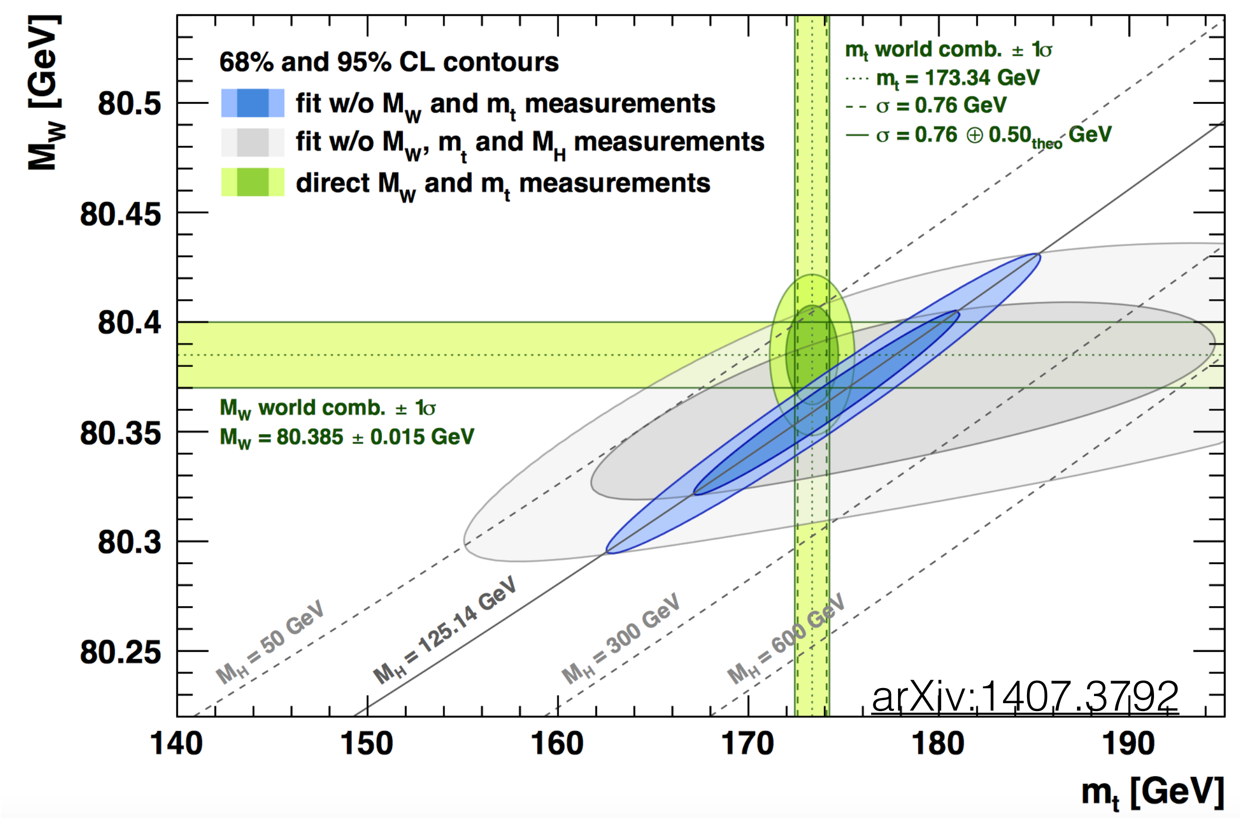
Run-1 ATLAS+CMS Top Mass Combination

Mark Owen
On behalf of ATLAS & CMS
LHCtopWG meeting 25 April 2024

See: arXiv:2402.08713 (accepted by PRL)

Why measure m_t?

- Top quark mass is a fundamental parameter of the Standard Model:
  - Precise measurement needed for checking consistency of the SM.



- LHC Run-1 produced large amount of top quarks and multiple top mass measurements - ATLAS and CMS individual combinations reached 0.5 GeV precision.

- Last LHC (preliminary) combination done as part of world average in 2013 [1] - misses most of the precise 8 TeV measurements.

[1] ATLAS-CONF-2014-008

Outline

- Methodology & systematic categorisation
- ATLAS & CMS input measurements
- LHC combination results
- Cross-checks

Methodology & systematic categorisation

Methodology

- Use Best Linear Unbiased Estimator method (BLUE) [1].
  - For measurements of uncertainty σ_i which have correlation coefficients ρ_{ij}, this provides the unbiased linear estimator of the physics parameter with the smallest uncertainty.

$$m_t = \sum_i w_i m_t^i; \ \ \sum w_i = 1$$

  - Uncertainty of each measurement is easy to get from the papers.
  - Must calculate / estimate the correlation between the measurements. BLUE then calculates the weight of each measurement and corresponding uncertainty on physics parameter.
- Correlation estimation:
  - Split systematic uncertainties into sources, assign / assess correlations between each pairs of measurements for each source.
- Procedure already used for ATLAS & CMS individual combinations & previous preliminary world combination.
  - Challenge here is the inter-experiment combinations.

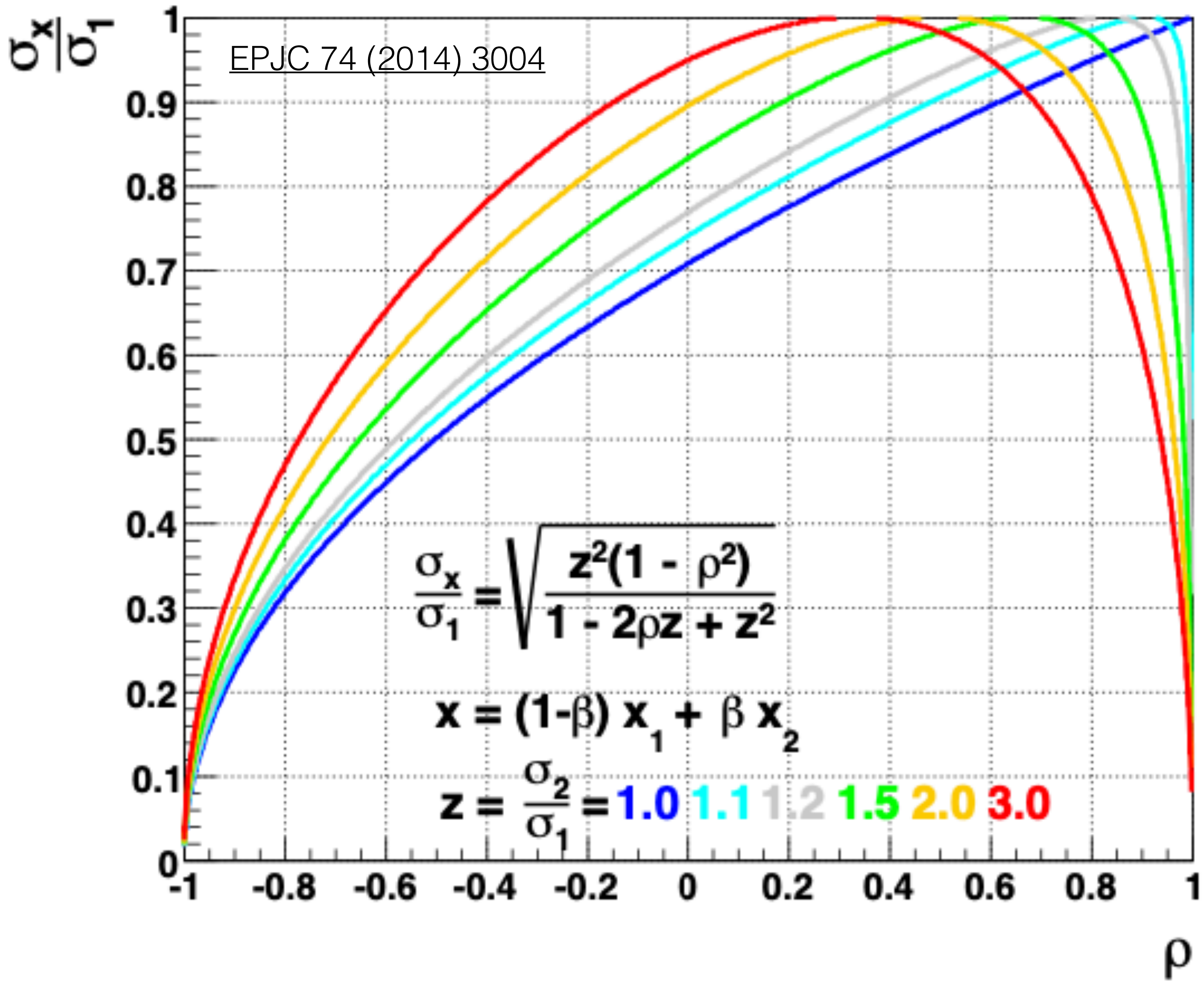
[1] Nucl. Instr. and Meth. A 270 (1988) 110

