CDF/ANAL/TOP/CDFR/8173
Version 1.1
October 17, 2006

Update on the Top Mass Measurement in the Lepton + Jets Channel using a Multivariate Method (MTM2)

John Freeman, Angela B. Galtieri, Paul J. Lujan, Jeremy Lys, Pedro Movilla Fernández, Jason Nielsen, Igor Volobouev
Lawrence Berkeley National Laboratory

Abstract

We update the top quark mass measurement reported at the summer 2004 conferences using a multivariate template method. This update improves upon that analysis by using a matrix element integration method for the signal. This allows us to use jet transfer functions. These functions carry more information than the average jet corrections used in the previous analysis. We use a multivariate variable to distinguish signal from backgrounds. We perform a 2D likelihood fit of M_{top} and JES, an overall factor that scales all jet energies. We report results from this method for the 940 pb^{-1} sample, using only events with 4 tight jets and 0 loose jets. Requiring $\geq$ 1 b-tagged jet, 116 events pass all the cuts. We find M = xxx.xx $\pm$ y.yy(stat) $\pm$ z.zz(JES) $\pm$ s.ss(sys).

Contents

1 Introduction

This note describes an update of the measurement of the top mass in the lepton plus jets topology made using the multivariate template method [1].

It has been known since Run I [2] that the corrections which bring the measured jet momenta to the parton level exhibit non-Gaussian shapes. In Run I we corrected the jets by taking the median of the distributions at different measured jet P_T and applying the corrections to all jets of that P_T, thus ignoring the presence of very large tails [3]. The Run II jet corrections follow the same methodology as for Run I, so they provide average corrections [4]. To improve resolution in the mass measurement it is desirable to take into account the tails of these distributions. This is done through the use of "transfer functions." The event-fitting method that we used for our previous result [1] did not allow non-Gaussian distributions, so we could not find any convenient way to redefine our fit χ^2 using transfer functions and still keep the advantage of standard linear algebra manipulations. Instead, we have switched to the integration over the $t\bar{t}$ matrix element and phase space. This integration fits naturally with the introduction of transfer functions. Thus, for each event, we compute a likelihood as a function of top mass under the assumption that the event comes from $t\bar{t}$ production. This method also provides natural permutation weighting by the phase space and matrix element. Previously [1] we were weighting each permutation by some incomplete matrix element, both for W decay amplitudes and $t\bar{t}$ spin correlations. We call this new analysis Multivariate Top Mass (MTM2).

Our previous analysis [1] introduced the concept of incorporating systematic errors due to uncertainties on the jet energies into the statistical uncertainties, by fitting the W mass in each event. A new factor by which to multiply the jet energy scale, JES, was found for each possible permutation in each event. This factor was used to alter the other two jets in the event before the fit to $t\bar{t}$ production was made. Eventually, we determined a parameter σ_s, which optimized the statistical uncertainty of the available sample of events. We now use a 2D likelihood of M(top) and JES, an overall factor that multiplies all jet, to optimize the mass measurement. This method has already been used by the TMT and the MEAT [12] analyses.

To suppress background contributions, we employ additional variables besides the top mass. Since these variables are designed to be independent of the top mass, we simplify the analysis by multiplying likelihoods rather then generating multivariate templates. We do not calculate specific matrix elements for background, as it is difficult to find a way to calculate them for QCD.

The ME integration method is discussed in references [5] and [6]. It was first used by the DØ collaboration [7], published in [8]. In Run II, several CDF analyses have used different versions of this method: DLM [9], MEAT [12, 13], and, for the dilepton case, MADCOW [14].

In section 2 we discuss the data samples. In section 3 we give an overview of the method used. In section 4 we discuss the derivation, parameterization and testing of the transfer functions. In section 5 we describe the matrix element used and the integration over phase space. In Section 6 we discuss the choice of kinematic variables used to separate background from signal. In section 7 we discuss how to handle the background information in the new scheme, test the method on MC events and discuss the statistical uncertainty (including the effect of the Δ_{JES} parameter) expected for the 940 pb^{-1} sample. In section 8 we present the results using the data, with statistical uncertainties only. In Section 9 we discuss the systematic uncertainties. Finally, in Section 10, we summarize our findings.

2 Data Samples

We discuss here the data sample used in this analysis. Appendix A gives details of the Gen6 challenge and the validation plots requested by the top mass group. We also show the expected background calculated by scaling the Gen5 results by the appropriate luminosity ratios.

In this analysis we restrict the top mass candidates to events with a W (decaying into a lepton and neutrino) and 4 tight jets ($E_T > 15$ GeV at level 4 and $|\eta| < 2$). We make the additional requirement that there be no loose jets in the event, where a loose jet is defined as a jet with 8 GeV $< E_T <$ 15 GeV with $|\eta| < 2$, so only jets with $E_T <$ 8 GeV at any value of $|\eta|$ are added to the unclustered energy to calculate the recoil to the $t\bar{t}$ system. We are making some studies on how to incorporate events with $N_{\text{loose jets}} > 0$.

Table 1 shows a summary of the candidate top events observed in the Gen5 sample of 318 pb^{-1}, the Gen6 winter sample of 362 pb^{-1}, the sample added for summer conferences (250 pb^{-1}) and the total sample of 940 pb^{-1}. Standard top group cuts are applied to the data before 4 tight jets are required. For the present analysis we also require that no loose jets are present.

Table 1: Top candidate events with only 4 tight jets found in the Gen5 and the Gen6 data samples. For each sample the candidates with four and only four tight jets are shown, as well as the subset with the $N_{\text{loose jets}}$=0. requirement. The last row shows the number of candidate events for each category and the totals.

detector	0d		0h		0i		Totals	
	N_j=4	N_l=0	N_j=4	N_l=0	N_j=4	N_l=0	N_j=4	N_l=0
CEM	40	25	38	24	32	26	110	75
CMUP	20	11	19	14	13	10	52	35
CMX	6	3	5	2	6	1	17	6
Totals	66	39	62	40	51	37	179	116

To calculate our backgrounds, we start from the backgrounds used by the DLM top mass analysis in the W+4 tight jets sample [9]. We use the values reported in Table

I of the PRD publication [10] which used 318 pb^{-1} of data. These values were derived using the technique employed in the $t\bar{t}$ cross section measurements with 162 pb^{-1} [15], published in [16] . The estimated backgrounds for 318 pb^{-1}, scaled to 940 pb^{-1}, and with the additional requirement of $N_{\text{loose jets}} = 0$ are shown in Table 2. The expected number of top events is evaluated using the theoretical cross-section.

Table 2: Backgrounds for the W+4 tight jet sample used by the DLM analysis. The backgrounds are rescaled for 940 pb^{-1} integrated luminosity. The third column shows the values in the second column rescaled by a fraction of 0.648, to account for the effect of our loose jet cut.

Background	318 pb^{-1}	940 pb^{-1}	0 loose jets
non-W QCD	3.07 ± 1.06	9.08 ± 3.13	5.88 ± 2.18
W + light (mistag)	2.27 ± 0.45	6.71 ± 1.33	4.34 ± 1.00
diboson (WW, WZ, ZZ)	0.39 ± 0.08	1.15 ± 0.24	0.75 ± 0.18
Sum of above 2	2.66 ± 0.53	7.86 ± 1.57	5.09 ± 1.18
W $b\bar{b}$	1.70 ± 0.79	5.03 ± 2.33	3.25 ± 1.24
W $c\bar{c}$, Wc	1.31 ± 0.63	3.87 ± 1.86	2.51 ± 1.21
Single top	0.41 ± 0.09	1.21 ± 0.27	0.78 ± 0.20
Sum of above 3	3.43 ± 1.41	10.11 ± 4.16	6.54 ± 2.80
Total Background	9.16 ± 1.83	27.05 ± 5.41	17.51 ± 4.05
Expected top (m_t=175)	42 ± 5	124 ± 15	82 ± 10
Events observed	63	179	116

To estimate the expected number of events with 0 loose jets we use the data, *i.e.*, we use a ratio of 116/179, which is the ratio of observed events with $N_{\text{loose jets}} = 0$ and the total number of events, respectively. This gives a ratio $N_{\text{nloose}=0}/N_{\text{total}} = 0.648 \pm 0.075$, which has a 12% uncertainty. We assume that this ratio is the same for both top and background events. We find that the total number of background events is $N_{\text{background}} = 17.5 \pm 4.1$ events.

We make extensive use of Monte Carlo events, for testing our method and for validation of the data (see A). The samples used are shown in Table 3.

3 Overview of the MTM2 Method

For each event we obtain a likelihood distribution as a function of the top pole mass, and a jet energy scale parameter (S) using the following expression:.

Table 3: Monte Carlo samples used

Sample Name	Description	Events	≥ 0 LJ	$= 0$ LJ
ttop4g	Herwig $t\bar{t}$, $m_t = 160$	382757	9057	5851
ttop7g	Herwig $t\bar{t}$, $m_t = 167.5$	385251	9723	6327
ttopag	Herwig $t\bar{t}$, $m_t = 175$	354003	9166	5946
ttopdg	Herwig $t\bar{t}$, $m_t = 182.5$	383003	10458	6803
ttopgg	Herwig $t\bar{t}$, $m_t = 190$	384216	10850	6998
atop7a	$W \to e\nu + 4p$ (mistag)	243427	138	83
atopfb	$W \to \mu\nu + 4p$ (mistag)	287271	160	89
atoppb	$W \to e\nu + b\bar{b} + 2p$	235221	546	320
atopjb	$W \to \mu\nu + b\bar{b} + 2p$	239255	537	354
atoptb	$W \to e\nu + c\bar{c} + 2p$	193991	150	112
atopmb	$W \to \mu\nu + c\bar{c} + 2p$	254511	160	101
atopkc	$W \to e\nu + c + 3p$	299172	256	152
bhel0d_ni	non-iso e data	1255715	54*	29*
bhmu0d_ni	non-iso μ data	552401	63*	30*

* – no tag required

$$L(\vec{y} \mid m_t, \vec{x}, S) = \frac{1}{N(m_t)} \int \frac{1}{A(m_t, S)} \frac{f(z_1) f(z_2)}{FF} \, TF(\vec{y}(S) \mid \vec{x}) \, |M(m_t, \vec{x})|^2 \, d\Phi(\vec{x}) \tag{1}$$

where $\vec{y}$ are the quantities we measure (the charged lepton and jet momenta); $\vec{x}$ are the parton-level quantities that define the kinematics of the event; $N(m_t)$ is a global normalization factor; S stands for JES, an overall factor that multiplies the momenta of each jet; $A(m_t, S)$ is the acceptance for top mass m_t and a given value of S; $f(z_1)$ and $f(z_2)$ are the momentum probability distributions, obtained from the PDFs, for incoming partons z_1 and z_2; FF is the flux factor for the PDF's; TF($S * \vec{y} \mid \vec{x}$) are the transfer functions that predict the measured jet momenta distributions from the parton-level quarks; M($m_t, \vec{x}$) is the matrix element for $t\bar{t}$ production and decay; and $d\Phi(\vec{x})$ indicates integration over phase space for the 6 partons in the $t\bar{t}$ decay. We describe in Section 5.5 our construction of "effective" propagators which modify the usual Breit-Wigner functions to take into account our simplifying assumptions. Detailed discussion of the normalization factor $\frac{1}{N(m_t)}$ can be found in Section 5.1.

The PDFs $f(z_1)$ and $f(z_2)$ are integrated over the appropriate combinations of q, $\bar{q}$ and gluons. The flux factor FF is given by

$$FF = \sqrt{(z_1 \cdot z_2)^2 - m_{z1}^2 m_{z2}^2} \quad \text{with} \quad FF = |z_1||z_2| \quad \text{for} \quad m_{z1} = m_{z2} = 0 \tag{2}$$

The acceptance $A(m_t, S)$ is obtained from Monte Carlo for different values of JES. Details are found in Section 5.2. The transfer functions are also obtained from MC, as discussed in Section 4.

We evaluate the likelihood for each event, both $t\bar{t}$ candidate and background events, and than use the following expression to evaluate the top mass.

$$\log L(m, S) = \sum_{i=1}^{N} \log((1 - f_i)L_{t\bar{t}}(m, S) + f_i U(m, S)) \tag{3}$$

where N is the total number of events; f_i is the probability for event i to be a background event; $L_{t\bar{t}}(m, S)$ is the normalized likelihood curve as a function of the top mass and S for event i assumed to be $t\bar{t}$ event; and $U(m, E)$ is the background likelihood as a function of the top pole mass. In this analysis, $U(m, S)$ is a uniform distribution integrated to one over our top mass range. The probability f_i for an event to be a background event is estimated by a single discriminant, which combines a number of kinematic variables, as discussed in Section 6. Details of the final fit will be discussed in section 7.

4 Transfer Functions

We compute the ratio of magnitude of the jet p, corrected to L5, to generated parton E as a function of parton p_T, where the parton is taken to be the "effective" on-shell parton of the sort used to construct the effective propagators described in Section 5.5. This procedure is the opposite of what the DLM method is using. Our approach yields transfer functions independent of the top mass; thus we can use all of the top mass Monte Carlo samples to obtain the transfer functions.

The transfer functions are obtained from tight reconstructed jets in the Monte Carlo (HERWIG) samples and depend on the matching requirements between HEPG partons and reconstructed jets. For our matching, we consider all possible jet-to-parton assignments, and for each assignment we calculate a matching χ^2 defined by

$$\chi^2 = \sum_{i=1}^{4} \left(\frac{\eta_i^j - \eta_i^p}{\sigma_\eta}\right)^2 + \left(\frac{\phi_i^j - \phi_i^p}{\sigma_\phi}\right)^2 \tag{4}$$

where σ_η and σ_ϕ, the η and ϕ resolutions, are both taken to be 0.05. The parton-jet matches taken from the event correspond to that permutation with the lowest χ^2.

We need to parameterize the transfer functions in order to use them effectively in the integration. It is desirable to use some functions described by quantities calculated from the transfer function distributions themselves, such as: mean (μ), standard deviation (σ), skewness (s), and kurtosis (k). For this reason we use the density functions proposed in 1949 by statistician Norman Johnson [17], described in Appendix A. Johnson's S_U and S_B distributions attain all possible values of skewness and kurtosis. By comparison the

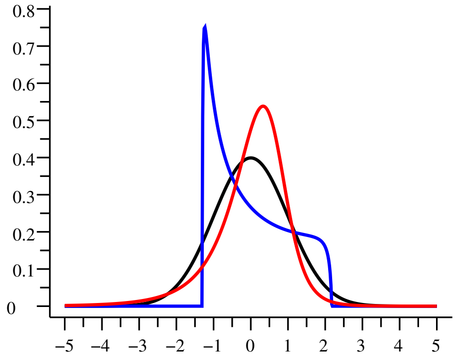

Figure 1: Black: $s = 0,\ k = 3$, Gaussian. Red: $s = -1.5,\ k = 10$, S_U Blue: $s = 0.5,\ k = 2$, S_B. For all three curves, the mean is 0 and the standard deviation is 1.

Pearson's system needs twelve different function types to do this. Examples of Johnson's density functions are shown in Fig 1.

A computer algorithm for calculating S_U and S_B parameters for given mean, standard deviation, skewness, and kurtosis can be found in [18]. Using such an algorithm, the complete Johnson's system can be treated as a single function with parameters μ, σ, s, and k. We have developed an improved version of this algorithm. The three parameters – mean, skewness, and kurtosis – are represented as quadratic functions of parton p_T. The standard deviation is represented by $\sigma = A + B \exp{(Cp_T + Dp_T^2)}$. The fit results of the transfer functions for the light flavor jets in the central calorimeter are shown in Figure 2. We fit separate transfer functions for four different $|\eta|$ regions: the dip region ($|\eta| < 0.15$), the central region ($0.15 \le |\eta| < 0.85$), the crack region ($0.85 \le |\eta| < 1.4$), and the plug region ($1.4 \le |\eta| < 2.0$). The fits are performed on a sample of Herwig Monte Carlo events at many top quark masses. The selection cuts used on the events are the same as those for our data, the only additional requirement being that it must actually be a lepton+jets event.

4.1 Transfer Function Crosscheck

To validate the transfer functions, we perform a crosscheck by reconstructing the hadronic W and top mass using only the transfer functions. This crosscheck uses a much simpler machinery than the full top mass integration, and it should be emphasized that the crosscheck is not intended to produce an independent W and top mass measurement – its purpose is to validate the transfer functions by showing that they can correctly reconstruct these masses.

Please note that this section uses the following conventions:

- $\vec{p_1}$ denotes a 3-vector.
- $p_1 = |\vec{p_1}|$ denotes the magnitude of a 3-vector.

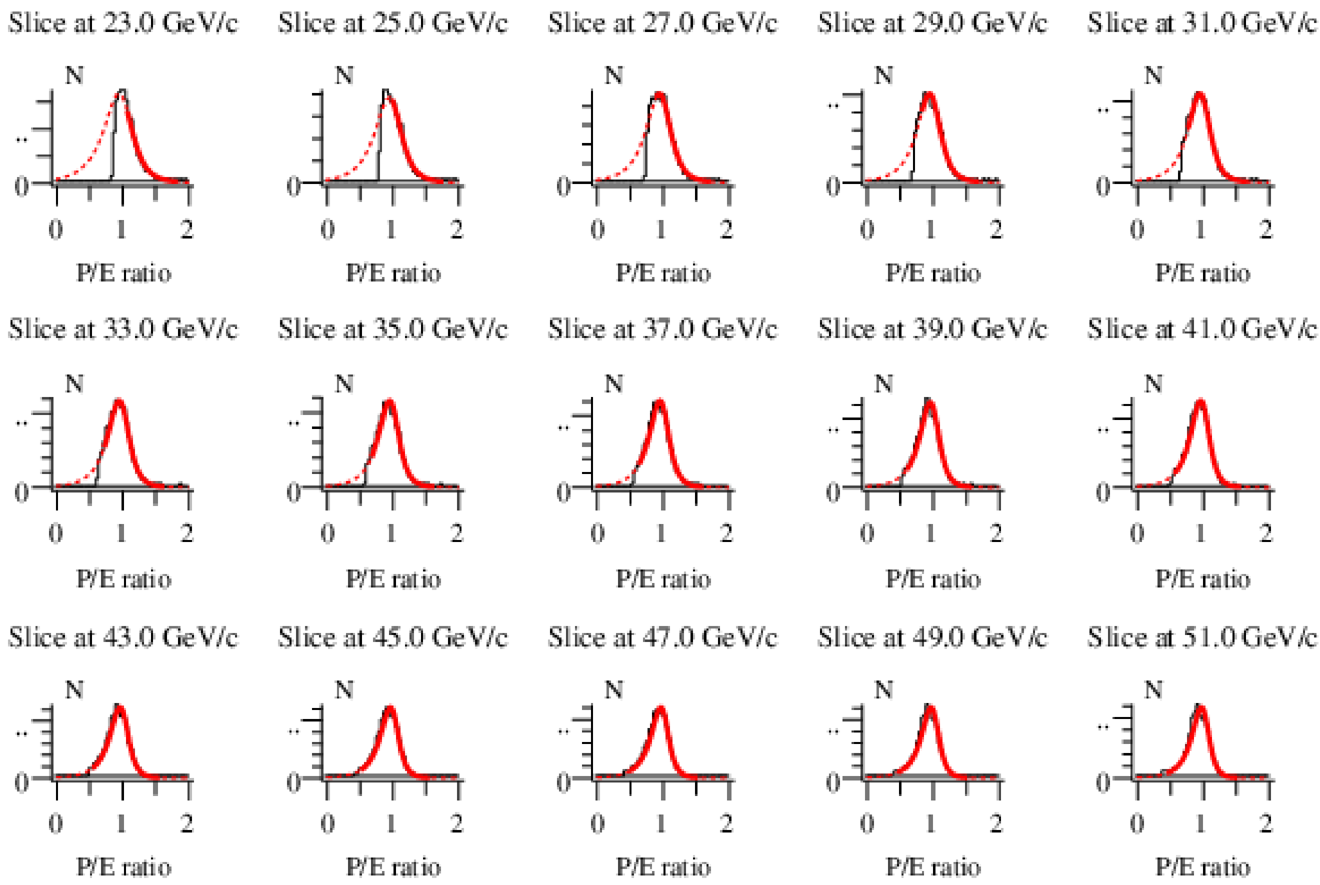

Figure 2: Fitted transfer functions for light quarks in the central calorimeter in 10 GeV/c p_T bins from 22 to 52 GeV/c. The dashed curves represent extrapolations of the fit functions to momenta which are below the cutoff value.

- P_1 denotes a 4-vector.
- $P^2 = P^\nu P_\nu$ denotes the invariant magnitude of a 4-vector.

Note incidentally that numeric constants of 2, π, etc. will be dropped throughout, as the overall normalization of the likelihood does not matter.

W mass reconstruction

We begin with the observation that for a given W mass m_W, the probability for observing jets with momenta $\vec{j_1}$ and $\vec{j_2}$ is given by

$$P(\vec{j_1}, \vec{j_2}) = \int \frac{d^3\vec{p_1}}{2E_1} \frac{d^3\vec{p_2}}{2E_2} \delta^4(P_1 + P_2 - P_W) w(\vec{j_1}|\vec{p_1}) w(\vec{j_2}|\vec{p_2}) f(\vec{p_W}) d^3\vec{p_W}$$

where w is the transfer function and f is the prior distribution of the W momentum.

Now, we assume that the transfer functions depend only on the magnitude of the momentum; i.e. $w(\vec{j}|\vec{p}) = w(j|p)\delta^2(\Omega_j - \Omega_p)$. Integrating over the angular variables and dropping the factors of 2 yields

$$P(j_1,j_2) \sim \int \frac{p_1^2 dp_1}{E_1}\frac{p_2^2 dp_2}{E_2}\delta^4(P_1+P_2-P_W)w(j_1|p_1)w(j_2|p_2)f(\vec{p_W})d^3\vec{p_W}$$

Now, we consider the prior $f(\vec{p_W})$. We do not expect the prior to have any ϕ-dependence, so, given that the prior is a density function, we write

$$f(\vec{p_W})d^3\vec{p_W} = f(\vec{p_W})p_W^2 dp_W d\cos\theta_W d\phi$$

so

$$f(\vec{p_W}) \sim f(p_W,\cos\theta_W)/p_W^2$$

Next, we assume a narrow W width, so we can write $\frac{d^3\vec{p_W}}{2E_W} = d^4P_W\delta(P_W^2 - m_W^2)$. Making these substitutions, we get

$$\begin{aligned}P(j_1,j_2) \sim \int\ & \frac{p_1^2 dp_1}{E_1}\frac{p_2^2 dp_2}{E_2}\delta^4(P_1+P_2-P_W)w(j_1|p_1)w(j_2|p_2)\frac{f(p_W,\cos\theta_W)}{p_W^2}\\ & 2E_W d^4P_W\delta(P_W^2-m_W^2)\end{aligned}$$

Now we can perform the P_W integration, getting rid of the δ-function enforcing the conservation of 4-momentum (so we now have $P_W = P_1 + P_2$) to yield

$$P(j_1,j_2) \sim \int \frac{p_1^2 dp_1}{E_1}\frac{p_2^2 dp_2}{E_2}\frac{E_W}{p_W^2}w(j_1|p_1)w(j_2|p_2)f(p_W,\cos\theta_W)\delta(P_W^2-m_W^2)$$

At this point we have two variables of integration and one δ-function left, so we change variables to P_W^2 and β, where β is defined as $\log\frac{p_1}{p_2}$.

