Available on CMS information server **CMS IN 2009/000**

CMS Internal Note

The content of this note is intended for CMS internal use and distribution only

14 January 2009

Correcting Missing Transverse Energy Using Tracks

F. Golf, D. Evans, J. Mülmenstädt, S. Padhi, F. Würthwein, A. Yagil

University of California, San Diego

J. Ribnik, P. Kalavase, D. Kovalskyi, V. Krutelyov, C. Campagnari

University of California, Santa Barbara

I. Bloch, O. Gutsche, I. Fisk, K. Burkett, L. Bauerdick

Fermi National Accelerator Laboratory, Batavia, Illinois

Abstract

Missing transverse energy (MET) is an important observable used to establish detector performance, observe basic SM signatures and search for new physics. An accurate determination of this observable is made difficult by the non-linearity of the CMS calorimeter and the strong encompassing magnetic field. Here an algorithm is developed that improves the measurement of MET in CMS by correcting for charged tracks while also improving the core resolution in events with real MET.

Preliminary version

1 Introduction

Missing Transverse Energy (MET) is an important observable used to establish proper detector performance, observe basic standard model (SM) signatures and search for new physics. A true measurement of $\not\!\!E_T$ requires understanding the physics as well as proper detector calibration. Of the basic physics objects, it is the most susceptible to detector calibration and noise. Two main properties characterize the $\not\!\!E_T$ performance of a detector: resolution and the non-Gaussian tail. The MET performance in CMS suffers because of

- the large material budget of the tracking system,
- the non-linear response of the calorimeter,
- the strong (3.8 T) magnetic field.

The last two points are the primary sources of MET mis-mesaurement. The energy a particle deposits is generally under-measured due to the non-linear response of the calorimeter. Mis-measurement of the magnitude of the deposition is not the only source of the performance deficit. The strong magnetic field in CMS bends the trajectories of charged particles, causing energy deposited in the calorimeter to be displaced in ϕ relative to the initial direction of the particle.

The issues of magnitude and direction have to be addressed in order to obtain an accurate determination of $\not\!\!E_T$. We have developed a method to improve performance in measuring MET by taking advantage of the CMS tracker. At the scale of interest ($\sim$1 GeV), the tracker has excellent resolution compared to the calorimeter. The proposal is to correct the measured $\not\!\!E_T$ by replacing, for all well reconstructed tracks, the average (or expected) energy deposition in the calorimeter by the measured momentum in the tracker.

2 The problem: Fake MET in Z $\rightarrow$ *ll* events

Any search for new physics will begin with the study of SM processes necessary to calibrate the detector and develop experimental techniques. These processes present significant backgrounds that distort signatures of new physics. Drell-Yan (DY) processes are among the most abundant. Rejection of these events should be possible by cutting on MET. However, mis-measurement of $\not\!\!E_T$ results in a large number of DY events with large fake MET. This mis-mesaurement poses significant difficulties for both BSM searches as well as the study of SM processes such as $t\bar{t}$ events.[1, 2, 3].

The n-jet distribution for di-leptons plus MET illustrates some of the processes we have been investigating and our current understanding of their relative contributions. Fig. 1 shows the n-jet distribution for the di-muon final state after requiring:

- lepton $p_T > 20$ GeV,
- MET > 30 GeV,
- lepton isolation $p_T/\left(p_T + S\right) > 0.92$,
- count jets with $|\eta| < 3$ and $E_T > 30$ GeV (corrected),

where S is the a quantity that characterizes the amount of hadronic activity in a cone of $\Delta R = 0.3$. Here the MET is not corrected for the jet energy scale (JES). However, applying these type-I corrections does not change the picture substantially. The sizable contribution of $Z \rightarrow \mu\mu$ is an indication of the problem CMS faces in measuring MET. At this level, it will be a serious background to any search for new physics whose signature involves $\not\!\!E_T$.

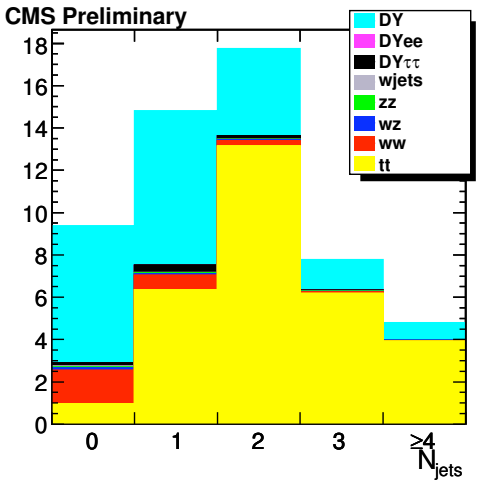

Figure 1: Relative contributions to the n-jet distribution for $t\bar{t}$ after cutting on lepton p_T and MET. (from Figure 11, [1]).

3 Releases and Data Sets

This work was done in the context of CMSSW release 1_6_12. The following samples were used.

1. Single π
   - /RelValSinglePiPlus0To100/CMSSW_1_6_7-RelVal-1202415807/GEN-SIM-DIGI-RECO
2. DY $\rightarrow ll$, $m_{ll} > 40\ GeV$
   - /DrellYan_ll_40/CMSSW_1_6_7-CSA07-1201885455/RECO
3. $W \rightarrow \mu\nu$
   - /Wmunu/CMSSW_1_6_7-CSA07-1192835438/RECO
4. $W \rightarrow e\nu$
   - /Wenu/CMSSW_1_6_7-CSA07-1197047869/RECO
5. QCD
   - /QCD_Pt_20_30/CMSSW_1_5_2-CSA07-2162/GEN-SIM-DIGI-RECO
   - /QCD_Pt_50_80/CMSSW_1_5_2-CSA07-2049/GEN-SIM-DIGI-RECO
   - /QCD_Pt_120_170/CMSSW_1_5_2-CSA07-2171/GEN-SIM-DIGI-RECO
   - /QCD_Pt_230_300/CMSSW_1_5_2-CSA07-2050/GEN-SIM-DIGI-RECO

4 Procedure

The goal is to construct a procedure to correct $\not\!\!E_T$ on a track-by-track basis. Beginning with a single pion sample, the track and calorimeter information is extracted and a set of filters is used to select a subset of well-reconstructed pions. From this subset, a lookup table of detector response in η, p_T is built.

The following section will discuss the implementation methodology that uses the response function to remove expected depositions in the calorimeter and replace them with values determined from the tracker. Special treatment is given to leptons. The performance evaluation of the track-corrected MET (tcMET) is discussed next. The effects of this correction on Drell-Yan samples with fake $\not\!\!E_T$ is compared to events with real $\not\!\!E_T$, such as W production.

