

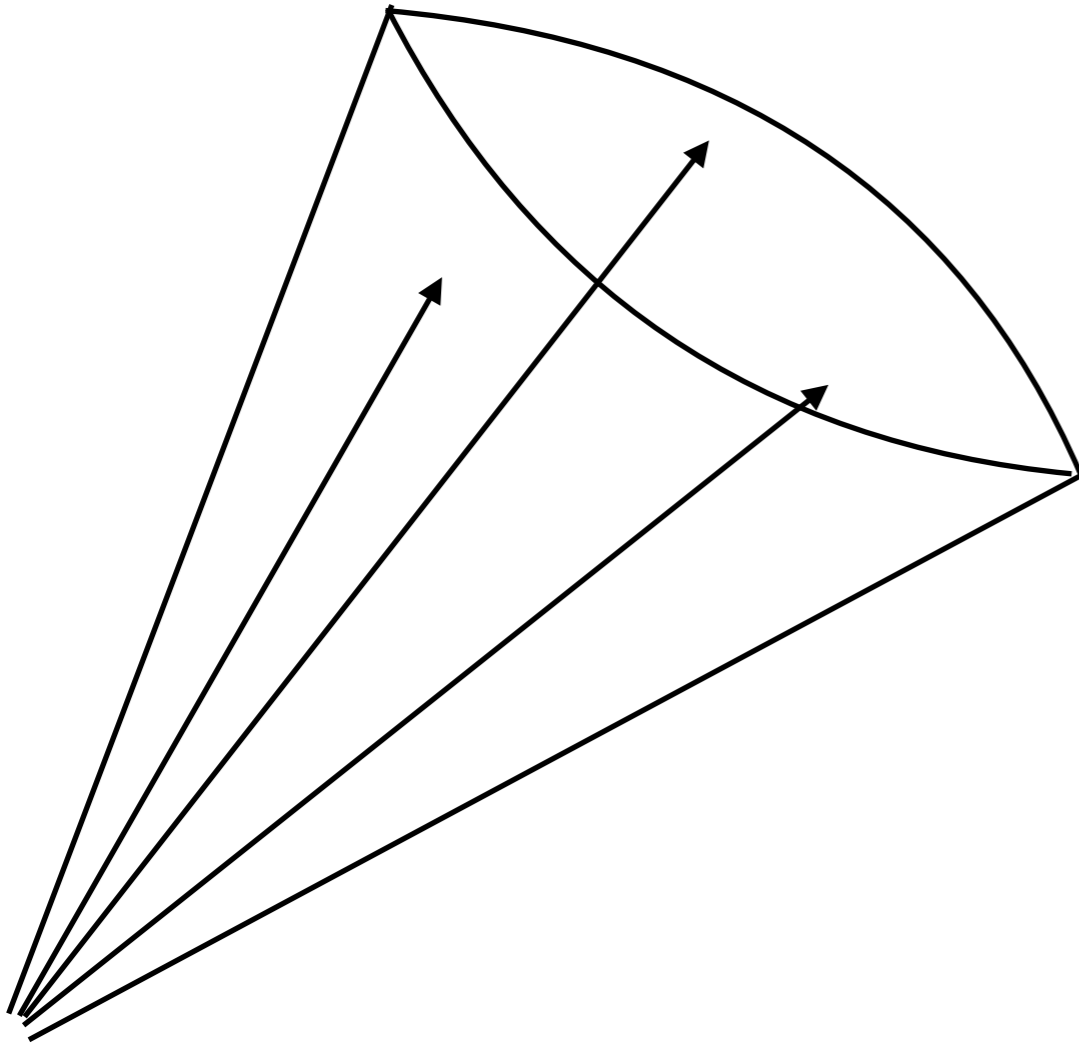
Tagger performance in Data

Cristina Mantilla Suarez on behalf of CMS



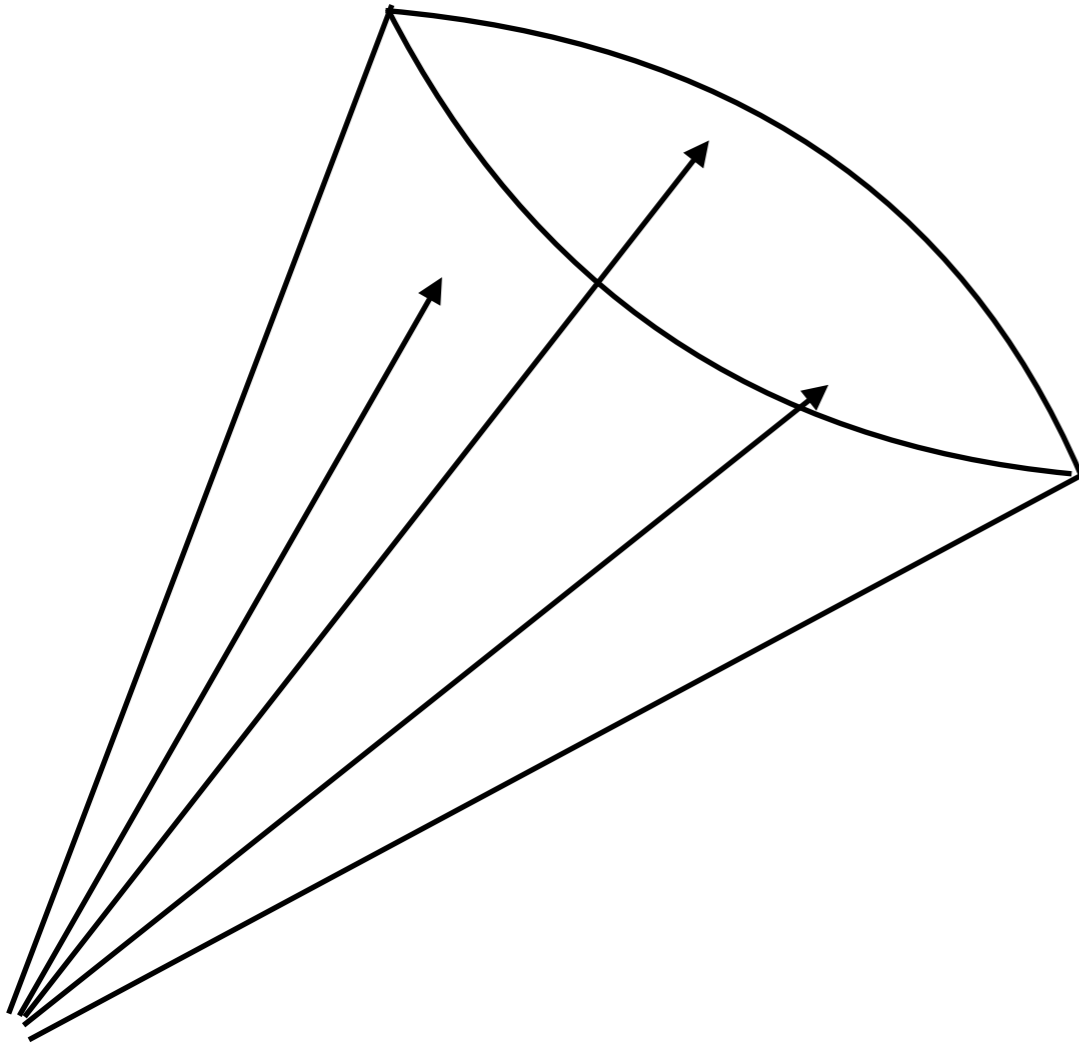
BOOST 2018

Outline



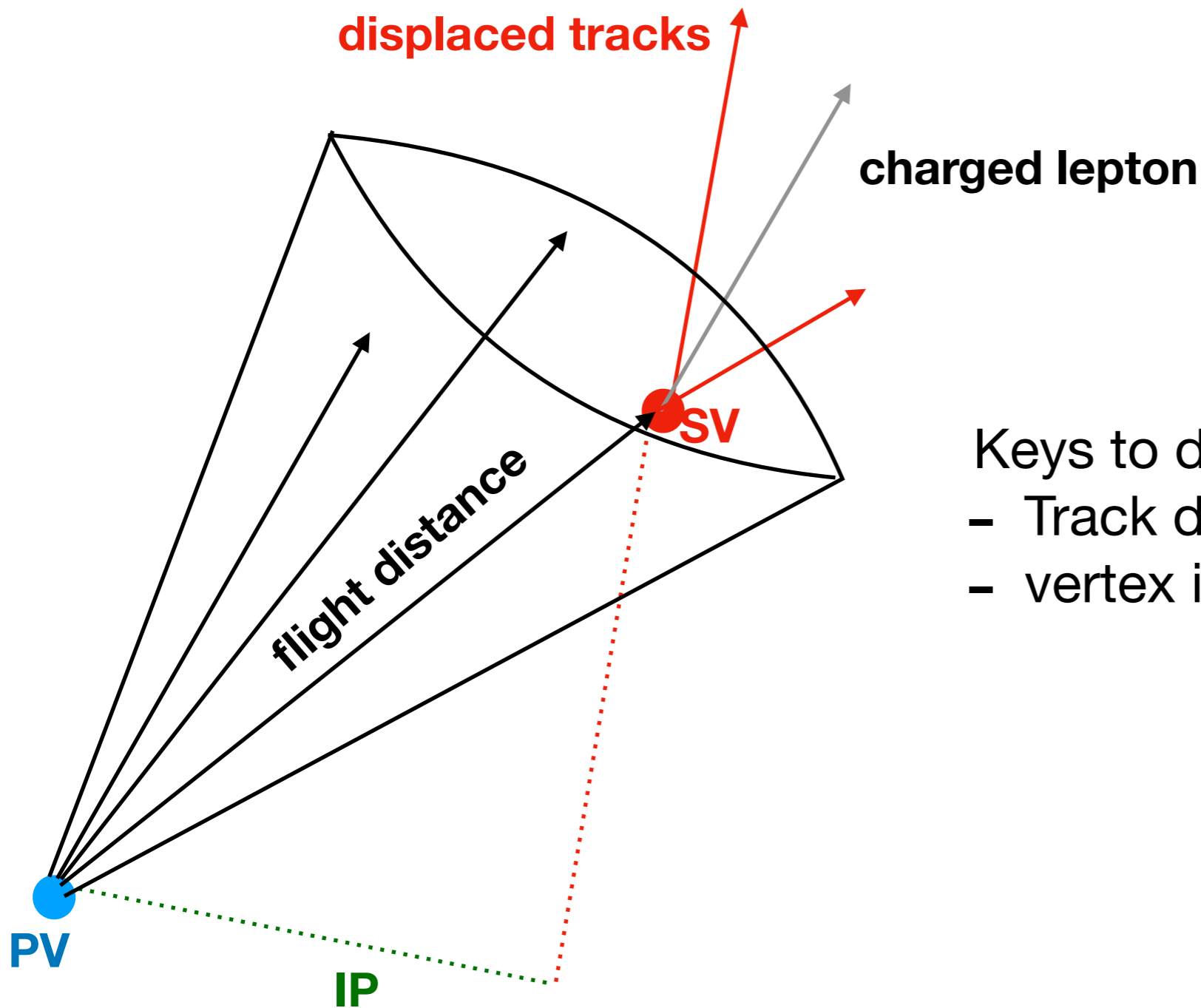
- Performance in data + limitations of current approaches
 - b-tagging
 - double b-tagging
 - W and top tagging
- + Open questions

For tagger details and new developments,
see Loukas's talk tomorrow!



Heavy flavour jets

Heavy flavour Jet identification



- Keys to distinguish from $c/(udsg)$ jets:
- Track displacement
 - vertex information

Ideal scenario to combine information:

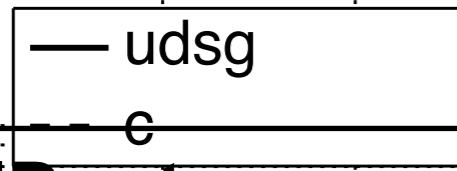
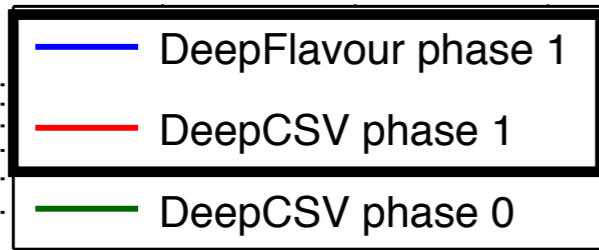
- **CSVv2** (track, SV info)
- **DeepCSV** (track, SV info+ more charged tracks)
- **DeepFlavour** (charged and neutral PF objects + SV info)

$\sqrt{s} = 13 \text{ TeV}$

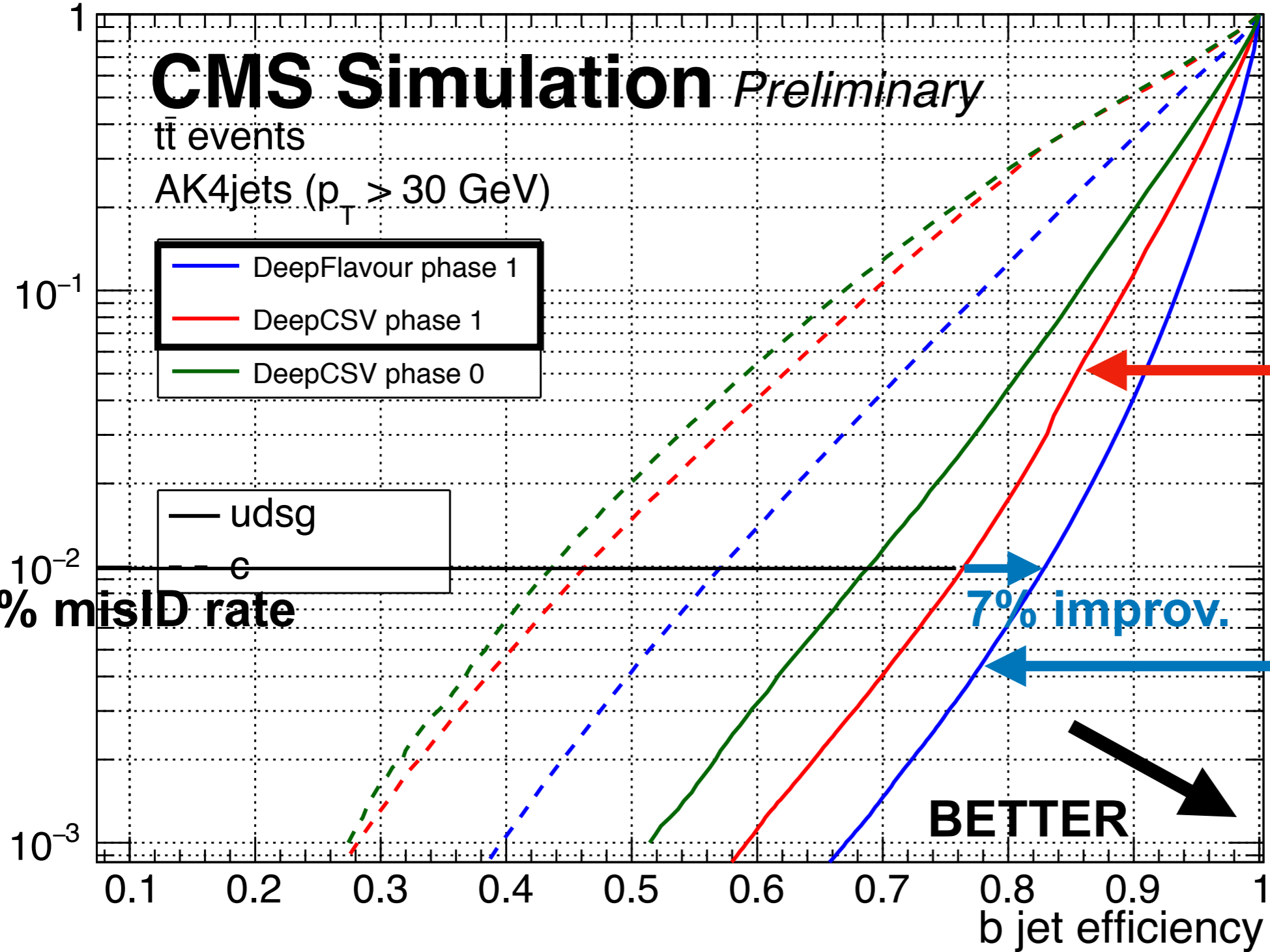
CMS Simulation *Preliminary*

$t\bar{t}$ events

AK4jets ($p_T > 30 \text{ GeV}$)



1% misID rate



CMS tagger (~65 human made inputs)

7% improv.

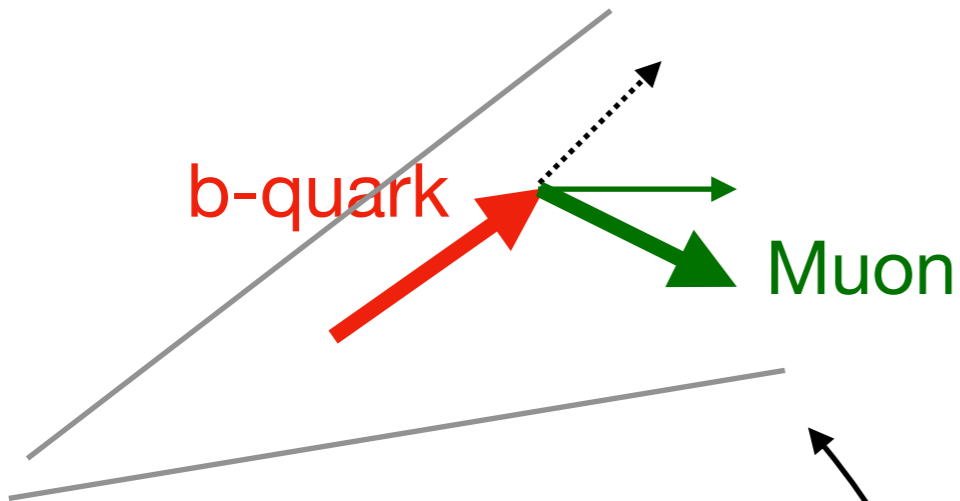
Physics inspired DNN (650 inputs)

BETTER

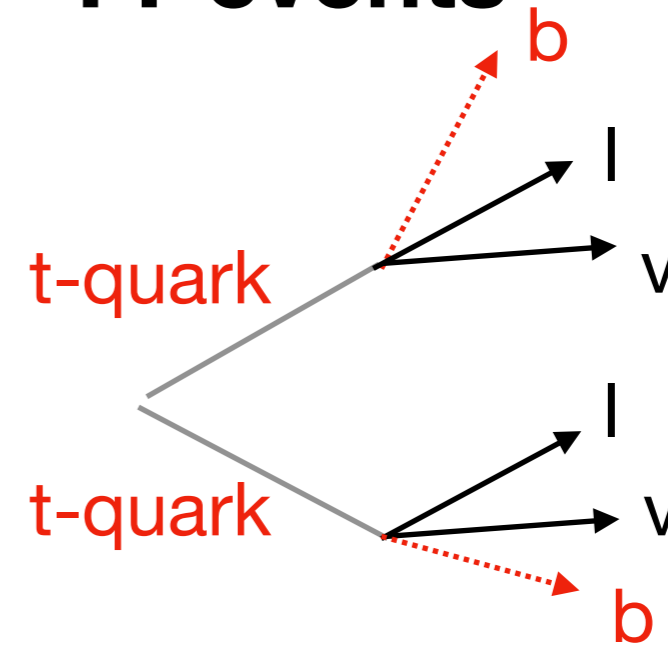
Heavy flavour in data

Where to look:

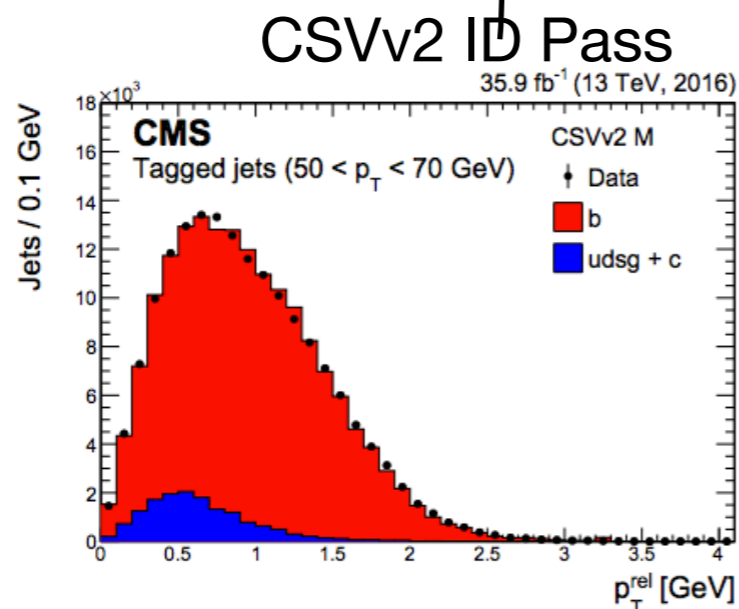
QCD events with muons



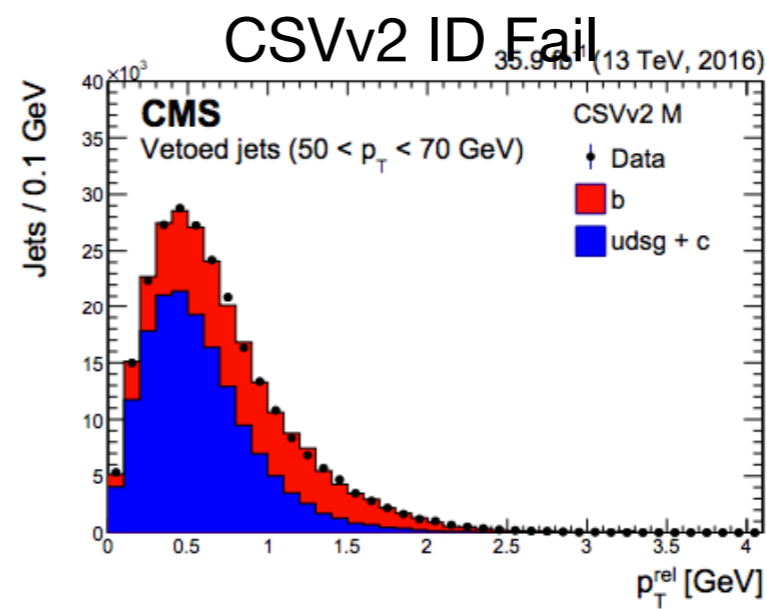
TT events



Look at discriminating variable e.g.:



