There is something about… gluon splitting

Luca Perrozzi (ETH Zurich)

CMS Heavy flavour tagging workshop

Bruxelles, April 10th 2018

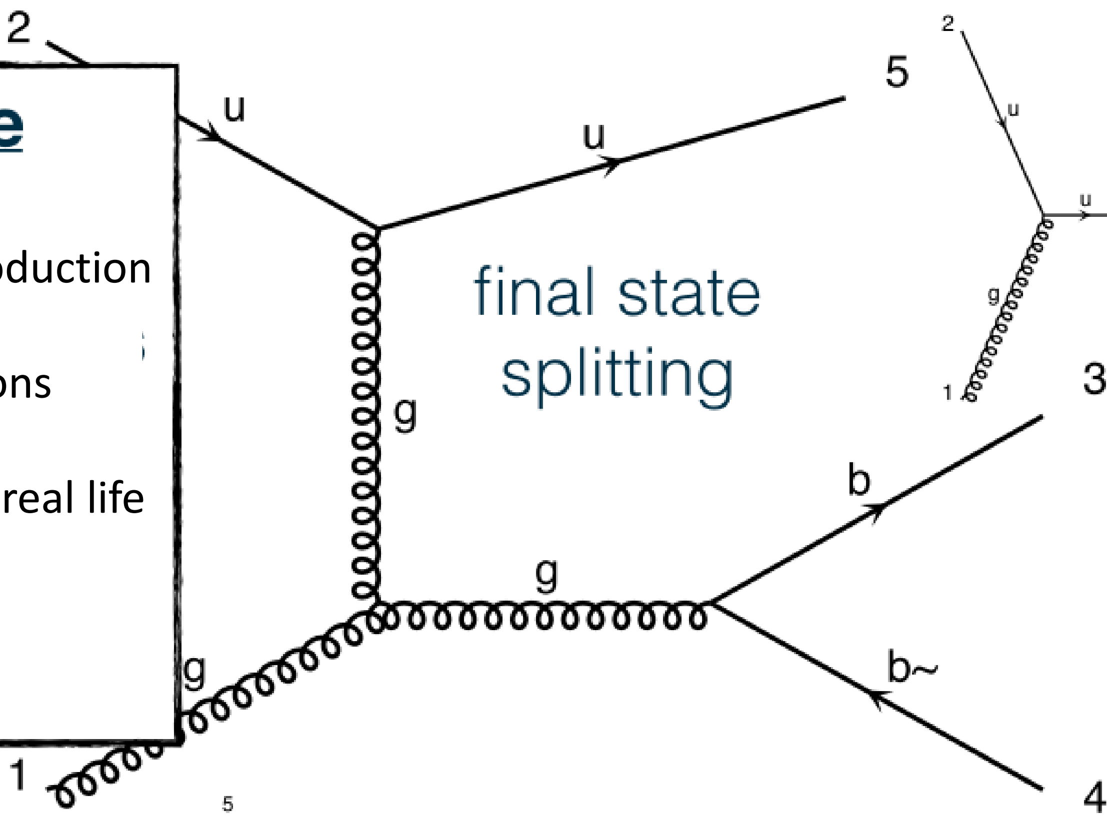
Gluon Splitting to bottom quarks at the LHC

Outline

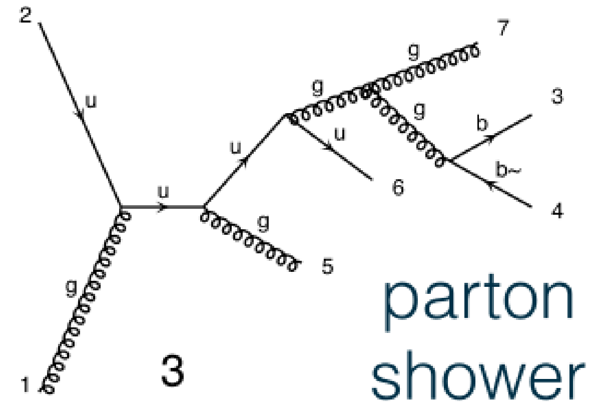
Bottom quark production

Angular correlations

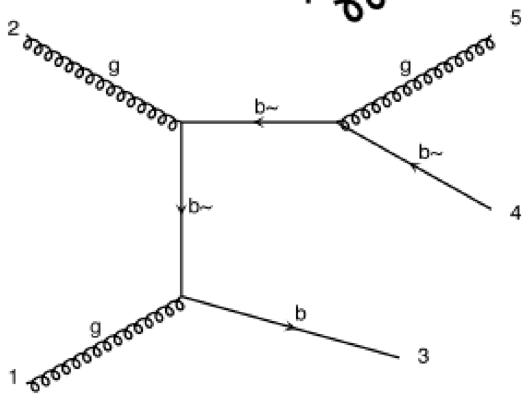
Gluon splitting in real life



final state splitting



parton shower splitting



initial state splitting

$$P_{q\leftarrow g}(x) = \frac{1}{2}\left[x^2 + (1-x)^2\right]$$

leading order splitting function

Diagram produced with MG5_aMC

Bottom quark production at the LHC

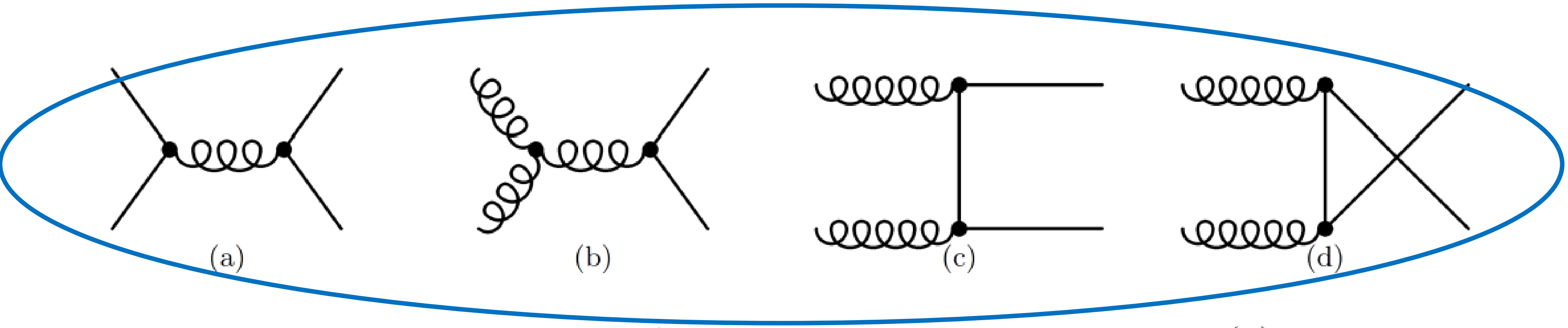


Figure 3.5: Leading order diagrams for heavy-quark pair production: (a) quark-antiquark annihilation $q\bar{q} \to Q\bar{Q}$, (b)-(d) gluon-gluon fusion $gg \to Q\bar{Q}$.

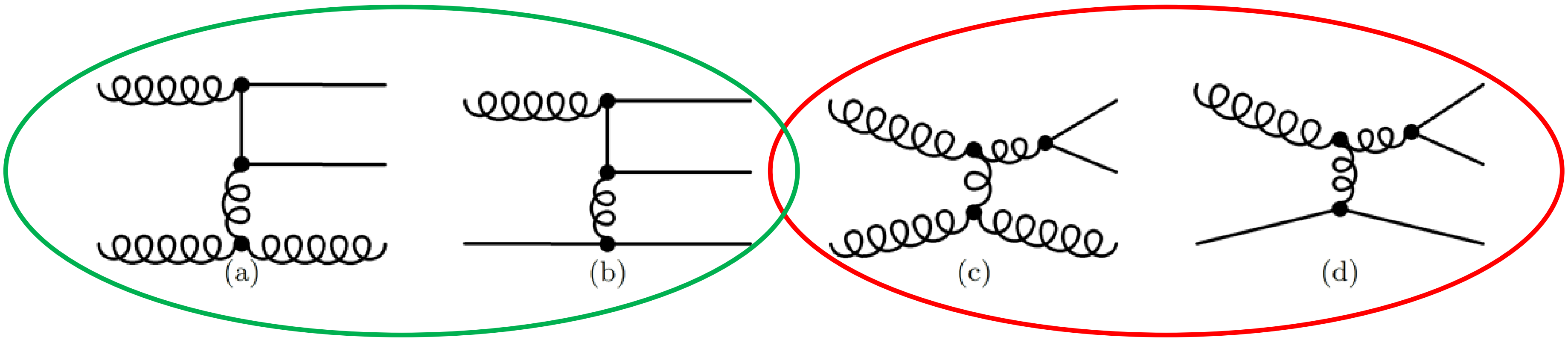
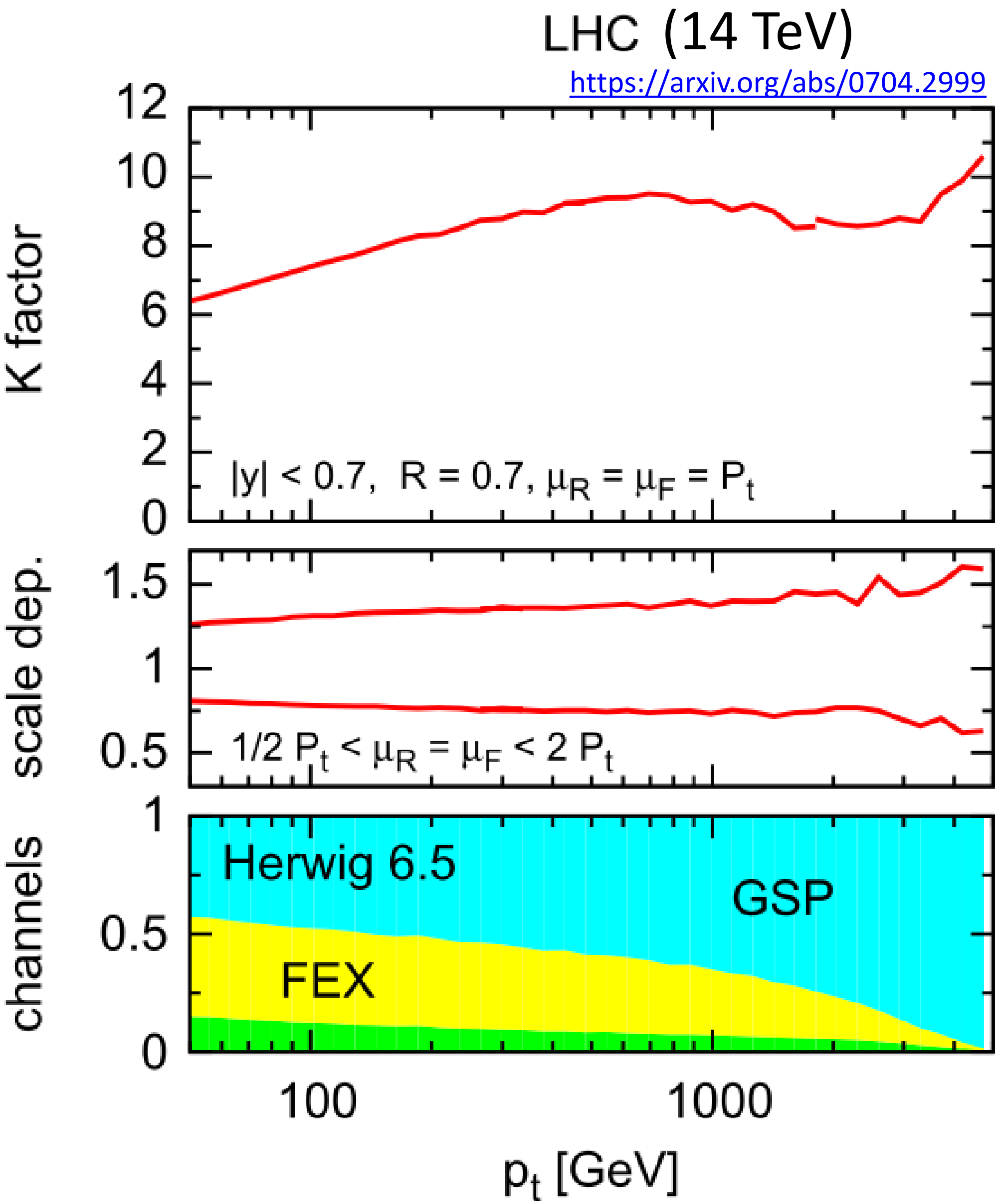
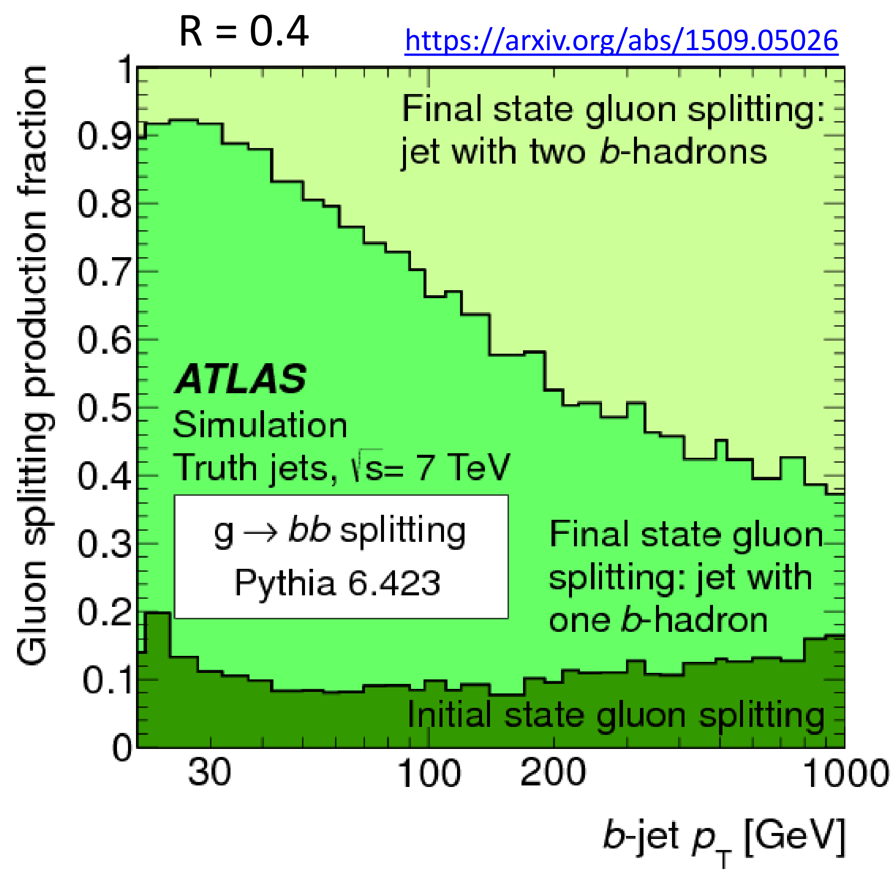


Figure 3.6: Next-to-leading order diagrams for heavy-quark pair production: (a),(b) flavor excitation; (c),(d) gluon splitting.

