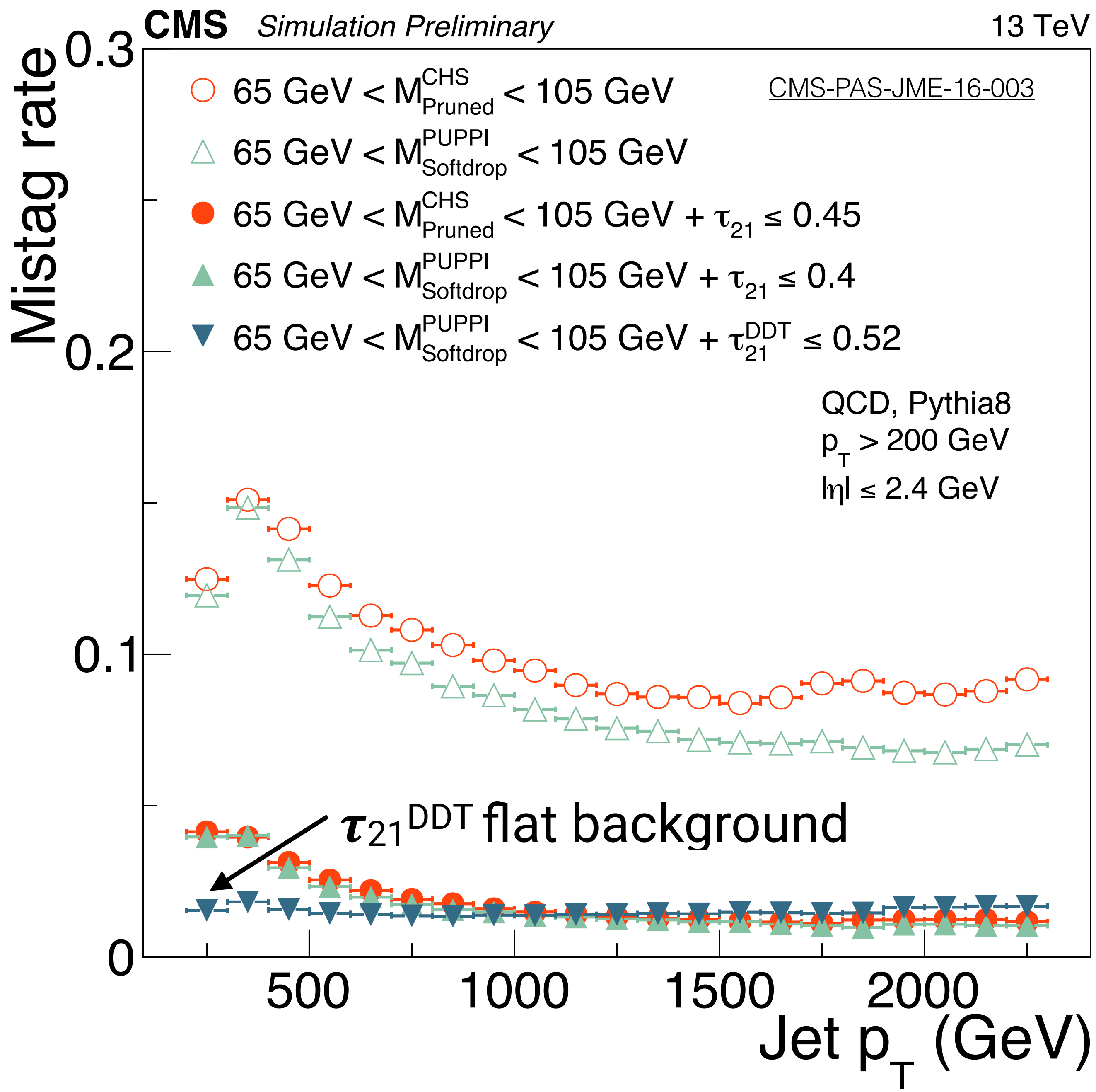


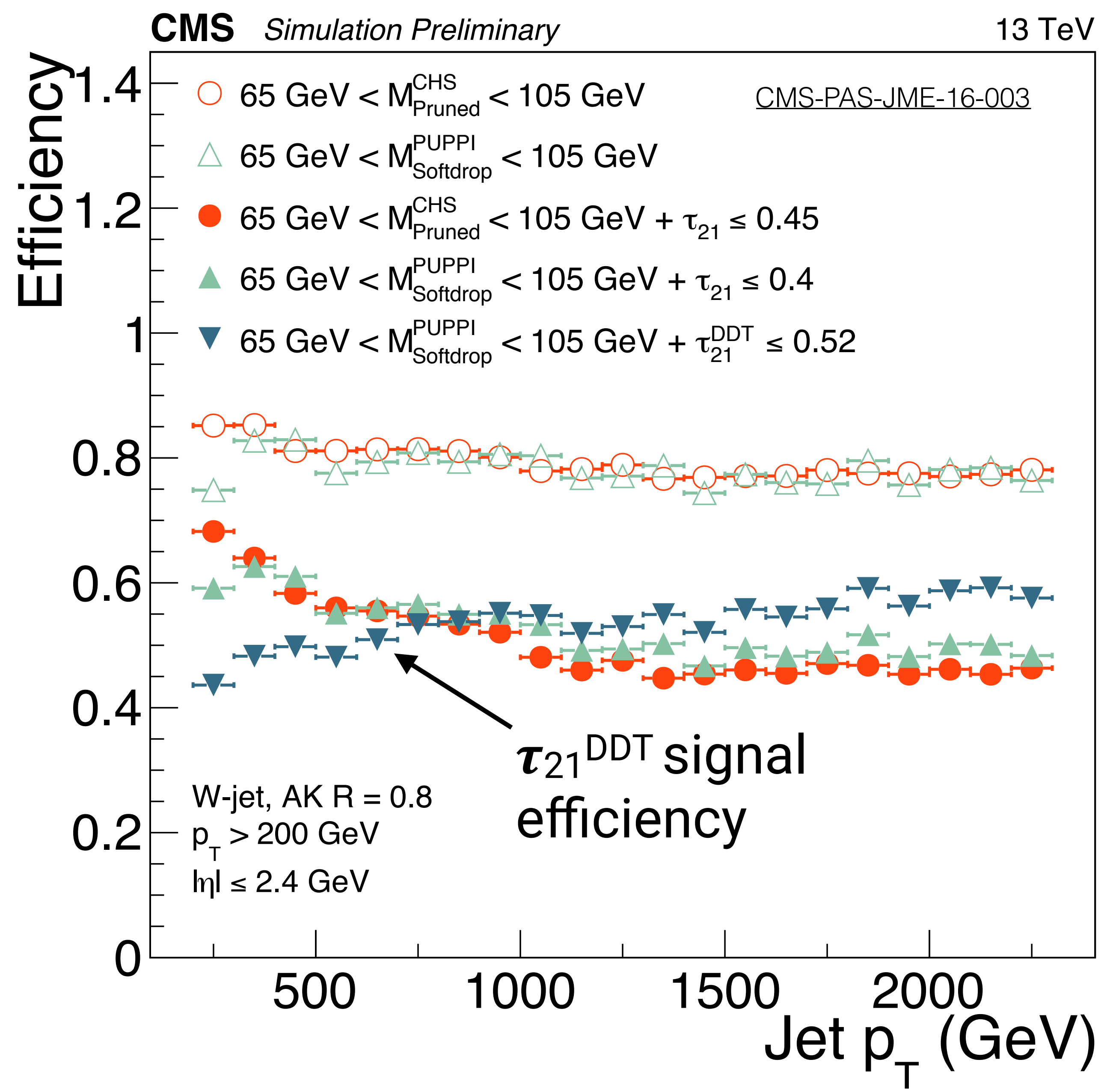
V-tagging performance





V-tagging performance





Mass: Pruning

Removes soft, large angle constituents from the jet

- Recluster jet using Cambridge-Achen algorithm, removing each recombination that has

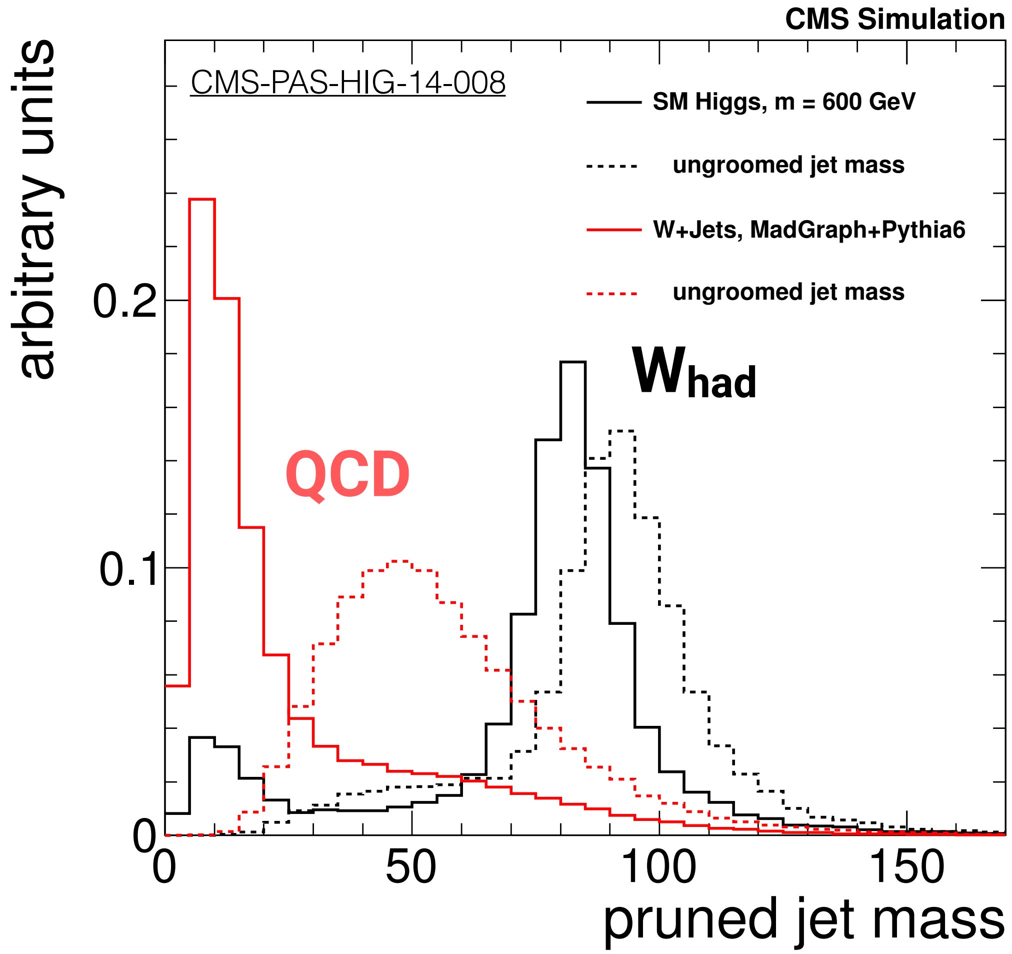
$$\Delta_{ab} > R_{\mathrm{prune}} = R_{\mathrm{fact}} \cdot \frac{2m}{p_T} \qquad \min(p_{\mathrm{Ta}}, p_{\mathrm{Tb}}) < z_{\mathrm{cut}} p_{\mathrm{T,(a+b)}}$$

