Determination of the top quark mass from leptonic observables

Stefano Frixione[a] **and Alexander Mitov**[b]

[a]*PH Department, TH Unit, CERN, CH-1211 Geneva 23, Switzerland*
[b]*Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, U.K.*

E-mail: Stefano.Frixione@cern.ch, adm74@cam.ac.uk

ABSTRACT: We present a procedure for the determination of the mass of the top quark at the LHC based on leptonic observables in dilepton $t\bar{t}$ events. Our approach utilises the shapes of kinematic distributions through their few lowest Mellin moments; it is notable for its minimal sensitivity to the modelling of long-distance effects, for not requiring the reconstruction of top quarks, and for having a competitive precision, with theory errors on the extracted top mass of the order of $0.8\,\mathrm{GeV}$. A novel aspect of our work is the study of theoretical biases that might influence in a dramatic way the determination of the top mass, and which are potentially relevant to all template-based methods. We propose a comprehensive strategy that helps minimise the impact of such biases, and leads to a reliable top mass extraction at hadron colliders.

KEYWORDS: QCD Phenomenology, Hadronic Colliders

ARXIV EPRINT: 1407.2763

Contents

1 Introduction

The current world average of the top quark mass [1]

$$m_t = 173.34 \pm 0.76 \, \text{GeV} \quad [\text{World Average}] \tag{1.1}$$

implies that m_t is known with a precision better than 0.5%. Such an accuracy is perfectly adequate for present collider-physics applications [2] including, notably, the global electroweak (EW) fits [3], which are saturated by the uncertainty on the W-boson mass, and not by that on m_t. Still, the accurate determination of the top quark mass at hadron colliders remains a subject of much activity and debate.

Two separate developments have been the main drivers behind the above mentioned activity: the outsize role played by the top quark mass in determining the stability of the electroweak vacuum (both in the Standard Model (SM) [4–6] and beyond [7]), and the recognition that the extraction of m_t at hadron colliders involves significant theoretical challenges, that might conceivably affect its value at the level of $\mathcal{O}(1\,\text{GeV})$ (see ref. [2] for detailed discussion).

The bottom-up extrapolation of EW-scale physics, based on eq. (1.1), implies either that the EW vacuum becomes unstable below the Planck scale, or that the result of eq. (1.1) deviates from the value needed for the stability of the SM EW vacuum up to the Planck scale by about two to four sigma's [6, 8]. If confirmed, such a conclusion might indirectly imply the existence of Beyond the SM (BSM) physics somewhere below the Planck scale. Given the non-observation of BSM signals so far, it would be hard to overstate the importance of this implication. We stress that these facts are mainly driven by the m_t value of eq. (1.1), and this because of the large parametric dependence of the stability condition on the top quark mass.

At this point one might wonder about the need for revisiting the subject of m_t determination, given the quite high precision of the top mass of eq. (1.1). To this end let us remind the reader that there are a number of high-precision measurements that marginally agree with the current world average. Examples are the very recent CMS [9] and D0 [10] measurements:

$$m_t = 172.04 \pm 0.77 \, \text{GeV} \quad [\text{CMS Collaboration}] \,,$$
$$m_t = 174.98 \pm 0.76 \, \text{GeV} \quad [\text{D0 Collaboration}] \,. \tag{1.2}$$

The above measurements have the same uncertainty as the combination in eq. (1.1), but notably different central values.[1] In particular, the CMS measurement [9] is consistent with the SM EW vacuum being stable up to the Planck scale, while the D0 measurement [10] implies a rather unstable SM EW vacuum. Therefore, the spread in the available m_t measurements alone warrants a closer inspection of the determination of the top quark mass. As we shall detail in the following, there are also strong theoretical reasons that motivate further studies of the extraction of this parameter from hadron collider data.

The determination of the top quark mass is as much dependent on theoretical assumptions as it is on measurements. The reason is that the top quark mass is not an observable and thus cannot be measured directly:[2] it is a theoretical concept, and its value is extracted from data in collider events that feature top quarks. Such an extraction depends on the definition of the mass (pole mass, running mass, and so forth), on the observables chosen, and on the various approximations made when computing those observables. Since measurements are insensitive to theory assumptions,[3] any modification in the theoretical modelling will result in a different value of the extracted top mass. If everything is consistent, i.e. if the estimated uncertainty is a realistic representation of the true uncertainty, then the differences in the returned values should fall within the corresponding theory errors. In reality, this may not be the case due to the presence of biases, whose very existence might be difficult to establish. With this important subtlety in mind, one of the main aspects of the present work is to devise a structured approach towards the identification of such hidden biases.

---

[1]The measurements in eq. (1.2) agree with the world average of eq. (1.1) at approximately 2σ.

[2]For this reason we do not speak of top mass *measurements* but of top mass *determinations* or *extractions*.

[3]Strictly speaking, this is never the case. For example, corrections for detector effects do depend on theory assumptions. In the first approximation, one can ignore these data-theory correlations.

A significant number of techniques for the determination of the top mass exist or are under study; see ref. [2] for a recent in-depth overview. Such techniques may be organised into two classes, whose definitions cannot be given in a rigorous way, but which are nevertheless based on clearly distinct physical principles. The first class includes all those approaches that use, in some form, the fact that the top is a particle that decays: the knowledge of the decay products (i.e. their identities and kinematic configurations) is then exploited to reconstruct some quantity which is directly related to the top, and thus bears information on its mass. The crucial characteristic is that, by emphasising the role of the decay, one factors out the details of the process in which the primary top(s) is(are) produced, so that the details of the production mechanism become irrelevant. The ideal (i.e. not realistic) procedure which belongs to this class is the one where the top virtuality is reconstructed exactly by measuring the invariant mass of its decay products, thus scanning its lineshape. In the approaches that belong to the second class the role of the top as a mother particle must not matter; the only important thing is that some observable(s) of a top-mediated process depend in a significant way on m_t, so that their measurements can be mathematically inverted (using suitable theoretical predictions) to return the top mass. We stress that the fact that the observables mentioned above are most likely constructed by using the top decay products is not relevant. The only important thing is that they depend on the top quark mass, a feature that might be possessed by other quantities as well (for example, the primary QCD radiation in the production process).

The approaches that belong to the first class are often perceived to be affected by smaller theoretical systematics than those of the second class, because by their very definition one assumes that many sources of uncertainties, such as PDF dependence, absence of higher-order perturbative corrections, and new-physics contributions, will drop out, being mostly associated with the production mechanism. Unfortunately, this is not really the case. Firstly, some of these sources might be relevant to decays as well. Secondly, different kind of uncertainties could become important: a good example is the so-called J/Ψ method [11] which, although experimentally very clean and theoretically well defined, is hampered by its sensitivity to the non-perturbative b-fragmentation. Thirdly, in these approaches one must start by *defining* what one means by "top", which introduces some auxiliary (if only intermediate) concept in the procedure, and renders it difficult to assign a proper theoretical error to it. Note that this necessity goes beyond what one must do in order to reconstruct the top quark experimentally, and is purely theoretical.

The bottom line is that, regardless of which class an m_t-extraction technique belongs to, some amount of theoretical modelling will be involved. In this paper, we follow an approach of the second class; we believe that not having to define the top as a final-state object is a virtue that more than compensates a larger dependence on the production process.

Another important motivation behind the procedure we are proposing is the use of observables that can be both reliably predicted within the SM, and cleanly measured. Thus, we employ kinematic distributions of leptons in dilepton $t\bar{t}$ events; more precisely, we are interested in their shapes. Furthermore, we find that the information on the top mass that such shapes encode can be very effectively provided by the Mellin moments of

the corresponding distributions, and it is such moments that will play a central role in our method. Our goal is the determination of m_t with competitive precision, supplemented by a detailed study of the various sources of theoretical systematics. Apart from not having to rely, directly or indirectly, on the reconstruction of top quarks, our approach has minimal sensitivity to the modelling of both perturbative and non-perturbative QCD effects.[4] We believe that the latter property is one of the chief advantages of the method we are pursuing.

In this paper we shall be working with the top quark pole mass, and shall not consider alternative mass definitions. Our viewpoint is that the intrinsic differences between any two of these definitions (renormalon-related effects are a good example) are largely below the present level of uncertainties, and therefore we do not see them as a reason for concern at present. A fuller discussion can be found in ref. [2].

We shall conclude that, with the procedure we employ, the extraction of the top pole mass can be achieved with a theoretical error of about $0.8\,\mathrm{GeV}$, and possibly smaller. While a significant number of $t\bar{t}$ dilepton events have been recorded during Run I of the LHC, no measurements are published of the Mellin moments that would allow us to apply our procedure to real data. We thus hope that this paper will encourage the LHC experimental collaborations to measure directly such moments, so that the present analysis could be repeated, and its results compared with those of eqs. (1.1) and (1.2). Furthermore, we are hopeful that the reliability and small theoretical systematics of the method proposed in this work will help shed light on the issue of the EW vacuum stability.

This paper is organised as follows: in section 2 we introduce our method in detail and define, in particular, its associated theoretical errors (section 2.2) and biases that may affect it (section 2.3). Our results are presented in section 3: those with the highest theoretical accuracy in section 3.2.3, with discussions on the effects due to parton showers, higher orders, and spin correlations in section 3.2.2, and explicit examples of theoretical biases in section 3.2.4. We give our conclusions in section 4. Some technical material is reported in the appendices.

2 The method

Our goal is to study the determination of the top quark pole mass m_t from several differential distributions of leptons in dilepton $t\bar{t}$ events:

$$pp \to t\bar{t} + X\,, \qquad t \to \ell^+ \nu_\ell b\,, \quad \bar{t} \to \ell^- \bar{\nu}_\ell \bar{b}\,, \qquad \ell = e, \mu\,. \tag{2.1}$$

Each of the observables that we consider features the following important properties:

- It does not require the reconstruction of the t and/or $\bar{t}$ quark; indeed, we do not even need to speak of top quarks.[5]

---

[4]The emphasis is on "modelling" here: we point out that in parton shower Monte Carlos several choices can be made (e.g. those of the so-called shower variables) that affect the perturbative part of the simulation, which are all compatible with the underlying perturbative description.

[5]We shall ignore backgrounds. In a more realistic analysis, some mild dependence on the definition of top quarks might enter through the subtraction of backgrounds.

Label	Observable
1	$p_T(\ell^+)$
2	$p_T(\ell^+\ell^-)$
3	$M(\ell^+\ell^-)$
4	$E(\ell^+) + E(\ell^-)$
5	$p_T(\ell^+) + p_T(\ell^-)$

Table 1. The set of observables used in this paper, and their labelling conventions.

- It is almost completely inclusive in hadronic radiation: the only possible dependence on strongly-interacting final-state objects is that due to selection cuts (on b-jets).

