$t\bar{t}H$ theory activities

Stefano Pozzorini

on behalf of the ttH conveners

13th Workshop of the LHC HXSWG, 14 July 2017

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

Universität Zürich[UZH]

Foreword

- 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 to $t\bar{t}H(b\bar{b})$: a very serious bottleneck of $t\bar{t}H(b\bar{b})$ searches!
- modern tools support automated $t\bar{t}b\bar{b}$ simulations, but $t\bar{t}b\bar{b}$ remains a highly nontrivial multi-particle multi-scale QCD process . . .
- . . . understanding of $t\bar{t}b\bar{b}$ QCD dynamics crucial to assess TH uncertainties

Outline

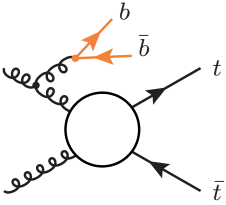
1. 5F vs 4F scheme for $t\bar{t} + b$-jets at NLO

2. $t\bar{t}b\bar{b}$ MC comparisons and open issues

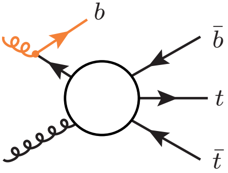
3. New Powheg 4F generator for $t\bar{t} + b$-jets

Option 1: NLOPS $t\bar{t}$ 5F (e.g. Powheg's hvq generator)

$t\bar{t}b\bar{b}$ 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

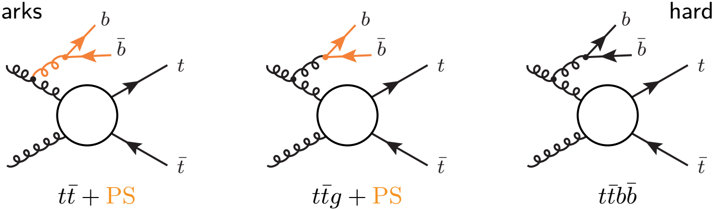
- crucial for more realistic TH uncertainties

Option 2: (N)LO merging $t\bar{t} + 0, 1, 2$ jets 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

softer b-quarks $\Longleftarrow$

harder b-quarks $\Longrightarrow$



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

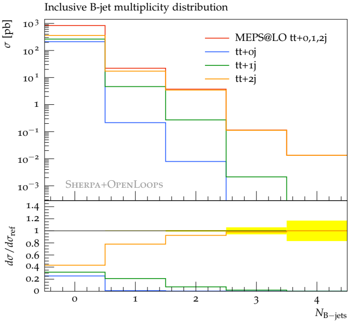
- separates ME regions ($k_T > Q_{\rm cut}$) from shower regions ($k_T < Q_{\rm cut}$)

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

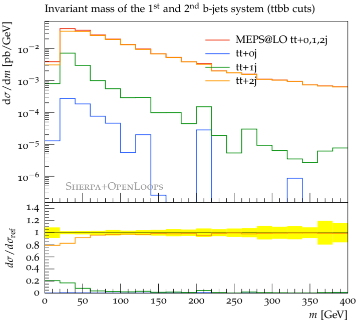
Amount of $t\bar{t}$+jets ME information

$t\bar{t} + 0, 1, 2$ jet LO merging with $Q_{\text{cut}} = 20\,\text{GeV}$

$N_{b\,\text{jets}}$



m_{bb} with ttbb cuts



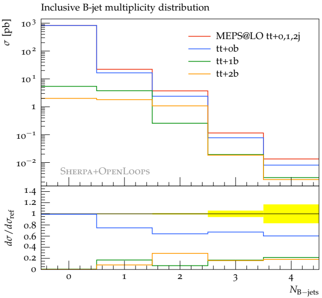
Observables with ≥ 1 additional b-jets

- dominated by $t\bar{t} + 2$ jet MEs (suggesting ME precision) ...

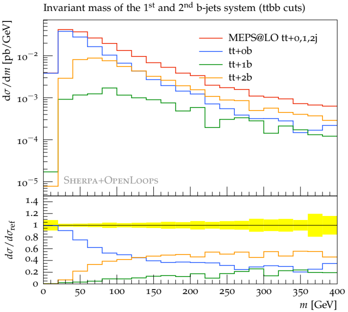
Amount of $t\bar{t} + b$-jets ME information

$t\bar{t} + 0, 1, 2$ jet LO merging with $Q_{\mathrm{cut}} = 20\,\mathrm{GeV}$

$N_{b\,\mathrm{jets}}$



m_{bb} with ttbb cuts

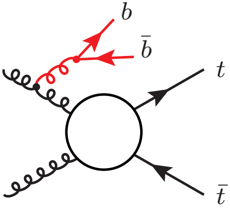


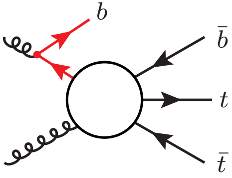
Observables with ≥ 1 additional b-jets

- actually dominated by MEs with 2 light jets and no b-jets (up to $Q \sim 100\,\mathrm{GeV}$)!

