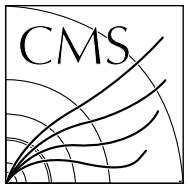
Available on CMS information server

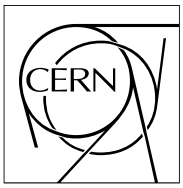
CMS AN -2008/089



The Compact Muon Solenoid Experiment

Analysis Note

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



18 October 2008 (v3, 11 June 2009)

MET Reconstruction, Performance, and Validation

Greg Landsberg and Filip Moortgat

Abstract

We discuss progress in reconstruction, optimization, and validation of Missing Transverse Energy (MET) in the CMS experiment, using combination of Monte Carlo and cosmics data. This note is an incremental update since AN/2007-041.

CMS 2008/089

2009/06/08
Archive Id: 1.18
Archive Date: 2009/06/07 23:01:05

$\not\!\!E_T$ Reconstruction, Performance and Validation

Jim Alexander[1], Colin Bernet[2], Anwar Bhatti[9], Freya Blekman[1], Claudio Campagnari[11], Rick Cavanaugh[3,4], Frank Chlebana[3], Costas Foudas[5], Bill Gary[6], Frank Golf[7], Alfredo Gurrola[8], Kenichi Hatakeyama[9], Patrick Janot[10], Teruki Kamon[8], Puneeth Kavalase[11], Aleko Khukhunaishvili[1], Vyacheslav Krutelyov[11], Shuichi Kunori[12], Florent Lacroix[4], Edward Laird[13], Greg Landsberg[14], Feng Liu[6], Filip Moortgat[15], Radek Ofierzynski[16], Ronald Remington[17], Alexei Safonov[8], Xin Shi[1], Ka Vang Tsang[14], Chris Tully[13], Avi Yagil[7], and John Yelton[17]

[1] Cornell University, Ithaca, NY, United States
[2] LLR/Ecole Polytechnique, Palaiseau, France
[3] Fermi National Laboratory, Batavia, IL, United States
[4] University of Illinois, Chicago, IL, United States
[5] Imperial College, London, UK
[6] University of California, Riverside, CA, United States
[7] University of California, San Diego, CA, United States
[8] Texas AM University, College Station, TX, United States
[9] Rockefeller University, New York, NY, United States
[10] CERN, Geneva, Switzerland
[11] University of California, Santa Barbara, CA, United States
[12] University of Maryland, College Park, MD, United States
[13] Princeton University, Princeton, NJ, United States
[14] Brown University, Providence, RI, United States
[15] ETH, Zurich, Switzerland
[16] Northwestern University, Evanston, IL, United States
[17] University of Florida, Gainesville, FL, United States

Abstract

We discuss progress in reconstruction, optimization, and validation of Missing Transverse Energy ($\not\!\!E_T$) in the CMS experiment, using a combination of Monte Carlo and Cosmics data. This note is an incremental update of AN/2007-041. The correction scheme for calorimeter based MET has been refined by improved muon and tau corrections, as well as through the addition of Type-2 corrections for unclustered and out-of-cone energy. Fundamental improvements to the calorimeter-based $\not\!\!E_T$ have been made through the development of "track-corrected MET", which exploits the precise measurement of charged particles in the tracker, and "Particle Flow MET", which makes use of both charged and neutral particles reconstructed using the Particle Flow algorithm. A Data Quality Monitoring (DQM) framework for $\not\!\!E_T$ has been put in place in order to identify good runs and luminosity sections for analyses that rely on the $\not\!\!E_T$ quantity. The functionality of the framework is illustrated on CRAFT data. The treatment of hot and bad cells is discussed and data from the CMS Global Runs (notably CRAFT) is used to study the actual $\not\!\!E_T$ measurements in the CMS detector, in particular the effect of instrumental noise. $\not\!\!E_T$ monitoring triggers are proposed to study noisy events and cleaning strategies in future runs. The impact of the forward calorimeter (HF) on the $\not\!\!E_T$ resolution has been studied. Finally, the status of the Level-1 $H\!\!\!/_T$ trigger is summarized.

Contents

1 Introduction

Neutrinos and other hypothetical weakly interacting particles cannot be directly detected as they do not interact with any detector material. However, information on the presence of these particles can be extracted by exploiting the conservation of the total transverse momentum in the plane perpendicular to the beam. The momentum imbalance in this plane due to invisible particles is called *Missing Transverse Momentum* and is a 2D-vector. Its magnitude is referred to as *Missing Transverse Energy* ($\not\!\!E_T$). $\not\!\!E_T$ plays a primordial role in many physics analyses at the LHC. First of all, in searches for physics Beyond the Standard Model, any model that contains a Dark Matter candidate will lead to a signature containing significant $\not\!\!E_T$ in the final state. Also for many Standard Model channels, such as $W \rightarrow l^{\pm}\nu$ or (semi-)leptonic $t\bar{t}$, the knowledge of the Missing Transverse Momentum plays an important role.

A basic description of the $\not\!\!E_T$ reconstruction and the correction scheme can be found in [1]. In the present note, we revisit the muon and tau corrections, introduce the "Type-2 corrections" for out-of-cone effects, document the recent software changes and discuss the inclusion of the MET objects in the Physics Analysis Toolkit (PAT).

Fundamental improvements to calorimeter-based $\not\!\!E_T$ can be made by exploiting other CMS subdetectors, in particular the tracker. *Track Corrected* $\not\!\!E_T$ aims to improve the $\not\!\!E_T$ resolution and tails by replacing the expected energy deposition in the calorimeter of all well reconstructed tracks by their measured momentum in the tracker. Since at the momentum scale of interest the tracker has a far superior resolution compared to the calorimeter this should lead to significant improvements in $\not\!\!E_T$ resolution and tails. An even more refined approach consists in using the individual particles that have been reconstructed using the Particle Flow technique, in which charged as well as neutral particles are replaced. In section 4 we discuss the performance of the $\not\!\!E_T$ constructed from Particle Flow objects ($PFMET$) and show the improvements that can be expected from this technique.

Since $\not\!\!E_T$ is constructed by summing up all visible particles in the detector, it is particularly sensitive to detector malfunctions and detector resolution effects. To make sure that malfunctions do not corrupt the $\not\!\!E_T$ quantity, a Data Quality Monitoring (DQM) framework has been set up; the basic features of this, along with the quantities that are being monitored, are described in this note. Also, the policy for handling of dead or hot calorimeter cells is discussed. To take into account the detector resolution effects, one can try to take into account the uncertainties on the inputs to the $\not\!\!E_T$ by defining an $\not\!\!E_T$ *significance*. An algorithm to calculate this significance, along with an illustration of its potential, is described in section 6.

We use data from CMS Global Runs (notably the CRAFT runs during which CMS was operated at full magnetic field, 3.8 T) to study the $\not\!\!E_T$ measurements in the actual CMS detector, in particular the effect of instrumental noise. $\not\!\!E_T$ monitoring triggers are proposed to study noisy events and cleaning strategies. We study the impact of the forward calorimeter (HF) on the $\not\!\!E_T$ resolution. Finally we report on the status of the Level-1 $H\!\!\!/_T$ trigger, where $H\!\!\!/_T$ refers to the missing transverse energy calculated using jets only.

2 Updates to Calorimetric $\not\!\!E_T$ Corrections

The description of the $\not\!\!E_T$ reconstruction algorithms and the correction scheme can be found in [1]. Corrections to the raw $\not\!\!E_T$ consists of Type-1 corrections such as Jet Energy Scale (JES), Muon and Tau corrections. Additional Type-2 corrections account for unclustered energy response and out-of-cone effects. Below we describe the updated Muon and Tau correction algo-

rithms and discuss a first implementation of the Type-2 corrections.

2.1 $\not\!\!E_T$ correction for muons

Muons, since they are minimum ionizing for much of the momentum range at the LHC, pass through the detector depositing only a small fraction of their energy in the calorimeter. However, their momentum is measured precisely in the muon system and the tracker. Since they can be significant contributors to the total $\not\!\!E_T$ in the event, their footprint needs to be taken into account. This is done by adding their deposits in the calorimeter to the $\not\!\!E_T$ sum and then subtracting their transverse momentum:

$$\vec{\not\!\!E}_T{}' = \vec{\not\!\!E}_T + \sum_{\mu} \vec{E}^{\mu}_{T,calo} - \sum_{\mu} \vec{p}^{\mu}_{T}, \tag{1}$$

where $\vec{\not\!\!E}_T$ and $\vec{\not\!\!E}_T{}'$ are the two-dimensional vectors of the missing transverse momentum before and after the muon correction, $\vec{E}^{\mu}_{T,calo}$ is a two-dimensional vector corresponding to the muon's calorimeter deposit pointing from the IP to the muon's entry point to the calorimeter, $\vec{p}^{\mu}_{T}$ is the component of the muon's momentum in xy-plane; the sum is taken over the selected reconstructed muons (see below).

2.1.1 Muon Corrections in CMSSW 2_2_X

The muon $\not\!\!E_T$ correction implemented in CMSSW 2_2_X corrects the $\not\!\!E_T$ for muons which pass the following quality cuts:

- muon must be a Global Muon,
- $p^{\mu}_{T} > 10.0$ GeV/c,
- $|\eta| < 2.5$,
- Number of valid hits in the silicon tracker > 5,
- Error on signed transverse curvature (qoverpError) < 0.5.

The above cuts are the defaults and are fully configurable. The global muons which pass the above cuts are used to make a new muon collection, *goodMuonsforMETCorrection*, which is then used to correct the $\not\!\!E_T$. This collection is not made persistent in the event however. The performance of the correction in is demonstrated in Fig. 1 using 20,000 Pythia $t\bar{t}$ events generated as part of the Summer08 production (`/TauolaTTbar/Summer08_IDEAL_V9_v1/GEN-SIM-RECO`).

Depending on the isolation quality of the muon, the $E^{\mu}_{T,calo}$ contribution to the $\not\!\!E_T$ correction is given by a sum over towers traversed by the muon if the muon is isolated, and by a median value of such sum expected from the muon alone if the muon is not isolated. Non-isolated muons should not be treated in the same way as isolated muons because the deposits in the calorimeter around a non-isolated muon would not be entirely due to the muon itself. We would then end up over-correcting the $\not\!\!E_T$. We define a muon to be isolated if: $\sum_{dR<0.3} p_T < 3$ GeV/c, and $\sum_{\mathrm{ECAL+HCAL}} E_T < 5$ GeV, where the first sum is taken over tracks in a cone with $dR < 0.3$ with respect to the muon, and the second sum is taken over towers in a cone of 0.3 taking into account only the ECAL and HCAL contributions from the towers. The median value is extracted from an analytic function of p_T and η of the muon. This function was derived using a high statistics single-muon MC sample reconstructed in CMSSW 2_0_9. The performance of the average correction is demonstrated in Fig. 2.

In addition to the implementation of the $\not\!\!E_T$ correction for muons using the full CMSSW framework, we provide an implementation of the corrections which does not require the full frame-

work and can be used either in FWLITE or in plain ROOT.

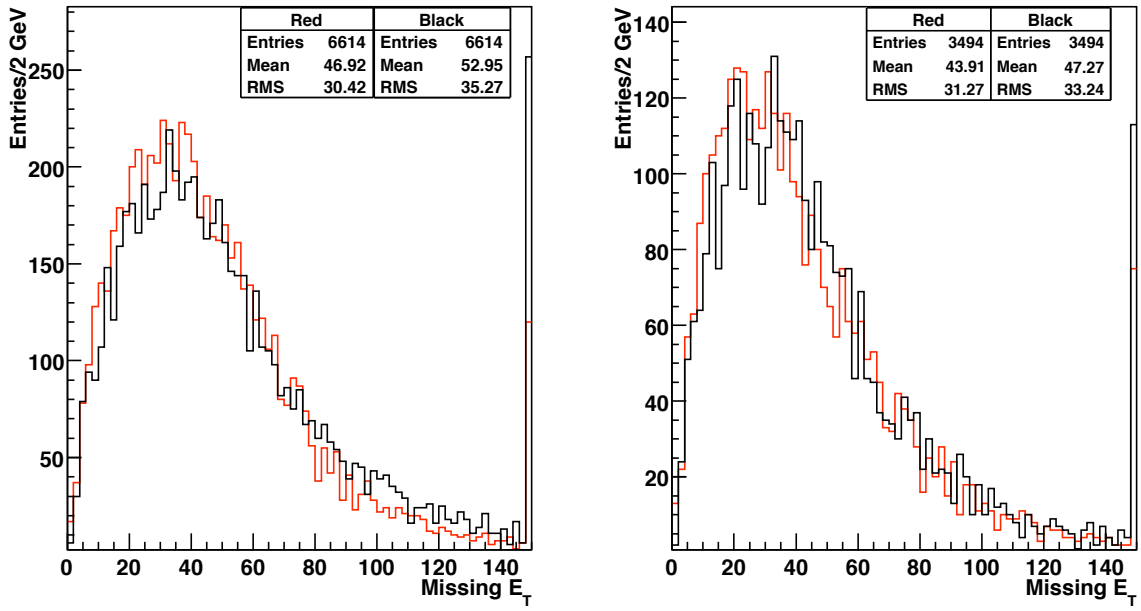


Figure 1: In black is plotted the raw $\not\!\!E_T$, and in red is the corrected $\not\!\!E_T$ for $t\bar{t}$ events. In the left we plot the $\not\!\!E_T$ for all events which have at least one muon in the *goodMuonsforMETCorrection* collection, while on the right in red is plotted the $\not\!\!E_T$ for all events that have at least one non-isolated muon in the *goodMuonsforMETCorrection* collection. The last bin is the overflow. The corrected and uncorrected $\not\!\!E_T$ have very similar shapes, but one sees that the tails extend much further out for the uncorrected $\not\!\!E_T$.

2.1.2 Muon Corrections in CMSSW 3_1_X

In CMSSW 3_1_X the muon $\not\!\!E_T$ correction described in Sec. 2.1.1 was improved in two ways:

1. The muon selection is impoved
2. Tools are provided in order to add flexibility

The need for an improved muon selection is driven by the observation that in very high $\hat{P}_T$ QCD events the correction of of Sec. 2.1.1 actually degrades the overall $\not\!\!E_T$ performance. In Fig. 3 we show that in such a sample the correction makes the $\not\!\!E_T$ worse unless the muon is real, *i.e.*, truth-matched to a GEN muon from b or c-quark decay. Note that only 5% of $P_T > 10$ GeV global muons in such a sample are real. For this reason the default selection of muons for $\not\!\!E_T$ correction in CMSSW 3_x is tightened to

- $P_T > 10$ GeV,
- $\chi^2/ndof$ of global fit < 10,
- number of valid hits of silicon fit > 10,
- impact parameter of silicon hit, corrected for beam spot < 2 mm,

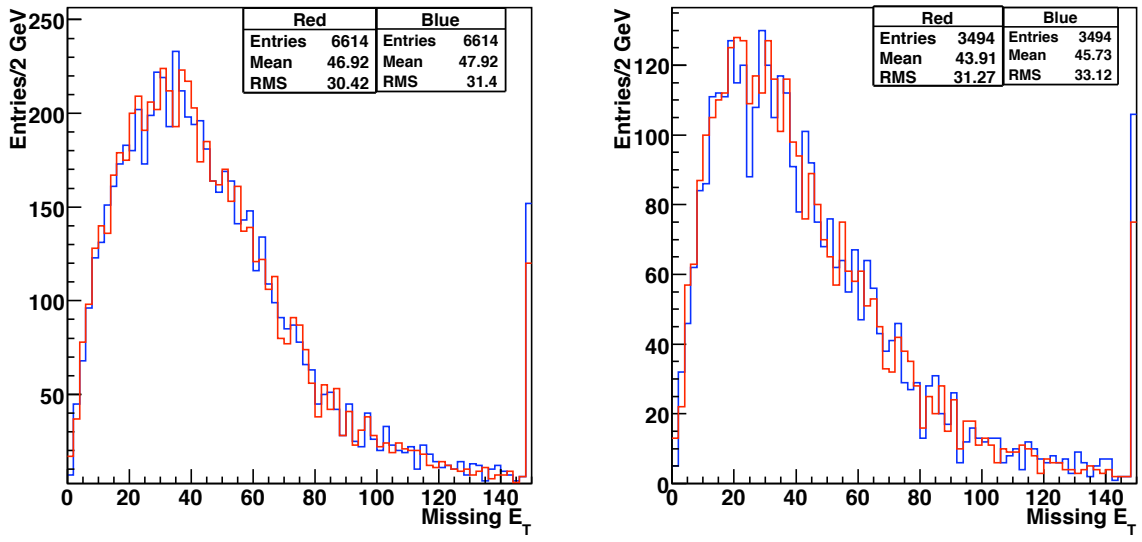


Figure 2: In red is plotted the $\not\!\!E_T$ using the average correction, i.e. we make a distinction between isolated and non-isolated muons. In blue is plotted the $\not\!\!E_T$ without the average correction, so both isolated and non-isolated muons are treated identically. On the left is the $\not\!\!E_T$ for all events which have at least one muon in the *goodMuonsforMETCorrection* collection. On the right is plotted the MET for all events with at least one non-isolated muon. The last bin is the overflow bin. Again, the shapes of the distributions are similar but when the average correction is not used, the tails of the $\not\!\!E_T$ distribution extend out further.

