Top quark mass measurement in lepton and jets final state at the CMS experiment

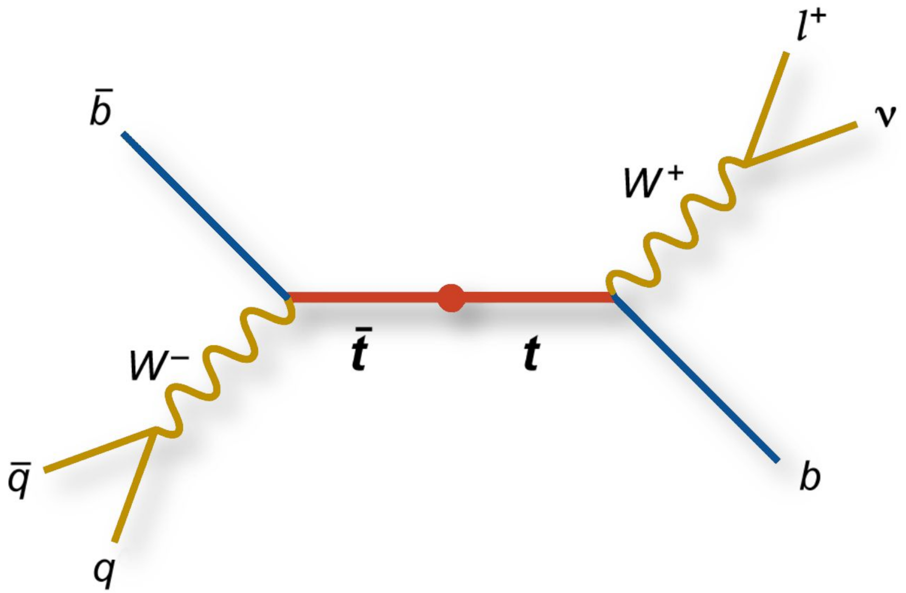
Mikael Myllymäki
Helsinki Institute of Physics

Spåtind 2023

January 6, 2023

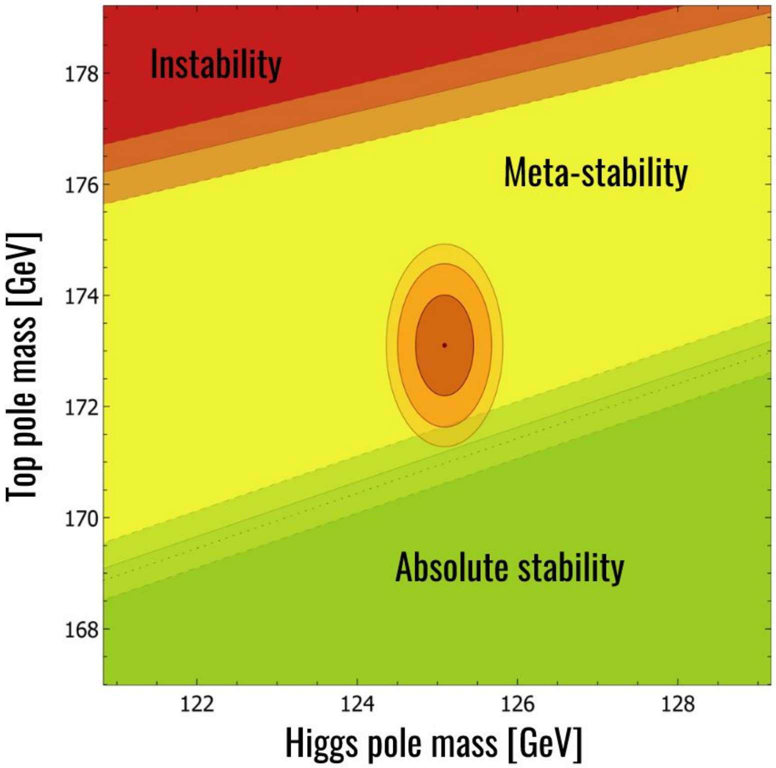
Overview

- Motivation
- Latest results
- Analysis strategy
  - event topology and the semileptonic channel
  - kinematic fit
  - binned profile likelihood method
- Future and current challenges
- Summary



Motivation

- top quark is the most massive particle in the SM
  - important probe for SM or ingredient for BSM models
- stability of the electroweak vacuum parameterized often as a function of the pole masses of Higgs and top
- “To rule out absolute stability to 3σ confidence, the uncertainty on the top quark pole mass would have to be pushed below 250 MeV” [1]



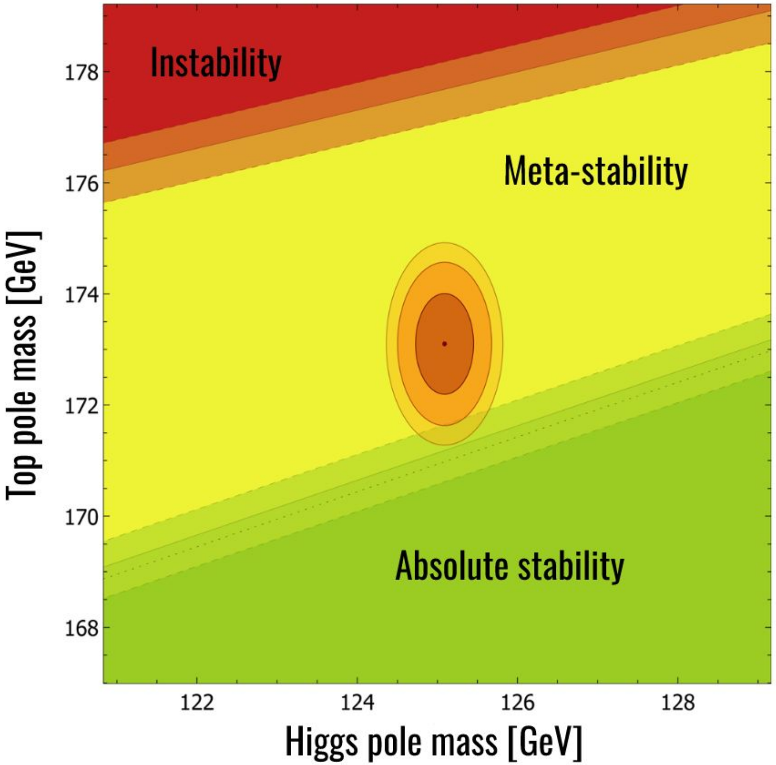
EW vacuum stability illustration [1]

[1] arxiv:1707.08124

Motivation

- top quark is the most massive particle in the SM
  - important probe for SM or ingredient for BSM models
- stability of the electroweak vacuum parameterized often as a function of the pole masses of Higgs and top
- “To rule out absolute stability to 3σ confidence, the uncertainty on the top quark pole mass would have to be pushed below 250 MeV” [1]

How close are we?