- Owing to this inclusiveness, the observable is minimally sensitive to the modelling of long-distance effects. This feature increases the reliability of the theoretical predictions.

The set of observables considered in this paper and their labelling conventions are given in table 1: $p_T(\ell^+)$ is the single-inclusive transverse momentum of the positively-charged lepton; $p_T(\ell^+\ell^-)$ and $M(\ell^+\ell^-)$ are the transverse momentum and the invariant mass, respectively, of the charged-lepton pair; finally, $E(\ell^+) + E(\ell^-)$ and $p_T(\ell^+) + p_T(\ell^-)$ are the scalar sums of the energies and transverse momenta of the two charged leptons, respectively. We point out that the latter two sums are computed event-by-event; in other words, observables #4 and #5 are *not* constructed a-posteriori given the single-inclusive energy and transverse momentum distributions of the leptons.

The extraction of the top quark mass utilises the sensitivity of the *shapes* of kinematic distributions to the value of m_t. It is cumbersome to work directly with differential distributions. Instead, we utilise their lower Mellin moments, whose precise definition is given in section 2.1. The idea of the method proposed in this paper is to predict the m_t dependence of the moments, and then to extract the value of m_t by comparing the predicted and measured values of those moments. The procedure is detailed in section 2.2.

The use of moments for the extraction of the top mass has been suggested previously in the context of the so-called J/Ψ method [11], or in connection with variables supposed to minimise the dependence on the jet-energy scale [12, 13]. To our knowledge, the most up-to-date theoretical treatment of this technique is in ref. [14]. All these papers consider only the first moment (of various distributions); in the case of m_t extraction from different observables, the results are either not combined [14], or limited to two observables [13]. These choices may lead to issues, as we shall discuss in sections 2.3, 3.2.2, and 3.2.4. In the case of the dilepton channel, ref. [14] also employs one of the observables considered in this paper ($E(\ell^+) + E(\ell^-)$); owing to the different choices made for cuts, jet algorithm, collider energy, and PDFs, we have refrained from making a direct comparison with those results. We also point out that in ref. [14] the simultaneous variation of the factorisation and renormalisation scales has been adopted, which leads to smaller scale uncertainties than those we find in this paper (where the two scales are varied independently, see section 3).

Finally, we remark that other discrete parameters of kinematic distributions, such as medians and maxima, might also be used for a top mass extraction. We have chosen to work with moments because of the ease of their calculation, and of the fact that the results they give can be systematically improved by including those of increasingly higher rank. For other previous theoretical approaches whose philosophy differs, in one or more aspects, w.r.t. the one adopted here, see e.g. refs. [15–18].

2.1 Definition of moments

We denote by σ and $d\sigma$ the total and fully-differential $t\bar{t}$ cross sections respectively (possibly within cuts), so that:

$$\sigma = \int d\sigma \,, \tag{2.2}$$

where the integral in understood over all degrees of freedom. Given an observable O (e.g. one of those in table 1), its normalised moments are defined as follows:

$$\mu_O^{(i)} = \frac{1}{\sigma} \int d\sigma \, O^i \,, \tag{2.3}$$

for any non-negative integer i. In this way, one has:

$$\mu_O^{(0)} = 1 \,, \qquad \mu_O^{(1)} = \langle O \rangle \,, \qquad \mu_O^{(2)} = \langle O^2 \rangle = \sigma_O^2 + \left(\mu_O^{(1)} \right)^2 \,, \tag{2.4}$$

and so forth. Note that, when selection cuts are applied (see eq. (3.6)) in the calculation of moments, they are applied exactly in the same manner in the denominator and in the numerator of eq. (2.3).

A short technical remark: the numerator of eq. (2.3) is usually derived from the result relevant to the differential distribution $d\sigma/dO$. On the other hand, it could also be computed directly (i.e., without using $d\sigma/dO$), which is a procedure affected by smaller uncertainties, as we explain in appendix A. The important thing to point out here is that such a direct calculation has a fully analogous experimental counterpart: Mellin moments can be measured directly, which is the procedure we recommend. See appendix A for more details.

2.2 Extraction of the top quark mass and its uncertainties

The method for extracting m_t from the first moment of any one of the observables of table 1 is given schematically in figure 1. The x and y axes of figure 1 are associated with the top pole mass m_t, and the first moment of O, $\mu_O^{(1)}$, respectively. The three lines f_C, f_U, and f_L represent the central, upper, and lower theoretical predictions for $\mu_O^{(1)}(m_t)$. The case of moments higher than the first, $\mu_O^{(i)}(m_t)$, $i > 1$, is identical, except for the fact that the x axis of figure 1 is associated with m_t^i (see eq. (3.8)). We have assumed that $\mu_O^{(i)}$ increases with m_t, which is indeed the case in the SM and for the observables considered here; the formulae given below can however be trivially extended to the case of $\mu_O^{(i)}$ decreasing with m_t. As figure 1 suggests, the functions $f_{C,U,L}$ are linear in m_t^i; although in general they
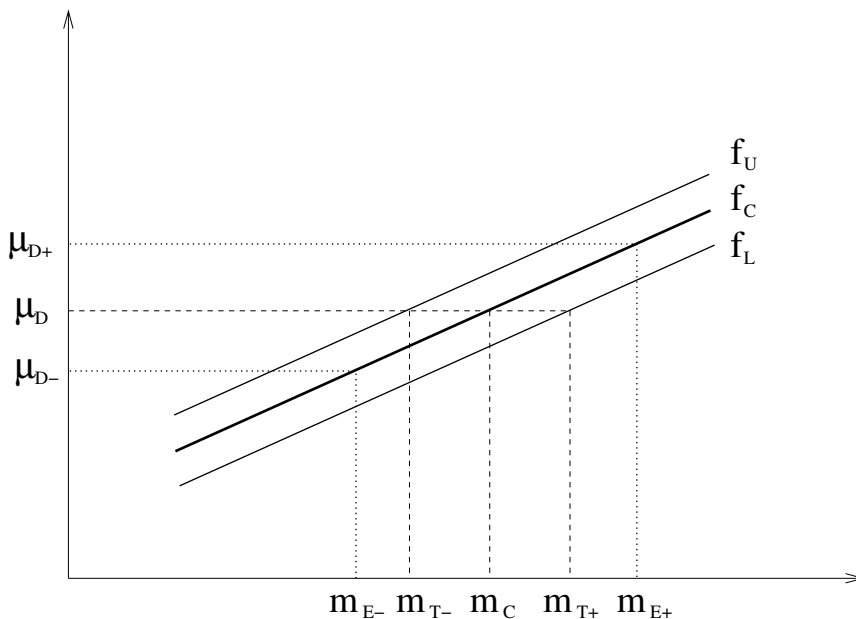
Figure 1. Graphic representation of the method used in this paper to extract the top mass from the first moment of any given observable. The case of higher moment is identical, except for the fact that the x axis is associated with m_t^i.

need not be so, we have found that straight lines are perfectly adequate to our purposes. We explain how such functions are computed in section 3.1.

Given the data[6]

$$\mu_D{}^{+\Delta_\mu^+}_{-\Delta_\mu^-}\,,\tag{2.5}$$

with

$$\Delta_\mu^- = \mu_D - \mu_{D-}\,,\qquad \Delta_\mu^+ = \mu_{D+} - \mu_D\,,\tag{2.6}$$

the extracted top mass will be (see figure 1):

$$m_t = m_C{}^{+\Delta_{mT}^+ + \Delta_{mE}^+}_{-\Delta_{mT}^- - \Delta_{mE}^-}\,.\tag{2.7}$$

We define the central value and theoretical uncertainties associated with such an extraction as follows:

$$\Delta_{mT}^- = m_C - m_{T-}\,,\qquad \Delta_{mT}^+ = m_{T+} - m_C\,,\tag{2.8}$$

with

$$m_C = f_C^{-1}\left(\mu_D\right)\,,\qquad m_{T-} = f_U^{-1}\left(\mu_D\right)\,,\qquad m_{T+} = f_L^{-1}\left(\mu_D\right)\,.\tag{2.9}$$

Since the functions $f_{C,U,L}$ are linear in m_t^i, their inversion is trivial; however, we point out that eq. (2.9) remains valid regardless of the particular (monotonic and increasing) forms of $f_{C,U,L}$. While the quantities introduced in eq. (2.8) are the theory errors that affect the top-mass extraction from any given observable and moment, there might be cases in which

---

[6]Despite the large number of $t\bar{t}$ dilepton events accumulated so far at the LHC, no measurement of these moments is available at present.

they are inadequate to measure the actual difference between the central value m_C and the physical top mass. This happens in the presence of what we call theory *biases*, which we shall discuss at length in section 2.3.

In keeping with figure 1, we define the experimental errors as:

$$\Delta_{mE}^- = m_C - m_{E-} , \qquad \Delta_{mE}^+ = m_{E+} - m_C , \qquad (2.10)$$

with

$$m_{E-} = f_C^{-1} \left(\mu_{D-} \right) , \qquad m_{E+} = f_C^{-1} \left(\mu_{D+} \right) . \qquad (2.11)$$

It is easy to convince oneself that the more conservative choice:

$$m_{E-} = f_U^{-1} \left(\mu_{D-} \right) , \qquad m_{E+} = f_L^{-1} \left(\mu_{D+} \right) , \qquad (2.12)$$

is not correct, since it leads to non-zero uncertainties also in the case of null experimental errors. In this paper, we shall not consider the experimental uncertainties any longer, and shall be concerned only with the theoretical ones. We point out that the size of these depend on two factors: the uncertainty on the theoretical predictions for $\mu_O^{(i)}$, which is $f_U(m_t) - f_C(m_t)$ or $f_C(m_t) - f_L(m_t)$ at a given m_t, and the slope of $f_C(m_t)$: the steeper the latter, the smaller the errors on the extracted values of m_t.

2.3 Theory biases

In this section we address the question whether there might be some biases in the method outlined in section 2.2, that would prevent the errors defined in eq. (2.8) from being a reliable representation of the true uncertainties underlying the m_t extraction.

