PHYSICAL REVIEW D 96, 032002 (2017)

Measurement of the top quark mass in the dileptonic $t\bar{t}$ decay channel using the mass observables $M_{b\ell\nu}$, $M_{T2}$, and $M_{b\ell\nu}$ in pp collisions at $\sqrt{s}=8$ TeV

A. M. Sirunyan et al.*
(CMS Collaboration)
(Received 20 April 2017; published 22 August 2017)

A measurement of the top quark mass ($M_t$) in the dileptonic $t\bar{t}$ decay channel is performed using data from proton-proton collisions at a center-of-mass energy of 8 TeV. The data was recorded by the CMS experiment at the LHC and corresponds to an integrated luminosity of $19.7 \pm 0.5$ fb$^{-1}$. Events are selected with two oppositely charged leptons ($\ell = e, \mu$) and two jets identified as originating from $b$ quarks. The analysis is based on three kinematic observables whose distributions are sensitive to the value of $M_t$. An invariant mass observable, $M_{b\ell\nu}$, and a “transverse mass” observable, $M_{T2}$, are employed in a simultaneous fit to determine the value of $M_t$ and an overall jet energy scale factor (JSF). A complementary approach is used to construct an invariant mass observable, $M_{b\ell\nu}$, that is combined with $M_{T2}$ to measure $M_t$. The shapes of the observables, along with their evolutions in $M_t$ and JSF, are modeled by a nonparametric Gaussian process regression technique. The sensitivity of the observables to the value of $M_t$ is investigated using a Fisher information density method. The top quark mass is measured to be $172.22 \pm 0.18$ (stat) $^{+0.82}_{-0.93}$ (syst) GeV.

DOI: 10.1103/PhysRevD.96.032002

I. INTRODUCTION

The top quark mass is a fundamental parameter of the standard model (SM), and an important component in global electroweak fits evaluating the self-consistency of the SM [1]. In addition, the value of $M_t$ has implications for the stability of the SM electroweak vacuum due to the role of the top quark in the quartic term of the Higgs potential [2]. Measurements of $M_t$ have been conducted by the CDF and D0 experiments at the Tevatron, and by the ATLAS and CMS experiments at the CERN LHC. These measurements are typically calibrated against the top quark mass parameter in Monte Carlo (MC) simulation. Studies suggest that this parameter can be related to the top quark mass in a theoretically well-defined scheme with a precision of about 1 GeV [3]. A combination of measurements including all four experiments and $t\bar{t}$ decay channels with zero, one, or two high-$p_T$ electrons or muons (all-hadronic, semileptonic, and dileptonic, respectively) gives a value of $173.34 \pm 0.36$ (stat) $\pm 0.67$ (syst) GeV [4] for the top quark mass. Currently, the most precise experimental determination of $M_t$ is provided by CMS using a combination of measurements in all $t\bar{t}$ decay channels, yielding a value of $172.44 \pm 0.13$ (stat) $\pm 0.47$ (syst) GeV [5]. In the dileptonic $t\bar{t}$ decay channel, the ATLAS [6] and CMS [5] collaborations have recently determined $M_t$ to be $172.99 \pm 0.41$ (stat) $\pm 0.74$ (syst) GeV and $172.82 \pm 0.19$ (stat) $\pm 1.22$ (syst) GeV, respectively. This paper presents a reanalysis of the dileptonic $t\bar{t}$ data set recorded in 2012, with a primary motivation of reducing the systematic uncertainties in $M_t$ determination.

The dileptonic top quark pair ($t\bar{t}$) decay topology, $t\bar{t} \rightarrow (b\ell^+\nu)(b\ell^-\bar{\nu})$, with $\ell = (e, \mu)$, presents a challenge in mass measurement arising primarily from the presence of two neutrinos in the final state. While the undetected $\bar{p}_T$ of a single final-state neutrino in a semileptonic $t\bar{t}$ decay can be inferred from the momentum imbalance in the event, the allocation of momentum imbalance between the two neutrinos in a dileptonic $t\bar{t}$ decay is unknown a priori. For this reason, the dileptonic $t\bar{t}$ system is kinematically underconstrained, and mass determination cannot be easily conducted on an event-by-event basis. Instead, the mass of the parent top quarks in the dileptonic $t\bar{t}$ system can be extracted from kinematic features over an ensemble of events, with the help of appropriate observables and reconstruction techniques.

The measurement reported in this paper is based on a set of observables that have been proposed specifically for mass reconstruction in underconstrained decay topologies. These observables include the invariant mass, $M_{b\ell\nu}$, of a $b\ell$ system, a “transverse mass” variable, $M_{T2}^{bb}$, constructed with the $b$ and $\bar{b}$ daughters of the $t\bar{t}$ system [7–9], and the invariant mass of a $b\ell\nu$ system, $M_{b\ell\nu}$, where the neutrino momentum is estimated by the $M_{T2}$-assisted on-shell (MAOS) reconstruction technique [10]. The MAOS reconstruction technique builds on $M_{T2}$, constructed with the $b$ and $\bar{b}$ daughters of the $t\bar{t}$ system [7–9], and the invariant mass of a $b\ell\nu$ system, $M_{b\ell\nu}$, where the neutrino momentum is estimated by the $M_{T2}$-assisted on-shell (MAOS) reconstruction technique [10]. The sensitivity of the $M_{b\ell\nu}$, $M_{T2}^{bb}$, and $M_{b\ell\nu}$

*Full author list given at the end of the article.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.
observables to the value of $M_t$ is investigated using a Fisher information density method. Distributions of $M_{b\ell}$ and $M_{T_2}$ in dileptonic events contain a sharp edge descending to a kinematic end point, the location of which is sensitive to the value of $M_t$. Recently, masses of the top quark, W boson ($M_W$), and neutrino ($M_\nu$) were extracted in a simultaneous fit using the end points of these distributions in dileptonic $t\bar{t}$ events [11]. The $M_{b\ell}$, $M_{T_2}$, and MAOS $M_{b\nu}$ observables are described in more detail in Sec. IV.

One of the dominant sources of systematic uncertainty limiting the precision of this measurement comes from the overall uncertainty in jet energy scale (JES). To address the JES uncertainty, we introduce a technique that uses the $M_{b\ell}$ and $M_{T_2}$ observables to determine an overall jet energy scale factor (JSF) simultaneously with the top quark mass, $M_t$, and $\Delta M_{bb}$, where the jet pair originating from a W boson decay is used to determine the JSF [5]. Because light-quark jets from the W boson decay are used to calibrate the energy scale of $b$ jets arising from the $t$ and $\bar{t}$ decays, these methods are sensitive to flavor-dependent uncertainties that emerge from differences in the response of $b$ jets and light-quark jets. In the method featured here, the JSF is determined in the dileptonic $t\bar{t}$ channel without relying on a $W$ boson decaying to jets. Instead, it achieves sensitivity to the JSF through the kinematic differences between $b$ jets, which are subject to JSF scaling, and leptons, which are not. Because it does not use light quarks from a hadronic W boson decay, this approach is insensitive to flavor-dependent JES uncertainties.

To model the $M_{b\ell}$, $M_{T_2}$, and MAOS $M_{b\nu}$ distribution shapes, we use a Gaussian process (GP) regression technique [12,13]. This technique is nonparametric, and thus largely model independent. It is effective in modeling distribution shapes when no theoretical guidance is available to specify a functional form. The distribution shapes can conveniently be modeled as functions of multiple variables. In this analysis, three variables are used: the value of the relevant observable ($M_{b\ell}$, $M_{T_2}$, or $M_{b\nu}$), $M_t$, and the JSF. The shapes are determined using simulated events generated with seven different values of $M_t$ ranging from 166.5 to 178.5 GeV, and with five values of JSF, ranging from 0.97 to 1.03, applied to the jets in each event. Each shape ultimately models the distributions of the observables together with their evolution in $M_t$ and in JSF.

II. THE CMS DETECTOR

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two end cap sections. The tracker has a track-finding efficiency of more than 99% for muons with transverse momentum $p_T > 1$ GeV and pseudorapidity $|\eta| < 2.4$. The ECAL is a fine-grained hermetic calorimeter with quasiprojective geometry, and is distributed in the barrel region of $|\eta| < 1.48$ and in two end caps that extend up to $|\eta| < 3.0$. The HCAL barrel and end caps similarly cover the region $|\eta| < 3.0$. In addition to the barrel and end cap detectors, CMS has extensive forward calorimetry. Muons are measured in gas-ionization detectors, which are embedded in the steel flux-return yoke outside of the solenoid. The silicon tracker and muon systems play a crucial role in the identification of jets originating from the hadronization of $b$ quarks [14]. Events of interest are selected using a two-tiered trigger system [15]. The first level, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a time interval of less than 4 $\mu$s. The second level, known as the high-level trigger, consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to less than 1 kHz before data storage. A more detailed description of the CMS detector, together with a definition of the coordinate system used, can be found in Ref. [16].

III. DATA SETS AND EVENT SELECTION

We select dileptonic $t\bar{t}$ events from a data set recorded at $\sqrt{s} = 8$ TeV during 2012 corresponding to an integrated luminosity of $19.7 \pm 0.5$ fb$^{-1}$ [17]. Events are required to pass one of several triggers that require at least two leptons, $ee$, $e\mu$, or $\mu\mu$, where the leading (higher-$p_T$) lepton satisfies $p_T > 17$ GeV and the subleading lepton satisfies $p_T > 8$ GeV.

A particle-flow (PF) algorithm [18,19] is used to reconstruct and identify each individual particle in an event by combining information from various subdetectors of CMS. Each event is required to have at least one reconstructed collision vertex, with the primary vertex selected as the one containing the largest $\sum p_T^2$ of associated tracks. Electron candidates are reconstructed by matching a cluster of energy deposits in the ECAL to a reconstructed track [20]. They are required to satisfy $p_T > 20$ GeV and $|\eta| < 2.5$. Muon candidates are reconstructed in a global fit that combines information from the silicon tracker and muon system [21], and must have $p_T > 20$ GeV and $|\eta| < 2.4$. A requirement on the relative isolation is imposed inside a cone $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$ around each lepton candidate, where $\phi$ is the azimuthal angle in radians. A parameter $I_{rel} = \sum p_{Ti}/p_T^2$ is defined, where the sum includes all reconstructed PF candidates inside the cone (excluding the lepton itself), and $p_T^i$ is the lepton $p_T$. Electron (muon) candidates are required to have $I_{rel} < 0.15$.
(0.2) with $\Delta R < 0.3$ (0.4). Events selected off-line are required to contain exactly two such leptons, $e^+e^-$, $e\mu$, or $\mu\mu$, with opposite charge. For events containing an $e^+e^-$ or $\mu^+\mu^-$ pair, contributions from low-mass resonances are suppressed by requiring an invariant mass of the lepton pair $M_{ll^\prime} > 20$ GeV, while contributions from $Z$ boson decays are suppressed by requiring that $|M_Z - M_{ll^\prime}| > 15$ GeV, where $M_Z = 91.2$ GeV [22].

