Monte Carlo Event Generator studies for pp $\rightarrow \mathrm{t\bar{t}b\bar{b}}$ events at $\sqrt{s}$ = 13 TeV at the LHC

Proseminar: Physik am LHC

Luca Spitzauer

Overview

I. Motivation for $t\bar{t}b\bar{b}$ differential cross section measurement

II. Previous $t\bar{t}b\bar{b}$ cross section measurements

III. Monte Carlo Event Generator introduction

IV. Analysis of MC Event Generator samples

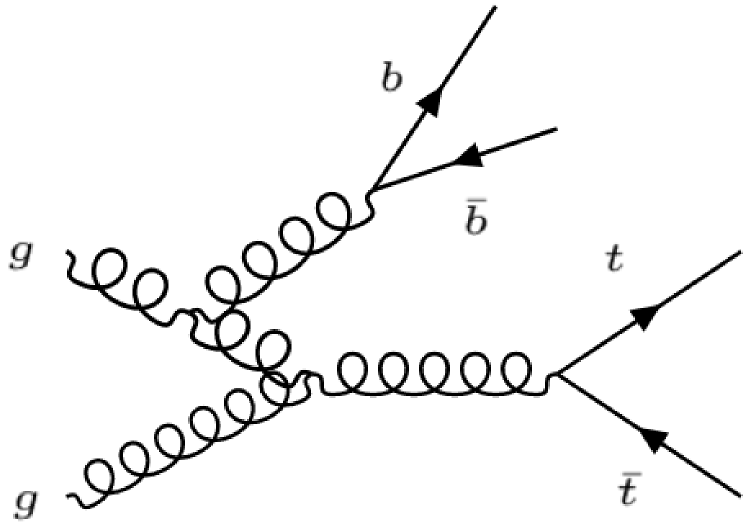
I. Motivation for $\mathrm{t\bar{t}b\bar{b}}$ differential cross section measurement

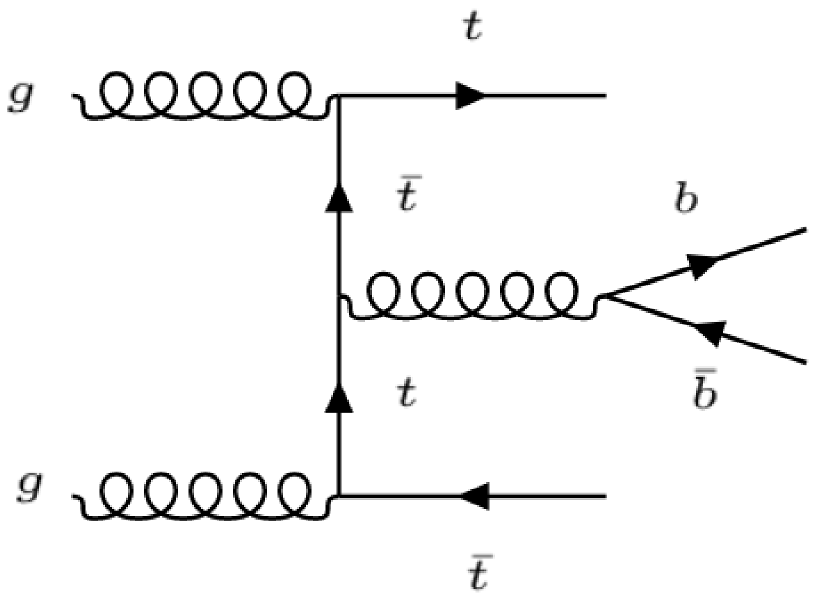
$\mathrm{t\bar{t}b\bar{b}}$

Proton-Proton collisions at the LHC have very large centre of mass energy (here: 13 TeV)

→ Production of multiple heavy quarks like top-quarks (∼173 GeV) or bottom-quarks (∼4.18 GeV) possible

One heavy quark production channel of interest is the production of a top-anti-top quark-pair and a bottom-anti-bottom quark-pair: $\mathbf{t\bar{t}b\bar{b}}$



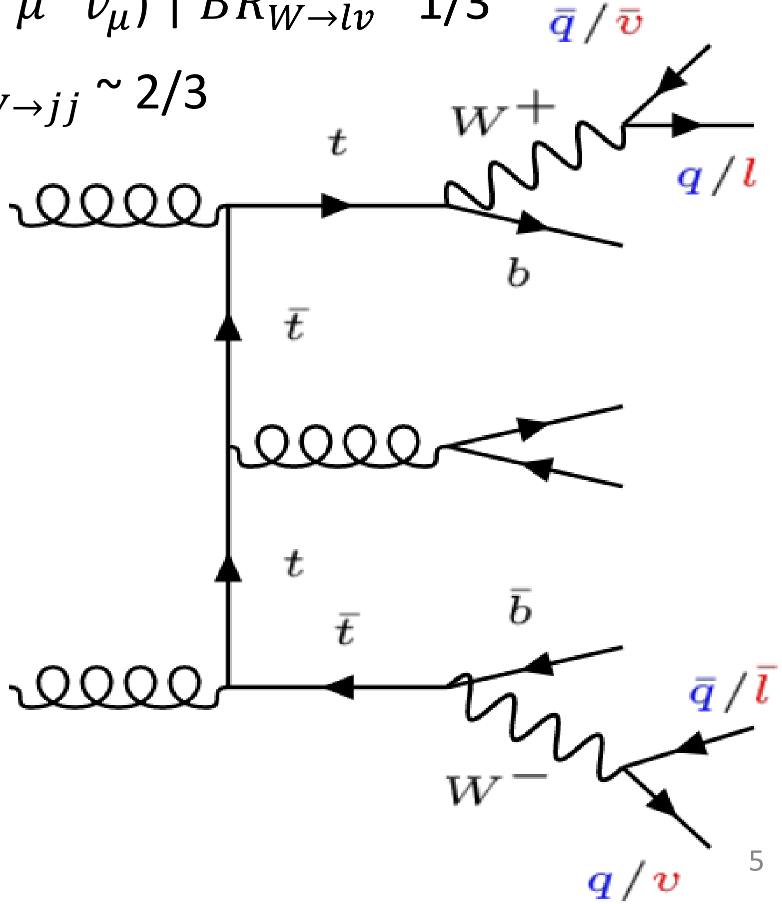


$t\bar{t}b\bar{b}$ final state

Because of their high mass top quarks decay before reaching the detector into a W^+/ W^- and a bottom quark

The W^+/ W^- from the top/antitop each decay as well into either:

- Lepton + Anti-Neutrino/Anti-Lepton + Neutrino pair (e.g. $e^-\bar{v}_e$, $\mu^+ v_\mu$) | $BR_{W\to lv}$~ 1/3
- Quark-anti-quark-pair of different flavour (e.g. $u\bar{d}$, $\bar{c}s$) | $BR_{W\to jj}$ ~ 2/3



Resulting possible final states:

- **All hadronic** ($b\bar{b}$ bjj $\bar{b}jj$)
- **semi-leptonic** ($b\bar{b}$ bjj $\bar{b}l\bar{v}$)
- **dileptonic** ($b\bar{b}$ $bl\bar{v}\bar{b}\bar{l}v$)

Motivations for $\mathrm{t\bar{t}b\bar{b}}$ analysis

Reasons for $\mathbf{t\bar{t}b\bar{b}}$ cross section measurement:

1) Improve understanding of high heavy-flavour production

2) $\mathrm{t\bar{t}b\bar{b}}$ is a major background for Higgs Boson production channels

3) $\mathrm{t\bar{t}b\bar{b}}$ production sensitive to new physics

Improve understanding of high heavy-flavour production

$\mathbf{t\bar{t}b\bar{b}}$ production is a relatively rare process:

NLO predictions for integrated $t\bar{t}$+ b-jets cross sections at 13 TeV*:

(acceptance region: p_T >25 GeV; η<|2.5|)

$t\bar{t}$ + >= 1b-jets cross section: ~12 pb

$t\bar{t}$ + >= 2b-jets cross section: ~2.3 pb

The fully hadronic channel of $\mathbf{t\bar{t}b\bar{b}}$ has not yet been measured by ATLAS

Advantage of full hadronic channel:

- All particles in the final state are detectable at ATLAS (neutrinos in the leptonic final states are not detectable here) so full reconstruction is possible

Disadvantage:

- Larger background from multijet production

Production processes involving multiple heavy quarks are very complex

Simulation studies of $\mathbf{t\bar{t}b\bar{b}}$ are showing discrepancies, requiring further investigation of the modelling of high-flavour and input from measurements

* (https://arxiv.org/pdf/1610.07922.pdf)

Improve understanding of high heavy-flavour production

LHC Cross Section Working Group report (2017):

Generators used:

Tools	Matching method	Shower	FNS	m_b [GeV]	Generation cuts
SHERPA 2.2.1+OPENLOOPS 1.2.3	S-MC@NLO	SHERPA	4FNS	4.75	fully inclusive
MADGRAPH5_AMC@NLO 2.3.2+PYTHIA8 2.1.0	MC@NLO	PYTHIA8	4FNS	4.75	fully inclusive
POWHEL+PYTHIA8 2.1.0	POWHEG	PYTHIA8	5FNS	0	$m_{bb} > 2m_b,\ p_{T,b} > m_b$

4FNS: Four-Flavour-Number-Scheme(Considering effects of the b-mass in the initial state)

5FNS: Five-Flavour-Number-Scheme(Ignoring effects of the b-mass in the initial state)

Comparison of simulated cross-sections

Selection	Tool	σ_{NLO} [fb]	$\sigma_{\mathrm{NLO+PS}}$ [fb]	$\sigma_{\mathrm{NLO+PS}}/\sigma_{\mathrm{NLO}}$
$n_b \geq 1$	SHERPA+OPENLOOPS	$12820^{+35\%}_{-28\%}$	$12939^{+30\%}_{-27\%}$	1.01
	MADGRAPH5_AMC@NLO		$13833^{+37\%}_{-29\%}$	1.08
	POWHEL		$10073^{+45\%}_{-29\%}$	0.79
$n_b \geq 2$	SHERPA+OPENLOOPS	$2268^{+30\%}_{-27\%}$	$2413^{+21\%}_{-24\%}$	1.06
	MADGRAPH5_AMC@NLO		$3192^{+38\%}_{-29\%}$	1.41
	POWHEL		$2570^{+35\%}_{-28\%}$	1.13

$t\bar{t}$ + >=1 b simulation:

Parton Shower contribution low

Consideration of b-mass has large contribution

(5FNS POWHEL 10073 fb vs 4FNS MG5 13833 fb/Sherpa 12939 fb)

$t\bar{t}$ + >=2b simulation:

Huge contribution from Parton Shower

Large discrepancy between generators

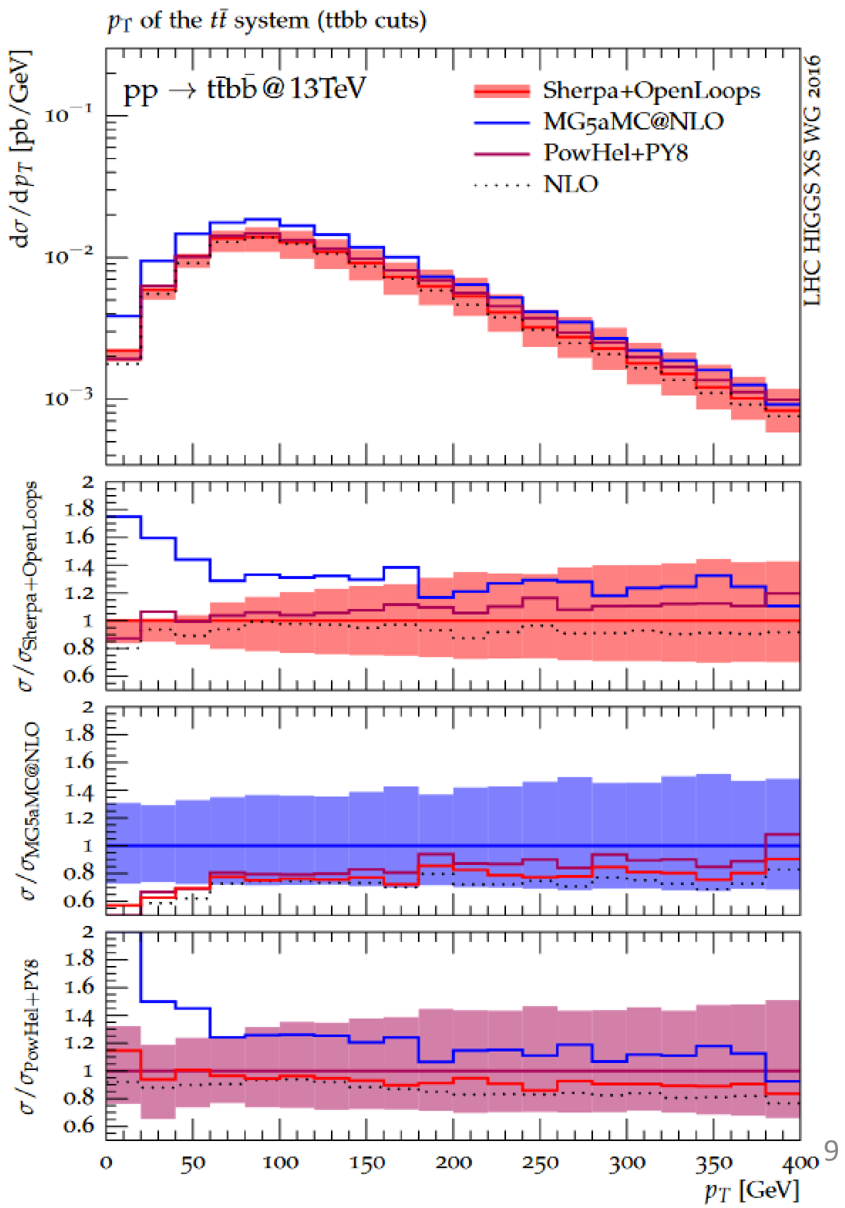
(MadGraph 3192 fb vs Sherpa 2413 fb)