It is not difficult to devise a scenario where the answer to the previous question is positive. Let us suppose that $t\bar{t}$ production, as is seen in LHC detectors, proceeds through both the well-known SM mechanisms, and the exchange of a hypothetical heavy non-SM resonance. The nature of the latter is irrelevant here; what matters is the fact that, owing to its being very massive, it will cause the t and $\bar{t}$, and hence their decay products, to be slightly more boosted on average than if only SM physics would exist. Thus, for example, the measured first moment of $p_T(\ell^+)$ (which is observable #1 in table 1) will have a larger value than what would be measured if only the SM were present. Let us further suppose that the theoretical predictions used to extract m_t with the procedure of section 2.2 are the pure-SM ones: what has been said above also implies that the functions f_C, f_U, and f_L will have lower values, for any given top mass, than their counterparts computed in the BSM theory that corresponds to the measured data. Figure 1 then leads immediately to the conclusion that the value of m_t extracted from BSM data using SM predictions will be larger than the "true" top mass value:[7] the difference between the extracted and the true m_t is then a theoretical bias. The crucial point is that the uncertainties associated with such an extraction will be essentially the same[8] as those one would obtain in the complete

---

[7]In keeping with section 2.2, we have used the fact that the first moment of $p_T(\ell^+)$ is a growing function of m_t. Qualitatively, the conclusions drawn in this section are independent of the slope of $f_{C,U,L}$.

[8]This is because f_U and f_L are, to a very good extent, parallel to f_C.

absence of BSM physics in the data: in other words, the errors of eq. (2.8) would fail to capture the presence of the existing theoretical bias.

The main lesson to be drawn from the previous example is the following. Given only one observable and one of its Mellin moments, the extraction of the top mass according to eq. (2.7) will always be possible with "small" theoretical errors,[9] regardless of whether the theory employed gives a correct representation of the physics model embedded in Nature. This observation, however, implicitly suggests a solution to the problem posed by theoretical biases. In fact, while the above indeed applies to each individual observable-and-moment choice, if the theoretical description is ultimately not compatible with Nature it is not likely that two values of m_t extracted from two different observables will be compatible with each other. Conversely, the probability that the extracted values of m_t be all mutually compatible (in the case of an incorrect underlying theoretical model) will decrease with the number of observables and moments considered.

Note that it is easier to establish the possible incompatibility of any two m_t results when their theoretical errors are small. Therefore, the property of the uncertainties of eq. (2.8) of being insensitive to theoretical biases is actually a positive feature in this context, and underlines the importance of accurate predictions. The bottom line is that, in order for the presence of theoretical biases to be clearly uncovered, it is of utmost importance to consider as many observables and moments as is possible. The choice of the set of table 1 reflects this view, but it is clear than any further addition to it will be beneficial.

We conclude this section by making various further observations. Firstly, it is not necessary to have a BSM-vs-SM scenario for theoretical biases to appear: it is sufficient that theory and data are not fully compatible. We shall give several examples of this in section 3, all of them within the SM. Secondly, although possibly biased, the m_t value extracted in a single-observable-and-moment procedure is not "wrong": it is, by construction, the result that, for the given data, will give the best prediction with the assumed theoretical model. Therefore, as long as one uses such m_t with that model and only for that observable, one is perfectly consistent. It is in the interpretation of the results, however, that one must be careful, since the larger the bias, the less the extracted top quark mass will have to do with the fundamental parameter so important e.g. for the stability of the vacuum. This stresses again the fact that it is always recommended, for example through the multi-observable approach advocated here, to determine the presence of theoretical biases. Thirdly and finally, the relationship between m_t-extraction and biases is by no means specific to the use of Mellin moments; it is common to all template-based methods. If anything, Mellin moments just render the discussion particularly transparent.

3 Results

Our predictions are obtained by simulating $t\bar{t}$ production in the SM, by treating the top quarks as stable, and by decaying them afterwards. We perform the calculations in the

---

[9]In a reasonable range of m_t. For too large or too small m_t values this statement is not necessarily true: the functions $f_{U,L,C}$ might simply be inadequate to describe correctly the dependence of the moments on the top mass.

Label	Extended name	Accuracy	Parton shower	Spin correlations
1	NLO+PS+MS	NLO	Yes	Yes
2	LO+PS+MS	LO	Yes	Yes
3	NLO+PS	NLO	Yes	No
4	LO+PS	LO	Yes	No
5	fNLO	NLO	No	No
6	fLO	LO	No	No

Table 2. Calculational scenarios considered in this paper. The rightmost column reports the inclusion of *production* spin correlations; decay spin correlations are included in all cases.

fully automated MADGRAPH5_AMC@NLO framework [19], where we can easily investigate the impact of the various approximations that may be employed; in particular, we shall consider both LO and NLO results, with or without their matching to parton showers, with or without including spin-correlation effects. We have thus several calculational scenarios, which we summarise in table 2. We shall refer to each of them interchangeably with either their labels or their extended names, the latter chosen in agreement with ref. [19]. NLO fixed-order computations are based on the FKS subtraction method [20, 21]. NLO results are matched to parton showers according to the MC@NLO formalism [22]; throughout this paper, we have used HERWIG6 [23, 24]. Spin-correlation effects in the computations matched to parton showers are accounted for with the method of ref. [25] through its implementation in MADSPIN [26] (shortened to MS henceforth), a package embedded in MADGRAPH5_AMC@NLO. As far as fixed-order results are concerned, only decay spin correlations (i.e. those described by the matrix elements relevant to $t \to \ell^+ \nu_\ell b$) are taken into account, whence the "No" in the rightmost entry of the last two rows of table 2.

We have used a five-light-flavour scheme, and the MSTW2008 (68% CL) PDF sets [27] and their associated errors, at the LO or the NLO depending on the perturbative accuracy of the various scenarios reported in table 2. We have included both PDF and scale uncertainties in our predictions; both have been computed with the reweighting method of ref. [28]. As far as the latter uncertainties are concerned, they have been obtained with an independent variation of the renormalisation and factorisations scales, subject to the constraints

$$0.5 \leq \xi_F \,, \; \xi_R \,, \; \xi_F / \xi_R \leq 2 \,, \tag{3.1}$$

where

$$\mu_F = \xi_F \hat{\mu} \,, \qquad \mu_R = \xi_R \hat{\mu} \,, \tag{3.2}$$

and $\hat{\mu}$ is a reference scale; the default values or central scale choices correspond to $\xi_F = \xi_R = 1$. We point out that eq. (3.1) is a conservative scale variation (as was done e.g. in ref. [29], and as opposed to setting the two scales equal to a common value), which estimates well the missing higher-order corrections to the total $t\bar{t}$ cross section at the NNLO [30, 31]. We have considered three different functional forms for the reference scale $\hat{\mu}$ in eq. (3.2):

$$\hat{\mu}^{(1)} = \frac{1}{2} \sum_i m_{T,i} \,, \qquad i \in \{t, \bar{t}\} \,, \tag{3.3}$$

$$\hat{\mu}^{(2)} = \frac{1}{2} \sum_i m_{T,i}\,, \qquad i \in \text{ final state}\,, \tag{3.4}$$

$$\hat{\mu}^{(3)} = m_t\,, \tag{3.5}$$

with the transverse masses $m_{T,i} = \sqrt{p_{T,i}^2 + m_i^2}$. We point out that, since in our calculations the top quarks are treated as stable particles at the level of hard matrix elements, the difference between eq. (3.3) and (3.4) is the contribution to the latter of the transverse momentum of the massless parton which is possibly present in the final state (owing to real-emission corrections); the scale of eq. (3.4) is nothing but $H_T/2$.

Our simulations are carried out at the 8 TeV LHC. Since we only consider the process of eq. (2.1), i.e. top-pair production without any background contamination, all of our events are $t\bar{t}$ ones by construction. On the other hand, in order to perform a more realistic analysis, we also impose the following event selection: on top of having two oppositely-charged leptons (electrons and/or muons), events are required to contain at least two b-flavored jets, with jets defined according to the anti-k_T algorithm [32] with $R = 0.5$, as implemented in FASTJET [33]. The events so selected are then subject to the following cuts:

$$\begin{aligned}
\left| \eta(\ell^\pm) \right| &\leq 2.4\,, & p_T(\ell^\pm) &\geq 20 \text{ GeV}\,, \\
\left| \eta(J_b) \right| &\leq 2.4\,, & p_T(J_b) &\geq 30 \text{ GeV}\,.
\end{aligned} \tag{3.6}$$

If more than two b-jets are present, the cuts above are imposed on the two hardest ones. In order to simplify our analysis, b-hadrons have been set stable in HERWIG6, so that the vast majority of the events just contain the two charged leptons arising from top decays. In addition to the cuts of eq. (3.6), we have also checked the effects of imposing lepton-jet isolation cuts: these being negligible, we shall not consider them any further in this paper.

3.1 Calculation of the moments and of the functions $f_{C,U,L}(m_t)$

With the settings described above, we have simulated $t\bar{t}$ production in all of the six calculational scenarios of table 2; in the case of NLO+PS+MS (which we believe to give the best description of SM physics, and is thus treated as our reference computation), results have been obtained with all of the three scales choices of eqs. (3.3)–(3.5), while in all the other cases only the scale of eq. (3.3) has been considered.

Each of these calculations has been performed eleven times, once for each value of the top quark mass chosen in the discrete set:

$$m_t = (168, 169, \ldots, 178) \text{ GeV}\,. \tag{3.7}$$

In each of these runs, we have computed the first four Mellin moments for all the observables listed in table 1, both without applying any cuts, and with the selection cuts of eq. (3.6); all moments are evaluated on the fly (i.e. not a-posteriori using the corresponding differential distribution), as explained in appendix A. At the end of the runs, we have the predictions for the Mellin moments that correspond to the central scales and PDF set, and to all non-central scales and PDFs that belong to the relevant error set; as already explained, all the non-central results do not require additional runs, but are obtained through reweighting.
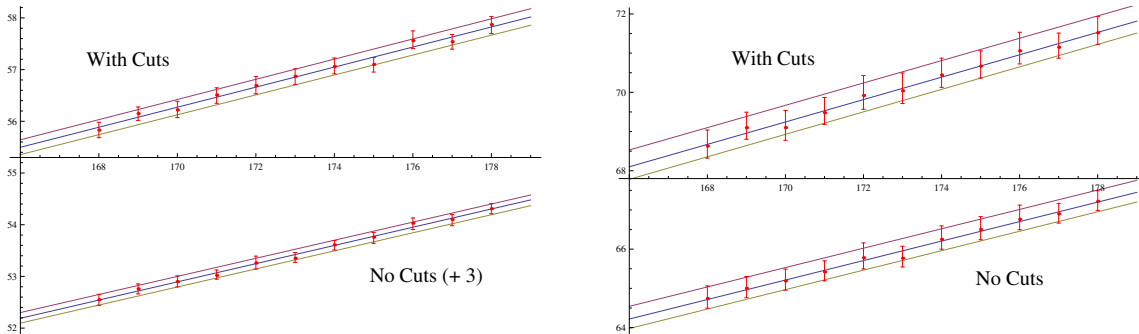
Figure 2. Results (red vertical lines) for the first moment of $p_T(\ell^+)$ (left panel) and $p_T(\ell^+\ell^-)$ (right panel), with and without the cuts of eq. (3.6), in an NLO+PS+MS simulation with the scale of eq. (3.3). In each case, the three lines are the best straight-line fits to the centres or to the upper/lower ends of the theoretical error band, and are thus identified with the functions $f_{C,U,L}$. The lower ("No Cuts") band on the left panel is shifted upwards by $3\,\mathrm{GeV}$ for improved readibility.