Hadronic jets are clustered from PF candidates with the infrared and collinear safe anti-$k_T$ algorithm [23], with a distance parameter $R$ of 0.5, as implemented in the FASTJET package [24]. The jet momentum is determined as the vectorial sum of all particle momenta in this jet. Corrections to the JES and jet energy resolution (JER) are derived using MC simulation, and are confirmed with measurements of the energy balance in quantum chromodynamics (QCD) dijet, QCD multijet, photon + jet, and $Z + \text{jet}$ events [25]. Muons, electrons, and charged hadrons originating from multiple collisions within the same or nearby bunch crossings (pileup), are not included in the jet reconstruction. Contributions from neutral hadrons originating from pileup are estimated and subtracted from the reconstruction. Contributions from neutral hadrons originating from pileup are estimated and subtracted from the simulation. The jet reconstruction is estimated in simulation to be 95% dileptonic $b\bar{b}$ jets.

Dileptonic $b\bar{b}$ jets production are generated with $\text{MADGRAPH}$ 5.1.3.30, and contributions from $WW$, $WZ$, $ZZ$ processes are simulated with $\text{PYTHIA}$. The CMS detector response to the simulated events is modeled with $\text{GEANT4}$ [36]. All background processes are normalized to their predicted cross sections [37–41].

With the requirements outlined previously, 41640 $t\bar{t}$ candidate events are selected in data. The sample composition is estimated in simulation to be 95% dileptonic $t\bar{t}$, 4% single top quark, and 1% other processes including diboson, $W + \text{jet}$, and Drell-Yan production, as well as semileptonic and all-hadronic $t\bar{t}$.

IV. OBSERVABLES

The observables featured in this study have been developed for physics scenarios where undetected particles, such as neutrinos, carry away a portion of the kinematic information necessary for full event reconstruction. In the dileptonic $t\bar{t}$ system, distributions in these observables contain end points, edges, and peak regions that are sensitive to the top quark mass. The observables are described in more detail below.

A. The $M_{ll^\prime}$ observable

The $M_{ll^\prime}$ observable is defined as

$$M_{ll^\prime} = \sqrt{(p_T^b + p_T^l)^2},$$

(1)

where $p_T^b$ and $p_T^l$ are four-vectors corresponding to a $b$ jet and lepton, respectively. The $b\ell$ pairs underlying each value of $M_{ll^\prime}$ are chosen out of four possible combinations by an algorithm described in Sec. IVA 1. The $M_{ll^\prime}$ observable contains a kinematic end point that occurs when the $b$ jet and lepton are directly back to back in the top quark rest frame. The location of this end point, $(M_{ll^\prime})_{\text{max}}$, is a function of the masses involved in the decay:

$$M_{ll^\prime})_{\text{max}} = \frac{\sqrt{(M_W^2 - M_W^2)(M_Z^2 - M_Z^2)}}{M_W^2}. $$

(2)

With $M_t = 172.5$ GeV, $M_W = 80.4$ GeV [22], and $M_\nu = 0$, we have $(M_{ll^\prime})_{\text{max}} = 152.6$ GeV. Although this end point is a theoretical maximum on the value of $M_{ll^\prime}$ at leading order, events are still observed beyond this value due to background contamination, resolution effects, and nonzero particle widths.

The $M_{ll^\prime}$ distribution is shown in data and MC simulation in Fig. 1 (upper), with a breakdown of signal and background events shown in the simulation. The “signal” category includes $t\bar{t}$ dilepton decays where both $b$ jets are correctly identified by the $b$ tagging algorithm. The background categories include: “mistag” dilepton decays where a light quark or gluon jet is incorrectly selected by the $b$...
the 8 TeV data set corresponding to an integrated luminosity of

GP regression technique, described in Sec.VI. tagging algorithm; “τ decays” where dilepton events include at least one τ lepton in the final state subsequently decaying leptonically; and “hadronic decays” that include events where at least one of the top quarks decays hadronically. The “non-τ bkg” category consists of single top quark, diboson, W + jets, and Drell-Yan processes. Events in which a top quark decays through a τ lepton contain extra neutrinos stemming from the leptonic τ decay. Although the extra neutrinos cause a small distortion to the kinematic distributions, these events still contribute to the sensitivity of the measurement.

The sensitivity of the $M_{b\tau}$ observable to the value of $M_t$ is demonstrated in Fig. 1 (lower), where $M_{b\tau}$ shapes corresponding to three values of the top quark mass in MC simulation ($M_t^{MC}$) are shown. The variation between these shapes reveals regions of the $M_{b\tau}$ distribution that are sensitive to the value of $M_t$, such as the edges to the left and right of the $M_{b\tau}$ peak, and regions that are not sensitive, such as the stationary point where the three shapes intersect. To provide a quantitative description of these effects, we introduce a “local shape sensitivity” function, also known as the Fisher information density, shown in Figs. 1, 3, and 4. This function conveys the sensitivity of an observable at a specific point on its shape. For the $M_{b\tau}$ observable, the local shape sensitivity function peaks near the kinematic end point ($M_{b\tau} \sim 150$ GeV), and has a zero value at the stationary point ($M_{b\tau} \sim 105$ GeV). The integral of this function over its range is proportional to $1/\sigma_{M_t}$, where $\sigma_{M_t}$ is the statistical uncertainty on a measurement of $M_t$. A full description of the local shape sensitivity function is given in Appendix A.

1. $b$ jet and lepton combinatorics

The two $b$ jets and two leptons stemming from each $t\bar{t}$ decay give rise to a twofold matching ambiguity, with two correct and two incorrect $b\tau$ pairings possible in each event. Pairings in which the $b$ jet and lepton emerge from different top quarks do not necessarily obey the upper bound described in Eq. (2), and thus do not have a clean kinematic end point in $M_{b\tau}$. Although $a$ priori it is experimentally difficult to distinguish between correct and incorrect pairings, one possible approach is to select the smallest two $M_{b\tau}$ values in each event. This way, the kinematic end point of the distribution is preserved—even if the smallest two $M_{b\tau}$ values do not correspond to the correct pairings, they are guaranteed to fall below the correct pairings, which do respect the end point. In this analysis, we employ a slightly more sophisticated matching technique, introduced in Ref. [11], where either two or three $b\tau$ pairs are selected in each event. By selecting either two or three $b\tau$ pairs in each event, the technique employed in this analysis has the benefit of increased statistical power, while preserving the kinematic end point of $M_{b\tau}$. Although they are not necessarily the correct pairs, the corresponding $M_{b\tau}$ values are guaranteed by construction to be less than or equal to those of the correct pairs. The matching technique is based on the following prescription:

1. match each $b$ jet with the lepton that produces the lower $M_{b\tau}$ value;

FIG. 1. (Upper) The $M_{b\tau}$ distribution in data and simulation with $M_t^{MC} = 172.5$ GeV, normalized to the number of events in the 8 TeV data set corresponding to an integrated luminosity of $19.7 \pm 0.5$ fb$^{-1}$. The lower panel shows the ratio between the data and simulation. Statistical and systematic uncertainties on the distribution in simulation are represented by the shaded area. A description of the systematic uncertainties is given in Sec. VIII.

(Lower) The $M_{b\tau}$ distribution shapes in simulation, normalized to unit area, corresponding to three values of $M_t^{MC}$ are shown together with the local shape sensitivity function, described in Appendix A. The $M_{b\tau}$ distributions include two or three values of $M_{b\tau}$ for each event. The distribution shapes are modeled with a GP regression technique, described in Sec. VI.

$M_{b\tau}$ distributions include at least one $\tau$ lepton in the final state subsequently decaying leptonically; and “hadronic decays” that include events where at least one of the top quarks decays hadronically. The “non-\tau bkg” category consists of single top quark, diboson, $W +$ jets, and Drell-Yan processes. Events in which a top quark decays through a $\tau$ lepton contain extra neutrinos stemming from the leptonic $\tau$ decay. Although the extra neutrinos cause a small distortion to the kinematic distributions, these events still contribute to the sensitivity of the measurement.

The sensitivity of the $M_{b\tau}$ observable to the value of $M_t$ is demonstrated in Fig. 1 (lower), where $M_{b\tau}$ shapes corresponding to three values of the top quark mass in MC simulation ($M_t^{MC}$) are shown. The variation between these shapes reveals regions of the $M_{b\tau}$ distribution that are sensitive to the value of $M_t$, such as the edges to the left and right of the $M_{b\tau}$ peak, and regions that are not sensitive, such as the stationary point where the three shapes intersect. To provide a quantitative description of these effects, we introduce a “local shape sensitivity” function, also known as the Fisher information density, shown in Figs. 1, 3, and 4. This function conveys the sensitivity of an observable at a specific point on its shape. For the $M_{b\tau}$ observable, the local shape sensitivity function peaks near the kinematic end point ($M_{b\tau} \sim 150$ GeV), and has a zero value at the stationary point ($M_{b\tau} \sim 105$ GeV). The integral of this function over its range is proportional to $1/\sigma_{M_t}$, where $\sigma_{M_t}$ is the statistical uncertainty on a measurement of $M_t$. A full description of the local shape sensitivity function is given in Appendix A.

1. $b$ jet and lepton combinatorics

The two $b$ jets and two leptons stemming from each $t\bar{t}$ decay give rise to a twofold matching ambiguity, with two correct and two incorrect $b\tau$ pairings possible in each event. Pairings in which the $b$ jet and lepton emerge from different top quarks do not necessarily obey the upper bound described in Eq. (2), and thus do not have a clean kinematic end point in $M_{b\tau}$. Although $a$ priori it is experimentally difficult to distinguish between correct and incorrect pairings, one possible approach is to select the smallest two $M_{b\tau}$ values in each event. This way, the kinematic end point of the distribution is preserved—even if the smallest two $M_{b\tau}$ values do not correspond to the correct pairings, they are guaranteed to fall below the correct pairings, which do respect the end point. In this analysis, we employ a slightly more sophisticated matching technique, introduced in Ref. [11], where either two or three $b\tau$ pairs are selected in each event. By selecting either two or three $b\tau$ pairs in each event, the technique employed in this analysis has the benefit of increased statistical power, while preserving the kinematic end point of $M_{b\tau}$. Although they are not necessarily the correct pairs, the corresponding $M_{b\tau}$ values are guaranteed by construction to be less than or equal to those of the correct pairs. The matching technique is based on the following prescription:

1. match each $b$ jet with the lepton that produces the lower $M_{b\tau}$ value;
The result is an event topology with two identical branches, invisible (\(t\) in the dileptonic system). Here, the invisible particle (see the final values of \(\vec{p}_T^2\) and \(\vec{p}_T^3\)) are typically treated as intermediate quantities in the \(M_{T_2}\) algorithm, they are employed as neutrino \(\vec{p}_T\) estimates in the MAOS reconstruction technique described in Sec. IV C.