This requires computing the Jacobian. For this, we note that

$$P_W^2 = (P_1+P_2)^2 = P_1^2 + P_2^2 + 2P_1\cdot P_2 = m_1^2 + m_2^2 + 2(E_1E_2 - p_1p_2\cos\theta)$$

where θ is the angle between the two light quarks. Expanding this out yields

$$P_W^2 = m_1^2 + m_2^2 + 2(\sqrt{p_1^2+m_1^2}\sqrt{p_2^2+m_2^2} - p_1p_2\cos\theta)$$

Now, we differentiate:

$$\begin{aligned}\frac{\partial P_W^2}{\partial p_1} &= 2(\frac{p_1\sqrt{p_2^2+m_2^2}}{\sqrt{p_1^2+m_1^2}} - p_2\cos\theta) = 2(\frac{p_1E_2}{E_1} - p_2\cos\theta)\\ \frac{\partial P_W^2}{\partial p_2} &= 2(\frac{p_2\sqrt{p_1^2+m_1^2}}{\sqrt{p_2^2+m_2^2}} - p_1\cos\theta) = 2(\frac{p_2E_1}{E_2} - p_1\cos\theta)\\ \frac{\partial\beta}{\partial p_1} &= \frac{1}{p_1}\\ \frac{\partial\beta}{\partial p_2} &= -\frac{1}{p_2}\end{aligned}$$

So, we integrate over P_W^2 to eliminate the last δ-function, and are left with

$$P(j_1, j_2) \sim \int \frac{p_1^2}{E_1}\frac{p_2^2}{E_2}\frac{E_W}{p_W^2} w(j_1|p_1) w(j_2|p_2) f(p_W, \cos\theta_W) \frac{d\beta}{J}$$

where J is the Jacobian computed from the above matrix of partial derivatives.

Note that this method assumes fixed parton masses, so these must be specified somehow in order for the integration to proceed.

t mass integration

The top mass integration proceeds in a way very similar to the integration method for the W mass outlined above. We start with the same integral as for the W mass, but with added terms for the b quark and the t parent:

$$P \sim \int \frac{d^3\vec{p_1}}{2E_1}\frac{d^3\vec{p_2}}{2E_2}\frac{d^3\vec{p_b}}{2E_b}\delta^4(P_1 + P_2 - P_W) w_1 w_2 w_3 f(\vec{p_W}) d^3\vec{p_W} f_t(\vec{p_t}) \delta^4(P_W + P_b - P_t) d^3\vec{p_t}$$

where $w_1 \equiv w(\vec{j_1}|\vec{p_1})$, and we have introduced another prior f_t for the top momentum.

As before we eliminate the angular integrals, substitute $d^3\vec{p_W} = 2E_W d^4P_W \delta(P_W^2 - m_W^2)$ and similarly $d^3\vec{p_t} = 2E_t d^4P_t \delta(P_t^2 - m_t^2)$, and integrate over d^4P_W and d^4P_t to eliminate those δ-functions to obtain

$$P \sim \int \frac{p_1^2 dp_1}{E_1}\frac{p_2^2 dp_2}{E_2}\frac{p_b^2 dp_b}{E_b} w_1 w_2 w_3 2E_W 2E_t f(\vec{p_W}) f_t(\vec{p_t}) \delta(P_W^2 - m_W^2) \delta(P_t^2 - m_t^2)$$

Now, we want to change variables again. This time we change from p_1, p_2, and p_3 to P_W^2, P_t^2, and β, thus introducing a 3x3 Jacobian. The treatment of P_W^2 and β is the same as before; as for P_t^2, we use $P_t^2 = m_W^2 + m_b^2 + 2(\sqrt{p_W^2 + m_W^2}\sqrt{p_b^2 + m_b^2} - 2p_W p_b \cos\theta_{Wb})$. As it turns out, we only need $\frac{\partial P_t^2}{\partial p_b}$, since $\frac{\partial P_W^2}{\partial p_b}$ and $\frac{\partial \beta}{\partial p_b}$ are both zero, so the other two derivatives of P_t^2 do not appear in the final Jacobian.

Computing out the derivatives yields:

$$
\begin{aligned}
\frac{\partial P_W^2}{\partial p_1} &= 2(\frac{p_1\sqrt{p_2^2+m_2^2}}{\sqrt{p_1^2+m_1^2}} - p_2\cos\theta) = 2(\frac{p_1E_2}{E_1} - p_2\cos\theta) \\
\frac{\partial P_W^2}{\partial p_2} &= 2(\frac{p_2\sqrt{p_1^2+m_1^2}}{\sqrt{p_2^2+m_2^2}} - p_1\cos\theta) = 2(\frac{p_2E_1}{E_2} - p_1\cos\theta) \\
\frac{\partial P_W^2}{\partial p_b} &= 0 \\
\frac{\partial P_t^2}{\partial p_b} &= 2(\frac{p_bE_W}{E_b} - p_W\cos\theta_{Wb}) \\
\frac{\partial \beta}{\partial p_1} &= \frac{1}{p_1} \\
\frac{\partial \beta}{\partial p_2} &= -\frac{1}{p_2} \\
\frac{\partial \beta}{\partial p_b} &= 0
\end{aligned}
$$

Thus, our final likelihood becomes

$$
P \sim \int \frac{p_1^2}{E_1}\frac{p_2^2}{E_2}\frac{p_b^2}{E_b}w_1w_2w_3\frac{E_W}{p_W^2}\frac{E_t}{p_t^2}f(p_W,\cos\theta_W)f_t(p_t,\cos\theta_t)\frac{d\beta}{J}
$$

again, where J is the Jacobian above.

There is still one problem left: this yields a two-dimensional likelihood curve $L(m_W, m_t)$. Scanning over both m_W and m_t would take a very long time, and it would leave us with the question of how exactly to interpret the resulting likelihood curve. Consequently, we merely take one slice of this resulting likelihood curve at the peak value of m_W obtained from the W mass integration. Our studies have shown that this produces comparable results to performing the full two-dimensional integration and fitting the results.

Integration Details

We select events using a similar procedure to that used to build the transfer functions described above: we use Herwig $t\bar{t}$ events with 4 or more tight jets, with the jets and partons matched using the χ^2-based algorithm described previously.

For the W mass integration, we scan over the W mass from 20 to 160 GeV at 0.5 GeV intervals. At each W mass point, we integrate over the β parameter from -1.95 to 1.95 at intervals of 0.1. The β parameter integration is centered at the value of β for the L5-corrected jets, although studies have shown that this yields no discernible difference between merely centering the range of β at 0.

After the scan is completed, the resulting likelihood curve is then convoluted with a Breit-Wigner with a width of the W width (2.1 GeV). (This is intended to re-introduce

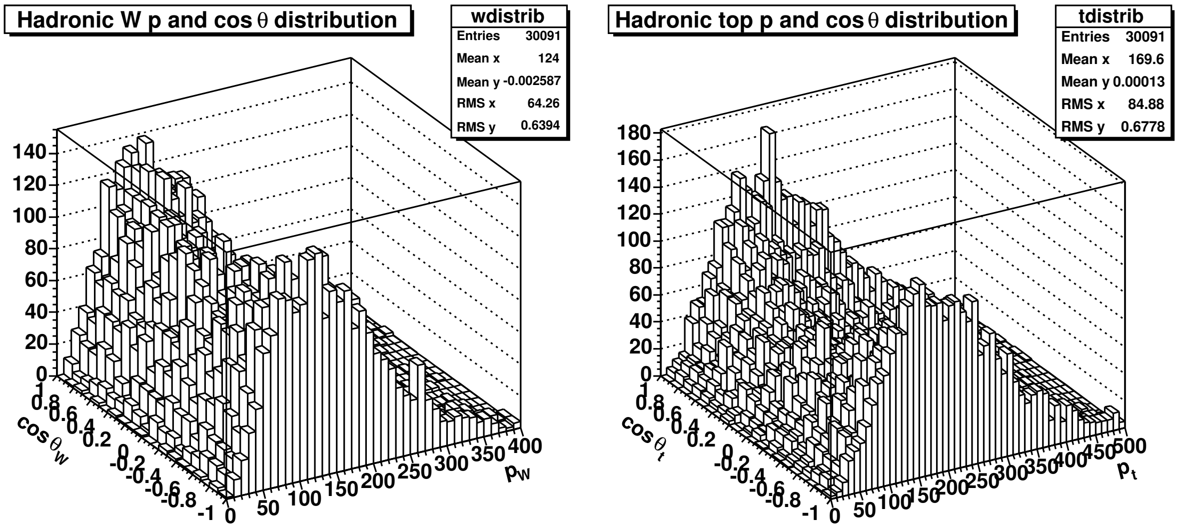

Figure 3: 2-dimensional prior distributions built from HEPG.

the finite W width, since the integration method, as described above, assumes a zero W width.) The function is then interpolated using a local polynomial fit to steps of 0.1 GeV.

The top mass integration scans from 80 to 280 at 0.5 GeV intervals, and uses the same β integration as in the W mass.

The resulting likelihood curves are then multiplied together, and we use the peak of the resulting likelihood curve as our final W and top mass results. We can also perform pseudoexperiments on the individual likelihood curves, which yields very similar answers.

Priors

To compute the prior distributions, we simply build a 2-D histogram of the W momentum and $\cos\theta$ and top momentum and $\cos\theta$ from the HEPG banks of selected events. This results in the priors shown in Fig. 3.

However, using these priors "as-is" results in limited resolution due to binning effects in the histograms. In order to obtain smoother priors, first we observe that from these histograms, the two variables appear to be sufficiently uncorrelated that we can separate them into two separate 1-D histograms. We can then fit the 1-D histograms with appropriate functions. For the momentum distribution, we choose a function of the form $y = Cx^3e^{-Ax}$. For the W $\cos\theta$ distribution, we use $y = C(x^2 + A)$, while the t $\cos\theta$ distribution uses a quartic, $y = C(x^4 + A)$. In all cases the C is an overall normalization which can be thrown away, so we only end up with one relevant constant, A. The resulting fits are shown in Fig. 4.

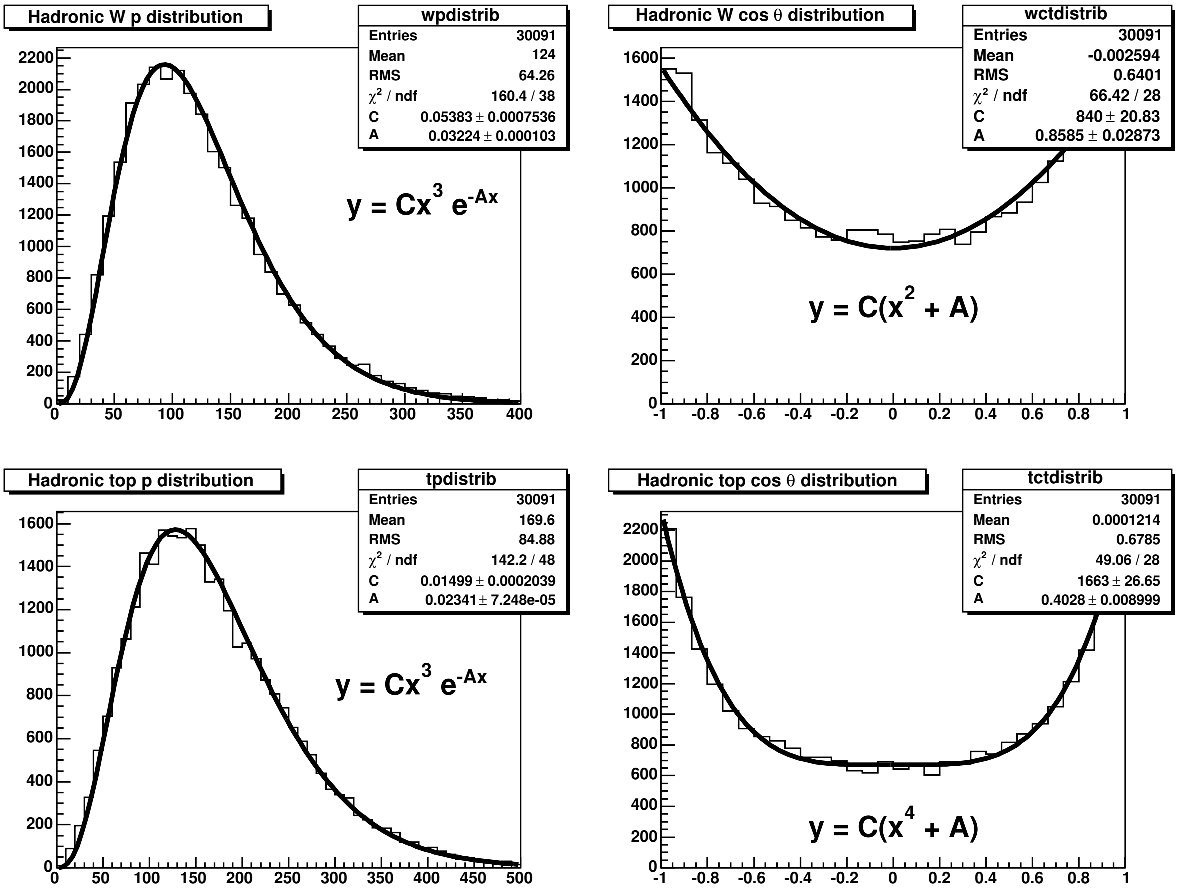

Figure 4: 1-dimensional prior distributions with fitted curves.

Result of the crosscheck

These results are produced using a Herwig $t\bar{t}$ sample (`ttophl`) with a nominal W mass of 80.4 GeV and a top mass of 178 GeV.

We first test the crosscheck by performing our integration using HEPG partons with a 20% Gaussian smearing applied as our "jets". For this test, the transfer functions used are also just Gaussian transfer functions. We also take the parton masses from HEPG in order to solve our kinematic equations. These results are displayed in Fig. 5, showing a W mass of 80.1 GeV and a top mass of 179.2 GeV.

For our final check, we use our fully-simulated jets and our actual transfer functions. In this case, the parton masses are set to 4.95 GeV for the hadronic b parton and 0 GeV for the W daughters when solving the kinematic equations. These results are displayed in Fig. 6. We obtain a W mass of 79.0 GeV and a top mass of 178.2 GeV.

The results for both the HEPG partons and the simulated jets are satisfactory, indicating that our transfer functions perform as desired.

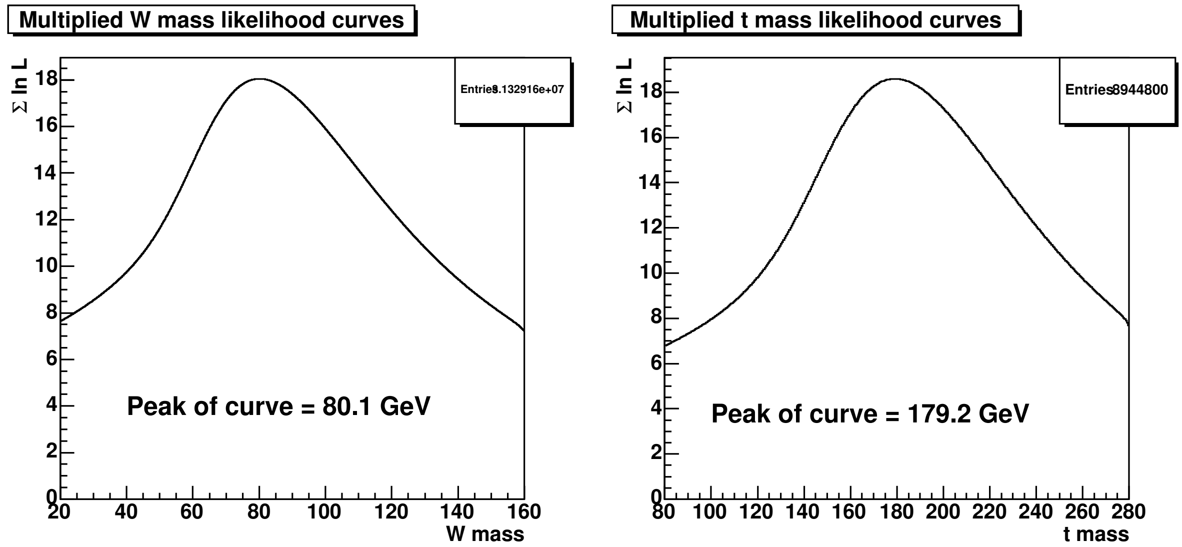

Figure 5: TF crosscheck results for hadronic W and top mass using smeared HEPG partons.

5 Signal Event Likelihood

The $t\bar{t}$ system to be solved in the lepton+jets topology has 6 particles in the final state: 2 leptons from the W, 2 light quarks from the hadronic W and two b quarks, one from the leptonic and one from the hadronic top. For leading order production of the $t\bar{t}$ system, we expect only 4 jets in the event. The number of variables is then $n_{\mathrm{var}} = 24 - 2$ if we assume that the electron and the neutrino masses are known. We reduce that number further if we make the following assumptions:

1. The lepton energy and angles are perfectly measured.
2. Quark angles are the same as the measured jet angles.
3. Hadronic side quarks are on mass shell, leptonic side b mass is zero.

We integrate over the variables that we do not know perfectly. Eq. 1 of Section 3, shown again below, uses a set of measured quantities $\vec{y}$.

$$L(\vec{y} \mid m_t, \vec{x}, S) = \frac{1}{N(m_t)} \int \frac{1}{A(m_t, S)} \frac{f(z_1)f(z_2)}{FF}\, TF(\vec{y}(S) \mid \vec{x})\, |M(m_t, \vec{x})|^2\, d\Phi(\vec{x}) \qquad (5)$$

The measured quantities, $\vec{y}$, are the 4 jet momenta and angles. It is convenient to use a different set of variables in our integration as discussed in Section 5.4.

This section describes the different components of the above integral. The normalization factor, the acceptances, the matrix element for $t\bar{t}$ production and decay. The phase space integration is discussed in Section 5.4, with details reserved for Appendix C.

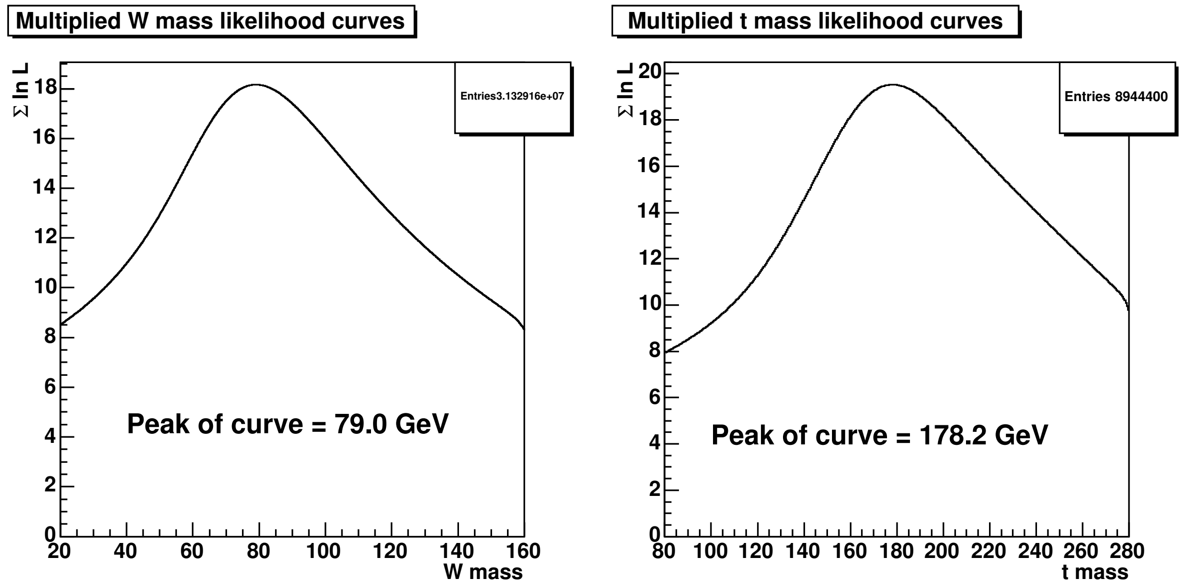

Figure 6: TF crosscheck results for hadronic W and top mass using simulated jets.

5.1 Likelihood Normalization

The normalization factor for the likelihood calculation is not simply the $t\bar{t}$ cross section for the following reasons: we use events that were generated with a fixed top width independent of m_t, we use effective propagators instead of simple Breit-Wigner functions, and we include mass and JES dependent acceptances which modify the integral. Any "global normalization" needs to account for all of these effects.

It is interesting to note that our calculated global normalization factor considering just the best permutation is also valid when considering all permutations. We can tell that not just the best/correct permutation dominates; when we permute the W daughters, the calculated normalization remains the same. This implies that the two permutations both contribute to the overall normalization.

The computed normalization is proportional to $\frac{\sigma_{t\bar{t}}\Gamma_t^2}{m_t^2}$, where m_t is the top mass, $\sigma_{t\bar{t}}$ is the tree-level $t\bar{t}$ cross-section, and Γ_t is the top width (which is itself proportional to m_t^3).

It's also interesting to note (Fig. 7) that the normalization peaks at a top scan mass near $175\,\mathrm{GeV}/c^2$. The factor is of course calculated for the same mass scan range as the final result.

5.2 Acceptance

In order to correct for the effect of acceptance (also referred to as "efficiency" in this note) on the top likelihood, we must compute the acceptance A as a function of m_t and JES. In the past, we computed the acceptance using Monte Carlo samples with different mass, and shifting the JES accordingly. However, we now use a slightly different method.

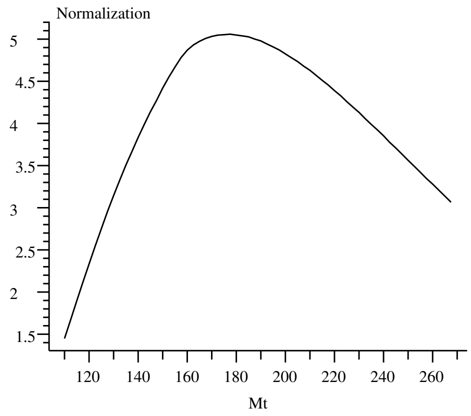

Figure 7: Global normalization factor calculated as a function of top scan mass.

Rather than use the fully-simulated Monte Carlo and the cuts on the observed quantities, we use quantities from the effective partons used in building our effective propagators. In doing this, we get an acceptance which is more in agreement with the model used in our integration procedure.

We have a large number of HEPG-only (unsimulated) Monte Carlo events generated at a variety of different masses. For each event, we take the "effective partons", built as described in Section 5.5, and then apply smearing according to our transfer functions. This should give us particles which look like Level 5 jets. The particle energies are then rescaled according to JES, and we apply the following cuts to determine if the event is accepted:

- The 4 partons must have an energy > 21 GeV (a cut of 21 GeV at level 5 being equivalent to a cut of 15 GeV at level 4) and an $|\eta| < 2$.
- The lepton must have a $P_T > 20$ GeV and an $|\eta| < 1$.
- The partons must be separated by a minimum ΔR of 0.5 (to mimic the effect of jet clustering).
- The lepton must be separated from each of the partons by a minimum ΔR of 0.5 (to account for the lepton isolation cut).
- The missing E_T, defined as the sum of the momenta of the 4 partons and the $t\bar{t}$ system, must be at least 20 GeV.

These cuts are designed to mimic the effect of the simulation cuts on HEPG-only Monte Carlo. We then obtain the acceptance by dividing the number of events accepted for a given m_t and JES by the total number of events in that m_t sample.

Figure 8 shows the 2-D efficiency as a function of m_t and JES. Note that the JES axis runs from -5 to 5 because it is measured in number of sigmas; this is to allow for the possibility of varying the sigma on a jet-by-jet basis in the future. Currently σ is fixed at 0.03, so effectively this represents a JES scale running from 0.85 to 1.15.

