IL NUOVO CIMENTO Vol. 123 B, N. 8-9 Agosto-Settembre 2008
DOI 10.1393/ncb/i2008-10665-8

World average top-quark mass

D. Glenzinski
Fermilab - Batavia, IL, USA

(ricevuto il 22 Ottobre 2008; pubblicato online il 12 Dicembre 2008)

Summary. — This paper summarizes a talk given at the Top2008 Workshop at La Biodola, Isola d'Elba, Italy. The status of the world average top-quark mass is discussed. Some comments about the challanges facing the experiments in order to further improve the precision are offered.

PACS 12.15.Ff – Quark and lepton masses and mixing.
PACS 14.65.Ha – Top quarks.

1. – Introduction

The world average top-quark mass includes measurements from the CDF and DØ experiments using data taken during Run-I and Run-II of the Fermilab Tevatron. Run-I ended in 1995 with each experiment collecting about $110\,\mathrm{pb}^{-1}$. Run-II began in 2001 and is schedule to end in 2009-2010 with each experiment collecting about $7000\,\mathrm{pb}^{-1}$ of data. At the time of this conference each experiment had collected close to $3000\,\mathrm{pb}^{-1}$. The results shown are based on a combination of the most up-to-date results available at the time of the 2008 winter conferences and includes Run-II results using a little over $2\,\mathrm{fb}^{-1}$. There are measurements from each of the $t\bar{t}$ final states, the dilepton $t\bar{t} \to WbW\bar{b} \to \ell\nu\ell\nu b\bar{b}$ (dil), the lepton-plus-jets $\ell\nu q\bar{q}' b\bar{b}$ (ljt), and the all hadronic $q\bar{q}' bq\bar{q}'\bar{b}$ (had).

The Tevatron experiments have been combining their measurements for a decade. The first world average is documented in ref. [1]. The most recent is documented in ref. [2]. The methodology is unchanged during that time. We employ the BLUE method, which is described in detail in ref. [3]. The combinations are performed by the Tevatron Electroweak Working Group [4], which is composed of experts from the CDF and DØ experiments. In this paper I will give some details about the present combination and briefly discuss some of the challenges which lay ahead.

2. – Combination methodology

The acronym BLUE stands for "Best Linear Unbiased Estimator" and is widely used throughout HEP for combining measurements. The methodology is straightforward and allows the combination of correlated measurements of one or more parameters. Assuming the inputs are unbiased, the combination itself also yields an unbiased estimate while minimizing the variance (that is the "Best" part). It has the nice feature that it produces a fit χ^2 which quantifies the consistency of the inputs and also provides a clean way to breakdown the total uncertainty into its constituent contributions. As with all combination procedures the correlations among the inputs are required as input. The BLUE methodology assumes all uncertainties are Gaussian distributed.

As input the method requires a list of measurements and their correlation matrix. The uncertainties on the measurements are broken down into several categories and the correlations among the input measurements are estimated for each category separately. The total correlation matrix is then obtained by summing over the correlation matrices determined separately for each of these error categories. For the statistical uncertainties, the correlations are rigorously determined—using MC pseudo-experiments if necessary. For the systematic uncertainties the correlations are taken to be either 0 or 1—variations are considered as a cross-check.

Once the total correlation matrix has been specified, the BLUE method is straightforward. After inverting the matrix, the rest of the combination is analytic and the resulting top mass is a weighted average of the input measurements. For the top-quark mass averages we perform the combination using an analytic code as well as performing a MINUIT minimization. The two results always have agreed to better than $0.001\,\mathrm{GeV}/c^2$.

3. – Combination inputs

The latest combination uses twelve inputs: five published Run-1 results (C1-dil, C1-ljt, C1-had, D1-dil, D1-ljt) and seven Run-II results most of which are the latest preliminary results available (C2-dil, C2-ljt, C2-had, C2-lxy, D2-dil, D2-ljt-a, D2-ljt-b). Most the labels should be obvious in what follows but I will point out a few exceptions. The "C2-lxy" input is a CDF measurement in the ljt channel which uses the average observed flight distance of vertex tagged b-hadrons to determine the top-quark mass. The "D2-ljt-a" and "D2-ljt-b" inputs are DØ measurements in the ljt channel using the Run-IIa and Run-IIb datasets, respectively. At the time of this conference, these were about $1\,\mathrm{fb}^{-1}$ each. The uncertainties are separated into twelve categories:

Statistical: uncertainty on M_t arising from the limited data statistics.