A few comments on the effect of the correction on $\not\!\!E_T$ resolution will follow. The final section will remark briefly on hybrid implementations of track-based $\not\!\!E_T$ corrections. Our work tries to address large tails in the Drell-Yan MET distribution, particularly in the 0-jet bin which contains unclustered energy. Although alternate approaches such as type-I corrections do not offer the general applicability or performance of tcMET, in certain scenarios they may complement track corrected MET. Future work will study situations where additional gains may be had, such as the use of a non zero suppressed (non-ZSP) response function in high multiplicity environments [4].

To summarize, the outline of what follows is:

- Using single pions, extract the calorimeter response as a function of $\eta, p_T \rightarrow$ response function (RF)
- Identify the set of tracks that are good candidates for correction - quality cuts, kinematic windows $\rightarrow$ correcting bad tracks will produce fake $\not\!\!E_T$
- Muons are minimum ionizing particles and thus cannot considered pions in the above set - correct caloMET using standard CMS methods
- Electrons have large electromagnetic fraction (EMF) and deposit a large fraction of their energy in the ECAL; remove from list of candidate tracks for which caloMET is corrected using response function
- Evaluate performance of tcMET - compare effect on W (real $\not\!\!E_T$), DY (fake MET), QCD (standard)
- Additional discussion on secondary considerations - resolution, etc.

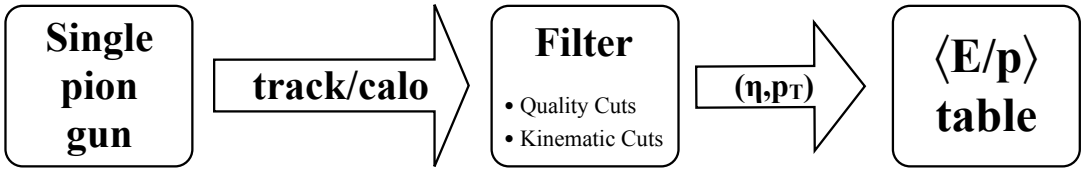

Figure 2: Conceptual process used to derive the calorimeter response function starting with a single π sample.

5 Derivation of the Response Function (RF)

In order to account for the non-linearity of the calorimeter we need to understand how the detector responds to charged hadrons (leptons receive special treatment and will be discussed later). The goal is to create a lookup table that allows for a simple determination of the expected detector response based only on track kinematics.

Fig. 2 shows the conceptual steps involved in the derivation of the response function. Starting with a single particle gun, the track and calorimeter information for each pion is extracted and passed through a set of filters. If the pion passes all cuts, the detector response is calculated and entered into the appropriate η, p_T bin of a 2D histogram. Finally, the response function is extracted from the histogram. More detail on each of these steps follows below.

5.1 Selection of pions

The starting point is a sample of single charged pions. Each pion in the sample is passed through a kinematic filter that selects those tracks with:

- $p_T \in$ (2,100) GeV,
- $|\eta| < 2.4$.

The requirement on η is imposed by the tracker. The current release of CMSSW provides less reliable tracking below 2 GeV p_T. As a result, these tracks are currently omitted from consideration for correction with the response function. Section 6 contains further details about the handling of these during the application of the correction. Tracks in this kinematic region will be reconsidered in later releases as tracking algorithms improve. Although real tracks with p_T above 100 GeV are typically well-measured, the relative number of fakes increases with increasing p_T. Thus, we have restricted ourselves to a kinematic region where we can confidently trust the fidelity of the track collection. Those tracks passing the kinematic cuts are then passed through additional quality filters. We require

- number of hits > 7,
- χ^2 / ndof < 5,
- $|d_0| < 0.01$,

where the impact parameter is given in centimeters measured from $(0, 0)$. In later releases the impact parameter will be calculated from the beamspot. The first two are standard cuts. The third, d_0, usually gets distorted to some large non-zero value if the estimate of the curvature is in error and hence is a good indicator of badly measured tracks. The $N - 1$ distributions for our sample of pions are shown in fig. 3, justifying the choice of cuts made.

Quality cuts are tight to ensure that only good tracks go into the lookup table and prevent outliers from distorting the response.

5.2 Definition of the response

The events that pass the kinematic and quality cuts form the working subset of pions from which the response function is built. For each pion the momentum *p* measured at the vertex serves as a seed to the analytical propagator which swims the trajectory out to the expected location of of the intersection of the particle with the calorimeter. Using this point as the center, a cone of ΔR = 0.5 is opened and the energy in each calorimeter tower within the cone is summed to determine the energy deposited in the detector.

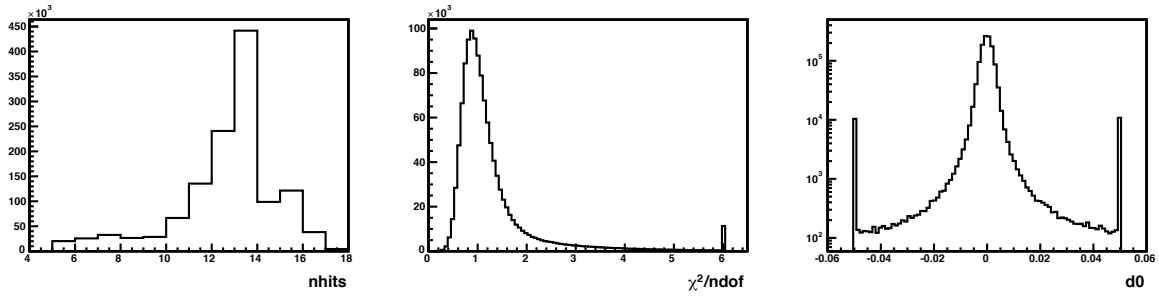

Figure 3: N-1 distributions of the three quality variables for the single pion sample.

The use of towers rather than rec hits is a deliberate and important one. MET is calculated as a vector sum of calorimeter towers, as verified in fig. 4. Thus, towers are a natural calorimeter segmentation with which to determine detector response. The difference goes beyond geometry though. Rec hits are subject to different thresholds than towers and thus can give different results. This is particularly important in the 0-jet bin where unclustered energy dominates. Put simply, a naive sum over rec hits in the cone can give a significantly different result than the corresponding sum over towers. This having been said, there are environments in which rec hits may be the more appropriate choice.