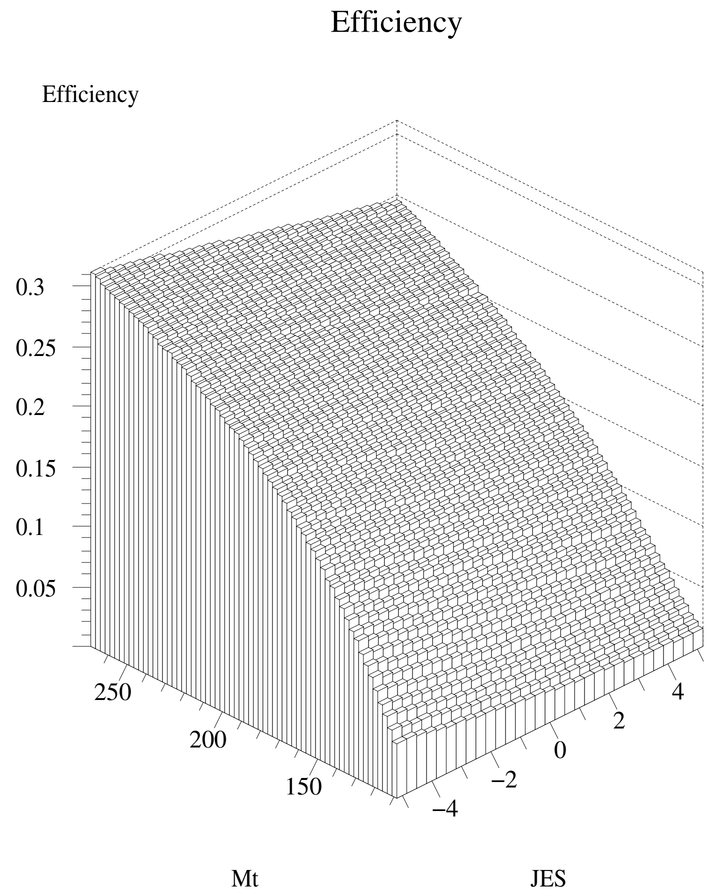

Figure 8: Efficiencies used in our integration. JES is expressed in terms fraction of σ=0.03, thus -5 to +5 corresponds to JES=0.85-1.15.

5.3 Matrix Elements and Parton Density Functions

In our previous analysis [1] we used the matrix elements by Mahlon and Parke [19, 20], which is a very compact analytical expression for the square of the amplitudes for the process $q\bar{q}\to t\bar{t}\to b q_1 \bar{q}_2 \bar{b} l \nu$ (valid also for the other $t\bar{t}$ decay modes). The calculation takes into account the spin correlations between the top productions and the decay products. However, the calculations were performed in the narrow width approximation for the W and assume no masses for the light quarks q_1, q_2 and the b quark.

For this update we consider the complete leading order matrix elements by Kleiss and Stirling [21] that includes both the $q\bar{q} \to t\bar{t}$ and $gg \to t\bar{t}$ process with complete spin correlations. It is expected that the gluon-gluon fusion fraction amounts about 15% of the total $t\bar{t}$ cross section at current Tevatron energies. The calculation also considers the finite widths of the W and effects of a non-zero b quark mass, thus we are more consistent with our phase space integrator which performs an integration over the W mass Breit-Wigner W mass and assumes massive b quark mass on the hadronic side of the $t\bar{t}$ decay.

The calculation is supplied by the authors as FORTRAN code[1] which we adapted for the needs of our phase space integrator. Due to the presence of the $t\bar{t}$ recoil there is an ambiguity in the choice of the initial $q\bar{q}$ or gg configuration. In our approach we evaluate the initial four-vectors from the four-vectors of the decay products given by the phase space integrator, assuming that the flight directions of the initial partons are parallel to the beam axis in the $t\bar{t}$ rest frame.

Since we do not experimentally distinguish between the two quarks from the hadronically decaying W, we take the average of the two possible configurations. Finally we weigh the matrix element squared for the qq part and gg part with the leading order CTEQ5L parton distribution probabilities [22] for the light quarks and for the gluon, respectively. The sum of both parts enters as a weight of the event within the integrator, for a given top quark mass, permutation and neutrino solution.

We have performed various checks, such as the reproduction of the Mahlon and Parke matrix elements using the $q\bar{q}$ part of the Kleiss-Stirling code, or by checking the rotational invariance of the matrix elements, and found that code is sane. The phenomenology of gluon-gluon fusion is discussed e.g. in [23].

With a small hit on the information content of the events, the analysis can be performed using just the top quark and W propagators. Unfortunately, a model independent analysis would be difficult since non-SM top mass MC samples are not available as yet.

Kleiss-Stirling formulae

The matrix elements calculations [21] consider the helicity amplitides for the two processes

$$\begin{aligned} \bar{q}(p_1) + q(p_2) &\rightarrow t + \bar{t}\,, & (6)\\ \bar{g}(p_1) + g(p_2) &\rightarrow t + \bar{t}\,, & (7) \end{aligned}$$

where both processes are followed by the top decays

$$\begin{aligned} \bar{t} &\rightarrow \bar{b}(p_3) + e^-(p_4) + \bar{\nu}_e(p_5)\,, & (8)\\ t &\rightarrow \bar{b}(p_6) + \nu_e(p_7) + e^+(p_8)\,. & (9) \end{aligned}$$

p_i, $i = 1 \ldots 8$ are the four-momenta of the initial and final state particles involved. Eq. (8) and (9) represent the semileptonic decay mode, but the following formulae are also valid for the other decay modes if the p_i are assigned to the corresponding isospin partners of the alternative decay products. In the calculations, production and decay are treated as one process, thus complete decay correlations are automatically taken into account.[2]

[1] Thanks to Stephen Parke.

[2] The authors do not suppose that the polarization of the massive b quarks is measured. Thus, their spins are summed over implicitly at amplitude level.

The helicity amplitudes are given in terms of spinor products

$$
\begin{aligned}
s_+(k_1,k_2) &= \bar{u}_+(k_1)u_-(k_2)\,, \\
s_-(k_1,k_2) &= \bar{u}_-(k_1)u_+(k_2)\,,
\end{aligned} \tag{10}
$$

where $u_+(k_1)$ and $u_-(k_2)$ denote the two helicity states with four-momentum k_i. All results follow directly by straightforward summation of the Feynman diagrams, squaring, spin summation and averaging.

The amplitudes $M_q(\nu)$ for the process (6) given the helicity ν for the incoming quark (neglecting the color structure for now) can be written as

$$
\begin{aligned}
M_q(\nu) &= i g_s^2 g_w^4 T_q(\nu)/D_q\,, \\
D_q &= P_W(p_4+p_5)P_W(p_7+p_8)P_t(q)P_t(q')(p_1+p_2)^2\,,
\end{aligned} \tag{11}
$$

where g_s and g_w are the weak and strong coupling constants, respectively, and $T_q(\nu)$ contains the spin information. q and q' denote the momenta of the top quark and the anti-top quark, respectively. P_W and P_t are propagator functions originally defined as

$$
\begin{aligned}
P_W(k) &= k\cdot k - m_W^2 + i m_W \Gamma_W\,, \\
P_t(k) &= k\cdot k - m_t^2 + i m_t \Gamma_t\,.
\end{aligned} \tag{12}
$$

For the spinor part, two auxiliary four-vectors p_9 and p_{10} need to be introduced. They are related to the momentum of the top quark q and that of the anti-top quark q' through[3]

$$
\begin{aligned}
p_9 &= q' - \left(\frac{q'\cdot q'}{2q'\cdot p_4}\right) p_4\,, \\
p_{10} &= q - \left(\frac{q\cdot q}{2q\cdot p_8}\right) p_8\,.
\end{aligned} \tag{13}
$$

Using useful representations of Eq. (10) and the substitution (13) reduces the numerator part T_q to spinor products:

$$
\begin{aligned}
T_q(+) &= 256\,[(p_3\cdot p_5)(p_6\cdot p_7)]^{1/2} \\
&\quad \cdot\{-s_+(p_8,p_{10})s_-(p_{10},p_2)s_+(p_1,p_9)s_-(p_9,p_4) + m_t^2 s_+(p_8,p_1)s_-(p_2,p_4)\}, \\
T_q(-) &= [T_q(+)](p_1 \leftrightarrow p_2)
\end{aligned} \tag{14}
$$

The two helicity amplitudes are squared and summed, and an overall factor is applied which takes into account the color sum and the averaging over the initial spins and colors. The final result for this process is

$$
\mathcal{R}_{qq} = \frac{1}{18}\{|M_q(+)|^2 + |M_q(-)|^2\} \tag{15}
$$

[3]The replacement $q \to p_{10}$, $q' \to p_9$ is necessary since the auxiliary vectors are lightlike and only lightlike vectors may appear as arguments of the spinor products.

The process (7) includes three diagrams instead of one. The corresponding amplitudes M_g^1, M_g^2 and M_g^3 for given helicities λ_1 and λ_2 and momenta p_1 and p_2 of the gluons can be written as

$$\begin{aligned} M_g^l(\lambda_1,\lambda_2) &= -ig_s^2 g_w^4 T_g^l(\lambda_1,\lambda_2)/(4D_g^l p_1\cdot p_2))\,,\ l=1,2,3\,, \\ D_g^1 &= P_W(p_4+p_5)P_W(p_7+p_8)P_t(q)P_t(q')P_t(q-p_1)\,, \\ D_g^2 &= P_W(p_4+p_5)P_W(p_7+p_8)P_t(q)P_t(q')P_t(q-p_2)\,, \\ D_g^3 &= P_W(p_4+p_5)P_W(p_7+p_8)P_t(q)P_t(q')(p_1+p_2)^2\,. \end{aligned} \tag{16}$$

Again we modify the original propagator terms $P_W(p_4+p_5)$, $P_W(p_7+p_8)$, $P_t(p_{10})$ and $P_t(p_9)$. In Eq.(16) a suitable gauge for the polarization vector[4] is already used which simplifies both the denominator and the spinorial parts.

In addition to the substitution (13), a third auxiliary vector

$$p_{11} = q - \left(\frac{q\cdot q}{2q\cdot p_1}\right)p_1 \tag{17}$$

is needed to evaluate the T's just as before. The resulting expressions read

[4]Since the top quarks are only intermediate states and not produced as stable, on-shell particles with zero width, the sum of the above Feynman diagrams are not gauge-invariant. Instead of introducing extra graphs to overcome this problem, the authors claim that by using the direction of the one top quark as polarization vector of the other top quark, the error made by restricting on the amplitudes Eq.(16) is negligible and less than $\mathcal{O}(\Gamma_t/m_t)$.

$$
\begin{aligned}
T_g^1(+,+) &= 512\left[(p_3\cdot p_5)(p_6\cdot p_7)\right]^{1/2}\{-2(q\cdot p_1) \\
&\quad \cdot s_+(p_8,p_{10})s_-(p_{10},p_2)s_+(p_2,p_9)s_-(p_9,p_4) \\
&\quad +m_t^2 s_+(p_8,p_{10})s_-(p_{10},p_2)s_+(p_1,p_2) \\
&\quad s_-(p_1,p_4) - m_t^2 s_+(p_8,p_1)s_-(p_2,p_1) \\
&\quad s_+(p_2,p_9)s_-(p_9,p_4) + 2m_t^2((q-p_1)\cdot p_2) \\
&\quad s_+(p_8,p_1)s_-(p_1,p_4)\}\,, \qquad (18)\\
T_g^1(+,-) &= 512\left[(p_3\cdot p_5)(p_6\cdot p_7)\right]^{1/2} \qquad (19)\\
&\quad \cdot\{-s_+(p_8,p_{10})s_-(p_{10},p_2)s_+(p_1,p_{11}) \\
&\quad s_-(p_{11},p_2)s_+(p_1,p_9)s_-(p_9,p_4) \\
&\quad +m_t^2 s_+(p_8,p_1)s_-(p_2,p_{11})s_+(p_{11},p_1)s_-(p_2,p_4)\}\,, \\
T_g^1(-,+) &= [T_g^1(+,-)](p_1\leftrightarrow p_2)\,, \qquad (20)\\
T_g^1(-,-) &= 512\left[(p_3\cdot p_5)(p_6\cdot p_7)\right]^{1/2} \qquad (21)\\
&\quad \cdot\{-2((q-p_1)\cdot p_2)s_+(p_8,p_{10})s_-(p_{10},p_1) \\
&\quad \cdot s_+(p_2,p_9)s_-(p_9,p_4) + m_t^2 s_+(p_8,p_{10}) \\
&\quad \cdot s_-(p_{10},p_1)s_+(p_2,p_1)s_-(p_2,p_4) \\
&\quad -m_t^2 s_+(p_8,p_2)s_-(p_1,p_2)s_+(p_1,p_9)s_-(p_9,p_4) \\
&\quad 2(q\cdot p_1)s_+(p_8,p_2)s_-(p_2,p_4)\}\,, \\
T_g^2(\lambda_1,\lambda_2) &= [T_g^1(\lambda_2,\lambda_1)](p_1\leftrightarrow p_2) \qquad (22)\\
T_g^3(+,+) &= 512\left[(p_3\cdot p_5)(p_6\cdot p_7)\right]^{1/2}(p_1\cdot p_2) \qquad (23)\\
&\quad \cdot\{s_+(p_8,p_{10})s_-(p_{10},p_1)s_+(p_1,p_9)s_-(p_9,p_4) \\
&\quad -s_+(p_8,p_{10})s_-(p_{10},p_2)s_+(p_2,p_9)s_-(p_9,p_4) \\
&\quad -m_t^2(s_+(p_8,p_1)s_-(p_1,p_4) - s_+(p_8,p_2)s_-(p_2,p_4))\}\,, \\
T_g^3(+,-) &= T_g^3(-,+) = 0\,, \qquad (24)\\
T_g^3(-,-) &= T_g^3(+,+)\,. \qquad (25)
\end{aligned}
$$

Adding the square of the amplitudes plus taking the color structure and initial spin configurations into account, the final result for this process is

$$
\begin{aligned}
\mathcal{R}_{gg} &= \frac{1}{256}\sum_{\lambda_1,\lambda_2}\{\frac{7}{3}|M_g^1(\lambda_1,\lambda_2)+M_g^2(\lambda_1,\lambda_2)|^2 \qquad (26)\\
&\quad +3|M_g^1(\lambda_1,\lambda_2)-M_g^2(\lambda_1,\lambda_2)+2M_g^3(\lambda_1,\lambda_2)|^2\}\,.
\end{aligned}
$$

5.4 Phase Space Integration

Given the four-momenta q_1 and q_2 of the initial parton configuration, $\rho_i = |\mathbf{p_i}|$ $(i = 1...4)$ the magnitudes of of the four jet momenta, and $\mathbf{p}_\nu$ the neutrino momentum, and assuming perfect angular resolution for jets, the "interesting" part of the phase space looks like this:

$$d\Phi_6 = \delta^4(q_1 + q_2 - \sum_{i=1}^{6} p_i) \prod_{i=1}^{4} \frac{\rho_i^2 d\rho_i}{(2\pi)^3 2E_i} \frac{d^3\mathbf{p}_\nu}{(2\pi)^3 2E_\nu} \tag{27}$$

In Appendix C, we re-derive the phase space factor for $t\bar{t}$ decay. The phase space has sharp peaks both at the regions near the propagator peaks and the phase space boundaries. By choosing efficient integration variables we avoid singularities in the integrand. Furthermore, in order to ensure a numerically more stable integration, we perform a change of variables, using the squares of the top quark and the W invariant masses as integration variables in the following way: given a function $g(x) > 0$ for all x then $z(x) = \left(\int_a^b g(y)dy\right)^{-1} \int_a^x g(y)dy$ is a monotonously increasing function of x. Therefore the inverse function $x(z)$ always exists, and it is single-valued.

$$\int_a^b f(x)g(x)dx = \int_a^b g(x)dx \int_0^1 f(x(z))\, dz \tag{28}$$

The right side has significantly better numerical convergence properties than the left side if g is a highly peaked function (such as Breit-Wigner) and f is relatively smooth (such as the rest of the matrix element). This variable substitution works very well if the inverse $x(z)$ is easy to calculate. The above formula is applied twice using W propagators in place of g (and the W mass in place of x). It is also applied to the product of the transfer functions on the hadronic side, in which case the inverse function is evaluated numerically.

For this update, we include a two-dimensional integration over the transverse momentum of the $t\bar{t}$ system. In total we have four trivial (i.e. absorbed by δ functions) and seven non-trivial phase space integration variables. Among the latter are the square of the top quark masses on the hadronic (M_1^2) and on the leptonic side (M_2^2) of the event; the square of the mass of the hadronically (m_1^2) and of the leptonically decaying W boson (m_2^2); the parameter $\beta = \log \frac{\rho_1}{\rho_2}$, the ratio of the scalar momenta of the two jets from the hadronically decaying W; and the P_T of the $t\bar{t}$ system.

5.5 Effective Propagators

The phase space integration variables M_{thad}^2, M_{tlep}^2, m_{Whad}^2, and m_{Wlep}^2 are distributed according to the Breit-Wigner function with appropriate parameters for the top and W masses and widths. These known distributions are what make those quantities attractive variables of integration. Unfortunately the assumptions which simplify the solution of the $t\bar{t}$ system also change the distributions for the W and top mass from the simple Breit-Wigner function to a more complicated *"effective"* propagator functions. In particular, we assume the quark masses are always the same, the measured jet direction is the parent parton's, and the lepton momentum is perfectly measured. We have tried to account for these assumptions and recalculate the distributions of the integration variables. These corrected distributions can then be used in the integration instead of the simple Breit-Wigner propagators. In principle, the corrections should improve the mass resolution from the probability calculation because they account for deviations from the assumed perfect jet angle and jet mass measurements.

The construction of the effective propagators uses Herwig events which have been recalculated according to the integration assumptions. Herwig accounts for hadronization by first decaying the parent to on-shell daughters, then sending the daughters off-shell while preserving the four-momentum of the parent particle. We can take advantage of this fact to adjust the Herwig $t\bar{t}$ events so that the kinematics match our assumptions. Essentially, we take the momentum of the decay daughters in the rest frame of the parent, and reboost the daughters as if the hadronization were turned off. For the hadronic top, we take the original massless W decay daughters in the W rest frame, which is affected by the b hadronization. Then we find the momenta of the b and W before the b hadronization. We assume that the four-momenta of the W daughters in the pre-b-hadronization W rest frame are the same as the four-momenta of the W daughters in the post-b-hadronization rest frame. Then boosting back to the lab frame yields quarks from a hypothetical $t\bar{t}$ decay in which everything decayed on-shell. After the boost, the new quarks have different angles in the lab compared to their hadronized counterparts. We preserve the masses and energies of these tree-level quarks and rotate them into the measured particle jet angles. A similar boost is calculated for the leptonic top and applied to the W daughters.

The resulting top and W masses are now different from the corresponding masses in the original Monte Carlo event. Specifically, the new masses reflect the assumptions made in the matrix element integration. The purpose of the effective propagator is to describe the difference between these new top and W mass distributions with respect to the simple Breit-Wigner function. The effective propagators are generated event-by-event based on the expected errors for the jet masses and angles as well as the event Jacobian calculation.

The average error on the measured muon p_T is $\delta p_T = 0.1\% p_T^2$ and on the electron p_T is $\delta p_T = 13.5\% \sqrt{p_T}$. The resolution on the jet mass squared is $\delta m^2 = 241.7\,\mathrm{GeV}^2/\mathrm{c}^4$ for b jets and $\delta m^2 = 201.8\,\mathrm{GeV}^2/\mathrm{c}^4$ for light jets. The jet angular resolutions are also estimated from simulated events and are shown in Figure 9.

Then the effective hadronic W propagator is constructed in two steps. First, we note that the effect of the jet masses on the W mass squared can be expressed as a universal function whose width is proportional to the partial derivative of the hadronic W mass squared with respect to the light jet masses. This width-adjusted distribution can be convoluted with a Breit-Wigner function to give an effective propagator taking into account the jet mass effects. Second, the broadening of the propagator due to jet angle errors is applied as a Gaussian distribution with width predicted by the partial derivatives of the hadronic W mass squared with respect to the light jet angles. This predictor variable and the Gaussian are recalculated for each point in the original M_W^2 integration grid because both β and M_W^2 are necessary to calculate the separate p_q and $p_{\bar{q}}$. In this way, the effects of the quark angles are convoluted with the effects of the quark masses and the original Breit-Wigner distribution.

If the values of the predictor variables are small, the effective propagators are quite similar to Breit-Wigner propagators; however, large values of the predictor variables are associated with long tails (due to jet mass effects) and/or an overall broadening (due to jet

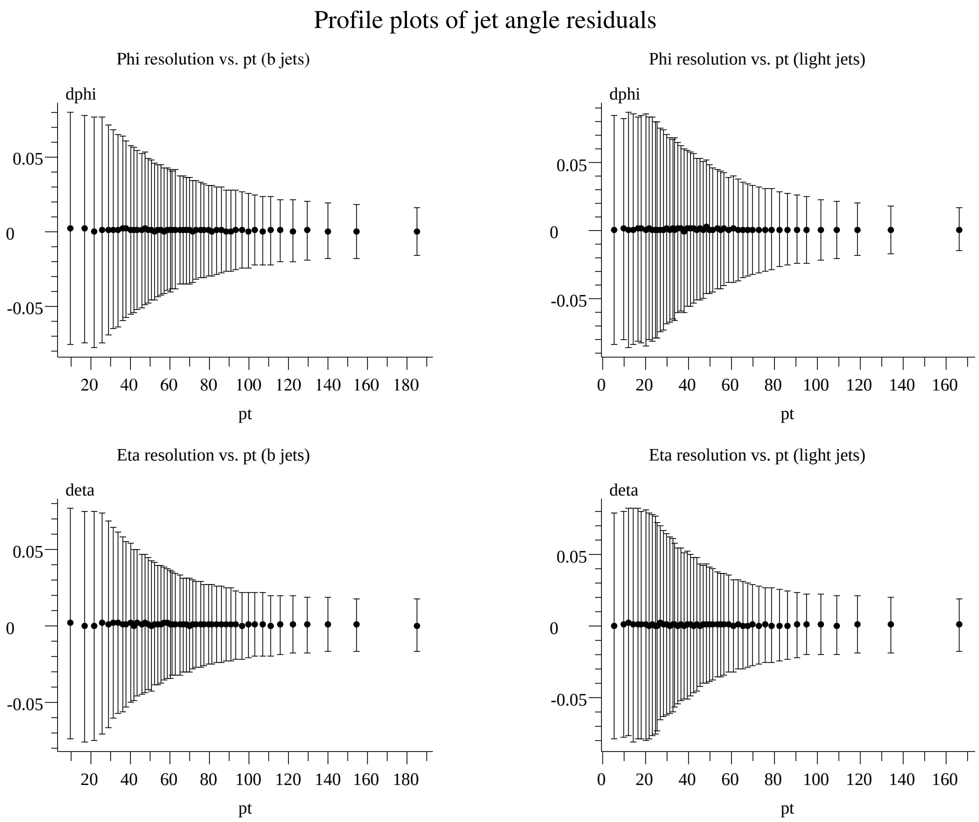

Figure 9: Profile plots of differences between measured jet and true parton angles for b jets (left) and light jets (right).

angle effects). Two such extreme examples are shown in Figure 10. The integration grid is generated as equidistant probability points which cover 97% of the derived propagator.