Signal: uncertainties arising from variations in modeling the $p\overline{p} \to t\bar{t}$ process (*e.g.*, ISR, FSR, PDF).

Background: uncertainties associated with the normalization and shape variations in the background processes (*e.g.*, arising from Q^2 variations).

Fit: uncertainties arising from the variations in the fit methodology including those which arise from the limited statistics of the Monte Carlo samples used to calibrate the fits.

Monte Carlo: uncertainties arising from variations in the choice of Monte Carlo generator by comparing Pythia, Herwig, Alpgen, and MC@NLO.

UN/MI: uncertainties unique to DØ arising from noise contributions in the Uranium calorimeter and in the handling of multiple interaction events.

Jet energy scale: uncertainties arising from variations of the jet energy scale (JES). The determinatoin of the JES is quite involved and its uncertainty includes contributions from detector effects as well as MC modeling assumptions. In order

TABLE I. – *Summary of the input measurements used to determine the world average top-quark mass. All numbers are in GeV/c^2. The error categories and their correlations are described in the text. The total systematic uncertainty and the total uncertainty are obtained by adding the relevant contributions in quadrature.*

	Run-I published					Run-II preliminary						
	CDF			DØ		CDF				DØ		
	had	ljt	dil	ljt	dil	ljt	dil	had	lxy	ljt-a	ljt-b	dil
Result	186.0	176.1	167.4	180.1	168.4	172.7	171.2	177.0	180.7	170.5	173.0	173.7
iJES	0.0	0.0	0.0	0.0	0.0	1.3	0.0	1.8	0.0	0.0	0.0	0.0
aJES	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.8	1.9
bJES	0.6	0.6	0.8	0.7	0.7	0.4	0.1	0.1	0.0	0.2	0.1	0.9
cJES	3.0	2.7	2.6	2.0	2.0	0.5	1.7	0.6	0.0	0.0	0.0	2.1
dJES	0.3	0.7	0.6	0.0	0.0	0.1	0.1	0.1	0.0	1.7	1.4	0.9
rJES	4.0	3.4	2.7	2.5	1.1	0.2	1.8	0.5	0.3	0.0	0.0	0.0
Signal	1.8	2.6	2.8	1.1	1.8	0.6	0.7	0.6	1.4	1.0	0.5	0.8
BG	1.7	1.3	0.3	1.0	1.1	0.6	0.4	1.0	7.2	0.5	0.4	0.6
Fit	0.6	0.0	0.7	0.6	1.1	0.2	0.6	0.6	4.2	0.1	0.2	0.9
MC	0.8	0.1	0.6	0.0	0.0	0.4	0.7	0.3	0.7	0.0	0.0	0.2
UN/MI	0.0	0.0	0.0	1.3	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Syst.	5.7	5.3	4.9	3.9	3.6	1.7	2.8	2.4	8.5	2.2	1.7	3.4
Stat.	10.0	5.1	10.3	3.6	12.3	1.2	2.7	3.3	14.5	1.9	1.3	5.4
Total	11.5	7.3	11.4	5.3	12.8	2.1	3.9	4.1	16.8	2.9	2.2	6.4

to more accurately account for the correlations between the two experiments the JES contribution to the top mass uncertainty is actually subdivided into six sub-categories labeled aJES, bJES, cJES, dJES, iJES, and rJES. These are discussed in more detail in ref. [2].

The twelve input measurements and their uncertainties are given in table I. The correlations used in the combination are detailed in ref. [2]. There are categories which are uncorrelated across all inputs (*e.g.*, Statistical), correlated across inputs from the same run and same experiment (*e.g.*, some of the JES sub-categories), correlated across all inputs from the same experiment but not across CDF and D0 (*e.g.*, UM/MI), correlated across all inputs in the same channel (*e.g.*, Background), and correlated everywhere across all inputs (*e.g.*, Signal modeling and Monte Carlo comparisons). Using the inputs in table I and their correlations yields the total correlation matrix in table II.