The envelopes of the non-central scale and PDF results are then separately constructed. Finally, the scale and PDF uncertainties are combined in quadrature.

The bottom line is that at the end of a given run for each Mellin moment we obtain three numbers: the central, upper, and lower predictions for that moment. Examples of such an outcome, for all the m_t values in the set (3.7), are given in figure 2 in the form of the usual error-bar layout. Both panels of figure 2 are relevant to NLO+PS+MS simulations with the scale choice of eq. (3.3); the one on the left (right) reports the first moment of $p_T(\ell^+)$ ($p_T(\ell^+\ell^-)$), both with and without selection cuts.

Having the above results, the set of the eleven central, or upper, or lower, values for each of the moments is then fitted with the following functional form:

$$\frac{\mu_O^{(i)}(m_t)}{(1\ \mathrm{GeV})^i} = \alpha_O^{(i)}\,173^i + \beta_O^{(i)}\left(\frac{m_t}{1\ \mathrm{GeV}}\right)^i . \tag{3.8}$$

The best fits to the central, upper, and lower moments are finally identified with the functions f_C, f_U, and f_L, respectively, introduced in section 2.2. In the examples of figure 2, these three functions are the three straight lines (for each of the four situations considered there); the analogy with figure 1 is evident. The actual fitted values of the coefficients of eq. (3.8) that correspond to the $f_{C,U,L}$ functions in the case of the selection cuts, together with their analogues for the other three observables not shown in figure 2, are reported in table 3. They will not be explicitly used in what follows, and simply constitute a benchmark for future applications.

We conclude this section by pointing out that statistical integration errors are completely neglected in the fitting procedure described above (which is equivalent to taking all of them equal). In fact, the main reason behind choosing such a large number (11) of m_t values for our simulations is that of minimising the impact of statistical fluctuations, without having to bother about the statistical errors, which are notoriously tricky in the case of the integration of NLO cross sections. The typical size of the statistical fluctuations can be gathered from figure 2; it tends to increase in the case of higher moments, but it

Observable	α (central)	β (central)	α (upper)	β (upper)	α (lower)	β (lower)
$p_T(\ell^+)$	0.1347	0.1939	0.1345	0.1950	0.1353	0.1925
$p_T(\ell^+\ell^-)$	0.1195	0.2857	0.1227	0.2850	0.1166	0.2868
$M(\ell^+\ell^-)$	0.3182	0.3109	0.3246	0.3073	0.3216	0.3046
$E(\ell^+) + E(\ell^-)$	0.5752	0.5100	0.5663	0.5258	0.5636	0.5154
$p_T(\ell^+) + p_T(\ell^-)$	0.2824	0.3755	0.2896	0.3701	0.2765	0.3796

Table 3. Values of the coefficients of eq. (3.8), for the first moments resulting from an NLO+PS+MS simulation with the scale of eq. (3.3); the selection cuts of eq. (3.6) are applied.

remains manageable up to the fourth moment, which is the largest we have considered. Obviously, statistical fluctuations can be reduced by increasing the accuracy of all runs performed. Given the large number of simulations relevant to the present paper, we have limited ourselves to work with 10^6 events (of which, about 30% pass the selection cuts of eq. (3.6)) in the case of computations matched to partons showers, and with a comparable accuracy in the case of fixed-order calculations. With this setup, we have found that over the interval $168 - 178\,\mathrm{GeV}$ the functional form of eq. (3.8) gives an excellent fit up to the fourth moment, and we believe that this conclusion applies regardless of the statistics; in other words, we see no reason for considering polynomials of higher orders in m_t^i in the fitting procedure.

3.2 Extraction of the top quark mass

We now use the predictions for $f_{C,U,L}(m_t)$, calculated as described in section 3.1, to extract the value of the top quark mass according to the procedure outlined in section 2.2. In this way, we shall obtain the main figure of merit relevant to the method proposed in this paper, namely the size of the theoretical errors, eq. (2.8), associated with the extraction. In addition, by comparing the results emerging from the different computational scenarios of table 2, we shall assess the presence and numerical impact of the possible theory biases that affect the various approximations.

In order to carry out the programme just described, we need to start from some data, as in eq. (2.5). In view of the fact that Mellin moments for the observables of table 1 have not been measured at the LHC, we shall generate them ourselves, by using the procedure to be described in section 3.2.1. We point out that theoretically-generated (pseudo)data are actually more advantageous than real data if one is interested in studying the performances of a given procedure, since they provide one with a fully-controlled environment.

All of the theory predictions and pseudodata used in this section have been subject to the selection cuts of eq. (3.6).

3.2.1 Pseudodata

Since we believe that our reference scenario, namely NLO+PS+MS, will give the best description of actual (SM) physics, it is natural to adopt it for the generation of the pseudodata. While well-motivated from a physics viewpoint, we stress that, for the sake of a purely theoretical exercise, this choice is completely arbitrary, and that the conclusions

we shall arrive at would be qualitatively unchanged had we chosen a different scenario. The pseudodata are generated by setting:

$$m_t^{\mathrm{pd}} = 174.32 \ \mathrm{GeV} \,,$$
$$\mu_F^{\mathrm{pd}} = 0.45 \left(m_{T,t} + m_{T,\bar{t}} + 0.5 p_{T,p}\right) \,,$$
$$\mu_R^{\mathrm{pd}} = 0.60 \left(m_{T,t} + m_{T,\bar{t}} + 0.3 p_{T,p}\right) \,, \tag{3.9}$$

where in the definitions of the scales use is made of the transverse masses of the top and the antitop, and of the transverse momentum of the massless parton possibly present in the final state at the hard-subprocess level.

With the choices in eq. (3.9), pseudodata generation does not correspond to any of the scenarios of table 2 and to any of the scales of eqs. (3.3)–(3.5). Therefore, one must not expect that the extractions of the top mass will return exactly m_t^{pd}, owing to the presence of the biases discussed in section 2.3. Having said that, we expect pseudodata to show a clear "preference" (i.e. smaller biases) towards simulations based on the NLO+PS+MS scenario, since in those cases the biases must be due only to scale choices. The verification that this is indeed the case will constitute a self-consistency check of the procedure we are following. Note that the information relevant to theory biases is encoded not in the actual value of the extracted m_t, but in its difference with m_t^{pd}. Because of the behaviour of the $f_{C,U,L}$ functions, such a difference is very much insensitive to the choice of m_t^{pd}, which allows one to pick an arbitrary value for the latter, as is done in eq. (3.9), and which is ultimately the reason for the robustness of the usage of pseudodata.

3.2.2 Shower, NLO, and spin-correlation effects

The scenarios of table 2 differ by the various approximations they are based upon, each of which may lead to biases in the extraction of the top mass. An interesting question is then whether the different sources of possible biases can be disentangled from each other (i.e. whether in a sense they factorise). This is not only relevant in the context of the present exercise, but also because it may help assess the impact of approximations not considered here (such as NNLO corrections), and which might become crucial in the presence of real data. Furthermore, it also sheds light on the characteristics of the various observables used in this paper, and in so doing suggests how to enlarge their set.

In order to address the items above, we proceed as follows. We select pairs of scenarios that differ in one and only one aspect of the approximations they involve; for example, scenarios #1 and #2 differ in the perturbative accuracy (NLO vs LO) of the underlying computations. The aspects that we shall be able to consider are three, namely parton-shower, NLO-correction, and spin-correlation effects, which we shall discuss in turn below. The top mass extracted within scenario #i will be denoted by:

$$m_t^{(i)} \,. \tag{3.10}$$

Let us then suppose to have chosen a pair of scenarios (#i, #j) that differ only by aspect

obs.	$m_t^{(3)} - m_t^{(5)}$	$m_t^{(3)} - m_t^{\mathrm{pd}}$	$m_t^{(4)} - m_t^{(6)}$	$m_t^{(4)} - m_t^{\mathrm{pd}}$
1	$-0.35^{+1.14}_{-1.16}$	$+0.12$	$-2.17^{+1.50}_{-1.80}$	-0.67
2	$-4.74^{+1.98}_{-3.10}$	$+11.14$	$-9.09^{+0.76}_{-0.71}$	$+14.19$
3	$+1.52^{+2.03}_{-1.80}$	-8.61	$+3.79^{+3.30}_{-4.02}$	-6.43
4	$+0.15^{+2.81}_{-2.91}$	-0.23	$-1.79^{+3.08}_{-3.75}$	-1.47
5	$-0.30^{+1.09}_{-1.21}$	$+0.03$	$-2.13^{+1.51}_{-1.81}$	-0.67

Table 4. Impact of parton showers on mass extractions. See the text for details.

A. What we may consider are the quantities:

$$m_t^{(i)} - m_t^{\mathrm{pd}}, \qquad m_t^{(j)} - m_t^{\mathrm{pd}}, \tag{3.11}$$

$$m_t^{(i)} - m_t^{(j)}. \tag{3.12}$$

While the differences in eq. (3.11) are sensitive to all theory biases that affect scenarios $\#i$ and $\#j$, we expect that the difference in eq. (3.12) is solely sensitive to the effect of **A** (if the factorisation property mentioned above holds to some extent). In the following, we report the differences that appear in eq. (3.11) and (3.12),[10] for all the relevant $(\#i, \#j)$ pairs and all the observables of table 1. We shall limit ourselves to considering the first moments, which are sufficient for the sake of the present exercise; all results are obtained with the scale of eq. (3.3). In the case of eq. (3.12), which is our main interest here, we also report the errors affecting the difference, which is computed by combining in quadrature the errors (determined according to eq. (2.8)) that affect the individual $m_t^{(i)}$ and $m_t^{(j)}$ values. The errors on the differences in eq. (3.11) are of comparable size, up to a factor $\sqrt{2}$ smaller since m_t^{pd} is assumed to be known with infinite precision.

We start with shower effects, and report the corresponding results in table 4. The relevant scenario pairs are $(3, 5)$ and $(4, 6)$, the latter being the LO counterpart of the former, which is accurate to NLO. Note that scenarios $\#1$ and $\#2$ have not been considered here, owing to the lack of fixed-order results that include production spin correlations. The first observation is that the $(3, 5)$ and $(4, 6)$ cases are rather consistent with each other; however, the results for eq. (3.12) of the latter are in absolute value systematically larger than those of the former. This is compatible with the expectation, corroborated by ample heuristic evidence in many different processes, that shower effects are milder if the underlying computations are NLO-accurate (as opposed to LO ones), for the simple reason that NLO results do already include part of the radiation to be generated by parton showers.[11] While in the case of NLO-based simulations all differences are statistically compatible with zero (within 1σ) except for observable $\#2$, in the case of LO-based simulations more significant deviations can be seen in the cases of observables $\#1$ and $\#5$ as well. The take-home message, then, is that shower effects are moderate if higher-order corrections are taken

---

[10]Owing to the linear dependence of these three quantities, only one of the differences in eq. (3.11) will be shown.

