Published for SISSA by Springer

Received: March 4, 2020
Accepted: May 3, 2020
Published: June 1, 2020

On the ATLAS top mass measurements and the potential for stealth stop contamination

Timothy Cohen,[a] **Stephanie Majewski,**[a] **Bryan Ostdiek**[a,b] **and Peter Zheng**[a]

[a] *Institute for Fundamental Science, Department of Physics, University of Oregon, 1371 E 13th Avenue, Eugene, OR 97403, U.S.A.*

[b] *Department of Physics, Harvard University, 17 Oxford Street, Cambridge, MA 02138, U.S.A.*

E-mail: tcohen@uoregon.edu, smajewsk@uoregon.edu, bostdiek@g.harvard.edu, pzheng@uoregon.edu

Abstract: The discovery of the stop — the Supersymmetric partner of the top quark — is a key goal of the physics program enabled by the Large Hadron Collider. Although much of the accessible parameter space has already been probed, all current searches assume the top mass is known. This is relevant for the "stealth stop" regime, which is characterized by decay kinematics that force the final state top quark off its mass shell; such decays would contaminate the top mass measurements. We investigate the resulting bias imparted to the template method based ATLAS approach. A careful recasting of these results shows that effect can be as large as 2.0 GeV, comparable to the current quoted uncertainty on the top mass. Thus, a robust exploration of the stealth stop splinter requires the simultaneous consideration of the impact on the top mass. Additionally, we explore the robustness of the template technique, and point out a simple strategy for improving the methodology implemented for the semi-leptonic channel.

Keywords: Hadron-Hadron scattering (experiments), Top physics, top squark

ArXiv ePrint: 1909.09670

https://doi.org/10.1007/JHEP06(2020)019

JHEP06(2020)019

Contents

1 Introduction

The top quark plays a critical role in understanding the structure of the Standard Model (SM) and its extensions. The measured value of the top quark mass m_t (and Yukawa coupling) is an important input for precision tests of the self consistency of the SM. If nature is Supersymmetric (SUSY), the top should have a partner — the stop — that

JHEP06(2020)019

tames the ultraviolet sensitivity implied by the coupling between the top quark and the Higgs boson. Since this is one of the most compelling ways to extend the SM, an extensive search program for the stop has been conducted by both ATLAS [1–3] and CMS [4, 5], yielding an impressive exclusion covering stop masses as high as $\sim 1.2\,\text{TeV}$. Although the narrow "splinter" region ($m(\tilde{t}_1) \sim m_t$) now appears to be closed, all current searches assume that m_t is known. As we will argue here, this assumption deserves further scrutiny.

The SUSY framework makes it manifest that as the mass of the stop becomes parametrically large with respect to the weak scale, the fundamental parameters become increasingly fine tuned in order to reproduce the measured Higgs vacuum expectation value. Thus, there remains significant interest in this inherently natural but notoriously difficult to explore "stealth" stop region of parameter space. The degeneracy of these mass parameters implies tight kinematic constraints such that the final state looks nearly identical to top pair production, albeit where the tops are off-shell. Thus, not only does the presence of copious SM top pair production obscure the presence of the stop, but if the stop exists with a mass in this regime, the precision measurements of the top mass itself would be biased due to the presence of stop decays. We build upon previous studies of the subtle phenomenological signals of stealth stops [6, 7] in a number of ways: we study the impact on all three channels (all-hadronic, semi-leptonic, and di-leptonic) tracking their correlated effects, we use the most up-to-date ATLAS measurements, and we recast the template method in detail and study its robustness. Our main results quantify the potential contamination of these measurements due to a stealth stop. Furthermore, we propose an improvement in the methodology for measuring the top mass by explicitly using both tops in the event; we highlight this in the semi-leptonic channel.

To achieve this goal, we carefully recast precise measurements of the top quark mass, choosing the ATLAS Collaboration's template method for its straightforward response to stop signal contamination; we expect our results to be generally applicable regardless of the method used.[1] The ATLAS Collaboration characterizes its measurements by the top decay products considered: all-hadronic [12] (with no leptons in the final state), semi-leptonic [13] (where one top quark decays to jets and the other decays via an electron or muon), and di-leptonic [14] (where both top quarks decay via an electron or muon). These measurements are based on the $\sqrt{s} = 8\,\text{TeV}$ dataset and are summarized in the left panel of figure 1; we also provide a crude combination[2] assuming uncorrelated Gaussian error bars, which is consistent with the sophisticated combination performed by ATLAS that includes the $\sqrt{s} = 7\,\text{TeV}$ dataset. A general review of the template method together with an illustrative toy example is presented in section 2. In section 3, we present a detailed analysis of the semi-leptonic channel and propose an improved strategy that requires minor modifications to the current ATLAS approach.

The potential contamination from a stealth stop is modeled using the "stop-neutralino" Simplified Model [15–17], which is inspired by the "more minimal SUSY SM" [18, 19]. Under the well-motivated assumption the lightest superpartner is a stable state, phenomeno-

[1] The top mass has also been precisely measured by the CDF [8] and D0 [9] Collaborations using the matrix element technique, and by the CMS Collaboration using the ideogram method [10, 11].

[2] The details of our naive approach for combining measurements are presented in section E.

JHEP06(2020)019

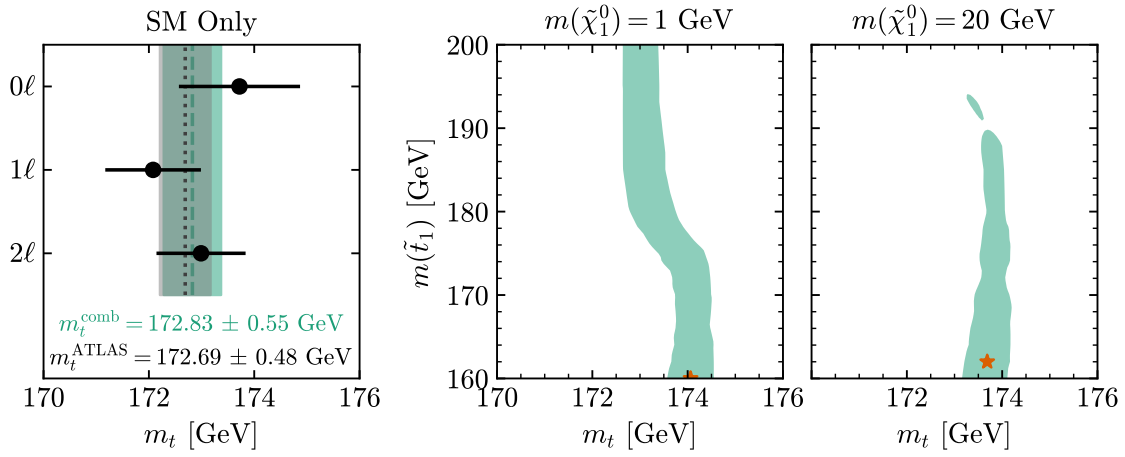

Figure 1. Parameters which give the best fit for the top mass measurement by ATLAS by combining the three different channels at $\sqrt{s} = 8$ TeV. The ATLAS measurements of the three different channels are shown in the left panel: all-hadronic [12] ($m_t = 173.72 \pm 1.15$ GeV), semi-leptonic [13] ($m_t = 172.08 \pm 0.91$ GeV), and di-leptonic [14] ($m_t = 172.99 \pm 0.84$ GeV). Assuming the SM only, the green band shows our crude best-fit value of $m_t^{\rm comb}$ with its associated uncertainty and the black band shows the ATLAS combination $\left(m_t^{\rm ATLAS}\right)$ given in ref. [13], taking into account 7 and 8 TeV data. The center and right panels illustrate the impact of stealth stop contamination, where we show the best fit point in the m_t-$m(\tilde{t}_1)$ plane (marked by the orange star), and the 1-σ confidence interval as shaded bands. When the stops must decay through an off-shell top, they shift the reconstructed template mass extraction to values that are smaller than the actual top mass chosen in the Monte Carlo event generation.

logical viability requires that the particle be neutral, thereby providing a dark matter candidate [20, 21], the so-called lightest neutralino $\tilde{\chi}_1^0$. The rate of direct stop pair production is fully specified by $m(\tilde{t}_1)$, and each stop subsequently decays to an on- or off-shell top quark and $\tilde{\chi}_1^0$, as illustrated in figure 2.[3] The stealth stop region of parameter space is thus more precisely defined by $m(\tilde{t}_1) - m_t \simeq m(\tilde{\chi}_1^0)$. The degeneracy of these mass parameters implies tight kinematic constraints such that the final state looks nearly identical to top pair production, perhaps with some additional missing energy due to the presence of the neutralino. There have been many phenomenological studies to constrain light or compressed stops, e.g. [15, 16, 22–46]. The stop is off-shell in much of this parameter space; a careful modeling of the angular distributions of the final state is needed since the kinematics can have a non-trivial impact on the resulting efficiencies. Therefore, one must abandon the narrow-width approximation [47] (depicted in the right panel of figure 2) and compute the full four-body kinematics (as illustrated in the left panel of figure 2). Here, we will follow the procedure developed in ref. [45] for simulating events including these effects.

[3]When the stop decays involve off-shell tops, the final state branching ratio for the stop can recieve non-trivial contributions from other channels as discussed in ref. [22]. Since this depends on the details of the underlying SUSY breaking flavor structure, we will ignore these subtleties and assume that the stop decays to an off-shell top and a neutralino 100% of the time.

JHEP06(2020)019

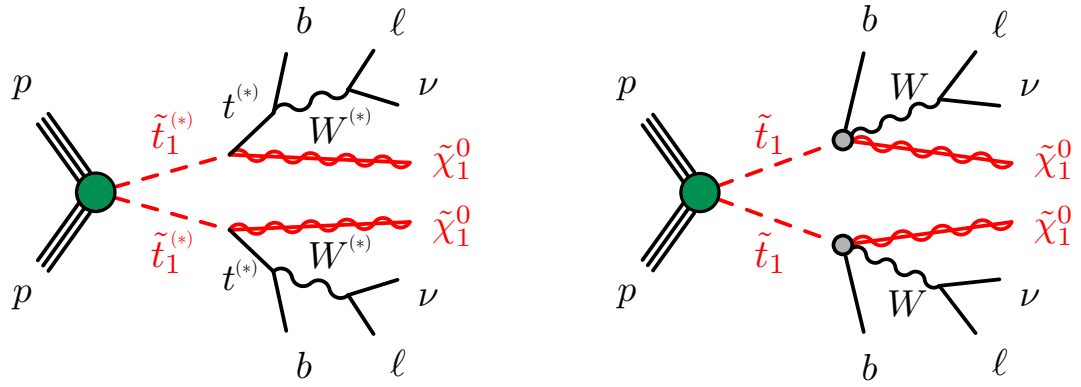

Figure 2. The left diagram illustrates the full process for the parameter space where $m(\tilde{t}_1) - m(\tilde{\chi}_1^0) < m_t$ including the off-shell propagators which encode the non-trivial kinematic correlations, and the right diagram illustrates the same process in the narrow width approximation. The green circles represent the full tree-level stop pair production matrix element, which is included in our simulations. The gray circles represent decays that do not include any matrix element information, i.e., the particles are decayed using phase space alone. The superscripts $(*)$ denote particles that can go off shell. This figure was adapted from diagrams given in ref. [48].

	All-hadronic	Di-leptonic	Semi-leptonic
$m(\tilde{t}_1)$	172.2 GeV	166.5 GeV	160.8 GeV
Bias	−0.5 GeV	−2.0 GeV	−1.3 GeV

Table 1. Summary of the maximum bias on the measured m_t due to stop contamination in each channel, assuming $m(\tilde{\chi}_1^0) = 1$ GeV. The top row shows the mass of the stop that maximally biases the experimentally measured mass from the Monte Carlo truth mass. The size of the bias in the measurement for each channel is shown in the bottom row.

The central and right panels of figure 1 provide a summary of our main results, which are described in detail in section 4. We introduce stop contamination into the recasted top mass measurements and provide a simple combination of the three channels for the 1-σ best-fit region in the $m(\tilde{t}_1)$-m_t plane. The best fit point is shown as an orange star. Two neutralino mass points are shown: $m(\tilde{\chi}_1^0) = 1$ GeV (center) and $m(\tilde{\chi}_1^0) = 20$ GeV (right) for a range of stop and top masses. The maximum bias for each channel is also summarized in table 1. The bias in the observed top mass depends on the mass of the contaminating stops, and it can be as large as 2.0 GeV in the di-leptonic channel.

Our results have an important impact on interpretations of stop exclusion in the stealth stop region. Both precision measurements and direct searches have attempted to whittle away the apparent available parameter space to a mere "splinter." An early ATLAS approach to examining this region relied on precision measurements of the top cross section [49], although only results for $m(\tilde{\chi}_1^0) = 1$ GeV were presented. This motivated our previous study [45], where we performed a careful recasting of the ATLAS exclusion to extend it into the full stop-neutralino mass plane. More recently, both ATLAS [50] and CMS [51] have exploited the clean signature and angular distributions of $e\mu$ events, nearly

JHEP06(2020)019

excluding the narrow splinter-like region. However, ref. [52] shows that observed limits on the stop mass using the $t\bar{t}$ cross section ratio at 7 and 8 TeV center-of-mass energy collisions at the LHC drop from around 180 GeV to 160 GeV if the top mass is changed from 172.5 to 175 GeV, indicating that $\mathcal{O}(1\,\text{GeV})$ shifts in the top mass can have an appreciable impact on the stop limits. In this paper we demonstrate that stealth stops can contaminate the top mass measurement at this level, which would lead one to infer that the top mass is lighter than its true underlying value. To know if we have actually closed the window on light stops, the interplay between the measured top mass and the stealth stop exclusion limits must be rigorously explored.

2 The template method

Any discussion relating the theoretical mass of a particle to an experimental observable requires care. From a quantum field theory point of view, the choice of scheme is defined by how one decides to remove the UV divergences when renormalizing perturbation theory. Two common choices yield what is referred to as the "pole" mass or the "$\overline{\text{MS}}$" mass. In the three measurements studied here, ATLAS avoids these issues and instead infers what is often called the "Monte Carlo" mass by comparing some observable that is sensitive to the top mass against Monte Carlo generator predictions as a function of the numerically implemented mass parameter. The MC top quark mass $m_{t,\text{MC}}$ is related to the field-theoretic pole mass $m_{t,\text{pole}}$ as

$$m_{t,\text{MC}} = m_{t,\text{pole}} \pm \delta m_t \,. \tag{2.1}$$

In the discrepancy $\delta m_t \sim \mathcal{O}\big(Q_0\ \alpha_s(Q_0)\big)$, Q_0 corresponds to the scale of the shower cut-off [53–57], and α_s is the strong coupling. Other studies suggest the uncertainty in this conversion is on the order of the hadronization scale [58, 59]; see ref. [60] for a study on reducing this ambiguity by means of jet grooming. We conclude that the difference is generally on the order of a few hundred MeV, which is comparable to typical modern experimental precision [61]. From here forward, we will put these issues aside and focus on the methodology employed by ATLAS — we emphasize that δm_t is another source of systematic uncertainty that must be tracked when comparing the value of m_t measured by ATLAS measurement to other approaches or as an input to a theory calculation.