$$R_{\mathrm{fact}} = 0.5, z_{\mathrm{cut}} = 0.1$$

  Push q/g mass to zero, increase V mass resolution

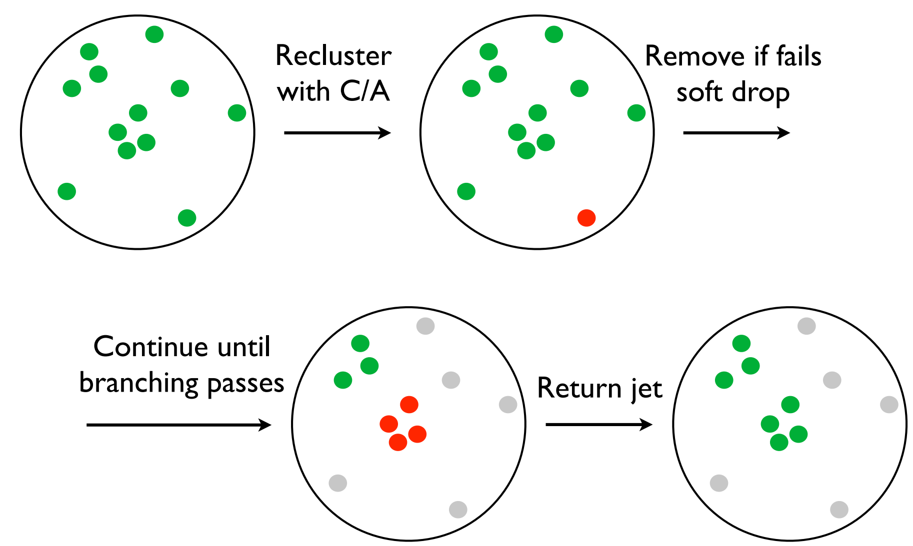
- but, do not fully remove soft emissions and cannot be analytically calculated due to non-global logs (e.g soft emissions entering jet cone from outside)

- want infrared and collinear safe jet observable!



Mass: Softdrop

Recluster jet with C-A algorithm. Then decluster and check if subjets pass

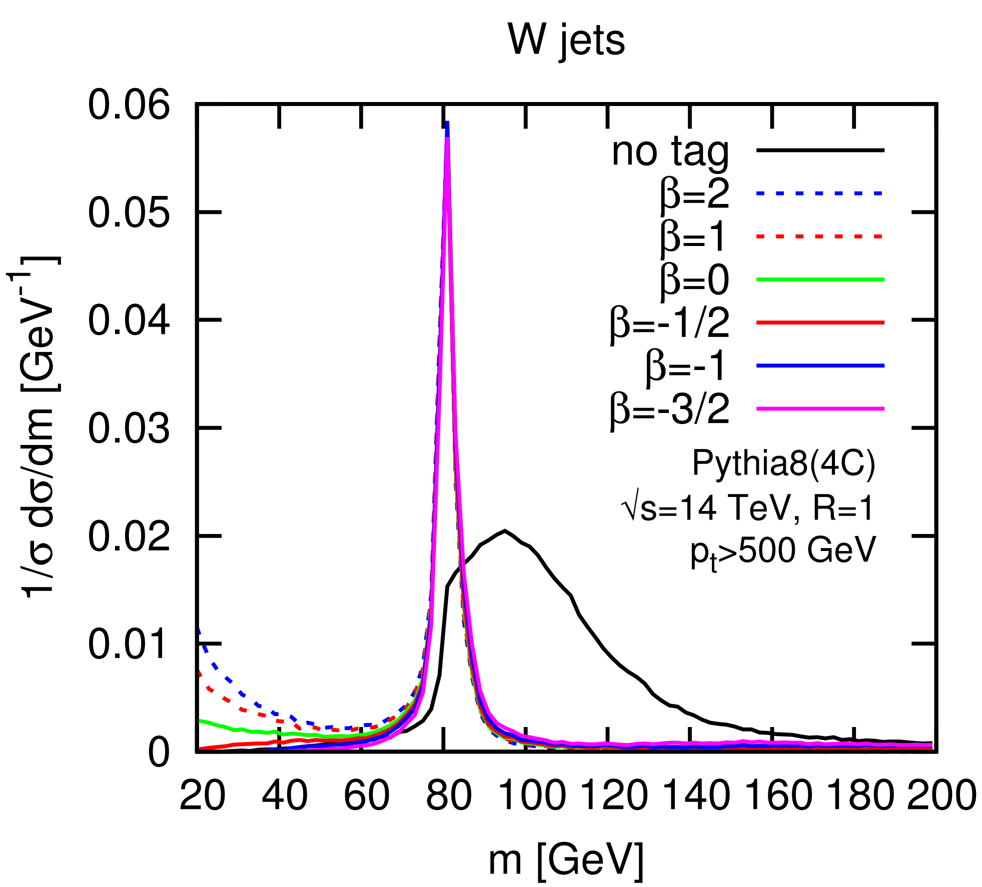


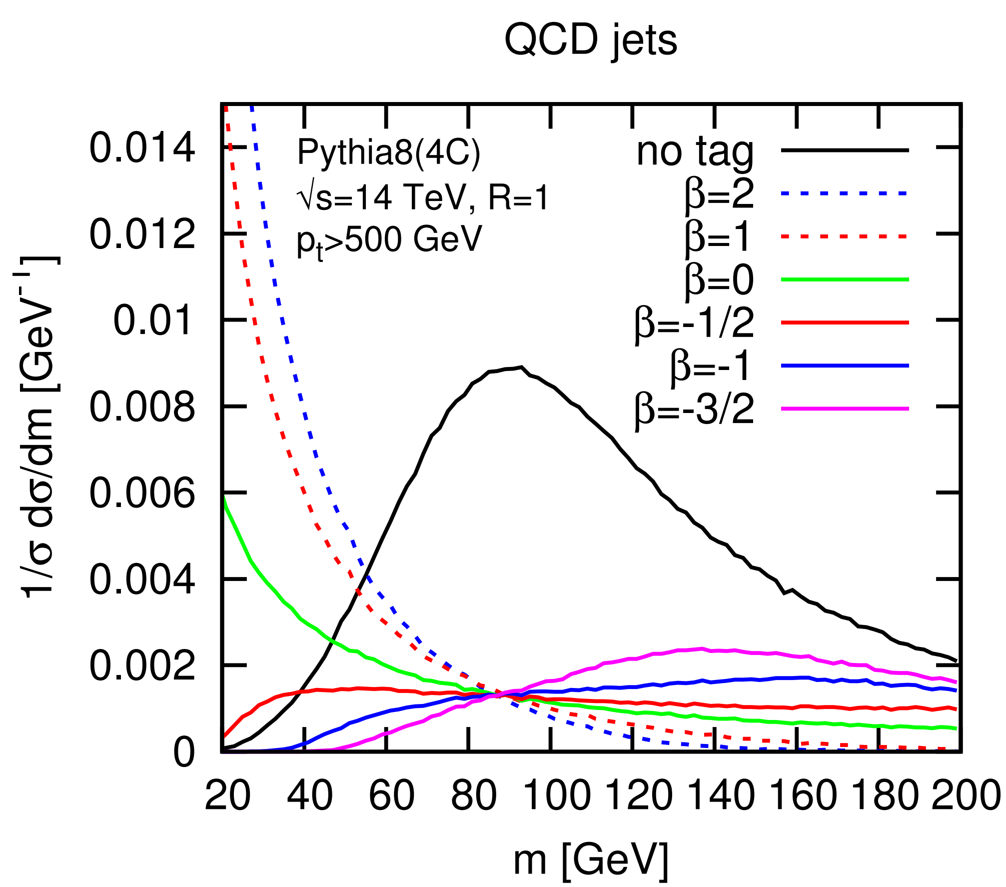
$$\frac{\min(p_{T1}, p_{T2})}{p_{T1}+p_{T2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

Soft threshold

Angular exponent

- in CMS β=0, z_{cut} = 0.1 (modified Mass Drop)





Tuned parameters: z_{cut} and β

β = ∞
no grooming

β > 0
soft, wide angle removed
some soft-collinear removed

β = 0
all soft emissions removed
modified Mass Drop limit

CMS default

β < 0
all soft and collinear emissions removed

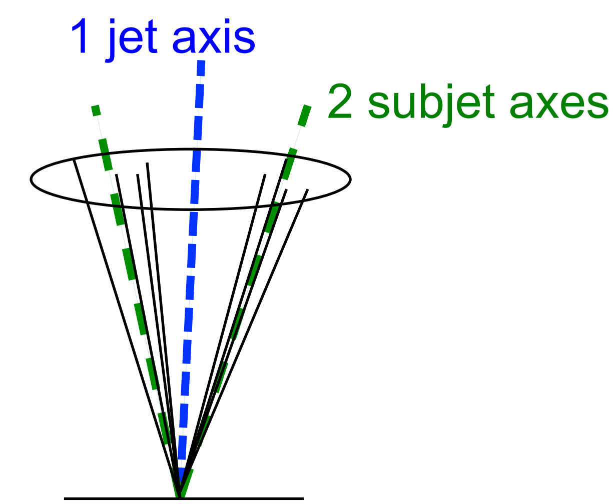
Substructure: N-subjettiness

p_T-weighted sum over all constituents of the distance w.r.t the closest of N axes in a jet

