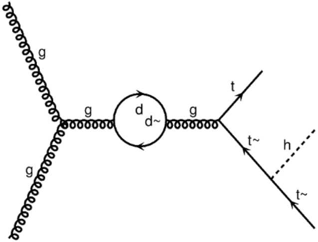
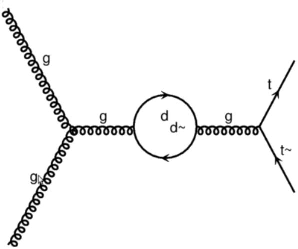
Report from ATLAS and CMS on $t\bar{t}$+b-jet background to $t\bar{t}H(bb)$

Common meeting on $t\bar{t}$+b-jet backgrounds to $t\bar{t}H(bb)$

Marco A. Harrendorf, Maria Moreno Llacer, Stefan Guindon | February 6th, 2017

INSTITUT FÜR EXPERIMENTELLE KERNPHYSIK (IEKP)





Outline of meeting

High priorities in the $t\bar{t}H(bb)$ community

- Handling of systematic uncertainties (especially treatment of $t\bar{t}$+HF uncertainties).
- Description of $t\bar{t}$+HF by event generators with major focus on a comparison of different event generators for the production of an exclusive $t\bar{t}+b\bar{b}$ sample.

Outline of today's meeting

- Summarizing the efforts and approaches towards $t\bar{t}$+b background estimation for the $t\bar{t}H(bb)$ search at ATLAS and CMS (my talk).
- Presenting a proposal/recommendation for treating systematic $t\bar{t}+b\bar{b}$ uncertainties in a common way in the future (Stefano's talk).

Monte Carlo samples

Common

- Nominal sample: Inclusive 5FS POWHEG Box V2 $t\bar{t}$ sample.

ATLAS: Until ICHEP 2016

- Nominal sample: POWHEG Box V2 +Pythia6 and CT10 PDF set.
- Re-weighted to match predictions of 4FS Sherpa+OpenLoops $t\bar{t}$+$b\bar{b}$ sample.

ATLAS: New baseline

- Nominal sample: POWHEG Box V2 +Pythia8 with hdamp=1.5·m(top) and NNPDF3.0 PDF set.

CMS

- Nominal sample: POWHEG Box V2 +Pythia8 and NNPDF3.0 PDF set.

MC systematics: ATLAS new baseline

Nominal sample: POWHEG Box V2 + Pythia8

- hdamp=1.5· m(top) and NNPDF3.0 PDF set
- Including event weights to evaluate uncertainties due to scale, PDF, and hdamp choices

Matrix element generator comparisons

- MG5aMC(NLO) + Pythia8 vs. POWHEG Box V2 + Pythia8

Parton shower and hadronization comparisons

- POWHEG Box V2 + Herwig7 vs. POWHEG Box V2 + Pythia8

Tune variations

- A14 eigentunes for Pythia8

MC systematics: CMS

Karlsruher Institut für Technologie

Nominal sample: POWHEG Box V2 + Pythia8

- NNPDF3.0 PDF set
- Including event weights to evaluate uncertainties due to scale, PDF. Planned: α_S and hdamp choices

Matrix element generator comparisons

- MG5aMC(NLO) + Pythia8 vs. POWHEG Box V2 + Pythia8 vs . MG5aMC(LO) + Pythia8

Parton shower and hadronization

- Pythia8 with ISR and FSR α_S variations
- CUETP8M2 tune based on Monash tune for Pythia8

Jet definitions

KIT
Karlsruher Institut für Technologie

Common

- anti-kT jet algorithm with R=0.4.

ATLAS

- Particle level: $p_T > 15$ GeV (ensuring $\approx$ 100 % $t\bar{t}$+$b\bar{b}$ reweighting for 15 GeV truth jets that can become 25 GeV reco jets).