Bottom quark production at the LHC





Bottom quark production at the LHC

- General properties of the (initial state) gluon splitting:
  - Low $\Delta\phi$
  - Low $\Delta\eta$
  - Mostly (but not only) similar p_T between the two quarks

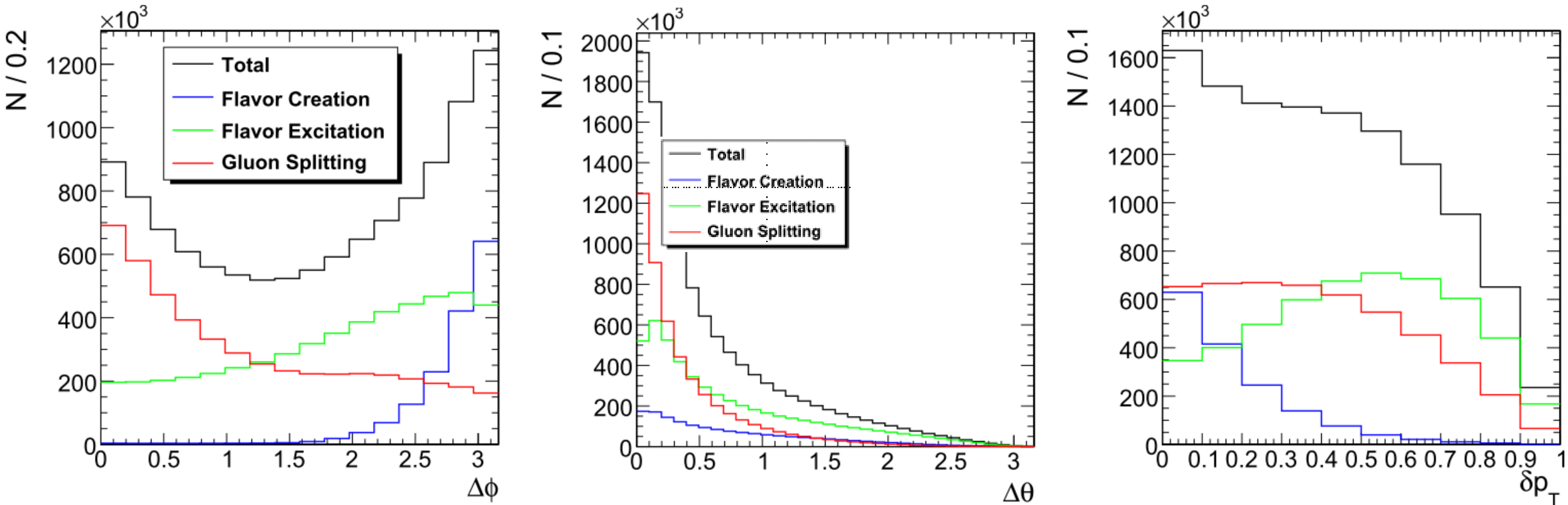


Figure 3.7: Correlated variables between simulated b-quark pairs produced from proton collisions at 10 TeV are also shown for the different mechanisms [108]: $\Delta\phi$ (top left), $\Delta\eta$ (top right), and the quark momentum asymmetry in the transverse plane δp_T (bottom).

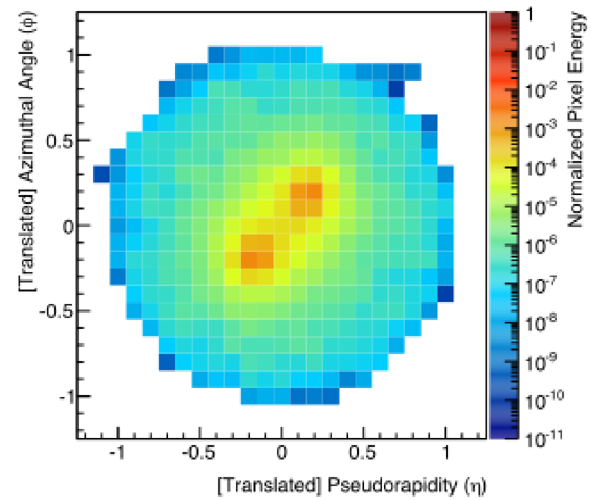
https://cdsweb.cern.ch/record/1311216/

Rich and diverse probe of the SM

Perturbative QCD

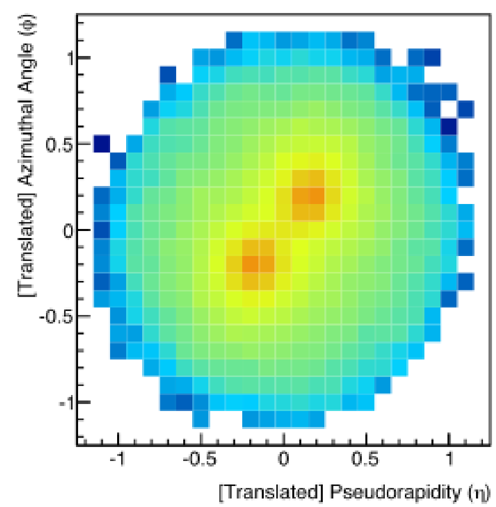
-essentially the only (nearly) direct measurement of a parton splitting function

-pure source of gluon jets (though complicated by B-hadrons)



Non-perturbative QCD

-pure source of color octets →

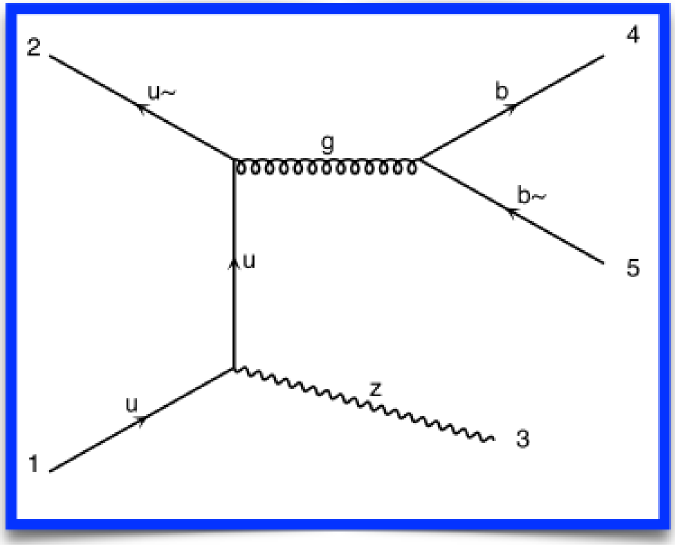


Higgs Boson (self-coupling)

-important background to many Higgs processed (VH, HH, BSM)

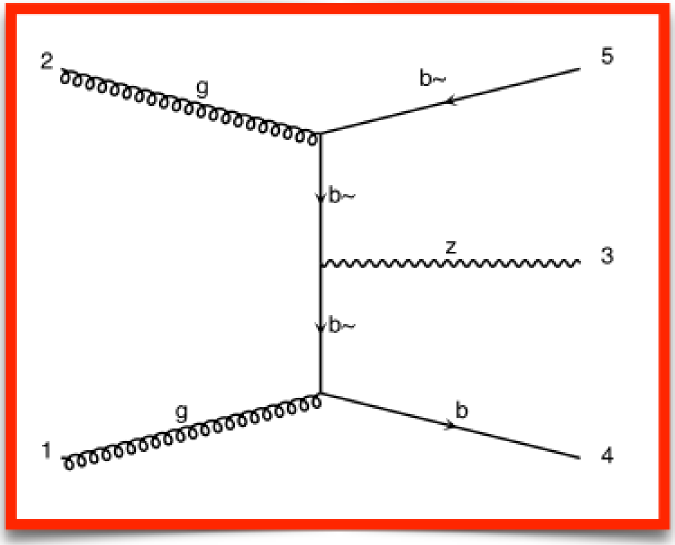
What have we learnt from LHC Run 1?

- So far, several 7 TeV measurements:
  - bb
  - Z+b(b) (gluon splitting diluted by gg-induced process)
  - W+b(b) (only gluon splitting contributes)



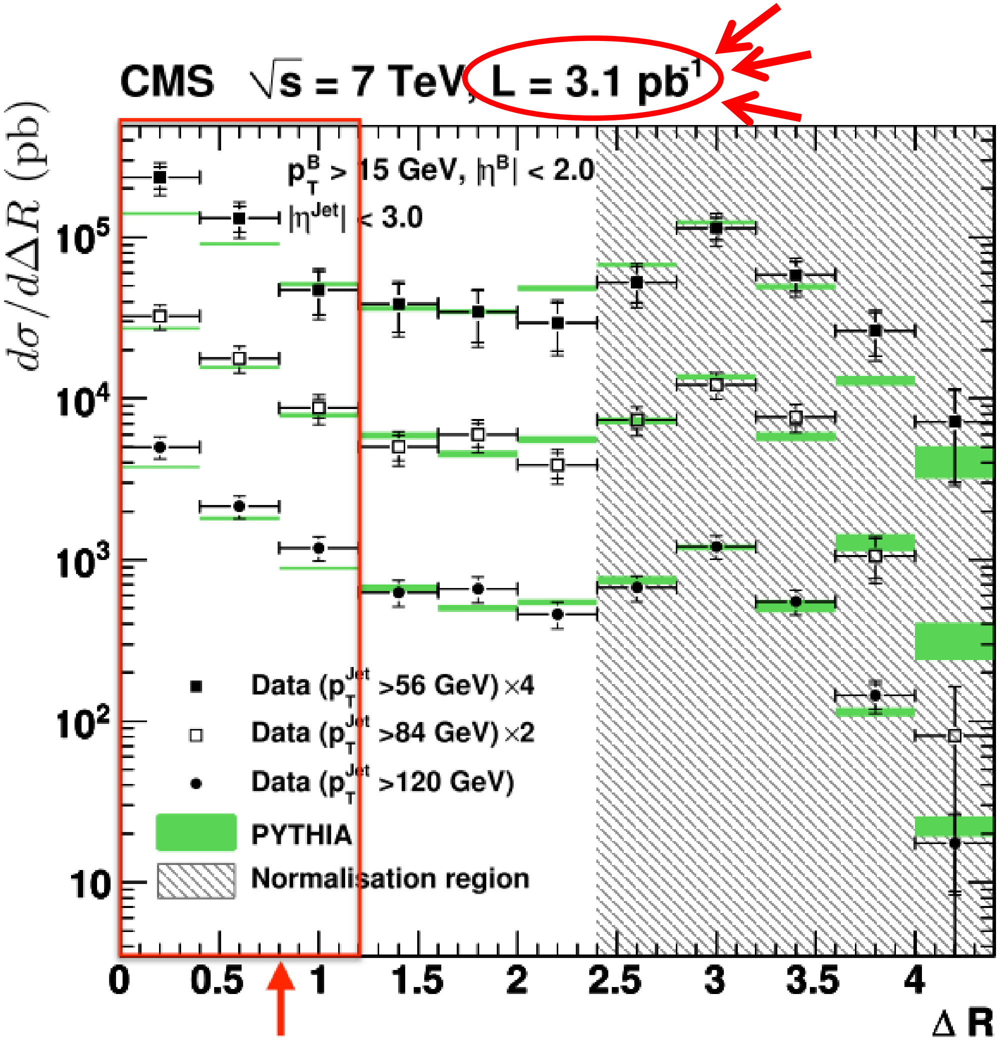
Not FS gluon splitting →

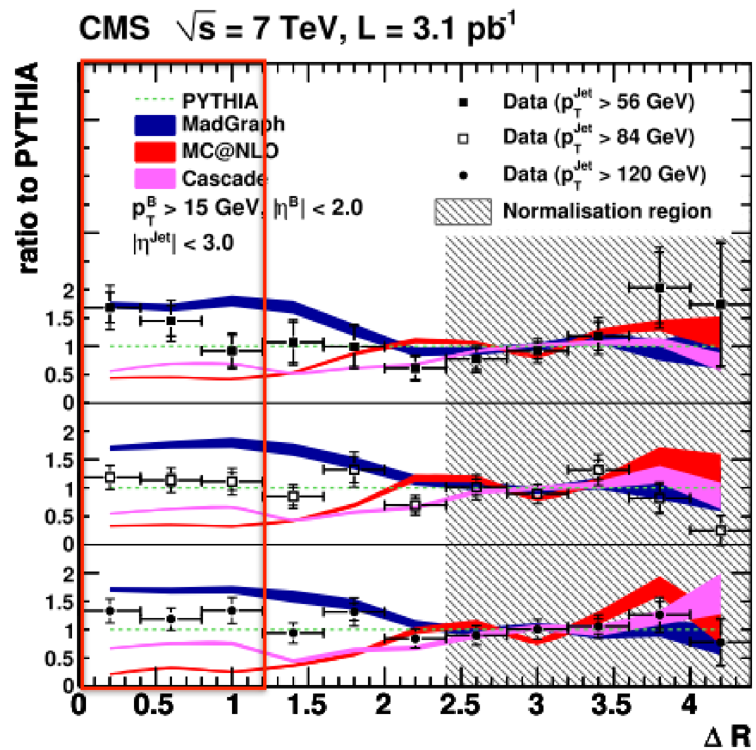
← enhance by looking at low DR



Why no gamma+bb? For gluon splitting, this is more interesting than Z+bb

CMS di-hadrons at 7 TeV





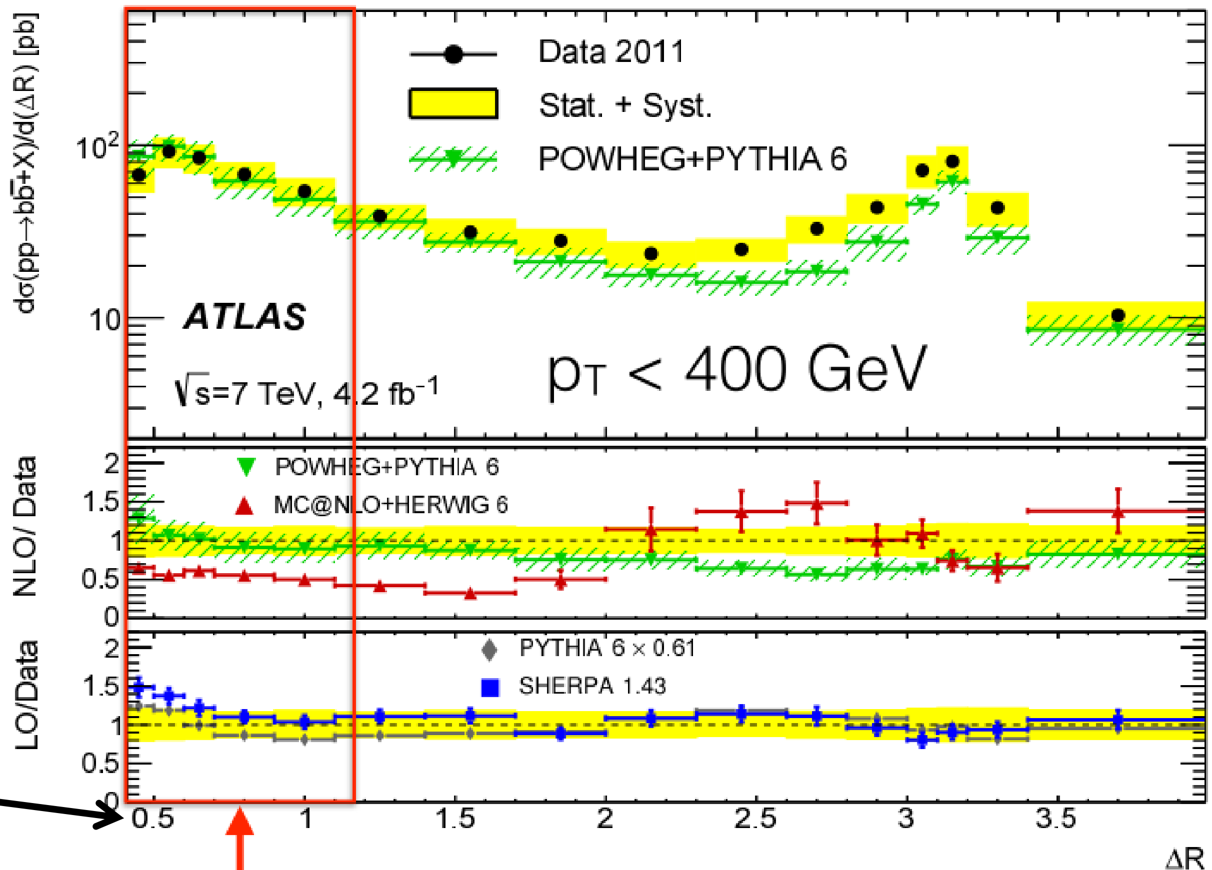
Starting to see the shape in the gluon-splitting dominated regime

As with the ATLAS result, significant differences with the MC (though Pythia is not so bad), though the comparisons there are by now outdated.