4. – Combination results

The combination yields a world average top-quark mass of

$$M_t = 172.6 \pm 1.4\,\mathrm{GeV}/c^2 \qquad (1)$$

TABLE II. – *The resulting matrix of total correlation coefficients used to determined the world average top-quark mass.*

		Run-I published					Run-II preliminary						
		CDF			DØ		CDF				DØ		
		ljt	dil	had	ljt	dil	ljt	dil	had	lxy	ljt-a	ljt-b	dil
CDF-I	ljt	1.00											
CDF-I	dil	0.29	1.00										
CDF-I	had	0.32	0.19	1.00									
DØ-I	ljt	0.26	0.15	0.14	1.00								
DØ-I	dil	0.11	0.08	0.07	0.16	1.00							
CDF-II	ljt	0.30	0.17	0.16	0.22	0.09	1.00						
CDF-II	dil	0.45	0.27	0.33	0.21	0.11	0.24	1.00					
CDF-II	had	0.17	0.11	0.15	0.09	0.05	0.11	0.17	1.00				
CDF-II	lxy	0.11	0.03	0.02	0.10	0.01	0.16	0.03	0.02	1.00			
DØ-II	ljt-a	0.16	0.09	0.06	0.11	0.05	0.16	0.07	0.06	0.10	1.00		
DØ-II	ljt-b	0.11	0.06	0.04	0.09	0.03	0.12	0.04	0.04	0.09	0.20	1.00	
DØ-II	dil	0.18	0.12	0.11	0.17	0.08	0.14	0.19	0.07	0.01	0.21	0.14	1.00

with a χ^2/d.o.f. = 6.9/11 which corresponds to a χ^2-probability of 81%. The relative uncertainty is now below the one percent level at 0.8% and far surpasses the Run-II goal. This is largely due to improvements in the analysis techniques—in particular the inclusion of an *in situ* JES calibration by using information from the $W \to q\bar{q}'$ decays, which significantly reduces the JES related uncertainties.

The total uncertainty can be broken down into three main sub-components,

$$\Delta M_t = \pm 0.8\,(\mathrm{sta}) \,\pm 0.7\,(W \to q\bar{q}) \,\pm 0.8\,(\mathrm{sys})\ \mathrm{GeV}/c^2, \quad (2)$$

$$= \pm 1.1\,(\mathrm{exp}) \,\pm 0.8\,(\mathrm{sys})\ \mathrm{GeV}/c^2. \quad (3)$$

The statistical uncertainties together with the component of the JES uncertainty arising from the *in situ* $W \to q\bar{q}$ calibration dominate the uncertainty, but both will continue to scale with the dataset. The largest contributions to the remaining systematic uncertainties arise from the Signal modeling ($\pm 0.5\,\mathrm{GeV}/c^2$) and the remaining JES uncertainties ($\pm 0.5\,\mathrm{GeV}/c^2$). It should be noted that uncertainties arising from Color Reconnection effects have not yet been quantified. Monte Carlo generators have only recently become available which include these effects for $p\bar{p}$ (and pp interactions). Both experiments are in the process of tuning these models to their data before utilizing them to quantify an uncertainty on the top-quark mass measurements.

A variety of cross-checks are performed. The version of BLUE we use accepts only symmetric uncertainties. For some of the input measurements asymmetric uncertainties are quoted. The combination is repeated using each extreme of the asymmetric uncertainties for these inputs. Some of the input measurements are potentially statistically correlated. While Monte Carlo pseudo-experiments determined that the real correlation is consistent with zero, we repeat the combination using the statistical uncertainty on the estimate of this correlation instead. For all the systematic related error categories for which off-diagonal correlations are used, we simultaneously vary the correlations by ±20% and repeat the combination. We also repeat the combination treating the Signal uncertainties as uncorrelated across the measurements since their determination is usually limited by Monte Carlo statistics. Lastly, the set of error categories presently used was developed after the Run-I results had been published. Consequently the Run-I uncertainties had to be reclassified and some ambiguities arise in so doing. Variations of the Run-I uncertainties are considered to address these ambiguities and the combination repeated. In all these cases both the central value and the total uncertainty vary by less that $0.15\,\mathrm{GeV}/c^2$. At present, we are largely insensitive to the details of these systematic uncertainties and their correlations. The experiments are continuing to work together to improve their understanding of some of these correlations since we anticipate they could become more important as the experimental uncertainties continue to shrink.