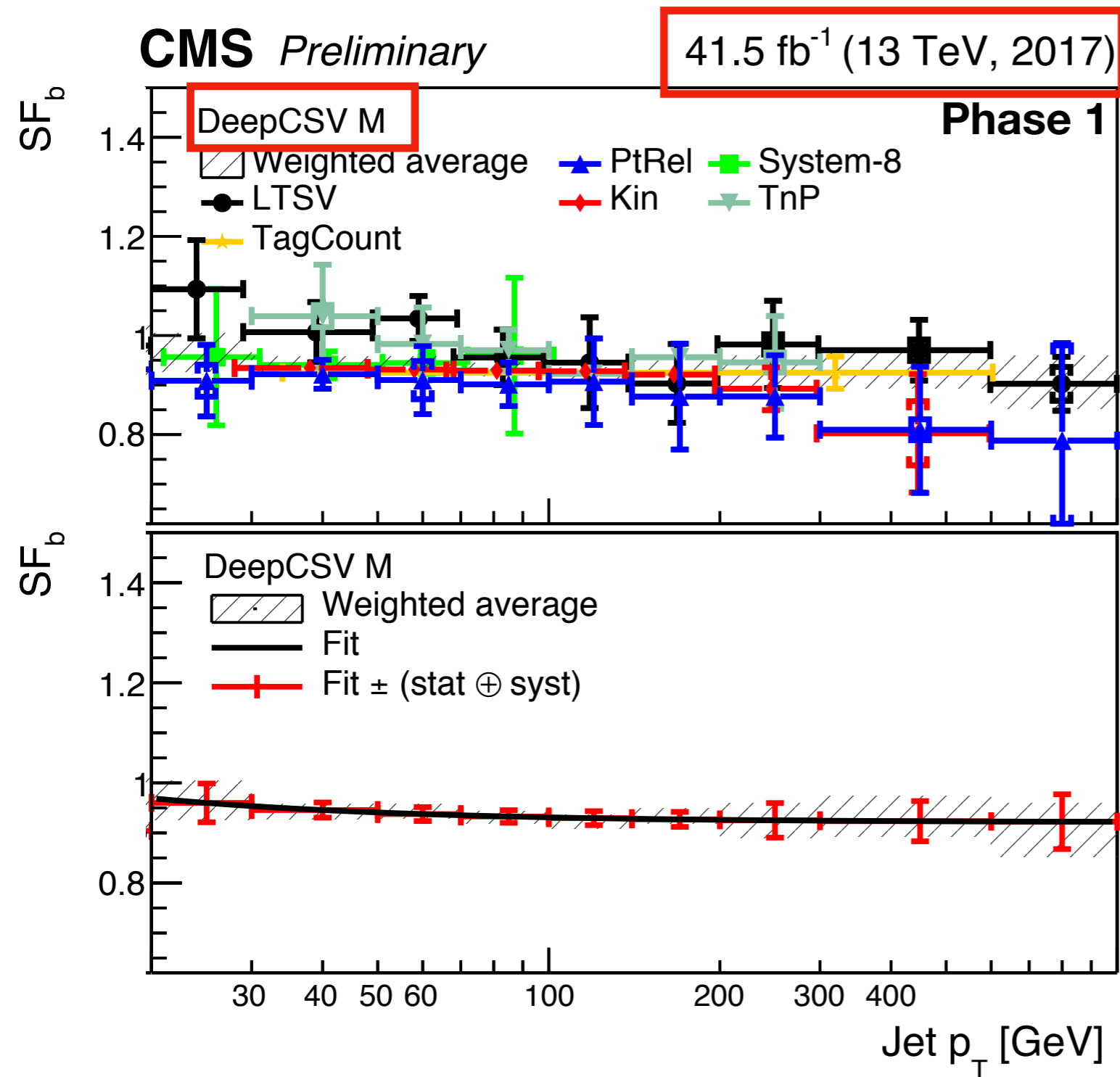
muon p_T relative to jet axis



6

Heavy flavour SF

Too many methods to go through them all: consistent within unc.



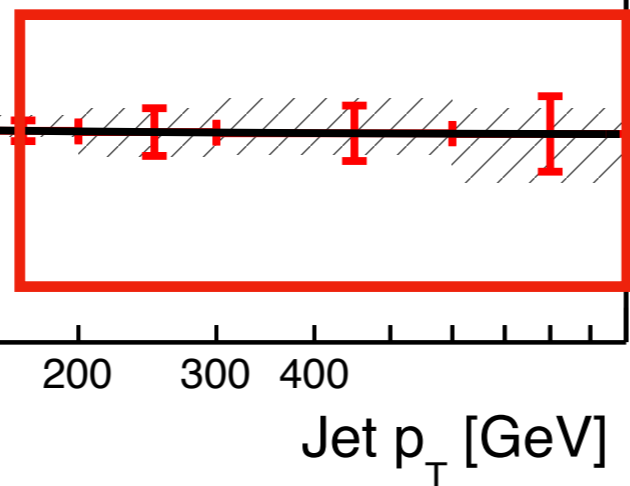
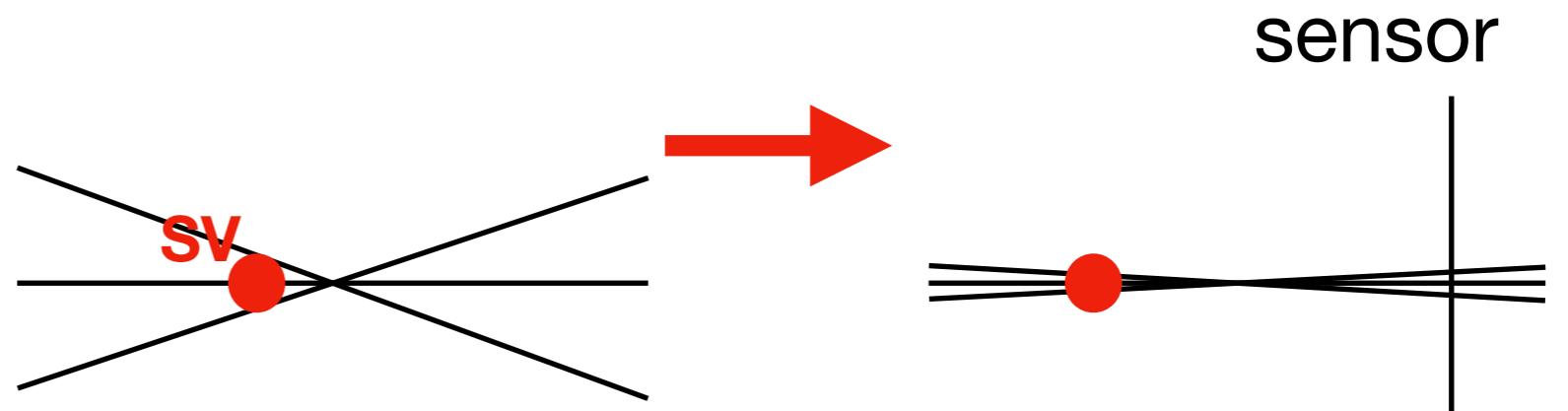
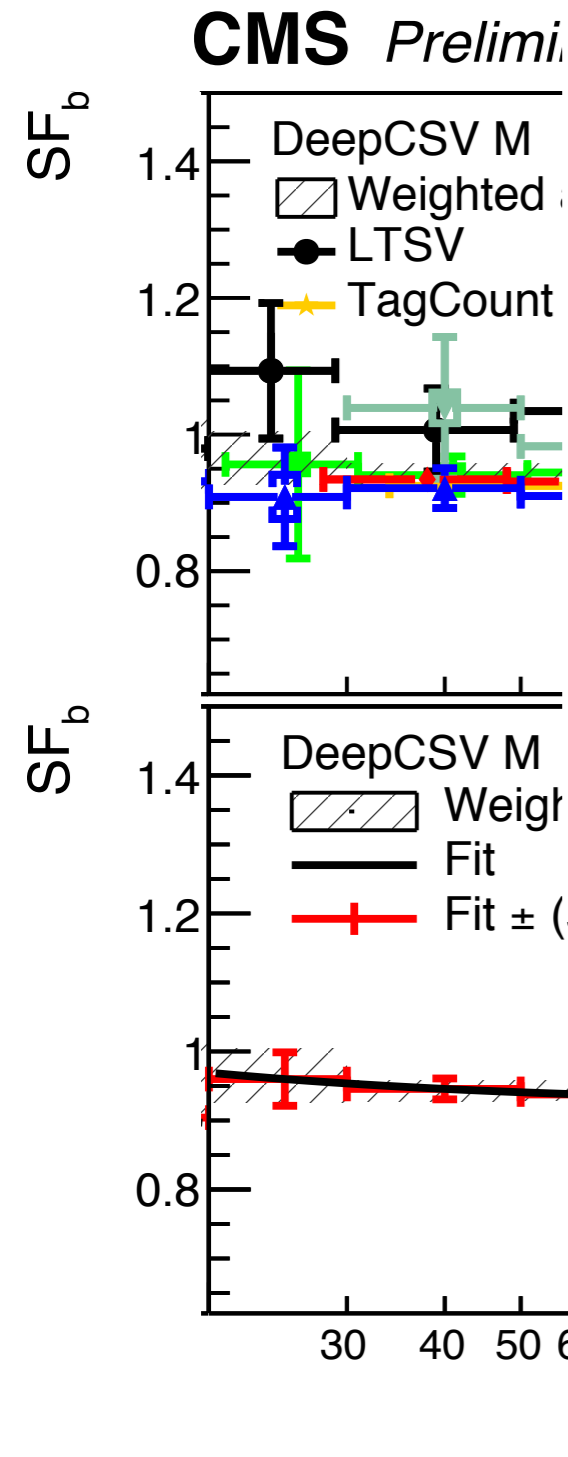
agreement between data/MC at 5% level (same for unc.)

What happens at high p_T : b-tagging

Tracking and vertexing increasingly difficult at high p_T

Tracks and # of pixel hits: more interesting

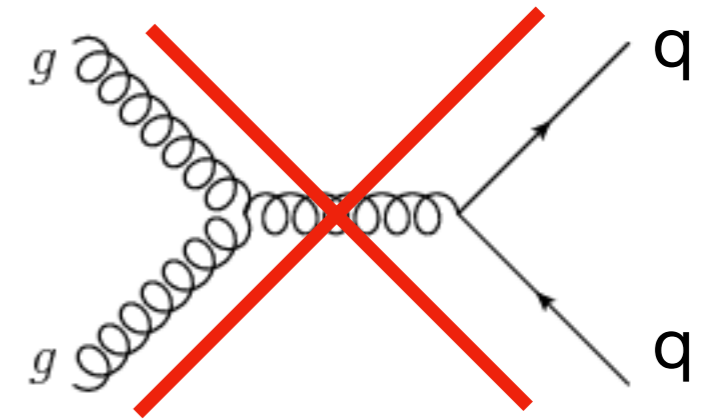
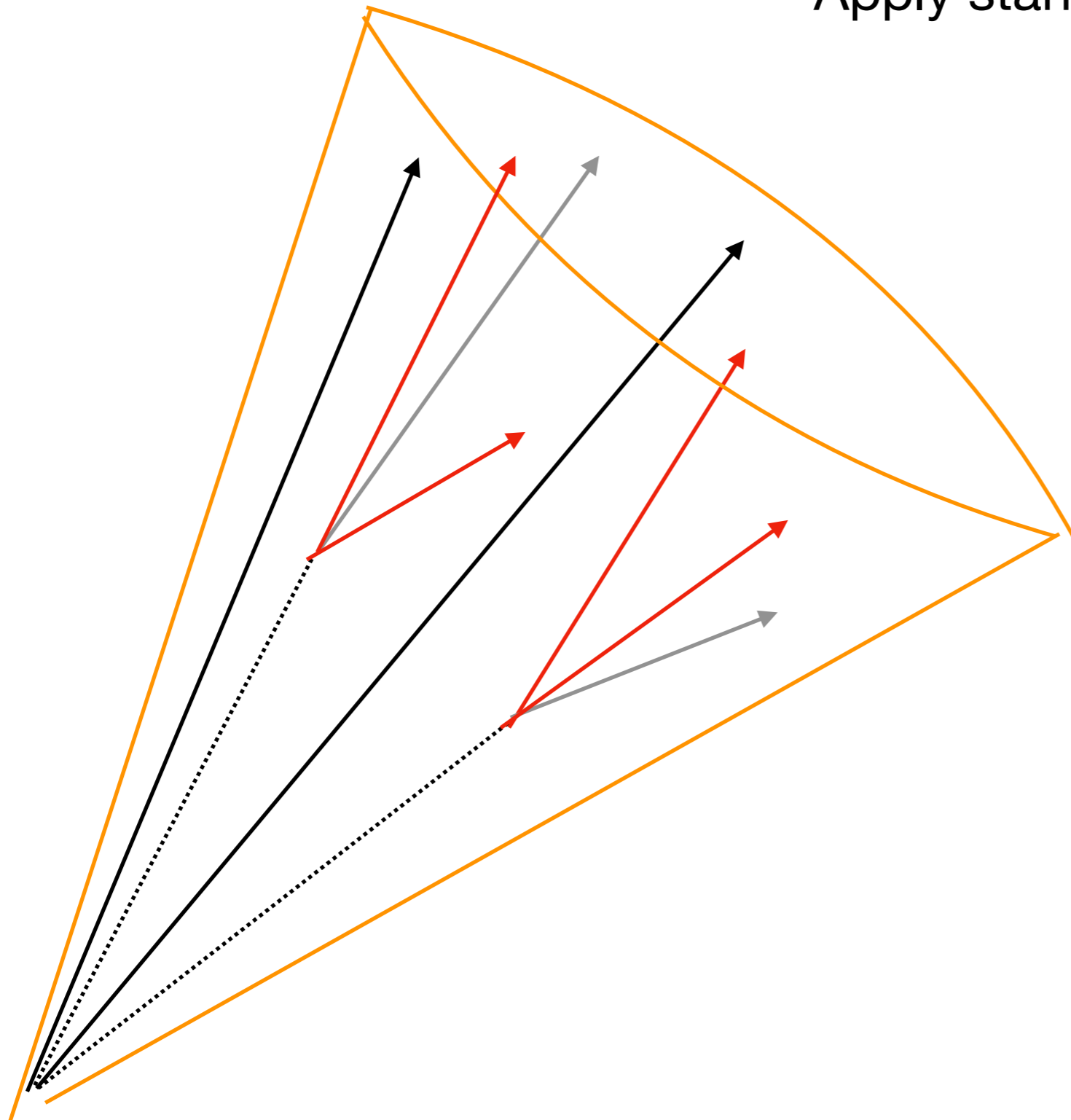
How to select tracks (and assign unc.) at high p_T ?



Heavy flavour Jet identification (boosted)

Either:

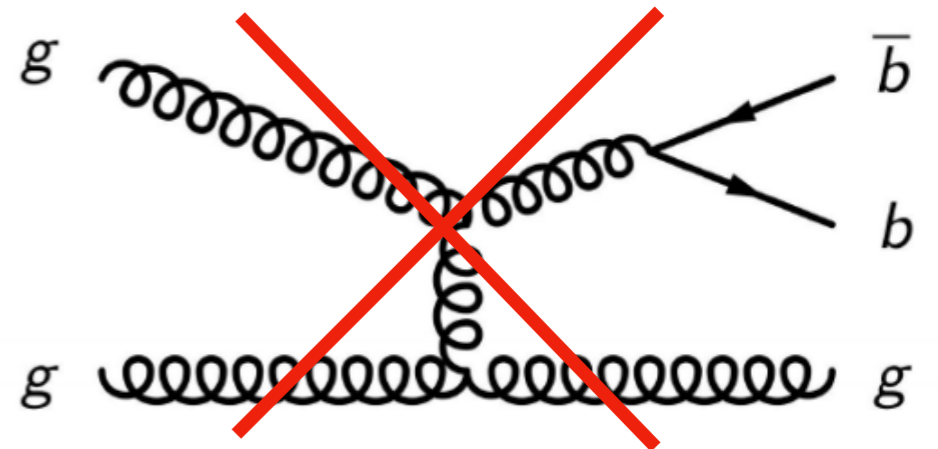
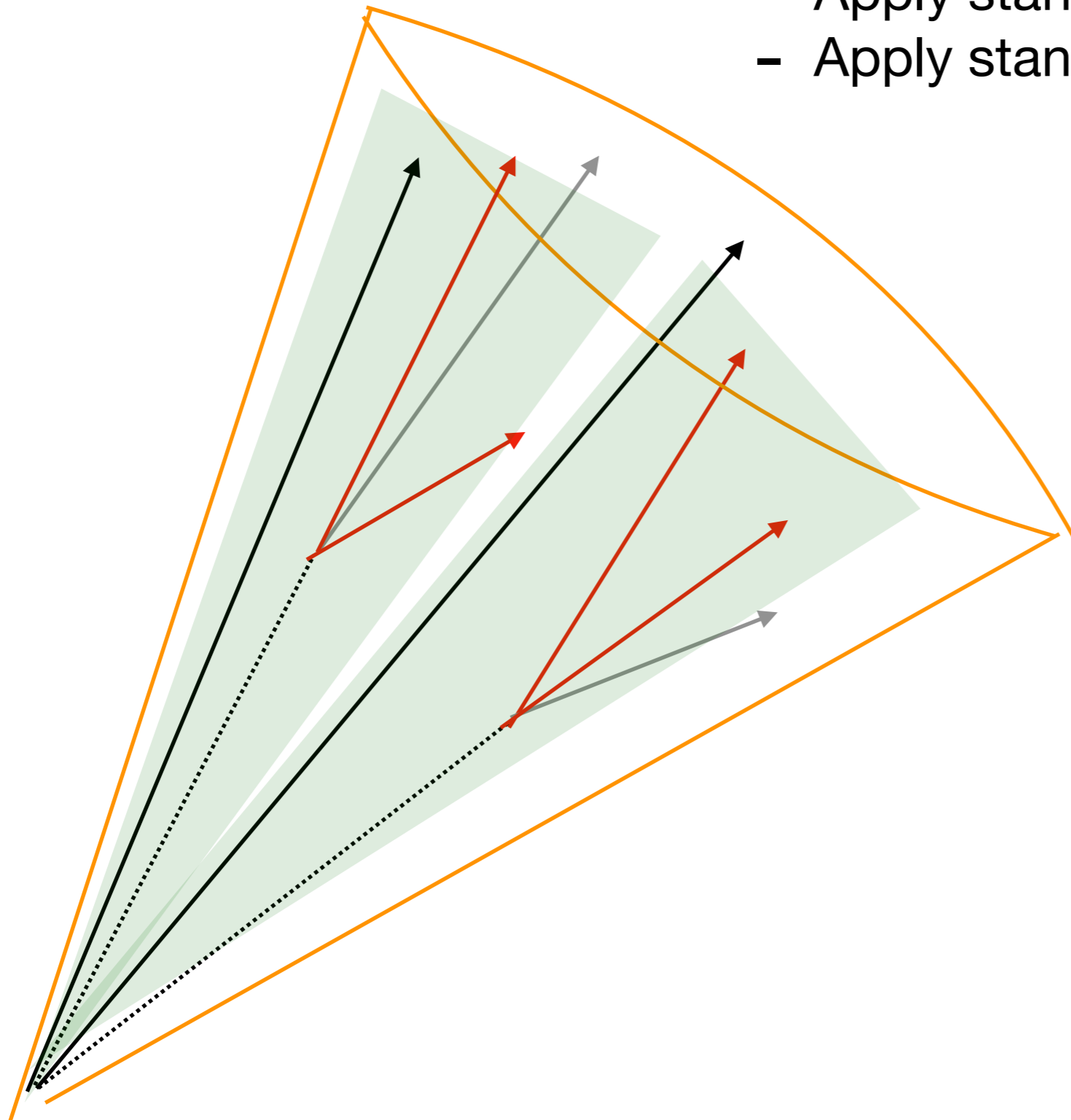
- Apply standard b-tagging to **whole jet**



Heavy flavour Jet identification (boosted)

Either:

