Julia Kim Summer 2022

MT2DC Analysis Documentation

The purpose of this document is to present the theory of the new kinematic variable $m_{\mathrm{T2}}^{\mathrm{DC}}$ and the work accomplished in its investigation and development. This document is divided into four parts, in which are presented the background theory, the topics of research, the codes used for analysis, and a selection of important plots that illustrate properties of the kinematic variables and the functionality of the codes.

Background

The ATLAS collaboration have developed a general purpose detector at the Large Hadron Collider (LHC) at CERN, used for the measurement of particles emerging from p-p and heavy-ion collisions. The detector is composed of a series of concentric, cylindrical subsystems, each sensitive to the detection of different kinds of particles (e.g., photons, hadrons, muons, electrons), excepting neutrinos. It is conventional to describe the trajectories of these particles with a right-handed co-ordinate system, whose origin is the nominal interaction point (IP) and whose z-axis and x-y plane are respectively defined by the beam direction and the plane *transverse* to it (Fig. 1). Its positive x-axis points from the interaction point to the centre of the LHC ring, whilst its positive y-axis points upwards. As the detector's geometry is cylindrical, it is common to measure the azimuthal angle ϕ of the particle, which sweeps around the beam in the transverse plane, as well as its **pseudo-rapidity** η, defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$.

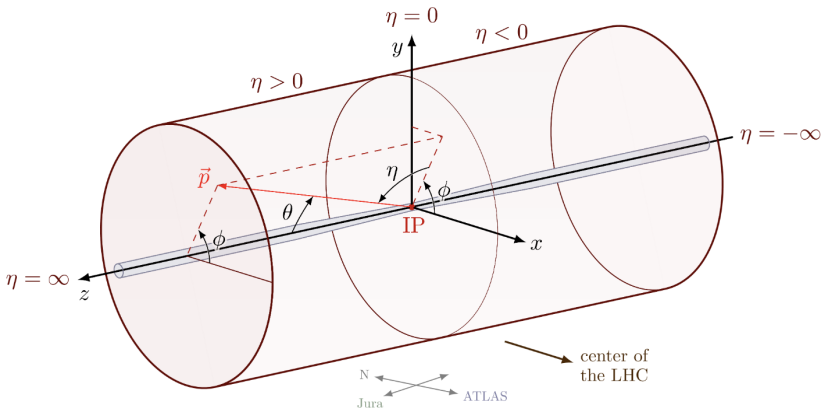

Figure 1: Cartesian and cylindrical co-ordinate systems used to describe events in the ATLAS detector.

A first instance of a semi-invisible particle event measurable in the ATLAS detector is a W-boson decay, in which a W-boson decays into a visible Standard Model particle ℓ and an invisible ν. As the neutrino neither interacts with the strong nor the electromagnetic force, it passes by undetected in ATLAS' subsystems and, therefore, is simply presumed to carry away the **missing transverse energy** $E_{\mathrm{T}}^{\mathrm{miss}}$ and **momentum** $p_{\mathrm{T}}^{\mathrm{miss}}$ of the event. From measurements performed upon the observable ℓ and measurements used to ascertain the transverse energy, the **transverse mass** m_{T} of the parent particle can be defined, as follows:

$$m_{\mathrm{T}}(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) \doteq m_{\mathrm{T}}(W) = \sqrt{m(\ell)^2 + m(E_{\mathrm{T}}^{\mathrm{miss}})^2 + 2[E_{\mathrm{T}}(\ell)E_{\mathrm{T}}^{\mathrm{miss}} - \vec{p}_{\mathrm{T}}(\ell)\vec{p}_{\mathrm{T}}^{\mathrm{miss}}]}, \tag{1}$$

where $E_{\mathrm{T}}(a) = \sqrt{m(a)^2 + \vec{p}_{\mathrm{T}}(a)^2}$. As the energy and momentum in (1) are determined in the transverse direction alone, discarding any component in the beam direction, the transverse mass will always be less than or equal to the *true* mass of the parent particle, $m_W \simeq 80.4$ GeV. In short,

$$m_{\mathrm{T}}(W) \leq m_W. \tag{2}$$

Also measurable in the ATLAS detector is a $t\bar{t}$ decay chain (DC), in which each t-quark of a top quark-antiquark pair decays into a b-quark and a W-boson, the latter of which subsequently decays into a final

state, comprising a lepton ℓ and an invisible ν (Fig. 2). We distinguish the two 'sides' resulting from the decays of t and $\bar{t}$ by subscripts 1 and 2, respectively. Owing to there being multiple invisible products ν_1 and ν_2 in the decay, the **stransverse mass** variable m_{T2}, in lieu of the **transverse mass**, will now be used to give a partial reconstruction[1] of the masses of the parent particles.

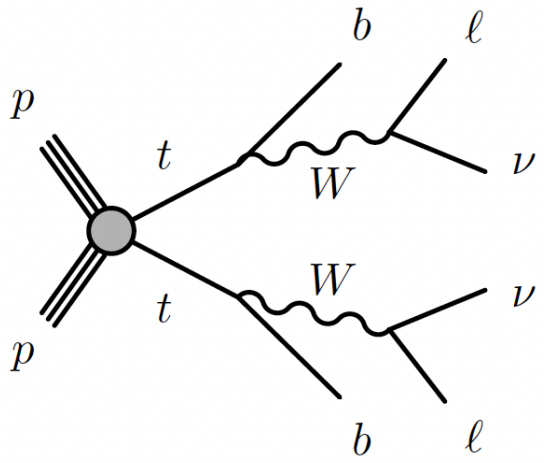

Figure 2: A Feynman diagram of a $t\bar{t}$ decay chain.

Accordingly, to give a measurement of the mass of each W-boson, the $m_{\mathrm{T2}}(W)$-variable finds the *greater* of two transverse masses $m_{\mathrm{T},1}(W)$ and $m_{\mathrm{T},2}(W)$, minimised over all possible combinations of the transverse momenta of the invisible particles (whose sum is necessarily $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$). In short,

$$\begin{aligned} m_{\mathrm{T2}}(W) &= \min_{\vec{p}_{\mathrm{T},1}^{\,\mathrm{invis}}+\vec{p}_{\mathrm{T},2}^{\,\mathrm{invis}}=\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}}[\max(m_{\mathrm{T},1}(W), m_{\mathrm{T},2}(W))] \\ &= \min_{\vec{p}_{\mathrm{T},1}^{\,\mathrm{invis}}+\vec{p}_{\mathrm{T},2}^{\,\mathrm{invis}}=\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}}[\max(m_{\mathrm{T}}(\ell_1, E_{\mathrm{T},\,1}^{\mathrm{miss}}), m_{\mathrm{T}}(\ell_2, E_{\mathrm{T},\,2}^{\mathrm{miss}}))]. \end{aligned} \tag{3}$$

Note that the $\vec{p}_{\mathrm{T},1}^{\,\mathrm{invis}}$ and $\vec{p}_{\mathrm{T},2}^{\,\mathrm{invis}}$ pair which minimise the quantity in square brackets may not be equal to the *true* transverse momenta of the neutrinos. To emphasise this distinction, this momentum pair will be called the momentum pair of the **substitute** neutrinos 1 and 2. If, instead, the *greater* of the two transverse masses were calculated using the *true* transverse momenta of ν_1 and ν_2, denoted by $m_{\mathrm{T},\nu_1}^{\mathrm{true}}(W)$ and $m_{\mathrm{T},\nu_2}^{\mathrm{true}}(W)$, respectively, then we would have

$$m_{\mathrm{T2}}(W) \leq \max(m_{\mathrm{T},\nu_1}^{\mathrm{true}}(W), m_{\mathrm{T},\nu_2}^{\mathrm{true}}(W)) \leq m_W. \tag{4}$$

Likewise, in order to give a measurement of the mass of each t parent particle,

$$\begin{aligned} m_{\mathrm{T2}}(t) &= \min_{\vec{p}_{\mathrm{T},1}^{\,\mathrm{invis}}+\vec{p}_{\mathrm{T},2}^{\,\mathrm{invis}}=\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}}[\max(m_{\mathrm{T},1}(t), m_{\mathrm{T},2}(t))] \\ &= \min_{\vec{p}_{\mathrm{T},1}^{\,\mathrm{invis}}+\vec{p}_{\mathrm{T},2}^{\,\mathrm{invis}}=\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}}[\max(m_{\mathrm{T}}(\ell_1 b_1, E_{\mathrm{T},\,1}^{\mathrm{miss}}), m_{\mathrm{T}}(\ell_2 b_2, E_{\mathrm{T},\,2}^{\mathrm{miss}}))], \end{aligned} \tag{5}$$

where $\ell_i b_i$ denotes the object in which the energy-momentum four-vectors (E, $\vec{p}$) of ℓ_i and b_i have been added together. Analogous to (4),

$$m_{\mathrm{T2}}(t) \leq \max(m_{\mathrm{T},\nu_1}^{\mathrm{true}}(t), m_{\mathrm{T},\nu_2}^{\mathrm{true}}(t)) \leq m_t. \tag{6}$$