1. The \(M_{T_2}\) subsystems

In the \(t\bar{t}\) system, there are several ways in which \(M_{T_2}\) can be computed, depending on how the decay products are grouped together. The \(M_{T_2}\) algorithm classifies them into three categories: upstream, visible, and child particles [42]. The child particles are those at the end of the decay chain that are unobservable or simply treated as unobservable. In the latter case, the child particle momenta are added to the \(\vec{p}_T^{\text{miss}}\) vector. The visible particles are those whose \(\vec{p}_T\) values are measured and used in the calculations; and the upstream particles are those from further up in the decay chain, including any initial-state radiation (ISR) accompanying the hard collision.

In general, the child, visible, and upstream particles may actually be collections of objects, creating three possible subsystems in the dileptonic \(t\bar{t}\) event topology. These subsystems are illustrated in Fig. 2. For simplicity, we refer to the corresponding \(M_{T_2}\) observables as \(M_{T_2}^{bb}, M_{T_2}^{\ell\ell}, \) and \(M_{T_2}^{\ell \bar{\ell}}\), where

(i) The \(M_{T_2}^{bb}\) observable uses the two leptons as visible particles, treating the neutrinos as invisible child particles, and combining the \(b\) jets with all other upstream particles in the event.

(ii) The \(M_{T_2}^{\ell\ell}\) observable uses the \(b\) jets as visible particles, and treats the \(W\) bosons as child particles, ignoring the fact that their charged daughter leptons are indeed observable. It considers only ISR jets as generators of upstream momentum.

(iii) The \(M_{T_2}^{\ell \bar{\ell}}\) observable combines the \(b\) jet and the lepton to form a single visible system, and takes the neutrinos as the invisible particles. A twofold matching ambiguity results from the matching of \(b\) jets to leptons in each event. In order to preserve the kinematic end point of the \(M_{T_2}^{\ell \bar{\ell}}\) distribution, the \(b\ell\) configuration highlights the difference between this recipe and the simpler approach of choosing the smallest two values of \(M_{T_2}\), which do not necessarily incorporate both \(b\) jets and both leptons in the event. For example, this could occur if both \(b\) jets are matched to a single lepton. In these cases, the next largest \(M_{T_2}\) value is also needed to ensure both \(b\) jets and both leptons from the event are used.

B. The \(M_{T_2}\) observable

The \(M_{T_2}\) transverse mass observable [7,8] is based on the transverse mass, \(M_T\). The transverse mass of the \(W\) boson in a \(W \to \ell \nu\) decay is given by

\[
M_T = \sqrt{m_T^2 + m_T^2 + 2(E_T\ell E_T\nu - \vec{p}_T\ell \cdot \vec{p}_T\nu)},
\]

where \(E_T\ell = m_T^2 + \vec{p}_T^2\) for \(x \in \{\ell, \nu\}\), \(m_x\) is the particle mass, and \(\vec{p}_T\ell\) is the particle momentum projected onto the plane perpendicular to the beams. This quantity exhibits a kinematic end point at the parent mass, \(M_w\), which occurs in configurations when both the lepton and neutrino momenta lie entirely in the transverse plane (up to a common longitudinal boost).

The dileptonic \(t\bar{t}\) system has two layers of decays, with \(t \to Wb\) in the first step followed by \(W \to \ell \nu\) in the second. The result is an event topology with two identical branches, \(t \to b\ell^+\nu\) and \(\bar{t} \to \bar{b}\ell^-\bar{\nu}\), each with a visible (\(b\ell\)) and invisible (\(\nu\)) component. In this case, one value of \(M_T\) can be computed for each branch. The invisible particle momentum associated with each branch, however, is not known. While for a semileptonic \(t\bar{t}\) decay, with only one \(W \to \ell \nu\) decay, the neutrino \(\vec{p}_T\ell\) is estimated from the \(\vec{p}_T^{\text{miss}}\) in the event, a dileptonic \(t\bar{t}\) decay includes two neutrinos, for which the allocation of \(\vec{p}_T^{\text{miss}}\) between them is unknown.

The \(M_{T_2}\) observable is an extension of \(M_T\) for a system with two identical decay branches, “a” and “b,” as those in the dileptonic \(t\bar{t}\) system. Here, the invisible particle momenta, \(\vec{p}_T^a\) and \(\vec{p}_T^b\), must add up to the total \(\vec{p}_T^{\text{miss}}\). The strategy of \(M_{T_2}\) is to impose this constraint on the invisible particle momenta, while also performing a minimization in order to preserve the kinematic end point of \(M_T\). For a general event with a symmetric decay topology, \(M_{T_2}\) is defined as

\[
M_{T_2} = \min_{\vec{p}_T^a + \vec{p}_T^b = \vec{p}_T^{\text{miss}}} \max \{M_T^a, M_T^b\},
\]

where \(M_T^a\) and \(M_T^b\) correspond to the two decay branches. If the invisible particle mass is known, it can be incorporated into the \(M_{T_2}\) calculation as well, yielding an end point at the parent particle mass.
pair with the smallest value of $M_{T_2}^{M_{1\ell}}$ is used in each event. These observables are identical, respectively, to $M_{T_2}^{(2.2,1)}$, $M_{T_2}^{(2.1,0)}$, and $M_{T_2}^{(2.2,0)}$ of Ref. [42], and $\mu_{h\ell\ell}, \mu_{l\ell\ell}, \mu_{o\ell\ell}$ of Ref. [11].

The subsystem observable $M_{T_2}^{M_{b\ell}}$ is employed in this study to complement the observable $M_{T_1\ell}$. The $M_{T_2}^{M_{b\ell}}$ observable contains an end point at the value of $M_{T_1}$, and can be combined with $M_{T_1\ell}$ to mitigate uncertainties due to the JES. This feature is discussed further in Sec. V. The distribution of $M_{T_2}^{M_{b\ell}}$ and its sensitivity to the value of $M_{T_1}$ are shown in Fig. 3. Although $M_{T_2}^{M_{1\ell}}$ is not directly sensitive to $M_{T_1}$, the neutrino $\vec{p}_T$ estimates that are a by-product of its computation are used as an input into the MAOS $M_{b\ell\ell}$ reconstruction technique described in Sec. IV C.

The $M_{T_2}^{M_{b\ell}}$ distribution employed in this analysis includes a kinematic requirement on the upstream momentum, defined as $\vec{p}_T^{\text{upst}} = \sum_{\text{jets}} \vec{p}_T - \sum_{\ell} \vec{p}_T - \sum_{\text{leptons}} \vec{p}_T$, where the sums are conducted over all reconstructed PF candidates, $b$ jets, and leptons in each event, respectively. The direction of $\vec{p}_T^{\text{upst}}$ is required to lie outside the opening angle between the two $b$ jet $\vec{p}_T$ vectors in the event. This requirement primarily impacts events at low values of $M_{T_2}^{M_{b\ell}}$, and its effect on the statistical sensitivity of the observable is small.

### C. The MAOS $M_{b\ell\ell}$ observable

The MAOS reconstruction technique employed in this analysis is based on the subsystem observable $M_{T_2}^{M_{1\ell}}$. In the $M_{T_2}^{M_{1\ell}}$ algorithm, an $M_{T_1}$ variable, defined in Eq. (3), is constructed from the $\ell^+\nu$ and $\ell^-\bar{\nu}$ pairs corresponding to each of the $t\bar{t}$ decay branches. Because the values of neutrino $\vec{p}_T$ are unknown, a minimization is conducted in Eq. (4) over possible values consistent with the measured $\vec{p}_T^{\text{miss}}$ in each event.

The MAOS technique employs the neutrino $\vec{p}_T$ values that are determined by the $M_{T_2}^{M_{1\ell}}$ minimization to construct full $b\ell\nu$ invariant mass estimates corresponding to each of the $t\bar{t}$ decay branches. Given the neutrino $\vec{p}_T$ values, the remaining $z$-components of their momenta are obtained by enforcing the W mass on-shell requirement [22]

$$M(\ell^+\nu) = M(\ell^-\bar{\nu}) = M_W = 80.4 \text{ GeV}.$$  

This yields a longitudinal momentum for each neutrino given by

$$p_{z\ell} = \frac{1}{E_{T\ell}^{\ell\nu}} \left[ p_{z\ell} A \pm \sqrt{p_{z\ell}^2 + E_{T\ell}^{\ell\nu} \sqrt{A^2 - (E_{T\ell}^{\ell\nu})^2}} \right],$$

where $A = \frac{1}{2}(M_W^2 + M_{\ell\nu}^2 + M_{T_1}^2) + \vec{p}_T \cdot \vec{p}_{T\ell}$ [10]. Given these estimates for the neutrino three-momenta together with $M_{\ell\nu} = 0$, we have the required four vectors to construct an $M_{b\ell\ell}$ invariant mass corresponding to the decay products of each top quark.

The quadratic equations in Eq. (6) underlying the W mass on-shell requirement provide up to two solutions for each value of $p_{z\ell}$, yielding a twofold ambiguity for each neutrino momentum. In addition, there is a twofold ambiguity resulting from the matching of $b$ jets to $\ell\nu$ pairs in the construction of $b\ell\nu$ invariant masses. No matching ambiguity exists between leptons and neutrinos, since the $\ell^+\nu$ and $\ell^-\bar{\nu}$ pairs have been fixed by the $M_{T_2}^{M_{1\ell}}$ algorithm. The combined fourfold ambiguity, along with the two top quark decays in each event, gives up to eight possible values of $M_{b\ell\ell}$. In the measurement, all of the available values are used: for each $\ell\nu$ pair, this includes up to two
neutralino $\nu$ solutions, and two $b\ell\nu$ matches. The distribution of MAOS $M_{bl\nu}$ and its sensitivity to the value of $M_t$ are shown in Fig. 4.

V. SIMULTANEOUS DETERMINATION OF $M_t$ AND JSF

To mitigate the impact of JES uncertainties on the precision of this measurement, we introduce a technique that allows a JSF parameter to be fit simultaneously with $M_t$. The JSF is a constant multiplicative factor that calibrates the overall energy scale of reconstructed jets. It is applied in addition to the standard JES calibration, which corrects the jet response as a function of $p_T$ and $\eta$. The dominant component of uncertainty in the JES calibration can be attributed to a global factor in jet response, which is captured in the JSF.

The challenge in determining the JSF simultaneously with $M_t$ stems from the large degree of correlation between these parameters. In the top quark decay, $t \to b\ell\nu$, the JSF directly affects the momentum of the $b$ jet, and indirectly, the inferred momentum of the neutrino, by scaling all jets entering the $p_T^{\text{miss}}$ sum. The $M_t$ parameter affects the momenta of these two particles in addition to the lepton produced in the top quark decay. In the context of observables and distribution shapes, variations in the $M_t$ and JSF parameters cause shape changes that are difficult to distinguish. For this reason, a shape-based analysis using a single observable can be implemented to determine either $M_t$ or JSF, but not both simultaneously.