In order to compare Monte Carlo predictions to data, ATLAS relies on a template method. An observable O is chosen such that it is sensitive to the top mass, and simulations are then used to compute distributions for multiple values of m_t. Clearly, the particular choice of O depends on the channel under consideration; for example, in the all-hadronic channel [12], ATLAS constructs the ratio between the 3- and 2-jet invariant masses as this minimizes sensitivity to the jet energy scale uncertainty. Samples of the distributions for O are generated over a range of values for the top mass (ATLAS does this for five m_t values from 167.5 GeV to 177.5 GeV). A set of preselection cuts are then applied, and each resulting distribution is fit with the same parametric curve. Then the resulting best-fit values are assumed to be linear functions of m_t, and an interpolation as a function of m_t is derived by linearly fitting the parameter variations as a function of m_t. This resulting

JHEP06(2020)019

object is the so-called template, which allows one to "predict" the shape of O as a function of m_t. To make this procedure more concrete, and to highlight some of its features, we work out a detailed toy example template in what follows.

2.1 A toy example

In this section, we present a toy model that illustrates how the template method works in practice. For now, we will assume that the distribution for the observable O has a characteristic peak followed by an extended tail. For concreteness, we model such a shape using a Gaussian for the peak and a Landau function for the tail, where the latter is defined as

$$P_{\text{Landau}}(x;\mu,c) = \frac{1}{\pi\, c}\int_0^\infty \mathrm{d}t\; e^{-t}\cos\left[t\left(\frac{x-\mu}{c}\right) + \frac{2\,t}{\pi}\log\left(\frac{t}{c}\right)\right], \tag{2.2}$$

where μ essentially controls the location of the peak and c controls the width of the distribution. This toy model is described by six parameters:

$$P(x;a,b,\chi,\sigma,\mu,c) = a\,P_{\text{Gaussian}}(x;\chi,\sigma) + b\,P_{\text{Landau}}(x;\mu,c)\,. \tag{2.3}$$

The parameters a and b control the relative normalizations of the Gaussian and the Landau components,[4] while χ and σ are the mean and standard deviation of the Gaussian, respectively. We choose to use a Gaussian plus a Landau as our toy distribution since this is the shape used by ATLAS for the di-leptonic measurement. The other two channels are fit to similar distributions as discussed below.

The key to choosing a good observable is that its shape (and ideally the location of a peak) must change as a function of the underlying parameter of interest — for the measurements of interest below, this parameter is the top mass, while in the toy model studied in this section, we will call this m_{toy}. We model the "truth-level" change in the underlying six parameters defined in eq. (2.3) as linear functions of m_{toy}, which are chosen to closely mimic those that ATLAS extracts from real data.

Once the observable O and the parametric model are chosen, the next step is to construct the templates. The ATLAS approach relies on Monte Carlo simulations for different choices of m_t. For our toy example, we draw samples from the truth-level probability distributions at five values of m_{toy} using the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm [62, 63]. A dataset of 10,000 elements is constructed for each choice of m_{toy}, which are subsequently binned and normalized. We then fit the resulting histograms to the distribution given in eq. (2.3). An example fit is shown in the left panel of figure 3, comparing the fitted distribution to the toy data, and the right panel displays the best-fit templates for three different values of m_{toy}.

In order to account for the statistical noise due to finite sample sizes, we generate 100 independent data sets from the truth-level distribution, and find the best fit parameters for each. The mean and the standard deviation for each of the parameters are shown as

[4] As a probability distribution, the values of a and b should be chosen such that the distribution integrates to unity. We do not enforce this constraint since we do not sample over the full allowed range of values for x, and furthermore, we found this more flexible form yields better fits.

JHEP06(2020)019

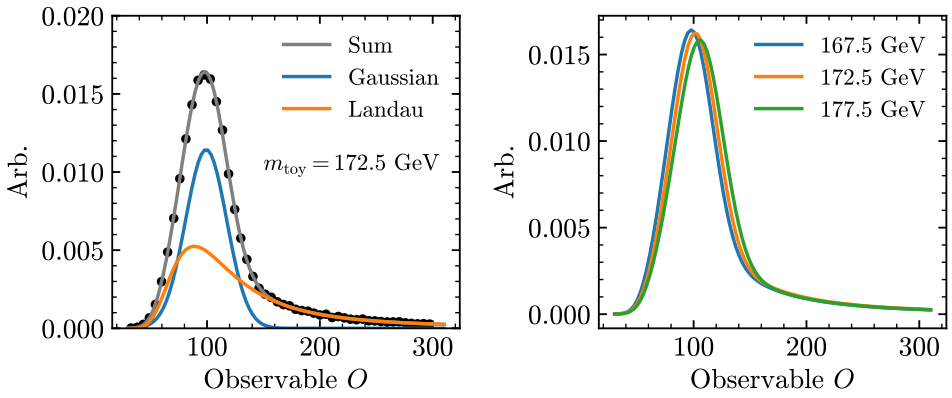

Figure 3. Our toy observable is modeled with a probability distribution which is the sum of a Gaussian for the peak and a Landau function for the tail. The gray curve in the left panel shows the fit to the black data points, while the blue and orange lines show the individual contributions of the Gaussian and Landau components, respectively. The right panel demonstrates how the shape of the observable changes as a function of m_{toy}: the location of the peak of the observable O is highly correlated with m_{toy}.

the data points with error bars in figure 4, and the linear functions are indicated by the dashed red lines. It is not surprising to see that the largest range occurs for the variable χ, since this determines location of the peak. Additionally, this parameter χ has the smallest fractional uncertainty of $\sim 1\%$, while the other parameter error bars vary from $\sim 4\%$ to as much as $\sim 8\%$, which can be traced back to its sensitivity to the position of the peak of the distribution. Each of these distributions is then fit to a line including the impact of the error bars on the fit, as shown by the blue lines in figure 4.

The final step for constructing a template is to use these linear fits as a function of m_{toy} to convert the parametric model of eq. (2.3) into a function of m_{toy} alone. Explicitly, the model becomes $P(x; a, b, \chi, \sigma, \mu, c) \to P(x; m_{\text{toy}})$, where each of the original parameters is determined by the appropriate best fit linear function of m_{toy}. Finally, one can use these templates to extract a mass measurement by fitting the template (which is now a function of the single parameter m_{toy}) to the experimentally determined distribution.

We identify two sources of uncertainty within the template method as implemented here: the first is the statistical uncertainty from using the derived template, and the second is the systematics associated with deriving the template itself. We use a closure test to assess the size of these uncertainties. An extra 100 sets of samples for a given mass point are generated; for each we find a template mass that best fits the distribution (using the templates of the blue lines of figure 4). The difference between the truth and extracted values are small, and the standard deviation gives us an estimate for the uncertainty of using the template, around 0.33 GeV for the toy model.

To measure the second source of uncertainty, which results from the assumption of the linear dependence of the template parameters, we repeat the closure test using the dotted lines of figure 4 denoting the uncertainty of the linear fits. Using the shifted template

JHEP06(2020)019

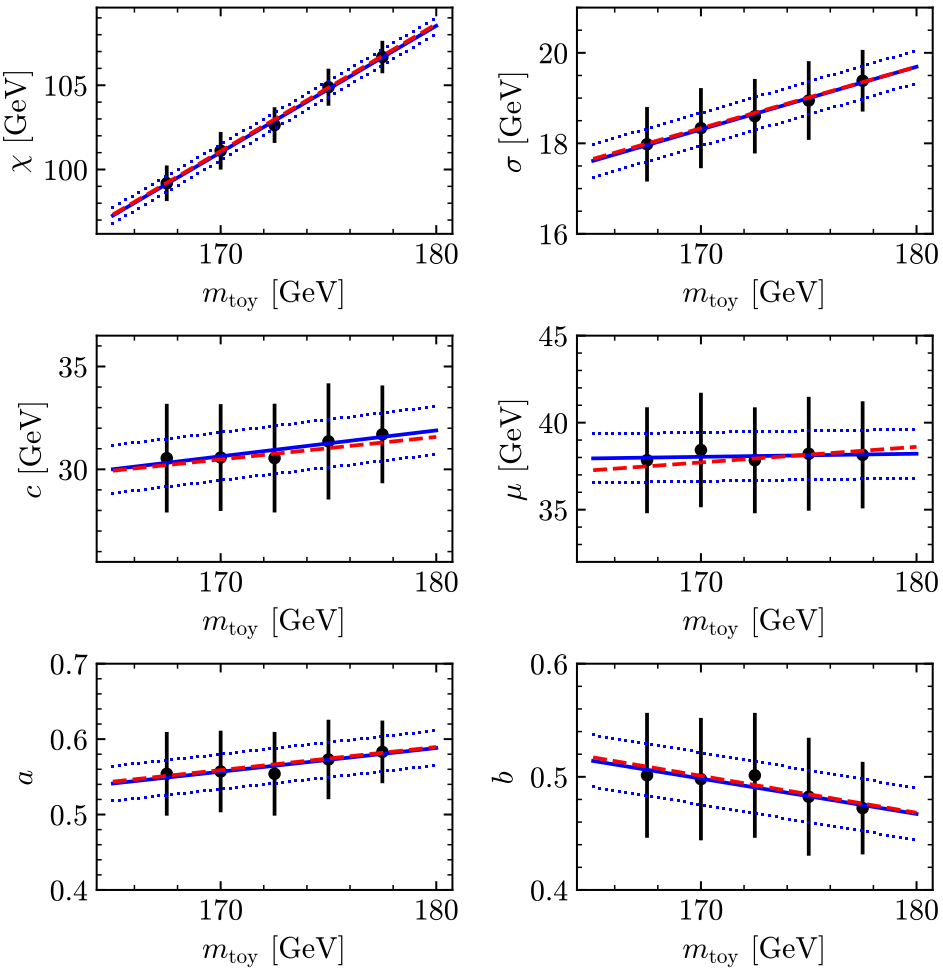

Figure 4. The toy observable is fit to the six-parameter function in eq. (2.3). We take the dependence of each parameter as a function of m_{toy} from the red dashed lines. For each value of m_{toy}, a dataset is drawn from these *truth*-level probability distributions, which is subsequently fit to the function in eq. (2.3). This process is repeated 100 times; the black dots with error bars provide the mean and standard deviation inferred for each parameter. The blue solid lines give the resultant linear fits to these points with associated errors. The upper and lower dotted lines in each panel are the result of systematically shifting the best fit line up and down by the covariance of the y-intercept as determined during the fitting procedure (see section 2.1). The envelope of dotted lines provides some insight into the systematic uncertainty associated with poorly modeled templates.

results in extracting $m_{\text{toy}} \sim m_{\text{true}} \pm 1.5\,\text{GeV}$, depending on if the upper or lower shift is considered. In the toy model, we find this systematic uncertainty from deriving the template is larger than the statistical uncertainty. When we perform the same tests for the ATLAS top mass measurements, we find that the two uncertainties are similar in size to each other and subdominant to other quoted experimental uncertainties, e.g. that come from the parton distribution functions or the jet energy scale.

JHEP06(2020)019

2.2 Dependence on the choice of fit function

The last section addressed some of the uncertainty associated with constructing a template. However, in performing those tests we used a parametric fit function that has the exact same form as the true underlying distribution. This is in contrast with the fit functions utilized by ATLAS, which are not necessarily determined from the underlying physics. As we will show here, the template method is quite robust as long as the model parameters are linearly dependent on m_{toy}, even if the model does not provide a particularly good fit to the distribution of the observable O.

To illustrate this point, we repeat the template analysis with the true data distributed according to the same toy model described by figure 4. We now fit the distributions by a Gaussian alone, which does not model the tail of the O distribution. As shown in figure 5, the best-fit Gaussian tracks the location of the peak, which is highly correlated with the underlying m_{toy}. We then generate a template in analogy with above, and perform a closure test, which yields the left panel of figure 6. As comparison, we provide the closure test result from the truth template on the right pannel. Surprisingly, the bias induced by this simple-yet-crude model for the shape of O is smaller than when we used the full model, and a similar trend is observed for the standard deviation. We conclude that although there is no a priori way to determine what parametric shape to use, the template procedure is not particularly sensitive to this choice.[5]

The fact that the template method does not require a model which accurately depicts the data can be seen as both a positive and a negative feature. On the positive side, it implies that one does not need to worry too much about the actual shape of the distribution when constructing a fit function, which is a plus since it is unknown how one might determine such shapes analytically (especially including the impact of pre-selection cuts). On the other hand, this opens the possibility that physically unmotivated observables can be used, as long as they are relatively correlated with the top mass. The fact that a good fit is not a necessary requirement for closure in the template approach implies that subtle effects could bias the final extracted value of the top mass without warning when testing the self consistency of the procedure. We will see an example of this kind of issue in section 4, where we investigate the impact of stealth stop contamination.

3 The semi-leptonic channel: a modified approach

Although the main focus of this study is to quantitatively investigate the impact that light stops could have on the measurement of the top quark mass, in this section we will critically evaluate the application of the template method to the semi-leptonic final state as currently implemented by ATLAS in ref. [13]. In particular, we showcase how defining a single observable for a pair produced event can obscure the physical interpretation of the observable. This can be corrected by a straightforward implementation of a two-dimensional template. The issue and its resolution are presented in what follows.

[5] In fact, we tried an even more radical example of assuming the shape of the parametric fit to O was simply a line, and the closure test again works out surprisingly well.

JHEP06(2020)019

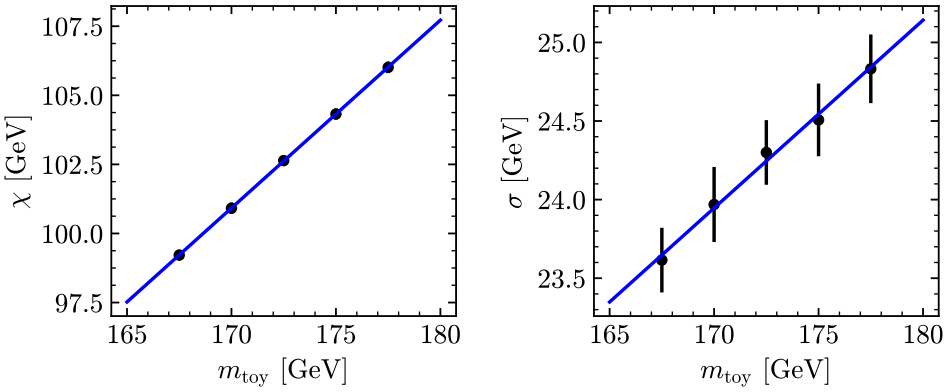

Figure 5. The results of a test using the toy model, where the parametric fitting function was assumed to be a Gaussian alone, which does not provide a good fit to the underlying O distribution. The left and right panels show the mean and standard deviations of the template Gaussian for different values of m_{toy} along with the linear fit used to generate the template. The data points and error bars come from the average and standard deviation of 100 fittings of random samples.