Upon inspection, one might notice there to be no constraint in the definitions of $m_{\mathrm{T2}}(W)$ and $m_{\mathrm{T2}}(t)$, requiring their substitute momentum pair to be the same. Knowing there to be, in reality, only one set of

[1] In this document, to give a 'partial reconstruction of the mass', a 'mass measurement' or 'mass information' are to be understood as giving a quantity or a distribution of some mass-related variable, from which can be extracted properties concerning the masses of *real* particles in the decay chain (by, e.g., inspection, fitting, template methods, event categorisation (see p. 5)...). In particular, though the stransverse mass decay chain in (7) does not describe the mass of any real particle, yet it is related to the masses of identifiable particles in the decay chain through theoretical inequalities, such as in (9), (13)-(16).

neutrinos, it is useful to define a new kinematic variable **stransverse mass decay chain** $m_{\mathrm{T2}}^{\mathrm{DC}}$, in which only one single minimisation is required to produce a mass measurement for the whole chain. For a specified $\alpha \in [0,1]$, it is defined, as follows:

$$m_{\mathrm{T2}}^{\mathrm{DC}}(b,\ell,E_{\mathrm{T}}^{\mathrm{miss}},\alpha) \doteq m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha) = \min_{\vec{p}_{\mathrm{T},1}^{\,\mathrm{invis}}+\vec{p}_{\mathrm{T},2}^{\,\mathrm{invis}}=\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}}[\xi], \tag{7}$$

where

$$\xi = \alpha\left(\max\left(m_{\mathrm{T}}(\ell_1,E_{\mathrm{T},\,1}^{\mathrm{miss}}),m_{\mathrm{T}}(\ell_2,E_{\mathrm{T},\,2}^{\mathrm{miss}})\right)\right) + (1-\alpha)\left(\max\left(m_{\mathrm{T}}(\ell_1 b_1,E_{\mathrm{T},\,1}^{\mathrm{miss}}),m_{\mathrm{T}}(\ell_2 b_2,E_{\mathrm{T},\,2}^{\mathrm{miss}})\right)\right). \tag{8}$$

Most importantly, we can assert

$$\alpha m_{\mathrm{T2}}(W) + (1-\alpha)m_{\mathrm{T2}}(t) \leq m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha) \leq \alpha m_W + (1-\alpha)m_t. \tag{9}$$

To be suggestive of the similarities between the expressions of $m_{\mathrm{T2}}^{\mathrm{DC}}$, $m_{\mathrm{T2}}(t)$ and $m_{\mathrm{T2}}(W)$, it is useful to write

$$m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha) = \alpha m_{\mathrm{T2}}(W|\alpha)' + (1-\alpha)m_{\mathrm{T2}}(t|\alpha)', \tag{10}$$

where the primed variables

$$m_{\mathrm{T2}}(W|\alpha)' = \max\left(m_{\mathrm{T}}(\ell_1,E_{\mathrm{T},\,1,\,\mathrm{opt.}}^{\mathrm{miss}}),m_{\mathrm{T}}(\ell_2,E_{\mathrm{T},\,2,\,\mathrm{opt.}}^{\mathrm{miss}})\right) \tag{11}$$

$$m_{\mathrm{T2}}(t|\alpha)' = \max\left(m_{\mathrm{T}}(\ell_1 b_1,E_{\mathrm{T},\,1,\,\mathrm{opt.}}^{\mathrm{miss}}),m_{\mathrm{T}}(\ell_2 b_2,E_{\mathrm{T},\,2,\,\mathrm{opt.}}^{\mathrm{miss}})\right) \tag{12}$$

respectively represent the two coloured terms in (8), obtained *after* optimisation. For brevity, we respectively refer to (11) and (12) as the **W- and t-terms** of $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha)$. As there is no constraint requiring the substitute momentum pair for $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha)$ to be equal to those for $m_{\mathrm{T2}}(W)$ and $m_{\mathrm{T2}}(t)$[2],

$$m_{\mathrm{T2}}(W) \leq m_{\mathrm{T2}}(W|\alpha)' \tag{13}$$

$$m_{\mathrm{T2}}(t) \leq m_{\mathrm{T2}}(t|\alpha)'. \tag{14}$$

Note that to satisfy the right inequality in (9), it is only necessary that *at most one* of $m_{\mathrm{T2}}(W|\alpha)'$ or $m_{\mathrm{T2}}(t|\alpha)'$ should be bounded above by m_W or m_t. For instance, if the properties of an event (e.g., the momenta of the leptons, the angle between the leptons and b-jets...) and the α-value chosen are such as to make $m_{\mathrm{T2}}(t|\alpha)' > m_t$, then one will find that $m_{\mathrm{T2}}(W|\alpha)' < m_W$, so that $\alpha m_{\mathrm{T2}}(W|\alpha)' + (1-\alpha)m_{\mathrm{T2}}(t|\alpha)' \leq \alpha m_W + (1-\alpha)m_t$. Accordingly, the upper bound relations

$$m_{\mathrm{T2}}(W|\alpha)' \lesssim m_W \tag{15}$$

$$m_{\mathrm{T2}}(t|\alpha)' \lesssim m_t, \tag{16}$$

are only approximately true— and whether or not the relations hold true in any given event must be distinguished on an event-by-event basis.

At last, it is at times instructive to *'normalise'* these mass variables, by dividing them by their theoretical upper bounds. In doing so, we can define the **normalised stransverse mass** quantities

$$\mu_{\mathrm{T2}}(W) = \frac{m_{\mathrm{T2}}(W)}{m_W}, \mu_{\mathrm{T2}}(t) = \frac{m_{\mathrm{T2}}(t)}{m_t} \text{ and } \mu_{\mathrm{T2}}^{\mathrm{DC}}(\alpha) = \frac{m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha)}{\alpha m_W + (1-\alpha)m_t}, \tag{17}$$

all of which are bounded above by unity.

[2]With the exception of two evident limiting cases: when $\alpha = 1$, $m_{\mathrm{T2}}(W|1)' \equiv m_{\mathrm{T2}}(W)$ and when $\alpha = 0$, $m_{\mathrm{T2}}(t|0)' \equiv m_{\mathrm{T2}}(t)$.

■ PROCEDURE

□ DATA

The investigation relies upon studying 65,100 14 TeV $t\bar{t}$ MC events, generated with MADGRAPH/HW6 and skimmed to be each composed of two tagged b-jets and two light leptons (downloadable from HepSim here). An 'ATLAS card' has been applied upon these events, in order that they might appear reconstructed from the ATLAS detector. These events are subsequently compared to the properties of 353,085 truth MC (here), skimmed to include the correct $t\bar{t}$-decays as well as $W \to e+\bar{\nu}$ or $W \to \mu+\bar{\nu}$ decays. The investigation would have been made more rigorous, by a comparison to the truth information from the 65K reconstructed ATLAS card events, but several difficulties (e.g., no event numbers, no reconstruction file) were encountered which made this information unavailable.

□ RESEARCH QUESTIONS

The research is turned specifically on these three questions:

1. *When $m_{\mathrm{T2}}(W)$ gives a poor mass estimate of m_W, is there still mass information to be salvaged? If so, can the alternative kinematic variable $m_{\mathrm{T2}}(W|\alpha)'$ be used to extract it?*
2. *The computational minimiser used to calculate m_{T2} and $m_{\mathrm{T2}}^{\mathrm{DC}}$ often supposes that the optimal transverse momentum of at least one of the substitute neutrinos is 0. Knowing this outcome to be physically improbable, we impose a pT-cut pT_{c} on the minimiser, so that the optimisation is performed under the constraint that the transverse momenta of the substitute particles exceed pT_{c}. Does there exist an ideal such pT_{c}? And what are the metrics by which such an ideal cut would be determined?*
3. *The computational minimiser used to calculate $m_{\mathrm{T2}}^{\mathrm{DC}}$ relies upon a pre-defined α value. What α value would give a 'best' mass measurement? And what are the metrics by which a 'best' mass measurement would be determined?*

We consider each question one by one. **Note that, where applicable, the figures in this document have been produced with numerical minimisers run in 'fast' mode** (*cf.* CODES, ITEM 2.). **It can be the subject of future study to establish how well the results can be improved in the alternative slower mode.**

1. QUESTION 1

   By inspecting the histogram of $m_{\mathrm{T2}}(W)$ (Fig. 3 *left*), it is clear that in a quarter or so of events, the numerical minimiser gives an unreasonably small mass estimate of $m_{\mathrm{T2}}(W) \sim 0$. It can be shown that such poor mass estimates are largely produced whenever the minimiser determines the optimal transverse momentum of at least one of the substitute neutrinos to be 0. However, the truth MC reveals this outcome to be very seldom the case (Fig. 3 *right*).