https://arxiv.org/abs/1102.3194

ATLAS di-bjets at 7 TeV

Higher p_T (though still relatively low) - trigger limited at low p_T



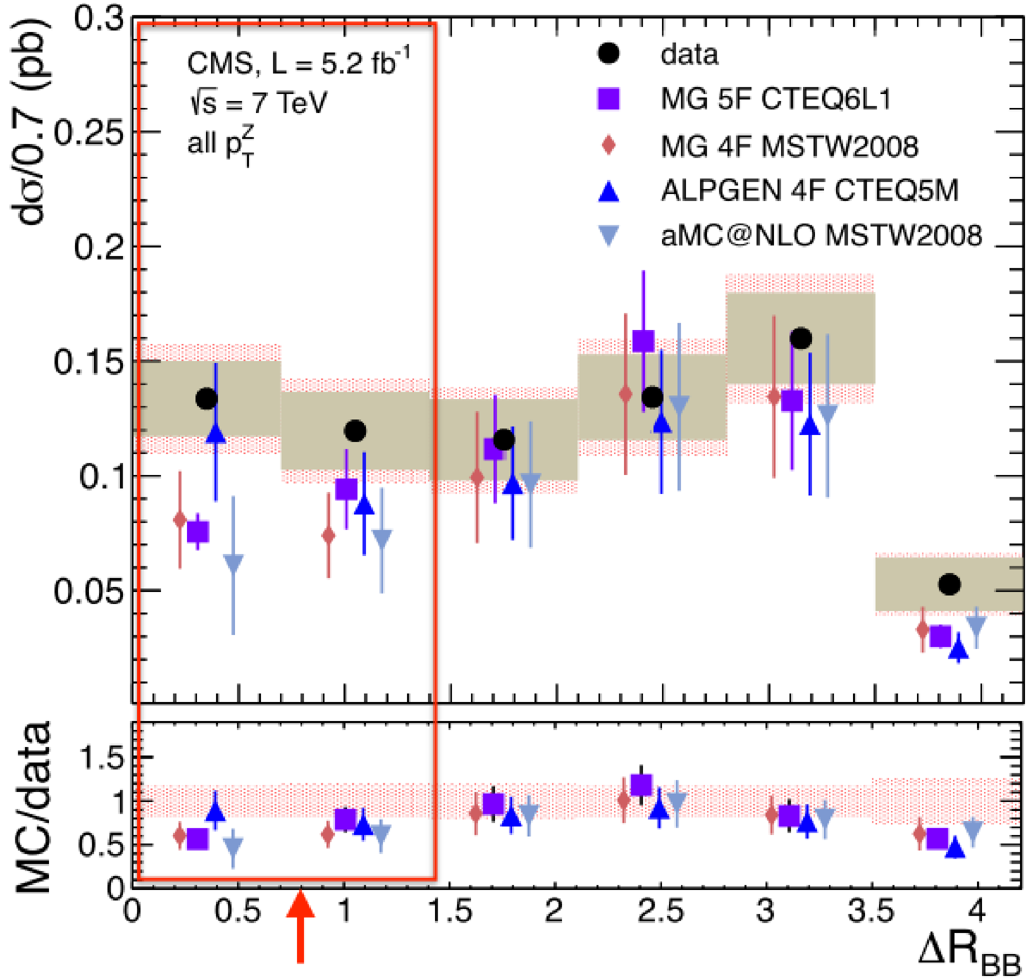
Penalty to pay when using jets

Starting to see the shape in the gluon-splitting dominated regime

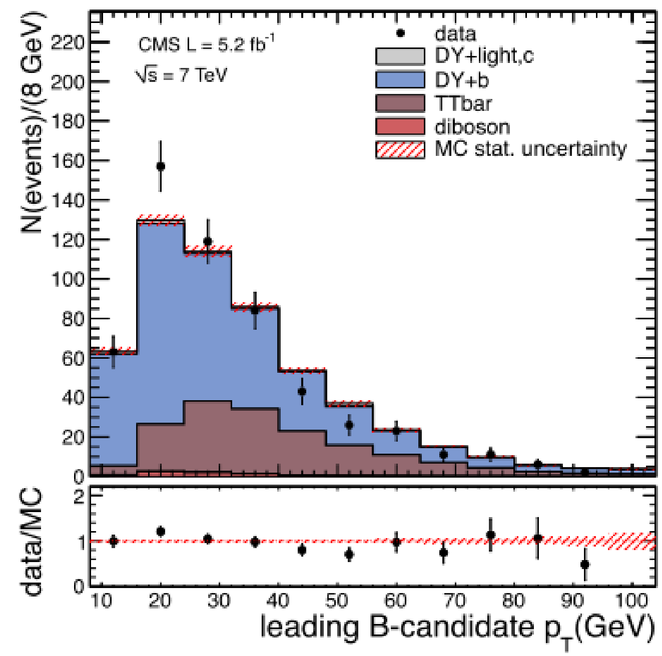
(note that the theory comparisons here are rather dated)

https://arxiv.org/abs/1607.08430

CMS Z+b(b) at 7 TeV



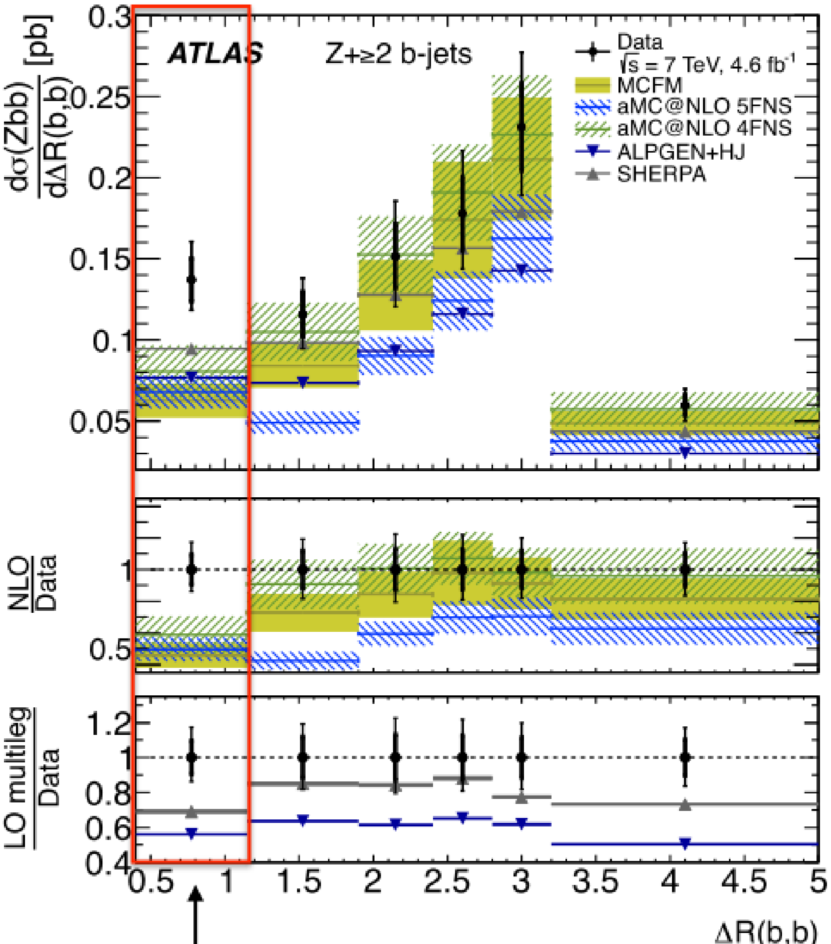
Dominated by gluon splitting

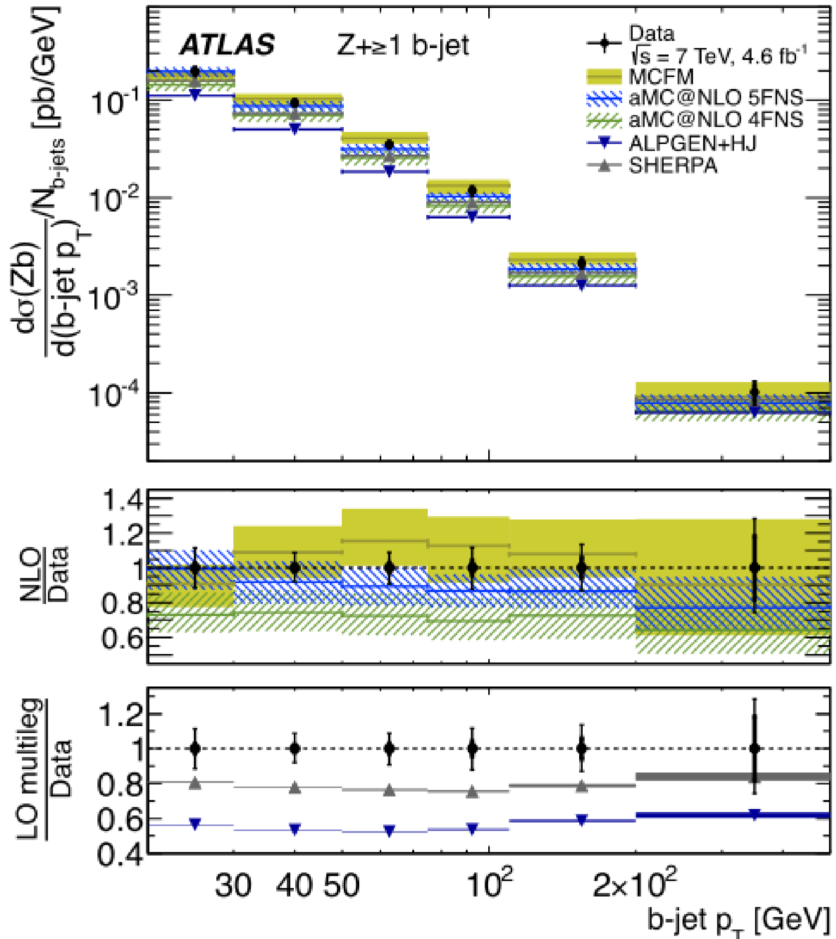


cfr Inclusive cross-section with two b-jets
https://arxiv.org/abs/1402.1521

https://arxiv.org/abs/1310.1349

ATLAS Z+b(b) at 7 TeV



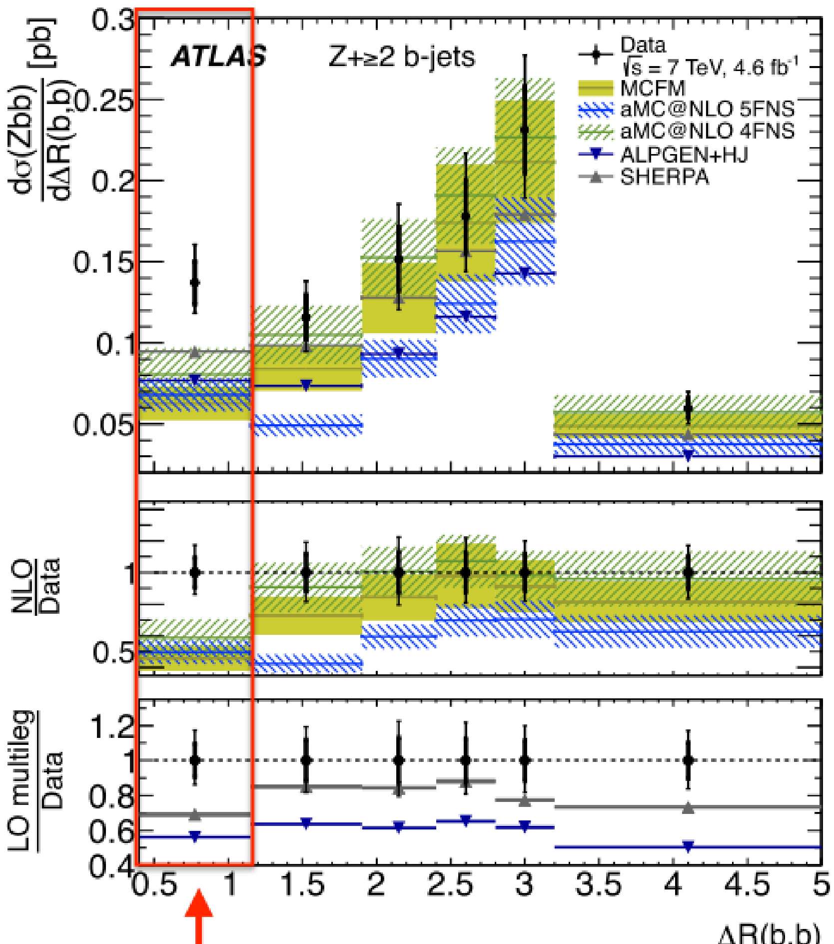


Approaching the PS phase space

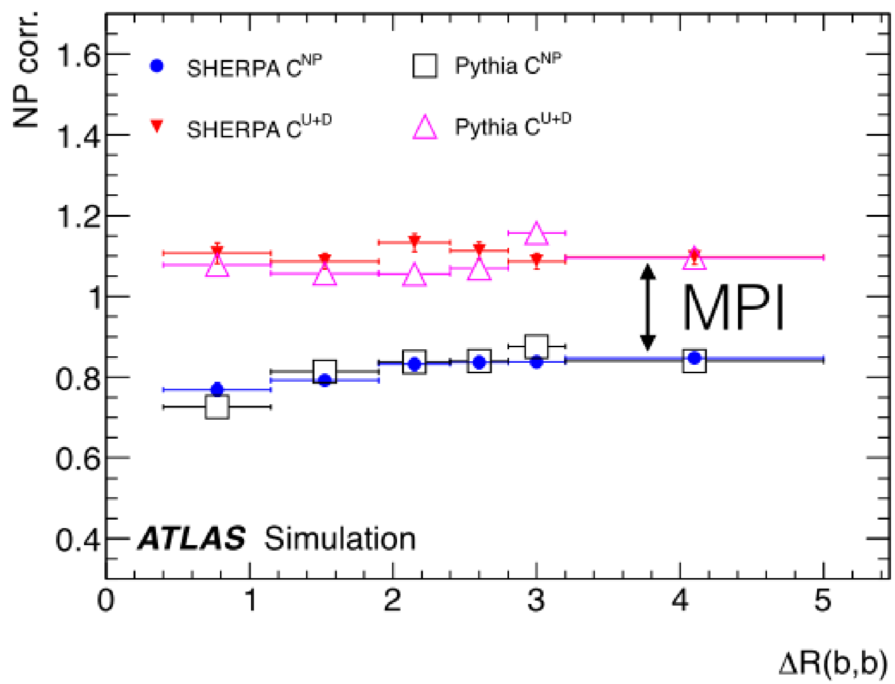
Relatively low p_T regime

https://arxiv.org/abs/1407.3643

ATLAS Z+b(b) at 7 TeV



Dominated by gluon splitting



Interestingly, MPI is a relatively big effect here.

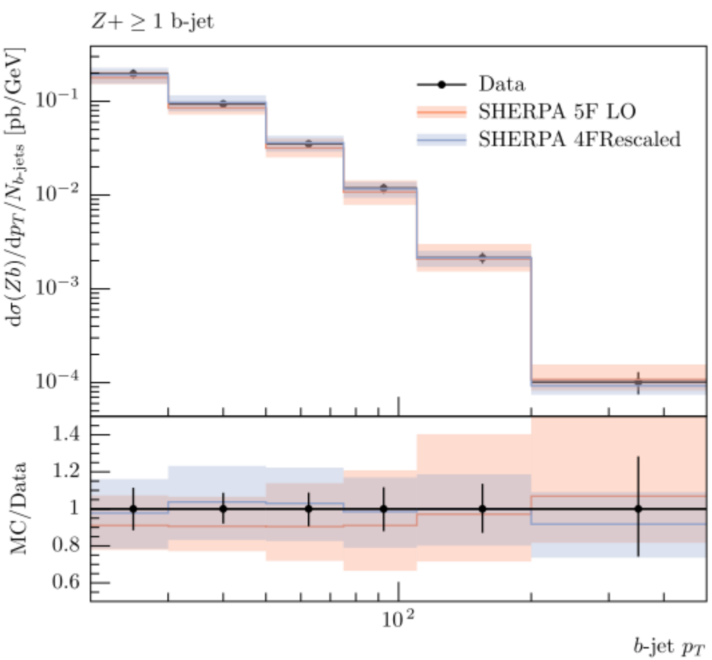
https://arxiv.org/abs/1407.3643

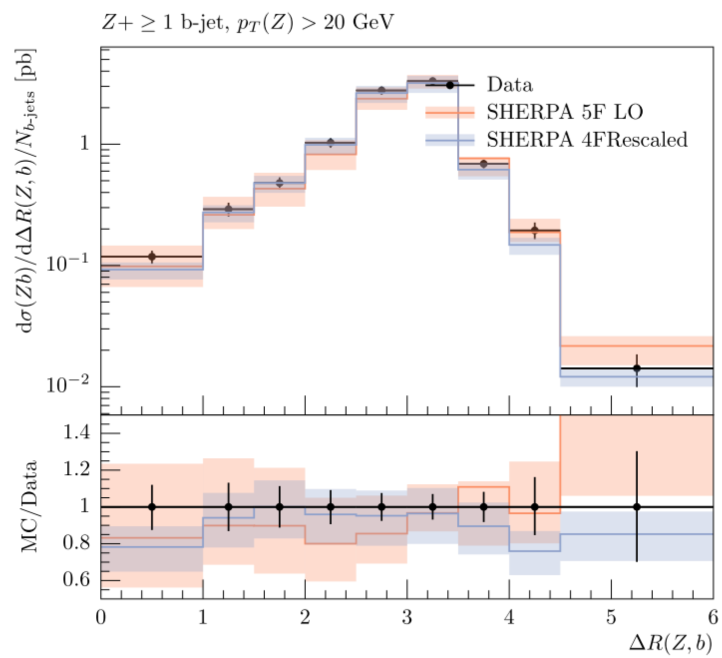
LH15 exercise on V+bb

6 Study of associated production of vector bosons and b-jets at the LHC

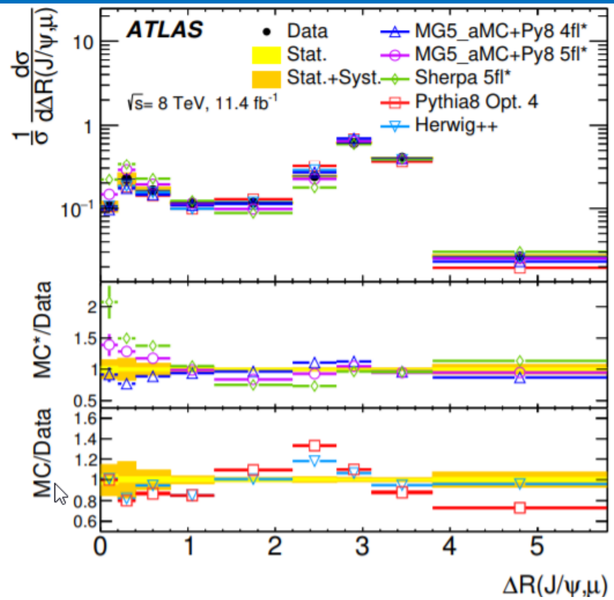
6.6 Conclusions

We presented a comparison of generators predictions using 4F and 5F scheme to most recent measurements of vector boson production in association with b-jets at the LHC. In the 4F scheme a good agreement is found among the different generators at NLO accuracy, and among different matrix-element to parton-shower matching algorithms. The agreement with data however is good only when two b-jets are tagged in the final state or, when one b-jet only is required, if a rescaling to the 5F integrated cross-section is applied. For Wbb, even taking into account the contribution from MPI, predictions seem to significantly undershoot the data. The Zb(b) production has been compared with predictions obtained in the 5F scheme with different setups,