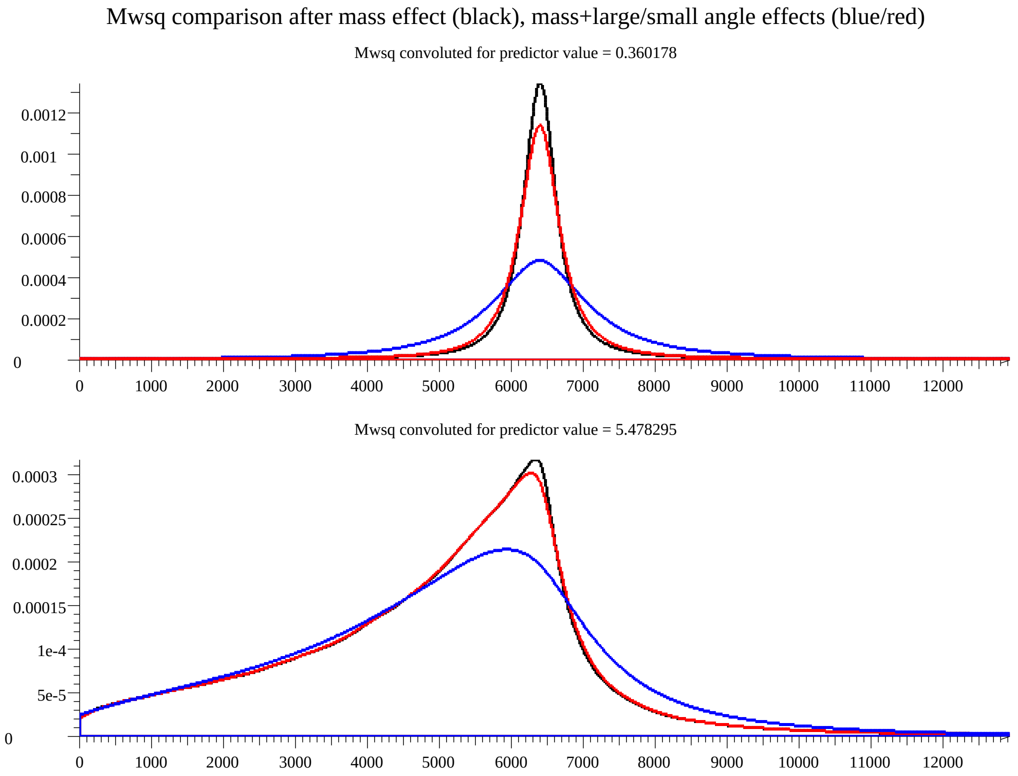

Figure 10: Effective hadronic W propagators under different conditions. The top plot is an example of the propagator for an event whose topology predicts a small smearing due to jet mass effects, as opposed to the bottom plot which shows the effective propagators for an event in which the jet mass effects are very large. In both cases, the effects of large and small angular smearing are shown with blue and red curves, respectively.

The hadronic side Jacobian between the non-integration variables and the integration variables, calculated analytically event-by-event, is used to derive the estimated errors on M_W^2 and M_t^2 from the input resolutions and event topology. The effective hadronic top propagators which relate M_t^2 to the pole mass m_t^2 are constructed from subsamples of events having the same estimated errors in M_t^2 and M_W^2. The distribution of physical hadronic top masses can change with different values of the hadronic W mass; this means that the hadronic top propagator is recalculated for each M_W^2 point in the integration.

Currently the leptonic side effective propagators are slightly simpler than the hadronic side. In this case the grid points in the leptonic side integration space are the same for each event. Nevertheless, these propagators are constructed from the same re-evolved Herwig events. The leptonic W integration grid is determined by the leptonic W mass for the particular slice of integration space – recall that the leptonic side is solved using input from the hadronic side. The difference in the leptonic and hadronic side propagators is most apparent from Figure 11.

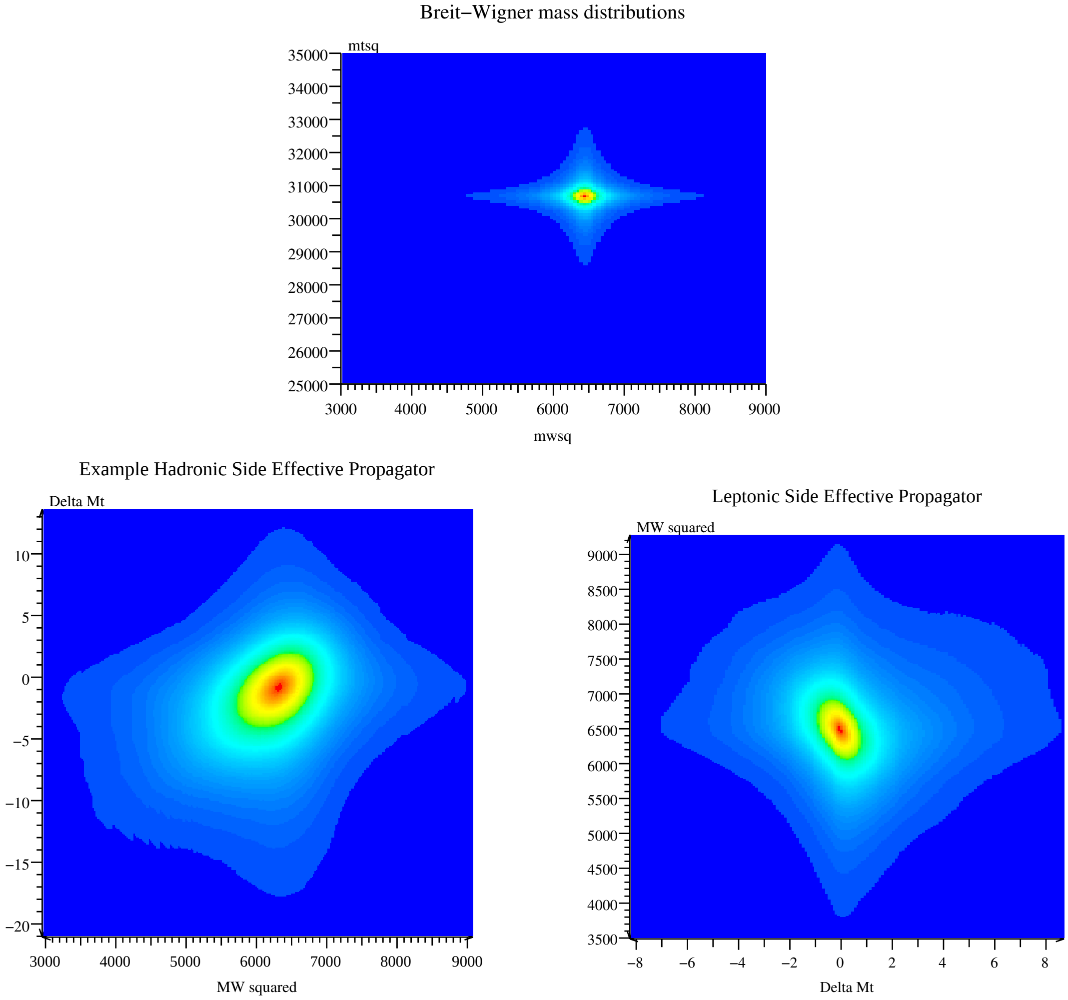

Figure 11: Distribution of physical W and top masses from Breit-Wigner function (top) and effective propagators for hadronic side (bottom left), leptonic side (bottom right).

It's worth noting that the transfer functions are defined using the transverse momenta of the recalculated quarks as well. The procedure described here has the effect of splitting up the usual transfer function definition into two parts: the transverse momentum transfer function (see Section 4) and the hadronization transfer function (jet masses and angles). The latter is treated by the effective propagators.

5.6 Trade-off of Statistical and Systematic Uncertainties

In our blessed MTM analysis [1] we used a permutation-by-permutation fit of the jet energy scale (JES). The idea of improving systematics by allowing JES to float within a constraint now took hold in other analyses. In our updated analysis we use a new approach to trade off statistical and systematic errors. We no longer perform an event-by-event JES fit because we are using all permutation, whereas before we only considered the "best" permutation (i.e. the solution with the lowest χ^2 output by our previous kinematic fitter). We could use a Bayesian approach and treat JES as a nuisance parameters by integrating over JES with a Gaussian prior. Frequentists call this approach "marginalization of the likelihood." Instead we build the profile likelihood, an approach which is used by the classic template analysis, as well as by the MEAT analysis.

5.7 Integration Parameters

We integrate over the seven variables specified above: the hadronic (M_1^2) and on the leptonic side (M_2^2) of the event, the squared of the mass of the hadronically (m_1^2) and of the leptonically decaying W boson (m_2^2), and the parameter $\beta = log\frac{\rho_1}{\rho_2}$, the ratio of the scalar momenta of the two jets from the hadronically decaying W. We also integrate over $p_T(t\bar{t})$. Finally, we sum over all jet-to-parton assignments ("permutations") as well as of the two solutions for the neutrino momenta. We weigh each permutation according to the probability that the jet assignment is correct.

The weights for each permutation are assigned as follows. We assume that the event is a $t\bar{t} \to$ lepton+jets event. Each jet in an event has a tag probability b if it is a b-jet, c if it is a c-jet, l if it is a light (u or d or s) jet. These probabilities depend on kinematic variables P_T and η. Then we can write down relative probabilities for each permutation. The probability for a tagged b-jet to be assigned to a parton that is a W decay daughter is non-zero , as is appropriate because of mistags and the non-zero probability of tagging a c-jet from a W decay. We assume that the two decays W $\to u\bar{d}$ and W $\to c\bar{s}$ have equal probability. Then these probabilities, properly normalized, are used as the weights. The efficiencies for tagging a b-jet as a function of PT and eta, as well as the mistgag probabilities are taken from the b-tagging group WEB page. The probability that the tagged jet is a charm jet is taken to be 22% of the tagged jets.

For each event we use 64 top mass points between 110 and 267.5 GeV. All of the integrals are evaluated numerically, with the number of points in each grid determined empirically. We have started with a large number of points in each grid and decreased the numbers until we see the effects in the output probabilities. The W mass on the hadronic side is sampled using 25 intervals equidistant in the cumulative Breit-Wigner probability; these intervals change from event to event. The W mass on the leptonic side is sampled at 5 points. The W mass integration, by design, covers 97% of the Breit-Wigner area, leaving out 1.5% tails at both ends. For the integration over the ratio of momenta from the hadronically decaying W, we use 17 points equally spaced in the parameter β. For the integration over $p_T(t\bar{t})$ we chose a very large grid, 301 points, distributed as shown in Fig 12. Earlier attempts to integrate over $p_T(t\bar{t})$ failed, because our grid was too small. This choice of grids would require a large amount of computer time for each event, so we reduce the grids by a "duty cycle", i.e., we take 1 in n points ($n \sim 2000$). We have tuned for this parameter. We use "Sobol sequences" [24] to pick up the points. These

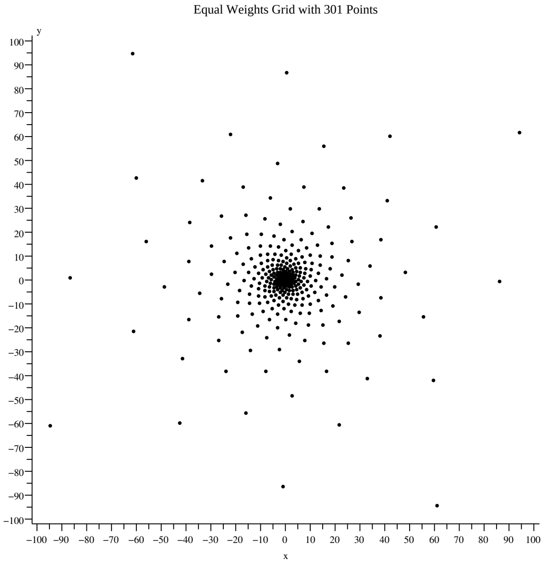

Figure 12: Visualization of the $t\bar{t}$ P_T integration grid. Each "ring" has a four-fold symmetry at an average radius. Consecutive "rings" are rotated by a constant relative phase.

are quasi-randon numbers, but are guaranteed to have coverage over the entire range. In practice we assign each grid point an integer index and evaluate that grid point if the index is generated in the Sobol sequence.

We repeat the integration 51 times, by modifying the jet momenta by a factor JES, which spans between 0.85 and 1.15. In practice we use as shift a fraction of sigma=0.03, a constant for each jet, with σ, varying between -5σ and +5σ. This is the systematic uncertainty for jets with $p_T > 70$ GeV, as provided by the Jet Correction Group [4]. The curves at different values of JES are then used for the 2D pseudoexperiments, described in Section 5.8.

5.8 Pseudo-experiments

The method by which we construct a PE is very simple. We take a random Poisson-distributed number of events, N_ev, from our sample of integrated events and multiply their likelihoods to obtain an ensemble likelihood. We sum the logs of the probabilities as follows:

$$\log L(m_t, \mathrm{JES}) = \sum_{\mathrm{events}} L_i(y_i|m_t, \mathrm{JES}) \tag{29}$$

This produces a two-dimensional likelihood curve. In order to reduce this to a single-dimensional likelihood curve, we eliminate the "nuisance variable" JES using the profile likelihood. The profile likelihood simply takes the maximum value along the JES axis for each m_t value. That is,

$$L_{\text{prof}}(m_t) = \max_{j \in \text{JES}} L(m_t, j) \tag{30}$$

We then take the peak of the profile likelihood as our measurement and its width (obtained by descending 1/2 unit from the peak value) as our measured error. The pull is given by the difference between the peak mass and the nominal value of the mass divided by the returned error for each event. The width of this distribution is the quoted pull width.

5.9 Test of the Method

The integration technique was tested using samples of simulated $t\bar{t}$ Monte Carlo events and looking at the correlation between the reconstructed and the generated top quark mass. We use realistic Johnson's transfer functions, the full Kleiss-Stirling matrix elements and perform the integration for 51 values of the JES parameter. We then perform pseudo-experiments of 116 events at each generated mass point and evaluate the profile likelihood for each top mass. The relation between generated and reconstructed top mass is shown in Figure 13. Results on bias and pulls are also shown in for the total sample of simulated events (left). The sample is contaminated by events for which the W decays into $\tau + \nu$. This can be a source of top mass bias. At this purpose we remove these events from the sample and plot the slope, bias and pulls for the remaining sample (right plots). We observe that bias and pulls are reduced.

More information can be obtained by looking at the 1D plots for the bias and the pulls. These are obtained at the JES =1, i.e. adding a shift value = 0.0 (i.e. no shift) for the mass plots and at a mass of 175 GeV for the JES plots. See Fig 14.

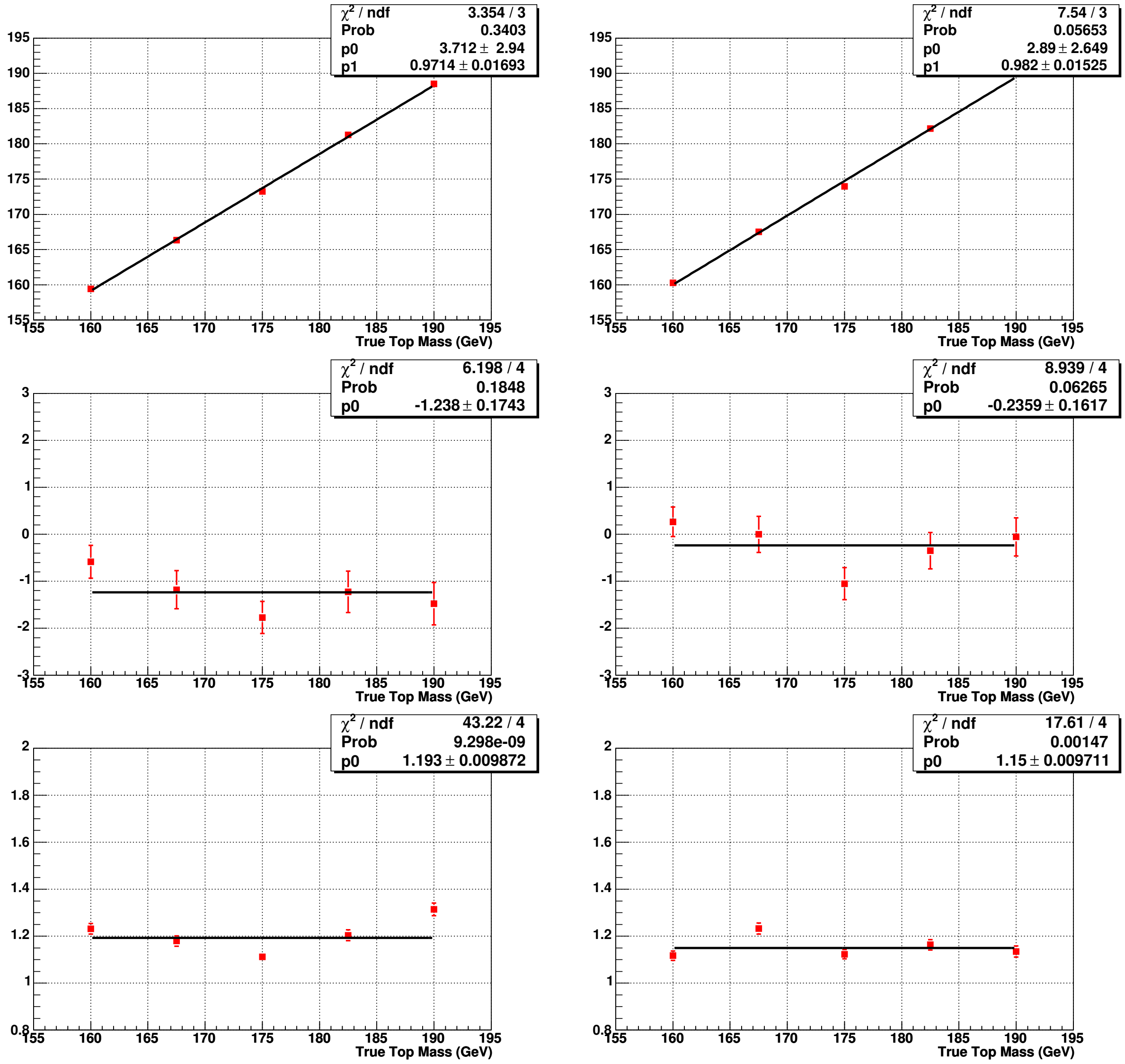

Figure 13: Reconstructed top mass versus top mass from pseudo-experiments using full matrix elements and realistic transfer functions. Full simulation is used. From top to bottom: reconstructed versus true top mass, bias *vs* true top mass, pulls of distribution vs true top mass. Left: all events. Right: events with W $\to \tau$ removed.

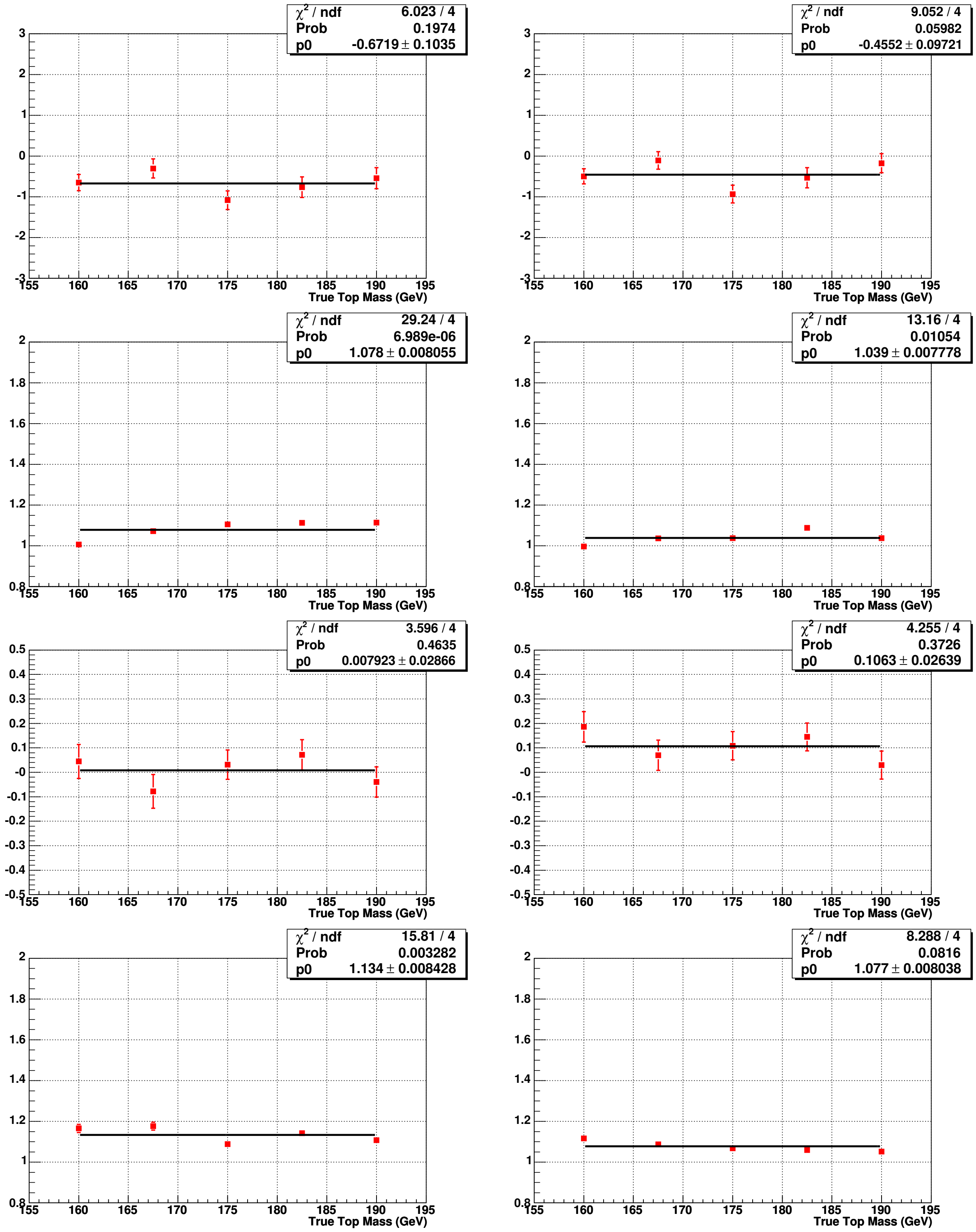

Figure 14: Bias and pulls versus top mass for the 1D mass and 1D JES fit from pseudo-experiments of 116 events: full sample (left), sample with events with $W \to \tau$ removed (right). Top two plots: bias *vs* true top mass, pull *vs* true top mass for the 1D mass fit. Bottom two: bias *vs* true top mass, and pulls *vs* true top mass for the 1D JES fit. Here JES is expressed in terms of a shift from 1. of a fraction of a σ with σ=0.03. It corresponds to JES=.85 to 1.15.

6 Optimization of an Additional Variable

In our previous studies [1] we have investigated various additional kinematic variables to be used to build multi-dimensional templates in order to add mass and signal/background information for the mass reconstruction. We have considered both pure energy quantities and dimensionless shape/angle variables. From MC studies of the global correlation between various sets of variables and the generated top mass it turned out that adding an additional variable doesn't increase the top mass sensitivity but might be very useful to discriminate signal from background events. Unfortunately it turned out that the variable candidates that we tried in the MTM analysis introduced additional systematic uncertainties in the top mass determination that might compensate the advantages of better S/B suppression.

In our present analysis the peak likelihood top mass m as described in Sect. 5 serves as our primary template variable. We have performed further optimization studies for a second template x variable which are led by the following objectives:

1. x should improve signal/background for *all* types of background.
2. x should not be affected by large systematic uncertainties.
3. Ideally we find a mass independent variable x which allows for a much simpler treatment of the signal likelihoods. In this case we can either switch to 1-d templates for x and m to be multiplied event-by-event, or we can multiply the top mass likelihood $\mathcal{L}(m) = (1 - f_b(x))L_{t\bar{t}}(m) + f_b(x)U(m)$ for all events (U is uniform distribution, f_b is the background fraction, N_b is the number of background events).

In the following we introduce a set of kinematic variables as a starting point and investigate their capability to discriminate signal from various backgrounds using histogram divergence measures (Sect. 6.3). We also discuss their mass dependence (Sect. 6.4) and their systematic uncertainties (Sect. 6.5). Based on these observation we propose a hybrid variable (Sect. 6.6) which we try to optimize for our needs (Sect. 6.7).