For this conference I performed a couple of variations of the fit. It should be noted that these are in no way official—I did them for fun.

First, I repeated the combination but fit for two observables, the CDF top-quark mass and the DØ top-quark mass. I found $M_t^{\mathrm{CDF}} = 172.5 \pm 1.8\,\mathrm{GeV}/c^2$ and $M_t^{\mathrm{D}\emptyset} = 172.7 \pm 1.8\,\mathrm{GeV}/c^2$ with a total correlation coefficient of $\rho_{\mathrm{C\text{-}D}} = 0.22$. These are clearly consistent and surprisingly only moderately correlated—mostly owing to the fact that the experimental uncertainties are still dominant. The consistency is not trivial since there are several important differences between the two experiments. For example, they calibrate their methods to different Monte Carlo generators and their calorimeters are very different and thus they each determine their JES in very different ways. Note that the combinations given above will not necessarily reproduce the combinations reported by each experiment separately. The reason is that the above fits include the effects of systematic correlations across the two experiments. Note that in order to properly compare the consistency between the two experiments, it is necessary to account for these cross-experiment correlations as done here.

I also repeated the combination for a Run-I and a Run-II top-quark mass. I found $M_t^{\mathrm{R1}} = 178.2 \pm 4.3\,\mathrm{GeV}/c^2$ and $M_t^{\mathrm{R2}} = 172.8 \pm 1.4\,\mathrm{GeV}/c^2$ with a total correlation coefficient of $\rho_{\mathrm{R1\text{-}R2}} = 0.43$. These are consistent at the 17% level. Again, the Run-I number reported here will not exactly reproduce the number reported in ref. [1] since it includes additional information through the Run-I / Run-II correlations([1]).

5. – Remarks concerning the combination

A few remarks are in order. CDF and DØ do not quantify their systematic uncertainties using exactly the same prescriptions. This has two consequences. First, the assumption that the systematic correlations are $\rho = 1$ is suspect. Second, a smaller uncertainty does not necessarily mean a particular input is less sensitive to that systematic variation—it could instead just reflect the different methodologies being used to quantify the systematic. This could bias the weights assigned. From the cross-checks described in sect. **4** we see that these caveats do not significantly affect the result, at least for the time being. Moving forward these things might become more important as those uncertainties

---

([1]) The final combined Run-I top-quark was $178.0 \pm 4.3\,\mathrm{GeV}/c^2$.

come to dominate the total uncertainty. The two experiments are working together to define common procedures, particularly for the uncertainties related to Signal modeling.

It is also worth noting that some components of the systematic uncertainties are presently being overestimated. For example, there is some double counting of ISR/FSR uncertainties. These are difficult to untangle experimentally. Moreover contributions from ISR/FSR variations are included in the Signal category as well as one of the JES sub-categories. Of course the comparison across different generators also implicitly includes ISR/FSR modeling variations as well. Another example is that many of the Signal uncertainties are actually limited by the Monte Carlo statistics available. The present samples are equivalent to over 250 times the available luminosity. This still limits the statistical precision with which the systematic comparisons can be made to the $\approx 0.2\,\mathrm{GeV}/c^2$ level per effect. While reweighting techniques are employed to quantify the PDF uncertainties, these are difficult to implement for many of the more important modeling uncertainties (*e.g.*, ISR/FSR) since these arise from non-perturbative QCD effects.

Finally it is worth reiterating that Monte Carlo generators including Color Reconnection effects for hadron collider processes have only recently become available. The systematic uncertainty arising from the introduction of these effects is in the process of being evaluated.

The main point is that we expect to make additional progress on understanding and quantifying the systematic uncertainties.

Looking ahead, the final Run-II dataset will be another factor of 3-4 times larger than the one included in the present combination. The total experimental uncertainties could reach the $\pm 0.6\,\mathrm{GeV}/c^2$ level. The current systematic uncertainties are at the $\pm 0.8\,\mathrm{GeV}/c^2$ level, but as mentioned above that will continue to evolve (with contributions in both directions). Still, it is possible for the Tevatron combined top-quark mass to reach a precision of about $1\,\mathrm{GeV}/c^2$ total uncertainty with the full Run-II dataset. To get there the two experiments need to continue working on the understanding of the modeling related systematic uncertainties.