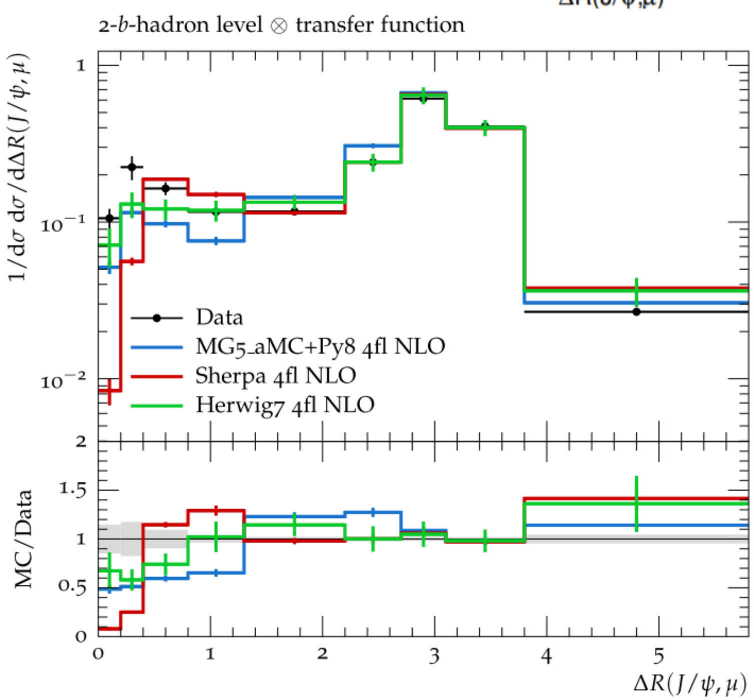


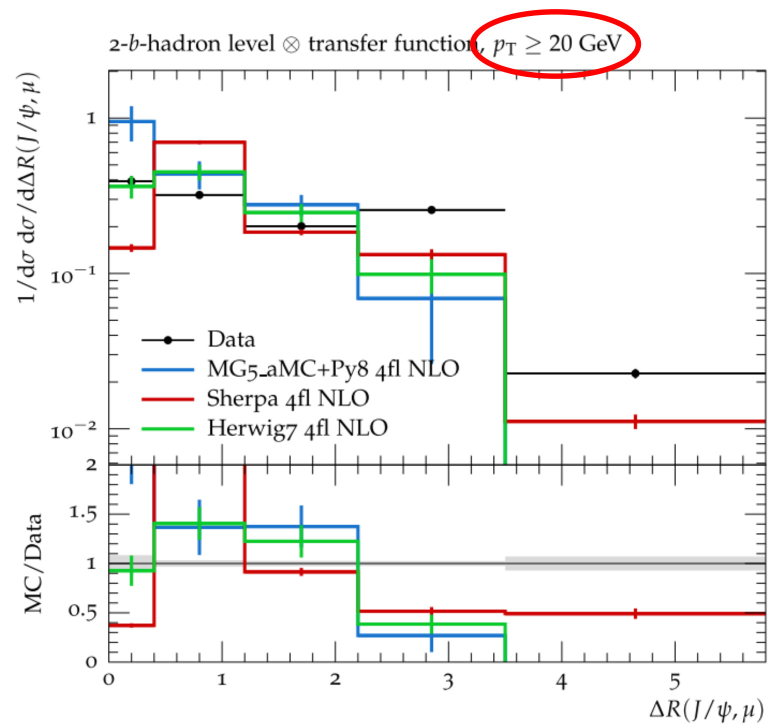


LH17 exercise on hadron correlations



- Using the ATLAS di-hadron correlation measurement (https://arxiv.org/abs/1705.03374) in the channel **B(→ J/ψ[→ μμ] + X)B(→ μ + X)**
  - Go down to ΔR~0 but need to unfold fragmentation





Where is this going?

- V+bb background modeling strategies for VH(bb)

CMS

- **V+(light-flavor) modeling**
  CRs defined by inverting b-tagging requirements (anti-2-btag)
- Data driven pre-fit systematics
- **V+(heavy-flavor) modeling**
  CRs defined by inverting M(jj)-window

  (b-tag $CMVA_{min}$ shape fitted from CRs)

Process	0-lepton	1-lepton	2-lepton low-p_T(V)	2-lepton high-p_T(V)
W0b	1.14 ± 0.07	1.14 ± 0.07	—	—
W1b	1.66 ± 0.12	1.66 ± 0.12	—	—
W2b	1.49 ± 0.12	1.49 ± 0.12	—	—
Z0b	1.03 ± 0.07	—	1.01 ± 0.06	1.02 ± 0.06
Z1b	1.28 ± 0.17	—	0.98 ± 0.06	1.02 ± 0.11
Z2b	1.61 ± 0.10	—	1.09 ± 0.07	1.28 ± 0.09
tt	0.78 ± 0.05	0.91 ± 0.03	1.00 ± 0.03	1.04 ± 0.05

ATLAS

- **V+(heavy-flavor) modeling**
  W: dedicated CR (large m-top, low m-bb) - yield only, no shape
  Z: no *dedicated* CR - full m-bb spectrum included in the SRs
  Theory driven pre-fit systematics
  V+hf = V+(bb, bc, bl, cc)

Process	Normalisation factor
$t\bar{t}$ 0- and 1-lepton	0.90 ± 0.08
$t\bar{t}$ 2-lepton 2-jet	0.97 ± 0.09
$t\bar{t}$ 2-lepton 3-jet	1.04 ± 0.06
W + HF 2-jet	1.22 ± 0.14
W + HF 3-jet	1.27 ± 0.14
Z + HF 2-jet	1.30 ± 0.10
Z + HF 3-jet	1.22 ± 0.09

Backgound reweighting corrections for V+jets:

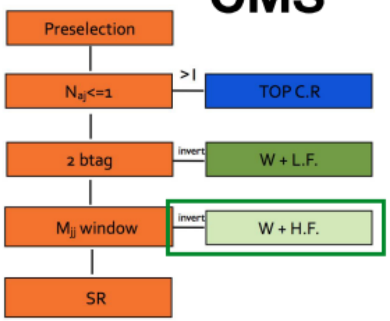
- $f(p_T^V)$ differential correction (up to 10% at 400GeV) accounting for EW corrections
- $f(p_T^V)$ dedicated 1-lepton correction on W+light, W+b(b), ttbar, single-t
- deltaEta(jj) correction from LO/NLO comparison (depending on #b-labeled jets)

https://indico.cern.ch/event/665524/contributions/2903789/attachments/1623131/2583678/2018-03-26_LHCHXSWG_WG1_VH_exp.pdf

Where is this going?

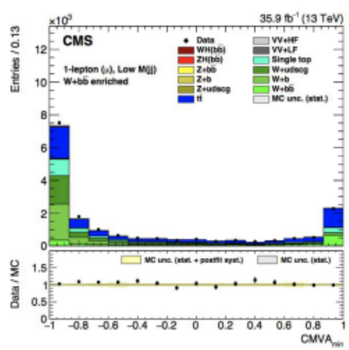
- More detailed example: W+b(b) background modeling for VH(bb)

CMS



- Define dedicated control region (CR)
- Scale factors applied from CR to Signal Regions (SR)
- Systematic uncertainties fully correlated between CR and SR

Variable	W+HF
$p_T(j_1)$	>25
$p_T(j_2)$	>25
$p_T(jj)$	>100
$p_T(V)$	>100
CMVA_{max}	$>\text{CMVA}_T$
N_{aj}	=0
$N_{a\ell}$	=0
$\sigma(p_T^{miss})$	>2.0
$\Delta\phi(\vec{p}_T^{miss}, \ell)$	<2
$M(jj)$	<90, [150, 250]



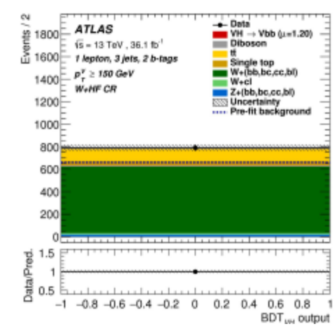
ATLAS

- **standard 1-lepton selection +**
  m(bb) < 75GeV
  m(top) > 225GeV

- Scale factor fitted directly in the SR
- extrapolation uncertainties from CR to SR obtained from theory
  - Sherpa 2.2.1 muR, muF, ckkw, qsf scale variations
  - Sherpa 2.2.1 comparison with Madgraph_aMC@NLO 2.2.2

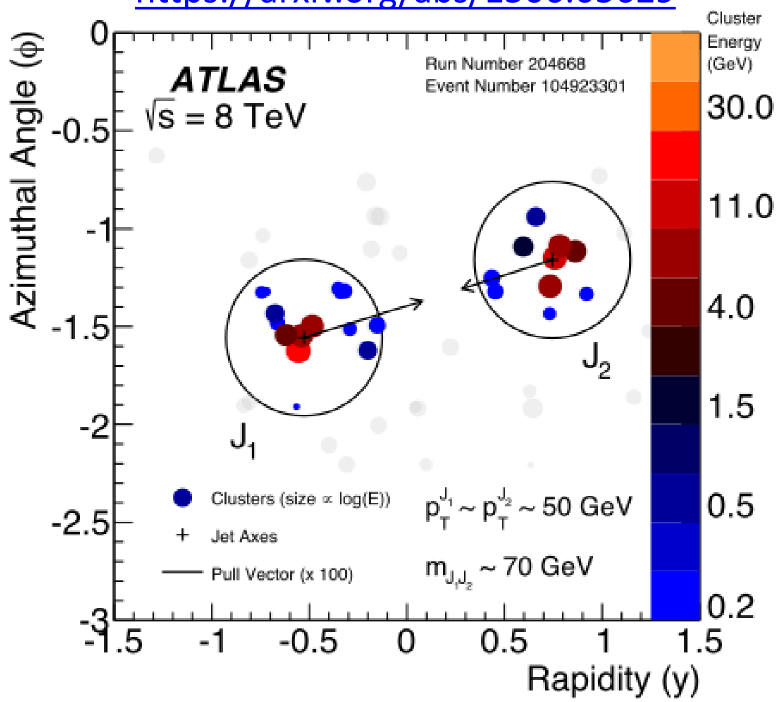
- Pre-fit theory modeling uncertainties

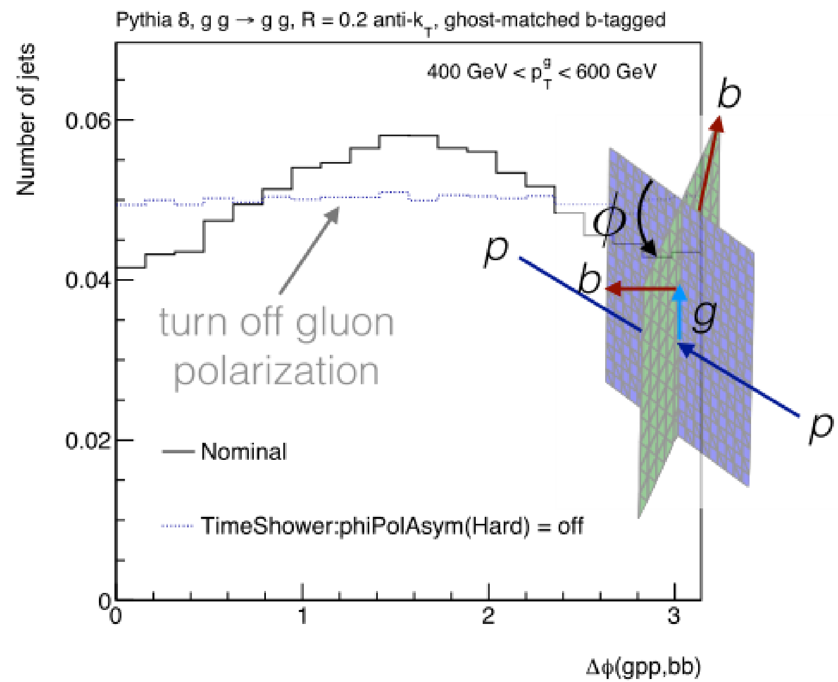
	W + jets
$W + ll$ normalisation	32%
$W + cl$ normalisation	37%
$W + bb$ normalisation	Floating (2-jet, 3-jet)
$W + bl$-to-$W + bb$ ratio	26% (0-lepton) and 23% (1-lepton)
$W + bc$-to-$W + bb$ ratio	15% (0-lepton) and 30% (1-lepton)
$W + cc$-to-$W + bb$ ratio	10% (0-lepton) and 30% (1-lepton)
0-to-1 lepton ratio	5%
W + HF CR to SR ratio	10% (1-lepton)
m_{bb}, p_T^V	S



What else can and should we measure?

https://arxiv.org/abs/1506.05629



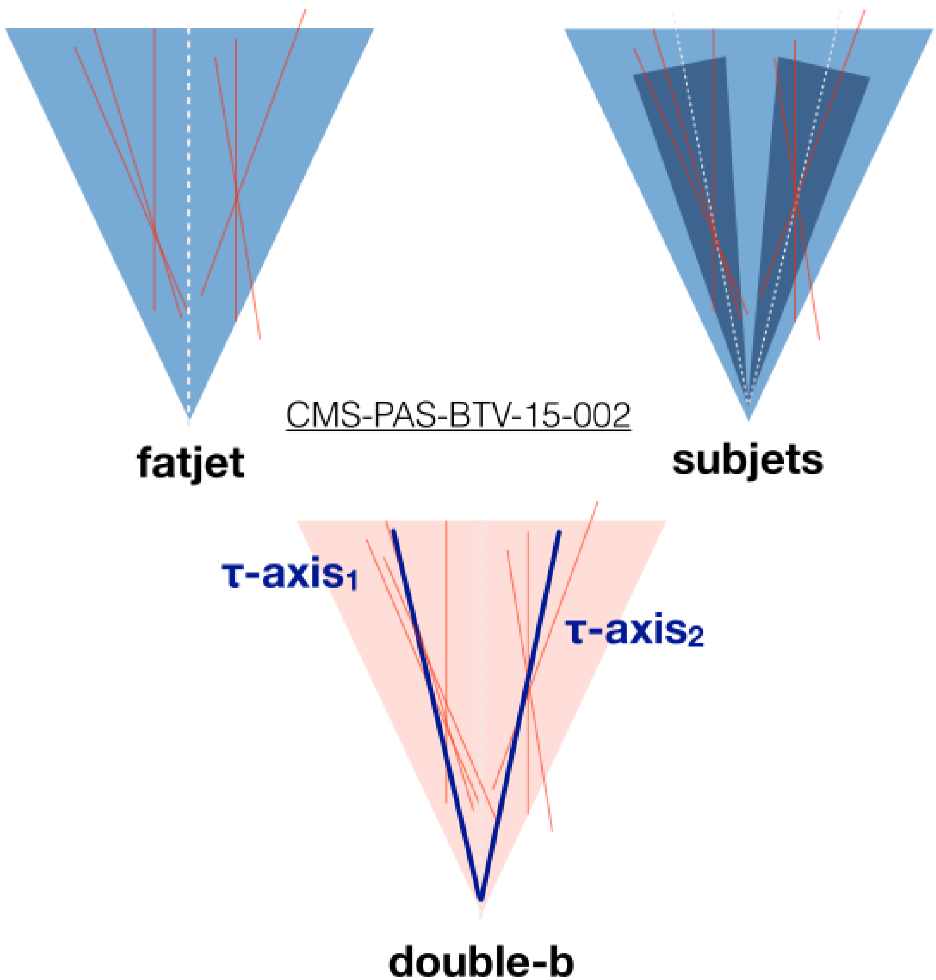


The singlet radiation pattern has been measured in W decays - should measure octet in g->bb!