FIG. 4. Following the conventions of Fig. 1, shown are the (upper) MAOS $M_{bl\nu}$ distribution in data and simulation with $M_t^{MC} = 172.5$ GeV, and (lower) the MAOS $M_{bl\nu}$ distribution shapes in simulation corresponding to three values of $M_t^{MC}$, along with the local shape sensitivity function. The MAOS $M_{bl\nu}$ distributions include up to eight values of $M_{bl\nu}$ for each event.

FIG. 5. The (upper) $M_{bl\nu}$ and (lower) $M_{bbT}^{T_2}$ distributions in simulation with $M_t = 172.5$ GeV for several values of JSF. Two or three values are included in the $M_{bl\nu}$ distribution for each event, and one value is included in the $M_{bbT}^{T_2}$ distribution if it satisfies the kinematic requirement outlined in Sec. IV B 1. The distributions are normalized to unit area. The three curves corresponding to each of the $M_{bl\nu}$ and $M_{bbT}^{T_2}$ distributions are obtained using a GP regression technique described in Sec. VI.
To determine the $M_t$ and JSF parameters simultaneously, we construct a likelihood function that contains two distributions corresponding to the $M_{b\ell}$ and $M_{bb}$ observables. In this configuration, variations in the parameters produce shifts in each individual distribution. They also create a relative shift between the distributions that provides the additional constraint needed for a simultaneous fit of $M_t$ and JSF. The dependence of the $M_{b\ell}$ and $M_{bb}$ distribution shapes on $M_t$ is shown in Figs. 1 and 3, and their dependence on the JSF is shown in Fig. 5. The difference in response between the $M_{b\ell}$ and $M_{bb}$ shapes to the JSF parameter is rootied in the reconstructed objects underlying the $M_{b\ell}$ and $M_{bb}$ observables—where each value of $M_{b\ell}$ uses one $b$ jet and one lepton, each value of $M_{bb}$ uses two $b$ jets and no leptons for the visible system. Thus, $M_{bb}$ exhibits a stronger dependence on the JSF. The likelihood fit used in this measurement is described in more detail in Sec. VII.

VI. GAUSSIAN PROCESSES FOR SHAPE ESTIMATION

In this analysis, the $M_{b\ell}$, $M_{bb}$, and $M_{b\ell}\nu$ distribution shapes are modeled with a GP regression technique that has two main advantages over other commonly used shape estimate methods. First, the GP shape is nonparametric, determined only by a set of training points and hyperparameters that regulate smoothing; and second, it can be easily trained as a function of several variables simultaneously. The latter feature allows one to capture the smooth evolution of the distribution shapes as the $M_t$ and JSF parameters are varied. A detailed introduction to GPs can be found in Refs. [12,13]. Here, we give a brief overview of the GP regression technique, with further discussion provided in Appendix B.

The likelihood fit described in Sec. VII uses distribution shapes of the form $f(x|M_t,\text{JSF})$, where $x$ is the value of an observable ($M_{b\ell}$, $M_{bb}$, or $M_{b\ell}\nu$), and $M_t$ and JSF are free parameters in the fit. The shapes $f$ are shown in Figs. 1, 3, and 4 for each observable, where the free parameters are set to $M_t = 166.5, 172.5$, or $178.5$ GeV and JSF = 1. In Fig. 5, shapes corresponding to the $M_{b\ell}$ and $M_{bb}$ observables are shown with the free parameters set to $M_{b\ell} = 172.5$ GeV and JSF = 0.97, 1.00, or 1.03. In the figures, these shapes are represented as functions of a single variable (the observable $x$) with $M_t$ and JSF fixed. In GP regression, however, each shape is treated as a function of all three quantities ($x, M_t$, and JSF), and can be described as a probability density in three dimensions.

Each GP shape is trained using binned distributions of the observable $x$ in MC simulation. For each observable, 35 binned distributions are used, corresponding to seven values of $M_{b\ell}^{\text{MC}}$ ranging from 166.5 to 178.5 GeV and five values of JSF ranging from 0.97 to 1.03. Each distribution has 75 bins in $x$, yielding a total of 2625 training points at which the value of $f$ is known and used as an input into the GP regression process. Each training point is specified by its values of $x$, $M_t$, and JSF. The GP regression technique interpolates between the discrete values of $x$, $M_t$, and JSF covered by these training points to provide a shape that is smooth over its range. The smoothness properties of each shape are determined by a kernel function that is set by the analyzer. The GP shapes in this analysis correspond to the kernel function given in Eq. (B5) of Appendix B.

The binned distributions used to construct each GP shape are normalized to unity. However, the normalization of the GP shape itself may deviate slightly from unity due to minor imperfections in shape modeling. To mitigate this effect, the GP shape normalization is recomputed for each value of $M_t$ and JSF at which the shape is evaluated. In a likelihood fit, the normalization is recomputed for every variation of the fit parameters.

VII. FIT STRATEGY

This measurement employs an unbinned maximum-likelihood fit using the $M_{b\ell}$, $M_{bb}$, and MAOS $M_{b\ell}\nu$ observables described in Sec. IV, along with the GP shape estimate technique described in Sec. VI. The MC samples used to train the GP shapes include the $t\bar{t}$ signal and background processes described in Sec. III.

The likelihood constructed from a single observable, $x$, is given by

$$\mathcal{L}(M_t, \text{JSF}) = \prod_i f(x_i|M_t, \text{JSF}).$$

Here, the distribution shape $f$ depends on the value of the free parameters $M_t$ and JSF, and expresses the likelihood of drawing some event $i$ where the value of the observable is $x_i$. It is normalized to unity over its range for all values of $M_t$ and JSF. The parameters $M_t$ and JSF are varied in the fit to maximize the value of the likelihood.

A likelihood containing two observables, $x_1$ and $x_2$, is constructed as a product of individual likelihoods:

$$\mathcal{L}(M_t, \text{JSF}) = \mathcal{L}(M_t, \text{JSF})\mathcal{L}(M_t, \text{JSF})$$

$$= \prod_i f(x_{1i}|M_t, \text{JSF})f(x_{2i}|M_t, \text{JSF}).$$

This analysis employs three different versions of the likelihood fit:

1. the 1D fit uses the $M_{b\ell}$ and $M_{bb}$ observables to determine $M_t$, and JSF is constrained to be unity;
2. the 2D fit also uses $M_{b\ell}$ and $M_{bb}$ but imposes no constraint on the JSF and determines $M_t$ and JSF simultaneously;
3. the MAOS fit uses the $M_{bb}$ and $M_{b\ell}\nu$ observables to determine $M_t$, and JSF is constrained to be unity.

Among these versions, the 1D fit provides the best precision on the value of $M_t$. The 2D fit mitigates the JES uncertainties, which are the largest source of systematic error in the 1D approach. The MAOS fit is expected to
FIG. 6. Likelihood fit results as a function of $M_{t}^{MC}$ corresponding to the (top) 2D, (center left) 1D, (center right) MAOS, and (bottom) hybrid fits. For each value of $M_{t}^{MC}$, the fit is conducted using 50 pseudo-experiments in MC simulation. The mean parameter values, $M_{t}^{fit}$ and JSF$^{fit}$, are represented by the points, with statistical uncertainties indicated by the error bars. A best-fit line of the form $y = ax + b$ is shown for each fit configuration.
yield results similar to the 1D fit, and is presented as a viable alternative that substitutes the $M_{he}$ observable for MAOS $M_{he}^{\text{MC}}$. The best overall precision on $M_t$ is given by a combination of the 1D and 2D fits, which is discussed in Sec. VII A. The fit results are discussed in Sec. IX.

The central value and statistical uncertainty on $M_t$ and JSF are determined using the bootstrapping technique [43]. This method is based on pseudo-experiments rather than the shape of the total likelihood defined in Eq. (8) near its maximum, and thus mitigates the effects of correlation between the two observables, $x_1$ and $x_2$, in the likelihood. The technique also mitigates possible correlations within the $M_{he}$ and $M_{he}^{\text{MC}}$ observables when multiple values of the observable occur in a single event. The bootstrapping technique is primarily relevant for statistical uncertainty determination, which may otherwise be affected by correlations in the likelihood. The technique has a negligible impact on the central values of $M_t$ and JSF. The bootstrap pseudo-experiments are constructed by resampling the full data set with replacement, where the size of each pseudo-experiment is fixed to have the number of events in data (41640 events). Events are selected at random from the full data set, so that a particular event has the same probability of being chosen at any stage during the sampling process. In this procedure, a single event may be selected more than once for any given pseudo-experiment. In data, all events have an equal probability to be selected. In simulation, the probability of selecting a particular event is proportional to its weight, containing the relevant cross sections, as well as corrections for MC modeling and object reconstruction efficiencies.

The performance of the likelihood fitting approach described above is evaluated using events in simulation, where the true values of $M_t$ and JSF are known. The fit is conducted using seven different values of $M_t^{\text{MC}}$ ranging from 166.5 to 178.5 GeV for each version of the likelihood fit. The results of this performance study are shown in Fig. 6. The likelihood fits are consistent with zero bias, showing that the GP shape modeling technique accurately captures the distribution shapes and their evolution over several values of $M_t^{\text{MC}}$. For this reason, no calibration of the fit is necessary for an unbiased determination of the $M_t$ and JSF parameters.

**A. Combination of 1D and 2D fits**

The 1D and 2D fits discussed above have differing sensitivities to various sources of systematic uncertainty in this measurement. Although the 2D fit successfully mitigates the JES uncertainties, which dominate in the 1D fit, other uncertainties in the 2D method are larger and cause the total precision to worsen (Sec. VIII). The best overall precision on the value of $M_t$ is provided by a hybrid fit, defined as a linear combination of the 1D and 2D fits. The measured value of $M_t$ in the hybrid fit is given by

$$M_t^{\text{hyb}} = w_{\text{hyb}} M_t^{1D} + (1 - w_{\text{hyb}}) M_t^{2D},$$

where the parameter $w_{\text{hyb}}$ determines the relative weight between the 1D and 2D fits in the combination. The value of $M_t^{\text{hyb}}$ and its statistical uncertainty are extracted using bootstrap pseudo-experiments, as described above. In each pseudo-experiment, the measured value of $M_t^{\text{hyb}}$ is given by the linear combination in Eq. (9) of the measured $M_t^{1D}$ and $M_t^{2D}$ values. A value of $w_{\text{hyb}} = 0.8$ is found to achieve the best precision on $M_t$ when both statistical and systematic uncertainties are taken into account. The performance of the hybrid fit, evaluated using MC samples corresponding to seven values of $M_t^{\text{MC}}$, is shown in Fig. 6.