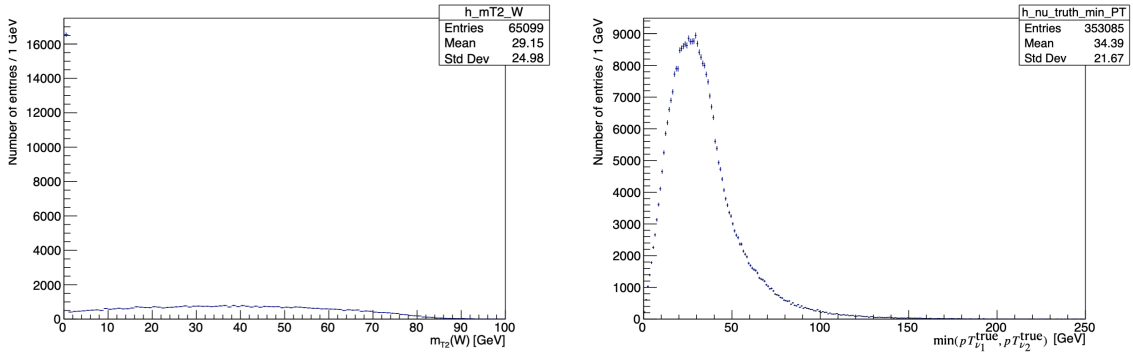

Figure 3: Histograms of $m_{\mathrm{T2}}(W)$ (*left*) and of the minimum of the true transverse momenta of the neutrinos $\min(p_{T\nu_1}^{\mathrm{true}}, p_{T\nu_2}^{\mathrm{true}})$, produced from $t\bar{t} \to WbWb \to \ell\nu b\ell\nu b$ decays (*right*).

By contrast, the new kinematic variable $m_{\mathrm{T2}}(W|\alpha)'$ appears to be able to salvage some mass information. Indeed, in Fig. 4, it is shown that, for events with an unreasonably small mass estimates

of $m_{\mathrm{T2}}(W) < 1$, the corresponding $m_{\mathrm{T2}}(W|\alpha = 0)'$ produces a much wider mass distribution, with a diminished number of entries in the first bin as well as a substantial mean and standard deviation.

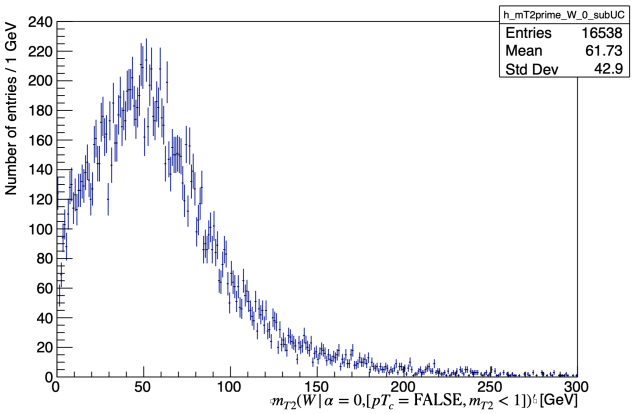

Figure 4: Histogram of $m_{\mathrm{T2}}(W|\alpha = 0)'$, calculated with the unconstrained minimiser ($p_{Tc} = \mathrm{FALSE}$) on events for which $m_{\mathrm{T2}}(W) < 1$.

With this broader mass distribution, there is perhaps a possibility of using the *template method* to reconstruct the W-mass, by generating simulated template distributions of $m_{\mathrm{T2}}(W|\alpha = 0)'$ as a function of m_W and applying a log-likelihood fit of the templates to the data distribution to extract the maximum likelihood estimator $\hat{m}_W$. What is more, in salvaging mass information, $m_{\mathrm{T2}}(W|\alpha)'$ may be more efficient a tool than $m_{\mathrm{T2}}(W)$ to use for *event categorisation*, that is, for the discrimination between two types of events on an event-by-event basis. For instance, cuts on $m_{\mathrm{T2}}(W)$ have been used in high-mass lepton-hadron analysis to categorise relevant events as being supersymmetry-like (> 120 GeV) or $t\bar{t}$-like ($t\bar{t} \to WbWb \to \ell\nu b\ell\nu b$) (20-80 GeV) with given efficiencies. If the events for which mass information was previously lost, i.e., $m_{\mathrm{T2}}(W) \sim 0$, are now distributed throughout the $m_{\mathrm{T2}}(W|\alpha)'$ distribution, similar cuts on the $m_{\mathrm{T2}}(W|\alpha)'$ variable may perhaps be used to more accurately and efficiently categorise them as being, for instance, $t\bar{t}$-like, supersymmetry-like and so forth. (For a more detailed discussion of the use of stransverse variables in event categorisation, see Ewan Hill's 2017 dissertation).

2. Question 2

Depending upon the main objective of one's investigation, there are various ways by which this question can be approached. If, for instance, one is considering backgrounds, then an ideal cut can be regarded as one for which there is the most significant signal-to-background separation in a pre-selected region. If one has truth information that can be compared to the simulated data in a 1-1 correspondence, then an ideal cut can be defined as one which produces the smallest difference between calculated kinematic quantities for the simulated and the truth data. In our case, having limited time and limited truth data, the question of an ideal cut is considered only for $m_{\mathrm{T2}}(W|\alpha = 1)'$ distributions, without regard to the truth MC and with the main objective of preserving the *shape* of the $m_{\mathrm{T2}}(W)$ distribution (or, more precisely, the $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = 0)'$ distribution, as we shall note later). Indeed, if such a shape is preserved, then the methods of analysis already developed for the stransverse variables m_{T2} might be easily and accurately applied to the newer prime variables m'_{T2}, facilitating their adoption. Accordingly, an *ideal* cut p_{Tc} applied to the numerical minimiser for the computation of $m_{\mathrm{T2}}(W|\alpha = 1)'$ is to be regarded as a positive lower bound to the transverse momenta of the substitute neutrinos, which best preserves the shape of $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = 0)'$, whilst diminishing the number of entries in the first bin to as low as number as possible.

Before passing on to a description of the procedure, a small note concerning the distinction between $m_{\mathrm{T2}}(W)$ and $m_{\mathrm{T2}}(W|\alpha = 1, pT_{c=0)'}$ is in order. Though in principle the two are equal, in practice, the use of a constrained minimiser to calculate $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = 0)'$ and the use of an unconstrained minimiser to calculate $m_{\mathrm{T2}}(W)$ produces non-negligible differences between their outcomes (Fig. 5).

To account for the differences which arise from the use of a constrained minimiser, it is, therefore, necessary to compare the shap to $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = 0)'$.

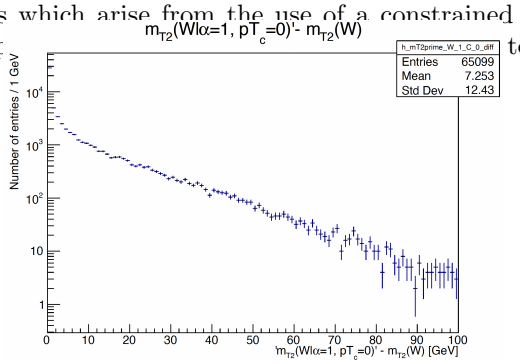

Figure 5: Histogram of $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = 0)' - m_{\mathrm{T2}}(W)$.

Knowing larger $p_{Tc} = X$ must cause the shapes of the $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = X)'$ distributions to differ more greatly from $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = 0)'$ (Fig. 6), it is reasonable to select for study a set of constraints that 'small' with respect to m_W, in our case being equal to 1, 5, 10 and 15 GeV (Figs. 7-8). It can be the subject of future to study a greater number of constraints, especially at small values between 0 and 5 GeV.

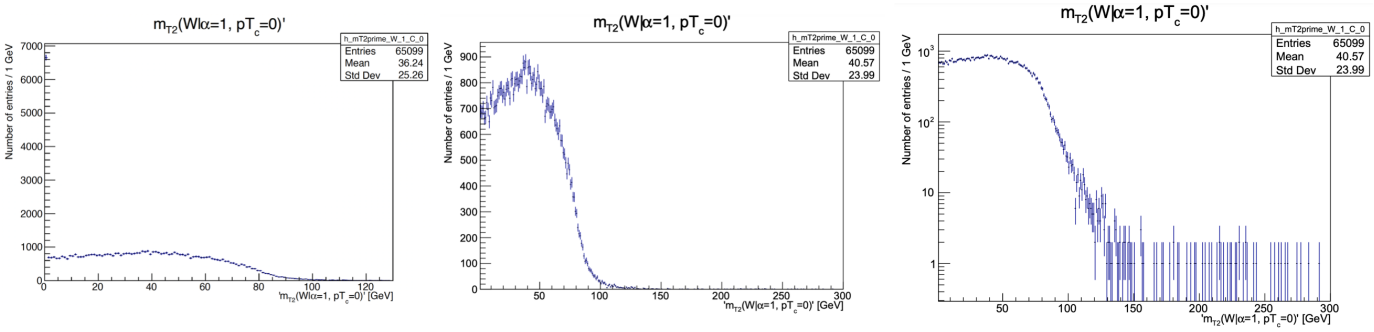

Figure 6: $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = 0)'$ distributions from [0, 140] GeV (*left*), from [1, 300] GeV on a linear (*middle*) scale and from [1, 300] GeV on a logarithmic (*right*) scale.