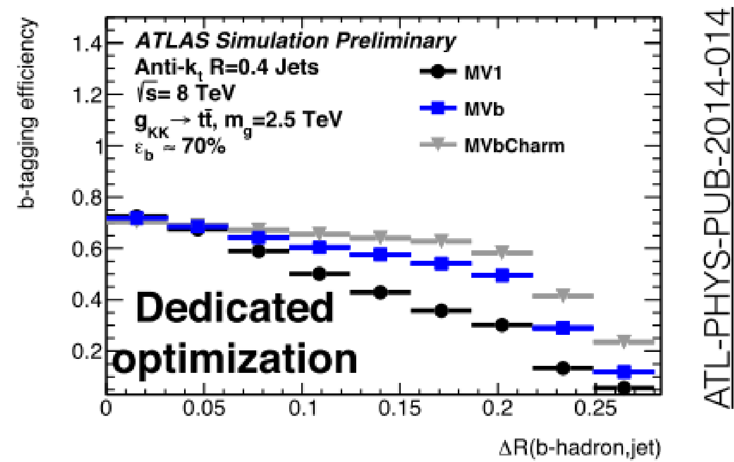
We should measure all aspects of the g->bb production (angles + energies)

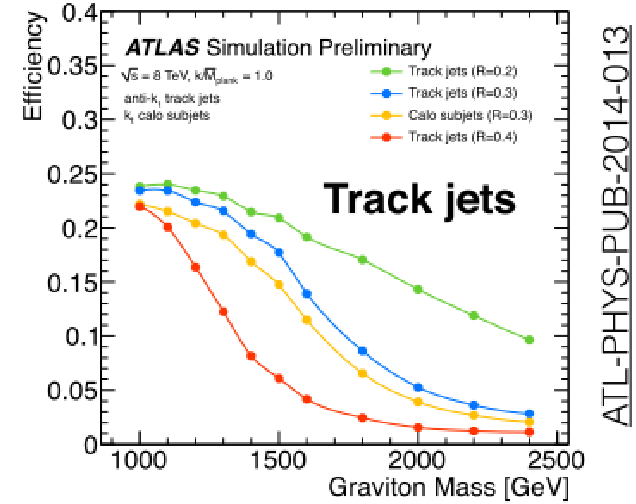
What have we been up to since early Run 1?

Since 7 TeV, there has been a lot of work to improve b-tagging inside jets and to measure the efficiency in data.



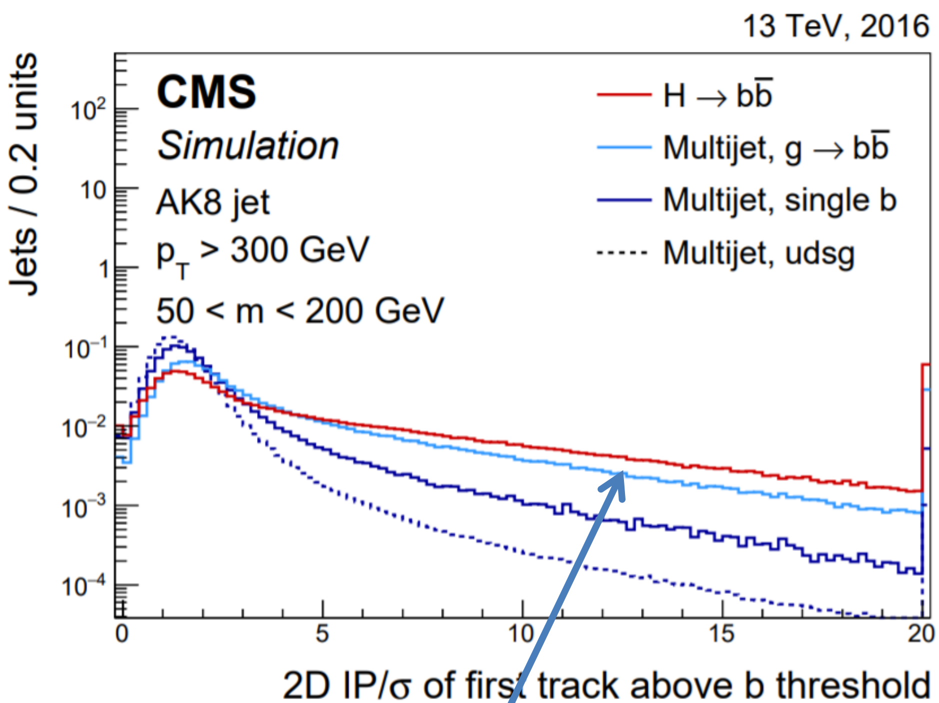
CMS-PAS-BTV-15-002

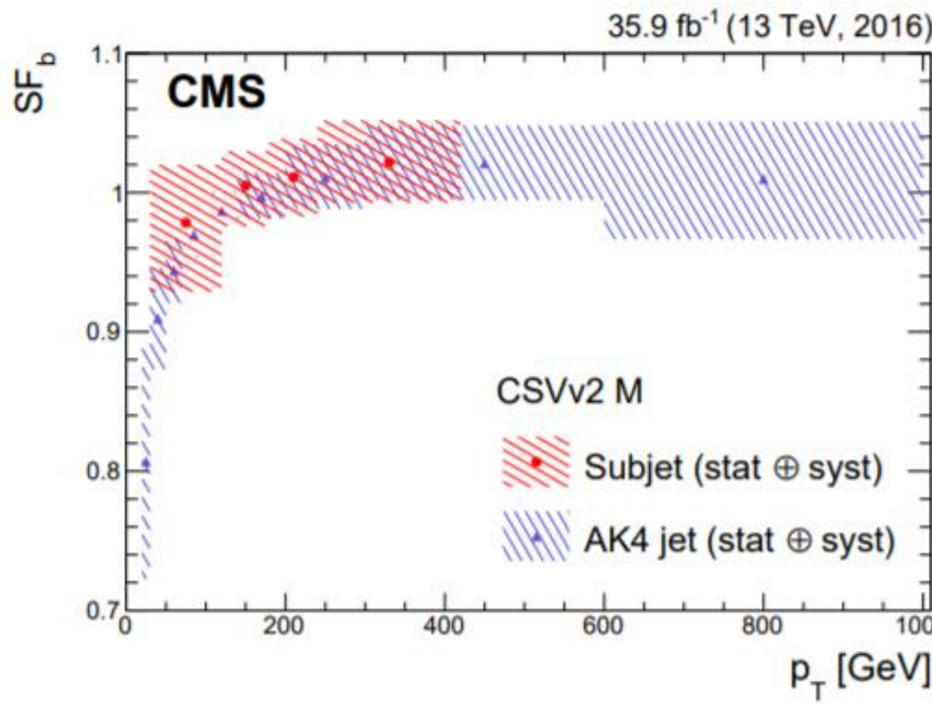




CMS g→bb performance

https://arxiv.org/abs/1712.07158





~10% uncertainty; stats limited

Higgs and g->bb are very similar!

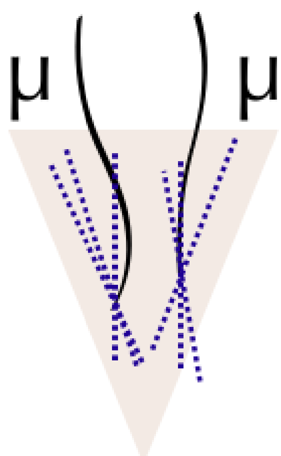
8 TeV analysis

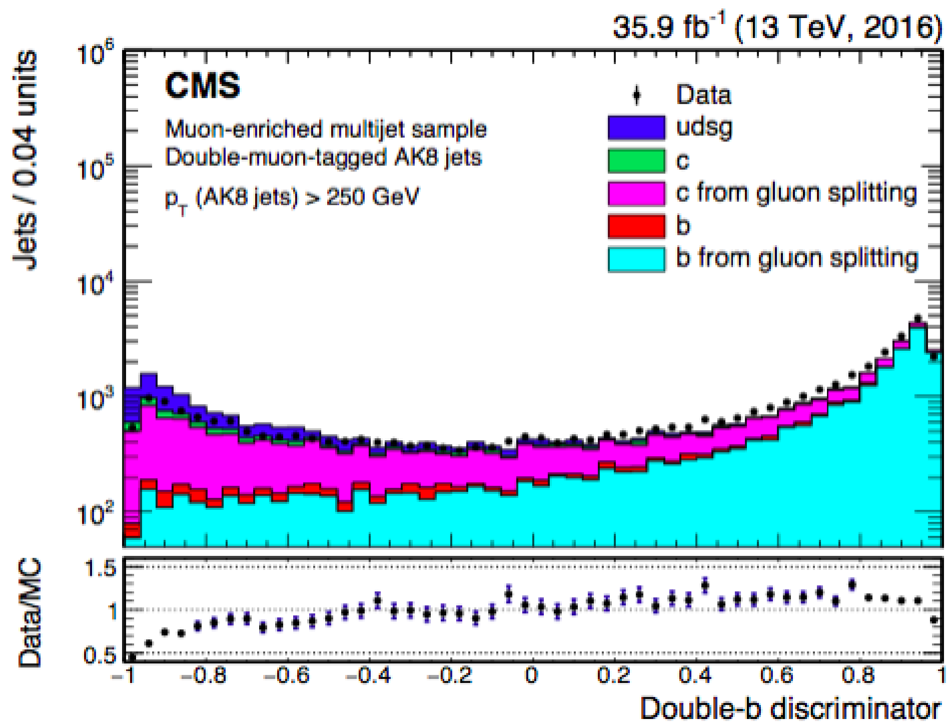
Refining the calibration with more data

https://arxiv.org/abs/1712.07158

Jet selection has been designed to ensure jets are signal-like

- High AK8 p_T jet (p_T > 250 GeV)
- **double-muon** tagged jets (muon p_T > 7 GeV)
- **mass cut** (>50 GeV)





Where is this going?

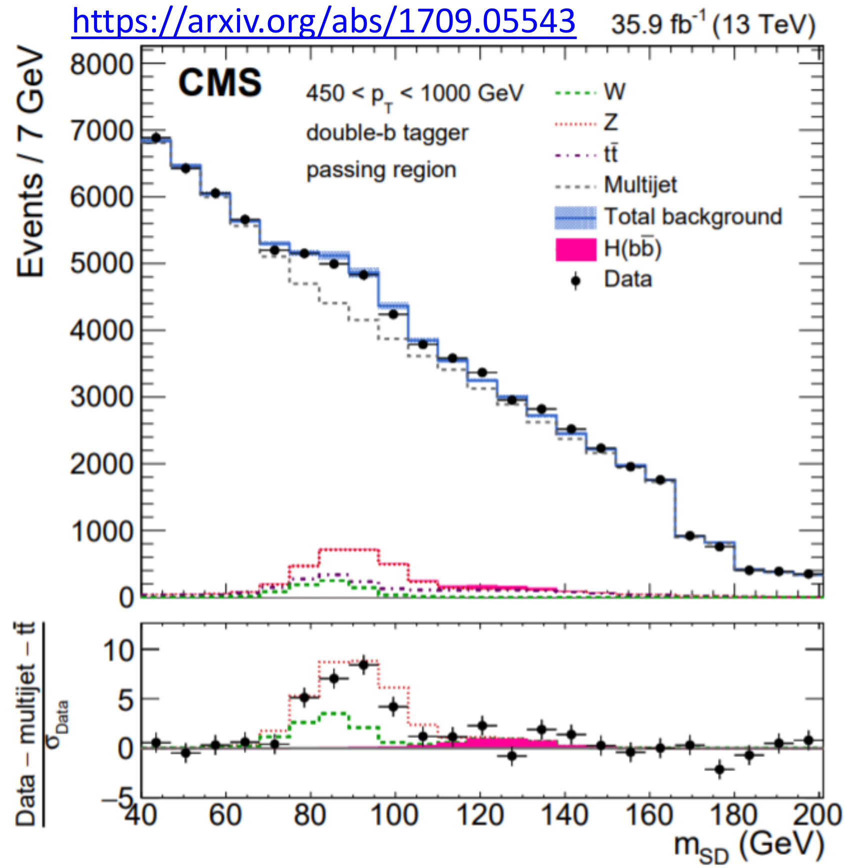
The Run 2 efforts are motivated by (boosted) Higgs tagging.

For many of these searches, g->bb is the main background.

Data-driven techniques are used because the MC is not reliable (still needed to check closure)

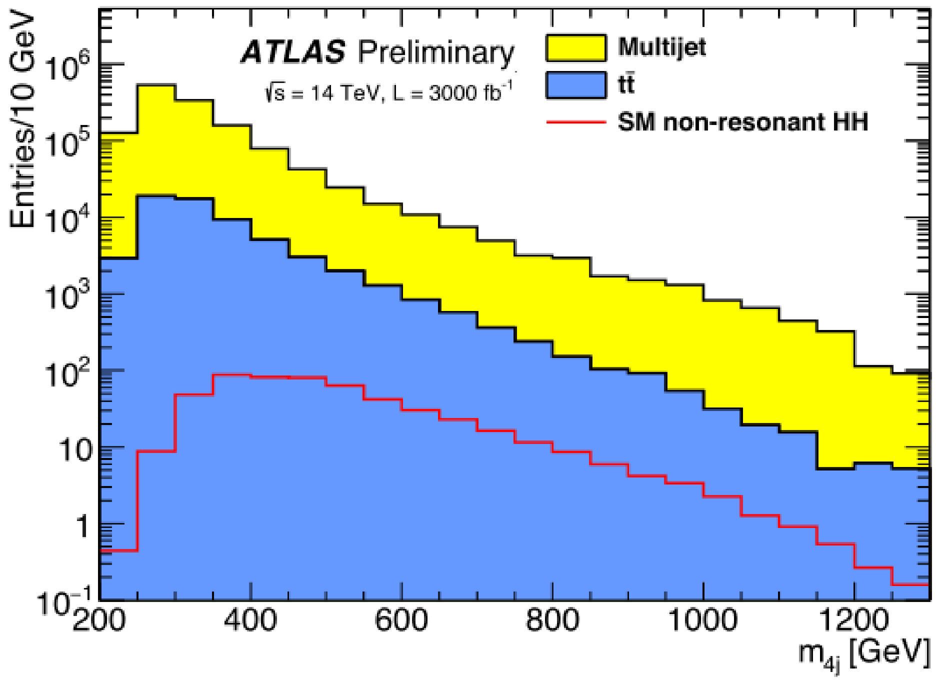
Is this something that could change in the next 10 years?

https://arxiv.org/abs/1709.05543



https://twiki.cern.ch/twiki/bin/view/CMSPublic/HighPtTrackingDP

HH @ HL-LHC



https://cds.cern.ch/record/2221658/

For example, one of the most challenging and important measurements is the Higgs self-coupling.

The g->bb background is complicated, but maybe a better understanding could be a game-changer here!

