All Hadronic $t\bar{t}H(b\bar{b})$ Analysis with the ATLAS Detector

Giovanni Bartolini

[1] Aix-Marseille-Université, CPPM, Marseille, France

December 17, 2019

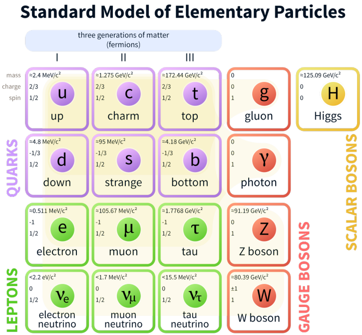
Outline

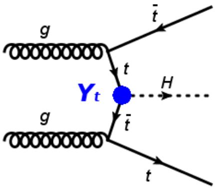
1. **Introduction**
   - Standard Model and Top Yukawa Coupling
   - The ATLAS Experiment at LHC
2. ***b*-jet Trigger Calibration**
3. **All Hadronic $t\bar{t}H(b\bar{b})$ Analysis**
   - TRF_{MJ}
   - Multi-Variate Analysis
   - Fit preliminary results
4. **Conclusion**

Introduction

Standard Model and Top Yukawa Coupling

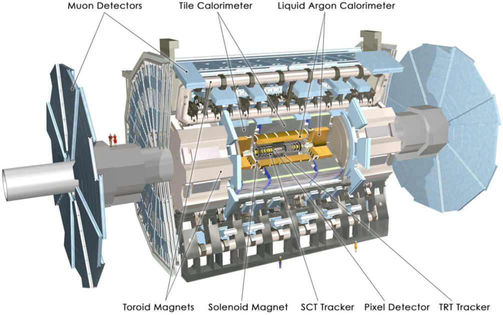
- The **Standard Model** (SM) of Particle Physics is a gauge theory that classifies all known elementary particle and describes Strong, Weak and Electromagnetic interaction forces
- Very successful theory, still many shortcomings: inclusion of gravity, neutrino masses, evidence of dark matter, ...
- Discovery of **Higgs** boson in 2012 completed the set of predicted elementary particle and started an effort on the precise measure of its properties
- The **top** quark is the heaviest elementary particle
  - has the highest Yukawa coupling: $Y_t \sim 1$
- Anomalous values for Y_t could hint for *Beyond the Standard Model* (BSM) Physics
- Associated production ($t\bar{t}H$) only way to directly measure Y_t

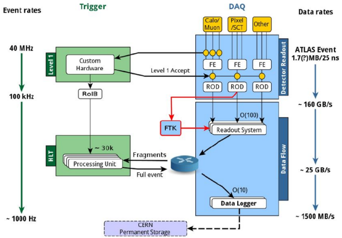




The ATLAS experiment at LHC

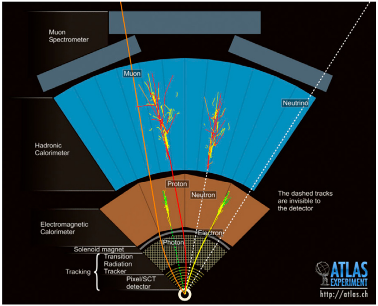
- The ATLAS detector is placed in one of the 4 interaction point of the Large Hadron Collider (LHC) in CERN experimental area and collects data from pp collisions at $\sqrt{s} = 13$ TeV
- The LHC provides $\sim 10^{34}$ cm^{-2} s^{-1} instantaneus luminosity
  - more than 40 million collision per second
  - trigger system to collect only interesting events (few hundreds per second)