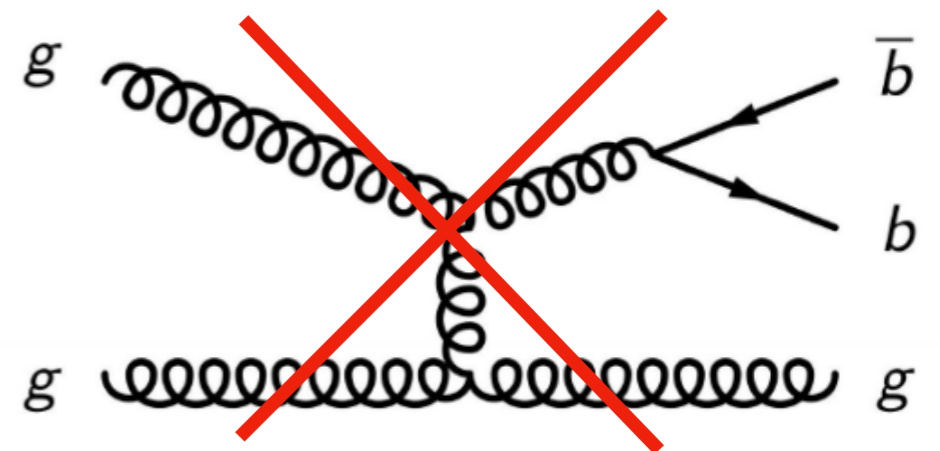
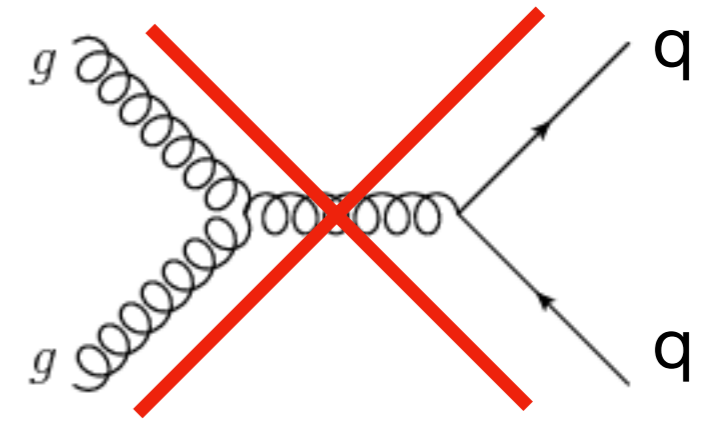
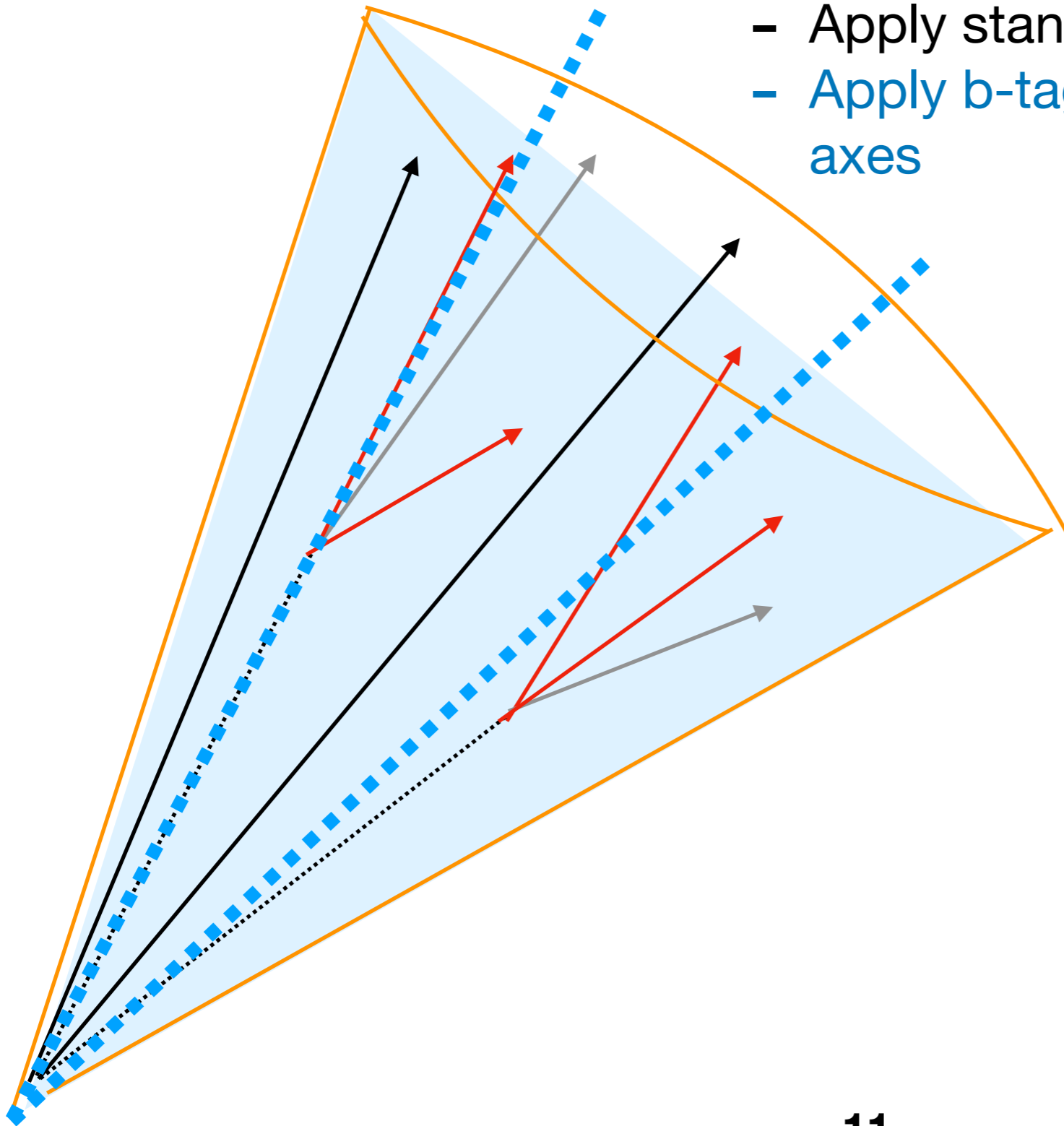
- Apply standard b-tagging to **whole jet**
- Apply standard b-tagging to **sub jet**

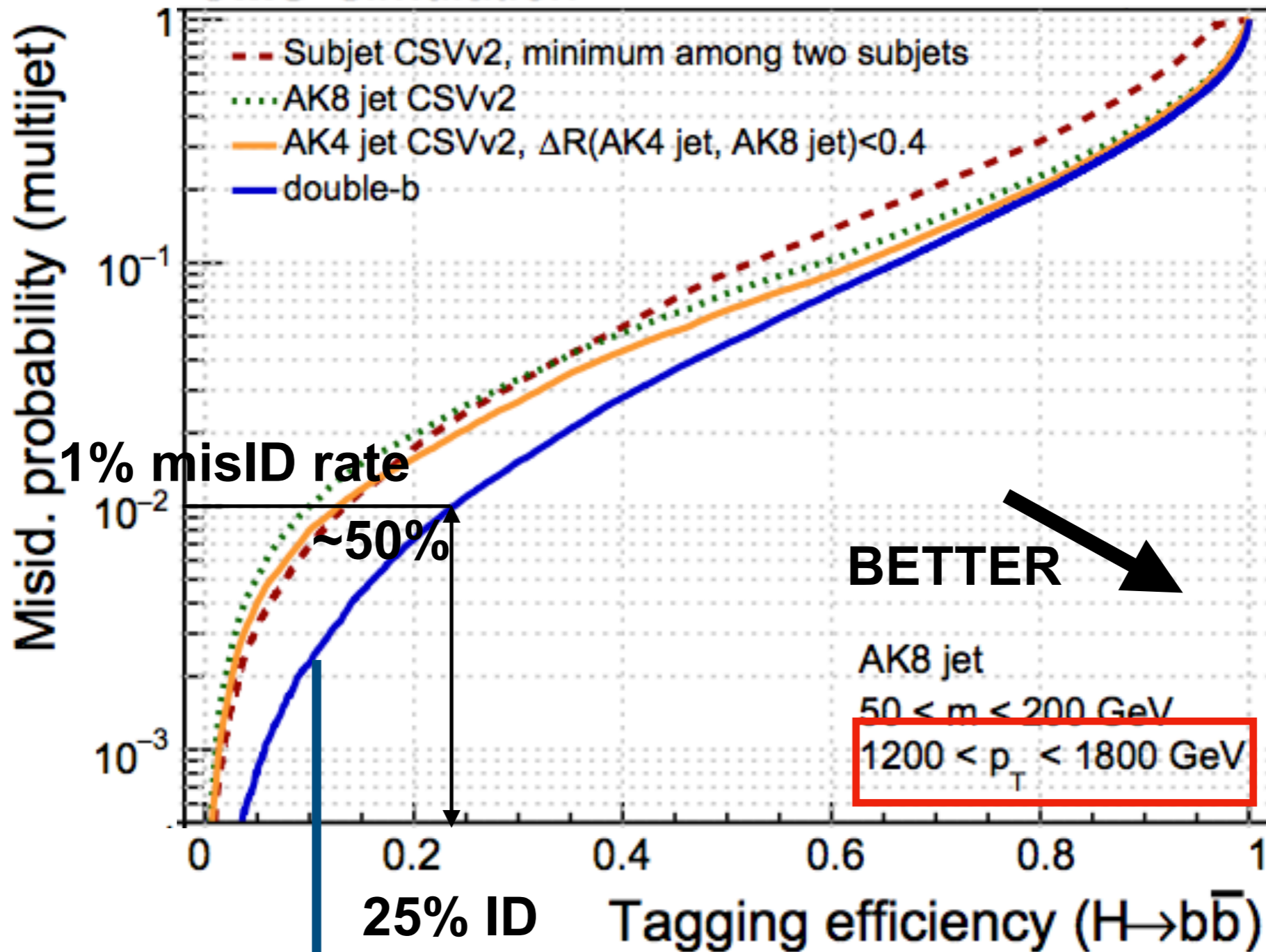


Heavy flavour Jet identification (boosted)

Either:

- Apply standard b-tagging to **whole jet**
- Apply standard b-tagging to **sub jet**
- Apply b-tagging + adding n-subjetiness axes



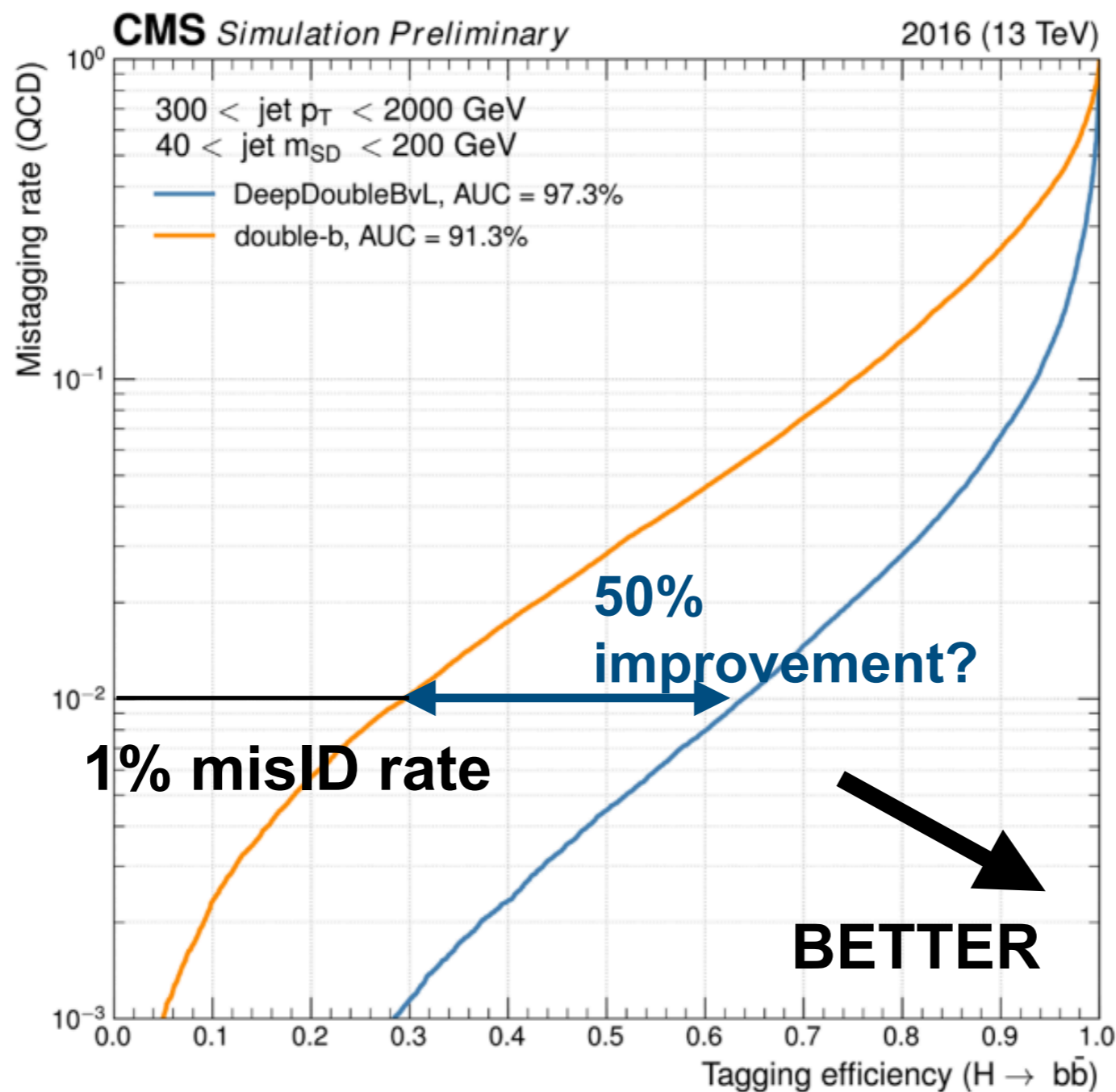


Double-b: Combination gives over 50% improvement over usual approach

Coming soon: **Deep Double-b/c**

NEW: What can DNN bring!
Not measured in data **yet**

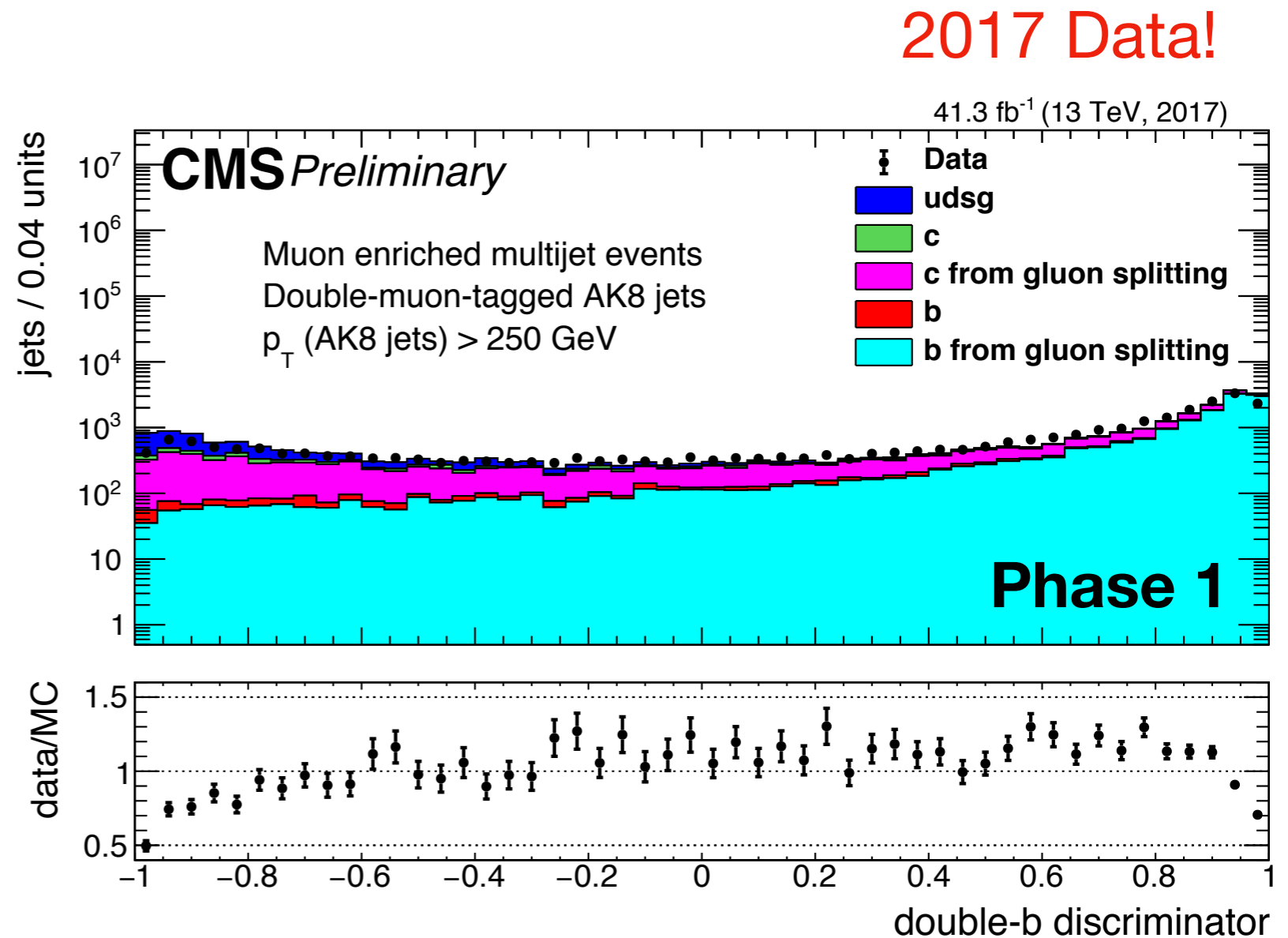
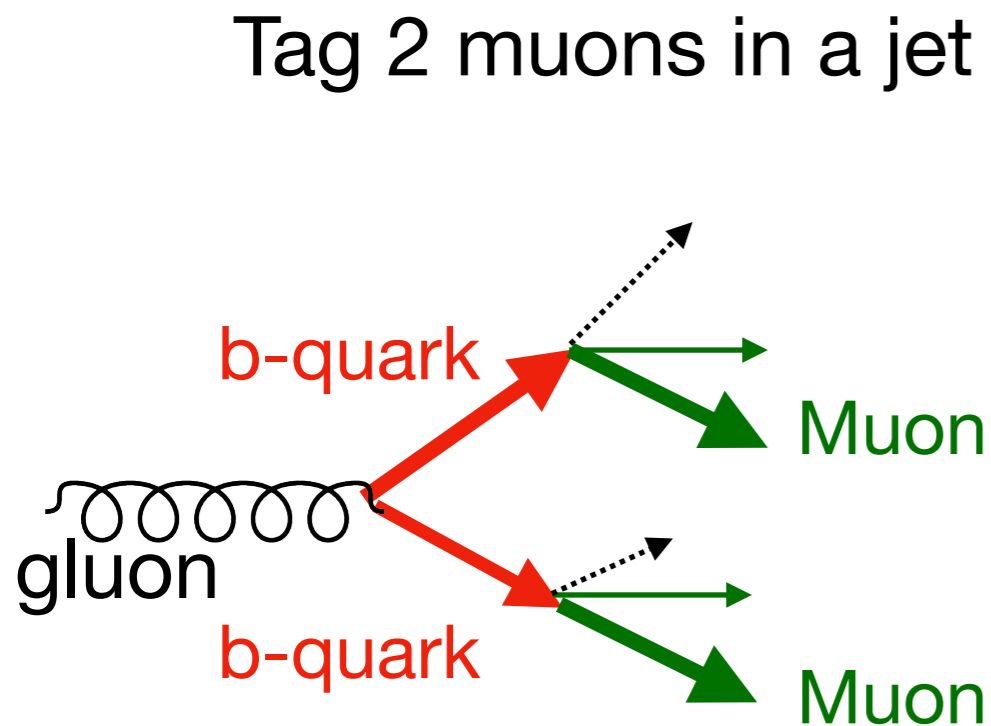
See Andrzej's POSTER!



+Cutting on double b-tag is invariant of mass

Heavy flavour in data

Where to look: $H \rightarrow bb$ would be ideal but...
get $g \rightarrow bb$ instead



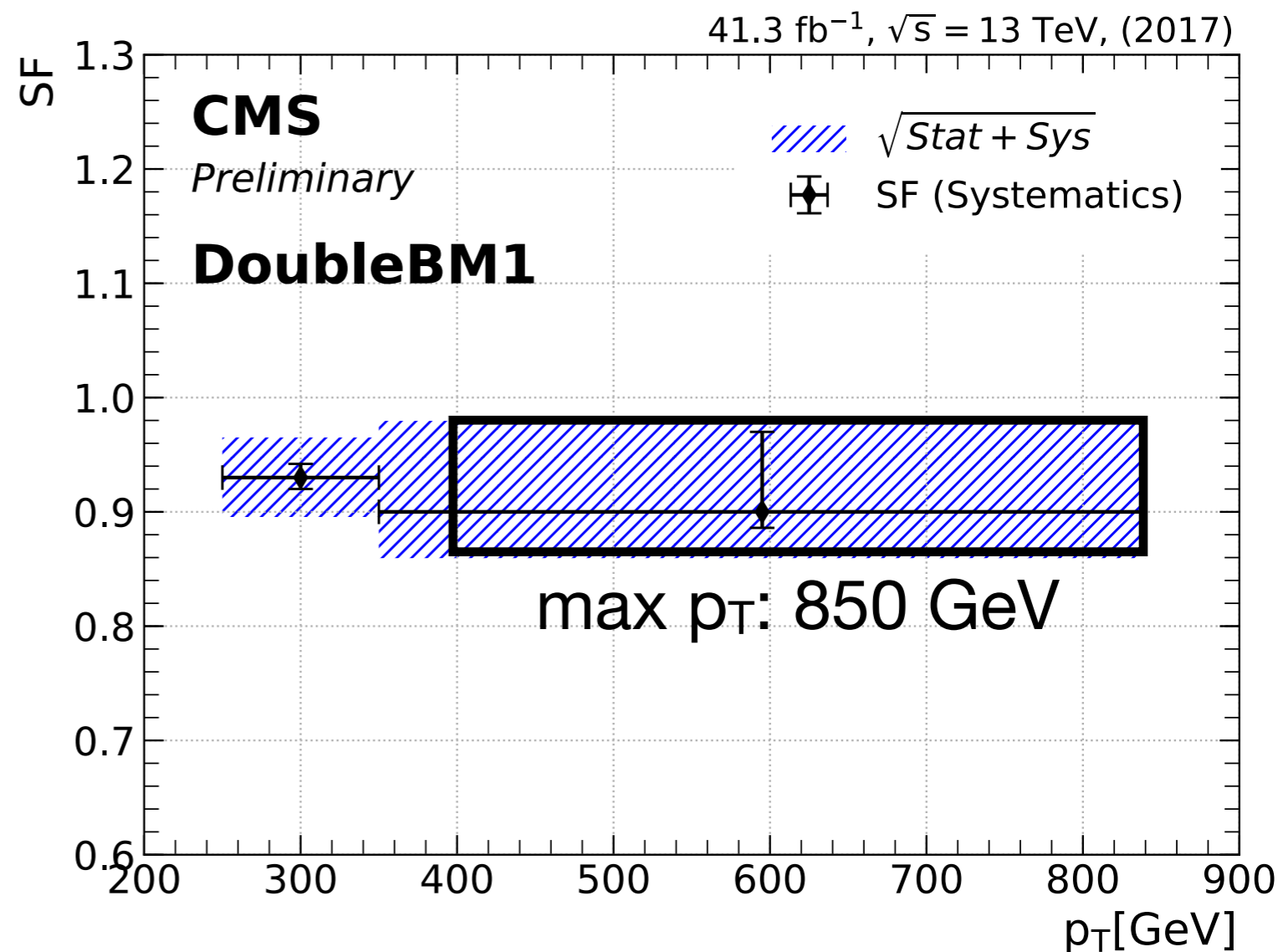
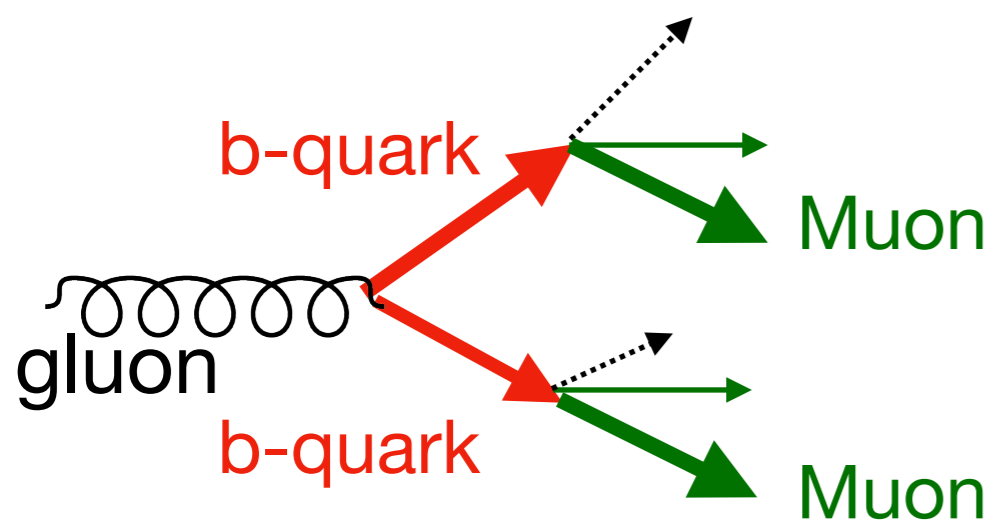
Well modeled w.r.t MC
Similar as for 2016 data.