[11]This also shows that NLO and shower effects do not factorise entirely; it remains true that they affect the m_t extraction for a given observable in different manners.

obs.	$m_t^{(1)} - m_t^{(2)}$	$m_t^{(1)} - m_t^{\mathrm{pd}}$	$m_t^{(3)} - m_t^{(4)}$	$m_t^{(3)} - m_t^{\mathrm{pd}}$	$m_t^{(5)} - m_t^{(6)}$	$m_t^{(5)} - m_t^{\mathrm{pd}}$
1	$+1.16^{+1.43}_{-1.60}$	$+0.41$	$+0.79^{+1.43}_{-1.60}$	$+0.12$	$-1.03^{+1.22}_{-1.43}$	$+0.47$
2	$-2.79^{+1.27}_{-1.65}$	-1.18	$-3.05^{+1.35}_{-1.64}$	$+11.14$	$-7.41^{+1.64}_{-2.72}$	$+15.87$
3	$-0.73^{+3.21}_{-3.45}$	$+0.84$	$-2.18^{+3.03}_{-3.30}$	-8.61	$+0.09^{+2.42}_{-2.91}$	-10.13
4	$+1.74^{+3.27}_{-3.78}$	$+0.16$	$+1.23^{+3.10}_{-3.61}$	-0.23	$-0.70^{+2.79}_{-3.09}$	-0.38
5	$+0.99^{+1.42}_{-1.72}$	$+0.25$	$+0.70^{+1.40}_{-1.72}$	$+0.03$	$-1.13^{+1.23}_{-1.33}$	$+0.33$

Table 5. Impact of NLO corrections on mass extractions. See the text for details.

into account, which is good news in view of the future availability of NNLO parton-level differential results; however, this conclusion does not apply to the transverse momentum of the charged-lepton pair, for which a proper matching with parton showers appears to be needed.

As far as the results for eq. (3.11) are concerned, table 4 shows that values significantly different from zero are obtained in the cases of observables #2 and #3. The size of the difference relevant to #2 is larger than that resulting from eq. (3.12), which implies that for such an observable other effects, on top of those due to showers, are sources of theory biases as well (both NLO and spin correlations, as we shall show later). A similar conclusion applies to the lepton-pair invariant mass #3, for which the absence of shower effects implies that biases are entirely due to some other mechanism (spin correlations, as it will turn out).

We next consider NLO effects, which we document in table 5, and for which the relevant scenario pairs are $(1,2)$, $(3,4)$, and $(5,6)$. As far as eq. (3.12) is concerned, the differences for all pairs and all observables except #2 are compatible with zero; thus, the first moments of such observables appear to be quite stable perturbatively, regardless of the matching to parton showers, and of the presence of spin correlations. For what concerns $p_T(\ell^+\ell^-)$, on top of the fact that NLO effects are significant in all scenarios, we observe that they are particularly strong when the matching to showers is not performed (pair $(5,6)$); this is again related to the fact that, in certain corners of the phase space, showers and NLO corrections affect the kinematics in a similar way. Coming to the absolute size of theory biases, eq. (3.11), we see that they are all rather small in the case of NLO+PS+MS predictions (second column); this is what we expect, as explained in section 3.2.1. For the other scenarios, large differences are observed in the case of observables #2 and #3, which was expected in view of table 4. For the latter observable, this fact, the absence of NLO effects, and the results of table 4 imply that the biases are solely due to spin correlations.

We finally turn to spin-correlations effects, whose results are reported in table 6, and for which the relevant scenario pairs are $(1,3)$ and $(2,4)$; these two pairs differ in the underlying perturbative accuracy, which is NLO and LO respectively. The conclusions that can be drawn from table 6 have already been anticipated. Namely, that the differences resulting from eq. (3.12) are significantly different from zero for both observables #2 and #3, while they are negligible in the other cases. The sizes of the former differences appear to be fairly insensitive to NLO corrections, which is an indirect confirmation of the factorisation of spin-correlation effects.

obs.	$m_t^{(1)} - m_t^{(3)}$	$m_t^{(1)} - m_t^{\rm pd}$	$m_t^{(2)} - m_t^{(4)}$	$m_t^{(2)} - m_t^{\rm pd}$
1	$+0.29^{+1.17}_{-1.14}$	$+0.41$	$-0.08^{+1.66}_{-1.96}$	-0.75
2	$-12.32^{+1.62}_{-2.13}$	-1.18	$-12.58^{+0.90}_{-0.94}$	$+1.60$
3	$+9.45^{+2.36}_{-2.16}$	$+0.84$	$+8.00^{+3.74}_{-4.26}$	$+1.57$
4	$+0.39^{+2.93}_{-3.16}$	$+0.16$	$-0.11^{+3.42}_{-4.16}$	-1.58
5	$+0.22^{+1.12}_{-1.28}$	$+0.25$	$-0.06^{+1.65}_{-2.07}$	-0.73

Table 6. Impact of spin correlations on mass extractions. See the text for details.

The general conclusion of this section is the following. Observables that are single-inclusive ($p_T(\ell^+)$), and that feature a mild correlation between the decay products of the top and antitop ($E(\ell^+) + E(\ell^-)$ and $p_T(\ell^+) + p_T(\ell^-)$), are rather stable against shower, NLO, and spin-correlations effects. This is not true for observables for which the correlation between the two charged leptons is stronger ($p_T(\ell^+\ell^-)$ and $M(\ell^+\ell^-)$): the fact that either shower or spin-correlation effects (or both) are relevant implies, among other things, that the computation of the $t\bar{t}$ cross section at the NNLO with stable tops will not be sufficient to give a good description of such observables, at the very least in the context of the top mass extraction considered in this paper.

3.2.3 Results for the top quark mass

In this section we present the results for the extraction of the top quark mass obtained with our reference computational scenario, NLO+PS+MS. We are specifically interested in checking the size of the theory uncertainty affecting such an extraction, and its behaviour (together with that of the central top quark mass) when the results emerging from the individual observables and moments are combined together. These findings will also serve as benchmarks for the studies that we shall carry out in section 3.2.4, where the extraction of the top mass will be performed by using the other scenarios of table 2. Furthermore, we want to study how the above results are influenced by the scale choice, and therefore we shall consider all of the three forms given in eqs. (3.3)–(3.5).

The general strategy is the following. For a given scale choice, we extract the top mass from each of the five observables of table 1 and their first three moments,[12] i.e. fifteen m_t values in total, each with its theory errors of eq. (2.8). These values, or any subset of them, are then combined to obtain the "best" result. The combination technique is briefly explained in appendix B, and is rather standard: basically, the central values are weighted with the inverse of the square of their errors. Since the various observables and their moments are correlated, it is necessary to take these correlations into account, lest one skew the final central value of m_t and underestimate its error.

The simplest case is that where one uses a single observable for extracting m_t; as was explained in section 2.3, this is far from being ideal, and we present it here only as a way to

---

[12]The fourth moments turn out not to be particularly useful in the extraction procedure, being affected by errors larger than those of the lower moments, and being rather strongly correlated with the third moments; these are the reasons why they are not taken into account.

scale	$i = 1$	$i = 1 \oplus 2$	$i = 1 \oplus 2 \oplus 3$
1	$174.73^{+0.80}_{-0.79}[0.2]$	$174.73^{+0.80}_{-0.79}[0.2]$	$174.72^{+0.80}_{-0.79}[0.2]$
2	$174.78^{+0.90}_{-0.90}[0.6]$	$174.78^{+0.90}_{-0.90}[0.6]$	$174.78^{+0.90}_{-0.90}[0.6]$
3	$172.73^{+2.0}_{-1.2}[0.5]$	$172.73^{+1.96}_{-1.19}[0.5]$	$172.73^{+1.96}_{-1.19}[0.5]$
$1 \oplus 2 \oplus 3$	$174.46^{+0.99}_{-0.92}$	$174.46^{+0.99}_{-0.92}$	$174.45^{+0.99}_{-0.92}$

Table 7. Top mass values extracted from observable #1, with up to three moments, and for three different scale choices. The last line reports the results obtained by combining the central m_t values relevant to the three scales. The numbers in square brackets are χ^2/n. The pseudodata top mass is $m_t^{\mathrm{pd}} = 174.32 \, \mathrm{GeV}$. See the text for details.

compare with the multi-observable results that will be shown later. We use observable #1 ($p_T(\ell^+)$) because it is the one whose top-mass extractions are affected by the smallest errors (in the case of the scale of eq. (3.3)). The values of m_t that we obtain are given in table 7, which should be read as follows (this layout will be used for the other tables of this section as well). Each one of the first three rows corresponds to one of the scales of eqs. (3.3)–(3.5) (i.e. the ith row is obtained with $\hat{\mu}^{(i)}$). The first, second, and third column reports the results obtained by considering only the first, up to the second, and up to the third Mellin moments, respectively. The results in the fourth row are obtained by combining the three results that appear in the first three rows of the same column. Such a combination is achieved by weighting those three results with the inverse of the square of their errors. Since the errors are asymmetric, one treats separately the $+$ and $-$ ones; the two resulting "central" m_t values are possibly different, and the single m_t reported in table 7 is then obtained again with a weighted average. Finally, the numbers in square brackets are the values of χ^2 per degree of freedom, computed by always considering the first four Mellin moments, regardless of how many of them had been actually used in the combination. One should not seek a deep meaning in this χ^2, in particular because of the way the errors that enter into it are obtained (i.e. their behaviour from a statistical viewpoint is unknown to us). On the other hand, while its precise value is not of particular significance, it represents a very useful reference for the performance of the extraction procedure, as we shall see in section 3.2.4.

