tt+jets/HF model and systematics in ATLAS and questions for CMS

Benasque Higgs Topping - Breakout Session: tt+jets/HF modeling

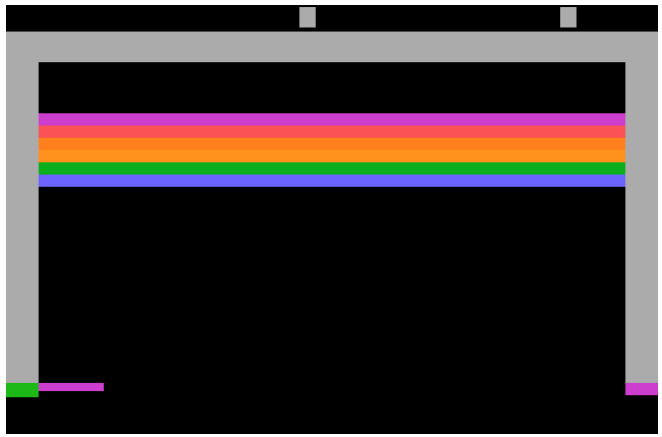
Timothée Theveneaux-Pelzer

HELMHOLTZ GEMEINSCHAFT

Deutsches Elektronen-Synchrotron (DESY) - Zeuthen

DESY

Thursday, May 29th 2018



Atari breakout (1976)

$t\bar{t}+\mathrm{HF}$ categorisation

- $t\bar{t}+\mathrm{jets}$ process subdivided at truth level depending on origin of the additional jets
- First match **not-from-top** hadrons to particle-level jets :

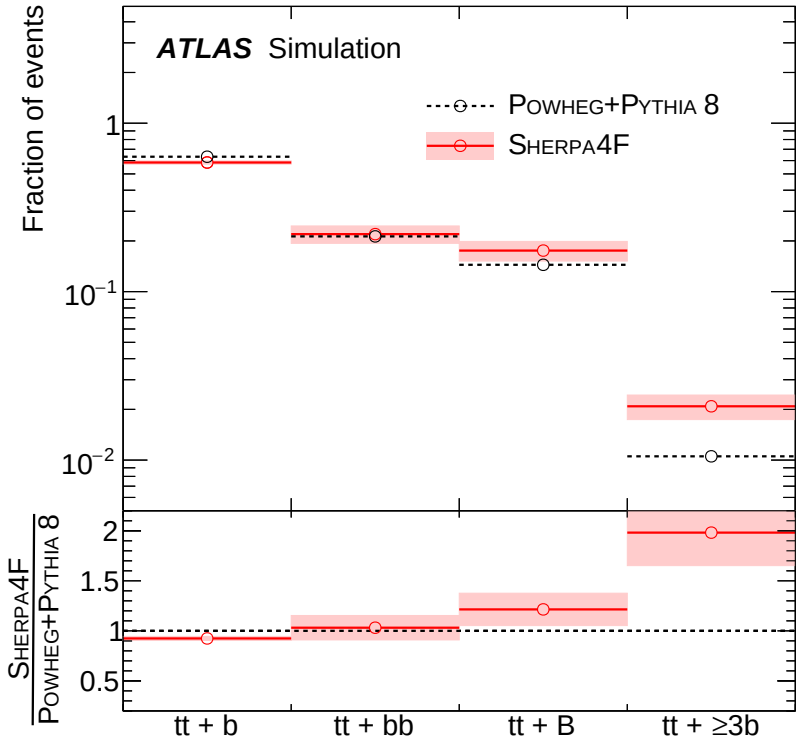
Particle-level jet $\mathrm{p_T}$	>15 GeV
Particle-level jet $\lvert\eta\rvert$	<2.5
Hadron $\mathrm{p_T}$	>5 GeV
Jet-Hadron matching	ΔR<0.3