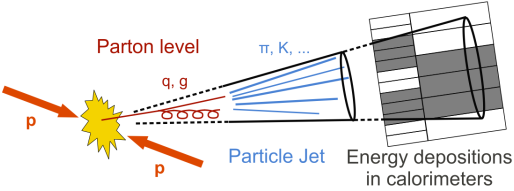


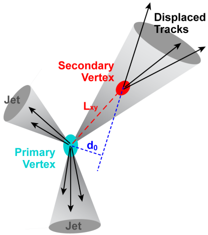
Particle detection with ATLAS

- **Electrons**: Energy deposition in calorimeter and charged track in ID
- **Photons**: Energy deposition in calorimeter, no track in ID
- **Muons**: Combined track in ID and MS
- **MET**: negative vectorial sum of selected physics objects and the soft term
- **JETS**: Quarks and gluons that are produced from a collision will *hadronize* producing a collimated flow of hadronic particles, that is reconstructed using anti-k_t algorithm

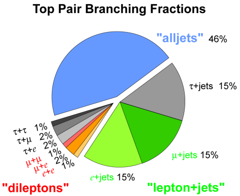


- **BJET**: b-hadrons travel few hundreds μm before decay, ATLAS has impact parameter track resolution of $\sim 10\ \mu$m: can reconstruct the Secondary Vertex (SV) to identify b-jet

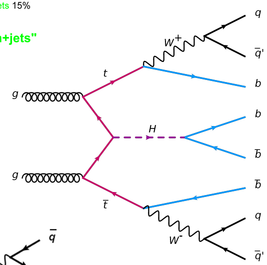


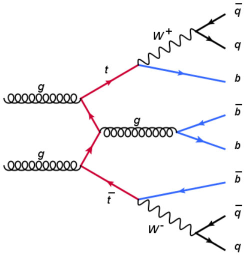


$t\bar{t}H$ Production in Fully Hadronic Final State



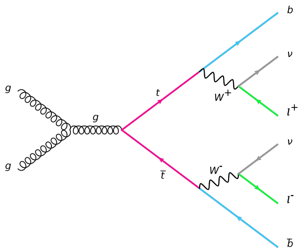
- $t\bar{t}H$ process has many accessible final states: $\gamma\gamma$, multi-lepton, lepton+jets, all hadronic
- ATLAS first $t\bar{t}H$ observation using Run 1 and Run 2 data published last year
  - with observed(expected) significance of 6.3(5.1)
- All hadronic has the largest brancing ratio:
  - $\sim$33% of total $t\bar{t}H$ production
- Ideal for differential analysis
  - explore the CP nature of Y_t
- Challenging experimental signature:
  - 8 quarks, 4 b-quarks
  - Large QCD multi-jet background
  - irreducible $t\bar{t} + b\bar{b}$ background



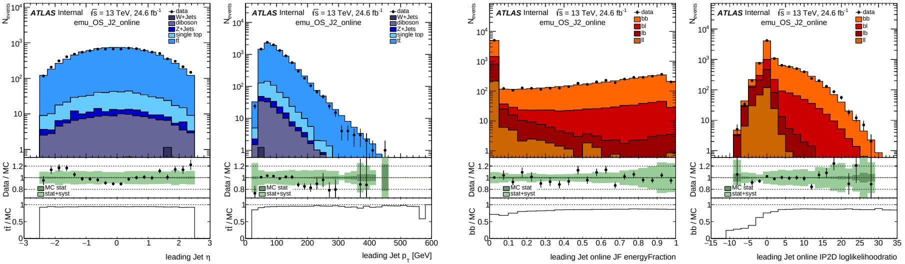


b-jet Trigger Calibration

b-jet Trigger Calibration

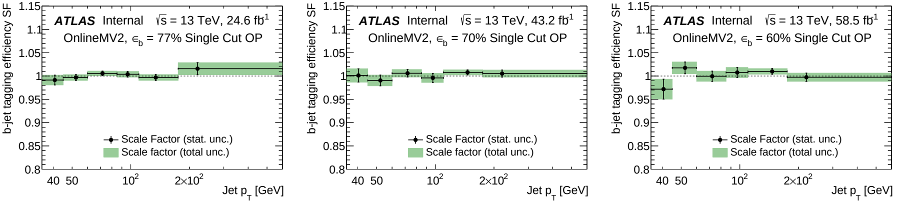


- b-tagging crucial to reduce multi-jet background
  - both at trigger level and at reconstruction level
  - since 2016 trigger level b-tagging use same algorithm as offline
- b-tagging efficiency is calibrated in $t\bar{t}$ dilepton events
  - sample with high purity of b-jets
- Event selection:
  - high p_T $e\mu$ with opposite charge
  - exactly 2 high p_T jets
- likelihood based method to extract b-tagging efficiency