**VIII. SYSTEMATIC UNCERTAINTIES**

The systematic uncertainties evaluated in this measurement are given in Table I. The uncertainties include experimental effects from detector calibration and object reconstruction, and modeling effects mostly arising from the simulation of QCD processes. All uncertainties are determined by conducting the likelihood fit using events from MC simulation with the relevant parameters varied by $\pm 1\Delta$, where $\Delta$ is the uncertainty on a particular parameter. The difference in the measured top quark mass ($\delta M_t$) or JSF ($\delta \text{JSF}$) is taken to be the corresponding systematic uncertainty. For uncertainties that are evaluated by comparing two or more independent MC samples, the values of $\delta M_t$ and $\delta \text{JSF}$ may be subject to statistical fluctuations. For this reason, if the value of $\delta M_t$ or $\delta \text{JSF}$ is smaller than its statistical uncertainty in a particular systematic variation, the statistical uncertainty is quoted as the systematic uncertainty. Finally, if a systematic uncertainty is one sided, where both $+\Delta$ and $-\Delta$ variations produce $\delta M_t$ or $\delta \text{JSF}$ shifts of the same sign, the larger shift is taken as the symmetric systematic uncertainty.

In the hybrid fit, the systematic uncertainties are evaluated according to the linear combination in Eq. (9). For each systematic variation, this gives $\delta M_t^{\text{hyb}} = w_{\text{hyb}} \delta M_t^{1D} + (1 - w_{\text{hyb}}) \delta M_t^{2D}$. This approach provides the smallest overall uncertainty, with the largest contributions stemming from the JES, $b$ quark fragmentation modeling, and hard scattering scale. The next most precise result is given by the 1D fit, also dominated by the same sources of uncertainty. The JES uncertainties are successfully mitigated in the 2D fit. The 2D fit, however, is more sensitive to the uncertainties in the top quark $p_T$ spectrum, matching scale, and underlying event tune, so the total systematic uncertainty for the 2D fit is larger than that of the 1D fit. The MAOS fit has a larger total systematic uncertainty than the 1D fit due to its sensitivity to the JES, top quark $p_T$ spectrum, and $b$ quark fragmentation modeling uncertainties. Further details on each source of systematic uncertainty are given below.

(i) Jet energy scale: The JES uncertainty is evaluated separately for four components, which are then added in quadrature [44]. The “Intercalibration” uncertainty arises from the modeling of radiation.
TABLE I. Systematic uncertainties for the 2D, 1D, hybrid, and MAOS likelihood fits. The breakdown of JES and $b$ quark fragmentation uncertainties into separate components is shown, where the components are added in quadrature to obtain the total. The “up” and “down” variations are given separately, with the sign of each variation indicating the direction of the corresponding shift in $M_t$ or JSF. The ⊗ character highlights the uncertainty sources that are large in at least one of the likelihood fits.

<table>
<thead>
<tr>
<th>Source</th>
<th>$\delta M^{2D}_t$ [GeV]</th>
<th>$\delta JSF^{2D}$</th>
<th>$\delta M^{1D}_t$ [GeV]</th>
<th>$\delta M^{b,02}_t$ [GeV]</th>
<th>$\delta M^{MAOS}_t$ [GeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>JES (total)</td>
<td>⊗</td>
<td>+0.06</td>
<td>+0.007</td>
<td>+0.54</td>
<td>+0.43</td>
</tr>
<tr>
<td>– In situ</td>
<td></td>
<td>−0.10</td>
<td>−0.006</td>
<td>−0.55</td>
<td>−0.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+0.04</td>
<td>−0.002</td>
<td>−0.22</td>
<td>−0.18</td>
</tr>
<tr>
<td>– Uncorrelated</td>
<td></td>
<td>+0.04</td>
<td>+0.003</td>
<td>+0.21</td>
<td>+0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+0.01</td>
<td>+0.001</td>
<td>−0.04</td>
<td>−0.03</td>
</tr>
<tr>
<td>– Flavor</td>
<td></td>
<td>+0.04</td>
<td>−0.005</td>
<td>−0.39</td>
<td>−0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>−0.04</td>
<td>+0.005</td>
<td>+0.39</td>
<td>+0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+0.02</td>
<td>+0.004</td>
<td>+0.31</td>
<td>+0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>−0.09</td>
<td>−0.003</td>
<td>−0.32</td>
<td>−0.27</td>
</tr>
<tr>
<td>$b$ quark fragmentation (total)</td>
<td>⊗</td>
<td>+0.39</td>
<td>−0.001</td>
<td>+0.40</td>
<td>+0.40</td>
</tr>
<tr>
<td>– Fragmentation function</td>
<td></td>
<td>+0.38</td>
<td>−0.001</td>
<td>+0.38</td>
<td>+0.38</td>
</tr>
<tr>
<td>– Branching fraction</td>
<td></td>
<td>+0.07</td>
<td>+0.001</td>
<td>+0.13</td>
<td>+0.12</td>
</tr>
<tr>
<td>JER</td>
<td></td>
<td>+0.03</td>
<td>−0.002</td>
<td>+0.01</td>
<td>+0.00</td>
</tr>
<tr>
<td>Unclustered energy</td>
<td></td>
<td>+0.10</td>
<td>+0.001</td>
<td>−0.02</td>
<td>−0.04</td>
</tr>
<tr>
<td>Pileup</td>
<td></td>
<td>+0.06</td>
<td>−0.001</td>
<td>−0.06</td>
<td>−0.06</td>
</tr>
<tr>
<td>Electron energy scale</td>
<td></td>
<td>+0.39</td>
<td>−0.002</td>
<td>−0.21</td>
<td>−0.24</td>
</tr>
<tr>
<td>Muon momentum scale</td>
<td></td>
<td>−0.11</td>
<td>+0.001</td>
<td>−0.06</td>
<td>−0.07</td>
</tr>
<tr>
<td>Electron Id/iso</td>
<td></td>
<td>+0.09</td>
<td>−0.001</td>
<td>+0.05</td>
<td>+0.06</td>
</tr>
<tr>
<td>Muon Id/iso</td>
<td></td>
<td>+0.07</td>
<td>−0.001</td>
<td>+0.03</td>
<td>+0.03</td>
</tr>
<tr>
<td>$b$ tagging</td>
<td></td>
<td>+0.02</td>
<td>−0.001</td>
<td>−0.01</td>
<td>−0.01</td>
</tr>
<tr>
<td>Top quark $p_T$ reweighting</td>
<td>⊗</td>
<td>+0.93</td>
<td>−0.007</td>
<td>+0.40</td>
<td>+0.40</td>
</tr>
<tr>
<td>Hard scattering scale</td>
<td>⊗</td>
<td>+0.36</td>
<td>+0.007</td>
<td>+0.31</td>
<td>+0.21</td>
</tr>
<tr>
<td>Matching scale</td>
<td>⊗</td>
<td>+0.20</td>
<td>−0.003</td>
<td>−0.49</td>
<td>−0.47</td>
</tr>
<tr>
<td>Underlying event tunes</td>
<td>⊗</td>
<td>+0.30</td>
<td>+0.008</td>
<td>+0.11</td>
<td>+0.12</td>
</tr>
<tr>
<td>Color reconnection</td>
<td></td>
<td>+0.56</td>
<td>+0.007</td>
<td>+0.08</td>
<td>+0.11</td>
</tr>
<tr>
<td>ME generator</td>
<td></td>
<td>−0.56</td>
<td>−0.007</td>
<td>−0.08</td>
<td>−0.11</td>
</tr>
<tr>
<td>PDFs</td>
<td></td>
<td>+0.18</td>
<td>−0.004</td>
<td>−0.19</td>
<td>−0.13</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>+0.39</td>
<td>−0.001</td>
<td>+0.17</td>
<td>+0.17</td>
</tr>
</tbody>
</table>

in the $p_T$- and $\eta$-dependent JES determination. The “In situ” category includes uncertainties stemming from the determination of the absolute JES using $\gamma/Z$+jet events. The “Uncorrelated” uncertainty includes uncertainties due to detector effects and pileup. Finally, the “Flavor” uncertainty stems from differences in the energy response between different jet flavors—it is a linear sum of contributions from the light quark, charm quark, bottom quark, and gluon responses, which are estimated by comparing the Lund string fragmentation in PYTHIA [29] and cluster fragmentation in HERWIG++ [45] for each type of jet. All JES uncertainties are propagated into the reconstructed $p_T^{\text{miss}}$ in each event.

(ii) $b$ quark fragmentation: The $b$ quark fragmentation uncertainty includes two components that are implemented using event weights. The first component stems from the $b$ quark fragmentation function, which can modeled using the Lund fragmentation model in the PYTHIA ZZ+ tune, or tuned to empirical results from the ALEPH [46] and DELPHI [47] experiments. This component is evaluated by comparing the measurement results in MC simulation using these two tunes of the $b$ quark fragmentation function, with the difference symmetrized to obtain the corresponding uncertainty. The second uncertainty component stems from the $B$ hadron semileptonic branching fraction, which has an impact on
the $b$ quark JES due to the production of a neutrino. The corresponding uncertainty is evaluated by repeating the measurement with branching fraction values of 10.05% and 11.27%, which are variations about the nominal value of 10.50% and encompass the range of values measured from B hadron decays and their uncertainties [22]. Both uncertainty components are combined in quadrature to obtain the total uncertainty.

(iii) Jet energy resolution: The energy resolution of jets is known to be underestimated in MC simulation compared to data. This effect is corrected with a set of scale factors that are used to smear the jet four-vectors to broaden their resolutions. The scale factors are determined in bins of $\eta$. Here, they are varied within their uncertainties, which are typically 2.5%–5%. The effect of these variations is also propagated into the $p_T^{\text{miss}}$.

(iv) Unclustered energy: The unclustered energy in each event comprises the low-$p_T$ hadronic activity that is not clustered into a jet. Here, the scale of the unclustered energy is varied by ±10% [26].

(v) Pileup: The uncertainty in the number of pileup interactions in MC simulation stems from the instantaneous luminosity in each bunch crossing and the effective inelastic cross section. In this analysis, the number of pileup interactions in MC is reweighted to match the data. The pileup uncertainty is evaluated by varying the effective inelastic cross section by ±5%.

(vi) Lepton energy scale: The electron energy scale is varied up and down by 0.6% in the ECAL barrel ($|\eta| < 1.48$) and by 1.5% in the ECAL end cap ($1.48 < |\eta| < 3.0$) [20]. The muon momentum scale is varied up and down by 0.2%. All variations are propagated into the $p_T^{\text{miss}}$.

(vii) Lepton identification and isolation: Event weights are applied to adjust the electron and muon yields in MC simulation to account for differences in the identification and isolation efficiencies between data and simulation. For muons, the uncertainty is taken to be 0.5% of the identification event weight, and 0.2% of the isolation event weight [21]. For electrons, the uncertainties are estimated in bins of $p_T$ and $\eta$, and are approximately 0.1%–0.5% of the combined event weight for identification and isolation [20].

(viii) $b$ tagging efficiency: Event weights are applied to adjust the $b$ jet yields in MC simulation to account for the difference in the $b$ tagging efficiency between data and MC simulation [14]. The uncertainties are evaluated in bins of $p_T$ and $\eta$.

(ix) Top quark $p_T$ reweighting: Event weights are applied in order to compensate for a difference in the top quark $p_T$ spectrum between data and MC simulation [48]. The uncertainty is evaluated by comparing the measurement in MC simulation with and without the weights applied. The event weights are not applied in the nominal result. This uncertainty is one sided by construction, and is not symmetrized.