$\Rightarrow$ direct description in terms of $t\bar{t}b\bar{b}$ MEs seems preferable

$t\bar{t}b\bar{b}$ 4F at NLO





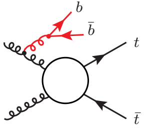
4F $t\bar{t}b\bar{b}$ MEs with $m_b > 0$ cover full b-quark phase space

- 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)?

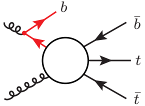
Dominant topologies in 4F $t\bar{t}b\bar{b}$ (FS vs IS $g \to b\bar{b}$)

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



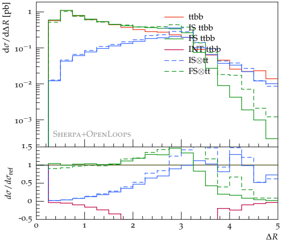
- dominant in full ttbb and ttb phase space
- notion of $g \to b\bar{b}$ splittings and IS/FS separation seems ill defined at large ΔR_{bb}, m_{bb}, $p_{T,b}$ due to sizable interferences

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

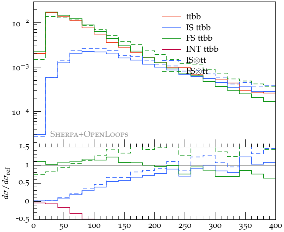


- mostly clearly subdominant (no need for 5F scheme resummation)

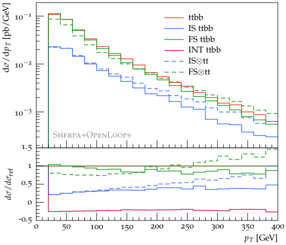
ΔR_{bb} with ttbb cut



m_{bb} with ttb cuts

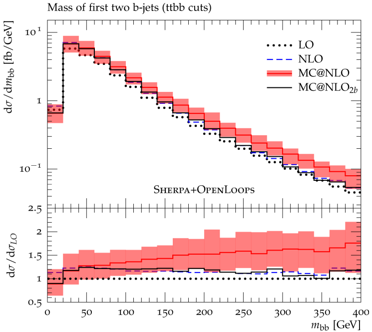


p_{T,b_1} with ttb cuts



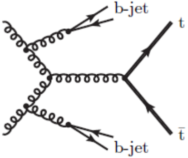
supports choice of 4F scheme with $m_b > 0$ and no b-quark PDF

NLOPS $t\bar{t}b\bar{b}$ 4F with SHERPA+OPENLOOPS [Cascioli et al '13]



Main NLOPS features to keep in mind

- similarly mild NLO K-factors for ttbb and ttb observables
- large matching/shower effects in Higgs region ($\sim$ 30%)
- due to double $g \to b\bar{b}$ splittings (one splitting from PS!)



⇒ TH uncertainties related to matching & shower crucial!

Outline

Tuned comparison of NLO+PS $t\bar{t}b\bar{b}$ simulations at 13 TeV

Different NLOPS methods, showers, and m_b treatments

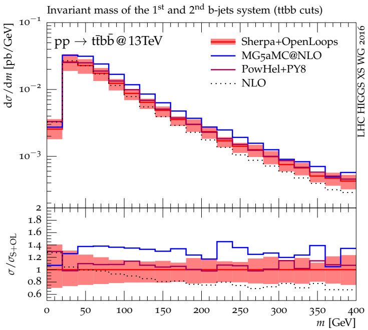
Tool	Matching	Shower	m_b[GeV]	gencuts
SHERPA2.1+OPENLOOPS	SMC@NLO	Sherpa 2.1	4.75 (4F)	no
MG5_aMC@NLO	MC@NLO	Pythia 8.2	4.75 (4F)	no
POWHEL	Powheg	Pythia 8.2	0 (5F)	$p_{T,b} > 4.75\,\text{GeV}$ $\frac{m_{bb}}{2} > 4.75\,\text{GeV}$

note: heuristic implementation of m_b effects in Powhel

Main idea: NLOPS parton level w.o. top decays and hadronisation

- transparent picture of key QCD dynamics and uncertainties
- bias from neglecting hadronisation effects is relatively small in Sherpa/MG5 comparison and zero in Powhel+PY8/MG5+PY8 comparison (see backup slides)