6.1 Variables

$t\bar{t}$ events are characterized by larger transverse energies and more spherical topologies compared to events from typical background processes such as W plus jets or non-W QCD processes. These gives reason to consider the following set of observables. All quantities are calculated at jet correction level 7.

- Kinematic energy variables:
  - The transverse momenta of the four leading jets: $p_{\mathrm{T}}^{(i)} (i = 1 \ldots 4)$.
  - Various combinations of sums of the transverse jet momenta among the four leading jets:
    $p_{\mathrm{T}}^{(i+j)} \equiv p_{\mathrm{T}}^{(i)} + p_{\mathrm{T}}^{(i+1)}$, $p_{\mathrm{T}}^{(1-3)} \equiv \sum_{i=1}^{3} p_{\mathrm{T}}(i)$, $p_{\mathrm{T}}^{(1-4)} \equiv \sum_{i=1}^{4} p_{\mathrm{T}}^{(i)}$.

- The transverse energy of the charged lepton: $E_{\mathrm{T}}^{(l\pm)}$.
- The transverse missing energy: $\not{E}_{\mathrm{T}}$.
- The scalar sum of the missing transverse energy, the transverse momentum of the charged lepton, and the transverse momenta of all jets:
  $H_{\mathrm{T}} \equiv \sum_{\mathrm{jets}} p_{\mathrm{T}}^{(i)} + E_{\mathrm{T}}^{(l\pm)} + \not{E}_{\mathrm{T}}$.
- The minimum di-jet mass among the four leading jets:$M_{ij}^{(\mathrm{min})}$.

• Shape/angle variables (generally calculated from the momenta $\vec{p_i}$ of the jets and the charged lepton):

  - $|\eta|^{(\mathrm{max})}$, the largest magnitude pseudo-rapidity value of the three leading jets.
  - The sphericity $S = \frac{3}{2}(Q_1 + Q_2)$ and the aplanarity $A = \frac{3}{2}Q_1$. Q_i are the eigenvalues $Q_1 < Q_2 < Q_3$ of the momentum tensor $\Theta_{ab} = \sum_i p_a^i p_b^i / \sum_i |\vec{p_i}|^2$. No missing momentum information is included in this version.
  - The thrust $T = \max_{\vec{n}}(\sum_i |\vec{p_i} \cdot \vec{n}| / \sum_i |\vec{p_i}|)$ is obtained by varying the unit vector $\vec{n}$ such that the expression in brackets is a maximum. Given the direction of the thrust axis $\vec{n_T}$ thus obtained, the trust major $T_{\mathrm{maj}} = \max_{\vec{n}\vec{n_T}=0}(\sum_i |\vec{p_i} \cdot \vec{n}| / \sum_i |\vec{p_i}|)$ is defined in a similar way by varying the unit vector within a plane perpendicular to the trust axis. Given the thrust major axis $\vec{n_{\mathrm{maj}}}$ thus obtained, the thrust minor is defined by $T_{\mathrm{min}} = \sum_i |\vec{p_i} \cdot \vec{n_{\mathrm{min}}}| / \sum_i |\vec{p_i}|)$ using the axis $\vec{n_{\mathrm{min}}}$ perpendicular to $\vec{n_T}$ and $\vec{n_{\mathrm{maj}}}$. The oblateness is defined by $O = T_{\mathrm{maj}} - T_{\mathrm{min}}$.
  - The acoplanarity $A = 4\min(\sum_i |\vec{p_i}|^{\mathrm{out}} / \sum_i |\vec{p_i}|)^2$. $\vec{p_i}^{\mathrm{out}}$ is the component of p_i perpendicular to the event plane. The plane is varied such that the expression in brackets is minimized. In the version used here, the "missing momentum" derived from the other particles is added as an extra particle.
  - $D_{\mathrm{R}} = \Delta R_{ij}^{(\mathrm{min})} \times \min(p_z^{(i,j)}) / p_{\mathrm{T}}(l^{\pm})$, i.e. the minimum separation $\Delta R_{ij}^{\mathrm{min}}$ in (η,ϕ) space between a jet pair weighted with the ratio of the lower p_{T} of the jet pair and the charged lepton p_{T}.
  - $H_{\mathrm{TZ}} = \sum_{i=2}^{4} |p_{\mathrm{T}}^{(i)}| / (\sum_{i=1}^{4} |p_z^{(i)}| + |p_z^{(l\pm)}| + |p_z^{(\nu)}|)$, i.e. the ratio of scalar sums transverse and longitudinal momenta. The numerator contains all jets except the leading jet. The denominator sums all jets, the charged lepton and the neutrino. The smaller $|p_z^{(\nu)}|$ solution given by the kinematic equation for the leptonic W decay (assuming $M_W = 80.4\mathrm{GeV}/c^2$) is taken.
  - We build a hybrid variable V as a linear combination of A, D_{R} and H_{TZ}, see Sect. 6.6.

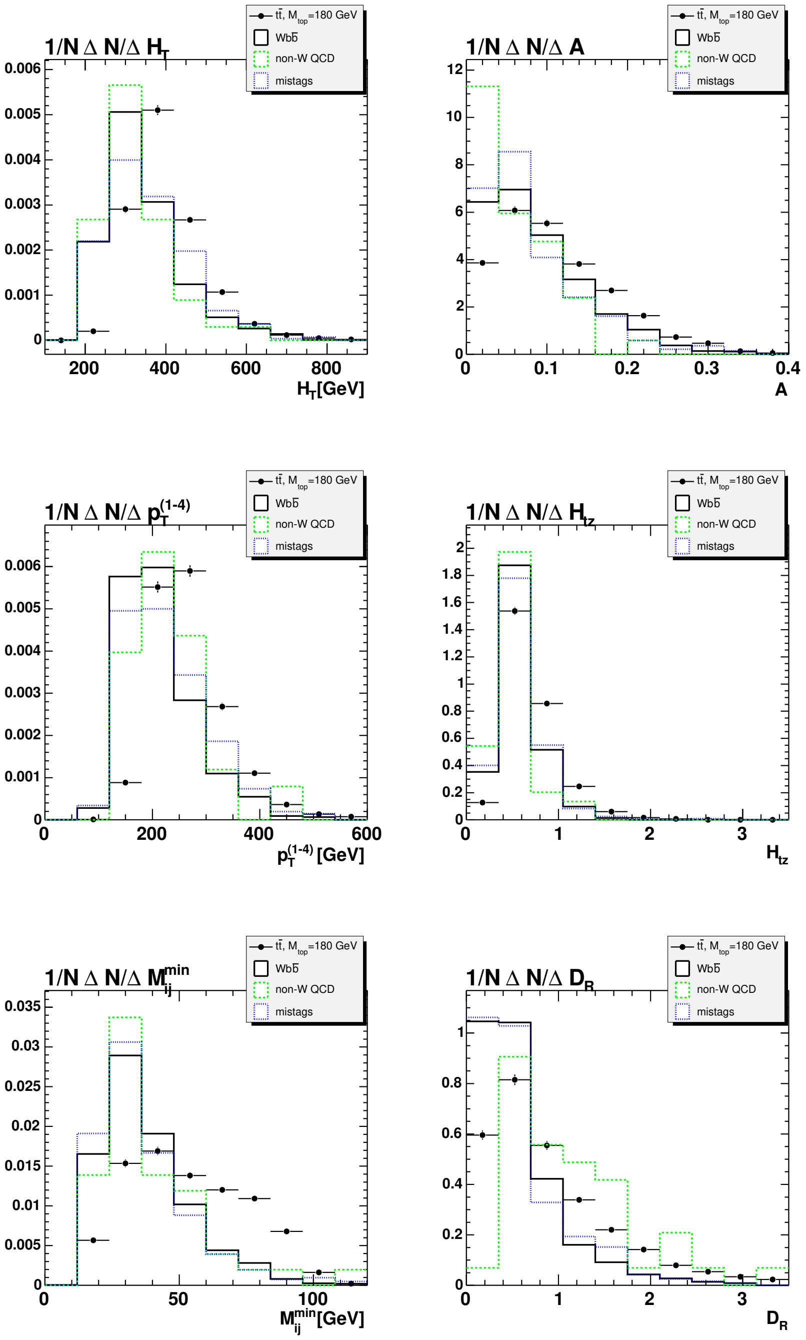

Figure 15: The normalized distributions of $p_{\mathrm{T}}^{(1-4)}$, H_{T} and $M_{ij}^{(\mathrm{min})}$ (upper row) and A, H_{TZ}, and D_{R} (lower row) calculated from tagged simulated events for signal $t\bar{t}$ (m_t=180 GeV, points), $Wb\bar{b}$ (black solid line), non-W QCD (green dashed line) and mistags (W+light jets, blue dotted line).

6.2 Samples for Signal and Background Distributions

For our signal/background separation studies we use Herwig $t\bar{t}$ samples and consider the following background samples [5]: $Wb\bar{b}$ ALPGEN+Herwig, non-W QCD (from data with $\not{E}_{\mathrm{T}} > 20\,\mathrm{GeV}$ and lepton isolation cut > 0.2) and mistags (from ALPGEN+HERWIG W+light jets). We impose the standard top selection criteria to the events and require exactly four tight jets ($E_{\mathrm{T}} > 15\,\mathrm{GeV}$, $|\eta| < 2.0$) and ≥ 0 loose jets ($8\,\mathrm{GeV} < E_{\mathrm{T}} < 15\mathrm{GeV}$, $|\eta| < 2.0$). We consider both tagged and non tagged events. In addition to the standard cuts, we reject non tagged events with $20\,\mathrm{GeV} < \not{E}_{\mathrm{T}} < 30\,\mathrm{GeV}$ if the azimuthal angle between the leading jet and the missing transverse momentum is smaller than 0.5 or larger than 2.5.

Fig. 15 shows the distributions thus obtained for typical energy variables ($p_{\mathrm{T}}^{(1-4)}$, H_{T} and $M_{ij}^{(\mathrm{min})}$) and some shape variables (A, H_{TZ} and D_{R}). The distributions are normalized to 1.

6.3 Quantifying the S/B distance

The differentness between signal and background distributions like those shown in Fig. 15 can be quantified using various divergence measures, see [1] for details. In the following we consider the K distance between two probability distributions $p_1(x)$, $p_2(x)$ of a variable x as defined by

$$\begin{aligned} d_K(p_1, p_2) &= \sum_{i=1}^{n} p_1(x_i) \log\left(\frac{2p_1(x_i)}{p_2(x_i) + p_1(x_i)}\right) \\ K &= d_K(p_1, p_2) + d_K(p_2, p_1) \end{aligned} \tag{31}$$

where n is the number of bins and x_i is the content of bin i. In order to compare the S/B distances for different variables whose numerical values x_i span different ranges, one has to map the typical range of values of a variable onto uniform kinematic range prior to evaluating Eq. 31. Since the distances explicitly depend on the number of bins, we also have to ensure that the probability distributions of all variables have the same binning. Due the limited statistics of our non-W QCD samples we have chosen a relatively coarse binning of 10 bins per distribution and also verified that our observations do not change if we double the number of bins. Since there is no accepted gold standard for distance measures, it is useful to consult further measures. Among other things we also look at the Kolmogorov-Smirnov distance (KS) which is the maximum bin-by-bin difference between the cumulative distributions $\int p_1(x)\mathrm{d}x$ and $\int p_2(x)\mathrm{d}x$.

The upper plot of Fig. 16 shows the distances between signal and $Wb\bar{b}$ background. The variables are sorted such that pure energy variables are arranged on the left side and shape variables on the right side. Generally energy variables do a better job than jet quantities, in agreement with our previous observations [1] based on looser selection

[5] We are still using generation 4 samples for this study, but the Gen5 background samples happen to be the same events. The largest possible difference is due to changes in the jet energy corrections.

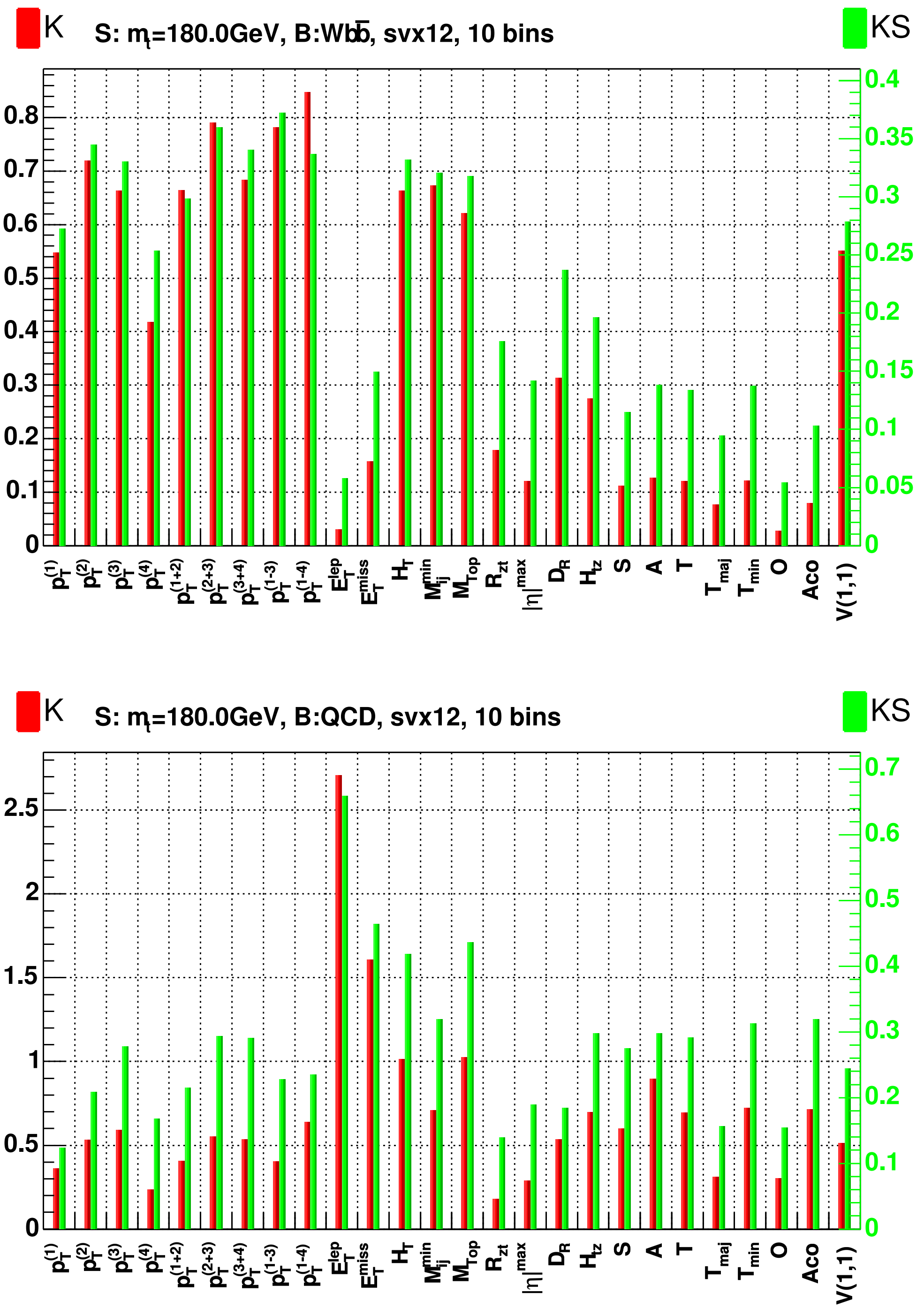

Figure 16: The K (red bars) and the KS distance (green bars) between $t\bar{t}$ signal with a 180 GeV generated top mass and background distributions (top: $Wb\bar{b}$, bottom: non-W QCD) of various variables for tagged events. Pure energy variables are grouped on the left side and shape variables on the right side.

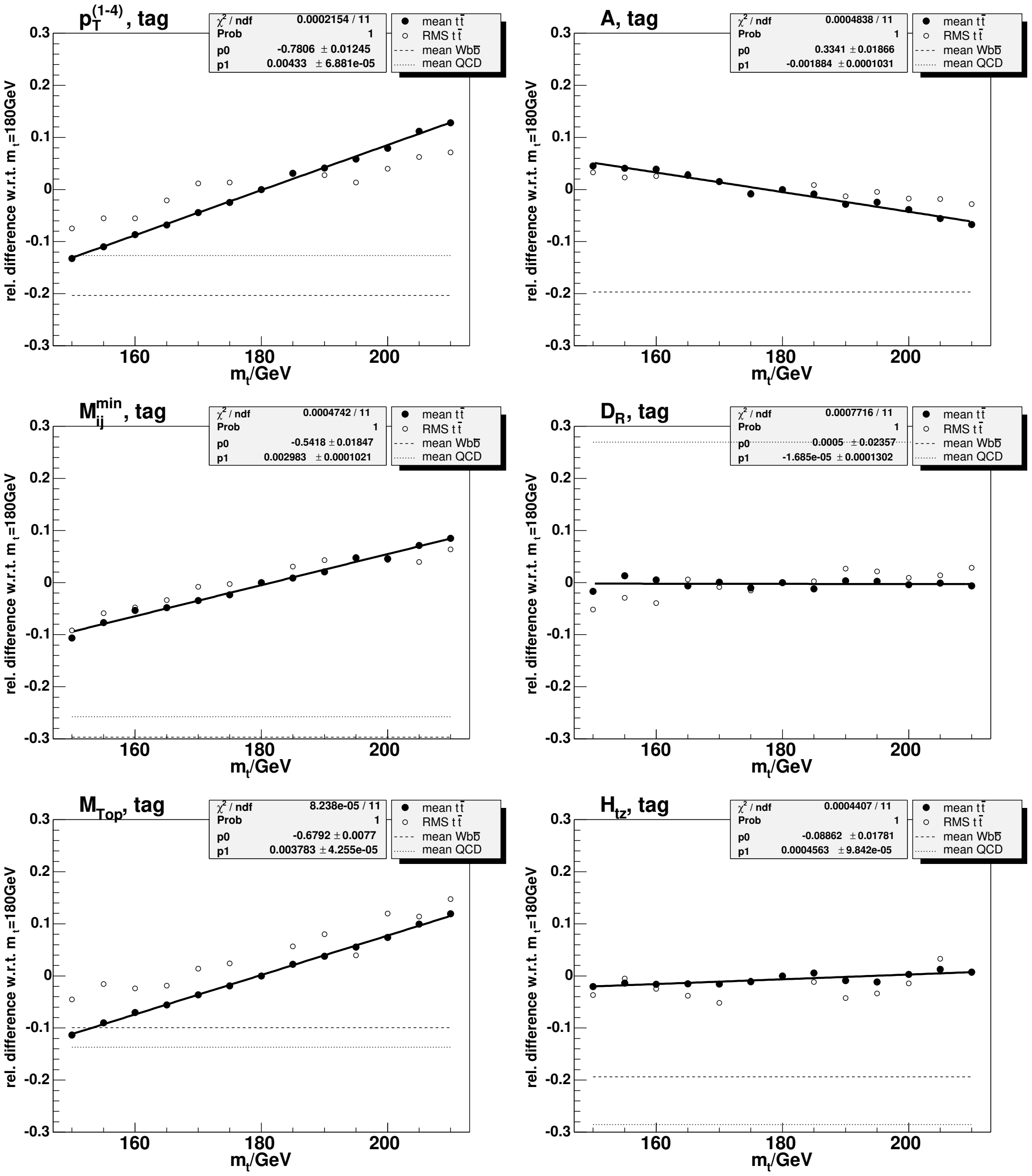

Figure 17: Mean values (solid circles) and the r.m.s. values (empty circles) of the signal distributions versus generated top mass m_t. Shown are the relative changes with respect to the value at m_t=180 GeV for $p_{\mathrm{T}}^{(1-4)}$,$M_{ij}^{(\min)}$, the reconstructed (obsolete) top mass M_{top} from our former kinematic fitter (left column) and for aplanarity A, D_{R} and H_{TZ} (right column). The solid lines represent linear fits to the mean values. For comparison, the mean values from the $Wb\bar{b}$ (dashed line) and non-W QCD background distributions (dotted line) are also shown (if inside range of values).

cuts. The best discriminator is the E_{T} sum of the four leading jets, but further sum combinations have similar performance. The best shape variables are D_{R} and the hybrid V (here we show the result for V using standard coefficients, see Sect. 6.6). It is notable that the hybrid variable performs as well as the former reconstructed top mass M_{top}. For the zero tag case we observe a similar pattern.

The lower plot of Fig. 16 summarizes the results for the non-W QCD background case. We observe that the signal distances to this background type is larger than to the $Wb\bar{b}$ background. Furthermore shape variables perform generally better S/B separation than energy variables. On exception is the charged lepton E_{T}, but this variable is correlated with the isolation cut and might hence not be very useful. Best discriminator among the shape variables is the aplanarity. The hybrid V is an average performer but can be optimized (see discussion later). Events with zero tags give a similar picture except that the aplanarity is more clearly favored than other shape variables.

6.4 Mass Dependence

The top mass dependence of the variables can be quantified using the mean and root mean squares (r.m.s.) of the signal distributions. In Fig. 17 we show the relative changes of theses values for $p_{\mathrm{T}}^{(1-4)}$, $M_{ij}^{(\mathrm{min})}$, A, D_{R} and H_{TZ} with respect to the value at m_t=180 GeV, as a function of generated top masses m_t between 150 GeV and 210 GeV. As a reference we also show the mass dependence of the reconstructed top mass M_{top} based on our previous kinematic fitter [1]. It is noticeable that e.g. $p_{\mathrm{T}}^{(1-4)}$ is more mass sensitive than M_{top}. Shape variables exhibit much less mass dependence than the energy variables. Hence it would be natural to start with shape variables for the construction of a variable with almost perfect mass independence. For example, Fig. 17 suggests that the aplanarity and the H_{TZ} variable can be combined in such a way that their slopes mutually cancel.

6.5 Systematic Uncertainties

Similarly we studied the change of the mean and the r.m.s. of the distributions when varying the jet energy scale by ±1 standard deviation of its total systematic uncertainty as given in [4]. Fig. 18 illustrates the results for the same variables as mentioned above. From these and other plots it turned out that energy variables are generally affected by larger uncertainties. However, by combining variables with opposite tendencies when varying JES (such as A and D_{R}) one can reduce systematic effects.