(x) Hard scattering scale: The factorization scale, $\mu_F$, determines the threshold separating the parton-parton hard scattering from softer interactions embodied in the PDFs. The renormalization scale, $\mu_R$, sets the energy scale at which matrix-element calculations are evaluated. Both of these scales are set to $\mu_F = \mu_R = Q$ in the matrix-element calculation and the initial-state parton shower of the MadGraph samples, where $Q^2 = M_T^2 + \sum p_T^2$. Here, the sum runs over all additional final state partons in the matrix element. The values of $\mu_F$ and $\mu_R$ are varied simultaneously up and down by a factor of 2 to estimate the corresponding uncertainty.

(xi) Matching scale: The matrix element-parton shower matching threshold is used to interface the matrix elements generated in MadGraph with parton showers simulated in PYTHIA. Its reference value of 20 GeV is varied up and down by a factor of 2.

(xii) Underlying event tunes and color reconnection: The underlying event tunes affect the modeling of soft hadronic activity that results from beam remnants and multiparton interactions in each event. The measurement is conducted with a $t\bar{t}$ sample from MC simulation using the “Perugia 2011” tune. It is compared to results using samples with the “Perugia 2011 mpiHi” and “Perugia 2011 Tevatron” tunes [49] in PYTHIA, corresponding to an increased and decreased underlying event activity, respectively. The largest difference is symmetrized to obtain the final uncertainty. The color reconnection (CR) uncertainty is evaluated by comparing measurement results using $t\bar{t}$ samples with the “Perugia 2011” and “Perugia 2011 no CR” tunes [49], where CR effects are not included in the latter. The difference is symmetrized to obtain the final uncertainty.

(xiii) Matrix-element generator: The measurement is repeated using MC samples produced with the POWHEG event generator, which provides a next-to-leading-order calculation of the $t\bar{t}$ production. These measurement results are compared with the reference $t\bar{t}$ MC sample, generated using MadGraph, to determine the corresponding uncertainty.

(xiv) Parton distribution functions: Initial-state partons are described by PDFs. The corresponding uncertainty is evaluated by applying event weights in the MC simulation to reflect the CT10 PDF set [50] with 50 error eigenvectors. The total PDF uncertainty is determined by adding the variations corresponding to these error sets in quadrature.

**IX. RESULTS AND DISCUSSION**

The results for each version of the likelihood fit, determined from 1000 bootstrap pseudo-experiments in each fit, are shown in Fig. 7. The 2D fit uses the $M_{bb}$ and $M_{T}^{ll}$ observables to simultaneously determine the values of $M_b$ and
FIG. 7. Likelihood fit results using 1000 bootstrap pseudo-experiments for the (top) 2D fit, (center left) 1D fit, and (center right) MAOS fit. (Bottom) Hybrid fit results given by the linear combination in Eq. (9) of the 1D and 2D fits. The error bars represent the statistical uncertainty corresponding to the number of pseudo-experiments in each bin.
yielding $M_{t}^{2D} = 171.56 \pm 0.46\,(\text{stat})^{+1.31}_{-1.25}\,(\text{syst})$ GeV and $\text{JSF}^{2D} = 1.011 \pm 0.006\,(\text{stat})^{+0.015}_{-0.014}\,(\text{syst})$. The correlation between the $M_{t}$ and JSF fit parameters in the 2D fit is shown in Fig. 8, with a correlation coefficient of $\rho = -0.94$. The $M_{b'}$ and $M_{bb}^{2D}$ distribution shapes corresponding to the fit results in a typical pseudo-experiment are shown in Fig. 9. The 2D fit is successful in mitigating the uncertainty due to the determination of JES, which is otherwise the largest source of systematic uncertainty in this measurement. In particular, this approach is insensitive to the flavor-dependent component of JES uncertainties—stemming from differences in the response between $b$ jets, light-quark jets, and gluon jets—since predominantly $b$ jets are used for the determination of both $M_{t}$ and JSF parameters. The underlying strategy, rooted in a simultaneous fit of two distributions with differing sensitivities to the JSF, does not rely on any specific assumptions about the event topology or final state. For this reason, it can be a viable option for JES uncertainty mitigation in a variety of physics scenarios.

The 1D fit is also based on the $M_{b'l}$ and $M_{bb}^{T2}$ observables, but constrains the JSF parameter to unity. The 1D fit gives a value of $M_{t}^{1D} = 172.39 \pm 0.17\,(\text{stat})^{+0.91}_{-0.95}\,(\text{syst})$ GeV. In this approach, the JES accounts for the largest source of uncertainty. However, other uncertainties are reduced with respect to the 2D fit, resulting in an improved overall precision.
The best overall precision is given by the hybrid fit, which is given by a linear combination of the 1D and 2D fit results. The 1D and 2D fits use the same set of events and an identical likelihood function constructed from the $M_{b\ell}$ and $M_{T2}^{bb}$ observables. These fits are fully correlated, with the only difference between them stemming from the treatment of the JSF parameter, which is fixed to unity in the 1D fit and acts as a free parameter in the 2D fit. The choice to fix the JSF parameter or allow it to float has an impact on the fit sensitivity to a variety of uncertainty sources in addition to the JES. A linear combination of the 1D and 2D fits with $w_{hyb} = 0.8$, as defined in Eq. (9), achieves an optimal balance between all uncertainty sources, thus providing the best overall precision. The hybrid fit gives

$$M_{t}^{hyb} = 172.22 ± 0.18(\text{stat}) ^{+0.89}_{−0.93} (\text{syst}) \text{ GeV}.$$  

The correlation between the $M_t$ and JSF fit parameters in the hybrid fit is shown in Fig. 8, with a correlation coefficient of $\rho = −0.40$.

The MAOS fit substitutes the $M_{b\ell}$ observable for an $M_{b\ell}^{i}$ invariant mass, yielding a value of $M_{t}^{\text{MAOS}} = 171.54 ± 0.19(\text{stat}) ^{+1.27}_{−1.02} (\text{syst}) \text{ GeV}$. The MAOS observable presents a new approach for mass reconstruction in a decay topology characterized by underconstrained kinematics. Here, the MAOS fit provides a determination of $M_t$ that is complementary to the 2D, 1D, and hybrid fits. The MAOS $M_{b\ell}^{i}$ distribution shape corresponding to the fit results in a typical pseudo-experiment is shown in Fig. 10.

The results for each version of the likelihood fit are summarized in Fig. 11.

X. SUMMARY

A measurement of the top quark mass ($M_t$) in the dileptonic $t\bar{t}$ decay channel is performed using proton-proton collisions at $\sqrt{s} = 8 \text{ TeV}$, corresponding to an integrated luminosity of $19.7 ± 0.5 \text{ fb}^{-1}$. The measurement is based on the mass observables $M_{b\ell}$, $M_{T2}^{bb}$, and $M_{b\ell}^{i}$, which allow for mass reconstruction in decay topologies that are kinematically underconstrained. The sensitivity of these observables to the value of $M_t$ is investigated using a Fisher information density technique. The observables are employed in three versions of an unbinned likelihood fit, where a Gaussian process technique is used to model the corresponding distribution shapes and their evolution in $M_t$ and an overall jet energy scale factor (JSF). The Gaussian process shapes are nonparametric, and allow for a likelihood fitting framework that gives unbiased results. The 2D fit provides the first simultaneous measurement of $M_t$ and JSF in the dileptonic channel. It is robust against uncertainties due to the determination of jet energy scale, including the flavor-dependent uncertainty component arising from differences in the response between $b$ jets, light-quark jets, and gluon jets. The fit yields $M_t = 171.56 ± 0.46(\text{stat}) ^{+1.31}_{−1.25} (\text{syst}) \text{ GeV}$ and JSF = $1.011 ± 0.006(\text{stat}) ^{+0.015}_{−0.014} (\text{syst})$. The most precise measurement of $M_t$ is given by a linear combination of this result with a fit in which the JSF is constrained to be unity, yielding a value.
of 172.22 ± 0.18(stat) ±0.89−0.93(syst) GeV. This measurement achieves a 25% improvement in overall precision on $M$, compared to previous dileptonic channel analyses using the 2012 data set at CMS. The improvement can be attributed to a reduction of the systematic uncertainties in the measurement.

ACKNOWLEDGMENTS

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMWFW and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COCIFI (Colombia); MSES and CSF (Croatia); RPF (Cyprus); SENESCYT (Ecuador); MoER, ERC IUT, and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and MTA (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); MSIP and NRF (Republic of Korea); LAS (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MBIE (New Zealand); PAEC (Pakistan); MSHE and NSC (Poland); FCT (Portugal); JINR (Dubna); MON, RosAtom, RAS, RFBR and RAEP (Russia); MESTD (Serbia); SEIDI, CPAN, PCTI and FEDER (Spain); Swiss Funding Agencies (Switzerland); MST (Taipei); ThEPCenter, IPST, STAR, and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU and SFFR (Ukraine); STFC (United Kingdom); DOE and NSF (USA). Individuals have received support from the Marie-SFFR (Ukraine); STFC (United Kingdom); DOE and NSF MST (Taipei); ThEPCenter, IPST, STAR, and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU and SFFR (Ukraine); STFC (United Kingdom); DOE and NSF (USA). Individuals have received support from the Marie-Curie program and the European Research Council and EPLANET (European Union); the Leventis Foundation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l’Industrie et dans l’Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Council of Science and Industrial Research, India; the HOMING PLUS program of the Foundation for Polish Science, cofinanced by European Union, Regional Development Fund, the Mobility Plus program of the Ministry of Science and Higher Education, the National Science Center (Poland), Contracts Harmonia No. 2014/14/M/ST2/00428, Opus No. 2014/13/B/ST2/02543, No. 2014/15/B/ST2/03998, and No. 2015/19/B/ST2/02861, Sonata-bis No. 2012/07/E/ST2/01406; the National Priorities Research Program by Qatar National Research Fund; the Programa Clarín-COFUND del Principado de Asturias; the Thalis and Aristeia programs cofinanced by EU-ESF and the Greek NSRF; the Rachadapisek Sompot Fund for Postdoctoral Fellowship, Chulalongkorn University and the Chulalongkorn Academic into Its 2nd Century Project Advancement Project (Thailand); and the Welch Foundation, Contract No. C-1845.

APPENDIX A: STATISTICAL SENSITIVITY OF KINEMATIC OBSERVABLES

The sensitivity of a kinematic observable to the value of a parameter such as $M$, can be quantified by its Fisher information [51,52]. The Fisher information of an observable is related to its likelihood function, $L$, which we have introduced in Eq. (7) and reproduce here:

$$\log L(m) = \sum_i^n \log f(x_i|m).$$ (A1)

where $f(x|m)$ is the distribution of observable $x$ normalized to unity over its range, $m$ is a free parameter, and $N$ is the number of observations of $x$. In this measurement, we have $x = M_{h\ell}$, $M_{W2}$, or $M_{h\ell_{\nu}}$, $m = M$, or JSF, and $N$ is a multiple of the total number of events. For simplicity we consider the distribution shape $f$ as a function of only one free parameter. The Fisher information corresponding to the shape $f(x|m)$ is given by

$$I(m) = \int \left( \frac{\partial}{\partial m} \log f(x|m) \right)^2 f(x|m) dx.$$ (A2)

The quantity $I(m)$ provides a measure of curvature near the likelihood maximum. It can be interpreted as the variance of the slope, $(\partial \log f(x|m)/\partial m)$, known as the “statistical score” of $f(x|m)$.