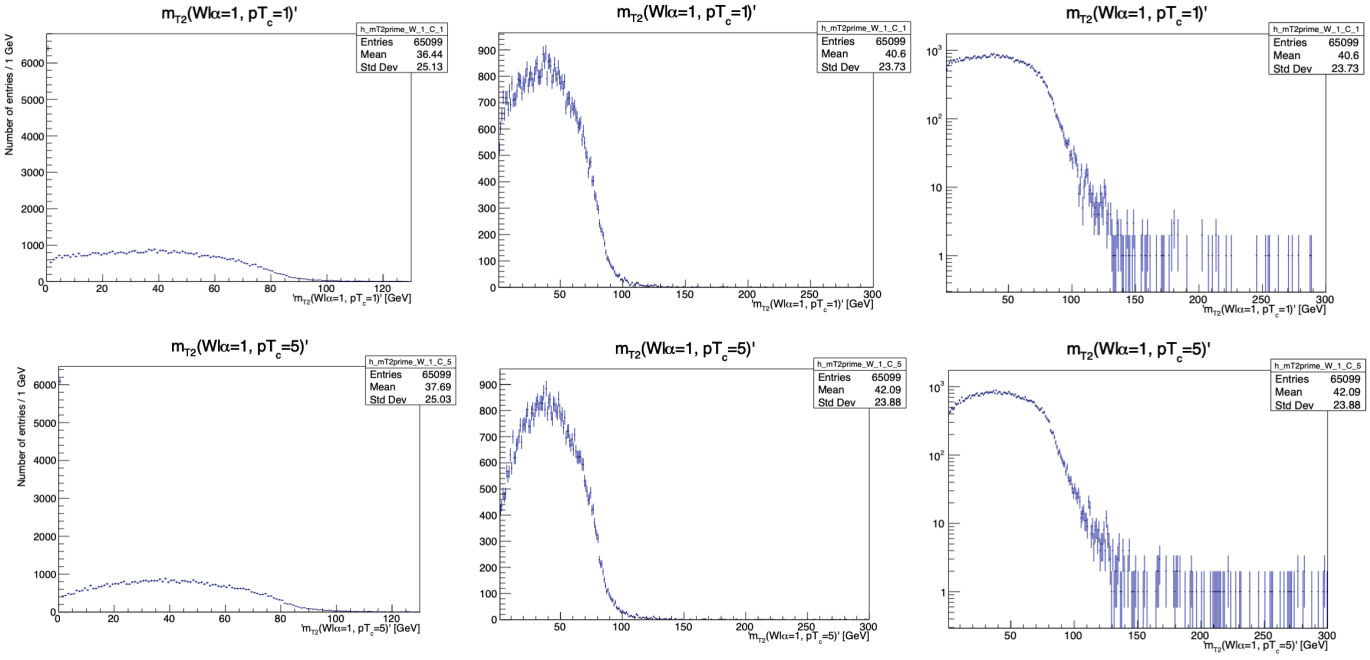

Figure 7: $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = X)'$ distributions from [0, 140] GeV (*left*), from [1, 300] GeV on a linear (*middle*) scale and from [1, 300] GeV on a logarithmic (*right*) scale, for $X = 1, 5$ GeV.

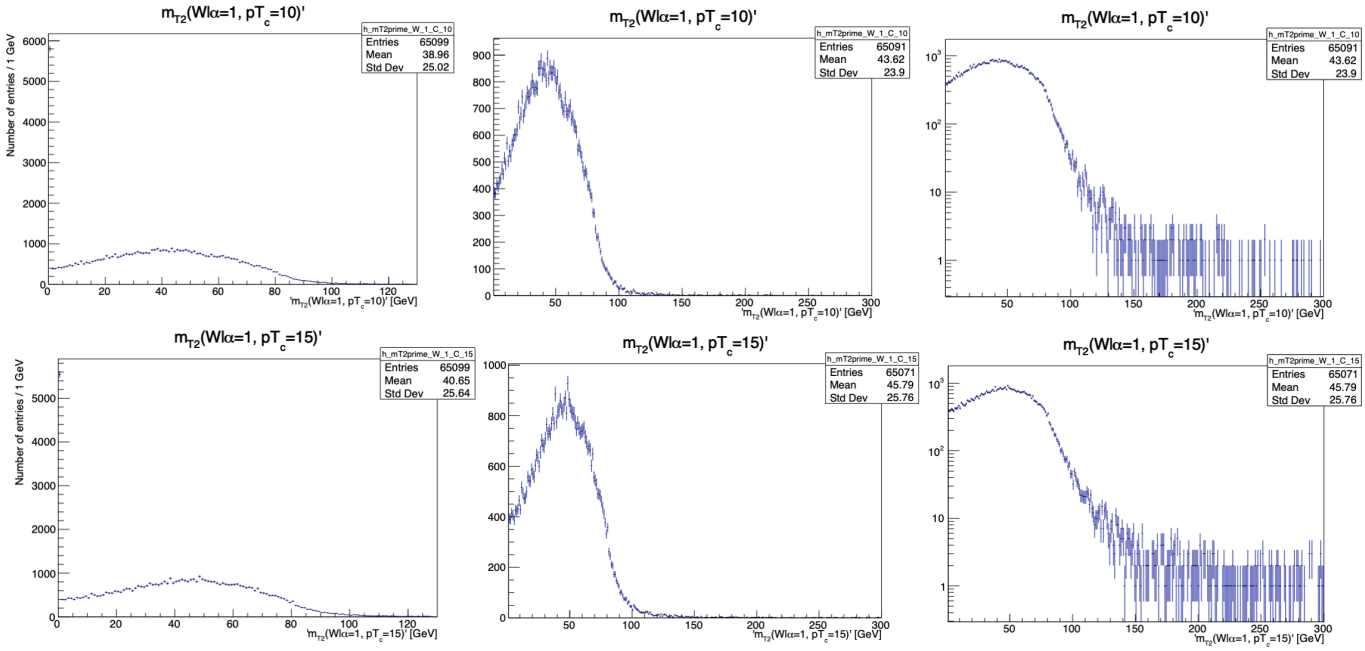

Figure 8: $m_{\mathrm{T2}}(W|\alpha = 1, p_{Tc} = X)'$ distributions from [0, 140] GeV (*left*), from [1, 300] GeV on a linear (*middle*) scale and from [1, 300] GeV on a logarithmic (*right*) scale, for $X = 10, 15$ GeV.

Counting entries in the zero bin n_0 and those in the tail n_t (at masses $> m_W$):

pT_c	n_0/n_{tot}	$n_t/n_{>1\text{GeV}}$
0	$6654/65099 = 10.2\%$	$2707/58445 = 4.7\%$
1	$6398/65099 = 9.8\%$	$3164/58701 = 5.4\%$
5	$6094/65099 = 9.4\%$	$5810/59005 = 9.8\%$
10	$5807/65099 = 8.9\%$	$8317/59292 = 14.0\%$
15	$5543/65099 = 8.5\%$	$9148/59556 = 15.4\%$

Accordingly, if the shape of the tail is the most important feature to preserve, then a constraint $\leq$ 1 GeV, which adds only a small proportion of entries to the tail, may be regarded as most optimal. If, instead, a method of analysis already developed for $m_{\text{T2}}(W)$ relies upon retaining the position of the peak, then a slightly larger constraint of ≤ 5 GeV may be used, with the additional advantage of reducing the number of entries in the 0 bin. Larger constraints of 10 or 15 GeV poorly preserve the shape of the peak and tail, rendering such values suboptimal, as they would make the adoption of the new prime variables using the frameworks developed for the stransverse variables difficult.