6.6 Hybrid Variable

We have tried to design a hybrid variable in which the JES systematics and top mass dependences of its individual part mutually cancel, and in which the signal to background discrimination is optimal for all types of background. Inspired by the observations men-

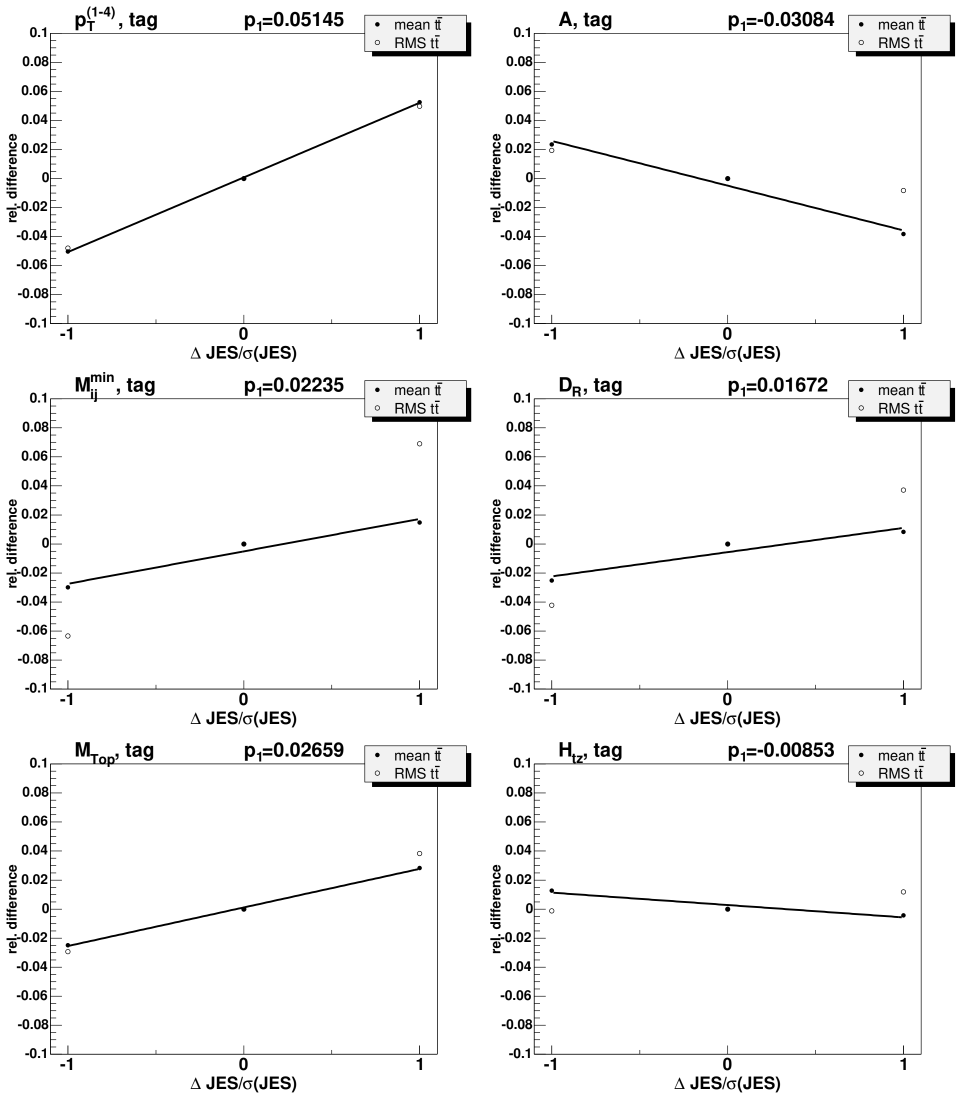

Figure 18: Relative changes of the mean values (solid circles) and the r.m.s. (empty circles) of the signal distributions versus variation of the jet energy scale by $\pm$ one standard deviation for $p_{\rm T}^{(1-4)}$,$M_{ij}^{(\rm min)}$, the reconstructed (obsolete) top mass $M_{\rm top}$ from our former kinematic fitter (left column) and for aplanarity A, $D_{\rm R}$ and $H_{\rm TZ}$ (right column). The solid lines represent linear fits to the mean values.

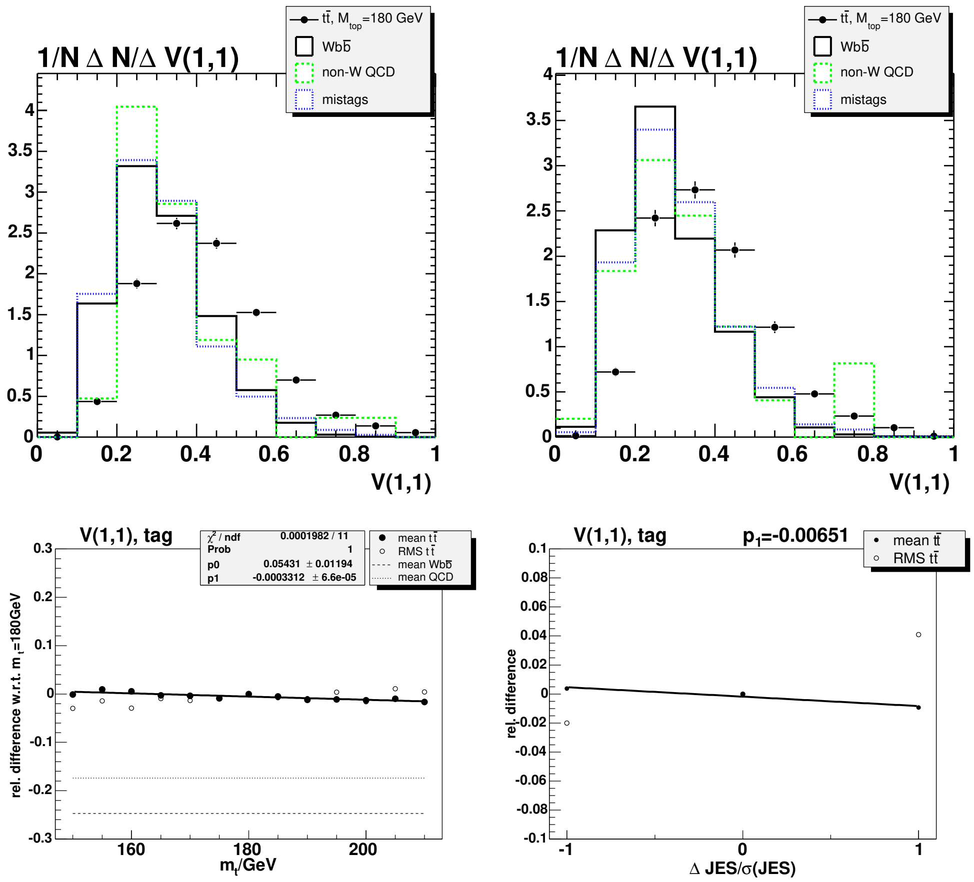

Figure 19: Hybrid variable $V(1,1,1)$. The upper row shows the distributions for the signal and the various backgrounds (normalized to 1) for tagged (left) and zero tag case (right). The lower left plot shows the relative change of mean (solid circles) and r.m.s. (empty circles) of signal events versus generated top mass, the lower right plot shows the corresponding relative changes versus one standard deviation variation of the jet energy scale.

tioned above we considered shape variables and define[6]

$$V(c_1, c_2, c_3) = \left(c_1 \frac{A}{0.4} + c_2 \frac{D_\mathrm{R}}{3.5} + c_3 \frac{H_\mathrm{TZ}}{2.0} \right) \times N(c_1, c_2, c_3) \tag{32}$$

with A, D_R and H_TZ defined in Sec. 6.1 and coefficients c_1, c_2 and c_3. Note that among the shape variables discussed previously this construction contains the best S/B discriminator (in terms of K distance) for both $Wb\bar{b}$ as well as for non-W QCD background. Each component of V is normalized by its typical kinematic range. N is a normalization factor chosen such that the typical range of V values are identical for different sets of coefficients c_i. This requirement is very important for the evaluation and comparison of the distances

[6] This was also a toy variable for top mass studies in Run I.

for different parameter sets, but not for the calculation of likelihoods itself. We define $N(c_1, c_2, c_3) = 1.4/(c_1 + c_2 + c_3)$ so that the typical range of V values is 0 to 1.

In Fig. 19 we show the V distribution for $c_1 = c_2 = c_3 = 1$. The S/B distances for this variable are shown by the rightmost bars in the plots of Fig. 16. For the $Wb\bar{b}$ case, the discriminative power of V is competitive with energy variables (and even better than the reconstructed top mass from our previous kinematic fitter). In the same time, the top mass dependence and the JES systematics (see bottom plots of Fig. 19) is much smaller compared to $M_{ij}^{(\min)}$or $p_{\mathrm{T}}^{(1-4)}$. For the non-W QCD case, our default V is only average performer, but we might find another parameter set that improves S/B separation for this case.

6.7 Optimization

We have two out of three c_i coefficients to play with because the normalization of the V range is not relevant for the templates, In the following we fix $c_3 \equiv 1$ and scan the (c_1, c_2) space, i.e. the weights for aplanarity and D_{R}, to find an optimal parameter set.

Fig. 20 shows the K distance of the signal V distribution and both the $Wb\bar{b}$ and the non-W QCD background V distributions versus (c_1, c_2) within $0 < c_1 < 6$, $0 < c_2 < 6$. The spectrum of colors covers blue, which encodes small S/B separation power, to red, which encodes maximum S/B separation power observed. Obviously regions of high discriminative power for one background type are disfavored by the other background type. Choosing large weights for the D_{R} part of the hybrid variable is helpful to suppress $Wb\bar{b}$ background (not surprising according to Fig. 16) but is not very useful to suppress non-W QCD background. The lower left quadrant of the parameter space covers a region with high discrimination power for both types of backgrounds. "Good" regions in terms of K distance are (roughly)

for $Wb\bar{b}$ background: $0 < c_1 < 0.8$, $0.3 < c_2 < 2.0$

for non-W background: $c_1 > 0.2$, $c_2 < 0.3$

In case of zero tag events we find a similar scenario, except that for non-W QCD suppression higher weights for A are preferred (e.g. $c_1 > 4.6$ in terms of K distance).

How much can we improve with respect to our default $V(1,1,1)$? For example, looking at tagged events for $c_1 = 0.4$, $c_2 = 0.8$ the signal K distance from $Wb\bar{b}$ background is $\sim 20\%$ higher compared to the default case $c_1 = c_2 = 1$ and about the same for non-W QCD background. If we instead choose $c_1 = 0.4$, $c_2 = 0.0$ we improve the K distance from non-W QCD background drastically by almost a factor 3 but at the same time deteriorate the distance from the $Wb\bar{b}$ by a roughly a factor 0.5. A good compromise parameter set might be $c_1 = 0.3$, $c_2 = 0.3$.

Keeping these observations in mind, we focus in the following on this region and study the mass dependence of V by looking at different aspects of the V distribution. It should be noted that the hybrid variable has by construction already a very little mass dependence (Fig.20), so the following should be considered as an additional "fine tuning" of the V coefficients. We quantify the mass evolution of the mean and the r.m.s. of V

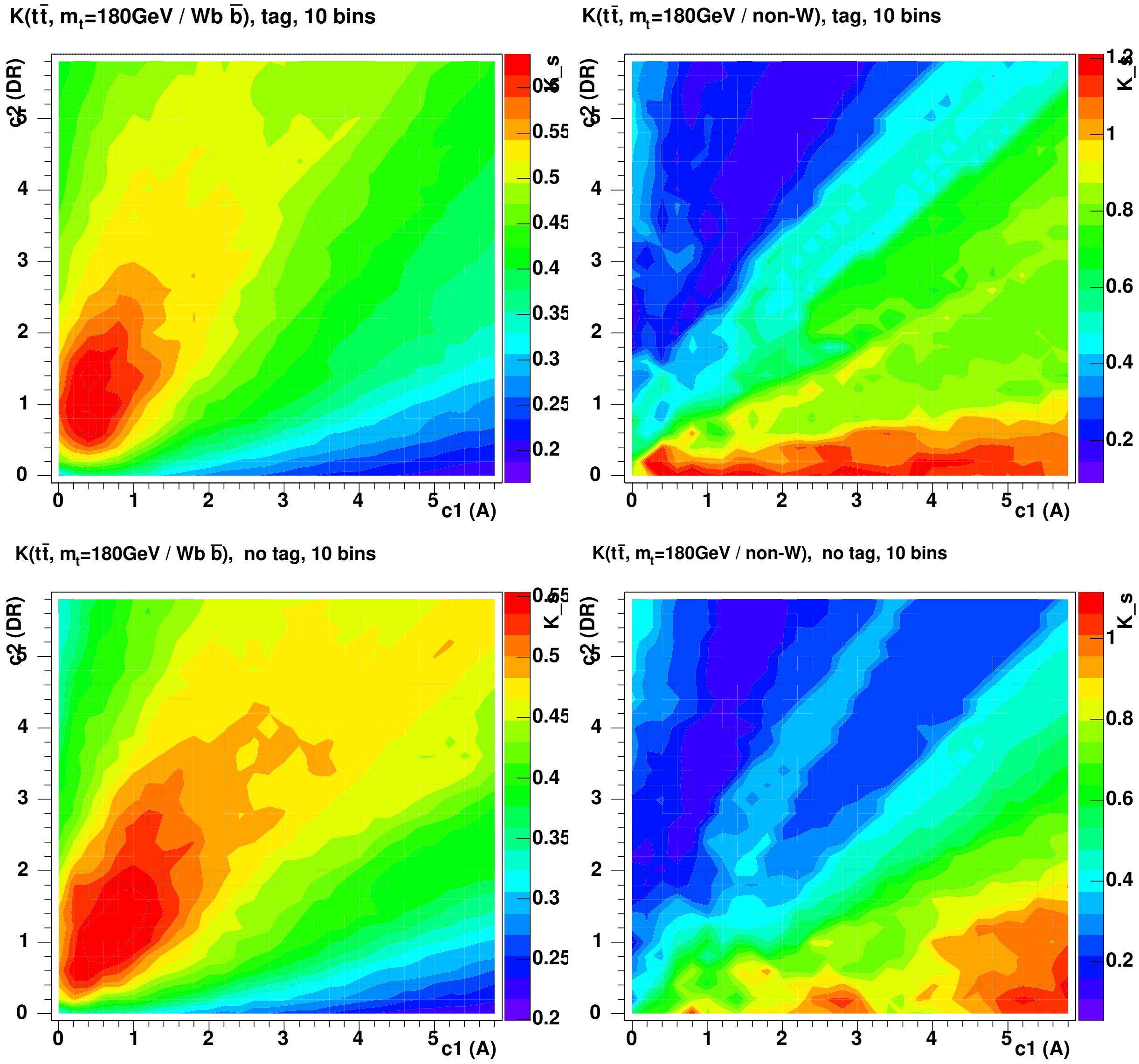

Figure 20: S/B discrimination power in terms of K distance between the signal distributions and the $Wb\bar{b}$ (left column) and non-W QCD background distributions (right column) of V for tagged for tagged (upper row) and non tagged events (lower row).

by using the slopes obtained from a line fit to the respective values versus m_t within the range $m_t = 150 - 210\,\mathrm{GeV}/c^2$. In addition we look at the mass evolution of the K and KS distances of V with respect to a reference distribution (we choose the distribution at m_t=150 GeV). In Fig. 21 we show the absolute values of the four slopes thus obtained versus (c_1, c_2) after we have zoomed into our region of interest $0 < c_1 < 2.5$, $0 < c_2 < 2.5$ with higher resolution. Blue colors encode zero or little mass, red color represent maximum mass dependence observed.

The diagram for the mean of V can be understood qualitatively by recalling Fig. 17.

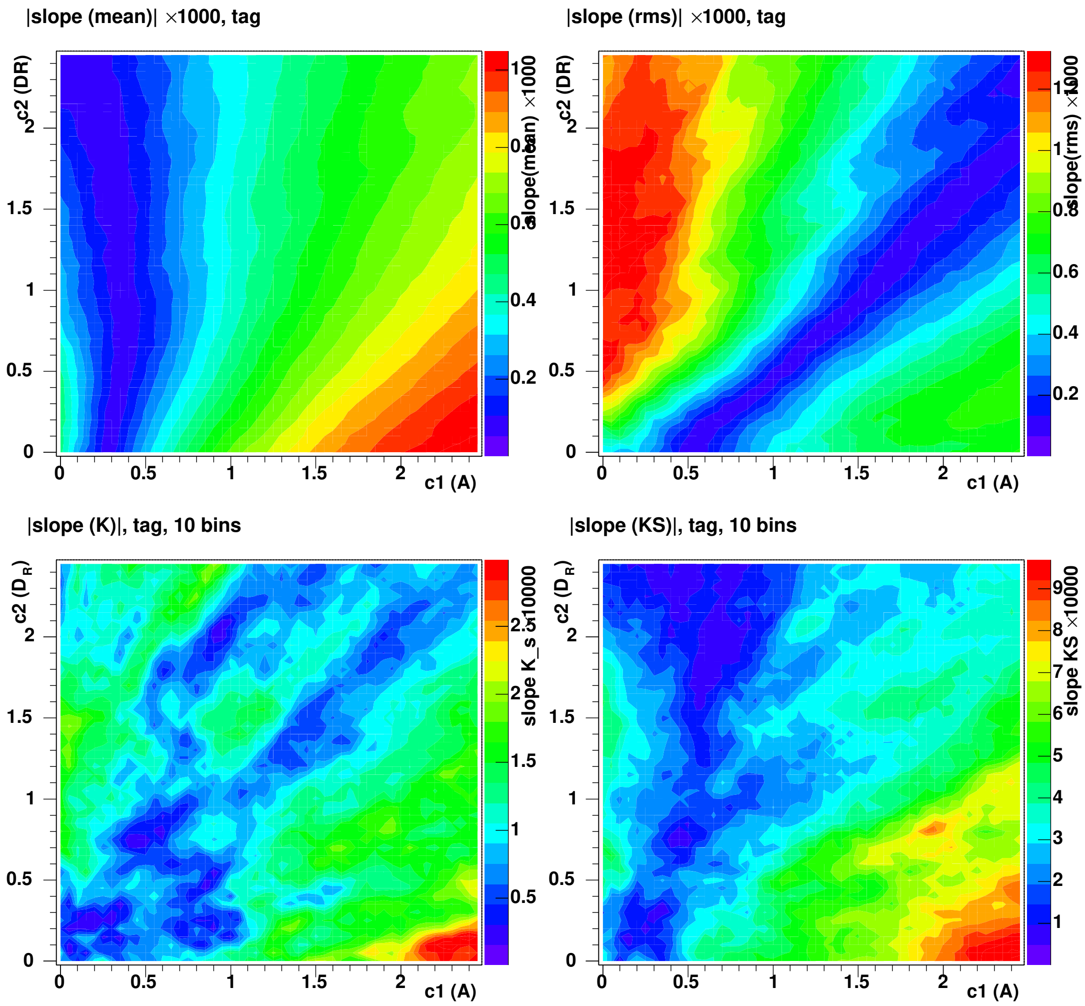

Figure 21: The top mass dependence of mean (top left) and r.m.s. (top right) of the V distribution and the K (bottom left) and KS (bottom right) distance of V with respect to the V distribution at 150 GeV. The colors encode the absolute slope values from line fits to the m_t evolution of these quantities.

The slopes for A and H_{TZ} have opposite sign which cancel exactly for a certain weight of A relative to H_{TZ}. Given no D_{R} contribution ($c_2 = 0$) this weight is $c_1 \sim 0.25$ because the absolute A slope is approximately 4 times larger than the H_{TZ} slope. Adding D_{R} with increasing weight c_2 while fixing $c_1 \sim 0.25$ doesn't change the situation much since the mass dependence of D_R alone is very small, which explains the vertical blue band in the upper left plot of Fig. 21.

A region with zero absolute slopes for the means is $0.2 < c_1 < 0.4$. A region with zero slopes for the r.m.s. is a diagonal starting from $c_2 = 0$, $c_1 \sim 0.5$, and is mostly disjoint with the result from the means. The distance measures are sensitive to both mean and r.m.s. simultaneously, and the plots for K and KS distance indeed reflect some kind of superposition of the two individual information. The distance measures suggest that a

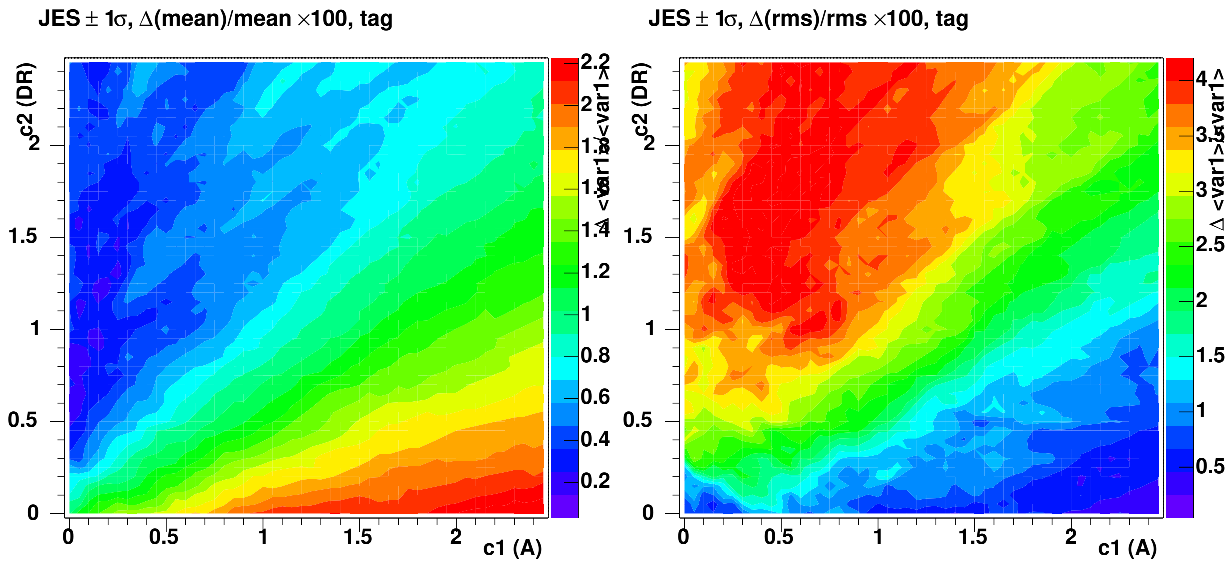

Figure 22: JES systematics is represented by the average absolute change of mean (left) and r.m.s. (right) of V when varying JES by ± 1 standard deviations of the total systematic uncertainties.

wider parameter space is usable as long as we avoid too high weights of A, but in order to be safe with our conclusions it is desirable to find a common good region in all four plots. A "good" region which also consider the interesting region of high S/B suppression is

$$0.2 < c_1 < 0.4, \ 0 < c_2 < 0.5$$

where the second requirement is needed to keep the mass dependence of r.m.s. moderately. For events with zero tags we draw similar conclusions.

Finally we show in Fig. 22 the JES systematics of the V distributions in terms of its mean and r.m.s. values. The colors encode the average absolute changes of respective values when varying JES by ± 1 standard deviations of the total systematic uncertainties (blue means low, red means large systematic effects). Optimal working points for mean and r.m.s. separately are

$$\text{mean}: c_1 \sim 0, \ c_2 \sim 0.6, \quad \text{r.m.s.}: c_1 \sim 2.3, \ c_2 \sim 0$$

but unfortunately the parameter sets are strongly disfavored by one another as well as by the mass sensitivity plots Fig. 21. On the other hand we observe that our region of interest pointed out previously has relatively moderate systematics for both mean and r.m.s. Similar results also apply for the zero tag case.

6.8 Conclusions

In view of using the hybrid variable $V(c_1, c_2, c_3 = 1)$ as a second variable for our multivariate template approach, we have narrowed the space of (c_1, c_2) down to a region of interest in which the S/B discrimination power of V can be optimized for $Wb\bar{b}$ and non-W QCD background such that both JES systematics and mass dependence of the V distribution are under control. It turned out that there is no common working point for V that performs an optimal suppression of both background types simultaneously, but our studies allows us to propose reasonable sets of V coefficients in order to explore the impact of different background types on the top mass uncertainties in a systematic way.

The expected statistical and systematic uncertainties of our top mass determination are discussed in detail in Sect. 7. For pseudo experiments involving the hybrid variable it is reasonable to fix the coefficient $c_1 = 0.3$ and to vary the coefficient c_2 from 0.0 to 0.5 in 0.05 steps. Thus we systematically scan the profile between the two optimal background suppression scenarios Fig. 20 such that the region of very low mass and moderate JES dependence of V is covered.

If we don't care of mass dependence, we might try larger c_2 values (D_{R} weights) up to $c_2 \sim 0.8$ where $Wb\bar{b}$ suppression is optimal. To maximize non-W QCD suppression one should try $c_2 = 0$ and large values for c_1 (A weights).

7 Top Mass Measurement Study

In this section we explain how we combine the signal and background information in order to extract a mass measurement. In addition, we compute the statistical and systematic errors we can expect our technique to yield through PEs. Finally, issues related to the bias and pull of the measurement will be discussed.