$t\bar{t}b\bar{b}$ distributions with $\geq 2b$-jets



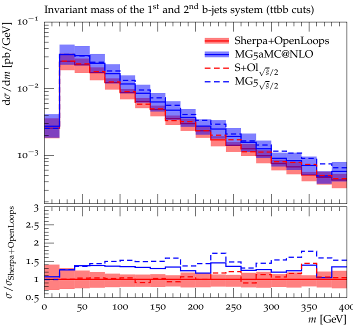
Sherpa+OpenLoops vs PowHel+PY8

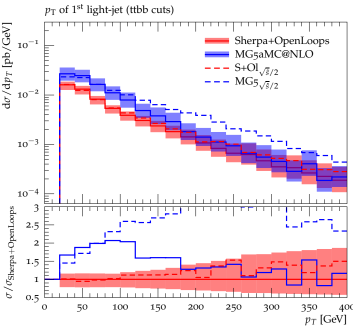
- well consistent also in observables that receive significant shower corrections
- confirmation of "double-splitting effects" (see e.g. m_{bb})

Sherpa+OpenLoops vs MG5aMC@NLO

- 40% enhancement of $t\bar{t} + 2$b XS & sizable differences in NLO radiation pattern
- related to strong sensitivity to resummation scale (shower starting scale) in MG5 . . .

Dependence on resummation scale μ_Q





Nominal MG5_aMC and Sherpa+OpenLoops predictions in YR4

- MG5_aMC supports only* $\mu_Q = f(\xi)\sqrt{\hat{s}}$ $\Rightarrow$ smearing function restricted to $0.1 < f(\xi) < 0.25$ to mimic recommended $\mu_Q = H_T/2$ implemented in Sherpa

μ_Q variations enhance the discrepancy

- $\mu_Q = \sqrt{\hat{s}}/2$ in Sherpa to mimic MG5_aMC default choice $0.1 < f(\xi) < 1$
- strong μ_Q-sensitivity of MG5_aMC $\Rightarrow$ much more pronounced deviations

* Ongoing studies with new MG5 version supporting $H_T/2$. See talks by Zaro & Neu.

Outline

$t\bar{t}b\bar{b}$ 4F with Powheg+OpenLoops [Jezo, Lindert, S.P., in preparation]

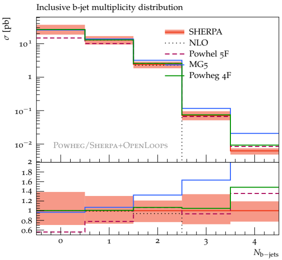
First $t\bar{t}b\bar{b}$ 4F Powheg simulation with $m_b > 0$

- consistent comparison against Sherpa+OpenLoops and MG5aMC@NLO possible

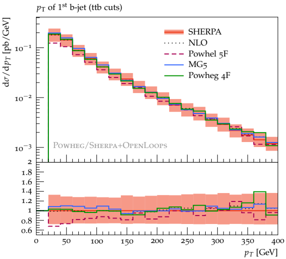
NLOPS Powheg+PY8 results with YR4 settings (preliminary)

- improved agreement with NLO wrt Powhel (especially for ttb cuts)
- good agreement with Sherpa with $ttbb$ and ttb cuts
- confirms tension with MG5 with $ttbb$ cuts