Improve understanding of high heavy-flavour production

Discrepancies in generated kinematics:

Example: Transverse Momentum p_T of $t\bar{t}$-system

MadGraph tends to generate events with lower p_T



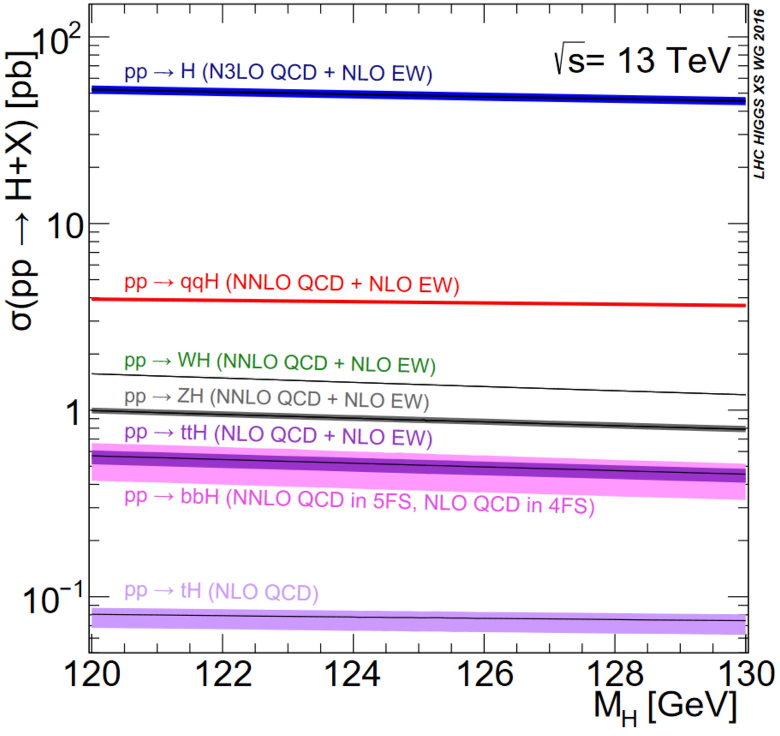
$t\bar{t}b\bar{b}$ in Higgs measurements: $t\bar{t}H$

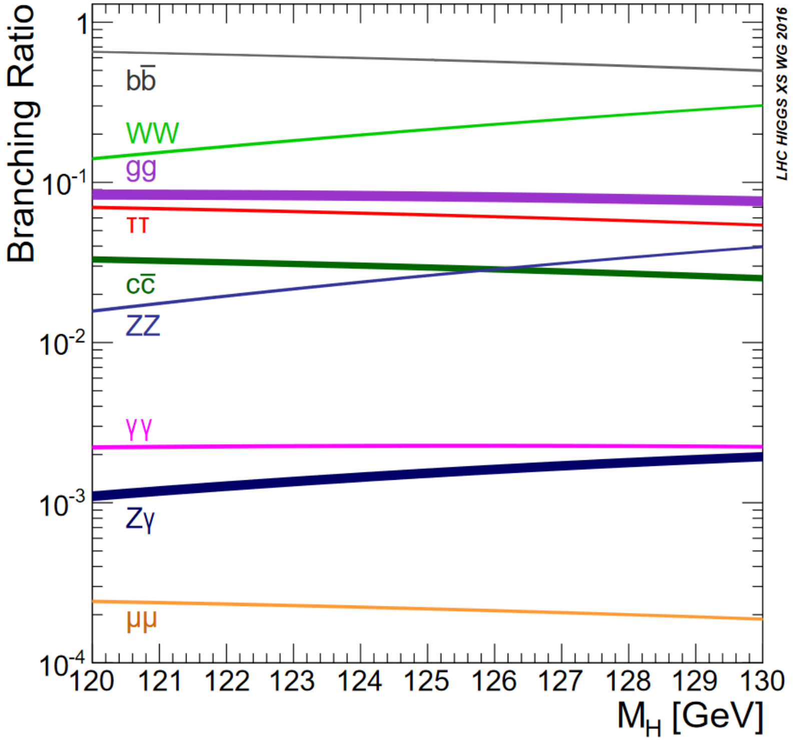
Associated $t\bar{t}H$ production:

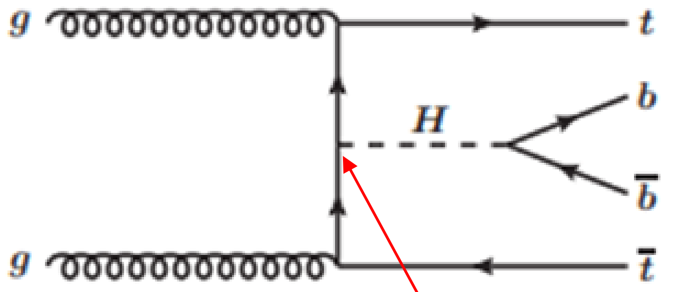
Important test of the SM by measuring Higgs to t coupling

$t\bar{t}H$ (~1% of total Higgs production cross section)

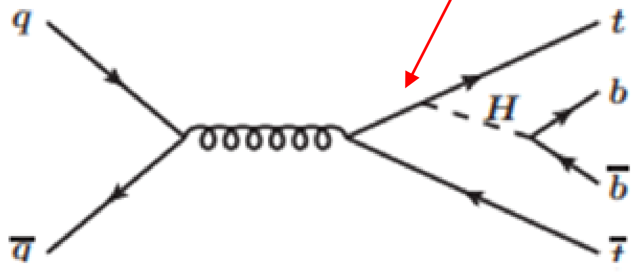
$H \to b\bar{b}$ decay with highest Branching Ratio







Yukawa coupling



https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWG?redirectedfrom=LHCPhysics.LHCHXSWG#Higgs_cross_sections_and_decay_b

$t\bar{t}b\bar{b}$ in Higgs measurements: $t\bar{t}H$

The $t\bar{t}$H production cross section in multiple decay channels has been measured by both ATLAS and CMS

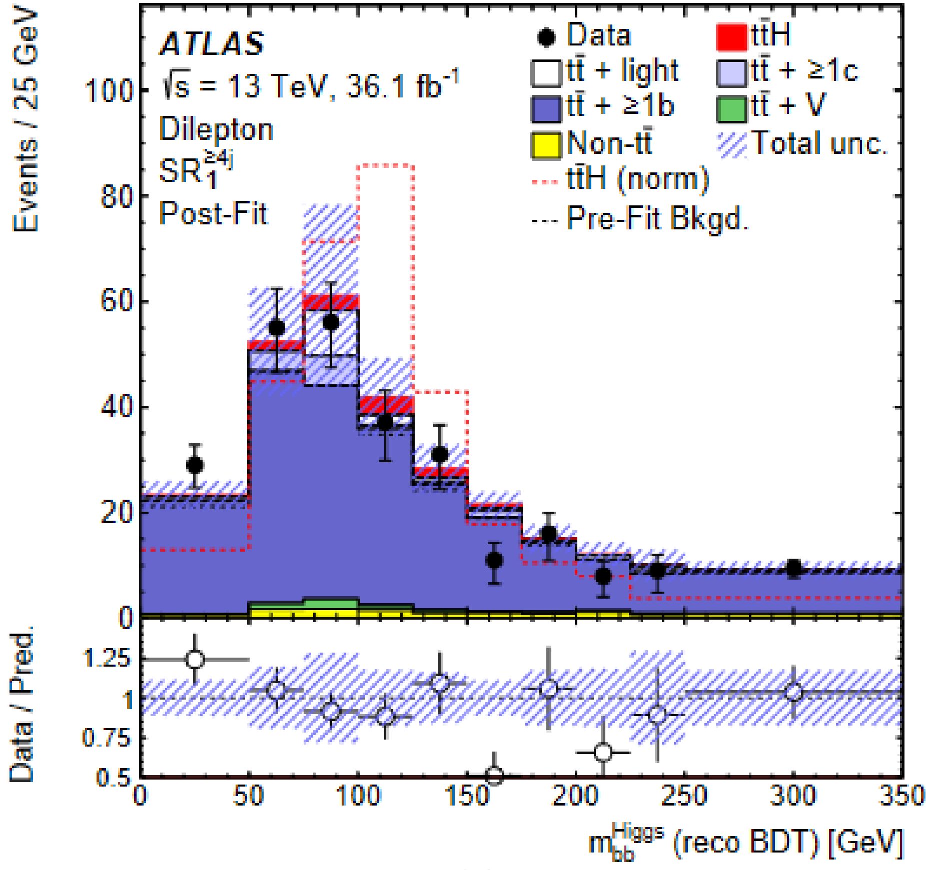
From the ATLAS $t\bar{t}$H measurement at $\sqrt{s} = 13$ TeV: (https://arxiv.org/pdf/1712.08895.pdf)

Samples are semi-leptonic + dileptonic
(at least one W from the tops decays into electron/muon + antineutrino)

$t\bar{t}$ + jets background generated by the Powheg-Box v2 NLO generator and then rescaled to match generated data from the more precise Sherpa event generator

Background contributions in the dileptonic channel

-> tt + >= 1b (blue) constitutes a large background with regards to the signal (red)



https://arxiv.org/pdf/1712.08895.pdf figure 12 (a)

$t\bar{t}b\bar{b}$ in Higgs measurements: $t\bar{t}H$

Uncertainty source	$\Delta\mu$	
$t\bar{t} + \geq 1b$ modeling	+0.46	−0.46
Background-model statistical uncertainty	+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

Contributions to uncertainty of the ratio between measured and predicted $t\bar{t}H$ crossection

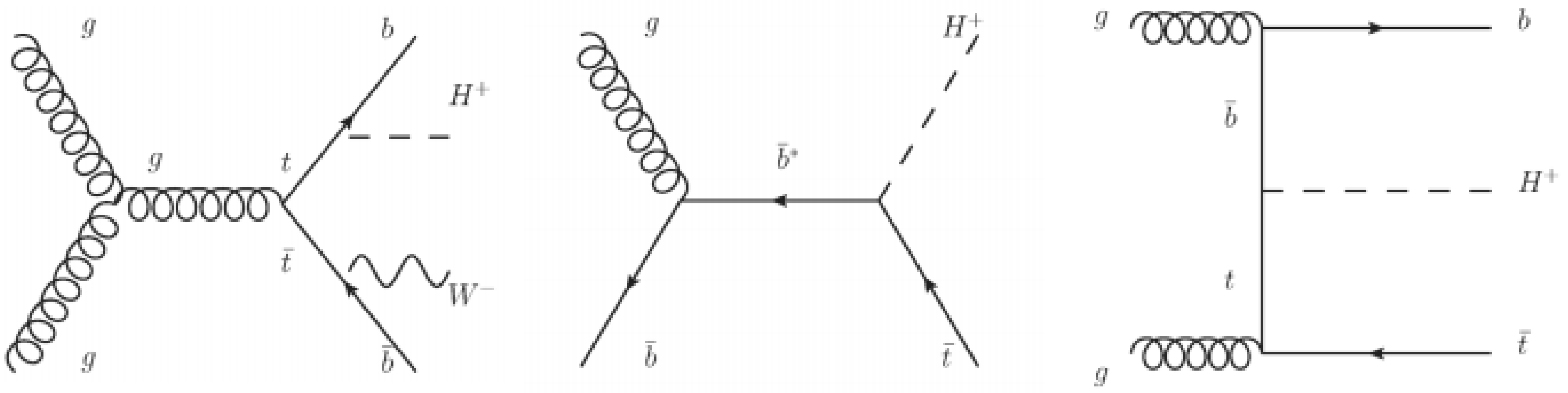
-> tt+ >=1b modelling produces largest uncertainty

https://arxiv.org/pdf/1712.08895.pdf table 2

$t\bar{t}b\bar{b}$ and new physics

$\mathbf{t\bar{t}b\bar{b}}$ is an important background for several searches for new physics beyound the SM