Heavy flavour in data

Where to look: $H \rightarrow bb$ would be ideal but...
get $g \rightarrow bb$ instead

2017 Data!

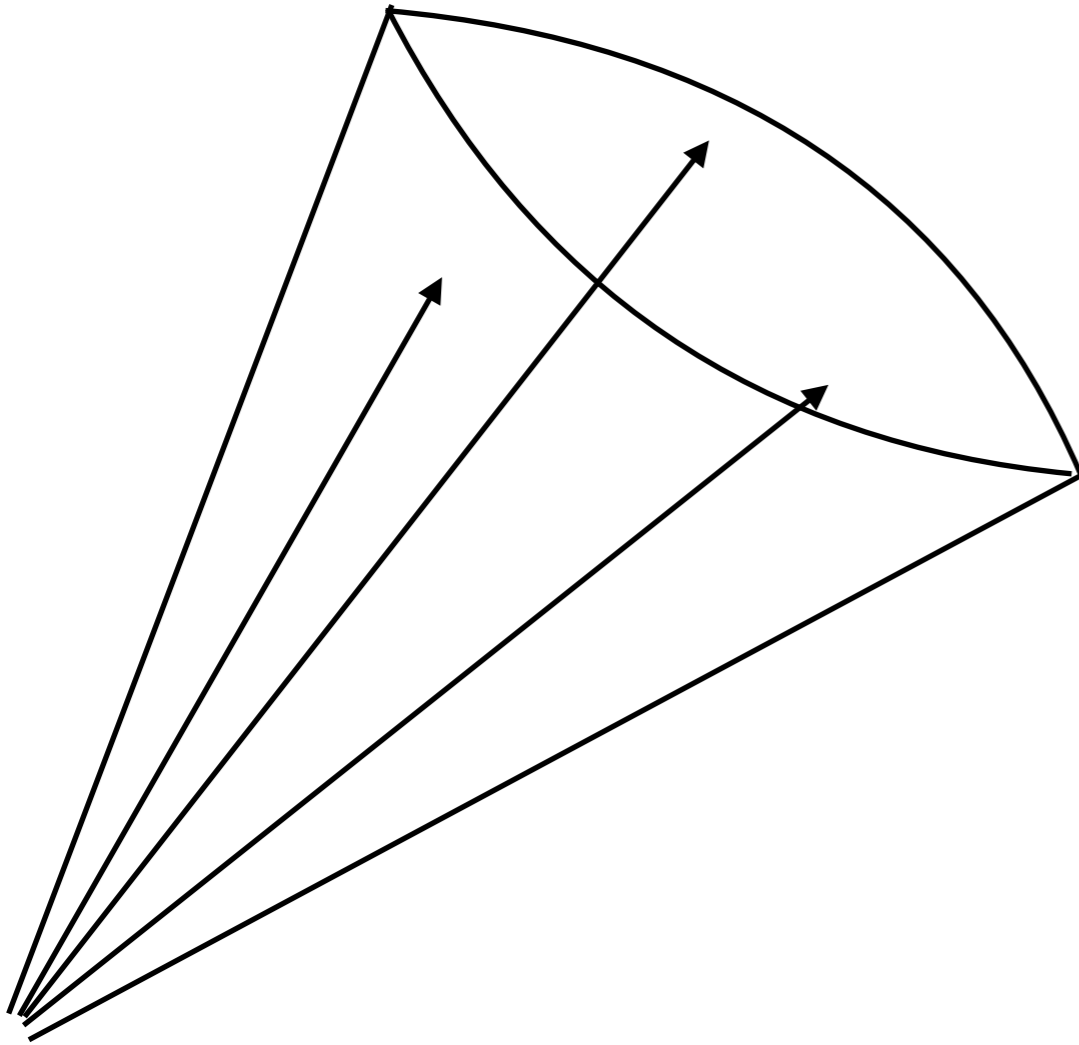
Tag 2 muons in a jet



SF close to unity

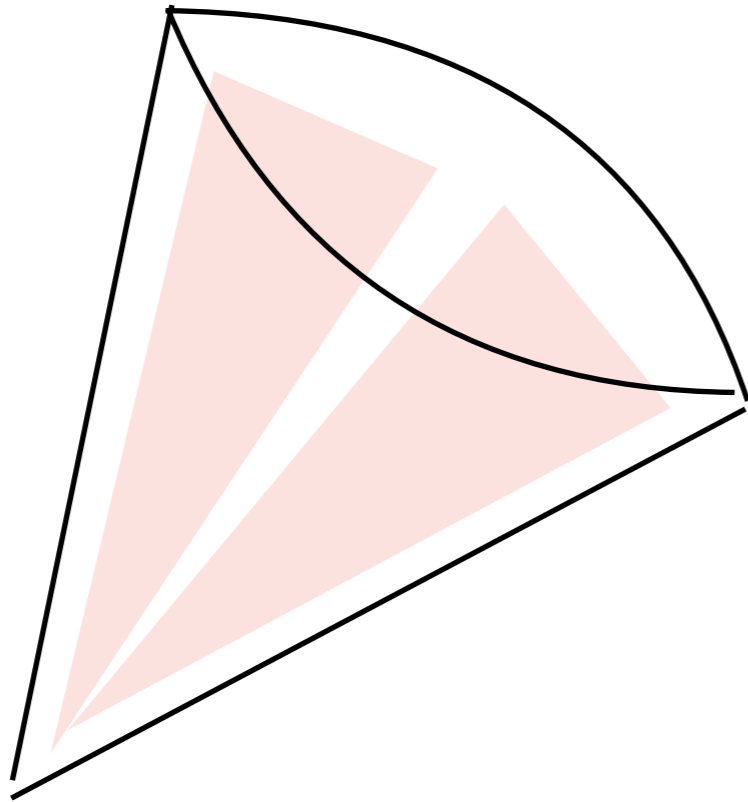
Associated data/MC uncertainty 3-5%

CMS DP-2018/033



Heavy resonances

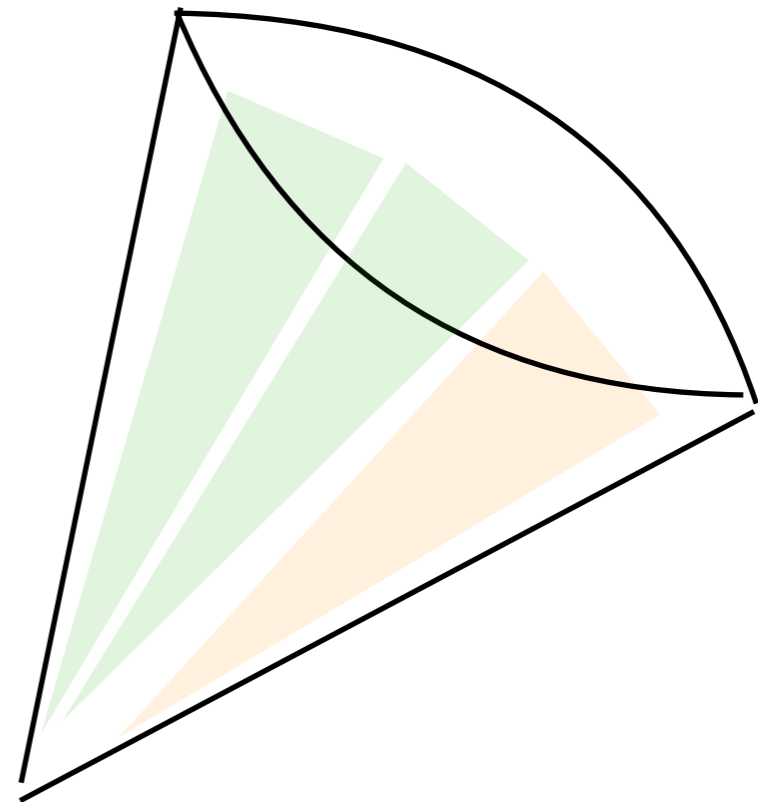
Heavy Resonances identification



Boosted:

$p_T > 200$ GeV for $V=W$, $R=0.8$

W/Z-tagging/ Higgs-tagging



Boosted:

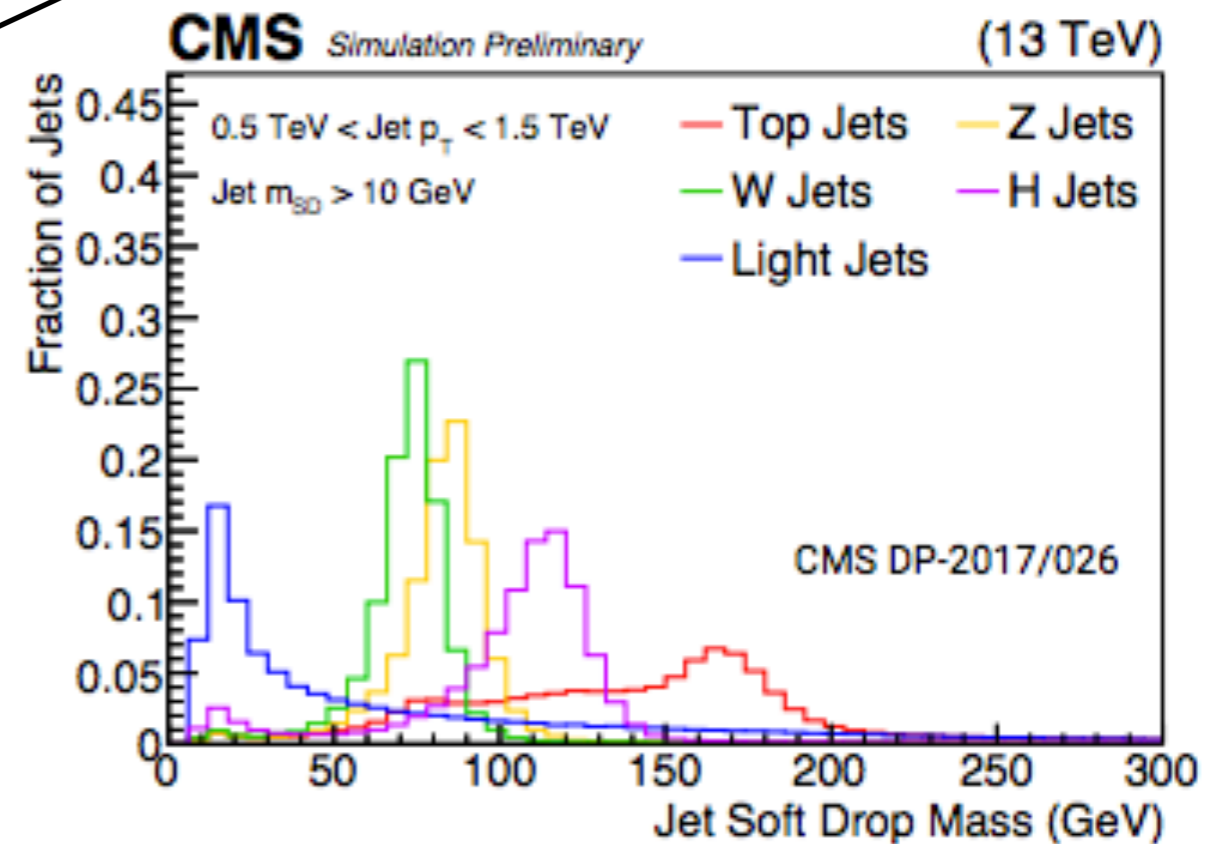
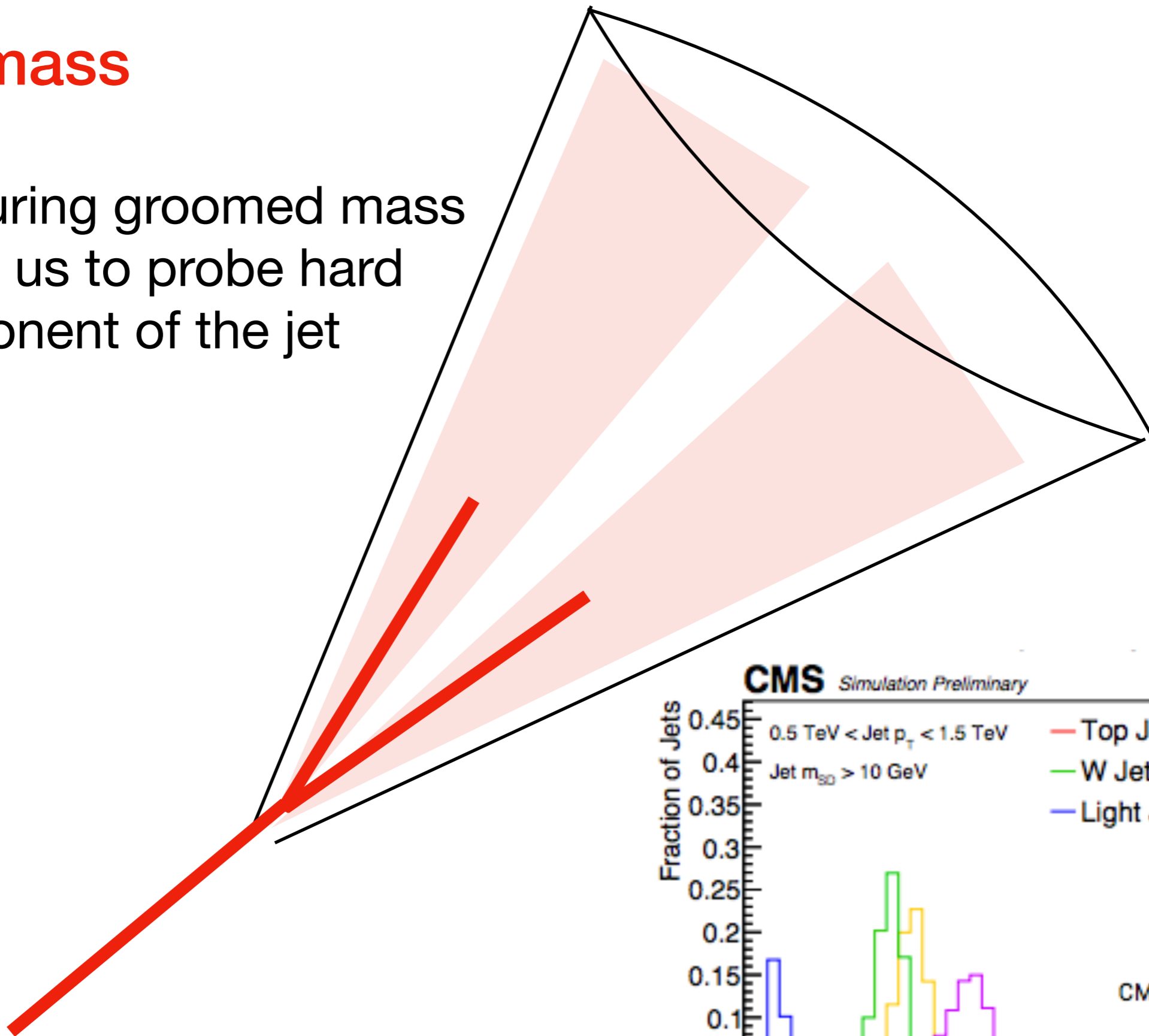
$p_T > 400$ GeV, $R=0.8, 1.5$

Top-tagging

- Jet mass
- Jet substructure
 - 2-prong ID/3-prong ID

Jet mass

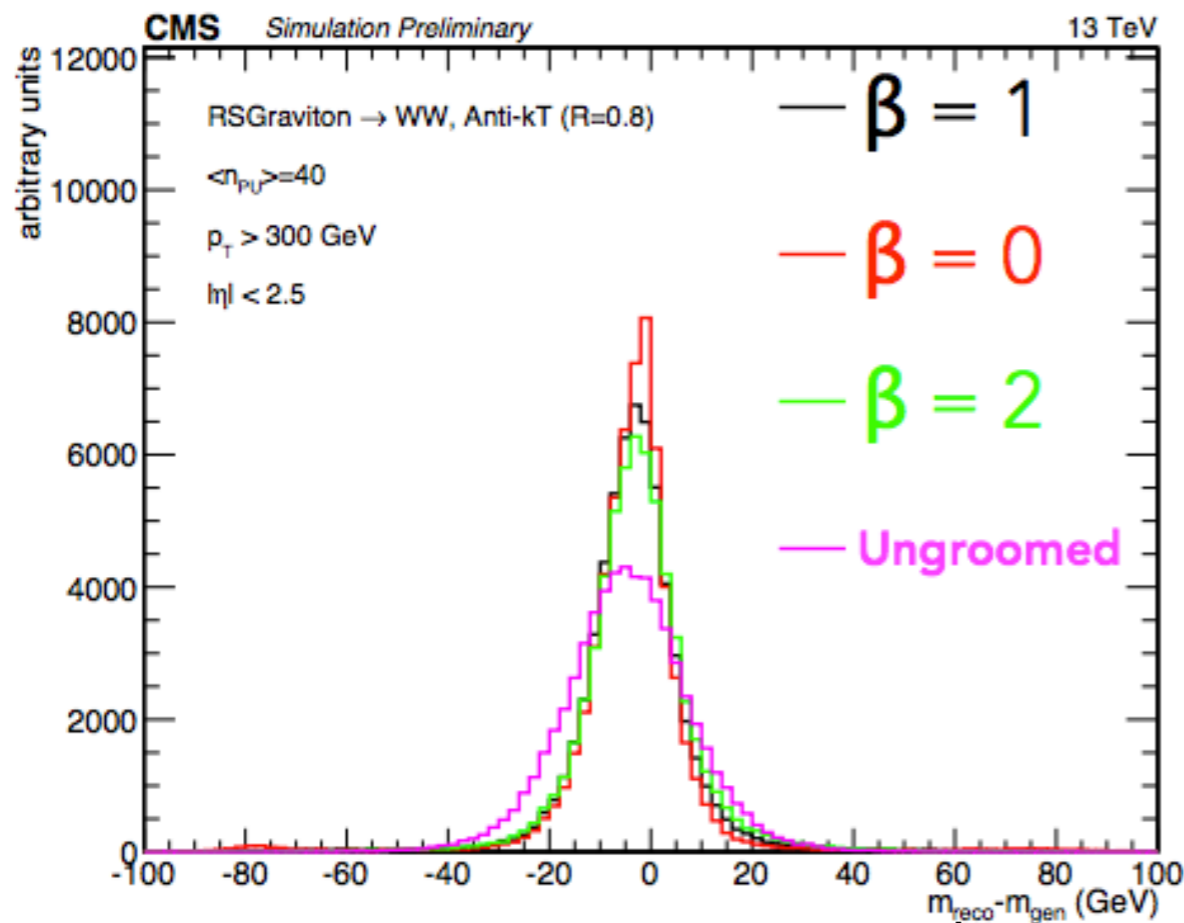
Measuring groomed mass allows us to probe hard component of the jet



Jet mass

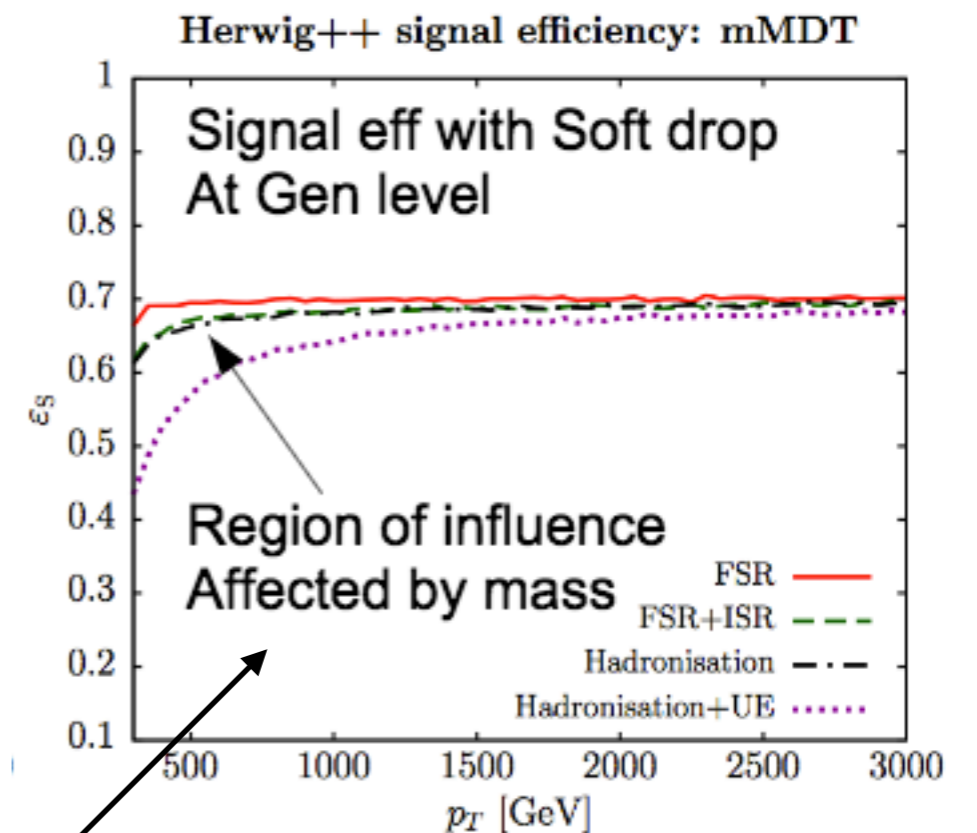
For CMS, default mMDT (soft drop $\beta=0$):