7.1 Overview of the Measurement Technique

After computing the signal likelihood for an event as described in Section 5, we can create a joint likelihood for all events by multiplying together the likelihood curves as discussed in Section 5.8. We obtain a two dimensional likelihood from which we extract a profile likelihood, which is the result of our analysis.

In Section 5.8 we have used this methodology to extract a mass value, an uncertainty and a pull value for Monte Carlo signal events. In practice, we have one set of top event candidates in our mass sample that includes signal as well as background (see Section 2. A final likelihood must take into account the presence of background.

7.2 Handling Background Events

In order to deal with background events, we first need to calculate the probability that a given event is background, $f_{\mathrm{bg}}(q_i)$, where q_i is the value of the discrimination variable for that event. We use the hybrid discrimination variable as discussed in Section 6. In order to calculate f_{bg} given a q_i, we create separate distributions of this variable from signal MC and background MC, and normalize the height of each distribution such that the total area of the signal distribution amounts to the expected signal fraction, and likewise for

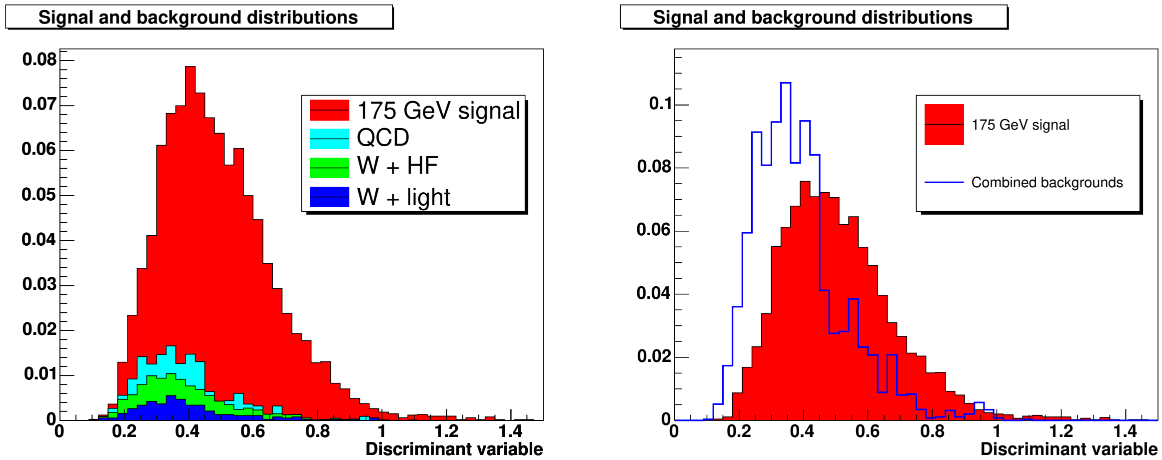

Figure 23: The signal and background probabilities obtained using our discriminant. The left figure shows the signal and types of background normalized to their expected fractions. The right figure shows the signal and background each normalized to 1 to show the discriminating power of our variable.

the background distribution. Then, for a given event's value of this variable, q, we find the values of the signal and background distributions, $S(q)$ and $B(q)$. We then simply set $f_{\rm bg}(q) = B(q)/(B(q) + S(q))$.

Obviously the discrimination variable needs to satisfy two criteria: it must have different distributions for signal and background, and it must be uncorrelated with the signal mass and JES parameter. Based on the studies in Section 6, we choose a hybrid variable with coefficients of $c_1 = 0.3, c_2 = 0.25$.

The signal and background distributions, calculated with these coefficients are shown in Figure 23. The signal MC is HERWIG (175 GeV), and the backgrounds are W+light jets, $W + b\bar{b}$, and non-W QCD.

We are not planning to use matrix elements for background because a significant fraction of the background comes from QCD and we do not know how to write a matrix element for it. Thus, the only information for a given event (whether that event is signal or background) is the likelihood calculated in the assumption that the event is a signal event, and the background fraction $f_{\rm bg}$.

In order to correct our signal likelihood for the effect of the background events, we need to consider what the effect of the background events on the total likelihood will be. We assume that, when treated as signal, the likelihood for background events will have a peak at a mass lower than the peak for the signal events, so that the likelihood contribution from the background events in the region of the signal peak is approximately linear (in log-likelihood space). Thus, the effect of adding background events is to *shift* the likelihood peak, but keep the *width* of the curve the same (to first approximation). Figure 24 shows a schematic of this process.

The question is, thus, how to remove the effect of this shift caused by the addition of background events. Suppose we know, perfectly, the number of background events $n_{\rm bg}$,

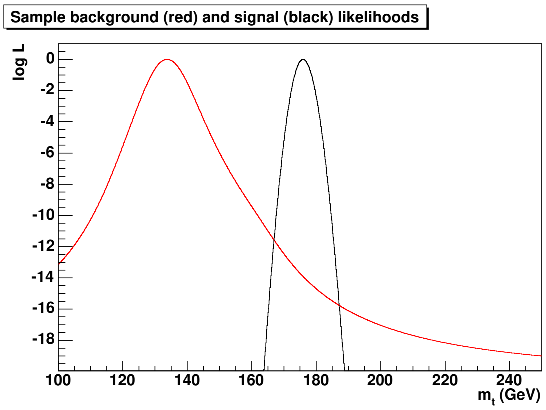

Figure 24: The expected effect of adding background events to our signal likelihood is to shift the peak of our likelihood curve to the left.

and that all background events have the same likelihood, so that we can use the average likelihood for background events, $\overline{L(\text{background})}$ (still computed under the assumption that they are signal.) In this case, computing the correct likelihood is simple. We merely subtract out the contribution from the background events:

$$L_{\text{tot}} = \sum_i (\log L_i) - n_{\text{bg}} \log \overline{L(\text{background})}$$

Since our background fraction f_{bg} is, by definition, derived such that $\sum_i f_{\text{bg}}(q_i) = n_{\text{bg}}$, we can write the same as:

$$L_{\text{tot}} = \sum_i (\log L_i - f_{\text{bg}}(q_i) \log \overline{L(\text{background})})$$

Of course, when we use this on real events, this formula will not work perfectly, because the likelihood of the actual background events that appear in a pseudoexperiment will have a scatter around the average background likelihood $\overline{L(\text{background})}$, and this will introduce another source of error, which we call the "background error" σ_{bg}.

In order to reduce the error, we can modify the likelihood curves as they are added into the total likelihood. (This modification is not physical – it is purely a mathematical modification of the curves to get the best peak and width.) Suppose the background curves exhibit a large degree of variation around the average background likelihood $\overline{L(\text{background})}$. In this case, we can reduce the error caused by this variation by "flattening" the curves. We perform this "flattening" by taking an average with the uniform distribution U. We want the "flattening" to be applied, ideally, only to the background events, so we have the "flattening" be in proportion to the background fraction f_{bg}. Thus:

$$\log L(m_t, \text{JES}) \Rightarrow \log[L(m_t, \text{JES})(1 - f_{\text{bg}}(q)) + f_{\text{bg}}(q) U]$$

However, we don't, a priori, know the correct amount of "flattening" to apply. As discussed above, "flattening" the background curves will reduce their contribution to the overall likelihood and hence reduce $\sigma_{\rm bg}$. However, inevitably, some "flattening" will be applied to the signal events as well, causing $\sigma_{\rm sig}$ to increase. Thus, we want to flatten by the proper amount to get a balance between decreased $\sigma_{\rm bg}$ and increased $\sigma_{\rm sig}$. In order to do this, we introduce an additional factor κ:

$$\log L(m_t, \mathrm{JES}) \Rightarrow \log[L(m_t, \mathrm{JES})(1 - f_{\rm bg}(q)\kappa) + f_{\rm bg}(q)\kappa U]$$

This κ factor can be empirically optimized to produce the smallest combined error. We can also consider the possibility that κ can be optimized as a function of m_t.

Thus, our final likelihood will look like this:

$$L_{mod}(m_t, \mathrm{JES}) = \sum_{\text{events}} [\log\{L(m_t, \mathrm{JES}\,|\mathrm{signal})(1 - f_{bg}(q)\kappa(m_t)) + f_{bg}(q)\kappa(m_t)U\} -$$

$$f_{bg}(q)\log\{\overline{L(m_t, \mathrm{JES}\,|\mathrm{background})}(1 - f_{bg}(q)\kappa(m_t)) + f_{bg}(q)\kappa(m_t)U\}]$$

Currently we use $\kappa(m) \equiv 1$.

MORE INFORMATION ON MEASURING BACKGROUND ERROR WILL GO HERE

7.3 Background plus Signal MC Results

We obtain likelihood distributions using the integration procedure outlined in Section 5. Some sample 2-D likelihoods for a variety of events are shown in Fig. **??**.

After the integration of signal and background samples we check the results with pseudo-experiments. Each PE is constructed from four sources of event: signal, W+$b\bar{b}$ and W+light MC, as well as QCD data events. Each source contributes a Poisson-averaged number of events, the average value computed from the expected fraction of the source in the sample (see Section 2 for the expected background fractions) with a total of 116 events.

Figure **??** shows results from 4000 pseudo-experiments of 116 events each at a top mass of 175 GeV, with the expected signal fraction, obtained using the profile likelihood method. Both the mass and the pull distributions are shown.

Plots of reconstructed mass *vs.* true mass, as well as bias and pulls *vs.* true mass are shown in Fig. 25

Table 4: PE results on signal + background events

Input used	Bias	slope	σ at 175 GeV	pull
Sig+background	-0.80 $\pm$ 0.12	0.975$\pm$0.012	2.93$\pm$0.05	1.30$\pm$0.01

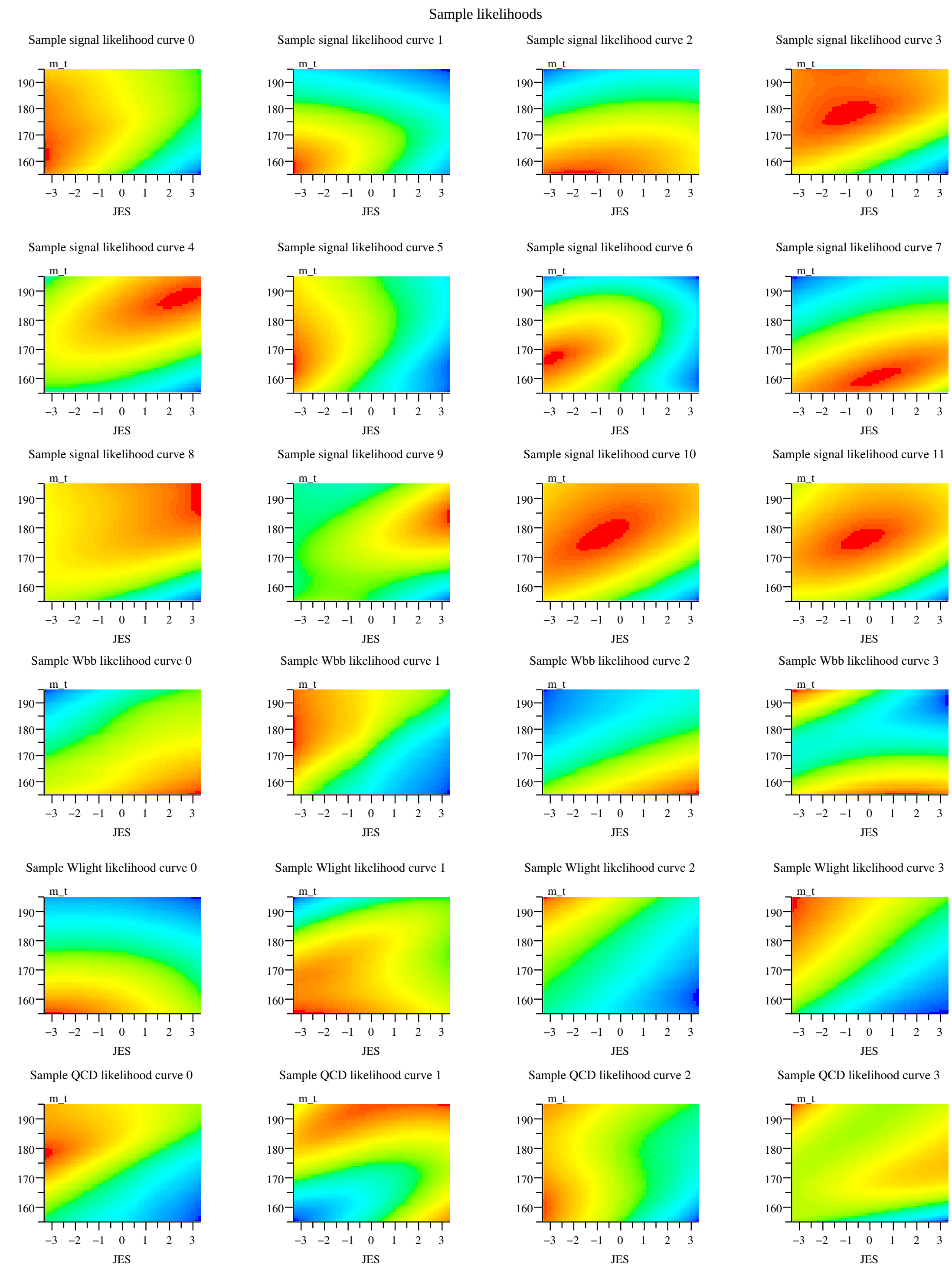

Figure 25: Examples of likelihood distributions obtained for signal events (rows 1-3), W+$b\bar{b}$ (row 4), W+light quarks (row 5), QCD background (row 6).

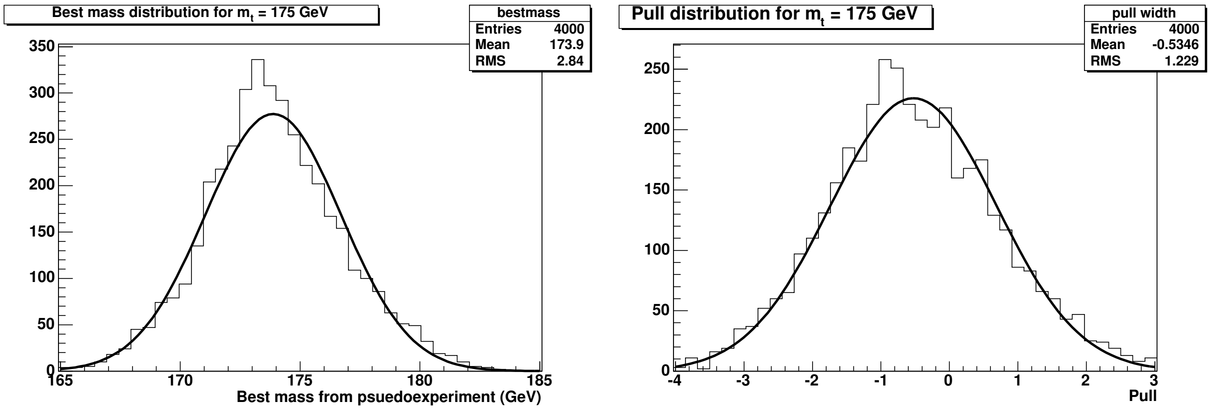

Figure 26: Left: Results of pseudo-experiments using the profile likelihood at a mass of 175 GeV. Left: top mass distribution. Right: pull distribution.

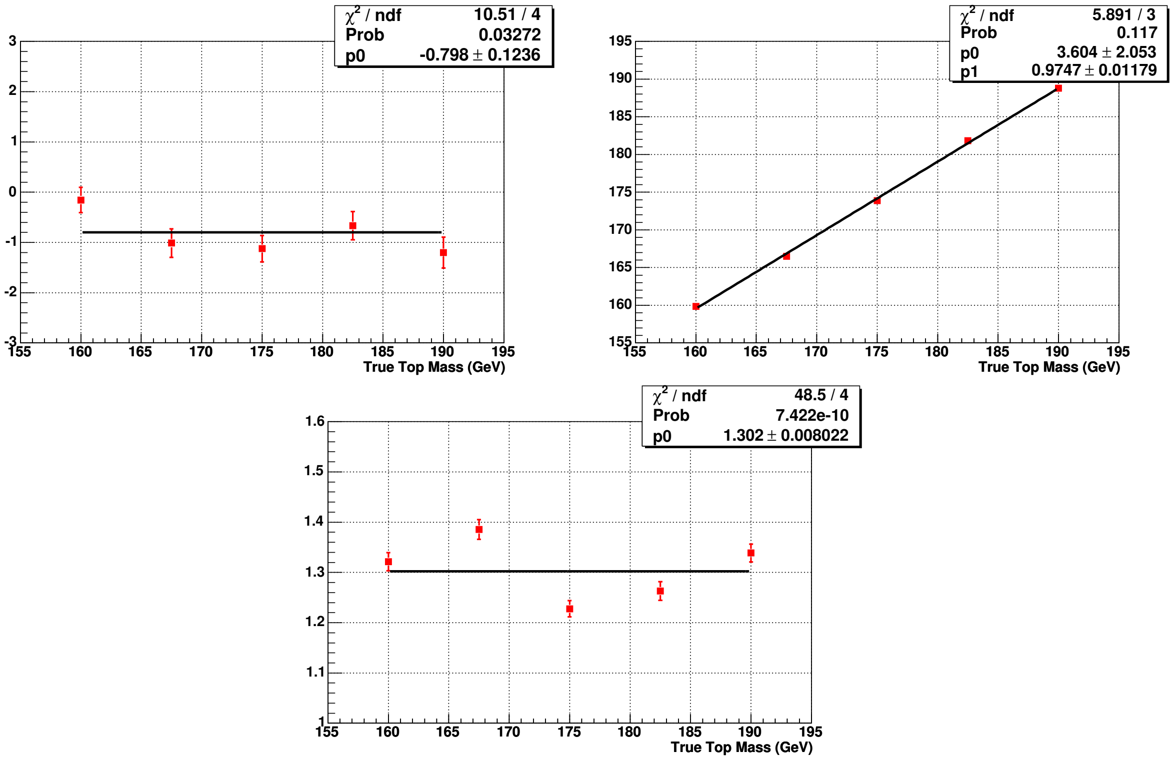

Figure 27: Top left: Bias vs. true mass. Top right: Reconstructed mass vs. true mass. Bottom: Pull width vs. true mass.

7.4 Expected Errors

In order to study the statistical errors we should expect from our analysis, we ran 4000 PEs each for a signal mass range of 160 to 200 GeV in 7.5 GeV steps.

Fig. ?? shows the expected statistical uncertainty from MC events using the 2-D profile likelihood method.

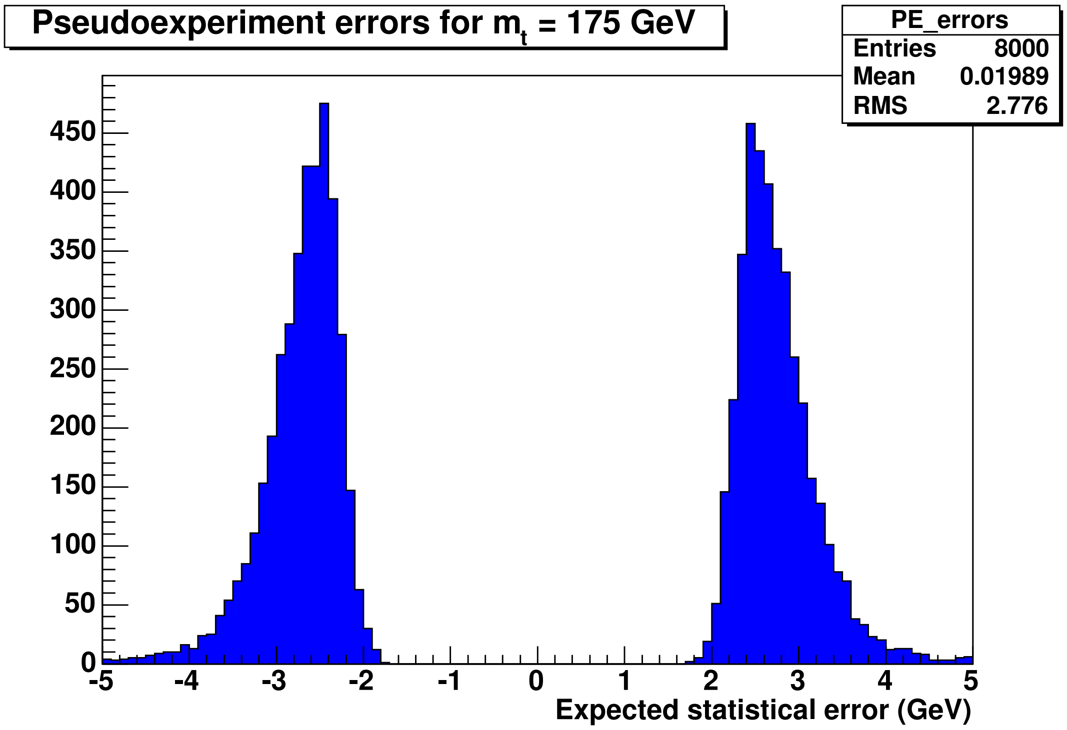

Figure 28: Expected statistical uncertainty on the 2-D profile likelihood method. The errors have been scaled by the pull width.

7.5 Checks on the Blind Mass Samples

7.6 Checks on the Blind JES Scale Samples

8 Results from data

We discuss here the results.

9 Systematic Uncertainties

We discuss here the systematic uncertainties.

1. Residual Jet Energy

2. b-jet Energy
3. parton distribution functions
4. ISR/FSR
5. gluon fraction

10 Conclusions

We have measured a value for the mass of

M = xxx $\pm$ yyy(stat) $\pm$ zzz(syst)

We expect to improve the measurement considerably for summer conferences, adding many features that are already being investigated, but are not fully checked at this point.

11 Acknowledgments

We are grateful to Stephen Parke for helpful discussions about the Kleiss-Stirling matrix element code.

A Appendix A: Gen6 Validation

This section discusses validation of the Gen6 data.

Fig. 26 and Fig. 27 show the distribution of various quantities for the data compared to the Monte Carlo for events passing all of our selection cuts, but without any loose jet cut. The three colors of dots represent the 0d, 0h, and 0i datasets. The solid line is the sum total of 175 GeV signal and background in the appropriate background fractions.

Fig. 26 shows the comparison of jet E_T for the first five jets. The agreement is good for the four leading jets. For the fifth jet the data is somewhat higher than the MC. Since we do not use the fifth jet in our events this is not an area of major concern.

Fig. 27 shows the MET, total number of jets, E_T for b-tagged jets, and lepton P_T. Again, the agreement is good for most quantities. The total number of jets is somewhat higher in the data in general, but again, since we are only interested in the 4 leading jets, this is not a major issue.

Fig. 28 and Fig. 29 show the same information for events on which the 0 loose jet cut has been applied. The agreement is similar.