EW vacuum stability illustration [1]

[1] arxiv:1707.08124

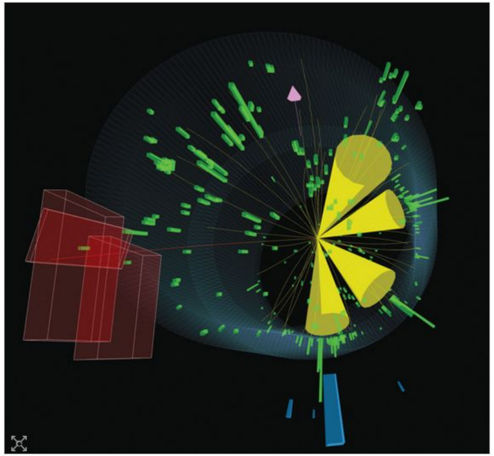
Previous results at LHC

- latest public result from 2022 with lepton+jets channel using 2016 data performed by the CMS Collaboration
  - $m_t = 171.77 \pm 0.38$ GeV [1]

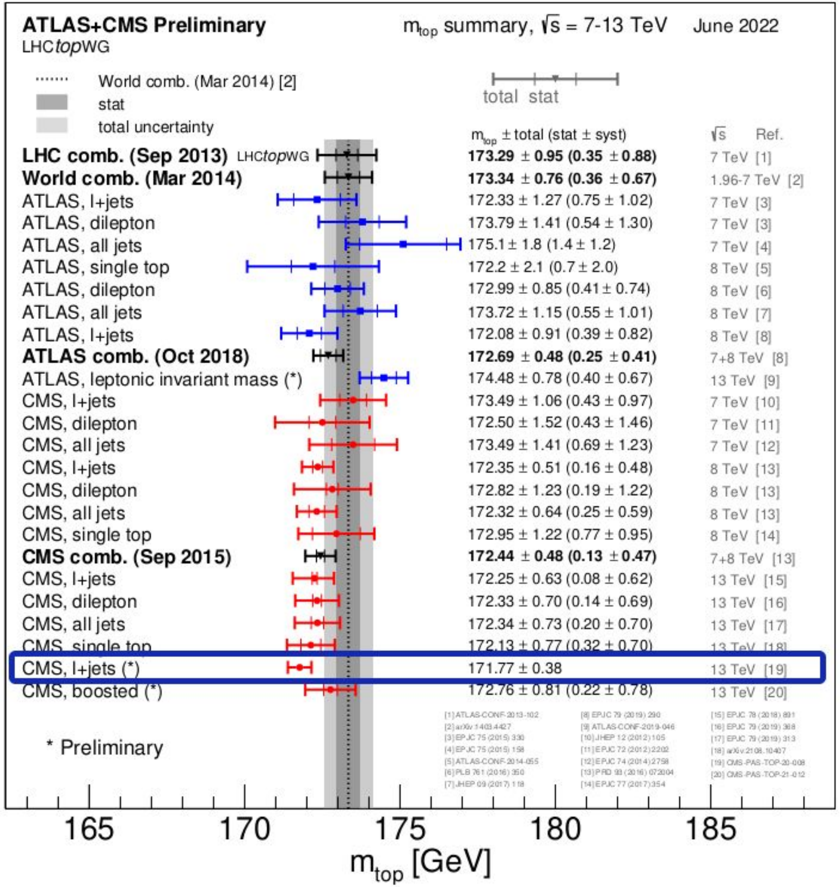
HIGGS AND ELECTROWEAK | NEWS

Top quark weighs in with unparalleled precision

1 July 2022



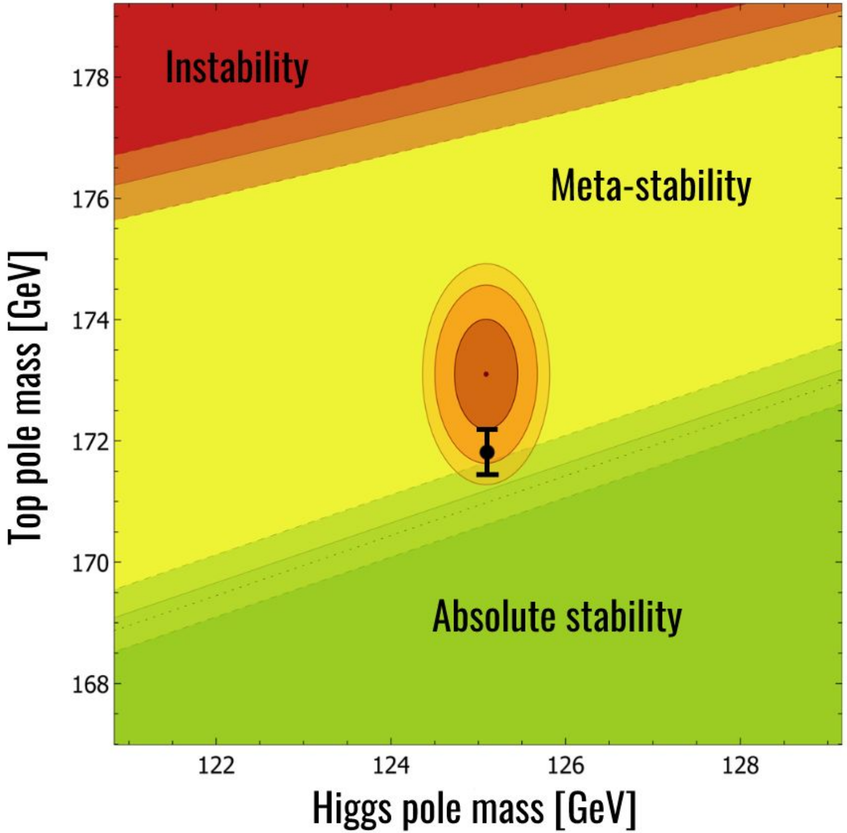
Top marks The classic signature of a top-quark pair at the LHC is four jets (yellow cones), one muon (red line and boxes) and missing energy from a neutrino (pink arrow). Credit: CMS



[1] [CMS-PAS-TOP-20-008]

Lepton + jets 2016 results

- 171.77 ± 0.38 GeV in the stability figure

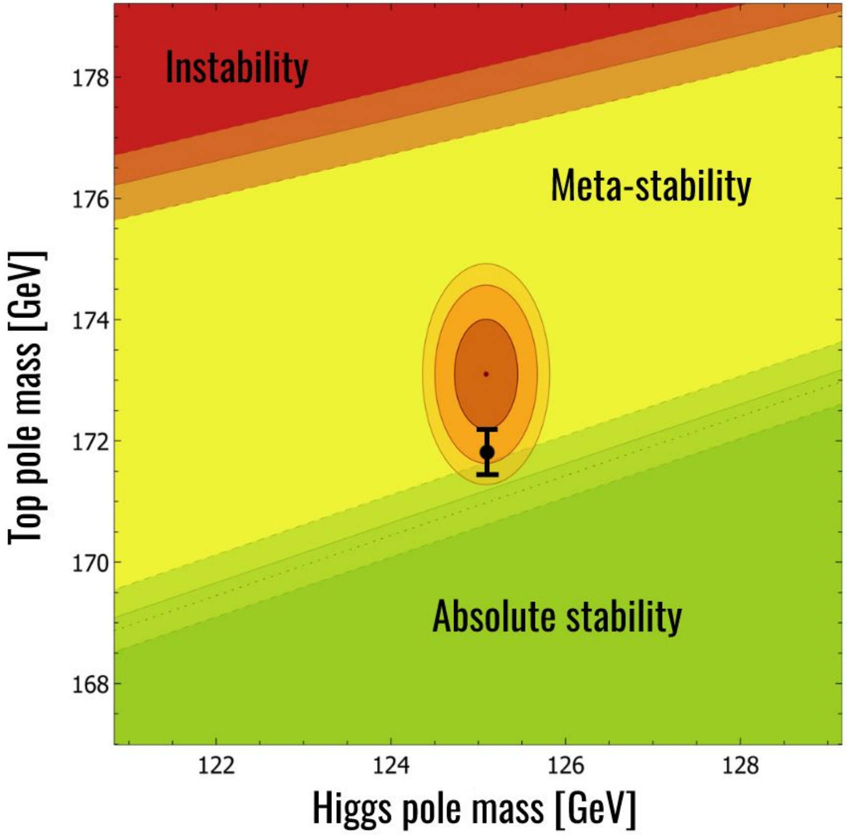


arxiv:1707.08124

Lepton + jets 2016 results

- 171.77 ± 0.38 GeV in the stability figure

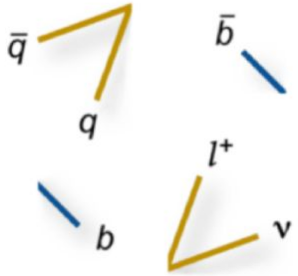
- our analysis in Helsinki extends this using data from 2017-2018
    - ~100 fb^{-1} → around 3 times more than 2016
    - systematic uncertainties can be reduced if well constrained by the data using **profile likelihood methods**