Source	$\Delta\mu$
Luminosity	0.05
Jet Energy	0.09
b-tagging	0.34
Theoretical	0.10
Multijet	1.85
$t\bar{t}$	2.83

>> CMS HH projections

http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/FTR-15-002/

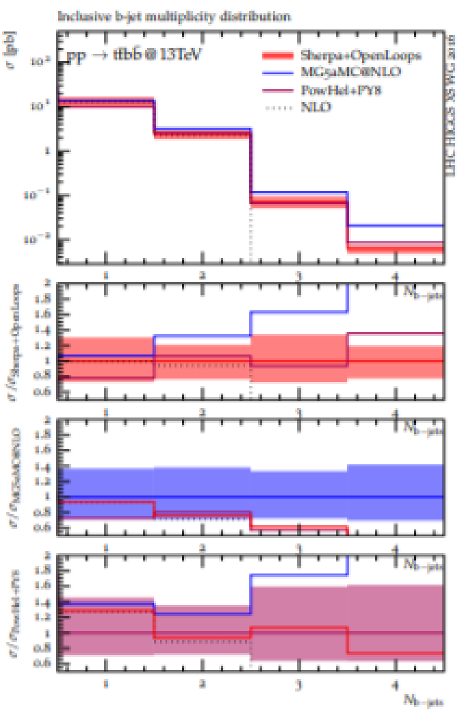
What about ttH(bb)?

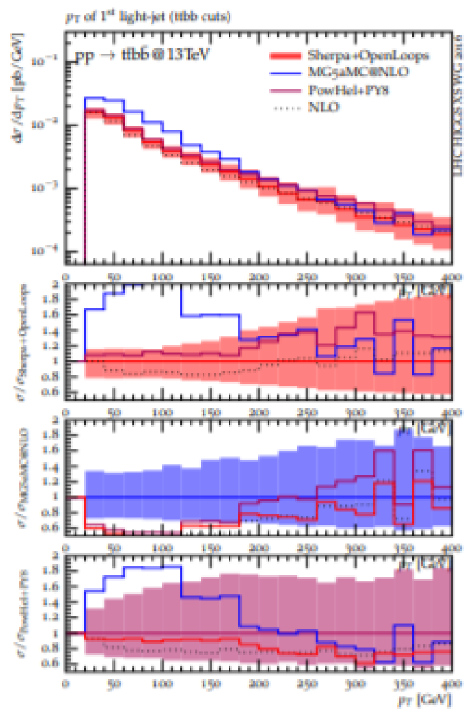
- after YR4 we decided to focus $t\bar{t}H/tH$ subgroup activities on highest priority TH issues in EXP analyses
- in the recent months we focussed on TH uncertainties of $t\bar{t}+b$-jet background, which dominate $t\bar{t}H(b\bar{b})$ systematics
- $t\bar{t}+b$-jet data help, but precise "extrapolation" to signal region calls for $t\bar{t}+b$-jet shape uncertainties at 10% level
- $pp \to t\bar{t}b\bar{b}$ remains a nontrivial multi-particle multi-scale QCD process
- better understanding of its QCD dynamics and NLOPS technicalities crucial for assessment of TH uncertainties

https://indico.cern.ch/event/407347/contributions/975965/attachments/1211342/1766869/hxswg16.pdf

What about ttH(bb)?

YR4 highlights: Validation of NLO+PS tools, $t\bar{t} + b$ jets





↪ switch off top decays, hadronization, UE

↪ To better compare the effect of

- different matchings
- different parton showers
- different flavor scheme

Discrepancies emerge that will have to be understood if we want to resolve the very large systematic uncertainties that affects experimental analyses

↪ **This is becoming a limiting factor.**

https://indico.cern.ch/event/555360/contributions/2274684/attachments/1353121/2044020/tth_theory_october_16.pdf

ttbb 5F at NLO

NLOPS ttbb 5F (e.g. POWHEG hvq)

$t\bar{t}b\bar{b}$ described through $t\bar{t}j$ tree MEs plus $g \to b\bar{b}$ shower splittings

$gg \to t\bar{t}g$ + FS splittings $\bar{b}g \to t\bar{t}\bar{b}$ + IS splittings

Precision vs accuracy

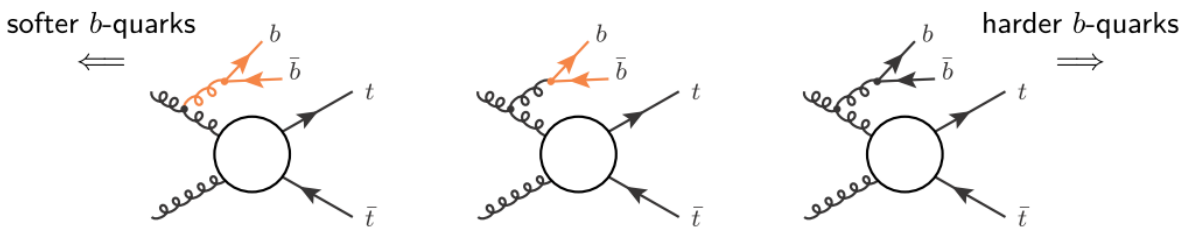
- precision lower than LO (parton shower allows for accurate tuning to data)

Calls for improved description based on $t\bar{t}b\bar{b}$ MEs

- crucial for more realistic TH uncertainties

(N)LO ttbar from merging tt+0,1,2j 5F

$t\bar{t}b\bar{b}$ described through $t\bar{t}+0,1,2$ jet MEs and $g \to b\bar{b}$ shower splittings



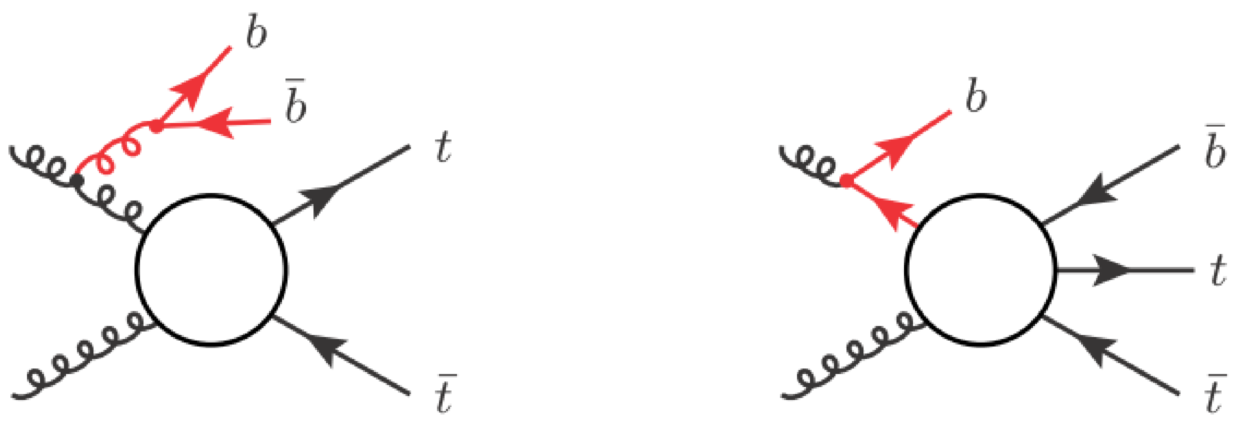
$t\bar{t}$ + PS $t\bar{t}g$ + PS $t\bar{t}b\bar{b}$

Precision and CPU cost strongly depend on choice of merging cut Q_{cut}

- separates ME regions ($k_T > Q_{\mathrm{cut}}$) from shower regions ($k_T < Q_{\mathrm{cut}}$)

Does this describe $t\bar{t}$+b-jet production mostly through $t\bar{t}b\bar{b}$ MEs?

ttbb 4F at NLO



4F $t\bar{t}b\bar{b}$ MEs with $m_b > 0$ cover full b-quark phase space

- NLO precision for $t\bar{t} + 2\,b$-jet and $1\,b$-jet! [Cascioli et al '13]
- 80% LO uncertainty reduced to 20–30% at NLO
- collinear $g \to b\bar{b}$ splittings and m_b effects very important

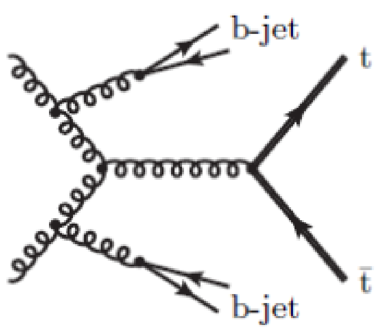
what about drawbacks of 4F scheme (e.g. no b-quark PDF)?

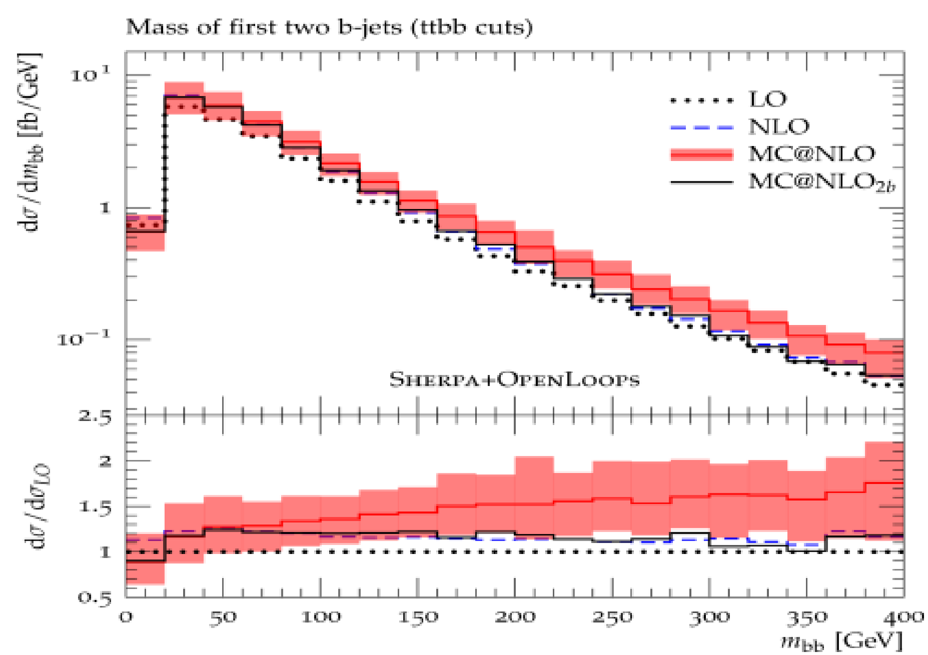
NLOPS ttbb 4F SHERPA+OPENLOOPS

Convergence of 4F scheme but unexpected MC@NLO enhancement

	ttb	$ttbb$	$ttbb\,(m_{bb} > 100)$
σ_{LO}[fb]	$2644^{+71\%+14\%}_{-38\%-11\%}$	$463.3^{+66\%+15\%}_{-36\%-12\%}$	$123.4^{+63\%+17\%}_{-35\%-13\%}$
σ_{NLO}[fb]	$3296^{+34\%+5.6\%}_{-25\%-4.2\%}$	$560^{+29\%+5.4\%}_{-24\%-4.8\%}$	$141.8^{+26\%+6.5\%}_{-22\%-4.6\%}$
$\sigma_{\mathrm{NLO}}/\sigma_{\mathrm{LO}}$	1.25	1.21	1.15
$\sigma_{\mathrm{MC@NLO}}$[fb]	$3313^{+32\%+3.9\%}_{-25\%-2.9\%}$	$600^{+24\%+2.0\%}_{-22\%-2.1\%}$	$181^{+20\%+8.1\%}_{-20\%-6.0\%}$
$\sigma_{\mathrm{MC@NLO}}/\sigma_{\mathrm{NLO}}$	1.01	1.07	1.28

Large enhancement (∼30%) in Higgs region from double $g \to b\bar{b}$ splittings





One $g \to b\bar{b}$ splitting from PS

⇒ TH uncertainties related to matching, shower and 4F/5F schemes crucial!

Conclusions

- **Usually, from the experimental perspective, gluon splitting is something you want to get rid of**
  - VH(bb)
  - gg→H(bb)
  - H(bb)H(bb)
  - ttH(bb)
  - …
- **But you must know very well**
  - Very abundantly produced at LHC
  - Enters in most of the analyses with final state b quarks
- **Very difficult to model theoretically**
  - Large uncertainties
- **Useful to calibrate b-tagging**

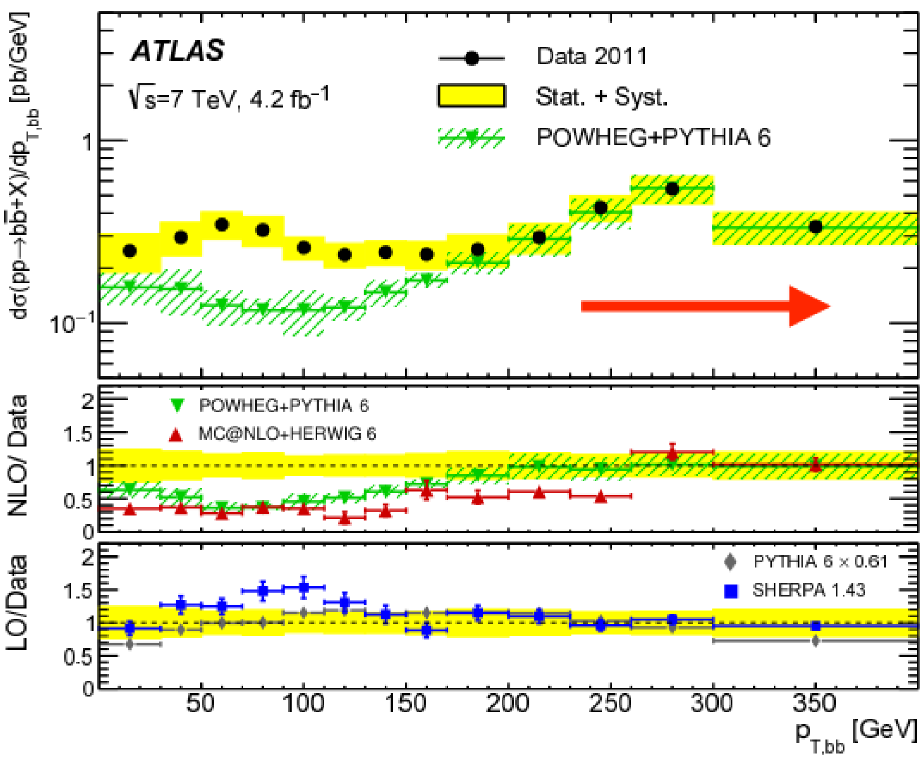
- **We should probably perform more SM measurements in different phase spaces, multi-differential etc**