W tagging $\sim 65 \text{ GeV} < m_{SD} < 140 \text{ GeV}$
 t tagging $\sim 105 \text{ GeV} < m_{SD} < 210 \text{ GeV}$



extra unc. comes from variations at low pT (U.E. reduces signal eff.)

mMDT corrections measured in signal applied as function of pT and eta

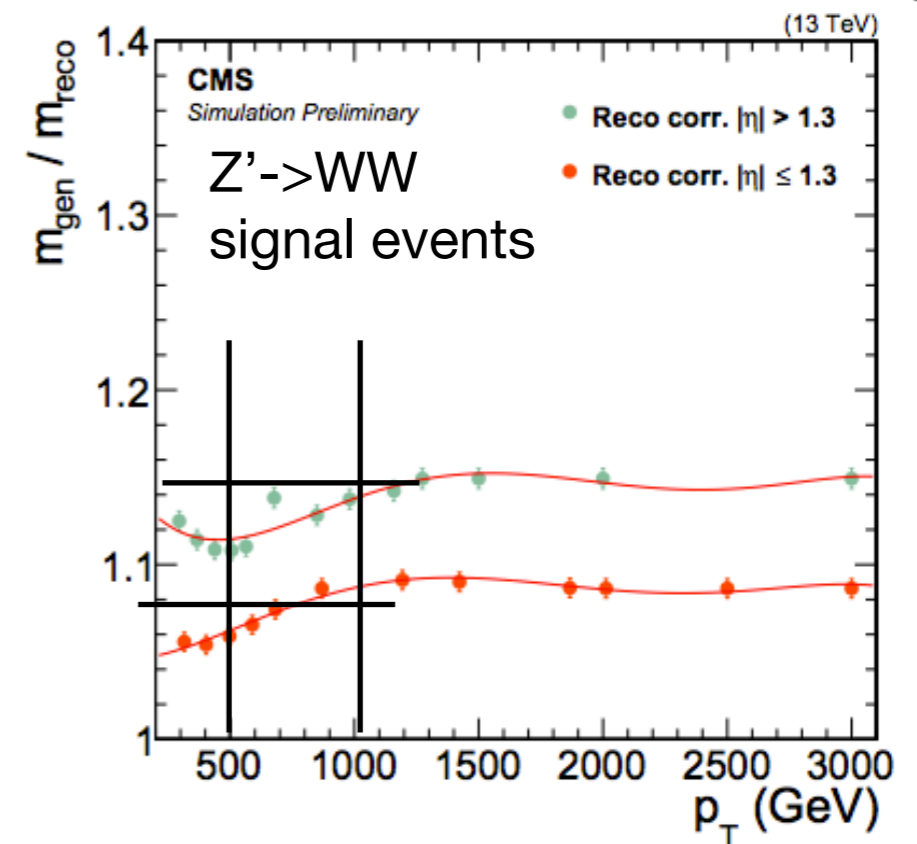
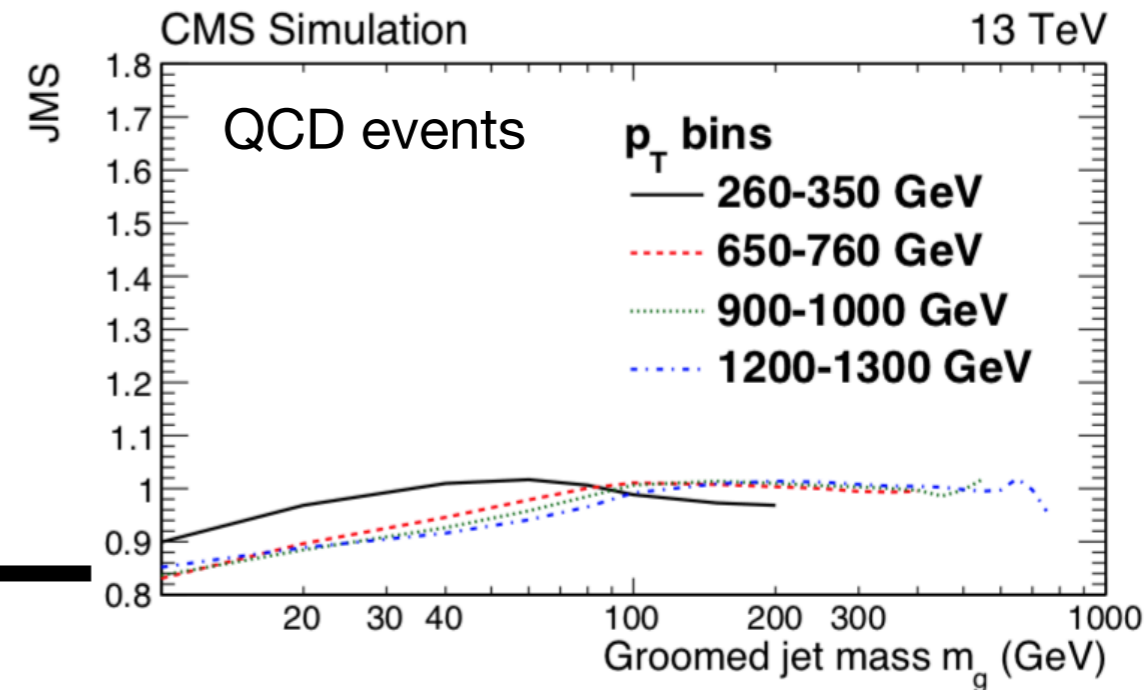
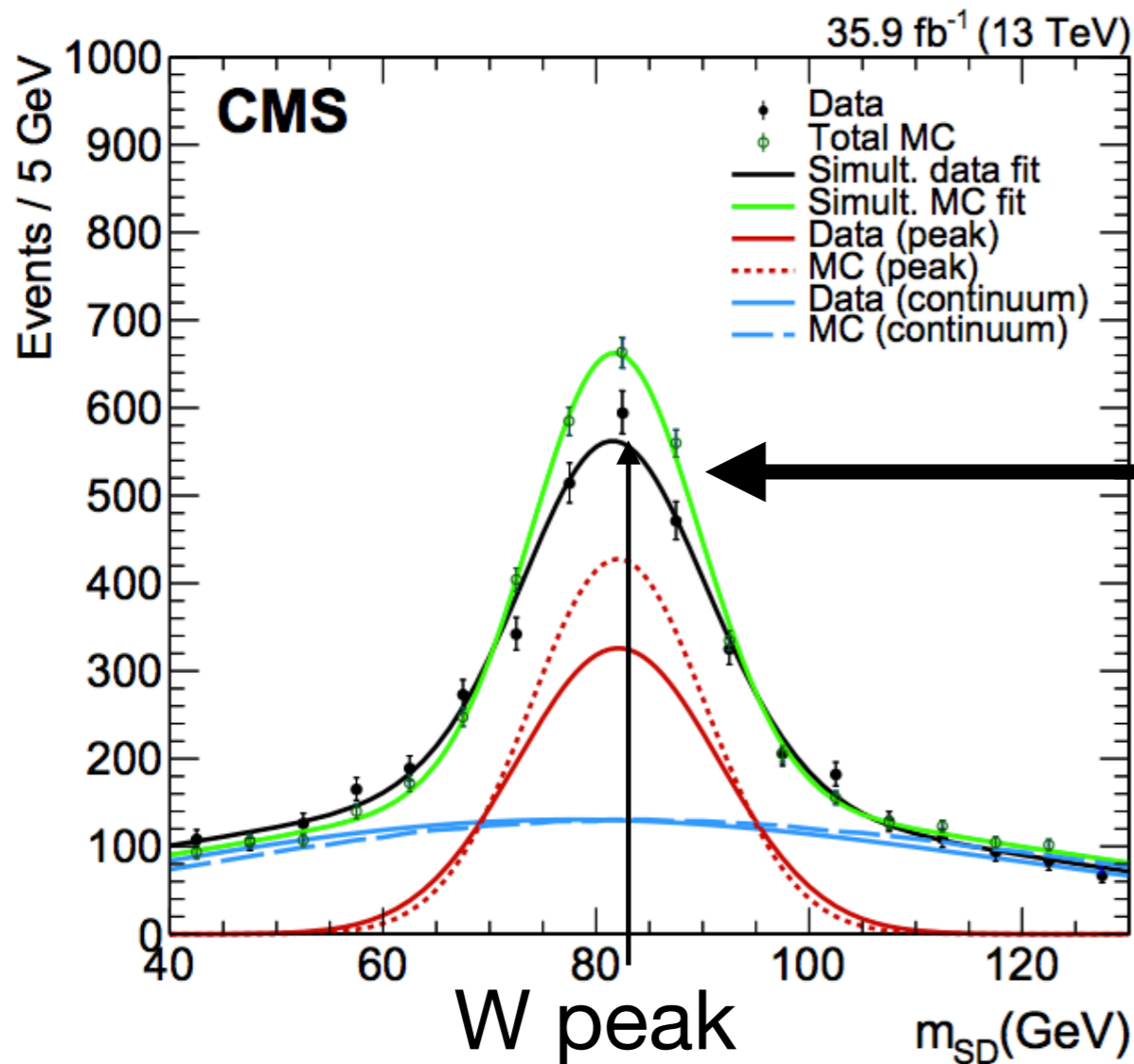


talk by A. Siodmok (BOOST 2015)

PUPPI + Softdrop keeps effect small

Jet mass scale

Jet mass scale and resolution obtained in **top events**



Unc.: stat. + JECs/JER, PU+ PDF and physics model.
 extrapolated to all phase space

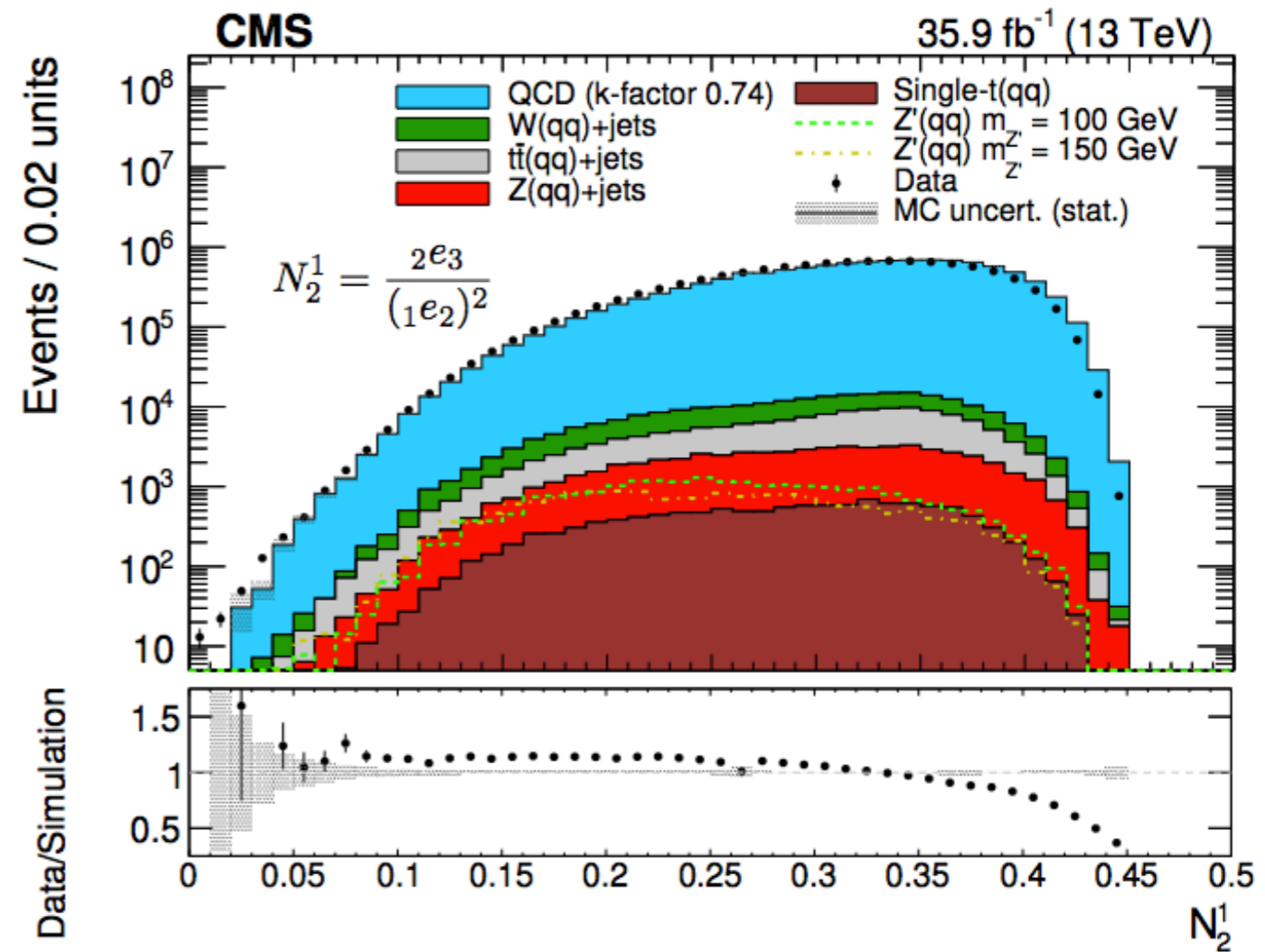
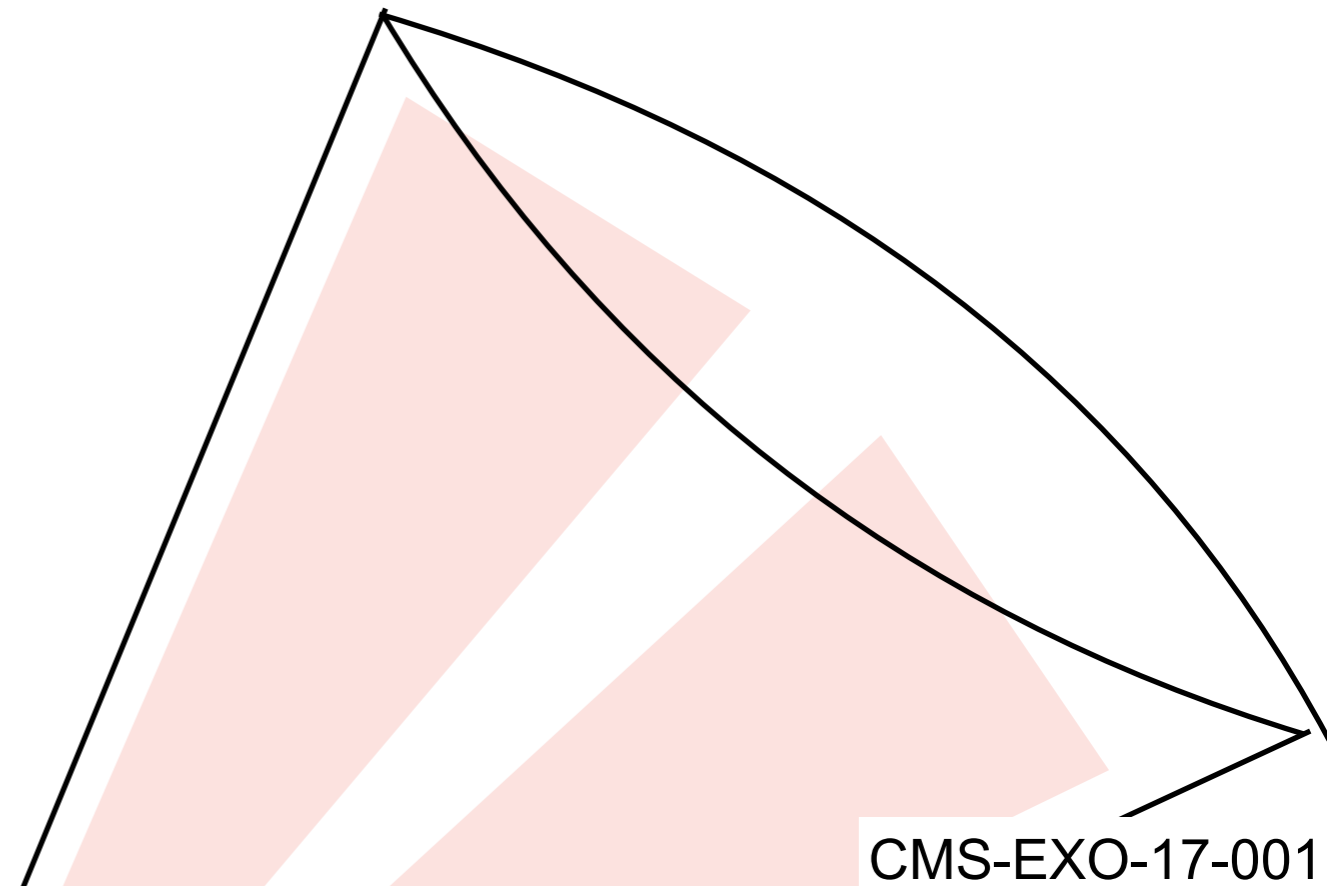
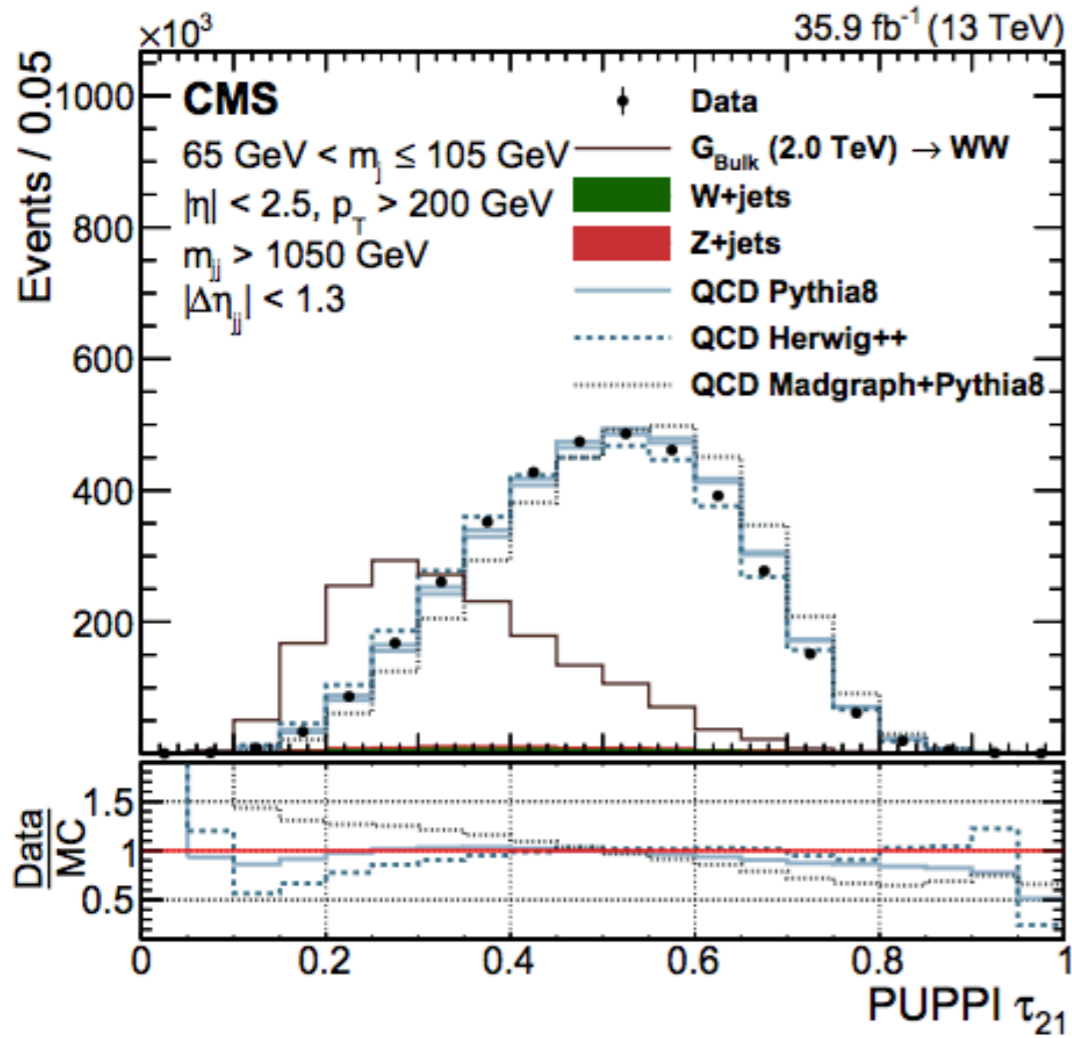
Mass shift dependent on MC
 modelling, not clear trend w.r.t p_T:

2% variation in mass scale over 500 GeV

Jet substructure

- 2 prong ID

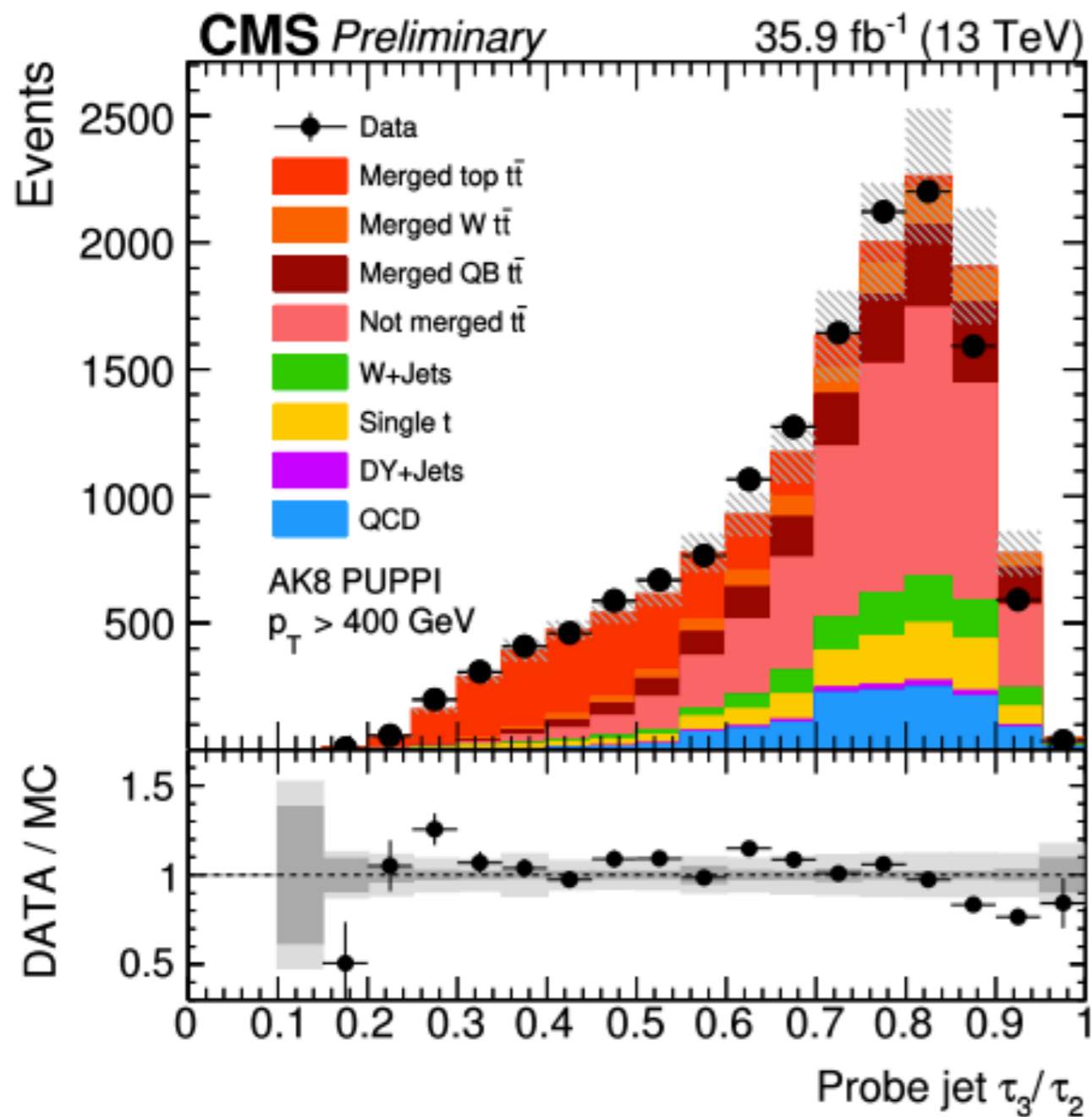
CMS-B2G-17-001



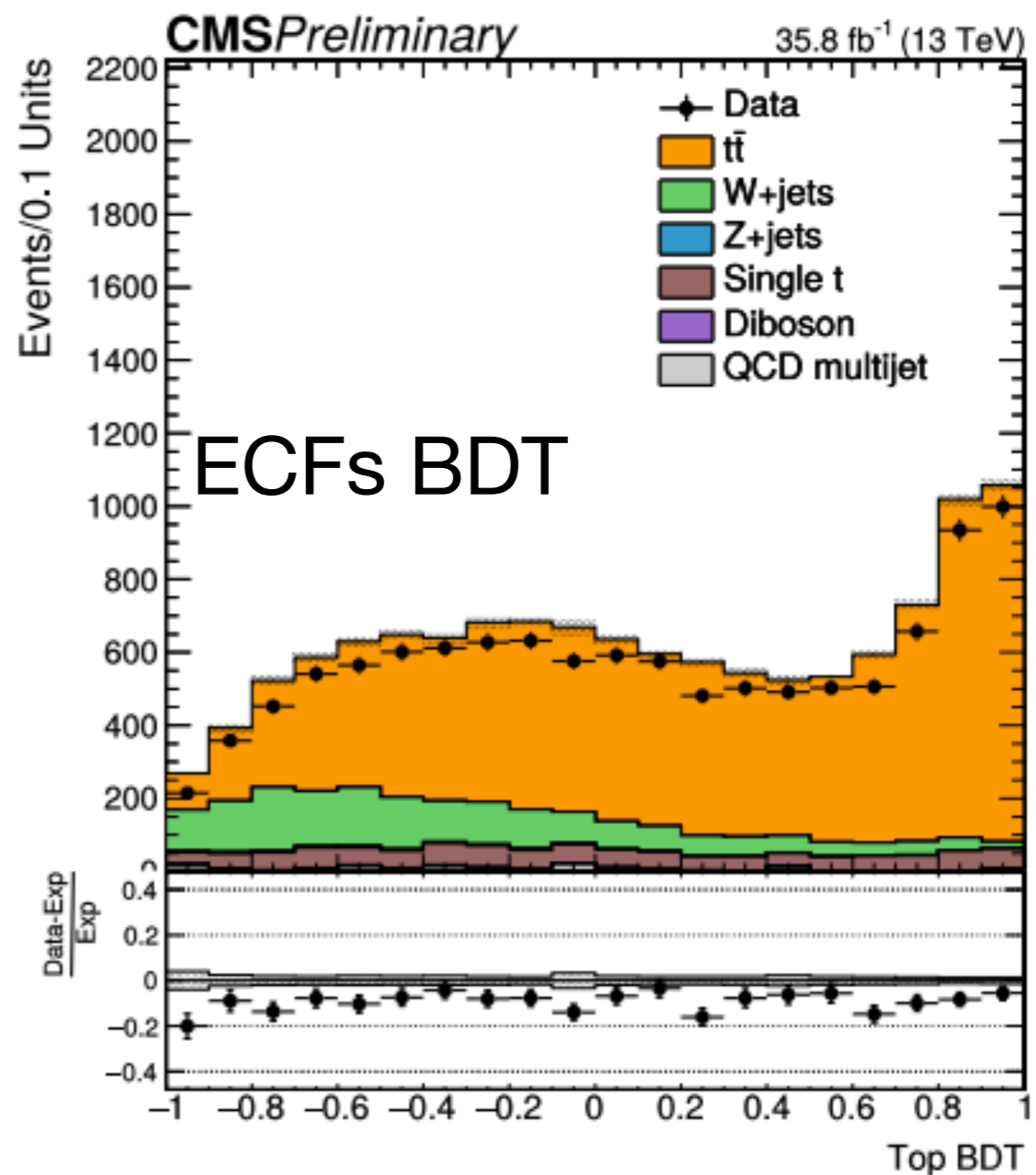
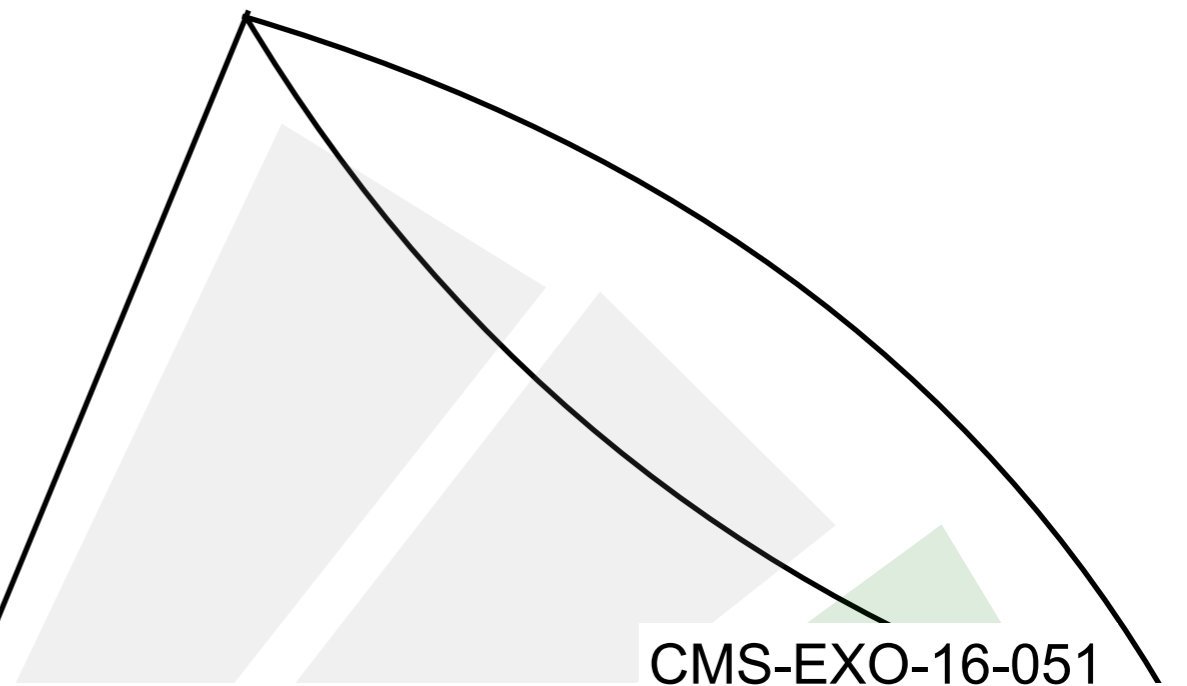
Jet substructure

- 3 prong ID

CMS-DPS-2017-026

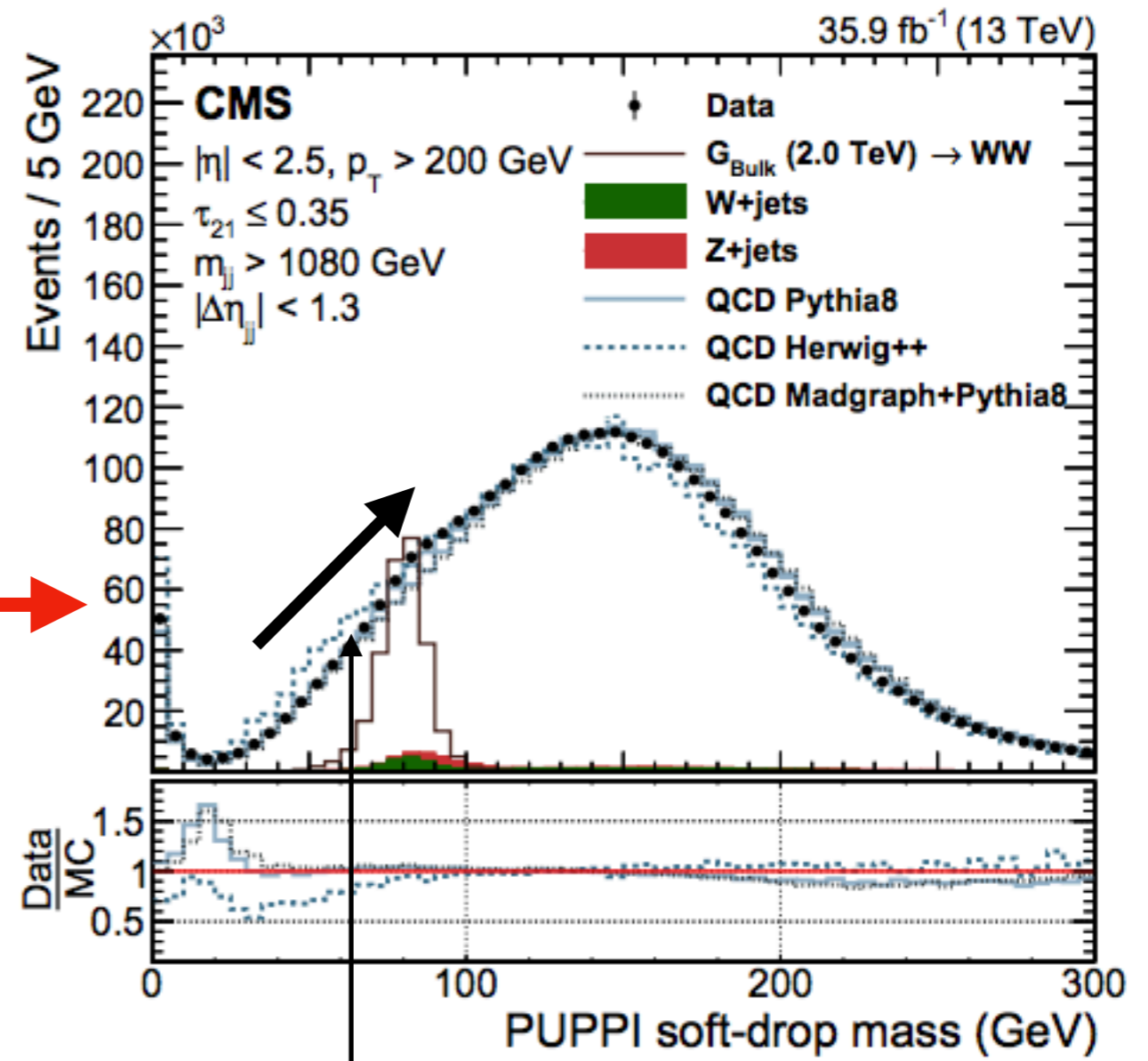
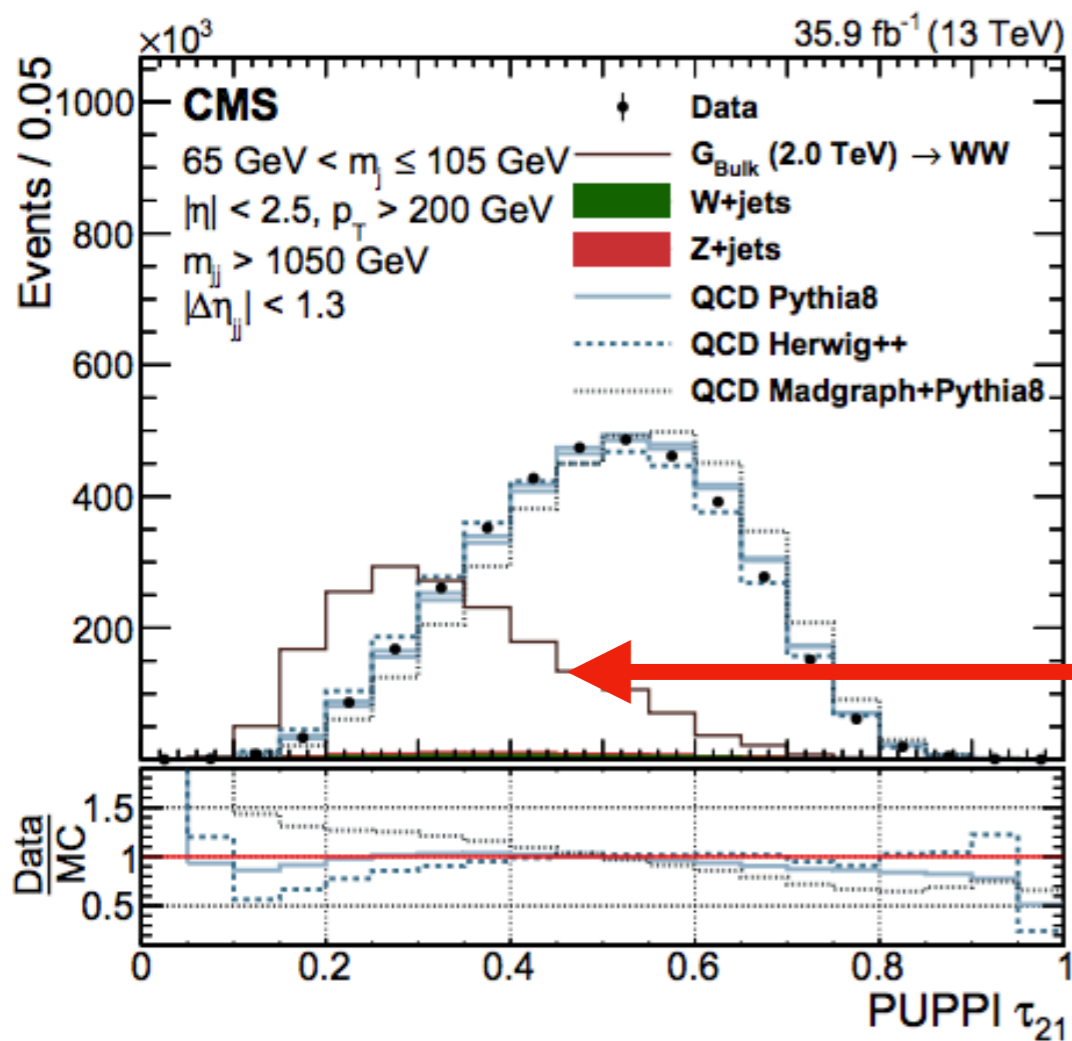


- ▶ Both τ_{32} and ECFs BDT \sim mass invariant



A note on: mass decorrelation

CMS-B2G-17-001



► How does this affect the search for a resonance

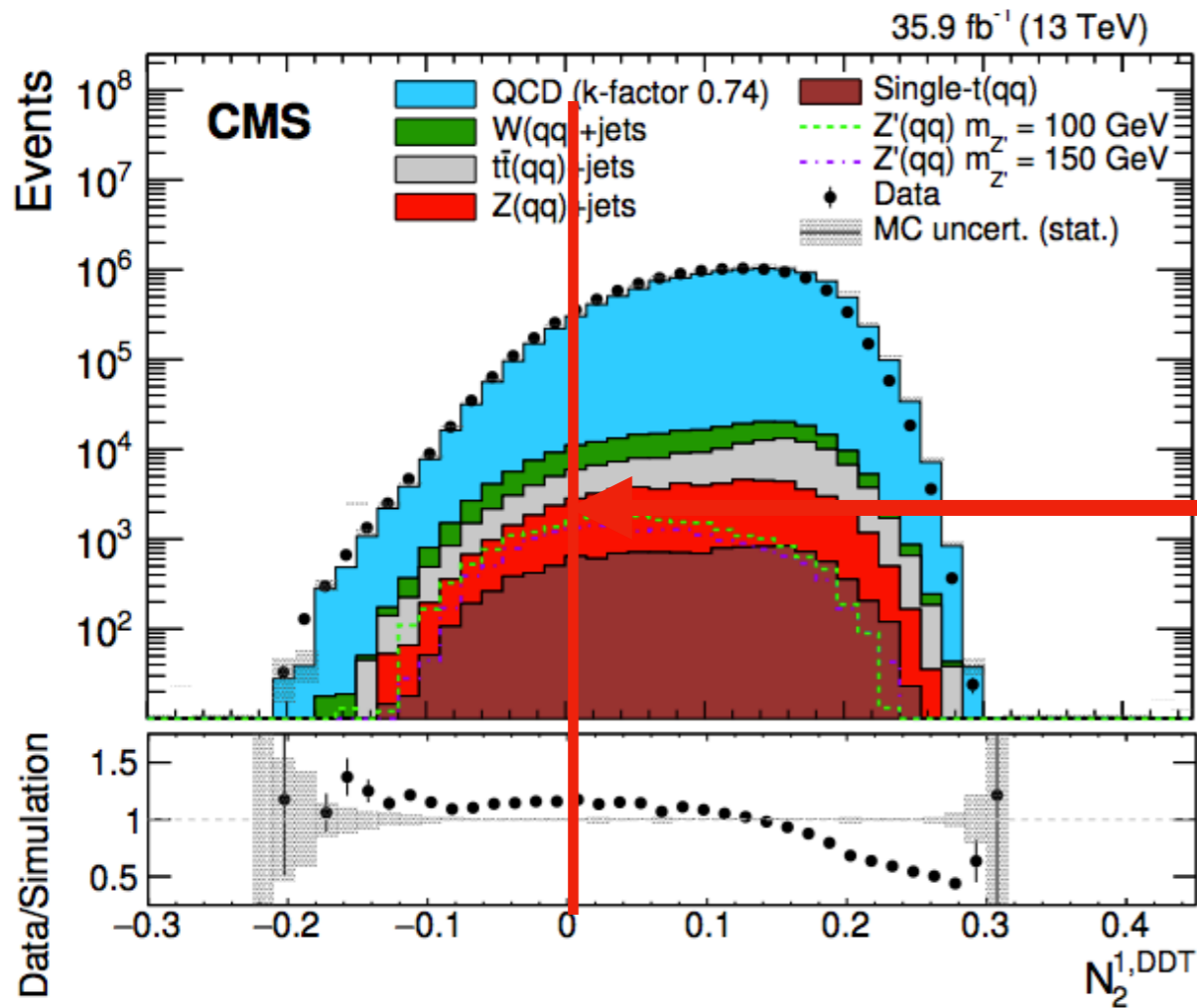
A note on: mass decorrelation

- Enter the DDT:

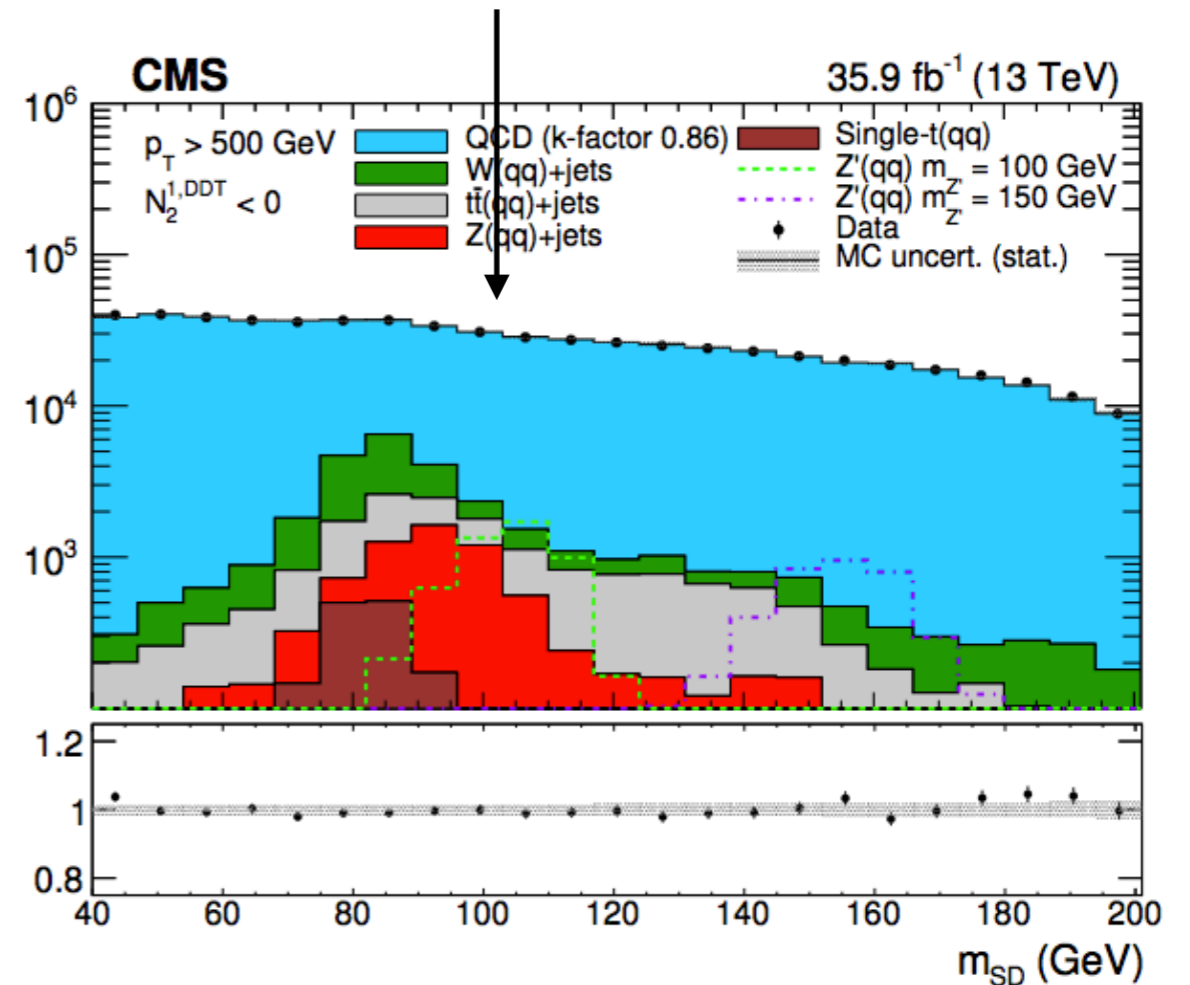
$$N_2^{\text{DDT}} = N_2 - N_2 (\epsilon_{\text{QCD}})$$