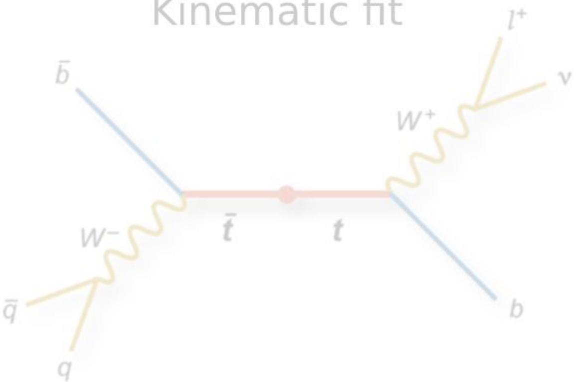
arxiv:1707.08124

Analysis strategy

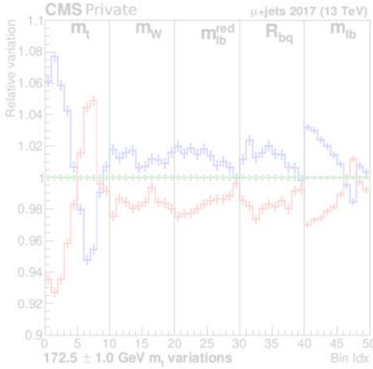
Object selections



Kinematic fit

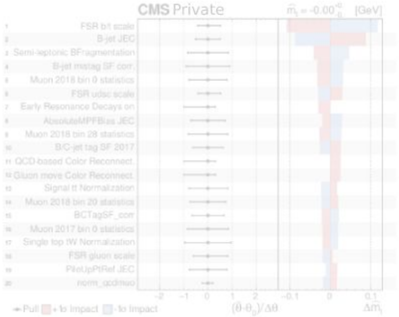


Binned profile likelihood fit



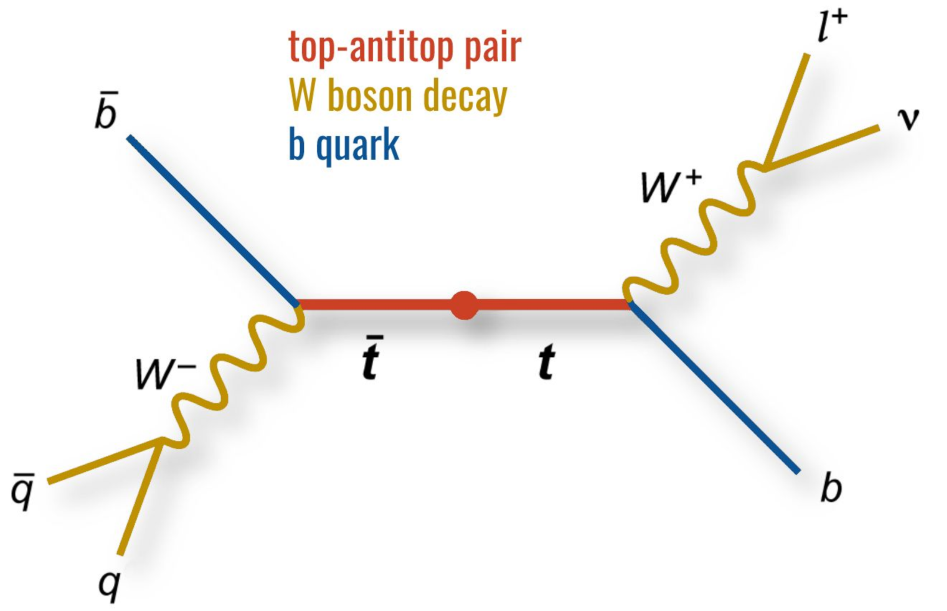
Results

17X.XX ± 0.XX GeV



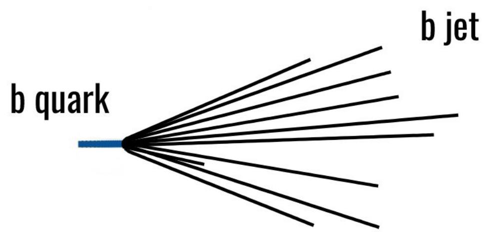
Mass measurement in ttbar topology

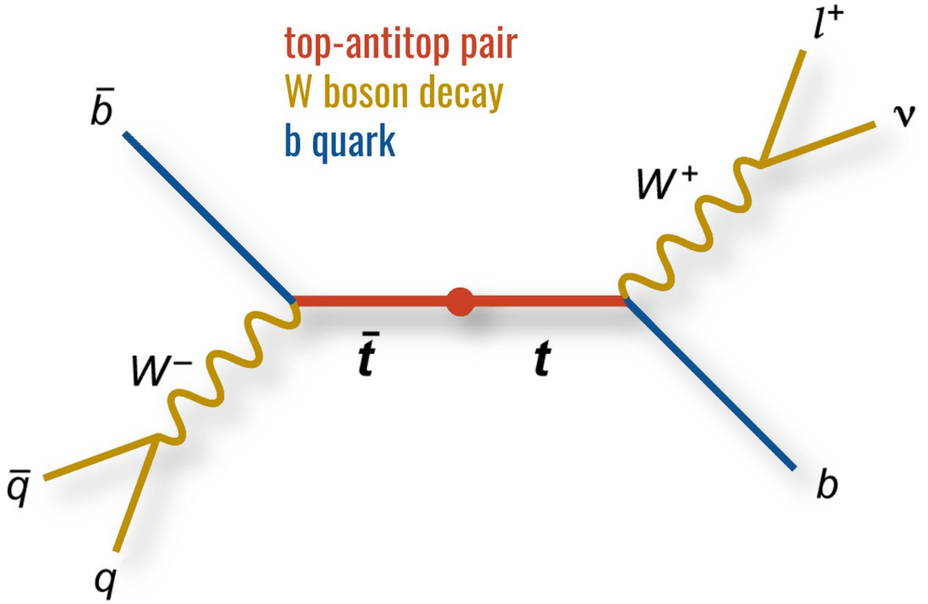
- direct mass measurement
  - mass from the reconstructed decay products
  - most precise method at LHC at the moment
- ttbar at 13 TeV LHC offers largest statistics
- top decays almost exclusively to W and b
- channels categorized by the W boson decays
  - dileptonic (W→l+ν, W→l+ν)
    + clear signal
    - missing momentum from neutrinos
  - hadronic (W→qq, W→qq)
    + largest statistics
    - difficult to distinguish jets
  - semileptonic channel (W→qq, W→l+ν)
    ± signal from the single lepton
    + only one neutrino
    ± decent statistics



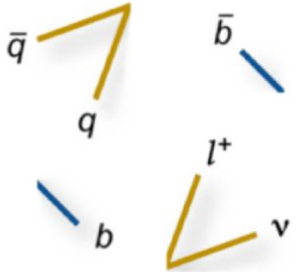
Semileptonic channel selections

- exactly one muon/electron
- at least 4 jets
  - of which 2 are b-jets
    - challenging to tag
  - of which 2 are light quark jets from W
    - boosted W complicates things
- neutrino inferred using missing momentum