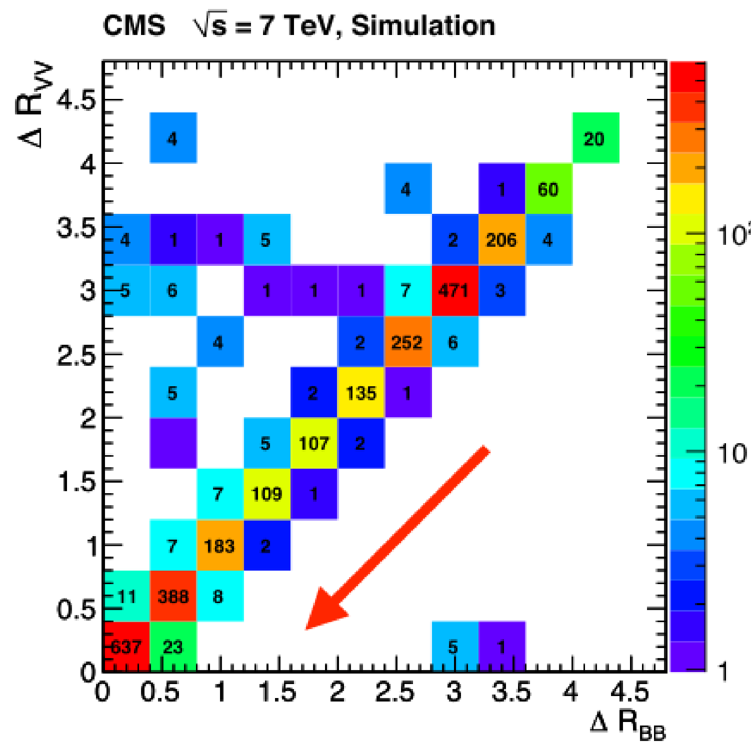
Backup

Partial list of references

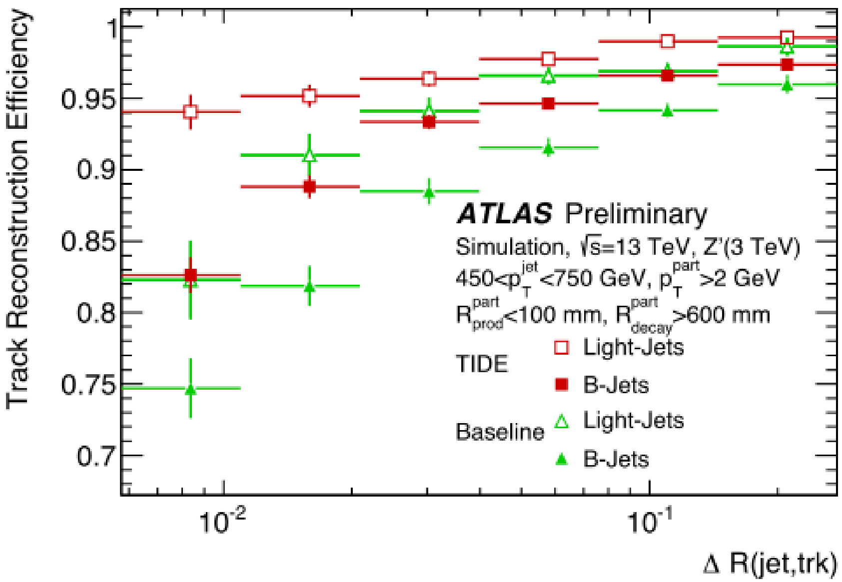
- B. Nachman - Gluon splitting to bottom quarks at the LHC
  - Parton Radiation and Fragmentation from LHC to FCC-ee https://indico.cern.ch/event/557400/timetable/
- J. A. McFayden, D. Napoletano, E. Re - Is there room for improvement in the description of the gluon splitting to heavy quarks?
  - Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report https://arxiv.org/abs/1803.07977
- R. M. Ralich - Study of b Quark Pair Production Mechanisms in pp Collisions with the CMS Experiment at LHC
  - PhD thesis https://cdsweb.cern.ch/record/1311216

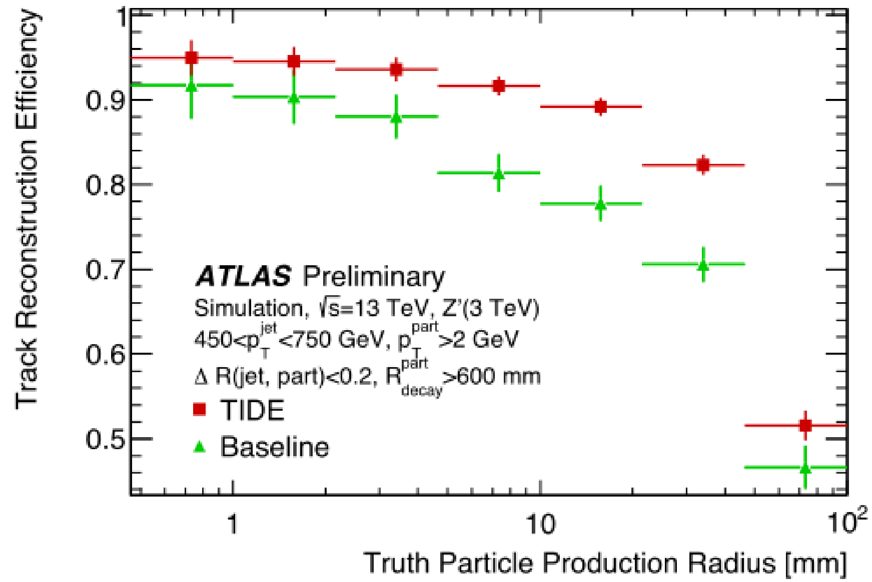
What does it look like with more data, higher p_T, and state-of-the-art simulation?





Note: in order to maintain b-tagging performance, it is critical to have dedicate methods for tracking inside jets



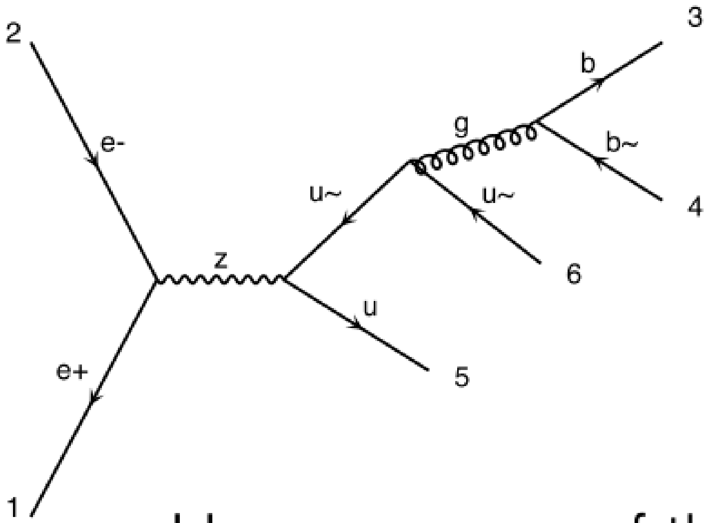


Algorithmic improvements are (much) cheaper than hardware ones - it is important to optimize performance when designing a new detector!

>> Similar studies in CMS

At a lepton collider, we would have an experimentally and theoretically clean environment for studying g->bb

>> no pileup, UE, MPI, etc. <<



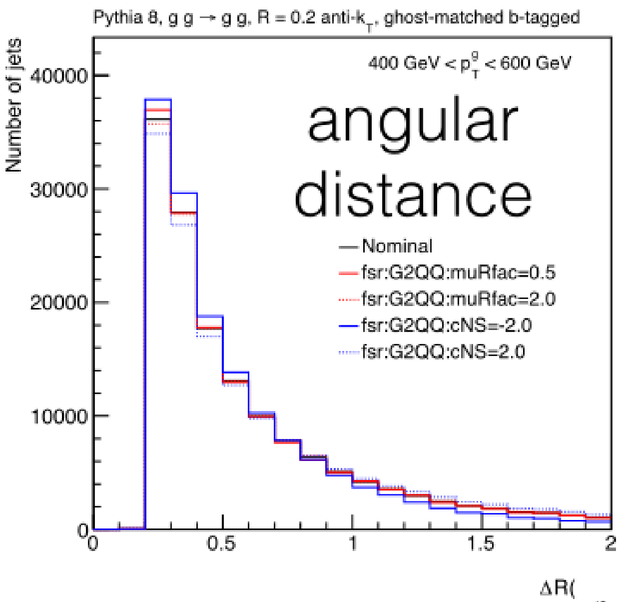
LEP measurements limited to inclusive rates of g->bb

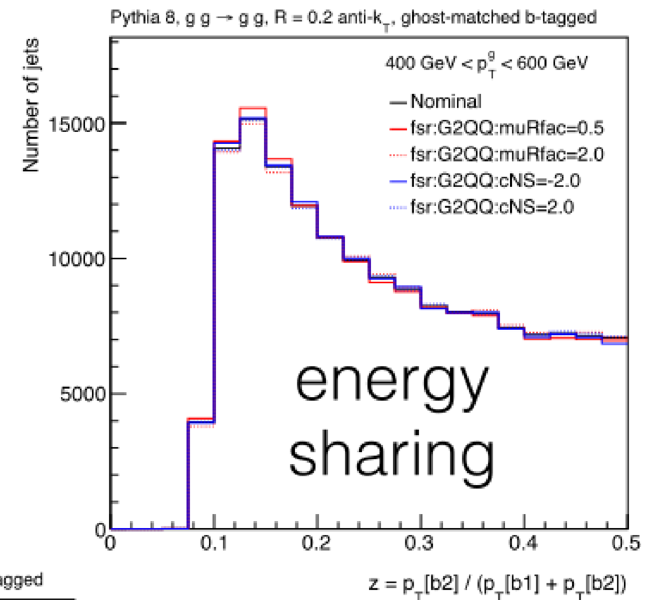
can probe pQCD to high precision by measuring properties of the splitting

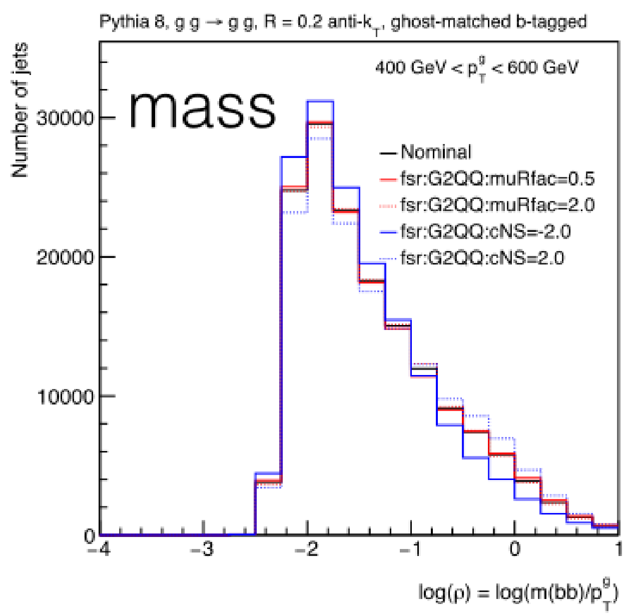
However, one of the most exciting prospects of FCC-ee is to perform a series of comparative measurements of g->bb, Z->bb, and H->bb

also, critical input for the Higgs at FCC-hh!

What else can and should we measure?

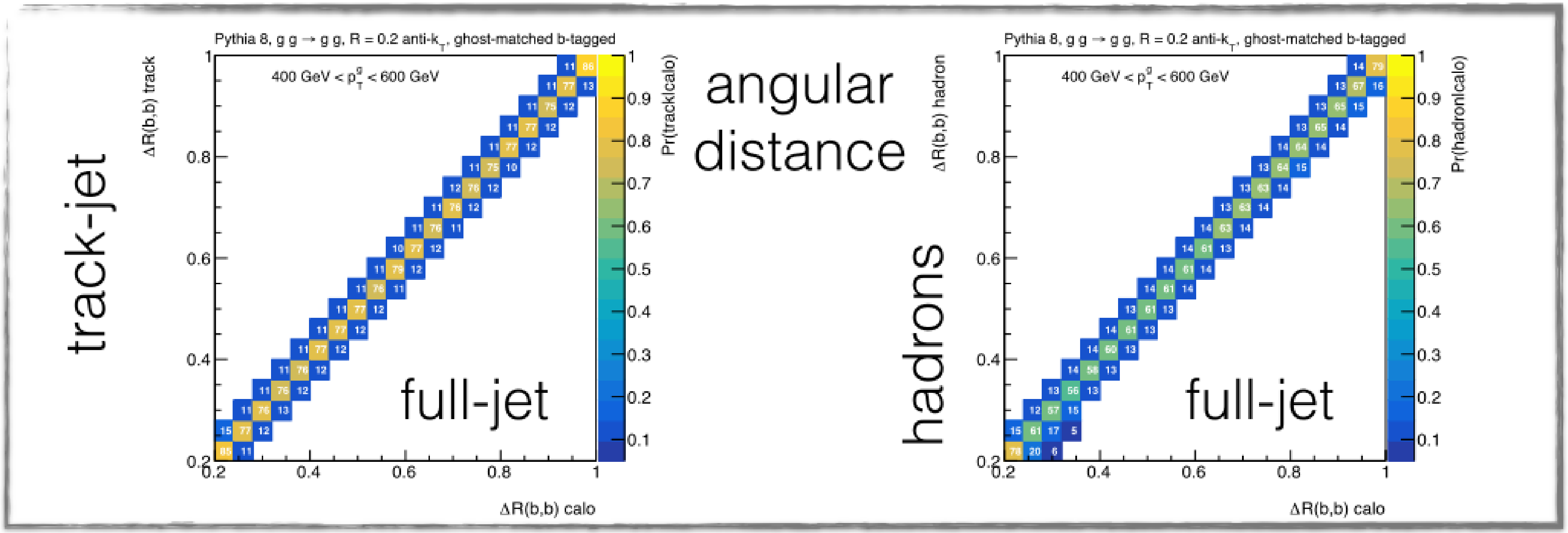


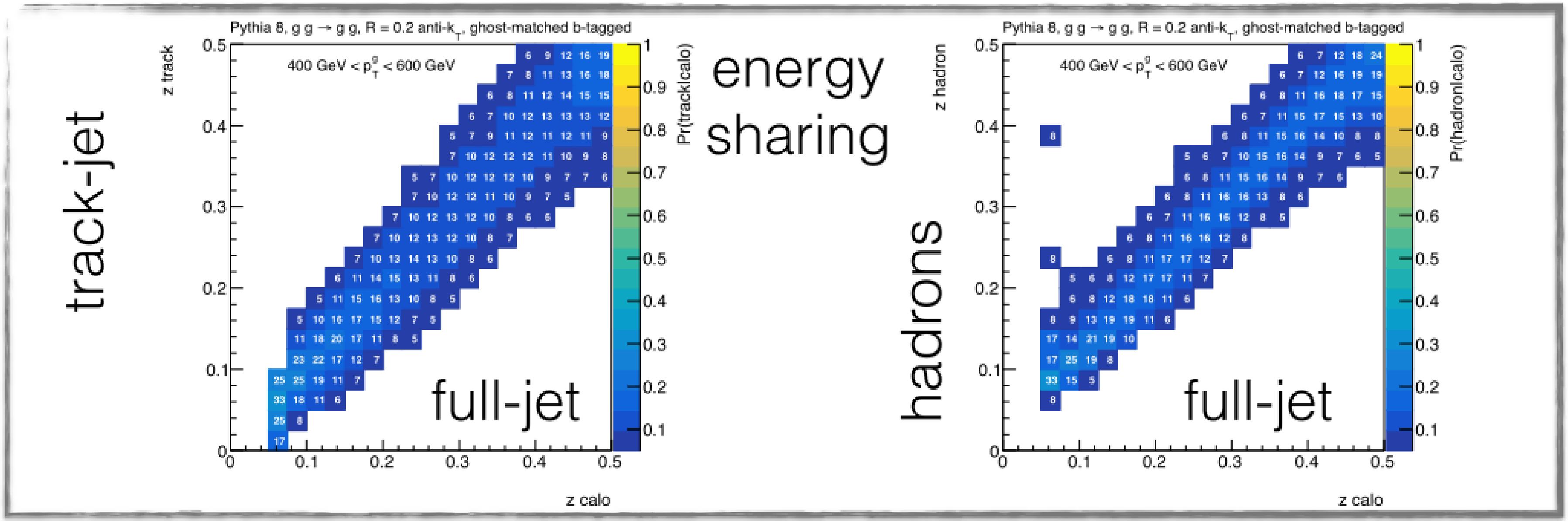




all the lines: tuning nobs in Pythia to see the sensitivity to the modeling of g->bb

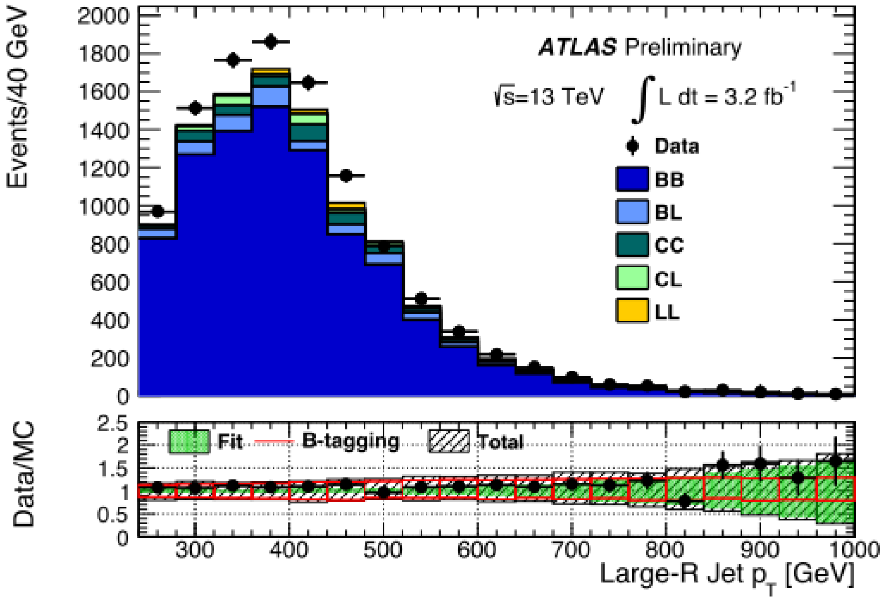
Caution: do we measure partons, hadrons, or (track) jets?