b-jet Trigger Calibration

- b-tagging efficiency is extracted using a likelihood based method
- Combined tagger: `AND` combination between online and offline tagger
  - $P_f(comb) = P_f(trig \text{ AND } tag) = P_f(trig|tag)P_f(tag)$
  - $P_f(\overline{comb}) = P_f(\overline{trig} \text{ OR } \overline{tag}) = (1 - P_f(trig|tag)P_f(tag))$
- $P_f(tag)$ is given by the offline calibration
- $P_f(trig|tag)$ is evaluated by calibrating the online tagger in events with all jets tagged by the offline tagger
- Results obtained for full Run2
  - on the way to be available for full ATLAS collaboration



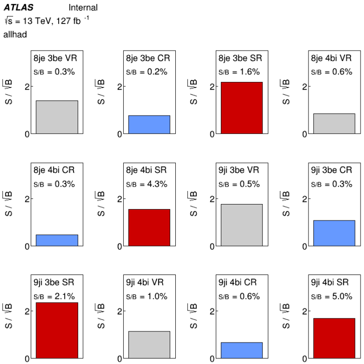
All Hadronic $t\bar{t}H(b\bar{b})$ Analysis

$t\bar{t}H(b\bar{b})$ Analysis Overview

Preselection requirements

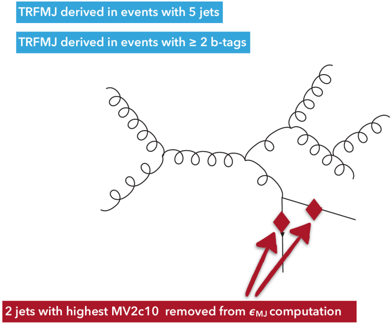
- b-jet trigger requiring ≥ 2 b-jet $+ \geq 2$ additional jets
- Lepton veto for orthogonality with other channels
- ≥ 5 high p_T jets
- ≥ 2 jets b-tagged by **combined** online+offline b-tagging

- Categorization in jet and b-tagged jet multiplicity
  - 4 regions considered
  - 8 or $\geq$9 jets
  - 3 or $\geq$4 b-jets
- Signal and background modeling
  - $t\bar{t}H$ signal: Powheg+Pythia8
  - $t\bar{t}$: Powheg+Pythia8
  - single top (Wt): Powheg+Pythia8
  - $t\bar{t}V$: aMC@NLO+Pythia8
  - **QCD multi-jet**: estimated with data-driven method TRF_{MJ}
    - ★ TRF_{MJ} is derived in region with exactly 5 jets and $\geq$2 combined b-tagged jets



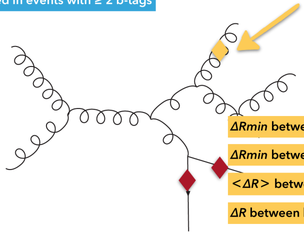
Tag-Rate-Function multi-jet: TRF_{MJ}

- TRF_{MJ} method is used to estimate the number of events with ($\geq$)k *b*-tagged jets from a sample with $\geq$n (n$\leq$k) *b*-tagged jets
- The probability of tagging a QCD jet, $\epsilon_{\mathbf{MJ}}$, is derived as a function of variables sensitive to heavy flavor production and *b*-tagging efficiency