Methodology

- BLUE is rather simple for two measurements:

$$x = (1-\beta)x_1 + \beta x_2 \qquad \beta = \frac{1-\rho z}{1-2\rho z+z^2} \qquad z = \frac{\sigma_2}{\sigma_1} \geq 1$$

E.g. gives weight = 1/2 for simple case $\rho = 0,\ z = 1$



- BLUE always gives combined uncertainty as good or better than best input measurement.
- Extent to which combination improves on individual measurements depends on precision (z) & correlation (ρ) of measurements.
- Weights can be negative.
- Note, that taking very strong +ve correlation is not necessarily conservative.
- More info in [1].

[1] EPJC 74 (2014) 3004

Uncertainty categorisation

- Ideally, would be able to map every potentially correlated ATLAS systematic uncertainty to a CMS one.
  - Not possible due to different methods, MC, detectors etc.
- Instead, setup categories that reflect common uncertainty sources and then use physics judgement to assign correlation across categories.
- Signs of uncertainties are tracked - where signs of impact of uncertainties are negative then these are kept (effective negative correlation). Was already the case in ATLAS combination, treatment is new for CMS (and ATLAS+CMS).
- Correlations generally not perfectly known, so then scan around nominal to test sensitivity to the assumptions made.

Uncertainty categorisation - JES

- Most well understood sector - real benefit of ATLAS+CMS JES correlation studies done as part of LHCtopWG [1,2]

Uncertainty category	ρ	Scan range
JES 1	0	—
JES 2	0	$[-0.25, +0.25]$
JES 3	0.5	$[+0.25, +0.75]$
b-JES	0.85	$[+0.5, +1]$
g-JES	0.85	$[+0.5, +1]$
l-JES	0	$[-0.25, +0.25]$

- JES 1: statistical, pileup and time-dependent variations expected to be uncorrelated.
- JES 2: absolute JES from $\gamma/Z+$ jets events. Significant differences between ATLAS and CMS - assume uncorrelated.

[1] ATL-PHYS-PUB-2015-049 / CMS PAS JME-15-001
[2] ATL-PHYS-PUB-2014-020 / CMS PAS JME-14-003

Uncertainty categorisation - JES

Uncertainty category	ρ	Scan range
JES 1	0	—
JES 2	0	$[-0.25, +0.25]$
JES 3	0.5	$[+0.25, +0.75]$
b-JES	0.85	$[+0.5, +1]$
g-JES	0.85	$[+0.5, +1]$
l-JES	0	$[-0.25, +0.25]$

- JES 3: relative η intercalibration. Uncertainties from generator modelling of radiation patterns - partially correlated.

- b-JES: jet energy response uncertainty for b-jets. Derived from similar MC comparisons - strong correlation.

Uncertainty categorisation - JES

Uncertainty category	ρ	Scan range
JES 1	0	—
JES 2	0	$[-0.25, +0.25]$
JES 3	0.5	$[+0.25, +0.75]$
b-JES	0.85	$[+0.5, +1]$
g-JES	0.85	$[+0.5, +1]$
l-JES	0	$[-0.25, +0.25]$

- g-JES: jet response for gluons (CMS), relative gluon-to-light quark jet response (ATLAS). Similar MC comparisons - strong correlation.
- l-JES: light-quark jet response (CMS), jet flavour composition (ATLAS). Different uncertainty sources - uncorrelated.

Uncertainty categorisation - MC modelling

- Non-trivial differences, plus nominal MC are different (Powheg vs Madgraph).

Category	ATLAS	CMS	Correlation
ME generator	Powheg vs MC@NLO	Madgraph vs Powheg	0.5
QCD radiation	ISR/FSR modelling variations in P+P6	Factorisation / renormalisation scale and matching scale variations in Madgraph	0.5
Hadronization	Powheg+Pythia vs Powheg+Herwig at analysis level	Vary b-fragmentation model in Pythia	0.5
Semi-leptonic BR	-	Vary semi-leptonic BR	-
Colour reconnection	Perugia2012-LoCR	Perugia2011-NoCR	0.5
Underlying event	Perugia2012 mpiHi tune	Perugia 2011 mpiHi & Perugia Tevatron tunes	0.5
PDF	PDF4LHC	PDF4LHC	0.85
Top pT	- (assumed covered by Herwig sample)	Reweighting to 8 TeV pT distribution (8 TeV results only)	-

Uncertainty categorisation - experimental

- Generally assume 0 correlation (different detector & independent calibrations).

Category	Correlation
Jet energy resolution	0
Lepton energy scale / resolution / efficiency	0
b-tagging	0.5
MET	0
Pileup	0.85
Trigger (non-lepton analyses)	0
Background (data)	0
Background (simulation)	0.85
Method / calibration	0

B-tagging calibrations both use similar methods using di-jet events, ATLAS also uses $t\bar{t}$ events

Pileup modelling similar between experiments - 7 & 8 TeV are uncorrelated due to different conditions.

Backgrounds from simulation (W+jets, Z+jets) are similar, take correlated.

ATLAS & CMS input measurements

ATLAS inputs

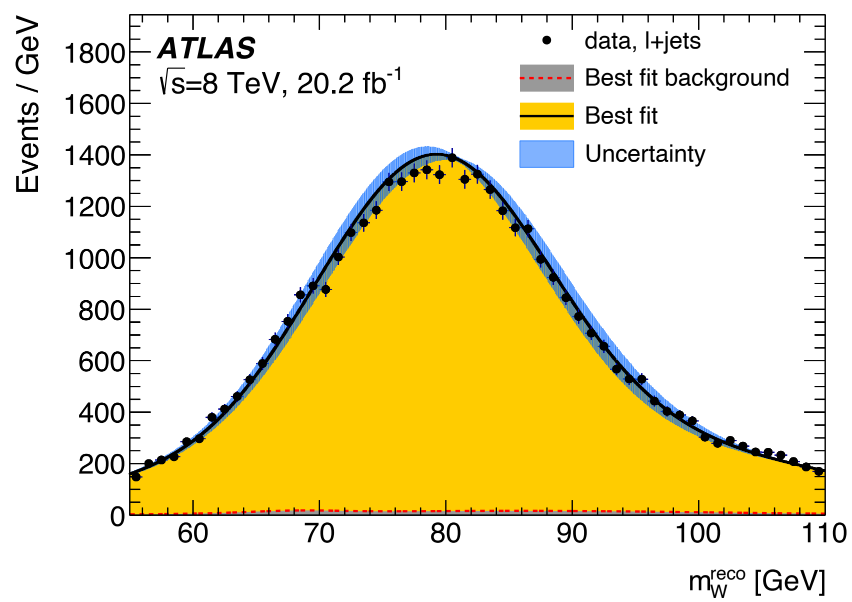
- Same 6 measurements as entered the published combination:
  - Lepton+jets, dilepton, all-jets at 7 & 8 TeV.