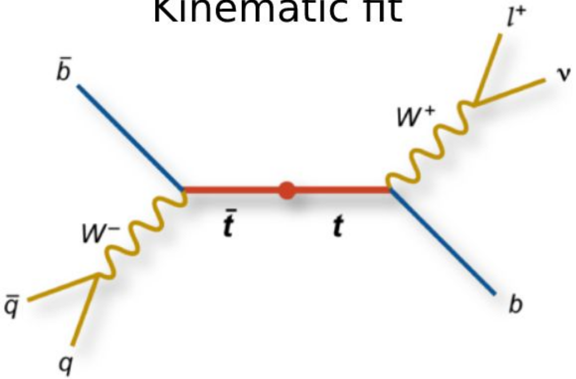




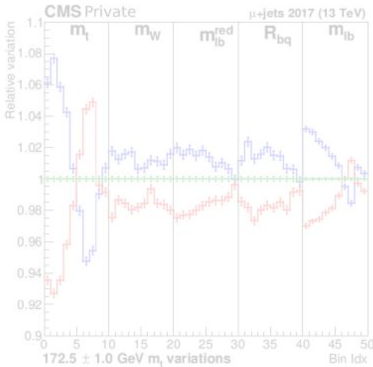
Object selections



Kinematic fit

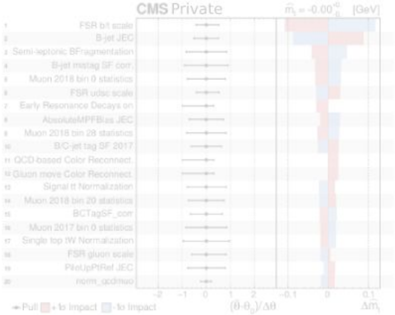


Binned profile likelihood fit



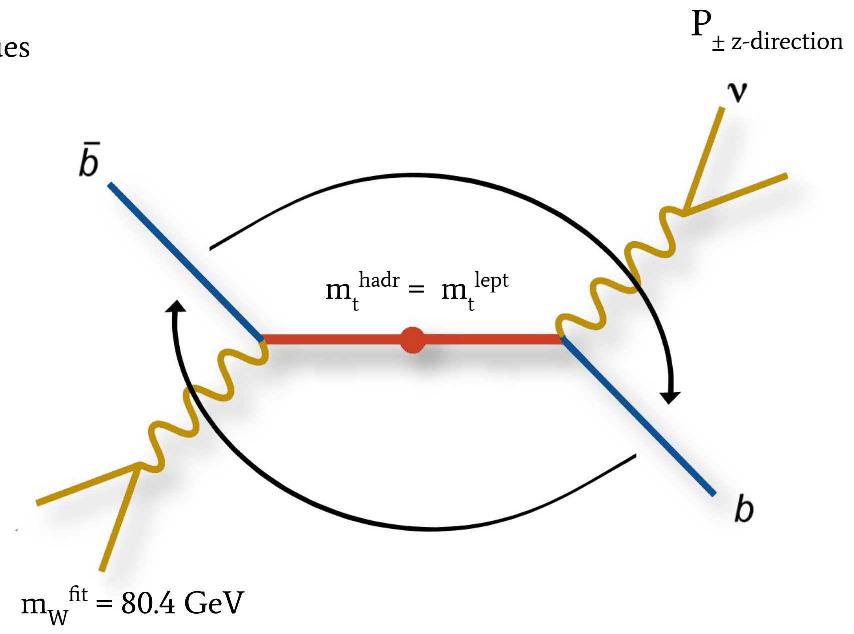
Results

17X.XX ± 0.XX GeV



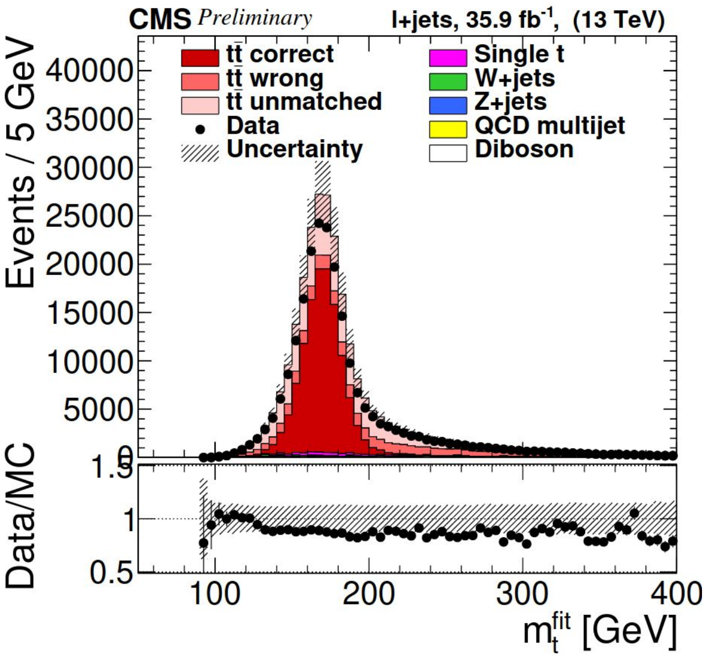
Kinematic fit

- semileptonic event hypothesis tested for combinations of selected objects
  - two possible combinations for b jets
  - neutrino momentum z-component has two possible values
- kinematic fit constraints:
  - $m_W^{fit} = 80.4$ GeV
  - $m_t^{hadr} = m_t^{lept}$
- gives χ^2 using object-parton resolution
  - example: further the reconstructed qq system from W mass, larger the χ^2
- goodness-of-fit for each permutation determined
  - $P_{gof} = \exp(-½\,\chi^2)$
  - hypothesis with the highest P_{gof} value is used