Tag-Rate-Function multi-jet: TRF_{MJ}

- TRF_{MJ} method is used to estimate the number of events with $(\geq)$k *b*-tagged jets from a sample with $\geq$n (n$\leq$k) *b*-tagged jets
- The probability of tagging a QCD jet, $\epsilon_{\mathbf{MJ}}$, is derived as a function of variables sensitive to heavy flavor production and *b*-tagging efficiency



Probabilities of having n+m *b*-tag jets (m=0,1,2)

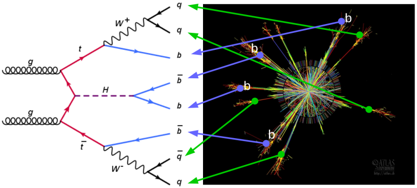
$$P_{m=0} = \prod_{i=1}^{N_{jet}-n} (1-\epsilon_i),$$

$$P_{m=1} = \sum_{j=1}^{N_{jet}-n} \left(\epsilon_j \prod_{i\neq j} (1-\epsilon_i) \right),$$

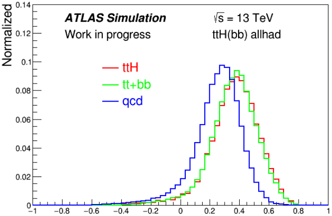
...

Reconstruction BDT

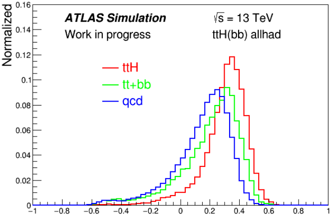
- **Goal**: find the best association between jets reconstructed in the detector and the final state partons
  - applied in events with $\geq$8 jets and $\geq$3 b-tag
  - large mutliplicites $\rightarrow$ large combinatorics: from 36 up to thousands of possible ways to reconstruct the $t\bar{t}H$ system



- 2 different BDTs using reconstructed resonances and angular correlations between jets
  - **recoBDT**: tries to reconstruct only $t\bar{t}$ system
    $\rightarrow$ no bias on the Higgs candidate mass
  - **recoBDT_withHiggs**: full $t\bar{t}H$ system reconstruction
    $\rightarrow$ higher reconstruction efficiency



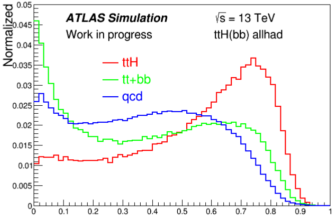
max recoBDT score in 9j,4b



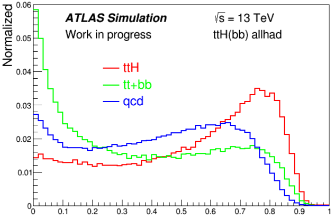
max recoBDT_withHiggs score in 9j,4b

Likelihood Discriminant Method

- Runs on all jet permutations to evaluate the event probability to be identified as signal ($t\bar{t}H$) or background ($t\bar{t} + b\bar{b}$)
- Perform a weighted average for the sig probability P_{sig} and bkg probability P_{bkg} of all permutations
- Final discinimant is obtained by the ratio: $LD = \frac{P_{sig}}{P_{sig}+P_{bkg}}$



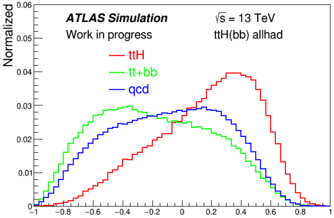
LHD in 8j,3b



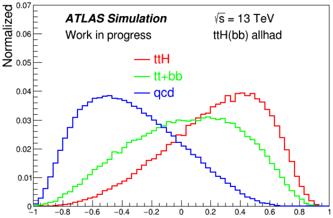
LHD in $\geq$9j,3b

2D Classification BDT

- **Goal**: perform signal vs background discrimination
- Combines reconstruction results from previous step with global event kinematics
- Trained two separate BDTs, optimized respectively against $t\bar{t}+b\bar{b}$ and QCD multi-jet backgrounds
- Variables optimization performed separately in each signal region with a recursive method
- Current use of the two BDTs:
  - Split each multiplicity region into control and signal regions with a cut at 0.3 on BDT_{MJ}
  - Perform final fit in $\text{BDT}_{t\bar{t}+b\bar{b}}$