The Fisher information is related to the precision of a measurement by the Crámer-Rao bound:

$$\sigma_m^2 \geq \frac{1}{NI(m)},$$ (A3)

where $\sigma_m$ is the statistical uncertainty on parameter $m$. In a likelihood with large $N$, the shape of the likelihood near its maximum is roughly Gaussian, and the bound approaches an equality. This expression confirms the expected relationship $\sigma_m \propto 1/\sqrt{N}$ between the statistical uncertainty and the value of $N$, but also reveals the proportionality factor as the reciprocal of the Fisher information. It expresses the uncertainty $\sigma_m$ in terms of the total number of events, the shape $f$, and the derivative $\partial f/\partial m$. 

032002-16
The Fisher information also provides a mathematical framework for quantifying the sensitivity of an observable at a specific point on its shape. In this analysis, the $M_{bb}$ and $M_{\ell_2}^{hb}$ observables have kinematic end points at approximately $\sqrt{M_t^2 - M_{bb}^2}$ and $M_{\ell_2}$, respectively; the MAOS $M_{b\ell\ell}$ observable is an invariant mass whose shape contains a peak near the value of $M_t$. Because these features carry a dependence on the value of $M_t$, the regions near the end points of $M_{bb}$ and $M_{\ell_2}^{hb}$ and the peak of $M_{b\ell\ell}$ are expected to contribute significantly to the sensitivity of these observables. To relate these local features to the Fisher information, we consider the integral in Eq. (A2) over the value of observable $x$. Here, the integrand of the Fisher information can be interpreted as the contribution to the total sensitivity stemming from a specific value of $x$. Rewriting the integrand in a more convenient form, we define the local shape sensitivity function by

$$s(x|m) \equiv \frac{1}{f(x|m)} \left[ \frac{\partial f(x|m)}{\partial m} \right]^2. \quad (A4)$$

This function is also known as the Fisher information density. It is shown for the $M_{b\ell\ell}$, $M_{\ell_2}^{bb}$, and MAOS $M_{b\ell\ell}$ observables in Figs. 1, 3, and 4, respectively, with $m = M_t$ and the JSF parameter fixed to unity. It is observed to peak near the kinematic end points of $M_{bb}$ and $M_{\ell_2}^{bb}$, and on the left-side edge of $M_{b\ell\ell}$. The values of $x$ where $s(x|m) = 0$ coincide with the stationary points at which the distribution shapes in Figs. 1, 3, and 4 intersect. This is a reflection of the fact that in a likelihood fit, events with a value of $x$ near a stationary point make little or no contribution to the determination of $m$. In general, the shape of $s(x|m)$ for each observable establishes a link between the underlying kinematic properties of the observable and regions of high and low sensitivity on its shape. In this analysis, it provides heuristic information about the $M_{b\ell\ell}$, $M_{\ell_2}^{bb}$, and MAOS $M_{b\ell\ell}$ distributions, and their sensitivity to the value of $M_t$.

In addition to providing heuristic information, the local shape sensitivity function is used in this analysis to identify potential overfitting effects in the Gaussian process (GP) shapes. Overfitting occurs when the interpolation between GP training points is not smooth, causing fluctuations in the shape that may be difficult to identify by eye. Such fluctuations can be a source of bias, both in the determination of $M_t$ and its corresponding uncertainties. A typical symptom of overfitting is an underestimated statistical uncertainty on the value of $M_t$. This can occur when fluctuations in the GP shape increase the value of the slope $\partial f(x|M_t)/\partial M_t$ appearing in Eq. (A2), thus artificially increasing the Fisher information of the corresponding shape. The issue is easily revealed by the shape of $s(x|M_t)$, which acquires visible fluctuations when overfitting is indeed present. In such cases, overfitting can be mitigated by increasing relevant GP hyperparameter values to improve the smoothness of the GP shape.

### Appendix B: Gaussian Process Regression Technique

The likelihood fit described in Sec. VII uses distribution shapes of the form $f(x|M_t, \text{JSF})$, where $x$ is the value of an observable ($M_{bb}$, $M_{\ell_2}^{bb}$, or $M_{b\ell\ell}$), and $M_t$ and JSF are free parameters in the fit. In this analysis, the distribution shapes $f$ are modeled with a Gaussian process (GP) regression technique. We define a point, $u_i$, on each distribution shape by its position in $x$, $M_t$, and JSF:

$$u_i \equiv (x_i, M_t, \text{JSF}). \quad (B1)$$

The value of the shape at $u_i$ is given by $f(u_i) = f(x_i|M_t, \text{JSF})$. The point $u_i$ can be a training point, at which the value of $f$ is known and used as an input into the GP regression process; or it can be a test point, at which the value of $f$ is to be determined. Each GP shape is trained using binned distributions of the observable $x$ in MC simulation. For each observable, 35 binned distributions are used, corresponding to seven values of $M_{bb}^{MC}$ ranging from 166.5 to 178.5 GeV and five values of JSF ranging from 0.97 to 1.03. Each distribution has 75 bins in $x$, yielding a total of 2625 training points. This binning scheme is chosen to provide an accurate modeling of the distribution shapes, while mitigating the effects of statistical fluctuations. The GP regression technique interpolates between the discrete values of $x$, $M_t$, and JSF covered by these training points to provide a shape that is smooth over its range.

The “Gaussian” in GP refers to the distribution of possible values of the shape $f$. The value at a single point, $f(u_i)$, is distributed according to a one-dimensional Gaussian function rather than being treated as an exact quantity. The mean of this Gaussian function is the most probable value of the shape at that point $u_i$, and it is the value used for likelihood fitting (Sec. VII); the variance stems from the modeling uncertainty inherent in the GP regression process. The values $f(u_i)$ and $f(u_j)$ at any two points follow a two-dimensional Gaussian distribution and are related by a covariance. The correlation between $f(u_i)$ and $f(u_j)$ determines the degree to which the GP shape is allowed to vary between the points $u_i$ and $u_j$. By extension, any $N$ values of the shape are described by an $N$-dimensional Gaussian distribution, and are related by an $N \times N$ covariance matrix. To determine the value of the shape at a test point $u_{N+1}$, an $(N+1)$-dimensional Gaussian distribution is constructed relating the training point values $f(u_1) \ldots f(u_N)$ to the test point value $f(u_{N+1})$. Then, $f(u_1) \ldots f(u_N)$ are fixed to their known values, and the $(N+1)$-dimensional Gaussian distribution is reduced to a one-dimensional conditional Gaussian distribution representing the possible values of $f(u_{N+1})$.

To demonstrate this process graphically, we consider a simple GP with one training point, $u_{\text{train}}$, at which the value
Here, \( \sigma \) is a noise parameter that accounts for the statistical uncertainty on the distribution bin underlying each training point, and \( \delta_{ij} \) is the Kronecker delta function. The terms inside the exponentials specify the covariance between any two values of the shape as a function of their corresponding \( x \), \( M_i \), and JSF. The hyperparameters \( \theta_1 \), \( \theta_2 \), and \( \theta_3 \) specify the length scales over which the GP shape is allowed to vary, and \( \rho \) is a correlation coefficient that couples the \( M \) and JSF parameters. The hyperparameter \( N_i \) specifies the
overall normalization of the kernel function, and \( N_2 \) determines the relative normalization between the Gaussian and noise terms.

The values of all hyperparameters are determined with the help of a cross-validation likelihood fit \([12]\), conducted for each observable separately. The length scale hyperparameters \((\theta_1, \theta_2, \text{and } \theta_3)\) must be small enough for the GP shape to pass through the training points, and large enough for the shape to interpolate smoothly between them. Hyperparameters that are underestimated satisfy the former criterion, but cause overfitting to occur in the resulting GP shape. This creates a noisy interpolation between training points, and may lead to bias in the measured value of \( M_t \) and its uncertainties. In this analysis, the GP shapes are checked for overfitting effects using the local shape sensitivity function described in Appendix A.