- muon must, in addition to being a global muon, also be reconstructed as a tracker muon.

These requirements, except for the last one, originate from the muon identification analysis note [2]; the last requirement is essentially 100% efficient for real muons. The combined efficiency of these requirements is about 93%. In addition, if the difference between the P_T from the global and silicon fits is more than 20% of the P_T of the silicon fit, the correction is performed using the P_T from the global fit. This minimizes the effects of non-Gaussian tails in the global muon fit.

With this improved muon selection, the muon $\not\!\!E_T$ correction improves the $\not\!\!E_T$ tails in QCD events. Moreover, events with fake muons still passing the selection described above do not degrade the $\not\!\!E_T$ resolution. This is illustrated in Fig. 4.

The tightening of the requirements on muons for $\not\!\!E_T$ correction reduces the tails in QCD events due to fake muons. At the same time it introduces problems for analyses like $Z \rightarrow \mu\mu +\not\!\!E_T$, where the muon requirements for Z reconstruction may be looser than the ones described above. In these cases not all the muons from the Z will be corrected for, causing unwanted tails in $\not\!\!E_T$. It is not possible to find a solution that perfectly fits all use-cases. For this reason the CMSSW 3_1_X code provides additional information to allow the user to easily change the behavior of the muon $\not\!\!E_T$ correction to fit her needs. The x and y-components of the calorimeter deposits associated with each muon are stored in the event (using an `edm::ValueMap`), as well as a flag to tell the user whether the $\not\!\!E_T$ was corrected for the muon, and what fit (global, silicon, or standalone) was used for the correction. A second value map is also provided to allow the same flexibility for tcMET (see later). In this case the value map contains the x and y-

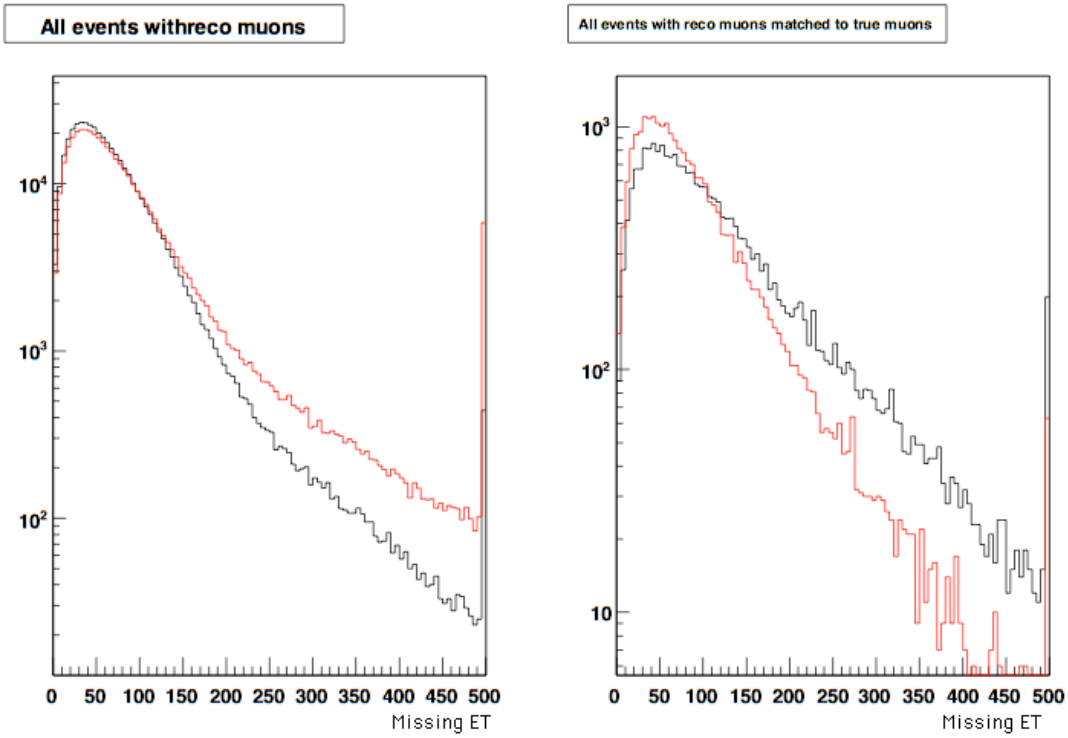


Figure 3: $\not\!\!E_T$ in QCD Pythia MC events with $\hat{P}_T > 800$ GeV for events with a global muon of $P_T > 10$ GeV. Left plot: $\not\!\!E_T$ including (red) and excluding (black) the CMSSW 2_2_x $\not\!\!E_T$ muon correction; Right plot: same as the left plot but only for event where the leading global muon is truth matched to a GEN muon from b or c-quark decay. The last bin is the overflow bin.

components of the expected energy deposition for a pion of the same momentum as the muon in question, and the flag also contains information on whether tcMET was calculated treating the muon as a pion.

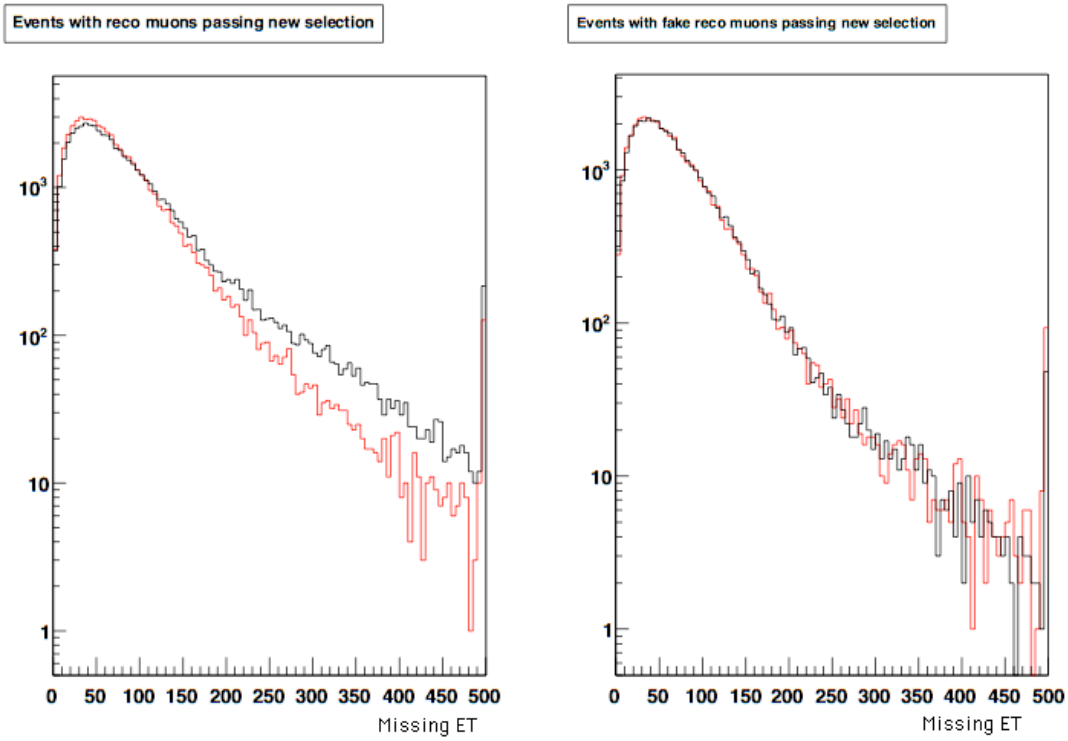


Figure 4: $\not\!\!E_T$ in QCD Pythia MC events with $\hat{P}_T > 800$ GeV for events with a global muon of $P_T > 10$ GeV passing the CMSSW 3_x selection. Left plot: $\not\!\!E_T$ including (red) and excluding (black) the $\not\!\!E_T$ muon correction; Right plot: same as the left plot but only for event where the leading global muon is not truth matched to a GEN muon from b or c-quark decay. The last bin is the overflow bin. In this case the $\not\!\!E_T$ muon correction differs from the CMSSW 3_x correction described in the text in that the global muon fit is always used and the muon energy deposition in the calorimeter is taken as a constant 3 GeV.

2.2 Correction to $\not\!\!E_T$ due to Hadronically Decaying Tau Leptons

$\not\!\!E_T$ corrections due to mismeasurements of hadronically decaying taus incorporates the effects associated with the non-compensating nature of the CMS calorimeter and magnetic field effects. Hadronically decaying tau jets generally have low particle multiplicity and fairly energetic products, while standard jets of the same energy typically have higher multiplicity with a large fraction of the energy being composed of low energy particles. These differences make $\not\!\!E_T$ corrections for events with taus different than the case of QCD jets.

Theoretically, the correction can be established by defining a region in the calorimeter that is large enough to include energy deposits of all decay products of a tau and calculating the correction by summing up the energy deposition in that region (using the same thresholds used to calculate the raw MET), removing contributions from underlying events and pile-up, and replacing the remainig energy deposition with the "true" visible tau transverse energy.

$$\Delta\vec{\not\!\!E}_T = \sum_{reg}\vec{E}_T^{cal} - \vec{E}_T^{\tau} - \sum_{reg}\vec{E}_T^{UE} - \sum_{reg}\vec{E}_T^{PU} \tag{2}$$

Defining the size of our particular region of interest is straight forward for the case of taus. Tau isolation criteria, which is the main discriminator against QCD jets, essentially requires 0 particles between a signal region (roughly 10 degrees) and an outer cone (roughly 30 degrees). Therefore, a region of size $\Delta R < 0.5$ typically used for jet reconstruction is a safe choice. The second term ("true" visible energy) can be approximated by the best available measurement of the tau, which at the moment involves the reconstruction of tau jets using particle flow candidates. It has been shown in [1] that tau jets reconstructed using particle flow candidates provide better measurements than both standard uncorrected calorimeter jets and corrected calorimeter jets. We define the correction for taus as:

$$\Delta\vec{\not\!\!E}_T = \sum\vec{E}_T^{cal\ jet\ 0.5} - \vec{E}_T^{PF\ \tau}, \tag{3}$$

Because we have maintained the "factorized" approach to correcting the $\not\!\!E_T$, we assume that corrections due to underlying events and pile-up will be applied elsewhere (e.g. Type II corrections). Furthermore, the minimal correction defined above is accurate in the low luminosity regime, where effects from UE/PU are small.

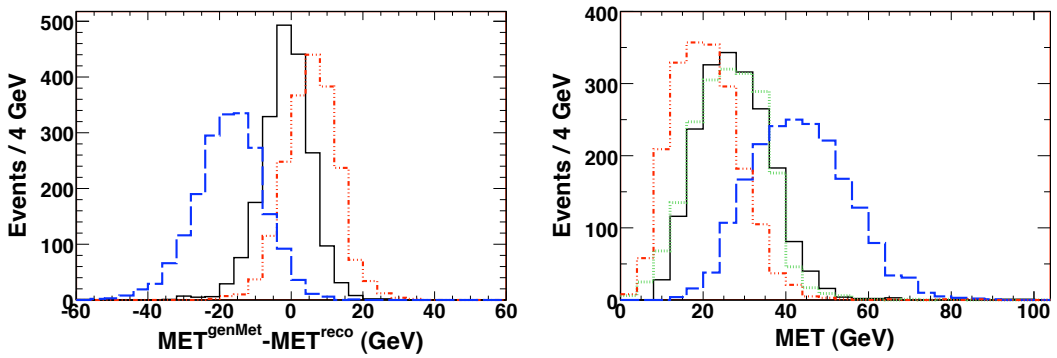


Figure 5: On the left:$(\not\!\!E_T{}^{true} - \not\!\!E_T{}^{reco})$ distribution for the case of no correction (red, dashed-dotted), standard jet correction (blue, dashed) and tau correction (black, solid). On the right: $\not\!\!E_T$ distribution for "true" generator level $\not\!\!E_T$ (green, dotted), raw uncorrected $\not\!\!E_T$ (red, dashed-dotted), standard jet based correction (blue, long-dash), and tau based correction (black, solid).

To validate the performance of the correction for taus and remove effects associated with mismeasurements of recoil jet energies, we selected a sample of $W \to \tau\nu$ events with the requirement that there be 0 extra jets with $E_T > 5$ GeV. Figure 5(a) shows $(\not\!\!E_T{}^{true} - \not\!\!E_T{}^{reco})$ distribution

for three cases: no correction (red, dashed-dotted), standard jet correction (blue, dashed) and tau (black, solid) correction. Figure 5(b) shows the $\not\!\!E_T$ distribution for four cases: "true" generator level $\not\!\!E_T$ (green, dotted), raw uncorrected $\not\!\!E_T$ (red, dashed-dotted), standard jet based correction (blue, long-dash), and tau based correction (black, solid). The figures were produced using Summer08 samples, the most current JetMETCorrections/Type1MET subsystem, and CMSSW_2_2_X. One can see from both figures that the correction defined above provides correct scale and best resolution. Furthermore, if standard jet corrections are used for events containing hadronically decaying taus, lower track multiplicity and more energetic tracks leads to a significant over-correction of $\not\!\!E_T$.

The applicability of the tau based corrections has been discussed in [1]. It is sufficient to note that tau corrections improve MET scale and resolution only in events that have real taus and only if these taus were identified as such. The limits of applicability of the tau corrections can vary significantly per analysis. Furthermore, the correct applicability can vary on an event by event basis. To take this into account, the current implementation in CMSSW is based on the code (TauMET.cc) that calculates the $\not\!\!E_T$ correction for each calorimeter passing a set of user defined criteria. The default use assumes as input CaloJets, PFTaus, and Type I jet corrected $\not\!\!E_T$. The "factorized" approach to the Type I corrections assumes that standard jet corrections are applied first. To compensate for real tau jets being corrected as standard jets, the current implementation of the code uncorrects the $\not\!\!E_T$ for those taus that have been corrected using Type I jet based corrections and then corrects them using the proper tau corrections defined above. Since tau tagging methods vary for each analysis, the current implementation provides the user with the option of choosing his/her own set of tau tagging criteria. Furthermore, since the corrections for taus can even vary on an event by even basis, the user can choose to produce as output (1) a METCollection of LorentzVectors (with the same indices as the PFTaus) containing the correction per tau jet or (2) a CaloMETCollection containing the corrected $\not\!\!E_T$ where each event has been corrected based on the set of user defined tau tagging criteria. An example of running the TauMET code can be found in: /JetMETCorrection/Type1MET/test .

2.3 Type II $\not\!\!E_T$ Corrections

Type II $\not\!\!E_T$ corrections take account of unclustered energy response, and correct out-of-cone energy for every jets, to remove biases in $\not\!\!E_T$ scale. Type II corrections are done on top of Type I corrections, according to the formula,

$$\vec{\not\!\!E}_T^{\text{ Type II}} = \vec{\not\!\!E}_T^{\text{ Type I}} + c \times \sum_j [\vec{p_{T_j}}^{\text{IC7}} - \vec{p_{T_j}}^{\text{IC5}}], \tag{4}$$

where $\vec{p_{T_j}}^{\text{IC7}}$ and $\vec{p_{T_j}}^{\text{IC5}}$ are the transverse momentums of reconstructed jet, by iterative cone algorithm of cone size 0.7 and 0.5 respectively. The sum j runs over all jets with $p_{T_j}^{\text{IC5}}$ greater than a threshold, which can be used, together with a constant c, for optimization. In this studies, the threshold is set to 20 GeV, which is a ad hoc choice to match the default threshold for type I $\not\!\!E_T$ corrections and may not be the optimal value.

Type I $\not\!\!E_T$ corrections are applied using the CMSSW default configuration
`JetMETCorrections/Type1MET/MetType1Corrections_cff`
with default jetPTthreshold 20 GeV and jetEMFracLimit 0.9.

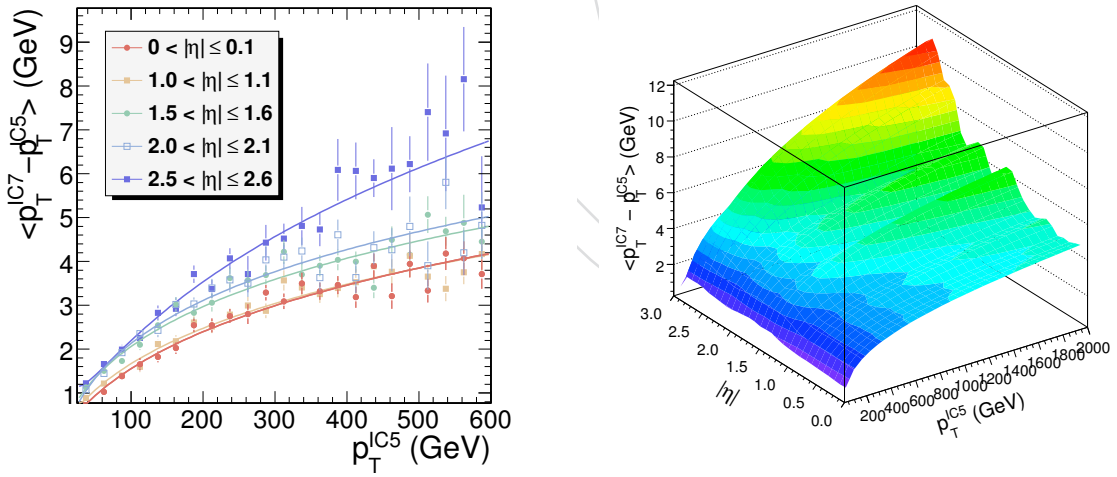
A correction function was obtained by fitting the mean energy of $< p_T^{\text{IC7}} - p_T^{\text{IC5}} >$, in bins of p_T^{IC5} and η (Fig. 6(a)). This function is interpreted as the mean energy leaking out of a 0.5 iterative cone. A set of QCD samples, RelValQCD CMSSW_2_2_3, with redoing 0.7/0.5 iterated cone jet reconstruction, was used in the correction function parameterization. Further constraints were

put to require all jets are not overlap, and well aligned in jet axis, i.e. $\Delta R < 0.1$ for matching between 0.7 and 0.5 iterative cone algorithms. A quadratic function in $\log p_T^{IC5}$ was fitted for every η bins. The correction function was extrapolated to high p_T up to 2000 GeV (Fig. 6(b)).