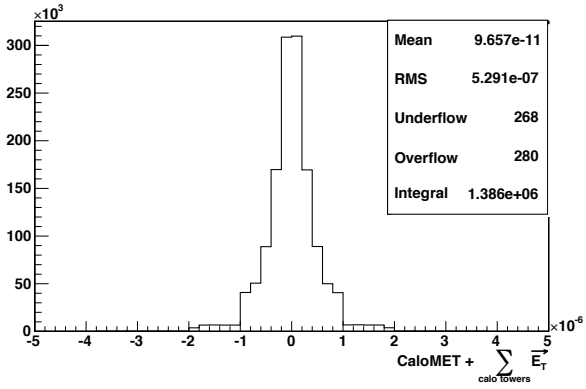

Figure 4: Difference between the value of the reconstructed MET object and the negative sum of calorimeter towers obtained from the Drell-Yan sample.

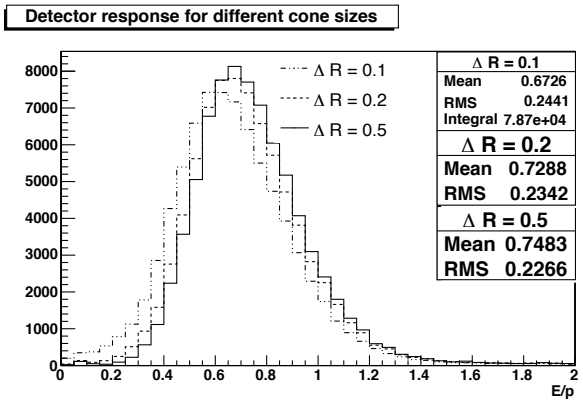

Figure 5: Comparison of detector response for different cone sizes. As a representative sample, we consider pions in the barrel with p_T between 20 and 30 GeV.

The detector response, E/p, is defined to be the energy deposited in the cone divided by the magnitude of the track momentum measured at the vertex. Fig. 5 shows the detector response for several different cone sizes. A cone of ΔR = 0.1 provides a lower response that do larger ones. As the cone size increases, there is a significant improvement for ΔR = 0.2 and then the response improves asymptotically for larger radii. This observation is not surprising for the clean environment of a single particle sample. Given this flexibility, a large cone size of ΔR = 0.5 was chosen.

5.3 Response function

The response for each good pion is determined as described in the previous section and entered into a two-dimensional histogram with variable bins in η, p_T. The exact binning is given below. Binning in η is symmetric and thus is only given for positive values.

- η = {0, 0.087, 0.174, 0.261, 0.348, 0.435, 0.522, 0.609, 0.696, 0.783, 0.879, 0.957, 1.044, 1.131, 1.218, 1.3 05, 1.392, 1.479, 1.566, 1.653, 1.740, 1.830, 1.930, 2.043, 2.172, 2.322, 2.5}
- p_T = {0, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100}

The use of variable bin sizes is an important and intentional decision. Bins in η coincide with towers in the calorimeter[6], logical given the choice of definition of the response. On the other axis, bins have 1 GeV size at low p_T and increase to 10 GeV at the high end of the spectrum. This last point deserves emphasis. A larger variation in detector response is expected at low p_T, thus requiring finer binning, than at higher p_T, where coarser separation will suffice. Once the histogram is filled, a one-dimensional histogram of the response in each bin is created and the mode (i.e. most probable value) is extracted. The extracted value fills the corresponding bin of the response function.

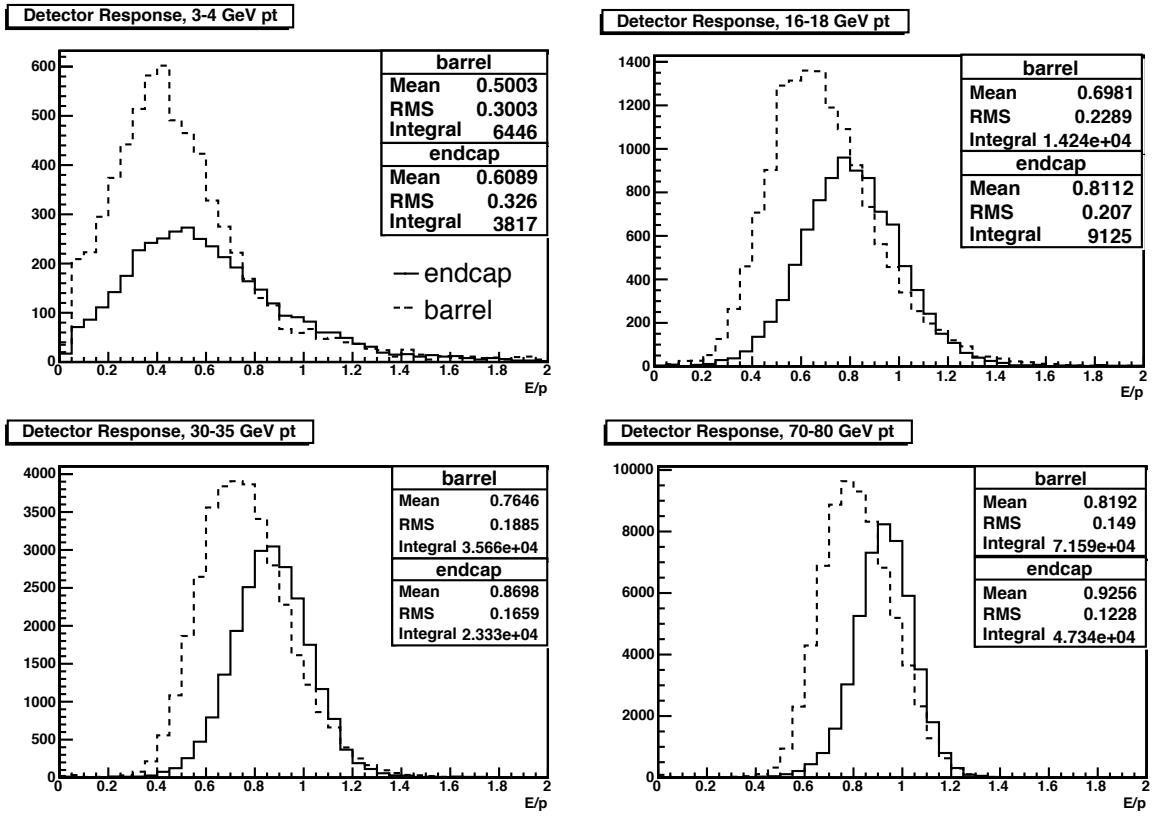

Figure 6: Detector response of single pions in the barrel and endcap for several p_T intervals.

Fig. 6 shows the detector response in the endcap and barrel for several different p_T intervals. Two general features are:

- for a given p_T interval, response is better in the endcap than in the barrel,
- response improves with increasing p_T.