It is worth noting that the modeling uncertainties will be exactly the same for the LHC experiments. Eventually we will want to add those top-quark mass measurements to the combination. A common methodology for quantifying the modeling related systematics as well as implementing a common set of error categories would greatly simplify this.

6. – An aside

Given the precision we are reaching, it is worth asking "What are we measuring?". This is not a simple question to answer. But here are some things to consider.

All the measurements reported here (and in the future) will calibrate themselves to Monte Carlo. This is not unique to the top-quark mass. It is standard experimental practice to demonstrate for any given measurement that the methodology employed yields an unbiased estimate of the physical observable of interest. In many cases a clean control sample is not available and thus Monte Carlo samples are used. The top-quark mass measurements to date are calibrated by varying a parameter in the Monte Carlo generator which roughly corresponds to the pole mass. This statement is based on numerous conversations with numerous authors of the generators in use. The approximation is good to of order $\Lambda_{\mathrm{QCD}} \sim 200\,\mathrm{MeV}/c^2$. It may turn out that the parameter should be more accurately described as a "jet mass". The implications for the Electroweak fits are unclear to me (and I think are unclear in general at this point).

Work continues with the theorists making the calculations and the Monte Carlo authors to further resolve this issue.

I would like to point out that this is a *theoretical* issue. That is, the theorists and the Monte Carlo authors need to resolve this. It seemed to me a lot of progress along those lines was made during this week. But I think the experiments are doing their job here. To measure the top-quark mass we need an experimental observable that is a) a color singlet, b) sensitive to M_t, c) is well defined at a hadron collider, and d) can be modeled theoratically in a well-defined manner. The first three of these are satisfied by construction. The last is challenging because the experimental observables available are affected by numerous non-perturbative (QCD) effects. This makes the theoretical interpretation more difficult since mapping the perturbative calculations to the observables requires non-perturbative models. But it should be made clear that the effect of varying these non-perturbative models is *exactly* what we aim to quantify in our systematic uncertainties. How far off might we be from the pole-mass? At present we estimate about $500\,\mathrm{MeV}/c^2$. To the extent the Monte Carlo generators properly include the most important corrections, incorporate reasonable non-perturbative models, and allow the experiments to adequately vary these choices, the systematic studies performed represent a reasonable estimate of these uncertainties.

7. – Conclusion

The Tevatron experiments CDF and D∅ have recently updated their Run-II measurements to include data sets as large as $2\,\mathrm{fb}^{-1}$. Combining these preliminary results with published Run-I results yields a world average top-quark mass of $M_t = 172.6 \pm 1.4\,\mathrm{GeV}/c^2$, which corresponds to a relative uncertainty of 0.8%. The uncertainty is dominated by experimental contributions ($\pm 1.1\,\mathrm{GeV}/c^2$) which are expected to scale with the data statistics available. The remaining systematic contributions ($\pm 0.8\,\mathrm{GeV}/c^2$) are the focus of on going work. It is conceivable that with the full Run-II dataset the Tevatron combined top-quark mass could reach a precision of about $1\,\mathrm{GeV}/c^2$. In parallel with this experimental effort a theoretical effort to more precisely define the parameter used in the Monte Carlo generators is necessary. At present I think it is fair to say that it is understood that this parameter corresponds to the top-quark pole mass within a few times $200\,\mathrm{MeV}/c^2$. It is also fair to note that there is some disagreement about this and that discussions are continuing.

* * *

The author would like to thank the organizers for a great workshop in which discussion flourished.

REFERENCES

[1] Demortier L. *et al.*, Fermilab-TM-2084 (2008).
[2] Tevatron Electroweak Working Group, arxiv:0803.1683 [hep-ex] (2008).
[3] Lyons L., Gibaut D. and Clifford P., *Nucl. Instrum. Methods A*, **270** (1988) 110; Velassi A., *Nucl. Instrum. Methods A*, **500** (2003) 391.
[4] For more information see, `http://tevewwg.fnal.gov`.