- Then, count the number of jets matched to not-from-top HF hadrons :

process	# jets	w/ # hadrons
$t\bar{t}+\geq 1b$	≥ 1	$\geq 1b$
$t\bar{t}+b(MPI/FSR)$	all b-jets from MPI/FSR	
$t\bar{t}+b$	$=1$	$=1b$
$t\bar{t}+bb$	$=2$	$=1b$
$t\bar{t}+B$	$=1$	$\geq 2b$
$t\bar{t}+\geq 3b$	other $t\bar{t}+\geq 1b$ events	
$t\bar{t}+\geq 1c$	≥ 1	$\geq 1c$
$t\bar{t}+\mathrm{light}$	other $t\bar{t}+\mathrm{jets}$ events	

Nominal $t\bar{t}+$jets model

- Nominal $t\bar{t}+$jets MC sample : Powheg+Pythia8
  - A14 tune, $\alpha_S(ISR) = \alpha_S(FSR) = 0.127$ (see ATL-PHYS-PUB-2014-021)
  - h_{damp}=1.5 $\mathrm{m_{top}}$ (optimised on 8 TeV data, see ATL-PHYS-PUB-2016-020)
  - normalised to NNLO+NNLL prediction $831.76\ \mathrm{pb}^{-1}$
- $t\bar{t}+\geq 1b$, $t\bar{t}+\geq 1c$ fractions taken from PP8, also when comparing to other samples
  - but are let free-floating in the fit
- $t\bar{t}+\geq 1b$ sub-categories fractions reweighted to Sherpa 4FS $t\bar{t}+b\bar{b}$
  - except $t\bar{t}+b(MPI/FSR)$ which isn't included in Sherpa 4FS $t\bar{t}+b\bar{b}$



Uncertainties on measured $t\bar{t}H$ cross-section

Uncertainty source	$\Delta\mu$	
$t\bar{t}+>1b$ modeling	$+0.46$	-0.46
Background-model stat. unc.	$+0.29$	-0.31
b-tagging efficiency and mis-tag rates	$+0.16$	-0.16
Jet energy scale and resolution	$+0.14$	-0.14
$t\bar{t}H$ modeling	$+0.22$	-0.05
$t\bar{t}+\geq 1c$ modeling	$+0.09$	-0.11
JVT, pileup modeling	$+0.03$	-0.05
Other background modeling	$+0.08$	-0.08
$t\bar{t}+$ light modeling	$+0.06$	-0.03
Luminosity	$+0.03$	-0.02
Light lepton (e,μ) id., isolation, trigger	$+0.03$	-0.04
Total systematic uncertainty	$+0.57$	-0.54
$t\bar{t}+\geq 1b$ normalization	$+0.09$	-0.10
$t\bar{t}+\geq 1c$ normalization	$+0.02$	-0.03
Intrinsic statistical uncertainty	$+0.21$	-0.20
Total statistical uncertainty	$+0.29$	-0.29
Total uncertainty	$+0.64$	-0.61