Both type I and type II $\not\!\!E_T$ corrections were applied to events with true $\not\!\!E_T$ using Summer08 $W(e\nu)$ + jets samples. The default thresholds, as discussed above, were used. Furthermore, all electrons were removed from the jet list and not corrected as jet, neither by type I nor II corrections. The parallel component of $\not\!\!E_T$ ($\not\!\!E_\parallel$), defined as the projection of reconstructed $\not\!\!E_T$ on the direction of "true" $\not\!\!E_T$ in generator level, gives the magnitude of measured $\not\!\!E_T$ on the same direction as neutrino's transverse momentum produced by W decay.

Similar to type I corrections, type II corrections remove bias significantly with reference to raw $\not\!\!E_T$. It performs slightly better than type I corrections alone (Fig. 7(a)). However, the corrections do not improve the $\not\!\!E_\parallel$absolute resolution $\sigma(\not\!\!E_\parallel)$, as the energy scale changed (Fig. 7(b). A significant improvement, as shown in Fig. 7(c), is achieved in relative resolution $\sigma(\not\!\!E_\parallel)/\not\!\!E_\parallel$. Type I/II corrections removes bias in $\not\!\!E_\parallel$and gives a better measured $\not\!\!E_\parallel$value. The increase in denominator $\not\!\!E_\parallel$, compared to no correction, improves the relative resolution.

The benefit of type II corrections, with respect to type I corrections, is to further remove bias (Fig. 7(a)) without any penalties of absolute/relative resolution (Figs. 7(b),7(c)). Again, the number shown is not optimal. All factors, such as the performance of type I $\not\!\!E_T$ corrections, parameterization of energy leaking, jet energy thresholds etc, should be consider in full optimization.



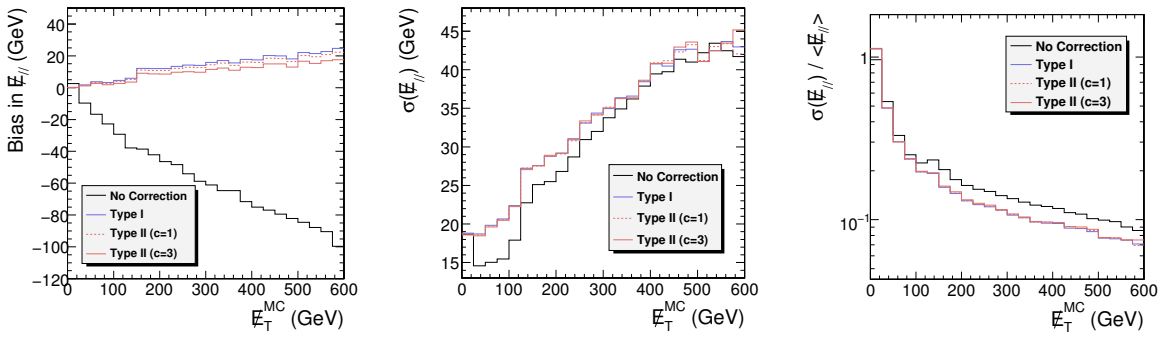
(a) Fitting to quadratic function of $\log p_T^{\mathrm{IC5}}$.

(b) Two dimensional parameterization.

Figure 6: Parameterization of energy leaking out of a 0.5 iterative cone as a function of jet energy p_T^{IC5} and $|\eta|$ using QCD samples.

3 Track-Corrected $\not\!\!E_T$

A detailed understanding of the standard model, especially the tails, is a precondition to any search for new physics. The details of any new physics drives the requirement on how well $\not\!\!E_T$ must be understood. However, if new physics exists close to the standard model, reducing



(a) Bias in $\not\!\!E_{\parallel}$. (b) Absolute resolution of $\not\!\!E_{\parallel}$. (c) Relative resolution of $\not\!\!E_{\parallel}$.

Figure 7: Comparison for raw $\not\!\!E_T$ and type I/II corrections in the $W(e\nu)$ + jets samples.

the tails of MET distributions will be vital. Track-corrected MET (tcMET) was developed with these scenarios in mind. The goal at the outset was to develop a robust algorithm to reduce the non-Gaussian tails of (fake) MET distributions. Robustness requires that the algorithm is independent of calorimetry specifics and stable with respect to varying (startup) conditions. The algorithm should also be independent of details of geometry and magnetic field and run on AOD so that it is usable at all data tiers.

Track-corrected MET has been developed to take advantage of the excellent resolution of the tracker compared to the calorimeter at energy scales $\mathcal{O}(1\,\text{GeV})$. The basic idea of tcMET is to compute MET by replacing the expected energy deposition of good tracks (assumed to be pions) with the corresponding momentum. Any attempt to correct $\not\!\!E_T$ using tracker information must take care to avoid "double-counting" by including energy from both the calorimeter and the tracker. Various approaches to this problem have been developed - tcMET addresses this by removing the expected energy, $\langle E \rangle$, of charged hadron deposits in the calorimeter. The expected deposition $\langle E \rangle$ is determined from a response function (RF) derived from a single pion monte carlo sample. This is one of the key differences between this approach and PFlow; tcMet relies on the *expected* deposition rather than the *actual* energy contained in a calorimeter cluster. Further details on the response function and its derivation can be found in [3].

Track-corrected MET is calculated using the CaloMET, muons, electrons, tracks, and the response function.

- Muons are minimum ionizing particles that leave little energy in the calorimeter. Correcting muons as pions using the RF results in under-correction and generates fake $\not\!\!E_T$. Instead, tcMET makes use of the standard muon corrections that have been discussed earlier in this note.
- Electrons are particles that deposit most of their energy in the ECAL - treating electrons as pions generates fake $\not\!\!E_T$ due to over-correction. To account for this, tracks matched to electrons are removed from the set of correctable tracks.
- The remaining good tracks in the event are corrected for using:

$$\text{tcMET} = \text{baseline } \not\!\!E_T + \sum_{\substack{\text{good} \\ \text{tracks}}} \langle \vec{E}_T \rangle - \sum_{\substack{\text{good} \\ \text{tracks}}} \vec{p}_T, \tag{5}$$

  where baseline $\not\!\!E_T$ is CaloMET corrected for muons in the event.

It is important to note that the correction for each good track involves two sets of coordinates:

the intersection of the track at the calorimeter face and the direction of the track at the primary vertex. The expected energy deposition for each track is removed from the calorimeter. This location is determined using a simple helical propagator. The track momentum that replaces it is measured at the vertex. To be explicit, the correction for a single component of $\not\!\!E_T$ takes the form:

$$(\text{tcMET}_x) = (\text{baseline } \not\!\!E_{T\ x}) + \sum_{\substack{\text{good}\\ \text{tracks}}} \langle E \rangle \sin\theta_c \cos\phi_c - \sum_{\substack{\text{good}\\ \text{tracks}}} p_T \cos\phi_v, \tag{6}$$

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.

Selection of good tracks is important for robustness and performance. Badly measured tracks can generate fake $\not\!\!E_T$ and correcting for such a track can make things 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 $\not\!\!E_T$ if the track is passed to the response function.

To test the performance of this algorithm, we evaluated how tcMET reduces the tail of the $\not\!\!E_T$ distribution for events with fake MET and accurately reconstructs the magnitude and direction of real $\not\!\!E_T$. Figure 8 shows the $\not\!\!E_T$ and $\Delta\not\!\!E_T$ distributions for events with fake and real MET, respectively. The left histogram displays the characteristic non-Gaussian tail of the Drell-Yan $\not\!\!E_T$ distribution. Correcting for tracks reduces the number of events above 30(50) GeV by a factor of 3.4(6.8) compared to correcting for muons only. The right histogram shows the difference between true MET and demonstrating and improvement in MET resolution of approximately 30%. A Z+jets sample was used to study fake $\not\!\!E_T$ while $t\bar{t}$ provided a standard model leptons+jets+MET signature. Comparisons in both cases were done for dilepton final states where the leptons were required to have $p_T > 20$ GeV, and pass ID and isolation cuts .

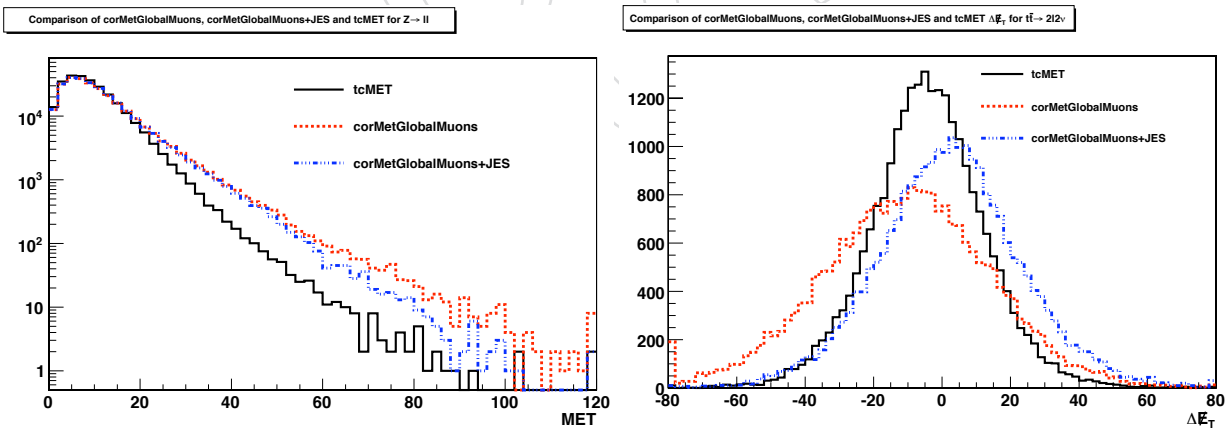


Figure 8: On the left is the MET distribution for Drell-Yan events with fake $\not\!\!E_T$. The number of events with MET$>$ 30(50) GeV is reduced by a factor of 3.4(6.8) compared to correcting for muons only. Compared to correcting MET for both muons and the jet energy scale(JES), reduction is by a factor of 3.0(4.3) for the same cuts. On the right is the distribution of $\Delta\not\!\!E_T$ = $\not\!\!E_T{}^{reco}$ - $\not\!\!E_T{}^{true}$ for $t\bar{t}$ events with real $\not\!\!E_T$. Correcting for tracks improves MET resolution by nearly 30% compared to correcting for muons only and by approximately 20% compared to muon+JES corrections.

The left plot of Fig. 9 shows the MET resolution for $t\bar{t}$ events as a function of the true $\not\!\!E_T$.

Here, true MET is calculated as the vector sum of neutrinos in the transverse plane. Except at the lowest values of true $\not\!\!E_T$ (below 10 GeV), tcMET significantly improves MET resolution compared to correcting for muons only. Above 20 GeV, tcMET agrees with true $\not\!\!E_T$ to better than 10%. No attempt was made to optimize performance at high $\not\!\!E_T$.

In addition to the magnitude of $\not\!\!E_T$, it is also interesting to compare the direction of the MET vector compared to that expected for events with real $\not\!\!E_T$. The right plot of Fig. 9 shows $\Delta\phi$ between true MET and corrected $\not\!\!E_T$ after correction for muons only, muon+JES corrections, and tcMET. Correcting for tracks improves determination of the $\not\!\!E_T$ direction by nearly 30% compared to correcting for muons only and by approximately 15% compared to correcting for muons and the JES.

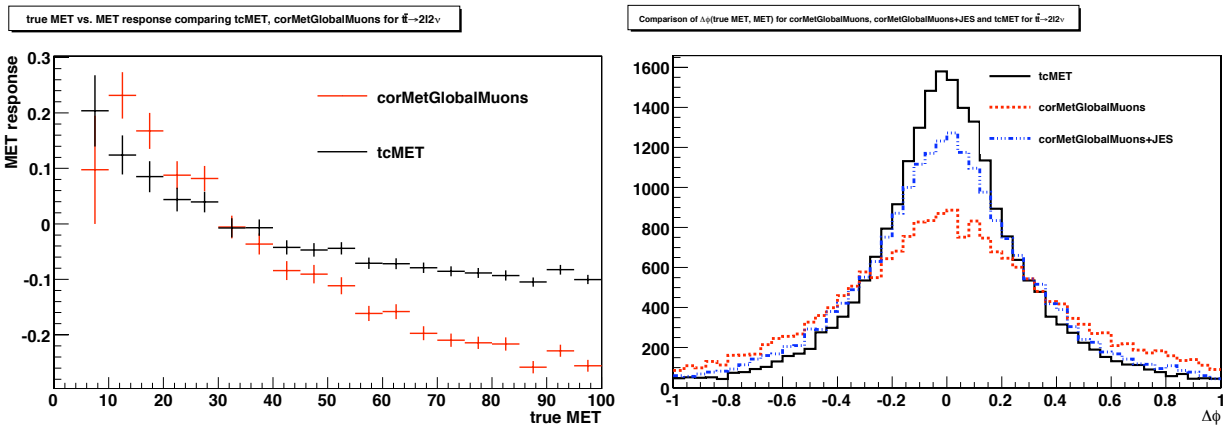


Figure 9: Left is $\not\!\!E_T$ resolution as a function of the true MET for $t\bar{t}$ events with real $\not\!\!E_T$. Except at the lowest values of true $\not\!\!E_T$, tcMET provides better resolution than correcting for muons only. Above 20 GeV, track-corrected $\not\!\!E_T$ agress with true MET to better than 10%. The right plot shows $\Delta\phi$ between corrected $\not\!\!E_T$ and true MET for the various correction schemes. Compared to correcting for muons only, tcMET provides nearly 30% better direction determination. This improvement is approximately 15% compared to correcting for both muons and the JES.

Correcting for tracks using tcMET reduces fake MET and improves the measurement of real $\not\!\!E_T$. The response function implemented here was derived using CMSSW_1_6_11. Its performance is an indication of the robustness of the algorithm, as it demonstrates that tcMET is not particularly sensitive to the details of the calorimetry simulation.

Track-corrected MET is available as external tags beginning with the 2_2 CMSSW releases. A tcMET collection will be available as part of the AOD beginning in the 3 series. Further refinements to tcMET continue to be investigated.

4 $\not\!\!E_T$ from Particle Flow

The particle-flow reconstruction algorithm aims to provide a global (i.e., complete and unique) event description at the level of individually reconstructed particles, with an optimal combination of information across all CMS sub-detectors. The reconstructed and identified individual particle list includes muons, electrons (with individual Bremsstrahlung photons), photons (either unconverted or converted), charged hadrons (without or with a nuclear interaction in the tracker material), as well as stable and unstable neutral hadrons. These particles are not necessarily isolated, and may even originate from an intricate overlap of reconstructed charged particles, ECAL and HCAL energy clusters, and hits in the muon chambers.

Determining $\not\!\!E_T$ using particle-flow is in principle simple: one merely forms the transverse vector sum over all reconstructed particles in the event resulting in an azimuthal, momentum two-vector. In this sense, $\not\!\!E_T$ from particle flow is analogous to $\not\!\!E_T$ determined at generator level via all visible "Monte Carlo truth" particles, except that the particles used to determine particle-flow $\not\!\!E_T$ are reconstructed with some efficiency and fake rate, and possess a finite momentum resolution. Hence, the particle-flow algorithm is fully general and neither specific nor tuned to $\not\!\!E_T$ determination. Consequently, it is important to note that the $\not\!\!E_T$ physics-object, determined directly from particles reconstructed via the general particle-flow algorithm, has to be calibrated (even if residually) and verified using a data-driven manner analogous to the calorimeter-based MET determination.

While the basics of the particle-flow algorithm are finished, many details remain work in progress. The full implementation in CMS will be described in a forthcoming paper; only the essential features are presented here. The particle-flow algorithm starts with reconstruction performed independently within each CMS sub-detector: clustering is conducted separately in each of the electromagnetic (ECAL) and hadronic (HCAL) calorimeters and track reconstruction takes place in the combined silicon and pixel tracker system.