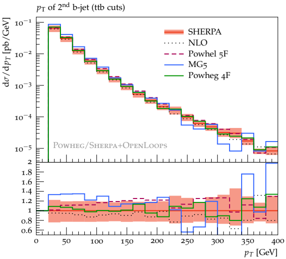
$N_{b\,\mathrm{jets}}$



p_{T,b_1} with ttb cuts*



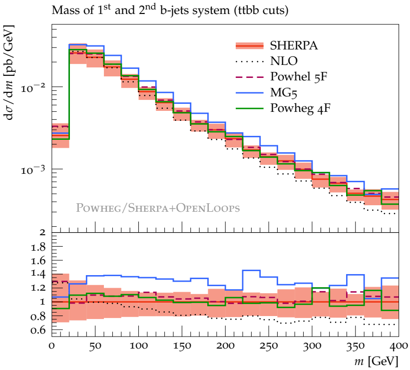
p_{T,b_2} with ttbb cuts



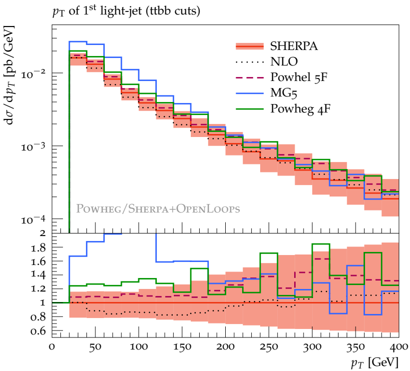
*note excellent consistency of all NLOPS 4F predictions with ttb cuts

Further Powheg+PY8 results

m_{bb} with ttbb cuts



p_{T,j_1} with ttbb cuts

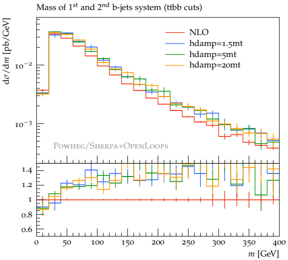


- "double $g \to b\bar{b}$ splittings" confirmed also by Powheg+PY8

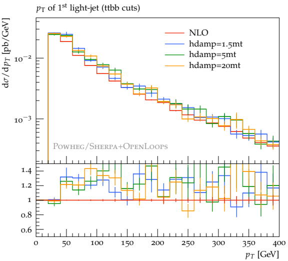
- Powheg+PY8 features enhancement in same direction as MG5+PY8
- but no strong distortion of spectrum

Stability of Powheg+PY8 wrt hdamp (very preliminary)

m_{bb} with ttbb cuts



p_{T,j_1} with ttbb cuts



Idea: compare hdamp in Powheg vs μ_Q in MC@NLO

- since both scales used to separate 1st emission into hard region (LO ME) and soft region (Sudakov resummation + local K-factor)

Weak hdamp dependence in Powheg+PY8

- probably because 1st Powheg emission 100% from ME (which also dictates scalup!)

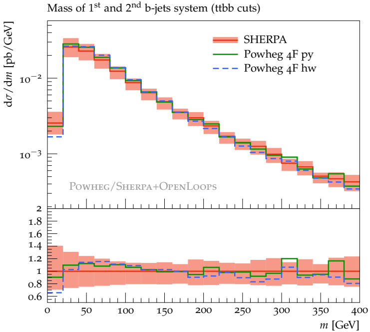
Strong μ_Q dependence in MG5aMC@NLO+PY8

- probably because 1st MC@NLO emission matched to PY8 below μ_Q=scalup

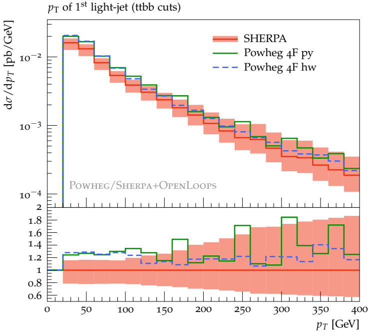
⇒ matching or PY8 issue? (cf. small μ_Q dependence in Sherpa)

Shower dependence of Powheg 4F (very preliminary)

m_{bb} with ttbb cuts



p_{T,j_1} with ttbb cuts



⇒ Negligible difference between Powheg+Pythia 8.2.1.0 vs Powheg+Herwig 7.1.0*

*Pythia with YR4 settings; Herwig with angular ordering and default settings (apart from YR4 top/bottom masses and PDFs)

Conclusions and Outlook

NLOPS simulations of $t\bar{t}+b$-jet production

- 4F scheme preferable since less sensitive to $g \to b\bar{b}$ splittings wrt 5F
- YR4 MC comparisons have revealed significant matching/shower dependence
- reliable estimate of TH uncertainty requires further in-depth studies (ongoing)
- can now be addressed with three independent $t\bar{t}b\bar{b}$ 4F generators: Sherpa, MG5aMC@NLO, Powheg (new!)

Todo: realistic estimates of $t\bar{t}b\bar{b}$ MC uncertainties in EXP analyses

- require extension of MC studies to particle level and detailed framework for TH uncertainties (matching, shower, hadronisation, . . .)
- try to exploit possible sinergies between $t\bar{t}H$ and $t\bar{t}b\bar{b}$ measurements

Todo: identify and address further TH priorities in $t\bar{t}H/tH$ searches

- $t\bar{t}H(WW)$ backgrounds (and signal modelling)?