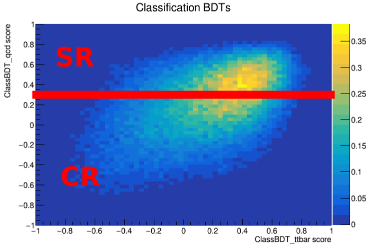


$\text{BDT}_{t\bar{t}+b\bar{b}}$ in 8j,3b

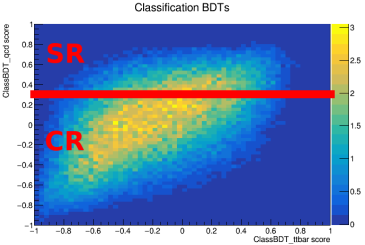


BDT_{MJ} in 8j,3b

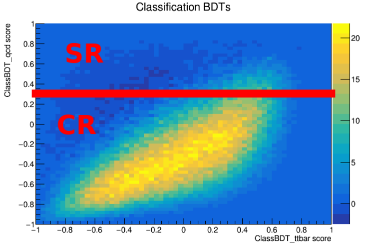
2D Classification BDT



$t\bar{t}H$ distribution in $\geq$9j,3b



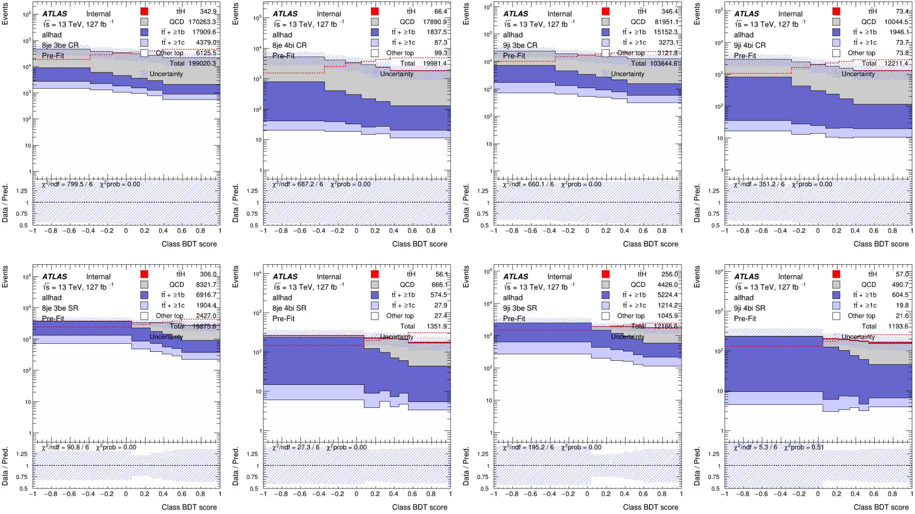
$t\bar{t} + b\bar{b}$ distribution in $\geq$9j,3b



QCD multi-jet distribution in in $\geq$9j,3b

- Cut at 0.3 on BDT_{MJ} is removing most of the QCD multi-jet background while keeping most of the signal

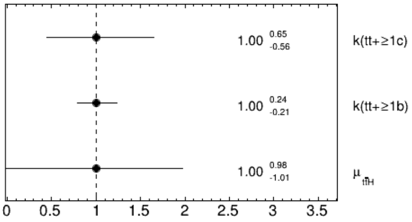
Run2 fit setup: BDT$_{t\bar{t}+b\bar{b}}$ Plots CR (top) and SR(bottom)