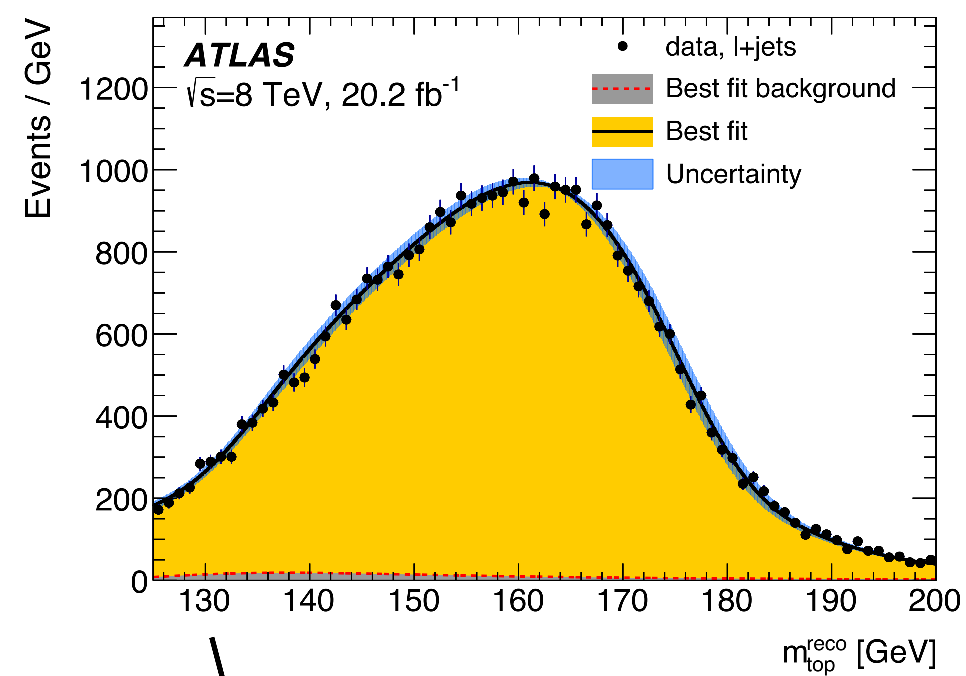
EPJC 79 (2019) 290

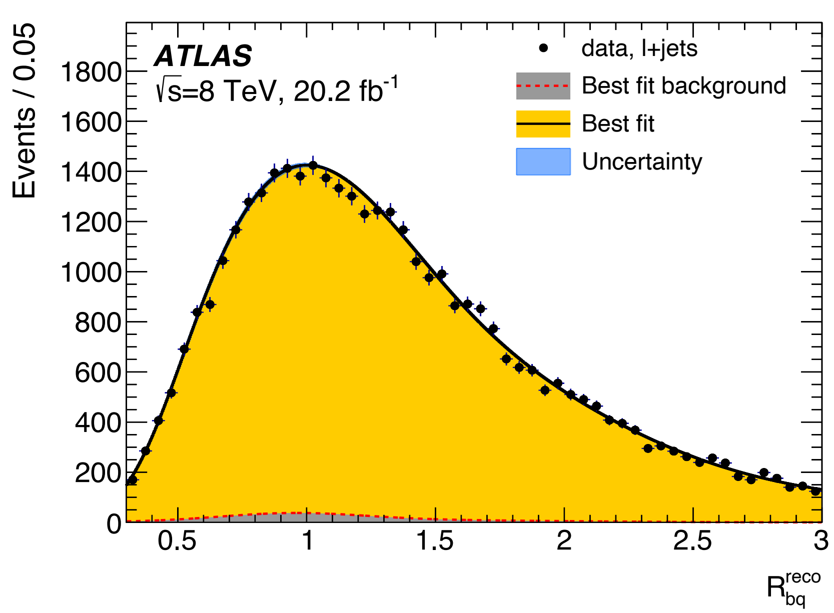
ATLAS inputs

- Same 6 measurements as entered the published combination:
  - **Lepton+jets**, dilepton, all-jets at 7 & **8** TeV.

EPJC 79 (2019) 290







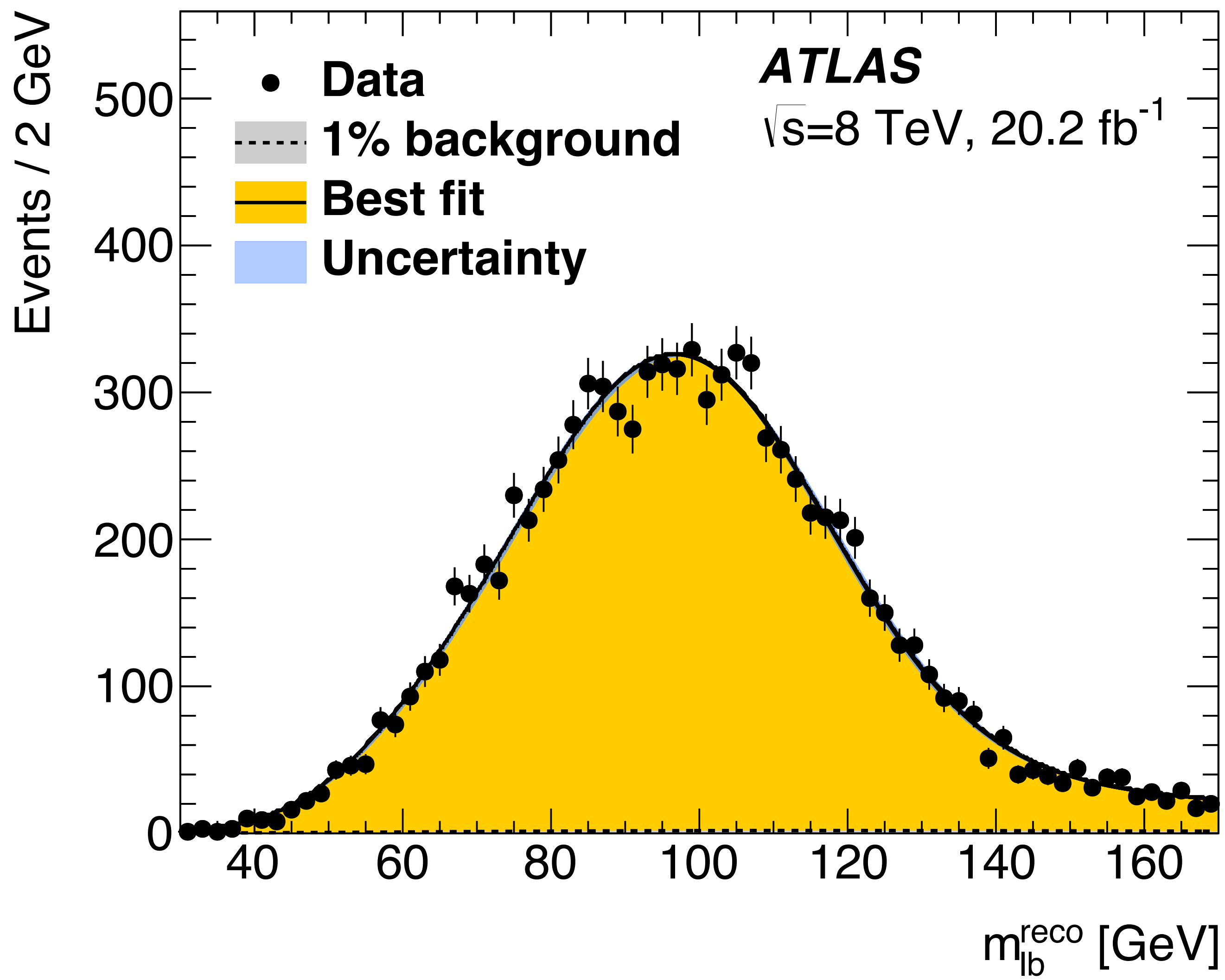
$$m_t = 172.08 \pm 0.91 \ \left[0.39\ (\text{stat}) \ \pm 0.82\ (\text{syst})\right] \ \text{GeV}$$