Links between ECAL clusters, HCAL clusters, and tracks are made, depending on the spatial and energy compatibility between the clusters and/or tracks. The particle-flow algorithm proceeds in stages, associating clusters and tracks with a newly reconstructed particle at each stage. First, tracks and clusters identified as being associated with hits and segments in the muon chambers are tagged as muons and removed from the list of unassociated objects. Next, tracks and clusters identified as being associated with electrons (and all the possible individual Bremsstrahlung photons) are tagged and removed from further processing.[1]

Following that, in the case of an HCAL cluster linked to a track, the calibrated HCAL cluster energy is compared with the track momentum. If the cluster energy is compatible with the track momentum, a charged hadron is created with energy determined from a weighted average of the track momentum and the cluster energy. If the difference between the cluster energy and the track momentum is significant (with respect to the expected calorimeter energy resolution and the measured track momentum uncertainty), a neutral hadron is created out of the excess cluster energy. In the case where an ECAL cluster and an HCAL cluster are linked together with a track, the calibrated combined energy of the ECAL cluster and HCAL cluster is compared with the track momentum. If the combined ECAL and HCAL calorimeter energy is compatible with the track momentum, a charged hadron is created as explained above; otherwise either a neutral hadron or a photon is created out of the excess calorimeter energy.

After removing the above tracks and clusters from the list of unassociated objects, only clusters not linked to any track remain uncleaned from the event. Any such ECAL clusters are assumed to be photons and any such HCAL clusters (or HCAL clusters linked with ECAL clusters) are assumed to be neutral hadrons. It is important to note that the entire event is partitioned into separate particles; that is, with the exception of inefficient or fake particles, there is no undercounting or double counting of energy in the event.

The performance of $\not\!\!E_T$ from particle flow (both uncorrected and corrected for the particle-flow jet energy scale) was studied using `CMSSW_2_2_3` and compared with $\not\!\!E_T$ from calorimeter towers (corrected for both the jet energy scale of calorimeter jets and the possible existence of muons). All jet energy corrections (particle-flow and calorimeter) were obtained using the

[1] While particles corresponding to electrons are included in the list of particles reconstructed by the particle-flow algorithm, the electron identification was not yet finalized for CMSSW version 2_2_3, and so the algorithm does not yet fully distinguish electrons from pions.

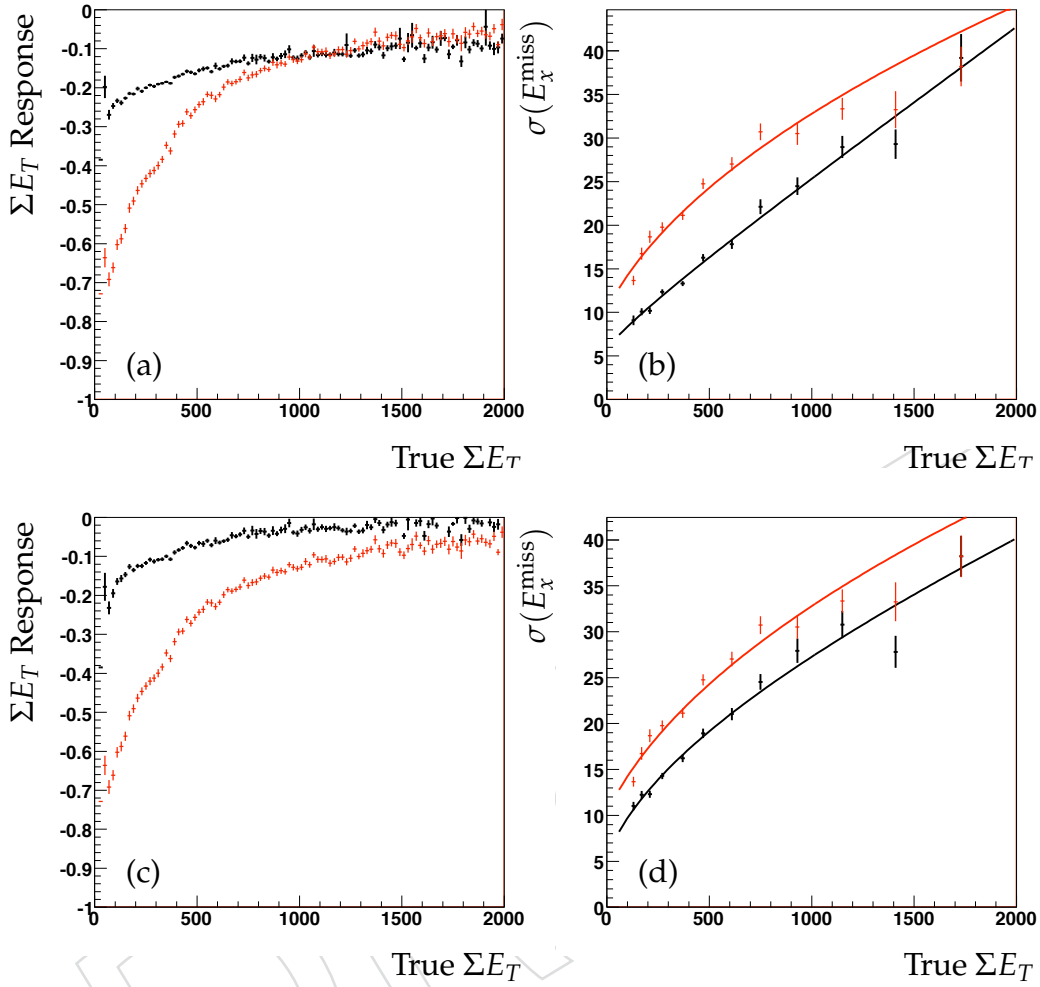


Figure 10: Fully simulated QCD dijet events using `CMSSW_2_2_3`. (a) Response of the total visible transverse energy of the event, defined to be $(\Sigma E_T^{\text{reco}} - \Sigma E_T^{\text{true}}])/\Sigma E_T^{\text{true}}$, as a function of the "true" total visible transverse energy of the event: upper (bold black) points represent "raw" particle-flow reconstruction (uncorrected for the particle-flow jet energy scale), whilst the lower (red) points represent calorimeter based reconstruction (corrected for the calorimeter jet energy scale); (b) Resolution of the x-projection of $\not\!\!E_T$ versus the total visible transverse energy of the event: the lower (bold black) curve represents "raw" particle-flow reconstruction (uncorrected for particle-flow jet energy scale), whilst the upper (red) curve represents calorimeter-tower reconstruction (corrected for the calorimeter-tower jet energy scale); For illustration purposes, (c), (d) are the same as (a), (b) except that corrections for the particle-flow jet energy corrections (taken from `L2L3Corrections_Summer08Redigi`) are applied to ΣE_T and $\not\!\!E_T$ determined from particle-flow (bold black points and curve) – the calorimeter-only based ΣE_T and $\not\!\!E_T$ (red points and curve) are unchanged with respect to (a) and (b).

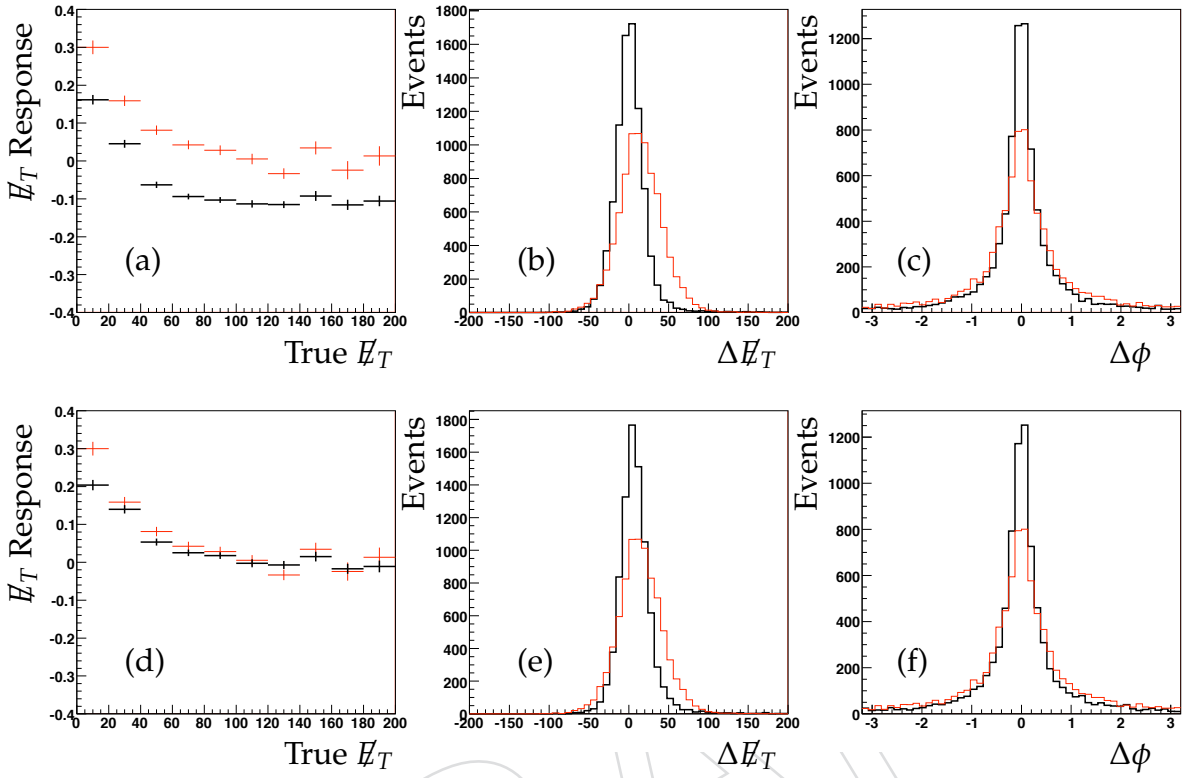


Figure 11: Fully simulated $t\bar{t}$ events using `CMSSW_2_2_3` and requiring that there be at least 5 GeV of "true" missing transverse energy. (a) Response of the missing transverse energy of the event, defined to be $(\not\!\!E_T{}^{\text{reco}} - \not\!\!E_T{}^{\text{true}})/\not\!\!E_T{}^{\text{true}}$, as a function of the "true" missing transverse energy of the event: the lower (bold black) points represent "raw" particle-flow based reconstruction (uncorrected for the particle-flow jet energy scale), whilst the upper (red) points represent calorimeter based reconstruction (corrected for both the jet energy scale of calorimeter-tower jets and the possible existence of muons); (b) $\not\!\!E_T$ resolution, $\not\!\!E_T{}^{\text{reco}} - \not\!\!E_T{}^{\text{true}}$, accounting for muons and detector acceptance: the narrow (bold black) histogram represents "raw" particle-flow based reconstruction (uncorrected for the particle-flow jet energy scale), whilst the shorter (red) histogram represents calorimeter reconstruction (corrected for both the jet energy scale of calorimeter-tower jets and the possible existence of muons); (c) $\not\!\!E_T$ angular resolution, $\phi^{\text{reco}} - \phi^{\text{true}}$: the narrow (bold black) histogram represents "raw" particle-flow reconstruction (uncorrected for the particle-flow jet energy scale), whilst the shorter (red) histogram represents calorimeter reconstruction (corrected for both the jet energy scale of calorimeter-tower jets and the possible existence of muons); For illustration purposes (d), (e), (f) are the same as (a), (b), (c) except that $\not\!\!E_T$ from particle-flow (bold black) has been corrected for the particle-flow jet energy scale (taken from `L2L3Corrections_Summer08Redigi`) – the calorimeter-only based $\not\!\!E_T$ (red) are unchanged with respect to (a), (b), (c).

L2L3Corrections_Summer08Redigi corrections (relative energy response across η and absolute energy scale as a function of $_T$) derived for the "Reprocessed Summer08" data samples.

Figures 10 (a) and (c) show that the particle-flow algorithm is able to measure the total visible transverse energy (ΣE_T) in QCD dijet events very well, compared with calorimeter-tower based reconstruction, particularly at low values of ΣE_T and, when corrected for the particle-flow jet energy scale, performs better than calorimeter-tower based reconstruction even at higher values of ΣE_T. Figures 10 (b) and (d) show that at low ΣE_T the particle-flow algorithm is also able to measure zero missing transverse energy in QCD dijet events very well, compared with calorimeter based $\not\!\!E_T$. It is interesting to note that jet energy scale corrections to $\not\!\!E_T$ from particle flow tend to worsen the $\not\!\!E_T$ resolution, particularly for low values of ΣE_T. This may be an indication that particle-flow jets could benefit from a more sophisticated energy scale correction, based on jet content such as the neutral hadronic fraction, than is currently done. A fit to both resolution curves in Figure 10 (b) gives:

$$\begin{aligned}
\sigma^2_{\mathrm{PF}}(E_x^{\mathrm{miss}}) &= \Big(10.6\ \mathrm{GeV}\Big)^2 + \Big(66.1\%\sqrt{\textstyle\sum \mathrm{E_T} - 207.6\ \mathrm{GeV}}\Big)^2 + \Big(1.7\%(\textstyle\sum \mathrm{E_T} - 207.6\ \mathrm{GeV})\Big)^2 \\
\sigma^2_{\mathrm{Cal}}(E_x^{\mathrm{miss}}) &= \Big(15.5\ \mathrm{GeV}\Big)^2 + \Big(98.5\%\sqrt{\textstyle\sum \mathrm{E_T} - 138.2\ \mathrm{GeV}}\Big)^2 + \Big(0.0\%(\textstyle\sum \mathrm{E_T} - 138.2\ \mathrm{GeV})\Big)^2
\end{aligned}$$

and, at $\Sigma E_T^{\mathrm{true}} = 500$ GeV, shows a roughly 35% improvement in the absolute resolution for particle-flow based, $\not\!\!E_T$ compared with calorimeter-based $\not\!\!E_T$ for events where no $\not\!\!E_T$ is expected. Larger values of ΣE_T have not yet been studied, however for very large values of ΣE_T, the performance of $\not\!\!E_T$ from particle-flow is designed to converge to the calorimeter-based $\not\!\!E_T$; this will be verified in future studies.

Figure 11 demonstrates the case of real missing transverse energy in $t\bar{t}$ events. Specifically, Figures 11 (a) and (d) show the effect of applying the particle-flow jet energy scale corrections to $\not\!\!E_T$ from particle flow. Without the energy scale corrections, $\not\!\!E_T$ from particle flow is currently measured to be roughly 10% too low; with the energy scale corrections, the response of particle-flow $\not\!\!E_T$ is seen to be very similar to calorimeter-tower $\not\!\!E_T$ (corrected for the calorimeter jet energy scale) and is unbiased above approximately 50 GeV. Additional developments in the particle-flow algorithm (not available in CMSSW_2_2_3) already demonstrate in improved response to jets and so further improvements in the particle-flow $\not\!\!E_T$ response are expected.

Figures 11 (b) and (c) show that a significant improvement in the energy and angular resolution using $\not\!\!E_T$ from particle flow compared with MET from calorimeter towers. A gaussian fit to the histograms in Figure 11 (b) indicate that "raw" particle-flow $\not\!\!E_T$ provides more than 35% improvement in absolute resolution compared with calorimeter-tower $\not\!\!E_T$ (corrected for jet energy scale and muons) in $t\bar{t}$ events containing real $\not\!\!E_T$ (σ = fitted width):

$$\begin{aligned}
\sigma_{\mathrm{PF}}(\not\!\!E_T\ ) &= (16.8 \pm 0.2)\ \mathrm{GeV} \\
\sigma_{\mathrm{Cal}}(\not\!\!E_T\ ) &= (26.8 \pm 0.2)\ \mathrm{GeV}
\end{aligned}$$

with less overall bias (μ = fitted mean):

$$\begin{aligned}
\mu_{\mathrm{PF}}(\not\!\!E_T\ ) &= (0.3 \pm 0.2)\ \mathrm{GeV} \\
\mu_{\mathrm{Cal}}(\not\!\!E_T\ ) &= (14.1 \pm 0.3)\ \mathrm{GeV}
\end{aligned}$$

The observed "zero" bias in the case of the "raw" (uncorrected) particle-flow $\not\!\!E_T$ can be seen in Figure 11 (a) to be due to an accidental cancellation of a roughly 10% too high response below 30 GeV and a 10% too low response above 30 GeV. Corrections to particle-flow $\not\!\!E_T$ for

the jet energy scale currently shift the fitted mean in Figure 11 (e) to larger values ($\mu_{\rm PF} = 6$ GeV), but still significantly less than the calorimeter-based $\not\!\!E_T$ fitted mean ($\mu_{\rm Cal} = 14$ GeV). Additional developments in the particle-flow algorithm (not available in `CMSSW_2_2_3`) already demonstrate an improved response to jets (both at low as well as at high p_T) and so further improvements an the particle-flow $\not\!\!E_T$ response are expected.

5 Recent changes to $\not\!\!E_T$ Software and $\not\!\!E_T$ in Physics Analysis Toolkit (PAT)

We describe the recent changes to the $\not\!\!E_T$ Software and its standalone Validation Suite. (Note that a common validation suite is being developed in the framework of the Data Quality Monitoring project – see later.) We also describe the $\not\!\!E_T$ oject in the Physics Analysis Toolkit (PAT), its access methods and its default behaviour.