Better response in the endcap is likely due to the increased material budget and relatively higher p for a given p_T at larger η. Binning in p rather than p_T was considered but no improvement in final performance was observed. Improved response at higher p_T is expected. Extracting the response using the mode is a common method, but not the only possibility. Several other options were also evaluated. Table 1 compares the mode with two other methods.

The response in the first column was found by calculating the mean of entries in each bin. The value in the second column was determined by fitting each bin with a gaussian on the response interval (0.1, 1.2) and extracting the mean. This acts like a truncated mean. The response in the last column is determined from the mode, or most probable value, in each bin. A couple of general features are evident. The mean and gaussian fit yield similar values, except at low p_T and large η. The mode tends to give a slightly lower response than the other methods. This is particularly evident at low p_T.

Table 1: Comparison of methods for deriving RF

Interval	Mean	Gaussian Fit	Mode (M.P.V.)
p_T = 2.5, η = 0.05	0.52	0.49	0.27
p_T = 2.5, η = 2.25	0.75	0.66	0.43
p_T = 17, η = 0.05	0.74	0.70	0.63
p_T = 17, η = 2.25	0.87	0.84	0.85
p_T = 32.5, η = 0.05	0.79	0.79	0.79
p_T = 32.5, η = 2.25	0.93	0.90	0.87
p_T = 75, η = 0.05	0.84	0.84	0.77
p_T = 75, η = 2.25	0.94	0.95	0.93

6 Application of the Response Function

Track-corrected MET (tcMET) is calculated for an event using:

- caloMET
- muon collection (*muons*)
- electron collection (*pixelMatchGsfElectrons*)
- track collection (*ctfWithMaterialTracks*)
- response function (RF)

For each event, the algorithm begins by identifying muons and correcting the caloMET for them. Muons are corrected at the outset by subtracting the p_T of each associated track from the caloMET. CaloMET corrected for muons will serve as the baseline for comparisons throughout this paper.

$$\begin{aligned} \text{baseline MET} &= \text{caloMET} - \sum_{\text{muons}} \vec{p}_T, \\ &= -\sum_{\text{towers}} \vec{E}_T - \sum_{\text{muons}} \vec{p}_T \end{aligned} \tag{1}$$

The algorithm next identifies electron-like objects and skips them. No correction is applied for tracks matched to these objects because they deposit a large fraction of their energy electromagnetically. Further discussion on electrons follows below. The remainder of the tracks are candidates for correction using the derived response function. The response function serves as a lookup table that takes as input the kinematics of a track and returns as output the expected energy deposited by the track in the calorimeter. This sequence is represented in fig. 7.

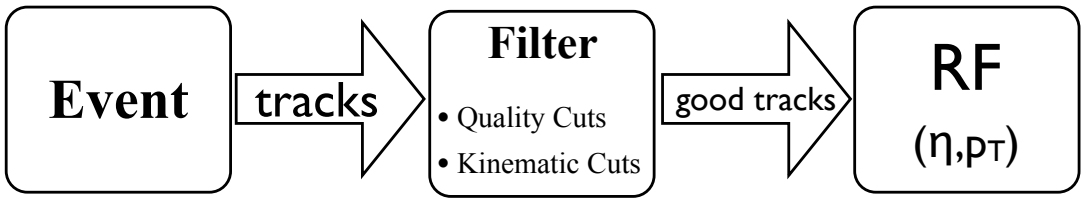

Figure 7: Conceptual process used to apply the response function to correct MET for tracks in an event. A set of quality cuts is used to select good tracks that are then passed to the response function from which a correction factor is extracted.

Tracks not matched to leptons are passed through a series of filters. Quality variables analogous to those in the derivation of the response function are used, but with less restrictive cuts required for the application of the response function. The subset of tracks that survive form the list of good tracks for which $\not\!\!E_T$ is corrected. The actual correction is implemented by removing the expected energy $\langle\vec{E}_T\rangle$deposited by each good track in the calorimeter, determined using the response function, and replacing it with the track momentum at the vertex. Tracks with $p_T <$

2 GeV are fully compensated for assuming no response from the calorimeter $\left(\text{i.e.}\langle\vec{E}_T\rangle = 0\right)$. Tracks with $p_T >$ 100 GeV receive no correction for reasons described previously.

$$\begin{aligned} \text{tcMET} &= \text{baseline MET} + \sum_{\substack{\text{good}\\\text{tracks}}} -\langle\vec{E}_T\rangle - \sum_{\substack{\text{good}\\\text{tracks}}} \vec{p}_T, \\ &= -\sum_{\text{towers}} \vec{E}_T - \sum_{\text{muons}} \vec{p}_T + \sum_{\substack{\text{good}\\\text{tracks}}} \langle\vec{E}_T\rangle - \sum_{\substack{\text{good}\\\text{tracks}}} \vec{p}_T \end{aligned} \tag{2}$$

It is important to note that the correction for each good track involves two sets of coordinates. The expected energy deposition for each track is removed from the calorimeter. This location is determined using the track at the vertex as a seed to the analytical propagator, as was done in the derivation of the response function. The track momentum that replaces it is taken at the vertex. To be explicit, the correction for a single component of MET takes the form:

$$(\text{tcMET})_x = (\text{baseline MET})_x + \sum_{\substack{\text{good}\\\text{tracks}}} \langle E\rangle \sin\theta_c \cos\phi_c - \sum_{\substack{\text{good}\\\text{tracks}}} p_T \cos\phi_v \tag{3}$$

where θ_c, ϕ_c are the polar and azimuthal position coordinates of the particle at the calorimeter face and ϕ_v is the azimuthal angle of the track at the vertex.

6.1 Track quality cuts - how <u>not</u> to generate fake MET

Correcting $\not\!\!E_T$ for a badly measured track will generate fake MET and make the $\not\!\!E_T$ determination worse. Consider a track that, due to a pattern recognition problem in a noisy environment, has a very large estimated momentum. The amount of this over-estimate will translate directly into excess MET if the track is passed to the response function. A set of kinematic and quality cuts is used to filter these mis-measurements from the track collection and avoid generating fake $\not\!\!E_T$.

It is necessary that the same kinematic cuts applied in the derivation of the response function are also used in its application. However, the quality cuts need not take the same values and are loosened here to be more inclusive. The application of the response function defines good tracks to have:

- number of hits > 6,
- χ^2 / ndof < 5,
- $|d_0| < 0.05$

The set of N-1 distributions in fig. 8 are from the Drell-Yan sample, justifying the choice of cuts.