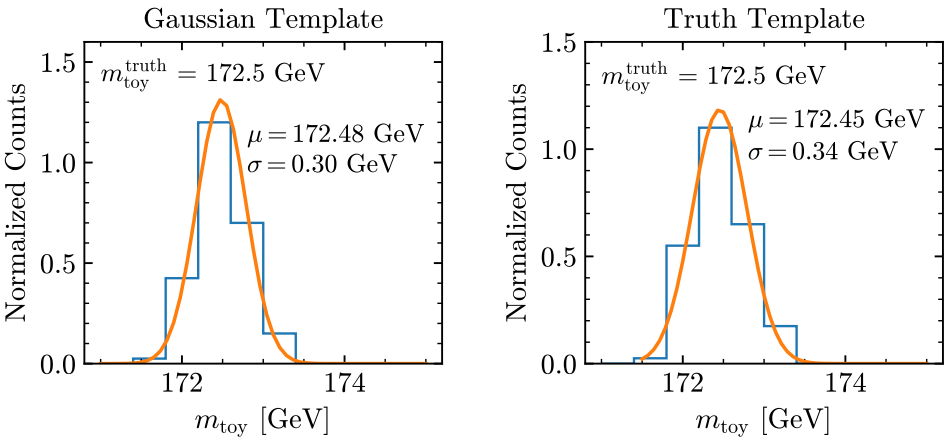

Figure 6. The left panel shows the value extracted from 100 independent data sets of $m_{\text{toy}} = 172.5\,\text{GeV}$ assuming this Gaussian based template while the right panel shows the same data with the template from eq. (2.3). The bias and uncertainty with the simpler model are smaller than when we used the *truth* model.

3.1 Pre-selection cuts

The defining characteristic of the semi-leptonic channel is that one of the top decays involves a lepton and the other decays fully hadronically. The final state of interest is then two b-jets, two light flavor jets, one charged lepton, and missing energy from the neutrino. This is a powerful channel since the QCD background is reduced due to the lepton requirement.

We recast the ATLAS measurement [13] as closely as possible. However, we encountered a subtle issue as discussed in what follows, which motivates our modified approach. The parton level events are generated using `Madgraph` [64], and are subsequently showered and hadronized using `Pythia 8` [65]. We use `Delphes` [66] to simulate detector effects,

JHEP06(2020)019

and we modified the `Delphes` detector card to match the b-tagging characteristics reported by ATLAS. More details regarding the event generation can be found in section A, and additional details and validation results for the semi-leptonic channel are given in section D.

In reconstructing objects for the analysis, we use the following set of definitions. Electron candidates are required to have a transverse momentum of $p_T > 25\,\text{GeV}$, $|\eta| < 2.47$ excluding the range (1.37, 1.52) due to the mismatch between the barrel and the end cap at ATLAS. Muon candidates must satisfy $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$. Jet candidates are reconstructed with the anti-k_t algorithm [67] with a radius of $R = 0.4$ and are required to satisfy $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$. Muons reconstructed within $\Delta R < 0.4$ of a jet candidate are considered to be part of the jet and are subsequently removed from the list of charged lepton candidates. Jet candidates are labeled as jets if they have $\Delta R > 0.2$ from all electron candidates, and otherwise they are removed. Finally, electron candidates within $\Delta R < 0.4$ of a valid jet are removed. We set a flat b-tagging efficiency of 0.7, and rejection factors of 5 and 140 for the charm quark and the light quarks, respectively.

To select events that are likely due to the semi-leptonic decay of a $t\bar{t}$ pair, the following pre-selection cuts are imposed:

- Exactly one charged lepton.
- The $\not{E}_T$and m_T^W cuts depend on the type of lepton:[6]
  - μ channel: $\not{E}_T > 20\,\text{GeV}$ and $\not{E}_T + m_T^W > 60\,\text{GeV}$.
  - e channel: $\not{E}_T > 30\,\text{GeV}$ and $m_T^W > 30\,\text{GeV}$.
- At least four jets with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$.
- Exactly two b-tagged jets.

Table 4 in section D shows the number of events that survive each of these successive cuts as predicted by our simulation.

3.2 A likelihood approach to inferring the neutrino momentum

A likelihood-based method is used to determine the missing neutrino momentum and address the combinatoric backgrounds as developed in ref. [68]. This methodology is the basis of the template approach as a function of the reconstructed top quark mass, $m_{t,\text{reco}}$, developed by ATLAS in the semi-leptonic channel as discussed in the next section. In order to recast this method, a likelihood function is built from Breit-Wigner (BW) distributions [69] defined as follows for each event that passes the preselection cuts:

$$\text{BW}\big(m \,|\, p\big) = \frac{1}{\big(p^2 - m^2\big)^2 + m^2\,\Gamma^2}\,, \tag{3.1}$$

where m is the particle mass and Γ is its width. The likelihood function is simply the product of four BWs, one for each of the two W bosons, one for each of the two top

[6] m_T^W is the transverse mass of the W and is defined as $m_T^W = \sqrt{2p_{T,\ell}\,\not{E}_T\Big(1 - \cos\phi\big(\ell, \vec{\not{E}}_T\big)\Big)}$.

JHEP06(2020)019

quarks:

$$L\big(m_{t,\mathrm{reco}}, m_{W,\mathrm{reco}}, p_{z,\nu} \,\big|\, p_{b_1}, p_{b_2}, p_{q_1}, p_{q_2}, p_\ell, \not{E}_T\big) = \mathrm{BW}\big(m_{t,\mathrm{reco}} \,\big|\, p_{b_1} + p_{q_1} + p_{q_2}\big) \times \mathrm{BW}\big(m_{W,\mathrm{reco}} \,\big|\, p_{q_1} + p_{q_2}\big) \times \mathrm{BW}\big(m_{t,\mathrm{reco}} \,\big|\, p_{b_2} + p_\ell + \not{E}_T + p_{z,\nu}\big) \times \mathrm{BW}\big(m_{W,\mathrm{reco}} \,\big|\, p_\ell + \not{E}_T + p_{z,\nu}\big)\,, \quad (3.2)$$

where $p_{b_{1,2}}$ are the four momenta of the two b-jets, $p_{q_{1,2}}$ are those of the untagged jets, and p_ℓ is the lepton four momentum. ATLAS additionally includes transfer functions in the likelihood, $W(p_{\mathrm{measured}}|p_{\mathrm{true}})$, for each jet, the lepton, and the missing energy. These transfer functions parameterize the mapping between the momenta of the detector-level objects and the momenta of the initial partons.[7] The effect of the transfer functions will be discussed in more detail below.

The inputs to eq. (3.2) are the lepton momentum, the missing transverse momentum, and the momenta for up to six jets. The x and y components of the neutrino momentum are assumed to be equal to the missing energy components. The z component, $p_{z,\nu}$ is unmeasurable at the LHC, and is therefore treated as a free parameter when maximizing the likelihood function, where the initial value provided to the maximizer is derived from $m_W^2 = (p_\ell + p_\nu)^2$. If the solutions of $p_{z,\nu}$ are complex, then the initial guess for the maximization is set to $p_{z,\nu}^{\mathrm{init}} = 0$. If there are two real solutions, then the solution resulting in the largest likelihood is used. The likelihood is then maximized for all possible assignments of the b-tagged jets to the leptonic side of the event, and all choices of two out of the possible four un-tagged jets. The choice which maximizes the likelihood is then taken to determine the assignment of decay products for both hadronic and leptonic tops. We have additionally checked that this approach does a reasonable job of reproducing the truth level assignments of final states with the appropriate top, and that it tends to find a very good approximation for the z-component of the neutrino momentum, as expected.

3.3 The ATLAS semi-leptonic template

After selecting events using the preselection cuts described above in section 3.1, ATLAS applies the likelihood method introduced in section 3.2. This provides a systematic way of assigning final state objects to either of the two top candidates, which is then used to construct a three dimensional template as a function of m_t, the jet energy scale (JES), and the b-JES. This is done by fitting to three observables O, $m_{t,\mathrm{reco}}, m_{W,\mathrm{reco}}$, and R_{bq}, where

$$R_{bq} = \frac{p_T^{b,\mathrm{had}} + p_T^{b,\mathrm{lep}}}{p_T^{q_1} + p_T^{q_2}}\,, \quad (3.3)$$

where $p_T^{b,\mathrm{had}}$ and $p_T^{b,\mathrm{lep}}$ are the momenta of the b-jets assigned to the hadronic and leptonic sides of the even respectively, and $q_{1,2}$ are the light flavor jets that are associated with the decay of the W.

[7] Note that these transfer functions are distinct from the jet energy scale.

JHEP06(2020)019

ATLAS finds that $m_{W,\text{reco}}$ largely constrains the JES, while R_{bq} constrains the b-JES relative to the JES. Given that our analysis relies on a simple parametrized detector simulation, we are not equipped to perform a realistic study of the impact of varying the JES or b-JES. Critically, we find that our $m_{W,\text{reco}}$ and R_{bq} distributions agree relatively well with those provided by ATLAS, as shown in section D in the left and center panels of figure 18. Therefore, we are confidant that the JES and b-JES dependence will not have a significant impact on our interpretation of the semi-leptonic mass measurement.

The extraction of the Monte Carlo top quark mass comes mostly from the $m_{t,\text{reco}}$ distribution. ATLAS finds that $m_{t,\text{reco}}$ peaks at values much lower than the top mass that is extracted. We find that we cannot reproduce the ATLAS distribution, in part because we neglect the transfer functions in the likelihood and ATLAS additionally uses a BDT to select events. This is shown in section D in the right panel of figure 18.

3.4 Impact of the transfer functions

It is surprising that ATLAS finds that the distribution for $m_{t,\text{reco}}$ peaks below the actual top mass. It is important to emphasize that the location of the peak for ATLAS *is not* the extracted value, which comes from finding the *best-fit template*. We reiterate that as shown in section 2, as long as the shape of the template varies linearly with the generator mass, the template procedure will close and the extraction of the best fit is expected to be robust. Despite this fact, this section is devoted to explaining the unexpected $m_{t,\text{reco}}$ distribution. Along the way, we will argue that $m_{t,\text{reco}}$ is not physically meaningful, which will motivate a physics-driven proposal for a modified approach presented in section 3.5.

In order to generate their distribution, ATLAS populates a histogram using the value of $m_{t,\text{reco}}$ that maximizes the likelihood function in eq. (3.2) for each event. The underlying assumption is that $m_{t,\text{reco}}$ captures the best fit top mass for the whole event. However, this is not always the appropriate interpretation of this variable, as can be made clear by studying the form of the likelihood function without the transfer functions, as we do now.

As discussed above, the likelihood-based approach provides a way to systematically assign the final state objects to the two top quarks in the event, while also solving for the z-component of the neutrino momentum. Assuming one has made all of these choices such that the maximum value for the likelihood can be achieved, we are left with a simple function of $m_{t,\text{reco}}$:

$$L \sim \frac{1}{\left[\left(\sum p^{\text{had}}\right)^2 - m_{t,\text{reco}}^2\right]^2 + m_{t,\text{reco}}^2\,\Gamma^2} \times \frac{1}{\left[\left(\sum p^{\text{lep}}\right)^2 - m_{t,\text{reco}}^2\right]^2 + m_{t,\text{reco}}^2\,\Gamma^2}\,, \tag{3.4}$$

where $\sum p^{\text{lep}}$ and $\sum p^{\text{had}}$ are the sum of the four momenta for the final states assigned to the leptonic and hadronic tops respectively, and the width Γ is set to the PDG value, 1.41 GeV.

The choice to use a Breit-Wigner shape when constructing the likelihood function that peaks at the best fit mass of the top quark is clearly physically motivated. However, while the product form in eq. (3.2) works very well as an approach to the combinatoric background and for determining p_z for the neutrino, it does not return an event-level "best fit" for the

JHEP06(2020)019

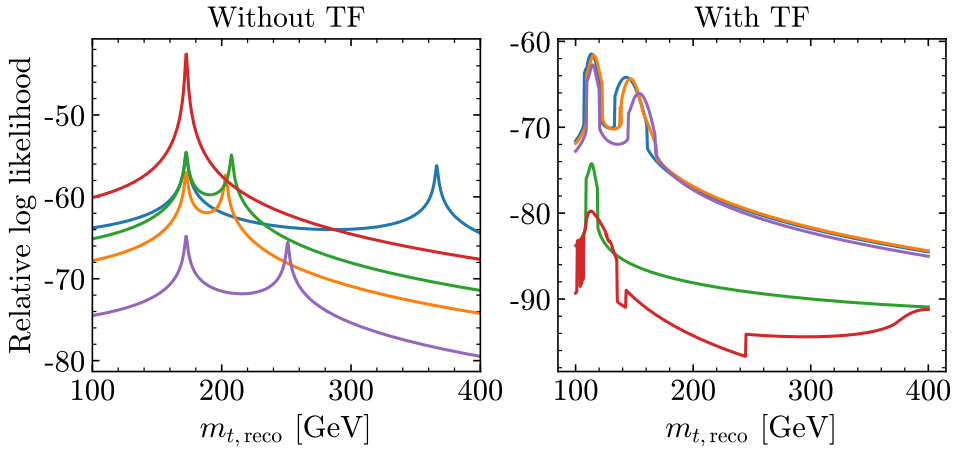

Figure 7. The left panel shows double peak structure of the log likelihood function for 5 random events at truth level, without including the transfer functions. A clear double peaked structure occurs for many events. As expected, the lower mass peak is always more likely than the higher mass one. The right panel shows the impact of including the transfer functions in the likelihood and optimizing over the z-component of the neutrino momentum for the same five events, as implemented in KLFitter. Although the transfer functions push the peaks together, they do not always result in a likelihood that has a single local maximum. Note that the absolute scale of the log likelihood is expected to differ between the two panels.

top mass. In particular, using the simplified expression in eq. (3.4), it is straightforward to see that the likelihood shape has two very sharp peaks, one for each choice of $m^2_{t,\mathrm{reco}}$ that equals $\left(\sum p^{\mathrm{had}}\right)^2$ and $\left(\sum p^{\mathrm{lep}}\right)^2$, rather than one peak between the two. This point is clearly illustrated in the left panel of figure 7, where we evaluate eq. (3.4) as a function of $m_{t,\mathrm{reco}}$ for five independent top pair production events (without the transfer functions) taken from the KLFitter semi-leptonic example file.[8] Note that the left peak will always be more likely than the right one; this is clear from the denominator of the Breit-Wigner.[9]

The transfer functions are implemented in part to help account for the two peaks, by having the ability to adjust the jet momenta to push the two tops to the same p^2. There are two minor issues with this. The first is that off-shell effects are relevant, and even at the partonic, truth-level, the two tops rarely have the same p^2. The second is related to the fact that the Breit-Wigner function has higher likelihoods for lower invariant mass peaks, which allows the transfer functions to consistently push $m_{t,\mathrm{reco}}$ to smaller values. The right

[8] https://github.com/KLFitter/KLFitter.

[9] ATLAS states that they are using a "Breit-Wigner" distribution, but they do not specify its functional form. If they had chosen to include the numerator factor

$$k = \frac{2\sqrt{2}\, m\, \Gamma \sqrt{m^2\left(m^2+\Gamma^2\right)}}{\pi \sqrt{m^2+\sqrt{m^2\left(m^2+\Gamma^2\right)}}},$$

as the distribution is sometimes given, the logic changes and the right peak is always more likely. We note that the ROOT implementation [70, 71] of the Breit-Wigner function has unit numerator.