b-JES only 50 MeV, non-zero statistical uncertainty due to 3D fit

ATLAS inputs

- Same 6 measurements as entered the published combination:
  - Lepton+jets, **dilepton**, all-jets at 7 & **8** TeV.

PLB 761 (2016) 350



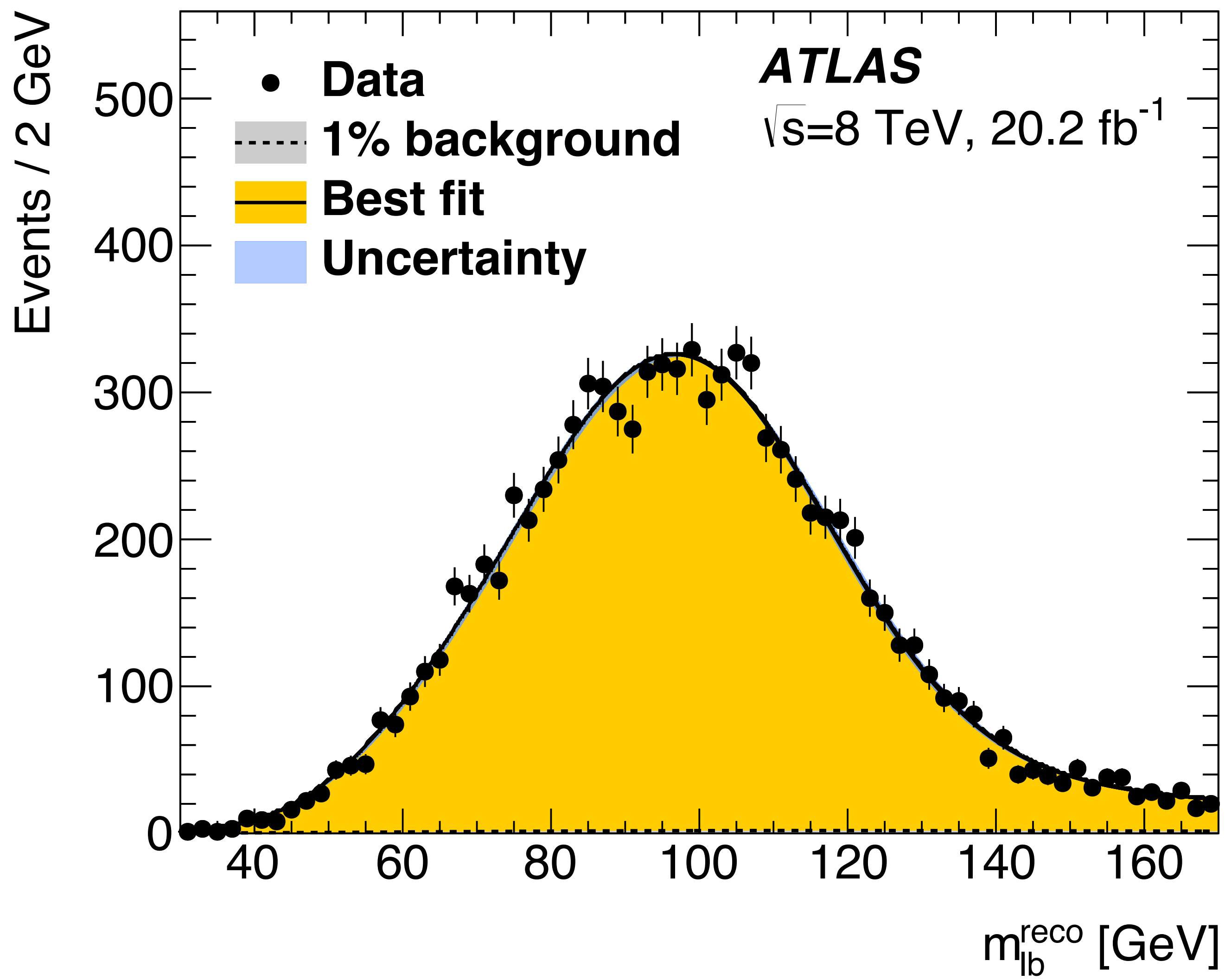
$$m_t = 172.84 \pm 0.84 \ \left[0.41\ \text{(stat)}\ \pm 0.74\ \text{(syst)}\right]\ \text{GeV}$$

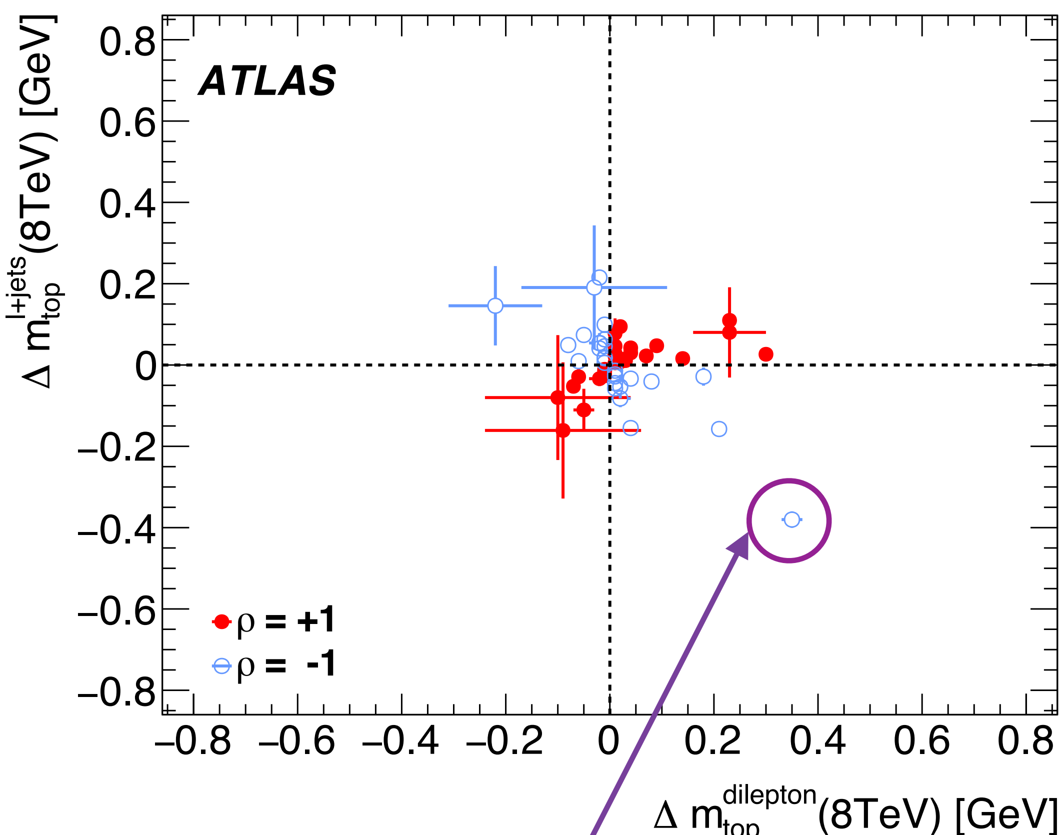
300 MeV b-JES, non-zero stat uncertainty due to selection requirements

ATLAS inputs

- Same 6 measurements as entered the published combination:
  - Lepton+jets, **dilepton**, all-jets at 7 & **8** TeV.