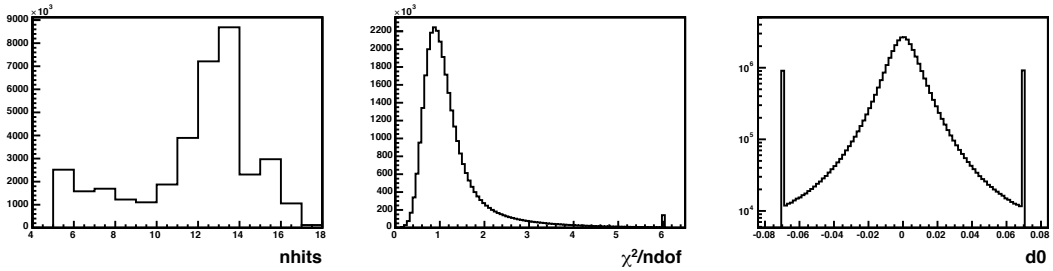

Figure 8: N-1 distributions of the three quality variables for the Drell-Yan sample.

6.2 Correction for electron-tracks

Applying the response function to an electron-like track, i.e. either an electron or a pion that showered predominantly in the ECAL, will generate fake MET. Thus, it is important to identify and avoid correcting such particles. As such, we consider objects found in the PixelMatchGsfElectronCollection as a function of the ratio of hadronic to electromagnetic energy deposited in the calorimeter. Tracks matched to an object whose h/e is sufficiently small are excluded from the list of candidate tracks corrected with the response function.

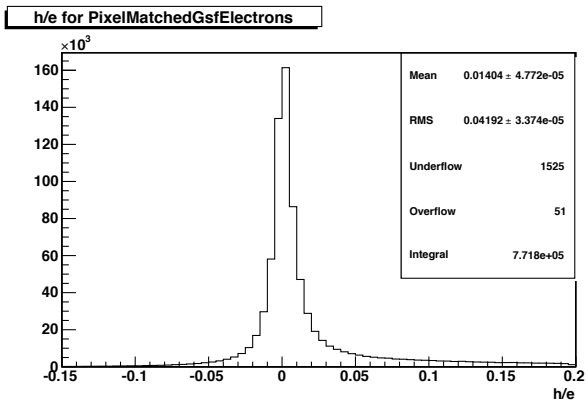

Figure 9: Distribution of *h/e* for objects identified as PixelMatchedGsfElectrons

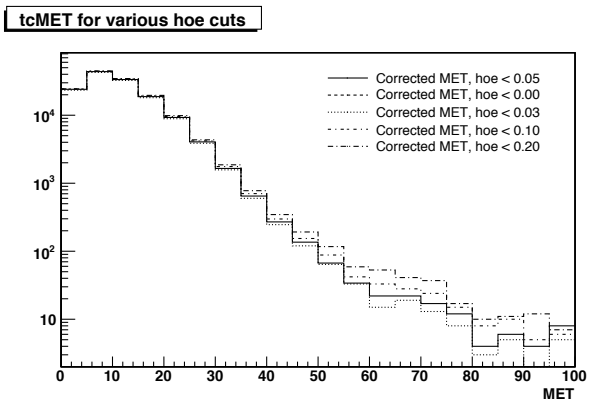

Figure 10: Comparison of effect of various *h/e* cuts on tcMET obtained from a subset of dielectron Drell-Yan events.

The discriminating variable h/e is shown in fig. 9. The ratio is defined to be the sum of rec hits in the hadronic calorimeter divided by the electromagnetic supercluster energy. The efficacy of a cut on h/e was studied on a subset of the Z $\rightarrow$ ee sample. Fig. 10 shows tcMET where the h/e cut has been varied. Objects from the PixelMatchGsfElectronCollection are associated with tracks by ΔR matching. Compared to not discriminating based on the h/e variable, even a very loose cut of 0.03 sees a factor of 2 improvment which improves to a factor 4 for tighter selection.

7 Performance Evaluation

The goal at the outset was to reduce the sizable tails of the Drell-Yan $\not\!\!E_T$ distribution. The correction is tested on a Drell-Yan sample with the metric being the number of events with MET > 30. In addition, the response function is also used to correct a sample of W leptonic decays. The use of a sample with real $\not\!\!E_T$ is an important cross check. Correcting the Drell-Yan tails with a naive additive or multiplicative factor will eliminate events with high $\not\!\!E_T$, but will also remove events with real $\not\!\!E_T$. Thus, it is necessary check that any attempt to correct for mis-measured MET does not significantly impact the ability to identify those events for which $\not\!\!E_T$ is an important and real signature.

7.1 Muon final states

Separate evaluations will be performed for two distinct cases: di-muon and di-electron final states. Consider the former. The baseline $\not\!\!E_T$ is defined to be caloMET plus muon corrections. Additionally, we define a *good muon*:

- $p_T > 20$ GeV,
- number of hits > 6,
- χ^2 / ndof < 5,
- $|d_0| < 0.25$,

where the cuts on the number of hits and the impact parameter are on the track matched to the muon and the third cut is on the normalized chi-squared of the global fit. Only Drell-Yan events with exactly two good muons and

Table 2: Performance of tcMET on Drell-Yan in the muon final state.

Case	0 jets	1 jet	2 jets	3+ jets
baseline	4006/120177	6278/32161	1859/5517	363/856
	3%	20%	34%	42%
tcMET	1687/120177	1981/32161	767/5517	165/856
	1%	6%	14%	19%
factor of improvement	2.4	3.2	2.4	2.2

W decays with at least one good muon in the final state are used for comparison. These requirements ensure the former contains predominantly events with fake MET and the latter with real.

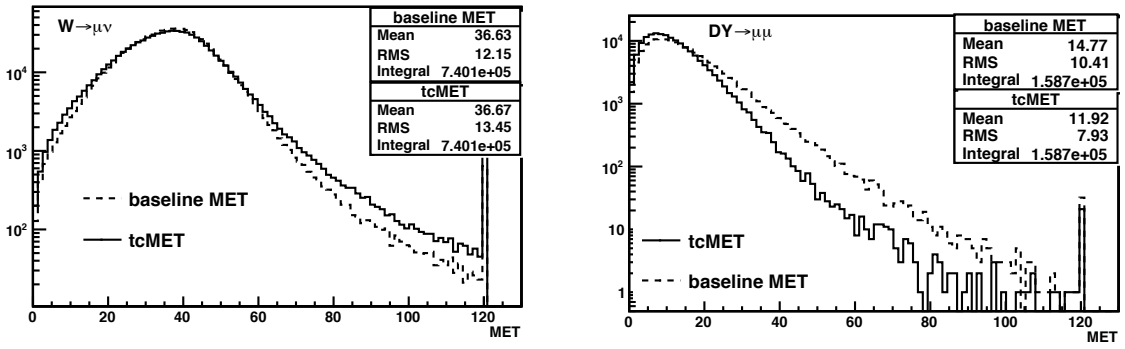

Figure 11: Comparison of baseline MET and tcMET for W decays and Drell-Yan events in the muon final state.