CMS

- Particle level: $p_T > 20$ GeV
- Reco level semilepton events: $p_T > 30$ GeV
- Reco level dilepton events: $p_T > 30$ GeV for first 2 dilepton jets, others $p_T > 20$ GeV. (Comparable to CMS reference dilepton $t\bar{t}$ analysis)

B-jet definition

Common

- Based on b-hadrons.

ATLAS

- Only b-hadrons with $p_T > 5$ GeV using ΔR-matching.

CMS

- Every existing b-hadron is used and determined by ghost-matching.

Labelling: $t\bar{t}$+b-jets subcategories

Common

- Using the following (different defined, see next slides) subcategories:
  - exclusive $t\bar{t}$+b.
  - exclusive $t\bar{t}$+$b\bar{b}$.
  - exclusive $t\bar{t}$ +B / $t\bar{t}$+2b (ATLAS / CMS terminology).
- Note: Subcategories based on number of additional b-jets (not from top decay).

ATLAS

- Treating uncertainties as fully correlated.
- Also uses $t\bar{t} + \geqq 3$b category.

CMS

- Treating uncertainties as fully uncorrelated.

Labelling: Further details

Karlsruher Institut für Technologie

	ATLAS	CMS
tt+b-jets	b-hadron (pT > 5 GeV) not from top or W within DR 0.3 of particle jet (pT >15 GeV) Re-weighted to 4F Sherpa+OL calculation	b-hadron not from top or W ghost match particle jet (20 GeV) No re-weighting applied
tt+cc	c-hadron not from top/W within DR 0.3 of particle jet (pT >15 GeV)	c-hadron not from top/W ghost match particle jet (20 GeV)

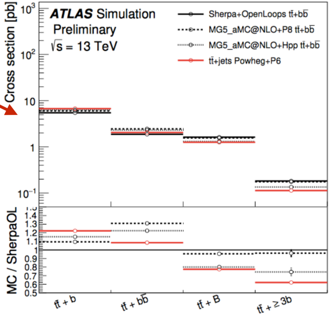
- tt+b -> one extra b-hadron within one truth jet per event
- tt+bb -> one extra b-hadron within one truth jet x2 per event
- tt+B -> two extra b-hadrons within one truth jet per event

Provided by Stefan Guindon and Maria Moreno Llacer

ATLAS Re-weighting procedure

ATLAS: Re-weighting procedure

- Some differences in *tt̄+bb* kinematics between generators
- **Re-weight Powheg+Pythia6 to Sherpa+OpenLoops** * Re-weighted to 4FS due to lack of 4FS reco samples
  - Powheg+Pythia6: default ttbar inclusive generator for ICHEP results
- Inclusive XS for tt+>=1b is kept from Powheg+Pythia6
- Correct normalization of different categories
- **Small correction to most HF categories**
  - Largest in unresolved gluon spitting to bb
- 2D re-weighting:
  - **top/ttbar pT (SherpaOL)**
- 2D re-weighting:
  - **ΔR(bb)/pT(bb)**
  - **OR: q1 pT/η (for tt+b/B)**



Additional plots of tt+>=1b here:
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-016/

Apr 13, 2016 Stefan Guindon University at Albany

Labelling: ATLAS systematics

Systematics are fully correlated across all $t\bar{t} \geq 1b$ categories

- One normalization uncertainty (no prior – fully floating nuisance parameter)
- 4FS systematics from Sherpa+OpenLoops (scales, PDF, MPI/FSR).
- 4FS Generator comparison – comparison of re-weighting to Sherpa+OpenLoops or MG5aMC + Pythia8.
  - Follows from YR4 and WG discussions regarding 4FS comparisons.
  - Same setup for MC samples as in YR4.
  - Larger impact on measurement than 4FS uncertainties of only Sherpa+OpenLoops.
- 5FS and PS residual systematics (alternative samples are re-weighted to Sherpa+OpenLoops and remaining difference is taken as the systematic uncertainty).