3. Question 3
A very preliminary method to address the question was developed for the analysis of $m_{\text{T2}}(W)'$ distributions for the mass measurement of the W-boson, but was not fully carried out on account of several difficulties encountered. This method proceeds, as follows: 21 normalised $m_{\text{T2}}(W|\alpha, pT_c = \text{FALSE})'$ histograms, each with a fixed α-value in $[0, 1]$, were produced. Three of these histograms with a low ($\alpha = 0.1$), medium ($\alpha = 0.5$) and large ($\alpha = 1$) were individually studied, so as to give reasonable conjectures of the parameters P_0, P_1, P_2, P_3, P_4 in

$$f_1(x) = (P_2 + P_3 x)\arctan((x - P_0)/P_1) + P_4 x + 1 \tag{18}$$

that it was hoped would best model the shape of their descending regions. (This function is taken from [1]). An example of how the parameters were chosen for the case of $\alpha = 0.1$ is contained in the Appendix). The fitted parameter P_0 and its error δP_0, responsible for centering the descending region, were considered to provide a mass measurement of the W boson. With the 'ideal' α, it was hoped that $P_0 \sim m_W$ and $\delta P_0/m_W \ll 1$, so that the centre of the descending region should be found as close as possible to m_W. However, in attempting to execute this method, several difficulties were encountered, the simplest and most 'fatal' of which being that the shapes of the mass distributions were simply *not* conducive to be fitted with an arctan function (Fig. 17). Whatever the specified bin size or α-value, there was no evident sequence of flattish and steep drop-off regions, characteristic of an arctan-like distribution. This lack of a proper shape was noted to be worse at smaller α-values than larger ones. Further, the quality of the fits were found to be *too* dependent upon the initial values of the parameters. A small change to these initial values, which were only very roughly estimated by eye, would significantly alter the fitted P_0 and δP_0 values, rendering it difficult to determine an 'optimal' α-value according to the above criteria.

■ Codes

The following section serves to document the purpose, functionality, inputs and outputs of the codes required for analysis, presented in the order in which they should be run. All codes and required input files are readily available for download from the GitHub Repository MT2DC_Analysis_FINAL.

1. `mt2dc_kinematic_functions.py`

- **Input ROOT File**: None
- **Output ROOT File**: None
- **Purpose**: An auxiliary file, in which are defined basic kinematic functions required in the numerical minimiser code (*cf.* 2 below). Such functions (1) extract the $x, y, z-$momenta and energy from p_T, η, ϕ and mass information, (2) calculate the transverse mass and transverse energy of single or parent particles, using scalar and 4-vector input variables and (3) extract the values of the W- and t-terms from the numerical minimisers after they have been run.

2. `mt2dc_analysis.py`

- **Input ROOT File**: 2022_03_March_07_skim_mg5_ttbar_jet_merged_001-716_ntuple_2l2b_v01.root

  N.B. The code used to calculate stransverse mass m_{T2} values are taken from the `lester_mt2_bisect.h` code, available from a wider code repository at the Oxbridge Kinetics Library.
- **Output ROOT File**: mt2dc_analysis_output.root
- **Output Directory:** /User/.../Documents/mT2DCAnalysisPlots/
- **Purpose**: The main analysis code, which calculates the kinematic variable $m_{\mathrm{T2}}^{\mathrm{DC}}$ for ∼65,000 simulated $t\bar{t}$ decay events using constrained and unconstrained numerical minimisers and a specified list of 21 α-values. Using information concerning side 1 and 2 leptons, b-jets and missing transverse energy from the input ROOT file, the minimisers calculate the $m_{\mathrm{T2}}^{\mathrm{DC}}$, $\mathrm{m_{T2}}(W|\alpha)$' and $\mathrm{m_{T2}}(t|\alpha)$' mass values and the p_Ts of the substitute invisible particles for every event.
- **Functionality**:
  - This code may be run in one of two computational speeds, 'fast' or 'slow.'
    - Fast: For every event, 21 conjectured side 1 lepton p_Ts are inputted into the objective function. The one which outputs the smallest *function value* is then used as the initial guess for minimisation.
    - Slow: For every event, these 21 conjectured p_Ts are inputted into the numerical minimiser, with the one producing the smallest *minimised output value* being retained for the final analysis.

    For efficiency, all histograms in this document are produced in 'fast' mode.
  - The constrained minimiser is subject to the requirement that the p_Ts of the substitute particles be greater than or equal to some p_T threshold, specified by the user (Fig. 9). The unconstrained minimiser uses no constraints whatsoever. Note there is a difference between using a pT threshold of 0 ($pT_c = 0$) and using no p_T threshold at all (pT_c = FALSE), as the former is still considered a constrained minimisation.

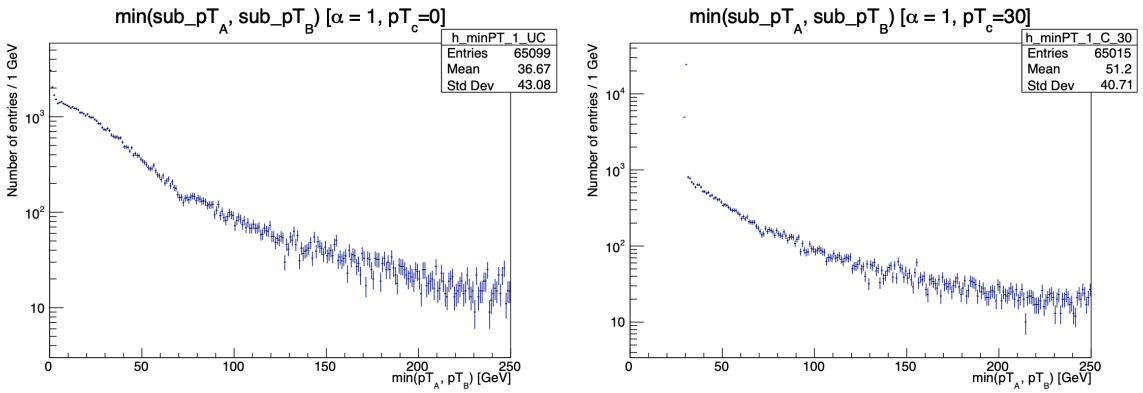

Figure 9: Minimum of the substitute neutrino pT in the case of $\alpha = 1, pT_c = 0$ (*left*) and $\alpha = 1, pT_c = 30$ GeV (*right*). In the right histogram, the number of events has diminished, as the constraint causes certain minimisations to fail.

  - Both minimisers use `scipy.optimize.minimize` in `SLSQP` mode. Of all the possible methods, `SLSQP` is chosen, as it requires no complicated Jacobian and as it produces $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0) - m_{\mathrm{T2}}(t)$ and $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 1) - m_{\mathrm{T2}}(W)$ (Fig. 14) values that are generally smaller (or more negative) than those produced with other methods. Whenever `SLSQP` fails to minimise the constrained or unconstrained objective function within 2000 iterations, a second mode, `COBYLA`, is called. If both methods fail, the iteration is flagged as a failure, so that the event may be later discarded.

3. `mt2dc_TH2F_plotter.py`

- **Input ROOT File**: mt2dc_analysis_output.root
- **Output ROOT File**: TH2F_plotter_output.root
- **Output Directory:** /Users/.../Documents/styledPlotsOutputs/
- **Purpose:** A plotting code, which takes the optimisation results from the analysis code and stores them as entries of the form (α, result) in TH2F histograms.
- **Functionality:**
  - The code 'cuts on success', so that only successful optimisation results are stored, and also sorts the constrained and unconstrained results into separate histograms.

4. `mt2dc_fitting_analysis.py`

- **Input ROOT File:** TH2F_plotter_output.root
- **Output Files:** mt2dc_fitting_analysis_output.root,
  /Users/.../Documents/styledPlotsOutputs/fit_functions_parameters.txt,
  /Users/.../Documents/styledPlotsOutputs/mT2prime_W_UC_f1_Parameter0.txt,
  /Users/.../Documents/styledPlotsOutputs/mT2prime_W_UC_f2_Parameter0.txt,
  /Users/.../Documents/styledPlotsOutputs/mT2prime_W_UC_f3_Parameter0.txt.
- **Output Directory:** /Users/.../Documents/styledPlotsOutputs/fit_functions/
- **Purpose:** The fitting code, which fits three pre-defined functions of the form (15) to the drop-off regions of 21 normalised $m_{\mathrm{T2}}(W|\alpha,\, p_{Tc} = \mathrm{FALSE})$' histograms, each of fixed α. Various data, including the fitted parameters, number of degrees of freedom, chi-square and derivatives, are saved to the first .txt file. The value of the fitted parameter P_0 and its error δP_0 for fits 1 (f1), 2 (f2) and 3 (f3) are respectively saved to the last three .txt files.
- **Functionality:**
  - The code projects the TH2F histogram `h2_mT2prime_W_UC`, in which are stored entries of the form (α, $m_{\mathrm{T2}}(W|\alpha,\, p_{Tc} = \mathrm{FALSE})$'), along the y-axis to create 21 normalised TH1F histograms, each of fixed α.
  - The initial values of the parameters are chosen according to what appears reasonable by inspection of the shapes of the graphs (*cf.* Appendix). The parameters of fit 1, 2 and 3 are chosen by inspecting the histograms for $\alpha = 0.1$, $\alpha = 0.5$ and $\alpha = 1.0$, respectively.