PLB 761 (2016) 350





$$m_t = 172.84 \pm 0.84 \ \left[0.41\ \text{(stat)} \ \pm 0.74\ \text{(syst)}\right] \ \text{GeV}$$

300 MeV b-JES, non-zero stat uncertainty due to selection requirements

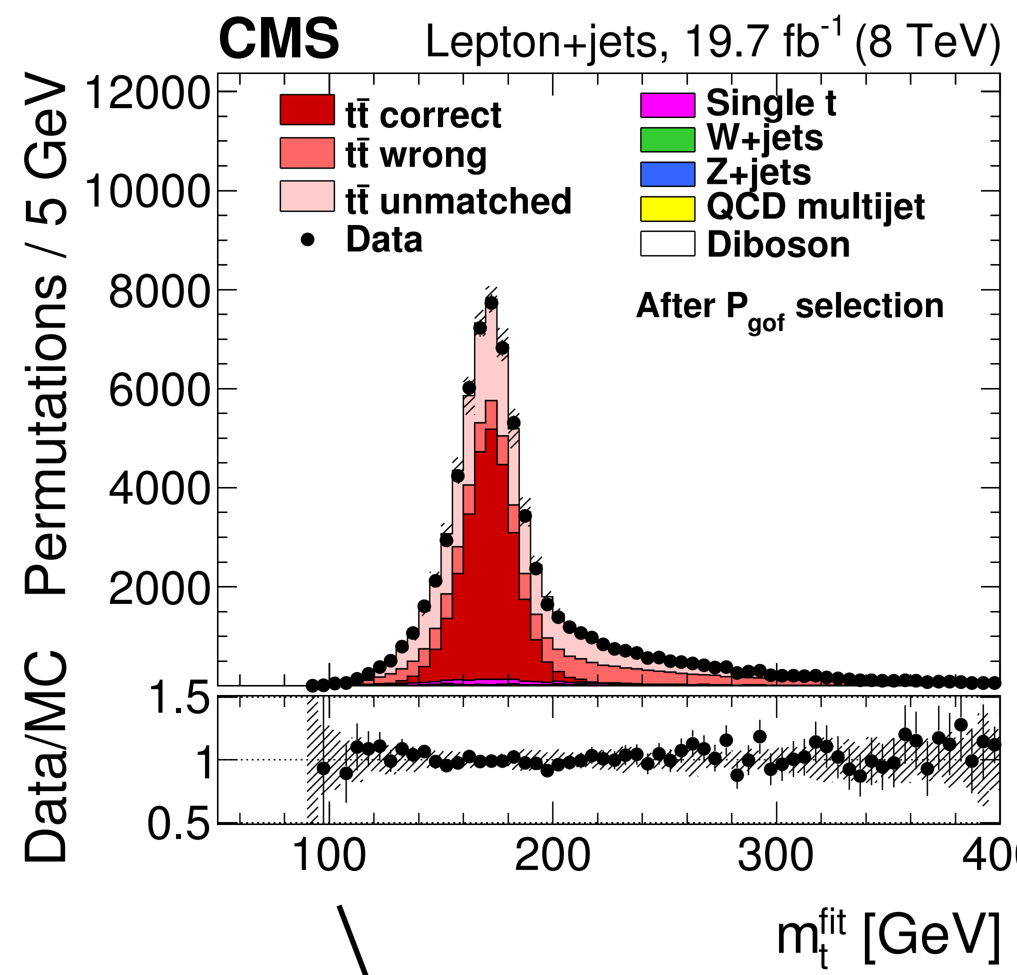
Important -ve correlations to l+jets analysis due to 1D vs 3D fit, e.g. for JES modelling NP1