For example: Charged Higgs $\mathbf{H^+}$ ($\mathbf{H^+ \rightarrow t\bar{b}}$)



https://arxiv.org/abs/2102.10076

$t\bar{t}b\bar{b}$ and new physics

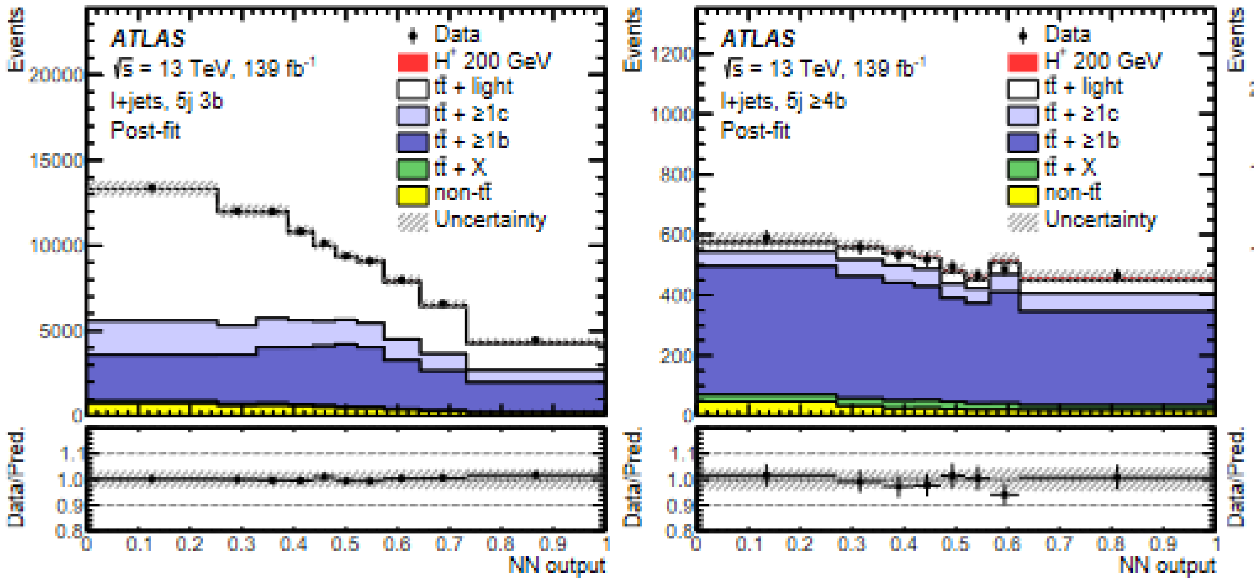
From ATLAS H^+ search at 13 TeV:

(https://arxiv.org/pdf/2102.10076.pdf)

Similar picture as with ATLAS Higgs measurement:

$t\bar{t}$ + >= 1b (blue) constitutes a large background

$t\bar{t}$ + >=1b modelling produces largest uncertainty



(a) 5j3b (b) 5j≥4b

Uncertainty source	$\Delta\mu(H^+_{200})$ [pb]	$\Delta\mu(H^+_{800})$ [pb]
$t\bar{t} + \geq 1b$ modelling	1.01	0.025
Jet energy scale and resolution	0.35	0.009
$t\bar{t} + \geq 1c$ modelling	0.32	0.006
Jet flavour tagging	0.20	0.025
Reweighting	0.22	0.007
$t\bar{t}$ + light modelling	0.33	0.009
Other background modelling	0.19	0.011
MC statistics	0.11	0.008
JVT, pile-up modelling	<0.01	0.001
Luminosity	<0.01	0.002
Lepton ID, isolation, trigger, E_T^{miss}	<0.01	<0.001
H^+ modelling	0.05	0.002
Total systematic uncertainty	1.35	0.049
$t\bar{t} + \geq 1b$ normalisation	0.23	0.007
$t\bar{t} + \geq 1c$ normalisation	0.045	0.015
Total statistical uncertainty	0.43	0.025
Total uncertainty	1.42	0.055

II. Previous $\mathbf{t\bar{t}b\bar{b}}$ cross section measurements

Previous $t\bar{t}b\bar{b}$ Cross Section measurements: ATLAS dileptonic + semi-leptonic

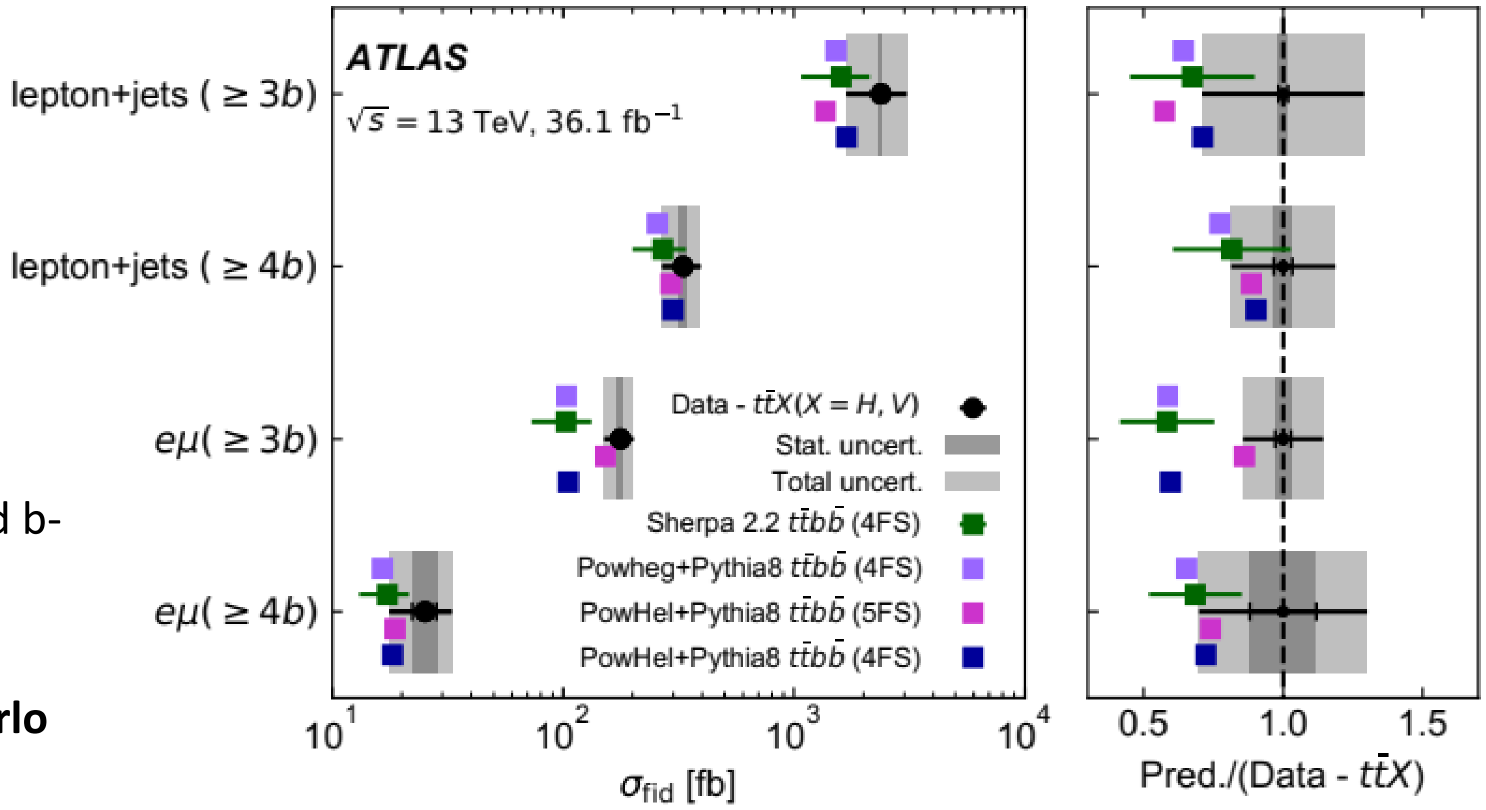
Centre of mass energy: 13 TeV   Luminosity: 36.1 $\mathbf{fb^{-1}}$

Final states:

- $b\bar{b}\ bjj\ \bar{b}lv$
- $b\bar{b}\ bev_e\ \bar{b}\mu v_\mu$

Inclusive and differential cross section are measured

4 signal regions depending on number of leptons and b-jets



Comparison between measured data and Monte Carlo simulation:

Discrepancies between cross section from measured data and different MC generators in the range of ~1-2 σ

https://link.springer.com/content/pdf/10.1007%2fjhep04%282019%29046.pdf figure 7

Previous $t\bar{t}b\bar{b}$ Cross Section measurements: ATLAS dileptonic + semi-leptonic

Differences between measured differential cross section and Monte Carlo prediction

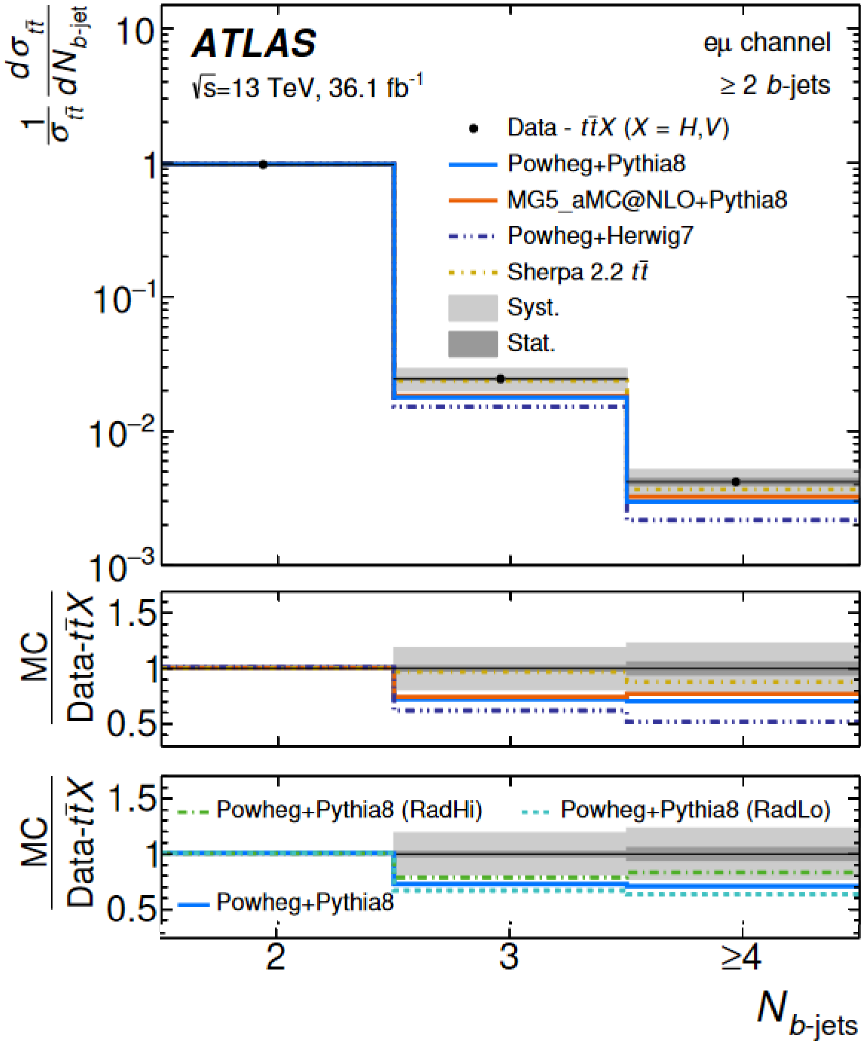
Example:

Number of b-jets

Event Generators underproduce b-quarks compared to data

Best performing generator in this comparison:

Sherpa 2.2 $t\bar{t}$ + jets sample



Adapted from: https://link.springer.com/content/pdf/10.1007%2fjhep04%282019%29046.pdf figure 8

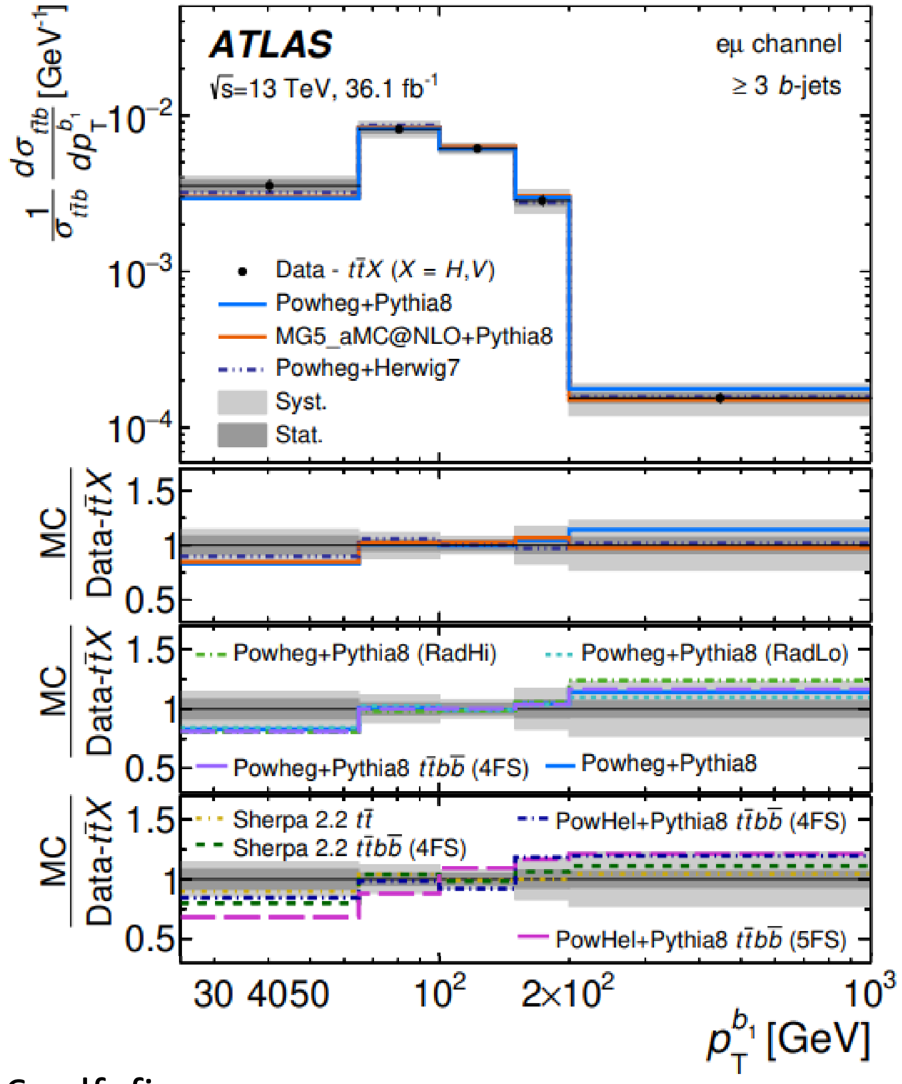
Previous $t\bar{t}b\bar{b}$ Cross Section measurements: ATLAS dileptonic + semi-leptonic

Differences in Pt distribution for >=3 b jets

E.g.: Powheg distribution predicts higher p_T than data

The effect is less pronounced for other generators

→ **Discrepancies between MC event generators and measured Data need to be understood**



https://link.springer.com/content/pdf/10.1007%2fjhep04%282019%29046.pdf fi

Previous $t\bar{t}b\bar{b}$ Cross Section measurements: CMS all hadronic $t\bar{t}b\bar{b}$

Centre of mass energy: 13 TeV   Luminosity: 35.9 $\mathbf{fb^{-1}}$

Final states:

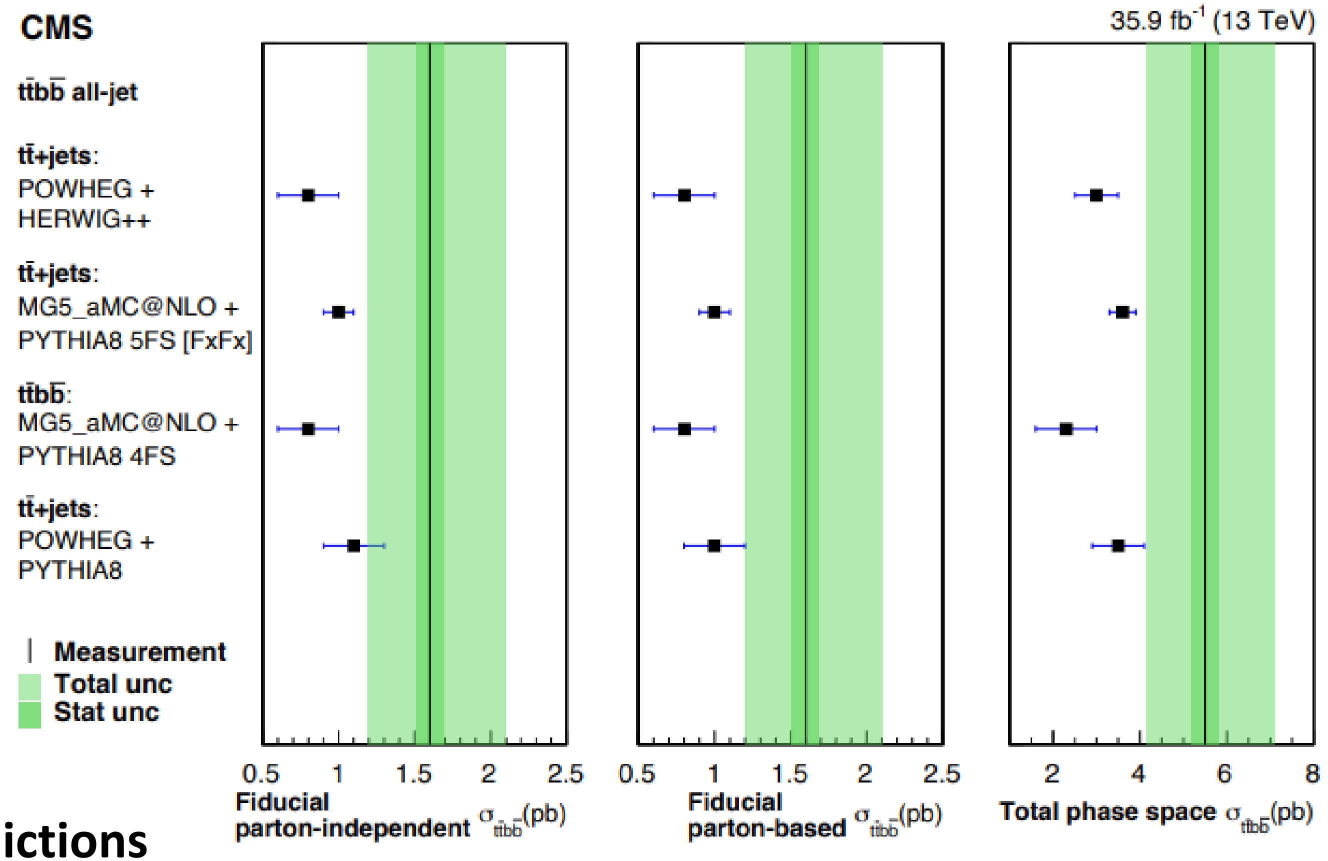
$b\bar{b}\ bjj\ \bar{b}jj$

Event selection:

Two triggers requiring at least 6 jets with

- $|\eta|<2.4$
- p_T > 40 (30) GeV
- Jet scalar sum of p_T (H_T) > 450 (400) GeV
- <= 1 (2) jets are b-tagged

Multivariate analysis to isolate signal



Once again discrepancies in cross section predictions between different MC event generators and measured data are observed

(General trend: MC generator < Data)

https://arxiv.org/pdf/1909.05306.pdf figure 3

III. Monte Carlo Event Generator introduction

Monte Carlo Event Generators

Description of collision events at the LHC involve multiple, highly complex physical processes

Analytic calculation of the expected kinematics is often too difficult and/or resource-consuming to be feasible

Instead:

- Model probability distributions for all processes involved
- Produce expected distributions generating pseudo-random events from the probability distributions

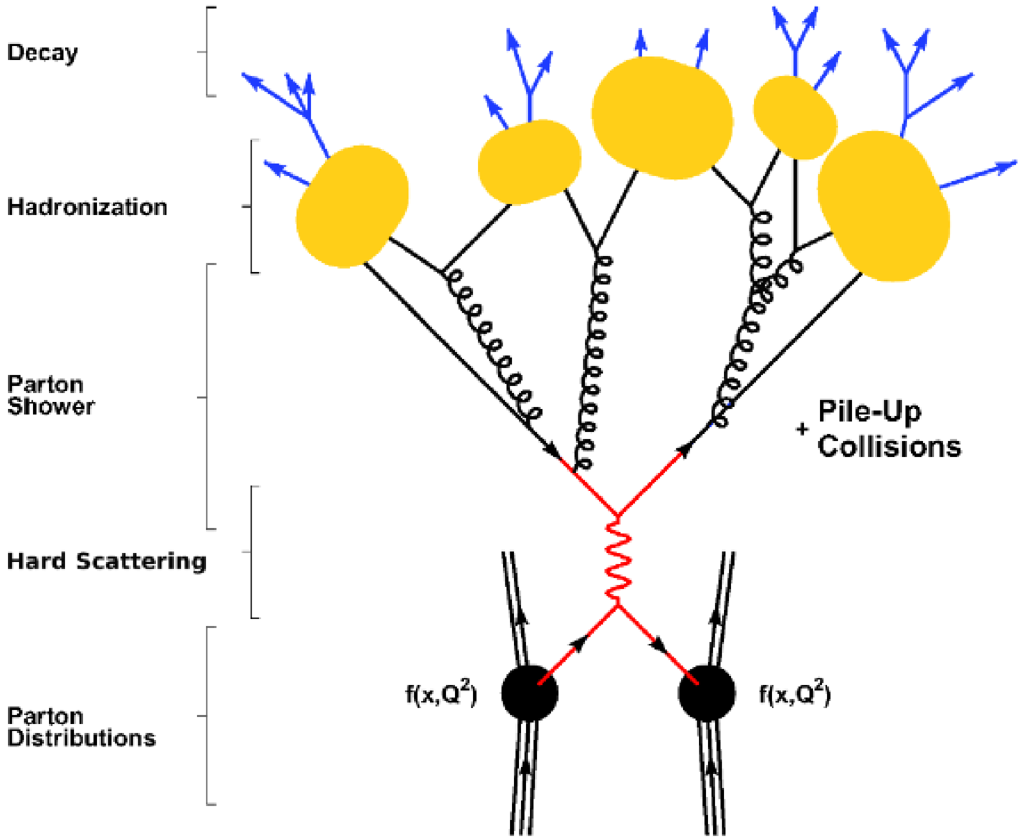
Calculation of differential cross-section by assuming distributions for the involved processes:

$$d\sigma_{final\ state} = d\sigma_{Hard\ Process}\, P_{QCD\ radiation}\, P_{Hadronization}\, P_{Decays} P_{QED\ radiation}\, P_{Multiple\ Parton\ Interactions}$$

Structure of generated events

The generation of a proton-proton collision event can be summarized in few general steps:

1. Hard Process
2. Parton Shower
3. Hadronization
4. Underlying Event
5. Particle Decay
6. Simulation of detector response (Geant4)