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\left((\Delta R_{1,k}), (\Delta R_{2,k}) \ldots (\Delta R_{N,k})\right)$$

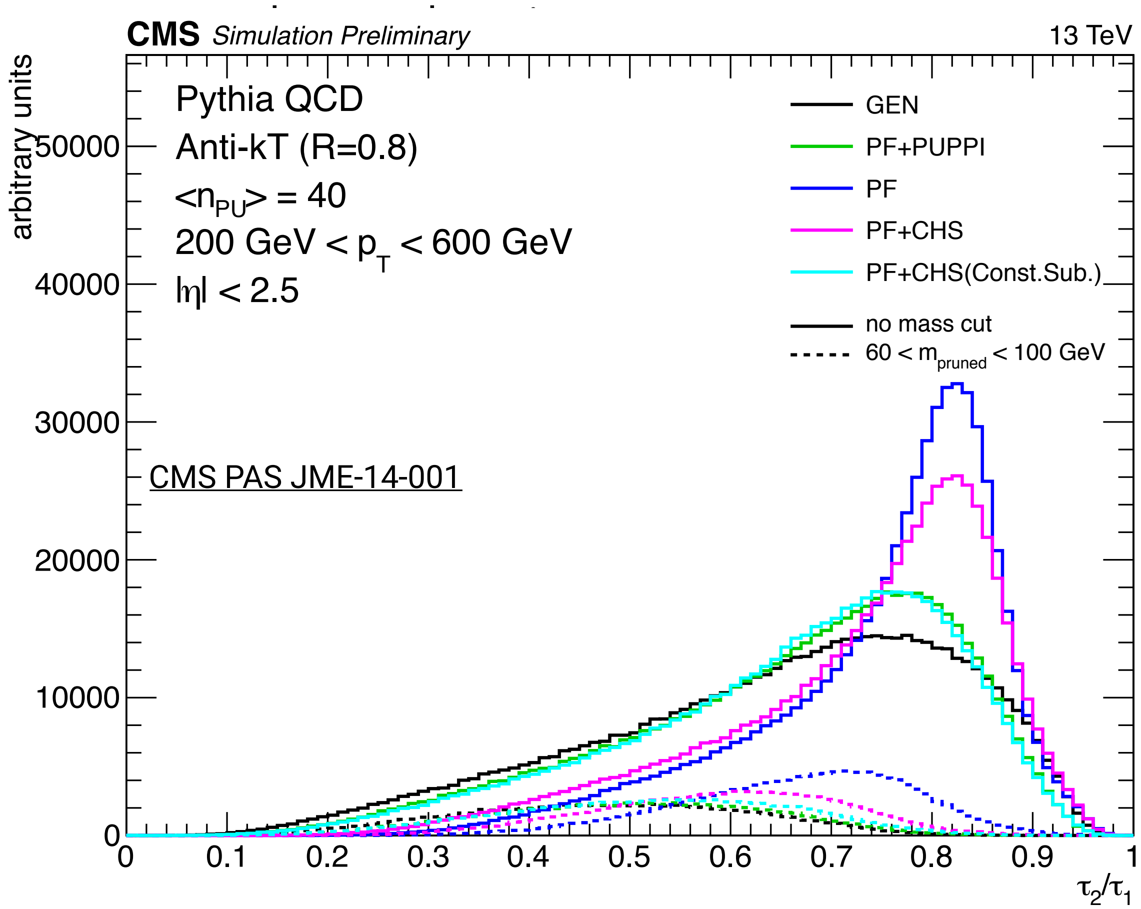
Distance between momentum of constituent k w.r.t momentum of rest-frame subjet N

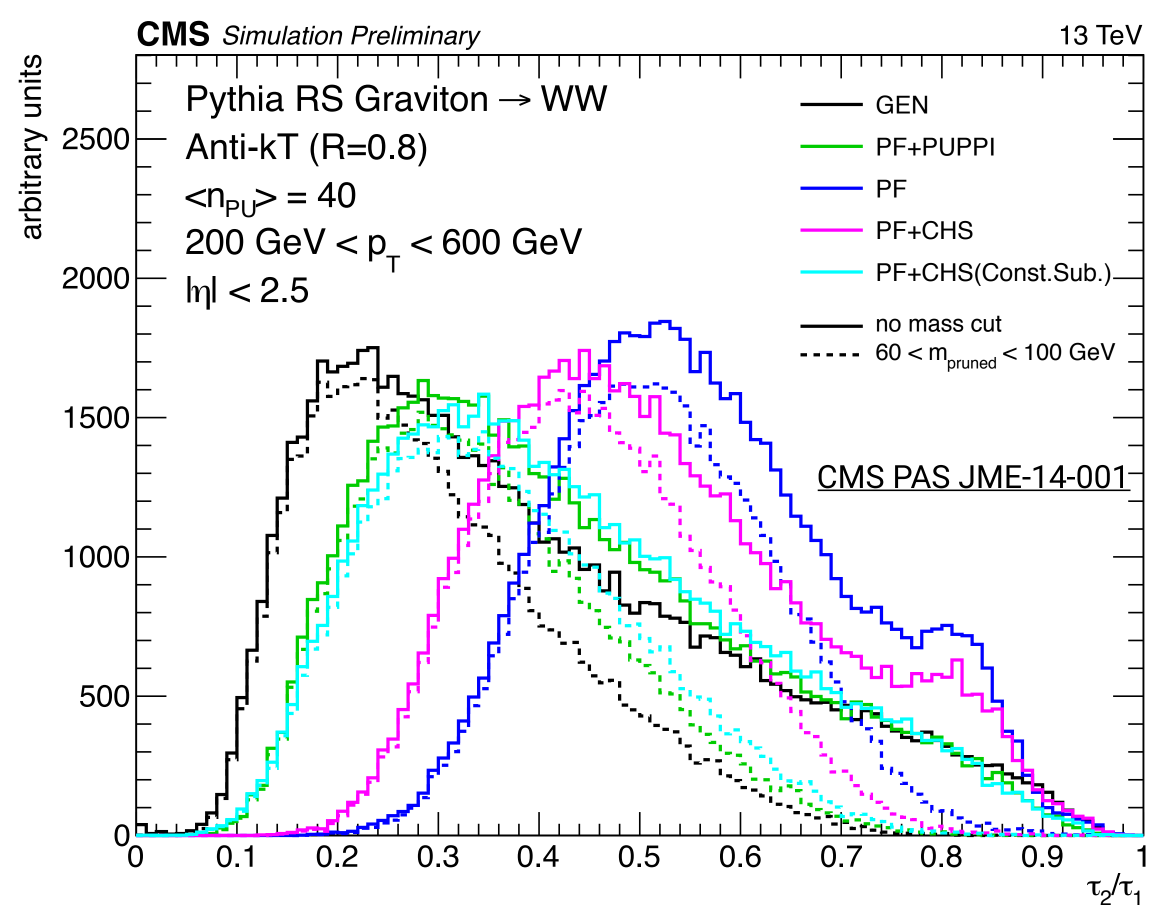
Each constituent assigned to nearest subjet!



- axis obtained by undoing last (N-1) steps of clustering algorithm

- small τ_N indicates compatibility with N





Energy correlation functions (EFCs)

Signal jets satisfy the inequality ${}_2e_3 \ll (e_2)^2$, explaining the definition of the N2 observable