The messages to be taken out of table 7 are the following. Firstly, the impact of the addition of moments beyond the first is extremely modest, if visible at all. This is due to the fact that the errors affecting m_t increase with higher moments, and to the non-negligible correlations between the moments (see appendix B). Secondly, the scales $\hat{\mu}^{(1)}$ and $\hat{\mu}^{(2)}$ tend to give central results larger than the "true" one of the pseudodata, $m_t^{\mathrm{pd}} = 174.32$ GeV, while the opposite applies to scale $\hat{\mu}^{(3)}$, where the effect is more evident (but still within 1σ). Let us then consider the latter case to be definite, and compare the functional form of eq. (3.5) with those of eq. (3.9). Because of the dependence on the transverse momenta of the scales used in the pseudodata, which is absent in the case of $\hat{\mu}^{(3)}$, the tails of the p_T-related distributions obtained with $\hat{\mu}^{(3)}$ will be less rapidly falling than those of the pseudodata (mainly because the p_T-dependence of μ_R in eq. (3.9) will induce a stronger α_S suppression, relative to the small-p_T region, than in the case of $\hat{\mu}^{(3)}$; this effect is only

scale	$i=1$	$i=1\oplus 2$	$i=1\oplus 2\oplus 3$
1	$174.67^{+0.75}_{-0.77}[3.0]$	$174.67^{+0.75}_{-0.77}[3.0]$	$174.61^{+0.74}_{-0.77}[3.2]$
2	$174.81^{+0.83}_{-0.80}[6.2]$	$174.80^{+0.82}_{-0.80}[6.2]$	$174.85^{+0.82}_{-0.80}[6.1]$
3	$172.63^{+1.85}_{-1.16}[0.2]$	$172.64^{+1.82}_{-1.15}[0.2]$	$172.58^{+1.81}_{-1.15}[0.2]$
$1\oplus 2\oplus 3$	$174.44^{+0.92}_{-0.87}$	$174.44^{+0.92}_{-0.87}$	$174.43^{+0.91}_{-0.87}$

Table 8. As in table 7, with the extractions performed by using observables #1, #4, and #5. The pseudodata top mass is $m_t^{\mathrm{pd}} = 174.32\,\mathrm{GeV}$.

scale	$i=1$	$i=1\oplus 2$	$i=1\oplus 2\oplus 3$
1	$174.48^{+0.73}_{-0.77}[5.0]$	$174.55^{+0.72}_{-0.76}[5.0]$	$174.56^{+0.71}_{-0.76}[5.1]$
2	$174.73^{+0.77}_{-0.80}[4.3]$	$174.74^{+0.76}_{-0.79}[4.3]$	$174.91^{+0.75}_{-0.79}[4.1]$
3	$172.54^{+1.03}_{-1.07}[1.6]$	$172.46^{+0.99}_{-1.05}[1.6]$	$172.22^{+0.95}_{-1.04}[1.4]$
$1\oplus 2\oplus 3$	$174.16^{+0.81}_{-0.85}$	$174.17^{+0.80}_{-0.84}$	$174.17^{+0.78}_{-0.84}$

Table 9. As in table 7, with the extractions performed by using all observables. The pseudodata top mass is $m_t^{\mathrm{pd}} = 174.32\,\mathrm{GeV}$.

mildly compensated by that due to μ_F). Thus, the moments computed with scale #3 will be slightly larger than their analogues in the pseudodata. For the reasons explained in section 2.3, this difference then results in a lower (than the input m_t^{pd}) value for the extracted top mass, which is what we see in the third row of table 7. The same effect, but (slightly) in the opposite direction, is at play in the case of scales #1 and #2. Here, the numerical values of such scales at large p_T's relative to their small p_T counterparts are closer to those relevant to the pseudodata scales than in the case of scale #3, whence closer-to-m_t^{pd} central results for the top mass. Given these opposite behaviours, not surprisingly the average of the three results is closer to m_t^{pd} than any of them; such an average is biased towards the results of $\hat{\mu}^{(1)}$ and $\hat{\mu}^{(2)}$, owing to their errors being smaller than those associated with the extractions with $\hat{\mu}^{(3)}$.

We now repeat the combination procedure that has led to the results of table 7, by including, on top of the m_t values obtained with observable #1, also those relevant to observables #4 and #5; the new combined results are presented in table 8. By far and large, all comments relevant to table 7 can be repeated here. There is a decrease (less than 10% for all scales) of the errors, which is not large because of two facts: observable #1 induces the smallest errors (in the present observable set), and the observables considered are sizably correlated, as documented in appendix B. By adding more observables one starts to see the effects of the inclusion of higher moments; although statistically not significant, there are trends in the central values which were not visible in the case of a single observable.

Finally, in table 9 we present the results obtained by combining the extractions of m_t from all observables; thus, according to the discussion given in section 2.3, these have to be considered our best estimates of the top mass, given the pseudodata of section 3.2.1. The

errors decrease further w.r.t. those of table 8 (not significantly in the case of $\hat{\mu}^{(1)}$ and $\hat{\mu}^{(2)}$, but by a large factor for $\hat{\mu}^{(3)}$; this is because, for such a scale, it is the p_T of the lepton pair that happens to be affected by the smallest errors). The trend induced by the addition of higher moments becomes more visible than before, and statistically significant (a 2σ effect) in the case of scale #3. However, the final results of the fourth row, obtained by combining the outcomes associated with the different scales, are quite stable. The case of the results associated with $\hat{\mu}^{(3)}$ is interesting, because it stresses again the importance of considering as many observables and as diverse as possible in order to expose potential theory biases in the top-mass extraction. Given that here all our predictions are based on the same computational scenario as the pseudodata, namely NLO+PS+MS, the only deviations from a perfect reconstruction can only be due to the different choice of scales, and $\hat{\mu}^{(3)}$ happens to be farther from the pseudodata ones of eq. (3.9) than either $\hat{\mu}^{(1)}$ or $\hat{\mu}^{(2)}$. The crucial point is that this observation is true regardless of the type of observables considered, but it is only when the lepton-pair correlations #2 and #3 enter the combination that the effects become more noticeable. This is related to the behaviour of these two observables discussed in section 3.2.2, which exhibit the strongest sensitivity to (among other things) extra radiation. A change of scale is an effective, if quite mild, way of probing some of these extra-radiation effects. As we shall see in section 3.2.4, the impact of the addition of these two observables on the theory biases is spectacular when the underlying calculational scenario is different w.r.t. that used in the generation of the pseudodata.

There are two conclusions that can be drawn from this section. The first is that the procedure proposed in this paper appears to be able to give theory errors on the extracted top mass of the order of $0.8\,\mathrm{GeV}$. While we have neglected background contaminations, we have also been conservative with the range of scale variations; on top of that, the addition of further observables may help reduce further those errors. The second conclusion is more general, in that it applies to any extraction method based on templates. Our exercise demonstrates that one thing is the variation of the scales induced by pre-factors that multiply a given functional form, and quite another the change of that functional form. Although the two procedures overlap, they are not equivalent. We have shown a practical way to probe the changes of the above functional form: the idea is that, by re-computing theoretical predictions for many different scale choices, and by performing a weighted average of their outcomes, one might effectively capture the scale settings which optimally describe Nature.

3.2.4 More on theory biases

The aim of this section is that of repeating what has been done in section 3.2.3, for scenarios other than NLO+PS+MS. In other words, all of the computations considered here are different w.r.t. that used in the generation of the pseudodata; we shall thus study the theory biases, whose sources we have already discussed in section 3.2.2, at the level of the combined results for the extracted top quark mass. All the calculations are performed by using the scale $\hat{\mu}^{(1)}$. We report the results in table 10, which is organised with the same conventions as those used in the tables of section 3.2.3. This table is split into two parts, relevant to the m_t extraction performed by using only three observables (#1, #4, and #5),

Scenario	$i = 1$	$i = 1 \oplus 2$	$i = 1 \oplus 2 \oplus 3$
	Observables #1, #4, #5		
LO+PS+MS	$173.61^{+1.10}_{-1.34}[1.0]$	$173.63^{+1.10}_{-1.34}[1.0]$	$173.62^{+1.10}_{-1.34}[1.0]$
NLO+PS	$174.40^{+0.75}_{-0.81}[3.5]$	$174.43^{+0.75}_{-0.81}[3.5]$	$174.60^{+0.75}_{-0.79}[3.2]$
LO+PS	$173.68^{+1.08}_{-1.31}[0.8]$	$173.68^{+1.08}_{-1.31}[0.9]$	$173.75^{+1.08}_{-1.31}[0.9]$
fNLO	$174.73^{+0.72}_{-0.74}[5.5]$	$174.72^{+0.71}_{-0.74}[5.6]$	$175.18^{+0.64}_{-0.71}[4.6]$
fLO	$175.84^{+0.90}_{-1.05}[1.2]$	$175.75^{+0.89}_{-1.05}[1.2]$	$175.82^{+0.89}_{-1.04}[1.2]$
	All observables		
LO+PS+MS	$175.98^{+0.63}_{-0.69}[16.9]$	$176.05^{+0.63}_{-0.68}[17.8]$	$176.12^{+0.61}_{-0.68}[18.9]$
NLO+PS	$175.43^{+0.74}_{-0.80}[29.2]$	$176.20^{+0.73}_{-0.79}[30.1]$	$175.67^{+0.73}_{-0.76}[31.2]$
LO+PS	$187.90^{+0.6}_{-0.6}[428.3]$	$187.71^{+0.60}_{-0.60}[424.2]$	$187.83^{+0.58}_{-0.60}[442.8]$
fNLO	$174.41^{+0.72}_{-0.73}[96.6]$	$174.82^{+0.71}_{-0.73}[93.1]$	$175.44^{+0.70}_{-0.68}[94.8]$
fLO	$197.31^{+0.42}_{-0.35}[2496.1]$	$197.19^{+0.42}_{-0.35}[2505.6]$	$197.48^{+0.36}_{-0.35}[3005.6]$

Table 10. Combined extracted values of m_t, for various scenarios and two choices of the set of observables. The pseudodata top mass is $m_t^{\mathrm{pd}} = 174.32\,\mathrm{GeV}$.

or all of them. These two parts thus are in one-to-one correspondence with (the first row of) tables 8 and 9, respectively.

From the upper part of table 10, we see that the use of observables #1, #4, and #5 leads to central m_t values which may not be in perfect agreement with the pseudodata value m_t^{pd}, but are not far from it either, irrespective of the calculational scenario considered. Furthermore, both the errors and the χ^2 values are totally reasonable, and rather consistent with those of table 8. These findings need not be surprising, because they could be anticipated in section 3.2.2, where observables #1, #4, and #5 have been shown to be fairly insensitive to shower, NLO, and spin-correlation effects. These effects are ultimately the difference between each of the scenarios considered here, and our reference one, NLO+PS+MS. It is therefore instructive to see what happens when observables #2 and #3 are used in the extractions as well (lower part of table 10). Not only the differences among the central results for the extracted top mass are much larger than before (and particularly so at the LO in absence of proper spin correlations), but it is especially the χ^2 values that increase dramatically, in spite of (and, in a sense, thanks to) the fact that the errors remain quite moderate. This is exactly the situation that has been described in section 2.3: the extraction of m_t from individual observables is always acceptable and affected by small errors; however, if the underlying theoretical description is incompatible with that of the (pseudo)data, the different results will be mutually incompatible. A (certainly non-unique) way of making explicit the presence of such incompatibilities is through the computation of a χ^2. The lower part of table 10 is thus another, very explicit way of showing why considering a large number of observables with different characteristics is always beneficial, in this or in other template-based methods.