5.1 Recent changes to $\not\!\!E_T$ Software and its Validation Suite

The $\not\!\!E_T$ reconstruction software has recently undergone a phase of rapid growth, both in size and complexity. The number of $\not\!\!E_T$ collections offered in AOD has increased by a factor of six in the course of one year. Before the release of CMSSW_2_0_0, three $\not\!\!E_T$ Collections persisted at the AOD level. With the impending release of CMSSW_3_1_0, this number will be nineteen.

While there have been many new features recently implemented in the MET reconstruction software, the overall infrastructure, which revolves around one centralized EDProducer, has remained largely unchanged. This software paradigm has allowed many developers to make contributions to the $\not\!\!E_T$ reconstruction algorithms in a way that aspires to opimize transparency and backwards compatibility with respect to previous versions of CMSSW.

The centralized EDProducer, which is located in the `RecoMET/METProducers` directory, makes a variety of different flavors of $\not\!\!E_T$. These flavors belong to one of four different datatypes. Each datatype is associated with a set of data members that offers specific information along with the $\not\!\!E_T$ calculation. The most basic datatype is `reco::MET`, which stores six basic $\not\!\!E_T$ related quantities: $\not\!\!E_T$, $\not\!\!E_x$, $\not\!\!E_y$, $\phi(\vec{\not\!\!E}_T)$, ΣE_T, and $\not\!\!E_T$ Significance. All of the flavors of $\not\!\!H_T$, which measure the jet transverse energy imbalance, are the stored to the event as a collection of `reco::MET` objects. A recent addition to the arsenal of `reco::MET` collections in the AOD is the track-corrected $\not\!\!E_T$ ("tcMet"), which is slated to make its debut in CMSSW_3_1_0. As part of this implementation, a novel and sophisticated algorithm has recently been developed which will be invoked by the METProducer during the standard reconstruction sequence to calculate this track-corrected $\not\!\!E_T$. This algorithm operates on several inputs which include tracks, muons, electrons, and the generic calorimeter based $\not\!\!E_T$ ("met"), to produce a more refined MET object [3].

The datatype which currently boasts the most flavors of $\not\!\!E_T$ in the AOD is `reco::CaloMET`. This datatype is used for all $\not\!\!E_T$ calculations that are based on CaloTowers. As such, the `CaloMET` datatype stores quantities that are specific to the calorimetry and relevant for the $\not\!\!E_T$ calculation on top of basic quantities that pertain to `reco::MET`. Some of these quantities include the event Hadronic E_T and Electromagnetic E_T reported from the respective subdetectors within the calorimetry.

In the past, it proved sufficient to provide only one flavor of type `CaloMET`, which consisted of the negative vector sum of all transverse energy components of all calotowers formed in the event. Historically, this flavor has been labeled "met". Starting with CMSSW_2_0_0, new

features were added to the reconstruction algorithms that allowed for $\not\!\!E_T$ to be calculated with only certain subsets of calotowers. A better understanding of the noise profile of the HCAL Forward made it desirable to add a flavor of `CaloMET` to the AOD that excluded HF from the $\not\!\!E_T$ calculation. At this same time optimization studies of CaloTower thresholds also prompted the addition of a $\not\!\!E_T$ calculation derived from a different set of CaloTowers, reconstructed with so called "Optimized" thresholds, aside from the one based on the default "Scheme B" thresholds. More recently, similar concerns of noise contamination led to an additional flavor of `CaloMET` that operated on CaloTowers that did not consider HCAL Outer Barrel deposits to the tower 4-vector. This was implemented in CMSSW_2_1_8, and brings the total number of `CaloMET` type $\not\!\!E_T$ calculations to eight.

A simple datatype, called `reco::GenMET`, is in place to accomodate the generator level $\not\!\!E_T$, which is obviously useful for Monte Carlo studies. Despite the existence of only one "true" $\vec{\not\!\!E}_T$, the GenMET datatype lends itself to several flavors. One flavor, "genMetCalo", serves as the generator level analog of the `CaloMET` flavor "met" and only considers generator particles that would be visible to the calorimetry.

Finally, a new datatype, called `reco::PFMET`, has been introduced to accomodate not only a Particle Flow-based calculation of the $\not\!\!E_T$, but also several other specific quantities produced by the Particle Flow-based algorithms. These include the neutral and charged electromagnetic energy fractions and the muon energy charged fraction to name a few. PFMET has debuted in the AOD in CMSSW_2_2_0, which is a Particle Flow dedicated release. It is expected that other Particle Flow specific quantities related to $\not\!\!E_T$ will be added to the PFMET datatype in the future. A summary of the various flavors of $\not\!\!E_T$ past, present, and future is provided in Table 1.

Table 1: A summary of reconstructed $\not\!\!E_T$ calculatations offered in the AOD

$\not\!\!E_T$ Collection InputTag	Data Type	CMSSW Version	Input to MET Algorithms
met	CaloMET	All Versions	SchemeB CaloTowers w/ HF, w/o HO
metNoHF	CaloMET	$\geq$ 2_0_0	SchemeB CaloTowers w/o HO,HF
metHO	CaloMET	$\geq$ 2_1_8	SchemeB CaloTowers w/ HO,HF
metNoHFHO	CaloMET	$\geq$ 2_1_8	SchemeB CaloTowers w/ HO, w/o HF
metOpt	CaloMET	$\geq$ 2_0_0	Optimized CaloTowers w/ HF, w/o HO
metOptNoHF	CaloMET	$\geq$ 2_0_0	Optimized CaloTowers w/o HO,HF
metOptHO	CaloMET	$\geq$ 2_1_8	Optimized CaloTowers w/ HO,HF
metOptNoHFHO	CaloMET	$\geq$ 2_1_8	Optimized CaloTowers w/ HO, w/o HF
pfMet	PFMET	$\geq$ 2_2_0	PFCandidates
tcMet	MET	$\geq$ 3_0_0	met, tracks, muons, pixelMatchGsfElectrons
htMetIC5	MET	$\geq$ 2_0_0	iterativeCone5CaloJets
htMetKT4	MET	$\geq$ 2_0_0	kt4CaloJets
htMetKT6	MET	$\geq$ 2_0_0	kt6CaloJets
htMetSC5	MET	$\geq$ 2_0_0	sisCone5CaloJets
htMetSC7	MET	$\geq$ 2_0_0	sisCone7Calojets
genMet	GenMET	$\leq$ 3_0_0	GenParticles excluding BSM's, ν's, and μ's
genMetNoNuBSM	GenMET	$\leq$ 3_0_0	GenParticles excluding prompt ν's and μ's
genMetFromIC5GenJets	GenMET	$\geq$ 2_0_0	iterativeCone5GenJets
genMetCalo	GenMET	$\geq$ 3_0_0	same as for genMet
genMetCaloAndNonPrompt	GenMET	$\geq$ 3_0_0	same as for genMetNoNuBSM
genMetTrue	GenMET	$\geq$ 3_0_0	GenParticles excluding BSM's and neutrinos

The health of the reconstruction algorithms is tested and validated regularly by the $\not\!\!E_T$ Validation software, a robust, html-based set of tools that allow for thumbnail access to well over ten thousand histograms. These plots help monitor not only $\not\!\!E_T$ and $\not\!\!E_T$ -related quantities, but all other relevant quantities that come from the subdetectors that are used to reconstruct $\not\!\!E_T$. These constituents include the ECAL Barrel, ECAL Endcaps, HCAL Barrel, HCAL Forward, HCAL Outer Barrel, and HCAL Endcaps, as well as the CaloTowers for which they combine to

form. These histograms are compared between releases and tested for compatibility using the Kolmogorov-Smirnov and χ^2 test statistics. A compatability score less than 10^{-6} marks a given distribution as incompatible. There may be many possible causes for incompatible distributions. Often times, they result from changes in the detector simulation. Less often, they result from bugs in the reconstruction algorithms. In either case, the $\not\!\!E_T$ Validation software proves to be a powerful tool to identify such effects.

An example of the output of the MET Validation software is shown in Figs. 12 (a) and (b), which show a comparison of (a) $\phi(\vec{\not\!\!E}_T)$ and (b) ΣE_T from two separate frontier conditions within the release of CMSSW_2_1_2. Distributions coming from frontier conditions tag IDEAL_V9_v1 are labeled "new" while those coming from tag IDEAL_V6_v1 are labeled "reference". The main difference between these two conditions comes from the new gains implemented in the HCAL. The IDEAL_V9_v1 tag reflects a change in the HF reponse to a value of 0.7 relative to HB/HE, as well as an increase in the HO gains by a factor of 3.5. The $\not\!\!E_T$ Validation software was able to detect an asymmetry in $\phi(\vec{\not\!\!E}_T)$ and a harder spectrum in the scalar transverse energy sum. A descrepancy in the latter distribution was perhaps expected but certainly none was expected in the former.

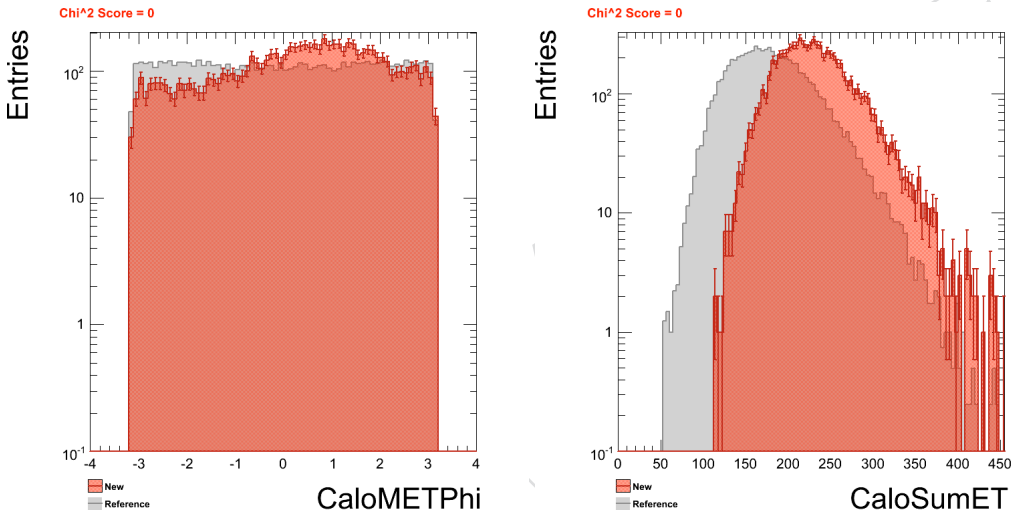


Figure 12: A comparison of (a) $\phi(\vec{\not\!\!E}_T)$ and (b) ΣE_T from two separate frontier conditions within the release of CMSSW_2_1_2.

Another triumph of the MET Validation software is its ability to discern incompatibilties at the subdetector level. This feature is exemplified in Figs. 13 (a) and (b), which show a summary of compatibilty scores for CaloTower occupancy distributions as a function of iη and a particular CaloTower occupancy distribution for iη = 25, repectively. A conspicuous range of η-rings which unanimously fail the χ^2 test for compatibility is clearly seen in Fig. 13(a). This η region coincides perfectly with the physical space spanned by the ECAL endcaps. Here this comparison involves CMSSW_2_1_0_pre9 ("new") versus CMSSSW_2_1_0_pre6 ("reference"). The failures in the endcaps result from a change in the geometry files, namely, the removal of the ECAL Pre-shower. Due to less material budget in the endcaps, more energy is deposited in this region of the calorimetry, resulting in more calotowers being reconstructed per event.

As part of the validation process, it is desirable to analyze the $\not\!\!E_T$ content of a healthy sampling of Standard Model processes, as well as processes that may contain BSM scale $\not\!\!E_T$. In the past, the $\not\!\!E_T$ Validation efforts have been limited to two QCD dijet samples (one low and high $\hat{p}_T$ range) and one $t\bar{t}$ sample. Now, one low and high p_{T} W+jets sample and a SUSY benchmark

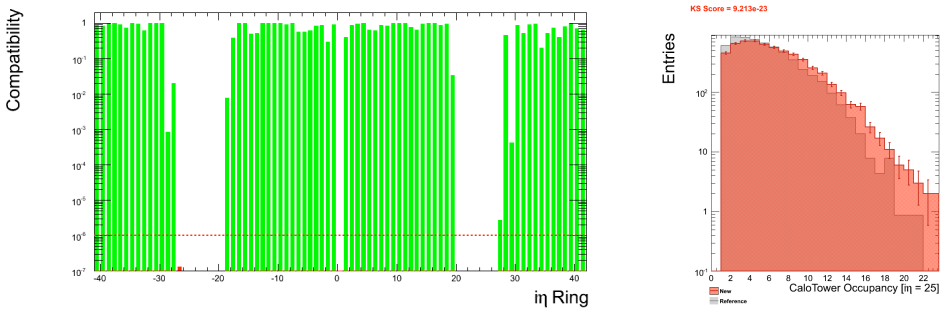


Figure 13: (a) A summary of compatibilty scores for CaloTower occupancy distributions as a function of iη and (b) a particular CaloTower occupancy distribution for iη = 25.

point have been added to the assortment for validation purposes.

5.2 $\not\!\!E_T$ in PAT

The *Physics Analysis Toolkit* (PAT) is a high-level analysis layer in CMSSW enabling the development of common analysis efforts across and within Physics Analysis Groups. It aims at providing both easy-of-use for the beginner and flexibility for the advanced user. Specifically for the $\not\!\!E_T$ object in PAT, the goals are

- to provide a fully configurable chain of sequential corrections to the $\not\!\!E_T$, with reasonable defaults;
- to provide an easy mechanism to retreive the $\not\!\!E_T$ object object after the various stages of correction.

There are three PAT "layers" with each a specific functionality: Layer 0 performs generic post-AOD selection tasks ("cleaning") on the standard reconstructed object, computes related information (such as Monte Carlo and trigger matching) and takes care of the low-level handling of references. Layer 1 collapses the output of Layer 0 into compact objects which contain all necessary information and are easily accessable. Layer 2 performs event-hypothesis dependent tasks and is therefore specific to each analysis.

By default, the Layer 1 $\not\!\!E_T$ object in PAT provides the $\not\!\!E_T$ quantities (a.o $\not\!\!E_T$, $\not\!\!E_x$, $\not\!\!E_y$, $\phi(\vec{\not\!\!E}_T)$, ΣE_T, and $\not\!\!E_T$ Significance) after Type-1 JES corrections (`corMetType1Icone5` by default) and muon corrections (`goodMuonsforMETCorrection` by default, as discussed before). Further optional corrections include tau corrections and Type-2 corrections.

It is possible to easily access the uncorrected or partially corrected $\not\!\!E_T$. Technically, the Layer 1 `pat::MET` object is a wrapper around the `reco::CaloMET` object. Therefore, all methods of the latter are accessable through the `pat::MET`. The `pat::MET` also provides a simplified access to the uncorrected or partially corrected $\not\!\!E_T$ using methods with arguments `UncorrectionType uncorType`. All available `UncorrectionType` values are:

1. `uncorrALL = 0` is used to get a raw $\not\!\!E_T$, prior to any recorded corrections
2. `uncorrJES = 1` is used to uncorrect the first correction, which is the Jet Energy Scale correction.
3. `uncorrMUON = 2` is used to uncorrect the second correction, which is the correction for muons.

4. `uncorrMAXN = 3` is the total number of uncorrections defined. Currently (as listed above) these are `uncorrALL`, `uncorrJES`, and `uncorrMUON`.

The most useful uncorrection methods of the `pat::MET` are expected to be `uncorrectedPt( uncorType)` and `uncorrectedPhi( uncorType)`, where `uncorType`, if not set, defaults to `uncorrAll`.

Updated information on the $\not\!\!E_T$ object in PAT can be found on
`https://twiki.cern.ch/twiki/bin/view/CMS/SWGuidePatLayer1#MET`

6 $\not\!\!E_T$ Significance

Missing transverse momentum may arise from many sources, some of which, like instrumental defects and measurement resolution, represent unwanted contributions that we would like to suppress. In this section we discuss a significance variable, $\mathcal{S}$, which assesses on an event by event basis the likelihood that the observed $\not\!\!E_T$ is a fluctuation from zero arising from finite measurement resolution. $\mathcal{S}$ provides information beyond the raw value of $\not\!\!E_T$ that the user may deploy to improve S/B in some physics cases.

6.1 Definition

The determination of missing transverse momentum involves computation of the vector sum of selected $\vec{\not\!\!E}_T$ vectors, and its significance requires evaluation of the total uncertainty associated with that sum. Using the notation $\vec{E}_{\rm T} = (\varepsilon_x, \varepsilon_y)$ to indicate a generic $\vec{E}_T$ vector, we will characterize the measurement uncertainty associated with $\vec{E}_{\rm T}$ by a function $\mathcal{L}(\vec{E}_{\rm T})$, which defines likelihood contours in the $(\varepsilon_x, \varepsilon_y)$ plane. To find the likelihood function associated with the sum over many $\vec{E}_{{\rm T}i}$, we consider first the case of two vectors, $_1$ and $\vec{E}_{\rm T2}$; the resultant $\vec{E}_{\rm T} = \vec{E}_{\rm T1} + \vec{E}_{\rm T2}$ has a likelihood distribution given by:

$$\mathcal{L}(\vec{E}_{\rm T}) = \int \mathcal{L}_1(\vec{E}_{\rm T1})\mathcal{L}_2(\vec{E}_{\rm T2})\delta(\vec{E}_{\rm T} - \vec{E}_{\rm T1} - \vec{E}_{\rm T2})\, d\vec{E}_{\rm T1}\, d\vec{E}_{\rm T2}. \tag{7}$$

For an arbitrary number of input vectors the final likelihood is obtained by applying Eq. (7) recursively.