$$N_2^\beta = \frac{{}_2e_3^\beta}{({}_1e_2^\beta)^2}$$

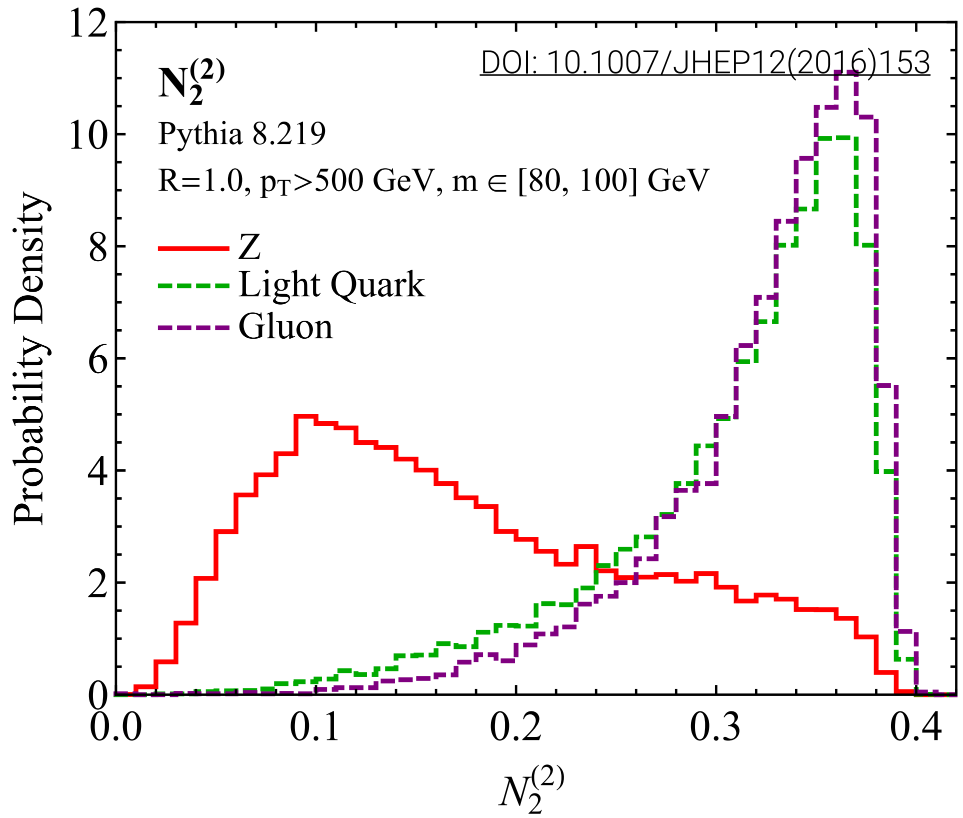
\# particles

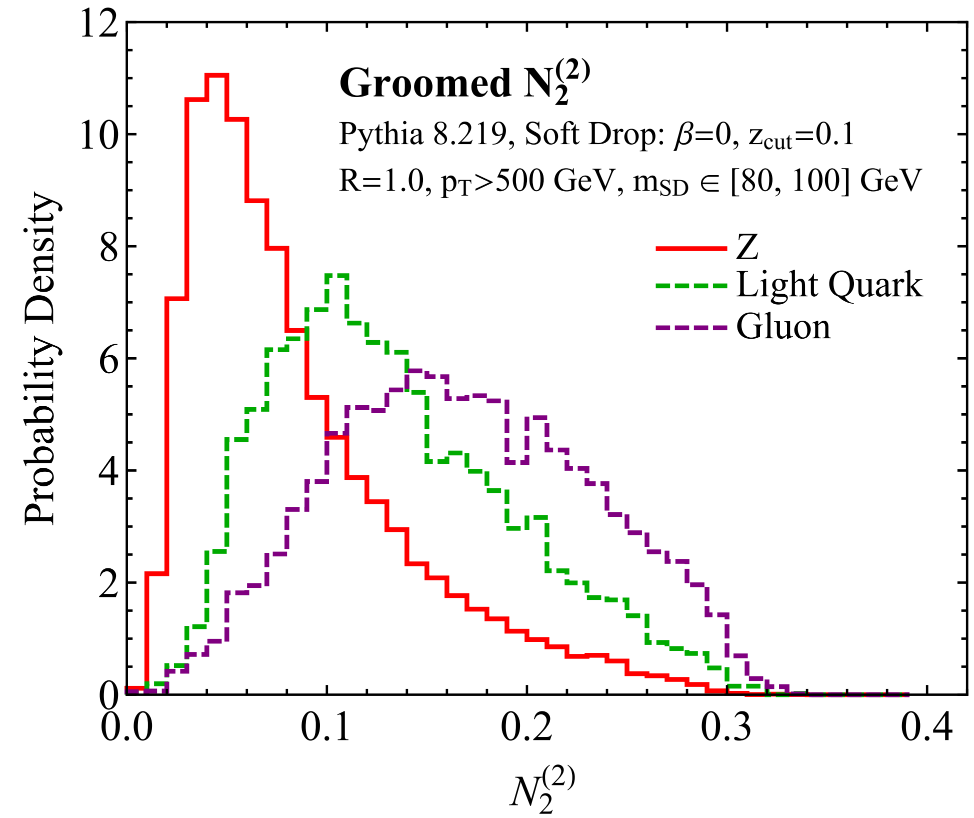
\# angles

Less discriminating power after grooming applied

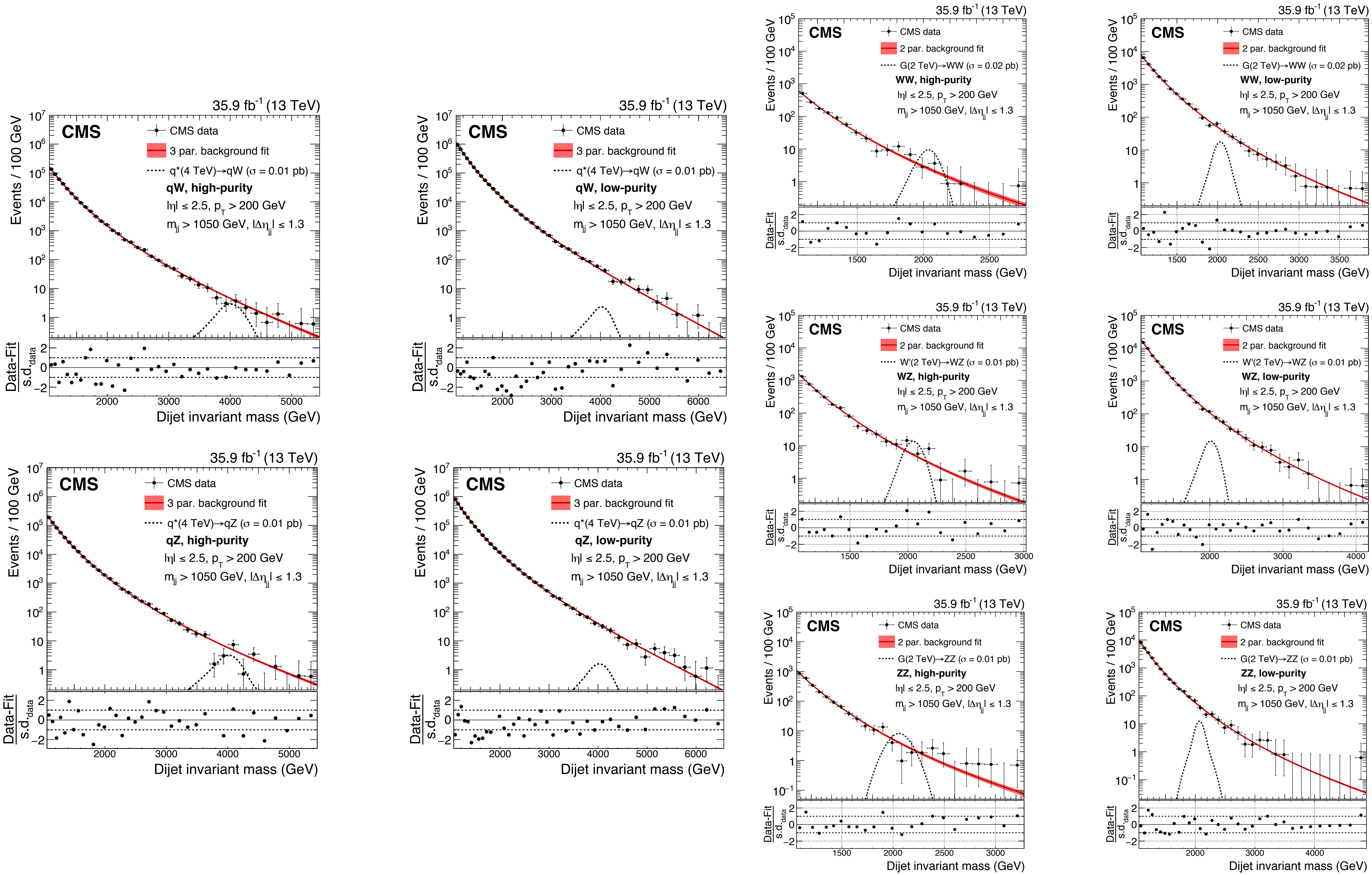
$${}_1e_2^1 = \sum_{1\le i<j\le n_J} z_i z_j \Delta R_{ij}$$