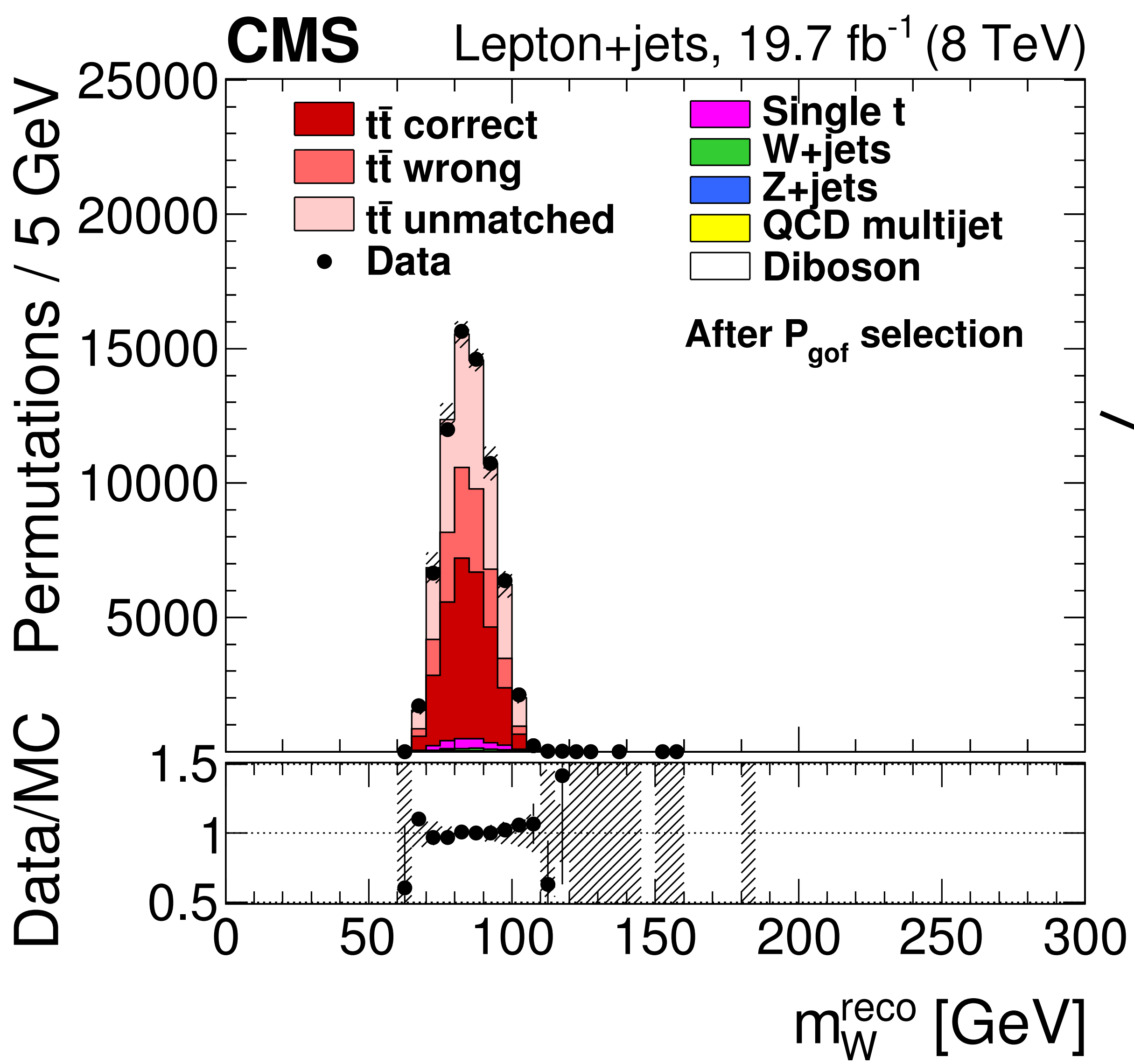
CMS Inputs

- Nine CMS measurements, six in same channels as ATLAS (lepton+jets, dilepton, all-jets at 7 & 8 TeV), plus: PRD 93 (2016) 072004 PRD 96 (2017) 032002
  - 8 TeV single-top measurement (1.2 GeV) EPJC 77 (2017) 354.
  - 8 TeV measurement using m(secondary vertex + lepton) (1.5 GeV) PRD 93 (2016) 092006.
  - 8 TeV muon + J/psi from m(3mu) mass (3.1 GeV) JHEP 12 (2016) 123.
- Relevant changes compared to last CMS combination:
  - No longer take max(stat on syst, syst) for stat limited systematics (as ATLAS) - small improvement in precision of each measurement.
  - Where possible the signs of systematic impacts were included.

CMS Inputs

- Most precise input is 8 TeV l+jets:





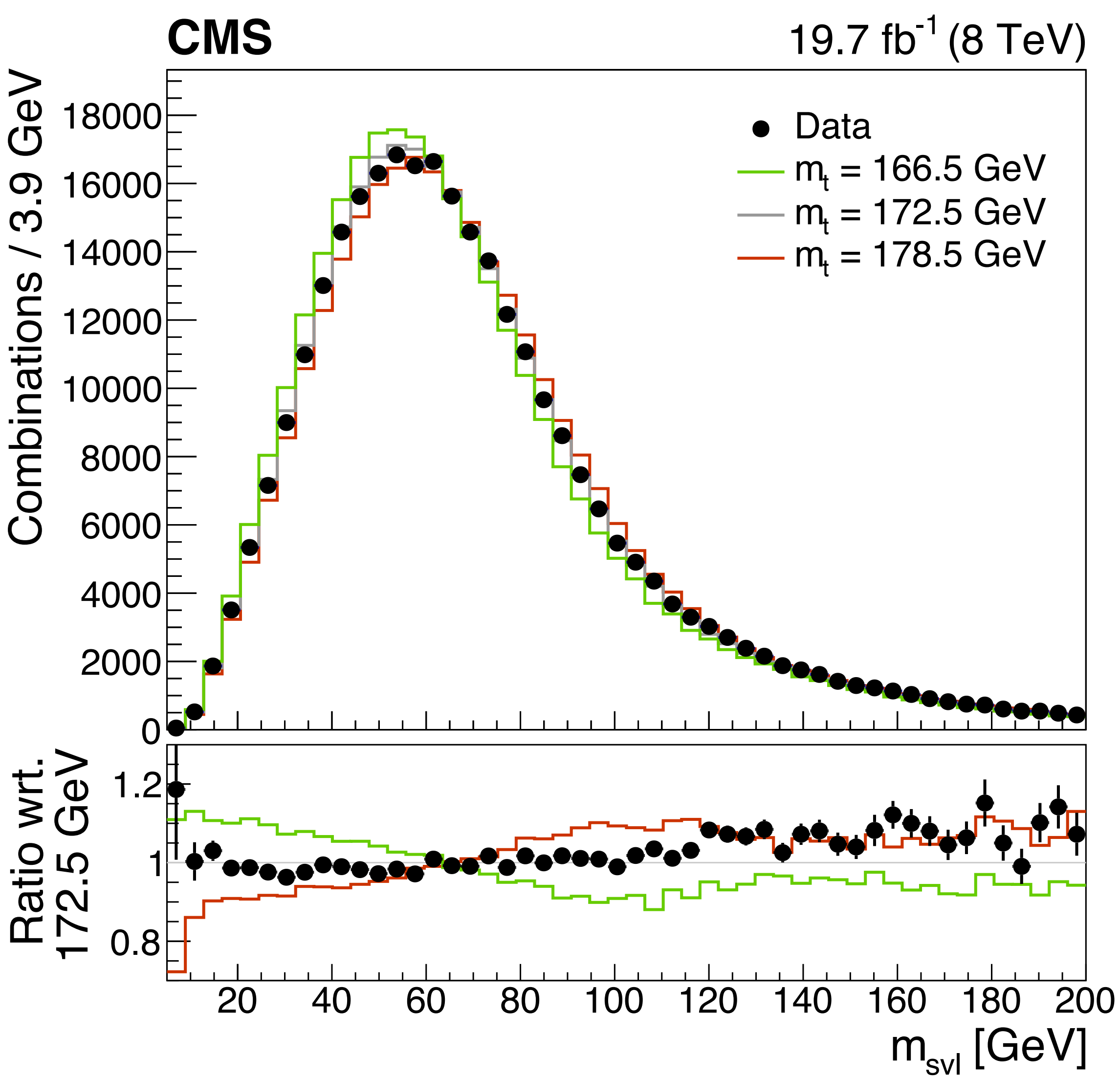
$$m_t = 172.35 \pm 0.48 \ \left[0.16\ (\text{stat}) \ \pm 0.45\ (\text{syst})\right] \ \text{GeV}$$

320 MeV b-JES, small statistical uncertainty

PRD 93 (2016) 072004

CMS Inputs

- Top mass measured from invariant mass of secondary vertex and lepton:



$$m_t = 173.68 \pm 1.12 \ \left[0.20\ (\text{stat}) \ \pm 1.11\ (\text{syst})\right] \ \text{GeV}$$

Sensitive to b-fragmentation,
but JES < 200 MeV

PRD 93 (2016) 092006

University of Glasgow

LHC combination

LHC combination

- $m_t = 172.52 \pm 0.33 \left[0.14 \text{ (stat)} \pm 0.30 \text{ (syst)}\right]$ GeV
  - Uncertainty of 0.33 GeV is 31% improvement on most precise input.
  - Excellent compatibility, $\chi^2 = 7.5$; $p(\chi^2) = 0.91$.
  - Most precise m_t result to date.

LHC combination

- $m_t = 172.52 \pm 0.33 \left[0.14 \text{ (stat)} \pm 0.30 \text{ (syst)} \right]$ GeV
  - Uncertainty of 0.33 GeV is 31% improvement on most precise input.
  - Excellent compatibility, $\chi^2 = 7.5$; $p(\chi^2) = 0.91$.
  - Most precise m_t result to date.
- Single most important input is CMS 8 TeV l+jets, followed by ATLAS 8 TeV l+jets / dilepton:

	ATLAS						CMS								
	2011 (7 TeV)			2012 (8 TeV)			2011 (7 TeV)			2012 (8 TeV)					
	dil	lj	aj	dil	lj	aj	dil	lj	aj	dil	lj	aj	t	J/ψ	vtx
Pull	+0.93	−0.15	+1.43	+0.61	−0.51	+1.09	−0.01	+0.96	+0.71	−0.33	−0.47	−0.37	+0.38	+0.31	+1.08
Weight	−0.02	+0.07	+0.00	+0.16	+0.17	+0.03	−0.08	−0.01	+0.03	+0.12	+0.34	+0.12	−0.03	+0.01	+0.08

- The CMS secondary vertex analysis has weight as high as any 7 TeV measurement -> value of “alternate” measurements which are sensitive to different systematics.