Labelling: ATLAS systematics in detail

Systematics

- **Systematics are correlated for all tt+>=1b categories**

Systematic source	How evaluated	$t\bar{t}$ categories
$t\bar{t}$ cross-section	±6%	All, correlated
NLO generator (*residual*)	Powheg-Box + Herwig++ vs. MG5_aMC + Herwig++	All, uncorrelated
Radiation (*residual*)	Variations of μ_R, μ_F, and *hdamp*	All, uncorrelated
PS & hadronisation (*residual*)	Powheg-Box + Pythia 6 vs. Powheg-Box + Herwig++	All, uncorrelated
NNLO top & $t\bar{t}$ p_T	Maximum variation from any NLO prediction	$t\bar{t}+\geq 1c$, $t\bar{t}$ +light, uncorr.
$t\bar{t}+b\bar{b}$ NLO generator *reweighting*	SherpaOL vs. MG5_aMC+ Pythia8	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ PS & hadronis. *reweighting*	MG5_aMC + Pythia8 vs. MG5_aMC + Herwig++	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ renorm. scale *reweighting*	Up or down a by factor of two	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ resumm. scale *reweighting*	Vary μ_Q from $H_T/2$ to μ_{CMMPS}	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ global scales *reweighting*	Set μ_Q, μ_R, and μ_F to μ_{CMMPS}	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ shower recoil *reweighting*	Alternative model scheme	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ PDF *reweighting*	CT10 vs. MSTW or NNPDF	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ MPI	Up or down by 50%	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ FSR	Radiation variation samples	$t\bar{t}+\geq 1b$
$t\bar{t}+c\bar{c}$ ME calculation	MG5_aMC + Herwig++ inclusive vs. ME prediction	$t\bar{t}+\geq 1c$

- **tt+>=1b:**
- **Residual inclusive 5F ttbar+jets uncertainties**
  - Remaining differences after re-weighting to SherpaOL
  - Uncorrelated with tt+light and tt+>=1c
  - *Will change with tuning of generators*
- **4F vs 4F Generator comparison**
  - Re-weighting done to MG5 and not SherpaOL
- **4F Systematics from Sherpa+OL**

Apr 13, 2016 Stefan Guindon University at Albany

Labelling: CMS systematics

Karlsruher Institut für Technologie

Systematics are fully uncorrelated across all $t\bar{t} \geq 1b$ categories

- General inclusive $t\bar{t}$ cross section normalization uncertainty
- General PDF uncertainty for gg initiated processes except $t\bar{t}H$ (e.g. $t\bar{t}$, $t\bar{t}Z$).
- Three independent and uncorrelated uncertainties per $t\bar{t}$+b-jets category:
  - Cross section normalization uncertainty (so far 50 %) for the $t\bar{t}$+b-jets subcategories in addition to inclusive $t\bar{t}$ cross section normalization uncertainty.
  - Renormalization and factorization scale shape uncertainty.
  - Parton shower ISR / FSR shape uncertainty.
- Note: Uncertainties can be either normalization / rate uncertainties or shape uncertainties.

Questions to discuss

MC event samples

- Sizable differences for the NLO 4FS $t\bar{t}+b\bar{b}$ predictions from various MC event generators are observed. How can we understand this difference better? In the meantime, which MC event generator should / could we use for the baseline MC samples?
- 4FS systematics from single event generators are seemingly small (in terms of using re-weighting) compared to 4FS vs. 4FS (MG5aMC vs. Sherpa+OpenLoops) generator comparison - does this make sense?
- Common use of MC event generator and PDF sets: Can we define a common PDF set and MC event generator in ATLAS and CMS which could be used by default for the $t\bar{t}H(bb)$ analysis?
- Should we and how could we integrate dedicated 4FS $t\bar{t}+b\bar{b}$ predictions into the used inclusive 5FS $t\bar{t}$ samples?