5. `mt2dc_alpha_plotter.py`

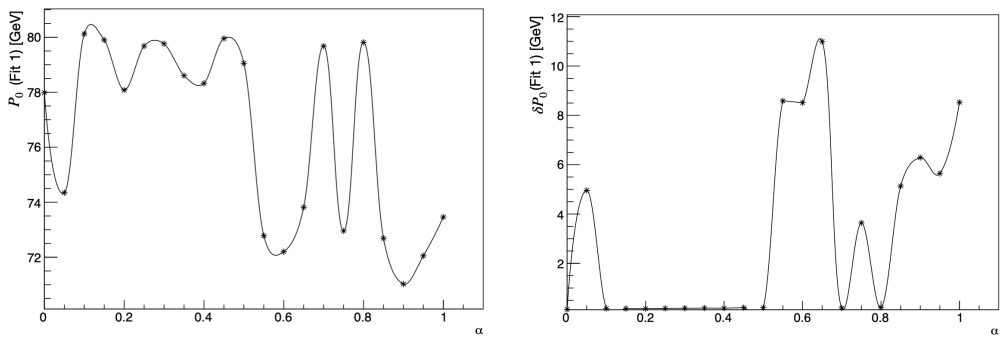

Figure 10: Graphs of P_0 (*left*) and δP_0 (*right*) against α for fit 1.

- **Input Files:** /Users/.../Documents/styledPlotsOutputs/mT2prime_W_UC_f1_Parameter0.txt,
  /Users/.../Documents/styledPlotsOutputs/mT2prime_W_UC_f2_Parameter0.txt,
  /Users/../Documents/styledPlotsOutputs/mT2prime_W_UC_f3_Parameter0.txt.
- **Output ROOT File:** mt2dc_alpha_plotter_output.root

- **Output Directory:** /Users/.../Documents/styledPlotsOutputs/fit_functions/
- **Purpose:** A plotting code, which produces graphs used for the study of which α gives the best mass measurement.
- **Functionality:**
  - Currently, for every fit, the values of P_0 and δP_0 are plotted as functions of α (Fig. 10).

The following two codes are also useful during any step of the analysis.

6. `mt2dc_truth_analysis.py`

- **Input ROOT File:** 2022_08_Aug_02_truthSkim__mg5_ttbar_jet_001-716_v01.root
- **Output ROOT File:** truth_analysis_output.root
- **Output Directory:** /Users/.../Documents/truthAnalysisPlots/
- **Purpose:** This code makes various truth histograms from a ROOT file of $\sim 350,000$ 14 TeV MC $t\bar{t}$ events. These plots may be compared to those produced using the `mt2dc_analysis` code, by means, for instance, of overlays (*cf.* `mt2dc_overlay_plotter`).
- **Functionality:**
  - This script can be used to make truth plots not directly found in the truth file. Currently, code is written to generate plots of the minimum p_T of the side 1 and 2 neutrinos and the maximum of the transverse masses of the side 1 and 2 W bosons.

7. `mt2dc_overlay_plotter.py`

- **Output Directory:** /Users/.../Documents/styledPlotsOutputs/
- **Purpose:** This code takes in two histograms from specified ROOT files in order to produce an overlay of them.

■ Important Plots

This section contains important plots, which either illustrate the functionality of the codes or demonstrate that the theoretical relations between kinematic variables (e.g., inequalities 4, 9, 11) have been met.

1. Comparison of $m_{\mathrm{T2}}^{\mathrm{DC}}$ and m_{T2}

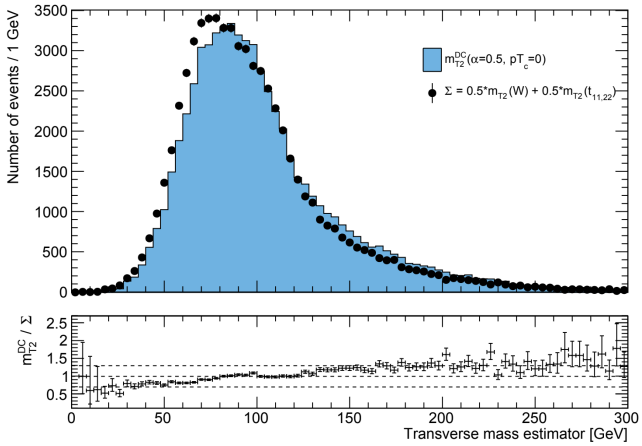

Figure 11: Overlay of $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0.5, pT_c = 0)$ and $\Sigma = 0.5m_{\mathrm{T2}}(W) + 0.5m_{\mathrm{T2}}(t_{11,22})$.

- Observe the distributions of Σ and $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0.5, p_{Tc} = 0)$ in Fig. 11 to be similar in shape, with the former shifted slightly to the left. These observations are consistent with inequality (9), which asserts $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0.5, pT_c = 0) = 0.5m_{\mathrm{T2}}(W|0.5)' + 0.5m_{\mathrm{T2}}(t|0.5)' \geq \Sigma$. Any event for which this inequality is not satisfied must owe to limitations in the resolution of the data.

- The subscript $t_{11,22}$ in $m_{\mathrm{T2}}(t_{11,22})$ indicates that the highest-p_T lepton 1 and highest-p_T b-jet 1 have been paired together, and the lowest-p_T lepton 2 and lowest-p_T b-jet 2 have been paired together in the calculation of $m_{\mathrm{T2}}(t_{11,22})$. As a matter of fact, it not known which lepton should be paired with b-jet, so the t-mass could have been instead calculated with the pairing $12, 21$, denoted $m_{\mathrm{T2}}(t_{12,21})$.

2. Comparison of m_{T2}, m'_{T2} with Truth Information

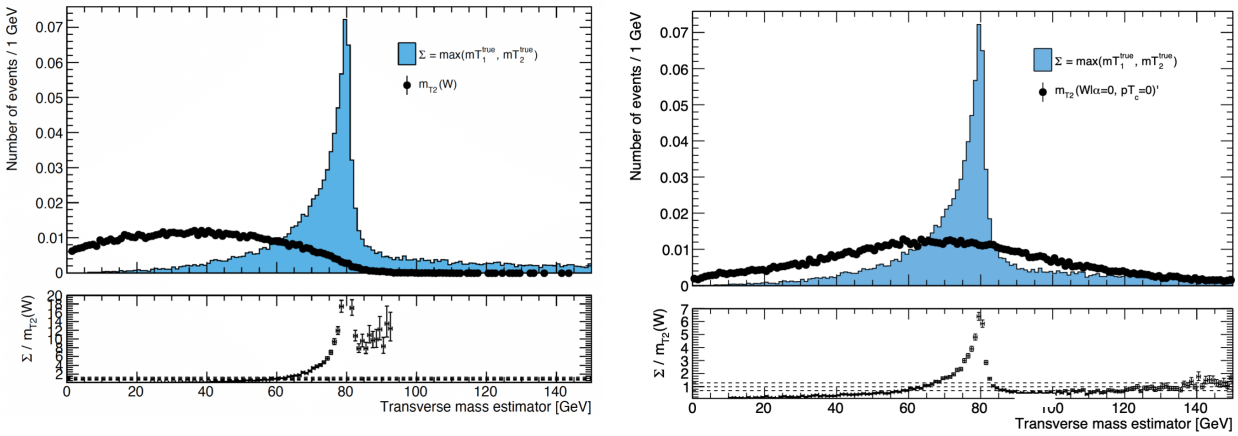

Figure 12: $\Sigma = \max(m^{\mathrm{true}}_{\mathrm{T},\nu_1}, m^{\mathrm{true}}_{\mathrm{T},\nu_2})$ and $m_{\mathrm{T2}}(W)$ (*left*); Σ and $m_{\mathrm{T2}}(W|\alpha = 0, pT_c = 0)'$ (*right*).

Fig. 12 compares primed and unprimed stransverse variables to $\max(m_{\mathrm{T},\nu_1}^{\mathrm{true}}, m_{\mathrm{T},\nu_2}^{\mathrm{true}})$, which represents the maximum of the true transverse masses of neutrinos 1 and 2 calculated with the truth data. In this figure, observe:

- The Σ-distribution generally satisfies the inequality $\Sigma \leq m_W \simeq 80.4$ GeV, consistent with (4).
- The $m_{\mathrm{T2}}(W)$-distribution in the left panel generally satisfies $m_{\mathrm{T2}}(W) \leq \Sigma$, consistent with (4).