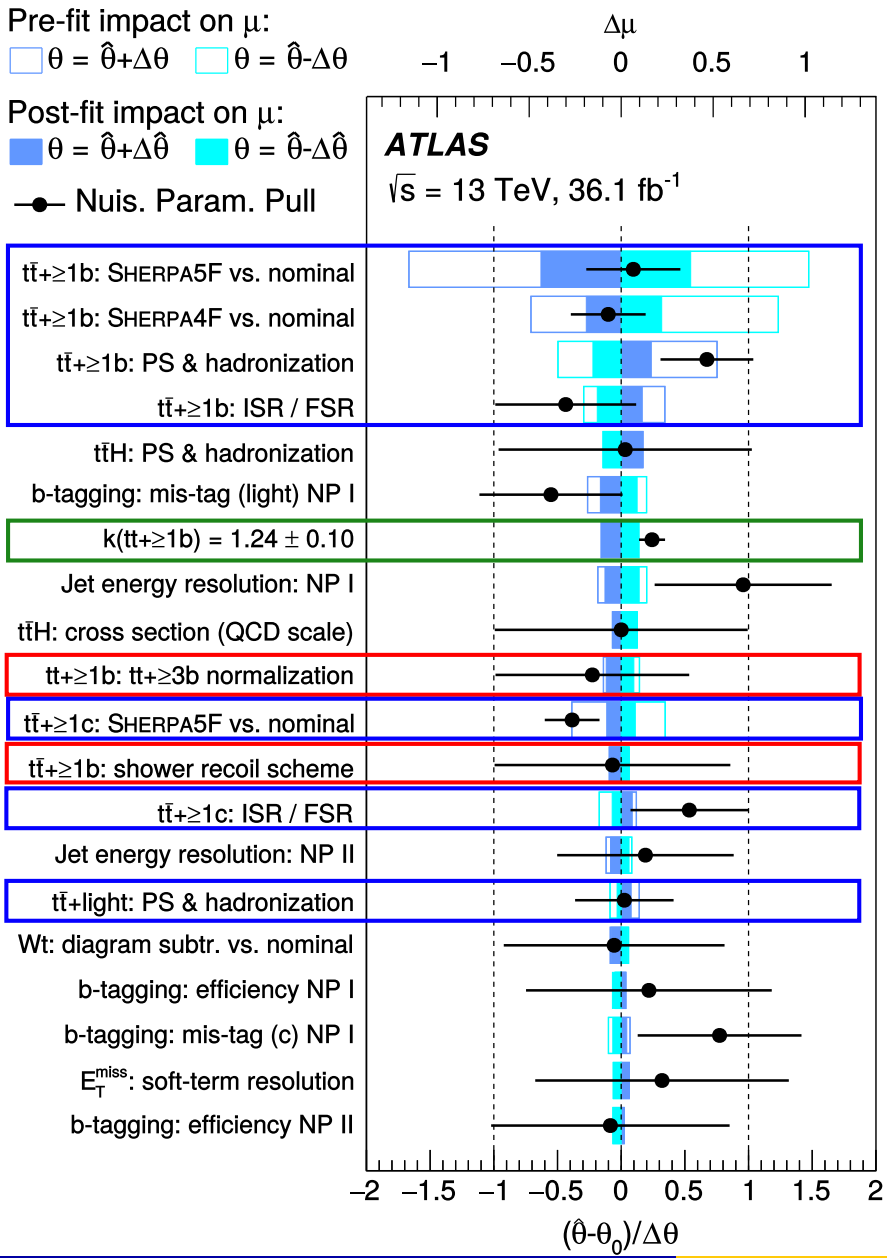
- Uncertainties on $t\bar{t}+\geq 1b$ are the largest (in case you didn't know)
- Also : size (of MC samples) does matter

$t\bar{t}+$jets normalisation (N) and shape (S) uncertainties

- N $t\bar{t}$ cross-section uncertainty correlated accross all $t\bar{t}+$jets components
  - $tt+\geq 1c/b$ pre-fit fractions from PP8, free floating in fit
- N $t\bar{t}+b/bb/B/\geq 3b$ fraction uncertainties from Sherpa (4fs) $t\bar{t}+b\bar{b}$ prediction
  - plus : additional 50% uncertainties on $t\bar{t}+\geq 3b$ and $t\bar{t}+b(MPI/FSR)$ rates
- S uncertainties on $t\bar{t}+$jets, uncorrelated accross $t\bar{t}+\geq 1b$, $t\bar{t}+\geq 1c$, and $t\bar{t}+$light
  - NLO matching : comparison of Powheg+P8 to Sherpa (5fs) $t\bar{t}$
  - PS and hadronisation model : comparison of Powheg+P8 to Powheg+H7
  - additional radiations : simultaneous variations of μ_R, μ_F, $\alpha_S(ISR)$, h_{damp} in Powheg+P8
- S $t\bar{t}+\geq 1b$: comparison of Powheg+P8 (5fs) to Sherpa (4fs)
- S $t\bar{t}+\geq 1c$: comparison of aMC@NLO+H++ (5fs) to aMC@NLO+H++ (3fs)