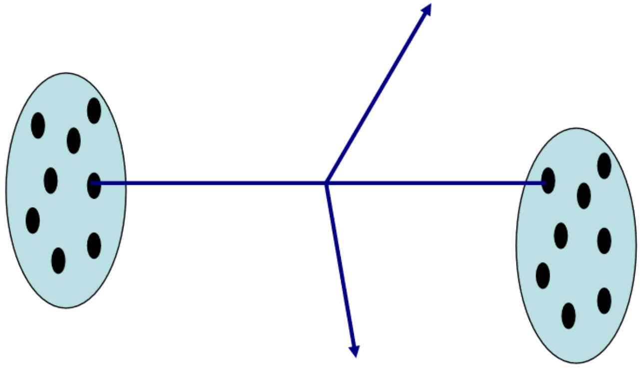


Hard Process

Incoming protons are essentially „clouds“ consisting of partons (quarks + gluons)

How likely these partons are to carry a certain fraction of momentum of the proton is described by the so called Parton Distribution Functions (PDFs)

During the high-energy proton-proton collisions at the LHC these partons interact



Since most processes of interest are very rare, the simulation starts from the particular parton scattering of interest with high momentum transfer (hard scattering)

The Matrix Element describes how likely a particular hard scattering between two partons is

The Matrix Element is calculated using perturbative expansion in powers of the coupling constant

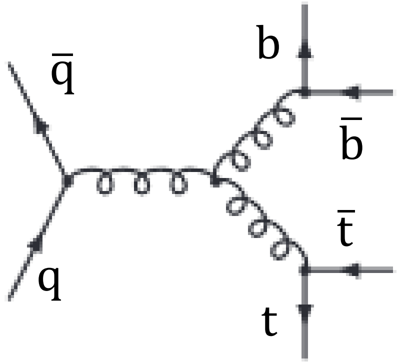
Hard Process

Feynman-Diagramms contributing to the $t\bar{t}b\bar{b}$ matrix element:

Leading Order (Least amount of verteces) production diagramm:

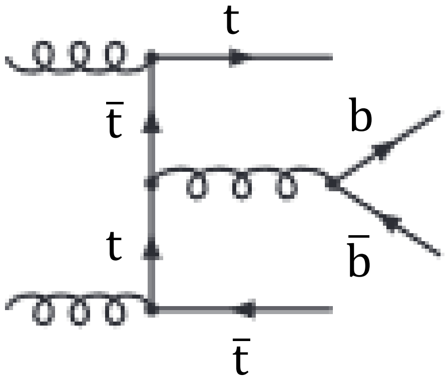
$q\bar{q} \rightarrow t\bar{t}b\bar{b}$ (7 tree diagramms)

Example:



$gg \rightarrow t\bar{t}b\bar{b}$ (36 tree diagramms)

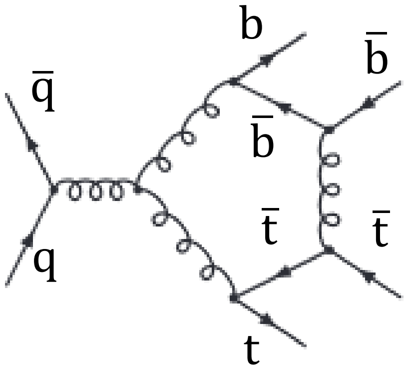
Example:



Next to Leading Order (One more vertex than minimally necessary) production diagramm:

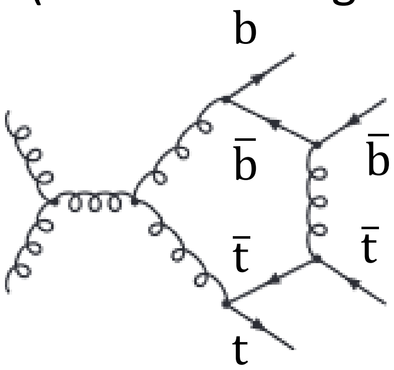
$q\bar{q} \rightarrow t\bar{t}b\bar{b}$ (188 tree diagramms)

Example:



$gg \rightarrow t\bar{t}b\bar{b}$ (1003 tree diagramms)

Example:



Parton Shower

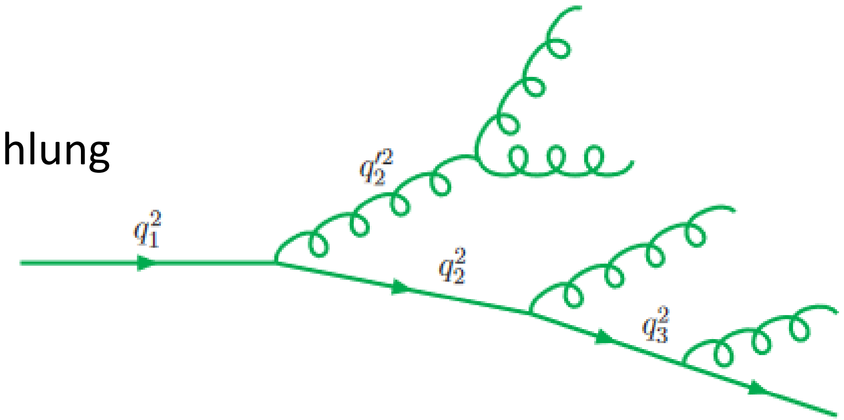
On top of the partons resulting from the Hard Scattering, additional partons are emitted

Just like particles carrying electrical charge emit photons as Bremsstrahlung when accelerated (QED), particles carrying colour charge emit gluons when accelerated (QCD)

But: Gluons carry color themselves

→ Gluons emit other, lower energy partons as Bremsstrahlung

→ Large number of partons generated (Parton Shower)



$$q_1^2 > q_2^2 > q_3^2 > \dots$$
$$q_1^2 > q_2'^2 \dots$$

Initial State Radiation (ISR): Parton emissions before hard scattering

Final state Radiation (FSR): Parton emissions after hard scattering

Parton Shower

Since the emitted gluons spiral down into lower and lower energy, fixed order perturbation theory starts to break down since higher order corrections become more significant at lower energy scales

→ A parametric function is used

$$d\sigma = \sigma_0 \frac{\alpha_s}{2\pi} \frac{d\theta^2}{\theta^2} dz\, P(z,\phi) \frac{d\phi}{2\pi},$$

σ_0 Total cross-section θ opening angle between quark/gluon

α_s strong coupling constant Φ splitting angle around parent parton

$P(z, \phi)$: Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP)splitting-function

Good approximations for collinear and soft (low energy) gluons

For harder and more sperated jets assistance from fixed order NLO perturbation calculations improve the results (to avoid double counting: Matrix Element matching)

Hadronization

Since gluons carry colour-charge themselves the strong field increases in intensity at longer distances

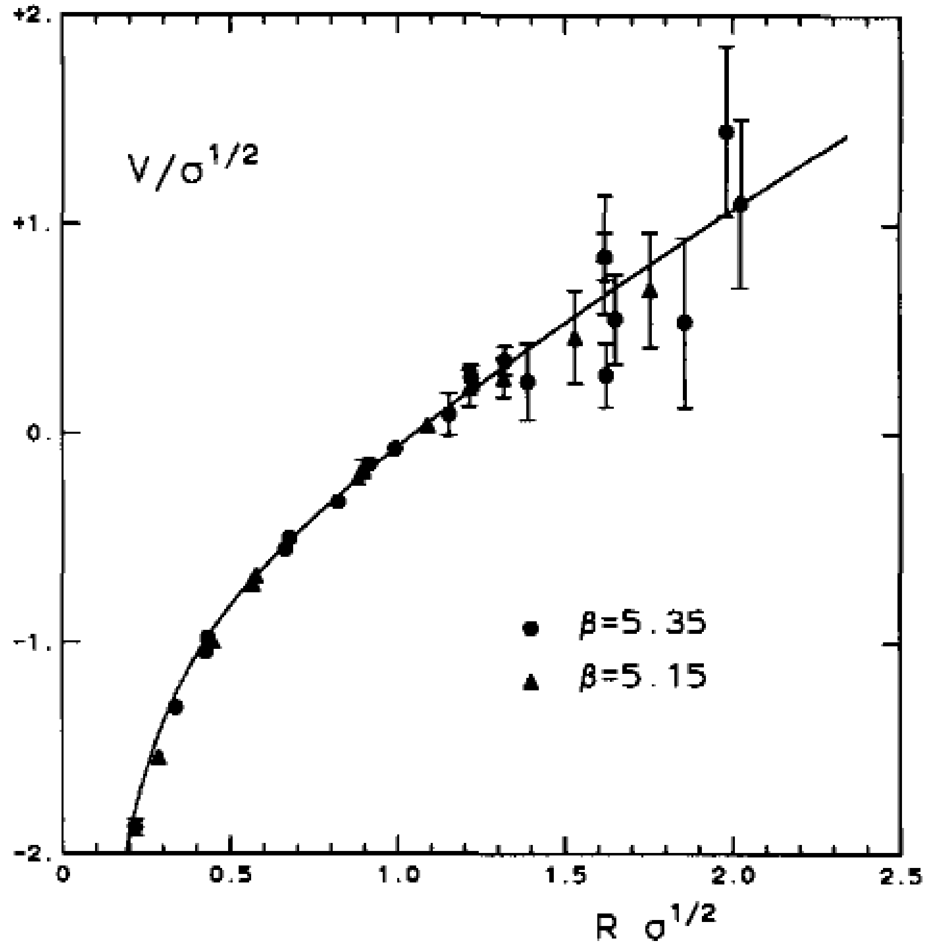
→ no free quarks (confinement)

Partons produced by the hard scattering and parton shower continue to move away from each other

→ Potential between them increases until it is energetically favorable to produce a quark-/antiquark-pair to form Hadrons

Since the coupling constant is large for these long range interactions perturbation theory breaks down

→ **Hadronization models required**



String Model	**Cluster Model**
Interquark Potential modelled as elastic strings that break from quark-anti-quark production	Gluons are treated as color-anticolor-pair forming mesonic-resonances (clusters) that decay into stable hadrons

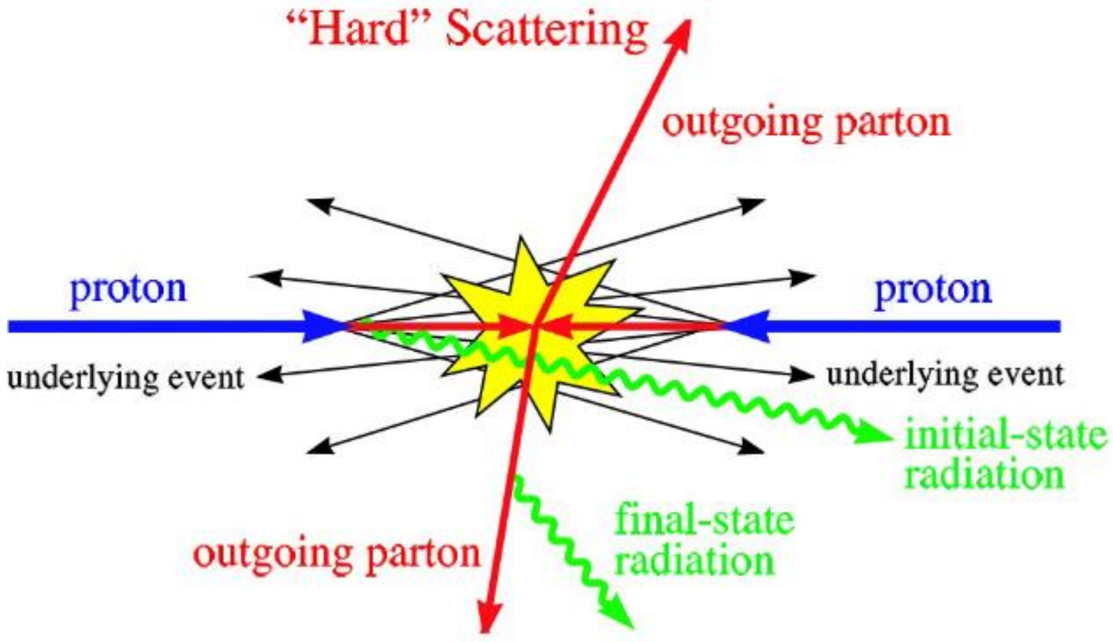
https://inspirehep.net/files/692a257de6c8fc73e2136312d179ac4e

Underlying Event

After hard scattering:

Other partons remain

Remaining partons can cause further soft scatterings



https://www.icts.res.in/sites/default/files/jlhc2017-2017-01-25-Sunil-bansal.pdf

IV. Monte Carlo event generator studies

Analysis Overview

Center of mass energy: 13 TeV   Luminosity: not final, 139 $\mathbf{fb^{-1}}$ at most

final state: $\mathrm{b\bar{b}}$ bjj $\mathrm{\bar{b}jj}$ (all hadronic, 8 jets, 4 are from b quarks)