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





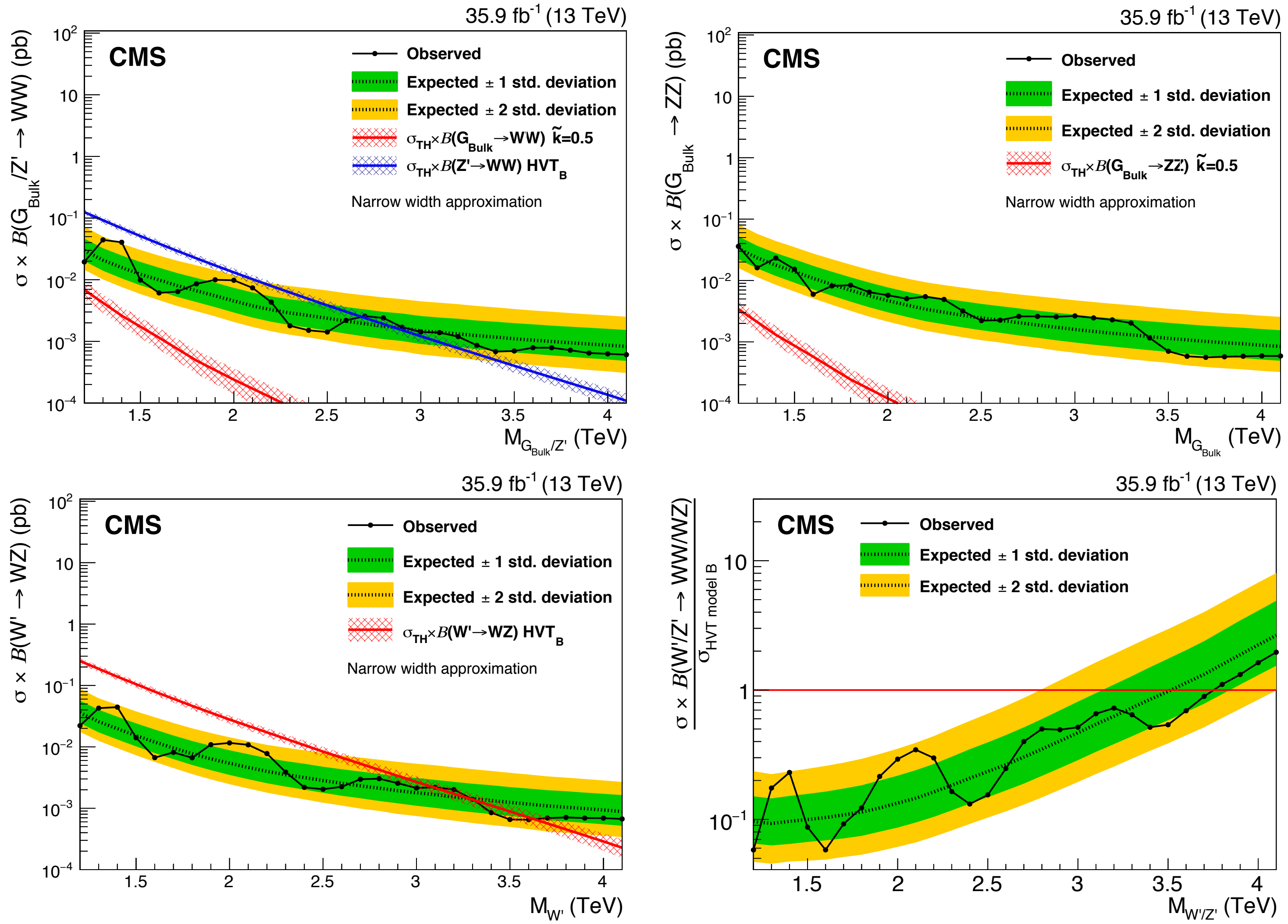
B2G-17-001



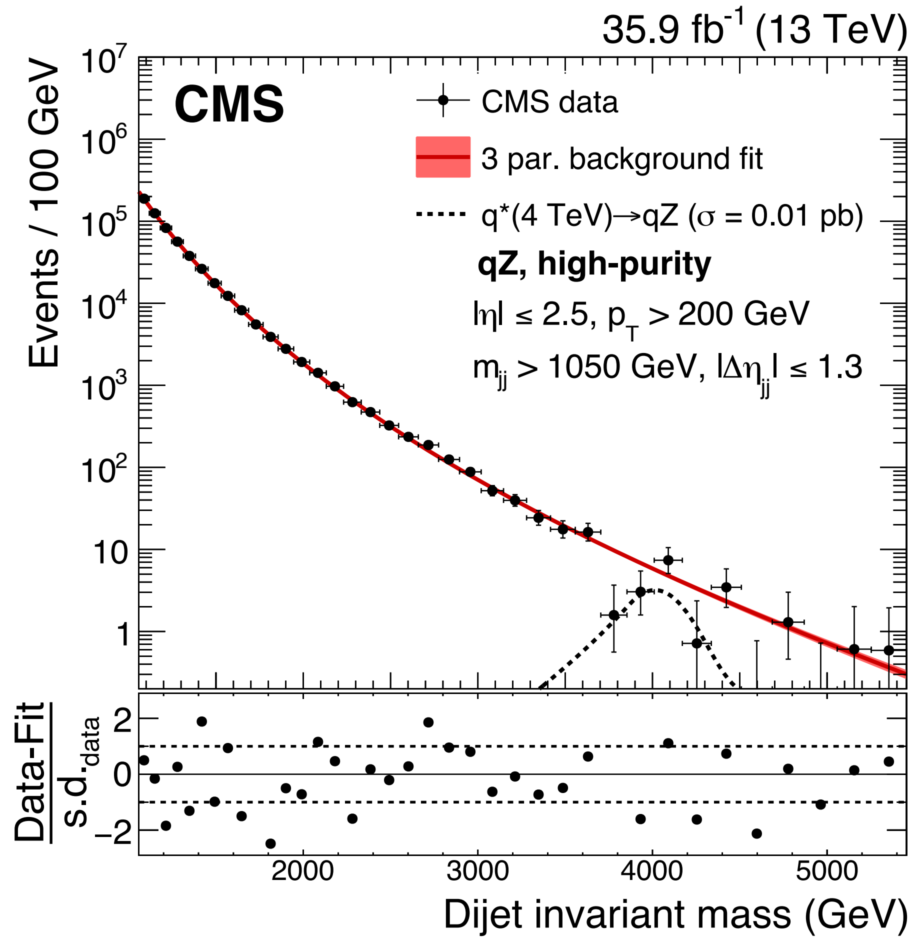
35.9 fb^{-1} (13 TeV)
CMS
CMS data
3 par. background fit
q*(4 TeV)→qW (σ = 0.01 pb)
qW, high-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)
35.9 fb^{-1} (13 TeV)
CMS
CMS data
3 par. background fit
q*(4 TeV)→qW (σ = 0.01 pb)
qW, low-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)
35.9 fb^{-1} (13 TeV)
CMS
CMS data
3 par. background fit
q*(4 TeV)→qZ (σ = 0.01 pb)
qZ, high-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)
35.9 fb^{-1} (13 TeV)
CMS
CMS data
3 par. background fit
q*(4 TeV)→qZ (σ = 0.01 pb)
qZ, low-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)
35.9 fb^{-1} (13 TeV)
CMS
CMS data
2 par. background fit
G(2 TeV)→WW (σ = 0.02 pb)
WW, high-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)
35.9 fb^{-1} (13 TeV)
CMS
CMS data
2 par. background fit
G(2 TeV)→WW (σ = 0.02 pb)
WW, low-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)
35.9 fb^{-1} (13 TeV)
CMS
CMS data
2 par. background fit
W'(2 TeV)→WZ (σ = 0.01 pb)
WZ, high-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)
35.9 fb^{-1} (13 TeV)
CMS
CMS data
2 par. background fit
W'(2 TeV)→WZ (σ = 0.01 pb)
WZ, low-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)
35.9 fb^{-1} (13 TeV)
CMS
CMS data
2 par. background fit
G(2 TeV)→ZZ (σ = 0.01 pb)
ZZ, high-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)
35.9 fb^{-1} (13 TeV)
CMS
CMS data
2 par. background fit
G(2 TeV)→ZZ (σ = 0.01 pb)
ZZ, low-purity
|η| ≤ 2.5, p_T > 200 GeV
m_{jj} > 1050 GeV, |Δη_{jj}| ≤ 1.3
Events / 100 GeV
Data-Fit / s.d.$_{data}$
Dijet invariant mass (GeV)

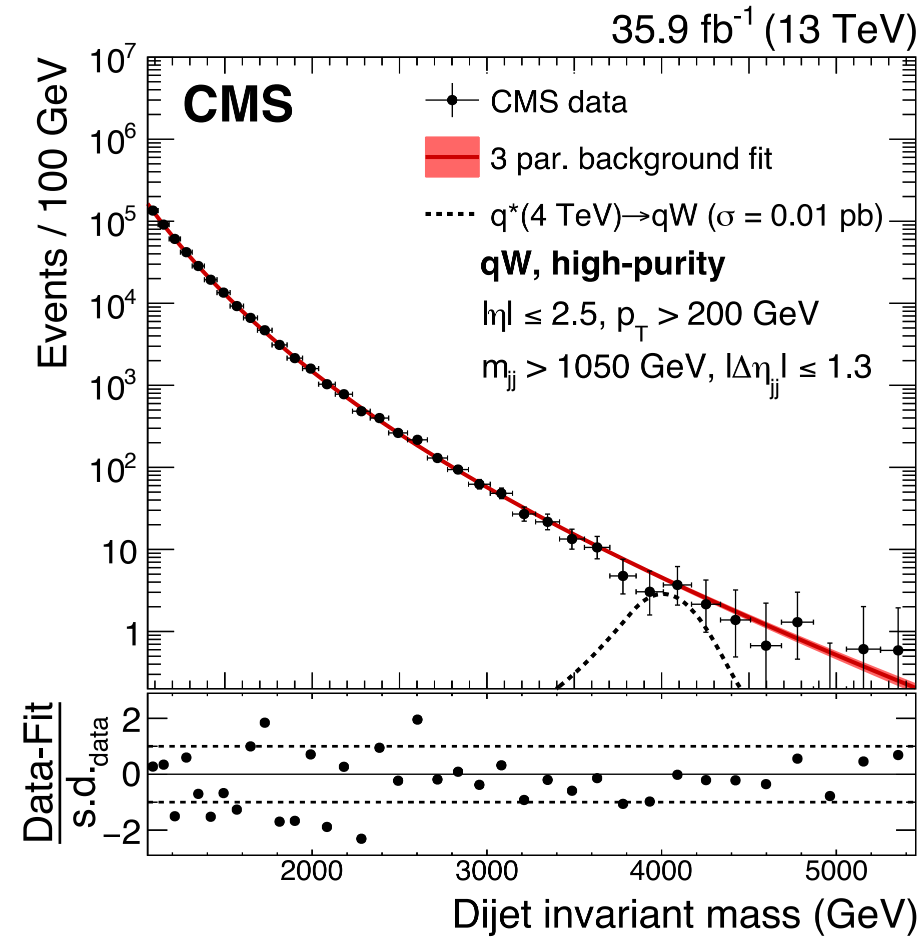
B2G-17-001

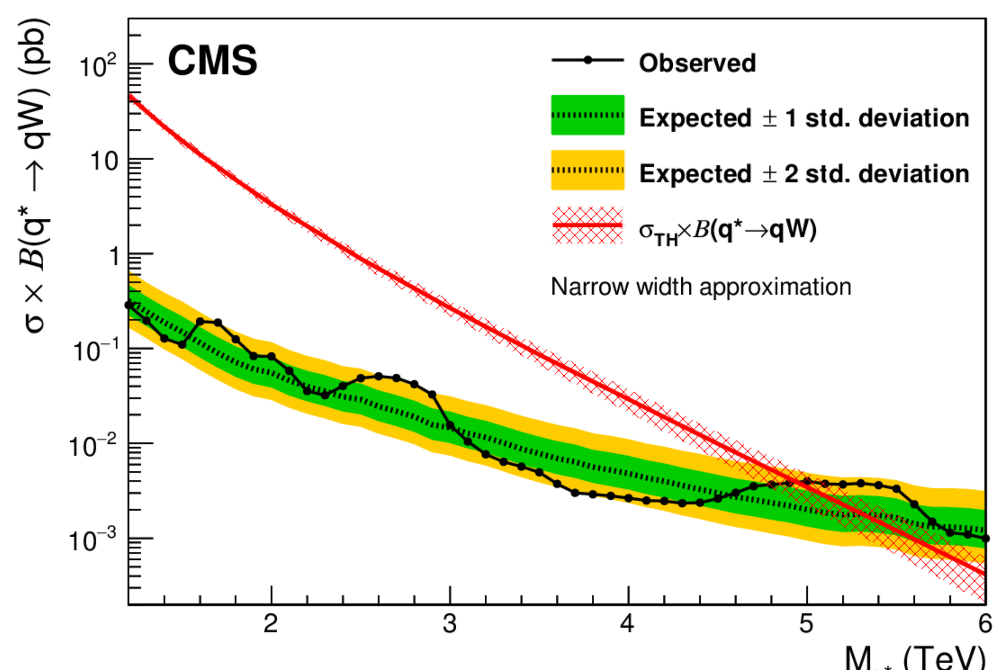
Source	Relevant quantity	Uncertainty (%)			
		Double-tag		Single-tag	
		HP+HP	HP+LP	HP+j	LP+j
Jet energy scale	Resonance shape	2	2	2	2
Jet energy resolution	Resonance shape	6	7	4	3
PDF	Resonance shape	5	7	13	8
Jet energy scale	Signal yield	<1		<1	
Jet energy resolution	Signal yield	<1		<1	
Jet mass scale	Signal yield	<2		<1	
Jet mass resolution	Signal yield	<6		<8	
Pileup	Signal yield	2			
PDF (acceptance)	Signal yield	2			
Integrated luminosity	Signal yield	2.5			
Jet mass scale	Migration	<36		<10	
Jet mass resolution	Migration	<25		<7	
V tagging τ_{21}	Migration	22	33	11	22
V tagging p_T-dependence	Migration	19–40	14–29	9–23	4–11
PDF and scales (W′ and Z′)	Theory	2–18			
PDF and scales (G_{bulk})	Theory	8–78			
PDF and scales (q*)	Theory			1–61	