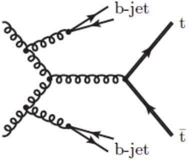
Backup slides

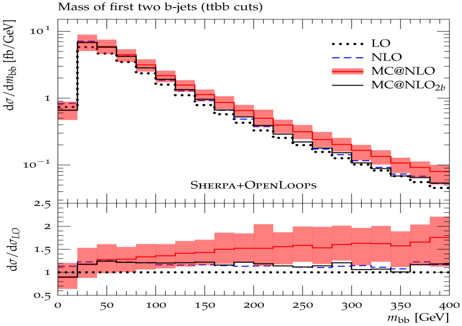
NLOPS $t\bar{t}b\bar{b}$ 4F with SHERPA+OPENLOOPS [Cascioli et al '13]

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!

Hadronisation effects in $t\bar{t}b\bar{b}$ MC comparisons

Motivation of theory studies w.o. top decays and hadornisation

- top decays are trivial (well understood EW interactions) but render the analysis of b-quark production in $WWb\bar{b}b\bar{b}$ final states quite cumbersome
- switching off top decays is very useful in order to investigate the QCD dynamics of b-production in $pp \to t\bar{t}b\bar{b}$ (which dominates TH uncertainties!)
- since top quarks carry SU(3) charge, also hadronisation needs to be switched off

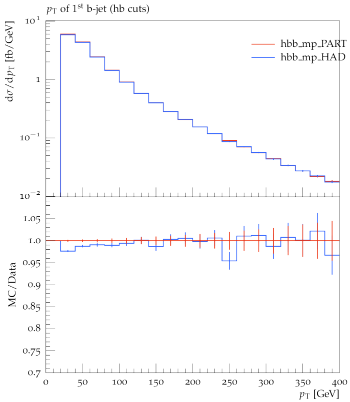
Possible bias of MC comparisions?

- switching off hadronisation could bias comparisons of different showers (Pythia, Sherpa, Herwig) due to dependencies on unphysical dependences (e.g. IR cutoff)
- irrelevant for Powheg+PY8 vs MG5+PY8 comparison (same shower)
- for Sherpa vs MG5+PY8 we have assessed this effect comparing LOPS simulations of $H + b$-jet production (as proxy of $t\bar{t}b\bar{b}$ production) finding non-negligible but rather small hadronisation effects wrt the observed differences in $t\bar{t}b\bar{b}$ production

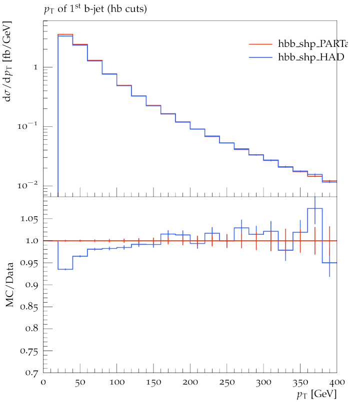
see https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LOpphHadronisation

Hadronisation corrections in MG5+PY8 vs Sherpa2.1

MG5+Pythia 8.2



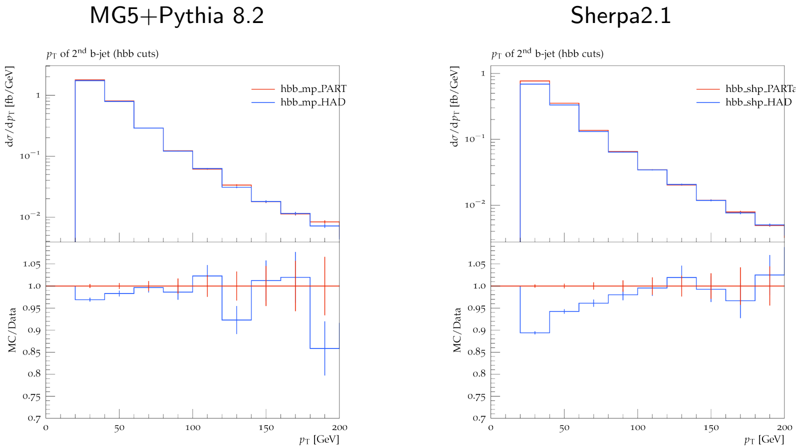
Sherpa2.1



p_T of leading b-jet in $H+\geq 1b$ selection

- moderate hadronisation corrections (up to 6%) in the soft region
- differences below 3–4%