Questions to discuss

Systematics

- How do we merge systematics from 5FS and / or 4FS estimations?
- ATLAS was using fully correlated uncertainties, while CMS was using fully uncorrelated uncertainties for the $t\bar{t}$+b-jets subcategories. How (much) should we correlate the subcategories?

Object definition

- Choice of p_T threshold used for additional true b-jets to define $t\bar{t} + \geq 1b$? (ATLAS: $p_T > 15$ GeV / 25 GeV, CMS: $p_T > 20$ GeV / 30 GeV)
- ATLAS currently labels hadronic b-jets wit $p_T \ll$ matched particle jet p_T as true b-jets. Better procedure?

Backup – Reweighting

ATLAS Re-weighting procedure for $t\bar{t}$ sample

1. Applying a kinematic reweighting separately in each of the $t\bar{t} + \geqq 1b$ sub-categories, such that the relative normalisation of the sub-categories and the kinematic distributions match the Sherpa+OpenLoops prediction.
2. In each sub-category, two-dimensional reweighting based on the p_T of the top quark and the p_T of the $t\bar{t}$-system.
3. In the $t\bar{t} + \geqq 3b$ and $t\bar{t}+b\bar{b}$ sub-categories, two-dimensional reweighting of the ΔR between the b-jets and the p_T of the system of b-jets.
4. In the $t\bar{t}$ + B and $t\bar{t}$+b sub-categories, the B or b-jet p_T and η are used.
5. Topologies which are not included in the NLO calculation but are labeled as $t\bar{t} + \geqq 1b$ are not reweighted. These include $t\bar{t}$ events with additional b-jets from multi-parton interactions and b-jets from final-state radiation.

Backup – Labelling

ATLAS $t\bar{t}$+b-jets categorization

1. Take match between b-hadrons not from top with $p_T > 5$ GeV and particle jets (truth jets) with $\Delta R < 0.4$.
2. If this particle jet has $p_T > 15$ GeV and $|\eta| < 2.5$, then the event has an additional b.
3. If there are two such particle jets, matched to two b-hadrons not from top, then the event is labeled "bb".
4. If two b-hadrons not from top are matched to one particle jet, it is labelled "B".
5. In the fit, all b categories are combined and all c categories are combined. No split between $t\bar{t}$+b and $t\bar{t}$+$b\bar{b}$ are explicitly formalized by a NP in the profile likelihood, however, the PS systematics, comparisons of generators and NLO systematics allow for the normalization of $t\bar{t}$+b and $t\bar{t}$+$b\bar{b}$ to vary.

CMS $t\bar{t}$+b-jets categorization

CMS is making use of the GenHFHadronMatcher together with ghost hadron matching:

1. Define kinematic criteria for additional jets: $p_T > 20$ GeV and $|\eta| < 2.4$.
2. Identify b jets containing b hadrons from t$\rightarrow$ Wb (regardless of their kinematics)
3. Create lists of remaining jets that contain b or c hadrons not from t$\rightarrow$ Wb, but originate from before or after top decay.
4. Skip b and c hadrons coming from the W decay. of the t$\rightarrow$ Wb.
5. Skip c hadrons that stem from the decay of a b hadron.
6. Count number of corresponding hadrons in each jet (Jets from overlapping hadrons can have higher b-tag output).
7. Jets are divided into additional (from hadrons before top decay) and pseudoadditional (from hadrons after top decay) jets.
8. Event is categorized based on number of additional/pseudoadditional b/c jets and number of hadrons in each of them.