A final comment on table 10. The errors that affect the extracted top mass do not follow the usual LO→NLO reduction pattern, and they need not to. Indeed, the relationship between the above errors, and those which are usually considered at the level of rates, is rather indirect. Furthermore, in the combination of the results obtained from different observables, a single m_t value affected by errors much smaller than the others will have a very large weight, with the picture being further complicated by the presence of strong correlations among the observables studied here. While the particular combination technique used in this paper (see appendix B) can certainly be refined, possibly leading to changes in the central values of m_t and their associated errors, the conclusions reached before will not change, being based on a few well-understood physics phenomena.

4 Conclusions

In this paper we have proposed a procedure for the determination of the top quark pole mass from dilepton $t\bar{t}$ events. Our main proposals and findings are the following:

- We use leptonic single-inclusive and correlation observables, which are clean and largely insensitive to the modelling of long-distance effects. Our method, based on Mellin moments, relies neither on the definition of the top quark as a pseudo-particle, nor on its reconstruction.

- The quality of the results for m_t and their reliability improves by increasing the number of observables and of their moments. It is important that the observables employed have different sensitivities to the various mechanisms relevant to $t\bar{t}$ production and decay, such as higher-order corrections, and shower and spin-correlation effects. Several theoretical simulations must be used that differ in the choice of the functional form for the hard scales, and the extracted m_t values must be combined. Thus, we consider the entry in the rightmost column and last row of table 9 as our "best" result.

- The errors associated with m_t may underestimate the difference between the extracted value and the actual pole mass, in the case of an inadequate theoretical description of the underlying production mechanism. A χ^2-type test is effective in identifying the presence of such biases, *provided* that a sufficiently large number of observables has been employed in the extraction procedure, as is documented in table 10.

We stress that the second and third items above apply to any template method that exploits the shapes of observables for the extraction of the top quark mass.

The most precise m_t determination that we have achieved with our method in the context of the purely-theoretical exercise performed here is affected by errors of the order of $0.8\,\text{GeV}$. It is probably possible to reduce this figure further, by using a set of observables larger than the one considered in this paper. On the other hand, we have not addressed two important aspects which will need to be taken into account in an extraction of m_t from real data, namely the contamination due to backgrounds, and the systematics due to the

choice of the parton shower Monte Carlo. For what concerns this Monte-Carlo systematics, it is worth pointing out that within our approach two different Monte Carlos must lead to two separate top quark mass values, which should eventually be combined on the basis of their respective errors and of the results of some χ^2 tests. An interesting, if not particularly desirable, case is that where two Monte Carlos would lead to statistically-incompatible m_t results, with two small χ^2 values similar to each other. This implies that the observables chosen in the extraction procedure do not constrain well enough the theoretical models adopted by the Monte Carlos, and it is thus doubtful which (if any) of the two m_t results best describes the "physical" pole mass.

While we believe that our approach has many competitive features, it remains true that the determination of the top quark mass will benefit from the use of many different techniques. For example, any BSM physics able to modify in a significant manner the kinematic distributions w.r.t. those predicted by the SM may induce large biases in template-based m_t extractions, unless the simulation of such BSM contribution is also taken into account. In this case, an approach insensitive to the production dynamics (which thus belongs to the first class introduced in section 1) would offer a valuable addition; one may mention here the CMS end-point method [34], or the promising energy-peak method suggested in ref. [35, 36], provided that it could be extended to include NLO QCD corrections to top decays.

The approach we have pursued here has many variants which do not change its essence. For example, one may start looking into b-jet variables in order to increase the sensitivity to m_t; this has the downside of introducing a larger dependence on long-distance modelling, and the balance between these two competing aspects must be carefully addressed. Conversely, one can try and select dilepton events of opposite flavour without imposing cuts on the b jets, in order to further reduce the impact of hadronisation; the problem then becomes that of the control of the backgrounds. Our method is also immediately applicable in the context of NNLO simulations. However, for this to be effective, a proper description of top decays, and in particular one that incorporates production-spin-correlation effects, must be included. The matching to parton shower would also be highly desirable.

Acknowledgments

A Computation of moments in the context of event generation

While the moments of an observable O can be computed by using the result for the differential distribution $d\sigma/dO$, there is actually a more direct way. During the course of an MC

simulation, the fully-differential cross section is expressed through a set of N kinematic configurations ("events") and their associated weights:

$$d\sigma \; \leftrightarrow \; \left\{ \mathcal{K}_k, W_k \right\}_{k=1}^{N} , \tag{A.1}$$

with

$$\sigma = \sum_{k=1}^{N} W_k \,. \tag{A.2}$$

Note that the W_k's need not necessarily be equal to each other (in absolute value); in other words, what follows is valid in the context of both unweighted and weighted event generation, these being typically relevant to calculations matched to parton-shower Monte Carlos and at fixed order, respectively. When one computes a differential cross section, one evaluates event-by-event the value of the observable of interest in the generated kinematic configuration, $O(\mathcal{K}_k)$; such a value determines, in turn, the bin of the corresponding histogram where the weight W_k must be stored. In a completely analogous manner, the calculation of the (unnormalised) moments can also be performed on the fly. In order to do so, for a given observable O one will book a histogram with bins of width one centered at non-negative integers. When the kth event is generated, one stores the weight:

$$W_k \times \left(O(\mathcal{K}_k) \right)^i \tag{A.3}$$

in the ith bin of the histogram; this must be done for all bins. By using eqs. (2.3), (A.2), and (A.3) , one sees that at the end of the run the ith bin will be equal to the normalised i^{th} moment, times σ, so that the normalised moments themselves can be obtained by dividing the content of each bin by that of the bin centered at zero.

We point out that this direct way of computing moments is exact in the $N \to \infty$ limit. On the other hand, the (indirect) calculation which uses the result of $d\sigma/dO$ is *not* exact even in the $N \to \infty$ limit, unless the limit of vanishing bin size (in $d\sigma/dO$) is taken as well, which is impossible in the context of an actual simulation, where one thus might have a residual bin-size inaccuracy. Furthermore, in the case where the range of the histogram in O does not cover the whole kinematically-accessible range for such an observable, another inaccuracy affects the indirect computation. For these reasons, and for its greater simplicity, in this paper we have always adopted the direct, event-by-event method outlined above in the calculation of the moments. We have checked, in the case of the first moments, that the results of the direct computations are very similar, but not identical, to those obtained a-posteriori by using the distributions. It must be stressed that the distributions we have used cover rather large ranges (up to $400\,\mathrm{GeV}$ for observables #1, #2, and #3, up to $1.2\,\mathrm{TeV}$ for observable #4, and up to $1\,\mathrm{TeV}$ for observable #5), and contain 100 bins. Therefore, in the context of e.g. an experimental analysis, where the use of large-size bins is typical at large momenta, the risk of inaccuracies affecting the moments computed from distributions may be non negligible.

B Combination of different top quark mass results

In this appendix we briefly outline the technique used to combine different m_t results and their errors. We denote these by:

$$m_t^{(\alpha)} = \overline{m}_t^{(\alpha)} \pm \delta m_t^{(\alpha)} \,, \tag{B.1}$$

where the index α identifies unambiguously a Mellin moment of a given observable (so that, for example, when considering all of the five observables of table 1 and their first three moments, as has been done in the rightmost column of table 9, α can take fifteen different values). Note that this notation does *not* have the same meaning of the very similar one used in section 3.2.2. The central value of the top mass that results from the combination of the values in eq. (B.1) and its standard deviation are taken to be:

$$m_t = \sum_{\alpha=1}^{M} w_\alpha \overline{m}_t^{(\alpha)} \,, \tag{B.2}$$

$$\sigma^2(m_t) = \sum_{\alpha,\beta=1}^{M} w_\alpha V_{\alpha\beta} w_\beta \,, \tag{B.3}$$

where the weights w and the covariance matrix V are defined as follows:

$$w_\alpha = \sum_{\beta=1}^{M} (V^{-1})_{\alpha\beta} \Big/ \sum_{\gamma,\delta=1}^{M} (V^{-1})_{\gamma\delta}, \quad 1 \le \alpha \le M\,, \quad M = \dim(V), \tag{B.4}$$

$$V_{\alpha\beta} = \delta_{\alpha\beta} \left(\delta m_t^{(\alpha)} \right)^2 + (1 - \delta_{\alpha\beta}) \, \min \left\{ \left(\delta m_t^{(\alpha)} \right)^2, \left(\delta m_t^{(\beta)} \right)^2, C_{\alpha\beta} \delta m_t^{(\alpha)} \delta m_t^{(\beta)} \right\}. \tag{B.5}$$

The latter definition has been adopted in keeping with what has been done in ref. [37], which in turn follows closely the prescriptions of the LEP QCD Working Group [38]. The correlation matrix $C_{\alpha\beta}$, given explicitly below[13] in eq. (B.8), has been computed at one given value of the top mass (173 GeV): we thus neglect effects possibly due to the dependence of such correlations on the top mass, since we expect them to be negligible, especially in the context of eq. (B.5). Given that the correlation between two variables X and Y is defined as

$$C(X,Y) = \frac{\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}{\sigma_X \sigma_Y} = \frac{\langle XY \rangle - \langle X \rangle \langle Y \rangle}{\sigma_X \sigma_Y} \,, \tag{B.6}$$

with σ_X and σ_Y the standard deviations, for any two observables O_r and O_s and their i^{th} and j^{th} moments $\mu_{O_r}^{(i)}$ and $\mu_{O_s}^{(j)}$, we use eq. (B.6) by identifying $X \equiv \mu_{O_r}^{(i)}$ and $Y \equiv \mu_{O_s}^{(j)}$ and proceed similarly to what is done in eq. (A.3); in particular, we have:

$$\langle XY \rangle = \frac{1}{N} \sum_{k=1}^{N} W_k \left(O_r(\mathcal{K}_k) \right)^i \left(O_s(\mathcal{K}_k) \right)^j \,. \tag{B.7}$$

---

[13]In order to facilitate the reading of that matrix, each row and column is labelled with the Mellin moment it corresponds to, in the notation introduced in eq. (2.3).

We also point out that the calculation of $C_{\alpha\beta}$ has been performed by choosing the scale of eq. (3.3), and in the context of an NLO+PS+MS simulations. Although, owing to the form of eq. (B.5), these choices have only a moderate impact on the central values of the combined top masses (as we have verified by setting $C_{\alpha\beta} = 0$), we emphasise again that a more refined procedure will lead exactly to the same conclusions: namely, the necessity of combining the results obtained with different observables and moments, and that of performing a χ^2-type test on the final outcome.