Hadronisation corrections in MG5+PY8 vs Sherpa2.1



p_T of 2nd b-jet in $H+\geq 2b$ selection

- hadronisation corrections up to 10% in the soft region
- differences up to 7%

Combination of $t\bar{t}+X$ and $t\bar{t}b\bar{b}$ MC samples

NLOPS 4F $t\bar{t}b\bar{b}$ sample

- can be applied in its full phase space (no generation cuts)

$\Rightarrow$ inclusive description of $t\bar{t}+\geq 1b$-quarks

- includes also contributions corresponding to $gb \to t\bar{t}b$ in the 5F scheme

Inclusive $t\bar{t}+X$ sample

- needs to be restricted to $t\bar{t}+0\,b$-quarks to avoid double counting

$\Rightarrow$ veto events containing b-quarks not arising from showered top decays or MPI or UE

Possible implementations

- $t\bar{t}+X$ and $t\bar{t}b\bar{b}$ samples independent samples
- reweighting of $t\bar{t}+X$ sample through $t\bar{t}b\bar{b}$ in the $t\bar{t}+\geq 1b$-quarks region

$t\bar{t}b\bar{b}/t\bar{t}+X$ combination: refinement at small $p_{T,b}$

Caveat

- $t\bar{t}b\bar{b}$ sample yields (small) contribution to $t\bar{t}+0\,b$-jet categories of EXP analysis
- $t\bar{t}+0\,b$-jet categories (dominated by $t\bar{t}$+gluons/light-quarks) can bias $t\bar{t}b\bar{b}$ fit

$\Rightarrow$ preferable to restrict $t\bar{t}b\bar{b}$ to $t\bar{t}+b$-jet categories

Proposal: smooth matching of $t\bar{t}+X$ and $t\bar{t}b\bar{b}$ samples

- using smearing function of leading b-jet p_T, such as

$$\xi(p_{T,b}) = \begin{cases} 0 & \equiv \text{ pure } t\bar{t}+0b & \text{for} \quad p_{T,b} < p_{T,\min} \\ \frac{1}{2}\left[1-\cos\left(\pi\frac{p_{T,b}-p_{T,\min}}{p_{T,\max}-p_{T,\min}}\right)\right] & & \text{for} \quad p_{T,\min} < p_{T,b} < p_{T,\max} \\ 1 & \equiv \text{ pure } t\bar{t}+\geq 1b & \text{for} \quad p_{T,b} > p_{T,\max} \end{cases}$$

- with transition region in the vicinity of experimental b-jet threshold, e.g. $[p_{T,\min}, p_{T,\max}] = [15, 25]$ GeV
- same matching procedure should be used in ATLAS and CMS for a transparent comparison and combination of EXP results

Scale choices (YR4) and uncertainties (no proposal yet)

Factorisation (μ_Q) and resummation (μ_Q) scales

$$E_{T_i} = \sqrt{m_i^2 + p_{T,i}^2}$$

$$\mu_F = \mu_Q = \frac{H_T}{2} = \frac{1}{2} \sum_{i=t,\bar{t},b,\bar{b}} E_{T,i}$$

$\mu_Q \equiv$ shower starting scale is a free paramater in MC@NLO (not in Powheg)

CKKW-like (softer) renormalisation scale

$$\mu_R = \mu_{\mathrm{CKKW}} = \prod_{i=t,\bar{t},b,\bar{b}} E_{T,i}^{1/4}$$

Scale variations (leading uncertainty) ~20-30%

- factor-2 variations of μ_R and μ_F ⇔ normalisation
- "kinematic" variations of μ_R, μ_F, μ_Q ⇔ shape
- variations of μ_Q in MC@NLO and h_{damp} in Powheg ⇔ NLOPS matching

Other variations

- PDF variations (only few percent)
- shower variations: tune variations, shower recoil scheme, . . .

Correlation of TH uncertainties between categories

Categories

- $t\bar{t}h(b\bar{b})$ analyses based on simultaneous fit of MC to data in various categories with different # of light- and b-jets
- correlations crucial to constrain background in signal region (with multiple b-jets)

Between $t\bar{t}$+light-jet and $t\bar{t} + b$-jet categories

- uncertainties should be uncorrelated

Between sub-categories (e.g. ttb, $ttbb$, ttB)

- uncertainties should be correlated

Motivation: independent shower, matching and ME variations account for different types of uncertainties (e.g. related to collinear $g \to b\bar{b}$ splittings or hard b-production) $\Rightarrow$ no need of separate categories with uncorrelated uncertainties