Fig. 11 show baseline MET and tcMET. The W distribution with real $\not\!\!E_T$ on the left shows no significant change after correction. The tail of the Drell-Yan distribution on the right shows a dramastic reduction in the number of events with $\not\!\!E_T$. The number of events with MET > 30 GeV is reduced by a factor of 2.7. The reduction is enhanced to a factor of 4.7 for a cut at 50 GeV. It is also useful to break these results into n-jet bins. Here, the jet counting parameters are $|\eta| < 3$ and $E_T > 15$ GeV (uncorrected).

Changes in the W sample are less than 5% in any n-jet bin. Table 2 shows the breakdown for Drell-Yan. Each entry shows the fraction of events with MET > 30 GeV in raw numbers and as a percentage. The final row shows the factor of improvement for tcMET over baseline MET.

The majority of Drell-Yan events fall into the 0-jet bin. These events are the target of the correction and tail is reduced by a factor of 2.4. Although this correction wasn't conceived with jets in mind, the large number of tracks in these events gives reason to expect similar reductions. The improvement in the 2-jet bin is commensurate with that seen in events with no jets. The 1-jet bin shows even better. This is not entirely unexpected. These events contain one sizable jet which provides additional tracks for correction. The presence of a single large jet that is typically under-measured produces significant asymmetry, providing excellent conditions for correction. The observation of a smaller improvement in the 2-jet bin is also reconciled in this view as the two jets typically have some spatial separation which results in corrections for tracks in one jet partially canceling out corrections for tracks in the other jet.

7.2 Electron final state

A similar comparison is also done for the electron final state. Only Drell-Yan events with exactly two electrons with $p_T > 20$ GeV are used for comparison. The requirement for W decays is the presence of at least one electron with the same p_T cut. There are no further requirements on the electrons.

Fig. 12 show baseline MET and tcMET. The W distribution with real $\not\!\!E_T$ on the left shows no significant change after correction. The tail of the Drell-Yan distribution on the right shows a significant reduction in the number of events with $\not\!\!E_T$. The number of events with MET > 30 GeV is reduced by a factor of 2.7, the same as in the muon final state. Table 3 below shows the breakdown by n-jet bin for Drell-Yan. The W sample is not

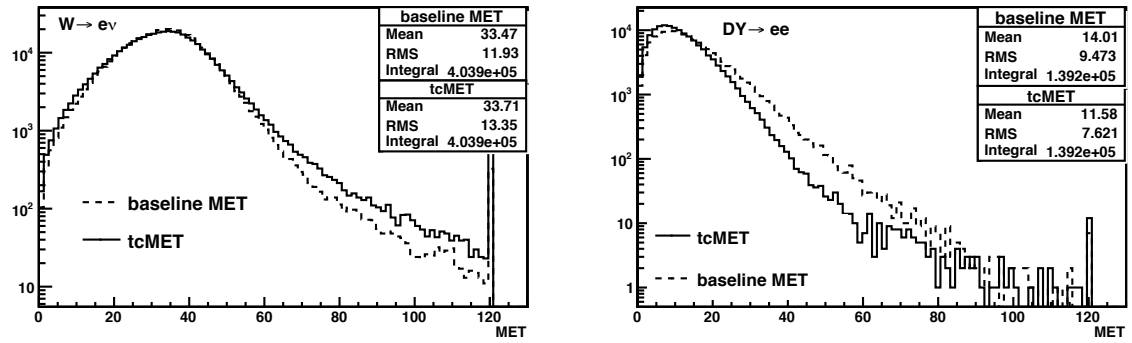

Figure 12: Comparison of baseline MET and tcMET for W decays and Drell-Yan events in the electron final state.

changed by more than $\sim 2\%$ in any bin. In addition to the previously jet counting criteria, additional care must be taken here to remove electrons from the list of jets. Here we remove jets that contain an object from the PixelMatchGsfElectronCollection with $h/e < 0.1$ and $p_T > 20$ GeV.

Table 3: Performance of tcMET on Drell-Yan in the electron final state.

Case	0 jets	1 jet	2 jets	3+ jets
baseline	3117/108383	4169/26291	1057/3968	205/515
	3%	16%	27%	40%
tcMET	1392/108383	1329/26291	378/3968	103/515
	1%	5%	10%	20%
factor of improvement	2.2	3.1	2.8	2.0

These results are similar to those seen for the muon final state. The correction for the electron final state is slightly less effective in the 0-jet bin and slightly more effective in the 2-jet bin than that observed for muons. The variation in performance amongst the various bins can be explained by the same reasons given in the case of muons above.

7.3 QCD processes

Before moving on it is worthwhile to consider the effect a track-based correction has on MET for QCD processes. These events are expected to contain little real MET and thus serve as a good platform for comparison. Here, inclusive QCD samples over several p_T ranges were tested. In all cases, tcMET is seen to improve both the tails and the resolution of the the $\not\!\!E_T$ distribution. This improvement increases with increasing p_T up to several hundred GeV.

8 Missing E_T Resolution

The track-based correction shown above is quite effective at reducing the tail of the $\not\!\!E_T$ distribution in Drell-Yan. No effort was made to improve resolution. Although reducing the tails is a desirable goal, tcMET may be of limited utility if it significantly worsens MET resolution. A crosscheck was performed by fitting the distribution of each component of $\not\!\!E_T$ with a Gaussian from which the the resolution determined using

$$\sigma_{\mathrm{MET}} = \sigma \cdot \sqrt{\frac{4-\pi}{2}} \tag{4}$$

as described in [5]. Here, σ is the width of the distribution of one of the components. For our determination, we define σ to be the average of the widths of the distributions of the two components. A detailed derivation of this expression can be found in Appendix A. Fig. 14 shows the distributions of the y-component of baseline MET and tcMET. Table 4 shows the width of each component.