JHEP06(2020)019

panel of figure 7 shows the likelihood when using the transfer functions for the same events as the left panel. The sharp double peaks are softened — and sometimes combined into one — but the peaks are also shifted to lower masses.

The "best-fit" momenta maximizing the likelihood are not physical. The procedure results in $m_{t,\text{reco}}$ well below the actual top mass; because of this, ATLAS uses the measured momenta for $m_{W,\text{reco}}$. Part of the motivation for the transfer functions was to combine the two tops into a single observable, at the expense of reducing the available information. Without the transfer functions, we find the value of $m_{t,\text{reco}}$ that maximizes the likelihood corresponds to the peak associated with the leptonic top $\sim 60\%$ of the time. This implies that the $m_{t,\text{reco}}$ distribution is a non-trivial mixture of hadronic and leptonic tops, with unknown implications for systematic effects on the distribution of $m_{t,\text{reco}}$. This motivates our proposal for a modified approach using independent information from both tops, which is presented in the next section.

3.5 A two-dimensional mass extraction template

Instead of using a one-dimensional template for $m_{t,\text{reco}}$, one would prefer an approach that takes advantage of the fact that there is both a leptonic and hadronic top decay in each event. We propose a modified approach in this section, relying on the same combined likelihood given in eq. (3.2) to control the combinatorics and to solve for the missing neutrino momentum. We use the configuration that maximizes the likelihood to generate the distribution shown in figure 8, where we give a two-dimensional density plot of the hadronic and leptonic top masses that result. One observes that the density is essentially symmetric about the diagonal $m_t^{\text{had}} = m_t^{\text{lep}}$, and that most of the time the values of the top masses from the two sides of the event are very similar. Along the diagonal, the density peaks near $m_t \sim 170\,\text{GeV}$ and then has an extended tail to larger masses. This 2D plane provides an excellent candidate for an improved observable O from which we construct a template.

For our parametric model, we want a function with a peak and a tail along the diagonal. We chose this to be a Gaussian plus a Landau function, following the ATLAS approach used to fit a one-dimensional $m_{t,\text{reco}}$ distribution. Then we model the spread orthogonal to the diagonal using a second independent Gaussian. Concretely, the two-dimensional template is

$$\begin{aligned} P\big(x,y;a,b,\chi_1,\sigma_1,\mu,c,\chi_2,\sigma_2\big) = &\Big(a\,P_{\text{Gaussian}}\big(x;\chi_1,\sigma_1\big) + b\,P_{\text{Landau}}\big(x;\mu,c\big)\Big) \\ &\times P_{\text{Gaussian}}\big(y;\chi_2,\sigma_2\big)\,, \end{aligned} \tag{3.5}$$

where x and y are the distance along the diagonal and distance away from the diagonal, respectively. This is a relatively crude model for the distribution shown in figure 8, and it does not take into account how the spread away from the diagonal changes as a function of the distance from the origin. We tested that a more precise fitting function did not lead to improved extractions of the Monte Carlo top mass, while drastically increasing the computational time to perform the two-dimensional fit. This makes sense given the discussion regarding the sensitivity of the template approach to the shape of the fit function, as discussed in section 2.2.

JHEP06(2020)019

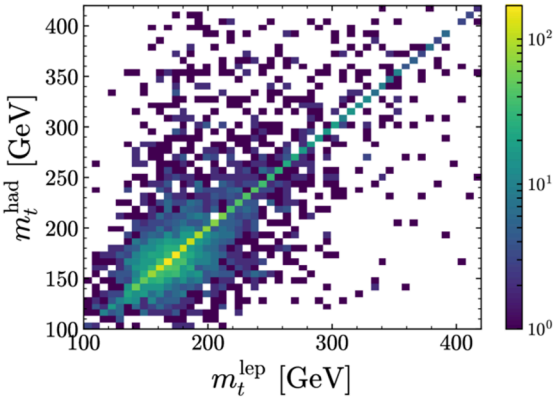

Figure 8. This figure provides a density plot of the reconstructed top mass for the semi-leptonic channel in the hadronic mass versus the leptonic mass plane. For the underlying simulations, the Monte Carlo truth mass is taken to be m_t=172.5 GeV. In the majority of events, the two masses are highly correlated and lie along the diagonal centered around the truth value of m_t. This shape motivates the form of the fitting function we are proposing that can be used to build a template for extracting the top mass in the semi-leptonic channel.

As in our toy model, we perform a closure test to validate the proposal of extracting the top mass from the two-dimensional template. Our new approach faithfully extracts the correct mass, and comes with a relatively small statistical uncertainty $\sim$ 0.1 GeV. We additionally checked that varying the linear fit of the templates up and down by an amount determined by the covariances led to a similar size uncertainty; see section 2.1 for a discussion of this test. Our determination of these sources of uncertainty due to the template method are subdominant to the JES uncertainties provided by ATLAS in [13].

With this modified procedure in hand, we are now ready to assess the impact of stop contamination on the top mass measurement. As we will emphasize below, our results in the semi-leptonic channel use the two-dimensional template method discussed here. As such, the results in the semi-leptonic channel are *not* a recasting, but can instead be interpreted as an estimate for how much contamination one could expect for this final state.

4 The impact of light stops

Now that we have explored the template method as it is used by ATLAS to extract the Monte Carlo mass of the top quark (along with our modified approach in the semi-leptonic channel), we will turn to the impact of stop contamination on the template mass extraction. This is important since attempts by ATLAS [49, 50, 72] to exclude the stealth stop region of parameter space utilizing properties of high purity $t\bar{t}$ samples assume the top mass is measured in an orthogonal channel. As we will show in this section, light stops can bias the extracted top mass by up to 2 GeV. This implies that any limit which claims to exclude stealth stops using aspects of the top pair kinematics must simultaneously account for the impact on the top mass measurement. While this may not seem like a major issue at first glance, we emphasize that the leading order cross section prediction for $t\bar{t}$ production at

JHEP06(2020)019

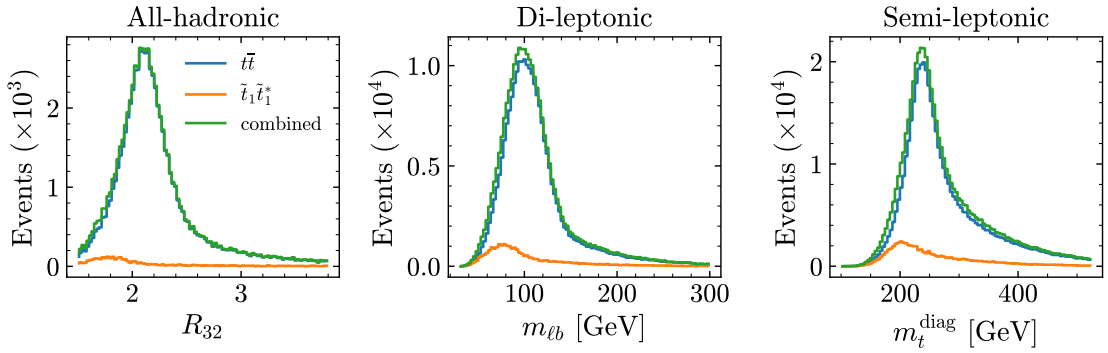

Figure 9. The distributions that are used as input to the template procedure. Moving from left to right shows all three channels for the $t\bar{t}$ sample (blue), the $\tilde{t}_1\tilde{t}_1^*$ sample (orange), and the combination (green), where the masses are taken to be $m_t = 172.5\,\mathrm{GeV}$ and $m(\tilde{t}_1) = 164\,\mathrm{GeV}$ with $m(\tilde{\chi}_1^0) = 1\,\mathrm{GeV}$. Stop contamination tends to shift the distributions slightly to the left.

$\sqrt{s} = 8\,\mathrm{TeV}$ drops from 160 pb for $m_t = 172\,\mathrm{GeV}$ to 150 pb for $m_t = 174\,\mathrm{GeV}$. Since the production of stealth stops is $\mathcal{O}(10\,\mathrm{pb})$, this could easily impact the boundaries of exclusion regions. This sensitivity to m_t has been demonstrated by ATLAS in ref. [52] where observed limits on the stop mass drop from around 180 GeV to 160 GeV if the top mass is changed from 172.5 to 175 GeV. Here, we will focus on demonstrating the quantitative impact of this contamination — assessing how this alters limits is left for future work.

For concreteness, we will work with the stop-neutralino Simplified Model framework. Given a choice of top mass, we then generate a suite of events for different values of the stop mass (and for two benchmark choices of the neutralino mass), including the full effects of the off-shell propagators following the procedure detailed in ref. [45]; more details regarding the event generation can also be found in section A.2. In particular, this approach self consistently computes the width of the stop and the top quark as the parameter space is varied. The pair production of stops is determined by its QCD interactions, and its subsequent decay $\tilde{t}_1 \to t^{(*)}\,\tilde{\chi}_1^0$ yields a (potentially off-shell) top quark and missing energy. Intuitively, the biggest impact on the top mass measurement will occur in the parameter space where the top that results from the stop decay is off-shell, since the reconstructed "top" in such events will have a "mass" that is smaller than m_t. We will see exactly this behavior in the quantitative results that follow.

To get a sense of the impact that stealth stops can yield, figure 9 shows the shape of the potential stop contribution to the observable O used to generate the template for each channel from top pair production with $m_t = 172.5\,\mathrm{GeV}$ (blue solid), stop pair production (orange solid) with $m(\tilde{t}_1) = 164\,\mathrm{GeV}$, $m(\tilde{\chi}_1^0) = 1\,\mathrm{GeV}$, and the combined distribution (green solid). Note that in the semi-leptonic channel, we use the two-dimensional observable introduced in section 3.5, but plot the one-dimensional slice along the $m_t^{\mathrm{diag}} \equiv m_t^{\mathrm{had}} = m_t^{\mathrm{lep}}$ diagonal. Each of these distributions are normalized using the production cross section times efficiency to pass the relevant pre-selection cuts, assuming an integrated luminosity of $\mathcal{L} = 20.2\,\mathrm{fb}^{-1}$. While the stop contribution is clearly subdominant, it peaks at a slightly

JHEP06(2020)019

lower value in each observable than $t\bar{t}$. This has the effect of biasing the combined sample such that the extracted Monte Carlo top mass that best fits the combined distribution is lower than the true value of m_t.

The results of our study for all three channels are presented in two different ways: the first representation is provided in figure 10, and the second is in figure 11.[10] The colored horizontal lines in figure 10 denote the mass that is extracted using the pure SM sample for true top masses of 167.5, 170, 172.5, and 175 GeV, respectively, as one moves from the bottom to top of each panel. The thick black lines show contours of constant reconstructed top mass when using the templates made using only $t\bar{t}$. Values of the contours are chosen to match the closure test values for the pure $t\bar{t}$ samples. The dotted diagonal line shows the kinematic boundary where $m(\tilde{t}_1) = m(\tilde{\chi}_1^0) + m_t$. Left of this dotted kinematic boundary, the tops are off-shell so the black lines are above the horizontal benchmark lines. This implies that the truth-level top mass (shown on the y-axis) is larger than the reconstructed value when using a SM only template. As the stops are taken to be heavier and cross the dotted line, two effects become important: the stops decay to on-shell tops removing the off-shell effects, and the stop production cross section decreases, thus explaining why the results asymptote to the pure SM in this limit.

The top row of figure 10 shows the results for the all-hadronic channel (for more details on the recast procedure for this channel see appendix B). For $t\bar{t}$ production, this channel does not result in any intrinsic missing energy, and the preselection cuts do not make any requirements on $\not{E}_T$. This implies that the distribution utilized for the all-hadronic channel is less sensitive to presence of the additional $\not{E}_T$ due to the final state neutralinos. Therefore, this channel is relatively insensitive to stop contamination; off-shell effects (left of the dotted line in the top left panel of figure 10) yield the dominant impact on the top mass extraction.

The middle row of figure 10 shows the results for the semi-leptonic channel. Due to our issues validating this channel as discussed in section 3 above, we have performed this analysis using our proposed 2D template approach. As with the all-hadronic case, when the stops are lighter (to the left of the blue dashed line) they decay through an off-shell top quark, which biases the templates to extract lower masses. However, there is an additional important effect, which makes the results in this channel even more striking. The SM contribution contains a neutrino, and so the pre-selection cuts explicitly rely on $\not{E}_T$. Furthermore, the likelihood procedure utilized for addressing the combinatoric background and the missing z-component of the neutrino momentum assumes that the measured $\not{E}_T$corresponds to the transverse components of the neutrino momentum. This implies that the neutralinos in the final state will have a non-trivial impact on the shape of the observable used for the template method. From the figures, it is clear that the impact of stealth stop contamination on the semi-leptonic channel is more dramatic than in the all-hadronic channel, yielding a bias as large as ~ 2 GeV. For the largest stop masses, the reconstructed top mass over-shoots the true value in this range due to the effect of the neutralinos on the observable, but eventually asymptotes to the SM-only true value.

[10]Both plots show contours of the mass which would be extracted from templates made using the SM only assumption when the real data is contaminated by stops. The contours are made in the same way for both plots, but different contours are shown to highlight different aspects of the contamination.