Primarily expected background is QCD multijet production

Event selection:

- 8 final jets with $\mathrm{p_T}$ > 25 GeV and |η| < 2.5
- No leptons
- Low missing transverse energy
- multiple choices of trigger being considered
- $\mathrm{t\bar{t}}$ reconstructed from decay products

Work is currently ongoing

Monte Carlo event generator studies

Due to the disagreements between multiple generators and measured $\mathbf{t\bar{t}b\bar{b}}$ cross-sections it is interesting to compare multiple signal sample and investigate different MC event generators

The study is done at parton level (before hadronization effects)

Samples used in this comparison:

- **NLO Powheg all hadronic $\mathbf{t\bar{t}b\bar{b}}$ sample (100000 Events)**
- **LO Madgraph5 all hadronic $\mathbf{t\bar{t}b\bar{b}}$ sample(700000 Events)**

Simulation of Initial State Radiation is turned off

(meaning there are no parton emissions before the hard scattering)

- **NLO MadGraph5 dileptonic + semi-leptonic $\mathbf{t\bar{t}jj}$ sample (67000000 Events)***

($t\bar{t}jj$ implies that b-quarks that are not from the top decay are expected to come from the parton shower)

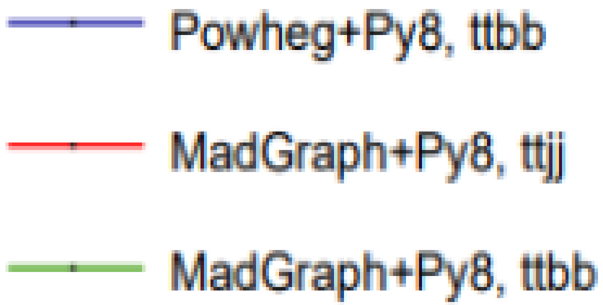
For all samples, Parton Shower is generated using Phytia8

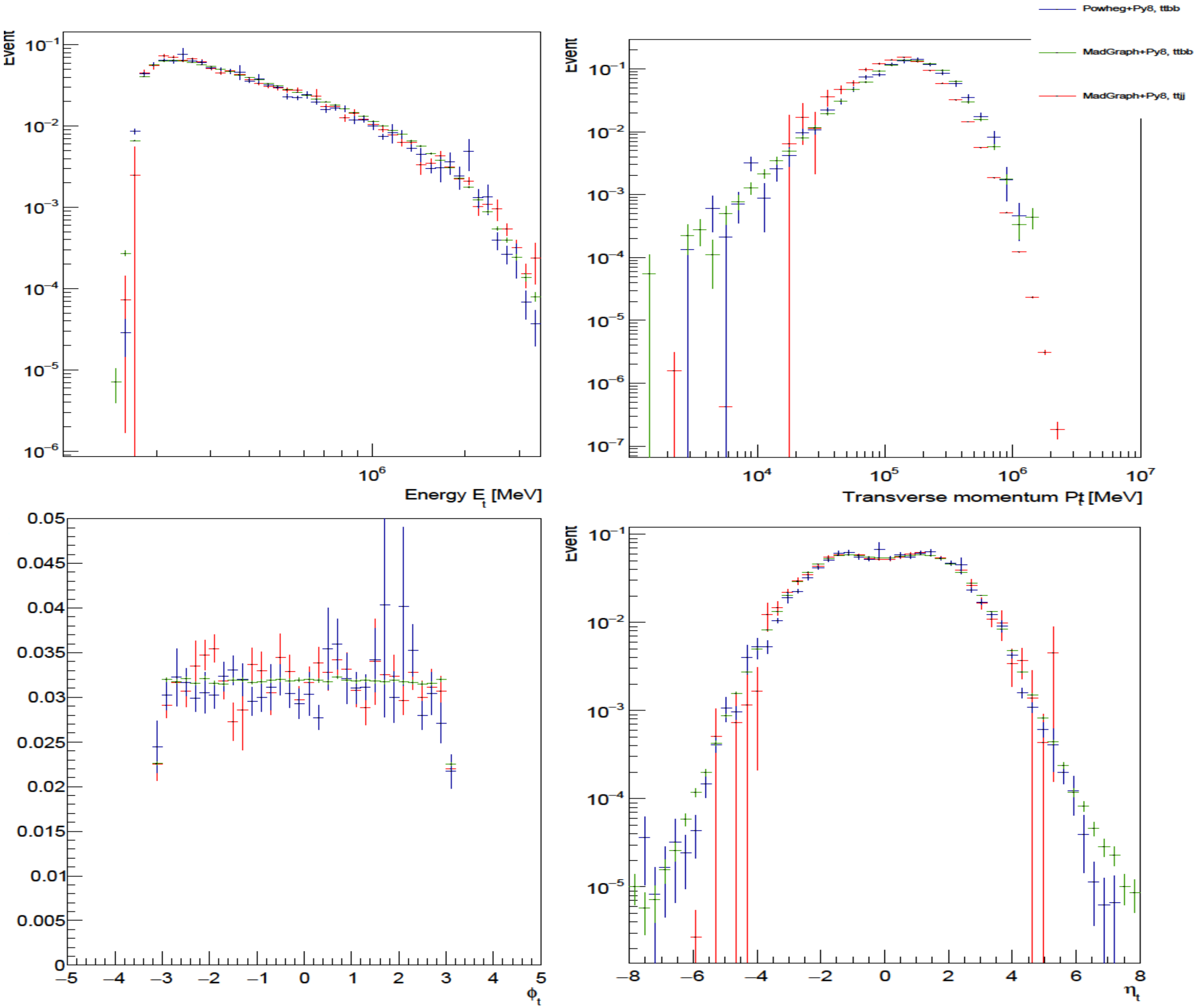
*An all hadronic ttjj sample was not yet available, but that is not a problem for a parton level study

t-quark kinematics

Generally very good agreement in shape between all three samples

(Plots are normalised to unity to compare shape)





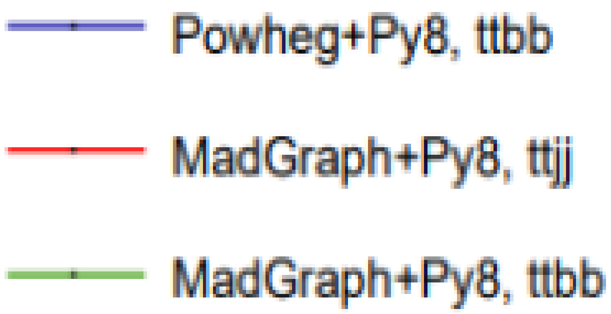
$t\bar{t}$ -quark-system

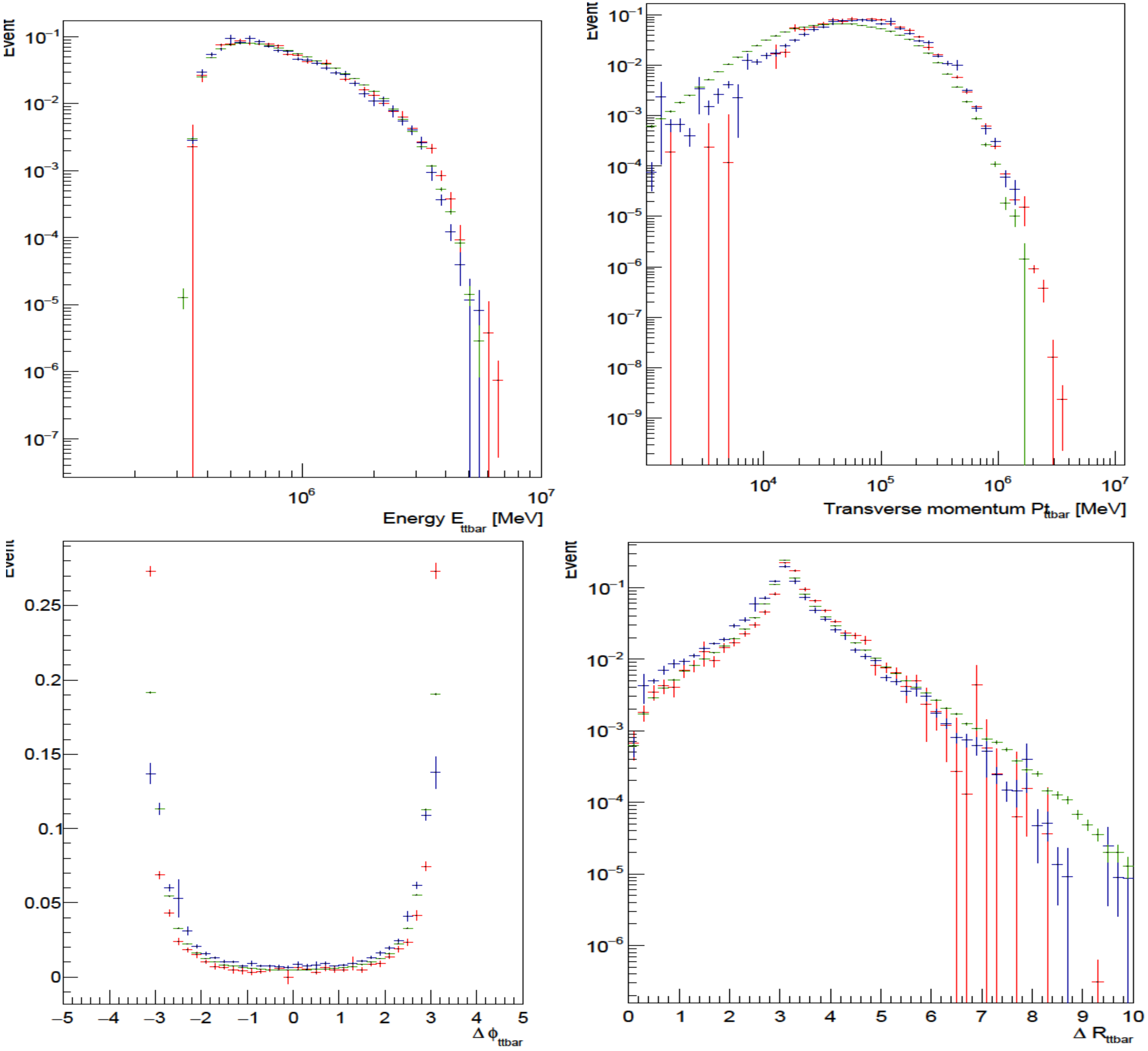
Small difference in $t\bar{t}$ kinematics

MG $t\bar{t}b\bar{b}$ sample somewhat shifted to lower p_T

(Could be caused by missing NLO corrections in the MG $t\bar{t}b\bar{b}$ sample)

Good overlap between MG $t\bar{t}jj$ and Powheg $t\bar{t}b\bar{b}$ sample





Extra b-quarks

Analysis of b-quarks that do not result from top-quark decays

Significantly lower ttbb events from $t\bar{t}jj$ sample

b-quark distribution mostly follows expectation:

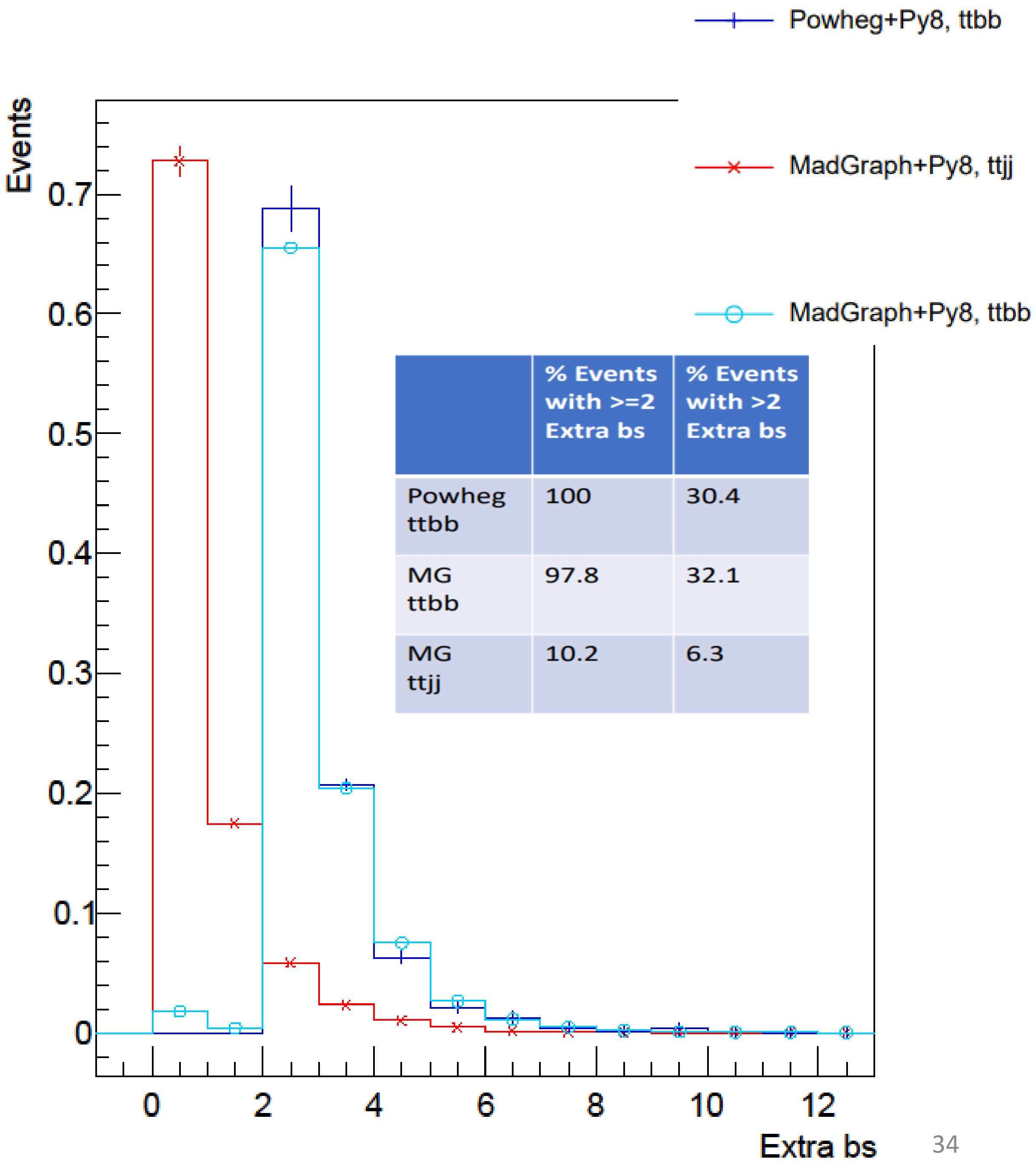
$t\bar{t}jj$ sample has mostly 0 additional b jets
(no contribution from matrix element)

$t\bar{t}b\bar{b}$ samples have mostly 2 additional b jets (2 b-quarks are expected from the matrix element since the b process was chosen as hard scattering)

The amount of bins with no b-quarks outside the top-decay (~2%) is unexpected since 2 b-quarks should be produced in the hard scattering

Origin of the bins containing no additional b-quarks remains an open question as of yet

Since there is a non-zero fraction containing more than 2 b-quarks it has to be made sure that the correct ones are chosen for the $t\bar{t}b\bar{b}$ system



	% Events with >=2 Extra bs	% Events with >2 Extra bs
Powheg ttbb	100	30.4
MG ttbb	97.8	32.1
MG ttjj	10.2	6.3

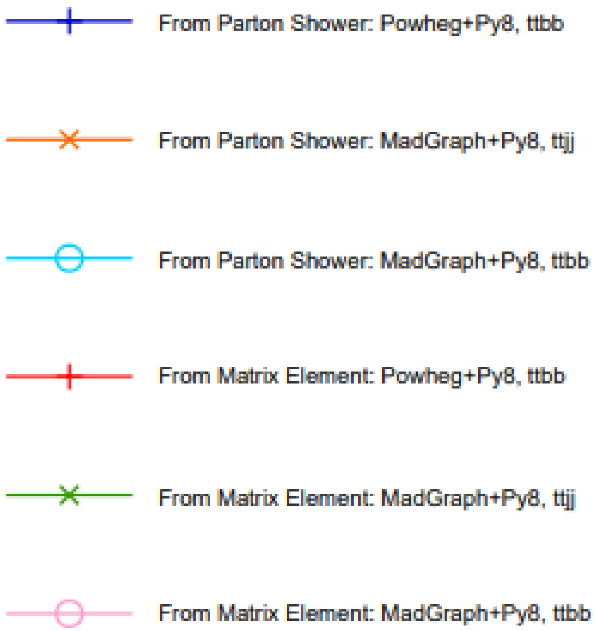
Additional b-quarks

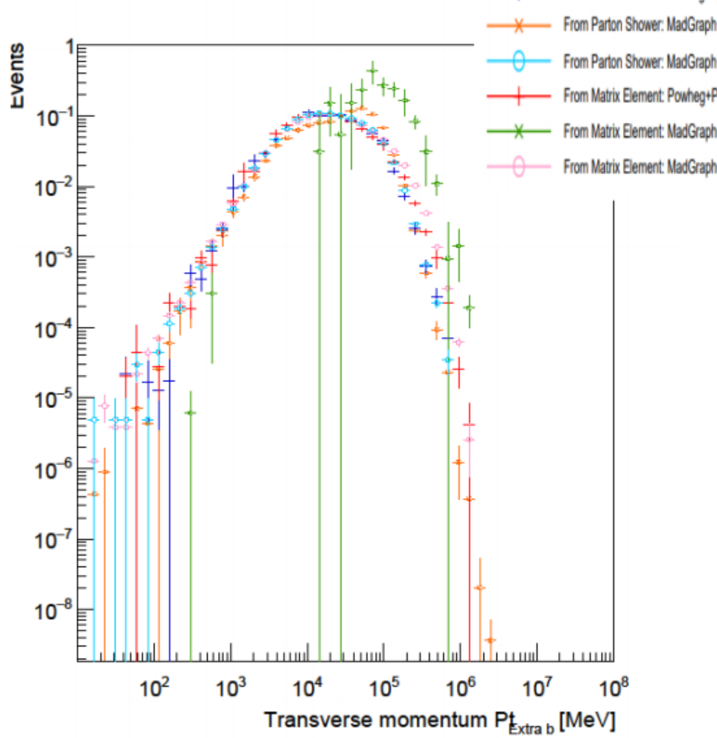
Kinematics of the b-quarks not produced by the top-decay are seperated into b-quarks from the matrix element and b-quarks form the parton shower

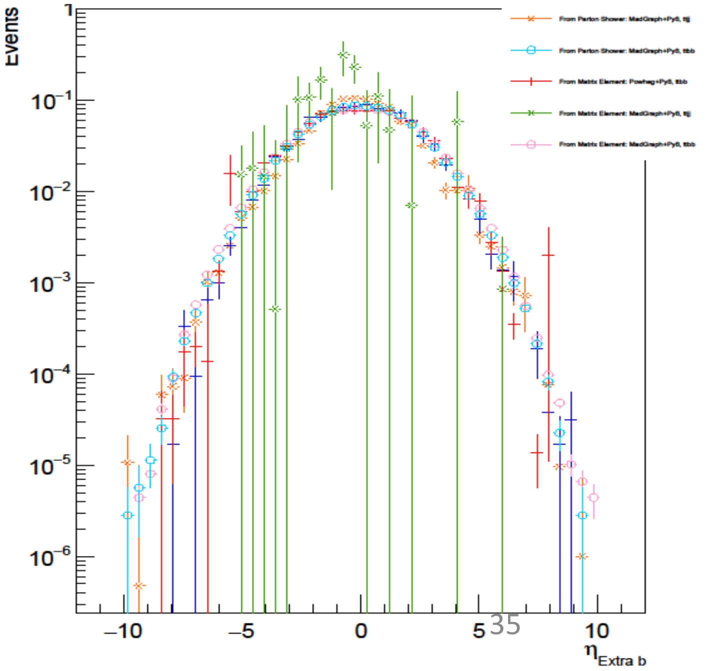
Good agreement in shape of kinematics between both ttbb samples

The b-quarks from the $t\bar{t}jj$ sample‘s parton shower seem to have a slightly shifted peak towards higher Pt

The b-quarks from the $t\bar{t}jj$ sample‘s matrix element show a more noticibely different form







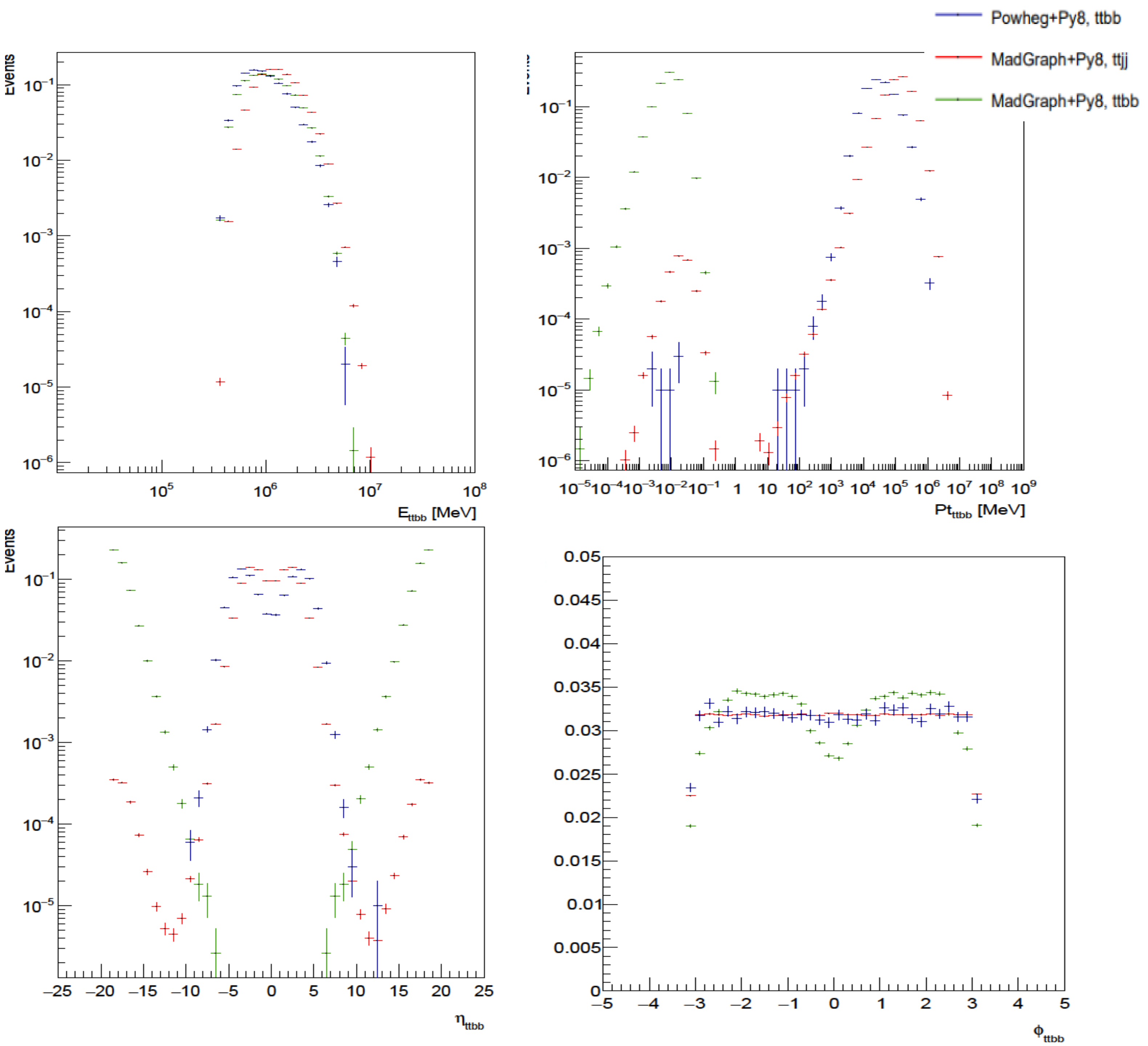
$t\bar{t}b\bar{b}$ system

The two b-quarks used to build the $t\bar{t}b\bar{b}$ system are selected by choosing the matrix element b-quarks for the $t\bar{t}b\bar{b}$ samples and by choosing the b-quarks with highest p_T for the $t\bar{t}jj$ sample