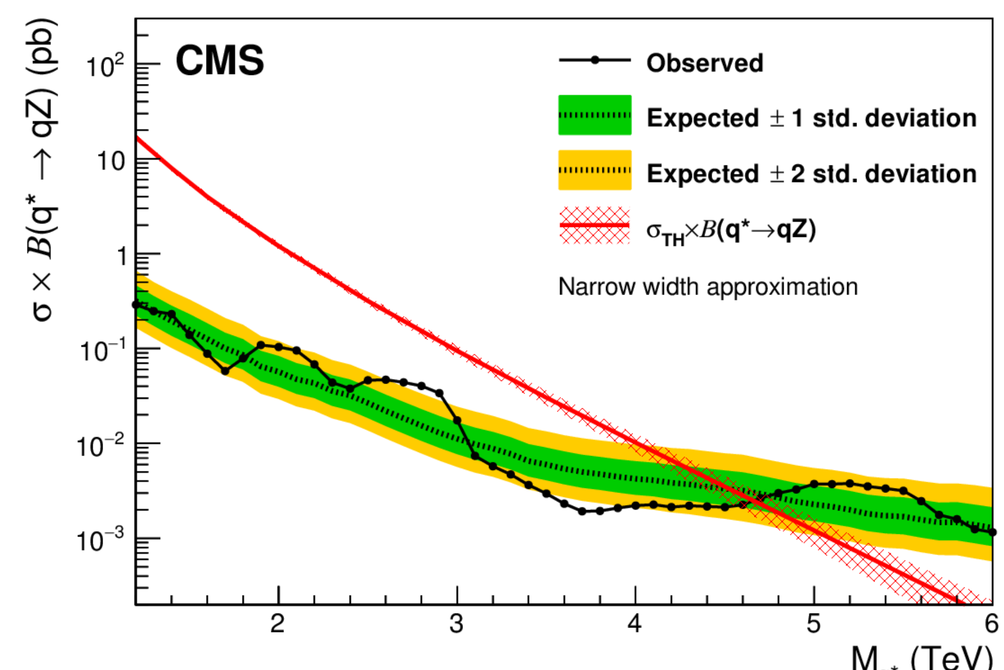


Results: Excited quarks

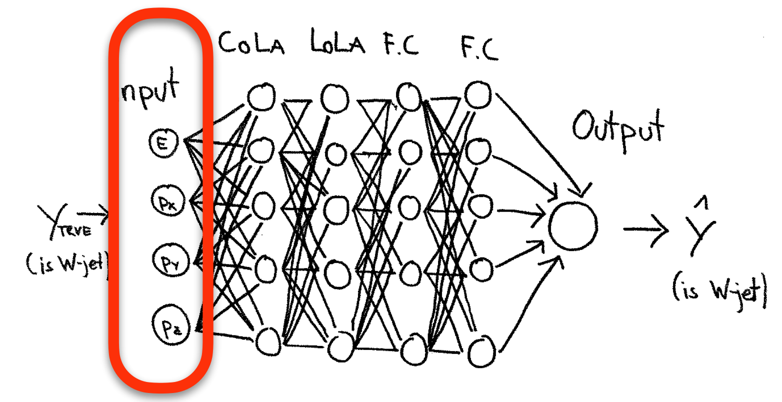








LoLa: Input



Signal

- Fully merged hadronic W-jets (AK8) from G→WW →4q ($M_{W'}$ = 0.6-4.5 TeV)
- Do not mix signal samples until one is understood (can change with W polarisation)

Background

- QCD Pythia 8 non-W jets
- Danger: Jet substructure strongly depends on shower generator (different

Four features per jet constituent

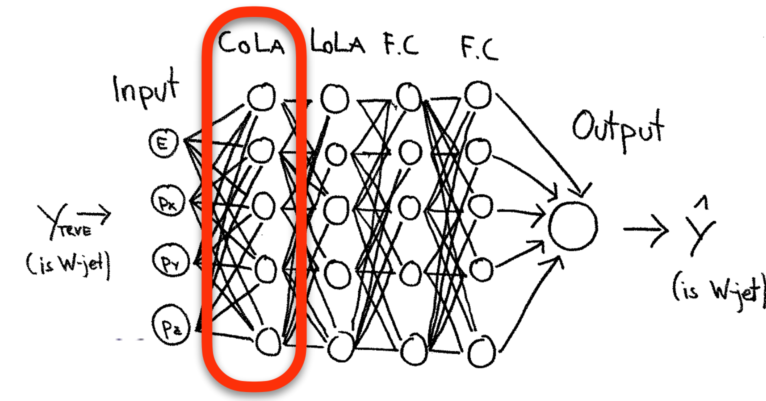
- 4-vectors of the N=20 highest-p_T jet constituents of AK8 jets

Input is 4x20 matrix $k_{\mu,i}$ per jet

$$(k_{\mu,i}) = \begin{pmatrix} k_{0,1} & k_{0,2} & \cdots & k_{0,N} \\ k_{1,1} & k_{1,2} & \cdots & k_{1,N} \\ k_{2,1} & k_{2,2} & \cdots & k_{2,N} \\ k_{3,1} & k_{3,2} & \cdots & k_{3,N} \end{pmatrix}$$

(4 Features , 20 constituents)

Combination Layer (CoLa)



COLA: E.g for 2 constituents

$$k_{\mu,i} = \underbrace{\begin{bmatrix} E_1 & E_2 \\ p_x^1 & p_x^2 \\ p_y^1 & p_y^2 \\ p_z^1 & p_z^2 \end{bmatrix}}_{(4,2)}$$

CoLa

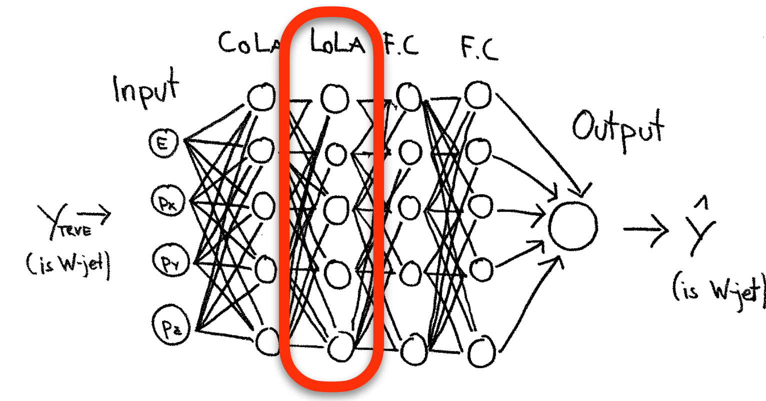
$$\tilde{k}_{\mu,j} = k_{\mu,i}\boxed{C_{ij}} = \overbrace{\underbrace{\begin{bmatrix} E_1 & E_2 \\ p_x^1 & p_x^2 \\ p_y^1 & p_y^2 \\ p_z^1 & p_z^2 \end{bmatrix}}_{(4,2)}}^{k_{\mu,i}} \cdot \overbrace{\underbrace{\begin{bmatrix} 1 & 1 & 0 & w_{4,1} \\ 1 & 0 & 1 & w_{4,2} \end{bmatrix}}_{(2,M=4)}}^{\text{CoLa}} = \begin{bmatrix} E_1+E_2 & E_1 & E_2 & w_{4,1}E_1 + w_{4,2}E_2 \\ p_x^1+p_x^2 & p_x^1 & p_x^2 & w_{4,1}p_x^1 + w_{4,2}p_x^2 \\ p_y^1+p_y^2 & p_y^1 & p_y^2 & w_{4,1}p_y^1 + w_{4,2}p_y^2 \\ p_z^1+p_z^2 & p_z^1 & p_z^2 & w_{4,1}p_z^1 + w_{4,2}p_z^2 \end{bmatrix}$$

$\sum E$, $\sum p_x$, $\sum p_y$, $\sum p_z$ — p_4^1 — p_4^2 — linear combinations of momenta

Linear combinations similar to jet-clustering

- Sum of all momenta
- Each original constituent momenta
- Linear combinations + **trainable weights**. Can "weight" constituents away, reconstruct substructure axes → groomer

Lorentz Layer (LoLa)



LoLA:

LoLa

$$\tilde{k}_{w_j} = \underbrace{\begin{bmatrix} \Sigma E & E_1 & E_2 & w_{4,1}E_1 + w_{4,2}E_2 \\ \Sigma p_x & p_x^1 & p_x^2 & w_{4,1}p_x^1 + w_{4,2}p_x^2 \\ \Sigma p_y & p_y^1 & p_y^2 & w_{4,1}p_y^1 + w_{4,2}p_y^2 \\ \Sigma p_z & p_z^1 & p_z^2 & w_{4,1}p_z^1 + w_{4,2}p_z^2 \end{bmatrix}}_{(4,4)} \rightarrow \hat{k}_j = \underbrace{\begin{bmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ w_{jm}^{E}\, E(\tilde{k}_m) \\ w_{jm}^{1d}\, \Sigma d^2_{jm} \\ w_{jm}^{2d}\, \Sigma d^2_{jm} \\ w_{jm}^{3d}\, \min d^2_{jm} \\ w_{jm}^{4d}\, \min d^2_{jm} \end{bmatrix}}_{(7,4)} = \begin{bmatrix} g_{\mu\nu} P^{\mu}_{\Sigma p_\mu} P^{\nu}_{\Sigma p_\nu} & \cdots \\ \left|\sqrt{\Sigma p_x^2 + \Sigma p_y^2}\right| & \cdots \\ w_{jm}\, \Sigma E & \cdots \\ w_1^{min} & \cdots \\ w_2^{min} & \cdots \\ w_1^{sum} & \cdots \\ w_2^{sum} & \cdots \end{bmatrix}$$

$$\left\{ (\tilde{k}_\Sigma - \tilde{k}_m)_\mu \, g^{\mu\nu} \, (\tilde{k}_\Sigma - \tilde{k}_m)_\nu \right. \cdots$$