In eq. (B.1) the errors affecting $\overline{m}_t^{(\alpha)}$ are symmetric. In the case when they are asymmetric, the procedure above, and in particular the construction of eqs. (B.4) and (B.5), is repeated twice, for the $+$ and $-$ errors. The two resulting central values for the top mass need not coincide; when this happens, the final central value is taken to be the weighted average of the two, with the weights defined as the inverse of the respective $\sigma^2(m_t)$'s as given in eq. (B.3).

$$
C = \begin{pmatrix}
1 & 0.91 & 0.65 & 0.39 & 0.41 & 0.33 & 0.58 & 0.53 & 0.39 & 0.53 & 0.48 & 0.37 & 0.76 & 0.70 & 0.51 \\
 & 1 & 0.89 & 0.38 & 0.45 & 0.42 & 0.53 & 0.56 & 0.48 & 0.48 & 0.48 & 0.42 & 0.71 & 0.77 & 0.68 \\
 & & 1 & 0.29 & 0.40 & 0.44 & 0.38 & 0.48 & 0.49 & 0.34 & 0.39 & 0.38 & 0.52 & 0.67 & 0.73 \\
 & & & 1 & 0.92 & 0.68 & 0.10 & 0.12 & 0.10 & 0.35 & 0.29 & 0.20 & 0.52 & 0.43 & 0.28 \\
 & & & & 1 & 0.90 & 0.15 & 0.16 & 0.13 & 0.36 & 0.32 & 0.24 & 0.54 & 0.50 & 0.38 \\
 & & & & & 1 & 0.14 & 0.16 & 0.13 & 0.29 & 0.28 & 0.23 & 0.45 & 0.48 & 0.42 \\
 & & & & & & 1 & 0.90 & 0.65 & 0.68 & 0.61 & 0.47 & 0.75 & 0.70 & 0.52 \\
 & & & & & & & 1 & 0.90 & 0.63 & 0.64 & 0.57 & 0.69 & 0.75 & 0.67 \\
 & & & & & & & & 1 & 0.46 & 0.54 & 0.55 & 0.51 & 0.65 & 0.70 \\
 & & & & & & & & & 1 & 0.93 & 0.74 & 0.70 & 0.62 & 0.44 \\
 & & & & & & & & & & 1 & 0.93 & 0.62 & 0.62 & 0.50 \\
 & & & & & & & & & & & 1 & 0.48 & 0.54 & 0.50 \\
 & & & & & & & & & & & & 1 & 0.92 & 0.68 \\
 & & & & & & & & & & & & & 1 & 0.90 \\
 & & & & & & & & & & & & & & 1
\end{pmatrix}
\begin{matrix}
\mu_1^{(1)} \\ \mu_1^{(2)} \\ \mu_1^{(3)} \\ \mu_2^{(1)} \\ \mu_2^{(2)} \\ \mu_2^{(3)} \\ \mu_3^{(1)} \\ \mu_3^{(2)} \\ \mu_3^{(3)} \\ \mu_4^{(1)} \\ \mu_4^{(2)} \\ \mu_4^{(3)} \\ \mu_5^{(1)} \\ \mu_5^{(2)} \\ \mu_5^{(3)}
\end{matrix}
$$

(column labels: $\mu_1^{(1)}\ \mu_1^{(2)}\ \mu_1^{(3)}\ \mu_2^{(1)}\ \mu_2^{(2)}\ \mu_2^{(3)}\ \mu_3^{(1)}\ \mu_3^{(2)}\ \mu_3^{(3)}\ \mu_4^{(1)}\ \mu_4^{(2)}\ \mu_4^{(3)}\ \mu_5^{(1)}\ \mu_5^{(2)}\ \mu_5^{(3)}$)

(B.8)

References

[1] ATLAS, CDF, CMS, D0 collaboration, *First combination of Tevatron and LHC measurements of the top-quark mass*, arXiv:1403.4427 [INSPIRE].

[2] A. Juste, S. Mantry, A. Mitov, A. Penin, P. Skands et al., *Determination of the top quark mass circa 2013: methods, subtleties, perspectives*, arXiv:1310.0799 [INSPIRE].

[3] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Kennedy et al., *The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC*, *Eur. Phys. J.* **C 72** (2012) 2205 [arXiv:1209.2716] [INSPIRE].

[4] F. Bezrukov, M.Y. Kalmykov, B.A. Kniehl and M. Shaposhnikov, *Higgs Boson Mass and New Physics*, *JHEP* **10** (2012) 140 [arXiv:1205.2893] [INSPIRE].

[5] G. Degrassi, S. Di Vita, J. Elias-Miro, J.R. Espinosa, G.F. Giudice et al., *Higgs mass and vacuum stability in the Standard Model at NNLO*, *JHEP* **08** (2012) 098 [arXiv:1205.6497] [INSPIRE].

[6] D. Buttazzo, G. Degrassi, P.P. Giardino, G.F. Giudice, F. Sala et al., *Investigating the near-criticality of the Higgs boson*, *JHEP* **12** (2013) 089 [arXiv:1307.3536] [INSPIRE].

[7] F.L. Bezrukov and M. Shaposhnikov, *The Standard Model Higgs boson as the inflaton*, *Phys. Lett.* **B 659** (2008) 703 [arXiv:0710.3755] [INSPIRE].

[8] A. Kobakhidze and A. Spencer-Smith, *The Higgs vacuum is unstable*, arXiv:1404.4709 [INSPIRE].

[9] CMS Collaboration, *Measurement of the top-quark mass in t t-bar events with lepton+jets final states in pp collisions at $\sqrt{s}$=8 TeV*, CMS-PAS-TOP-14-001 (2014)

[10] D0 collaboration, V.M. Abazov et al., *Precision measurement of the top-quark mass in lepton+jets final states*, *Phys. Rev. Lett.* **113** (2014) 032002 [arXiv:1405.1756] [INSPIRE].

[11] A. Kharchilava, *Top mass determination in leptonic final states with J/ψ*, *Phys. Lett.* **B 476** (2000) 73 [hep-ph/9912320] [INSPIRE].

[12] C.S. Hill, J.R. Incandela and J.M. Lamb, *A Method for measurement of the top quark mass using the mean decay length of b hadrons in t$\bar{t}$ events*, *Phys. Rev.* **D 71** (2005) 054029 [hep-ex/0501043] [INSPIRE].

[13] J.F. Garberson, J. Incandela, S. Koay, R. Rossin and C. Hill, *Measurement of the Top Quark Mass using Quantities with Minimal Dependence on the Jet Energy Scale*, arXiv:0808.0050 [INSPIRE].

[14] S. Biswas, K. Melnikov and M. Schulze, *Next-to-leading order QCD effects and the top quark mass measurements at the LHC*, *JHEP* **08** (2010) 048 [arXiv:1006.0910] [INSPIRE].

[15] R. Frederix and F. Maltoni, *Top pair invariant mass distribution: A Window on new physics*, *JHEP* **01** (2009) 047 [arXiv:0712.2355] [INSPIRE].

[16] S. Kawabata, Y. Shimizu, Y. Sumino and H. Yokoya, *Boost-Invariant Leptonic Observables and Reconstruction of Parent Particle Mass*, *Phys. Lett.* **B 710** (2012) 658 [arXiv:1107.4460] [INSPIRE].

[17] S. Alioli, P. Fernandez, J. Fuster, A. Irles, S.-O. Moch et al., *A new observable to measure the top-quark mass at hadron colliders*, *Eur. Phys. J.* **C 73** (2013) 2438 [arXiv:1303.6415] [INSPIRE].

[18] S. Kawabata, Y. Shimizu, Y. Sumino and H. Yokoya, *Weight function method for precise determination of top quark mass at Large Hadron Collider*, arXiv:1405.2395 [INSPIRE].

[19] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[20] S. Frixione, Z. Kunszt and A. Signer, *Three jet cross-sections to next-to-leading order*, *Nucl. Phys.* **B 467** (1996) 399 [hep-ph/9512328] [INSPIRE].

[21] S. Frixione, *A General approach to jet cross-sections in QCD*, *Nucl. Phys.* **B 507** (1997) 295 [hep-ph/9706545] [INSPIRE].

[22] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029 [hep-ph/0204244] [INSPIRE].

[23] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010 [hep-ph/0011363] [INSPIRE].

[24] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri et al., *HERWIG 6.5 release note*, hep-ph/0210213 [INSPIRE].

[25] S. Frixione, E. Laenen, P. Motylinski and B.R. Webber, *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*, *JHEP* **04** (2007) 081 [hep-ph/0702198] [INSPIRE].

[26] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *JHEP* **03** (2013) 015 [arXiv:1212.3460] [INSPIRE].

[27] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J.* **C 63** (2009) 189 [arXiv:0901.0002] [INSPIRE].

[28] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau et al., *Four-lepton production at hadron colliders: aMC@NLO predictions with theoretical uncertainties*, *JHEP* **02** (2012) 099 [arXiv:1110.4738] [INSPIRE].

[29] M. Cacciari, S. Frixione, M.L. Mangano, P. Nason and G. Ridolfi, *Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC*, *JHEP* **09** (2008) 127 [arXiv:0804.2800] [INSPIRE].

[30] M. Czakon, P. Fiedler and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$*, *Phys. Rev. Lett.* **110** (2013) 252004 [arXiv:1303.6254] [INSPIRE].

[31] M. Czakon, P. Fiedler, A. Mitov and J. Rojo, *Further exploration of top pair hadroproduction at NNLO*, arXiv:1305.3892 [INSPIRE].

[32] M. Cacciari, G.P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[33] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[34] CMS collaboration, *Measurement of masses in the $t\bar{t}$ system by kinematic endpoints in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J.* **C 73** (2013) 2494 [arXiv:1304.5783] [INSPIRE].

[35] K. Agashe, R. Franceschini and D. Kim, *Using Energy Peaks to Measure New Particle Masses*, arXiv:1309.4776 [INSPIRE].

[36] K. Agashe, R. Franceschini and D. Kim, *Simple "invariance" of two-body decay kinematics*, *Phys. Rev.* **D 88** (2013) 057701 [arXiv:1209.0772] [INSPIRE].

[37] R. Frederix, S. Frixione, K. Melnikov and G. Zanderighi, *NLO QCD corrections to five-jet production at LEP and the extraction of $\alpha_s(M_Z)$*, *JHEP* **11** (2010) 050 [arXiv:1008.5313] [INSPIRE].

[38] R.W.L. Jones, *Final α_s combinations from the LEP QCD Working Group*, *Nucl. Phys. Proc. Suppl.* **152** (2006) 15 [INSPIRE].