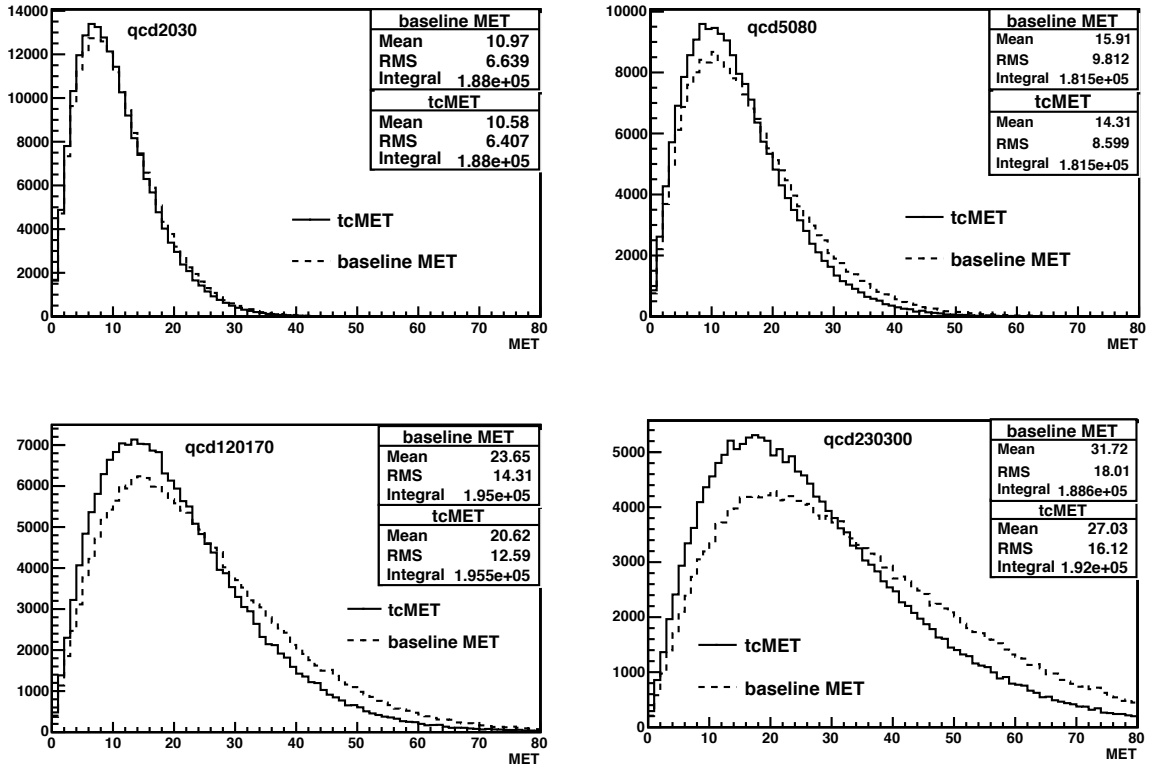

Figure 13: Comparison of tcMET to baseline MET. Four different p_T ranges are shown, where qcdXY refers to the lower value, X, and upper value, Y, of the p_T interval. In all cases, tcMET improves both the tail and resolution of the MET distribution with the degree of improvement increasing with p_T. All plots were made from CSA07 inclusive QCD samples.

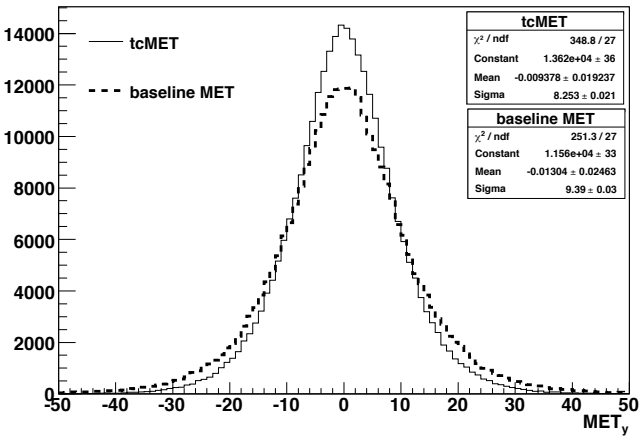

Table 4: tcMET Resolution

Case	σ_x	σ_y	σ_{MET}
baseline	9.43	9.49	6.2
tcMET	8.29	8.32	5.4

Figure 14: Distribution of y-component of baseline MET and tcMET for DY→ee. The table shows an improvement in MET resolution of $\sim$ 15 %.

MET resolution gets better after applying the track-based correction. The resolution improves by $\sim$ 15%. This is an unintended bonus - we set out to reduce the tail and along the way also improved the resolution. The presence of this enhancement validates procedure and indicates that the algorithm developed is a reasonable one.

9 Type-I JES Corrections and MET

Correcting MET for tracks shows large reductions of the tails of the MET distributions for Drell-Yan. In particular, the algorithm corrects for unclustered energy in the 0-jet bin, a feature that does not currently exist in the standard $\not\!\!E_T$ corrections. However, type-I JES corrections [5] do correct for clustered energy and thus it is interesting to

ask what effect, if any, do these have on the Drell-Yan sample.

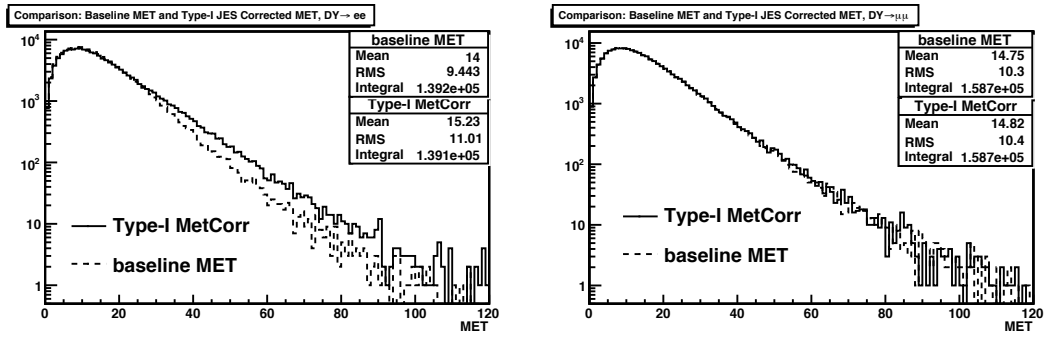

Figure 15: Comparison of caloMET+muons+type-I JES corrections to baseline MET over all jet bins.

A comparison is made between baseline MET and baseline MET + type-I JES corrections. Only events with two good muons or electrons were kept. The standard type-I JES corrections as implemented in CMSSW_1_6_X were used. Fig. 15 shows the comparison for the electron (left) and muon (right) final states. No improvement over the baseline is observed in the muon final state. Type-I JES corrections make the tail considerably worse in the electron final state. The suspicion is that the JES implementation is correcting electrons as jets, thus generating fake MET. However, according to the documentation, this should not be the cause as electrons are explicitly removed and there is an additional cut of $|h/e| < 0.1$. This discrepancy has been reported to the JetMet group and is under investigation. In summary, to reduce the tails of the MET distribution of Drell-Yan, tcMET performs significantly better than the type-I JES corrections.