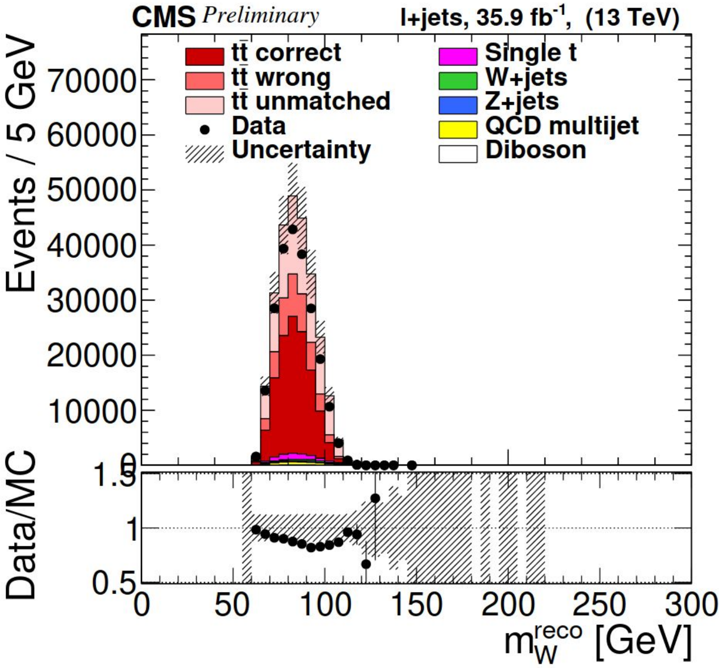


Control plots after kinematic fit

- data-simulation agreement can be confirmed

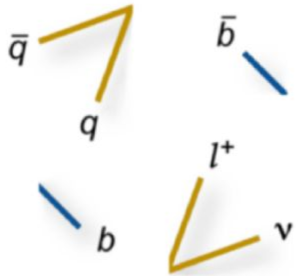


top mass from the fit

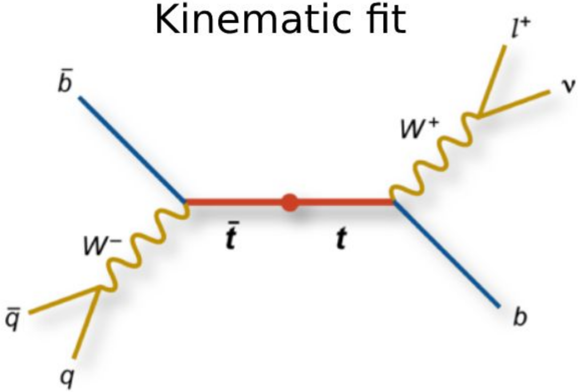


W mass

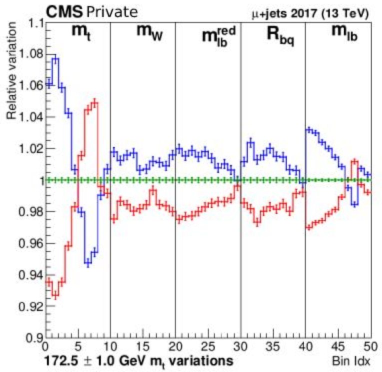
Object selections



Kinematic fit

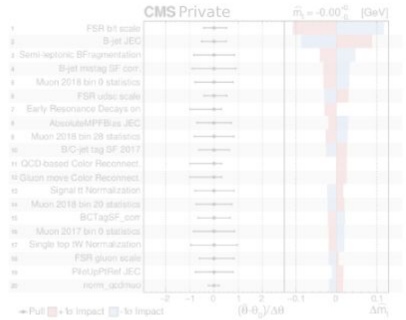


Binned profile likelihood fit



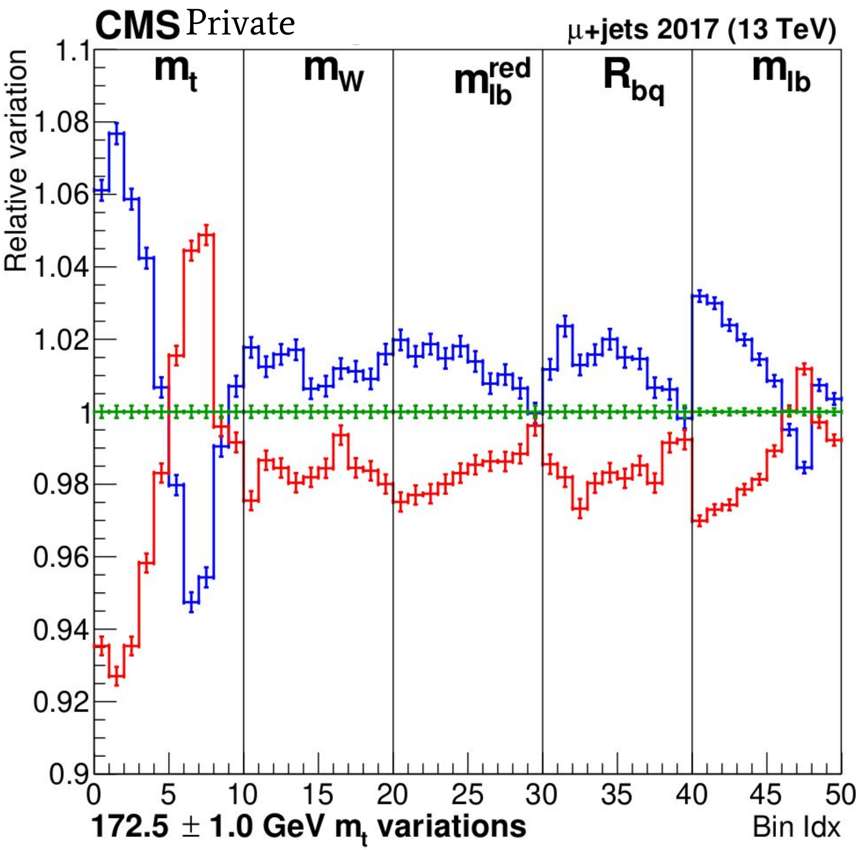
Results

17X.XX ± 0.XX GeV



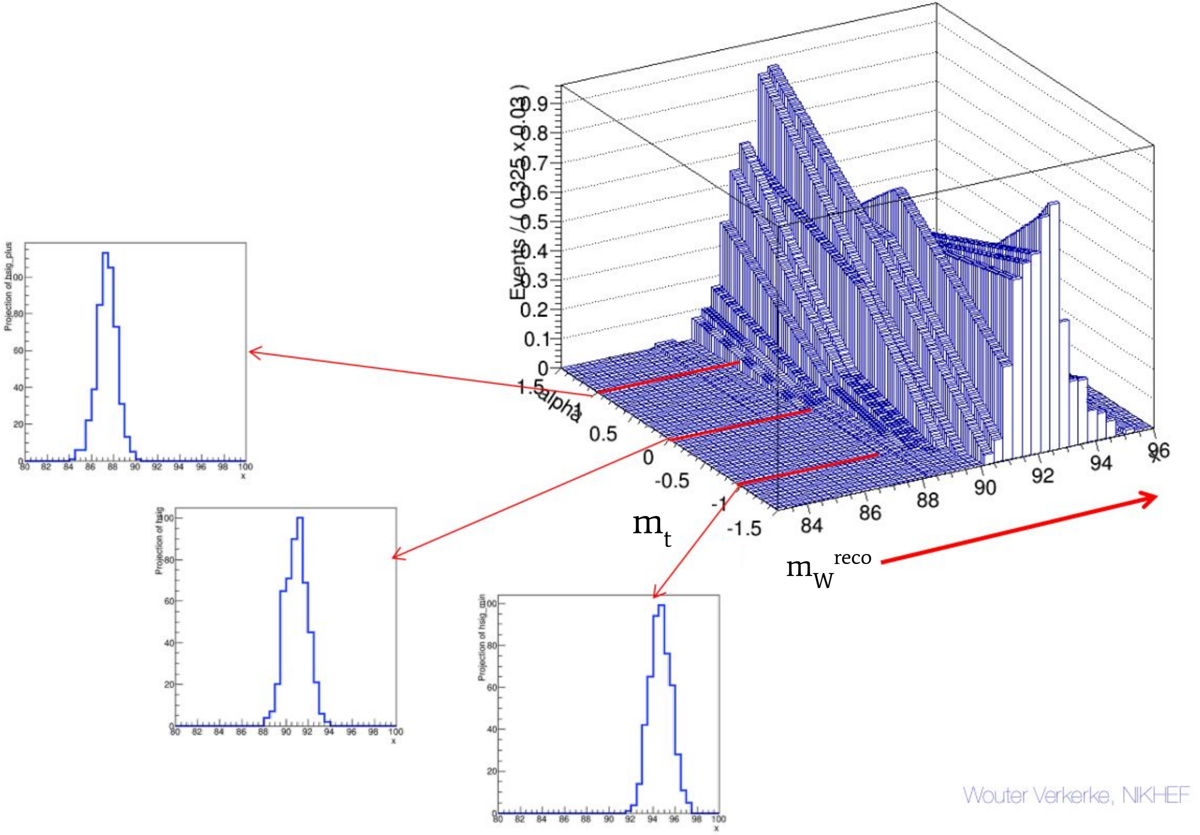
Binned profile likelihood fit

- m_t is the parameter of interest
- binned fit in five dimensions
    1. m_t^{fit} = obvious choice for m_t
    2. m_W^{reco} = jet energy scale of light quarks
    3. m_{lb}^{red}
    4. R_{bq}^{reco} = b-jet energy scale (originally from ATLAS)
    5. m_{lb}^{reco} $(P_{gof} < 0.2)$
- **Combine** tool used for the fitting
    - handles well year combinations
    - used extensively by Higgs community
    - nowadays also in precision measurements
- m_t inferred from the maximized likelihood