This formulation in terms of likelihoods is completely general and will accomodate any likelihood distributions $\mathcal{L}(\vec{E}_{{\rm T}i})$, but in practice we can often assume gaussian errors for measured quantities – in which case the integrals of Eq. (7) may be done analytically. Given a 2×2 covariance matrix $\mathbf{V}_i$ describing the measurement uncertainties in $\vec{E}_{{\rm T}i}$, a gaussian likelihood distribution is given by[2]

$$\mathcal{L}_i(\vec{E}_{{\rm T}i}) \sim \exp(-\frac{1}{2}\vec{E}_{{\rm T}i}{}^T \mathbf{V}_i^{-1} \vec{E}_{{\rm T}i})$$

and the integration of Eq. 7 yields

$$\mathcal{L}(\vec{E}_{\rm T}) \sim \exp(-\frac{1}{2}(\vec{E}_{\rm T} - \vec{E}_{\rm T1} - \vec{E}_{\rm T2})^T \mathbf{V}^{-1} (\vec{E}_{\rm T} - \vec{E}_{\rm T1} - \vec{E}_{\rm T2}))$$

[2]The superscript T indicates the transpose of a matrix or column vector; subscript T indicates "transverse", as in "transverse energy".

with $\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2$. This is just ordinary propagation of errors. For many contributing vectors the expression becomes

$$\mathcal{L}(\vec{E}_{\mathrm{T}}) \sim \exp(-\frac{1}{2}(\vec{E}_{\mathrm{T}} - \sum_i \vec{E}_{\mathrm{T}i})^T (\sum_i \mathbf{V}_i)^{-1} (\vec{E}_{\mathrm{T}} - \sum_i \vec{E}_{\mathrm{T}i}). \tag{8}$$

In the case of $\not\!\!E_T$ significance, the sums are taken over detector elements (such as calotowers) or higher level objects (such as jets). Each $\vec{E}_{\mathrm{T}i}$ is the $-\vec{\not\!\!E}_T$ of the i^{th} element or object, and $\mathbf{V}_i$ its corresponding covariance matrix. Each covariance matrix in turn is most easily specified in a natural coordinate system having one axis aligned with the object's $\vec{E}_T$ vector, $\vec{E}_T \equiv (E_T \cos\phi,\ E_T \sin\phi)$:

$$\mathbf{U}_i = \begin{pmatrix} \sigma^2_{E_{Ti}} & 0 \\ 0 & E^2_{Ti}\,\sigma^2_{\phi_i} \end{pmatrix}. \tag{9}$$

Once specified in this way, the matrix must be rotated into the standard CMS xy reference frame with a rotation matrix $R(\phi_i)$,

$$\mathbf{V}_i = R(\phi_i)\mathbf{U}_i R^{-1}(\phi_i) \tag{10}$$

before summing with others.

We then define the significance as the log-likelihood ratio,

$$\mathcal{S} \equiv -2\ln\left(\frac{\mathcal{L}(\vec{E}_{\mathrm{T}} = \vec{\not\!\!E}_T^{\,\mathrm{observed}})}{\mathcal{L}(\vec{E}_{\mathrm{T}} = 0)}\right), \tag{11}$$

which compares the likelihood of the observed $\vec{\not\!\!E}_T$ with the likelihood of the null hypothesis $\vec{\not\!\!E}_T = 0$. Combining Eqs. 8, 10, and 11 we obtain

$$\mathcal{S} = (\sum_i \vec{E}_{\mathrm{T}i})^T \left(\sum_i R(\phi_i)\mathbf{U}_i R^{-1}(\phi_i)\right)^{-1} (\sum_i \vec{E}_{\mathrm{T}i}). \tag{12}$$

Note that

$$\vec{\not\!\!E}_T = -\sum_i \vec{E}_{\mathrm{T}i}.$$

Equation 12 is the basis upon which we build the significance algorithm. In general $\mathcal{S}$ will be small when the $\not\!\!E_T$ can be attributed to measurement resolution, and large otherwise. The value of $\mathcal{S}$ depends on resolutions which are presently not well known, but it may still be used effectively as a discriminant to distinguish hypothesis pairs ("signal" and "background") in appropriate cases. The discrimination power is something the user should evaluate for his or her own analysis.

Two comments are in order:

- It is useful to realize that in the gaussian case $\mathcal{S}$ is nothing more than a χ^2 value. In fact if we choose to work in a coordinate system with the x axis aligned with the $\vec{\not\!\!E}_T$ axis, instead of the CMS horizontal axis, then Eq. 12 simplifies to a one-dimensional statement, $\mathcal{S} = E^2_T/\sigma^2_{E_T}$. This clarifies the essential meaning of $\mathcal{S}$, but it tends also to obscure an important fact – namely that through its denominator, $\mathcal{S}$ embodies

the full topological information in the event. Essential features such as the angles between the measured MET vector and the various jets and unclustered calotowers in the event are embedded in the definition of $\sigma^2_{E_T}$. This fact is more evident in the representation of Eq. 12.

- The specialization to gaussian likelihood functions adopted here is somewhat less restrictive than it may appear. Eq. (7) is linear in the likelihoods and consequently any likelihood that can be expressed as linear combination of gaussians is easily handled. This is the situation commonly seen, as for example in cases where resolutions are parametrized by combinations of "narrow" and "wide" gaussians. Asymmetric distributions can also sometimes be handled in this way, using displaced gaussians. In any case Eq. (7) remains valid for arbitrary distributions, even if numerical integration may be required in difficult cases.

6.2 Implementation

To apply Eq. (12) one must specify the domain of the sums; different flavors of significance can be constructed by choosing different sets of objects to sum over. For succinctness, we restrict ourselves here to discussing MHT Significance, $\mathcal{S}_H$, in which the sums are taken over jets, electrons, and muons. Another obvious possibility is $\not\!\!E_T$ Significance, $\mathcal{S}_E$, in which one sums directly over calotowers. $\mathcal{S}_E$ carries additional information embedded in the distribution of unclustered energy in the event, and in the long run may prove to be the more useful formulation, but $\mathcal{S}_H$ has the advantage of relying only on well-studied objects whose resolutions are known. In Monte Carlo studies (of CSA07 vintage) $\mathcal{S}_E$ is somewhat more effective than $\mathcal{S}_H$. Nevertheless, unless specifically noted otherwise, all further discussion is about $\mathcal{S}_H$.

In the current implementation the algorithm sums jets, electrons, and muons. We require the jets to pass loose criteria, $p_T > 20\,\text{GeV}$, $\eta < 5$, and $EMF < 0.9$, where EMF is the electromagnetic fraction in the jet, the ratio of ecal to hcal energy deposition. Electrons must satisfy $p_T > 10\,\text{GeV}$, $\eta < 3$, and muons, $p_T > 10\text{GeV}$, $\eta < 2.5$. These criteria are settable[3] and may be subject to future optimization, as well as individual user control. The algorithm uses `SelectedLayer1Jets`, `SelectedLayer1Electrons`, and `SelectedLayer1Muons` as its input so the user may impose additional and/or more restrictive criteria through the PAT Layer1 selection.

In the implementation in CMSSW 1.6.12 reported here, we parametrize the jet resolutions with an expansion in $1/\sqrt{E_T}$ of the reconstructed jet,

$$\left(\frac{\sigma_{E_T}}{E_T}\right)^2 \;=\; \left(\frac{a}{E_T}\right)^2 + \left(\frac{b}{\sqrt{E_T}}\right)^2 + c^2 \tag{13}$$

with $a = 5.6$, $b = 1.25$, and $c = 0.033$ as given in the Physics TDR (Ch. 11.4). The units are GeV. Similarly the jet angular resolution σ_ϕ is taken to have the same form, with $a = 4.75$, $b = 0.426$, and $c = 0.023$. These values are set in the `mhtProducer.cfi` file and enter the formalism here through the covariance matrix $\mathbf{U}_i$ of Eq. (9). For each jet, the parameterization of Eq. 13 is evaluated at the measured value of the jet E_T and is then taken to be constant, *i.e.*, it is not a variable in the integration of Eq. 7. The entries of the total covariance matrix $\mathbf{V} \equiv \sum_i R(\phi_i)\mathbf{U}_i R^{-1}(\phi_i)$ are usually dominated by the jet resolutions, which go roughly as $\sigma_{E_T} \sim 1.25\sqrt{E_T}$ per jet, while the contributions from electrons and muons, characterized by $\sigma_{p_T} \sim 0.02 p_T$, are significantly smaller. We therefore treat the lepton resolutions as negligible (*i.e.*, zero) for now. This is approximation cannot be justified at large p_T and/or η, and will need

[3] `/CMSSW/PhysicsTools/PatAlgos/data/producersLayer1/mhtProducer.cfi`

to be improved upon in the future. Note that this *only* applies to resolutions used in computing the covariance matrix, and *not* to the lepton $\vec{E}_{\mathrm{T}}$ vectors entering the $\vec{\not\!\!E}_T$ sum!

6.3 Performance

To illustrate the use of the significance variable, we explore its performance in two quite different physics cases: $W \to e\nu$ and SUSY (LM1, all hadronic). In the first case the typical $\not\!\!E_T$ is around 40 GeV; while in the latter case it is generally above 200 GeV.

For the $W \to e\nu$ analysis we follow the strategy similar to that laid out in CMS Note 2007/026, and for LM1 we follow CR2007/053. $W \to e\nu$ must first pass the single isolated electron HLT trigger, and we then require GSF electrons of $p_T > 20\,\mathrm{GeV}$ with mild isolation and ID requirements. In addition, for the purpose of exploring MHT Significance, we require the event to contain at least one jet because $\mathcal{S}_H$ isn't defined otherwise. For LM1 there must be three or more jets ($p_{T1,2,3} > 180,\ 110,\ 30\,\mathrm{GeV}$). The analysis cuts are deliberately kept simple to illustrate generic characteristics. Signal samples are CSA07, background samples are CSA07 Gumbo and Chowder:

```
/Wenu/CMSSW_1_6_7-ReRecoIdeal-1198082363/RECO
/CSA07Electron/CMSSW_1_6_7-CSA07-Tier0-A1-Gumbo/RECO
/LM1_sftsdkpyt/CMSSW_1_6_7-CSA07-1200560744/RECO
/CSA07JetMET/CMSSW_1_6_7-CSA07-Tier0-A1-Chowder/RECO
```

In each of the two physics cases we examine the distributions for $\mathcal{S}_H$ and $\not\!\!H_T$, and construct plots of signal efficiency versus background efficiency for two different analysis strategies: cutting only on $\not\!\!H_T$, or cutting only on $\mathcal{S}_H$ (above and beyond the cuts noted above). The results are shown in the four panels of Fig. 14. For $W \to e\nu$ the scatterplot of $\mathcal{S}_H$ versus $\not\!\!H_T$ in panel (a) indicates that $W \to e\nu$ signal events exhibit large $\mathcal{S}_H$ at the expected $\not\!\!H_T \sim 40\,\mathrm{GeV}$, and correspondingly the signal efficiency versus background efficiency shown in panel (b) confirms that the $\mathcal{S}_H$ cut is the (slightly) more favorable choice. The non-smoothness in the efficiency curves in panel (b) is due to the use of weighted events in the background sample. For LM1, the scatter plot in panel (c) reveals that $\mathcal{S}_H$ will have little impact on the analysis as $\not\!\!H_T$ is by itself an effective variable and significance has little to add. The efficiency plot in panel (d) shows that in fact the $\mathcal{S}_H$ cut is less useful than the direct cut on $\not\!\!H_T$ for this mode.

7 $\not\!\!E_T$ Data Quality Monitoring and Cleanup

We describe the plans for the $\not\!\!E_T$ Data Quality Monitoring and for the treatment of dead and hot calorimeter cells. We also discuss the $\not\!\!E_T$ distribution due to calorimeter noise as observed in the CMS Global Runs.

7.1 $\not\!\!E_T$ Data Quality Monitoring

In this section, we discuss the data quality monitoring (DQM) for missing E_T which has been developed as a part of JetMET DQM.

At CMS, DQM consists of (1) online DQM and (2) offline DQM. Online DQM runs right after data taking with limited statistics, and monitors detector-level quantities. Offline DQM runs on the full data statistics as part of standard reconstruction at Tier-0 ("prompt reconstruction") and Tier-1 ("re-reconstruction"). It confirms the findings of online DQM, and in addition it monitors the reconstructed physics objects such as photons, electrons, muons, jets and missing

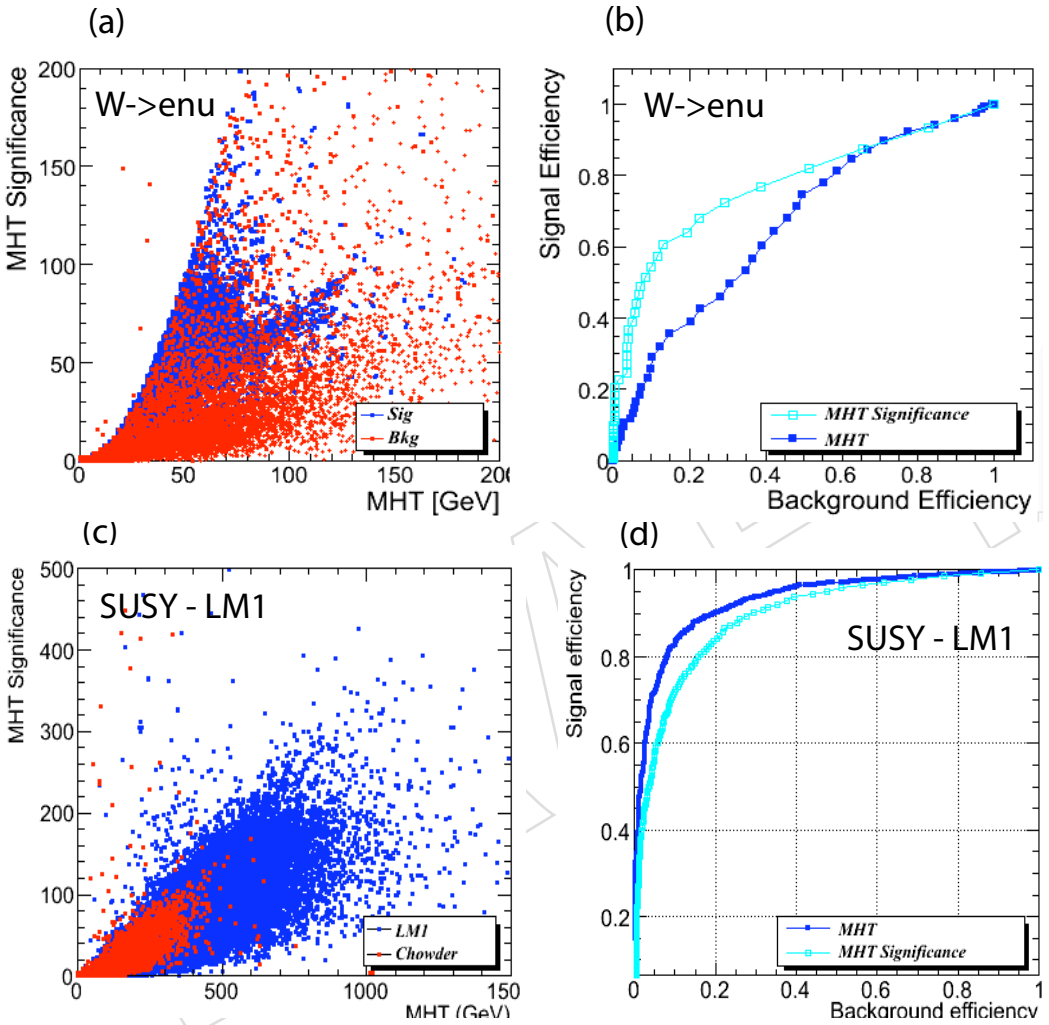


Figure 14: (a) $\mathcal{S}_H$ versus $\not{H}_T$ for $W \to e\nu$; (b) signal efficiency versus background efficiency for $W \to e\nu$; (c) $\mathcal{S}_H$ versus $\not{H}_T$ for LM1; (d) signal efficiency versus background efficiency for LM1.