JHEP06(2020)019

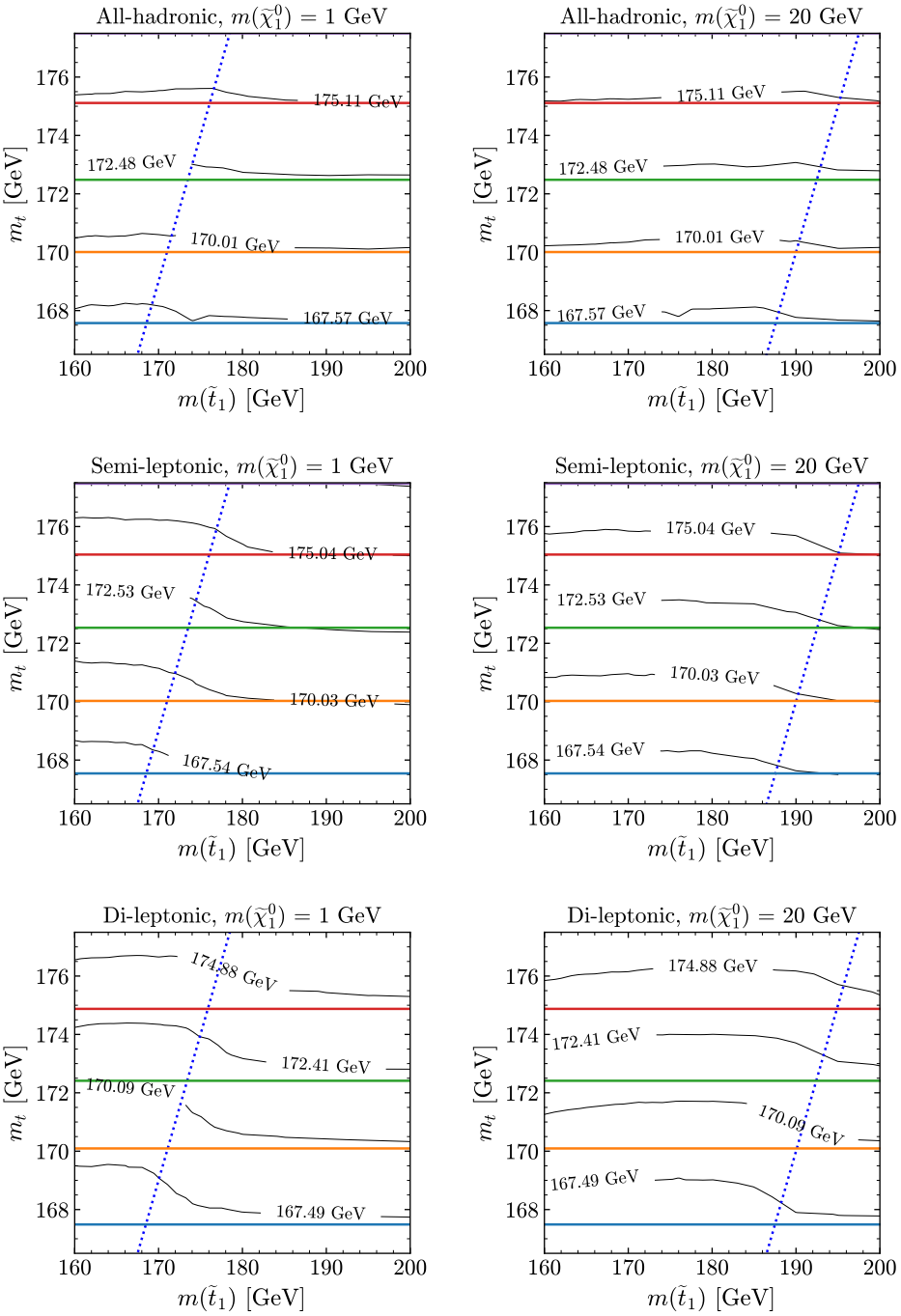

Figure 10. Effects of stop contamination on m_t compared to pure top simulations. The top, middle, and bottom rows correspond to the all-hadronic, semi-leptonic, and di-leptonic channels, respectively. The plots in the left column correspond to the choice $m(\tilde{\chi}_1^0) = 1\,\text{GeV}$, while those in the right column are for $m(\tilde{\chi}_1^0) = 20\,\text{GeV}$. In each panel, the colored lines represent the m_t reconstructed from pure $t\bar{t}$ benchmark samples. The blue dotted line indicates the kinematic boundary where the tops from the stop decays start to be produced on-shell. The black curves are contours of constant reconstructed top mass when using SM only templates. The specific contours are chosen to match the SM only closure tests.

JHEP06(2020)019

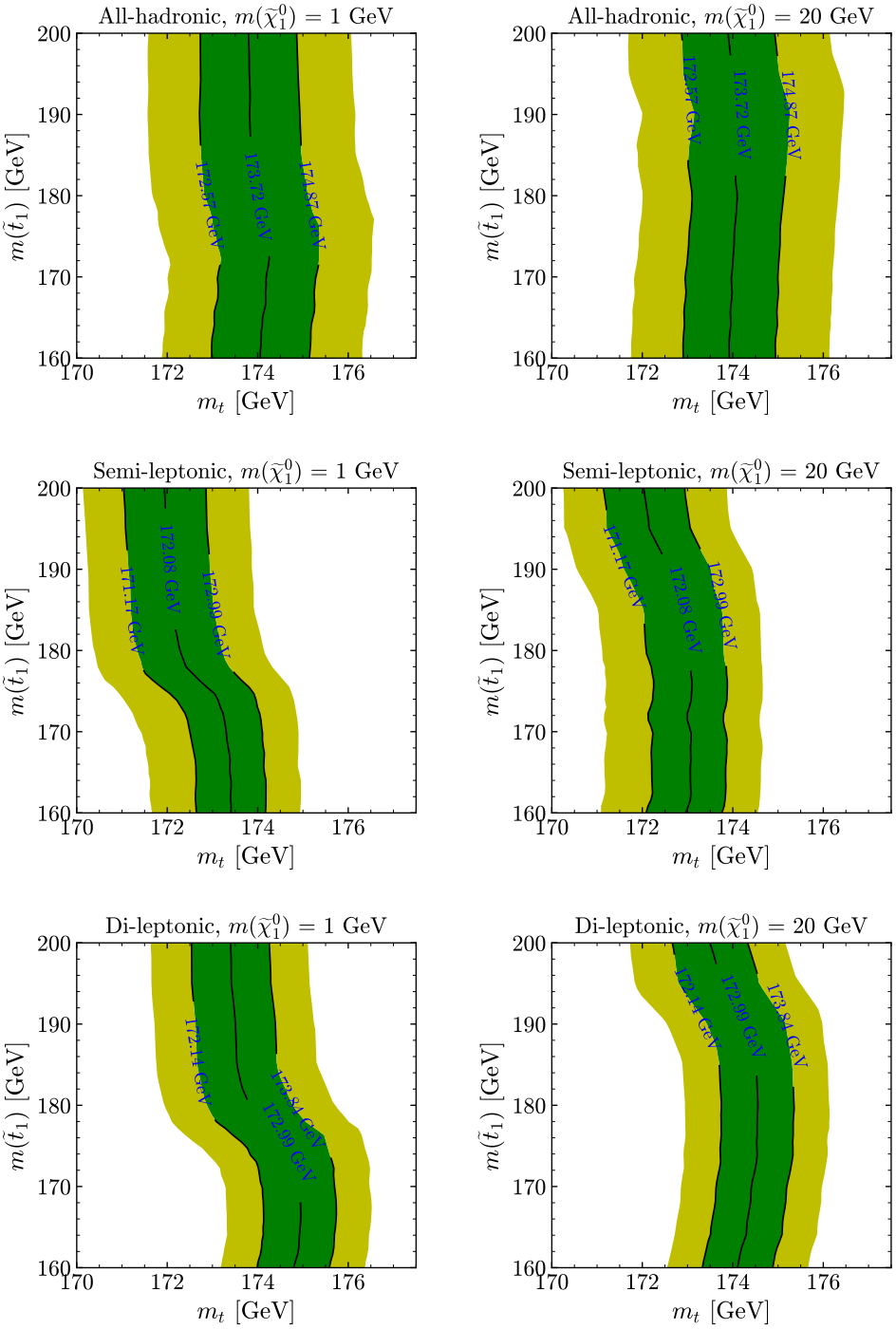

Figure 11. Contours of the mass extracted when using templates constructed in the SM only assumption. The top, middle, and bottom rows correspond to the all-hadronic, semi-leptonic, and di-leptonic channels, respectively. The plots in the left column correspond to the choice $m(\tilde{\chi}_1^0) = 1\,\text{GeV}$, while those in the right column are for $m(\tilde{\chi}_1^0) = 20\,\text{GeV}$. In each panel, the central line denotes the value measured by ATLAS in the given channel. The green (yellow) band illustrates the regions in parameter space where the reconstructed mass is within 1-σ (2-σ) of the measurement using the reported uncertainty.

JHEP06(2020)019

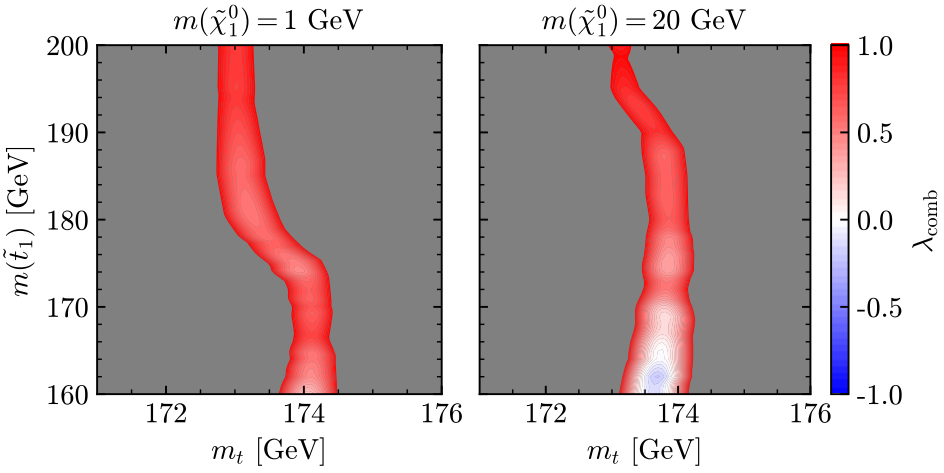

Figure 12. Values of the test statistic given in eq. (E.1), which is designed to compare the goodness of fit for the SM alone versus the BSM scenario studied here. The red regions provide a fit which is within 1σ of the SM alone hypothesis, the white regions give the same quality of fit as the SM, and the blue regions give a better fit than the SM alone. The overall best fit is at $m_t = 173.7\,\text{GeV}$ and $m(\tilde{t}_1) = 162.0\,\text{GeV}$ for the panel with $m(\tilde{\chi}_1^0) = 20\,\text{GeV}$; the fit to the top mass measurement at this point is only 0.1σ better than can be achieved with the SM only hypothesis.

The di-leptonic channel is shown in the bottom row of figure 10. In this case, the SM final state contains two neutrinos, and so the preselection includes a cut on $\not{E}_T$. As opposed to the semi-leptonic case, the observable used in this channel is simply the invariant mass of the lepton and b-jet pairs $m_{\ell b}$, and so the distribution should not be impacted by new sources of $\not{E}_T$. However, $m_{\ell b}$ it is not fixed by the mass of a parent particle, and so the resulting distribution is more sensitive to details such as the spin of the top quarks and the kinematics of the top pairs. This explains why the bias in the reconstructed top mass for this channel is the most dramatic of the three, including the fact that the result asymptotes to the SM value even more slowly as the stop mass increases.

Now that we have a sense of how large the bias from stealth stop contamination can be, figure 11 illustrates the consistency of the BSM parameter space with the observations performed by ATLAS. In this figure, the axes have been rotated with respect to figure 10, and we plot the *truth-level* Monte Carlo top mass used to generate events along the horizontal axis, while the input stop mass is on the vertical axis. For each point in the truth parameter space, we extract the reconstructed top mass. The black line at the center of the bands denotes the parameters that yield a reconstructed top mass which is equal to the value observed by ATLAS in each channel assuming the SM alone, while the green (yellow) bands are the 1-σ (2-σ) uncertainties taken directly from the ATLAS papers [12–14]. This allows one to visualize the non-trivial shapes that result from stealth stop contamination in each channel, and provides some insight into what parameter choices could yield the best consistency.

To quantitatively explore the consistency between the three channels, we performed a naive combination of these channels in the BSM parameter space; the methodology is

JHEP06(2020)019

described in section E and the results are provided in the middle and right panels of figure 1. At each point in the m_t-$m(\tilde{t}_1)$ plane (keeping $m(\tilde{\chi}_1^0)$ fixed), we compute the χ^2 for the extracted template mass in each channel as compared to the observations. The orange stars in the middle and right panels of figure 1 show the best-fit point, and the shaded area shows the 1-σ region. For the model with $m(\tilde{\chi}_1^0) = 1\,\text{GeV}$, the best fit point is found to be $m_t = 174.0\,\text{GeV}$ and $m(\tilde{t}_1) = 160\,\text{GeV}$,[11] and when $m(\tilde{\chi}_1^0) = 20\,\text{GeV}$, the best fit point is at $m_t = 173.7\,\text{GeV}$ and $m(\tilde{t}_1) = 162.0\,\text{GeV}$. Both panels show that the data fit best using lighter stops; the 1-σ uncertainty band for the right panel does not even extend to the top of the panel. In addition, the entire region results in masses larger than the SM-only assumption. If there are light stops, we may not know the mass of the top quark as accurately as we think we do.

As an amusement, we note that the all-hadronic and semi-leptonic uncertainty bands only slightly overlap in the SM alone assumption. With that, it may be possible that light stops could improve the consistency of the experimental results. In order to naively explore the extent to which the BSM model is a better fit than the SM alone, we compute the test statistic defined in eq. (E.1) and the result is presented in figure 12. While there is no particular overall improvement in the fit for $m(\tilde{\chi}_1^0) = 1\,\text{GeV}$, the heavier choice $m(\tilde{\chi}_1^0) = 20\,\text{GeV}$, does has a very mild preference for the BSM scenario. Although we simply take this to be a coincidence given the current state of the top mass measurement, it does demonstrate that if the top mass measurements became discrepant between the different channels, light stops could bias the mass measurements enough to provide a resolution.

5 Conclusions

In this work, we have investigated the stealth stop contamination of the $t\bar{t}$ sample that can potentially bias the measurements of the top mass at ATLAS by up to 2 GeV. Three decay channels are studied in detail: all-hadronic, di-leptonic, and semi-leptonic. The top mass measurement in the all-hadronic channel is the least sensitive to stop contamination, while the di-leptonic channel is the most sensitive. The combination of results suggests that the heavy neutralino case is slightly favored in terms of overall consistency among the three channels. Furthermore, we have proposed a modified method to better measure the top mass in the semi-leptonic channel.

While our focus here was on the template method, there are many approaches that could be used to measure the top mass, which could respond differently to stop contamination. For example, CMS has published a measurement of m_t using $5\,\text{fb}^{-1}$ of 7 TeV data utilizing the $m_{\ell b}$ endpoint in the di-leptonic channel [73]. Since this is a kinematic feature (as opposed to relaying on the shape of the entire distribution) the relevant kink is unlikely to be modified by the presence of light stops. It would be interesting to investigate how contamination would impact non-template based approaches like this one.

[11] Note that the best fit occurs at the lower edge of the $m(\tilde{t}_1)$ range simulated here; to find the true best fit, the region should be extended. However, the purpose of this study is to show the effect of stop contamination, which is clearly demonstrated, as opposed to taking any particular improvement in the fit seriously.

JHEP06(2020)019

In conclusion, $\mathcal{O}(1\,\text{GeV})$ shifts in the top mass measurement due to stop contamination are possible and can have $\mathcal{O}(10\,\text{GeV})$ impacts on the stealth stop exclusion limits [52]. Thus, we advocate that the LHC experiments perform an analysis of the full three-dimensional Simplified Model parameter space spanned by $m(\tilde{t}_1)$-m_t-$m(\tilde{\chi}_1^0)$ in order to make a definitive statement on the potential existence of stealth stops.

Acknowledgments

The authors thank Walter Hopkins for useful comments on the manuscript. We are particularly grateful to Merve Karacasu, Andrea Knue, Richard Nisius, and Javier Pena for discussions on the semi-leptonic channel. TC and PZ are supported by the U.S. Department of Energy under grant number DE-SC0018191. SM is supported by the U.S. Department of Energy under grant numbers DE-SC0012008 and DE-SC0017996. BO is supported by the U.S. Department of Energy under grant numbers DE-SC0018191, DE-SC0011640, and DE-SC0013607. This work utilized the University of Oregon Talapas high-performance computing cluster.

A Event generation

A.1 Top event generation

The 8 TeV $t\bar{t}$ sample is generated at the parton level using `MadGraph5_aMC@NLO 2.6.1` [64], and is passed to `Pythia 8.2` [65] for showering and hadronization. Detector effects are approximated using `Delphes 3.4.1` [66], which relies on `Fastjet` [74, 75] to cluster the jets with the anti-k_T algorithm [67]. We use the default `Delphes` ATLAS card, except that the b-tagging efficiency is set to be 0.57 for all-hadronic channel and 0.7 for the other two channels, in accordance with ATLAS [12–14]. We generated 5 million events for each of 5 top masses: 167.5 GeV, 170 GeV, 172.5 GeV, 175 GeV, 177.5 GeV.