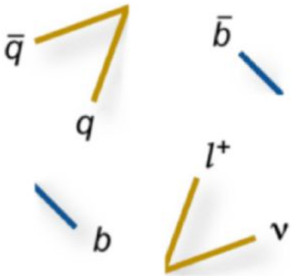
Interpolation example

- interpolation between central simulation and variations by histogram morphing
- linear interpolation between -1,0,1 variations

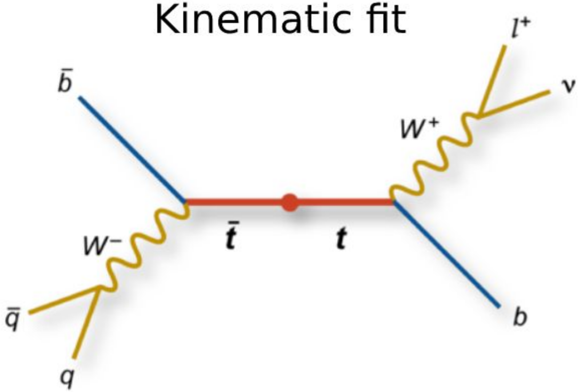


Source: [Wouter Verkeke - Advanced Statistics, Systematic uncertainties and profiling]

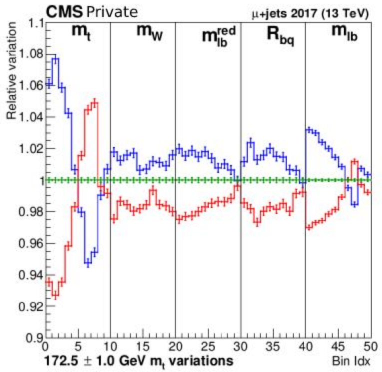
Object selections



Kinematic fit

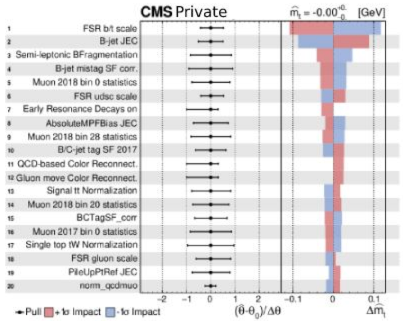


Binned profile likelihood fit



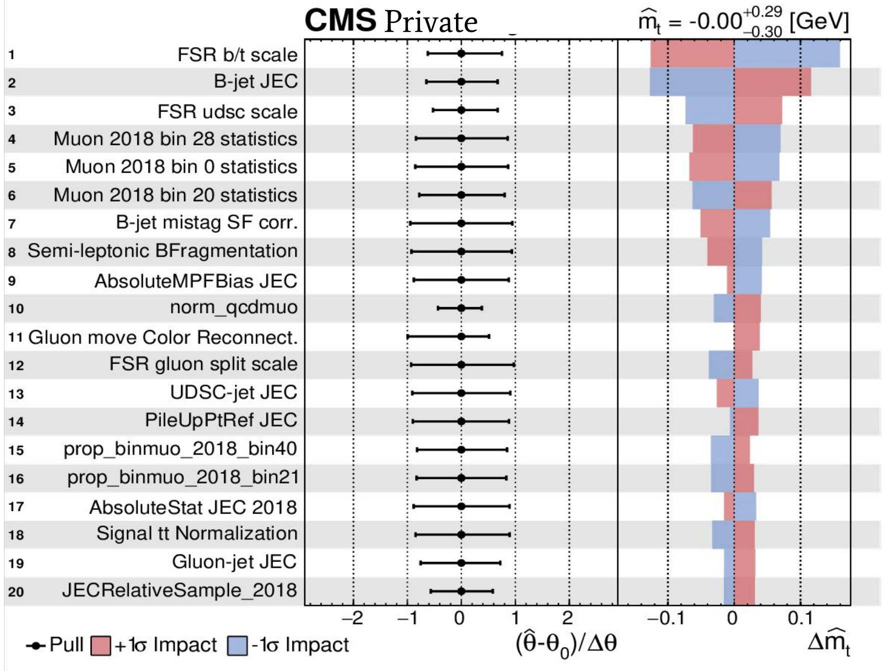
Results

17X.XX ± 0.XX GeV



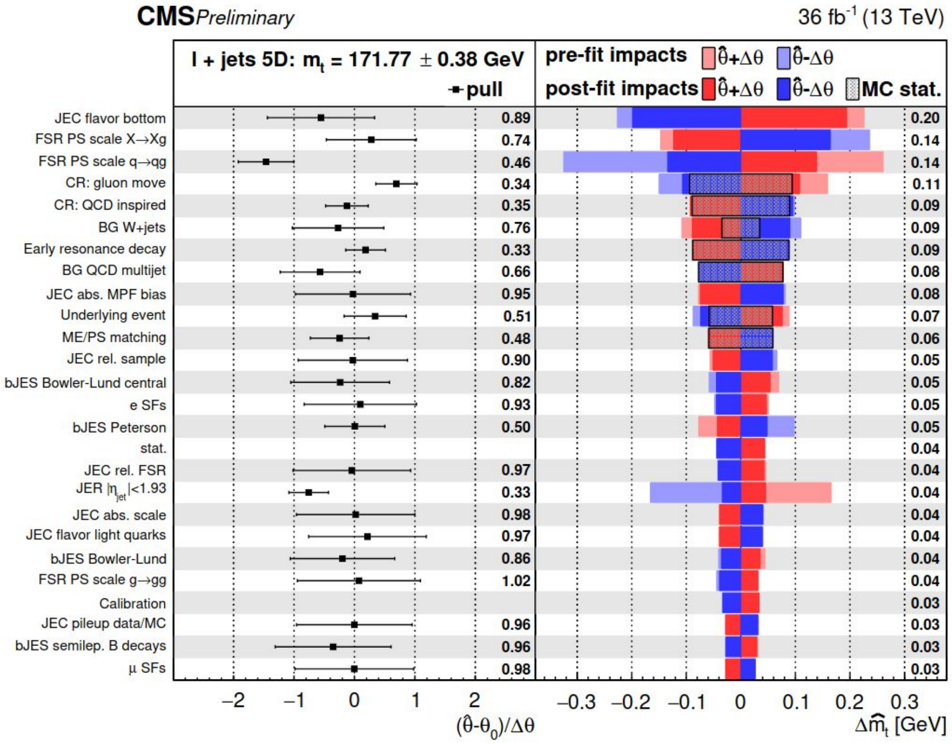
Impact plot for 2018

- uncertainty on the m_t by each nuisance parameter
- impact plotting
  - fit all nuisance parameters together
  - fixing single nuisance to ±1 value and refitting
  - impact = difference in the value of the m_t after fit
- in example plot we have blinded impact where the simulation is used as a target instead of data
  - impacts still estimate the uncertainty
- bJEC and b/t quark final state radiation (FSR) leading systematics