E_T which the online DQM will not monitor in detail. In offline DQM, we also check for cross-detector effects, e.g., electron E_T from the electromagnetic calorimeter compared with electron p_T from the tracking system; this is not done by the online DQM. The work flow of the JetMET DQM is shown in Fig. 15, which is a part of the offline DQM. Details of each step in this work

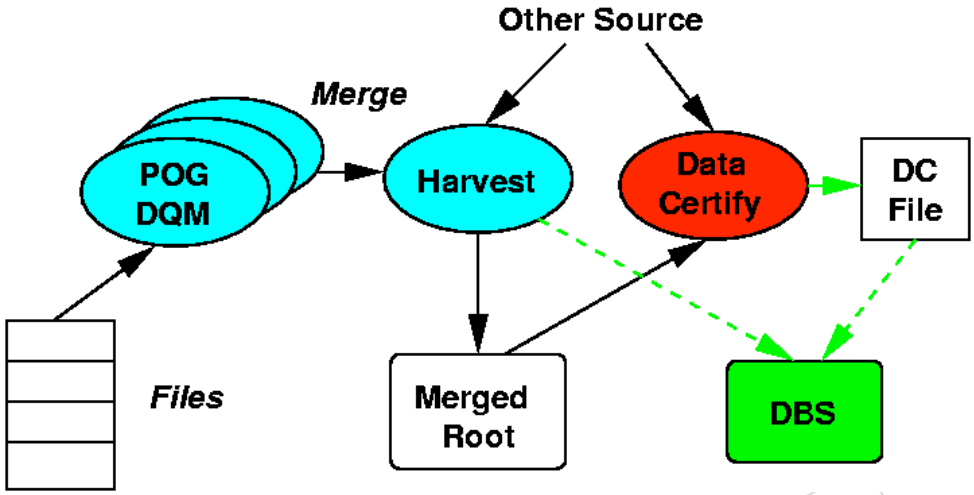


Figure 15: Work flow of the JetMET DQM.

flow are discussed below.

The goal of the missing E_T DQM is to identify good runs and luminosity sections [4] for use by physics analyses that relies on missing E_T related observables. This will provide data quality information for the Barrel, Endcap and Forward calorimeters, and will complement the detector-level quality monitoring for these calorimeters.

The code for missing E_T DQM is implemented in CMSSW under `DQMOffline/JetMET`. The DQM for $\not\!\!E_T$ is quite advanced among the physics object groups, in that the first version of the monitoring code which produce the histograms necessary for data quality evaluation is already integrated in the CMS DQM framework and tested with the CRAFT data reconstruction. Reconstruction run in many parallel jobs, and harvesting will merge outputs from those parallel jobs and produce a single root file for each run containing a set of histograms for monitoring. The resulting histograms for each run from the tier-0 prompt reconstruction and tier-1 re-reconstruction are available on the tier-0 monitoring web pages [5].

Those histograms include distributions for $\not\!\!E_x$, $\not\!\!E_y$, $\not\!\!E_T$, METPhi, $\sum E_T$, etc from `met` and `metNoHF`. In order to evaluate the data quality for each luminosity section, the 2-dimensional histograms of $\not\!\!E_x(\not\!\!E_y)$ versus luminosity section are also made. The inclusion of the same set of histograms from `metHO` and `metNoHFHO` is under consideration.

Fig. 16 shows distributions of $\not\!\!E_x$, $\not\!\!E_y$ and METPhi for two CRAFT runs; one run is a typical "good" run and the other run is a typical "bad" run. In good runs, it is expected that the $\not\!\!E_x$ and $\not\!\!E_y$ distributions are symmetric and peaked at zero, and the METPhi distribution is smooth. Such feature is observed in the top plots in Fig. 16. The *sinus* structure in the METPhi distribution is presumed to be due to cosmic ray muons coming through the shaft and hitting the CMS detector. In bad runs with a hot area in the calorimeters, the $\not\!\!E_x$ and $\not\!\!E_y$ distributions peak at a non-zero value, and the METPhi is peaked at some value corresponding to 180° away

[4] One luminosity section is about 90 seconds.

[5] `https://cmsweb.cern.ch/dqm/tier-0/` and `https://cmsweb.cern.ch/dqm/tier-1/`

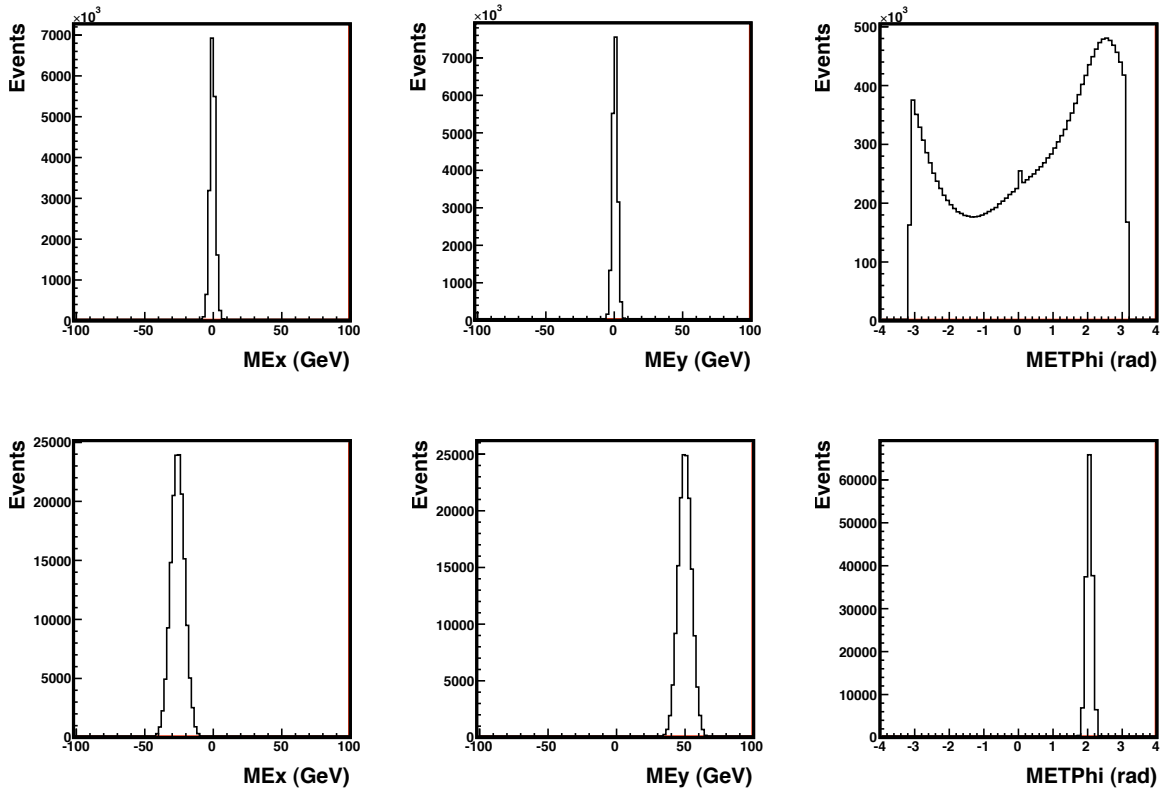


Figure 16: $\not\!\!E_x$, $\not\!\!E_y$, and METPhi distributions for a typical "good" run (top, run=68021) and a "bad" run with a hot tower(s) (bottom, run=67124).

from the hot spot. Such feature is seen in the bottom plots in Fig. 16.

These histograms have to be evaluated by a "data certification" (DC) algorithm to determine the goodness of data. The first version of the algorithm is already implemented as a standalone code, and it is being integrated into CMSSW_2_2_X. The version being integrated into CMSSW_2_2_X simply determines the goodness of data based on mean and RMS of the $\not\!\!E_x$ and $\not\!\!E_y$ distributions. More sophisticated algorithm is being prepared which makes use of histograms other than the $\not\!\!E_x$ and $\not\!\!E_y$ distributions and makes comparisons with the reference histograms using the χ^2 and Kolmogorov-Smirnov tests. We are waiting for the central DQM system to support this scheme that uses comparisons with reference histograms.

In the central DQM system, this DC step will be performed in the harvesting step. In the harvesting step, we can incorporate information from information from "other source" such as DCS (detector control system) and DAQ. Finally, the outputs of the data certification will be written to the DC file and DBS (database).

We are working toward making the monitoring code fully functional by next summer, when we expect to have the first proton-proton collision data.

7.2 Bad Cell Treatment

The reconstruction code will need to be able to take into account dead or noisy cells caused by detector problems. The CaloTower object consists of multiple cells. Jets and $\not\!\!E_T$ objects are formed from a collection of CaloTowers. The user will need to be able to identify how many

Bad Cells are contained in reconstructed objects such as Jets and $\not\!\!E_T$. Bad Cells can be the result of malfunctioning hardware and can lead to no energy being read out, incorrect energy, or to excessive noise. Over time, additional hardware failures may occur while some hardware failures may be fixed and it will be essential to keep track of the current list of Bad Cells in the database. It will also be important to simulate the changing conditions.

The fine granularity of the ECAL allows the possibility of estimating the dead cell energy by using information from adjacent cells. Different algorithms to estimate the energy can be used and we will need to study the impact of the particular algorithm on the reconstructed object. For HCAL it is not anticipated that a recovery algorithm will be used since the cells are too large to make this feasible.

The software should be able to accommodate the following cases:

- Case 1: Understanding the $\not\!\!E_T$ tails and events with high $\not\!\!E_T$.
  Check if the CaloTowers used to construct the $\not\!\!E_T$ object contain hot, dead, recovered cells, or anomalous cells.
- Case 2: Bad Cell information is used to define $\not\!\!E_T$ Significance.
  $\not\!\!H_T$ is calculated in PAT after reconstruction. Need to know the number of bad cells and the type that are included in the $\not\!\!E_T$.
- Case 3: Detailed studies of the recovery algorithms.
  Since the recovery algorithms may depend on information that has been dropped from the AOD it will be necessary to go back the RECO data in order to test different recovery algorithms.

Several algorithms, developed for the online and offline Data Quality Monitoring, are available to identify Bad Cells. In addition to Bad Cells that are persistent for the entire run we have observed a significant rate of intermittent noise. The noise has distinct characteristics and comes from discharge of the Hybrid Photo Diode (HPD), noise associated with the Readout Boxes (RBX), and observed nuclear effects in the HF. In addition, there are problems that lead to the wrong energy being read out. Cells with intermittent problems are referred to as Anomalous Cells. A new word was added to the rechit object in order to save information used to classify the event. This classification will be used by the reconstruction code to determine how the cell is treated when reconstructing physics objects. The CaloTower has a status word which will be used to count the number of bad cells belonging to each category and this information will be available to the user.

The overlap of the intermittent noise with collision data is expected to be small and we can either 1) mark the Anomalous Cells on an event by event basis or 2) develop filters to identify events with noise overlap and reject the event. If we chose to use option 1, we will have to run a anomalous event algorithm during reconstruction. The result of the algorithm would be saved in the event record and not be saved in the database. We would have to track the version of the algorithm. Some algorithms to identify intermittent noise use detailed information that is only available at the digi level and is later dropped from the event saved in the AOD.

The different categories of problems are:

- "Bad"
  The cell is off, dead/hot (as determined by DQM), or the energy is reconstructed unreliably. This category includes all cases when the RecHit energy was not added to the CaloTower energy because it is unavailable or grossly mismeasured. Some problematic cells will be put in this category for CaloTower accounting purposes,

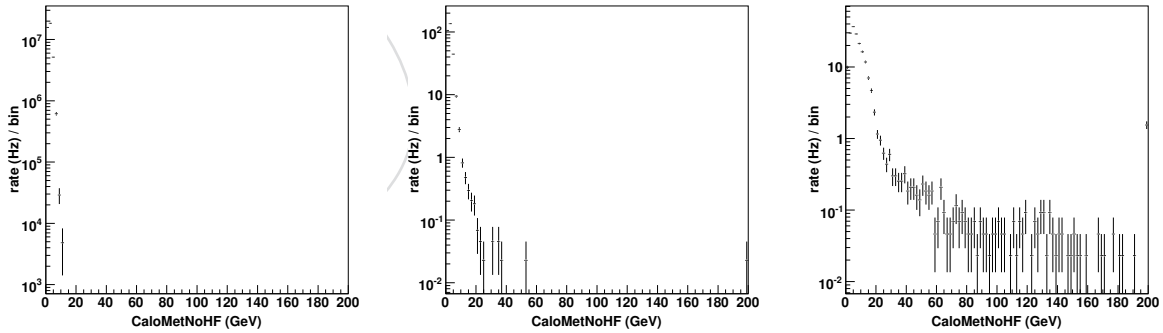
based on the severity of the problem.

- "Recovered"
  Cells where the energy was not taken from the recorded information, but "recovered" using some algorithm. Example is the recovery performed by ECAL based on information of neighboring cells. The assigned energy is used when calculating the energy in the CaloTower. If a cell can otherwise be classified as bad (has severe problems), but a "recovery" was performed AND recovered cells are set to be used, it will be removed from the Bad cell count and counted here. By definition "recovered" cells have some problem, but they will not be classified as "problematic".
- "Problematic"
  Cell/RecHit has some issues, but the severity of the problem is not high-enough to warrant its rejection. It is still flagged for detailed studies by the JetMEt or physics groups. It contains the original energy, not a "recovered" energy.

Details of how the cells will be included in CaloTower will be specified in the configuration file in order to allow flexibility for studies before fixing the default behavior.

7.3 $\not\!\!E_T$ in Global Runs

Data from CMS Global Runs enable us to study $\not\!\!E_T$ measurement with the actual CMS detector. We look at $\not\!\!E_T$ in CRAFT run 67141, during which the CMS magnet was operated at full magnetic field (3.8 T). We consider three different level one triggers: random, cosmic muon, and calorimeter noise. The cosmic muon trigger was provided by the muon subsystems. The predominant contribution to the calorimeter trigger was noise in the HCAL barrel and endcaps (HBHE). The trigger was fired when and only when the sum of the (total) energy measured in any two neighboring HBHE towers exceeded ∼10 GeV. Figure 7.3 shows the distributions of $\not\!\!E_T$ for these three triggers. Table 7.3 gives the means and standard deviations of the $\not\!\!E_T$ components in these samples. We reach two conclusions: 1) the standard deviation of $\not\!\!E_T$ when CMS is quiet is ∼2 GeV, and 2) the rate of high-$\not\!\!E_T$ noise in HBHE is ∼2 Hz.



(a) random trigger (b) cosmic muon trigger (c) HBHE noise trigger

Figure 17: Absolutely normalized $\not\!\!E_T$ distributions for three different triggers in CRAFT run 67141. Overflows are included in the rightmost visible bins.

In Figures 7.3 and 18 two structures emerge: (1) real $\not\!\!E_T$ due to muon energy deposits in the calorimeter ($\not\!\!E_T < 20$ GeV) and (2) "fake" $\not\!\!E_T$ residing in the tails of the distributions ($\not\!\!E_T > 50$ GeV). It has been shown that the main effects that contribute to "fake" high $\not\!\!E_T$ events are due to HCAL noise, where electronics failure in a HPD or readout box produces large rechit energies, and EB abnormal events, where noise in ECAL produces extremely large rechit energies.

trigger	$\mu(\not{E}_x)$	$\mu(\not{E}_y)$	$\sigma(\not{E}_x)$	$\sigma(\not{E}_y)$
random	-1.0	0.8	1.9	1.7
cosmic muon	-1.0	0.8	2.5	2.1
HBHE noise	0.0	2.2	6.5	8.6

Table 2: Means and standard deviations (in GeV) of $\not{E}_T$ components in CRAFT run 67141.

Since $\not{E}_T$ will be a main discriminator for physics searches beyond the standard model, it is extremely important to categorize and identify the different types of high $\not{E}_T$ events so that one can gain some insight on how to remove or "clean" these events. More specifically, one of the main goals is to use the digi level properties of "noisy" events to develop algorithms that can be used to tag these anomalous events for the purposes of (1) supressing them at the trigger level and (2) "cleaning" events where real signal events due to pp collisions coincide or overlap with noise. Therefore, a short term goal is to develop a path to collect RAW data of potentially anomalous events. These potentially anomalous events can then be used to study their digi level properties. However, in choosing a specific path, one must keep in mind that anomalous events should be selected such that the digi level properties are unbiased by the requirements used in selecting them. Thus, we propose the following $\not{E}_T$ triggers for monitoring purposes:

(1) $\not{E}_T > 10$ with PreScale=500
(2) $\not{E}_T > 20$ with PreScale=100
(3) $\not{E}_T > 80$ with PreScale=35
(4) $\not{E}_T > 180$ with PreScale=2

The main objectives are to try to exploit the two structures, achieve an acceptable overall rate, and obtain an unbiased high $\not{E}_T$ sample ($\not{E}_T > 100$ or 200 GeV). If a situation occurs where the noise rate increases, then we can impose a higher prescale on the lower $\not{E}_T$ triggers.

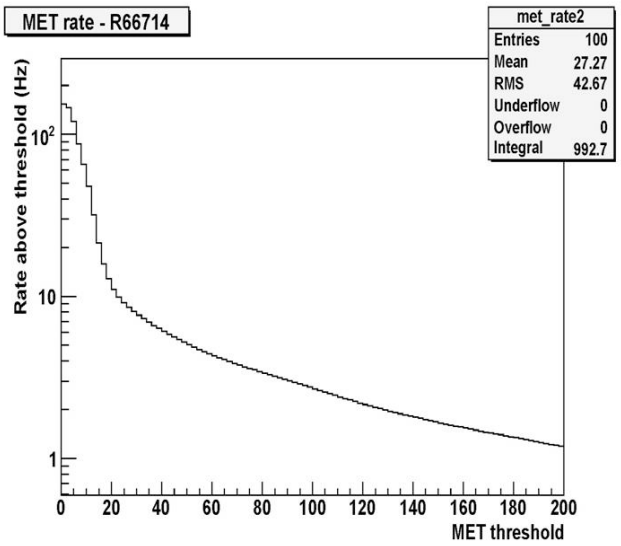


Figure 18: The event rate vs. $\not{E}_T$ threshold for CRAFT Run 66714 data.