AK8

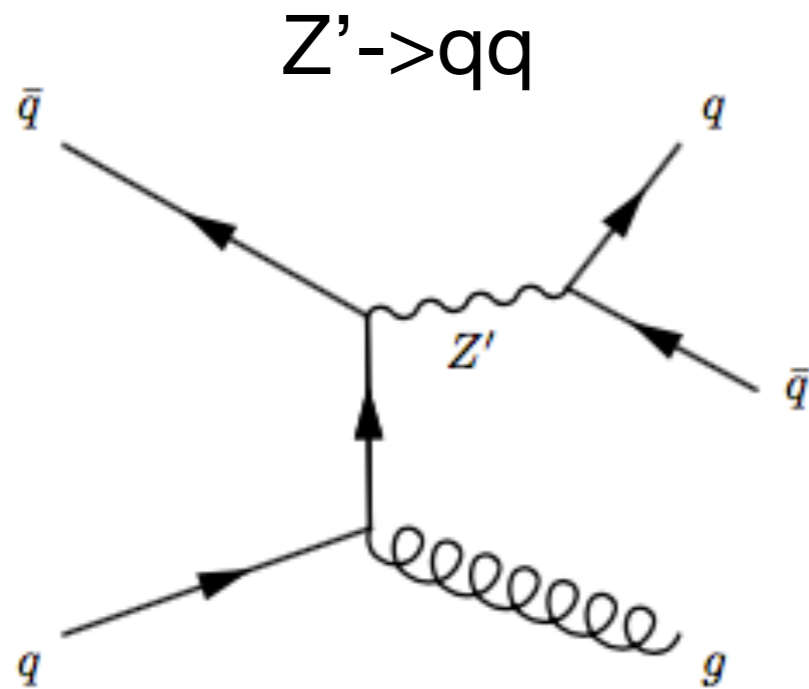
minimize the variation
across jet mass and pT



Events / 7 GeV



Inclusive channel and **W/Z peak**



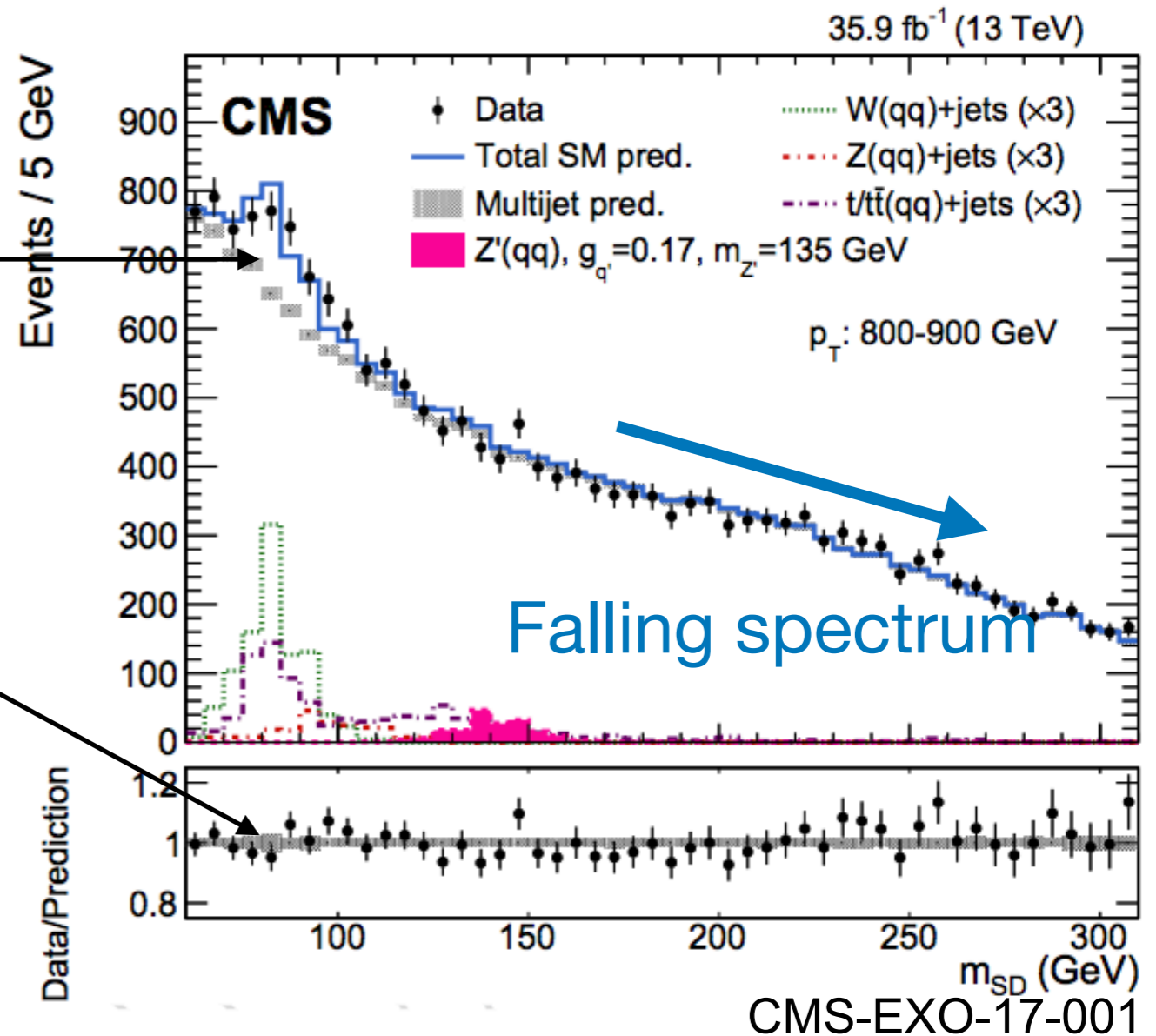
- N_2
- + Puppi
- + m_{SD}
- + DDT



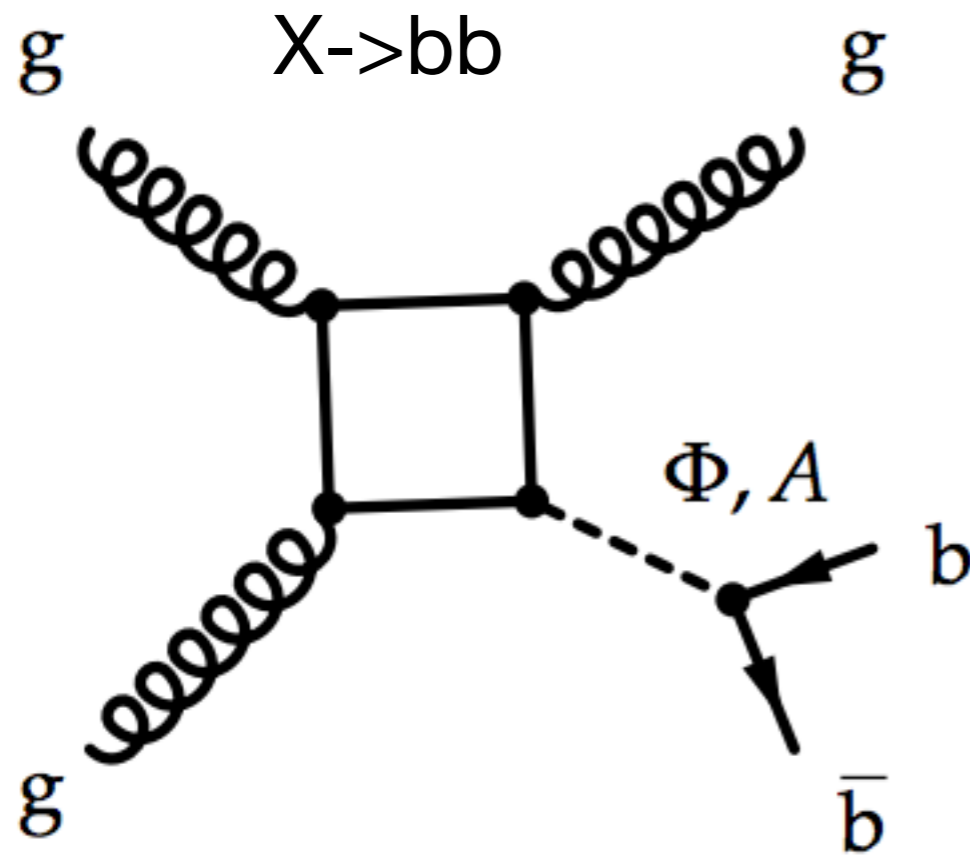
W/Z standard candles

Precision to 7% level

Unc: signal efficiency and misID calculations



Inclusive channel and **H peak**

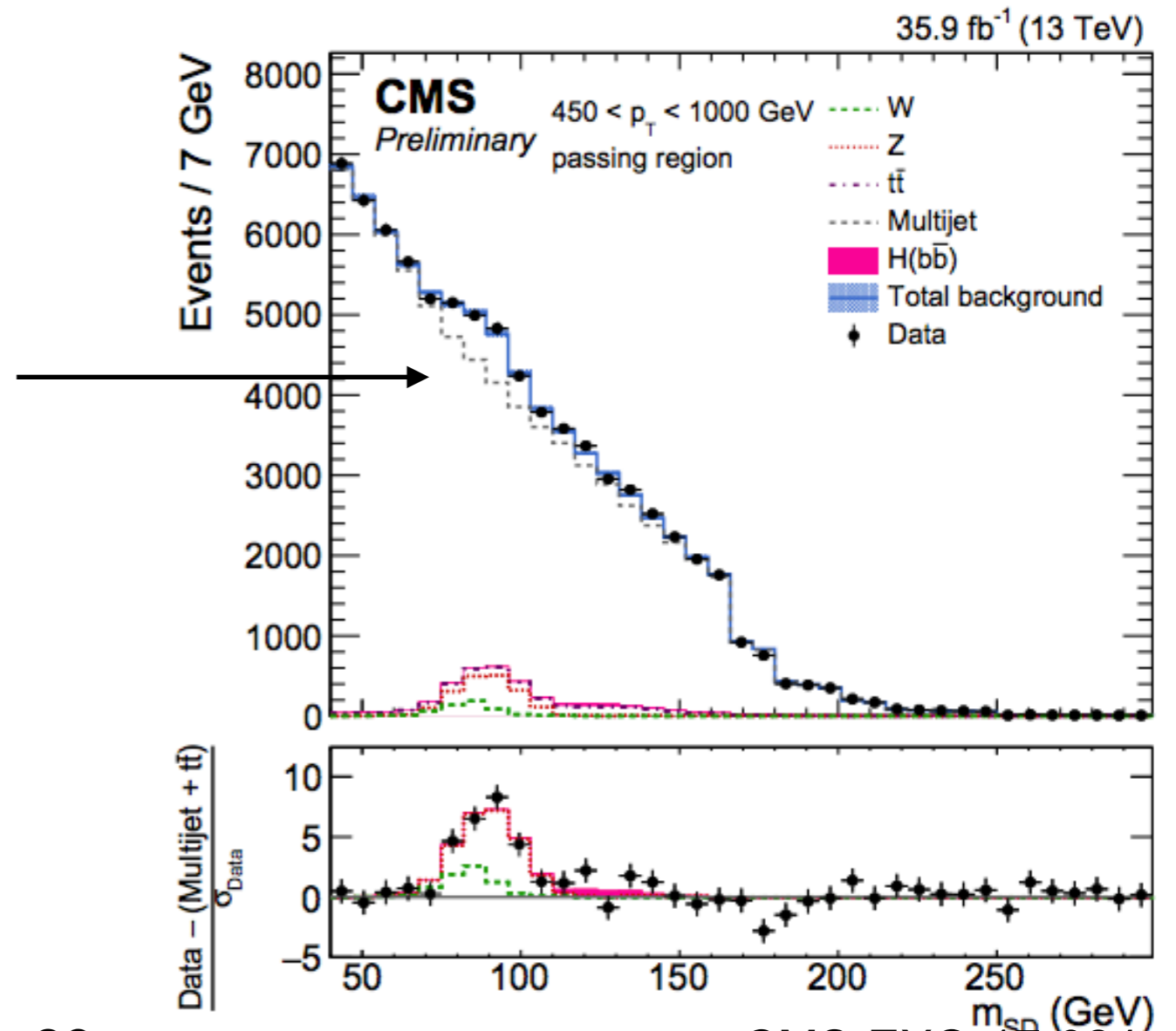


- + N_2
- + Puppi
- + m_{SD}
- + DDT
- + b-tagging



Z(bb) for calibration

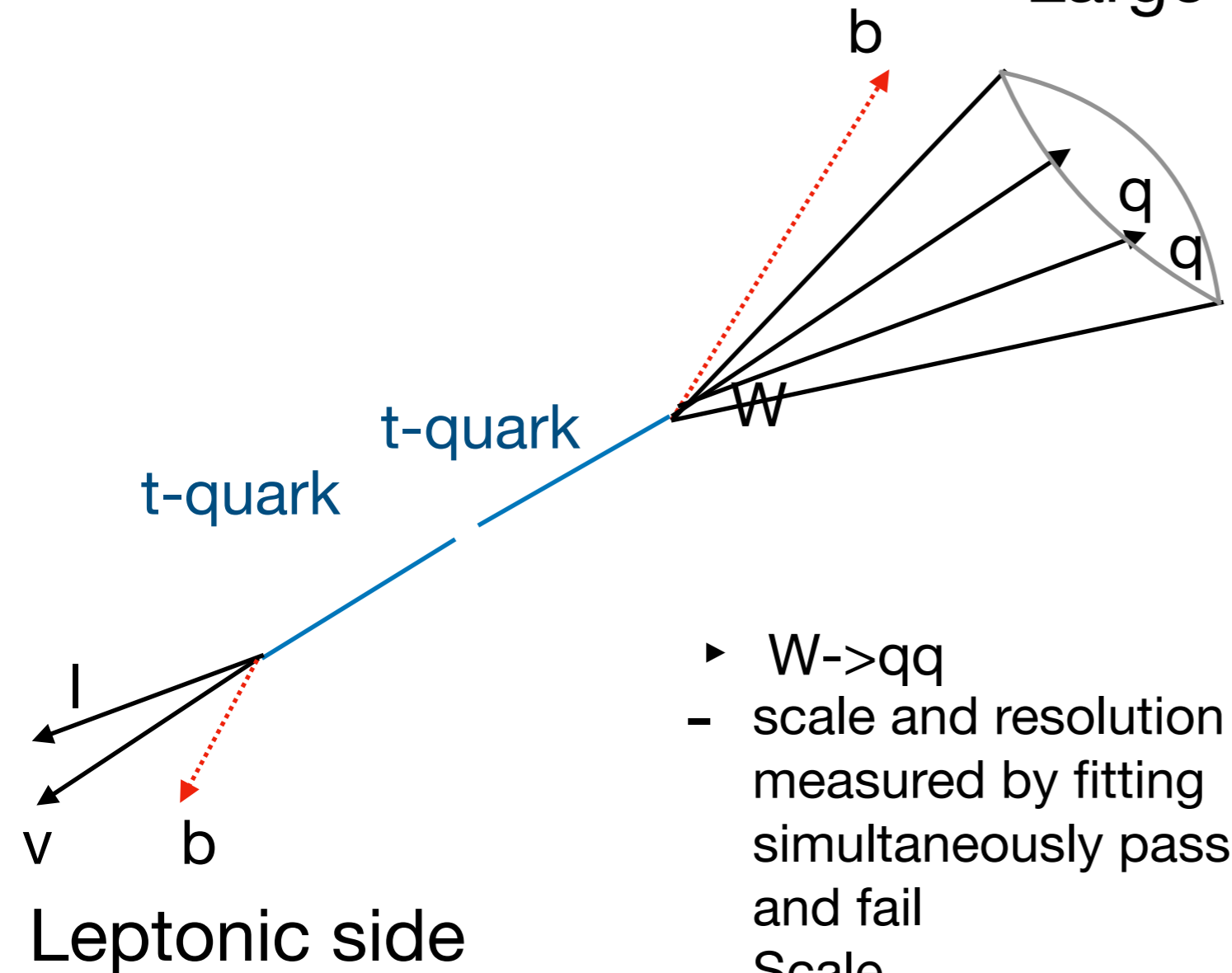
At which lumi (fb^{-1}) will be able use $H(bb)$ instead?



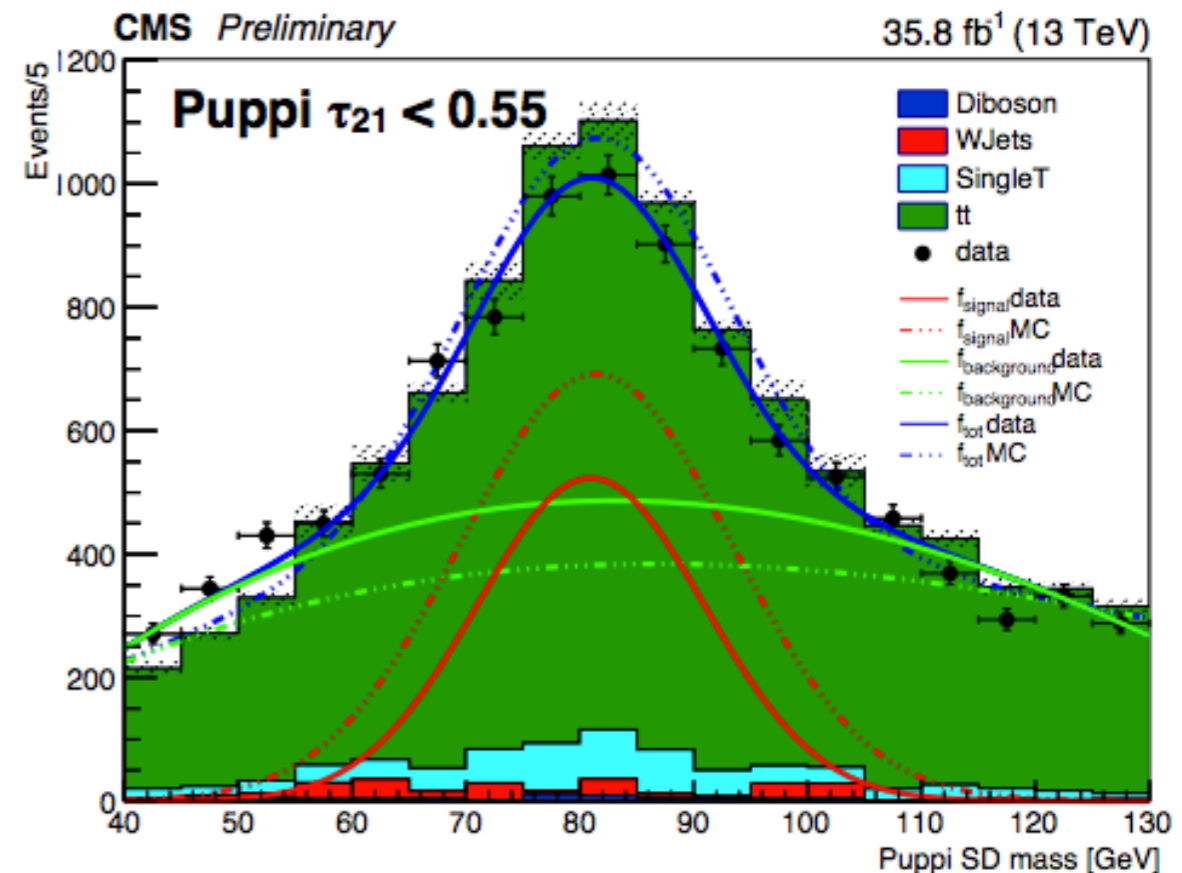
Heavy Resonances in data:

Semi-leptonic $t\bar{t}$ topology:

Large R-jet: $W(qq)$, $t \rightarrow bqq$



- Leading uncertainties
 - $t\bar{t}$ modelling (Powheg+pythia)
 - Which pythia tune to use?
 - JEC, JER, Pu, Muon SFs,
 - PDF and scale