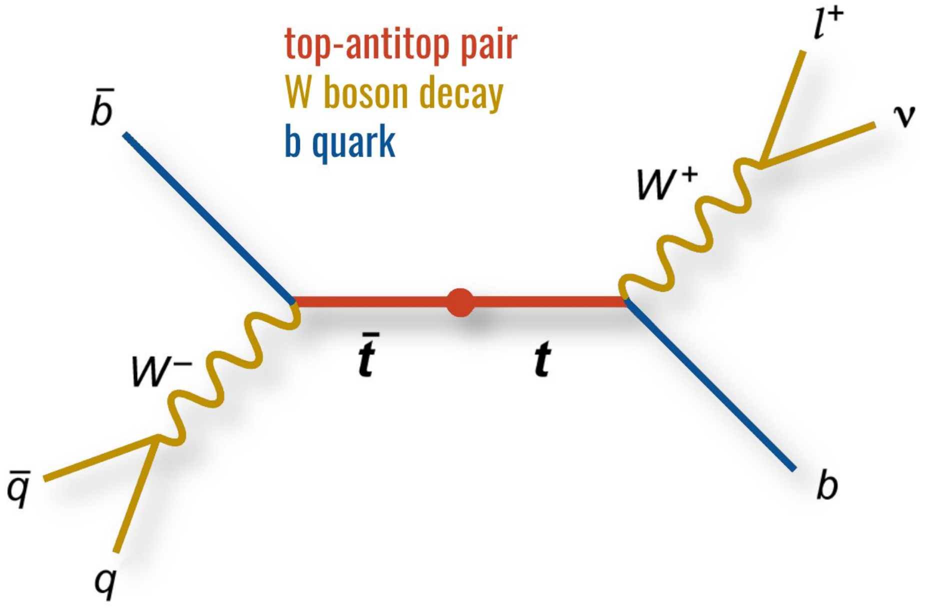
Impacts for 2016

- now fitted against real data
  - **pulls** indicate how different systematics should be shifted to best model the data
  - ideally pulls are below one sigma
- large pull in qFSR, opposite to bFSR
  - qFSR -1.5 σ
  - bFSR +0.3 σ
- uses split FSR scheme
  - light quarks and heavy quarks not correlated
- however, there is indication that
  - α_s^{FSR} **should be bigger in simulation**
  - **qFSR and bFSR are correlated**
  - these are being studied for 2017-2018 analysis



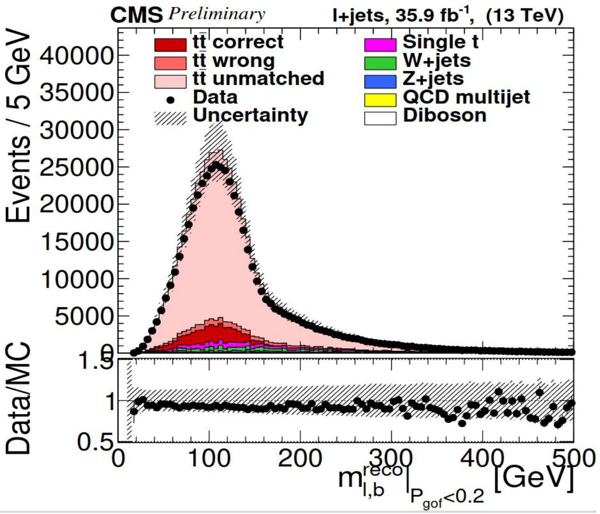
Summary

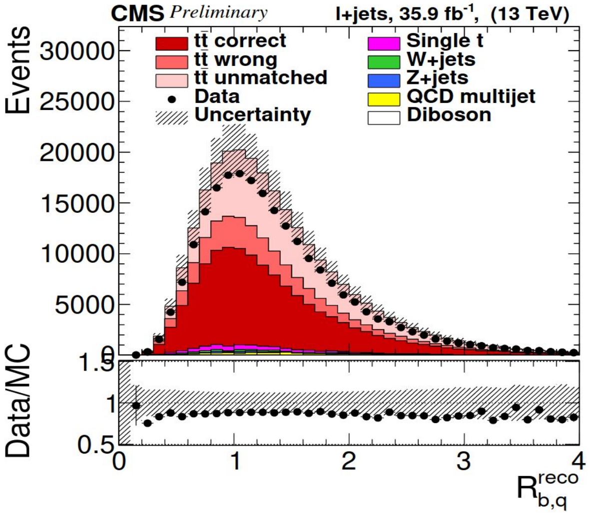
- latest public lepton + jets 2016 results presented
  - **most precise top mass measurement yet**
  - open questions related to FSR modeling
- we will aim to publish 2017-2018 results in 2023
  - around 3 times more data than 2016
  - some differences in the details of the analysis
- direct top mass measurement in ttbar semileptonic channel
  - event selections
  - kinematic fit
  - binned profile likelihood method

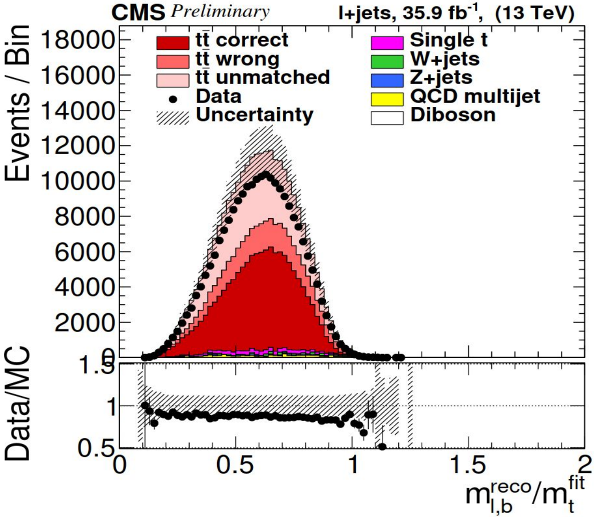


Backup

Backup: rest of the fit dimensions

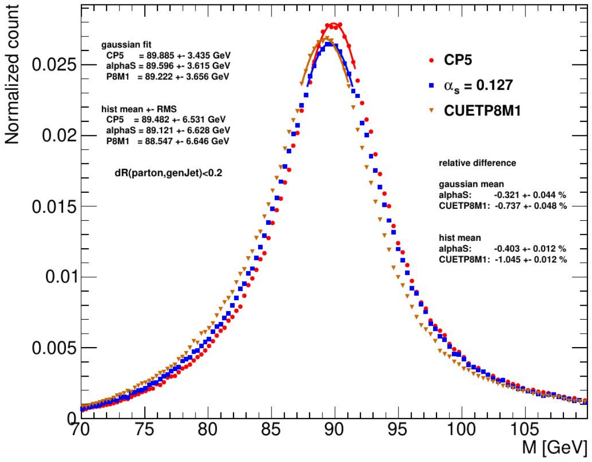




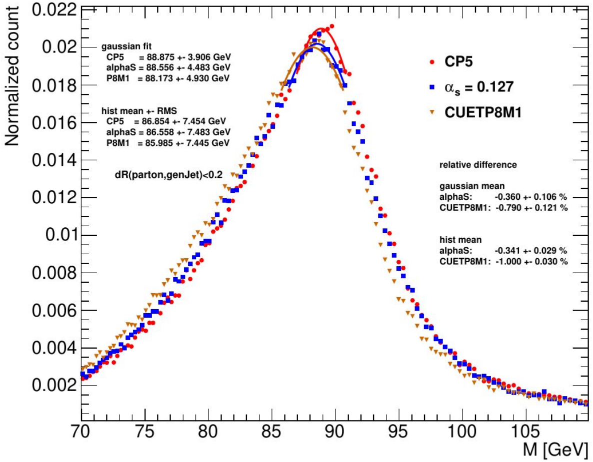


Backup: FSR correlation

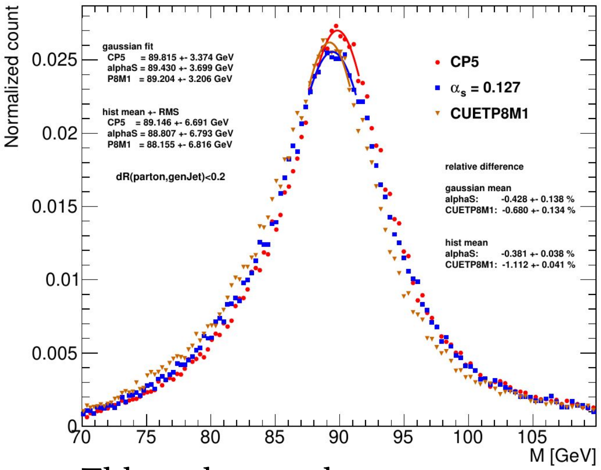
- new studies indicating correlation between light and b quark FSR in Pythia
- performed using Z->qq' decays and reconstructing Z-mass from the simulated quark jets