---

A. M. SIRUNYAN et al.


MEASUREMENT OF THE TOP QUARK MASS IN THE ...
<table>
<thead>
<tr>
<th>Institution Name</th>
<th>City, Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institute for Nuclear Research and Nuclear Energy, Sofia, Bulgaria</td>
<td>Sofia, Bulgaria</td>
</tr>
<tr>
<td>University of Sofia, Sofia, Bulgaria</td>
<td></td>
</tr>
<tr>
<td>Beihang University, Beijing, China</td>
<td></td>
</tr>
<tr>
<td>Institute of High Energy Physics, Beijing, China</td>
<td></td>
</tr>
<tr>
<td>State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, China</td>
<td></td>
</tr>
<tr>
<td>Universidad de Los Andes, Bogota, Colombia</td>
<td></td>
</tr>
<tr>
<td>University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia</td>
<td></td>
</tr>
<tr>
<td>University of Split, Faculty of Science, Split, Croatia</td>
<td></td>
</tr>
<tr>
<td>Institute Rudjer Boskovic, Zagreb, Croatia</td>
<td></td>
</tr>
<tr>
<td>University of Cyprus, Nicosia, Cyprus</td>
<td></td>
</tr>
<tr>
<td>Charles University, Prague, Czech Republic</td>
<td></td>
</tr>
<tr>
<td>Universidad San Francisco de Quito, Quito, Ecuador</td>
<td></td>
</tr>
<tr>
<td>Academy of Scientific Research and Technology of the Arab Republic of Egypt, Egyptian Network of High Energy Physics, Cairo, Egypt</td>
<td></td>
</tr>
<tr>
<td>National Institute of Chemical Physics and Biophysics, Tallinn, Estonia</td>
<td></td>
</tr>
<tr>
<td>Department of Physics, University of Helsinki, Helsinki, Finland</td>
<td></td>
</tr>
<tr>
<td>Helsinki Institute of Physics, Helsinki, Finland</td>
<td></td>
</tr>
<tr>
<td>Lappeenranta University of Technology, Lappeenranta, Finland</td>
<td></td>
</tr>
<tr>
<td>IRFU, CEA, Université Paris-Saclay, GIF-sur-Yvette, France</td>
<td></td>
</tr>
<tr>
<td>Laboratoire Leprince-Ringuet, Ecole polytechnique, CNRS/IN2P3, Université Paris-Saclay</td>
<td></td>
</tr>
<tr>
<td>Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France</td>
<td></td>
</tr>
<tr>
<td>Centre de Calcul de l’Institut National de Physique Nucléaire et de Physique des Particules, CNRS/IN2P3, Villeurbanne, France</td>
<td></td>
</tr>
<tr>
<td>Université de Lyon, Université Claude Bernard Lyon 1, CNRS-IN2P3, Institut de Physique Nucléaire de Lyon, Villeurbanne, France</td>
<td></td>
</tr>
<tr>
<td>Georgian Technical University, Tbilisi, Georgia</td>
<td></td>
</tr>
<tr>
<td>Tbilisi State University, Tbilisi, Georgia</td>
<td></td>
</tr>
<tr>
<td>RWTH Aachen University, I. Physikalisches Institut, Aachen, Germany</td>
<td></td>
</tr>
<tr>
<td>RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany</td>
<td></td>
</tr>
<tr>
<td>RWTH Aachen University, III. Physikalisches Institut B, Aachen, Germany</td>
<td></td>
</tr>
<tr>
<td>Deutsches Elektronen-Synchrotron, Hamburg, Germany</td>
<td></td>
</tr>
<tr>
<td>University of Hamburg, Hamburg, Germany</td>
<td></td>
</tr>
<tr>
<td>Institut für Experimentelle Kernphysik, Karlsruhe, Germany</td>
<td></td>
</tr>
<tr>
<td>Institute of Nuclear and Particle Physics (INPP), NCSR Demokritos, Aghia Paraskevi, Greece</td>
<td></td>
</tr>
<tr>
<td>National and Kapodistrian University of Athens, Athens, Greece</td>
<td></td>
</tr>
<tr>
<td>National Technical University of Athens, Athens, Greece</td>
<td></td>
</tr>
<tr>
<td>University of Ioánnina, Ioánnina, Greece</td>
<td></td>
</tr>
<tr>
<td>MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary</td>
<td></td>
</tr>
<tr>
<td>Wigner Research Centre for Physics, Budapest, Hungary</td>
<td></td>
</tr>
<tr>
<td>Institute of Nuclear Research ATOMKI, Debrecen, Hungary</td>
<td></td>
</tr>
<tr>
<td>Institute of Physics, University of Debrecen</td>
<td></td>
</tr>
<tr>
<td>Indian Institute of Science (IISC)</td>
<td></td>
</tr>
<tr>
<td>National Institute of Science Education and Research, Bhubaneswar, India</td>
<td></td>
</tr>
<tr>
<td>Panjab University, Chandigarh, India</td>
<td></td>
</tr>
<tr>
<td>University of Delhi, Delhi, India</td>
<td></td>
</tr>
<tr>
<td>Saha Institute of Nuclear Physics, Kolkata, India</td>
<td></td>
</tr>
<tr>
<td>Indian Institute of Technology Madras, Madras, India</td>
<td></td>
</tr>
<tr>
<td>Bhabha Atomic Research Centre, Mumbai, India</td>
<td></td>
</tr>
<tr>
<td>Tata Institute of Fundamental Research-A, Mumbai, India</td>
<td></td>
</tr>
<tr>
<td>Tata Institute of Fundamental Research-B, Mumbai, India</td>
<td></td>
</tr>
<tr>
<td>Indian Institute of Science Education and Research (IISER), Pune, India</td>
<td></td>
</tr>
<tr>
<td>Institute for Research in Fundamental Sciences (IPM), Tehran, Iran</td>
<td></td>
</tr>
<tr>
<td>University College Dublin, Dublin, Ireland</td>
<td></td>
</tr>
<tr>
<td>INFN Sezione di Bari, Bari, Italy</td>
<td></td>
</tr>
<tr>
<td>Università di Bari, Bari, Italy</td>
<td></td>
</tr>
<tr>
<td>Politecnico di Bari, Bari, Italy</td>
<td></td>
</tr>
<tr>
<td>INFN Sezione di Bologna, Bologna, Italy</td>
<td></td>
</tr>
</tbody>
</table>
A. M. SIRUNYAN et al.

PHYSICAL REVIEW D 96, 032002 (2017)

65b Università di Bologna, Bologna, Italy
65c INFN Sezione di Catania, Catania, Italy
65d Università di Catania, Catania, Italy
65e INFN Sezione di Firenze, Firenze, Italy
65f Università di Firenze, Firenze, Italy
65g INFN Laboratori Nazionali di Frascati, Frascati, Italy
65h INFN Sezione di Genova, Università di Genova, Genova, Italy
65i INFN Sezione di Genova, Genova, Italy
65j Università di Genova, Genova, Italy
65k Università di Milano-Bicocca, Milano, Italy
65l INFN Sezione di Napoli, Napoli, Italy
65m Università di Napoli 'Federico II', Napoli, Italy
65n Università della Basilicata, Rom, Italy
65o Università G. Marconi, Roma, Italy
65p INFN Sezione di Padova, Padova, Italy
65q Università di Padova, Padova, Italy
65r Università di Trento, Trento, Italy
65s INFN Sezione di Pavia, Pavia, Italy
65t Università di Pavia, Pavia, Italy
65u INFN Sezione di Perugia, Perugia, Italy
65v Università di Perugia, Perugia, Italy
65w INFN Sezione di Pisa, Pisa, Italy
65x Università di Pisa, Pisa, Italy
65y Scuola Normale Superiore di Pisa, Pisa, Italy
65z INFN Sezione di Roma (a); Sapienza Università di Roma (b)
66a INFN Sezione di Roma, Roma, Italy
66b Università di Roma, Roma, Italy
66c INFN Sezione di Torino, Torino, Italy
66d Università di Torino, Torino, Italy
66e Università del Piemonte Orientale, Novara, Italy
66f INFN Sezione di Trieste, Trieste, Italy
66g Università di Trieste, Trieste, Italy
67a Kyungpook National University, Daegu, Korea
67b Chonbuk National University, Jeonju, Korea
67c Chonnam National University, Institute for Universe and Elementary Particles, Kwangju, Korea
67d Hanyang University, Seoul, Korea
67e Korea University, Seoul, Korea
67f Seoul National University, Seoul, Korea
67g University of Seoul, Seoul, Korea
67h Sungkyunkwan University, Suwon, Korea
67i Vilnius University, Vilnius, Lithuania
68a National Centre for Particle Physics, Universiti Malaya, Kuala Lumpur, Malaysia
68b Centro de Investigación y de Estudios Avanzados del IPN, Mexico City, Mexico
68c Universidad Iberoamericana, Mexico City, Mexico
68d Benemerita Universidad Autónoma de Puebla, Puebla, Mexico
68e Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico
68f University of Auckland, Auckland, New Zealand
68g University of Canterbury, Christchurch, New Zealand
68h National Centre for Physics, Quaid-I-Azam University, Islamabad, Pakistan
68i National Centre for Nuclear Research, Swierk, Poland
68j Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw, Poland
68k Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal
68l Joint Institute for Nuclear Research, Dubna, Russia
68m Petersburg Nuclear Physics Institute, Gatchina (St. Petersburg), Russia
68n Institute for Nuclear Research, Moscow, Russia
68o Institute for Theoretical and Experimental Physics, Moscow, Russia
68p Moscow Institute of Physics and Technology, Moscow, Russia
68q National Research Nuclear University 'Moscow Engineering Physics Institute' (MEPhI), Moscow, Russia
68r P. N. Lebedev Physical Institute, Moscow, Russia
68s Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia

032002-28
A. M. SIRUNYAN et al.

PHYSICAL REVIEW D 96, 032002 (2017)

167 Princeton University, Princeton, USA
168 University of Puerto Rico, Mayaguez, USA
169 Purdue University, West Lafayette, USA
170 Purdue University Northwest, Hammond, USA
171 Rice University, Houston, USA
172 University of Rochester, Rochester, USA
173 Rutgers, The State University of New Jersey, Piscataway, USA
174 University of Tennessee, Knoxville, USA
175 Texas A&M University, College Station, USA
176 Texas Tech University, Lubbock, USA
177 Vanderbilt University, Nashville, USA
178 University of Virginia, Charlottesville, USA
179 Wayne State University, Detroit, USA
180 University of Wisconsin—Madison, Madison, Wisconsin, USA

*Deceased.

† Also at Vienna University of Technology, Vienna, Austria.
‡ Also at State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, China.
§ Also at Universidade Estadual de Campinas, Campinas, Brazil.
∥ Also at Universidade Federal de Pelotas, Pelotas, Brazil.
¶ Also at Université Libre de Bruxelles, Bruxelles, Belgium.
‖ Also at Universidad de Antioquia, Medellin, Colombia.
¶¶ Also at Joint Institute for Nuclear Research, Dubna, Russia.
¶§ Also at Suez University, Suez, Egypt.
¶¶ Also at British University in Egypt, Cairo, Egypt.
¶¶¶ Also at Ain Shams University, Cairo, Egypt.
¶¶¶¶ Also at Helwan University, Cairo, Egypt.
¶¶¶¶¶ Also at Université de Haute Alsace, Mulhouse, France.
¶¶¶¶¶¶ Also at Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia.
¶¶¶¶¶¶¶ Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland.
¶¶¶¶¶¶¶¶ Also at RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany.
¶¶¶¶¶¶¶¶¶ Also at University of Hamburg, Hamburg, Germany.
¶¶¶¶¶¶¶¶¶¶ Also at Brandenburg University of Technology, Cottbus, Germany.
¶¶¶¶¶¶¶¶¶¶¶ Also at Institute of Nuclear Research ATOMKI, Debrecen, Hungary.
¶¶¶¶¶¶¶¶¶¶¶¶ Also at MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary.
¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Institute of Physics, University of Debrecen.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at IIT Bhubaneswar.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at University of Visva-Bharati, Santiniketan, India.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Institute of Physics, Bhubaneswar, India.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at University of Ruhuna, Matara, Sri Lanka.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Isfahan University of Technology, Isfahan, Iran.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Yazd University, Yazd, Iran.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Plasma Physics Research Center, Science and Research Branch, Islamic Azad University, Tehran, Iran.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Università degli Studi di Siena, Siena, Italy.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Purdue University, West Lafayette, USA.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at International Islamic University of Malaysia, Kuala Lumpur, Malaysia.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Malaysian Nuclear Agency, MOSTI, Kajang, Malaysia.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Consejo Nacional de Ciencia y Tecnología, Mexico city, Mexico.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Warsaw University of Technology, Institute of Electronic Systems, Warsaw, Poland.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Institute for Nuclear Research, Moscow, Russia.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at National Research Nuclear University ’Moscow Engineering Physics Institute’ (MEPhI), Moscow, Russia.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at St. Petersburg State Polytechnical University, St. Petersburg, Russia.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at University of Florida, Gainesville, USA.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at P. N. Lebedev Physical Institute, Moscow, Russia.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at California Institute of Technology, Pasadena, USA.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Budker Institute of Nuclear Physics, Novosibirsk, Russia.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Faculty of Physics, University of Belgrade, Belgrade, Serbia.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at INFN Sezione di Roma (a); Sapienza Università di Roma (b).
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at University of Belgrade, Faculty of Physics and Vinca Institute of Nuclear Sciences, Belgrade, Serbia.
¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ Also at Scuola Normale e Sezione dell’INFN, Pisa, Italy.