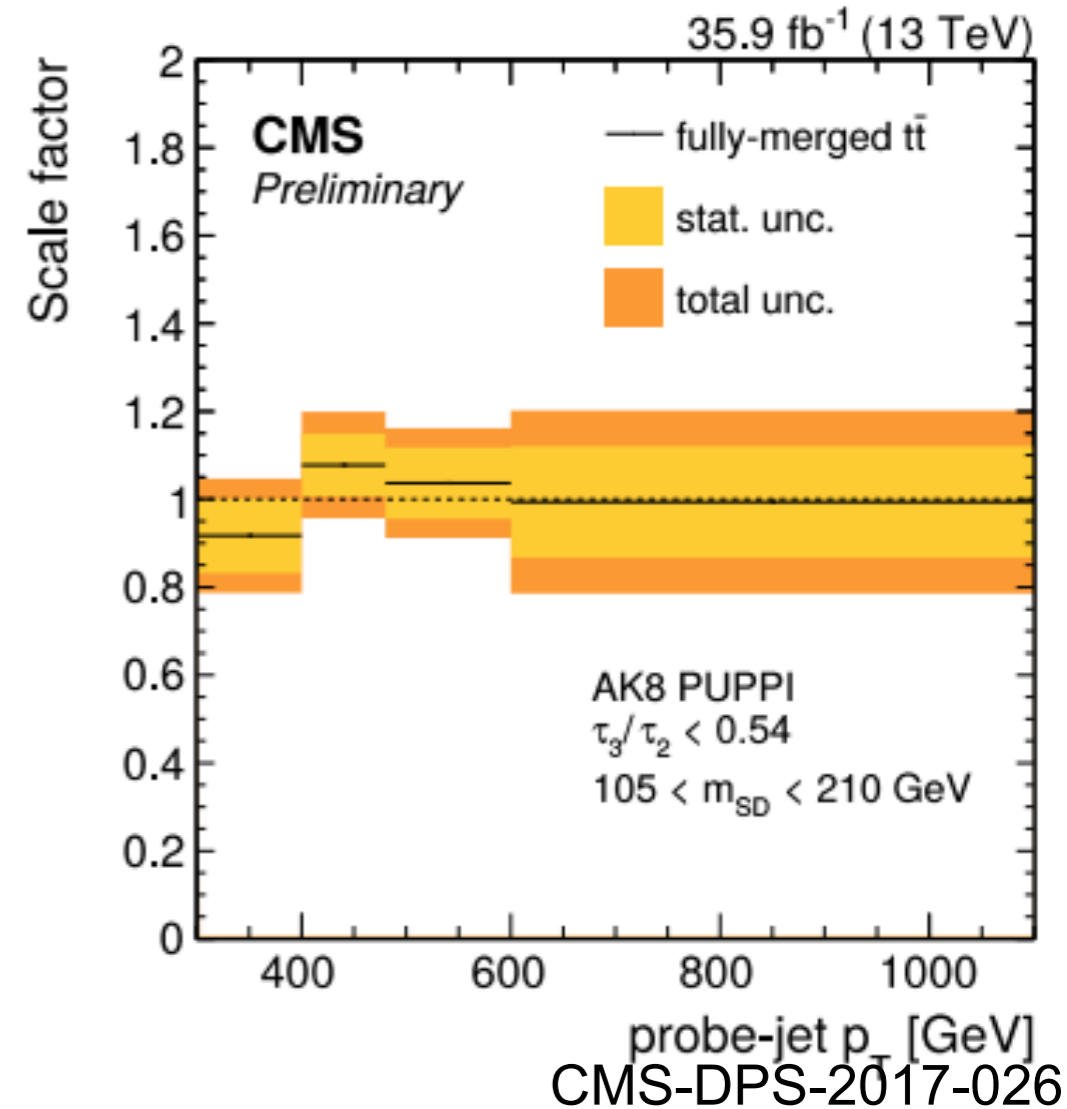
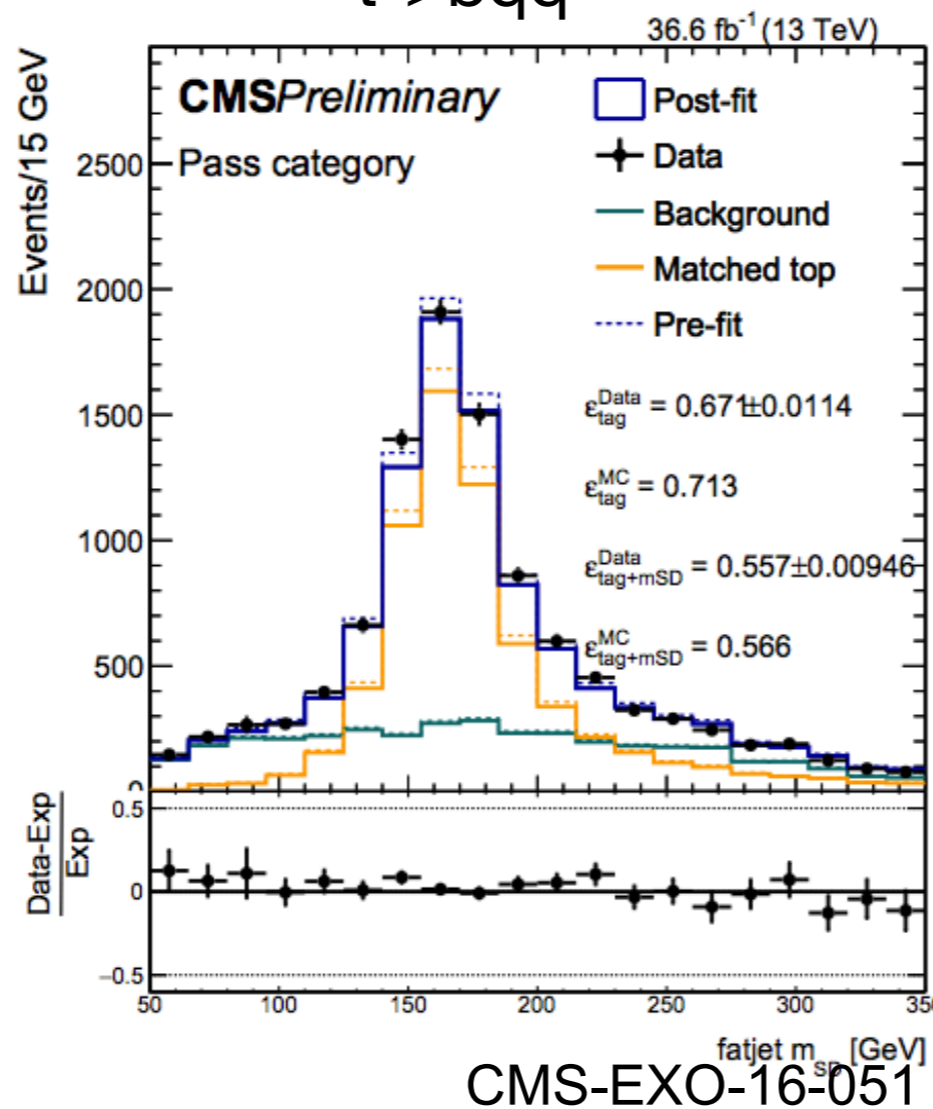
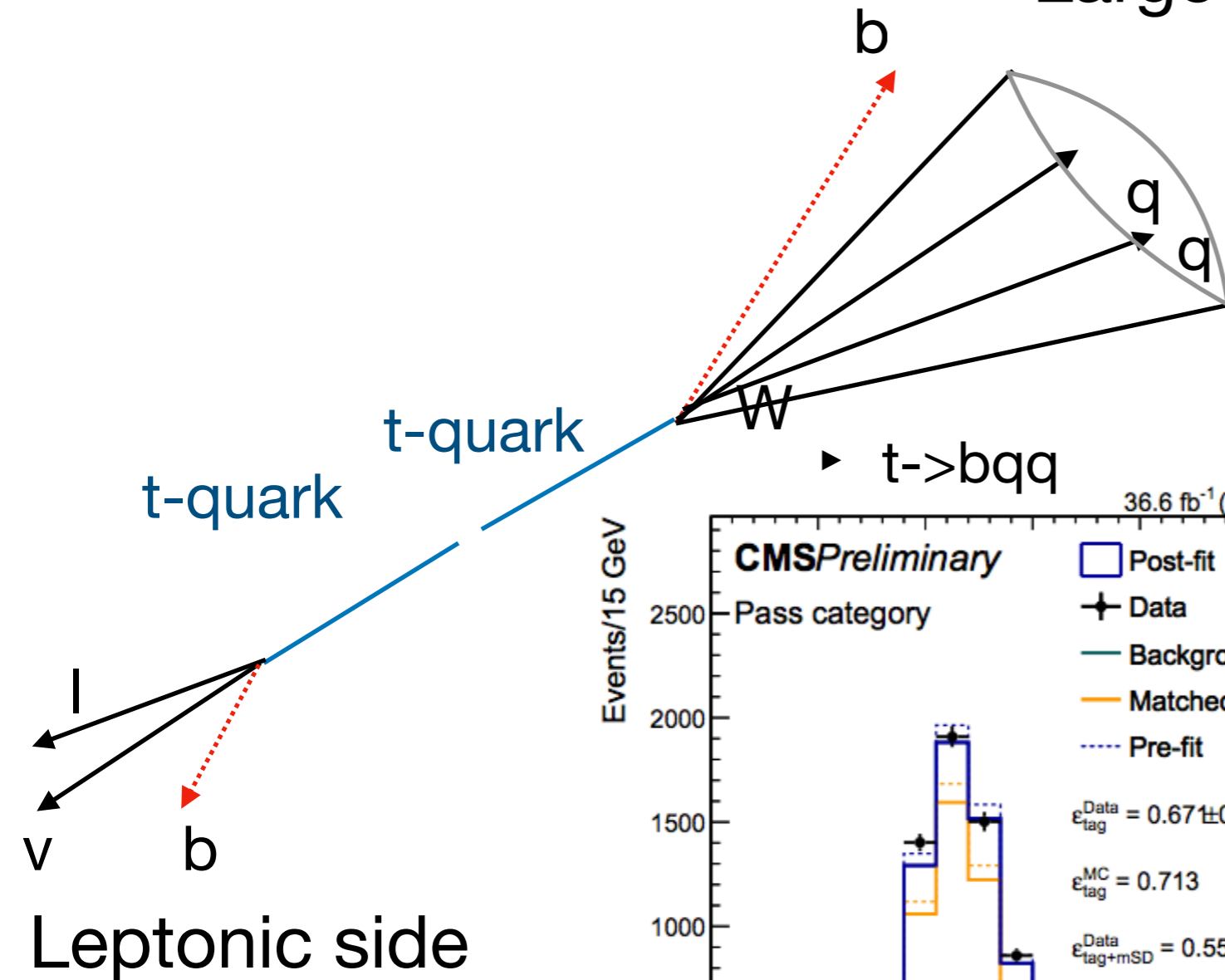


Heavy Resonances in data:

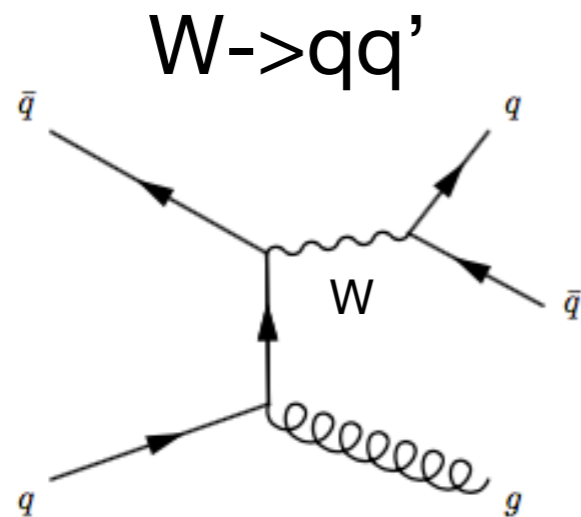
Semi-leptonic $t\bar{t}$ topology:

Large R-jet: $W(qq)$, $t \rightarrow bqq$

- Leading uncertainties
 - $t\bar{t}$ modelling (Powheg+pythia)
 - Which pythia tune to use?
 - JEC, JER, Pu, Muon SFs,
 - PDF and scale



Inclusive channel and **W/Z peak**



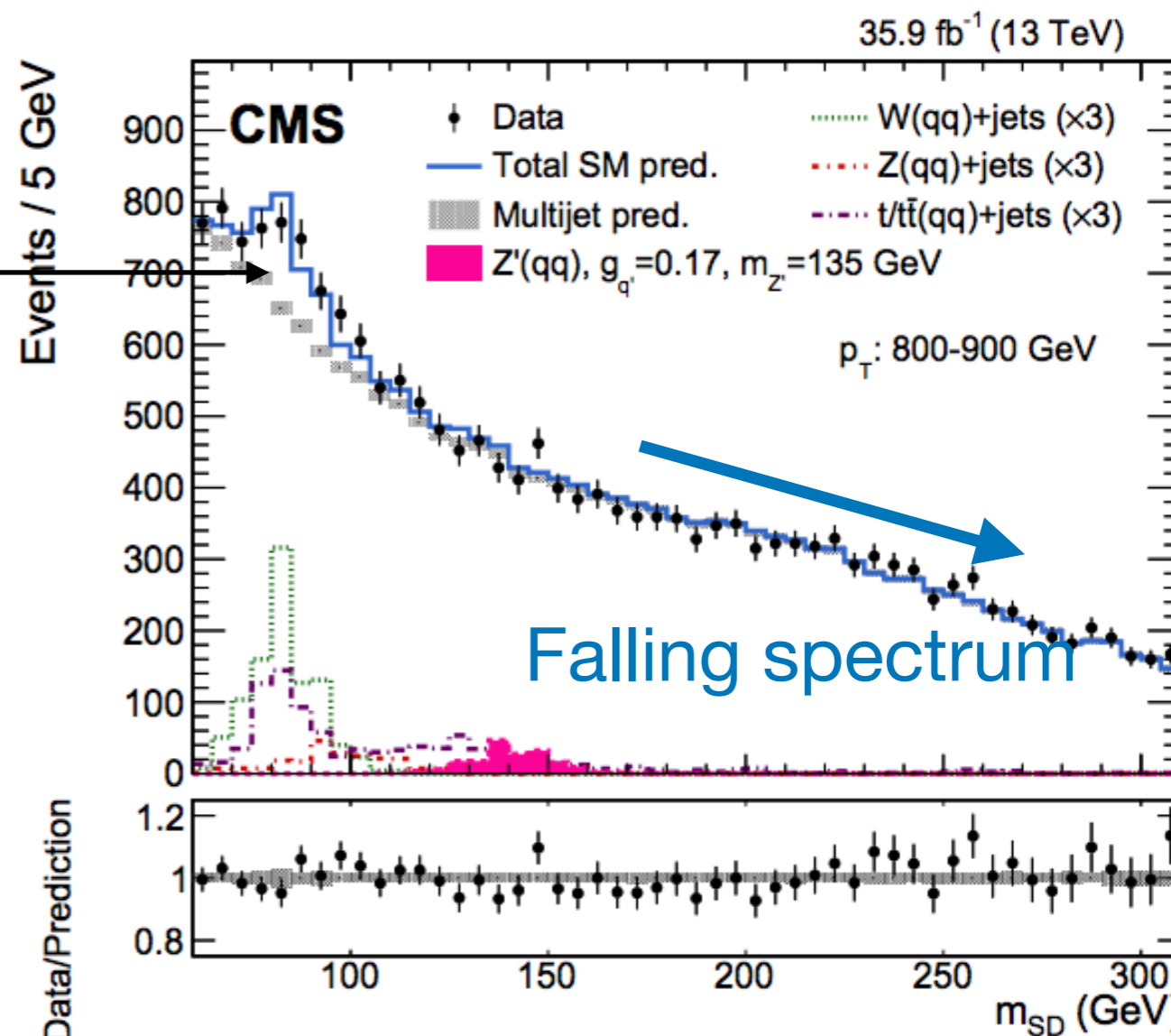
- + N_2
- + Puppi
- + m_{SD}
- + DDT



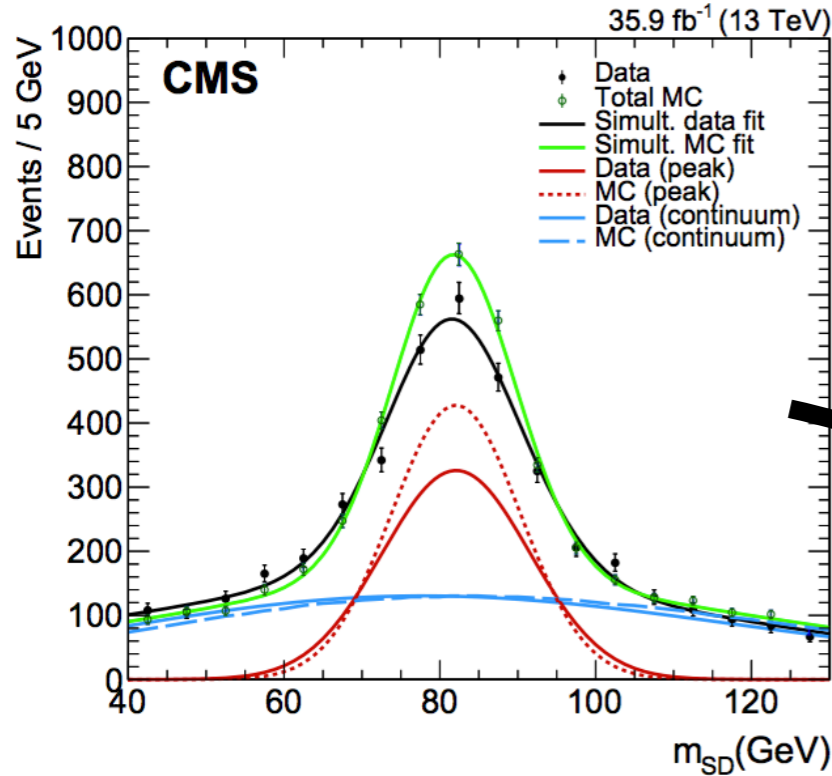
~~W/Z standard candles~~

Use **Z** as standard candle instead

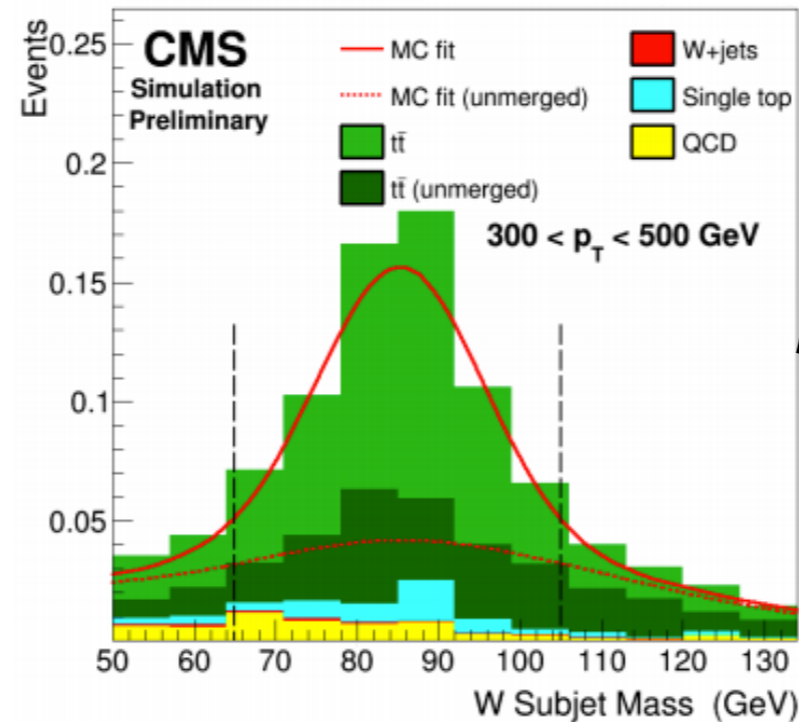
Can 3000 fb^{-1} be enough to measure W mass?
See Andreas's talk!



Inclusive channel and **W-tagging efficiency**

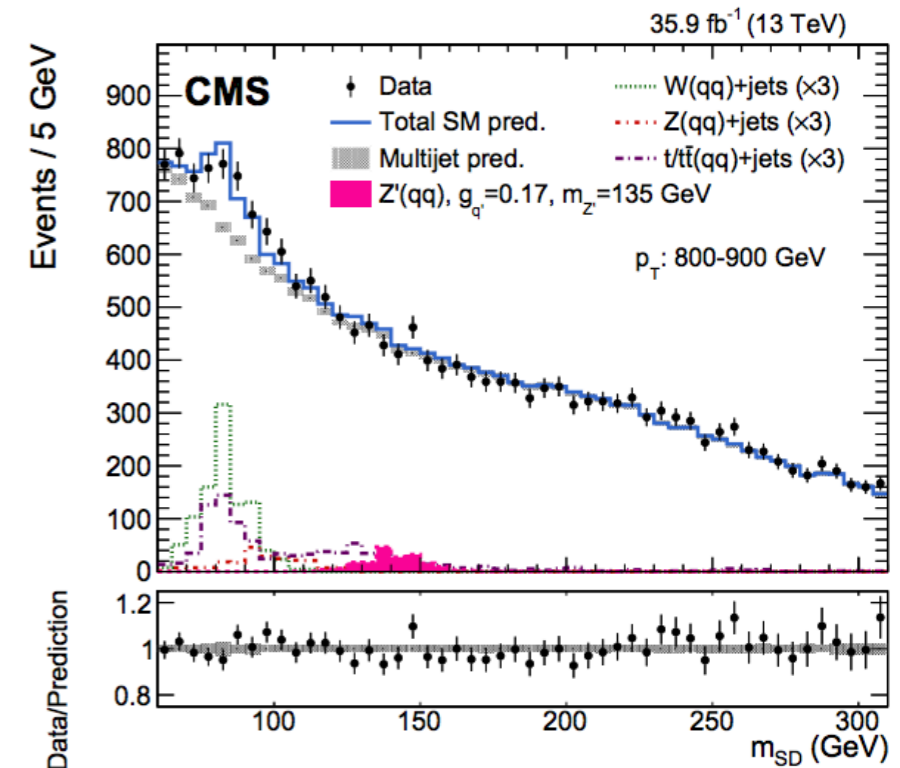


- Top events: p_T : 200-500 GeV
- 7% unc.



- Merged Top events:
- 4% unc. (p_T : 300-500 GeV)
- 14% unc. ($p_T > 500$ GeV)

- Inclusive channel
 - in-situ measurement
 - Efficiency gets constrained by the fit
 - 5% unc. (p_T : 500-1000 GeV)
 - Can we constrain W jet mass scale + resolution in-situ?



Summary

- “Shallow” approach of tagging still strong:
 - b-tagging moving fast to the DNN lane
 - questions for the future:
 - tracking at high p_T , how to treat unc.
 - vertexing at HL-LHC PU conditions
 - W/Z/H/t-tagging can still rely on
 - jet mass+ jet substructure + (b-tagging)
 - expect to check DNN data performance soon!
- Main takeaway from SF measurements:
 - Many methods/topologies
 - rely on control regions (can run low stat at high p_T)
 - try in-situ efficiency measurements in data?
 - measure W mass, scale, at high p_T ?