A.2 Stop event generation

We work with a stop-neutralino Simplified Model, where the stop has the couplings appropriate for being right-handed. To cover the stealth stop region, events are generated for two choices of $m(\tilde{\chi}_1^0)$: 1 GeV and 20 GeV, and for a range of stop masses: $m(\tilde{t}_1)$ from 160 GeV to 180 GeV in steps of 2 GeV, and $m(\tilde{t}_1)$ from 180 GeV to 200 GeV in steps of 5 GeV. At each parameter point, we use `MadGraph5_aMC@NLO` to calculate the stop decay width. One must be very careful to account for all finite width effects during the generation of events when the top can be off-shell, see [45] for a detailed discussion. To this end, we ensure that the top and W widths are defined consistently for the decay and production in `MadGraph5_aMC@NLO`. Given the appropriate widths, we again use `MadGraph5_aMC@NLO` to calculate the matrix elements and generate 500,000 events for stop production and subsequent decay to each final states. We emphasize that this approach does not require any particle to appear on shell, and keeps track of all spin correlations and finite width effects.

To mix the stop and top samples so that we can investigate the impact of the stop contamination, we weight the events from the two samples according to their leading order

JHEP06(2020)019

cross sections, appropriately normalized by the total number of events generated. The stop production cross section is approximately $O(10\%)$ of the top, when the stop mass is within the range we scan.

B The all-hadronic channel

In the all-hadronic channel, the final states is characterized by two b-jets and four light-flavor jets. While this channel has the largest branching ratio (45.7%) of the three final states, it suffers from a large QCD multi-jet background and from large uncertainties in the JES. This channel is the most challenging to measure, which explains why it has the largest error bar.

B.1 Pre-selection cuts

The following preselection cuts are required before applying the template procedures. Events with isolated e/μ are excluded. At least 6 jets with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$ are required, and at least 5 of these jets must have $p_T > 60\,\text{GeV}$. For any pair of jets, an isolation requirement is applied such that $\Delta R(j_i, j_k) > 0.6$, where ΔR is the angular distance between two objects. An event must contain at least 2 b-tagged jets, with an azimuthal separation of $\Delta\phi(b_i, b_j) > 1.5$. To remove events with neutrinos, a missing transverse energy cut of $\not{E}_T < 60\,\text{GeV}$ is applied.

The all-hadronic channel has large combinatoric background, due to the homogeneity of the final state. To associate the jets with a particular top decay, a minimum χ^2 approach is utilized. One keeps the permutation that gives the lowest χ^2 among all possible permutations of jets in an event, where the χ^2 is defined as

$$\chi^2 = \frac{\left(m_{b_1 j_1 j_2} - m_{b_2 j_3 j_4}\right)^2}{\sigma^2_{\Delta m_{bjj}}} + \frac{\left(m_{j_1 j_2} - m_W^{\text{MC}}\right)^2}{\sigma^2_{m_W^{\text{MC}}}} + \frac{\left(m_{j_3 j_4} - m_W^{\text{MC}}\right)^2}{\sigma^2_{m_W^{\text{MC}}}}\,, \tag{B.1}$$

where the m_W^{MC} is taken to be $81.18 \pm 0.04\,\text{GeV}$ and the widths $\sigma_{\Delta m_{bjj}}$ and $\sigma_{m_W^{\text{MC}}}$ are taken from [12]: $\sigma_{\Delta m_{bjj}} = 21.60 \pm 0.16\,\text{GeV}$ and $\sigma_{m_W^{\text{MC}}} = 7.89 \pm 0.05\,\text{GeV}$. Then the preselection requires $\chi^2 < 11$. Finally, a cut is applied to the azimuthal angle between b-jets and their associated W boson: the average of the two angular separations between the b and the W for each event must satisfy $\langle \Delta\phi(b, W) \rangle < 2$. For validation, we present figure 13, which gives distributions for the three and two jet invariant masses, after the pre-selection cuts are applied.

B.2 R_{32} templates

The observable R_{32} is defined as the ratio of the three-jet mass to the di-jet mass, where the three-jet is a proxy for the top decay and di-jet is associated with the W boson decay. One reason this observable is chosen for building the templates is that it partially reduces the systematic errors due to uncertainties in the JES. Following ATLAS, we fit the R_{32} distributions for each of the 5 MC m_t choices to the sum of a Landau function as defined

JHEP06(2020)019

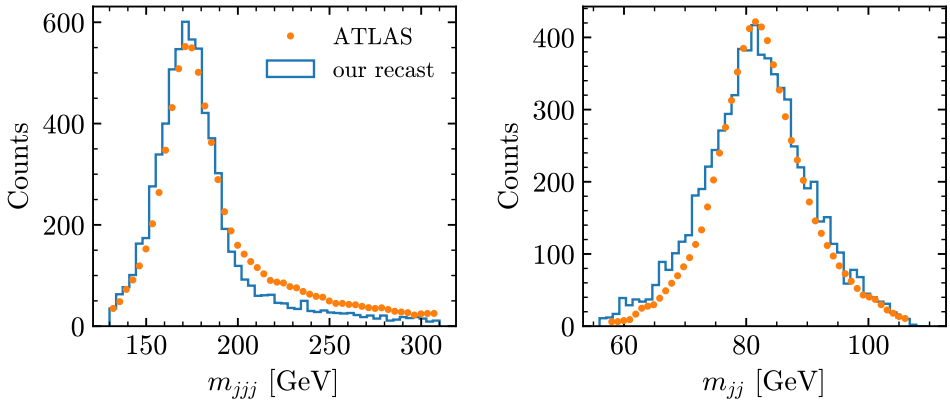

Figure 13. Two validation plots are provided for the all-hadronic channel. Three-jet invariant mass distributions for top candidates (left) and di-jet invariant mass for W-boson candidates (right), assuming $m_t = 172.5\,\text{GeV}$. The blue distributions correspond to our simulation and the orange dots represent the ATLAS results [12], which have been rescaled to so that the normalizations agree.

Cut	All-hadronic
Before cuts	108433
e/μ isolation	108367
≥ 6 jets with $p_T > 25\,\text{GeV}$	71173
$N_{\text{btag}} \geq 2$	23008
$\Delta R(j_i, j_k) > 0.6$	18241
≥ 5 jets with $p_T > 60\,\text{GeV}$	4499
$E_T^{\text{miss}} < 60\,\text{GeV}$	4003
$\Delta\phi(b_i, b_j) > 1.5$	1268
$\chi^2 < 11$	829
$\langle\Delta\phi(b, W)\rangle < 2$	780

Table 2. The cut-flow table for all-hadronic sample with $m_t = 172.5\,\text{GeV}$. The initial number of events is determined using the ATLAS integrated luminosity $L = 20.2\,\text{fb}^{-1}$ and the production cross section as calculated using `MadGraph5_aMC@NLO`.

in eq. (2.2), and a Novosibirsk function defined as

$$F(x) = N \exp\left(-\frac{1}{2\,\sigma_0^2} \log^2\left(1 - \frac{x - x_p}{\sigma_E}\,\eta\right) - \frac{\sigma_0^2}{2}\right), \tag{B.2}$$

with

$$\sigma_0 = \frac{1}{\sqrt{\log 4}} \sinh^{-1}\left(\sqrt{\log 4}\,\eta\right) \tag{B.3}$$

where x_p is the peak, σ_E is the width and η is the asymmetry tail factor. As described in section 2.1 above, we account for the statistical uncertainties associated with having a finite

JHEP06(2020)019

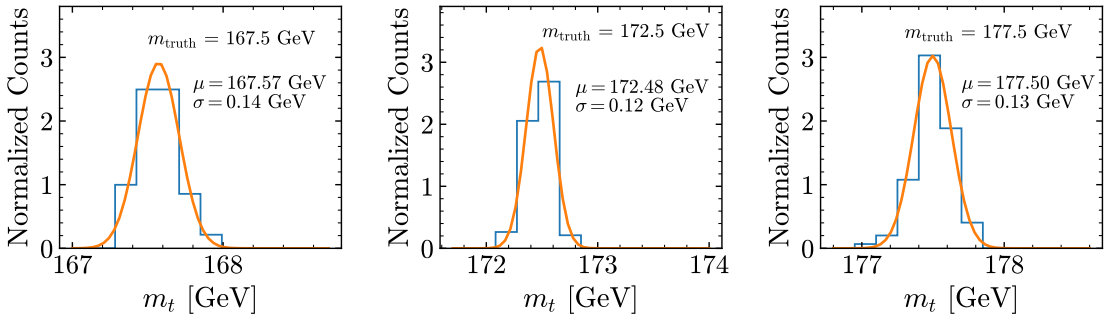

Figure 14. All-hadronic template closure tests for m_t=167.5 GeV, 172.5 GeV and 177.5 GeV. The distribution of bootstrapped results are given by the histogram, along with an accompanying Gaussian fit, whose mean and standard deviation are inset within each panel.

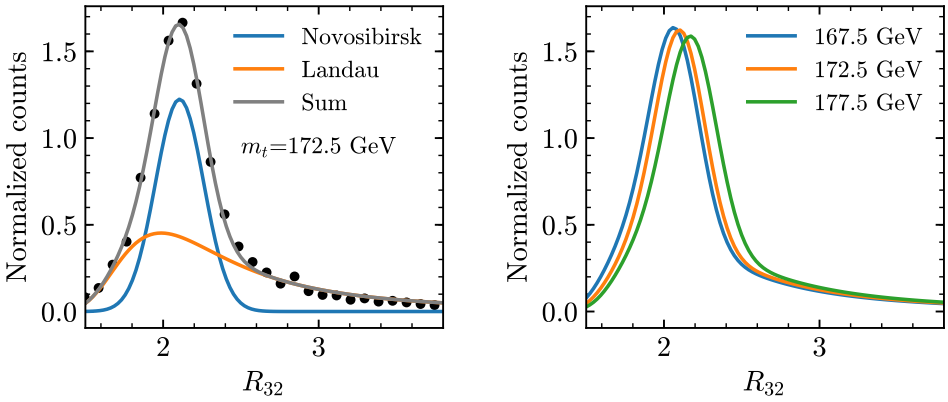

Figure 15. The left panel gives the individual contributions of the Novosibirsk and Landau function using the best fit parameters for the m_t=172.5 GeV sample. The right panel gives the R_{32} template fits for m_t=167.5 GeV, 172.5 GeV, and 177.5 GeV, respectively, in the all-hadronic channel.

data set by bootstrapping 100 samples which are taken to be 3/5 of the full data. We then repeat the R_{32} fit for each of these datasets, and use these as input to generate a template as a function of m_t. We then test that this procedure closes, which gives the histogram plotted in figure 14. Fitting these distributions to a Gaussian gives a quantitative measure of the closure goodness in the form of the mean and standard deviation given in each panel.

C The di-leptonic channel

In the di-leptonic channel, each of the W bosons decays into a charged lepton and a neutrino. The final state is characterized by two b-jets, two leptons (e or μ), and $\not{E}_T$. One advantage of this channel is that the background is relatively low, especially in the $e\mu$ final state where there is no contribution from Z boson decays. Some drawbacks are that the branching ratio is only 10.5%, and that it is not possible to reconstruct the top mass directly since there are two neutrinos that contribute to the $\not{E}_T$.

JHEP06(2020)019

Cut	$e\,e$	$\mu\,\mu$	$e\,\mu$
Before cuts	7796	13151	20229
$\not{E}_T > 60\,\text{GeV}$	3993	6784	–
$m_{\ell\ell} > 15\,\text{GeV}$	3942	6713	–
$> 10\,\text{GeV}$ from Z pole	3365	5734	–
$p_T(e+\mu+jets) > 130\,\text{GeV}$	–	–	19705
≥ 2 valid jets	2818	4783	16812
≥ 1 b-tagged jet	2450	4143	14628

Table 3. Cut-flow table for di-leptonic sample with $m_t = 172.5\,\text{GeV}$, separated by the flavor of leptons in the final state. The initial number of events is determined using the ATLAS integrated luminosity $L = 20.2\,\text{fb}^{-1}$, and the production cross section is calculated using `MadGraph5_aMC@NLO`.

C.1 Pre-selection cuts

The physics object definitions are given as follows. Electron candidates are required to have a transverse momentum of $p_T > 25\,\text{GeV}$ and a rapidity $|\eta| < 2.47$ excluding range (1.37, 1.52). Muon candidates must satisfy $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$. Muons must additionally satisfy an isolation requirement: muons within a $\Delta R = 0.4$ cone about the axis of a jet that has $p_T > 25\,\text{GeV}$ are not considered. Jets must satisfy an isolation requirement: events with jets that lie within a $\Delta R = 0.2$ cone about the axis of an electron candidate are removed. Then, an electron isolation requirement discards events where electrons are found within $\Delta R = 0.4$ cone about any of the remaining jets. The b-tagging efficiency is set to 0.7, and rejection factors of 5 and 137 are taken for c and light-flavor quarks respectively.

Now that we have defined our objects, we will walk through the pre-selection requirements. Events are required to have a signal from the single-electron or single-muon trigger and at least one primary vertex with at least five associated tracks (we assume the trigger efficiency is 100% for events that have an isolated electron or muon). An event must have exactly two oppositely charged leptons, where at least one of them must match the object that fired the corresponding trigger. In the same lepton flavor channels, $\not{E}_T > 60\,\text{GeV}$ is required. The invariant mass of the lepton pair is must be $m_{\ell\ell} > 15\,\text{GeV}$, excluding a window within 10 GeV of the Z boson mass. In the different lepton flavor channels, the scalar sum of p_T of the two selected leptons and all jets is required to be larger than 130 GeV. There must be at least two valid jets, and at least one of these jets must be b-tagged. Finally, a cut on $p_{T_{\ell b}} > 120\,\text{GeV}$ is required.[12]

[12]To compute $p_{T_{\ell b}}$, we need to pair up the b with one of the leptons. ATLAS does this using a multi-variate approach (MV): the two jets carrying the highest MV1 weight are taken as the two b-jets originating from the decays of the two top quarks. In our analysis, we simply obtain the b-jets by selecting the two b-tagged jets with the highest and second highest p_T. If there is only 1 b-tagged jet in an event, then we take the second b to be the un-tagged jet with the highest p_T. Whichever way of pairing up the lepton and the b-jet, gives a lower invariant mass is used to calculate $p_{T_{\ell b}}$, and the observable used in the template procedure $m_{\ell b}$.