Zqq (no bb)

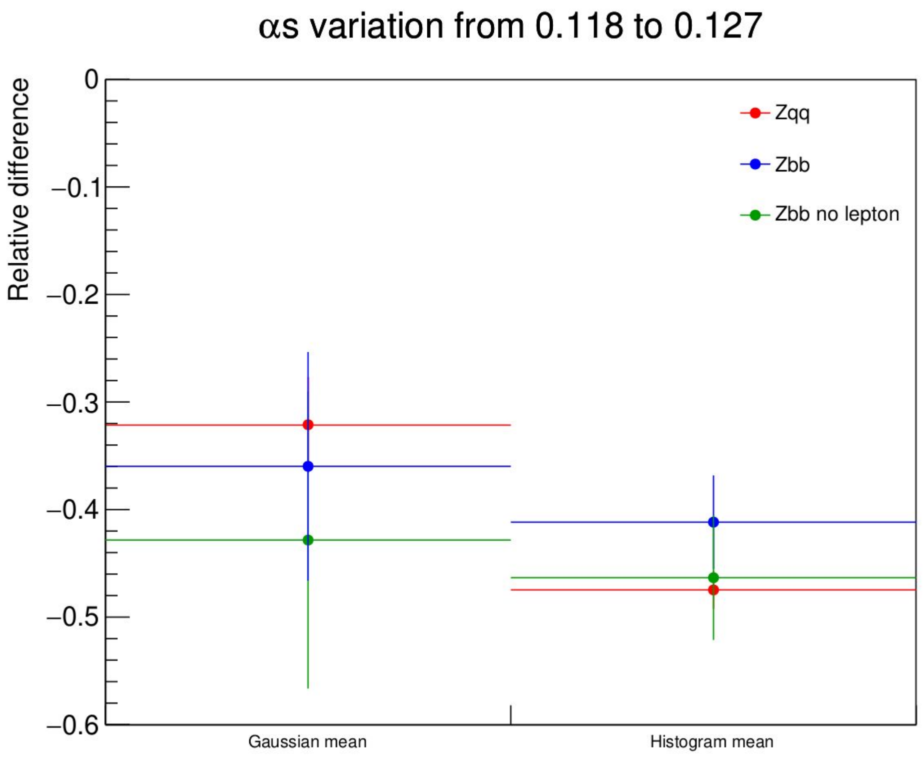


Zbb



Zbb no lepton decays

Backup: FSR correlation



Backup: Binned likelihood function

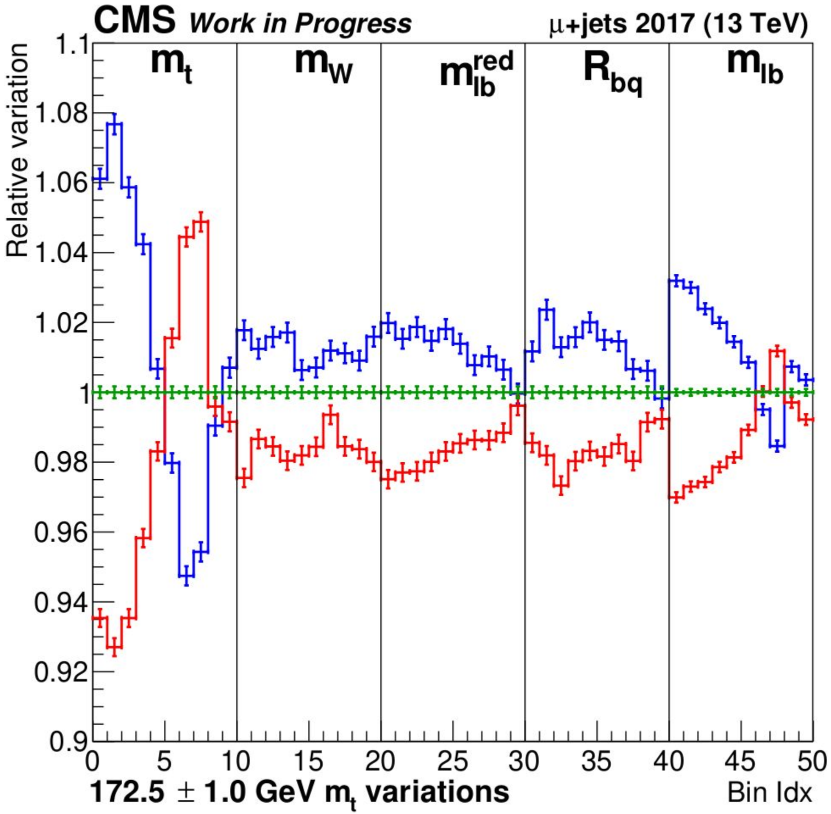
$$\mathcal{L}(\vec{n}) = \prod_{i\in\text{bins}} \mathcal{P}\left(n_i \middle| \sum_{j\in\text{samples}} (1+\kappa_j)^{\eta_j} \times \nu_i^j\left(\vec{\theta}, m_t\right)\right) \times \prod_{k\in\text{nuisances}} \mathcal{G}(\theta_k) \times \prod_{j\in\text{samples}} \mathcal{G}(\eta_j) \,.$$

- n_i = number of events in data in bin i
- ν_i^j = expected number of events for simulated sample j
- $P(n|\lambda)$ = Poisson probability to observe n events when λ predicted
- θ_k = nuisance parameter corresponding to a systematic uncertainty k
  - $\theta = 0$ is the central value and -1 and +1 are the down and up variations
- k_j = scaling uncertainty of simulated sample j
  - minimally cross-section uncertainty

- systematics interpreted as nuisance parameters
- nuisances constrained by normalized Gaussians $\mathcal{G}(\theta_k)$
  - up/down variations shift the likelihood by one sigma

Backup: Likelihood fit dimensions

- binned fit in five dimensions
  - m_t^{fit}
  - m_W^{reco}
  - $m_{lb}^{red} = m_{lb}^{reco} / m_t^{fit}$
  - $R_{bq}^{reco} = (p_T^{b\text{-jet 1}} + p_T^{b\text{-jet 2}}) / (p_T^{q\text{-jet 1}} + p_T^{q\text{-jet 2}})$
  - m_{lb}^{reco} ($P_{gof} < 0.2$)



Backup: top mass definition

- top quark mass is theoretically complex to define
- what is measured in the direct measurements is the so-called Monte Carlo mass m_t^{MC}
- relationship with field theoretic mass definitions such as the pole mass m_t^{pole} is not straightforward

“top mass interpretation problem”

- theoretical work (A. Hoang [1]) has been conducted on this topic and as the precision of the measurements increases this additional uncertainty must be given serious thought

[1] http://dx.doi.org/10.1146/annurev-nucl-101918-023530