LHC combination

Uncertainty category	Uncertainty impact [GeV] LHC	ATLAS	CMS
b-JES	0.18	0.17	0.25
b tagging	0.09	0.16	0.03
ME generator	0.08	0.13	0.14
JES 1	0.08	0.18	0.06
JES 2	0.08	0.11	0.10
Method	0.07	0.06	0.09
CMS b hadron $\mathcal{B}$	0.07	—	0.12
QCD radiation	0.06	0.07	0.10
Leptons	0.05	0.08	0.07
JER	0.05	0.09	0.02
CMS top quark p_T	0.05	—	0.07
Background (data)	0.05	0.04	0.06
Color reconnection	0.04	0.08	0.03
Underlying event	0.04	0.03	0.05
g-JES	0.03	0.02	0.04
Background (MC)	0.03	0.07	0.01
Other	0.03	0.06	0.01
l-JES	0.03	0.01	0.05
CMS JES 1	0.03	—	0.04
Pileup	0.03	0.07	0.03
JES 3	0.02	0.07	0.01
Hadronization	0.02	0.01	0.01
p_T^{miss}	0.02	0.04	0.01
PDF	0.02	0.06	$<$0.01
Trigger	0.01	0.01	0.01
Total systematic	0.30	0.41	0.39
Statistical	0.14	0.25	0.14
Total	0.33	0.48	0.42

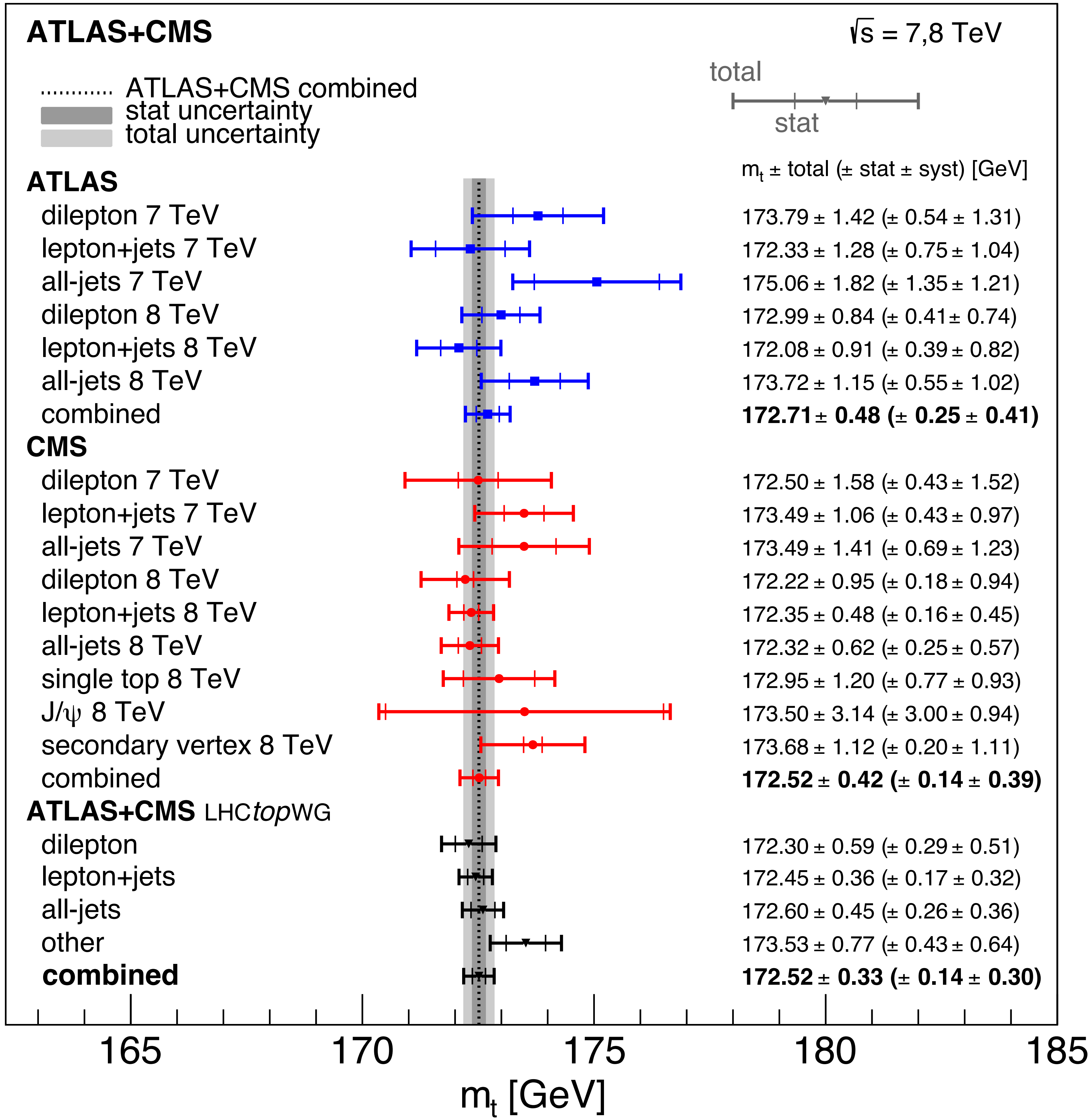
- b-JES is single most important uncertainty.
- JES and b-tagging also relevant.
- Modelling of $t\bar{t}$ events is also crucial.

Correlation scans

Uncertainty category	ρ	Scan range	$\Delta m_t/2$ [MeV]	$\Delta\sigma_{m_t}/2$ [MeV]
JES 1	0	—	—	—
JES 2	0	[−0.25, +0.25]	8	7
JES 3	0.5	[+0.25, +0.75]	1	<1
b-JES	0.85	[+0.5, +1]	26	5
g-JES	0.85	[+0.5, +1]	2	<1
l-JES	0	[−0.25, +0.25]	1	<1
CMS JES 1	—	—	—	—
JER	0	[−0.25, +0.25]	5	1
Leptons	0	[−0.25, +0.25]	2	2
b tagging	0.5	[+0.25, +0.75]	1	1
p_T^{miss}	0	[−0.25, +0.25]	<1	<1
Pileup	0.85	[+0.5, +1]	2	<1
Trigger	0	[−0.25, +0.25]	<1	<1
ME generator	0.5	[+0.25, +0.75]	<1	4
QCD radiation	0.5	[+0.25, +0.75]	7	1
Hadronization	0.5	[+0.25, +0.75]	1	<1
CMS b hadron $\mathcal{B}$	—	—	—	—
Color reconnection	0.5	[+0.25, +0.75]	3	1
Underlying event	0.5	[+0.25, +0.75]	1	<1
PDF	0.85	[+0.5, +1]	1	<1
CMS top quark p_T	—	—	—	—
Background (data)	0	[−0.25, +0.25]	8	2
Background (MC)	0.85	[+0.5, +1]	2	<1
Method	0	—	—	—
Other	0	—	—	—

- Combination very stable.
- Correlations relevant for uncertainty: JES, bJES, MC modelling.
- Central value has mild dependence on bJES correlation.