JHEP06(2020)019

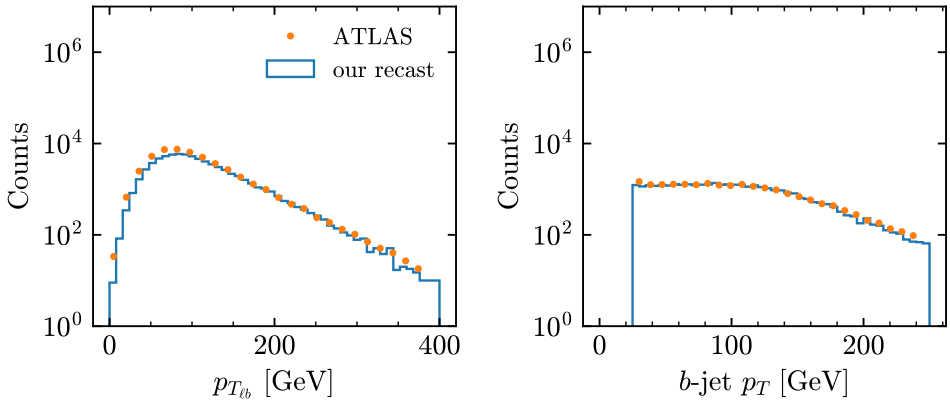

Figure 16. Two validation plots are provided for the di-leptonic channel. The $p_{T_{\ell b}}$ distribution is given in the left panel, and the b-jet p_T distribution is given on the right, assuming $m_t = 172.5\,\text{GeV}$. The blue distributions correspond to our simulation and the orange dots represent the ATLAS results [14], which have been rescaled to so that the normalizations agree.

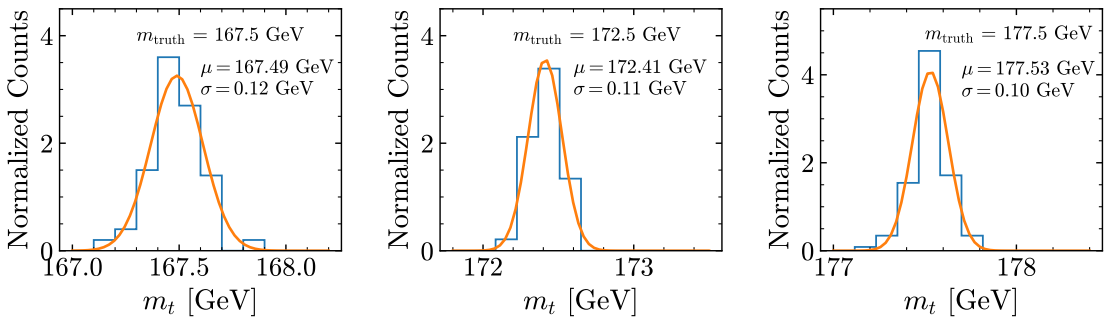

Figure 17. Di-leptonic template closure tests for m_t=167.5 GeV, 172.5 GeV and 177.5 GeV. The distribution of bootstrapped results are given by the histogram, along with an accompanying Gaussian fit, whose mean and standard deviation are inset within each panel.

C.2 $m_{\ell b}$ templates

In figure 16, we show two distributions computed using our samples after the pre-selection cuts have been applied, along with the comparisons to those given by ATLAS. The observable $m_{\ell b}$ is used to generate templates, where the parametric fit is now chosen to be a Landau function as defined in eq. (2.2) and a Gaussian. As described in section 2.1 above, we account for the statistical uncertainties associated with having a finite data set by bootstrapping 100 samples which are taken to be 3/5 of the full data. We then repeat the $m_{\ell b}$ fit for each of these datasets, and use these as input to generate a template as a function of m_t. We then test that this procedure closes by fitting the resulting histograms and comparing the fitted mean to the input m_{truth} value, as shown in figure 17 .

JHEP06(2020)019

Cut	e
Before cuts	89162
1 lepton	19437
$\not{E}_T > 30\,\text{GeV}$	14926
$m_T^W > 30\,\text{GeV}$	13075
≥ 4 jets	7793
2 b-tagged jets	3173

Cut	μ
Before cuts	89162
1 lepton	25907
$\not{E}_T > 20\,\text{GeV}$	23089
$\not{E}_T + m_T^W > 60\,\text{GeV}$	21791
≥ 4 jets	12883
2 b-tagged jets	5262

Table 4. Cutflow table for semi-leptonic sample with $m_t = 172.5\,\text{GeV}$, separated by the identity of the lepton. The initial number of events is determined using the ATLAS integrated luminosity $L = 20.2\,\text{fb}^{-1}$, and the production cross section as calculated using `MadGraph5_aMC@NLO`.

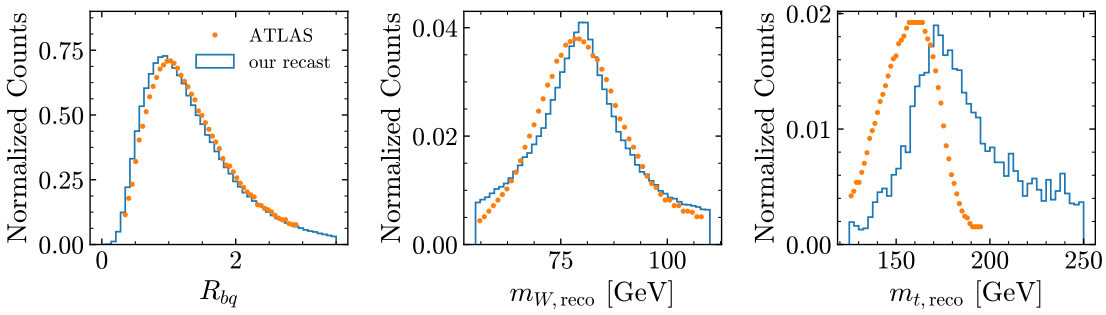

Figure 18. This figure provides distributions of the three different observables used in the template analysis for the measurement of the top quark mass by ATLAS in the semi-leptonic channel. The panels from left to right show R_{bq} which is defined in eq. (3.3), $m_{W,\text{reco}}$, and $m_{t,\text{reco}}$, where the later two are the determined from maximizing the event-by-event likelihood defined in eq. (3.2). The distributions for our analysis are shown by the solid blue histogram, while the orange dots are provided by ATLAS in [13]. We observe good agreement for the $m_{W,\text{reco}}$ and R_{bq} distributions, but we are unable to reproduce the $m_{t,\text{reco}}$ distribution.

D More on the semi-leptonic channel

This section provides extra validation information for the semi-leptonic channel. In section 3.1, we summarized the pre-selection cuts for ATLAS's semi-leptonic analysis. The detailed cutflow table is given by table 4.

Distributions of the observables used by ATLAS in this channel are shown in figure 18. Note that we do not include the transfer functions in our likelihood, which specifically affects the $m_{t,\text{reco}}$ distribution, as explained in section 3.4.

To make our proposed two-dimensional templates, we generate parton-level events with `MadGraph5_aMC@NLO` that are subsequently passed to `Pythia8` for showering and hadronization, and then to `Delphes` to model detector effects. Five million events are generated for each of five choices for the top mass; a table providing the number of events that pass the preselection cuts is given in table 4. For each event, we use the likelihood defined in eq. (3.2) to chose the assignment of the jets, b-jets, $p_{z,\nu}$, and $m_{W,\text{reco}}$. The resulting two-

JHEP06(2020)019

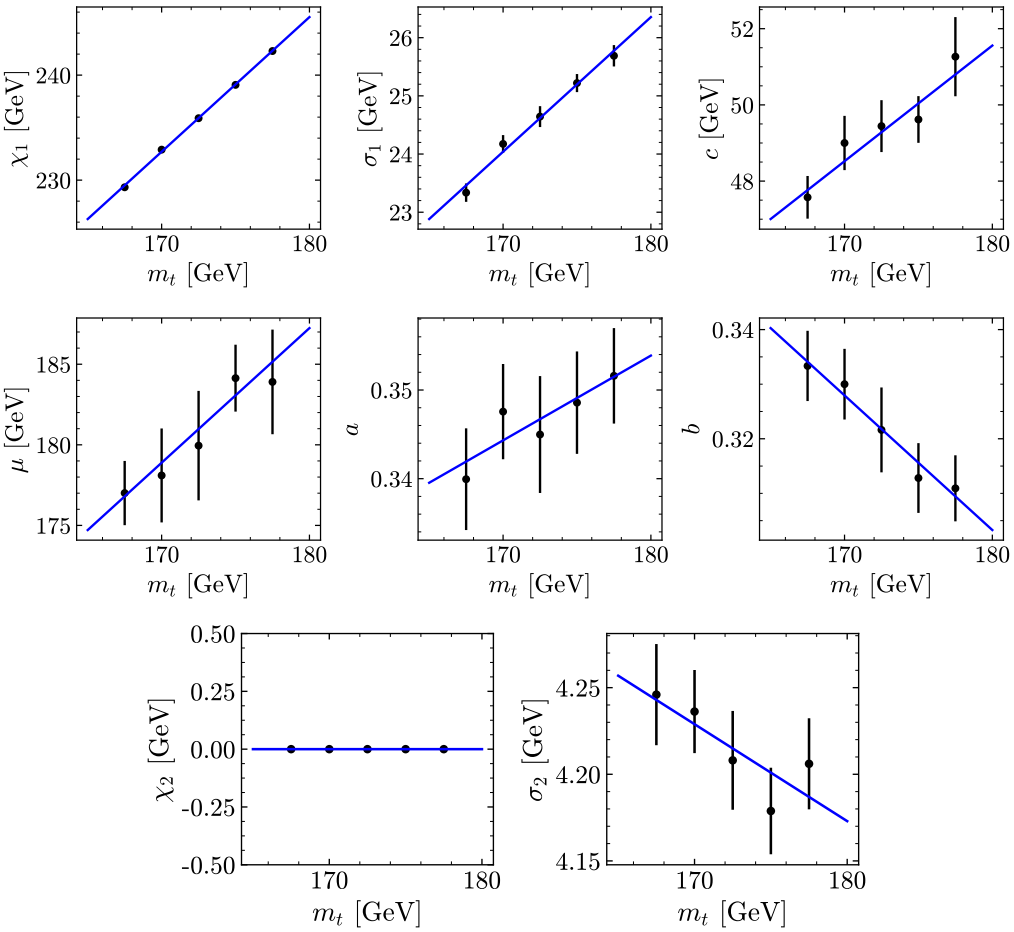

Figure 19. Here we provide the mean and standard deviation of the best fit parameters for the 100 samples derived using the bootstrap procedure. These distributions are then fit to a line as a function of m_t in order to generate the semi-leptonic template.

dimensional distributions in the m_t^{had} versus m_t^{lep} plane are then fit using the parametric function defined in eq. (3.5).

In order to follow the procedure discussed in section 2, we would like to have a set of ~ 100 statistically independent samples to work with. However, it is computationally to expensive to re-generate the 5 million events many times. Therefore, we circumvent this issue using the statistical bootstrap, see e.g. [76]. Specifically, we random draw 3/5 of the 5 million events 100 times, allowing for replacement such that some events can be drawn more than once. We then find the best fit using our parametric function to each bootstrapped data set, providing us with an ensemble of 100 best fit parameters. The results are shown in figure 19, where the points and error bars show the mean and standard deviation, respectively, of the best fit value for each parameter at each of the five top mass choices. Finally, we fit a line to each of these parameters as a function of the top mass, which is the input needed to define our two-dimensional template as a function of a single top mass parameter.

JHEP06(2020)019

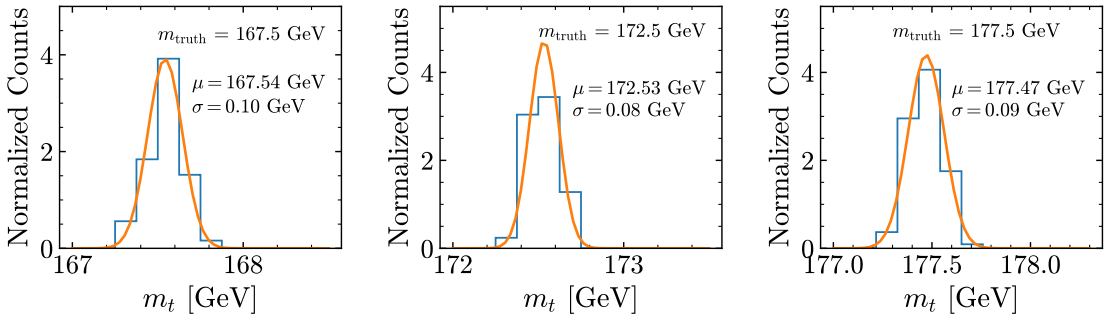

Figure 20. The semi-leptonic template closure tests for m_t=167.5 GeV, 172.5 GeV and 177.5 GeV, using our modified approach. The distribution of bootstrapped results are given by the histogram, along with an accompanying Gaussian fit, whose mean and standard deviation are inset within each panel.

The closure test is performed by taking another independent bootstrapped sample of 3 million events and fitting this data to our template to derive a best fit value of m_t. This is repeated 100 times for each of the five truth top mass choices. The results are shown in figure 20, we show the results of this closure test which extracts the correct mass with a relatively small statistical uncertainty ~ 0.1 GeV.

E Combining measurements

Since we are interested in the global impact of stop contamination, it is useful to develop a simple framework for combining the measurements made in multiple independent channels. In particular, the results in figure 1 of the main text show parameters which best fit the mass measurements combining the three channels. In the SM only assumption (left panel of figure 1), the top mass which minimizes the χ^2 error is $m_t^{\text{comb}} = 172.83$ GeV, and the uncertainty band is determined by finding the contour where the χ^2 is larger than the minimum by 1.0. We note that both the all-hadronic and the semi-leptonic central values lie outside this best-fit uncertainty band. The regions shown in the BSM parameter space are computed in a similar fashion, see the middle and right panels of figure 1.

These regions only show the best-fit, and in particular they do not tell us how good the fit is. Therefore, it is amusing to ask if light stops can actually improve the fit to distributions measured by ATLAS. To perform a quantitative test, we compute the likelihood ratio for observing the measured values in the three channels in the BSM scenarios as compared to the SM-only assumption. The test statistic is given by

$$\lambda_{\text{comb}} = -2 \log \frac{\prod_{i\in\text{channels}} P_{\text{Gaussian}}\left(m_{t,i}^{\text{pred}};\ m_{t,i}^{\text{Obs}}, \sigma_i^{\text{Obs}}\right)}{\prod_{i\in\text{channels}} P_{\text{Gaussian}}\left(m_t^{\text{comb}};\ m_{t,i}^{\text{Obs}}, \sigma_i^{\text{Obs}}\right)}\,, \tag{E.1}$$

where $m_{t,i}^{\text{Obs}}$ and σ_i^{Obs} are the channel specific measurement and uncertainty given by ATLAS, and $m_{t,i}^{\text{pred}}$ is the value predicted in the BSM model for a given Monte Carlo m_t,

JHEP06(2020)019

$m(\tilde{t}_1)$, and $m(\tilde{\chi}_1^0)$. This test statistic is constructed so that when $\lambda_{\text{comb}} > 0$ the SM is a better fit, while when $\lambda_{\text{comb}} < 0$ the BSM scenario is preferred.

The values of λ_{comb} computed in the $m(\tilde{t}_1)$-m_t plane are shown in figure 12. The red regions correspond to values of $\lambda_{\text{comb}} > 0$, indicating that the SM alone provides a better fit to the data. We gray out any parameter space with $\lambda_{\text{comb}} > 1$ for brevity, since this region has a much stronger preference for the SM alone (and is of course additionally constrained by direct searches for stops). The white regions have a similar fit between the models, giving $\lambda_{\text{comb}} = 0$. Intriguingly, we find a small region in the $m(\tilde{\chi}_1^0) = 20\,\text{GeV}$ panel with $\lambda_{\text{comb}} < 0$. However, our analysis yields that this parameter point is a mere 0.1 σ more consistent with the data than the SM alone. We *do not* take this to be evidence for a BSM contribution to the top mass measurements.