Systematic source	Description	$t\bar{t}$ categories
$t\bar{t}$ cross-section	Up or down by 6%	All, correlated
$k(t\bar{t}+\geq 1c)$	Free-floating $t\bar{t}+\geq 1c$ normalization	$t\bar{t}+\geq 1c$
$k(t\bar{t}+\geq 1b)$	Free-floating $t\bar{t}+\geq 1b$ normalization	$t\bar{t}+\geq 1b$
SHERPA5F vs. nominal	Related to the choice of NLO event generator	All, uncorrelated
PS & hadronization	POWHEG+HERWIG 7 vs. POWHEG+PYTHIA 8	All, uncorrelated
ISR / FSR	Variations of μ_{R}, μ_{F}, h_{damp} and A14 Var3c parameters	All, uncorrelated
$t\bar{t}+\geq 1c$ ME vs. inclusive	MG5_aMC@NLO+HERWIG++: ME prediction (3F) vs. incl. (5F)	$t\bar{t}+\geq 1c$
$t\bar{t}+\geq 1b$ SHERPA4F vs. nominal	Comparison of $t\bar{t}+b\bar{b}$ NLO (4F) vs. POWHEG+PYTHIA 8 (5F)	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ renorm. scale	Up or down by a factor of two	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ resumm. scale	Vary μ_{Q} from $H_{\text{T}}/2$ to μ_{CMMPS}	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ global scales	Set μ_{Q}, μ_{R}, and μ_{F} to μ_{CMMPS}	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ shower recoil scheme	Alternative model scheme	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ PDF (MSTW)	MSTW vs. CT10	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ PDF (NNPDF)	NNPDF vs. CT10	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ UE	Alternative set of tuned parameters for the underlying event	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ MPI	Up or down by 50%	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 3b$ normalization	Up or down by 50%	$t\bar{t}+\geq 1b$

The fit and the fit checks



- We have large constrains on the large $t\bar{t}+\geq 1b$ modelling shape systematics
- $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ normalisation pulled
  - $k_{t\bar{t}+\geq 1b} = 1.24 \pm 0.10$
  - $k_{t\bar{t}+\geq 1c} = 1.63 \pm 0.23$

  $\Rightarrow$ reflected in post-fit yields
- Several checks
  - post-fit distributions of MVA inputs

    $\Rightarrow$ no mismodelling, agreement improved post-fit
  - alternatives for $t\bar{t}+\geq 1b$ model and systs

    $\Rightarrow$ give compatible results
  - fit on pseudo-data with Powheg+P6 for $t\bar{t}$

    $\Rightarrow$ no bias in μ extraction
  - decorrelate NPs accross regions

    $\Rightarrow$ pulls found to correct bkg. mis-modellings

    $\Rightarrow$ same pulls are seen in bkg.-only fits in bkg.-dominated regions
  - fix NPs to their pre-fit values, and re-do fit

    $\Rightarrow$ μ shift smaller than corresponding uncert.
  - fit channels or categories separately

    $\Rightarrow$ fitted μ is compatible

Questions for CMS

- CMS fit changes $t\bar{t}+\mathrm{HF}$ yields significantly : $t\bar{t}+\mathrm{light}/c\bar{c}/b$ go **up**, $t\bar{t}+b\bar{b}/2b$ go **down**
  - why is it changing that much while the prefit Data/MC seems good ?
  - is it the same when using Sherpa $t\bar{t}+b\bar{b}$ 4fs ?
  - is it the same when fitting with a different variable ?
  - NB : ATLAS reweights $t\bar{t}+\geq 1b$ sub-fractions to Sherpa $t\bar{t}+b\bar{b}$ 4fs, and $t\bar{t}+\geq 1b/c$ go **up**
- CMS test with Sherpa $t\bar{t}+\geq 1b$ 4FS as pseudo-data doesn't change μ (much)
  - is Sherpa 4FS that different from PP8, and closer to data ?
  - are the changes in $t\bar{t}+\mathrm{HF}$ yields the same as in the fit on data ?
- Any other check on fit quality ?
  - e.g. post-fit distributions of MVA inputs