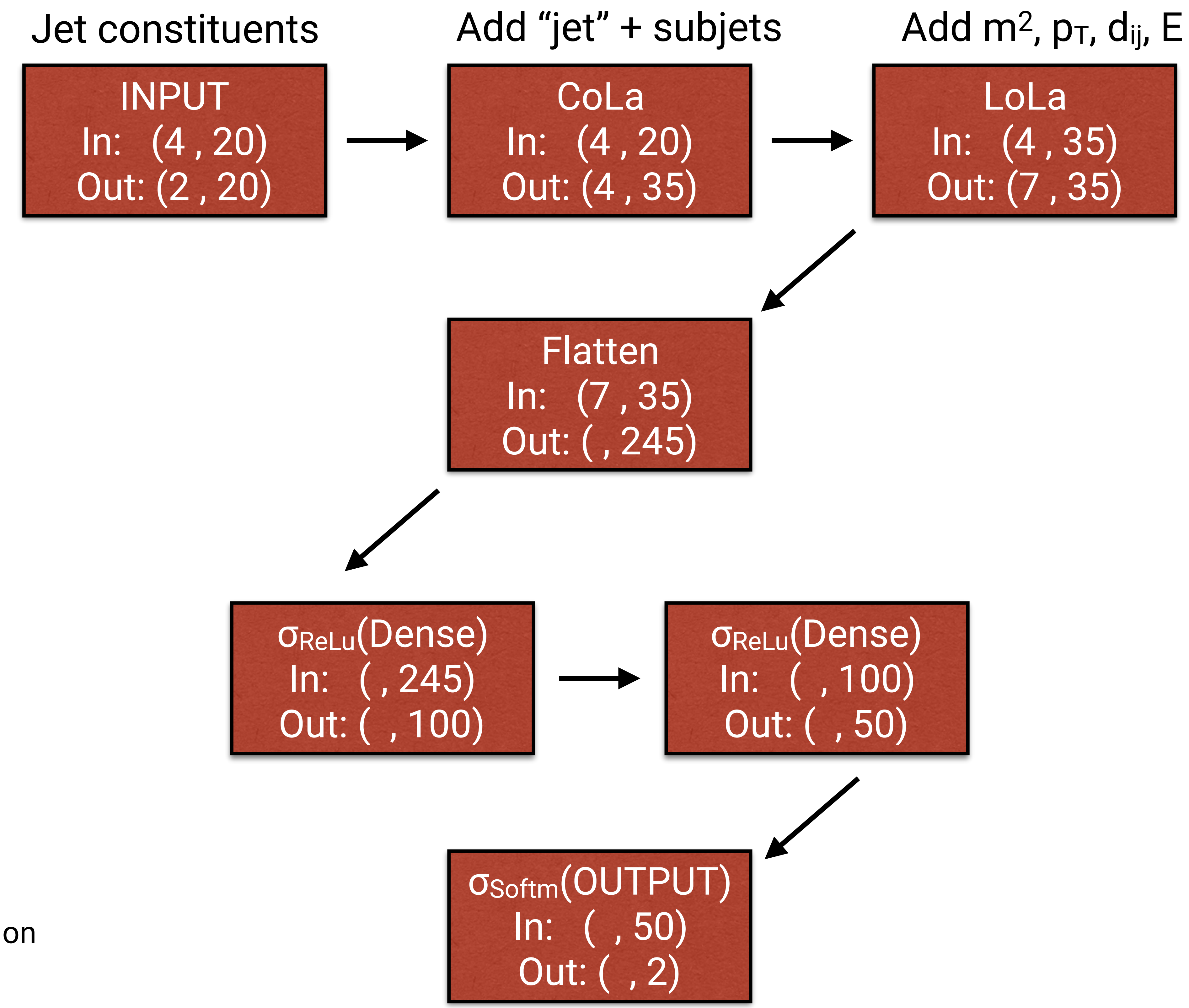
Minkowski metric explicitly used for m^2 and d

Maps CoLa output onto

- $m^2 + p_T$ of each 4-vector ("jet", constituents, subjets)
- Energy of all 4-vectors (with trainable weight)
- Distance between all 4-vectors in Minkowski space (2*min+ 2*sum)
  → n-subjetiness

Model summarised

- 4 layer DNN doing supervised learning with fixed-size input vectors
    - feed forward sequential network
    - Two novel layers (CoLa and LoLa) implementing Minkowski metric and “substructure” calculations (see later) and two fully connected layers
- Technicalities
    - Keras with Theano backend
    - Loss function: categorical crossentropy
    - ADAM optimiser (adapt learning rate of model parameters during training)
- Train 200k + Test 60k + Val 60k on AWS



LoLa: Beyond the ROC

LoLa: Beyond the ROC

Three things to consider when making a DNN tagger:

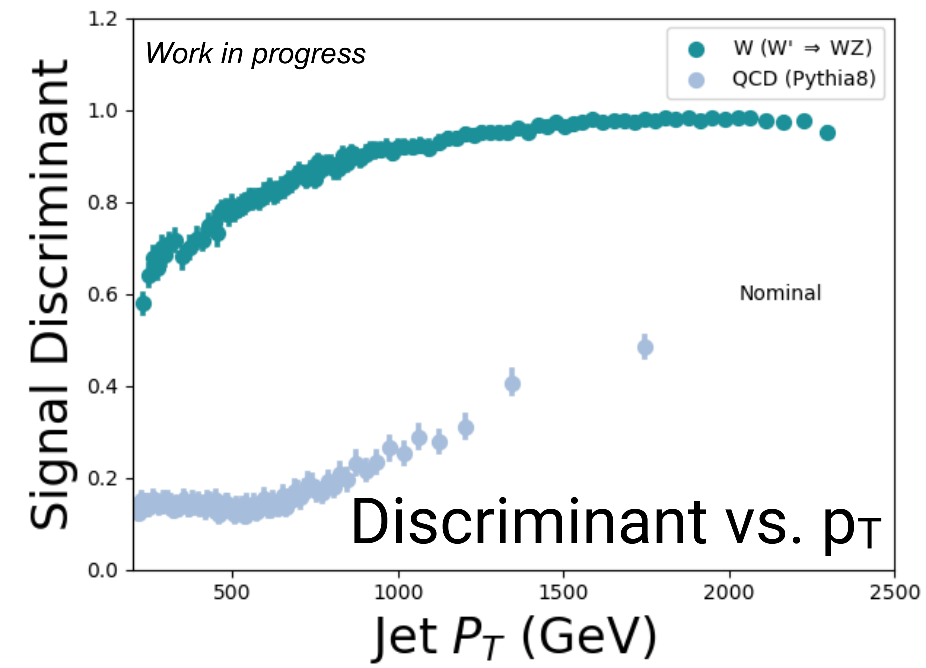
- is the absolute performance better (compared to common methods, a standard BDT)?

LoLa: Beyond the ROC

Three things to consider when making a DNN tagger:

- is the absolute performance better (compared to common methods, a standard BDT)?
- is the tagger p_T-dependent?

Output strongly correlated with p_T/mass

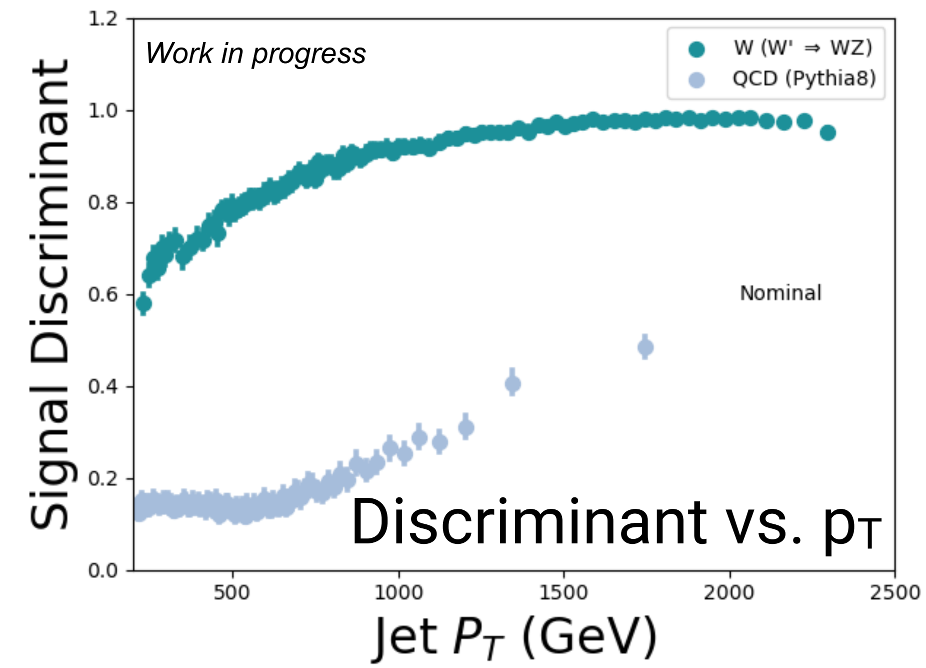


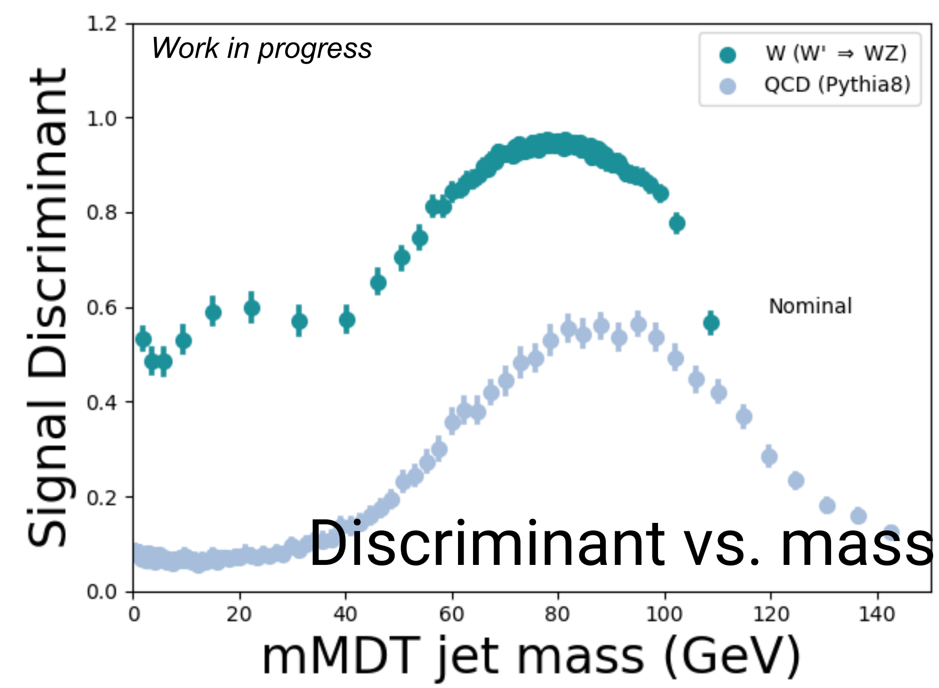
LoLa: Beyond the ROC

Three things to consider when making a DNN tagger:

- is the absolute performance better (compared to common methods, a standard BDT)?
- is the tagger p_T-dependent?
- does the tagger sculpt the mass spectrum?