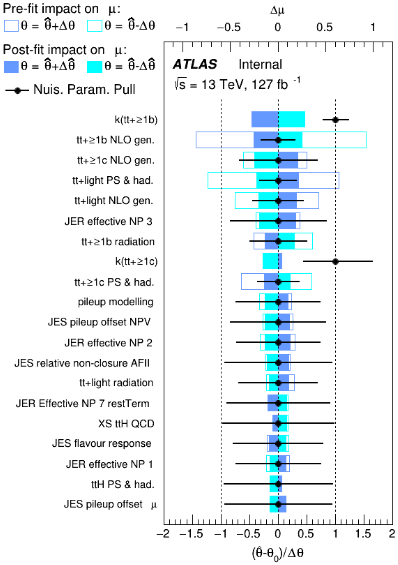


- In both CR and SR the fit is performed on the BDT$_{t\bar{t}+b\bar{b}}$
- Only simulation is used: obtain expected results before looking at data

Run2 Fit Preliminary Results

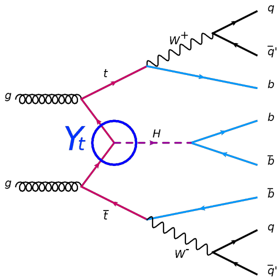
- Obtained expected signal strenght $\mu_{t\bar{t}H} = 1.00^{+0.98}_{-1.01}$
- Systematics are ranked based on their impact on the signal strenght $\mu_{t\bar{t}H}$
- Leading systematics from modeling of $t\bar{t}$ backgounds
- TRF_{MJ} systematics for QCD normalinzation highly constrained by the fit
  - currently low ranking
  - shape uncertainty may become leading one





Conclusion

Conclusion



- $t\bar{t}H$ production only way to directly measure the top-Higgs coupling Y_t
  - can be an important window for New Physics
- Fully hadronic $t\bar{t}H(b\bar{b})$ analysis:
  - Large statistic available and event fully reconstructable
  - but dominated by large QCD multi-jet background
- My contribution to improve the analysis:
  - Calibration of trigger *b*-tagging efficiency
    - performed for full Run2, results on the way to be available for full ATLAS collaboration
  - Implemented a 2 steps strategy for MVA based signal/background discrimination
    - **Reconstruction step** to resolve combinatorics: reconstruction BDT to find best combination and LHD to evaluate signal probability
    - **Classification step** with two separate optimizations: BDT$_{MJ}$ for QCD multi-jet and BDT$_{t\bar{t}}$ for $t\bar{t}$ background discrimination
- Obtained preliminary results for full Run2 Analysis
  - Expected signal strenght $\mu_{t\bar{t}H} = 1.00^{+0.98}_{-1.01}$
  - Leading systematics from modeling of $t\bar{t}$ backgounds
    - but still need to add shape uncertainties to TRF$_{MJ}$ predictions
- Next: look at p_T differential $\sigma_{t\bar{t}H}$ where this channel can give an important contribution

END

BACKUPS

Classification BDT versus $t\bar{t} + b\bar{b}$

- 8 jets 3 btag
  - LHD_Discriminant_merged
  - RecoBDT_withH_maxscore
  - Mass
  - M
  - DeltaRavgbb
  - St
  - HighestEt
  - Mbb_minDeltaR
- ≥9 jets 3 btag
  - LHD_Discriminant_allmatched
  - RecoBDT_withH_maxscore
  - Mass
  - M
  - DeltaRavgbb
  - St
  - HighestEt
  - Mbb_minDeltaR
  - Htjets
- 8 jets ≥4 btag
  - LHD_Discriminant_merged
  - RecoBDT_withH_maxscore
  - Mass
  - M
  - RecoBDT_ttbar_best_Higgs_mass
  - DeltaRavgbb
  - HighestEt
  - Mbb_minDeltaR
  - N30Higgs
- ≥9 jets ≥4 btag
  - LHD_Discriminant_allmatched
  - RecoBDT_withH_maxscore
  - Mass
  - M
  - DeltaRavgbb
  - Mbb_minDeltaR
  - RecoBDT_ttbar_best_Higgs_mass
  - Deltaetajjmax
  - TransverseMass