Systematics

ATLAS systematics

Impact on Analysis I

Uncertainty source	$\Delta\mu$	
$t\bar{t}+\geq 1b$ modelling	+0.53	−0.53
Jet flavour tagging	+0.26	−0.26
$t\bar{t}H$ modelling	+0.32	−0.20
Background model statistics	+0.25	−0.25
$t\bar{t}+\geq 1c$ modelling	+0.24	−0.23
Jet energy scale and resolution	+0.19	−0.19
$t\bar{t}$+light modelling	+0.19	−0.18
Other background modelling	+0.18	−0.18
Jet-vertex association, pileup modelling	+0.12	−0.12
Luminosity	+0.12	−0.12
$t\bar{t}Z$ modelling	+0.06	−0.06
Light lepton (e, μ) ID, isolation, trigger	+0.05	−0.05
Total systematic uncertainty	+0.90	−0.75
$t\bar{t}+\geq 1b$ normalisation	+0.34	−0.34
$t\bar{t}+\geq 1c$ normalisation	+0.14	−0.14
Statistical uncertainty	+0.49	−0.49
Total uncertainty	+1.02	−0.89

- **tt+>=1b modelling dominant systematic**
- Normalization uncertainty still large, **but not leading in Run-2**

Event reconstruction

Full event reconstruction

Karlsruher Institut für Technologie

Common

- Removing the b-jets belonging to top decays based on origin of B-hadron.

ATLAS: Full event reconstruction

- Applying BDT to reconstruct full event simultaneously.

CMS: Top reconstruction

- In most categories: χ^2-based reconstruction.
- Applying matrix element method in almost all categories as a full event likelihood to identify top candidates.

Backup – Generator settings

ATLAS Generator settings

ttbar Generator Settings

ME gen. PS/UE gen.	Powheg-Box Pythia 6.428	Powheg-Box Herwig++2.7.1	MG5_aMC Herwig++2.7.1	Powheg-Box Pythia 6.428	Powheg-Box Pythia 6.428
Ren. scale	$\sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$\sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$\sqrt{m_t^2 + \frac{1}{2}(p_{\mathrm{T,t}}^2 + p_{\mathrm{T,\bar{t}}}^2)}$	$\frac{1}{2} \cdot \sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$2 \cdot \sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$
Fact. scale	$\sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$\sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$\sqrt{m_t^2 + \frac{1}{2}(p_{\mathrm{T,t}}^2 + p_{\mathrm{T,\bar{t}}}^2)}$	$\frac{1}{2} \cdot \sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$2 \cdot \sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$
hdamp	m_t	m_t	–	$2 \cdot m_t$	m_t
ME PDF	CT10	CT10	CT10	CT10	CT10
PS/UE PDF	CTEQ6L1	CTEQ6L1	CTEQ6L1	CTEQ6L1	CTEQ6L1
Tune	P2012	UE-EE5	UE-EE5	P2012 radHi	P2012 radLo

Table 3: Summary of the settings used for the simulation of the inclusive $t\bar{t}$ samples. For the renormalisation and factorisation scales, $p_{\mathrm{T,t}}$ ($p_{\mathrm{T,\bar{t}}}$) indicates the transverse momentum of the top (anti-top) quark in the $t\bar{t}$ centre-of-mass reference frame.

ME gen. PS/UE gen.	MG5_aMC Herwig++ 2.7.1	MG5_aMC Pythia 8.210	SherpaOL Sherpa
Renorm. scale	μ_{CMMPS}	μ_{CMMPS}	μ_{CMMPS}
Fact. scale	$H_{\mathrm{T}}/2$	$H_{\mathrm{T}}/2$	$H_{\mathrm{T}}/2$
Resumm. scale	$f_{\mathrm{Q}}\sqrt{\hat{s}}$	$f_{\mathrm{Q}}\sqrt{\hat{s}}$	$H_{\mathrm{T}}/2$
ME PDF	NNPDF3.0 4F	NNPDF3.0 4F	CT10 4F
PS/UE PDF	CTEQ6L1	NNPDF2.3	
Tune	UE-EE-5	A14	Author's tune

Table 4: Summary of the settings used for the simulation of $t\bar{t} + b\bar{b}$ 4F NLO samples. For the resummation scale, $\sqrt{\hat{s}}$ is the Born-level partonic centre-of-mass energy, while f_{Q} is a random pre-factor allowed to vary in the range [0.1, 0.25].