ATLAS-CMS compatibility

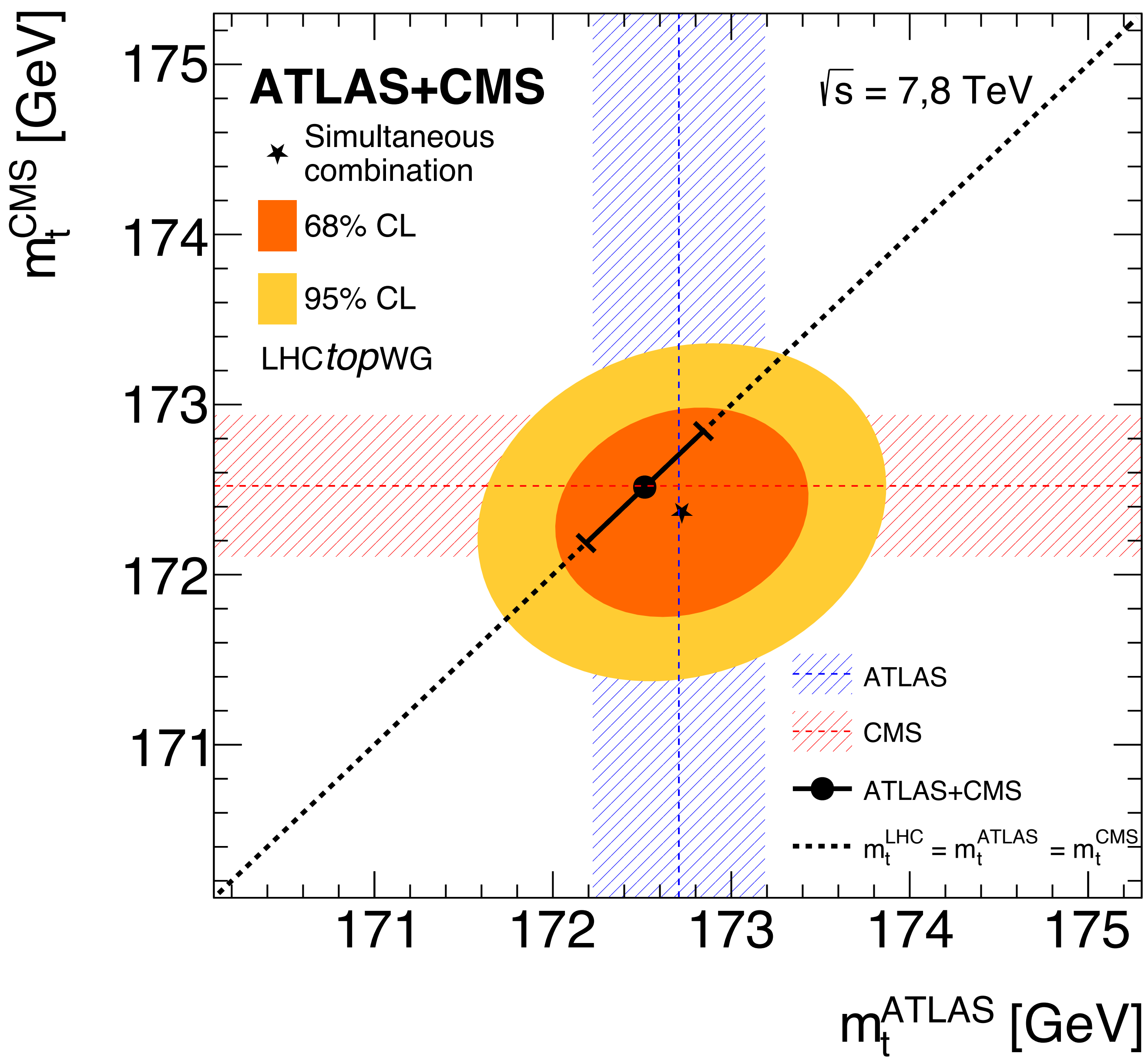


ATLAS-CMS compatibility

- To see ATLAS-CMS compatibility, run "simultaneous" BLUE combination with two m_t parameters, m_t^{ATLAS}, m_t^{CMS} :
  - Distinct from individual combinations: $m_t^{\text{ATLAS}} = \sum_i^{\text{ATLAS}} \lambda_i m_i + \sum_j^{\text{CMS}} \kappa_j m_j; \sum_i \lambda_i = 1; \sum_j \kappa_j = 0$

ATLAS-CMS compatibility

- To see ATLAS-CMS compatibility, run "simultaneous" BLUE combination with two m_t parameters, m_t^{ATLAS}, m_t^{CMS} :
  - Distinct from individual combinations: $m_t^{ATLAS} = \sum_i^{ATLAS} \lambda_i m_i + \sum_j^{CMS} \kappa_j m_j; \sum_i \lambda_i = 1; \sum_j \kappa_j = 0$



- Excellent agreement between the two experiments.

Lessons learnt

- Documentation is key & more is better.
  - Easiest job we had was when full breakdown of systematics is already public in paper / on webpage.
    - The level of info for a combination is typically more than a reader wants (e.g. I want every JES component, while a reader probably cares about the overall impact of JES). We can (& should?) digitise this information e.g. into HepData.
    - ATLAS results generally had finer grained information available than CMS result.
  - Digging through internal notes also worked, but more difficult.

Lessons learnt

- Documentation is key & more is better.
  - Easiest job we had was when full breakdown of systematics is already public in paper / on webpage.
    - The level of info for a combination is typically more than a reader wants (e.g. I want every JES component, while a reader probably cares about the overall impact of JES). We can (& should?) digitise this information e.g. into HepData.
    - ATLAS results generally had finer grained information available than CMS result.
  - Digging through internal notes also worked, but more difficult.
- We also need to know if “up” is the same, e.g. we spotted fairly late on that “up” for b-JES meant higher response in one experiment and lower response in the other.

Lessons learnt

- Documentation is key & more is better.
  - Easiest job we had was when full breakdown of systematics is already public in paper / on webpage.
    - The level of info for a combination is typically more than a reader wants (e.g. I want every JES component, while a reader probably cares about the overall impact of JES). We can (& should?) digitise this information e.g. into HepData.
    - ATLAS results generally had finer grained information available than CMS result.
  - Digging through internal notes also worked, but more difficult.
- We also need to know if "up" is the same, e.g. we spotted fairly late on that "up" for b-JES meant higher response in one experiment and lower response in the other.
- Modelling uncertainties are challenging.
  - More granularity is probably better, e.g. for the ATLAS Pythia vs Herwig comparison we can only correlate that with one of the CMS equivalent uncertainties.
  - Harmonisation would help with correlation assignments (in many places we took 0.5).
  - Personal comment: Harmonisation is good, but it is also risky to be in a place where both experiments have identical MC setups -> potentially lose some robustness.

Lessons learnt

- Documentation is key & more is better.
  - Easiest job we had was when full breakdown of systematics is already public in paper / on webpage.
    - The level of info for a combination is typically more than a reader wants (e.g. I want every JES component, while a reader probably cares about the overall impact of JES). We can (& should?) digitise this information e.g. into HepData.
    - ATLAS results generally had finer grained information available than CMS result.
  - Digging through internal notes also worked, but more difficult.
- We also need to know if “up” is the same, e.g. we spotted fairly late on that “up” for b-JES meant higher response in one experiment and lower response in the other.
- Modelling uncertainties are challenging.
  - More granularity is probably better, e.g. for the ATLAS Pythia vs Herwig comparison we can only correlate that with one of the CMS equivalent uncertainties.
  - Harmonisation would help with correlation assignments (in many places we took 0.5).
  - Personal comment: Harmonisation is good, but it is also risky to be in a place where both experiments have identical MC setups -> potentially lose some robustness.
- Having different analyses with different sensitivity to the systematics matters.

Summary

- The run-1 combination for the top-quark mass yields:

$$m_t = 172.52 \pm 0.33 \left[0.14 \text{ (stat)} \pm 0.30 \text{ (syst)}\right] \text{ GeV}$$

- This is the most precise result to data.
- The ATLAS and CMS run-1 measurements are highly consistent and the result is stable against variations of the correlations.
- We learnt a lot which can hopefully aid future combinations.