Classification BDT versus QCD

- 8 jets 3 btag
  - RecoBDT_ttbar_maxscore
  - LHD_log10ProbSig_merged
  - CentralityMass
  - AverageEtSinThetaStarNotTwoHighestEt
  - Ht5
  - MbbmaxPt
  - MinDR
  - TwobjetsMass
  - TransverseMass
  - MbbmaxM
  - Drbb_MaxPt
  - MbjmaxPt
  - Deltaetajjmax

- $\geq$9 jets 3 btag
  - RecoBDT_ttbar_maxscore
  - LHD_log10ProbSig_merged
  - CentralityMass
  - AverageEtSinThetaStarNotTwoHighestEt
  - Ht5
  - MbbmaxPt
  - MinDR
  - TwobjetsMass
  - TransverseMass
  - MbbmaxM
  - MbjmaxPt
  - MbjmaxM
  - MjjmaxM

- 8 jets $\geq$4 btag
  - RecoBDT_ttbar_maxscore
  - LHD_log10ProbSig_merged
  - CentralityMass
  - AverageEtSinThetaStarNotTwoHighestEt
  - Ht5
  - MbbmaxPt
  - MinDR
  - TwobjetsMass
  - MbbmaxM
  - MbjmaxPt
  - St
  - Htjets
  - Aplanarity

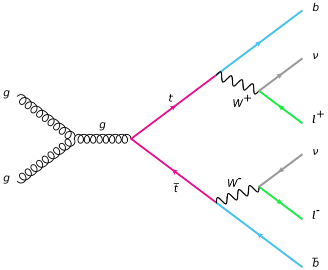
- $\geq$9 jets $\geq$4 btag
  - RecoBDT_ttbar_maxscore
  - LHD_log10ProbSig_merged
  - CentralityMass
  - AverageEtSinThetaStarNotTwoHighestEt
  - Ht5
  - MbbmaxPt
  - MinDR
  - TwobjetsMass
  - MbbmaxM
  - MbjmaxPt
  - N30Higgs
  - St
  - Aplanarity
  - Njet40
  - MjjmaxM

Run2 fit setup: Modeling and Systematics

- $t\bar{t}H$ signal: Powheg+Pythia8
- Backgrounds:
  - $t\bar{t}$: Powheg+Pythia8
  - single top (Wt): Powheg+Pythia8
  - $t\bar{t}V$: aMC@NLO+Pythia8
  - **QCD multi-jet**: TRF_{MJ} data-driven
- Instrumental systematics:
  - Luminosity, pileup modelling, JVT, JES, JER, flavour tagging on all MC
- Theoretical systematics:
  - cross section of $t\bar{t}+c$ and $t\bar{t}+b$ used as normalization factors
  - Uncertainties of cross section of MC backgrounds
  - Radiation: $t\bar{t}H$, $t\bar{t}$
  - Generator: aMC@NLO+Pythia8 $t\bar{t}H$, $t\bar{t}$ and single-top
  - PS+had: Powheg+Herwig7 $t\bar{t}H$, $t\bar{t}$ and single-top
- TRF_{MJ} unclosure systematics
  - Uncorrelated across jet and b-tag multiplicity
  - Normalization only, temporary set to 50%
  - Shape systematic has to be added

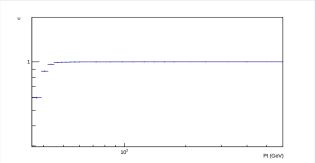
$t\bar{t}$ dilepton PDF Method: Event Selection

- Event selection:
  - activate one *b*-offperf trigger
  - exactly 2 tight leptons with $p_T > 28$ GeV and opposite charge
  - exactly 2 anti-kt4 calo jets with $p_T > 35$ GeV and $\eta < 2.5$
  - $e\mu$ channel
  - m_{lj} cuts
  - both jets matched