  Wherever this inequality (4) is not strictly obeyed must owe to limitations in the minimiser or the resolution of the data.
- The $m_{\mathrm{T2}}(W)$- and $m_{\mathrm{T2}}(W|\alpha = 0, pT_c = 0)'$-distributions in the left and right panels generally satisfy $m_{\mathrm{T2}}(W) \leq m_{\mathrm{T2}}(W|\alpha = 0, pT_c = 0)'$, consistent with (13). In addition to theory, there are perhaps two more reasons for which this inequality holds. First, if $\alpha = 0$, the W-term in the expression of $m_{\mathrm{T2}}^{\mathrm{DC}}$ vanishes, allowing the numerical minimiser to 'focus' on determining substitute pT values which specifically diminish the t-term (rather than the W-term) to as small a value as possible. Second, the optimisation is performed by supposing lepton 1 and b-jet 1 as well as lepton 2 and b-jet 2 to be paired together, which may not always be the case in practice. It may be instructive, in the future, to incorporate the possibility of the pairing $12, 21$ in the optimiser or to determine a means to select between the pairings $11, 22$ and $12, 21$ with good efficiency, to see whether its mass outputs are generally diminished.
- The relation $m_{\mathrm{T2}}(W|\alpha = 0, p_{Tc} = 0)' \lesssim \Sigma$ is more loosely obeyed, as there is no strict theoretical inequality between the two variables. Indeed, the most which can be said from inequalities (4) and (13) is that $m_{\mathrm{T2}}(W|\alpha = 0, pT_c = 0)' \lesssim m_W \sim \Sigma$.

3. Correlation Between Substitute Transverse Momentum and $m_{\mathrm{T2}}^{\mathrm{DC}}$

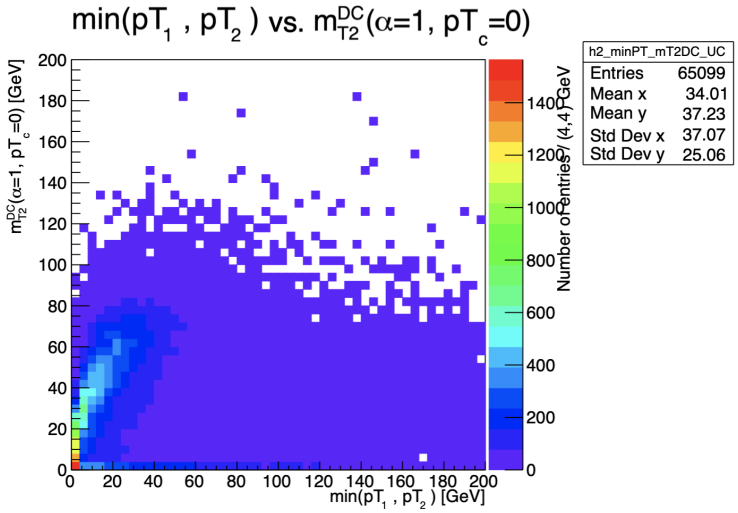

Figure 13: TH2F histogram of $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 1, p_{Tc} = 0)$ against $\min(pT_1, pT_2)$.

- In Fig. 13, observe that, in 25% of $t\bar{t}$ events, $m_{\mathrm{T2}}^{\mathrm{DC}} \sim 0$ arises from (at least) one of the substitute neutrinos having $p_T \sim 0$.
- Such a spike represents the failure of the minimiser to make a reasonable estimate of the transverse momentum pair of the neutrinos and, therefore, of the transverse mass of the W-boson. Accordingly, to diminish the number of events for which this unreasonable estimate is made, it is useful to impose pT-constraints, which require that the transverse momentum of each substitute neutrino should be greater than some chosen positive number.

4. Functionality of the Numerical Minimiser

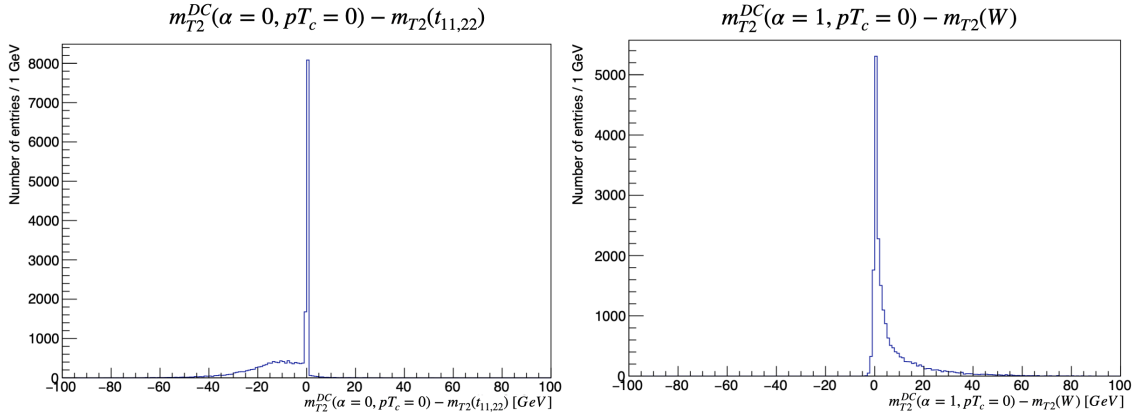

Figure 14: Histograms of $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0, pT_c = 0) - m_{\mathrm{T2}}(t_{11,22})$ (*left*) and $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 1, pT_c = 0) - m_{\mathrm{T2}}(W)$ (*right*).

- When the numerical minimiser was developed, its functionality was tested by determining whether it succeeded in producing 'small' differences, $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha) - [\alpha m_{\mathrm{T2}}(W) + (1-\alpha) m_{\mathrm{T2}}(t)]$, for $\alpha \in \{0, 1\}$. Various types of minimiser modes were tried, the number of iterations and the required precision of the solution were varied, and various initial side 1 neutrino pT guesses were implemented, until a combination could be found, which diminished these differences, but did not excessively lengthen the computation time. In Fig. 14, the differences between $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0)$ and $m_{\mathrm{T2}}(t_{11,22})$ (*left*) or $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 1)$ and $m_{\mathrm{T2}}(W)$ (*right*) are plotted, with a numerical minimiser run in `SLSQP` and `COBYLA` modes, constrained to a maximum number of iterations of `2000` and implemented with one of 21 initial side 1 neutrino pT guesses (selected differently for every event, according to the 'fast' mode algorithm).
- The extent to which the leftmost inequality in (9) is obeyed depends upon the α-value. When $\alpha = 0$, a proportion of $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0)$ values are *smaller* than $m_{\mathrm{T2}}(t_{11,22})$, inconsistent with (9). When $\alpha = 1$, however, $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 1)$ is generally larger than $m_{\mathrm{T2}}(W)$, consistent with the expected inequality. Does this inconsistency, in large part, owe to limited resolution in the data; or must the minimiser be subject to future revision, in order that it should generally produce $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha) - [\alpha m_{\mathrm{T2}}(W) + (1-\alpha) m_{\mathrm{T2}}(t)] > 0$, for any arbitrary α?

5. Comparison of m'_{T2} and m_{T}

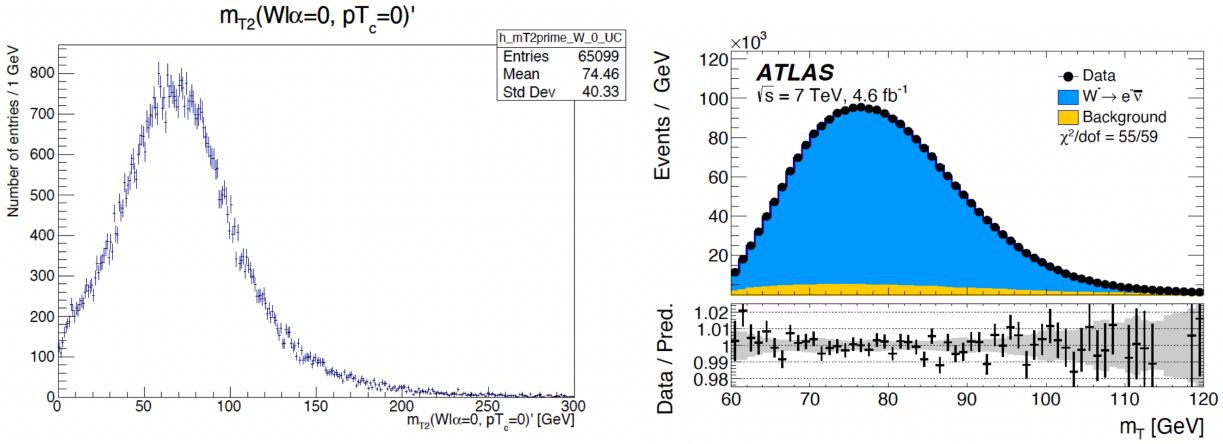

Figure 15: Histograms of $m_{\mathrm{T2}}(W|\alpha = 0, pT_c = 0)'$ (*left*) and $m_{\mathrm{T}}(W)$ (right) [2].

- In Fig. 15, the distribution $m_{\mathrm{T2}}(W|\alpha = 0, pT_c = 0)'$ (*left*) panel resembles $m_{\mathrm{T}}(W)$ (*right*), calculated for the single-neutrino decay $W \to e{+}\bar{\nu}$, excepting a peak shifted slightly leftwards and a more extended tail. The similarity between the two distributions is desirable, as the m_{T2} variables are attempting to reproduce the characteristics of m_{T} for a di-neutrino events.