References

[1] ATLAS collaboration, *Search for a scalar partner of the top quark in the jets plus missing transverse momentum final state at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **12** (2017) 085 [arXiv:1709.04183] [INSPIRE].

[2] ATLAS collaboration, *Search for top-squark pair production in final states with one lepton, jets and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector*, *JHEP* **06** (2018) 108 [arXiv:1711.11520] [INSPIRE].

[3] ATLAS collaboration, *Search for a scalar partner of the top quark in the all-hadronic $t\bar{t}$ plus missing transverse momentum final state at $\sqrt{s} = 13$ TeV with the ATLAS detector*, arXiv:2004.14060 [INSPIRE].

[4] CMS collaboration, *Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **10** (2017) 005 [arXiv:1707.03316] [INSPIRE].

[5] CMS collaboration, *Search for top squark pair production in pp collisions at $\sqrt{s} = 13$ TeV using single lepton events*, *JHEP* **10** (2017) 019 [arXiv:1706.04402] [INSPIRE].

[6] M. Czakon, A. Mitov, M. Papucci, J.T. Ruderman and A. Weiler, *Closing the stop gap*, *Phys. Rev. Lett.* **113** (2014) 201803 [arXiv:1407.1043] [INSPIRE].

[7] T. Eifert and B. Nachman, *Sneaky light stop*, *Phys. Lett.* **B 743** (2015) 218 [arXiv:1410.7025] [INSPIRE].

[8] D0 collaboration, *Precision measurement of the top-quark mass in lepton + jets final states*, *Phys. Rev.* **D 91** (2015) 112003 [arXiv:1501.07912] [INSPIRE].

[9] D0 collaboration, *Measurement of the top quark mass using the matrix element technique in dilepton final states*, *Phys. Rev.* **D 94** (2016) 032004 [arXiv:1606.02814] [INSPIRE].

[10] CMS collaboration, *Measurement of the top quark mass using proton-proton data at $\sqrt{s} = 7$ and 8 TeV*, *Phys. Rev.* **D 93** (2016) 072004 [arXiv:1509.04044] [INSPIRE].

JHEP06(2020)019

[11] CMS collaboration, *Measurement of the top quark mass in the dileptonic $t\bar{t}$ decay channel using the mass observables $M_{b\ell}$, M_{T2} and $M_{b\ell\nu}$ in pp collisions at $\sqrt{s} = 8$ TeV*, *Phys. Rev.* **D 96** (2017) 032002 [arXiv:1704.06142] [INSPIRE].

[12] ATLAS collaboration, *Top-quark mass measurement in the all-hadronic $t\bar{t}$ decay channel at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **09** (2017) 118 [arXiv:1702.07546] [INSPIRE].

[13] ATLAS collaboration, *Measurement of the top quark mass in the $t\bar{t} \to$ lepton+jets channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results*, *Eur. Phys. J.* **C 79** (2019) 290 [arXiv:1810.01772] [INSPIRE].

[14] ATLAS collaboration, *Measurement of the top quark mass in the $t\bar{t} \to$ dilepton channel from $\sqrt{s} = 8$ TeV ATLAS data*, *Phys. Lett.* **B 761** (2016) 350 [arXiv:1606.02179] [INSPIRE].

[15] M. Papucci, J.T. Ruderman and A. Weiler, *Natural SUSY endures*, *JHEP* **09** (2012) 035 [arXiv:1110.6926] [INSPIRE].

[16] R. Essig, E. Izaguirre, J. Kaplan and J.G. Wacker, *Heavy flavor simplified models at the LHC*, *JHEP* **01** (2012) 074 [arXiv:1110.6443] [INSPIRE].

[17] LHC New Physics Working Group collaboration, *Simplified models for LHC new physics searches*, *J. Phys.* **G 39** (2012) 105005 [arXiv:1105.2838] [INSPIRE].

[18] S. Dimopoulos and G.F. Giudice, *Naturalness constraints in supersymmetric theories with nonuniversal soft terms*, *Phys. Lett.* **B 357** (1995) 573 [hep-ph/9507282] [INSPIRE].

[19] A.G. Cohen, D.B. Kaplan and A.E. Nelson, *The more minimal supersymmetric Standard Model*, *Phys. Lett.* **B 388** (1996) 588 [hep-ph/9607394] [INSPIRE].

[20] P. Fayet, *Supersymmetry and weak, electromagnetic and strong interactions*, *Phys. Lett.* **B 64** (1976) 159 [INSPIRE].

[21] G. Jungman, M. Kamionkowski and K. Griest, *Supersymmetric dark matter*, *Phys. Rept.* **267** (1996) 195 [hep-ph/9506380] [INSPIRE].

[22] K. Krizka, A. Kumar and D.E. Morrissey, *Very light scalar top quarks at the LHC*, *Phys. Rev.* **D 87** (2013) 095016 [arXiv:1212.4856] [INSPIRE].

[23] D.S.M. Alves, M.R. Buckley, P.J. Fox, J.D. Lykken and C.-T. Yu, *Stops and $\not{E}_T$: the shape of things to come*, *Phys. Rev.* **D 87** (2013) 035016 [arXiv:1205.5805] [INSPIRE].

[24] Z. Han, A. Katz, D. Krohn and M. Reece, *(Light) stop signs*, *JHEP* **08** (2012) 083 [arXiv:1205.5808] [INSPIRE].

[25] C. Kilic and B. Tweedie, *Cornering light stops with dileptonic m_{T2}*, *JHEP* **04** (2013) 110 [arXiv:1211.6106] [INSPIRE].

[26] P. Agrawal and C. Frugiuele, *Mixing stops at the LHC*, *JHEP* **01** (2014) 115 [arXiv:1304.3068] [INSPIRE].

[27] B. Dutta et al., *Probing compressed top squark scenarios at the LHC at 14 TeV*, *Phys. Rev.* **D 90** (2014) 095022 [arXiv:1312.1348] [INSPIRE].

[28] J. Aebischer, A. Crivellin and C. Greub, *One-loop SQCD corrections to the decay of top squarks to charm and neutralino in the generic MSSM*, *Phys. Rev.* **D 91** (2015) 035010 [arXiv:1410.8459] [INSPIRE].

[29] W.S. Cho et al., *Improving the sensitivity of stop searches with on-shell constrained invariant mass variables*, *JHEP* **05** (2015) 040 [arXiv:1411.0664] [INSPIRE].

JHEP06(2020)019

[30] M.R. Buckley, T. Plehn and M.J. Ramsey-Musolf, *Top squark with mass close to the top quark*, *Phys. Rev.* **D 90** (2014) 014046 [arXiv:1403.2726] [INSPIRE].

[31] R. Gröber, M.M. Mühlleitner, E. Popenda and A. Wlotzka, *Light stop decays: implications for LHC searches*, *Eur. Phys. J.* **C 75** (2015) 420 [arXiv:1408.4662] [INSPIRE].

[32] J. Fan and M. Reece, *A new look at Higgs constraints on stops*, *JHEP* **06** (2014) 031 [arXiv:1401.7671] [INSPIRE].

[33] H. An and L.-T. Wang, *Opening up the compressed region of top squark searches at* 13 *TeV LHC*, *Phys. Rev. Lett.* **115** (2015) 181602 [arXiv:1506.00653] [INSPIRE].

[34] B. Batell, M. McCullough, D. Stolarski and C.B. Verhaaren, *Putting a stop to di-Higgs modifications*, *JHEP* **09** (2015) 216 [arXiv:1508.01208] [INSPIRE].

[35] B. Batell and S. Jung, *Probing light stops with stoponium*, *JHEP* **07** (2015) 061 [arXiv:1504.01740] [INSPIRE].

[36] A. Belyaev, V. Sanz and M. Thomas, *Towards model-independent exclusion of light stops*, *JHEP* **01** (2016) 102 [arXiv:1510.07688] [INSPIRE].

[37] J. Fan, R. Krall, D. Pinner, M. Reece and J.T. Ruderman, *Stealth supersymmetry simplified*, *JHEP* **07** (2016) 016 [arXiv:1512.05781] [INSPIRE].

[38] G. Ferretti, R. Franceschini, C. Petersson and R. Torre, *Spot the stop with a b-tag*, *Phys. Rev. Lett.* **114** (2015) 201801 [arXiv:1502.01721] [INSPIRE].

[39] K.-I. Hikasa, J. Li, L. Wu and J.M. Yang, *Single top squark production as a probe of natural supersymmetry at the LHC*, *Phys. Rev.* **D 93** (2016) 035003 [arXiv:1505.06006] [INSPIRE].

[40] S. Macaluso, M. Park, D. Shih and B. Tweedie, *Revealing compressed stops using high-momentum recoils*, *JHEP* **03** (2016) 151 [arXiv:1506.07885] [INSPIRE].

[41] H. An, J. Gu and L.-T. Wang, *Exploring the nearly degenerate stop region with sbottom decays*, *JHEP* **04** (2017) 084 [arXiv:1611.09868] [INSPIRE].

[42] H.-C. Cheng, L. Li and Q. Qin, *Second stop and sbottom searches with a stealth stop*, *JHEP* **11** (2016) 181 [arXiv:1607.06547] [INSPIRE].

[43] P. Konar, T. Mondal and A.K. Swain, *Demystifying the compressed top squark region with kinematic variables*, *Phys. Rev.* **D 96** (2017) 095011 [arXiv:1612.03269] [INSPIRE].

[44] J. Aebischer, A. Crivellin, C. Greub and Y. Yamada, *The MSSM without gluinos; an effective field theory for the stop sector*, *Eur. Phys. J.* **C 77** (2017) 740 [arXiv:1703.08061] [INSPIRE].

[45] T. Cohen, W. Hopkins, S. Majewski and B. Ostdiek, *Magnifying the ATLAS stealth stop splinter: impact of spin correlations and finite widths*, *JHEP* **07** (2018) 142 [arXiv:1804.00111] [INSPIRE].

[46] T. Roxlo and M. Reece, *Opening the black box of neural nets: case studies in stop/top discrimination*, arXiv:1804.09278 [INSPIRE].

[47] M.A. Gigg and P. Richardson, *Simulation of finite width effects in physics beyond the Standard Model*, arXiv:0805.3037 [INSPIRE].

[48] ATLAS collaboration, *Search for direct top squark pair production in final states with two leptons in* $\sqrt{s} = 13$ *TeV pp collisions with the ATLAS detector*, *Eur. Phys. J.* **C 77** (2017) 898 [arXiv:1708.03247] [INSPIRE].

JHEP06(2020)019

[49] ATLAS collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector*, *Eur. Phys. J.* **C 74** (2014) 3109 [*Addendum ibid.* **C 76** (2016) 642] [arXiv:1406.5375] [INSPIRE].

[50] ATLAS collaboration, *Measurements of top-quark pair spin correlations in the $e\mu$ channel at $\sqrt{s} = 13$ TeV using pp collisions in the ATLAS detector*, arXiv:1903.07570 [INSPIRE].

[51] CMS collaboration, *Search for the pair production of light top squarks in the $e^{\pm}\mu^{\mp}$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **03** (2019) 101 [arXiv:1901.01288] [INSPIRE].

[52] ATLAS collaboration, *ATLAS run 1 searches for direct pair production of third-generation squarks at the Large Hadron Collider*, *Eur. Phys. J.* **C 75** (2015) 510 [*Erratum ibid.* **C 76** (2016) 153] [arXiv:1506.08616] [INSPIRE].

[53] S. Fleming, A.H. Hoang, S. Mantry and I.W. Stewart, *Jets from massive unstable particles: top-mass determination*, *Phys. Rev.* **D 77** (2008) 074010 [hep-ph/0703207] [INSPIRE].

[54] A.H. Hoang and I.W. Stewart, *Top mass measurements from jets and the Tevatron top-quark mass*, *Nucl. Phys. Proc. Suppl.* **185** (2008) 220 [arXiv:0808.0222] [INSPIRE].

[55] M. Butenschoen, B. Dehnadi, A.H. Hoang, V. Mateu, M. Preisser and I.W. Stewart, *Top quark mass calibration for Monte Carlo event generators*, *Phys. Rev. Lett.* **117** (2016) 232001 [arXiv:1608.01318] [INSPIRE].

[56] A.H. Hoang, S. Mantry, A. Pathak and I.W. Stewart, *Extracting a short distance top mass with light grooming*, *Phys. Rev.* **D 100** (2019) 074021 [arXiv:1708.02586] [INSPIRE].

[57] A.H. Hoang, S. Plätzer and D. Samitz, *On the cutoff dependence of the quark mass parameter in angular ordered parton showers*, *JHEP* **10** (2018) 200 [arXiv:1807.06617] [INSPIRE].

[58] P. Nason, *Theory summary*, PoS(TOP2015)056 (2016) [arXiv:1602.00443] [INSPIRE].

[59] P. Nason, *The top mass in hadronic collisions*, in *From my vast repertoire. . . : Guido Altarelli's legacy*, A. Levy, S. Forte and G. Ridolfi eds., World Scientific, Singapore (2019), pg. 123 [arXiv:1712.02796] [INSPIRE].

[60] A. Andreassen and M.D. Schwartz, *Reducing the top quark mass uncertainty with jet grooming*, *JHEP* **10** (2017) 151 [arXiv:1705.07135] [INSPIRE].

[61] G. Corcella, *The top-quark mass: challenges in definition and determination*, *Front. Phys.* **7** (2019) 54 [arXiv:1903.06574] [INSPIRE].

[62] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, *Equation of state calculations by fast computing machines*, *J. Chem. Phys.* **21** (1953) 1087.

[63] W.K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, *Biometrika* **57** (1970) 97.

[64] J. Alwall 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].

[65] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A brief introduction to PYTHIA* 8.1, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [INSPIRE].

[66] DELPHES 3 collaboration, *DELPHES 3, a modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [arXiv:1307.6346] [INSPIRE].

JHEP06(2020)019

[67] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[68] J. Erdmann et al., *A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework*, *Nucl. Instrum. Meth.* **A 748** (2014) 18 [arXiv:1312.5595] [INSPIRE].

[69] G. Breit and E. Wigner, *Capture of slow neutrons*, *Phys. Rev.* **49** (1936) 519 [INSPIRE].

[70] L. Moneta, I. Antcheva and D. Gonzalez Maline, *Recent improvements of the root fitting and minimization classes*, PoS(ACAT08)075 (2009).

[71] *ROOT data analysis framework webpage*, http://root.cern.ch/.

[72] ATLAS collaboration, *Measurement of spin correlation in top-antitop quark events and search for top squark pair production in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Phys. Rev. Lett.* **114** (2015) 142001 [arXiv:1412.4742] [INSPIRE].

[73] 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].

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

[75] M. Cacciari and G.P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, *Phys. Lett.* **B 641** (2006) 57 [hep-ph/0512210] [INSPIRE].

[76] Wikipedia contributors, *Bootstrapping (statistics)*, https://en.wikipedia.org/wiki/Bootstrapping_(statistics), (2019).

JHEP06(2020)019