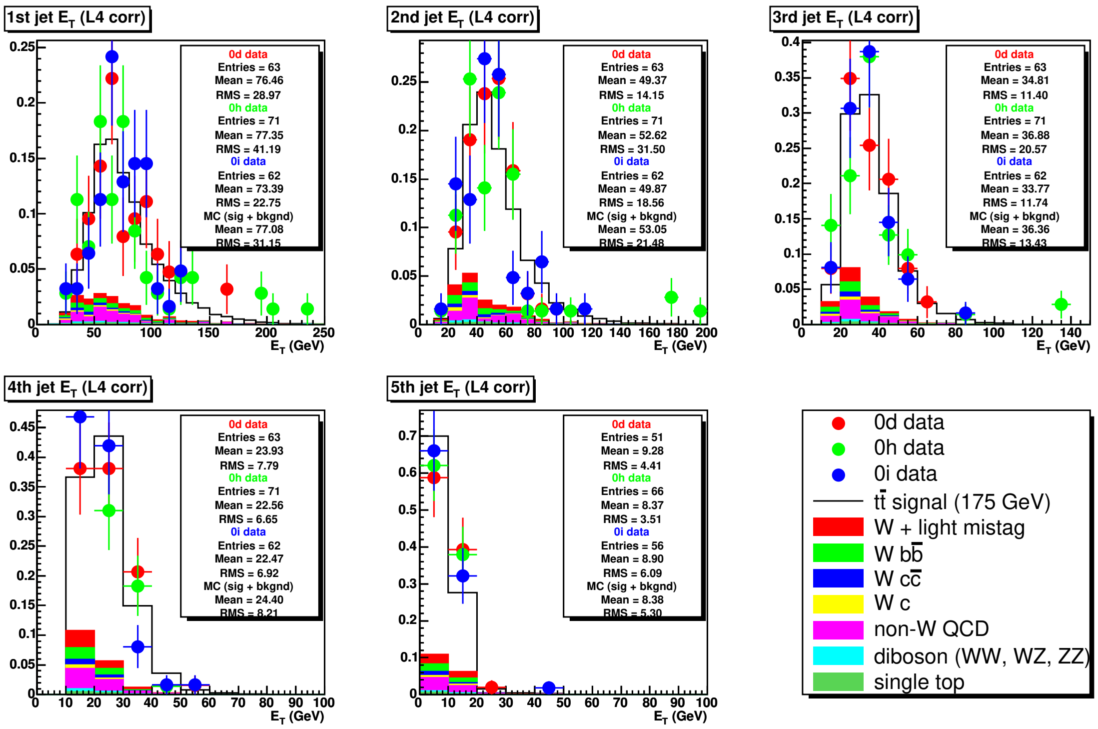

Figure 29: Comparison of data and MC for selected events with no cut on loose jets.

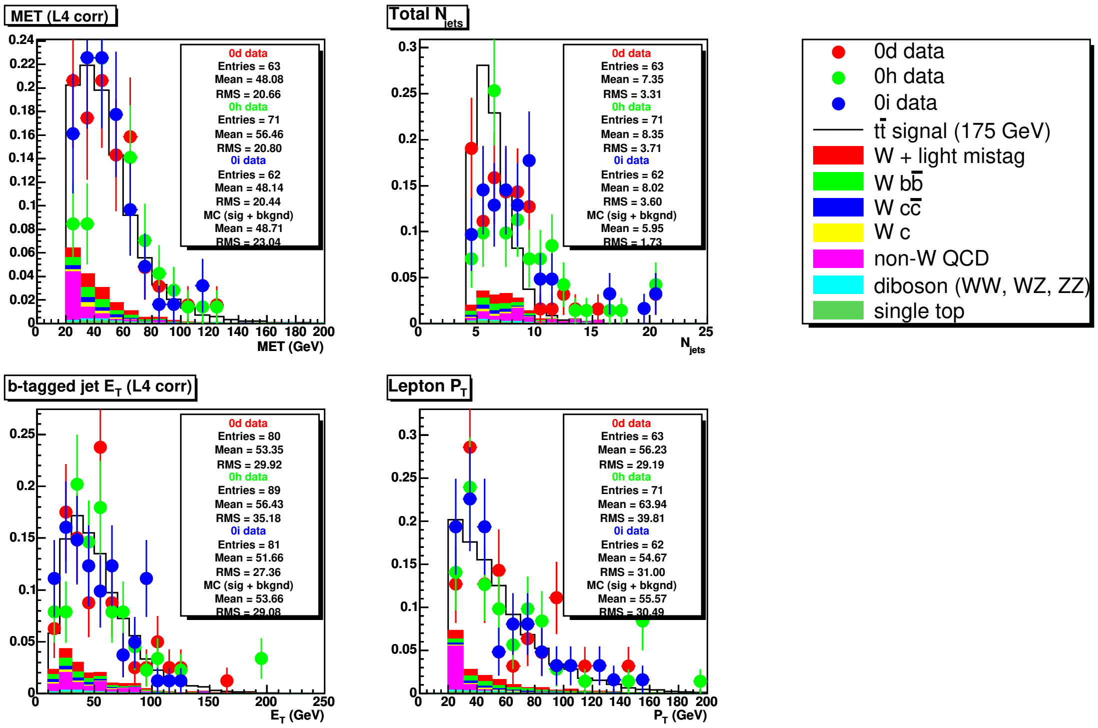

Figure 30: Comparison of data and MC for selected events with no cut on loose jets.

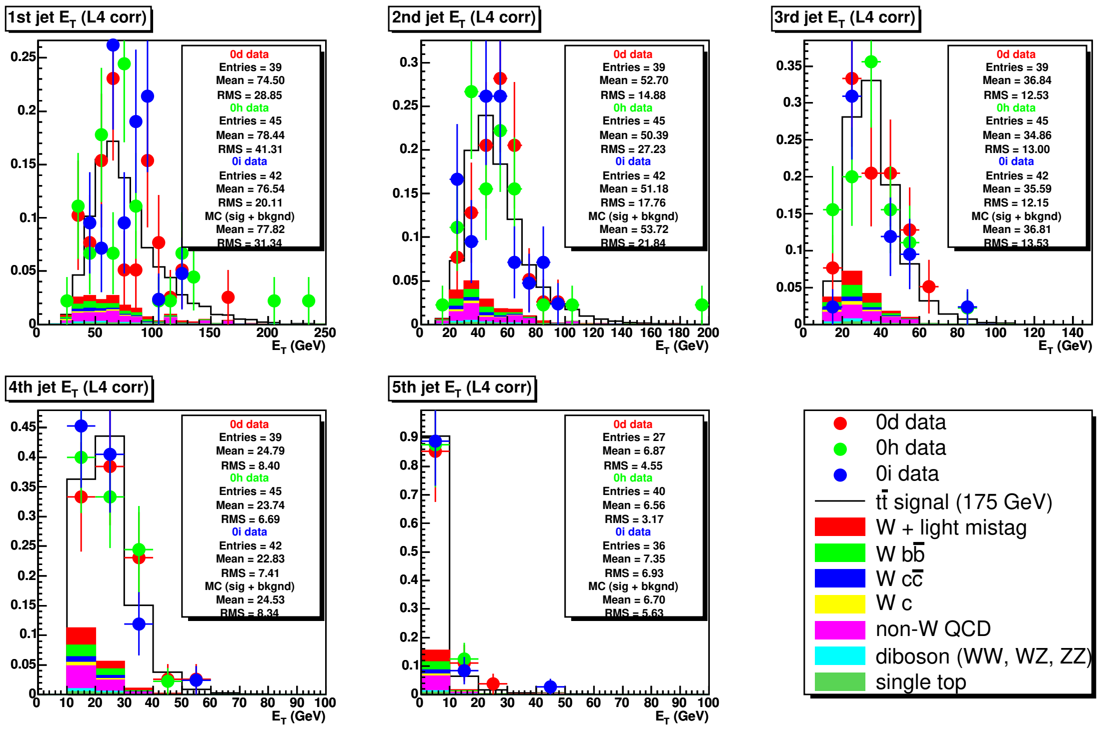

Figure 31: Comparison of data and MC for selected events with 0 loose jets required.

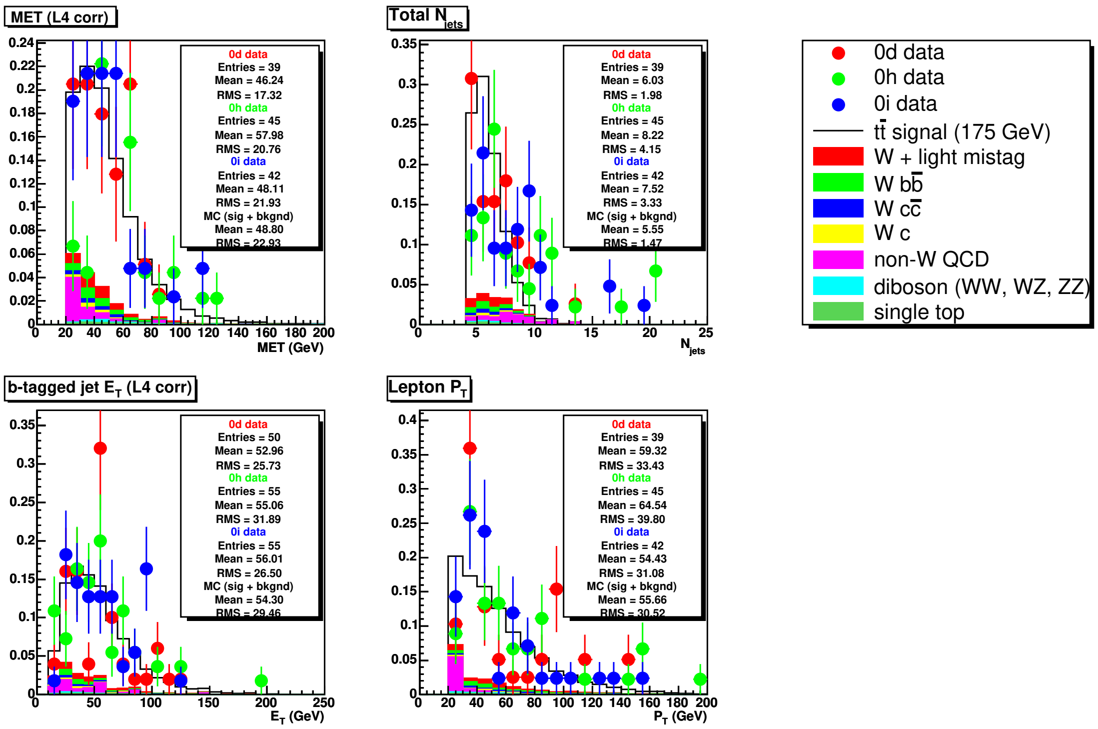

Figure 32: Comparison of data and MC for selected events with 0 loose jets required.

B Appendix B: Johnson's Functions

The Johnson's density functions are as follows:

$$S_U: \quad p(x) = \frac{\delta}{\lambda\sqrt{2\pi\left(1+\left(\frac{x-\xi}{\lambda}\right)^2\right)}}\, e^{-\frac{1}{2}\left(\gamma+\delta\sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)\right)^2}$$

$$S_B: \quad p(x) = \frac{\delta}{\lambda\sqrt{2\pi}\left(\frac{x-\xi}{\lambda}\right)\left(1-\frac{x-\xi}{\lambda}\right)}\, e^{-\frac{1}{2}\left(\gamma+\delta\log\left(\frac{x-\xi}{\xi+\lambda-x}\right)\right)^2}$$

They are related to the normal distribution by simple variable transformations. Variable z distributed according to $N(0,1)$ can be obtained from Johnson's variates x by transformation $z = \gamma + \delta f\left(\frac{x-\xi}{\lambda}\right)$:

$$\begin{array}{rlll} f(y) = & \sinh^{-1}(y) & & S_U \text{ system} \\ f(y) = & \log\left(\frac{y}{1-y}\right) & (0 < y < 1) & S_B \text{ system} \end{array}$$

Both systems become arbitrarily close to the lognormal distribution $z = \gamma + \delta\log(x-\xi)$ in the limit $\gamma \to \infty$ and to the normal distribution in the limit $\delta \to \infty$ (ξ and λ need to be adjusted accordingly).

Together with their limiting cases, Johnson's S_U and S_B distributions attain all possible values of skewness and kurtosis. The parameters of S_U and S_B have no relation to each other. (Crossing the lognormal boundary requires a complete change of parameter values, and this is very inconvenient for practical data fitting purposes.)

C Appendix C: On the $t\bar{t}$ Decay Phase Space

C.1 Introduction

The structure of the $t\bar{t}$ decay matrix element is determined predominantly by t and W propagators. Therefore, to improve efficiency, the integration over the $t\bar{t}$ phase space is performed using a set of variables which includes top and W masses. Here we examine the relevant variable transformation and derive the correct Jacobian.

C.2 Phase Space Formulae

The phase space for the 6 final state particles in the $q_1q_2 \to t\bar{t} \to 4\,\text{jets} + e(\text{or } \mu)\nu$ process can be written as

$$d\Phi_6(q_1+q_2;\, p_1,p_2,p_3,p_4,p_e,p_\nu) = \delta^4(q_1+q_2-\sum_{i=1}^{6}p_i)\prod_{i=1}^{6}\frac{d^3\mathbf{p}_i}{(2\pi)^3 2E_i}$$

To reflect good angular resolution for jets, it is convenient to express the $\frac{d^3\mathbf{p}_i}{2E_i}$ term for each jet as $\frac{\rho^2 d\rho_i d\Omega_i}{2E_i}$ where $\rho_i = |\mathbf{p}_i|$. DØ assumes that the detector jet angular resolution function is $\delta^2(\Omega - \Omega_i)$. Although it would be more logical to assume angular resolution function in the form $\delta^2(\phi - \phi_i, \eta - \eta_i)$, the term $\frac{\partial \eta}{\partial \cos\theta}$ for massless jets is just $\sin^{-2}\theta$, and does not depend on the jet momentum. Therefore, all such terms do not affect the shape of the top mass likelihood (they are just factors independent from top mass), and they are not important as long as the likelihood curves are renormalized and jet masses are small in comparison with their momenta. After integrating over the jet angles with angular resolution functions, the phase space becomes

$$d\Phi_6 = \delta^4(q_1 + q_2 - \sum_{i=1}^{6} p_i) \prod_{i=1}^{4} \frac{\rho_i^2 d\rho_i}{(2\pi)^3 2E_i} \frac{d^3\mathbf{p}_e}{(2\pi)^3 2E_e} \frac{d^3\mathbf{p}_\nu}{(2\pi)^3 2E_\nu}$$

It is also reasonable to assume that the lepton momentum is well measured, and that the integration over the lepton phase space with lepton resolution function results in a likelihood term which is independent from the top mass. Then the "interesting" part of the phase space is just

$$d\Phi_6 = \delta^4(q_1 + q_2 - \sum_{i=1}^{6} p_i) \prod_{i=1}^{4} \frac{\rho_i^2 d\rho_i}{(2\pi)^3 2E_i} \frac{d^3\mathbf{p}_\nu}{(2\pi)^3 2E_\nu} \tag{33}$$

C.3 Variable Transformation

Assume that initial partons q_1 and q_2 are massless and neglect their transverse momenta, and assume that q_1 moves in the positive z direction, so that $E_1 = q_{1z}$ and $E_2 = -q_{2z}$. Then the differential phase space in Equation 33 is a function of 9 variables: q_{1z}, q_{2z}, ρ_1, ρ_2, ρ_3, ρ_4, $p_{\nu x}$, $p_{\nu y}$, and $p_{\nu z}$. Our goal will be to transform the variables so that the integration is performed over P_E, P_x, P_y, P_z, β, M_1^2, m_1^2, M_2^2, and m_2^2, where

$$\begin{aligned}
P_E &= E_1 + E_2 - \sum_{i=1}^{6} E_i \\
P_x &= -\sum_{i=1}^{6} p_{ix} \\
P_y &= -\sum_{i=1}^{6} p_{iy} \\
P_z &= q_{1z} + q_{2z} - \sum_{i=1}^{6} p_{iz} \\
\beta &= \log(\rho_1/\rho_2) \\
m_1^2 &= (p_1 + p_2)^2 \\
M_1^2 &= (p_1 + p_2 + p_3)^2 \\
M_2^2 &= (p_4 + p_e + p_\nu)^2 \\
m_2^2 &= (p_e + p_\nu)^2
\end{aligned}$$

m_1 is the invariant mass of the particles which form the hadronically decaying W, M_1 is the invariant mass of the top decay products on the hadronic side, while m_2, M_2 are the corresponding quantities for the leptonic side. These definitions are the same as in [7]. Integration over P_E, P_x, P_y, and P_z will manifestly take care of the delta function in Equation 33, and integration over mass variables can be performed efficiently using points equidistant in the cumulative probability distributions of the corresponding propagators.

The complete Jacobian matrix for performing the change of variables can be written as

$$\begin{aligned} J &= \frac{\partial(P_E, P_z, \beta, m_1^2, M_1^2, M_2^2, m_2^2, P_x, P_y)}{\partial(q_{1z}, q_{2z}, \rho_1, \rho_2, \rho_3, \rho_4, p_{\nu x}, p_{\nu y}, p_{\nu z})} \\ &= \begin{pmatrix} 1 & -1 & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} & a_{19} \\ 1 & 1 & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} & a_{29} \\ 0 & 0 & a_{33} & a_{34} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{43} & a_{44} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{53} & a_{54} & a_{55} & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} & a_{69} \\ 0 & 0 & a_{73} & a_{74} & a_{75} & 0 & a_{77} & a_{78} & a_{79} \\ 0 & 0 & a_{83} & a_{84} & a_{85} & a_{86} & -1 & 0 & 0 \\ 0 & 0 & a_{93} & a_{94} & a_{95} & a_{96} & 0 & -1 & 0 \end{pmatrix} \end{aligned}$$

Fortunately, we do not need to evaluate all a_{ij} partial derivatives in order to find the determinant of this Jacobian matrix. It is easy to see that the determinant can be factorized into the product of the following terms, representing the initial state, the hadronic top, and the leptonic top:

$$|J| = |J_1|\,|J_2|\,|J_3|$$

where

$$\begin{aligned} |J_1| &= \left|\frac{\partial(P_E,P_z)}{\partial(q_{1z},q_{2z})}\right| &&= 2 \\ |J_2| &= \left|\frac{\partial(\beta,m_1^2,M_1^2)}{\partial(\rho_1,\rho_2,\rho_3)}\right| &&= (a_{33}a_{44} - a_{34}a_{43})a_{55} \\ |J_3| &= \left|\frac{\partial(M_2^2,m_2^2,P_x,P_y)}{\partial(\rho_4,p_{\nu x},p_{\nu y},p_{\nu z})}\right| &&= a_{66}a_{79} + (a_{67}a_{79} - a_{69}a_{77})a_{86} + (a_{68}a_{79} - a_{69}a_{78})a_{96} \end{aligned}$$

The partial derivatives actually needed to evaluate the determinants are

$$
\begin{aligned}
a_{33} &= \tfrac{\partial\beta}{\partial\rho_1} &&= 1/\rho_1 \\
a_{34} &= \tfrac{\partial\beta}{\partial\rho_2} &&= -1/\rho_2 \\
a_{43} &= \tfrac{\partial m_1^2}{\partial\rho_1} &&= 2\left(E_{W1}\rho_1/E_1 - \mathbf{p}_1\cdot\mathbf{p}_{W1}/\rho_1\right) \\
a_{44} &= \tfrac{\partial m_1^2}{\partial\rho_2} &&= 2\left(E_{W1}\rho_2/E_2 - \mathbf{p}_2\cdot\mathbf{p}_{W1}/\rho_2\right) \\
a_{55} &= \tfrac{\partial M_1^2}{\partial\rho_3} &&= 2\left(E_{t1}\rho_3/E_3 - \mathbf{p}_3\cdot\mathbf{p}_{t1}/\rho_3\right) \\
a_{66} &= \tfrac{\partial M_2^2}{\partial\rho_4} &&= 2\left(E_{t2}\rho_4/E_4 - \mathbf{p}_4\cdot\mathbf{p}_{t2}/\rho_4\right) \\
(a_{67}, a_{68}, a_{69}) &= \tfrac{\partial M_2^2}{\partial\mathbf{p}_\nu} &&= 2\left(E_{t2}/E_\nu\,\mathbf{p}_\nu - \mathbf{p}_{t2}\right) \\
(a_{77}, a_{78}, a_{79}) &= \tfrac{\partial m_2^2}{\partial\mathbf{p}_\nu} &&= 2\left(E_e/E_\nu\,\mathbf{p}_\nu - \mathbf{p}_e\right) \\
a_{86} &= \tfrac{\partial P_x}{\partial\rho_4} &&= -\sin\theta_4\cos\phi_4 \\
a_{96} &= \tfrac{\partial P_y}{\partial\rho_4} &&= -\sin\theta_4\sin\phi_4
\end{aligned}
$$

D References

References

[1] I. Volobouev *et al.*, "Top Mass Measurement in the Lepton + Jets Channel Using a Multivariate Template Method", CDF-6970.

[2] A. Caner *et al.*, "Jet Corrections for Top Mass Fitting," CDF-2469.

[3] T. Affolder *et al.*, "Measurement of the Top Quark Mass with the Collider Detector at Fermilab," Phys. Rev. D, **63** (2001) 032003.

[4] A. Bhatti *et al.*, "Determination of the Jet Energy Scale at the Collider Detector at Fermilab," Nucl. Instr. Meth. A. **566**,375-412 (2006). Also *hep-ex/0510047*.

[5] K. Kondo, J. Phys. Soc. Jpn. **57** (1988) 4126.

[6] R. H. Dalitz and G. R. Goldstein, Phys. Rev. **D45**, 1531 (1992).

[7] J. Estrada, Ph.D. thesis, University of Rochester (see `http://www-d0.fnal.gov/results/publications_talks/thesis/estrada/thesis.ps`)

[8] B. Abbott *et al.* (D0 Collaboration), Nature **429**, 638 (2004).

[9] K. Yorita and K. Kondo, "Top quark mass measurement using the DLM method in the lepton + jets channel with SECVTX," CDF-6829 (for the 168 pb^{-1} sample) and CDF-7642, updated fyor the 318 pb^{-1} sample.

[10] A. Abulencia *et al.*, "Top quark mass measurement using the DLM method in the lepton + jets channel with displaced vertices b-tagging", Phys. Rev. **D73**, 092002 (2006)

[11] mtm) Erik et al CDF note + PRD MTM

[12] F. Canelli and B. Mohr, "Measurement of the Top Quark Mass using the Matrix Element Analysis Technique (MEAT)," CDF-8038.

[13] A. Gibson *et al.*, "A Measurement of the Top Quark Mass with a Matrix Element Method in the Lepton+Jets Channel," CDF-7720.

[14] B. Jayatilaka *et al.*, "Measurement of the Top Quark Mass in the Dilepton Channel in $\int Ldt = 750\,\mathrm{p}b^{-1}$," CDF-8058.

[15] H. Bachacou, J. Nielsen, W-M Yao, "Optimized Measurement of the $t\bar{t}$ Cross Section in the SECVTX Tagged Sample," CDF-6902.

[16] D. Acosta *et al.*, "Measurement of the $t\bar{t}$ Cross Section Measurement at $\sqrt{s}$=1.96 TeV, using Lepton + Jets events with Secondary Vertex b-tagging," Phys. Rev. **D71**, 052003 (2005).

[17] See the following books: *Systems of frequency Curves* by W.P. Elderton and *Statistical Models in Engineering* by G.J. Hahn and S.S. Shapiro.

[18] Applied Statistics **25**, 180 (1976).

[19] G. Mahlon and S. Parke, "Angular correlations in top quark pair production and decay at hadron colliders," Phys.Rev. **D 53**, 4886 (1996).

[20] G. Mahlon and S. Parke, "Maximizing spin correlations in top quark pair production at the Tevatron," Phys.Lett. **B 411**, 173 (1997).

[21] R. Kleiss and W.J.Stirling, "Top quark production at hadron colliders: some useful formulae," Z.Phys. **C 40**, 419 (1988).

[22] H.L.Lai *et al.*, *Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions*,Eur.Phys.J. **C 12**, 375 (2000).

[23] S. Pashapour and P.K. Sinervo, "$gg \to t\bar{t}$ vs. $q\bar{q} \to t\bar{t}$: A Generator-Level Study," CDF-6218.

[24] I. Sobol and Y.L. Levitan, "The Production of Points Uniformly Distributed in a Multidimensional Cube," Preprint IPM Akad. Nauk SSSR, Number 40, Moscow 1976;
`http://www.scs.fsu.edu/~burkardt/cpp_src/sobol/sobol.html`