6. Calculation of Stranverse Mass Variables for $\alpha = 0.5$.

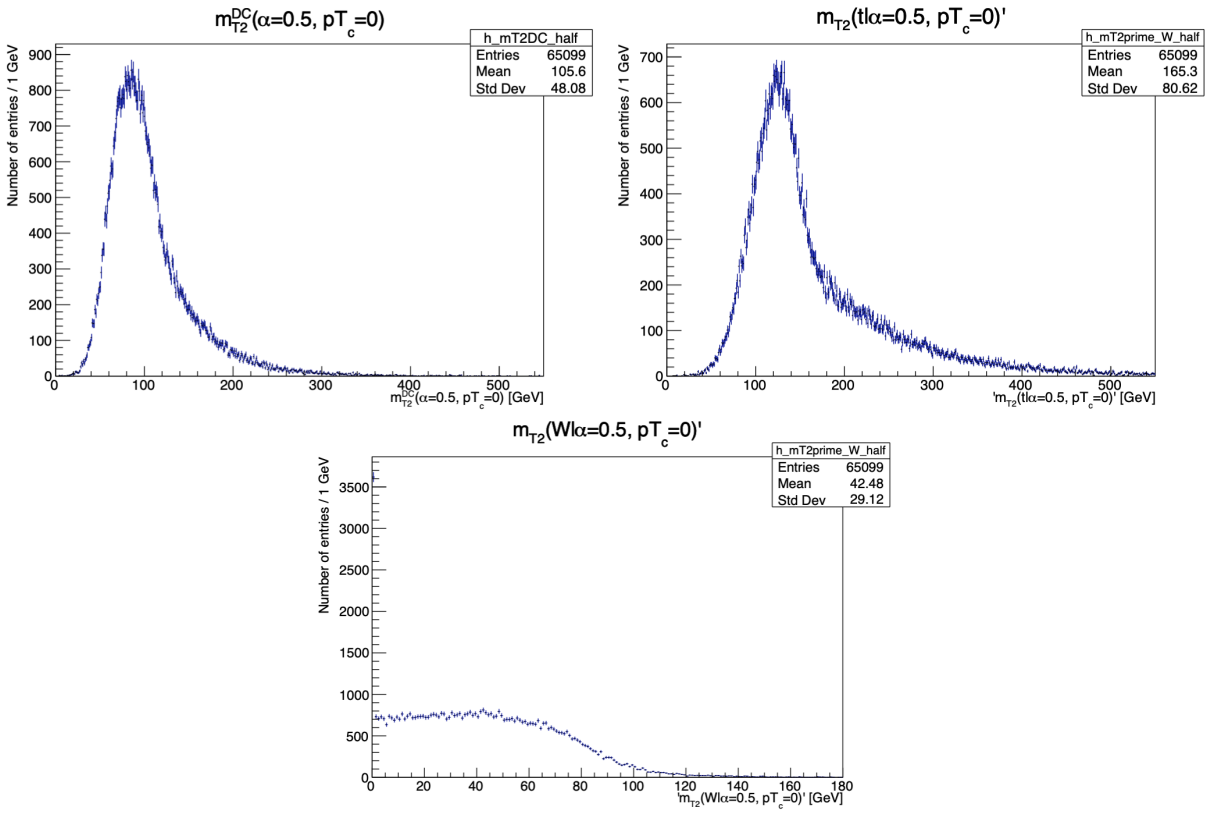

Figure 16: Histograms of $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0.5, pT_c = 0)$ (*top left*), $m_{\mathrm{T2}}(t|\alpha = 0.5, pT_c = 0)'$ (*top right*) and $m_{\mathrm{T2}}(W|\alpha = 0.5, pT_c = 0)'$ (*bottom*).

- In Fig. 16, every event in the $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0.5)$ histogram is a weighted average of the corresponding event in the top-prime and W-prime histograms, according to $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0.5) = 0.5 m_{\mathrm{T2}}(t|\alpha = 0.5)' + 0.5 m_{\mathrm{T2}}(W|\alpha = 0.5)'$. As the t-quark has a much larger mass of $m_t \sim 172.8$ GeV and as the resolution with which the b-jets are measured is much poorer than that of the leptons, the point of drop-off and tail of the top-prime distribution are much extended with respect to the W-prime distribution and so largely govern the shape of the $m_{\mathrm{T2}}^{\mathrm{DC}}$ distribution at larger masses.
- The approximate upper limits of the particle masses that can be extracted from these distributions appear to all be reasonable: the points of drop-off of the W- and top-prime distribution are near their true particle masses (*i.e.,* near 80 and 170 GeV, respectively) and that of the $m_{\mathrm{T2}}^{\mathrm{DC}}(\alpha = 0.5)$-distribution appears to be situated close to the average of the two (*i.e.,* near 130 GeV).

References

[1] Abudinén, C. *et al.* (Belle II Collaboration). τ lepton mass measurement at Belle II (2021). arXiv:2008.04665v3 [hep-ex]

[2] Aaboud, M. *et al.* Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Eur. Phys. J. C **78**, 110 (2018). https://doi.org/10.1140/epjc/s10052-017-5475-4

■ Appendix

The following section serves to document the procedure by which the initial parameter values of the fit function (18) are chosen to best model the shape of $m_{\mathrm{T2}}(W|\alpha = 0.1,\ \mathrm{pT}_c = \mathrm{FALSE})$'). Such a fitting was carried out, notwithstanding there to be no evident sequence of flattish and drop-off regions, characteristic of an arctan function (Fig. 17).

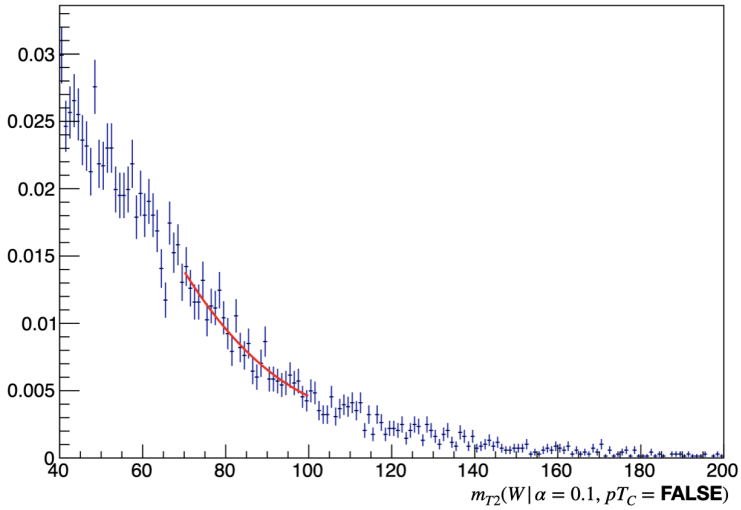

Figure 17: Normalised histogram of $\mathrm{m}_{\mathrm{T2}}(W|\alpha = 0.1,\ \mathrm{pT}_c = \mathrm{FALSE})$'. In red is the fitted function (18), whose initial parameters are chosen according to the below procedure.

1. The mass bins from [0, 40] are omitted, in order to dispense with the large spike at 0 and to more closely study the descending region around $m_W \sim 80$ GeV. The histogram is then normalised.
2. To conjecture reasonable parameters for the first fit function $f_1(x) = (P_2 + P_3 x)\arctan((x - P_0)/P_1) + P_4 x + 1$, we first consider a simpler fit $g(x) = a\arctan((x - h)/k) + C$. For this simplification, observe:

   - h governs the horizontal shift of the graph. Wishing the drop to be anti-symmetric about $m_W \sim 80$ GeV, let $P_0 = 80$.
   - C governs the vertical shift of the graph. When $x \sim 80$, $g(80) \sim 0.0075$.
   - a, k govern the vertical and horizontal stretch of the graph. To obtain the proper shape of arctan, $a < 0$. Here, observe the shape of arctan to diminish from $(70, 0.0125)$ to $(100, 0.005)$. Accordingly, one good initial guess might be $(a, k) \simeq (-0.05, 95)$. (One may play around with this Desmos graph to satisfy oneself that this is the case).

   Next, observe of the original function:

   - P_0 plays the role of h, so $P_0 = 80$.
   - P_1 plays the role of k, so $P_1 = 95$.
   - For large $x > 0$, recall arctan to approach a constant value (namely, $\pi/2$). Accordingly, for large $x > 0$, $f_1(x) \rightsquigarrow \{(P_2 + 1) + (P_3 + P_4)x\}$. Wishing for flat tails after the drop-off, $(P_3 + P_4) \sim 0$. Further, as $f_1(80) = 0 + P_4(80) + 1 \sim 0.0075$, we can deduce that $P_4 = -0.0124$ and $P_3 = 0.0124$.
   - As P_3 is small, a possible approximation to P_2 may be that $P_2 \sim a = -0.05$.