10 Conclusion

MET is a simple variable whose measurement involves many complications. The design of CMS involved many trade-offs which strongly impact the ability to accurately measure the energy balance of the event. The strong magnetic field, large material budget, and the non-linearity of the calorimeter all contribute to the difficulties that exist in properly determining the $\not\!\!E_T$. The superior resolution of the CMS tracker provides a useful tool for improving the tails of MET distributions, as well as the resolution, by correcting for charge tracks. The algorithm presented reduces the tail of the MET distribution for Drell-Yan by a factor of 2.7 above 30 GeV while preserving the distribution for W events with real $\not\!\!E_T$. This improvement increases as one increases the threshold for $\not\!\!E_T$. The algorithm also improves MET resolution by $\sim$15 % for both Drell-Yan events as well as QCD.

References

[1] **CMS AN-2008/015**, V. Krutelyov et al., *"Expectations for $t\bar{t} \to ll$ in the early phase of CMS"*.

[2] **CMS PAS TOP-08-001**, CMS collaboration, *"Expectations for observation of top quark pair production in the dilepton nal state with the rst 10 pb$^{?1}$ of CMS data"*.

[3] CMS collaboration, *"https://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopQuarkNotesSlava"*.

[4] **CMS AN-2004/015**, O. Kodolova et al., *"Jet energy correction with charged particle tracks in CMS"*.

[5] **CMS AN-2007/041**, CMS collaboration, *"$\not\!\!E_T$ performance in CMS"*.

[6] CMS Collaboration, *"CMS Physics Technical Design Report v1.: Detector Performance and Software,"* CERN LHCC/2006-001 (2006)

A Definition of MET Resolution

Section 8 stated the MET resolution to be

$$\sigma_{\text{MET}} = \sigma \cdot \sqrt{\frac{4-\pi}{2}} \tag{5}$$

as defined in [5], where σ is the r.m.s. of either transverse component of MET. Although this expression is sufficient for the purposes of this note, the derivation is not difficult and it is informative to understand the assumptions behind it. Begin by assuming an ideal detector. In this limit, the transverse components $\not\!\!E_x, \not\!\!E_y$ take Gaussian distributions with mean $\mu = 0$ and r.m.s. σ. The distribution of MET is then a combination of the two one-dimensional Gaussians. In addition, we impose the requirement that the distribution is non-negative and properly normalized. Rewriting the resulting two-dimensional distribution in one-dimension with appropriate Jacobian yields

$$P(\not\!\!E_T, \sigma) = \frac{\sqrt{2\pi}}{\sigma}\theta(\not\!\!E_T) \cdot \not\!\!E_T \cdot G(\not\!\!E_T, 0, \sigma), \tag{6}$$

where

$$G(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{7}$$

The term linear in $\not\!\!E_T$ is the Jacobian factor and the Heaviside function enforces the non-negativity of the distribution. The normalization of the distribution is easy to verify.

$$\begin{aligned}
\int d\not\!\!E_T \; P(\not\!\!E_T, \sigma) &= \frac{1}{\sigma^2}\int d\not\!\!E_T \; \theta(\not\!\!E_T) \; \not\!\!E_T \; e^{-\frac{\not\!\!E_T^2}{2\sigma^2}} \\
&= e^{-\frac{\not\!\!E_T^2}{2\sigma^2}}\Big|_0^\infty \\
&= 1
\end{aligned} \tag{8}$$

The mean of the distribution in simply

$$\begin{aligned}
\langle \not\!\!E_T \rangle &= \int d\not\!\!E_T \; \not\!\!E_T \; P(\not\!\!E_T, \sigma) \\
&= \frac{1}{\sigma^2}\int d\not\!\!E_T \; \theta(\not\!\!E_T) \; \not\!\!E_T^2 \; e^{-\frac{\not\!\!E_T^2}{2\sigma^2}} \\
&= \frac{1}{\sigma^2}\int_0^\infty d\not\!\!E_T \; \not\!\!E_T^2 \; e^{-\frac{\not\!\!E_T^2}{2\sigma^2}}
\end{aligned}$$

Define $\alpha = \frac{1}{2\sigma^2}$. Then, the integral can be rewritten

$$\langle \not\!\!E_T \rangle = -2\alpha \; \frac{d}{d\alpha}\int_0^\infty d\not\!\!E_T \; e^{-\alpha \not\!\!E_T^2}$$

Making the substitution $\xi = \sqrt{\alpha}\not\!\!E_T$ and making use of the fact that $\int_{-\infty}^{\infty} d\xi \; e^{-\xi^2} = \sqrt{\pi}$ yields

$$\begin{aligned}
\langle \not\!\!E_T \rangle &= -2\alpha \; \frac{d}{d\alpha}\left(\frac{1}{\sqrt{\alpha}}\frac{\sqrt{\pi}}{2}\right) \\
&= \frac{1}{2}\sqrt{\frac{\pi}{\alpha}} \\
&= \sigma \cdot \sqrt{\frac{\pi}{2}}
\end{aligned} \tag{9}$$

The r.m.s. of the distribution can be calculated by way of $\sqrt{\langle \not{E}_T^2 \rangle - \langle \not{E}_T \rangle^2}$.

$$\begin{aligned}
\langle \not{E}_T^2 \rangle &= \int d\not{E}_T \; \not{E}_T^2 \; P(\not{E}_T, \sigma) \\
&= \frac{1}{\sigma^2} \int d\not{E}_T \; \theta(\not{E}_T) \; \not{E}_T^3 \; e^{-\frac{\not{E}_T^2}{2\sigma^2}} \\
&= \frac{1}{\sigma^2} \int_0^\infty d\not{E}_T \; \not{E}_T^3 \; e^{-\frac{\not{E}_T^2}{2\sigma^2}} \\
&= -2\alpha \; \frac{d}{d\alpha} \int_0^\infty d\not{E}_T \; \not{E}_T \; e^{-\alpha \not{E}_T^2} \\
&= -2\alpha \; \frac{d}{d\alpha} \left(-\frac{1}{2\alpha} e^{\alpha \not{E}_T^2} \right) \Bigg|_0^\infty \\
&= 2\sigma^2 \qquad (10)
\end{aligned}$$

Combining gives

$$\begin{aligned}
\sigma_{\text{MET}} &= \sqrt{\langle \not{E}_T^2 \rangle - \langle \not{E}_T \rangle^2} \\
&= \sqrt{2\sigma^2 - \left(\sigma \sqrt{\frac{\pi}{2}} \right)^2} \\
&= \sigma \cdot \sqrt{\frac{4-\pi}{2}} \qquad (11)
\end{aligned}$$