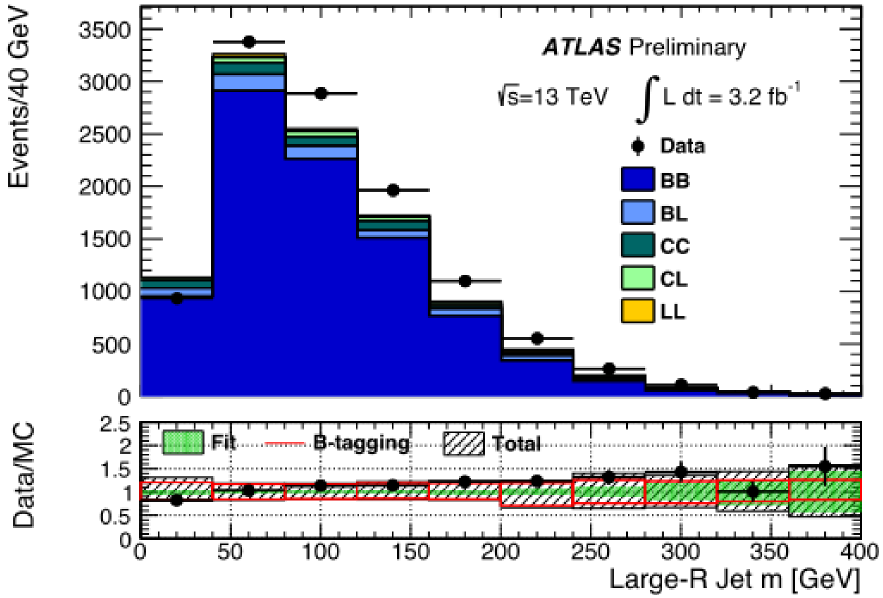




ATLAS g→bb performance

R = 0.2 anti-k_t track jets





With the full Run 1 and Run 2 datasets, there are plenty of gluon jets for studying the modeling of double b-tagging

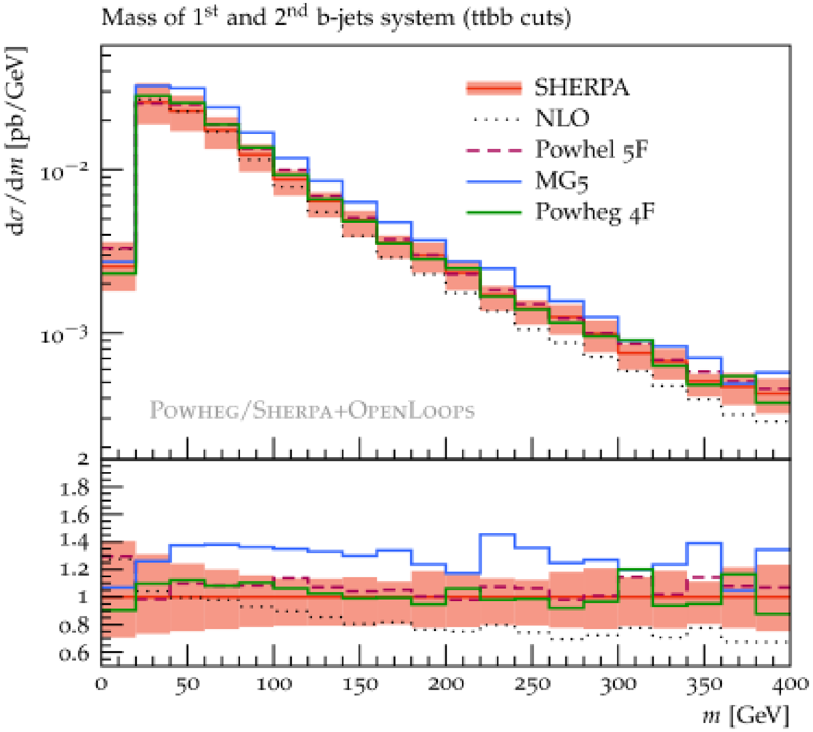
(in this case, use muons to increase purity)

[7 TeV double b-tagger] [8 TeV analysis]

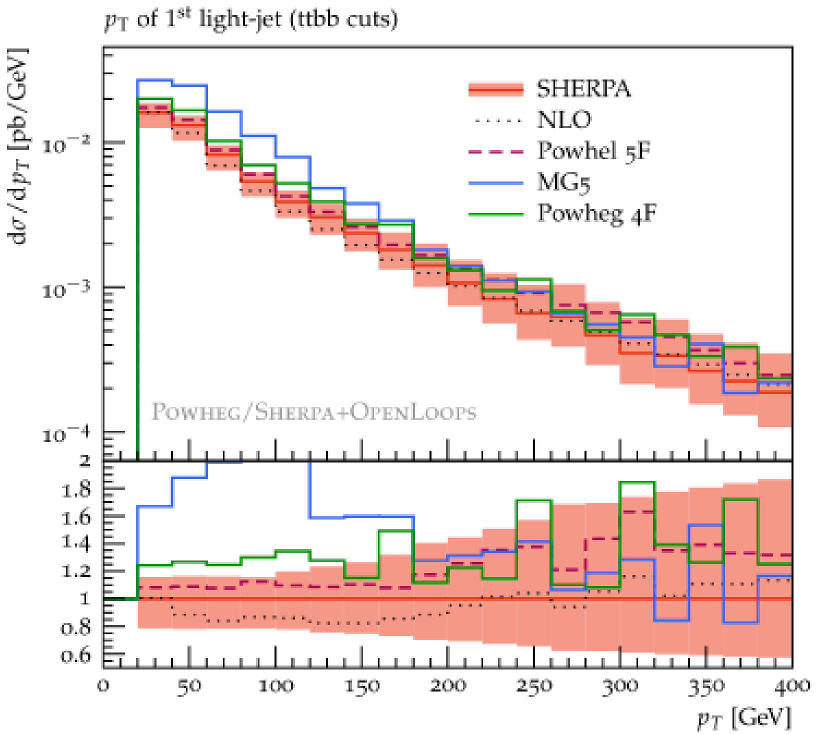
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2016-039/

Further Powheg+Pythia8 results

m_{bb} with ttbb cuts



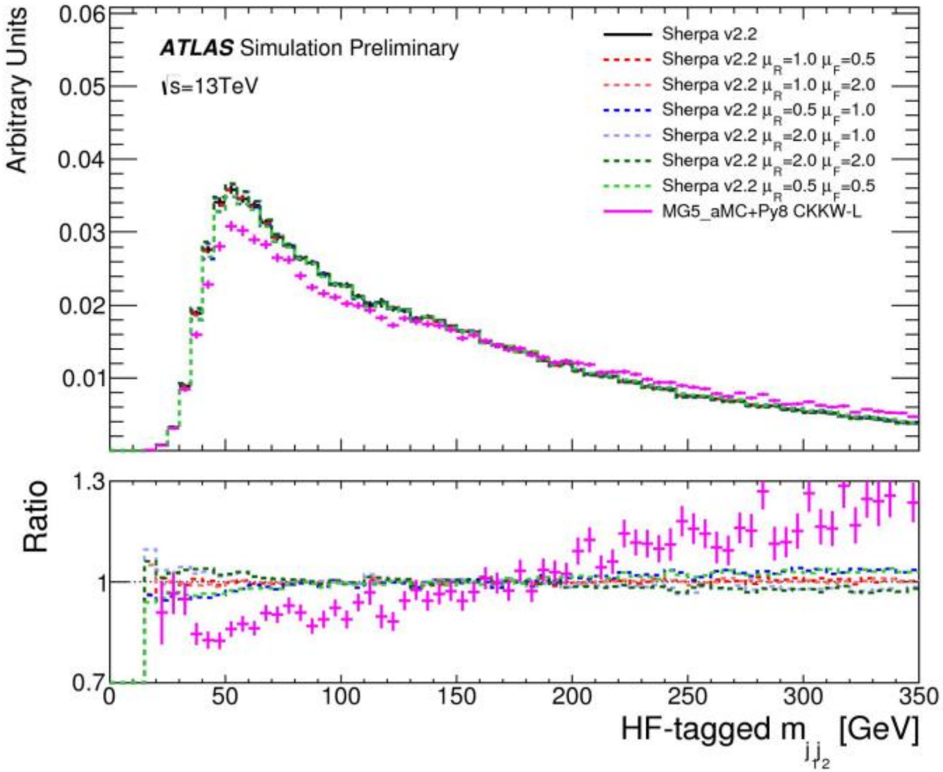
p_{T,j_1} with ttbb cuts



- "double $g \to b\bar{b}$ splittings" confirmed also by Powheg+PY8
- Powheg+PY8 features enhancement in same direction as MG5+PY8
- but no strong distortion of spectrum

V+jets background modeling

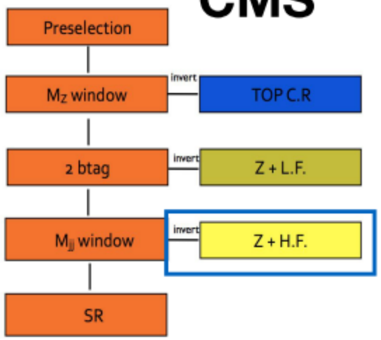
ATLAS PUB note on V+jets modeling and MC simulation
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-006/



- **selection close to nominal VH(bb) analysis regions**
- no W+hf CR/SR separation

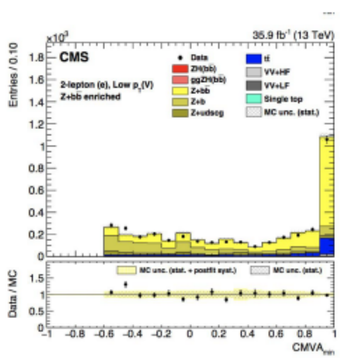
Z+heavy flavors (0-/2-lepton channel)

CMS



- Define dedicated control region (CR)
- Scale factors applied from CR to Signal Regions (SR)
- Systematic uncertainties fully correlated between CR and SR

Variable	Z+HF
$p_T(jj)$	—
$p_T(V)$	[50,150],>150
$CMVA_{max}$	>$CMVA_T$
$CMVA_{min}$	>$CMVA_L$
N_{aj}	—
$N_{a\ell}$	—
p_T^{miss}	<60
$\Delta\phi(V,jj)$	>2.5
$M(\ell\ell)$	[85,97]
$M(jj)$	$\notin$ [90,150]



ATLAS

- **no dedicated control region for Z+hf**
- no m(bb) window selection applied in the nominal analysis selection

- **m(bb) and pTV** shape systematic derived from data/MC in Z+hf enriched-region
  (2-lepton) x (1-btag)
  (2-lepton) x (2-btag) x (remove events with m(jj) around

- Pre-fit theory modeling uncertainties

	Z + jets
$Z + ll$ normalisation	18%
$Z + cl$ normalisation	23%
$Z + bb$ normalisation	Floating (2-jet, 3-jet)
$Z + bc$-to-$Z + bb$ ratio	30 – 40%
$Z + cc$-to-$Z + bb$ ratio	13 – 15%
$Z + bl$-to-$Z + bb$ ratio	20 – 25%
0-to-2 lepton ratio	7%
m_{bb}, p_T^V	S

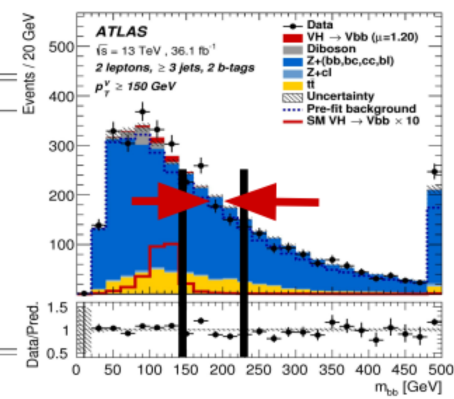


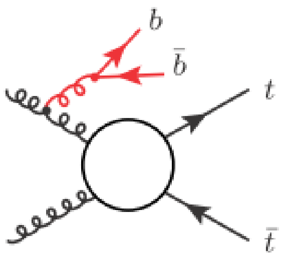
Table 8: The total numbers of events in each channel, for the rightmost 20% region of the event BDT output distribution, are shown for all background processes, for the SM Higgs boson VH signal, and for data. The yields from simulated samples are computed with adjustments to the shapes and normalizations of the BDT distributions given by the signal extraction fit. The signal-to-background ratio (S/B) is also shown.

Process	0-lepton	1-lepton	2-lepton low-$p_T(V)$	2-lepton high-$p_T(V)$
Vbb	216.8	102.5	617.5	113.9
Vb	31.8	20.0	141.1	17.2
V+udscg	10.2	9.8	58.4	4.1
$t\bar{t}$	34.7	98.0	157.7	3.2
Single top quark	11.8	44.6	2.3	0.0
VV(udscg)	0.5	1.5	6.6	0.5
VZ(bb)	9.9	6.9	22.9	3.8
Total background	315.7	283.3	1006.5	142.7
VH	38.3	33.5	33.7	22.1
Data	334	320	1030	179
S/B	0.12	0.12	0.033	0.15

Dominant topologies in 4F ttbb (FS vs IS)

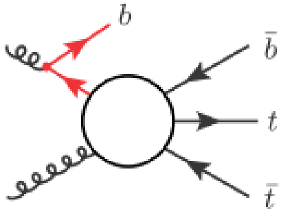
$t\bar{t}b\bar{b}$ **topologies with FS** $g \to b\bar{b}$ **splittings**

- dominant in full ttbb and ttb phase space
- notion of $g \to b\bar{b}$ splittings and IS/FS separation seems ill defined at large ΔR_{bb}, m_{bb}, $p_{T,b}$ due to sizable interferences

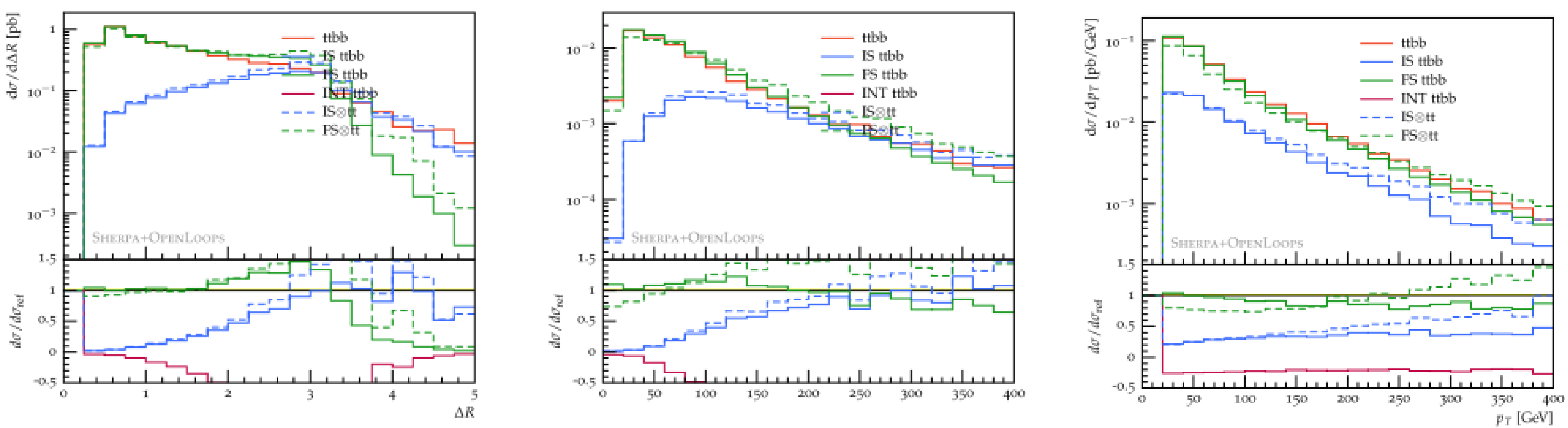


$t\bar{t}b\bar{b}$ **topologies with IS** $g \to b\bar{b}$ **splittings**

- mostly clearly subdominant (no need for 5F scheme resummation)



ΔR_{bb} with ttbb cut | m_{bb} with ttb cuts | p_{T,b_1} with ttb cuts



supports choice of 4F scheme with $m_b > 0$ and no b-quark PDF