Very different shapes between samples

Deviations of MG ttbb sample are expected due to lack of inital state radiation

- Since there is no initial state radiation the partons just move in the direction of the beam line and have negliable transverse momentum
- Because conservation of momentum the resulting transverse momentum of the system emerging from the hard scattering also has to be very low and they are more likely to move in the direction of the beam pipe



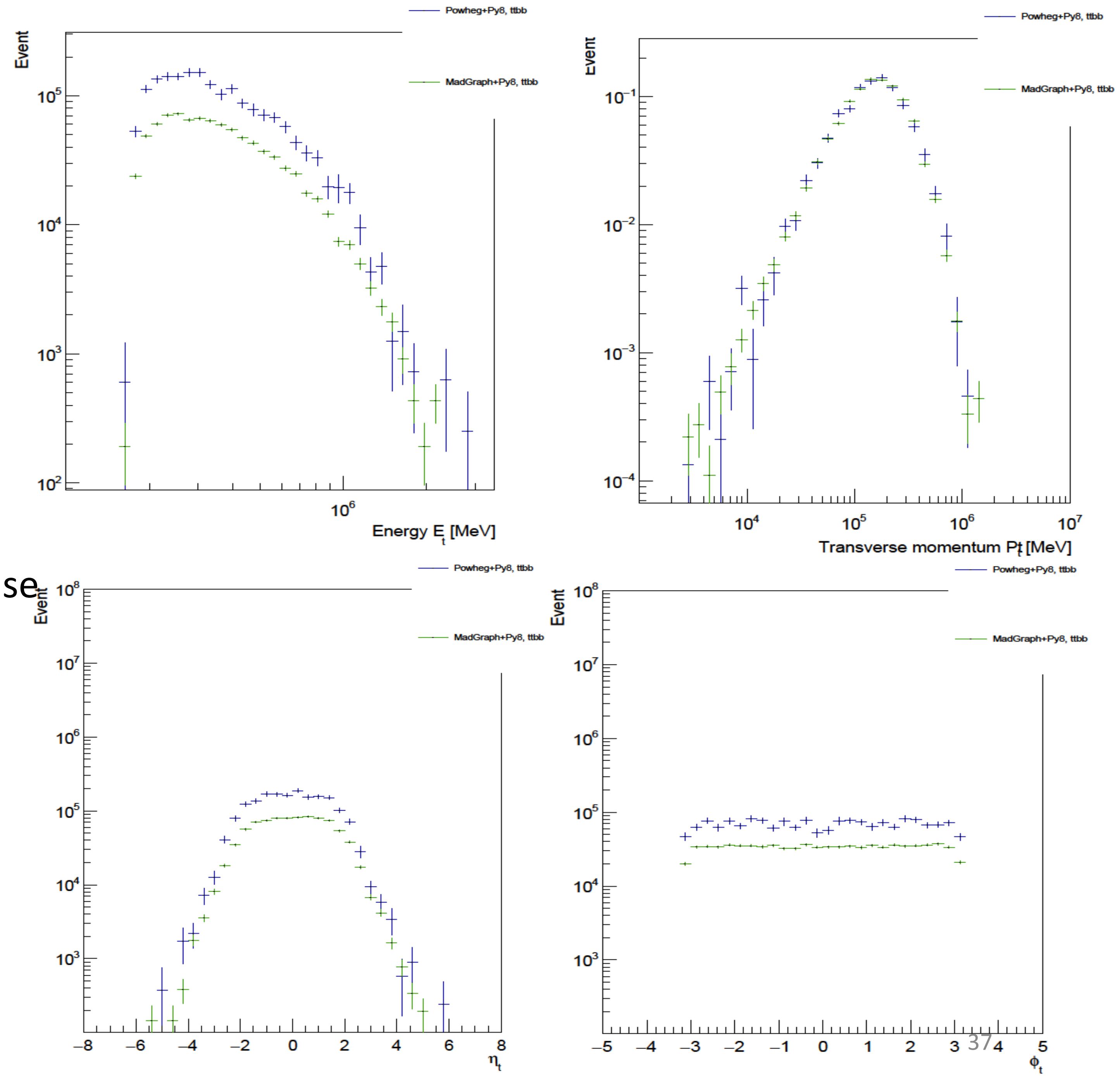
8 jet selection

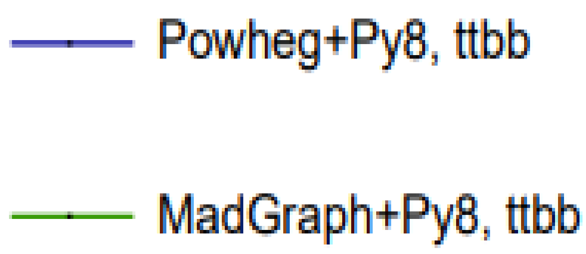
All 8-Jet particles (b-quarks from top-quark decay + W decay products + extra bs) p_T >25 GeV; η<|2.5|

$t\bar{t}jj$ sample is not included here since the dileptonic + leptonic data has not yet been adjusted to be compatible with these cuts

Good agreement between the $t\bar{t}b\bar{b}$ samples

Factor two in normalisation is expected because of the difference between LO (MadGraph sample) and NLO (Powheg sample)



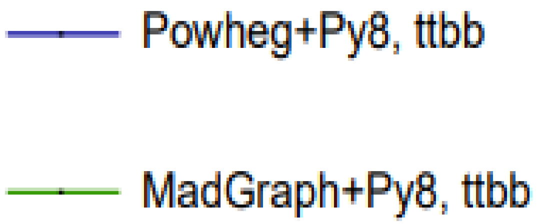


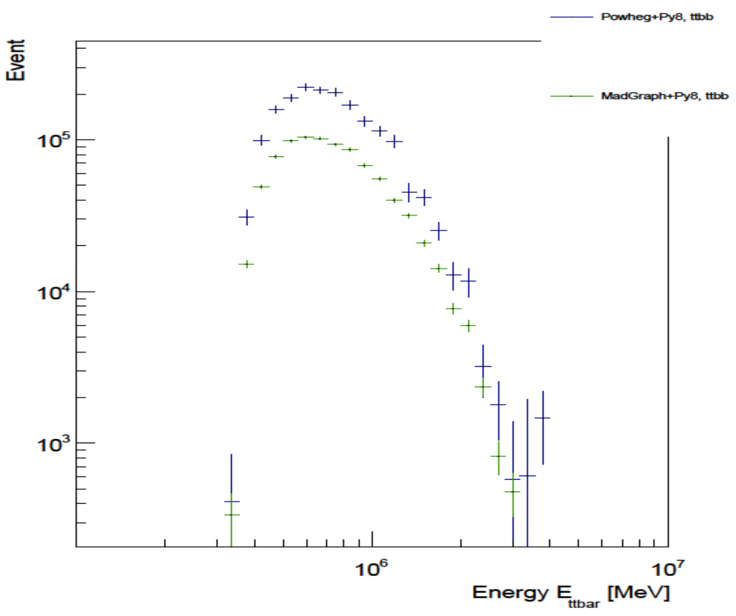
8 jet selection

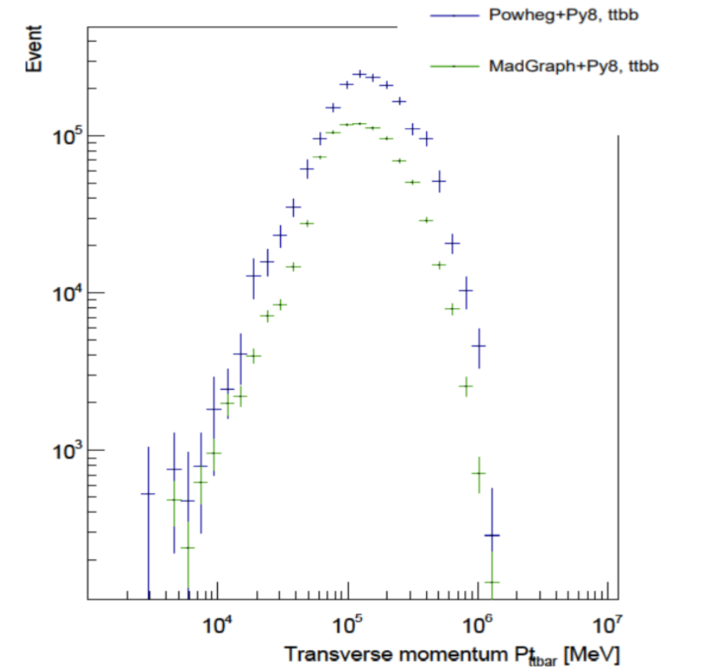
$t\bar{t}$ system:

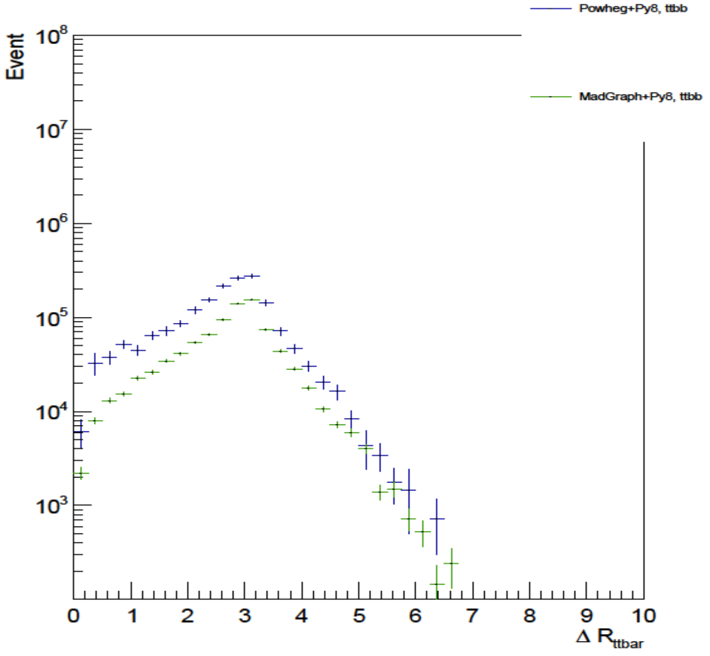
$t\bar{t}jj$ sample is not included here since formatting of the dileptonic + leptonic simulated data has not yet been adjusted to these cuts

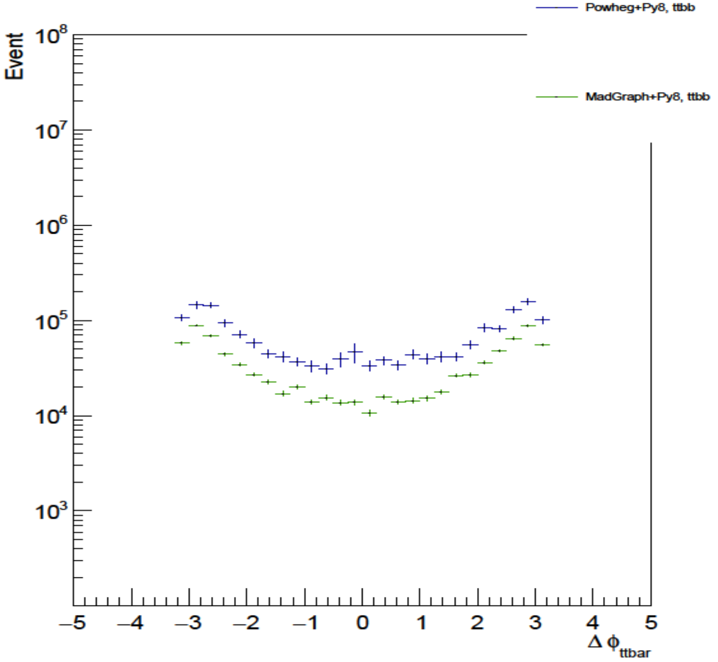
Good agreement between the $t\bar{t}b\bar{b}$ samples











Summary

- $t\bar{t}b\bar{b}$ production is a process of considerable interest with important implications for different areas of research at the LHC
- Monte Carlo event generators are an important tool to describe complex collision processes
- Discrepancies between the simulated data from event generators and measured data need to be understood

Summary

- In general good agreement was found between both $t\bar{t}b\bar{b}$ samples as well as the $t\bar{t}jj$ sample
- The contribution of inital state radiation and the effect of NLO vs LO was also nicely demonstrated
- Some open questions remain about some unexpected behaviour in the histograms
- Study is currently being finalised

Thank you for your attention

Special thanks to Prof. Kroha and Dr. Margherita Spalla