- Geometrical Matching
  - associate offline AntiKt4EMTopo jets to the corresponding online SplitJet jets with the geometrical requirement $\Delta R < 0.2$

Matching Efficiency

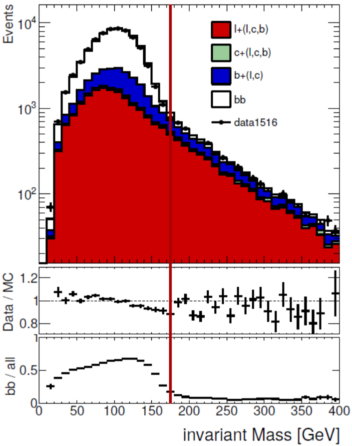


m_{lj} cuts

*slide from Julian Constantin Schmoeckel

- Idea: improve bb-purity by finding Jet + lepton combinations which corresponds to the top quarks.
- For *b*-Jets the invariant mass of the combination should be smaller then the top mass.
- The combination we found which seems to be the most promising in reducing background is the one which minimizes the sum of the squared invariant mass of both possible "lj-combinations" in the event.

lj-Combination, invariant Mass



- Veto events with one $m_{\mathrm{lj}} > 175\mathrm{GeV}$ ($\approx$ top-mass), or constrain flavor fractions

$t\bar{t}$ dilepton PDF Method: Likelihood Fit

*slide from Julian Constantin Schmoeckel

$$\begin{aligned}\mathcal{L}_{\mathrm{E}}(p_{\mathrm{T},1}, p_{\mathrm{T},2}, w_1, w_2|\mathcal{P}_{\mathrm{b}}(w|p_{\mathrm{T}})) =& [f_{\mathrm{bb}}(p_{\mathrm{T},1}, p_{\mathrm{T},2})\mathcal{P}_{\mathrm{b}}(w_1|p_{\mathrm{T},1})\mathcal{P}_{\mathrm{b}}(w_2|p_{\mathrm{T},2}) \\ &+ f_{\mathrm{bl}}(p_{\mathrm{T},1}, p_{\mathrm{T},2})\mathcal{P}_{\mathrm{b}}(w_1|p_{\mathrm{T},1})\mathcal{P}_{\mathrm{l}}(w_2|p_{\mathrm{T},2}) \\ &+ f_{\mathrm{lb}}(p_{\mathrm{T},1}, p_{\mathrm{T},2})\mathcal{P}_{\mathrm{l}}(w_1|p_{\mathrm{T},1})\mathcal{P}_{\mathrm{b}}(w_2|p_{\mathrm{T},2}) \\ &+ f_{\mathrm{ll}}(p_{\mathrm{T},1}, p_{\mathrm{T},2})\mathcal{P}_{\mathrm{l}}(w_1|p_{\mathrm{T},1})\mathcal{P}_{\mathrm{l}}(w_2|p_{\mathrm{T},2})]\end{aligned} \tag{1}$$

$f_{\mathrm{f}_1,\mathrm{f}_2}(p_{\mathrm{T},1}, p_{\mathrm{T},2})$ = fraction of flavour combination $[\mathrm{f}_1, \mathrm{f}_2]$. (Extracted from Simulation)
$\mathcal{P}_{\mathrm{f}}(w_1|p_{\mathrm{T},1})$ = pdf for a b-tagging weight w of jet with flavour f and a given $pT, 1$.

$$\mathcal{L}(\mathcal{P}_{\mathrm{b}}(w|p_{\mathrm{T}})) = \prod_{\mathrm{data}} \mathcal{L}_{\mathrm{E}}(\mathrm{data}|\mathcal{P}_{\mathrm{b}}(w|p_{\mathrm{T}})) \tag{2}$$

$$\epsilon_{\mathrm{b}}(p_{\mathrm{T}}) = \int_{w_{\mathrm{cut}}}^{\infty} dw' \mathcal{P}_{\mathrm{b}}(w\prime|p_{\mathrm{T}}) \tag{3}$$