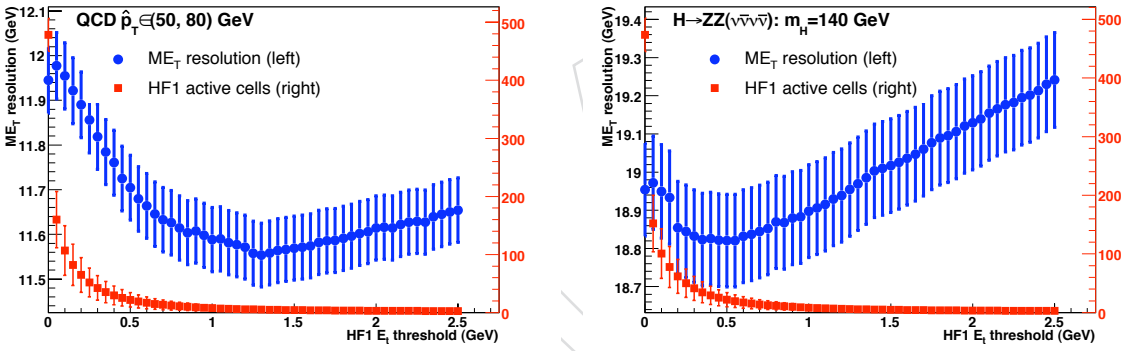
8 Impact of HF on $\not\!\!E_T$ resolution

Here, we study how HF impacts the $\not\!\!E_T$ resolution. The MC studies are based on CSA07 samples or the equivalent.

Figure 19(a) shows the $\not\!\!E_T$ resolution in QCD Monte Carlo events as a function of cell transverse energy (E_T) thresholds applied to the HF long fiber cells (see [1] for a definition of the $\not\!\!E_T$ resolution). The analogous results for Higgs events produced through vector-boson-fusion (VBF), with the subsequent invisible decay of the Higgs, are shown in Fig. 19(b). The $\not\!\!E_T$ resolution is minimized with an E_T threshold around 1.3 GeV for the QCD events and 0.5 GeV for the Higgs events. The overall estimate for the optimal threshold is 0.7 GeV. A list of optimized cell E_T thresholds for all HCAL sub-detectors is given in Table 8. The corresponding plots based on cell energy thresholds, rather than E_T thresholds, have the same shape as Figs. 19(a)-19(b) but are stretched out by a factor of about 40, i.e., the minimum in $\not\!\!E_T$ resolution occurs for a cell energy threshold around 50 GeV (20 GeV) for the QCD (Higgs) events.

Table 3: Optimized cell E_T thresholds for HCAL sub-detectors.

sub-detector	HB	HO	HE (5°)	HE (10°)	HF (long)	HF (short)
E_T threshold (GeV)	0.4	0.5	0.3	0.4	0.7	1.0



(a) QCD events.

(b) VBF Higgs events with invisible decay of the Higgs.

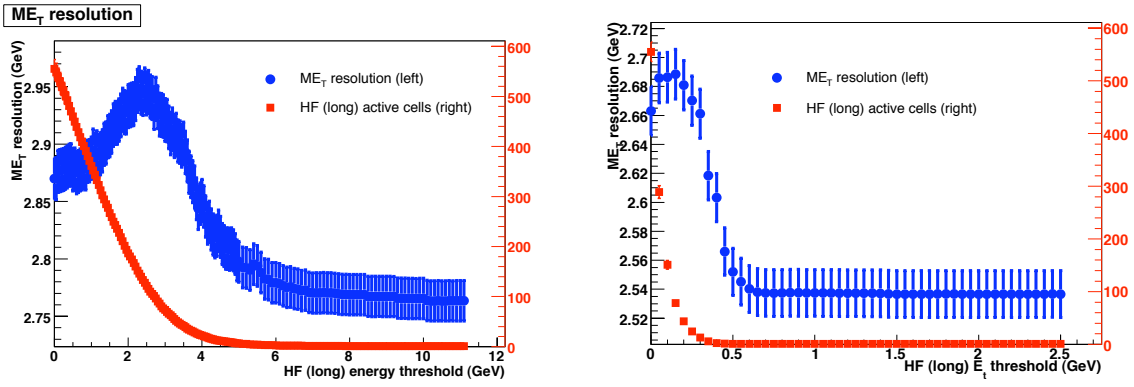
Figure 19: $\not\!\!E_T$ resolution versus HCAL cell E_T threshold in the HF long fibers.

Columns (1) and (2) in Table 4 compare the $\not\!\!E_T$ resolutions obtained with Scheme-B cell energy thresholds and optimized cell E_T thresholds for representative physics samples. The Scheme-B thresholds are the current default in CMSSW. The optimized E_T thresholds yield $\not\!\!E_T$ resolutions that are 3-13% smaller than the Scheme-B thresholds, depending on the process.

The $\not\!\!E_T$ resolution observed in Global Run cosmic data is shown in Figs. 20(a) and 20(b) as a function of cell energy and E_T thresholds in the HF long fibers. From the perspective of $\not\!\!E_T$, these data represent noise runs. The energy threshold required to eliminate the noise is seen to be around 6 GeV (Fig. 20(a)), larger than the Scheme-B threshold of 1.2 GeV while at the same time smaller than the 20-50 GeV required to minimize the $\not\!\!E_T$ resolution in the physics events. In contrast, the noise is eliminated with a cell E_T threshold of about 0.7 GeV (Fig. 20(b)) that is the same as required to minimize $\not\!\!E_T$ resolution in the physics events. Thus HF cell thresholds based on E_T seem to be more sensible than those based on energy. We find similar conclusions for the HE 10° cells.

Table 4: Comparison of the $\not\!\!E_T$ resolution in representative physics samples, with different treatments of HCAL cell thresholds or else different schemes to incorporate HF information.

Thresholds	(1) Scheme B E	(2) Optimized E_T	(3) E_T, exclude HF	(4) E_T, exclude HF, include forward jets
QCD	13.4	11.6	16.8	13.4
W + jets	11.4	10.6	12.2	11.3
VBF Higgs	19.8	18.8	35.1	19.6
SUSY LM1	42.9	41.4	43.8	41.5



(a) Cell threshold in E.

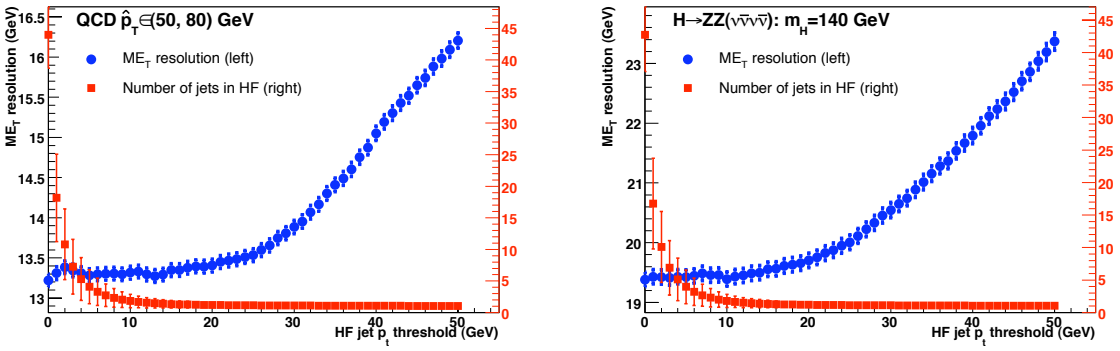
(b) Cell threshold in E_T.

Figure 20: $\not\!\!E_T$ resolution versus HCAL energy or transverse energy cell threshold for cosmic ray data.

The forward energy flow in LHC data, or the noise levels at start-up, could be larger than expected. We therefore consider backup strategies for the incorporation of HF information into $\not\!\!E_T$. As an extreme, column (3) in Table 4 lists the $\not\!\!E_T$ resolutions obtained when HF is excluded from the $\not\!\!E_T$ calculation. These results are based on optimized cell E_T thresholds that are somewhat adjusted compared to the values in Table 8. For physics processes with significant forward energy, such as the QCD and Higgs VBF events, the $\not\!\!E_T$ resolution without HF is seen to be much worsened.

As a less extreme possibility, we investigate the use of jets in the forward region, with HCAL cell information used elsewhere. The jets are defined by the iterative cone algorithm with a cone size of 0.5, and are taken from the standard jet collections. Jets are selected if they have $|\eta| > 2.853$, where the cutoff corresponds to the minimum $|\eta|$ of HF cells. The negative of the sum of the jet transverse momenta assuming massless jets (MHT) is then subtracted from the $\not\!\!E_T$ determined without HF. Contributions from HE cells that fall within the jet cones are subtracted to avoid double counting.

The results for QCD and VBF Higgs events are shown in Figs. 21(a) and 21(b). These figures demonstrate that the $\not\!\!E_T$ resolution is not much affected until the cut on the p_T of the forward jets exceeds 20 GeV. Column (4) of Table 4 gives the $\not\!\!E_T$ resolutions in the physics samples for a 20 GeV cut on the forward jet p_T. For the QCD and Higgs samples, the $\not\!\!E_T$ resolutions are much improved compared to those obtained if no HF information is used at all [column (3)].



(a) QCD events.

(b) VBF Higgs events with invisible decay of the Higgs.

Figure 21: $\not\!\!E_T$ resolution based on jets in the forward region in conjunction with HCAL cell information elsewhere.

9 Missing H_T at Level 1

Current Level 1 (L1) trigger in CMS is capable only of calculating $\not\!\!E_T$ based on the entire calorimeter. Similar to the way $\not\!\!E_T$ is calculated offline, L1 $\not\!\!E_T$ is based on the calorimeter regions (typically, 3×3 calorimeters towers), rather than individual towers. This coarser granularity at L1 does not affect either angular or energy resolution significantly, thus making L1 $\not\!\!E_T$ information similar to that offline.

However, the flip-side of this algorithm is that any instrumental problem with the calorimeter, or its part, would immediately translate in L1 $\not\!\!E_T$, thus making this object and the corresponding trigger term not robust against detector malfunction. While $\not\!\!E_T$ contribution due to possible detector malfunctions can be typically accounted for offline, this is all but impossible at the low trigger level. Consequently, if the L1 $\not\!\!E_T$ trigger rate goes up significantly due to such malfunctions, this trigger would be immediately prescaled, thus making it inefficient. Since $\not\!\!E_T$ is a crucial signature for many searches for new physics, the reduced $\not\!\!E_T$ trigger efficiency would imply reduced sensitivity to new physics signals.

To remedy this obvious deficiency and to make the L1 trigger more robust against detector problems, it has been suggested [4] to use only Level 1 jets, instead of individual calorimeter regions to calculate $\not\!\!H_T$[6]. Such an additional $\not\!\!H_T$ trigger can be used either as a designated trigger, or as a backup for the $\not\!\!E_T$ trigger in case of detector malfunctions.

Using jets, rather than the entire calorimeter, to calculate missing transverse energy has several other advantages. First of all, one could adjust minimum jet E_T threshold for the jet to be counted toward $\not\!\!H_T$ to minimize the effect of pile-up on the missing transverse energy resolution. One can also restrict jets to certain pseudorapidity range, thus eliminating more noisier forward part of the calorimeter. Finally, one could choose whether to use raw jet energies, or to correct them for jet energy scale for $\not\!\!H_T$ calculations.

As a result of this proposal, a designated $\not\!\!H_T$ trigger is currently being developed in CMS. Since both Level 1 jets and Level 1 $\not\!\!E_T$ are formed in the Global Calorimeter Trigger (GCT, [5]), it is the proper part of the trigger to supply $\not\!\!H_T$ information in addition to its current role. The GCT L1 MHT algorithm is currently being implemented and tested by the GCT team and should be

[6]Since energetic electrons and phtons are also reconstructed as jets by Level 1 trigger, this definition is sensitive to jets, electrons, and photons.

available in time for the first LHC collision data.

A brief description of the algorithm is as follows:

1. The Leaf cards send the rank of the jets found to the Wheel cards. Jet rank is in general a monotonic function of the jet transverse energy.

2. The Wheel cards take the rank and derive an azimuthal angle (ϕ) of the jet with 5-bit dynamic range. A look-up table is then used to calculate $\sin(\phi)$ and $\cos(\phi)$ and to compute the x- and y-contributions to $\not{H}_T$ from this jet. The sums over all jets seen by the Wheel cards are sent to the Concentrator card.

3. The Concentrator card sums positive and negative rapidity sums and returns the result as 7 bits of $\not{H}_T$ rank and 5 bits of the azimuthal angle of the $\not{H}_T$ vector.

As jet energy scale corrections in the GCT are currently applied in the Concentrator card, the $\not{H}_T$ won't have this scheme produces $\not{H}_T$ without Type I corrections. It is possible, in principle, to correct jets in the Leaf cards; if this scheme is implemented, the L1 $\not{H}_T$ will be also corrected with Type I corrections. Note that moving jet energy scale corrections to the Leaf card would require significant changes in firmware and is not expected in early running, or at a later time, unless it is shown that such corrections could improve the performance of the $\not{H}_T$ trigger appreciably. However, if partial jet energy scale corrections are implemented at the Regional Calorimeter Trigger (RCT) level, they would automatically propagate into L1 $\not{H}_T$ without any changes to the latter.

Once L1 $\not{H}_T$ is implemented in hardware and GCT emulator, we plan to compare its performance with that for L1 $\not{E}_T$ using both cosmics and first collider data.

10 Conclusions

In this note, the progress in reconstuction, optimization and validation of the Missing Transverse Energy since AN/2007-041 [1] is discussed.

During the last year the $\not{E}_T$ reconstruction software has grown significantly, notably through the addition of several new $\not{E}_T$ collections at the AOD level such as $\not{E}_T$ including/excluding the HF and/or HO detectors, the "track-corrected MET ("tcmet") and the "particle flow MET". Also the $\not{E}_T$ validation suite has been upgraded to include more collections and benchmark samples. The access to the $\not{E}_T$ quantities in the Physics Analysis Toolkit (PAT) is described.

The $\not{E}_T$ correction for muons has been improved by taking into account their small energy deposit in the calorimeter. For isolated muons, the calorimetric deposit is completely removed, while for non-isolated muons a median value of the expected deposit of a single muon is subtracted. A Type-2 correction, taking into account unclustered and out-of-cone energy, has been derived using a template method. It is shown that the bias of Type-1 corrected $\not{E}_T$ is further reduced after applying these corrections.

Fundamental improvements to the calorimeter-based $\not{E}_T$ have been made through the "track-corrected MET" (*tcmet*) in which, for all well reconstructed tracks, the expected energy depostition in the calorimeter is replaced by the measured momentum in the tracker. This leads to reduced $\not{E}_T$ tails and improved resolution. The "particle flow MET" (*pfmet*) goes a step further and exploits both charged and neutral particles that have been reconstructed using the Particle Flow algorithm.

A Data Quality Monitoring (DQM) framework for $\not\!\!E_T$ has been put in place in order to identify good runs and luminosity sections for analyses that rely on the $\not\!\!E_T$ quantity. The DQM is illustrated by comparing good and bad runs in the CRAFT data. The treatment of hot and dead cells is discussed. A $\not\!\!E_T$ significance algorithm has been developed, which assesses on an event by event basis the likelihood that the observed $\not\!\!E_T$ is a fluctuation from zero arising from finite measurement resolution.

Data from the CMS Global Runs (notably CRAFT) has been used to study the actual $\not\!\!E_T$ measurements in the CMS detector, in particular the effect of instrumental noise. $\not\!\!E_T$ monitoring triggers are proposed to study noisy events and cleaning strategies in future runs. The impact of the forward calorimeter (HF) on the $\not\!\!E_T$ resolution has been studied. Finally, the status of the Level-1 $\not\!\!H_T$ trigger was summarized.

Amongst others, the following future developments are planned:

- further study of instrumental effects (HPD noise, ...) and cleaning procedures
- study of $\not\!\!E_T$ in future Global Runs
- impact of punch-through in the calorimeter
- workflows for $\not\!\!E_T$ commissioning and calibration using SM candles

References

[1] S. Esen *et al.*, "Missing E_T Performance in CMS," *CMS Analysis Note* **AN/2007-041** (2007).

[2] M. Mulders *et al.*, "Muon identification in CMS," *CMS Analysis Note* **AN/2008-098** (2007).

[3] F. Golf and A. Yagil, "Track Corrected Missing Transverse Energy in CMS," *CMS Analysis Note* **AN/2009-022** (2009).

[4] G. Landsberg, "Missing H_T Trigger Proposal," *Talk at the CMS TriDAS Meeting* **August 28, 2007** (2007).

[5] G. Heath and U. Schager, "CMS Global Calorimeter Trigger Conceptual Design," *CMS Technical Note* **TN/1995-049** (1995).