Output strongly correlated with p_T/mass





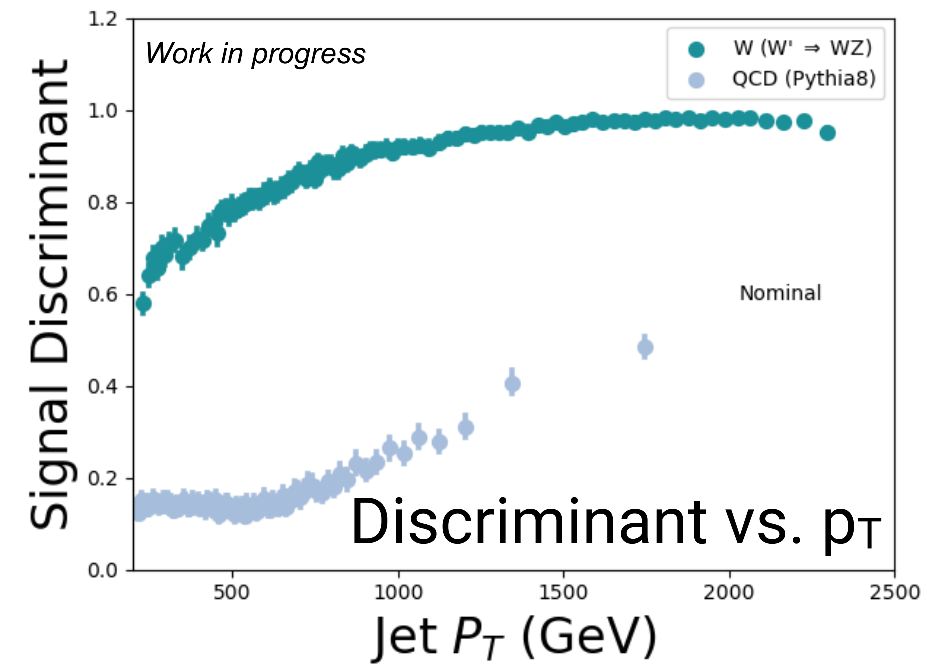
LoLa: Beyond the ROC

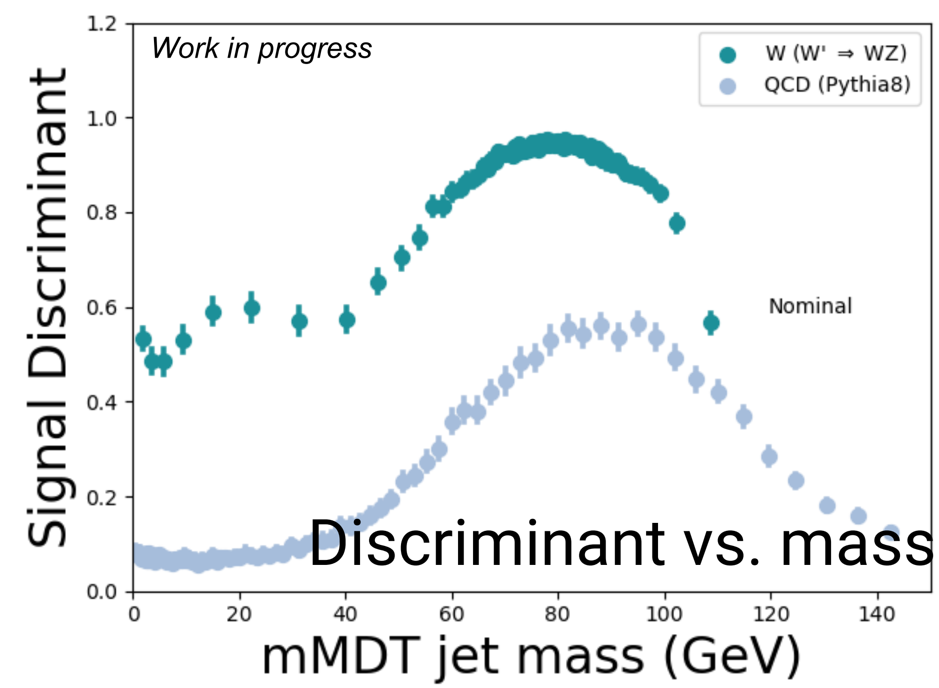
Three things to consider when making a DNN tagger:

- is the absolute performance better (compared to common methods, a standard BDT)?
- is the tagger p_T-dependent?
- does the tagger sculpt the mass spectrum?

These three measures are equally important in quantifying performance

Output strongly correlated with p_T/mass





LoLa: Beyond the ROC

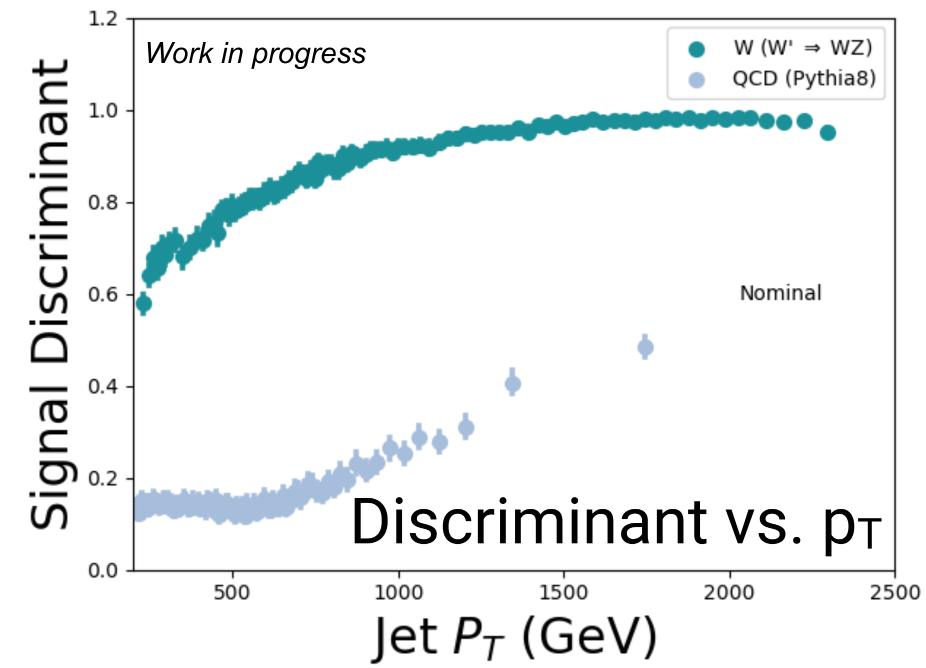
Three things to consider when making a DNN tagger:

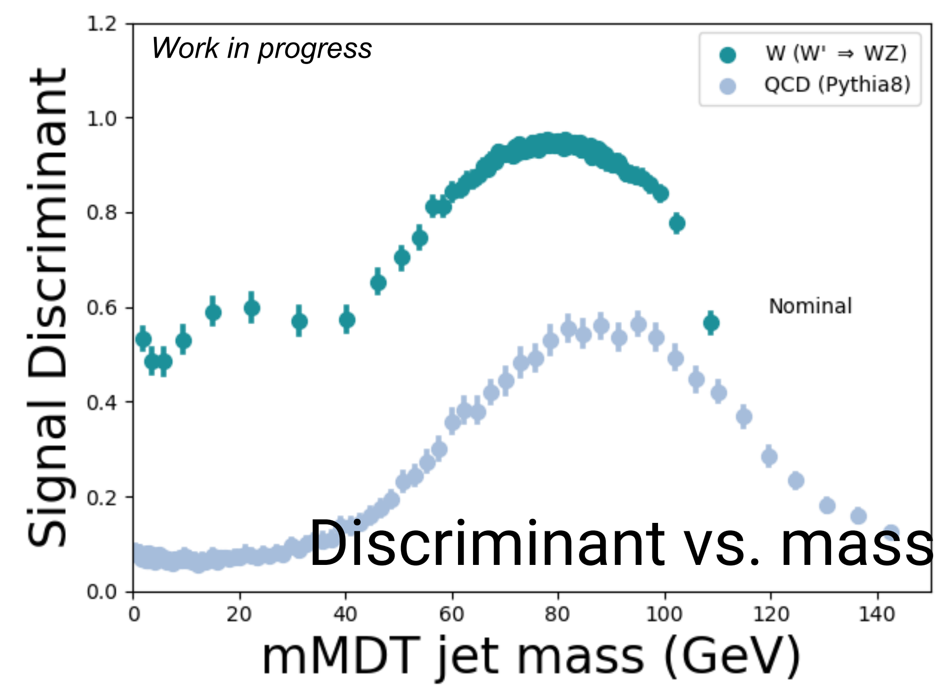
- is the absolute performance better (compared to common methods, a standard BDT)?
- is the tagger p_T-dependent?
- does the tagger sculpt the mass spectrum?

These three measures are equally important in quantifying performance

A DNN will naturally learn that p_T and mass are discriminating variables unless its penalised for it

Output strongly correlated with p_T/mass





LoLa: Beyond the ROC

Three things to consider when making a DNN tagger:

- is the absolute performance better (compared to common methods, a standard BDT)?
- is the tagger p_T-dependent?
- does the tagger sculpt the mass spectrum?

These three measures are equally important in quantifying performance

A DNN will naturally learn that p_T and mass are discriminating variables unless its penalised for it

Output strongly correlated with p_T/mass